A scalable approach for solving chemical master equations based on modularization and filtering

Zhou Fang[a], **Ankit Gupta**[a], **and Mustafa Khammash**[a,1]

[a]Department of Biosystems Science and Engineering, ETH Zurich, CH-4058 Basel, Switzerland

This manuscript was compiled on October 28, 2022

Solving the chemical master equation (CME) that characterizes the probability evolution of stochastically reacting processes is greatly important for analyzing intracellular reaction systems. Conventional methods for solving CMEs include the simulation-based Monte-Carlo methods, the direct approach (e.g., the finite state projection), and so on; however, they usually do not scale very well with the system dimension either in terms of accuracy or efficiency. To mitigate this problem, we propose a new computational method based on modularization and filtering. Our method first divides the whole system into a leader system and several conditionally independent follower subsystems. Then, we solve the CME by applying the Monte Carlo method to the leader system and the direct approach to the filtered CMEs that characterize the conditional probabilities of the follower subsystems. The system decomposition involved in our method is optimized so that all the subproblems above are low dimensional, and, therefore, our approach scales more favorably with the system dimension. Finally, we demonstrate the efficiency and accuracy of our approach in high-dimensional estimation and inference problems using several biologically relevant examples.

stochastic reaction systems | chemical master equation | stochastic filtering | modularization | Rao-Blackwellization

Gene expression systems are inevitably random due to the low molecular counts of the involved chemical species (1–3). To describe this phenomenon, researchers usually model an intracellular reaction system by a continuous-time Markov chain (4), where each node in the chain corresponds to a specific cellular state, and the transitions between these nodes correspond to the reactions happening in the cell. The key to analyzing such systems lies in the computation of their time-dependent probability distribution, which is characterized by a set of linear differential equations, called the chemical master equation (CME) or known as the Kolmogorov forward equation. By solving the CME, scientists can gain deep insight into the mechanisms of complicated biological processes (e.g., the chemotaxis (5) and the cell cycle (6, 7)) and apply the acquired knowledge to the rational design (e.g., the genetic repressilator (8) and the antithetic integral feedback controller (9)). Moreover, the CME is also important for the inference and filtering of biochemical reaction systems, where this equation is used to predict the performance of candidate models (10, 11).

Usually, the CME does not have an explicit solution, with some exceptions (12–15); therefore, efficient and accurate solvers for the CME are needed. The most popular methods for this problem are the Monte-Carlo methods and the direct approach, but neither scales well with the system dimension. To solve the CME, the Monte-Carlo methods first simulate many trajectories by the Gillespie algorithm (16, 17), tau-leaping (18–20), or hybrid models (21, 22), and then they use the empirical distribution to approximate the exact probability distribution. This approach is computationally efficient even for high-dimensional systems, as long as the simulation algorithm is appropriately chosen. However, its accuracy drops quickly as the system dimension grows, which implies it performs poorly in high-dimensional settings. In contrast, the finite state projection (FSP) approach (23), a direct method that straightforwardly solves the CME on a large but finite state space, is always very accurate for chemical reaction systems. However, since the size of the state space scales exponentially with the system dimension, the FSP is computationally infeasible in high-dimensional settings.

In some situations, high-dimensional CMEs can be effectively solved by some parametric methods. When the system size is large, the linear noise approximation (24) fits the exact probability distribution by a Gaussian distribution whose mean and variance are the undetermined parameters that need to be computed based on the underlying system. The moment closure method closes the moment equations based on a chosen family of distributions and then tracks the first few moments (25–32). Also, the solutions of CMEs can be represented in the tensor-train format and solved accordingly (33, 34). More recently, some deep learning approaches (11, 35) were established for solving CMEs thanks to the universal approximation properties of neural networks. These parametric methods are quite successful in many applications. However, their validity largely depends on the suitable choice of the family of distributions, which is not trivial for generic systems.

Significance Statement

Due to the inherent randomness in microscale systems, an intracellular reaction system is often stochastic and modeled by a chemical master equation, which characterizes the probability evolution of this system. Solving this equation for high-dimensional problems is challenging: conventional methods, including the Monte-Carlo methods and the direct approach, are usually inaccurate or inefficient in these cases. To mitigate this problem, we propose a new method for solving CMEs based on modularization and filtering. Our method works by an optimized decomposition of the system, which transforms the high-dimensional problem into several low-dimensional ones; consequently, our approach scales more favorably with the system dimension. Finally, we use several biologically relevant examples to illustrate our approach in high-dimensional estimation and inference problems.

Author contributions: Z.F., A.G, and M.K designed research; Z.F. performed research; Z.F., A.G, and M.K wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence should be addressed. E-mail: mustafa.khammash@bsse.ethz.ch

When the system is multiscale, this curse of dimensionality can be somehow rescued by the method of timescale separation (e.g., (22, 36–38)), which derives reduced dimensional CMEs by applying the quasi-stationary assumption. Though this method is not always applicable, it is effective for analyzing many biological processes, which are usually multiscale.

Such a dimension reduction idea has also been applied to some hybrid methods for solving CMEs. The method of conditional moments (39) derives a reduced dimensional CME for several selected species by integrating out the moments of other species conditioned on the state of these selected ones using moment closure methods. Similarly, the uncoupled simulation method (40, 41) generates Monte-Carlo samples only for several selected species by integrating out the effect of other species using moment closure methods. These hybrid approaches (consisting of the moment closure method and some others) are computationally efficient and accurate in estimating the selected species (39–41). However, since these approaches use the moment closure method to approximate the conditional distributions of non-selected species, their accuracy and reliability are not always guaranteed, especially when Hill-type kinetics are involved. Moreover, to estimate those non-selected species, these methods have to convert the approximated moments to distributions, leading to another source of error. Also, it is still unclear how to systematically decompose the reacting system so that a hybrid approach performs optimally.

Overall, solving high-dimensional CMEs is still very challenging. To mitigate this problem, we propose in this paper a new computational method based on modularization and stochastic filtering, which turns out to scale more favorably with the system dimension.

Our method is inspired by the idea of Rao-Blackwellization (originating in (42, 43)), which decomposes a classical estimation problem into the estimation of sufficient statistics and a computational problem of integrating out other fluctuations. Such a Rao-Blackwellized estimator can be significantly more accurate than the conventional empirical estimator according to the Rao-Blackwell theorem (a variant of the law of total variance). In the past few decades, this idea has been successfully applied to sampling problems (44), stochastic filtering (45, 46), simultaneous localization and mapping problems (47), navigation and tracking problems (48, 49), biological inference problems (50), etc.

Following the idea of Rao-Blackwellization, our method first divides the whole chemical reacting system into a leader system and several conditionally independent follower subsystems in a principled manner (see Figure 1). Then, we solve the CME by applying the Monte-Carlo method to the leader system and the direct approach to the filtered CMEs that characterize the conditional probabilities of the follower subsystems. More importantly, the system decomposition involved in our method is well-designed so that the leader system and all the follower subsystems are low dimensional. Consequently, our method transforms the original high-dimensional problem into several low-dimensional ones so that each subproblem is more tractable, and their accumulated cost is much more reduced. Our method can also be viewed as a hybrid method combining both the Monte-Carlo method (applied to the leader system) and the direct approach (applied to the follower subsystems). In one extreme, when all the species are classified as leader-level species, our approach becomes the conventional Monte-Carlo method; in another extreme, when all the species are classified as follower-level species, our method is equivalent to the FSP approach. In this sense, our optimized system-decomposition algorithm can be seen as a powerful tool to create a good synergy between the Monte-Carlo method and the direct approach. Compared with other system-decomposition-based algorithms (39–41), our method avoids the moment closure step (by using the direct approach for the follower subsystems) so that the estimate of every species is always accurate and reliable. In this paper, we use several biologically relevant models to test the performance of our approach in solving CMEs and the associated stochastic filtering problems. The results show that our method can efficiently and accurately deal with high-dimensional chemical reaction systems.

