

Global fitting without a global model: Regularization based on the continuity of the evolution of parameter distributions

Jason T. Giurleo and David S. Talaga^{a)}

*Department of Chemistry and Chemical Biology and BIOMAPS Institute, Rutgers,
The State University of New Jersey, Piscataway, New Jersey 08854, USA*

(Received 12 October 2007; accepted 3 January 2008; published online 21 March 2008)

We introduce a new approach to global data fitting based on a regularization condition that invokes continuity in the global data coordinate. Stabilization of the data fitting procedure comes from probabilistic constraint of the global solution to physically reasonable behavior rather than to specific models of the system behavior. This method is applicable to the fitting of many types of spectroscopic data including dynamic light scattering, time-correlated single-photon counting (TCSPC), and circular dichroism. We compare our method to traditional approaches to fitting an inverse Laplace transform by examining the evolution of multiple lifetime components in synthetic TCSPC data. The global regularizer recovers features in the data that are not apparent from traditional fitting. We show how our approach allows one to start from an essentially model-free fit and progress to a specific model by moving from probabilistic to deterministic constraints in both Laplace transformed and nontransformed coordinates. © 2008 American Institute of Physics.
[DOI: 10.1063/1.2837293]

I. INTRODUCTION

Sample heterogeneity is nearly unavoidable in spectroscopy. In many simple systems the heterogeneity can be reduced to the point where it can be ignored. However, in complex biological systems there is much to be gained from understanding the heterogeneity. The biological machinery in the cell, for example, relies on the dynamic nature of proteins and protein assemblies.¹ Intermediate species have been identified in protein (mis)folding mechanisms.^{2–6} Multiple conformations of protein-ligand complexes have been discovered because of signal heterogeneity.⁷ The evolution of multiple binding sites for β -lactoglobulin and the time-dependence of the site-binding entropy in response to the sudden presence of a strong dipole was elucidated by the heterogeneity in a time-dependent Stokes-shift measurement.⁸ Heterogeneity is a crucial element of interpreting single molecule measurements.^{3,9–12} The presence of multiple species in misfolded β -lactoglobulin leads to very heterogeneous signals prior to^{13,14} and during the assembly of amyloid.¹⁵ To fully understand the underlying physics of such complex systems requires approaches to data reduction that can accommodate their heterogeneity.

Data reduction, or fitting, always requires a model. For any phenomenon being measured there is a set, or space, of models that could reasonably be expected to explain the data. The fitting procedure should eliminate all parts of the model space inconsistent with the data. From the remaining possibilities, the algorithm should allow selection of the most likely model given the experimental information. Experimental information in this context includes both the explicit data that comes from the instrumentation as well as knowledge that comes from the experimental design and under-

standing of the physics of the system. These two sources of information we will call “data” and “prior knowledge.”¹⁶

Prior knowledge is used to determine the model space for the problem. Many experiments—including dynamic light scattering,^{17–20} circular dichroism,²¹ dynamic nuclear magnetic resonance,²² and fluorescence lifetime²³—can be related to their underlying physics by Fredholm integral equations of the first kind

$$F(t,y) = \int_a^b A(t,k,y)f(k,y)dk. \quad (1)$$

Data inversion seeks to find the function, $f(k,y)$, from the noisy signal, $F(t,y)+\epsilon(t,y)$, given the continuous kernel function, $A(t,k,y)$, that connects t and k , with y remaining untransformed.

The mathematical model space comprises all possible functions $f(k,y)$. In most circumstances the physical interpretation of $f(k,y)$ is that it represents a distribution of populations of species that differ in their values of k . The goal of the experiment is usually to determine the systematic behavior of those populations with experimental changes in y , which can represent any experimental condition such as pH, incubation time, solvent polarity, etc. For example, determining the species population changes with temperature allows thermodynamic parameters to be extracted.

For experiments that monotonically decay such as fluorescence lifetime and dynamic light scattering measurements, the kernel in Eq. (1) is $A(t,k,y)=\exp(-kt)$ and $a=0$, and $b=\infty$ provide the integration limits for a Laplace transform

^{a)}Electronic mail: talaga@rutgers.edu.

$$F(t, y) = \mathcal{L}\{f(k, y)\} = \int_0^{\infty} e^{-kt} f(k, y) dk. \quad (2)$$

Solving for $f(k, y)$ requires inversion of integral Eq. (2).

In the presence of noise, the inverse Laplace transform is not unique. An explicit expression for the inverse Laplace transform²⁴ illustrates the difficulty of direct inversion

$$f(k, y) = \mathcal{L}^{-1}\{F(t, y)\} = \lim_{i \rightarrow \infty} \frac{(-1)^i}{i!} \left[\frac{\cdot}{k} \right]^{i+1} F^{(i)}\left(\frac{i}{k}, y\right), \quad (3)$$

where $F^{(i)}$ is the i th derivative of F . The inverse Laplace transform is sensitive to high order derivatives that do not exist in real data in any meaningful way; each subsequent derivative increases the appearance of noise. As a result, all functions, $f(k, y)$, that have Laplace transforms, $F(t, y)$, with similar first few derivatives, but divergent higher order derivatives, will be valid solutions to the inverse Laplace transform.

Since direct inversion is problematic, an indirect approach based on least-squares optimization is typical^{25–27}

$$\min \left[\left(\int_a^b A(t, k, y) f(k, y) dk - F(t, y) \right)^2 \right]. \quad (4)$$

Expression (4), though more stable, still suffers from the problem that multiple solutions for $f(k, y)$ will be statistically equivalent. As a result, the strategy that a particular optimization algorithm uses to obtain the inversion, $f(k, y)$, can influence which solution is found, or even if a solution is found.

Many algorithms exist to solve the general least-squares problem defined by Eq. (4) and determine the inverse transform, $F(k, y)$.^{25,26} A direct solution to this problem can be obtained using methods such as singular value decomposition.²⁶ However, these approaches generate solutions that are highly sensitive to the details of the noise and typically give many negative values for the parameters. Therefore, approaches that allow constraint of the solution to positive values are desirable.

One of the most common methods for least-squares minimization is Levenberg–Marquardt (LM).^{28,29} This method is fast for small to medium sized data sets and is appropriate for both linear and nonlinear models, but is subject to potential nonlinear instabilities.^{26,30} A gridded representation of the kernel $A(k, t, y)$ is called a design matrix and eliminates nonlinear instabilities because it allows the use of linear least-squares algorithms like the active-set (AS) and the interior-point gradient (IPG) methods.^{31,32} Though the setup is common in both methods, each minimizes the general least-squares expression very differently.

The AS method enforces non-negativity on $f(k, y)$ by adding and removing basis functions from the kernel (design matrix).³¹ This corresponds to expanding and contracting the dimensionality of that solution space, in effect finding the “right number” of exponentials on the grid to represent the fit. This method is computationally expensive for very large scale problems because of the continual matrix factorizations necessary to identify the active set.^{32,33}

The IPG method has an adaptive step size that accelerates convergence while limiting itself to steps that maintain the non-negativity of the solution at all points in the search, unlike the AS method.³² IPG is a typical gradient method in that it uses the entire set of basis functions and requires only matrix-vector multiplications.^{34,35} Though these methods are considered to be less accurate than the traditional AS method,³² the decreased convergence time of IPG for large problems makes it very useful. IPG accomplishes this by exploiting the totally non-negative structure of the problem to calculate the step and direction giving fits that converge faster than the classic active-set method.³²

Because many different elements of the model space will give equally good fits to the noisy data,³⁰ all three methods we mention must constrain the model space, in some way, to stabilize the inverse Laplace transform. Since many of the solutions may not be physically reasonable, constraints can be selected based on prior knowledge. When $f(k, y)$ is a population then $f(k, y) \geq 0$. Instrumental limitations can limit the range of k that can be determined. When y is varied systematically then the global behavior of the individual species populations can be invoked to limit the possible $f(k, y)$. For example, the conservation of matter could provide a normalization condition for $f(k, y)$ at each value of y . The systematic variation of y could provide an expectation of a particular functional dependence of $f(k, y)$ along y for particular values of k representing different species. Limiting the model space with prior knowledge can be done with deterministic or probabilistic constraints.

Constraints of both types may be applied to the transformed variable, k , and the nontransformed variable, y . Constraints in the transformed variable are formed from hypotheses regarding the nature of the system, $f(k, y)$, across k . This is illustrated by the upper right panel of Fig. 1 where continuity has been invoked along k . Constraints in the nontransformed variable (such as time, temperature, solvent, etc.) require hypotheses that generate global models leading to specific functions or continuity conditions describing the evolution of the system across y . This is illustrated by the lower right panel of Fig. 1.

Deterministic constraints consider only some very small subset of functions in the model space and forbid all nonconforming solutions. Such empirical functions make the tacit prior assumption of a specific physical model for the system that gives analytical forms for $f(k, y)$ that transform trivially. A least-squares fit to $F(t, y)$ with these functions is equivalent to the inversion. The most common way is to use a small number (i.e., 1–3) exponential function to fit the data. This limits the model space $f(k, y)$ to a small number of delta functions. All other possible solutions have their likelihood set to zero *a priori*. Global models are constructed by assigning specific functions to describe the evolution of $f(k, y)$ along y for all of the discrete values of k that represent a species. A drawback is that this can forbid the “true” solution from being obtained if it is not, by happenstance, consistent with the deterministic constraints.

By contrast, probabilistic constraints reduce the likelihood of solutions in proportion to their departure from the constraint, but do not set the likelihood of any solutions in

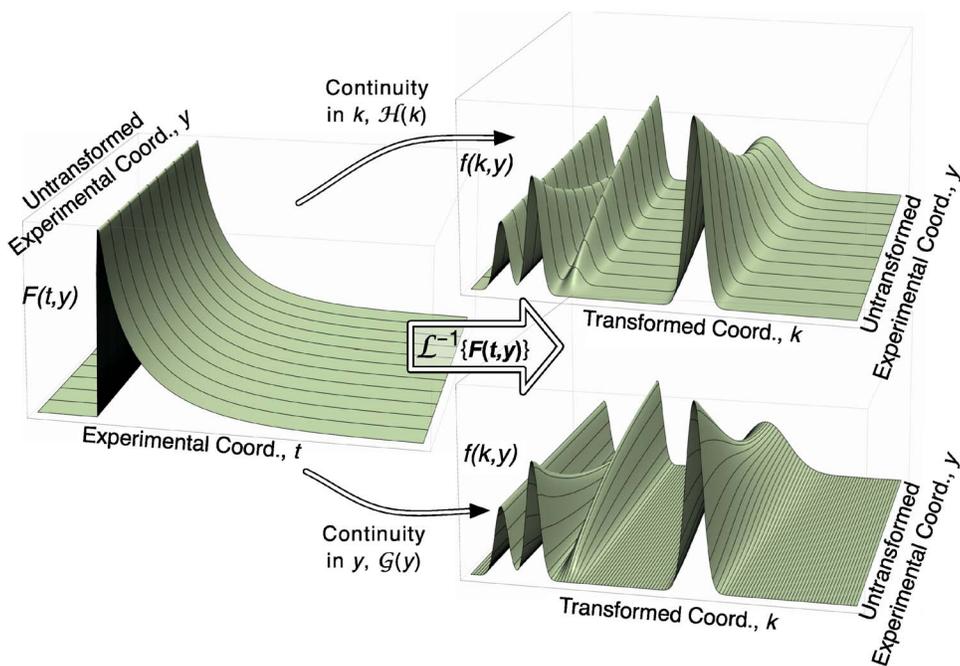


FIG. 1. (Color online) Experimental prior knowledge. With global fitting available, prior knowledge in the form of a regularizer can either be applied in the Laplace transform dimension, k , or in the experimental evolution dimension, y . In the left panel the individual data transients are indicated by the black mesh lines. In the right panels, the direction of continuity conditions is indicated by the black mesh lines. The top right panel emphasizes the continuity in the k dimension implying continuous distributions of properties. The bottom right panel emphasizes continuity in the y dimension implying continuous evolution of population.

the model space to zero. Probabilistic constraints are usually imposed by penalty functions or regularization conditions. The new fitting functional including regularization is

$$\Xi^2 = \chi^2 + \gamma\Omega^2, \quad (5)$$

where χ^2 is the usual sum of the weighted squared residuals, Ω^2 is any arbitrary regularizer functional, and γ is the strength of the regularization.

