# Geometry of nonequilibrium reaction networks

Sara Dal Cengio, Vivien Lecomte, Matteo Polettini

# Geometry of nonequilibrium reaction networks

Sara Dal Cengio* and Vivien Lecomte
*Université Grenoble Alpes, CNRS, LIPhy, FR-38000 Grenoble, France*

Matteo Polettini
*Complex Systems and Statistical Mechanics, Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg, Luxembourg*

(Dated: August 2, 2022)

Building on Kirchhoff's treatment of electrical circuits, Hill and Schnakenberg – among others – proposed a celebrated theory for the thermodynamics of Markov processes and linear biochemical networks that exploited tools from graph theory to build fundamental nonequilibrium observables. However, such simple geometrical interpretation does not carry through for arbitrary chemical reaction networks because reactions can be many-to-many and are thus represented by a hypergraph, rather than a graph. Here we generalize some of the geometric intuitions behind the Hill–Schnakenberg approach to arbitrary reaction networks. In particular, we give simple procedures to build bases of cycles (encoding stationary nonequilibrium behavior) and cocycles (encoding relaxation), to interpret them in terms of circulations and gradients, and to use them to properly project nonequilibrium observables onto the relevant subspaces. We develop the theory for chemical reaction networks endowed with mass-action kinetics and enrich the description with insights from the corresponding stochastic models. Finally, basing on the linear regime assumption, we deploy the formalism to propose a reconstruction algorithm for metabolic networks consistent with Kirchhoff's Voltage and Current Laws.

## CONTENTS



* sara.dal-cengio@univ-grenoble-alpes.fr

## I. INTRODUCTION

### A. Context and motivations

Our understanding of nonequilibrium physics is built on identifying the nontrivial forces which drive a system out of equilibrium and the currents governing how energy is consumed and dissipated – and this both inside the system and in its interaction with the outside environment.

Perhaps the simplest example is that of a driven Brownian particle moving on a ring. The system is described by the Langevin equation $\dot{x}(t) = F(x(t)) + \eta(t)$ for the particle's position $x(t)$, where $F(x)$ is the deterministic force and $\eta(t)$ is the thermal noise (white and Gaussian). The one-dimensional nature of the problem makes it particularly easy to decompose the force field into conservative and non-conservative contributions: $F(x) = -V'(x) + f$, where $V(x) = V(x + L)$ is a periodic potential and $f$ is defined by $f = \int_0^L \mathrm{d}x\, F(x)$ with $L$ the ring length. The source of the drive is then identified in the scalar parameter $f$, which characterizes irreversibility and drives a net current on the ring (when $f \neq 0$) [1]. For $f = 0$ the system relaxes, with a vanishing current, to an equilibrium steady state governed by the Boltzmann probability distribution $P(x) \propto \exp[-V(x)/T]$ (with the temperature $T$ identified from the noise amplitude). Already in this simple example, decomposing the force shows how topology plays a role: in the absence of periodic boundary conditions the system can not sustain a nonequilibrium steady state. It is the geometry of the ring which supports a non-vanishing constant current. As a matter of fact, the centrality of cycles to nonequilibrium was manifest already in the early days of thermodynamics: engines perform in cycles [4] and cycles are crucial in the operation of molecular motors [5].

The model of the particle on a ring has arguably little relevance in real-world settings. However, there are alternative scenarios where cycles are encountered naturally in link with nonequilibrium. It is the case of chemical reaction networks (CRNs) [6–10], in particular biochemical networks, such as those involved in gene regulation, cellular sensing and metabolic functions of living systems. A chemical network consists of a set of reactions involving chemical species. Each reaction has its inherent chemical activity: it transforms (a combination of) reactants into products giving rise to a net flux of matter, named current, in response to an intrinsic chemical force, named affinity. At equilibrium both the currents and the affinities vanish, which is the domain of standard thermodynamics. Thermodynamic feasibility then specifies that a closed system, i.e. a system which does not exchange matter (or energy) with the outside environment, always reaches an equilibrium state determined by the initial conditions. Thereupon, external currents can be injected in the system through external chemostats which then foster nonequilibrium behavior. Let us consider for example a minimal model of glycolysis [11] for the consumption of ATP in the cell:

$$\begin{aligned} \emptyset &\overset{1}{\rightleftharpoons} \mathrm{Gly} \\ \mathrm{Pyr} &\overset{2}{\rightleftharpoons} \emptyset \\ \mathrm{Gly} + 2\mathrm{ADP} &\overset{3}{\rightleftharpoons} 2\mathrm{ATP} + 2\mathrm{Pyr} \\ \mathrm{ATP} &\overset{4}{\rightleftharpoons} \mathrm{ADP} \end{aligned} \tag{1}$$

Here the first two reactions stand for the couplings with external chemostats, depicted with the symbol $\emptyset$. In the model, the cell imports and expels glucose (Gly) and pyruvate (Pyr) through the membrane in order to maintain their concentration levels. Within the cell (reaction 3), a molecule of Gly is used to convert two molecules of low-energy adenosine diphosphate (ADP) into two molecules of high-energy adenosine triphosphate (ATP). The chemical energy stored in ATP is then released during the spontaneous dephosphorylation of ATP (reaction 4), and used to fuel the physiological activity of the cell. One sees that whenever the four reactions are performed in proportion $(1, 2, 1, 2)$, the number of molecules of each species is preserved. This is what in the literature is called a chemical cycle [10, 12], that is, a sequence of reactions which does not alter the overall state of the system. The cycle's preferential direction is the consequence of chemostats – which sustain a net consumption of glucose, supplied through nutrition, and a net production of pyruvate, expelled in the form of lactose. As a consequence, the cell is maintained in a nonequilibrium state, consuming energy for its metabolic activity.

Going back to where we started, we may ask what are the conservative and the non-conservative forces in this case. The presence of a chemical cycle here plays a role similar to the periodic boundary conditions for the Brownian particle on a ring. However, the analogy is hindered by the fact that interactions in this case are inherently discrete and dependent on the overall topology of the network. Thus, the challenge is primarily conceptual: what does it mean to decompose chemical forces into conservative and non-conservative contributions for a reaction network?

Based on works by Kirchhoff on electrical circuits, and Kolmogorov on Markov chains, Hill and Schnakenberg [6, 8, 13] – among others – proposed a generic framework to describe the source of irreversibility in the steady state as stemming from chemical cycles. Analyzing the graph at a population level (each node representing a given set of numbers of molecules for all species), he put forward a method to identify macroscopic currents and affinities. First, using the graph-theory notion of spanning tree, he identified a fundamental sets of cycles defined on the population graph. Then, he

[1] The current-force relation is found by solving the corresponding 1D stationary Fokker–Planck equation [1–3]: for $f > 0$ the steady current is $j^* = L / \int_0^\infty \mathrm{d}y \int_0^L \mathrm{d}x \exp[-fy + V(y + x) - V(x)]$. We learn (*i*) that for $f \to 0$ the current goes to 0 (and therefore it makes sense to call this limit equilibrium); (*ii*) That the current-force relation is complicated and highly nonlinear; (*iii*) That the quantity $j^* f$ is nonnegative (second law).

proved that chemical forces acquire a direct physical interpretation when defined from those cycles.

In Ref. [14], one of us showed that, from the same notion of spanning tree, the picture can be extended by considering graph cocycles, a notion complementary to that of cycles and related to it via graph-theoretical duality. Interestingly, the notion of cocycle was the missing piece in Schnakenberg's analysis to understand the structure of chemical forces beyond steady state (i.e. so as to fully encompass the finite-time dynamics). Using graph cycles and cocycles, affinities and currents can be respectively decomposed into (*i*) conservative versus non-conservative affinities and (*ii*) transient versus steady-state currents. The notion of cycle, familiar to the statistical physicist, has been exploited in a number application of modern thermodynamics [15–19].

Albeit elegant and powerful, the analyses of Hill–Schnakenberg and the one in Ref. [14] in practice apply only to noninteracting networks which can be represented as simple graphs [20], for which a spanning tree can be constructed. This is the case for instance of resistor (or flow) networks [21], or unimolecular chemical networks [22], where each reaction involves only two species [e.g. reaction 4 in Eq. (1)].

However, in real-world networks, such as the autocatalytic networks involved in life functions [23], reactions typically involve more than two chemical components [e.g. reaction 3 in Eq. (1)]. Such interactions give rise to nonlinearities at the level of chemical concentrations, resulting in a spectrum of dynamical behaviors not displayed by noninteracting networks. Furthermore, the non-pairwise nature of these reactions make CRNs best represented as hypergraphs [24], that is, generalized graphs where hyper-edges connect more than two nodes (the uninitiated reader may refer to Figs. 6 and 7 for examples). Hypergraphs have recently emerged as a new challenge in network science [25–27], and, contrary to simple graphs, they currently lack a comprehensive theoretical understanding. In particular, no notion of spanning tree exists for hypergraphs, precluding one to formulate a 'geometrical' analysis based on the graph cycles and cocycles we just described. For this reason, the possibility of generalizing such analysis to interacting networks has remained an open issue.

