NaCl nucleation from brine in seeded simulations: Sources of uncertainty in rate estimates

Nils. E. R. Zimmermann (iD), Bart Vorselaars (iD), Jorge R. Espinosa, David Quigley (iD), William R. Smith (iD), Eduardo Sanz, Carlos Vega (iD), and Baron Peters

COLLECTIONS

Paper published as part of the special topic on Ions in Water

ARTICLES YOU MAY BE INTERESTED IN

Forward flux sampling calculation of homogeneous nucleation rates from aqueous NaCl solutions
The Journal of Chemical Physics **148**, 044505 (2018); https://doi.org/10.1063/1.5016554

A simulation study of homogeneous ice nucleation in supercooled salty water
The Journal of Chemical Physics **148**, 222811 (2018); https://doi.org/10.1063/1.5008889

Communication: Nucleation rates of supersaturated aqueous NaCl using a polarizable force field
The Journal of Chemical Physics **149**, 141102 (2018); https://doi.org/10.1063/1.5053652
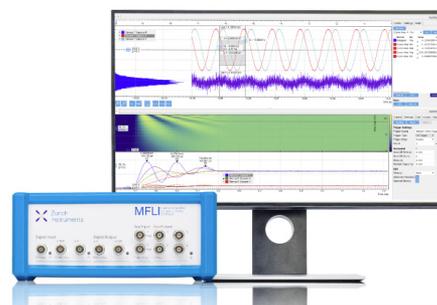
NaCl nucleation from brine in seeded simulations: Sources of uncertainty in rate estimates

Nils. E. R. Zimmermann,[1,a)] Bart Vorselaars,[2] Jorge R. Espinosa,[3] David Quigley,[4,a)] William R. Smith,[5,6,a)] Eduardo Sanz,[3] Carlos Vega,[3] and Baron Peters[7,8,a)]

[1]*Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[2]*School of Mathematics and Physics, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, United Kingdom*
[3]*Departmento de Quimica-Fisica I, Facultad de Ciencias Quimicas, Universidad Complutense de Madrid, 28040 Madrid, Spain*
[4]*Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom*
[5]*Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G2W1, Canada*
[6]*Department of Chemistry, University of Guelph, Guelph, Ontario N1G2W1, Canada*
[7]*Department of Chemical Engineering, University of California, Santa Barbara, California 93106, USA*
[8]*Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA*

This work reexamines seeded simulation results for NaCl nucleation from a supersaturated aqueous solution at 298.15 K and 1 bar pressure. We present a linear regression approach for analyzing seeded simulation data that provides both nucleation rates and uncertainty estimates. Our results show that rates obtained from seeded simulations rely critically on a precise driving force for the model system. The driving force vs. solute concentration curve need not exactly reproduce that of the real system, but it should accurately describe the thermodynamic properties *of the model system*. We also show that rate estimates depend strongly on the nucleus size metric. We show that the rate estimates systematically increase as more stringent local order parameters are used to count members of a cluster and provide tentative suggestions for appropriate clustering criteria. © *2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).* https://doi.org/10.1063/1.5024009

I. INTRODUCTION

Nucleation is important in many contexts, yet both theory and simulation have largely failed to quantitatively predict experimental nucleation rates.[1–5] Recently, some promising results[6,7] have been obtained using a new seeded simulation technique.[8,9] The seeding approach was developed to circumvent certain problems with other rare event methods[1] in studies of solute precipitate nucleation. It was not originally expected to provide highly accurate rates,[8] and as outlined above there are many cases where it *should not* provide accurate rates. The seeding approach infers interfacial free energies,[10] attachment frequencies, and ultimately nucleation rates from the rate at which seeded nuclei of different initial sizes grow or shrink. These ideas are also present in the classical nucleation theory (CNT), and yet seeding avoids several of the most problematic assumptions within CNT.[11] By using a cluster size coordinate, the seeding approach avoids defining an explicit surface around the nucleus and the associated capillarity approximations.[12,13] It also avoids commonly used assumptions about diffusion limited attachment by extracting the attachment frequency from short trajectory data. Seeding results have previously been combined with a spherical nucleus assumption to obtain an effective interfacial free energy, but in principle the seeding equations only require the interfacial free energy and the shape factor as a single lumped parameter.

The seeded simulation approach relies on accurate estimates of the critical nucleus size and the driving force $\Delta\mu$. The nucleus size metric is integral to the data analyses and to the interpretation of results, e.g., different nucleus sizes ascribed to the same nucleus lead to different inferred interfacial free energies. The driving force is a potential source of error for two reasons. First, small nuclei may have very different compositions[14] and/or structures,[15–17] etc. from their bulk counterpart phases. The seeding approach cannot accurately predict rates for nucleation processes that involve such non-classical nuclei.[18] Second, even when nuclei do have the structure and composition of the stable phase, nucleation rates are highly sensitive to the driving force.[19] Precise estimates of driving forces for solute precipitation pose major challenges in simulations.[20]

Despite these limitations, Sanz *et al.* demonstrated that seeding results are surprisingly accurate for crystal nucleation in supercooled liquids of hard-spheres, water, NaCl, and a Lennard-Jones fluid.[6,7] The findings by Sanz *et al.* motivated Zimmermann *et al.* to examine the quantitative predictions of seeded simulation rates for solute precipitate nucleation. Zimmermann *et al.* computed NaCl nucleation rates from supersaturated brine.[11] Salt nucleation from electrolyte solutions is important in geochemistry,[21] and it presents unique fundamental questions about electroneutrality of nuclei,

---
a)Authors to whom correspondence should be addressed: nerz@lbl.gov; d.quigley@warwick.ac.uk; bilsmith@uoguelph.ca; and baronp@engineering.ucsb.edu

non-ideal activities, and desolvation barriers during ion attachment. Aqueous NaCl was specifically chosen because it is a simple 1-1 electrolyte, because there are extensively characterized force fields,[20] and because experimental homogeneous nucleation rates have been reported.[22–24] For simulations, homogeneous nucleation is an advantageous starting point for several reasons. Homogeneous nucleation is a well-defined rate process, whereas heterogeneous nucleation can occur at many types of sites with unknown characteristics and populations. Additionally, properties like attachment frequencies, interfacial free energies, and driving forces that govern homogeneous nucleation rates also influence heterogeneous nucleation rates.

To compare predicted and measured solute precipitate nucleation rates, Zimmerman *et al.* compared theoretical and experimental rates *at the same driving forces*, i.e., comparisons of $J_{expt}(\Delta\mu)$ vs. $J_{sim}(\Delta\mu)$, where $(\Delta\mu)_{expt} = (\Delta\mu)_{sim} = \Delta\mu$. These comparisons require precise knowledge of the driving force vs. composition for the experimental system *and for the model system*.[8] The classical theory suggests that comparisons performed in this way will largely remove the effects of force field imperfections. Based on the classical theory, all residual differences in the $J_{expt}(\Delta\mu)$ vs. $J_{sim}(\Delta\mu)$ comparison stem from errors in the interfacial free energy, from errors in the attachment/detachment kinetics (prefactors), and from non-classical mechanisms. For rates computed with seeding, we show in this work that additional errors may also arise from the chosen nucleus size metric.

