

Chemical master equation parameter exploration using DMRG

John P. Zima, Schuyler B. Nicholson, and Todd R. Gingrich^{a)}

Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA

Well-mixed chemical reaction networks (CRNs) contain many distinct chemical species with copy numbers that fluctuate in correlated ways. While those correlations are typically monitored via Monte Carlo sampling of stochastic trajectories, there is interest in systematically approximating the joint distribution over the exponentially large number of possible microstates using tensor networks or tensor trains. We exploit the tensor network strategy to determine when the steady state of a seven-species gene toggle switch CRN model supports bistability as a function of two decomposition rates, both parameters of the kinetic model. We highlight how the tensor network solution captures the effects of stochastic fluctuations, going beyond mean field and indeed deviating meaningfully from a mean-field analysis. The work furthermore develops and demonstrates several technical advances that will allow steady-states of broad classes of CRNs to be computed in a manner conducive to parameter exploration. We show that the steady-state distributions can be computed via the ordinary density matrix renormalization group (DMRG) algorithm despite having a non-Hermitian rate operator with a small spectral gap, we illustrate how that steady-state distribution can be efficiently projected to an order parameter that identifies bimodality, and we employ excited-state DMRG to calculate a relaxation timescale for the bistability.

I. INTRODUCTION

Markovian master equations appear in diverse disciplines¹. Though the models are ubiquitous, closed-form analytical solutions are rare and limited to very simplified models. Complex systems therefore often require Monte Carlo sampling of representative stochastic trajectories. Unfortunately, naïve sampling methods like the Gillespie algorithm² can suffer from slow convergence, particularly when used to compute properties of systems with a separation of timescales between frequent and rare events. In that case, for a rare event to be observed, orders of magnitude more frequent events must first be sampled. Generating an outsized sampling of frequent events can be prohibitively costly, especially because the typical time steps taken by the Gillespie algorithm scale inversely with system size.

The computational expense is dramatically amplified when one wants to solve inverse problems. Then one hopes to explore the forward problem—given a model and its rate parameters, what stochastic dynamics emerges—for a large number of parameters^{1,3}, demanding ways to efficiently extract the sensitivity to the parameters⁴. Ideally, one should leverage the computational effort invested for one parameter set to more cheaply study the subsequent parameter sets, but brute-force Gillespie sampling does not easily lend itself to cost savings. Reweighting samples from one set of system parameters to provide insight about a different but similar set of parameters costs essentially the same effort as re-sampling from scratch.

When the forward problem features rapid but rare events, there are enhanced sampling techniques that focus on slow-timescale transitions between different

metastable regions of configuration space. Transition path sampling (TPS)^{5–7} and forward flux sampling (FFS)^{8–10} can both generate representative trajectories of those transitions, but both classes of methods bring other complications. TPS requires an initial transition path and a sometimes fickle set of Monte Carlo moves to hop from one reactive trajectory to another; FFS requires identification of a “good” low-dimensional reaction coordinate.

Given the challenges, it is worthwhile to reconsider whether sampling approaches are truly necessary. Could one bypass the realizations and instead compute the full steady-state distribution? Initially, this idea seems impractical. The number of microstates scales exponentially with system size, and calculations involving the steady-state distribution require that one compute the probability for each of those exponentially many microstates. Suppose, however, that one sought instead a controllable approximation to that distribution. Due to the modern success of machine learning, one might look to neural networks as the natural approximation tool. Perhaps one could use stochastic gradient descent to fit a billion- or trillion-parameter nonlinear function to a large number of sampled trajectories, for example.

As we detail in this work, we can achieve a parameterized, controllable approximation to the distribution while avoiding both trajectory sampling and stochastic gradient descent. However, the key is not the use of a neural network. Rather, we fit a high-dimensional *multilinear* function using a *non-gradient* optimization procedure that draws directly from the rules of the stochastic dynamics (the rates in the master equation), not from sampled data. The multilinear functions to be fitted are graphically represented as tensor networks^{11,12}, most often a one-dimensional chain of contracted tensors known in physics communities as a matrix product state (MPS) and in the applied math community as a tensor train (TT)¹³. Parameterization of the distribution

^{a)}Electronic mail: todd.gingrich@northwestern.edu.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0276591

boils down to fitting the elements of all tensors in the chain, an optimization problem that can be performed by sweeping through the tensors and optimizing a single tensor at a time. The core tensor network ideas and optimization strategies are very well developed and broadly known in the quantum many-body^{11,12,14–16}, quantum chemistry^{17–21}, and numerical linear algebra^{13,22–25} communities. They have begun to be used for classical master equations^{26–37}, including applications for chemical master equations (CMEs)^{25,38–44}. An advantage of these techniques over sampling-based approaches lies in the fact that once the distribution has been approximated for one parameter set, that approximated distribution can be used as a very effective seed when studying parameters with a similar steady-state distribution.

Gelß et al. leveraged this advantage when studying a heterogeneous catalysis model of CO oxidation on a RuO₂(110) surface^{41,45}. As the master equation stiffened, rare events became rarer and the computational cost of kinetic Monte Carlo (Gillespie) scaled worse than the cost of converging an MPS approximation using the so-called Alternating Linear Scheme (ALS)⁴⁵. Liao et al. have made significant strides in applying this class of methods to biochemical kinetics, tackling multiple chemical reaction networks, including a cell cycle model⁴⁰. Their approach uses tensor networks to solve the chemical Fokker-Planck equation (CFPE) that arises from truncating the Kramers-Moyal expansion of the CME at second order. They discretized the problem on a grid and approximated the steady state by converging an MPS with the Alternating Minimal Energy Method (AMEn)²⁴. The procedure is flexible in that it can be performed with different discretized grids, but passing from the CME to the CFPE is problematic in the low-copy-number regime if the particular CME exhibits “discreteness-induced multimodality”^{46,47}. That low-copy-number regime is essential for some kinetic models of biological processes. Take, for instance, models of a gene toggle switch (GTS)⁴⁸ with fluctuating amounts of proteins A and B, each suppressing the activation of the gene that produces the other protein. With many copies of DNA, proteins A and B coexist since, at any given time, some DNA molecules would be in an A-producing state and others in a B-producing state. Inside a single cell, however, one cannot ignore the discreteness of the DNA, which could cause the cell to flip between A- and B-rich states.

To handle both small- and large-copy-number regimes, we build the tensor network approach directly from the CME using the Doi-Peliti formalism^{44,49,50}. This formalism, a classical analog of second quantization methods, allows a set of elementary reactions to be written in terms of raising and lowering operators that act on each species. By writing each reaction of the GTS model in this second-quantized form, the rate operator that generates the CME dynamics can be numerically constructed as a matrix product operator (MPO) using standard software packages^{51,52}. The steady-state distribution over

microstates is the top eigenvector of that MPO. In other words, solving for the steady-state distribution is analogous to using the Density Matrix Renormalization Group (DMRG) algorithm to find the ground state of a Hamiltonian. Unlike in the usual quantum mechanical context, the rate operator MPO is typically not Hermitian, but we appeal to the Time Dependent Variational Principle (TDVP) to rationalize why the ordinary DMRG algorithm can nevertheless compute the steady state. That steady-state distribution can be easily marginalized to obtain a distribution for an order parameter that detects bimodality, allowing for the efficient construction of a kinetic phase diagram.

The paper is structured as follows. Section II introduces chemical reaction networks and the chemical master equation, specializing in the case of the exclusive gene toggle switch model of Ref.⁴⁸. Section III reviews the tensor network tools, both the MPS used for the distribution and the MPO used for the rate operator. Section IV discusses the DMRG algorithm, highlighting why standard DMRG algorithms can target the steady-state distribution for CMEs even though the time propagator is not Hermitian. Section V shows how to use the tensor network tools to assess the bimodality and bistability of the CME. Section VI presents our main numerical results, demonstrating how reseeded DRMG calculations can help explore the parameter space. We close with a discussion of the outlook and challenges.

II. CHEMICAL REACTION NETWORKS AND THE GENE TOGGLE SWITCH MODEL

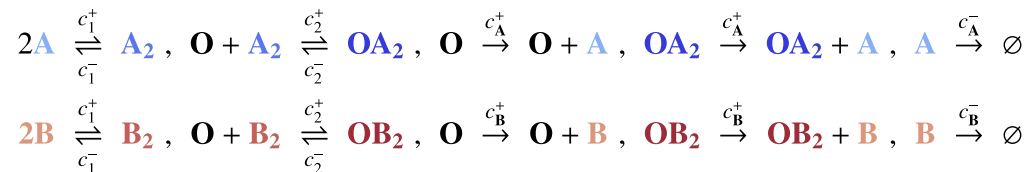
In this paper, we focus on well-mixed chemical reaction networks (CRNs), which are comprised of a set of N species and K elementary reactions. This paper is organized around a specific example CRN, the GTS model. In that example, the 7 species and 14 reactions are depicted explicitly in Fig. 1. CRNs like the GTS model become rich and interesting because the outputs of some elementary reactions are inputs for others, coupling together the elementary steps into a complex kinetic network. The dynamics of the network is monitored by tracking the microstate of the CRN, specified by an N -element vector \mathbf{n} whose elements give the copy number of each species. For GTS, $\mathbf{n} = (n_A, n_{A_2}, n_{OA_2}, n_O, n_{OB_2}, n_{B_2}, n_B)$. The action of elementary reaction k is encoded in an N -element vector $\delta\mathbf{n}_k$ that reports the change in copy number for each species when reaction k fires. Let reaction 1 be $2A \rightarrow A_2$ as in Fig. 1. Then $\delta\mathbf{n}_1 = (-2, 1, 0, 0, 0, 0, 0)$. The propensity for a particular reaction to fire is a product of two terms, a combinatorial term counting the number of unique sets of reaction k 's reactants and a rate for any one of those sets to react. This propensity $w_k(\mathbf{n})$ depends on the microstate, since the combinatorial factor depends on \mathbf{n} . The probability of occupying that microstate at time t evolves according to the chemical

Stochastic Chemical Reaction Networks (CRNs)

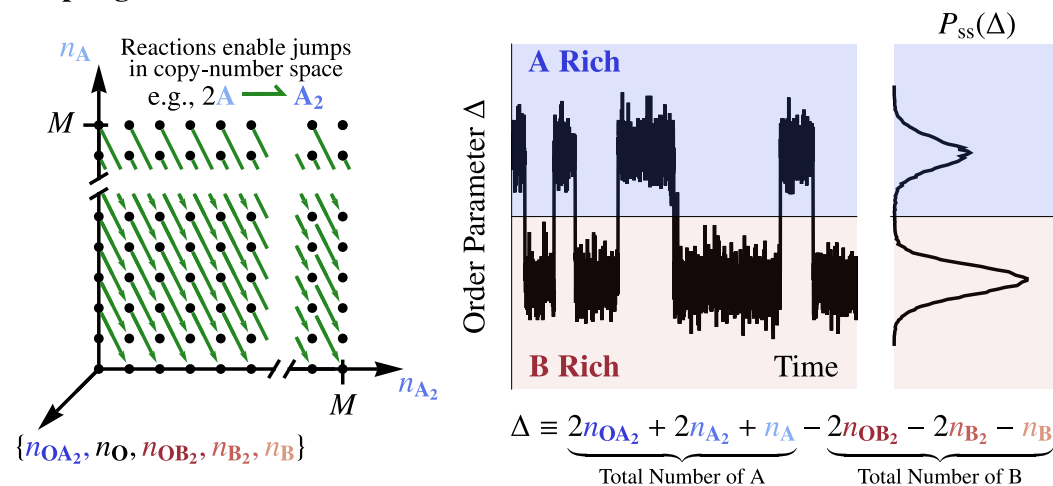
7 Chemical Species for the Gene Toggle Switch Model:

Protein A Species: $\underbrace{A, A_2, OA_2, O}_{\text{Bound and Unbound DNA Species}}$
 Protein B Species: $\underbrace{OB_2, B_2, B}_{\text{Bound and Unbound DNA Species}}$

Set of 14 Elementary Reactions and Associated Rates:



Sampling Realizations



Computing Distributions

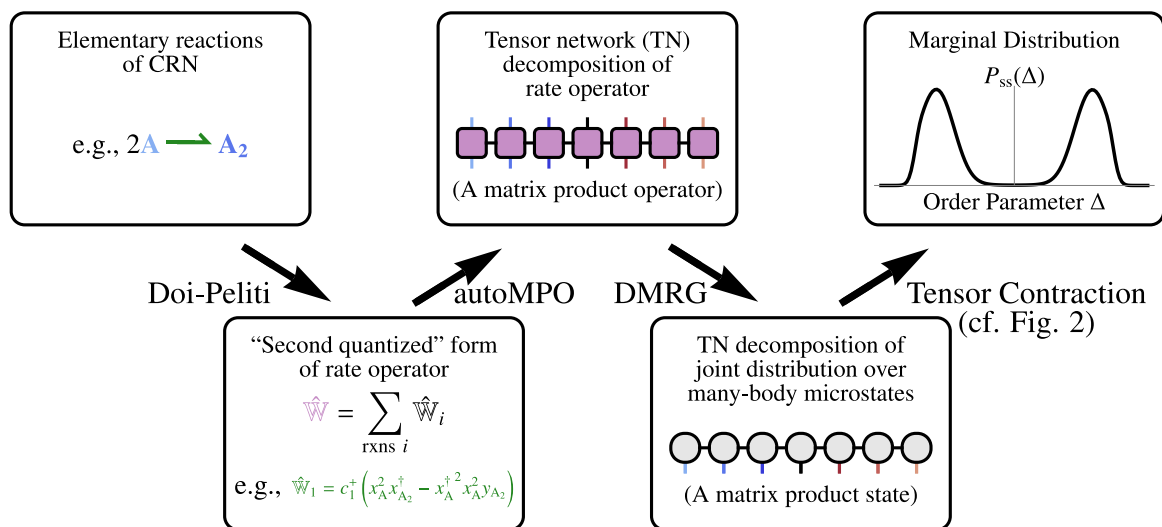


FIG. 1. Two approaches to stochastic chemical kinetics, illustrated with the gene toggle switch (GTS) model. **(Top)** The 14 elementary reactions of the model couple together fluctuations in the copy number of the seven chemical species. The model is especially notable for generating a bistable switch. **(Middle)** Traditionally, the time-evolution for those copy-number fluctuations is sampled by a kinetic Monte Carlo (Gillespie) trajectory that jumps between a set of microstates too numerous to practically enumerate. With sufficiently long trajectories, one can build converged histograms for low-dimensional order parameters like Δ to sample $P_{ss}(\Delta)$. **(Bottom)** This manuscript describes an approach to numerically compute those same distributions without sampling. Rather, the elementary reactions are converted into a raising and lowering operator form (Doi-Peliti), allowing the chemical master equation to be expressed in terms of a tensor network operator. The steady-state distribution can be controllably approximated as a matrix product state.

master equation (CME):

$$\frac{dp(\mathbf{n}, t)}{dt} = \sum_{k=1}^K [w_k(\mathbf{n} - \delta\mathbf{n}_k)p(\mathbf{n} - \delta\mathbf{n}_k, t) - w_k(\mathbf{n})p(\mathbf{n}, t)], \quad (1)$$

with the first term of the summand giving the rate of gaining probability in microstate \mathbf{n} from reaction k and the second term giving the rate of losing probability from \mathbf{n} due to that same elementary reaction.

By choosing an ordering of microstates, Eq. (1) can be cast in matrix form. We combine the probabilities $p(\mathbf{n}, t)$ into a vector $|p(t)\rangle$ with component i corresponding to the probability of the i^{th} microstate at time t , and we define the rate matrix \mathbb{W} to have off-diagonal elements \mathbb{W}_{ij} that sum w_k over all reactions k which transform microstate j into i . The off-diagonal elements of \mathbb{W} capture the “gain” terms of Eq. (1). The negative “loss” terms are accounted for by setting diagonal terms such that the columns sum to zero: $\mathbb{W}_{ii} = -\sum_{j \neq i} \mathbb{W}_{ji}$, a condition that ensures conservation of probability. Eq. (1) then becomes simply

$$\frac{d|p(t)\rangle}{dt} = \mathbb{W}|p(t)\rangle. \quad (2)$$

This master equation is formally solved by $|p(t)\rangle = e^{\mathbb{W}t}|p(0)\rangle$. Under the relatively weak condition of irreducibility (all microstates can be reached by mixing together a sequence of elementary reactions), the long-time steady-state probability distribution is given by the top right eigenvector of \mathbb{W} . The simplicity of the formal solutions to the master equation hides the complexity brought by the particular structure of \mathbb{W} and the need to compute spectral properties of that very large matrix.

We emphasize this challenge with our focus on the GTS of Warren and ten Wolde⁴⁸, chosen as the illustrative CRN for three reasons. Firstly, the model is well understood, having been studied using mean field, Gillespie, and even advanced sampling procedures⁸. Secondly, while the model’s state space is too large to treat directly with dense linear algebra, it sits at the boundary of what is possible using sparse matrix methods⁵³, offering a useful comparison when assessing errors (see Appendix D). Thirdly, the model supports a regime where the discreteness of the species’ copy numbers is materially important, making GTS a challenging test of methodology.

The GTS model comprises seven species and fourteen reactions, as shown in Fig. 1. Two of those species, A and B, represent proteins which can be synthesized by DNA and which decompose through different elementary reactions with rates c_A^- and c_B^- . A defining characteristic of the GTS model is that both proteins suppress the production of the other. That production requires DNA, which exists in one of three forms: unbound (species O), bound to a dimer of protein A (species OA_2), or bound to a dimer of protein B (species OB_2). All three forms can synthesize protein. Although O can synthesize both A and B, OA_2 can only synthesize A, and OB_2 can only

synthesize B. The final two species are the protein dimers A_2 and B_2 , which reversibly bind to the DNA and which reversibly form from the respective monomers. These rules and the associated rate constants of Fig. 1 are sufficient to prescribe the form of \mathbb{W} , but that does not mean that it is trivial to practically construct the matrix or compute the steady state.

The core problem is the curse of dimensionality. A microstate of the model is given by the copy numbers $(n_A, n_{A_2}, n_{OA_2}, n_O, n_{OB_2}, n_{B_2}, n_B)$. Each copy number must be a nonnegative counting number, so microstates are lattice points in a seven-dimensional space. Conservation laws in CRNs can slightly reduce the size of that space. For example, none of the 14 elementary reactions creates or destroys DNA; they only interconvert between its three forms (O, OA_2 , and OB_2). As a consequence, the dynamics is only irreducible within a manifold of states with a conserved value of $n_{\text{DNA}} \equiv n_{OA_2} + n_O + n_{OB_2}$. In this paper, we study the $n_{\text{DNA}} = 1$, single-DNA regime, which means that microstates span a five-dimensional space. That single DNA exists in one of three states, but the copy numbers of A, A_2 , B, and B_2 can all vary from 0 to ∞ . Gillespie sampling executes a random walk on that infinite set of copy numbers, randomly drawing a next reaction and following that reaction to get to the next lattice point (see Fig. 1). Though the random walk occurs in the high-dimensional space of all copy numbers, the trajectory can be projected down onto a lower dimensional order parameter. For GTS, it is natural to look at the evolution of $\Delta \equiv 2n_{OA_2} + 2n_{A_2} + n_A - 2n_{OB_2} - 2n_{B_2} - n_B$, which counts the excess of protein A relative to protein B. Fig. 1 shows a representative Gillespie trajectory that illustrates the toggle-switch nature of the GTS model; the system switches between two metastable macrostates, an A-rich state characterized by positive Δ and a B-rich state with negative Δ . Rare, stochastic fluctuations trigger the flips between macrostates. With enough sampling, the steady-state distribution for the order parameter, $P_{\text{ss}}(\Delta)$, can be constructed as a histogram, but many of the rare flips would need to be observed to converge.

An alternative way to study the GTS is to numerically compute the steady-state distribution from the CME without sampling. Doing so clearly requires a truncation to allow only a finite number of microstates. By disallowing copy numbers for A, A_2 , B, and B_2 from growing larger than M , we limit the number of microstates (and thus the dimension of the rate matrix \mathbb{W}) to $3(M+1)^4$. For the elementary-reaction rate constants studied in this work, the protein degradation rates c_A^- and c_B^- were sufficiently large that $M = 20$ was a reasonable truncation. Incorporating this truncation and the $n_{\text{DNA}} = 1$ conservation law, numerically solving for $|\pi\rangle \equiv \lim_{t \rightarrow \infty} |p(t)\rangle$ amounts to finding the largest right eigenvector of a sparse \mathbb{W} with dimension 583,433. That problem is accessible to sparse linear algebra methods like the generalized minimal residual method (GMRES), but it is easy to anticipate how increases in M or the number of

of dimensionality, but we can build a more tractable order seven tensor from the contraction of smaller-order tensors, one per species:

$$\approx \text{---} \quad (5)$$

This one-dimensional chain of tensors is a factorization into a matrix product state (MPS) or tensor train (TT). For a sufficiently large bond dimension, the steady-state distributions could always be written in this factorized form, but the bond dimension would need to grow very large. The computational advantage emerges because real problems appear to involve steady-state distributions which can be very well approximated by an MPS with a modest bond dimension. In the quantum mechanical context, this ability to compress is well justified in terms of the so-called area laws^{11,54–57}. In the CME context, we are unaware of comparatively satisfactory formal justifications for the observed ability to massively truncate the bond dimension, but our GTS calculations in this work required a maximum bond dimension no greater than 100 (see also Appendix D). Before discussing those results, we must address how, for a particular choice of bond dimension, we can fit the tensor elements of each MPS tensor to best approximate $|\pi\rangle$. Doing so requires that we first demonstrate how to systematically convert the CME's elementary reactions into a TN form.

B. Rate Operators in a Matrix Product Form

Elements of the rate matrix \mathbb{W} are specified by both the initial microstate \mathbf{n} and the final microstate \mathbf{n}' . In our TN notation, a rate operator thus takes the form:

$$\hat{\mathbb{W}} = \text{---} \quad (6)$$

This $\hat{\mathbb{W}}$ is very high-order, so it is appealing to try to factorize it into lower-order tensors, as we did with the

approximation for the steady-state distribution:

$$\approx \text{---} \quad (7)$$

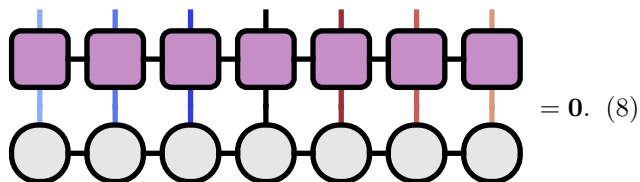
The factorized form is a matrix product operator (MPO). Each of the rounded rectangles acts on the physical copy number space (colored legs) of a single species, tracking the rates for increasing or decreasing the copy number of that species. But the species do not change in isolation.

Take, for example, our first reaction, $2A \xrightarrow{c_1^+} A_2$. This reaction will simultaneously decrease n_A by two while increasing n_{A_2} by one. This correlation is naturally expressed in a second-quantized representation in terms of a raising operator x_l^\dagger that increases the copy number of species l by one and a lowering operator x_l that lowers it by one. For example, the “gain” term of the rate operator is simply $c_1^+ x_A^2 x_{A_2}^\dagger$, which acts on the two species in conjunction. The Doi-Peliti formalism systematically converts elementary reactions into a contribution to the second-quantized rate operator, including both the “gain” terms and the “loss” terms^{49,50}. The second-quantized version of the GTS model is discussed further in Appendix A of this work and Appendix C of Ref.⁵³.

The black horizontal bond between the first two tensors in the factorization of Eq. (7) allows us to build an operator that correlates the action on the two different species. In the case of reaction 1, the correlations are built into a single horizontal bond between neighboring tensors, but the bond dimension only needs to grow linearly to pass correlations down the chain. Consequently, the TN representation of the rate operator $\hat{\mathbb{W}}$ can be very efficient from the perspective of both the memory needed to store the operator and the computational effort to perform operations on it. When we first introduced the rate matrix \mathbb{W} , we conceived of it as a very large matrix with dimension equal to the number of microstates. Even for a sparse matrix, storing that object in memory grows problematic. With the factorization, we see that the rate operator $\hat{\mathbb{W}}$ could be alternatively expressed by storing the values of seven factorized tensors that have order 3 or 4. Building those tensors from the second-quantized operator is essentially a solved problem⁵⁸ that is implemented in tensor network software packages. Our calculations used the autoMPO function of the ITensor software package to turn Eq. (A9) into the MPO^{51,52}.

IV. STEADY-STATE DISTRIBUTIONS VIA DMRG

Section III showed how to construct the seven tensors of the MPO from elementary reactions and their associated rates. The section furthermore introduced the idea that an MPS factorization could approximate a distribution over microstates. It remains to compute the values of those MPS tensors which approximate a particular distribution, the steady-state. The key is to recognize that for the irreducible CME, the steady state is the unique largest eigenvector of $\hat{\mathbb{W}}$, which must have eigenvalue zero. In terms of the TNs, we therefore need to solve for the values of the circular tensors to yield:



$$= \mathbf{0}. \quad (8)$$

where $\mathbf{0}$ represents an order-7 tensor of the appropriate dimensions in which each entry is zero. This equation can be viewed as an optimization problem, where the optimized circular tensors will approximate π .

