



PERSPECTIVE | JANUARY 02 2018

Perspective: Maximum caliber is a general variational principle for dynamical systems FREE

Purushottam D. Dixit; Jason Wagoner; Corey Weistuch ; Steve Pressé ; Kingshuk Ghosh; Ken A. Dill



J. Chem. Phys. 148, 010901 (2018)

<https://doi.org/10.1063/1.5012990>

 CHORUS



View
Online



Export
Citation

Articles You May Be Interested In

Modeling of biomolecular machines in non-equilibrium steady states

J. Chem. Phys. (December 2021)

Multiscale modeling of genome organization with maximum entropy optimization

J. Chem. Phys. (July 2021)

Nanostructured TiO₂ for light-driven CO₂ conversion into solar fuels

APL Mater. (April 2020)

Challenge us.

What are your needs for periodic
signal detection?



Zurich
Instruments

Find out more



Perspective: Maximum caliber is a general variational principle for dynamical systems

Purushottam D. Dixit,¹ Jason Wagoner,² Corey Weistuch,^{2,3} Steve Pressé,⁴ Kingshuk Ghosh,⁵ and Ken A. Dill^{2,6,7}

¹Department of Systems Biology, Columbia University, New York, New York 10032, USA

²Laufer Center for Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, USA

³Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York 11794, USA

⁴Department of Physics and School of Molecular Sciences, Arizona State University, Tempe, Arizona 85281, USA

⁵Department of Physics and Astronomy, University of Denver, Denver, Colorado 80208, USA

⁶Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, USA

⁷Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794, USA

(Received 9 November 2017; accepted 17 December 2017; published online 2 January 2018)

We review here *Maximum Caliber* (Max Cal), a general variational principle for inferring distributions of paths in dynamical processes and networks. Max Cal is to dynamical trajectories what the principle of *maximum entropy* is to equilibrium states or stationary populations. In Max Cal, you maximize a path entropy over all possible pathways, subject to dynamical constraints, in order to predict relative path weights. Many well-known relationships of non-equilibrium statistical physics—such as the Green-Kubo fluctuation-dissipation relations, Onsager’s reciprocal relations, and Prigogine’s minimum entropy production—are limited to near-equilibrium processes. Max Cal is more general. While it can readily derive these results under those limits, Max Cal is also applicable far from equilibrium. We give examples of Max Cal as a method of inference about trajectory distributions from limited data, finding reaction coordinates in bio-molecular simulations, and modeling the complex dynamics of non-thermal systems such as gene regulatory networks or the collective firing of neurons. We also survey its basis in principle and some limitations. © 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5012990>

I. INTRODUCTION: NON-EQUILIBRIUM STATISTICAL PHYSICS—HISTORY AND BACKGROUND

We review here Maximum Caliber (Max Cal), a principle for inferring stochastic dynamics from limited data. As an example, consider a complex biochemical reaction network. Suppose you know the average abundances of some of the species and you know a few average rates. Max Cal gives a way to use that limited information to estimate the probabilities of all the possible reactions within the network. As a principle, Max Cal resembles that of *Maximum Entropy* (Max Ent) used in equilibrium statistical mechanics for predicting the properties of materials and also for inferring probabilities, in general, of states based on incomplete information, often used in model building.

There has long been interest in establishing a variational principle for interpreting the statistical properties of dynamical systems.^{1–4} Just as equilibrium statistical mechanics was formulated around a variational principle of maximizing entropy, a seemingly natural approach for nonequilibrium systems was to formulate corresponding variational principles of *entropy production* or dissipation. Examples include Prigogine’s principle of minimum entropy production⁵ and Onsager’s principle of least dissipation.⁶ For example, the minimum-entropy-production principle (Min EPP) states that the stationary-state distribution of a system interacting with

multiple baths is the one that minimizes the total entropy production rate.⁷ These near-equilibrium principles have been important in providing the quantitative underpinnings for continuum theories of flows of heat, particles, electrical currents, and other conserved quantities. This constitutes today’s field of nonequilibrium thermodynamics, as it is now expressed in standard textbooks.⁸ However, these principles are limited to near-equilibrium processes, and “closeness to equilibrium” is often not well defined. Consequently, the range of application of the theoretical development to model experiments is unclear.

Efforts to generalize these near-equilibrium principles to far-from-equilibrium situations have been largely unsuccessful. But, starting in the 1990s, there has been tremendous success in identifying principles in the form of *fluctuation theorems* (FT). Briefly, fluctuation theorems relate the ratio of probability of observing a given trajectory (say Γ) and its time-reversed counterpart Γ' to a dissipation-based quantity, such as the total entropy production during the trajectory.^{9–16} The Crooks’ FT¹⁷ and its corollary the Jarzynski relation¹⁸ are notable in that they pertain to work distributions. These relationships have practical value, for example, in constructing free energy profiles in single-molecule pulling experiments¹⁹ and in clarifying the relationship between thermodynamics and information.^{20–22} Good reviews of these relationships are given in Refs. 23 and 24.

Here, we describe recent efforts towards a more general variational principle for predicting distributions in dynamical processes. We describe the principle of Maximum Caliber (Max Cal). It applies both near and far from equilibrium. However, the program of developing and testing this principle, particularly in complex situations, remains in early stages and incomplete.

II. VARIATIONAL PRINCIPLES PREDICT PROBABILITY DISTRIBUTIONS IN STATISTICAL PHYSICS

A. The Boltzmann distribution for equilibria as an application of Max Ent

The second law of thermodynamics predicts that matter tends to vary its degrees of freedom to reach the maximum of *entropy* at equilibrium. The second law applies on the macroscale to explain how heat exchanges from hot to cold bodies, particles diffuse from crowded to low-density regions, or pressures tend to equalize. On the microscale, the second law takes a statistical form, as first recognized by Maxwell and Boltzmann.

In particular, the most probable probability distribution of a system at equilibrium is that which maximizes the entropy. This microscopic variational principle proposed by Boltzmann was later generalized by Gibbs. Briefly, the probability distribution $\{p_i\}$ over microscopic states $\{i\}$ of a macroscopic system, say in thermodynamic equilibrium with a surrounding heat bath, can be obtained by maximizing the entropy S ,

$$S = - \sum_i p_i \log p_i, \quad (1)$$

over $\{p_i\}$, subject to a constraint of average energy

$$\sum_i p_i E_i = \bar{E}, \quad (2)$$

where E_i is the energy of microscopic state i , \bar{E} is the average energy of the system, and $k_B T$ are Boltzmann's constant and temperature and that probabilities are normalized quantities

$$\sum_i p_i = 1. \quad (3)$$

The resulting Gibbs-Boltzmann distribution after maximizing the entropy is

$$p_i^* \propto e^{-E_i/k_B T}, \quad (4)$$

where p_i^* are the probabilities that satisfy these conditions. The Gibbs-Boltzmann distribution is at the heart of equilibrium statistical physics. In short, given a model for the energies, E_i of a system, and given a measured value of the first moment of the energy $\langle E \rangle = k_B T$ (which, in this case, means a known value of the temperature T), Eq. (4) predicts the probability distribution over the states $i = 1, 2, 3, \dots$

B. Maximum caliber: A variational principle for dynamical systems

As expressed above, Max Ent appears as a principle of thermal material equilibrium. However, as noted later in

this review, maximum entropy has also been widely developed as a broader principle of inference and model-making for probability distributions in general. It is not restricted to predictions of thermodynamics, materials, or equilibrium. It can also be used in dynamical modeling to infer distributions of path probabilities. When used in dynamics, it has been called Maximum Caliber (Max Cal).²⁵ Max Cal is a trajectory-based method of dynamics.^{26–33} Max Cal seeks the probabilities of paths or trajectories of individual particles, molecules, or agents. It does so by finding the path distribution that maximizes a path entropy, subject to imposed constraints.

