

RESEARCH ARTICLE | SEPTEMBER 10 2024

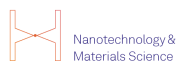
## A microscopic approach to crystallization: Challenging the classical/non-classical dichotomy

James F. Lutsko   ; Cédric Schoonen 



*J. Chem. Phys.* 161, 104502 (2024)

<https://doi.org/10.1063/5.0225658>



Nanotechnology &  
Materials Science



Optics &  
Photonics



Impedance  
Analysis



Scanning Probe  
Microscopy



Sensors



Failure Analysis &  
Semiconductors



Unlock the Full Spectrum.  
From DC to 8.5 GHz.

Your Application. Measured.

[Find out more](#)



# A microscopic approach to crystallization: Challenging the classical/non-classical dichotomy

Cite as: J. Chem. Phys. 161, 104502 (2024); doi: 10.1063/5.0225658

Submitted: 26 June 2024 • Accepted: 26 August 2024 •

Published Online: 10 September 2024



View Online



Export Citation



CrossMark

James F. Lutsko<sup>a)</sup>  and Cédric Schoonen 

## AFFILIATIONS

Center for Nonlinear Phenomena and Complex Systems CP 231, Université Libre de Bruxelles, Blvd. du Triomphe, 1050 Brussels, Belgium

<sup>a)</sup> Author to whom correspondence should be addressed: [jlutsko@ulb.be](mailto:jlutsko@ulb.be). URL: <http://www.lutsko.com>

## ABSTRACT

We present a fundamental framework for the study of crystallization based on a combination of classical density functional theory and fluctuating hydrodynamics that is free of any assumptions regarding order parameters and that requires no input other than molecular interaction potentials. We use it to study the nucleation of both droplets and crystalline solids from a low-concentration solution of colloidal particles using two different interaction potentials. We find that the nucleation pathways of both droplets and crystals are remarkably similar at the early stages of nucleation until they diverge due to a rapid ordering along the solid pathways in line with the paradigm of “non-classical” crystallization. We compute the unstable modes at the critical clusters and find that despite the non-classical nature of solid nucleation, the size of the nucleating clusters remains the principle order parameter in all cases, supporting a “classical” description of the *dynamics* of crystallization. We show that nucleation rates can be extracted from our formalism in a systematic way. Our results suggest that in some cases, despite the non-classical nature of the nucleation pathways, classical nucleation theory can give reasonable results for solids but that there are circumstances where it may fail. This contributes a nuanced perspective to recent experimental and simulation work, suggesting that important aspects of crystal nucleation can be described within a classical framework.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0225658>

## I. INTRODUCTION

Crystallization is an everyday phenomenon for which a fundamental physical description remains elusive. The key first step is nucleation: the formation, via thermal fluctuations, of a sufficiently large cluster of solid as to be thermodynamically stable. Classical Nucleation Theory (CNT),<sup>1</sup> based on macroscopic concepts, such as surface tension and the assumption that the size of a cluster is the only order parameter, was developed in the first half of the twentieth century and continues to define the way we approach the phenomena. However, its extension to crystallization from solution—where both ordering and cluster-building are important—has always been problematic since, after the size of a cluster, any choice of additional order parameters is somewhat arbitrary. Furthermore, modern advances in experimental and simulation capabilities have revealed many so-called non-classical aspects of solid nucleation,<sup>2–9</sup> such as multistep nucleation pathways (e.g., crystallization via the formation of dense unstructured clusters followed by ordering with little mass change) and the frequent

occurrence of nucleation precursors—small clusters that appear to play a role in nucleation but which are not expected to exist based on the classical considerations.

Theoretical efforts to go beyond the standard CNT go back a long way and include mathematical developments, such as the seminal work of Langer,<sup>10,11</sup> Talkner,<sup>12</sup> and others on multidimensional barrier crossing due to thermal fluctuations (the Kramers problem<sup>13</sup>), as well as physically motivated extensions of CNT to include order parameters beyond the cluster size<sup>7,14–17</sup> and attempts to accommodate multistep pathways within its framework.<sup>4,5,18–20</sup> Recently, it has been shown that the combination of classical Density Functional Theory (cDFT) and fluctuating hydrodynamics offers a relatively fundamental, if still mesoscopic, description of nucleation able to predict *a priori* the structure of small clusters (including crystalline clusters) and of non-classical pathways.<sup>21,22</sup> The advantages of this approach are that (1) the structure of small clusters is accurately described by cDFT even down to molecular length scales<sup>23</sup> and (2) no order parameters are introduced. In particular, there are no *a priori* assumptions about the crystalline structure, nor the nature

of critical clusters, nor the form of the nucleation pathways—all emerge as outputs of a theory only based on a molecular interaction potential.

While this approach, which we call mesoscopic nucleation theory, has many advantages, it remains incomplete as it does not address the aspects of nucleation that are of most practical importance and are the focus of most experimental and simulation investigations: namely, the dynamics of the process and, most particularly, the nucleation rates. The goal of the present work is to fill this gap and to thus give a complete theory of nucleation that characterizes nucleation pathways, nucleation rates, and the emergent order parameters, resulting in a complete description of nucleation of any system, including crystals, at a much more fundamental level than that of CNT. This involves three important contributions. The first is to describe the calculation of the unstable modes: the unstable eigenvalue and corresponding eigenvectors of the critical cluster. We note that the relevant quantities are not those of the Hessian of the free energy as is often assumed but rather those of the dynamical matrix consisting of the product of the kinetic coefficients (determined by the dynamics of fluctuations) and the free energy Hessian. In our examples, we demonstrate the remarkable result that, taking this into account, the natural order parameter—i.e., the dynamics projected onto the unstable mode—corresponds almost entirely to the mass of the cluster for the nucleation of both liquid and crystalline clusters, thus suggesting that both processes are (at least dynamically) “classical.” This is puzzling as the pathways for the solid clusters are clearly non-classical, starting first as droplets and with crystalline ordering only forming late in the nucleation process as previously reported.<sup>22</sup>

Our second contribution, which allows us to make sense of this apparent paradox, is the introduction of what we term the “kinetic distance” as the natural reaction coordinate of the process. This is a natural distance measure in density space dictated by, and therefore closely related to, the dynamics of the process so that large kinetic distances imply long times and small distances imply short times. We show that when the pathways are expressed in terms of this quantity, the pathways leading to both droplets and solid clusters at the same thermodynamic conditions are nearly indistinguishable except that the solids show almost instantaneous jumps corresponding to the development of crystalline order at constant mass. Thus, the kinetic distance reveals that the dynamics are largely classical (i.e., as assumed in CNT) with very short bursts of ordering characterizing the non-classical parts of the process. In our earlier, non-dynamical, work on the nucleation pathways, the kinetic-distance was not known and we instead used the Euclidean distance in density space, which is impossible to interpret physically.<sup>22</sup> The dynamics of the process are therefore dominated by the long “classical” periods of mass accumulation with the short non-classical bursts of ordering being of secondary importance. The process thus possesses both classical and non-classical characteristics simultaneously.

The third and final contribution is a means of calculating nucleation rates. We do this with two different methods. In the first, we follow the reasoning introduced by Auer and Frenkel<sup>24</sup> and give an expression that makes use of the information obtained for the unstable eigenvectors and eigenvalues of the critical cluster. This is expected to be reasonable for droplet nucleation, which is already an advance beyond previous work. However, we note that for solids,

this expression can be misleading if the critical cluster happens to appear during one of the bursts of ordering, and so we also introduce a more heuristic expression involving integration along the entire nucleation pathway and demonstrate that it is consistent with the first method for droplets and gives more physically reasonable results for solids too. An incidental, but satisfying result of our analysis is a theoretical expression for the attachment and detachment frequencies for diffusion-limited nucleation that play a central role in CNT but for which no such theoretical expression has previously been developed.

These elements—the unstable modes, the kinetic distance, and the expressions for nucleation rates—together with the basic framework of cDFT and fluctuating hydrodynamics<sup>21</sup> and the determination of the nucleation pathway as the most likely path of the stochastic dynamics<sup>22</sup> give all of the tools necessary for a theoretical description of not only homogeneous nucleation of droplets and crystals as studied in this work but also heterogeneous nucleation. Being based on hydrodynamics, non-equilibrium effects, such as flows and temperature gradients, can also be described in a natural manner.

## II. MESOSCOPIC NUCLEATION THEORY

We focus attention here on a two-species system consisting of a larger species in a bath of small molecules: for example, macromolecules in solution or colloidal systems. Our considerations are, in particular, relevant to the important problem of nucleation from solution, including precipitation of crystals. Treating the small molecules implicitly and working in the overdamped limit,<sup>21</sup> the dynamics of the local number density  $n_t(\mathbf{r})$  of the larger species is modeled with the stochastic differential equation,

$$\frac{\partial}{\partial t} \widehat{n}_t(\mathbf{r}) = D \nabla \cdot \widehat{n}_t(\mathbf{r}) \nabla \frac{\delta \beta F^{(\text{dyn})}[\widehat{n}_t]}{\delta \widehat{n}_t(\mathbf{r})} + \nabla \cdot \sqrt{2D \widehat{n}_t(\mathbf{r})} \widehat{\xi}_t(\mathbf{r}), \quad (1)$$

where the caret indicates stochastic quantities,  $D$  is the tracer diffusion constant for a single monomer of the large species in the bath,  $F^{(\text{dyn})}[\widehat{n}_t]$  is the dynamical free energy functional for the large species, and  $\widehat{\xi}_t(\mathbf{r})$  is Gaussian-distributed white noise, representing the effect of the bath. In addition,  $\beta = 1/k_B T$ , where  $k_B$  is Boltzmann's constant and  $T$  is the temperature. Note that here and below, we use the generic term “monomer” to refer to what could be an atom, molecule, or colloid particle. This minimal description can be derived from the microscopic equations of motion by first using projection operator techniques to obtain fluctuating hydrodynamics<sup>25,26</sup> and then adding additional dissipative and fluctuating terms to take account of the bath followed by the overdamped limit.<sup>27</sup> The local density appearing in Eq. (1) is then understood to be a coarse-grained density, e.g., the instantaneous microscopic density averaged over small volumes. What we have termed the “the dynamical free energy” is usually the thermodynamic free energy of a system constrained (e.g., by the presence of external fields) to have the specified coarse-grained density, which can also be thought of as a local equilibrium free energy. An interpretation of this model is that density fluctuations on length scales smaller than the coarse-graining scale are accounted for in the dynamical free energy, and Eq. (1) describes fluctuations of the density on larger length scales.

Only formal expressions for the dynamical free energy are known, and to proceed, one must introduce a model. For this, we turn to classical Density Functional Theory (cDFT) where one of the main goals is the development of constrained free energy functionals.<sup>28,29</sup> The sophisticated models developed over several decades are routinely used to determine the free energy, structure, and thermodynamic properties of inhomogeneous systems, including solid clusters.<sup>23,30</sup> The only inputs required are the interaction potentials of the molecules, and no *a priori* assumptions are made about crystal structures, lattice parameters, etc. Despite their somewhat different formal origins (cDFT is a strictly equilibrium theory), the cDFT free energy models are the best available examples of constrained free energy functionals capable of describing multiphase systems down to molecular length scales and as such are the only feasible option to use in the present context.

In an early discussion of this combination of fluctuating hydrodynamics and cDFT for the study of nucleation,<sup>21,31</sup> it was noted that this type of dynamical model has been used to model fluctuations in Brownian systems, in general,<sup>32,33</sup> and, particularly, the Dean–Kawasaki<sup>34,35</sup> model used in the study of the glass transition; see, e.g., Ref. 36. In this context, our main innovation is using a microscopic free energy functional capable of describing the crystal structure at the atomistic level and our approach to its analysis as discussed below. Another closely related class of models are the diffuse interface model for multi-phase systems,<sup>37</sup> the Cahn–Hilliard model<sup>38,39</sup> for binary systems, and the Phase-Field (PF) and Phase-Field Crystal (PFC) models<sup>40</sup> of which PFC is the closest in spirit to ours. These can be understood in different ways: either as heuristic models for which one postulates a Landau-like free energy involving gradient terms as well as an appropriate dynamics or as approximations derived from more fundamental approaches, such as Eq. (1). For PFC, this connection has recently been developed in detail by van Teefelen *et al.*<sup>41</sup> where it was noted that a crucial assumption is that the free energy can be expanded in gradients of the density and the expansion truncated at low order. This can only be justified when the density does not vary too fast over microscopic length scales. For crystallization, as illustrated below, one necessarily finds density changes of orders of magnitude over an atomic diameter, thus stressing this assumption. Furthermore, in order to stabilize a crystalline structure,<sup>42,43</sup> various other simplifications are introduced, such as the use of a small selection of reciprocal lattice vectors in order to stabilize the solid phase with the further implication of fixed lattice constants (no lattice relaxation at surfaces, etc.). As a result of these approximations, its predictive power, at least at the microscopic level, is compromised.<sup>41,44</sup> Furthermore, the representation of three phase systems within this framework is non-trivial and the subject of on-going research.<sup>45</sup> One of our main goals is to work with the full microscopic model so as to avoid these issues, and in Sec VI, we return to the differences between our approach and PFC.

