

***MULTIVARIATE ANALYSIS OF HOMOGENEOUS NUCLEATION  
RATE MEASUREMENTS: II. TEMPERATURE AND VAPOR  
CONCENTRATION DEPENDENCE***

R. McGraw

Submitted for publication in  
*Journal of Chemical Physics*

May 2007

*Environmental Sciences Department/Atmospheric Sciences Division*

**Brookhaven National Laboratory**

P.O. Box 5000  
Upton, NY 11973-5000  
[www.bnl.gov](http://www.bnl.gov)

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

# Multivariate analysis of homogeneous nucleation rate measurements: II. Temperature and vapor concentration dependence

Robert McGraw

Environmental Sciences Department, Atmospheric Sciences Division  
Brookhaven National Laboratory, Upton, NY 11973

## Abstract

The multivariate analysis of nucleation rate dependency on vapor concentration, initiated in Part I [McGraw and Zhang, 2007], is extended to include temperature. Nucleation rate sensitivity to changes in vapor concentration and temperature are described using kinetic extensions of the nucleation theorem (KNTs). The nucleation rate is found to be highly multi-linear in selected KNT-suggested temperature and concentration coordinates with little curvature over the tested range of measurements. Bivariate analyses of measurements for the homogeneous nucleation rates of water, methanol, hexanol, and nonane vapors are presented. Estimates for critical cluster energy, and number of molecules present in the critical nucleus are compared with predictions from the liquid drop model of classical nucleation theory. Surprisingly a simple Arrhenius temperature dependence provides an excellent description of the rate measurements for each of the studied cases. The quality of the planar fits to the full data sets are also excellent with  $R^2$  values ranging from 0.89 (methanol) to 0.97 (hexanol). These results yield a physically based parameterization of the measurements, valid over about five orders of magnitude in nucleation rate, as well as a measurement-based determination of critical cluster properties suitable for comparing with molecular simulations and phenomenological nucleation theories. Comparison with the capillary drop model of classical nucleation theory reveals generally good agreement for critical cluster size. Nevertheless, with the exception of water vapor, the drop model tends to grossly overestimate critical cluster energy.

## 1. Introduction

In [McGraw and Zhang, 2007], hereafter I, we introduced a general multivariate framework in which to apply the first nucleation theorem. This resulted in a new method for analyzing nucleation rate measurements and for studying the first-order rate sensitivities to changes in concentrations of the different condensable species present in a multicomponent vapor. A key feature of the method is that only directly measured quantities such as vapor species concentrations and the nucleation rate itself are used in the analysis. Thus the need to make estimates of the bulk thermodynamic properties of the nucleated phase (e.g., surface tension, density, and equilibrium vapor pressures of the various species present – all required for any analysis based on the capillary drop model of classical nucleation theory) is avoided. The method further yields the molecular content of the critical nucleus, thus providing an essentially model-free and molecular-level interpretation of nucleation mechanism, together with a compact but highly accurate parameterization for the nucleation rate.

This paper will extend the multivariate framework, applying both the first and second nucleation theorems so as to include temperature as well as concentration dependence. Remarkably, it is found that a simple Arrhenius temperature dependence provides an accurate description of the nucleation rate over the full range (circa 40°K) of a typical set of measurements. Although the analysis of I was limited to a fixed temperature, it was applied to a multicomponent vapor system: a ternary mixture of p-toluic acid, sulfuric acid, and water vapors. Characteristic of the complex mixed organic-inorganic systems thought to be important for atmospheric new particle formation, the bulk thermodynamic properties needed to apply classical nucleation theory are largely unavailable and other approaches, e.g., to interpretation of the aerosol chamber rate measurements that have made on these systems [Zhang et al., 2004], are required. This was our primary motivation in [I] for developing the new methods. The present paper follows a complementary tack by re-examining homogeneous nucleation rates in several simpler single-component vapor systems (water, methanol, hexanol, and nonane) for which measurements over a range of temperature and vapor concentration are available to test the extended theory. Because the bulk properties of these simpler systems have also

been measured, a comparison can be made (in Sec. 3) with the critical cluster properties, specifically size and energy, predicted by the liquid drop model of classical nucleation theory. Whereas the analysis of these simpler systems requires only a bivariate formulation of the theory – with temperature and vapor concentration as variables – this extension, over previous generally univariate application of nucleation theorems, already represents a significant advance.

For generality the temperature-extended theory is formulated in Sec. 2 in fully multivariate form in anticipation of future applications to more complex systems. In order to first eliminate any dependence on vapor saturation ratio,  $S$ , a necessary task in cases where equilibrium vapor pressures are unavailable, the second nucleation theorem (giving temperature sensitivity) is recast in terms of vapor concentration (Appendix). Both composition coordinates are utilized in the calculations of Sec. 3 and yield different temperature rate sensitivities. These are interpreted in terms of different reference states (vapor or bulk liquid) for the critical cluster energy. Contributions from higher-order rate sensitivities (higher-order nucleation theorems) are found to cause slight but systematic departure from Arrhenius behavior. These are examined statistically and also given a molecular level interpretation in terms of critical cluster properties. Section 4 concludes with a summary and discussion of results.

## **2. Kinetic extensions of the nucleation theorem for multi-component vapors: Selection of composition and temperature coordinates**

In a single component system the steady-state nucleation rate is given by the usual Becker-Döring summation over reciprocal cluster growth rates [e.g., Abraham, 1974]:

$$J_{BD} = \left( \sum_{g=1}^G \frac{1}{\beta_g n_g} \right)^{-1}. \quad (2.1)$$

Here  $\beta_g$  is the rate constant for monomer addition to a single cluster containing  $g$  monomer units:

$$\beta_g = \alpha_c \frac{\bar{c}_1}{4} s_g n_1 = \alpha_c \left( \frac{kT}{2\pi m_1} \right)^{1/2} s_g n_1 \quad (2.2)$$

$n_g$  is the constrained equilibrium concentration of  $g$ -clusters per unit volume, and  $n_1$  is the concentration of condensable monomer in the vapor phase.  $s_g$  is the surface area of a single  $g$ -cluster,  $\bar{c}_1$  is the mean molecular speed of a monomer of mass  $m_1$ , and  $\alpha_c$  is the accommodation coefficient. The summation in Eq. 2.1 is dominated by terms near the critical cluster size  $g^*$ , where  $n_g$  is smallest, and terminates at a size  $G$  sufficiently large that returns to pre-critical size can be neglected.

Differentiating Eq. 2.1 at constant temperature yields a form of the nucleation theorem (NT) first put forth by Ford [Ford, 1997]:

$$\left( \frac{\partial \ln J}{\partial \ln n_1} \right)_T = \left( \frac{\partial \ln J}{\partial \ln S_1} \right)_T = 1 + \bar{g} \approx 1 + g^* \quad (2.3)$$

The first equality shows equivalent results regardless of whether the concentration variable is expressed as  $n_1$  or in terms of the saturation ratio  $S_1 \equiv n_1 / n_1^{eq}$  where  $n_1^{eq} = n_1^{eq}(T)$  is the vapor concentration in equilibrium with bulk phase. The overbar denotes averaging over the distribution of reciprocal cluster growth rates appearing in the Becker- Döring sum:

$$\bar{g} \equiv \frac{\sum_{g=1}^G \frac{1}{\beta_g n_g} g}{\sum_{g=1}^G \frac{1}{\beta_g n_g}} = J_{BD} \sum_{g=1}^G \frac{g}{\beta_g n_g} \equiv \sum_{g=1}^G g P(g). \quad (2.4)$$

$P(g) \equiv J_{BD} / \beta_g n_g$  is the normalized distribution. The approximate equality of Eq. 2.3 is accurate when  $P(g)$  is symmetric and/or sharply peaked at a critical cluster size,  $g^*$ .

An exact expression is also available for the higher-order derivatives of  $J_{BD}$  [McGraw and Wu, 2003]. Using Eq. 2.1 and repeatedly differentiating at constant temperature it can be shown that:

$$\left( \frac{\partial^n \ln J_{BD}}{\partial (\ln n_1)^n} \right)_T = (-1)^{n+1} \kappa_n \quad (n \geq 2) \quad (2.5)$$

where  $\kappa_n$  is the  $n^{\text{th}}$  cumulant of  $P(g)$ . For example, for the second derivative,

$$\left( \frac{\partial^2 \ln J_{BD}}{\partial (\ln n_1)^2} \right)_T = \left( \frac{\partial \bar{g}}{\partial \ln n_1} \right)_T = -\kappa_2 \equiv -(\overline{g^2} - \bar{g}^2). \quad (2.6)$$

This result is similar to the Kelvin relation in that it gives the change in critical cluster size with respect to change in vapor concentration at constant  $T$ . To the extent that the variance of  $P(g)$ , equal to  $\kappa_2$ , is small, so too will be the change in critical cluster size - thus limiting departure from the local linearity implied by Eq. 2.3.

