On the design of experiments for determining ternary mixture free energies from static light scattering data using a nonlinear partial differential equation

Chris W. Wahle,¹ David S. Ross,¹ and George M. Thurston²

¹*School of Mathematical Sciences, Rochester Institute of Technology, Rochester, New York 14623, USA*

²*Department of Physics, Rochester Institute of Technology, Rochester, New York 14623, USA*

(Received 15 December 2011; accepted 13 June 2012; published online 20 July 2012)

We mathematically design sets of static light scattering experiments to provide for model-independent measurements of ternary liquid mixing free energies to a desired level of accuracy. A parabolic partial differential equation (PDE), linearized from the full nonlinear PDE [D. Ross, G. Thurston, and C. Lutzer, *J. Chem. Phys.* **129**, 064106 (2008)], describes how data noise affects the free energies to be inferred. The linearized PDE creates a net of spacelike characteristic curves and orthogonal, timelike curves in the composition triangle, and this net governs diffusion of information coming from light scattering measurements to the free energy. Free energy perturbations induced by a light scattering perturbation diffuse along the characteristic curves and towards their concave sides, with a diffusivity that is proportional to the local characteristic curvature radius. Consequently, static light scattering can determine mixing free energies in regions with convex characteristic curve boundaries, given suitable boundary data. The dielectric coefficient is a Lyapunov function for the dynamical system whose trajectories are PDE characteristics. Information diffusion is heterogeneous and system-dependent in the composition triangle, since the characteristics depend on molecular interactions and are tangent to liquid-liquid phase separation coexistence loci at critical points. We find scaling relations that link free energy accuracy, total measurement time, the number of samples, and the interpolation method, and identify the key quantitative trade-offs between devoting time to measuring more samples, or fewer samples more accurately. For each total measurement time there are optimal sample numbers beyond which more will not improve free energy accuracy. We estimate the degree to which many-point interpolation and optimized measurement concentrations can improve accuracy and save time. For a modest light scattering setup, a sample calculation shows that less than two minutes of measurement time is, in principle, sufficient to determine the dimensionless mixing free energy of a non-associating ternary mixture to within an integrated error norm of 0.003. These findings establish a quantitative framework for designing light scattering experiments to determine the Gibbs free energy of ternary liquid mixtures.

© 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4731694>]

I. INTRODUCTION

The long-term goal of this work is to provide experimental means of determining the free energies of mixing of multicomponent liquids in a non-parametric manner, using light scattering intensity measurements as a function of mixture composition. Such a method can supplement standard static light scattering determination of molecular weights, virial coefficients, and susceptibilities by determining the free energy itself in a more general fashion, throughout the liquid-phase region of composition space.

Direct free energy data vs. composition can provide a sound, long-lasting foundation for guiding and testing statistical thermodynamic models of a given experimental system. The light scattering intensity has the great value of providing a non-invasive, essentially non-perturbing way to gain such free energy data. Nature is full of surprises, so direct free energy data can be particularly useful when a chosen, prior mathematical model such as the virial expansion, an association model, or a liquid-state theory approximation does not work

well for an experimental system of interest. From the measured free energy function vs. composition, chemical potentials, and thermodynamic susceptibilities vs. composition can also be calculated, and used for quantifying driving forces for chemical change and diffusion, as well as for checking consistency of measured tie lines and phase boundaries.

In the previous work, we showed that solution of a nonlinear partial differential equation (PDE), with theoretical light scattering efficiency vs. composition as input, could be used to determine the mixing free energy of ternary liquid mixtures.¹ The PDE used follows from the statistical-thermodynamic theory of the efficiency of the light scattering, or Rayleigh ratio, from multicomponent, isotropic liquid mixtures, within the approximation that the molecular volumes are additive.¹

The present work focuses on the practical experimental design problem for light scattering free energy determination. This problem is mathematically novel and interesting in its own right, given the need to solve a PDE to connect the data

FIG. 1. Precise measurements of the light scattering Rayleigh ratio (upper left) can be used as the basis for integration of a nonlinear partial differential equation to yield the underlying mixing Gibbs free energy of a liquid mixture (upper right), as shown in Ross *et al.*¹ But real measurements (lower left) will be made only at a finite number of compositions, measured with noise, and interpolated or smoothed before integration. Each practical aspect will contribute to the error in inferred free energy, as depicted by the difference function at lower right, that corresponds to the simulated measurements at lower left. In this paper, we analyze experimental designs that target a desired degree of accuracy of measured free energies.

to the free energy. For ternary mixtures, we show that this connection involves a dynamical system of characteristics in the composition triangle that forms the basis for understanding the nature of a diffusion of information from the light scattering data to the inferred free energy. By analysis and simulation of binary and ternary mixtures, we study the influence of finite sample numbers, noise in light scattering data, choice of interpolation method, and finite measurement time on the expected accuracy of the measured free energy. The analysis results in scaling laws and guidelines that can help in planning sample composition spacing and measurement time requirements for a given instrument.

The essential question addressed in this work is illustrated in Fig. 1. If light scattering efficiency were to be measured very accurately at closely spaced compositions, as depicted at upper left in Fig. 1, PDE integration would yield the mixing free energy perfectly, as shown in the upper right plot. To what degree do the number and spacing of samples, measurement noise, and a chosen interpolation or smoothing method, as illustrated at lower left, make the inferred free energy differ from the true function? This is illustrated in Fig. 1 at lower right, by the free energy difference function corresponding to the simulated data at lower left.

We first set up a general framework for addressing the design considerations that follow from the need for integration of a differential equation, for both binary and ternary liquid mixtures. Then, to make numerical estimates about the key requirements, we primarily consider solutions well described by a regular solution free energy model. It should be noted that optimal experimental designs will be both system- and equipment-dependent.

It turns out that a diffusion equation on an orthogonal network of curves governs the flow of information from light scattering perturbations to corresponding changes in the inferred free energy. The network of curves itself follows from a dynamical system whose trajectories are the characteristics of both the original PDE and its linearized version. The dynamical system, in turn, depends on both the underlying free energy and the dependence of the dielectric coefficient on composition. Interestingly, for an ideal solution, the dynamical system whose trajectories are PDE characteristics is mathematically the same as a competing species model in ecology.

Through analysis of this information diffusion in combination with numerical simulation, limiting scaling relations between free energy error and experimental design parameters are obtained, together with more general plots that identify regions of parameter space within which different primary sources of error limit accuracy.

With the experimental design information so obtained, together with knowledge about efficiency and noise for a given light scattering setup, one can estimate the concentration spacing of samples and the light scattering measurement time needed to determine the free energy to within a chosen tolerance. For a given, new system such estimates must be regarded as preliminary; for instance, in the case of incipient molecular association such as micelle formation, or near phase separation spinodal boundaries, light scattering efficiency might change rapidly with composition and call for finer grid spacing in parts of the composition triangle.

In its present form the method assumes, in addition to additive molecular volumes, that the single-phase portions of the system are isotropic liquids, and that the single-phase part of the ternary composition triangle is both simply connected and extends to include all or part of all three binary axes. We expect that some of these restrictions can eventually be removed.

On the experimental side, it is assumed that the efficiency with which light is singly scattered is measured. Also, roughly speaking, the scattering angle should be low enough to probe concentration fluctuations at a spatial scale larger than that of the spatial correlations of polarizability in the solution. Such correlations could be due to characteristic molecular or aggregate sizes, or to collective concentration fluctuations. Thus the theory will not readily apply near a critical locus without additional considerations, nor will it apply to solutions that contain molecules or aggregates that are too large for a given light wavelength and scattering angle.

It is important to note that both small-angle x-ray scattering (SAXS) and small-angle or ultra small-angle neutron scattering (SANS or USANS) could very readily be used to determine the free energy, by applying the same mathematical framework used here for light scattering to the x-ray or neutron scattering cross sections in the near-forward direction. To do so, one would replace the dielectric coefficient at optical frequencies with the electron density for x-rays, or with the scattering length density for neutrons. Either x-ray or neutron scattering would have the advantage over light scattering of giving more information about liquid structure and particle structure from the higher angle data.

An outline of the paper is as follows. In Sec. II, we recall the nonlinear light scattering equation presented previously and recast it in a form that puts it among a recognized, studied class of nonlinear PDEs, a form that also allows one to readily deduce some key properties of the PDE characteristics. In Sec. III, we linearize the light scattering equation around its exact solution to obtain a parabolic PDE that describes information diffusion within the composition space from light scattering data to the free energy, and we describe the characteristics of the light scattering equation, which govern the diffusion of free energy information. We illustrate how the characteristics respond to changes in the dielectric gradient, and to intermolecular interactions that can result in proximity to spinodal and critical loci for liquid-liquid phase separation. In Sec. IV, we formulate the design problem of minimizing uncertainty in the inferred free energy due to light scattering data noise, discrete sample compositions, and numerical interpolation or smoothing, under the constraint of finite total experiment time. We first do this in a general setting, independent of the number of mixture components, then consider distinct features of binary and ternary mixture design. We also show how to calculate the actual experimental light scattering time needed for a given instrument and provide an estimate of this time for a typical laboratory light scattering instrument. We then describe our numerical results for the relationships among free energy error, sample number, and composition choice, dimensionless time for experimentation, and properties of the underlying free energy itself, including proximity to phase separation. These results include scaling relationships that can be used to help plan time and sample requirements for practical light scattering free energy determinations. We then state important open questions and summarize the work.

II. CLASSIFICATION AND CHARACTERISTICS OF THE NONLINEAR LIGHT SCATTERING PDE

We have previously shown how to find the underlying Gibbs free energy of mixing of an isotropic liquid formed by three liquid components, from measurements of the binary and ternary composition dependence of the absolute efficiency of the light scattering at forward angles.¹ We briefly give an overview of the derivation of the main governing equation here; the detailed steps may be found in Ref. 1. That analysis starts from the final, general expression for the Rayleigh ratio for multicomponent mixtures presented as Eq. (12) in Kirkwood and Goldberg,² which is cast in terms of the matrix of each of the concentration derivatives of each of the chemical potentials of the components of the mixture. The analysis in Ref. 1 makes use of the fact that the chemical potential derivatives may be re-expressed in terms of second concentration derivatives of the intensive free energy, which provides an algebraic route for recasting the Rayleigh ratio in terms of the derivatives of a single function. The analysis applies to the restricted case of liquids that can be modeled as having partial molecular volumes that do not vary with composition. With that assumption, the Appendix of Ref. 1 details the steps needed to express the light scattering turbidity and forward scattering angle Rayleigh ratio, in excess of the Rayleigh ratio associated with the compressibility of the

system, in terms of the Hessian matrix of second derivatives of the free energy per unit volume with respect to the number densities of the components. We term the resulting light scattering cross section the excess Rayleigh ratio. Included in Ref. 1 are the needed relationships between the chemical potentials and the free energy per unit volume, and the relevant change of variables from those used by Kirkwood and Goldberg, to number densities. The resulting expression for the excess Rayleigh ratio reduces to zero for each of the pure components, as it must since for each of the pure components the light scattering intensity can be expressed solely in terms of its compressibility, which is not included in the present analysis. Clearly, this is an opportunity for generalization of what follows.

The result of the analysis in Ref. 1 is the following nonlinear partial differential equation for the mixing free energy g , with excess Rayleigh ratio R ,

$$\frac{\varepsilon_y^2 g_{xx} - 2\varepsilon_x \varepsilon_y g_{xy} + \varepsilon_x^2 g_{yy}}{g_{xx} g_{yy} - g_{xy}^2} = R(x, y) = \nabla \varepsilon \cdot (H[g])^{-1} \cdot \nabla \varepsilon. \quad (1)$$

In Eq. (1), g denotes the intensive, dimensionless Gibbs free energy $\Omega_0 G / (V k_B T)$, where G is the extensive Gibbs free energy, V is the volume, k_B is Boltzmann's constant, T is absolute temperature, $\Omega_0 = \partial V / \partial N_0$ is the volume associated with each molecule of component 0, and $x = \Omega_1 N_1 / V$ and $y = \Omega_2 N_2 / V$ are the volume fractions of components 1 and 2. Consistent with the derivation of Eq. (1) from the multicomponent light scattering formulation of Kirkwood and Goldberg,¹ we take each Ω_i to be constant. Also in Eq. (1), $g_{xx} = \partial^2 g / \partial x^2$, and likewise for g_{xy} and g_{yy} . (Throughout this paper we use subscripts to denote partial derivatives.) The function $\varepsilon(x, y)$ is the real part of the dimensionless dielectric coefficient at the optical frequency used for measuring $R(x, y)$. The function $R(x, y)$ is a convenient, dimensionless form of the excess Rayleigh ratio, $\Delta R(0)$, the latter having dimensions of inverse length. Specifically, $R(x, y) = \lambda^4 \Delta R(0) / (\pi^2 \Omega_0)$, where λ is the vacuum wavelength of the incident and the scattered light. The matrix $H[g]$ is the Hessian of g with respect to the volume fraction variables, while $\nabla \varepsilon$ is the gradient of ε in the same variables; the expression on the right side of Eq. (1) holds for any number of composition variables greater than or equal to two.

