Finite state projection based bounds to compare chemical master equation models using single-cell data

Zachary Fox,¹ Gregor Neuert,^{2,3,4} and Brian Munsky^{1,5}

¹*School of Biomedical Engineering, Colorado State University, Fort Collins, Colorado 80523, USA*

²*Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA*

³*Department of Pharmacology, School of Medicine, Vanderbilt University, Nashville, Tennessee 37232, USA*

⁴*Department of Biomedical Engineering, Vanderbilt University School of Engineering, Nashville, Tennessee 37232, USA*

⁵*Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, Colorado 80523, USA*

(Received 17 May 2016; accepted 25 July 2016; published online 15 August 2016)

Emerging techniques now allow for precise quantification of distributions of biological molecules in single cells. These rapidly advancing experimental methods have created a need for more rigorous and efficient modeling tools. Here, we derive new bounds on the likelihood that observations of single-cell, single-molecule responses come from a discrete stochastic model, posed in the form of the chemical master equation. These strict upper and lower bounds are based on a finite state projection approach, and they converge monotonically to the exact likelihood value. These bounds allow one to discriminate rigorously between models and with a minimum level of computational effort. In practice, these bounds can be incorporated into stochastic model identification and parameter inference routines, which improve the accuracy and efficiency of endeavors to analyze and predict single-cell behavior. We demonstrate the applicability of our approach using simulated data for three example models as well as for experimental measurements of a time-varying stochastic transcriptional response in yeast. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4960505>]

I. INTRODUCTION

Many physical, chemical, and biological processes are characterized by discrete particles that randomly fluctuate in space, time, or number. These microscopic fluctuations often provide the key to understand and modify mechanisms that control macroscopic phenomena. As one example, stochastic fluctuations in discrete numbers of specific genes, RNA, or proteins across genetically identical populations of cells play an increasingly important role in the understanding of gene regulation.¹ Emerging experimental techniques, such as flow cytometry, single-cell RNA sequencing, and single-molecule fluorescence *in-situ* hybridization (smFISH),^{2–4} allow for the precise quantification of these fluctuations at the single-cell level. Several approaches have been developed to fit models to the statistical moments,^{5–7} stochastic trajectories,⁸ or full probability distributions⁴ of data collected with these experimental techniques. However, substantial work remains to establish more rigorous and efficient approaches to integrate stochastic analyses with single-cell experimental data.

A little over ten years ago, the finite state projection (FSP⁹) approach was introduced to approximate the solution of the Chemical Master Equation (CME^{10,11}) and to capture the dynamics of discrete molecular events that control single-cell gene regulation. Since that time, the FSP has received substantial attention, has seen numerous computational improvements, and has become a benchmark tool in the analysis of stochastic gene regulation. Most recently, the FSP

has been used to fit and predict experimental measurements of RNA transcription in yeast, bacteria, and human cells.¹² The main utility of the FSP is to provide precise bounds on the accuracy of its approximation as well as a systematic approach to improve that accuracy. However, improved accuracy comes with increased computational cost, and no attention has been given to how one could optimize this tradeoff. Careful evaluation of this tradeoff is needed to improve the rigor and efficiency with which FSP models can be matched to experimentally measured data. In this work, we develop new FSP-based bounds on the likelihood of single-cell data given a stochastic model. We show how these bounds can be used to reduce computational costs without compromising accuracy. Finally, we use a combination of simulated and experimentally collected single-cell data to demonstrate how the co-design of FSP tools and experimental data can lead to efficient inference of discrete stochastic models.

II. MATHEMATICAL BACKGROUND

Like many single-molecule kinetic events, gene expression is often modeled as a Markov process, where each discrete state $\mathbf{x}_i = [\xi_1 \ \xi_2 \ \xi_3 \ \cdots \ \xi_N]^T_i$ corresponds to the integer numbers of N chemical species (e.g., RNA or protein). Transition events between states are different reactions such as transcription, translation, or degradation, and these reactions can be indexed by $\mu \in \{1, 2, \dots, M\}$. Each reaction occurs

with a propensity $w_\mu(\mathbf{x}_i)dt$, which is the probability that the μ th reaction will occur in the next infinitesimal time step $(t, t + dt)$ given the current state $\mathbf{x}_i$. State transitions are described as $\mathbf{x}_i \rightarrow \mathbf{x}_j = \mathbf{x}_i + \mathbf{s}_\mu$, where $\mathbf{s}_\mu$ is the stoichiometry vector that describes the change in population after the μ th reaction. In such models, each node has a continuous valued probability $P(\mathbf{x}_i, t)$ that evolves in time according to the linear ODE known as the chemical master equation (CME),^{10,11}

$$\frac{dP(\mathbf{x}_i, t)}{dt} = \sum_{\mu=1}^M w_\mu(\mathbf{x}_i - \mathbf{s}_\mu) P(\mathbf{x}_i - \mathbf{s}_\mu, t) - \sum_{\mu=1}^M w_\mu(\mathbf{x}_i) P(\mathbf{x}_i, t). \quad (1)$$

By enumerating all possible $\{\mathbf{x}_1, \mathbf{x}_2, \dots\} \in \mathbf{X}$ and corresponding probabilities, $\mathbf{P} = [P(\mathbf{x}_1, t) \ P(\mathbf{x}_2, t) \ \dots]^T$, the CME can be posed in matrix form as $\frac{d}{dt}\mathbf{P}(t) = \mathbf{A}\mathbf{P}(t)$, where $\mathbf{A}$ is known as the *infinitesimal generator* (examples of $\mathbf{A}$ are provided in Secs. IV and V).

The CME dimension is often infinite, making it impossible to solve directly for most systems. The finite state projection (FSP) approach allows one to approximate the CME solution within strict error bounds.^{4,9,13} In its formulation, the FSP approach selects a finite set of indices, $J = \{j_1, \dots, j_L\}$ with which it separates the full state space $\mathbf{X}$ into two exhaustive and disjoint sets, $\mathbf{X}_J = \{\mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_L}\}$ and its complement $\mathbf{X}_{J'}$. Under this reorganization, the full master equation can be written in matrix form as

$$\frac{d}{dt} \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{JJ} & \mathbf{A}_{JJ'} \\ \mathbf{A}_{J'J} & \mathbf{A}_{J'J'} \end{bmatrix} \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix}. \quad (2)$$

To approximate the CME, the FSP approach then forms a finite state Markov process, where all nodes in $\mathbf{X}_{J'}$ are aggregated to one or more sink states, denoted by g , which record the probability mass that leaves $\mathbf{X}_J$. However, all probability mass within g is now restricted to remain in g as time proceeds. The new, reduced FSP-CME becomes

$$\frac{d}{dt} \begin{bmatrix} \mathbf{P}_J^{\text{FSP}} \\ g(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{JJ} & \mathbf{0} \\ -\mathbf{1}^T \mathbf{A}_{JJ} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}_J^{\text{FSP}} \\ g(t) \end{bmatrix}. \quad (3)$$

The resulting approximation in Eq. (3) provides three key insights into the exact CME solution. First, it provides a lower bound on the true solution,

$$\begin{bmatrix} \mathbf{P}_J^{\text{FSP}}(t) \\ \mathbf{0} \end{bmatrix} \leq \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix} \quad \text{for all } t > 0. \quad (4)$$

This can be interpreted easily by noting that probability can only leave $\mathbf{X}_J$ in the FSP-CME (Eq. (3)), but can return from $\mathbf{X}_{J'}$ to $\mathbf{X}_J$ in the original CME (Eq. (2)). Second, the FSP provides an exact measure of the approximation error,

$$\left\| \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix} - \begin{bmatrix} \mathbf{P}_J^{\text{FSP}}(t) \\ \mathbf{0} \end{bmatrix} \right\|_1 = g(t), \quad (5)$$

where $\|\cdot\|_1$ denotes the absolute sum of a vector. Finally, as states are added to the set $\mathbf{X}_J$, the error $g(t)$ decreases

monotonically. Proofs of these three results are provided in Refs. 13 and 14.

