Information thermodynamics of transition paths between multiple mesostates

Miranda D. Louwerse [*] and David A. Sivak [†]

Department of Physics, Simon Fraser University, Burnaby, British Columbia, Canada V5A1S6

A central concern across the natural sciences is a quantitative understanding of the mechanism governing rare transitions between two metastable states. Recent research has uncovered a fundamental equality between the time-reversal asymmetry of the ensemble of such transition paths and the informativeness of system dynamics about the reactivity of a given trajectory, immediately leading to quantitative criteria for judging the importance of distinct system coordinates for the transition. Here we generalize this framework to multiple mesostates. We find that the main system-wide and coordinate-specific results generalize intuitively, while the combinatorial diversity of pairwise transitions raises new questions and points to new concepts. This work increases the previous framework's generality and applicability and forges connections to enhanced-sampling and coarse-grained dynamical approaches such as milestoning and Markov-state models.

I. INTRODUCTION

Biomolecular dynamics are studied and understood on several length scales, from atomic details to more coarse-grained aspects [1]. The mechanisms by which reactions occur are typically couched in terms of a global picture of the dynamics, including distinguishable metastable states, their relative stability, and transitions between them. At the same time, microscopic details that affect the stability of mesostates or the transitions between them are of interest, especially in applications where tuning molecular components affects the function of the molecule [2]. It is therefore desirable to build models of dynamics that capture both global and local features of the system.

Much of the theory of reaction dynamics is developed in the context of a reaction between two endpoints, which from a coarse-grained perspective is essentially two state, parameterized by forward and reverse rate constants, which can be derived from microscopic dynamics using transition-path theory [3,4]. In previous work [5], we used tools from stochastic thermodynamics to adapt the transition-path theory description to propose an information-theoretic approach, in which an irreversible entropy-production rate quantifies the informativeness of motion in relevance of microscopic degrees of freedom to describing the mesoscopic reaction dynamics, i.e., to determining the rate constant. Through this, we showed that in this information-theoretic sense the committor function, a central quantity in transition-path theory, is a sufficient reaction coordinate (capturing all available system information about trajectory reactivity) and therefore the true reaction coordinate [2].

Here, we extend these ideas to reactions between multiple mesostates. These types of reaction networks are commonly seen in biomolecular systems and are frequently modeled using coarse-grained Markov-state models [1,6–9]. We define an information-theoretic description of a multimesostate reaction network and relate the informativeness of system dynamics to the entropy-production rate for reactive dynamics, in close analogy to the relation derived for two endpoints in our previous work [5]. However, the multimesostate model has some nuances: e.g., the committor (defining the probability of next reaching a particular mesostate given an initial microstate) becomes a vector [8]; we show that this committor is a sufficient statistic for relating the reaction network to the microscopic dynamics. We also derive an entropy-production rate for transition-path ensembles between each pair of mesostates and derive a local reaction coordinate for each pairwise reaction that fully captures this entropy-production rate. We discuss how these information-theoretic quantities could be used to build informative global models of the reaction network while at the same time highlighting system features important to each local reaction. Finally, we show that the equating of entropy production and information generation also holds for subensembles of trajectories that leave a given mesostate [10] (without specifying the destination mesostate). With this multimesostate framework, the main results from our previous work generalize simply and intuitively, increasing their applicability.

II. CATEGORICAL OUTCOME AND ORIGIN VARIABLES AND TRAJECTORY SUBENSEMBLES

We first generalize our theoretical framework to accommodate more than two mesostates of interest.

A. System dynamics

A system with discrete microstates $\boldsymbol{\phi}$ in configuration space $\boldsymbol{\Phi}$ evolves under master-equation dynamics

$$\mathrm{d}_t p(\boldsymbol{\phi}) = \sum_{\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} p(\boldsymbol{\phi}') \tag{1}$$

———————
[*]Present address: AbCellera Biologics Inc., Vancouver, British Columbia, Canada.
[†]dsivak@sfu.ca

with probability $p(\boldsymbol{\phi})$ of microstate $\boldsymbol{\phi}$ and $\boldsymbol{\phi}' \to \boldsymbol{\phi}$ transition rate $T_{\boldsymbol{\phi}\boldsymbol{\phi}'}$. The system has reached the equilibrium Boltzmann distribution

$$\pi(\boldsymbol{\phi}) = e^{\beta F - \beta E(\boldsymbol{\phi})} \tag{2}$$

for energy $E(\boldsymbol{\phi})$ of microstate $\boldsymbol{\phi}$ and equilibrium free energy $F \equiv -\ln \sum_{\boldsymbol{\phi}} e^{-\beta E(\boldsymbol{\phi})}$. At equilibrium the probability of microstate $\boldsymbol{\phi}$ does not change,

$$0 = \mathrm{d}_t \pi(\boldsymbol{\phi}) = \sum_{\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}'), \tag{3}$$

and detailed balance

$$T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') = T_{\boldsymbol{\phi}'\boldsymbol{\phi}} \pi(\boldsymbol{\phi}) \tag{4}$$

holds for each individual $\boldsymbol{\phi}' \to \boldsymbol{\phi}$ transition.

Define a set of N_{meso} mesostates $\mathcal{M} = \{m_i\}$, each of which is a distinct subset of microstates (e.g., in the two-mesostate transition-path theory [4], $N_{\mathrm{meso}} = 2$, $m_1 = A$, and $m_2 = B$). (To simplify the analysis, we assume there are no direct transitions between distinct mesostates.) These mesostates could be chosen as the basins of metastable states [6], at transient milestones during a transition [11], or in other ways. For each time in a long ergodic equilibrium supertrajectory comprising a sequence of many transitions between each of the different mesostates, we define the trajectory outcome S_+ and origin S_- as, respectively, the next and the most recent mesostate visited by the system, which are categorical random variables with values drawn from the list of mesostates: $s_+, s_- \in \{m_i\}$. The joint distribution of microstate $\boldsymbol{\phi}$, trajectory origin s_-, and trajectory outcome s_+ can be decomposed as

$$p(\boldsymbol{\phi}, s_-, s_+) = \pi(\boldsymbol{\phi}) p(s_- | \boldsymbol{\phi}) p(s_+ | \boldsymbol{\phi}). \tag{5}$$

The right-hand-side decomposition relies on the conditional independence of trajectory outcome and origin given a microstate $\boldsymbol{\phi}$, which follows from the Markovianity of the underlying dynamics [3,4].

By marginalizing over all microstates $\boldsymbol{\phi}$ the joint distribution, the supertrajectory is partitioned into N_{meso}^2 trajectory subensembles $\boldsymbol{s} = (s_-, s_+)$ with marginal probability distribution

$$p(s_-, s_+) = \sum_{\boldsymbol{\phi}} \pi(\boldsymbol{\phi}) p(s_- | \boldsymbol{\phi}) p(s_+ | \boldsymbol{\phi}). \tag{6}$$

There are N_{meso} stationary subensembles (where the system leaves and returns to the same mesostate) and $N_{\mathrm{meso}}^2 - N_{\mathrm{meso}}$ reactive subensembles (where the system leaves one mesostate and next reaches another). The forward and reverse transition-path ensembles between any pair of mesostates are time-reversed counterparts with equal-magnitude flux of reactive trajectories at any given point in configuration space; there are therefore $\binom{N_{\mathrm{meso}}}{2} = N_{\mathrm{meso}}(N_{\mathrm{meso}} - 1)/2$ distinct reactive subensemble pairs.

B. Forward and reverse multi-mesostate committor vectors

The conditional probability of a trajectory outcome (origin) given a microstate $\boldsymbol{\phi}$, $p(S_+ = m_i | \boldsymbol{\phi})$ [$p(S_- = m_i | \boldsymbol{\phi})$]

[appearing in Eq. (5)], is a multimesostate generalization of the forward (backward) committor function that appears in the two-mesostate transition-path theory: it gives the probability that a system initiated at $\boldsymbol{\phi}$ will next reach (last visited) mesostate m_i [8]. The forward (backward) committor $p(s_+ | \boldsymbol{\phi})$ [$p(s_- | \boldsymbol{\phi})$] for multiple mesostates is a vector function of dimension N_{meso}, with each element giving the probability the system will next visit (last visited) mesostate m_i [8], with normalization

$$\sum_{i=1}^{N_{\mathrm{meso}}} p(S_+ = m_i | \boldsymbol{\phi}) = 1 \tag{7a}$$

$$\sum_{i=1}^{N_{\mathrm{meso}}} p(S_- = m_i | \boldsymbol{\phi}) = 1, \tag{7b}$$

for each microstate $\boldsymbol{\phi}$. The traditional committor for a two-mesostate model can also be considered a vector function with two elements, but the constraints in Eq. (7) imply that both elements are determined by a scalar (because they sum to 1), thus the two-mesostate committor is conventionally described as a scalar function.

