

## Dynamic information transfer in stochastic biochemical networks

Anne-Lena Moor<sup>1,2</sup> and Christoph Zechner<sup>1,2,3,\*</sup>

<sup>1</sup>Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany

<sup>2</sup>Center for Systems Biology Dresden, 01307 Dresden, Germany

<sup>3</sup>Cluster of Excellence Physics of Life, TU Dresden, 01062 Dresden, Germany



(Received 9 August 2022; accepted 8 December 2022; published 23 January 2023)

We develop numerical and analytical approaches to calculate mutual information between complete paths of two molecular components embedded into a larger reaction network. In particular, we focus on a continuous-time Markov chain formalism, frequently used to describe intracellular processes involving lowly abundant molecular species. Previously, we have shown how the path mutual information can be calculated for such systems when two molecular components interact directly with one another with no intermediate molecular components being present. In this paper, we generalize this approach to biochemical networks involving an arbitrary number of molecular components. We present an efficient Monte Carlo method as well as an analytical approximation to calculate the path mutual information and show how it can be decomposed into a pair of transfer entropies that capture the directed flow of information between two network components. We apply our methodology to study information transfer in a simple three-node feedforward network, as well as a more complex positive-feedback system that switches stochastically between two metastable modes.

DOI: [10.1103/PhysRevResearch.5.013032](https://doi.org/10.1103/PhysRevResearch.5.013032)

### I. INTRODUCTION

Information theory provides a powerful mathematical framework to study information transfer in complex dynamical systems. Originating from man-made communication systems [1], information theory has made its way into the biological sciences, where it has helped in the understanding of diverse biological processes, ranging from tissue patterning [2] to signal transduction [3].

A central concept in information theory is that of *mutual information*, a quantity that captures the amount of information that is shared between two random objects  $X$  and  $Y$  [4]. As an example,  $X$  could correspond to the input of a system which is transformed into a corresponding output  $Y$  through a set of mathematical operations. In this case, mutual information provides a quantitative measure of how efficiently the system propagates information from input  $X$  to output  $Y$ . In most biological systems, both  $X$  and  $Y$  as well as the system that relates the two are inherently dynamic. Prime examples can be found, for instance, in gene regulation, where different time-varying transcription factor profiles are converted into distinct gene expression patterns through specific promoter activation and transcription dynamics [5,6]. In these situations,  $X$  and  $Y$  denote complete trajectories of the input and output processes on a considered time interval  $[0, t]$ , and the

corresponding mutual information quantifies the cumulative amount of information exchanged along these trajectories.

While trajectory variants of mutual information are well established for Gaussian processes [7,8], they remain very difficult to calculate for continuous-time Markov chains (CTMCs), which are used to describe biochemical processes involving low-copy-number molecules [9]. Recently, some progress has been made towards addressing this challenge. In Ref. [10], for instance, the authors derive exact expressions for the trajectory-level mutual information and channel capacity for a simple Markov chain model of transcription. For more complex systems, however, analytical solutions are generally not available, and one has to resort to approximation techniques. We have previously proposed one such technique, which estimates the mutual information between complete trajectories of two molecular species by combining stochastic simulations with a moment-closure approximation of a stochastic filtering problem [11]. While this approach is computationally efficient, it is so far limited to two-component systems with no additional intermediate molecular species. A related approach has been proposed in Ref. [12], where the authors use a generic Hawkes process approximation to solve the underlying filtering problem.

An orthogonal method, referred to as *path weight sampling* (PWS), has been developed [13]. The key advantages of PWS are that the resulting estimates are exact up to sampling variance and that it applies also to networks involving intermediate components. On the downside, however, it is challenging to apply PWS when the information flow between the two components  $X$  and  $Y$  is bidirectional (e.g., due to feedback between output and input).

The goal of the present work is to develop an efficient and general approach to quantify mutual information between complete trajectories of any two molecular components of a

\*zechner@mpi-cbg.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

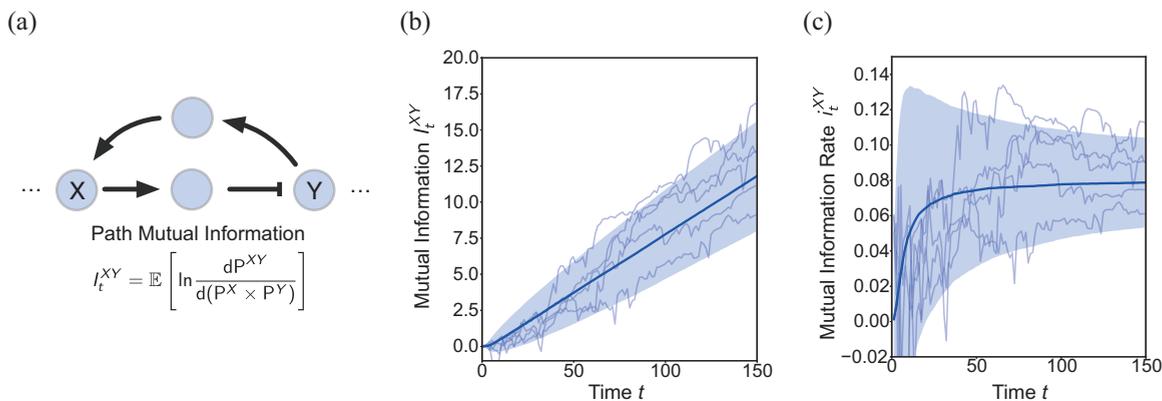
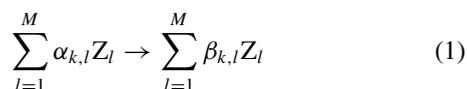


FIG. 1. (a) Schematic representation of a reaction network through which two molecular components X and Y exchange information. The path mutual information quantifies the cumulative amount of information exchanged between X and Y on a time interval  $[0, t]$ . (b) and (c) Example trajectories of the path mutual information  $I_t^{XY}$  (b) and the corresponding path mutual information rate  $i_t^{XY}$  (c). Calculations were performed using reaction network (7) considering  $X = A$  and  $Y = C$  and using parameters  $\{k_1, k_2, k_3, k_4, k_5, k_6\} = \{1, 0.1, 1, 0.1, 1, 0.1\}$ . Mutual information rates were estimated as  $i_t^{XY} \approx I_t^{XY}/t$ . Thick solid lines denote averages calculated over  $n = 10\,000$  samples, and shaded areas mark one standard deviation above and below the mean.

chemical reaction network. To this end, we build on our previously proposed theoretical approach but, importantly, lift the assumption that no intermediate components are present. We present two Monte Carlo schemes to estimate mutual information, an exact but computationally demanding one and a much more efficient one that employs moment-closure approximations to solve the underlying filtering problem. Additionally, we propose an analytical approximation, which provides direct insight into how information transfer depends on the underlying system parameters. We use our methodology to study information processing in two archetypical network motifs, a feedforward three-component system and a positive-feedback system that switches between two metastable states. Our analyses demonstrate the utility of our approach and reveal insights into how information propagates through networks of chemical reactions.

## II. STOCHASTIC REACTION NETWORKS

We consider a well-mixed reaction network consisting of  $M$  chemical species  $Z_1, \dots, Z_M$  and  $K$  reaction channels. Each reaction  $k$  is defined by a stoichiometric equation



with  $\alpha_{k,l}$  and  $\beta_{k,l}$  as reactant and product multiplicities. We denote the stochastic state vector of the system by  $(Z(t))_{t \geq 0}$ , which tracks the copy numbers of all species over time. Each reaction channel is associated with a rate function  $\lambda_k(Z(t))$ , which defines how likely a reaction fires within a small amount of time given the current state of the system. Typically,  $\lambda_k(Z(t))$  is given by the law of mass action, but nonelementary rate laws such as Michaelis-Menten kinetics could also be considered. When a reaction of type  $k$  happens at time  $t^*$ , the system state changes instantaneously from  $Z(t^*)$  to  $Z(t^*) + \nu_k$ , where  $\nu_k = (\beta_{k,l} - \alpha_{k,l})_{l=1, \dots, M}$ . The dynamics of  $Z(t)$  satisfies a Markov jump process, which can be de-

scribed at the level of individual trajectories using the random time-change representation [14]

$$Z(t) = Z(0) + \sum_{k=1}^K N_k \left( \int_0^t \lambda_k(Z(s)) ds \right) \nu_k \quad (2)$$

with  $N_1(t), \dots, N_K(t)$  being independent unit Poisson processes and  $Z(0)$  being the initial state of the system. We denote by  $Z_0^t$  a complete trajectory of  $Z(t)$ , collecting all reaction times and types within the time interval  $[0, t]$ .

## III. PATH MUTUAL INFORMATION

We are interested in the dynamic exchange of information between two arbitrary chemical species  $X = Z_l$  and  $Y = Z_j$  that are part of system (2) [Fig. 1(a)]. To this end, we define trajectories  $X_0^t \subset Z_0^t$  and  $Y_0^t \subset Z_0^t$  which contain only the reaction times and types that modify X and Y, respectively. For simplicity, we exclude reactions that change X and Y simultaneously such as  $X \rightarrow Y$ . In this case, the two subtrajectories  $X_0^t$  and  $Y_0^t$  are disjoint sets, which simplifies the notation required in what follows. We remark, however, that the overall approach applies analogously to the more general case, where X and Y can change simultaneously.

The cumulative amount of information transfer on the interval  $[0, t]$  can be quantified by the path mutual information

$$I_t^{XY} = \mathbb{E} \left[ \ln \frac{dP^{XY}}{d(P^X \times P^Y)} \right] \quad (3)$$

with  $P^{XY}$  being the *joint* path measure associated with the combined trajectory  $\{X_0^t, Y_0^t\}$  and  $P^X$  and  $P^Y$  being *marginal* path measures corresponding to  $X_0^t$  and  $Y_0^t$ , respectively. Note that also  $P^{XY}$  is technically a marginal measure because all chemical species apart from X and Y have been integrated out. The term inside the logarithm of (3) denotes the Radon-Nikodym derivative [15] between  $P^{XY}$  and  $P^X \times P^Y$ . Evaluating the latter for paths satisfying (2) [15,16], taking

the logarithm, and simplifying (Appendix A) leads to

$$I_t^{XY} = \sum_{k \in R_X} \int_0^t \mathbb{E} [\lambda_k^{XY}(s) \ln [\lambda_k^{XY}(s)] - \lambda_k^X(s) \ln [\lambda_k^X(s)]] ds + \sum_{k \in R_Y} \int_0^t \mathbb{E} [\lambda_k^{XY}(s) \ln [\lambda_k^{XY}(s)] - \lambda_k^Y(s) \ln [\lambda_k^Y(s)]] ds. \quad (4)$$

In (4), the sets  $R_X$  and  $R_Y$  comprise all reactions that modify  $X$  [or  $Y$ ], except those whose propensity depends exclusively on  $X(t)$  [or  $Y(t)$ ]. Only these reactions lead ultimately to an exchange of information between  $X$  and  $Y$ , which will be illustrated more concretely later in our case studies. The functions  $\lambda_k^{XY}(t)$ ,  $\lambda_k^X(t)$ , and  $\lambda_k^Y(t)$  denote *marginal propensities* [17,18], that is, the rate functions with which  $(X(t), Y(t))$ ,  $X(t)$ , and  $Y(t)$  evolve if the states of all other species are unknown. A marginal propensity is defined as a conditional expectation  $\lambda_k^A(t) = \mathbb{E}[\lambda_k(Z(t)) | A_0^t]$ , corresponding to the optimal causal estimate of  $\lambda_k(Z(t))$  given some partial path  $A_0^t \subset Z_0^t$ . This demonstrates that the amount of information transferred between sender and receiver depends on how well the sender’s signal can be reconstructed from measurements taken by the receiver (and vice versa in the presence of feedback) [19,20]. We remark that while only the reactions in  $R_X$  and  $R_Y$  show up explicitly in (4), the other reactions also contribute implicitly to  $I_t^{XY}$  through the inner and outer expectations in (4).

Note that the first and second lines on the right-hand side of Eq. (4) can be identified as transfer entropies  $H_t^{Y \rightarrow X}$  and

$H_t^{X \rightarrow Y}$ , which correspond to the fraction of information that is transferred from  $X$  to  $Y$  and from  $Y$  to  $X$ , respectively. In contrast to the mutual information, which by definition is symmetric in  $X$  and  $Y$ , the transfer entropy provides useful insights into the directed flow of information in dynamical systems [21].

In many situations, it is helpful to study information transfer at stationary states. Since  $I_t^{XY}$ ,  $H_t^{Y \rightarrow X}$ , and  $H_t^{X \rightarrow Y}$  generally increase with time [e.g., Fig. 1(b)], they typically diverge as  $t \rightarrow \infty$ . In these cases, one can resort to the corresponding *information rates*, which for the quantities of interest can be defined as  $i^{XY} = \lim_{t \rightarrow \infty} I_t^{XY} / t$ ,  $h^{X \rightarrow Y} = \lim_{t \rightarrow \infty} H_t^{X \rightarrow Y} / t$ , and  $h^{Y \rightarrow X} = \lim_{t \rightarrow \infty} H_t^{Y \rightarrow X} / t$  [e.g., Fig. 1(c)].

