Phase diagrams—why they matter and how to predict them

Pin Yu Chew and Aleks Reinhardt
Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

(Dated: 28 November 2022)

Understanding the thermodynamic stability and metastability of materials can help us to gauge for example whether crystalline polymorphs in pharmaceutical formulations are likely to be durable. It can also help us to design experimental routes to novel phases with potentially interesting properties. In this article, we provide an overview of how thermodynamic phase behaviour can be quantified both in computer simulations and using machine-learning approaches to determining phase diagrams, as well as combinations of the two. We review the basic workflow of free-energy computations for condensed phases, including some practical implementation advice, ranging from the Frenkel–Ladd approach to thermodynamic integration and to direct-coexistence simulations. We illustrate the applications of such methods on a range of systems from materials chemistry to biological phase separation. Finally, we outline some challenges, questions and practical applications of phase-diagram determination which we believe are likely to be possible to address in the near future using such state-of-the-art free-energy calculations, which may provide fundamental insight into separation processes using multicomponent solvents.

I. INTRODUCTION

Knowing a material's thermodynamic stability under different conditions, as summarised in a phase diagram, is of considerable practical importance with numerous technological applications, for instance in separation processes.[1] In simple cases, an experimental determination of phase diagrams is often sufficient, but as the building blocks become more complex, it is not always clear which polymorphs may be stable, and if they are not, what the limits of their metastability may be. For example, the phase diagram of water has been studied for over a hundred years with progressively more solid phases discovered and characterised,[2,3] while the famous case of HIV drug ritonavir suddenly converting into a largely insoluble more stable polymorph is thought not to be an isolated case amongst pharmaceutical compounds.[4]

Questions of thermodynamic stability and metastability can in principle be addressed using statistical mechanics and computer simulations, as first shown by Hoover and Ree in the context of hard-sphere melting back in 1968.[5] On the one hand, computing phase diagrams in silico can be very helpful under conditions that are difficult to achieve experimentally, such as when studying high-pressure behaviour relevant to planetary cores[6–9] or the properties of synthetic elements,[10] and can thus help guide experimental efforts. On the other hand, phase diagrams are very sensitive to the potentials used to describe the building blocks themselves[11–14] and a potential that captures phase behaviour accurately will often also describe other properties of a substance well. Classical potentials can therefore be parameterised by fine-tuning their ability to reproduce phase behaviour; such an approach has fruitfully been used to design ever better water[15–25] and protein[26–28] models.

Indeed, when computing phase diagrams with computer simulations, the first step is generally to choose how to simulate the material of interest. This can range from quantum-mechanically accurate potentials[29] based on either wavefunction or density-functional theory (DFT) electronic structure calculations, to classical potentials parameterised either from quantum simulations[30] or from experiment,[31] to coarse-grained[32–35] and 'toy model' potentials.[36–44] Such an approach has been used to investigate the phase behaviour of water,[7–9,45,46] gallium,[47] supercritical hydrogen[6] and titanium dioxide.[48]

The computation of phase diagrams has been thoroughly reviewed in an excellent paper by Vega and co-workers.[18] In this perspective, we briefly summarise the fundamental principles, as well as review some more recent work on determining phase diagrams in computer simulations, and identify some of the trickier aspects of the process and caveats involved. We also provide a short overview of some machine-learning methods that have recently been used to predict phase diagrams without the need for expensive molecular simulations. Finally, we speculate about some possible future applications of the methods that have been developed to investigate phase behaviour.

II. STRATEGIES FOR COMPUTING CHEMICAL POTENTIALS

When the temperatures, pressures and chemical potentials of two phases are equal, they are at equilibrium and are said to coexist. It is straightforward in computer simulations to fix the temperature and pressure, but the chemical potential is often more difficult to compute because it entails a 'thermal' (i.e. non-mechanical) component, namely the entropy, that cannot be estimated by sampling. Perhaps the simplest approach to computing the phase behaviour of a system is therefore to assume that the chemical potential is completely dominated by the enthalpic term, and that the entropy differences between phases are negligible. Such an approach is advantageous because the enthalpy is a simple mechanical observable that can readily be determined, and is often not unreasonable when dealing with phases of a similar structure and thus similar entropy. In the context of solid phases, the phase behaviour is in many cases completely dominated by enthalpic terms and the entropy can be relatively unimportant.[12,49] For example, many such 'absolute-zero' phase diagrams have been reported for titanium dioxide.[50–57] However, when competing phases have relatively similar enthalpies, entropy differences can be crucial in controlling the phase behaviour.[58] Sometimes, approximations can be made to estimate the free energy as a function of temperature.[56,58–63]

Even if an approximate zero-temperature phase diagram is to be calculated, a further challenge remains when predicting phase diagrams of solid phases: how can we ensure that all potentially stable polymorphs have been identified? We can consider any phases known experimentally, but a significant advantage of computer simulation is precisely that we may be

able to predict phases that are not yet known experimentally and thus help steer experimental exploration of new phases. There are numerous approaches that can help us to identify potentially stable crystal phases.[64–68] Most such approaches usually entail some stochastic step to identify potentially unseen phases and a subsequent energy minimisation step, often followed by identification of the crystal space-group symmetry, for example with tools like FINDSYM.[69] Broadly speaking, the larger a crystal's unit cell, the more difficult it is to find by random searching. Since energy minimisation is usually involved, these approaches are often able to identify polymorphs that are competitive at very low temperatures, but may be less adept at identifying high-temperature phases that are only entropically stabilised. In random searches, low-enthalpy structures have been shown to occur more frequently and so have a larger basin of attraction;[70] indeed, the volume of the basin of attraction of each minimum can be approximated by the number of times each structure is found in a random search, and this can provide a crude first approximation of the relative entropies of the competing polymorphs.[49]

A. Direct estimation of chemical potentials

If we wish to compute the chemical potential in computer simulations, we can note that, although the chemical potential cannot directly be determined as a canonical average over the phase space, there are several approaches that we can take to compute it. Using the Widom insertion method, we can relate the excess chemical potential to the thermal average of a Boltzmann factor for the change in energy $\Delta U(N, N+1)$ when an additional particle is randomly inserted into a system of N particles,[71] i.e.

$$\mu^{\mathrm{ex}} = -k_{\mathrm{B}}T \ln\left[\int \langle \exp(-\beta \Delta U(N, N+1))\rangle_N \, \mathrm{d}\boldsymbol{r}_{N+1}\right]. \quad (1)$$

This and similar methods, such as the Gibbs ensemble approach,[72,73] can be used to determine the phase behaviour of systems that are sufficiently dilute to enable additional particles to be inserted with a non-negligible probability. Phase switching[74,75] is an approach based on a similar idea, but is also applicable to solid systems.

B. Thermodynamic integration

For denser systems, however, we can also note that although the chemical potential is a thermal quantity, its derivatives are mechanical observables. For example, from the Gibbs–Duhem relation $\mathrm{d}\mu = v\,\mathrm{d}P - s\,\mathrm{d}T$ (where μ is the chemical potential, v is the volume per particle, P is the pressure, s is the entropy per particle and T is the absolute temperature), we can write $(\partial\mu/\partial P)_T = v$. By integrating this equation numerically from some initial pressure P_0, we obtain

$$\mu(P) = \mu(P_0) + \int_{P_0}^{P} v(P')\,\mathrm{d}P'. \quad (2)$$

The analogous derivative with respect to temperature ($(\partial\mu/\partial T)_P = s$) is unhelpful, since the entropy is also not a mechanical observable, but we can instead use the Gibbs–Helmholtz equation,

$$\left(\frac{\partial(G/T)}{\partial T}\right)_{N,P} = -H/T^2, \quad (3)$$

where G is the Gibbs energy, N is the number of particles and H is the enthalpy. Since for a one-component system $G = N\mu$, this thus relates the derivative of the chemical potential to another mechanical observable, the enthalpy. Similar derivatives can be constructed where variables other than the pressure or the temperature are held fixed, e.g. along iso-(βP) lines.[76,77]

1. Choice of reference state

In principle, if we can determine how the enthalpy and the density of a system change as a function of the thermodynamic variables of interest, we can also determine how the chemical potential of a phase changes relative to the starting point. However, this does not yet give us an 'absolute' chemical potential of the phase in question, since the starting point is arbitrary. We cannot in general integrate different phases to the same starting point, since paths along which we perform thermodynamic integration must be reversible to maintain numerical stability. In order to ensure that chemical potentials of different phases are expressed relative to a common origin, there are several possible strategies. The simplest strategy would be to determine a coexistence point of each phase of interest with another common phase (e.g. the liquid or the vapour); this is the basic idea behind direct-coexistence simulations (Sec. II D). It is also possible to perform thermodynamic integration along a non-physical reversible path from a liquid to a solid,[78] such that the chemical potential is well defined throughout the process.

Alternatively, we can relate the chemical potential to a state whose free energy can be computed analytically. There are not many such states, but examples include the perfect gas and the Einstein and Debye crystals.[18,79–82] However, such reference states with analytically computable partition functions are simple non-interacting systems, and to relate them to an interacting system, we need to find a reversible path from the reference state to the conditions of interest. For fluid systems, this can often readily be achieved simply by reducing the density until the particles are on average too far to interact; if the system has long-ranged electrostatic interactions, these can often be switched off gradually (see Sec. II B 2 below). In principle, the Helmholtz energy could be integrated in volume using the standard relation $(\partial A/\partial V)_{N,T} = -P$ expressed as an integral in density,

$$\beta a(\rho_1) = \beta a(\rho_0) + \int_{\rho_0}^{\rho_1} \frac{\beta P}{\rho^2}\,\mathrm{d}\rho, \quad (4)$$

where $\beta = 1/k_{\mathrm{B}}T$ and $a = A/N$ is the Helmholtz energy per particle. We could then use the ideal Helmholtz energy for the reference state $\beta a(\rho_0) = \beta a^{\mathrm{id}} = \ln(\rho\Lambda^3) - 1$, where Λ is the de Broglie thermal wavelength. The difficulty with this expression is that the integrand scales as $1/\rho$ at low densities (for which $P \to P^{\mathrm{id}} = \rho k_{\mathrm{B}}T$), whilst $\lim_{x\to 0} \ln x = -\infty$. To maintain numerical stability, it is therefore usual to compute the excess Helmholtz energy, $a^{\mathrm{ex}} \equiv a - a^{\mathrm{id}}$, using the relation[18]

$$\beta a^{\mathrm{ex}}(\rho_1) = \int_{\rho_0}^{\rho_1} \left[\frac{\beta P}{\rho^2} - \frac{1}{\rho}\right] \mathrm{d}\rho, \quad (5)$$

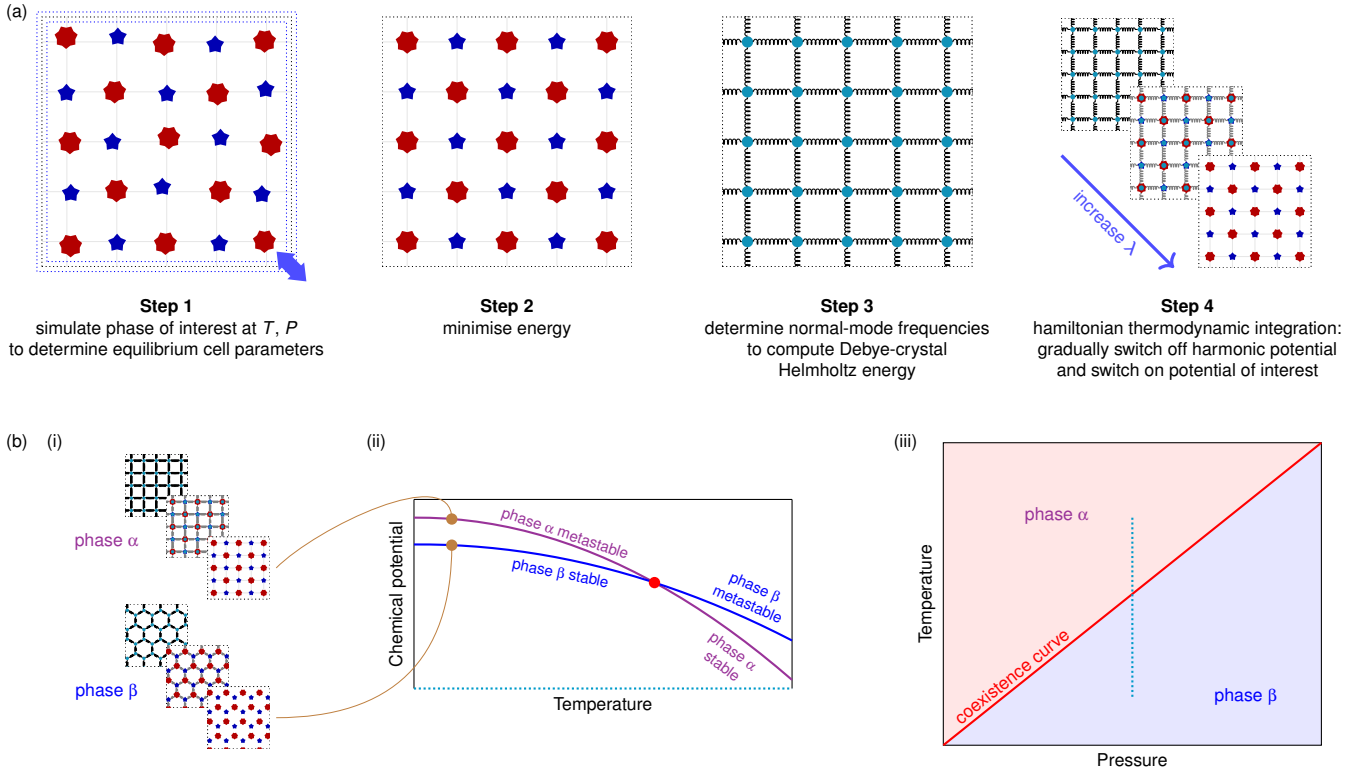

Figure 1. (a) A schematic illustration of the steps involved in determining the initial chemical potential of a solid phase. (b) A schematic of the procedure to determine the phase diagram. The process in panel (a) can be applied to different phases, here labelled α and β [panel (i)]. This leads to two initial chemical potentials at the initial conditions investigated, shown as brown points in panel (ii). Thermodynamic integration at a constant pressure from these initial points can be used to compute the chemical potential at other temperatures [panel (ii)]. In this illustration, at low temperatures, phase β is thermodynamically stable and phase α is metastable, while at higher temperatures the converse holds. The point labelled in red is the point at which the temperatures, pressures and chemical potentials of the two phases are equal, and is therefore part of the coexistence curve [panel (iii)]. The data in panel (ii) correspond to a slice of panel (iii), as indicated by the dotted cyan line. The rest of the coexistence curve can be obtained for example by repeating the entire calculation at other pressures; by using thermodynamic integration as a function of pressure and then repeating step (ii) to give another coexistence point; or by using a method such as Gibbs–Duhem integration.

with $a^{\mathrm{id}}(\rho_1)$ added on at the density of interest ρ_1, rather than in the limit of low density, in the final step. In this integration, as the density tends to zero, the integrand tends to the second virial coefficient,[18,38] which can be computed independently in a Monte Carlo (MC) calculation.