### B. Main results and structure of the paper

In the present work, we generalize the geometrical analysis of Hill–Schnakenberg to the case of interacting reaction networks, allowing one to identify their driving forces and macroscopic currents – beyond the case of noninteracting systems. Crucially, we propose a new notion of cycle and cocycle, which does not rely on graph theory (and on having a spanning tree), but follows from an alternative algebraic approach. This requires a conceptual shift: from graph-theoretical objects to vectorial spaces and to (a special choice of) their bases. We demonstrate how the newly defined cycles and cocycles possess many of the features of the geometrical counterparts. In particular, we show that, for linear networks, they reduce to the cycles and cocycles defined from graph theory, thus justifying keeping the same terminology.

A key step in the theory consists in analyzing the dependent and independent degrees of freedom that the stoichiometric matrix $\mathbb{S}$, which prescribes the topology of the network, possesses. This allows one to identify a matrix $\mathbb{T}$, central to our theory, which connects to $\mathbb{S}$ via the following matrix relation:

$$-\mathbb{S}^\top \mathbb{G}^\top = \begin{pmatrix} \mathbb{0} & \mathbb{1}_M \\ \mathbb{0} & \mathbb{T}^\top \end{pmatrix} . \tag{2}$$

Linear algebra guarantees the existence of an invertible matrix $\mathbb{G}$ satisfying this relation, together with the uniqueness of the matrix $\mathbb{T}$. The different actors of Eq. (2) all carry a physical interpretation, which is discussed throughout the paper. In particular, the integer $M$ represents both the number of independent reactions and independent species, the matrix $\mathbb{T}$ contains the exact amount of information describing both cycles and cocycles, while the matrix $\mathbb{G}$ allows one to connect the chemical affinities to the thermodynamic potentials.

Notice that, for noninteracting networks, the matrix $\mathbb{S}$ coincides with the incidence matrix relating the nodes and edges of an oriented graph (see Fig. 1). Then, Eq. (2) possesses a geometrical interpretation, with $\mathbb{S}^\top$ being the discrete gradient on the oriented graph and the matrix $\mathbb{G}^\top$ being the corresponding discrete integrator along a spanning tree (such description will be detailed in the following). We extend this picture to the case of interacting networks, by defining an integration operator on the hypergraph through a geometrical notion of escape routes. In this way, we are able to connect the notion of reversibility, for the nonlinear dynamics, to a potential (or integrability) condition for the chemical forces.

The notion of cycles and cocycles, now constructed via Eq. (2), allow us to decompose affinities and current into thermodynamically relevant contributions. Namely, for the chemical affinities, we put forward a decomposition into conservative and non-conservative forces which resembles the Helmholtz–Hodge decomposition of vector calculus in $\mathbf{R}^3$. For the unfamiliar reader, we recall that Helmholtz–Hodge decomposition of a vector field, $\mathbf{F} = -\nabla V + \nabla \times \mathbf{A}$, provides, in three dimensions, the corresponding separation of the force into two components: (*i*) a gradient force that is conservative, (*ii*) a non-gradient force with zero divergence (i.e. of zero total flux through any closed surface) and drives irreversible behavior. We discuss the analogy with our decomposition of chemical affinities and its implication for nonequilibrium physics [28, 29]. A complementary decomposition for the current is also introduced, in terms of (*i*) cycle currents which survive in the steady state and (*ii*) cocycle currents, which are transient and correspond to 'tidal' relaxation.

Throughout the paper, we put particular emphasis on the notion of cocycle, which has been somehow overlooked in the literature compared to that of cycle. Albeit less intuitive, the notion of cocycle is fundamental to depict how relaxation occurs in the system. In fact, it enables us to explore some direct consequences of the formalism: the linear response of interacting CRNs, which turns out to be controlled by the matrix $\mathbb{T}$; and the slow modes of non-linear relaxation, which are controlled by the cocycles when a timescale separation occurs.

Finally, as a practical application of the formalism, we consider the problem of thermodynamically consistent re-

construction of the CRNs involved in various biochemical processes, e.g. metabolism [30–37] (but other multi-omics datasets could also be considered). The problem is, roughly, the following: DNA sequencing grants knowledge of the enzymes possibly present in a cell, and enzyme specificity identifies the substrates (metabolites) that bind and interact. Thus the stoichiometry of the metabolite network is known. However, the rates or even the directions of the reactions are not, and one would like to have some principles to make informed guesses about the overall metabolism, or to be able to move in the metabolic landscape. Such landscape is constrained by 1) conservation of chemical elements, and 2) thermodynamic feasibility. While the first is a simple linear constraint, the second is non-linear, and it has proven difficult to implement it in reconstruction algorithms [38–40].

While metabolic reconstruction is a much broader enterprise, dealing with further constraints [41] such as concentrations [42], or biomass or other thermodynamic target functions [43], abundance of conservation laws [44] or of enzymes, we trivialize it as the identification and navigation of a landscape of possible currents with some known values of the rates of some externally injected metabolites. Then our approach grants a simple linear-regime approximation of such landscape that easily allows one to explore it. The best feature of this reconstruction method is that the only free parameters are some positive real numbers, one per internal reaction. Once these parameters are given and some intuition about which are the independent external currents is built, the reconstruction is just a simple linear formula that allows one to explore a bulk landscape of solutions.

The paper is organized as follows. For the sake of clarity, we dedicate Part II to noninteracting networks, whose configuration space is a graph, and review the analysis of Hill–Schnakenberg, extending it to include finite-time relaxation (in the spirit of Ref. [14]). In Secs. II A-II B we discuss the complete mapping between the graph-theoretical analysis and the algebraic framework. Then, In Secs. II D-II E we connect to standard thermodynamics both for closed and open (chemostatted) systems. In particular, we show how the chemostatting procedure results in non-conservative chemical affinities, thus driving the system out of equilibrium. Part II has the primary goal to make the paper self-contained, providing as much physical intuition as possible and making it accessible to a broad spectrum of readers. Any reader already familiar with graph theory, chemical kinetics, network thermodynamics and all that, might directly go to Part III, the real core of the paper. Therein, we show how to fully extend the theory to interacting CRNs. In Sec. III D we explain how the stochastic level of description is related to thermodynamics. In Sec. III E we study the linear response of interacting CRNs and unveil a spectral symmetry between the equilibrium-relaxation and the drive-steady state responses. Finally, in Sec. III F we apply the formalism to the problem of metabolic reconstruction and propose an algorithmic-like procedure to find feasible values for the internal currents in metabolic networks. Sec. IV contains our final remarks.

### C. Setup and notations

In this paper we work with CRNs with mass action kinetics. More precisely, we consider a dilute, well-stirred mixture of $N$ chemical species interacting through $R$ reactions. We group the chemical species in the vector $\boldsymbol{X} = (X_1, ..., X_N)^\top$ and for each reaction $\rho$ we write the corresponding stoichiometric equation as

$$\boldsymbol{\nu}^{+\rho} \cdot \boldsymbol{X} \underset{k_\rho^-}{\overset{k_\rho^+}{\rightleftharpoons}} \boldsymbol{\nu}^{-\rho} \cdot \boldsymbol{X}\,, \tag{3}$$

where the vectors $\boldsymbol{\nu}^{+\rho}$ and $\boldsymbol{\nu}^{-\rho}$ contain the numbers of particles per species being consumed and produced by reaction $\rho$, and $\cdot$ is the scalar product. Each reaction is strictly reversible, that is, can occur both in the forward and backward direction with reaction rate constants $k_\rho^\pm > 0$. Thus, for each reaction we introduce a pair of velocities $\lambda_\rho^\pm(\boldsymbol{x})$ describing the rate of change of the chemical concentrations $\boldsymbol{x} = (x_1, ..., x_N)^\top$ in the corresponding direction. For large number of particles (i.e. negligible fluctuations), the velocities are proportional to the concentrations of the species partaking to the reaction $\pm\rho$,

$$\lambda_\rho^\pm(\boldsymbol{x}) = k_\rho^\pm \boldsymbol{x}^{\boldsymbol{\nu}^{\pm\rho}} \quad \forall\rho\,, \tag{4}$$

with the notation convention $\boldsymbol{a}^{\boldsymbol{b}} = \prod_i a_i^{b_i}$. Then, we define the net current $J_\rho$ of reaction $\rho$ as the difference between the two reaction velocities

$$J_\rho(\boldsymbol{x}) = \lambda_\rho^+(\boldsymbol{x}) - \lambda_\rho^-(\boldsymbol{x})\,. \tag{5}$$

