Directed information flow in reaction networks under energy constraints: A framework for communication and optimal design applications

Maximilian Gehri,[1] Lukas Stelzl,[2] and Heinz Koeppl[1]

[1]*Centre for Synthetic Biology, Technical University of Darmstadt, 64283 Darmstadt, Germany*
[2]*Institute of Molecular Physiology, Johannes Gutenberg University Mainz, 55122 Mainz, Germany*
(*heinz.koeppl@tu-darmstadt.de)
(Dated: January 21, 2026)

Biological information processing is constrained by energetic costs, making it natural to treat information flow and dissipation jointly. We develop a unified framework for continuous-time chemical reaction networks (CRNs) that couples trajectory-level mutual and directed information between disjoint species sets with process-based stochastic thermodynamics for open, multi-reservoir systems. The formulation covers causal conditioning and indistinguishable reactions at the subnetwork level arising from multiple reservoir coupling or projection. To ensure stringent use of stochastic thermodynamics across disciplines, we compile a unified account that treats coarse-graining and multi-reservoir modeling correctly when estimating dissipation. We also give a conversion from species–reaction graphs to local-independence graphs on reactions and link local independence to causally conditioned directed information. Applications are twofold: First, the framework yields a rigorous continuous-time communication model over a CRN, with messages source-encoded by time-dependent chemostat protocols. Capacity is posed over trajectory distributions under principled thermodynamic costs. Graph tools enable a functionality-based classification that distinguishes between encoding reactions and reactions that constitute the transmission path. This facilitates characterizing degrees of freedom in the chemical noisy channel coding problem. Second, we advocate directed information as an objective for naturally evolved and engineered biochemical circuits under finite energy budgets. Case studies with minimal promoter-switching models illustrate framework application and visualize trade-offs between information flow and dissipation.

I. INTRODUCTION

Energy consumption is integral to gene regulatory mechanisms and intracellular signal transduction in both prokaryotes and eukaryotes. For example, chromatin remodeling complexes in eukaryotes use ATP to maintain gene expression profiles [1, 2], while in prokaryotes, ATP conversion to cAMP provides a cofactor for transcription factors such as the catabolite-activating protein (CAP) [3]. These processes indicate that biological information processing is fundamentally constrained by energetic costs. This observation motivates a closer investigation of how upper bounds on energy dissipation shape the ability of cellular systems to process and transmit information.

Gene regulatory networks form an essential part of intracellular information processing [4], and information-theoretic tools are increasingly used to analyze their signaling capabilities and regulatory architectures [5, 6]. There is growing evidence that cells encode information in the spatiotemporal variation of populations of molecular species when responding to internal or external stimuli [7–10]. Classical information theory further implies that in systems with feedback and network-mediated memory, meaningful communication can only be properly characterized by information-theoretic quantities defined over entire trajectories, as only these capture the causal temporal structure required to model such signaling processes consistently [11]. Markovian stochastic chemical reaction networks (CRNs) offer a natural modeling framework in this context, capturing both the discrete and stochastic nature of molecular reactions. Any static or dynamical representation of information is thus accurately captured in the trajectories of a CRN. In complement, CRNs also allow an accurate characterization of thermodynamic costs.

In particular, we advocate the use of directed information (DI) between continuous-time stochastic processes [11–13] as a natural measure of causal directional information flow in CRNs. Especially in the context of network-based signal processing and causal feedback encoding of information we will show that the DI represents an upper bound on the information transmission rate between an external stimulus trajectory and the molecular representation of the processed signal. We generalize existing expressions for mutual information (MI) between trajectories to quantify the total information exchange between subnetworks, i.e., subpopulations of species within a CRN. In addition, we derive the corresponding DI, which captures the causal information flow from one subnetwork to another over time. Such subnetwork formulations are especially relevant in settings where multiple external signals are sensed by distinct molecular species and subsequently integrated through internal molecular interactions to produce a response involving multiple downstream species [14, 15]. While we do not require the entire CRN to be bipartite or multipartite, we impose the condition that the two subnetworks of interest do not undergo simultaneous state changes [16]. As a result, the joint dynamics of these two subnetworks form a bipartite, but potentially non-Markovian, marginal process, even though the full CRN is assumed to be Markovian and may exhibit non-bipartite structure. For example, consider the species X, A, and Y in a Markovian conversion network $\mathsf{X} \rightleftarrows \mathsf{A} \rightleftarrows \mathsf{Y}$. The full system is Markovian and non-bipartite, as both $\{\mathsf{X}, \mathsf{A}\}$ and $\{\mathsf{Y}, \mathsf{A}\}$ can change simultaneously. However, the marginal dynamics of the subnetwork $\{\mathsf{X}, \mathsf{Y}\}$ are bipartite and non-Markovian.

As a harmonizing model formulation, we use Kurtz's process-based description of CRNs to align stochastic thermodynamics (sTD) with trajectory-level information theory. This

facilitates the analysis of trade-offs between trajectory-level information processing and energy dissipation of metabolic resources in thermodynamic non-equilibrium systems. Such trade-offs between energy dissipation and precision in cellular decision making and adaptation have been described as an emerging general design principle of biological systems [17].

We provide a comprehensive introduction to stochastic thermodynamics for CRNs. This introduction is certainly longer than is appropriate for a research article. We justify its extent by aiming to serve as a tutorial and thereby promote rigorous thermodynamic modeling across scientific disciplines. An important feature of our exposition is the explicit integration of internal microscopic degrees of freedom, which allows us to distinguish between the physically measurable dissipated heat and the related, but not equivalent, mesoscopic heat. We emphasize the necessity of accurately accounting for multiple driven reaction mechanisms and avoiding violations of underlying coarse-graining assumptions. In time-independent settings, the non-equilibrium steady-state (NESS) heat dissipation rate can be identified with the entropy production rate (EPR). In contrast, under time-dependent driving, internal microscopic degrees of freedom must be considered explicitly to correctly quantify entropic balances.

Finally, we explore unified applications of information theory and stochastic thermodynamics to CRNs. First, we propose a thermodynamically consistent chemical communication model in which external information is encoded in the temporal dynamics of a subnetwork and processed through intermediate species, allowing for memory, feedback, and physically principled constraints on energy consumption. Second, we propose energetically constrained DI as an objective function for deriving energy-efficient designs of biological circuits. This can serve both as a testable prediction of optimal information flow as an emergent evolutionary trait and as a foundation for synthetic biology design.

A key contribution is the identification of network decomposition rules that allow reactions to be classified into "input/output modulating", "jointly modulating", "feedforward" and "feedback" reactions. This provides a more detailed understanding of the specific function of a reaction in terms of information processing and makes information theoretic optimization problems practically accessible.

We anticipate that continued developments in this direction will support the construction of genetic circuits that optimally balance information processing and energetic cost.

A. Previous related work

A powerful formalization of the description of Markovian chemical reaction networks was provided by Kurtz and Anderson [18]. They demonstrated how counting processes relevant to these networks can be represented using Poisson-type processes, leading to a random time-change representation. Therewith, they obtain a stochastic equation for continuous-time Markov jump processes, i.e., their state-level description. Our introduction to CRNs largely follows their work, supplemented by some thermodynamically relevant concepts.

Information theory was pioneered by Claude E. Shannon and laid the foundation for the digital age. One of his landmark contributions was the first channel coding theorem [19], which determines the maximum rate at which information can be reliably transmitted despite noise corrupting the received signal. A key aspect of Shannon's work was his abstraction of noise corruption through the concept of a transmission channel, characterized by the conditional distribution of the output given the encoded message. He introduced mutual information, a functional of the joint distribution of input and output of the channel, to quantify the stochastic dependency between input X and output Y and thereby the information exchange between them. Channel coding theorems identify the supremum of the mutual information for a given transmission channel, known as information (channel) capacity, with the maximum information encoding rate at which the probability of a decoding error of the sent message from the observed signal Y asymptotically approaches zero in the limit of infinite transmission duration. This limiting rate is known as the coding (channel) capacity. Following Shannon, channel coding theorems and other fundamental theoretical results like source-channel separation theorems [20] have been derived for various communication channels.

Marko developed a theory of bidirectional communication, highlighting that communication, unlike mere transmission, is intrinsically bidirectional [21]. Building on his ideas, Massey established a consistent theory of information transmission that incorporates the use of feedback by emphasizing the distinction between statistical and causal dependence [11]. Feedback is not an inherent feature of the channel itself; rather, it is integrated into the encoding of the signals transmitted through the channel. In particular, Massey introduced the concept of directed information and demonstrated that, when used in place of mutual information, it could sometimes accurately extend the channel coding theorems for causal communication systems to transmission with the use of feedback. For channels that are used without feedback, directed information from X to Y reduces to the mutual information between them. In channel coding theorems for causal communication systems with feedback, the coding capacity is identified with the information channel capacity, which is defined either as the supremum of the mutual information between the message and the output sequence, or sometimes as the supremum of the directed information between the message-encoding sequence and the output sequence [20, 22]. A generalization of the notion of directed information to continuous-time trajectories is due to Weissman *et al.* [12]. The supremum for the continuous-time information capacity is taken over the set of distributions of the channel encoding X given the message and then the message representations themselves [23, 24]. To the best of our knowledge, no full communication channel model for CRNs has been proposed to date. Uda, for instance, notes that the concept of information transmission in systems biology remains underdeveloped [25].

Although Barlow [26] has already hypothesized that biological signal processing systems may function as communication systems in the early 60s, applications of information theory in systems biology have only started to develop into a

research field for around two decades [25]. Research in this field has addressed a range of problems, including the analysis of connectivity, sensitivity, and cross-talk in signaling pathways [27]; the identification of optimal cellular adaptation strategies and their adaptation targets [28]; and the study of gene regulatory motifs at steady state [6], at specific time points [29], or in terms of dynamic features such as amplitude and frequency [14]. Tostevin and ten Wolde first considered the MI between trajectories in the biological context, recognizing that relevant information may be contained in the temporal dynamics of the input signal [30]. They provide the mutual information rate (MIR), i.e., the asymptotic differential information gain, under a linear-noise Gaussian approximation. Lestas *et al.* used MI between trajectories to lower bound the cost of suppressing molecular fluctuations via feedback under a Langevin approximation [31], later extended by Nakahira *et al.* who identified continuous-time DI as the appropriate measure [32]; however, the stringency of these lower bounds has been questioned by Parag [33].

A measure-theoretic account for the mutual information between the trajectories of two species of a class of biochemical reaction networks was first given by Duso and Zechner [34] and later generalized by Moor and Zechner [35]. Moor and Zechner further showed that the trajectory mutual information decomposes into a pair of measure-theoretically defined transfer entropies [36]. The transfer entropy is in fact identical to the DI under certain conditions. See Appendix C for a brief comparison. In this paper, we further generalize the expression for the mutual information in [35] to encompass the information flow between disjoint subnetworks without simultaneous reactions in the two sets. Without this bipartite-type requirement for the subnetworks of interest, the MI has been argued to diverge [16]. In contrast to Spinney *et al.* [36], we equivalently derive the continuous-time DI for CRNs via the structural extremal-based approach of Weissman *et al.* [12] instead of explicitly constructing a Radon-Nikodym derivative. Directed information remains well defined without the bipartite assumption, whereas mutual information generally does not.

Sinzger *et al.* and Gehri *et al.* advanced techniques for analytically and numerically computing the mutual information of the Poisson-type channel modulated by a telegraph process, which can be regarded as a minimal module of a CRN [37–39]. Moor and Zechner [35] proposed several methods in the context of CRNs: a combination of stochastic simulation (SSA) [40] and moment closure approximation, a quasi-exact approach combining SSA and numerical integration, and an analytical approximation. Lastly, Reinhardt *et al.* developed an exact Monte Carlo method called path weight sampling to compute the mutual information between trajectories [41], which has recently been adapted for the computation of transfer entropies [42]. Analytical derivations of information channel capacities for the Poisson-type channel under peak and average-constraints on the message waveforms have been put forth by Kabanov [23], Davis [43] and Frey [24], where Frey also considered marked Poisson-type channels. For small biology-motivated systems, channel capacities have been analytically derived by Thomas and Eckford [44–46].

The combination of thermodynamics and information theory of biological information processing has been discussed in several studies. Some early works proposed the entropy production rate (EPR) as a measure of the energetic cost of cellular information processing and decision-making [47, 48]. Later, models of diffusive transport were used to assess the thermodynamic cost of moving information-carrying molecules across cellular compartments [49]. More recently, stochastic thermodynamics has been combined with communication theory to analyze finite-state bipartite Markov systems at nonequilibrium steady state, treating each channel use as a readout of a steady-state output symbol [50]. In that framework, entropy production is not a constraint but a quantity co-optimized alongside information capacity, revealing trade-offs in the steady state. Another approach considers a discrete memoryless channel driven by an i.i.d. binary input, where the output is observed only once sufficient correlation between input and output is established [51].

The idea that information flow may be an emergent optimization principle in nervous systems, shaped by evolution, was first proposed by Attneave [52] and later formalized by Barlow [26] as the "efficient coding hypothesis," suggesting that neurons evolve to reduce redundancy and maximize information about the environment. This notion was given a concrete mathematical formulation by Linsker through the principle of maximum information preservation (infomax) [53]. The optimization paradigm was subsequently extended to systems biology, particularly in the context of gene regulation, by Tkačik *et al.* [54]. Numerous follow-up studies have built on this foundation, which we do not review here for brevity.

B. Outline and notation

The paper is organized as follows. In Sec. II we review the process-based description of microscopically reversible Markovian CRNs as developed by Kurtz and Anderson [18].

The concepts of mutual information and directed information between continuous-time processes are both introduced in a measure-theoretic way in Sec. III. In particular, the MI is defined via a Radon-Nikodym derivative and the DI and its conditional versions are defined via the extremal-based approach of Weismann *et al.* [12]. We then use a classical expression [55] that links the mutual information between jump processes to the characterization of the counting processes of the distinct reactions. To obtain the MI between disjoint bipartite-type subnetworks of the CRN, we propose a notion of indistinguishability of different reactions from the perspective of the subnetworks. We also provide a proof that, without the bipartite assumption, mutual information is ill-defined due to the non-existence of the corresponding Radon–Nikodym derivative. A similar expression for the directed information between subnetworks is derived rigorously and linked to McFadden's point process entropy [56]. The DI expressions presented are also valid in the case that certain reactions change input and output species simultaneously.

The basics of thermodynamically consistent modeling for closed and open chemical reaction networks are outlined Sec.

IV A and IV B. Then, in Sec. IV C we introduce thermodynamic notions on the trajectory-level in line with coarse-graining of a microscopic system (cf. Supp. Sec. S2.1). The common identification of the EPR of the NESS with the heat dissipation rate is given in Sec. IV D. Common decompositions of heat dissipation and entropy production are briefly outlined in Sec. IV E. Precautions to prevent underestimation of the EPR are detailed in Sec. IV F. In Sec. IV G the CRN description of stochastic thermodynamics is linked to the equivalent multi-reservoir description for general CTMCs. As an addition, the relation of thermodynamic irreversibility and time reversal is outlined in Supp. Sec. S2.4, together with a summary of equivalent equilibrium conditions.

Section V introduces local independence graphs for CRNs, which will later be used to formulate a chemical communication system, categorize its the reactions into different functional classes and separate design degrees of freedom in a chemical reaction network from a fixed channel.

Combined applications of information theory and stochastic thermodynamics will be discussed in Section VI. The primary focus is laid on a Shannon-theoretic communication model in Sec. VI A, where a chemical reaction network acts both is used to encode and transmit information in continuous-time from one set of species to another. First, the model of a chemical communication system is introduced. Second, we discuss the noisy channel coding problem in the context of chemical communication systems, which entails rigorous mathematical definitions of chemical codes, information rates and capacity. Third, we introduce quantities from stochastic thermodynamics as physically principled constraints on the allowed energy dissipation rate or chemical work of the system. Forth, we introduce the dual perspective on information capacity under energy constraints, namely, the minimum energy dissipation required per information unit as a function of the code rate. Lastly, we apply the introduced graph-based tools to decompose a chemical communication system into encoding reactions, which belong to the design degrees of freedom, and those reactions that are associated with the invariant channel law.

In Section VI B we discuss how directed information can be used to test the hypothesis that information flow is an emergent optimization principle followed by evolution.

Finally, in Sec. VI C, we showcase how directed information and stochastic thermodynamics are combined for two analytically tractable minimal promoter models as case studies. The first is a nonequilibrium two-state promoter model with parallel reaction channels, and the second is a three-state promoter model. Both models describe the switching between a single transcriptionally active state and an inactive state. We derive analytical expressions for the directed information (DI) between the promoter's switching activity and the RNA copy number, and we optimize the directed information rate (DIR) under an upper bound on the promoter's entropy production rate. In the process show that the thermodynamic modeling pitfalls pointed out lead to inconsistent results.

In the following let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with complete filtration $\mathscr{F} = \{\mathscr{F}_t\}_{t\geq 0}$. The symbol $\vee$ denotes the join operation for sigma-algebras that yields the smallest sigma-algebra that contains all elements of the individual sigma-algebras [57]. For any σ-algebra $\mathfrak{S}$ on Ω let $\overline{\mathfrak{S}}$ denote its completion with $\mathbb{P}$-null sets. With $\int_s^t$ we denote the integration over the interval $(s,t]$, where $s < t$. This convention is important when considering counting process integrals. All stochastic pure jump processes, like counting processes and jump processes on a chemical copy number space, are assumed to have càdlàg trajectories. In the information-theoretic context we refer to X and Y as the channel input and output, respectively. In the context of communication channels the symbol X is typically used for the message-encoding signal, but otherwise we do not explicitly fix the meaning of X. In the notation of quantities in stochastic thermodynamics we proximally follow the IUPAC conventions [58].

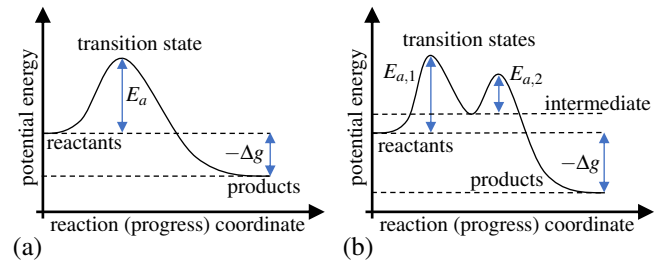


FIG. 1. Exemplary reaction coordinate diagrams for (a) elementary reactions and (b) a non-elementary reaction with a single intermediate. E_a denotes the activation energy of the reaction, while $-\Delta g$ is amount of energy released by the reaction.

II. MARKOVIAN STOCHASTIC CHEMICAL REACTION NETWORKS

A stochastic chemical reaction network (CRN) consists of a finite set of chemical species $\mathscr{S} = \{1, \ldots, |\mathscr{S}|\}$ with identities $\mathsf{U} = \{\mathsf{U}_1, \ldots, \mathsf{U}_{|\mathscr{S}|}\}$ and a finite set of reaction channels $\mathscr{R} = \{1, \ldots, |\mathscr{R}|\}$ with identities $\mathfrak{R}_1, \ldots, \mathfrak{R}_{|\mathscr{R}|}$ with stoichiometric balance equations

$$\mathfrak{R}_r : \quad \sum_{d=1}^{|\mathscr{S}|} a_d^{(r)} \mathsf{U}_d \quad \underset{\lambda_r^-}{\overset{\lambda_r^+}{\rightleftarrows}} \quad \sum_{d=1}^{|\mathscr{S}|} b_d^{(r)} \mathsf{U}_d,$$

wherein $a^{(r)}, b^{(r)} \in \mathbb{Z}_{\geq 0}^{|\mathscr{S}|}$ represent the number of reacting molecules of each species and the number of product molecules, respectively, for the forward reaction, and vice versa for the backward reaction. The difference of product and reactant species of $\mathfrak{R}_r$ forms the stoichiometric change vector $\nu_r = b^{(r)} - a^{(r)}$ of the forward reaction. For each reaction channel, the backward reaction microscopically reverses the forward reaction. We assume that all reactions are elementary [59], i.e., there are no reaction intermediates and each reaction has only a single transition state in the reaction coordinate diagram as illustrated in Fig. 1 (a).

Let $\{U(t) \in \mathbb{Z}_{\geq 0}^{|\mathscr{S}|}\}_{t\geq 0}$ be the continuous-time Markov chain of the copy number vector of the chemical species U. Throughout, we sometimes assume that the rates of the reactions, commonly referred to as propensity functions, obey

stochastic mass action kinetics [18, 60]

$$\lambda_r^+(u) = k_r^+ \prod_{d=1}^{|\mathscr{S}|} \binom{u_d}{a_d^{(r)}},\ \lambda_r^-(u) = k_r^- \prod_{d=1}^{|\mathscr{S}|} \binom{u_d}{b_d^{(r)}} \tag{1}$$

with $u \in \mathbb{Z}_{\geq 0}^{|\mathscr{S}|}$ and stochastic rate constants $k_r^+ > 0$ and $k_r^- \geq 0$. A conversion of the discrete molecule numbers to concentrations in a constant, finite solution volume $\mathsf{V} > 0$ gives rise to another pair of rate constants $\kappa_r^+ > 0$, $\kappa_r^- \geq 0$, such that

$$k_r^+ = \kappa_r^+ \frac{\mathsf{V} \prod_{d=1}^{|\mathscr{S}|} a_d^{(r)}!}{\prod_{d=1}^{|\mathscr{S}|} \mathsf{V}^{a_d^{(r)}}}, \qquad k_r^- = \kappa_r^- \frac{\mathsf{V} \prod_{d=1}^{|\mathscr{S}|} b_d^{(r)}!}{\prod_{d=1}^{|\mathscr{S}|} \mathsf{V}^{b_d^{(r)}}}, \tag{2}$$

due to consistency of physical units in both the deterministic ODE model (of the concentrations) and the probability evolution equation of the stochastic model [18, 61]. κ_r^+ and κ_r^- can be identified with the rate constants of the deterministic law of mass action by taking the macroscopic limit. The mass action assumption can be dropped in the following if it is not explicitly used. A reaction (channel) $\mathfrak{R}_r$ is called microscopically reversible if $k_r^- > 0$, or more generally if $\lambda_r^+(u) > 0$ implies $\lambda_r^-(u + \nu_r) > 0$ and vice versa for all u. A CRN is called microscopically reversible if all reactions are microscopically reversible. In the following, a CRN is assumed to be microscopically reversible unless otherwise noted.

For each reactions of channel $\mathfrak{R}_r$ we denote the counting processes of reactions in the forward and reverse direction as $\{R_r^+(t) \in \mathbb{Z}_{\geq 0}\}_{t \geq 0}$ with $R_r^+(0) = 0$ and $\{R_r^-(t) \in \mathbb{Z}_{\geq 0}\}_{t \geq 0}$ with $R_r^-(0) = 0$, respectively. These can be represented via random time changes [18]

$$R_r^\varepsilon(t) = N_r^\varepsilon \left(\int_0^t \lambda_r^\varepsilon(U(s))\, \mathrm{d}s \right)$$

of $2|\mathscr{R}|$ independent unit-rate Poisson processes $\{N_r^\varepsilon(t) \in \mathbb{Z}_{\geq 0}\}_{t \geq 0}$ with $N_r^\varepsilon(0) = 0$ for $r \in \mathscr{R}$ and $\varepsilon \in \{+, -\}$. Then the random state $U(t)$ of the CRN with random initial value $U(0)$ obeys

$$U(t) = U(0) + \sum_{r=1}^{|\mathscr{R}|} \nu_r (R_r^+(t) - R_r^-(t)).$$

The probability mass function $p(u,t) = \mathbb{P}(U(t) = u)$ follows the chemical master equation [60, 62, 63]

$$\begin{aligned} \partial_t p(u,t) = \sum_{r=1}^{|\mathscr{R}|} & \lambda_r^+(u - \nu_r) p(u - \nu_r, t) \\ & + \lambda_r^-(u + \nu_r) p(u + \nu_r, t) \\ & - (\lambda_r^+(u) + \lambda_r^-(u)) p(u,t) \end{aligned} \tag{3}$$

Before proceeding, we recapitulate the model assumptions made so far, as they are fundamental for stochastic thermodynamics of CRNs. Assuming a Markov process description means that the future dynamics of the mesoscopic system is sufficiently described by the current state and does not depend on the past. In particular, the future dynamics do also not depend on the particular microstate, the system is in, but just on the current mesostate u.

The above assumptions about the system state U have implications for the environment in which the reaction network is embedded. Since the propensity functions are independent of time and space, the embedding environment must be seen as thermally equilibrated and spatially homogeneous from the perspective of the chemical species [64]. That is, the surrounding of each molecule of U homogenizes spatially on a much faster time scale than the reactions occur. In addition, the temperature of the solvent is constant over time and the volume V. Similarly, fluctuations in the solvent's molecular composition, such as changes in pH or ionic strength, are assumed to be small enough to leave the mesoscopic rate constants effectively time-invariant.

In later sections, isothermal–isobaric conditions will be adopted as a natural assumption for biological cells. Under such conditions, the stochastic rate constants $k_r^\pm$ in (2) could, in principle, depend on the mesostate u due to volume fluctuations. However, we assume these fluctuations are negligible, allowing $k_r^\pm$ to remain effectively constant, except in scenarios involving slow, time-dependent changes such as cellular growth.

According to Gillespie, in the light of collision theory the elementary reaction assumption is plausible up to bimolecular reactions and the dynamics is rigorously shown to follow stochastic mass action kinetics under the above assumptions on the environment [64]. In other cases, in particular when the chemical species represent different conformations of polymers as is common in gene expression models, a more careful justification of the elementary reaction assumption and the Markov assumption may be required. The relevance of the elementary reaction assumption for stochastic thermodynamics of CRNs is discussed at the end of Sec. IV A.

III. MUTUAL AND DIRECTED INFORMATION BETWEEN SUBNETWORK TRAJECTORIES

A. Definitions of mutual and directed information between continuous-time trajectories

Let $(X_{[0,t]}, Y_{[0,t]}) = \{X(s) \in \mathscr{X},\ Y(s) \in \mathscr{Y}\}_{s \in [0,t]}$ be a pair of $\mathscr{F}$-adapted [57] stochastic processes with discrete state spaces $\mathscr{X}$ and $\mathscr{Y}$ over the finite time interval $[0,t]$, $t > 0$. Let Σ_X and Σ_Y denote the power sets of $\mathscr{X}$ and $\mathscr{Y}$, respectively. These are the smallest sigma-algebras that contain all singleton sets $\{x\}$, $x \in \mathscr{X}$ or $\{y\}$, $y \in \mathscr{Y}$. Then, the measurable space of trajectories $X_{[0,t]}$ is given by the product space $(\mathscr{X}^{[0,t]}, \Sigma_X^{\otimes[0,t]})$, where $\mathscr{X}^{[0,t]} = \bigtimes_{s \in [0,t]} \mathscr{X}_s$, with $\mathscr{X}_s = \mathscr{X}$, and $\Sigma_X^{\otimes[0,t]} = \bigotimes_{s \in [0,t]} \Sigma_{X,s}$ the product-sigma-algebra (or cylinder-sigma-algebra) with $\Sigma_{X,s} = \Sigma_X$. The product-sigma-algebra is the smallest sigma-algebra on the product set $\mathscr{X}^{[0,t]}$, that can resolve all singleton sets $\{x\} \in \Sigma_X$ at every time-point [65]. In turn, the path space of $(X,Y)_{[0,t]}$ is given by the product space $(\mathscr{X}^{[0,t]} \times \mathscr{Y}^{[0,t]},\ \Sigma_X^{\otimes[0,t]} \otimes \Sigma_Y^{\otimes[0,t]})$. This measurable space can be equipped with induced path-probability measures, either the joint $\mathbb{P}_{[0,t]}^{XY}$ or the product of marginal mea-

sures $\mathbb{P}^X_{[0,t]} \otimes \mathbb{P}^Y_{[0,t]}$ defined as

$$\begin{aligned}\mathbb{P}^{XY}_{[0,t]} &:= \mathbb{P} \circ (X,Y)^{-1}_{[0,t]},\\ \mathbb{P}^{X}_{[0,t]} &:= \mathbb{P} \circ X^{-1}_{[0,t]}, \text{ and}\\ \mathbb{P}^{Y}_{[0,t]} &:= \mathbb{P} \circ Y^{-1}_{[0,t]}.\end{aligned}$$

Here, $(X,Y)^{-1}_{[0,t]} \colon \Sigma_X^{\otimes[0,t]} \otimes \Sigma_Y^{\otimes[0,t]} \to \mathscr{F}$, $X^{-1}_{[0,t]} \colon \Sigma_X^{\otimes[0,t]} \to \mathscr{F}$, and similarly $Y^{-1}_{[0,t]}$ denote the preimages of the respective stochastic process trajectories. By $\mathscr{F}^{XY}$ we denote the internal/natural filtration of the process (X,Y), such that for all $t \geq 0$ it holds

$$\mathscr{F}^{XY}_t = \sigma\left((X_s,Y_s)^{-1}(A) \colon s \in [0,t],\, A \in \Sigma_X \otimes \Sigma_Y\right).$$

Note that if (X,Y) has càdlàg trajectories, then $\mathbb{P}^{XY}_{[0,t]}$ can also be defined as the induced probability measure on the càdlàg space $\mathscr{D}([0,t], \mathscr{X} \times \mathscr{Y}) \subset \mathscr{X}^{[0,t]} \times \mathscr{Y}^{[0,t]}$ with the trace σ-algebra of $\Sigma_X^{\otimes[0,t]} \otimes \Sigma_Y^{\otimes[0,t]}$, i.e., the σ-algebra only contains subsets of $\mathscr{D}([0,t], \mathscr{X} \times \mathscr{Y})$.

The MI between the processes $X_{[0,t]}$ and $Y_{[0,t]}$ is then defined via the Radon-Nikodym derivative [57] between the joint and product of the marginal path-probability measures

$$\mathbb{I}(X_{[0,t]};Y_{[0,t]}) := \mathbb{E}\left[\ln \frac{\mathrm{d}\mathbb{P}^{XY}_{[0,t]}}{\mathrm{d}(\mathbb{P}^X_{[0,t]} \otimes \mathbb{P}^Y_{[0,t]})}\right]$$

if $\mathbb{P}^{XY}_{[0,t]}$ is absolutely continuous with respect to $\mathbb{P}^X_{[0,t]} \otimes \mathbb{P}^Y_{[0,t]}$, typically denoted as $\mathbb{P}^{XY}_{[0,t]} \ll \mathbb{P}^X_{[0,t]} \otimes \mathbb{P}^Y_{[0,t]}$. The mutual information rate (MIR) is defined as its asymptotic slope

$$\bar{\mathbb{I}}(X;Y) := \lim_{t\to\infty} \frac{1}{t} \mathbb{I}(X_{[0,t]};Y_{[0,t]}),$$

with $(X,Y) = \{X(t),Y(t)\}_{t\geq 0}$.

For discrete sequences, the DI was originally defined by Massey [11] as

$$\mathbb{I}_M(X_{1:N} \to Y_{1:N}) = \sum_{n=1}^{N} \mathbb{I}\left(X_{1:n};Y_n \mid Y_{1:n-1}\right), \tag{4}$$

where $X_{1:N} = \{X(n)\}_{n\in\{1,\dots N\}}$ and $Y_{1:N} = \{Y(n)\}_{n\in\{1,\dots N\}}$ are random sequences for all $N \in \mathbb{N}$. As noted by Newton [13], this definition was developed in the context of communication channels, where $Y_n = Y(n)$ is received after $X_n = X(n)$ is transmitted. Thus, the enumeration index n does not represent physical time for the joint process $\{(X(n),Y(n))\}_{n\in\{1,\dots N\}}$. If the index n does denote physical time for the joint sequence, then (4) includes undirected "instantaneous information exchange" terms [66] of the form $\mathbb{I}(X_n;Y_n \mid (X,Y)_{1:n-1})$, which should be excluded when aiming to quantify the causal influence of $X_{1:N}$ on $Y_{1:N}$.

To address this, Newton proposed a causally faithful definition of DI for discrete sequences, where the index represents physical time:

$$\mathbb{I}(X_{1:N} \to Y_{1:N}) = \sum_{n=2}^{N} \mathbb{I}\left(X_{1:n-1};Y_n \mid Y_{1:n-1}\right)$$

To generalize this definition to continuous-time trajectories, we follow the structural extremal-based approach of Weissman *et al.* [12], who defined the directed information for continuous processes by analogy with Massey's formulation as

$$\begin{aligned}&\mathbb{I}_M(X_{[0,t)} \to Y_{[0,t)})\\ &:= \inf_{\boldsymbol{t}\in\mathscr{T}(0,t)} \sum_{i=1}^{n(\boldsymbol{t})} \mathbb{I}\left(X_{[0,t_i)};Y_{[t_{i-1},t_i)} \mid Y_{[0,t_{i-1})}\right),\end{aligned}$$

where $\mathscr{T}(0,t) = \{\boldsymbol{t} = (t_i)_{i\in\{1,\dots,n\}} \colon 0 < t_1 < \dots < t_n = t\}$ is the set of all finite partitions of the interval $[0,t)$, and $n(\boldsymbol{t})$ denotes the length of the partition. However, we modify the definition to follow the causal structure of Newton's proposal instead. Thus, for continuous-time processes on the right-open interval $[0,t)$, we define

$$\begin{aligned}&\mathbb{I}(X_{[0,t)} \to Y_{[0,t)})\\ &:= \sup_{\boldsymbol{t}\in\mathscr{T}(0,t)} \sum_{i=2}^{n(\boldsymbol{t})} \mathbb{I}\left(X_{[0,t_{i-1})};Y_{[t_{i-1},t_i)} \mid Y_{[0,t_{i-1})}\right).\end{aligned} \tag{5}$$

The use of the supremum, in contrast to the infimum, is justifed in Supp. Lemma S1.1.

Directed information on a closed interval $[0,t]$ is given by the right-limit

$$\mathbb{I}(X_{[0,t]} \to Y_{[0,t]}) = \lim_{h\searrow 0} \mathbb{I}(X_{[0,t+h)} \to Y_{[0,t+h)}),$$

and the directed information rate (DIR) is

$$\bar{\mathbb{I}}(X \to Y) := \lim_{t\to\infty} \frac{1}{t} \mathbb{I}(X_{[0,t)} \to Y_{[0,t)}). \tag{6}$$

In other works, e.g., [13, 16], directed information and transfer entropy have been extended to continuous time by taking the continuity limit of their discrete-time forms to construct a time-local rate, which is then integrated. The drawback of this approach for general continuous-time processes is, that boundary terms or discontinuous jumps may be overlooked (although discontinuities do not occur for the typically considered jump processes). In contrast, the approach of Weissman *et al.* offers a structurally sound alternative, as it is rooted in measure theory and thus defined directly in terms of continuous-time path space probability measures, similar to explicitly measure theoretic definitions (cf. Appendix C or the Appendix of [36]).