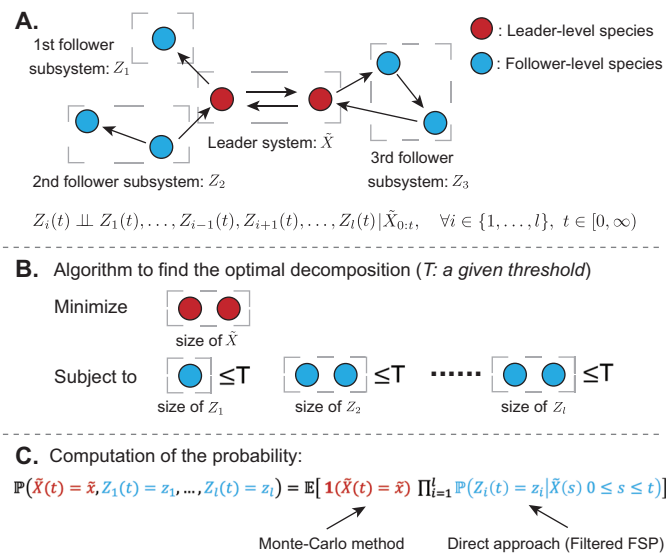

Fig. 1. Illustration of our approach. (A) System decomposition. Our approach works by a system decomposition, which first divides a chemical reaction system into a leader system (red species) and a follower system (blue species). Then, the follower system is further decomposed into several subsystems according to some topological conditions such that the follower subsystems are conditionally independent given the trajectory of the leader system. (B) Algorithm to find the optimal decomposition. Our algorithm chooses the optimal decomposition that minimizes the size of the leader system (or maximizes the size of the whole follower system) while keeping the size of each follower subsystem below a given threshold. By this algorithm, all the subsystems are low dimensional. (C) The computation of the probability distribution. Our approach solves the CME by applying the Monte-Carlo method to the leader system and a direct approach (filtered FSP (51)) to each follower subsystem. Since each subsystem is low dimensional by our system decomposition algorithm, our computational approach scales more favorably with the system dimension.

1. Theoretical results

A. Chemical master equations and the curse of dimensionality. We consider an intracellular reaction system that has r reactions:

$$v_{1,j} S_1 + \cdots + v_{n,j} S_n \to v'_{1,j} S_1 + \cdots + v'_{n,j} S_n \quad j = 1, \ldots, r,$$

where $S_1, \ldots, S_n$ are n different chemical species, and $v_{i,j}$ and $v'_{i,j}$ are the stoichiometric coefficients. Due to the low molecular counts, this chemical reaction system is usually

modeled by a continuous time Markov chain (4)

$$X(t) = X(0) + \sum_{j=1}^{r} \zeta_j R_j \left(\int_0^t \lambda_j(X(s)) \mathrm{d}s \right) \qquad [1]$$

where $X(t)$ is an n-dimensional vector representing the molecular count for each species at time t, the vector ζ_j equals to $\left(v'_{1,j} - v_{1,j}, \ldots, v'_{1,n} - v_{1,n}\right)^\top$ indicating the state change after a firing of the j-th reaction, $R_j(t)$ are independent unit rate Poisson processes, and $\lambda_j(\cdot)$ are the propensities indicating the rates of these reactions. In this paper, we consider several mild technical assumptions for this dynamical system (see *SI Appendix*, section S1) so that the process $X(t)$ is regular enough. Under these assumptions, the probability distribution of the system [1] is characterized by the chemical master equation (CME) (4)

$$\frac{\mathrm{d}p(t,x)}{\mathrm{d}t} = \sum_{j=1}^{r} \lambda_j(x - \zeta_j) p(t, x - \zeta_j) - \sum_{j=1}^{r} \lambda_j(x) p(t,x),$$

where $x \in \mathbb{Z}^n_{\geq 0}$ is the value of the state, and $p(t,x) \triangleq \mathbb{P}(X(t) = x)$ is the probability at x. By solving the CME, scientists can gain many insights into the considered biological processes.

Usually, CMEs are difficult to solve explicitly. Conventional methods to numerically solve the CME include the simulation-based Monte-Carlo methods and the finite state projection (FSP) method. However, both of them scale poorly with the system dimension.

A simulation-based Monte Carlo method first simulates N trajectories of the system [1], denoted by $\mathrm{x}_1(t), \ldots, \mathrm{x}_N(t)$, and then uses the empirical distribution $p_{\mathrm{MC}}(t,x) \triangleq \sum_{j=1}^{N} \mathbb{1}\left(\mathrm{x}_1(t) = x\right)$* to approximate the exact probability distribution. The error of this method can be evaluated by the L_1 distance between $p_{\mathrm{MC}}(t,\cdot)$ and $p(t,\cdot)$, which upper bounds the largest possible error of the Monte Carlo in estimating any particular probability. Mathematically, this error converges at the rate of $1/\sqrt{N}$, and its pre-convergence rate factor (defined by $\lim_{N\to\infty} \sqrt{N}\mathbb{E}\left[\|\hat{p}_{\mathrm{MC}}(t,\cdot) - p(t,\cdot))\|_1\right]$) lies in a particular range shown as follows (*SI Appendix*, section S2.A):

$$\lim_{N\to\infty} \sqrt{N}\mathbb{E}\left[\|\hat{p}_{\mathrm{MC}}(t,\cdot) - p(t,\cdot))\|_1\right] = \sqrt{\frac{2}{\pi}} \left[\left(\sum_{x \in \mathbb{Z}^n_{\geq 0}} \sqrt{p(t,x)} \right) \pm 1 \right]. \qquad [2]$$

This states that given a large sample size N, the error of the Monte Carlo largely depends on the value of $\sum_{x\in\mathbb{Z}^n_{\geq 0}} \sqrt{p(t,x)}$. Note that this sum is no less than $1/\sqrt{\max_x p(t,x)}$, where $\max_x p(t,x)$ can be seen as a measure of how concentrated the distribution $p(t,x)$ is, and the value of this measure tends to decay exponentially with the system dimension. So the quantity $\sum_{x\in\mathbb{Z}^n_{\geq 0}} \sqrt{p(t,x)}$ tends to scale exponentially with n. Particularly, when the molecular counts of all the chemical species are independent and identically distributed, this sum grow exponentially with n exactly, as its value equals to $\prod_{i=1}^{n} \left(\sum_{x_i \in \mathbb{Z}_{\geq 0}} \sqrt{p_i(t,x_i)} \right)$ with $p_i(t,\cdot)$ the marginal distribution for the i-th species. In summary, the error of a Monte-Carlo method tends to scale exponentially with n, and this method usually performs poorly in high-dimensional problems.

In contrast, the FSP approach (23), which directly solves the CME on a large but finite state space, is very accurate for arbitrary reaction systems. However, since the size of the finite state space scales exponentially with n, the FSP is computationally demanding for high-dimensional problems.

In conclusion, both of the methods above suffer the curse of dimensionality.

B. Rao-Blackwell method for solving CMEs. To mitigate the aforementioned problem, we use the idea of Rao-Blackwellization to propose a new computational method that solves CMEs by modularization and stochastic filtering. Specifically, our method first uses an automated algorithm to decompose the system into a leader system $\tilde{X}(t)$ and a follower system $Z(t)$ (see Figure 1). (The detailed decomposition strategy will be discussed later in this subsection.) Based on this leader-follower decomposition, we also divide reaction vectors ζ_j into $\zeta_j^{\tilde{X}}$ and ζ_j^{Z}, where $\zeta_j^{\tilde{X}}$ indicates the state change of the leader system, and ζ_j^{Z} indicates the state change of the follower system. Moreover, we further decompose the follower system $Z(t)$ into several lower-dimensional subsystems, $Z_1(t), \ldots, Z_l(t)$, such that the following topological conditions are satisfied.

C1 Each reaction can alter at most one follower subsystem's state.

C2 These follower subsystems do not directly influence each other's dynamics. In other words, all the inferences between them must flow through the leader system state.

C3 The reactions having the same non-zero $\zeta_j^{\tilde{X}}$ directly interact with at most one follower subsystem. (Here, the direct interaction means the reaction can alter the state of the follower system, and the follower system state affects the reaction's propensity.)