#### IV. STOCHASTIC FILTERING

The central step in evaluating (4) is the calculation of the conditional expectations that are required for determining the marginal propensities. In the case of  $\lambda_k^X(t)$ , for instance, we have to average  $\lambda_k(Z(t)) = \lambda_k(\bar{z}, X(t))$  with respect to the conditional probability distribution  $\pi^X(\bar{z}, t) = P(\bar{Z}(t) = \bar{z} | X_0^t)$ , where  $\bar{Z}(t)$  is a vector containing all copy numbers of  $Z(t)$  except  $X(t)$ . It can be shown that such a conditional probability distribution satisfies a stochastic differential equation termed a *filtering equation* [22]. In the case of  $\pi^X(\bar{z}, t)$ , this equation reads

$$d\pi^X(\bar{z}, t) = \left[ \mathcal{A}^{\bar{Z}|X} \pi^X(\bar{z}, t) - \sum_{k \in R_{\bar{X}}} [\lambda_k(\bar{z}, X(t)) - \lambda_k^X(t)] \pi^X(\bar{z}, t) \right] dt + \sum_{k \in R_{\bar{X}}} \left[ \frac{\lambda_k(\bar{z} - v_k^{\bar{Z}}, X(t))}{\lambda_k^X(t)} \pi^X(\bar{z} - v_k^{\bar{Z}}, t) - \pi^X(\bar{z}, t) \right] dN_k(t), \quad (5)$$

where  $\mathcal{A}^{\bar{Z}|X}$  is an operator that is related to the generator of the original, unconditional process [see Appendix B, Eq. (B2), for more details],  $v_k^{\bar{Z}}$  is the part of the state change vector of reaction  $k$  acting upon  $\bar{Z}$ , and  $R_{\bar{X}}$  collects all reactions that modify  $X$ . Note that in the two sums of Eq. (5), only reactions in  $R_{\bar{X}}$  that involve species in  $\bar{Z}$  (as reactants and/or products) will have a nonzero contribution.

There are two major challenges in evaluating Eq. (4) in practice. First, analytical solutions of Eq. (5) and the corresponding conditional expectations are available only in exceptional cases. Second, the outer expectation in (4) is taken with respect to a distribution that is generally not known analytically. Even if that were the case, expectations over the  $x \ln x$  terms in (4) are most likely intractable in practice. To address these problems, we propose three different approaches which differ in scope and computational efficiency.

##### A. Quasixact method

The quasixact approach numerically integrates Eq. (5) on a finite grid. This is analogous to the finite-state projection algorithm that is commonly applied to numerically solve conventional (unconditional) master equations [23]. The outer expectation of (4) is calculated as a Monte Carlo average over  $n$  independent path realizations generated using Gillespie’s stochastic simulation algorithm [24]. The main advantage of this approach is that its error is fully controllable and negligible when the grid size and sample size are sufficiently large. As with other finite-state projection approaches, however, the efficiency of this approach suffers from the combinatorial explosion of states in larger reaction networks. We will use this technique to calculate ground-truth solutions for comparison with our approximate techniques described below.

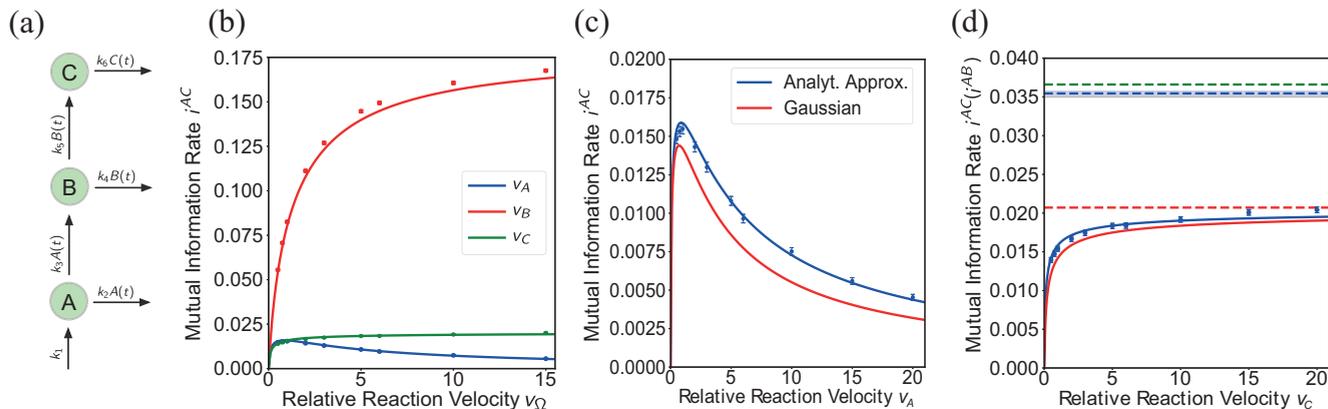


FIG. 2. Information transfer in a three-node feedforward network. (a) Schematic illustration of the considered network. (b) Stationary path mutual information rate between species A and C ( $i^{AC}$ ) as a function of the relative reaction velocity  $v_\Omega$ ,  $\Omega \in \{A, B, C\}$  calculated via the analytical approximation (Analyt. Approx.; solid lines) and the moment-approximation method (dots). (c) Stationary path mutual information rate  $i^{AC}$  as a function of the relative reaction velocity  $v_A$  as in (b) and comparison with Gaussian process theory [8] (red). (d) Stationary path mutual information rate  $i^{AC}$  as a function of the relative reaction velocity  $v_C$  as in (b) (solid lines) and comparison with the mutual information rate between nodes A and B,  $i^{AB}$ , calculated via the analytical approximation (dashed green line), the quasixact method (dashed blue line), and Gaussian process theory (dashed red line). Simulations were performed with parameters  $\{k_1, k_2, k_3, k_4, k_5, k_6\} = \{1, 0.1, 0.1, 0.1, 1, 0.1\}$ , whereas individual pairs of parameters were varied as described in the main text. Monte Carlo averages were calculated using  $n = 10000$  samples. Error bars and shaded areas correspond to 2.5 times the standard error.

### B. Moment-approximation method

In principle, we can derive an equation for the marginal propensity  $\lambda_k^X(t)$  by multiplying Eq. (5) with  $\lambda_k(\bar{z}, X(t))$  and summing over all  $\bar{z}$ . If all propensity functions are polynomial (as is the case for mass-action kinetics), this leads to a system of moment differential equations, which in general, however, is infinite dimensional. This problem can be addressed by imposing distributional assumptions on the conditional distribution which can then be used to express moments higher than a certain order as functions of lower-order moments [25]. While moment-closure approximations are generally *ad hoc*, we found that the conditional distribution (5) is typically very well approximated by those techniques. Throughout our case studies, we found the multivariate Gamma closure proposed in Ref. [26] to yield excellent results. We remark that while the moment-approximation method requires polynomial rate functions, it may also be applied to more complex (e.g., rational) rate laws if suitable polynomial approximations are available (see Sec. V, Case Study II).

### C. Analytical approximation

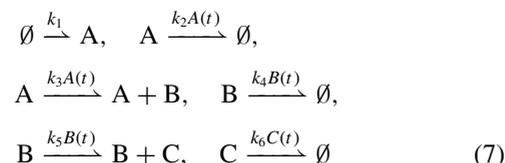
Once we have obtained a closed system of conditional moment equations, we can approximate the outer expectation in (4) by employing a second-order Taylor expansion. In particular, if we perform an expansion around the respective expectations  $\mathbb{E}[\lambda_k^{XY}(t)] = \mathbb{E}[\lambda_k^X(t)] = \mathbb{E}[\lambda_k^Y(t)] = \mathbb{E}[\lambda_k(Z(t))]$ , we obtain

$$I_t^{XY} \approx \sum_{k \in R_X} \int_0^t \frac{\text{Var}[\lambda_k^{XY}(t)] - \text{Var}[\lambda_k^X(t)]}{2\mathbb{E}[\lambda_k(Z(t))]} ds + \sum_{k \in R_Y} \int_0^t \frac{\text{Var}[\lambda_k^{XY}(t)] - \text{Var}[\lambda_k^Y(t)]}{2\mathbb{E}[\lambda_k(Z(t))]} ds. \quad (6)$$

This equation involves variances of the marginal propensities for which approximate differential equations can be derived (see Appendix B 2). Solving these equations provides a direct way to calculate the path mutual information and its rate.

### V. CASE STUDIES

*Case study I: Three-node feedforward network.* The first system we want to study is a simple feedforward reaction network



with  $k_1, \dots, k_6$  as rate constants. We have chosen this network because it resembles an elementary motif where an input (species A) transmits information to an output (species C) through an intermediate component (species B). Moreover, all first- and second-order moments are analytically tractable, which will be useful to compare our results with predictions obtained from Gaussian process theory [8]. Using Eq. (4), the mutual information between paths  $A_0^t$  and  $C_0^t$  is given by

$$I_t^{AC} = \int_0^t \mathbb{E}[k_5 \mathbb{E}[B(s)|A_0^t, C_0^t] \ln(k_5 \mathbb{E}[B(s)|A_0^t, C_0^t])] ds - \int_0^t \mathbb{E}[k_5 \mathbb{E}[B(s)|C_0^t] \ln(k_5 \mathbb{E}[B(s)|C_0^t])] ds. \quad (8)$$

Note that (8) involves terms associated with reaction  $B \rightarrow B + C$  only. This is because only through this reaction does component C ultimately receive information about component B and, hence, A. As mentioned previously, however, other

reactions contribute implicitly to  $I_t^{AC}$  through the expectation values in (8).

To study information transmission in the considered network, we calculate the stationary path mutual information rate  $i^{AC}$  for different parameter regimes. To this end, we introduce reaction velocities  $v_A$ ,  $v_B$ , and  $v_C$  that set the time scales of production and degradation of A, B, and C without changing their average abundance. In the case of A, for instance, we set  $k_1 = \tilde{k}_1 v_A$  and  $k_2 = \tilde{k}_2 v_A$  such that  $\mathbb{E}[A(t)] = \tilde{k}_1/\tilde{k}_2$  for any value of  $v_A$ . To verify the accuracy of our approach, we compared the moment-approximation method with the analytical and quasixact methods and found good agreement between all three approaches [Fig. 2(b) and Appendix C, Fig. 5].

Our analysis shows that while the mutual information rate  $i^{AC}$  increases monotonically with both  $v_B$  and  $v_C$ , it scales nonmonotonically with  $v_A$ , exhibiting a maximum at an intermediate value of  $v_A$ . Intuitively, this is because for small  $v_A$ , species B is fast enough to track changes in A but, at the same time, A produces little information per unit time. Correspondingly, increasing  $v_A$  will initially increase  $i^{AC}$  because more information is generated by A. However, when A becomes fast in comparison to B, a substantial amount of information is lost between A and B, causing  $i^{AC}$  to decrease for large  $v_A$ . In other words, there exists an optimal time scale of A that strikes a balance between the amount of information produced by A and the fraction of it that can be transferred to C via intermediate species B as has been similarly observed in other types of systems [13,27]. In contrast, varying either  $v_B$  or  $v_C$  does not affect the information content in A but only how effectively this information can be propagated forward, leading to a simple monotonic relationship between  $i^{AC}$  and  $v_B$  and  $v_C$ , respectively.

We next use the same three-node network motif to study whether and to what extent discrete-state biochemical systems differ from Gaussian processes in terms of information transfer. To this end, we consider a continuous variant of network (7) where the abundances of A, B, and C are described by a three-dimensional Gaussian process with identical first- and second-order statistics. The corresponding path mutual information rate  $i_G^{AC}$  can be obtained from Gaussian channel theory (see Appendix C1 and Refs. [7,8]). As can be proven analytically, Gaussian theory provides a lower bound on mutual information for non-Gaussian scenarios as long as the first- and second-order statistics are known [28]. This is reflected also by our analysis, which compares the path mutual information rate  $i^{AC}$  with  $i_G^{AC}$  across different  $v_A$  [Fig. 2(c)]. While Gaussian theory predicts a very similar scaling and optimum of  $i_G^{AC}$  with respect to  $v_A$ , it generally underestimates the information transfer between A and C.

We next considered the limit  $v_C \rightarrow \infty$  such that information in B is expected to propagate to C in a “perfect” manner. In this case, Gaussian theory predicts that

$$\lim_{v_C \rightarrow \infty} i_G^{AC} = -\frac{k_2}{2} + \frac{1}{2}\sqrt{k_2(k_2 + k_3)}, \quad (9)$$

which coincides with the Gaussian mutual information rate between A and B,  $i_G^{AB}$ . In other words, the three-node network reduces to a two-node network for Gaussian processes when  $v_C \rightarrow \infty$ , as might be expected intuitively. To compare these results with the discrete-state network, we determined

$\lim_{v_C \rightarrow \infty} i^{AC}$  analytically based on (6), which happens to coincide exactly with (9) (see Appendix C 1 for a derivation). In the case of a discrete-state system, however, this asymptotic limit is lower than the mutual information rate between A and B, which based on (6) is approximated as

$$i^{AB} \approx -\frac{k_2}{2} + \frac{1}{2}\sqrt{k_2(k_2 + 2k_3)}. \quad (10)$$

Numerical simulations show that both  $i^{AB}$  and  $i^{AC}$  are approximated accurately through (6) [Fig. 2(d) and Appendix C, Fig. 5(c)]. Calculating the ratio between  $i^{AB}$  and  $i^{AC}$  and taking the respective limits with respect to  $k_2$  suggests that  $i^{AC}$ , and correspondingly  $i_G^{AB}$ , are lower than  $i^{AB}$  by a factor of at least  $\sqrt{2}$  ( $k_2 \rightarrow 0$ ) and at most 2 ( $k_2 \rightarrow \infty$ ) for any  $k_3 > 0$ . In other words, the amount of information transferred between A and B is systematically larger in the discrete-state case, but this additional amount of information is inevitably lost between B and C, even when  $v_C \rightarrow \infty$ . This is in contrast to the Gaussian scenario where the information transfer between A and B is lower to begin with, but the additional loss of information through intermediate species B can be made arbitrarily small by increasing  $v_C$ . This discrepancy seems to originate from the fact that for discrete-state biochemical networks, fluctuations in the copy number of B can be uniquely attributed to either the birth or death reactions, whereas only the former carry information about the dynamics of A. In the Gaussian approximation, by contrast, birth and death reactions are coarse grained into an effective continuous process, where information about individual events is lost. In summary, our results show that discrete biochemical systems can exhibit not only quantitative but also even qualitative differences when compared with equivalent Gaussian approximations.