For each microstate $\boldsymbol{\phi}$, each element of the forward committor vector $p(S_+ = m_i | \boldsymbol{\phi})$ satisfies [3,8]

$$0 = \sum_{\boldsymbol{\phi}'} p(S_+ = m_i | \boldsymbol{\phi}') T_{\boldsymbol{\phi}'\boldsymbol{\phi}}, \tag{8}$$

and the reverse committor vector $p(S_- = m_i | \boldsymbol{\phi})$ satisfies

$$0 = \sum_{\boldsymbol{\phi}'} p(S_- = m_i | \boldsymbol{\phi}') T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \frac{\pi(\boldsymbol{\phi}')}{\pi(\boldsymbol{\phi})} \tag{9a}$$

$$= \sum_{\boldsymbol{\phi}'} p(S_- = m_i | \boldsymbol{\phi}') T_{\boldsymbol{\phi}'\boldsymbol{\phi}} \tag{9b}$$

with boundary conditions $p(S_+ = m_i | \boldsymbol{\phi}) = p(S_- = m_i | \boldsymbol{\phi}) = 1$ for $\boldsymbol{\phi} \in m_i$ and $p(S_+ = m_i | \boldsymbol{\phi}) = p(S_- = m_i | \boldsymbol{\phi}) = 0$ for $\boldsymbol{\phi} \in \mathcal{M} \setminus m_i$. Equation (9b) uses the detailed-balance relation (4). For our time-reversible dynamics, the system of equations for $p(S_- = m_i | \boldsymbol{\phi})$ and $p(S_+ = m_i | \boldsymbol{\phi})$ are identical, which implies $p(S_- = m_i | \boldsymbol{\phi}) = p(S_+ = m_i | \boldsymbol{\phi})$.

III. INFORMATION THEORY

The coarse-grained mesostate model just described retains some information about the system's microstate. This is quantified by the mutual information between system microstate and trajectory type,

$$I[\boldsymbol{\Phi}; \boldsymbol{S}] = \sum_{\boldsymbol{\phi}, \boldsymbol{s}} p(\boldsymbol{\phi}, \boldsymbol{s}) \ln \frac{p(\boldsymbol{\phi}, \boldsymbol{s})}{\pi(\boldsymbol{\phi}) p(\boldsymbol{s})} \tag{10a}$$

$$= H(\boldsymbol{S}) - H(\boldsymbol{S} | \boldsymbol{\Phi}) \tag{10b}$$

$$= H(\boldsymbol{\Phi}) - H(\boldsymbol{\Phi} | \boldsymbol{S}). \tag{10c}$$

This is the reduction of uncertainty about the current trajectory type upon observation of the system microstate, or conversely the reduction of uncertainty about the microstate from knowledge of the trajectory type. For example, simulation methods for importance sampling of transition paths

between two mesostates generate a set of trajectories [12] or sample the flux of trajectories [11,13,14] of a given trajectory type. This generates microstate probability distributions conditioned on each trajectory type (e.g., $A \to A$, $A \to B$, $B \to B$, $B \to A$) whose average entropy over all subensembles is less than the system entropy in the full equilibrium ensemble; the reduction in entropy is the mutual information between the system microstate and trajectory type [Eq. (10c)]. By resolving one of these variables, either the current trajectory type or the system microstate, information is gained about the other.

A major focus of chemical-physics research is the identification of a small set of coordinates that—despite vastly reduced dimensionality—can both distinguish between mesostates of interest and describe the progress of important reactions in the system [2]. For two-mesostate systems, it is well recognized that this amounts to finding a small set of collective variables (functions of microstate) that can parametrize the committor function. Posed information theoretically, the mutual information between the reduced space of collective variables and trajectory type measures the sufficiency of the chosen collective variables to quantify the statistical relationship between the microstate and trajectory type (which determines the reactivity or stationarity of the current trajectory segment). From the data-processing inequality [15], the mutual information between collective variables and trajectory type is less than or equal to the mutual information between microstate and trajectory type, and therefore information about the reactivity of the system is lost upon insufficient coarse graining. An optimal set of collective variables preserves the mutual information upon coarse graining. The forward committor vector has N_{meso} elements, with a constraint that all elements must sum to unity; due to time-reversal symmetry, no further collective variables are needed to determine the reverse committor vector. An N_{meso}-mesostate model therefore requires at minimum $N_{\text{meso}} - 1$ degrees of freedom to sufficiently quantify the information between underlying microstate and trajectory type.

Let $q = \{q_i\}$ be an N_{meso}-dimensional vector with each entry the forward (or equivalently reverse) committor for mesostate i, constrained to values such that $\sum_i q_i = 1$. We show that the committor is a sufficient statistic for the mutual information between the microstate and trajectory type, as follows:

$$I[\mathbf{\Phi}; S] = \sum_{\phi, s} \pi(\phi) p(s|\phi) \ln \frac{p(s|\phi)}{p(s)} \tag{11a}$$

$$= \int d\boldsymbol{q} \sum_{\phi, s} \pi(\phi) p(s|\phi) \delta[p(s_+|\phi) - \boldsymbol{q}] \ln \frac{p(s|\phi)}{p(s)} \tag{11b}$$

$$= \int d\boldsymbol{q} \sum_{i,j} q_i q_j \ln \frac{q_i q_j}{p(S_- = m_i, S_+ = m_j)}$$
$$\times \sum_{\phi} \pi(\phi) \delta[p(s_+|\phi) - \boldsymbol{q}], \tag{11c}$$

where we used $p(S_- = m_i, S_+ = m_j|\phi) = q_i q_j$ and grouped together all microstates with the same committor-vector

value. For marginal committor probability $\pi(\boldsymbol{q}) = \sum_{\phi} \pi(\phi) \delta[p(s_+|\phi) - \boldsymbol{q}]$, this further simplifies to

$$I[\mathbf{\Phi}; S] = \int d\boldsymbol{q} \sum_{i,j} \pi(\boldsymbol{q}) q_i q_j \ln \frac{\pi(\boldsymbol{q}) q_i q_j}{\pi(\boldsymbol{q}) p(S_- = m_i, S_+ = m_j)} \tag{12a}$$

$$= I[\boldsymbol{Q}; S], \tag{12b}$$

where we expressed the mutual information between microstate and trajectory type in terms of the sufficient statistic $\boldsymbol{Q}$, the random variable given by the committor of the microstate.

When the mesostates fully partition the configuration space, the committor vector of each microstate has unity for the mesostate to which the microstate belongs and zeros for all other mesostates [$p(s = (m_i, m_i)|\phi) = 1$ for $\phi \in m_i$ and 0 otherwise], so the mutual information between system microstate and trajectory type reduces to the trajectory-type entropy:

$$I[\mathbf{\Phi}; S] = \sum_{\phi} \sum_s \pi(\phi) p(s|\phi) \ln \frac{p(s|\phi)}{p(s)} \tag{13a}$$

$$= -\sum_{i=1}^{N_{\text{meso}}} \left[\sum_{\phi \in m_i} \pi(\phi) \right] \ln \left[\sum_{\phi \in m_i} \pi(\phi) \right] \tag{13b}$$

$$= -\sum_{i=1}^{N_{\text{meso}}} \pi(m_i) \ln \pi(m_i) \tag{13c}$$

$$= H(\boldsymbol{S}), \tag{13d}$$

for Boltzmann probability $\pi(m_i)$ of the mesostate. In this case, observation of system microstate fully determines the current trajectory type, thus $H(S|\mathbf{\Phi}) = 0$. If the number N_{meso} of mesostates approaches the number of system microstates, knowing the microstate fully determines the trajectory type and vice versa, so the mutual information between microstate and trajectory type approaches the equilibrium system entropy:

$$\lim_{N_{\text{meso}} \to N_{\text{micro}}} I[\mathbf{\Phi}; S] = H(\mathbf{\Phi}). \tag{14}$$

A. Construction of one-dimensional reaction coordinates

In two-mesostate transition-path theory, the (scalar) committor plays an important role as the one-dimensional reaction coordinate that is fully informative about the reaction. The committor can be used to coarse grain the configuration space into isocommittor surfaces [4], and determining properties of the system (e.g., mean values of particular system coordinates that are relevant to the reaction) on the isocommittor surfaces yields insight into the reaction mechanism. A particularly interesting surface is the $q = 0.5$ isocommittor, the transition-state ensemble of the reaction.