In Supp. Sec. S1.2 we show that, under certain continuity conditions, the DI satisfies

$$\mathbb{I}(X_{[0,t]} \to Y_{[0,t]}) = \int_0^t \lim_{h\searrow 0} \frac{1}{h} \mathbb{I}\left(X_{[0,s)};Y_{[s,s+h)} \mid Y_{[0,s)}\right) \mathrm{d}s, \tag{7}$$

which is easier to evaluate in practice than definition (5). Similarly, for the continuous generalization of Massey's DI, we obtain

$$\begin{aligned}\mathbb{I}_M(X_{[0,t]} \to Y_{[0,t]}) =& \mathbb{I}(X_0;Y_0)\\ &+ \int_0^t \lim_{h\searrow 0} \frac{1}{h} \mathbb{I}\left(Y_{[s,s+h)};X_{[0,s+h)} \mid Y_{[0,s)}\right) \mathrm{d}s,\end{aligned}$$

under the same continuity conditions. A similar expression was previously proposed by Weissman *et al.* [12], though without accounting for potential discontinuities or including the MI of the initial values, and without a formal proof.

If $\mathbb{I}(X_{[0,t)};Y_{[0,t)}) < \infty$, then mutual and directed information are related through the information conservation law

$$\begin{aligned}\mathbb{I}(X_{[0,t)};Y_{[0,t)}) =& \mathbb{I}(X_{[0,t)} \to Y_{[0,t)}) + \mathbb{I}(Y_{[0,t)} \to X_{[0,t)}) \\ &+ \mathbb{I}\left(X_{[0,t)} \leftrightarrow Y_{[0,t)}\right),\end{aligned} \tag{8}$$

where the continuous-time instantaneous information exchange is defined as

$$\begin{aligned}&\mathbb{I}\left(X_{[0,t)} \leftrightarrow Y_{[0,t)}\right) \\ &:= \inf_{\boldsymbol{t}\in\mathscr{T}(0,t)} \sum_{i=1}^{n(\boldsymbol{t})} \mathbb{I}\left(X_{[t_{i-1},t_i)};Y_{[t_{i-1},t_i)} \mid X_{[0,t_{i-1})},Y_{[0,t_{i-1})}\right),\end{aligned} \tag{9}$$

generalizing the corresponding discrete-time definition [66]. This conservation law has already been addressed in [16] (cf. Eq. (38)) and follows from its discrete counterpart in [13], as we show for Supp. Proposition S1.1.

In contrast, the Massey DI obeys the more asymmetric conservation law [12]

$$\mathbb{I}(X_{[0,t)};Y_{[0,t)}) = \mathbb{I}_M(X_{[0,t)} \to Y_{[0,t)}) + \lim_{\delta\searrow 0} \mathbb{I}_M(Y_{[0,t-\delta)} \to X_{[0,t)}),$$

where the delayed-input Massey DI is defined by

$$\begin{aligned}&\mathbb{I}_M(Y_{[0,t-\delta)} \to X_{[0,t)}) \\ &:= \inf_{\boldsymbol{t}\in\mathscr{T}(\delta,t)} \sum_{i=1}^{n(\boldsymbol{t})} \mathbb{I}\left(Y_{[0,t_i-\delta)};X_{[t_{i-1},t_i)} \mid X_{[0,t_{i-1})}\right)\end{aligned}$$

for any delay $\delta \in (0,t]$.

The conditional DI given a random variable Z, and the causally conditioned DI given an $\mathscr{F}$-adapted stochastic process $Z_{[0,t)} = \{Z(s)\}_{s\in[0,t)}$, are defined respectively as

$$\begin{aligned}&\mathbb{I}\left(X_{[0,t)} \to Y_{[0,t)} \mid Z\right) \\ &:= \sup_{\boldsymbol{t}\in\mathscr{T}(0,t)} \sum_{i=1}^{n(\boldsymbol{t})} \mathbb{I}\left(X_{[0,t_{i-1})};Y_{[t_{i-1},t_i)} \mid Y_{[0,t_{i-1})},Z\right).\end{aligned}$$

and

$$\begin{aligned}&\mathbb{I}\left(X_{[0,t)} \to Y_{[0,t)} \| Z_{[0,t)}\right) \\ &:= \mathbb{I}\left(X_{[0,t)},Z_{[0,t)} \to Y_{[0,t)}\right) - \mathbb{I}\left(Z_{[0,t)} \to Y_{[0,t)}\right)\end{aligned} \tag{10}$$

if $\mathbb{I}\left(X_{[0,t)},Z_{[0,t)} \to Y_{[0,t)}\right) < \infty$ and $\mathbb{I}\left(Z_{[0,t)} \to Y_{[0,t)}\right) < \infty$. Definition (10) is motivated by the chain rules for the discrete-time Newton DI, as detailed in Appendix A. An extremal-based definition is not viable since

$$\sum_{i=2}^{n(\boldsymbol{t})} \mathbb{I}\left(X_{[0,t_{i-1})};Y_{[t_{i-1},t_i)} \mid Y_{[0,t_{i-1})},Z_{[0,t_{i-1})}\right)$$

is not necessarily monotonous under refinement of the partition $\boldsymbol{t} \in \mathscr{T}(0,t)$. However, there exists a sequence of partitions that simultaneously approximates $\mathbb{I}\left(X_{[0,t)},Z_{[0,t)} \to Y_{[0,t)}\right)$ and $\mathbb{I}\left(Z_{[0,t)} \to Y_{[0,t)}\right)$, and hence also $\mathbb{I}\left(X_{[0,t)} \to Y_{[0,t)} \| Z_{[0,t)}\right)$. This can be shown in analogy to the proof of Supp. Lemma S1.2. Similarly, we define

$$\begin{aligned}&\mathbb{I}\left(X_{[0,t)} \to Y_{[0,t)} \|_A Z_{[0,t)}\right) \\ &:= \mathbb{I}\left(X_{[0,t)} \to Y_{[0,t)},Z_{[0,t)}\right) - \mathbb{I}\left(X_{[0,t)} \to Z_{[0,t)} \| Y_{[0,t)}\right)\end{aligned}$$

if the r.h.s. exists. Here, $\|_A$ denotes "anticipatory" causal conditioning. This definition is again motivated in Appendix A, and it can be shown that there exists a sequence of partitions $\boldsymbol{t} \in \mathscr{T}(0,t)$ such that $\mathbb{I}\left(X_{[0,t)} \to Y_{[0,t)} \|_A Z_{[0,t)}\right)$ is approximated by

$$\sum_{i=2}^{n(\boldsymbol{t})} \mathbb{I}\left(X_{[0,t_{i-1})};Y_{[t_{i-1},t_i)} \mid Z_{[0,t_i)},Y_{[0,t_{i-1})}\right).$$

Lemma 1. *It holds that*

$$\begin{aligned}&\mathbb{I}\left(X_{[0,t)} \to Y_{[0,t)} \| Z_{[0,t)}\right) \geq 0 \\ &\mathbb{I}\left(X_{[0,t)} \to Y_{[0,t)} \|_A Z_{[0,t)}\right) \geq 0,\end{aligned}$$

if the quantity on the l.h.s exists.

This lemma follows directly from the nonnegativity of the approximating sequences.

B. Trajectory Information between subnetworks of a CRN

To simplify notation, we define the set of directed reaction $\mathscr{R}' := \{1,\dots,|\mathscr{R}|\} \cup \{-r \in \mathscr{R} \colon k_r^- > 0\}$, which has cardinality $|\mathscr{R}'|$. The directed reaction channels $\mathfrak{R}'_1,\dots,\mathfrak{R}'_{|\mathscr{R}'|}$ have stoichiometric balanced equations

$$\mathfrak{R}'_{r'}\colon \quad \sum_{d=1}^{|\mathscr{S}|} \tilde{a}_d^{(r')} \mathsf{U}_d \quad \xrightarrow{\lambda'_{r'}} \quad \sum_{d=1}^{|\mathscr{S}|} \tilde{b}_d^{(r')} \mathsf{U}_d$$

and change vectors $\nu'_{r'} = \tilde{b}^{(r')} - \tilde{a}^{(r')}$. For clarity, we will omit the prime notation in the remainder of this section.

Let $\mathscr{S}_X, \mathscr{S}_Y \subseteq \mathscr{S}$ be disjoint subsets in the index set of chemical species U. They represent the subsets of species between which we want to characterize information exchange. Hence, we have $X = \{U_d(s) \colon d \in \mathscr{S}_X\}_{s\geq 0}$ and $Y = \{U_d(s) \colon d \in \mathscr{S}_Y\}_{s\geq 0}$ with the respective state spaces $\mathscr{X} \subseteq \mathbb{Z}_{\geq 0}^{|\mathscr{S}_X|}$ and $\mathscr{Y} \subseteq \mathbb{Z}_{\geq 0}^{|\mathscr{S}_Y|}$. Let

$$\mathscr{V} := \{\nu_1,\dots,\nu_{|\mathscr{R}|}\} \setminus \{0\} \subset \mathbb{Z}^{|\mathscr{S}|} \tag{11}$$

be the set of non-zero change vectors and

$$\pi_X : \mathscr{V} \to \mathbb{Z}^{|\mathscr{S}^X|},\ \nu \mapsto \nu_X = (\tilde{b}_d^{(r)} - \tilde{a}_d^{(r)} \colon d \in \mathscr{S}_X) \tag{12}$$

$$\mathscr{V}_X := \{\pi_X(\nu) \mid \nu \in \mathscr{V}\} \setminus \{0\} \tag{13}$$

be the coordinate map [65] to the coordinates of the subnetwork of X and the corresponding set of non-zero change vectors. For Y we analogously define π_Y and $\mathscr{V}_Y$.

Definition 1. *The directed reaction channels $\mathfrak{R}_r$ and $\mathfrak{R}_{\tilde{r}}$ are called X-indistinguishable if*

$$\pi_X(\nu_r) = \pi_X(\nu_{\tilde{r}})$$

and analogously for Y.

As an example, consider the set of species X, Y, A and the X-indistinguishable reactions $\emptyset \to \mathsf{X}$, $\mathsf{X} \to 2\mathsf{X}$, $\mathsf{Y}+\mathsf{A} \to \mathsf{Y}+\mathsf{X}$, which all increase the amount of X by one. Definition 1 allow us to account for the indistinguishability of changes in the marginal subnetwork compared to the entire network. This is important for the definition of the reaction counting processes that generate the trajectories associated with the joint and marginal probability measures $\mathbb{P}^{XY}_{[0,t]}$, $\mathbb{P}^{X}_{[0,t]}$ and $\mathbb{P}^{Y}_{[0,t]}$.

Let $V\colon \mathscr{R} \to \mathscr{V} \cup \{0\}$, $r \mapsto \nu_r$ denote the mapping from reaction indices to their corresponding change vectors. This function is not necessarily injective, as we will see in Sec. IV B. We now distinguish between reaction counters of changes in X or Y. Set

$$\begin{aligned} R^X_s(\nu_X) &:= \sum_{r\in V^{-1}(\pi_X^{-1}(\nu_X))} R_r(s), \\ R^Y_s(\nu_Y) &:= \sum_{r\in V^{-1}(\pi_Y^{-1}(\nu_Y))} R_r(s) \end{aligned} \tag{14}$$

for all $\nu_X \in \mathscr{V}_X$ and $\nu_Y \in \mathscr{V}_Y$. In this notation the time dependence is shifted to the lower index and the functional arguments are the respective marginal change vectors. The corresponding propensity functions of subsystem changes are

$$\begin{aligned} \lambda^X(\nu_X,u) &:= \sum_{r\in V^{-1}(\pi_X^{-1}(\nu_X))} \lambda_r(u), \\ \lambda^Y(\nu_Y,u) &:= \sum_{r\in V^{-1}(\pi_Y^{-1}(\nu_Y))} \lambda_r(u) \end{aligned} \tag{15}$$

With this notation, we obtain for all $s \geq 0$ the process equations

$$\begin{aligned} X(s) =& X(0) + \sum_{\nu_X\in\mathscr{V}_X} \nu_X\, R^X_s(\nu_X), \\ Y(s) =& Y(0) + \sum_{\nu_Y\in\mathscr{V}_Y} \nu_Y\, R^Y_s(\nu_Y). \end{aligned} \tag{16}$$

To compute the MI, we must represent the evolution of X and Y relative to the histories of the marginalized and joint subsystems, specifically $\mathscr{F}^X_s$, $\mathscr{F}^Y_s$, and $\mathscr{F}^{XY}_s$, rather than relying on a representation that depends on the current state of U. This is done by changing the propensity functions of the reaction counters $R^X_s(\nu_X)$, $R^Y_s(\nu_Y)$ to stochastic processes, that depend on the history of the respective marginal system. Abusing notation, we say that $R^X_s(\nu_X)$ has the $\mathscr{F}^{XY}$-intensity process [57] $\lambda^X(\nu_X)$ and the $\mathscr{F}^X$-intensity $\hat{\lambda}^X(\nu_X)$ with

$$\lambda^X_s(\nu_X) = \mathbb{E}\left[\lambda^X(\nu_X, U(s^-)) \mid \mathscr{F}^{XY}_{s-}\right] \tag{17}$$

$$\hat{\lambda}^X_s(\nu_X) = \mathbb{E}\left[\lambda^X(\nu_X, U(s^-)) \mid \mathscr{F}^{X}_{s-}\right], \tag{18}$$

where $U(t-) = \lim_{s\nearrow t} U(s)$ (or $U(t^-)$) denotes the left limit in time. Analogously, $R^Y_s(\nu_Y)$ has the $\mathscr{F}^{XY}$-intensity $\lambda^Y(\nu_Y)$ and the $\mathscr{F}^Y$-intensity $\hat{\lambda}^Y(\nu_Y)$ with

$$\lambda^Y_s(\nu_Y) = \mathbb{E}\left[\lambda^Y(\nu_Y, U(s^-)) \mid \mathscr{F}^{XY}_{s-}\right] \tag{19}$$

$$\hat{\lambda}^Y_s(\nu_Y) = \mathbb{E}\left[\lambda^Y(\nu_Y, U(s^-)) \mid \mathscr{F}^{Y}_{s-}\right]. \tag{20}$$

By definition, all jump processes are non-explosive and all intensities are left-continuous processes and therefore satisfy the predictability property [55] w.r.t. their particular filtration.

We now introduce two notions of bipartiteness.

Definition 2. *Let X and Y be the subnetwork processes as described above. We say that X and Y are*

(a) weakly bipartite if the probability of a simultaneous jump vanishes, i.e.,

$$\mathbb{P}(\exists t > 0\colon \Delta X(t) \neq 0 \text{ and } \Delta Y(t) \neq 0) = 0,$$

where $\Delta X(s) := X(s) - X(s^-)$ for all $s \geq 0$ and analogously for Y.

(b) strongly bipartite if $\mathscr{R}$ does not contain reactions that simultaneously change X and Y, i.e.,

$$\pi_Y\left(\pi_X^{-1}(\nu_X)\right) = \{0\} \quad \text{and} \quad \pi_X\left(\pi_Y^{-1}(\nu_Y)\right) = \{0\}$$

for all $\nu_X \in \mathscr{V}_X$, $\nu_Y \in \mathscr{V}_Y$.

The weakly bipartite case aligns with the usual bipartiteness definition for (non-)Markovian processes [16, cf. p. 7], i.e., fo all $s \geq 0$ and for $h > 0$ it holds

$$\begin{aligned} &\mathbb{P}(X(s+h) = x, Y(s+h) = y \mid \mathscr{F}^{XY}_s) \\ =&\mathbb{P}(X(s+h) = x \mid \mathscr{F}^{XY}_s)\,\mathbb{P}(Y(s+h) = y \mid \mathscr{F}^{XY}_s) + O(h^2). \end{aligned}$$

That weak bipartiteness does not imply strong bipartiteness can be seen at the trivial case with $\mathbb{P}((X,Y)(0) = 0) = 1$. However, strong bipartiteness implies weak bipartiteness as can be seen by the discussion in Supp. Sec. S1.1. The reaction counters (14) comprise a multivariate point process in the strongly bipartite case, which directly implies weak bipariteness.

Both X and Y, individually, and (X,Y) are "fundamental processes" of the kind as presented in §2.13 of [55], even without bipartiteness. Given these definitions, the following theorem provides an expression for the mutual information, similar to the previously stated [35] and excludes a generalization to the relevant non-bipartite cases.

Theorem 1. *If the conditions*

(i) $\mathbb{I}(X(0);Y(0)) < \infty$,

(ii) X and Y are strongly bipartite,

(iii) the intensities meet the requirement

$$\mathbb{E}\left[\sum_{\nu_X\in\mathscr{V}_X} \int_0^t \left|\ln\left(\lambda^X_s(\nu_X)/\hat{\lambda}^X_s(\nu_X)\right)\right| \lambda^X_s(\nu_X)\mathrm{d}s\right] < \infty,$$

and analogously for Y,

are satisfied, then the MI is finite and satisfies

$$\begin{aligned}&\mathbb{I}(X_{[0,t]};Y_{[0,t]})\\ =&\mathbb{I}(X_0;Y_0)+\sum_{\nu_X\in\mathscr{V}_X}\int_0^t\mathbb{E}\left[\phi(\lambda_s^X(\nu_X))-\phi(\hat{\lambda}_s^X(\nu_X))\right]\mathrm{d}s\\ &+\sum_{\nu_Y\in\mathscr{V}_Y}\int_0^t\mathbb{E}\left[\phi(\lambda_s^Y(\nu_Y))-\phi(\hat{\lambda}_s^Y(\nu_Y))\right]\mathrm{d}s,\end{aligned}$$

where $\phi(a):=\ln(a)\,\mathbb{1}_{(0,\infty)}(a)$ with $a\geq 0$ and $\mathbb{1}$ denoting the indicator function of the measurable set in its lower index.

If X and Y are not weakly bipartite and exists $r\in\mathscr{R}$ with $\pi_X(\nu_r),\pi_Y(\nu_r)\neq 0$ such that $\mathbb{E}[R_r(s)]>0$ for all $s>0$, then $\mathbb{P}^{XY}_{[0,t]}$ is not absolutely continuous with respect to $\mathbb{P}^X_{[0,t]}\otimes\mathbb{P}^Y_{[0,t]}$ for all $t>0$.

If absolute continuity is not granted, then the Radon-Nikodym derivative does not exist and it is often said that $\mathbb{I}(X_{[0,t]};Y_{[0,t]})=\infty$ by convention. A proof of the Theorem is provided in Supp. Sec. S1.1.

The MI naturally decomposes into structurally equivalent parts, which we identify as the directed informations.

Theorem 2. *Under the conditions of Theorem 1 (strongly bipartite case), it holds*

$$\mathbb{I}(X_{[0,t]}\to Y_{[0,t]})=\sum_{\nu_Y\in\mathscr{V}_Y}\int_0^t\mathbb{E}\left[\phi(\lambda_s^Y(\nu_Y))-\phi(\hat{\lambda}_s^Y(\nu_Y))\right]\mathrm{d}s, \tag{21}$$

$$\mathbb{I}(Y_{[0,t]}\to X_{[0,t]})=\sum_{\nu_X\in\mathscr{V}_X}\int_0^t\mathbb{E}\left[\phi(\lambda_s^X(\nu_X))-\phi(\hat{\lambda}_s^X(\nu_X))\right]\mathrm{d}s, \tag{22}$$

$$\mathbb{I}(X_{[0,t]}\leftrightarrow Y_{[0,t]})=\mathbb{I}(X(0);Y(0)).$$

and for the Massey DI

$$\begin{aligned}\mathbb{I}_M(X_{[0,t]}\to Y_{[0,t]})=&\mathbb{I}(X(0);Y(0))+\mathbb{I}(X_{[0,t]}\to Y_{[0,t]})\\ \mathbb{I}_M(Y_{[0,t]}\to X_{[0,t]})=&\mathbb{I}(X(0);Y(0))+\mathbb{I}(Y_{[0,t]}\to X_{[0,t]})\end{aligned}$$

A formal proof is given in Supp. Sec. S1.3. The expressions for Newton's DI are well-defined even in the non-bipartite case, but require a different method of proof, which we do not provide here. The instantaneous information exchange and the Massey DI both contain simultaneous jump contributions, which cause them to diverge in this case.

For the causally conditioned DI, a similar expression, in terms of intensity processes, follows directly from (10). Let $\mathscr{S}_Z\subseteq\mathscr{S}$ be a set of species and define $Z:=\{U_d\colon d\in\mathscr{S}_Z\}$. Coordinate maps to the subsystem and projected non-zero change vectors be defined analogously to (12) and (13).

Corollary 1. *Let $(\mathscr{S}_X\cup\mathscr{S}_Z)\cap\mathscr{S}_Y=\emptyset$. Then*

$$\begin{aligned}&\mathbb{I}\left(X_{[0,t)}\to Y_{[0,t)}\|Z_{[0,t)}\right)\\ =&\sum_{\nu_Y\in\mathscr{V}_Y}\int_0^t\mathbb{E}\left[\phi(\check{\lambda}_s^Y(\nu_Y))-\phi(\hat{\lambda}_s^Y(\nu_Y))\right]\mathrm{d}s,\end{aligned}$$

where $\check{\lambda}^Y(\nu_Y)$ is the $\mathscr{F}^{XZY}$-intensity of $R_s^Y(\nu_Y)$ and

$$\begin{aligned}\mathscr{F}_s^{XZY}=&\sigma(U_d(0)\colon d\in\mathscr{S}_X\cup\mathscr{S}_Z\cup\mathscr{S}_Y)\\ &\vee\left(\bigvee_{\nu_{XZY}\in\mathscr{V}_{XZY}}\sigma(R_u^{XZY}(\nu_{XZY}),0\leq u\leq s)\right)\end{aligned}$$

for all $s\geq 0$.

For completeness, we also introduce the point process entropies and conditional point process entropies, which generalize the concepts of Kramer [67] to the considered class of continuous-time processes. Kramer first introduced the concept of causally conditioned probabilities, which Spinney *et al.* [36] generalized to continuous time. These probabilities will be of importance in Sec. VI A, as they define the chemical communication channel. We denote $\mathbb{P}^{Y\|X}_{[0,t]}$ for the probability of $Y_{[0,t]}$, causally conditioned on $X_{[0,t]}$ (cf. Appendix C for a definition). The concept of point process entropy was first introduced by McFadden [56] via the likelihood of a univariate counting process, given in terms of local Janossy densities [68] (cf. p. 213). Similarly, a multivariate point process likelihood can be defined for the reaction counters $(R^Y(\nu_Y),\nu_Y\in\mathscr{V}_Y)$ of a chemical (sub)network via the set of local Janossy densities on $[0,t]\times\mathscr{V}_Y$ [68] (cf. Ch. 7.3). Such likelihoods are defined on the space $\bigcup_{n=0}(\mathbb{R}_{\geq 0}\times\mathscr{V}_Y)^n$, conditionally on the initial value $Y(0)$. The likelihood associated with $\mathbb{P}^{Y|Y(0)}_{[0,t]}$ is

$$\begin{aligned}L^Y_{[0,t]}=&\exp\left(-\sum_{\nu_Y\in\mathscr{V}_Y}\int_0^t\hat{\lambda}_s^Y(\nu_Y)\,\mathrm{d}s\right)\\ &\times\exp\left(\sum_{\nu_Y\in\mathscr{V}_Y}\int_0^t\ln\left(\hat{\lambda}_s^Y(\nu_Y)\right)\mathrm{d}R_s^Y(\nu_Y)\right)\end{aligned}$$

and the likelihood associated with $\mathbb{P}^{Y\|X}_{[0,t]}$ is

$$\begin{aligned}L^{Y\|X}_{[0,t]}=&\exp\left(-\sum_{\nu_Y\in\mathscr{V}_Y}\int_0^t\lambda_s^Y(\nu_Y)\,\mathrm{d}s\right)\\ &\times\exp\left(\sum_{\nu_Y\in\mathscr{V}_Y}\int_0^t\ln\left(\lambda_s^Y(\nu_Y)\right)\mathrm{d}R_s^Y(\nu_Y)\right).\end{aligned} \tag{23}$$

The marginal point process entropy is then defined as

$$h_p(Y_{[0,t]}):=-\mathbb{E}\left[\ln\left(L^Y_{[0,t]}\right)\right]$$

and the causally conditioned point process entropy as

$$h_p(Y_{[0,t]}\|X_{[0,t]}):=-\mathbb{E}\left[\ln\left(L^{Y\|X}_{[0,t]}\right)\right].$$

Consequently, the DI satisfies the familiar relation (cf. [67])

$$\mathbb{I}(X_{[0,t]}\to Y_{[0,t]})=h_p(Y_{[0,t]})-h_p(Y_{[0,t]}\|X_{[0,t]}).$$

Finally, we emphasize the importance of the merging of reaction counters of indistinguishable reactions. Instead of following the above projection procedure, define the index set

$$\mathscr{R}^X:=\{r\mid\pi_X(V(r))\neq 0\}=V^{-1}\left(\pi_X^{-1}(\mathscr{V}_X)\right), \tag{24}$$

i.e., the set of all reactions that modify X. While

$$\mathscr{F}_s^X = \sigma(X(0)) \vee \left(\bigvee_{\nu_X \in \mathscr{V}_X} \sigma(R_u^X(\nu_X), 0 \le u \le s) \right)$$

the newly constructed set of reaction counters only satisfies

$$\mathscr{F}_s^X \subseteq \sigma(X(0)) \vee \left(\bigvee_{r \in \mathscr{R}^X} \sigma(R_r^X(u), 0 \le u \le s) \right).$$

Analogous definitions can be made for Y and (X,Y).

To build intuition, consider a given trajectory $x_{[0,s]}$. From $x_{[0,s]}$, one can uniquely construct the corresponding counting process realizations $r^X_{[0,s]}(\nu_X)$ for all $\nu_X \in \mathscr{V}_X$. Conversely, given these counting process realizations together with the initial condition $x(0)$, the trajectory $x_{[0,s]}$ can be uniquely reconstructed. Similarly, $x_{[0,s]}$ can also be reconstructed from $x(0)$ and the realizations of the reaction counters R_r, $r \in \mathscr{R}^X$, on the interval $[0,s]$. However, the reverse is not generally true: if the system includes X-indistinguishable reactions, the trajectory $x_{[0,s]}$ does not contain enough information to determine which specific parallel reaction r caused each observed stoichiometric change. A more formal discussion of this equivalence of jump process representations has been provided in Appendix A of [39]. In conclusion, not merging X-indistinguishable reactions can lead to misrepresentation of $\mathscr{F}^X$ and thus of the intensity processes $\lambda^X(\nu_X)$, $\hat{\lambda}^X(\nu_X)$. For instance, it must be taken into account in the stochastic filtering equations used to evaluate the various intensity processes [69]. The corresponding equations are provided in Appendix B.

Next, consider the $\mathscr{F}^{XY}$- and $\mathscr{F}^X$-intensity processes defined for all $r \in \mathscr{R}^X$ by

$$\kappa_s^X(r) := \mathbb{E}\left[\lambda_r(U(s^-)) \middle| \mathscr{F}_{s-}^{XY}\right],$$
$$\hat{\kappa}_s^X(r) := \mathbb{E}\left[\lambda_r(U(s^-)) \middle| \mathscr{F}_{s-}^{X}\right].$$

Similarly, for all $r \in \mathscr{R}^Y$, define the $\mathscr{F}^{XY}$- and $\mathscr{F}^Y$-intensity processes as

$$\kappa_s^Y(r) := \mathbb{E}\left[\lambda_r(U(s^-)) \middle| \mathscr{F}_{s-}^{XY}\right],$$
$$\hat{\kappa}_s^Y(r) := \mathbb{E}\left[\lambda_r(U(s^-)) \middle| \mathscr{F}_{s-}^{Y}\right].$$

Note that

$$\lambda_s^X(\nu_X) = \sum_{r:\, \pi_X(\nu_r)=\nu_X} \kappa_s^X(r),$$
$$\hat{\lambda}_s^X(\nu_X) = \sum_{r:\, \pi_X(\nu_r)=\nu_X} \hat{\kappa}_s^X(r),$$

and analogous relations hold for Y. By the log-sum inequality,

$$\lambda_s^X(\nu_X) \ln\left(\frac{\lambda_s^X(\nu_X)}{\hat{\lambda}_s^X(\nu_X)}\right) \le \sum_{r:\, \pi_X(\nu_r)=\nu_X} \kappa_s^X(r) \ln\left(\frac{\kappa_s^X(r)}{\hat{\kappa}_s^X(r)}\right), \tag{25}$$

with equality if the sum consists of a single term. It follows that the directed information satisfies

$$\mathbb{I}\left(Y_{[0,t]} \to X_{[0,t]}\right) \le \sum_{r \in \mathscr{R}^X} \int_0^t \mathbb{E}\left[\phi\left(\kappa_s^X(r)\right) - \phi\left(\hat{\kappa}_s^X(r)\right)\right] \mathrm{d}s,$$

with equality holding when $|\mathscr{V}_X| = |\mathscr{R}^X|$. Thereby the tower property of the conditional expectation was used to recast both sides of (25) in the familar representation, that includes ϕ. An analogous inequality holds for $\mathbb{I}\left(X_{[0,t]} \to Y_{[0,t]}\right)$.

IV. STOCHASTIC THERMODYNAMICS

A. Closed stochastic chemical reaction networks

A thermodynamic system is considered closed if it exchanges energy, but not matter, with its environment. In contrast, a chemical reaction network (CRN) is typically called closed if its core species U do not exchange molecules with the surrounding solvent, even if the solvent itself interacts with an external environment. The species in the solution are thus partitioned into two disjoint sets: the core species U, which participate in the CRN, and the solvent species S, which do not. That is, solvent species S do not appear as reactants or products in any reaction involving the core species U. Under the assumption that the solvent is maintained in thermal and chemical equilibrium at a constant temperature T, it effectively serves as both a heat bath and a chemically equilibrated environment for the CRN.

The solution $\mathsf{U} \cup \mathsf{S}$ corresponds to an isothermal closed system whose classical (non-reactive) microscopic dynamics are governed by a Hamiltonian $\mathfrak{H}$. A mesoscopic description, as used in stochastic thermodynamics, becomes valid under the assumption of a time-scale separation: fast equilibration of microstates relative to the slower dynamics of mesoscopic state changes. This assumption is fundamental and enables a coarse-graining procedure in which each mesoscopic state u is assigned a well-defined free energy [70]. Under constant volume, this potential corresponds to the Helmholtz free energy $f(u)$; under constant pressure, as typical for biological systems, it is replaced by the Gibbs free energy $g(u)$. The pressure dependence is implicitly encoded in $g(u)$ and need not appear explicitly in the mesoscopic description. The derivation of these mesoscopic free energies via coarse-graining from an underlying Hamiltonian model is reviewed in Supp. Sec. S2.1. Alternatively, $g(u)$ can be obtained using equilibrium statistical mechanics of ideal dilute solutions [60], provided that all solute species are accounted for. Since the systems of interest in this work are biological and typically operate under isothermal–isobaric conditions, we use the Gibbs free energy g as the relevant mesoscopic thermodynamic potential throughout.

The zeroth law of thermodynamics for chemical reaction networks dictates that a closed, microscopically reversible CRN that accounts for all solute species will always relax to equilibrium [60]. The equilibrium distribution depends on the initial state $U(0)$ of the system and differs between stoichiometric compatibility classes [71], the sets of all states that are connected via the reaction channels $1, \ldots, |\mathscr{R}|$. In the following we assume that any initial distribution assigns positive mass only to states of a single stoichiometric compatibility

class, since otherwise the limiting distribution

$$\pi(u) := \lim_{t\to\infty} \mathbb{P}(U(t) = u)$$

can become a mixture of the equilibrium distrubutions of individual compatibility classes. Equilibrium statistical mechanics further dictates that the equilibrium distribution of a closed CRN under isothermal-isobaric conditions is (i) a Gibbs-Boltzmann distribution $\pi(u) \propto e^{-\beta g(u)}$, where $\beta = (k_\mathrm{B}T)^{-1}$, and (ii) it satisfies detailed balance

$$\lambda_r^+(u)\,\pi(u) = \lambda_r^-(u+\nu_r)\,\pi(u+\nu_r) \tag{26}$$

for all admissible states u and all microscopically reversible reactions $\mathfrak{R}_r$ [60].

To describe the energetics of each reaction $\mathfrak{R}_r$ we introduce difference operators for functions of mesoscopic states.

Definition 3. *For each reaction channel $\mathfrak{R}_r$ we introduce the forward and backward difference operator Δ_r^+ and Δ_r^-, such that for any function $f\colon \mathbb{Z}_{\geq 0}^{|\mathscr{S}|} \to \mathbb{R}$ we define*

$$\Delta_r^\pm f\colon \{u \in \mathbb{Z}_{\geq 0}^{|\mathscr{S}|} \mid u \pm \nu_r \in \mathbb{Z}_{\geq 0}^{|\mathscr{S}|}\} \to \mathbb{R}$$
$$(\Delta_r^\pm f)(u) = \pm(f(u\pm\nu_r) - f(u)).$$

The definition implies $(\Delta_r^+ f)(u) = -(\Delta_r^- f)(u+\nu_r)$. Starting from state u in forward direction, the change in Gibbs free energy of the state associated with reaction $\mathfrak{R}_r$ is given by

$$\Delta_r^+ g(u) = g(u+\nu_r) - g(u) = -\Delta_r^- g(u+\nu_r).$$

Plugging the Boltzmannn distribution into (26) yields the local detailed balance condition, or thermodynamic consistency relation, for closed CRNs [60]:

$$\ln\left(\frac{\lambda_r^+(u)}{\lambda_r^-(u+\nu_r)}\right) = -\beta\Delta_r^+ g(u) \tag{27}$$

for all admissible states u and all microscopically reversible reactions $\mathfrak{R}_r$. Equation (27) relates kinetic and energetic constants, and therefore it is also satisfied during the transient nonequilibrium and independent of the initial conditions.

However, the consistency relation can be broken if reactions are not elementary, which violates the fundamental coarse-graining assumption that mesostates u define local equilibrium states. Any reaction coordinate of a potentially non-elementary reaction that does not decay instantaneously must be accounted for as a species of the CRN. Deviations from the well-mixed assumption caused by individual elementary reaction events are neglected under the implicit assumption of instantaneous diffusion, which ensures local equilibrium [72] within the finite reaction volume. Consequently, the local equilibrium assumption of stochastic thermodynamics aligns with the kinetic discrete-state Markov process model.

