Maximum-Entropy Analysis of Photon Correlation Spectroscopy Data

Su-Long Nyeo† and Benjamin Chu*,†,‡

Department of Chemistry and Department of Materials Science and Engineering, State University of New York at Stony Brook, Stony Brook, New York 11794-3400.
Received November 11, 1988; Revised Manuscript Received March 13, 1989

ABSTRACT: The maximum-entropy formalism is briefly reviewed and then applied as a data analysis tool to estimating characteristic line-width distributions from photon correlation spectroscopy data. Two methods for solving the nonlinear maximum-entropy constrained optimization problem for maximal entropy characteristic line-width distributions are described. The reliability of the formalism is then tested by analyzing several sets of numerically simulated correlation data, which correspond to very broad unimodal characteristic line-width distributions, bimodal distributions, and a negatively skewed unimodal distribution. Experimental data of semidilute ternary solutions consisting of a four-arm star polystyrene (PS) and a linear poly(methyl methacrylate) (PMMA) in toluene are analyzed by the maximum-entropy formalism. From the behavior of the reconstructed characteristic line-width distributions, we can show a definite influence of the invisible (isorefractive) but entangled PMMA on the visible PS coils.

1. Introduction

Photon correlation spectroscopy (PCS) has been used extensively for characterizing various polymer systems,[1-5] colloidal suspensions,[6-8] and microemulsions.[9,10] It has proved more effective than many other known techniques, such as gel permeation chromatography (GPC), mainly for its noninvasiveness and ability to determine very wide molecular size ranges. Moreover, PCS permits characterizations of many otherwise intractable polymers in solution.[3-5] Yet, the effectiveness of this technique is often offset by various intrinsic limitations of the mathematical techniques used for analyzing correlation data. To extract size information about a polymer system, it is necessary to estimate its characteristic line-width distribution $G(\Gamma)$ in the absence of particle internal motions and interparticle interactions, where $\Gamma = DK^2$, with D being a translational diffusion constant and K the magnitude of a scattering vector. The estimation of $G(\Gamma)$ is a difficult problem, which involves solving a Fredholm integral equation of the first kind

$$|g^{(1)}(\tau_i)| = \int_0^\infty G(\Gamma) \exp(-\Gamma\tau_i)\, d\Gamma \qquad (1)$$

from the given electric field autocorrelation function $g^{(1)}(\tau_i)$, where τ_i denotes the delay time of channel i. This function is related to a measured self-beating intensity–intensity autocorrelation function $G_{exp}^{(2)}(\tau_i)$ by the Siegert relation[11]

$$G_{exp}^{(2)}(\tau_i) = B(1 + \beta|g_{exp}^{(1)}(\tau_i)|^2) \qquad (2)$$

where B is a measured base line, β is an instrumental coherence factor, and $g_{exp}^{(1)}(\tau_i)$ denotes the "measured" normalized electric field autocorrelation function.

Equation 1 is a well-known ill-posed linear problem, in which the number of available data $g_{exp}^{(1)}(\tau_i)$ is always less than that needed to describe $G(\Gamma)$ uniquely. In particular, eq 1 is very unstable to noise in the sampled data, to round-off in computations,[12-19] and often to the bandwidth limitation of the photon correlation instrument used. Let $G(\Gamma)$ be a solution to eq 1. Then for a bounded function $N(\omega)$, $|N(\omega)| < \omega$, $G(\Gamma) + N(\omega) \sin(\omega\Gamma)$ is also a solution if the parameter ω is chosen to be sufficiently large that the Laplace transform of $N(\omega) \sin(\omega\Gamma)$ is within the error deviations of the experimental data.[12-19] Thus, to recover a distribution from noisy and incomplete data, it is necessary to define eq 1 as a mathematically well-posed one by introducing, for example, some regularization (smoothing) procedure or some finite number of known functions to represent $G(\Gamma)$. Many approaches were proposed in the past decade but shown to have certain intrinsic limitations. For example, the most widely used cumulants method[20] can only provide an approximate mean and variance of a reasonably narrow unimodal distribution. Clearly, this method cannot be used to extract information on complicated, multimodal distributions. We shall refrain from giving details on all the available analysis methods but refer to several recent reviews.[21-23]

For instructive purposes, however, we briefly mention the conventional regularization procedure. In the context of the mathematical theory of regularization, the ill-conditioning of an inverse problem, such as eq 1, can be quenched by solving, in the weighted least-squares or the χ-square (χ^2) sense, for example, the *regularized* problem[12-19]

$$\text{minimize } Q = \left\{ \left\| \frac{\hat{d}_i - d_i}{\sigma_i} \right\|^2 + \alpha \|L\hat{G}\|^2 \right\} \qquad (3)$$

(or maximize $-Q$). Here $\hat{d}_i = \beta^{1/2}|g^{(1)}(\tau_i)|$ is an estimate of $d_i = \beta^{1/2}|g_{exp}^{(1)}(\tau_i)|$, which is given by eq 2, where the factor β is taken as a parameter in the fitting procedure, so that the estimate $\hat{G}(\Gamma)$ of the true $G(\Gamma)$ is normalized to $\beta^{1/2}$. The quantity σ_i denotes the error deviation of d_i, and $\|...\|^2$, a least-squares quantity. (For the definition of a least-squares $\|...\|^2$, we refer to eq 20 below.) The positive parameter α, which is generally called a regularization parameter, is an undetermined Lagrange multiplier. Its value is to be specified by some statistic criterion[17-19,24,25] and has the effect of controlling the size of the second term in eq 3. In essence, the purpose of the second term is to ensure that the estimate $\hat{G}(\Gamma)$ is unique in the sense determined by some differential operator L, which is chosen either for convenience or after some preliminary analysis on simulated data. The most common choice is to define

$$L = d^n/d\Gamma^n \qquad (4)$$

giving an nth order regularization. (We note that since $\hat{G}(\Gamma)$ in eq 3 is necessarily discretized, L is expressed as a difference operator.) Thus, with $n = 2$, for example, we control the size of the second derivative of $\hat{G}(\Gamma)$. Equation 3 can be solved for $\hat{G}(\Gamma)$ by a quadratic programming algorithm,[26,27] together with the prior constraint that $\hat{G}(\Gamma)$ be positive.

This regularization method has been used for handling many ill-posed physical problems and tested with rea-

* Author to whom all correspondence should be addressed.
† Department of Chemistry.
‡ Department of Materials Science and Engineering.

0024-9297/89/2222-3998$01.50/0

sonable success and reliability. It has been applied to various inverse problems ranging from heat conduction[28] and tomography[29,30] to PCS.[17-19] (For general purposes, Provencher[31] has developed a program, which uses a second-order operator ($n = 2$).) The major drawback of using this type of regularization procedure is the lack of any physical justification or criterion for choosing one over many other possible operators. Clearly we should not apply this procedure indiscriminately to every inverse problem. In PCS, for example, this procedure was shown to give inadequate descriptions for certain distributions, negatively skewed ones in particular.[32] Thus, it is crucial to use a more efficient formalism and to test its reliability with various known distributions, among which is a negatively skewed one.

The principle of maximum entropy[33-35] can provide a general, information theoretic approach to solving inverse problems.[36,37] It has been shown capable of handling various ill conditions of inverse problems, and proven[38] to be the only objective and consistent method of selecting a solution from many that fit the data. This formalism has been used in many areas of applications.[39-53] After providing many very convincing and remarkable results, notably in the areas of image reconstruction,[39,43,44] the maximum-entropy method (MEM) has found many other applications, which include structure determination of liquid from neutron diffraction data[51] and colloidal suspensions from PSC data [5-8,52] in recent years. For reviews of the various applications of MEM, we refer to ref 47 and 48. Last but not least, the MEM has also been found useful in the field of urban and regional modeling.[53]

The objectives of this paper are (a) to describe in detail the maximum-entropy formalism and (b) to apply it to analyzing PCS data. This paper is thus organized as follows. In section 2 we shall give a brief introduction to the maximum entropy formalism and then in section 3 apply it to estimating characteristic line-width distributions from PCS data. For pedagogical reasons, two methods for solving the nonlinear maximum-entropy problem formulated in section 3 will be described in section 4. In section 5, we shall apply the MEM to simulated data. In order to evaluate the reliability of the MEM, it is essential to perform numerical simulation of correlation data corresponding to some known line-width distribution functions, which should include, in our opinion, very broad unimodal characteristic line-width distributions, bimodal distributions, and also a negatively skewed unimodal distribution. We shall confine our study to these simple distributions for practical and illustrational purposes. Broad distributions and bimodal distributions are commonly encountered in polymer systems and should therefore be included in the test for the reliability of the MEM. On the other hand, a negatively skewed distribution should provide a more stringent test of the formalism, since the conventional regularization method is known to give inadequate result for negatively skewed distributions. In section 6, experimental data of semidilute ternary solutions consisting of a four-arm star polystyrene (PS) and a linear poly(methyl methacrylate) (PMMA) in toluene will be analyzed by the MEM. Finally, the results of the paper will be discussed in section 7.

2. The Maximum-Entropy Formalism

The first application of the principle of entropy maximization was made by Jaynes,[33,34] who was concerned with providing a firm basis for the procedures of statistical mechanics, where one induces by the principle a certain probability structure of the states of some physical system and then calculates therefrom the averages of the various thermodynamic quantities. The induced probability distribution of the system is the only one that has maximal entropy and corresponds to minimum assumptions about the system.

In statistical mechanics, probability theory plays an important role. Unlike particle mechanics, statistical mechanics is the study of mechanical systems on the basis of incomplete data. To describe a system with a large number of degrees of freedom, for example, we consider incomplete data, such as ensemble averages, of physical quantities of the system and make predictions about the system from the data. However, since the given information is insufficient to specify the system, statistical mechanical predictions necessarily contain an element of uncertainty and can only be based on *inductive* reasoning. For inductive reasoning, the essential formalism is probability theory, which hence constitutes the mathematical backbone of statistical mechanics.

To describe a statistical mechanical system, in which inductive probabilities are useful, and for which given information is insufficient to permit deductive inferences, information theory can be used to provide a useful quantitative measure of the missing information in a probability distribution. Thus, the concept of using information theory is relevant to statistical mechanics.

A reasonable measure of missing information in a probability assignment $(p_1, p_2, ..., p_N)$ was first introduced by Shannon[54] and has the form

$$S = -\sum_{i=1}^{N} p_i \log (p_i) \tag{5}$$

Since eq 5 has the maximal solution p_i = constant, it applies only to cases in which all events are equally likely to occur prior to observation being made. (Simple derivation of eq 5 can be found in many statistical mechanics texts.[55-57]) This so-called entropy function was subsquently generalized by Kullback[58] to

$$S = -\sum_{i=1}^{N} p_i \log (p_i/g_i) \tag{6}$$

where g_i is a prior probability assignment, which plays the role of an initial model of the distribution of a physical system before any measurements are made. This generalization includes the important fact that prior probabilities are not necessarily equal. (For a quantum statistical system, g_i may be identified as a degeneracy weighting factor for quantum state i.) They are constructed on the basis of all available theoretical information about the system, and it is well-known that prior information can lead to improvements and more reliability in estimating the parameters of a system.[59]

The corresponding entropy of a continuous univariate probability distribution $P(x)$ of a mechanical system is defined by

$$S = -\int_{\Omega} P(x) \log (P(x)/m(x))\, dx \tag{7}$$

Here $m(x)$ is a prior probability distribution, which has the same transformation property as $P(x)$ and the same meanings as g_i in eq 6, and Ω denotes the domain of the distributions $P(x)$ and $m(x)$. (An entropy expression for a multivariate probability distribution can also be defined.) This entropy functional is an invariant with respect to the transformation of the variable x. If we express eq 7 in terms of another variable y, $y = y(x)$, then we must have

$$S = -\int_{\Omega} P(x) \log (P(x)/m(x))\, dx = -\int_{\Omega^*} Q(y) \log (Q(y)/Q^0(y))\, dy \tag{8}$$

where Ω^* denotes the domain of the posterior distribution $Q(y)$ and the prior distribution $Q^0(y)$ of the system described in terms of the variable y. This invariant property provides a useful relation between distributions of two variables that we can use to describe a physical system. In PCS, for example, a characteristic line-width distribution $G(\Gamma)$ or a relaxation time distribution $H(t)$ can be used to describe a polymer system, where the characteristic line width Γ and the relaxation time t are related by $\Gamma = 1/t$. In this case, by the invariance definition of entropy of the system, we should identify $P(x)$ and $Q(y)$ in eq 8 with $G(\Gamma)$ and $H(t)$, respectively, and introduce their corresponding prior distributions.