Early work on probabilistic constraints^{17,30} used a second derivative regularizer functional of the form

$$\Omega^2 = \mathcal{H}(k) = \int \left[\frac{\partial^2 f(k, y)}{\partial k^2} \right]^2 dk, \quad (6)$$

with the assumption that this would provide the most “parsimonious” discretized solution to the inversion problem.¹⁷ This regularizer implies piecewise linearity in k , strongly biasing against any discontinuities. There will be many systems for which this will not be valid, such as those having discrete distributions across k . Other regularizers have been used to probabilistically reduce the available solution space for the inversion by imposing expectations of the nature of the solution in k .^{36,37}

Maximum entropy is another criterion that is often used as a regularizer,^{23,36,38,39}

$$\mathcal{W} = \iint f(k, y) - m(k, y) - f(k, y) \log \left(\frac{f(k, y)}{m(k, y)} \right) dk dy, \quad (7)$$

where $m(k, y)$ is the uniform distribution that gives maximum entropy. The entropy functional satisfies two conditions that serve to define it. Entropy is maximized for flat distributions. This has the physical meaning that all values of k are not only possible but equally likely. Entropy also has the property that it is maximized when, in a joint distribution, the likelihood is independent.^{16,40} That is, when $f(k, y) = f(k) \times f(y)$. This has the physical meaning that all species have identical behavior under the influence of the experi-

mental variable. This is an excellent example of how assumptions about the behavior of the system can creep in without the explicit knowledge of the investigator. Also, since entropy is an extensive property, performing a maximum entropy analysis globally is identical to performing parallel local analyses, so long as the likelihood at each value of y is normalized across k . This suggests that though maximum entropy is a useful condition for probabilistic constraint of fitting, it is of limited benefit for global fitting in this context.

In this paper we introduce a method to exploit the global behavior of a system (e.g., continuity) across an experimental coordinate, y , to define a global regularization condition that does not make any assumptions about the shape or continuity in the Laplace transform dimension, k . In a typical experiment, k is directly related to some property of the different species present and the concentrations of the different species change with the experimental coordinate, y , (e.g., over time by some kinetic rate law). Therefore, the solution set, $f(k, y)$, will often be piecewise continuous in y , implying smooth changes in the population of individual species with respect to y . To represent this behavior mathematically, we introduce a global regularizer that favors solutions that satisfy the continuity/smoothness condition in y ,

$$\mathcal{G}(y) = \int \left[\frac{\partial^2 f(k, y)}{\partial y^2} \right]^2 dy. \quad (8)$$

Implementation of Eq. (8) into Eq. (5) requires a general least-squares optimization algorithm that can accommodate the inclusion of a regularizer term. The IPG method minimizes Eq. (4) by choosing the scale of the step and direction to be as close as possible the exact minimizer without crossing into the non-negative region. We augment the IPG with a regularizer functional such that the step and direction favor solutions with continuity in the either the k , local, or y , global, dimensions.

To demonstrate utility of global regularization on an

evolving data set, we simulate a test data set, evaluate fitting methods and compare and contrast existing method with methods introduced here. We also evaluate the effect of noise on each method's ability to reproduce the test parameter set. Last, we show how to use global regularization in the experimental domain with a traditional physical model for the species domain to do global fitting using LM.

II. METHODS

A. IPG regularizer

The computer implementation of Eq. (4) replaces continuous variables t , y , and k with discrete variables t_i , k_j , and y_l with $i \in \{1 \cdots T\}$, $j \in \{1 \cdots K\}$, and $l \in \{1 \cdots Y\}$. Matrix notation replaces continuous notation as follows. Each transient, $F(t, y_l)$ has a design matrix

$$\mathbf{A}_l^{[T \times K]} \equiv A(t, k, y_l) / s(t, y_l),$$

a solution vector

$$\mathbf{x}_l^{[K \times 1]} \equiv f(k, y_l),$$

and a data vector

$$\mathbf{b}_l^{[T \times 1]} \equiv F(t, y_l) / s(t, y_l).$$

The design matrix and the data vector are scaled by the standard deviations of the data, which were estimated by $s(t, y_l) = \sqrt{F(t, y_l) + 1}$.

We momentarily drop the subscripts in our notation when considering one transient at a time. The minimum of Eq. (4) occurs when the derivatives with respect to the parameters equal zero, yielding the exact minimizing equation

$$\mathbf{A}^T \mathbf{A} \mathbf{x} - \mathbf{A}^T \mathbf{b} = 0. \quad (9)$$

The details of the fitting algorithm can be found in Ref. 32. The key feature of the algorithm is the way it determines the scaling vector for calculating each iteration

$$\mathbf{d} = \frac{\mathbf{x}}{\mathbf{A}^T \mathbf{A} \mathbf{x}}. \quad (10)$$

Equation (10) is the core of the IPG algorithm and makes it less sensitive to ill-conditioned problems. Reference 32 shows that with a condition number of $\mathbf{A}^T \mathbf{A} \approx 10^{16}$, the relative error of the fits between IPG and the active-set after 10^3 iterations is $\approx 10^{-5}$ and nearly five orders of magnitude better than other scaling methods. For our global fits, typical condition numbers range from 10^{18} to 10^{22} making the insensitivity of IPG to condition number particularly valuable.

1. Locally regularized IPG

The same minimization principle can be applied when considering the cost function, Ξ^2 , in [Eq. (5)],

$$(\mathbf{A}^T \mathbf{A} + \gamma \mathbf{H}) \mathbf{x} - \mathbf{A}^T \mathbf{b} = 0, \quad (11)$$

where the regularizer \mathcal{H} is now represented as the regularization matrix \mathbf{H} ,

$$\mathbf{H}^{[K \times K]} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ -2 & 5 & -4 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \ddots & & & & \\ 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & \cdots & 0 & 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{pmatrix}, \quad (12)$$

where again K is the number of parameters for each transient. This second derivative regularizer matrix measures departure from piecewise linearity in $f(k)$. Other regularizers will be appropriate depending on the available prior knowledge. Details for generating different types of regularization matrices are found in chapter 18–5 of Ref. 26. For locally regularized IPG, we substitute $(\mathbf{A}^T \mathbf{A} + \gamma \mathbf{H})$ for all occurrences of $\mathbf{A}^T \mathbf{A}$ in the IPG algorithm.

2. Globally regularized IPG

The global fits simultaneously consider all the available Y transients. Setting up globally regularized IPG is analogous to the locally regularized case, except that the entire data set is fit at once. The matrices and vectors must accommodate the global nature of the problem. This requires concatenation of local data transients and parameters into global vectors

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_Y \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_Y \end{pmatrix}. \quad (13)$$

The design matrices, \mathbf{A}_l are placed in a block diagonal global-design matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{A}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}_Y \end{pmatrix}. \quad (14)$$

The regularization matrix is also expanded to incorporate continuity across the nontransformed dimension. Equation (8) specifies a global regularization matrix, $\mathbf{G}^{[KY \times KY]}$. To obtain $\mathbf{G}^{[KY \times KY]}$, each element $(h_{1, \dots, Y, 1, \dots, Y})$ of Eq. (12), $(\mathbf{H}^{[Y \times Y]})$ is replaced by a $K \times K$ diagonal matrix

$$\mathbf{G} = \begin{pmatrix} h_{1,1} \mathbf{I} & h_{1,2} \mathbf{I} & \cdots & h_{1,Y} \mathbf{I} \\ h_{2,1} \mathbf{I} & h_{2,2} \mathbf{I} & \cdots & h_{2,Y} \mathbf{I} \\ \vdots & & \ddots & \vdots \\ h_{Y,1} \mathbf{I} & h_{Y,2} \mathbf{I} & \cdots & h_{Y,Y} \mathbf{I} \end{pmatrix}, \quad (15)$$

where \mathbf{I} is a $K \times K$ identity matrix.

For global regularization, we substitute $(\mathbf{A}^T \mathbf{A} + \gamma \mathbf{G}) \rightarrow \mathbf{A}^T \mathbf{A}$ in the IPG algorithm.

B. Data simulations

Ten ($Y=10$) time-correlated single-photon counting (TCSPC) measurements were synthesized with $T=4096$ bins over a 50 ns time-window range to evaluate the utility of the global regularization method. Each transient, $F(t, y_i)$, is generated by the matrix multiplication of an exponential decay matrix \mathbf{S} [with each element, $S_{ij}=\exp(-k_j t_i)$] by a $K=1200$ synthesized solution set (the exemplar) $\tilde{\mathbf{f}}=\tilde{f}(k, y_i)$. The synthesized solution set was evenly spaced in k^{-1} , from 0.01 to 12 ns. The grid for the simulation was tenfold denser than that used for the fit to better simulate the presence of continuous distributions.

The exemplar solution set, $\tilde{f}(k, y)$, was generated by representing four evolving species, each by a Gaussian distribution

$$\tilde{f}(k, y) = \sum_{i=1}^4 C_i (\sigma_i \sqrt{2\pi})^{-1} e^{-1/2[(k - k_i)/\sigma_i]^2}. \quad (16)$$

Such fluorescence decay components commonly arise and are difficult to fit.^{36,37,41,42} Species 1 was centered at $k_1^{-1}=7$ ns, concentration $C_1=1$ and had an evolving width

$$\sigma_1(y) = \left(0.2 + \frac{0.7}{e^{-(5-y)/2} + 1} \right) \text{ ns}.$$

Species 2 ($k_2^{-1}=2$ ns, $\sigma_2^{-1}=0.2$ ns) has an initial concentration of $C_2=0.5$ and converts into species 3 ($k_3^{-1}=3$ ns, $\sigma_3^{-1}=0.2$ ns) with a first order rate of 0.2/experimental unit in y . Species 4 was nonevolving in both k and y , ($k_4^{-1}=1$ ns, $\sigma_4^{-1}=0.2$ ns) with constant concentration of $C_4=0.3$. The right two panels in Fig. 1 show the exemplar solution, $\tilde{f}(k, y)$, for the evolving test case, $F(t, y_i)$.

The intensity decays generated were convoluted with an instrument response function typical of microchannel plate photomultipliers.^{7,43} We examined the influence of noise on the fitting algorithms by generating data sets with different intensity levels. The peak convoluted decay intensities were scaled to 11 values: $\mathcal{I}=10^2$, 2×10^2 , 5×10^2 , 10^3 , 2×10^3 , 5×10^3 , 10^4 , 2×10^4 , 5×10^4 , 10^5 , and 10^8 . We included uncorrelated background (e.g., dark counts) at an intensity level of 30 counts for every bin for all signal-to-noise (S:N) levels. The noise of photon counting was simulated from a Poisson distribution at the intensity of each bin in $F(t, y)$. For example, the case with an intensity of $\mathcal{I} \sim 10^4$ peak counts and 30 background counts has a S:N level of $\sim 100:1$ at the peak.

C. Fitting mechanics

All fits were performed with Igor Pro 6.01 (Wavemetrics, Inc.) running on a 2.16 GHz Intel DuoCore MacBook Pro under Mac OS X 10.4 (Tiger) with 2 GB of RAM. Active-set and IPG were implemented as user-defined functions. Nonlinear least-squares fitting was performed implementing IGOR's Levenberg–Marquardt (LM) curvefit package.

1. Levenberg–Marquardt

We performed four types of exponential, instrument-response-convoluted fits utilizing the LM method: Three exponential model, four exponential model, global model, and a regularized global model fit. For the global model fit, the entire data set was fit simultaneously. The global model was chosen based on the global IPG fits, the full rationale for which appears in the results.

Three and four exponential fits were of the form $F(t, y) = F_0(y) + \sum_{i=1}^n P_i(y) e^{-k_i t}$, where $n=3$ or 4, respectively. All parameters F_0 , P_i , and k_i were unconstrained. The data were weighted by the estimated standard errors. Each data transient was fit individually. To provide the best likelihood of successful convergence, the initial guesses were determined from a successful multiexponential fit for the high S:N limit ($\mathcal{I}=10^8$). For each S:N ratio the initial guesses were scaled according to the intensity and used for the first transient. The resulting fit parameters were then used as the initial guess for the next transient in the data set. Convergence of the algorithm occurred when either one of the two conditions were met: The number iterations reached a maximum of 100, or fractional decrease of sum of the weighted residuals, χ^2 , from one iteration to the next was less than 0.001.

Global models were performed on an entire data set for a particular signal-to-noise ratio. The chosen model for the global fit was represented by one stretched exponential plus three exponentials

$$F(t, y) = F_0(y) + P_1(y) e^{-\{k_1(y)\}^{\beta(y)} t} + \sum_{i=2}^4 P_i(y) e^{-k_i t}, \quad (17)$$

where k_2 , k_3 , and k_4 were global parameters and $\beta(y)$, $k_1(y)$, $P_2(y)$, $P_3(y)$, $P_4(y)$, and $F_0(y)$ were local. Initial guesses were determined using the same method as the local fits. All parameters were constrained to be positive and the following parameters were constrained by $0.8 < k_4^{-1} < 1.2$ ns, $1.5 < k_2^{-1} < 2.5$ ns, $2.8 < k_3^{-1} < 3.5$ ns, and $\beta \leq 1$. The rationale for the constraints is in Sec. III. A perturbation coefficient applied to stabilize the estimates of the numerical derivatives calculated by LM was necessary for a successful fit. In IGOR Pro, the coefficient is implemented via an “epsilon wave” and was set to 0.01. The convergence criteria were the same as the local LM fits.

