

# Assessment of Anharmonicities in Clusters: Developing and Validating a Minimum-Information Partition Function

Roope Halonen

*Center for Joint Quantum Studies and Department of Physics,  
School of Science, Tianjin University, 92 Weijin Road, Tianjin 300072, China\**

(Dated: January 28, 2024)

Precise thermodynamic calculations are essential for understanding the dynamics of cluster systems and new particle formation. However, the widely employed harmonic statistical mechanical approach often falls short in terms of accuracy. In this study, we present an improved statistical model that incorporates vibrational anharmonicity via a novel partition function that requires only one additional system-specific input parameter. In addition to considering vibrational aspects, we also account for anharmonicity related to the configurational space. The role of anharmonicities is thoroughly examined in the case of general clusters, where the complete sets of conformers, mechanically stable spatial arrangements, are known up to clusters composed of 14 monomers. By performing consistent Monte Carlo simulations on these systems, we benchmark the statistical model's efficacy in reproducing key thermodynamic properties (formation free energy and potential energy) in the classical limit. The model exhibits exceptional alignment with simulations, accurately reproducing free energies within a precision of  $2k_{\text{B}}T$  and reliably capturing cluster melting temperatures. Furthermore, we demonstrate the significance and applicability of the model by reproducing thermodynamic barriers in homogeneous gas-phase nucleation of larger clusters. The transferability of our developed approach extends to more complex molecular systems and bears relevance for atmospheric multicomponent clusters, in particular.

## I. INTRODUCTION

Undoubtedly, the key quantity in thermodynamic calculations is the partition function, serving as the foundational basis for deriving virtually every thermodynamic function. When computing a system's thermochemistry through ab initio molecular methods, accurate direct approaches for obtaining partition functions are lacking. Consequently, resorting to approximations based on basic statistical mechanical principles becomes necessary[1, 2]. In essence, this "statistical approach" assumes the uncoupling of translational, rotational, vibrational motions, and electronic excitation, facilitating the separability of the partition function into distinct factors. Importantly, the use of this approach extends beyond quantum chemistry and single molecule computations, finding application in diverse domains such as studies involving atomic and molecular clusters[3–5], colloidal microcrystals[6, 7], biomolecules[3, 8], and supercooled liquids[9].

The possible shortcomings of the statistical approach can be addressed through computer simulations, including methods like Monte Carlo or molecular dynamics, providing a convenient and accurate means to obtain thermodynamic properties. Nevertheless, these methods come with their own set of limitations. Firstly, even with the computation of interatomic forces through electronic structure calculations, the system's dynamics are treated classically[10]. As a consequence, the simulated vibrational properties do not precisely reflect those of a system in the quantum regime. Secondly, although simulations provide valuable quantitative data, they offer limited physical insight into thermodynamic properties and dynamic processes[9, 11]. This underscores the continued relevance of the statistical model for both quantum and classical systems, thereby warranting further development.

For calculations using this model, the initial step involves characterizing the multidimensional energy landscape, which represents the energetic dependencies of the system based on the positions of the considered atoms. While creating an accurate map of the entire topography of this landscape, complete with mountain ranges, valleys, and passes, is often challenging for chemically complex species, available techniques allow for the identification of the most crucial areas. This mapping primarily focuses on local minima and their adjacent saddle points, which serve as the main sources of structural and thermodynamic information. Each minimum, usually referred to as a conformer or an "inherent structure" [9], resides in a basin of attraction, confining the system's thermally agitated vibrational motion within this specific region. Assumed under the ergodicity principle, each basin must be connected to at least one other basin via a finite-height saddle point. Leveraging information about the landscape, the statistical model evaluates

---

\* roope@tju.edu.cn

the partition function for the system by summing up contributions from microstates related to individual minima and the basins surrounding them.

For large systems, numerous conformers with distinct structures and energies exist, typically representing various states and forms of matter (e.g., crystal, liquid, amorphous glass). Despite the multitude of conformers, it is sometimes assumed that the global minimum energy conformer sufficiently represents all inherent structures. This assumption becomes problematic at elevated temperatures, where close-packed crystal-like structures become entropically unfavorable. Another frequently employed approximation involves treating intrabasin vibrational motion as harmonic oscillation. However, as temperature rises, vibrational excursions away from the minimum intensify, leading to increasingly anharmonic motion. Together, these two assumptions constitute the “harmonic approximation”.

Following the terminology from gas-phase cluster chemistry[12–14], the limitations of the harmonic approximation can be mitigated by incorporating *global* and *local* anharmonic corrections. The former involves properly accounting for the presence of different conformers[13, 15–18], while the latter entails adjusting the conformer’s vibrational frequencies[14, 19–21] or the applied partition function[11, 22–25].

In this study, our objective is to shed light on the anharmonic behavior within cluster systems. To achieve this, we first develop a vibrational partition function using minimal amount of information, namely, the interconformer transition state energies in addition to vibrational frequencies. We focus on simple clusters composed of point-like particles due to their relatively weak intermolecular interactions and a high number of conformers. For instance, considering a cluster of 14 sticky-hard spheres, there are 895,478 unique arrangements[26], implicating substantial global anharmonicity at elevated temperatures within this size range of clusters. By confining our analysis to classical systems, we can rigorously validate and benchmark the developed anharmonic statistical model against Monte Carlo simulations.

Our arguments are related to nucleation phenomena and cluster free energies in particular, reminiscent of early elementary cluster studies[4, 15–18], while also remaining relevant to present-day theoretical investigations into atmospheric new particle formation[5, 13, 20].

## II. STATISTICAL CLUSTER THERMODYNAMICS

The standard statistical approach centers around individual cluster configurations which correspond to mechanical equilibria on the energy landscape. These conformers are indexed by  $k = 0, 1, 2, \dots$ , with  $k = 0$  representing the conformer with the lowest known binding energy. In this study, we only consider clusters comprising  $N$  identical monomers and their respective conformers, and thus, the relevant partition function is written as  $z_{N,k}$ .

In the following, we introduce classical statistical models tailored for harmonic, globally anharmonic, and locally anharmonic cluster formation free energies. It is worth noting that the presented models do not account for certain potential factors, such as vibration-rotation coupling and cluster-cluster interactions. However, existing evidence suggests that their impact is likely minor[15] or can be mitigated through careful analysis.

### A. Harmonic Formation Free Energy of a Conformer

Pioneering work by McGinty [15] established a comprehensive expression for the formation free energy related to the  $k$ th conformer of an  $N$ -cluster,  $W_{N,k}$ , considering an ideal gas environment with a specific monomer density,  $\mathcal{N}_1$ . The formulation is given by

$$\frac{W_{N,k}}{k_B T} = N \ln z_1 - \ln z_{N,k} - (N - 1) \ln \mathcal{N}_1, \quad (1)$$

where  $T$  is temperature and  $k_B$  the Boltzmann constant. While this expression can be extended to multicomponent systems[27], our focus here remains on one-component systems.

For point-like monomers, such as atoms, exhibiting only translational motion, the partition function  $z_1 = z_1^{\text{tr}}$ . And consequently, the formation free energy of a single monomer,  $W_1$ , is inherently zero. The total partition function of a cluster,  $N \geq 2$ , involves contributions from different modes of motion and the binding energy,  $E_{b,k}$ , of the equilibrium configuration:

$$z_{N,k} = \exp\left(-\frac{E_{b,k}}{k_B T}\right) z_N^{\text{tr}} z_{N,k}^{\text{rot}} \prod_i^{N_{\text{dof}}} z_{N,k,i}^{\text{vib}}, \quad (2)$$

where the product is over all vibrational degrees of freedom (for linear and nonlinear systems  $N_{\text{dof}} = 3N - 5$  and  $3N - 6$ , respectively).

Motion within a vibrational degree of freedom, denoted by index  $i$ , is characterized by its normal-mode frequency  $\omega_i$ . This frequency is linked to a positive eigenvalue  $\lambda_i$  derived from the diagonalization of the Hessian matrix:  $\omega_i = \sqrt{\lambda_i/\mu}$ , where  $\mu$  is the monomer mass. Since our analysis focuses solely on systems exhibiting classical behavior, the system's thermal energy significantly surpasses the energy associated with a single phonon, i.e.,  $k_B T \gg \hbar\omega_i$  ( $\hbar$  being the Planck constant). The classical partition function for a harmonically oscillating mode is

$$z_{k,i}^{\text{vib}} = z_{k,i}^{\text{hrm}} = \frac{k_B T}{\hbar\omega_{k,i}}. \quad (3)$$

Utilizing also standard partition functions for translational and rotational motion[1, 28], the formation free energy of a non-linear one-component conformer  $k$  within the harmonic model can be conveniently expressed as:

$$\frac{W_{N,k}^{\text{hrm}}}{k_B T} = \frac{E_{b,k}}{k_B T} - (N-1) \ln \mathcal{N}_1 - \frac{3}{2} \ln N + \frac{1}{2} \sum_{i=1}^{N_{\text{dof}}} \ln \left( \frac{\lambda_{k,i}}{2\pi k_B T} \right) - \ln \left( \frac{8\pi^2}{s_k} \sqrt{\frac{I_{xx,k} I_{yy,k} I_{zz,k}}{\mu^3}} \right). \quad (4)$$

Notably, the Planck constant's contribution cancels out in this formulation. Here  $I_{xx}$ ,  $I_{yy}$  and  $I_{zz}$  represent the three principal moments of inertia, and  $s_k$  is the symmetry number of the conformer. For a linear dimer ( $N = 2$ ,  $s = 2$ ), the final term in Eq. 4 is  $-\ln(\pi r_e^2)$ , where  $r_e$  signifies the equilibrium distance between the two monomers. The parameters involved in Eq. 4 are routinely obtainable from optimization calculations.

It is essential to highlight that Eq. 4 is formulated for clusters in an ideal gas, thus disregarding potential interactions with surrounding monomers or other clusters that could influence  $W_N$ . In high-density systems, however, the work imposed upon the cluster by neighboring particles can have an impact[29–31].

In the context of this study, the specific term *harmonic cluster formation free energy*, denoted as  $W_N^{\text{hrm}}$ , is attributed to the free energy of the global minimum energy conformer,  $k = 0$ , i.e.,  $W_N^{\text{hrm}} = W_{N,k=0}^{\text{hrm}}$ . Consequently, when exploring conformers  $k > 0$ , we delve into the realm of global anharmonicity.

## B. Global Anharmonicity

Global anharmonicity stems from the comprehensive consideration of all conformers within a system. According to the principles of statistical mechanics, the partition function of an  $N$ -cluster system, covering various distinct regions of the energy landscape represented by different conformers, is  $z_N = \sum_k z_{N,k}$ [3, 11, 13, 32].

The collective influence of different conformers of the same cluster size manifests in an additive manner when determining free energies. Despite their relatively low stability at  $T = 0$  K, higher energy conformers can contribute significantly to the overall entropy production of the system. And, the free energy consistently decreases with the inclusion of more conformers. The globally anharmonic cluster formation free energy is expressed as

$$\frac{W_N^{\text{Gahrm}}}{k_B T} = - \ln \left[ \sum_k \exp \left( - \frac{\Delta W_{N,k}}{k_B T} \right) \right]. \quad (5)$$

Importantly, this expression holds true irrespective of the barrier heights separating the conformers on the energy landscape[13].

Given its theoretical unambiguousness, the recognition of the significance of Eq. 5 dates back to the early stages of computational cluster studies[18, 32, 33]. However, as pointed out by Partanen *et al.* [13], a notable number of more recent studies have overlooked proper considerations for global anharmonicity.

## C. Local Anharmonicity in Classical Systems

Incorporating global anharmonicity into systems with multiple conformers is a straightforward task. However, due to the nontrivial landscape features associated with vibrational degrees of freedom, estimating the true impact of local anharmonicity is a more challenging endeavor. Particularly significant is the possible undercounting of the vibrational density of states in the harmonic approximation, where the assumption that all energy levels are equally spaced leads to insufficient entropy production. This becomes a concern for clusters at relatively high temperatures, posing a substantial issue in accurately capturing their thermodynamic behavior.