Zimmermann *et al.* chose NaCl concentrations in their seeded simulations to approximately match the driving forces in the available experiments. Unfortunately, the nucleation rates predicted by Zimmermann *et al.* were 15 to 30 orders of magnitude larger than the experimental rates.[11] Because the computed homogeneous rates were *faster* than the experiments, the discrepancy could not be ascribed to experiments that inadvertently probed heterogeneous nucleation. Zimmermann *et al.* speculated that the Joung-Cheatham force field with SPC/E water (JC/SPC/E) was not able to describe interfacial properties of the nuclei and even questioned the assumptions used to extract rates from the experimental measurements.[11]

Imperfect force fields and experimental interpretations are potentially important, but more importantly the electrolyte chemical potentials for the JC/SPC/E force field differ from those which Zimmermann *et al.* used to define the driving force. Joung and Cheatham estimated the solubility limit for their model as $7.2m$,[25] where m refers to molality in (mol NaCl/kg of water). Aragones *et al.* estimated the solubility as $4.8m$ by directly computing the chemical potentials of the solid and the electrolyte solution.[26] They recommended the similar and independently obtained value of $5.1m$ based on long time scale coexistence simulations. Recent studies starting with Moucka *et al.*[27–30] and followed by Mester and Panagiotopoulos,[31,32] Kolafa,[33] Benavides *et al.*,[34] and Espinosa *et al.*[35] have all arrived at estimates near $3.7m$.

This work reviews the seeded simulation approach and revises our predicted NaCl nucleation rates to account for the revised driving force estimates. We illustrate an analysis of variance (ANOVA) procedure[36] to quantify how errors in

the driving force and in the seeded simulation data influence the predicted rates. We also examine the effects of different nucleus size metrics in the seeding analysis.

II. BRIEF REVIEW OF SEEDING APPROACHES

The seeded simulation approach relies on some, but not all, aspects of CNT:

1. The seeding approach assumes that the reaction coordinate is nucleus size, an assumption that is justified by several recent analyses.[37–39]

2. The seeding approach assumes that the free energy (in units of $k_B T$) to form a nucleus is

$$\beta F(n) = -n\beta\Delta\mu + (\beta\phi\gamma)_{eff}n^{2/3}, \quad (1)$$

where $(\beta\phi\gamma)_{eff}$ is an effective dimensionless parameter that lumps together the shape factor ϕ, the interfacial free energy γ, and $\beta = 1/(k_B T)$. The lumped parameter is not required to match the value for nuclei with spherical shapes and/or macroscopic interfacial properties.[40]

3. The commonly used CNT rate expression is

$$J = \rho Z D_+ \Gamma \exp\left[-B/(\beta\Delta\mu)^2\right], \quad (2)$$

where ρ is the density of monomers, Z is the Zeldovich factor,[41] where D_+ is the attachment frequency, Γ is the Girshick-Chiu correction[42] $\{$i.e., $\Gamma = \exp[F(1)/(k_B T)]\}$, and the B parameter is[13]

$$B = \left[2(\beta\phi\gamma)_{eff}/3\right]^3/2. \quad (3)$$

Equations (2) and (3) emerge from the Frenkel-Zeldovich equation[13] or equivalently from an overdamped Langevin model in which the diffusivity along the (continuous) nucleus size coordinate is the attachment frequency.[43] By analogy, the seeding approach assumes that short trajectory swarms are realizations of an overdamped Langevin equation. After integrating away the noise, one obtains an equation that describes how the average nucleus size in a swarm of trajectories should evolve from some initial size n_0,[8]

$$\left.\frac{d\langle n\rangle}{dt}\right|_{n_0} = \left.\left\{-D_+\beta\Delta\mu + \frac{2}{3}D_+(\beta\phi\gamma)_{eff}n^{-1/3}\right\}\right|_{n_0}. \quad (4)$$

The quantity in brackets is proportional to $\partial F/\partial n$ as predicted by CNT. The expression should only be used at short times. If nuclei evolve for too long, the forces that influence their growth/dissolution will change.[44]

A seeding calculation begins with the creation of carefully solvated and annealed seeds ranging from very small pre-critical sizes to very large post-critical sizes. See the work of Zimmermann *et al.* for details on seed preparation.[11] For each group of initial seed sizes, many independent configurations are initiated and allowed to run for a short time. The seeded trajectories are used to estimate $d\langle n\rangle/dt$ for each initial nucleus size. Given $\beta\Delta\mu$, the seeded simulation data enable an estimate of all remaining properties that are needed to compute the rate. For example, Zimmermann *et al.* obtained D_+ and $(\beta\phi\gamma)_{eff}$ by fitting their data to Eq. (4) and then computed the nucleation rate according to Eq. (2).[11] Zimmermann *et al.* used five seeds per initial size and a seeded trajectory duration of

4 ns.[11] 4 ns is sufficiently short to avoid any appreciable depletion of the overall saturation level in the simulation box. The data presented by Zimmermann *et al.* suggest that 4 ns is also sufficiently long to identify an initial slope as $d\langle n\rangle/dt$. Seeding enables a straightforward uncertainty analysis (see below), but the most appropriate seed duration time for extracting an attachment frequency is not clear.

Consider, for example, swarm evolution from a post-critical seed. Because of the annealing procedure, the initial concentration around the seed should begin at the bulk supersaturation. During the first moments of swarm evolution, i.e., immediately after release of constraints, the surface concentration should drop rapidly to that given by the Ostwald-Freundlich equation.[45] As time progresses, the post-critical nucleus and the depleted boundary layer thickness around the nucleus should both grow. Estimates of the attachment frequency at both short and long times are complicated by the moving phase boundary and the changing depletion layer thickness.

Seeding approaches enable a prediction of trends, while methods like forward flux sampling[46] and direct simulation estimates of mean first passage times[47] provide the rate only at the simulated supersaturation. In particular, by predicting the barrier parameter B, seeded simulation results from one $\beta\Delta\mu$-value can be used to predict nucleation rates at other supersaturations.[48] But before a seeding calculation can be completed, we also require an accurate model for $\Delta\mu$. Section III reviews the challenges and recent evolution of chemical potential driving force estimates for nucleation in a supersaturated NaCl solution.

III. DRIVING FORCES

The seeded simulations of Zimmermann *et al.* examined NaCl concentrations from 8*m* to 12*m*, but directly computed electrolyte chemical potentials were only available up to 6*m*. An accurate model of the driving force in the metastable zone is critical for most versions of the seeded simulation approach (see the work of Lifanov *et al.* for the exception).[18] As noted in the Introduction, the driving force and solubility involve difficult calculations, and a consensus for the JC/SPC/E model has only recently emerged. Moucka *et al.*[27,29,30] obtained the first accurate solubility estimates at 298.15 K and 1 bar pressure using their osmotic ensemble Monte Carlo method, reporting a value of 3.64 ± 0.20 *m* for the JC/SPC/E force field. Mester and Panagiotopoulos subsequently used MD particle insertion and thermodynamic integration techniques to estimate the solubility as 3.59 ± 0.04 *m* or 3.71 ± 0.04 *m*[31,32] (the second estimate incorporates an improved treatment of the infinite dilution limit). Benavides *et al.* then obtained a value of 3.71 ± 0.20 *m* using a "global" MD thermodynamic integration calculation.[34,49] Finally, Kolafa[33] and Espinosa *et al.*[35] separately implemented a correct version of the direct MD coexistence method to estimate the solubility as 3.6 ± 0.4 *m* and 3.7 ± 0.4 *m*, respectively.