Consider a system whose coordinates are collectively described by the variable σ . For simplicity, we assume that the system evolves in a discrete-time and discrete-state fashion. Let $\{\Gamma\}$ be the set of all possible trajectories, individually given by $\Gamma = \{\sigma_{T_i}, \sigma_{T_{i+1}}, \dots, \sigma_{T_f}\}$, that the system can take between time points T_i and T_f . Finally, let p_Γ be the probability distribution defined over the ensemble $\{\Gamma\}$ of paths.

Let $F(\Gamma)$ be a functional defined on the space of paths. Examples of F include the total flux of mass/heat carried by the path, the average dissipation along the path, or the average energy along the path. Analogous to the equilibrium problem, imagine a situation where we want to infer the distribution p_Γ over the paths while constraining the average

$$\langle F \rangle = \sum_\Gamma p_\Gamma F(\Gamma). \quad (5)$$

Note that there are potentially infinitely many probability distributions p_Γ that are consistent with such constraints. Analogous to the equilibrium situation, we maximize the entropy

$$S = - \sum_\Gamma p_\Gamma \log \frac{p_\Gamma}{q_\Gamma} \quad (6)$$

now defined as a *distribution over paths*, subject to constraint in Eq. (5) and normalization. Here, q_Γ is some reference/prior distribution over paths. An implicit assumption in the equilibrium maximum entropy methods is that priors are the same for all micro-states.

The constrained maximization problem is solved by introducing Lagrange multipliers. We write the unconstrained optimization function, popularly known as the caliber C , as

$$C = - \sum_\Gamma p_\Gamma \log \frac{p_\Gamma}{q_\Gamma} - \gamma \left(\sum_\Gamma p_\Gamma F(\Gamma) - \langle F \rangle \right) + \alpha \left(\sum_\Gamma p_\Gamma - 1 \right). \quad (7)$$

In Eq. (7), γ is a Lagrange multiplier that tunes the ensemble average $\langle F \rangle$ and δ ensures normalization. After maximization, we find

$$p_\Gamma = \frac{q_\Gamma e^{-\gamma F(\Gamma)}}{\mathcal{Z}}, \quad (8)$$

where

$$\mathcal{Z} = \sum_\Gamma q_\Gamma e^{-\gamma F(\Gamma)}, \quad (9)$$

a sum of weights over paths, is the dynamical equivalent of a partition function. Equations (8) and (9) are not particularly useful for computations as they stand. They are rendered practical, for example, when the value of γ is known and related to an average flux or rate by the derivative relationship,

$$-\frac{\partial}{\partial \gamma} \log \mathcal{Z} = \langle F \rangle. \quad (10)$$

While the expressions above follow from using constraints with no associated uncertainty (i.e., hard constraints), Ref. 30 discusses generalizations of the results of this section to problems involving constraints with associated uncertainty.

III. AS A BASIC TEST, THE MAX CAL PRINCIPLE GENERATES KNOWN RESULTS OF NONEQUILIBRIUM STATISTICAL PHYSICS

A. Max Cal modeling of non-equilibrium stationary states

First consider the flow of particles between two baths (see Fig. 1).³⁴ On the right is a bath with higher density of particles. It is connected *via* a small conduit (the “system”) to a bath on the left having a lower density of particles. After an initial period, the system reaches a steady state with a constant flux of particles from the right to left baths.

When the system is macroscopic, Fick’s law of diffusion describes both the initial transient dynamics and the system’s steady-state flux. When the number of particles is small, there are frequent violations of the average flux direction; some particles may climb up the concentration gradient. We can compute the distribution of fluxes by invoking the principle of Max Cal over trajectories, constrained by known average fluxes.

Consider a system with time-dependent fluxes (mass, heat, etc.) (see Fig. 1). We want to describe the distribution p_Γ over the trajectories Γ of this system. The ensemble average flux of some quantity a (say heat or mass) over the ensemble of trajectories at a fixed time t is given by

$$J_a(t) = \langle j_{a\Gamma}(t) \rangle = \sum_\Gamma p_\Gamma j_{a\Gamma}(t). \quad (11)$$

In Eq. (11), $j_{a\Gamma}(t)$ is the flux of quantity a at time t in the trajectory Γ .

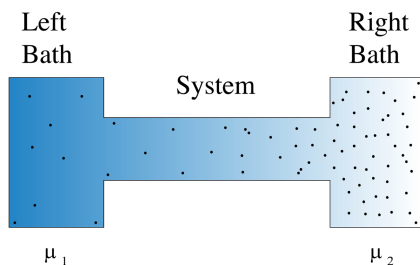


FIG. 1. Consider a system connecting two large baths of particles. Imagine that the density of “stuff” (particles/heat) in the two baths is different leading to a constant flow from the right bath to the left bath. Reprinted with permission from Hazoglou *et al.*, J. Chem. Phys. **143**, 051104 (2015). Copyright 2015 AIP Publishing LLC.

Suppose we want to infer the distribution p_Γ over trajectories that is consistent with two macroscopic fluxes $J_a(t)$ and $J_b(t)$ where a and b characterize two types of fluxes such as particles and heat or two types of particles. Potentially, there are infinitely many distributions p_Γ that are consistent with these two constraints. We choose the one that has the maximum path-entropy (caliber). The caliber is

$$C = - \sum_\Gamma p_\Gamma \log \frac{p_\Gamma}{q_\Gamma} + \sum_t \lambda_a(t) \left(\sum_\Gamma p_\Gamma j_{a\Gamma}(t) - J_a(t) \right) + \sum_t \lambda_b(t) \left(\sum_\Gamma p_\Gamma j_{b\Gamma}(t) - J_b(t) \right) + \alpha \left(\sum_\Gamma p_\Gamma - 1 \right). \quad (12)$$

In Eq. (12), $\lambda_a(t)$ and $\lambda_b(t)$ are time dependent Lagrange multipliers enforcing the constraint regarding known fluxes. Maximizing the caliber gives

$$p_\Gamma = \frac{q_\Gamma}{Z} \exp [\lambda_a(t) j_{a\Gamma}(t) + \lambda_b(t) j_{b\Gamma}(t)], \quad (13)$$

$$Z = \sum_\Gamma q_\Gamma \exp [\lambda_a(t) j_{a\Gamma}(t) + \lambda_b(t) j_{b\Gamma}(t)]. \quad (14)$$

Note that at equilibrium we have $\lambda_{a,b}(t) = 0 \forall t$.

Observable rate quantities, which are ensemble-averaged over pathways, can be obtained as derivatives of the dynamical distribution function,

$$\frac{\partial \log Z}{\partial \lambda_a(t)} = J_a(t), \quad (15)$$

$$\frac{\partial^2 \log Z}{\partial \lambda_a(t) \partial \lambda_b(\tau)} = \langle j_{a\Gamma}(t) j_{b\Gamma}(\tau) \rangle - \langle j_{a\Gamma}(t) \rangle \langle j_{b\Gamma}(\tau) \rangle. \quad (16)$$

We stress that the distribution predicted by Eq. (13) is a “model prediction” based on only average-flux constraints. There may be situations that require different, or additional constraints, in Max Cal. However, we show below that this treatment does capture at least key well-known near-equilibrium results.