Consider how this picture changes in multimesostate transition-path theory [8]. The $i \to j$ and $j \to i$ transition-path ensembles are characterized by two committor components $q_i(\phi) = p(S_- = m_i|\phi) = p(S_+ = m_i|\phi)$ and $q_j(\phi) = p(S_- = m_j) = p(S_+ = m_j|\phi)$. In two-mesostate transition-path theory (where without loss of generality $i = A$ and $j = B$), these components are related by $q_A(\phi) + q_B(\phi) = 1$ and there is an obvious choice of scalar reaction coordinate;

such a relationship does not exist with multiple mesostates where the i and j committor components do not sum to unity $[q_i(\boldsymbol{\phi}) + q_j(\boldsymbol{\phi}) \leqslant 1]$ and their sum may change throughout configuration space $[q_i(\boldsymbol{\phi}) + q_j(\boldsymbol{\phi}) \neq q_i(\boldsymbol{\phi}') + q_j(\boldsymbol{\phi}')$ for all $\boldsymbol{\phi}$ and $\boldsymbol{\phi}']$. Nevertheless, a local reaction coordinate $r_{ij}(\boldsymbol{\phi})$ can be constructed as a function of both committor components

$$r_{i \to j}(\boldsymbol{\phi}) = \frac{q_j(\boldsymbol{\phi})}{q_i(\boldsymbol{\phi}) + q_j(\boldsymbol{\phi})}, \tag{15}$$

which we term the pairwise committor [16] for the $i \to j$ reactive subensemble. In line with the two-mesostate committor function, the pairwise committor has values in [0,1], where microstates with pairwise committor closer to 0 are dynamically closer to m_i (reactantlike) and microstates with pairwise committor closer to 1 are dynamically closer to m_j (productlike). At equilibrium, the reaction coordinate for the reverse transition (with reactant m_j and product m_i) is $r_{j \to i} = 1 - r_{i \to j}$. Analogous to the two-mesostate transition-state ensemble, the $i \to j$ transition-state ensemble is identified as microstates with $r_{i \to j}(\boldsymbol{\phi}) = 0.5$, where the system has equal probability to next reach m_i or m_j. Other isosurfaces of the pairwise committor span microstates that have the same odds of next reaching m_j before m_i. One can construct $\binom{N_{\text{meso}}}{2}$ such local reaction coordinates to describe each of the reactive subensemble pairs. These reaction coordinates are not independent, since only $N_{\text{meso}} - 1$ committor functions are needed to construct them.

IV. SYSTEM-SUBENSEMBLE DYNAMICS

A. Microstate dynamics in trajectory-outcome and trajectory-origin subensembles

We consider the dynamics in the joint space of microstates and trajectory types, which yields the stationary joint distribution (5). The joint transition rates for the categorical outcome and origin and the system microstate are [5,17]

$$T_{\boldsymbol{\phi}\boldsymbol{\phi}'}^{ss'} = \begin{cases} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}^s \equiv T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \frac{p(s_+|\boldsymbol{\phi})}{p(s_+|\boldsymbol{\phi}')}, & s' = s \\[2mm] T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \, p(S_+ = m_j|\boldsymbol{\phi}), & \begin{cases} \boldsymbol{\phi}' \in m_i, \, \boldsymbol{\phi} \notin m_i, \\ s' = (m_i, m_i), \\ s = (m_i, m_j) \end{cases} \\[4mm] T_{\boldsymbol{\phi}\boldsymbol{\phi}'} / p(S_+ = m_i|\boldsymbol{\phi}), & \begin{cases} \boldsymbol{\phi}' \notin m_i, \, \boldsymbol{\phi} \in m_i, \\ s' = (m_j, m_i), \\ s = (m_i, m_i) \end{cases} \\[4mm] -\sum\limits_{\substack{\boldsymbol{\phi}'' \neq \boldsymbol{\phi}' \\ s'' \neq s'}} T_{\boldsymbol{\phi}''\boldsymbol{\phi}'}^{s''s'}, & \boldsymbol{\phi} = \boldsymbol{\phi}', \, s = s' \\[4mm] 0 & \text{otherwise} \end{cases} \tag{16}$$

The top transition rate is for transitions that do not change the trajectory type s, the second is for transitions that leave a mesostate m_i to start a reactive trajectory to m_j, the third is for transitions that finish a reactive trajectory from m_j to m_i, and the fourth ensures that at steady state the probability is conserved for each microstate. Note that while the marginal dynamics on the microstate space alone (described by $T_{\boldsymbol{\phi}\boldsymbol{\phi}'}$) are at equilibrium, the joint dynamics of microstate and trajectory type ($T_{\boldsymbol{\phi}\boldsymbol{\phi}'}^{ss'}$) are at nonequilibrium steady state, due to the

fundamental time-reversal asymmetry of the trajectory-type variable.

B. Information flows

As the system evolves at equilibrium, the respective changes in each of the random variables contribute to changes in the mutual information (10a) between system microstate and trajectory type, such that they on average cancel to result in no change to the mutual information. We consider separately the changes in mutual information during changes of trajectory type, $i^S[\boldsymbol{\Phi}; S]$, and during microstate transitions within a fixed trajectory type, $i^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S]$. These two contributions are equal in magnitude and have opposite sign due to stationarity of total mutual information at equilibrium,

$$0 = \mathrm{d}_t I[\boldsymbol{\Phi}; S] = i^S[\boldsymbol{\Phi}; S] + i^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S]. \tag{17}$$

We now show now that the two contributions in fact vanish separately, due to time-reversal symmetry.

The rate of change in mutual information during microstate transitions that do not change the trajectory type is

$$i^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S] = \sum_{s_-} \sum_{s_+} \sum_{\boldsymbol{\phi},\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}^s p(\boldsymbol{\phi}', s) \ln \frac{p(s_+|\boldsymbol{\phi})p(s_-|\boldsymbol{\phi})}{p(s_+|\boldsymbol{\phi}')p(s_-|\boldsymbol{\phi}')} \tag{18a}$$

$$= \sum_{s_-} \sum_{\boldsymbol{\phi},\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(s_-|\boldsymbol{\phi}') \ln \frac{p(s_-|\boldsymbol{\phi})}{p(s_-|\boldsymbol{\phi}')}$$

$$+ \sum_{s_+} \sum_{\boldsymbol{\phi},\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(s_+|\boldsymbol{\phi}) \ln \frac{p(s_+|\boldsymbol{\phi})}{p(s_+|\boldsymbol{\phi}')} \tag{18b}$$

$$= i^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S_-] + i^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S_+] \tag{18c}$$

$$= 0, \tag{18d}$$

which we split into information generated about the outcome, $i^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S_+] \geqslant 0$, and about the origin, $i^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S_-] \leqslant 0$, that are equal in magnitude and have opposite sign due to time-reversal symmetry of trajectory outcome and origin at equilibrium. The rate of change in mutual information during changes of trajectory type is

$$i^S[\boldsymbol{\Phi}; S] = \sum_i \sum_{j \neq i} \sum_{\boldsymbol{\phi} \notin m_i} \sum_{\boldsymbol{\phi}' \in m_i} \left[T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(S_+ = m_j|\boldsymbol{\phi}) \right.$$

$$\times \ln \frac{p(S_+ = m_j|\boldsymbol{\phi})p(S_- = m_i|\boldsymbol{\phi})p(m_i, m_i)}{p(m_i, m_j)}$$

$$+ T_{\boldsymbol{\phi}'\boldsymbol{\phi}} \pi(\boldsymbol{\phi}) p(S_- = m_j|\boldsymbol{\phi}) \tag{19a}$$

$$\times \ln \frac{p(m_j, m_i)}{p(S_+ = m_i|\boldsymbol{\phi})p(S_- = m_j|\boldsymbol{\phi})p(m_i, m_i)} \right]$$

$$= 0, \tag{19b}$$

where the first term is the contribution from transitions where the outcome changes from m_i to m_j (system starts a reactive trajectory from m_i to m_j) and the second term is the contribution from transitions where the origin changes from m_j to

m_i (system finishes a reactive trajectory from m_j to m_i). These terms are equal in magnitude and have opposite sign since the two sets of transitions leading to changes of trajectory type are time-reversed counterparts.