### 2.1 Temperature variation at constant vapor composition

Equations 2.3 and 2.5 comprise the kinetic extension of the nucleation theorem, which in its original and most common form is a purely thermodynamic relation:  $d(W^*/kT)/d \ln n_1 = -g^*$ , involving the reduced nucleation barrier height,  $W^*/kT$ , rather than  $\ln J$  [Kashchiev, 1982; 2006]. Extensions yielding the sensitivity of the barrier height to changes in temperature (the so-called ‘second nucleation theorem’) have also been developed [Oxtoby and Kashchiev, 1994; Ford, 1997]. The kinetic nucleation theorem (KNT) for the sensitivity of  $\ln J$  with respect to temperature at constant saturation ratio is [Ford, 1997; McGraw and Wu, 2003]:

$$\left( \frac{\partial \ln J}{\partial T} \right)_{s_1} = \frac{1}{2T} + \frac{\bar{E}_g - \bar{g} E_1^b}{kT^2} + \frac{E_1 - E_1^b}{kT^2} \approx \frac{E_{g^*} - g^* E_1^b}{kT^2} + \frac{E_1 - E_1^b}{kT^2} \quad (2.7)$$

where  $\bar{E}_g$  is the cluster energy averaged over the distribution  $P(g)$ , as in Eq. 2.4,

$$\bar{E}_g = \sum_g E_g P(g) \quad (2.8)$$

and  $E_1^b$  is the energy per molecule in the bulk liquid phase. The lead term on the right hand side of Eq. 2.7 results from differentiation of the mean speed of monomer in the vapor phase:  $d \ln c_1 / dT = 1/(2T)$ . This contribution is small over the temperature range of interest and will be neglected. Averages over  $P(g)$  have been replaced in the approximate equality by their most probable values as in Eq. 2.3.

The right-hand side of Eq. 2.7 was obtained by Ford using a Gaussian approximation for the distribution  $P(g)$  [Ford, 1997]. Other expressions of this useful relation are obtained on changing the variable held constant during differentiation. For example, Oxtoby and Kashchiev held free energy ( $=kT \ln S_1$ ) constant, yielding a result in terms of cluster entropy [Oxtoby and Kashchiev, 1994]. Noteworthy is that the right-hand-side of Eq. 2.7 contains bulk properties of the liquid phase,  $E_1^b$  and energy of vaporization, while the nucleation rate refers here to the rate of forming molecular clusters of critical size from the *vapor* phase. Reflection shows that these bulk quantities enter through the saturation ratio, which is held constant during differentiation. A modification of the KNT, useful when equilibrium vapor pressures are unavailable, is obtained holding vapor concentration,  $n_1$ , constant on the left-hand-side of Eq. 2.7 instead of  $S_1$ . Application of the chain rule (see the Appendix) yields:

$$\left( \frac{\partial \ln J}{\partial T} \right)_{n_1} = \frac{\bar{E}_g - \bar{g} E_1}{kT^2} \approx \frac{E_{g^*} - g^* E_1}{kT^2}. \quad (2.9)$$

Here  $E_1 \equiv E(1)$  is the monomer energy in the vapor phase. Bulk phase properties have cancelled, along with the vaporization energy, via the Clapreyron relation (Appendix):

$$\frac{d \ln n_1^{eq}}{dT} = \frac{E_1 - E_1^b}{kT^2}. \quad (2.10)$$

Equation 2.9 is readily extended to multi-component vapors and is the form of the second nucleation theorem used in much of the present study.

## 2.2 Extension to multi-component systems

Multi-component extension of Eq. 2.3 is complicated by the lack of simple closed-form expressions for the nucleation rate analogous to Eq. 2.1. Thus rigorous extensions of the KNT have been limited to special model cases. Following (I) we achieve extension here using a simple transition state model wherein the nucleation rate is equated to the barrier crossing rate, itself equal to the sum of the net fluxes contributed by each condensable vapor species to growth beyond the critical cluster size. For example, the contribution from species  $i$  has the form:

$$J_i = \kappa_i \beta_i n(g^*_1, g^*_2, \dots) \quad (2.11)$$

where  $g^*_j$  is the number of molecules of species  $j$  present in the critical nucleus, and  $n(g^*_1, g^*_2, \dots) \equiv n_{g^*_1, g^*_2, \dots}$  is the constrained equilibrium concentration of critical clusters. (To avoid complicated subscripts, parenthesis will henceforth be used to denote cluster composition when multiple species are present.)

$$\beta_i = \alpha_{c,i} \left( \frac{kT}{2\pi m_i} \right)^{1/2} s(g^*_1, g^*_2, \dots) n_i$$

is the accommodation rate of species  $i$  with the critical cluster of surface area  $s(g^*_1, g^*_2, \dots)$  where  $\alpha_{c,i}$  is the accommodation coefficient, generally size and species dependent.  $m_i$  and  $n_i \equiv [n_i]$ , are the monomer mass and monomer number concentration, respectively, for species  $i$ , and  $\kappa_i$  is a barrier transmission coefficient, assumed here to be constant, as is  $\alpha_{c,i}$ , along the growth coordinate of species  $i$ . For homogeneous nucleation of a single component,  $\kappa$  is the Zeldovich correction factor and Eq. 2.11 is the Becker-Döring-Zeldovich expression for the nucleation rate [Abraham, 1974]. Approximating the total nucleation rate  $J = \sum J_i$  we write the differential of  $\ln J$ :

$$d \ln J = \sum_i \frac{J_i}{J} d \ln J_i = \sum_i \delta_i d \ln J_i \quad (2.12)$$

where  $\delta_i \equiv J_i / J$  is the fraction of total flux along growth coordinate  $i$ . Because  $J_i \geq 0$  we obtain the following properties for the  $\{\delta_i\}$ :  $0 \leq \delta_i \leq 1$  and  $\sum \delta_i = 1$ .

Equations 2.11 and 2.12 yield the generalization of Eq. 2.3 that we seek (additional details of the derivation are described in I):

$$\left( \frac{\partial \ln J}{\partial \ln n_i} \right)_{T, \{n_j\}} = g^*_i + \delta_i \quad (2.13)$$

This last result suggests a local multi-linear expansion for  $\ln J$ , which on integration gives:

$$\ln J \approx \ln J_0 + \sum_i (g^*_i + \delta_i) (\ln n_i - \ln n_i^0) \quad (2.14a)$$

and equivalently on exponentiation

$$J = J_0 \prod_i \left( \frac{n_i}{n_i^0} \right)^{g_i^* + \delta_i} \quad (2.14b)$$

The “0” refers to a specific set of reference steady-state conditions, such as those at the centroid of an experimental data set expressed in nucleation theorem-motivated coordinates as described below.

Including temperature requires extending Eq. 2.9 to the multicomponent case. Consider the equilibrium constant for  $g_1$  and  $g_2$  molecules of species 1 and 2 to combine to form a  $(g_1, g_2)$  cluster in the vapor phase. For two components:

$$K_{eq}(T) = \frac{n(g_1, g_2)}{(n_1)^{g_1} (n_2)^{g_2}} . \quad (2.15)$$

The law of mass action – proportionality between vapor species activity and concentration – conveniently gives concentrations in place of activities on the right hand side. Application of the Gibbs-Helmholtz relation gives:

$$\frac{d \ln K_{eq}(T)}{dT} = \frac{E(g_1, g_2) - g_1 E_1 - g_2 E_2}{kT^2} = \left[ \frac{\partial \ln n(g_1, g_2)}{\partial T} \right]_{n_1, n_2} , \quad (2.16)$$

which reduces to the single component result (c.f. Eq. 3.7 below) for  $g_2 = 0$ . The binary extension of Eq. 2.9 follows Eq. 2.11 and the second equality of Eq. 2.16 evaluated at the critical cluster size. Further extension to systems having multiple components is straightforward and gives the following general result:

$$\left( \frac{\partial \ln J}{\partial T} \right)_{n_1, n_2, \dots} = \frac{E(g_1^*, g_2^*, \dots) - \sum_i g_i^* E_i}{kT^2} \equiv \frac{\Delta E(g_1^*, g_2^*, \dots)}{kT^2} . \quad (2.17)$$

Again, the small temperature dependence arising from the mean molecular speed has been neglected. Equation 2.17 is the multi-component extension of the second nucleation theorem for temperature dependence of the nucleation rate that we seek. As in Eq. 2.9, with vapor concentrations held fixed on differentiation the relevant  $\Delta E$  is the energy of critical cluster formation from the vapor phase.

The combination of Eqs 2.14 and 2.17 suggest a local multilinear expansion for  $\ln J$  in the nucleation theorem-motivated coordinates  $\{1/T, \ln[n_i]\}$ :

$$\begin{aligned}
\ln J &= \ln J_0 + \sum_i \left( \frac{\partial \ln J}{\partial \ln n_i} \right)_0 d(\ln n_i) + \left( \frac{\partial \ln J}{\partial T} \right)_0 dT \\
&= \ln J_0 + \sum_i (g_i^* + \delta_i) d(\ln n_i) + \frac{\Delta E(g_1^*, g_2^*, \dots)}{kT^2} dT \quad . \quad (2.18) \\
&= \ln J_0 + \sum_i (g_i^* + \delta_i) d(\ln n_i) - \frac{\Delta E(g_1^*, g_2^*, \dots)}{k} d\left(\frac{1}{T}\right)
\end{aligned}$$

Integration gives

$$\ln J \approx \ln J_0 + \sum_i (g_i^* + \delta_i) (\ln n_i - \ln n_i^0) - \frac{\Delta E(g_1^*, g_2^*, \dots)}{k} \left( \frac{1}{T} - \frac{1}{T_0} \right) \quad (2.19a)$$

and on exponentiation

$$J = J_0 \prod_i \left( \frac{n_i}{n_i^0} \right)^{g_i^* + \delta_i} \text{Exp} \left[ - \frac{\Delta E(g_1^*, g_2^*, \dots)}{k} \left( \frac{1}{T} - \frac{1}{T_0} \right) \right] \quad (2.19b)$$

showing the Arrhenius temperature dependence. Equation 2.19 provides the generalization of the results of I needed to include temperature dependence. It is expected to be valid throughout the range of temperature and vapor concentrations for which the coefficients  $g_i^* + \delta_i$  and  $\Delta E$  can be regarded constant. This local behavior should be contrasted with that of the full nucleation rate surface, which will be much more complicated as critical cluster size, composition, and energy all vary with larger changes in environmental conditions. The utility of Eq. 2.19 derives empirically from the observed strong propensity for linearity over a surprisingly wide range of temperature, vapor concentration and nucleation rate. This is demonstrated in the following section.