There is a terse reformulation of Eq. (1) that is convenient and informative, which we derive here for the ternary mixture case. The expression

$$\frac{\cos^2(\theta) g_{xx} - 2 \cos(\theta) \sin(\theta) g_{xy} + \sin^2(\theta) g_{yy}}{(\cos(\theta) \varepsilon_x + \sin(\theta) \varepsilon_y)^2} \quad (2)$$

is the ratio of the second directional derivative of g to the square of the first directional derivative of ε in the direction of the unit vector $(\cos(\theta), \sin(\theta))$. Informally, it is easy to see this ratio as a measure of how much scattering would be caused by thermal perturbations in the direction of this unit vector; if the numerator is large, small thermal perturbations will produce little change in composition, and if the denominator is small, small changes in composition will produce little variation in the index of refraction. So if this

ratio is large, we would expect little scattering because of composition fluctuations in the direction of the unit vector. The most significant direction for scattering is the direction in which this ratio is smallest. A calculus computation shows that the minimum value of this ratio over all possible directions is $(g_{xx}g_{yy} - g_{xy}^2)/(\varepsilon_y^2 g_{xx} - 2\varepsilon_x \varepsilon_y g_{xy} + \varepsilon_x^2 g_{yy})$, the reciprocal of the expression on the left-hand side of Eq. (1). This observation yields a reformulation of Eq. (1),

$$\min_s \frac{g_{ss}}{\varepsilon_s^2} = \frac{1}{R(x, y)}, \quad (3)$$

in which the minimum is taken with respect to directions s in the (x, y) -plane that are not orthogonal to $\nabla \varepsilon$. Equation (3) is a fully nonlinear degenerate elliptic PDE, a variant of the Hamilton-Jacobi-Bellman equation.³ Although the derivation above referred only to the ternary case, a more general derivation given in Sec. 1 of the Appendix shows that Eq. (3) applies independently of the number of components in the mixture.

One implication of Eq. (3) is that experimental measurements of R and $\nabla \varepsilon$ give the ratio of the directional second derivative of g to the square of the directional derivative of ε , in the direction $s_{\min}$ in which that ratio is a minimum. This is true despite the fact that, before solving either Eq. (1) or Eq. (3), the actual direction s that minimizes Eq. (2) is not yet known.

A second implication of Eq. (3) is that $s_{\min}$ is also a characteristic direction of the light scattering PDE. A characteristic direction has the property that one cannot arbitrarily specify g along it. This is indeed the case for $s_{\min}$, since Eq. (3) shows that g_{ss} is determined along $s_{\min}$ by the input functions $R(x, y)$ and $\varepsilon(x, y)$.

A third implication of Eq. (3) is that the direction $s_{\min}$ cannot become orthogonal to $\nabla \varepsilon$. Otherwise, the denominator of Eq. (2) would vanish, and hence $s_{\min}$ could not correspond to a minimum value for Eq. (2). Thus, along characteristics ε must be a monotonic function of arc length and consequently, the dielectric coefficient is a Lyapunov function for the dynamical system, analyzed below, whose solution is the characteristics.

We now further analyze Eq. (3) in coordinates defined by the second derivatives of g , in order to give more insight into the geometry of the characteristics, and to help show that the characteristics cannot enter certain vertices of the composition triangle. To this end let λ_+ and λ_- be the eigenvalues of the Hessian of $g(x, y)$,

$$H = \begin{pmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{pmatrix},$$

with $\lambda_+ \geq \lambda_-$, and let $\mathbf{v}_+$ and $\mathbf{v}_-$ be the corresponding orthonormal eigenvectors. Let ξ and η be the coordinates in the direction of $\mathbf{v}_+$ and $\mathbf{v}_-$, respectively. Consider Eq. (3) in the $\xi\eta$ -coordinate system. Let $\mathbf{s}$ be a unit vector in the unknown s -direction, and let θ be the angle between $\mathbf{s}$ and $\mathbf{v}_+$. Furthermore, let α be the angle between $\nabla \varepsilon$ and $\mathbf{v}_+$, as shown in

FIG. 2. A small change in measured Rayleigh ratio changes inferred free energy in proportion to a toothlike surface (left – in profile; right – from above). Information from light scattering diffuses to the free energy in a spacelike direction along characteristics (left – colored, right – black; one red), and in a timelike direction towards concave characteristic sides. Colors indicate free energy change magnitudes resulting from integration of a tentlike function $T_j(\mathbf{x})$ (Eq. (12)). For linear interpolation of light scattering data onto the PDE solution grid, $T_j(\mathbf{x})$ has a hexagonal base as shown. (Top right inset): Key vectors at a point on the red characteristic. In accord with Eq. (4), $\pi/2 > \theta > \alpha$, so that ε is monotonic along the characteristic. (Bottom) Characteristic directions and endpoints depend on $\nabla \varepsilon$ direction.

Fig. 2. Then, we have

$$g_{\xi\xi} = \mathbf{v}_+ H \mathbf{v}_+^T = \lambda_+ \mathbf{v}_+ \mathbf{v}_+^T = \lambda_+,$$

$$g_{\xi\eta} = \mathbf{v}_+ H \mathbf{v}_-^T = \lambda_- \mathbf{v}_+ \mathbf{v}_-^T = 0,$$

$$g_{\eta\eta} = \mathbf{v}_- H \mathbf{v}_-^T = \lambda_- \mathbf{v}_- \mathbf{v}_-^T = \lambda_-,$$

so in the $\xi\eta$ -coordinate system, the Hessian of $g(\xi, \eta)$ is given by

$$\tilde{H} = \begin{pmatrix} g_{\xi\xi} & g_{\xi\eta} \\ g_{\xi\eta} & g_{\eta\eta} \end{pmatrix} = \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix},$$

and $\mathbf{s} = \langle \cos \theta, \sin \theta \rangle$. Then,

$$g_{ss} = \mathbf{s} \tilde{H} \mathbf{s}^T = \lambda_+ \cos^2 \theta + \lambda_- \sin^2 \theta = (\lambda_+ - \lambda_-) \cos^2 \theta + \lambda_-$$

and

$$\begin{aligned} \varepsilon_s &= \nabla \varepsilon \cdot \mathbf{s} = |\nabla \varepsilon| \langle \cos \alpha, \sin \alpha \rangle \cdot \langle \cos \theta, \sin \theta \rangle \\ &= |\nabla \varepsilon| \cos(\theta - \alpha). \end{aligned}$$

Thus,

$$\frac{g_{ss}}{\varepsilon_s^2} = \frac{(\lambda_+ - \lambda_-) \cos^2 \theta + \lambda_-}{|\nabla \varepsilon|^2 \cos^2(\theta - \alpha)}.$$

This expression is minimized when θ satisfies the equation

$$\lambda_+ \tan \alpha = \lambda_- \tan \theta. \quad (4)$$

Since $\lambda_+ \geq \lambda_-$, it follows from Eq. (4) that $\tan \theta \geq \tan \alpha$. Therefore, the acute angle between the tangent lines to the characteristic curves and $\mathbf{v}_+$ is always greater than or equal to the acute angle between $\mathbf{v}_+$ and lines parallel to the dielectric gradient vector, with equality holding at umbilic points of the free energy surface, at which $\lambda_+ = \lambda_-$.

Now, of the two directed rays that are tangent to a characteristic curve at a point, consider the ray that forms an acute

angle with $\nabla\epsilon$, the angle given by $\theta - \alpha$ in Fig. 2. As the characteristic curve is traversed, this tangent ray cannot change direction discontinuously, and therefore Eq. (4) constrains $\theta - \alpha$ to remain acute. Therefore, the directional derivative of ϵ along such a tangent ray is positive, consistent with the conclusion presented above, that the dielectric coefficient itself is a Lyapunov function for the dynamical system that yields the characteristics.

We now show that the characteristics cannot enter a vertex of the composition triangle at which a normal direction to $\nabla\epsilon$ points into the triangle. To do so, we make use of the following useful expression for g from statistical thermodynamics, which correctly gives the asymptotic behavior of g for nonelectrolyte solutions as at least one of the three volume fractions x , y or $1 - x - y$ approaches 0 (Ref. 4): $g(x, y) = (\Omega_0 x / \Omega_1) \log(x) + (\Omega_0 y / \Omega_2) \log(y) + \log(1 - x - y) + \tilde{\phi}(x, y)$, in which the dimensionless excess free energy of mixing, $\tilde{\phi}(x, y)$, as well as its composition derivatives, have finite values throughout the closed ternary composition triangle.

In the interest of clarity we present the argument for the origin; it applies, mutatis mutandis, to all vertices. Near the origin, the Hessian

$$H = \begin{bmatrix} \frac{\Omega_0}{\Omega_{1x}} & 0 \\ 0 & \frac{\Omega_0}{\Omega_{2y}} \end{bmatrix} + \mathcal{O}(1).$$

We fix the ratio of y to x . The characteristic direction corresponds to a unit vector $(\cos(\sigma), \sin(\sigma))$ that minimizes

$$\frac{\frac{\Omega_0}{\Omega_{1x}} \cos^2(\sigma) + \frac{\Omega_0}{\Omega_{2y}} \sin^2(\sigma) + \mathcal{O}(1)}{\epsilon_x^2 \cos^2(\sigma) + 2\epsilon_x \epsilon_y \cos(\sigma) \sin(\sigma) + \epsilon_y^2 \sin^2(\sigma)}.$$

As we approach the origin the term $\mathcal{O}(1)$ in the numerator become negligible and the characteristic direction corresponds to the value of σ that minimizes the expression that remains. If the components of $\nabla\epsilon$ differ in sign, then so will the components of $(\cos(\sigma), \sin(\sigma))$ that correspond to the minimum; thus in such a case characteristics do not point into the first or fourth quadrants.

III. THE LINEARIZED LIGHT SCATTERING PDE: DIFFUSION OF FREE ENERGY INFORMATION

In order to provide a basis for analyzing how noise in light scattering measurements affects the inferred free energy, we now linearize Eq. (1). For this purpose we introduce the small perturbations f and ρ , and let $g \rightarrow g + f$, $R \rightarrow R + \rho$, where both f/g , $\rho/R \ll 1$, to obtain

$$af_{xx} - 2bf_{xy} + cf_{yy} = -(\det H)^2 \rho(x, y), \quad (5)$$

where

$$a = (\nabla g_y \times \nabla \epsilon)^2, \quad b = (\nabla g_x \times \nabla \epsilon)(\nabla g_y \times \nabla \epsilon), \quad \text{and} \\ c = (\nabla g_x \times \nabla \epsilon)^2.$$

Equation (5) is parabolic, or degenerate elliptic, since $b^2 = ac$. This immediately implies that noise in light scattering measurements at a particular composition, which is later to be represented by a random variable $\rho(x, y)$, produces an effect

f on the free energy g that “diffuses” within the composition triangle from the initial point, in a sense that we will define and analyze further below.

However, it is important to recognize, first, that both the space and time directions that occur in normal one-dimensional diffusion are to be found within the ternary concentration triangle. That is, the free energy perturbation $f(x, y)$ due to a light scattering perturbation $\rho(x, y)$ is analogous to a record of both the time- and space-dependence of the diffusion of particles in one dimension. Second, the complicated composition dependence of the coefficients H , a , b and c means that the effective diffusion coefficient, as well as the spacelike and timelike directions, are at once functions of position in the ternary composition triangle, functions of the direction of the dielectric coefficient gradient, and functions of the components of the Hessian of the free energy g . It is important to bear in mind that there is no temporal dimension and no spatial dimension in this problem; the terms timelike and spacelike refer to the purely mathematical structure of the PDE.

Each of these points is illustrated in Figs. 2–4 by calculated free energy perturbations resulting from light scattering perturbations at selected composition points. Each light scattering perturbation has first been linearly interpolated onto the grid used to solve the light scattering PDE, as described below in the formulation of the design problem. Figure 2 introduces the basic structure of the toothlike free energy perturbation surface, and Fig. 3 illustrates the variety of such free energy perturbations that occur as a function of composition. To construct both Figs. 2 and 3 an ideal solution free energy g was used, while Fig. 4 used instead a regular solution free energy near liquid-liquid criticality on the y axis, to illustrate how free energy perturbations are also quite sensitive to changes in the underlying free energy g . These considerations show explicitly what is naturally to be expected, namely, that the experimental design problem for determining the free energy from light scattering measurements, to be analyzed below, is an intricate one in which the error analysis is sensitive to the system under investigation.

