# UC Irvine
## UC Irvine Previously Published Works

**Title**

Learning dynamic Boltzmann distributions as reduced models of spatial chemical kinetics

**Permalink**

https://escholarship.org/uc/item/8x54v853

**Journal**

The Journal of Chemical Physics, 149(3)

**ISSN**

**Authors**

Ernst, Oliver K
Bartol, Thomas
Sejnowski, Terrence
et al.

**Publication Date**

**DOI**

Peer reviewed

# Learning dynamic Boltzmann distributions as reduced models of spatial chemical kinetics

Oliver K. Ernst, Thomas Bartol, Terrence Sejnowski, and Eric Mjolsness

# Learning dynamic Boltzmann distributions as reduced models of spatial chemical kinetics

Oliver K. Ernst,[1,a)] Thomas Bartol,[2,b)] Terrence Sejnowski,[2,3,c)] and Eric Mjolsness[4,d)]

[1]*Department of Physics, University of California at San Diego, La Jolla, California 92093, USA*
[2]*Salk Institute for Biological Studies, La Jolla, California 92037, USA*
[3]*Division of Biological Sciences, University of California at San Diego, La Jolla, California 92093, USA*
[4]*Departments of Computer Science and Mathematics, and Institute for Genomics and Bioinformatics,
University of California at Irvine, Irvine, California 92697, USA*

Finding reduced models of spatially distributed chemical reaction networks requires an estimation of which effective dynamics are relevant. We propose a machine learning approach to this coarse graining problem, where a maximum entropy approximation is constructed that evolves slowly in time. The dynamical model governing the approximation is expressed as a functional, allowing a general treatment of spatial interactions. In contrast to typical machine learning approaches which estimate the interaction parameters of a graphical model, we derive Boltzmann-machine like learning algorithms to estimate directly the functionals dictating the time evolution of these parameters. By incorporating analytic solutions from simple reaction motifs, an efficient simulation method is demonstrated for systems ranging from toy problems to basic biologically relevant networks. The broadly applicable nature of our approach to learning spatial dynamics suggests promising applications to multiscale methods for spatial networks, as well as to further problems in machine learning. *Published by AIP Publishing.* https://doi.org/10.1063/1.5026403

## I. INTRODUCTION

### A. Model reduction of statistical many-body systems

Master equations are broadly applicable to stochastic systems in biology. For reaction-diffusion systems, the solution to the chemical master equation (CME) fully characterizes the probability distribution over system states and all observables at all times.[1] However, solving the CME for relevant moments is challenging when the interactions of two or more reagents lead to non-linear differential equation systems for the moments and even more challenging when considering spatially distributed systems. A wealth of analytical and numerical approaches has been developed in pursuit of approximate solutions, each of which is optimally suited for a distinct dynamical regime.[2] For example, at the low and spatially heterogeneous concentrations of molecules present in dendritic spines in synapses, particle-based methods may best describe the highly stochastic signaling activity.[3–5] In the larger volumes such as the dendritic shaft, simpler geometries and higher concentrations allow more efficient partial differential equation (PDE) methods. It remains an open problem to develop a modeling framework that is able to flexibly transition across different dynamical regimes or to describe their coexistence in the same spatial domain.

One key problem is that the number of states appearing in the CME increases exponentially with the number of underlying random variables describing the system. This system state space explosion poses a computational challenge for Monte Carlo algorithms such as the popular Gillespie stochastic simulation algorithm (SSA),[6] requiring the sampling of a sufficiently large number of trajectories to estimate observables.

The direct estimation of observables also poses a challenge. Generally, many-body systems result in a hierarchy of moments [analogous to a Bogoliubov–Born–Green–Kirkwood–Yvon (BBGKY) hierarchy[7–9]], where the differential equation for any given moment depends on the current value of higher order ones. This requires the use of a moment closure technique (see Ref. 10 for review), but a poor choice here can unduly restrict modeling of the rich correlation structures available.

Machine learning approaches present an opportune setting for addressing these problems because a central goal of these approaches is the estimation of the structures underlying complex correlations. For example, machine learning has recently been proposed to approximate quantum systems.[11] Previous applications to chemical systems include the predictions of molecular reactions[12] and synaptic activity.[13] However, to our knowledge, no general formulation for learning chemical dynamics exists that incorporates the complex spatial interactions central to many problems in biology. In this work, we present such a framework and derive algorithms for simulating reaction-diffusion systems in continuous space. This has promising general implications for both moment closure and model reduction of the CME and of other generally spatially distributed systems.

a)Electronic mail: oernst@ucsd.edu.
b)Electronic mail: bartol@salk.edu.
c)Electronic mail: terry@salk.edu.
d)Electronic mail: emj@uci.edu.

## B. Inferring Markov random fields for reduced dynamics

Previous work has shown the applicability of machine learning to model reduction. In particular, the "Graph Constrained Correlation Dynamics" (GCCD) framework[10] uses a Markov Random Field (MRF) of plausible state variables and interactions as input, incorporating human expertise into the model reduction process. The probability distribution associated with this MRF is written in a form that separates the time evolution $\mu(t)$ from the graph structure $V_\alpha(s)$,

$$\tilde{p}(s, t; \{\mu\}) = \frac{1}{\mathcal{Z}(\mu(t))} \exp\left[-\sum_\alpha \mu_\alpha(t) V_\alpha(s)\right]. \qquad (1)$$

This time-evolving mean field model can be learned separately at each time-point using the well-known Boltzmann Machine (BM) learning algorithm[14] and then approximated with its own dynamics for $\mu(t)$. For a suitably chosen MRF, the resulting temporal dynamics of $\mu(t)$ result in a large degree of model reduction, as reported in Ref. 10.

In Sec. II, we formalize these ideas and extend to the spatial domain a general variational problem for estimating the dynamical model dictating the time evolution of a Boltzmann distribution. The importance of spatial networks has been widely studied,[15,16] with continued interest in mean-field methods, such as for evolving networks.[17] Our formulation using functionals presents a flexible framework for capturing the dynamics of any desired spatial correlations in the system. This leads to algorithms closely related to a BM, with a modified learning rule for directly estimating the functionals dictating the time evolution of the mean-field model. We anticipate that such an approach is broadly applicable to other spatially organized networks and will have further practical applications in machine learning.

## II. LEARNING ALGORITHMS FOR MODEL REDUCTION

This section is organized as follows. Section II A reviews recent developments that have enabled the derivation of stochastic simulation algorithms for chemical kinetics in the Doi-Peliti formalism and introduces further extensions for describing spatial dynamics. Section II B introduces time-evolving Boltzmann distributions as reduced models, and Sec. II C sets up the associated variational problem for their dynamics. For the case of well-mixed systems, this is solvable by an algorithmic approach, as shown in Sec. II D. Section II E treats some analytic solutions for simple systems used to guess parameterizations needed to algorithmically solve the general spatially heterogeneous case in Sec. II F.

### A. Field theoretic approaches to deriving stochastic simulation algorithms

An equivalent description parallel to the CME is the quantum-field-theoretic Doi-Peliti[18–20] operator algebra formalism (see, e.g., Ref. 21 for review). Extensions to this formalism have developed it as a natural framework for deriving stochastic simulation algorithms of chemical kinetics. These extensions include:

1. The introduction of parameterized objects has generalized bare molecules to allow the description of macromolecular complexes such as phosphorylation states[22,23] or other structures with size, type, and other internal parameters that affect their dynamics.
2. Dynamically graph-linked collections of objects[24] have allowed collections of inter-related objects and extended objects to associate and dissociate according to specified rules and propensities.
3. Differential operators have been introduced that express differential equations and stochastic differential equations,[22,23] as in the Lie Series.[25]

These innovations naturally lead to the rederivation of the popular Gillespie SSA from the CME and from there to extensions to the parametric, graph-matching, and mixed-dynamics cases.[23] Here, we consider further extensions to this formalism to develop *model reduction* techniques for spatial reaction-diffusion systems.

The raising and lowering operators, $\hat{a}$ and $a$, create and destroy identical particles of a single species. For states consisting of a single species distributed on a discrete lattice $|\{n\}\rangle$, where $\{n\}$ describes the occupancy of each lattice site, the action of these operators on the $i$-th lattice site (in some ordering) is

$$\hat{a}_i|\{n\}\rangle = |\{\ldots, n_{i-1}, n_i + 1, n_{i+1}, \ldots\}\rangle,$$
$$a_i|\{n\}\rangle = n_i|\{\ldots, n_{i-1}, n_i - 1, n_{i+1}, \ldots\}\rangle. \qquad (2)$$

Further, they satisfy the Heisenberg algebra commutation relationship $[a_i, \hat{a}_j] = \delta_{ij}$, where $\delta_{ij}$ is the Kronecker delta function. We note that these are different from the ladder operators in quantum mechanics in which $a_i$ is not conjugate to $\hat{a}_i$. However, in the present context, they are of key importance as they capture mass action chemical kinetics.

These operators admit an equivalent generating function representation

$$|\{n\}\rangle \rightarrow \prod_i z_i^{n_i}, \qquad (3)$$

where the product runs over all spatial lattice sites. Then the operators may be represented as

$$\hat{a}_i \rightarrow z_i,$$
$$a_i \rightarrow \frac{\partial}{\partial z_i}. \qquad (4)$$

In the spatially continuous case, let the state of the system be denoted by $|n, \boldsymbol{\alpha}, \boldsymbol{x}\rangle$, consisting of $n$ particles at locations $\boldsymbol{x}$, consisting of $n$ positions in 3D space, with species labels $\boldsymbol{\alpha}$, also of length $n$. The equivalent generating function representation is

$$|n, \boldsymbol{\alpha}, \boldsymbol{x}\rangle \rightarrow \prod_{i=1}^n z(\alpha_i, x_i). \qquad (5)$$

The raising and lowering operators are now

$$\hat{a}(\alpha, x) \rightarrow z(\alpha, x),$$
$$a(\alpha, x) \rightarrow \frac{\delta}{\delta z(\alpha, x)}, \tag{6}$$

where, switching from the discrete to the continuous case, partial derivatives for the annihilation operator turn into functional derivatives. Importantly, the CME

$$\dot{p}(n, \alpha, x, t) = \mathbf{W} p(n, \alpha, x, t) \tag{7}$$

can still be written in an equivalent form where the time-evolution operator $\mathbf{W}$ is a polynomial in the ladder operators, encoding the set of reactions and rates. We make use of these extensions in Secs. III and IV where analytic forms for differential equations of moments are required.

## B. Reduced states in a dynamic Boltzmann distribution

Let $|n, \alpha, x, t\rangle$ denote the true state of the system at time $t$. In the spirit of a MRF, construct states in a coarse-scale model,

$$|\{v_k\}_{k=1}^K, t\rangle = \frac{1}{\mathcal{Z}[\{v_k\}_{k=1}^K]} \sum_{n=0}^{\infty} \sum_{\alpha}$$
$$\times \int dx \exp\left[-\sum_{k=1}^K \sum_{\langle i \rangle_k^n} v_k(\alpha_{\langle i \rangle_k^n}, x_{\langle i \rangle_k^n}, t)\right] |n, \alpha, x, t\rangle, \tag{8}$$

where $\langle i \rangle_k^n = \{i_1 < i_2 < \ldots < i_k : i \in [1, n]\}$ denotes ordered subsets of $k$ indexes each in $\{1, 2, \ldots, n\}$, and $v_k(\alpha_{\langle i \rangle_k^n}, x_{\langle i \rangle_k^n}, t)$ are $k$-particle interaction functions up to a cutoff order $K$. We note that $\{\ldots\}_{k=1}^K$ is used to denote an index-ordered set in this context. This expansion of $n$-body interactions is a specific case of more general dimension-wise decompositions, such as analysis of least variance (ANOVA).[26] The probability of being in a state $|n, \alpha, x, t\rangle$ is given by a dynamic and instantaneous Boltzmann distribution

$$\tilde{p}(n, \alpha, x, t) = \langle n, \alpha, x, t | \{v_k\}_{k=1}^K, t\rangle$$
$$= \frac{\exp\left[-\sum_{k=1}^K \sum_{\langle i \rangle_k^n} v_k(\alpha_{\langle i \rangle_k^n}, x_{\langle i \rangle_k^n}, t)\right]}{\mathcal{Z}[\{v_k\}_{k=1}^K]}. \tag{9}$$

The true probability distribution $p(n, \alpha, x, t)$ evolves according to the CME (7). To describe the time evolution of the reduced model, introduce a set of functionals $\{\mathcal{F}_k\}_{k=1}^K$, forming a differential equation system for the interaction functions $\{v_k\}_{k=1}^K$,

$$\frac{d}{dt} v_k(\alpha_{\langle i \rangle_k^n}, x_{\langle i \rangle_k^n}, t) = \mathcal{F}_k[\{v(\alpha, x, t)\}], \tag{10}$$

where

$$\{v(\alpha, x, t)\} = \left\{ v_{k'}(\alpha_{\langle j \rangle_{k'}^n}, x_{\langle j \rangle_{k'}^n}, t) \ \forall \ \langle j \rangle_{k'}^n : 1 \le k' \le K \right\} \tag{11}$$

denotes all possible $v$ functions evaluated at the given arguments. Here, the right hand side $\mathcal{F}_k$ may be a global functional, in the sense that the arguments $\{v(\alpha, x, t)\}$ are not restricted to the arguments appearing on the left hand side of (10). We

consider particular local parameterizations of this general form in Sec. II F.