### 3. Calculations

At constant temperature ( $T = T_0$ ) Eq. 2.19 reduces to Eq. 2.14, the utility of which was demonstrated for analysis of several ternary nucleation systems in I and recently for the cis-pinonic acid/sulfuric acid/water system in [Zhang et al., 2007]. Ideally one would like to test Eq. 2.19 through comparisons with laboratory measurements of

nucleation rate that span a significant range of both temperature and multi-component vapor composition coordinates. In lieu of such comprehensive measurements, the comparisons will be made here using reported measurements for several different single component vapors. The temperature-dependent methods introduced in Sec. 2 are demonstrated first for the homogeneous nucleation of water vapor through comparisons with the nucleation pulse chamber measurements of Wölk and Strey [Wölk and Strey, 2001]. Measurements for methanol, hexanol, and nonane - systems that unlike water are known to depart strongly from the predictions of classical nucleation theory – are examined in Sec. 3.2.

### *3.1 Illustrative calculations for water vapor*

For single-component vapors, Eq. 2.19a reduces to:

$$\ln J \approx \ln J_0 + (g^* + 1)(\ln n_1 - \ln n_1^0) - \frac{\Delta E(g^*)}{k} \left( \frac{1}{T} - \frac{1}{T_0} \right). \quad (3.1)$$

Equation 3.1 is expected to be valid over a locally flat region of the nucleation rate surface near the reference condition. Its range of validity is determined by the curvature of the nucleation rate surface, an estimation of which requires extension to include quadratic terms in the selected coordinates. Theoretical analysis of curvature along the vapor concentration coordinate follows Eq. 2.6 and direct estimates from data analysis of measurements, along both the  $1/T$  and concentration coordinates, are presented below. Remarkably these higher-order terms appear to have little effect thus supporting the accuracy of Eq. 3.1 over the range of measurements.

Wölk and Strey provide extensive measurements of homogeneous nucleation in both  $H_2O$  and  $D_2O$  vapors [Wölk and Strey, 2001]. The present analysis is limited to the  $H_2O$  measurements. Originally reported in terms of the saturation ratio  $S = n_1 / n_1^{eq}$ , the measurements are converted here to vapor concentration ( $n_1$ ), using the equilibrium vapor pressure formula provided by the authors (Table 1 of [Wölk and Strey, 2001]). This will demonstrate the utility of working in concentration units, which is necessary in other more complex systems for which equilibrium vapor pressures are unavailable. The measurements are next expressed in KNT-motivated coordinates  $\{x_i, y_i, z_i\}$

where  $x = 1000/T$ ,  $y = \text{Log}_{10}[n_1, \text{ molecules cm}^{-3}]$ ,  $z = \text{Log}_{10}[J \text{ cm}^{-3} \text{ s}^{-1}]$  for  $i = 1, \dots, N$  where  $N = 343$  is the number of reported measurements.

In I we introduced principal component analysis (PCA) as a general framework for statistical analysis and interpretation of nucleation rate measurements based on the covariance matrix. PCA can be used with any number of coordinates and offers additional advantages for data analysis and compression in highly multivariate applications. For present purposes the linear minimum variance estimator (LMVE), or best planar fit to the dataset, is sufficient and although this is also computable through the covariance matrix (see I) it is most easily obtained using standard linear regression programs such as are available in *Mathematica* [Wolfram, 1999]. The result:

$$z_{LMVE} = -674.3 + 48.30x + 28.09y \quad (3.2)$$

provides the LMVE planar fit to the  $N$ -point water data set.

Figure 1 shows the reported measurements (points), which fall into five classes clustering near the indicated temperatures in the figure. The parallel lines are projections from the LMVE plane, Eq. 3.2, onto the  $(y,z)$  plane for the five constant  $x$ -values evaluated at the mean temperature for each grouping of points. These are, from left to right in the figure: 218.94, 229.52, 239.58, 249.75, and 259.87 degrees K. Within each grouping the standard deviation in temperature is of order only 0.05K, justifying assignment of a single, mean value, to  $x$  for each grouping. The LMVE description agrees remarkably well with the data, except at the lowest temperature (near 220K) where there appears to be some curvature to the measurements that is missed by the planar fit. An estimate of quality of fit is given by  $R^2$ :

$$R^2 = 1 - \frac{\sum_{i=1}^N (z_i - z)^2}{\sum_{i=1}^N (z_i - \mu_z)^2}, \quad (3.3)$$

with  $R^2 = 0.94$  for the water dataset. Here  $z \equiv z(x_i, y_i)$ , from Eq. 3.2,  $\mu_z$  is the mean value of  $z$ , and  $N = 343$  is the number of measurements in the dataset.

Equation 3.2 gives linear minimum variance estimates for the coefficients of  $x$  and  $y$  (equal to 48.30 and 28.09, respectively) in terms of which physical properties of the critical cluster can be inferred from the nucleation theorems. Thus the  $y$ -coefficient gives  $g^* \approx 27$  for the number of molecules in the critical nucleus (c.f. Eq. 3.1). The cluster energy, relative to the vapor is calculated as follows:

$$\left(\frac{\partial \ln J}{\partial T}\right)_{n_1} = -2.303 \times \frac{1000}{T^2} \times \left(\frac{\partial \text{Log}_{10} J}{\partial (1000/T)}\right)_{n_1} = -\frac{2303}{T^2} \times \left(\frac{\partial z}{\partial x}\right)_y = \frac{\Delta E(g^*)}{kT^2} \quad (3.4)$$

where the second equality includes the coefficient of  $x$  from the regression analysis and the last equality follows Eq. 3.1. Evaluation using the numerical value of the  $x$ -coefficient from Eq. 3.2 gives

$$-\frac{\Delta E^*}{kT_0} \equiv \frac{g^* E_1 - E(g^*)}{kT_0} = 464.52 \quad (3.5)$$

at the centroid temperature,  $T_0 = 1000/\bar{x} = 239.5K$ , indicating that the cluster energy is lower than that of the vapor by this amount.

Uncertainty in the regression coefficients of the LMVE (Eq. 3.2) implies essentially proportional uncertainty in corresponding estimates of critical cluster size and energy relative to the vapor. The regression coefficients are characterized by confidence intervals that depend on the size and quality of the data set and the quality of the model fit. These can be generated using the standard regression package in *Mathematica* [Wolfram, 1999]. The horizontal and vertical dashed lines of Fig. 2 show the 95% confidence intervals for each coefficient computed independently. The corresponding joint confidence interval has a highly eccentric elliptical shape indicating that the uncertainties in the  $x$ - and  $y$ -coefficient estimators are highly correlated. The coefficients appearing in Eq. 3.2 are the most probable values located at the center of the ellipse as indicated by the dot in the figure.

Figure 3 shows an Arrhenius plot of the measurements where  $K_{eq}$  is the equilibrium constant of Eq. 2.15 evaluated at  $g = g^*$  and specialized to the case of a single component:

$$K_{eq} = \frac{n_{g^*}}{(n_1)^{g^*}} = \frac{J}{\kappa \alpha_c \left( \frac{kT}{2\pi m_1} \right)^{1/2} s_{g^*} (n_1)^{g^*+1}} \quad (3.6)$$

The last equality follows Eq. 2.11, again specialized to the case of a single component. Kinetic factors appearing in Eq. 3.6 are eliminated by plotting the ratio  $K_{eq} / K_{eq}^0$  where  $K_{eq}^0$  is the equilibrium constant evaluated under reference temperature and concentration conditions. The slope gives the cluster energy relative to vapor according to the Gibbs-Helmholtz relation:

$$\frac{d \ln K_{eq}(T)}{dT} = \frac{E(g^*) - g^* E_1}{kT^2}. \quad (3.7)$$

This is found to have the same value (equal to 48.30) as the  $x$ -coefficient in Eq. 3.2, previously obtained from the bi-variate regression - thus yielding an energy of formation identical to Eq. 3.5. The high degree of linearity seen along the  $1/T$  coordinate (the line shown in the figure is the linear least squares fit to the data) is further analyzed in the residual plot shown in the lower panel of the figure. Here the regression line is subtracted from ordinate value of each measurement and plotted as a residual. The mean values for each temperature grouping are then fit with a quadratic yielding the dashed parabolic curve shown in the figure. The degree of curvature provides an estimate of the difference in heat capacity between the vapor and cluster states – the larger the difference in heat capacity, the greater the curvature expected in an Arrhenius plot.