2. *Artificial thermodynamic integration*

For solids and systems with long-ranged interactions, another approach is to define a reference potential U_0 corresponding to a simpler system whose partition function is calculable, and relate it to U_1, the potential of interest, often with a simple linear scaling such as $U(\lambda) = \lambda U_1 + (1-\lambda)U_0$.[83] When the parameter λ is zero, this potential is equivalent to the reference potential, whilst when it is unity, it is equivalent to the potential of interest. The canonical partition function of this potential, $Q(\lambda) = \int \exp(-\beta U(\boldsymbol{r}^N;\lambda)\,\mathrm{d}\boldsymbol{r}^N$, is of course also a function of λ, and is in principle not feasible to compute analytically. However, as we have noted, derivatives of the free energy are often mechanical observables, and using the product rule and the bridge relation $A = -k_{\mathrm{B}}T\ln Q$, we can readily compute that

$$\left(\frac{\partial A(\lambda)}{\partial\lambda}\right)_{NVT} = \frac{\int\left(\frac{\partial U(\boldsymbol{r}^N;\lambda)}{\partial\lambda}\right)\exp\left[-\beta U(\boldsymbol{r}^N;\lambda)\right]\mathrm{d}\boldsymbol{r}^N}{\int\exp\left[-\beta U(\boldsymbol{r}^N;\lambda)\right]\mathrm{d}\boldsymbol{r}^N} = \left\langle\frac{\partial U(\lambda)}{\partial\lambda}\right\rangle_\lambda. \tag{6}$$

Since by construction we can compute $A_0 \equiv A(\lambda=0)$ analytically, we can compute the Helmholtz energy of the potential of interest as

$$A_1 = A_0 + \int_0^1 \langle U_1 - U_0\rangle_\lambda\,\mathrm{d}\lambda, \tag{7}$$

provided the path is reversible. This approach is known as artificial or hamiltonian thermodynamic integration,[79] and with it, we can compute an initial Helmholtz energy of the system with the potential of interest. If we add a suitable PV term,[80] we can also determine the Gibbs energy and hence the absolute chemical potential.

3. *Einstein and Debye crystals*

For crystalline phases, a possible reference system may be the Einstein crystal, a very simple model of a crystal where particles

are tethered to their lattice positions by harmonic springs. It is straightforward to compute the canonical partition function of such a crystal,[18,79] even for non-spherical molecules.[84] The spring constants can then gradually be made weaker whilst the potential of interest is switched on using hamiltonian thermodynamic integration, as discussed above. In practice, to reduce the scope for numerical issues in the integration, an additional constraint of a fixed centre of mass is introduced; this can be corrected for analytically, but with reasonable system sizes, the correction arising from this constraint is in any case very small.[80] An alternative approach is to constrain the position of a single molecule.[85] Another possible crystalline reference state is the Debye crystal, where harmonic springs connect atoms in such a way that their frequencies correspond to the normal-mode frequencies of the original crystal.

In the normal-mode approximation,[86] we first expand the potential energy in a Taylor series to second (harmonic) order,

$$U(\boldsymbol{R}) \approx U_{\text{harm}}(\boldsymbol{R}) = U(\boldsymbol{R}_{\text{min}}) + \frac{1}{2}\sum_{i,j}^{3N} \delta R_i \delta R_j \left(\frac{\partial^2 U}{\partial R_i \partial R_j}\right)_{\text{min}}, \tag{8}$$

where $U(\boldsymbol{R})$ is the potential energy of the system as a function of all the $3N$ co-ordinates of each atom, denoted by $\boldsymbol{R}$. The second derivatives are evaluated at the energy minimum, and make up matrix elements H_{ij} of the hessian matrix $\boldsymbol{H}$. Since heavier atoms move less than lighter ones in each normal mode, it is convenient to introduce mass-weighted displacements with components $q_i = (m_i)^{1/2}\delta R_i$, and a mass-weighted hessian matrix with matrix elements $K_{ij} = U_{ij}/(m_i m_j)^{1/2}$. In this representation, the potential energy, relative to the minimum energy, can be written to harmonic order as

$$U_{\text{harm}}(\boldsymbol{q}) - U(\boldsymbol{R}_{\text{min}}) = \frac{1}{2}\sum_{i,j}^{3N} q_i K_{ij} q_j. \tag{9}$$

Although all the $3N$ harmonic oscillators are coupled, since the matrix $\boldsymbol{K}$ is symmetric, we can find a similarity transform to diagonalise it, $\widetilde{\boldsymbol{C}}\boldsymbol{K}\boldsymbol{C}$, where $\boldsymbol{C}$ is the matrix of normalised eigenvectors of $\boldsymbol{K}$, and $\widetilde{\boldsymbol{C}}$ is its transpose. In this 'normal mode' representation, $\widetilde{\boldsymbol{Q}} = \widetilde{\boldsymbol{q}}\boldsymbol{C}$, the potential energy is made up of $3N$ uncoupled harmonic oscillator terms,

$$U_{\text{harm}}(\boldsymbol{Q}) - U(\boldsymbol{R}_{\text{min}}) = \frac{1}{2}\sum_{i}^{3N} \lambda_i Q_i^2, \tag{10}$$

where $\lambda_i = \omega_i^2$ are the eigenvalues of the matrix $\boldsymbol{K}$, and ω_i are the angular frequencies of the corresponding oscillators. In the Debye-crystal approach, in the first instance we therefore need to compute the normal-mode frequencies. To achieve this, we first equilibrate a system at the conditions of interest [Fig. 1(a), Step 1]. We then minimise the potential energy of an example configuration of the target solid at the correct density for the temperature of interest [Fig. 1(a), Step 2]. We can achieve this using steepest descent, conjugate gradient minimisation[87] or similar optimisation methods.[88] To determine the matrix of second partial derivatives [Fig. 1(a), Step 3], we can use a finite-difference approach; namely, for each pair of degrees of freedom i and j, we can estimate the hessian matrix element H_{ij} by moving the particles in question by some small distance $\pm\delta$,

$$H_{ij} \approx \frac{1}{4\delta^2}\big[U(R_i+\delta,\, R_j+\delta) + U(R_i-\delta,\, R_j-\delta) - U(R_i+\delta,\, R_j) - U(R_i,\, R_j+\delta)\big], \tag{11}$$

where the remaining degrees of freedom are unchanged. Once the full mass-weighted hessian matrix is obtained, we find its eigenvalues $\lambda_i = \omega_i^2$, i.e. the squares of the normal-mode angular frequencies. Since each of the normal-mode harmonic oscillators is uncoupled by construction, we can easily find the corresponding partition function and in turn the Helmholtz energy. In particular, apart from the three translational modes with zero eigenvalues, we can obtain the classical harmonic Helmholtz energy of the harmonic crystal by summing over the Helmholtz energies of each normal-mode oscillator [Fig. 1(a), Step 3],

$$\beta A_{\text{harm}} = \beta U(\boldsymbol{R}_{\text{min}}) + \sum_{i=1}^{3N-3} \ln\frac{\hbar\omega_i}{k_{\text{B}}T}. \tag{12}$$

Finally, we can use the hessian matrix to define the energy function of the Debye crystal, i.e. Eq. (8), and use hamiltonian thermodynamic integration [Eq. 7] to couple it to the overall potential energy [Fig. 1(a), Step 4]. Since the Debye crystal energy is just the harmonic part of the overall potential energy, at least at reasonably low temperatures the difference with the true potential is small, and so the final hamiltonian thermodynamic integration is usually smooth and results in a small anharmonic correction to the free energy. The Debye-crystal route has therefore been suggested to offer an approach that is often numerically better behaved than the Einstein crystal route,[89,90] although the final free energy should in principle be the same with either method.

Once the absolute chemical potential is known at one set of conditions, we can use thermodynamic integration to determine the form of the chemical potential as a function of, for example, the pressure or the temperature. Coexistence points can then be identified by finding where chemical potential curves of different phases cross [Fig. 1(b)].

At sufficiently low temperatures, the entropic part of the chemical potential is always negligible. In principle, if all solid phases of interest were stable at low enough temperatures, one could simply use thermodynamic integration to determine the change in ΔG as a function of temperature to determine high-temperature phase behaviour. However, simulations at very low temperature are often slow and require long equilibration times. The harmonic approximation is usually reasonable[80] at somewhat higher temperatures, where equilibration is usually less problematic, and so it is again often possible to use thermodynamic integration alone, without the full Einstein-crystal formalism. This approach however does not work if the phase of interest is not (meta)stable at temperatures at which the approximation is reasonable.

In analytical reference states for both fluid and solid phases, the canonical partition function and hence the chemical potential depend on the de Broglie wavelength once momentum degrees of freedom have been integrated over. In principle, in classical statistical mechanics, the choice of the de Broglie wavelength cannot affect thermodynamic properties, as it is shifts the free energy by the same amount across all phases under given conditions. Even though $\Lambda^2 = h^2/2\pi m k_{\text{B}}T$ depends on temperature, it is in practice often chosen to be a

fixed value, such as $\Lambda = 1$ Å, for numerical convenience.[18] If this is the case, care must be taken to exclude the translational kinetic energy from other thermodynamic integrations (e.g. in the enthalpy when integrating Eq. (3)) in order to ensure thermodynamic consistency.[12] Finally, it ought to be borne in mind that the above discussion applies to systems governed by classical mechanics. If the momentum degrees of freedom are not factorisable, i.e. in systems where quantum effects may be significant, the translational part of the partition function can no longer straightforwardly be integrated out. The quantum kinetic energy depends on the environment[91–93] and is therefore no longer phase-independent. When exploring the role of nuclear quantum effects on phase behaviour, it is thus often convenient to compute the chemical potentials of the corresponding classical system and then add a quantum correction in a separate step.[91,94]

C. Estimating the density of states

Instead of using thermodynamic integration to known reference states, it is also possible to estimate the density of states in a Wang–Landau simulation[95–98] and in turn to compute the free energy of a system. In a usual Wang–Landau calculation in the canonical ensemble, the density of states of a given system, $g(E)$, is approximated by constructing a histogram with a random walk biassed towards visiting previously unsampled states. The initial estimate for the density of states can be $g(E) = 1$ if no information is known about the system, although a reasonable initial guess can speed up convergence in some cases.[99,100] A Monte Carlo acceptance probability between states 1 and 2 is computed as $\min[1, g(E_1)/g(E_2)]$. Each time a state with energy E is visited, its corresponding density-of-states histogram value is multiplied by a modifier $f > 1$ so that $g(E) \leftarrow g(E)f$, and a histogram of how many times each energy state was visited is incremented by one. The process is repeated until the histogram shows a uniform sampling of relevant states. Since $g(E)$ changes between Monte Carlo steps, the Wang–Landau algorithm does not obey detailed balance,[99] although if f is progressively decreased towards unity to minimise the final error, this is not usually a significant concern. The choice of how f is scaled can significantly affect the algorithm's performance.[101] Finally, thermodynamic properties such as the Helmholtz energy can be computed as[96]

$$A = -k_\mathrm{B}T \ln Q \approx -k_\mathrm{B}T \ln\left[\sum_{\text{bins } i} g(E_i) \exp(-\beta E_i)\right], \qquad (13)$$

where Q is the canonical partition function that we approximate by summing over the bins of the histogram. All other thermodynamic quantities can then be computed from the Helmholtz energy. A similar approach can be used in the grand ensemble.[98]

The Wang–Landau sampling approach and its analogues have been very successful in modelling the behaviour of lattice models such as the Ising model,[96,97] the XY model[102] and even lattice-based liquid crystals[103] and alloys,[104] as well as in modelling the phase behaviour of fluids[98,105,106] and solutions.[107] In lattice models with discrete energy levels, the minimum and maximum energy to consider is usually well defined. By contrast, in systems with continuous energy levels, the Wang–Landau approach of binning energies into histograms and imposing a minimum and maximum energy to consider can lead to inefficiencies and errors; systematic discretisation methods[108] have been proposed to address this problem.

Moreover, since solid-state densities of states are usually much narrower than those of liquid phases, approaches based on densities of states are likely to find supercooled liquid structures instead of solid phases in a random walk.[109] It has been proposed that starting from low-energy structures and partitioning the energy space upwards permits the calculation of the density of states of known solid structures.[110] It is also possible to bias the sampling of solid states with a suitable order parameter, for example using umbrella sampling.[111]

Another strategy for estimating the density of states is known as nested sampling.[112–114] The approach relies on estimating the degeneracy of a given energy level by generating a pool of K random configurations, identifying the configuration with the highest energy, and replacing it with a new random configuration that has a lower energy. In each step of the procedure, we can estimate, relative to an unimportant origin, the volume of phase space at the maximum energy as $\Phi(i) = [K/(K+1)]^i$,[114] where i is the iteration number. The density of states can then be computed as $g(i) \approx \Phi(i) - \Phi(i-1)$. This approach does not require the binning of energies; however, its ability to find solid phases is similarly limited as the Wang–Landau-based approaches mentioned above. In particular, a random sampling of position co-ordinates corresponds effectively to sampling at a very high temperature, where the large entropy of vapour and liquid states completely dominates the system's behaviour, and it is very unlikely, unless a very large number of configurations is sampled, that potential energy wells of solid phases will be adequately represented. However, such states are usually possible to obtain by energy minimisation; a way of computing the density of states that captures solid phases that dominate the phase behaviour at low temperatures entails a combination of optimisation and nested sampling.[115,116]

D. Direct coexistence

As an alternative to computing 'absolute' chemical potentials relative to known analytical models, we can instead determine phase coexistence analogously to its experimental determination: by simulating the two coexisting phases explicitly in a 'direct-coexistence' simulation.[117,118] In this approach, we typically simulate a system in a slab geometry in an elongated periodic box with two explicit interfaces between the phases in question, one on each side of the two coexisting bulk phases. We usually run simulations at a given pressure over a range of temperatures and then bracket regions where a given phase shrinks or grows to determine the coexistence temperature, although simulations in other ensembles (e.g. microcanonical or canonical) are also possible.[119] This approach was introduced in the 1970s as a way of investigating both the interfacial structure[117] and the thermodynamic behaviour[118] of Lennard-Jones (LJ) particles. As computational power has increased and much larger system sizes could be simulated, the method has been used in progressively more contexts since, from hard spheres[120–122] to metals[123,124] to water.[125,126]

The main advantage of direct-coexistence simulations is their relative simplicity: often, they are very easy to implement, since in principle the system is just evolved using normal MC or molecular dynamics (MD) algorithms. However, there are

also several possible disadvantages. The simulation relies on the stable phase growing rapidly at the expense of the unstable phase when away from their coexistence temperature. If the two phases in question have very similar chemical-potential gradients, the driving force for the phase transition may not be sufficient to enable the coexistence point to be bracketed accurately. Moreover, even if the driving force is significant but the phase transition is dynamically slow, direct-coexistence methods are unlikely to be productive.

There are further practical considerations that may limit the applicability of direct-coexistence simulations. When computing chemical potentials directly, pure phases are simulated, and relatively small system sizes are often sufficient to obtain accurate results. By contrast, in direct-coexistence simulations, an explicit interface forms between the bulk phases. Although in the thermodynamic limit, the role of the interface is immaterial, typical simulation-box sizes are much smaller and the proportion of particles at the interface is non-negligible. The interface can thus dominate the system's apparent behaviour. With large enough system sizes, such finite-size effects in theory disappear, and so probing the behaviour as a function of simulation-box size is especially important in such simulations. In practice, even if the system is in principle large enough to simulate coexistence, large blocks of the two coexisting phases must still be brought into direct contact and the interface properly equilibrated at each set of conditions being considered, which can be a laborious process. The larger the system size, the longer such initial equilibration will take, while for smaller systems, an unfortunate initial, pre-equilibration choice of interface structuring may result in such large restructuring of the coexisting phases that they disappear altogether even if they are not far from coexistence. When explicit interfaces are present, particular care must be exercised in interpreting simulation results when using truncated potentials.[14,127]

When one of the phases in question is a solid, further difficulties arise, since the simulation box must be compatible with the unit cell at the conditions studied to prevent placing stress on the solid.[128–130] Practically, this can be addressed by (i) determining the crystal's unit cell parameters as a function of pressure and temperature, (ii) scaling the box with an interface to these dimensions at the pressure and temperature studied, and (iii) applying a barostat only in the direction orthogonal to the interface.[18] This approach relies on the coexisting phase being able to adapt to the box shape and size consistent with the solid phase, which in general will only be possible with a fluid phase. However, this does not preclude direct-coexistence simulations from being used in determining solid–solid coexistence lines as long as both solids can coexist with a fluid phase, even if they are only metastable.[126]

Despite some of these potential complications with using direct-coexistence simulations in determining phase diagrams, direct-coexistence simulations have especially recently been very popular for studying what is effectively vapour–liquid coexistence in biomolecular mixtures.[26–28,131–135] In such systems, proteins and nucleic acids are typically modelled using fast, coarse-grained potentials with implicit solvents, enabling any interfaces to equilibrate readily and coexistence properties to be determined easily. Typically, such simulations are performed in the canonical ensemble at a fixed overall density held somewhere in the range where phase separation occurs; this typically corresponds to very low coexistence pressures.[14] Since phase diagrams are increasingly being reported in the experimental literature,[136] the computation of phase diagrams, primarily using direct coexistence, has been instrumental in both designing and benchmarking numerous coarse-grained potentials[26,28,137] which have subsequently been used to probe the fundamentals of protein phase behaviour, including the effects of salt concentration,[138] valency,[134,139–141] composition (of both the individual proteins and of the entire mixture) and patterning/distribution of amino acids,[140,142,143] as well as evolving protein sequences to promote or inhibit their capacity to phase-separate.[133,135] As such models evolve and account for more complex systems, it may well be that the limitations of direct-coexistence simulations become more apparent, especially in the light of large systems sizes that may need to be simulated to avoid artefacts. Other free-energy methods, such as thermodynamic integration, may thus become more broadly used in this field in due course.