In addition to the connection to MRFs, we note several advantages of the form of this reduced model (9) and (10):

1. Since the states $|\{v_k\}_{k=1}^K, t\rangle$ define a grand canonical ensemble (GCE), (9) exactly describes equilibrium systems and is expected to reasonably approximate systems approaching equilibrium.

2. If the interactions between two groups of particles are independent, their joint probability distribution equals the product of their probabilities and their interaction functions $v_k$ in (9) sum. The Boltzmann distribution thus preserves the locality of interactions.

3. A further important result pertains to linearity, as stated in the following proposition.

*Proposition 1. Given a reaction network and a fixed collection of $K$ interaction functions $\{v_k\}_{k=1}^K$, the linearity of the CME in reaction operators $\dot{p} = \sum_r \mathbf{W}^{(r)} p$ extends to the functionals $\mathcal{F}_k = \sum_r \mathcal{F}_k^{(r)}$.*

*Proof.* The dynamic Boltzmann distribution $\tilde{p}(n, \alpha, x, t)$ is a maximum entropy (MaxEnt) distribution, where each interaction function $v_k(\alpha_{\langle i \rangle_k^n}, x_{\langle i \rangle_k^n}, t)$ controls a corresponding moment $\mu_k(\alpha_{\langle i \rangle_k^n}, x_{\langle i \rangle_k^n}, t)\}$ given by

$$\mu_k(\alpha_{\langle i \rangle_k^n}, x_{\langle i \rangle_k^n}, t) = \sum_{n'=0}^{\infty} \sum_{\alpha'} \int dx' \ p(n', \alpha', x', t)$$
$$\times \sum_{\langle j \rangle_k^{n'}} \delta\left(x_{\langle i \rangle_k^n} - x'_{\langle j \rangle_k^{n'}}\right) \delta\left(\alpha_{\langle i \rangle_k^n} - \alpha'_{\langle j \rangle_k^{n'}}\right). \tag{12}$$

Here, $\delta(x)$ denotes a multi-dimensional Dirac delta function. Note that there are $L = \sum_{k=1}^K \binom{n}{k}$ interaction terms and equally many moments they control. Switching to vector notation, let $\mathbf{v}$ of length $L$ denote the interaction functions and $\boldsymbol{\mu}$ denote the corresponding moments.

Relating the interaction functions to the moments constitutes an inverse Ising problem. Let the solution to this problem be

$$v_l = \phi_l(\boldsymbol{\mu}) \tag{13}$$

for some functions $\phi_l$ for $l = 1, \ldots, L$. This solution depends only on the interaction functions and not on the reaction operators appearing in the CME. For a single reaction process, let the differential equations for the moments be $\dot{\boldsymbol{\mu}}^{(r)}$, resulting from $\dot{p}^{(r)} = \mathbf{W}^{(r)} p^{(r)}$, where $\dot{x}$ denotes a time derivative. Taking the time derivatives of both sides of (13) gives

$$\dot{v}_l^{(r)} = \sum_{l'=1}^L \frac{\partial \phi_l(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_{l'}} \dot{\boldsymbol{\mu}}_{l'}^{(r)} \tag{14}$$

for the full network of reactions, then

$$\dot{v}_l = \sum_{l'=1}^L \frac{\partial \phi_l(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_{l'}} \dot{\boldsymbol{\mu}}_{l'} = \sum_r \sum_{l'=1}^L \frac{\partial \phi_l(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_{l'}} \dot{\boldsymbol{\mu}}_{l'}^{(r)} = \sum_r \dot{v}_l^{(r)} \tag{15}$$

gives the desired linearity property.

$\square$

Due to Proposition 1, the functionals $\mathcal{F}$ will be referred to as *basis functionals*. In Sec. III C, the utility of this property is explored further in a machine learning context.

## C. Formulation of general problem to determine functionals governing spatial dynamics

We next formulate a general problem to determine the functionals leading at all times to the MaxEnt Boltzmann distribution. Define the action as the Kullback-Leibler (KL) divergence between the true and reduced models (extending Ref. 10),

$$
S = \int_0^\infty dt \, \mathcal{D}_{\mathcal{KL}}(p||\tilde{p}),
$$

$$
\mathcal{D}_{\mathcal{KL}}(p||\tilde{p}) = \sum_{n=0}^\infty \sum_{\boldsymbol{\alpha}} \int d\boldsymbol{x} \, p(n, \boldsymbol{\alpha}, \boldsymbol{x}, t) \ln \frac{p(n, \boldsymbol{\alpha}, \boldsymbol{x}, t)}{\tilde{p}(n, \boldsymbol{\alpha}, \boldsymbol{x}, t)}.
$$

(16)

Next, we introduce a notation to define a higher-order variational problem. Since the interaction functions are defined specifying by the set of functionals $\mathcal{F}_k[\{\nu(\boldsymbol{\alpha}, \boldsymbol{x}, t)\}]$ for all $k = 1, \ldots, K$, we use the notation $\nu_k[\![\{\mathcal{F}\}]\!]$ to denote that $\nu_k$ is a higher-order generalization of a functional. The action is a functional of the set of all interaction functions, which we denote by $S[\{\nu[\![\{\mathcal{F}\}]\!]\}]$, where $\{x\} = \{x_k\}_{k=1}^K$.

The higher-order variational problem for the basis functionals is given by the chain rule

$$
\frac{\hat{\delta} S[\{\nu[\![\{\mathcal{F}\}]\!]\}]}{\hat{\delta} \mathcal{F}_k[\{\nu(\boldsymbol{\alpha}, \boldsymbol{x}, t)\}]} = \sum_{k'=1}^K \sum_{\boldsymbol{\alpha}'} \int d\boldsymbol{x}' \int dt' \frac{\delta S[\{\nu\}]}{\delta \nu_{k'}(\boldsymbol{\alpha}', \boldsymbol{x}', t')}
$$
$$
\times \frac{\hat{\delta} \nu_{k'}(\boldsymbol{\alpha}', \boldsymbol{x}', t')}{\hat{\delta} \mathcal{F}_k[\{\nu(\boldsymbol{\alpha}, \boldsymbol{x}, t)\}]} = 0,
$$

(17)

where we use the notation $\hat{\delta}$ to denote that this is not an ordinary variational problem, in the sense that a variation with respect to a functional is implied. Equation (17) should therefore be regarded as a purely notation solution, generalizing the well-known chain rule for functionals where a variational derivative is taken of a functional: $\frac{\delta F[G[\phi]]}{\delta \phi(y)} = \int dx \frac{\delta F[G]}{\delta G(x)} \frac{\delta G[\phi]}{\delta \phi(y)}$. The first term is a variational derivative analogous to that appearing in the derivation of the BM learning algorithm,[14] giving

$$
\frac{\hat{\delta} S[\{\nu[\![\{\mathcal{F}\}]\!]\}]}{\hat{\delta} \mathcal{F}_k[\{\nu(\boldsymbol{\alpha}, \boldsymbol{x}, t)\}]} = \sum_{k'=1}^K \sum_{\boldsymbol{\alpha}'} \int d\boldsymbol{x}' \int_0^\infty dt' \Big( \mu_{k'}(\boldsymbol{\alpha}', \boldsymbol{x}', t')
$$
$$
- \tilde{\mu}_{k'}(\boldsymbol{\alpha}', \boldsymbol{x}', t') \Big) \frac{\hat{\delta} \nu_{k'}(\boldsymbol{\alpha}', \boldsymbol{x}', t')}{\hat{\delta} \mathcal{F}_k[\{\nu(\boldsymbol{\alpha}, \boldsymbol{x}, t)\}]} = 0,
$$

(18)

where the moments $\mu$ are defined in (12), with $\tilde{\mu}$ having $p$ replaced by $\tilde{p}$. Next, we consider well-mixed systems where the de-escalation from functionals to ordinary functions $F_k$ makes this problem (18) well-defined. In Sec. II F, we parameterize the functional form of $\mathcal{F}$ to consider spatially distributed systems.

## D. Learning algorithm for reduced dynamics of well-mixed systems in one species

In the case of well-mixed systems in one species, the state of the system is entirely characterized by the number of

individuals $|n, t\rangle$. Dropping the species and position labels in the dynamic Boltzmann distribution gives

$$
\tilde{p}(n, t) = \frac{1}{\mathcal{Z}(\{\nu\})} \exp \left[ -\sum_{k=1}^K \binom{n}{k} \nu_k(t) \right],
$$

(19)

where we use the notation $\{\nu\} = \{\nu_{k'}\}_{k'=1}^K$. The time evolution is now described by basis functions forming the autonomous differential equation system

$$
\frac{d}{dt} \nu_k(t) = F_k(\{\nu\}),
$$
$$
\text{with I.C.:} \quad \nu_k(t = 0) = \eta_k,
$$

(20)

where $F_k$ are now functions rather than functionals $\mathcal{F}_k$. The variational problem (18) for the basis functions becomes

$$
\frac{\delta S[\{\nu[\{F\}]\}]}{\delta F_k(\{\nu\})} = \sum_{k'=1}^K \int_0^\infty dt' \left( \left\langle \binom{n}{k'} \right\rangle_{p(t')} - \left\langle \binom{n}{k'} \right\rangle_{\tilde{p}(t')} \right) \frac{\delta \nu_{k'}(t')}{\delta F_k(\{\nu\})}
$$
$$
= 0,
$$

(21)

where $\langle X \rangle_p(t') = \sum_{n=0}^\infty X p(n, t')$ and similarly for $\tilde{p}$.

The variational term on the RHS of (21) may be determined by a number of methods, including an ordinary differential equation (ODE) formulation derived in Appendix A 1, a PDE formulation derived from applying the chain rule at the initial condition, and using a Lie series approach (see the supplementary material). The first of these and arguably the most practical is

$$
\frac{d}{dt'} \left( \frac{\delta \nu_{k'}(t')}{\delta F_k(\{\nu\})} \right) = \sum_{l=1}^K \frac{\partial F_{k'}(\{\nu(t')\})}{\partial \nu_l(t')} \frac{\delta \nu_l(t')}{\delta F_k(\{\nu\})}
$$
$$
+ \delta_{k',k} \delta(\{\nu\} - \{\nu(t')\}),
$$
$$
\text{with I.C.:} \quad \frac{\delta \nu_{k'}(t' = 0)}{\delta F_k(\{\nu\})} = 0.
$$

(22)

An algorithmic solution to (21) is therefore possible in the form of a PDE-constrained optimization problem: Solve (21) and (22) subject to the PDE-constraint (20). An example algorithm using a simple gradient descent is given by Algorithm 1.