The state space of the FSP, $\mathbf{X}_J$, can be defined through use of polynomial projection shapes,¹³

$$\mathbf{X}_J = \{\mathbf{x}_i \text{ such that } f_k(\mathbf{x}_i) \leq c_k \text{ for } k = 1, \dots, K\}. \quad (6)$$

Here $\{f_k(\mathbf{x}_i)\}$ is a set of polynomials of the population counts, and the constraints $\{c_k\}$ are weights on these polynomials that may be increased (decreased) to include more (fewer) states. In practice, each k th constraint is associated with its own sink, g_k , which aggregates all states that satisfy the $\{1, \dots, (k-1)\}$ th constraints, but not the k th constraint. The value of $g_k(t_f)$ then quantifies probability of violation for the k th constraint, which in turn guides the systematic increase of c_k . With this expansion, the FSP algorithm presented in Refs. 9 and 13 can be used to select an $\mathbf{X}_J$ to make $g(t) = \sum_k g_k(t)$ small for a specified finite time. However, lower error will require more states and greater computational expense. Until now, there has been little attention given to how small $g(t)$ should be, especially for comparison to a given set of experimental data. In this article, we will show how the combination of experimental data and the FSP results in Eqs. (4) and (5) combines to reveal a natural requirement for the accuracy of $g(t)$.

III. FSP-DERIVED BOUNDS ON THE LIKELIHOOD

To quantify how experimental data affect the accuracy requirement for the FSP, we consider single-molecule, single-cell data, such as that obtained using the technique of smFISH. This technique allows experimentalists to count the number of specific RNA molecules in individual cells. Let $\mathbf{D} = \{d_1, d_2, \dots\}$ be the set of data corresponding to d_i cells with measured populations corresponding to each state $\mathbf{x}_i$. Define $\mathcal{I}_{\mathbf{D}}$ as the subset of all i such that $d_i \neq 0$ (i.e., the support of the observed data). Assuming independence of individual cells, the logarithm of the likelihood of observing data $\mathbf{D}$ is

$$\begin{aligned} \log L(\mathbf{D}) &= \log \left(\prod_i P(\mathbf{x}_i)^{d_i} \right) \\ &= \sum_i d_i \log P(\mathbf{x}_i) = \sum_{i \in \mathcal{I}_{\mathbf{D}}} d_i \log P(\mathbf{x}_i). \end{aligned} \quad (7)$$

In many cases only partial state observations are available (e.g., only one mRNA species is measured), and with no loss of generality, $\mathbf{P}$ may be projected onto the observable dimension prior to the summation in Eq. (7). With these definitions in place, we now explore how the FSP bounds affect the computation of the likelihoods of experimental data.

The solution for the FSP is $\mathbf{P}_J^{\text{FSP}}$ and is guaranteed to be a lower bound on the model's true solution $\mathbf{P} = [P(\mathbf{x}_1), P(\mathbf{x}_2), \dots]$ by Eq. (4). The log-likelihood in Eq. (7) is monotonic in each $P(\mathbf{x}_i)$; therefore $\mathbf{P}_J^{\text{FSP}}$ provides a lower bound on the log-likelihood of $\mathbf{D}$ given the model

$$\text{LB}_J(\mathbf{D}) \equiv \sum_{i \in \mathcal{I}_{\mathbf{D}}} d_i \log P_J^{\text{FSP}}(\mathbf{x}_i) \leq \sum_{i \in \mathcal{I}_{\mathbf{D}}} d_i \log P(\mathbf{x}_i). \quad (8)$$

However, in Eq. (3) the FSP also provides the exact error in the solution of a particular model described by the CME. By redistributing the known FSP error back onto the CME solution in an optimal manner, an upper bound on $\log L(\mathbf{D})$ can be derived,

$$\begin{aligned} \text{UB}_J(\mathbf{D}) &\equiv \max_{\{\varepsilon_i\}} \sum_{i \in \mathcal{I}_D} d_i \log \left(P_J^{\text{FSP}}(\mathbf{x}_i) + \varepsilon_i \right) \\ &\quad \text{such that: } \sum_i \varepsilon_i = g \text{ and } \varepsilon_i \geq 0, \\ &\geq \sum_{i \in \mathcal{I}_D} d_i \log P(\mathbf{x}_i) \end{aligned} \quad (9)$$

where ε_i is the probability error redistributed to state $\mathbf{x}_i$. To optimize the redistribution of g and determine $\text{UB}_J(\mathbf{D})$, we developed a modified water-filling algorithm similar to those used to determine the amount of power to send to different channels in communication systems.^{15,16} To simplify notation, we define the FSP probability for each state as $p_i \equiv P_J^{\text{FSP}}(\mathbf{x}_i)$ and the corresponding portion of the total objective due to state $\mathbf{x}_i$ as $f_i \equiv d_i \log(p_i + \varepsilon_i)$. Clearly, adding probability mass to the $\mathbf{x}_i$ with the greatest $\partial f_i / \partial \varepsilon_i$ will have the largest increase in the $\sum_i f_i$. When a set of two or more $\{\partial f_i / \partial \varepsilon_i\}$ are maximal and equal, probability mass must be redistributed to equally increase all members of that set. At the optimal redistribution, all $\mathbf{x}_i$ to which probability mass has been added will have identical values for $\partial f_i / \partial \varepsilon_i$. This approach, where extreme values are first equalized before adjusting more moderate values, is analogous to raising the water level over a rough surface, filling holes to equal levels before covering hills.

The derivative of f_i with respect to ε_i is computed from Eq. (9) to get

$$\left. \frac{\partial f_i}{\partial \varepsilon_i} \right|_{\varepsilon_i=0} = \begin{cases} \frac{d_i}{p_i + \varepsilon_i} = \frac{d_i}{p_i} & \text{for } i \in \mathcal{I}_D, \\ 0 & \text{for } i \notin \mathcal{I}_D \end{cases}, \quad (10)$$

and we define N_D as the number of distinct observations (i.e., the size of $\mathcal{I}_D$). These values are then ranked in decreasing magnitude according to

$$\frac{\tilde{d}_1}{\tilde{p}_1} \geq \frac{\tilde{d}_2}{\tilde{p}_2} \geq \dots \geq \frac{\tilde{d}_r}{\tilde{p}_r} \geq \dots \geq \frac{\tilde{d}_{N_D}}{\tilde{p}_{N_D}}, \quad (11)$$

where the notation $(\tilde{\cdot})$ refers to the data-ordered state space $\tilde{\mathbf{X}}_D = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_r, \dots, \tilde{\mathbf{x}}_{N_D}\}$. An optimal redistribution of g equalizes the first n terms of Eq. (11) and satisfies the linear constraints

ALGORITHM 1. FSP error redistribution algorithm.

```

Rank  $\left. \frac{\partial f_i}{\partial \varepsilon_i} \right|_{\varepsilon_i=0} \forall i$  in decreasing order.
 $n = 1, \varepsilon_1 = g$ 
while  $\left. \frac{\partial f}{\partial \varepsilon_{n+1}} \right|_{\varepsilon_{n+1}=0} > \left. \frac{\partial f}{\partial \varepsilon_n} \right|_{\varepsilon_n=0}$  and  $n < N_D$  do
     $n \rightarrow n + 1$ 
    Solve for  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  using Eq. (12)
end while
 $\varepsilon = [\varepsilon_1, \dots, \varepsilon_n]$  is the optimal redistribution of  $g$ .