The dynamical evolution of the concentration vector $\boldsymbol{x}$ is given by a deterministic rate equation:

$$\frac{d}{dt}\boldsymbol{x}(t) = \mathbb{S}\boldsymbol{J}(\boldsymbol{x}(t))\,, \tag{6}$$

where $\mathbb{S}$ is the stoichiometric matrix, of dimensions $N \times R$, whose columns describe the net stoichiometry of each reaction $\mathbb{S}_\rho = \boldsymbol{\nu}^{-\rho} - \boldsymbol{\nu}^{+\rho}$. As such, $\mathbb{S}$ encodes the topology of the network, and acts as a discrete divergence in Eq. (6), which can be seen as a continuity equation. Finally, for each reaction $\rho$ we introduce the chemical affinity $A_\rho$ defined as:

$$A_\rho(\boldsymbol{x}) = \log\frac{\lambda_\rho^+(\boldsymbol{x})}{\lambda_\rho^-(\boldsymbol{x})} = \log\left(\frac{k_\rho^+}{k_\rho^-}\boldsymbol{x}^{-\mathbb{S}_\rho}\right). \tag{7}$$

The following constitutive equation between current $J_\rho$ and conjugated affinity $A_\rho$ holds:

$$J_\rho(\boldsymbol{x}) = \lambda_\rho^+(\boldsymbol{x})\left[1 - \exp\left(-A_\rho(\boldsymbol{x})\right)\right], \tag{8}$$

which quantifies the chemical drive, i.e. how an imbalance in the concentrations of reactants and products results in a net reaction current.

We highlight that, although the deterministic rate equation (6) applies only in the infinite system-size limit, there exists, underlying, a microscopic (molecular) and stochastic

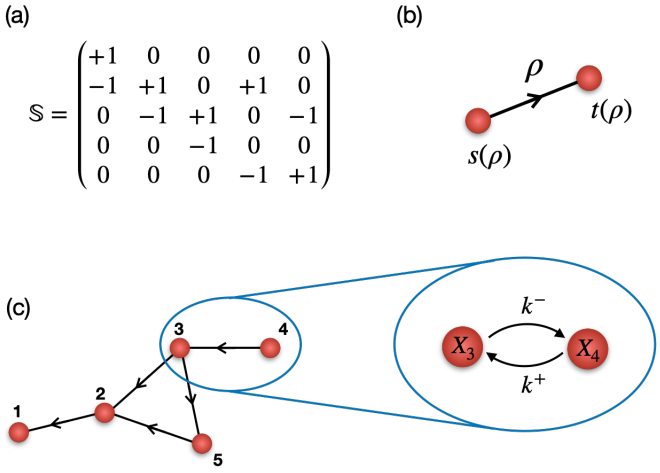



FIG. 1. Example of graphical representation for a noninteracting reaction network of the type of Eq. (9). **(a)** The stoichiometric matrix of a noninteracting network is an incidence matrix describing the relation between the species/nodes and the reactions/edges. Each column $\mathbb{S}_\rho$ has exactly a $+1$ entry corresponding to the species produced by the reaction $\rho$ and a $-1$ entry corresponding to the species consumed by the reaction $\rho$. **(b)** Accordingly, one can represent each column of $\mathbb{S}$ as an edge between a source node $s(\rho)$ and a target node $t(\rho)$. **(c)** By doing so for all the reactions in the network, one obtains a planar graphical representation where each node is a species and each edge is a reaction. Since the reactions are reversible, the orientation of the edges is conventional.

level of description where individual molecules transform one into each other, with transition rates $\kappa_\rho^\pm$ that are directly related to the reaction rate constants $k_\rho^\pm$ in Eq. (3). Namely, for every reaction $\rho$, reactants transform into products with a transition rate $\kappa_\rho^+$ (resp. $\kappa_\rho^-$) in the forward (resp. backward) direction, with $\kappa_\rho^\pm = \Omega^{1-\sum_i \nu_i^{\pm\rho}} k_\rho^\pm$ (where $\Omega$ is the system volume – see Appendix B for a complete treatment). As it will appear, several of our results also apply to this stochastic level of description.

## II. NONINTERACTING REACTION NETWORKS

We dedicate this section to noninteracting networks where each reaction involves a single molecule converting from one species to another:

$$X_i \underset{k_\rho^-}{\overset{k_\rho^+}{\rightleftharpoons}} X_j \,. \tag{9}$$

In this case the stoichiometric matrix $\mathbb{S}$ takes the form of an incidence matrix, namely:

$$\mathbb{S}_{i\rho} = \begin{cases} -1 & \text{if } i \text{ is the species consumed by } \rho \\ +1 & \text{if } i \text{ is the species produced by } \rho \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

We first consider isolated systems, and the description of chemostatted open systems is done in Sec. II E. The objective of this first part is to relate the physical, algebraic and graph-theoretical pictures underlying such set of reactions, in view of extending these to the case of interacting CRNs (as done in Sec. III). We adopt in this section the language of chemical reactions, but Eq. (9) also describes a Markov chain between states labeled by the $X_i$'s (this is one of Schnakenberg's standpoints). The implication of our results in the language of Markov chains is detailed in Sec. IV.

### A. A handful of graph theory

Noninteracting CRNs like Eq. (9) admit a graphical representation in terms of nodes (or vertices) and edges (or links): the incidence matrix Eq. (10) describes the topology of an oriented graph $\mathcal{G}$ where each reaction $\rho$ is a two-ended edge pointing from a source node $s(\rho)$ to a target node $t(\rho)$, and each node represents a chemical species (see Fig. 1 for an example). Without loss of generality, we consider simply connected graphs. We now present the building blocks used in the present section, referring the reader to Fig. 2 for illustrations. Following the Hill–Schnakenberg approach [6, 8, 13], we introduce the concept of spanning tree, defined as a connected subgraph of $\mathcal{G}$, containing every node but no closed paths (see Fig. 2b). Clearly in general there are several spanning trees and their number depends on the topology of $\mathcal{G}$. We fix one that we call $T_\mathcal{G}$. Choosing $T_\mathcal{G}$ corresponds to splitting the edges of $\mathcal{G}$ into edges that are excluded from the spanning tree and edges that belong to it. In graph theory [20], these distinct edges are respectively named chords and cochords [2] and we associate them with two indices, $\alpha \notin T_\mathcal{G}$ spanning the set of chords and $\gamma \in T_\mathcal{G}$ spanning the set of cochords.

Adding a chord $\alpha$ back in $T_\mathcal{G}$ generates a closed path. Removing a cochord $\gamma$ from $T_\mathcal{G}$ generates a cut, i.e. a splitting of the nodes of $\mathcal{G}$ into two disconnected islands/components (see Fig. 2). Such closed paths and cuts can be given an orientation: that of the closed path prescribes the direction of going along it; and that of the cut is a choice of source and target among the two disconnected islands. We thus define the cycle $C(\alpha)$ as the closed path generated by restoring the chord $\alpha$ into $T_\mathcal{G}$, and oriented in the same direction as the generating chord $\alpha$ (Fig. 2b). We define the cocycle $C(\gamma)$ as the set of edges that reconnect $\mathcal{G}$ after removing $\gamma$ from $T_\mathcal{G}$. Conventionally, the source island is chosen to be the island containing the source node $s(\gamma)$ (Fig. 2c). Notice that, if the cochord $\gamma$ is a bridge, i.e. if it does not belong to any closed path, the corresponding cocycle contains only the cochord $\gamma$. This occurs when the reconnection of $\mathcal{G}$ is obtained by reintroducing $\gamma$ only. On the other hand, if the cochord $\gamma$ belongs to one (or more) closed path in $\mathcal{G}$, the corresponding cocycle contains all the chords associated with those closed paths. Namely, for any pair of chord and cochord $(\alpha, \gamma)$ we have that $\gamma \in C(\alpha) \iff \alpha \in C(\gamma)$, with the pair of edges $(\alpha, \gamma)$ oriented parallel to each other in $C(\gamma)$ and antiparallel (head-tail orientation) in $C(\alpha)$. This encodes the fact that whenever

[2] Notice that the mapping of $\mathcal{G}$ to the dual graph $\mathcal{G}*$ in graph theory corresponds to the mapping of chords to cochords and vice-versa.