We note the implicit connection between this approach and using Boltzmann machines, such as in GCCD, by the algorithm's objective function. Here, the whole trajectory of moments from stochastic simulations is used to directly estimate time evolution operators, rather than estimating the interaction parameters at each time step. We make this connection explicit in Algorithm 2 in Sec. III C.

Further improvements to Algorithm 1 are possible, such as to replace ordinary gradient descent by an accelerated version, e.g., Nesterov accelerated gradient descent.[27] Furthermore, the wealth of methods available to solve PDE-constrained optimization problems, e.g., adjoint methods,[28] offers rich possibilities for further development.

### 1. Example: Mean of the Galton-Watson branching process

As a simple illustrative example, consider a reduced model that captures the time-evolving mean of the Galton-Watson branching process consisting of the birth process $A \rightarrow A + A$ with rate $k_b$ and death $A \rightarrow \emptyset$ with rate $k_d$.

Algorithm 1. Gradient descent for learning basis functions governing well-mixed dynamics.

---

1: **Initialize**
2:    Grid of values $\{\nu\}$ to solve over.
3:    $F_k(\{\nu\})$ for $k = 1, \ldots, K$.
4:    Max. integration time $T$.
5:    A formula for the learning rate $\lambda$.
6: **while** not converged **do**
7:    Initialize $\Delta F_k(\{\nu\}) = 0$ for all $k$, $\{\nu\}$.
8:    Generate a sample of random initial conditions $\{\eta\}$.
9:    **for** $\eta_i \in \{\eta\}$ **do**
10:       ▷*Generate trajectory in reduced space* $\{\nu\}$:
11:       Solve the PDE constraint (20) with IC $\eta_i$
              for $0 \leq t \leq T$.
12:       Solve (22) for variational term $\delta \nu_{k'}(t)/\delta F_k(\{\nu\})$.
13:       ▷*Sampling step:*
14:       Evaluate moments $\left\langle \binom{n}{k'} \right\rangle_{\tilde{p}(t')}$ of the Boltzmann
              distribution by sampling or analytically.
15:       Evaluate true moments $\left\langle \binom{n}{k'} \right\rangle_{p(t')}$ by stochastic
              simulation or analytic solution.
16:    ▷*Evaluate the objective function:*
17:       Update $\Delta F_k(\{\nu\})$ as the cumulative moving
              average of (21) over initial conditions $\{\eta\}$.
18:    ▷*Update to decrease objective function:*
19:       Update $F_k(\{\nu\})$ to decrease the objective function:
              $F_k(\{\nu\}) \rightarrow F_k(\{\nu\}) - \lambda \Delta F_k(\{\nu\})$.

---

In this case, there are only self-interactions ($K = 1$) described by $\nu(t)$ with basis function $F(\nu(t))$. The dynamic Boltzmann distribution is

$$\tilde{p}(n, t) = \frac{1}{\mathcal{Z}} \exp[-n\nu(t)]. \qquad (23)$$

Using the fact that

$$\langle n \rangle_{\tilde{p}} = \frac{1}{e^\nu - 1} \qquad (24)$$

and from the CME,

Algorithm 2. Boltzmann machine-style learning of dynamics.

---

1: **Initialize**
2:    Initial $\theta^{(r)}$ for all $r$.
3:    Max. integration time $T$.
4:    A formula for the learning rate $\lambda$.
5:    Time-series of lattice spins $\{s\}(t)$ from stochastic
              simulations from some known IC $h_0, J_0$.
6:    Fully visible MRF with NN connections and as many
              units as lattice sites $N$.
7: **while** not converged **do**
8:    ▷*Generate trajectory in reduced space:*
9:       Solve the PDE constraint (52) with IC $h_0, J_0$
              for $0 \leq t \leq T$.
10:   ▷*Awake phase:*
11:      Evaluate true moments $\mu(t), \Delta(t)$ from the
              Stochastic simulation data $\{s\}(t)$.
12:   ▷*Asleep phase:*
13:      Evaluate moments $\tilde{\mu}(t), \tilde{\Delta}(t)$ of the Boltzmann
              distribution by Gibbs sampling.
14:   ▷*Update to decrease objective function:*
15:      Solve (54) for derivative terms.
16:      Update $\theta^{(s)}$ to decrease the objective function
              for all $s$ by taking: $\theta^{(s)} \rightarrow \theta^{(s)} - \lambda \times$ (53).
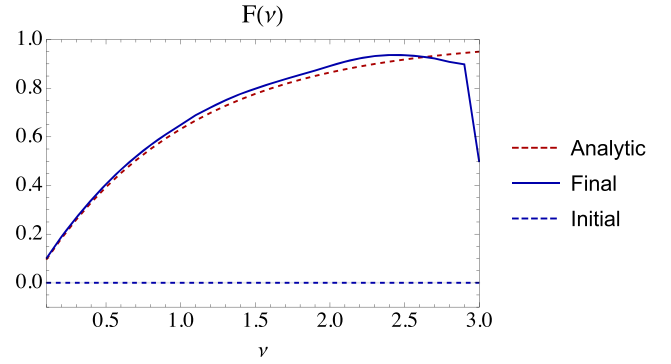
---



FIG. 1. Learned basis function for the simple annihilation process $A \rightarrow \emptyset$ after 40 iterations from a uniform initial condition.

$$\frac{d\langle n \rangle_p}{dt} = (k_b - k_d)\langle n \rangle_p \qquad (25)$$

gives the analytic solution for the basis functions

$$F(\nu) = (k_b - k_d)(e^{-\nu} - 1). \qquad (26)$$

This solution is reproduced using Algorithm 1, as shown in Fig. 1 for $k_d = 3k_b/2$. Here, the solution is constructed on a grid of $\nu \in [0^+, 3.0]$ with spacing $\Delta\nu = 0.1$, with maximum integration time $T = 1$ (arbitrary units). The learning rate is decreased exponentially over iterations to improve convergence. The convergence of the algorithm is shown in Fig. 2.

Figure 3 shows the variational terms (22) evaluated at a single point $\nu = 2$ for several initial conditions $\eta$. These resemble step functions, where the position in time of the step corresponds to the time at which the solution trajectory crosses $\nu = 2$, with value $\delta\nu(t)/\delta F(\nu = 2) \approx 0$ before this time, and non-zero value afterwards. These terms assume different characteristic shapes as the differential equation model (10) is varied, another example of which is shown in Sec. II F 1.

### 2. Example: Two basis functions controlling mean and variance

Consider again the process of Sec. II D 1, but with $K = 2$ basis functions $\nu_1(t)$ and $\nu_2(t)$ controlling the mean and variance in the number of particles. The dynamic Boltzmann distribution is
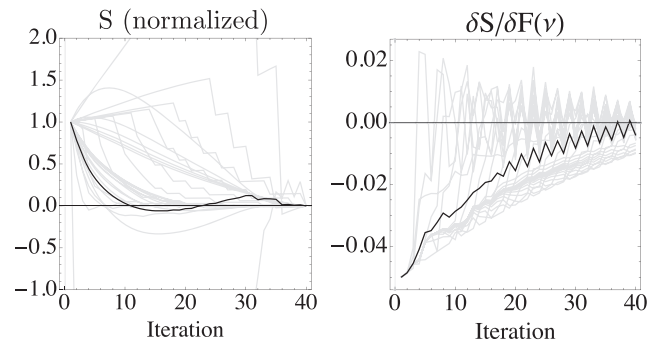


FIG. 2. *Left:* The convergence of the action $S$ as it is minimized over iterations following Algorithm 1. Trajectories (gray) for individual $\eta$, normalized to start at one and end at zero and their mean (black). *Right:* The minimization of the variation in the action $\delta S/\delta F(\nu)$, as a function of the position to vary $\nu$. Trajectories (gray) at each $\nu$ and their mean (black).
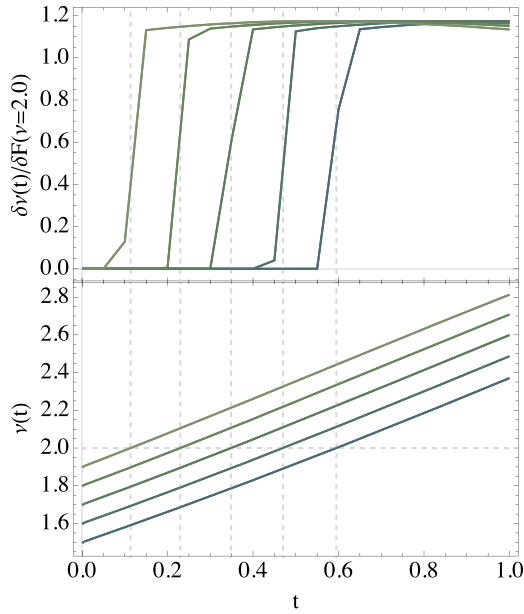
FIG. 3. *Top:* The variational term $\delta v(t)/\delta F(v = 2.0)$ for several initial conditions $\eta = 1.5, \ldots, 1.9$ as a function of time, obtained by solving (22) numerically. *Bottom:* The solution trajectories $v(t)$ starting from these $\eta$. Before the solution trajectory crosses $v = 2$, the value of the variational terms is approximately zero. After this crossing time, the value is non-zero and approximately constant, resembling a step function.

$$\tilde{p}(n, t) = \frac{1}{\mathcal{Z}} \exp\left[-n v_1(t) - \binom{n}{2} v_2(t)\right]. \quad (27)$$

This may be interpreted as a Gaussian distribution in the number of particles, provided we treat $n$ as continuous and extend its range to $\pm\infty$ or consider systems with means far from $n = 0$. In this case, the mean $\mu$ and variance $\sigma^2$ can be related to the interaction functions as $\mu = 1/2 - v_1/v_2$ and $\sigma^2 = 1/v_2$. The differential equations derived from the CME for the moments of this system are

$$\frac{d\mu}{dt} = (k_b - k_d)\mu,$$

$$\frac{d\sigma^2}{dt} = 2(k_b - k_d)\sigma^2 + (k_b + k_d)\mu, \quad (28)$$

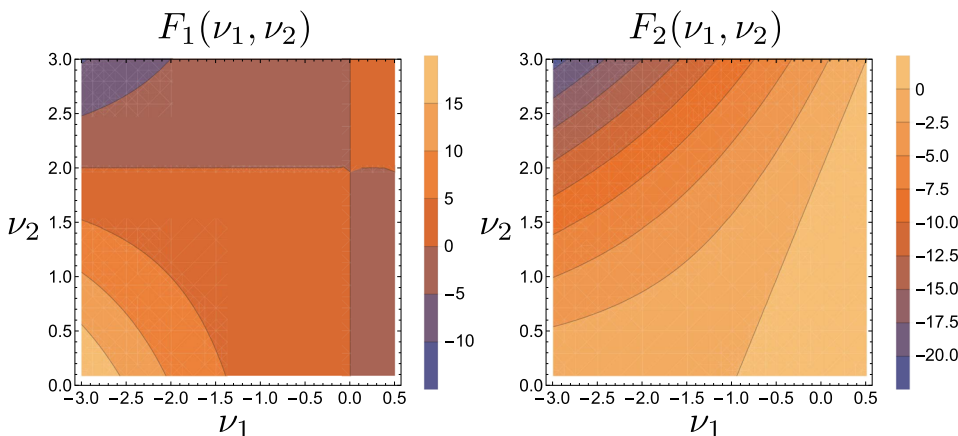which can be converted to analytic solutions for the basis functions

$$F_1(v_1, v_2) = v_1(k_d - k_b + (k_b + k_d)v_1)$$

$$- \frac{v_2}{2}\left(k_b - k_d + (k_b + k_d)v_1\right), \quad (29)$$

$$F_2(v_1, v_2) = -\frac{v_2}{2}\left(k_d(-4 - 2v_1 + v_2) + k_b(4 - 2v_1 + v_2)\right).$$

These are shown in Fig. 4.