The globally regularized model applies the second derivative global continuity condition on same local parameters across the nontransformed coordinate. The regularizer for any set of local parameters, $\mathbf{P}^{(Y \times 1)}$, is

$$\Omega_p^2 = \mathbf{P}^T \mathbf{H} \mathbf{P}, \quad (18)$$

where $\mathbf{H} \equiv \mathbf{H}^{(Y \times Y)}$ and $\mathbf{P} \equiv (\text{local}_1, \text{local}_2, \dots, \text{local}_Y)^T$. A regularizer value for any set of parameters can be calculated. For this paper, the regularizer value for local parameters $P_1(y)$, $P_2(y)$, $P_3(y)$, $P_4(y)$, $k_1(y)$, and $\beta(y)$ was calculated and summed into the total regularizer value for the model, Ω_M^2 .

In order to use the same LM minimizer used to fit the three and four exponential fits, Igor's built in curve fitting operation was used to minimize the new cost function in Eq. (5). This was accomplished by expanding the data and fit by

one data point (a total of $T \times Y + 1$ points). By placing the square root of $\gamma \Omega_M^2$ in the last fit point and a zero in the last data point, the fitting algorithm will add the squared difference to the residuals giving the proper cost function.

2. Active-set and interior point gradient

AS and IPG require as input a design matrix (\mathbf{A}_i), scaled data (\mathbf{b}_i), and an initial guess for the solution \mathbf{x}_i . AS, IPG, and locally regularized IPG fit each transient in a data set individually. For TCSPC data simulations in this paper the design matrix \mathbf{A} has elements $\exp(-k_j t_i)$. The design matrix was convoluted column-by-column with an instrument response function typical of microchannel plate photomultipliers. The total number of grid points per local fit was set to $K=120$. The grid was evenly spaced in k^{-1} from 0.1 to 12.0 ns. For globally regularized IPG, these local matrices were organized as previously described in the methods for the global case.

The IPG method requires a scaling term $\lambda \in (0, 1)$. The scaling term is used when IPG attempts to take step into a nonfeasible region of the parameter state space; λ is used to scale the maximum allowable step distance as to get close, but not reach the non-negative boundary. We set $\lambda=0.9$ which is 90% of the largest step. In the current data fits λ was set to 0.9 because it seemed to converge the fastest while also avoiding local minima.

Convergence of the IPG algorithm was allowed a maximum of 10^6 iterations. Every 2000 iterations, χ^2 was evaluated. If the fractional decrease from one evaluation χ^2 to the next was less than 10^{-7} , the fit was considered converged. The high S:N limit data, $\mathcal{I}=10^8$, required 10^7 iterations due to the large condition number of the global design matrix. Initial guesses were set to 10^{-32} .

We chose the linear continuity condition for both local regularization and for global regularization in Eqs. (6) and (8), respectively. Adding the regularizer to the minimization for IPG modifies the condition needed to maintain the parameters in the non-negative region. To maintain the totally non-negative condition, regularized IPG requires $(\mathbf{A}^T \mathbf{A} + \gamma \mathbf{H}) \mathbf{x} > 0$, so γ is constrained by

$$\gamma \leq \frac{\mathbf{H} \mathbf{x}}{\mathbf{A}^T \mathbf{A} \mathbf{x}}. \quad (19)$$

For all S:N, γ was set many orders of magnitude below this maximum threshold.

3. Maximum entropy method fits

Maximum entropy fits were performed using a grid of exponential rates and minimized by the Levenberg–Marquardt method using the same convergence and constraint criteria as described earlier. Each transient was fit locally using the same design matrices, \mathbf{A}_i , as employed for the AS and IPG methods

$$F(t, y_i) = \sum_{i=1}^T \sum_{j=1}^K A_{ij} x_{ij}. \quad (20)$$

Maximum entropy regularizer, $\gamma \mathcal{W}$, was by calculated using Eq. (7) and were implemented using Igor's built in curve fitting operation in similar fashion as the global regularized model fits.

D. Model similarity criteria

Two perspectives can be taken to compare and contrast the fitting methods presented in this paper. The first compares the quality of fit and the second compares the quality of parameters. A model is considered acceptable when the quality-of-fit, χ^2 , is below some statistical threshold. Multiple acceptable models (usually differing in model simplicity) can be differentiated from each other via statistical testing. However, many models will be statistically indistinguishable based solely on quality of fit to the data, even when χ^2 is evaluated globally across all Y transients, frustrating the search for the “best” global model. The ultimate goal is to get the set of parameters that best represents the physics of the system assuming all candidate models give adequate fits to the data. We use the Kullback–Leibler divergence and species population deviation to evaluate the quality of the parameters as compared to the exemplar parameter set.

1. Quality-of-fit

The sum of the weighted, squared residuals, χ^2 , is normally reduced by the expected variance or value of χ^2 based on normal statistics. However, because we have synthesized data, we can directly compare the true $F(t, y)$ (i.e., noiseless data) to the noisy data to determine the actual variance of the data and reduce the fit χ^2 by this value

$$\chi_r^2 = \frac{\chi_{\text{fit}}^2}{\chi_{\text{true}}^2}. \quad (21)$$

χ_r^2 values are calculated for the entire data set. We consider an adequate fit to be $1.1 > \chi_r^2 > 1.0$. χ_r^2 with values above or below this range underfitting or overfitting the data, respectively.

The F test is used to calculate the probability-to-reject the hypothesis that two fits are the same based on χ_r^2 .¹⁷ Values close to 1 indicate the values of χ_r^2 are statistically different. For this paper, we use the F test for two different purposes: To set convergence criteria and to compare the statistical significance of different model fits. The convergence criteria for IPG fits provided a probability-to-reject < 0.01 for all S:N when compared to the AS fits. The regularizer strength, γ , for local and globally regularized IPG fits was chosen such that probability-to-reject < 0.01 when compared to the AS fits. With very little regularization, we will show that even though fits are not statistically different, the underlying parameters distributions are. The F test was used compare traditional global model fitting with regularized global fitting, again choosing the regularization strength such that probability-to-reject < 0.01 . We also used the F test to compare three exponential fit to the four exponential, traditional global, and regularized global fits.

2. Quality-of-parameters

Kullback–Leibler divergence is an information theory approach to quantify the difference between a “true” probability distribution and an arbitrary probability distribution and represented by

$$D_{\text{KL}} = \sum_k \sum_y f(k,y) \ln \left(\frac{f(k,y)}{\tilde{f}(k,y)} \right). \quad (22)$$

D_{KL} is a measure of the relative entropy of the two distributions. Perfect overlap of a test set would result in a value of $D_{\text{KL}}=0$.

A common goal in fitting data is to determine the evolution of the populations of different species. Attribution of the populations from gridded fits such as the AS and IPG methods described earlier is not trivial. Therefore, we define regions of the grid that are attributed to each species. The region for each species is summed for a given value of y to determine the total population of that species. The evolution of species i in a particular data set was defined as $S_i(y)$,

$$S_i(y) = \sum_{k=k_{i,\min}}^{k_{i,\max}} f(k,y), \quad (23)$$

where $k_{i,\min}$ and $k_{i,\max}$ were set based on the features of the exemplar parameter set. Specifically, $k_{4,\min}^{-1}=0.6$ ns, $k_{4,\max}^{-1}=1.4$ ns, $k_{2,\min}^{-1}=1.5$ ns, $k_{2,\max}^{-1}=2.4$ ns, $k_{3,\min}^{-1}=2.5$ ns, $k_{3,\max}^{-1}=3.5$ ns, $k_{1,\min}^{-1}=5.0$ ns, and $k_{1,\max}^{-1}=9.0$ ns. The species parameters are calculated for every S:N ratio, but we will only explicitly compare S_i 's for different fitting methods at $\mathcal{I}=10^4$.

An overall score for reproducing the populations over the evolution dimension y was determined by the mean-squared difference of the populations

$$\text{POP} = \sum_{i=1}^4 (\tilde{S}_i(y) - S_i(y))^2, \quad (24)$$

where $\tilde{S}_i(y)$ is the population vectors calculated from the exemplar model. This metric, along with the Kullback–Leibler divergence, allowed us to score the overall performance of different methods for all S:N ratios in their ability to reproduce the underlying physics of the system.

E. Error estimates

Parameter errors in ill-posed problems are unbounded.¹⁷ This arises because many potentially very different parameter sets can fit the data equally well. However, for a particular fit one can estimate the errors in the parameters and the degree to which they are coupled from diagonal and off diagonal elements of the covariance matrix, respectively. Taking the derivative of the Jacobian, Eq. (11), gives the Hessian matrix. The matrix inverse of the Hessian matrix gives the covariance matrix

$$\mathbf{C} = ((\mathbf{A}^T \mathbf{A} + \gamma \mathbf{H}))^{-1}. \quad (25)$$

The square root of the diagonal elements of the covariance matrix provides an estimate of the errors in the parameters. Not all parameters will have a value of zero in the Jacobian

given by Eq. (11). This is a consequence of the constraint of the parameters to be non-negative. These parameters are not active and should not be included in error estimates. Therefore, only parameters with values of the Jacobian that are close to zero should be included in the Hessian matrix. In our experience the covariance matrix gives large estimates for the errors in these ill-posed problems because of the large degree of anticorrelation between parameters.

III. RESULTS

In this section we examine the results of several approaches to fitting the data generated by the exemplar model at different levels of photon counting noise.

A. Levenberg–Marquardt

A common method for multiexponential fitting of TCSPC data is nonlinear least-squares optimization. This method implies a weighted sum of delta functions at the different values of the decay rates in Laplace space. The number of delta functions is static which helps to stabilize the inversion. Levenberg–Marquardt is one of the most common algorithms for accomplishing this. For the data sets generated by the exemplar model, less than three exponentials did not give satisfactory fits.

1. Three-exponential model fits

Three exponential fitting was fast, robust, and relatively insensitive to initial guesses with respect to convergence, except for the high S:N limit ($\mathcal{I}=10^8$) which converged much more slowly. Only at this high S:N limit did a fit yield a value of χ_r^2 that indicated that the fit failed to reproduce the data (see Fig. 11).

Contour plots of the parameter values resulting from three exponential LM fits appear in Fig. 2. The discrete functions are unable to reproduce the width or evolution of species 1 at any S:N. The conversion of species 2–3 is never resolved. At high S:N ($\mathcal{I} \geq 10^5$) the conversion is represented by a single species with a decay rate at approximately the weighted average of the contributions of species 2 and 3. Species 4 should not evolve in either intensity or position and is only properly reproduced at high S:N ($\mathcal{I} \geq 10^5$).

The population evolutions for species 1–4 at a moderate S:N, $\mathcal{I}=10^4$, are compiled in Fig. 14. For this three exponential fit, $S_1(y)$, represented by brown circles, follows a similar trend to that of the exemplar evolution (shown by the solid line of the same color). $S_2(y)$ (green diamonds) is not present, suggesting the need for another exponential term. $S_3(y)$ (blue triangles) follows the correct increasing population evolution trend over the experiment coordinate, but is overestimated. $S_4(y)$ (red squares) decreases until it reaches the true population, $y=2$, then flattens. The initial overestimation of species 4 population evolution and total overestimation of species 3 is a compensation for the inadequacy of the three exponential model to describe species 2.