```

$$\left. \begin{aligned} \tilde{d}_{r+1} \varepsilon_r - \tilde{d}_r \varepsilon_{r+1} &= \tilde{d}_r \tilde{p}_{r+1} - \tilde{d}_{r+1} \tilde{p}_r \\ \varepsilon_r &\geq 0 \end{aligned} \right\} \quad \text{for } r \in \{1, \dots, n-1\}. \quad (12)$$

$$\sum_{j=1}^n \varepsilon_j = g.$$

For example, if the $n = 4$ greatest $\partial f_i / \partial \varepsilon_i$ can be equalized with $\sum \varepsilon_i = g$, the optimal redistribution solution can be found by the linear algebraic Eq. (12), which can be written in matrix form as follows:

$$\begin{bmatrix} \tilde{d}_2 & -\tilde{d}_1 & 0 & 0 \\ 0 & \tilde{d}_3 & -\tilde{d}_2 & 0 \\ 0 & 0 & \tilde{d}_4 & -\tilde{d}_3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} = \begin{bmatrix} \tilde{d}_1 \tilde{p}_2 - \tilde{d}_2 \tilde{p}_1 \\ \tilde{d}_2 \tilde{p}_3 - \tilde{d}_3 \tilde{p}_2 \\ \tilde{d}_3 \tilde{p}_4 - \tilde{d}_4 \tilde{p}_3 \\ g \end{bmatrix}. \quad (13)$$

In this formulation, the number of states to which probability is redistributed, n , is the largest dimension for which the solution of Eq. (13) is strictly positive for all ε_i . If the states $\mathbf{X}_J$ used by the approximation do not span the support of the distribution of data, the corresponding values of p_i are zero and $\left. \frac{\partial f_i}{\partial \varepsilon_i} \right|_{\varepsilon_i=0}$ are infinite. Therefore, the corresponding redistribution of $\{\varepsilon_i\}$ will always include some mass for those states. Algorithm 1 provides pseudocode for the proposed error redistribution approach. At most, the FSP error redistribution algorithm requires N_D iterations, and in practice computation of the upper bound on the likelihood takes only a fraction of the time needed for the FSP solution itself, especially for cases where the data correspond to partial state observations.

The FSP-derived bounds on likelihoods have several important implications for the comparison of stochastic models to single-cell data. Let $\mathbf{\Lambda}$ denote a particular combination of a model and its parameters, and let $L(\mathbf{D}|\mathbf{\Lambda})$ denote the likelihood of $\mathbf{D}$ given $\mathbf{\Lambda}$. In the case when $\mathbf{X}_J = \emptyset$ (i.e., the FSP set is empty), all of the probability mass must be redistributed, and the FSP-derived upper bound is given by

$$[\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_{N_D}] = \frac{1}{\sum_{i=0}^{N_D} \tilde{d}_i} [\tilde{d}_1 \ \tilde{d}_2 \ \dots \ \tilde{d}_{N_D}]. \quad (14)$$

This result is understood easily—the maximum of the log-likelihood function occurs when the distribution of the data exactly matches the model, and in this case the FSP upper bound describes the best any potential model can ever do. To interpret bounds for non-trivial FSP projections, we make use of the facts that (i) the FSP approximation lower bound p_i increases monotonically as $\mathbf{X}_J$ is expanded⁹ and (ii) the likelihood increases monotonically as each p_i increases. As a result, $\text{LB}_J(\mathbf{D})$ and $\text{UB}_J(\mathbf{D})$ are guaranteed to be monotonically increasing and decreasing functions of the projection size. Figure 1 illustrates the converging upper- and lower-bounds for the likelihoods of two FSP models as the size of the index set J , or equivalently the size of $\mathbf{X}_J$, is increased.

FIG. 1. Schematic of discriminatory projection sizes. Monotonically converging upper and lower FSP bounds on the likelihood are shown for two parameter sets, Λ_1 and Λ_2 . The red and green circles illustrate two pairs of projections in $\Phi(\Lambda_1, \Lambda_2)$ that enable exact ranking of the two parameter sets. The blue circle illustrates the minimal symmetric discriminatory projection, $\phi_s(\Lambda_1, \Lambda_2)$. The filled green circle illustrates the minimal nonsymmetric discriminatory projection, $\phi_2(\Lambda_1, \Lambda_2)$, needed to discriminate the system given the previous analysis of Λ_1 (open green circle).

A. Using FSP-derived bounds for model discrimination

For any two models and their parameter sets, Λ_i and Λ_j , we define the set of sufficient discriminating projections, $\Phi(\Lambda_i, \Lambda_j)$, as any pair of projection index sets, J_i and J_j , that guarantees the correct ranking of likelihoods for the two models,

$$\Phi(\Lambda_i, \Lambda_j) \equiv \{J_i, J_j\} \text{ such that } \text{UB}_{J_i}(\Lambda_i) < \text{LB}_{J_j}(\Lambda_j) \text{ or } \text{UB}_{J_j}(\Lambda_j) < \text{LB}_{J_i}(\Lambda_i). \quad (15)$$

Intuitively, these are any two projections such that the worst possible likelihood for one model is greater than best possible likelihood of the other. In Fig. 1, the red and green circles denote pairs of projections sufficient to guarantee that parameter set Λ_2 is more likely than Λ_1 . Because $\Phi(\Lambda_1, \Lambda_2)$ can contain an infinite set of such pairs with varying projection sizes, we define a minimal symmetric discriminatory projection, $\phi_s(\Lambda_i, \Lambda_j)$, as

$$\phi_s(\Lambda_i, \Lambda_j) \equiv \text{smallest set } J \text{ such that } \{J, J\} \in \Phi(\Lambda_i, \Lambda_j). \quad (16)$$

In Fig. 1, the blue circle denotes $\phi_s(\Lambda_1, \Lambda_2)$. Finally, in many sequential parameter searches, previous FSP models may already be computed to high accuracy, and it may not be necessary to demand the same accuracy for subsequent models. For this case, we define a minimum non-symmetric projection size such that

$$\phi_i(\Lambda_i, \Lambda_j) \equiv \text{smallest set } J_i \text{ such that } \{J_i, J_j\} \in \Phi(\Lambda_i, \Lambda_j). \quad (17)$$

The utility of this particular discriminatory projection size definition becomes important in parameter search problems, where it enables sequential likelihood evaluations to be conducted at minimal projection sizes. In Fig. 1, the green circles denote $\phi_2(\Lambda_1, \Lambda_2)$ a particular combination where the size of $\mathbf{X}_{J_2}$ is minimized given a previous computation for $\mathbf{X}_{J_1}$.