Figure 5 shows the variational terms $\delta v_1(t)/\delta F_1(v_1, v_2)$ and $\delta v_2(t)/\delta F_1(v_1, v_2)$, resulting from Algorithm 1 and determined by (22). Interestingly, the self-varying term $\delta v_1(t)/\delta F_1(v_1, v_2)$ more closely resembles the multivariate delta-function appearing in (22), while the cross term $\delta v_2(t)/\delta F_1(v_1, v_2)$ shows a greater temporal memory of the solution trajectory.

### E. Analytic MaxEnt solutions

We next consider special cases where analytic solutions for the basis functionals are possible, to motivate a parameterization leading to a solvable version of the variational problem (18).

#### 1. Gaussian distributions

The well-mixed case (19) is the MaxEnt distribution consistent with $\left\langle \binom{n}{k} \right\rangle_p$ for $k = 1, \ldots, K$. If $K = 2$, then (19) may be interpreted as a Gaussian distribution in continuous $n$, as discussed in Sec. II D. Generalizing these results, the basis functions are generally given by

$$F_1(v_1, v_2) = -v_2 \frac{d\mu}{dt} - v_1 v_2 \frac{d\sigma^2}{dt},$$

$$F_2(v_1, v_2) = -v_2^2 \frac{d\sigma^2}{dt}, \quad (30)$$

where $d\mu/dt$, $d\sigma^2/dt$ are evaluated from the CME and expressed in terms of $v_1, v_2$. Here, a moment closure approximation must be applied if the reactions have a total stoichiometry greater than one on the starting (input) side. For example, the higher order moments appearing in the CME may be approximated by those of the reduced model $\tilde{p}$ and expressed in terms of lower order $\mu, \sigma^2$ following the well-known property of Gaussian distributions. This closure technique is described further in Sec. III A.



FIG. 4. The true basis functions (29) for the Galton-Watson system. *Left:* $F_1(v_1, v_2)$. *Right:* $F_2(v_1, v_2)$. The reaction rates used are $k_d = 3k_b/2 = 1.5$.
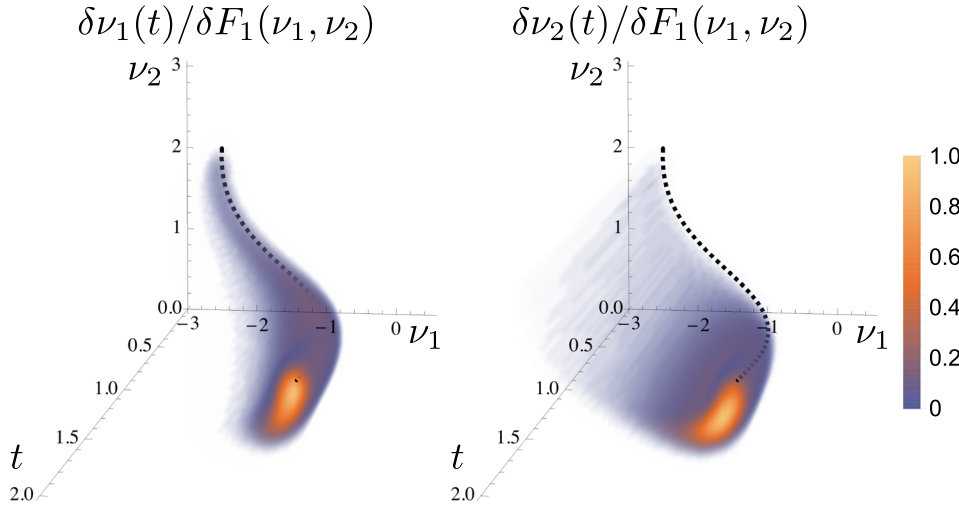
FIG. 5. Two of the four variational terms for the Galton-Watson system using two basis functions. *Left:* $\delta\nu_1(t)/\delta F_1(\nu_1, \nu_2)$. *Right:* $\delta\nu_2(t)/\delta F_1(\nu_1, \nu_2)$. The black dashed line shows the solution trajectory. The initial conditions are $(\eta_1, \eta_2) = (-2.5, 2.0)$, with reaction rates $k_d = 3k_b/2 = 1.5$. The effect of the mixing term $\delta\nu_2(t)/\delta F_1(\nu_1, \nu_2)$ is lower in magnitude but persists longer over time. Note that the absolute magnitude of the spread is related to the approximation chosen for the delta function in (22), a normalized multivariate Gaussian with a variance of 0.1 in both $\nu_1, \nu_2$.

### 2. Diffusion from point source

In the spatial case, consider a diffusion process of a fixed number of particles $n$ with diffusion constant $D$ spreading out from a point source at $x_0$. The analytic solution to the CME is

$$p(\boldsymbol{x}, t) = (4\pi Dt)^{-n/2} \exp\left[-\sum_{i=1}^{n} \frac{(x_i - x_0)^2}{4Dt}\right], \quad (31)$$

reflecting that only self-interactions ($K = 1$) are necessary to describe the process. The reduced model (9) becomes

$$\tilde{p}(\boldsymbol{x}, t) = \frac{1}{\mathcal{Z}} \exp\left[-\sum_{i=1}^{n} \overline{\nu}(x_i, t)\right]. \quad (32)$$

It is straightforward to verify that $p(\boldsymbol{x}, t) = \tilde{p}(\boldsymbol{x}, t)$ if

$$\overline{\nu}(y, t) = \ln\left(1 + \frac{1}{n}\right) + \frac{1}{2}\ln(4\pi Dt) + \frac{(y - x_0)^2}{4Dt}. \quad (33)$$

Consequentially, from $\partial_t\overline{\nu}(y, t)$, the basis functional is

$$F[\overline{\nu}(y, t)] = D\partial_y^2\overline{\nu}(y, t) - D\left(\partial_y\overline{\nu}(y, t)\right)^2. \quad (34)$$

### 3. Unimolecular reaction-diffusion

For reaction networks that involve only diffusion and unimolecular reactions, two key properties hold for the CME solution:

1. Separable spatial and particle number distributions $p(n, \boldsymbol{x}) = p(n)p(\boldsymbol{x})$ where each distribution is normalized $\sum_n p(n) = 1$ and $\int d\boldsymbol{x}\, p(\boldsymbol{x}) = 1$.
2. Independence of spatial distribution $p(\boldsymbol{x}) = p(x_1)p(x_2) \dots p(x_n)$ where each $\int d\boldsymbol{x}\, p(\boldsymbol{x}) = 1$ is normalized. This assumes that initial $p(x_i)$ are independent—otherwise, a fixed mixture of independent components must be considered.

Analogous to the purely diffusive process above, this allows analytic solutions to the inverse Ising problem by imposing these conditions upon the dynamic Boltzmann distribution $\tilde{p}$. Here, we exploit the fact that multiplication of

Boltzmann distributions results in addition of the energy functions.

Introduce a single interaction function $\overline{\nu}(x, t)$ to capture the diffusion process and the usual $\nu_1(t), \dots, \nu_K(t)$ to describe the reactions (for brevity, omit further time arguments in this section). Furthermore, impose the normalization $\int dx\, \exp[-\overline{\nu}(x)] = 1$. The dynamic Boltzmann distribution becomes

$$\tilde{p}(n, \boldsymbol{x}) = \tilde{p}(n)\tilde{p}(\boldsymbol{x}) = \tilde{p}(n)\tilde{p}(x_1)\tilde{p}(x_2)\dots\tilde{p}(x_n),$$

$$\tilde{p}(n) = \frac{1}{\mathcal{Z}}\exp\left[-\sum_{k=1}^{K}\binom{n}{k}\nu_k\right], \quad (35)$$

$$\tilde{p}(x) = \exp[-\overline{\nu}(x)],$$

where the partition function is

$$\mathcal{Z} = \sum_n \int d\boldsymbol{x}\, \tilde{p}(n, \boldsymbol{x}) = \sum_n \tilde{p}(n). \quad (36)$$

The distribution $\tilde{p}(n, \boldsymbol{x})$ is the MaxEnt distribution consistent with the moments $\langle\binom{n}{k}\rangle$ for all $k = 1, \dots, K$, as well as the spatial moment

$$\left\langle\sum_{i=1}^{n}\delta(y - x_i)\right\rangle = \sum_n \int d\boldsymbol{x}\, \sum_{i=1}^{n}\delta(y - x_i)\tilde{p}(n, \boldsymbol{x})$$
$$= \exp[-\overline{\nu}(y)]\langle n\rangle \quad (37)$$

such that the solution to the inverse Ising problem is

$$\overline{\nu}(y) = \ln\left(\frac{\langle n\rangle}{\left\langle\sum_{i=1}^{n}\delta(y - x_i)\right\rangle}\right). \quad (38)$$

The solution for the inverse Ising problem for $\langle\binom{n}{k}\rangle$ is independent of this spatial moment and is analytically possible, e.g., for $K = 1$ or 2, as demonstrated in Secs. II D 1 and II E 1 above.

Taking the time derivatives of these solutions $\dot{\overline{\nu}}$ and $\dot{\nu}_k$ and using the CME to derive differential equations for the moments give the basis functionals. For unimolecular reactions, the diffusion process does not affect the reactions such

that the functional controlling $\bar{\nu}$ is always that of diffusion (34). For example, for a branching random walk consisting of diffusion from a point source and the Galton-Watson process with $K = 2$, the basis are the functional (34) and the functions (29).

## F. Parameterizations for spatially heterogeneous systems

For reaction-diffusion systems that involve reactions which have a total stoichiometry greater than one on the starting (input) side, it generally becomes difficult to analytically solve the inverse Ising problem and identify basis functionals. However, an algorithmic solution remains possible, where we guess a local parameterization of the functional (10) based on the analytic solutions presented above.

Let $\boldsymbol{\beta}, \boldsymbol{y}$ be of length $k$ and use the notation

$$\{\nu(\boldsymbol{\beta},\boldsymbol{y},t)\} = \left\{\nu_{k'}(\boldsymbol{\beta}_{\langle i \rangle_{k'}^k},\boldsymbol{y}_{\langle i \rangle_{k'}^k},t) \ \forall \ \langle i \rangle_{k'}^k : 1 \le k' \le k\right\}. \quad (39)$$

Then choose the *spatially local* parameterization of $\mathcal{F}_k$ in (10),

$$\frac{d}{dt}\nu_k(\boldsymbol{\beta},\boldsymbol{y},t) = F_k^{(0)}(\{\nu(\boldsymbol{\beta},\boldsymbol{y},t)\}) + \sum_{\lambda=1}^{k}\left(F_k^{(1,\lambda)}(\{\nu(\boldsymbol{\beta},\boldsymbol{y},t)\}) \sum_{\langle i \rangle_{\lambda}^k} \sum_{m=1}^{\lambda}\left(\partial_m \nu_\lambda(\boldsymbol{\beta}_{\langle i \rangle_{\lambda}^k},\boldsymbol{y}_{\langle i \rangle_{\lambda}^k},t)\right)^2\right.$$

$$\left. + F_k^{(2,\lambda)}(\{\nu(\boldsymbol{\beta},\boldsymbol{y},t)\}) \sum_{\langle i \rangle_{\lambda}^k} \sum_{m=1}^{\lambda} \partial_m^2 \nu_\lambda(\boldsymbol{\beta}_{\langle i \rangle_{\lambda}^k},\boldsymbol{y}_{\langle i \rangle_{\lambda}^k},t)\right),$$

$$\text{with I.C.: } \nu_k(\boldsymbol{\beta},\boldsymbol{y},t=0) = \eta_k(\boldsymbol{\beta},\boldsymbol{y}), \quad (40)$$

where $\partial_m$ denotes the derivative with respect to the $m$-th component of $\boldsymbol{y}_{\langle i \rangle_{\lambda}^k}$, and $F_k^{(\gamma)}(\{\nu(\boldsymbol{\beta},\boldsymbol{y},t)\})$ for $(\gamma) = (0), (1,\lambda), (2,\lambda)$ are local functions, i.e., functions of the arguments on the left hand side of (40).