E. Interface pinning

In direct-coexistence simulations, the coexistence point is usually determined by varying the thermodynamic parameters until one phase grows at the expense of the other. However, rather than needing to rely on potentially slow equilibration under different conditions, we can ensure that the system remains in a two-phase state by applying a bias to the system in an approach known as interface pinning.[144–146] The bias applied is related to the free-energy difference between the two phases.

An order parameter must first be identified that can distinguish between the two phases of interest. Often, this is achieved by probing the local environment around each particle, for example with spherical-harmonic functions.[147–153] The spherical harmonics form a complete orthonormal system and any real function on the unit sphere – such as the particle neighbour density – can be expressed using the spherical harmonics as a basis set in a Laplace series.[154] It can often be computationally more efficient to use real spherical harmonics instead of their more usual complex analogues.[155] In typical simulation studies, since order-parameter calculations often need to be repeated many times, typically only one term in the series is picked that can best distinguish between the phases of interest[147] to help reduce the computational expense. Other order parameters may also be suitable,[156–158] depending on the system studied, and for simple solid–fluid interfaces, a simple density-field approach is often sufficient.[145]

As long as an order parameter has been found that can distinguish between two phases, say α and β, we can define the total number of particles as $N = N_\alpha + N_\beta + N_{\text{int}}$, where N_{int} is the number of particles at the interface. In a slab geometry, like with direct-coexistence simulations, the interfacial area is minimised by forming two planar interfaces across the smallest box dimension, and as long as the interface remains planar, the number of interfacial particles and the resulting interfacial contribution to the free energy should not change as the interface moves. The total Gibbs energy is thus given by $G = N_\alpha\mu_\alpha + N_\beta\mu_\beta + G_{\text{int}}$ or equivalently $G = N_\alpha\Delta\mu + \text{constant}$, where $\Delta\mu = \mu_\alpha - \mu_\beta$ and the constant accounts for terms that do not depend on N_α. The chemical potential difference $\Delta\mu$ is the quantity we wish to determine; it is zero at coexistence, while for non-zero values the sign tells us which phase, α or β, is thermodynamically more stable.

To compute $\Delta\mu$, we can apply an additional bias potential to the system that depends on the global order parameter of the simulation box. In the following, we will use as an example N_α, the number of particles in phase α, as the order parameter for simplicity; however,it is not actually necessary for us to be able to determine the phase of each molecule individually: a global order parameter that changes between the bulk phase α and phase β is sufficient.[144,145]

One possibility is to compute the free energy directly using a method such as umbrella sampling.[111] If we apply a bias to the global order parameter such that we force the number of particles of phase α to change (i.e. the interface moves in a direction we bias it towards) and repeat the procedure at different target values of N_α, and then account for the bias, we find $G(N_\alpha)$ relative to an arbitrary origin for a range of N_α, which permits us to determine $\Delta\mu$ as the gradient. Alternatively, suppose the bias is of the harmonic form $U_{\text{bias}}(N_\alpha) = (\kappa/2)[N_\alpha - a]^2$, with a chosen to be close to the number of particles in phase α at the start of the simulation, and κ a tuneable parameter. The associated Gibbs energy is $G_{\text{biassed}} = G(N_\alpha) + U_{\text{bias}}(N_\alpha) = (\kappa/2)[N_\alpha - a + \Delta\mu/\kappa]^2 + \text{constant}$, where we have completed the square and ignored terms independent of N_α.[144,145] The Boltzmann probability for observing a particular number of particles in phase α is $P(N_\alpha) \propto \exp\{-\beta G_{\text{biassed}}\}$. Assuming that κ is chosen to be sufficiently large so that the interface does not move appreciably, and that the number of particles is sufficiently large to treat it as a continuous variable, we can normalise this probability by extending the integration limits in N_α to $\pm\infty$ without introducing significant error. The probability $P(N_\alpha)$ then becomes of a standard gaussian form. Finally, we can find

$$\langle N_\alpha \rangle = \int_{-\infty}^{\infty} N_\alpha P(N_\alpha)\, \mathrm{d}N_\alpha = a - \Delta\mu/\kappa. \tag{14}$$

We can therefore find $\Delta\mu$ simply by determining $\langle N_\alpha \rangle$ in a simulation with a bias.[144,145]

By changing the thermodynamic conditions and repeating this calculation, we can determine the point at which $\Delta\mu = 0$, i.e. where the phases α and β coexist. Since the two-phase system is stable if a sufficient bias is applied, such simulations are at equilibrium even when the phases are not at coexistence, and are therefore less sensitive to how they are initialised than direct-coexistence simulations. However, as with direct-coexistence simulations, since an explicit interface is present in the simulation box, the method usually requires considerably larger system sizes than approaches based on thermodynamic integration, and finite-size effects must be investigated carefully.[144] Similarly, because of the explicit interface required, it is not usually possible to use this approach to study solid–solid phase coexistence directly.

F. Determining coexistence curves

Coexistence curves on a phase diagram correspond to loci of points where two (or more) phases have the same temperatures, pressures and chemical potentials [see Fig. 1(b)(iii)]. We can determine a single coexistence point using the methods outlined so far, either by explicitly evaluating chemical potentials or by using a proxy method such as direct coexistence. Other points may then be determined by repeating the procedure at different initial conditions. Alternatively, the coexistence curve can itself be integrated. Perhaps the simplest approach is to run a series of simulations for a pair of pure phases that are known to coexist at a given point and, starting from this known point, numerically integrate the Clapeyron equation,

$$\frac{\mathrm{d}P}{\mathrm{d}T} = \frac{\Delta H}{T\Delta V}. \tag{15}$$

This is often known as Gibbs–Duhem integration,[18,159,160] although, potentially confusingly, it entails integrating the Clapeyron equation rather than the Gibbs–Duhem equation. Such an approach is often reasonable for integrating over relatively small ranges of temperature or pressure, but there is no inbuilt error-checking mechanism: small errors in the enthalpy or volume change are cumulative. Other approaches, such as multistate reweighting, are far more accurate.[161–163] Although their implementation may be more challenging, several open-source tools exist that make their use straightforward.

G. Analysing errors

Since there are many potential sources of error in free-energy calculations, both numerical and systematic, it can be difficult to determine and propagate errors.[14,18] It is usual in such calculations to employ a posteriori consistency checks[12,18,41,89] to ensure thermodynamic consistency. This is especially important if coexistence curves are directly integrated with methods like Gibbs–Duhem integration, but consistency checks are a valuable tool for finding implementation errors in all cases. For example, if one performs an Einstein-crystal calculation at two different pressures and temperatures for a given phase, it should be possible to integrate the chemical potential from one of these points to the other along many different isotherms and isobars, and all such reversible paths should give the same chemical potential (within numerical accuracy) as the Einstein-crystal calculation itself. At least a few coexistence points should also ideally be checked to ensure that coexistence temperatures and pressures determined from chemical-potential crossovers can be reproduced in direct-coexistence simulations. Of course it is not impossible to estimate errors in the chemical potential itself; for example, we can compute the chemical potential in several completely independent simulations and determine the corresponding standard deviation,[8] or use polynomial fitting to isotherm or isobar data to give prediction bands for the chemical potential.[9] However, such error estimation cannot account for any systematic errors, and consistency checks are usually rather more revealing than numerical error analysis.

H. Capturing long-timescale entropies

One significant potential difficulty with calculating chemical potentials by thermodynamic integration from known reference states is that some features of phases may be impossible to equilibrate at computationally accessible time- and length-scales. A well-studied example of this is proton disorder in ice. Many ice phases exist as proton-ordered and proton-disordered analogues,[8,164–170] with the former dominating at low temperatures where the stable phase is determined primarily by low enthalpy, and the latter dominating at high temperatures at which the higher entropic favourability of (partially) disordered

phases takes over. However, different manifestations of proton disorder, which contribute to the overall entropy of such phases, are not readily accessible to computer simulations. Moreover, if a particular proton-disordered configuration that is not fully representative of the phase in question is chosen, this can significantly affect the calculated phase behaviour; the calculated phase diagram of such phases can easily be wrong without any clear signs that anything is amiss.[126] Although in principle a direct-coexistence simulation should result in correct phase behaviour, in practice the proton disorder in the crystalline phases is largely locked in at all computationally accessible time scales. In calculations of water phase diagrams, proton disorder is therefore usually accounted for by assuming that the experimentally determined proton disorder is correct[126,171] and adding a suitable analytically computed residual entropy to the chemical potential.

Another example of a system whose entropy is difficult to capture are quasicrystals, which have no long-ranged translational order, and so cannot be described by a periodic lattice, but which often have long-ranged orientational ordering.[172] They are often thought of as projections of a higher-dimensional crystal into a lower dimension. Quasicrystals often, though not always,[173] arise because of a competition between multiple lengthscales, either in multicomponent mixtures of differently sized particles or in single-component mixtures with different intrinsic lengthscales[174,175] or in systems with explicit orientational order.[77,176,177] A knowledge of the phase behaviour of quasicrystals from simple models can help guide experimental design; for example, computer simulations of quasicrystals made from DNA nanostar motifs[178] have in part led to such soft binary quasicrystals being realised experimentally.[179]

When it comes to determining a quasicrystal's stability, the configurational disorder clearly entropically stabilises the quasicrystal, but is not straightforward to determine directly. Although simple models sometimes have such fast dynamics that an equation of state that accounts for bulk quasicrystalline phase behaviour can readily be determined in brute-force simulations,[180] in other systems, the timescale at which quasicrystalline structures can rearrange themselves can vastly exceed the simulation time available. To determine thermodynamic stability, sometimes the free energy of an approximant crystal – i.e. a periodic crystal with a large unit cell featuring motifs identified in the quasicrystalline phase of interest – is computed.[181–183] Although the quasicrystal's configurational entropy is not captured,[184] it can be subsequently added using a random-tiling approximation.[185,186] Another possible approach is to use the thermodynamic integration formalism introduced above to compute the free energy of a particular quasicrystal configuration and account for the configurational entropy using an approximation of uncorrelated phason flips.[187,188] However, in simulations of systems with intermediately fast dynamics, a convenient approach would again be to use a direct-coexistence simulation with an explicit interface with the fluid phase, in a system that is sufficiently large that the free-energy cost of defects arising from the periodic box are negligible. Once one point of coexistence is known, the chemical potential of the quasicrystal must by construction be equal to that of the fluid phase, which can easily be determined using thermodynamic integration from an ideal gas. The chemical potential of the quasicrystal can then be integrated along isotherms or isobars to other conditions of interest, and the full phase diagram can be determined this way.[77] Such an approach is only feasible if simulations are fast enough to be tractable even with systems so large that the quasicrystal–fluid equilibrium is dominated by bulk terms, which should be explicitly checked. More broadly, a method can work very well for one system, whilst being inappropriate to use for another, highlighting the importance of having a range of tools at our disposal when investigating different systems.

III. MACHINE-LEARNING APPROACHES TO PREDICTING PHASE-DIAGRAMS

As an alternative to free-energy calculations, machine-learning methods have been gaining traction in recent years due to their potential for high-accuracy predictions with low computational cost. Machine-learning models can often make accurate predictions quickly once trained on sufficient data, and can moreover usually be updated or retrained relatively easily when more data become available. However, on the flip side, the process of developing and training such models can often be complex and time consuming. One first has to decide on the choice of model architecture, as well as the type and source of data to use to train the model. These data need to be either generated or extracted from repositories and then processed into a suitable and consistent format to be used as training inputs for the model. To this end, suitable descriptors that in some way quantify features relating to the properties of interest must be identified. Finally, the training process itself requires considerations such as hyperparameter tuning, cross-validation and regularisation techniques to prevent overfitting, as well as benchmarking of the model with other methods. The appropriate route to follow in the case of machine-learned methods is therefore often somewhat less clear than is the case with traditional methods.

Here, we provide a brief summary of some machine-learning methods that have recently been applied to predict phase diagrams.

A. Machine-learned potentials

One possible route that fruitfully exploits such techniques is to use machine-learned potentials (MLPs)[6–9,45–48,190–193] that can simulate systems with the accuracy of ab initio methods but at a fraction of the computational cost [Fig. 2(a)]. Such approaches permit the determination of phase diagrams at the level of the underlying electronic structure theory; indeed, despite some possible limitations of machine-learned approaches in quantifying longer-ranged interactions,[45,190] it has been shown that the phase diagrams computed with MLPs are about as different from the underlying DFT phase diagrams as different DFT functionals are to one another.[8] When coupled with MLPs to make the computations tractable,[45] thermodynamic phase behaviour could thus be used to obtain better approximate DFT functionals and to understand the nature of the approximations made in such functionals.

One advantage of such MLPs is that they provide an intuitive molecular-level description of the material in question, as well as likely mechanisms by which phase transformations can occur. However, the full machinery outlined so far in this article is required to determine the phase behaviour of such potentials.