As the examples below will demonstrate, utilization of these minimal discriminating projections can substantially reduce the computational effort in parameter inference by eliminating much of the potential parameter space with smaller projection sizes.

B. Relationship of FSP bounds to other CME truncations

The FSP upper and lower bounds present an opportunity to better understand relationships between the FSP and other CME approximations in the context of single-cell data. For example, several groups have imposed limits of species¹⁷ or total molecule populations,¹⁸ which result in truncated master equations with reflecting boundary conditions (in contrast to the FSP's absorbing boundary condition). In a similar vein, one could renormalize the FSP solution to make use of the original FSP computation. For this renormalization, the CME can be written in terms similar to the FSP formulation,

$$\frac{d}{dt} \mathbf{P}_J^{\text{renorm}} = \mathbf{A}_{JJ} \mathbf{P}_J^{\text{renorm}} + \alpha \mathbf{P}_J^{\text{renorm}}, \quad (18)$$

where $\alpha = |\mathbf{A}_{JJ} \mathbf{P}_J^{\text{renorm}}|_1$ is the rate of flow of probability out of $\mathbf{X}_J$, which is now redistributed back into $\mathbf{X}_J$ according to the probability $\mathbf{P}_J^{\text{renorm}}$. However, the solution of this non-Markovian system is identical to simply renormalizing $\mathbf{P}_J^{\text{FSP}}$ at all times,

$$\mathbf{P}_J^{\text{renorm}} = \frac{\mathbf{P}_J^{\text{FSP}}}{|\mathbf{P}_J^{\text{FSP}}|_1}. \quad (19)$$

This can be shown easily by substituting Eq. (19) into Eq. (18).

Because the FSP likelihood bounds provide the best- and worst-case redistribution of the probability absorbed in g , likelihoods computed by reflection, renormalization, or any other arbitrary strategy are guaranteed to lay between the computed FSP bounds. Therefore, it is possible that reflecting boundaries may provide an improved approximation of the true likelihood for a particular combination of data and model. Unfortunately, the likelihoods of reflected or renormalized solutions are not necessarily monotonic, and the likelihoods of renormalized solutions for two different parameter sets or models may change rank depending on the projection size. These issues will be addressed further in Section IV C.

IV. APPLICATION OF FSP-DERIVED BOUNDS

To demonstrate the application of the FSP bounds, this section uses simulated data and three examples of stochastic gene regulation: an unregulated birth-death model, a genetic toggle switch, and a non-linear self-activating gene.

A. Birth-death model

The variability in mRNA copy numbers for housekeeping genes is well captured by the standard single-species birth

and death model, where $\mathbf{x}_i = x_i$ is the discrete number of RNA. This model consists of two reactions that describe transcription and degradation as shown in Fig. 2(a),

where the propensity functions, $\mathbf{w} = \{w_1, w_2\}$, are

$$w_1 = k_r, \quad w_2 = \gamma x.$$

Table I shows three different parameter sets for this example. A well-known analytical solution for this model, assuming $x(0) = 0$, is the time-varying Poisson distribution

$$P(x_i | k_r, \gamma) = \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}, \quad (20)$$

$$\text{where } \lambda = \frac{k_r}{\gamma} (1 - e^{-\gamma t}).$$

The infinitesimal generator for this model can be written as

$$\mathbf{A}_{ji} = \begin{cases} -w_1(x_i) - w_2(x_i) & \text{for } i = j \\ w_1(x_i) & \text{for } (i, j) \text{ such that } x_j = x_i + 1 \\ w_2(x_i) & \text{for } (i, j) \text{ such that } x_j = x_i - 1 \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

The FSP formulation for this model can be written from Eq. (21) as

$$\frac{d}{dt} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_N \\ g(t) \end{bmatrix} = \begin{bmatrix} -k_r & \gamma & 0 & \cdots & 0 \\ k_r & -k_r - \gamma & 2\gamma & \ddots & 0 \\ 0 & k_r & -k_r - 2\gamma & \ddots & 0 \\ \vdots & \ddots & \ddots & N_m \gamma & \vdots \\ 0 & \ddots & k_r & -k_r - N_m \gamma & 0 \\ 0 & \cdots & 0 & k_r & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_N \\ g(t) \end{bmatrix}, \quad (22)$$

where N_m is the size of the FSP truncation (i.e., $\mathbf{X}_J = \{x_0, x_1, \dots, x_{N_m}\}$). In this case, we used a single constraint function in Eq. (12) $f_1(x_i) = x_i \leq c_1$, where $c_1 = N_m$. To

expand the state space for this model, c_1 is simply increased by one in each iteration.

For this model and the parameters provided in the bottom row of Table I, we simulated 500 trajectories of the stochastic simulation algorithm,¹⁹ and plotted the resulting “data” in Fig. 2(b) (black). For a projection size defined by $N_m = 50$, Fig. 2(b) also shows the FSP lower bound computed using Eq. (4) in red and the FSP upper bound from Eq. (9) in blue. Figure 3 demonstrates the convergence of the upper and lower FSP bounds as c_1 is increased. Results are plotted for two different parameter sets. Increasing c_1 adds more states to $\mathbf{X}$ and monotonically decreases the error g . In turn, less error is available to be redistributed back onto the FSP solution, and $\text{UB}_J(\Lambda)$ and $\text{LB}_J(\Lambda)$ converge monotonically to the analytical value of $\log L(\mathbf{D}|\Lambda)$.

FIG. 2. FSP bound for the birth/death model. (a) Schematic of the model. (b) Probability distributions for the model at $t = 1$. Simulated data (500 runs of the SSA¹⁹) are in black. The lower bound Eq. (4) is shown in red, and the upper bound Eq. (9) is shown in blue. The shaded region denotes the redistribution of FSP error to maximize the likelihood of data. See Table I for parameters.

TABLE I. Birth and death model parameters. Production rate is k_r , degradation rate is γ , and N is the unit for discrete number RNA. Parameter set Λ_1 is used for Fig. 2. Parameter sets Λ_1 and Λ_2 are compared in Fig. 3. Parameter set $\hat{\Lambda}$ is used as the reference parameter set in Fig. 9.

Λ	$k_r(Nt^{-1})$	$\gamma(t^{-1})$
Λ_1	50.0	0.5
Λ_2	49.65	0.5
$\hat{\Lambda}$	45.0	0.5

FIG. 3. Demonstration of the converging bounds for the birth and death model. Upper and lower bounds on the likelihood of a simulated data set given two different parameter sets, Λ_1 and Λ_2 , as a function of the number of states. As the number of states increases, the upper and lower bounds monotonically converge to the true likelihood of each parameter set. The dotted horizontal lines are the likelihood values found using the analytical solutions in Eq. (20). ϕ_s indicates the minimum symmetric projection size that guarantees correct discrimination between the two parameter sets. See Table I for parameters.