### III. IMPLEMENTATION

Our calculations are based on a discretization of this model. We label the Cartesian grid points with a super-index  $I \leftrightarrow (i_x, i_y, i_z)$  and the values of the density on the grid as  $\widehat{n}_t^I$ . The grid indices take on values  $0 \leq i_a < N_a$  for  $a = x, y, z$ , the super-index has values  $0 \leq I < N_x N_y N_z$ , and the grid spacing is taken to be  $\Delta$ . The

operator  $\nabla \cdot \widehat{n}(\mathbf{r}) \nabla$  becomes a matrix, which we call  $-g^{IJ}(\widehat{n})$  that plays the role of a position-dependent diffusivity (or, more abstractly, kinetic prefactor; see Appendix A for details). The stochastic differential equation becomes

$$\frac{\partial}{\partial t} \widehat{n}_t^I = -\frac{D}{\sigma^2} g^{IJ}(\widehat{n}_t) \frac{\partial \beta \Omega(\widehat{n}_t)}{\partial \widehat{n}_t^J} + \sqrt{\frac{2D}{\sigma^2}} \mathbf{q}^{IK}(\widehat{n}_t) \cdot \widehat{\xi}_t^K, \quad (2)$$

where we use the Einstein convention of summing over repeated indices and we let  $\widehat{n}_t$ , with no indices, represent the entire collection of values  $\widehat{n}_t^I$  and where  $\sigma$  is a length scale associated with the interaction potential (see Appendix D). The first term on the right now refers to  $\Omega \equiv F - \mu N$ , which is the result of our use of fixed-density boundary conditions to model an open system (see Appendix B for more details). In the noise term, the amplitude and noise are both three-vectors at each lattice point. The amplitude of the noise is related to the coefficient of the gradient via  $\mathbf{q}^{IK} \cdot \mathbf{q}^{JK} = g^{IJ}$ , which is a fluctuation–dissipation relation, assuring that in equilibrium the system obeys a canonical distribution  $P(n) \sim \exp(-\beta \Omega(n))$ . Finally, and importantly, we will consider here only open systems for which the density on the boundaries of the computational cell is held constant: this is equivalent to setting the chemical potential for the system. In this representation, a pure solution corresponds to constant density throughout the system.<sup>21</sup>

As is commonly done, we will restrict ourselves to the weak-noise limit, which in some sense corresponds to low temperatures<sup>46</sup> for which the following properties have been established:<sup>21</sup> if the system makes a transition from one local minimum of the free energy to another, then the most likely path (MLP) (given the stochastic dynamics) passes through a saddle point and corresponds to the two steepest descent paths that result from perturbing the system at the saddle point in the direction of the unstable eigenvector (in positive and negative directions). In this way, the concept of the critical cluster (a saddle point on the free energy surface) and its critical role in the phase transition emerges naturally. The critical cluster  $n^*$  is defined by

$$\left. \frac{\partial \beta \Omega(n)}{\partial n^I} \right|_{n^*} = 0, \quad (3)$$

and expanding the stochastic equation about this value gives

$$\frac{\partial}{\partial t} \delta \widehat{n}_t^I = -\frac{D}{\sigma^2} g^{IJ}(n^*) \left( \frac{\partial^2 \beta \Omega(n)}{\partial n^J \partial n^K} \right)_{n^*} \delta \widehat{n}_t^K + \sqrt{\frac{2D}{\sigma^2}} \mathbf{q}^{IJ}(n^*) \cdot \widehat{\xi}_t^J \quad (4)$$

so that the unstable direction is a solution of the eigenvalue problem

$$g^{IJ}(n^*) \left( \frac{\partial^2 \beta \Omega(n)}{\partial n^J \partial n^K} \right)_{n^*} v^K = \lambda v^I \quad (5)$$

for  $\lambda < 0$ . The matrix in parentheses is the Hessian of the free energy, and the product of the kinetic factor times this matrix will be referred to henceforth as the dynamical matrix. Because the dynamical matrix is real, but not symmetric, there is, for each eigenvalue, a corresponding left eigenvector that solves

$$u_I g^{IJ} (n^*) \left( \frac{\partial^2 \beta \Omega(n)}{\partial n^I \partial n^K} \right)_{n^*} = \lambda u_K, \quad (6)$$

and the collection of left and right eigenvectors forms a bi-orthogonal set. While distinct, it is easy to see that the two eigenvectors are related by  $u_I g^{IJ} (n^*) \propto v^J$ . In fact, as discussed in Appendix C, we will use a normalization throughout this work according to which  $u_I g^{IJ} = v^J$  and  $u_I v^I = 1$ . It has been discussed elsewhere<sup>21</sup> that the structure of the stochastic equation imposes a Riemannian geometry on density space, with the matrix inverse of the diffusivity matrix, denoted with subscripts as  $g_{IJ}$ , playing the role of the metric on the space. For this reason, we use the standard notation of covariant and contravariant vectors related via the metric. This just means that for any (covariant) vector with lower indices,  $v_I$ , there is a corresponding (contravariant) vector with upper indices defined as  $v^I = g^{IJ} v_J$ .

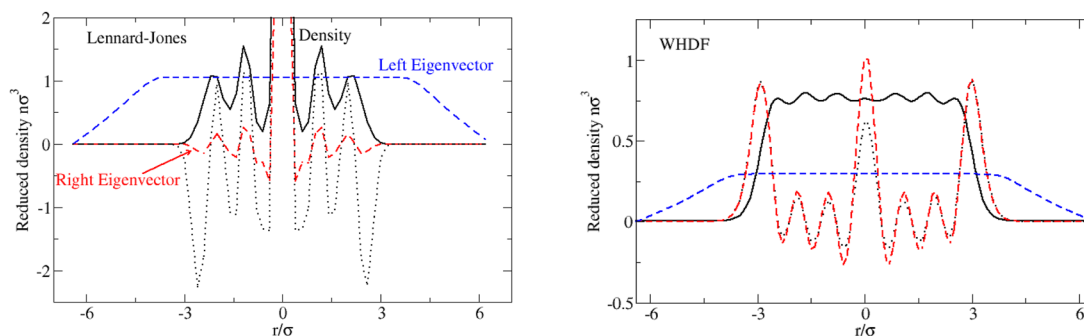
#### IV. NUCLEATION OF DROPLETS AND SOLIDS: ORDER PARAMETERS AND PATHWAYS

We will first illustrate this framework with calculations for the nucleation of dense droplets from a low-concentration, or “weak,” solution of particles interacting via a Lennard-Jones (LJ) potential and via a more short-ranged pair interaction introduced by Wang *et al.* to model colloidal interactions<sup>47</sup> and which we refer to as the WHDF potential (the explicit forms of both potentials are given in Appendix D). The WHDF interaction gives generally lower surface tensions than the LJ, thus allowing us to sample both the high (with LJ) and low (with WHDF) supersaturation regimes at similar computational expense. As stated above, what we have called the dynamical free energy is modeled by a cDFT functional, specifically the explicitly stable fundamental measure theory model<sup>48</sup> with the attractive part of the potential described by a mean field term. Our grid has 64 points in each dimension for the LJ potential and 128 for the WHDF potential, giving more than 250 thousand grid points overall in the former case and more than  $2 \times 10^6$  in the latter.

The discretization and constant-density boundary conditions used to mimic an open system are described in Appendixes A and C. While the dynamical matrix is far too large to be calculated explicitly, it turns out that the structure of the cDFT model allows for the computation of its product with any given vector to be performed efficiently using fast Fourier transforms. This is enough to allow for the use of sophisticated libraries for the calculation of a limited number of eigenpairs (including both left and right eigenvectors), and our calculations use SLEPC.<sup>49–52</sup>

Critical clusters were constructed (as described in Appendix D) for the Lennard-Jones system under conditions of high supersaturation ( $S \equiv \beta \Delta \mu = 4.00$ , where  $\Delta \mu$  is the excess chemical potential) and conditions of low supersaturation ( $S = 0.51$ ) for the WHDF system. Then, the largest several (up to 10) eigenvalues and eigenvectors were calculated. In the examples presented here, as well as many others, we find a single unstable (negative) eigenvalue. Figure 1 shows the droplet critical clusters as well as the left and right eigenvectors in the unstable direction. The density of the critical clusters oscillates, which is typical any time a liquid with strong short-ranged repulsion is confined: it represents layering of the fluid in alternating shells of high and low density. In the LJ droplet, a large density at the center of the cluster simply indicates a high probability of finding a particle there. The right eigenvectors, the unstable directions in density space, are complicated, with many structures corresponding to that of the density itself. However, the left eigenvectors are remarkably simple, being nearly flat within the clusters and then decreasing outside the clusters until vanishing at the boundary. (That it is not completely devoid of structure is shown in the supplementary material.)

The meaning of this simplicity becomes clear when one considers two points. First, since the eigenvectors of the dynamical matrix form a complete set, any local density,  $n$ , can be written as an expansion of the form  $n = \sum_{\alpha} c^{(\alpha)} v^{(\alpha)}$ , where the index  $\alpha$  labels the different eigenvectors (so it is not lattice index) and the coefficients are calculated from the corresponding left eigenvectors,  $c^{(\alpha)} = v_I^{(\alpha)} n^I$ . We let  $\alpha = -$  correspond to the unstable (negative



**FIG. 1.** The black line shows that density in a critical droplet as a function of distance from the center, for an open Lennard-Jones fluid with temperature  $k_B T/\epsilon = 0.3$  and supersaturation  $S = 4.00$  (left panel) and for a WHDF fluid with  $k_B T/\epsilon = 0.275$  and  $S = 0.51$  (right panel), where  $\epsilon$  and  $\sigma$  are the energy and length scales of the respective potentials. The oscillations are signs of packing of the fluid due to the confinement, and the large peak in the Lennard-Jones density indicates a high probability that a particle is in the center of the droplet. The dashed lines show the left and right eigenvectors for the unstable mode of the dynamical matrix, and the dotted line shows the same for the Hessian of the free energy, with no dynamics (left and right eigenvectors are therefore the same). The right eigenvectors (and those based on the Hessian) reflect the structure of the droplet, whereas the (dynamic) left eigenvector is very simple, corresponding to a probe of the mass of the droplet, a result only possible because of the mass-conserving dynamics. Note that the droplets are spherically symmetric, so the figures give complete information about the various quantities.