For the ease of notations, we rearrange the order of species such that $X(t) = \left(\tilde{X}(t), Z_1(t), \ldots, Z_l(t)\right)$; also, for every state x, we write $x = (\tilde{x}, z_1, \ldots, z_l)$, where $\tilde{x}$ is the state of the leader system, and z_i $(i = 1, \ldots, l)$ is the state of the follower system. Under the conditions above, the follower subsystems are conditionally independent given the trajectory of the leader system (see the proof in *SI Appendix*, Section S3.A). Moreover, the conditional probability distribution $\pi_{Z_i|\tilde{X}}(t,z_i) \triangleq \mathbb{P}\left(Z_i(t) = z_i \middle| \tilde{X}(s),\ 0 \leq s \leq t\right)$ is characterized by a set of differential equations with jumps (heuristically derived in (41, 52) and rigorously verified in (51)):

$$\mathrm{d}\pi_{Z_i|\tilde{X}}(t,z_i) = f\left(X(t), z, \pi_{Z_i|\tilde{X}}(t,\cdot)\right) \mathrm{d}t + \sum_{j=1}^{r_1} \mathbb{1}\left(X(t) - X(t^-) = \zeta_j^{\tilde{X}}\right) g_j\left(X(t^-), z, \pi_{Z_i|\tilde{X}}(t^-,\cdot)\right) \qquad [3]$$

whose detailed expression is given in *SI Appendix*, Section S3.A. In this paper, we call [3] the filtered CME because it estimates the conditional distribution, which filters out much noise compared to the unconditional distribution that solves the CME. A schematic illustration of the decomposition introduced above is presented in Figure 1.A. Here, we should note that this conditional independence does not necessarily

*$\mathbb{1}(\cdot)$ is the indicator function whose value equals to 1 if its argument is true, otherwise 0.

hold if we are only given the current leader state rather than its whole trajectory.

By this decomposition and the law of total expectation, we can rewrite the probability distribution by $p(t,x) = \mathbb{E}\left[\mathbb{1}\left(\tilde{X}(t)=\tilde{x}\right)\prod_{i=1}^{l}\pi_{Z_i|\tilde{X}}(z_i)\right]$. Based on it, we design a Rao-Blackwell method for CMEs, which applies Monte Carlo to the leader system and a direct approach to each follower subsystems. Concretely, we first generate N simulations of the system [1] and denote their leader-system parts by $\tilde{\mathrm{x}}_1(t),\ldots,\tilde{\mathrm{x}}_N(t)$, respectively. Then, for each simulated trajectory and each follower subsystem, we calculate the conditional probability distribution $q_j^i(t,z_i) \triangleq \mathbb{P}\left(Z_i(t)=z_i \,\middle|\, \tilde{X}(s)=\tilde{\mathrm{x}}_j(s),\ 0\le s\le t\right)$ using a direct approach, called the filtered FSP (51), which solves [3] directly on a large but finite state space. Finally, we approximate the exact probability distribution by the quantity

$$\hat{p}_{\mathrm{RB}}(t,x) = \frac{1}{N}\sum_{j=1}^{N}\left[1\left(\tilde{\mathrm{x}}_j(t)=\tilde{x}_j\right)\prod_{i=1}^{l}q_j^i(t,z)\right]. \qquad [4]$$

We name this algorithm the Rao-Blackwellized CME solver (RB-CME solver). Mainly, this Rao-Blackwellized method transforms the original problem of solving the CME into several potentially low dimensional subproblems. Therefore, our method tends to scale more favorably with the system dimension.

The RB-CME solver can be seen as a combination of the Monte-Carlo method and the direct approach, as it generates random variables for the leader system and solves the filtered CME by the direct approach. Particularly, when all the species are classified as leader-level species, this method becomes a Monte-Carlo method; when all the species are classified as follower-level species, the filtered CME becomes the CME, and the RB-CME solver is equivalent to the FSP method. By classifying chemical species differently, one can choose whether this method is similar to the direct approach or to the Monte-Carlo method. In high dimensional cases, our method is more computationally practical than the FSP approach, as our method applies the direct approach to each follower subsystem separately.

Given the same samples size N, the RB-CME solver is no less accurate than the conventional Monte-Carlo method. Basically, the RB-CME solver is also a kind of Monte-Carlo method[†]; therefor, its L_1 error has a convergence rate $\sqrt{N}$ and its pre-convergence rate factor depends on the variance of the random variable it generates. The expression of this pre-convergence factor is given in *SI Appendix*, Section S3.C. Note that the variance of $\mathbb{1}\left(\tilde{X}(t)=\tilde{x}\right)\prod_{i=1}^{l}\pi_{Z_i|\tilde{X}}(z_i)$ (used for the RB-CME solver) is no greater than that of $\mathbb{1}\left(X(t)=x\right)$ (used for the conventional Monte-Carlo method) due to the law of total variance. So, we can conclude the superior performance of our method in the sense of the pre-convergence rate factor, i.e., (see *SI Appendix*, Section S3.C)

$$\lim_{N\to\infty}\sqrt{N}\mathbb{E}\left[\|\hat{p}_{\mathrm{RB}}(t,\cdot)-p(t,\cdot))\|_1\right] \le \lim_{N\to\infty}\sqrt{N}\mathbb{E}\left[\|\hat{p}_{\mathrm{MC}}(t,\cdot)-p(t,\cdot))\|_1\right]$$

Particularly, if the conditional probability distribution $\pi_{Z_i|\tilde{X}}(z_i)$ only depends on the final state of the leader system $\tilde{X}(t)$ and is irrelevant to any of its historical information, then the RB-CME solver is equivalent to the time scale separation method, which eliminates all the follower-level species using quasi-stationary assumption. In this case, our method's error equals to the Monte-Carlo method's error for estimating $\tilde{X}(t)$ solely *SI Appendix*, Section S3.C, and its pre-convergence rate factor is far less than [2] when the dimension of $\tilde{X}(t)$ is much less than that of $X(t)$. This gives a clue that the RB-CME solvers can be far more accurate than the Monte-Carlo method in general high-dimensional problems.

Despite this error reduction, given the same sample size N, the RB-CME solver is more computationally demanding than the Monte-Carlo method, as our method needs to additionally solve the filtered CMEs. Therefore, at the same time cost, the superior performance of our method is not guaranteed. To achieve this goal, we appropriately design the system decomposition involved in our method as follows.

Recall that for each $\tilde{\mathrm{x}}_j(t)$ and each subsystem $Z_i(t)$, we solve the filtered CME directly on a finite state space. We denote the size of this space by $\mathtt{SS}_i$. Intuitively, the RB-CME solver is more accurate when the whole follower system inferred by the direct approach has a large size, i.e., $\prod_{i=1}^{l}\mathtt{SS}_i$ is large. On the other hand, since we solve the filtered CME for each follower subsystem separately, the time cost of our method can be measured by the size of the largest follower subsystem, i.e., $\max_i \mathtt{SS}_i$, when the computations are well parallelized. Consequently, to optimize the RB-CME solver, we choose a leader-follower decomposition (among all) that maximizes the size of the whole follower system ($\prod_{i=1}^{l}\mathtt{SS}_i$) while keeping the size of each individual follower subsystem ($\mathtt{SS}_i$) below a given threshold. The detailed algorithm following this strategy is provided in *SI Appendix*, Section S3.D. With this decomposition algorithm, we do divide the problem of solving CMEs into several lower-dimensional computational subproblems, and, consequently, our approach scales more favorably with the system dimension in accuracy and efficiency.

Finally, we summarize the procedure of the RB-CME solver as follows (also see a schematic illustration in Figure 1).

1. Decompose the system into a leader system and several follower subsystems using the algorithm in *SI Appendix*, Section S3.D.

2. Generate N simulations of the leader system, and denote them by $\tilde{\mathrm{x}}_1(t),\ldots,\tilde{\mathrm{x}}_N(t)$.

3. For each trajectory $\tilde{x}_j(\cdot)$ and each subsystem $Z_i(t)$, solve $q_j^i(t,z_i) \triangleq \mathbb{P}\left(Z_i(t)=z_i \,\middle|\, \tilde{X}(s)=\tilde{\mathrm{x}}_j(s),\ 0\le s\le t\right)$ by applying the filtered FSP to the filtered CME [3].

4. Compute the result by [4].

C. Rao-Blackwell method for solving the filtering problem. With modern microscope, scientists can measure fluorescent reporters and use these measurements to infer the dynamical states of unobserved species (e.g., the gene state). This whole procedure is called stochastic filtering for intracellular reaction systems. By solving this problem, scientists can gain knowledge about the chemical species that are unobserved, and this can potentially lead to better control strategies for the reacting process. Here, we apply the proposed RB-CME solver to this filtering problem.

[†] In the rest of this paper, we refer the Monte-Carlo method to the conventional one we introduced in the previous subsection.