The variational problem (18) now becomes

$$\frac{\delta S[\{\nu[\{F\}]\}]}{\delta F_k^{(\gamma)}(\{\nu(\boldsymbol{\beta},\boldsymbol{y})\})} = \sum_{k'=1}^{K}\sum_{\boldsymbol{\beta}'}\int d\boldsymbol{y}' \int dt' \left(\mu_{k'}(\boldsymbol{\beta}',\boldsymbol{y}',t')\right.$$

$$\left. - \tilde{\mu}_{k'}(\boldsymbol{\beta}',\boldsymbol{y}',t')\right)\frac{\delta \nu_{k'}(\boldsymbol{\beta}',\boldsymbol{y}',t')}{\delta F_k^{(\gamma)}(\{\nu(\boldsymbol{\beta},\boldsymbol{y})\})} = 0$$

$$(41)$$

for $(\gamma) = (0), (1,\lambda), (2,\lambda)$, where $\boldsymbol{\beta}', \boldsymbol{y}'$ are of length $k'$.

Analogously to the well-mixed case, it is possible to derive a PDE system governing the variational term $\delta \nu_{k'}(\boldsymbol{\beta}',\boldsymbol{y}',t')/\delta F_k^{(\gamma)}(\{\nu(\boldsymbol{\beta},\boldsymbol{y})\})$. In Appendix A 2, an illustrative example is derived for a diffusion process.

Equations (40) and (41) together form a PDE-constrained optimization problem, which may be solved analogously to Algorithm 1, with additional spatial axes.

### 1. Example: Branching random walk

Consider a branching random walk consisting of the Galton-Watson process and diffusion from a point source with rate $D$ in one spatial dimension and one species. From the true solutions for the basis functionals (34) and (29), use one spatial interaction function $\bar{\nu}(y,t)$ and two purely temporal $\nu_1(t)$, $\nu_2(t)$ and further restrict the parameterization (40) of the basis functionals to be

$$\frac{d\bar{\nu}(y,t)}{dt} = \overline{F}[\bar{\nu}(y,t)] = \overline{F}^{(1)}(\bar{\nu}(y,t))\left(\partial_y \bar{\nu}(y,t)\right)^2$$

$$+ \overline{F}^{(2)}(\bar{\nu}(y,t))\partial_y^2\bar{\nu}(y,t), \quad (42)$$

$$\frac{d\nu_k(t)}{dt} = F_k[\nu_1(t),\nu_2(t)] = F_k^{(0)}(\nu_1(t),\nu_2(t)) \quad (43)$$

for $k = 1, 2$. The variational problem for $\gamma = 1, 2$ is

$$\frac{\delta S}{\delta \overline{F}^{(\gamma)}(\bar{\nu})} = \int dy' \int dt' \ (\mu_1(y',t') - \tilde{\mu}_1(y',t'))\frac{\delta \bar{\nu}(y',t')}{\delta \overline{F}^{(\gamma)}(\bar{\nu})}$$

$$= 0, \quad (44)$$

$$\frac{\delta S}{\delta F_k^{(0)}(\nu_1,\nu_2)} = \sum_{k'=1}^{2}\int dt' \left(\left\langle\!\!\left\langle\binom{n}{k'}\right\rangle\!\!\right\rangle_{p(t')} - \left\langle\!\!\left\langle\binom{n}{k'}\right\rangle\!\!\right\rangle_{\tilde{p}(t')}\right)$$

$$\times \frac{\delta \nu_{k'}(t')}{\delta F_k^{(0)}(\nu_1,\nu_2)} = 0. \quad (45)$$

Differential equations governing the variational terms $\delta \bar{\nu}(y',t')/\delta \overline{F}^{(\gamma)}(\bar{\nu})$ for $\gamma = 1, 2$ are derived in Appendix A 2, given by (A15). Differential equations governing $\delta \nu_{k'}(t')/\delta F_k^{(0)}(\nu_1,\nu_2)$ are given by (22).

The optimization problems (44) and (45) subject to the PDE-constraints (42) and (43) may be solved algorithmically using Algorithm 1 in each $F_k, \overline{F}^{(1)}, \overline{F}^{(2)}$ analogously to the well-mixed case. We note that the true solutions are given by (34) and (29), in particular: $\overline{F}^{(1)} = D$ and $\overline{F}^{(2)} = -D$.

Figure 6 plots the spatial variational terms resulting from the true basis functionals. Here, the reaction rates used are as before $k_d = 3k_b/2 = 1.5$, with a diffusion constant of $D = 1$. Contrary to the well-mixed case, these terms do not resemble step functions but rather exhibit some extended temporal dynamics.
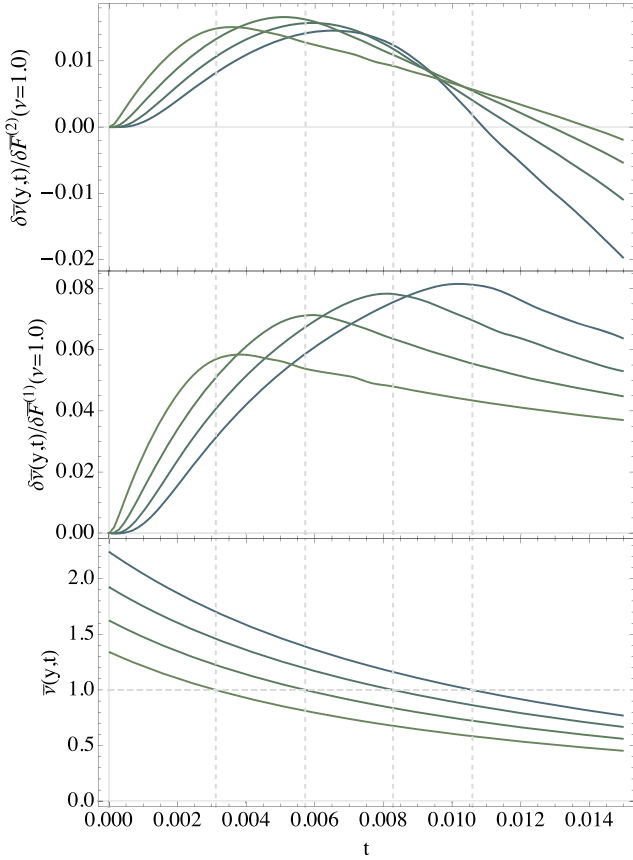
FIG. 6. The solution to Algorithm 1 for a branching random walk with diffusion in 1D. *Top:* The variational term $\delta \overline{v}(y,t)/\delta \overline{F}^{(2)}(v=1.0)$ as a function of time at several spatial locations $y = 0.25, 0.5, 0.75, 1$. Here, $\overline{F}^{(1)} = D, \overline{F}^{(2)} = -D$ are the true solutions. *Middle:* $\delta \overline{v}(y,t)/\delta \overline{F}^{(1)}(v=1.0)$. *Bottom:* The solution trajectories $\overline{v}(y,t)$, starting from a point source. Contrary to the well-mixed case, the variational terms do not resemble step functions at $v = 1.0$ but rather exhibit some extended temporal dynamics.

## III. ESTIMATING EFFECTIVE REDUCED DYNAMICS IN 1D

The PDE-constrained optimization problems given above are the general solution for finding the basis functionals that govern the time evolution of the reduced MaxEnt model. Here, we present a more efficient machine learning approach for learning the basis functions from the solutions of simple, analytically solvable models. In Sec. III A, we present a method for finding such analytic solutions in the discrete lattice limit and present examples for a variety of simple processes in Sec. III B. In Sec. III C, we demonstrate the utility of using such analytic solutions in a Boltzmann machine-like learning algorithm and further in Sec. III D to learn non-linear combinations of solutions using artificial neural networks (ANNs).

### A. Mapping to spin glass systems in 1D

At low particle densities, a feasible model of a reaction-diffusion system in one spatial dimension and one species is a 1D lattice in the single occupancy limit. Let the spin values occupying each lattice site be $s_i \in \{0, 1\}, i = 1, \ldots, N$, denoting the absence or presence of a particle.

The reduced model (9) now becomes the discrete analog. We note that this model is consistent with the continuous

version in some parameter regime where the separation between molecules is large compared to the interaction radius. By including only self-interactions described by an interaction function $h(t)$, and two particle nearest-neighbor interactions $J(t)$, we obtain the well-known Ising model, with partition function

$$\mathcal{Z} = \sum_{\{s\}} \exp\left[ h(t) \sum_{i=1}^{N} s_i + J(t) \sum_{i=1}^{N-1} s_i s_{i+1} \right]. \quad (46)$$

This may be evaluated explicitly using the standard transfer matrix method. In the thermodynamic limit, $\ln \mathcal{Z} \approx \lambda_+^N$ is analytically accessible, where $\lambda_+$ is the largest eigenvalue of the transfer matrix.

The inverse Ising problem has the solution

$$\begin{pmatrix} \left\langle \sum_{i=1}^{N} s_i \right\rangle(t) \\ \left\langle \sum_{i=1}^{N-1} s_i s_{i+1} \right\rangle(t) \end{pmatrix} = \begin{pmatrix} \partial_h \ln \mathcal{Z} \\ \partial_J \ln \mathcal{Z} \end{pmatrix}. \quad (47)$$

Taking the derivatives of both sides of (47),

$$\begin{pmatrix} \frac{d}{dt}\left\langle \sum_{i=1}^{N} s_i \right\rangle \\ \frac{d}{dt}\left\langle \sum_{i=1}^{N-1} s_i s_{i+1} \right\rangle \end{pmatrix} = \begin{pmatrix} \partial_h^2 \ln \mathcal{Z} & \partial_h \partial_J \ln \mathcal{Z} \\ \partial_h \partial_J \ln \mathcal{Z} & \partial_J^2 \ln \mathcal{Z} \end{pmatrix} \begin{pmatrix} \frac{dh}{dt} \\ \frac{dJ}{dt} \end{pmatrix}. \quad (48)$$

The time derivatives of the moments on the left may be obtained directly from the CME $\dot{p} = \mathbf{W}p$ using the Doi-Peliti formalism described in Sec. II A. If the system is *linear*, these may be expressed further in terms of $h, J$ using (47), and the basis functions are given directly by inverting (48). If the system is *non-linear*, the presence of a moment hierarchy requires an approximation in the form of a moment closure technique. Here, we choose to express the higher order moments that appear through the CME in terms of $h$, $J$, which is possible for any higher order moment since the partition function (46) is analytically accessible. As a result of inverting (48),

$$\begin{pmatrix} \tilde{F}_h(h, J) \\ \tilde{F}_J(h, J) \end{pmatrix} = \begin{pmatrix} \partial_h^2 \ln \mathcal{Z} & \partial_h \partial_J \ln \mathcal{Z} \\ \partial_h \partial_J \ln \mathcal{Z} & \partial_J^2 \ln \mathcal{Z} \end{pmatrix}^{-1} \begin{pmatrix} \frac{d}{dt}\left\langle \sum_{i=1}^{N} s_i \right\rangle \\ \frac{d}{dt}\left\langle \sum_{i=1}^{N-1} s_i s_{i+1} \right\rangle \end{pmatrix}, \quad (49)$$

where the RHS has been expressed in terms of $h, J$ as described above, and we use the notation $\tilde{F}_h, \tilde{F}_J$ to indicate that these are generally only approximations to the true basis functions $F_h$, $F_J$ and only exact for systems with closed moments. Effectively, we have replaced the probability distribution $p$ in the CME $\dot{p} = Wp$ by the dynamic Boltzmann distribution $\tilde{p}$ and evaluated the effect of the operator on the RHS on this new distribution. The analytic solution to the 1D inverse Ising problem therefore provides an elegant approach to moment closure (see Refs. 10 and 29 for related MaxEnt approaches to moment closure). Similar extensions to 2D Ising models[30] are likewise possible and possibly to 3D as well.[31]

Furthermore, we note that analogous to the continuous case proven in Proposition 1, the linearity of reaction operators in the CME extends to the basis function approximations $\tilde{F}$ (regardless of whether $\mathcal{Z}$ is analytically accessible as in the 1D case). This requires that the inverse Ising problem has not changed, as discussed further in Sec. III C.
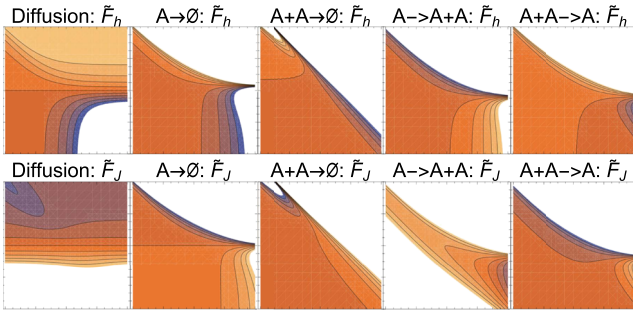
FIG. 7. Basis functions (49) for several simple reaction schemes in one species. Horizontal, vertical axis: $h$, $J \in [-4, 4]$. The magnitudes have been scaled to $[-1, 1]$ from blue to manila since the reaction rate/diffusion constant provides an arbitrary scaling factor.