A common modeling oversight, only recently emphasized in [73], is the misidentification of forward and backward reactions that in fact correspond to distinct microscopic channels. In such cases, the thermodynamic consistency relation

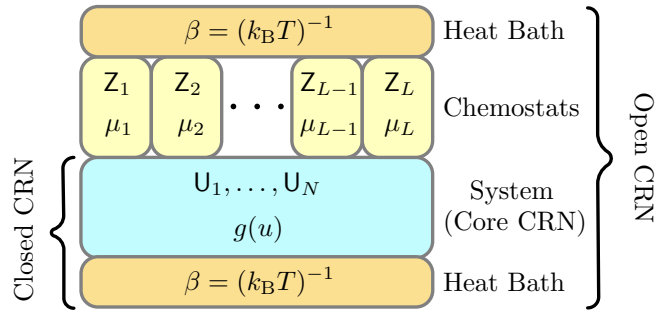


FIG. 2. Visualization of an open CRN, where the core system U is coupled to a number of chemostat species Z_l of chemical potential μ_l via reactions. The solvent acts as a common heat bath with temperature T for the core system and all chemostats. Thus, each horizontal layer in the illustration can exchange energy with its vertically adjacent layers.

is, quite naturally, not valid. This issue is particularly relevant for models of gene expression, where the production and decay of polymers such as mRNA and proteins typically do not represent elementary reactions and are not microscopic reverses of one another.

With a slight abuse of notation, we denote the heat dissipated into the environment due to reaction $\mathfrak{R}_r$ by $\Delta_r^\pm Q(u)$. Strictly speaking, $Q(u)$ is not always a well-defined state function, as heat is a path-dependent quantity. We incorporate the operator into the definition since it satisfies the property $\Delta_r^+ Q(u) = -\Delta_r^- Q(u+\nu_r)$.

As reviewed in Supp. Sec. S2.1, the mesoscopic Gibbs free energy $g(u)$ of the solution can be decomposed into its mesoscopic enthalpy $\mathfrak{h}(u)$ and the internal entropy $h_\mathrm{i}(u)$ of the mesostate u, averaging over the fast equilibrated degrees of freedom (cf. [60], Eq. (58)):

$$g(u) = \mathfrak{h}(u) - k_\mathrm{B}T\,h_\mathrm{i}(u).$$

Following the mesoscopic formulation of the first law of thermodynamics under the local equilibrium assumption (cf. Supp. Sec. S2.1), the heat released to the environment during a state transition can be identified with the decrease in enthalpy:

$$\Delta_r^+ Q(u) = -\Delta_r^+ \mathfrak{h}(u). \tag{28}$$

Substituting this into the free energy change yields

$$\Delta_r^+ g(u) = -\Delta_r^+ Q(u) - k_\mathrm{B}T\,\Delta_r^+ h_\mathrm{i}(u), \tag{29}$$

which shows that the local detailed balance condition (27) links the kinetic parameters of the model with both external and internal entropy changes associated with each reaction.

B. Open chemical reaction networks – Coupling to chemostats

An open CRN allows the reactive exchange of molecules with external reservoirs. Here, we assume that these reservoirs are part of the solvent, which itself needs to be driven

by its external environment to comply with the steady state assumption inherent to all reservoirs.

The core system U be coupled to L chemostats $\mathsf{Z}_1,\dots,\mathsf{Z}_L$, which are part of the solvent species S and immediately equilibrate after participation in reactions of the core system U. The energetic coupling is illustrated in Figure 2. We assume that the copy number of chemostat molecules in the solvent is sufficiently high to be represented by their real-valued concentrations $z=(z_1,\dots,z_L)$, which are held constant over time. In our case, we imagine a biological cell that strives to maintain homeostasis. Others have discussed chemostats that vary in time according to a predetermined deterministic protocol [60, 63, 74]. For simplicity, we first introduce constant chemostats and include the possibility of temporal variations at the end of this section.

The open CRN has stoichiometric balance equations

$$\mathfrak{R}_r: \quad \sum_{l=1}^{L} s_l^{(r)}\mathsf{Z}_l + \sum_{n=1}^{|\mathscr{S}|} a_n^{(r)}\mathsf{U}_n \quad \underset{\lambda_r^-}{\overset{\lambda_r^+}{\rightleftarrows}} \quad \sum_{l=1}^{L} t_l^{(r)}\mathsf{Z}_l + \sum_{n=1}^{|\mathscr{S}|} b_n^{(r)}\mathsf{U}_n,$$

where $\gamma_r = t^{(r)} - s^{(r)}$ is the stoichiometric change vector of the chemostat species associated with $\mathfrak{R}_r$. Immediately after each reaction, the chemostat is assumed to correct the copy number of all chemostat species. Consequently, under the assumption of stochastic mass action kinetics, the propensity functions of the coupled reactions maintain the form (1) with the adjustment that the stochastic rate constants now depend on the chemostat concentrations insofar as they incorporate a deterministic law of mass action

$$\begin{aligned} k_r^+(z) &= \kappa_r^+ \frac{\mathsf{V}\prod_{n=1}^{|\mathscr{S}|} a_n^{(r)}!}{\prod_{n=1}^{|\mathscr{S}|}\mathsf{V}^{a_n^{(r)}}} \prod_{l=1}^{L} z_l^{s_l^{(r)}}, \\ k_r^-(z) &= \kappa_r^- \frac{\mathsf{V}\prod_{n=1}^{|\mathscr{S}|} b_n^{(r)}!}{\prod_{n=1}^{|\mathscr{S}|}\mathsf{V}^{b_n^{(r)}}} \prod_{l=1}^{L} z_l^{t_l^{(r)}}, \end{aligned} \tag{30}$$

where the units of the deterministic rate constants κ_r^+, κ_r^- are adapted accordingly. For any set of Markovian kinetic laws of the form $u \mapsto \lambda_r^{\varepsilon}(u)$ (i.e., not necessarily mass action kinetics), hence z is a fixed parameter of the propensity functions. Therefore, the probability evolution is still described by (3).

To describe the energetics of each reaction $\mathfrak{R}_r$ we denote by μ_l the chemical potential of Z_l, i.e., the energy associated with a single molecule of species Z_l. The chemical work $W_{\mathrm{c}}^{(r)}$ done on the system by the chemostat via reaction $\mathfrak{R}_r$ is then

$$W_{\mathrm{c}}^{(\pm r)} = \mp \sum_l (t_l^{(r)} - s_l^{(r)})\mu_l,$$

where $(-r)$ denotes the backward reaction. Hence the energy dissipated into the environment per reaction is

$$\Delta_r^{\pm} Q(u) = W_{\mathrm{c}}^{(\pm r)} - \Delta_r^{\pm}\mathfrak{h}(u). \tag{31}$$

Intuitively, the dissipated heat is the chemical work done on the system but not "stored" in the systems enthalpy. Further, expression (31) is consistent with (28) when the joint set of species $\mathsf{U}\cup\mathsf{Z}$ is regarded as the core system of a closed CRN instead of regarding Z as a controlled part of S. In open CRNs the thermodynamic consistency relation then becomes (cf. [60], Eq. (76))

$$\ln\left(\frac{\lambda_r^+(u)}{\lambda_r^-(u+\nu_r)}\right) = \beta\,\Delta_r^+ Q(u) + \Delta_r^+ h_{\mathrm{i}}(u) \tag{32}$$

for all admissible states u and all reversible reactions $\mathfrak{R}_r$. The local detailed balance relation (27) together with (29) follows the same expression as (32), where however the identities of the dissipated energies $\Delta_r^+ Q(u)$ per reaction differ. Thus, (32) is more general in the sense that potentially accounts for chemical work through (31) in addition to enthalpy changes. A detailed mesoscopic derivation of (32), including multiple reservoirs, is given in Supp. Sec. S2.2.

In Supp. Sec. S2.3 we review the consistency of (32) in the macroscopic limit for an ideal dilute solution, obeying mass action kinetics.

Lastly, we note that beyond coupling a CRN to a chemostat with constant concentrations, more general time-dependent mechanisms can perturb the system away from equilibrium. Following the terminology of [70], these perturbations fall into two (not mutually exclusive) categories: (i) "Driving", which refers to coupling the system to an external agent that exchanges energy with it during state transitions, as in the case of chemostats; and (ii) "Manipulation", which involves time-dependent control of system parameters, such as changing chemostat concentrations or variations in the solution volume, for example due to cellular growth. Note that this terminology is not used consistently across the literature.

In the presence of a time-varying protocol $\Gamma(t)$, the propensity functions become explicitly time-dependent, $\lambda_r^{\varepsilon}(u,t)$, resulting in a time-inhomogeneous Markov jump process. The structure of the chemical master equation (3) remains valid upon replacing the propensity functions with their time-dependent counterparts. Likewise, thermodynamic quantities such as $g(u,t)$ and $\mu_l(t)$ acquire explicit time dependence, yielding the generalized local detailed balance condition

$$\ln\left(\frac{\lambda_r^+(u,t)}{\lambda_r^-(u+\nu_r,t)}\right) = \beta\,\Delta_r^+ Q(u,t) + \Delta_r^+ h_{\mathrm{i}}(u,t) \tag{33}$$

together with the reaction-wise enthalpy balance relation

$$\Delta_r^{\pm} Q(u,t) = W_{\mathrm{c}}^{(\pm r)}(t) - \Delta_r^{\pm}\mathfrak{h}(u,t). \tag{34}$$

Eq. (33) follows by introducing the instantaneous stationary distribution $\pi(u,t)$, defined as the limiting distribution obtained by hypothetically freezing the protocol at $\Gamma(t)$ and allowing the system to relax. This distribution can then be used to derive (33) via the approach outlined in Supp. Sec. S2.2.

This framework naturally extends to stochastic protocols, where the external control parameters are themselves governed by a stochastic process. In such cases, it is essential that the influence of the core system on the protocol, such as feedback on chemostat concentrations, remains negligible.

C. Thermodynamics of random trajectories

A central appeal of stochastic thermodynamics is that quantities like heat dissipation, work, and entropy production can be defined along individual random trajectories. In the following, we review a priori definitions of these quantities, consistent with macroscopic thermodynamic principles. The presentation builds on Kurtz's process-based formulation of CRNs, which naturally accounts for the instantaneous, finite changes at the mesoscopic scale caused by chemical reactions of U.

Let $0 < \tau_1^{(\pm r)} < \tau_2^{(\pm r)} < \cdots \leq t$ be the random times, when reaction $\mathfrak{R}_r$ happens in forward or backward direction.

Heat is dissipated only during reaction events due to "excess" chemical work provided, even if the system is externally manipulated. The heat dissipated along the random trajectory $U_{[0,t]} := \{U(s)\}_{s\in[0,t]}$ of the open CRN within the time interval $(0,t]$ is then

$$q(t) = \sum_{r=1}^{|\mathscr{R}|} \left\{ \left(\sum_{k:0<\tau_k^{(r)}\leq t} \Delta_r^+ Q(U(\tau_k^{(r)}-), \tau_k^{(r)}-) \right) + \left(\sum_{k:0<\tau_k^{(-r)}\leq t} \Delta_r^- Q(U(\tau_k^{(-r)}-), \tau_k^{(-r)}-) \right) \right\} \tag{35}$$

where $U(t-) = \lim_{s\nearrow t} U(s)$ (or $U(t^-)$) denotes the left limit in time, such that $U(\tau_k^{(r)}-)$ is the state of the CRN just before the k-th occurrence of reaction $\mathfrak{R}_r$ in forward direction [63]. The left limit in the second argument of $\Delta_r^\pm Q$ is irrelevant if Γ is a continuous function, but may be relevant for stochastic manipulation protocols. We restrict the discussion to deterministic protocols Γ that are piecewise-differentiable and hence drop the left limit in the second argument.

We rewrite (35) via stochastic integrals of the reaction counting processes

$$q(t) = \sum_{r=1}^{|\mathscr{R}|} \left\{ \int_0^t \Delta_r^+ Q(U(s^-), s)\, \mathrm{d}R_r^+(s) + \int_0^t \Delta_r^- Q(U(s^-), s)\, \mathrm{d}R_r^-(s) \right\}. \tag{36}$$

Defining the forward, backward, and net probability fluxes of reaction $\mathfrak{R}_r$ as

$$\begin{aligned} J_r^+(u,t) &= \mathbb{P}(U(t)=u)\, \lambda_r^+(u,t), \\ J_r^-(u,t) &= \mathbb{P}(U(t)=u+\nu_r)\, \lambda_r^-(u+\nu_r,t), \text{ and} \\ J_r(u,t) &= J_r^+(u,t) - J_r^-(u,t), \end{aligned} \tag{37}$$

the mean heat dissipation rate (also called thermal power [58]) is

$$\dot{Q}(t) = \sum_u \sum_{r=1}^{|\mathscr{R}|} \Delta_r^+ Q(u)\, J_r(u,t), \tag{38}$$

as rigorously derived in Supp. Sec. S2.5.

In a similar fashion, the chemical work performed on the trajectory $U_{[0,t]}$ is

$$w^{\mathrm{c}}(t) = \sum_{r=1}^{|\mathscr{R}|} \int_0^t W_{\mathrm{c}}^{(r)}(s)(\mathrm{d}R_r^+(s) - \mathrm{d}R_r^-(s)), \tag{39}$$

which gives rise to the mean rate of chemical work

$$\dot{W}_{\mathrm{c}}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[w^{\mathrm{c}}(t)] = \sum_u \sum_{r=1}^{|\mathscr{R}|} W_{\mathrm{c}}^{(r)}(t) J_r(u,t),$$

by following the steps in the proof of (38). $\dot{W}_{\mathrm{c}}(t)$ can also be understood as the amount of energy required to maintain the chemostat concentrations, i.e., the energy expenditure of the external agent. In a manipulated system, the change of the free energy of the system amounts to another work contribution along $U_{[0,t]}$. For simplicity we assume that the internal energy is a continuous function of time, such that [70]

$$w^{\Gamma}(t) = \int_0^t \partial_s \mathfrak{h}(U(s), s)\, \mathrm{d}s.$$

and, hence, the mean rate of manipulated work is

$$\dot{W}_{\Gamma}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[w^{\Gamma}(t)] = \sum_u p(u,t)\, \partial_t \mathfrak{h}(u,t).$$

The entire work, associated with manipulation and driving, is then

$$w(t) = w^{\Gamma}(t) + w^{\mathrm{c}}(t). \tag{40}$$

In this account of work, we have ignored volume changes under isobaric conditions, since these are already accounted for in the enthalpy. The enthalpy change along the trajectory can be expressed via integrals, which account for both continuous evolution between reactions and changes due to reactions:

$$\begin{aligned} &\mathfrak{h}(U(t),t) - \mathfrak{h}(U(0),0) \\ &= w^{\Gamma}(t) + \sum_{r=1}^{|\mathscr{R}|} \left\{ \int_0^t \Delta_r^+ \mathfrak{h}(U(s^-), s)\, \mathrm{d}R_r^+(s) + \int_0^t \Delta_r^- \mathfrak{h}(U(s^-), s)\, \mathrm{d}R_r^-(s) \right\}. \end{aligned} \tag{41}$$

and likewise also the internal entropy change

$$h_{\mathrm{i}}(U(t),t) - h_{\mathrm{i}}(U(0),0) = \beta q^{\mathrm{i}}(t) + \int_0^t \partial_s h_{\mathrm{i}}(U(s),s)\, \mathrm{d}s. \tag{42}$$

where we denote the reactive change of the internal entropy as

$$\beta q^{\mathrm{i}}(t) = \sum_{r=1}^{|\mathscr{R}|} \left\{ \int_0^t \Delta_r^+ h_{\mathrm{i}}(U(s^-), s)\, \mathrm{d}R_r^+(s) + \int_0^t \Delta_r^- h_{\mathrm{i}}(U(s^-), s)\, \mathrm{d}R_r^-(s) \right\}. \tag{43}$$

Plug (34) into (36) and (41) into (40), which together yields the enthalpy balance relation

$$w(t) - q(t) = \mathfrak{h}(U(t),t) - \mathfrak{h}(U(0),0),$$

referred to as the first law of stochastic thermodynamics on the trajectory-level (cf. [60], Eq. (69)). Through the development of stochastic thermodynamics, this law was often reported in a different form, which is sometimes called "first law of stochastic thermodynamics" [70, 75]

$$w^{\mathrm{mes}}(t) - q^{\mathrm{mes}}(t) = g(U(t),t) - g(U(0),0)$$

with $w^{\mathrm{mes}}_{\Gamma}(t) := \int_0^t \partial_s g(U(s),s)\,\mathrm{d}s$, $w^{\mathrm{mes}}(t) := w^{\mathrm{mes}}_{\Gamma}(t) + w^{\mathrm{c}}(t)$ and $q^{\mathrm{mes}}(t) := q(t) + q^{\mathrm{i}}(t)$. It was first noted by Sekimoto [76] that the first-law-like equation was not consistent with the macroscopic first law, since the designated "mesoscopic heat exchange" $q^{\mathrm{mes}}(t)$ contains an entropic part that does not account for heat flow, but a change in the system entropy. Since then, $q(t)$ is distinctly called measured heat [76] or calorimetric heat [70].

We now turn two a trajectory-level account of the total entropy production of the open system. The total entropy production along $U_{[0,t]}$ comprises entropy changes of the system $\mathsf{U}\cup\mathsf{S}$ and the entropy exchange with the environment:

$$h^{\mathrm{tot}}(t) = h^{\mathrm{sys}}(t) + h^{\mathrm{ex}}(t).$$

The system entropy itself contains two parts: (i) changes in the mesoscopic self-entropy

$$h_{\mathsf{U}}(u,s) := -\ln\left(p(U(s),s)\right),$$

whose mean is simply the population (Shannon) entropy of the core system $H[U(s)] = \mathbb{E}[h_{\mathsf{U}}(U(s),s)] = \sum_u -p(U(s),s)\ln\left(p(U(s),s)\right)$ at time $s \geq 0$, and (ii) changes in the internal entropy of the mesoscopic states h_{i}. We denote the change of the population entropy along $U_{[0,t]}$ as

$$h^{\mathsf{U}}(t) := h_{\mathsf{U}}(U(t),t) - h_{\mathsf{U}}(U(0),0) = \ln\left(\frac{p(U(0),0)}{p(U(t),t)}\right),$$

and, in turn, define the system entropy as

$$h^{\mathrm{sys}}(t) = h^{\mathsf{U}}(t) + h_{\mathrm{i}}(U(t),t) - h_{\mathrm{i}}(U(0),0).$$

Not much can be said about the change of internal entropy, but the population entropy is a quantity, which is given in terms of the kinetic model of the open CRN. Following an analogous decomposition as (42) we obtain

$$\begin{aligned} h^{\mathsf{U}}(t) = &-\int_0^t \partial_s \ln(p(U(s),s))\,\mathrm{d}s \\ &+ \sum_{r=1}^{|\mathscr{R}|}\int_0^t \ln\left(\frac{p(U(s^-),s)}{p(U(s^-)+\nu_r,s)}\right)\mathrm{d}R_r^+(s) \\ &+ \sum_{r=1}^{|\mathscr{R}|}\int_0^t \ln\left(\frac{p(U(s^-),s)}{p(U(s^-)-\nu_r,s)}\right)\mathrm{d}R_r^-(s), \end{aligned} \tag{44}$$

where by the use of the CME (3) we may substitute

$$\begin{aligned} &\int_0^t \partial_s \ln(p(U(s),s))\,\mathrm{d}s \\ &= \sum_{r=1}^{|\mathscr{R}|}\int_0^t \Big\{\lambda_r^+(U(s)-\nu_r,s)\frac{p(U(s)-\nu_r,s)}{p(U(s),s)} \\ &\qquad - (\lambda_r^+(U(s),s) + \lambda_r^-(U(s),s)) \\ &\qquad + \lambda_r^-(U(s)+\nu_r,s)\frac{p(U(s)+\nu_r,s)}{p(U(s),s)}\Big\}\,\mathrm{d}s. \end{aligned} \tag{45}$$

We further have $\mathbb{E}\left[\int_0^t \partial_s \ln(p(U(s),s))\,\mathrm{d}s\right] = 0$, which follows by pulling the expectation into the integral (Fubini's theorem) and applying a shift of summation. Using the CME (3) we can express the derivative of the mean population entropy as

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}H[U(t)] &= -\sum_u (\ln(p(u,t)) + 1)\,\partial_t p(u,t) \\ &= \sum_u \sum_{r=1}^{|\mathscr{R}|} \ln\left(\frac{p(u,t)}{p(u+\nu_r,t)}\right) J_r(u,t) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[h^{\mathsf{U}}(t)\right], \end{aligned}$$

where the last equality follows from (44) and the proof method for Eq. (38) in Supp. Sec. S2.5.

According to macroscopic thermodynamic principles [72], the heat dissipated into the environment $q(t)$ must be consistent with the increase of entropy in the evironment h^{ex}, such that

$$h^{\mathrm{ex}}(t) = \beta q(t).$$

Now apply the generalized consistency relation (33) to the expression for the mesoscopic heat exchange (36), which yields

$$\begin{aligned} \beta q^{\mathrm{mes}}(t) &= \beta q(t) + \beta q^{\mathrm{i}}(t) \\ &= \sum_{r=1}^{|\mathscr{R}|}\left\{\int_0^t \ln\left(\frac{\lambda_r^+(U(s^-),s)}{\lambda_r^-(U(s^-)+\nu_r,s)}\right)\mathrm{d}R_r^+(s)\right. \\ &\qquad \left. + \int_0^t \ln\left(\frac{\lambda_r^-(U(s^-),s)}{\lambda_r^+(U(s^-)-\nu_r,s)}\right)\mathrm{d}R_r^-(s)\right\}. \end{aligned} \tag{46}$$

This representation shows that mesoscopic heat exchange captures all entropy flow associated with reaction events that is not accounted for by changes in the statistical distribution over mesostates or by external manipulation. As it depends explicitly on the propensity functions, it encodes the local directional bias of each reaction channel $\mathfrak{R}_r$, relating it to thermodynamic reversibility [77] as further discussed in Supp. Sec. S2.4.

Again following the proof method for Eq. (38), the mean mesoscopic heat exchange rate (in units of $k_{\mathrm{B}}T$) satisfies

$$\beta\dot{Q}^{\mathrm{mes}}(t) = \sum_u \sum_{r=1}^{|\mathscr{R}|} \ln\left(\frac{\lambda_r^+(u,t)}{\lambda_r^-(u+\nu_r,t)}\right) J_r(u,t), \tag{47}$$

and thus depends only on the kinetic model of the open CRN.

In conclusion, the total entropy production satisfies

$$h^{\mathrm{tot}}(t) = \underbrace{h^{\mathsf{U}}(t) + \beta q^{\mathrm{mes}}(t)}_{\text{population part}} + \underbrace{\int_0^t \partial_s h_{\mathrm{i}}(U(s),s)\,\mathrm{d}s}_{\text{microscopic part}}, \tag{48}$$

which decomposes into a part that depends only on the time-dependent population distribution and on the propensity functions, and a term that accounts for internal entropic changes in the microscopic degrees of freedom due to external manipulation. The βq^{i} terms cancel out.

In the absence of external manipulation, e.g., when chemostat concentrations remain constant, the total entropy production of the open CRN depends solely on experimentally accessible quantities: the probability distribution over states and the reaction propensities. This key observation has already been emphasized before [63]. Using equations (44), (46), and (37), we obtain

$$\begin{aligned} h^{\mathrm{tot}}(t) =& h^{\mathsf{U}}(t) + \beta q^{\mathrm{mes}}(t) \\ =& \sum_{r=1}^{|\mathscr{R}|} \int_0^t \Big\{ \lambda_r^+(U(s)-\nu_r) \frac{p(U(s)-\nu_r,s)}{p(U(s),s)} \\ & - (\lambda_r^+(U(s)) + \lambda_r^-(U(s))) \\ & + \lambda_r^-(U(s)+\nu_r) \frac{p(U(s)+\nu_r,s)}{p(U(s),s)} \Big\} \, \mathrm{d}s \quad (49) \\ & + \sum_{r=1}^{|\mathscr{R}|} \Big\{ \int_0^t \ln\left(\frac{J_r^+(U(s^-),s)}{J_r^-(U(s^-),s)} \right) \mathrm{d}R_r^+(s) \\ & + \int_0^t \ln\left(\frac{J_r^-(U(s^-)-\nu_r,s)}{J_r^+(U(s^-)-\nu_r,s)} \right) \mathrm{d}R_r^-(s) \Big\}, \end{aligned}$$

where we used (44), (46) and (37). The (total) entropy production rate (EPR) then follows with $\mathbb{E}\left[\int_0^t \partial_s \ln(p(U(s),s))\,\mathrm{d}s\right] = 0$ and the method of proof of (38):

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} H^{\mathrm{tot}}(t) :=& \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[h^{\mathrm{tot}}(t)\right] \\ =& \sum_u \sum_{r=1}^{|\mathscr{R}|} \ln\left(\frac{J_r^+(u,t)}{J_r^-(u,t)} \right) J_r(u,t), \end{aligned} \quad (50)$$

which can also be related to other mean rates

$$\frac{\mathrm{d}}{\mathrm{d}t} H^{\mathrm{tot}}(t) = \frac{\mathrm{d}}{\mathrm{d}t} H[U(t)] + \beta \dot{Q}^{\mathrm{mes}}(t). \quad (51)$$

This decomposition motivated earlier inconsistent identifications of $\beta \dot{Q}^{\mathrm{mes}}$ with the time-dependent entropy exchange rate [62, 78].

By virtue of $\ln(x)(x-1) \geq 0$ for all $x \geq 0$ (with convention $\ln(0) := \infty$), each term on the r.h.s. of (50) is non-negative and hence

$$\frac{\mathrm{d}}{\mathrm{d}t} H^{\mathrm{tot}}(t) \geq 0$$

shows consistency with the second law of macroscopic thermodynamics.

D. Energy dissipation and entropy production at stationarity

Lastly, we look at the entropy production rate at stationarity, which in the following will be identified with the mean energy dissipation rate in a non-equilibrium steady state (NESS).

At stationarity, we have $\lim_{t\to\infty} H[U(t)] = -\sum_u \pi(u) \ln(\pi(u))$, implying $\lim_{t\to\infty} \frac{\mathrm{d}}{\mathrm{d}t} H[U(t)] = 0$. Consequently, with (51) the entropy production rate at equilibrium or in a NESS satisfies

$$e_p := \lim_{t\to\infty} \frac{\mathrm{d}}{\mathrm{d}t} H^{\mathrm{tot}}(t) = \lim_{t\to\infty} \beta \dot{Q}^{\mathrm{mes}}(t).$$

The EPR at NESS hence quantifies the entropy changes in the whole system, consisting of $\mathsf{U} \cup \mathsf{S}$ and the environment, with all entropic changes that are associated with the dynamics of the CRN and not just changes in the population distribution.

In a second step we note that stationarity of $U(t)$ together with the local equilibrium condition Supp. Eq. (S2.5) implies that also the process of all microscopic degrees of freedom $\Xi(t)$ is stationary (cf. Supp. Sec. S2.1). Then the mean rate of reactive changes of the internal entropy (43) also vanishes:

$$\begin{aligned} & \lim_{t\to\infty} \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[q^{\mathrm{i}}(t)\right] \\ =& \lim_{t\to\infty} \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[q^{\mathrm{i}}(t)\right] + \lim_{t\to\infty} \frac{\mathrm{d}}{\mathrm{d}t} H[U(t)] \\ =& \lim_{t\to\infty} \frac{\mathrm{d}}{\mathrm{d}t} H[\Xi(t)] = 0. \end{aligned}$$

Applying (46) then establishes the identification of the EPR with the energy dissipation rate (in units of $k_{\mathrm{B}}T$) at stationarity:

$$e_p = \lim_{t\to\infty} \dot{Q}(t) \quad (52)$$

with mean heat exchange rate $\dot{Q}(t) := \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[q(t)]$. Eq. (52) is key to the applications discussed in Sec. VI A and VI B.

E. Decompositions of heat dissipation and entropy production

The central equality (52) motivates the definition of the so-called housekeeping heat [63, 79, 80] at trajectory level

$$\begin{aligned} & \beta q^{\mathrm{hk}}(t) \\ :=& \sum_{r=1}^{|\mathscr{R}|} \int_0^t \ln\left(\frac{\lambda_r^+(U(s^-),s)\pi(U(s^-),s)}{\lambda_r^-(U(s^-)+\nu_r,s)\pi(U(s^-)+\nu_r,s)} \right) \mathrm{d}R_r^+(s) \\ & + \sum_{r=1}^{|\mathscr{R}|} \int_0^t \ln\left(\frac{\lambda_r^-(U(s^-),s)\pi(U(s^-),s)}{\lambda_r^+(U(s^-)-\nu_r,s)\pi(U(s^-)-\nu_r,s)} \right) \mathrm{d}R_r^-(s), \end{aligned} \quad (53)$$

where we again generalized to time-dependent protocols Γ, using the instantaneous stationary distribution $\pi(u,s)$. This quantity captures the portion of the total heat dissipation $q(t)$ along a trajectory $U_{[0,t]}$ that is required to maintain the instantaneous NESS of the system at any time $s \in [0,t]$. Notably, the definition remains valid even away from stationarity, allowing one to distinguish the excess heat, defined as $q^{\mathrm{ex}}(t) = q(t) - q^{\mathrm{hk}}(t)$, which quantifies the additional heat dissipated during relaxation toward the steady state.

Similarly, there is the notion of adiabatic entropy production $h^{\mathrm{a}}(t)$ [80], which is also defined by the r.h.s. of (53), and is supposed to describes the entropy produced by an adiabatic protocol Γ^{a} that satisfies the ad-hoc assumption of instantaneous stationarity $p(\cdot,s) = \pi(\cdot,s)$ for all $s \in [0,t]$ and hence $\mathbb{E}[h^{\mathrm{tot}}(t)] = \mathbb{E}[h^{\mathrm{a}}(t)]$. In practice such a protocol is only realizable by a constant Γ at stationarity [70], where even $h^{\mathrm{tot}}(t) = h^{\mathrm{a}}(t)$. The adiabatic entropy production allows for the decomposition of the total entropy production

into an adiabatic term and the non-adiabatic entropy production $h^{\mathrm{na}}(t) := h^{\mathrm{tot}}(t) - h^{\mathrm{a}}(t)$. For a slowly manipulated closed system (i.e., without driving) following an adiabatic protocol, instantaneous detailed balance is approximately satisfied at all times. Using detailed balance as an ad-hoc assumption, yields $h^{\mathrm{a}}(t) = 0$ and hence $h^{\mathrm{tot}}(t) = h^{\mathrm{na}}(t)$. While the intuition for this entropy production decomposition seems to be somewhat unripe, the two quantities exhibit important statistical properties [80].

F. Effective entropy production rate

In practice, the chemostat species of an open CRN are often omitted when formulating the stoichiometric balance equations and the corresponding propensity functions (1). As (30) shows, this simplification does not necessarily affect the definition of the stochastic dynamics of the core process U. However, in the context of stochastic thermodynamics, it is essential to keep track of all species involved in each reaction. In particular, using such "effective" reactions leads to a systematic underestimation of the entropy production rate.

In a properly defined open CRN it typically holds that $(\nu_{r_1}, \gamma_{r_1}) \neq (\nu_{r_2}, \gamma_{r_2})$ and $(\nu_{r_1}, \gamma_{r_1}) \neq (-\nu_{r_2}, -\gamma_{r_2})$ if $r_1 \neq r_2$. However, the same changes in the core species U can be driven with different chemostat species.

Definition 4. *Two reaction channels $\mathfrak{R}_{r_1}$ and $\mathfrak{R}_{r_2}$ are called* parallel*, if $\nu_{r_1} = \nu_{r_2}$ and $\gamma_{r_1} \neq \gamma_{r_2}$, or $\nu_{r_1} = -\nu_{r_2}$ and $\gamma_{r_1} \neq -\gamma_{r_2}$.*

Let $\mathscr{C} := \{1, \ldots, |\mathscr{C}|\}$ with $|\mathscr{C}| \leq |\mathscr{R}|$ be the index set of the parallel reaction classes $\mathscr{C}_\alpha$. All reactions within a class be properly aligned such that $\nu_{r_1} = \nu_{r_2}$ for all $r_1, r_2 \in \mathscr{C}_\alpha$ with $r_1 \neq r_2$. Then we define effective reaction counters $\{\overline{R}^{\pm}_\alpha(t)\}_{t \geq 0}$ with effective propensity functions

$$\overline{\lambda}^{\pm}_\alpha(u) := \sum_{r \in \mathscr{C}_\alpha} \lambda^{\pm}_r(u)$$

and effective probability fluxes

$$\overline{J}^{\pm}_\alpha(u,t) := \sum_{r \in \mathscr{C}_\alpha} J^{\pm}_r(u,t) = \overline{\lambda}^{\pm}_\alpha(u) p(u,t).$$

Applying (50) with these effective fluxes yields the "effective EPR" of the model with effective propensities:

$$\frac{\mathrm{d}}{\mathrm{d}t} H^{\mathrm{tot,eff}}(t) := \sum_u \sum_{\alpha=1}^{|\mathscr{C}|} \ln\left(\frac{\overline{J}^{+}_r(u,t)}{\overline{J}^{-}_r(u,t)}\right) \overline{J}_r(u,t),$$

The effective EPR is a lower bound on the true EPR, i.e.,

$$0 \leq \frac{\mathrm{d}}{\mathrm{d}t} H^{\mathrm{tot,eff}}(t) \leq \frac{\mathrm{d}}{\mathrm{d}t} H^{\mathrm{tot}}(t),$$

as has been shown in [70] (cf. Eq. (3.81)), using the multi-reservoir description provided in the following section. We exemplify this inequality Sec. VI C with a two-state Markov model exhibiting two reaction channels. While in this case, $\frac{\mathrm{d}}{\mathrm{d}t} H^{\mathrm{tot,eff}}(t) = 0$ for any choice of rate constants $\frac{\mathrm{d}}{\mathrm{d}t} H^{\mathrm{tot}}(t)$ can be positive.

G. Entropy Production Rate of driven conversion networks and general CTMCs with multiple reservoirs

Much of the literature on entropy production rate employs a formulation based on the infinitesimal generator (transition rate matrix) Λ of a CTMC, rather than a reaction network perspective [62, 70, 81]. A bridge between these two pictures can be constructed by introducing an index set $\mathscr{J} := \{1, \ldots, |\mathscr{U}|\}$, where each mesoscopic state $u \in \mathscr{U}$ is uniquely assigned an index $j \in \mathscr{J}$ (with $|\mathscr{U}| = \infty$ allowed). This indexing defines a one-hot encoding of the state space, $\mathscr{U}' = \{e_j \colon j \in \mathscr{J}\}$, where e_j denotes the unit vector with a single non-zero entry at position j.