It should be realized that the use of an entropy expression for inference includes the introduction of a prior probability distribution, in which we summarize the prior information of the parameters of a system. Therefore, any sensible subsequent estimation of parameters is based on some problem-dependent choice of a prior distribution. Clearly, some choice may bias the results of the estimation process so as to favor parameter values for which the prior distribution is relatively large. Of course, any bias should diminish as the number of experiments is increased, except for values of that parameter for which the probabilities are zero. We note that a prior distribution $m(x)$ is uniform or rectangular if x is the canonical or natural variable of a system. In this instance, the system of interest can be described unambiguously by the parameter x. However, there are cases, referred to as noninformative, in which the scale of the variables of such systems is unknown, then the prior distributions should not be uniform.[60-62] For such cases, Jeffreys[60] proposed a scheme which requires that $m(x)$ be invariant with respect to *powers* in x. Namely, we should choose

$$m(x) \propto x^{-1} \tag{9}$$

so that for $y = x^n$

$$Q^0(y) \propto y^{-1} \tag{10}$$

This proposition was subsequently substantiated by Jaynes[61] by a group invariance principle argument. This assignment is essential if the concept of using the maximum-entropy formalism is to obtain objective and consistent solutions. PCS provides a typical example where such an assignment is needed (see below).

According to the formalism, we should introduce into the optimization problem constraints, such as data, whose effect is to introduce dependency among the parameters of the system, and then maximize eq 6 or 7 subject to these constraints. This maximization then leads to the most probable distribution, which satisfies the constraints and corresponds to minimum assumptions about the system, beyond what is surely known about the system.

For instructive purposes, consider an example in statistical mechanics, where the probabilities p_i of N possible quantum states are to be determined from some given information, such as the average energy of a system, E:

$$E = \sum_{i=1}^{N} \epsilon_i p_i \qquad \sum_{i=1}^{N} p_i = 1 \tag{11}$$

where ϵ_i is the energy of quantum state i. This is an ill-posed problem, because there are more unknown variables p_i than the given one E. Thus, to estimate p_i, we maximize the so-called Shannon–Jaynes entropy (6) subject to the given information (11):

$$S = -\sum_{i=1}^{N} p_i \log (p_i/g_i) - \lambda_1 \sum_{i=1}^{N} \epsilon_i p_i - \lambda_2 \sum_{i=1}^{N} p_i \tag{12}$$

where λ_1 and λ_2 are Lagrange multipliers. Here g_i denotes the degeneracy weighting factor of quantum state i. Maximizing (12) then leads to the following probabilities:

$$p_i = g_i \exp[-1 - \lambda_1 \epsilon_i - \lambda_2] \tag{13}$$

where the multipliers are fixed by the conditions (11). Given these p_i, we can then estimate any function of ϵ_i in a least biased manner with respect to unavailable data.

The above is a prototype ill-posed problem. For general mathematically ill-posed inverse problems, Jaynes[37] suggests that they should be treated as problems of inference and that the principle of maximum entropy can provide the least biased estimates on the basis of the available information about the problems. Such an approach is clearly conceptually distinct from the conventional differential-operator regularization approach mentioned in the Introduction. The major difference is that in the maximum-entropy (regularization) formalism, correlations in the maximal entropy solution are dictated purely by data and by the nature of a problem, while in the differential-operator formalism, correlations are also effected by our choice of the order of an operator. In the next section, we shall describe how the maximum-entropy formalism can be applied to analyzing PCS data.

3. Application of MEM to PCS

Although the maximum-entropy formalism is usually applied to inducing a probability distribution on the basis of partial information, it can[36,37] also be used as an inversion procedure for making inductive inferences on the best estimates of general variables, or on the least-biased solutions of inverse problems. Thus, in this section we describe how the maximum-entropy formalism can be applied to PCS.

To get an estimate $\hat{G}(\Gamma)$ of the true characteristic line-width distribution $G(\Gamma)$ from the data $d_i = \beta^{1/2}|g_{exp}^{(1)}(\tau_i)|$, we first consider the discretized relation between $\hat{G}(\Gamma_j)$ and $\hat{d}_i$:

$$\hat{d}_i = \sum_{j=1}^{N} \mathbf{K}_{ij} \hat{G}(\Gamma_j) \Delta, \qquad i = 1, ..., M \tag{14}$$

Here $\mathbf{K}_{ij} = \exp(-\Gamma_j \tau_i)$ is the curvature matrix, M is the number of data, N is the number of function values $\hat{G}(\Gamma_j)$ used in the discretization approximation (14), and Δ is a constant-Γ spacing. Then we expand $\hat{G}(\Gamma)$ in the set of random variables $\vec{X} = \{X_j\}$ as all possible solutions:[36,49]

$$\hat{G}(\Gamma_j) = \int X_j P(\vec{X})\, d\vec{X}, \qquad \int P(\vec{X})\, d\vec{X} = 1 \tag{15}$$

where $P(\vec{X})$ is a multivariate probability distribution. We note that by defining $\hat{G}(\Gamma_j)$ as averages we get the best estimates in the sense that $\hat{G}(\Gamma_j)$ have minimal deviations from the true values $G(\Gamma_j)$. Substituting the first equation in (15) into (14) then gives

$$\hat{d}_i = \int \Phi_i(\vec{X}) P(\vec{X})\, d\vec{X} \tag{16}$$

where $\Phi_i(\vec{X}) = \sum_j \mathbf{K}_{ij} X_j$. By the philosophy of the maximum-entropy formalism, we can now set up an objective function for the probability distribution as follows:

$$S = -\int P(\vec{X}) \log (P(\vec{X})/m(\vec{X}))\, d\vec{X} - \mu \int P(\vec{X})\, d\vec{X} - \sum_{i=1}^{M} \lambda_i \int \Phi_i(\vec{X}) P(\vec{X})\, d\vec{X} \tag{17}$$

Here, as usual, $m(\vec{X})$ is a prior distribution and μ and λ_i are Lagrange multipliers, whose values are fixed by conditions (15). Maximizing (17) with respect to $P(\vec{X})$ then leads to the Euler–Lagrange equation[33-36]

$$P(\vec{X}) = m(\vec{X}) \exp[-1 - \mu - \sum_{i=1}^{M} \lambda_i \Phi_i(\vec{X})] \quad (18)$$

or the partition function

$$Z[\alpha_j = \sum_i \lambda_i \mathbf{K}_{ij}] = \exp(1 + \mu) = \int m(\vec{X}) \exp[-\sum_{j=1}^{N} \alpha_j X_j] \, d\vec{X} \quad (19)$$

Equation 19 is in fact a general expression for an inverse problem. For an inverse problem, we only have to define a prior distribution $m(\vec{X})$ and a kernel $\mathbf{K}_{ij}$. By using eq 19, we can calculate $\hat{G}(\Gamma_j)$ and $\hat{d}_i$ by eq 15 and 16, respectively. But, we should remember to relate the experimental data d_i to the data estimates $\hat{d}_i$ by some statistic criterion, such as a χ^2 constraint.

For our problem, it is important to choose a relevant statistic criterion. A reasonable and practical assumption is that the errors of our data are independently normally distributed with zero means and square deviations σ_i^2, so that the errors obey a χ^2 constraint. (The use of a χ^2 constraint is also computationally very convenient.) In many circumstances, other constraints, such as L_1 and L_∞ norms,[63] may be more relevant. The L_1 norm, for example, is a natural choice to use in the deconvolution of seismic data, whereas the L_∞ norm has been used in the design of digital band-pass filters and the suppression of sidelobes in beam-forming techniques, where the objective is to design models that have certain specified tolerance limits. The χ^2 condition that we shall use is characterized by

$$\chi^2 = \left\| \frac{\hat{d}_i - d_i}{\sigma_i} \right\|^2 = \sum_{i=1}^{M} \left[\frac{\hat{d}_i - d_i}{\sigma_i} \right]^2 = M \quad (20)$$

which simply indicates that the estimated or predicted and the observed spreads in the observations are, on the average, equal for each channel value. Other χ^2 values have also been considered.[24,25] Since χ^2 has a mean M and a deviation $(2M)^{1/2}$ for large M, any χ^2 values in the interval $(M - (2M)^{1/2}, M + (2M)^{1/2})$ may be considered acceptable. Nevertheless, it should be noted that larger χ^2 values correspond to fewer structures in the characteristic line-width distribution but to higher confidence levels. For instance, a χ^2 value of $(M + 3.29M^{1/2})$ corresponds to about a 99% confidence level,[64] implying a 99% certainty that the true solution contains at least as much structure as that contained in the maximal entropy solution, while a χ^2 value of $(M + 1.64M^{1/2})$ corresponds to about a 90% confidence level.

Equally important is to provide an estimate of the error square deviations σ_i^2 of data d_i. A data analysis method without some sort of an error estimate is useless. Since a general analytic expression for σ_i^2, which includes all possible experimental conditions, is not accessible, an approximation is necessary. We shall assume that the square deviations of the correlation in eq 2 have a Poisson profile, so that the deviations of d_i are given by[65,66]

$$\sigma_i^2 \approx \frac{1 + d_i^2}{4Bd_i^2} \quad (21)$$

Equation 21 exhibits a reasonable feature of a correlation function that data at longer delay times have larger σ_i^2. However, since this approximation is valid only for sufficiently low photon count rate cases,[67,68] it often provides poor fittings to data if the measured base line B is used. Thus, it is necessary to perform a preliminary analysis to obtain a smooth curve, which is then used to estimate the deviations of data. Of course, we should remember that some preliminary analysis method may bias the estimation process. Furthermore, since there exist uncertainties in the estimated deviations, the equality statistic criterion (20) is inexact. Thus, eq 20 is only an approximation, which depends on the preliminary analysis method used to estimate the error deviation of correlation data. Alternatively, we may use a different statistic criterion such as that used in Provencher's CONTIN algorithm.