The nature of diffusion of information about the free energy is analyzed in further detail in Sec. 2 of the Appendix. There, we identify the spacelike and timelike directions, that is, the analogues in the composition triangle of the spatial and temporal dimensions in a literal, one-dimensional diffusion problem. Specifically, one can take the spacelike coordinate, $s(x, y)$, to be equal to the optical dielectric coefficient, $\epsilon(x, y)$, and the timelike coordinate to be $t = \epsilon_y g_x - \epsilon_x g_y = \nabla g \times \nabla \epsilon$. It also turns out that the nature of the diffusion is governed by the shapes of the characteristics of the PDE, which we analyze further, below. In these terms, the analysis in Sec. 2 of the Appendix shows, further, that the diffusion of information equation can be written as (Eq. (A3) in Sec. 2 of the Appendix)

$$\sigma \kappa |\nabla t|^3 f_t = (\nabla \epsilon \times \nabla t)^2 f_{ss} + (\det H)^2 \rho(s, t). \quad (6)$$

Equation (6) is in the form of the heat equation with a source term, and the diffusion coefficient is proportional to $1/\kappa$, the local radius of curvature of the characteristics, defined in Sec. 2 of the Appendix. σ , also defined there, specifies that

FIG. 3. Diffusion of free energy information during PDE solution is a strong function of ternary composition, and consequently light scattering noise has composition-dependent effects on inferred free energy. See Fig. 5 (left) for relative sizes of intrinsic free energy error contributions vs. composition. Colors as in Fig. 2; ideal solution; $\nabla\epsilon$ parallel to $(\sqrt{3}/2, 1/2)$.

the direction of the diffusion is towards the concave side of the characteristics. We note that Sec. 2 of the Appendix also recasts Eq. (6) in terms of the angles θ and α that are shown in Fig. 2.

To finish this analysis of the linearized light scattering equation we now present explicit forms for its characteristics, by analyzing a free energy that is the sum of the ideal free energy $x \ln x + y \ln y + z \ln z$ plus an excess free energy $\phi(x, y)$,

$$g(x, y) = x \ln x + y \ln y + z \ln z + \phi(x, y),$$

where $z = 1 - x - y$. For simplicity we take $\Omega_0 = \Omega_1 = \Omega_2$, and have distinguished ϕ from $\tilde{\phi}$ (above) for this reason. We have

$$t(x, y) = \ln \left[\left(\frac{x}{z} e^{\phi_x} \right)^{\epsilon_y} \left(\frac{y}{z} e^{\phi_y} \right)^{-\epsilon_x} \right],$$

or equivalently,

$$t(x, y) = \ln \left[\frac{x^{\epsilon_y}}{y^{\epsilon_x}} (1 - x - y)^{\epsilon_x - \epsilon_y} \right] + \epsilon_y \phi_x - \epsilon_x \phi_y,$$

FIG. 4. (Left) Free energy information flows in a molecular interaction-dependent fashion during PDE solution, since light scattering PDE characteristics (black) can develop inflections (green), shown here near liquid-liquid criticality on the y axis for a regular solution free energy.^{11,12} Free energy magnitudes (colors) from integration of the two hexagonal interpolation tents illustrate that information flows towards concave characteristic sides. (Right) A characteristic (red) tangent to liquid-liquid coexistence (green) at a critical point (blue star); the model spinodal is the curve inside the green region. (Inset) Hessian eigenvectors and $\nabla\epsilon$ at the critical point. Colors as in Fig. 2; parameters (left) $C_{xy} = 0$, $C_{xz} = 0$, $C_{yz} = 1.9$, $\epsilon_x = \sqrt{3}/2$, $\epsilon_y = 1/2$. (right) $C_{xy} = 0$, $C_{xz} = 0$, $C_{yz} = 2.2$, $\epsilon_x = \sqrt{3}/2$, $\epsilon_y = 1/2$.

and the characteristics, $t(x, y) = C$, are defined by

$$\frac{x^{\epsilon_y}}{y^{\epsilon_x}} (1 - x - y)^{\epsilon_x - \epsilon_y} e^{\epsilon_y \phi_x - \epsilon_x \phi_y} = \frac{\left(\frac{x}{z} e^{\phi_x} \right)^{\epsilon_y}}{\left(\frac{y}{z} e^{\phi_y} \right)^{\epsilon_x}} = K \geq 0. \quad (7)$$

For the regular solution model, $\phi(x, y) = C_{xy}xy + C_{xz}xz + C_{yz}yz$,

$$\phi_x = C_{xz}(1 - 2x) + (C_{xy} - C_{xz} - C_{yz})y,$$

$$\phi_y = C_{yz}(1 - 2y) + (C_{xy} - C_{xz} - C_{yz})x.$$

While the form of the characteristics can be readily obtained, as just done, by setting $t(x, y) = C$, it is also interesting to note that the characteristic curves can also be considered as trajectories of a weakly nonlinear system of ordinary differential equations. In particular, with use of

$$\frac{dy}{dx} = \frac{y}{x} \cdot \frac{\epsilon_y(1 - y + xz\phi_{xx}) - \epsilon_x x(1 + z\phi_{xy})}{\epsilon_x(1 - x + yz\phi_{yy}) - \epsilon_y y(1 + z\phi_{xy})}$$

we find that

$$\dot{x} = f(x, y) = x[\epsilon_x(1 - x + yz\phi_{yy}) - \epsilon_y y(1 + z\phi_{xy})], \quad (8)$$

$$\dot{y} = g(x, y) = y[\epsilon_y(1 - y + xz\phi_{xx}) - \epsilon_x x(1 + z\phi_{xy})]. \quad (9)$$

Furthermore, for the ideal form of the free energy, the second composition derivatives of ϕ are all 0, and provided $\nabla\epsilon$ is a constant vector, the dynamical system for the light scattering PDE characteristics is a special case of the competing species model.⁵

IV. FORMULATION OF THE DESIGN PROBLEM

In this section, we investigate the free energy consequences of perturbations in the light scattering intensity, considered as being measured with noise at particular points in the composition space. The error in the free energy results in part from the fact that it is impractical to measure the light scattering at every point within the region, and in part from the

fact that the light scattering process itself will involve measurement error. We envision an overall process in which the light scattering values at isolated measurement points are subject to some form of interpolation to create values on a grid to be used for numerical integration of the light scattering partial differential equation. Since there is an integration of the light scattering intensity to obtain the free energy, an isolated error in light scattering will have consequences for the free energy over a region of concentrations, not just at the perturbation site. We begin with general considerations before proceeding to the binary and ternary cases, respectively, which turn out to have different characters, being described by ordinary and partial differential equations, respectively.

A. General consequences of measurement uncertainty and method optimization

We begin by expressing the interpolated light scattering noise function, $\tilde{\rho}(\mathbf{x})$, as a sum over distinct regions of the composition space. As a result of interpolation of the light scattering data prior to integration of the light scattering PDE, the interpolated light scattering at any given point $\mathbf{x}$ within the composition space will typically have contributions from several nearby measurements. Specifically, within each region k , there will be a set of measurement points, $\{j(k)\}$, for which nonzero noise ρ_k would make a nonzero contribution to $\tilde{\rho}(\mathbf{x})$. For each of the interpolation methods used in this paper, namely, first-, third-, and seventh-degree Lagrange interpolation in the case of binary mixtures, and linear interpolation as well as quadratic and cubic Shepard interpolation⁶ in the case of ternary mixtures, the contributions can be written in the following form:

$$\tilde{\rho}(\mathbf{x}) = \sum_k \tilde{\rho}_k(\mathbf{x}) = \sum_k \sum_{\{j\}(k)} c_{kj}(\mathbf{x}, \mathbf{x}_j) \rho_j, \quad (10)$$

in which $c_{kj}(\mathbf{x}, \mathbf{x}_j)$ symbolizes the contribution to region k for a unit error at measurement point j , and the notation $\{j\}(k)$ symbolizes the set of points, $\{j\}$ for which the interpolation functions $c_{kj}(\mathbf{x}, \mathbf{x}_j)$ are nonzero within the given region, k .

By switching the order of summation, we can then identify the total contribution to $\tilde{\rho}(\mathbf{x})$ associated with each measurement point, which we term the tent functions $T_j(\mathbf{x})$,

$$\tilde{\rho}(\mathbf{x}) = \sum_j \rho_j T_j(\mathbf{x}) = \sum_{\text{points } j} \rho_j \sum_{\text{regions } k(j)} c_{kj}(\mathbf{x}, \mathbf{x}_j), \quad (11)$$

$$T_j(\mathbf{x}) = \sum_{\text{regions } k(j)} c_{kj}(\mathbf{x}, \mathbf{x}_j). \quad (12)$$

For the interpolation methods used here, the tent functions $T_j(\mathbf{x})$ have height 1 at their central points $\mathbf{x}_j$ and are zero at other measurement points, so that $T_j(\mathbf{x}_k) = \delta_{jk}$. In between measurement points, $T_j(\mathbf{x}) \neq 0$ only within a small number of intervals or regions neighboring $\mathbf{x}_j$. Each such tent function $T_j(\mathbf{x})$ will be multiplied by the local light scattering error ρ_j and then integrated according to Eq. (5) to generate an ad-

ditive contribution to the error in the free energy. Hence, it is convenient to find the error contribution of each tent function by itself, which we call $\phi_j(\mathbf{x})$, by solving the linearized light scattering PDE, Eq. (5), formally as follows:

$$a(\mathbf{x})\phi_{j,xx}(\mathbf{x}) + 2b(\mathbf{x})\phi_{j,xy}(\mathbf{x}) + c(\mathbf{x})\phi_{j,yy}(\mathbf{x}) = -T_j(\mathbf{x}), \quad (13)$$

$$f(\mathbf{x}) = \sum_j \rho_j \phi_j(\mathbf{x}). \quad (14)$$

Thus the $\phi_j(\mathbf{x})$ give the contributions to the free energy that result from a light scattering error or from light scattering noise at the particular measurement point $\mathbf{x}_j$. In the ternary case, functions $\phi_j(\mathbf{x})$ obtained by integrating Eq. (13) using hexagonal, linear interpolation tents $T_j(\mathbf{x})$ are the ones shown in Figs. 2–4, the tooth functions. Very similar functions can be generated by the same procedure, for the quadratic and cubic interpolation methods that prove below to yield more practical ternary mixture time and number of measurement composition requirements.

The development leading to Eq. (14) now permits us to estimate the size of the error in the free energy. This error has a particularly convenient form when it is expressed in terms of the expectation value of the square of the $\mathcal{L}_2$ -norm, which we use. It is important to note that in making such an analysis, we assume there is no systematic error in the measurement process, through the mathematical assumption that the light scattering errors at different measurement points are not correlated and that each has average value 0, i.e., $\langle \rho_i \rho_j \rangle = \langle \rho_i \rangle \langle \rho_j \rangle = 0$ for $i \neq j$; providing methods for detecting systematic errors, for example through various free energy self-consistency checks, is beyond the scope of the present paper. We then have

$$E[\mathcal{L}_2[f]^2] = E \left[\int (f(\mathbf{x}))^2 d\mathbf{x} \right] = \sum_j \langle \rho_j^2 \rangle \lambda_j, \quad (15)$$

in which

$$\lambda_j = \int \phi_j(\mathbf{x})^2 d\mathbf{x}. \quad (16)$$

To summarize, λ_j is the integral of the square of the tooth function ϕ_j , itself proportional to the free energy contribution that would result from the interpolated light scattering error or noise at the measurement point $\mathbf{x}$, which in turn is proportional to one of the tent functions. Therefore, λ_j is a way to keep track of the contribution to the overall free energy error, as measured by the square of the $\mathcal{L}_2$ -norm, due to such a light scattering error or to noise. Examples of the functions λ_j , which clearly play a key role in the error analysis, are shown in Fig. 5 for three cases of ternary mixtures, described in the caption.

Two comments relevant to Eqs. (15) and (16) are in order. First, note that $E[\mathcal{L}_2[f]] \leq \sqrt{E[\mathcal{L}_2[f]^2]}$, so that the expectation value in Eq. (15) can also be used to bound that of the $\mathcal{L}_2$ -norm of f , itself. Second, note also that to use a different measure of goodness of fit, say, the $\mathcal{L}_4$ -norm, we would instead have needed to evaluate different integrals of the ϕ_j

FIG. 5. Intrinsic free energy error contributions from light scattering noise depend strongly on composition, molecular interactions, and $\nabla\epsilon$ direction. Each bar height, $\sqrt{\lambda_j}$ (Eq. (16)), is the $\mathcal{L}_2$ -norm of the PDE integral of a linear interpolation tent function centered at that point; that is, the $\mathcal{L}_2$ -norm of a function like those in Figs. 2–4. (Left) $\sqrt{\lambda_j}$ vs. composition for an ideal free energy; $\nabla\epsilon$ along $(\sqrt{3}/2, 1/2)$. (Center) $\sqrt{\lambda_j}$ vs. composition for a regular solution free energy with $C_{yz} = 1.9$, $C_{xy} = 0$, and $C_{xz} = 0$; $\nabla\epsilon$ along $(\sqrt{3}/2, 1/2)$. (Right) $\sqrt{\lambda_j}$ vs. composition for an ideal free energy; $\nabla\epsilon$ along $(1/2, \sqrt{3}/2)$. Bar cross-sections show linear interpolation tent function domain shapes, half-width for clarity.

functions

$$\left\langle \int (f(\mathbf{x}))^4 d\mathbf{x} \right\rangle = \sum_j \langle \rho_j^4 \rangle \int \phi_j(\mathbf{x})^4 d\mathbf{x} + \sum_{i,j} \langle \rho_i^2 \rangle \langle \rho_j^2 \rangle \int \phi_i(\mathbf{x})^2 \phi_j(\mathbf{x})^2 d\mathbf{x}. \quad (17)$$

In this paper, we restrict ourselves to use of the $\mathcal{L}_2$ -norm as a measure of error. An $\mathcal{L}_n$ -norm with $n > 2$ or other measures of goodness of fit could nevertheless prove useful in cases of rapid change of free energy with composition, for example, near a critical micellar concentration, or near a critical point for liquid-liquid phase separation.