*Case study II: Bistable switch.* As a second example, we consider a variant of the previous network that exhibits more complex nonlinear dynamics. In particular, we introduce positive feedback between species C and A by replacing the constant production rate of A with one that increases with the abundance of C [Fig. 3(a)]. More precisely, we choose a Hill-type rate law of the form

$$k_1(C(t)) = \mu \frac{C(t)^{n_H}}{K^{n_H} + C(t)^{n_H}} + \epsilon \quad (11)$$

with  $\mu$ ,  $n_H$ ,  $K$ , and  $\epsilon$  being positive constants. For certain parameter regimes, this system is bistable where individual trajectories switch stochastically between two modes (see caption of Fig. 3 for specific parameter values). The goal of this case study is to understand how such bistability affects information transfer in biochemical systems. To this end, we applied our approach to estimate stationary mutual information rates between species A and C. Note that due to the nonpolynomial form of the rate  $k_1$ , an additional Taylor expansion had to be used to make our moment-approximation method applicable (see Appendix C 2 for more details).

We first analyzed how the mutual information rate changes for varying feedback strengths  $\mu$ . For low and high values of  $\mu$ , the system exhibits a single mode, while intermediate values of  $\mu$  lead to bimodal behavior [Fig. 3(b)]. This is resembled qualitatively by a corresponding mean-field model (Appendix C, Fig. 6). Figure 3(c) shows that, the mutual information rate first increases with  $\mu$  until it reaches a certain

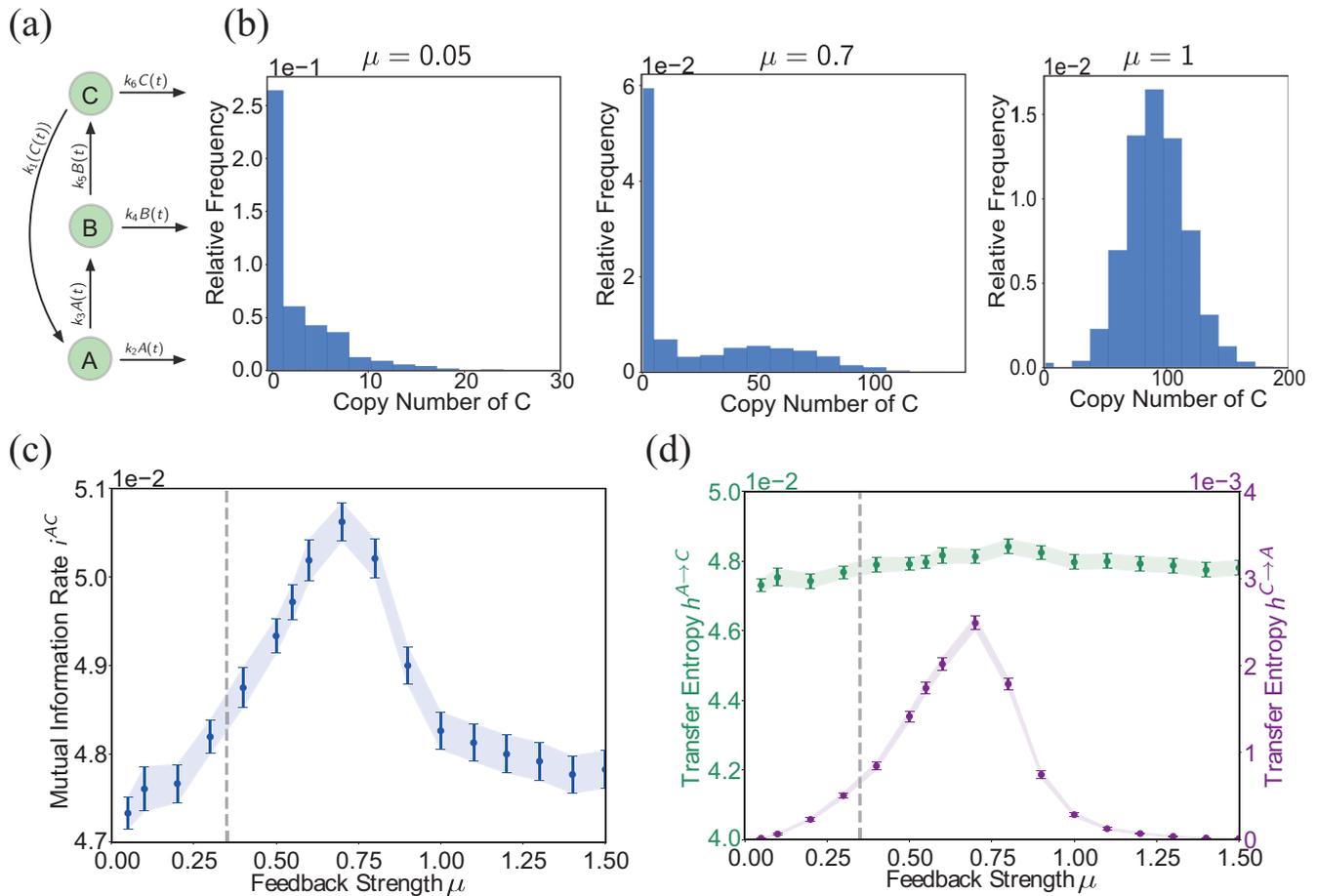


FIG. 3. Information transfer in a bistable system. (a) Schematic illustration of the considered network. (b) Stationary copy number distributions of species C for parameters  $\mu = 0.05$ ,  $\mu = 0.7$ , and  $\mu = 1$ . (c) Stationary path mutual information rate  $i^{AC}$  between species A and C. The dashed gray line indicates the bifurcation point predicted from mean-field theory. (d) Decomposition of the path mutual information rate into forward and backward transfer entropies  $h^{A \rightarrow C}$  (green) and  $h^{C \rightarrow A}$  (magenta), respectively. Simulations were performed with the parameters  $\{k_2, k_3, k_4, k_5, k_6\} = \{0.1, 1, 0.1, 0.1, 0.1\}$  and  $\{K, n_H, \epsilon\} = \{30, 3, 0.03\}$ . The system was simulated for  $T = 10\,000$  time units to reach steady state, whereas only the last 1500 time units were used to estimate stationary information rates. Ensemble averages were calculated using  $n = 5000$  samples. Error bars correspond to 2.5 times the standard error.

maximum. At this point, the system fluctuates between two equilibrium points as can be seen from the copy numbers of the species being bimodally distributed [Fig. 3(b)]. Beyond this maximum, the bimodality becomes less pronounced, and  $i^{AC}$  decreases again. Interestingly, this suggests that information transfer is most effective in regimes where the system as a whole is very noisy.

To understand this better, we decomposed the mutual information rate into the transfer entropy rates  $h^{A \rightarrow C}$  and  $h^{C \rightarrow A}$ , quantifying the directed flow of information from A to C and from C to A, respectively [Fig. 3(d)]. Interestingly, this shows that the forward contribution  $h^{A \rightarrow C}$  is more or less the same for all considered feedback strengths  $\mu$ , regardless of whether the system exhibits one or two modes. By contrast, the backward contribution  $h^{C \rightarrow A}$  is approximately zero for very small  $\mu$  but then shows a peak that is located within the bimodal regime. From this point on, the second equilibrium becomes more and more populated, and the backward contribution  $h^{C \rightarrow A}$  decreases again. For very large  $\mu$ , the rate  $k_1$  is strongly satu-

rated, such that the system effectively reduces to the simpler feedforward motif discussed in the previous section, where no information flows along the backward direction  $C \rightarrow A$ .

In summary, this shows that information transfer between two molecular species can be significantly enhanced by positive feedback. This is the case when the feedback strength is in a regime where it generates multiple metastable equilibria that the system can attain. In this situation, both the forward contribution and the backward contribution to the mutual information are significantly different from zero, leading to large values overall. More generally, this analysis illustrates information processing in a complex and strongly nonlinear dynamical system that is beyond Gaussian theory.

### VI. CONCLUSIONS

In this paper we have developed a general method to quantify information transmission in biochemical networks via the path mutual information. This method exploits a fundamental

relationship between mutual information and filtering theory [19,20]. We have first introduced a quasixact Monte Carlo scheme that combines conventional stochastic simulations with a brute-force numerical solution of the underlying filtering equations. While this approach was needed to calculate ground-truth solutions, it quickly becomes computationally infeasible as the system size grows. As we have shown in our earlier work [11], this problem can be addressed using moment-closure approximations that project the filtering distribution onto a finite set of (approximate) moments, which are obtained by solving a system of differential equations. In our numerical experiments, we found the approximate Monte Carlo method based on the Gamma closure to be in very good agreement with the exact path mutual information, although different closures may be required for other types of systems.

We have further shown how the outer expectation in the calculation of the path mutual information can be approximated analytically. In this way, Monte Carlo sampling can be avoided entirely, and the path mutual information becomes analytically accessible. Although the proposed approximation appears relatively coarse, it was in surprisingly close agreement with Monte Carlo estimation. Deriving similar approximations in a more principled manner will be an interesting avenue for future research.

When applied to our case studies, the path mutual information revealed interesting insights into how information propagates across cascades of chemical reactions. For instance, we found that discrete biochemical processes exhibit quantitative and even qualitative differences in information transmission when compared with equivalent Gaussian processes with identical first- and second-order statistics. Fundamentally, these differences appear to result from a different notion of path in each of these two situations: For discrete-state systems, a path contains detailed information about reaction types and time points. In the Gaussian process approximation, fluctuations in a species' copy number are coarse grained into effective stochastic increments to which multiple reactions contribute. Since individual events can no longer be resolved, this introduces additional uncertainty and, correspondingly, a loss of information. Gaining a deeper understanding of these discrepancies and their implications for cellular information processing is an important goal for the future.

In our second case study, we have studied information transfer in a nonlinear, positive-feedback system. Our analysis revealed that positive feedback—and the resulting bistability—can enhance information transmission between input and output. By decomposing the mutual information into the respective transfer entropies, we found that this enhancement is due to an increased backward contribution to the mutual information (i.e., from output to input) while the forward contribution remains largely unaffected by the presence of feedback. Interestingly, the backward contribution is maximal in the bimodal regime, when the system switches randomly and evenly between the two modes.

In summary, our results highlight the need for information theoretical concepts that are compatible with the discrete and nonlinear dynamics of biochemical networks. The methodology outlined in this paper aims to fill this gap, and we envision several interesting applications in the future. For

instance, it could be used to identify network architectures and parameter regimes that are optimal in terms of information transfer and understand how those compare with evolved intracellular systems. Along those lines, it may be possible to derive bounds on mutual information in small networks embedded into larger ones, and how those limit the precision of biochemical pathways [29]. Beyond this, mutual information and transfer entropy rates play an important role in the context of stochastic thermodynamics, for instance, to derive second-law-like inequalities for feedback-controlled systems [30]. Understanding how the information processing capabilities of biochemical systems are limited by thermodynamic constraints will be an interesting subject for future work.

The PYTHON code underlying all our simulations is available [31].

### ACKNOWLEDGMENT

The authors thank L. Duso, C. Modes, D. Busiello, T. Bianucci, and the ten Wolde group for useful insights and critical feedback on the work. This work was supported by core funding of the Max Planck Institute of Molecular Cell Biology and Genetics.

### APPENDIX A: DERIVATION OF THE PATH MUTUAL INFORMATION

The following derivation is based on our previous work [11]. As discussed in the main text, the mutual information between two trajectories  $X_0^t \subset Z_0^t$  and  $Y_0^t \subset Z_0^t$  can be defined as

$$I_t^{XY} = \mathbb{E} \left[ \ln \frac{dP^{XY}}{d(P^X \times P^Y)} \right], \quad (\text{A1})$$

where  $dP^{XY}/d(P^X \times P^Y)$  is the Radon-Nikodym derivative between the joint path measure  $P^{XY}$  and the product of the marginal path measures  $P^X$  and  $P^Y$ . Intuitively, the Radon-Nikodym derivative corresponds to a likelihood ratio between two competing probability laws, in this case  $P^{XY}$  and  $P^X \times P^Y$ , respectively. In other words, it assesses how much more likely a joint path  $\{X_0^t, Y_0^t\}$  originates from the (true) joint probability measure  $P^{XY}$  than from the product measure  $P^X \times P^Y$ , in which  $X_0^t$  and  $Y_0^t$  are considered to be independent.