Quantitative measure of the similarity [see D_{KL} and population overlap parameter (POP) in Figs. 12 and 13, respectively] of the three exponential fit parameters and the exemplar model show little improvement with increasing

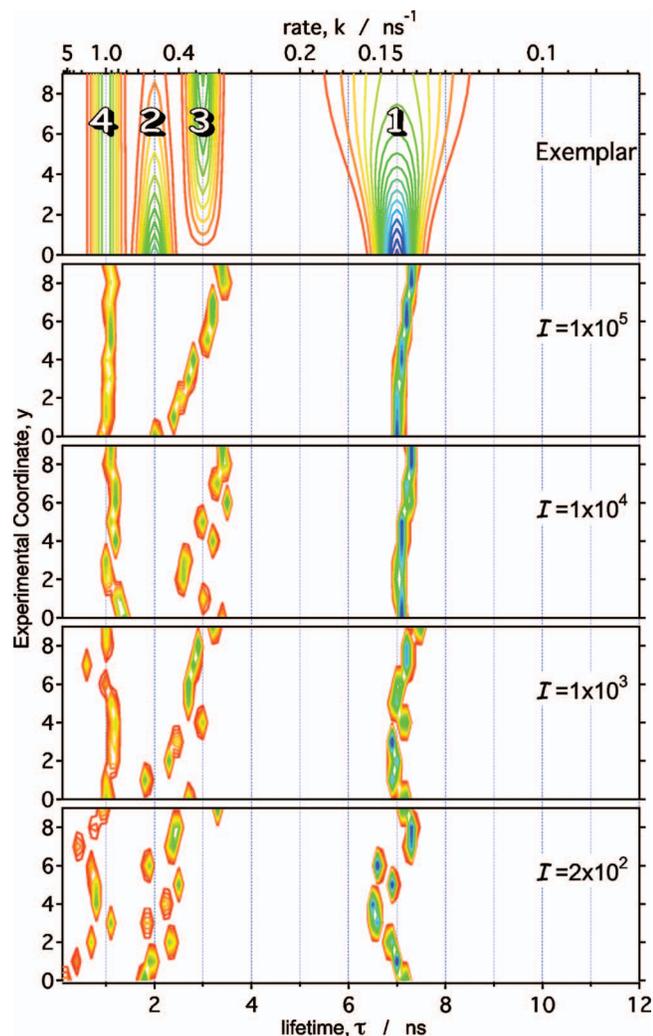


FIG. 2. (Color) $f(k, y)$ for three exponential fits. Starting from the top panel down: Exemplar solution parameters, fit solution parameters synthesized with $\mathcal{I}=10^5$, 10^4 , 10^3 , and 200 peak mean photons. Superimposed numbers correspond to species described in Sec. II.

S:N, even at the high limit ($\mathcal{I}=10^8$). This is a consequence of the lack of flexibility in the model to represent the trends in the exemplar model. In effect, the correct model was excluded prior to fitting. A normal next step would be to add an additional exponential term to the model and compare the fits to the three exponential results.

2. Four exponential model fits

Contour plots of the parameter values resulting from four exponential LM fits appear in Fig. 3. For all S:N, the resulting parameters were sensitive to initial guesses, even though the χ^2 statistic was not. At the S:N limit ($\mathcal{I}=10^8$), convergence became extremely sensitive to the initial guesses for the parameters. We note that this prompted our use of similar initial guesses for each S:N ratio as described in Sec. II. According to the F test, the χ_r^2 values were not significantly different from those obtained for three exponential fits until high S:N (probability-to-reject for $\mathcal{I} \leq 2 \times 10^4$ was < 0.01). Only the high S:N limit ($\mathcal{I}=10^8$) gave a value of χ_r^2 that indicated that the fit failed to reproduce the data (see Fig. 11).

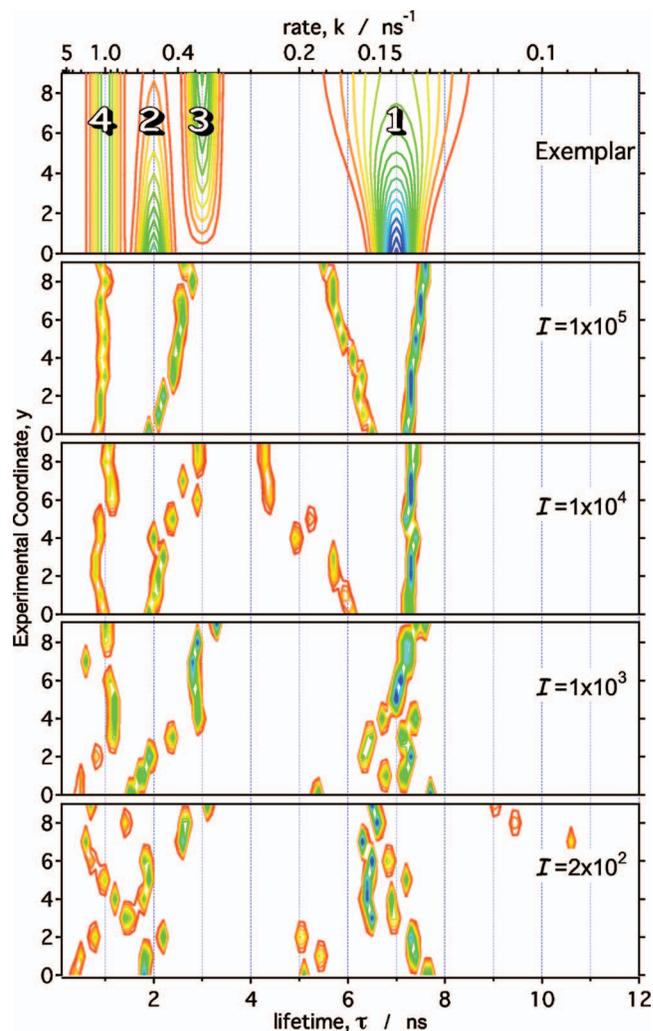


FIG. 3. (Color) $f(k, y)$ for four exponential fits. Panels are described in Fig. 2.

The four exponential model fails to reproduce the exemplar model as illustrated in Fig. 3. In the region of the distribution associated with species 1, the fit typically used two of the four available exponentials. There was no systematic trend in these components until high S:N ($\mathcal{I} \geq 10^5$), where the pair of exponentials are split and increases in separation over the nontransformed coordinate, y . Though this might have been hailed as a success, there is little to distinguish this pattern from that generated in the species 2, 3, and 4 region. Here species 4 is resolved and the exchange of species 2 and 3 is again reduced to a single exponential at the weighted mean.

These trends can be more clearly seen when we consider the species population evolutions in Fig. 14. The $S_1(y)$ decreases and diverges from the exemplar at $y=6$ as the one of the two exponential rates, attempting to shape the edges of this distribution, overshifted to about $k^{-1}=4.5$ ns. There was an abrupt exchange of population between $S_2(y)$ and $S_3(y)$ at $y=6$ as the single exponential component shifts to longer lifetime. Finally, $S_4(y)$ evolved by first decreasing ($y \leq 5$) then, after a sharp increase, decreased again ($y > 6$).

The overall POP was overall better than the three exponential fits and gradually, if inconsistently, trended toward

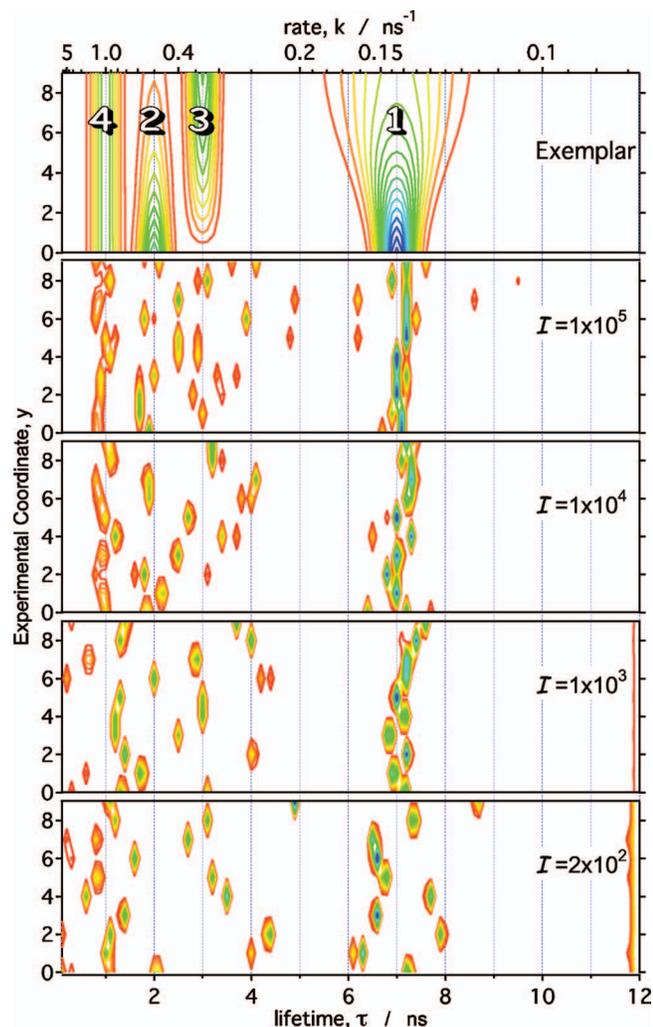


FIG. 4. (Color) $f(k, y)$ for active-set method fits. Panels are described in Fig. 2.

improvement with increasing S:N. The Kullback–Liebler divergence also improved slightly with increasing S:N. Approaching the high S:N limit there was no further improvement in POP or in D_{KL} . Interestingly, χ_r^2 greatly deviated from one but was two orders of magnitude smaller than three exponential χ_r^2 at the same high S:N limit.

At low S:N the fits are of such poor stability that the addition of another exponential makes it even more difficult to identify a trend in the fits, most likely because of the necessary trade-off between representing the different numbers of species and widths at the early versus late stages of the evolution across y . For example, in the first few transients, three species were being represented by four exponentials. Later there were four species active, but one of them required a width that needs at least a fifth exponential to represent it. All this would suggest that an adaptive method (such as the active-set method) that could change the number of exponential contributions as needed might provide better fits and better reconstruction of the exemplar model.

B. Active-set

Contour plots of the parameter values resulting from active-set method fits appear in Fig. 4. The active-set method

converged substantially slower than the three and four exponential LM fits. This difference became more pronounced with increasing S:N. The χ_r^2 steadily decreases with increasing S:N.

At low S:N, the fits are unstable and correct trends are difficult to ascertain. Species 1 is represented by two or three components, but no systematic trend appears. Likewise species 2, 3, and 4 are represented by three or four components, but are not resolved until the high S:N limit ($\mathcal{I}=10^8$ parameters not shown).

As can be seen in Fig. 4, the active-set method typically selects four to eight exponential contributions for a given position along the nontransformed (y) coordinate. However, as can be seen in Fig. 11 this has not translated into a substantially better match with the exemplar; it is not only difficult to ascertain a trend for species 2, 3, or 4 at any S:N, but there is little evidence of distribution of rates for species 1.

The lack of any discernible trend has translated into erratic behavior of the species evolutions in Fig. 14. Again, $S_1(y)$ slowly deviates from the exemplar as the width of the distribution of species 1 increases. Populations $S_2(y)$ and $S_3(y)$ randomly exchange across the experimental coordinate while $S_4(y)$ fluctuates about exemplar population values.

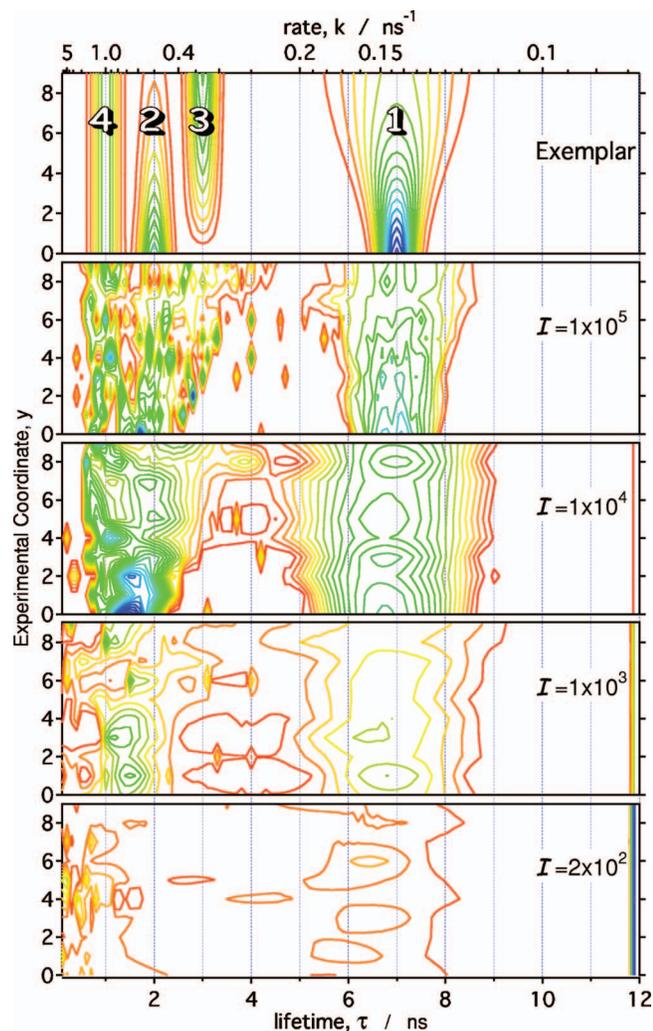
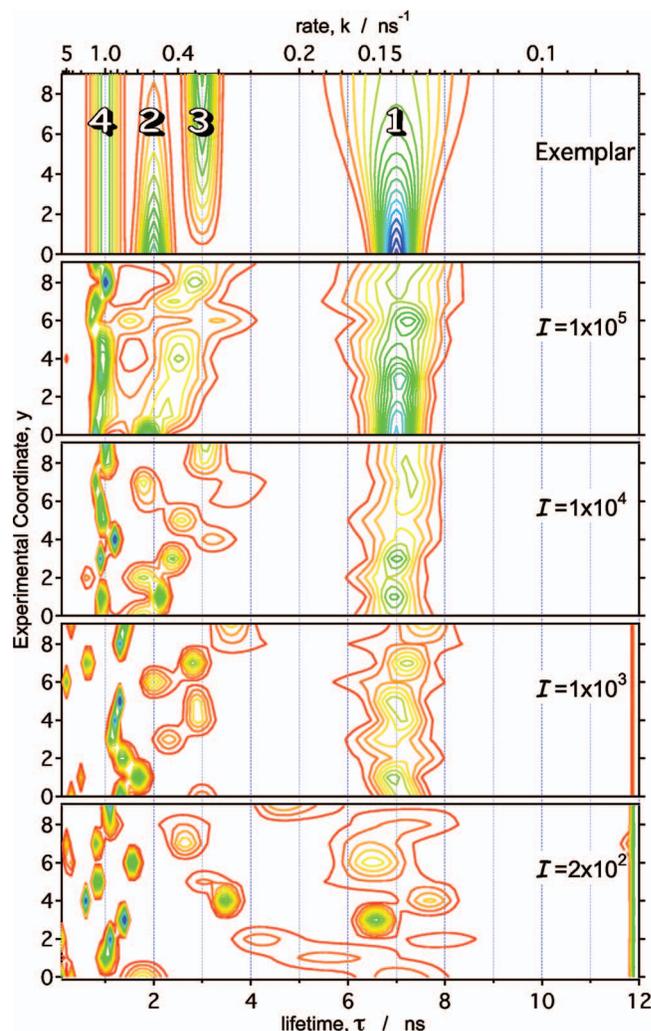
Though performed on a grid, the active-set method tends to give discrete distributions. This is a consequence of the algorithm used to expand and contract the basis set. During the search for the best fit the algorithm explores regions of parameter space that include negative populations. To correct this, these negative values are pruned from the active set of basis functions. It is usually the closely related basis functions that get pruned. This results in the suppression of solutions with continuous distributions. This tends to make the active-set method not fit distributions as well as one might expect given that the distribution is explicitly allowed by virtue of the procedure being performed on a grid.