C. Entropy changes

The microstate transitions that do not change trajectory type s are in general time-reversal asymmetric, which can be quantified by a local detailed-balance relation,

$$\frac{T^s_{\phi\phi'}p(\phi',s)}{T^s_{\phi'\phi}p(\phi,s)} = \frac{p(S_+ = m_j|\phi)p(S_- = m_i|\phi')}{p(S_+ = m_j|\phi')p(S_- = m_i|\phi)}. \tag{20}$$

For stationary subensembles, $p(S_+ = m_i|\phi) = p(S_- = m_i|\phi)$ and the ratio equals unity, so there is no time-reversal asymmetry. For reactive subensemble pairs, the ratio can be expressed as a change in the reaction coordinate for that

subensemble,

$$\ln \frac{p(S_+ = m_j|\phi)p(S_- = m_i|\phi')}{p(S_+ = m_j|\phi')p(S_- = m_i|\phi)}$$

$$= \ln \frac{r_{i\to j}(\phi)}{r_{j\to i}(\phi)} - \ln \frac{r_{i\to j}(\phi')}{r_{j\to i}(\phi')} \tag{21a}$$

$$= \ln \frac{r_{i\to j}(\phi)}{1 - r_{i\to j}(\phi)} - \ln \frac{r_{i\to j}(\phi')}{1 - r_{i\to j}(\phi')}. \tag{21b}$$

Thus we can consider time-asymmetric reactive dynamics as arising from a nonconservative potential $\ln r_{i\to j}(\phi')/[1 - r_{i\to j}(\phi')]$ on the system, that is a function of the reaction coordinate.

To understand the effect of this nonconservative potential on the system thermodynamics, consider the joint entropy $H(\mathbf{\Phi}, \mathbf{S}) \equiv -\sum_{\phi,s} p(\phi,s) \ln p(\phi,s)$. The change in joint entropy can be decomposed into three terms [18]:

$$0 = d_t H(\mathbf{\Phi}, \mathbf{S}) \tag{22a}$$

$$= \underbrace{\sum_{\phi,\phi',s} T^s_{\phi\phi'}p(\phi',s) \ln \frac{T^s_{\phi\phi'}p(\phi',s)}{T^s_{\phi'\phi}p(\phi,s)}}_{\dot{H}^{\mathrm{irr}}(\mathbf{\Phi},\mathbf{S})} - \underbrace{\sum_{\phi,\phi',s} T^s_{\phi\phi'}p(\phi',s) \ln \frac{T^s_{\phi\phi'}}{T^s_{\phi'\phi}}}_{\dot{H}^{\mathrm{env}}(\mathbf{\Phi},\mathbf{S})} + \underbrace{\sum_{\phi,\phi',s\neq s'} T^{ss'}_{\phi\phi'}p(\phi',s') \ln \frac{p(\phi',s')}{p(\phi,s)}}_{\dot{H}^{\mathrm{sub}}(\mathbf{\Phi},\mathbf{S})}, \tag{22b}$$

where $\dot{H}^{\mathrm{irr}}(\mathbf{\Phi}, \mathbf{S})$ is the irreversible entropy production, $\dot{H}^{\mathrm{env}}(\mathbf{\Phi}, \mathbf{S})$ is the environmental entropy change for transitions that do not change the trajectory type, and $\dot{H}^{\mathrm{sub}}(\mathbf{\Phi}, \mathbf{S})$ is the change in joint entropy due to transitions that change the trajectory type.

Substituting the transition rates that change trajectory type (16), the change in joint entropy due to trajectory-type changes is

$$\dot{H}^{\mathrm{sub}}(\mathbf{\Phi}, \mathbf{S}) = \sum_{\phi,\phi',s\neq s'} T^{ss'}_{\phi\phi'}p(\phi',s') \ln \frac{p(\phi',s')}{p(\phi,s)} \tag{23a}$$

$$= \sum_i \sum_{j\neq i} \sum_{\phi\notin m_i} \sum_{\phi'\in m_i} \tag{23b}$$

$$\times \left[T_{\phi\phi'}\pi(\phi')p(S_+ = m_j|\phi) \ln \frac{p[\phi',(m_i,m_i)]}{p[\phi,(m_i,m_j)]} + T_{\phi'\phi}\pi(\phi)p(S_- = m_j|\phi) \ln \frac{p[\phi,(m_j,m_i)]}{p[\phi',(m_i,m_i)]} \right]$$

$$= 0. \tag{23c}$$

In Eq. (23b), we write the change in joint entropy for all transitions that leave mesostate m_i and begin a transition path to m_j and for all transitions that enter mesostate m_i after completing a transition path from m_j. Due to time-reversal symmetry, $p(S_+ = m_j|\phi) = p(S_- = m_j|\phi)$ and $p[\phi,(m_i,m_j)] = p[\phi,(m_j,m_i)]$ and thus the two entropy changes within square brackets cancel.

Substituting the joint transition rates (16) into the expression for the environmental entropy change gives

$$\dot{H}^{\mathrm{env}}(\mathbf{\Phi}, \mathbf{S}) = \sum_{\phi,\phi',s} T^s_{\phi\phi'}p(\phi',s) \left[\ln \frac{T_{\phi\phi'}}{T_{\phi'\phi}} + 2\ln \frac{p(s_+|\phi)}{p(s_+|\phi')} \right] \tag{24a}$$

$$= -\sum_s p(s)\dot{Q}^s + 2\dot{I}^{\mathbf{\Phi}}[\mathbf{\Phi}; S_+] \tag{24b}$$

$$= 2\dot{I}^{\mathbf{\Phi}}[\mathbf{\Phi}; S_+], \tag{24c}$$

where $\dot{I}^{\mathbf{\Phi}}[\mathbf{\Phi}; S_+] \geqslant 0$ is the rate of change in mutual information between the trajectory outcome and system microstate due to system dynamics in a fixed trajectory type, and $\dot{Q}^s$ is the mean rate of change of system energy [see Eq. (2)] during a fixed trajectory type s. Equation (24c) follows from summing over s to cancel average energy changes during forward and reverse transition-path ensembles that are equal in magnitude and have opposite sign.

The irreversible entropy production [19] of all subensemble dynamics is

$$\dot{H}^{\mathrm{irr}}(\mathbf{\Phi}, S_-, S_+) = \sum_s p(s)\dot{\Sigma}_s \tag{25a}$$

$$= \sum_{s_-, s_+} \sum_{\phi, \phi'} T^s_{\phi\phi'} p(\phi', s) \left[\ln \frac{T_{\phi\phi'}\pi(\phi')}{T_{\phi'\phi}\pi(\phi)} + \ln \frac{p(s_-|\phi')}{p(s_-|\phi)} + \ln \frac{p(s_+|\phi)}{p(s_+|\phi')} \right] \tag{25b}$$

$$= \sum_{s_+} \sum_{\phi\phi'} T_{\phi\phi'}\pi(\phi')p(s_+|\phi) \ln \frac{p(s_+|\phi)}{p(s_+|\phi')} \tag{25c}$$

$$+ \sum_{s_-} \sum_{\phi\phi'} T_{\phi\phi'}\pi(\phi')p(s_-|\phi') \ln \frac{p(s_-|\phi')}{p(s_-|\phi)} \tag{25d}$$

$$= \dot{I}^{\mathbf{\Phi}}[\mathbf{\Phi}; S_+] - \dot{I}^{\mathbf{\Phi}}[\mathbf{\Phi}; S_-], \tag{25e}$$

which equals the rate of information generation about the trajectory outcome, $\dot{I}^{\mathbf{\Phi}}[\mathbf{\Phi}; S_+] \geqslant 0$, minus the rate of information loss about the trajectory origin, $\dot{I}^{\mathbf{\Phi}}[\mathbf{\Phi}; S_-] \leqslant 0$.