B. Toggle model

We next explore the application of the FSP bounds on a model of the toggle switch²⁰ in which two genes λcI and $lacI$ mutually inhibit one another, as illustrated in Fig. 4(a). Here we consider a simple model similar to that presented by Tian and Burrage.²¹ In this model, the system state is defined by the discrete populations of the two proteins, $\mathbf{x} = [\lambda cI \text{ } lacI]$. The four reactions are

where the propensity functions $\mathbf{w} = \{w_1, w_2, w_3, w_4\}$, are given by

$$\begin{aligned} w_1 &= b_{\lambda cI} + \frac{k_{\lambda cI}}{1 + \alpha_{\lambda cI} LacI^{\eta_{\lambda cI}}}, & w_2 &= \gamma_{\lambda cI} \cdot \lambda cI, \\ w_3 &= b_{LacI} + \frac{k_{LacI}}{1 + \alpha_{\lambda cI} \lambda cI^{\eta_{\lambda cI}}}, & w_4 &= \gamma_{LacI} \cdot LacI. \end{aligned}$$

Table II provides the toggle model parameters that have been used to generate simulated data shown in black in Fig. 4(b).

The infinitesimal generator for this toggle model can be written as

FIG. 4. FSP bounds for the toggle model. (a) Schematic of the toggle model with two mutually repressing proteins, LacI and λcI . (b) Marginal probability distributions of LacI at $t = 8$ h. Simulated data are in black. The lower bound Eq. (4) is shown in red and the upper bound Eq. (9) is shown in blue. The shaded region denotes the redistribution of FSP error to maximize the likelihood of data. See Table II for parameters.

$$\mathbf{A}_{ji} = \begin{cases} -\sum_{\mu=1}^4 w_{\mu}(\mathbf{x}_i) & \text{for } i = j \\ w_1(\mathbf{x}_i) & \text{for } (i, j) \text{ such that } \mathbf{x}_j = \mathbf{x}_i + [1, 0] \\ w_2(\mathbf{x}_i) & \text{for } (i, j) \text{ such that } \mathbf{x}_j = \mathbf{x}_i + [-1, 0] \\ w_3(\mathbf{x}_i) & \text{for } (i, j) \text{ such that } \mathbf{x}_j = \mathbf{x}_i + [0, 1] \\ w_4(\mathbf{x}_i) & \text{for } (i, j) \text{ such that } \mathbf{x}_j = \mathbf{x}_i + [0, -1] \\ 0 & \text{elsewhere} \end{cases} \quad (23)$$

To apply the FSP to truncate the toggle model CME, we consider three constraint functions from Eq. (6), where c_1 , c_2 , and c_3 define the projection as

$$\mathbf{X}_J = \{\mathbf{x}_i\} \text{ such that } \begin{cases} f_1(\mathbf{x}_i) = (LacI - 4)(\lambda cI - 4) \leq c_1 \\ f_2(\mathbf{x}_i) = LacI \leq c_2 \\ f_3(\mathbf{x}_i) = \lambda cI \leq c_3 \end{cases}. \quad (24)$$

These constraints are illustrated in Fig. 5. By monotonically increasing c_k , more states are included, and the error $g(t)$ decreases. For simplicity in presentation, c_2 and c_3 were initially set at high values of 95 and 55, respectively, and the expansion only modifies the criteria c_1 . Similar results can

TABLE II. Toggle model parameters. Parameters are defined in the main text, and N is the unit for discrete molecule number. Parameter set Λ_1 is used for Fig. 4. Parameter sets Λ_1 and Λ_2 are compared in Fig. 6. Parameter set $\hat{\Lambda}$ is used as the reference parameter set in Fig. 9.

Λ	$b_{\lambda cI}$ (s^{-1})	$k_{\lambda cI}$ (s^{-1})	$\alpha_{\lambda cI}$ ($N^{-\eta_{\lambda cI}}$)	$\eta_{\lambda cI}$ —	$\gamma_{\lambda cI}$ ($N^{-1} s^{-1}$)	b_{LacI} (s^{-1})	k_{LacI} (s^{-1})	$\alpha_{\lambda cI}$ ($N^{-\eta_{\lambda cI}}$)	$\eta_{\lambda cI}$ —	γ_{LacI} ($N^{-1} s^{-1}$)
Λ_1	6.8×10^{-5}	1.6×10^{-2}	6.1×10^{-3}	2.1	6.7×10^{-4}	2.2×10^{-3}	1.7×10^{-2}	2.6×10^{-3}	3.0	3.8×10^{-4}
Λ_2	6.8×10^{-5}	1.6×10^{-2}	6.1×10^{-3}	2.1	8.0×10^{-4}	2.2×10^{-3}	1.7×10^{-2}	2.6×10^{-3}	3.0	3.8×10^{-4}
$\hat{\Lambda}$	6.8×10^{-5}	1.4×10^{-2}	6.1×10^{-3}	2.1	6.7×10^{-4}	2.2×10^{-3}	1.6×10^{-2}	2.6×10^{-3}	3.0	3.8×10^{-4}

FIG. 5. FSP state space expansion. Maximum species counts, as in Eq. (24) is $c_2 = N_{\text{LacI}} = 95$ and $c_3 N_{\text{LacI}} = 55$. States included within a truncation of $c_1 = 500$ are in gray. As c_1 is increased, the boundary (dashed lines) increases to include more states, the FSP error decreases, and the FSP bounds converge to one another (see also Fig. 6).

be obtained with more general expansion routines, provided that c_k are constant or monotonically increasing for each k . Figure 4(b) shows the marginal probability distribution for LacI with $c_1 = 150$, where the FSP lower and upper bounds are in red and blue, respectively. Although Algorithm 1 distributes the error onto the joint probability distribution of both species, results are plotted only for the marginal distribution. Figure 6 shows the converging likelihood bounds for two parameter sets as c_1 is increased to expand the number of states captured by the projection. Because these bounds are monotonic, the analysis reveals a minimal symmetric projection size, ϕ_s , at which one can guarantee that the model in green has a greater likelihood than the model in blue.

C. Comparing FSP bounds to other CME truncation approaches

To illustrate potential issues that arise through application of other CME truncation approaches, we consider a simple

FIG. 6. Demonstration of the converging bounds for a two dimensional system. Upper and lower bounds on the likelihood of a simulated data set given two different parameter sets, Λ_1 and Λ_2 , as a function of the number of states included in the toggle model. The vertical dashed line denotes the size of the minimal symmetric discriminating projection, $\phi_s(\Lambda_1, \Lambda_2)$. Parameters are given in Table II.

FIG. 7. Positive self-regulated gene expression. First order degradation (black line) and positive feedback in production (red, green, dashed) can give rise to bistable (red) or monostable dynamics (green, dashed). Blue dots represent stable equilibria and the cross represents an unstable equilibrium. See Table III for parameters and Fig. 8(c) for examples of the corresponding distributions.

gene regulation model with nonlinear self-activation. In this model, the propensity of production is given by the positive feedback function

$$w_1(x) = k_1 + k_2 \left(\frac{x^n}{m^n + x^n} \right), \quad (25)$$

and the propensity of degradation is a first order process given by $w_2 = \gamma x$. In this formulation, k_1 is the rate of production for small values of x , $k_1 + k_2$ is the rate of production for large values of x , m is the value of x at which the rate of production is halfway between k_1 and $k_1 + k_2$, and the cooperativity factor n determines the steepness of the activation function. In the deterministic regime, this model of self-regulated gene expression can lead to one or two stable equilibria as illustrated in Fig. 7. In the stochastic regime, this corresponds to bimodal distributions in some parameter regimes, and unimodal distributions in other regimes. Table III provides three possible parameter sets for this model. We simulated data from the first of these models, Λ_{True} , which admits a single stable point (see the dashed line in Fig. 7) and yields a unimodal distribution of data as shown in black in Fig. 8.