With some prescribed $m(\vec{X})$, the Laplace inversion problem now consists of solving eq 16, 19, and 20 for the parameters $\{\lambda_i\}$ or $\{\alpha_i\}$. This problem in its present form is numerically less efficient.[69] So instead of solving these equations, we shall use an equivalent but simpler maximum-entropy formalism.[36,49] Since a characteristic line-width distribution is semipositive and bounded, we can take it as a probability density. Thus, we can define

$$S = -\sum_{j=1}^{N} \{p_j \log (p_j/m_j)\}, \qquad p_j = \hat{G}(\Gamma_j)\Delta / \sum_j \hat{G}(\Gamma_j)\Delta \quad (22)$$

as the entropy of our system, where p_j is the representative fraction of particles (or macromolecules) with characteristic line width in the interval $(\Gamma_j - \Delta/2, \Gamma_j + \Delta/2)$. Here Δ is a constant spacing and Γ_j are equally spaced. However, in this case, the prior model distribution $\{m_j\}$ is not uniform. As mentioned in section 2, the scale of the parameter Γ is a priori unknown. So we should choose $m_j \propto 1/\Gamma_j$, which is an objective choice. We can take either a characteristic line width Γ or a relaxation (or decay) time t $(=1/\Gamma)$ as a parameter for describing a system. The objective choice for the prior distribution is then either $m_j \propto 1/\Gamma_j$ or $m_j^0 \propto 1/t_j$ (cf. eq 9 and 10). Otherwise, the choice of m_j = constant would lead to $m_j^0 \propto 1/(t_j)^2$ and not to m_j^0 = constant, and vice versa. If we now describe the problem in terms of log Γ or log t, we can use objectively a uniform prior distribution. This choice also makes the description of a characteristic line-width distribution more effective, so that very wide distributions can be better specified. Accordingly, we should define the fractions of particle sizes in the interval $(\log \Gamma_j - \Delta/2, \log \Gamma_j + \Delta/2)$ as $F_j = \hat{G}(\Gamma_j)\Gamma_j\Delta$, where $\Delta = (\log \Gamma_N - \log \Gamma_1)/(N - 1)$ is a spacing in log Γ and N is the number of logarithmically spaced Γ_j values satisfying $\log \Gamma_{j+1} = \log \Gamma_j + \Delta$. Normally, to describe very broad and multimodal distributions, we may take any value of N between 30 and 100. (In our program $N = 81$ is used.) The lower and upper bounds Γ_1 and Γ_N are chosen such that $\hat{G}(\Gamma_1)$ and $\hat{G}(\Gamma_N)$ are negligibly small. It should be emphasized that in order to define $\hat{G}(\Gamma)$, sample times should be chosen such that the relations $\Gamma_N = 1/\tau_1$ and $\Gamma_1 = 0.01/\tau_M$ are approximately satisfied,[17] where τ_1 and τ_M are the delay times of the first and the last channel, respectively. We can now use F_j as our "probabilities" and a uniform prior distribution to define an entropy expression. In this case, the Laplace transform eq 1 simply reads

$$\hat{d}_i = \sum_{j=1}^{N} F_j \mathbf{K}_{ij}, \qquad \mathbf{K}_{ij} = \exp(-\Gamma_j \tau_i) \quad (23)$$

where $\mathbf{K}_{ij}$ is the curvature matrix (kernel). We should note that since in PCS characteristic line-width distributions are only normalized to $\beta^{1/2}$, i.e., $\sum_j F_j = \beta^{1/2}$, the probabilities should be $p_j = F_j/\sum_j F_j$. But, entropy depends only on the form of $\{F_j\}$ and not on its normalization, thus it is computationally advantageous to use the following definition:

$$S = -\sum_{j=1}^{N} \{F_j \log (F_j/b)\} = -\sum_{j=1}^{N} \{F_j \log (F_j/A_0) - F_j\} \quad (24)$$

where b and A_0 $(=b/e, e = 2.718 \ldots)$ are constants, with

A_0 being a default or predetermined value,[43,47] which in the absence of data constraint is the maximal solution of eq 24. Here we choose $A_0 = \beta^{1/2}/N$, so that F_j and A_0 are normalized to $\beta^{1/2}$. The inverse problem now amounts to maximizing eq 24 subject to the χ^2 constraint (20) and the transform (23).

4. The Maximal Entropy Solution

In this section, we shall briefly describe two methods for solving the nonlinear maximum-entropy problem and how the solution can be attained.

The problem can be stated as follows: maximize the objective function with respect to the distribution $\{F_j\}$

$$S = -\sum_{j=1}^{N}\{F_j \log (F_j/A_0) - F_j\} \tag{24}$$

subject to the χ^2 constraint

$$\chi^2 = \sum_{i=1}^{M}\left[\frac{\hat{d}_i - d_i}{\sigma_i}\right]^2 = M \tag{20}$$

where $\hat{d}_i = \beta^{1/2}|g^{(1)}(\tau_i)|$ are the estimates on the data d_i and are related to F_j by eq 23. Equations 20, 23, and 24 describe a nonlinearly equality-constrained optimization problem and can be solved by several standard approaches.[70-73] For instructive purposes, we briefly describe two: (a) the Newton–Raphson iterative approach and (b) the quadratic models approximation method described by Burch et al.[43] (see also ref 44 and 48).

(a) The Newton–Raphson Iterative Approach. In this approach, we first formulate the constrained problem by using the method of Lagrange multipliers[70-73] as an unconstrained one and consider the new objective function or Lagrangian:

$$\tilde{S} = S - \frac{\alpha}{2}\chi^2, \qquad \alpha > 0 \tag{25}$$

where α is an undetermined Lagrange multiplier. (The factor $^1/_2$ in eq 25 is chosen for later convenience.) Then we directly maximize eq 25 with respect to F_j, and thus obtain the Euler–Lagrange equations

$$\partial S/\partial F_j - \frac{\alpha}{2}\partial\chi^2/\partial F_j = 0 \tag{26}$$

with

$$\partial S/\partial F_j = -\log (F_j/A_0) \tag{27}$$

$$\partial\chi^2/\partial F_j = 2\sum_{i=1}^{M}(\hat{d}_i - d_i)\mathbf{K}_{ij}/\sigma_i^2 \tag{28}$$

Solving (26) for F_j, we get

$$F_j = A_0 \exp\{-\alpha\sum_{i=1}^{M}(\hat{d}_i - d_i)\mathbf{K}_{ij}/\sigma_i^2\} \tag{29}$$

Equation 29 is highly nonlinear but can be solved for $\hat{d}_i$ by the Newton–Raphson iterative approach. It involves writing eq 29 as a set of coupled equations:

$$\phi_i = \hat{d}_i - \sum_{j=1}^{N}F_j\mathbf{K}_{ij} = 0 \tag{30}$$

where F_j is given by (29). To solve eq 30, we first expand it about $\hat{d}_i^{(0)}$, with $\hat{d}_i^{(1)} = \hat{d}_i^{(0)} + \delta d_i$ and $\hat{d}_i^{(0)} = d_i$:

$$\phi_i^{(1)} \approx \phi_i^{(0)} + \sum_{k=1}^{M}[\mathbf{M}]_{ik}\delta d_k \approx 0 \tag{31}$$

Here the matrix [M] is the derivative of (30) with respect to $\hat{d}_k$ and given explicitly by

$$[\mathbf{M}]_{ik} = \delta_{ik} + \frac{\alpha}{\sigma_i^2}\sum_{j=1}^{N}A_0 \exp\{-\alpha\sum_{l=1}^{M}(\hat{d}_l - d_l)\mathbf{K}_{lj}/\sigma_l^2\}\mathbf{K}_{ij}\mathbf{K}_{kj} \tag{32}$$

which is evaluated at $\hat{d}_i^{(0)} = d_i$, where δ_{ik} is the Kronecker delta. The corrections δd_i can now be obtained by matrix inversion:

$$\delta d_i = -\sum_{k=1}^{M}[[\mathbf{M}]^{-1}]_{ik}\phi_k^{(0)} \tag{33}$$

giving better estimates $\hat{d}_i^{(1)} = \hat{d}_i^{(0)} + \delta d_i$. The procedure is then iterated by expanding eq 30 about $\hat{d}_i^{(1)}$ and by evaluating eq 31–33 until the corrections at some iteration are negligibly small.

Many criteria can be implemented for terminating an iterative search. The following are two generally recommended criteria:[71,74]

$$\max_i |\hat{d}_i^{(n+1)} - \hat{d}_i^{(n)}| \le \delta_i \tag{34}$$

$$\max_j |F_j^{(n+1)} - F_j^{(n)}| \le \delta_j \tag{35}$$

where δ_i and δ_j are some tested tolerance levels. For our problem, $\delta_i = 10^{-4}(\hat{d}_i^{(n)} + 10^{-3})$ and $\delta_j = 10^{-4}(F_j^{(n)} + 10^{-3})$ are recommended.

Although this approach is known to be quadratically convergent about the "true" solution, it is numerically very unstable. This instability arises because the exponent in (32) can become very large (especially when the experimental data are very noisy or when the parameter α is necessarily large). Moreover, this approach is computationally very slow because one matrix inversion requires about M^3 mathematical operations. In general, it takes about as much computation time to generate the matrix [M] in eq 32 as to inverse it. For a set of 136 data points, about 2 min are required to perform an iteration on an IBM PS/2 Model 80 microcomputer with Microsoft Fortran (Version 4.1) or on a VAX 11-780 computer. On a microcomputer using both a 12-MHz 80286 math coprocessor and the Microsoft Fortran, about 5 min are needed for an iteration. For a given α value, roughly 10–20 iterations are generally needed for reaching an acceptable (not necessarily the final) solution. To obtain the final solution, the parameter α is adjusted from its initial value of 10^{-5}, say, such that if χ^2 is greater than the required confidence level, α is increased by a factor of 10 or less; otherwise, it is decreased by a comparable factor. Typically, as many as five α values, corresponding to five solutions, may be required for an iterative search. Since this approach demands considerable computer time, it is more favorable to adopt an alternative approach, to which we now turn.

(b) The Quadratic Model Approximation Approach. This second approach amounts to first approximating eq 20 and 24 as quadratic models:

$$\hat{S} = S_0 + \sum_{j=1}^{N}(\partial S/\partial F_j)\delta F_j + {}^1/_2\sum_{j,k}^{N}(\partial^2 S/\partial F_j\partial F_k)\delta F_j\delta F_k \tag{36}$$

$$\hat{\chi}^2 = {\chi_0}^2 + \sum_{j=1}^{N}(\partial\chi^2/\partial F_j)\delta F_j + {}^1/_2\sum_{j,k}^{N}(\partial^2\chi^2/\partial F_j\partial F_k)\delta F_j\delta F_k \tag{37}$$

where δF_j are the corrections to $F_j^0 = A_0$ (the default value), at which S_0, ${\chi_0}^2$, and the derivatives of S and χ^2 are evaluated. Note that the χ^2 expansion terminates at the second derivative and is therefore exact. However, since the entropy function is highly nonlinear, the quadratic expansion of it is only approximately valid locally about $F_j^0 = A_0$; that is, the quadratic term in eq 36 must

be smaller than its preceding terms. It is necessary to set an upper bound on the quadratic term

$$-\sum_{j,k}^{N}(\partial^2 S/\partial F_j \partial F_k)\delta F_j \delta F_k \leq 0.1\sum_{j}^{N} F_j^0 \tag{38}$$

which is chosen on practical experience.