Separately analyzing each individual trajectory subensemble, the entropy production rate within a stationary $m_i \to m_i$ subensemble is zero, while the entropy production rate within a reactive $m_i \to m_{j\neq i}$ subensemble is

$$p(S_- = m_i, S_+ = m_j)\dot{\Sigma}_{i\to j}$$

$$= \sum_{\phi\phi'} T_{\phi\phi'}\pi(\phi')p(S_+ = m_j|\phi)p(S_- = m_i|\phi') \ln \frac{p(S_+ = m_j|\phi)p(S_- = m_i|\phi')}{p(S_+ = m_j|\phi')p(S_- = m_i|\phi)} \tag{26a}$$

$$= \sum_{\phi\phi'} T_{\phi\phi'}\pi(\phi')p(S_+ = m_j|\phi)p(S_- = m_i|\phi') \left[\ln \frac{r_{i\to j}(\phi)}{1 - r_{i\to j}(\phi)} - \ln \frac{r_{i\to j}(\phi')}{1 - r_{i\to j}(\phi')} \right] \tag{26b}$$

$$\geqslant 0, \tag{26c}$$

which equals the flux-weighted average of changes to the local reaction coordinate. The forward and reverse transition-path ensembles for a given pair of mesostates have equal entropy production rates,

$$p(m_i, m_j)\dot{\Sigma}_{i\to j} = p(m_j, m_i)\dot{\Sigma}_{j\to i} \tag{27a}$$

$$\dot{\Sigma}_{i\to j} = \dot{\Sigma}_{j\to i}, \tag{27b}$$

where we use $p(m_i, m_j) = p(m_j, m_i)$.

Recall that the system (without reference to trajectory origin or outcome) dynamically evolves along the supertrajectory at equilibrium with no entropy production. The entropy production rate $\dot{\Sigma}_{i\to j}$ physically quantifies the necessary dissipation in a hypothetical system evolving according to the detailed-balance-breaking dynamics of that subensemble of trajectories that originate in mesostate m_i and next reach m_j.

V. MULTIPARTITE ENTROPY PRODUCTION RATES AND COORDINATE RELEVANCE TO THE REACTION

We now consider how the entropy production rates in a given subensemble inform us about system degrees of freedom relevant to the local reaction mechanism. Suppose the system microstate $\mathbf{\Phi}$ is described by d degrees of freedom $X = \{X_1, X_2, \ldots, X_d\}$. We further assume that the dynamics are multipartite, such that only one X_ν changes in a given time step,

$$T_{xx'} = \begin{cases} T_{xx'} & x_\nu \neq x'_\nu, \, x_\mu = x'_\mu \, \forall \mu \neq \nu \\ -\sum_{x'' \neq x'} T_{x''x'} & x = x' \\ 0 & \text{otherwise} \end{cases}. \tag{28}$$

The irreversible entropy production of all reactive subensemble pairs can be split into contributions from each degree of freedom X_ν,

$$\sum_{s_+} \sum_{s_- \neq s_+} p(s_-, s_+)\dot{\Sigma}_{s_-, s_+} = \sum_{s_+} \sum_{s_- \neq s_+} \sum_\nu p(s_-, s_+)\dot{\Sigma}^{X_\nu}_{s_-, s_+}, \tag{29}$$

and the entropy production rate of coordinate X_ν equals the information generation rate due to that coordinate's dynamics,

$$\sum_{s_+} \sum_{s_- \neq s_+} p(s_-, s_+)\dot{\Sigma}^{X_\nu}_{s_-, s_+} = 2\dot{I}^{X_\nu}[\mathbf{\Phi}; S_+] \tag{30a}$$

$$= \dot{I}^{X_\nu}[\mathbf{\Phi}; S_+] - \dot{I}^{X_\nu}[\mathbf{\Phi}; S_-], \tag{30b}$$

following analogous steps as in Eq. (25).

The entropy production rate due to X_ν dynamics quantifies the contribution of this coordinate to the irreversibility of all reactive subensemble pairs,

$$\dot{\Sigma}^{X_\nu}_{s_-, s_+} = \sum_{x, x'} \frac{T_{xx'}\pi(x')p(s_+|x)p(s_-|x')}{p(s_-, s_+)} \ln \frac{p(s_+|x)p(s_-|x')}{p(s_+|x')p(s_-|x)}, \tag{31}$$

where the summation runs over all $\boldsymbol{x}$ and $\boldsymbol{x}'$ such that $x_\nu \neq x'_\nu$, $x_\mu = x'_\mu \, \forall \mu \neq \nu$.

One can find coordinates that are relevant to characterizing all reactive subensemble pairs in the network, or focus on a specific transition and identify a coordinate with maximal entropy production in that particular transition-path ensemble of interest. A one-dimensional coordinate with maximal entropy production rate in the $i \rightarrow j$ transition-path ensemble is the local reaction coordinate [or any invertible transformation thereof, e.g., the pairwise committor Eq. (15)], such that

$$\dot{\Sigma}_{s_-, s_+} = \dot{\Sigma}_{s_-, s_+}^{r_{i \rightarrow j}}. \tag{32}$$

The irreversible entropy production thus provides a meaningful quantitative criterion that determines the informative value of any system coordinate to describing a particular transition.

A. Information geometry

The committor vector suggests an N_{meso}-dimensional reaction space, where each axis corresponds to the conditional probability $p(S_+ = m_i | \boldsymbol{\phi})$ [or equivalently $p(S_- = m_i | \boldsymbol{\phi})$] for each mesostate m_i, ranging from 0–1. The committor vector of a microstate can be interpreted as a point in the reaction space, with mesostates making up the vertices [$p(S_- = m_i | \boldsymbol{\phi} \in m_i) = 1$]. Since the conditional probabilities of each origin must sum to unity, the accessible reaction space is the $(N_{\mathrm{meso}}-1)$-dimensional probability simplex. Here, we show that an information metric exists in this space that is related to the average entropy production rate of all transition-path ensembles.

In Appendix A, we show that the entropy production rate (25) for each forward and reverse transition-path ensemble pair can be rewritten as a contribution to a Kullback-Leibler divergence D_{KL} for each transition [15]:

$$\sum_s p(s)\dot{\Sigma}_s = i^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S_+] - i^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S_-] \tag{33a}$$

$$= \sum_{\boldsymbol{\phi}\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}\pi(\boldsymbol{\phi}')D_{\mathrm{KL}}[p(s|\boldsymbol{\phi}')||p(s|\boldsymbol{\phi})] \tag{33b}$$

$$\approx \frac{1}{2}\sum_{\boldsymbol{\phi}\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}\pi(\boldsymbol{\phi}')[x_\nu - x'_\nu]\mathcal{I}_{\mu\nu}(\boldsymbol{\phi}')[x_\mu - x'_\mu], \tag{33c}$$

where x_ν is the νth component of the vector $\boldsymbol{X}$ parametrizing the multidimensional system microstate, and

$$\mathcal{I}_{\mu\nu}(\boldsymbol{\phi}) \equiv \sum_s p(s|\boldsymbol{\phi}) \frac{\partial \ln p(s|\boldsymbol{\phi})}{\partial x_\mu} \frac{\partial \ln p(s|\boldsymbol{\phi})}{\partial x_\nu} \tag{34a}$$

$$= \sum_i \frac{2}{q_i(\boldsymbol{\phi})} \frac{\partial q_i(\boldsymbol{\phi})}{\partial x_\mu} \frac{\partial q_j(\boldsymbol{\phi})}{\partial x_\nu} \tag{34b}$$

is the Fisher information metric [15] at microstate $\boldsymbol{\phi}$, quantifying the curvature of the probability simplex at that point.

The Fisher information metric defines a square distance

$$\mathrm{d}l^2_{\boldsymbol{\phi}\boldsymbol{\phi}'} \equiv \frac{1}{2}[x_\mu - x'_\mu]\mathcal{I}_{\mu\nu}(\boldsymbol{\phi}')[x_\nu - x'_\nu] \tag{35}$$

in the reaction space for a transition from $\boldsymbol{\phi}' \rightarrow \boldsymbol{\phi}$. Substituting this square distance into Eq. (33c) gives

$$\sum_s p(s)\dot{\Sigma}_s \approx \sum_{\boldsymbol{\phi}\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}\pi(\boldsymbol{\phi}')\mathrm{d}l^2_{\boldsymbol{\phi}\boldsymbol{\phi}'}. \tag{36}$$

The distance $\mathrm{d}l^2_{\boldsymbol{\phi}\boldsymbol{\phi}'}$ between two microstates is the mean transition-path entropy-production rate averaged over all subensembles, a measure of the contribution of that microstate transition to the irreversibility of the joint dynamics of system and trajectory type.