Most of the above solubility estimates required the electrolyte chemical potential $\mu_{\text{NaCl}}(m)$ at a discrete series of molalities up to a supersaturated value of 6*m*. The computed values of $\mu_{\text{NaCl}}(m)$ are used to parameterize analytic models

for the chemical potential. According to convention, the reference chemical potential $\mu^{\dagger}_{\text{NaCl}}$ is obtained by extrapolating Henry's law to a hypothetically ideal 1.0*m* NaCl solution.[50] Deviations from the ideal Henry's law behavior are captured in the activity coefficient $\gamma_{\pm}$:

$$\mu_{\text{NaCl}} = \mu^{\dagger}_{\text{NaCl}} + 2k_{\text{B}}T \ln m + 2k_{\text{B}}T \ln \gamma_{\pm}, \qquad (5)$$

where $\gamma_{\pm}$ is the overall NaCl electrolyte activity coefficient and $\mu^{\dagger}_{\text{NaCl}}$ for the JC/SPC/E model is approximately −391.6 kJ/mol.[31] Different investigators have used different models for the $\ln\gamma_{\pm}$ part of $\mu_{\text{NaCl}}(m)$. Several of the proposed models accurately fit the computed $\mu_{\text{NaCl}}(m)$ values, but in our study $\mu_{\text{NaCl}}(m)$ must be extrapolated beyond 6*m* for comparisons to experimental nucleation rates. The functional form of $\ln\gamma_{\pm}$ that is used to extrapolate $\mu_{\text{NaCl}}(m)$ into the metastable supersaturation zone influences the estimated driving force in this concentration range. A satisfactory model of $\ln\gamma_{\pm}$ for extrapolation deep into the metastable zone must account for the finite size of ions and the concentration dependence of the dielectric constant.[51] We have used an extension of the Debye-Hückel limiting law[52] similar to the expression of Hamer and Wu[53] used to fit experimental data, but restricted to a linear term appended to the Debye-Hückel term,

$$\log_{10}\gamma_{\pm} = -\frac{a\sqrt{m}}{1 + b\sqrt{m}} + cm. \qquad (6)$$

We used the theoretical expression for a depending on the H_2O dielectric constant and density,[52] yielding $a = 0.568$ and fitted b and c to the combined simulation data of Moucka *et al.*[27] and of Mester and Panagiotopoulos[31] along with new high-*m* data points from this work using the same methodology as in Ref. 34. The fit yields values $b = 1.177\,69\ m^{-1/2}$ and $c = 0.177\,157\ m^{-1}$. Figure 1 shows the computed solid NaCl chemical potential as well as the computed and extrapolated electrolyte chemical potentials.

The revised calculations of $\mu_{\text{NaCl}}(m)$ leave almost no uncertainty in the solubility, but there remains considerable uncertainty in the driving force for molalities deep in the metastable zone. Specifically, we determined that the uncertainties increased with increasing molality: 0.111, 0.125, 0.168, 0.226, and 0.363 kJ/mol for m = 3.7, 8, 10, 12, and 16, respectively. These uncertainties are calculated (i) by constructing a new data set by perturbing each data point around its initial position according to a normal distribution with width 0.2 kJ/mol (the reported simulation uncertainty[31]) and (ii) then refitting the new data set to Eq. (6); this procedure is repeated 1000 times to yield average chemical potentials and standard deviations at varying molalities.

The driving force basis should be consistent with the definition of nucleus size. In our case, nucleus size is counted in ions, not in formula units of NaCl. The driving force for nucleation *on a per-ion basis* is

$$\Delta\mu \equiv 0.5\,\Delta\mu_{\text{NaCl}} = 0.5(\mu_{\text{NaCl}}(m) - \mu_{\text{solid}}), \qquad (7)$$

i.e., the difference between the electrolyte chemical potential in solution and the solid NaCl chemical potential, $\mu_{\text{solid}} = \mu_{\text{NaCl}}(m_{\text{sat}})$, at the same temperature and pressure.
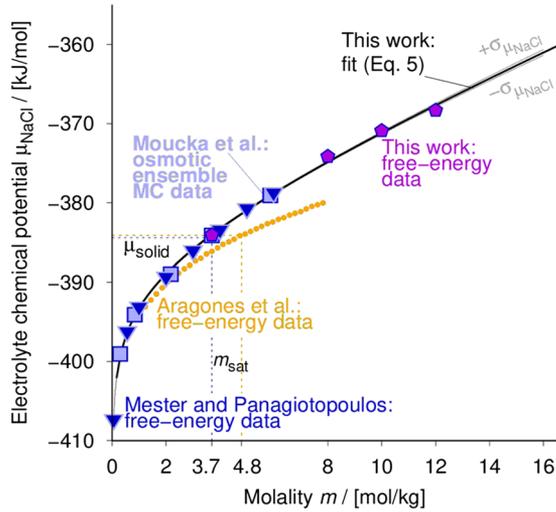
FIG. 1. Electrolyte chemical potential of sodium chloride, μ_{NaCl}, as a function of molality, m; $T = 298$ K and $p = 1$ bar. Blue squares and blue triangles show the electrolyte chemical potential as computed by Moucka *et al.*[27] and Mester and Panagiotopoulos,[31] whereas purple pentagons are new data from this work. The black curve extrapolates the data points using Eqs. (5) and (6) to $m = 16$, and the gray curves give the uncertainty as 1 standard deviation. The gold circles show the early results of Aragones *et al.*[26] which were used by Zimmermann *et al.*[11] Dashed horizontal lines show the solid NaCl chemical potential. The estimated solubility limits are also shown as vertical lines.

IV. RATES AND UNCERTAINTIES FROM SEEDING

Equation (4) suggests that one can directly regress the seeded simulation data ($n^{-1/3}$ and $d\langle n\rangle/dt$) to infer D_+ and $(\beta\phi\gamma)_{\text{eff}}$ from the slope and intercept of the best fit line. The resulting estimates of D_+ and $(\beta\phi\gamma)_{\text{eff}}$ can be used to estimate the rate via Eqs. (2) and (3). D_+ and $(\beta\phi\gamma)_{\text{eff}}$ are both random variables (estimates from linear regression), and their uncertainties propagate into the rate calculation. The slope and intercept estimates in linear regression [and therefore D_+ and $(\beta\phi\gamma)_{\text{eff}}$] are correlated to one another.[36] The complicated dependence of the rate on D_+ and $(\beta\phi\gamma)_{\text{eff}}$, as well as their interdependence, complicates efforts to quantify uncertainty in the rate estimate.