For nonlinear models, it is necessary to solve for solution iteratively by calculating the corrections δF_j to the preceding approximate distribution. Thus, in the first iteration we calculate

$$F_j = F_j^0 + \delta F_j \tag{39}$$

and use them as the second estimates at which a new quadratic models approximation can be constructed. Then we define δF_j in terms of several search directions $\mathbf{e}^\mu$ (normally three is sufficient) in which we make the expansion:

$$\delta F_j = \sum_{\mu=1}^{3} x_\mu \mathbf{e}_j^\mu \tag{40}$$

To achieve efficiently the maximization of $\hat{S}$ under the χ^2 (20), the following simple unit vectors are used:[43]

$$\mathbf{e}_j^1 \propto F_j \partial\chi^2/\partial F_j \tag{41}$$

$$\mathbf{e}_j^2 \propto F_j[\alpha_1(\partial S/\partial F_j) - \alpha_2(\partial\chi^2/\partial F_j)] \tag{42}$$

$$\mathbf{e}_j^3 \propto \sum_{k=1}^{N} F_j(\partial^2\chi^2/\partial F_j \partial F_k)\mathbf{e}_k^2 \tag{43}$$

where α_1 and α_2 are given by

$$\alpha_1 = [\sum_j F_j(\partial S/\partial F_j)^2]^{-1/2} \qquad \alpha_2 = [\sum_j F_j(\partial\chi^2/\partial F_j)^2]^{-1/2} \tag{44}$$

so that the first and second quantities in eq 42 are normalized with respect to the metric tensor $\mathbf{g}_{jk} = \delta_{jk}/F_j$ (δ_{jk} being the Kronecker delta). The metric tensor $\mathbf{g}_{jk}$, which is (minus) the second derivative of the entropy function, $-(\partial^2 S/\partial F_j \partial F_k)$, provides a local measure of the magnitude and direction of the corrections (cf. eq 38) about some map $\{F_j\}$. Thus, it is effective to use this metric tensor for defining local quantities. For instance, the unit vectors (41)–(43) are normalized according to

$$\|\mathbf{e}^\mu\|^2 = \sum_{j,k}^{N} \mathbf{g}_{jk}\mathbf{e}_j^\mu \mathbf{e}_k^\mu = 1, \qquad \mu = 1, 2, 3 \tag{45}$$

Substituting the three vectors into eq 36 and 37, we can rewrite them in terms of x_μ ($\mu = 1, 2, 3$), respectively, as

$$\hat{S} = S_0 + \sum_{\mu=1}^{3} S_\mu x_\mu - \tfrac{1}{2}\sum_{\mu,\nu}^{3} \mathbf{h}_{\mu\nu} x_\mu x_\nu \tag{46}$$

$$\hat{\chi}^2 = \chi_0^2 + \sum_{\mu=1}^{3} C_\mu x_\mu + \tfrac{1}{2}\sum_{\mu,\nu}^{3} \mathbf{t}_{\mu\nu} x_\mu x_\nu \tag{47}$$

where

$$S_\mu = \sum_{j=1}^{N}(\partial S/\partial F_j)\mathbf{e}_j^\mu, \qquad \mathbf{h}_{\mu\nu} = -\sum_{j,k}^{N}(\partial^2 S/\partial F_j \partial F_k)\mathbf{e}_j^\mu \mathbf{e}_k^\nu \tag{48}$$

$$C_\mu = \sum_{j=1}^{N}(\partial\chi^2/\partial F_j)\mathbf{e}_j^\mu, \qquad \mathbf{t}_{\mu\nu} = \sum_{j,k}^{N}(\partial^2\chi^2/\partial F_j \partial F_k)\mathbf{e}_j^\mu \mathbf{e}_k^\nu \tag{49}$$

The distance constraint (38) reduces simply to

$$l = \sum_{\mu,\nu}^{3} \mathbf{h}_{\mu\nu} x_\mu x_\nu \leq 0.1\sum_{j}^{N} F_j^0 \tag{50}$$

A Lagrangian can now be set up: $L(x) = \alpha S - \chi^2 - \beta l$, with α and β being positive parameters. (We note that we can use instead the Lagrangian $L'(x) = S - \alpha'\chi^2 - \beta' l$, with positive parameters α' and β', which amounts simply to the rescalings: $L(x) \to \alpha L'(x)$, $\alpha' = 1/\alpha$, $\beta' = \beta/\alpha$. Thus, here the use of $L(x)$ is purely for convenience.) Maximizing $L(x)$ leads to a set of coupled equations for x_μ. For computational efficiency, we first decouple the components x_μ by transforming[73,75] them to an orthogonal set $\{y_\mu\}$, so that the matrices or tensors $\mathbf{h}_{\mu\nu}$ and $\mathbf{t}_{\mu\nu}$ become diagonal. By so doing, the Lagrangian now reads

$$L(y) = \alpha S_0 - \chi_0^2 + \sum_{\mu=1}^{3}(\alpha\tilde{S}_\mu - \tilde{C}_\mu)y_\mu - \frac{1}{2}\sum_{\mu,\nu}^{3}((\alpha + 2\beta)\delta_{\mu\nu} + \Lambda_{\mu\mu})y_\mu y_\nu \tag{51}$$

where tildes denote quantities that are orthogonally transformed, $\delta_{\mu\nu}$ is a Kronecker delta, and $\Lambda_{\mu\mu}$ ($\mu = 1, 2, 3$) are the diagonal elements of the diagonalized matrix of $\mathbf{t}_{\mu\nu}$. Maximization of this Lagrangian can be carried out in many ways. For example, the fact that eq 51 is stationary with respect to the variations of y_μ leads to the Euler–Lagrange equations for y_μ:

$$y_\mu = (\alpha\tilde{S}_\mu - \tilde{C}_\mu)((\alpha + 2\beta) + \Lambda_{\mu\mu})^{-1} \tag{52}$$

Equation 52 gives an infinitely many maximal entropy solutions which are classified by the parameters α and β. Of all the possible solutions we want to find one that satisfies our statistic criterion and the distance constraint (50), which is now given by $l = \sum_\mu y_\mu y_\mu$. But since at each iteration the minimum attainable χ^2 value is (cf. eq 51)

$$\chi_{\min}^2 = \chi_0^2 - \tfrac{1}{2}\sum_{\mu=1}^{3} C_\mu C_\mu/\Lambda_{\mu\mu} \tag{53}$$

which can be larger than our required value M, a slightly higher χ^2 value is imposed to provide some flexibility in the iterative procedure:

$$\chi_*^2 = \chi_0^2 - \tfrac{1}{3}\sum_{\mu=1}^{3} C_\mu C_\mu/\Lambda_{\mu\mu} \tag{54}$$

Thus, the refined χ^2 value in the procedure is given by

$$\tilde{\chi}^2 = \max\,(\chi_*^2, \chi^2 = M) \tag{55}$$

How the solution is chosen is briefly given as follows. At each iteration we first set $\beta = 0$ and find the largest α value such that the required χ^2 value (55) is attained. Then the parameter β is increased until the subsequent solution y_μ obeys the distance constraint. If the constraint is not satisfied, the $\tilde{\chi}^2$ value is increased to χ_0^2. The corrections δF_j are then obtained by backward substitution.

The required solution $\{F_j\}$ is the one that must satisfy some terminating criteria, e.g., (34) and (35). Alternatively, the maximal entropy solution can be selected by using the following criterion

$$\sum_{j=1}^{N} F_j[\alpha_1(\partial S/\partial F_j) - \alpha_2(\partial\chi^2/\partial F_j)]^2 < 2 \times 10^{-3} \tag{56}$$

where α_1 and α_2 are given in (44). Equation 56 is a measure of the degrees of entropy maximization, where the tolerance value of 2×10^{-3} is determined on practical grounds. We note that a zero tolerance value corresponds to the ideal maximal solution.

This approach is numerically very stable and computationally faster than the Newton–Raphson approach. However, this approach is not without any undesirable feature. In particular, the use of quadratic model ap-

proximation unavoidably loses the positivity property of the solution in the iterative procedure. Thus, it is necessary to reset any negative values in each iteration to a small but positive number.

For a set of 136 data points, about 6 s is needed for an iteration by using a microcomputer with a 12-MHz 80286 math coprocessor and Microsoft Fortran, and about 100 iterations are required. By comparison, 1 s per iteration is needed if an IBM PS/2 Model 80 microcomputer is used.

5. Simulation

Before applying the MEM to experimental data, it is crucial to test it with computer-simulated data. To do this, we have generated several sets of first-order correlation data d_i corresponding to reasonably broad characteristic line-width distributions, bimodal distributions, and a negatively skewed distribution. We used a log-normal distribution function for simulating both the broad distribution and the bimodal distribution, while we used a β function as the negatively skewed distribution. For the purpose of comparison, the simulated data are also analyzed by Provencher's CONTIN algorithm.

5.1. Data Generations. a. Broad Unimodal and Bimodal Distributions. To generate correlation data for reasonably broad unimodal distributions, we used the following log-normal function

$$G(\Gamma) = G_0(\Gamma) - D \tag{57}$$

$$G_0(\Gamma) = \frac{C}{\Gamma} \exp\left\{-\frac{1}{\beta^2}(\log \Gamma/\Gamma_0)^2\right\} \tag{58}$$

where $\beta = (2 \log (\sigma + 1))^{1/2}$ and $\Gamma_0 = \bar{\Gamma}\ (\sigma + 1)^{-1/2}$, with $\bar{\Gamma}$ being an approximate mean and σ ($=\mu_2/\bar{\Gamma}^2$) an approximate variance:

$$\mu_2 = \int_{\Gamma_{min}}^{\Gamma_{max}} (\Gamma - \bar{\Gamma})^2 G(\Gamma)\, d\Gamma \tag{59}$$

The upper and lower bounds, $\Gamma_{max} = \Gamma_0(\sigma + 1)^{-1}e^{3\beta}$ and $\Gamma_{min} = \Gamma_0(\sigma + 1)^{-1}e^{-3\beta}$, of $G(\Gamma)$ were chosen in order to generate very broad distributions. The ratios $G_0(\Gamma_{max})/G_0(\Gamma_{peak})$ and $G_0(\Gamma_{min})/G_0(\Gamma_{peak})$ are of the order 10^{-9}, with $G_0(\Gamma_{peak})$ denoting the peak value. In eq 57, D is given by $G_0(\Gamma_{max})$ and C is a normalization constant determined according to

$$\int_{\Gamma_{min}}^{\Gamma_{max}} G(\Gamma)\, d\Gamma = 1 \tag{60}$$

We should note that $\bar{\Gamma}$ is an exact mean and σ an exact variance if $\Gamma_{min} = 0$ and $\Gamma_{max} = \infty$. In our simulation, the mean, variance, and also skewness (see below for our definition of skewness) are numerically recalculated.

The first-order correlation function was generated by numerical integration defined as

$$|g^{(1)}(\tau_i)| = \int_{\Gamma_{min}}^{\Gamma_{max}} G(\Gamma) \exp(-\Gamma\tau_i)\, d\Gamma \tag{61}$$

$$|g^{(1)}(\tau_i)| \approx \sum_{j=1}^{N} \omega_j G(\Gamma_j) \Delta \exp(-\Gamma_j \tau_i) \tag{62}$$

where $\Delta = (\Gamma_{max} - \Gamma_{min})/(N - 1)$ and N is the total number of equally spaced Γ_j values. For our simulation, we took $N = 1001$ and approximate the integral by Simpson's integration quadrature with coefficients $\{\omega_j\} = \{{}^1/_3, {}^4/_3, {}^2/_3, {}^4/_3, ..., {}^2/_3, {}^4/_3, {}^1/_3\}$. (In fact, integral (61) can be reasonably well approximated by using histograms, with all coefficients $\omega_j = 1$.) The delay times τ_i were chosen so as to mimic the Brookhaven Instruments BI-2030AT 136 channel correlator, in which the 136 channels are partitioned into four groups according to (32, 32, 32, 40). These four groups have sample times $\Delta\tau_1$, $\Delta\tau_2 = 2^{n_1}\Delta\tau_1$, $\Delta\tau_3 = 2^{n_2}\Delta\tau_1$, and $\Delta\tau_4 = 2^{n_3}\Delta\tau_1$, respectively. We chose $\Delta\tau_1$ and the integers $\{n_1, n_2, n_3; n_1 \le n_2 \le n_3\}$ such that the simulated noise-free first correlation channel value $|g^{(1)}(\Delta\tau)|^2$ is greater than 0.998, and the last channel value $|g^{(1)}(\text{last channel})|^2$ is less than 0.005.