In a straightforward statistical-mechanics model, incorporating an accurate local anharmonic correction for vibrational motion is challenging due to the inherent difficulty in precisely mapping out the basin surrounding a local

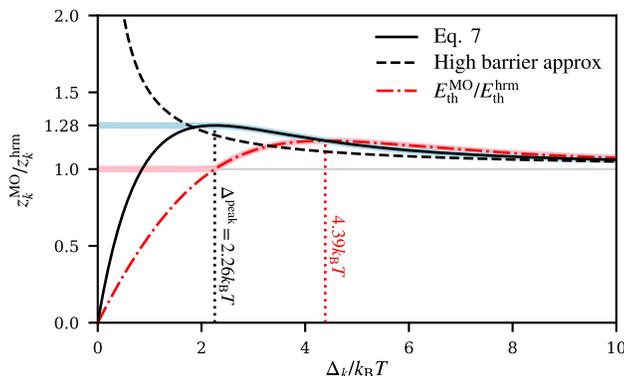


FIG. 1. Anharmonic correction on vibrational partition function, as described by Eq. 7, as a function of dissociation energy  $\Delta_k$  (black solid line). The dashed black line represents the high-barrier estimate from prior studies[11, 35]. The red dash-dotted line illustrates the predicted ratio between anharmonic and harmonic thermal energy, derived from Eqs. 11 and 10, respectively. Dotted vertical lines mark the positions of maximum corrections within the anharmonic model. The effective corrections, as determined by the maximum entropy scheme, Eq. 9, are visually emphasized through thick blue and red lines.

minimum using only basic knowledge about the sampled stationary points (minima and saddle points). In practice, such mapping is achievable for the simplest classical systems through direct phase-space sampling, provided the computational cost is reasonable, and sufficient ergodicity is attained.

Thus a convenient and more realistic analytical alternative to the harmonic approximation is to model the potential wells as the Morse oscillators (MOs)[11, 23–25]. In principle, MOs offer a more accurate representation of the flat, high-energy regions of the energy landscape compared to harmonic oscillators. The energy of MO's  $j$ th energy level is given by

$$\epsilon_i = \hbar\omega_i \left( j + \frac{1}{2} \right) \left[ 1 - \left( j + \frac{1}{2} \right) \frac{\hbar\omega_i}{4\Delta_k} \right], \quad (6)$$

where  $\Delta_k$  is the oscillator's characteristic dissociation energy, assumed to be equal across all vibrational modes of a  $k$ -conformer, a reasonable assumption for small clusters. Essentially this energy sets an upper limit for the vibrational energy states. By taking the Laplace transform of the classical level density,  $\rho(\epsilon) = dj/d\epsilon$ , the anharmonic correction on the vibrational partition function can be expressed as

$$\frac{z_k^{\text{MO}}}{z_k^{\text{hrm}}} = \frac{\int_0^{\Delta_k} d\epsilon_i \rho(\epsilon_i) \exp\left(-\frac{\epsilon_i}{k_B T}\right)}{k_B T / \hbar\omega_i} = 2\sqrt{\frac{\Delta_k}{k_B T}} \mathcal{D}\left(\sqrt{\frac{\Delta_k}{k_B T}}\right), \quad (7)$$

with  $\mathcal{D}(x)$  representing the Dawson integral[34]. Notably, contrary to the general convention[28], the integration in Eq. 7 spans only up to  $\Delta_k$  rather than infinity to confine the energy levels within the well. The locally anharmonic formation free energy of a conformer is thus given by

$$\frac{W_{N,k}^{\text{Lahrm}}}{k_B T} = \frac{W_{N,k}^{\text{hrm}}}{k_B T} - N_{\text{dof}} \times \ln \frac{z_k^{\text{MO}}}{z_k^{\text{hrm}}}. \quad (8)$$

As demonstrated in Fig. 1, the adjustment provided by Eq. 7 peaks at approximately 1.28 for  $\Delta_k^{\text{peak}} \approx 2.26k_B T$ . This translates into a maximum additional entropy of about  $k_B/4$  per vibrational degree of freedom.

In a prior study, Doye and Wales [11] (and subsequently Ball and Berry [35]) employed a similar approach to analyze the thermodynamics of various atomic and molecular clusters. They replaced harmonic oscillators with MOs in their framework, focusing specifically on deriving the anharmonic partition function for cases where  $\Delta_k \gg k_B T$ . Similarly,  $\mathcal{D}(x)$  can be approximated as  $\mathcal{D}(x) \approx x^{-1}/2 + x^{-3}/4$  in this high-barrier limit, and the introduced Eq. 7 simplifies to  $z_k^{\text{MO}}/z_k^{\text{hrm}} = 1 + k_B T/2\Delta_k$  used in refs. 11 and 35. As illustrated in Fig. 1, the high-barrier approximation exhibits a monotonic decrease with  $\Delta_k$ , emphasizing the necessity for  $\Delta_k$  to be several times higher than  $k_B T$  to achieve satisfactory agreement with the exact expression provided by Eq. 7.

The crucial question then arises: what should be considered as the value for  $\Delta_k$ ? Fundamentally,  $\Delta_k$  signifies a point situated on the surfaces that separates a basin from its adjacent basins. The lower bound of  $\Delta_k$  is determined by the smallest transition state barrier separating conformer  $k$  from its nearby conformers  $\{l\}$ :  $\Delta_k \geq \Delta E_k^{\ddagger \text{min}} =$

$\min(E_{k,\{l\}}^\ddagger - E_{b,k})$ . Given the assumed ergodicity of the system, nonstationary states above this minimum saddle point are equally accessible and able to carry statistical weight. Consequently, while the minimum energy path follows through the saddle point, the entropy-favored trajectory might intersect the dividing surface elsewhere.

We conjecture that, among all accessible states, the considered dissociation state maximizes vibrational entropy designated by Eq. 7. Thus, the chosen  $\Delta_k$  in our model is determined as follows:

$$\Delta_k = \max\left(\Delta^{\text{peak}}, \Delta E_k^{\ddagger\text{min}}\right). \quad (9)$$

Effectively, this rule prevents  $z_k^{\text{MO}}$  from being smaller than  $z_k^{\text{hrm}}$ . Conceptually, adopting the maximum entropy scheme results in Eq. 3 when applied to harmonic potential. If the harmonic partition function is derived with an energy-level limitation, the obtained function, integrated from 0 to  $\Delta_k$ , monotonically approaches  $k_B T / \hbar \omega$  from below, leading to  $\Delta_k = \infty$ .

Beyond free energies, the vibrational thermal energy can be directly derived from the partition function, and in a relaxed system this energy is evenly divided to potential and kinetic energy. The thermal energy of a harmonically oscillating mode of motion is given by

$$E_{\text{th},k}^{\text{hrm}} = k_B T^2 \left( \frac{\partial \ln z_k^{\text{hrm}}}{\partial T} \right) = k_B T. \quad (10)$$

Similarly, for the energy related to MOs given by Eq. 7, it is expressed as

$$E_{\text{th},k}^{\text{MO}} = k_B T \left( \frac{1}{2} + \frac{\Delta_k}{k_B T} - \frac{\sqrt{\Delta_k/k_B T}}{2\mathcal{D}(\sqrt{\Delta_k/k_B T})} \right). \quad (11)$$

The maximum value of  $E_{\text{th},k}^{\text{MO}}$  is about  $1.18k_B T$  at  $\Delta_k \approx 4.39k_B T$ , and beyond this point,  $E_{\text{th},k}^{\text{MO}}$  correctly converges to  $E_{\text{th},k}^{\text{hrm}}$  as  $\Delta_k$  increases. Additionally, at the limiting value of  $\Delta^{\text{peak}}$ ,  $E_{\text{th},k}^{\text{MO}} = k_B T$ . The ratio between anharmonic and harmonic potential energy is further illustrated in Fig. 1. Thus, within the proposed scheme for local anharmonicity, the internal thermal energy is also consistently determined.

### III. SYSTEMS AND COMPUTATIONAL METHODS

#### A. Mie fluids

To examine the distinct and combined influences of global and local anharmonicities on cluster thermodynamics, our focus is directed towards the investigation of monatomic complexes featuring diverse ranges of interatomic interactions. The pairwise forces are characterized by the Mie potential, a generalized form of the Lennard-Jones potential:

$$U(r) = C\epsilon \left[ \left( \frac{\sigma}{r} \right)^n - \left( \frac{\sigma}{r} \right)^m \right], \quad (12)$$

where  $r$  is the interatomic distance. The prefactor, determined by

$$C = \frac{n}{n-m} \left( \frac{n}{m} \right)^{\frac{m}{n-m}}, \quad (13)$$

ensures that the pairwise minimum energy is  $-\epsilon$ . For these potentials, the equilibrium distance, denoted as the location of the minimum, is given by  $r_e/\sigma = (n/m)^{1/(n-m)}$ . The parameters  $m$  and  $n$  ( $> m > 3$ ) correspond to the attractive and repulsive ranges of the potential, respectively. The condition  $m > 3$  is set to prevent nonextensive energies near the continuum limit[9]. In this study, only potentials corresponding to systems with  $n = 2m$  are considered, specifically for  $n = 8, 12$ , and  $24$ . For such systems, the constants are  $C = 4$  and  $r_e = 2^{2/n}$ . The respective shapes of the analyzed potentials are illustrated in Fig. 2(a).

As we exclusively consider classical systems where the Planck constant is inconsequential, we adopt standard dimensionless units for Mie system parameters:  $\epsilon = \sigma = \mu = k_B = 1$ . Despite the uniformity in length and energy scales across various potentials, the thermodynamic properties are not directly comparable. However, following the principle of corresponding states[37], the relationships between different thermodynamic properties remain consistent among systems with apparent differences when normalized against their critical points. Therefore, for a more robust physicochemical comparison, it is sometimes more preferable to examine different Mie systems in terms of the temperature scale  $T/T_c$ , where  $T_c$  is the critical temperature (provided in Table I). In general,  $T_c$  exhibits a significant decrease with an increase in  $n$ .

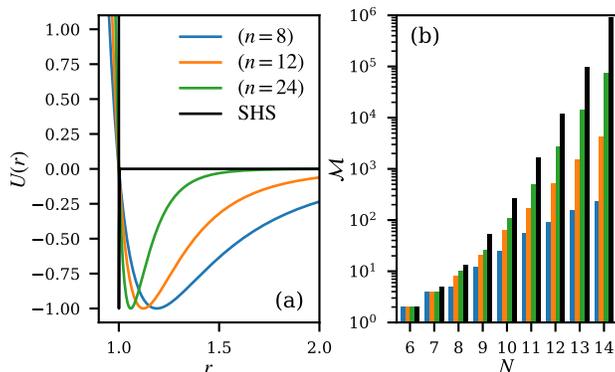


FIG. 2. (a) Mie potentials depicted for various systems under consideration (colored lines). The black line corresponds to the sticky hard-sphere (SHS) system. (b) Number of distinct local minima,  $\mathcal{M}$ , determined through geometry optimization. The black bars signify the results for the SHS clusters[26, 36].

TABLE I. Characteristics of studied Mie Systems ( $n = 2m$ ) including critical temperature ( $T_c$ ), equilibrium pairwise distance ( $r_e$ ), square-well equivalent range of attraction ( $L$ ), and enumeration constants ( $\alpha$  and  $\beta$ ). All values are presented in standard dimensionless units.

$n$	$m$	$T_c$	$r_e$	$L$	$\alpha$	$\beta$
8	4	3.461 <sup>a</sup>	1.189	2.178	0.49	1.33
12	6	1.316 <sup>b</sup>	1.122	1.479	1.07	6.61
24	12	0.560 <sup>b</sup>	1.059	1.150	1.66	12.05

<sup>a</sup> From ref. 38.

<sup>b</sup> From ref. 39.