It is computationally taxing to perform operations on order-7 tensors like Eq. (4), but Eq. (5)'s factorization offers a smarter route. We want to perform operations on single-species circular tensors, updating them sequentially to solve the global optimization problem of Eq. (8). In the quantum mechanical context, the Density Matrix Renormalization Group (DMRG) algorithms are popular and effective methods to find the extremal eigenvector (ground state) of quantum many-body systems by sweeping through the chain performing optimizations on factorized tensors^{14–16}. It is tempting to simply apply the same algorithms to the CME, but we must remember that the CME is typically non-Hermitian whereas usual quantum mechanical problems seek the ground state of a Hermitian Hamiltonian.

A. Ground states of non-Hermitian operators

There has been some activity trying to utilize DMRG for non-Hermitian systems. One approach is to construct a Gram matrix $\hat{G} = \hat{\mathbb{W}}^\dagger \hat{\mathbb{W}}$, where † here denotes a conjugate transpose. \hat{G} and $\hat{\mathbb{W}}$ share the same right zero eigenvector, but \hat{G} is Hermitian even if $\hat{\mathbb{W}}$ is non-Hermitian^{59,60}. Applying the usual DMRG algorithms to \hat{G} therefore offers a valid route to construct π for the non-Hermitian GTS, but there is a drawback. The convergence rate of DMRG is sensitive to the spectral gap between the first and second eigenvalues. The closer the gap is to zero, the more problematic DMRG convergence becomes. The eigenvalues of \hat{G} are not those of $\hat{\mathbb{W}}$ but instead are the singular values squared of $\hat{\mathbb{W}}$. Then collections of small singular value gaps for $\hat{\mathbb{W}}$ can lead to even

smaller gaps in \hat{G} ⁵⁹. Another naïve method of restoring Hermiticity would be to use the operator $\hat{\mathbb{W}} + \hat{\mathbb{W}}^\dagger$ instead of $\hat{\mathbb{W}}$. However, this operator will no longer have the same ground state as $\hat{\mathbb{W}}$, thus requiring a more complicated cost function to restore the existence of a variational principle⁶¹. A third route is to develop alternative DMRG algorithms that are specifically built for non-Hermitian operators^{62–65}. This route, which holds long-term promise, cannot immediately utilize all of the software tools already developed for Hermitian DMRG.

There are, however, reports of success simply applying the usual DMRG procedure to the non-Hermitian operator^{66–68}. We have similarly found that we can compute π for our GTS model by applying a naïve DMRG scheme to our non-Hermitian $\hat{\mathbb{W}}$. One might anticipate that DMRG would remain stable if there is a small parameter introducing a non-Hermitian perturbation on top of a Hermitian operator. In our case, one cannot argue that $\hat{\mathbb{W}}$ is particularly close to Hermitian, so we sought a justification to understand why the convergence works in practice.

B. Rayleigh-Ritz justification for DMRG

The usual justification for the DMRG algorithm originates with the Rayleigh-Ritz variational principle. An MPS with a modest bond dimension cannot exactly represent the ground state in general, but the MPS could nevertheless be expanded as a superposition of the exact right eigenvectors $|\phi_i\rangle$: $|\Phi_{\text{MPS}}\rangle = \sum_i \alpha_i |\phi_i\rangle$. The conjugate transpose of that MPS can also be expanded as a superposition of the exact left eigenvectors $\langle\phi_i|$: $\langle\Phi_{\text{MPS}}| = \sum_i \beta_i \langle\phi_i|$. When $\hat{\mathbb{W}}$ is non-Hermitian, α_i and β_i are not simply related to each other, but when $\hat{\mathbb{W}}$ is Hermitian, they are complex conjugates, $\alpha_i = \beta_i^*$. For a Hermitian rate operator $\hat{\mathbb{W}}$, the Rayleigh-Ritz quotient is a weighted average of $\hat{\mathbb{W}}$'s eigenvalues, λ_i , so the quotient must fall between the extremal eigenvalues λ_{\min} and $\lambda_{\max} = 0$:

$$\lambda_{\min} \leq \frac{\sum_i |\alpha_i|^2 \lambda_i}{\sum_i |\alpha_i|^2} \leq 0. \quad (9)$$

For the Rayleigh-Ritz quotient to approach 0, $|\Phi_{\text{MPS}}\rangle$ must approach a state with α_i equal to 1 for the ground state and 0 for all other eigenstates. That $|\Phi_{\text{MPS}}\rangle$ is parameterized by many tensor elements defining the tensors of each single-species tensor of the MPS. The Rayleigh-Ritz principle tells us that a perturbation to each of those single-species tensors which succeeds in moving the Rayleigh-Ritz quotient closer to zero will have also moved $|\Phi_{\text{MPS}}\rangle$ closer to the true ground state.

There are many flavors of DMRG algorithm that iteratively pass through those MPS tensors to perform perturbations that nudge the Rayleigh-Ritz quotient upward toward zero. For example, one can attempt to optimize a single-species tensor (one grey circle) while hold-

ing the other six single-species tensors fixed. The six fixed tensors can be contracted with the MPO to produce a new effective operator that acts on the gray circle to be optimized. Optimization of the single-species tensor is performed efficiently as an Arnoldi⁶⁹ maximum eigenvector calculation of that low-rank effective operator. The DMRG algorithm then moves to the next gray circle down the chain, optimizing it subject to all the others being fixed. These sweeps through the MPS tend to increase the Rayleigh-Ritz quotient, which therefore pushes $|\Phi_{\text{MPS}}\rangle$ closer to the ground state. Many more advanced DMRG algorithms exist, some of which incorporate noise⁷⁰, subspace expansion⁷¹, or optimizations of more than one tensor at a time^{14–16}. For example, our results utilize a two-site DMRG algorithm that sweeps through the MPS doing local maximal eigenvector calculations on the joint space given by two neighboring gray circles.

C. Time evolution justification for DMRG

Making \hat{W} non-Hermitian does not fundamentally disrupt the idea that the global optimization of Eq. (8) can be decomposed into a sequence of local optimizations, but it does fundamentally alter the Rayleigh-Ritz inequality of Eq. (9). Without the relationship between α_i and β_i , the Rayleigh-Ritz quotient is

$$\frac{\sum_i \alpha_i \beta_i \lambda_i}{\sum_i \alpha_i \beta_i}, \quad (10)$$

and $\alpha_i \beta_i$ may no longer be a positive real number. As a consequence, the Rayleigh quotient might now rise above $\lambda_{\text{max}} = 0$, meaning perturbations to $|\Phi_{\text{MPS}}\rangle$ which yield a Rayleigh-Ritz quotient closer to zero might not correspond to states which are closer to the ground state. One can perform the same DMRG sweeps—construct an effective operator for a one- or two-site tensor, perform a maximum eigenvector calculation, sweep to the next sites, and iterate—but without the assurances of the Rayleigh-Ritz principle it is unclear if such an algorithm actually targets the ground state for a non-Hermitian operator.

Here, we provide a different argument that supports that naïve application of DMRG to our particular non-Hermitian problems. Specifically, we note that the Perron-Frobenius theorem implies that any probability distribution over microstates tends to the unique $|\pi\rangle$ in the long-time limit, that is $|\pi\rangle = \lim_{t \rightarrow \infty} e^{\hat{W}t} |p(0)\rangle$. In essence, this long-time propagation is the solution of the maximal eigenvector by the power method. The propagation for infinite time can be thought of as an infinite number of discrete timesteps of length Δt ,

$$|\pi\rangle = e^{\hat{W}\Delta t} \dots e^{\hat{W}\Delta t} |p(0)\rangle. \quad (11)$$

For sufficiently small Δt , the tensor network implementation of the time-dependent variational principle (TN TDVP) gives a controllable approximation for the time

evolution from one MPS into another. If the initial $|p(0)\rangle$ is itself an MPS, that TN TDVP algorithm constructs another MPS for the time-propagated state $|p(\Delta t)\rangle$. In lieu of the DMRG procedure, one could therefore find $|\pi\rangle$ by the power method through the sequential application of the TN TDVP algorithm with a sufficiently small Δt ⁴⁴. This TN TDVP solution for the ground state may require many timesteps, but even without the Rayleigh-Ritz method, Perron-Frobenius guarantees that $|\pi\rangle$ will eventually be reached.

To make TN TDVP reach the steady-state distribution faster, one could try simply increasing Δt , but in a loose sense, this will cause some MPS tensors to “get out ahead of the others”. TN TDVP requires matrix exponentials for time Δt using an effective rate operator that acts on one or two sites. One can think of those steps of the algorithm as propagating the distribution for some sites to time $t + \Delta t$ while leaving the others behind at time t . Once the full sweep has been completed, all tensors have been advanced to $t + \Delta t$. Within a sweep, some tensors capture behavior at time t while others have already advanced to $t + \Delta t$. For finite Δt , a timestep error emerges because an individual tensor is propagated with an effective rate operator built from a mixture of time t and time $t + \Delta t$ tensors. Accurate evolution of the distribution requires the small Δt limit.

Reaching the steady-state distribution, however, does not require accurate evolution. One could update $|p\rangle$ in a manner that has significant timestep errors but nevertheless approaches the same fixed point. We note that for our master equation problems, the fixed point of a TN TDVP sweep does not depend on the timestep Δt . To see this fact, imagine that we first evolved an initial distribution $|p(0)\rangle$ using TN TDVP with a very small Δt to approximate $|\pi\rangle$ by the power method. We could take the final MPS and apply one more step of TN TDVP with a very large Δt timestep. The old concern about timestep error ceases to matter when all of the site tensors have already reached their steady-state ($t \rightarrow \infty$) values. Now, each site tensor will already be a zero eigenvector for its effective rate operator, meaning the time propagation by Δt will not change the site tensor. In other words, once an MPS representation of $|\pi\rangle$ is found, further time evolution *with any choice of Δt* will stick at $|\pi\rangle$.

This observation justifies running TN TDVP steps with large timesteps. The path to a converged MPS may be less smooth than running TN TDVP with a small timestep, but a large timestep means the MPS can be significantly changed in a single sweep. Some site tensors could get way out ahead of the others, leading to a thrashing as the site tensors are each propagated based on outdated values of the other tensors. That thrashing could shrink a radius of convergence, but if we can cleverly seed each calculation to fall within some convergence region, we can be confident that the final stationary state of the TN TDVP algorithm is $|\pi\rangle$.

Procedurally, the $\Delta t \rightarrow \infty$ limit of a TDVP timestep is particularly nice. A TDVP timestep ordinarily involves

matrix exponentials of an effective rate operator acting for time Δt . In the long-time limit, that matrix exponential is simply related to the maximal eigenvalue and the associated eigenvector. Indeed, Haegeman et al. have noted that the DMRG algorithm can be understood as a special case of the TDVP algorithm in the infinite-timestep limit⁷². The same DMRG algorithm that targets a Hermitian ground state by a Rayleigh-Ritz justification therefore targets the stationary state $|\pi\rangle$ of our non-Hermitian rate operator \hat{W} ⁷³. Convergence from an arbitrary starting distribution $|p(0)\rangle$ is of course not certain given the thrashing, but it is reasonable to hope that there exists a radius of convergence of a non-negligible size. The numerical experiments which constitute the main results of this paper confirm this optimistic notion.

V. ASSESSING BISTABILITY

We know that the GTS model can exhibit bistability for certain values of the elementary rates in the CRN. There are two different ways to extract information about that bistability from the DMRG calculations. The first route, which directly utilizes the steady-state distribution $|\pi\rangle$, is a static characterization of the bimodality of the stationary distribution. The second route, which requires excited-state DMRG calculations, more directly probes the timescale of dynamics.

Let us start by directly using $|\pi\rangle$, found as the DMRG ground state. The order parameter Δ sorts microstates based on how many excess A proteins there are compared to B, so the marginal distribution $P_{\text{ss}}(\Delta)$ shows whether $|\pi\rangle$ is dominated by A-rich states, B-rich states, or some mixture. Building $P_{\text{ss}}(\Delta)$ from $|\pi\rangle$ is a relatively straightforward marginalization, one which is particularly easy to compute using tensor networks. Figure 2 shows how sparse tensors can be constructed from Kronecker deltas to contract over the various species involving protein A, transforming to a new tensor that is indexed by the total number of A, n_A . An equivalent Kronecker delta tensor characterizes the microstates by the total number of B, n_B . A final Kronecker delta tensor, $\delta_{\Delta, n_A - n_B}$, combines all mixtures of n_A and n_B that produce identical values of Δ . The contraction scheme of Fig. 2 builds $P(\Delta)$ from any $|p\rangle$, but when applied to $|\pi\rangle$ it gives the stationary $P_{\text{ss}}(\Delta)$. The bimodality of that $P_{\text{ss}}(\Delta)$ can be assessed from the skewness γ and kurtosis κ in terms of Sarle's bimodality coefficient⁷⁴

$$\text{BC} = \frac{\gamma^2 + 1}{\kappa}. \quad (12)$$

By construction, this BC ranges from 0 to 1. Values that exceed the BC of a uniform distribution (5/9) typically indicate a bimodal or multimodal distribution⁷⁴. In Sec. VI, we numerically show how that BC varies as a function of elementary GTS rates, building the kinetic phase diagram in Fig. 3.