VI. SYSTEM DYNAMICS WITHIN TRAJECTORY-ORIGIN SUBENSEMBLES

Throughout this work, we have focused on trajectory subensembles defined by both trajectory outcome and origin, which highlight reactive trajectory segments in the underlying dynamics. However, a more typical dynamical coarse graining specifies only the trajectory origin [10]. This type of coarse graining arises in milestoning, a method originally developed to sample the transition region between two metastable states by choosing mesostates (milestones) at discrete intervals on a one-dimensional reaction coordinate [11], or as edges of a grid in multidimensional spaces [20]. By initiating trajectories at one mesostate and collecting statistics on the time to reach the next mesostate and the identity of the next mesostate, one builds a coarse-grained model of the dynamics in the transition regions, which yields the overall rate constant for a two-mesostate reaction. Further, milestoning can be used to build a coarse-grained model of dynamics between mesostates representing metastable states of the system [8,9], similar to those expressed in Markov-state models [1,6,7].

Importantly, states in a coarse-grained model built from milestoning are defined by the trajectory origin [10] and not the current system microstate. Therefore the marginal dynamics in the space of the trajectory origin constitutes a coarse-graining of trajectories and not a coarse graining of microstates. This contrasts with Markov-state models where the configuration space is coarse grained to fully partition it into mesostates [6,21], and coarse-grained dynamics are constructed by observing the frequency of transitions across the boundaries between mesostates.

We adapt the framework presented here to trajectory-origin dynamics. Partitioning the long equilibrium trajectory into segments with the same trajectory origin yields N_{meso} subensembles, each defined by a particular trajectory origin. Since the underlying dynamics are Markovian and do not depend on trajectory origin, the joint transition rate for a $(\boldsymbol{\phi}', s'_-) \rightarrow (\boldsymbol{\phi}, s_-)$ transition is $T_{\boldsymbol{\phi}\boldsymbol{\phi}'}$, unmodified from the original dynamics, in contrast to Eq. (16), which also specifies the trajectory outcome. The subensembles with common trajectory origin have a net flux between microstates, since trajectories that were at mesostate m_i will eventually reach another mesostate. In Appendix B, we expand the change in

the joint entropy $H(\boldsymbol{\phi}, s_-)$ to quantify the irreversible entropy production,

$$\dot{H}^{\text{irr}}(\boldsymbol{\Phi}, S_-) = \sum_{\boldsymbol{\phi}, \boldsymbol{\phi}'} \sum_{s_-} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(s_- | \boldsymbol{\phi}') \ln \frac{T_{\boldsymbol{\phi}\boldsymbol{\phi}'} p(\boldsymbol{\phi}', s_-)}{T_{\boldsymbol{\phi}'\boldsymbol{\phi}} p(\boldsymbol{\phi}, s_-)}$$

$$(37\text{a})$$

$$= \sum_i \sum_{\boldsymbol{\phi}, \boldsymbol{\phi}' \notin m_{j \neq i}} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(S_- = m_i | \boldsymbol{\phi}')$$

$$\times \ln \frac{p(S_- = m_i | \boldsymbol{\phi}')}{p(S_- = m_i | \boldsymbol{\phi})} \qquad (37\text{b})$$

$$= -\dot{I}^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S_-] \qquad (37\text{c})$$

$$\geqslant 0, \qquad (37\text{d})$$

where $\dot{I}^{\boldsymbol{\Phi}}[\boldsymbol{\Phi}; S_-]$ is the change (here a decrease) in information between system state and trajectory origin due to system transitions within the same trajectory-origin subensemble. Thus the equivalence between irreversible entropy production and information flows in trajectory subensembles is maintained when only the trajectory origin is used to partition the system's dynamics. A similar equality can be found for trajectory-outcome dynamics.

The entropy production rate within specific subensemble $S_- = m_i$ is

$$p(S_- = m_i) \dot{\Sigma}_{m_i} = \sum_{j \neq i} \sum_{\boldsymbol{\phi} \notin m_j} \sum_{\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(S_- = m_i | \boldsymbol{\phi}')$$

$$\times \ln \frac{p(S_- = m_i | \boldsymbol{\phi}')}{p(S_- = m_i | \boldsymbol{\phi})}. \qquad (38)$$

The origin-subensemble entropy production rate can be interpreted as indicating the irreversibility of dynamics for the system leaving mesostate m_i, and can be used to find coordinates whose dynamics are most responsible for loss of information about the trajectory origin. Coarse graining by partitioning a long trajectory rather than partitioning configuration space therefore allows identification of underlying microscopic dynamics that are informative about the coarse-grained transitions, providing a deeper understanding that spans both mesoscopic and microscopic scales.

VII. DISCUSSION

For reaction networks with many distinct conformations, determining the relative stability of key species (thermodynamics) and the rate of transitions between them (kinetics) allows a global understanding of the function of the system. In this paper, we investigated how the underlying microscopic dynamics provide information about the reactions (mechanisms) in the system by combining tools from information theory and stochastic thermodynamics with multimesostate transition-path theory.

The main significance of the results is a quantitative criterion for reaction-coordinate identification. In two-mesostate transition-path theory, it is well known that the committor is the perfect reaction coordinate, which we clarify as a sufficient statistic in an information-theoretic sense. In a multimesostate transition-path network, we find that the statements are slightly modified. An N_{meso}-dimensional committor vector is fully informative about the entire reaction network, and committor components can be combined to yield $\binom{N_{\text{meso}}}{2}$ reaction coordinates, one for each conjugate pair of mesostate transitions. Each is a local reaction coordinate that provides a distinctive measure of the system components relevant for a particular transition-path ensemble.

The transition-path entropy production quantifies the information that system dynamics generate, resolving whether the system is heading to m_i or m_j next, and the rate at which information about the most recent mesostate is lost. This can help with identifying coordinates that describe one transition, allowing for directed manipulation of a small number of system components that can have an impact on coarse-grained kinetics. While we express this entropy production in terms of committor-vector elements, we stress that determination of the committor is not necessary to calculate entropy production. Entropy production simply measures the imbalance in forward and reverse flux in an ensemble of trajectories; therefore, one can generate an ensemble of transition paths for the reaction of interest and count the number of forward and reverse transitions that occur between discrete regions of configuration space; this information is sufficient to determine the entropy production. To identify a one-dimensional coordinate, the ensemble can be projected onto a single coordinate, and the flux asymmetry on that coordinate can be estimated; making modifications to that coordinate to maximize entropy production is a general procedure to identify the reaction coordinate for that transition-path ensemble.

In addition to identification of local reaction coordinates, the mutual information between system microstate and trajectory type quantifies the information the coarse-grained model resolves about the underlying microstates. This information depends on the choice of system mesostates and could potentially provide a criterion for optimizing their location in configuration space. Current procedures for selecting mesostates or reduced-coordinate representations typically depend on capturing the longest relaxation modes in the system, which can be quantified by, e.g., the VAMP score [22] or Kemeny constant [23]. Using static mutual information and information-generation rates could provide an alternative perspective on optimizing the choice of mesostates that focuses more on capturing informative aspects of the system entropy rather than capturing the longest timescales.

The multimesostate transition-path description represents a coarse graining of trajectories rather than of microstates. This builds time asymmetry into the dynamics on a fundamental level. We considered here systems at equilibrium, but the results could be generalized to coarse-grain trajectories from nonequilibrium steady-state distributions with underlying time asymmetry. The key distinction from our present results is that the forward and reverse committor vectors would no longer be identical, and some of the simplifications in this paper would no longer apply. For example, the forward and reverse transition-path ensembles between two mesostates would no longer be time-reversed twins and would generally have distinct features, which can be captured by different forward and reverse reaction coordinates. However, the transition-path ensemble entropy productions

from the two ensembles would still allow for identification of these coordinates. This is an interesting future direction of study.

We derived this information thermodynamics for a discrete microstate system with master-equation dynamics. In principle, these results can be generalized to continuous configuration spaces and dynamics, which extends their applicability. The information metric in Sec. V A can be formally introduced in a continuous space, and provides a measure of information distance traveled by a particular path in configuration space. This may be of interest in finding an optimal path that represents a transition-path ensemble, currently

typically represented by minimum-free-energy or maximum-flux transition paths [24,25].