Viewed as a population representation, this encoding interprets each state e_j as indicating the presence of a single effective species U'_j, with all other species absent. This reformulation recasts the original reaction network into a conversion network: a collection of microscopically reversible transitions between discrete one-hot states. Each conversion $i \to j$ (with change vector $\nu_{(i,j)} = e_j - e_i$) may be coupled to one or more reservoirs r, corresponding to the reaction channels $\mathfrak{R}_r$, as discussed in Supp. Sec. S2.2.

For each reservoir r, we denote the set of reversible transitions it mediates as

$$\mathscr{R}^{(r)} := \{(i,j) \in \mathscr{J}^2 \colon i < j,\ \lambda^{(+r)}_{(i,j)}(e_i) > 0\}.$$

with stoichiometric balance equations

$$\mathfrak{R}^{(r)}_{(i,j)} : \quad \mathsf{U}_i \underset{\lambda^{(-r)}_{(i,j)}}{\overset{\lambda^{(+r)}_{(i,j)}}{\rightleftarrows}} \mathsf{U}_j, \quad \text{with } (i,j) \in \mathscr{R}^{(r)}$$

and reservoir-specific propensity functions $\lambda^{(\pm r)}_{(i,j)}(u')$.

The effective Markov generator Λ of such a network, coupled to multiple reservoirs r is then given as the sum of reservoir-wise generators $\Lambda^{(r)}$, such that the transition rates for $i \to j$ (and also diagonal elements) satisfy [82]

$$\Lambda_{ij} = \sum_{r=1}^{|\mathscr{R}|} \Lambda^{(r)}_{ij}.$$

For each $(i,j) \in \mathscr{R}^{(r)}$ the reservoir-wise propensity functions and generators are related as

$$\begin{aligned} \lambda^{(+r)}_{(i,j)}(u') &= \Lambda^{(r)}_{ij}\, \mathbb{1}\{u' = e_i\}, \\ \lambda^{(-r)}_{(i,j)}(u') &= \Lambda^{(r)}_{ji}\, \mathbb{1}\{u' = e_j\}. \end{aligned} \tag{54}$$

The reservoir-wise generators then satisfy the local detailed balance relation for $(i,j) \in \mathscr{R}^{(r)}$ (cf. Supp. Sec. S2.2 and [82], Eq. (4))

$$\ln\left(\frac{\Lambda^{(r)}_{ij}}{\Lambda^{(r)}_{ji}}\right) = \ln\left(\frac{\lambda^{(+r)}_{(i,j)}(e_i)}{\lambda^{(-r)}_{(i,j)}(e_j)}\right) = \beta\, \Delta^{+}_r Q(e_i) + \Delta^{+}_r h_{\mathrm{i}}(e_i).$$

By (50) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}H^{\mathrm{tot}}(t) = \sum_{i\in\mathscr{J}}\sum_{r=1}^{|\mathscr{R}|}\sum_{(k,j)\in\mathscr{R}^{(r)}} \ln\left(\frac{p(e_i,t)\,\lambda^{(+r)}_{(k,j)}(e_k)}{p(e_i+e_j-e_k,t)\,\lambda^{(-r)}_{(k,j)}(e_j)}\right) \times (\lambda^{(+r)}_{(k,j)}(e_k)p(e_i,t) - \lambda^{(-r)}_{(k,j)}(e_j)p(e_i+e_j-e_k,t))$$

and hence substituting (54) yields the multi-reservoir EPR of a consistently modeled CTMC [82]

$$\frac{\mathrm{d}}{\mathrm{d}t}H^{\mathrm{tot}}(t) = \sum_{i\in\mathscr{J}}\sum_{r=1}^{|\mathscr{R}|}\sum_{j>i} \ln\left(\frac{p(e_i,t)\,\Lambda^{(r)}_{ij}}{p(e_j,t)\,\Lambda^{(r)}_{ji}}\right) \times (\Lambda^{(r)}_{ij}\,p(e_i,t) - \Lambda^{(r)}_{ji}\,p(e_j,t)).$$

Importantly, neglecting the contribution of individual reservoirs can result in an underestimation of the EPR (cf. Ref. 70, Eq. (3.81)).

V. SPECIES-REACTION GRAPHS AND LOCAL INDEPENDENCE GRAPHS FOR CRNS

A. Species reaction graphs

We define a species-reaction graph (SR-graph) slightly different from its original notion [83]. This is necessary to faithfully represent local independence. A reversible SR-graph is a biparite graph $(\mathscr{S},\mathscr{R},\mathscr{E})$ with edge set $\mathscr{E}$ and two vertex types: (i) the species nodes $\mathscr{S}$ for all U_d, and (ii) the reversible reaction channel nodes $\mathscr{R}$ for all reaction channels $\mathfrak{R}_r$. Here, the SR-graph attribute "reversible" means that we do not resolve forward and backward direction of a single reaction channel $\mathfrak{R}_r$ if the reaction channel is reversible. However, the reversible SR-graph may actually comprise irreversible reaction channels. In contrast, we refer to an SR-graph as irreversible, if all reaction channels are considered irreversible.

In a biparite graph, only vertices of different type can be connected by an edge $e \in \mathscr{E}$, i.e.,

$$\mathscr{E} \subseteq (\mathscr{S}\times\mathscr{R}) \cup (\mathscr{R}\times\mathscr{S}),$$

where we assume distinct enumeration $\mathscr{S}\cap\mathscr{R} = \emptyset$ for notational simplicity. A directed edge $(d,r) \in \mathscr{S}\times\mathscr{R}$ is drawn from U_d to $\mathfrak{R}_r$ if either λ_r^+ or λ_r^- explicitly depends on u_d. Such edges encapsulate local (in)dependencies of the reaction counters. A directed edge $(r,d) \in \mathscr{R}\times\mathscr{S}$ is drawn from $\mathfrak{R}_r$ to U_d if ν_r is non-zero in the coordinate of U_d. These edges encapsulate the time-point wise algebraic equation

$$U_d(s) = U_d(0) + \sum_{r\in\mathscr{R}^d} \pi_d(\nu_r)\,(R_r^+(s) - R_r^-(s)),$$

where π_d is the coordinate map to $d \in \mathscr{S}$ and $\mathscr{R}^d := \{r \in \mathscr{R} \mid \pi_d(\nu_r) \neq 0\}$. If directed edges in both directions exist, they can be replaced by a single bidirected edge "$\longleftrightarrow$" in the visualization of the graph.

An edge-simple walk (often called a 'trail' in classical graph theory [84]) of length $N \in \mathbb{N}$, in the SR-graph is sequence of directed edges $\vec{e} = (e_1,\ldots,e_N)$ which joins a sequence of vertices $(v_0,\ldots,v_N)$, such that all directed edges $e_k = (v_{k-1},v_k)$ are pairwise distinct. A edge-simple walk $\vec{e}$ is said to be from v_0 to v_N. We say that $\vec{e}$ passes through some vertex v, if exists $i \in \{1,\ldots,N-1\}$ such that $v = v_i$. For edge-simple walks between sets instead of single vertices we use the following definition.

Definition 5. *Let $\mathscr{R}' \subseteq \mathscr{R}$ and consider the disjoint species sets $\mathscr{S}_X,\mathscr{S}_Y \subset \mathscr{S}$. An edge-simple walk of length $N \in \mathbb{N}$ is said to be*

(a) from $\mathscr{S}_X$ to $\mathscr{S}_Y$ if $v_0 \in \mathscr{S}_X$, $v_N \in \mathscr{S}_Y$ and $v_1,\ldots,v_{N-1} \notin \mathscr{S}_X \cup \mathscr{S}_Y$.

(b) from $\mathscr{R}'$ to $\mathscr{S}_Y$ if $v_0 \in \mathscr{R}'$, $v_N \in \mathscr{S}_Y$ and $v_1,\ldots,v_{N-1} \notin \mathscr{R}' \cup \mathscr{S}_Y$.

By this definition, a 'trail' from $\mathscr{S}_X$ to $\mathscr{S}_Y$ is always a 'trail' on the edge-deleted subgraph $(\mathscr{S},\mathscr{R},\mathscr{E}_{X\to Y})$ with

$$\mathscr{E}_{X\to Y} := \mathscr{E}\setminus(\{(d,r)\in\mathscr{S}\times\mathscr{R} \mid d\in\mathscr{S}_Y\} \cup\{(r,d)\in\mathscr{R}\times\mathscr{S} \mid d\in\mathscr{S}_X\}),$$

which does neither contain edges into $\mathscr{S}_X$ nor edges that originate from $\mathscr{S}_Y$. Similarly, a 'trail' from $\mathscr{R}'$ to $\mathscr{S}_Y$ is a 'trail' on the edge-deleted subgraph $(\mathscr{S},\mathscr{R},\mathscr{E}_{\mathscr{R}'\to Y})$ with

$$\mathscr{E}_{\mathscr{R}'\to Y} := \mathscr{E}\setminus(\{(d,r)\in\mathscr{S}\times\mathscr{R} \mid d\in\mathscr{S}_Y\} \cup\{(d,r)\in\mathscr{S}\times\mathscr{R} \mid r\in\mathscr{R}'\}).$$

B. Local independence graphs for multivariate counting processes

We briefly introduce the concepts of local independence graphs and dynamic Markov properties for multivariate counting processes, as developed by Didelez [85].

Temporarily assume that $\mathscr{R}$ is a set of irreversible reaction indices. We introduce bold notation for the multivariate counting process $\boldsymbol{R} := (R_r\colon r\in\mathscr{R})$. For any $A \subseteq \mathscr{R}$ denote the subprocess $\boldsymbol{R}^A := (R_r\colon r\in A)$ and its internal filtration $\mathscr{F}_s^A := \bigvee_{r\in A}\sigma(R_r(\tau)\colon \tau\le s)$. Further let $s\mapsto\kappa_s(r)$ denote the $(\mathscr{F},\mathbb{P})$-intensity process of R_r. Thereby $\kappa_s(r)$ may depend on the entire past $\mathscr{F}_{s-}$, i.e., $\boldsymbol{R}$ may be a non-Markovian process. The process $(\boldsymbol{R},\mathbb{P})$ is always associated with a probability measure, which determines its intensity processes. Further, it is also associated with a finite observation interval $[0,t]$, $t>0$. In the following, we simply write $\boldsymbol{R}$ unless required.

Definition 6 ([85])**.** *Let $A,B,C \subseteq \mathscr{R}$ be pairwise disjoint. We say that a subprocess $\boldsymbol{R}^B$ is locally independent of $\boldsymbol{R}^A$ given $\boldsymbol{R}^C$ over $[0,t]$ if for all $r\in B$ the $\mathscr{F}^{A\cup B\cup C}$-intensities $\bar{\lambda}_s(r) := \mathbb{E}[\kappa_s(r) \mid \mathscr{F}_{s-}^{A\cup B\cup C}]$ are $\mathscr{F}^{B\cup C}$-intensities. This is denoted by $A \not\to B \mid C$. Otherwise, $\boldsymbol{R}^B$ is locally dependent on $\boldsymbol{R}^A$ given $\boldsymbol{R}^C$, i.e., $A\to B\mid C$. If $C=\emptyset$, then $\boldsymbol{R}^B$ is marginally locally (in)dependent of $\boldsymbol{R}^A$.*

Although this is not necessary in general, the above definition assumes that local independence is "reflexive", i.e., each subprocess depends at least on its own past, which is the case for most practical applications. $A,B,C \subseteq \mathscr{R}$, not necessarily disjoint,

Now let $\mathscr{G} := (\mathscr{R},\mathscr{A})$ be a directed graph (possibly containing cycles) with directed edge set

$$\mathscr{A} \subseteq \{(r',r) \in \mathscr{R}\times\mathscr{R} \mid r' \neq r\}.$$

Definition 7 ([85]). *$\mathscr{G}$ is called local independence graph of $\boldsymbol{R}$ if $\boldsymbol{R}$ satisfies the pairwise dynamic Markov property with respect to $\mathscr{G}$, i.e., for all $r',r \in \mathscr{R}$, $r' \neq r$*

$$(r',r) \notin \mathscr{A} \quad \Longrightarrow \quad \{r'\} \not\to \{r\} \mid \mathscr{R}\setminus\{r',r\}.$$

In the following, we need a few more notions of graph theory [86] and will mainly adhere to the conventions of Didelez [87] (cf. Sec. 1.1). To keep these notions abstract, we temporarily denote the finite vertex set as $\mathscr{V}$ and the edge set as $\mathscr{E}$, where the edge set may contain both directed and undirected edges (denoted as sets), i.e.,

$$\mathscr{E} \subseteq \left\{(v',v) \in \mathscr{V}^2 \mid v' \neq v\right\} \cup \left\{\{v',v\} \subseteq \mathscr{V} \mid v' \neq v\right\}.$$

We consider only reciprocal graphs, where $\{v',v\} \in \mathscr{E}$ implies $(v',v),(v,v') \notin \mathscr{E}$. Let $\mathscr{G} := (\mathscr{V},\mathscr{E})$ denote the graph. For $A \subset \mathscr{V}$ we call $\mathscr{G}_A := (A,\mathscr{E}|_A)$ the induced subgraph, where $\mathscr{E}|_A := \mathscr{E} \cap ((A\times A) \cup \{\{v',v\} \subseteq A\})$. For $B \subset \mathscr{V}$ we denote with $\mathscr{G}^B := (\mathscr{V},\mathscr{E}^B)$ the edge-deleted subgraph with

$$\mathscr{E}^B := \mathscr{E} \setminus \{(v,v') \in \mathscr{E} \mid v \in B\}.$$

A tuple $\vec{v} = (v_0,\ldots,v_N)$ of distinct vertices $v_k \in \mathscr{V}$ (except possibly $v_0 = v_N$) is called a path of length $N \in \mathbb{N}$ from v_0 to v_N if $\{v_{k-1},v_k\} \in \mathscr{E}$ or $(v_{k-1},v_k) \in \mathscr{E}$ for all $k \in \{1,\ldots,N\}$. A path $\vec{v}$ is called semidirected if exists k with $(v_{k-1},v_k) \in \mathscr{E}$ and directed the latter holds for all $k \in \{1,\ldots,N\}$. Note that Didelez [87] also introduced a generalized notion of trails for the given class of graphs [88]. A chain graph is a (reciprocal) graph that does not contain any (semi)directed cycles.

Let $a,b \in \mathscr{V}$, $a \neq b$ and $A,B,C \subset \mathscr{V}$. A set C is called an (a,b)-separator if all paths from a to b intersect C. We say that C separates A from B if C is an (a,b)-separator for all $a \in A$ and $b \in B$. a is a parent of b ($a \in \mathrm{pa}(b)$) if $(a,b) \in \mathscr{E}$ and $\mathrm{pa}(A) := \cup_{v\in A}\mathrm{pa}(v) \setminus A$. b is a child of a ($b \in \mathrm{ch}(a)$) if $(a,b) \in \mathscr{E}$ and $\mathrm{ch}(A) := \cup_{v\in A}\mathrm{ch}(v) \setminus A$. a is a neighbor of/adjacent to b ($a \in \mathrm{nb}(b)$) if $\{a,b\} \in \mathscr{E}$ and $\mathrm{nb}(A) = \cup_{v\in A}\mathrm{nb}(v) \setminus A$. a and b are non-adjacent if $\{(a,b),(b,a),\{a,b\}\} \cap \mathscr{E} = \emptyset$. The union of parents and neighbors is called the boundary $\mathrm{bd}(A) = \mathrm{pa}(A) \cup \mathrm{nb}(A)$ and $\mathrm{cl}(A) := \mathrm{bd}(A) \cup A$ is the closure of A. If there is a path from a to b, then a is an ancestor of b ($a \in \mathrm{an}(b)$) and b is a descendant of a ($b \in \mathrm{de}(a)$). For sets $\mathrm{an}(A) := \cup_{v\in A}\mathrm{an}(v) \setminus A$, $\mathrm{de}(A) := \cup_{v\in A}\mathrm{de}(v) \setminus A$, and additionally the non-descendants $\mathrm{nd}(A) := \mathscr{V} \setminus (\mathrm{de}(A) \cup A)$. A set A is called ancestral if $\mathrm{bd}(A) = \emptyset$. The smallest ancestral set, containing A is denoted as $\mathrm{An}(A)$ and satisfies $\mathrm{An}(A) = \mathrm{an}(A) \cup A$.

The moralized graph $\mathscr{G}^{\mathrm{m}} := (\mathscr{V},\mathscr{E}^{\mathrm{m}})$ is an undirected graph, such that $\{v',v\} \in \mathscr{E}^{\mathrm{m}}$ if $(v',v) \in \mathscr{E}$ or $(v,v') \in \mathscr{E}$, or if exists $\tilde{v} \in \mathscr{V}$ such that $\{v',v\} \subseteq \mathrm{pa}_{\mathscr{G}}(\Upsilon(\tilde{v}))$, where $\Upsilon(\tilde{v}) := \{v \in \mathscr{V} \mid \exists \text{ undirected path form } \tilde{v} \text{ to } v\}$ and $\mathrm{pa}_{\mathscr{G}}$ means "parent with respect to graph $\mathscr{G}$".

With these notions, and switching back to the graph $\mathscr{G} = (\mathscr{R},\mathscr{A})$, an asymmetric separation criterion can be defined.

Definition 8 ([85]). *For pairwise disjoint $A,B,C \subset \mathscr{R}$ we say that C δ-separates A from B in $\mathscr{G}$ if C separates A from B in the undirected graph $(\mathscr{G}^B_{\mathrm{An}(A\cup B\cup C)})^{\mathrm{m}}$.*

For $A,B,C \subset \mathscr{R}$ not pairwise disjoint we say that C δ-separates A from B if $C \setminus B$ δ-separates $A \setminus (B \cup C)$ from B. Further, by convention $\emptyset$ is always δ-separated from B and $\emptyset$ δ-separates A from B if A and B are unconnected in $(\mathscr{G}^B_{\mathrm{An}(A\cup B)})^{\mathrm{m}}$.

Note that the moralization criterion in the above definition applies only in the disjoint case. With the given graph theory notions two further dynamic Markov properties can be defined.

Definition 9 ([85]). *$\boldsymbol{R}$ is said to satisfy the local dynamic Markov property with respect to $\mathscr{G}$ if for all $r \in \mathscr{R}$*

$$\mathscr{R} \setminus \mathrm{cl}(r) \not\to \{r\} \mid \mathrm{pa}(r),$$

i.e., the $\mathscr{F}^{\mathscr{R}}$-intensity $\bar{\lambda}(r)$ is a $\mathscr{F}^{\mathrm{cl}(r)}$-intensity.

$\boldsymbol{R}$ is said to satisfy the global dynamic Markov property with respect to $\mathscr{G}$ if for all pairwise disjoint $A,B,C \subseteq \mathscr{R}$

$$C \ \delta\text{-separates } A \text{ from } B \quad \Longrightarrow \quad A \not\to B \mid C.$$

Note that the generalization of the global dynamic Markov property to the non-disjoint case is straightforward. Didelez proved two theorems – a criterion for (symmetric) conditional independence and a criterion for the equivalence of all dynamic Markov properties – which we restate in the following.

Proposition 1 ([85, 89]). *Let $A,B,C \subseteq \mathscr{R}$ be pairwise disjoint.*

(a) Let $\mathscr{G}$ be the local independence graph of $\boldsymbol{R}$. If C separates A from B on $(\mathscr{G}_{\mathrm{An}(A\cup B\cup C)})^{\mathrm{m}}$, then

$$\mathscr{F}^A_s \perp\!\!\!\perp \mathscr{F}^B_s \mid \mathscr{F}^C_s \quad \text{for all } s \in [0,t].$$

(b) If $\boldsymbol{R}$ satisfies conditional measurable separability [90], i.e., for all $D,E \subseteq \mathscr{R}$ and all $s \in [0,t]$

$$\overline{\mathscr{F}^D_s} \cap \overline{\mathscr{F}^E_s} = \overline{\mathscr{F}^{D\cap E}_s}$$

with convention $\mathscr{F}^{\emptyset} := \{\emptyset,\Omega\}$, then (under mild regularity conditions) all dynamic Markov properties are equivalent.

Hence, given that $\mathscr{G}$ is a local independence graph of $\boldsymbol{R}$, Proposition 1 allows the use of graphical separation criteria to obtain "global" local independence relations. Further, (a) has several implications: (i) $A \not\to B \mid C$, (ii) $B \not\to A \mid C$, and (iii) for each $r \in C$ the $\mathscr{F}^{A\cup B\cup C}$-intensity $\bar{\lambda}(r)$ is either a $\mathscr{F}^{A\cup C}$- or a $\mathscr{F}^{B\cup C}$-intensity. The application of local independence graphs to the reaction counters of CRNs requires the extension with a non-dynamic node for each species in $\mathscr{S}$, which

jointly represent the initial state $U(0)$. In general, local independence graphs can be extended with time-fixed covariates, whose information must then be included in the filtration at the start $\mathscr{F}_0$. The subgraph induced by non-dynamic nodes must be either a directed acyclic graph or a chain graph [85]. Edges from dynamic to non-dynamic nodes are not allowed. This construction ensures that the induced subgraph of non-dynamic nodes can faithfully represent conditional independence relations and (non-dynamic) Markov properties. In the following, we formalize

Let $\mathscr{I}$ be the set of $\mathscr{F}_0$-measurable baseline covariates θ_i, $i \in \mathscr{I}$, such that $\mathscr{I} \cap \mathscr{R} = \emptyset$. These covariates can, for example, be (random) initial conditions or time-dependent parametrizations of intensity processes, which are fully determined at time zero. Again, we introduce a multivariate random variable on a subset $I \subseteq \mathscr{I}$ with notation $\boldsymbol{\theta}^I := \{\theta_i \colon i \in I\}$. As before let $\mathscr{R}$ be the set of counting processes. We define the extended graph $\mathscr{G} := (\mathscr{I} \cup \mathscr{R}; \mathscr{B}, \mathscr{A})$ with the set of undirected edges between baseline covariates

$$\mathscr{B} \subseteq \{\{i, i'\} \subseteq \mathscr{I} \mid i \neq i'\}$$

and directed edges

$$\mathscr{A} \subseteq \{(j, j') \in (\mathscr{I} \cup \mathscr{R})^2 \colon j \neq j', \{j, j'\} \notin \mathscr{B}\}$$

with $\mathscr{A} \cap (\mathscr{R} \times \mathscr{I}) = \emptyset$. The induced subgraph $\mathscr{G}_{\mathscr{I}}$ be a chain graph. The graph $\mathscr{G}$ is equivalent to a dynamic graph [87]. For any set $G \subseteq \mathscr{I} \cup \mathscr{R}$ we assign the subfiltration $\mathscr{F}^G$ such that for all $s \in [0, t]$

$$\mathscr{F}_s^G := \left(\bigvee_{r \in G \cap \mathscr{R}} \sigma(R_r(\tau) \colon \tau \in [0, s]) \right) \vee \left(\bigvee_{i \in G \cap \mathscr{I}} \sigma(\theta_i) \right).$$

On the extended local independence graph $\mathscr{G}$ we need to generalize the notion of local independence and Markov properties. In particular, dynamic properties can only be defined for subsets of $\mathscr{R}$. Definition 6 (local independence) and 8 (δ-separations) and all dynamic Markov properties can be readily generalized for $A, C \subseteq \mathscr{I} \cup \mathscr{R}$ and $B \subseteq \mathscr{R}$ pairwise disjoint, using the extended subfiltrations.

Definition 10. *$\boldsymbol{R}$ is said to satisfy the local dynamic Markov property with respect to $\mathscr{G}$ if for all $r \in \mathscr{R}$*

$$(\mathscr{I} \cup \mathscr{R}) \setminus \mathrm{cl}(r) \not\to \{r\} \mid \mathrm{pa}(r),$$

i.e., the $\mathscr{F}^{\mathscr{I} \cup \mathscr{R}}$-intensity $\bar{\lambda}(r)$ is a $\mathscr{F}^{\mathrm{cl}(r)}$-intensity.

$\boldsymbol{R}$ is said to satisfy the global dynamic Markov property with respect to $\mathscr{G}$ if for all $A, C \subseteq \mathscr{I} \cup \mathscr{R}$ and $B \subseteq \mathscr{R}$, pairwise disjoint,

$$C \ \delta\text{-separates } A \text{ from } B \quad \Longrightarrow \quad A \not\to B \mid C.$$

As one of our key results, we show that local independence is equivalent to a vanishing causally conditioned directed information. The proof is provided in Supp. Sec. S1.4.

Theorem 3. *Let $A, C \subseteq \mathscr{I} \cup \mathscr{R}$ and $B \subseteq \mathscr{R}$ be pairwise disjoint. Then $A \not\to B \mid C$ if and only if*

$$\mathbb{I}\left(\boldsymbol{\theta}^{A \cap \mathscr{I}}, \boldsymbol{R}_{[0,t]}^{A \cap \mathscr{R}} \to \boldsymbol{R}_{[0,t]}^{B} \| \boldsymbol{\theta}^{C \cap \mathscr{I}}, \boldsymbol{R}_{[0,t]}^{C \cap \mathscr{R}} \right) = 0.$$

We conclude by providing an extension of Definition 7, which is sufficient for our purposes.

Definition 11. *$\mathscr{G}$ is called an (extended) local independence graph of $(\boldsymbol{\theta}, \boldsymbol{R}, \mathbb{P})$ if the following is satisfied:*

(i) $\mathbb{P} \circ \boldsymbol{\theta}^{-1}$ satisfies the pairwise chain Markov property [86], relative to $\mathscr{G}_{\mathscr{I}}$, i.e., for any pair i, i' of non-adjacent vertices with $i' \in \mathrm{nd}(i)$ satisfies $i \perp\!\!\!\perp i' \mid \mathrm{nd}(i) \setminus \{i'\}$;

(ii) For all $j \in \mathscr{I} \cup \mathscr{R}$, $r \in \mathscr{R}$, $j \neq r$ it holds that

$$(j, r) \notin \mathscr{A} \quad \Longrightarrow \quad \{j\} \not\to \{r\} \mid (\mathscr{I} \cup \mathscr{R}) \setminus \{j, r\}.$$

Given that all deterministic parameters are fixed (and hence do not appear as nodes in the local independence graph), we assume conditional measurable separability for every relevant case in this paper. Hence, we always have equivalence of dynamic Markov properties for the generalized case with random baseline covariates.

C. Local independence graphs for Markovian CRNs

We have thus far claimed that the SR-graph of the Markovian CRN U faithfully represents local (in)dependencies. Now we derive the extended local independence graph of $(U(0), \boldsymbol{R})$ from the "reversible" SR-graph. Here, any process node $r \in \mathscr{R}$ represents the collection $\boldsymbol{R}^{\{r\}} = (R_r^+, R_r^-)$ if the reaction channel $\mathfrak{R}_r$ is reversible and $\boldsymbol{R}^{\{r\}} = R_r^+$ otherwise. For simplicity of notation, assume distinct enumeration of species and reactions, i.e., $\mathscr{S} \cap \mathscr{R} = \emptyset$. The local independence graph of the reactions of a CRN is complementary to the related concept of a kinetic independence graph [91]. The latter is formulated on the level of species instead of reactions.

In the following, let the subscript 0 for parents, children etc. stand for the respective functions with respect to the SR-graph and not the local independence graph. That is, for any $A \subseteq \mathscr{R}$, $\mathrm{pa}_0(A) \subset \mathscr{S}$ are the set of species that modulate reactions in A and $\mathrm{ch}_0(A) \subset \mathscr{S}$ are the set of species changed by reactions in A. We construct the edge sets $\mathscr{A}$ and $\mathscr{B}$ of the graph

$$\mathscr{G} = (\mathscr{S} \cup \mathscr{R}; \mathscr{B}, \mathscr{A})$$

from the edge set $\mathscr{E}$ by iteratively adding elements. For each $r \in \mathscr{R}$ add to $\mathscr{A}$ the in-arrows from its parent's parents, ignoring reflexive edges. To account for the dependence of reaction counters on the initial copy numbers of species $\mathscr{S}$, add directed edges from the parents. Consequently,

$$\begin{aligned} \mathscr{A} = &\left\{ (r', r) \in \mathscr{R}^2 \mid r' \in \mathrm{pa}_0(\mathrm{pa}_0(r)) \setminus \{r\} \right\} \\ &\cup \left\{ (d, r) \in \mathscr{S} \times \mathscr{R} \mid d \in \mathrm{pa}_0(r) \right\}. \end{aligned}$$

Further, we assume the most general case of dependences between initial copy numbers of all species by assuming that $\mathscr{S}$, the covariate set of the graph, is a complete subset relative to $\mathscr{G}$, i.e., every pair of vertices on the subspace is joined by an undirected edge, such that

$$\mathscr{B} = \left\{ \{d, d'\} \subset \mathscr{S} \mid d \neq d' \right\}.$$

We verify that $\mathscr{G}$ is a local independence graph in accordance with Def. 11. The pairwise chain Markov property, relative to $\mathscr{G}_{\mathscr{S}}$ is hence trivially satisfied due to completeness. Next, consider any $(j,r) \notin \mathscr{A}$. By construction we have

$$j \notin \begin{cases} \mathrm{pa}_0(\mathrm{pa}_0(r)) \setminus \{r\}, & \text{if } j \in \mathscr{R} \\ \mathrm{pa}_0(r), & \text{if } j \in \mathscr{S}. \end{cases}$$

Hence, if $j \in \mathscr{R}$, no species that is changed by j modulates r (i.e., $\mathrm{ch}_0(j) \cap \mathrm{pa}_0(r) = \emptyset$), and if $j \in \mathscr{S}$, it does not modulate r since $j \notin \mathrm{pa}_0(r)$. Consequently, $\{j\} \not\to \{r\} \mid (\mathscr{S} \cup \mathscr{R}) \setminus \{j,r\}$, i.e., the pairwise dynamic Markov property is satisfied.

VI. DISCUSSION OF COMBINED APPLICATIONS

We explore two application areas, namely continuous-time communication systems and and optimal design principles, where continuous-time path mutual information and directed information are combined with stochastic thermodynamic constraints. Other research traditions also merge information theory with thermodynamics, most notably information thermodynamics and information dynamics. A detailed comparison with these literatures lies outside the scope of the present work; for a comparison with the concept of learning rates, see Chetrite and Gupta [16], and for related contributions in information dynamics, see Spinney *et al.* [92].

We first adopt a classical information-theoretic viewpoint by interpreting deterministically manipulated chemical reaction networks as communication channels, where information encoding and transmission of information entail a thermodynamic cost in the form of heat dissipation, entropy production or work. Then we discuss the utility of the combined frameworks in evaluating the hypothesis that information flow functions as an emergent optimization principle, implicitly followed by natural selection in biological evolution [6, 93, 94]. As a complementary perspective, we also emphasize the relevance for synthetic biology, where these tools can inform the design of chemical reaction networks that are optimized for energy-efficient information processing.

In general, one should be careful with continuous approximations of CRNs, such as the chemical Langevin equation, in both information theory and thermodynamics. While heat dissipation may still be well approximated by diffusion models, stochastic consistency (cf. Supp. Sec. S2.4) is typically lost [77]. In a similar way, path-information measures are not consistent under continuous approximations because the continuum limit eliminates "discrete reaction information" [35].

A. Chemical reaction networks as communication channels

1. *A causal chemical communication model*

We introduce the communication model of interest, closely following the marked Poisson-type channel described by Frey [24]. Let $t > 0$ be the transmission duration and $M \colon \Omega \to \mathscr{M}$ be a random variable, $\mathscr{F}_0$-measurable, on a finite index set $\mathscr{M}$. We call M the message index (or simply message) and $\mathscr{M}$ the message set of a $(|\mathscr{M}|,t)$-code. For any $q \in (1,\infty)$ denote with $R_0 := \frac{1}{t}\log_q |\mathscr{M}|$ the maximum q-ary code rate of $(|\mathscr{M}|,t)$-codes. If M is uniformly distributed, then its entropy satifies $\frac{1}{t}H(M) = R_0$. Traditionally the message M is assumed to be uniformly distributed because the channel coding problem can be reduced to the case of a uniform message by a particular choice of random encoding [95] (p. 73). For biochemical systems, however, we acknowledge that the meaning of a message may be tied to its encoding, which precludes the traditional reduction method. Therefore, we allow for arbitrary message distributions and define the mean code rate [96]

$$R := \frac{1}{t}H(M) \leq R_0. \tag{55}$$

The symbol R, without any index, is reserved for code rates in this section. For notational simplicity we assume that all parallel reactions $r_1, r_2 \in \mathscr{R}$, $r_1 \neq r_2$, satisfy $\nu_{r_1} = \nu_{r_2}$, i.e., the change vectors have the same orientation for $\varepsilon = +$. For a CRN U with $|\mathscr{R}|$ potentially reversible reaction channels we redefine $\mathscr{V}$ (cf. (11)) to contain only forward change vectors, i.e.,

$$\mathscr{V} := \{\nu_r \mid r \in \mathscr{R}\} \setminus \{0\} \subset \mathbb{Z}^{|\mathscr{S}|}.$$

As before, assume $\mathscr{S}_X, \mathscr{S}_Y$ are disjoint subsets of species. The marginal reaction counters of X are

$$R_s^X(\varepsilon \nu_X) := \sum_{r \in V^{-1}(\pi_X^{-1}(\nu_X))} R_r^{\varepsilon}(s)$$

for all $\nu_X \in \mathscr{V}_X$ and $\varepsilon \in \{+,-\}$. Reaction counters for Y are defined analogously. Consequently, we have

$$X(s) = X(0) + \sum_{\nu_X \in \mathscr{V}_X} \nu_X \left(R_s^X(+\nu_X) - R_s^X(-\nu_X)\right)$$

for all $s \in [0,t]$ and analogously for Y. At time zero, the system U be always prepared with a well defined initial distribution $p(u,0)$ for every independent use of the communication channel (and hence independent of M).