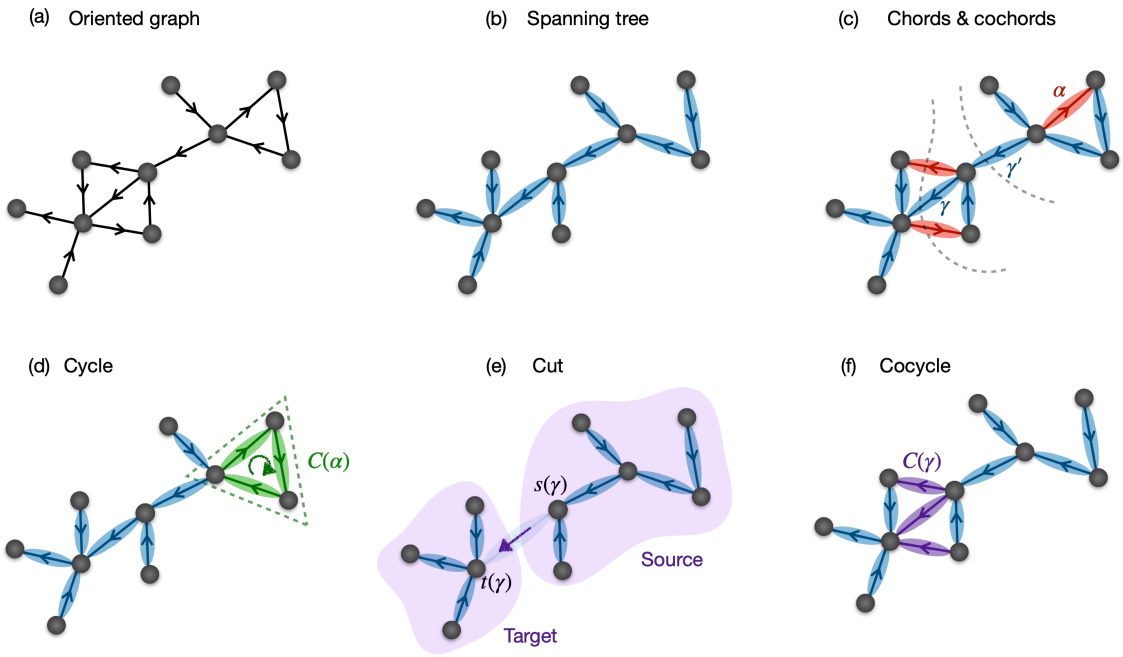



FIG. 2. A summary of the notions from graph theory which are discussed in the text. **(a)** Example of an oriented graph $\mathcal{G}$ obtained from the stoichiometric/incidence matrix $\mathbb{S}$ of a noninteracting CRN: each node represents a species and each edge a reaction. **(b)** A spanning tree $T_{\mathcal{G}}$, here represented with blue-marked edges, is obtained by pruning edges from the original graph so as to remove every closed path (while keeping a connected tree). **(c)** Picking $T_{\mathcal{G}}$ corresponds to (a choice of) splitting of the edges in $\mathcal{G}$ between the cochords (in blue), forming $T_{\mathcal{G}}$, and the chords (in red). The set of chords is spanned by the index $\alpha$ and the set of cochords by the index $\gamma$. Cochords like $\gamma'$ are named bridge as they do not belong to any closed path, in contrast for instance to the cochord $\gamma$ which belongs to two closed paths. **(d)** Reintroducing a chord $\alpha$ into the spanning tree generates a cycle $C(\alpha)$ which is oriented according to the orientation of the chord $\alpha$. By construction, each cycle only contains a single chord (the generating one). **(e)** The removal of a cochord $\gamma$ from the spanning tree generates a cut, i.e. a splitting of the full graph in two disconnected islands/components. In particular, we name 'source island' (resp. 'target island') the component containing the source node $s(\gamma)$ (resp. target node $t(\gamma)$) of the generating cochord. This allows one to establish an orientation of the cut. **(f)** The cocycle $C(\gamma)$ is defined as the set of edges that reconnect the source island to the target island after removing the cochord $\gamma$. In the example it has three elements: the cochord $\gamma$ and the two chords associated to cycles that contain $\gamma$. All the three edges are oriented parallel to the cochord $\gamma$, in order to connect the source to the target.

$\gamma \in C(\alpha)$, the cycle $C(\alpha)$ possesses nodes both in the source and in the target of the cut corresponding to the cocycle $C(\gamma)$, and the edges of this cycle must present a zero flux in total between source and target. We stress that this 'duality' property between $C(\alpha)$ and $C(\gamma)$ is non-trivial, and its manifestation will be met in multiple forms in the following [3].

Cocycles and cycles are the central graph-theoretical ingredients of this paper. In the following, they will serve as a basis to decompose the chemical forces governing the reactions of Eq. (9) into conservative and non-conservative forces, in analogy with the Helmholtz–Hodge decomposition in continuous space. Before detailing how cycles and cocycles are algebraically related to $\mathbb{S}$, we explain how they are physically meaningful. A cycle is a closed path in the space of reactions. As such, it corresponds to a sequence of transformations which brings a molecule back to its original state/node. One already sees the analogy with the chemical cycles defined in the introduction. In the spirit of Hill–Schnakenberg's theory, we will use them to characterize nonequilibrium steady states. The interpretation of cocycles is less intuitive but will play a central role. Those are sets of reactions associated to a binary splitting of the system into two separate sub-systems; as will become clear in the following sections, they are associated (*i*) to fluxes of matter with no circulation, relating to the modes of relaxation of the dynamics, and (*ii*) to conservative chemical forces, that 'derive' from a potential.

[3] A proof of this property can be found in §2-2 of Nakanishi's book [20]. Anticipating on algebraic notions, it encodes that a same matrix $\mathbb{T}$ describes both the dependencies between reactions and the cycles of CRNs; in this framework, a purely algebraic proof of this can be found at the end of our Appendix A 1.

### B. From graph theory to algebra

We now explain how to build an alternative characterization of cycles and cocycles by means of linear algebra. Namely, one associates to cycles and cocycles two families of vectors which not only retain the same properties as on the graph but also, algebraically, occur to be bases of two complementary (real valued) vector spaces,

$$\mathrm{Ker}\,\mathbb{S} \perp \mathrm{Im}\,\mathbb{S}^{\top}, \tag{11}$$

namely the kernel $\mathrm{Ker}\,\mathbb{S}$ and the coimage $\mathrm{Im}\,\mathbb{S}^\top$ of the stoichiometric matrix.

To start with, we count the number of independent degrees of freedom. We know from the graph construction that the total number of cycles and cocycles is equal to the number of reactions/edges $R$ and, likewise, to the number of columns of $\mathbb{S}$. The latter is in turn related to the dimensions of the image and kernel of $\mathbb{S}$ via the rank-nullity theorem:

$$R = \mathrm{rank}\,\mathbb{S} + \dim \mathrm{Ker}\,\mathbb{S}\,. \tag{12}$$

It is known that the rank $M$ of the incidence matrix of any connected graph is $N-1$ [4] with the matrix possessing a sole left nullvector $\boldsymbol{\ell}_0 = (1, 1, \cdots, 1)$. This reflects the fact that the sum of the entries in any given column of $\mathbb{S}$ is zero. Physically, $\boldsymbol{\ell}_0$ has the status of a mass conservation law and one verifies that the dynamics in Eq. (6) indeed conserves the quantity $\boldsymbol{\ell}_0 \cdot \boldsymbol{x}(t)$: in a closed system, i.e. in the absence of fluxes in and out of the system, Lavoisier's law of mass conservation is satisfied. Algebraically, the $N$ species hence present $M = N-1$ independent degrees of freedom. Accordingly, one can use Euler's formula together with Eq. (12) to relate the number of cycles and cocycles to the fundamental subspaces of $\mathbb{S}$: one finds that the cycles are in number equal to the dimension of the kernel, $\dim \mathrm{Ker}\,\mathbb{S} = R - M$; and that the cocycles are in number equal to $M = \mathrm{rank}\,\mathbb{S}$, the number of independent columns of $\mathbb{S}$ (i.e. independent reactions).