Random noise was then added to the generated second-order correlation data, $G^{(2)}(\tau_i)$. We used the subprograms RANDOM and RGAUSS in Provencher's CONTIN algorithm to provide random noise with zero mean and square deviations $G^{(2)}(\tau_i)$ and estimated the errors of data by eq 21. To represent the achievable values of the background B and the coherence factor β for correlation data, we chose $B = 10^7$ and $\beta = 0.3$. More specifically, the noisy correlation data are simulated according to the expression in the CONTIN subprogram USERSI:

$$d_i^2(\text{with noise}) = d_i^2 + \text{ERROR} \times [1 + d_i^2]^{1/2} \tag{63}$$

where $d_i^2 = \beta|g^{(1)}(\tau_i)|^2$ are evaluated by using eq 61 and $\text{ERROR} = RN/B^{1/2}$ is a normal deviate with zero mean and deviation $1/B^{1/2}$.

To generate correlation data corresponding to bimodal distributions, we simply combined two log-normal functions and the noisy data were then calculated according to the procedures given by eq 57–63.

b. Negatively Skewed Distribution. To simulate correlation data for a negatively skewed distribution, we used the following distribution function:

$$G(\Gamma) = C\Gamma^m(100 - \Gamma)^n \tag{64}$$

where C is a normalization constant. We chose $m = 4.0$, $n = 0.5$ to represent a negatively skewed distribution. This choice gives a distribution with $\bar{\Gamma} = 77.9$, variance $\mu_2/\bar{\Gamma}^2 = 0.04$, and skewness $\mu_3/\mu_2^{1.5} = -0.83$, where μ_2 is defined by eq 59, while μ_3 is given by

$$\mu_3 = \int_{\Gamma_{min}}^{\Gamma_{max}} (\Gamma - \bar{\Gamma})^3 G(\Gamma)\, d\Gamma \tag{65}$$

Again, the correlation data were generated as described by eq 57–63.

5.2. Results and Discussions. The generated data corresponding to the different distributions are analyzed by both the CONTIN and the MEM algorithm. The reconstructions of $G(\Gamma)$ by both methods are shown in Figures 1–3.

We have used $\mu_2/\bar{\Gamma}^2 \ge 0.5$ to represent *broad* characteristic line-width distributions because in our PCS data analysis the $\bar{\Gamma}$ values of such $G(\Gamma)$ distributions cannot be retrieved reliably by the cumulants method without an elaborate extrapolation procedure. The main aims of the present analysis by using simulated data are (1) to illustrate conditions whereby broad $G(\Gamma)$ distributions can be determined by either method, (2) to demonstrate the delicate limitation of both methods in bimodal analysis, and (3) to point out the problem related to the negative skewness in even a fairly narrow $G(\Gamma)$ distribution by both CONTIN and MEM.

Figure 1a shows $G(\Gamma)$ results from CONTIN and MEM from a broad characteristic line-width distribution according to eq 57–63 with $\bar{\Gamma} = 750\ s^{-1}$ and variance $\mu_2/\bar{\Gamma}^2 = 0.50$. Both CONTIN (denoted by hollow diamonds) and MEM (denoted by the dashed line) yield $\bar{\Gamma}$ values within ≈1% and $\mu_2/\bar{\Gamma}^2$ values within ≈10%. The shape of the $G(\Gamma)$ distribution is well represented by the $G(\Gamma)$ reconstruction from either CONTIN or MEM. The only observable deviation is the sharper cutoff of $G(\Gamma)$ from MEM. However, the uncer-

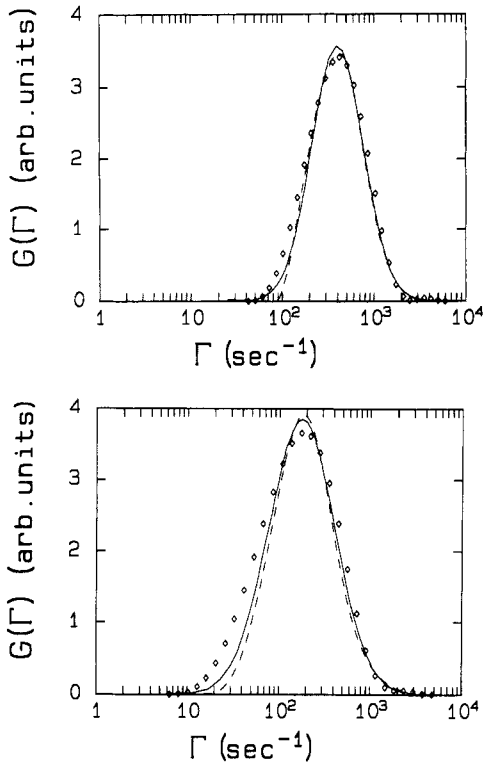

Figure 1. (a, Top) CONTIN and MEM results of a broad line-width distribution with input $\bar{\Gamma}$ = 750 s^{-1} and variance $\mu_2/\bar{\Gamma}^2$ = 0.50. Hollow diamonds denote the CONTIN result with $\bar{\Gamma}$ = 755 s^{-1} and variance $\mu_2/\bar{\Gamma}^2$ = 0.56, while the dashed line denotes the MEM result with $\bar{\Gamma}$ = 752 s^{-1} and variance $\mu_2/\bar{\Gamma}^2$ = 0.50. (b, Bottom) CONTIN and MEM results of a broad line-width distribution with input $\bar{\Gamma}$ = 494 s^{-1} and variance $\mu_2/\bar{\Gamma}^2$ = 0.90. Hollow diamonds denote the CONTIN result with $\bar{\Gamma}$ = 498 s^{-1} and variance $\mu_2/\bar{\Gamma}^2$ = 0.96, while the dashed line denotes the MEM result with $\bar{\Gamma}$ = 492 s^{-1} and variance $\mu_2/\bar{\Gamma}^2$ = 0.82.

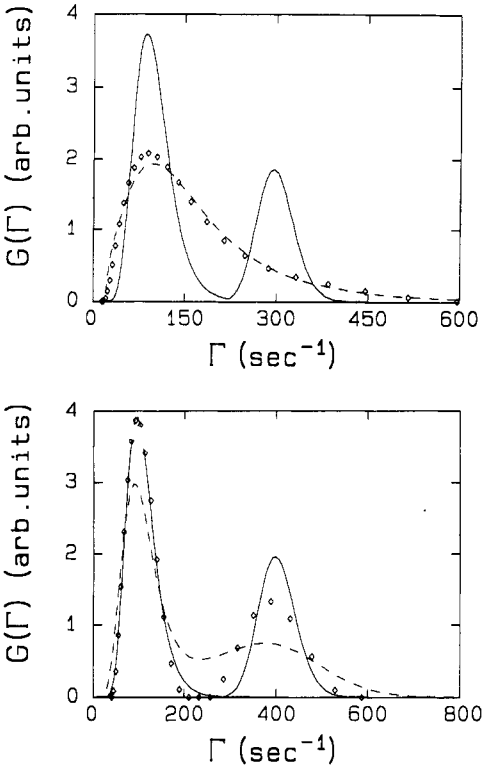

Figure 2. (a, Top) Reconstruction of a bimodal distribution with $\bar{\Gamma}_2/\bar{\Gamma}_1$ = 3, where $\bar{\Gamma}_1$ = 100 s^{-1} and $\bar{\Gamma}_2$ = 300 s^{-1} are the line-width averages of the first and second peaks, respectively. The input variance for the first peak is 0.1, while that for the second peak is 0.01. The ratio of areas A_2/A_1 is 0.528, $\bar{\Gamma}_{total}$ = 169 s^{-1} and an overall variance = 0.35. The CONTIN output ($\bar{\Gamma}_{total}$ = 170 s^{-1}, overall variance = 0.4) is represented by hollow diamonds and the MEM output ($\bar{\Gamma}_{total}$ = 170 s^{-1}, overall variance = 0.38) is shown by the dashed line. (b, Bottom) Reconstruction of a bimodal distribution with $\bar{\Gamma}_2/\bar{\Gamma}_1$ = 4, where $\bar{\Gamma}_1$ = 100 s^{-1} and $\bar{\Gamma}_2$ = 400 s^{-1} are the line-width averages of the first and second peaks, respectively. The input variance for the first peak is 0.1, while that for the second peak is 0.01. The ratio of areas A_2/A_1 is 0.70, $\bar{\Gamma}_{total}$ = 224 s^{-1}, and an overall variance = 0.46. The CONTIN output ($\bar{\Gamma}_2$ = 388 s^{-1}, $\bar{\Gamma}_1$ = 106 s^{-1}, ratio of areas A_2/A_1 = 0.7, $\mu_{2,2}/\bar{\Gamma}_2^2$ = 0.02, $\mu_{2,1}/\bar{\Gamma}_1^2$ = 0.073, $\bar{\Gamma}_{total}$ = 225 s^{-1}, overall variance = 0.44) is represented by hollow diamonds and the MEM output ($\bar{\Gamma}_{total}$ = 226 s^{-1}, overall variance = 0.46) is shown by the dashed line.

tainties in the lower bound (Γ_{min}) are usually fairly large due to noises in the base line as well as the tail section of the intensity–intensity time-correlation function measurements. Furthermore, values of the measured Γ_{min} could be influenced by arbitrary procedures in setting up the initial boundary conditions for the $G(\Gamma)$ distribution. Therefore, errors in the lower bound of the cumulative $G(\Gamma)$ could reach up to a few percent of the total value.

Figure 1b shows $G(\Gamma)$ reconstructions from CONTIN and MEM using a broader characteristic line-width distribution with $\bar{\Gamma}$ = 494 s^{-1} and $\mu_2/\bar{\Gamma}^2$ = 0.90. The $\bar{\Gamma}$ and $\mu_2/\bar{\Gamma}_2$ values are again in reasonable agreement, i.e., ≈1% and ≈10%, respectively, with the input values. The above demonstrations suggest that CONTIN (hollow diamonds) and MEM (dashed line) are equally applicable for both narrow and broad unimodal distributions. Although it may appear that we are repeating on what has already been considered as acceptable facts, to our knowledge this exercise has not been performed or reported in the literature. What we have done is to establish a base line for comparison under more controversial conditions, as to be illustrated in Figures 2 and 3.