We remark that all three probability measures,  $P^{XY}$ ,  $P^X$ , and  $P^Y$ , are marginal measures because all remaining components are “integrated out.” In other words, they are the probability laws that capture how  $(X(t), Y(t))$ ,  $X(t)$ , and  $Y(t)$  evolve if no knowledge about all other components in  $Z(t)$  is available. We and others have previously shown how the dynamics of such marginal processes can be obtained using the theory of stochastic filtering [17,18,32]. To illustrate this, we focus on the case of  $Y(t)$ , but  $(X(t), Y(t))$  and  $X(t)$  can be handled analogously. If knowledge about the whole state  $Z(t)$  is available, the dynamics of  $Y(t)$  satisfies

$$Y(t) = Y(0) + \sum_{k \in R_Y} N_k \left( \int_0^t \lambda_k(Z(t)) \right) v_k^Y, \quad (\text{A2})$$

where  $v_k^Y$  is the element of the stoichiometric change vector  $v_k$  that acts upon  $Y(t)$ . With  $R_{\bar{Y}}$  and  $R_{\bar{X}}$  we denote the set of reactions that modify  $Y(t)$  and  $X(t)$ , respectively. Note that we focus here on the case where there are no reactions that change  $X(t)$  and  $Y(t)$  simultaneously (e.g.,  $X \rightarrow Y$ ), such that  $R_{\bar{X}}$  and  $R_{\bar{Y}}$  are disjoint sets. Reactions that change more than one  $Z_i(t) \notin \{X(t), Y(t)\}$  at the same time and reactions that change  $Z_i(t)$  simultaneously with  $X(t)$  or  $Y(t)$  are included. As can be seen from (A2), the dynamics of  $Y(t)$  depends on the complete state  $Z(t)$  and as such is not self-contained. In more technical terms, (A2) describes the dynamics of  $Y(t)$  relative to the natural filtration of  $Z(t)$ , that is, the complete history of events  $Z_0^t$  that brought the full system from  $Z(0)$  to  $Z(t)$ . To derive the marginal dynamics of  $Y(t)$ , one then requires  $Y(t)$  to depend no longer on the complete history  $Z_0^t$ , but only on its own history  $Y_0^t \subset Z_0^t$ . The latter contains information about all reactions  $R_{\bar{Y}}$  that modify species  $Y(t)$ . The innovation theorem [33] then states that the dynamics of  $Y(t)$  relative to  $Y_0^t$  satisfies

$$Y(t) = Y(0) + \sum_{k \in R_{\bar{Y}}} N_k \left( \int_0^t \lambda_k^Y(s) \right) v_k^Y, \quad (\text{A3})$$

with  $\lambda_k^Y(t) = \mathbb{E}[\lambda_k(Z(t)) | Y_0^t]$ . In other words, the original propensities are replaced by their expectation conditionally

on  $Y_0^t$ , which we refer to as *marginal propensity*. Importantly, Eq. (A3) is now self-contained because the dependency on all components except  $Y(t)$  has been integrated out through the conditional expectation. We remark that (A3) is exact: Solving (A3)—for instance, through stochastic simulation—will generate paths  $Y_0^t$  consistent with the marginal probability measure  $\mathbb{P}^Y$ . Analogous constructions can be performed for  $\mathbb{P}^{XY}$  and  $\mathbb{P}^X$ , respectively.

To determine the Radon-Nikodym derivative in (A1), we focus on the joint trajectories  $\{X_0^t, Y_0^t\}$ , which contain information about all reactions that modify  $X(t)$  and  $Y(t)$ . We can now employ Jacod's formula for the Radon-Nikodym derivative [16], which for a counting process  $\Xi(t)$  consisting of  $K$  counting processes [e.g., such as Eq. (2) in the main text] and with known initial state  $\Xi(0)$  takes the form

$$\frac{dW}{dQ} = \frac{\prod_{k \in K} \left( \prod_{j=1}^{N_k(t)} \phi_k(\tau_{k,j}) \right) \exp \left[ - \int_0^t \phi_k(s) ds \right]}{\prod_{k \in K} \left( \prod_{j=1}^{N_k(t)} \tilde{\phi}_k(\tau_{k,j}) \right) \exp \left[ - \int_0^t \tilde{\phi}_k(s) ds \right]}, \quad (\text{A4})$$

where  $W$  and  $Q$  are the measures under which the process has propensity functions  $\phi_k$  and  $\tilde{\phi}_k$ , respectively. Note that  $\phi_k$  and  $\tilde{\phi}_k$  can in general depend on the state of  $\Xi(t)$  or even the whole process history  $\Xi_0^t$ . The symbol  $\tau_{k,j}$  denotes the time point right before the  $j$ th reaction of type  $k$  happens. Instantiating (A4) for the derived marginal processes [e.g., Eq. (A3) in the case of  $Y(t)$ ] yields

$$\frac{d\mathbb{P}^{XY}}{d(\mathbb{P}^X \times \mathbb{P}^Y)} = \frac{\prod_{k \in R_{\bar{X}} \cup R_{\bar{Y}}} \left( \prod_{j=1}^{N_k(t)} \lambda_k^{XY}(\tau_{k,j}) \right) \exp \left[ - \int_0^t \lambda_k^{XY}(s) ds \right]}{\prod_{k \in R_{\bar{X}}} \left( \prod_{j=1}^{N_k(t)} \lambda_k^X(\tau_{k,j}) \right) \exp \left[ - \int_0^t \lambda_k^X(s) ds \right] \times \prod_{k \in R_{\bar{Y}}} \left( \prod_{j=1}^{N_k(t)} \lambda_k^Y(\tau_{k,j}) \right) \exp \left[ - \int_0^t \lambda_k^Y(s) ds \right]} \quad (\text{A5})$$

with  $\lambda_k^{XY}(t) = \mathbb{E}[\lambda_k(Z(t)) | X_0^t, Y_0^t]$ ,  $\lambda_k^X(t) = \mathbb{E}[\lambda_k(Z(t)) | X_0^t]$ , and  $\lambda_k^Y(t) = \mathbb{E}[\lambda_k(Z(t)) | Y_0^t]$  being the marginal propensities of processes  $(X(t), Y(t))$ ,  $X(t)$ , and  $Y(t)$ . Equation (A5) can be further simplified by realizing that reactions in  $R_{\bar{X}}$  for which  $\lambda_k^{XY}(t) = \lambda_k^X(t)$  cancel out (and equivalently for  $R_{\bar{Y}}$ ). This is the case for reactions whose propensity function depends exclusively on  $X(t)$  [or  $Y(t)$ ]. Therefore we only have to consider reactions in  $R_{\bar{X}}$  (or  $R_{\bar{Y}}$ ) that depend on species other than  $X(t)$  [or  $Y(t)$ ] such as  $Z_k \rightarrow Z_k + X$ . We refer to these reactions as  $R_X$  and  $R_Y$ , respectively. With these definitions, and realizing that  $\prod_{j=1}^{N_k(t)} f(\tau_{k,j}) = \exp \left[ \int_0^t \ln f(s) dN_k(s) \right]$  yields for the path mutual information

$$I_t^{XY} = \mathbb{E} \left[ \sum_{k \in R_X} \left( \int_0^t (\ln [\lambda_k^{XY}(s)] - \ln [\lambda_k^X(s)]) dN_k(s) - \int_0^t (\lambda_k^{XY}(s) - \lambda_k^X(s)) ds \right) + \sum_{k \in R_Y} \left( \int_0^t (\ln [\lambda_k^{XY}(s)] - \ln [\lambda_k^Y(s)]) dN_k(s) - \int_0^t (\lambda_k^{XY}(s) - \lambda_k^Y(s)) ds \right) \right]. \quad (\text{A6})$$

Moreover, we can decompose the reaction counters  $N_k(t)$  into a predictable part and a martingale such that  $dN_k(t) = \lambda_k^{XY}(t) dt + d\tilde{N}_k(t)$ , where  $d\tilde{N}_k(t)$  is a *centered* process which is zero on average. In addition, we observe that

$$\mathbb{E}[\lambda_k^X(t)] = \mathbb{E}[\mathbb{E}[\lambda_k(Z(t)) | X_0^t]] = \mathbb{E}[\lambda_k(Z(t))]. \quad (\text{A7})$$

Then, by changing the order of integration and expectation, we further obtain

$$I_t^{XY} = \sum_{k \in R_X} \int_0^t \mathbb{E}[(\ln [\lambda_k^{XY}(s)] - \ln [\lambda_k^X(s)])(\lambda_k^{XY}(s) ds + d\tilde{N}(s))] - \int_0^t \mathbb{E}[\lambda_k^{XY}(s) - \lambda_k^X(s)] ds + \sum_{k \in R_Y} \int_0^t \mathbb{E}[(\ln [\lambda_k^{XY}(s)] - \ln [\lambda_k^Y(s)])(\lambda_k^{XY}(s) ds + d\tilde{N}(s))] - \int_0^t \mathbb{E}[\lambda_k^{XY}(s) - \lambda_k^Y(s)] ds$$

$$\begin{aligned}
 &= \sum_{k \in R_X} \int_0^t \mathbb{E}[(\ln[\lambda_k^{XY}(s)] - \ln[\lambda_k^X(s)])\lambda_k^{XY}(s)]ds \\
 &+ \sum_{k \in R_Y} \int_0^t \mathbb{E}[(\ln[\lambda_k^{XY}(s)] - \ln[\lambda_k^Y(s)])\lambda_k^{XY}(s)]ds,
 \end{aligned} \tag{A8}$$

where the second equality follows from the martingality of  $\tilde{N}_k(t)$  and from  $\mathbb{E}[\lambda_k^{XY}(t)] = \mathbb{E}[\lambda_k^X(t)] = \mathbb{E}[\lambda_k(Z(t))]$  [see Eq. (A7)]. Finally, by realizing that

$$\begin{aligned}
 \mathbb{E}[\lambda^{XY}(t) \ln(\lambda^X(t))] &= \mathbb{E}[\mathbb{E}[\lambda^{XY}(t) \ln[\lambda^X(t)] | X_0^t]] \\
 &= \mathbb{E}[\underbrace{\mathbb{E}[\lambda^{XY}(t) | X_0^t]}_{\lambda_k^X(t)} \ln[\lambda^X(t)]] \\
 &= \mathbb{E}[\lambda_k^X(t) \ln[\lambda_k^X(t)]],
 \end{aligned} \tag{A9}$$

we arrive at

$$\begin{aligned}
 I_t^{XY} &= \sum_{k \in R_X} \int_0^t \mathbb{E}[\lambda_k^{XY}(s) \ln[\lambda_k^{XY}(s)] - \lambda_k^X(s) \ln[\lambda_k^X(s)]]ds \\
 &+ \sum_{k \in R_Y} \int_0^t \mathbb{E}[\lambda_k^{XY}(s) \ln[\lambda_k^{XY}(s)] - \lambda_k^Y(s) \ln[\lambda_k^Y(s)]]ds.
 \end{aligned} \tag{A10}$$

Note that even though Eq. (A10) is compact and beneficial for analytical purposes, it is numerically more difficult to handle than Eq. (A6). In particular, it would require evaluating the marginal propensities, calculating Monte Carlo averages over the corresponding  $x \ln x$  terms and numerically integrating the resulting averages over time. Using (A6), by contrast, one first calculates the time integral (for which effective numerical solvers can be used) and subsequently averages over many Monte Carlo samples. We therefore use Eq. (A6) instead for all our numerical simulations.

### APPENDIX B: STOCHASTIC FILTERING

To calculate mutual information between paths  $X_0^t$  and  $Y_0^t$ , we require the marginal propensities  $\lambda_k^{XY}(t)$ ,  $\lambda_k^X(t)$ , and  $\lambda_k^Y(t)$

as we have seen in Appendix A. We will from now on focus on the calculation of  $\lambda_k^X(t)$ , but we remark that  $\lambda_k^{XY}(t)$  and  $\lambda_k^Y(t)$  can be obtained analogously. Recall that a marginal propensity is defined as

$$\begin{aligned}
 \lambda_k^X(t) &= \mathbb{E}[\lambda_k(Z(t)) | X_0^t] = \mathbb{E}[\lambda_k(\bar{Z}(t), X(t)) | X_0^t] \\
 &= \sum_{\bar{z}} \lambda_k(\bar{z}, X(t)) \pi^X(\bar{z}, t),
 \end{aligned} \tag{B1}$$

where  $\bar{Z}(t)$  is a vector collecting all molecular abundances except  $X(t)$ . The average is taken with respect to a conditional probability distribution  $\pi^X(\bar{z}, t) := P(\bar{Z}(t) = \bar{z} | X_0^t)$ , also referred to as a *filtering distribution*. As derived in detail in Ref. [17] the distribution  $\pi^X(\bar{z}, t)$  satisfies the stochastic differential equation

$$\begin{aligned}
 d\pi^X(\bar{z}, t) &= \underbrace{\sum_{k \in R_Z} [\lambda_k(\bar{z} - v_k^{\bar{Z}}, X(t)) \pi^X(\bar{z} - v_k^{\bar{Z}}, t) - \lambda_k(\bar{z}, X(t)) \pi^X(\bar{z}, t)] dt}_{\mathcal{A}^{\bar{Z}|X} \pi^X(\bar{z}, t)} \\
 &- \sum_{k \in R_{\bar{X}}} [\lambda_k(\bar{z}, X(t)) - \lambda_k^X(t)] \pi^X(\bar{z}, t) dt + \sum_{k \in R_{\bar{X}}} \left[ \frac{\lambda_k(\bar{z} - v_k^{\bar{Z}}, X(t))}{\lambda_k^X(t)} \pi^X(\bar{z} - v_k^{\bar{Z}}, t) - \pi^X(\bar{z}, t) \right] dN_k(t),
 \end{aligned} \tag{B2}$$

where  $v_k^{\bar{Z}}$  is the part of the stoichiometric change vector  $v_k$  that acts on  $\bar{Z}(t)$  and  $dN_k(t)$  is the differential version of the counting process  $N_k(t)$ , which is 1 exactly at the time points where reaction  $k$  happens and zero otherwise. The reaction set  $R_Z$  contains all reaction indices that modify exclusively components in  $\bar{Z}(t)$ . Note that in the second sum of Eq. (B2), only reactions in  $R_{\bar{X}}$  that involve species in  $\bar{Z}$  as reactants will have a nonzero contribution and, in the third sum, reactions involving  $\bar{Z}$  as reactants and/or products will have a

nonzero contribution. In all other cases, we have that  $v_k^{\bar{Z}} = 0$  and, furthermore,  $\lambda_k(\bar{z}, X(t)) = \lambda_k(X(t)) = \lambda_k^X(t)$  such that the respective summands cancel out.