As depicted in Fig. 2(a), the width of the potential narrows with  $n$ . This *range of attraction*,  $L$ , is a convenient parameter to generalize the phase behavior of simple fluids[37]. It has been established[37] that the width of the potential can be quantified by determining the square-well (SW) equivalent  $L$  (at  $T_c$ ) for a Mie system using the second virial coefficients  $B_2^{\text{SW}}$  and  $B_2^{\text{Mie}}$ :

$$B_2^{\text{Mie}}(T_c) \equiv B_2^{\text{SW}}(T_c) = \frac{2\pi}{3} \left[ \exp\left(\frac{1}{T_c}\right) - L^3 \left( \exp\left(\frac{1}{T_c}\right) - 1 \right) \right]. \quad (14)$$

While  $B_2^{\text{Mie}}(T_c) \approx -3.3$  (in reduced units) remains consistent irrespective of the interaction parameters[39], we precisely compute it using a closed-form expression[40]. The examined systems, along with their respective  $T_c$ ,  $r_e$ , and  $L$ , are detailed in Table I.

The Mie model offers a computational advantage due to its efficiency and capacity to simplify and generalize realistic molecular systems[38, 41]. Particularly relevant to this study, it allows for the optimization of the complete set of close-packed local minimum energy structures and facilitates thorough phase-space sampling for clusters comprising more than a mere handful of monomers, tasks often deemed too demanding for models of higher chemical specificity.

In theoretical chemical physics, the Lennard-Jones (LJ) fluid with  $n = 12$  and  $m = 6$  stands out as perhaps the most extensively studied model system. In practical chemical modeling, Eq. 12 has demonstrated its prowess as a potent functional for parameterizing intermolecular interactions in various polyatomic compounds, relying on the tunable parameters  $n$  and  $m$ [41]. While the range of obtained  $n$  values can be considerable,  $m \approx 6$  generally aligns with the decay of the dispersion interaction in non-polar systems, corresponding to a  $r^{-6}$  decay pattern. For more cohesive systems, lower  $m$  parameters are involved (e.g., for dipole-dipole interaction  $m = 3$ ), prompting interest in exploring systems at the limit of  $m = 4$ [38]. Furthermore, mesoscale particles with a notably short attractive range, such as colloidal spheres, find representation through the Mie potential[6, 42, 43], albeit with a large  $m$ .

For conciseness, here we denote specific  $n = 2m$  Mie-systems and clusters as  $(n = X)$  and  $(n = X)_{N=Y}$ , respectively. As an example, a cluster of 10 monomers with interaction parameters  $n = 12$  and  $m = 6$  is represented as  $(n = 12)_{N=10}$ .

## B. Local minimum and transition state structures

In this study, the primary set of optimized structures for Mie clusters is sourced from a recent study by Trombach *et al.* [36]. They successfully enumerated nearly every minimum for  $(12 \leq n \leq 96)_{8 \leq N \leq 14}$  systems by applying a basin-hopping global optimization algorithm[3, 44].

Comprehensively enumerating local minima becomes practically infeasible for  $N > 14$  due to the staggering number of distinct conformers,  $\mathcal{M}$ , a challenge amplified, especially in the case of short-range potentials. Despite the availability of the extensive database compiled by Trombach *et al.* [36], our calculations are constrained to systems with  $n \leq 24$  due to the computational demands associated with transition state sampling and the subsequent evaluation of  $\Delta E_k^{\ddagger \text{min}}$ . This limitation arises from the observation that the number of distinct transition states is roughly proportional to  $N\mathcal{M}$ [3, 45]. As an alternative, we conducted complementary sampling, incorporating a long-range system with  $n = 8$  and cluster sizes ranging from  $N = 2$  to 7 into the set of optimized structures.

In this study, sampling of minima and connecting transition states was carried out using the Pele package[46]. This software was also employed to generate conformer-specific values for  $E_{b,k}$ ,  $E_{k,l}^{\ddagger}$ ,  $I_{xx,k}$ ,  $I_{yy,k}$ ,  $I_{zz,k}$ , and  $\lambda_{k,i}$ . Throughout the transition state sampling process, additional minima were discovered and subsequently included in the set of analyzed clusters

To address rare instances of mechanically stable yet physically unattached clusters, a lenient acceptance criterion was introduced, stipulating that all  $3N - 6$  eigenvalues  $\lambda$  must exceed 0.01. The total number of minima  $\mathcal{M}$  for  $N \geq 6$  in each studied system is depicted in Fig. 2(b), along with the limiting values found for sticky hard-sphere (SHS) clusters[26].

In our TS sampling protocol, we systematically explore pairs of minima, attempting to locate saddle points between them using the Nudged Elastic Band (NEB) method implemented in the Pele package. This search is specifically focused on *interconfigurational* connections. However, certain conformers exhibit transitions while maintaining their inherent geometry, with notable examples being the defected icosahedral clusters  $N = 12$  and 14, featuring a surface vacancy and an adatom, respectively. These icosahedral structures represent the minimum energy conformers ( $k = 0$ ) for the studied systems. To identify *intraconfigurational* TS of these conformers and calculate  $\Delta E_{\text{ico}}^{\ddagger \text{min}}$ , additional NEB molecular dynamics (MD) simulations were conducted using the LAMMPS MD code[47]. While interconfigurational transitions are assumed to involve all monomers (meaning all vibrational modes of a conformer are treated with the same  $\Delta_k$ ) the intraconfigurational transitions of these icosahedral structures are well-localized. Specifically, the  $N = 14$  adatom transition involves only one monomer (three vibrational modes), while the  $N = 12$  vacancy transition involves five monomers (15 modes). Consequently, in our locally anharmonic model, the remaining vibrational modes of these specific conformers are computed based on the interconfigurational TS barriers instead of  $\Delta E_{\text{ico}}^{\ddagger \text{min}}$ .

## C. Monte Carlo phase-space sampling

To establish a representative benchmark for the discussed statistical models, we conducted a detailed exploration of the classical phase space for each studied system through Metropolis Monte Carlo (MC) sampling. Specifically, we employed a semi-grand canonical simulation scheme developed by Vehkamäki and collaborators[48–50]. In these simulations, canonical ensemble averages of the grand canonical growth and decay probabilities,  $\overline{G}_N$  and  $\overline{D}_N$  for each cluster size, were calculated at some reference vapor monomer density  $\mathcal{N}'_1$ . Based on these probabilities, the formation cluster free energy can be determined recursively and scaled to any desired  $\mathcal{N}_1$  according to the formula[50]

$$\frac{W_N^{\text{MC}}(\mathcal{N}_1)}{k_{\text{B}}T} = \frac{W_{N-1}^{\text{MC}}(\mathcal{N}_1)}{k_{\text{B}}T} - \ln \left( \frac{\mathcal{N}_1 \overline{G}_{N-1}}{\mathcal{N}'_1 \overline{D}_N} \right). \quad (15)$$

Despite the broad applicability of these MC simulations, two fundamental challenges persist:

Firstly, the Metropolis algorithm's sufficiency diminishes at the lower end of the temperature range. In the presence of exceedingly high barriers (relative to the thermal energy  $k_{\text{B}}T$ ) between different adjacent basins, certain states become inaccessible within a reasonable sampling time, rendering proper ergodicity unattainable. Consequently, flawed sampling leads to overestimated free energies. Rather than employing specialized techniques or algorithm modifications to enhance ergodicity[51], we opt for the standard Metropolis scheme, exercising caution when interpreting formation free energies at very low temperatures.

The second notable issue in the simulations pertains to the definition of the phase boundary, specifically, what criteria are used to identify a cluster. In the selected MC simulation protocol, the conventional Stillinger connectivity distance  $r_{\text{S}}$  is employed to determine whether a monomer is part of the cluster[52]. The selection of  $r_{\text{S}}$  is crucial, as its value needs to be tailored for each potential and simulated temperature to avoid a systematic misrepresentation of the formation free energy. An excessively small  $r_{\text{S}}$  results in an overestimation of  $W_N^{\text{MC}}$  (as attractive interactions are

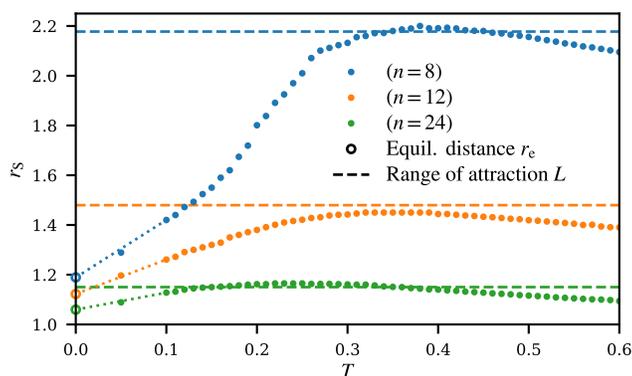


FIG. 3. Temperature dependence of Stillinger connectivity distances employed in our Monte Carlo simulations,  $r_S$ , illustrated for various systems (full dots). Dotted lines show the linear interpolation from  $r_S(T = 0.1)$  to the equilibrium distance  $r_e$ , marked by open dots at  $T = 0$ . System-specific ranges of attraction,  $L$ , are indicated by horizontal dashed lines.

cut short), while an excessively large  $r_S$  leads to the formation of loosely connected, fluffy complexes with an unclear phase boundary, resulting in an underestimated  $W_N^{\text{MC}}$ .

Following the approach proposed by Merikanto *et al.* [53], consistent values of  $r_S$  are determined by comparing the obtained  $W_N^{\text{MC}}$  with those derived from virial coefficients and the Mayer cluster expansion[54]. Given the computational simplicity of the second virial coefficients and the shared nature of  $r_S$  among different cluster sizes (at least for  $(n = 12)_{N \leq 6}$  [53]), we computed commensurate connectivity distances by conducting a series of MC simulations for dimers with varying  $r_S$ . For an appropriate  $r_S$ ,  $W_2^{\text{MC}}(\mathcal{N}_1, r_S)$  should align with  $-k_B T \ln(-\mathcal{N}_1 B_2^{\text{Mie}})$ . The Mayer-equivalent  $r_S$  values obtained are presented in Fig. 3 for all systems as a function of temperature.

Crucially, the thermodynamic reach offered by the Mayer expansion surpasses that of the MC framework and Eq. 1. Mayer clusters existing in an imperfect vapor interact with other clusters and unbound monomers. Consequently, the determined connectivity distance is not a monotonic function of temperature: Overall, a system necessitates a larger  $r_S$  as thermal motion intensifies with increasing temperature. However, the vapor pressure around the cluster also rises with  $T$ , counteracting this trend by imposing constraints on  $r_S$  due to the destabilizing influence from nearby particles.

As shown in Fig. 3, the interplay between thermal motion and the effects arising from vapor-cluster/cluster-cluster interactions is manifested through the appearance of an inflection point on the  $(r_S, T)$ -curve. Despite the possibility of analytically estimating the contribution of surrounding particles to  $W_N$ [31], we have made the deliberate choice to exclusively incorporate the calculated  $r_S$  in our free energy simulations below the inflection-point temperature. Beyond this temperature threshold, we opt for the maximum value of  $r_S$ . This decision is further substantiated by the observation that these maxima align closely with the system-specific ranges of attraction  $L$  provided in Table I (indicated by horizontal dashed lines in Fig. 3).

## IV. RESULTS

### A. Conformers and Transition States Analysis

#### 1. Cluster Minima and Binding Energies

Firstly, we conduct a concise analysis of the general trends in the number of distinct minima identified for various systems, depicted earlier in Fig. 2(b). While this aspect is not directly tied to free energy calculations, its findings will be useful, as demonstrated subsequently. Additionally, we present the characteristics of cluster binding energy distributions for the examined systems.

The analyzed conformer energies and their distributions are illustrated in Fig. 4. The energy data cover all three Mie interaction types ( $n = 8, 12$ , and  $24$ ) within the range of  $N = 6$  to  $14$  and are presented per vibrational degree of freedom, i.e.,  $E_b/(3N - 6)$ . This is in accordance with Maxwell's rule, which states that a rigid frame of  $N$  points necessitates at least  $3N - 6$  connecting lines[26, 55]. Although the analyzed clusters systematically follows this rule, systems with larger  $n$  (and consequently shorter  $L$ ) may exhibit conformers with more contacts (evident in the SHS system for  $N \geq 10$ ).