To facilitate the rate calculation and error analysis, note that the free energy barrier within the exponential part of rate expression (2) can be written as[4]

$$B/(\beta\Delta\mu)^2 = \frac{1}{2}n_{\ddagger}\beta\Delta\mu. \qquad (8)$$

Equation (8) is an identity that emerges from classical nucleation theory.[12,13] To exploit Eq. (8), we can estimate $n_{\ddagger}$ and D_+ directly from the seeded simulation data, with $\beta\Delta\mu$ being separately computed using methods listed in Sec. III. To estimate $n_{\ddagger}$, rearrange Eq. (4) to give $n^{-1/3}$ as a function of dn/dt rather than dn/dt as a function of $n^{-1/3}$. The rearrangement still allows identification of D_+ and $(\beta\phi\gamma)_{\text{eff}}$ from the slope and intercept so that the prefactor and B-parameter can be determined. However, the regression line also provides an estimate of $n_{\ddagger}$ (where $dn/dt = 0$) and its uncertainty for the rate calculation [see Fig. 2(b)]. An alternative seeding procedure[9] uses an algorithm somewhat like bisection to identify the size at which $dn/dt = 0$.
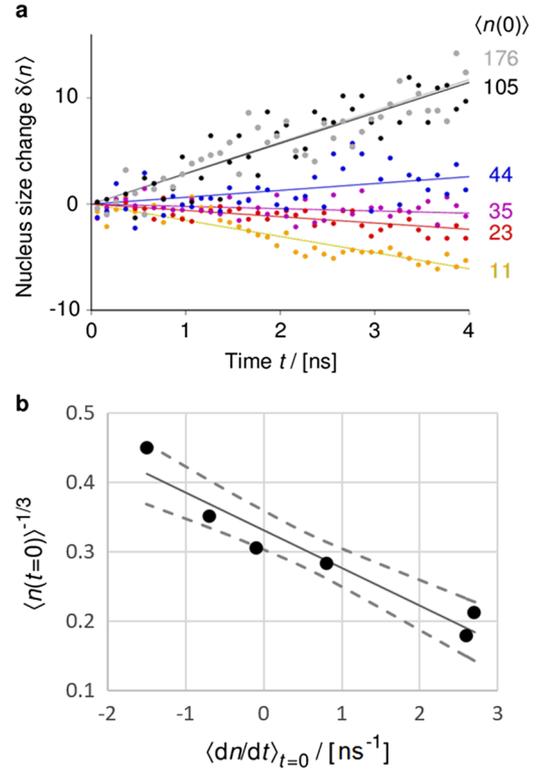
FIG. 2. (a) Seeded simulation results for an aqueous solution of NaCl with molality $8m$. The initial nucleus sizes are 11, 23, 35, 44, 105, and 176 ions. (b) Linear regression results. Each data point corresponds to five trajectories initiated from carefully solvated nuclei of a specific initial size. The dashed curves show a 90% confidence interval on the regression model.

Figure 2(a) shows seeded simulation data from the work of Zimmermann *et al.*[11] on rocksalt NaCl nucleation from an aqueous solution of sodium chloride at a molality of $8m$. Figure 2(b) plots these data as $n^{-1/3}$ vs. $d\langle n\rangle/dt$. This procedure assumes an n-independent attachment frequency, but the attachment frequency should actually depend on nucleus size.[1] Zimmermann *et al.*[11] showed that ion-attachment to NaCl nuclei is desolvation limited, so D_+ should be proportional to $n^{2/3}$. More elaborate regression procedures could account for n-dependence in D_+, but the effects are expected to be small. As the molality changes from m = 8, 10, to 12, the critical nucleus size changes from 31, 14, to 6 ions. The corresponding attachment frequencies 1.03/ns, 1.47/ns, and 1.47/ns, obtained from $\langle(\delta n)^2\rangle$ vs t,[54] are nearly constant.

The separate estimate of $\beta\Delta\mu$ is needed to interpret the linear regression results, to predict the nucleation rate, *and* to place predictions and experimental measurements on the $\ln J$ vs. $1/(\beta\Delta\mu)^2$ plot as shown in Fig. 3. The original rate predictions of Zimmermann *et al.*[11] using $\mu_{\text{NaCl}}(m)$ and m_{sat} from the work of Aragones *et al.* were between 10^{15} and 10^{30} orders of magnitude faster than the experimental rates. Figure 3 shows that most of the error is reconciled by calculations based on the revised chemical potential driving forces.

Unfortunately, the revised driving forces used in the simulations no longer match those used in the experiments. It would be useful to study nucleation at lower NaCl molalities in future work. According to CNT, the critical nucleus size $n_{\ddagger}$ scales as $n_{\ddagger} \sim (\beta\Delta\mu)^{-3}$. Based on directly computed critical sizes in $8m$,
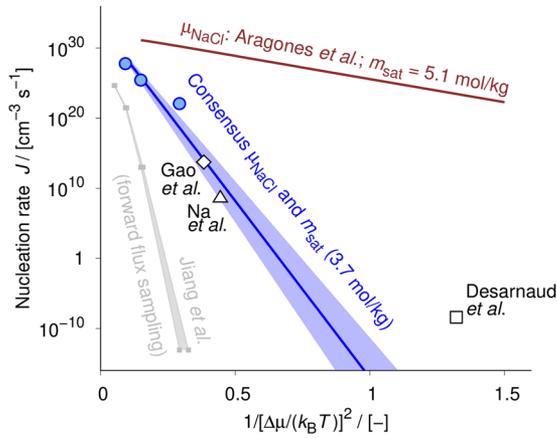
FIG. 3. Homogeneous nucleation rate, J, of NaCl in water as a function of the squared inverse driving force, $1/[\Delta\mu/(k_BT)]^2$. The red line gives predicted rates from seeded simulation analysis using $\Delta\mu$ based on the m_{sat} and $\mu_{NaCl}(m)$ calculations of Aragones *et al.*[26] The predictions in blue are based on the electrolyte chemical potential model of Eqs. (5) and (6) and the consensus solubility: 3.7 mol/kg. The three solid blue circles are from separate analyses of the seeding data at 8m, 10m, and 12m each giving separate D_+ and $(\beta\phi\gamma)_{eff}$ parameters. The solid blue line uses one molality-independent interfacial free energy, a mean attachment frequency, and Eq. (2) to extrapolate the trends of the nucleation rate. The open symbols (diamond, triangle, and square) are three independent experimental estimates of the homogeneous nucleation rate.[22–24] Nucleation rates from forward flux sampling simulations by Jiang *et al.*[55] are shown in gray.

10m, and 12m solutions, the critical nuclei at 6m would have sizes near 200 ions (see the supplementary material). These exceed our largest simulations, but it should be possible to reach such sizes in future seeded simulations.