The terms in the first row of (B2) can be summarized as  $\mathcal{A}^{\bar{Z}|X} \pi^X(\bar{z}, t)$ , where  $\mathcal{A}^{\bar{Z}|X}$  is a forward operator acting on  $\pi^X(\bar{z}, t)$ . This part is essentially the same as what we encounter on the right-hand side of a conventional master equation with state  $\bar{z}$  and rates  $\lambda_k(\bar{z}, X(t))$  for all  $k \in R_Z$  that depend not only on the internal state  $\bar{z}$ , but also on some

external process  $X(t)$ . More specifically, it describes how the components  $\bar{Z}(t)$  would evolve if the reactions in  $R_{\bar{X}}$  were switched off. The second row accounts for the knowledge about  $\bar{Z}(t)$  that is gained through observing (or not observing) reactions in  $R_{\bar{X}}$ . Concrete examples will be given later in this Appendix.

### 1. Moment approximations

A direct numerical solution of (B2) is computationally demanding due to the combinatorial explosion of states in larger reaction networks. More effective solutions of (B2) can be obtained using moment-closure techniques as we have also shown previously [11,17,18]. This technique is suitable when the reaction propensities are of polynomial form such as encountered with mass-action kinetics. If this is the case, Eq. (B2) can be readily transformed into an equivalent system of moment equations, which is subsequently truncated to obtain a finite-dimensional moment hierarchy. The first step can be achieved by multiplying (B2) with a polynomial in  $\bar{z}$  (corresponding to a certain desired moment) and summing over all possible values of  $\bar{z}$ . This generally leads to an infinite-dimensional system of moment equations which needs to be truncated at a certain order. This can be achieved by assuming the underlying distribution  $\pi^X(\bar{z}, t)$  to belong to a certain family of probability distributions, which can be described by a finite (and small) number of degrees of freedom (e.g., mean  $\mu$  and standard deviation  $\sigma$  in the case of a univariate Gaussian distribution). This can then be exploited to express higher-order moments as a function of lower-order moments, leading to a closed system of differential equations. In our previous works [11,18], we have used a third-order Gamma closure to approximate conditional moments of a one-dimensional filtering equation and have found very good accuracy. Since we consider multidimensional filtering equations in this paper, we analogously use a multivariate extension of the third-order Gamma closure [26], for which

$$\mathbb{E}[\bar{Z}_j^2 \bar{Z}_l | X_0^t] = 2 \frac{\mathbb{E}[\bar{Z}_j^2 | X_0^t] \mathbb{E}[\bar{Z}_j \bar{Z}_l | X_0^t]}{\mathbb{E}[\bar{Z}_j | X_0^t]} - \mathbb{E}[\bar{Z}_j^2 | X_0^t] \mathbb{E}[\bar{Z}_l | X_0^t]. \quad (\text{B3})$$

Note that (B3) agrees with the univariate Gamma closure if  $j = l$ . A concrete application of the multivariate Gamma closure will be presented later in Appendix C 1.

Figure 4 shows example trajectories of the first and second conditional moments obtained by direct numerical integration of (B2) and the Gamma closure.

### 2. Analytical approximation

In this section, we derive a simple analytical approximation of the path mutual information as shown in Eq. (6) in the main text. For that purpose, we focus on a simplified scenario where all reactions have linear propensities (i.e., they are unimolecular in the case of mass-action kinetics) and where no other species change simultaneously with X or Y. As can be seen from Eq. (A10), the path mutual information involves inner expectations that are conditional on paths  $X_0^t, Y_0^t$ , and  $\{X_0^t, Y_0^t\}$ , as well as an outer expectation that integrates over different realizations of these paths. In our moment approximation, the inner expectations are handled

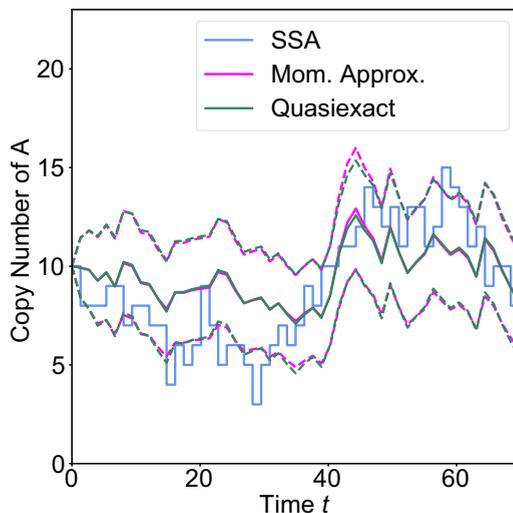


FIG. 4. Solution of the filtering equation obtained by direct integration on a finite grid (*quasixact method*) and moment closure (*moment-approximation method*). The considered example is a two-node feedforward network, for which we calculated  $P(A(t) = a | B_0^t)$  (see Appendix C 1). Shown are means (solid lines) and standard deviations (dashed lines) as well as the underlying realization of  $A(t)$  (blue stair plot). Parameters were set to  $\{k_1, k_2, k_3, k_4\} = \{1, 0.1, 0.1, 0.1\}$  and  $[A_0, B_0] = \{10, 10\}$ . Mom. Approx. moment-approximation; SSA, stochastic simulation algorithm.

by finding approximate moments of the underlying filtering distribution, whereas the outer expectations are obtained by averaging over many (exact) stochastic simulations. To obtain analytical results, we further expand the  $x \ln x$  terms inside the outer expectations into a Taylor series. In particular, we obtain

$$\lambda \ln(\lambda) = \tilde{\lambda} \ln(\tilde{\lambda}) + (\ln(\tilde{\lambda}) + 1)(\lambda - \tilde{\lambda}) + \frac{(\lambda - \tilde{\lambda})^2}{2\tilde{\lambda}} + O(\lambda^3), \quad (\text{B4})$$

where  $\tilde{\lambda}$  is the point at which we perform the expansion. If we now consider  $\lambda$  to be a random variable, set  $\tilde{\lambda} = \mathbb{E}[\lambda]$ , and take the expectation, we find

$$\begin{aligned} \mathbb{E}[\lambda \ln(\lambda)] &\approx \mathbb{E}[\lambda] \ln(\mathbb{E}[\lambda]) \\ &+ (\ln(\mathbb{E}[\lambda]) + 1)(\mathbb{E}[\lambda] - \mathbb{E}[\lambda]) \\ &+ \frac{\mathbb{E}[(\lambda - \mathbb{E}[\lambda])^2]}{2\mathbb{E}[\lambda]}, \end{aligned} \quad (\text{B5})$$

which simplifies to

$$\mathbb{E}[\lambda \ln(\lambda)] \approx \mathbb{E}[\lambda] \ln(\mathbb{E}[\lambda]) + \frac{\text{Var}[\lambda]}{2\mathbb{E}[\lambda]}. \quad (\text{B6})$$

Applying this approximation to (A10) for the considered scenario, we finally obtain

$$\begin{aligned} I_t^{XY} &\approx \sum_{k \in R_X} \int_0^t \frac{\text{Var}[\lambda_k^{XY}(t)] - \text{Var}[\lambda_k^X(t)]}{2\mathbb{E}[\lambda_k(Z(t))]} ds \\ &+ \sum_{k \in R_Y} \int_0^t \frac{\text{Var}[\lambda_k^{XY}(t)] - \text{Var}[\lambda_k^Y(t)]}{2\mathbb{E}[\lambda_k(Z(t))]} ds. \end{aligned} \quad (\text{B7})$$

This shows that up to second order, the path mutual information can be approximated by the variances of the marginal propensities such as  $\text{Var}[\lambda_k^{XY}(t)] = \text{Var}[\mathbb{E}[\lambda_k(Z(t)) | X_0^t, Y_0^t]]$ . The general idea is then to derive ordinary differential equations for these variances from the previously obtained conditional moment equations. In the following, we show how such equations can be derived. Concrete examples will be given later in Appendix C 1.

For the sake of illustration, we consider the variance of the marginal propensity  $\lambda_k^X(t) = \mathbb{E}[\lambda_k(Z(t)) | X_0^t]$ . Since all reactions have linear rates, we have  $\lambda_k(Z(t)) = c_k \bar{Z}_{j(k)}(t)$  with  $\bar{Z}_{j(k)}(t)$  being the copy number of the particular species in  $\bar{Z}$  that drives reaction  $k$  and  $c_k$  being a constant. As a consequence of the linear propensity functions, finding the marginal propensity  $\lambda_k^X(t)$  is equivalent to finding the first conditional moment of  $\bar{Z}_{j(k)}(t)$ , which for compactness we denote by  $M_k(t) = \mathbb{E}[\bar{Z}_{j(k)}(t) | X_0^t]$  in the following. Note that  $M_k(t)$  is a functional of the path  $X_0^t$  and as such is stochastic. The goal is now to find the variance of this stochastic process  $\text{Var}[M_k(t)] = \text{Var}[\mathbb{E}[\bar{Z}_{j(k)}(t) | X_0^t]]$  from which we subsequently obtain  $\text{Var}[\lambda_k^X(t)] = c_k^2 \text{Var}[M_k(t)]$ .

The variance of  $M_k(t)$  can be written as  $\text{Var}[M_k(t)] = \mathbb{E}[M_k(t)^2] - \mathbb{E}[M_k(t)]^2 = \mathbb{E}[M_k(t)^2] - \mathbb{E}[\bar{Z}_{j(k)}(t)]^2$ . Therefore, since the average  $\mathbb{E}[\bar{Z}_{j(k)}(t)]$  is straightforward to obtain for linear networks, the calculation of the variance boils down to finding the second noncentral moment of  $M_k(t)$ . To this end, we consider the differential equation of  $M_k(t)$ , which can be obtained by multiplying (B2) with  $\bar{z}_{j(k)}$  and summing over all  $\bar{z}$ :

$$\begin{aligned} dM_k(t) &= \left( D_k(t) - \sum_{i \in R_X} c_i (M_{i,k}(t) - M_i(t)M_k(t)) \right) dt \\ &+ \sum_{i \in R_X} \frac{M_{i,k}(t) - M_i(t)M_k(t)}{M_i(t)} dN_i(t) \\ &= \left( D_k(t) - \sum_{i \in R_X} c_i C_{i,k}(t) \right) dt + \sum_{i \in R_X} \frac{C_{i,k}(t)}{M_i(t)} dN_i(t), \end{aligned} \quad (\text{B8})$$

where  $D_k(t) = \sum_{\bar{z}_{j(k)}} \mathcal{A}^{\bar{z}_{j(k)}} \pi^X(\bar{z}, t) z_{j(k)}$  summarizes the fluxes originating from the forward operator  $\mathcal{A}^{\bar{z}_{j(k)}}$  and where we have defined the second noncentral conditional moment  $M_{i,k}(t) := \mathbb{E}[Z_{j(i)}(t)Z_{j(k)}(t) | X_0^t]$  as well as the conditional covariance  $C_{i,k}(t) := \text{Cov}[Z_{j(i)}(t), Z_{j(k)}(t) | X_0^t] = M_{i,k}(t) - M_i(t)M_k(t)$ . To obtain an equation for the second moment of  $M_k(t)$ , we first derive an equation for  $M_k(t)^2$  using Ito's rule for counting processes [34], i.e.,

$$d(M_k(t)^2) = 2M_k(t)dM_k(t) + \sum_{i \in R_X} \frac{C_{i,k}(t)^2}{M_i(t)^2} dN_i(t). \quad (\text{B9})$$

Decomposing the reaction counters  $dN_i(t)$  into a predictable part and a martingale, i.e.,  $dN_i(t) = c_i M_i(t) dt + d\tilde{N}_i(t)$ , and

inserting Eq. (B8), we further obtain

$$\begin{aligned} d(M_k(t)^2) &= 2M_k(t)D_k(t)dt - \sum_{i \in R_X} 2c_i M_k(t)C_{i,k}(t)dt \\ &+ \sum_{i \in R_X} 2c_i M_k(t)C_{i,k}(t)dt \\ &+ \sum_{i \in R_X} 2M_k(t) \frac{C_{i,k}(t)}{M_i(t)} d\tilde{N}_i(t) \\ &+ \sum_{i \in R_X} c_i \frac{C_{i,k}(t)^2}{M_i(t)} dt + \sum_{i \in R_X} \frac{C_{i,k}(t)^2}{M_i(t)^2} d\tilde{N}_i(t) \\ &= 2M_k(t)D_k(t)dt + \sum_{i \in R_X} 2M_k(t) \frac{C_{i,k}(t)}{M_i(t)} d\tilde{N}_i(t) \\ &+ \sum_{i \in R_X} c_i \frac{C_{i,k}(t)^2}{M_i(t)} dt + \sum_{i \in R_X} \frac{C_{i,k}(t)^2}{M_i(t)^2} d\tilde{N}_i(t). \end{aligned} \quad (\text{B10})$$

By taking the expectation, all terms involving the martingale increments  $d\tilde{N}_i(t)$  vanish, which after some simplifications leads to

$$d\mathbb{E}[M_k(t)^2] = 2\mathbb{E}[M_k(t)D_k(t)]dt + \sum_{i \in R_X} c_i \mathbb{E} \left[ \frac{C_{i,k}(t)^2}{M_i(t)} \right] dt. \quad (\text{B11})$$

Since the considered network architecture has linear propensities, the first term on the right-hand side of (B11) will result in moments of at most order 2, regardless of the specific form of  $D_k(t)$ . However, the second term involves a more complex expectation that is in general difficult to calculate. This expectation can be approximated when the underlying filtering distribution is one dimensional (i.e.,  $\bar{z}$  is a scalar that drives a single reaction  $\bar{Z} \rightarrow \bar{Z} + X$ ) and assumed to be a Gamma distribution. In this case, the sum in (B11) involves only a single term of the form  $\mathbb{E}[C_{i,i}(t)^2/M_i(t)]$ , whereas  $C_{i,i}(t)$  is a conditional variance. Based on the assumption of a Gamma distribution, we have previously suggested an approximation of the form  $\mathbb{E}[C_{i,i}(t)^2/M_i(t)] \approx \mathbb{E}[C_{i,i}(t)]^2/\mathbb{E}[M_i(t)]$  [18]. Inspired by this, and in line with the multivariate Gamma closure (B3), we thus use the approximation  $\mathbb{E}[C_{i,k}(t)^2/M_i(t)] \approx \mathbb{E}[C_{i,k}(t)]^2/\mathbb{E}[M_i(t)]$  to close Eq. (B11). While this approximation is so far heuristic, it yielded accurate results in the case studies we considered.