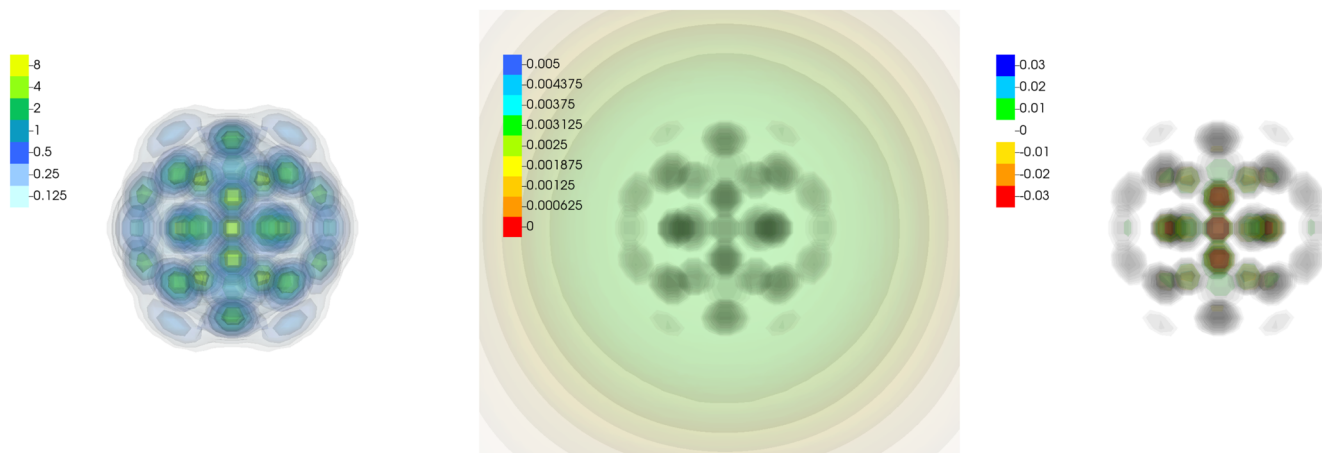
eigenvalue) mode so that the projection of the density onto the unstable direction is the coefficient  $c^{(-)}$ . The second observation is that the total mass of the system is calculated as the sum of the density at each point multiplied by the volume element,  $\sum_I n^I \Delta^3$ , and this can be written as  $M_I n^I$ , where the covariant vector  $M_I$  has all elements equal to  $\Delta^3$ , which would simply be a horizontal line in traversing the system in Fig. 1. The excess mass of the cluster is then  $\Delta N = M_I (n^I - n_0^I)$ , where  $n_0^I$  is the uniform density of the mother phase. Since the left eigenvector is (nearly) constant within the cluster, and the density outside the cluster is virtually the same as in the mother phase, it follows that  $v_i^{(-)} (n^I - n_0^I)$  is very nearly proportional to the excess mass as well. Thus, the projection of the local density onto the unstable direction is, up to an irrelevant additive constant, a measure of the excess mass of the cluster. Since the unstable direction is also the MLP for nucleation, the identification of the projection of the density onto this direction as the order parameter is intuitively appealing. Furthermore, Berezhkovskii and Szabo<sup>16</sup> (see also the discussion in ch. 18 of Peters<sup>17</sup>) have shown that this is the only choice of order parameter that leads to the correct nucleation rate in a one-dimensional description of nucleation. All of this together leads to the conclusion that the emergent order parameter at the critical cluster is, in both cases, the excess mass of the cluster as assumed in CNT. This conclusion is reassuring since one expects CNT to describe droplet nucleation.

Again using the procedures described in Appendix D, we have generated solid clusters under the same thermodynamic conditions as for the droplets. As previously observed,<sup>22</sup> crystallization pathways involve an initial critical cluster followed by a sequence of shallow minima and maxima corresponding to the formation of complete shells of solid as the cluster grows. Figure 2 shows the *first* critical cluster and the unstable eigenvectors associated with it for the LJ potential (see the [supplementary material](#) for corresponding images for the WHDF case). Roughly, half the mass is localized into “particles” and half is delocalized liquid-like fluid.

We recall that in this formalism there is no explicit representation of colloidal particles or of crystal lattices and that all such structure observed forms spontaneously during minimization of the free energy so that “particles” are really just locations where the local density (the probability of finding a colloidal particle) is very high, surrounded by a region with very low probability, thus indicating a localized particle. In this cluster, as well as in the WHDF case, these “particles” are located at the vertices of an icosahedron, with additional peaks in front of each one of its 20 facets. This hcp-like stacking of particles on the icosahedron facets is the first shell of what is sometimes referred to as an “anti-Mackay” structure. Structures with icosahedral symmetry have been observed in super-cooled liquids and glasses<sup>53–55</sup> and can be obtained when a liquid crystallizes in a confined environment.<sup>56,57</sup> The eigenvectors are shown in Fig. 2 as three-dimensional contour plots overlaid on the cluster image. One clearly sees the same structure as for the droplet: the right eigenvectors reflect the density of the cluster, whereas the left eigenvector is, quite surprisingly given the complexity of the structure, almost constant within the cluster, as in the case of the droplet. Once again, this suggests that the order parameter is closely related to the excess mass of the clusters.

We have calculated nucleation pathways (i.e., the most likely path) from the initial low-concentration state to the various critical clusters using the string method.<sup>22</sup> In brief, one creates an initial guess at the path consisting of some number (we typically use 30) images of the system with the density interpolated between the initial, uniform, state and the final critical cluster, which is then relaxed based on the deterministic dynamics, under the constraint that the distance between the images remains fixed.

To present the results, we introduce a natural measure of distance in density space. Note that the deterministic dynamics is driven by the gradient of the free energy, combined with the kinetic factor,  $g^J(n)$ , in Eq. (2). This can be interpreted as gradient descent in a curved space in which one measures the distance



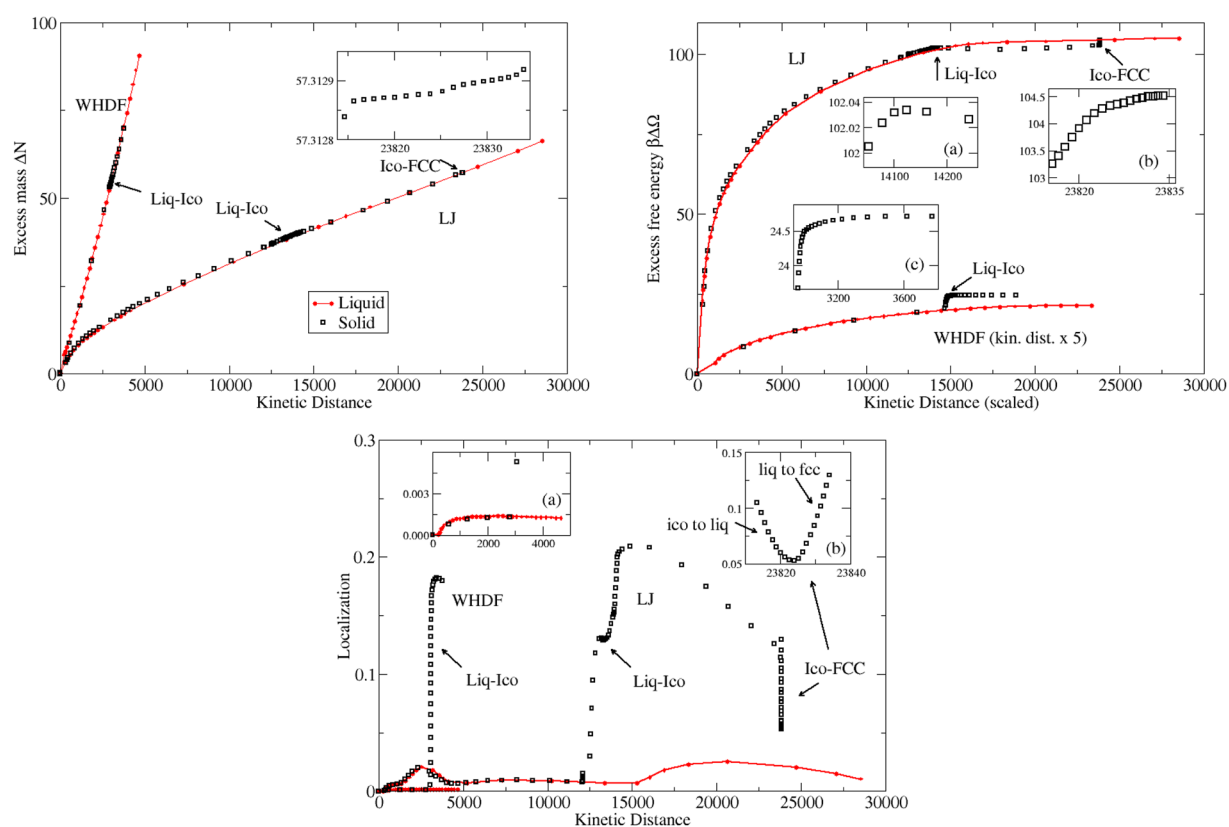
**FIG. 2.** Local density of the solid-like critical cluster (left, displayed on a log-scale) and the left and right eigenvectors (center and right, both linear scales) for the unstable mode of the LJ cluster. The cluster has an excess mass of about 40 particles of which about half are localized into fixed “atoms” and about half in a fluid state. In the eigenvector images, the local density is represented in gray and the eigenvectors in color. The eigenvectors are qualitatively very similar as for the droplet: the left eigenvector is nearly constant inside the solid cluster, while the right eigenvector is similar to the density as evidenced by the fact that one only observes color on the localized gray sites.

between a density distribution  $n^l$  and another  $n^l + dn^l$ , infinitesimally close, as  $ds^2 = dn^l g_{ij} dn^l$  (we will call this dimensionless quantity the “kinetic distance” and details concerning its calculation are given in Appendix F).

The excess mass (number of particles) and free energy of the clusters as they evolve along the nucleation pathway are shown in Fig. 3 using the kinetic distance along the pathway as the independent variable. The nucleation pathway for the LJ cluster is complex: first, an icosahedral critical cluster is formed followed by a solid–solid transition to an FCC crystalline lattice. For the WHDF interaction, our calculation terminates at an icosahedral structure and further calculations would be necessary to determine if an eventual transition to a close-packed structure occurs (as one would expect).

These pathways exhibit several remarkable features. In particular, the kinetic distance along the pathways corresponds

monotonically, and almost linearly, to excess mass in both cases. Even more surprising, the excess energy of the eventual solid and liquid clusters is almost identical until, at a given point the (eventual) solid-like cluster shows an abrupt jump in free energy, almost a discontinuity. During this process, the excess mass is almost constant and the main difference is that the localization into “particles” occurs. This leads to the unexpected conclusion that while ordering does occur independently of mass accumulation, in a kind of three-step process of mass increase, then ordering, and then mass increase, the ordering is very fast (in the sense that the kinetic distance between the ordered and unordered clusters is very small). Nevertheless, the unstable eigenvalues at the critical cluster differ significantly for the two processes as can be seen from the values given in Table I. Table I also gives the unstable eigenvalues for the corresponding free energy Hessians and illustrates the differences between the two calculations.



**FIG. 3.** Excess mass (total number of particles compared to the initial uniform state, left panel), excess free energy (center panel), and “localization” (right panel) as functions of kinetic distance along the most likely paths for both the solid cluster and the liquid droplet for both potentials. Black symbols are the results for the “solid” pathways, and red lines are the droplet pathway. The kinetic distance for the WHDF system is multiplied by a factor of 5 in the middle panel so as to improve visibility. All pathways end in a critical cluster; in both solid pathways, an initially liquid-like droplet first transforms into an icosahedral cluster. In the case of LJ, this is followed by a solid–solid transition to the FCC structure. The localization is a measure of order within the cluster (as described in Appendix E) and shows that the near discontinuity in the free energy corresponds to a sharp increase into solid-like localization of the density and then the reorganization into an FCC structure. The left and center panels both include insets showing a detailed view of the reorganization of the liquid droplet into an icosahedral solid (inset a). The center panel also shows similar zooms of the ico–fcc transition in LJ (b) and the liq–ico transition in the WHDF system (c). The insets in the right hand panel show (a) a close-up of the divergence between the droplet and solid pathways for the WHDF interaction and (b) the extreme point of the transition from icosahedral order to FCC order via a dis-ordered, fluid-like intermediate state.

**TABLE I.** Unstable eigenvalues of the Hessian of the free energy and of the dynamical matrix for the two interaction potentials. For the dynamical matrix, the eigenvalues are dimensionless, while for the Hessian, they have units of  $(length)^6$  and are expressed in terms of the length scale of the potential. The ratio of the droplet to the solid is also given in each case and illustrates how different are the static and dynamic values.