FIG. 2. DMRG returns a tensor network which assigns a probability to each microstate (gray MPS). Summing over those microstates to get the marginal distribution of the order parameter Δ requires summing over all microstates that yield the particular value of Δ . As n_O does not influence Δ , that coordinate is traced over (contracted with the identity vector $\mathbf{1}$). Sums over the occupation numbers of the other species are efficiently realized by building a tensor network of Kronecker delta tensors. Upon contraction, one is left with an order-one tensor (brown circle) whose index ranges over the allowed values of Δ and whose elements are the marginal steady-state probabilities.

The particular shape of $P_{\text{ss}}(\Delta)$ cannot simply yield the timescale for switching between A-rich and B-rich states, but bimodality is a good proxy for bistability. A more direct measure of the bistability requires dynamical timescales, which come from the excited-state eigenvalues of \hat{W} . Let us order the eigenvalues of \hat{W} with $0 = \lambda_0 > \lambda_1 > \lambda_2 \dots$ and with $|\pi\rangle \equiv |\phi_0\rangle, |\phi_1\rangle, |\phi_2\rangle, \dots$ being the associated right eigenvectors. An initial probability distribution can be expanded in the eigenbasis $|p(0)\rangle = \sum_i \alpha_i |\phi_i\rangle$. Time evolution is simple in the eigenbasis:

$$|p(t)\rangle = e^{\hat{W}t} |p(0)\rangle = \sum_i \alpha_i e^{\lambda_i t} |\phi_i\rangle \quad (13)$$

Because all excited-state λ_i values are negative and ordered, the long-time distribution's deviations from the ground state are dominated by the first excited state, $|\phi_1\rangle$:

$$\lim_{t \rightarrow \infty} |p(t)\rangle \approx |\pi\rangle + \frac{\alpha_1}{\alpha_0} |\phi_1\rangle e^{\lambda_1 t}, \quad (14)$$

meaning $\tau = 1/|\lambda_1|$ is the slowest timescale for relaxation. For two-state kinetics with rate k_{AB} for switching from state A to B and rate k_{BA} for switching back, the relaxation timescale is $\tau = (k_{AB} + k_{BA})^{-1}$. When the two-state kinetics is symmetric, $k_{AB} = k_{BA} = \lambda_1/2$. Hence, the typical time to switch from an A-rich to B-rich state, τ_{AB} , follows from the eigenvalue of the first excited state:

$$\tau_{AB} = \frac{1}{k_{AB}} = \frac{2}{\lambda_1}. \quad (15)$$

By “projecting out” the ground-state, one can modify the rate operator \hat{W} to solve for λ_1 and $|\phi_1\rangle$ as a ground state of the modified rate operator

$$\hat{W}' = \hat{W} - \xi |\phi_0\rangle \langle \phi_0|, \quad (16)$$

where ξ is a constant that shifts the eigenvalue of the original ground state down from 0⁷⁵. The penalty ξ is chosen to be sufficiently large that the first excited-state of \hat{W} becomes the ground state of \hat{W}' . Because our \hat{W} is typically not Hermitian, $|\phi_0\rangle$ and $\langle \phi_0|$ are not typically conjugate transposes of each other. As we have seen, $|\phi_0\rangle = |\pi\rangle$, and conservation of probability requires that $\langle \phi_0| = \langle \mathbb{1}|$ is a vector of ones.

VI. NUMERICAL EXPLORATION OF THE GTS MODEL

The GTS model is parameterized by eight distinct kinetic parameters, and the bistability is strongly influenced by those parameters. For any choice of the elementary rates, Gillespie sampling can be run with or without the maximum copy number M truncation. For example, the model acts as a bistable switch when $c_1^+ = c_1^- = c_2^+ = 5$, $c_2^- = c_A^+ = c_B^+ = 1$, $c_A^- = c_B^- = 0.25$, in line with prior studies of the model^{48,76}. Throughout the paper, we will keep the first six rate parameters fixed, focusing only on the dependence on the decomposition rates. Increasing the decomposition rates to $c_A^- = c_B^- = 0.8$ causes the distribution $P_{ss}(\Delta)$ to approach a unimodal one with a rapid relaxation rate λ_1 . Gillespie sampling offers a useful confirmation that truncation at $M = 20$ does not significantly alter bistability or bistability, but Gillespie samples must be regenerated for every set of rate parameters that one wishes to study.

Because parameter exploration would require so many repeated Gillespie simulations, prior work has explored whether mean-field analyses can capture the stochastic kinetics⁷⁷. For the GTS model, Warren and ten Wolde solved a system of seven differential equations at a negligible computational cost to generate a kinetic phase diagram mapping whether the GTS will act as a bistable switch as a function of the depletion rates of the two proteins, c_A^- and c_B^- . The analysis, briefly reproduced in Appendix B, divides the parameter space (gray line in Fig. 3) into regions with and without anticipated bistability. That mean-field analysis is computationally cheap, requiring only minutes on a laptop, but there are limitations. The mean-field calculations return a single average value for the copy number of each species, so it is not possible to obtain a distribution of outcomes like $P_{ss}(\Delta)$. Moreover, the mean value will typically be noninteger, an especially egregious limitation for the O, OA₂, and OB₂ species, which in reality always have either one or zero copies present.

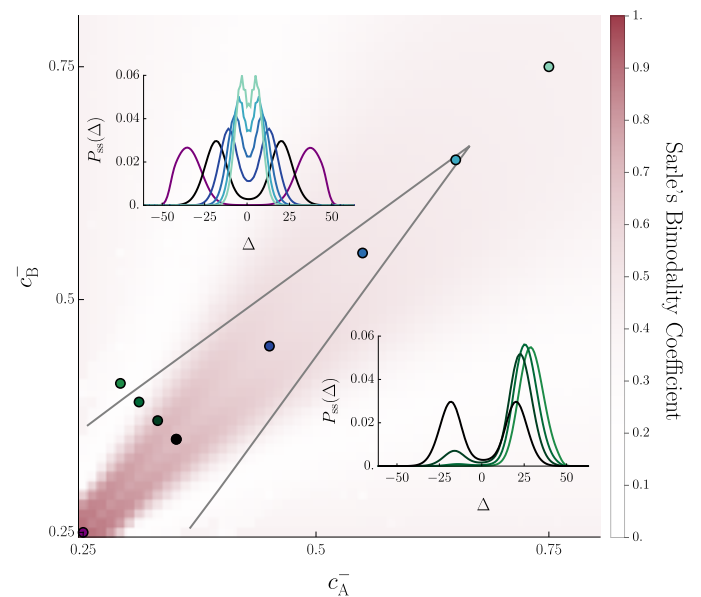


FIG. 3. A kinetic phase diagram shows the GTS model bistability as a function of protein decomposition rates. Prior mean-field analysis identified a bistable region inside the gray wedge. DMRG calculations, repeated over a grid of c_A^- and c_B^- values yield the steady-state distribution $|\pi\rangle$ and the marginal $P_{ss}(\Delta)$. Strongly bimodal $P_{ss}(\Delta)$ distributions arise for kinetic parameters within the gray wedge, but stochastic fluctuations cause the bimodal region to shrink relative to the mean-field prediction. Insets highlight two distinct ways the distribution can lose bimodality. The top left inset tracks $P_{ss}(\Delta)$ moving along a diagonal ($c_A^- = c_B^-$), showing a merging of the two peaks at high decomposition rates, akin to a second order phase transition. Note also that the discreteness of the copies of proteins is most apparent in the case of high decomposition rates (i.e., the sea green curve). The bottom right inset tracks $P_{ss}(\Delta)$ moving perpendicular to that diagonal, showing a first-order-like switch between a unimodal A-dominated distribution and a unimodal B-dominated distribution. The kinetic phase diagram is generated with fixed parameters $c_1^+ = c_1^- = c_2^+ = 5$ and $c_2^- = c_A^+ = c_B^+ = 1$ using DMRG convergence thresholds described in Appendix D.

A. DMRG Exploration

Having laid out the TN tools, we are in a position to illustrate how useful they are for the parameter exploration that has previously been studied with sampling and mean-field methods. By repeating DMRG calculations of the ground state on a grid of c_A^- and c_B^- values, we computed Sarle's bimodality coefficient for each ground state distribution. Figure 3 shows that strongly bimodal steady states roughly align with the mean-field calculation, though there are some notable deviations from the mean-field analysis. The bimodality coefficient demonstrates that incorporating fluctuations narrows the bimodal regime for low decomposition rates, as compared with the mean-field analysis. Graphically, the misidentification of the bimodal areas is identified by the white

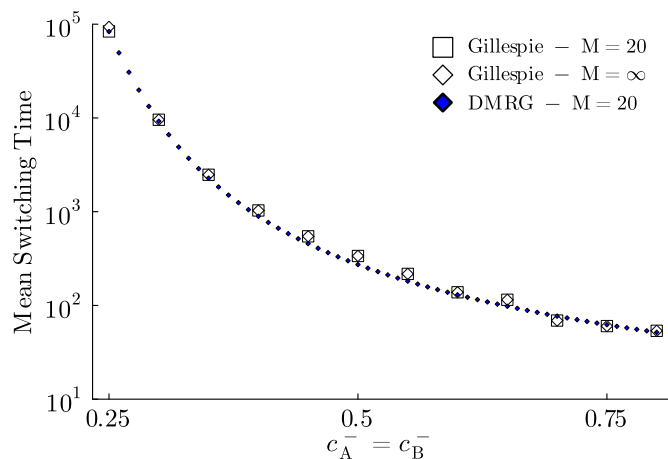


FIG. 4. Along the diagonal of Fig. 3, $c_A^- = c_B^-$ and $P_{ss}(\Delta)$ supports two peaks. The timescale for switches between positive and negative Δ spans orders of magnitude. That timescale is estimated from long Gillespie trajectories with or without a maximal occupation number ($M = 20$ and $M = \infty$), showing that the finite truncation at $M = 20$ only barely begins to impact the switching time for the smallest decomposition rates of $c_A^- = c_B^- = 0.25$. Excited-state DMRG calculations reproduce the $M = 20$ timescale via Eq. 15. Gillespie estimates of switching time require defining two basins with a threshold d , one with $\Delta \geq d$ and the other with $\Delta \leq -d$. The mean switching time is the average time between switches, averaged over 10^9 units of time. The mean switching time is weakly impacted by the chosen threshold d , and an appropriately chosen threshold must shift as a function of decomposition rate (see the top left inset of Fig. 3). Those shifting basin definitions create the impression that the Gillespie estimates are slightly noisy about the DMRG curve.

patches located inside the gray wedge of Fig. 3. This deviation from mean field is further visualized in the lower-right inset of Fig. 3, where we plot how $P_{ss}(\Delta)$ responds to a simultaneous decrease of one decomposition rate and increase of the other. Moving from black to green, $P_{ss}(\Delta)$ shows the A-rich states growing at the expense of the B-rich, with the two humps equal in height for the black $c_A^- = c_B^- = 0.35$ distribution. We can additionally view how $P_{ss}(\Delta)$ changes when the decomposition rates are increased in concert. The upper-left inset of Fig. 3 shows the two distinct humps at positive and negative Δ merging into each other at $\Delta = 0$, much like a second-order phase transition in an Ising model.

Even for parameters lying outside the mean-field-theory bistable region, e.g., $c_A^- = c_B^- = 0.75$, $P_{ss}(\Delta)$ retains a bimodal character. That light blue $P_{ss}(\Delta)$ reveals two humps merging so close to $\Delta = 0$ that stochastic fluctuations readily enable flips between A-rich and B-rich states. This set of parameters requires us to be cautious about differentiating between bimodality of the steady-state distribution and bistability, with the latter carrying dynamical information about switching from one set of states at one time to another set at a later time. That

bistability is better probed, not from the bimodality coefficient, but rather from the slowest timescale of relaxation, λ_1 , coming from the excited-state DMRG calculations. Figure 4 plots τ_{AB} as a function of the decomposition rate, explicitly showing that when $P_{ss}(\Delta)$ has a small “barrier” separating the two peaks, interconversion between positive and negative Δ is rapid. When performing excited-state DMRG with $c_A^- = c_B^-$, as in Fig. 4, we find some numerical advantage when using a penalty method with a modified rate operator that differs slightly from the \tilde{W} of Eq. (16) (see Appendix C).