To analyze further how measurement error affects the inferred free energy, we now make use of the fact that for typical measurement noise, $\langle \rho_j^2 \rangle$ will be closely proportional to the inverse of the measurement time at point j , τ_j .⁷ We now analyze this proportionality in more detail, in order to be able to make calculations of actual time requirements for measurement. To do so we make use of the total Rayleigh ratio, R_{tot} , which quantifies the efficiency with which light is singly scattered by a medium. For vertically polarized incident light and vertically polarized scattered light, detected in the plane perpendicular to the incident beam polarization, R_{tot} may be expressed as

$$(R_{tot})_{v,v}(\theta) = \frac{(P_{sca})_{v,v}(\theta)}{I_0} \frac{1}{\Delta\Omega_{det} V_{sca}}, \quad (18)$$

in which $(P_{sca})_{v,v}(\theta)$ is the power of the light scattered at angle θ from the incident beam direction into the detected solid angle $(\Delta\Omega)_{det}$, from the scattering volume V_{sca} , and I_0 is the irradiance of the input beam incident on the scattering volume. From Eq. (18) it is evident that R_{tot} has the dimensions of inverse length; we continue to denote the dimensionless excess Rayleigh ratio introduced above by R .

We now relate R_{tot} to a typical experimental situation. We suppose that the light scattering instrument records the numbers of photocounts n_i scattered during successive, equal time

increments $(\Delta t)_i$, and after subtracting a predetermined average dark count rate $\langle n_{dark} \rangle$, each is divided by a measured signal proportional to the irradiance (power per unit area) of the light incident on the scattering volume, $I_{0,i}$, during that same time period. Suppose this is done for a total time period $\tau_j = N_j \Delta t$. We assume that the quantities $((n_i - \langle n_{dark} \rangle)/(I_{0,i}))$ can be regarded as samples from a wide-sense stationary random process.⁷ We also assume that the instrument is in a suitable, linear operating range, such that the mean excess photocount number $E[n - \langle n_{dark} \rangle] = (\epsilon f_{opt}/h\nu)(P_{sca})_{v,v} \Delta t$, where ϵ is the quantum efficiency of the detector, f_{opt} is an optical factor of order 1, h is Planck's constant and ν is the frequency of the scattered light. Taking account of the possibility that the input intensity I_0 can itself fluctuate, but assuming that it does so on a time scale much larger than Δt , we can use Eq. (18) to relate R_{tot} to $E[n - \langle n_{dark} \rangle]$ as follows:

$$E[n_i - \langle n_{dark} \rangle] = \frac{\epsilon f_{opt} \Delta\Omega_{det} V_{sca} \Delta t}{h\nu} (R_{tot})_{v,v} I_{0,i} = \beta (R_{tot})_{v,v} I_{0,i}, \quad (19)$$

where we have defined

$$\beta = \frac{\epsilon f_{opt} \Delta\Omega_{det} V_{sca} \Delta t}{h\nu}. \quad (20)$$

In practice, it is convenient to determine β with use of measurements of a light scattering standard, such as toluene, for which the Rayleigh ratio has been previously measured at the appropriate wavelength. In so doing, it is important to correct for the fact that, depending on instrument design, both $\Delta\Omega_{det}$ and V_{sca} will typically depend on the index of refraction of the scattering medium. We also note that in Eq. (19) we have suppressed the dependence on θ . The present analysis is valid for scattering angles θ that probe concentration fluctuations of a wavelength sufficiently larger than molecular sizes or their spatial correlation ranges. A suitable range for θ will vary from system to system.

At measurement composition j the expectation value of the mean $\mathcal{M}_j$ of the N_j individual, normalized sample photocount numbers is

$$E[\mathcal{M}_j] = E \left[(1/N_j) \sum_{i=1}^{N_j} (n_i - \langle n_{\text{dark}} \rangle) / (I_{0,i}) \right] = \beta R_{j,\text{tot}}. \quad (21)$$

Consistent with Eq. (21), we use the quantity $E[\mathcal{M}_j/\beta]$ as an estimate of $R_{j,\text{tot}}$. The variance of this estimate, for composition j , is given by $\text{Var}[R_{j,\text{tot}}] = (1/\beta^2)\text{Var}[\mathcal{M}_j]$. To estimate $\text{Var}[\mathcal{M}_j]$, we assume that suitable measurement time intervals Δt have been chosen to be larger than the typical correlation time of the scattered light intensity, so that samples within different time intervals Δt are uncorrelated. Specifically, for $i \neq j$ we take $\langle ((n_i - \langle n_{\text{dark}} \rangle)/(I_{0,i}))((n_j - \langle n_{\text{dark}} \rangle)/(I_{0,j})) \rangle = \beta^2 R_{j,\text{tot}}^2$. For measurement composition j , we set the variance of the normalized sample numbers to be

$$\begin{aligned} \text{Var}[(n_i - \langle n_{\text{dark}} \rangle)/I_{0,i}] \\ = E[(n_i - \langle n_{\text{dark}} \rangle)/I_{0,i}]^2 - (\beta R_{j,\text{tot}})^2 = \sigma_j^2. \end{aligned}$$

Typically, for a dust-free sample, a beam that is stable for much longer than Δt , and scattered laser light that is well in excess of the dark count rate, one will further have the Poisson statistics case $\text{Var}[n_i] = E[n_i] = \beta R_{j,\text{tot}} I_{0,i}$, where Eq. (19) was used. This gives $\text{Var}[n_i/I_{0,i}] = (1/I_{0,i})^2 \text{Var}[n_i] = \beta R_{j,\text{tot}}/I_{0,i} = (\sigma_j^2)_{\text{Poisson}}$.

The variance of the mean of the normalized photocount numbers is now (e.g., Ref. 7)

$$\begin{aligned} \text{Var}[\mathcal{M}_j] = E \left[\left((1/N_j) \sum_i (n_i - \langle n_{\text{dark}} \rangle) / (I_{0,i}) \right) \right. \\ \left. - \beta R_{j,\text{tot}} \right]^2 = \sigma_j^2 / N_j = \sigma_j^2 \Delta t / \tau_j, \end{aligned} \quad (22)$$

and it follows that

$$\text{Var}[R_{j,\text{tot}}] = (1/\beta^2) \text{Var}[\mathcal{M}_j] = \frac{\sigma_j^2 \Delta t}{\beta^2 \tau_j}. \quad (23)$$

For Poisson counting statistics we now have

$$\begin{aligned} \text{Var}[R_{j,\text{tot}}]_{\text{Poisson}} &= \frac{R_{j,\text{tot}}}{(\beta I_{0,i} / \Delta t) \tau_j} \\ &= \frac{R_{j,\text{tot}}}{(I_{0,i} / h\nu) \varepsilon f_{\text{opt}} \Delta \Omega_{\text{det}} V_{\text{sca}} \tau_j} \\ &= \frac{R_{j,\text{tot}}}{\alpha \tau_j}, \end{aligned} \quad (24)$$

where $\alpha = (I_{0,i} / h\nu) \varepsilon f_{\text{opt}} \Delta \Omega_{\text{det}} V_{\text{sca}}$ is an instrument throughput parameter that is the product of the number of photons per second per unit area in the incident beam, the quantum efficiency of the detector, the optical loss factor, the detected solid angle, and the scattering volume. Hence α has dimensions *length/time*, consistent with dimensions $1/\text{length}$ for $R_{j,\text{tot}}$ and $1/\text{length}^2$ for $\text{Var}[R_{j,\text{tot}}]$.

To relate $\text{Var}[R_{j,\text{tot}}]$ to $\text{Var}[\rho_j]$ we take the scattering from the pure components to be well determined, so that $\text{Var}[R_{j,\text{tot}}] = \text{Var}[\Delta R_j]$, and use the relationship

$R(x, y) = \lambda^4 \Delta R(0) / (\pi^2 \Omega_0)$ between the dimensionless excess Rayleigh ratio $R(x, y)$, and ΔR . This gives

$$\begin{aligned} \text{Var}[\rho_j] &= \text{Var}[R_j] = \left(\frac{\lambda^4}{\pi^2 \Omega_0} \right)^2 \text{Var}[\Delta R_j] \\ &= \left(\frac{\lambda^4}{\pi^2 \Omega_0} \right)^2 \text{Var}[R_{j,\text{tot}}], \end{aligned} \quad (25)$$

which for Poisson statistics becomes

$$\text{Var}[\rho_j] = \left(\frac{\lambda^4}{\pi^2 \Omega_0} \right)^2 \frac{R_{j,\text{tot}}}{\alpha \tau_j} = \left(\frac{\lambda^4}{\pi^2 \Omega_0} \right) \frac{R_T(x_j, y_j)}{\alpha \tau_j}, \quad (26)$$

where in the last equality we have used the symbol $R_T(x_j, y_j) = \lambda^4 R_{j,\text{tot}} / (\pi^2 \Omega_0)$ to denote the dimensionless total Rayleigh ratio. In the simulations and for the associated figures, we use the dimensionless times

$$\tilde{\tau}_j = \frac{\pi^2 \Omega_0 \alpha \tau_j}{\lambda^4} = \frac{\tau_j}{\tau_{\text{inst}}}, \text{ and } T = \frac{\pi^2 \Omega_0 \alpha \sum_j \tau_j}{\lambda^4} = \frac{\sum_j \tau_j}{\tau_{\text{inst}}} \quad (27)$$

to stand for the time spent at point j and the total time for all of the measurements, respectively. The quantity τ_{inst} is discussed in Sec. IV B.

B. The instrumental time constant

Clearly, the quantity

$$\tau_{\text{inst}} = \frac{\lambda^4}{\pi^2 \Omega_0 \alpha} = \frac{\lambda^4}{\pi^2 \Omega_0 (I_{0,i} / h\nu) \varepsilon f_{\text{opt}} \Delta \Omega_{\text{det}} V_{\text{sca}}} \quad (28)$$

is the time needed for a given light scattering instrument, in a set configuration, to acquire data corresponding to a single unit of the dimensionless time T . τ_{inst} also yields the time needed for a given instrument to achieve a set value of the variance in the experimental estimate of the dimensionless light scattering intensity, for a given overall dimensionless intensity, for by combining Eqs. (26) and (27) we have

$$\text{Var}[\rho_j] = R_T(x_j, y_j) / \tilde{\tau}_j. \quad (29)$$

With use of values for the relevant instrumental parameters, τ_{inst} can be calculated directly from Eq. (28). Alternatively, τ_{inst} can be measured using a light scattering standard, such as toluene, that has a previously known Rayleigh ratio.^{8,9} To do so, it is convenient to note that Eqs. (19) and (28) can be combined to yield

$$\tau_{\text{inst}} = \frac{(\lambda^4 / (\pi^2 \Omega_0)) (R_{\text{tot}})_{v,v} \Delta t}{E[n_i - \langle n_{\text{dark}} \rangle]}. \quad (30)$$

As an example, for the commonly used light scattering standard toluene,⁸ estimating $\Omega_0 = 0.176 \text{ nm}^3$ and using $\lambda = 632.8 \text{ nm}$ and the corresponding value $(R_{\text{tot}})_{v,v} = 1.40 \times 10^{-5} \text{ cm}^{-1}$, we obtain

$$\tau_{\text{inst}} = \frac{0.129}{(E[n_i - \langle n_{\text{dark}} \rangle] \text{ for toluene}) / (\Delta t \text{ in seconds})}. \quad (31)$$

Now, a modest laboratory light scattering setup can very readily attain a count rate of $7 \times 10^4/\text{s}$ for a toluene standard, leaving headroom for higher intensity measurements. For such a

count rate, Eq. (31) yields $\tau_{inst} = 1.8 \mu\text{s}$. As will be seen in the analysis to follow, a value of $T = 10^9$ for the total dimensionless light scattering measurement time can give excellent results for the $\mathcal{L}_2$ -norm of the expected error in the composition dependence of the mixing free energy, given enough samples in a ternary liquid mixture composition triangle. $T = 10^9$ and $\tau_{inst} = 1.8 \mu\text{s}$ correspond to a half-hour of total light scattering measurement time, divided appropriately between the samples as described below.

C. Optimal distribution of measurement times

We now show that for a given choice of sample compositions $\mathbf{x}_j$, there is in principle an optimal distribution of the total measurement time T into the time devoted to measuring each sample, τ_j . While this distribution will not typically be known in advance for a given, new experimental system, it can nevertheless inform the basis for feedback that improves the efficiency of a measurement process.