C. Maximum entropy method

A commonly used method to improve the stability of multiexponential and distribution-of-exponentials fitting is to regularize using maximum entropy (MEM). The parameter values of the resulting fits are shown in Fig. 5. The MEM converged substantially slower than the active-set fits. Good fits were only obtainable when the regularizer parameter was increased to a value consistent with a tenfold higher probability-to-reject the regularized solution as compared to the typical IPG regularized fits (*vide infra*). The χ_r^2 steadily decreases with increasing S:N and displayed a consistent trend with AS fits even at the highest S:N.

At low S:N ($\mathcal{I}=2 \times 10^2$), none of the species is resolvable. At modest S:N ($\mathcal{I}=10^3$) the distribution of species 1 is apparent and is the first example in this paper of fitting method depicting a distribution. Though species 1 distribution becomes better defined as high S:N, species 2, 3, or 4 are still undecipherable. Only in the high S:N limit ($\mathcal{I} \leq 10^8$) are all the species resolvable.

Kullback–Liebler divergence and the POP systematically improves with increasing S:N. At the (unrealistically) high S:N limit, these parameters are extremely good.

FIG. 5. (Color) $f(k, y)$ for MEM fits. Panels are described in Fig. 2.FIG. 6. (Color) $f(k, y)$ for IPG fits. Panels are described in Fig. 2.

D. Interior point gradient method

The interior point gradient method improves the grid-based solution by performing the fit search in such a way that physically forbidden (negative population) values of the parameters are never accessed. Execution time of the algorithm was substantially faster than that of the active-set method, but slower than LM. Adding local regularization to IPG increased the average iteration time by less than 10%. Globally regularized IPG was the slowest to converge because this method fits all the data simultaneously. For all reasonable values of S:N ($\mathcal{I} \leq 10^5$) the resulting χ_r^2 were consistent with a good fit as seen in Fig. 11.

1. Unregularized IPG

Contour plots of the parameter values resulting from unregularized IPG fits appear in Fig. 6. The IPG fits do a better job reproducing distributions, such as that of species 1. Even at modest S:N ($\mathcal{I} \geq 10^3$) the distribution of species 1 is well defined, though the proper evolution of its width is not resolved except for high S:N ($\mathcal{I} \geq 10^5$). The contributions of species 2, 3, or 4 are unstable at low S:N.

At high S:N ($\mathcal{I} = 10^5$) species 4 is resolved from species 2 and 3. However, the conversion of species 2 to 3 is a

shifting distribution rather than separate exchanging distributions. Only at the high S:N limit ($\mathcal{I} = 10^8$ parameters not shown) does the unregularized IPG method reproduce the contributions from all four species and their evolution with y .

Even though population distributions are being reproduced, the species populations for the IPG method evolve erratically over the experiment coordinate. $S_1(y)$ decreases with y , but the deviation from the exemplar, \tilde{S}_1 , is less than seen in AS. $S_2(y)$ and $S_3(y)$ readily exchange population. $S_4(y)$ is far from constant evolution across y .

Kullback–Liebler divergence systematically improves with increasing S:N. The overall population parameter, POP, also improves with increasing S:N. Fitting data with high S:N, $\mathcal{I} \geq 2 \times 10^4$, there is a big improvement in POP. Comparing parameters in Fig. 6 with moderately high to high S:N, i.e., $\mathcal{I} = 10^4$ to $\mathcal{I} = 10^5$, IPG gains the ability to separate species 2 from species 3, albeit, with minimal, success. To improve the IPG's inability to accurately reproduce parameters at low S:N, a regularizer may be used to stabilize the fitting procedure.

2. Locally regularized IPG

The effect of the local regularizer on the fits can be seen in Fig. 7 as effectively broadening the distributions as com-

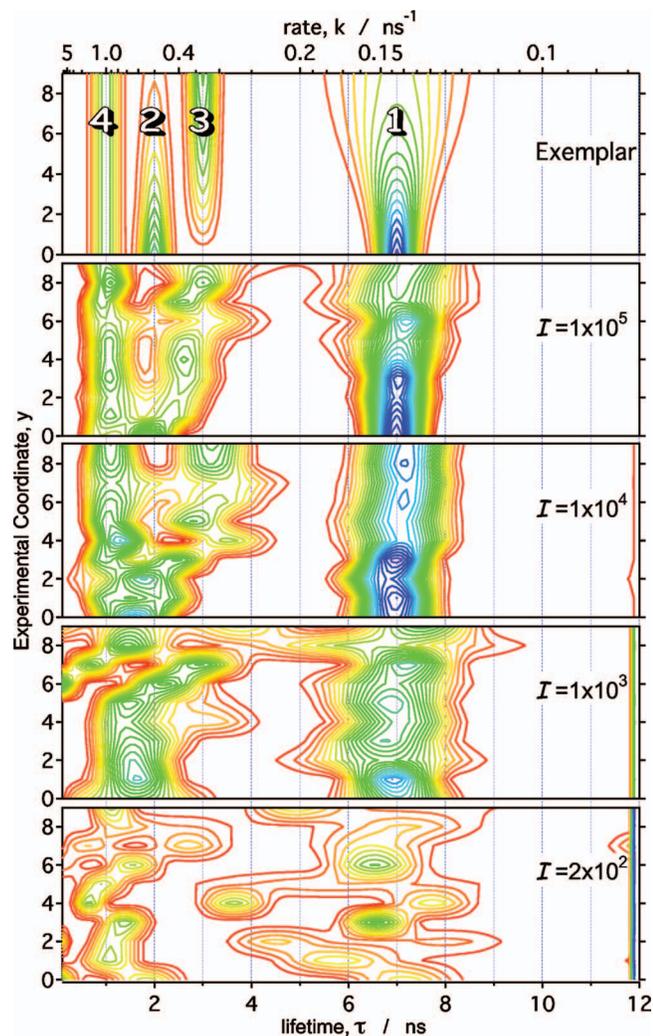


FIG. 7. (Color) $f(k, y)$ for locally regularized IPG fits. Panels are described in Fig. 2.

pared to unregularized IPG. The imposition of the continuity condition in the transformed dimension k has essentially given very similar solutions to that of the unmodified IPG except that the distributions have broadened beyond the exemplar widths. At very low S:N, the correct trend is impossible to ascertain in any reasonable way. Fitting moderate S:N data ($\mathcal{I}=10^4$), the distribution of species 1 is resolved but does not correctly evolve in width. Concurrently, species 2, 3, and 4 are merged into an undecipherable single species. At high S:N ($\mathcal{I}=10^5$), species 1 is resolved with a similar width evolution as the exemplar. However, for $y < 4$ species 2, 3, and 4 are merged making it difficult to establish initial species. Even at this S:N, species 1 has been overestimated in width.

The species populations for locally regularized IPG were far more stable than the AS or unregularized IPG. $S_1(y)$ decreased away from the exemplar population after $y=6$. Even though the distributions of species 2 and 3 were unresolved, the evolution of $S_2(y)$ and $S_3(y)$ indicated that the correct populations were beginning to be recovered, though they were still unstable. There was an apparent trend of decrease

in population of species 2 and a simultaneous increase in species 3. $S_4(y)$ was well behaved about its correct exemplar value.

Though the qualitative trends comparing the locally regularized IPG fit parameters to the exemplar model suggested worse fits, the Kullback–Liebler divergence was smaller than the unregularized case. This is a result of broader distributions being more forgiving in terms of overlap with the exemplar model. Even the population is slightly better reproduced in the locally regularized IPG fits. Similar to the unregularized case there is a big improvement in this metric occurring at low S:N, $\mathcal{I}=2 \times 10^3$ to $\mathcal{I}=10^4$. As seen in Fig. 7, species 1 becomes better defined at higher S:N, as well as separating species 3 from species 1 for $y \geq 7$.

Overall, the regularization of IPG stabilized the solution set and may allow the investigator to establish trends at high S:N. Though the evolution of species populations showed reasonable overlap with the exemplar in Fig. 14, it is unlikely that local regularization would lead the investigator to correctly identify the number and properties of species present.

3. Globally regularized IPG

The globally regularized IPG method was far superior for reproducing the exemplar model. Even at level of S:N appropriate for a single molecule measurement, ($\mathcal{I}=2 \times 10^2$), the fit demonstrates that the long lifetime species was a broad distribution whereas the short lifetime components were narrower. Species 1 broadens with y . The presence and conversion of species 2 and 3 were beginning to be resolved. At modest S:N ($\mathcal{I}=10^3$), species 2 and 3 were clearly two separate populations converting from one to the other, adjacent to a third constant species 4. The mean of the population distribution of species 1 was about 7 ns, increasing in width along the experimental coordinate.

The species population evolution was quite similar to that of the exemplar. Species populations $S_2(y)$ and $S_3(y)$ both slowly exchanged populations without affecting $S_4(y)$. Notice that populations followed a more linear trend than the exemplar populations for species 2 and 3. This was due to the linear nature of the regularization condition applied in the fitting algorithm.

The quality-of-fit improved systematically with increasing S:N. The quality-of-parameters also became increasing better with S:N. POP decreases abruptly between $\mathcal{I}=10^2$ and $\mathcal{I}=2 \times 10^2$, then again between $\mathcal{I}=5 \times 10^3$ and $\mathcal{I}=10^4$. Though the globally regularized IPG method reproduced a distribution for species 1, at the lowest S:N there was an inability to distinguish between species 2, 3, and 4 (data not shown). At a slightly higher S:N, this was no longer a problem. When comparing a moderate to a high S:N, not only were species 2 and 3 better separated, their populations more closely matched the exemplar. This represented the second drop in POP. The Kullback–Liebler divergence also steadily decreased with increasing S:N. Fits for data with low to modest S:N, $\mathcal{I}=5 \times 10^2$ to $\mathcal{I}=10^3$, also showed a drop in the divergence parameter. This small increase in S:N allowed the globally regularized IPG method to clearly define all species and populations.