Assigned to each message index $m \in \mathscr{M}$ is a deterministic manipulation protocol Γ_m^{X} of the core CRN U with bounded, piecewise-differentiable càdlàg trajectories on the interval $[0,t]$. This protocol is turned "on" at $t = 0$. Importantly, $m \mapsto \Gamma_m^{\mathsf{X}}$ be a bijective mapping. Next, we define the $\mathscr{F}_0$-measurable message process θ, such that for all $s \in [0,t]$

$$\theta(\omega, s) := \Gamma_{M(\omega)}^{\mathsf{X}}(s).$$

The manipulation protocols Γ_m^{X} have the superscript X as the effect of the manipulation is assumed to solely affect certain reaction channels $r \in \mathscr{R}_{\leftarrow}^{\circ}$ that exclusively change species X:

$$\mathscr{R}_{\leftarrow}^{\circ} \subseteq \{r | \pi_X(V(r)) \neq 0,\ \pi_X^{\mathsf{c}}(V(r)) = 0\} \subseteq \mathscr{R}^X,$$

where $\pi_X^{\mathsf{c}} \colon \mathscr{V} \to \mathbb{Z}^{|\mathscr{S}_X^{\mathsf{c}}|}$ is the coordinate map to coordinates of $\mathscr{S}_X^{\mathsf{c}} := \mathscr{S} \setminus \mathscr{S}_X$, complementary to $\mathscr{S}_X$.

The propensity functions for reaction channels $r \in \mathscr{R}_{\leftarrow}^{\circ}$ on $[0,t]$ depend on $m \in \mathscr{M}$ through the protocol Γ_m^{X}, which we

denote as $\lambda_r^{\varepsilon}(u,\Gamma_m^{\mathsf{X}}(s))$. If the protocols are non-constant, then the propensities are explicitly time-dependent, given m. All other reactions $r' \notin \mathscr{R}_{\leftarrow}^{\circ}$ are assumed to have no explicit time-dependence for simplicity.

While θ can be regarded a (noiseless) source encoding, we refer to X as the (noisy) channel encoding with channel encoding propensities

$$\lambda^X(\varepsilon\nu_X,u,\Gamma_m^{\mathsf{X}}(s)) := \sum_{r\in V^{-1}(\pi_X^{-1}(\nu_X))} \lambda_r^{\varepsilon}(u,\Gamma_m^{\mathsf{X}}(s))$$

analogous to (15), where we do not distinguish between m-dependent and m-independent reactions on the r.h.s. for brevity. We call Y the channel output with output propensities

$$\lambda^Y(\varepsilon\nu_Y,u) := \sum_{r\in V^{-1}(\pi_Y^{-1}(\nu_Y))} \lambda_r^{\varepsilon}(u).$$

Given the above specifications for the CRN, we introduce intensity processes that depend solely on the history of the general communication model M, θ, X and Y. The $\mathscr{F}^{\theta XY}$-intensities

$$\lambda_s^X(\varepsilon\nu_X) := \mathbb{E}\left[\lambda^X(\varepsilon\nu_X,U(s^-),\theta(s^-))\middle|\mathscr{F}_{s-}^{\theta XY}\right],$$

are called feedback channel encoding intensities, and the $\mathscr{F}^{MXY}$-intensities

$$\lambda_s^{M\to X}(\varepsilon\nu_X) := \mathbb{E}\left[\lambda^X(\varepsilon\nu_X,U(s^-),\theta(s^-))\middle|\mathscr{F}_{s-}^{MXY}\right],$$

are called feedback source-channel encoding intensities. Further, the $\mathscr{F}^{XY}$-intensities

$$\lambda_s^Y(\varepsilon\nu_Y) := \mathbb{E}\left[\lambda^Y(\varepsilon\nu_Y,U(s^-))\middle|\mathscr{F}_{s-}^{XY}\right],$$

are called (causal) channel output intensities. As we will discuss later, these channel output intensities represent the causal communication channel. For the evaluation of the relevant mutual information of a channel with feedback (and memory) yet another intensity is required. The $\mathscr{F}^{MY}$-intensities

$$\lambda_s^{M\to Y}(\varepsilon\nu_Y) := \mathbb{E}\left[\lambda^Y(\varepsilon\nu_Y,U(s^-))\middle|\mathscr{F}_{s-}^{MY}\right],$$

we call source-output intensities.

For the given setup with noiseless source encoding it can be shown that the source-channel encoding intensities are equal to the channel encoding intensities. For the formal statement of this equivalence we need the notion of the $\mathscr{F}$-predictable σ-algebra on $(0,\infty)\times\Omega$ [57](p. 8):

$$\mathscr{P}(\mathscr{F}) := \sigma\left((s,\infty)\times A\colon A\in\mathscr{F}_s, s\geq 0\right)$$

Then $((0,\infty)\times\Omega,\mathscr{P}(\mathscr{F}),\mu\otimes\mathbb{P})$ is a measure space, where μ denotes the Lebesgue measure. For our purpose we only look at the restricted space $(0,t]\times\Omega$ and denote restricted filtrations as $\mathscr{F}_{[0,t]}$.

Proposition 2. *If θ is a noiseless source encoding, i.e., it is defined as specified above, then the channel encoding intensities are equivalent to the source-channel encoding intensities. That is, for all $\varepsilon\in\{+,-\}$ and $\nu_X\in\mathscr{V}_X$ it holds that*

$$\lambda^X(\varepsilon\nu_X) = \lambda^{M\to X}(\varepsilon\nu_X) \quad (\mu\otimes\mathbb{P})\text{-a.e. on } \mathscr{P}(\mathscr{F}_{[0,t]}^{XY}\vee\sigma(M)).$$

Further, the source-output intensities satisfy

$$\lambda_s^{M\to Y}(\varepsilon\nu_Y) = \mathbb{E}\left[\lambda^Y(\varepsilon\nu_Y,U(s^-))\middle|\mathscr{F}_{s-}^{\theta Y}\right],$$

$(\mu\otimes\mathbb{P})$-a.e. on $\mathscr{P}(\mathscr{F}_{[0,t]}^{Y}\vee\sigma(M))$, for all $\varepsilon\in\{+,-\}$ and $\nu_Y\in\mathscr{V}_Y$.

A proof is presented in Supp. Sec. S1.6 A.

Lastly, let

$$D\colon \mathscr{D}([0,t],\mathscr{Y})\to\mathscr{M}\cup\{0\}$$

be a deterministic decoding function, which maps every output trajectory back to either of the messages in $\mathscr{M}$ or declares an erasure, by mapping to zero. We may associate two distinct conceptions with this decoding function: (i) the *extrinsic decoder* is a (computationally) learned function by an external observer, based on experimental observations of $Y_{[0,t]}$ given M; (ii) the *intrinsic decoder* refers to the conception that each $m\in\mathscr{M}$ stands for a pair of input and output of a bijective function, where Γ_m^{X} is the input and the output is a deterministic downstream action implemented by the biological cell. However, in the following we consider an *abstract decoder* which just determines the function space required for information theoretic analyses of chemical communication models. The (mean) probability of a decoding error be defined as

$$P_e := \mathbb{P}(M\neq D(Y_{[0,t]})) \tag{56}$$

While other performance measures, like the maximum probability of error $\max_{m\in\mathscr{M}}\mathbb{P}(M\neq D(Y_{[0,t]})\mid M=m)$ exist, we focus on the definition (56).

In general, the relevant mutual information for a continuous-time chemical communication channel with memory and use of feedback is

$$\mathbb{I}(M;Y_{[0,t]}) = \sum_{\varepsilon}\sum_{\nu_Y\in\mathscr{V}_Y}\int_0^t \mathbb{E}\left[\phi(\lambda_s^{M\to Y}(\varepsilon\nu_Y)) - \phi(\hat{\lambda}_s^Y(\varepsilon\nu_Y))\right]\mathrm{d}s,$$

where we used that $\mathbb{I}(M;U(0))=0$ by construction. Consistently with Proposition 2 we have

$$\mathbb{I}(M;Y_{[0,t]}) = \mathbb{I}(\theta_{[0,t]};Y_{[0,t]}) = \mathbb{I}(\theta_{[0,t]}\to Y_{[0,t]})$$

for a noiseless source-encoding $\theta_{[0,t]}$.

Now let us elaborate on the construction of the communication system by deriving its (extended) local independence graph $\mathscr{G}=(\mathscr{I}\cup\mathscr{R};\mathscr{B},\mathscr{A})$. Here, $\mathscr{I}:=\{\mathsf{M}\}\cup\mathscr{S}$, where M is the label for the node of the baseline variable M, whose realization is equivalent to a full trajectory $\theta_{[0,t]}$. Mainly, $\mathscr{G}$ is derived from the SR-graph of U as outlined in Sec. V C, augmented with the baseline vertex M. Hence, the least-assumptious edge sets are

$$\begin{aligned}\mathscr{A} = \bigcup_{r\in\mathscr{R}} \big(&\{(r',r)\in\mathscr{R}^2 \mid r'\in\mathrm{pa}_0(\mathrm{pa}_0(r))\setminus\{r\}\}\\ &\cup\{(d,r)\in\mathscr{S}\times\mathscr{R}\mid d\in\mathrm{pa}_0(r)\}\big)\\ &\cup\{(\mathsf{M},r)\mid r\in\mathscr{R}_{\leftarrow}^{\circ}\}\end{aligned}$$

and

$$\mathscr{B} = \left\{ \{d,d'\} \subset \mathscr{S} \mid d \neq d' \right\}.$$

In the absence of further restrictions on the chemical communication system, choosing $(\mathscr{S},\mathscr{B})$ as a complete subgraph is consistent with the generic conditional independence structure induced by chemical reaction dynamics. A time-unrolled description of U as a dynamic Bayesian network, with species population values $(\ldots, U(-2\,\mathrm{d}s), U(-\mathrm{d}s), U(0))$ at previous infinitesimal time layers, is sparsely connected across adjacent layers, but marginalization over the unobserved reaction history typically produces a fully connected conditional independence graph on $U(0)$ for any non-infinitesimal time horizon. Postulating a missing edge $\{d,d'\} \notin \mathscr{B}$ therefore amounts to requiring that $U_d(0) \perp\!\!\!\perp U_{d'}(0) \mid U_{\mathscr{S}\setminus\{d,d'\}}(0)$ holds uniformly over the admissible choices of stoichiometry, reaction propensities, and preparation protocols for $p(u,0)$. A sparser $\mathscr{B}$ thus presupposes dynamically decoupled modules (e.g., bivariate subsystems) or a specific ergodic preparation protocol that asymptotically enforces such separations. To keep the framework applicable to general open CRNs with arbitrary preprocessing, we therefore adopt the structurally least restrictive choice and take $(\mathscr{S},\mathscr{B})$ to be complete.

In a proper communication model, the message M must be locally independent of U, which we have achieved through source-encoding via $\mathscr{F}_0$-measurable manipulation protocols. Note that M and $U_{[0,t]}$ are not independent random variables since $M \to U$. Now assume that

$$\begin{aligned} \mathscr{R} \cap \mathrm{pa}(\mathrm{ch}(\mathsf{M})) &\subseteq \mathscr{R}^{\circ}_{\leftarrow}, \\ \mathscr{S} \cap \mathrm{pa}(\mathrm{ch}(\mathsf{M})) &\subseteq \mathscr{S}_X \cup \mathscr{S}_Y \end{aligned}$$

i.e., all reaction-co-parents of M are in $\mathscr{R}^{\circ}_{\leftarrow} \subseteq \mathscr{R}^X$ and all species-co-parents are in $\mathscr{S}_X \cup \mathscr{S}_Y$. Then, by construction and the global dynamic Markov property, we have that

$$\mathsf{M} \not\to \tilde{\mathscr{R}} \mid \mathscr{R}^{\circ}_{\leftarrow} \cup \mathscr{S}_X \cup \mathscr{S}_Y.$$

for all non-empty sets $\tilde{\mathscr{R}} \subseteq \mathscr{R} \setminus \mathscr{R}^{\circ}_{\leftarrow}$. In particular, we then have

$$\mathsf{M} \not\to \mathscr{R}^Y \mid \mathscr{R}^X \cup \mathscr{S}_X \cup \mathscr{S}_Y.$$

If additionally, there are no X- and no Y-indistinguishable reactions, then Theorem 3 implies

$$\mathbb{I}(\theta_{[0,t]} \to Y_{[0,t]} \| X_{[0,t]}) = \mathbb{I}(M \to Y_{[0,t]} \| X_{[0,t]}) = 0, \tag{57}$$

where the first equality follows since source-encoding is noiseless and bijective. The property in (57) is actually an important classification of a communication system that was first introduced by Massey [11] for discrete-time models. We provide an adaptation of this definition to our model class.

Definition 12. *For a $(|\mathscr{M}|,t)$-source code with message M, let $\theta_{[0,t]}$ be a (potentially noisy) source encoding and $Y_{[0,t]}$ be the channel output. A chemical communication system $(M,\theta,U,\mathscr{S}_X,\mathscr{S}_Y)$ with designated channel encoding $X_{[0,t]}$ is called* causal *if*

$$(s,\omega) \mapsto \mathbb{E}\left[\lambda^Y(\varepsilon \nu_Y, U(s^-)) \middle| \mathscr{F}^{XY}_{s-} \vee \mathscr{F}^{\theta}_t \right]$$

is a $\mathscr{F}^{XY}_{[0,t]}$-intensity for all $\varepsilon \in \{+,-\}$ and all $\nu_Y \in \mathscr{V}_Y$.

A direct consequence is that causal communication systems satisfy

$$\mathbb{I}(\theta_{[0,t]} \to Y_{[0,t]} \| X_{[0,t]}) = 0. \tag{58}$$

For a noiseless source-encoding Definition 12 is actually equivalent to $\mathbb{I}(M \to Y_{[0,t]} \| X_{[0,t]}) = 0$ by $\sigma(M) = \mathscr{F}^{\theta}_t$ and Theorem 3. Similar to Massey, we obtain the following inequality for any causal chemical communication system.

Proposition 3. *Let $(\theta_{[0,t]}, X_{[0,t]}, Y_{[0,t]})$ be a causal communication system for a $(|\mathscr{M}|,t)$-source code $\theta_{[0,t]}$. Then*

$$\mathbb{I}(\theta_{[0,t]}; Y_{[0,t]}) \leq \mathbb{I}(X_{[0,t]} \to Y_{[0,t]})$$

with equality if and only if $\mathbb{I}(X_{[0,t]} \to Y_{[0,t]} \| \theta_{[0,t]}) = 0$.

Proof. Since $U(0)$ is prepared independently of $\theta(0)$ it holds that $\mathbb{I}(\theta(0); Y(0)) = 0$. Further, $\theta_{[0,t]}$ is locally independent from $Y_{[0,t]}$ by construction. Hence $\mathbb{I}(Y_{[0,t]} \to \theta_{[0,t]}) = 0$. Consequently,

$$\begin{aligned} &\mathbb{I}(\theta_{[0,t]}; Y_{[0,t]}) = \mathbb{I}(\theta_{[0,t]} \to Y_{[0,t]}) \\ \leq &\mathbb{I}(\theta_{[0,t]} \to Y_{[0,t]}) + \mathbb{I}(X_{[0,t]} \to Y_{[0,t]} \| \theta_{[0,t]}) \\ = &\mathbb{I}(\theta_{[0,t]}, X_{[0,t]} \to Y_{[0,t]}) \\ = &\mathbb{I}(X_{[0,t]} \to Y_{[0,t]}) + \mathbb{I}(\theta_{[0,t]} \to Y_{[0,t]} \| X_{[0,t]}) \\ = &\mathbb{I}(X_{[0,t]} \to Y_{[0,t]}), \end{aligned}$$

where we used (8) in the first line, non-negativity of the causally conditioned DI in the second line, the chain rule for DI in the third line and (58) in the last line. □

In that, $\mathbb{I}(X_{[0,t]} \to Y_{[0,t]} \| \theta_{[0,t]})$ quantifies the surplus of causal information provided by X to Y without the causal information already provided by θ. We can think of this as the amount of channel encoding noise that is causally transmitted to Y. Since the $\mathscr{F}^{\theta Y}$-intensities

$$\mathbb{E}\left[\lambda^Y(\varepsilon \nu_Y, U(s^-)) \middle| \mathscr{F}^{\theta Y}_{s-} \right]$$

usually differ from the $\mathscr{F}^{\theta XY}$-intensities

$$\mathbb{E}\left[\lambda^Y(\varepsilon \nu_Y, U(s^-)) \middle| \mathscr{F}^{\theta XY}_{s-} \right]$$

for any $\nu_Y \in \mathscr{V}_Y$, $\varepsilon \in \{+,-\}$, this surplus information typically has a positive value. Such positive surplus information can be exemplarily verified with the small models in Sec. VI C under the assumption of just two different constant chemostat concentration values as Γ^{X}_m-trajectories.

2. *The noisy channel coding problem in chemical communication*

In the standard noisy channel coding problem for a causal communication system, the causally conditioned channel law $\mathbb{P}^{Y\|X}_{[0,t]}$ is taken as fixed (see Appendix C for a definition). Here, we discuss the notion of channel capacity in the context of chemical communication.

The channel capacity is a property of a causal chemical channel $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$ with fixed input and output species identities and a fixed channel law. Communicators may choose a code design – an encoder (the dynamics of X) and a decoder (post-processing of Y) – but they cannot alter the physical medium or its intrinsic randomness. Accordingly, the fluctuations that drive Y beyond those induced by X are modeled as the channel's internal noise, which is not under the communicators' control. This setup cleanly separates "code design" from "channel control". Allowing $\mathbb{P}^{Y\|X}_{[0,t]}$ to change turns the problem into a joint code–channel co-design task. While such co-design may be a reasonable modeling choice for biochemical systems in some cases, we deliberately focus on the traditional fixed-channel formulation within our chemical communication model.

The channel law $\mathbb{P}^{Y\|X}$ is equivalently represented by the collection of predictable intensities $\{\lambda^Y(\varepsilon\nu_Y) \mid \varepsilon \in \{+,-\}, \nu_Y \in \mathscr{V}_Y\}$, which is unique up to stochastic equivalence [57] (p. 31). This equivalence is readily apparent from the path likelihood (23). Whereas much of the literature formulates such likelihoods on the marked-jump space $\bigcup_{n=0} ([0,t] \times (\mathscr{V}_Y \cup -\mathscr{V}_Y))^n$, we work with an equivalent, and more intuitive, representation on the space of càdlàg trajectories $\mathscr{D}([0,t], \mathscr{X} \times \mathscr{Y})$. To this end, we introduce a collection of predictable non-negative functionals

$$\{\Lambda^Y_{\varepsilon,\nu_Y} \colon [0,t] \times \mathscr{D}([0,t], \mathscr{X} \times \mathscr{Y}) \to \mathbb{R}_{\geq 0} \mid \varepsilon \in \{+,-\}, \nu_Y \in \mathscr{V}_Y\}$$

chosen such that

$$\lambda^Y(\varepsilon\nu_Y) = \Lambda^Y_{\varepsilon,\nu_Y}(\cdot, (X,Y)_{[0,t]}), \qquad (\mu \otimes \mathbb{P})\text{-a.e.}$$

on the predictable space $\left([0,t] \times \Omega, \mathscr{P}(\mathscr{F}^{XY}_{[0,t]})\right)$. Their existence is guaranteed by the factorization lemma [97], and they can be chosen as the regenerative version specified by Brémaud [57] (p. 63).

With this notation, the causally conditioned path likelihood admits the form

$$\begin{aligned} &L^{Y\|X}_{[0,t]}(x_{[0,t]}, y_{[0,t]}) \\ &= \exp\left(-\sum_{\varepsilon}\sum_{\nu_Y \in \mathscr{V}_Y} \int_0^t \Lambda^Y_{\varepsilon,\nu_Y}(s, (x,y)_{[0,t]})\, \mathrm{d}s\right) \\ &\quad \times \exp\left(\sum_{\varepsilon}\sum_{\nu_Y \in \mathscr{V}_Y} \int_0^t \ln\left(\Lambda^Y_{\varepsilon,\nu_Y}(s, (x,y)_{[0,t]})\right) \mathrm{d}\rho^Y_s(\varepsilon\nu_Y)\right) \end{aligned}$$

where the counting (process) trajectories $\rho^Y(\varepsilon\nu_Y)$ are fully determined by $y_{[0,t]}$.

For coding with transmission duration t and message M, and under the constraint that the channel law $\mathbb{P}^{Y\|X}$ is kept invariant, the remaining degrees of freedom are: (i) the $(|\mathscr{M}|, t)$ source code, which specifies the functional form of the protocols Γ^{X}_m on $[0,t]$; (ii) the feedback channel encoding, i.e., the causal law $\mathbb{P}^{X\|\theta,Y}$; and (iii) the initial distribution $p(u,0)$, if it can be manipulated.

As in the output case, the feedback encoding law $\mathbb{P}^{X\|\theta,Y}$ can be represented by the collection of predictable input intensities $\{\lambda^X(\varepsilon\nu_X) \mid \varepsilon \in \{+,-\}, \nu_X \in \mathscr{V}_X\}$ (unique up to stochastic equivalence). In particular, we may choose predictable functionals

$$\{\Lambda^X_{\varepsilon,\nu_X} \mid \varepsilon \in \{+,-\}, \nu_X \in \mathscr{V}_X\}$$

such that, on the predictable space $\left([0,t] \times \Omega, \mathscr{P}(\mathscr{F}^{\theta XY}_{[0,t]})\right)$,

$$\lambda^X(\varepsilon\nu_X) = \Lambda^X_{\varepsilon,\nu_X}\big(\cdot, (\theta, X, Y)_{[0,t]}\big), \qquad (\mu \otimes \mathbb{P})\text{-a.e.}$$

Equivalently, for any fixed protocol Γ^{X}_m and history $(x,y)_{[0,t]}$, the map $s \mapsto \Lambda^X_{\varepsilon,\nu_X}(s, \Gamma^{\mathsf{X}}_m, (x,y)_{[0,t]})$ specifies a version of the input intensity along that trajectory. Accordingly, the design space of admissible channel codes can be identified with the class of such predictable functionals (subject to regularity constraints). Replacing this family by an alternative collection

$$\{\check{\Lambda}^X_{\varepsilon,\nu_X} \mid \varepsilon \in \{+,-\}, \nu_X \in \check{\mathscr{V}}_X\}$$

induces, in general, a different joint measure $\check{\mathbb{P}}$ on $(\Omega, \mathscr{F}^{M\theta XY})$. The invariance constraint on the physical channel is precisely that the conditional output law remains unchanged, i.e., "$\mathbb{P}^{Y\|X} = \check{\mathbb{P}}^{Y\|X}$". A class of such admissible channel codes will be derived in Sec. VI A 5, Theorem 5.

Finally, degrees (i) and (ii) may be combined into a single causal feedback source–channel encoding law $\mathbb{P}^{X\|M,Y}_{[0,t]}$ on $[0,t]$, which is equivalently represented by predictable intensities for the message-driven input process. Concretely, one may introduce predictable functionals

$$\{\Lambda^{M\to X}_{\varepsilon,\nu_X} \mid \varepsilon \in \{+,-\}, \nu_X \in \mathscr{V}_X\}$$

such that

$$\lambda^{M\to X}_s(\varepsilon\nu_X) = \Lambda^{M\to X}_{\varepsilon,\nu_X}\big(s, M, (X,Y)_{[0,t]}\big), \qquad (\mu \otimes \mathbb{P})\text{-a.e.}$$

A more formal notion of codes is presented in Appendix E. Next we turn to the definition of a suitable class of code families, which is required for the formulation of asymptotic limiting properties.

In classical channel coding, longer block codes are often built by concatenating codewords in time. Given a $(|\mathscr{M}_a|, t_a)$-code and a $(|\mathscr{M}_b|, t_b)$-code, one sends a message $(m_a, m_b) \in \mathscr{M}_a \times \mathscr{M}_b$ by sending a $(|\mathscr{M}_a|, t_a)$-codeword and, after a channel and feedback reset (or under a memoryless channel), sending a $(|\mathscr{M}_b|, t_b)$-codeword. Such a code scheme enforces that single codewords are sent via independent uses of the channel. While this works well for memoryless channels, it may result in practically low code rates for channels with memory as the guard/relaxation intervals for resets to the a prescribed channel initialization p_0 can be substantial and may dominate the communication time [98]. Previous continuous-time capacity results have typically assumed a memoryless channel [24, 99] and feedback with instantaneous resets [12]. For channels with memory, where small guard times lead to intersymbol interference across code blocks [98], it is therefore favorable, both theoretically and practically, to use a single code with a larger codebook, spanning the whole transmission, instead of

a sequence of codewords that are decoded independently under intersymbol interference. Continuous-time collections of messages and codes of practical and theoretical interest will hence be in the following class.

Definition 13. *A collection $(M_t)_{t\geq 0}$ of messages $M_t\colon \Omega \to \mathscr{M}_t$ is called monotonous if $|\mathscr{M}_{t'}| \geq |\mathscr{M}_t|$ for all $t' > t > 0$. Any collection of codes $(\mathsf{C}_t)_{t\geq 0}$ for $(M_t)_{t\geq 0}$ is also called monotonous.*

While this definition does not ensure that the code rate $t \mapsto R(t) = \frac{1}{t}H(M_t)$ is a monotonously increasing function, it formally accounts for the time-concatenation idea in continuous-time for a single channel use and prevents the maximum code rate $R_0(t)$ to vanish at finite times. For instance, the time-concatenation of potentially interference-aware $(|\mathscr{M}_0|, \Delta t)$ code-slots with $|\mathscr{M}_0| \geq 2$ results in a time-dependent maximum code rate

$$t \mapsto R_0(t) = \frac{\log(|\mathscr{M}_0|)}{t}\left\lfloor \frac{t}{\Delta t}\right\rfloor \leq \frac{\log(|\mathscr{M}_0|)}{\Delta t} = \lim_{t\to\infty} R_0(t).$$

Since, unlike in discrete-time lossless coding (where the per-symbol rate is bounded), a continuum-valued protocol can in principle encode countably many messages on an arbitrarily short interval $[0, \Delta t)$, one could make the apparent code rate arbitrarily large by choosing Δt very small.

Having introduced practical issues of code design for chemical communication, we now turn to the notions of achievability and capacity required for coding theorems. The following definitions are adapted from Csiszár and Körner [100].

Definition 14. *Let $\varepsilon \in [0,1)$ and the code C be consistent with $\mathbb{P}$. The code C is called an $(|\mathscr{M}|, t, \varepsilon)$-code for the message M if*

$$P_e = \mathbb{P}\left(M \neq D(Y_{[0,t]})\right) \leq \varepsilon.$$

Definition 15 (ε-achievable rates and ε-capacity)**.** *For $\varepsilon \in [0,1)$, a nonnegative number R^* is an* ε-achievable rate *for the channel $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$ if for all $\delta > 0$ exists $t_0 > 0$ such that for all $t \geq t_0$ exists a message $M_t\colon \Omega \to \mathscr{M}_t$ such that*

(i) $R^* - \delta \leq \frac{1}{t}H(M_t)$ and

(ii) exists an $(|\mathscr{M}_t|, t, \varepsilon)$-code for M_t.

A number R^* is an achievable rate if it is ε-achievable for all $\varepsilon \in (0,1)$.

The supremum of ε-achievable rates, respectively, achievable rates is called ε-capacity C_ε, respectively, capacity C of the channel $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$.

Since C_ε decreases as $\varepsilon \searrow 0$, we have

$$C = \inf_{\varepsilon\in(0,1)} C_\varepsilon = \lim_{\varepsilon\searrow 0} C_\varepsilon.$$

In line with the existing literature, we explicitly point out that the collection of messages $(M_t)_{t\geq 0}$ is subject to design and is not fixed in the noisy channel coding problem. However, as we pointed out when defining code rates, semantic aspects in biochemical systems may constrain admissible message random variables (and their distributions) in practice.

The information capacity $C_{\mathbb{I}}$ for any *information stable* channel $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$ be defined as

$$C_{\mathbb{I}} := \liminf_{t\to\infty} C_{\mathbb{I}}(t) \quad \text{with} \tag{59}$$

$$c_{\mathbb{I}}(t) := \sup_{M_t} \sup_{\theta_{[0,t]}} \sup_{X_{[0,t]}} \frac{1}{t}\mathbb{I}(M_t; Y_{[0,t]}) \tag{60}$$

where $\sup_{X_{[0,t]}}$ optimizes over the families of functions $\{\Lambda^X_{\varepsilon,\nu_X} \mid \varepsilon \in \{+,-\}, \nu_X \in \mathscr{V}_X\}$, and thereby implicitly over the feedback channel encoding distribution $\mathbb{P}^{X\|\theta,Y}_{[0,t]}$, as well as the initial distribution $p(u,0)$ under the constraint that it leaves the channel law invariant. Further, $\sup_{\theta_{[0,t]}}$ optimizes over the space of source encoding protocol functions Γ^{X}_m and $\sup_{M_t}$ optimizes over the cardinality of $\mathscr{M}_t$ and well as the distribution of M_t.

The limit in (59) is essential for defining information capacity because, in principle, it enables identification with the operational channel capacity, i.e., the supremum of rates R for which there exists a sequence of codes with $P_e \to 0$ as $t \to \infty$. Establishing this equivalence lies beyond the scope of the present work and we confine ourselves to the converse channel coding theorem (cf. Theorem 4). In classical definitions of the channel capacity for discrete-time systems with memory and feedback, the latter two suprema are typically not involved in the definition. Optimizing over $\theta_{[0,t]}$ and M_t can actually be subsumed into the optimization over $X_{[0,t]}$ for causal systems both channel encoding and source encoding are noiseless (i.e., one-to-one mappings). Since the channel encoding is always noisy for chemical communication, only $\sup_{M_t}$ can be subsumed into $\sup_{\theta_{[0,t]}}$ for noisless source coding. Similarly, if a (joint) source-channel code is used instead of a split source-channel code, then the remaining optimization problems are $\sup_{M_t}\sup_{X_{[0,t]}}$ since θ is not explicitly modeled.

For intuition, we propose a generalization of information stability (as defined in [101]) to continuous-time channels, but we leave proving its applicability to future work.

Definition 16. *The channel $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$ is called* information stable *if there exists a sequence $\left((M_t^{(n)}, \theta^{(n)}_{[0,t]})_{t\geq 0}, X^{(n)}\right)$, $n \in \mathbb{N}$ such that for all $\delta > 0$*

$$\limsup_{n\to\infty} \mathbb{P}^{(n)}\left(\left|\frac{1}{n}\ln\frac{\mathrm{d}\mathbb{P}^{Y|M_t^{(n)}}_{[0,n]}}{\mathrm{d}\mathbb{P}^{Y^{(n)}}_{[0,n]}} - C_{\mathbb{I}}\right| > \delta\right) = 0$$

(i.e., with convergence in probability).

Here, we added the sequence of processes/distributions since the capacity is typically not achieved by any element of the optimization space. Information stability is typically satisfied for ergodic communication system processes.

What is not explicit in (60) is that the function spaces over which we optimize are typically restricted by additional con-

straints, such as upper bounds on reaction rates, a total maximum energy budget and confinements on the practical degrees of freedom in $\mathbb{P}^{X\|\theta,Y}_{[0,t]}$ for chemical systems. We make energetic constraints explicit in the section VI A 3 and give a converse coding theorem, while practical degrees of freedom are discussed in VI A 5.

3. Channel coding under energetic constraints

Let us now introduce a physically principled constraint on the search space of $X_{[0,t]}$ and $\theta_{[0,t]}$. For each $m \in \mathscr{M}_t$ we consider the heat $q_m(t)$ dissipated by the trajectory $U_{[0,t]}$. Let $Q_0 \geq 0$ be the upper bound on the time-averaged heat, such that

$$\frac{1}{t}\mathbb{E}[q_m(t)] \leq Q_0. \tag{61}$$

Such a constraint is also called input-output cost function [95], because it considers the dissipative cost of the whole CRN U and not only the energy dissipated by channel encoding into X. Evaluating the finite-time information capacity under this constraint yields the thermodynamic capacity-cost functions

$$Q_0 \mapsto C_{\mathbb{I}}(Q_0), \quad Q_0 \mapsto c_{\mathbb{I}}(t, Q_0).$$

Alternatively, we can also constrain the chemical work or the total entropy production

$$\frac{1}{t}\mathbb{E}[w^{\mathrm{c}}(t)] \leq W_0, \quad \frac{1}{t}\mathbb{E}[h^{\mathrm{tot}}_m(t)] \leq \varepsilon_0$$

with $W_0, \varepsilon_0 \geq 0$, leading to other thermodynamic capacity-cost functions. If the protocols Γ^{X}_m are constant protocols, where m just specifies a different concentration of chemostat species, then both the heat dissipation and the entropy production constraint can be rewritten in the limit $t \to \infty$, such that

$$e_p \leq \beta Q_0 \quad \text{or} \quad e_p \leq \varepsilon_0$$

by virtue of (52).

Further alternatives, that aim to (i) approximate the cost, if no information about the internal entropy h_{i} is available, or (ii) isolate the thermodynamic cost of the channel encoding reactions (excluding the cost of "pure" transmission) may use the above constraints, but with (46), (39) and (49). In the case of pure encoding cost, we would just sum over reaction channels $\mathfrak{R}_r$ that are directly coupled to the protocols Γ^{X}_m.

For the given communication channel we can also provide a converse (channel coding) theorem [20, 24], which is valid for any of the discussed energy constraints. Importantly, it is also valid for information unstable channels, since it only provides an upper bound on the operational capacity that is not necessarily tight.

Lemma 2. *Consider an $(|\mathscr{M}|,t)$-code for the channel $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$ with message $M\colon \Omega \to \mathscr{M}$. If the code exhibits the code rate R, then*

$$R \leq \frac{1}{t}\mathbb{I}(M; Y_{[0,t]}) + \frac{1}{t}H_B(P_e) + P_e R_0,$$

where $H_B(p) = -p\log(p) - (1-p)\log(1-p)$ denotes the binary entropy function (for an arbitrary log-base).

Theorem 4. *Let R be an achievable rate of the channel $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$ under a given set of energy constraints on admissible codes. For simplicity, we exemplarily use the heat dissipation constraint with upper bound Q_0. Then it holds*

$$R \leq C(Q_0) \leq C_{\mathbb{I}}(Q_0),$$

where $C(Q_0)$ denotes the operational capacity under the dissipation constraint.

Proofs for both, the Lemma and the Theorem, is provided in Supp. Sec. S1.6 A.

Lastly, although we formulate these optimization problems under thermodynamic constraints, they should often be complemented by kinetic constraints. For instance, at thermodynamic equilibrium, where the rate constants satisfy (27), one could in principle let all rate constants diverge while maintaining fixed propensity ratios. This, however, is clearly unphysical. Additional assumptions, such as lower bounds on mean sojourn times [37] or explicit models of the reaction constants [64], are thus necessary to ensure physical plausibility. In stochastic thermodynamics, the interplay between thermodynamic and kinetic constraints has been explored in the context of thermodynamic and kinetic uncertainty relations [102].