## B. Analytic approximations to basis functions of simple reaction motifs

Figure 7 shows the basis function approximations calculated using the 1D Ising model (49) for several simple unimolecular reaction processes. Note that the reaction rates/diffusion constants provide an overall multiplicative factor to each process. Computer algebra systems can be used to determine these analytic forms, which contain sums on the order of ten to a hundred terms in length, depending on the operator (see the supplementary material for the code used to generate Figs. 7 and 8).

Generalizing these simple systems, we solve for the basis function approximations of the trivalent reaction $A + B \rightarrow C$ with its reverse process $C \rightarrow A + B$. This process is fundamentally important as a generalization of many simple biochemical processes and has been studied extensively.[23,32] For example, it is the building block of the broadly applicable substrate-enzyme-product (SEP) motif $S + E \rightleftharpoons C \rightarrow P + E$, where $S$, $E$, $P$ denote the substrate, enzyme, and product (see Sec. III D).

In the Ising model, the description of this process involves 9 time dependent interaction functions $h_A$, $h_B$, $h_C$, $J_{AA}$, $J_{AB}$, $J_{AC}$, $J_{BB}$, $J_{BC}$, $J_{CC}$,

$$\mathcal{Z} = \sum_{\{s\}} \sum_{\{\alpha\}} \exp\left[ \sum_{i=1}^{N} h_{\alpha_i}(t)s_i + \sum_{i=1}^{N-1} J_{\alpha_i,\alpha_{i+1}}(t)s_i s_{i+1} \right], \quad (50)$$

where the species label $\alpha_i \in \{A, B, C\}$, and we implicitly note that the sum $\sum_{\{\alpha\}}$ runs only over occupied sites
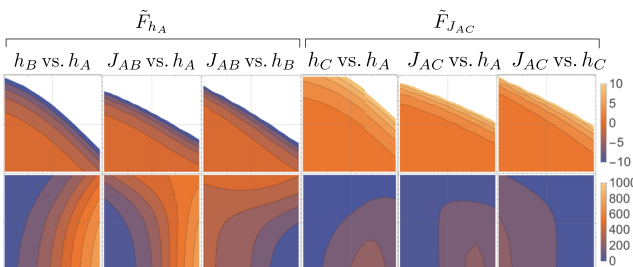


FIG. 8. Basis function approximations $\tilde{F}_{h_A}, \tilde{F}_{J_{AC}}$ corresponding to the forward trivalent reaction $A + B \rightarrow C$ with rate $k = 1$. Each is a 9 dimensional function, of which 2D slices are shown, holding all other parameters at zero. The top row shows the basis function, while the bottom row shows the corresponding moments controlled by these parameters $h_A$, $J_{AC}$. The chain length used is $N = 1000$. The ranges for all horizontal and vertical axes are $[-2, 2]$.

$s_i = 1$. Figure 8 shows several 2D slices of three of the nine basis function approximations for the forward process $A + B \rightarrow C$.

Due to the species labels, (49) leads to analytic expressions containing on the order of hundreds of terms. Here, we use a numerical strategy as described in Appendix B for evaluating the basis functions over the chosen domain. While a computer algebra system may be employed as before, this strategy is computationally faster.

## C. Boltzmann machine-style learning algorithm for dynamics

The basis function approximations derived above constitute a space of possible reduced dynamics. Here, we consider using these analytic insights to describe large spatially distributed reaction networks in 1D. This approach faces two key problems:

1. For non-linear systems, $\tilde{p}$ obeying (49) will over the time diverge from the MaxEnt distribution consistent with the CME moments due the moment closure approximation made. As a fundamental consequence of this moment hierarchy, it is not possible to find exact basis functions over the *entire* interaction parameter space (e.g., $h$, $J$). Another way to see this is that trajectories of the CME system will intersect in $h$, $J$ space.

   However, we postulate that it may be possible to learn approximately well the basis functions for a *single trajectory* (from a single initial condition) which does not self-intersect over some domain. This model may be used for extrapolation with a reasonable accuracy close to the stochastic trajectory.

2. For large reaction networks, the basis functions are generally not linear in the basis functions of individual processes because the collection of interaction functions is not fixed, violating the assumption in Proposition 1. For example, consider the process $A \rightarrow B \rightarrow C$. Here, nine basis functions are required to capture all means and nearest neighbor (NN) correlations such that (49) is nine dimensional. Denote these by $\boldsymbol{b} = \mathbf{A}^{-1}\boldsymbol{m}$ where $\boldsymbol{b}$ denotes the basis functions, $\boldsymbol{m}$ denotes the time evolving moments, and $\mathbf{A}$ denotes the matrix of partition function derivatives.

   Next, consider the separate processes $A \rightarrow B$ and $B \rightarrow C$, described by five basis functions each. Let these be denoted by $\boldsymbol{b}^{(r)} = (\mathbf{A}^{(r)})^{-1}\boldsymbol{m}^{(r)}$ for each of the two reactions $r$. Clearly, not all nine basis functions in $\boldsymbol{b}$ are present in each $\boldsymbol{b}^{(r)}$. Furthermore, for those that are present in both, it is not necessarily true that the $i$-th basis function is expressible as $\boldsymbol{b}_i \neq \boldsymbol{b}_j^{(1)} + \boldsymbol{b}_k^{(2)}$ for appropriate $j$, $k$.

Generally, a reaction network involves more interaction parameters than each of the individual processes such that Proposition 1 does not apply. It is only for a subset of networks, such as reaction networks in one species, where the linearity in the CME extends *exactly* to the basis functions. Regardless, we postulate that many networks may be described *approximately* well by linear combinations of basis functions corresponding to individual processes.

In light of these observations, we return to the variational problem (21) and its PDE-constraint. In the discrete lattice case considered in Sec. III A, it becomes for each $\gamma = h, J$,

$$\int_0^\infty dt' \, (\tilde{\mu}(t') - \mu(t')) \frac{\delta h(t')}{\delta F_\gamma(h, J)}$$
$$+ \int_0^\infty dt' \, (\tilde{\Delta}(t') - \Delta(t')) \frac{\delta J(t')}{\delta F_\gamma(h, J)} = 0, \quad (51)$$

where we have used the notation $\mu$, $\Delta$ to denote the average number of particles, nearest neighbors (NN) over $p$, and similarly $\tilde{\mu}, \tilde{\Delta}$ to denote averages over $\tilde{p}$.

Here, we exploit the analytic results derived above to simplify this problem and derive an efficient Boltzmann-machine type learning algorithm for the dynamics. In particular, we assume that the true basis functions are linear combinations of the approximations derived in Sec. III B above, given by
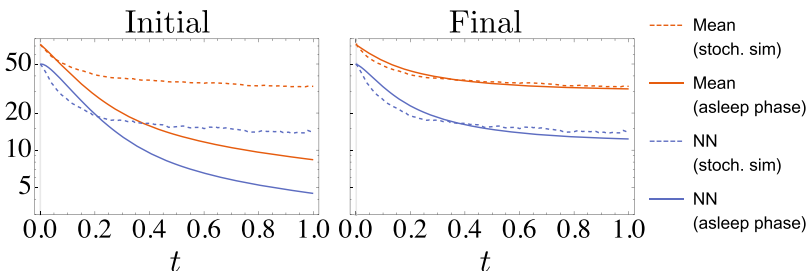
$$\frac{dh}{dt} = F_h(h, J) = \sum_r \theta^{(r)} \tilde{F}_h^{(r)},$$
$$\frac{dJ}{dt} = F_J(h, J) = \sum_r \theta^{(r)} \tilde{F}_J^{(r)}. \quad (52)$$

Here, the reaction rates and diffusion constant are all set to unity such that the coefficients $\theta$ indicate the rates. The variational problem now turns into a regular optimization problem for the coefficients $\theta$ that will yield at all times the MaxEnt distribution consistent with the CME moments. The resulting optimization problem is: Subject to the PDE constraint (52), solve:

$$\int_0^\infty dt' \, (\tilde{\mu}(t') - \mu(t')) \frac{\partial h(t')}{\partial \theta^{(s)}}$$
$$+ \int_0^\infty dt' \, (\tilde{\Delta}(t') - \Delta(t')) \frac{\partial J(t')}{\partial \theta^{(s)}} = 0, \quad (53)$$

where the derivative terms are given by the solution to the ordinary differential equation system

$$\frac{\partial}{\partial t'}\left(\frac{\partial h(t')}{\partial \theta^{(s)}}\right) = \tilde{F}_h^{(s)} + \frac{\partial h(t')}{\partial \theta^{(s)}} \sum_r \theta^{(r)} \frac{\partial \tilde{F}_h^{(r)}}{\partial h}$$
$$+ \frac{\partial J(t')}{\partial \theta^{(s)}} \sum_r \theta^{(r)} \frac{\partial \tilde{F}_h^{(r)}}{\partial J},$$
$$\frac{\partial}{\partial t'}\left(\frac{\partial J(t')}{\partial \theta^{(s)}}\right) = \tilde{F}_J^{(s)} + \frac{\partial h(t')}{\partial \theta^{(s)}} \sum_r \theta^{(r)} \frac{\partial \tilde{F}_J^{(r)}}{\partial h}$$
$$+ \frac{\partial J(t')}{\partial \theta^{(s)}} \sum_r \theta^{(r)} \frac{\partial \tilde{F}_J^{(r)}}{\partial J}, \quad (54)$$

with initial condition $\partial h(0)/\partial \theta^{(s)} = \partial J(0)/\partial \theta^{(s)} = 0$.

Parameter estimation is greatly simpler to solve than the function estimation (51). Furthermore, the variational problem (54) is significantly simplified since $\tilde{F}^{(r)}$ and consequentially its derivatives are analytically accessible. We capitalize upon these practical qualities in Algorithm 2, which solves this problem in a Boltzmann-machine learning style approach.

As an illustrative example, we apply Algorithm 2 to a branching and annihilating random walk (BARW) on a 1D lattice, described by the following processes: $A \to A + A$ with rate $k_b$, $A + A \to 0$ with reaction probability $p_a$ upon encounters, and diffusion. BARWs have been studied in the context of universality classes, in particular, the directed percolation universality class.[33,34]

Stochastic simulations are used to generate training data for this system on a chain of length $N = 100$ for maximum time of $T = 1$ with time step $\Delta t = 0.01$. Here, we follow the numerical procedure described in Ref. 33, with particles attempting to hop to neighboring lattice sites at every time step and reacting with $k_b = 10$ and $p_a = 0.1$. The basis functions used in (52) are those of the three processes present, as shown in Fig. 7. The initial coefficients $\theta_0^{(s)}$ used are $k_b$ for branching, $p_a/\Delta t$ for annihilation, and 10 for diffusion.

Figure 9 shows the moments of the BARW system. Due to the moment closure problem, the system predicted by solving the constraint equations (52) diverges from the stochastic simulations. After running 400 iterations of Algorithm 2, the new coefficients lead to much closer agreement to the true system.