	$\lambda^{(-)}\sigma^{-6}$ (free energy Hessian)			$\lambda^{(-)}$ (dynamical matrix)		
	Droplet	Solid	Ratio	Droplet	Solid	Ratio
LJ (Ico)	-0.005 05	-0.002 80	1.8	$-1.61 \times 10^{-5}$	$-6.88 \times 10^{-3}$	0.0024
LJ (FCC)	-0.005 05	$-3.71 \times 10^{-5}$	136	$-1.61 \times 10^{-5}$	$-5.51 \times 10^{-2}$	0.000 29
WHDF	$-1.93 \times 10^{-4}$	$-6.44 \times 10^{-6}$	30	$-9.42 \times 10^{-4}$	$-1.75 \times 10^{-4}$	5.4

## V. NUCLEATION RATES

Classical nucleation theory is based on the assumption that the number of particles in a cluster is a good order parameter for the transition from vapor to liquid or to solid. One expects that in the general context of MeNT, the order parameter should in some sense correspond to movement along the nucleation pathway, which would mean that at the critical cluster, it would involve the projection of the density onto the unstable direction,  $v_i^{(-)}n^l$ , and the fact that  $v_i^{(-)}(n^l - n_0^l) \propto \Delta N$  indicates that the excess mass is indeed a good order parameter. The same assumption underlies the “parameter-free” estimations of nucleation rates from simulation developed by Auer and Frenkel.<sup>24</sup> The idea is to begin with the CNT expression for the nucleation rate,

$$J = n_0 f(N^*) \sqrt{\frac{1}{2\pi} |\beta\Omega''(N^*)|} e^{-\beta\Delta\Omega^*}, \quad (7)$$

where (in our version)  $\Omega(N)$  is the grand-canonical free energy (which in CNT only depends on the order parameter, namely, the mass of the cluster),  $\Delta\Omega^* = \Omega(N^*) - \Omega_0$  is the excess free energy of the critical cluster with  $\Omega_0$  that of the mother phase,  $f(N^*)$  is the rate of attachment of monomers to the critical cluster (which, by definition of the critical cluster, is exactly equal to the rate of detachment of monomers), and  $n_0$  is the density (concentration) of the mother phase. Surprisingly, the CNT attachment rate for diffusive systems is not agreed upon: although some works use the Smolochowski rate for coalescence,<sup>1</sup> this is not well-accepted<sup>17</sup> and more often one resorts to ill-defined concepts, such as typical jump

distance and typical jump time to make estimates. Auer and Frenkel evaluated these various elements directly from simulation and then used the values to get a model-free determination of the nucleation rate (assuming that the size of the cluster is the correct order parameter).

We show in Appendix G that the attachment and detachment frequencies at the critical cluster can be extracted from our model with the result

$$f(N^*) = g(N^*) = \frac{D}{\sigma^2} (M_J v^{(-)J})^2 \quad (8)$$

and that this results in the CNT nucleation rate

$$J = n_0 \frac{D}{\sigma^2} |M_J v^{(-)J}| \sqrt{\frac{|\lambda^{(-)}|}{2\pi}} e^{-\beta\Delta\Omega^*}, \quad (9)$$

which involves only quantities accessible from the model free energy functional and the corresponding dynamical matrix. Evaluations using this are given in Table II. In the case of the WHDF interaction, the kinetic prefactor for the solid and the droplet is almost the same and the difference in the rates of the processes is entirely due to the Arrhenius factors. In contrast, the kinetic factors for the LJ solids are much larger (indicating faster processes) than for the droplet, which seems paradoxical since the solid pathway involves the same accumulation of mass as for the droplet plus the additional time needed for the ordering: even if the latter occurs very fast, the mass accumulation should be the same.

**TABLE II.** Calculated nucleation rates for the LJ icosahedral saddle (Ico), the LJ FCC critical cluster, and the WHDF critical cluster. It also shows the free energy barriers, the prefactors  $Z \equiv J/e^{-\beta\Delta\Omega}$  of the exponentials in Eq. (9), and the ratio of the full rates calculated from the same expression. The final section of this table shows the Z-factors and the ratios of the full rates as calculated from the heuristic one-dimensional model involving the full nucleation pathway.

	$\beta\Delta\Omega^*$		CNT rates			Full path rates		
	Droplet	Solid	Z-droplet	Z-solid	$\frac{J_{\text{droplet}}}{J_{\text{solid}}}$	Z-droplet	Z-solid	$\frac{J_{\text{droplet}}}{J_{\text{solid}}}$
LJ (Ico)	104.9	102.0	$1.76 \times 10^{-9}$	$1.65 \times 10^{-6}$	$6 \times 10^{-5}$	$3.38 \times 10^{-10}$	$5.56 \times 10^{-10}$	0.03
LJ (FCC)	104.9	104.5	$1.76 \times 10^{-9}$	$3.03 \times 10^{-7}$	$4 \times 10^{-3}$	$3.38 \times 10^{-10}$	$1.74 \times 10^{-9}$	0.13
WHDF	21.4	24.7	$9.24 \times 10^{-5}$	$1.67 \times 10^{-4}$	15	$5.33 \times 10^{-5}$	$4.64 \times 10^{-5}$	31



We pause to observe that the application of the CNT rate formulas to crystallization, even with our expressions for the attachment and detachment rates, carries with it various uncomfortable assumptions. It has previously been noted<sup>22</sup> that in crystallization, one expects to see a sequence of free energy minima and maxima along the crystallization pathway due to the formation of complete shells (free energy minima) and the nucleation of new layers (involving free energy maxima) as the crystal grows. Hence, the CNT picture of crossing a unique energy barrier cannot really be correct. Indeed, in the WHDF case, our critical cluster is icosahedral, but the final thermodynamically stable phase is FCC, so we know that there will be at least one further solid–solid transition along the pathway. Thus, the use of the CNT-type analysis must, at some level, be quite heuristic for the crystallization and one of the advantages of our formalism is that it provides the tools to analyze this in the future. For now, we simply note that this assumption is also made in simulations whenever this route to estimating nucleation rates is used. We believe it can be justified, at least as a first approximation, if the critical cluster has the highest free energy of any of those occurring along the full pathway, but formalizing this intuition will require further analysis.

To further understand these results, we have also developed a more elaborate evaluation of the rates based on a one-dimensional theory derived from our analysis and that involves integration along the entire nucleation pathway [see [Appendix H](#), Eq. (H7)]. Results using this expression are also given in [Table II](#). We cannot claim that this full-path value is more accurate than the CNT result, given the limited resolution of the nucleation pathway using the string method. Nevertheless, it is interesting to note that for WHDF, the full-path rate is very similar to the CNT result, differing only by a factor of two. However, while the kinetic factor for the LJ droplet is again similar to the CNT value, the values for the solids are quite different from their CNT values and indeed of similar magnitude as for the droplet. Referring to [Fig. 3](#) and [Table I](#) provides some insights. In the case of the WHDF solid, the critical cluster appears after a period of mass accumulation that in turns occurs after the abrupt ordering: the eigenvalues being computed and used in the CNT rate expression are therefore characteristic of the mass-accumulation process. In contrast, the LJ solid critical clusters both appear during the ordering process so that the eigenvalues being computed are related to the time-scale of the ordering: the large dynamic eigenvalues obtained for these cases indicate the relative speed of the ordering. Only by integrating over the entire path is the much slower mass-accumulation accounted for, in which case the kinetic factors are much more comparable to that of the droplet.

Finally, we note that the fact that for *identical thermodynamics conditions* the nucleation rate for droplets in the WHDF system is much higher than for direct nucleation of crystals suggests the preferential formation of the latter with the possibility that crystal nucleation may then occur within the droplets. In contrast, no such preference is suggested for the LJ system, thus implying the preferential formation of crystals directly from the solution.

## VI. DISCUSSION

We have shown how the powerful combination of cDFT and fluctuating hydrodynamics can give unique insights into the

process of crystallization at the microscopic level. Our results highlight the absolute necessity of a realistic dynamical description of fluctuations. In the present case, if one simply computed the unstable mode of the free energy, it would lead to the erroneous conclusion that the solid and droplet clusters are governed by very different order parameters, reflecting their very different density distributions, and that the order parameters having little physical interpretation. In contrast, the fact that MeNT is based on hydrodynamics that conserves mass leads both to different left and right eigenvectors and to the critical and surprising fact that the projection of the density onto the unstable mode is simply a measure of the excess mass of the system.

For the examples of solid-like critical clusters discussed here, it was found that the development of order, which is a necessary part of the formation of crystalline structures, seems to have little effect on the kinetics of the process and that, contrary to most expectations, taking account of a separate order parameter corresponding to crystalline order is unnecessary. This provides independent support for the idea of using CNT-like expressions for the nucleation rate for crystallization, as underlies the original work of Auer and Frenkel,<sup>24</sup> and the subsequent works based on it. However, we found that if the critical cluster appears during the ordering process, the kinetics sampled in its neighborhood are not typical of mass accumulation and can therefore give deceptively large kinetic factors. In fact, our results suggest that the kinetic factors for the solid should, in general, be comparable to that for a droplet.

Despite these observations, we emphasize that the nucleation pathway for crystallization studied here is multistep in nature: formation and growth of a liquid-like droplet—which seems indistinguishable whether the pathway results in a droplet, a crystalline cluster, or an amorphous cluster—is followed by rapid localization into solid-like regions, followed once more by mass accumulation until criticality is reached. This is consistent with the pathways described previously for Lennard-Jones systems.<sup>22</sup> The suddenness of the ordering was not evident in the earlier work because paths were plotted as functions of the *Euclidean* distance in density space and not the *kinetic* distance that is relevant for the dynamics (see additional figures in the [supplementary material](#)). Our unexpected conclusion is that despite this multistep process, the nucleation rate appears to be quite insensitive to the ordering—other than via the jump in free energy—and so can be described in the language of CNT. This provides a line of evidence, completely independent of experiment and simulation, for the idea<sup>4,5,18</sup> that while the details of the nucleation pathways of crystallization may be (or, perhaps, usually are) highly nonclassical—with independent phases of mass accumulation and ordering and possibly multiple intermediate structures—the dynamics is dominated by the diffusive, and therefore slow, rate of mass accumulation, which is occasionally interrupted by brief periods of ordering or re-ordering, thus giving it a mostly classical appearance.

This scenario—of mass accumulation followed by ordering—is consistent with some of the earliest discussions of two-step nucleation, including the seminal simulations of ten Wolde and Frenkel<sup>7</sup> for LJ and early experimental work by Vekilov<sup>58</sup> on protein crystallization from solution. Notably, in contradiction to our results here, Vekilov's experiments indicated that the crystallization rate of the proteins was much *lower* than expected from a naive application of CNT. However, in the following decade, Vekilov and co-workers

refined their understanding of the process into one in which the initial liquid clusters that precede the solid phase are large (containing  $10^4$  or more molecules) and, for reasons still not well understood, metastable; see, e.g., the work of Maes *et al.*<sup>9</sup> In that work, the slow crystallization dynamics is attributed to high viscosity within the dense droplet. Thus, one possibility is that, even though the over-damped model accounts for slower diffusive dynamics in denser regions via the free energy functional, it may be that the over-damped approximation is too crude inside the droplets and one should instead use the full hydrodynamic model.<sup>21</sup> In recent years, the surprising metastability of the droplets has been investigated by Vekilov's group, and they have proposed that it is due to the presence of transient oligomers that are stabilized in the concentrated clusters,<sup>59</sup> although details of this scenario have been challenged.<sup>60</sup> Whatever the explanation, the existence of metastable high concentration droplets indicates that the protein results involve physical elements that remain poorly understood and are not at present included in our calculations.

We note also the important role played by the icosahedral clusters, which seem to occur frequently as the first manifestation of an incipient solid-like structure. Strikingly, in other calculations that we will describe elsewhere and in which the droplet has a higher free energy barrier than the solid, we see a droplet go through an icosahedral intermediate step, which then transforms back to a droplet. Our work seems to indicate, in general, very complex landscapes with many local maxima and minima and correspondingly complicated pathways.

One notable result of our analysis is the role played by the geometry of density space. As discussed in detail by Graham,<sup>61,62</sup> a diffusive stochastic dynamics induces a metric in the space of the stochastic field, which is to say in our case a natural way of calculating the distance between two local densities. This is exploited here in the definition of the "kinetic distance," which is calculated using this metric as well as by the fact that the weak-noise approximation to the nucleation pathway is steepest descent under this metric.<sup>21</sup> (We recall that the geometric interpretation of steepest descent is that one draws a circle around the starting position, calculates the energy at each point on the circle, and then moves to the point with the lowest energy: the metric tells us how to draw a circle.) One potential danger of less fundamental approaches is that the metric, which in our case is the operator  $D\nabla \cdot \hat{n}(\mathbf{r})\nabla$ , is replaced, either for simplicity or for technical reasons,<sup>41,45</sup> to  $D\nabla^2$ , thus fundamentally changing the geometry of density space. An example of the latter is when the logarithms in the ideal gas contribution to the free energy are replaced by polynomial approximations, as is often the case in phase field crystal models, since one wishes to preserve Fickian diffusion in the low-density limit. In the language of Riemannian geometry, this is a shift from a curved to a flat metric and it would be interesting in the future to investigate the importance of such a qualitative change.