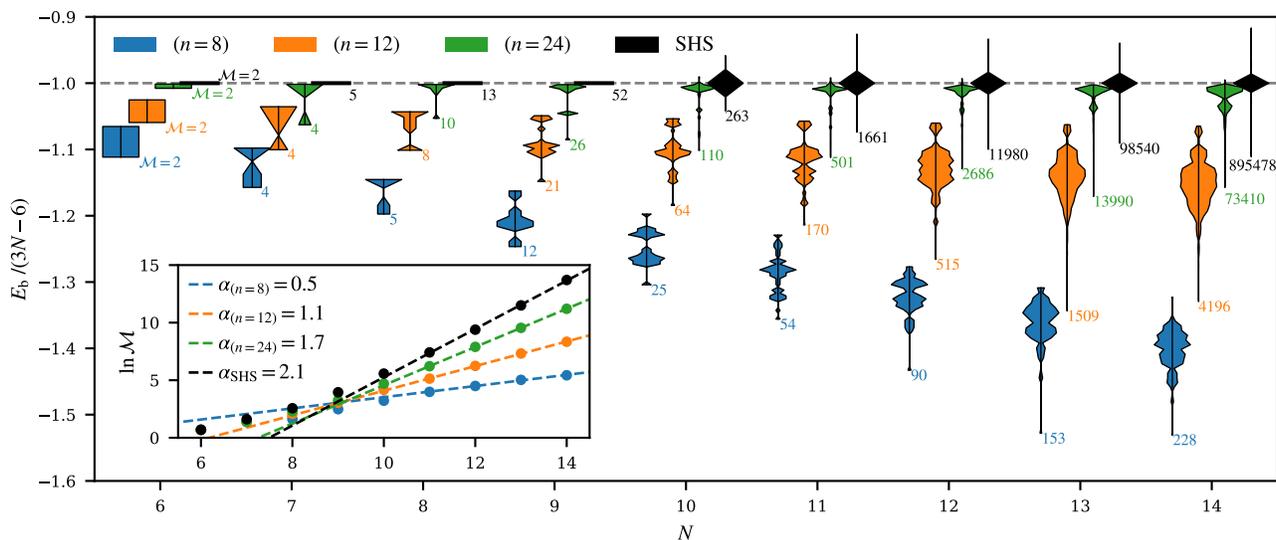


FIG. 4. (Main figure) Normalized binding energy distributions presented for clusters exhibiting multiple conformers. The respective number of conformers,  $\mathcal{M}$ , is provided next to the corresponding distribution curve. A dashed horizontal line represents the “Maxwell limit”. (Inset)  $\ln \mathcal{M}$  is plotted against the cluster size  $N$ . Linear fits are applied to points with  $N \geq 11$ , and the resulting slopes ( $\alpha$ ) are given in the inset legend. The black markers correspond to clusters within the sticky hard-sphere system.

As observed from the distributions, despite the abundance of conformers, a significant portion of short-range systems is concentrated on narrow high-energy bands, primarily at  $-E_b = 3N - 6$  or  $3N - 5$ . In contrast, the binding energy distributions of ( $n = 8$ ) and ( $n = 12$ ) systems are more dispersed and extend to lower energies, reflecting contributions from remote monomers. The shapes of the distributions from  $N = 6$  to  $N = 14$  roughly resembles types of cocktail glasses: from old-fashioned tumblers to martini glasses and eventually to long-stem champagne flutes. The latter referring to large clusters with a uniquely low-energy  $k = 0$  conformer.

Apart from the smallest clusters, the number of minima  $\mathcal{M}$  shows an exponential increase with size  $N$ [56]:

$$\ln \mathcal{M} \approx \alpha N - \beta, \quad (16)$$

where  $\alpha$  and  $\beta$  are interaction-specific constants. In the inset of Fig. 4,  $\ln \mathcal{M}$  is presented as a function of  $N$  for the studied systems (and the SHS system as well) with corresponding linear fits (for  $N \geq 11$ ). The obtained  $\alpha$  and  $\beta$  are provided in Table I. Since each distinct equilibrium corresponds to a unique set of points in three-dimensional space,  $\mathcal{M}$  can be considered proportional to a volume associated with the width of the attraction ( $L^3$ ). For example, a SHS system facilitates more minima ( $L = 1$ ,  $\alpha = \alpha_{\text{SHS}} = 2.10$ ) than a system interacting through a long-range Mie potential ( $n = 12$ ,  $L = 1.479$ ,  $\alpha = 1.07$ ). Based on this rationale and the results presented in the inset of Fig. 4, a straightforward relationship emerges:

$$\alpha \approx \alpha_{\text{SHS}} - 3 \ln L. \quad (17)$$

This relation holds also well for the other ( $n = 2m > 24$ ) systems studied by Trombach *et al.* [36]. Given that  $\alpha > 0$ , the rule expressed by Eq. 17 is applicable to systems with  $L < \exp(\alpha_{\text{SHS}}/3) \approx 2$ . Therefore, systems with an exceptionally long range, such as ( $n = 8$ ), do not conform to this relationship.

## 2. Minimum transition state barriers

While the number of unique minima becomes substantial as  $N$  surpasses a certain size and  $L$  decreases, the count of transition states (TS) exhibits an even stronger dependence on  $N$ . Consequently, a comprehensive exploration of transition states, crucial for determining  $\Delta E^{\ddagger \text{min}}$  for each considered minimum, presents a challenge of its own right.

The search for interconfigurational connections is meticulously conducted for the following systems: Given the relatively small  $\mathcal{M}$  across the size range, the TS sampling is efficiently performed for the ( $n = 8$ ) system without substantial computational effort. For the ( $n = 12$ ) system, we guide our sampling using previously obtained TS networks by Doye *et al.* [45, 57]. An extensive sampling of TS is also carried out for ( $n = 24$ ) $_{N < 12}$  and it has

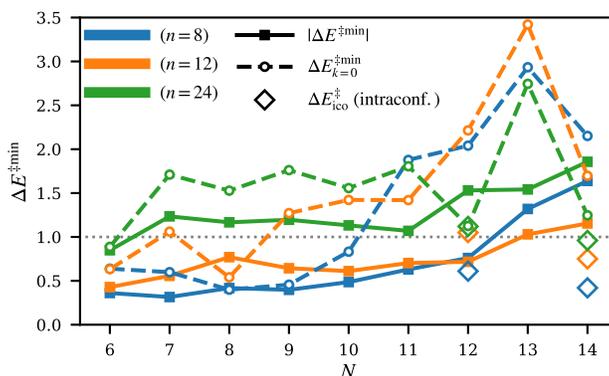


FIG. 5. Minimum transition state barrier height,  $\Delta E^{\ddagger\min}$ , as a function of cluster size  $N$ . Solid lines with square markers show the averaged  $E^{\ddagger\min}$  across various conformers of a given size. Dashed lines with open circles depict the barriers corresponding to the global minimum energy conformer,  $k = 0$ . Additionally, open diamonds represent intraconfigurational transition state energies obtained from NEB-MD simulations specifically conducted for the icosahedral  $N = 12$  and  $14$  cluster structures.

been ensured that at least one connection is found for each minimum. For  $(n = 24)_{N \geq 12}$ , the search focuses on TS connecting the minimum energy conformer to other minima. Additionally, the search extends to some low-free-energy conformers, selected based on  $W_{N,k}^{\text{hrm}}$ . In the subsequent thermodynamic calculations, conformers without any assigned TS are set to  $\Delta E^{\ddagger\min} = 1$ . This also applies to clusters with only one conformer (i.e.,  $N < 6$ ).

In Fig. 5, both the average  $\Delta E^{\ddagger\min}$  over all identified TS (squares, solid lines) and  $\Delta E_{k=0}^{\ddagger\min}$  related to the minimum energy conformer (circles, dashed lines) are presented for the systems with multiple conformers. Additionally, for the defected icosahedra  $N = 12$  and  $14$ , the intraconfigurational  $\Delta E_{\text{ico}}^{\ddagger\min}$  obtained via the NEB-MD simulations are indicated with large diamonds.

As shown in Fig. 4, the global minimum energy conformer typically has a uniquely low binding energy, and  $\Delta E_{k=0}^{\ddagger\min} \geq E_{b,k=1} - E_{b,k=0}$ . It is thus expected that  $\Delta E_{k=0}^{\ddagger\min}$  is higher than  $|\Delta E^{\ddagger\min}|$ . We observe two main exceptions to this pattern:  $(n = 12)_{N=8}$  and  $(n = 24)_{N=12}$ . In both cases, the conformers  $k = 0$  and  $k = 1$  are energetically very close to each other; for  $(n = 12)_{N=8}$ , this is manifested as a wide brim of the “cocktail glass” in Fig. 4. Moreover, in an SHS system, performing a transition between two conformers requires the breakage of at least one bond. Thus,  $\Delta E^{\ddagger\min} \gtrsim 1$  when  $n$  is large, a behavior followed by the  $(n = 24)$  system. Also, generally, when approaching the “champagne flute” systems, i.e.,  $N \gtrsim 12$ ,  $\Delta E_{k=0}^{\ddagger\min}$  increases.

The transition state sampling uncovered some anomalies in the conformer data, where certain energetically adjacent conformers were connected by an extremely shallow barrier, i.e.,  $\Delta E_{k,l}^{\ddagger} \approx \Delta E_{l,k}^{\ddagger} \approx 0$ . A closer examination revealed that these instances were mere duplicates of the same conformer but were erroneously identified as separate entities due to limited numerical precision. Fortunately, after a thorough check of the data, we found only a few occurrences of this issue: the original conformer sets of systems  $(n = 8)_{N=10}$ ,  $(n = 8)_{N=12}$ ,  $(n = 24)_{N=13}$ , and  $(n = 24)_{N=14}$  contained only 1, 1, 3, and 100 duplicates, respectively. These false positives have been excluded from the conformer data presented in Figs. 2(b) and 4.

In the context of statistical thermodynamics and our model for local anharmonicity, the relatively high  $\Delta E_{k=0}^{\ddagger\min}$  holds significant implications. If the TS barriers are approximately equal throughout  $k$ , the adjustment for local anharmonicity would not significantly influence the relative population of various conformers within a specific cluster size. However, the presence of distinctively high barriers indicates diminished vibrational entropy, consequently influencing the abundance of specific conformers. This aspect will be briefly elaborated upon in Sec. VB.

## B. Effect of Local and Global Anharmonicities on Cluster Thermodynamics

### 1. Magnitude of the effect

Now that we have gathered the essential data on the minima, we can proceed to calculate the formation free energies for each system. Our initial focus is on examining the relative impact of both global (Eq. 5) and local (Eq. 8) anharmonicity on the thermodynamics of the systems across different temperatures.