For comparison with results obtained in the following section it is useful to obtain an estimate of the energy of cluster formation relative to bulk liquid using Eq. 3.5 and the energy of vaporization. The energy of vaporization per molecule of water is obtained from the vapor pressure equation using Eq. 2.10. Fitting values from the vapor pressure formula provided in Table I of [Wölk and Strey, 2001] over the temperature range of the measurements gives:

$$\frac{E_1 - E_1^b}{kT_0} = 22.43. \quad (3.8)$$

Multiplying this result by  $g^*$  and combining Eqs. 3.5 and 3.8 yields the following estimate for the critical cluster energy relative to the bulk liquid phase:

$$\frac{E(g^*) - g^* E_1^b}{kT_0} = g^* \left( \frac{E_1 - E_1^b}{kT_0} \right) - \left( \frac{g^* E_1 - E(g^*)}{kT_0} \right) = 143.02. \quad (3.9)$$

The energy scale represented by Eqs. 3.7-3.9 is depicted in Fig. 4.

The law of mass action, implicit in Eq. 3.6, results in elimination of concentration dependence, making the ordinate of Fig. 3 a function of temperature alone. In similar fashion it is possible to eliminate temperature dependence to obtain a temperature-scaled form of the nucleation rate that is a function of vapor concentration alone. Consider the following scaled nucleation rate defined so as to eliminate the Arrhenius temperature dependence in the ratio  $J/J_0$ :

$$\ln \left( \frac{J}{J_0} \right)_{scaled} \equiv \ln \left( \frac{J}{J_0} \right) + \frac{\Delta E^*}{k} \left( \frac{1}{T} - \frac{1}{T_0} \right) = (g^* + 1) \ln \left( \frac{n_1}{n_1^0} \right). \quad (3.10)$$

To the extent that the temperature dependence follows the Arrhenius form it has been eliminated, the resulting scaled-rate ratio depending only on vapor composition. From this one can estimate curvature and change in critical cluster size with vapor concentration.

Figure 5 shows evaluation of the middle terms of Eq. 3.10 for the water vapor measurements using the already established value of  $\Delta E^*/k$  from Eq. 3.5. The slope of the regression line ( $= 28.09$ ), shown in the top panel of the figure, is the same as the coefficient of  $y$  in the LMVE planar regression of Eq. 3.2 and thus gives the same value for  $g^* + 1$ . The lower panel in the figure shows the residuals, defined as the differences (log scale) between the data point ordinates and the regression line. The dashed curve is a quadratic fit in  $\text{Log}_{10}(n_1/n_1^0)$  to the residuals. This predicts a small systematic change in critical cluster size from about 27.3 (at 220K) to 29.0 (at 260K), which is close to the range of  $y$ -coordinate values seen in the joint confidence region of Fig. 2. Comparing the lower panels of Figs. 3 and 5 one sees, noting the reversal of temperature order between the two figures, a negative correlation between cluster size and energy of formation

relative to the vapor, i.e., the larger the cluster the more negative its energy of formation. This negative correlation is also reflected in the orientation of the joint confidence region of Fig. 2, which shows the positive correlation between the two bi-variate regression coefficients.

### ***3.2 Comparing estimated critical cluster properties for several vapors with predictions based on the capillary drop model of classical nucleation theory***

A similar analysis was carried out for methanol, hexanol, and nonane, and repeated for water, using  $S$  instead of  $n$  to describe vapor concentration. As these quantities are related through the vapor pressure equation, similar qualities of fit are expected to the extent that the vapor pressure itself follows the Arrhenius form. The model equation in these coordinates is:

$$\ln J = a + b(1000/T) + c \ln S. \quad (3.11)$$

Bi-variate fits based on Eq. 3.11 were tested for each of four single-component vapors with results summarized in Fig. 6 and Table 1. Comparison measurements are from [Wölk and Strey, 2001] for water vapor, [Strey et al., 1986] for methanol and n-hexanol, and [Wagner and Strey, 1984] for nonane. Figure 6 shows excellent quality of fits, with  $R^2$  values ranging from a low of 0.89 for methanol to a high of 0.97 for n-hexanol. No systematic departure from Eq. 3.11 is observed. The first row of Table 1 gives the number of measurement points. Rows 2 and 3 give the temperature,  $T_0$ , and saturation ratio,  $S_0$ , at the centroid of each set of measurements. The next four rows give the bi-variate regression coefficients  $a$ ,  $b$ , and  $c$  appearing in Eq. 3.11, and the quality of fit.

Unlike with the mixed ternary systems studied in I, there is sufficient thermodynamic data available here that comparisons can be made with the capillary drop model of classical nucleation theory (CNT). For methanol and hexanol, the thermodynamic measurements needed for the comparison are available as follows: saturation vapor pressures, from [Schmeling and Strey, 1983], surface tensions [Strey and Schmeling, 1983], and liquid densities [Yaws, 2003]. For water and nonane, vapor pressures, surface tensions and liquid densities are available from [Wölk, and Strey, 2001] and [Wagner and Strey, 1984] respectively.

Comparisons are made first for the critical cluster size, given in the drop model by the Kelvin relation:

$$g^*_{CNT} = \frac{32\pi}{3\rho_l^2} \frac{\sigma_\infty^3}{(kT \ln S)^3} \quad (3.12)$$

where  $\sigma_\infty$  is the bulk interfacial tension for a flat surface and  $\rho_l$  ( $cm^{-3}$ ) is the bulk liquid density. Equation 3.12 was evaluated at the data centroid conditions  $T = T_0$  and  $S = S_0$  for each set of measurements with the results shown in Table 1. Notice the favorable comparison between  $g^* = c - 1$  from the bi-variate planar fit (Eq. 3.11) and from the Kelvin relation in all cases, except for methanol where the critical cluster size is significantly underestimated by the Kelvin relation.

Similar comparisons are made for the critical cluster energy. With  $\ln S$  as independent variable it is natural to compare cluster energy relative to bulk liquid. This requires first subtraction of the energy of vaporization of a single molecule as obtained from the vapor pressure following Eq. 2.10 (c.f., Eq. 2.7). Single-molecule vaporization energies, in reduced form  $\Delta E_1 / kT_0 = (E_1 - E_1^b) / kT_0$ , are included in Table 1. Next, the  $b$  values (row 5, Table 1) are used to evaluate the left-hand-side of Eq. 2.7 to obtain the energy difference of interest:  $\Delta E_b^* / kT_0 = (E(g^*) - g^* E_1^b) / kT_0$ . These are listed in the next-to-last row of Table 1. Note that the estimate  $\Delta E_b^* / kT_0 = 141.40$  for water is in good agreement with the (albeit not completely independent) energy cycle estimate,  $\Delta E_b^* / kT_0 = 143.02$  from Fig. 4. Critical cluster energy is defined in the capillary drop model relative to an amount of the bulk liquid having the same number of molecules as the nucleus. It is thus a pure surface property. The following equation applies at the equimolecular dividing surface for which the surface excess density  $\Gamma = 0$  [Rowlinson and Widom, 2002]:

$$\frac{\Delta E_{CNT}^*}{kT} \equiv \frac{E_{CNT}(g^*) - g^* E_1^b}{kT} = \frac{s_{g^*}}{kT} \left( \sigma_\infty - T \frac{d\sigma_\infty}{dT} \right). \quad (3.13)$$

For water, the right hand side of Eq. 3.13 evaluates to  $\Delta E_{CNT}^* / kT = 140.97$  at  $T = T_0$  and values for all four vapors are given in the last row of Table 1.  $s_{g^*}$  is the cluster

surface area, now assuming the spherical shape, bulk liquid density, and abrupt interface that are the hallmarks of the capillarity approximation. Consistent with these properties, the radius at the equimolecular dividing surface equals the drop radius of the classical theory and the effect of surface curvature on interfacial tension is neglected [McGraw and Laaksonen, 1997]. The second equality of Eq. 3.13 can be seen as follows:  $s_{g^*}\sigma_\infty$  is the Helmholtz free energy, or reversible work, needed to fashion a drop from the bulk liquid [Reiss, 1996]. Similarly, the term in parenthesis is the interfacial energy per unit area, and neglecting change in surface area from thermal expansion, the middle expression gives the reduced droplet energy relative to the bulk liquid for a drop of critical size. The numerical values included in the last row of Table 1 were obtained taking the  $g^*$  values for use in Eq. 3.13 from the regression ( $g^* = c - 1$ ), rather than from  $g_{CNT}^*$ , so as to eliminate size as a variable when comparing entries in the last two rows of Table 1. With the exception of water, for which agreement is excellent, the drop model tends to grossly overestimate the critical cluster energy.