Any theoretical approach to this phenomenon is going to involve heuristic elements due to the underlying complexity and intrinsically multiscale nature of the problem and ours is no exception. The principle assumption is our use of cDFT free energy functionals as the dynamic free energy. This is an assumption that could be explored at a formal level, but here, we only note that, in practice, the most common cDFT free energy functionals are just mean field models, and as such, it could be argued that it makes

more sense to use them as we do here, as part of a dynamic description, than as true thermodynamic free energies, as in cDFT, which should contain renormalization effects that go beyond the mean field.

The applications of our framework are much broader than the specific results given here. In principle, there is no necessity to work in the over-damped limit<sup>21</sup> and, by keeping the full hydrodynamic description, the effect of flows, heat transport, and other important physical phenomena could be studied as well. Indeed, Gallo *et al.*<sup>63</sup> have recently studied bubble nucleation using this approach. Exactly the same concepts can also be used in the study of heterogeneous nucleation, which is of not only more practical importance than homogeneous nucleation but also more problem-specific. It is also possible to determine the MLP without resorting to the weak noise limit,<sup>21</sup> which would allow for nucleation without necessarily passing through the critical cluster.<sup>46</sup>

Crystallization is a prime example of a complex, multiscale phenomenon and as such is approachable at various levels from the most microscopic level (atomistic simulations) to the most coarse-grained (CNT), and each provides unique insights and is useful for addressing different questions. Our theory fits into this spectrum. Its strengths are that, like simulation, it requires only interaction potentials as input and can describe structures down to atomic length scales. At the same time, like all theoretical descriptions, it is phrased in a different conceptual language and so gives direct access to different quantities: in our case, things such as path probabilities<sup>21,64</sup> and the microscopic order parameters. Compared to other theoretical approaches, it involves fewer assumptions and has more expressive power but at the cost of being conceptually and computationally more costly. More coarse-grained methods, such as phase field crystal theory, give up the detailed description of structures and dynamics at the smallest length scales in exchange for being computationally simpler and can therefore be applied to larger and more complex systems. The most coarse-grained approaches, CNT and its many elaborations, are computationally inexpensive and can be adapted to almost any system: they provide the language for discussing, understanding, and interpreting all nucleation phenomena at the macroscopic scale. In the theoretical hierarchy, each level of description can be derived from the more fundamental ones as approximations with well-defined regimes of applicability. As such, we view them not as being in competition with one another but as mutually complementary tools.

## SUPPLEMENTARY MATERIAL

The [supplementary material](#) includes several additional figures: 1. The final FCC cluster on the LJ nucleation pathway discussed in the main text. 2. A zoomed view of the left eigenvectors displayed in [Fig. 1](#) of the main text revealing the fine-scale structure of the left eigenvector. 3. The analog to [Fig. 2](#) of the main text for the WHDF potential. 4. The nucleation pathways plotted using both the kinetic distance, as in [Fig. 3](#) of the main text, and the naive Euclidean distance used in a previous study.<sup>22</sup>

## ACKNOWLEDGMENTS

We thank Bortolo Moggetti for a critical reading of this paper and for several insightful comments that helped us improve it.

The work of J.F.L. was supported by the European Space Agency (ESA) and the Belgian Federal Science Policy Office (BELSPO) in the framework of the PRODEX Program, Contract No. ESA AO-2004-070. That of C.S. was funded by the Belgian National Fund for Scientific Research (FRS-FNRS) under FRIA Grant No. FC 38825. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**James F. Lutsko:** Conceptualization (lead); Formal analysis (lead); Methodology (lead); Software (lead); Visualization (lead); Writing – original draft (lead). **Cédric Schoonen:** Conceptualization (supporting); Formal analysis (supporting); Investigation (equal); Software (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting).

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## APPENDIX A: DISCRETIZED MODELS

The points on the computational lattice are labeled  $I \equiv I(i_x, i_y, i_z)$  with  $0 \leq i_x < N_x$ , etc. and  $0 \leq I \leq N_x N_y N_z$ . We denote the set of all points as  $\mathcal{V}$  (for volume), and we denote the boundary points as  $\partial\mathcal{V} \equiv \{I(i_x, i_y, i_z) : i_x = 0 \text{ or } i_y = 0 \text{ or } i_z = 0\}$ . Our calculations always use periodic boundaries so that, e.g.,  $I(i_x + N_x, i_y, i_z) = I(i_x, i_y, i_z)$ . If  $\vec{n} = (n_1, \dots, n_{N_x N_y N_z})$  be the local density, then the free energy  $\beta\Omega(\vec{n}) = \beta F(\vec{n}) - \beta\mu N(\vec{n})$ , where  $N(\vec{n})$  is the total mass, calculated as described in the main text,  $\mu$  is the chemical potential, and  $F(\vec{n})$  is the Helmholtz functional of cDFT. The latter is written as a sum of an ideal gas contribution, a hard-sphere contribution, and a mean-field contribution. Our calculations of the free energy use the discretized forms for these quantities described in Ref. 65.

Given the explicit form of the free energy functional, the discretization of the dynamics only requires specification of the kinetic prefactor,  $g^{IJ}(n)$ . The calculations presented here were performed using

$$g^{IJ} = \frac{\sigma^2}{\Delta^5} \sum_{a=x,y,z} \left\{ \begin{array}{l} \frac{n^{I+e^{(a)}} + n^I}{2} (\delta^{I+e^{(a)}I} - \delta^{IJ}) \\ - \frac{n^I + n^{I-e^{(a)}}}{2} (\delta^{IJ} - \delta^{I-e^{(a)}I}) \end{array} \right\},$$

where  $e^{(a)}$  is the unit vector in the direction  $a$  so that, e.g.,  $I + e^{(x)} = (I_x + 1, I_y, I_z)$ . We note that corresponding to this we have

$$q^{IK(a)} = \frac{\sigma}{\Delta^{5/2}} (\delta^{I+e^{(a)}K} - \delta^{IK}) \sqrt{\frac{n^K + n^{K-e^{(a)}}}{2}}, \quad (\text{A1})$$

and it is straightforward to confirm that these satisfy the fluctuation–dissipation relation, which, incidentally, assures that  $g^{IJ}$  is positive semi-definite.

## APPENDIX B: BOUNDARY CONDITIONS

In calculating the free energy and its derivatives, we always use periodic boundaries (for reasons of efficiency: the calculations are only feasible with fast Fourier transforms). In order for the dynamics to model an open system, we take  $n^I$  to be constant on the border of the computational cell (defined as values of  $I$  for which one or more components is zero). In order that the uniform density be a stationary solution of the deterministic dynamics, we must also set the forces  $\partial F / \partial n^I = \Delta^3 \mu$  for all values of  $I$  on the boundary. The constant  $\mu$  is simply the chemical potential corresponding to the vapor. Examination of the resulting difference equations shows that a mathematically equivalent formulation, that is perhaps more physically useful, is to replace the free energy by  $F(n) - \mu N(n)$  and to set  $F_I = 0$  for the boundary points. This formulation shows that the boundary conditions force the use of a “grand canonical” free energy functional. We also note that the fixed density along the borders breaks translational symmetry (which is otherwise present with periodic boundaries), and as a result, the matrix  $g^{IJ}$  has no zero eigenvalues and is invertible.

To avoid any confusion, we note that the periodic boundaries constitute a *spatial* boundary condition used to calculate the free energy and its derivatives, whereas the fixed density on the borders is a *dynamical* boundary condition pertaining to the solution of the time-dependent equations. The two are therefore independent and compatible as long as the fixed boundary densities are also periodic, as is the case in our calculations.

## APPENDIX C: CALCULATION OF EIGENVALUES AND EIGENVECTORS

We determine the eigenvalues and eigenvectors using the SLEPC library using the Jacobi–Davidson method for the Hessian of the free energy and the Krylov–Schur method for the dynamical matrix. In both cases, the Generalized Minimal Residual method (KSPGMRES) linear equation solver was used. When diagonalizing the dynamical matrix, we used two-sided balancing and demanded that the left and right eigenvectors be calculated.

Suppose that we determine a right eigenvector of the dynamical matrix,  $v^I$ . Then, it is easy to see that  $v_I \equiv g_{IJ} v^J$  is a left eigenvector having the same eigenvalue. We can therefore create a normalized, conjugate pair of eigenvectors by defining

$$\begin{aligned} \tilde{v}^I &= \frac{1}{\sqrt{v^J g_{JK} v^K}} v^I, \\ \tilde{v}_I &= \frac{1}{\sqrt{v^J g_{JK} v^K}} g^{IL} v_L, \end{aligned} \quad (\text{C1})$$

satisfying  $\tilde{v}_I = g_{IJ} \tilde{v}^J$  and  $\tilde{v}_I \tilde{v}^I = 1$ . Similarly, given any left eigenvector  $u_I$ , an analogous pair

$$\begin{aligned}\tilde{u}_I &= \frac{1}{\sqrt{u_J g^{JK} u_K}} u_I, \\ \tilde{u}^I &= \frac{1}{\sqrt{u_J g^{JK} u_K}} g^{II} u_L\end{aligned}\quad (\text{C2})$$

also exists. All of our results are written with the assumption that the eigenvectors obey this normalization. Finally, if we have independently determined left and right eigenvectors,  $u_I$  and  $v^I$ , then they should be related by  $v^I = \alpha g^{II} u_I$  for some constant,  $\alpha$ , where contraction with  $u_I$  gives

$$\alpha = \frac{v^I u_I}{u_J g^{JK} u_K},$$

and so  $v^J g_{JK} v^K = \alpha u_K v^K = \frac{(v^I u_I)^2}{u_J g^{JK} u_K}$ , giving

$$\begin{aligned}\tilde{v}^I &= \frac{\sqrt{u_J g^{JK} u_K}}{u_L v^L} v^I, \\ \tilde{v}_I &= \frac{1}{\sqrt{u^J g^{JK} u_K}} u_I.\end{aligned}\quad (\text{C3})$$

#### APPENDIX D: CALCULATION DETAILS

The spherically symmetric pair-potentials used in this study were a cutoff Lennard-Jones (LJ) potential,

$$v_{\text{LJ}}(r) = 4\epsilon \left( \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right) - 4\epsilon \left( \left( \frac{\sigma}{r_{\text{cut}}} \right)^{12} - \left( \frac{\sigma}{r_{\text{cut}}} \right)^6 \right), \quad (\text{D1})$$

$r < r_{\text{cut}},$

with cutoff  $r_{\text{cut}} = 3\sigma$  and the potential of Wang *et al.*<sup>47</sup> or WHDF potential,

$$v_{\text{WHDF}}(r) = 114.11\epsilon \left( \left( \frac{\sigma}{r} \right)^2 - 1 \right) \left( \left( \frac{r_{\text{cut}}}{r} \right)^2 - 1 \right)^2, \quad r < r_{\text{cut}}, \quad (\text{D2})$$

with cutoff  $r_{\text{cut}} = 1.2\sigma$ . The computational lattice spacing was  $\Delta = 0.2\sigma$  for the Lennard-Jones and  $\Delta = 0.1\sigma$  for the WHDF potential. Both potentials have a nearly hard-core repulsion at short distances ( $r < \sigma$ ) and an attractive well outside this. The difference between them is that the ratio of the width of the attractive part of the potential to the width of the hard-core region is much smaller for WHDF than for the LJ. This is intended to model typical colloidal interactions.