Figure 10 shows the coefficients converge over the iterations. In particular, the effective rates for bimolecular annihilation and branching have decreased, while the effective diffusion constant has increased. Since the final values are sensitive to the initial $\theta_0^{(s)}$ chosen, an $L_2$ regularization term is included in the action. A further constraint in Algorithm 2 to keep $\theta^{(s)}$ positive enforces the connection to effective reaction rates.

### D. ANNs for learning non-linear combinations of basis functions

As a more general approach than linear combinations, we use ANNs (artificial neural networks) to describe non-linear combinations of basis functions. Consider the SEP system diffusing on a 1D lattice described by

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \overset{k_2}{\to} P + E. \quad (55)$$
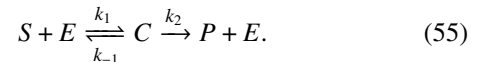
FIG. 9. The 1st and 2nd moments (mean and NN) of the BARW system obtained from stochastic simulation (dashed) and by integrating the PDE constraint (52) and using Gibbs sampling in the asleep phase of Algorithm 2 (solid). *Left:* using initial $\theta_0^{(s)}$ reveals the limitations of moment closure approximation. *Right:* after 400 iterations, the coefficients have adjusted to more accurately capture the true CME dynamics.
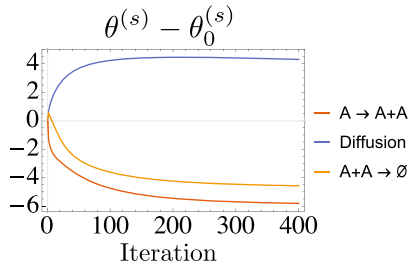
FIG. 10. The coefficients in (52) converging over 400 iterations of Algorithm 2 applied to the BARW system.

The full Ising model for this system consists of four external fields and 10 NN coupling parameters.

Figure 11 shows several moments of this system evolving in time from stochastic simulations. Here, the parameters used are $k_{-1} = 0.1$, $k_2 = 0.5$, bimolecular reaction probability $p_1 = 0.1$, maximum time $T = 2$ with time step 0.01, and lattice length $N = 100$. The system evolves from an initial lattice generated by Gibbs sampling with parameters $h_S = 0.5$, $h_E = 1$, $h_C = -1$, $h_P = -1$ and all NN terms set to zero.

The inputs to the ANN are the basis functions for the three separate processes, each of which belongs to the trivalent reaction motif of Fig. 8, thereby contributing 9 basis functions. Additionally, the two basis functions for the diffusion of each of the four species are included from Fig. 7 for a total of 35 inputs. The other layers in the ANN are two layers of 40 units, and an output layer of 14 units, with tanh activation functions between each layer. One third of the total length $T$ of the timeseries is used for training. These are converted to trajectories in interaction parameter space using Boltzmann machine learning and smoothed using a low-pass filter before being used to evaluate the 35 input basis functions. The outputs to be learned are the time derivatives of these 14 parameters also smoothed by a low-pass filter.

The network learns the dynamics of these parameters to high precision. We infer from the fast training times that the usage of these analytic solutions as input greatly reduces the difficulty of training the network from the interaction parameters directly.

Figure 11(a) shows the extrapolated parameters and corresponding moments, compared to the remaining third of the simulation time. These extrapolations are approximately linear in interaction space and may diverge quickly, such as $h_P$. However, the moments show considerable robustness to these variations. In Fig. 11(b), we compare the symmetric relative error of the interaction parameters and the moments. For each, the error $\varepsilon$ is calculated by comparing the extrapolated values from the ANN and the remaining two thirds of the stochastic simulation data, averaged over the $K = 14$ moments,

$$\varepsilon = \frac{1}{K} \sum_{\alpha=1}^{K} \frac{2|\mu_\alpha^{(\text{ANN})} - \mu_\alpha^{(\text{stoch. sim.})}|}{|\mu_\alpha^{(\text{ANN})}| + |\mu_\alpha^{(\text{stoch. sim.})}|} \tag{56}$$

and similarly for the interaction parameters. The error in the moments grows considerably slower than in the parameters, suggesting that using ANNs for extrapolation is possible for practical applications.

We compare the learned model to a well-mixed model described by mass action kinetics for (55). The reaction rates appearing in the ODEs for this model are fit by non-linear regression to the training data, obtaining $k_1 \approx 0.09$, $k_{-1} \approx 1.34$, $k_2 \approx 0.66$. These parameters are used to further extrapolate the model over the remaining 2/3 of the timeseries, as shown in Fig. 11(a). The symmetric relative error shown in Fig. 11(b) is at all times larger than for the ANN model. The better performance of the ANN method results from it explicitly capturing spatial correlations, rather than implicitly as in the well-mixed model. The tradeoff for this accuracy is the computational cost of training the ANN. However, as typical for ANNs, forward passes through the network are fast, and
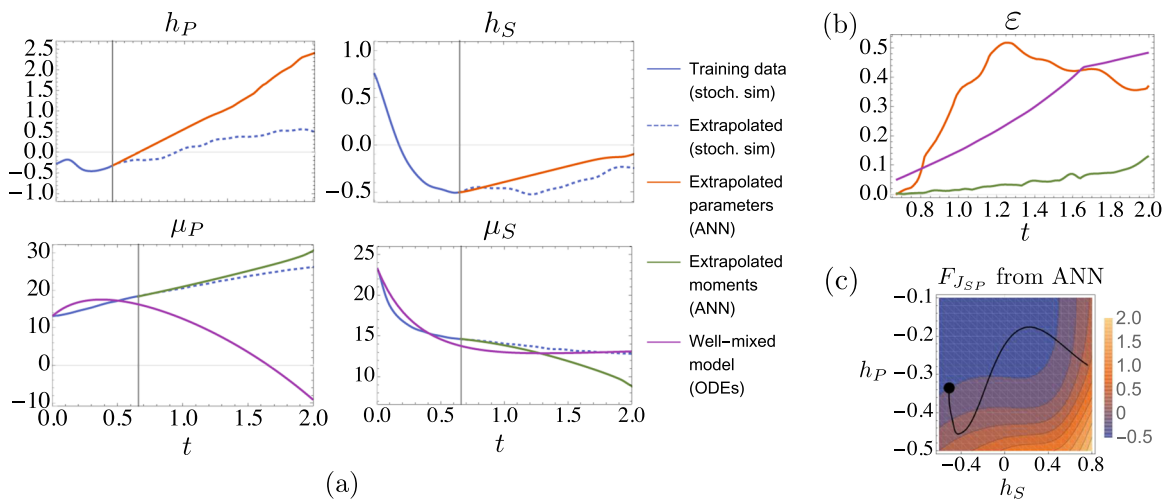


FIG. 11. (a) Trajectories of the SEP system from stochastic simulations, and extrapolated values from the trained ANN, shown for $h_P$, $h_S$ and corresponding moments $\mu_P$, $\mu_S$. Also shown is a well-mixed model described by mass action kinetics, whose reaction rates are fit to the training data. (b) The symmetric relative error $\varepsilon$ between the extrapolated values from the ANN and the stochastic simulations grows faster for the parameters than for the moments, suggesting a stability in the observable quantities of the model to small errors. Here, the error is averaged over all 14 interaction parameters or corresponding moments. The error for the well-mixed model (averaged over the 4 means) grows faster than the error in the moments of the ANN model. The legend is shared with panel (a). (c) $F_{J_{SP}}$ learned by the ANN. A two dimensional slice is shown through this 14 dimensional function. The black line shows the trajectory of the training data, while the dot indicates the evaluation point for this slice, chosen at the end of the training data [gray vertical line in (a)]. All other parameters other than $h_S$, $h_P$ are held fixed at this point.

the cost of solving the resulting differential equations after training is comparable between the two models. This ANN approach is therefore ideal for stochastic spatial modeling applications where the training time is not a bottleneck.

A further feature learned by the ANN is a closure approximation for the dynamics of $J_{SP}$ and its corresponding NN moment. This parameter is not included in any of the basis functions or inputs to the ANN. The basis function learned, $F_{J_{SP}}$, shown in Fig. 11(b), expresses the dynamics of this moment in terms of the interactions made available as input to the network. Similar extensions to higher order moments are likewise possible.

## IV. DISCUSSION AND CONCLUSIONS

This paper presents a new approach to model reduction of spatial chemical systems. Slowly time-evolving MaxEnt models are employed to capture the key correlations in the system. This approach is particularly useful for multiscale problems, where different spatial and temporal correlations become more or less relevant over time to accurately describe the system. For example, in synaptic level neuroscience, the stochastic influx of signaling molecules in the post-synaptic spine produces complex spatial correlations between ion channels and downstream targets, but these are less relevant during quiescent periods. We anticipate that such problems stand to benefit greatly from modeling approaches that are able to adjust which correlations are included to optimize simulation efficiency and accuracy.

A general model that is functional in nature is introduced to describe dynamic Boltzmann distributions. This extends and formalizes ideas originally developed in GCCD in Ref. 10—in particular:

1. A general variational problem has been formulated to determine the functions in the dynamical system controlling the interaction parameters, taking the form of a PDE-constrained optimization problem.
2. The reduced model has been extended to capture spatial correlations, with particular relevance to biological applications. By motivating parameterizations of the functionals from analytically solvable cases, practical optimization algorithms for learning spatial systems are made possible.
3. ANNs have been employed to learn *non-linear* combinations of basis functions, derived for individual reaction processes using computer algebra systems.

Mapping the chemical system onto a spin lattice allows a direct connection to the more traditional formulation of a Boltzmann machine. Here, the connection to the new learning algorithm is evident in (51), and we anticipate this will suggest numerous further applications to diverse areas of machine learning where estimating the dynamics of a time series is required.

The learning rule for dynamic Boltzmann distributions developed in Sec. II applies to systems of any spatial dimension, generally 3D. The results of Sec. III for chemical kinetics on 1D lattices can be extended to treat 2D and 3D systems. For example, for 2D systems with no external field, the basis function approximations may be derived as before since exact

solutions for observables of the Ising model exist.[30] In general, the inverse Ising problem is not solvable. Instead, approximations such as mean-field theory may be used as an input to the ANNs or as initialization for further learning.

Numerous strategies are possible for improving the efficiency of the PDE-constrained optimization problem formulated here, such as adjoint methods.[28] In this work, we have shown that the complexity of this problem can be greatly reduced by instead learning linear and non-linear combinations of analytically accessible approximations. Deconstructing the problem in this way can offer physical insight into a complex reaction system, such as in Sec. III C, where effective reaction rates are learned. Future work in this direction may further explore these principled methods for integrating human intuition with machine inference in the model reduction process.

## SUPPLEMENTARY MATERIAL

See supplementary material for alternate derivations of the differential equation system (22) and for the code used to implement Algorithms 1 and 2.