APPENDIX A: INFORMATION GEOMETRY

The entropy production rate for each forward and reverse transition-path ensemble pair (26a) can be rewritten as a contribution to a Kullback-Leibler divergence D_{KL} for each transition [15]:

$$\sum_s p(s)\dot{\Sigma}_s = i^{\mathbf{\Phi}}[\mathbf{\Phi}; S_+] - i^{\mathbf{\Phi}}[\mathbf{\Phi}; S_-] \tag{A1a}$$

$$= \sum_{\boldsymbol{\phi},\boldsymbol{\phi}',s_+} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}\pi(\boldsymbol{\phi}')p(s_+|\boldsymbol{\phi})\ln\frac{p(s_+|\boldsymbol{\phi})}{p(s_+|\boldsymbol{\phi}')} - \sum_{\boldsymbol{\phi},\boldsymbol{\phi}',s_-} T_{\boldsymbol{\phi}'\boldsymbol{\phi}}\pi(\boldsymbol{\phi})p(s_-|\boldsymbol{\phi})\ln\frac{p(s_-|\boldsymbol{\phi}')}{p(s_-|\boldsymbol{\phi})} \tag{A1b}$$

$$= \sum_{\boldsymbol{\phi},\boldsymbol{\phi}',s_+,s_-} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}\pi(\boldsymbol{\phi}')p(s_+|\boldsymbol{\phi})p(s_-|\boldsymbol{\phi})\ln\frac{p(s_+|\boldsymbol{\phi})}{p(s_+|\boldsymbol{\phi}')} - \sum_{\boldsymbol{\phi},\boldsymbol{\phi}',s_+,s_-} T_{\boldsymbol{\phi}'\boldsymbol{\phi}}\pi(\boldsymbol{\phi})p(s_-|\boldsymbol{\phi})p(s_+|\boldsymbol{\phi})\ln\frac{p(s_-|\boldsymbol{\phi}')}{p(s_-|\boldsymbol{\phi})} \tag{A1c}$$

$$= \sum_{\boldsymbol{\phi},\boldsymbol{\phi}',s_+,s_-} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}\pi(\boldsymbol{\phi}')p(s_+|\boldsymbol{\phi})p(s_-|\boldsymbol{\phi})\ln\frac{p(s_+|\boldsymbol{\phi})p(s_-|\boldsymbol{\phi})}{p(s_+|\boldsymbol{\phi}')p(s_-|\boldsymbol{\phi}')} \tag{A1d}$$

$$= \sum_{\boldsymbol{\phi}\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}\pi(\boldsymbol{\phi}')D_{\mathrm{KL}}[p(s|\boldsymbol{\phi}')||p(s|\boldsymbol{\phi})] \tag{A1e}$$

$$\approx \frac{1}{2}\sum_{\boldsymbol{\phi}\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}\pi(\boldsymbol{\phi}')[x_\nu - x_\nu']\mathcal{I}_{\mu\nu}(\boldsymbol{\phi}')[x_\mu - x_\mu'], \tag{A1f}$$

where x_ν is the νth component of the vector $\boldsymbol{X}$ parametrizing the multidimensional system microstate, and

$$\mathcal{I}_{\mu\nu}(\boldsymbol{\phi}) \equiv \sum_s p(s|\boldsymbol{\phi})\frac{\partial \ln p(s|\boldsymbol{\phi})}{\partial x_\mu}\frac{\partial \ln p(s|\boldsymbol{\phi})}{\partial x_\nu} \tag{A2a}$$

$$= \sum_{ij} q_i(\boldsymbol{\phi})q_j(\boldsymbol{\phi})\frac{\partial \ln q_i(\boldsymbol{\phi})q_j(\boldsymbol{\phi})}{\partial x_\mu}\frac{\partial \ln q_i(\boldsymbol{\phi})q_j(\boldsymbol{\phi})}{\partial x_\nu} \tag{A2b}$$

$$= \sum_{ij} \frac{1}{q_i(\boldsymbol{\phi})q_j(\boldsymbol{\phi})}\frac{\partial q_i(\boldsymbol{\phi})q_j(\boldsymbol{\phi})}{\partial x_\mu}\frac{\partial q_i(\boldsymbol{\phi})q_j(\boldsymbol{\phi})}{\partial x_\nu} \tag{A2c}$$

$$= \sum_{ij}\left[\frac{q_i(\boldsymbol{\phi})}{q_j(\boldsymbol{\phi})}\frac{\partial q_j(\boldsymbol{\phi})}{\partial x_\mu}\frac{\partial q_j(\boldsymbol{\phi})}{\partial x_\nu} + \frac{\partial q_i(\boldsymbol{\phi})}{\partial x_\mu}\frac{\partial q_j(\boldsymbol{\phi})}{\partial x_\nu} + \frac{\partial q_j(\boldsymbol{\phi})}{\partial x_\mu}\frac{\partial q_i(\boldsymbol{\phi})}{\partial x_\nu} + \frac{q_j(\boldsymbol{\phi})}{q_i(\boldsymbol{\phi})}\frac{\partial q_i(\boldsymbol{\phi})}{\partial x_\mu}\frac{\partial q_i(\boldsymbol{\phi})}{\partial x_\nu}\right] \tag{A2d}$$

$$= \sum_i \frac{2}{q_i(\boldsymbol{\phi})}\frac{\partial q_i(\boldsymbol{\phi})}{\partial x_\mu}\frac{\partial q_j(\boldsymbol{\phi})}{\partial x_\nu} \tag{A2e}$$

is the Fisher information metric [15] at microstate $\boldsymbol{\phi}$.

APPENDIX B: TRAJECTORY-ORIGIN SUBENSEMBLES

The transition rate for a $\boldsymbol{\phi}' \to \boldsymbol{\phi}$ transition within a trajectory-origin subensemble is unchanged from the original system dynamics $T_{\boldsymbol{\phi}\boldsymbol{\phi}'}$ since they are Markovian. The local detailed-balance relation that quantifies time asymmetry is

$$\frac{T^{s_-}_{\boldsymbol{\phi}\boldsymbol{\phi}'} p(\boldsymbol{\phi}', s_-)}{T^{s_-}_{\boldsymbol{\phi}'\boldsymbol{\phi}} p(\boldsymbol{\phi}, s_-)} = \frac{p(s_-|\boldsymbol{\phi}')}{p(s_-|\boldsymbol{\phi})}. \tag{B1}$$

Therefore a $\boldsymbol{\phi}' \to \boldsymbol{\phi}$ transition where $p(s_-|\boldsymbol{\phi})$ changes has some time asymmetry in the trajectory-origin subensembles. The net flux for a $\boldsymbol{\phi}' \to \boldsymbol{\phi}$ transition within subensemble s_- is the difference between the forward and reverse transition frequencies:

$$T^{s_-}_{\boldsymbol{\phi}\boldsymbol{\phi}'} p(\boldsymbol{\phi}', s_-) - T^{s_-}_{\boldsymbol{\phi}'\boldsymbol{\phi}} p(\boldsymbol{\phi}, s_-) = T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(s_-|\boldsymbol{\phi}') - T_{\boldsymbol{\phi}'\boldsymbol{\phi}} \pi(\boldsymbol{\phi}) p(s_-|\boldsymbol{\phi}) \tag{B2a}$$

$$= T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}')[p(s_-|\boldsymbol{\phi}') - p(s_-|\boldsymbol{\phi})]. \tag{B2b}$$