4. Energy-per-bit perspective

A dual perspective to information capacity is the energy-per-bit-rate function [95]. Define the average energy per information unit (given the code rate R in an arbitrary log-base) as

$$E_{\mathrm{b}}(R) := \frac{Q_0}{R}.$$

Instead of the heat dissipation bound Q_0, we can of course also use any other bound. Since the capacity-cost functions $Q_0 \mapsto C_{\mathbb{I}}(Q_0)$ and $Q_0 \mapsto c_{\mathbb{I}}(t, Q_0)$ are monotonously increasing, the minimum-energy-per-bit-rate function can be defined as their generalized inverse:

$$E^{\min}_{\mathrm{b}}(R) := \inf\{E_{\mathrm{b}} \geq 0 | C_{\mathbb{I}}(E_{\mathrm{b}}R) \geq R\}, \tag{62}$$

$$e^{\min}_{\mathrm{b}}(t,R) := \inf\{E_{\mathrm{b}} \geq 0 | c_{\mathbb{I}}(t, E_{\mathrm{b}}R) \geq R\}. \tag{63}$$

Given Theorem 4, $E^{\min}_{\mathrm{b}}(R)$ is actually a lower bound on the minimum amount of heat that must be dissipated per unit of information sent to reliably communicate at code rate R over a given channel. This bound is conjectured to be tight for information stable channels. The following Lemma makes the asymptotic version numerically more accessible.

Lemma 3. *It holds that*

$$E^{\min}_{\mathrm{b}}(R) = \limsup_{t\to\infty} e^{\min}_b(t,R).$$

A proof is provided in Supp. Sec. S1.6 B.

Similarly, we can define the minimum average heat dissipation required for a minimum information rate R:

$$q_{\mathbb{I}}^{\min}(t,R) := \inf_{\theta_{[0,t]}} \inf_{X_{[0,t]}} \frac{1}{t}\mathbb{E}[q(s)], \tag{64}$$

subject to the constraint $\mathbb{I}(\theta_{[0,t]};Y_{[0,t]}) \geq Rt$ for both infima. It's asymptotic version is

$$Q_{\mathbb{I}}^{\min}(R) := \limsup_{t\to\infty} q_{\mathbb{I}}^{\min}(t,R).$$

The two presented notions of minimum energy dissipation per information unit are shown to be equivalent.

Proposition 4.

$$e_{\mathrm{b}}^{\min}(t,R) = \frac{1}{R} q_{\mathbb{I}}^{\min}(t,R)$$
$$E_{\mathrm{b}}^{\min}(R) = \frac{1}{R} Q_{\mathbb{I}}^{\min}(R)$$

As before, the definitions and equivalence results work analogously with total entropy production, etc. A proof is provided in Supp. Sec. S1.6 B.

5. *Degrees of freedom in channel coding*

Chemical reactions cannot be changed arbitrarily in reality. To find the supremum $\sup_{X_{[0,t]}}$ in the capacity definition, we have just a few options. First, we may apply "thermodynamic parameter tuning" by exchanging certain species with stoichiometrically equivalent ones that nonetheless exhibit different reaction affinities. That is, we change g, such that $\Delta_r^+ g$ is modified for a subset of reaction channel. Otherwise, we can change the fixed chemostat concentrations z (not the source encoding protocols Γ_m^{X}), such that only those reaction channels are changed, that leave the channel probabilities invariant. Lastly, we can vary the reaction network topology by introducing or erasing reactions, i.e., adding and removing new stoichiometric vectors (including their chemostat couplings). This may also include a change in the number and identities of species that are not in $\mathscr{S}_X \cup \mathscr{S}_Y$. Again, all these modification must not change the channel $\mathbb{P}_{[0,t]}^{Y\|X}$. This leaves us with the question of which reaction channels are actually allowed to be modified?

We address this problem with a network-topological classification based on the species-reaction graph.

Definition 17. *Consider the reversible species-reaction graph of the open CRN U. A reaction channel $\mathfrak{R}_r$ is classified as ...*

(a) feedforward ($\mathscr{R}_{\mathrm{ff}}$) if there is a 'trail' from $\mathscr{S}_X$ to $\mathscr{S}_Y$ that passes through node $\mathfrak{R}_r$.

(b) feedback ($\mathscr{R}_{\mathrm{fb}}$) if there is a 'trail' from $\mathscr{S}_Y$ to $\mathscr{S}_X$ that passes through node $\mathfrak{R}_r$.

(c) modulating ($\mathscr{R}_{\mathrm{c}}$) if either of the following is true:

(i) $\mathfrak{R}_r$ is purely modulating ($\mathscr{R}_{\mathrm{c}}^{\circ}$) if it is neither a feedforward nor a feedback reaction and there exists a 'trail' from node $\mathfrak{R}_r$ to $\mathscr{S}_X \cup \mathscr{S}_Y$.

(ii) $\mathfrak{R}_r$ is feedforward, but not feedback, and there exists a 'trail' from $\mathfrak{R}_r$ to $\mathscr{S}_X$ that avoids $\mathscr{S}_Y$.

(iii) $\mathfrak{R}_r$ is feedback, but not feedforward, and there exists a 'trail' from $\mathfrak{R}_r$ to $\mathscr{S}_Y$ that avoids $\mathscr{S}_X$.

(d) disconnected ($\mathscr{R}_{\mathrm{d}}$) if there is no 'trail' from node $\mathfrak{R}_r$ to $\mathscr{S}_X \cup \mathscr{S}_Y$.

(e) input modulating ($\mathscr{R}_{\mathrm{Tx}}$) if it is modulating and exists a 'trail' from $\mathfrak{R}_r$ to $\mathscr{S}_X$ that avoids $\mathscr{S}_Y$.

(f) output modulating ($\mathscr{R}_{\mathrm{Rx}}$) if it is modulating and exists a 'trail' from $\mathfrak{R}_r$ to $\mathscr{S}_Y$ that avoids $\mathscr{S}_X$.

(g) purely input modulating if $r \in \mathscr{R}_{\mathrm{Tx}}^{\circ} := \mathscr{R}_{\mathrm{c}}^{\circ} \cap \mathscr{R}_{\mathrm{Tx}} \setminus \mathscr{R}_{\mathrm{Rx}}$.

(h) purely output modulating if $r \in \mathscr{R}_{\mathrm{Rx}}^{\circ} := \mathscr{R}_{\mathrm{c}}^{\circ} \cap \mathscr{R}_{\mathrm{Rx}} \setminus \mathscr{R}_{\mathrm{Tx}}$.

The given classification rules are designed to emphasize two relevant functionalities that a reaction channel can exhibit. First, the directional information exchange between X and Y (feedforward and feedback), and second, modulation of either X or Y that does not exploit any information about the other. Disconnected reaction channels do not influence X and Y at all and are hence irrelevant to the communication system. The subclassifications within $\mathscr{R}_{\mathrm{c}}$ serve the practical interpretability. In Appendix D we additionally dissect the above rules in terms of binary "adjacency" matrices that indicate whether directed trails between $\mathfrak{R}_r$ and either $\mathscr{S}_X$ or $\mathscr{S}_Y$ exist.

We are led by the following intuition: reaction channels for which no trail from $\mathfrak{R}_r$ to $\mathscr{S}_Y$ exists that avoids $\mathscr{S}_X$ should be modifiable without changing the channel $\mathbb{P}_{[0,t]}^{Y\|X}$. Only trails that carry influence from a reaction channel to Y without passing through X can alter how inputs are mapped to outputs. Using the matrix representations in Appendix D, one verifies that the following set collects exactly these reactions, and its decomposition makes the functional roles transparent.

Definition 18. *Consider the reversible species-reaction graph of the open CRN U. Reaction channels in the set*

$$\mathscr{R}_{\leftarrow} := \mathscr{R}_{\mathrm{d}} \cup [\mathscr{R}_{\mathrm{fb}} \setminus (\mathscr{R}_{\mathrm{ff}} \cup \mathscr{R}_{\mathrm{Rx}})] \cup \mathscr{R}_{\mathrm{Tx}}^{\circ}$$

are called encoding channels.

Thus far, we have taken a reaction-centered perspective, but the species-centered perspective is always implicit. For example, changing a chemostat concentration, changes the propensities of a set of reactions, and erasing a core species for the network corresponds to erasing all reactions in which it participates.

A degree of freedom in the optimization of $\sup_{X_{[0,t]}}$ that we did not address so far, is the initial distribution $p(u,0)$. At least for finite transmission duration t we can expect that a

well-chosen initial distribution can reduce the decoding error P_e for a given source encoding $\theta_{[0,t]}$. The initial values of certain species have an effect only on $X_{[0,t]}$, while others also have a direct effect on $\mathbb{P}^{Y\|X}_{[0,t]}$, e.g., because they are part of the species that implement feedforward or output modulation. Species that are directly connected to $\mathscr{S}_Y$ in the SR-graph can be identified with $\mathrm{pa}_0(\mathscr{R}^C_{\leftarrow})$, where we introduce "$(\dots)^C$" (in contrast to "$(\dots)^c$") as the notation for the complement on either of the sets $\mathscr{R}$ or $\mathscr{S}$. That is, $\mathscr{R}^C_{\leftarrow} = \mathscr{R} \setminus \mathscr{R}_{\leftarrow}$ and $\left[\mathrm{pa}_0(\mathscr{R}^C_{\leftarrow})\right]^C = \mathscr{S} \setminus \mathrm{pa}_0(\mathscr{R}^C_{\leftarrow})$ are the species that do not have a direct connection to $\mathscr{S}_Y$. Since the Janossy density of $(X,Y)_{[0,t]}$, represented as a function of trajectories, admits the factorization

$$\begin{aligned}\mathscr{P}^{X,Y}_{[0,t]}(x_{[0,t]},y_{[0,t]}) =& \mathbb{P}\left(X(0)=x(0), Y(0)=y(0)\right) \\ &\times \mathscr{P}^{Y\|X}_{[0,t]}(x_{[0,t]},y_{[0,t]})\ \mathscr{P}^{X\|Y}_{[0,t]}(x_{[0,t]},y_{[0,t]}),\end{aligned} \tag{65}$$

the channel law is independent of the choice of initial distribution of the species $\mathscr{S}_X \cup \mathscr{S}_Y$.

Definition 19. *Consider the reversible species-reaction graph of the open CRN U. The species in the set*

$$\mathscr{S}_{\leftarrow} := \left[\mathrm{pa}_0(\mathscr{R}^C_{\leftarrow})\right]^C \cup \mathscr{S}_X \cup \mathscr{S}_Y,$$

are called precoding (or, reset-encoding) species, where the complements are taken on $\mathscr{R}$ and $\mathscr{S}$, respectively.

The following theorem asserts, that the intuition about $\mathscr{R}_{\leftarrow}$ and $\mathscr{S}_{\leftarrow}$ is correct for a particular class of systems.

Let $(M,\theta,U,\mathscr{S}_X,\mathscr{S}_Y)$ describe a causal chemical communication system, as just outlined, but with the extension that manipulation protocols Γ^{X}_m may affect not only $\mathscr{R}^X_{\circ}$, but all of $\mathscr{R}_{\leftarrow}$. Without restriction of generality, we assume that all propensity functions $\lambda^{\varepsilon}_r(u,m,t)$ have a non-empty support, i.e., they must admit positive values for some arguments. For all $r \in \mathscr{R}_{\leftarrow}$ let $\eta^{\varepsilon}_r(u,m,t)$ be alternative (reversible) reaction propensities, which may differ from the original $\lambda^{\varepsilon}_r(u,m,t)$ such that the stoichiometry of the reaction is unmodified. Either the reaction intensity is set to zero, i.e., $\eta^{\varepsilon}_r = 0$ for all $\varepsilon \in \{+,-\}$ for all arguments, which corresponds to an erasure of the reaction from the network. Else, the reaction remains active, i.e., $\eta^{\varepsilon}_r(u,m,t) > 0$ for some arguments. Then it needs to satisfy two properties:

(i) Reaction stoichiometry of $\mathfrak{R}_r$ is unchanged, i.e., for all $u \neq u'$ with $(u_d\colon \pi_d(\nu_r) \neq 0) = (u'_d\colon \pi_d(\nu_r) \neq 0)$

$$\eta^{\varepsilon}_r(u,m,t) = \eta^{\varepsilon}_r(u',m,t).$$

(ii) If (u,m,t) satisfies $\eta^{\varepsilon}_r(u,m,t) > 0$, then $\lambda^{\varepsilon}_r(u,m,t) > 0$. This particularly restricts possible variations of t-dependence, given m.

All other reactions keep their original propensities. Now denote $S_p := (U_d(0)\colon d \in \mathscr{S}_{\leftarrow})$ and $S_c := (U_d(0)\colon d \in \mathscr{S} \setminus \mathscr{S}_{\leftarrow})$, such that $U(0) = (S_p, S_c)$. The set of alternative propensities, together with a yet to be defined alternative initial distribution and an alternative message distribution, defines an alternative process on the probability space $(\Omega, \mathscr{F}, \breve{\mathbb{P}})$, such that the alternative path measure is given by $\breve{\mathbb{P}}^U_{[0,t]}$. We stipulate that $\breve{\mathbb{P}} \circ S_c^{-1} = \mathbb{P} \circ S_c^{-1}$ and allow variations of $\breve{\mathbb{P}}(S_p = \cdot \mid S_c = s_c)$ for any s_c, as long as any s_p with $\mathbb{P}(S_p = s_p \mid S_c = s_c) > 0$ also satisfies $\breve{\mathbb{P}}(S_p = s_p \mid S_c = s_c) > 0$.

Our specifications of the alternative propensities and allowed variations of the initial distribution ensure that the alternative path measure is absolutely continuous with respect to the original path measure. With the alternative probability measure, we define channel output intensities

$$\eta^Y_s(\varepsilon\nu_Y) := \breve{\mathbb{E}}\left[\lambda^Y(\varepsilon\nu_Y, U(s^-)) \middle| \mathscr{F}^{XY}_{s^-}\right].$$

for all $s \in [0,t]$ and all $\nu_Y \in \mathscr{V}_Y$, $\varepsilon \in \{+,-\}$.

Theorem 5. *Consider two CRNs that differ only in reaction propensities of encoding reactions $\mathscr{R}_{\leftarrow}$ and the conditional initial distribution of precoding species $\mathscr{S}_{\leftarrow}$, as described above. If $\mathscr{S}^C_{\leftarrow} = \emptyset$, $\mathrm{pa}(\mathscr{R}_{\leftarrow}) \cap \mathscr{R} \subseteq \mathscr{R}^Y \cup \mathscr{R}^X$ and there are no X- or Y-indistinguishable reactions, then*

$$\lambda^Y_s(\varepsilon\nu_Y) = \eta^Y_s(\varepsilon\nu_Y), \quad \breve{\mathbb{P}}\text{-a.s.}$$

for all $\nu_Y \in \mathscr{V}_Y$, $\varepsilon \in \{+,-\}$ and all $s \in [0,t]$. In particular, the causally conditional path measures satisfy

$$\breve{\mathbb{P}}^{Y\|X}_{[0,t]} = \mathbb{P}^{Y\|X}_{[0,t]}, \quad \breve{\mathbb{P}}\text{-a.s.}$$

A proof is provided in Supp. Sec. S1.6 C.

The equalities in Theorem 5 are stated $\breve{\mathbb{P}}$-a.s. To assert them also $\mathbb{P}$-a.s., an additional dominance condition is required: for every $r \in \mathscr{R}_{\leftarrow}$ and $\varepsilon \in \{+,-\}$, the alternative propensity η^{ε}_r must not vanish on the support of the corresponding original propensity λ^{ε}_r. Equivalently, this ensures $\mathbb{P} \ll \breve{\mathbb{P}}$ (in addition to the already guaranteed $\breve{\mathbb{P}} \ll \mathbb{P}$), i.e., mutual absolute continuity. This subtlety has little practical impact. One may view $(\theta, U, \mathbb{P})$ as a "super-network" that contains at least all reactions and species present under $(\theta, U, \breve{\mathbb{P}})$. Then every $\breve{\mathbb{P}}$-possible path is also $\mathbb{P}$-possible, so the set of events with positive probability may be larger under $\mathbb{P}$. What matters for the capacity problem is that, on every event to which both measures assign positive probability, the channel law coincides, and thus the operational conclusion remains unchanged. However, Theorem 5 does not, by itself, certify that changes in $\mathscr{R}_{\leftarrow}$ exhaust the design space relevant to $\sup_{X_{[0,t]}}$. A completeness statement would require a converse: any admissible change of propensities for reactions in $\mathscr{R}^C_{\leftarrow}$ alters the channel law $\mathbb{P}^{Y|X}_{[0,t]}$ on a set of positive measure. This appears natural in light of the graphical intuition developed above, but establishing the converse lies beyond the scope of this work.

Finally, while we did neither stipulate a fully reversible CRN nor elementary reactions in the definition of the reversible SR-graph, we finally assess the consequences of thermodyamically consistent modeling on the set of encoding channels.

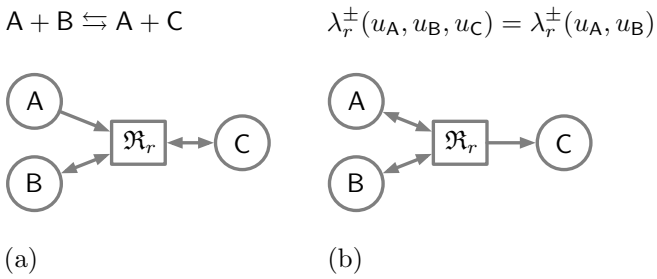


FIG. 3. Species-Reaction subgraphs depicting distinct types of reaction channels that ensure unidirectional information flow. (a) Example of an effective enzymatic reaction, where A modulates the propensities, but is not changed. (b) Example of a (elementary) reaction with bidirectional zero-order propensity w.r.t. species C.

Theorem 6. *For any causal chemical communication model that is based on a reversible CRN it holds that*

$$\mathscr{R}_{\mathrm{ff}} = \mathscr{R}_{\mathrm{fb}}$$

if all reactions are elementary (and physically meaningful).

In particular,

$$\mathscr{R}_{\leftarrow} \setminus \mathscr{R}_{\mathrm{d}} = \mathscr{R}^{\circ}_{\mathrm{Tx}}.$$

Proof. The necessary condition for a reversible reactions channel $\mathfrak{R}_r$ to have unidirectional information flow is that it follows either of the structural assumptions depicted in Fig. 3. However, any effective enzymatic reaction is a non-elementary approximation of at least two elementary reactions. Hence, we can focus on the remaining case of bidirectional zero-order propensities in some species d with $\pi_d(\nu_r) \neq 0$. The case of an elementary one-directional zero-order reaction, where the forward reaction does not depend on d, but the reverse reaction does, may arise from exchanges with a reservoir. That is, however, not a bidirectional zero-order case, which we deem not physically meaningful. In conclusion, unidirectional information flow does not happen in a reversible CRN consisting of elementary, physically meaningful reactions. Then, $\mathscr{R}_{\mathrm{ff}} = \mathscr{R}_{\mathrm{fb}}$ follows immediately and hence implies

$$\mathscr{R}_{\leftarrow} = \mathscr{R}_{\mathrm{d}} \sqcup \mathscr{R}^{\circ}_{\mathrm{Tx}}.$$

□

A major consequence is that variation of feedback is not a degree of freedom that can be used to increase the channel capacity of a thermodynamically consistent chemical communication model.

B. Information flow as an optimization principle

Using information flow in chemical reaction networks as an optimization objective function yields proposals for the properties of optimized information processing structures or their inputs. These proposals can be used as testable predictions to evaluate the hypothesis that information flow is an emergent optimization objective of evolution. Alternatively, they can be used as principled design suggestions for engineering chemical reaction networks optimized for information transmission.

In this work, we introduce trajectory-level information measures between subnetworks that extend earlier steady-state ensemble-level metrics [103] and Gaussian channel approximations [104]. By formulating the information measures on the paths probabilities of the underlying molecular dynamics, the framework captures the full temporal structure of the process. Thermodynamically consistent modeling enables the replacement of previous ad-hoc cost functions, which constrained mean molecular counts or their variances [103, 104], with energetic bounds derived from first principles. Taken together, these extensions allow the paradigm to be explored at a finer molecular-level and temporal resolution while remaining anchored in fundamental physical laws. Further, directed information and causally conditioned path measures add a notion of causal directionality to the mathematical formulation of the optimization principle, distinguishing causal forward information flow from causal feedback.

The paradigm of information optimization in biological systems has largely been discussed using mutual information, often motivated by the argument that mutual information is a bias-free measure of statistical dependence and thus serves as a minimal requirement for the presence of information flow [6]. While this rationale supports mutual information as a suitable objective function, it does not capture the causal structure present in many information-processing tasks. Earlier work introduced predictive information as a trajectory-level measure quantifying how much the past output dynamics reveal about future input dynamics [105]. Although not always framed in explicitly directional terms, this formulation inherently distinguishes the predicted quantity (future input) from the given information (past output), implying a temporal and functional asymmetry. Directed information formalizes such asymmetries by explicitly accounting for causal direction, and thus serves as a natural extension of mutual information when distinguishing forward from feedback pathways is essential.

Recalling Proposition 3 and adopting the previously discussed communication system perspective, the directed information $\mathbb{I}(X_{[0,t]} \to Y_{[0,t]})$ can be interpreted as an upper bound, and thus a proxy, for the information flow from an inaccessible external stimulus $\theta_{[0,t]}$ to the output $Y_{[0,t]}$, effectively excluding feedback contributions that would otherwise be included in the mutual information $\mathbb{I}(X_{[0,t]}; Y_{[0,t]})$.

Moreover, Definition 18 of encoding channels can be fruitfully applied in the information-optimization paradigm. It allows specific subsets of the reaction network to be held invariant, enabling the targeted design or analysis of other network components.

Recall that the Janossy density, represented as a function of trajectories, admits the factorization in Eq. (65). Assuming a fixed initial distribution, then implies that fixing $\mathbb{P}^{Y\|X}_{[0,t]}$ leaves $\mathbb{P}^{X\|Y}_{[0,t]}$ as degrees of freedom for optimization, and vice versa. Consequently, optimizing over the causally conditioned dis-

tribution or the marginal path distribution is equivalent if only one of the causally conditional probabilities is held fixed. In this case, Theorem 5 applies equivalently, when $\mathbb{P}^{X\|Y}_{[0,t]}$ is held invariant the encoding channels are replaced with

$$\mathscr{R}_{\rightarrow} := \mathscr{R}_{\mathrm{d}} \cup [\mathscr{R}_{\mathrm{ff}} \setminus (\mathscr{R}_{\mathrm{fb}} \cup \mathscr{R}_{\mathrm{Tx}})] \cup \mathscr{R}^{\circ}_{\mathrm{Rx}}.$$

If, however, the marginal law $\mathbb{P}^{X}_{[0,t]}$ is fixed, the situation becomes more difficult. Then, we must ensure that every change in the $Y_{[0,t]}$-dynamics is either "compensated for" or "not fed back into" the $X_{[0,t]}$-dynamics. If no change in Y is fed back to X, then $\mathbb{P}^{X}_{[0,t]} = \mathbb{P}^{X\|Y}_{[0,t]}$, which maps the optimization problem to the one, where the causal conditional probability is fixed. The case where compensation is required is beyond the scope of this work.

Bialek [93] has summarized two mathematical approaches to test the hypothesis that evolutionary processes optimize information flow: (1) distribution matching and (2) network architecture. We extend these frameworks to the setting of path-level directed information subject to thermodynamic energy constraints.

In the distribution matching framework, the input distribution is optimized for a given channel. This can be interpreted as an inverse problem to evolutionary network optimization: given the channel $\mathbb{P}^{Y\|X}_{[0,t]}$, does the actual input distribution $\mathbb{P}^{X}_{[0,t]}$ or the feedback-encoding distribution $\mathbb{P}^{X\|Y}_{[0,t]}$ maximize forward information flow?

To eliminate the degree of freedom associated with the initial distribution, we assume fixed initial conditions $X(0) = x(0) \in \mathscr{X}$ and $Y(0) = y(0) \in \mathscr{Y}$. The two equivalent formulations of the distribution matching problem are then

$$\begin{aligned}
\check{\mathbb{P}}^{X\|Y}_{[0,t]} &= \underset{X_{[0,t]}\|Y_{[0,t]}}{\arg\sup}\ \mathbb{I}(X_{[0,t]} \to Y_{[0,t]}),\\
\check{\mathbb{P}}^{X}_{[0,t]} &= \underset{X_{[0,t]}}{\arg\sup}\ \mathbb{I}(X_{[0,t]} \to Y_{[0,t]}),\\
\text{s. t.}\quad & \frac{1}{t}\mathbb{E}\left[q(t)\right] \le Q_0,
\end{aligned}$$

if the optimal distribution exists. Otherwise the solution can be represented as a sequence of distributions achieving the optimum asymptotically. The supremum is taken under any physically principled constraint of the form (61), which can involve the energetic contributions of the entire network or only a subset.

The optimization involves modifications such rescaling or removal of encoding reactions $\mathscr{R}_{\leftarrow}$, consistent with Theorem 5. Since the path measures are fully determined by the network's propensity functions, optimization may be carried out directly over these function spaces. In this way, the distribution matching framework becomes one of "matching propensity functions (and possibly network topology) under energetic constraints". If $X_{[0,t]}$ is causally independent of $Y_{[0,t]}$, that is, $\mathbb{I}(Y_{[0,t]} \to X_{[0,t]}) = 0$, then mutual information equals directed information under fixed initial conditions. Similarly, if X models an external signal such as DNA-damaging photon arrival statistics, rather than a part of the CRN, then energetic constraints on $X_{[0,t]}$ are irrelevant, and kinetic constraints may be more appropriate.

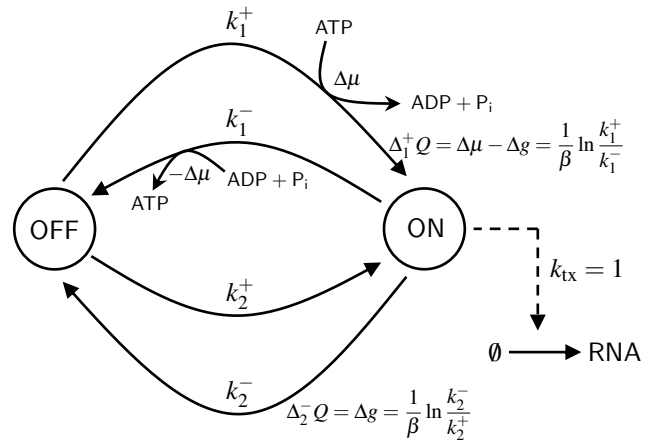


FIG. 4. State transition diagram of a two-state promoter model with two microscopically reversible reaction channels. The first reaction channel involves ATP hydrolysis and is therefore coupled to the chemostats of ATP, ADP and $\mathsf{P_i}$.

In contrast, the network architecture framework keeps the input process, or feedback encoding, fixed and optimizes the kinetic parameters and topology of the channel. This corresponds more directly to the evolutionary hypothesis, wherein the internal network evolves to optimally process a given environmental input. Again assuming fixed initial conditions and fixing $\mathbb{P}^{X\|Y}_{[0,t]}$ (not $\mathbb{P}^{X}_{[0,t]}$), we consider the problems

$$\begin{aligned}
\check{\mathbb{P}}^{Y\|X}_{[0,t]} &= \underset{Y_{[0,t]}\|X_{[0,t]}}{\arg\sup}\ \mathbb{I}(X_{[0,t]} \to Y_{[0,t]}),\\
\check{\mathbb{P}}^{Y}_{[0,t]} &= \underset{Y_{[0,t]}}{\arg\sup}\ \mathbb{I}(X_{[0,t]} \to Y_{[0,t]}),\\
\text{s. t.}\quad & \frac{1}{t}\mathbb{E}\left[q(t)\right] \le Q_0,
\end{aligned}$$

where again the supremum may only be achieved asymptotically by a sequence of distributions. The resulting channel process, defined by its topology and propensity functions, yields a principled design proposal under varying energy budgets, or a testable prediction in support of the information-optimization hypothesis.

Both frameworks can also be formulated in terms of the information rate (6). This information rate perspective may even be more appropriate for the discussed application, as it eliminates the effect of an arbitrarily chosen finite time-window. In the following, we illustrate this optimization framework in two small-scale, analytically tractable examples, where the capacity of biochemical channels is evaluated under energetic constraints using promoter-level models of transcriptional regulation.

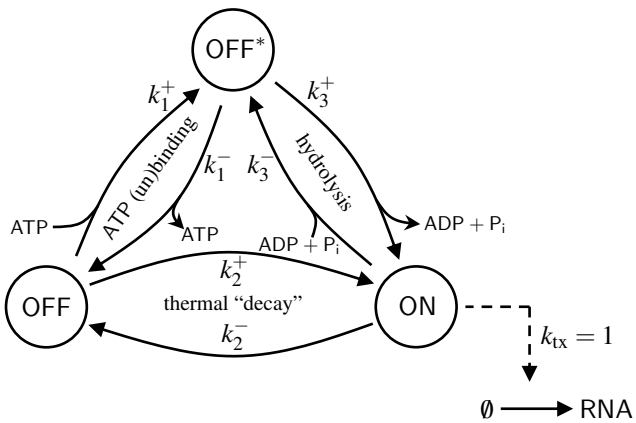


FIG. 5. State transition diagram of a three-state promoter model with three microscopically reversible reaction channels. The upper path from OFF to ON involves ATP hydrolysis. Hence, OFF* is the intermediate of a non-elementary hydrolysis reaction.

C. Case studies: Accurate energy-accounting and information-dissipation trade-off in small promoter models

We now apply the optimization principle to two analytically tractable promoter models. In each model the promoter state trajectory acts as the channel input X and the mRNA transcription record as the output Y. Our task is to maximize the directed information rate $\bar{\mathbb{I}}(X \to Y)$ subject to an upper bound on the promoter's entropy-production rate e_p. Transcription events are excluded from the thermodynamic budget because (i) their energetic cost is invariant for a fixed gene and does not affect the optimal promoter design, and (ii) the transcription events are not modeled thermodynamically consistent for simplicity and tractability. The resulting constrained problems are solved with `fmincon` (MATLAB R2022a), providing quantitative examples of the distribution-matching framework in which the input kinetics are tuned to exploit the full dynamic range of the transcriptional output [93].

Consider a two-state Markov jump process with states OFF and ON, which modulate transcriptional activity by switching between the transcription rates 0 and k_{tx}. For simplicity, we set $k_{\text{tx}} = 1$ in our analysis, thereby fixing the system's time scale. The state transition diagram of this model, along with transition rates and thermodynamic assumptions, is shown in Fig. 4. The model features a pair of parallel reaction channels: the first channel involves ATP hydrolysis with $\Delta\mu = \mu_{\text{ADP+P}_\text{i}} - \mu_{\text{ATP}} < 0$ and is coupled to corresponding chemostats, while the second reaction channel is thermally "driven" and not coupled to a chemostat. We denote the Gibbs free energy difference between the states by $\Delta g = g(\text{ON}) - g(\text{OFF})$. The entropy production rate at the NESS is

$$e_p = J \ln\left(\frac{k_1^+ k_2^-}{k_2^+ k_1^-}\right) \tag{66}$$

with net probability flux $J = \frac{k_1^+ k_2^- - k_2^+ k_1^-}{k_1^+ + k_1^- + k_2^+ + k_2^-}$, as we show in appendix F. This model provides a minimal example of the underestimation of the true entropy production by the effective entropy production (cf. Section IV F), as the effective two-state model always satisfies detailed balance at stationarity, such that its effective entropy production rate vanishes.

The three-state promoter model in Figure 5 extends the two-state model by adding an intermediate state, OFF*, which represents a step in a non-elementary reaction involving ATP hydrolysis. In this model, the first reaction channel handles ATP binding and unbinding between OFF and OFF*, while the second channel accounts for the actual ATP hydrolysis, driving the transition from OFF* to ON. The third reaction channel allows direct transitions between OFF and ON and is not coupled to any chemostat species. Distinguishing ATP binding/unbinding and hydrolysis may thereby be a more accurate account of the elementary reaction assumption than the single driven channel in the two-state model. Thermodynamically, the three-state promoter is characterized by

$$\begin{aligned}
\Delta_3^+ Q =& -\mu_{\text{ADP+P}_\text{i}} - \Delta_3^+ g = \frac{1}{\beta}\ln\frac{k_3^+}{k_3^-},\\
\Delta g :=& \Delta_2^+ g = \Delta_1^+ g + \Delta_3^+ g \text{ and}\\
\Delta_2^- Q =& \Delta g = \frac{1}{\beta}\ln\frac{k_2^-}{k_2^+}.
\end{aligned}$$

The EPR satisfies

$$e_p = J \ln\left(\frac{k_1^+ k_3^+ k_2^-}{k_1^- k_3^- k_2^+}\right)$$

with $J = k_2^- \pi(\text{ON}) - k_2^+ \pi(\text{OFF})$, which is related to the general form of the EPR for circular state transition diagrams [81].

We now formulate the maximization problem and use the (information) capacity notation for the maximum DIR. For the two-state promoter we compute

$$C_{\mathbb{I}}(\varepsilon_0) = \sup_{\substack{0 \le \bar{k}^+, \bar{k}^- \\ 0 \le e_p(\bar{k}^+, \bar{k}^-) \le \varepsilon_0}} \bar{\mathbb{I}}(X \to Y) \tag{67}$$

where $\bar{k}^+ = k_1^+ + k_2^+$, $\bar{k}^- = k_1^- + k_2^-$ are the effective rates of the transitions between OFF and ON.

The physical boundary conditions $-\Delta\mu, k_1^+, k_2^- > 0$ constrain the admissible region of the effective rates:

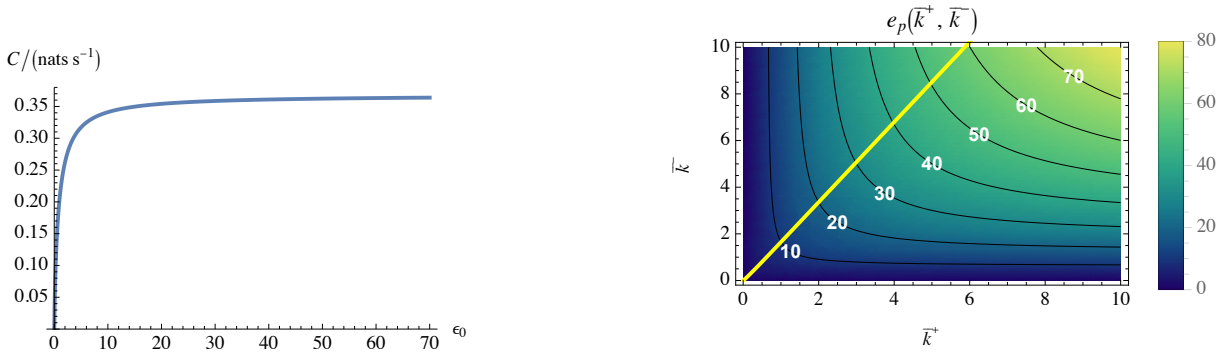
$$\bar{k}^+ > \bar{k}^- e^{-\beta\Delta g}, \quad \bar{k}^- > \bar{k}^+ e^{\beta(\Delta g + \Delta\mu)}$$

For this admissible region the EPR is thus expressed in terms of the effective rates via

$$e_p(\bar{k}^+, \bar{k}^-) = -\beta\Delta\mu \frac{\left(\bar{k}^+ - \bar{k}^- e^{-\beta\Delta g}\right)\left(\bar{k}^- e^{-\beta\Delta\mu} - \bar{k}^+ e^{-\beta\Delta g}\right)}{\left(\bar{k}^+ + \bar{k}^-\right)\left(e^{-\beta\Delta\mu} - 1\right)}, \tag{68}$$

as we show in Appendix F.