B. The Green-Kubo relations from Max Cal

Suppose the system described above is near equilibrium, i.e., the fluxes are small. We expand fluxes at some arbitrary time (say $t = 0$) around Lagrange multipliers $\lambda(t) = 0$. That is, we expand $\langle j_{a\Gamma}(t) \rangle = J_a(t)$ at $t = 0$ to first order around $\lambda_a(\tau)$, $\lambda_b(\tau) = 0$ for all τ in the past,

$$J_a(0) \approx \sum_\tau \left[\left. \frac{\partial \langle j_{a\Gamma}(0) \rangle}{\partial \lambda_a(\tau)} \right|_{\lambda=0} \lambda_a(\tau) + \left. \frac{\partial \langle j_{a\Gamma}(0) \rangle}{\partial \lambda_b(\tau)} \right|_{\lambda=0} \lambda_b(\tau) \right]. \quad (17)$$

At steady state, the Lagrange multiplier does not depend on time. We have $\lambda_a(t) = \lambda_a \forall t$. Thus,

$$J_a(0) \approx \lambda_a \sum_\tau \langle j_{a\Gamma}(0) j_{a\Gamma}(\tau) \rangle|_{\lambda=0} + \lambda_b \sum_\tau \langle j_{a\Gamma}(0) j_{b\Gamma}(\tau) \rangle|_{\lambda=0}. \quad (18)$$

In Eq. (18), λ ’s can be interpreted as the driving forces. Moreover, the cross correlations $\sum_\tau \langle j_{a\Gamma}(0) j_{a\Gamma}(\tau) \rangle|_{\lambda=0}$ quantify “flux” fluctuations at equilibrium and thus can be identified as

the transport coefficients. With this identification of terms, the result is just the Green-Kubo relationship.^{35–37}

C. Onsager's reciprocal relations from Max Cal

Max Cal can also capture Onsager's reciprocal relationships. Onsager considered near-equilibrium systems in which fluxes are linearly proportional to the imposed forces,^{38,39}

$$J_a = L_{aa}\lambda_a + L_{ab}\lambda_b, \quad (19)$$

$$J_b = L_{ba}\lambda_a + L_{bb}\lambda_b. \quad (20)$$

Using Eqs. (16)–(18), we have

$$L_{ab} = \sum_{\tau} \frac{\partial^2 Z}{\partial \lambda_a(0) \partial \lambda_b(\tau)} \Big|_{\lambda=0}. \quad (21)$$

At the same time, we have at $\lambda = 0$ (equilibrium),

$$\begin{aligned} \sum_{\tau} \langle j_{a\Gamma}(0) j_{b\Gamma}(\tau) \rangle|_{\lambda=0} &= \sum_{\tau} \langle j_{a\Gamma}(\tau) j_{b\Gamma}(0) \rangle|_{\lambda=0} \\ &= \sum_{\tau} \frac{\partial^2 \log Z}{\partial \lambda_b(0) \partial \lambda_a(\tau)} \Big|_{\lambda=0} = L_{ba}. \end{aligned} \quad (22)$$

In Eq. (22), we have assumed that both fluxes have the same parity under time reversal (symmetric or anti-symmetric) and invoked microscopic reversibility of trajectories at equilibrium state. As a result, we have $L_{ab} = L_{ba}$ which is exactly Onsager's reciprocal relationship.

D. Prigogine's principle of minimum entropy production can be derived from Max Cal

An interpretation of Prigogine's principle of minimum entropy production is as follows. Consider a near-equilibrium system with two coupled flows. Imagine that one of the flows (flow of a) is driven by a force while the flow of b is unconstrained. The flux of b at steady state is predicted to be that which has the minimum rate of entropy production.^{40,41} First, we consider the standard derivation of the principle. If S is the state entropy, the rate dS/dt , of entropy production, in a system carrying two fluxes J_a and J_b is given by

$$\sigma = \frac{dS}{dt} = J_a \lambda_a + J_b \lambda_b, \quad (23)$$

where λ_a and λ_b are driving gradients. Now, near equilibrium, the Onsager relationships give

$$\sigma = L_{aa}\lambda_a^2 + 2L_{ab}\lambda_a\lambda_b + L_{bb}\lambda_b^2. \quad (24)$$

The minimal entropy production rate with respect to variations in λ_b is given by

$$\frac{\partial \sigma}{\partial \lambda_b} = 2(L_{ab}\lambda_a + L_{bb}\lambda_b) = 2J_b = 0, \quad (25)$$

which correspondingly also predicts that $J_b = 0$.⁴²

The same principle can also be derived from Max Cal. First, we express the caliber as

$$\begin{aligned} C &= - \sum_{\Gamma} p_{\Gamma} \ln \left(\frac{p_{\Gamma}}{q_{\Gamma}} \right) \\ &= \ln Z - \sum_{\Gamma} [\lambda_a(t) J_a(t) + \lambda_b(t) J_b(t)]. \end{aligned} \quad (26)$$

Maximizing C with respect to λ_b ,

$$\begin{aligned} \frac{\partial C}{\partial \lambda_b(\tau)} &= - \sum_{\Gamma} \left[\lambda_a(t) \frac{\partial J_a(t)}{\partial \lambda_b(\tau)} + \lambda_b(t) \frac{\partial J_b(t)}{\partial \lambda_b(\tau)} \right] \\ &\approx -\lambda_a L_{ab} - \lambda_b L_{bb} + \mathcal{O}(\lambda^2) = -J_b = 0. \end{aligned} \quad (27)$$

Thus, the force-flux relationship derived using Max Cal is the same as the entropy-production argument above. Notably, Max Cal makes useful predictions beyond the linear regime that can be explicitly tested. The caliber is maximized when

$$\sum_{\Gamma} \left[\lambda_a(t) \frac{\partial J_a(t)}{\partial \lambda_b(\tau)} + \lambda_b(t) \frac{\partial J_b(t)}{\partial \lambda_b(\tau)} \right] = 0. \quad (28)$$

So, given how J_a and J_b depend on the imposed thermodynamic gradients λ_a and λ_b , solving Eq. (28) gives the gradient λ_b to which the system adjusts itself when it is not constrained.

E. Max Cal gives Fick's law of diffusion, including the "few-molecule" limit

The maximization of path entropy predicts diffusion and the Fokker-Planck equation.^{27,43} Fick's law expresses that the macroscopic average rate of particle diffusion, driven by a gradient, is

$$\langle J \rangle = -D \frac{dc}{dx}, \quad (29)$$

where D is the diffusion constant, $\langle J \rangle$ is the average flux, and dc/dx is the macroscopic gradient of particle concentration. But, we are interested here in more microscopic detail. What is the full rate distribution? For example, what is the second moment of flux, $\langle J^2 \rangle - \langle J \rangle^2$? This is a simple problem that can be solved in various ways, including using the Boltzmann transport equation.⁴⁴ The difference is that Max Cal focuses on distributions of paths, rather than particle concentrations, and is not limited to near-equilibrium assumptions.^{30,45,46} While we do not give the details here, we note that Max Cal gives the full rate distribution, which is predicted to be Gaussian and verified experimentally (see Fig. 2). The dynamical constraint that is imposed here is a

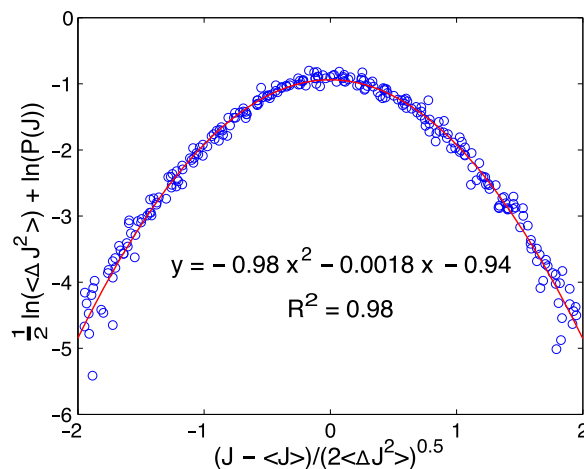


FIG. 2. Max Cal predicts the distribution of microscopic fluxes to be Gaussian (red line). Experimental data are shown as blue circles. See Ref. 46 for details. Reprinted with permission from Seitaridou *et al.*, J. Phys. Chem. B **111**(9), 2288–2292 (2007). Copyright 2007 American Chemical Society.

hopping probability between discretized space which is equivalent to knowing the value of the diffusion constant D .^{45,46} Max Cal's trajectory-based approach predicts the flux distribution, including variances, which cannot be done using Fick's law or the diffusion equation or Boltzmann's transport equation, which are based on concentrations/density. This Max Cal modeling gave the new result that Fick's Law holds even for gradients down to as small as a few molecules. Closely related is a treatment that maximizes the path entropy while constraining the action (A) averaged over all paths.⁴³ Wang has also used that approach to derive other phenomenological laws such as Ohm's law and Fourier's law of heat flow.⁴³ Earlier work²⁷ has shown that path entropy maximization can be used to derive generalized Fokker-Planck equation using first and second moments of coordinate variables—other than action—as constraints.