Additional strong support for the Arrhenius temperature dependence is found when Eq. 3.11 is examined vis a vis classical nucleation theory. It has been noticed, even since the earliest days of reliable rate measurements, that experimental nucleation rates, although often orders of magnitude different from classical theory prediction tend to differ by only a temperature-dependent amount. Specifically it has been found that the logarithm of the measured-to-predicted rate ratio tends to be linear in  $1/T$  [Strey et al., 1986]:

$$\ln\left(\frac{J}{J_{CNT}}\right) \approx a' + \frac{b'}{T} \quad (3.14)$$

where  $a'$  and  $b'$  are constants. With some reasonable restrictions on  $J_{CNT}$  it can be shown that  $J$  from Eq. 3.11 also satisfies this relation. The only significant restriction is that the surface tension be linear in temperature:  $\sigma_\infty(T) \approx A + BT$ , which form reduces the right-hand side of Eq. 3.13 to  $s_{g^*}A/kT$ , thereby conveying Arrhenius temperature dependence to  $J_{CNT}$ . (A linear fit to the surface tension data for liquids is generally quite successful over a wide range of temperatures below the region of interest for critical

phenomena [Hale, 1986].) Less significant, but worth noting, is the fact that  $J_{CNT}$  doesn't quite cancel the  $S$  dependence in Eq. 3.11. This is due to the well-known failure of classical theory to satisfy mass action and the first nucleation theorem – for example, substitution of  $J_{CNT}$  into Eq. 2.3 yields  $g^*+2$  on the right-hand side instead of the correct result,  $g^*+1$ . As shown by Courtney [Courtney, 1961],  $J_{CNT}$  requires only a (negligibly small)  $1/S$  correction factor to get things right – and complete cancellation of  $S$  from Eq. 3.14 results. Thus, within the very reasonable approximation of linear temperature dependence for the surface tension in  $J_{CNT}$ , the  $J$  from Eq. 3.11 can well be expected to satisfy Eq. 3.14. This establishes an important property of  $J$  from Eq. 3.11 that appears to be shared by measured nucleation rates.

#### **4. Summary and discussion:**

The main theoretical development of this series leads directly to the multi-component and temperature-dependent expression for nucleation rate given by Eq. 2.19. This result has a number of applications that include development of physically-based parameterizations likely to be useful for modeling nucleation in the atmosphere and elsewhere, interpolation of sparse data sets, and interpretation of nucleation measurements in terms of molecular pathways and critical cluster properties. In principle one can now obtain a linearized approximation to the local rate surface using just  $n+2$  measurements for an  $n$ -component system with temperature dependence. Measurement noise and uncertainty will generally necessitate more measurements, but these are not restricted to controlled constant temperature or constant concentration conditions. Indeed the capability for flexible data acquisition is likely to become one of the most useful features of the multivariate analysis especially for analysis of field measurements and in other applications where there is little control over environmental conditions. All of these advances have potential for extension to highly multivariate systems using modern statistical approaches that include regression and principal components analysis for dimensional reduction and automated coordinate selection – i.e., automatically identifying the most important gas-phase species and/or combinations of species participating in the nucleation process.

A variety of techniques are available for homogeneous nucleation measurement, but each of these is typically limited in coverage to a range of 3 - 5 orders-of-magnitude in nucleation rate [Iland et al., 2004]. For the diversity of cases examined here and in I, the multi-linear rate approximation (Eq. 2.19) was found to be highly accurate over the measurement range. Higher-order terms, beginning with the quadratic, will be needed to describe the curvature seen in similar log rate plots when the data from several methods are combined. Significant curvature is seen, for example, in Fig. 9 of Kim et al. [Kim et al., 2004], which combines measurements for D<sub>2</sub>O from the nucleation pulse chamber (rate coverage from about  $10^4$  to  $10^9$   $cm^{-3}s^{-1}$ ) and from a supersonic nozzle (rate coverage from about  $10^{16}$  to  $10^{18}$   $cm^{-3}s^{-1}$ ). Nevertheless, in the present more limited context of representing the data from a single measurement source, it takes careful analysis of residuals to uncover any systematic departure from the linearized rate expressions of Sec. 2. In I we examined measurements using an aerosol chamber for which rates covering a range of  $10^{-1}$  to  $10^4$   $cm^{-3}s^{-1}$  were reported and here we examined the nucleation pulse chamber measurements of Strey and co-workers with coverage from about  $10^5$  to  $10^{10}$   $cm^{-3}s^{-1}$ . A good fit to multivariate nucleation data over the range of the aerosol chamber measurements is probably sufficient for most applications to atmospheric particle formation, and this now seem feasible using the linearized expressions of Sec. 2 without significant correction for curvature required.

Although a local Arrhenius temperature dependence for the nucleation rate is expected from the second nucleation theorem, its persistence over the full temperature range of the studied measurements, about 40 degrees Kelvin, is surprising and suggestive of a small difference in heat capacity between the critical cluster and its dissociated vapor. As noted in Sec. 3, Strey et al. [Strey et al., 1986] found a similar - linear in  $1/T$  - dependence for  $Log[J/J_{CNT}]$  on comparing their experimental measurements with classical nucleation theory, but did not apply this function to the rate itself. In this case the slope gives the *difference in energy* between the critical cluster and its model representation in classical nucleation theory as a capillary drop having bulk properties. In a study of nucleation in condensed phase systems, the saturation vapor pressure of water over supercooled liquid at the homogeneous ice nucleation threshold, and at the

efflorescence point for heterogeneous nucleation of ammonium sulfate on small particles of calcium carbonate dispersed in a supersaturated aqueous solution of ammonium sulfate, were each found to follow an Arrhenius temperature dependence. In these last two cases the measurements were of nucleation threshold conditions and not of nucleation rate [Onasch et al., 2000].

The linear in  $1/T$  behavior of  $\text{Log}[J/J_{CNT}]$  is a special case of a more general scaling result for the nucleation rate derived from the nucleation theorem and Kelvin relation (Eq. 3.12). According to the general scaling theory  $\text{Log}[J/J_{CNT}] \approx D(T)$ , where  $D(T)$  is *some* function of  $T$ , if and only if the critical cluster size satisfies the Kelvin relation [McGraw and Laaksonen, 1997]. In this case, because  $J_{CNT}$  automatically satisfies the Kelvin relation, both  $J$  and  $J_{CNT}$  should have the same critical cluster size. Any discrepancy between  $J$  and  $J_{CNT}$  is more likely due to a difference in cluster energy. This scaling is supported here by the generally better agreement between  $g^*$  and  $g_{CNT}^*$  than between  $\Delta E_b^*$  and  $\Delta E_{CNT}^*$  found in Table 1. Finally, the specific (linear in  $1/T$ ) form for  $D(T)$  was shown to be consistent with the Arrhenius temperature dependence of the parameterized rate expressions developed in the present study. These connections will be further explored in future studies.

### **Acknowledgements.**

This research was supported by the DOE Atmospheric Sciences Program. The authors thank Profs. Renyi Zhang and Dr. Stephen E. Schwartz for their insightful comments. This manuscript has been authorized by Brookhaven Science Associates, LLC under Contract NO. DE-AC02-98CH1-886 with the U. S. Department of Energy.

### **References:**

- F. F. Abraham, *Homogeneous Nucleation Theory* (Academic, New York, 1974).
- W. G. Courtney, *J. Chem. Phys.* **35**, 2249 (1961).
- K. I. Diamantaras, and S. Y. Kung, *Principal component neural networks: Theory and applications*, (Wiley, New York, 1996).

- I. J. Ford, J. Chem. Phys. **105**, 8324 (1996); J. Phys. Chem. B **105**, 11649 (2001).
- I. J. Ford, Phys. Rev. E **56**, 5615 (1997).
- B. N. Hale, Phys. Rev. A **33**, 4156 (1986).
- K. Iland, J. Wederkind, J. Wölk, P.E. Wagner, and R. Strey, J. Chem. Phys. 121, 12259 (2004).
- A. Jaeger-Voirol and P. Mirabel, Atmospheric Environment **21**, 2053 (1989).
- V. I. Kalikmanov and M. E. H. van Dongen, Phys. Rev. E **51**, 4391 (1995).
- D. Kashchiev, J. Chem. Phys. **76**, 5098 (1982).
- D. Kashchiev, J. Chem. Phys. **125**, 014502 (2006).
- Y. J. Kim, B. E. Wyslouzil, G. Wilemski, J. Wölk, and R. Strey, J. Phys. Chem. A **108**, 4365 (2004).
- P. Korhonen, M. Kulmala, A. Laaksonen, Y. Viisanen, R. McGraw, and J. H. Seinfeld, J. Geophysical Res. **104**, 26349 (1999).
- M. Kulmala and Y. Viisanen, J. Aerosol Sci. **22**, Supplement 1, S97-100 (1991).
- R. McGraw and W. H. Marlow, J. Chem. Phys. **78**, 2542 (1983).
- R. McGraw and A. Laaksonen, J. Chem. Phys. **106**, 5284 (1997).
- R. McGraw and D. Wu, J. Chem. Phys., **118**, 9337 (2003).
- P. H. McMurry, M. Fink, H. Sakurai, M. R. Stolzenburg, R. L. Mauldin III, J. Smith, F. Eisele, K. Moore, S. Sjostedt, D. Tanner, L. G. Huey, J.B. Nowak, E. Edgerton, and D. Voisin, J. Geophys. Res. 110, D22S02,doi:10.1029/2005JD005901 (2005).
- I. Napari, M. Noppel, H. Vehkamäki, and M. Kulmala, Journal of Geophysical Research D19, **107**, 4381 (2002).
- T. B. Onasch, R. McGraw and D. Imre, J. Phys. Chem. A **104**, 10797 (2000).
- D. W. Oxtoby and D. Kashchiev, J. Chem. Phys. **100**, 7665 (1994).
- H. Reiss, Methods of Thermodynamics, (Dover, New York, 1996).
- J. S. Rowlinson and B. Widom, Molecular theory of capillarity, (Dover, New York, 2002) pg. 36.