## APPENDIX C: CASE STUDIES

### 1. Case study I

**Path mutual information.** We consider a three-node feed-forward network as presented in the main text. To obtain an expression for the path mutual information between species A and C, we realize that species C receives information about species A only through reaction  $R_C = \{5\}$ , which depends on intermediate species B. Using Eq. (A10), the path mutual

information takes the form

$$I_t^{AC} = \mathbb{E} \left[ \int_0^t k_5 (\mathbb{E}[B(s) | A_0^s, C_0^s] \ln [k_5 \mathbb{E}[B(s) | A_0^s, C_0^s]] - \mathbb{E}[B(s) | C_0^s] \ln [k_5 \mathbb{E}[B(s) | C_0^s]]) ds \right]. \quad (\text{C1})$$

In this case, two filtering equations are required, one that describes  $B(t)$  conditionally on  $A_0^t$  and  $C_0^t$  and one that describes  $(A(t), B(t))$  conditionally on  $C_0^t$ . Based on Eq. (B2) these two filtering equations are given by

$$\begin{aligned} d\pi^{AC}(b, t) = & k_3 A(t) (\pi^{AC}(b-1, t) - \pi^{AC}(b, t)) dt \\ & + k_4 ((b+1)\pi^{AC}(b+1, t) - b\pi^{AC}(b, t)) dt \\ & - k_5 (b - \mathbb{E}[B(t) | A_0^t, C_0^t]) \pi^{AC}(b, t) dt \\ & + \frac{b - \mathbb{E}[B(t) | A_0^t, C_0^t]}{\mathbb{E}[B(t) | A_0^t, C_0^t]} \pi^{AC}(b, t) dN_5(t) \end{aligned} \quad (\text{C2})$$

and

$$\begin{aligned} d\pi^C(a, b, t) = & k_1 (\pi^C(a-1, b, t) - \pi^C(a, b, t)) dt \\ & + k_2 ((a+1)\pi^C(a+1, b, t) - a\pi^C(a, b, t)) dt \\ & + k_3 a (\pi^C(a, b-1, t) - \pi^C(a, b, t)) dt \\ & + k_4 ((b+1)\pi^C(a, b+1, t) - b\pi^C(a, b, t)) dt \\ & - k_5 (b - \mathbb{E}[B(t) | C_0^t]) \pi^C(a, b, t) dt \\ & + \frac{b - \mathbb{E}[B(t) | C_0^t]}{\mathbb{E}[B(t) | C_0^t]} \pi^C(a, b, t) dN_5(t), \end{aligned} \quad (\text{C3})$$

which in principle can be solved numerically on a finite grid using the quasiexact method. Moreover, differential equations can be obtained from these equations for the conditional means  $\mathbb{E}[B(t) | A_0^t, C_0^t]$  and  $\mathbb{E}[B(t) | C_0^t]$  which will, however, depend on moments of higher order. In the case of Eq. (C2), for instance, we obtain for the first two conditional moments

$$\begin{aligned} d\mathbb{E}[B(t) | A_0^t, C_0^t] = & k_3 A(t) dt - k_4 \mathbb{E}[B(t) | A_0^t, C_0^t] dt - k_5 \mathbb{E}[B(t)^2 | A_0^t, C_0^t] dt + k_5 \mathbb{E}[B(t) | A_0^t, C_0^t]^2 dt \\ & + \frac{\mathbb{E}[B(t)^2 | A_0^t, C_0^t] - \mathbb{E}[B(t) | A_0^t, C_0^t]^2}{\mathbb{E}[B(t) | A_0^t, C_0^t]} dN_5(t), \\ d\mathbb{E}[B(t)^2 | A_0^t, C_0^t] = & 2k_3 A(t) \mathbb{E}[B(t) | A_0^t, C_0^t] dt + k_3 A(t) dt - 2k_4 \mathbb{E}[B(t)^2 | A_0^t, C_0^t] dt + k_4 \mathbb{E}[B(t) | A_0^t, C_0^t] dt \\ & - k_5 \mathbb{E}[B(t)^3 | A_0^t, C_0^t] dt + k_5 \mathbb{E}[B(t)^2 | A_0^t, C_0^t] \mathbb{E}[B(t) | A_0^t, C_0^t] dt \\ & + \frac{\mathbb{E}[B(t)^3 | A_0^t, C_0^t] - \mathbb{E}[B(t)^2 | A_0^t, C_0^t] \mathbb{E}[B(t) | A_0^t, C_0^t]}{\mathbb{E}[B(t) | A_0^t, C_0^t]} dN_5(t). \end{aligned} \quad (\text{C4})$$

This implies that one would need an infinite amount of differential equations in order to calculate the moment dynamics. Applying a third-order Gamma closure allows us replace the third conditional moment  $\mathbb{E}[B(t)^3 | A_0^t, C_0^t]$  by moments of order 1 and 2 using (B3). Then, the closed equation for the second moment becomes

$$\begin{aligned} d\mathbb{E}[B(t)^2 | A_0^t, C_0^t] = & 2k_3 A(t) \mathbb{E}[B(t) | A_0^t, C_0^t] dt + k_3 A(t) dt - 2k_4 \mathbb{E}[B(t)^2 | A_0^t, C_0^t] dt + k_4 \mathbb{E}[B(t) | A_0^t, C_0^t] dt \\ & - 2k_5 \frac{\mathbb{E}[B(t)^2 | A_0^t, C_0^t]^2}{\mathbb{E}[B(t) | A_0^t, C_0^t]} dt + 2k_5 \mathbb{E}[B(t) | A_0^t, C_0^t] \mathbb{E}[B(t)^2 | A_0^t, C_0^t] dt \\ & + 2 \frac{\mathbb{E}[B(t)^2 | A_0^t, C_0^t]^2}{\mathbb{E}[B(t) | A_0^t, C_0^t]^2} dN_5(t) - 2\mathbb{E}[B(t)^2 | A_0^t, C_0^t] dN_5(t). \end{aligned} \quad (\text{C5})$$

In the same way, we can get equations for the set of differential equations conditionally on  $C_0^t$  to evaluate  $\mathbb{E}[B(t) | C_0^t]$ .

In this example, we also made use of the analytical approximation of the path mutual information [i.e., Eq. (B7)], by which we obtain

$$I_t^{AC} \approx \int_0^t \frac{k_5}{2} \frac{\text{Var}[\mathbb{E}[B(s) | A_0^s, C_0^s]] - \text{Var}[\mathbb{E}[B(s) | C_0^s]]}{\mathbb{E}[B(s)]} ds. \quad (\text{C6})$$

In order to demonstrate how to obtain the variances of the respective conditional means, we derive  $\text{Var}[\mathbb{E}[B(s) | A_0^s, C_0^s]]$  as discussed in Appendix B 2. We start with defining the differential

$$d\text{Var}[\mathbb{E}[B(t) | A_0^t, C_0^t]] = d\mathbb{E}[\mathbb{E}[B(t) | A_0^t, C_0^t]^2] - d\mathbb{E}[\mathbb{E}[B(t) | A_0^t, C_0^t]]^2 = d\mathbb{E}[\mathbb{E}[B(t) | A_0^t, C_0^t]^2] - 2\mathbb{E}[B(t)] d\mathbb{E}[B(t)].$$

The dynamics of  $\mathbb{E}[B(t)]$  can be derived from a conventional master equation and is given by

$$d\mathbb{E}[B(t)] = (k_3\mathbb{E}[A(t)] - k_4\mathbb{E}[B(t)])dt. \tag{C7}$$

Using Eq. (B11) and realizing that

$$\mathbb{E}[A(t)\mathbb{E}[B(t)|A_0^t, C_0^t]] = \mathbb{E}[\mathbb{E}[A(t)B(t)|A_0^t, C_0^t]] = \mathbb{E}[A(t)B(t)], \tag{C8}$$

we get an approximate expression for the differential of the squared moment

$$\frac{d}{dt}\mathbb{E}[\mathbb{E}[B(t)|A_0^t, C_0^t]^2] = 2k_3\mathbb{E}[A(t)B(t)] - 2k_4\mathbb{E}[\mathbb{E}[B(t)|A_0^t, C_0^t]^2] + k_5\frac{\mathbb{E}[\text{Var}[B(t)|A_0^t, C_0^t]]^2}{\mathbb{E}[B(t)]}, \tag{C9}$$

where we have used the approximation  $\mathbb{E}[\text{Var}[B(t)|A_0^t, C_0^t]^2]/\mathbb{E}[B(t)|A_0^t, C_0^t] \approx \mathbb{E}[\text{Var}[B(t)|A_0^t, C_0^t]^2]/\mathbb{E}[B(t)]$ . Now, we can combine Eqs. (C7) and (C9) and by subtracting  $2\mathbb{E}[B(t)]d\mathbb{E}[B(t)]/dt = 2\mathbb{E}[\mathbb{E}[B(t)|A_0^t, C_0^t]]d\mathbb{E}[\mathbb{E}[B(t)|A_0^t, C_0^t]]/dt$ , we find that the requested variance satisfies the differential equation

$$\frac{d}{dt}\text{Var}[\mathbb{E}[B(t)|A_0^t, C_0^t]] = 2k_3\text{Cov}[A(t), B(t)] - 2k_4\text{Var}[\mathbb{E}[B(t)|A_0^t, C_0^t]] + k_5\frac{\mathbb{E}[\text{Var}[B(t)|A_0^t, C_0^t]]^2}{\mathbb{E}[B(t)]}. \tag{C10}$$

Equation (C10) involves additional moments, for which equations can be obtained in a similar way. In total, this leads to a system of differential equations

$$\begin{aligned} \frac{d}{dt}\text{Var}[\mathbb{E}[B(t)|A_0^t, C_0^t]] &= 2k_3\text{Cov}[A(t), B(t)] - 2k_4\text{Var}[\mathbb{E}[B(t)|A_0^t, C_0^t]] + k_5\frac{\mathbb{E}[\text{Var}[B(t)|A_0^t, C_0^t]]^2}{\mathbb{E}[B(t)]}, \\ \frac{d}{dt}\mathbb{E}[A(t)] &= k_1 - k_2\mathbb{E}[A(t)], \\ \frac{d}{dt}\mathbb{E}[B(t)] &= k_3\mathbb{E}[A(t)] - k_4\mathbb{E}[B(t)], \\ \frac{d}{dt}\mathbb{E}[\text{Var}[B(t)|A_0^t, C_0^t]] &= k_3\mathbb{E}[A(t)] + k_4\mathbb{E}[B(t)] - 2k_4\mathbb{E}[\text{Var}[B(t)|A_0^t, C_0^t]] - k_5\frac{\mathbb{E}[\text{Var}[B(t)|A_0^t, C_0^t]]^2}{\mathbb{E}[B(t)]}, \\ \frac{d}{dt}\text{Cov}[A(t), B(t)] &= -(k_2 + k_4)\text{Cov}[A(t), B(t)] + k_3\text{Var}[A(t)], \\ \frac{d}{dt}\text{Var}[A(t)] &= k_1 + k_2\mathbb{E}[A(t)] - 2k_2\text{Var}[A(t)], \end{aligned} \tag{C11}$$

whose steady state can be found by setting the left-hand side to zero. The same procedure can be followed for calculating the variance  $\text{Var}[\mathbb{E}[B(t) | C_0^t]]$ , which in combination with  $\text{Var}[\mathbb{E}[B(t) | A_0^t, C_0^t]]$  allows us to approximate the path mutual information via Eq. (C6) at steady state. Note that for stationary systems, the path mutual information rate can be obtained simply by dropping the time integral in (C6). Doing so, and considering the limit  $v_c \rightarrow \infty$  (or equivalently  $k_5$ ), leads to the steady-state mutual information rate shown in Eq. (9) in the main text.

In Fig. 2 in the main text, we compared the analytical approximation with the moment-approximation method and found very good agreement. However, since the moment-approximation method is also approximate, we carried out a similar comparison with the quasixact method (Fig. 5). Also in this case, we found very good agreement between the two approaches. Note that due to the computational complexity of the quasixact method, this analysis was restricted to a smaller parameter region than the one considered in Fig. 2 in the main text.