F. Markov models give the dynamics that maximize the caliber for particular data

Dynamical processes are often modeled as a Markovian, where the probability of transition to a state depends only on its previous state. This is also the basis for hidden Markov models in data analysis. What is the justification for the Markov assumption? It has been found that Markovian dynamics uniquely maximize the caliber, depending on the nature of the form of the measured rate data.^{47,48} Depending on the data that is used as constraints, the probability over paths partitions into a product over transition probabilities depending on the state occupied at the previous time point.^{26,47–51} For example, if constraints are defined using average number of transitions $N_{i,j}$ between two consecutive time steps— i is the state at some instant of time and j is the state at the next time step—we recover a traditional master equation⁴⁷ and Lagrange multipliers relate to reaction rates.⁵⁰ This also applies when the measured jump statistics include multiple time steps⁴⁸ or memory.⁵²

IV. MAX CAL INFERS DYNAMICAL DISTRIBUTIONS FROM LIMITED INFORMATION

A. Inferring the full rate matrix of a network from its state populations

In single-molecule experiments and simulations of biomolecules, the populations at the nodes of a network may be known or assumed, while node-to-node jump dynamics may not be so well characterized. Consider a situation in which a single biomolecule visits a set of conformations or states $\{a, b, c, \dots\}$, including metastable states. The limited information can include the stationary distribution $\{p_a\}$, stationary state averages $\langle E \rangle = \sum p_a E_a$ over nodes, or path ensemble averages $\langle J \rangle = \sum_{a,b} p_a k_{ab} J_{ab}$, where $\{k_{ab}\}$ are the transition probabilities (or equivalently, transition rates, when considering a continuous time Markov process) between these states. Our goal is to infer the set of transition rates $\{k_{ab}\}$ from infinitely many possible Markov processes that would be consistent with such limited data.

What then is the “best” Markov model that we can infer? Recently, Dixit *et al.* used Max Cal to derive a functional form of the rate constants that reproduce a known stationary distribution $\{p_a\}$ and also satisfy a few dynamical path-based constraints $\langle J \rangle$ ^{31–33} (see Fig. 3). Dixit *et al.* predicted that the transition rates at equilibrium (i.e., those that also satisfy *detailed balance*) are proportional to the square root of the stationary populations between those nodes; $k_{ab} \propto \sqrt{p_b/p_a}$ (so, also $k_{ba} \propto \sqrt{p_a/p_b}$). This square-root relationship has also been derived by Bicout and Szabo using a discretized Smoluchowski equation.⁵³ What is the basis for this square-root relationship? It is because Max Cal makes the most parsimonious choice when the rates are underdetermined by the data. Knowing only a state of equilibrium, $K = p_a/p_b = k_{ba}/k_{ab}$, tells us only about the *ratio of rates*, not the individual rates themselves. So, if a model is required to reproduce the equilibrium constant, then the square-root law assigns an equal parsing of that equilibrium constant into forward and reverse processes,

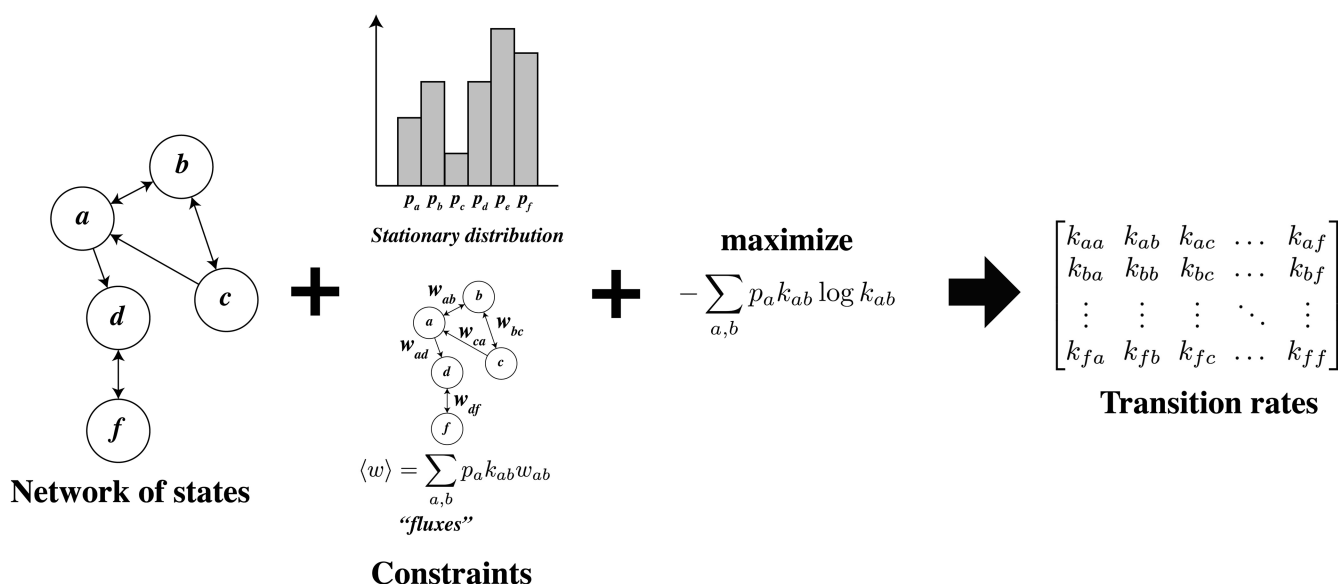


FIG. 3. The Max Cal approach predicts the Markovian transition rates k_{ab} over all routes between states a and b , given only the steady-state populations at the nodes, in addition to an average global rate observable.

$\log k_{ba} = 1/2 \log K$ and $\log k_{ab} = -1/2 \log K$, in the absence of additional information. This square-root relationship has been validated in the dynamics of small peptides and genetic networks.^{31–33}

Wan *et al.* have used Max Cal to modify the Markov state model describing the dynamics among metastable states of wild type peptides to capture the effect of mutations on folding dynamics⁵⁴ and Zhou *et al.* used it to study the effect of protein-protein interactions on transitions among the metastable states of proteins.⁵⁵

B. Finding good reaction coordinates in molecular simulations

When performing molecular simulations of chemical or physical processes, it is often challenging to learn the dominant reaction paths. These are of interest because reaction paths are the essential coordinates which define chemical “mechanisms.” Several methods have been developed to find reaction paths,⁵⁶ for example, metadynamics is an adaptive simulation technique to explore free energy landscapes along a few collective variables, such as reaction coordinates. However, the challenges are to choose good collective variables and to identify “slowly changing” collective variables that are suitable for sampling rare events, for example, in barrier crossing. Based on the Max Cal method of Dixit *et al.*, Tiwary *et al.* developed novel metadynamics-based algorithms for fast identification of reaction coordinates by maximizing timescale separation.^{57–60} In this way, Tiwary shed light on the molecular mechanisms orchestrating unbinding of streptavidin from the biotin-streptavidin complex.⁶⁰

C. Modeling networks that are biochemical or social

Cells in an isogenic population often have widely fluctuating protein copy numbers due to stochastic gene expression.⁶¹ The noisy time profiles of protein expression are the stochastic trajectories to which Max Cal can be applied to infer a predictive model (see Fig. 4). These problems are ubiquitous in genetic networks and are particularly challenging when there is feedback. Feedback often involves interactions between multiple species not directly observable in experiments. Typically, only one or two types of proteins can be seen—far too few

compared with the actual number of molecular actors involved. These are examples of underdetermined problems with limited information. How do we infer microscopic parameters for these models? In normal, “forward,” modeling in physics, a model is assumed and dynamical equations of motion are written. Predictions are then made and compared to the data. For the example of the toggle switch (TS), you could start with master equations describing the dual-negative feedback loop and subsequently make predictions on the basis of this model. The challenge with this approach is that the model parameters are either introduced in an *ad hoc* fashion or adjusted to fit the data. Thus predictions are sensitive to the choice of parameters.