Mathematically, we can model the observation channels by

$$Y(t_i) = h(X(t_i)) + \sigma W_i$$

where t_i are the observation time points, $Y(t_i)$ is a vector of observations with each element corresponding to a particular light frequency, $h(\cdot)$ is a vector-valued function indicating the ideal relation between the measurement and the state, σ is a diagonal matrix indicating observation noise intensities, and W_i are vectors of independent standard Gaussian noise. Stochastic filtering aims to compute the conditional probability distribution of the state $X(t_i)$ (for every $i = 1, 2, \dots$) given the observations up to time t_i, i.e., $\pi_{t_i}(x) \triangleq \mathbb{P}\left(X(t_i) = x \middle| Y(t_s),\ 1 \le s \le i\right)$. By Bayes' rule, $\pi_{t_i}(x)$ satisfies the following recursive formulas (53)

$$\rho_{t_{i+1}}(x) = \sum_{x' \in \mathbb{Z}^n_{\ge 0}} \mathbb{P}\left(X(t_{i+1}) = x \middle| X(t_i) = x'\right) \pi_{t_i}(x') \qquad [5]$$

$$\pi_{t_{i+1}}(x) \propto L\left(Y(t_{i+1}) \middle| x\right) \rho_{t_{i+1}}(x) \qquad [6]$$

where $\rho_{t_{i+1}}(x) \triangleq \mathbb{P}\left(X(t_{i+1}) = x \middle| Y(t_s),\ 1 \le s \le i\right)$ and $L(y|x)$ is the density function of $\mathbb{P}\left(Y(t_{i+1}) \in \cdot \middle| X(t_{i+1}) = x\right)$ (usually called the likelihood function). We can interpret [5] as the prediction of the state at the next time point t_{i+1} using the observation up to the current time t_i. We interpret [6] as the adjustment of the prediction according to the new observation. Note that the prediction step [5] is actually solving a CME with $\pi_{t_i}(\cdot)$ being the initial probability distribution and $\rho_{t_{i+1}}(\cdot)$ being the final solution. Consequently, the filtering problem can be seen as a combination of the usual CME and an adjustment step.

Following the idea above, one can solve the filtering problem by applying various CME solvers to the prediction step. Conventional methods include the particle filter that uses the Monte-Carlo method for [5] (e.g., (54–56), to cite a few) and the direct approach that uses the FSP for [5]. Similar to the situation in solving CMEs, these two approaches scale poorly with the system dimension when solving the filtering problem. Specifically, the particle filter has a similar L_1 error to the Monte-Carlo method (for solving CMEs) when they are applied to the same system (*SI Appendix*, Section S2.B), and, therefore, the particle filter can be inaccurate when the system dimension is large. Besides, applying the direct approach to the filtering problem can be time-consuming because the size of the state space scales exponentially with n.

Here, we solve the filtering problem by applying the RB-CME solver to [5]; we call this approach the Rao-Blackwellized particle filter (RB-PF). In this case, the follower subsystems need to be conditionally independent given the trajectories of both the leader system and the observation. To this end, we require the following condition for the system decomposition in addition to C1, C2, and C3. (The proof is given in *SI Appendix*, Section S4.A.)

C4 Fluorescent reporters from different follower subsystems have different colors. In other words, each entry of $h(\cdot)$ can be affected by one follower subsystem at most (besides the leader system).

This requirement automatically holds in many practical applications because experimentalists usually give different colors to different genes. We provide the modified leader-follower decomposition algorithm for the filtering problem in *SI Appendix*, Section S4.A, and the detailed algorithm of the RB-PF in *SI Appendix*, Section S4.B.

We found that when applied to the same chemical reacting system, the RB-PF (for the filtering problem) usually has similar accuracy to the RB-CME solver (for solving CMEs) *SI Appendix*, Section S4.C, which suggests that the RB-PF also scales favorably with the system dimension. Also, this means that for a given reaction system, if the RB-CME solver can accurately solve its CME, then the RB-PF can also accurately solve its filtering problem and vice versa. To conclude, this consistency result guarantees good performance of the RB-PF.

2. Case studies

Now, we illustrate our approach using several biologically relevant models. Unless otherwise indicated, the experiments are performed on a high-performance cluster where each node has a 2.25-GHz, 15-core CPU. The code is available on GitHub: "https://github.com/ZhouFang92/Rao-Blackwellized-CME-solver".

A. A class of linear reaction systems. To demonstrate the scalability of the RB-CME solver, we applied our approach to a class of expandable linear networks shown in Figure 2A and compared its performance with the traditional approaches. Specifically, the linear network consists of n chemical species and three types of reactions: the production ($\emptyset \to S_i$), the degradation ($S_i \to \emptyset$), and the conversion of S_i into S_{i+1} ($i \le n-1$). We set the reactions to follow the mass-action kinetics with the rate constants presented in the caption of Figure 2. At the initial time, the molecular counts of different species are independent and have a Poisson probability with mean 0.5. In this setting, the associated CME can be solved by a multivariate Poisson distribution whose mean evolves according to the deterministic dynamics of the system (12).

We first compared the scalability of different approaches in terms of accuracy and efficiency. For the FSP method, we truncated the state space by $\bigotimes_{i=1}^{n}\{0, 1, \dots, 9\}$, which contains most of the probability. For the RB-CME solver, we similarly truncated the state space for each follower-level species by $\{0, 1, \dots, 9\}$ and required each follower subsystem to contain no more than 100 states. Our leader-follower decomposition algorithm consistently classifies species S_{3i} ($i = 1, 2, \dots$) as leader-level species and the rest as follower-level species. Moreover, each pair of S_{3i-2} and S_{3i-1} ($i = 1, 2, \dots$) form a follower subsystem. Also, we set the RB-CME solver and the Monte-Carlo method to have respectively 10^4 samples and 10^5 samples so that their computational time is relatively the same. The experimental results are shown in Figure 2B.

Figure 2B indicates that the computational time of the RB-CME solver scales well with the system dimension (n), and its accuracy is much better than that of the Monte-Carlo method at the same time cost. Specifically, the first block in Figure 2B shows that the computational time of the Monte-Carlo method grows quadratically with n, which is because the Gillespie method has the computational complexity $\mathcal{O}(\texttt{\#reaction channels} \times \texttt{\#reaction firing events})$ (see (57, Section III)), and both of these quantities grow linearly with n in this example. For the FSP method, the algorithm's time-complexity is linear with the truncated space size, which scales exponentially with the dimension n in

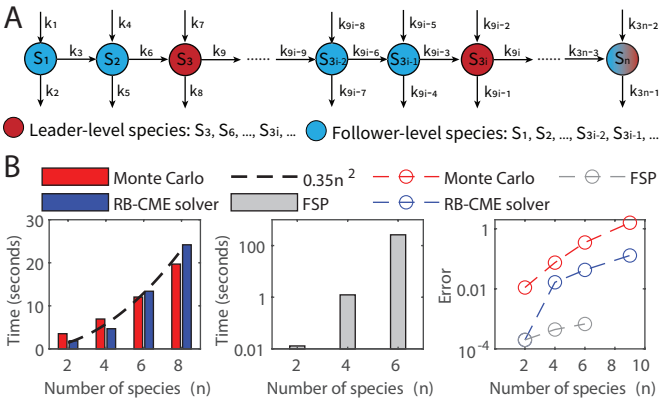

Fig. 2. Scalability of the RB-CME solver in a class of linear networks. (A) A class of linear networks that consists of three types of reactions: the production, degradation, and conversion of S_i into S_{i+1} ($i = 1, \ldots, n-1$). All the reactions follow the mass-action kinetics, and their reaction constants are $k_1 = 2.4$, $k_{3n-1} = 1.6$, $k_{3i+1} = 0.9$, $k_{3i-1} = 0.6$, and $k_{3i} = 1$ ($i = 1, \ldots, n-1$). At the initial time, each species has a Poisson probability with mean 0.5, and all of them are independent. (B) The scalability of the Monte-Carlo method (with 10^5 samples), RB-CME solver (with 10^4 samples), and finite state projection approach (with the truncated space $\bigotimes_{i=1}^{n}\{0, 1, \ldots, 9\}$) when solving the CME at time 10. In the third block, the error is evaluated by the L_1 distance between the numerical solution and the exact solution of the CME. This panel tells that the computational time of the Monte-Carlo method and that of the RB-CME solver both grow quadratically with the system dimension (n), whereas the computational time of the FSP method grows exponentially with n. Moreover, the error of the Monte-Carlo method and that of the RB-CME solver both grow exponentially with n, but the latter grows much slower than the former (see the slopes of the linear-like curves in the log-domain). Notably, the RB-CME solver is as accurate as the FSP method when $n = 2$; it is because in this case, no leader-level species exists, and, therefore, the RB-CME solver is equivalent to the FSP method.