The DIR for the model with a two-state promoter has been



(a) Maximum DIR for fixed $-\Delta g/\Delta\mu = 0.5$ as a function of the upper bound $\varepsilon_0 \geq e_p$.

(b) Density plot of e_p together with its level sets (black lines) and the DIR–maximizing rates on the sets with $e_p \leq \varepsilon_0$ for varying $\varepsilon_0 > 0$ (yellow line).

FIG. 6. Maximum DIR for the two-state promoter model and fixed $-\beta\Delta\mu = 16$ and $-\Delta g/\Delta\mu = 0.5$.

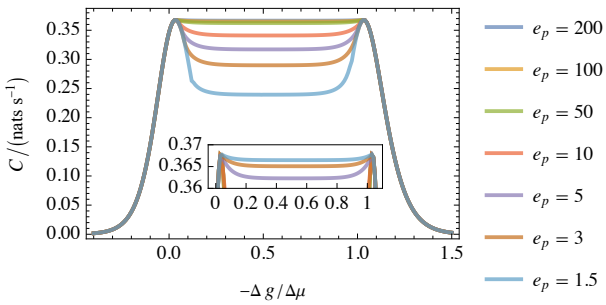


FIG. 7. Maximum DIR of the two-state promoter model with $-\beta\Delta\mu = 16$ as a function of $-\Delta g/\Delta\mu$ for different level sets $e_p = \varepsilon_0$.

shown to satisfy [39]

$$\begin{aligned}\bar{\mathbb{I}}(X \to Y) =& \frac{1}{\mathsf{E}[\tau]}(\ln(k_{\text{tx}}) + h(\tau) - 1)\\ =& k_{\text{tx}}\pi(\mathsf{ON})(\ln(k_{\text{tx}}) + h(\tau) - 1)\\ =& k_{\text{tx}}\frac{h(\tilde{\tau}) - 1}{\mathsf{E}[\tilde{\tau}]},\end{aligned}$$

where $h(\tau)$ is the differential entropy of the random time τ between transcription events in the dynamics of Y. $\tilde{\tau} = k_{\text{tx}}\tau$ is its dimensionless version. The probability density of τ is

$$p(\tau = t) = \frac{k_{\text{tx}} - w_1}{w_2 - w_1} w_2 e^{-w_2 t} + \frac{w_2 - k_{\text{tx}}}{w_2 - w_1} w_1 e^{-w_1 t},$$

which is an affine combination of exponentials and where $w_1, w_2 > 0$ are the roots of $w \mapsto (\bar{k}^- + k_{\text{tx}} - w)(\bar{k}^+ - w) - \bar{k}^+\bar{k}^-$ with $w_1 < w_2$.

Figure 6 presents the numerical solution of (67) for a fixed value of $-\beta\Delta\mu = 16$, corresponding to the chemical work done on the system via ATP hydrolysis [106], and a fixed ratio of $-\Delta g/\Delta\mu = 0.5$. Thereby, fixing $\Delta\mu$ corresponds to fixing the ratio of concentrations $\frac{[\mathsf{ATP}]}{[\mathsf{ADP}][\mathsf{P_i}]}$ and fixing Δg essentially fixes the identities of core species that participate in the promoter switching. The only degree of freedom left for practical parameter tuning is then the absolute concentration $[\mathsf{ATP}]$, keeping the ratio fixed. In Figure 6a the information capacity is evaluated for different $\varepsilon_0 > 0$, given that $-\Delta g/\Delta\mu = 0.5$. Complementarily, Figure 6b displays a density plot of the EPR e_p in the $(\bar{k}^+, \bar{k}^-)$-plane together with (i) its hyperbolic level sets and (ii) the curve of DIR maximizing effective rates for varying $\varepsilon_0 > 0$. The density plot illustrates that upper bounding the EPR does not upper bound the kinetic rates.

In contrast, Figure 7 shows the information capacity for varying $-\Delta g/\Delta\mu$ and different level sets of e_p. Hence, we do not only vary the absolute ATP concentration, but also explore the maximum for a continuous range of possible core species identities. In the range $-\Delta g/\Delta\mu \in [0,1]$ the maximum DIR increases monotonously with increasing e_p.

For $-\Delta g/\Delta\mu \notin [0,1]$ the level set $e_p = 0$ in the $(\bar{k}^+, \bar{k}^-)$-plane always intersects with the DIR maximizing ratio $\bar{k}^-/\bar{k}^+ \approx e - 1$ [37]. However, $-\Delta g/\Delta\mu < 0$ can be regarded as implausible since ON would then have less Gibbs free energy than OFF. $-\Delta g/\Delta\mu > 1$ on the other hand corresponds to $\Delta_1^+ Q < 0$, i.e., a chemically driven reaction, which additionally absorbs heat from the environment. Both scenarios are biophysically implausible.

For the three-state promoter let X represent the subnetwork $\{\mathsf{ON}\}$, i.e., the switching dynamics between a transcriptionally active and any inactive promoter state $\{\mathsf{OFF}, \mathsf{OFF}^*\}$. We define the capacity problem as

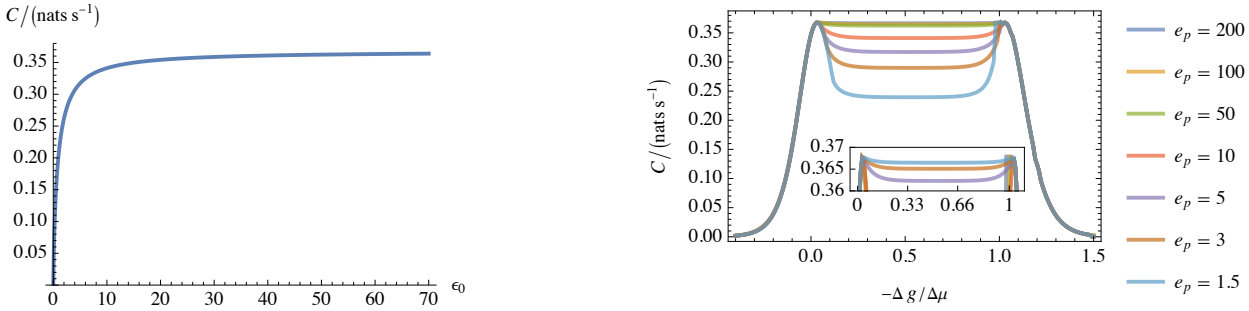
$$C_{\mathbb{I}}(\varepsilon_0) = \sup_{\substack{0 \leq k_1^+, k_1^-, k_2^-, k_3^+ \\ 0 \leq e_p \leq \varepsilon_0}} \bar{\mathbb{I}}(X \to Y),$$

where Δg, $\Delta\mu$ and k_{tx} are fixed. The DIR of this model can be shown to have the form [39]

$$\bar{\mathbb{I}}(X \to Y) = k_{\text{tx}} \ln(k_{\text{tx}})\pi(\mathsf{ON}) + \frac{h(\tau) - 1}{\mathsf{E}[\tau]}. \tag{69}$$

The corresponding information capacity curves for $k_{\text{tx}} = 1$ in Figure 8 are analogous to the curves of the two-state promoter.

In the light of information flow as an optimization principle, the two systems U (one with two-state promoter and the other



(a) Maximum DIR for fixed $-\Delta g/\Delta\mu = 0.5$ as a function of the upper bound $\varepsilon_0 \geq e_p$.

(b) Maximum DIR as a function of $-\Delta g/\Delta\mu$ for different level sets of $e_p = \varepsilon_0$.

FIG. 8. Maximum DIR for the three-state promoter model with fixed $-\beta\Delta\mu = 16$.

with three-state promoter) hence, have the same information-dissipation trade-off. Note that the channel $\mathbb{P}^{Y\|X}$ equals in the two models and the two optimizations (and their combination) can hence be regarded as small examples of the "distribution matching framework". The two distinct input topologies do not change the information capacity with respect to kinetic rate optimization under an entropy production rate constraint.

VII. CONCLUSION AND OUTLOOK

This work establishes a trajectory-level framework that links information processing and thermodynamic cost in Markovian chemical reaction networks (CRNs). On the information-theoretic side, we formulate mutual information and directed information between continuous-time trajectories and extend these notions to information exchange and causal information flow between arbitrary subnetworks, clarifying when mutual information is ill-defined. On the thermodynamic side, we provide a thermodynamically consistent account for open CRNs in Kurtz's process formulation and highlight that, at stationarity, the entropy production rate identifies the mean dissipation rate, while parallel reaction channels can invalidate common coarse-grained identifications if not modeled carefully.

Building on these foundations, we introduce, to the best of our knowledge, the first Shannon-style communication channel model for CRNs that explicitly separates message, encoding, and decoding and that permits principled energetic constraints on admissible codes. In this setting, the finite-time information capacity and its capacity–cost variants are formulated as optimizations over causal encoding laws subject to physical bounds (e.g., on dissipated heat or entropy production), and we provide a converse coding theorem under such constraints. A key structural contribution is the reaction-network classification that isolates design degrees of freedom from an invariant physical channel law. Modifying propensities within the class of encoding reactions preserves the causal channel kernel, whereas changes outside these classes generically alter it. Finally, analytically tractable promoter models illustrate how energetic constraints shape achievable information flow, yielding explicit information-dissipation trade-offs.

Several limitations delineate immediate directions for future work. First, we do not prove a full channel coding theorem for the CRN channel class; accordingly, the identification of information capacity with operational capacity is made under an information-stability assumption, and we only provide a converse bound. Establishing sufficient conditions for information stability in chemically relevant regimes remains open. Second, while the channel model allows time-dependent message protocols, thermodynamic constraints for general time-dependent driving are not available in closed form at the same level of generality. Obtaining them requires a microscopic derivation from the underlying Hamiltonian and is beyond the present scope. Third, the present results are exact at the level of elementary reaction networks. For full gene-expression networks, even a single transcriptional module may correspond to thousands of elementary reactions, and experimental knowledge at this resolution can be limited. An important challenge is therefore to make the framework accessible for non-elementary (coarse-grained) descriptions without losing control of the relevant information and energetic quantities.

Conceptually, we view the developed tools as a foundation for a reaction-network-based theory of signal processing: the principled design and analysis of biochemical signal encoding and processing under explicit causal, information-theoretic, and thermodynamic criteria. In this light, the reaction classification that enables a separation between "causal channel" and "causal (feedback) encoding" is best viewed as a tool for importing the classical noisy channel coding problem into the CRN setting. Thereby we have clarified what is meant by "communication through a CRN" and how this notion is invoked in an operational Shannon-sense, rather than a mere metaphorical one. We envision that, prospectively, our tools can serve as a starting point for progressing from the classical noisy channel coding problem, in which the channel is fixed, to more realistic biochemical coding problems. Natural next steps include extending the present methods to scenarios where parts of the "channel" are tunable (i.e., joint code-channel co-design) or performance metrics are goal-oriented, as well as developing controlled approximation schemes that bridge elementary and effective network descriptions.

ACKNOWLEDGMENTS

We thank Nicolai Engelmann for helpful discussions.

M.G. developed the methodology, led the formal analysis, contributed to conceptualization, and wrote the original draft; M.G., L.S., and H.K. reviewed and edited the manuscript. L.S. contributed to conceptualization. H.K. supervised the research and led the conceptualization.

The authors declare no competing financial interest.

DATA AVAILABILITY

The code used to perform the optimizations is available from the corresponding author upon reasonable request.

Appendix A: Chain rules for directed information

The following chain rules for the Newton DI are derived similarly to the original proofs given by Kramer [67] for the Massey DI.

Lemma 4. *For the random discrete sequences $X_{1:N}$, $Y_{1:N}$ and $Z_{1:N}$ the DI satisfies the chain rules*

$$\mathbb{I}(X_{1:N}, Z_{1:N} \to Y_{1:N}) = \mathbb{I}(X_{1:N} \to Y_{1:N}) + \mathbb{I}(Z_{1:N} \to Y_{1:N} \| X_{1:N}),$$

$$\begin{aligned}\mathbb{I}(X_{1:N} \to Y_{1:N}, Z_{1:N}) =& \mathbb{I}(X_{1:N} \to Y_{1:N} \| Z_{1:N}) \\ &+ \mathbb{I}\left(X_{1:N} \to Z_{1:N} \| D^{-1} Y_{1:N}\right),\end{aligned}$$

where D^{-1} denotes an anticipatory shift such that

$$\mathbb{I}\left(X_{1:N} \to Z_{1:N} \| D^{-1} Y_{1:N}\right) = \sum_{n=2}^{N} \mathbb{I}(X_{1:n-1}; Z_n \mid Y_{1:n}, Z_{1:n-1}).$$

Proof.

$$\begin{aligned}&\mathbb{I}(X_{1:N}, Z_{1:N} \to Y_{1:N}) = \sum_{n=2}^{N} \mathbb{I}(X_{1:n-1}, Z_{1:n-1}; Y_n \mid Y_{1:n-1}) \\ &= \sum_{n=2}^{N} \mathbb{I}(Z_{1:n-1}; Y_n \mid Y_{1:n-1}, X_{1:n-1}) + \mathbb{I}(X_{1:n-1}; Y_n \mid Y_{1:n-1}) \\ &= \mathbb{I}(Z_{1:N} \to Y_{1:N} \| X_{1:N}) + \mathbb{I}(X_{1:N} \to Y_{1:N})\end{aligned}$$

$$\mathbb{I}(X_{1:N} \to Y_{1:N}, Z_{1:N}) = \sum_{n=2}^{N} \mathbb{I}(X_{1:n-1}; Y_n, Z_n \mid Y_{1:n-1} Z_{1:n-1})$$

$$= \sum_{n=2}^{N} \mathbb{I}(X_{1:n-1}; Y_n \mid Y_{1:n-1}, Z_{1:n-1}) + \mathbb{I}(X_{1:n-1}; Z_n \mid Y_{1:n}, Z_{1:n-1})$$

$$= \mathbb{I}(X_{1:N} \to Y_{1:N} \| Z_{1:N}) + \mathbb{I}\left(X_{1:N} \to Z_{1:N} \| D^{-1} Y_{1:N}\right)$$

□

Appendix B: Stochastic filtering equations for the intensity processes

In order to compute the MI or DI from the expressions in the Theorems 1 and 2, we need to evaluate the intensity processes λ^X, $\hat{\lambda}^X$, λ^Y and $\hat{\lambda}^Y$. Computing the MI from similar expressions and approximations thereof has been discussed extensively in [35]. If, however, there are (X,Y)-, X- or Y-indistinguishable reactions, then this needs to be accounted for [69] (see also the discussion in Sec. III B). To improve accessibility, we provide the relevant expressions (Eq. (21) and (22) in [69]) in our notation.

We maintain all assumptions from Sec. III B. Let $\mathscr{S}_A \subset \mathscr{S}$ be an arbitrary subset of species indices of the core system U, and its complement $\mathscr{S}_B := \mathscr{S} \setminus \mathscr{S}_A$. The set $\mathscr{S}_A$ represents the set of species with respect to whom's history we want to filter the propensity functions in Eq (15). Hence, we choose $\mathscr{S}_A \in \{\mathscr{S}_X, \mathscr{S}_Y, \mathscr{S}_X \cup \mathscr{S}_Y\}$ in this context. Define $A := \{U_d(s) \colon d \in \mathscr{S}_A\}_{s \geq 0}$ and $B := \{U_d(s) \colon d \in \mathscr{S}_B\}_{s \geq 0}$, together with the natural filtration $\mathscr{F}^A$, as well as the coordinate maps π_A, π_B and the non-zero subnetwork change vectors $\mathscr{V}_A$, $\mathscr{V}_B$. In the following we substitute $U(s) = (A(s), B(s))$.

Exemplary, we want to evaluate the $\mathscr{F}^A$-intensity process of the reaction counter $R^X(\nu_X)$, $\nu_X \in \mathscr{V}_X$:

$$\begin{aligned}&\mathbb{E}\left[\lambda^X(\nu_X, U(s^-)) \middle| \mathscr{F}^A_{s-}\right] \\ =& \mathbb{E}\left[\lambda^X(\nu_X, A(s^-), B(s^-)) \middle| \mathscr{F}^A_{s-}\right] \\ =& \sum_b \Pi_b(s^-)\, \lambda^X(\nu_X, A(s^-), b)\end{aligned}$$

for all $s \in [0,t]$, where $\Pi_b(s)$ denotes the conditional probability mass function $s \mapsto \Pi_b(s) := \mathbb{P}\left(B(s) = b \middle| \mathscr{F}^A_s\right)$ as a function of time, which is actually a piecewise-deterministic stochastic process on the state space $[0,1]$ for all $b \in \mathbb{Z}_{\geq 0}^{|\mathscr{S}_B|}$. We refer to $\{\Pi_b(s), b \in \mathbb{Z}_{\geq 0}^{|\mathscr{S}_B|}\}$ as the filtering distribution at time $s \geq 0$. It follows the stochastic differential equation

$$
\begin{aligned}
\mathrm{d}\Pi_b(s) = &\left(-\lambda^A(0, A(s^-), b)\,\Pi_b(s^-) + \sum_{\nu \in \pi_A^{-1}(0) \cap \pi_B^{-1}(\mathscr{V}_B)} \lambda_{V^{-1}(\nu)}(A(s^-), b - \pi_B(\nu))\,\Pi_{b-\pi_B(\nu)}(s^-) \right) \mathrm{d}s \\
&- \sum_{\nu_A \in \mathscr{V}_A} \left(\lambda^A(\nu_A, A(s^-), b) - \hat{\lambda}_s^A(\nu_A) \right) \Pi_b(s^-)\,\mathrm{d}s \\
&+ \sum_{\nu_A \in \mathscr{V}_A} \left(\frac{\sum_{\nu \in \pi_A^{-1}(\nu_A)} \lambda_{V^{-1}(\nu)}(A(s^-), b - \pi_B(\nu))\,\Pi_{b-\pi_B(\nu)}(s^-)}{\hat{\lambda}_s^A(\nu_A)} - \Pi_b(s^-) \right) \mathrm{d}R_s^A(\nu_A)
\end{aligned}
$$

such that $\Pi_b(s) = \int_0^s \mathrm{d}\Pi_b(u)$, where

$$
\begin{aligned}
\lambda^A(\nu_A, a, b) &:= \sum_{\nu \in \pi_A^{-1}(\nu_A)} \lambda_{V^{-1}(\nu)}(a, b) \\
\hat{\lambda}_s^A(\nu_A) &:= \mathbb{E}\left[\lambda^A(\nu_A, A(s^-), B(s^-)) \middle| \mathscr{F}_{s-}^A \right] \\
&= \sum_b \lambda^A(\nu_A, A(s^-), b)\,\Pi_b(s^-)
\end{aligned}
$$

Importantly, the stochastic updates at jumps of the counters $R^A(\nu_A), \nu_A \in \mathscr{V}_A$ depend on the correct conditioning. Using the reaction counters $R_r, r \in \mathscr{R}^X$ (cf. Eq. (24)) instead would add the surplus information, which of some X-indistinguishable reaction has happened. In particular, the enumerator and denominator of the update equation satisfy

$$
\sum_{\nu \in \pi_A^{-1}(\nu_A)} \lambda_{V^{-1}(\nu)}(A(s^-), b - \pi_B(\nu))\,\Pi_{b-\pi_B(\nu)}(s^-)
$$

$$
= \sum_{r \in V^{-1}(\pi_A^{-1}(\nu_A))} \lambda_r(A(s^-), b - \pi_B(\nu_r))\,\Pi_{b-\pi_B(\nu_m)}(s^-)
$$

$$
\hat{\lambda}_s^A(\nu_A) = \sum_{r \in V^{-1}(\pi_A^{-1}(\nu_A))} \sum_b \lambda_r(A(s^-), b)\Pi_b(s^-)
$$

with $V^{-1}(\pi_A^{-1}(\nu_A)) \subseteq \mathscr{R}^X$ being a set of X-indistinguishable reaction indices. The continuous part of the evolution equation does not depend on the choice of conditioning, discussed here [69].

Appendix C: Transfer entropy between subnetworks and causally conditioned probabilities

While the primary focus of this paper is on mutual and directed information, we also provide an expression for the transfer entropy between subnetworks to facilitate comparison for an interdisciplinary readership. Following conventional notation, we refer to X as the target and Y as the source. To accommodate the definition of transfer entropy, we extend the definition of stochastic processes from the finite interval $[0,t]$ to the real line $\mathbb{R}$, such that the CRN is assumed to have started at time $-\infty$ rather than at time zero.

Consider the families of sigma-algebras $\{\mathscr{G}_s^{XY,(\alpha,\beta)}\}_{s\in\mathbb{R}}$ and $\{\mathscr{G}_s^{X,(\alpha)}\}_{s\in\mathbb{R}}$, defined by

$$
\mathscr{G}_s^{XY,(\alpha,\beta)} := \sigma\big(X(\tau), \tau \in [s-\alpha, s]\big) \vee \sigma\big(Y(\tau), \tau \in [s-\beta, s]\big),
$$

$$
\mathscr{G}_s^{X,(\alpha)} := \sigma\big(X(\tau), \tau \in [s-\alpha, s]\big),
$$

which represent the sliding-window partial histories of the joint process (X,Y) and the marginal process X, respectively. The parameters $\alpha, \beta > 0$ specify the history lengths for X and Y. Note that these families are generally not filtrations, since the fixed-length window definitions imply $\mathscr{G}_s^{XY,(\alpha,\beta)} \not\subset \mathscr{G}_{s'}^{XY,(\alpha,\beta)}$ and $\mathscr{G}_s^{X,(\alpha)} \not\subset \mathscr{G}_{s'}^{X,(\alpha)}$ for $s < s'$ in general. They become filtrations only in the limit $\alpha, \beta \to \infty$.

Spinney *et al.* [36] generalize the discrete-time definition of transfer entropy to arbitrary stochastic processes indexed by a strictly ordered and uncountable set $\mathbb{T}$. For continuous-time processes, they assume $\mathbb{T} \supseteq [t_0 - \max(\alpha,\beta), t)$ with $t_0 < t$. Their approach constructs a regular version of the probability measure

$$
\mathbb{P}\left(X_A \in \mathscr{A}_A \middle\| \{\mathscr{G}_s^{XY,(\alpha,\beta)}\}_{s\in A} \right) := \prod_{s\in A} \mathbb{P}\left(X(s) \in \mathscr{A}_s \middle| \mathscr{G}_{s-}^{XY,(\alpha,\beta)} \right) \tag{C1}
$$

for every finite subset $A \subset [t_0, t)$ and every product set $\mathscr{A}_A = \bigtimes_{s\in A} \mathscr{A}_s \in \Sigma_X^{\otimes A}$. This family of finite-dimensional distributions determines a unique probability measure on the path space over $[t_0, t)$ by Kolmogorov's Extension Theorem. The notation $\mathbb{P}(\cdot \| \cdot)$ is adopted from the "causally conditioned probability" as originally defined by Kramer [107]. Note that the definition in (C1) can be applied for arbitrary families of filtrations. Consequently, a causally conditioned probability for continuous-time systems, generalizing Kramers definition in the sense of Newton, is given by $\mathbb{P}^X_{[0,t)}\left(\cdot \middle\| \{\mathscr{F}_s^{XY}\}_{s\in[0,t)}\right)$ with short notation $\mathbb{P}^{X\|Y}_{[0,t)}$.

The transfer entropy is then defined measure-theoretically via the Radon-Nikodym derivative of this probability measure with respect to the conditional probability, i.e.,

$$
T_{y\to x}^{(\alpha,\beta)}\big|_{t_0}^{t} := \mathbb{E}\left[\ln \frac{\mathrm{d}\mathbb{P}^X_{[t_0,t)}\left(\cdot \middle\| \{\mathscr{G}_s^{XY,(\alpha,\beta)}\}_{s\in[t_0,t)}\right)}{\mathrm{d}\mathbb{P}^X_{[t_0,t)}\left(\cdot \middle\| \{\mathscr{G}_s^{X,(\alpha)}\}_{s\in[t_0,t)}\right)} \right]. \tag{C2}
$$

This framework also permits a pathwise, random-variable ver-

sion of transfer entropy, as is common in stochastic thermodynamics.

For comparison, we provide a related generalization of Schreiber's discrete-time transfer entropy [108] based on the structural, extremal-based approach of Weissman *et al.* [12]

$$T_{y\to x}^{(\alpha,\beta)}|_{t_0}^{t} := \sup_{\boldsymbol{t}\in\mathscr{T}(t_0,t)} \sum_{n=1}^{n(\boldsymbol{t})} \mathbb{I}\left(Y_{[t_{i-1}-\beta,t_{i-1})}; X_{[t_{i-1},t_i)} \mid X_{[t_{i-1}-\alpha,t_{i-1})}\right).$$

Here, we simply conjecture that the two definitions are equivalent for sufficiently regular path spaces. Using the same approach outlined in Supp. Sec. S1.3, one can show that the latter definition yields the same expression for the transfer entropy of jump processes as that derived by Spinney *et al.* [36].

Using the subnetwork propensities (15) we define the conditional intensity processes

$$\begin{aligned}\lambda_s^{X,(\alpha,\beta)}(\nu_X) =& \mathbb{E}\left[\lambda^X(\nu_X, U(s^-)) \mid \mathscr{G}_{s-}^{XY,(\alpha,\beta)}\right]\\ \hat{\lambda}_s^{X,(\alpha)}(\nu_X) =& \mathbb{E}\left[\lambda^X(\nu_X, U(s^-)) \mid \mathscr{G}_{s-}^{X,(\alpha)}\right].\end{aligned}$$

Then we obtain

$$\begin{aligned}T_{y\to x}^{(\alpha,\beta)}|_{t_0}^{t} =& \int_{t_0}^{t} \sum_{\nu_X\in\mathscr{V}_X} \mathbb{E}\left[\phi\left(\lambda_s^{X,(\alpha,\beta)}(\nu_X)\right)\right]\\ &- \mathbb{E}\left[\phi\left(\hat{\lambda}_s^{X,(\alpha)}(\nu_X)\right)\right] \mathrm{d}s.\end{aligned}$$

In the limit of full history this yields the relation

$$\lim_{\alpha,\beta\to\infty} T_{y\to x}^{(\alpha,\beta)}|_{t_0}^{t} = \mathbb{I}(Y_{(-\infty,t)} \to X_{[t_0,t)} \mid X_{(\infty,t_0)}).$$

There are two ways to make the transfer entropy equal to the DI. Either artificially set the path of U to a constant for all $s<0$ and randomly select $U(0)$. Then

$$\lim_{\alpha,\beta\to\infty} T_{y\to x}^{(\alpha,\beta)}|_{0}^{t} = \mathbb{I}(Y_{[0,t)} \to X_{[0,t)}).$$

Alternatively, the parameters α,β can be interpreted as time-dependent functions $\alpha(s),\beta(s)$. Setting $\alpha(s)=\beta(s)=s$ we obtain $\mathscr{G}_s^{XY,(s,s)} = \mathscr{F}_s^{XY}$ for all $s\geq 0$, such that

$$\mathbb{I}(Y_{[0,t)} \to X_{[0,t)}) = \int_0^t \dot{T}_{y\to x}^{(s,s)}(s)\, \mathrm{d}s.$$

Appendix D: Species reaction graph classification

To provide intuition for the reaction channel classification in Def. 17 we provide the following binary "adjacency" matrices on the set $\{\mathscr{S}_X, \mathfrak{R}_r, \mathscr{S}_Y\}$ that indicate whether directed trails between $\mathfrak{R}_r$ and either $\mathscr{S}_X$ or $\mathscr{S}_Y$ exist. Thereby, we ignore trails between species sets and cycles. In total, there are 16 distinct configurations. For example,

$$\begin{pmatrix} & 1 & \\ 0 & & 0 \\ & 0 & \end{pmatrix}$$

indicates that a trail from $\mathscr{S}_X$ to $\mathfrak{R}_r$ ($\mathscr{S}_X \to \mathfrak{R}_r$) exists, but neither a trail in the reverse direction nor any trail between $\mathfrak{R}_r$ and $\mathscr{S}_Y$ exists. Hence, for the given example, the connectivity via trails may be visualized as

$$\mathscr{S}_X \rightleftarrows \mathfrak{R}_r \not\rightleftarrows \mathscr{S}_Y,$$

where we use $\rightleftarrows$ instead of $\to$, since we reserve the latter arrowhead for the local (in)dependence notation.

Then, for categories (a) - (d), we obtain

$$\mathscr{R}_{\mathrm{ff}}: \quad \begin{pmatrix} & 1 & \\ 1 & & 1 \\ & 1 & \end{pmatrix}, \begin{pmatrix} & 1 & \\ 0 & & 1 \\ & 1 & \end{pmatrix}, \begin{pmatrix} & 1 & \\ 1 & & 1 \\ & 0 & \end{pmatrix}, \begin{pmatrix} & 1 & \\ 0 & & 1 \\ & 0 & \end{pmatrix}$$

$$\mathscr{R}_{\mathrm{fb}}: \quad \begin{pmatrix} & 1 & \\ 1 & & 1 \\ & 1 & \end{pmatrix}, \begin{pmatrix} & 0 & \\ 1 & & 1 \\ & 1 & \end{pmatrix}, \begin{pmatrix} & 1 & \\ 1 & & 0 \\ & 1 & \end{pmatrix}, \begin{pmatrix} & 0 & \\ 1 & & 0 \\ & 1 & \end{pmatrix}$$

$$\mathscr{R}_{\mathrm{c}}^{\circ}: \quad \begin{pmatrix} & 1 & \\ 1 & & 0 \\ & 0 & \end{pmatrix}, \begin{pmatrix} & 0 & \\ 0 & & 1 \\ & 1 & \end{pmatrix}, \begin{pmatrix} & 0 & \\ 1 & & 0 \\ & 0 & \end{pmatrix}, \begin{pmatrix} & 0 & \\ 0 & & 1 \\ & 0 & \end{pmatrix},$$

$$\begin{pmatrix} & 0 & \\ 1 & & 1 \\ & 0 & \end{pmatrix}$$

$$\mathscr{R}_{\mathrm{c}} \setminus \mathscr{R}_{\mathrm{c}}^{\circ}: \quad \overbrace{\begin{pmatrix} & 1 & \\ 1 & & 1 \\ & 0 & \end{pmatrix}}^{\in\mathscr{R}_{\mathrm{ff}}\setminus\mathscr{R}_{\mathrm{fb}}}, \overbrace{\begin{pmatrix} & 0 & \\ 1 & & 1 \\ & 1 & \end{pmatrix}}^{\in\mathscr{R}_{\mathrm{fb}}\setminus\mathscr{R}_{\mathrm{ff}}}$$

$$\mathscr{R}_{\mathrm{d}}: \quad \begin{pmatrix} & 0 & \\ 0 & & 0 \\ & 0 & \end{pmatrix}, \begin{pmatrix} & 1 & \\ 0 & & 0 \\ & 0 & \end{pmatrix}, \begin{pmatrix} & 0 & \\ 0 & & 0 \\ & 1 & \end{pmatrix}, \begin{pmatrix} & 1 & \\ 0 & & 0 \\ & 1 & \end{pmatrix}.$$

Note that $\mathscr{R} = \mathscr{R}_{\mathrm{ff}} \cup \mathscr{R}_{\mathrm{fb}} \cup \mathscr{R}_{\mathrm{c}}^{\circ} \cup \mathscr{R}_{\mathrm{d}}$. For the intersection between feedforward and feedback, i.e., $\mathscr{S}_X \rightleftarrows \mathfrak{R}_r \rightleftarrows \mathscr{S}_Y$, as well as the input and output modulating cases, we identify

$$\mathscr{R}_{\mathrm{ff}} \cap \mathscr{R}_{\mathrm{fb}}: \quad \begin{pmatrix} & 1 & \\ 1 & & 1 \\ & 1 & \end{pmatrix}$$

$$\mathscr{R}_{\mathrm{Tx}}: \quad \overbrace{\begin{pmatrix} & 1 & \\ 1 & & 0 \\ & 0 & \end{pmatrix} \begin{pmatrix} & 0 & \\ 1 & & 0 \\ & 0 & \end{pmatrix}}^{\mathscr{R}_{\mathrm{Tx}}^{\circ}}, \begin{pmatrix} & 0 & \\ 1 & & 1 \\ & 0 & \end{pmatrix}, \begin{pmatrix} & 1 & \\ 1 & & 1 \\ & 0 & \end{pmatrix}$$

$$\mathscr{R}_{\mathrm{Rx}}: \quad \overbrace{\begin{pmatrix} & 0 & \\ 0 & & 1 \\ & 1 & \end{pmatrix}, \begin{pmatrix} & 0 & \\ 0 & & 1 \\ & 0 & \end{pmatrix}}^{\mathscr{R}_{\mathrm{Rx}}^{\circ}}, \begin{pmatrix} & 0 & \\ 1 & & 1 \\ & 0 & \end{pmatrix}, \begin{pmatrix} & 0 & \\ 1 & & 1 \\ & 1 & \end{pmatrix}$$

$$\mathscr{R}_{\mathrm{fb}} \cap \mathscr{R}_{\mathrm{Rx}}: \quad \begin{pmatrix} & 0 & \\ 1 & & 1 \\ & 1 & \end{pmatrix}$$

$$\mathscr{R}_{\mathrm{ff}} \cap \mathscr{R}_{\mathrm{Tx}}: \quad \begin{pmatrix} & 1 & \\ 1 & & 1 \\ & 0 & \end{pmatrix}.$$

Appendix E: Formal definition of chemical channel codes

We may formalize the notions of the chemical channel coding problem. The key technical point in our setting is that a change of input intensities typically induces a change of the underlying path measure. Hence, admissible encoders should be defined as *pathwise* (measure-free) objects, and only subsequently be related to probability measures via a consistency requirement.