Figure 2a shows a reconstruction of a bimodal distribution using eq 57–63 with area ratio A_2/A_1 = 0.528, $\bar{\Gamma}_2/\bar{\Gamma}_1$ = 3, $\bar{\Gamma}_2$ = 300 s^{-1}, and $\bar{\Gamma}_1$ = 100 s^{-1}; $\mu_{2,2}/\bar{\Gamma}_2^2$ = 0.01 and $\mu_{2,1}/\bar{\Gamma}_1^2$ = 0.1. Both CONTIN (hollow diamonds) and MEM (dashed line) yield an almost identical $G(\Gamma)$ without any bimodal characteristics. However, $\bar{\Gamma}_{CONTIN}$ = 170 s^{-1} and $\bar{\Gamma}_{MEM}$ = 170 s^{-1} are in excellent agreement with the input $\bar{\Gamma}$ value of 169 s^{-1}, and the $\mu_2/\bar{\Gamma}^2$ values from CONTIN and MEM are, respectively, 0.40 and 0.38, which are in reasonable agreement with the input $\mu_2/\bar{\Gamma}^2$ value of 0.35. At a higher $\bar{\Gamma}_2/\bar{\Gamma}_1$ value of 4 ($\bar{\Gamma}_2$ = 400 s^{-1}, $\bar{\Gamma}_1$ = 100 s^{-1}, ratio of areas A_2/A_1 = 0.7, $\mu_{2,2}/\bar{\Gamma}_2^2$ = 0.01, $\mu_{2,1}/\bar{\Gamma}_1^2$ = 0.1, $\bar{\Gamma}_{total}$ = 224 s^{-1}, overall variance = 0.46) the CONTIN algorithm shows a clear separation of the bimodal characteristics in $G(\Gamma)$ ($\bar{\Gamma}_2$ = 388 s^{-1}, $\bar{\Gamma}_1$ = 106 s^{-1}, ratio of areas A_2/A_1 = 0.7, $\mu_{2,2}/\bar{\Gamma}_2^2$ = 0.02, $\mu_{2,1}/\bar{\Gamma}_1^2$ = 0.073, $\bar{\Gamma}_{total}$ = 225 s^{-1}, overall variance = 0.44), while MEM (dashed line, $\bar{\Gamma}_{total}$ = 226 s^{-1} and overall variance = 0.46) appears to have less resolving power as shown in Figure 2b. The reason is that to demonstrate the importance of appropriately estimating the error deviations associated with the data in the Laplace inversion problem, we have intentionally used eq 21 in MEM in evaluating the error deviations in the measured data (d_i) *without prior smoothing of measured data*, in contrast to CONTIN, which utilizes prior smoothing of measured data. Clearly, prior knowledge of the error deviations of data can improve the resolution in $G(\Gamma)$. Nevertheless, from bimodal analysis with $\bar{\Gamma}_2/\bar{\Gamma}_1$ = 3 and 4, we see that claims of resolution of better than $\bar{\Gamma}_2/\bar{\Gamma}_1 < 3$ becomes questionable, whenever the bimodal distributions are not made up of two δ functions or various optimal experimental conditions, and any knowledge of associated error deviations of data are not executed properly. Thus, Laplace inversion of dynamic light scattering data represents a technique of fairly limited resolution.

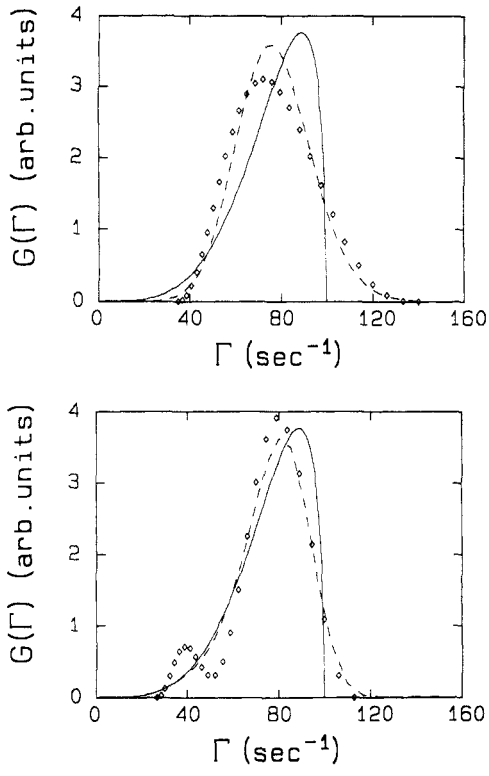

Figure 3. Reconstruction of the negatively skewed distribution in eq 64 by CONTIN (hollow diamonds) and MEM (dashed line) from the corresponding simulated correlation data (a, top) with noise and (b, bottom) without noise added. (a) CONTIN: $\bar{\Gamma} = 77.1\ s^{-1}$, $\mu_2/\bar{\Gamma}^2 = 0.053$, skewness = 0.29. MEM: $\bar{\Gamma} = 77.2\ s^{-1}$, $\mu_2/\bar{\Gamma}^2 = 0.045$, skewness = 0.22. (b) CONTIN: $\bar{\Gamma} = 76.9\ s^{-1}$, $\mu_2/\bar{\Gamma}^2 = 0.04$, skewness = −0.77. MEM: $\bar{\Gamma} = 76.9\ s^{-1}$, $\mu_2/\bar{\Gamma}^2 = 0.04$, skewness = −0.54.

Figure 3 shows reconstruction of a negatively skewed distribution by using the CONTIN (hollow diamonds) and the MEM (dashed line). Figure 3a shows the $G(\Gamma)$ results using the simulated data with noise added (eq 63, with $B = 10^7$ and $\beta = 0.3$). While both $\bar{\Gamma}$ and $\mu_2/\bar{\Gamma}^2$ values (77.9 s^{-1} and 0.04) can be retrieved by using either CONTIN ($\bar{\Gamma} = 77.1\ s^{-1}$, $\mu_2/\bar{\Gamma}^2 = 0.053$) or MEM ($\bar{\Gamma} = 77.2\ s^{-1}$, $\mu_2/\bar{\Gamma}^2 = 0.045$), the shape of $G(\Gamma)$ (skewness = −0.83) has clearly been distorted by either method with a skewness of 0.29 and 0.22 for CONTIN and MEM, respectively. In fact, both algorithms cannot retrieve $G(\Gamma)$ with a negative skewness even using input data (where round-off errors are still present and correspond to an experimentally inaccessible base-line value B greater than 10^9) without added noise as shown in Figure 3b. However, the MEM reconstructed $G(\Gamma)$ does not exhibit the bimodal characteristics as does the CONTIN reconstructed $G(\Gamma)$. This example demonstrates that the MEM approach can be used to provide better reconstructions than the conventional regularization procedure.

6. Application of MEM to Experimental Data

In this section, we shall apply the MEM algorithm to experimental photon correlation data of a set of ternary solutions consisting of a four-arm star polystyrene (PS) and a linear poly(methyl methacrylate) (PMMA) in toluene. In these solutions, PMMA is essentially isorefractive with toluene at room temperatures.

6.1. Polymer Solution Characteristics. The star polystyrene has a weight-averaged molecular weight of $M_w = 3.0 \times 10^5$ g/mol and a polydispersity $M_w/M_n < 1.1$, while the linear PMMA has $M_w = 9.9 \times 10^6$ g/mol and $M_w/M_n \approx 1.35$. These solutions with toluene as solvent were prepared with approximately the same concentrations of PS, C_{PS}, but with different concentrations of PMMA, C_{PMMA}. The overlap concentration C^* for PS is 5.3×10^{-2} g/mL, while that for PMMA is 2.5×10^{-3} g/mL. The characteristics of the solutions are summarized in Table I.

Table I
Properties of PS/PMMA/TOL Polymer Solutions

soln	C_{PS}, 10^{-2} g/g	C_{PMMA}, 10^{-2} g/g	$10^{-3}\bar{\Gamma}$, s^{-1}	$\mu_2/\bar{\Gamma}^2$
PSPMT0	0.23	0	1.63	0.076
PSPMT1	0.24	0.26	1.34	0.312
PSPMT2	0.26	1.00	0.88	0.556

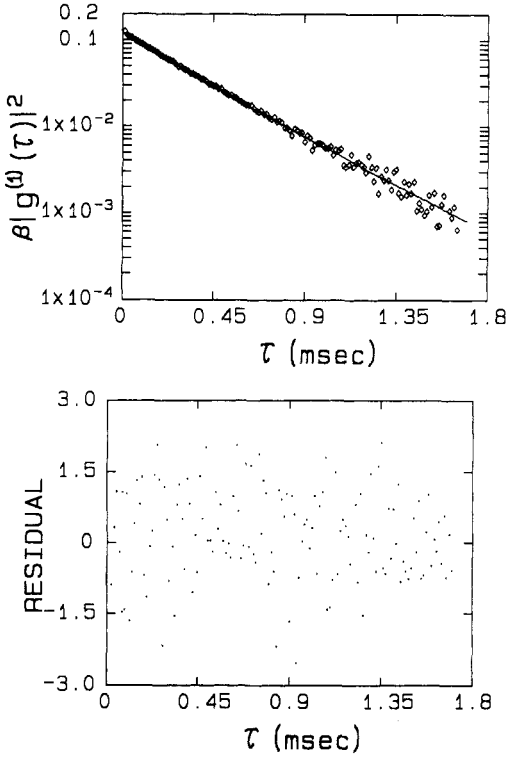

Figure 4. (a, Top) Fitting of time-correlation data by using MEM. Hollow diamonds denote experimental data. Solid line denotes MEM fitting with $G(\Gamma)$ shown in Figure 7 (solid line). (b, Bottom) Plot of residuals of the fitting for solution PSPMT0 as a function of delay time with residuals defined as $(d_i - \hat{d}_i)/\sigma_i$. Here d_i denotes the measured value $\beta^{1/2}|g^{(1)}(\tau_i)|$ and is related to $G(\Gamma)$ by eq 1, and σ_i, which is given by eq 21, is the error deviation of d_i.

6.2. Experimental Measurements. To measure the photon correlation data of these solutions, a Lexel Model 95 argon ion laser operated at wavelength $\lambda_0 = 514.5$ nm and at ≈200 mW and a Brookhaven Instruments BI2030AT 136-channel digital correlator were used. All measurements were performed at 25 °C and at a scattering angle of 22.5°. More detailed experiment methods are given in ref 76.

6.3. MEM Analysis. The correlation data were analyzed by using the MEM algorithm. Figures 4–6 show fittings of correlation data of the effect of isorefractive PMMA on PS in toluene by using MEM. In the absence of PMMA, PS shows a narrow variance as expected for a narrow molecular weight distribution PS sample in a good solvent (Figure 4). In the presence of isorefractive PMMA (Figure 5 and 6), the dynamics of the PS was influenced by the entangled PMMA coils and $\mu_2/\bar{\Gamma}^2$ was increased to reflect the motions of PMMA (Table I). Finally, as shown in Figure 7, a distinctive coupled motion of PS appears when $C_{PMMA} >> C^*_{PMMA}$, indicating a definitive influence of invisible but entangled PMMA on the visible PS coils. Thus, we see that the MEM can be used to analyze correlation data with broad and bimodal characteristic line-

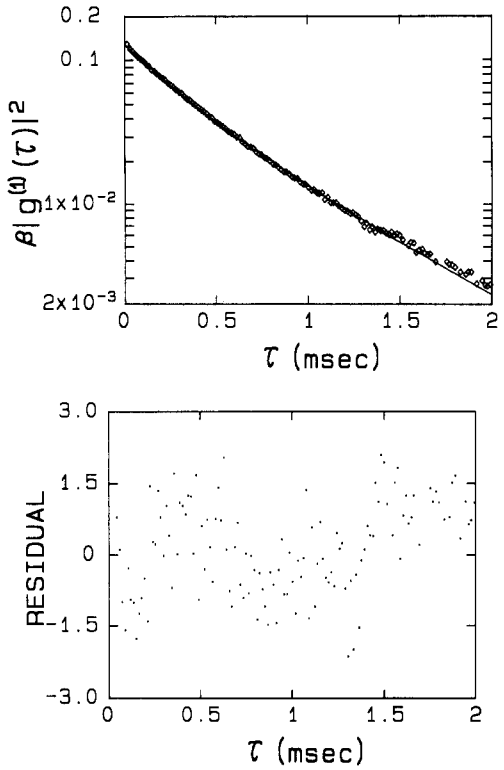

Figure 5. (a, Top) Fitting of time-correlation data by using MEM. Hollow diamonds denote experimental data. Solid line denotes MEM fitting with $G(\Gamma)$ shown in Figure 7 (dashed line). (b, Bottom) Plot of residuals of the fitting for solution PSPMT1 as a function of delay time with residuals defined in Figure 4b.