In the main text, we compared the mutual information rate between species A and C with the mutual information rate between A and B. According to Eq. (A10), the latter is given by

$$I_t^{AB} = \mathbb{E}\left[\int_0^t k_3(A(s) \ln[k_3A(s)] - \mathbb{E}[A(s)|B_0^s] \ln[k_3\mathbb{E}[A(s)|B_0^s]])ds\right], \tag{C12}$$

where information transfer is mediated through reaction  $R_B = \{3\}$ . Note that the conditional expectations vanish in the first term on the right-hand side of (C12), because conditioning on  $\{A_0^t, B_0^t\}$  provides complete knowledge about the history of the network (i.e.,  $\mathbb{E}[A(t) | A_0^t, B_0^t] = A(t)$ ). The analytical approximation of the path mutual information then becomes

$$I_t^{AB} \approx \int_0^t \frac{k_3}{2} \frac{\text{Var}[A(s)] - \text{Var}[\mathbb{E}[A(s)|B_0^s]]}{\mathbb{E}[A(s)]} ds. \tag{C13}$$

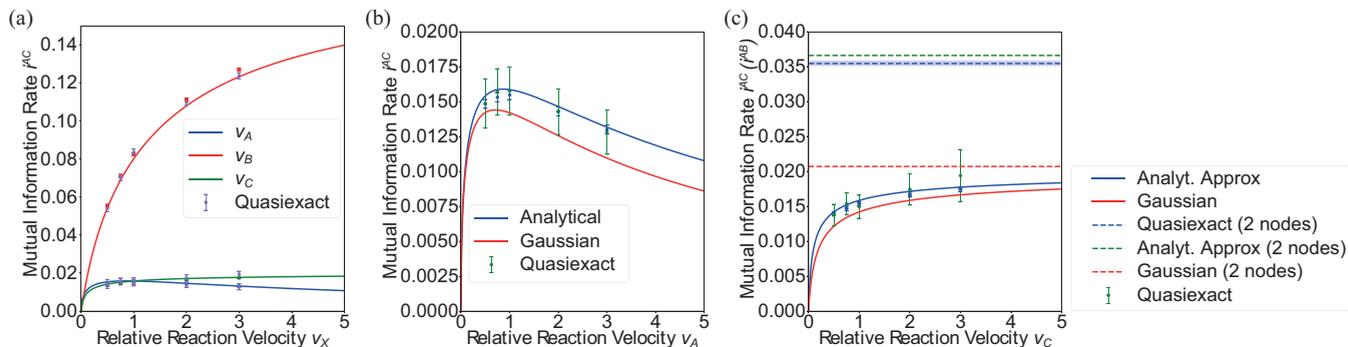


FIG. 5. (a) Analytical approximation of the path mutual information rate  $i^{AC}$  as a function of the relative reaction velocity  $v_\Omega$ ,  $\Omega \in [A, B, C]$  and comparison with the quasiexact solution (purple points). (b) Path mutual information rate  $i^{AC}$  as a function of  $v_A$ . The quasiexact solution is given by the green points. (c) Path mutual information rate  $i^{AC}$  as a function of  $v_C$  for the network consisting of three nodes and the network consisting of two nodes. The simulations are performed with the same set of parameters as for Fig. 2 in the main text. The quasiexact calculation is performed with a sample size of  $n = 600$ . Results obtained by the moment-approximation method ( $n = 10\,000$ ) are shown as points in the same color as the corresponding analytical approximation. The grid sizes for the quasiexact method for  $\pi^{AC}$  and  $\pi^C$  were chosen to be  $50 \times 50$ , respectively.

At stationarity, the path mutual information rate is then given by the integrand of Eq. (C13). Moreover, by the law of total variance, it holds that  $\text{Var}[A(s)] - \text{Var}[\mathbb{E}[A(s) | B_0^s]] = \mathbb{E}[\text{Var}[A(t) | B_0^t]]$ , such that we obtain

$$\lim_{t \rightarrow \infty} i_t^{AB} = i^{AB} = \frac{k_3}{2} \frac{\mathbb{E}[\text{Var}[A(t) | B_0^t]]}{\mathbb{E}[A(t)]}. \quad (\text{C14})$$

The steady state of  $\mathbb{E}[A(t)]$  can be obtained directly from the underlying master equation and is given by  $\lim_{t \rightarrow \infty} \mathbb{E}[A(t)] = k_1/k_2$ . Therefore the only term that needs to be determined is the expected conditional variance  $\mathbb{E}[\text{Var}[A(t) | B_0^t]]$  that satisfies the differential equation

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[\text{Var}[A(t) | B_0^t]] &= k_1 + k_2 \mathbb{E}[A(t)] - 2k_2 \mathbb{E}[\text{Var}[A(t) | B_0^t]] \\ &\quad - k_3 \mathbb{E} \left[ \frac{\text{Var}[A(t) | B_0^t]^2}{\mathbb{E}[A(t) | B_0^t]} \right], \end{aligned} \quad (\text{C15})$$

which can be derived using Ito's rule for counting processes as demonstrated before. Making use of the approximation  $\mathbb{E}[\text{Var}[A(t) | B_0^t]^2 / \mathbb{E}[A(t) | B_0^t]] \approx \mathbb{E}[\text{Var}[A(t) | B_0^t]^2] / \mathbb{E}[A(t)]$ , solving for the stationary solution, and inserting it into (C14) finally yields

$$i^{AB} = -\frac{k_2}{2} + \frac{1}{2} \sqrt{k_2(k_2 + 2k_3)}. \quad (\text{C16})$$

*Gaussian approximations.* For comparison, we calculated stationary mutual information rates using Gaussian theory as proposed by Tostevin and ten Wolde [8]. To this end, we employ the relationship

$$i_G^{XY} = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \ln \left[ 1 - \frac{|S_{XY}(\omega)|^2}{S_{XX}(\omega)S_{YY}(\omega)} \right] d\omega, \quad (\text{C17})$$

where  $S_{XY}(\omega)$  is the cross power spectral density of signals  $X(t)$  and  $Y(t)$  and  $S_{XX}(\omega)$  and  $S_{YY}(\omega)$  are the power spectral densities of  $X(t)$  and  $Y(t)$ , respectively. Since all propensities in the considered two- and three-node networks are linear,  $S_{XY}(\omega)$ ,  $S_{XX}(\omega)$ , and  $S_{YY}(\omega)$  can be calculated in closed form by first deriving the respective cross covariance and autocovariance functions and calculating their Fourier transforms.

For the considered two-node network, we obtain

$$\begin{aligned} \frac{d\text{Cov}[A(0), A(t)]}{dt} &= -k_2 \text{Cov}[A(0), A(t)], \\ \frac{d\text{Cov}[A(0), B(t)]}{dt} &= k_3 \text{Cov}[A(0), A(t)] - k_4 \text{Cov}[A(0), B(t)], \\ \frac{d\text{Cov}[B(0), B(t)]}{dt} &= k_3 \text{Cov}[B(0), A(t)] - k_4 \text{Cov}[B(0), B(t)], \\ \frac{d\text{Cov}[B(0), A(t)]}{dt} &= -k_2 \text{Cov}[B(0), A(t)]. \end{aligned} \quad (\text{C18})$$

For the three-node network, we obtain

$$\begin{aligned} \frac{d\text{Cov}[A(0), A(t)]}{dt} &= -k_2 \text{Cov}[A(0), A(t)], \\ \frac{d\text{Cov}[A(0), B(t)]}{dt} &= k_3 \text{Cov}[A(0), A(t)] - k_4 \text{Cov}[A(0), B(t)], \\ \frac{d\text{Cov}[A(0), C(t)]}{dt} &= k_5 \text{Cov}[A(0), B(t)] - k_6 \text{Cov}[A(0), C(t)], \\ \frac{d\text{Cov}[C(0), A(t)]}{dt} &= -k_2 \text{Cov}[C(0), A(t)], \\ \frac{d\text{Cov}[C(0), A(t)]}{dt} &= k_3 \text{Cov}[C(0), A(t)] - k_4 \text{Cov}[C(0), B(t)], \\ \frac{d\text{Cov}[C(0), C(t)]}{dt} &= k_5 \text{Cov}[C(0), B(t)] - k_6 \text{Cov}[C(0), C(t)]. \end{aligned} \quad (\text{C19})$$

Equations (C18) and (C19) define systems of ordinary differential equations, which can be Fourier transformed directly, giving rise to systems of linear algebraic equations. Solving these equations yields  $S_{AB}(\omega)$ ,  $S_{AA}(\omega)$ , and  $S_{BB}(\omega)$  in the case of the two-node network and  $S_{AC}(\omega)$ ,  $S_{AA}(\omega)$  and  $S_{CC}(\omega)$  in the case of the three-node network, respectively. In that way, the fraction  $F_{XY}(\omega) = |S_{XY}(\omega)|^2 / (S_{XX}(\omega)S_{YY}(\omega))$  inside (C17) becomes

$$F_{AB}(\omega) = \frac{k_2 k_3}{\omega^2 + k_2(k_2 + k_3)}, \quad (\text{C20})$$

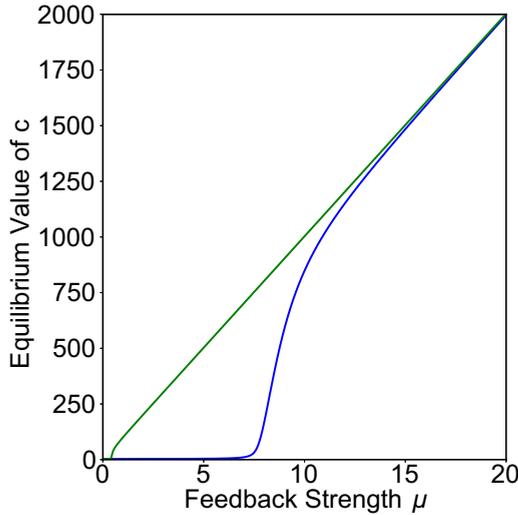


FIG. 6. Equilibrium points of the positive-feedback system as a function of  $\mu$  predicted by mean-field theory. The bifurcation point is at  $\mu \approx 0.35$ .

in the case of the two-node network and

$$F_{AC}(\omega) = \frac{k_2 k_3 k_4 k_5}{k_2 k_4 (k_3 k_5 + k_2 (k_4 + k_5)) + (k_2^2 + k_4 (k_4 + k_5)) \omega^2 + \omega^4} \tag{C21}$$

in the case of the three-node network. In the former case, the remaining integral in (C17) can be solved analytically, leading to

$$i_G^{AB} = -\frac{k_2}{2} + \frac{1}{2} \sqrt{k_2 (k_2 + k_3)}. \tag{C22}$$

Taking the limit  $k_5 \rightarrow \infty$  in (C21) directly leads to (C20), showing that  $\lim_{k_5 \rightarrow \infty} i_G^{AC} = i_G^{AB}$ . In other words, the three-node network reduces to the simpler two-node network in terms of information transfer as  $k_5 \rightarrow \infty$ . Note that in the case of the three-node network, inserting (C21) into (C17) led to an integral which could not be solved analytically, which is why an explicit expression of  $i_G^{AC}$  could not be provided. For the results shown in Figs. 2(b)–2(d) in the main text, this integral was solved numerically.

### 2. Case study II

The second system considered in the main text is a bistable switch. The system is similar to a three-node feedforward network, but contains a positive feedback between species C and A. This feedback is incorporated by letting  $k_1$  depend on the abundance of C. In particular, we chose a Hill-type propensity of the form  $k_1(C(t)) = \mu C(t)^{n_H} / (K^{n_H} + C(t)^{n_H}) + \epsilon$  with Hill coefficient  $n_H = 3$  and  $\mu$  and  $\epsilon$  being positive constants. Using Eq. (A10), the path mutual information between A and C takes the form

$$I_t^{AC} = \mathbb{E} \left[ \int_0^t k_1(C(s)) \ln[k_1(C(s))] - \mathbb{E}[k_1(C(s)) | A_0^s] \ln[\mathbb{E}[k_1(C(s)) | A_0^s]] ds \right]$$

$$+ k_5 \int_0^t \mathbb{E}[B(s) | A_0^s, C_0^s] \ln[\mathbb{E}[B(s) | A_0^s, C_0^s]] - \mathbb{E}[B(s) | C_0^s] \ln[\mathbb{E}[B(s) | C_0^s]] ds. \tag{C23}$$

In contrast to the previous cases, information transmission is now bidirectional, mediated through reactions  $R_A = \{1\}$  and  $R_C = \{5\}$ . In this case, we require three filtering equations to calculate the path mutual information:

$$d\pi^{AC}(b, t) = k_3 A(t) (\pi^{AC}(b-1, t) - \pi^{AC}(b, t)) dt + k_4 ((b+1)\pi^{AC}(b+1, t) - b\pi^{AC}(b, t)) dt - k_5 (b - \mathbb{E}[B(t) | A_0^t, C_0^t]) \pi^{AC}(b, t) dt + \frac{b - \mathbb{E}[B(t) | A_0^t, C_0^t]}{\mathbb{E}[B(t) | A_0^t, C_0^t]} \pi^{AC}(b, t) dN_5(t), \tag{C24}$$

$$d\pi^A(b, c, t) = k_3 A(t) (\pi^A(b-1, c, t) - \pi^A(b, c, t)) dt + k_4 ((b+1)\pi^A(b+1, c, t) - b\pi^A(b, c, t)) dt + k_5 b (\pi^A(b, c-1, t) - \pi^A(b, c, t)) dt + k_6 ((c+1)\pi^A(b, c+1, t) - c\pi^A(b, c, t)) dt - [k_1(c) - \mathbb{E}[k_1(C(t)) | A_0^t]] \pi^A(b, c, t) dt + \frac{k_1(c) - \mathbb{E}[k_1(C(t)) | A_0^t]}{\mathbb{E}[k_1(C(t)) | A_0^t]} \pi^A(b, c, t) dN_1(t), \tag{C25}$$

$$d\pi^C(a, b, t) = k_1(C(t)) (\pi^C(a-1, b, t) - \pi^C(a, b, t)) dt + k_2 ((a+1)\pi^C(a+1, b, t) - a\pi^C(a, b, t)) dt + k_3 a (\pi^C(a, b-1, t) - \pi^C(a, b, t)) dt + k_4 ((b+1)\pi^C(a, b+1, t) - b\pi^C(a, b, t)) dt - k_5 (b - \mathbb{E}[B(t) | C_0^t]) \pi^C(a, b, t) dt + \frac{b - \mathbb{E}[B(t) | C_0^t]}{\mathbb{E}[B(t) | C_0^t]} \pi^C(a, b, t) dN_5(t). \tag{C26}$$