The impact of global anharmonicity is shown in Fig. 6(a,c,e) through a comparison of  $W_N^{\text{Gahrm}}$  with the harmonic cluster formation free energy  $W_N^{\text{hrm}}$ . Also here, the resulting differences,  $\Delta W$ , are presented per  $N_{\text{dof}}$  to enhance

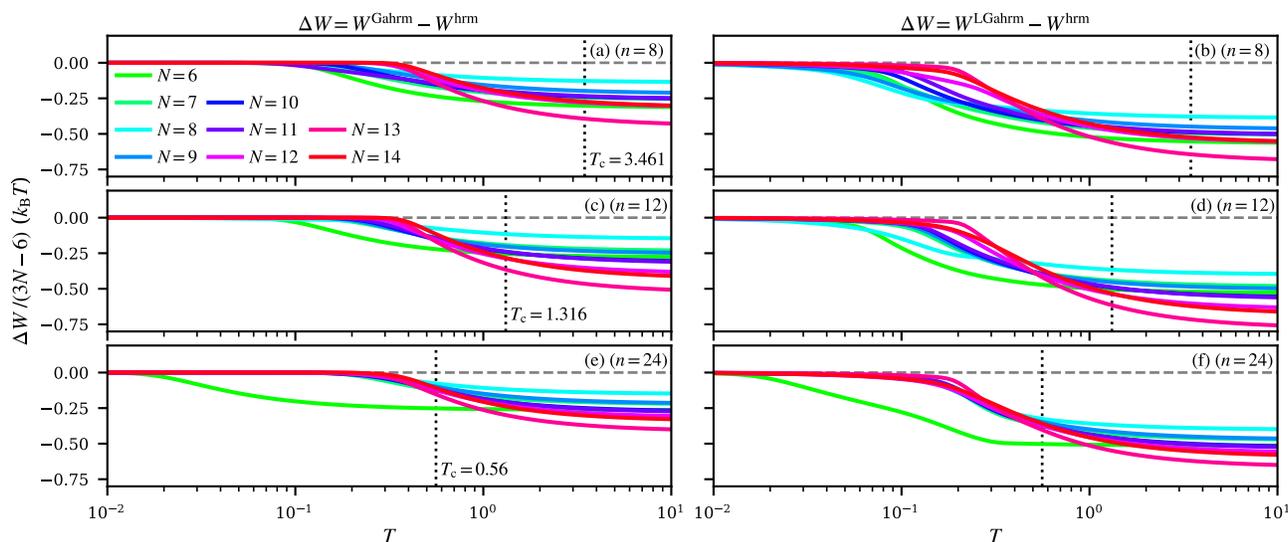


FIG. 6. (Left-hand side column; (a), (c), (e)) Temperature-dependent difference between globally anharmonic cluster formation free energies,  $W_N^{\text{Gahrm}}$ , and the corresponding harmonic formation free energy,  $W_N^{\text{hrm}}$ , for cluster sizes  $N \geq 6$ . (Right-hand side column; (b), (d), (f)) Differences between locally and globally anharmonic cluster formation free energies,  $W_N^{\text{LGahrm}}$ , and  $W_N^{\text{hrm}}$ . The values are normalized with respect to the number of vibrational degrees of freedom and presented in units of thermal energy ( $k_B T$ ). Vertical dotted lines mark system-specific critical temperatures, while horizontal dashed lines highlight the zero level.

the clarity of the comparison. The temperature at which  $W_N^{\text{Gahrm}}$  begins to deviate from  $W_N^{\text{hrm}}$  signifies the cluster’s “melting point”, a temperature that varies with cluster size and among systems. Due to their high surface-to-volume ratio, finite-size particles experience melting at lower temperatures than the bulk melting temperature.

In all systems, the hexamer ( $N = 6$ ), characterized by two conformers, demonstrates the most notable melting behavior. The impact of global anharmonicity on the hexamer is pronounced already at very low temperatures; the transition is driven by differences in rotational symmetry ( $s = 24$  for the octahedral  $k = 0$  and 2 for the polytetrahedral  $k = 1$  conformer) and small binding energy gaps ( $\Delta E_b \approx 0.6, 0.4,$  and  $0.1$  for  $(n = 8)_{N=6}$ ,  $(n = 12)_{N=6}$ , and  $(n = 24)_{N=6}$ , respectively). Another notable cluster size is  $N = 8$ , where the effect of global anharmonicity is relatively small. Unlike the hexamer, the  $k = 0$  conformer’s symmetry number is only 2, while some high-energy conformers are more symmetrical, leading to a less significant entropic gain from transitioning between high and low symmetry. Generally, for other cluster sizes, based on  $W_N^{\text{Gahrm}}$ , melting occurs around  $T \approx 0.3$ , irrespective of the system. Although similar on the absolute temperature scale, the melting conditions vary significantly concerning the systems’  $T_c$ .

Overall, the impact of global anharmonicity per vibrational degree of freedom remains relatively uniform at very high temperatures, approximately  $(-0.3 \pm 0.1)k_B T/N_{\text{dof}}$ . At high temperatures, where  $E_{k>0} - E_{k=0} \ll k_B T$ , global anharmonicity is primarily associated with the conformer space size  $\mathcal{M}$ . Equation 18 provides an approximation:

$$\frac{W_N^{\text{Gahrm}}}{k_B T} \approx \frac{W_N^{\text{hrm}}}{k_B T} + N_{\text{dof}} \ln \left| \sqrt{\left\langle \frac{\lambda_k}{\lambda_{k=0}} \right\rangle} \right| - \ln \mathcal{M} - \ln s_{k=0}. \quad (18)$$

Here,  $|\dots|$  and  $\langle \dots \rangle$  denote the arithmetic mean over conformers and geometric mean over conformer’s degrees of freedom, respectively. Equation 18 assumes that conformers  $k > 0$  effectively lack notable rotational symmetry, i.e.,  $s_k \approx 1$ . For most clusters studied here, the averaged eigenvalue ratio  $\langle \lambda_k / \lambda_{k=0} \rangle$  is close to unity. Consequently, for large  $N$ ,  $\Delta W_N / k_B T / N_{\text{dof}} \approx -\alpha/3$ . This behavior, requiring a small  $\beta$ , is rather evident in Fig. 6 for  $(n = 8)$  and  $(n = 12)$ , corresponding to  $\alpha/3 = 0.16$  and  $0.36$ , respectively.

The results computed with both the global correction and the proposed correction for local anharmonicity are presented in Fig. 6(b,d,f). We use notation  $W^{\text{LGahrm}}$  for free energies calculated using Eq. 5 with anharmonic conformer formation energies from Eq. 8. Firstly, the cluster-specific melting temperatures decrease due to anharmonic vibrational motion, at  $T \approx 0.15 - 0.2$ . As indicated in Sec. IV A 2 and in Fig. 5, the  $k = 0$  conformers are typically separated from other minima by high barriers, resulting in non-minimum conformers generating more vibrational entropy than the global minimum. Interestingly, the melting points of  $(n = 24)_{N>6}$  remain relatively constant, about  $T = 0.18$ , while other systems exhibit more significant deviations. This consistency in  $(n = 24)$  is attributed to both the uniform binding energy distributions and the consistent  $E^{\ddagger \text{min}}$  per size.

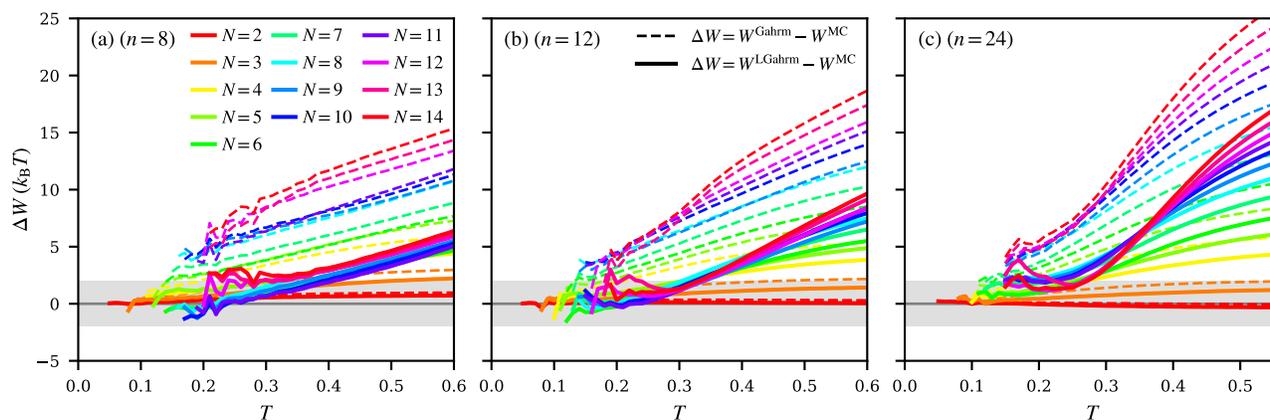


FIG. 7. Comparison between  $W_N^{\text{Gahrm}}$  (dashed lines) and  $W_N^{\text{LGahrm}}$  (solid lines) against the Monte Carlo simulated formation free energies,  $W_N^{\text{MC}}$ , for the studied systems of (a) ( $n = 8$ ), (b) ( $n = 12$ ), and (c) ( $n = 24$ ). The shaded gray area highlights the defined tolerance interval set to  $\pm 2k_B T$ .

Our locally harmonic model results in a maximum additional vibrational entropy of  $k_B/4$  per vibrational degree of freedom at high temperatures. Therefore, the approximation given in Eq. 18 can be further expanded to

$$\frac{W_N^{\text{LGahrm}}}{k_B T} \approx \frac{W_N^{\text{hrm}}}{k_B T} - \frac{N_{\text{dof}}}{4} - \ln \mathcal{M} - \ln s_{k=0}. \quad (19)$$

While this relation is built on many assumptions and approximations, it holds significant potential. As mentioned, the complete enumeration of conformers (not to mention the transition states) is only feasible for the smallest cluster sizes. Therefore, meticulous anharmonic cluster thermodynamic calculations, as presented here, are beyond reach in most circumstances related to e.g., nucleation. Thus, by only knowing the properties of the minimum energy conformer and the system-specific function for  $\ln \mathcal{M}$ , Eq. 19 enables us to at least estimate the magnitude of anharmonicity at high temperatures. This will be demonstrated later in Sec. V A.

## 2. Comparing Anharmonic Model Predictions with Monte Carlo Simulations

The introduced local anharmonicity correction in Sec. II C requires further scrutiny. The validity of the global anharmonicity treatment in statistical thermodynamics is much less ambiguous. Since we operate in the classical limit where the effects arising from discrete states are absent, proper validation of the anharmonic correction can be established by comparing the statistical model with direct phase space sampling, i.e., the Monte Carlo simulations discussed in Sec. III C. Assuming that the setup of the simulations is adequate, i.e., ergodicity is achieved and the applied cluster criterion is satisfactory, the obtained  $W^{\text{MC}}$  should reflect the true thermodynamics of the studied clusters in a perfect gas. In our simulations, we have probed a temperature range up to 0.6 or to  $T_c$  if  $T_c < 0.6$ .

Figures 7(a)–(c) show the comparison between the statistical models and the MC simulations for each studied system, ranging from  $N = 2$  to  $N = 14$ . The differences in free energies are displayed for both the model with the local anharmonicity correction ( $\Delta W = W_N^{\text{LGahrm}} - W_N^{\text{MC}}$ , solid lines) and the model without the local anharmonicity correction ( $W_N^{\text{Gahrm}} - W_N^{\text{MC}}$ , dashed lines). To account for potential uncertainties, we have considered an arbitrary margin of  $2k_B T$  for the free energies, which corresponds to Pople’s chemical accuracy[58] (approximately 1kcal/mol at room temperature). This tolerance interval is represented by the gray area in Fig. 7.

Owing to the limitations of the standard Metropolis Monte Carlo approach at low temperatures and the recursive nature of free energy computation (Eq. 15), Fig. 7 presents  $\Delta W$  only from a certain temperature onwards, and this cutoff temperature increases with  $N$ . Nevertheless, the data unmistakably reveals the temperatures at which the MC simulations encounter challenges, leading to excessively elevated free energies. Despite these inherent simulation issues, values from the simulations closely align with those predicted by the statistical models at the lowest temperatures where ergodicity is maintained.

While the dimer ( $N = 2$ ) and the trimer ( $N = 3$ ) are represented fairly accurately in all systems, the statistical models, on the whole, appear to overestimate the cluster formation free energy. Specifically, the divergence from the MC simulated free energies begins at relatively low temperatures, around  $T \approx 0.15$ , when only global anharmonicity is considered, and this discrepancy amplifies with increasing  $N$ . Some deviation from  $W^{\text{MC}}$  persists even when

both local and global anharmonicities are taken into account, albeit occurring at higher temperatures. Notably, the local anharmonicity correction rectifies the size-dependency issue by bringing together the curves shown in Fig. 7, particularly when  $T/T_c$  is low and  $N$  is large. This consolidation is a significant outcome: the systematic errors associated with  $N_{\text{dof}}$  in the statistical model can be alleviated, and the obtained  $W_N^{\text{LGahrm}}$  exhibit consistency across different cluster sizes.