B. Sampling versus DMRG

Although we generated Fig. 3 using the DMRG calculations, the figure could also be generated with sufficient Gillespie sampling. For every set of rate parameters, new Gillespie trajectories could be generated to sample $P_{ss}(\Delta)$ and to estimate τ_{AB} . We close the paper by reporting on performance in a manner that should help illustrate conditions for which DMRG is worth it and conditions for which sampling clearly makes more sense. The conversation is necessarily nuanced. One important reason why the discussion cannot be as simple as a comparison of timing of the two calculations is that the comparison is between apples and pears. While DMRG calculations were used to extract $P_{ss}(\Delta)$, the DMRG was actually converged to yield the more fine-grained $|\pi\rangle$. To estimate the steady-state probability for *every* microstate from Gillespie sampling is clearly impractical, so the DMRG is always computing an object with extra information. To the extent that the extra information is relevant to the question at hand, DMRG is always “worth it”.

If, however, one is only running DMRG to get $|\pi\rangle$ as a means to $P_{ss}(\Delta)$, one must ask whether that $P_{ss}(\Delta)$ should have just been sampled. Even this question has an annoying amount of nuance. The answer depends on whether the rate parameters are in a slow- or fast-switching regime. It depends on whether one intends to sample with advanced sampling tricks or in an approximate way with, for example, τ -leaping. And it especially depends on whether one aims to learn $P_{ss}(\Delta)$ for a single set of rate parameters or for a collection of many parameters as we did in Fig. 3. We focus on the last case, assuming that the goal is that parameter scan.

Sampling offers the advantage that each set of rate parameters is an independent calculation, so these calculations are trivially parallelized. For each value of the rate parameters, we estimate τ_{AB} by averaging over Gillespie trajectories with 1,000 switching events. The cost to converge $P_{ss}(\Delta)$ depends on the degree of convergence one hopes to achieve. On the way to convergence, the relative probabilities of the positive- and negative- Δ peaks will not be exactly equal. A coin flip calculation reflects that each peak will be $50 \pm 3\%$ after 10^3 switching events, $50 \pm 1\%$ after 10^4 switching events, and $50 \pm 0.3\%$ after

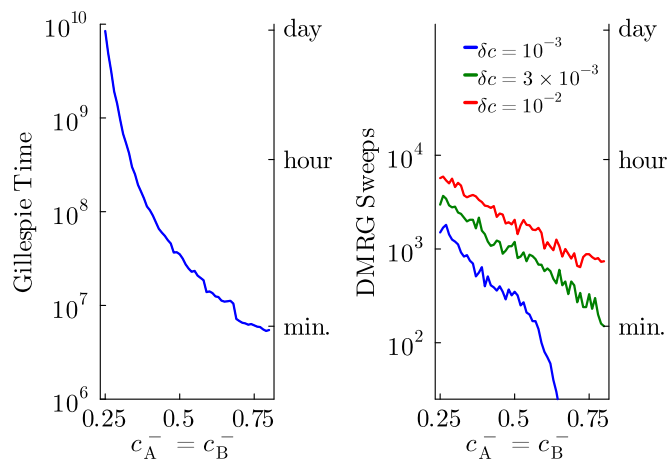


FIG. 5. Converging $P_{ss}(\Delta)$ from Gillespie samples requires many switching events, and the Gillespie time needed to generate 100,000 switching events grows dramatically longer as the decomposition rates $c_A^- = c_B^-$ decrease. An advantage of DMRG is that a converged DMRG calculation can seed the calculation for new decomposition rates. Starting with a converged $|\pi\rangle$ at one value of decomposition rates, we report the number of DMRG sweeps needed to converge and find $|\pi\rangle$ for a new value of decomposition rates which have been decreased by a step size of δc . By starting at large $c_A^- = c_B^-$ and decreasing with a small δc , one can more cheaply construct $P_{ss}(\Delta)$ down through the rare-event regime. Wall time is essentially proportional to the length of the Gillespie time and the number of DMRG sweeps. To guide the eye, an approximate conversion for the single CPU calculations is provided on the right axis, using the fact that a single DMRG sweep took about 0.4 seconds and 100,000 units of Gillespie time could be simulated in about 1 second of wall time.

10^5 switching events. We define that final threshold as our converged Gillespie sampling and estimate the time to converge by $10^5 \times \tau_{AB}$, plotted in Fig. 5. As anticipated, the rare-event regime with small c_A^- and c_B^- is dramatically more expensive to exhaustively sample by brute force Gillespie.

DMRG can offer computational advantages, particularly in this rare-events regime. Quantitatively assessing the benefit is challenging because any benefit depends on the particular way we choose to leverage one DMRG calculation to seed the next. The essential strategy is to start by converging DMRG in a corner of the parameter space. It has been noted that the convergence time of DMRG typically scales inversely with the spectral gap of the system⁷⁸, making it advantageous to begin with the fast-switching parameters $c_A^- = c_B^- = 0.8$. Having converged $|\pi\rangle$ for those rates, we should perturb the rates by a small step and reconverge. The expense of such a scheme is dominated by how many DMRG sweeps are necessary for each reconvergence. If you take small steps in parameter space, you have to reconverge many times. If you take large steps in parameter space, however, each reconvergence could grow much more expensive because the new $|\pi\rangle$ could be significantly dif-

ferent from the seed. To quantify these effects, we report in Fig. 5 the number of DMRG sweeps necessary to reconverge $|\pi\rangle$ at $(c_A^- - \delta c, c_B^- - \delta c)$ when seeded with the converged MPS for $|\pi\rangle$ at (c_A^-, c_B^-) for three different step sizes: $\delta c = 10^{-2}$, 3×10^{-3} , and 10^{-3} . To evaluate when the DMRG sweeps should terminate, we first ran an expensive, conservative convergence with extra DMRG sweeps. We then calculated the 1-norm between that “converged” distribution and the converging distribution, terminating the DMRG sweeps once the 1-norm dropped below 3×10^{-3} . Technical details of the convergence are discussed in Appendix D, but here we highlight that our 1-norm threshold is chosen so that the $P_{ss}(\Delta)$ from DMRG is about as converged as a histogram from trajectories of length $10^5 \times \tau_{AB}$.

Figure 5 shows that the expense of both types of calculations depends on the parameter regime, with fast decomposition rates always being easy. DMRG, however, scales better as the problem starts to become hard. While the cost to converge brute-force Gillespie histograms varies over three orders of magnitude, the number of DMRG sweeps varies over only one. That DMRG expense is really the marginal cost of getting the next distribution by dragging the last one over to the new rate parameters. How hard it is to perform that reconvergence depends on how dramatically $|\pi\rangle$ changes as a function of the rate parameters. Taking smaller steps δc , naturally results in smaller changes in $|\pi\rangle$ and fewer required sweeps.

VII. DISCUSSION

Ours is not the first work to suggest that the raising and lowering operator techniques (Doi-Peliti) are a powerful way to analyze the chemical master equation^{79–86}, nor is it the first to apply tensor network tools to CRNs^{25,27,36–43,87}. Our aim here has been to unite these two ideas, using the GTS model to demonstrate the systematic workflow from elementary reactions to Doi-Peliti representations to matrix product operators to steady-states. We believe that the case for approximating $|\pi\rangle$ without sampling is strongest for sensitivity analysis. Reseeding DMRG calculations with $|\pi\rangle$ for nearby parameters provides a natural way to reuse information from previously solved distributions, affording a computational advantage when one hopes to build a kinetic phase diagram like Fig. 3. Sensitivity analysis in the form of gradients should be even more easily computed if one measures the small changes to $|\pi\rangle$ in response to a small δc . Recently, a differentiable Gillespie algorithm has been developed to compute those gradients from sampling⁴. It will be interesting to compare and contrast in the future.

The GTS model has a modest enough state space that its steady state can actually be found with sparse-matrix solvers⁵³. We used that fact to aid in the convergence analysis (see Appendix D), but we hope the TN tools

will extend far beyond seven chemical species. TN tools are routinely used to study quantum spin chains with hundreds of spins, so it is not unreasonable to think that DMRG will be able to numerically investigate large CRNs with tens or even hundreds of species. Contrasting the quantum spin chains with the CRNs, however, highlights some important challenges that remain.

Firstly, large quantum spin chains often involve sites with low physical dimension (spins are up or down), whereas our chemical species have occupation numbers ranging from 0 to M . That difference grows the size of the MPS tensors, meaning 100 chemical species is a much harder problem than 100 spins. Clever ways to reduce the physical dimension will be important. One possibility would be to use quantized tensor trains (QTTs), which break down each physical dimension into the digits of its binary form, thus potentially reducing the scaling of storage cost in terms of the sizes of the individual state spaces from linear to logarithmic³⁹. Another possibility could be to find a better set of basis functions to represent the distributions^{87–89}. This approach would be analogous to the way the problem of large physical dimensionality is confronted in quantum chemistry⁶². Notably, it has already been shown that the Doi-Peliti formalism can be expressed in terms of Charlier or Hermite polynomials⁹⁰. Finally, for systems not exhibiting “discreteness-induced multimodality,” it may prove advantageous to use the CFPE instead of the CME, as done by Liao et al.⁴⁰, which would allow one to use a discretization grid of one’s choosing.

Secondly, unlike in spin chains, there is usually no notion of chemical species being connected along a one-dimensional chain. In applying the MPS ansatz to the GTS model, we are not assuming that species only interact with their neighbor in the chain, but correlations between distant species must be passed down the chain. Capturing those correlations requires a larger bond dimension than if the strongly interacting species had been connected to each other for the TN ansatz. As long as calculations with a large enough bond dimension can be performed, it is not necessary to optimize the topology of the TN ansatz, but it stands to reason that it could be advantageous to approximate the joint steady-state distribution by a TN whose connectivity mirrors the connectivity of the chemical species in the CRN. For technical reasons, tensor networks with loops lack an easily accessible “canonical form”⁹¹, frustrating the ability to fit a loopy tensor network via DMRG. By breaking apart loops in the CRN’s natural graph, one could instead approximate the joint distributions with tree tensor networks, which are amenable to DMRG⁹². Recent progress developing techniques for the “regauging” of loopy tensor networks^{93–98} offers hope that a DMRG-style approach may eventually be possible with a TN that matches the CRN topology.

Finally, while we successfully applied the ordinary DMRG algorithm for the GTS model, its performance was more delicate than we would have liked. Unless we

seeded the distributions well, we experienced numerical instability. Appendix D describes a slightly complicated reseeded scheme that we needed to engineer to confirm that DMRG sweeps did not get lost far from a converged MPS. We hope that DMRG algorithms specifically tailored for non-Hermitian systems^{64,65} will help improve the robustness and speed of convergence.

VIII. ACKNOWLEDGMENTS

We thank Cathryn Murphy and Nils E. Strand for helpful discussions. The material presented in this manuscript is based upon work supported by the National Science Foundation under Grant No. 2239867. T.R.G. was additionally supported by the Camille Dreyfus Teacher-Scholar Program.