In the previous section we introduced two indices, $\alpha$ and $\gamma$, to span the cycle and cocycle sets respectively. The same labeling can be introduced for the columns of $\mathbb{S}$. From the definition of $T_{\mathcal{G}}$, the $M$ columns labeled with $\gamma$ are (a choice of) linearly independent columns of $\mathbb{S}$. For convenience, we order them in such a way that $1 \leq \gamma \leq M$ and $M+1 \leq \alpha \leq R$. Then, we introduce two families of column vectors in $\mathbf{R}^R$, respectively denominated $\{\boldsymbol{c}^\alpha\}$ and $\{\boldsymbol{c}^\gamma\}$ and defined as:

$$(..., \boldsymbol{c}^\alpha, ...) = \begin{pmatrix} -\mathbb{T} \\ \mathbb{1}_{R-M} \end{pmatrix}, \quad (..., \boldsymbol{c}^\gamma, ...) = \begin{pmatrix} \mathbb{1}_M \\ \mathbb{T}^\top \end{pmatrix}. \tag{13}$$

Here $\mathbb{1}_n$ is the $n \times n$ identity matrix and $\mathbb{T}$ is a $M \times R-M$ rectangular matrix defined from the graph $\mathcal{G}$ as:

$$\mathbb{T}_{\gamma\alpha} = \begin{cases} +1\,, & \text{if cochord } \gamma \in C(\alpha) \text{ and } \parallel \text{ to the chord } \alpha \\ -1\,, & \text{if cochord } \gamma \in C(\alpha) \text{ and } \nparallel \text{ to the chord } \alpha \\ 0\,, & \text{otherwise,} \end{cases} \tag{14}$$

where $\parallel$ (resp. $\nparallel$) indicates that edges share (resp. do not share) the same orientation. As we now explain, the $\{\pm 1\}$ components of the vectors defined in Eq. (13) correspond to the oriented edges that compose the graph-theoretical cycles and cocycles defined in the previous paragraph. Indeed, the vector $\boldsymbol{c}^\alpha$ contains a non-zero entry for any edge that belongs to the cycle $C(\alpha)$: $c^\alpha_\rho \neq 0$ if and only if $\rho \in C(\alpha)$. By construction, $c^\alpha_{\alpha'} = \delta_{\alpha\alpha'}$ since in the chord set only the generating chord $\alpha$ belongs to $C(\alpha)$ and dictates its orientation. (Here and below, $\delta_{ij}$ denotes the Kronecker delta.) Analogously, the vector $\boldsymbol{c}^\gamma$ contains non-zero entries for any edge that belongs to the cocycle $C(\gamma)$, such that $c^\gamma_\rho \neq 0$ if and only if $\rho \in C(\gamma)$. By construction, $c^\gamma_{\gamma'} = \delta_{\gamma\gamma'}$ since the only cochord contained in $C(\gamma)$ is the generating one. Albeit not obvious, the same matrix $\mathbb{T}$ (up to a sign) controls the composition of both cycles and cocycles, as expressed by Eq. (13). This is the algebraic encoding of the duality discussed in the previous section.

Due to the identity matrices in Eq. (13) all vectors $\boldsymbol{c}^\alpha$ and $\boldsymbol{c}^\gamma$ are linearly independent and one easily checks that they span orthogonal subspaces, since:

$$\boldsymbol{c}^\gamma \cdot \boldsymbol{c}^\alpha = 0 \quad \forall\, \gamma, \alpha\,. \tag{15}$$

Furthermore, the geometric construction ensures that the vectors $\boldsymbol{c}^\alpha$ belong to the kernel of $\mathbb{S}$, that is

$$\sum_\rho \mathbb{S}_{i\rho}\, c^\alpha_\rho = 0 \quad \forall\, \alpha, i\,. \tag{16}$$

This represents the fact that any node $i$ in a cycle has exactly one incoming and one outgoing edge. As a consequence, cycles and cocycles form a basis for, respectively, the kernel of $\mathbb{S}$ and its orthogonal complement $\mathrm{Im}\,\mathbb{S}^\top$, i.e. the coimage of $\mathbb{S}$. This is the algebraic interpretation of cycles and cocycles, which complements their definition from graph theory. Likewise, a vectorial representation holds for the chords and the cochords. Those are the canonical vectors $e^\gamma_\rho = \delta_{\gamma\rho}$ and $e^\alpha_\rho = \delta_{\alpha\rho}$ in $\mathbf{R}^R$. All in all, we have identified two alternative bases for $\mathbf{R}^R$, that we can merge in the following two matrices:

$$(\boldsymbol{e}^\gamma, \boldsymbol{e}^\alpha) = \begin{pmatrix} \mathbb{1}_M & \mathbb{0} \\ \mathbb{0} & \mathbb{1}_{R-M} \end{pmatrix}, \quad (\boldsymbol{c}^\gamma, \boldsymbol{c}^\alpha) = \begin{pmatrix} \mathbb{1}_M & -\mathbb{T} \\ \mathbb{T}^\top & \mathbb{1}_{R-M} \end{pmatrix}. \tag{17}$$

The left-hand matrix is the canonical basis in $\mathbf{R}^R$ obtained from the chords/cochords vectorial representation. The right-hand matrix is the non-orthogonal basis formed by the vectorial representation of cycles and cocycles. One easily verifies the orthogonality relations:

$$\begin{aligned} \boldsymbol{e}^\alpha \cdot \boldsymbol{c}^{\alpha'} &= \delta_{\alpha\alpha'} \\ \boldsymbol{e}^\gamma \cdot \boldsymbol{c}^{\gamma'} &= \delta_{\gamma\gamma'}\,. \end{aligned} \tag{18}$$

For convenience, we will call cycles (resp. cocycles) both the vectors $\{\boldsymbol{c}^\alpha\}$ (resp. $\{\boldsymbol{c}^\gamma\}$) and the graph-theoretical objects defined in the previous paragraph, as they are equivalent. We designate by chemical cycles the space generated by the $\boldsymbol{c}^\alpha$'s.

### C. A physical decomposition for affinities and currents

Equipped with both the geometrical and the algebraic frameworks, we now turn our attention to their use to represent the currents $\boldsymbol{J}$ and the affinities $\boldsymbol{A}$ in noninteracting CRNs. Those are the vectors defined in Eqs. (5)-(7) and that control

[4] Indeed, from the rank-nullity theorem applied to $\mathbb{S}^\top$, one has $\mathrm{rank}\,\mathbb{S} = \mathrm{rank}\,\mathbb{S}^\top = N - \dim \mathrm{Ker}\,\mathbb{S}^\top = N-1$ since $\dim \mathrm{Ker}\,\mathbb{S}^\top = 1$, as indeed $\mathrm{Ker}\,\mathbb{S}^\top$ is spanned by $\boldsymbol{\ell}_0^\top = (1, \cdots, 1)^\top$. It contains no other independent vector: *ad absurdum*, if such a vector would exist, one could build, by linear combination with $\boldsymbol{\ell}_0^\top$, a non-zero vector $\boldsymbol{\ell}^\top \in \mathrm{Ker}\,\mathbb{S}^\top$ containing a 0 component $\ell_i = 0$; this is impossible, since using $\boldsymbol{\ell}\mathbb{S} = 0$ by recursion along the connected graph $\mathcal{G}$ starting from node $i$, we find $\ell_j = 0, \forall j$. See also for instance Ref. [45].

the dynamics and the thermodynamics. Following Ref. [14], we introduce a decomposition of the microscopic affinities $\boldsymbol{A} \in \mathbf{R}^R$ in terms of cocycles and chords as:

$$\boldsymbol{A} = \sum_\gamma A^c_\gamma \boldsymbol{c}^\gamma + \sum_\alpha A^e_\alpha \boldsymbol{e}^\alpha \,. \tag{19}$$

Notice that this decomposition is not orthogonal since $\boldsymbol{c}^\gamma \cdot \boldsymbol{e}^\alpha \neq 0$. Nevertheless, it bears a clear physical interpretation. As we will show, the first term in Eq. (19) embeds the conservative part of $\boldsymbol{A}$, while the second term contains the non-conservative part. As such, Eq. (19) can be viewed as a generalization of the Helmholtz–Hodge decomposition of the affinities on a graph. In continuous space, a test for conservativeness is Stokes' theorem, which states that the circulation – the line integral along any closed path – of a conservative force field vanishes. Consider now a generic closed path defined on the graph. Any such path can be expressed as a linear combination of the basis of cycles, as it belongs to $\mathrm{Ker}\,\mathbb{S}$ since applying its reactions leave the system invariant [5]. Thus to compute the circulation of $\boldsymbol{A}$ it is sufficient to compute its scalar product with the cycles $\boldsymbol{c}^\alpha$'s. Using the orthogonality conditions Eqs. (15) and (18), one gets:

$$\boldsymbol{c}^\alpha \cdot \boldsymbol{A} = A^e_\alpha \quad \forall \alpha \,. \tag{20}$$