The results presented in Fig. 7 reveal certain peculiarities, although the overall data appears to be quite consistent. The statistical free energies of the icosahedral conformers ( $N = 12, 13, 14$ ) seem to be slightly elevated. However, beyond  $T \approx 0.25$ , this erratic behavior dissipates, and the values align with other sizes. Additionally, while  $W_N^{\text{LGahrm}} - W_N^{\text{MC}} \approx 0$  for all cluster sizes in systems ( $n = 8$ ) and ( $n = 12$ ) at the lowest temperatures, the values corresponding to system ( $n = 24$ ) show a shift of approximately  $1k_B T$  or  $2k_B T$ . These peculiarities may stem from the applied statistical methods, input parameters (e.g.,  $E^\ddagger$ ), Monte Carlo simulations, or a combination of these factors. Nonetheless, given that the size- or system-specific discrepancies are relatively small, within the range of  $\pm 2k_B T$ , our focus will be on investigating the general temperature-related disagreement evident in Fig. 7.

The observation that  $W_N^{\text{MC}} < W_N^{\text{LGahrm}}$  at high temperatures suggests that the phase boundary set in the MC simulations might be too loose. Previous investigations on Lennard-Jones clusters also support this notion[59, 60]. An elementary analysis of the studied conformers indicates that each monomer has at least 3 neighbors (2 for  $N = 3$ , 1 for  $N = 2$ ) within the connectivity distance, while the Stillinger criterion in the MC simulation protocol is satisfied with a single neighbor. Rather than modifying the MC algorithm and conducting new simulations with an alternative cluster criterion, we opted to reevaluate the simulated cluster configurations using the ten Wolde-Frenkel cluster definition[61]. This involves selecting configurations that meet the ten Wolde-Frenkel criterion (all monomers have 3 or more neighbors within  $r_s$ ). The categorization into Stillinger (S-) and ten Wolde-Frenkel (tWF-) cluster configurations, where all S-clusters are also tWF-clusters, across various temperatures enables us to estimate the relative free energies through thermal integration from a given temperature  $T_0$  to  $T_1$ :

$$\frac{W_N(T_1)}{k_B T_1} - \frac{W_N(T_0)}{k_B T_0} = - \int_{T_0}^{T_1} dT \frac{U_N(T)}{k_B T^2}. \quad (20)$$

The cluster's internal energy  $U_N$ , comprising  $E_{\text{b},N}$  and  $E_{\text{th},N}$ , is not directly comparable with the potential energy calculated from a simulated configuration. However, assuming that the binding energies of both S and tWF-clusters are equal (i.e., the difference between these clusters lies only in thermal motion) and that  $W_N^{\text{MC/tWF}}(T_0) = W_N^{\text{MC/S}}(T_0)$ , the criterion-adjusted formation free energy can be computed at  $T_1$  as follows

$$\frac{W_N^{\text{MC/tWF}}(T_1)}{k_B T_1} = \frac{W_N^{\text{MC/S}}(T_1)}{k_B T_1} - 2 \int_{T_0}^{T_1} dT \frac{E_{\text{pot},N}^{\text{MC/tWF}} - E_{\text{pot},N}^{\text{MC/S}}}{k_B T^2}. \quad (21)$$

The potential energy difference between the two cluster types,  $E_{\text{pot},N}^{\text{MC/tWF}} - E_{\text{pot},N}^{\text{MC/S}}$ , is shown in Fig. 8(a). Given the extensive nature of the original MC data, even after excluding all non-tWF configurations, the statistical uncertainties of  $E_{\text{pot},N}^{\text{MC/tWF}}$  are deemed insignificant. The difference between  $W_N^{\text{LGahrm}}$  and  $W_N^{\text{MC/tWF}}$  for the ( $n = 12$ ) system is presented in Fig. 8(b). The reference temperature,  $T_0$ , is set to 0.25. Following the criterion adjustment, the temperature-related deviation diminished and fell within the assigned error margin. This coarse analysis already elucidates that the ambiguity related to the cluster criterion induces most of the discrepancy between the statistical model and the simulations explicitly probing the phase space.

### C. Potential energy

Finally, we compare the statistical model predictions with the simulated potential energies of the tWF-clusters. For the harmonic approximation,  $E_{\text{pot}}^{\text{hrm}} = E_{\text{b},k=0} + \frac{N_{\text{dof}}}{2} k_B T$ . The globally anharmonic potential energies (with or without the local anharmonicity) are obtained through Boltzmann averaging:

$$E_{\text{pot},N}^{(\text{L})\text{Gahrm}} = \frac{\sum_k \left( E_{\text{b},k} + \frac{N_{\text{dof}}}{2} E_{\text{th},k}^{(\text{MO})} \right) \exp \left( -W_{N,k}^{(\text{La})\text{hrm}} / k_B T \right)}{\exp \left( -W_N^{(\text{L})\text{Gahrm}} / k_B T \right)}. \quad (22)$$

In Fig. 9, we illustrate the impact of anharmonicity on potential energy in comparison to the harmonic approximation for the ( $n = 12$ ) system. Since  $\mathcal{M} = 1$  and  $E^{\ddagger\text{min}} = 1$  for  $N < 6$ , global anharmonicity has no effect on potential energies, and the thermal energy elevation due to local anharmonicity is size-independent with respect to  $N_{\text{dof}}$ . We have presented these results separately in the inset of Fig. 9(a). While the statistical model may not precisely capture

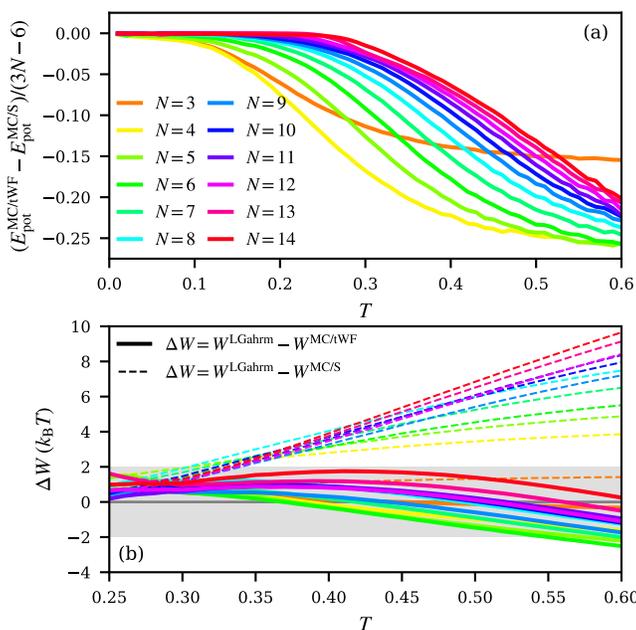


FIG. 8. (a) Comparative analysis of the Monte Carlo (MC) simulated potential energies for ten Wolde-Frenkel (tWF) and Stillinger (S) clusters within the ( $n = 12$ ) system. The values are normalized per the number of degrees of freedom and presented in units of  $\epsilon$ . (b) Comparison of  $W_N^{\text{LGahrm}}$  against the formation free energies obtained through thermal integration for the analyzed tWF-clusters (solid lines). The dashed lines represent analogous values for Stillinger clusters.

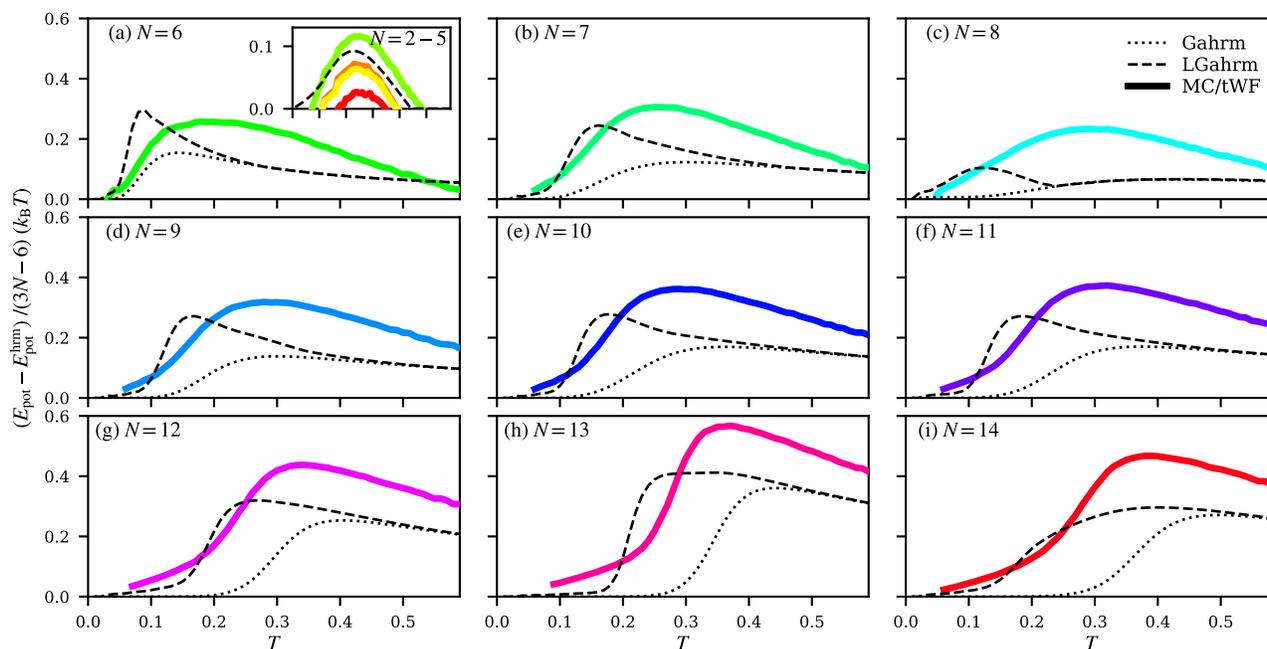


FIG. 9. (a)–(i) Deviation from the harmonic potential energy presented as a function of temperature for  $N \geq 6$  within the ( $n = 12$ ) system. Dashed lines represent the potential energy differences calculated for locally and globally anharmonic systems,  $E_{\text{pot}}^{\text{LGahrm}} - E_{\text{pot}}^{\text{hrm}}$ . Dotted lines depict those without local anharmonicity,  $E_{\text{pot}}^{\text{Gahrm}} - E_{\text{pot}}^{\text{hrm}}$ . The thick colored lines correspond to Monte Carlo simulated energies for the analyzed ten Wolde-Frenkel clusters. The inset in panel (a) shows the comparison for clusters with only one conformer, where our locally anharmonic model yields a uniform deviation for  $2 \leq N \leq 5$ .

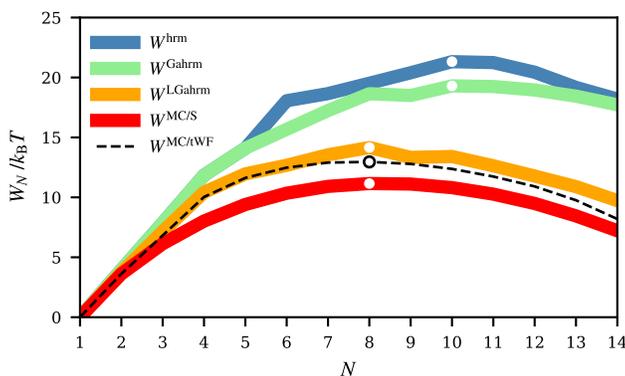


FIG. 10. Illustration of nucleation barriers for ( $n = 12$ ) clusters in a supersaturated vapor at  $\mathcal{N}_1 = 7 \times 10^{-4}$  and  $T = 0.35$ , derived from both harmonic and anharmonic statistical models, as well as Monte Carlo (MC) simulations. Critical cluster sizes, i.e., the peaks of the barriers, are highlighted with white dots. The black dashed line corresponds to the MC values recalculated for ten Wolde-Frenkel (tWF) clusters.

the magnitude of the simulated anharmonic potential energy, it effectively identifies the temperature at which the effect is maximized. According to Eq. 11, the peak at  $\sim 0.09k_B T$  corresponds to  $T = \Delta_k/4.39 \approx 0.23$ .