Appendix A: The Doi-Peliti Formalism

The Doi-Peliti formalism turns a CME into its second-quantized representation, meaning it expresses elementary reactions in terms of the raising and lowering operators:

$$x_l^\dagger |n_l\rangle = |n_l + 1\rangle \quad (\text{A1})$$

$$x_l |n_l\rangle = n_l |n_l - 1\rangle \quad (\text{A2})$$

$$x_l |0\rangle = 0. \quad (\text{A3})$$

The subscripts indicate on which sites the operator acts, with l being A, A₂, OA₂, O, OB₂, B₂, or B in the case of the GTS model. When chemical species l is capped at a maximum occupation number of M , the operators take the form of $M + 1$ by $M + 1$ matrices:

$$x_l = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 2 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & M \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \text{ and } x_l^\dagger = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \quad (\text{A4})$$

In Appendix C of Ref.⁵³, we have shown that it is convenient to additionally define matrices

$$y_l = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix} \text{ and } z_l = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}, \quad (\text{A5})$$

which are the identity provided one is not within one (or two) counts of the maximum occupancy M . Reaction

1 in Fig. 1 involves the loss of two A coupled to the production of a dimer A₂. One can demonstrate that reaction 1 brings to \hat{W} a contribution

$$\hat{W} = c_1^+ \left(x_A^2 x_{A_2}^\dagger - x_A^\dagger x_A^2 y_{A_2} \right), \quad (\text{A6})$$

where the positive part is the gain term and the negative part the loss term⁵³. As is typical, we adopt a shorthand whereby only non-identity terms acting on $|\mathbf{n}\rangle$ are explic-

itly specified. The identities are implicit. For example,

$$x_A^2 x_{A_2}^\dagger = x_A^2 \otimes x_{A_2} \otimes \mathbb{I}_{OA_2} \otimes \mathbb{I}_O \otimes \mathbb{I}_{OB_2} \otimes \mathbb{I}_{B_2} \otimes \mathbb{I}_B \quad (\text{A7})$$

While operators acting on A, A₂, B, and B₂ are $M + 1$ by $M + 1$ matrices, those acting on OA₂, O, and OB₂ are only 2 by 2 matrices because our $n_{\text{DNA}} = 1$ manifold restricts the maximum occupancy of those species to 1. Note that for species involving O, our prior definitions of the operators become:

$$x_l = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad x_l^\dagger = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \text{and } y_l^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (\text{A8})$$

Separately carrying out the Doi-Peliti procedure on all fourteen elementary reactions yields:

$$\begin{aligned} \hat{W}_1 &= c_1^+ \left(x_A^2 x_{A_2}^\dagger - x_A^\dagger x_A^2 y_{A_2} \right) & \hat{W}_8 &= c_1^+ \left(x_B^2 x_{B_2}^\dagger - x_B^\dagger x_B^2 y_{B_2} \right) \\ \hat{W}_2 &= c_1^- \left(x_A^\dagger x_{A_2} - z_A x_{A_2}^\dagger x_{A_2} \right) & \hat{W}_9 &= c_1^- \left(x_B^\dagger x_{B_2} - z_B x_{B_2}^\dagger x_{B_2} \right) \\ \hat{W}_3 &= c_2^+ \left(x_{A_2} x_{OA_2}^\dagger x_O - x_{A_2}^\dagger x_{A_2} c_{OA_2} x_O^\dagger x_O \right) & \hat{W}_{10} &= c_2^+ \left(x_{B_2} x_{OB_2}^\dagger x_O - x_{B_2}^\dagger x_{B_2} c_{OB_2} x_O^\dagger x_O \right) \\ \hat{W}_4 &= c_2^- \left(x_{A_2}^\dagger x_{OA_2} x_O^\dagger - x_{OA_2}^\dagger x_{OA_2} c_O + \right. & \hat{W}_{11} &= c_2^- \left(x_{B_2}^\dagger x_{OB_2} x_O^\dagger - x_{OB_2}^\dagger x_{OB_2} c_O + \right. \\ & \quad \left. (\mathbb{I}_{A_2} - y_{A_2}) x_{OA_2}^\dagger x_{OA_2} c_O \right) & & \quad \left. (\mathbb{I}_{B_2} - y_{B_2}) x_{OB_2}^\dagger x_{OB_2} c_O \right) \\ \hat{W}_5 &= c_A^+ \left(x_A^\dagger x_O^\dagger x_O - y_A x_O^\dagger x_O \right) & \hat{W}_{12} &= c_B^+ \left(x_B^\dagger x_O^\dagger x_O - y_B x_O^\dagger x_O \right) \\ \hat{W}_6 &= c_A^+ \left(x_A^\dagger x_{OA_2}^\dagger x_{OA_2} - y_A x_{OA_2}^\dagger x_{OA_2} \right) & \hat{W}_{13} &= c_B^+ \left(x_B^\dagger x_{OB_2}^\dagger x_{OB_2} - y_B x_{OB_2}^\dagger x_{OB_2} \right) \\ \hat{W}_7 &= c_A^- \left(x_A - x_A^\dagger x_A \right) & \hat{W}_{14} &= c_B^- \left(x_B - x_B^\dagger x_B \right). \end{aligned} \quad (\text{A9})$$

The rate operator \hat{W} for the full CRN is simply the sum of the individual contributions from each reaction:

$$\hat{W} = \sum_{i=1}^{14} \hat{W}_i. \quad (\text{A10})$$

Appendix B: The Mean Field Approach

As a point of comparison, we replicated the mean-field results presented in⁴⁸. In the large-system-size limit, the mass-action kinetics form of the 14 elementary reactions gives seven coupled differential equations for the mean concentration of the species. Steady-state kinetics requires that the concentration of each species has become stationary, so

$$\begin{aligned} 0 &= \frac{d[A]}{dt} = -2c_1^+ [A]^2 + 2c_1^- [A_2] + c_A^+ ([O] + [OA_2]) - c_A^- [A] \\ 0 &= \frac{d[A_2]}{dt} = c_1^+ [A]^2 - c_1^- [A_2] - c_2^+ [O][A_2] + c_2^- [OA_2] \\ 0 &= \frac{d[OA_2]}{dt} = c_2^+ [O][A_2] - c_2^- [OA_2] \\ 0 &= \frac{d[O]}{dt} = -c_2^+ [O]([A_2] + [B_2]) + c_2^- ([OA_2] + [OB_2]) \\ 0 &= \frac{d[OB_2]}{dt} = c_2^+ [O][B_2] - c_2^- [OB_2] \\ 0 &= \frac{d[B_2]}{dt} = c_1^+ [B]^2 - c_1^- [B_2] - c_2^+ [O][B_2] + c_2^- [OB_2] \\ 0 &= \frac{d[B]}{dt} = -2c_1^+ [B]^2 + 2c_1^- [B_2] + c_B^+ ([O] + [OB_2]) - c_B^- [B]. \end{aligned} \quad (\text{B1})$$

We selected two different initial conditions, $(n_A, n_{A_2}, n_{OA_2}, n_O, n_{OB_2}, n_{B_2}, n_B) = (10, 10, 1, 0, 0, 0, 0)$ and $(0, 0, 0, 0, 1, 10, 10)$, the first A-rich and the second

B-rich. From these seeds, we numerically integrated the mass action kinetics with the Euler scheme using $\delta t = 0.01$ until the time derivatives of all species concentrations had fallen below 10^{-10} . If one trajectory resulted in $\Delta > 1$ while the other resulted in $\Delta < -1$, the elementary rate parameters were considered to be in a bistable regime. We scanned a grid of c_A^- and c_B^- values, identifying the bistable points then plotted in Fig. 3 the convex hull of those bistable parameters, a plot that aligns with Fig. 2 of⁴⁸.

Appendix C: Excited-state DMRG for Symmetric Decomposition Rates

We find that, for symmetric parameter sets ($c_A^- = c_B^-$), the penalty-method calculation of the first excited state is even more robust when DMRG is applied to a modified rate operator with a slightly different penalty (one that would have been natural for a Hermitian problem):

$$\hat{W}'' = \hat{W} - \xi |\pi\rangle \langle \pi|, \quad (\text{C1})$$

where $\langle \pi|$ is now the conjugate transpose of $|\pi\rangle$. In general, the maximal right eigenvector of this \hat{W}'' is not the targeted first excited state of \hat{W} , $|\phi_1\rangle$. When $c_A^- = c_B^-$, we show (next paragraph) that $\langle \pi|\phi_1\rangle = 0$. Eq. (C1) then implies $\hat{W}''|\phi_1\rangle = \hat{W}|\phi_1\rangle$, so $|\phi_1\rangle$ is an eigenstate of the original rate operator and the new modified operator. Because $\langle \pi|\pi\rangle$ does not vanish, that $|\phi_1\rangle$ becomes the ground state of \hat{W}'' for a sufficiently large ξ .

Completing the argument requires us to prove that $\langle \pi|\phi_1\rangle = 0$ when $c_A^- = c_B^-$. We do so with a symmetry consideration, namely $\langle \pi|$ is symmetric under the exchange of the protein labels A and B while $|\phi_1\rangle$ is antisymmetric under the same exchange. The inner product between symmetric and antisymmetric vectors necessarily vanishes. To be slightly more explicit, if $(n_A, n_{A_2}, n_{OA_2}, n_O, n_{OB_2}, n_{B_2}, n_B) = (n_1, n_2, n_3, n_4, n_5, n_6, n_7)$ has steady-state probability p , so too must microstate $(n_7, n_6, n_5, n_4, n_3, n_2, n_1)$. In contrast, $|\phi_1\rangle$'s antisymmetry is required because the orthonormality of \hat{W} 's left and right eigenvectors gives $\langle \phi_0|\phi_1\rangle = \langle \mathbb{1}|\phi_1\rangle = 0$. For the sum of the elements in $|\phi_1\rangle$ to be zero, microstates $(n_1, n_2, n_3, n_4, n_5, n_6, n_7)$ and $(n_7, n_6, n_5, n_4, n_3, n_2, n_1)$ have equal magnitude and opposite sign.

Appendix D: DMRG convergence

1. Ground state scan

To scan the kinetic parameter space and construct Fig. 3, DMRG calculations had to be repeatedly initiated with seeds, and then DMRG sweeps were run until a convergence threshold was reached. The sweeps were performed using two-site DMRG with a maximum bond

dimension of 100 and a singular value cutoff of 10^{-15} . Throughout the full scan of the parameter space, the singular value cutoff was the limiting factor—the largest bond dimension of a converged ground state MPS was 54. Because we have non-Hermitian rate operators, in place of a symmetric solver like the Davidson algorithm, the Arnoldi method⁶⁹ was used with a maximum Krylov space dimension of 3 to solve the local eigenvalue problems.

We sought to initiate the parameter scan in the easiest-to-compute corner of Fig. 3. Since DMRG convergence tends to be faster when spectral gaps are larger, we thus started with the largest values of c_A^- and c_B^- . There, the DMRG readily converges when seeded from a uniform distribution. That convergence is monitored most simply by tracking the “energy”, $\langle \phi_{\text{trial}} | \hat{W} | \phi_{\text{trial}} \rangle / \langle \phi_{\text{trial}} | \phi_{\text{trial}} \rangle$, where $|\phi_{\text{trial}}\rangle$ is the MPS state and $\langle \phi_{\text{trial}}|$ is the complex conjugate. We refer to this quantity as the energy because it would be the expected energy if \hat{W} were a Hermitian Hamiltonian. Even for the non-Hermitian rate operator, the quantity makes sense to inspect because it would vanish for the true ground state $|\pi\rangle$, at which $\hat{W}|\pi\rangle = 0$. We considered 10^{-5} to be an appropriate threshold for identifying the ground state. For $c_A^- = c_B^- = 0.8$, this level of convergence was reached from the uniform seed within 10^4 DMRG sweeps.

After converging the steady-state distribution in the top right corner of Fig. 3, we wanted to scan in such a manner that the steady-state distribution would move as gently as possible. The gentlest move came from simultaneously decreasing both decomposition rates by the same small step, so we reseeded the DMRG convergences by passing the converged distribution at c_A^- and c_B^- down the diagonal to $c_A^- - \delta c$ and $c_B^- - \delta c$. At the boundaries of the scanned region, horizontal and vertical steps were taken, as necessary, to fully scan a grid of values c_A^- and c_B^- from 0.25 to 0.8 in increments of 0.01. DMRG can be reconverged with steps in the elementary rates of size 0.01 (see Fig. 5), but we preferred to take those steps via 200 smaller steps ($\delta c = 5 \times 10^{-5}$). In practice, we would run 100 DMRG sweeps for each new parameter and took the MPS with the lowest energy. Notice that we did not need to solve for the ground state on a 5×10^{-5} grid to benefit from the smaller step sizes. We only needed to connect points on the 0.01 grid by finely discretized lines along which the ground state distribution was “adiabatically dragged”.