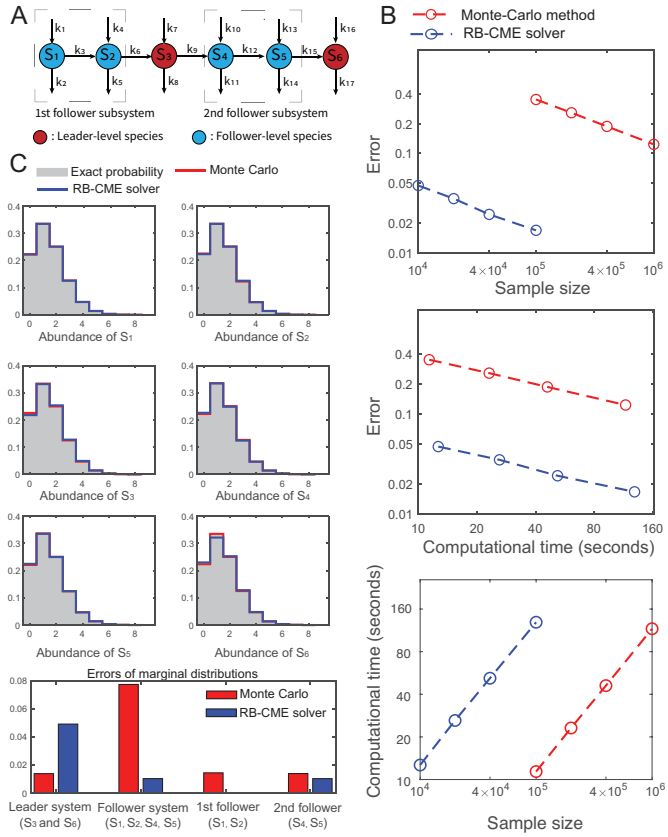

Fig. 3. Comparison of the RB-CME solver and Monte-Carlo method in the linear network with six species. (A) The diagram of the linear network with six species: all the settings are the same as that in Figure 2. (B) Convergence of both approaches in terms of the L_1 error. Both methods converge at the rate of $1/\sqrt{N}$ or $1/\sqrt{T}$, where N and T are the sample size and computational time, respectively. This also implies that for both approaches, the computational time is proportional to the sample size (see the last block). With the same sample size, the RB-CME solver is 20 times more accurate than the Monte-Carlo method, and at the same time cost, the RB-CME solver is 8 times more accurate. (C) Performance of the RB-CME solver (with 10^4 samples) and Monte-Carlo method (with 10^5 samples) in estimating marginal distributions. Both methods accurately approximate the marginal distributions for individual species, but they perform quite differently in estimating joint probabilities. Specifically, the RB-CME solver is more accurate in estimating the follower system, particularly the first follower subsystem consisting of S_1 and S_2, but it is less accurate in estimating the leader system.

this example, so its cost also grows exponentially with n (see the second block of Figure 2B). In contrast, though the RB-CME solver utilizes the direct approach to the follower system whose size also grows exponentially with n, its computational time still scales quadratically with n (see the first block of Figure 2B). This reduced computational complexity is because we apply the direct approach to each follower subsystems separately rather than in whole; therefore, in our algorithm, the computational cost of the FSP part becomes $\mathcal{O}(\texttt{size of the largest follower subsystem} \times \texttt{\#follower subsystems})$ where the first term is fixed ($= 100$), and the second term scales linearly with n. As for the accuracy, the third block of Figure 2B tells that the error of the Monte-Carlo method and that of the RB-CME solver both scale exponentially with n, but the latter grows much slower than the former. Notably, when the system dimensional equals two, i.e., all the species are follower-level species, the RB-CME solver is equivalent to the FSP method, and both approaches are equivalently accurate. All these results indicate that the RB-CME solver is a good compromise between the Monte-Carlo method and the FSP approach, and it is more favorable for high-dimensional problems.

To further understand the benefit of the RB-CME solver, we investigated the linear network with six species in more detail (see Figure 3). From the results, we can observe that both the Monte-Carlo method and the RB-CME solver converge to the exact probability distribution at the rate of $1/\sqrt{N}$, which agrees with the law of large numbers (see Figure 3B). Moreover, given the same sample size (resp., the same computational time), the RB-CME solver is significantly more accurate than the Monte-Carlo method with an improvement of 20 times (resp., 8 times), respectively. Note that the improvement of the RB-CME solver given the same sample size is greater than the improvement at the same time cost. This is because the RB-CME solver needs to additionally solve the filtered CME for follower subsystems and, therefore, is more time-consuming given the same sample size. We also studied the performance of both approaches in estimating the marginal distributions when the time costs are relatively the same (see Figure 3C). The result shows that both methods accurately approximate the marginal distribution for individual species, but their performance in estimating joint probabilities is very different. Concretely, the RB-CME solver is more accurate in estimating the follower system, especially the first follower subsystem consisting of S_1 and S_2, but it is less accurate in estimating the leader systems (see the last block in Figure 3B). The relative inaccuracy of the RB-CME solver for the leader system is attributed to the fact that both the RB-CME solver and the

Monte-Carlo method use the same protocol to estimate the leader system, and at the same time cost, the RB-CME solver has fewer samples than the Monte-Carlo method. Despite this issue, the RB-CME solver still has a much better performance in estimating the whole system because there are more follower-level species than leader-level ones.

B. Repressilator. To demonstrate the performance of the RB-CME solver for nonlinear systems, we consider a well-known genetic circuit called the repressilator (see Figure 4), where three gene expression systems cyclically repress each other. The repressilator was first engineered by Elowitz and Leibler in 2000 (8), and its name comes from the cyclical repression topology and its oscillatory dynamic behavior. In this system, the nonlinearity comes from the mRNA production processes, whose propensities are hill functions. More details about the modeling are presented in *SI Appendix*, Section S5.A.

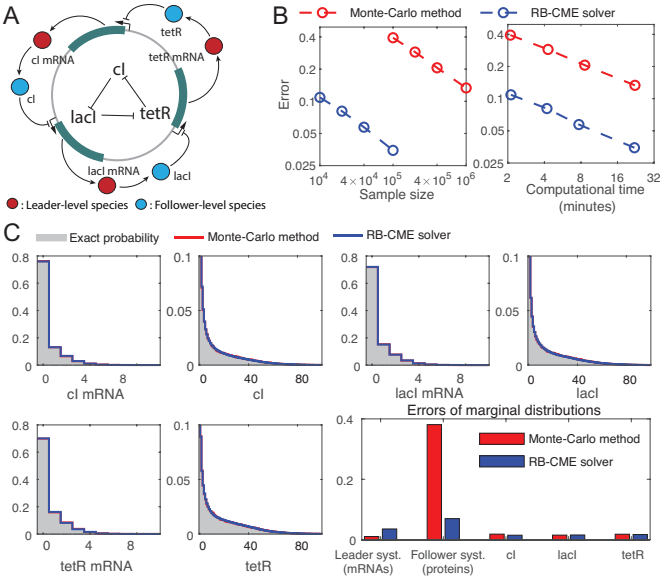

Fig. 4. Performance of the RB-CME solver in the repressilator. (A) Diagram of the repressilator with three gene expression systems whose proteins cyclically repress each other. Our method classifies all the mRNAs as leader-level species and all the proteins as follower-level species. (B) Convergence of the Monte-Carlo method and the RB-CME solver. We depict the error by the sum of the L_1 errors in estimating the leader system and the follower system. The exact probability distribution is approximated by the Monte-Carlo method with 3×10^9 samples. Given the same sample size or the same time cost, the RB-CME solver is much more accurate than the Monte-Carlo method. (C) Performance of the RB-CME solver (with 10^4 samples) and the Monte-Carlo method (with 10^5 samples) in estimating the marginal distributions. Both approaches have relatively the same computational time, and they both accurately estimate the marginal distributions of individual species. The superior performance of the RB-CME solver is attributed to the estimation of the follower system, which dominates in the whole estimation problem.