Thus, the coefficient $A^e_\alpha$ corresponds to the affinities integrated along cycle $\boldsymbol{c}^\alpha$, and quantifies the deviation from Stokes' theorem. One remarks that the cycle affinities $\{A^e_\alpha\}$ do not depend on the system concentrations [13]: They constitute a set of parameters that are intrinsic to the dynamics and quantify the nonequilibrium drive. We express the conservative condition for $\boldsymbol{A}$ as the requirement for all coefficients $A^e_\alpha$ to vanish:

$$\boldsymbol{c}^\alpha \cdot \boldsymbol{A} = A^e_\alpha = 0 \;\; \forall \alpha \,, \tag{21}$$

which is an instance of Kirchhoff Voltage Law (KVL) in electric circuits [13]. Let us assume for a moment that $\boldsymbol{A}$ is conservative and analyze what it means algebraically and geometrically. Algebraically, the conservative condition Eq. (21) is equivalent to $\boldsymbol{A} \in (\mathrm{Ker}\,\mathbb{S})^\perp = \mathrm{Im}\,\mathbb{S}^\top$ [see Eq. (11)], i.e. to the potential condition $\exists \boldsymbol{V} : \boldsymbol{A} = -\mathbb{S}^\top \boldsymbol{V}$. Noting that $\mathrm{Im}\,\mathbb{S}^\top$ is the space spanned by the cocycles, these conditions are encoded in the decomposition of Eq. (19) which, in this case, reduces to $\boldsymbol{A} = \sum_\gamma A^c_\gamma \boldsymbol{c}^\gamma$ with the affinity being a conservative force with zero circulation. Since cocycles are a basis for $\mathrm{Im}\,\mathbb{S}^\top$, one can interpret $\mathbb{S}^\top$ as a discrete 'gradient' on the graph, with the coefficients $A^c_\gamma = \boldsymbol{e}^\gamma \cdot \boldsymbol{A}$ corresponding to local potential drops along each cochord. Here the main difficulty to solve for the potential $\boldsymbol{V}$ in $\boldsymbol{A} = -\mathbb{S}^\top \boldsymbol{V}$ is the non-invertibility of $\mathbb{S}$, which prevents the identification of a discrete 'integrator' associated to $\mathbb{S}^\top$. Once again, graph theory comes in handy. Upon fixing arbitrarily a reference root node, we orient all edges in $T_\mathcal{G}$ towards it and define $\forall i$ the subset $\mathcal{U}(i)$ containing all the nodes that are upstream the node $i$ along the spanning tree, including $i$ itself; also, we number the reactions starting from the root (see Appendix A for details). Thus, we introduce the $N \times N$ square matrix $\mathbb{G}$ defined as

$$\mathbb{G}_{ij} = \begin{cases} 1\,, & \text{if node } j \in \mathcal{U}(i) \\ 0\,, & \text{otherwise.} \end{cases} \tag{22}$$

which, is invertible. See Fig. 3 for an example. In Appendix A we prove that the matrix $\mathbb{G}^\top$ then takes the form of a lower triangular integration operator on $T_\mathcal{G}$: namely, if $\boldsymbol{A}$ is conservative, a solution to $\boldsymbol{A} = -\mathbb{S}^\top \boldsymbol{V}$ is given by the matrix $\mathbb{G}^\top$ in the following way. For an arbitrary affinity vector $\boldsymbol{A}$, let us define a potential $V_i$ on each node from the set of coefficients $A^c_\gamma$ as:

$$V_i[A^c_\gamma] = \sum_{1 \leq \gamma \leq M} \left(\mathbb{G}^\top\right)_{i\, s(\gamma)} A^c_\gamma = \sum_{1 \leq \gamma \leq M} A^c_\gamma \, \delta_{i \in \mathcal{U}(s(\gamma))} \,. \tag{23}$$

The sum in Eq. (23) runs over all the edges in $T_\mathcal{G}$. Using the definition of $\mathbb{G}$ in Eq. (22) one sees that, for fixed $i$, $\sum_j (\mathbb{G}^\top)_{ij}$

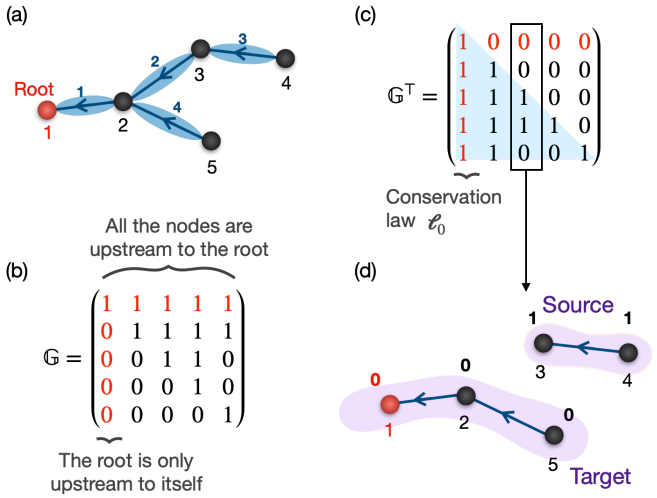



FIG. 3. **(a)** Spanning tree for the example of CRN in Fig. 1 where we have picked node 1 (red) as the root and oriented all edges towards it. The enumeration of the nodes (in black) and the edges (bold blue) follows the natural convention detailed in Appendix A. It entails a simple one-to-one correspondence between the nodes $j \neq$ root (in black) and the cochords $\gamma \in T(\mathcal{G})$: namely, $\forall j \neq$ root $\exists \gamma$ s.t. $j = \gamma + 1 = s(\gamma)$. **(b)** The $N \times N$ matrix $\mathbb{G}$ for the aforementioned spanning tree, constructed as detailed in the main text. The first line and first column (in red) refer to the root and reflect the fact that we have oriented all the edges towards it. As a consequence, all the nodes $\neq$ root are upstream to it, hence the first line full of one's. **(c)** The matrix $\mathbb{G}^\top$ presents the low triangular structure of an integral operator on the spanning tree. The first column (in red) coincides to the mass conservation law $\boldsymbol{\ell}_0$ and the zeroes in the first line (red) make that the potential of the root defined from Eq. (23) is zero: $V_{\text{root}} = 0$. **(d)** The last $M = N - 1$ columns of $\mathbb{G}^\top$ are in one-to-one correspondence with the $M$ cocycles. In particular, one reads from the $+1$ entries of the column $j = \gamma + 1$ the source island of the cocycle $C(\gamma)$. For example, if we remove the cochord $\gamma = 2$ from the spanning tree in panel (a), we generate a cut with the nodes 3 and 4 being completely disconnected from the rest of the graph. Algebraically, the appearance of a source island corresponds to the emergence of a new conservation law. Indeed the last $M$ columns of $\mathbb{G}^\top$ correspond to the conservation laws obtained by removing the cocycles.

[5] Notice that a closed path may pass through bridges, but since any bridge must be crossed an even number of times in opposite directions along such path, the bridges do not contribute.

runs over all the nodes which have $i$ among their upstreamers. Moreover, for every node $j$, except the root, there exists exactly one $\gamma$ such that $j = s(\gamma) = \gamma + 1$. Thus, the sum in Eq. (23) runs over the unique path on $T(\mathcal{G})$ between node $i$ and the root. Hence, it corresponds to a discrete integration of the cochord affinities $A^c_\gamma$ along the spanning tree. Notice that, from Eq. (23), the potential of the root is zero since the root, by convention, only has incoming edges. This means that the potential in Eq. (23) is uniquely defined up to a constant shift (fixed by $V_{\text{root}} = 0$), in analogy to the constant of integration in standard calculus. Accordingly, the coefficients $A^c_\gamma$ are expressed as a local potential difference:

$$A^c_\gamma = V_{s(\gamma)} - V_{t(\gamma)} = -\left(\mathbb{S}^\top \boldsymbol{V}\right)_\gamma . \tag{24}$$

Eqs. (24) and (20) provide a direct physical interpretation for the set of affinity coefficients $A^c_\gamma$ and $A^e_\alpha$ in Eq. (19). The former are the conservative affinities expressed as a discrete potential difference along the corresponding cochord. The latter are the non-conservative affinities expressed as the non-local integration of the full affinity along the cycles. As such we refer to them as macroscopic affinities, in contrast to the affinities $A_\rho$ defined for each reaction by Eq. (7).

Notably, the matrix product $\mathbb{S}^\top \mathbb{G}^\top$ takes the form (see Appendix A for a proof), discussed in the introduction,

$$-\mathbb{S}^\top \mathbb{G}^\top = \begin{pmatrix} \mathbf{0} & \mathbb{1}_M \\ \mathbf{0} & \mathbb{T}^\top \end{pmatrix} \tag{25}$$

where the first column is full of zeroes and the matrix $\mathbb{T}$ is the same matrix as defined in Eq. (14). This special structure encodes the fact that the stoichiometric matrix is not full-rank, but it contains some built-in redundancy. The $M \times M$ square identity matrix represents the inversion procedure between $\mathbb{S}^\top$ and $\mathbb{G}^\top$, illustrated in Eqs. (23)-(24). The first column $(0 \dots 0)^\top$ reflects the existence of the conservation law $\boldsymbol{\ell}_0$ and the $\mathbb{T}$ matrix reflects the interdependence among reactions. Namely, only $M$ out of the $R$ columns of $\mathbb{S}$ are linearly independent while the remaining columns, labeled with $\alpha$ and associated to the cycles $\boldsymbol{c}^\alpha$, can be obtained as a linear combination of the former. This is what the matrix $\mathbb{T}$ encodes: denoting by $\mathbb{S}_M$ the first $M$ (independent) columns of $\mathbb{S}$ and $\mathbb{S}_{\text{dep}}$ the last ones (corresponding to $R - M$ dependent reactions), one reads from Eq. (25) that $\mathbb{S}_{\text{dep}} = \mathbb{S}_M \mathbb{T}$. The relation in Eq. (25) is key in our analysis because it lies at the core of its extension from noninteracting to interacting CRNs, presented in Sec. III.