We found that the DMRG would occasionally get knocked off course, finding a distribution so far from the ground state that a very large number of sweeps was required to converge. To improve the stability of the scan, we used $P_{\text{ss}}(\Delta)$ as a cheap-to-compute indicator of severe numerical instabilities. For example, if the MPS for a point on the grid with spacing 0.01 had an energy exceeding 10^{-4} or gave a marginal distribution $P_{\text{ss}}(\Delta)$ with negative or complex values, it was clear that the MPS had veered too far from $|\pi\rangle$. In these cases, we would attempt up to 10^4 DMRG sweeps to reconverge to

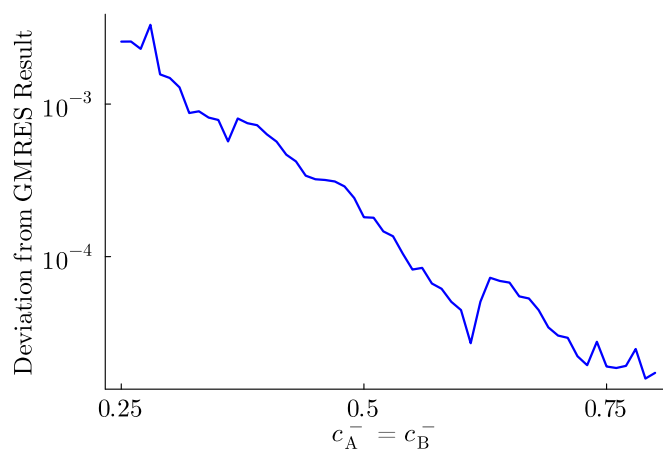


FIG. 6. The deviations in 1-norm of the joint steady-state probability distributions obtained from DMRG (with singular value cutoffs of 10^{-15}) from the results obtained from GMRES. The distributions are normalized to have a probabilistic interpretation (i.e., 1-norms are 1.0).

below the energy threshold of 10^{-5} . Occasionally, even those extra sweeps were insufficient. If the sweeps had been sufficient to at least return a MPS corresponding to a $P_{ss}(\Delta)$ without negative or complex values, the MPS with the lowest absolute energy during the extra sweeps was used. If even this was not achieved, we returned the MPS with the last $\delta c = 5 \times 10^{-5}$ grid point with positive real $P_{ss}(\Delta)$ probabilities.

As the energy gets closer to zero, it is a reflection that a TDVP evolution of the MPS state would move by smaller and smaller amounts. Arrested dynamics, as reflected by a tiny “energy”, is necessary at the steady state, but dynamics could become very slow even far away from the ground state. A more thorough validation of the DMRG convergence is offered by a comparison with a sparse linear algebra solver that does not use the compression of the MPS ansatz. Because the GTS state space is small enough⁵³, we used GMRES to converge $|\pi^{\text{GMRES}}\rangle$ for $c_A^- = c_B^-$. Fig. 6 plots the 1-norm of the difference $|\pi^{\text{GMRES}}\rangle - |\pi^{\text{DMRG}}\rangle$, reflecting that our “energy” threshold indeed returns a very good approximation to the ground state. Summing over all microstates, the absolute value of the greatest deviation between the GMRES probability and the DMRG probability we observed was around 3×10^{-3} . Suppose one summed only over the errors in the positive- Δ microstates. As a worst-case scenario, the $c_A^- = c_B^-$ DMRG ground state would over- or under-state the positive- Δ probability as $0.5 \pm 3 \times 10^{-3}$, a level of convergence that requires the 10^5 Gillespie switching events we discussed in the main text.

Comparisons with GMRES furthermore showcase how the MPS approximation acts as a type of data compression. The matrix product states obtained are intended to be low-parametric functions capable of returning an accurate estimate of the probability associated with each

of the 583,433 possible microstates in the model. Thus, it is of interest to compare the number of parameters (i.e., tensor elements) in the matrix product states obtained with the number of microstates. Among the 3,136 possible parameter sets considered, the number of tensor elements required to fit the distribution ranged from 4,718 to 34,893. This corresponds to between 0.8% and 6.0% of the number of microstates in the model. The median number of required tensor elements was 12,909, or 2.2% of the number of microstates. These results indicate that the MPS ansatz was usually able to provide a roughly 50-fold compression of the data representing the joint probability distribution. We believe that once we scale our approach to larger systems and use tensor networks with topologies more directly mimicking the topology of the system’s dynamics, it will be possible to obtain much larger compression ratios.

2. Timing calculations

For the experiments concerning convergence times, small steps were taken from each of the parameter sets with $c_A^- = c_B^-$ on a grid with a spacing of 0.01 between 0.25 and 0.8 to the corresponding parameter set with both c_A^- and c_B^- reduced by some step δc . For each such step, 20,000 sweeps at a cutoff of 10^{-15} were provided for complete convergence. Every 10 sweeps a copy of the MPS was saved. It was then possible to track the convergence with respect to the number of sweeps by monitoring the deviation in 1-norm between the distribution obtained at a given number of sweeps and the final distribution obtained after 20,000 sweeps. All distributions were normalized to have a probabilistic interpretation (i.e., a 1-norm of 1.0). The number of sweeps for convergence was defined to be the minimum number of sweeps required to reduce the 1-norm deviation to the fully converged distribution to below 3×10^{-3} . To obtain the data required to produce Fig. 5, this procedure was then repeated for $\delta c = 10^{-2}, 3 \times 10^{-3}$, and 10^{-3} . The number of sweeps required for the $\delta c = 10^{-3}$ series falls off for larger values of $c_A^- = c_B^-$ as the induced change in 1-norm due to the step taken eventually falls below 3×10^{-3} in that region of parameter space. One can gain confidence in the degree of convergence after 20,000 sweeps by comparing the change in the 1-norm deviation to the final distribution in the final 10 sweeps with the change resulting from the 10 sweeps directly after our threshold of 3×10^{-3} had been reached. We found that the changes in 1-norm deviation from the final distribution for the last 10 sweeps were always less than 10% (and in fact almost always less than 1%) of those resulting from the 10 sweeps directly after “convergence” had been reached.

3. Excited state scan

Our reseeded scheme could also be directly generalized to solve for excited states along the diagonal where $c_A^- = c_B^-$. For every small step in parameter space, $|\pi\rangle$ was updated first using \tilde{W} for the appropriate kinetic parameters. Then $|\phi_1\rangle$ was updated using \tilde{W}' as defined in Eq. (C1) using the same kinetic parameters and the $|\pi\rangle$ just found. Both $|\pi\rangle$ and $|\phi_1\rangle$ were reseeded such that the $|\pi\rangle$ at the previous parameter set was used as the seed for the calculation of $|\pi\rangle$ for the current parameter set and the $|\phi_1\rangle$ at the previous parameter set was used as the seed for the calculation of $|\phi_1\rangle$ for the current parameter set. However, to obtain accurate switching times from excited-state DMRG in the rare-events regime, it was necessary to further increase the accuracy used. Thus, the singular cutoff was reduced to 10^{-20} , and a further 10,000 sweeps were provided at each grid point to ensure full convergence of the excited states at this higher accuracy.

- ¹J. Goutsias and G. Jenkinson, "Markovian dynamics on complex reaction networks," *Physics Reports* **529**, 199–264 (2013).
- ²D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry* **81**, 2340–2361 (1977).
- ³G. Minas and D. A. Rand, "Parameter sensitivity analysis for biochemical reaction networks," *Mathematical Biosciences and Engineering* **16**, 3965–3987 (2019).
- ⁴K. Rijal and P. Mehta, "A differentiable Gillespie algorithm for simulating chemical kinetics, parameter estimation, and designing synthetic biological circuits," *bioRxiv* (2024), [10.1101/2024.07.07.602397](https://doi.org/10.1101/2024.07.07.602397).
- ⁵P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, "Transition path sampling: Throwing ropes over rough mountain passes, in the dark," *Annual Review of Physical Chemistry* **53**, 291–318 (2002).
- ⁶G. E. Crooks and D. Chandler, "Efficient transition path sampling for nonequilibrium stochastic dynamics," *Physical Review E* **64**, 026109 (2001).
- ⁷T. R. Gingrich and P. L. Geissler, "Preserving correlations between trajectories for efficient path sampling," *The Journal of Chemical Physics* **142**, 234104 (2015).
- ⁸R. J. Allen, P. B. Warren, and P. R. ten Wolde, "Sampling rare switching events in biochemical networks," *Physical Review Letters* **94**, 018104 (2005).
- ⁹R. J. Allen, D. Frenkel, and P. R. ten Wolde, "Simulating rare events in equilibrium or nonequilibrium stochastic systems," *The Journal of Chemical Physics* **124**, 024102 (2006).
- ¹⁰R. J. Allen, C. Valeriani, and P. R. ten Wolde, "Forward flux sampling for rare event simulations," *Journal of Physics: Condensed Matter* **21**, 463102 (2009).
- ¹¹R. Orús, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states," *Annals of Physics* **349**, 117–158 (2014).
- ¹²R. Orús, "Tensor networks for complex quantum systems," *Nature Reviews Physics* **1**, 538–550 (2019).
- ¹³I. V. Oseledets, "Tensor-train decomposition," *SIAM Journal on Scientific Computing* **33**, 2295–2317 (2011).
- ¹⁴U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states," *Annals of Physics* **326**, 96–192 (2011).
- ¹⁵S. R. White, "Density matrix formulation for quantum renormalization groups," *Physical Review Letters* **69**, 2863 (1992).
- ¹⁶S. R. White, "Density-matrix algorithms for quantum renormalization groups," *Physical Review B* **48**, 10345 (1993).

- ¹⁷G. K.-L. Chan and S. Sharma, "The density matrix renormalization group in quantum chemistry," *Annual Review of Physical Chemistry* **62**, 465–481 (2011).
- ¹⁸S. Wouters and D. Van Neck, "The density matrix renormalization group for ab initio quantum chemistry," *The European Physical Journal D* **68**, 1–20 (2014).
- ¹⁹S. Sharma and G. K.-L. Chan, "Spin-adapted density matrix renormalization group algorithms for quantum chemistry," *The Journal of Chemical Physics* **136**, 124121 (2012).
- ²⁰S. R. White and R. L. Martin, "Ab initio quantum chemistry using the density matrix renormalization group," *The Journal of Chemical Physics* **110**, 4127–4130 (1999).
- ²¹G. K.-L. Chan and M. Head-Gordon, "Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group," *The Journal of Chemical Physics* **116**, 4462–4476 (2002).
- ²²T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review* **51**, 455–500 (2009).
- ²³L. Grasedyck, D. Kressner, and C. Tobler, "A literature survey of low-rank tensor approximation techniques," *GAMM-Mitteilungen* **36**, 53–78 (2013).
- ²⁴S. V. Dolgov and D. V. Savostyanov, "Alternating minimal energy methods for linear systems in higher dimensions," *SIAM Journal on Scientific Computing* **36**, A2248–A2271 (2014).
- ²⁵S. Dolgov and B. Khoromskij, "Simultaneous state-time approximation of the chemical master equation using tensor product formats," *Numerical Linear Algebra with Applications* **22**, 197–219 (2015).
- ²⁶J. P. Garrahan, "Classical stochastic dynamics and continuous matrix product states: gauge transformations, conditioned and driven processes, and equivalence of trajectory ensembles," *Journal of Statistical Mechanics: Theory and Experiment* **2016**, 073208 (2016).
- ²⁷P. Helms, U. Ray, and G. K.-L. Chan, "Dynamical phase behavior of the single- and multi-lane asymmetric simple exclusion process via matrix product states," *Physical Review E* **100**, 022101 (2019).
- ²⁸N. E. Strand, H. Vroylandt, and T. R. Gingrich, "Using tensor network states for multi-particle Brownian ratchets," *The Journal of Chemical Physics* **156**, 221103 (2022).
- ²⁹N. E. Strand, H. Vroylandt, and T. R. Gingrich, "Computing time-periodic steady-state currents via the time evolution of tensor network states," *The Journal of Chemical Physics* **157**, 054104 (2022).
- ³⁰J. P. Garrahan and F. Pollmann, "Topological phases in the dynamics of the simple exclusion process," *Physical Review E* **109**, L032105 (2024).
- ³¹L. Causer, M. C. Bañuls, and J. P. Garrahan, "Optimal sampling of dynamical large deviations via matrix product states," *Physical Review E* **103**, 062144 (2021).
- ³²L. Causer, M. C. Bañuls, and J. P. Garrahan, "Finite time large deviations via matrix product states," *Physical Review Letters* **128**, 090605 (2022).
- ³³L. Causer, I. Lesanovsky, M. C. Bañuls, and J. P. Garrahan, "Dynamics and large deviation transitions of the XOR-Fredrickson-Andersen kinetically constrained model," *Physical Review E* **102**, 052132 (2020).
- ³⁴L. Causer, M. C. Bañuls, and J. P. Garrahan, "Optimal sampling of dynamical large deviations in two dimensions via tensor networks," *Physical Review Letters* **130**, 147401 (2023).
- ³⁵L. Causer, J. P. Garrahan, and A. Lamacraft, "Slow dynamics and large deviations in classical stochastic Fredkin chains," *Physical Review E* **106**, 014128 (2022).
- ³⁶W. Merbis, M. Geurts, C. de Mulatier, and P. Corboz, "Effective dimensional reduction of complex systems based on tensor networks," *arXiv preprint* (2024), [10.48550/arXiv.2411.13364](https://arxiv.org/abs/10.48550/arXiv.2411.13364).
- ³⁷S. Dolgov and D. Savostyanov, "Tensor product approach to modelling epidemics on networks," *Applied Mathematics and Computation* **460**, 128290 (2024).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0276591