- T. Schmeling and R. Strey, *Ber. Bunsenges. Phys. Chem.* **87**, 871 (1983).
- W. J. Shugard, R. H. Heist, and H. Reiss, *J. Chem. Phys.* **61**, 5298 (1974).
- R. Strey and T. Schmeling, *Ber. Bunsenges. Phys. Chem.* **87**, 324 (1983).
- R. Strey, P. E. Wagner, and T. Schmeling, *J. Chem. Phys.* **84**, 2325 (1986).
- H. Vehkamäki, M. Kulmala, I. Napari, K.E.J. Lehtinen, C. Timmreck, M. Noppel, and A. Laaksonen, *Journal of Geophysical Research D22*, **107**, 4622 (2002).
- Y. Viisanen, R. Strey, and H. Reiss, *J. Chem. Phys.* **99**, 4680 (1993).
- P. E. Wagner and R. Strey, *J. Chem. Phys.* **80**, 5266 (1984).
- S. Wolfram, *The Mathematica Book* (4<sup>th</sup> ed.), (Wolfram Media/Cambridge University Press, 1999), pg 106.
- J. Wölk, and R. Strey, *J. Phys. Chem. B* **105**, 11683 (2001).
- J. Wölk, R. Strey, C. H. Heath, and B. E. Wyslouzil, *J. Chem. Phys.* **117**, 4954 (2002).
- C. L. Yaws, *Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds* (Knovel, 2003).
- R. Zhang, I. Suh, J. Zhao, D. Zhang, E. C. Fortner, X. Tie, L. T. Molina, and M. J. Molina, *Science* **304**, 1487 (2004).
- R. Zhang, R. McGraw, J. Zhao, I. Suh, A. Kholizov, L. T. Molina, and M. J. Molina, In preparation (2007).

## Appendix: Derivation of Eq. 2.9

Consider the following version of the “second nucleation theorem” (Ford, 1997; McGraw and Wu, 2003):

$$\left(\frac{\partial \ln J}{\partial T}\right)_S \approx \frac{E(g^*) - g^* E_1^b}{kT^2} + \frac{E_1 - E_1^b}{kT^2}, \quad (\text{A1})$$

which is Eq. 2.7. Evaluating the left hand side of A1 at constant number concentration, instead of constant  $S$ , gives Eq. 2.9. The derivation follows the chain rule using the identities:

$$\begin{aligned} x &= T \\ y &= \ln S = \ln(n_1) - \ln(n_1^{eq}) \\ z &= \ln J \\ w &= \ln(n_1) \end{aligned} \quad (\text{A2})$$

we seek  $(\partial z / \partial x)_w$  where  $w = w(x, y)$ . The chain rule gives the following well-known identity:

$$\left(\frac{\partial z}{\partial x}\right)_w = \left(\frac{\partial z}{\partial x}\right)_y + \left(\frac{\partial z}{\partial y}\right)_x \left(\frac{\partial y}{\partial x}\right)_w \quad (\text{A3})$$

with

$$\left(\frac{\partial y}{\partial x}\right)_w = -\frac{\left(\frac{\partial w}{\partial x}\right)_y}{\left(\frac{\partial w}{\partial y}\right)_x}. \quad (\text{A4})$$

Substitution of A2 into A4 and that result into A3 gives the result:

$$\begin{aligned}
\left(\frac{\partial \ln J}{\partial T}\right)_{n_1} &= \left(\frac{\partial \ln J}{\partial T}\right)_{\ln S} - \left(\frac{\partial \ln J}{\partial \ln[n_1]}\right)_T \left(\frac{\partial \ln[n_1^{eq}]}{\partial T}\right)_{n_1} \\
&\approx \left(\frac{\partial \ln J}{\partial T}\right)_{\ln S} - (g^* + 1) \left(\frac{E_1 - E_1^b}{kT^2}\right) \approx \left(\frac{E(g^*) - g^* E_1^b}{kT^2} + \frac{E_1 - E_1^b}{kT^2}\right) \\
&\quad - \frac{g^* E_1 - g^* E_1^b}{kT^2} - \frac{E_1 - E_1^b}{kT^2} = \frac{E(g^*) - g^* E_1}{kT^2}
\end{aligned} \tag{A5}$$

which is Eq. 2.9. The second equality uses the isothermal nucleation theorem and the temperature derivative of the equilibrium vapor density, which follows the Clapeyron equation. At constant vapor density the latter is given as a difference in monomer energy between the vapor and bulk phases, rather than the more familiar difference in enthalpy encountered for the constant pressure case. Bulk information is present in the partial derivative taken at constant  $S$ , the determination of which requires knowledge of the equilibrium vapor pressure over a bulk sample of the daughter phase. This is reflected in the appearance of the energy of vaporization and of the cluster energy relative to the monomer energy in the bulk phase. On the other hand, no bulk property is required to specify the vapor number concentration,  $n_1$ , and none enters into the final result wherein the cluster energy is given relative to the energy of the vapor from which it is formed. Only quantities directly measured in a nucleation experiment appear on the left-hand-side of Eq. A5.

An alternative derivation, not requiring the second nucleation theorem, follows the steps leading up to Eq. 2.17. This makes use of the equilibrium constant for critical cluster formation, Eq. 2.15, the Gibbs-Helmholtz relation, Eq. 2.16, and the transition-state model for  $J$ , Eq. 2.11.

## Figure captions

Figure 1. Planar fit to the water nucleation data set of Wölk and Strey (WS). The points represent the homogeneous nucleation rate data plotted in here in terms of vapor concentration, as computed from the reported saturation values using the vapor pressure formula given in Table 1 of WS. The lines are projections of the linear minimum variance plane (Eq. 3.2) onto the y-z plane of the figure. The  $R^2$  quality of fit is 0.94.

Figure 2. Joint 95% confidence intervals for the coefficients of x and y in Eq. 3.2. Dashed lines give the 95% confidence intervals for the x and y coefficients separately. The central point indicates the linear minimum variance estimator coefficients used in the planar fit of Fig. 1 and for estimation of critical cluster properties in Sec. 3.1 and Fig. 4.

Figure 3. Arrhenius plot for water measurements. Top: logarithm of the equilibrium constant for cluster formation relative to its value at the reference condition. The slope gives the energy of critical cluster formation. Bottom: residuals are defined as the differences (log scale) between the data point ordinates and the regression line. The dashed curve is a quadratic fit in  $1/T$  to the mean residuals for each temperature grouping of measurements.

Figure 4. Critical cluster energy scale for water relative to the bulk liquid and vapor phases at the reference temperature,  $T_0 = 239.5$ . The numbers give energy differences in units of  $kT_0$  for the transitions shown. Calculation is for  $g^* = 27.09$ , which is the best-fit estimate of the critical cluster size.

Figure 5. Arrhenius-scaled J plot for water measurements. Top: logarithm of scaled J relative to its value at the reference condition. The slope gives the number (plus 1) of molecules present in the critical nucleus. Measurements tend to cluster according to temperature at the values indicated in the figure. Bottom: residuals are defined as the differences (log scale) between the data point ordinates and the regression line. The dashed curve is a quadratic fit in  $\text{Log}_{10}(n_1/n_1^0)$  to the residuals set.

Figure 6. Comparison of modeled and measured nucleation rates.  $R^2$  is a measure of the quality of fit as defined by Eq. 3.3. See Table 1 for model rate function and fit parameters.