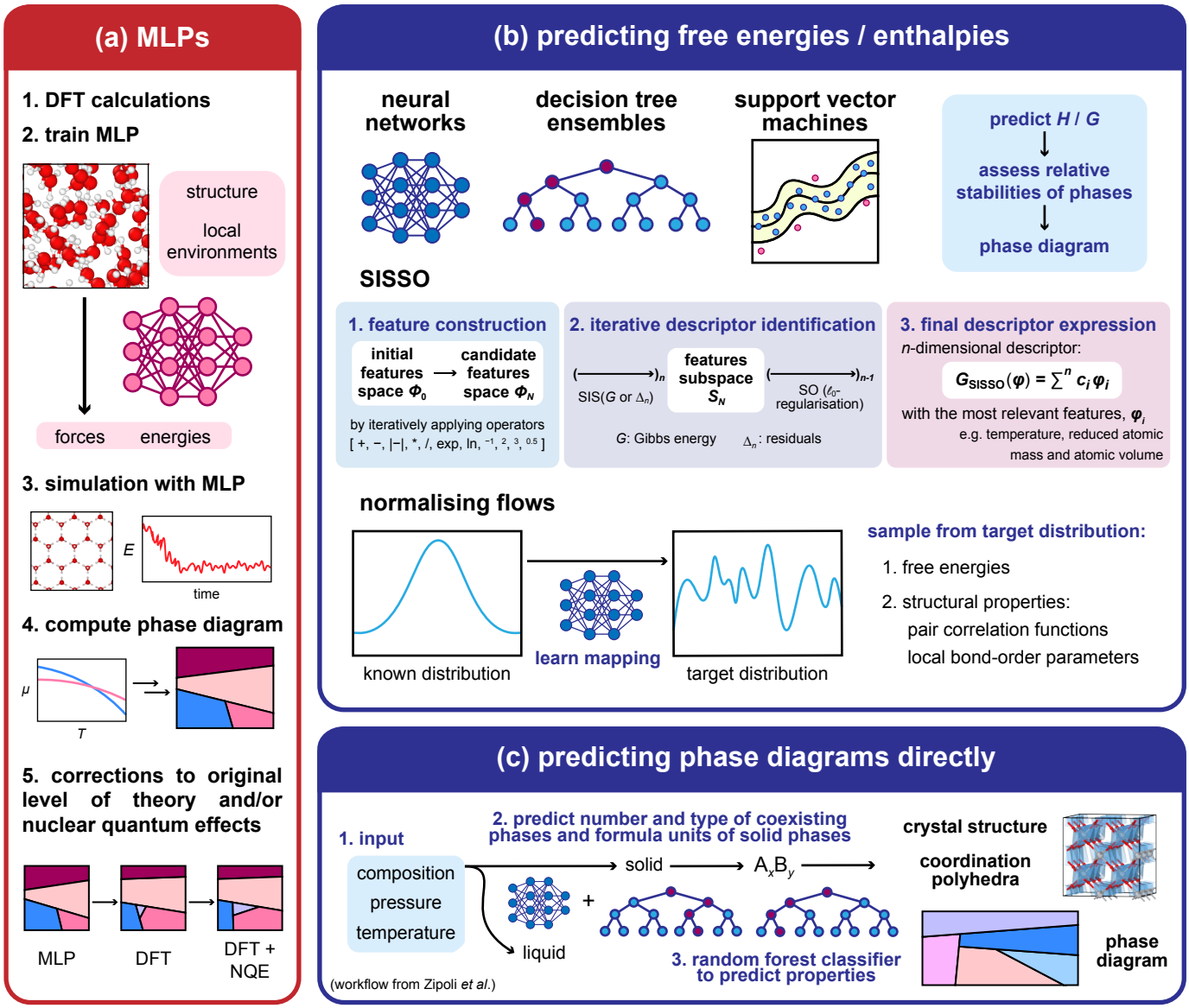

Figure 2. Summary of different machine-learning approaches to determining phase diagrams. (a) Machine-learned potentials (MLPs). (b) Predicting the enthalpies or free energies of different phases, which then allows one to assess their relative stabilities at different conditions (e.g. via a convex hull construction) to construct the phase diagram. This can be done via a variety of methods, ranging from more common machine-learning algorithms like neural networks and decision tree ensembles, to the sure-independence-screening sparsifying operator (SISSO) approach and normalising flow models. (c) Obtaining phase diagrams by directly predicting the number and types of coexisting phases at a given condition. The workflow presented here is from Zipoli and co-workers.[189]

B. Predicting enthalpies and free energies

Various machine-learning methods have also been used to predict phase diagrams and associated thermodynamic properties of materials directly, without requiring expensive molecular simulations. The success of such approaches depends heavily on the availability and quality of data of the relevant properties and existing phase diagrams for a number of materials to train the models. The amount of experimental and DFT data in various materials repositories that have become widely available in recent years has been instrumental in the development of these data-driven approaches using machine-learning methods. These include the Materials Project[194] or the Open Quantum Materials Database[195] for DFT-calculated quantities like the formation enthalpy, the Inorganic Crystal Structure Database,[196] the Cambridge Structural Database or Pearson's Crystal Data for crystal structure data, or the NIST 31 database for experimental and computational phase diagrams.

One approach to constructing phase diagrams is to predict the relevant thermodynamic quantities of different phases directly from their composition, and sometimes structure, and assess the relative stability of the phases using these quantities [Fig. 2(b)]. Such an approach has, for example, been used in materials discovery to identify if materials with certain elemental compositions are synthetically accessible by developing models to predict their formation enthalpies.[197] Models are trained on existing data of thermodynamic quantities obtained either experimentally or from DFT calculations, and then extrapolated to make predictions for new materials. The stability of a certain material relative to others with similar elemental compositions

can then be assessed with a convex hull construction.[198] Although differences between their relative entropies have often been approximated to be negligible when dealing with solid–solid transitions, models have also been trained to predict Gibbs energies.[63,199] For example, neural networks have frequently been used to predict formation enthalpies[197,200,201] and free energies[63] of materials. In general, neural networks are made up of interconnected nodes in a layered structure: the input layer first takes in and processes the input information before passing it on to the next layer, and the subsequent hidden layers analyse and process the output from the previous layer before it reaches the output layer, where the final 'result' is given. Connections between nodes are usually associated with a given weight and bias, and during the training process the neural network learns to adjust these parameters continuously via feedback loops to minimise the discrepancy between the predicted and actual quantity to improve predictions. Neural networks are thus able to learn complex relationships between the input data and the target quantity of interest, and they have the ability to make very accurate predictions, especially when trained on large amounts of data. Other models including decision tree ensembles[202,203] and support vector regressors[204–206] have also been used to predict thermodynamic quantities. Such machine-learning approaches have been shown to be able to predict the formation enthalpies of materials with an accuracy comparable to DFT; however, predicting relative stabilities of related compounds from their predicted formation enthalpies using convex hull constructions is less accurate, as the advantageous systematic error cancellation in DFT predictions does not apply to these models.[207] Additionally, the majority of such models only use compositional information and hence cannot make predictions for polymorphic transitions. Models that include both compositional and structural descriptors have been shown to produce more accurate stability predictions;[197,207] however, the structure is not always reported in the underlying data set.

One common issue with machine-learning models is their interpretability. For example, in neural network models, as the input data are passed through multiple hidden layers and processed, the predictions often become difficult to understand at a physical level and we are often not able to get a sense of what kinds of features make an important contribution to the accuracy of the predictions, and hence have physical significance.[197] In order to gain a better understanding of what properties are the most relevant for predicting thermodynamic quantities, we could make use of alternative methods which are explicitly able to select features to construct descriptors that result in the best predictions. An example is the sure-independence-screening sparsifying operator (SISSO)[208] method, which searches a space of mathematical expressions of selected features to find an optimal solution for an accurate descriptor of the quantity we want to predict. In this approach, to obtain a range of candidate descriptors, a combination of different mathematical operations is first recursively applied to an initial feature space consisting of properties that are potentially relevant for capturing the predicted quantity, while ensuring that only sensible combinations with physical units are permitted. From the set of candidate descriptors, a subspace of the best-performing features with the highest correlation with the target property can then be selected. Within the selected subspace, l_0-norm regularisation (or similar approximations) can be used to identify the best one-dimensional descriptor, which is then used to predict the target property and the associated residuals for the training set. This descriptor identification process can then be repeated to consider higher-dimensional solutions with each additional iteration, with the residuals from the previous iteration as the new target property. The goal is to identify the lowest-dimensional solution with acceptable errors below a desired threshold, and the features which are ultimately selected in the descriptor would be those that are the most important for capturing the property of interest.

The SISSO approach has been used by Bartel and co-workers to derive an accurate descriptor of the Gibbs energy of inorganic crystalline solids.[199] As a starting point for the initial feature space, they consider several quantities that are potentially relevant for predicting the Gibbs energy of a material. These include the atomic volume and band gap derived from the Materials Project database, experimental formation enthalpy and temperature, as well as tabulated elemental properties (ionisation energy, electron affinity, covalent radius, electronegativity and atomic mass), which were then transformed into compound-specific versions by finding suitable stoichiometrically weighted arithmetic and geometric means. Despite the larger set of properties being considered as input features for the model, the SISSO-learned descriptor depends only on temperature, reduced atomic mass and atomic volume, hence showing that these properties are the most significant in predicting the Gibbs energy. Their descriptor is benchmarked against Gibbs energies calculated via the quasiharmonic approximation (i.e. assuming harmonic normal modes [cf. Sec. II B 3], but with a changeable lattice constant) and then used to predict reaction energetics and equilibrium product distributions, as well as to assess temperature-dependent stability via a convex hull construction using the Gibbs formation energies of the materials. However, since this descriptor only predicts the free energies for solid phases, it is unable to determine the melting point of solid–liquid transitions directly. Transitions between different polymorphs are also beyond the scope of this predictor, since the only input feature that describes structure is the atomic volume, and errors from the descriptor are typically larger than the free-energy differences between polymorphs.[199]

In addition to training models to predict energies or free energies directly, machine-learning methods can also allow us to develop models to sample complex probability distributions, such as the Boltzmann distribution of a given system with some known potential energy function, from which we can sample to compute equilibrium energies and structures. Sampling from such distributions is traditionally done using methods like MC and MD, which can become expensive for complicated systems or when ergodicity is locally broken and rare events need to be simulated. Recently, alternative methods like flow-based sampling have been applied to various systems.[209–212] A normalising flow model[213] transforms an analytically tractable base distribution, such as a gaussian distribution, which we can easily sample from and evaluate probability densities for, into a more complex target distribution of the system of interest, through a series of invertible and differentiable functions. We can sample from the target distribution ρ_x by first sampling from the initial known base distribution ρ_z, and then applying this series of functions, i.e. taking $x = f(z)$. The probability density of such a transformed sample can be calculated using the product of the density of the original sample under the base distribution ρ_z and the associated change in volume from the sequence of transformations, which is the product of the

absolute values of the jacobian determinants for each inverse transformation, as given in the change of variable formula

$$\rho_{\mathrm{x}}(x) = \rho_{\mathrm{z}}(z) \left|\det J_f(z)\right|^{-1}, \tag{16}$$

where J_f is the jacobian matrix of f. In theory, as long as the two distributions are of the same dimension, any target distribution ρ_{x} can be obtained from any base distribution ρ_{z} given a sufficiently complex transformation function f. Such a function can comprise of a sequence of N invertible functions,

$$f = f_N \circ f_{N-1} \circ \cdots \circ f_1; \tag{17}$$

this function composition successively constructs more complex functions from the previous one, and the base distribution can be said to 'flow' through the series of functions.

Deep-learning algorithms can help us find a suitable composition of functions which can transform the base distribution into the target distribution. This approach has been used by Wirnsberger and co-workers to estimate the free energy and structural properties of atomic solids from a flow-based model.[210,214] Here, the flow-based model is trained to approximate the Boltzmann distribution of various atomic solids, starting from a base distribution of a lattice model with spherically truncated gaussian noise added to the atom at each lattice site, followed by a random permutation of all atoms. In this manner, we can target specific crystal structures (such as cubic and hexagonal ice) by using an appropriate choice of lattice for the base distribution to guide the model towards the phase of interest, without changing the potential energy function or using ground-truth samples. To transform the base distribution into the target distribution, each f_n in the series of functions transforms element-wise one or two co-ordinates of all the atoms as a function of the remaining atom co-ordinates, and each f_n is parameterised by a separate neural network whose parameters are trained by minimising the difference between the target and base distribution (as calculated by the Kullback–Leibler divergence[215] as the loss function). The trained flow-based model is able to reproduce structural properties and free-energy estimates with good accuracy: energy histograms, pair correlation functions and local bond-order parameters computed from sampling from the trained model for the truncated and shifted LJ face-centred cubic phase and for ice I using the mW water model were consistent with traditional methods, and free-energy estimates obtained using 'learned' free-energy perturbation methods with the trained model were also shown to be in good agreement with multistate-reweighting methods for both ice and LJ systems. Unlike traditional methods like thermodynamic integration or multistate reweighting, where samples from intermediate thermodynamic states are needed, this flow-based approach to obtain free energies does not require sampling from the target distribution for training the model or computing free-energy estimates. However, although once trained, generating samples and computing probability densities and various estimates from the trained model is efficient since samples can be obtained in parallel unlike in trajectory-based methods, one downside to this approach is that constructing and training the model is difficult and computationally expensive.[210]

C. Predicting phase diagrams directly

The methods discussed above involve first predicting thermodynamic quantities like free energies from the trained models, and the energies can then be used to assess the relative stabilities of the different phases considered to obtain a prediction for the phase diagram. An alternative approach may be to use machine-learning models to construct phase diagrams directly by training the models to predict the number and types of coexisting phases at a given condition, with or without predicting, estimating or computing free energies of the phases as an intermediate step.[63,189,216] This approach has recently been adopted by Zipoli and co-workers, who used machine-learning models to predict the polymorphs that are thermodynamically stable at certain conditions[189] [Fig. 2(c)]. For a certain condition defined by the overall chemical composition, temperature and pressure, we can first predict the types of phases (i.e. solid, liquid or vapour) that are present and the formula units of any solid phases using neural networks or random forest classifiers and regressors trained on data extracted from experimental and computational phase diagrams. Using the chemical formulae of the solid phases obtained from the first step, we can then use classifiers trained on crystal structure data of inorganic crystalline materials to predict structural properties of the solid phases in question (e.g. Bravais lattice types, crystal system, space group, structure type and local atomic environments characterised by the co-ordination polyhedra present). Such a data-driven approach with the workflow proposed by Zipoli and co-workers in principle has the potential to be able to build up phase diagrams directly and relatively quickly, using just the overall elemental composition, temperature and pressure to make predictions from the trained model, without any sort of feature construction required. However, the different components of the overall workflow may have varying levels of success: for example, even though structural properties of a particular solid phase can be predicted using classifiers from their chemical composition with good levels of accuracy,[189,217] obtaining the formula units of the different phases from the overall composition of the mixture when there are multiple phases present has proved to be much more difficult.[189]

Although machine-learning models for predicting phase behaviour and thermodynamic quantities have mostly focussed on inorganic crystalline solids, such approaches have also been used on soft materials, for example to predict the melting temperatures and the ternary phase diagrams of lipid mixtures, by predicting the types of coexisting phases from the mole fraction and properties of the lipids present.[218] Other recent approaches in predicting phase behaviour for biomolecules include sequence-based predictions of the propensity for proteins to phase separate.[219–222] In addition to the protein sequence, some models also take experimental conditions (protein concentration, salt concentration, temperature, pH and presence of crowding agents) into account when making predictions of phase-separation propensity;[223] however, such models are in general not able to predict quantitative measures such as the critical temperature or the physicochemical properties of the coexisting phases directly.

IV. FUTURE OUTLOOK

As evidenced by the large amount of work done in this field, the computational prediction of phase diagrams has already been very successful. Considerable effort has gone into developing a range of methods for studying phase coexistence, and even more to applications to a wide range of systems, from

inorganic materials to biological systems. A knowledge of the thermodynamic phase behaviour of a given computational model is useful both on the applied side, for example for understanding the underlying physics of a given system that may be difficult to study experimentally, and from the point of view of model development and refinement. The various methods outlined in this paper all have their advantages, but also some drawbacks.

For example, some machine-learning methods are able to produce predicted phase diagrams quickly, and so may be useful in guiding experiments towards promising phases with interesting properties. However, the 'best' approach in terms of the model architecture and training process is often problem-specific and might not be straightforward to decide on a priori. Indeed, often several models are first trained and the performance of the different models is then used to decide what kind of model architecture might work better for this one particular context considered. Despite the large amount of data available, the accuracy of machine-learning-based models is also limited by the quality and diversity of the data. It is often difficult to train such models in an unbiassed way, since it is not uncommon for data-sets to contain missing or inaccurate information, and unbalanced data sets with some underrepresented features might bias the model. Additionally, a large amount of data is needed to train a transferable model that can be applied to different system types whilst still achieving a high accuracy of predictions compared to more system-specific models, in which a smaller subset of components reduces the types of correlations that the model has to learn. Splitting a large data set into different subsets of similar materials could improve accuracy at the cost of generalisability. For soft materials in particular, one challenge in developing appropriate models is that fewer data are available compared to inorganic materials in general, and there is a wider variety of structures and behaviours that soft materials can exhibit, so approaches will need to be more tailored to specific contexts and system types. The effects of entropy may also be more significant for soft-matter systems, which may prove more challenging to capture with simple machine-learning approaches. Moreover, a data-driven approach only sees its input data – for example structures and energies from DFT calculations or from experiment –, but does not know anything about the underlying interactions between particles. Although predictions of behaviours under given conditions can often be surprisingly good, whether they are good for the right reasons is less clear: they are entirely phenomenological and even a significantly improved understanding or description of the physics of the building blocks would not directly help with the accuracy of the predictions of such models.