To calculate the path mutual information between species A and C for various parameter regimes, we made use of the moment-approximation method. However, since  $k_1(C(t))$  is not of polynomial form, this method is not directly applicable to the third filtering equation, Eq. (C25), because the right-hand side of the resulting moment dynamics would involve expectations of rational functions (and thus no longer depend on moments only). To address this issue, we linearize  $k_1(C(t))$  around the conditional expectation  $\mathbb{E}[C(t) | A_0^t]$ , which yields

$$k_1(C(t)) \approx \frac{\mu \mathbb{E}[C(t) | A_0^t]^3}{K^3 + \mathbb{E}[C(t) | A_0^t]^3} + \epsilon + \left( \frac{3\mu \mathbb{E}[C(t) | A_0^t]^2}{K^3 + \mathbb{E}[C(t) | A_0^t]^3} - \frac{3\mu \mathbb{E}[C(t) | A_0^t]^5}{(K^3 + \mathbb{E}[C(t) | A_0^t]^3)^2} \right) (C(t) - \mathbb{E}[C(t) | A_0^t]). \tag{C27}$$

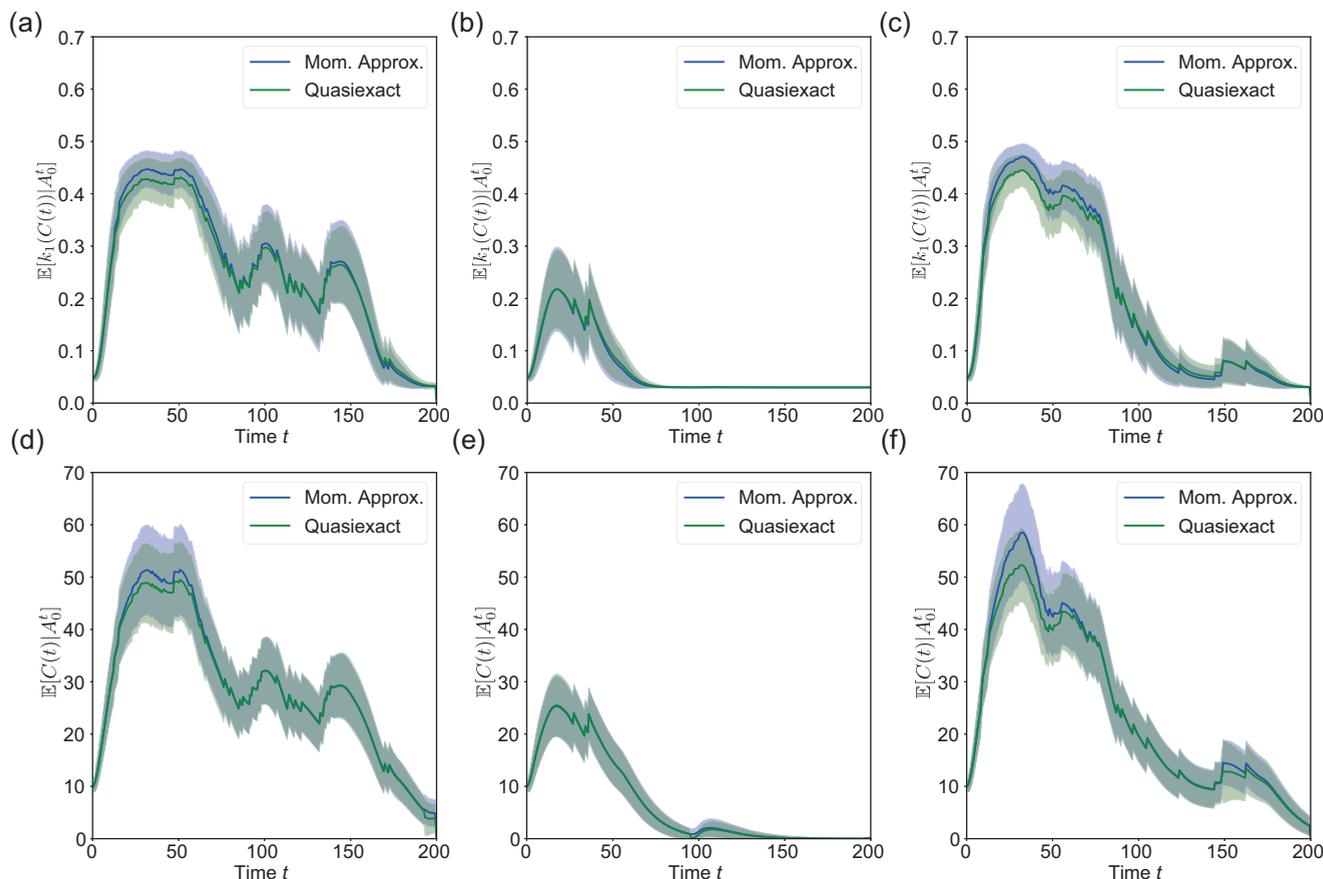


FIG. 7. (a)–(f) Comparison of the quasiexact method with the moment-approximation method for individual trajectories. The simulation was performed for parameters  $\{k_2, k_3, k_4, k_5, k_6\} = \{0.1, 1, 0.1, 0.1, 0.1\}$  and  $\{K, n_H, \epsilon\} = \{30, 3, 0.03\}$ . The grid size for the quasiexact method was chosen to be  $70 \times 70$ . The shaded areas indicate conditional standard deviations obtained by the respective methods.

Moments can then be derived from (C25) as described before. This approximation is expected to be accurate when the filtering distribution shows little variation in  $C(t)$  around its mean  $\mathbb{E}[C(t) | A_0^t]$ , which should be the case when reaction  $R_A = \{1\}$  fires frequently [i.e., frequent observation of this reaction reveals more information about  $C(t)$ ]. We remark that this approximation needs to be performed only for (C25) since the other two equations are conditioned on  $C_0^t$ , such that no averaging over  $C(t)$  has to be performed (which would otherwise give rise to expectations over rational functions). In order to simplify the identification of parameter regimes where this system exhibits two metastable states, we analyzed the corresponding macroscopic mean-field equations given by

$$\begin{aligned} \frac{d}{dt} a(t) &= \frac{\mu c(t)^3}{K^3 + c(t)^3} + \epsilon - k_2 a(t), \\ \frac{d}{dt} b(t) &= k_3 a(t) - k_4 b(t), \\ \frac{d}{dt} c(t) &= k_5 b(t) - k_6 c(t). \end{aligned} \tag{C28}$$

Figure 6 shows the equilibrium points of this system as a function of the feedback strength  $\mu$  for the parameters  $\epsilon = 0.03, K = 30, k_2 = 0.1, k_3 = 1, k_4 = 0.1, k_5 = 0.1,$  and

$k_6 = 0.1$ . This shows that there are two (stable) equilibrium points once  $\mu$  exceeds a threshold ( $\mu \approx 0.35$ ), and these points subsequently approach each other when  $\mu$  increases further.

Note that the filtering equation (C25) involves a strongly nonlinear state dependency via rate  $k_1(c)$ , which we linearized to obtain closed moment dynamics. It is thus not obvious whether the resulting conditional moments are approximated accurately by our moment-based approach. To test whether this is the case, we used the quasiexact method similarly to the previous case study. However, due to the computational complexity of the quasiexact method, we found a comprehensive numerical analysis of the stationary mutual information rate to be exceedingly expensive. Instead, we compared the moment-approximation method with the quasiexact method for a few individual trajectories for  $\mu = 0.5$ , where individual trajectories switch between the two metastable states. We performed this comparison for  $\mathbb{E}[C(t) | A_0^t]$  as well as  $\mathbb{E}[k_1(C(t)) | A_0^t]$ , whereas the latter is used for calculating the mutual information. For the quasiexact method, we solved (C25) on a finite-dimensional grid of size  $70 \times 70$ . The results show that the conditional moments are approximated well by the moment-approximation method (Fig. 7). Some degree of mismatch can be observed when  $\mathbb{E}[C(t) | A_0^t]$  assumes larger values [Figs. 7(a), 7(c), 7(d), and 7(f)]. This is likely an

inaccuracy of the quasixact method, which forces the probability mass to remain within the grid boundaries. Consequently,  $\mathbb{E}[C(t) | A'_0]$  and  $\mathbb{E}[k_1(C(t)) | A'_0]$  are estimated less accurately when  $\mathbb{E}[C(t) | A'_0]$  gets close to the grid bound-

aries. Note that this analysis was performed only for (C25), because the other two filtering equations are identical to the previous case study with the only difference that (C26) now exhibits a time-dependent rate  $k_1$ .

- 
- [1] C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [2] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, Positional information, in bits, *Proc. Natl. Acad. Sci. USA* **110**, 16301 (2013).
- [3] J. Selimkhanov, B. Taylor, J. Yao, A. Pilko, J. Albeck, A. Hoffmann, L. Tsimring, and R. Wollman, Accurate information transmission through dynamic biochemical signaling networks, *Science* **346**, 1370 (2014).
- [4] T. M. Cover, *Elements of Information Theory* (Wiley, New York, 1999).
- [5] J. E. Purvis, K. W. Karhohs, C. Mock, E. Batchelor, A. Loewer, and G. Lahav, p53 dynamics control cell fate, *Science* **336**, 1440 (2012).
- [6] A. S. Hansen and E. K. O'Shea, Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression, *Mol. Syst. Biol.* **9**, 704 (2013).
- [7] R. M. Fano, Transmission of information: A statistical theory of communications, *Am. J. Phys.* **29**, 793 (1961).
- [8] F. Tostevin and P. R. ten Wolde, Mutual Information between Input and Output Trajectories of Biochemical Networks, *Phys. Rev. Lett.* **102**, 218101 (2009).
- [9] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 1992), Vol. 1.
- [10] M. Sinzger, M. Gehri, and H. Koepl, Poisson channel with binary Markov input and average sojourn time constraint, in *Proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, NJ, 2020), pp. 2873–2878.
- [11] L. Duso and C. Zechner, Path mutual information for a class of biochemical reaction networks, in *Proceedings of the 2019 IEEE 58th Conference on Decision and Control (CDC)* (IEEE, Piscataway, NJ, 2019), pp. 6610–6615.
- [12] M. Sinzger-D'Angelo and H. Koepl, ACID: A low dimensional characterization of Markov-modulated and self-exciting counting processes, [arXiv:2205.07011](https://arxiv.org/abs/2205.07011).
- [13] M. Reinhardt, G. Tkačik, and P. R. ten Wolde, Path weight sampling: Exact Monte Carlo computation of the mutual information between stochastic trajectories, [arXiv:2203.03461](https://arxiv.org/abs/2203.03461).
- [14] D. F. Anderson and T. G. Kurtz, Continuous time Markov chain models for chemical reaction networks, in *Design and Analysis of Biomolecular Circuits* (Springer, New York, 2011), pp. 3–42.
- [15] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes II: Applications*, Stochastic Modelling and Applied Probability Vol. 6 (Springer, Berlin, Heidelberg, 2001).
- [16] P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding, *Statistical Models Based on Counting Processes*, Springer Series in Statistics (Springer, New York, 1993).
- [17] L. Duso and C. Zechner, Selected-node stochastic simulation algorithm, *J. Chem. Phys.* **148**, 164108 (2018).
- [18] C. Zechner and H. Koepl, Uncoupled analysis of stochastic reaction networks in fluctuating environments, *PLoS Comput. Biol.* **10**, e1003942 (2014).
- [19] D. Guo, S. Shamai, and S. Verdu, Mutual information and minimum mean-square error in Gaussian channels, in *Proceedings of the 2005 IEEE Transactions on Information Theory* (IEEE, Piscataway, NJ, 2005), pp. 1261–1282.
- [20] R. Atar and T. Weissman, Mutual information, relative entropy, and estimation in the Poisson channel, in *Proceedings of the 2011 IEEE International Symposium on Information Theory* (IEEE, Piscataway, NJ, 2011), pp. 708–712.
- [21] R. E. Spinney, M. Prokopenko, and J. T. Lizier, Transfer entropy in continuous time, with applications to jump and neural spiking processes, *Phys. Rev. E* **95**, 032319 (2017).
- [22] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering, Stochastic Modelling and Applied Probability* Vol. 60 (Springer, New York, 2009).
- [23] B. Munsky and M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, *J. Chem. Phys.* **124**, 044104 (2006).
- [24] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J. Comput. Phys.* **22**, 403 (1976).
- [25] D. Schnoerr, G. Sanguinetti, and R. Grima, Approximation and inference methods for stochastic biochemical kinetics—A tutorial review, *J. Phys. A: Math. Theor.* **50**, 093001 (2017).
- [26] E. Lakatos, A. Ale, P. D. Kirk, and M. P. Stumpf, Multivariate moment closure techniques for stochastic kinetic models, *J. Chem. Phys.* **143**, 094107 (2015).
- [27] M. Meijers, S. Ito, and P. R. ten Wolde, Behavior of information flow near criticality, *Phys. Rev. E* **103**, L010102 (2021).
- [28] P. P. Mitra and J. B. Stark, Nonlinear limits to the information capacity of optical fibre communications, *Nature (London)* **411**, 1027 (2001).
- [29] I. Lestas, G. Vinnicombe, and J. Paulsson, Fundamental limits on the suppression of molecular fluctuations, *Nature (London)* **467**, 174 (2010).
- [30] J. M. Horowitz and H. Sandberg, Second-law-like inequalities with information and their interpretations, *New J. Phys.* **16**, 125007 (2014).
- [31] <https://github.com/zechnerlab/PathMutualInformation/>.
- [32] L. Bronstein and H. Koepl, Marginal process framework: A model reduction tool for Markov jump processes, *Phys. Rev. E* **97**, 062147 (2018).
- [33] O. Aalen, O. Borgan, and H. Gjessing, *Survival and Event History Analysis: A Process Point of View* (Springer, New York, 2008).
- [34] B. Øksendal and A. Sulem, *Applied Stochastic Control of Jump Diffusions*, Universitext (Springer, Berlin, Heidelberg, 2007).