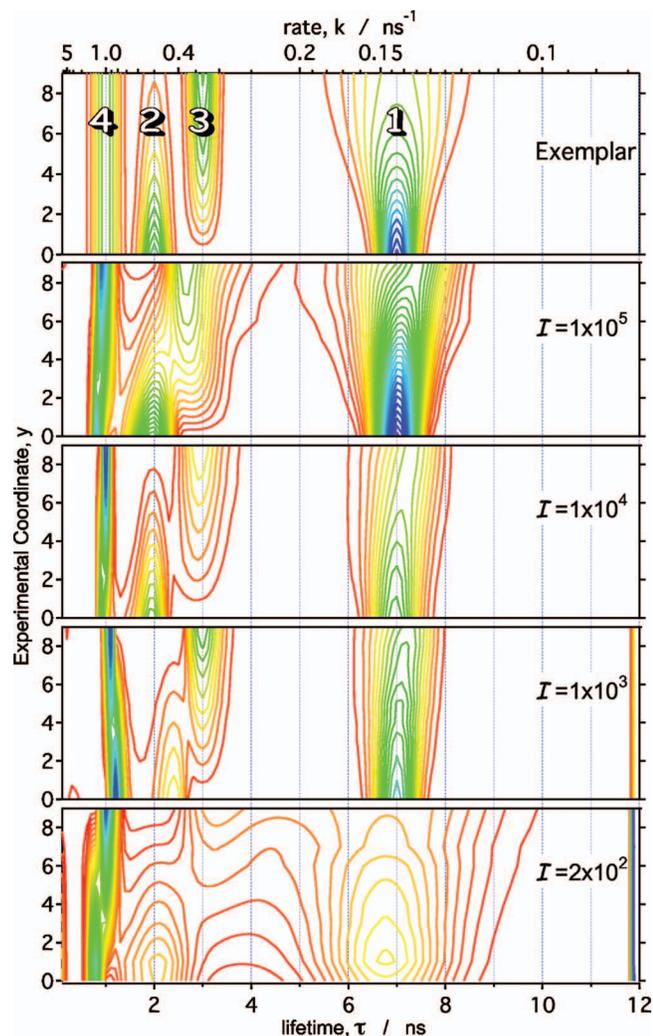


FIG. 8. (Color) $f(k, y)$ for globally regularized IPG fits. Panels are described in Fig. 2.

The globally regularized IPG fitting procedure does a great job reproducing the evolution of the exemplar model. However, the exact evolution of the population of the species present must be interpreted by extracting them from the gridded fits. Moreover, there is no way of directly interpreting the physics or chemistry implied by the evolution from these fits. This is the consequence of the probabilistic constraints as opposed to specifying a particular physical model. The advantage is that since models were not eliminated *a priori*, one can use the globally regularized IPG fits to determine an appropriate physical model for traditional global fitting as appears in the next subsection.

E. Global model LM

A global model puts deterministic constraints on the model space in the transformed coordinate. We determined a global fitting model using the distribution of rates from the globally regularized IPG fit at $\mathcal{I}=10^4$ in Fig. 8. From this fit, it is clear that there are four species. Species 1 had a significant width, that slightly evolved and was centered at $k^{-1}=7$ ns. Conventional global fitting might treat this as a stretched exponential expression [second term in Eq. (17)] in

order to model a distribution of rates. Both the stretch parameter, $\beta(y)$, and the rate parameter, $k_1(y)$, must be local parameters since $\beta(y)$ changes not only the width of the distribution but also the mean.

The global fit also showed that species 2, 3, and 4 were fairly narrow distributions thus represented by single exponentials, each with global rates, k_2 , k_3 and k_4 . Though the populations of these species had fairly consistent trends (i.e., species 2 descended, 3 ascended, and 4 remained nearly a constant in amplitude) the model was kept general enough to allow variability in populations across the evolution, by employing local pre-exponential factors $P_2(y)$, $P_3(y)$, $P_4(y)$, respectively. The fit model also contained a base line term, $F_0(y)$, local for each data transient.

Part of the instability of nonlinear least-squares fitting of multiexponential functions arises from the lack of constraints on the exponential parameters. Even with the proper discrete global model, the fit procedure must be constrained. If they are left unconstrained, the exponential parameters can attain the same values. When this occurs, singular matrix errors are encountered by the LM algorithm causing the fit to fail. To forbid the global rate parameters from combining, a problem common with nonlinear multiexponential fitting, constraints were used. Constraints were placed upon global parameters, k_2 , k_3 , and k_4 , and they were chosen based on the trends in the $\mathcal{I}=10^4$ globally regularized fits.

1. Traditional global model LM

For all reasonable S:N levels ($\mathcal{I} \leq 10^5$), good fits were obtained as determined from χ_r^2 . In fact, there was no significant statistical difference in quality-of-fit across nearly all S:N levels $\mathcal{I} \leq 5 \times 10^4$, comparing the three exponential model and the stretched model (probability-to-reject ≤ 0.01). At a high S:N limit, $\mathcal{I}=10^8$, the χ_r^2 value increased suggesting that the discretization of the model was beginning to make a difference in the quality of the fit at that limit. For all S:N, fits were robust but at low S:N, $\mathcal{I} \leq 2 \times 10^2$, the global rates were often limited by the constraint (not all data are shown).

For the low S:N fits, the stretched exponential term was not able to consistently reproduce the species 1 distribution. The $\beta(y)$ parameter converged to 1 in several cases resulting in a discrete exponential for some values of y . Even though species 1, 2, and 3 were centered at nearly the correct lifetime, the local amplitudes fluctuate greatly at low S:N. By a moderate S:N level, $\mathcal{I}=10^4$, the fits nicely reproduce the width and evolution of the species 1 distribution. The instability of the β parameter for the fit to species 1 appeared to interfere with the fitting of the conversion of species 2 to 3, which was not correctly reproduced until $\mathcal{I}=10^4$. The species 4 region was properly reproduced even at low S:N, $\mathcal{I}=2 \times 10^2$.

As anticipated, the ability of the global model to reproduce the populations was excellent. The populations trends appeared to be only slightly “noisy” across y . Moreover, the POP metric shows global model thus far producing the best overlap with each species population from $\mathcal{I} \geq 5 \times 10^2$. Because all population for a particular species were determined from subranges of the grid of k 's, the POP parameter was

less sensitive to the discrete nature of model used to describe species 2, 3, and 4. The plateau in D_{KL} divergence at $\mathcal{I} \sim 10^4$ suggests that above this value the deterministic constraints on the transformed coordinate dimensions of the model space limit the quality of overlap with the exemplar model.

The local nature of the amplitude parameters allowed them to fluctuate greatly along y , especially for a low to modest S:N. A specific deterministic model along y related to some physical theory could help stabilize this situation. We note that since there was no improvement in Kullback–Liebler divergence or POP for the high S:N limit ($\mathcal{I}=10^8$), we expect little gain in parameter stability at a moderately high S:N ($\mathcal{I}>10^4$) with a regularized model. Nonetheless, in the spirit of first using probabilistic models, we attempted to regularize the model fit along the y coordinate as a way to use global information without a specific model.

2. Regularized global model LM

Regularization of the global LM fitting increased convergence time by no more than a factor of 2 and increased the stability of the fitting procedure. The values of χ_r^2 indicated good fits across all reasonable values S:N and were not statistically different from the traditional global model fits as measured by the F test (probability-to-reject <0.05).

The addition of the regularizer stabilizes the distribution due to species 1. This, in turn, stabilized the evolution of species 2 and 3 since species 1 was no longer interfering with them. Overall good fits were obtained at moderate S:N, $\mathcal{I} \geq 10^3$.

The globally regularized model fit was a considerably better method for reproducing the distribution of species 1 at very low S:N. Species 1 had a clear distribution across the experimental coordinate, but lacked an evolving trend in width. Species 2, 3, and 4 had the correct population trends but are close to the employed bounds. For data with $\mathcal{I} \geq 10^3$, all species were well within the imposed bounds, while showing an evolving width associated with species 1. The ability of this method to resolve species 2, 3, and 4 did not change much with increasing S:N $\mathcal{I} \geq 10^3$, while species 1 improved with better S:N.

The addition of the regularizer to an otherwise traditional global fit improved the ability of the fit to reproduce systematic changes in species population evolution as seen in Fig. 14. Specifically, the evolutions for the regularized global model were noticeably smoother. As with the globally regularized IPG fits, $S_2(y)$ and $S_3(y)$ were more linear than their respective exemplar evolutions. The constant species, $S_4(y)$, also did not have a consistent increasing trend, as it did for the globally regularized IPG. Though $S_1(y)$ was slightly overestimated, the population evolution was nearly constant.

The match with the exemplar model was limited because of the discrete character of the model in the transformed dimension as seen in the Kullback–Liebler divergence. However, when comparing the Kullback–Liebler divergence for the regularized global model fits to the traditional global model, the regularizer model reached a plateau sooner ($\mathcal{I} = 10^3$ versus $\mathcal{I} = 5 \times 10^3$). As seen in Figs. 9 and 10, the regularized global model stabilized the evolving distribution of

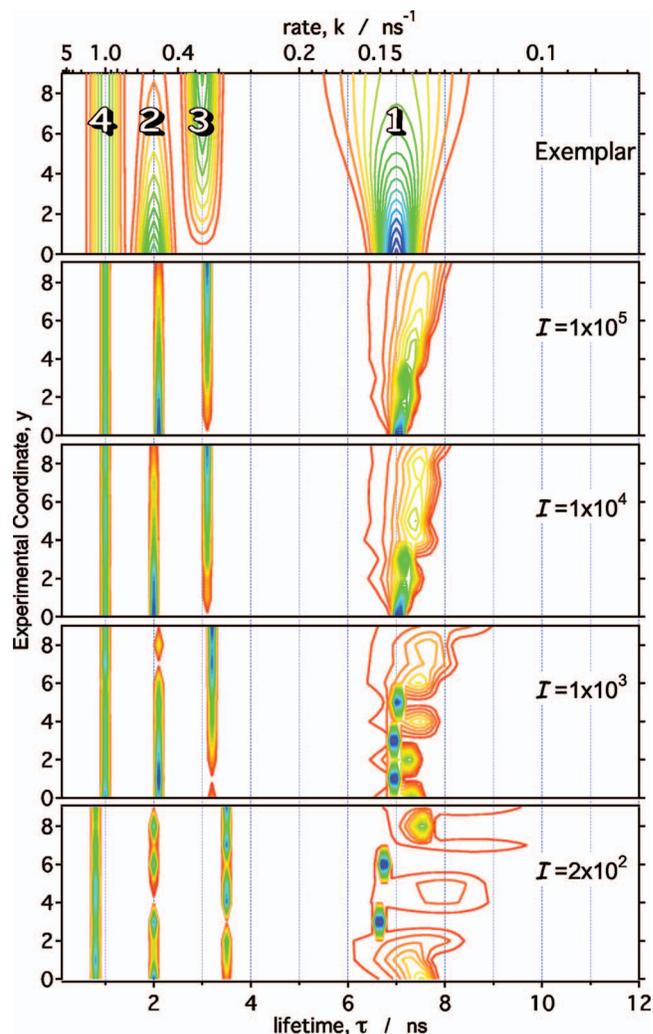


FIG. 9. (Color) $f(k, y)$ for global model fits. Panels are described in Fig. 2.

species 1 with modest S:N, $\mathcal{I} = 10^3$. There was no improvement in Kullback–Liebler divergence in the high S:N limit ($\mathcal{I} = 10^8$).

The POP metric steadily decreased until a plateau was reached for a moderate S:N, $\mathcal{I} = 2 \times 10^3$. This indicated that from a species population point-of-view, nothing else can be gained from better S:N. This can be seen in the increase in the POP value for fits to the highest S:N data ($\mathcal{I} = 10^8$), thereby showing the addition of a regularizer is not necessary for a discrete global model fit to virtually infinite S:N data.

IV. DISCUSSION

A. Comparison of methods

In the results we evaluated how the S:N ratio influenced the success of each fitting method using χ_r^2 , the Kullback–Liebler divergence, and the population parameters. We now compare the different methods to each other and discuss why they gave different results.