First, Eqs. (23) and (24) show explicitly the inverse proportionality between $\langle \rho_j^2 \rangle$ and τ_j stated above. For the Poisson case, combining Eqs. (15), (24) and (29) we now have

$$\begin{aligned} E[(\mathcal{L}_2[f])^2] &= E\left[\int (f(\mathbf{x}))^2 d\mathbf{x}\right] = \sum_j \langle \rho_j^2 \rangle \lambda_j \\ &= \sum_j \frac{R_{j,tot}}{\alpha^* \tau_j} \lambda_j = \sum_j \frac{R_T(x_j, y_j)}{\tilde{\tau}_j} \lambda_j, \end{aligned} \quad (32)$$

in which $\alpha^* = (\pi^2 \Omega_0 / \lambda^4)^2 = (I_{0,i} / h\nu) \varepsilon_{f, opt} \Delta \Omega_{det} V_{sca} (\pi^2 \Omega_0 / \lambda^4)^2$. Note that α^* has dimensions $[1/\text{length}] \times [1/\text{time}]$. Since the Rayleigh ratio $R_{j,tot}$ has the dimensions $[1/\text{length}]$, α^* is thus clearly proportional to the rate at which a given instrumental setup can measure Rayleigh ratios.

Thus, the physical content of Eq. (32) is that the contribution to the error in the free energy resulting from each measurement is proportional to the integral of the square of the tent function corresponding to the chosen interpolation method, times the variance of the measurement of the dimensionless excess Rayleigh ratio at that point.

In Eq. (32), the factors contributed by each measurement point to $E[\int (f(\mathbf{x}))^2 d\mathbf{x}]$, f_j and λ_j , have differing origins that are important for the analysis below. First, the unknown free energy enters into $\langle \rho_j^2 \rangle$ in Eq. (32) only through $R_{j,tot}$, and $\langle \rho_j^2 \rangle$ is inversely proportional to the instrumental setup parameter α^* and the time τ_j to measure sample j . On the other hand, the quantity λ_j , the integral of the square of $\phi_j(\mathbf{x})$, the free energy consequence of the interpolation unit-height tent function $T_j(\mathbf{x})$, depends on the free energy $g(\mathbf{x})$, the dielectric coefficient $\varepsilon(\mathbf{x})$, and the interpolation method chosen, but not on the instrument setup. While in practice one can typically vary α^* for convenience, for simplicity we take α^* to be fixed.

For a total measurement time, T , that is regarded as fixed in advance, and for fixed α^* , Eq. (32) enables one to formally derive the optimal measurement time proportions $(\tau_j)_{opt}/T$ to spend at each of a chosen set of measurement compositions $\mathbf{x}_j$, to minimize $E[(\mathcal{L}_2[f])^2]$. The result is formal since it is expressed in terms of quantities such as λ_j that are not fully known in advance. However, such an optimization is never-

theless helpful, since it shows that the mathematical aspects of the process of optimizing experimental design can be separated into time optimization and composition optimization. Specifically, with use of Lagrange multipliers one finds that

$$(\tau_j)_{opt} = T \frac{\sqrt{R_{j,tot} \lambda_j}}{\sum_k \sqrt{R_{k,tot} \lambda_k}}, \quad (33)$$

and that the time-optimized expectation value of the square of the $\mathcal{L}_2$ -norm of f is

$$E[\mathcal{L}_2[f]^2]_{opt} = \frac{1}{\alpha^* T} \sum_k \sqrt{R_{k,tot} \lambda_k}. \quad (34)$$

Equations (10) through (34) form a general basis for analysis of the expected experimental error in light scattering determination of the mixing free energy, that is largely independent of the number of mixture components. We now analyze and simulate λ_k and $E[\mathcal{L}_2[\delta g]]$ for binary and ternary mixtures and address the question of optimizing measurement locations.

D. Specific error considerations for binary mixtures

The case of binary mixtures is particularly instructive, since in the one-dimensional, binary mixture case, the relationship between the excess Rayleigh ratio R and the free energy simplifies to a second-order, ordinary differential equation,

$$g_{xx} = \frac{\varepsilon_x^2}{R}, \quad g(0) = g(1) = 0. \quad (35)$$

The linearized equation for binary mixtures can be found either by perturbing the equation above, or by working from the expressions below Eq. (5), making use of the fact that g_{yy} is asymptotic to $1/y$ as one approaches the x -axis to show that $a = R/g_{xx} = (R/\varepsilon_x)^2$, $b = 0$, and $c = 0$ in the expressions below Eq. (5). The result is

$$\left(\frac{R(x)}{\varepsilon_x(x)}\right)^2 f_{xx} = -\rho, \quad f(0) = f(1) = 0. \quad (36)$$

As a consequence of the simplicity of Eq. (36), many of the steps given formally above in the general case can be carried out analytically in the case of binary mixtures, as we now proceed to do.

Consider a grid of light scattering measurement points $x_0 < x_1 < \dots < x_n$, where $x_0 = 0$ and $x_n = 1$, and with $n - 1$ measurements taken at the points $x_1, x_2, \dots, x_{n-1}$. We illustrate the application of the analysis above by considering the simplest, linear interpolation approach here. To this end, we seek an approximate solution of the linearized equation by replacing the right-hand side by its linear interpolant

$$-\sum_{j=1}^{n-1} \rho_j T_{j,lin}(x),$$

in which $T_{j,lin}$ is the piecewise linear function that is 1 at x_j , 0 at neighboring grid points, and nonzero between x_j and neighboring grid points, and 0 otherwise. We also assume that the grid is fine enough that we can approximate $R(x)/\varepsilon_x(x)$ by $R(x_j)/\varepsilon_x(x_j) = (R/\varepsilon_x)_j$ throughout the range of x for which each

linear interpolant is nonzero. Then, as above letting $\phi_{j,lin}(x)$ be the function which satisfies

$$\left(\frac{R(x_j)}{\varepsilon_x(x_j)}\right)^2 (\phi_{j,lin})_{xx} = -T_{j,lin}(x), \quad \phi_{j,lin}(0) = \phi_{j,lin}(1) = 0$$

and integrating Eq. (36) twice yields

$$f(x) = \sum_{j=1}^{n-1} \rho_j \phi_{j,lin}(x) = \sum_{j=1}^{n-1} \left(\frac{\varepsilon_x}{R}\right)_j^2 \rho_j \tilde{\phi}_{j,lin}(x), \quad (37)$$

in which $\phi_{j,lin}(x) = (\varepsilon_x/R)_j^2 \tilde{\phi}_{j,lin}(x)$. The expressions for the $\tilde{\phi}_{j,lin}(x)$ in the case of the linear interpolation being described here are given in the Appendix. We now calculate the $\mathcal{L}_2$ -norm of the deviation, assuming as above that the ρ_j are independent random variables each with mean 0, to obtain

$$\left\langle \int_{x_0}^{x_n} (f(x))^2 dx \right\rangle = \sum_{j=1}^{n-1} \left(\frac{\varepsilon_x^4}{R^4}\right)_j \langle \rho_j^2 \rangle \tilde{\lambda}_{j,lin}, \quad (38)$$

in which $\tilde{\lambda}_{j,lin} = \int_{x_0}^{x_n} \tilde{\phi}_{j,lin}^2(x) dx$. With use of $\langle \rho_j^2 \rangle = R_T(x)/\tilde{\tau}_j$ from Eq. (29), this yields the following expression for the expected value of the square of the $\mathcal{L}_2$ norm of the error f ,

$$\left\langle \int_{x_0}^{x_n} (f(x))^2 dx \right\rangle = \sum_{j=1}^{n-1} \left(\frac{\varepsilon_x^4}{R(x)^4}\right)_j \left(\frac{R_T(x_j)}{\tilde{\tau}_j}\right) \tilde{\lambda}_{j,lin}. \quad (39)$$

To proceed further towards using Eq. (39) for making numerical estimates of the required times for accurate free energy measurement, we have first made a convenient, mathematically simple assumption about the relationship between the dimensionless excess Rayleigh ratio $R(x)$ and the dimensionless total Rayleigh Ratio $R_T(x)$, namely, that $R(x) = R_T(x)$. Such a situation could be close to the case for a macromolecule in aqueous solution or near a critical point, but is unlikely for a mixture of small organic molecules well away from criticality. An analysis such as that given below could of course be performed under various other assumptions, specific to particular experimental systems.

With the assumption $R(x) = R_T(x)$, we now have

$$\left\langle \int_{x_0}^{x_n} (f(x))^2 dx \right\rangle = \sum_{j=1}^{n-1} \left(\frac{\varepsilon_x^4}{R(x)^3}\right)_j \left(\frac{\tilde{\lambda}_{j,lin}}{\tilde{\tau}_j}\right). \quad (40)$$

The Lagrange multiplier method may now be used to minimize the net error subject to the constraint $T = \sum_{j=1}^{n-1} \tilde{\tau}_j$. The net error is minimized by taking

$$\tilde{\tau}_j = T \frac{\left(\frac{\varepsilon_x^2}{(R(x))^{3/2}}\right)_j \sqrt{\tilde{\lambda}_j}}{\sum_{k=1}^{n-1} \left(\frac{\varepsilon_x^2}{(R(x))^{3/2}}\right)_k \sqrt{\tilde{\lambda}_k}}. \quad (41)$$

Then, the optimal value of the integral in Eq. (39) is given by

$$\int_{x_0}^{x_n} (f(x))^2 dx = \frac{1}{T} \left(\sum_{j=1}^{n-1} \left(\frac{\varepsilon_x^2}{(R(x))^{3/2}}\right)_j \sqrt{\tilde{\lambda}_j} \right)^2.$$

E. Additional error considerations for ternary mixtures

For ternary mixtures the determination of the λ_j involves the integration of Eq. (5), starting from an appropriate tent-like interpolation function $T_j(\mathbf{x})$ placed at each light scattering measurement point, to produce a corresponding toothlike functions like those depicted in the figures. Then one squares and integrates the toothlike functions $\phi_j(\mathbf{x})$ to evaluate λ_j , as noted above. As described above and shown in the figures, the shapes of the λ_j are sensitive to composition, molecular interactions, and dielectric gradient direction.

Figure 5 shows examples of the resulting dependence of λ_j on composition, changes in molecular interaction strengths, and dielectric gradient direction, with use of the regular solution model free energy applied to linear interpolation on an evenly spaced ternary mixture measurement grid. The large values of λ_j near certain binary mixture edges imply that excellent data on those binary mixtures is a pre-requisite for a good determination of the ternary mixture free energy. However, which binary axes are most crucial for this purpose depends not only on molecular interactions, as shown by comparing the left-most and center plots, but also on the direction of $\nabla\varepsilon$, as shown by the comparing the left-most and right-most plots.

The complicated nature of the ternary λ_j implies that it is very useful to know something in advance about the system of interest, in order to plan an efficient set of experiments to determine the ternary free energy accurately. While this is indeed the case, we have also found by numerical integration that there are approximate scaling relations between the values of λ and the local dimensionless excess Rayleigh ratio. For ternary mixtures, a typical such scaling relationship is depicted in Fig. 6 (right). Since numerical integration to find ternary λ_j proved to be costly in computation time, we used this approximate scaling relationship between ternary λ_j and local excess Rayleigh ratio to perform the simulations giving error estimates for ternary free energies using light scattering, that are described further below and are shown in Fig. 12.

V. RESULTS: DEPENDENCE OF FREE ENERGY ACCURACY ON MEASUREMENT TIME AND COMPOSITION SAMPLING

In this section, we quantify the uncertainty in inferred free energy due to use of a finite number of light scattering samples measured with noise, for both binary and ternary mixtures. To do so we use Monte Carlo simulation of light scattering from mixtures that are assumed to be well modeled by the regular solution free energy.

The basic result of the simulations is that there are distinct regions of sample number and total measurement time space, (m, T) , in which interpolation error or light scattering noise limit free energy accuracy. For each interpolation method, the boundary between these regions is at the optimal value of m for fixed T , while at fixed m the boundary gives a value of T beyond which free energy accuracy would not increase. Note that T is dimensionless; the actual measurement

FIG. 6. Approximate power-law dependence of $\sqrt{\lambda}$ on excess light scattering intensity, for binary (left, $m = 20$) and ternary (right, $m = 10$; $N = 63$) ideal solutions. The exact binary mixture relation is given in Sec. 4 of the Appendix by Eq. (A8), and the exponent near -1 is consistent with Eq. (A10). The more approximate ternary scaling is likely due to the more complicated ternary free energy errors given by Eq. (5) and illustrated in Figs. 2–4. With Eq. (33), such approximate scaling laws for $\sqrt{\lambda}$ can be useful for planning of experiments.

time is $\tau_{inst}T$ for a fixed instrument configuration, where τ_{inst} can be obtained from Eqs. (28) through (31).