Next, we consider two perturbations of this true parameter set,

$$\begin{aligned} \Lambda_A &= [k_1(1 + \varepsilon) \quad k_2 \quad n \quad m \quad \gamma], \\ \Lambda_B &= [k_1 \quad \alpha k_2 \quad n \quad m \quad \gamma], \end{aligned} \quad (26)$$

TABLE III. Effective parameters for self-activating gene expression model. Parameters are defined in the main text. Parameter sets Λ_A and Λ_B are compared to the reference set Λ_{True} in Fig. 8.

Λ	k_1 (1/s)	k_2 (1/s)	m (N)	n –	γ (1/s)
Λ_{True}	20	40	70	4	1
Λ_A	22.5	40	70	4	1
Λ_B	20	125	70	4	1

FIG. 8. Comparing FSP bounds with other CME truncation approaches. (a) Distributions of response of the self-regulated gene. Simulated data are in black. The renormalized FSP solutions for Λ_B are shown in red and in green for Λ_A . Both methods use a projection size of 45. (b) Likelihood versus projection size for different CME truncations. The renormalized scheme is shown with dashed lines, and the reflecting scheme is shown with dotted lines. All schemes lie within the FSP bounds (solid lines) and eventually approach the correct likelihood values. Results for parameter set Λ_A are in green and for Λ_B are in red. At moderate projection sizes, the renormalized and reflecting boundary schemes appear to converge to a higher likelihood for Λ_B than for Λ_A . At higher projection sizes, the trend is switched. (c) The same as (a), but now the projection size has been increased to 285.

where Λ_A and Λ_B correspond to the red and green lines, respectively, in Figs. 7 and 8, and their parameters are given in Table III. For Λ_B the system has a bimodal response and for Λ_A , the response is unimodal. It is interesting to explore how the application of the renormalization scheme in Eq. (19) would affect comparison of these two models to simulated data. For a projection size of 45, Fig. 8(a) suggests that Λ_B provides a better match to the data than Λ_A . Moreover, the likelihood of the data given Λ_A and Λ_B appears to be nearly constant over a substantial portion of the projection space as shown by the dashed lines and the shaded region of Fig. 8(b). Based on this information, it would be easy to conclude that Λ_B is the more appropriate parameter set. However, only at large projections that include the second peak in the distribution for Λ_B , does it become apparent that the Λ_A is the better choice. This scenario illustrates how the renormalization scheme can complicate parameter discrimination for certain combinations of models and data. Similar cautions also apply for reflecting boundary approximations of the CME (see dotted lines in Fig. 8). The strict upper and lower bounds provided by the FSP provide a rigorous approach to eliminate this ambiguity as a function of projection size.

It should also be noted that likelihood computations using reflection or renormalization based truncations require the support of the CME to include that of the experimental data. Otherwise, these approaches will match the FSP lower bound, which suggests that the data are infinitely unlikely. Such a lower bound may appear useless at first, but as we will see in Sec. V, the FSP upper bound may still be sufficient for rigorous and efficient model selection.

V. FSP LIKELIHOOD BOUNDS IN PARAMETER SEARCHES

The FSP's constricting upper and lower bounds on the likelihood enable rigorous discrimination between two parameter sets, Λ_1 and Λ_2 , without using unnecessarily strict error tolerances. The following examples will demonstrate the utility of the sufficient discriminatory projections, $\Phi(\Lambda_i, \Lambda_j)$, $\phi_s(\Lambda_i, \Lambda_j)$, and $\phi_i(\Lambda_i, \Lambda_j)$.

A. Parameter search for the birth-death and toggle models

We return to the simulated data presented in Figs. 2 and 4, but this time we apply the FSP for many different parameter combinations and for many different projections. Figure 9 illustrates the practical strength of the minimal symmetric discriminating projection, $\phi_s(\Lambda_i, \Lambda_j)$. For the birth/death model in Figs. 9(a) and 9(b), parameter $\Lambda = k_r$ is allowed to vary. Figure 9(a) shows the likelihood of the data as a function of parameter Λ , and Fig. 9(b) shows the size of the sufficient symmetric projection, $\phi_s(\Lambda, \hat{\Lambda})$, needed to discriminate between Λ and a fixed parameter $\hat{\Lambda}$. Similarly, for the toggle model, both the maximum rates of production for LacI and λ CI were varied, $\Lambda = [k_{\lambda CI} \ k_{LacI}]$. In this case, Fig. 9(c) shows contour plots of the likelihoods, and Fig. 9(d) shows the corresponding contours of the size of $\phi_s(\Lambda, \hat{\Lambda})$. For models whose likelihood is better or worse than $\hat{\Lambda}$, Figs. 9(b) and 9(d) show that the comparison can be made with smaller projection sizes. Considering that the solution of Eq. (3) has a complexity that increases with order $O(n^2)$ or worse depending on the

FIG. 9. Likelihoods and sufficient projection sizes. (a) Log-likelihood versus RNA production rate k_r . The horizontal line denotes the likelihood at the fixed comparison parameter $\hat{\Lambda}$. (b) The symmetric projection size ϕ_s required to compare parameter set Λ to $\hat{\Lambda}$. For Λ such that $\log L(D|\Lambda) \cong \log L(D|\hat{\Lambda})$, larger projections are needed. The green (red) region represents parameter sets that are better (worse) than $\hat{\Lambda}$. (c,d) Same as (a,b), except for the toggle model and two variable parameters, k_{LacI} and $k_{\lambda cI}$. Parameters inside (outside) the dashed contour represent parameter sets are better (worse) than the comparison set denoted with the black dots. In all plots, dashed lines denote parameter combinations with likelihoods that are equivalent to the reference set.

solution scheme,²² such reductions can lead to substantial computational savings. In past studies, the FSP has been solved to uniformly strict error tolerances such as 10^{-6} in Neuert *et al.*,⁴ yet consideration of the FSP bounds allows for error tolerances that are relaxed by several orders of magnitude.