Notice also that the non-zero columns of the right-hand side (r.h.s.) of Eq. (25) correspond precisely to the cocycle basis [Eq. (13)] for the conservative affinities $\in \text{Im}\,\mathbb{S}^\top$. This means that the columns of $\mathbb{G}^\top$, except the first one associated to the root, can be seen as the $M = N - 1$ potential landscapes which, upon 'differentiation' via $\mathbb{S}^\top$, give the cocycle vectors $\boldsymbol{c}^\gamma$'s. We thus define the potential vector $\boldsymbol{v}^\gamma = (\mathbb{G}^\top)_{s(\gamma)}$, where $(\mathbb{G}^\top)_{s(\gamma)}$ indicates the column $s(\gamma)$ of $\mathbb{G}^\top$. By construction, $\boldsymbol{c}^\gamma = -\mathbb{S}^\top \boldsymbol{v}^\gamma$ and the entries of $\boldsymbol{v}^\gamma$ are only zeroes or ones, so that $\boldsymbol{v}^\gamma$ is a characteristic potential landscape defined on the nodes of $\mathcal{G}$. It is in fact characteristic of the cut generated by the removal of the cocycle $\boldsymbol{c}^\gamma$, as discussed previously. In particular, the target island, containing the root, is the sub-graph held at zero potential while the source island, corresponding to the $+1$ entries in $\boldsymbol{v}^\gamma$, is held at unit potential (see Fig. 3d).

Let us now introduce the complementary decomposition for the vector of currents $\boldsymbol{J} \in \mathbf{R}^R$ in terms of cochords and cycles:

$$\boldsymbol{J} = \sum_\gamma J^e_\gamma \boldsymbol{e}^\gamma + \sum_\alpha J^c_\alpha \boldsymbol{c}^\alpha . \tag{26}$$

Once again, one can identify an analog of it for vector calculus in continuous space. Any cocycle is associated with a cut of the full graph into two sub-graphs, in the same way as a closed surface splits $\mathbf{R}^3$ into an inner and an outer region. In particular the cocycle $\boldsymbol{c}^\gamma$ generates a splitting of the graph into a source island, the sub-graph containing the source node $s(\gamma)$, and a target island, the sub-graph containing the target node $t(\gamma)$ (see Fig. 2c). By construction, any flux between the source and the target must flow through the cocycle itself. If we take the scalar product between Eq. (26) and a cocycle $\boldsymbol{c}^\gamma$ one gets:

$$\boldsymbol{c}^\gamma \cdot \boldsymbol{J} = J^e_\gamma \quad \forall \gamma . \tag{27}$$

Thus, the coefficient $J^e_\gamma$ in the proposed decomposition represents the total current flowing from the source to the target along the corresponding cocycle $\boldsymbol{c}^\gamma$. The scalar product in Eq. (27) corresponds to a surface integral, i.e. a 'flux' across the 'boundary' of the source. Using the definition of $\boldsymbol{v}^\gamma$, one has

$$\boldsymbol{c}^\gamma \cdot \boldsymbol{J} = (-\mathbb{S}^\top \boldsymbol{v}^\gamma) \cdot \boldsymbol{J} = -\boldsymbol{v}^\gamma \cdot (\mathbb{S}\boldsymbol{J}) , \tag{28}$$

where on the r.h.s. we recognize the graph divergence $\mathbb{S}\boldsymbol{J}$ entering in Eq. (6). As such, Eq. (28) is the analog of a divergence theorem for the graph: the current across the boundary of the source is equal to the volume integral over the source of the divergence of the current. Finally, the coefficient $J^c_\alpha = \boldsymbol{e}^\alpha \cdot \boldsymbol{J}$ corresponds to the local current flowing along chord $\alpha$.

Assuming that in the long time limit the dynamics reaches a stationary state, $\boldsymbol{x}^* = \lim_{t\to\infty} \boldsymbol{x}(t)$, then the stationary current, $\boldsymbol{J}^* = J(\boldsymbol{x}^*)$, belongs to the kernel of $\mathbb{S}$ [see Eq. (6)]:

$$\mathbb{S}\boldsymbol{J}^* = 0 . \tag{29}$$

Eq. (29) corresponds to the Kirchhoff Current Law (KCL) [13] and is encoded in the decomposition of Eq. (26) which reduces to a linear combination of the cycles, $\boldsymbol{J}^* = \sum_\alpha J^{c,*}_\alpha \boldsymbol{c}^\alpha$. It follows that the currents $J^e_\gamma$ flowing from the source to the target of each cocycle are transient and they vanish at steady state. Notice that, from the graph-theoretical viewpoint, Eq. (29) ensures the balance between all currents entering and exiting at each node.

We stress that the advantage of the two decompositions expressed in Eqs. (19) and (26) lies in their direct physical interpretation. On the one side the condition of thermodynamic feasibility is expressed in the vanishing (at all times) of the non-conservative affinities, Eq. (21). On the other, the condition of stationarity is expressed in the vanishing (in the long

time limit) of the transient currents, Eq. (27). For this reason, the coefficients of the two decompositions are endowed with the status of macroscopic observables.

Importantly, the decomposition of affinity also applies to the underlying stochastic level of description. Consider now that the affinity $\boldsymbol{A}$ and current $\boldsymbol{J}$ are evaluated in a vector $\boldsymbol{x}$ whose components $x_i$ are the number of molecules of species $i$, divided by the volume [instead as above of evaluating them in $\boldsymbol{x}(t)$ solution of the deterministic rate equation, Eq. (6)]. One sees that having a conservative affinity is equivalent to having zero cycle affinities [$A^c_\alpha = 0\,, \forall\alpha$ in Eq. (19)], which, thanks to Kolmogorov's criterion [46], is equivalent to detailed balance for the microscopic rates, and thus to stochastic reversibility of the underlying process. Complete proofs are given in Appendix B (for generic interacting CRNs). Hence, although the condition to have a conservative affinity seems to only pertain to the deterministic level, it applies also to the stochastic one. Such characterizations of reversibility are analogous to that of a Langevin equation of the form $\partial_t \mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t)) + \boldsymbol{\eta}(t)$ (with $\mathbf{F}(\mathbf{x})$ the force and $\boldsymbol{\eta}(t)$ a centered Gaussian white noise). Indeed, there the process $\mathbf{x}(t)$ is stochastically reversible if and only if the force derives from a potential, if and only if its circulation along any loop is zero. Connections between the stochastic level of description and thermodynamics are discussed in Sec. III D.

We conclude the section by recalling the standard definition of entropy production rate $\sigma$ in terms of currents and affinities [47]:

$$\sigma = \boldsymbol{J}\cdot\boldsymbol{A} = \sum_\gamma J^e_\gamma A^c_\gamma + \sum_\alpha J^c_\alpha A^e_\alpha\,, \tag{30}$$

where in the second equality we made use of Eqs. (19) and (26). Notice that the first contribution vanishes at steady state while the second contribution vanishes for reversible dynamics.