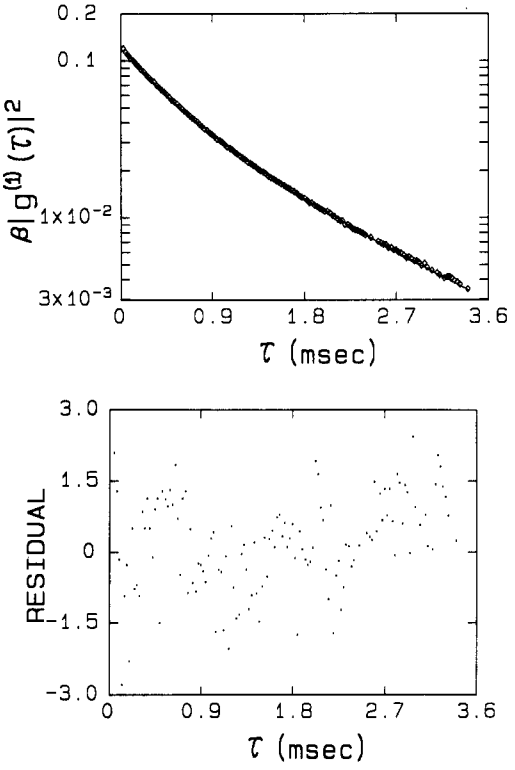

Figure 6. (a, Top) Fitting of time-correlation data by using MEM. Hollow diamonds denote experimental data. Solid line denotes MEM fitting with $G(\Gamma)$ shown in Figure 7 (dots). (b, Bottom) Plot of residuals of the fitting for solution PSPMT2 as a function of delay time with residuals defined in Figure 4b.

width distributions reliably, comparable to the capability of the CONTIN algorithm. The reconstructed characteristic line-width distributions corresponding to these solutions are displayed in Figure 7.

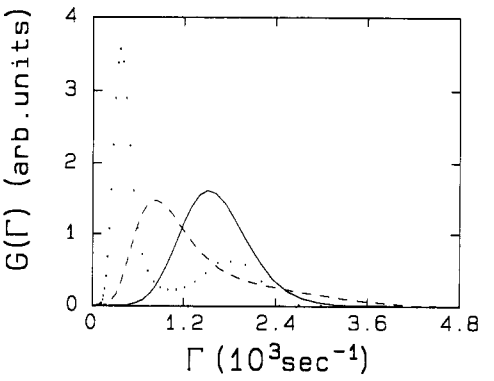

Figure 7. Plots of the MEM reconstructions of $G(\Gamma)$ corresponding to the time-correlation data (Figures 4–6) of solutions PSPMT0 (solid line), PSPMT1 (dashed line), and PSPMT2 (dots).

7. Concluding Remarks

In this paper we have pedagogically (1) introduced the maximum-entropy formalism, (2) outlined its use in estimating solutions to inverse problems in general and to the Laplace inversion problem in PCS in particular, (3) described two approaches for solving the nonlinearly constrained maximum-entropy optimization problem in PCS, and specified how its solution was achieved, and (4) tested the reliability of the formalism by using several sets of numerically simulated time correlation function data corresponding to known characteristic line-width distributions.

The main thrust of using the maximum-entropy (MEM) formalism lies in its proven fact of objectivity. In this formalism, information in the maximum-entropy solution is dictated purely by the available data and not by an arbitrarily introduced differential operator as in the regularization formalism. The maximum-entropy formalism requires introduction of a prior probability distribution in which we can summarize our prior information, such as any theoretical information about the physical problem of interest. Such an introduction depends on the nature of the problem. For most problems, the corresponding prior distributions are uniform. In the Laplace inverse problem in PCS, however, the prior distribution is not uniform but $\propto 1/\Gamma$ or $\propto 1/t$, depending on whether the characteristic line-width distribution $G(\Gamma)$ or the relaxation time distribution $H(t)$ is used for describing the system of interest. This choice of prior distribution leads to adopting a log Γ (or log t) description of $G(\Gamma)$ (or $H(t)$) and hence to specifying very broad $G(\Gamma)$ (or $H(t)$) more effectively.

We have used two methods for solving the nonlinear maximum-entropy problem, namely, the Newton–Raphson iterative approach (4a) and the quadratic model approximation (4b). While the Newton–Raphson iterative approach provides a positive definite solution $G(\Gamma)$, it is slow and relatively unstable. The quadratic model approximation is more efficient and yields a solution much faster, which is, unfortunately, not positively definite. Better approaches to solving the MEM problem could undoubtedly be devised. It should, however, be emphasized that knowledge of the error deviations of the experimental data can substantially improve the solution $G(\Gamma)$, including the resolution of $G(\Gamma)$. This error knowledge on the experimental data depends on subjective evaluation as well as systematic errors, which could be difficult to estimate. Preliminary analysis methods used for estimating or refining error deviations of measured data may introduce unintended bias to our $G(\Gamma)$ reconstruction. It remains to be seen as to what preliminary analysis method can best provide the least biased data deviations. As the analysis

must necessarily depend on the physical nature of the experiment, a generalized approach may be difficult to achieve. In addition to the noise in the measured data, we have alleviated the band-width limitation of the experimentally measured time correlation function by insisting on predetermining the value of β and by letting the first channel of the correlation function $|g^{(1)}(\Delta t)|^2$ be greater than 0.998 and the last channel value $|g^{(1)}(\text{last channel})|^2$ be less than 0.005.

In our comparison of MEM and CONTIN algorithms with simulated data, we have emphasized their utilities for unimodal $G(\Gamma)$ and the limitations related to negatively skewed $G(\Gamma)$ distributions and to bimodal $G(\Gamma)$ distributions, which are not δ functions. Finally, we demonstrated the application of MEM to experimental data and showed its utility comparable to that of the CONTIN algorithm.

In making an actual comparison of CONTIN and MEM methods using experimentally measured photon correlation data, it should be noted that with normal unimodal characteristic line-width distributions $G(\Gamma)$, both methods yield very good results. For skewed $G(\Gamma)$, the capability of CONTIN and MEM excels, as shown in Table I of ref 77. A comparison of two polystyrene species ($M_w = 4.48 \times 10^6$ g/mol, $M_w M_n = 1.14$ at 4.7×10^{-5} g/g and $M_w = 2.33 \times 10^5$ g/mol, $M_w/M_n = 1.06$ at 9.5×10^{-4} g/g) in benzene at 25 °C showed comparable findings by CONTIN and MEM.[77] Thus, at this time, even with our present algorithm, we can claim MEM to be at least comparable with CONTIN.

Acknowledgment. We should like to thank Dr. J. A. Potton for providing a maximum-entropy subprogram and Z.-L. Wang for supplying the experimental data. The financial support of this work by the Polymers Program, National Science Foundation (DMR 8617820), is gratefully acknowledged.

Registry No. PS, 9003-53-6; PMMA, 9011-14-7.