1. Reduced chi-square

The four deterministically constrained models (all optimized with LM) showed little systematic change in χ_r^2 until

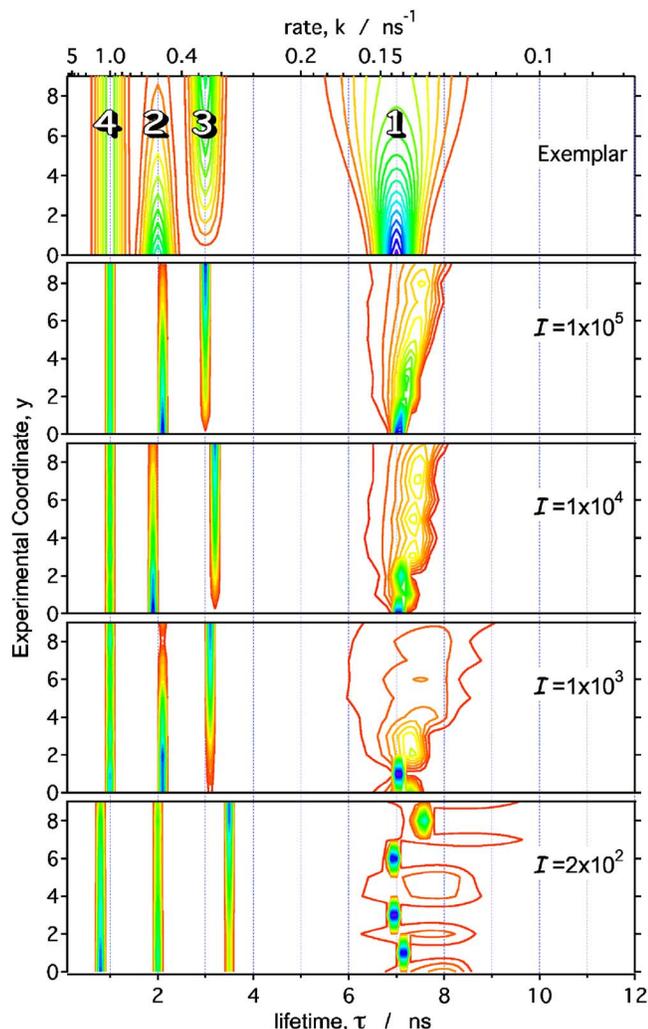


FIG. 10. (Color online) $f(k, y)$ for regularized global model fits. Panels are described in Fig. 2.

$\mathcal{I} \geq 5 \times 10^4$, at which point χ_r^2 began to get worse. (See Fig. 11). By comparison, the gridded methods (optimized by AS, MEM, or IPG) showed steady improvement in χ_r^2 until $\mathcal{I} = 10^5$. There was no statistical difference of fit between the AS fit and any of the MEM or IPG fits except in the high S:N limit ($\mathcal{I} = 10^8$). The MEM fits gave slightly higher χ_r^2 because of the higher value of the regularizer strength parameter required to get reasonable fits. One interpretation of the higher χ_r^2 for the gridded methods would be that they were done on too coarse of a grid. However, the improvement in χ_r^2 with increasing S:N suggests otherwise, since a constraint in the model space would be reflected as a lower boundary for the χ_r^2 . Furthermore, improving the grid resolution did not change χ_r^2 . Grid methods have more model flexibility and can take advantage of the increased S:N, however, the model space they are performed in does not include terms that adapt to the noise.

The global model fits were not statistically different than the three exponential fits until they began to diverge in terms of χ_r^2 at $\mathcal{I} > 10^4$ because they have less model flexibility than three exponential and four exponential fits. According to χ_r^2 the three exponential model was adequate to describe the

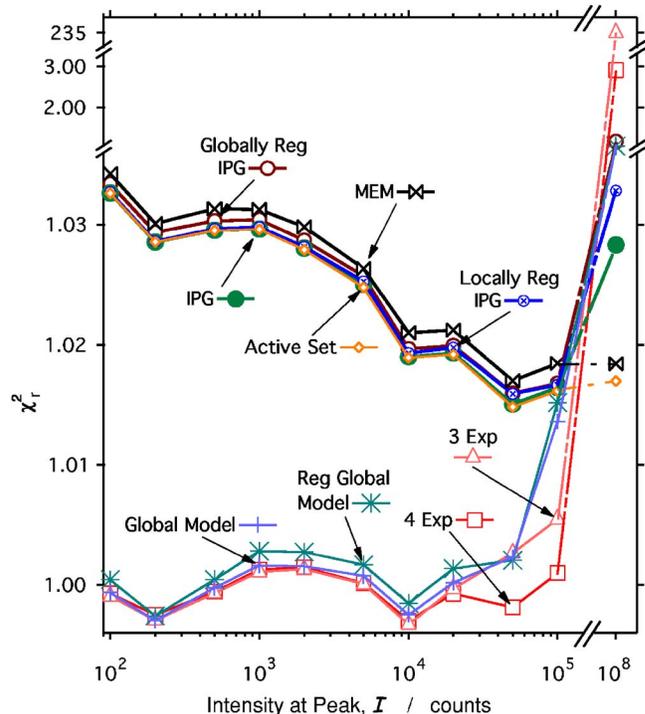


FIG. 11. (Color online) χ_r^2 for all \mathcal{I} using different fitting methods. (Δ) Three exponential fits, (\square) four exponential fits, (\diamond) active-set method fits, (\times) maximum entropy method fits, (\bullet) IPG fits, (\otimes) locally regularized IPG fits, (\circ) globally regularized IPG fits, ($+$) global model fits, and ($*$) regularized global model fits.

data for these fits until high S:N. In the high S:N limit it was quite clear that neither the three nor the four exponential fit was a good τ model for the system.

2. Kullback–Liebler divergence

The Kullback–Liebler divergences comparing the exemplar parameters to the three exponential, four exponential, AS, global model, and regularized global model fits did not improve much with increasing signal-to-noise (see Fig. 12). The deterministic constraints implied by these fitting methods forbid the true solution.

At low S:N ($\mathcal{I} = 10^2$) the methods were all clustered. In this limit the AS method was worst and the three exponential model was best. This suggests that too much model flexibility is not good when there is little information in the data, as the model flexibility will be used to fit noise. The MEM and IPG methods did a better job than AS because they tended to give continuous distributions that better represented the broad part of species 1 and were less able to adapt to fit details of the noise.

With a modest increase in S:N ($\mathcal{I} = 10^3$) the IPG and global models all showed a lower divergence from the exemplar model than did the three and four exponential, MEM, and AS method. This was primarily a consequence of their ability to represent the width of species 1. The MEM fits continued to improve, however, not as quickly as the IPG and global-model methods. The Kullback–Liebler divergence for the three and four exponential and AS methods reached a plateau here; they could not improve because the model was fixed, and they could not take advantage of the

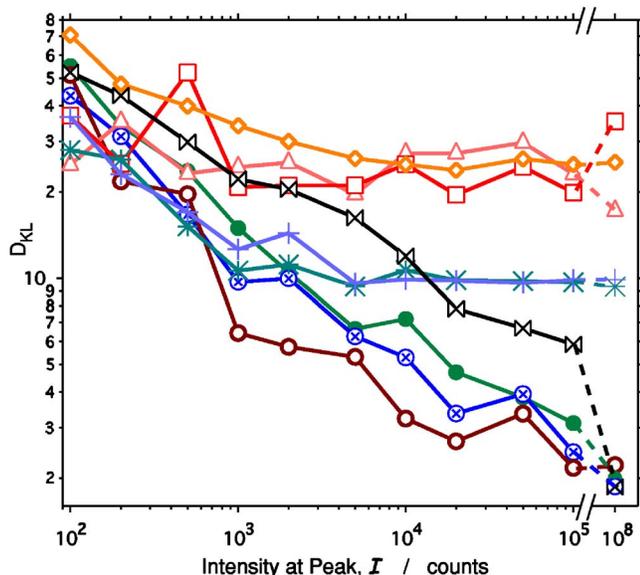


FIG. 12. (Color online) Kullback–Liebler divergence for all I using different fitting methods. Fitting methods are represented in the same markers as in Fig. 11.

increased information available at higher S:N. Multiexponential fitting optimizes the exponential parameters after assuming the size of the solution space is drastically limited. Multiexponential fitting never considers solutions of the form that IPG uses to find a best fit. In the case of the AS method, the fitting algorithm is such that it chooses multiple-discrete representations of the distribution of species 1. AS disfavors smooth distributions since they include more basis functions and usually give negative values in the matrix factorization step of the algorithm. The fundamental problem with the AS method is that in its search it allows intermediate solutions that have nonphysical negative amplitudes and deletes basis functions to prevent the negative amplitudes from being realized, but in the process also eliminates possible correct solutions. This tends to split smooth distributions into multiple discrete components. As a result AS chooses what could be considered a minimal entropy solution.

By $\mathcal{I} \geq 2 \times 10^3$ all the IPG methods are better than the global model fits. This is because the global model fits are representing the narrow width of each of the species 2, 3, and 4 by discrete exponentials. This restriction on the model space resulted in a limitation of how well it could reproduce the exemplar model. Rather than adjusting the fit model to include a width, as in the IPG methods, the differences between a discrete versus a narrow distribution were compensated by minor adjustments of other parameters and do not greatly affect χ_r^2 .

Across the range of $10^3 \leq \mathcal{I} \leq 10^5$ the globally regularized IPG fitting was superior to the other methods. This is because the regularizer favors those solutions that are continuous in y , resulting in better overlap with the exemplar parameters. IPG evolves the fit distribution within the full solution space until the distribution reaches a best fit.

3. Population evolution

The POP parameter in Fig. 13 measures the overall ability of the fit to extract the evolution of the populations, while

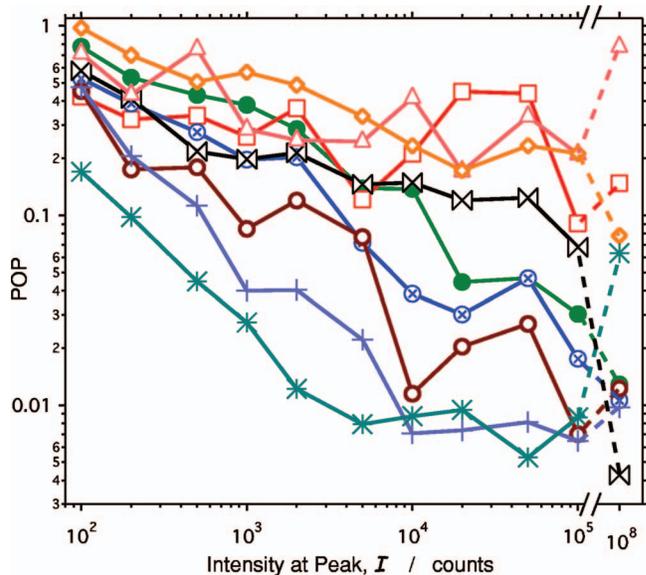


FIG. 13. (Color) POP for all \mathcal{I} using different fitting methods. Fitting methods are represented in the same markers as in Fig. 11.

the individual populations of the different species appear in Fig. 14. The globally regularized IPG began improving at a lower signal level than the other IPG methods, however, at higher signal levels the other methods eventually caught up. The MEM fits were comparable to the locally regularized fits for low S:N but after $\mathcal{I} = 2 \times 10^3$ the MEM was worse and did not catch up until the higher accuracy of our MEM implementation became important in the high S:N limit ($\mathcal{I} = 10^8$). The global regularizer could not do a perfect job tracking the population evolution because it inherently measures deviation from a linear dependence of the evolution in y . As a result the evolution was “flatter” than it should have been as shown in Fig. 14 right column, third row. Once it is known that there are a certain number of species and that they are evolving, the regularizer could be modified to allow more curvature in the evolution by measuring departure from piecewise quadratic behavior with a third derivative regularizer. However, at the level of reconstruction afforded by the second-order global regularizer, one would be better served forming a phenomenological model describing the evolution and fitting to that global model to extract physical parameters from the data.⁷ Prior knowledge in this case arises from the formation of a physical hypothesis and allows the global models to get a better solution. The traditional global model and regularized global model illustrate this point as they both did an excellent job reconstructing the populations and their evolution.

B. Prior knowledge and probabilistic constraints

For all reasonable S:N levels, an investigator considering the value of χ_r^2 and the F test would be led to choose the three exponential model. If one also considered the details of the experiment and used that prior knowledge in the fitting procedure, one would obtain better results in terms of identification of the species present, their properties in terms of the transformed coordinate, k , as well as evolution of their populations across the nontransformed coordinate, y . Stabi-

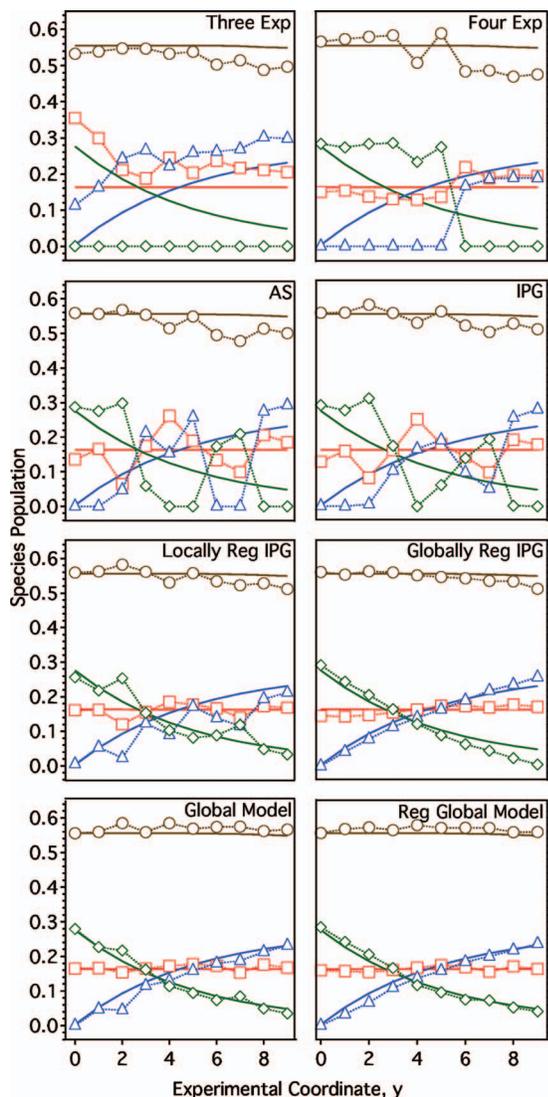


FIG. 14. (Color) Species populations, S_i , at $T=10^4$. Species 1–4 are represented by brown \circ , green \diamond , blue \triangle , and red \square , respectively. The solid lines of the same respective color, represent the exemplar species populations, \tilde{S}_i .

lization by global fitting works because the model space is being limited. To start with a deterministic global model would be to start with an assumed answer and eliminate all other regions of model space.