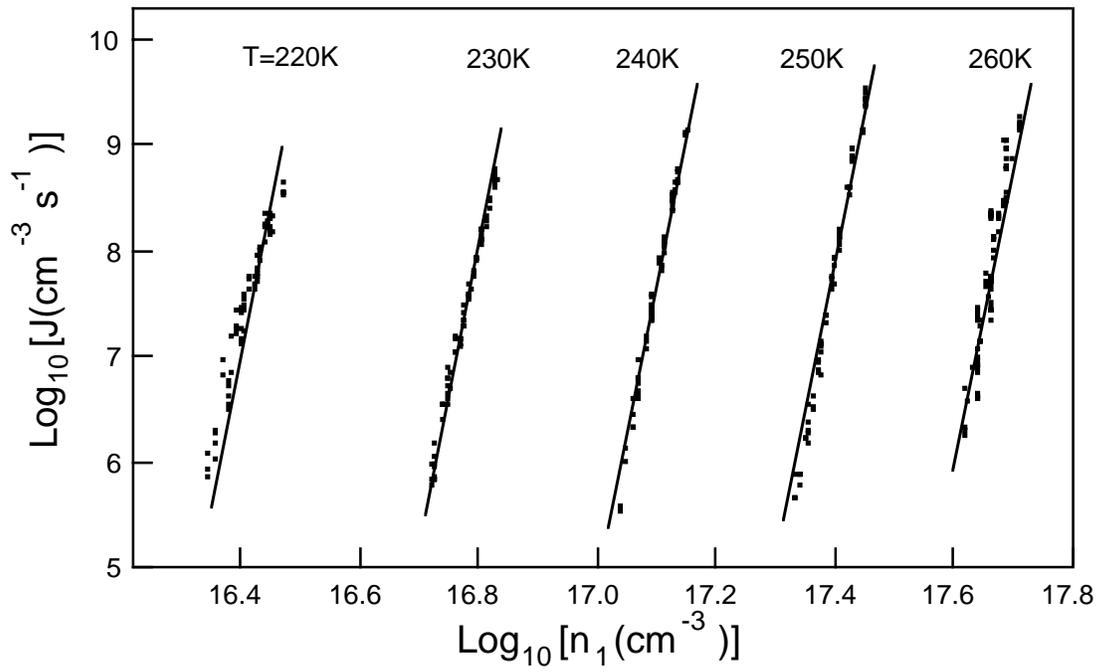


Figure 1

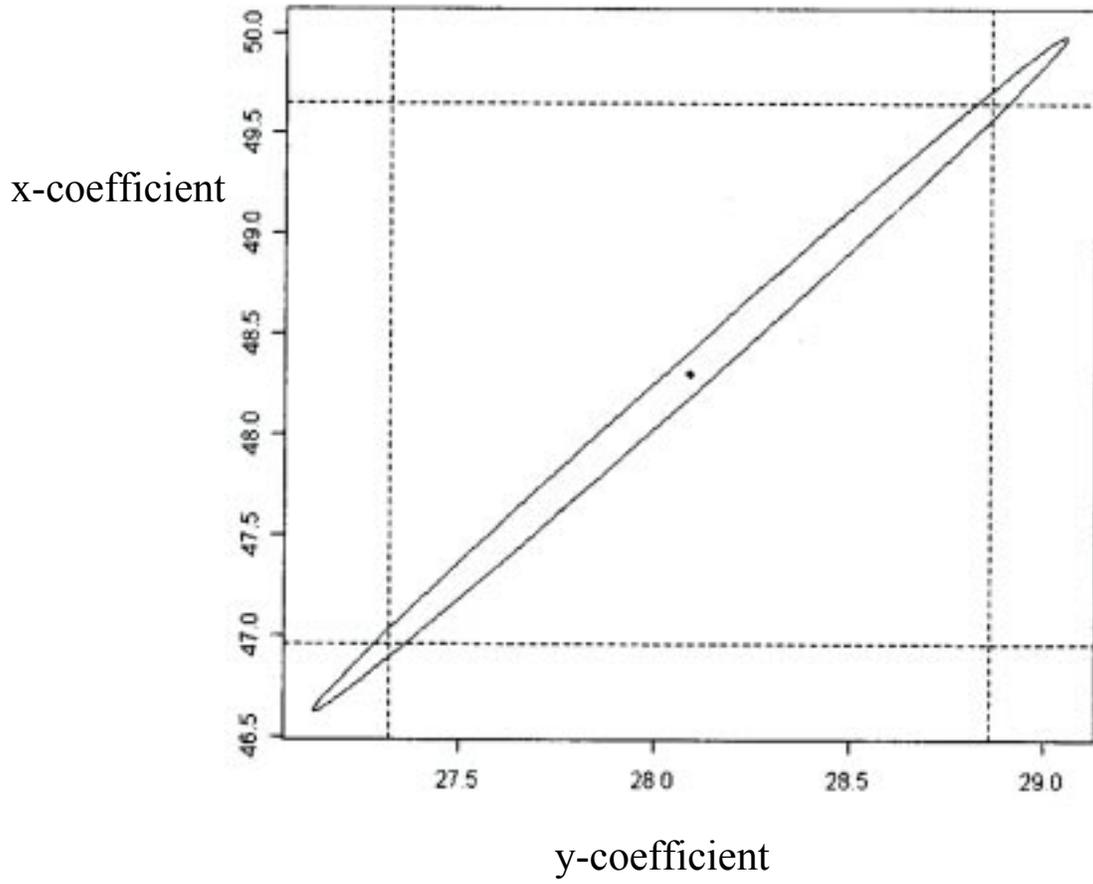


Figure 2

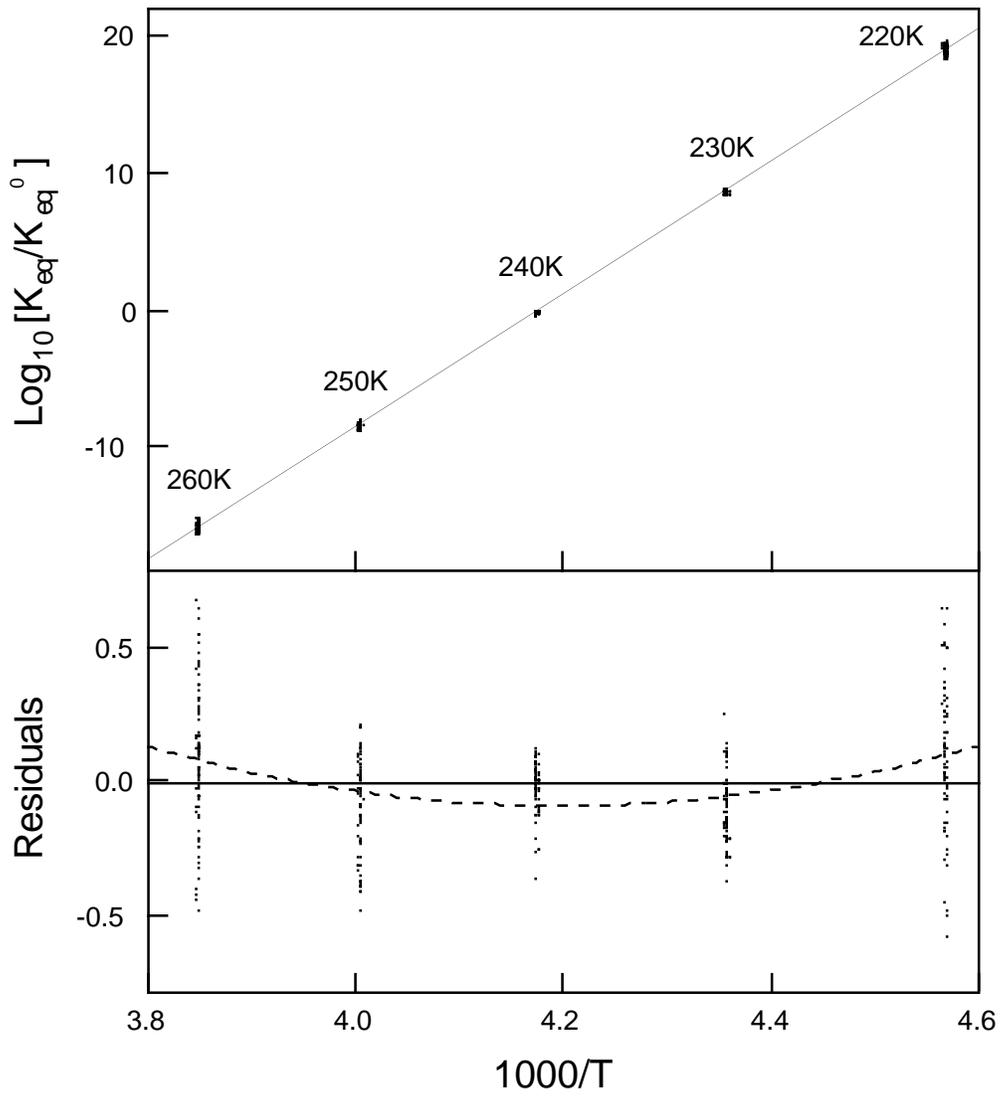


Figure 3

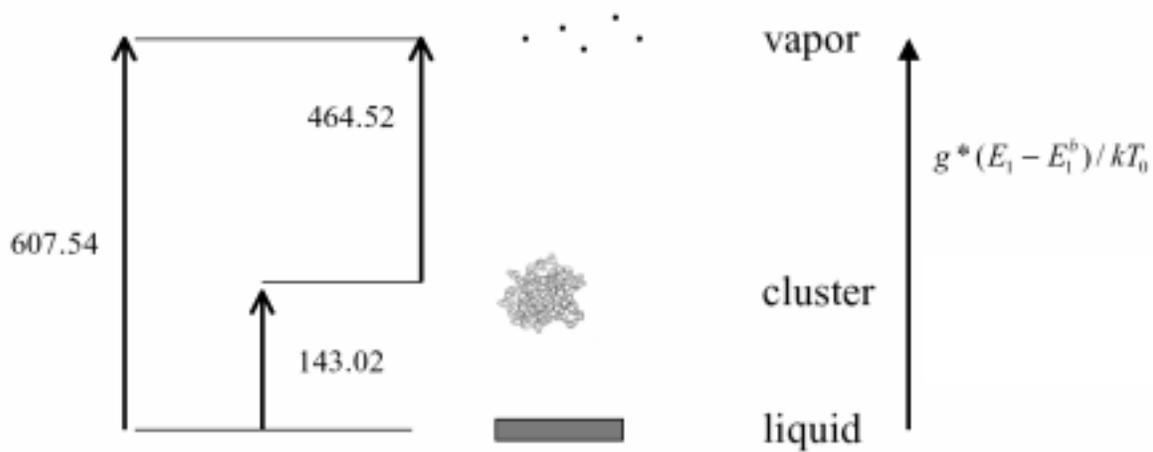


Figure 4

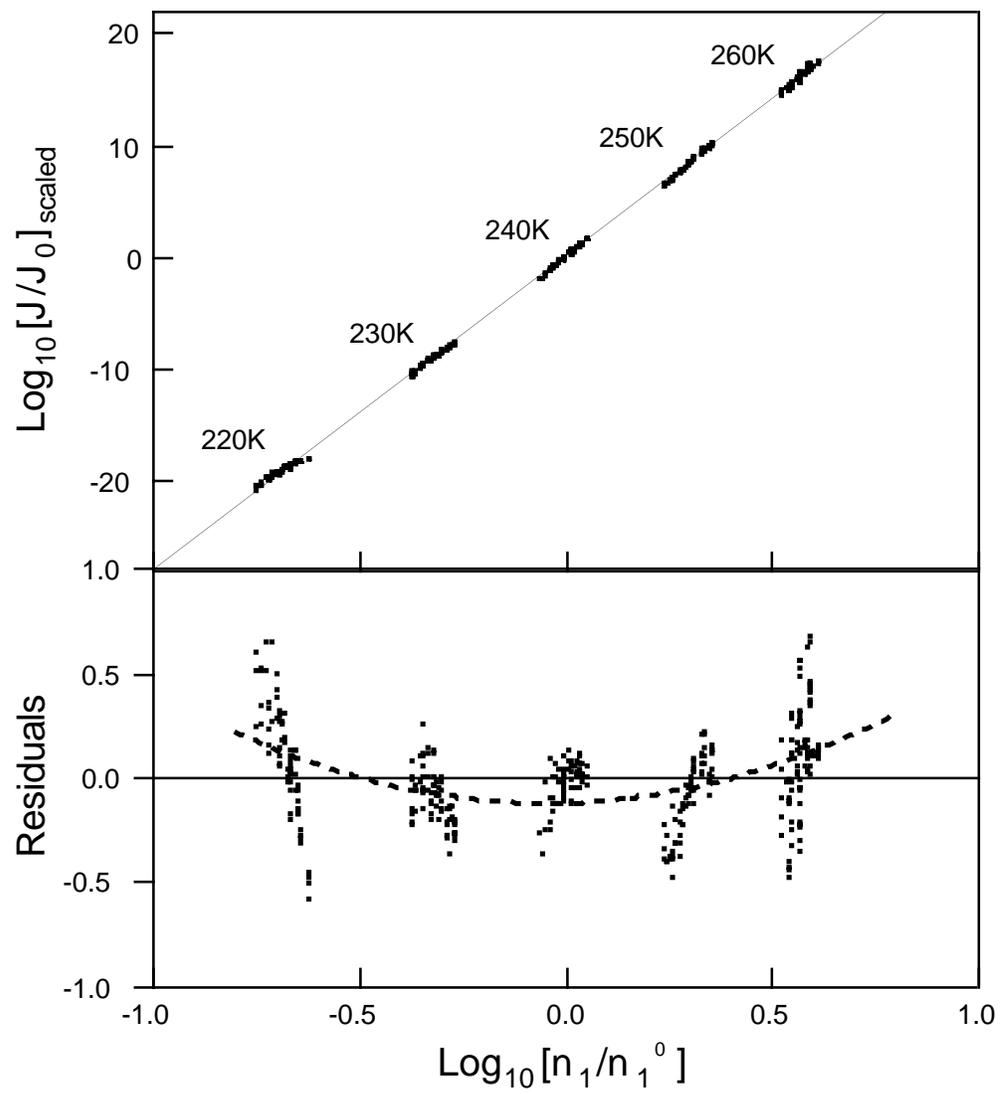


Figure 5

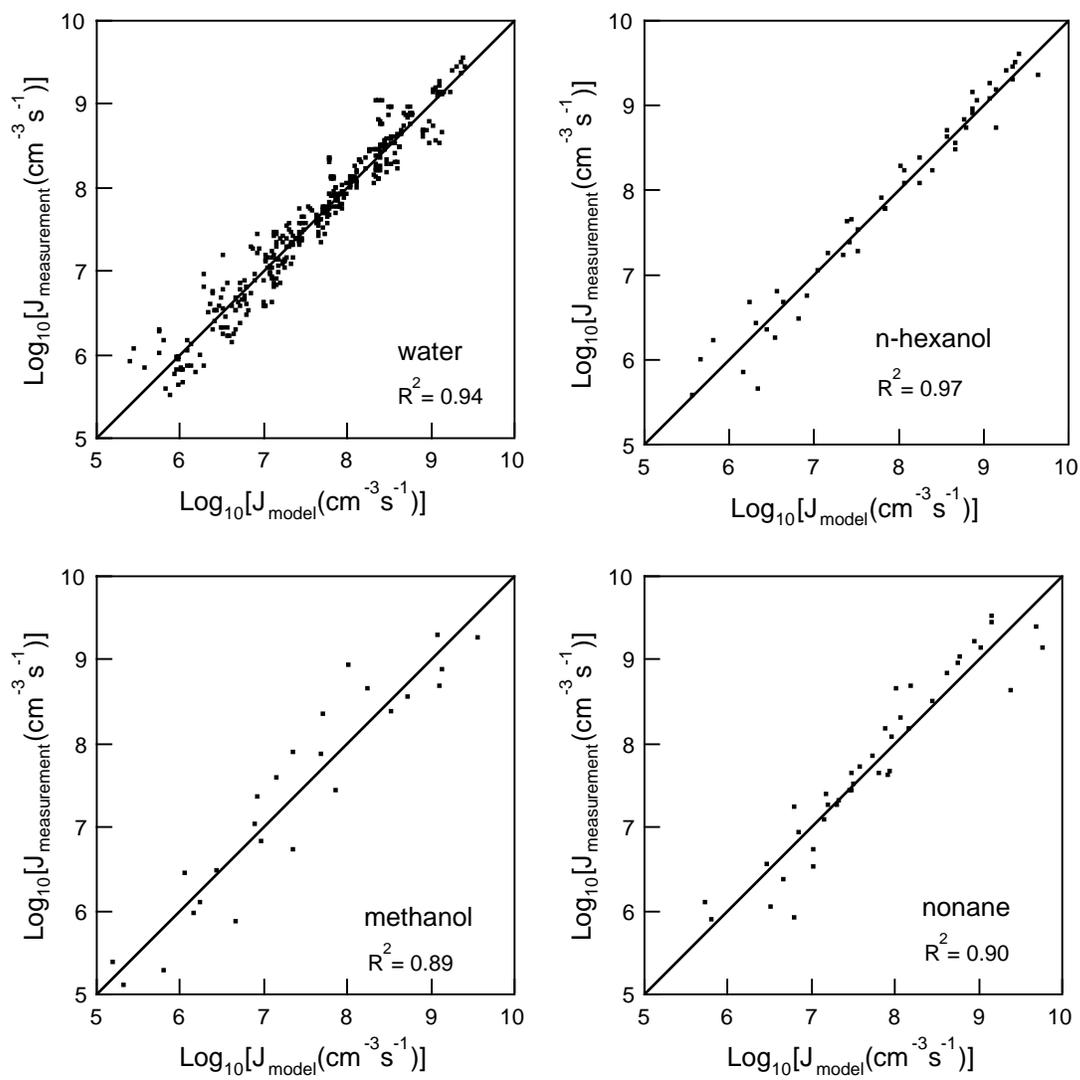


Figure 6

Table 1. Number of measurements (N), centroid conditions for data points  $\{x_i, y_i\}$  and model fit parameters  $z = a + bx + cy$  where  $z = \ln[J/cm^{-3}s^{-1}]$ ,  $x = 1000/T$ , and  $y = \ln S$ . Remaining rows give the quality of fit ( $R^2$ ); critical cluster size as inferred from the measurements ( $g^*$ ) and predicted using the capillary drop approximation of classical nucleation theory ( $g_{CNT}^*$ ); energy to vaporize a single molecule from the bulk liquid ( $\Delta E_1/kT_0$ ); critical cluster energy relative to bulk liquid as inferred from the measurements ( $\Delta E_b^*/kT_0$ ) and predicted using the capillary drop approximation ( $\Delta E_{CNT}^*/kT_0$ ). Results are given for each of the four substances included in Fig. 5.

Property	water	methanol	n-hexanol	nonane
N	343	25	50	41
$T_0$ (K)	239.5	250.3	275.6	222.7
$S_0$	11.7	3.10	11.15	245.7
a	112.831	70.0472	103.425	127.512
b	-39.251	-28.0212	-47.0352	-41.3059
c	27.8899	52.0461	34.6337	13.7828
$R^2$	0.94	0.89	0.97	0.90
$g^*$	27	51	34	13
$g_{CNT}^*$	30	42	36	16
$\Delta E_1/kT_0$	22.43	19.05	26.24	13.12
$\Delta E_b^*/kT_0$	141.40	92.92	144.45	172.34
$\Delta E_{CNT}^*/kT_0$	140.97	152.55	232.77	210.38