By contrast, the traditional methods of statistical and classical thermodynamics, such as thermodynamic integration, Gibbs–Duhem integration, histogram reweighting or Widom insertion, tend to be much slower because they require a number of well-equilibrated simulations to be run. The results obtained with such methods can be drastically improved with a better description of the building blocks, indeed perhaps even with a machine-learned interatomic potential, but the methods are often both computationally and labour intensive, and are not easy to automate for larger-scale applications. A combination of data-driven methods for an initial screening followed by traditional statistical-mechanical approaches to obtain accurate results may be the key to future applications of computational thermodynamics.

Although we have focussed primarily on predictions of phase diagrams in this article, several of the approaches we have outlined can provide details that go beyond mere thermodynamic stability. In particular, by determining the chemical potentials of potentially competitive phases, we can determine regions of metastability in addition to pure thermodynamic stability. This can be especially useful in understanding the likely experimental routes to synthesising metastable phases. For example, it may be more productive to try to synthesise a given metastable phase under conditions where the thermodynamically stable phase is only very marginally more stable.[8] Metastable phases can sometimes be relatively more stable in the initial stages of a phase transformation, even if the final bulk structure is a different one. An investigation of such metastable phases can thus help to reveal the possible mechanisms by which phases can transform into one another. For example, the metastable 'ice 0' structure was first proposed in the context of the mechanism of ice I nucleation.[224] We expect that, as calculations of chemical potentials become more routine, the importance of metastable phases will become clearer, and many more such pathways are likely to be identified.

Of course understanding the thermodynamics, stability and metastability is only the first step in understanding a phase transition. Studying the way phase transitions occur under different conditions is perhaps an even more challenging problem.[225] Broadly speaking, relatively close to coexistence, there are two main mechanisms by which phase transitions occur: heterogeneous nucleation, where an external nucleation seed is available, and homogeneous nucleation, where a spontaneous fluctuation has to overcome a free-energy barrier. Even though the phase diagram of substances like water close to atmospheric conditions is well known, a huge amount of work has gone into understanding the dynamics of ice I nucleation,[153,224,226–252] and there is doubtless more work to be done. Indeed, Oxtoby remarked back in 1998 that the study of nucleation is 'one of the few areas of science in which agreement of predicted and measured rates to within several orders of magnitude is considered a major success',[253] and more than two decades later, the same could be argued to hold. Nucleation is a rare event and this makes it difficult to study in computer simulations. Not only does it happen infrequently, but in order to track its progress, and potentially to drive it with rare-event methods, as with the interface pinning method, a suitable order parameter must first be identified to distinguish between the two phases in question, as discussed in Sec. II E. As systems become more complicated, distinguishing between phases becomes more difficult, and machine-learning approaches might again be helpful.[254–258] For example, Statt and co-workers have used unsupervised machine-learning techniques for local environments[259] to identify and classify different aggregate morphologies formed from model copolymer sequences[260] which the more conventional local order parameters failed to distinguish.[142] Techniques such as principal component analysis have also been used to find order parameters for detecting the freezing transition of hard spheres and ellipses, liquid–vapour phase separation of patchy particles and compositional demixing in the Widom–Rowlinson model.[261,262] Such order parameters will likely prove to be a useful tool in investigating nucleation mechanisms in more complex systems, including possible transitions to metastable phases discussed above.

In this article, we have largely considered simple systems made up of only one or two components. Even for such systems,

it is challenging to determine their phase behaviour well. The primary reason for this is still the quality of interatomic and intermolecular potentials, which are generally not applicable far beyond the range over which they were parameterised. We anticipate that ever-better interaction potentials will become available in the near future as machine-learned potentials are better parameterised.[263] However, there are still limitations in what can be simulated. For example, computer simulations often comprise relatively small systems, and it can be difficult to study defect formation. Schottky-style defects can be accounted for,[80] as can interstitial defects of neutral atoms;[264,265] however, interstitial defects involving charged species pose more of a challenge. Experimentally, it is possible to determine the energy of interstitial formation, but not usually the free energy: the effect of such defects on the phase behaviour is consequently very difficult to ascertain using anything other than computer simulations. Refining the methodology to facilitate the study of such defects would enable us to determine accurate high-pressure phase behaviour of technologically and geologically important materials even at high temperatures at which entropically driven defects are likely to play a significant role.

Much recent work has been done on extending the 'standard' free-energy methods to study both defects and interfaces.[266,267] Yeandel and co-workers have, for example, used the Einstein crystal framework to compute interfacial free-energy densities of solid–liquid mixtures, including systems where the solid contains miscible species that diffuse in the liquid.[267] In particular, they used the fact that Einstein crystal particles do not interact with one another and therefore the precise location of each individual molecule in either the bulk crystal phase or a crystalline slab in contact with the liquid is immaterial. This simplifies the calculation of interfacial properties considerably, and is another illustration of the power of thermodynamic integration and the construction of clever thermodynamic cycles, which should enable a progressively simpler calculation of interfacial properties at coexistence, and in particular the thermodynamics of systems in contact with non-pure solutions.[267]

Finally, there has been significant recent progress in simulating multicomponent systems.[268] Compositional phase diagrams involving fluid phases could be obtained using chemical potential measurements,[269,270] and such ideas may be important when investigating liquid–solid solutions. For example, deep eutectic solvents are thought to be promising for green-chemistry separation methods.[271–273] Computational modelling of such systems has often focussed on small systems that could be treated with density-functional theory, empirical equations of state or statistical associated fluid theory;[274,275] as discussed above, a determination of accurate phase diagrams for such systems could help in the development of transferable empirical potentials, allowing new insights to be gained into these technologically important systems from a molecular perspective. This challenging problem is only just beginning to be addressed.[276]

ACKNOWLEDGMENTS

This work was supported by the University of Cambridge Ernest Oppenheimer Fund and the Winton Programme for the Physics of Sustainability.

REFERENCES

[1] C. J. King, *Separation processes*, 2nd ed. (McGraw-Hill, 1980).
[2] C. G. Salzmann, P. G. Radaelli, B. Slater, and J. L. Finney, 'The polymorphism of ice: Five unresolved questions,' Phys. Chem. Chem. Phys. **13**, 18468–18480 (2011).
[3] C. G. Salzmann, 'Advances in the experimental exploration of water's phase diagram,' J. Chem. Phys. **150**, 060901 (2019).
[4] M. A. Neumann and J. van de Streek, 'How many ritonavir cases are there still out there?' Faraday Discuss. **211**, 441–458 (2018).
[5] W. G. Hoover and F. H. Ree, 'Melting transition and communal entropy for hard spheres,' J. Chem. Phys. **49**, 3609–3617 (1968).
[6] B. Cheng, G. Mazzola, C. J. Pickard, and M. Ceriotti, 'Evidence for supercritical behaviour of high-pressure liquid hydrogen,' Nature **585**, 217–220 (2020).
[7] B. Cheng, M. Bethkenhagen, C. J. Pickard, and S. Hamel, 'Phase behaviours of superionic water at planetary conditions,' Nat. Phys. **17**, 1228–1232 (2021).
[8] A. Reinhardt and B. Cheng, 'Quantum-mechanical exploration of the phase diagram of water,' Nat. Commun. **12**, 588 (2021).
[9] A. Reinhardt, M. Bethkenhagen, F. Coppari, M. Millot, S. Hamel, and B. Cheng, 'Thermodynamics of high-pressure ice phases explored with by atomistic simulations,' Nat. Commun. **13**, 4707 (2022).
[10] E. Florez, O. R. Smits, J.-M. Mewes, P. Jerabek, and P. Schwerdtfeger, 'From the gas phase to the solid state: The chemical bonding in the superheavy element flerovium,' J. Chem. Phys. **157**, 064304 (2022).
[11] D. Weidler and J. Gross, 'Transferable anisotropic united-atom force field based on the Mie potential for phase equilibria: Aldehydes, ketones, and small cyclic alkanes,' Ind. Eng. Chem. Res. **55**, 12123–12132 (2016).
[12] A. Reinhardt, 'Phase behavior of empirical potentials of titanium dioxide,' J. Chem. Phys. **151**, 064505 (2019).
[13] X. Wang, S. Ramírez-Hinestrosa, J. Dobnikar, and D. Frenkel, 'The Lennard-Jones potential: when (not) to use it,' Phys. Chem. Chem. Phys. **22**, 10624–10633 (2020).
[14] D. Atherton, A. Michaelides, and S. J. Cox, 'Can molecular simulations reliably compare homogeneous and heterogeneous ice nucleation?' J. Chem. Phys. **156**, 164501 (2022).
[15] J. L. F. Abascal and C. Vega, 'A general purpose model for the condensed phases of water: TIP4P/2005,' J. Chem. Phys. **123**, 234505 (2005).
[16] C. Vega, E. Sanz, and J. L. F. Abascal, 'The melting temperature of the most common models of water,' J. Chem. Phys. **122**, 114507 (2005).
[17] C. Vega, J. L. F. Abascal, E. Sanz, L. G. MacDowell, and C. McBride, 'Can simple models describe the phase diagram of water?' J. Phys.: Condens. Matter **17**, S3283 (2005).
[18] C. Vega, E. Sanz, J. L. F. Abascal, and E. G. Noya, 'Determination of phase diagrams via computer simulation: Methodology and applications to water, electrolytes and proteins,' J. Phys.: Condens. Matter **20**, 153101 (2008).
[19] C. Vega, J. L. F. Abascal, M. M. Conde, and J. L. Aragones, 'What ice can teach us about water interactions: A critical comparison of the performance of different water models,' Faraday Discuss. **141**, 251–276 (2009).
[20] C. Vega and J. L. F. Abascal, 'Simulating water with rigid non-polarizable models: A general perspective,' Phys. Chem. Chem. Phys. **13**, 19663–19688 (2011).
[21] P. T. Kiss and A. Baranyai, 'Sources of the deficiencies in the popular SPC/E and TIP3P models of water,' J. Chem. Phys. **134**, 054106 (2011).
[22] E. G. Noya, C. Menduiña, J. L. Aragones, and C. Vega, 'Equation of state, thermal expansion coefficient, and isothermal compressibility for ices I_h, II, III, V, and VI, as obtained from computer simulation,' J. Phys. Chem. C **111**, 15877–15888 (2007).
[23] H. L. Pi, J. L. Aragones, C. Vega, E. G. Noya, J. L. F. Abascal, M. A. Gonzalez, and C. McBride, 'Anomalies in water as obtained from computer simulations of the TIP4P/2005 model: Density maxima, and density, isothermal compressibility and heat capacity minima,' Mol. Phys. **107**, 365–374 (2009).
[24] C. McBride, E. G. Noya, J. L. Aragones, M. M. Conde, and C. Vega, 'The phase diagram of water from quantum simulations,' Phys. Chem. Chem. Phys. **14**, 10140–10146 (2012).
[25] L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande, 'Systematic improvement of a classical molecular model of water,' J. Phys. Chem. B **117**, 9956–9972 (2013).
[26] J. A. Joseph, A. Reinhardt, A. Aguirre, P. Y. Chew, K. O. Russell, J. R. Espinosa, A. Garaizar, and R. Collepardo-Guevara, 'Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative

accuracy,' Nat. Comput. Sci. **1**, 732–743 (2021).

[27] T. Dannenhoffer-Lafage and R. B. Best, 'A data-driven hydrophobicity scale for predicting liquid–liquid phase separation of proteins,' J. Phys. Chem. B **125**, 4046–4056 (2021).

[28] G. Tesei, T. K. Schulze, R. Crehuet, and K. Lindorff-Larsen, 'Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties,' Proc. Natl Acad. Sci. U.S.A. **118**, e2111696118 (2021).

[29] R. Iftimie, P. Minary, and M. E. Tuckerman, '*Ab initio* molecular dynamics: Concepts, recent developments, and future trends,' Proc. Natl. Acad. Sci. U. S. A. **102**, 6654–6659 (2005).

[30] F. Ercolessi and J. B. Adams, 'Interatomic potentials from first-principles calculations: The force-matching method,' EPL–Europhys. Lett. **26**, 583–588 (1994).

[31] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, 'Comparison of simple potential functions for simulating liquid water,' J. Chem. Phys. **79**, 926–935 (1983).

[32] D. Reith, M. Pütz, and F. Müller-Plathe, 'Deriving effective mesoscale potentials from atomistic simulations,' J. Comput. Chem. **24**, 1624–1636 (2003).

[33] F. Romano, J. Russo, and H. Tanaka, 'Novel stable crystalline phase for the Stillinger-Weber potential,' Phys. Rev. B **90**, 014204 (2014).

[34] L. Lu, J. F. Dama, and G. A. Voth, 'Fitting coarse-grained distribution functions through an iterative force-matching method,' J. Chem. Phys. **139**, 121906 (2013).

[35] T. D. Potter, J. Tasche, and M. R. Wilson, 'Assessing the transferability of common top-down and bottom-up coarse-grained molecular models for molecular mixtures,' Phys. Chem. Chem. Phys. **21**, 1912–1927 (2019).

[36] E. G. Noya, C. Vega, J. P. K. Doye, and A. A. Louis, 'Phase diagram of model anisotropic particles with octahedral symmetry,' J. Chem. Phys. **127**, 054501 (2007).

[37] F. Romano, E. Sanz, and F. Sciortino, 'Role of the range in the fluid–crystal coexistence for a patchy particle model,' J. Phys. Chem. B **113**, 15133 (2009).

[38] F. Romano, E. Sanz, and F. Sciortino, 'Phase diagram of a tetrahedral patchy particle model for different interaction ranges,' J. Chem. Phys. **132**, 184501 (2010).

[39] F. Romano, E. Sanz, and F. Sciortino, 'Crystallization of tetrahedral patchy particles *in silico*,' J. Chem. Phys. **134**, 174502 (2011).

[40] F. Romano and F. Sciortino, 'Two dimensional assembly of triblock Janus particles into crystal phases in the two bond per patch limit,' Soft Matter **7**, 5799–5804 (2011).

[41] A. Reinhardt, A. J. Williamson, J. P. K. Doye, J. Carrete, L. M. Varela, and A. A. Louis, 'Re-entrant phase behavior for systems with competition between phase separation and self-assembly,' J. Chem. Phys. **134**, 104905 (2011).

[42] G. Doppelbauer, E. G. Noya, E. Bianchi, and G. Kahl, 'Self-assembly scenarios of patchy colloidal particles,' Soft Matter **8**, 7768–7772 (2012).

[43] E. G. Noya, I. Kolovos, G. Doppelbauer, G. Kahl, and E. Bianchi, 'Phase diagram of inverse patchy colloids assembling into an equilibrium laminar phase,' Soft Matter **10**, 8464–8474 (2014).

[44] P. Teixeira and J. Tavares, 'Phase behaviour of pure and mixed patchy colloids — Theory and simulation,' Curr. Op. Colloid Interface Sci. **30**, 16–24 (2017).

[45] B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, 'Ab initio thermodynamics of liquid and solid water,' Proc. Natl Acad. Sci. U. S. A **116**, 1110–1115 (2019).

[46] L. Zhang, H. Wang, R. Car, and W. E, 'Phase diagram of a deep potential water model,' Phys. Rev. Lett. **126**, 236001 (2021).

[47] H. Niu, L. Bonati, P. M. Piaggi, and M. Parrinello, 'Ab initio phase diagram and nucleation of gallium,' Nat. Commun. **11**, 2654 (2020).

[48] J. G. Lee, C. J. Pickard, and B. Cheng, 'High-pressure phase behaviors of titanium dioxide revealed by a δ-learning potential,' J. Chem. Phys. **156**, 074106 (2022).