For clusters with multiple conformers, the departure from harmonic potential energies becomes more pronounced, as both thermal energy and binding energy are influenced by anharmonic corrections. The outset of the energy deviation in Fig. 9 coincides with the melting phenomenon in Fig. 6. Overall, the simulated onset of anharmonicity aligns well with the locally anharmonic model. In some cases, the effect is slightly overestimated at very low temperatures, but the onset temperature is surprisingly accurately predicted by the anharmonic model. At high temperatures, however,  $E_{\text{pot}}^{\text{MC/tWF}}$  is notably higher than the values provided by the models. When approaching low-density systems with extreme thermal agitation, statistical approaches based on equilibrium configurations are not expected to be very accurate methods[9].

## V. DISCUSSION

In summarizing the comparisons drawn between potential and free energies, our presented results underscore the adequacy and predictive power of the proposed anharmonic model. By exploring interaction-wise diverse systems and different properties, we aimed to mitigate the possibility of the observed agreement between the statistical model and direct phase-space sampling being merely coincidental. Although such an exhaustive benchmarking is challenging and often implausible for more complex systems, particularly those involving quantum aspects, our detailed analysis provides valuable insights for these systems as well. In the following, we will briefly discuss the implications that this model brings to the fields of cluster and nucleation science.

### A. Free energy barrier in homogeneous nucleation

Formation free energy is the crucial parameter that dictate the nucleation process, governing cluster growth from a supersaturated vapor or solution[4, 62]. Due to supersaturation,  $W_N$  has a shape of a downward parabola: the transition of the unstable system to the respective stable phase is thereby hindered by a *nucleation barrier*. Overcoming this barrier marks the initiation of nucleation: the position of the barrier's peak determines the *critical* cluster size, and the free energy corresponding to this size enters the exponent in the nucleation rate expression. As phenomenological theories are destined to fail at nanoscale[5, 62], it is preferable to employ atomistic modeling for cluster thermodynamics without reference to macroscopically defined quantities.

As discussed in Sec. IV B 2, without assessing local anharmonicity, the statistical model suffers from size-dependent overestimation of  $W_N$ . We have demonstrated this failure in a heavily supersaturated ( $n = 12$ ) system:  $T = 0.35$  and  $\mathcal{N}_1 = 7 \times 10^{-4}$ . As shown in Fig. 10, the critical cluster sizes according to  $W^{\text{hrm}}$  and  $W^{\text{Gahrm}}$  is 10, whereas both  $W^{\text{LGahrm}}$  and  $W^{\text{MC/S}}$  render the critical size at  $N = 8$ . Moreover, if the MC-values are reestimated for the tWF-clusters,  $W^{\text{LGahrm}}$  and  $W^{\text{MC/tWF}}$  align almost perfectly, the difference at  $N = 8$  is only  $1.2k_B T$ . The peaks of  $W^{\text{hrm}}$  and  $W^{\text{Gahrm}}$  are  $8.3k_B T$  and  $6.4k_B T$  above  $W^{\text{MC/tWF}}$ , translating into about three orders of magnitude lower nucleation rate.

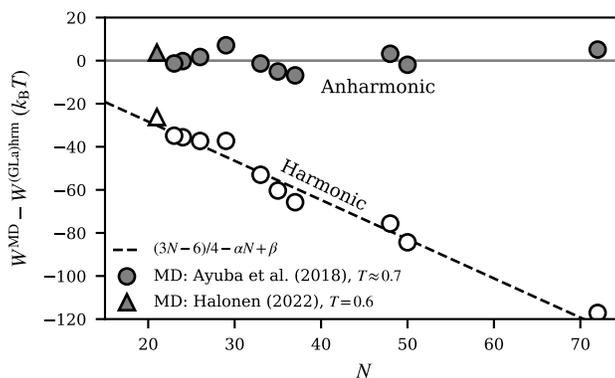


FIG. 11. Comparison between molecular dynamics (MD) nucleation data[63, 64] for ( $n = 12$ ) and the statistical model. Open symbols represent the difference between MD-simulated critical cluster formation free energy,  $W_N^{\text{MD}}$ , and the harmonic free energy,  $W_N^{\text{hrm}}$ . Filled symbols show the comparison against locally and globally anharmonic values,  $W_N^{\text{Lahrm}}$ , estimated using Eq. 19 with parameters  $\alpha = 1.07$  and  $\beta = 6.61$ . The adjustment provided by Eq. 19 is further depicted by the dashed line.

Under typical nucleation conditions, especially at high temperatures, the anticipated critical cluster is composed of more than 8 monomers. Since global minimum energy structures for ( $n = 12$ ) $N \leq 1000$  are readily available[36],  $W_N^{\text{hrm}}$  can be easily computed for all relevant cluster sizes. We thus test the validity of both the harmonic approximation and the approximation given in Eq. 19 against recent molecular dynamics (MD) simulation results by Ayuba *et al.* [63] and Halonen [64] carried out at  $T \approx 0.7$  and  $T = 0.6$ , respectively. In these direct nucleation simulation the cluster growth take place unrestricted by following the time evolution of a large gaseous system. Once a number of post-critical clusters have emerged, the formation free energies can be retrieved from the cluster distribution data. In Fig. 11, the MD simulated critical cluster formation free energies are compared against the corresponding  $W_N^{\text{hrm}}$  and  $W_N^{\text{Lahrm}}$ . Throughout the critical cluster range, from  $N = 21$  up to 72, the huge discrepancy between the harmonic approximation and the simulations is almost completely amended by the introduced high temperature expression, Eq. 19, using the system parameters  $\alpha$  and  $\beta$ . The anharmonic approximation, of course, is coarse compared to the model employing explicit  $k > 0$  conformers and their properties, but statistically the estimated values deviate only by  $\pm 4k_B T$  from  $W^{\text{MD}}$ .

## B. Colloidal cluster populations

A very different class of clusters is colloidal clusters made of equal-size micrometer-scale particles[7]. Even here, at mesoscale, the globally anharmonic statistical model can be used to interpret observed abundance spectra of colloidal conformers. Due to their characteristically narrow range of attraction[6, 7], the binding energies of colloidal clusters are rather close to the number of close contacts with particles and the TS barriers are roughly equal to multiples of the potential well depth (i.e.,  $\Delta E^{\ddagger \text{min}} = 1\epsilon, 2\epsilon, \dots$ ). While vibrational anharmonicity has relatively small effect on overall free energy of a colloidal cluster, due to the differences in conformer-specific  $\Delta E^{\ddagger \text{min}}$  one can expect that the local anharmonicity affects the conformer populations and their distributions.

So, another possible classical-regime test for the theory is to reproduce experimentally observed populations for colloidal clusters between sizes  $N = 6$  and 10[6].

## C. Atmospheric and molecular clusters

In Earth's atmosphere, certain vapor molecules have the ability to forge sufficiently stable intermolecular bonds, forming clusters that grow into larger particles and impact the planet's climate in various ways[65]. Unlike the systems discussed in this paper, atmospheric clusters comprise multiple distinct molecules and exhibit quantum mechanical properties. Nevertheless, the thermochemical quantities for these clusters are derived using an approach outlined in Sec. II A[5], using conformer data obtained from quantum chemical calculations. Ensuring the accuracy of free energy calculations is crucial, particularly given the high sensitivity of climate models to the predicted stability of nanometer-sized particles[66]. Moreover, atmospheric clustering rates, also, are exponentially dependent on the free energy of a critical cluster composition[27].

If computational resources allow for high-level electronic structure calculations for multiple conformers, global anharmonicity in atmospheric clusters can be and has been assessed[13]. However, due to their inherent complexity, atmospheric systems are likely dominated by the global minimum energy conformer, resulting in a global anharmonic effect of about 1 kcal/mol[13]. The significance of local anharmonicity, on the other hand, is even more pronounced for molecular clusters, given that the number of intermolecular degrees of freedom is  $6N - 6$ . It is worth cautioning that some of these degrees of freedom might be better treated as free rotors[67]. While the model for local anharmonicity in this work is limited to classical systems, the approach given in Sec. II C is equally applicable to discrete energy states (allowing determination of partition functions, e.g. through direct counting). However, intermolecular interactions in atmospheric clusters are not as strong as chemical bonds, and their frequencies are relatively low ( $\omega \lesssim 1000 \text{ cm}^{-1}$ ). Consequently, vibrational motion can be considered within the classical limit ( $k_B T/\hbar \approx 1300 \text{ cm}^{-1}$  at room temperature), and Eq. 7 most likely provides a good estimate for the anharmonic adjustment.

Obtaining precise conformer-wise transition state information for atmospheric clusters requires specific calculations that may not always be readily available. Nevertheless, even in the absence of such detailed data, our suggested approach provides an upper limit for the potential overestimation of cluster free energies. This knowledge is particularly relevant for making predictions in environments with low partial vapor pressures, where critical clusters are composed of multiple molecules ( $\sim 10$ ). Moreover, as quantum chemical calculations extend to even larger and more complex cluster sizes – demonstrated by recent studies, such as those by Engsvang *et al.*[68] examining clusters up to 30 or 60 molecules – the question of the level of local anharmonicity becomes increasingly pertinent.

## VI. CONCLUSIONS

The influence of anharmonicity becomes particularly pronounced in large cluster systems, where the exponential growth in the number of unique molecular arrangements amplifies global anharmonicity. Additionally, the weak intermolecular bonds within these systems may display notable local anharmonicity. While one approach to address the anharmonicity of individual vibrations involves adjusting frequencies using scaling factors[14, 19–21], it is crucial to recognize that, in the presence of significant thermal agitation on relatively low-frequency modes, the thermodynamic contribution of vibrations must be accurately assessed through the corresponding partition function.

In this study, we systematically characterize the global and local anharmonicity of small clusters interacting via varying ranges of attraction. With our findings, we introduce and validate a novel statistical mechanical approach for calculating the partition function, allowing us to accurately assess anharmonicities in the modes of vibrational motion. Our tests reveal that our model outperforms, both qualitatively and quantitatively, the harmonic approximation in predicting cluster free energies and potential energy. Two gratifying results underscore the efficacy of our approach: The improvement is achieved with the addition of only one input parameter, the interconfigurational transition state barrier height,  $\Delta E^\ddagger_{\text{min}}$ . And, for systems with  $\Delta E^\ddagger_{\text{min}}/k_B T \lesssim 2$ , our model predicts a maximum entropy enhancement of approximately  $k_B/4$  per vibrational mode.

To achieve even closer agreement with experiments, incorporating local anharmonicity, as outlined here, is anticipated. Work in progress is investigating the impact of anharmonicity on the modeling of both simple systems, such as homogeneous argon and water nucleation, which have abundant controlled experimental data, and atmospheric systems with broader scientific significance.

- 
- [1] K. K. Irikura and D. J. Frurip, Computational thermochemistry (ACS Publications, 1998).
  - [2] J. W. Ochterski, Thermochemistry in gaussian, Gaussian Inc **1**, 1 (2000).
  - [3] D. Wales *et al.*, *Energy landscapes: Applications to clusters, biomolecules and glasses* (Cambridge University Press, 2003).
  - [4] M. Hoare and P. Pal, Physical cluster mechanics: statistical thermodynamics and nucleation theory for monatomic systems, *Adv. Phys.* **24**, 645 (1975).
  - [5] J. Elm, J. Kubečka, V. Besel, M. J. Jääskeläinen, R. Halonen, T. Kurtén, and H. Vehkamäki, Modeling the formation and growth of atmospheric molecular clusters: A review, *J. Aerosol Sci.* **149**, 105621 (2020).
  - [6] G. Meng, N. Arkus, M. P. Brenner, and V. N. Manoharan, The free-energy landscape of clusters of attractive hard spheres, *Science* **327**, 560 (2010).
  - [7] M. E. Cates and V. N. Manoharan, Celebrating Soft Matter’s 10th anniversary: Testing the foundations of classical entropy: colloid experiments, *Soft Matter* **11**, 6538 (2015).
  - [8] J. A. Joseph, K. Röder, D. Chakraborty, R. G. Mantell, and D. J. Wales, Exploring biomolecular energy landscapes, *Chem. Comm.* **53**, 6974 (2017).
  - [9] F. H. Stillinger, *Energy landscapes, inherent structures, and condensed-matter phenomena* (Princeton University Press, 2015).