Two of the experimental rates shown in Fig. 3, from the studies of Na *et al.*[24] and Gao *et al.*,[23] were measured in droplets that were levitated and passed through an exposure chamber, respectively. The levitation experiments were designed to favor homogeneous nucleation by eliminating all potential heterogeneous nucleation sites except for an air-water interface. By contrast, Desarnaud *et al.* examined NaCl nucleation within a high surface area mesoporous silica.[22] Because of the large disordered silica-water interface, heterogeneous nucleation may explain the results of Desarnaud *et al.* which are between 10 and 20 orders of magnitude faster than the trend suggested by Gao *et al.*,[23] Na *et al.*,[24] and the revised calculations.

The revised computational results suggest that seeded simulations combined with accurate driving force calculations can be a powerful combination for predicting accurate nucleation rates. However, the striking difference between the previous rates of Zimmermann *et al.*[11] and the revised predictions of this study also illustrates the tremendous importance of a precise $\Delta\mu$ calculation for the molecular model. Apparently, much of the original 10^{15}-10^{30} error resulted from errors in m_{sat} and $\mu_{NaCl}(m)$ that correspond to just fractional k_BT errors in the free energy calculations.[20,34]

Why do such small errors lead to giant errors in the nucleation rate? Qualitatively, we can think of nucleation as a high order association reaction, and it is well known that reactions of high order are extremely sensitive to concentration. Quantitatively, we can estimate the uncertainty in the predicted rates

by examining two key parts of the exponential term, $n_{\ddagger}$ and $\beta\Delta\mu$. Starting from the differential relationship

$$d(n^{-1/3}) = -\frac{1}{3}n^{-4/3}dn \qquad (9)$$

and then using typical error propagation rules to relate variances in estimates of $n_{\ddagger}$ and $n_{\ddagger}^{-1/3}$,

$$\sigma_{n_{\ddagger}^{-1/3}}^2 = \frac{1}{9}n_{\ddagger}^{-8/3}\sigma_{n_{\ddagger}}^2. \qquad (10)$$

Inverting Eq. (10) gives $\sigma_{n_{\ddagger}}^2 = 9n_{\ddagger}^{8/3}\sigma_{n_{\ddagger}^{-1/3}}^2$. The variance in $n_{\ddagger}^{-1/3}$ can be obtained from analysis of variance rules for the linear regression, as shown in Fig. 2(b).[36]

To assess uncertainty in $\beta\Delta\mu$, we begin from the per-ion basis driving force

$$\beta\Delta\mu = 0.5(\beta\mu_{NaCl}(m) - \beta\mu_{NaCl}(m_{sat})). \qquad (11)$$

The uncertainty in the driving force in Eq. (7) is given by standard uncertainty propagation rules as

$$\sigma_{\beta\Delta\mu}^2 = (0.5)^2\left[\sigma_{\beta\mu_{NaCl}}^2(m) + \sigma_{\beta\mu_{NaCl}}^2(m_{sat})\right]. \qquad (12)$$

We use $\sigma_{\beta\mu_{NaCl}}(m_{sat}) = (0.2\text{ kJ/mol})/RT$ from the work of Nezbeda *et al.*[20] and values of $\sigma_{\mu_{NaCl}}(m)$ as given in Sec. III. We assume that the uncertainty in $\ln J$ is dominated by the term $n_{\ddagger}\beta\Delta\mu/2$ in the exponential. Accordingly, the uncertainty in $\ln[J/\rho ZD_+\Gamma]$ can be estimated from $\sigma_{\beta\Delta\mu}$ and $\sigma_{n_{\ddagger}}$ as

$$\sigma_{\ln[J/\rho D_+Z\Gamma]} = \frac{1}{2}n_{\ddagger}\beta\Delta\mu\sqrt{\frac{\sigma_{n_{\ddagger}}^2}{n_{\ddagger}^2} + \frac{\sigma_{\beta\Delta\mu}^2}{(\beta\Delta\mu)^2}}. \qquad (13)$$

Table I shows the estimates and uncertainties of the driving force, the critical nucleus size, and $\ln[J/\rho ZD_+\Gamma]$.

Based on the results in Table I, the error bars on the revised computational estimates of $\ln J$ are approximately four natural log units. In other words, the rate estimates (*within assumptions of the seeding approach*) are numerically accurate to within a factor of ca. 40. Because the error contributions from $\beta\Delta\mu$ have probably been overestimated, the dominant source of numerical error in the revised calculations is most likely from the estimate of $n_{\ddagger}$. Of course, tremendous computational effort has been invested in precise driving force estimates for the JC/SPC/E model of NaCl and water. For most other solutes and solvents, the dominant source of error will likely be uncertainty in the driving forces. Moreover, Sec. V shows that there are additional sources of uncertainty related to assumptions in the seeding approach.

TABLE I. Values and uncertainties of $\beta\Delta\mu$, $n_{\ddagger}$, and $\ln[J/\rho ZD_+\Gamma]$ from regression of seeding data and error propagation analysis. The rows labeled "due to $\sigma_{n_{\ddagger}}$" and "due to $\sigma_{\beta\Delta\mu}$" show that the uncertainty in $\sigma_{n_{\ddagger}}$ makes a larger contribution than the uncertainty in $\sigma_{\beta\Delta\mu}$ to the overall uncertainty in $\ln J$.

	8 m	10 m	12 m
$\beta\Delta\mu$	1.85 ± 0.05	2.59 ± 0.05	3.30 ± 0.06
$n_{\ddagger}$	27 ± 3	15 ± 3	9 ± 2
B	87 ± 12	129 ± 25	158 ± 38
$\ln[J/\rho ZD_+\Gamma]$	-25.3 ± 3.1	-19.3 ± 3.5	-14.5 ± 3.4
Due to $\sigma_{n_{\ddagger}}$	± 3.0	± 3.5	± 3.4
Due to $\sigma_{\beta\Delta\mu}$	± 0.7	± 0.4	± 0.3

V. INFLUENCE OF THE NUCLEUS SIZE METRIC

For single component nucleation processes, the free energy (and equilibrium population) as a function of nucleus size can be obtained from traditional rare event sampling methods.[56,57] Suppose that the free energy calculation is performed with two different nucleus size metrics, one that overestimates nucleus size and one that underestimates nucleus size. The two calculations will give two different free energy vs. nucleus size relationships with one peaked at a large size and one peaked at a small size. On the other hand, the two free energy *barriers* will be similar, as long as one metric consistently overestimates nucleus size while the other consistently underestimates nucleus size. The reason is that the barrier reflects the (small) population of critical clusters at equilibrium, regardless of the size ascribed to them in the clustering algorithm.

Typical nucleus size metrics use a local order parameter to identify atoms in a solute-rich and/or ordered local environment. Then a clustering algorithm groups these atoms into clusters and reports the cluster size. For the same nucleus, using different local order parameters can yield different cluster sizes. For example, consider a $3 \times 3 \times 3$ nucleus of "solutes" on a cubic lattice and suppose that all surrounding sites are occupied by "solvents" as shown in Fig. 4. One could use coordination numbers (CNs) to determine whether individual solute atoms are part of the nucleus. Let the coordination number (CN) for a given solute be the number of solute nearest neighbors. If our local order parameter requires CN ≥ 3, then the clustering algorithm will identify a contiguous cluster of size $n = 27$ solutes. If the local order parameter requires CN ≥ 4, then corner sites on the nucleus are not counted and clustering will give $n = 19$. If a stringent threshold of CN ≥ 6 is used for the local order parameter, then only $n = 1$ solute (at the center of the $3 \times 3 \times 3$ cube) will be counted.