The curves shown in Fig. 7 illustrate the transition between interpolation-dominated and light scattering-dominated noise. The average of the square of the $\mathcal{L}_2$ -norm of the error in the free energy, $\langle(\mathcal{L}_2[\delta g])^2\rangle$, is plotted vertically vs. m , the number of different compositions at which the Rayleigh ratio was measured. Composition samples were evenly spaced for generating Fig. 7; optimizing sample spacing can dramatically improve accuracy as illustrated in Fig. 10 and discussed below.

Each curve in Fig. 7 is for a fixed T , and the time spent on each sample followed the optimal distribution of Eq. (41). Each point on a given curve is an average of 20 simulation runs. For each run, a sample photon count at each composi-

tion was chosen from a normal distribution with mean R_j and standard deviation $\sqrt{R_j/\bar{\epsilon}_j}$, following Eq. (29), giving us that run's estimate of the Rayleigh ratio for that particular composition. Then, the data were interpolated onto a much finer grid of 4000 points in order to yield a negligible computational error and used to solve Eq. (35). Finally, we estimated the error from the integral

$$\begin{aligned} \langle(\mathcal{L}_2[\delta g])^2\rangle &= \langle\|g - g_{\text{exact}}\|_2^2\rangle \\ &= \left\langle \iint_{\text{triangle}} [g(x, y) - g_{\text{exact}}(x, y)]^2 dx dy \right\rangle. \end{aligned} \quad (42)$$

Since the normal distribution does not capture the correct photon counting situation when $\sqrt{R_j/\bar{\epsilon}_j}$ becomes an appreciable fraction of R_j , the runs for $T = 10^4$ and $T = 10^6$ only include m values for which the maximum value of $\sqrt{R_j/\tau_j}$ was less than $0.05R_j$.

The canonical form of the constant- T expected free energy error curves is as follows. For low values of m interpolation error dominates, as the measurement error for each sample is negligible. This means, in essence, that at such low values of m one would be wasting time getting unnecessarily precise Rayleigh ratios, time that could be better spent measuring other compositions. As m increases, the free energy error curve suddenly branches from the interpolation curve and stays flat. Beyond this branch or boundary point $m = m_b$, intrinsic noise in the experimental light scattering measurements overtakes interpolation error as a source of free energy error. Using more samples than m_b would not hurt free energy accuracy, but would mean that sample material and preparation time, rather than measurement time, may be being wasted. Of course, in real experiments one may well want more samples to aid in checking for systematic errors, but such errors are so specific to given setups and procedures that it is impractical to evaluate them quantitatively in this general setting.

For $m > m_b$, the loss in precision in the measurement of each Rayleigh ratio is compensated by the improvement

FIG. 7. Canonical free energy error norm, $\langle\mathcal{L}_2^2[\delta g]\rangle$, vs. number of samples, m , and dimensionless measurement time, T , illustrated here for linear interpolation of binary mixture light scattering simulated data from evenly spaced concentration samples. Actual measurement time is $\tau_{inst}T$ (Eqs. (27)–(31)). Dashed lines show theoretical asymptotic values from Eq. (45). The lower red curve is for hypothetical infinite time, for which free energy error is due solely to integrated interpolation error. As m increases at fixed T , interpolation error first dominates and the error norms follow the infinite-time curve, then branch off at $m = m_b$ and flatten when light scattering noise dominates. Thus, at fixed T , too few samples ($m < m_b$) will not make good use of available time, while too many samples $m > m_b$ will not reduce $\langle\mathcal{L}_2^2[\delta g]\rangle$. Parameters: $C = 1.5$, $\epsilon_x = \sqrt{3}/2$.

in coverage one obtains by finer sampling of the composition space, and these factors combine to make the constant T free energy error curves flat. To find the corresponding asymptotic values of the free energy error for large m , we first note that the form of the binary mixture λ integrals given by Eq. (A6) becomes a better approximation at the smaller Δx values that result when m increases. With use of this approximation, we note further that $\sqrt{\lambda} \propto \Delta x = (1/m)$, while for fixed T , one also has $\tau_j \propto (1/m)$. Thus, for sufficiently large m , $\langle (\mathcal{L}_2[\delta g])^2 \rangle$ becomes independent of m , as observed, and furthermore is given by the integral

$$\langle (\mathcal{L}_2[\delta g])^2 \rangle = \frac{1}{3\alpha T} \left(\int_0^1 \frac{\varepsilon_x^2(x)x(1-x)}{(R(x))^{3/2}} dx \right)^2. \quad (43)$$

For the regular solution model (see Sec. 4 of the Appendix just under Eq. (A8)), we have

$$R(x) = \frac{\varepsilon_x^2 x(1-x)}{1-2Cx(1-x)}, \quad (44)$$

in which case, Eq. (43) becomes

$$\langle (\mathcal{L}_2[\delta g])^2 \rangle = \frac{1}{3\alpha T} \left(\int_0^1 \frac{[1-2Cx(1-x)]^{3/2}}{\varepsilon_x(x)\sqrt{x(1-x)}} dx \right)^2.$$

For the ideal solution free energy ($C = 0$) with ε_x constant, the integral can be evaluated analytically to yield

$$\langle (\mathcal{L}_2[\delta g])^2 \rangle = \frac{\pi^2}{3\alpha T \varepsilon_x^2}.$$

Observe that the integral decreases as C increases from 0. The plots in Fig. 7 correspond to the regular solution model with $C = 1.5$ and ε_x constant, in which case we numerically evaluate the integral to obtain

$$\langle (\mathcal{L}_2[\delta g])^2 \rangle = \frac{2.7523}{3\alpha T \varepsilon_x^2}.$$

With $\alpha = 1$ and $\varepsilon_x = \sqrt{3}/2$, we obtain the expression used for Fig. 7,

$$\log_{10} \langle (\mathcal{L}_2[\delta g])^2 \rangle = \log_{10} \left(\frac{1.2232}{T} \right). \quad (45)$$

Figure 7 shows excellent agreement between Eq. (45) and the asymptotic values of the simulated $\langle (\mathcal{L}_2[\delta g])^2 \rangle$.

Figure 8 shows that high order data interpolation dramatically reduces the time and samples needed to obtain a given free energy accuracy, shown here for binary mixtures. Here we use the $\mathcal{L}_2$ -norm itself, given by

$$\begin{aligned} \langle \mathcal{L}_2[\delta g] \rangle &= \langle \|g - g_{\text{exact}}\|_2 \rangle \\ &= \left\langle \left[\iint_{\text{triangle}} [g(x, y) - g_{\text{exact}}(x, y)]^2 dx dy \right]^{1/2} \right\rangle. \end{aligned} \quad (46)$$

The gray points show the scatter in the simulations of free energy error, to indicate the range of results for individual experiments.

The simulations yield power laws for the dependence of $\langle \mathcal{L}_2[\delta g] \rangle$ on m (Fig. 9, left) and $\langle \mathcal{L}_2[\delta g] \rangle$ on T (Fig. 9, right)

FIG. 8. Higher order data interpolation considerably reduces time and sample numbers. $\langle \mathcal{L}_2[\delta g] \rangle$ vs. m for linear (red), 3rd-degree Lagrange (green), and 7th-degree Lagrange (blue) interpolation, for binary mixtures. Lower solid curves show interpolation error alone. The vertical dotted-dashed line is the $\langle \Delta \mathcal{L}_2 \rangle$ vs. m cut of Fig. 9 (right). Gray dots show representative scatter of $\mathcal{L}_2[\delta g]$ for 20 runs at each m used to compute $\langle \mathcal{L}_2[\delta g] \rangle$.

for interpolation- and light-scattering-noise-limited cases discussed above. By combining these power laws, we derive approximate formulas for the optimal sample numbers m_b for a given T . Specifically, we obtain

$$m_b(\text{linear, binary}) = 0.4445 T^{0.2679},$$

$$m_b(\text{3rd degree Lagrange, binary}) = 1.4304 T^{0.1179}, \text{ and}$$

$$m_b(\text{7th degree Lagrange, binary}) = 3.8771 T^{0.06077}.$$

We have found that optimizing sample concentrations using the Nelder-Mead method dramatically increases free energy accuracy for a given measurement time T and/or sample number m . In the example shown in Fig. 10, accuracy increases by a factor of one hundred. The corresponding alterations in sample concentration are shown in Fig. 10 (right).

We have also investigated the dependence of the time required to deduce the free energy from light scattering on the magnitude of effective intermolecular interactions, as modeled by the regular solution model. Figure 11 shows that light-scattering free energy accuracy is system- and temperature-dependent. The figure shows that accuracy is especially sensitive to C , the regular solution mixing parameter, for high-order interpolation at low error values. Thus the present estimates of accuracy will need modification in order to apply to systems in which the free energy has more intricate dependence on concentration than is observed in mixtures approximately described by the regular solution model. As an example, the dashed green line shows that $m = 20$ evenly spaced samples and 3rd-degree Lagrange interpolation can yield integrated error less than 0.001, regardless of C in the range shown. On the other hand, for 7th-degree Lagrange interpolation, there is a more dramatic dependence on C , but this occurs for lower free energy error values. The right-hand plot shows that optimal sample numbers m_b for given free energy accuracy also depend on C and increase near the critical value $C = 2$ for liquid-liquid phase separation.

Figure 12 shows the expected ternary mixture error norms vs. sample index m from simulations of light scattering data from $N = ((m+2)(m+1)/2) - 3$ evenly spaced mixture

FIG. 9. Power laws for $\langle \mathcal{L}_2[\delta g] \rangle$ vs. m (left) and $\langle \mathcal{L}_2[\delta g] \rangle$ vs. T (right) for interpolation- and light scattering noise-limited cases, respectively, from simulations. (Left) Power-law exponents of -1.9 , -4.3 , and -8.3 closely describe, respectively, the linear (red), 3rd-degree Lagrange (green), and 7th-degree Lagrange (blue) infinite-time $\langle \mathcal{L}_2[\delta g] \rangle$ vs. m curves of Fig. 8. (Right) Light scattering noise-limited $\langle \mathcal{L}_2[\delta g] \rangle \propto 1/\sqrt{T}$ along blue dotted-dashed line in Fig. 8, consistent with Eq. (34). Parameters as in Fig. 7.

FIG. 10. Optimizing binary sample concentrations dramatically increases free energy accuracy. (Left) Free energy error norms vs. m for [uniform] and (optimal) sample spacing using data interpolation that is linear [red] (magenta), 3rd-degree [light green] (dark green), and 7th-degree [blue] (turquoise). (Right) Original even spacings (blue) on the horizontal axis for $m = 16$, $T = 10^{12}$, 7th-degree interpolation (blue circle, left); the vertical axis shows concentrations (turquoise) after Nelder-Mead optimization, which leads to a hundred-fold reduction in $\langle \mathcal{L}_2[\delta g] \rangle$, from $10^{-3.9}$ to $10^{-5.9}$ (left). Parameters and even spacing results are as in Fig. 7.

FIG. 11. Light-scattering free energy accuracy is dependent on molecular interaction strengths, especially for high-order interpolation at low error values. (Left) Binary mixture interpolation-limited $\langle \mathcal{L}_2[\delta g] \rangle$ vs. regular solution interaction parameter C for linear (red), 3rd-degree Lagrange (green) and 7th-degree Lagrange (blue) interpolation, with and without concentration grid optimization. (Right) Optimal sample numbers m_b (circled) for given free energy accuracy also depend on C . Asymptotically $\langle \mathcal{L}_2[\delta g] \rangle \approx 10^{-5}$ at $T = 10^{12}$ using 7th-degree Lagrange data interpolation, but $m_b \approx 8$ for $C = 0.5$, $m_b \approx 20$ for $C = 1.5$, and $m_b \approx 32$ for $C = 1.95$, the latter being near the critical value $C = 2$ for liquid-liquid phase separation.

FIG. 12. Ternary mixture free energy error norms vs. sample index m , from simulations of light scattering data from $N = ((m + 2)(m + 1)/2) - 3$ evenly spaced mixture samples. $\langle \mathcal{L}_2[\delta g] \rangle$ vs. m ; see Eq. (46), for linear, quadratic, and cubic interpolation. Lower curves are limited by interpolation error. Optimal sample number estimates are given in the text. Each colored point averages 20 runs; black dots show representative scatter. Parameters: $C_{xy} = 0$, $C_{xz} = 1$, $C_{yz} = 0$, $\nabla \varepsilon = (\sqrt{3}/2, 1/2)$.

samples. As an example from Fig. 12, the cubic interpolation (blue) point near $m = 5$, or $N = 18$ and $T = 10^8$ corresponds to $\langle \mathcal{L}_2[\delta g] \rangle = 10^{-2.5} = 0.0032$. Therefore, with $\tau_{inst} = 1 \mu s$ and cubic interpolation, the measurement time $\tau_{inst}T$ to obtain δg to within three parts in a thousand is 100 s, about 5 s per sample. With use of Eq. (28) for the instrumental factors that determine τ_{inst} , one can design instruments to lower τ_{inst} from this value in a straightforward manner and thereby speed free energy measurement.