To find critical clusters, we first minimize the free energy of a system consisting of a droplet in a background of low-density fluid under fully periodic boundary conditions (e.g., no fixed density at the boundary): this implies constant total number of particles and so a canonical-like minimization. It has previously been shown that such a cluster is a stationary point when the system is opened: that is, when the density is held constant at the boundary and the total number of particles allowed to vary.<sup>22,23</sup> In this grand-canonical system, the cluster may be either a local minimum or a saddle point—there

is no way to know except to test it, e.g., by calculating the eigenvalues. Notice that when this is done at sufficiently low temperatures, the clusters spontaneously form solid-like, amorphous or crystalline, order.<sup>23</sup>

Regardless of the nature of the cluster so obtained, we use it as the endpoint for the determination of the nucleation pathway using the string method as described in detail in Ref. 22. If the cluster anchoring the string was a local minimum, a critical cluster appears somewhere along this pathway. However, because of the limited resolution of the string method—i.e., the fact that the pathway is characterized by a finite number of images—the “critical cluster” identified is not in fact exactly a saddle point of the free energy surface. We therefore further refine it using various methods (e.g., an eigenvector-following method<sup>66</sup>) to be described in a future publication. This refined critical cluster is then used as the endpoint for a new string calculation and, in the case of LJ reported in the main text, a separate string calculation was used to generate the solid–solid pathway from the refined critical cluster to the endpoint of the initial string calculation.

#### APPENDIX E: MEASURE OF ORDER WITHIN THE CLUSTERS

We track the development of order in the clusters using a quantity that we call “localization.” It is ignorant of the specific type of solid structure that is appearing and simply measures how much the density is peaked around the “atomic” positions. We define it in terms of the local packing fraction  $\eta^I(R)$ , which measures the amount of mass within a radius  $R$  around a lattice position  $I$ . The evaluation of these packing fractions on the computational lattice is not trivial and is explained in Ref. 65. The radius  $R$  is the effective hard sphere radius of our cDFT model. We estimate how much the density is localized around a position  $I$  by computing the excess of mass  $\eta^I(R/2) - \eta^I(R)/8$  within a smaller radius  $R/2$ , which we then normalize to get

$$l^I = \frac{\eta^I(R/2) - \eta^I(R)/8}{\eta^I(R) - \eta^I(R)/8}. \quad (\text{E1})$$

This (local) measure of localization is zero if the mass  $\eta^I(R)$  is uniformly distributed in the sphere of radius  $R$  and one if this entire mass is located in the smaller sphere. The density-weighted average of  $l^I$  over the entire computational lattice is our (global) measure of localization within the cluster,

$$\bar{l} = \frac{1}{N} \sum_I n^I l^I \Delta^3, \quad (\text{E2})$$

where  $N = \sum_I n^I \Delta^3$  is the total number of particles.

#### APPENDIX F: CALCULATING DISTANCE ALONG THE STRING PATH

The string method results in a series of snapshots of the system along the nucleation pathway. We label these consecutively as  $n_\alpha$  for  $\alpha = 0, 1, \dots$  with  $\alpha = 0$  corresponding to the initial uniform system as shown in Fig. 3. We approximate the distance



between successive points along the pathway by defining the difference  $\Delta n_{\alpha+1/2}^I = n_{\alpha+1}^I - n_{\alpha}^I$  and the average  $\bar{n}_{\alpha+1/2}^I = \frac{1}{2}(n_{\alpha+1}^I + n_{\alpha}^I)$  and evaluating

$$\Delta s(\alpha, \alpha + 1) = \sqrt{\Delta n_{\alpha+1/2}^I g_{II}(\bar{n}_{\alpha+1/2}^I) \Delta n_{\alpha+1/2}^I} \quad (\text{F1})$$

and then using  $s(\alpha) = \sum_{\gamma=1}^{\alpha} \Delta s(\gamma - 1, \gamma)$ . Since we only have an explicit representation for the matrix  $g^{IJ}$  and we require the inverse for this evaluation, we write

$$g^{IJ} \phi_J = \Delta n_{\alpha+1/2}^I \quad (\text{F2})$$

and solve this system of linear equations for  $\phi_I$  using tools from the PETSC library.<sup>67,68</sup> We can then replace  $\Delta n^I g_{IJ} \Delta n^J$  by  $\Delta n^I \phi_I$ .

### APPENDIX G: USING MeNT TO DETERMINE THE CNT NUCLEATION RATE

The CNT nucleation rates require as input the free energy at the critical cluster, the second derivative of the free energy with respect to particle number at the critical cluster, and the attachment rate. The free energies are of course directly given from our model. To determine the CNT attachment and detachment frequencies (which are only needed at the critical cluster in order to calculate the CNT nucleation rates), we compare the stochastic model near the critical cluster to the assumed dynamics of CNT. To this end, the stochastic equation is expanded to first order to get

$$\frac{d}{dt} \delta \widehat{n}_t^I = -\frac{D}{\sigma^2} g^{IJ}(n^*) \left( \frac{\partial^2 \beta \Omega(n)}{\partial n^I \partial n^K} \right)_{n^*} \delta \widehat{n}_t^K + \sqrt{\frac{2D}{\sigma^2}} \mathbf{q}^{IJ}(n^*) \cdot \widehat{\xi}_t^J. \quad (\text{G1})$$

Using the fact that  $\delta N = M_I v^{(-)I} \delta \widehat{n}_t^I \simeq (M_I v^{(-)I}) v_I^{(-)} \delta \widehat{n}_t^I$ , we find

$$\begin{aligned} \frac{d}{dt} \delta N &= -\frac{D}{\sigma^2} (M_L v^{(-)L}) v_I^{(-)} g^{IJ}(n^*) \left( \frac{\partial^2 \beta \Omega(n)}{\partial n^I \partial n^K} \right)_{n^*} \delta \widehat{n}_t^K \\ &\quad + \sqrt{\frac{2D}{\sigma^2}} (M_L v^{(-)L}) v_I^{(-)} \mathbf{q}^{IJ}(n^*) \cdot \widehat{\xi}_t^J \\ &= -\frac{D}{\sigma^2} \lambda^{(-)} (M_L v^{(-)L}) v_I^{(-)} \delta \widehat{n}_t^I \\ &\quad + \sqrt{\frac{2D}{\sigma^2}} (M_L v^{(-)L}) v_I^{(-)} \mathbf{q}^{IJ}(n^*) \cdot \widehat{\xi}_t^J \\ &\equiv -\frac{D}{\sigma^2} \lambda^{(-)} \delta N + \sqrt{\frac{2D}{\sigma^2}} (M_L v^{(-)L}) \widehat{\xi}_t, \end{aligned} \quad (\text{G2})$$

where, in the last step, we have replaced the white noise by a simpler form having the same covariance and therefore corresponding to the same Fokker–Planck equation.<sup>69</sup>

In CNT, it is assumed that the size of clusters changes by the attachment and detachment of monomers and that these occur with rates  $f(N)$  and  $g(N)$  for a cluster of size  $N$ . This means that the probability that a cluster has size  $N$  at time  $t + \tau$  is, for small  $\tau$ ,

$$\begin{aligned} P(N; t + \tau) &= P(N; t) + \tau f(N - 1) P(N - 1; t) \\ &\quad - \tau (f(N) + g(N)) P(N; t) \\ &\quad + \tau g(N + 1) P(N + 1; t), \end{aligned} \quad (\text{G3})$$

which can be rearranged as

$$\frac{P(N; t + \tau) - P(N; t)}{\tau} = \frac{\delta_2((g(N) - f(N)) P(N; t))}{2} + \frac{\delta_1^2((f(N) + g(N)) P(N; t))}{2}, \quad (\text{G4})$$

where we have used the standard notation for centered finite differences, viz.,  $\delta_h f(x) \equiv f(x + \frac{h}{2}) - f(x - \frac{h}{2})$ . Taking the limit that  $\tau \rightarrow 0$  leads to a differential equation in time, treating  $N$  as a continuous variable, and replacing the finite differences by derivatives give a partial differential equation for  $P(N; t)$  that is first order in time and second order in  $N$ ,

$$\frac{\partial}{\partial t} P(N, t) = \frac{\partial}{\partial N} \left( (g(N) - f(N)) P(N, t) \right) + \frac{1}{2} \frac{\partial^2}{\partial N^2} [(f(N) + g(N)) P(N, t)], \quad (\text{G5})$$

and is recognized as a Fokker–Planck equation. (This is essentially the same as the Tunitskii equation of CNT.<sup>1</sup>) That Fokker–Planck equation is in turn equivalent to the (Ito) stochastic equation,<sup>69</sup>

$$\frac{d}{dt} \widehat{N}_t = (f(\widehat{N}_t) - g(\widehat{N}_t)) + \sqrt{f(\widehat{N}_t) + g(\widehat{N}_t)} \widehat{\xi}_t. \quad (\text{G6})$$

Comparing this to Eq. (G2) and evaluating both at the critical cluster, we can identify the attachment and detachment frequencies as

$$f(N^*) = g(N^*) = \frac{D}{\sigma^2} (M_I v^{(-)I})^2 = \frac{D}{\sigma^2} \left( \frac{dN}{ds} \right)_{n^*}^2. \quad (\text{G7})$$

To relate these quantities to the thermodynamics, as in CNT, we turn to the evaluation of the derivative of the free energy along the MLP. The derivative of the free energy with respect to the particle number along the MLP is evaluated using the fact that the MLP points along the deterministic driving force (in the weak noise approximation), so near the given point on the MLP,  $n^{(\text{MLP})}$ , the path can be parameterized as

$$n^{(\text{MLP})I}(\alpha) = n^{(\text{MLP})I} + \alpha g^{IJ} \beta \Omega_J(n^{(\text{MLP})}), \quad (\text{G8})$$

and so

$$\begin{aligned} \left. \frac{d\beta\Omega}{dN} \right|_{n^{(\text{MLP})}} &= \lim_{\alpha \rightarrow 0} \left( \frac{d\beta\Omega(n^{(\text{MLP})}(\alpha))}{d\alpha} \right) / \left( \frac{dn^{(\text{MLP})}(\alpha)}{d\alpha} \right) \\ &= \frac{\beta \Omega_I(n^{(\text{MLP})}) g^{IJ}(n^{(\text{MLP})}) \beta \Omega_J(n^{(\text{MLP})})}{M_K g^{KL}(n^{(\text{MLP})}) \beta \Omega_L(n^{(\text{MLP})})}. \end{aligned} \quad (\text{G9})$$



At the critical point, the first derivative of the free energy vanishes, and so, near the critical point, the steepest descent equation for  $\delta n^I = n^I - n^{*I}$  takes the form

$$\frac{d\delta n^I}{dt} = \frac{D}{\sigma^2} g^{IJ}(n^*) \beta \Omega_{JK}(n^*) \delta n^K \quad (\text{G10})$$

and the displacement along the MLP obeys

$$\delta n^I \propto g^{IJ}(n^*) \beta \Omega_{JK}(n^*) \delta n^K, \quad (\text{G11})$$

or in other words, it is the direction of the unstable eigenvector. Hence, we now have

$$n^{(\text{MLP})I}(\alpha) = n^{*I} + \alpha v^{(-)I}, \quad (\text{G12})$$

and straightforward evaluation gives

$$\left. \frac{d\beta\Omega}{dN} \right|_{n^*} = 0, \quad \left. \frac{d^2\beta\Omega}{dN^2} \right|_{n^*} = \lambda^{(-)} \frac{v_j^{(-)} v^{(-)j}}{(M_I v^{(-)I})^2} = \frac{\lambda^{(-)}}{(M_I v^{(-)I})^2}. \quad (\text{G13})$$

It is also useful below to note that along the MLP,

$$ds = \sqrt{dn^I g_{IJ} dn^J} = \sqrt{\Omega_I g^{IJ} d\Omega_J} d\alpha, \quad (\text{G14})$$

and so

$$\frac{dN}{ds} = \frac{M_K g^{KL} \beta \Omega_L}{\sqrt{\Omega_I g^{IJ} d\Omega_J}}. \quad (\text{G15})$$

At the critical cluster, these become  $ds = d\alpha$  and

$$\left( \frac{dN}{ds} \right)_{n^*} = M_I v^{(-)I}, \quad (\text{G16})$$

respectively.

Finally, using Eqs. (G7) and (G13), to evaluate the CNT expression for the nucleation rate, Eq. (7) gives Eq. (9).