This section shows that the seeding approach, in contrast to traditional rare events methods, gives results which are highly sensitive to the nucleus size metric. This sensitivity of the results to the nucleus size metric increases as the size of inserted seed decreases, given that the ratio interface/volume increases. In Fig. 5, the data from Zimmermann *et al.* have been reanalyzed to obtain rate estimates with several different nucleus size metrics. Zimmermann *et al.* used a stringent local order parameter that requires an octahedral arrangement of counterions around each ion (open blue circles). The rates

FIG. 5. Homogeneous nucleation rate, J, of NaCl in water as a function of the squared inverse driving force, $1/[\Delta\mu/(k_B T)]^2$ obtained from different definitions for the cluster size, n. CN refers to a coordination number and q_{oct} is the octahedral order parameter from the work of Zimmermann *et al.* One outlier data point for the CN ≥ 6 criterion (8 m: 4×10^{-37}/cm³/s) falls outside the plotted range of rates (see the supplementary material for more information). We also show rate estimates from forward flux sampling simulations by Jiang *et al.*[55] and from brute-force simulations at high molality.[58]

obtained by Zimmermann *et al.* are similar to rates obtained from analysis with a stringent CN ≥ 6 requirement (blue diamonds). By systematically varying the CN-threshold, we can see how the computed nucleation rates (from the seeding procedure) change. Different CN-thresholds from CN ≥ 1 to CN ≥ 6 yield rate estimates that span 30 orders of magnitude. The rate estimate increases as the CN threshold becomes more stringent.

Figure 5 also shows results from the simulations of Lanaro and Patey. These authors directly observe nucleation in brute force simulations of a 15*m* NaCl solution. Lanaro and Patey[58] used the same JC/SPC/E force field to study nucleation in direct large scale simulations at 15.6*m*, 22.2*m*, and beyond. They did not report nucleation rates, but at 15.6*m* and 22.2*m* they reported the duration of their simulations and the number of nuclei generated. We have used their data with the approximation

$$J \geq \frac{\left(\begin{array}{c}\text{total number of}\\\text{nucleation events}\end{array}\right)}{\left(\begin{array}{c}\text{simulation}\\\text{box volume}\end{array}\right) \times \left(\begin{array}{c}\text{total duration}\\\text{of simulations}\end{array}\right)}. \quad (14)$$

Equation (14) gives an approximate lower bound on J because multiple nucleation events occurred in the simulations by Lanaro and Patey. Growth of the earliest nuclei consumes the supersaturation and tends to suppress further nucleation. As shown in Fig. 3, Eq. (14) applied to the data of Lanaro and Patey yields $J \geq 6.5 \times 10^{24}$/cm³ s at 15.7m and $J \geq 1.2 \times 10^{25}$/cm³ s at 22.2m.

FIG. 4. For a $3 \times 3 \times 3$ cluster of solutes on a cubic lattice, different coordination number (CN) thresholds for clustering lead to different nucleus sizes. (Left) Using CN ≥ 3 includes solutes at corners, edges, faces, and interior sites. (Middle) Using CN ≥ 4 includes solutes at edges, faces, and interior sites. (Right) CN ≥ 5 includes solutes at faces and interior sites.

We can understand the effects of different nucleus size metrics from Eqs. (2) and (8). Given a value of the driving force, the estimated free energy barrier for nucleation is directly proportional to the estimated critical nucleus size,

$$J = \rho Z D_+ \Gamma \exp\left[-\frac{1}{2} \frac{n_{\ddagger} \Delta \mu}{k_B T} \right]. \qquad (15)$$

Lax order parameters cause clustering algorithms to overestimate the nucleus size, to overestimate the free energy barrier, and therefore to underestimate the rate. Stringent order parameters cause clustering to underestimate the nucleus size, to underestimate the barrier, and to overestimate the rate. The dramatic effect of different nucleus size metrics is disconcerting because there is no straightforward way to choose the most suitable cluster-size definition.

Seeding analyses use a computational estimate of the critical nucleus size within the CNT rate expression; therefore, the ideal nucleus size metric should be consistent with the assumptions of CNT. Identifying an appropriate nucleus size metric for seeding analysis is non-trivial[7] because a true critical cluster is discrete, granular, and fluxional, while the CNT model has sharp interfaces that separate continuous stable and metastable phases. As yet, there are no generally established principles to select an optimal CNT-consistent metric, but we offer the following list of considerations:

(1) A stringent clustering threshold on the local order parameter ensures that a cluster of size n contains at least n atoms with bulk crystal-like structure, consistent with the interpretation of the $n\Delta\mu$ term in the CNT free energy expression. However, stringent local order parameter thresholds omit surface atoms from the nucleus. Because so many atoms in a small nucleus are part of the surface, seeding calculations must decide whether surface atoms are "on" or "in" the nucleus—a question that is avoided by the sharp interface assumption in CNT.

(2) A lax local order parameter threshold, e.g., CN ≥ 1 or CN ≥ 2, will include adsorbates that are clearly "on" (not "within") the surfaces of nuclei. These adsorbates may be dynamically committed to the nucleus and still exterior to a CNT-consistent dividing surface. Extremely lax order parameter thresholds (CN ≥ 1) with highly concentrated solutes also risk crossing the percolation limit in clustering, which may lead to extremely large nuclei.

(3) CNT can be formulated in numerous ways, but for droplet nucleation an equimolar dividing surface leads to a convenient formulation with no surface excess atoms.[13] An equimolar dividing surface is difficult to define for solid-fluid interfaces because the density profile is a series of sharp peaks rather than a smooth sigmoid curve.[59] Nevertheless, the analogy to droplet nucleation theories suggests that the nucleus should include atoms whose CN is approximately halfway between the large CN for atoms in the bulk solid and the small CN for atoms in the metastable solution.

Theories of crystal growth at low supersaturation focus on the half-lattice positions, i.e., "kink sites," on surfaces. Adding one atom to a kink site adds one atom to the bulk crystal lattice by translational symmetry arguments. For a cubic "Kossel" crystal, the kink sites have CNs that are half of the bulk coordination number.[60,61] For a large crystal, these considerations suggest including kink sites and more strongly coordinated atoms in the nucleus size n and excluding atoms with lower CNs. In the case of a cubic lattice, the half-lattice CN prescription suggests that CN ≥ 3 is an optimal clustering criterion. Interestingly, the forward flux sampling results[55] (which should be independent of the clustering criterion) agree most closely with the CN ≥ 3 and CN ≥ 4 results.