Throughout, fix a transmission duration $t > 0$ and write μ for Lebesgue measure on $[0,t]$. Let $\mathscr{D}([0,t], \mathscr{X} \times \mathscr{Y})$ denote the space of càdlàg trajectories equipped with its Borel σ-field. For $s \in [0,t]$, denote by r_s the restriction operator $\mathrm{r}_s(z) := z_{[0,s)}$. For a path $z \in \mathscr{D}([0,t],\cdot)$ let $\mathbb{J}(z) := \{s \in (0,t] : z(s) \neq z(s-)\}$ be its set of jump times. We write $L^1([0,t],[0,\infty))$ for integrable functions modulo μ-a.e. equality [109], and use the natural restriction map $\mathrm{r}_s : L^1([0,t],[0,\infty)) \to L^1([0,s),[0,\infty))$ induced by restricting representatives.

Let $\Pi \subseteq \mathscr{D}([0,t], \mathbb{R}^N_{\geq 0})$ be a set of admissible protocols, for some fixed $N \in \mathbb{N}$, viewed as deterministic control trajectories on $[0,t]$. We assume that every $\Gamma \in \Pi$ is bounded and piecewise differentiable, and has càdlàg sample paths on $[0,t]$. For $s \in [0,t]$ we write $\Pi|_{[0,s)} := \{\mathrm{r}_s(\Gamma) \mid \Gamma \in \Pi\}$.

A (feedback) channel encoder will be defined as an intensity-process map into $L^1([0,t],[0,\infty))$; this encodes exactly the degree of identification relevant for compensators and likelihoods (integration against μ), without reference to any ambient probability measure. Predictability and sample-path regularity are imposed deterministically via the existence of suitable representatives.

Definition 20. *A $(|\mathscr{M}|,t)$-code C for the message M and the causal chemical channel $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$ consists of the following objects.*

(a) For any protocol book Π a noiseless source code is a bijective mapping

$$\mathsf{c}: \mathscr{M} \to \Pi.$$

(b) A channel initialization is a probability mass function $p_0: \mathscr{U} \to [0,1]$.

(c) Let c be a source code with protocol book Π. A feedback channel encoder is a collection

$$\mathsf{f} := \big(\mathsf{f}_{\varepsilon,\nu_X}\big)_{\varepsilon \in \{+,-\},\, \nu_X \in \mathscr{V}_X}$$

of maps

$$\mathsf{f}_{\varepsilon,\nu_X}: \Pi \times \mathscr{D}([0,t], \mathscr{X} \times \mathscr{Y}) \to L^1([0,t],[0,\infty)).$$

These maps are required to be predictable in the following sense:
For every $s \in [0,t]$ and all $\Gamma, \Gamma' \in \Pi$ and $(x,y),(x',y') \in \mathscr{D}([0,t], \mathscr{X} \times \mathscr{Y})$,

$$\begin{aligned} &\mathrm{r}_s(\Gamma) = \mathrm{r}_s(\Gamma') \text{ and } \mathrm{r}_s(x,y) = \mathrm{r}_s(x',y') \\ &\implies \mathrm{r}_s\big(\mathsf{f}_{\varepsilon,\nu_X}(\Gamma,(x,y))\big) = \mathrm{r}_s\big(\mathsf{f}_{\varepsilon,\nu_X}(\Gamma',(x',y'))\big), \end{aligned}$$

as elements of $L^1([0,s),[0,\infty))$. Moreover, for each (ε,ν_X) there exists a representative family (a "predictable regenerative version")

$$\Lambda^X_{\varepsilon,\nu_X}: [0,t] \times \Pi \times \mathscr{D}([0,t], \mathscr{X} \times \mathscr{Y}) \to [0,\infty)$$

such that for all $(\Gamma,(x,y))$,

$$\Lambda^X_{\varepsilon,\nu_X}(\cdot,\Gamma,(x,y)) \in \mathsf{f}_{\varepsilon,\nu_X}(\Gamma,(x,y)),$$

where $s \mapsto \Lambda^X_{\varepsilon,\nu_X}(s,\Gamma,(x,y))$ is càglàd on $[0,t]$ and satisfies the (regenerative) jump-time constraint

$$\mathbb{J}\Big(\Lambda^X_{\varepsilon,\nu_X}(\cdot,\Gamma,(x,y))\Big) \subseteq \mathbb{J}(\Gamma) \cup \mathbb{J}(x) \cup \mathbb{J}(y).$$

(d) A feedback source–channel encoder is a collection

$$\mathsf{f} := \big(\mathsf{f}_{\varepsilon,\nu_X}\big)_{\varepsilon \in \{+,-\},\, \nu_X \in \mathscr{V}_X}$$

of maps

$$\mathsf{f}_{\varepsilon,\nu_X}: \mathscr{M} \times \mathscr{D}([0,t], \mathscr{X} \times \mathscr{Y}) \to L^1([0,t],[0,\infty)),$$

satisfying the analogous deterministic predictability condition (with $m \in \mathscr{M}$ in place of Γ) and admitting càglàd representatives $s \mapsto \Lambda^{M\to X}_{\varepsilon,\nu_X}(s,m,(x,y))$ whose jump times obey

$$\mathbb{J}\Big(\Lambda^{M\to X}_{\varepsilon,\nu_X}(\cdot,m,(x,y))\Big) \subseteq \mathbb{J}(x) \cup \mathbb{J}(y).$$

(e) A channel decoder is a mapping

$$D: \mathscr{D}([0,t], \mathscr{Y}) \to \mathscr{M} \cup \{0\}.$$

(f) A code $\mathsf{C} = (\mathsf{c}, p_0, \mathsf{f}, D)$, where f is a channel encoder, is called a split source-channel code. *A code $\mathsf{C} = (p_0, \mathsf{f}, D)$, where f is a source-channel encoder, is called a* source-channel code.

Note that the variation of $\mathscr{V}_X$ is explicitly allowed. However, the presented definition need not be consistent with all the requirements of the Markovian CRN $U \mid M$. This level of detail is lost due to projection and marginalization to (M,θ,X,Y). Therefore, while we have just defined the general functional class of codes, not all of those codes can necessarily be realized via a finite CRN. The question, which codes can be realized with such CRNs is addressed in Sec. VI A 5.

The following definition relates the code design to its induced probability law for the noisy channel coding problem.

Definition 21. *Let* C *be a split source-channel code for the chemical channel* $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$*. A probability measure* $\mathbb{P}^{\mathsf{C}}$ *on* $(\Omega, \mathscr{F})$ *is said to be* consistent *with* C *if:*

(i) $\mathbb{P}^{\mathsf{C}}(U(0) = u) = p_0(u)$ *for all* $u \in \mathscr{U}$*;*

(ii) the protocol-valued process θ *satisfies* $\theta = \mathsf{c}(M)$*;*

(iii) for each (ε, ν_X)*, the reaction-counting process* $R^X(\varepsilon\nu_X)$ *admits an* $\mathscr{F}^{\theta XY}$*- intensity process* $\lambda^X(\varepsilon\nu_X)$ *under* $\mathbb{P}^{\mathsf{C}}$*, which satisfies*

$$\lambda_s^X(\varepsilon\nu_X) = \Lambda^X_{\varepsilon,\nu_X}\big(s, \theta, (X,Y)_{[0,t]}\big), \qquad (\mu \otimes \mathbb{P}^{\mathsf{C}})\text{-a.e.}$$

on $([0,t] \times \Omega, \mathscr{P}(\mathscr{F}^{\theta XY}_{[0,t]}))$ *for the "predictable regenerative version"* $\Lambda^X_{\varepsilon,\nu_X}$ *of* $\mathsf{f}_{\varepsilon,\nu_X}$ *(as in Def. 20 (c));*

(iv) $\mathbb{P}^{\mathsf{C}} \ll \mathbb{P}$ *and* $\mathbb{P}^{\mathsf{C}Y\|X}_{[0,t]} = \mathbb{P}^{Y\|X}_{[0,t]}$*,* $\mathbb{P}^{\mathsf{C}}$*-a.s. (where* $\mathbb{P}$ *acts as a reference measure).*

Consistency for a source–channel code $\mathsf{C} = (p_0, \mathsf{f}, D)$ is defined analogously, with $\Lambda^{M\to X}_{\varepsilon,\nu_X}(s, M, (X,Y)_{[0,t]})$ in place of $\Lambda^X_{\varepsilon,\nu_X}(s, \theta, (X,Y)_{[0,t]})$ and $\mathscr{F}^{MXY}_{s-}$-predictability.

Definition 22. *The set* $\mathscr{C}$ *of all* $(|\mathscr{M}|, t)$*-codes* C *for which there exists at least one consistent measure* $\mathbb{P}^{\mathsf{C}}$ *is called the set of* $(|\mathscr{M}|, t)$*-codes admissible for the causal chemical channel* $(\mathscr{S}_X, \mathbb{P}^{Y\|X}, \mathscr{S}_Y)$*.*

Note that $\mathscr{C}$ is not necessarily a product set of admissible initializations, encoders and decoders.

Appendix F: EPR of a non-equilibrium two-state model

To derive the EPR for the two-state model we use Equation (47) for the entropy exchange rate, which is equal to the EPR at the NESS:

$$\begin{aligned}
e_p =& \pi(OFF)k_1^+ \ln\left(\frac{k_1^+}{k_1^-}\right) + \pi(ON)k_1^- \ln\left(\frac{k_1^-}{k_1^+}\right) \\
&+ \pi(OFF)k_2^+ \ln\left(\frac{k_2^+}{k_2^-}\right) + \pi(ON)k_2^- \ln\left(\frac{k_2^-}{k_2^+}\right) \\
=& \ln\left(\frac{k_1^+}{k_1^-}\right)(\pi(OFF)k_1^+ - \pi(ON)k_1^-) \\
&+ \ln\left(\frac{k_2^+}{k_2^-}\right)(\pi(OFF)k_2^+ - \pi(ON)k_2^-) \\
=& \ln\left(\frac{k_1^+}{k_1^-}\right)\frac{(k_1^- + k_2^-)k_1^+ - (k_1^+ + k_2^+)k_1^-}{k_1^+ + k_1^- + k_2^+ + k_2^-} \\
&+ \ln\left(\frac{k_2^+}{k_2^-}\right)\frac{(k_1^- + k_2^-)k_2^+ - (k_1^+ + k_2^+)k_2^-}{k_1^+ + k_1^- + k_2^+ + k_2^-}
\end{aligned}$$

with $\pi(OFF) = \frac{k_1^- + k_2^-}{k_1^+ + k_1^- + k_2^+ + k_2^-}$ and $\pi(ON) = \frac{k_1^+ + k_2^+}{k_1^+ + k_1^- + k_2^+ + k_2^-}$. The stationary distribution is the kernel of the generator matrix corresponding to the model. Simplifying the above expression yields (66).

To obtain the EPR as a function of the effective rates $\overline{k}^+ = k_1^+ + k_2^+$, $\overline{k}^- = k_1^- + k_2^-$, we we express the rates in the state transition diagram via the effective rates and the potential differences Δg, $\Delta\mu$. We have

$$\begin{aligned}
k_1^- &= k_1^+ e^{\beta(\Delta g + \Delta\mu)}, \quad k_2^+ = k_2^- e^{-\beta\Delta g}, \\
\overline{k}^- &= k_1^+ e^{\beta(\Delta g + \Delta\mu)} + k_2^-, \quad \overline{k}^+ = k_1^+ + k_2^- e^{-\beta\Delta g},
\end{aligned}$$

which implies

$$k_1^+ = \frac{\overline{k}^+ - \overline{k}^- e^{-\beta\Delta g}}{1 - e^{\Delta\mu}}, \quad k_2^- = \frac{\overline{k}^- - \overline{k}^+ e^{\beta(\Delta g + \Delta\mu)}}{1 - e^{\Delta\mu}}$$

and therefore also

$$\begin{aligned}
k_1^- &= \frac{\overline{k}^+ - \overline{k}^- e^{-\beta\Delta g}}{1 - e^{\Delta\mu}} e^{\beta(\Delta g + \Delta\mu)} \\
k_2^+ &= \frac{\overline{k}^- - \overline{k}^+ e^{\beta(\Delta g + \Delta\mu)}}{1 - e^{\Delta\mu}} e^{-\beta\Delta g}
\end{aligned}$$

The physical boundary conditions $-\Delta\mu, k_1^+, k_2^- > 0$ constrain the admissible region of the effective rates to

$$\overline{k}^+ > \overline{k}^- e^{-\beta\Delta g}, \quad \overline{k}^- > \overline{k}^+ e^{\beta(\Delta g + \Delta\mu)}.$$

For this region we obtain (68) by substituting the original rates with the above expressions and simplifying.

Appendix G: Interarrival time between transcription events for the circular three-state promoter

The DI between the promoter dynamics and the RNA copy number has was linked to the differential entropy of the interarrival time in (69). Here, we apply Anderson's filtering theorem for semi Markov processes similarly as described in [39]. Define the state space $E := \{\mathrm{J}, \mathrm{ON}, \mathrm{OFF}^*, \mathrm{OFF}\}$, where J is the jump state, which mimics ON in terms of leaving transitions. The Laplace version of the semi-Markov kernel density is

$$q^*(s) = \begin{pmatrix}
\frac{k_{\mathrm{tx}}}{u_{ON}+s} & 0 & \frac{k_3^-}{u_{ON}+s} & \frac{k_2^-}{u_{ON}+s} \\
\frac{k_{\mathrm{tx}}}{u_{ON}+s} & 0 & \frac{k_3^-}{u_{ON}+s} & \frac{k_2^-}{u_{ON}+s} \\
0 & \frac{k_3^+}{u_{OFF*}+s} & 0 & \frac{k_1^-}{u_{OFF*}+s} \\
0 & \frac{k_2^+}{u_{OFF}+s} & \frac{k_1^+}{u_{OFF}+s} & 0
\end{pmatrix},$$

where $u_{ON} = k_2^- + k_3^- + k_{\mathrm{tx}}$, $u_{OFF} = k_1^+ + k_2^+$ We filter E to the reduced state space $\{\mathrm{J}\}$ which yields the interarrival time distribution f_τ of the output. Since the geometric series of a 3×3-matrix is hard to evaluate, we do the filtering state by

state, from right to left. In the first step we identify

$$a^*(s) = \begin{pmatrix} \frac{k_{\text{tx}}}{u_{ON}+s} & 0 & \frac{k_3^-}{u_{ON}+s} \\ \frac{k_{\text{tx}}}{u_{ON}+s} & 0 & \frac{k_3^-}{u_{ON}+s} \\ 0 & \frac{k_3^+}{u_{OFF*}+s} & 0 \end{pmatrix}$$

$$b^*(s) = \left(\frac{k_2^-}{u_{ON}+s}, \frac{k_2^-}{u_{ON}+s}, \frac{k_1^-}{u_{OFF*}+s} \right)^T,$$

$$c^*(s) = \left(0, \frac{k_2^+}{u_{OFF}+s}, \frac{k_1^+}{u_{OFF}+s} \right) \text{ and}$$

$$d^* = 0.$$

Thus

$$b^*(s)\,c^*(s) = \begin{pmatrix} 0 & \frac{k_2^+k_2^-}{(u_{OFF}+s)(u_{ON}+s)} & \frac{k_1^+k_2^-}{(u_{OFF}+s)(u_{ON}+s)} \\ 0 & \frac{k_2^+k_2^-}{(u_{OFF}+s)(u_{ON}+s)} & \frac{k_1^+k_2^-}{(u_{OFF}+s)(u_{ON}+s)} \\ 0 & \frac{k_1^-k_2^+}{(u_{OFF*}+s)(u_{OFF}+s)} & \frac{k_1^+k_1^-}{(u_{OFF*}+s)(u_{OFF}+s)} \end{pmatrix}$$

and

$$\begin{aligned} \tilde{q}^*(s) =& a^*(s) + b^*(s)\,c^*(s) \\ =& \begin{pmatrix} \frac{k_{\text{tx}}}{u_{ON}+s} & \frac{k_2^+k_2^-}{(u_{OFF}+s)(u_{ON}+s)} & \frac{k_1^+k_2^- + k_3^-(u_{OFF}+s)}{(u_{OFF}+s)(u_{ON}+s)} \\ \frac{k_{\text{tx}}}{u_{ON}+s} & \frac{k_2^+k_2^-}{(u_{OFF}+s)(u_{ON}+s)} & \frac{k_1^+k_2^- + k_3^-(u_{OFF}+s)}{(u_{OFF}+s)(u_{ON}+s)} \\ 0 & \frac{k_1^-k_2^+ + k_3^+(u_{OFF}+s)}{(u_{OFF*}+s)(u_{OFF}+s)} & \frac{k_1^+k_1^-}{(u_{OFF*}+s)(u_{OFF}+s)} \end{pmatrix} \end{aligned}$$

In the second step we identify

$$\begin{aligned} a^*(s) =& \begin{pmatrix} \frac{k_{\text{tx}}}{u_{ON}+s} & \frac{k_2^+k_2^-}{(u_{OFF}+s)(u_{ON}+s)} \\ \frac{k_{\text{tx}}}{u_{ON}+s} & \frac{k_2^+k_2^-}{(u_{OFF}+s)(u_{ON}+s)} \end{pmatrix}, \\ b^*(s) =& \frac{k_3^-(r_2+s)}{(u_{OFF}+s)(u_{ON}+s)}\,(1,1)^T, \\ c^*(s) =& \frac{k_3^+(r_1+s)}{(u_{OFF*}+s)(u_{OFF}+s)}\,(0,1), \\ \sum_{k=0}^{\infty} (d^*(s))^k =& \frac{(u_{OFF*}+s)(u_{OFF}+s)}{(u_{OFF*}+s)(u_{OFF}+s) - k_1^+k_1^-} \end{aligned}$$

with $r_1 := \frac{k_1^-k_2^+}{k_3^+} + u_{OFF}$ and $r_2 := \frac{k_1^+k_2^-}{k_3^-} + u_{OFF}$. We denote the roots of $s \mapsto (u_{OFF*}-s)(u_{OFF}-s) - k_1^+k_1^-$ by w_1, w_2. Then

$$\begin{aligned} & b^*(s) \sum_{k=0}^{\infty} (d^*(s))^k\, c^*(s) \\ =& \frac{k_3^+k_3^-(r_1+s)(r_2+s)}{(u_{OFF}+s)(u_{ON}+s)(w_1+s)(w_2+s)} \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \\ & \tilde{\tilde{q}}^*(s) = a^*(s) + b^*(s) \sum_{k=0}^{\infty} (d^*(s))^k\, c^*(s) \\ =& \begin{pmatrix} \frac{k_{\text{tx}}}{u_{ON}+s} & \frac{k_3^+k_3^-(r_1+s)(r_2+s) + k_2^+k_2^-(w_1+s)(w_2+s)}{(u_{OFF}+s)(u_{ON}+s)(w_1+s)(w_2+s)} \\ \frac{k_{\text{tx}}}{u_{ON}+s} & \frac{k_3^+k_3^-(r_1+s)(r_2+s) + k_2^+k_2^-(w_1+s)(w_2+s)}{(u_{OFF}+s)(u_{ON}+s)(w_1+s)(w_2+s)} \end{pmatrix}. \end{aligned}$$

Denoting the roots of $\mathbb{C} \ni s \mapsto k_3^+k_3^-(r_1-s)(r_2-s) + k_2^+k_2^-(w_1-s)(w_2-s)$ by q_1, q_2 we rewrite

$$\tilde{\tilde{q}}^*(s) = \begin{pmatrix} \frac{k_{\text{tx}}}{u_{ON}+s} & \frac{(k_3^+k_3^- + k_2^+k_2^-)(q_1+s)(q_2+s)}{(u_{OFF}+s)(u_{ON}+s)(w_1+s)(w_2+s)} \\ \frac{k_{\text{tx}}}{u_{ON}+s} & \frac{(k_3^+k_3^- + k_2^+k_2^-)(q_1+s)(q_2+s)}{(u_{OFF}+s)(u_{ON}+s)(w_1+s)(w_2+s)} \end{pmatrix}.$$

For the last filtering step we use that $\frac{a}{b}\frac{1}{1-\frac{a}{b}} = \frac{a}{b-a}$ and $1 + \frac{a}{b-a} = \frac{b}{b-a}$ for any $a \neq b$. We obtain

$$\begin{aligned} f_\tau^*(s) =& \{k_{\text{tx}}(u_{OFF}+s)(w_1+s)(w_2+s)\} \\ & \times \{u_{OFF}+s)(u_{ON}+s)(w_1+s)(w_2+s) \\ & - (k_3^+k_3^- + k_2^+k_2^-)(q_1+s)(q_2+s)\}^{-1}. \end{aligned}$$

Denote $R(s) := k_{\text{tx}}(u_{OFF}+s)(w_1+s)(w_2+s)$ and let $v_1, \ldots, v_4$ be the roots of $U(s) := (u_{OFF}+s)(u_{ON}+s)(w_1+s)(w_2+s) - (k_3^+k_3^- + k_2^+k_2^-)(q_1+s)(q_2+s)$. Then

$$f_\tau(t) = \sum_{k=1}^{4} \frac{R(v_k)}{U'(v_k)} e^{v_k t}.$$

[1] M. Magaña-Acosta and V. Valadez-Graham, Frontiers in Genetics **11**, 600615 (2020).
[2] S. L. Johnson and G. J. Narlikar, Journal of Molecular Biology **434**, 167653 (2022).
[3] W. S. Klug, M. R. Cummings, C. A. Spencer, M. A. Palladino, and D. J. Killian, Concepts of genetics (2020).
[4] U. Alon, *An introduction to systems biology: design principles of biological circuits* (Chapman and Hall/CRC, 2019).
[5] G. Rieckh and G. Tkačik, Biophysical journal **106**, 1194 (2014).
[6] G. Tkačik and A. M. Walczak, Journal of Physics: Condensed Matter **23**, 153102 (2011).
[7] M. Behar and A. Hoffmann, Current opinion in genetics & development **20**, 684 (2010).
[8] J. E. Purvis and G. Lahav, Cell **152**, 945 (2013).
[9] D. Friedrich, L. Friedel, A. Finzel, A. Herrmann, S. Preibisch, and A. Loewer, Molecular Systems Biology **15**, e9068 (2019).
[10] S. Sasagawa, Y.-i. Ozaki, K. Fujita, and S. Kuroda, Nature cell biology **7**, 365 (2005).
[11] J. Massey *et al.*, in *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, Vol. 2 (1990).
[12] T. Weissman, Y.-H. Kim, and H. H. Permuter, IEEE Transactions on Information Theory **59**, 1271 (2012).
[13] N. J. Newton, arXiv preprint arXiv:1604.01969 (2016).
[14] A. S. Hansen and E. K. O'Shea, Elife **4**, e06559 (2015).
[15] J. J. Tyson and B. Novák, Annual review of physical chemistry **61**, 219 (2010).

[16] R. Chetrite, M. Rosinberg, T. Sagawa, and G. Tarjus, Journal of Statistical Mechanics: Theory and Experiment **2019**, 114002 (2019).
[17] P. R. Ten Wolde, Nature Physics **8**, 361 (2012).
[18] D. F. Anderson and T. G. Kurtz, in *Design and analysis of biomolecular circuits: engineering approaches to systems and synthetic biology* (Springer, 2011) pp. 3–42.
[19] C. E. Shannon, The Bell system technical journal **27**, 379 (1948).
[20] J. A. T. Thomas M. Cover, Channel capacity, in *Elements of Information Theory* (John Wiley & Sons, Ltd, 2005) Chap. 7, pp. 183–241.
[21] H. Marko, IEEE Transactions on communications **21**, 1345 (1973).
[22] Y.-H. Kim, IEEE Transactions on Information Theory **54**, 1488 (2008).
[23] Y. M. Kabanov, Theory of Probability & Its Applications **23**, 143 (1978).
[24] M. R. Frey, *Capacity of the Poisson communications channel*, Doctoral dissertation, The University of North Carolina at Chapel Hill (1990).
[25] S. Uda, Biophysical reviews **12**, 377 (2020).
[26] H. B. Barlow *et al.*, Sensory communication **1**, 217 (1961).
[27] C. Waltermann and E. Klipp, Biochimica et Biophysica Acta (BBA)-General Subjects **1810**, 924 (2011).
[28] I. Nemenman, Quantitative biology: from molecular to cellular systems **4**, 73 (2012).
[29] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, science **334**, 354 (2011).
[30] F. Tostevin and P. R. Ten Wolde, Physical review letters **102**, 218101 (2009).
[31] I. Lestas, G. Vinnicombe, and J. Paulsson, Nature **467**, 174 (2010).
[32] Y. Nakahira, F. Xiao, V. Kostina, and J. C. Doyle, in *2018 Annual American Control Conference (ACC)* (IEEE, 2018) pp. 2707–2714.
[33] K. V. Parag, Journal of theoretical biology **480**, 262 (2019).
[34] L. Duso and C. Zechner, in *2019 IEEE 58th Conference on Decision and Control (CDC)* (IEEE, 2019) pp. 6610–6615.
[35] A.-L. Moor and C. Zechner, Physical Review Research **5**, 013032 (2023).
[36] R. E. Spinney, M. Prokopenko, and J. T. Lizier, Physical Review E **95**, 032319 (2017).
[37] M. Sinzger, M. Gehri, and H. Koeppl, in *2020 IEEE International Symposium on Information Theory (ISIT)* (IEEE, 2020) pp. 2873–2878.
[38] M. Sinzger-D'Angelo and H. Koeppl, IEEE Transactions on Information Theory (2023).
[39] M. Gehri, N. Engelmann, and H. Koeppl, in *2024 IEEE International Symposium on Information Theory (ISIT)* (2024) pp. 1931–1936.
[40] D. T. Gillespie, The journal of physical chemistry **81**, 2340 (1977).
[41] M. Reinhardt, G. Tkačik, and P. R. Ten Wolde, Physical Review X **13**, 041017 (2023).
[42] A. Das and P. R. Ten Wolde, Physical Review Letters **135**, 107404 (2025).
[43] M. Davis, IEEE Transactions on Information Theory **26**, 710 (1980).
[44] P. J. Thomas and A. W. Eckford, IEEE Transactions on information Theory **62**, 7358 (2016).
[45] P. J. Thomas and A. W. Eckford, in *2016 IEEE International Symposium on Information Theory (ISIT)* (IEEE, 2016) pp. 1804–1808.
[46] A. W. Eckford and P. J. Thomas, IEEE Transactions on Molecular, Biological and Multi-Scale Communications **4**, 27 (2018).
[47] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu, Nature physics **8**, 422 (2012).
[48] P. Mehta and D. J. Schwab, Proceedings of the National Academy of Sciences **109**, 17978 (2012), https://www.pnas.org/doi/pdf/10.1073/pnas.1207814109.
[49] S. J. Bryant and B. B. Machta, Physical review letters **131**, 068401 (2023).
[50] F. Tasnim, N. Freitas, and D. H. Wolpert, Physical Review E **110**, 034101 (2024).
[51] A. Yadav and D. Wolpert, Minimal thermodynamic cost of communication (2025), arXiv:2410.14920 [cond-mat.stat-mech].
[52] F. Attneave, Psychological review **61**, 183 (1954).
[53] R. Linsker, Computer **21**, 105 (1988).
[54] G. Tkačik, C. G. Callan Jr, and W. Bialek, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **78**, 011910 (2008).
[55] R. Boel, P. Varaiya, and E. Wong, SIAM Journal on Control **13**, 1022 (1975).
[56] J. McFadden, Journal of the society for industrial and applied mathematics **13**, 988 (1965).
[57] P. M. Brémaud, *Point Processes and Queues: Martingale Dynamics*, Vol. 50 (Springer, 1981).
[58] T. Renner, E. R. Cohen, T. Cvitas, J. G. Frey, B. Holström, K. Kuchitsu, R. Marquardt, I. Mills, F. Pavese, M. Quack, J. Stohner, H. L. Strauss, M. Takami, and A. J. Thor, *Quantities, units and symbols in physical chemistry* (The Royal Society of Chemistry, 2007).
[59] International Union of Pure and Applied Chemistry (IUPAC), Elementary reaction, `https://doi.org/10.1351/goldbook.E02035` (2019), IUPAC Compendium of Chemical Terminology (Gold Book).
[60] R. Rao and M. Esposito, The Journal of Chemical Physics **149** (2018).
[61] T. G. Kurtz, The Journal of Chemical Physics **57**, 2976 (1972).
[62] H. Qian and H. Ge, *Stochastic Chemical Reaction Systems in Biology* (Springer International Publishing, 2021).
[63] T. Schmiedl and U. Seifert, The Journal of Chemical Physics **126** (2007).
[64] D. T. Gillespie, Physica A: Statistical Mechanics and its Applications **188**, 404 (1992).
[65] A. Klenke and A. Klenke, Probability Theory: A Comprehensive Course , 303 (2020).
[66] P.-O. Amblard and O. J. Michel, Entropy **15**, 113 (2012).
[67] G. Kramer, *Directed information for channels with feedback*, ETH Series in Information Processing, Vol. 11 (Hartung-Gorre Verlag, Konstanz, Switzerland, 1998).
[68] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods* (Springer Science & Business Media, 2006).
[69] L. Bronstein and H. Koeppl, Physical Review E **97**, 062147 (2018).
[70] L. Peliti and S. Pigolotti, *Stochastic thermodynamics: an introduction* (Princeton University Press, 2021).
[71] D. F. Anderson, G. Craciun, and T. G. Kurtz, Bulletin of mathematical biology **72**, 1947 (2010).
[72] D. Kondepudi and I. Prigogine, *Modern thermodynamics: from heat engines to dissipative structures* (John Wiley & Sons, 2014) Chap. 15.
[73] A. Kolchinsky, The Journal of Chemical Physics **161** (2024).

[74] U. Seifert, Physica A: Statistical Mechanics and its Applications **504**, 176 (2018).

[75] C. Van den Broeck and M. Esposito, Physica A: Statistical Mechanics and its Applications **418**, 6 (2015).

[76] K. Sekimoto, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **76**, 060103 (2007).

[77] J. M. Horowitz, The Journal of chemical physics **143** (2015).

[78] J. L. Lebowitz and H. Spohn, Journal of Statistical Physics **95**, 333 (1999).

[79] Y. Oono and M. Paniconi, Progress of Theoretical Physics Supplement **130**, 29 (1998).

[80] M. Esposito and C. Van den Broeck, Physical review letters **104**, 090601 (2010).

[81] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).

[82] M. Esposito, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **85**, 041125 (2012).

[83] M. Feinberg, Foundations of chemical reaction network theory , 205 (2019).

[84] D. B. West, Introduction to graph theory (2001).

[85] V. Didelez, Journal of the Royal Statistical Society Series B: Statistical Methodology **70**, 245 (2008).

[86] S. L. Lauritzen, *Graphical models*, Vol. 17 (Clarendon Press, 1996).

[87] V. Didelez, *Graphical models for event history analysis based on local independence* (Logos-Verlag, 2001).

[88] The notion of a trail in the field of probabilistic graphical models differs from the edge-simple walk, the 'trail' in classical graph theory. In earlier publications (and their follow ups) the trail is referred to as 'chain' [86].

[89] J.-P. Florens, M. Mouchart, and J.-M. Rolin, Elements of bayesian statistics (1990).

[90] The completion of σ-algebras in the conditional measurable separability definition has been dropped by Didelez [85, 87], but was part of the original definition by Florens *et al.* [89] (cf. Definition 5.2.2). The formulation is weaker, but still sufficient to prove the left intersection property for local independence. Hence, local independence still satisfies the asymmetric graphoid realations under this weaker assumption.

[91] C. G. Bowsher, The Annals of Statistics **38** (2010).

[92] R. E. Spinney, J. T. Lizier, and M. Prokopenko, Physical Review E **98**, 032141 (2018).

[93] W. Bialek, SciPost Phys. Lect. Notes, 084 (2024), doi:10.21468/SciPostPhysLectNotes.84.

[94] G. Tkačik and P. R. t. Wolde, Annual Review of Biophysics **54** (2025).

[95] A. El Gamal and Y.-H. Kim, *Network Information Theory* (Cambridge university press, 2011).

[96] To provide intuition, we'll briefly outline the idea behind information units, using the bit as an example. A bit is the unit of information that can be stored (capacity) in a data storage device with two (equally addressable) states (i.e., no state is a priori more likely than the other). That is, any state of this storage device provides 1 bit of information. Consequently, a data storage device with N equally addressable states can store $\log_2(N)$ bits of information (which equals the number of binary storage units required to achieve the same capacity). Shannon extended this notion to the information content of data or signals. For a message $M \sim p(m)$ with discrete alphabet $\mathscr{M}$ the symbol m contains $I(m) := -\log_2(p(m))$ bits of information. Now, $I(m)$ is the number of equally addressable binary information units required to represent the same amount of information. That is, we would be equally surprised to find this state in a storage device of $I(m)$ bits, if all states were equally likely. in this sense, $H(M)/t = \sum_m p(m)I(m)/t$ is the mean rate of information of M. The inequality in (55) is a standard result [110].

[97] A. Klenke, *Probability Theory : A Comprehensive Course*, 3rd ed. (Springer International Publishing, Cham, 2020).

[98] G. Genc, Y. E. Kara, H. B. Yilmaz, and T. Tugcu, IEEE Communications Letters **20**, 1729 (2016).

[99] A. D. Wyner, IEEE Transactions on Information Theory **34**, 1449 (2002).

[100] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems* (Cambridge University Press, 2011).

[101] S. Vembu, S. Verdu, and Y. Steinberg, IEEE Transactions on Information Theory **41**, 44 (2002).

[102] T. Van Vu, Y. Hasegawa, *et al.*, Journal of Physics A: Mathematical and Theoretical **55**, 405004 (2022).

[103] E. Ziv, I. Nemenman, and C. H. Wiggins, PloS one **2**, e1077 (2007).

[104] G. Tkačik, A. M. Walczak, and W. Bialek, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **80**, 031920 (2009).

[105] W. Bialek, I. Nemenman, and N. Tishby, Neural computation **13**, 2409 (2001).

[106] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 6th ed. (Garland Science, New York, NY, 2017).

[107] G. Kramer, IEEE Transactions on Information Theory **49**, 4 (2003).

[108] T. Schreiber, Physical review letters **85**, 461 (2000).

[109] The elements of $L^1([0,t],[0,\infty))$ are equivalence classes (i.e., sets) of integrable functions, that differ on at most countably many points [97]. the elements of equivalence classes are called representatives or versions.

[110] J. A. T. Thomas M. Cover, Entropy, relative entropy, and mutual information, in *Elements of Information Theory* (John Wiley & Sons, Ltd, 2005) Chap. 2, pp. 13–55.