## APPENDIX H: HEURISTIC ONE-DIMENSIONAL MODEL

To give an idea of the effect of the entire nucleation pathway on the nucleation rates, we consider a 1D model based on the information derived from the full theory. In the main text, we used a projection of the stochastic equation onto the unstable mode at the critical cluster to extract the attachment rate. Here, we follow a similar idea and, as in the discussion above (derivatives along the path), we project in the direction of the deterministic force, giving

$$\begin{aligned} \frac{\beta \Omega_I}{\sqrt{\Omega_I g^{IJ} \beta \Omega_J}} \frac{d}{dt} \tilde{n}_I &= -\frac{D}{\sigma^2} \frac{\beta \Omega_I}{\sqrt{\Omega_I g^{IJ} \beta \Omega_J}} g^{IJ} \beta \Omega_J \\ &+ \sqrt{\frac{2D}{\sigma^2}} \frac{\beta \Omega_I}{\sqrt{\Omega_I g^{IJ} \beta \Omega_J}} \mathbf{q}^{IJ} \cdot \tilde{\xi}_t^I, \end{aligned} \quad (\text{H1})$$

where we use a short-hand notation in which  $\beta \Omega_I \equiv \partial \beta \Omega / \partial n^I$ . To make sense of this, consider the deterministic evolution of the distance, which obeys

$$\frac{ds}{dt} = \sqrt{\frac{dn^I}{dt} g^{IJ} \frac{dn^J}{dt}} = \frac{D}{\sigma^2} \sqrt{\beta \Omega_I g^{IJ} \beta \Omega_J}, \quad (\text{H2})$$

so

$$\frac{\beta \Omega_I}{\sqrt{\Omega_I g^{IJ} \beta \Omega_J}} \frac{dn^I}{dt} = -\frac{ds}{dt}. \quad (\text{H3})$$

This suggests, then, the model (Stratonovich<sup>69</sup>) stochastic differential equation,

$$\frac{ds}{dt} = -\frac{D}{\sigma^2} \sqrt{\beta \Omega_I g^{IJ} \beta \Omega_J} + \sqrt{\frac{2D}{\sigma^2}} \frac{\beta \Omega_I}{\sqrt{\Omega_I g^{IJ} \beta \Omega_J}} \mathbf{q}^{IJ} \cdot \tilde{\xi}_t^I, \quad (\text{H4})$$

and, again replacing the noise term by an equivalent one with the same covariance, this becomes

$$\frac{ds}{dt} = -\frac{D}{\sigma^2} \sqrt{\beta \Omega_I g^{IJ} \beta \Omega_J} + \sqrt{\frac{2D}{\sigma^2}} \tilde{\xi}_t. \quad (\text{H5})$$

Finally, since all motion is along the MLP,  $\sqrt{\beta \Omega_I g^{IJ} \beta \Omega_J} = \frac{d\beta\Omega}{ds}$ ,

$$\frac{ds}{dt} = -\frac{D}{\sigma^2} \frac{d\beta\Omega(s)}{ds} + \sqrt{\frac{2D}{\sigma^2}} \tilde{\xi}_t. \quad (\text{H6})$$

We pause to note that the equilibrium distribution for this stochastic process will be  $P(s) \propto e^{-\beta\Omega(s)}$  as one would expect if the variable  $s$  is the correct order parameter.

A detailed analysis of this model will be given elsewhere. Here, we note that the mean first passage time,  $T(s)$ , for a cluster in state  $s$  to eventually reach the critical state,  $s^*$ , corresponding to the critical cluster can be determined exactly,<sup>69</sup> and from this, the nucleation rate is given by the population over the flux,<sup>13</sup> giving the general expression

$$J \simeq \frac{D}{\sigma^2} \frac{1}{2} \int_{s_0}^{s^*} \frac{c(s)}{\int_s^{s^*} e^{\beta\Omega(s')} \left( \int_0^{s'} e^{-\beta\Omega(s'')} ds'' \right) ds'} ds, \quad (\text{H7})$$

where  $c_1$  is the concentration of monomers. If we assume that the system begins in a (near-) equilibrium state,  $c(N) \approx c_1 e^{-\beta(\Omega(N) - \Omega(1))}$ , then  $c(s) \approx c_1 e^{-\beta(\Omega(s) - \Omega(0))} \frac{dN}{ds}$  and this can be evaluated. (Note that the factor of  $\frac{dN}{ds}$  is necessary as  $c(N)$  and  $c(s)$  are densities. For example, integrating  $c(N)$  over  $N = 0$  to  $N^*$  has to be the same as integrating  $c(s)$  over  $s = 0$  to  $s^*$  since both are the total number of sub-critical clusters.) Standard approximations give the intermediate form,

$$J \simeq \frac{D}{\sigma^2} \frac{1}{2} \left( \frac{dN}{ds} \right)_{s=0} \frac{c_1}{\int_0^{s^*} e^{\beta\Delta\Omega(s)} ds}, \quad (\text{H8})$$

and a final evaluation using Laplace's method gives

$$J \simeq \frac{D}{\sigma^2} c_1 \left( \frac{dN}{ds} \right)_{s=0} \sqrt{\frac{1}{2\pi} \left| \frac{d^2}{ds^2} \beta\Omega(s) \right|_{s=s^*}} e^{-\beta\Delta\Omega^*} \quad (\text{H9})$$

or, changing variables,

$$J \simeq \frac{D}{\sigma^2} c_1 \left( \frac{dN}{ds} \right)_{N=1} \left( \frac{dN}{ds} \right)_{N^*} \sqrt{\frac{1}{2\pi} \left| \frac{d^2}{dN^2} \beta\Omega(N) \right|_{N=N^*}}. \quad (\text{H10})$$

Not surprisingly, this only agrees with the CNT result, given in Eq. (9) of the main text, if  $\left(\frac{dN}{ds}\right)_{N=1} = \left(\frac{dN}{ds}\right)_{N^*}$  or, more generally, if  $\frac{dN}{ds}$  is constant, which means that  $s$  and  $N$  are equivalent as order parameters.

It is interesting to note that we can use this model to identify the attachment and detachment frequencies for any sized cluster. First, we multiply Eq. (H6) by  $dN/ds$  to get the Stratonovich stochastic differential equation,

$$\frac{dN}{dt} = -\frac{D}{\sigma^2} \left( \frac{dN}{ds} \right)^2 \frac{d\beta\Omega(N)}{ds} + \sqrt{\frac{2D}{\sigma^2}} \frac{dN}{ds} \widehat{\xi}_t, \quad (\text{H11})$$

which is equivalent to the Fokker–Planck equation,

$$\frac{\partial}{\partial t} P(N, t) = \frac{D}{\sigma^2} \frac{\partial}{\partial N} \left( \left( \frac{dN}{ds} \right)^2 \frac{d\beta\Omega(N)}{dN} + \frac{dN}{ds} \frac{\partial}{\partial N} \frac{dN}{ds} \right) P(N, t), \quad (\text{H12})$$

which can also be written as

$$\frac{\partial}{\partial t} P(N, t) = \frac{D}{\sigma^2} \frac{\partial}{\partial N} \left( \left( \frac{dN}{ds} \right)^2 \frac{d\beta\Omega(N)}{dN} - \frac{dN}{ds} \frac{d}{ds} \frac{dN}{ds} + \frac{\partial}{\partial N} \left( \frac{dN}{ds} \right)^2 \right) P(N, t), \quad (\text{H13})$$

or, more succinctly,

$$\frac{\partial}{\partial t} P(N, t) = \frac{D}{\sigma^2} \frac{\partial}{\partial N} \left( \left( \frac{dN}{ds} \right)^2 \frac{d(\beta\Omega(N) - \ln \frac{dN}{ds})}{dN} + \frac{\partial}{\partial N} \left( \frac{dN}{ds} \right)^2 \right) P(N, t). \quad (\text{H14})$$

Comparing to the CNT result, Eq. (G5), we can identify

$$f(N) - g(N) = -\frac{D}{\sigma^2} \left( \frac{dN}{ds} \right)^2 \frac{d(\beta\Omega(N) + \ln \frac{ds}{dN})}{dN}, \quad (\text{H15})$$

$$f(N) + g(N) = 2 \frac{D}{\sigma^2} \left( \frac{dN}{ds} \right)^2.$$

At the critical cluster, by definition,  $f(N^*) = g(N^*)$ , so this gives  $f(N^*) = \frac{D}{\sigma^2} \left( \frac{dN}{ds} \right)_{N^*} = \frac{D}{\sigma^2} \left( M_{IV}^{(-)I} \right)^2$  as in the main text. Note, however, that the critical cluster is not what one might expect: it is not the derivative of the free energy that is zero but of the shifted free energy  $\beta\Omega + \frac{ds}{dN}$ . This is related to an assumption that permeates CNT although it is seldom discussed. To illustrate within

the present context, let us imagine a sub-critical—and therefore, equilibrium—fluid. In this case, the probability that a cluster has a given size should be constant, and in fact, it is easy to determine the equilibrium distribution from our model. Referring to Eq. (G5), one easily finds that the probability that a cluster has size  $N$ ,  $P(N)$ , is  $P(N) = \mathcal{N} e^{-(\beta\Omega(N) + \ln \frac{ds}{dN})}$ , where the prefactor is a normalization constant. Hence, the effective free energy is shifted. The origin of this shift is apparent if we change variables from the size of the cluster to its order parameter. Assuming an invertible relation,  $s(N)$ , the probability that a cluster has order parameter  $s$ ,  $\tilde{P}(s) = P(N(s)) \frac{dN}{ds}$ , where the second factor on the right is needed to preserve normalization (integrating over all  $s$  must be the same as integrating over all  $N$ ). Working this out gives  $\tilde{P}(s) = \mathcal{N} e^{-\beta\tilde{\Omega}(s)}$ , where  $\tilde{\Omega}(s) \equiv \Omega(N(s))$ . Hence, the probability has the canonical form when written in terms of the natural order parameter and it is only when making a nonlinear change of variable that the “shift” occurs. In CNT, one sometimes sees presentations in which the order parameter is not taken to be  $N$  but, rather, the radius of a cluster  $R$ . In that case, it is often assumed that an equilibrium distribution is proportional to  $e^{-\beta\Omega(R)}$  rather than  $e^{-\beta\Omega(N)}$  and it is clear that both cannot be true. The fact that the free energy is shifted in these expressions reflects the fundamental observation that the “classical” picture really only fully holds when  $s$  and  $N$  are linearly related: curvature in their relation indicates that changes in the local density occur, which are not directly related to changes in total mass, and this violates the fundamental assumptions of CNT. We will explore these issues more fully in a future publication.

## REFERENCES

- D. Kashchiev, *Nucleation: Basic Theory with Applications* (Butterworth-Heinemann, Oxford, 2000).
- P. G. Vekilov, *Nonclassical Nucleation, in Crystallization via Nonclassical Pathways Volume 1: Nucleation, Assembly, Observation and Application, in ACS Symposium Series*, edited by X. Zhang (American Chemical Society, 2020).
- J. J. De Yoreo, P. U. P. A. Gilbert, N. A. J. M. Sommerdijk, R. L. Penn, S. Whitelam, D. Joester, H. Zhang, J. D. Rimer, A. Navrotsky, J. F. Banfield, A. F. Wallace, F. M. Michel, F. C. Meldrum, H. Cölfen, and P. M. Dove, “Crystallization by particle attachment in synthetic, biogenic, and geologic environments,” *Science* **349**, aaa6760 (2015).
- P. J. M. Smeets, A. R. Finney, W. J. E. M. Habraken, F. Nudelman, H. Friedrich, J. Laven, J. J. De Yoreo, P. M. Rodger, and N. A. J. M. Sommerdijk, “A classical view on nonclassical nucleation,” *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7882 (2017).
- C. Jia, A. Xiao, J. Zhao, P. Wang, X. Fang, H. Zhang, and B. Guan, “A new perspective on crystal nucleation: A classical view on non-classical nucleation,” *Cryst. Growth Des.* **24**, 601 (2024).
- D. Gebauer, J. D. Gale, and H. Cölfen, “Crystal nucleation and growth of inorganic ionic materials from aqueous solution: Selected recent developments, and implications,” *Small* **18**, 2107735 (2022).
- P. R. T. Wolde and D. Frenkel, “Enhancement of protein crystal nucleation by critical density fluctuations,” *Science* **277**, 1975 (1997).
- R. P. Sear, “The non-classical nucleation of crystals: Microscopic mechanisms and applications to molecular crystals, ice and calcium carbonate,” *Int. Mater. Rev.* **57**, 328 (2012).
- D. Maes, M. A. Vorontsova, M. A. Potenza, T. Sanvito, M. Sleutel, M. Giglio, and P. G. Vekilov, “Do protein crystals nucleate within dense liquid clusters?,” *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **71**, 815 (2015).
- J. S. Langer, “Statistical theory of the decay of metastable states,” *Ann. Phys.* **54**, 258 (1969).