References and Notes

(1) Burchard, W. *Adv. Polym. Sci.* **1983**, *48*, 1.
(2) Chu, B. *J. Polym. Sci., Polym. Symp.* **1985**, *73*, 137.
(3) Ying, Q.; Chu, B. *Macromolecules* **1986**, *19*, 1580.
(4) Wu, C.; Buck, W.; Chu, B. *Macromolecules* **1987**, *20*, 98.
(5) Chu, B.; Wu, C.; Zuo, J. *Macromolecules* **1987**, *20*, 700.
(6) Livesey, A. K.; Licinio, P.; Delaye, M. *J. Chem. Phys.* **1986**, *84*, 5102.
(7) Licinio, P.; Delaye, M.; Livesey, A. K.; Leger, L. *J. Phys.* **1987**, *48*, 1217.
(8) Livesey, A. K.; Delaye, M.; Licinio, P.; Brochon, J.-C. *Faraday Discuss. Chem. Soc.* **1987**, *14*, 83.
(9) Zulauf, M.; Eicke, H.-F. *J. Phys. Chem.* **1979**, *83*, 480.
(10) Nicholson, J. D.; Doherty, J. V.; Clarke, J. H. R. In *Microemulsions*; Robb, I. D., Ed.; Plenum Press: New York, 1982.
(11) Siegert, A. J. F. *MIT Rad. Lab. Report* **1943**, *No. 465*.
(12) Rust, B. W.; Burrus, W. R. *Mathematical Programming and the Numerical Solution of Linear Equations*; American Elsevier: New York, 1972.
(13) Lewis, B. A. *J. Inst. Maths. Applics.* **1975**, *16*, 207.
(14) Tikhonov, A. N.; Arsenin, V. Y. *Solutions of Ill-Posed Problems*; V. H. Winston & Sons: Washington, DC, 1977.
(15) Golberg, M. A., Ed. *Solution Methods for Integral Equations: Theory and Applications*; Plenum Press: New York and London, 1979.
(16) Lavrent'ev, M. M.; Romanov, V. G.; Shishat-skii, S. P. *Ill-Posed Problems of Mathematical Physics and Analysis*; American Mathematical Society: Providence, 1986.
(17) Provencher, S. W. *Makromol. Chem.* **1979**, *180*, 201.
(18) Provencher, S. W. *Comp. Phys. Commun.* **1982**, *27*, 213.
(19) Provencher, S. W. *Comp. Phys. Commun.* **1982**, *27*, 229.
(20) Koppel, D. E. *J. Chem. Phys.* **1972**, *57*, 4814.
(21) Chu, B. In *The Application of Laser Light Scattering to the Study of Biological Motion*; Earnshaw, J. C., Steer, M. W., Eds.; Plenum Pub: New York, 1983.
(22) Stock, R. S.; Ray, W. H. *J. Polym. Sci., Polym. Phys. Ed.* **1985**, *23*, 1393.
(23) Chu, B.; Ford, J. R.; Dhadwal, H. S. "Enzyme Structure Part J". *Methods Enzymol.* **1987**, *117*, 256.
(24) Reinsch, C. H. J. *Numerische Mathematik* **1967**, *10*, 177.
(25) Graham, N. Y. *Bell System Technol. J.* **1983**, *62*, 101.
(26) Ravindran, A. *Commun. ACM* **1972**, *15*, 818.
(27) Proll, L. G. *Commun. ACM.* **1974**, *17*, 590.
(28) Beck, J. V.; Blackwell, B.; St. Clair, C. R., Jr. *Inverse Heat Conduction: Ill-Posed Problems*; Wiley-Interscience: New York, 1985.
(29) Grachev, I. D.; Salakhov, M. K.; Fishman, I. S. *Comp. Enhanced Spectrosc.* **1984**, *2*, 1.
(30) Salakhov, M. Kh.; Grachev, I. D.; Latipov, R. Z.; Fishman, I. S. *Comp. Enhanced Spectrosc.* **1984**, *2*, 117.
(31) Provencher, S. *CONTIN (Version 2) Users Manual*; Technical Report EMBL-DA07: Heidelberg, 1984.
(32) Danovich, G. D.; Serdyuk, I. N. In *Photon Correlation Techniques in Fluid Mechanics*; Schulz-DuBois, E. O., Ed.; Springer-Verlag: Berlin, 1983.
(33) Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620.
(34) Jaynes, E. T. *Phys. Rev.* **1957**, 108, *171*.
(35) Levine, R. D.; Tribus, M., Eds. *The Maximum Entropy Formalism*, A Conference Held at the MIT on May 2–4, 1978; MIT Press: Cambridge, 1979.
(36) Levine, R. D. *J. Phys.* **1980**, *A13*, 91.
(37) Jaynes, E. T. In *Inverse Problems*; McLaughlin, D. W., Ed.; American Mathematical Society: Providence, 1984.
(38) Shore, J. E.; Johnson, R. W. *IEEE Trans. Inform. Theory* **1980**, *IT-26*, 26.
(39) Frieden, B. R. *J. Opt. Soc. Am.* **1972**, *62*, 511.
(40) Rietsch, E. *J. Geophys.* **1977**, *42*, 489.
(41) Collins, D. M. *Nature* **1982**, *298*, 49.
(42) Sibisi, S. *Nature* **1983**, *301*, 134.
(43) Burch, S. F.; Gull, S. F.; Skilling, J. *J. Comp. Vis. Grap. Image Proc.* **1983**, *23*, 113.
(44) Skilling, J.; Bryan, R. K. *Mon. Not. R. Astr. Soc.* **1984**, *211*, 111.
(45) Bricogne, G. *Acta Crystallogr.* **1984**, *A40*, 410.
(46) Cottrell, G. A. *Rev. Sci. Instrum.* **1984**, *55*, 1401.
(47) Gull, S. F.; Skilling, J. *IEE Proc.* **1984**, *131*, 646.
(48) Skilling, J.; Gull, S. F. In *Maximum-Entropy and Bayesian Methods in Inverse Problems*; Smith, C. R., Grandy, W. T., Jr., Eds.; D. Reidel: Dordrecht, 1985.
(49) Steenstrup, S. *Aust. J. Phys.* **1985**, *38*, 319.
(50) Livesey, A. K.; Skilling, J. *Acta Crystallogr.* **1985**, *A41*, 113.
(51) Root, J. H.; Egelstaff, P. A.; Nickel, B. In *Workshop on Neutron Scattering Data Analysis 1986*; Johnson, M. W., Ed.; J. W. Arrowsmith: Bristol, 1986.
(52) Brown, W.; Johnsen, R.; Stepanek, P.; Jakes, J. *Macromolecules* **1988**, *21*, 2859.
(53) Wilson, A. G.; Coelho, J. D.; Macgill, S. M.; Williams, H. C. W. L. *Optimization in Locational and Transport Analysis*; John Wiley & Sons: Chichester, 1981.
(54) Shannon, C. E. *Bell System Technol. J.* **1948**, *27*, 379.
(55) Katz, A. *Principles of Statistical Mechanics: The Information Theory Approach*; W. H. Freeman & Co.: San Francisco & London, 1967.
(56) Hobson, A. *Concepts in Statistical Mechanics*; Gordon & Breach Science Pub.: New York, 1971.
(57) Landsberg, P. T. *Thermodynamics and Statistical Mechanics*; Oxford University: Oxford, 1978.
(58) Kullback, S. *Information Theory and Statistics*; Wiley: New York, 1959.
(59) Toutenburg, H. *Prior Information in Linear Models*; John Wiley & Sons: Chichester, 1982.
(60) Jeffreys, H. *Theory of Probability*, 3rd ed.; Oxford University Press: London, 1961.
(61) Jaynes, E. T. *IEEE Trans. Systems Sci. Cybernetics* **1968**, *SSC-4*, 227.
(62) Tarantola, A.; Valette, B. *J. Geophys.* **1982**, *50*, 159.
(63) Barrodale, I.; Zala, C. In *Numerical Algorithms*; Mohamed, J. L., Walsh, J., Eds.; Oxford University Press: Oxford, 1986.
(64) Cramer, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, 1946.
(65) Bartlett, M. S. *Biometrics* **1947**, *3*, 39.
(66) Kendall, M. G.; Stuart, A. "The Advanced Theory of Statistics". *Design and Analysis, and Time-Series*, 2nd ed.; Hafner Pub. Co.: New York, 1968; Vol. 3.
(67) Jakeman, E.; Pike, E. R.; Swain, S. *J. Phys.* **1971**, *A4*, 519.
(68) Schätzel, K. *Opt. Acta* **1980**, *27*, 45.
(69) Alhassid, Y.; Agmon, N.; Levine, R. D. *Chem. Phys. Lett.* **1978**, *53*, 22.
(70) Wismer, D. A.; Chattergy, R. *Introduction to Nonlinear Optimization: A Problem Solving Approach*; North-Holland: New York, 1978.
(71) Harley, P. J. In *Numerical Algorithms*; Mohamed, J. L., Walsh, J., Eds.; Oxford University Press: Oxford, 1986.
(72) Freeman, T. L. In *Numerical Algorithms*; Mohamed, J. L., Walsh, J., Eds.; Oxford University Press: Oxford, 1986.

(73) Beightler, C. S.; Phillips, D. T.; Wilde, D. J. *Foundations of Optimization*, 2nd ed.; Prentice-Hall: Engelwood Cliffs, NJ, 1979.
(74) Marquardt, D. W. *SIAM J. Appl. Math.* **1963**, *11*, 431.
(75) Birkhoff, G.; Mac Lane, S. M. *A Survey of Modern Algebra*, 3rd ed.; Macmillan: New York, 1965.
(76) Wang, Z.-L.; Wang, Q.-W.; Fetters, L.; Chu, B. In *Proceedings of the Chemistry International Symposium on New Trends in Physics and Physical Chemistry of Polymers Honoring Professor P. G. de Gennes*, in press.
(77) Chu, B.; Xu, R.-L.; Nyeo, S.-L. *Part. & Part. Syst. Charact.* **1989**, *6*, 34.

Formation and Torsion Dynamics of the Stereocomplex of Isotactic and Syndiotactic Poly(methyl methacrylates) Studied by the Fluorescence Depolarization Method

Takashi Sasaki and Masahide Yamamoto*

Department of Polymer Chemistry, Kyoto University, Kyoto 606, Japan.
Received November 2, 1988; Revised Manuscript Received March 15, 1989

ABSTRACT: The dynamics of the stereocomplex of poly(methyl methacrylate) (PMMA) was studied by the fluorescence depolarization method. Anthracene-labeled syndiotactic PMMA was mixed with isotactic PMMA in toluene dilute solutions to form the anthracene-labeled stereocomplex. The formation of the stereocomplex was followed over the period of 3 weeks by the steady-state measurement. The stoichiometry of the complex formation was revealed to be isotactic/syndiotactic = 1/2 for this system. Time-resolved fluorescence emission anisotropy data, $r(t)$, were well explained by the intermediate zone formula of the torsion dynamics theory for stiff macromolecules. This fact supports the idea of the double-stranded helical structure for the stereocomplex. The torsional rigidity of the stereocomplex under the gel structure was estimated from $r(t)$ to be $\alpha = 2.4 \times 10^{-10}$ dyn cm and $C = 6.0 \times 10^{-18}$ dyn cm^2 at 25 °C, which are values larger than those obtained for DNA.

I. Introduction

Isotactic and syndiotactic stereoregular poly(methyl methacrylates) (i-PMMA and s-PMMA) associate in certain proper solvents, in a so-called stereocomplex.[1-3] The complexation occurs at the weight ratio of i-PMMA/s-PMMA = 1/2 in certain solvents.[4] The association is said to be caused by nonbonded interactions of CH_3 groups and ester groups.[5] Recent investigations on this phenomenon have revealed the structure and the complexation mechanism to some extent.[6-8] Also the stereospecific polymerization of methyl methacrylate (MMA) in the presence of stereoregular PMMA as a template is closely related to the stereocomplex formation.[9,10] Recently, the mechanical properties of the PMMA stereocomplex have been studied as an application.[11]

Bosscher et al.[7] investigated the structure of the stereocomplex of i-PMMA and s-PMMA by X-ray diffraction and by conformational energy calculations. Their results revealed a double-stranded helical structure consisting of an isotactic 30/4 helix surrounded by a syndiotactic 60/4 helix, with a real fiber period of 73.6 Å. The structure is illustrated by contours of i-PMMA and s-PMMA in Figure 1. The helical pitch is 18.4 Å, which corresponds to 7.5 monomer units of i-PMMA and 15 monomer units of s-PMMA. In this structure, one monomer unit of i-PMMA meets two monomer units of s-PMMA, and this idea is consistent with the weight ratio of the stereocomplex that i-PMMA/s-PMMA = 1/2. The dynamic behavior in solutions of the stereocomplex of such a helical structure may resemble that of a double-stranded helix of DNA rather than that of synthetic flexible polymer chains. However, no investigations on the dynamics of the PMMA stereocomplex have been done so far.

The fluorescence method was utilized to investigate the formation of other polymer complexes. Morawetz and his co-workers[12,13] investigated the kinetics of the complex formation of poly(acrylic acid) (PAA) and poly(oxyethylene) (POE) by detecting the fluorescence intensity of the dansyl group attached to the PAA chain. Anufrieva et al.[14] examined the exchange of the components of the PAA–POE complex by the fluorescence depolarization method. Uchida[15] also investigated the complexation of stereoregular PMMA, which was labeled with perylene, by the steady-state measurement of fluorescence depolarization. In general, the fluorescence depolarization measurement provides the information of local mobility, and this method is expected to be fairly sensitive to the complexation phenomenon, which is associated with drastic changes in the local environment.

As for the dynamics of rigid macromolecules such as DNA, theoretical treatment of a deformable filament model of twisting and bending has been developed,[16,17] and it has become possible to investigate the torsion dynamics of DNA from experimental data of fluorescence polarization anisotropy,[18,19] electric birefringence,[20,21] and transient photodichroism.[22] In the time region within 100 ns, the relaxation phenomena are mainly attributed to the torsional (twisting) motions in the case of a sufficiently long chain. The theory also treats internal relaxations such as libration or wobbling of the intercalated dye in a very rapid time region as in the picosecond fluorescence depolarization experiments.

In the present work, we investigated the formation and dynamics of the stereocomplex of isotactic and syndiotactic PMMAs in toluene solutions by the fluorescence depolarization method in the nanosecond time region. For this purpose, s-PMMA labeled with the anthracene group in the middle of the chain was mixed with unlabeled i-PMMA in dilute toluene solutions to form the anthracene-labeled stereocomplex filament. The labeled s-PMMA has also been used to investigate the local conformational motions in dilute solutions,[23] and the transition vector of the anthracene label lies parallel to the backbone of the s-PMMA chain. The theory of torsion dynamics was applied to the

0024-9297/89/2222-4009$01.50/0