By contrast, in the inverse modeling approach, the goal is to learn a model from the data, with otherwise minimal assumptions. In this case, using Max Cal with observed particle-number fluctuations is sufficient to predict the dynamics, minimizing assumptions and adjustable parameters (see Fig. 4). This was demonstrated in two synthetic gene circuits: (i) in a positive feedback (PF) circuit in which a gene auto-activates itself⁶² and (ii) in genetic toggle switch (TS) in which two genes repress each other.^{30,63} The information input was (1) protein synthesis, (2) protein turnover, and (3) effective coupling between species (positive feedback in the case of PF and negative feedback in the case of a TS; see Fig. 4 for the PF circuit). The success of the method was shown using synthetic time traces generated from known models (with seven parameters for PF and four parameters for the TS) using a Gillespie simulation. Max Cal was found to capture the same qualitative and quantitative information with few Lagrange multipliers. Moreover, Max Cal inferred underlying rates accurately and produced an effective feedback parameter.⁶² The Max Cal framework works directly on the trajectory space and is readily amenable to further including raw trajectories that are likely to be given in fluorescence (observed in typical experiments) instead of particle numbers.⁶² This approach can help analyze raw noisy fluorescence trajectories instead of protein numbers.

More complex biochemical networks can also be treated by Max Cal. For many biochemical networks (for example, bacterial chemotaxis⁶⁴ and mammalian growth factor⁶⁵), the network structure is known, but the individual rate

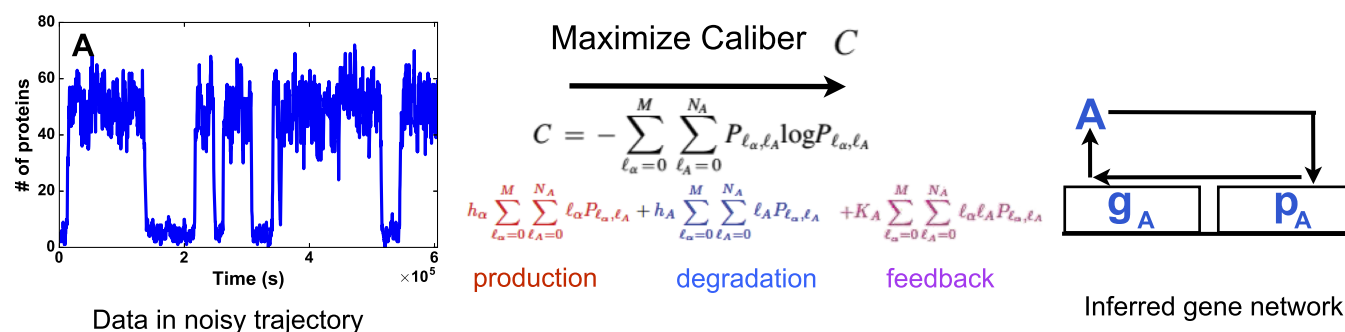


FIG. 4. Max Cal can be used to infer details of the underlying gene network using experimentally measured noisy gene expression trajectory. Max Cal develops a model by maximizing the caliber (C) constructed from the path entropy and three constraints of production, degradation, and feedback for a single gene auto-activating circuit. Details of the model and methodology can be found in Ref. 62. Reprinted with permission from Firman *et al.*, Biophys. J. **113**(9), 2121–2130 (2017). Copyright 2017 Elsevier.

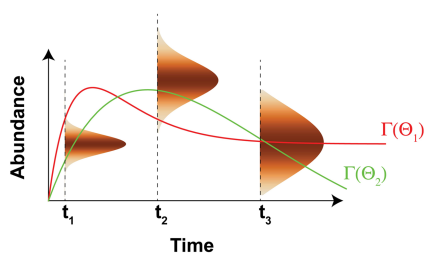


FIG. 5. Cell-to-cell variability in the abundance of a chemical species measured at 3 time points t_1 , t_2 , and t_3 . From the experimental data, we determine the fraction ϕ_{ik} of cells that populate the k th abundance bin at the i th time point by binning the cell-to-cell variability data in B_i bins. The signaling network model can predict trajectories of species abundances as a function of network parameters [see $\Gamma(\Theta_1)$, $\Gamma(\Theta_2)$, and $\Gamma(\Theta_3)$]. Dixit *et al.*^{66,67} derived the analytical expression for the parameter distribution $P(\Theta)$ as well as the distribution over trajectories $P(\Gamma(\Theta))$ consistent with experimentally estimated populations.

parameters are not. And, the challenge is to infer them because data collected on only a few species at a few experimental time points and the parameters themselves vary substantially from cell to cell in a population (called *extrinsic variability*). Dixit *et al.*^{66,67} have developed a Max-Cal-based framework to infer probability distributions over network parameters and species abundance trajectories in biochemical networks from experimental data. They inferred the distribution $P(\Theta)$ of network parameters Θ of a biochemical network as well as the distribution $P[\Gamma(\Theta)]$ over trajectories of species abundances $\Gamma(\Theta)$ that are consistent with histograms of experimentally measured cell-to-cell variability, for example, by flow cytometry or immunofluorescence (see Fig. 5). They showed that the framework can be used to quantify extrinsic noise in both stochastic gene expression networks⁶⁶ as well as in signaling networks.⁶⁷

Some probability distributions are not exponential and have power-law tails, particularly in social and economic systems. They include distributions of incomes, wealth, city sizes, journal citations, terrorist attacks, protein-protein interactions, and others. An entropy variational principle can be applied in some of these cases too, but they use energy-like cost functions that are non-extensive.^{68,69}

V. THE FOUNDATIONS OF MAX ENT AND MAX CAL: MODEL-MAKING AND PHYSICS

What is the justification for Max Cal? Why should populations of fluxes be computable by maximizing a path entropy, subject to a few dynamical constraints? And, when might it fail? We view the principle of Max Cal as having the same foundation and justification as other entropy variational principles, but simply applied to pathways, rather than states. Below, we divide the history of different justifications for this principle into 3 eras: (1) the Boltzmann-Gibbs idea (starting, the late 1800's)^{70,71} that statistics and probabilities can be leveraged to compute macro thermo from the micro of the mechanics of particle collisions, then (2) the Jaynes-Shannon idea (starting around 1950's)^{72,73} that entropy maximization is an informational procedure in which you aim to “minimize your ignorance with respect to all except what the data tell you explicitly,” then (3) Shore-Johnson (1980)⁷⁴ and its

interpretation³⁰ that entropy variation is about making self-consistent models that satisfy the laws of probability.