B. Using FSP bounds to compare CME models using experimental data

To demonstrate the application of the FSP bounds on real data, we examine a recent model and single-cell experimental data for Mitogen Activated Protein Kinase (MAPK) control of *STL1* gene regulation in *S. cerevisiae* (budding yeast). Yeast activate a variety of regulatory pathways to mitigate the osmotic pressure difference that arises from solute imbalance across the cell membrane. One such mechanism is the high-osmolarity glycerol (HOG) pathway.²³ Upon osmotic shock, the Hog1 kinase phosphorylates and localizes to the nucleus of the cell, where it initiates transcription of several genes, including *STL1*. After cells adapt to the new condition, the kinase leaves the nucleus, and the transcription pathways turn back off. Interestingly, while the nuclear localization and transcription initialization are a largely deterministic temporal signal, transcript abundance varies considerably between isogenic cells. This variability has been quantified in detail, using the smFISH technique to quantify transcript abundances in single cells at sixteen

different time points from zero to 60 min after osmotic shock.⁴

The current time-varying CME model of the *STL1* regulation process allows the gene to switch between four

FIG. 10. Gene regulation in the HOG-*STL1* system. (a) The four-state model of Hog1p-induced *STL1* gene regulation, in which each gene state ($S_1 \dots S_4$) has a distinct transcription rate. (b) The parameterized nuclear enrichment signal, Hog1p(t), that controls the rate of transition from S_2 to S_1 . This signal was parameterized from experimental measurements at 0.4M NaCl by Neuert *et al.*⁴

possible states with different transcription rates as shown in Fig. 10(a). Reactions that change the gene from state i to j occur with propensities $\{k_{ij}\}$, and the transcription rates are given by k_{r_i} , for each of the $i = \{1, 2, 3, 4\}$ th gene states. In this model, one particular transition rate k_{21} varies as a function of the Hog1p kinase in the nucleus as

$$k_{21}(t) = \max\{0, \alpha - \beta \text{Hog1p}(t)\}, \quad (27)$$

where the temporal signal profile for Hog1p was measured experimentally⁴ and is reproduced in Fig. 10(b). As a result of this dependence on a time-varying parameter, the infinitesimal generator for the CME is a function of time. The FSP truncation of the CME can be written as

$$\frac{d}{dt} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_{N_m} \\ g(t) \end{bmatrix} = \begin{bmatrix} \mathbf{S} - \mathbf{T} & \mathbf{\Gamma} & \mathbf{0} & \cdots & 0 \\ \mathbf{T} & \mathbf{S} - \mathbf{T} - \mathbf{\Gamma} & 2\mathbf{\Gamma} & \ddots & 0 \\ \mathbf{0} & \mathbf{T} & \mathbf{S} - \mathbf{T} - 2\mathbf{\Gamma} & \ddots & 0 \\ \vdots & \ddots & \ddots & \mathbf{N}_m \mathbf{\Gamma} & 0 \\ \mathbf{0} & \cdots & \mathbf{T} & \mathbf{S} - \mathbf{T} - \mathbf{N}_m \mathbf{\Gamma} & 0 \\ \mathbf{0} & \cdots & \cdots & \mathbf{1}^T \mathbf{T} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_{N_m} \\ g(t) \end{bmatrix}, \quad (28)$$

where N_m is the maximum number of RNA in the FSP approximation, and the matrices $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{\Gamma}$ are given by

$$\begin{aligned} \mathbf{S}(t) &= \begin{bmatrix} -k_{12} & k_{21}(t) & 0 & 0 \\ k_{12} & -k_{21}(t) - k_{23} & k_{32} & 0 \\ 0 & k_{23} & -k_{32} - k_{34} & k_{43} \\ 0 & 0 & k_{34} & -k_{43} \end{bmatrix}, \\ \mathbf{T} &= \begin{bmatrix} k_{r_1} & 0 & 0 & 0 \\ 0 & k_{r_2} & 0 & 0 \\ 0 & 0 & k_{r_3} & 0 \\ 0 & 0 & 0 & k_{r_4} \end{bmatrix}, \\ \mathbf{\Gamma} &= \begin{bmatrix} \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & \gamma \end{bmatrix}. \end{aligned} \quad (29)$$

Parameters for the model were found in Ref. 4 and are given in Table IV.

Figure 11 shows examples of distributions for six time points during the osmotic shock response. Experimental data⁴ (black) were collected using smFISH, and the FSP lower bound for a moderate projection size is shown in red. At

each experimentally measured time point during the dynamic process, the FSP error $g(t)$ is computed based on the FSP truncation, and the FSP upper bound on $\log L(\mathbf{D}|\mathbf{\Lambda})$ (shown in blue) is computed using Algorithm 1. In this illustrative example, we note that the projection size is substantially smaller than the support of the experimental data, yet the reduced FSP adequately captures the distribution, especially for the earlier time points. Because a good model must capture early and late time points, this observation suggests that smaller projections may be quite informative for model discrimination.

To explore the impact that experimental data has on the required projection for the FSP, we use the non-symmetric minimal projection to determine the necessary projection size needed for parameter discrimination. In this case, each sequential comparison of two models begins with a previous FSP model that is already solved to a known precision. For example, in Fig. 12, we plot the FSP error bounds versus the projection size for two Hog-*STL* model parameter sets, $\mathbf{\Lambda}_N$ and $\mathbf{\Lambda}_{N+1}$. If the likelihood is already known for the N th parameter set (horizontal dashed line in Figs. 12(a) and 12(b)), then the FSP model for the $(N+1)$ th parameter set need only be solved with a projection size corresponding to ϕ_{N+1} . Figure 12 represents this minimal non-symmetric

TABLE IV. Hog model parameters. Parameters are defined in the main text and in Neuert *et al.*⁴ Parameter sets $\mathbf{\Lambda}_{N-1}$, $\mathbf{\Lambda}_N$, and $\mathbf{\Lambda}_{N+1}$ are compared in Fig. 12. Parameter set $\mathbf{\Lambda}_{\text{fixed}}$ is used as the reference parameter set in Fig. 13. In the units, N_r refers to the discrete number of RNA and [A.U.] refers to the Hog1p nuclear signal in arbitrary units.

$\mathbf{\Lambda}$	k_{12} (1/s)	α (1/s)	β (1/(s*[A.U.]))	k_{23} (1/s)	k_{32} (1/s)	k_{34} (1/s)	k_{43} (1/s)	k_{r1} (N_r /s)	k_{r2} (N_r /s)	k_{r3} (N_r /s)	k_{r4} (N_r /s)	γ (1/s)
$\mathbf{\Lambda}_{N-1}$	2.096	5406.3	25 116.6	0.009 79	0.008 68	0.0448	0.465	9.16×10^{-4}	0.012 32	0.1372	1.953	5.53×10^{-3}
$\mathbf{\Lambda}_N$	2.096	5406.3	18 116.6	0.007 79	0.006 68	0.0448	0.465	9.16×10^{-4}	0.012 32	0.1072	1.953	5.53×10^{-3}
$\mathbf{\Lambda}_{N+1}$	2.096	5406.3	30 116.6	0.013 79	0.010 68	0.0448	0.465	9.16×10^{-4}	0.012 32	0.1372	1.953	5.53×10^{-3}
$\mathbf{\Lambda}_{\text{fixed}}$	2.096	5406.3	18 116.6	0.02	0.006 68	0.0448	0.465	9.16×10^{-4}	0.012 32	0.1072	1.953	5.53×10^{-3}

FIG. 11. FSP bounds on *STL1* mRNA distributions. Experimentally measured distributions of *STL1* transcripts are in black for each time point.⁴ The FSP lower bounds are shown in red and upper bounds are shown in blue. The Hog1-*STL1* pathway is activated at $t = 0$ with a 0.4M treatment of NaCl (see also Fig. 10). Parameters are given as Λ_N in Table IV.

projection size with black circles. In many cases, such as that shown in Fig. 12(b), the new parameter is worse than the previous case, and the necessary discriminating projection size can be much smaller than the support of the experimental data. Such situations where the next parameter set is worse than the current set are the norm in a typical parameter search.