- [10] M. E. Tuckerman, *Statistical mechanics: theory and molecular simulation* (Oxford university press, 2023).
- [11] J. P. Doye and D. J. Wales, Calculation of thermodynamic properties of small Lennard-Jones clusters incorporating anharmonicity, *J. Chem. Phys.* **102**, 9659 (1995).
- [12] S. Kathmann, G. Schenter, and B. Garrett, The critical role of anharmonicity in aqueous ionic clusters relevant to nucleation, *J. Phys. Chem. C* **111**, 4977 (2007).
- [13] L. Partanen, H. Vehkamäki, K. Hansen, J. Elm, H. Henschel, T. Kurtén, R. Halonen, and E. Zapadinsky, Effect of conformers on free energies of atmospheric complexes, *J. Phys. Chem. A* **120**, 8613 (2016).
- [14] L. Partanen, V. Hänninen, and L. Halonen, Effects of global and local anharmonicities on the thermodynamic properties of sulfuric acid monohydrate, *J. Chem. Theory Comput.* **12**, 5511 (2016).
- [15] D. J. McGinty, Vapor phase homogeneous nucleation and the thermodynamic properties of small clusters of argon atoms, *J. Chem. Phys.* **55**, 580 (1971).
- [16] J. Burton, Vibrational frequencies and entropies of small clusters of atoms, *J. Chem. Phys.* **56**, 3133 (1972).
- [17] J. Burton, The configurational contribution to the free energy of small face centered cubic clusters, *Chem. Phys. Lett.* **17**, 199 (1972).
- [18] J. Burton, Free energy of small face centred cubic clusters of atoms, *J. Chem. Soc. Farad. T. 2* **69**, 540 (1973).
- [19] J. Pople, H. Schlegel, R. Krishnan, D. Defrees, J. Binkley, M. Frisch, R. Whiteside, R. Hout, and W. Hehre, Molecular orbital studies of vibrational frequencies, *Int. J. Quantum Chem.* **20**, 269 (1981).
- [20] T. Kurten, M. Noppel, H. Vehkamäki, M. Salonen, and M. Kulmala, Quantum chemical studies of hydrate formation of H<sub>2</sub>SO<sub>4</sub> and HSO<sub>4</sub>, *Boreal Environment Research* **12**, 431 (2007).
- [21] S. M. Kathmann, C. J. Mundy, G. K. Schenter, T. Autrey, P. C. Aeberhard, B. David, M. O. Jones, and T. Ramirez-Cuesta, Understanding vibrational anharmonicity and phonon dispersion in solid ammonia borane, *J. Phys. Chem. C* **116**, 5926 (2012).
- [22] A. D. Isaacson and D. G. Truhlar, Vibrational partition functions calculated from limited information, *J. Chem. Phys.* **80**, 2888 (1984).
- [23] P. Haarhoff, The density of vibrational energy levels of polyatomic molecules, *Mol. Phys.* **7**, 101 (1964).
- [24] T. Ishida, Vibrational reduced partition function of the morse oscillator, *J. Chem. Phys.* **61**, 3009 (1974).
- [25] M. Strelakov, An accurate closed-form expression for the partition function of morse oscillators, *Chem. Phys. Lett.* **439**, 209 (2007).
- [26] M. C. Holmes-Cerfon, Enumerating rigid sphere packings, *SIAM Rev.* **58**, 229 (2016).
- [27] R. Halonen, A consistent formation free energy definition for multicomponent clusters in quantum thermochemistry, *J. Aerosol Sci.* **162**, 105974 (2022).
- [28] G. Herzberg, *Molecular Spectra and Molecular Structure-Volume 2: Infrared and Raman Spectra of Polyatomic Molecules* (Princeton, NJ, Van Nostrand Reinhold Company, 1960).
- [29] K. Oh and X. C. Zeng, A small-system ensemble Monte Carlo simulation of supersaturated vapor: Evaluation of barrier to nucleation, *J. Chem. Phys.* **112**, 294 (2000).
- [30] J. Wedekind, A.-P. Hyvärinen, D. Brus, and D. Reguera, Unraveling the “pressure effect” in nucleation, *Phys. Rev. Lett.* **101**, 125703 (2008).
- [31] R. Halonen, E. Zapadinsky, and H. Vehkamäki, Deviation from equilibrium conditions in molecular dynamic simulations of homogeneous nucleation, *J. Chem. Phys.* **148**, 164508 (2018).
- [32] D. McGinty, The single-configuration approximation in the calculation of the thermodynamic properties of microcrystalline clusters, *Chem. Phys. Lett.* **13**, 525 (1972).
- [33] J. Burton, Configuration, energy, and heat capacity of small spherical clusters of atoms, *J. Chem. Phys.* **52**, 345 (1970).
- [34]  $\mathcal{D}(x) = \exp(-x^2) \int_0^x dy \exp(y^2)$ .
- [35] K. D. Ball and R. S. Berry, Realistic master equation modeling of relaxation on complete potential energy surfaces: Partition function models and equilibrium results, *J. Chem. Phys.* **109**, 8541 (1998).
- [36] L. Trombach, R. S. Hoy, D. J. Wales, and P. Schwerdtfeger, From sticky-hard-sphere to Lennard-Jones-type clusters, *Phys. Rev. E* **97**, 043309 (2018). The cluster configurations can be accessed online: The Cambridge Cluster Database, D. J. Wales, J. P. K. Doye, A. Dullweber, M. P. Hodges, F. Y. Naumkin F. Calvo, J. Hernández-Rojas and T. F. Middleton, URL <http://www-wales.ch.cam.ac.uk/CCD.html>.
- [37] M. G. Noro and D. Frenkel, Extended corresponding-states behavior for particles with variable range attractions, *J. Chem. Phys.* **113**, 2941 (2000).
- [38] R. J. Sadus, Vapor-liquid equilibria and cohesive  $r^{-4}$  interactions, *J. Chem. Phys.* **153**, 204504 (2020).
- [39] G. A. Vliegthart and H. N. W. Lekkerkerker, Predicting the gas-liquid critical point from the second virial coefficient, *J. Chem. Phys.* **112**, 5364 (2000), <https://doi.org/10.1063/1.481106>.
- [40] A. Garrett, A closed form for the second virial coefficient of the Lennard-Jones gas, *J. Phys. A: Math. Gen.* **13**, 379 (1980).
- [41] T. Lafitte, A. Apostolou, C. Avendaño, A. Galindo, C. S. Adjiman, E. A. Müller, and G. Jackson, Accurate statistical associating fluid theory for chain molecules formed from Mie segments, *J. Chem. Phys.* **139** (2013).
- [42] G. Vliegthart, J. Lodge, and H. Lekkerkerker, Strong weak and metastable liquids structural and dynamical aspects of the liquid state, *Physica A* **263**, 378 (1999).
- [43] M. Holmes-Cerfon, S. J. Gortler, and M. P. Brenner, A geometrical approach to computing free-energy landscapes from short-ranged potentials, *P. Natl. A. Sci. USA* **110**, E5 (2013).
- [44] D. J. Wales and J. P. Doye, Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms, *J. Phys. Chem. A* **101**, 5111 (1997).
- [45] J. P. Doye and D. J. Wales, Saddle points and dynamics of Lennard-Jones clusters, solids, and supercooled liquids, *J.*

- Chem. Phys. **116**, 3777 (2002).
- [46] J. Stevenson, V. Rühle, S. Martiniani, and K. Schrenk, Python energy landscape explorer, GitHub <https://github.com/pele-python/pele> (2016).
- [47] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. **117**, 1 (1995).
- [48] H. Vehkamäki and I. J. Ford, Critical cluster size and droplet nucleation rate from growth and decay simulations of Lennard-Jones clusters, J. Chem. Phys. **112**, 4193 (2000).
- [49] J. Merikanto, H. Vehkamäki, and E. Zapadinsky, Monte Carlo simulations of critical cluster sizes and nucleation rates of water, J. Chem. Phys. **121**, 914 (2004).
- [50] A. Lauri, J. Merikanto, E. Zapadinsky, and H. Vehkamäki, Comparison of Monte Carlo simulation methods for the calculation of the nucleation barrier of argon, Atmos. Res. **82**, 489 (2006).
- [51] J. Neirotti, F. Calvo, D. L. Freeman, and J. Doll, Phase changes in 38-atom Lennard-Jones clusters. I. A parallel tempering study in the canonical ensemble, J. Chem. Phys. **112**, 10340 (2000).
- [52] F. H. Stillinger Jr, Rigorous basis of the Frenkel-Band theory of association equilibrium, J. Chem. Phys. **38**, 1486 (1963).
- [53] J. Merikanto, E. Zapadinsky, A. Lauri, I. Napari, and H. Vehkamäki, Connection between the virial equation of state and physical clusters in a low density vapor, J. Chem. Phys. **127**, 104303 (2007).
- [54] H. W. Woolley, The representation of gas properties in terms of molecular clusters, J. Chem. Phys. **21**, 236 (1953).
- [55] J. C. Maxwell, L. on the calculation of the equilibrium and stiffness of frames, Philos. Mag. **27**, 294 (1864).
- [56] F. H. Stillinger, Exponential multiplicity of inherent structures, Phys. Rev. E **59**, 48 (1999).
- [57] J. P. Doye, Network topology of a potential energy landscape: A static scale-free network, Phys. Rev. Lett. **88**, 238701 (2002); J. P. Doye and C. P. Massen, Characterizing the network topology of the energy landscapes of atomic clusters, J. Chem. Phys. **122**, 084105 (2005). The transition state network data is available online: <http://doye.chem.ox.ac.uk/networks/LJn.html>.
- [58] J. A. Pople, Nobel lecture: Quantum chemical models, Rev. Mod. Phys. **71**, 1267 (1999).
- [59] J. Wedekind and D. Reguera, What is the best definition of a liquid cluster at the molecular scale?, J. Chem. Phys. **127** (2007).
- [60] C. Nie, J. Geng, and W. Marlow, Formation free energy of an  $i$ -mer at spinodal, J. Chem. Phys. **154** (2021).
- [61] P. R. ten Wolde and D. Frenkel, Computer simulation study of gas-liquid nucleation in a Lennard-Jones system, J. Chem. Phys. **109**, 9901 (1998).
- [62] H. Vehkamäki, *Classical nucleation theory in multicomponent systems* (Springer Science & Business Media, 2006).
- [63] S. Ayuba, D. Suh, K. Nomura, T. Ebisuzaki, and K. Yasuoka, Kinetic analysis of homogeneous droplet nucleation using large-scale molecular dynamics simulations, J. Chem. Phys. **149** (2018).
- [64] R. Halonen, Comments on “The proper view of cluster free energy in nucleation theories” by Runlong Cai and Juha Kangasluoma, Aerosol Sci. Tech. **56**, 976 (2022), <https://doi.org/10.1080/02786826.2022.2114310>.
- [65] M. Kulmala, J. Kontkanen, H. Junninen, K. Lehtipalo, H. E. Manninen, T. Nieminen, T. Petäjä, M. Sipilä, S. Schobesberger, P. Rantala, *et al.*, Direct observations of atmospheric aerosol nucleation, Science **339**, 943 (2013).
- [66] J. Tröstl, W. K. Chuang, H. Gordon, M. Heinritzi, C. Yan, U. Molteni, L. Ahlm, C. Frege, F. Bianchi, R. Wagner, *et al.*, The role of low-volatility organic compounds in initial particle growth in the atmosphere, Nature **533**, 527 (2016).
- [67] S. Grimme, Supramolecular binding thermodynamics by dispersion-corrected density functional theory, Chem-Eur. J. **18**, 9955 (2012).
- [68] M. Engsvang and J. Elm, Modeling the binding free energy of large atmospheric sulfuric acid-ammonia clusters, ACS Omega **7**, 8077 (2022); M. Engsvang, J. Kubečka, and J. Elm, Toward modeling the growth of large atmospheric sulfuric acid-ammonia clusters, *ibid.* **8**, 34597 (2023); H. Wu, M. Engsvang, Y. Knattrup, J. Kubečka, and J. Elm, Improved configurational sampling protocol for large atmospheric molecular clusters, *ibid.* **8**, 45065 (2023).