We saw how the addition of a regularization condition in the transformed dimension, k , perturbs not only the narrow distributions of species 2, 3, and 4 but also increases the width of the broad distribution of species 1. The broadening from the regularizer masked the evolution of the species 1 width. We see that the regularized IPG gave fits that were no better and were in some ways worse than those given by unregularized IPG. This is because the transformed-coordinate regularizer is adding incorrect information to the fitting procedure.

In the case of the three exponential, four exponential, and AS fitting there was also incorrect information being added to the fitting procedure. In the case of the three and four exponential fits, the incorrect information was that the distributions are discrete. This limits the solution space avail-

able for the fits to the point where no correct solution can be found because the correct solution is not included in the space of possible solutions. The space of possible solutions for the AS method fits includes the correct solution, however, the search space includes solutions that are not in the space of possible solutions. The algorithmic mechanism used to return the search from the forbidden space to the allowed space is what biases the fit away from the correct solution. In locally regularized IPG the local solutions are biased away from solutions that include discrete or sharp features and thus, though the available solution space includes the correct solution, the regularizer puts an *a priori* bias against it because it includes narrow features. When viewed in this light, it is understandable why regularization in the transformed dimension is not a physically reasonable thing to do.

The global regularization worked well because there was an evolution of the system across y . If there were no systematic trend with the independent variable then it would not be sensible to regularize the data with the condition of piecewise linearity implied by Eq. (8). This method of global fitting may also be applied to systems without a second independent variable by using replicate measurements and a regularizer that is appropriate for a nonevolving system

$$\mathcal{G}(y) = \int \left[\frac{\partial f(k, y)}{\partial y} \right]^2 dy. \quad (26)$$

This regularizer measures departure of $f(k, x)$ from a constant in the y direction.

In systems that are strongly evolving the third derivative regularizer can be useful as it measures piecewise deviation from a parabola

$$\mathcal{G}(y) = \int \left[\frac{\partial^3 f(k, y)}{\partial y^3} \right]^2 dy. \quad (27)$$

Other regularizers could be used as well, so long as they can be placed into matrix form as in Eq. (12).

Though the different regularizers imply different prior knowledge of the experimental design, they do not forbid nonconforming solutions. They only favor conforming solutions to the degree dictated by the strength of the regularizer parameter, γ . Since γ is restricted to be weaker than any level that causes an increase in χ_r^2 that is significant as determined by the F test, one can be assured that the fit obtained is statistically indistinguishable from that of an unregularized fit. The prior knowledge of the experiment is only biasing within the set of solutions that are statistically the same.

C. Global fitting strategy

The goal of global analysis is to identify the species present and characterize their evolution with the nontransformed coordinate y . An additional benefit of global analysis is that a given local solution, when fit globally, is actually using information from all the decay curves and therefore behaves as if it has a better S:N. The global analysis essentially allows the amount of signal (information) for the entire data set to be used for the global parameters. For local parameters the benefit is indirect, however, having more stable

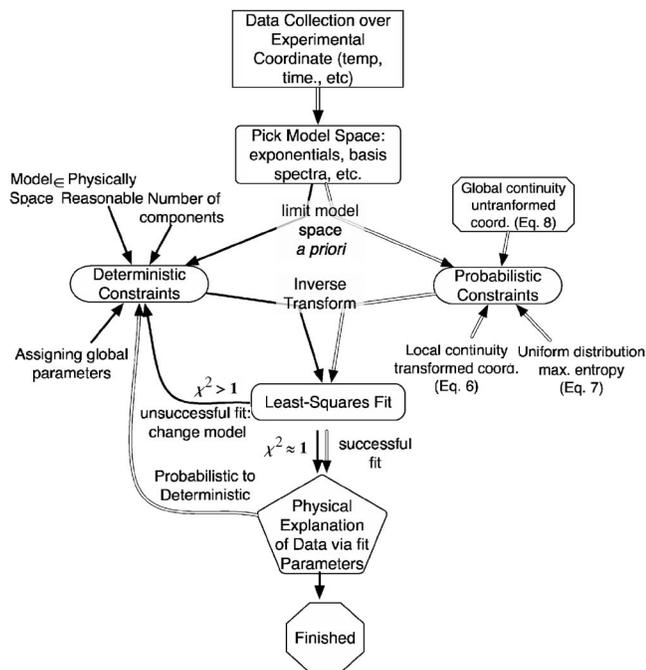


FIG. 15. Model selection vs model discovery. Current paradigm of global analysis follows the thin solid lines. In this case, a model is selected first with deterministic constraints *a priori*, the quality-of-fit is evaluated and a new model is chosen if necessary. Following the double lines, we propose using globally regularized method to discover the model of the evolution of the system based on the data and prior knowledge. Once a model is discovered, deterministic constraints can be applied in order to reach conclusions about the underlying physical system.

values for the global parameters results in more stability in the local parameters as well. Global regularization allows this to occur without requiring a specific model and without enforcing a value on a global parameter. Even when a traditional global model is used a regularizer can allow partial globalization of the nonglobal parameters. The nonglobal parameters are essentially free of the model. In the present example the global model would have needed to be expanded to include a specific functional form for the evolution of the population parameters.

Figure 15 illustrates a fitting strategy that stepwise reduces the available model space according to the data and prior knowledge. This allows fitting of the global data set without sacrificing the range of the model space while still allowing quantitative conclusions to be drawn. Initially one should use prior knowledge to determine the range of valid k for the inversion. Alternatively this could be done with a traditional multiexponential LM fit or a large-grid AS fit. Once the model space has been identified, one should use prior knowledge of the experimental design to determine if there should be some expectation of continuity or other regularization condition in either or both of the coordinates k and y . Once a probabilistically constrained model space is available, one can perform a global fit that includes all the model space using the globally regularized IPG method. From the globally regularized IPG fits one should identify the species present and form a deterministically constrained global model for the transformed (species) coordinate. A traditional global model can then fit the data either with or without a

probabilistic constraint on the nontransformed (evolution) coordinate. At this point a global model for both coordinates should be apparent and further reduction based on a fully deterministic physical model with a small number of parameters should be feasible.

V. CONCLUSIONS

Current physical measurements include complex systems that do not give simple homogeneous signals. Large heterogeneous data sets measured over many conditions have become the norm. Extracting physical information from such data sets requires handling large model spaces in the fitting procedure. Fitting with large numbers of basis functions such as is done in this work has only recently become feasible. The complexity of the global regularization method would have been prohibitive given the computational facilities commonly available 10 years ago. Many of the assumptions for approaches to and algorithms for data fitting and reduction were shaped by technological limitations of 20 or more years ago. Modern computers allow direct fitting of large data sets and should be exploited to allow better insight into the processes behind the data.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Health Grant No. R01GM071684. J.T.G. was supported by a Graduate Assistantship in Areas of National Need grant to the Department of Chemistry and Chemical Biology.

- ¹B. Alberts, *Cell* **92**, 291 (1998).
- ²J. Kaylor, N. Bodner, S. Edridge, G. Yamin, D.-P. Hong, and A. L. Fink, *J. Mol. Biol.* **353**, 357 (2005).
- ³J. H. Werner, R. Joggerst, R. B. Dyer, and P. M. Goodwin, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11130 (2006).
- ⁴H. Frauenfelder, S. Sligar, and P. Wolynes, *Science* **254**, 1598 (1991).
- ⁵B. Ma and R. Nussinov, *Biophys. J.* **90**, 3365 (2006).
- ⁶A. Ahmad, V. N. Uversky, D. Hong, and A. L. Fink, *J. Biol. Chem.* **280**, 42669 (2005).
- ⁷T. C. Messina and D. S. Talaga, *Biophys. J.* **93**, 579 (2007).
- ⁸J. Pronchik and D. S. Talaga (unpublished).
- ⁹A. A. Deniz, S. Mukhopadhyay, and E. A. Lemke, *J. R. Soc., Interface* **1**, 15 (2008).
- ¹⁰F. Ritort, *J. Phys.: Condens. Matter* **18**, R531 (2006).
- ¹¹H. P. Lu, L. Xun, and X. S. Xie, *Science* **282**, 1877 (1998).
- ¹²D. S. Talaga, W. L. Lau, H. Roder, J. Tang, Y. Jia, W. F. DeGrado, and R. M. Hochstrasser, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13021 (2000).
- ¹³W. Miller and K. A. Dill, *Protein Sci.* **6**, 2166 (1997).
- ¹⁴L. D'Alfonso, M. Collini, and G. Baldini, *Biochemistry* **41**, 326 (2002).
- ¹⁵J. T. Giurleo, X. He, and D. S. Talaga (unpublished).
- ¹⁶E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 2003).
- ¹⁷S. W. Provencher, *Comput. Phys. Commun.* **27**, 213 (1982).
- ¹⁸*Dynamic Light Scattering, The Method and Some Applications*, edited by W. Brown (Oxford Science, New York, 1993).
- ¹⁹B. J. Berne and R. Pecora, *Dynamic Light Scattering* (Wiley, New York, 1976).
- ²⁰A. K. Livesey, P. Licinio, and M. Delaye, *J. Chem. Phys.* **84**, 5102 (1986).
- ²¹S. W. Provencher and J. Glöckner, *Biochemistry* **20**, 33 (1981).
- ²²S. Sibisi, *Nature (London)* **301**, 134 (1983).
- ²³A. K. Livesey and J. C. Brochon, *Biophys. J.* **52**, 693 (1987).
- ²⁴E. Post, *Trans. Am. Math. Soc.* **32**, 723 (1930).
- ²⁵A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM: Society for Industrial and Applied Mathematics, Philadelphia, 2004).
- ²⁶W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Reci-*

- pes in C*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
- ²⁷ J. G. McWhirter and E. R. Pike, *J. Phys. A* **11**, 1729 (1978).
- ²⁸ K. Levenberg, *Q. Appl. Math.* **2**, 164 (1944).
- ²⁹ D. Marquardt, *SIAM J. Appl. Math.* **11**, 431 (1963).
- ³⁰ A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problem* (Wiley, New York, 1977).
- ³¹ C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
- ³² M. Merritt and Y. Zhang, *J. Optim. Theory Appl.* **126**, 191 (2005).
- ³³ M. H. Van Benthem and M. R. Keenan, *J. Chemom.* **18**, 441 (2004).
- ³⁴ J. B. Rosen, *SIAM J. Appl. Math.* **8**, 181 (1960).
- ³⁵ A. A. Goldstein, *Bull. Am. Math. Soc.* **70**, 709 (1964).
- ³⁶ A. Siemiarczuk, B. D. Wagner, and W. R. Ware, *J. Phys. Chem.* **94**, 1661 (1990).
- ³⁷ D. R. James and W. R. Ware, *Chem. Phys. Lett.* **120**, 455 (1985).
- ³⁸ M. Vincent, J. Gallay, and A. P. Demchenko, *J. Phys. Chem.* **99**, 14931 (1995).
- ³⁹ G. Wang, Y. Gao, and M. L. Geng, *Biochim. Biophys. Acta* **1760**, 1125 (2006).
- ⁴⁰ D. Jones, *Elementary Information Theory* (Oxford University Press, Oxford, 1979).
- ⁴¹ J. R. Alcala, E. Gratton, and F. G. Prendergast, *Biophys. J.* **51**, 597 (1987).
- ⁴² J. R. Alcala, E. Gratton, and F. G. Prendergast, *Biophys. J.* **51**, 925 (1987).
- ⁴³ G. Hungerford and D. J. S. Birch, *Meas. Sci. Technol.* **7**, 121 (1996).