We emphasize that our work can provide only tentative recommendations, as many potentially important factors have not been considered. Complex crystal structures may present many types of kink sites and many different building units. Moreover, the facets of nuclei are not infinitely large, so any rules inspired by large facets may require amendments to account for corners and edges. Indeed, the use of an equimolar dividing surface for droplet nucleation leads to a size-dependent surface tension because of the Tolman correction.[62,63] Beyond questions about dividing surfaces and nucleus size metrics, the procedure for solvating and annealing seeds at non-equilibrium sizes may influence their stability and dynamics. The many sources of uncertainty in our results, some not easily resolved, point to the need for more robust rare event methods that can circumvent the special difficulties in simulating solute precipitate nucleation.

VI. CONCLUSIONS

This work examines the sources of error in seeded simulation results for NaCl nucleation from aqueous solution. Our results demonstrate that the seeding approach relies critically on a precise driving force for the model system. For accurate rate estimates, the driving force vs. solute concentration curve, i.e., $\Delta\mu(m)$, need not exactly reproduce that of the real system, but it should accurately describe the thermodynamic properties *of the model system*. We have demonstrated these calculations and presented uncertainty analyses for the rates of NaCl nucleation from a supersaturated NaCl electrolyte solution.

We have further demonstrated that the nucleus size metric also has a strong influence on the predicted nucleation rates. Because the seeding approach uses the critical nucleus size in a CNT-like rate expression, the ideal nucleus size metric should be consistent with the equimolar dividing surface definition used in classical nucleation theory. Size metrics that overestimate nucleus sizes lead to underestimated nucleation rates, and size metrics that underestimate nucleus sizes lead to overestimated rates. Further work is needed to determine the optimal local order parameters and clustering rules for computing nucleus size, but our results point to some tentative suggestions. A comparison between our results and independent forward flux sampling results[55] suggests that the optimal nucleus size metric should include all atoms in the bulk, surface, edge, and perhaps corner sites. Solute atoms with lower coordination to other solute atoms should be viewed as adsorbates and excluded from the nucleus size estimate. The numerous and potentially gigantic sources of error in rates

obtained by the seeding approach should motivate the development of more robust rare event methods to overcome the unique challenges encountered in studies of solute precipitate nucleation.

SUPPLEMENTARY MATERIAL

See supplementary material for how rates were estimated from brute force simulations. It also shows how a single outlier data point in the reprocessed 8*m* trajectories leads to an extremely small rate estimate. Data associated with this manuscript is available for download at http://wrap.warwick.ac.uk/101981.

[1]V. Agarwal and B. Peters, "Solute precipitate nucleation: A review of theory and simulation advances," Adv. Chem. Phys. **155**, 97–160 (2014).
[2]R. J. Davey, S. L. M. Schroeder, and J. H. ter Horst, "Nucleation of organic crystals—A molecular perspective," Angew. Chem., Int. Ed. **52**, 2166–2179 (2013).
[3]D. Kashchiev and G. M. Van Rosmalen, "Review: Nucleation in solutions revisited," Cryst. Res. Technol. **38**, 555–574 (2003).
[4]K. F. Kelton and A. L. Greer, *Nucleation in Condensed Matter: Applications in Materials and Biology* (Elsevier, Oxford, 2010).
[5]J. W. P. Schmelzer, *Nucleation Theory and Applications* (Wiley, Wienheim, 2006).
[6]J. R. Espinosa, E. Sanz, C. Valeriani, and C. Vega, "Homogeneous ice nucleation evaluated for several water models," J. Chem. Phys. **141**(18), 18C529 (2014).
[7]J. R. Espinosa, C. Vega, C. Valeriani, and E. Sanz, "Seeding approach to crystal nucleation," J. Chem. Phys. **144**, 034501 (2016).
[8]B. C. Knott, V. Molinero, M. F. Doherty, and B. Peters, "Homogeneous nucleation of methane hydrates: Unrealistic under realistic conditions," J. Am. Chem. Soc. **134**, 19544–19547 (2012).
[9]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).
[10]X. M. Bai and M. Li, "Calculation of solid-liquid interfacial free energy: A classical nucleation theory based approach," J. Chem. Phys. **124**, 124707 (2006).
[11]N. E. R. Zimmermann, B. Vorselaars, D. Quigley, and B. Peters, "Nucleation of NaCl from aqueous solution: Critical sizes, ion-attachment kinetics, and rates," J. Am. Chem. Soc. **137**, 13352–13361 (2015).
[12]V. I. Kalikmanov, *Nucleation Theory* (Springer, Dordrecht, 2013).
[13]D. Kashchiev, *Nucleation* (Butterworth-Heinemann, Oxford, 2000).
[14]G. Wilemski, "Composition of the critical nucleus in multicomponent vapor nucleation," J. Chem. Phys. **80**, 1370–1372 (1984).
[15]N. Duff and B. Peters, "Nucleation in a Potts lattice gas model of crystallization from solution," J. Chem. Phys. **131**, 184101 (2009).
[16]P. R. ten Wolde and D. Frenkel, "Enhancement of protein crystal nucleation by critical density fluctuations," Science **277**, 1975–1978 (1997).
[17]P. G. Vekilov, "Two-step mechanism for the nucleation of crystals from solution," J. Crystal Growth **275**, 65–76 (2005).
[18]Y. Lifanov, B. Vorselaars, and D. Quigley, "Nucleation barrier reconstruction via the seeding method in a lattice model with competing nucleation pathways," J. Chem. Phys. **145**, 211912 (2016).
[19]A. Haji-Akbari and P. G. Debenedetti, "Direct calculation of ice homogeneous nucleation rate for a molecular model of water," Proc. Natl. Acad. Sci. U. S. A. **112**, 10582–10588 (2015).
[20]I. Nezbeda, F. Moucka, and W. R. Smith, "Recent progress in molecular simulation of aqueous electrolytes: Force fields, chemical potentials and solubility," Mol. Phys. **114**, 1665–1690 (2016).
[21]M. Babel and B. Schrieber, "Geochemistry of evaporites and evolution of seawater," in *Treatise on Geochemistry*, 2nd ed., edited by F. Mackenzie (Elsevier, 2014), Chap. 9, pp. 483–560.
[22]J. Desarnaud, H. Derluyn, J. Carmeliet, D. Bonn, and N. Shahidzadeh, "Metastability limit for the nucleation of NaCl crystals in confinement," J. Phys. Chem. Lett. **5**, 890–895 (2014).
[23]Y. Gao, L. E. Yu, and S. B. Chen, "Efflorescence relative humidity of mixed sodium chloride and sodium sulfate particles," J. Phys. Chem. A **111**, 10660–10666 (2007).
[24]H.-S. Na, S. Arnold, and A. S. Myerson, "Cluster formation in highly supersaturated solution droplets," J. Crystal Growth **139**, 104–112 (1994).
[25]I. S. Joung and T. E. Cheatham III, "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations," J. Phys. Chem. B **112**, 9020–9041 (2008).
[26]J. L. Aragones, E. Sanz, and C. Vega, "Solubility of NaCl in water by molecular simulation revisited," J. Chem. Phys. **136**, 244508 (2012).
[27]F. Moucka, I. Nezbeda, and W. R. Smith, "Molecular simulation of aqueous electrolytes: Water chemical potential results and Gibbs-Duhem equation consistency tests," J. Chem. Phys. **139**, 124505 (2013).
[28]F. Moucka, I. Nezbeda, and W. Smith, "Molecular force field development for aqueous electrolytes. I. Incorporating appropriate experimental data and the inadequacy of simple electrolyte force fields based on Lennard-Jones and point charge interactions with Lorentz–Berthelot rules," J. Chem. Theory Comput. **9**, 5076–5085 (2013).
[29]F. Moucka, I. Nezbeda, and W. R. Smith, "Chemical potentials, activity coefficients, and solubility in aqueous NaCl solutions: Prediction by polarizable force fields," J. Chem. Theory Comput. **11**, 1756–1764 (2015).
[30]F. Moucka, M. Lísal, J. Škvor, J. Jirsák, I. Nezbeda, and W. R. Smith, "Molecular simulation of aqueous electrolyte solubility. II. Osmotic ensemble Monte Carlo methodology for free energy and solubility calculations and application to NaCl," J. Phys. Chem. B **115**, 7849–7861 (2011).
[31]Z. Mester and A. Z. Panagiotopoulos, "Mean ionic activity coefficients in aqueous NaCl solutions from molecular dynamics simulations," J. Chem. Phys. **142**, 044507 (2015).
[32]Z. Mester and A. Z. Panagiotopoulos, "Temperature-dependent solubilities and mean ionic activity coefficients of alkali halides in water from molecular dynamics simulations," J. Chem. Phys. **143**, 044505 (2015).
[33]J. Kolafa, "Solubility of NaCl in water and its melting point by molecular dynamics in the slab geometry and a new BK3-compatible force field," J. Chem. Phys. **145**, 204509 (2016).
[34]A. L. Benavides, J. L. Aragones, and C. Vega, "Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route," J. Chem. Phys. **144**, 124504 (2016).
[35]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).
[36]J. S. Milton and J. C. Arnold, *Probability and Statistics in the Engineering and Computing Sciences* (McGraw-Hill, New York, 2003).
[37]S. Jungblut, A. Singraber, and C. Dellago, "Optimising reaction coordinates for crystallisation by tuning the crystallinity definition," Mol. Phys. **111**, 3527–3533 (2013).
[38]Y. Bi, A. Porras, and T. Li, "Free energy landscape and molecular pathways of gas hydrate nucleation," J. Chem. Phys. **145**, 211909 (2016).
[39]L. Lupi, V. Molinero, and B. Peters, "Pre-ordering of interfacial water in the pathway of heterogeneous ice nucleation does not lead to a two-step crystallization mechanism," J. Chem. Phys. **145**, 211910 (2016).
[40]E. M. Blokhuis and J. Kuipers, "Thermodynamic expressions for the Tolman length," J. Chem. Phys. **124**, 074701 (2006).
[41]J. Zeldovich, "Theory of nucleation and condensation," Sov. Phys. JETP **12**, 525 (1942) (in English).
[42]S. L. Girshick and C. P. Chiu, "Kinetic nucleation theory: A new expression for the rate of homogeneous nucleation from an ideal supersaturated vapor," J. Chem. Phys. **93**, 1273–1277 (1990).