In this model, we intended to compare the RB-CME solver with other approaches. With the parameters set in *SI Appendix*, Section S5.A, most of the probability is contained in the state space where each mRNA has no more than 20 copies, and each protein has no more than 200 copies. Notice that storing a probability distribution on this state space requires about 500 GB (8 bytes/state $\times$ $\left(20^3 \times 200^3\right)$ states), so applying the FSP approach to this example is impractical. Alternatively, to get an accurate solution of the CME, we simulated 3×10^9 trajectories of this dynamical system and used the empirical distribution to approximate the exact probability distribution. This whole procedure took 10 graphics processing units (GPUs) about 24 hours! Also, to avoid the storage problem, we only stored the marginal distribution of the mRNAs and the marginal distribution of the proteins. For the RB-CME solver, we set the truncated state space for each mRNA to be $\{0, 1, \ldots, 19\}$ (if the mRNA is classified as a follower-level species), the truncated state space for each protein to be $\{0, 1, \ldots, 199\}$ (if the protein is classified as a follower-level species), and each follower subsystem to contain no more than 200 states. In this setting, the leader-follower decomposition algorithm classifies all the mRNAs as the leader-level species and all the proteins as the follower-level species. Finally, we applied both the RB-CME solver and Monte-Carlo method to the repressilator; the results are shown in Figure 4.

Again, the results show that the RB-CME solver is efficient and accurate in this example. Specifically, Figure 4B tells that given the same sample size, the RB-CME solver is about 15 times more accurate than the Monte-Carlo method, and given the same time cost, the RB-CME solver is 4 times more accurate. Notably, by only taking one CPU and about half an hour, our method can provide a very accurate solution with an L_1 error of 3% in estimating marginal distributions of the mRNAs and the proteins. Furthermore, we compared the performance of both approaches in estimating the marginal distributions when their computational time is relatively the same (see Figure 4C). From this panel, we can observe that both methods accurately estimate the marginal distribution of individual species, and the superior performance of the RB-CME solver is attributed to the estimation of the follower system. More specifically, the proteins have much more molecular counts than the mRNAs, and, therefore, the marginal probability distribution of the proteins tends to be more dispersive than that of the mRNAs. According to our analysis in Section 1.A, this explains why the estimation error of the follower system (the proteins) dominates in the Monte-Carlo method (see Figure 4C). In contrast, the RB-CME solver has a much lower error in estimating the proteins because it applies the direct approach to the follower system, and this advantage greatly dominates the disadvantage of the RB-CME solver in estimating the mRNAs (see Figure 4C). More interestingly, when estimating each follower subsystem, the improvement of the RB-CME solver is not that significant compared with that in estimating the whole follower system (see Figure 4C), which again shows that our approach scales more favorably with the system dimension. These observations also imply that the RB-CME solver can accurately estimate other genetic circuits if all the species with large molecular counts are classified as follower-level species.

C. Stochastic filtering for the genetic toggle switch. Now, we consider the filtering problem for another well-known genetic circuit, called the genetic toggle switch, which was first engineered by Gardner, Cantor, and Collins in 2000 (58). This circuit consists of two gene expression systems, whose protein products repress each other's expression (see Figure 5A), and their trajectories exhibit switching behaviors (see Figure 5B). We assume that the first protein is fluorescent and measured by a microscope at several time points, and our goal is to infer the hidden dynamical state given the observations. More details about the modeling are presented in *SI Appendix*, Section S5.B.

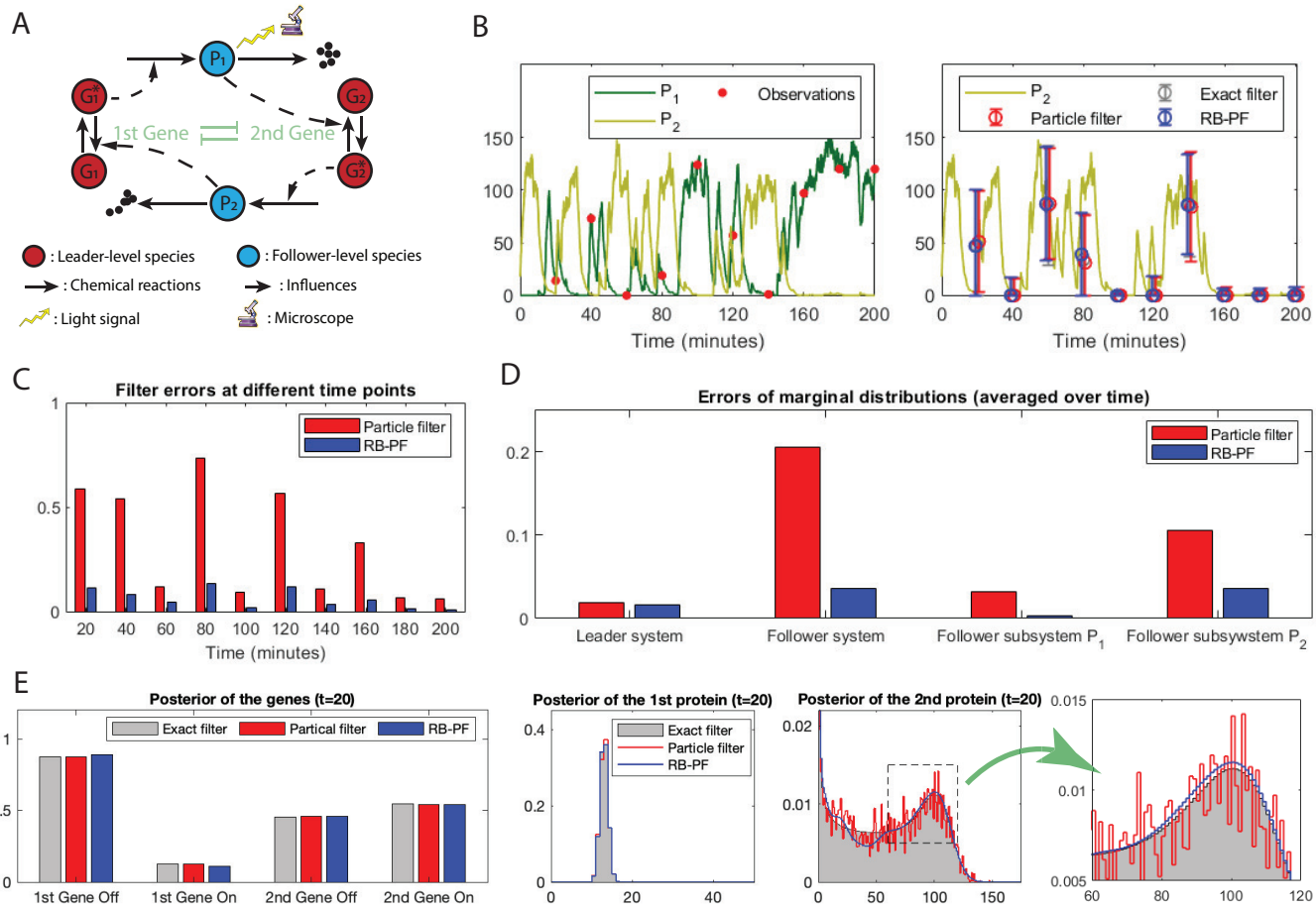

Fig. 5. Filtering results for the genetic toggle switch. (A) Description of the genetic toggle switch with two gene expression systems whose proteins deactivate each others' genes. The protein of the first gene is fluorescent and can be measured by a microscope. The observation noise intensity is set to be 1. In this setting, our approach classifies all the genes as the leader-level species and all the proteins as the follower-level species. (B) Performance of different filters in estimating the conditional mean of the second protein given a discrete-time trajectory of the first protein. In this example, the exact filter is approximated by the direct method. The particle filter and the Rao-Blackwellized particle filter (RB-PF) have, respectively, 10^5 samples and 10^4 samples, and, in this setting, they consume relatively the same computational time. This panel tells that all the filters have similar performance in estimating the conditional mean of the second protein. (C) L_1 errors of the PF and the RB-PF at different time points. It shows that the RB-PF outperforms the PF at every observation time point, and the improvement ratio is quite significant. (D) Average L_1 errors of the PF and the RB-PF in estimating the marginal distributions. (E) Evaluation of marginal posteriors at the first observation time ($t = 20$). The last two panels show that the major difference between the PF and the RB-PF lies in the estimates of the follower system.

In this example, we intend to compare the RB-PF with other filtering approaches. First, we simulated a trajectory of the system and generated observations at 10 different time points. Our task is to infer the hidden states based on these generated observations. To get an accurate approximation of the exact filter, we applied the direct approach to the problem with a state space where each protein has fewer than 200 copies. This procedure took 6.5 hours! For the RB-PF, we set the truncated state space for each protein to be $\{0, 1, \ldots, 199\}$ (if it is classified as a follower-level species) and the truncated state space for genes to be $\{0, 1\}$ (if it is classified as a follower-level species). By letting each follower subsystem have no more than 200 states, our approach classifies all the proteins as the follower-level species and the rest as leader-level species. Finally, we applied both the RB-PF and conventional particle filter to the filtering problem, and the results are shown in Figure 5 and Figure 6.