## ACKNOWLEDGMENTS

## APPENDIX A: DERIVATION OF DIFFERENTIAL EQUATION SYSTEM FOR VARIATIONAL TERM

### 1. Well-mixed case

Consider the differential equation system (20). Represent the solution as a functional of the basis functions $F$ using the notation

$$\nu_{k'}(t') = J_{k'}[\{F\}], \tag{A1}$$

where $\{F\} = \{F_l \,|l = 1, \ldots, K\}$, and $J$ results from solving (20). Further, let $\{J[\{F\}]\} = \{J_l[\{F\}] \,|l = 1, \ldots, K\}$, then (20) is

$$\frac{d}{dt'}J_{k'}[\{F\}] = F_{k'}(\{J[\{F\}]\}). \tag{A2}$$

To find the variational term $\delta\nu_{k'}(t')/\delta F_k(\{\nu\})$, let $F_k \to F_k + \epsilon\eta$ using the notation

$$\{F'\} = \{F_l + \delta_{l,k}\epsilon\eta|l = 1, \ldots, K\}, \tag{A3}$$

then,

$$\frac{d}{dt'}J_{k'}[\{F'\}] = F_{k'}(\{J[\{F'\}]\}) + \delta_{k',k}\epsilon\eta(\{J[\{F'\}]\}). \tag{A4}$$

Differentiating with respect to $\epsilon$ at $\epsilon = 0$ gives

$$\frac{d}{dt'}\left(\frac{dJ_{k'}[\{F'\}]}{d\epsilon}\bigg|_{\epsilon=0}\right) = \sum_{l=1}^{K}\frac{\partial F_{k'}(\{\nu(t')\})}{\partial\nu_l(t')}\left(\frac{dJ_l[\{F'\}]}{d\epsilon}\bigg|_{\epsilon=0}\right)$$
$$+ \delta_{k',k}\eta(\{\nu(t')\}). \tag{A5}$$

Substitute the definition of the functional derivative

$$\frac{dJ_{k'}[\{F'\}]}{d\epsilon}\bigg|_{\epsilon=0} = \int d\nu_1 \ldots \int d\nu_K \frac{\delta\nu_{k'}(t')}{\delta F_k(\{\nu\})}\eta(\{\nu\}) \tag{A6}$$

to obtain (22)

$$\frac{d}{dt'}\left(\frac{\delta \nu_{k'}(t')}{\delta F_k(\{\nu\})}\right) = \sum_{l=1}^{K} \frac{\partial F_{k'}(\{\nu(t')\})}{\partial \nu_l(t')} \frac{\delta \nu_l(t')}{\delta F_k(\{\nu\})}$$
$$+ \delta_{k',k}\delta(\{\nu\} - \{\nu(t')\}). \quad \text{(A7)}$$

### 2. Spatially heterogeneous example: Diffusion in 1D

Consider a diffusion process in 1D, described by a single basis functional parameterized according to

$$\frac{d}{dt'}\nu(y',t') = F[\nu(y',t')]$$
$$= F^{(1)}(\nu(y',t'))\left(\partial_{y'}\nu(y',t')\right)^2$$
$$+ F^{(2)}(\nu(y',t'))\left(\partial_{y'}^2 \nu(y',t')\right). \quad \text{(A8)}$$

Use the functional notation

$$\nu(y',t') = J[\{F\}], \quad \text{(A9)}$$

where $\{F\} = \{F^{(1)}, F^{(2)}\}$ and $J$ results from solving (40), then,

$$\frac{d}{dt}J[\{F\}] = F^{(1)}(J[\{F\}])\left(\partial_{y'}J[\{F\}]\right)^2 + F^{(2)}(J[\{F\}])\partial_{y'}^2 J[\{F\}]. \quad \text{(A10)}$$

To find the variational term $\delta \nu(y',t')/\delta F^{(\gamma)}(\omega)$ for $\gamma = 1, 2$, let $F^{(\gamma)} \rightarrow F^{(\gamma)} + \epsilon\eta$. Use the notation

$$\{F'\} = \{F^{(1)} + \delta_{\gamma,1}\epsilon\eta, F^{(2)} + \delta_{\gamma,2}\epsilon\eta\}, \quad \text{(A11)}$$

then

$$\frac{d}{dt}J[\{F'\}] = F^{(1)}(J[\{F'\}])\left(\partial_{y'}J[\{F'\}]\right)^2 + F^{(2)}(J[\{F'\}])$$
$$\times \partial_{y'}^2 J[\{F'\}] + \delta_{\gamma,1}\epsilon\eta(J[\{F'\}])\left(\partial_{y'}J[\{F'\}]\right)^2$$
$$+ \delta_{\gamma,2}\epsilon\eta(J[\{F'\}])\partial_{y'}^2 J[\{F'\}]. \quad \text{(A12)}$$

Take the derivative with respect to $\epsilon$ at $\epsilon = 0$,

$$\frac{d}{dt}\left(\frac{dJ[\{F'\}]}{d\epsilon}\bigg|_{\epsilon=0}\right)$$
$$= \left(\frac{\partial F^{(1)}(\nu)}{\partial \nu}\left(\partial_{y'}\nu\right)^2 + \frac{\partial F^{(2)}(\nu)}{\partial \nu}\partial_{y'}^2 \nu\right)\left(\frac{dJ[\{F'\}]}{d\epsilon}\bigg|_{\epsilon=0}\right)$$
$$+ \left(\delta_{\gamma,1}\left(\partial_{y'}\nu\right)^2 + \delta_{\gamma,2}\partial_{y'}^2 \nu\right)\eta(\nu)$$
$$+ 2F^{(1)}(\nu)\partial_{y'}\nu\frac{\partial}{\partial y'}\left(\frac{dJ[\{F'\}]}{d\epsilon}\bigg|_{\epsilon=0}\right)$$
$$+ F^{(2)}(\nu)\frac{\partial^2}{\partial y'^2}\left(\frac{dJ[\{F'\}]}{d\epsilon}\bigg|_{\epsilon=0}\right), \quad \text{(A13)}$$

where $\nu = \nu(y',t')$ everywhere. Substituting the definition of the functional derivative

$$\frac{dJ[\{F'\}]}{d\epsilon}\bigg|_{\epsilon=0} = \int d\omega \frac{\delta \nu(y',t')}{\delta F^{(\gamma)}(\omega)}\eta(\omega) \quad \text{(A14)}$$

gives

$$\frac{d}{dt}\left(\frac{\delta \nu}{\delta F^{(\gamma)}(\omega)}\right) = \left(\frac{\partial F^{(1)}(\nu)}{\partial \nu}\left(\partial_{y'}\nu\right)^2 + \frac{\partial F^{(2)}(\nu)}{\partial \nu}\partial_{y'}^2 \nu\right)\left(\frac{\delta \nu}{\delta F^{(\gamma)}(\omega)}\right)$$
$$+ \left(\delta_{\gamma,1}\left(\partial_{y'}\nu\right)^2 + \delta_{\gamma,2}\partial_{y'}^2 \nu\right)\delta(\nu - \omega)$$
$$+ 2F^{(1)}(\nu)\partial_{y'}\nu\frac{\partial}{\partial y'}\left(\frac{\delta \nu}{\delta F^{(\gamma)}(\omega)}\right)$$
$$+ F^{(2)}(\nu)\frac{\partial^2}{\partial y'^2}\left(\frac{\delta \nu}{\delta F^{(\gamma)}(\omega)}\right). \quad \text{(A15)}$$

### APPENDIX B: EVALUATING BASIS FUNCTIONS NUMERICALLY

To compute the basis functions numerically using (49), an efficient method is possible if the eigenvalues of the transfer matrix $M$ are singular. Let the eigenvalues be $\lambda_i$ with corresponding eigenvectors $\boldsymbol{u}_i$. Define

$$p_{ij}(\alpha) = \boldsymbol{u}_i^{\mathsf{T}}(\partial_\alpha M)\boldsymbol{u}_j^{\mathsf{T}} \quad \text{(B1)}$$

for $\alpha = h, J$, where $\partial_\alpha M$ denotes component-wise differentiation of $M$. Also note that $p_{ij}(\alpha) = p_{ji}(\alpha)$ is symmetric. Then the derivatives of the eigenvalues are given by[35]

$$\partial_\alpha \lambda_i = p_{ii}(\alpha),$$
$$\partial_\alpha \partial_\beta \lambda_i = \boldsymbol{u}_i^{\mathsf{T}}(\partial_\alpha \partial_\beta M)\boldsymbol{u}_i + 2\sum_{j\neq i}\frac{p_{ij}(\alpha)p_{ij}(\beta)}{\lambda_i - \lambda_j}, \quad \text{(B2)}$$

for $\beta = h, J$. The principle advantage of this approach lies in the fact that the analytic expressions for $\partial_\alpha M$ and $\partial_\alpha \partial_\beta M$ are simpler to derive than differentiating the analytic expressions for the eigenvalues $\lambda$.

It is now straightforward to numerically evaluate the components $\partial_\alpha \partial_\beta \ln \mathcal{Z}$ of (49) in the thermodynamic limit $\ln \mathcal{Z} \approx N \ln \lambda_+$, where $\lambda_+$ is the largest eigenvalue of the transfer matrix and $N$ the length of the chain.

[1]C. W. Gardiner, K. J. McNeil, D. F. Walls, and I. S. Matheson, J. Stat. Phys. **14**, 307 (1976).

[2]D. T. Gillespie, A. Hellander, and L. R. Petzold, J. Chem. Phys. **138**, 170901 (2013).

[3]T. M. Bartol, D. X. Keller, J. P. Kinney, C. L. Bajaj, K. M. Harris, T. J. Sejnowski, and M. B. Kennedy, Front. Synaptic Neurosci. **7**, 17 (2015).

[4]J. Stiles and T. Bartol, "Monte Carlo methods for simulating realistic synaptic microphysiology using MCell," in *Computational Neuroscience* (CRC Press, 2000).

[5]R. A. Kerr, T. M. Bartol, B. Kaminsky, M. Dittrich, J.-C. J. Chang, S. B. Baden, T. Sejnowski, and J. R. Stiles, SIAM J. Sci. Comput. **30**, 3126 (2008), a publication of the Society for Industrial and Applied Mathematics.

[6]D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).

[7]N. N. Bogoliubov, J. Phys. U. S. S. R. **10**, 265 (1946).

[8]J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946).

[9]J. G. Kirkwood, J. Chem. Phys. **15**, 72 (1947).

[10]T. Johnson, T. Bartol, T. Sejnowski, and E. Mjolsness, Phys. Biol. **12**, 045005 (2015).

[11]G. Carleo and M. Troyer, Science **355**, 602 (2017).

[12]M. A. Kayala, C.-A. Azencott, J. H. Chen, and P. Baldi, J. Chem. Inf. Model. **51**, 2209 (2011).

[13]J. Montes, E. Gomez, A. Merchán-Pérez, J. DeFelipe, and J.-M. Peña, PLoS One **8**, e68888 (2013).

[14]D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, Cognit. Sci. **9**, 147 (1985).

[15]R. Durrett and S. Levin, Theor. Popul. Biol. **46**, 363 (1994).

[16] M. Barthélemy, Phys. Rep. **499**, 1 (2011).

[17] J. P. Taylor-King, D. Basanta, S. J. Chapman, and M. A. Porter, Phys. Rev. E **96**, 012301 (2017).

[18] M. Doi, J. Phys. A: Math. Gen. **9**, 1465 (1976).

[19] M. Doi, J. Phys. A: Math. Gen. **9**, 1479 (1976).

[20] L. Peliti, J. Phys. France **46**, 1469 (1985).

[21] D. C. Mattis and M. L. Glasser, Rev. Mod. Phys. **70**, 979 (1998).

[22] E. Mjolsness and G. Yosiphon, Ann. Math. Artif. Intell. **47**, 329 (2006).

[23] E. Mjolsness, Phys. Biol. **10**, 035009 (2013).

[24] E. Mjolsness, in *Proceedings of the 26th Conference on the Mathematical Foundations of Programming Semantics (MFPS 2010)* [Electron. Notes Theor. Comput. Sci. **265**, 123 (2010)].

[25] M. F. Singer, J. Symb. Comput. **10**, 59 (1990).

[26] M. Griebel, in Proceedings of the Conference on Foundations of Computational Mathematics (FoCM05), Santander, 2005.

[27] Y. Nesterov, Dokl. Akad. Nauk SSSR **269**, 543–548 (1983).

[28] M. B. Giles and N. A. Pierce, Flow, Turbul. Combust. **65**, 393 (2000).

[29] P. Smadbeck and Y. N. Kaznessis, Proc. Natl. Acad. Sci. U. S. A. **110**, 14261 (2013).

[30] L. Onsager, Phys. Rev. **65**, 117 (1944).

[31] S. El-Showk, M. F. Paulos, D. Poland, S. Rychkov, D. Simmons-Duffin, and A. Vichi, Phys. Rev. D **86**, 025022 (2012).

[32] E. Mjolsness and U. Prasad, J. Chem. Phys. **138**, 104111 (2013).

[33] H. Takayasu and A. Y. Tretyakov, Phys. Rev. Lett. **68**, 3060 (1992).

[34] J. Cardy and U. C. Täuber, Phys. Rev. Lett. **77**, 4780 (1996).

[35] P. Lancaster, Numer. Math. **6**, 377 (1964).