- <sup>11</sup>J. S. Langer and L. A. Turski, "Hydrodynamic model of the condensation of a vapor near its critical point," *Phys. Rev. A* **8**, 3230 (1973).
- <sup>12</sup>P. Talkner, "Mean first passage time and the lifetime of a metastable state," *Z. Phys. B: Condens. Matter* **68**, 201 (1987).
- <sup>13</sup>P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: Fifty years after Kramers," *Rev. Mod. Phys.* **62**, 251 (1990).
- <sup>14</sup>J. Russo and H. Tanaka, "Crystal nucleation as the ordering of multiple order parameters," *J. Chem. Phys.* **145**, 211801 (2016).
- <sup>15</sup>T. Karmakar, M. Invernizzi, V. Rizzi, and M. Parrinello, "Collective variables for the study of crystallisation," *Mol. Phys.* **119**, e1893848 (2021).
- <sup>16</sup>A. Berezhkovskii and A. Szabo, "One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions," *J. Chem. Phys.* **122**, 014503 (2004).
- <sup>17</sup>B. Peters, *Reaction Rate Theory and Rare Events Simulations* (Elsevier, Amsterdam, 2017).
- <sup>18</sup>W. Gispen and M. Dijkstra, "Finding the differences: Classical nucleation perspective on homogeneous melting and freezing of hard spheres," *J. Chem. Phys.* **160**, 141102 (2024), <https://arxiv.org/abs/2401.05855>.
- <sup>19</sup>M. Iwamatsu, "Steady-state nucleation rate and flux of composite nucleus at saddle point," *J. Chem. Phys.* **136**, 204702 (2012).
- <sup>20</sup>M. Iwamatsu, "Nucleation and growth of a core-shell composite nucleus by diffusion," *Phys. Rev. E* **95**, 042803 (2017).
- <sup>21</sup>J. F. Lutsko, "A dynamical theory of nucleation for colloids and macromolecules," *J. Chem. Phys.* **136**, 034509 (2012).
- <sup>22</sup>J. F. Lutsko, "How crystals form: A theory of nucleation pathways," *Sci. Adv.* **5**, eaav7399 (2019).
- <sup>23</sup>J. F. Lutsko and J. Lam, "Classical density functional theory, unconstrained crystallization, and polymorphic behavior," *Phys. Rev. E* **98**, 012604 (2018).
- <sup>24</sup>S. Auer and D. Frenkel, "Prediction of absolute crystal-nucleation rate in hard-sphere colloids," *Nature* **409**, 1020 (2001).
- <sup>25</sup>H. Grabert, *Projection Operator Techniques in Nonequilibrium Statistical Mechanics* (Springer Berlin, Berlin, 1982).
- <sup>26</sup>D. Zubarev and V. Morozov, "Statistical mechanics of nonlinear hydrodynamic fluctuations," *Physica A* **120**, 411 (1983).
- <sup>27</sup>B. D. Goddard, G. A. Pavliotis, and S. Kalliadasis, "The overdamped limit of dynamic density functional theory: Rigorous results," *Multiscale Model. Simul.* **10**, 633 (2012).
- <sup>28</sup>R. Evans, "The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids," *Adv. Phys.* **28**, 143 (1979).
- <sup>29</sup>J. F. Lutsko, "Recent developments in classical density functional theory," *Adv. Chem. Phys.* **144**, 1 (2010).
- <sup>30</sup>C. Schoonen and J. F. Lutsko, "Crystal polymorphism induced by surface tension," *Phys. Rev. Lett.* **129**, 246101 (2022).
- <sup>31</sup>J. F. Lutsko, *J. Chem. Phys.* **135**, 161101 (2011).
- <sup>32</sup>P.-H. Chavanis, "Hamiltonian and Brownian systems with long-range interactions: V. Stochastic kinetic equations and theory of fluctuations," *Physica A* **387**, 5716 (2008).
- <sup>33</sup>P.-H. Chavanis, "Brownian particles with long- and short-range interactions," *Physica A* **390**, 1546 (2011).
- <sup>34</sup>D. S. Dean, *J. Phys. A: Math. Gen.* **29**, L613 (1996).
- <sup>35</sup>K. Kawasaki, *J. Stat. Phys.* **93**, 527 (1998).
- <sup>36</sup>N. Bidhoo and S. P. Das, *Phys. Rev. E* **92**, 012325 (2015).
- <sup>37</sup>A. G. Lamorgese, D. Molin, and R. Mauri, "Diffuse interface (D.I.) model for multiphase flows," in *Multiphase Microfluidics: The Diffuse Interface Model*, edited by R. Mauri (Springer Vienna, Vienna, 2012), pp. 1–72.
- <sup>38</sup>J. W. Cahn and J. E. Hilliard, "Free energy of a nonuniform system. I. Interfacial free energy," *J. Chem. Phys.* **28**, 258 (1958).
- <sup>39</sup>J. W. Cahn and J. E. Hilliard, "Free energy of a nonuniform system. III. Nucleation in a two-component incompressible fluid," *J. Chem. Phys.* **31**, 688 (1959).
- <sup>40</sup>L. Gránásy, G. I. Tóth, J. A. Warren, F. Podmaniczky, G. Tegze, L. Rátkai, and T. Pusztai, "Phase-field modeling of crystal nucleation in undercooled liquids—A review," *Prog. Mater. Sci.* **106**, 100569 (2019).
- <sup>41</sup>S. van Teeffelen, R. Backofen, A. Voigt, and H. Löwen, "Derivation of the phase-field-crystal model for colloidal solidification," *Phys. Rev. E* **79**, 051404 (2009).
- <sup>42</sup>H. Emmerich, H. Löwen, R. Wittkowski, T. Gruhn, G. I. Tóth, G. Tegze, and L. Gránásy, "Phase-field-crystal models for condensed matter dynamics on atomic length and diffusive time scales: An overview," *Adv. Phys.* **61**, 665 (2012).
- <sup>43</sup>M. Salvalaglio and K. R. Elder, "Coarse-grained modeling of crystals by the amplitude expansion of the phase-field crystal model: An overview," *Modell. Simul. Mater. Sci. Eng.* **30**, 053001 (2022).
- <sup>44</sup>A. J. Archer, D. J. Ratliff, A. M. Rucklidge, and P. Subramanian, "Deriving phase field crystal theory from dynamical density functional theory: Consequences of the approximations," *Phys. Rev. E* **100**, 022140 (2019).
- <sup>45</sup>D. L. Coelho, D. Burns, E. Wilson, and N. Provatas, "Generalizing the structural phase field crystal approach for modeling solid-liquid-vapor phase transformations in pure materials," [arXiv:2408.10992v1](https://arxiv.org/abs/2408.10992v1) (2024).
- <sup>46</sup>J. F. Lutsko and M. A. Durán-Olivencia, "A two-parameter extension of classical nucleation theory," *J. Phys.: Condens. Matter* **27**, 235101 (2015).
- <sup>47</sup>X. Wang, S. Ramírez-Hinestrosa, J. Dobnikar, and D. Frenkel, "The Lennard-Jones potential: When (not) to use it," *Phys. Chem. Chem. Phys.* **22**, 10624 (2020).
- <sup>48</sup>J. F. Lutsko, "Explicitly stable fundamental-measure-theory models for classical density functional theory," *Phys. Rev. E* **102**, 062137 (2020).
- <sup>49</sup>V. Hernandez, J. E. Roman, and V. Vidal, "SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems," *ACM Trans. Math. Software* **31**, 351 (2005).
- <sup>50</sup>V. Hernandez, J. E. Roman, and V. Vidal, "SLEPc: Scalable library for eigenvalue problem computations," *Lect. Notes Comput. Sci.* **2565**, 377 (2003).
- <sup>51</sup>J. E. Roman, C. Campos, L. Dalcin, E. Romero, and A. Tomas, *SLEPc Users Manual*, Tech. Rep. DSIC-II/24/02 - Revision 3.19 (D. Sistemas Informàtics I Computació Universitat Politècnica de València, 2023).
- <sup>52</sup>V. Hernández, J. E. Román, and A. Tomás, "Parallel Arnoldi eigensolvers with enhanced scalability via global communications rearrangement," *Parallel Comput.* **33**, 521 (2007).
- <sup>53</sup>H. Jónsson and H. C. Andersen, "Icosahedral ordering in the Lennard-Jones liquid and glass," *Phys. Rev. Lett.* **60**, 2295 (1988).
- <sup>54</sup>A. Hirata, L. J. Kang, T. Fujita, B. Klumov, K. Matsue, M. Kotani, A. R. Yavari, and M. W. Chen, "Geometric frustration of icosahedron in metallic glasses," *Science* **341**, 376 (2013).
- <sup>55</sup>N. Singh, Z. Zhang, A. K. Sood, W. Kob, and R. Ganapathy, "Intermediate-range order governs dynamics in dense colloidal liquids," *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2300923120 (2023).
- <sup>56</sup>B. de Nijs, S. Dussi, F. Smalenburg, J. D. Meeldijk, D. J. Groenendijk, L. Filion, A. Imhof, A. van Blaaderen, and M. Dijkstra, "Entropy-driven formation of large icosahedral colloidal clusters by spherical confinement," *Nat. Mater.* **14**, 56 (2015).
- <sup>57</sup>C. F. Mbah, J. Wang, S. Englisch, P. Bommineni, N. R. Varela-Rosales, E. Spiecker, N. Vogel, M. Engel, and M. Engel, "Early-stage bifurcation of crystallization in a sphere," *Nat. Commun.* **14**, 5299 (2023).
- <sup>58</sup>P. G. Vekilov, "Two-step mechanism for the nucleation of crystals from solution," *J. Cryst. Growth* **275**, 65 (2005), part of Special Issue: Proceedings of the 14th International Conference on Crystal Growth and the 12th International Conference on Vapor Growth and Epitaxy.
- <sup>59</sup>P. G. Vekilov and M. A. Vorontsova, "Nucleation precursors in protein crystallization," *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **70**, 271 (2014).
- <sup>60</sup>J. F. Lutsko, "Mechanism for the stabilization of protein clusters above the solubility curve: The role of non-ideal chemical reactions," *J. Phys.: Condens. Matter* **28**, 244020 (2016).
- <sup>61</sup>R. Graham, "Path integral formulation of general diffusion processes," *Z. Phys. B* **26**, 281 (1977).
- <sup>62</sup>R. Graham, "Covariant formulation of non-equilibrium statistical thermodynamics," *Z. Phys. B* **26**, 397–405 (1977).
- <sup>63</sup>M. Gallo, F. Magaletti, D. Cocco, and C. M. Casiola, "Nucleation and growth dynamics of vapour bubbles," *J. Fluid Mech.* **883**, A14 (2020).

- <sup>64</sup>J. F. Lutsko, “Nucleation of colloids and macromolecules: Does the nucleation pathway matter?,” *J. Chem. Phys.* **136**, 134502 (2012).
- <sup>65</sup>J. F. Lutsko and C. Schoonen, “Classical density-functional theory applied to the solid state,” *Phys. Rev. E* **102**, 062136 (2020).
- <sup>66</sup>D. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
- <sup>67</sup>S. Balay, S. Abhyankar, M. F. Adams, S. Benson, J. Brown, P. Brune, K. Buschelman, E. Constantinescu, L. Dalcin, A. Dener, V. Eijkhout, J. Faibusowitsch, W. D. Gropp, V. Hapla, T. Isaac, P. Jolivet, D. Karpeev, D. Kaushik, M. G. Knepley, F. Kong, S. Kruger, D. A. May, L. C. McInnes, R. T. Mills, L. Mitchell, T. Munson, J. E. Roman, K. Rupp, P. Sanan, J. Sarich, B. F. Smith, S. Zampini, H. Zhang, H. Zhang, and J. Zhang, “PETSc/TAO users manual,” Tech. Rep. ANL-21/39 - Revision 3.20 (Argonne National Laboratory, 2023).
- <sup>68</sup>S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith, “Efficient management of parallelism in object oriented numerical software libraries,” in *Modern Software Tools in Scientific Computing*, edited by E. Arge, A. M. Bruaset and H. P. Langtangen (Birkhäuser Press, 1997), pp. 163–202.
- <sup>69</sup>C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. (Springer, Berlin, 2004).