[49] A. Reinhardt, C. J. Pickard, and B. Cheng, 'Predicting the phase diagram of titanium dioxide with random search and pattern recognition,' Phys. Chem. Chem. Phys. **22**, 12697–12705 (2020).

[50] V. Swamy, J. D. Gale, and L. S. Dubrovinsky, 'Atomistic simulation of the crystal structures and bulk moduli of TiO_2 polymorphs,' J. Phys. Chem. Solids **62**, 887–895 (2001).

[51] J. Muscat, V. Swamy, and N. M. Harrison, 'First-principles calculations of the phase stability of TiO_2,' Phys. Rev. B **65**, 224112 (2002).

[52] X. G. Ma, P. Liang, L. Miao, S. W. Bie, C. K. Zhang, L. Xu, and J. J. Jiang, 'Pressure-induced phase transition and elastic properties of TiO_2 polymorphs,' Phys. Status Solidi B **246**, 2132–2139 (2009).

[53] H. Dekura, T. Tsuchiya, Y. Kuwayama, and J. Tsuchiya, 'Theoretical and experimental evidence for a new post-cotunnite phase of titanium dioxide with significant optical absorption,' Phys. Rev. Lett. **107**, 045701 (2011).

[54] V. Swamy and N. C. Wilson, 'First-principles calculations of the pressure stability and elasticity of dense TiO_2 phases using the B3LYP hybrid functional,' J. Phys. Chem. C **118**, 8617–8625 (2014).

[55] Q.-J. Liu, Z. Ran, F.-S. Liu, and Z.-T. Liu, 'Phase transitions and mechanical stability of TiO_2 polymorphs under high pressure,' J. Alloys Compd. **631**, 192–201 (2015).

[56] Y. Luo, A. Benali, L. Shulenburger, J. T. Krogel, O. Heinonen, and P. R. C. Kent, 'Phase stability of TiO_2 polymorphs from diffusion Quantum Monte Carlo,' New J. Phys. **18**, 113049 (2016).

[57] J. Trail, B. Monserrat, P. López Ríos, R. Maezono, and R. J. Needs, 'Quantum Monte Carlo study of the energetics of the rutile, anatase, brookite, and columbite TiO_2 polymorphs,' Phys. Rev. B **95**, 121108 (2017).

[58] Z. Fu, Y. Liang, S. Wang, and Z. Zhong, 'Structural phase transition and mechanical properties of TiO_2 under high pressure,' Phys. Status Solidi B **250**, 2206–2214 (2013).

[59] M. A. Blanco, E. Francisco, and V. Luaña, 'GIBBS: Isothermal-isobaric thermodynamics of solids from energy curves using a quasi-harmonic Debye model,' Comput. Phys. Commun. **158**, 57–72 (2004).

[60] Y. Jing-Xin, F. Min, J. Guang-Fu, and C. Xiang-Rong, 'Phase transition and thermodynamic properties of TiO_2 from first-principles calculations,' Chinese Phys. B **18**, 269 (2009).

[61] Y.-F. Hu, G. Jiang, D.-Q. Meng, and F.-J. Kong, 'Phase transition and thermodynamic properties of TiO_2,' Acta Phys.-Chim. Sin. **26**, 1664–1668 (2010).

[62] Z.-G. Mei, Y. Wang, S. Shang, and Z.-K. Liu, 'First-principles study of the mechanical properties and phase stability of TiO_2,' Comput. Mater. Sci. **83**, 114–119 (2014).

[63] S. Srinivasan, R. Batra, D. Luo, T. Loeffler, S. Manna, H. Chan, L. Yang, W. Yang, J. Wen, P. Darancet, and S. K. Sankaranarayanan, 'Machine learning the metastable phase diagram of covalently bonded carbon,' Nat. Commun. **13**, 3251 (2022).

[64] D. J. Wales and H. A. Scheraga, 'Global optimization of clusters, crystals, and biomolecules,' Science **285**, 1368–1372 (1999).

[65] S. M. Woodley and R. Catlow, 'Crystal structure prediction from first principles,' Nat. Mater. **7**, 937–946 (2008).

[66] L. Filion, M. Marechal, B. van Oorschot, D. Pelt, F. Smallenburg, and M. Dijkstra, 'Efficient method for predicting crystal structures at finite temperature: Variable box shape simulations,' Phys. Rev. Lett. **103**, 188302 (2009).

[67] Y. Wang, J. Lv, L. Zhu, and Y. Ma, 'Crystal structure prediction via particle-swarm optimization,' Phys. Rev. B **82**, 094116 (2010).

[68] C. J. Pickard and R. J. Needs, '*Ab initio* random structure searching,' J. Phys.: Condens. Matter **23**, 053201 (2011).

[69] H. T. Stokes and D. M. Hatch, '*FINDSYM*: program for identifying the space-group symmetry of a crystal,' J. Appl. Cryst. **38**, 237–238 (2005).

[70] V. Stevanović, 'Sampling polymorphs of ionic solids using random superlattices,' Phys. Rev. Lett. **116**, 075503 (2016).

[71] B. Widom, 'Some topics in the theory of fluids,' J. Chem. Phys. **39**, 2808–2812 (1963).

[72] A. Z. Panagiotopoulos, 'Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble,' Mol. Phys. **61**, 813–826 (1987).

[73] A. Z. Panagiotopoulos, N. Quirke, M. Stapleton, and D. J. Tildesley, 'Phase equilibria by simulation in the Gibbs ensemble – Alternative derivation, generalization and application to mixture and membrane equilibria,' Mol. Phys. **63**, 527–545 (1988).

[74] N. B. Wilding and A. D. Bruce, 'Freezing by monte carlo phase switch,' Phys. Rev. Lett. **85**, 5138–5141 (2000).

[75] N. B. Wilding, 'Phase switch Monte Carlo,' AIP Conf. Proc. **690**, 349–355 (2003).

[76] F. Romano, E. Sanz, P. Tartaglia, and F. Sciortino, 'Phase diagram of trivalent and pentavalent patchy particles,' J. Phys.: Condens. Matter **24**, 064113 (2012).

[77] A. Reinhardt, F. Romano, and J. P. K. Doye, 'Computing phase diagrams for a quasicrystal-forming patchy-particle system,' Phys. Rev. Lett. **110**, 255503 (2013).

[78] G. Grochola, 'Constrained fluid λ-integration: Constructing a reversible thermodynamic path between the solid and liquid state,' J. Chem. Phys. **120**, 2122–2126 (2004).

[79] D. Frenkel and A. J. C. Ladd, 'New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres,' J. Chem. Phys. **81**, 3188–3193 (1984).

[80] B. Cheng and M. Ceriotti, 'Computing the absolute Gibbs free energy in atomistic simulations: Applications to defects in solids,' Phys. Rev. B **97**, 054102 (2018).

[81] J. L. Aragones, C. Valeriani, and C. Vega, 'Note: Free energy calculations for atomic solids through the Einstein crystal/molecule methodology using GROMACS and LAMMPS,' J. Chem. Phys. **137**, 146101 (2012).

[82] J. L. Aragones, E. G. Noya, C. Valeriani, and C. Vega, 'Free energy calculations for molecular solids using GROMACS,' J. Chem. Phys. **139**, 034104 (2013).

[83] J. G. Kirkwood, 'Statistical mechanics of fluid mixtures,' J. Chem. Phys. **3**, 300–313 (1935).

[84] D. Frenkel and B. M. Mulder, 'The hard ellipsoid-of-revolution fluid I. Monte Carlo simulations,' Mol. Phys. **55**, 1171–1192 (1985).

[85] C. Vega and E. G. Noya, 'Revisiting the Frenkel-Ladd method to compute the free energy of solids: The Einstein molecule approach,' J. Chem. Phys. **127**, 154113 (2007).

[86] J. F. Cornwell, *Group theory in physics, Volume 1*, Techniques of physics 7 (Academic Press, London, 1986).

[87] M. Hestenes and E. Stiefel, 'Methods of conjugate gradients for solving linear systems,' J. Res. Natl Bur. Stand. (U.S.) **49**, 409 (1952).

[88] J. Guénolé, W. G. Nöhring, A. Vaid, F. Houllé, Z. Xie, A. Prakash, and E. Bitzek, 'Assessment and optimization of the fast inertial relaxation engine (FIRE) for energy minimization in atomistic simulations and its implementation in LAMMPS,' Comput. Mater. Sci. **175**, 109584 (2020).

[89] G. T. Gao, X. C. Zeng, and H. Tanaka, 'The melting temperature of proton-disordered hexagonal ice: A computer simulation of 4-site transferable intermolecular potential model of water,' J. Chem. Phys. **112**, 8534–8538 (2000).

[90] S. Habershon and D. E. Manolopoulos, 'Free energy calculations for a flexible water model,' Phys. Chem. Chem. Phys. **13**, 19714–19727 (2011).

[91] R. Ramírez and C. P. Herrero, 'Quantum path integral simulation of isotope effects in the melting temperature of ice Ih,' J. Chem. Phys. **133**, 144511 (2010).

[92] M. Ceriotti and T. E. Markland, 'Efficient methods and practical guidelines for simulating isotope effects,' J. Chem. Phys. **138**, 014112 (2013).

[93] M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland, 'Nuclear quantum effects in water and aqueous systems: Experiment, theory, and current challenges,' Chem. Rev. **116**, 7529–7550 (2016).

[94] B. Cheng, A. T. Paxton, and M. Ceriotti, 'Hydrogen diffusion and trapping in α-iron: The role of quantum and anharmonic fluctuations,' Phys. Rev. Lett. **120**, 225901 (2018).

[95] J. Lee, 'New Monte Carlo algorithm: Entropic sampling,' Phys. Rev. Lett. **71**, 211–214 (1993).

[96] F. Wang and D. P. Landau, 'Efficient, multiple-range random walk algorithm to calculate the density of states,' Phys. Rev. Lett. **86**, 2050–2053 (2001).

[97] F. Wang and D. P. Landau, 'Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram,' Phys. Rev. E **64**, 056101 (2001).

[98] G. Ganzenmüller and P. J. Camp, 'Applications of Wang–Landau sampling to determining phase equilibria in complex fluids,' J. Chem. Phys. **127**, 154504 (2007).

[99] A. J. Williamson, *Methods, rules and limits of successful self-assembly*, Ph.D. thesis, University of Oxford, Oxford (2011).

[100] V. Egorov and B. Kryzhanovsky, 'Influence of initial guess on the convergence rate and the accuracy of Wang–Landau algorithm,' Opt. Mem. Neural Netw. **30**, 284–290 (2021).

[101] P. Poulain, F. Calvo, R. Antoine, M. Broyer, and P. Dugourd, 'Performances of Wang-Landau algorithms for continuous systems,' Phys. Rev. E **73**, 056704 (2006).

[102] S. Sinha and S. K. Roy, 'Performance of wang–landau algorithm in continuous spin models and a case study: Modified XY-model,' Phys. Lett. A **373**, 308–314 (2009).

[103] B. K. Latha and V. S. S. Sastry, 'Phase diagram of a general biaxial nematic model based on density of states computation,' Liq. Cryst. **45**, 2197–2213 (2018).

[104] M. Borg, C. Stampfl, A. Mikkelsen, J. Gustafson, E. Lundgren, M. Scheffler, and J. N. Andersen, 'Density of configurational states from first-principles calculations: The phase diagram of Al–Na surface alloys,' ChemPhysChem **6**, 1923–1928 (2005).

[105] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, 'Generalization of the Wang-Landau method for off-lattice simulations,' Phys. Rev. E **66**, 056703 (2002).

[106] C. Desgranges and J. Delhommelle, 'Phase equilibria of molecular fluids via hybrid Monte Carlo Wang–Landau simulations: Applications to benzene and *n*-alkanes,' J. Chem. Phys. **130**, 244109 (2009).

[107] S. Boothroyd, A. Kerridge, A. Broo, D. Buttar, and J. Anwar, 'Solubility prediction from first principles: a density of states approach,' Phys. Chem. Chem. Phys. **20**, 20981–20987 (2018).

[108] H. Do, J. D. Hirst, and R. J. Wheatley, 'Rapid calculation of partition functions and free energies of fluids,' J. Chem. Phys. **135**, 174105 (2011).

[109] H. Do and R. J. Wheatley, 'Density of states partitioning method for calculating the free energy of solids,' J. Chem. Theory Comput. **9**, 165–171 (2013).

[110] H. Do and R. J. Wheatley, 'Reverse energy partitioning—an efficient algorithm for computing the density of states, partition functions, and free energy of solids,' J. Chem. Phys. **145**, 084116 (2016).

[111] G. Torrie and J. Valleau, 'Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling,' J. Comput. Phys. **23**, 187–199 (1977).

[112] J. Skilling, 'Nested sampling,' AIP Conf. Proc. **735**, 395–405 (2004).

[113] L. B. Pártay, A. P. Bartók, and G. Csányi, 'Efficient sampling of atomic configurational spaces,' J. Phys. Chem. B **114**, 10502–10512 (2010).

[114] L. B. Pártay, G. Csányi, and N. Bernstein, 'Nested sampling for materials,' Eur. Phys. J. B **94**, 159 (2021).

[115] S. Martiniani, J. D. Stevenson, D. J. Wales, and D. Frenkel, 'Superposition enhanced nested sampling,' Phys. Rev. X **4**, 031034 (2014).

[116] R. J. N. Baldock, L. B. Pártay, A. P. Bartók, M. C. Payne, and G. Csányi, 'Determining pressure-temperature phase diagrams of materials,' Phys. Rev. B **93**, 174108 (2016).

[117] A. C. L. Opitz, 'Molecular dynamics investigation of a free surface of liquid argon,' Phys. Lett. A **47**, 439–440 (1974).

[118] A. J. C. Ladd and L. V. Woodcock, 'Triple-point coexistence properties of the Lennard-Jones system,' Chem. Phys. Lett. **51**, 155–159 (1977).

[119] J. R. Morris and X. Song, 'The melting lines of model systems calculated from coexistence simulations,' J. Chem. Phys. **116**, 9352–9358 (2002).

[120] R. L. Davidchack and B. B. Laird, 'Simulation of the hard-sphere crystal–melt interface,' J. Chem. Phys. **108**, 9452–9462 (1998).

[121] E. G. Noya, C. Vega, and E. de Miguel, 'Determination of the melting point of hard spheres from direct coexistence simulation methods,' J. Chem. Phys. **128**, 154507 (2008).

[122] J. R. Espinosa, E. Sanz, C. Valeriani, and C. Vega, 'On fluid-solid direct coexistence simulations: The pseudo-hard sphere model,' J. Chem. Phys. **139**, 144502 (2013).

[123] B. J. Jesson and P. A. Madden, 'Structure and dynamics at the aluminum solid–liquid interface: An *ab initio* simulation,' J. Chem. Phys. **113**, 5935–5946 (2000).

[124] A. B. Belonoshko, R. Ahuja, and B. Johansson, 'Quasi-*ab initio* molecular dynamic study of Fe melting,' Phys. Rev. Lett. **84**, 3638–3641 (2000).

[125] R. García Fernández, J. L. F. Abascal, and C. Vega, 'The melting point of ice I_h for common water models calculated from direct coexistence of the solid-liquid interface,' J. Chem. Phys. **124**, 144506 (2006).

[126] M. M. Conde, M. A. Gonzalez, J. L. F. Abascal, and C. Vega, 'Determining the phase diagram of water from direct coexistence simulations: The phase diagram of the TIP4P/2005 model revisited,' J. Chem. Phys. **139**, 154505 (2013).

[127] M. Fitzner, L. Joly, M. Ma, G. C. Sosso, A. Zen, and A. Michaelides, 'Communication: Truncated non-bonded potentials can yield unphysical behavior in molecular dynamics simulations of interfaces,' J. Chem. Phys. **147**, 121102 (2017).

[128] U. Tartaglino and E. Tosatti, 'Strain effects at solid surfaces near the melting point,' Surf. Sci. **532–535**, 623–627 (2003).