Figure 13(a) shows the likelihood of the experimental data versus two parameters in the Hog-*STL1* model, and Fig. 13(b)

FIG. 13. Likelihoods and sufficient projection sizes for Hog1-*STL1* data. (a) The likelihood of smFISH data for 2500 different parameter combinations of k_{r3} and k_{23} . (b) The size of ϕ_{N+1} versus $\Lambda_{N+1} = [k_{r3}, k_{23}]$, where $\Lambda_N = \hat{\Lambda}$. Over most of the parameter space, sufficient discrimination does not require the full support of the experimental data ($N_m = 107$), with a median sufficient projection size of 70.

shows the size of the necessary discriminating projection for the new Λ_{N+1} given that the old Λ_N is at the black circle. For new parameters that give smaller likelihoods than Λ_N (i.e., those outside of the halo), parameter discrimination can be achieved with projection sizes that are a fraction of the support of the data. In fact the median projection needed to compare the old and new models is 70, compared to a data support size of 107.

FIG. 12. Using FSP-bounds to search Hog1-*STL1* models. (a) FSP upper and lower bounds versus projection size for old parameters Λ_{N-1} and new parameters Λ_N . In this case, Λ_N is better, and sufficient discrimination is made at ϕ_s , which corresponds to the support of the experimental data. (b) Comparison of the bounds for old parameters Λ_N and new parameters Λ_{N+1} . In this case, reusing the FSP bounds for Λ_N makes it possible to reject Λ_{N+1} at a projection size that is less than the support of the experimental data in Fig. 11. Parameters Λ_{N-1} , Λ_N , and Λ_{N+1} are given in Table IV.

VI. SUMMARY AND CONCLUSIONS

In recent years, substantial interest has arisen to integrate discrete stochastic models with single-cell experimental data. This has motivated many approaches to solve the chemical master equation, including stochastic simulations, moment closure analyses, and the finite state projection approach. Progress in this arena will continue in the future to open new biochemical processes for discrete, stochastic computational analyses. However, until now there has been little discussion of how accurate CME models need to be in order to adequately interpret experimental data. In this article, we have explored the benefit by which careful consideration of experimental data can help to reduce computational complexity and enable more efficient and rigorous comparison of multiple models in the context of experimental data. We have shown that this advance can substantially reduce the complexity of model identification for single-cell gene regulation models using real data, and we believe this approach opens new doors for gene regulation models in many pathways and organisms.

In light of the above results, it would be interesting to re-examine other approaches to fit stochastic models to single-cell data. For example, a common and highly flexible tool for this task is the stochastic simulation algorithm (SSA¹⁹). As one runs more and more SSA trajectories, the collected statistics converge to the solution of the CME, and the computed likelihood of the data given the model will also converge to the correct value. Unlike the FSP approach derived here, convergence of the SSA or other kinetic Monte Carlo schemes will not be monotonic, and long distribution tails can be very difficult to estimate. However, although the SSA does not provide a direct computation or bounds on its computational error, one can estimate the rate of convergence with increasing numbers of trajectories. With these one could imagine that the insight gained from the optimal redistribution of the FSP error could be adapted to explore similar *ad hoc* redistribution methods for the SSA. Such analyses provide intriguing paths for future theoretical, computational, and experimental investigations.

In our examples, we assumed perfect measurements of the number of RNA per cell. While the RNA-FISH approach provides a more precise quantification of molecule numbers than flow cytometry or single-cell sequencing approaches, all experiments are subject to some measurement error. Drawing insight from past efforts to predict flow cytometry data in the presence of background fluorescence,²⁴ the analyses described above can be extended to account for experimental uncertainties. Specifically, we are currently exploring experimental controls to estimate how variations in experiment or image analysis affect the probability to under- or over-count RNA spots. Such analyses will provide a baseline to calibrate how the distribution of *counted* spots conditions upon the *actual* spot number. Assuming the measurement error is identical and independently distributed for each cell,

models could be convolved with the calibrated conditional distribution to predict the probability distribution for observed spot counts, and the FSP error bounds derived above will hold with no loss in generality.

In conclusion, it is now well established that stochastic models can help to better understand single-cell gene regulatory responses. Here, we have complemented this fact by showing how single-cell data may inform the design of rigorous and yet more efficient computational analyses. Together, these insights offer further motivation for tighter integration and co-design of computational and experimental investigations of biological phenomena.

ACKNOWLEDGMENTS

This study was partially supported by the NIH-NIGMS under Award No. R25GM105608 and by the W.M. Keck Foundation. Z.F. was funded under NSF-NRT Grant No. DGE-1450032. G.N. was supported by the NIH under Award No. DP2GM114849. Contributions were as follows: B.M. designed the study; Z.F. and B.M. developed the computational approaches; Z.F. performed the numerical analyses; G.N. performed the experimental studies and advised on interpretation of the experimental data; Z.F. and B.M. wrote the manuscript.

¹B. Munsky, G. Neuert, and A. van Oudenaarden, *Science* **336**, 183 (2012).

²A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi, *Nat. Methods* **5**, 877 (2008).

³A. M. Femino, F. S. Fay, K. Fogarty, and R. H. Singer, *Science* **280**, 585 (1998).

⁴G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden, *Science* **339**, 584 (2013).

⁵A. Singh and J. P. Hespanha, *IEEE Trans. Autom. Control* **56**, 414 (2011).

⁶C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koepl, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8340 (2012).

⁷J. Ruess, A. Miliadis-Argeitis, and J. Lygeros, *J. R. Soc., Interface* **10**, 20130588 (2013).

⁸G. Lillacci and M. Khammash, *PLoS Comput. Biol.* **6**, e1000696 (2010).

⁹B. Munsky and M. Khammash, *J. Chem. Phys.* **124**, 044104 (2006).

¹⁰D. A. McQuarrie, *J. Appl. Probab.* **4**, 413 (1967).

¹¹N. G. Van Kampen and N. Godfried, *Stochastic Processes in Physics and Chemistry* (Elsevier, 1992).

¹²B. Munsky, Z. Fox, and G. Neuert, *Methods* **85**, 12 (2015).

¹³B. Munsky, in *Quantitative Biology: From Molecular to Cellular Systems*, edited by M. E. Wall (CRC Press, 2012), Chap. 11.

¹⁴B. Munsky and M. Khammash, *IET Syst. Biol.* **2**, 323 (2008).

¹⁵R. G. Gallager, *Information Theory and Reliable Communication* (Springer, 1968), Vol. 2.

¹⁶W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, *IEEE Trans. Inf. Theory* **50**, 145 (2004).

¹⁷V. Wolf, R. Goel, M. Mateescu, and T. A. Henzinger, *BMC Syst. Biol.* **4**, 42 (2010).

¹⁸Y. Cao and J. Liang, *BMC Syst. Biol.* **2**, 1 (2008).

¹⁹D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).

²⁰T. S. Gardner, C. R. Cantor, and J. J. Collins, *Nature* **403**, 339 (2000).

²¹T. Tian and K. Burrage, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8372 (2006).

²²C. Moler and C. Van Loan, *SIAM Rev.* **45**, 3 (2003).

²³S. Pelet, F. Rudolf, M. Nadal-Ribelles, E. de Nadal, F. Posas, and M. Peter, *Science* **332**, 732 (2011).

²⁴B. Munsky, B. Trinh, and M. Khammash, *Mol. Syst. Biol.* **5**, 318 (2009).