A. Boltzmann and the mechanics of collisions

How can you predict the macroscopic observables of gases and materials from the molecular properties of materials? Boltzmann and others sought answers in the mechanics and collisions in gases. Today's molecular simulations typically apply Newton's equations of motion, even for complex systems like protein molecules in water, and then collect statistical properties. However, Boltzmann's great insight was that while he could calculate, in principle, all the properties of gases, collision by collision, fully mechanically, he realized that it is rarely ever practical or useful.

Rather, his view was that one could capture all the micro-details of the collisions, without computing them, by thinking of them in probabilistic terms instead. This is sufficient to describe equilibrium and evolution of average properties. This key insight was captured by a single relation, $S = k \ln W$, that harnesses “statistics” as a practical approximation for computing the “mechanics.” This same idea is expressed equivalently as $S = -\sum p_i \ln p_i$ that holds in all different ensembles. S , on the left, is related to the macroscopic property (heat). W , on the right, is related to the microscopic details yielding number of arrangements/multiplicity leading to probabilities and distributions. The right-hand side is where the statistical interpretation or approximation of mechanics gets embodied. The logarithm can be intuitively understood assuming entropy (like energy) being extensive should add, while multiplicities—like probabilities—should multiply. This can be then satisfied by $S = k \ln W$. The idea that free energies are related to populations through $\Delta F_i = -kT \ln p_i$, where p_i is the population of state i , derives from Boltzmann's $S = k \ln W$ combined with the basic thermodynamic relation $F = U - TS$. This joining together of statistics with mechanics is the basis for practical uses of statistical mechanics today.

However, as noted in a more extensive history elsewhere,³⁰ Boltzmann's arguments—which converted time-dependent mechanical trajectories to static probability distribution functions—required the assumption of the *ergodic hypothesis*. However, even as early as the 1890's, ergodicity was argued to be problematic, by Loschmidt, Zermelo, Poincare, and others. As a modern example of the problem, if you simulate a protein molecule in water (even around its native state), it is usually essentially impossible to sample configurations sufficiently to be certain of seeing all states in proportion to their populations.⁷⁵ Lack of convergence is just another term for lack of ergodicity. Gibbs' method, described below, avoided the ergodicity premise.

B. Gibbs and the ensemble method

Willard Gibbs generalized Boltzmann's work so that it could apply to systems of interacting particles. Gibbs devised the logic of *ensembles* and computed probabilities using multinomial statistics, essentially envisioning configurations of a system as samples of a distribution, just like in a dice problem. There are two issues with Gibbs's method. First, the ensemble is an artificial construct to motivate probabilities using a

frequentist interpretation. Second, his predictions were contingent upon an axiom of *equal a priori probabilities* (EAP) over all states of the system regardless of their energy.

C. Shannon's information entropy and Jaynes' predictive statistical mechanics

In 1957, Jaynes expressed the view that maximization of the entropy function, subject to a first-moment constraint (on the average energy $\langle E \rangle$), embodies the idea of being maximally ignorant about all the details of the distribution except for that which is needed to satisfy the constraint $\langle E \rangle = k_B T$.⁷³ Jaynes' view was a major shift from thinking about statistical mechanics as a physical theory to thinking about statistical mechanics as *information-theoretic*. On the one hand, this solved key problems—of no longer requiring ergodicity or the EAP. Jaynes' derivation of the Boltzmann distribution law is also very attractive for didactics; it is simple to teach. Furthermore, it no longer relied on the frequentist argument. However, there were objections to Jaynes' view too. If entropy is about ignorance, then whose ignorance? And, why is ignorance relevant at all for problems in physical sciences where entropy can be measured.

D. The Shore-Johnson case about entropy variation

In 1980, Shore and Johnson (SJ) proposed a different, and axiomatic, argument. They showed that maximizing the Boltzmann-Gibbs entropy is the only procedure that draws inferences about probability distributions that are consistent with the basic rules of probability. SJ interpreted statistical physics³⁰ as the procedure that maximizes $S = -\sum p_i \ln p_i$ subject to constraints when presented with some sort of a model of physical reality that has unknown parameters or features, and the goal is to learn from given data to ultimately infer the posterior distribution that obeys the laws of probability. It asserts that entropy maximization is the only self-consistent logical pipeline from premises (model, data, and a prior distribution) to conclusions (a posterior distribution).

Shore and Johnson established entropy variational principles on a strong footing by showing it is all about self-consistent inferences about models, not about ignorance. An important value of the inference-based view is that it gives insight into why entropy variation is such a universal idea beyond just material equilibria. This allows application of statistical mechanical ideas in diverse areas well outside of the scope of material physics, for example, in ecology,⁷⁶ sociology,⁶⁹ and biology.⁷⁷ And, it is not limited to equilibria. It follows that paths are legitimate objects of probabilities, over which entropies can be maximized, in order to infer dynamics. This is the basis for Max Cal.

VI. THE ISSUES, CAVEATS, AND CHALLENGES WITH MAXIMUM CALIBER

It follows from Shore and Johnson that Max Ent or Max Cal is a general principle for making self-consistent probabilistic models, not principles of physics, *per se*. If a model is being used that does not accurately represent the physical situation being treated, then inferences from that model can be

wrong. In such cases, what is flawed is not the logical pipeline of entropy maximization, but rather the premises that are put into that pipeline.⁷⁸ This can be either the ignorance about the prior or the lack of data used as constraints. Such instances can be ways to learn from, and improve, models of real world situations by gathering more data.

A. Accounting for measurement errors in the constraints

One persistent question is whether entropy maximization can incorporate errors in measurements. In short, both Max Ent and Max Cal can include them. First, we discuss how the question arises. Within the Bayesian framework, entropy maximization can be interpreted as a maximum likelihood problem.³⁰ Suppose we want to infer a distribution $p(\Gamma)$ over trajectories Γ of a system. Suppose we are constraining an average \bar{f} of a path-based quantity $f(\Gamma)$. Now, suppose that there are no errors in the estimate of \bar{f} , we know that value precisely. Then, we can recast the Max Cal problem as the following maximum likelihood problem:

$$\text{maximize } e^{\alpha S[p(\Gamma)]} \times \delta \left(\sum_{\Gamma} p(\Gamma) f(\Gamma) - \bar{f} \right). \quad (30)$$

In Eq. (30), the first term $e^{\alpha S[p(\Gamma)]}$ is the *entropic prior*^{30,79} distribution that weighs different candidate distributions $p(\Gamma)$ according to their entropy. The second term is the likelihood. The Dirac delta function here enforces the idea that the estimate of \bar{f} has no errors or variance associated with it. Now on the other hand, consider a situation in which we know that there is uncertainty and it is represented by a standard deviation σ_f associated with \bar{f} . In this case, the maximum likelihood problem now becomes

$$\text{maximize } e^{\alpha S[p(\Gamma)]} \times e^{-\frac{(\sum_{\Gamma} p(\Gamma) f(\Gamma) - \bar{f})^2}{2\sigma_f^2}}. \quad (31)$$

In Eq. (31), a Gaussian error distribution is taken into account. Moreover, this Bayesian viewpoint not only allows us to find the distribution $p(\Gamma)$ having the maximum likelihood but also allows us to express the full posterior distribution $\phi[p(\Gamma)]$ over the distributions $p(\Gamma)$. We have

$$\phi[p(\Gamma)] \propto e^{\alpha S[p(\Gamma)] - \frac{(\sum_{\Gamma} p(\Gamma) f(\Gamma) - \bar{f})^2}{2\sigma_f^2}}. \quad (32)$$

B. What constraints are appropriate? Some situations entail size dependence and some do not

What are the requirements for choosing constraints for Max Ent and Max Cal? In some classes of problems, only first moments are used. For example, for macroscopic equilibria, the canonical ensemble is predicted by using a first-moment constraint, namely, the average energy $\langle E \rangle$. In those cases, higher moments are not imposed. For other situations, higher moments are appropriate (such as modeling gene networks). How can we rationalize these differences?