In Figure 5, we compare the performance of the RB-PF (10^4 samples) and PF (10^5 samples), which have the same time cost (see Figure 6B). The numerical results show that though the RB-PF and the PF have similar performance in estimating the conditional mean and variance of the second protein (Figure 5B), the RB-PF is far more accurate in estimating the whole conditional probabilities. Specifically, the RB-PF consistently performs better than the PF at different observation time points, and the improvement ratio is always significant (Figure 5C). When looking at the marginal distributions, we can observe that the superior performance of the RB-PF also comes from the estimation of the follower system (Figure 5D). Particularly, in this example, the estimation error of the follower system (the proteins) dominates in the PF (see Figure 5D), as its conditional probability distribution is more dispersive than that of the leader system. Thanks to the direct approach that our method applies to the follower system, the RB-PF is more accurate than the PF in estimating the follower system (see Figure 5D), and, therefore, the RB-PF is significantly more accurate in estimating the whole conditional probability distribution. From Figure 5E,

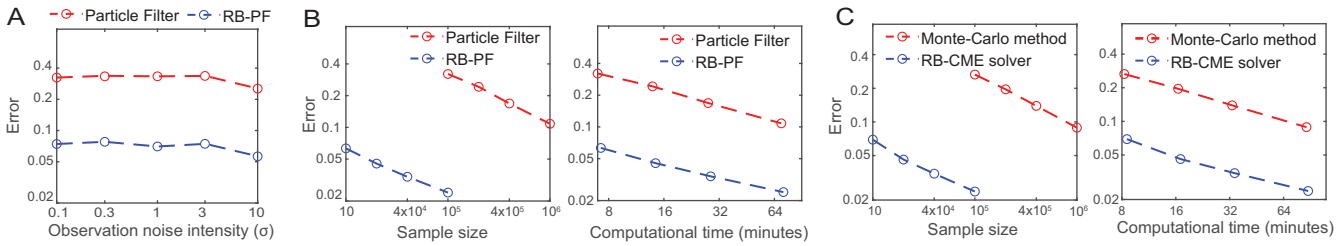

Fig. 6. Performance of the Rao-Blackwellized particle filter (RB-PF) in different settings. In this figure, we still consider the filtering problem for the genetic toggle switch (see Figure 5) and evaluate the performance of the RB-PF under different problem settings. (A) Performance of the particle filter (PF) and the RB-PF under different observation noise intensity. The PF and RB-PF have 10^5 samples and 10^4 samples, respectively, and the performance is evaluated by the average L_1 error over 10 observation time points. The panel shows that the performance of both filters are robust to the intensity of the observation noise. (B) Convergence of the PF and the RB-PF under a fixed observation noise intensity ($\sigma = 1$). Again, the performance is evaluated by the average L_1 error over 10 observation time points. (C) Convergence of the Monte-Carlo method and the RB-CME solver in computing the CME of the genetic toggle switch at time 200. In this panel, the performance is evaluated by the L_1 error between the numerical result and the exact solution, where the exact solution is approximated by the FSP method. The last two panels show that the RB-CME solver and the RB-PF have similar performance in their associated problems given the same sample size or the same computational time. The same is also true for the Monte-Carlo method and PF.

we can observe that for the follower system, the solution of the RB-PF is more smooth and accurate, whereas the result of the PF is fuzzier and more inaccurate, which again shows the advantage of RB-PF. More interestingly, in this filtering problem, the RB-PF and the PF have similar performance in estimating the leader system (Figure 5D and Figure 5E), which is quite different from the situation in solving the CME where the Rao-Blackwell method is less accurate than the conventional Monte-Carlo method (see Figure 3 and Figure 4 in previous examples). The reason for this phenomenon is that in the correction step, the conditional probability distribution of the leader system is influenced by the conditional probability distribution of the follower system via the likelihood function, and, therefore, the leader system can also benefit from Rao-Blackwellization.

We also compared the performance of the RB-PF and the PF under various different conditions. First, the performance of both filters is quite robust to the variation of the observation noise intensity Figure 6A. Moreover, the RB-PF and the RB-CME solver perform similarly in their associated problems (see Figure 6B and Figure 6C); the same is true for the PF and the Monte-Carlo method. These results agree well with our analysis in the theoretical part (also see *SI Appendix*, Section S2.B and Section S4.C). Also, they suggest a consistency result that for one system, if the RB-CME solver can accurately solve its CME, then the RB-PF should also perform well in its filtering problem and vice versa. Besides, similar to the situation in solving the CME, the RB-PF is orders of magnitude more accurate than the PF given the same sample size or the same time cost Figure 6B. All these conclusions indicate the reliability of our approach.

3. Discussion

Due to the inherent randomness in microscopic organisms, an intracellular reaction system is usually modeled by a chemical master equation (CME), which characterizes the probability evolution of this dynamical system. Currently, solving this equation in high-dimensional settings is still very challenging. To mitigate this problem, we proposed a new approach, named Rao-Blackwellized CME solver (RB-CME solver), based on modularization and stochastic filtering. Specifically, the RB-CME solver works by an optimized decomposition of the system, which transforms the high-dimensional problem into several lower dimensional ones and then solves each of them using either the Monte-Carlo method or the filtered FSP method. We showed that our method can be remarkably efficient and accurate in solving high-dimensional problems. Compared with the traditional Monte-Carlo method, our approach is far more accurate in estimating the follower system, which we solve using the direct approach, and it is less accurate in estimating the leader system, which we solve using the Monte-Carlo method. Thanks to the well-designed system decomposition, the additional costs for the direct approach are more than compensated by the accuracy benefits, and our method compares favorably to the classical Monte Carlo method for estimating CME solutions. We also developed the method for the stochastic filtering problem (named the Rao-Blackwellized particle filter or the RB-PF) and showed its superior performance over conventional approaches. In RB-PFs, we found that the leader system can also benefit from Rao-Blackwellization, as the conditional probability distribution of the leader interacts with the probability distribution of the follower in the correction step. More interestingly, the RB-CME solver (for solving CMEs) and the RB-PF (for the filtering problem) have relatively the same performance when dealing with the same system. To conclude, our method is a powerful tool to solve CMEs and the associated stochastic filtering problems for high-dimensional chemical reaction systems.

Since the CME plays the central role in the stochastic analysis of biochemical reaction systems, our method can also be generalized to consider other computational problems in biology. For instance, it is worth extending our method for the parameter sensitivity analysis and power spectrum analysis of stochastically reaction systems, which could also be computationally demanding in high-dimensional cases (59–62). Also, our method might be helpful to estimate the probability of rare events, which is difficult to estimate using conventional weighted simulation algorithms (63). Moreover, the divide-and-conquer idea used in our method can also be applied to computing other equations that capture the probability evolution of stochastic systems, e.g., the Fokker-Planck equation.

Finally, our method needs to be further improved in terms of generality and computational efficiency. First, to consider more complicated systems, our method needs to be further developed by including more biological mechanisms, e.g., the time delay, multiscale phenomena, model parameter uncertainty, and cell-cell interactions. Second, it is worth investigating how to speed

up the RB-CME solver, e.g., by applying deep neural networks to the filtered CMEs or establishing a neural network-aided approach based on our method.

Materials and Methods

Mathematical derivations and additional information are provided in *SI Appendix*.

ACKNOWLEDGMENTS. We acknowledge funding from the Swiss National Science Foundation under grant 182653.