We have also compared our results for the $\mathcal{L}_2$ -norm with those obtained by using a norm that measures the fractional deviation of the inferred free energy from its true value, given by

$$\begin{aligned} \langle \mathcal{L}_2[\delta g/g_{\text{exact}}] \rangle &= \left\langle \left\| \frac{g - g_{\text{exact}}}{g_{\text{exact}}} \right\|_2 \right\rangle \\ &= \left\langle \left[\iint_{\text{triangle}} \left[\frac{g(x, y) - g_{\text{exact}}(x, y)}{g_{\text{exact}}(x, y)} \right]^2 dx dy \right]^{1/2} \right\rangle. \end{aligned} \quad (47)$$

FIG. 13. Power laws for interpolation-limited (left) and light scattering noise-limited (right) free energy error, for ternary mixtures, analogous to those for binary mixtures shown in Fig. 9. The right-hand plot is for $m = 40$ along the blue vertical dashed-dotted line in Fig. 12.

We find that the curves of $\langle \mathcal{L}_2[\delta g/g_{\text{exact}}] \rangle$ vs. m are qualitatively similar to those shown in Fig. 12, but its numerical values for a given m are slightly higher than those of $\langle \mathcal{L}_2[\delta g] \rangle$.

Figure 13 shows the power laws for interpolation-limited (left) and light-scattering-noise-limited (right) free energy error, for ternary mixtures, analogous to those for binary mixtures shown in Fig. 9. As was done above for binary mixtures, these power laws can be solved simultaneously to yield estimates for the optimal sample numbers for a given dimensionless measurement time. The results are

$$m_b(\text{linear, ternary}) = 0.03125T^{0.3790},$$

$$m_b(\text{quadratic, ternary}) = 1.3187T^{0.1247}, \text{ and}$$

$$m_b(\text{cubic, ternary}) = 1.101T^{0.1098}.$$

Just as for binary mixtures, these scaling relationships can provide useful guides for the design of practical experiments for determining ternary mixture free energies with use of light scattering.

VI. DISCUSSION

The essential physical consideration underlying the present work is that light scattering does not determine the free energy at a single composition point, but instead gives information about the free energy in neighborhoods of a point. This is a simple consequence of the fact that the light scattering intensity can be expressed in terms of a nonlinear combination of the second compositional derivatives of the free energy.

As a consequence, the goal of determining free energies of mixing using light scattering intensity measurements has led to a novel experimental design situation that involves integration of a nonlinear partial differential equation to go from the data to the quantity to be determined.¹ We have set up a framework for the appropriate error analysis by first linearizing Eq. (1) to obtain a PDE that governs the diffusion of information from light scattering to the free energy. The flow of information is governed by a set of characteristic curves, which are solutions of a set of ordinary differential equations,

that determine spacelike and timelike information diffusion directions.

The analysis of design for light scattering determination of the free energy leads to error expressions and scaling relationships that permit optimizing sample number, sample concentrations, and measurement times so as to attain a desired free energy accuracy. A primary result of this analysis is the identification of a canonical form of the relationship between free energy error, number of samples, and measurement time.

Specifically, for each dimensionless measurement time T there is an optimum number of samples m_b below which time is wasted and above which sample is wasted. Approximate power-law relationships between m_b and T were determined for evenly spaced samples, in the case of solutions well modeled by the regular solution free energy. In the process of experimental planning, these power-law relationships, while approximate, can nevertheless provide a useful estimate of the needed number of samples, which can of course be refined following selected light scattering observations.

Continuing along this line of reasoning, the sensitivity of accuracy to choice of sample locations shown in Fig. 10, combined with the sensitivity of accuracy to molecular interactions shown in Fig. 11, underlines a result that is familiar to experimentalists, namely, that feedback is essential or at least can dramatically improve efficiency during almost any measurement process, including the one envisioned here of determining the free energy using light scattering. In this work we have not attempted the mathematical analysis of an efficient experimental design feedback mechanism, though we expect that it will be practical to do so. Furthermore, while we have assumed that the instrument throughput parameter τ_{inst} is fixed, this assumption is neither realistic, nor necessary, and remains to be examined.

A number of extensions of the present method would be useful. First, in real experimental mixtures it will be useful to determine free energies for liquid mixtures when the isotropic liquid phase covers only part of the ternary composition triangle. We envision that this problem can be tackled with use of light scattering at multiple wavelengths, or a combination of light scattering with x-ray and/or neutron scattering, or a combination of light scattering with other means of determining the free energy along certain cuts within the ternary triangle, techniques including osmometry, analytical ultracentrifugation, and equilibrium vapor-phase composition determination.

In this connection, we note that the present work has not examined the error analysis for determining derivatives of the free energy by light scattering. Such an analysis for the accuracy of chemical potentials will be important not only for comparison with these other methods, but also for determining driving forces for chemical reactions and diffusion, as well as for helping to constrain the properties of nearby phase boundary equilibria.

To accurately measure the free energies of mixtures of biological macromolecules has major importance for understanding both normal physiology and disease. The difference between health and disease, or for that matter even life and

death, can hinge on interaction strength differences corresponding to single amino-acid substitutions. Such substitutions often correspond to free energy differences that are quite small, comparable to thermal energy $k_B T$ or even less, and need to be measured.

In view of this importance, extending the present method to apply to selected portions of the composition phase diagram would be particularly useful for many systems relevant to biology and physiology, including micellar and microemulsion solutions, and mixtures of biological macromolecules, such as protein and DNA, that are ubiquitous within living cells. Biological macromolecular solutions are not liquid in all proportions and are also often at high concentrations at which dilute solution thermodynamics does not accurately apply, exactly a situation for which accurate free energy measurements are needed to help construct a truly molecular understanding.

It would also be useful to extend the present method so as to determine ternary composition tie lines and tie triangles. We have previously shown¹ that with prior knowledge of the compositional boundaries (cloud-point loci) between single-phase and multi-phase equilibrium coexistence of isotropic liquid phases, it is possible to use light scattering in the single-phase region to deduce the tie lines and tie triangles. The present error analysis has not addressed the accuracy of the tie lines that can in principle be so deduced with use of the combination of measured cloud points and real light scattering single-phase data, both measured with noise. The possibility of extending the method to actually determine the tie lines without use of the cloud points is intriguing and remains to be examined. In particular, it is conceivable that the chemical potential equalities required for multi-phase equilibrium coexistence can be grafted onto the present nonlinear PDE in the spirit of a “shock” condition, in the language of PDE study.

Multicomponent mixtures occur in a very broad variety of systems of interest in biology, chemistry, chemical engineering, and beyond. Thus it would quite clearly be useful to extend the method to quaternary and higher order mixtures. We have found that the general equation given above, Eq. (3), applies to such mixtures, as noted above, and that a parabolic partial differential equation similar to Eq. (5) is in general the linear form of such an equation. Thus we expect that the present analysis will go through in higher dimensions with suitable modification.

Finally, it would be useful to extend the method to cover a variety of other, related situations, each of which requires a mathematical reformulation of the basic problem. One example would be mixtures that do not satisfy the additive volume approximation, which include mixtures important in supercritical fluid technology. Another would be extending the method to apply to effectively two-dimensional mixtures such as those within the membranes of living cells, in which liquid-liquid criticality is being intensively studied. The idea in both cases is to continue to exploit the same basic physical relationship, namely, the well-established connection between the magnitude of spontaneous local composition fluctuations and second composition derivatives of thermodynamic potentials.

VII. SUMMARY AND CONCLUSIONS

The present method lays the groundwork for high-throughput means of determining multicomponent free energies directly and in a model-independent fashion, by the non-invasive technique of static light scattering. Once samples are ready, the actual light scattering measurement time needed to obtain a very accurate mixing free energy of a single-phase isotropic, ternary liquid mixture, over its entire composition triangle, is just minutes for a modest laboratory light scattering setup, and with dedicated instrumentation could be seconds or less.

The ability to determine the Gibbs free energies of multicomponent mixtures rapidly could considerably aid physical chemical investigation of systems ranging from the macromolecular mixtures found inside living cells, to polymer solutions, microemulsion and micellar solutions, and liquid mixtures of small molecules important in an extremely broad array of contexts. Experimental investigations of multicomponent mixtures at equilibrium have long shown that the whole is not the sum of the parts, and that one should expect surprises. To work towards understanding the molecular origin of the observed equilibrium behavior one needs to know the Gibbs free energy. Not only does the Gibbs free energy determine the probability distribution of spontaneous local composition fluctuations as used here in the context of relating light scattering to free energy, it also determines the chemical potentials of each constituent, that drive chemical change and spatial diffusion, and sets the position and nature of the phase boundaries and critical points. Experimental measurement of the Gibbs free energy can also help to test more fundamental theories for the statistical-thermodynamic partition function of the system, from which the free energy follows.

Furthermore, determining the Gibbs free energy experimentally by light scattering or by any other means has the great appeal that it provides model-independent results that can stand the test of time, and thus serve as a beacon for the development and continued refinement of detailed molecular theories of its origin. As an analogy, the boiling temperature of water at atmospheric pressure has been known for centuries, but predicting it to the known experimental accuracy directly from the Schrödinger equation, while conceivable in principle for a large collection of water molecules, remains to be accomplished, owing to the extraordinary computational and theoretical difficulties involved in dealing with so many electrons.

In essence the present work aspires to help create a new kind of imaging technology. Traditionally imaging refers to the determination of spatially varying structures in two- or three-dimensional space. But the conceptual spaces needed for understanding molecular liquid mixtures extend far beyond three dimensions and require considerable work to construct. A small part of the needed effort can be aided by rapidly creating what are effectively images of the multidimensional Gibbs free energy, a function that summarizes a large part of the how and why of the molecular basis of macroscopic properties.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH) Grant No. EY018249. We thank Michael Margitus for his help with Nelder-Mead calculations.

APPENDIX: FURTHER MATHEMATICAL DETAILS

1. Multicomponent generalization of Eq. (3)

Theorem 1: Suppose that H is a symmetric positive definite $N \times N$ matrix and suppose that q is a nonzero N -vector. For any nonzero vector p that is not orthogonal to q , $\frac{p^T H p}{p^T (q q^T) p} \geq \frac{1}{q^T H^{-1} q}$.

Proof. Because H is symmetric and positive definite it has a well-defined positive square root, $\sqrt{H}$. The inequality $(p^T H p)(q^T H^{-1} q) \geq (p^T (q q^T) p)$ is equivalent to the Schwarz inequality applied to the dot product of $\sqrt{H} p$ and $\sqrt{H^{-1}} q$. $\square$

Note that if $p = H^{-1} q$, then

$$\frac{p^T H p}{p^T (q q^T) p} = \frac{1}{q^T H^{-1} q}.$$

More important for our purpose, by letting H be the Hessian of G and letting $q = \nabla \varepsilon$ we find that we have established that

$$\min \frac{G_{ss}}{\varepsilon_s^2} = \frac{1}{\nabla \varepsilon^T H^{-1} \nabla \varepsilon}.$$

2. Analysis of the linearized light scattering PDE

We now study the nature of the diffusion of information about the free energy further, so as to identify and characterize a suitable spacelike direction s , a timelike direction t , and the corresponding diffusion coefficient. To do so, we first seek a change of independent variables to cast Eq. (5) into the corresponding canonical form for a parabolic second-order linear PDE, $f_{ss} = u(s, t, f, f_s, f_t)$. With $s = s(x, y)$ and $t = t(x, y)$, Eq. (5) becomes

$$\begin{aligned} & (as_x^2 - 2bs_xs_y + cs_y^2)f_{ss} + 2(as_xt_x - b(s_xt_y + s_yt_x) \\ & + cs_yt_y)f_{st} + (at_x^2 - 2bt_xt_y + ct_y^2)f_{tt} \\ & + (as_{xx} - 2bs_{xy} + cs_{yy})f_s + (at_{xx} - 2bt_{xy} + ct_{yy})f_t \\ & = -\rho(s, t), \end{aligned}$$

from which the canonical form results provided that s and t satisfy

$$\begin{aligned} as_xt_x - b(s_xt_y + s_yt_x) + cs_yt_y &= 0 \quad \text{and} \\ at_x^2 - 2bt_xt_y + ct_y^2 &= 0, \end{aligned}$$

so as to make the coefficients of f_{st} and f_{tt} , respectively, vanish. The first equation yields

$$\frac{t_x}{t_y} = \frac{b}{a},$$

and with this, the second equation is also satisfied. In the parabolic case, the PDE can be brought to canonical form by appropriately defining only one of the two new coordinates. Thus, there is only one family of characteristics,

defined by the level curves $t(x, y) = C$; in contrast to the hyperbolic and elliptic cases, each of which have two families of characteristics (complex in the elliptic case), second-order parabolic PDEs have only a single family of characteristics. Locally, we may express y as a function of x along the characteristics, in which case we have $t(x, y(x)) = C$. Differentiation with respect to x yields

$$t_x + t_y \frac{dy}{dx} = 0,$$

and therefore, the slope of the characteristics is given by

$$\frac{dy}{dx} = -\frac{t_x}{t_y} = -\frac{b}{a} = -\frac{c}{b},$$

or equivalently,

$$\frac{dy}{dx} = -\frac{\varepsilon_y g_{xx} - \varepsilon_x g_{xy}}{\varepsilon_y g_{xy} - \varepsilon_x g_{yy}}.$$

With t defined in this way, Eq. (5) reduces to the canonical form for a parabolic second-order linear PDE,

$$\begin{aligned} (as_x^2 - 2bs_xs_y + cs_y^2)f_{ss} \\ = -(\det H)^2\rho(s, t) - (as_{xx} - 2bs_{xy} + cs_{yy})f_s \\ - (at_{xx} - 2bt_{xy} + ct_{yy})f_t. \end{aligned} \quad (\text{A1})$$

Here, $s(x, y)$ may be any coordinate such that $\nabla s \times \nabla t \neq 0$ for all (x, y) in the domain. That is, $s(x, y)$ is any coordinate such that the transformation $(x, y) \rightarrow (s, t)$ is invertible throughout the domain.