At the most basic level, Shore and Johnson showed that any constraint is suitable for the Max Ent/Max Cal inference that is linear in the probabilities. But, this is not very

restrictive. A more precise further division is useful, into *scalable* (*size-dependent*) systems, examples of which include equilibria or kinetics of material systems composed of atoms or molecules), vs. *non-scalable* problems, where the system size is not a relevant concept (typically problems of data inference or model-making).

Nonscalable problems are typically situations involving making inferences or models: we start with some prior distribution function, we then learn some data, and we want to infer the new posterior distribution. In these nonscalable inference problems, the systems size is not a relevant variable. In non-scalable problems, we are free to use whatever experimental information we have to make inference. In such cases, we can use first or higher moments or other knowledge.^{62,63,67,77,80} If constraints beyond first moment are negligible it will be seen from the data which will make Lagrange multipliers vanishing for these higher moments. However we cannot assume that apriori.

In scalable problems, however, we have the extra predictive power to impose the first moment and ignore all the higher moments at the onset of model building. This is because of the particular nature of decomposibility into subsystems that those systems obey. Consider a glass of water as the combination of two half glasses of water, each half-glass having identical intensive properties (T , p , μ) but each half-glass having half the value of the extensive properties (E , V , N). Extensive properties scale in proportion to the system size. Taken to the limit of divisibility, scalable systems can be made up of elemental atoms or molecules or agents or elemental units of some kind.⁸¹

What extra power does this decomposability provide? First, entropy extensivity is the basis for some celebrated results of thermodynamics—the ability to define equilibria between subsystems by equalities of the intensive variables, the Maxwell relationships, and others. Second, of interest here, extensivity defines what constraints are appropriate for Max Ent predictions of large systems. Consider bringing two subsystems a and b together in a way that allows the exchange of an extensive property, such as the energy E , resulting in a combined system. In typical bulk equilibria, the subsystems are large enough to have precise average values of $\langle E \rangle$, which obey conservation in the exchange, $\langle E \rangle_{\text{total}} = \langle E \rangle_a + \langle E \rangle_b$.⁸² In such situations, fluctuations are negligible. Consequently, higher moments are not relevant because they scale sub-linearly with system size, so they become unimportant compared to $\langle E \rangle$. Similarly, vector properties—such as the momentum $\langle mv \rangle$ —are not relevant either because they vector-average to zero for large systems. By contrast, for sufficiently small systems, or where scalability is otherwise not applicable, higher-moment constraints often result in more accurate models compared to the traditional statistical mechanical ensembles.^{83–89}

Similarly, first-moment constraints are appropriate to model dynamics and Max Cal for systems where flows can be regarded as sums of component flows and large enough that average fluxes are well defined and where $\langle J \rangle_{\text{total}} = \langle J \rangle_a + \langle J \rangle_b$. This is the basis for the results above with Green-Kubo, Onsager reciprocal relations, and minimum entropy production.³⁴ Likewise, the first moment in number of transitions

between discrete states yields Markov processes and Master equation. Additional scalable constraints are sometimes applied, such as the condition of detailed balance, for kinetics that occurs at equilibrium.

However, note that while scalability allows us to discard second moment constraints, and simplify, it does not necessarily guarantee that first moment constraints are sufficient to describe system dynamics. Beyond the first moment constraints used, there may be additional scalable quantities that are relevant. For example, if grand canonical ensemble is described using only energy as a constraint, we will get erroneous result because of ignoring the other extensive quantity of particle number. Going a step further, even after knowing all the relevant extensive quantities and using their first moment as a constraint we may be at error, if there are other non-scalable observables at play. Not enough is yet known about what additional constraints are appropriate for far-from-equilibrium dissipative situations.^{90,91} As a result, our current lack of complete understanding in how to establish appropriate constraints in those cases should not be taken to imply the failure of the entropy variational principle itself, Max Ent or Max Cal, rather it indicates the need for more experience with more complex dissipative systems and the construction of constraints derived from likelihoods directly motivated from experiments.³⁰

VII. SUMMARY

We have discussed maximum caliber, a path-entropy-maximization principle for inferring dynamical distributions. It is quite general—applicable both near and far from equilibrium, and not limited to material systems in contact with baths. We show that it recovers, as a general principle should, well-known results of near-equilibrium dynamics—including the Green-Kubo fluctuation-dissipation relations, Onsager's reciprocal relations, and Prigogine's minimum entropy production. We describe examples of path-entropy variation results in inferring trajectory distributions from limited data, finding reaction coordinates in bio-molecular simulations, and modeling the complex dynamics of non-thermal systems such as gene regulatory networks or neuronal firing.

ACKNOWLEDGMENTS

K.D. appreciates support from the National Science Foundation (Grant No. 1205881) and from the Laufer Center. S.P. acknowledges the support for an ARO grant from the Mechanical Sciences Division (No. 66548-EG for Complex Dynamics and Systems) and K.G. acknowledges support from the National Science Foundation (Grant No. 1149992), Research Corporation for Science Advancement, and PROF grant from the University of Denver.

¹S. R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (Dover, NYC, 1984).

²H. J. Kreuzer, *Non-Equilibrium Thermodynamics and Its Statistical Foundations* (Oxford University Press, Oxford, 1981).

³B. H. Lavenda, *Nonequilibrium Statistical Thermodynamics* (John Wiley, USA, 1985).

⁴D. Kondepudi and I. Prigogine, *Modern Thermodynamics From Heat Engines to Dissipative Structures*, 2nd ed. (John Wiley, USA, 2015).