- ³⁸M. Hegland and J. Garcke, “On the numerical solution of the chemical master equation with sums of rank one tensors,” *ANZIAM Journal* **52**, C628–C643 (2010).
- ³⁹V. Kazeev, M. Khammash, M. Nip, and C. Schwab, “Direct solution of the chemical master equation using quantized tensor trains,” *PLOS Computational Biology* **10**, e1003359 (2014).
- ⁴⁰S. Liao, T. Vejchodský, and R. Erban, “Tensor methods for parameter estimation and bifurcation analysis of stochastic reaction networks,” *Journal of The Royal Society Interface* **12**, 20150233 (2015).
- ⁴¹P. Gelß, S. Matera, and C. Schütte, “Solving the master equation without kinetic Monte Carlo: Tensor train approximations for a CO oxidation model,” *Journal of Computational Physics* **314**, 489–502 (2016).
- ⁴²H. D. Vo and R. B. Sidje, “An adaptive solution to the chemical master equation using tensors,” *The Journal of Chemical Physics* **147**, 044102 (2017).
- ⁴³I. G. Ion, C. Wildner, D. Loukrezis, H. Koepl, and H. De Gersem, “Tensor-train approximation of the chemical master equation and its application for parameter inference,” *The Journal of Chemical Physics* **155**, 034102 (2021).
- ⁴⁴S. B. Nicholson and T. R. Gingrich, “Quantifying rare events in stochastic reaction-diffusion dynamics using tensor networks,” *Physical Review X* **13**, 041006 (2023).
- ⁴⁵S. Holtz, T. Rohwedder, and R. Schneider, “The alternating linear scheme for tensor optimization in the tensor train format,” *SIAM Journal on Scientific Computing* **34**, A683–A713 (2012).
- ⁴⁶A. Duncan, S. Liao, T. Vejchodský, R. Erban, and R. Grima, “Noise-induced multistability in chemical systems: Discrete versus continuum modeling,” *Physical Review E* **91**, 042111 (2015).
- ⁴⁷R. Erban and S. J. Chapman, *Stochastic modelling of reaction-diffusion processes*, Vol. 60 (Cambridge University Press, 2020).
- ⁴⁸P. B. Warren and P. R. ten Wolde, “Enhancement of the stability of genetic switches by overlapping upstream regulatory domains,” *Physical Review Letters* **92**, 128101 (2004).
- ⁴⁹M. Doi, “Stochastic theory of diffusion-controlled reaction,” *Journal of Physics A: Mathematical and General* **9**, 1479 (1976).
- ⁵⁰L. Peliti, “Path integral approach to birth-death processes on a lattice,” *Journal de Physique* **46**, 1469–1483 (1985).
- ⁵¹M. Fishman, S. R. White, and E. M. Stoudenmire, “The ITensor Software Library for Tensor Network Calculations,” *SciPost Physics Codebases* , 4 (2022).
- ⁵²M. Fishman, S. R. White, and E. M. Stoudenmire, “Codebase release 0.3 for ITensor,” *SciPost Physics Codebases* , 4–r0.3 (2022).
- ⁵³N. E. Strand, S. B. Nicholson, H. Vroylandt, and T. R. Gingrich, “From high-dimensional committers to reactive insights,” *The Journal of Chemical Physics* **161**, 224109 (2024).
- ⁵⁴M. Cramer, J. Eisert, M. B. Plenio, and J. Dreissig, “Entanglement-area law for general bosonic harmonic lattice systems,” *Physical Review A* **73**, 012309 (2006).
- ⁵⁵M. Srednicki, “Entropy and area,” *Physical Review Letters* **71**, 666 (1993).
- ⁵⁶M. B. Plenio, J. Eisert, J. Dreissig, and M. Cramer, “Entropy, entanglement, and area: Analytical results for harmonic lattice systems,” *Physical Review Letters* **94**, 060503 (2005).
- ⁵⁷G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, “Entanglement in quantum critical phenomena,” *Physical Review Letters* **90**, 227902 (2003).
- ⁵⁸C. Hubig, I. McCulloch, and U. Schollwöck, “Generic construction of efficient matrix product operators,” *Physical Review B* **95**, 035129 (2017).
- ⁵⁹J. Cui, J. I. Cirac, and M. C. Bañuls, “Variational matrix product operators for the steady state of dissipative quantum systems,” *Physical Review Letters* **114**, 220601 (2015).
- ⁶⁰H. P. Casagrande, D. Poletti, and G. T. Landi, “Analysis of a density matrix renormalization group approach for transport in open quantum systems,” *Computer Physics Communications* **267**, 108060 (2021).
- ⁶¹Z. Guo, Z.-T. Xu, M. Li, L. You, and S. Yang, “Variational matrix product state approach for non-Hermitian system based on a companion Hermitian hamiltonian,” *arXiv preprint* (2022), 10.48550/arXiv.2210.14858.
- ⁶²G. K.-L. Chan and T. Van Voorhis, “Density-matrix renormalization-group algorithms with nonorthogonal orbitals and non-Hermitian operators, and applications to polyenes,” *The Journal of Chemical Physics* **122**, 204101 (2005).
- ⁶³Y.-K. Huang, “Biorthonormal matrix-product-state analysis for the non-Hermitian transfer-matrix renormalization group in the thermodynamic limit,” *Journal of Statistical Mechanics: Theory and Experiment* **2011**, P07003 (2011).
- ⁶⁴F. Verstraete, J. J. Garcia-Ripoll, and J. I. Cirac, “Matrix product density operators: Simulation of finite-temperature and dissipative systems,” *Physical Review Letters* **93**, 207204 (2004).
- ⁶⁵P. Zhong, W. Pan, H. Lin, X. Wang, and S. Hu, “Density-matrix renormalization group algorithm for non-Hermitian systems,” *arXiv preprint* (2024), 10.48550/arXiv.2401.15000.
- ⁶⁶D.-W. Zhang, Y.-L. Chen, G.-Q. Zhang, L.-J. Lang, Z. Li, and S.-L. Zhu, “Skin superfluid, topological Mott insulators, and asymmetric dynamics in an interacting non-Hermitian Aubry-André-Harper model,” *Physical Review B* **101**, 235150 (2020).
- ⁶⁷K. Liao, H. Zhai, E. M. C. Christlmaier, T. Schraivogel, P. L. Ríos, D. Kats, and A. Alavi, “Density matrix renormalization group for transcorrelated Hamiltonians: Ground and excited states in molecules,” *Journal of Chemical Theory and Computation* **19**, 1734–1743 (2023).
- ⁶⁸P.-X. Shen, Z. Lu, J. L. Lado, and M. Trif, “Non-Hermitian Fermi-Dirac distribution in persistent current transport,” *Physical Review Letters* **133**, 086301 (2024).
- ⁶⁹W. E. Arnoldi, “The principle of minimized iterations in the solution of the matrix eigenvalue problem,” *Quarterly of Applied Mathematics* **9**, 17–29 (1951).
- ⁷⁰S. R. White, “Density matrix renormalization group algorithms with a single center site,” *Physical Review B* **72**, 180403 (2005).
- ⁷¹C. Hubig, I. P. McCulloch, U. Schollwöck, and F. A. Wolf, “Strictly single-site DMRG algorithm with subspace expansion,” *Physical Review B* **91**, 155115 (2015).
- ⁷²J. Haegeman, C. Lubich, I. Oseledets, B. Vandereycken, and F. Verstraete, “Unifying time evolution and optimization with matrix product states,” *Physical Review B* **94**, 165116 (2016).
- ⁷³Here we take the defining step of a DMRG algorithm to be computing a maximal eigenvector of a local operator, not necessarily that such an eigenvector is found using the Davidson algorithm. When we say the DMRG algorithm is unchanged from the Hermitian case, the local problem’s maximal eigenvectors must of course be found with an appropriate solver that does not assume a symmetric local operator.
- ⁷⁴R. Pfister, K. A. Schwarz, M. Janczyk, R. Dale, and J. B. Freeman, “Good things peak in pairs: a note on the bimodality coefficient,” *Frontiers in Psychology* **4**, 700 (2013).
- ⁷⁵E. M. Stoudenmire and S. R. White, “Studying two-dimensional systems with the density matrix renormalization group,” *Annual Review of Condensed Matter Physics* **3**, 111–128 (2012).
- ⁷⁶P. B. Warren and P. R. ten Wolde, “Chemical models of genetic toggle switches,” *The Journal of Physical Chemistry B* **109**, 6812–6823 (2005).
- ⁷⁷R. Blossey and C. Giuraniuc, “Mean-field versus stochastic models for transcriptional regulation,” *Physical Review E* **78**, 031909 (2008).
- ⁷⁸F. Verstraete and J. I. Cirac, “Matrix product states represent ground states faithfully,” *Physical Review B* **73**, 094423 (2006).
- ⁷⁹P. Grassberger and M. Scheunert, “Fock-space methods for identical classical objects,” *Fortschritte der Physik* **28**, 547–578 (1980).
- ⁸⁰J. L. Cardy and P. Grassberger, “Epidemic models and percolation,” *Journal of Physics A: Mathematical and General* **18**, L267 (1985).
- ⁸¹F. Van Wijland, K. Oerding, and H. Hilhorst, “Wilson renormalization of a reaction-diffusion process,” *Physica A: Statistical Mechanics and its Applications* **251**, 179–201 (1998).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0276591

- ⁸²M. Sasai and P. G. Wolynes, “Stochastic gene expression as a many-body problem,” *Proceedings of the National Academy of Sciences* **100**, 2374–2379 (2003).
- ⁸³Y. Lan, P. G. Wolynes, and G. A. Papoian, “A variational approach to the stochastic aspects of cellular signal transduction,” *The Journal of Chemical Physics* **125**, 124106 (2006).
- ⁸⁴A. M. Walczak, A. Mugler, and C. H. Wiggins, “A stochastic spectral analysis of transcriptional regulatory cascades,” *Proceedings of the National Academy of Sciences* **106**, 6529–6534 (2009).
- ⁸⁵M. F. Weber and E. Frey, “Master equations and the theory of stochastic path integrals,” *Reports on Progress in Physics* **80**, 046601 (2017).
- ⁸⁶J. J. Vastola, “Solving the chemical master equation for monomolecular reaction systems and beyond: a Doi-Peliti path integral view,” *Journal of Mathematical Biology* **83**, 48 (2021).
- ⁸⁷M. Prugger, L. Einkemmer, and C. F. Lopez, “A dynamical low-rank approach to solve the chemical master equation for biological reaction networks,” *Journal of Computational Physics* **489**, 112250 (2023).
- ⁸⁸I. Kryven, S. Röblitz, and C. Schütte, “Solution of the chemical master equation by radial basis functions approximation with interface tracking,” *BMC Systems Biology* **9**, 1–17 (2015).
- ⁸⁹R. Cai and Y. Lan, “A modified variational approach to noisy cell signaling,” *The Journal of Chemical Physics* **161**, 165103 (2024).
- ⁹⁰J. Ohkubo, “One-parameter extension of the Doi-Peliti formalism and its relation with orthogonal polynomials,” *Physical Review E* **86**, 042102 (2012).
- ⁹¹J. Tindall and M. Fishman, “Gauging tensor networks with belief propagation,” *SciPost Physics* **15**, 222 (2023).
- ⁹²N. Nakatani and G. K. Chan, “Efficient tree tensor network states (TTNS) for quantum chemistry: Generalizations of the density matrix renormalization group algorithm,” *The Journal of Chemical Physics* **138**, 134113 (2013).
- ⁹³G. Evenbly, “Gauge fixing, canonical forms, and optimal truncations in tensor networks with closed loops,” *Physical Review B* **98**, 085155 (2018).
- ⁹⁴R. Alkabetz and I. Arad, “Tensor networks contraction and the belief propagation algorithm,” *Physical Review Research* **3**, 023073 (2021).
- ⁹⁵J. Tindall, M. Fishman, E. M. Stoudenmire, and D. Sels, “Efficient tensor network simulation of IBM’s eagle kicked Ising experiment,” *PRX Quantum* **5**, 010308 (2024).
- ⁹⁶R. Haghshenas, M. J. O’Rourke, and G. K.-L. Chan, “Conversion of projected entangled pair states into a canonical form,” *Physical Review B* **100**, 054404 (2019).
- ⁹⁷Y. Wang, Y. E. Zhang, F. Pan, and P. Zhang, “Tensor network message passing,” *Physical Review Letters* **132**, 117401 (2024).
- ⁹⁸G. Evenbly, N. Pancotti, A. Milsted, J. Gray, and G. K.-L. Chan, “Loop series expansions for tensor networks,” *arXiv preprint* (2024), 10.48550/arXiv.2409.03108.