[129] J. Q. Broughton and G. H. Gilmer, 'Molecular dynamics of the crystal–fluid interface. V. Structure and dynamics of crystal–melt systems,' J. Chem. Phys. **84**, 5749–5758 (1986).

[130] J. Wang, S. Yoo, J. Bai, J. R. Morris, and X. C. Zeng, 'Melting temperature of ice I_h calculated from coexisting solid-liquid phases,' J. Chem. Phys. **123**, 036101 (2005).

[131] A. S. Holehouse and R. V. Pappu, 'Functional implications of intracellular phase transitions,' Biochemistry **57**, 2415–2423 (2018).

[132] S. Alberti and D. Dormann, 'Liquid–liquid phase separation in disease,' Annu. Rev. Genet. **53**, 171–194 (2019).

[133] S. M. Lichtinger, A. Garaizar, R. Collepardo-Guevara, and A. Reinhardt, 'Targeted modulation of protein liquid–liquid phase separation by evolution of amino-acid sequence,' PLOS Comput. Biol. **17**, e1009328 (2021).

[134] I. Sanchez-Burgos, J. R. Espinosa, J. A. Joseph, and R. Collepardo-Guevara, 'Valency and binding affinity variations can regulate the multilayered organization of protein condensates with many components,' Biomolecules **11**, 278 (2021).

[135] P. Y. Chew, J. A. Joseph, R. Collepardo-Guevara, and A. Reinhardt, 'Design-

ing multiphase biomolecular condensates by coevolution of protein mixtures,' bioRxiv (2022), 10.1101/2022.04.22.489187.
[136] A. Bremer, M. Farag, W. M. Borcherds, I. Peran, E. W. Martin, R. V. Pappu, and T. Mittag, 'Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains,' Nat. Chem. **14**, 196–207 (2021).
[137] G. Tesei and K. Lindorff-Larsen, 'Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range,' Open Research Europe **2** (2022), 10.12688/openreseurope.14967.1.
[138] G. Krainer, T. J. Welsh, J. A. Joseph, J. R. Espinosa, S. Wittmann, E. de Csilléry, A. Sridhar, Z. Toprakcioglu, G. Gudiškytė, M. A. Czekalska, W. E. Arter, J. Guillén-Boixet, T. M. Franzmann, S. Qamar, P. S. George-Hyslop, A. A. Hyman, R. Collepardo-Guevara, S. Alberti, and T. P. Knowles, 'Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions,' Nat. Commun. **12**, 1085 (2021).
[139] J. R. Espinosa, J. A. Joseph, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, 'Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components,' Proc. Natl. Acad. Sci. U. S. A. **117**, 13238–13247 (2020).
[140] E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, and T. Mittag, 'Valence and patterning of aromatic residues determine the phase behavior of prion-like domains,' Science **367**, 694–699 (2020).
[141] Y. Zhang, B. Xu, B. G. Weiner, Y. Meir, and N. S. Wingreen, 'Decoding the physical principles of two-component biomolecular phase separation,' eLife **10**, e62403 (2021).
[142] A. Statt, H. Casademunt, C. P. Brangwynne, and A. Z. Panagiotopoulos, 'Model for disordered proteins with strongly sequence-dependent liquid phase behavior,' J. Chem. Phys. **152**, 75101 (2020).
[143] Y.-H. Lin, J. P. Brady, J. D. Forman-Kay, and H. S. Chan, 'Charge pattern matching as a 'fuzzy' mode of molecular recognition for the functional phase separations of intrinsically disordered proteins,' New J. Phys. **19**, 115003 (2017).
[144] U. R. Pedersen, F. Hummel, G. Kresse, G. Kahl, and C. Dellago, 'Computing Gibbs free energy differences by interface pinning,' Phys. Rev. B **88**, 94101 (2013).
[145] U. R. Pedersen, 'Direct calculation of the solid–liquid Gibbs free energy difference in a single equilibrium simulation,' J. Chem. Phys. **139**, 104102 (2013).
[146] U. R. Pedersen, F. Hummel, and C. Dellago, 'Computing the crystal growth rate by the interface pinning method,' J. Chem. Phys. **142**, 44104 (2015).
[147] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, 'Bond-orientational order in liquids and glasses,' Phys. Rev. B **28**, 784–805 (1983).
[148] P.-R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, 'Simulation of homogeneous crystal nucleation close to coexistence,' Faraday Discuss. **104**, 93–110 (1996).
[149] B. Senger, P. Schaaf, D. S. Corti, R. Bowles, J.-C. Voegel, and H. Reiss, 'A molecular theory of the homogeneous nucleation rate. I. Formulation and fundamental numbers,' J. Chem. Phys. **110**, 6421–6437 (1999).
[150] W. Lechner and C. Dellago, 'Accurate determination of crystal structures based on averaged local bond order parameters,' J. Chem. Phys. **129**, 114707 (2008).
[151] S. Jungblut and C. Dellago, 'Crystallization of a binary Lennard-Jones mixture,' J. Chem. Phys. **134**, 104501 (2011).
[152] E. B. Moore, E. de la Llave, K. Welke, D. A. Scherlis, and V. Molinero, 'Freezing, melting and structure of ice in a hydrophilic nanopore,' Phys. Chem. Chem. Phys. **12**, 4124 (2010).
[153] A. Reinhardt, J. P. K. Doye, E. G. Noya, and C. Vega, 'Local order parameters for use in driving homogeneous ice nucleation with all-atom models of water,' J. Chem. Phys. **137**, 194504 (2012).
[154] G. B. Arfken and H. J. Weber, *Mathematical methods for physicists*, 6th ed. (Elsevier Academic, London, 2005).
[155] M. A. Blanco, M. Flórez, and M. Bermejo, 'Evaluation of the rotation matrices in the basis of real spherical harmonics,' J. Mol. Struc.–Theochem. **419**, 19–27 (1997).
[156] P.-L. Chau and A. J. Hardwick, 'A new order parameter for tetrahedral configurations,' Mol. Phys. **93**, 511–518 (1998).
[157] J. R. Errington and P. G. Debenedetti, 'Relationship between structural order and the anomalies of liquid water,' Nature **409**, 318–321 (2001).
[158] E. E. Santiso and B. L. Trout, 'A general set of order parameters for molecular crystals,' J. Chem. Phys. **134**, 064109 (2011).
[159] D. A. Kofke, 'Gibbs–Duhem integration: A new method for direct evaluation of phase coexistence by molecular simulation,' Mol. Phys. **78**, 1331–1336 (1993).
[160] D. A. Kofke, 'Direct evaluation of phase coexistence by molecular simulation via integration along the saturation line,' J. Chem. Phys. **98**, 4149–4162 (1993).
[161] M. R. Shirts and J. D. Chodera, 'Statistically optimal analysis of samples from multiple equilibrium states,' J. Chem. Phys. **129**, 124105 (2008).
[162] N. P. Schieber, E. C. Dybeck, and M. R. Shirts, 'Using reweighting and free energy surface interpolation to predict solid-solid phase diagrams,' J. Chem. Phys. **148**, 144104 (2018).
[163] G. Bauer and J. Gross, 'Phase equilibria of solid and fluid phases from molecular dynamics simulations with equilibrium and nonequilibrium free energy methods,' J. Chem. Theory Comput. **15**, 3778–3792 (2019).
[164] L. Pauling, 'The structure and entropy of ice and of other crystals with some randomness of atomic arrangement,' J. Am. Chem. Soc. **57**, 2680–2684 (1935).
[165] J. D. Bernal and R. H. Fowler, 'A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions,' J. Chem. Phys. **1**, 515–548 (1933).
[166] J. A. Hayward and J. R. Reimers, 'Unit cells for the simulation of hexagonal ice,' J. Chem. Phys. **106**, 1518–1529 (1997).
[167] V. Buch, P. Sandler, and J. Sadlej, 'Simulations of H_2O solid, liquid, and clusters, with an emphasis on ferroelectric ordering transition in hexagonal ice,' J. Phys. Chem. B **102**, 8641–8653 (1998).
[168] C. P. Herrero and R. Ramírez, 'Configurational entropy of hydrogen-disordered ice polymorphs,' J. Chem. Phys. **140**, 234502 (2014).
[169] L. G. MacDowell, E. Sanz, C. Vega, and J. L. F. Abascal, 'Combinatorial entropy and phase diagram of partially ordered ice phases,' J. Chem. Phys. **121**, 10145–10158 (2004).
[170] M. Matsumoto, T. Yagasaki, and H. Tanaka, 'GenIce: Hydrogen-disordered ice generator,' J. Comput. Chem. **39**, 61–64 (2017).
[171] J. L. Aragones, L. G. MacDowell, and C. Vega, 'Dielectric constant of ices and water: A lesson about water interactions,' J. Phys. Chem. A **115**, 5745–5758 (2011).
[172] W. Steurer, 'Twenty years of structure research on quasicrystals. Part I. Pentagonal, octagonal, decagonal and dodecagonal quasicrystals,' Z. Kristallogr. **219**, 391–446 (2004).
[173] M. Zu, P. Tan, and N. Xu, 'Forming quasicrystals by monodisperse soft core particles,' Nat. Commun. **8**, 2089 (2017).
[174] C. R. Iacovella, A. S. Keys, and S. C. Glotzer, 'Self-assembly of soft-matter quasicrystals and their approximants,' Proc. Natl. Acad. Sci. U. S. A. **108**, 20935–20940 (2011).
[175] P. F. Damasceno, S. C. Glotzer, and M. Engel, 'Non-close-packed three-dimensional quasicrystals,' J. Phys.: Condens. Matter **29**, 234005 (2017).
[176] M. N. van der Linden, J. P. K. Doye, and A. A. Louis, 'Formation of dodecagonal quasicrystals in two-dimensional systems of patchy particles,' J. Chem. Phys. **136**, 054904 (2012).
[177] D. F. Tracey, E. G. Noya, and J. P. K. Doye, 'Programming patchy particles to form three-dimensional dodecagonal quasicrystals,' J. Chem. Phys. **154**, 194505 (2021).
[178] A. Reinhardt, J. S. Schreck, F. Romano, and J. P. K. Doye, 'Self-assembly of two-dimensional binary quasicrystals: A possible route to a DNA quasicrystal,' J. Phys.: Condens. Matter **29**, 014006 (2017).
[179] L. Liu, Z. Li, Y. Li, and C. Mao, 'Rational design and self-assembly of two-dimensional, dodecagonal DNA quasicrystals,' J. Am. Chem. Soc. **141**, 4248–4251 (2019).
[180] G. Malescio and F. Sciortino, 'Self-assembly of quasicrystals and their approximants in fluids with bounded repulsive core and competing interactions,' J. Mol. Liq. **349**, 118209 (2022).
[181] A. Haji-Akbari, M. Engel, A. S. Keys, X. Zheng, R. G. Petschek, P. Palffy-Muhoray, and S. C. Glotzer, 'Disordered, quasicrystalline and crystalline phases of densely packed tetrahedra,' Nature **462**, 773–777 (2009).
[182] A. Haji-Akbari, M. Engel, and S. C. Glotzer, 'Degenerate quasicrystal of hard triangular bipyramids,' Phys. Rev. Lett. **107**, 215702 (2011).
[183] A. Haji-Akbari, M. Engel, and S. C. Glotzer, 'Phase diagram of hard tetrahedra,' J. Chem. Phys. **135**, 194101 (2011).
[184] K. Jiang and P. Zhang, 'Numerical methods for quasicrystals,' J. Comput. Phys. **256**, 428–440 (2014).
[185] M. Oxborrow and C. L. Henley, 'Random square-triangle tilings: A model for twelvefold-symmetric quasicrystals,' Phys. Rev. B **48**, 6966–6998 (1993).
[186] H. Pattabhiraman, A. P. Gantapara, and M. Dijkstra, 'On the stability of a quasicrystal and its crystalline approximant in a system of hard disks with a soft corona,' J. Chem. Phys. **143**, 164905 (2015).
[187] A. Kiselev, M. Engel, and H.-R. Trebin, 'Confirmation of the random tiling hypothesis for a decagonal quasicrystal,' Phys. Rev. Lett. **109**, 225502 (2012).
[188] M. Engel, 'Entropic stabilization of tunable planar modulated superstructures,' Phys. Rev. Lett. **106**, 095504 (2011).

[189] F. Zipoli, V. Viterbo, O. Schilter, L. Kahle, and T. Laino, 'Prediction of phase diagrams and associated phase structural properties,' Ind. Eng. Chem. Res **61**, 8378–8389 (2022).

[190] J. Behler, 'Constructing high-dimensional neural network potentials: A tutorial review,' Int. J. Quantum Chem. **115**, 1032–1050 (2015).

[191] J. Behler, 'Perspective: machine learning potentials for atomistic simulations,' J. Chem. Phys. **145**, 170901 (2016).

[192] J. Behler, 'First principles neural network potentials for reactive simulations of large molecular and condensed systems,' Angew. Chem., Int. Ed. **56**, 12828–12840 (2017).

[193] J. Behler, 'Four generations of high-dimensional neural network potentials,' Chem. Rev. **121**, 10037–10072 (2021).

[194] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, 'Commentary: The Materials Project: A materials genome approach to accelerating materials innovation,' APL Mater. **1**, 011002 (2013).

[195] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, 'The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies,' npj Comput. Mater. **1**, 15010 (2015).

[196] G. Bergerhoff, R. Hundt, R. Sievers, and I. D. Brown, 'The Inorganic Crystal Structure Data Base,' J. Chem. Inf. Comput. Sci. **23**, 66–69 (1983).

[197] G. G. C. Peterson and J. Brgoch, 'Materials discovery through machine learning formation energy,' J. Phys. Energy **3**, 022002 (2021).

[198] A. Anelli, E. A. Engel, C. J. Pickard, and M. Ceriotti, 'Generalized convex hull construction for materials discovery,' Phys. Rev. Materials **2**, 103804 (2018).

[199] C. J. Bartel, S. L. Millican, A. M. Deml, J. R. Rumptz, W. Tumas, A. W. Weimer, S. Lany, V. Stevanović, C. B. Musgrave, and A. M. Holder, 'Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry,' Nat. Commun. **9**, 4168 (2018).

[200] D. Jha, L. Ward, A. Paul, W.-k. Liao, A. Choudhary, C. Wolverton, and A. Agrawal, 'ElemNet: Deep learning the chemistry of materials from only elemental composition,' Sci. Rep. **8**, 17593 (2018).

[201] R. E. A. Goodall and A. A. Lee, 'Predicting materials properties without crystal structure: deep representation learning from stoichiometry,' Nat. Commun. **11**, 6280 (2020).

[202] L. Ward, A. Dunn, A. Faghaninia, N. E. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. A. Persson, G. J. Snyder, I. Foster, and A. Jain, 'Matminer: An open source toolkit for materials data mining,' Comput. Mater. Sci. **152**, 60–69 (2018).

[203] S. Gong, S. Wang, T. Xie, W. H. Chae, R. Liu, Y. Shao-Horn, and J. C. Grossman, 'Calibrating DFT formation enthalpy calculations by multifidelity machine learning,' JACS Au **2**, 1964–1977 (2022).

[204] S. Lotfi, Z. Zhang, G. Viswanathan, K. Fortenberry, A. Mansouri Tehrani, and J. Brgoch, 'Targeting productive composition space through machine-learning-directed inorganic synthesis,' Matter **3**, 261–272 (2020).

[205] K. K. Yalamanchi, M. Monge-Palacios, V. C. O. van Oudenhoven, X. Gao, and S. M. Sarathy, 'Data science approach to estimate enthalpy of formation of cyclic hydrocarbons,' J. Phys. Chem. A **124**, 6270–6276 (2020).

[206] S. Ubaru, A. Międlar, Y. Saad, and J. R. Chelikowsky, 'Formation enthalpies for transition metal alloys using machine learning,' Phys. Rev. B **95**, 214102 (2017).