- ⁵I. Prigogine and D. Kondepudi, *West Sussex* (John Wiley & Sons Ltd, England, 1998).
- ⁶L. Onsager and S. Machlup, *Phys. Rev.* **91**, 1505 (1953).
- ⁷M. J. Klein and P. H. Meijer, *Phys. Rev.* **96**, 250 (1954).
- ⁸E. N. L. R. B. Bird and W. E. Stewart, *Transport Phenomena*, 2nd ed. (John Wiley, USA, 2001).
- ⁹D. Evans, E. Cohen, and G. Morriss, *Phys. Rev. Lett.* **71**, 2401 (1993).
- ¹⁰D. Evans and D. Searles, *Phys. Rev. E* **50**, 1645 (1994).
- ¹¹G. Gallavotti and E. Cohen, *J. Stat. Phys.* **80**, 931 (1995).
- ¹²G. Gallavotti and E. Cohen, *Phys. Rev. Lett.* **74**, 2694 (1995).
- ¹³J. L. Lebowitz and H. Spohn, *J. Stat. Phys.* **95**, 333 (1999).
- ¹⁴C. Maes, *J. Stat. Phys.* **95**, 367 (1998).
- ¹⁵U. Seifert, *Phys. Rev. Lett.* **95**, 040602 (2005).
- ¹⁶U. Seifert, *Rep. Prog. Phys.* **75**, 126001 (2012).
- ¹⁷G. E. Crooks, *J. Stat. Phys.* **90**, 1481 (1998).
- ¹⁸C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
- ¹⁹G. Hummer and A. Szabo, *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3658 (2001).
- ²⁰J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, *Nat. Phys.* **11**, 131 (2015).
- ²¹T. Sagawa and M. Ueda, *Phys. Rev. Lett.* **109**, 180602 (2012).
- ²²D. Mandal and C. Jarzynski, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11641 (2012).
- ²³C. Jarzynski, *Annu. Rev. Condens. Matter Phys.* **2**, 329 (2011).
- ²⁴C. Bustamante, J. Liphardt, and F. Ritort, *Phys. Today* **58**(7), 43 (2005).
- ²⁵E. T. Jaynes, *Annu. Rev. Phys. Chem.* **31**, 579 (1980).
- ²⁶A. A. Filyukov and V. Karpov, *Inzh.-Fiz. Zh.* **13**, 798 (1967).
- ²⁷H. Haken, *Z. Phys. B: Condens. Matter* **63**, 505 (1986).
- ²⁸E. Smith, *Rep. Prog. Phys.* **74**, 046601 (2011).
- ²⁹C. Monthus, *J. Stat. Mech.: Theory Exp.* **2011**, P03008.
- ³⁰S. Presse, K. Ghosh, J. Lee, and K. Dill, *Rev. Mod. Phys.* **85**, 1115 (2013).
- ³¹P. D. Dixit and K. A. Dill, *J. Chem. Theory Comput.* **10**, 3002 (2014).
- ³²P. D. Dixit, A. Jain, G. Stock, and K. A. Dill, *J. Chem. Theory Comput.* **11**, 5464 (2015).
- ³³P. D. Dixit, *Phys. Rev. E* **92**, 042149 (2015).
- ³⁴M. Hazoglou, V. Walther, P. Dixit, and K. Dill, *J. Chem. Phys.* **143**, 051104 (2015).
- ³⁵M. S. Green, *J. Chem. Phys.* **20**, 1281 (1952).
- ³⁶M. S. Green, *J. Chem. Phys.* **22**, 398 (1954).
- ³⁷R. Kubo, *J. Phys. Soc. Jpn.* **12**, 570 (1957).
- ³⁸L. Onsager, *Phys. Rev.* **37**, 405 (1931).
- ³⁹L. Onsager, *Phys. Rev.* **38**, 2265 (1931).
- ⁴⁰U. Seifert, *Eur. Phys. J. B* **64**, 423 (2008).
- ⁴¹T. Tomé and M. J. de Oliveira, *Phys. Rev. Lett.* **108**, 020601 (2012).
- ⁴²D. Kondepudi and I. Prigogine, *From Heat Engines to Dissipative Structures* (John Wiley & Son, 1998).
- ⁴³Q. Wang, *Astrophys. Space Sci.* **305**, 273 (2006).
- ⁴⁴K. Huang, *Statistical Mechanics* (John Wiley, USA, 1987).
- ⁴⁵K. Ghosh, K. Dill, M. M. Inamdar, E. Seitaridou, and R. Phillips, *Am. J. Phys.* **74**, 123 (2006).
- ⁴⁶E. Seitaridou, M. Inamdar, R. Phillips, K. Ghosh, and K. Dill, *J. Phys. Chem. B* **111**(9), 2288–2292 (2007).
- ⁴⁷H. Ge, S. Presse, K. Ghosh, and K. Dill, *J. Chem. Phys.* **134**, 064108 (2012).
- ⁴⁸J. Lee and S. Presse, *J. Chem. Phys.* **137**, 074103 (2012).
- ⁴⁹A. Filyukov and V. Karpov, *Inzh.-Fiz. Zh.* **13**, 624 (1967).
- ⁵⁰G. Stock, K. Ghosh, and K. Dill, *J. Chem. Phys.* **128**, 194102 (2008).
- ⁵¹K. Ghosh, *J. Chem. Phys.* **134**(19), 195101 (2011).
- ⁵²M. Otten and G. Stock, *J. Chem. Phys.* **133**, 034119 (2010).
- ⁵³D. Bicout and A. Szabo, *J. Chem. Phys.* **109**, 2325 (1998).
- ⁵⁴H. Wan, G. Zhou, and V. A. Voelz, *J. Chem. Theory Comput.* **12**, 5768 (2016).
- ⁵⁵G. Zhou, G. A. Pantelopulos, S. Mukherjee, and V. A. Voelz, *Biophys. J.* **113**, 785 (2017).
- ⁵⁶P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- ⁵⁷P. Tiwary and B. Berne, *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2839 (2016).
- ⁵⁸P. Tiwary and B. Berne, *J. Chem. Phys.* **145**, 054113 (2016).
- ⁵⁹P. Tiwary and B. Berne, *J. Chem. Phys.* **147**, 152701 (2017).
- ⁶⁰P. Tiwary, *J. Phys. Chem. B* **121**(48), 10841 (2017).
- ⁶¹J. Paulsson, *Nature* **427**, 415 (2004).
- ⁶²T. Firman, G. Balázsi, and K. Ghosh, *Biophys. J.* **113**(9), 2121–2130 (2017).
- ⁶³S. Presse, K. Ghosh, and K. Dill, *J. Phys. Chem. B* **115**, 6202 (2011).
- ⁶⁴G. H. Wadhams and J. P. Armitage, *Nat. Rev. Mol. Cell Biol.* **5**, 1024 (2004).
- ⁶⁵R. S. Herbst, *Int. J. Radiat. Oncol. Biol., Phys.* **59**, S21 (2004).
- ⁶⁶P. D. Dixit, *Biophys. J.* **104**, 2743 (2013).
- ⁶⁷P. Dixit, E. Lyashenko, M. Niepel, and D. Vitkup, *bioRxiv:137513* (2017).
- ⁶⁸G. Peterson, S. Presse, and K. Dill, *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16023 (2010).
- ⁶⁹J. Peterson, P. D. Dixit, and K. A. Dill, *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20380 (2013).
- ⁷⁰L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, 1968).
- ⁷¹D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, 1987).
- ⁷²C. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
- ⁷³E. T. Jaynes, *Phys. Rev.* **106**, 620 (1957).
- ⁷⁴J. Shore and R. Johnson, *IEEE Trans. Inf. Theory* **26**, 26 (1980).
- ⁷⁵L. Sawle and K. Ghosh, *J. Chem. Theory Comput.* **12**, 861 (2016).
- ⁷⁶S. J. Phillips, R. P. Anderson, and R. E. Schapire, *Ecol. Modell.* **190**, 231 (2006).
- ⁷⁷E. Schneidman, M. Berry, R. Segev, and W. Bialek, *Nature* **440**, 1007 (2006).
- ⁷⁸R. C. Dewar, *Entropy* **11**, 931 (2009).
- ⁷⁹A. Caticha and R. Preuss, *Phys. Rev. E* **70**, 046127 (2004).
- ⁸⁰T. Mora, A. M. Walczak, W. Bialek, and C. G. Callan, *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5405 (2010).
- ⁸¹K. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology* (Garland Science, New York, 2003).
- ⁸²We limit consideration here to only those where the range of interactions is smaller than the system size.
- ⁸³T. Hill and R. Chamberlin, *Nano Lett.* **2**, 609 (2002).
- ⁸⁴R. Chamberlin and G. Wolf, *Eur. Phys. J. B* **67**, 495 (2009).
- ⁸⁵R. Chamberlin, J. Vermaas, and G. Wolf, *Eur. Phys. J. B* **71**, 1 (2009).
- ⁸⁶P. D. Dixit, *J. Chem. Phys.* **138**, 184111 (2013).
- ⁸⁷P. D. Dixit, *Phys. Chem. Chem. Phys.* **17**, 13000 (2015).
- ⁸⁸P. D. Dixit, A. Bansal, W. G. Chapman, and D. Asthagiri, *J. Chem. Phys.* **147**, 164901 (2017); preprint [arXiv:1706.08998](https://arxiv.org/abs/1706.08998) (2017).
- ⁸⁹T. Dauxois, S. Ruffo, E. Arimondo, and M. Wilkens, *Lecture Notes in Physics* (Springer, New York, Berlin, 2002), p. 1.
- ⁹⁰R. L. Jack and R. Evans, *J. Stat. Mech.: Theory Exp.* **2016**, 093305.
- ⁹¹C. Maes, preprint [arXiv:1603.05147](https://arxiv.org/abs/1603.05147) (2016).