With Eq. (A1), we are now in a position to show more explicitly that the linearized light scattering PDE in canonical form has the nature of a diffusion equation, and that the relevant, local diffusion constant is proportional to the local radius of curvature of the characteristics. To do so, we first make the simplifying assumption that $\nabla \varepsilon$ is a constant vector. With this assumption, we have

$$\frac{t_x}{t_y} = \frac{b}{a} = \frac{\varepsilon_y g_{xx} - \varepsilon_x g_{xy}}{\varepsilon_y g_{xy} - \varepsilon_x g_{yy}} = \frac{(\varepsilon_y g_x - \varepsilon_x g_y)_x}{(\varepsilon_y g_x - \varepsilon_x g_y)_y}.$$

Thus, t has the form $t(x, y) = u(\varepsilon_y g_x - \varepsilon_x g_y)$ for any function u such that $u' \neq 0$ throughout the domain. We take

$$t = \varepsilon_y g_x - \varepsilon_x g_y = \nabla g \times \nabla \varepsilon,$$

in which case, the coefficients of the linearized PDE (5) can be expressed as $a = t_y^2$, $b = t_x t_y$, and $c = t_x^2$, and Eq. (A1) can be written as

$$\begin{aligned} (\nabla s \times \nabla t)^2 f_{ss} + (t_y^2 s_{xx} - 2t_x t_y s_{xy} + t_x^2 s_{yy}) f_s \\ + (t_y^2 t_{xx} - 2t_x t_y t_{xy} + t_x^2 t_{yy}) f_t = -(\det H)^2 \rho(s, t). \end{aligned} \quad (\text{A2})$$

The coefficient of f_t in Eq. (A2) can be expressed in terms of the curvature κ of the characteristics as follows. It is known¹⁰ that the curvature of a level curve $t(x, y) = C$ is given by

$$\kappa = \frac{|t_y^2 t_{xx} - 2t_x t_y t_{xy} + t_x^2 t_{yy}|}{|\nabla t|^3} = -\sigma \frac{t_y^2 t_{xx} - 2t_x t_y t_{xy} + t_x^2 t_{yy}}{|\nabla t|^3},$$

where

$$\sigma = -\text{sgn}(t_y^2 t_{xx} - 2t_x t_y t_{xy} + t_x^2 t_{yy}),$$

and therefore,

$$t_y^2 t_{xx} - 2t_x t_y t_{xy} + t_x^2 t_{yy} = -\sigma \kappa |\nabla t|^3.$$

With an appropriate choice of the s -coordinate, Eq. (A2) can now be cast into the form of the diffusion or heat equation with a source term. As we have shown above that the dielectric coefficient ε varies monotonically along characteristics, we take one natural choice for the s -coordinate, simply $s(x, y) = \varepsilon(x, y)$. With this choice, and under the assumption that $\nabla \varepsilon$ is a constant vector, the coefficient of f_s in Eq. (A2) is equal to zero, and Eq. (A2) becomes

$$(\nabla \varepsilon \times \nabla t)^2 f_{ss} + 0 - \sigma \kappa |\nabla t|^3 f_t = -(\det H)^2 \rho(s, t),$$

or equivalently,

$$\sigma \kappa |\nabla t|^3 f_t = (\nabla \varepsilon \times \nabla t)^2 f_{ss} + (\det H)^2 \rho(s, t). \quad (\text{A3})$$

Equation (A3), identical to Eq. (6) in the text, is in the form of the heat equation with a source term, and the diffusion coefficient is proportional to $1/\kappa$, the local radius of curvature of the characteristics.

The coefficient of the diffusion term f_{ss} can be written as

$$(\nabla \varepsilon \times \nabla t)^2 = (|\nabla \varepsilon| |\nabla t| \sin \phi)^2,$$

where ϕ is the angle between $\nabla \varepsilon$ and ∇t . The angle ϕ is related to the angles θ and α , shown in Fig. 2, through the relation $\phi = (\theta - \alpha) + \pi/2$, and therefore, we have

$$(\nabla \varepsilon \times \nabla t)^2 = (|\nabla \varepsilon| |\nabla t| \cos(\theta - \alpha))^2.$$

The linear PDE can then be written as

$$\sigma \kappa |\nabla t|^3 f_t = |\nabla \varepsilon|^2 |\nabla t|^2 \cos^2(\theta - \alpha) f_{ss} + (\det H)^2 \rho(s, t),$$

or equivalently,

$$\sigma \kappa |\nabla t| f_t = \cos^2(\theta - \alpha) |\nabla \varepsilon|^2 f_{ss} + \frac{(\det H)^2}{|\nabla t|^2} \rho(s, t). \quad (\text{A4})$$

In addition to the influence of the characteristic curvature κ already noted, Eq. (A4) also shows that $\det H$ has a pronounced effect on the magnitude of the source term containing $\rho(s, t)$. In particular, proximity to the spinodal condition $\det H = 0$ can considerably reduce the influence of $\rho(s, t)$ on f , as is further illustrated below in connection with Fig. 5. We note that the quantity $|\nabla t| f_t$ in Eq. (A4) is the directional derivative of f with composition in the t direction, while $|\nabla \varepsilon|^2 f_{ss}$ is the second derivative of f with composition in the direction of $\nabla \varepsilon$, under the assumption, used in deriving Eq. (A4), that $\nabla \varepsilon$ is a constant vector.

In Eq. (A4), for comparison with the figures it is useful to note further that for the case in which $\varepsilon_x > \varepsilon_y$, the characteristics can be described with y being a function of x . In this case, we have

$$\frac{d^2 y}{dx^2} = \frac{t_y^2 t_{xx} - 2t_x t_y t_{xy} + t_x^2 t_{yy}}{-t_y^3}$$

and $t_y < 0$. Therefore,

$$\sigma = -\text{sgn}(t_y^2 t_{xx} - 2t_x t_y t_{xy} + t_x^2 t_{yy}) = -\text{sgn}\left(\frac{d^2 y}{dx^2}\right).$$

For the ideal free energy and $\varepsilon_x > \varepsilon_y$, as illustrated in Figs. 2 and 3, the characteristics are concave down everywhere, and therefore, $\sigma = 1$.

3. Lambda integrals for binary mixtures with linear interpolation

$$\phi_j(x) = \begin{cases} c_j(x - x_0), & x_0 \leq x \leq x_{j-1} \\ c_j(x - x_0) + \frac{(x - x_{j-1})^3}{6(x_j - x_{j-1})}, & x_{j-1} < x \leq x_j \\ d_j(x_n - x) + \frac{(x_{j+1} - x)^3}{6(x_{j+1} - x_j)}, & x_j < x \leq x_{j+1} \\ d_j(x_n - x), & x_{j+1} < x \leq x_n \end{cases},$$

$$c_j = -\frac{x_{j+1} - x_{j-1}}{2(x_n - x_0)} \left(x_n - \frac{x_{j-1} + x_j + x_{j+1}}{3} \right),$$

$$d_j = -\frac{x_{j+1} - x_{j-1}}{2(x_n - x_0)} \left(\frac{x_{j-1} + x_j + x_{j+1}}{3} - x_0 \right),$$

$$d_j + c_j = -\frac{x_{j+1} - x_{j-1}}{2},$$

$$\begin{aligned} \lambda_j &= \int_0^1 \phi_j^2(x) dx = c_j^2 \frac{(x_j - x_0)^3}{3} + c_j \frac{(x_j - x_{j-1})^3}{60} \\ &\quad \times (x_{j-1} + 4x_j - 5x_0) + \frac{(x_j - x_{j-1})^5}{252} \\ &\quad + d_j^2 \frac{(x_n - x_j)^3}{3} + d_j \frac{(x_{j+1} - x_j)^3}{60} \\ &\quad \times (5x_n - 4x_j - x_{j+1}) + \frac{(x_{j+1} - x_j)^5}{252}. \end{aligned} \quad (\text{A5})$$

The leading order form for λ_j for small $\Delta x = x_j - x_{j-1}$ is obtained from Eq. (A5) as

$$\lambda_j \simeq \frac{\Delta x^2}{3} x_j^2 (1 - x_j)^2. \quad (\text{A6})$$

4. Derivation of binary mixture scaling of error norm with intensity

With a single tent centered at $x = x_*$, Eq. (37) becomes

$$f(x) = \rho_* \phi_{*,lin}(x) = \left(\frac{\varepsilon_x}{R} \right)_*^2 \rho_* \tilde{\phi}_{*,lin}(x), \quad (\text{A7})$$

where, to leading order,

$$\tilde{\phi}_{*,lin}(x) = \begin{cases} \Delta x (1 - x_*) x, & x < x_* \\ \Delta x (1 - x) x_*, & x > x_* \end{cases}.$$

Then,

$$\begin{aligned} \int_0^1 f^2(x) dx &= \int_0^1 \left(\frac{\varepsilon_x}{R} \right)_*^4 \rho_*^2 \tilde{\phi}_{*,lin}^2(x) dx \\ &= \left(\frac{\varepsilon_x}{R} \right)_*^4 \rho_*^2 \Delta x^2 \left[\int_0^{x_*} (1 - x_*)^2 x^2 dx \right. \\ &\quad \left. + \int_{x_*}^1 (1 - x)^2 x_*^2 dx \right] \\ &= \left(\frac{\varepsilon_x}{R} \right)_*^4 \rho_*^2 \Delta x^2 \left[(1 - x_*)^2 \frac{x_*^3}{3} + x_*^2 \frac{(1 - x_*)^3}{3} \right] \\ &= \left(\frac{\varepsilon_x}{R} \right)_*^4 \frac{\rho_*^2 \Delta x^2}{3} x_*^2 (1 - x_*)^2 \\ &= \frac{\rho_*^2 \Delta x^2}{3} \frac{\varepsilon_x^4 x_*^2 (1 - x_*)^2}{R_*^4}, \end{aligned}$$

and therefore,

$$\mathcal{L}_2[f] = \left[\int_0^1 f^2(x) dx \right]^{1/2} = \frac{\rho_* \Delta x}{\sqrt{3}} \cdot \frac{\varepsilon_x^2 x_* (1 - x_*)}{R_*^2}. \quad (\text{A8})$$

For the regular solution model, $g(x) = x \ln x + (1 - x) \ln(1 - x) + Cx(1 - x)$, we have

$$R_* = \frac{\varepsilon_x^2 x_* (1 - x_*)}{1 - 2Cx_*(1 - x_*)}$$

and

$$\mathcal{L}_2[f] = \frac{\rho_* \Delta x}{\sqrt{3}} \cdot \frac{1 - 2Cx_*(1 - x_*)}{R_*}. \quad (\text{A9})$$

For the ideal solution free energy ($C = 0$), this expression reduces to

$$\mathcal{L}_2[f] = \frac{\rho_* \Delta x}{\sqrt{3}} \cdot \frac{1}{R_*}. \quad (\text{A10})$$

¹D. Ross, G. Thurston, and C. Lutzer, *J. Chem. Phys.* **129**, 064106 (2008).

²J. Kirkwood and R. Goldberg, *J. Chem. Phys.* **18**, 54 (1950).

³M. G. Crandall, H. Ishii, and P.-L. Lions, *Bull., New Ser., Am. Math. Soc.* **27**, 1 (1992).

⁴B. Widom, *J. Chem. Phys.* **39**, 2808 (1963).

⁵M. Hirsch, S. Smale, and R. Devaney, *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, 2nd ed. (Academic, 2003).

⁶W. Thacker, J. Zhang, L. Watson, J. Birch, and M. Iyer, *ACM Trans. Math. Softw.* **37**(3), 34 (2010).

⁷W. Davenport, Jr. and W. Root, *An Introduction to the Theory of Random Signals and Noise* (Wiley-IEEE, 1987).

⁸J. Kaye and J. McDaniel, *Appl. Opt.* **13**, 1934 (1974).

⁹T. Bender, R. Lewis, and R. Pecora, *Macromolecules* **19**, 244 (1986).

¹⁰R. Goldman, *Comput. Aided Geom. Des.* **22**, 632 (2005).

¹¹J. Meijering, *Philips Res. Rep.* **5**, 333 (1950).

¹²J. Meijering, *Philips Res. Rep.* **6**, 183 (1951).