[207] C. J. Bartel, A. Trewartha, Q. Wang, A. Dunn, A. Jain, and G. Ceder, 'A critical examination of compound stability predictions from machine-learned formation energies,' npj Comput. Mater. **6**, 97 (2020).

[208] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, and L. M. Ghiringhelli, 'SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates,' Phys. Rev. Materials **2**, 083802 (2018).

[209] F. Noé, S. Olsson, J. Köhler, and H. Wu, 'Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning,' Science **365**, eaaw1147 (2019).

[210] P. Wirnsberger, G. Papamakarios, B. Ibarz, S. Racanière, A. J. Ballard, A. Pritzel, and C. Blundell, 'Normalizing flows for atomic solids,' Mach. Learn.: Sci. Technol. **3**, 025009 (2022).

[211] M. S. Albergo, G. Kanwar, and P. E. Shanahan, 'Flow-based generative models for Markov chain Monte Carlo in lattice field theory,' Phys. Rev. D **100**, 034515 (2019).

[212] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, 'Equivariant flow-based sampling for lattice gauge theory,' Phys. Rev. Lett. **125**, 121601 (2020).

[213] E. G. Tabak and C. V. Turner, 'A family of nonparametric density estimation algorithms,' Comm. Pure Appl. Math. **66**, 145–164 (2013).

[214] P. Wirnsberger, A. J. Ballard, G. Papamakarios, S. Abercrombie, S. Racanière, A. Pritzel, D. Jimenez Rezende, and C. Blundell, 'Targeted free energy estimation via learned mappings,' J. Chem. Phys. **153**, 144112 (2020).

[215] S. Kullback and R. A. Leibler, 'On information and sufficiency,' Ann. Math. Stat. **22**, 79–86 (1951).

[216] G. Deffrennes, K. Terayama, T. Abe, and R. Tamura, 'A machine learning–based classification approach for phase diagram prediction,' Mater. Des. **215**, 110497 (2022).

[217] Y. Zhao, Y. Cui, Z. Xiong, J. Jin, Z. Liu, R. Dong, and J. Hu, 'Machine learning-based prediction of crystal systems and space groups from inorganic materials compositions,' ACS Omega **5**, 3596–3606 (2020).

[218] M. Aghaaminiha, S. A. Ghanadian, E. Ahmadi, and A. M. Farnoud, 'A machine learning approach to estimation of phase diagrams for three-component lipid mixtures,' Biochim. Biophys. Acta, Biomembr. **1862**, 183350 (2020).

[219] G. van Mierlo, J. R. Jansen, J. Wang, I. Poser, S. J. van Heeringen, and M. Vermeulen, 'Predicting protein condensate formation using machine learning,' Cell Rep. **34**, 108705 (2021).

[220] K. L. Saar, A. S. Morgunov, R. Qi, W. E. Arter, G. Krainer, A. A. Lee, and T. P. J. Knowles, 'Learning the molecular grammar of protein condensates from sequence determinants and embeddings,' Proc. Natl. Acad. Sci. U. S. A. **118**, e2019053118 (2021).

[221] X. Chu, T. Sun, Q. Li, Y. Xu, Z. Zhang, L. Lai, and J. Pei, 'Prediction of liquid–liquid phase separating proteins using machine learning,' BMC Bioinform. **23**, 72 (2022).

[222] M. Hardenberg, A. Horvath, V. Ambrus, M. Fuxreiter, and M. Vendruscolo, 'Widespread occurrence of the droplet state of proteins in the human proteome,' Proc. Natl. Acad. Sci. U. S. A. **117**, 33254–33262 (2020).

[223] D. Raimondi, G. Orlando, E. Michiels, D. Pakravan, A. Bratek-Skicki, L. Van Den Bosch, Y. Moreau, F. Rousseau, and J. Schymkowitz, 'In silico prediction of in vitro protein liquid–liquid phase separation experiments outcomes with multi-head neural attention,' Bioinformatics **37**, 3473–3479 (2021).

[224] J. Russo, F. Romano, and H. Tanaka, 'New metastable form of ice and its role in the homogeneous crystallization of water,' Nat. Mater. **13**, 733–739 (2014).

[225] K. E. Blow, D. Quigley, and G. C. Sosso, 'The seven deadly sins: When computing crystal nucleation rates, the devil is in the details,' J. Chem. Phys. **155**, 040901 (2021).

[226] L. Vrbka and P. Jungwirth, 'Homogeneous freezing of water starts in the subsurface,' J. Phys. Chem. B **110**, 18126–18129 (2006).

[227] D. Quigley and P. M. Rodger, 'Metadynamics simulations of ice nucleation and growth,' J. Chem. Phys. **128**, 154518 (2008).

[228] A. V. Brukhno, J. Anwar, R. Davidchack, and R. Handel, 'Challenges in molecular simulation of homogeneous ice nucleation,' J. Phys.: Condens. Matter **20**, 494243 (2008).

[229] E. Pluhařová, L. Vrbka, and P. Jungwirth, 'Effect of surface pollution on homogeneous ice nucleation: A molecular dynamics study,' J. Phys. Chem. C **114**, 7831–7838 (2010).

[230] P. Pirzadeh and P. G. Kusalik, 'On understanding stacking fault formation in ice,' J. Am. Chem. Soc. **133**, 704–707 (2011).

[231] T. Li, D. Donadio, G. Russo, and G. Galli, 'Homogeneous ice nucleation from supercooled water,' Phys. Chem. Chem. Phys. **13**, 19807 (2011).

[232] E. B. Moore and V. Molinero, 'Is it cubic? Ice crystallization from deeply supercooled water,' Phys. Chem. Chem. Phys. **13**, 20008–20016 (2011).

[233] T. L. Malkin, B. J. Murray, A. V. Brukhno, J. Anwar, and C. G. Salzmann, 'Structure of ice crystallized from supercooled water,' Proc. Natl. Acad. Sci. U. S. A. **109**, 1041–1045 (2012).

[234] S. J. Cox, S. M. Kathmann, J. A. Purton, M. J. Gillan, and A. Michaelides, 'Non-hexagonal ice at hexagonal surfaces: The role of lattice mismatch,' Phys. Chem. Chem. Phys. **14**, 7944–7949 (2012).

[235] B. J. Murray, D. O'Sullivan, J. D. Atkinson, and M. E. Webb, 'Ice nucleation by particles immersed in supercooled cloud droplets,' Chem. Soc. Rev. **41**, 6519–6554 (2012).

[236] A. Reinhardt and J. P. K. Doye, 'Free energy landscapes for homogeneous nucleation of ice for a monatomic water model,' J. Chem. Phys. **136**, 054501 (2012).

[237] G. Bullock and V. Molinero, 'Low-density liquid water is the mother of ice: On the relation between mesostructure, thermodynamics and ice crystallization in solutions,' Faraday Discuss. **167**, 371–388 (2013).

[238] A. Reinhardt and J. P. K. Doye, 'Note: Homogeneous TIP4P/2005 ice nucleation at low supercooling,' J. Chem. Phys. **139**, 096102 (2013).

[239] E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L. F. Abascal, and C. Valeriani, 'Homogeneous ice nucleation at moderate supercooling from molecular simulation,' J. Am. Chem. Soc. **135**, 15008–15017 (2013).

[240] A. Reinhardt and J. P. K. Doye, 'Effects of surface interactions on hetero-

geneous ice nucleation for a monatomic water model,' J. Chem. Phys. **141**, 084501 (2014).
[241] J. R. Espinosa, E. Sanz, C. Valeriani, and C. Vega, 'Homogeneous ice nucleation evaluated for several water models,' J. Chem. Phys. **141**, 18c529 (2014).
[242] L. Lupi, A. Hudait, and V. Molinero, 'Heterogeneous nucleation of ice on carbon surfaces,' J. Am. Chem. Soc. **136**, 3156–3164 (2014).
[243] L. Ickes, A. Welti, C. Hoose, and U. Lohmann, 'Classical nucleation theory of homogeneous freezing of water: thermodynamic and kinetic parameters,' Phys. Chem. Chem. Phys. **17**, 5514–5537 (2015).
[244] T. L. Malkin, B. J. Murray, C. G. Salzmann, V. Molinero, S. J. Pickering, and T. F. Whale, 'Stacking disorder in ice i,' Phys. Chem. Chem. Phys. **17**, 60–76 (2015).
[245] J. R. Espinosa, J. M. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. G. Debenedetti, and A. Z. Panagiotopoulos, 'On the calculation of solubilities via direct coexistence simulations: Investigation of nacl aqueous solutions and lennard-jones binary mixtures,' J. Chem. Phys. **145**, 154111 (2016).
[246] B. Cheng, C. Dellago, and M. Ceriotti, 'Theoretical prediction of the homogeneous ice nucleation rate: disentangling thermodynamics and kinetics,' Phys. Chem. Chem. Phys. **20**, 28732–28740 (2018).
[247] M. Fitzner, G. C. Sosso, S. J. Cox, and A. Michaelides, 'Ice is born in low-mobility regions of supercooled liquid water,' Proc. Natl. Acad. Sci. U. S. A. **116**, 2009–2014 (2019).
[248] S. Hussain and A. Haji-Akbari, 'Role of nanoscale interfacial proximity in contact freezing in water,' J. Am. Chem. Soc. **143**, 2272–2284 (2021).
[249] G. C. Sosso, P. Sudera, A. T. Backes, T. F. Whale, J. Fröhlich-Nowoisky, M. Bonn, A. Michaelides, and E. H. G. Backus, 'The role of structural order in heterogeneous ice nucleation,' Chem. Sci. **13**, 5014–5026 (2022).
[250] M. B. Davies, M. Fitzner, and A. Michaelides, 'Accurate prediction of ice nucleation from room temperature water,' Proc. Natl. Acad. Sci. U. S. A. **119**, e2205347119 (2022).
[251] F. Martelli and J. C. Palmer, 'Signatures of sluggish dynamics and local structural ordering during ice nucleation,' J. Chem. Phys. **156**, 114502 (2022).
[252] F. N. Isenrich, N. Shardt, M. Rösch, J. Nette, S. Stavrakis, C. Marcolli, Z. A. Kanji, A. J. deMello, and U. Lohmann, 'The Microfluidic Ice Nuclei Counter Zürich (MINCZ): a platform for homogeneous and heterogeneous ice nucleation,' Atmos. Meas. Tech. **15**, 5367–5381 (2022).
[253] D. W. Oxtoby, 'Homogeneous nucleation: theory and experiment,' Acc. Chem. Res. **10**, 897 (1998).
[254] P. Geiger and C. Dellago, 'Neural networks for local structure detection in polymorphic systems,' J. Chem. Phys. **139**, 164105 (2013).
[255] E. Boattini, M. Ram, F. Smallenburg, and L. Filion, 'Neural-network-based order parameters for classification of binary hard-sphere crystal structures,' Mol. Phys. **116**, 3066–3075 (2018).
[256] C. Dietz, T. Kretz, and M. H. Thoma, 'Machine-learning approach for local classification of crystalline structures in multiphase systems,' Phys. Rev. E **96**, 011301 (2017).
[257] R. S. DeFever, C. Targonski, S. W. Hall, M. C. Smith, and S. Sarupria, 'A generalized deep learning approach for local structure identification in molecular simulations,' Chem. Sci. **10**, 7503–7515 (2019).
[258] A. J. Mukhtyar and F. A. Escobedo, 'Computing free energy barriers for the nucleation of complex network mesophases,' J. Chem. Phys. **156**, 034502 (2022).
[259] W. F. Reinhart, 'Unsupervised learning of atomic environments from simple features,' Comput. Mater. Sci. **196**, 110511 (2021).
[260] A. Statt, D. C. Kleeblatt, and W. F. Reinhart, 'Unsupervised learning of sequence-specific aggregation behavior for a model copolymer,' Soft Matter **17**, 7697–7707 (2021).
[261] R. B. Jadrich, B. A. Lindquist, and T. M. Truskett, 'Unsupervised machine learning for detection of phase transitions in off-lattice systems. I. Foundations,' J. Chem. Phys. **149**, 194109 (2018).
[262] R. B. Jadrich, B. A. Lindquist, W. D. Piñeros, D. Banerjee, and T. M. Truskett, 'Unsupervised machine learning for detection of phase transitions in off-lattice systems. II. Applications,' J. Chem. Phys. **149**, 194110 (2018).
[263] B. Monserrat, J. G. Brandenburg, E. A. Engel, and B. Cheng, 'Liquid water contains the building blocks of diverse ice phases,' Nat. Commun. **11**, 5757 (2020).
[264] S. Chiesa, P. M. Derlet, and S. L. Dudarev, 'Free energy of a ⟨110⟩ dumbbell interstitial defect in bcc Fe: Harmonic and anharmonic contributions,' Phys. Rev. B **79**, 214109 (2009).
[265] J. Luo, C. Zhou, Q. Li, and L. Liu, 'Thermodynamic formation properties of point defects in germanium crystal,' Materials **15**, 4026 (2022).
[266] R. K. R. Addula and S. N. Punnathanam, 'Computation of solid–fluid interfacial free energy in molecular systems using thermodynamic integration,' J. Chem. Phys. **153**, 154504 (2020).
[267] S. R. Yeandel, C. L. Freeman, and J. H. Harding, 'A general method for calculating solid/liquid interfacial free energies from atomistic simulations: Application to $CaSO_4 \cdot xH_2O$,' J. Chem. Phys. **157**, 084117 (2022).
[268] R. S. DeFever and E. J. Maginn, 'Computing the liquidus of binary monatomic salt mixtures with direct simulation and alchemical free energy methods,' J. Phys. Chem. A **125**, 8498–8513 (2021).
[269] A. Rahbari, R. Hens, D. Dubbeldam, and T. J. H. Vlugt, 'Improving the accuracy of computing chemical potentials in CFCMC simulations,' Mol. Phys. **117**, 3493–3508 (2019).
[270] B. Cheng, 'Computing chemical potentials of solutions from structure factors,' J. Chem. Phys. **157**, 121101 (2022).
[271] A. P. Abbott, D. Boothby, G. Capper, D. L. Davies, and R. K. Rasheed, 'Deep eutectic solvents formed between choline chloride and carboxylic acids: versatile alternatives to ionic liquids,' J. Am. Chem. Soc. **126**, 9142–9147 (2004).
[272] N. Singh, S. Das, N. Singh, and T. Agrawal, 'Computer simulation, thermodynamic and microstructural studies of benzamide–benzoic acid eutectic system,' J. Cryst. Growth **310**, 2878–2884 (2008).
[273] A. González de Castilla, J. P. Bittner, S. Müller, S. Jakobtorweihen, and I. Smirnova, 'Thermodynamic and transport properties modeling of deep eutectic solvents: A review on gE-models, equations of state, and molecular dynamics,' J. Chem. Eng. Data **65**, 943–967 (2020).
[274] P. V. A. Pontes, E. A. Crespo, M. A. R. Martins, L. P. Silva, C. M. S. S. Neves, G. J. Maximo, M. D. Hubinger, E. A. C. Batista, S. P. Pinho, J. A. Coutinho, G. Sadowski, and C. Held, 'Measurement and PC-SAFT modeling of solid-liquid equilibrium of deep eutectic solvents of quaternary ammonium chlorides and carboxylic acids,' Fluid Phase Equilib. **448**, 69–80 (2017).
[275] M. A. R. Martins, E. A. Crespo, P. V. A. Pontes, L. P. Silva, M. Bülow, G. J. Maximo, E. A. C. Batista, C. Held, S. P. Pinho, and J. A. P. Coutinho, 'Tunable hydrophobic eutectic solvents based on terpenes and monocarboxylic acids,' ACS Sustainable Chem. Eng. **6**, 8836–8846 (2018).
[276] P. Vainikka, S. Thallmair, P. C. T. Souza, and S. J. Marrink, 'Martini 3 coarse-grained model for type III deep eutectic solvents: Thermodynamic, structural, and extraction properties,' ACS Sustainable Chem. Eng. **9**, 17338–17350 (2021).