[43]B. Peters, *Reaction Rate Theory and Rare Events* (Elsevier, Amsterdam, 2017).

[44]B. C. Knott, N. Duff, M. F. Doherty, and B. Peters, "Estimating diffusivity along a reaction coordinate in the high friction limit: Insights on pulse times in laser-induced nucleation," J. Chem. Phys. **131**, 224112 (2009).

[45]B. Peters, "On the coupling between slow diffusion transport and barrier crossing in nucleation," J. Chem. Phys. **135**, 044107 (2011).

[46]R. J. Allen, D. Frenkel, and P. R. ten Wolde, "Forward flux sampling-type schemes for simulating rare events: Efficiency analysis," J. Chem. Phys. **124**, 024102 (2006).

[47]J. Wedekind, R. Strey, and D. Reguera, "New method to analyze simulations of activated processes," J. Chem. Phys. **126**, 134103 (2007).

[48]B. Peters, "Common features of extraordinary rate theories," J. Phys. Chem. B **119**, 6349–6356 (2015).

[49]A. L. Benavides, M. A. Portillo, J. L. F. Abascal, and C. Vega, "Estimating the solubility of 1:1 electrolyte aqueous solutions: the chemical potential difference rule," Mol. Phys. **115**, 1301 (2017).

[50]K. G. Denbigh, *The Principles of Chemical Equilibrium* (Cambridge University Press, Cambridge, 1981).

[51]J. M. G. Barthel, H. Krienke, and W. Kunz, *Physical Chemistry of Electrolyte Solutions: Modern Aspects* (Springer, New York, 1998).

[52]P. Debye and E. Huckel, "Zur Theorie der Elektrolyte I. Gefrierpunktserniedrigung und verwandte Erscheinungen," Phys. Z. **9**, 185–206 (1923).

[53]W. J. Hamer and Y. C. Wu, "Osmotic coefficients and mean activity coefficients of uni-univalent electrolytes in water at 25 °C," J. Phys. Chem. Ref. Data **1**, 1047 (1972).

[54]S. Auer and D. Frenkel, "Quantitative prediction of crystal-nucleation rates for spherical colloids: A computational approach," Ann. Rev. Phys. Chem. **55**, 333–361 (2004).

[55]H. Jiang, A. Haji-Akbari, P. G. Debenedetti, and A. Z. Panagiotopoulos, "Forward flux sampling calculation of homogeneous nucleation rates from aqueous NaCl solutions," J. Chem. Phys. **148**, 044505 (2018).

[56]G. Bussi and D. Branduardi, "Free-energy calculations with metadynamics: Theory and practice," Rev. Comput. Chem. **28**, 1–49 (2015).

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

[58]G. Lanaro and G. N. Patey, "Birth of NaCl crystals: Insights from molecular simulations," J. Phys. Chem. B **120**, 9076–9087 (2016).

[59]A. J. C. Ladd and L. V. Woodcock, "Interfacial and co-existence properties of the Lennard-Jones system at the triple point," Mol. Phys. **36**, 611–619 (1978).

[60]W. K. Burton, N. Cabrera, and F. C. Frank, "The growth of crystals and the equilibrium structure of their surfaces," Philos. Trans. R. Soc., A **243**, 299–358 (1951).

[61]P. Dandekar, Z. B. Kuvadia, and M. F. Doherty, "Engineering crystal morphology," Annu. Rev. Mater. Res. **43**, 359–386 (2013).

[62]M. N. Joswiak, M. F. Doherty, and B. Peters, "Size dependent surface free energy and Tolman corrected droplet nucleation of TIP4P/2005 water," J. Phys. Chem. Lett. **4**, 4267 (2013).

[63]B. Cheng, G. A. Tribello, and M. Ceriotti, "Solid-liquid interfacial free energy out of equilibrium," Phys. Rev. B **92**, 180102 (2015).