1. HH McAdams, A Arkin, Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci.* **94**, 814–819 (1997).
2. A Arkin, J Ross, HH McAdams, Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells. *Genetics* **149**, 1633–1648 (1998).
3. N Fedoroff, W Fontana, Small numbers of big molecules. *Science* **297**, 1129–1131 (2002).
4. DF Anderson, TG Kurtz, *Stochastic analysis of biochemical systems*. (Springer) Vol. 674, (2015).
5. E Korobkova, T Emonet, JM Vilar, TS Shimizu, P Cluzel, From molecular noise to behavioural variability in a single bacterium. *Nature* **428**, 574–578 (2004).
6. S Kar, WT Baumann, MR Paul, JJ Tyson, Exploring the roles of noise in the eukaryotic cell cycle. *Proc. Natl. Acad. Sci.* **106**, 6471–6476 (2009).
7. R Perez-Carrasco, C Beentjes, R Grima, Effects of cell cycle variability on lineage and population measurements of messenger rna abundance. *J. Royal Soc. Interface* **17**, 20200360 (2020).
8. MB Elowitz, S Leibler, A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
9. SK Aoki, et al., A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature* **570**, 533–537 (2019).
10. D Schnoerr, G Sanguinetti, R Grima, Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J. Phys. A: Math. Theor.* **50**, 093001 (2017).
11. Q Jiang, et al., Neural network aided approximation and parameter inference of non-markovian models of gene expression. *Nat. communications* **12**, 1–12 (2021).
12. T Jahnke, W Huisinga, Solving the chemical master equation for monomolecular reaction systems analytically. *J. mathematical biology* **54**, 1–26 (2007).
13. JJ Vastola, Solving the chemical master equation for monomolecular reaction systems and beyond: a doi-peliti path integral view. *J. Math. Biol.* **83**, 1–82 (2021).
14. Z Cao, R Grima, Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proc. Natl. Acad. Sci.* **117**, 4682–4692 (2020).
15. V Shahrezaei, PS Swain, Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci.* **105**, 17256–17261 (2008).
16. DT Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. computational physics* **22**, 403–434 (1976).
17. DT Gillespie, Exact stochastic simulation of coupled chemical reactions. *The journal physical chemistry* **81**, 2340–2361 (1977).
18. DT Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems. *The J. chemical physics* **115**, 1716–1733 (2001).
19. Y Cao, DT Gillespie, LR Petzold, Efficient step size selection for the tau-leaping simulation method. *The J. chemical physics* **124**, 044109 (2006).
20. M Rathinam, LR Petzold, Y Cao, DT Gillespie, Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *The J. Chem. Phys.* **119**, 12784–12794 (2003).
21. EL Haseltine, JB Rawlings, Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *The J. chemical physics* **117**, 6959–6969 (2002).
22. HW Kang, TG Kurtz, Separation of time-scales and model reduction for stochastic reaction networks. *The Annals Appl. Probab.* **23**, 529–583 (2013).
23. B Munsky, M Khammash, The finite state projection algorithm for the solution of the chemical master equation. *The J. chemical physics* **124**, 044104 (2006).
24. NG Van Kampen, *Stochastic processes in physics and chemistry*. (Elsevier) Vol. 1, (1992).
25. L Bronstein, H Koeppl, A variational approach to moment-closure approximations for the kinetics of biomolecular reaction networks. *The J. chemical physics* **148**, 014105 (2018).
26. CA Gomez-Uribe, GC Verghese, Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *The J. chemical physics* **126**, 024109 (2007).
27. I Nåsell, An extension of the moment closure method. *Theor. population biology* **64**, 233–239 (2003).
28. J Hespanha, Moment closure for biochemical networks in *2008 3rd International Symposium on Communications, Control and Signal Processing*. (IEEE), pp. 142–147 (2008).
29. MJ Keeling, Multiplicative moments and measures of persistence in ecology. *J. Theor. Biol.* **205**, 269–281 (2000).
30. A Singh, JP Hespanha, Lognormal moment closures for biochemical reactions in *Proceedings of the 45th IEEE Conference on Decision and Control*. (IEEE), pp. 2063–2068 (2006).
31. P Smadbeck, YN Kaznessis, A closure scheme for chemical master equations. *Proc. Natl. Acad. Sci.* **110**, 14261–14265 (2013).
32. J Ruess, A Milias-Argeitis, S Summers, J Lygeros, Moment estimation for chemically reacting systems by extended kalman filtering. *The J. chemical physics* **135**, 10B621 (2011).
33. V Kazeev, M Khammash, M Nip, C Schwab, Direct solution of the chemical master equation using quantized tensor trains. *PLoS computational biology* **10**, e1003359 (2014).
34. IG Ion, C Wildner, D Loukrezis, H Koeppl, H De Gersem, Tensor-train approximation of the chemical master equation and its application for parameter inference. *The J. Chem. Phys.* **155**, 034102 (2021).
35. A Gupta, C Schwab, M Khammash, Deepcme: A deep learning framework for computing solution statistics of the chemical master equation. *PLoS computational biology* **17**, e1009623 (2021).
36. J Goutsias, Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *The J. chemical physics* **122**, 184102 (2005).
37. Y Cao, DT Gillespie, LR Petzold, The slow-scale stochastic simulation algorithm. *The J. chemical physics* **122**, 014116 (2005).
38. LD EW, E Vanden-Eijnden, Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *J Chem Phys* **123**, 194107 (2005).
39. J Hasenauer, V Wolf, A Kazeroonian, FJ Theis, Method of conditional moments (mcm) for the chemical master equation. *J. mathematical biology* **69**, 687–735 (2014).
40. C Zechner, H Koeppl, Uncoupled analysis of stochastic reaction networks in fluctuating environments. *PLoS computational biology* **10**, e1003942 (2014).
41. L Duso, C Zechner, Selected-node stochastic simulation algorithm. *The J. chemical physics* **148**, 164108 (2018).
42. CR Rao, Information and the accuracy attainable in the estimation of statistical parameters in *Breakthroughs in statistics*. (Springer), pp. 235–247 (1992).
43. D Blackwell, Conditional expectation and unbiased sequential estimation. *The Annals Math. Stat.* pp. 105–110 (1947).
44. G Casella, CP Robert, Rao-blackwellisation of sampling schemes. *Biometrika* **83**, 81–94 (1996).
45. A Doucet, N De Freitas, K Murphy, S Russell, Rao-blackwellised particle filtering for dynamic bayesian networks in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*. pp. 176–183 (2000).
46. HA Blom, EA Bloem, Exact bayesian and particle filtering of stochastic hybrid systems. *IEEE Transactions on Aerosp. Electron. Syst.* **43**, 55–70 (2007).
47. G Grisetti, C Stachniss, W Burgard, Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE transactions on Robotics* **23**, 34–46 (2007).
48. F Gustafsson, et al., Particle filters for positioning, navigation, and tracking. *IEEE Transactions on signal processing* **50**, 425–437 (2002).
49. A Giremus, A Doucet, V Calmettes, JY Tourneret, A rao-blackwellized particle filter for ins/gps integration in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. (IEEE), Vol. 3, pp. iii–964 (2004).
50. C Zechner, M Unger, S Pelet, M Peter, H Koeppl, Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat. methods* **11**, 197–202 (2014).
51. ES D'Ambrosio, Z Fang, A Gupta, M Khammash, Filtered finite state projection method for the analysis and estimation of stochastic biochemical reaction networks. *bioRxiv* (2022).
52. M Rathinam, M Yu, State and parameter estimation from exact partial state observation in stochastic reaction networks. *The J. Chem. Phys.* **154**, 034103 (2021).
53. D Crisan, Particle filters—a theoretical perspective in *Sequential Monte Carlo methods in practice*. (Springer), pp. 17–41 (2001).
54. Z Fang, A Gupta, M Khammash, Stochastic filters based on hybrid approximations of multiscale stochastic reaction networks in *2020 59th IEEE Conference on Decision and Control (CDC)*. (IEEE), pp. 4616–4621 (2020).
55. Z Fang, A Gupta, M Khammash, Stochastic filtering for multiscale stochastic reaction networks based on hybrid approximations. *J. Comput. Phys.* **467**, 111441 (2022).
56. Z Fang, A Gupta, M Khammash, Convergence of regularized particle filters for stochastic reaction networks. *SIAM J. on Numer. Analysis* (to appear).
57. KR Sanft, HG Othmer, Constant-complexity stochastic simulation algorithm with optimal binning. *The J. chemical physics* **143**, 074108 (2015).
58. TS Gardner, CR Cantor, JJ Collins, Construction of a genetic toggle switch in escherichia coli. *Nature* **403**, 339–342 (2000).
59. DF Anderson, An efficient finite difference method for parameter sensitivities of continuous time markov chains. *SIAM J. on Numer. Analysis* **50**, 2237–2258 (2012).
60. P Dürrenberger, A Gupta, M Khammash, A finite state projection method for steady-state sensitivity analysis of stochastic reaction networks. *The J. Chem. Phys.* **150**, 134101 (2019).
61. A Gupta, M Khammash, Unbiased estimation of parameter sensitivities for stochastic chemical reaction networks. *SIAM J. on Sci. Comput.* **35**, A2598–A2620 (2013).
62. A Gupta, M Khammash, Frequency spectra and the color of cellular noise. *Nat. Commun.* **13**, 1–18 (2022).
63. PJ Thomas, M Ahmadi, H Zheng, C Myers, A comparison of weighted stochastic simulation methods. *bioRxiv* (2021).