Thus there is net flux within the subensemble toward microstates with lower values of $p(s_-|\boldsymbol{\phi})$; that is, a net flux away from the origin mesostate $S_- = m_i$ towards the $M - 1$ other mesostates, which act as absorbing boundary conditions for dynamics in subensemble $S_- = m_i$. To maintain steady state within the subensemble, trajectories are regenerated in mesostate m_i with flux equal to the total flux of trajectories departing $S_- = m_i$ to all other mesostates,

$$\sum_{j \neq i} \nu_{i \to j}. \tag{B3}$$

Now consider the change in joint entropy $H(\boldsymbol{\Phi}, S_-)$. As before, the joint entropy is split into three contributions

$$0 = \mathrm{d}_t H(\boldsymbol{\Phi}, S_-) \tag{B4a}$$

$$= \sum_{\boldsymbol{\phi},\boldsymbol{\phi}'} \sum_{s_-,s'_-} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(s'_-|\boldsymbol{\phi}') \ln \frac{p(\boldsymbol{\phi}', s'_-)}{p(\boldsymbol{\phi}, s_-)} \tag{B4b}$$

$$= \underbrace{\sum_{\boldsymbol{\phi},\boldsymbol{\phi}'} \sum_{s_-} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(s_-|\boldsymbol{\phi}') \ln \frac{T_{\boldsymbol{\phi}\boldsymbol{\phi}'} p(\boldsymbol{\phi}', s_-)}{T_{\boldsymbol{\phi}'\boldsymbol{\phi}} p(\boldsymbol{\phi}, s_-)}}_{\dot{H}^{\mathrm{irr}}} - \underbrace{\sum_{\boldsymbol{\phi},\boldsymbol{\phi}'} \sum_{s_-} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(s_-|\boldsymbol{\phi}') \ln \frac{T_{\boldsymbol{\phi}\boldsymbol{\phi}'}}{T_{\boldsymbol{\phi}'\boldsymbol{\phi}}}}_{\dot{H}^{\mathrm{env}}}$$

$$+ \underbrace{\sum_{\boldsymbol{\phi},\boldsymbol{\phi}'} \sum_{s_- \neq s'_-} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(s'_-|\boldsymbol{\phi}') \ln \frac{p(\boldsymbol{\phi}', s'_-)}{p(\boldsymbol{\phi}, s_-)}}_{\dot{H}^{\mathrm{sub}}}. \tag{B4c}$$

The subensemble transitions contribute

$$\dot{H}^{\mathrm{sub}}(\Phi, S_-) = \sum_{\boldsymbol{\phi},\boldsymbol{\phi}'} \sum_{s_- \neq s'_-} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(s'_-|\boldsymbol{\phi}') \ln \frac{\pi(\boldsymbol{\phi}') p(s'_-|\boldsymbol{\phi}')}{\pi(\boldsymbol{\phi}) p(s_-|\boldsymbol{\phi})} \tag{B5a}$$

$$= \sum_i \sum_{\boldsymbol{\phi} \in m_{j \neq i}} \sum_{\boldsymbol{\phi}' \notin m_{j \neq i}} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(S_- = m_i|\boldsymbol{\phi}') \ln \frac{\pi(\boldsymbol{\phi}') p(S_- = m_i|\boldsymbol{\phi}')}{\pi(\boldsymbol{\phi})} \tag{B5b}$$

$$= \sum_i \sum_{\boldsymbol{\phi} \in m_{j \neq i}} \sum_{\boldsymbol{\phi}' \notin m_{j \neq i}} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(S_- = m_i|\boldsymbol{\phi}') \ln \frac{\pi(\boldsymbol{\phi}')}{\pi(\boldsymbol{\phi})}$$

$$+ \sum_i \sum_{\boldsymbol{\phi} \in m_{j \neq i}} \sum_{\boldsymbol{\phi}' \notin m_{j \neq i}} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') p(S_- = m_i|\boldsymbol{\phi}') \ln p(S_- = m_i|\boldsymbol{\phi}'). \tag{B5c}$$

In Eq. (B5b), we explicitly write the change in entropy for transitions leaving the trajectory origin subensemble for m_i by entering mesostate $m_{j \neq i}$. In Eq. (B5c), we identify two contributions, one arising from the change in system entropy (internal energy) during the transition and the other a change in conditional entropy of the current subensemble, which is reduced since a new mesostate is reached in the transition, thereby eliminating uncertainty about the trajectory type.

The environment entropy change for transitions within a subensemble is

$$\dot{H}^{\text{env}}(\Phi, S_-) = \sum_{\phi,\phi'} \sum_{s_-} T_{\phi\phi'} \pi(\phi') p(s_-|\phi') \ln \frac{T_{\phi\phi'}}{T_{\phi'\phi}} \tag{B6a}$$

$$= \sum_i \sum_{\phi,\phi' \notin m_{j\neq i}} T_{\phi\phi'} \pi(\phi') p(S_- = m_i|\phi') \ln \frac{T_{\phi\phi'}}{T_{\phi'\phi}}, \tag{B6b}$$

which is the average energy change in the system for dynamics in each trajectory-origin subensemble. This cancels the first term in the expression for subensemble entropy change (B5c), since these two contributions account for all energy changes at steady state, which by definition is zero.

The irreversible entropy production for transitions within a subensemble is

$$\dot{H}^{\text{irr}}(\Phi, S_-) = \sum_{\phi,\phi'} \sum_{s_-} T_{\phi\phi'} \pi(\phi') p(s_-|\phi') \ln \frac{T_{\phi\phi'} p(\phi', s_-)}{T_{\phi'\phi} p(\phi, s_-)} \tag{B7a}$$

$$= \sum_i \sum_{\phi,\phi' \notin m_{j\neq i}} T_{\phi\phi'} \pi(\phi') p(S_- = m_i|\phi') \ln \frac{p(S_- = m_i|\phi')}{p(S_- = m_i|\phi)} \tag{B7b}$$

$$= -\dot{I}^{\Phi}[\Phi; S_-] \tag{B7c}$$

$$\geqslant 0, \tag{B7d}$$

where $\sum_{\phi,\phi' \notin m_{j\neq i}}$ indicates a sum over microstates ϕ and ϕ', neither of which is in any mesostate m_j other than m_i, and $\dot{I}^{\Phi}[\Phi; S_-]$ is the information change due to system transitions within the same trajectory-origin subensemble.

The irreversible entropy production within a subensemble $S_- = m_i$ is

$$p(S_- = m_i)\dot{\Sigma}_{m_i} = \sum_{\phi,\phi' \notin m_{j\neq i}} T_{\phi\phi'} \pi(\phi') p(S_- = m_i|\phi') \ln \frac{p(S_- = m_i|\phi')}{p(S_- = m_i|\phi)}. \tag{B8}$$

[1] J.-H. Prinz, B. Keller, and F. Noe, Phys. Chem. Chem. Phys. **13**, 16912 (2011).

[2] B. Peters, Annu. Rev. Phys. Chem. **67**, 669 (2016).

[3] P. Metzner, C. Schutte, and E. Vanden-Eijnden, Multiscale Model. Simul. **7**, 1192 (2009).

[4] W. E and E. Vanden-Eijnden, J. Stat. Phys. **123**, 503 (2006).

[5] M. D. Louwerse and D. A. Sivak, Phys. Rev. Lett. **128**, 170602 (2022).

[6] J. D. Chodera and F. Noé, Curr. Opin. Struct. Biol. **25**, 135 (2014).

[7] V. S. Pande, K. Beauchamp, and G. R. Bowman, Methods **52**, 99 (2010).

[8] C. Schutte, F. Noé, J. Lu, M. Sarich, and E. Vanden-Eijnden, J. Chem. Phys. **134**, 204105 (2011).

[9] N.-V. Buchete and G. Hummer, J. Phys. Chem. B **112**, 6057 (2008).

[10] A. M. Berezhkovskii and A. Szabo, J. Chem. Phys **150**, 054106 (2019).

[11] A. K. Faradjian and R. Elber, J. Chem. Phys. **120**, 10880 (2004).

[12] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).

[13] T. S. Van Erp, D. Moroni, and P. G. Bolhuis, J. Chem. Phys. **118**, 7762 (2003).

[14] R. J. Allen, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett. **94**, 018104 (2005).

[15] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. (John Wiley & Sons, Hoboken, 2006).

[16] Z. W. Tan, E. Guarnera, and I. N. Berezovsky, PLoS Comput. Biol. **14**, e1006686 (2018).

[17] E. Vanden-Eijnden, in *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, edited by G. R. Bowman, V. S. Pande, and F. Noe (Springer, Berlin, 2014), Chap. 7, pp. 91–100.

[18] D. M. Busiello, D. Gupta, and A. Maritan, Phys. Rev. Res. **2**, 023011 (2020).

[19] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).

[20] R. Elber, Annu. Rev. Biophys. **49**, 69 (2020).

[21] D. Hartich and A. Godec, Phys. Rev. X **11**, 041047 (2021).

[22] A. Mardt, L. Pasquali, H. Wu, and F. Noé, Nat. Commun. **9**, 5 (2018).

[23] V. Koskin, A. Kells, J. Clayton, A. K. Hartmann, A. Annibale, and E. Rosta, J. Chem. Phys. **158**, 104112 (2023).

[24] L. Maragliano and E. Vanden-Eijnden, Chem. Phys. Lett. **426**, 168 (2006).

[25] R. Zhao, J. Shen, and R. D. Skeel, J. Chem. Theory Comput. **6**, 2411 (2010).