### D. Closed CRNs: connection to thermodynamics

In order to make contact with equilibrium thermodynamics, one wants the evolution in Eq. (6) for a closed system, i.e. in the absence of couplings with external reservoirs [48], to relax to an equilibrium state $\boldsymbol{x}^{\rm eq} = \lim_{t\to\infty}\boldsymbol{x}(t)$, fixed by the initial conditions. In this scenario, the internal currents are driven by nonequilibrium initial conditions and are expected to vanish at steady state, $\boldsymbol{J}^{\rm eq} = \boldsymbol{J}(\boldsymbol{x}^{\rm eq}) = 0$. Such thermodynamic behavior is guaranteed by choosing the reaction rates in accordance with the Wegscheider criterion [49], which states that the product of the forward rates along any cycles $\boldsymbol{c}^\alpha$ is equal to that of the backward rates:

$$\prod_{\rho=1}^{R}\left(\frac{k^+_\rho}{k^-_\rho}\right)^{c^\alpha_\rho} = 1\,, \quad \forall\alpha\,. \tag{31}$$

Eq. (31) is a necessary and sufficient condition for the dynamics in Eq. (6) to relax to an equilibrium steady state with $\boldsymbol{A}^{\rm eq} = \boldsymbol{A}(\boldsymbol{x}^{\rm eq}) = 0$ and is equivalent to the existence of a (Poissonian) distribution with detailed balance for the underlying population dynamics (see Appendix B for a proof). Moreover, from a graph-theoretical perspective, Eq. (31) implies the existence of a time-independent potential defined for each node/species $i$ taking the form:

$$\exp\left(\mu^\ominus_i\right) = \prod_{\rho\,\in\,[\text{root}\to i]}\frac{k^+_\rho}{k^-_\rho} \tag{32}$$

where the product is taken along any arbitrary path on the graph $\mathcal{G}$ from the root to node $i$. Combining Eqs. (31) and (32) one obtains a local detailed balance condition for noninteracting CRNs [9, 12]:

$$\frac{k^+_\rho}{k^-_\rho} = \exp\left(\mu^\ominus_{s(\rho)} - \mu^\ominus_{t(\rho)}\right) = \exp\left[-(\mathbb{S}^\top\boldsymbol{\mu}^\ominus)_\rho\right], \tag{33}$$

where one recognizes $\boldsymbol{\mu}^\ominus$ to play the role of the (dimensionless) standard chemical potential of equilibrium thermodynamics [48]. Notice that Eq. (33) is also the standard condition of detailed balance for Markov chains with respect to a configuration probability $\mathcal{P}_i \propto \exp\left(-\mu^\ominus_i\right)$. The analogy with classical thermodynamics is completed by introducing a Hopf–Cole-like transformation for the species concentrations $x_i(t) = \exp\left[\mu_i(t) - \mu^\ominus_i\right]$ in terms of the chemical potential $\mu_i(t)$. By replacing it in Eq. (7) together with Eq. (33) one gets an expression for the affinities of a closed system at all times:

$$A_\rho = -\left(\mathbb{S}^\top\boldsymbol{\mu}\right)_\rho = \mu_{s(\rho)} - \mu_{t(\rho)}\,. \tag{34}$$

Interestingly, one sees that the potential $V_i$ introduced in the previous section and constructed from integrating the cocycle affinities $A^c_\gamma$ along the spanning tree coincides (up to a constant shift) with the usual chemical potential of thermodynamics. Thus, we have seen how the potentials of the chemical setting can be constructed directly from the graph representation.

From Eq. (34) one sees that, at the level of macroscopic affinities, the Wegscheider condition ensures $A^e_\alpha = 0$ at all times, which is KVL Eq. (21). Hence, for closed systems (i.e. reversible dynamics) only the cocycle affinities $A^{\rm c}_\gamma$ survive at finite time, and vanish in the long time limit $A^{\rm c}_\gamma(t\to\infty) = 0$; this describes the process of relaxation to equilibrium. Algebraically this implies that the equilibrium chemical potential $\boldsymbol{\mu}^{\rm eq} = \lim_{t\to\infty}\boldsymbol{\mu}(t)$ is a left nullvector of the stoichiometric matrix, $\boldsymbol{\mu}^{\rm eq}\cdot\mathbb{S} = 0$. For closed noninteracting CRNs, $\boldsymbol{\mu}^{\rm eq}$ is therefore proportional to the mass conservation law $\boldsymbol{\ell}_0$ and the equilibrium state is reached when all the chemical species have relaxed to the same value of chemical potential.

### E. Open CRNs: explicit chemostatting

We now suppose that the reaction rate constants $k^\pm_\rho$ are chosen arbitrarily and do not fulfill the Wegscheider condition in Eq. (31). This is often the case in phenomenological models of evolutionary games [50], gene regulatory networks [51] or theoretical ecology [52] where effective reactions are typically irreversible. As a result, the dynamics evolves irreversibly towards a nonequilibrium steady state or a limit cycle. In order

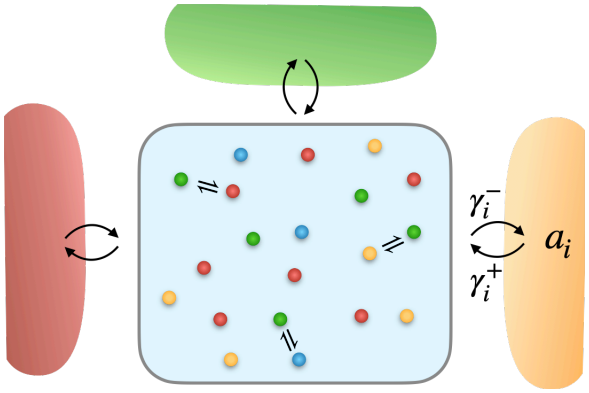



FIG. 4. Sketch of a system of chemical species put in contact with external chemostats. The chemostats are treated as infinite reservoirs of a single chemical species. For illustration, the chemostat $i$ (here in orange) exchanges orange particles with the system and it is characterized by a pair of reaction rates $\gamma_i^\pm$ and a driving parameter $a_i$.

to make connection with thermodynamics while still being irreversible, in CRNs, the breakdown of the Wegscheider condition is usually prescribed through the coupling with different chemostats which drive the system out of equilibrium [53]. Each chemostat is depicted as a reservoir of a given chemical species which is put in contact with the system to favor the exchange of molecules (Fig. 4). In this picture, an external current builds up between the chemostat and the system affecting the internal concentration of the chemostatted species [6]. In the model, this is done by introducing a new reaction for each of the chemostatting couplings:

$$\emptyset \underset{\gamma_i^-}{\overset{\gamma_i^+}{\rightleftharpoons}} Y_i \,, \tag{35}$$

where we indicate with $Y_i$ the chemostatted species and with $\gamma_i^\pm$ the rate of particles interchange with the chemostat. Since reactions like Eq. (35) involve exactly one chemical species we make use of the same index $i = 1, ..., |Y|$ for both the chemostatted species and the chemostatting reaction. In the microscopic picture, a molecule of $Y_i$ is injected into the system at rate $\Omega\gamma_i^+$ and expelled from it at rate $\gamma_i^-$. Those rates reflect the action of the chemostat and, contrarily to the bulk rates $\{k_\rho^\pm\}$, are not thermodynamically constrained by the Wegscheider condition. The extent of the Wegscheider violation is then quantified via a driving parameter $a_i$, for each chemostat

$$\frac{\gamma_i^-}{\gamma_i^+} = \exp\left(\mu_i^\ominus - a_i\right) , \tag{36}$$

which plays the role of an external parameter characterizing the coupling with the $i-$chemostat.

Following Refs. [10, 12], we label with $Y$ the set of chemostatted species and with $X$ the remaining non-chemostatted chemical species, such that the union $X \cup Y$ forms the set of all chemical species in the systems. Accordingly, the concentrations of the chemostatted species are grouped in the vector $\boldsymbol{y}(t)$ and that of the non-chemostatted species in $\boldsymbol{x}(t)$. Then, internal reactions Eq. (9) and external reactions Eq. (35) are combined into an $N \times (R + |Y|)$ extended stoichiometric matrix $\mathbb{S}_{\rm res}$, namely:

$$\mathbb{S}_{\rm res} = \begin{pmatrix} \mathbb{1}_{|Y|} & \mathbb{S}_Y \\ \mathbb{0} & \mathbb{S}_X \end{pmatrix} \quad \text{with} \quad \mathbb{S} = \begin{pmatrix} \mathbb{S}_Y \\ \mathbb{S}_X \end{pmatrix} , \tag{37}$$

where the first $|Y|$ columns correspond to the chemostatting reactions and we split the stoichiometric matrix in Eq. (10) into chemostatted and non-chemostatted species. The evolution of the concentrations now reads:

$$\frac{d}{dt}\begin{pmatrix} \boldsymbol{y}(t) \\ \boldsymbol{x}(t) \end{pmatrix} = \mathbb{S}_{\rm res} \boldsymbol{J}(\boldsymbol{x}(t), \boldsymbol{y}(t)) \,, \tag{38}$$

where the first $|Y|$ component of the vector $\boldsymbol{J}$ are the external currents of the chemostatting reactions:

$$\begin{aligned} J_i &= \gamma_i^+ - \gamma_i^- y_i \\ &= \gamma_i^+ \left[1 - \exp\left(\mu_i - a_i\right)\right] , \quad \text{for } 1 < i < |Y| \,, \end{aligned} \tag{39}$$

expressed as a function of the chemical potential $\mu_i$ and the driving parameter $a_i$. The set of affinities $\boldsymbol{A}$ associated to $\boldsymbol{J}$ can be conveniently expressed as:

$$\boldsymbol{A} = -\mathbb{S}_{\rm res}^\top \boldsymbol{\mu} + \boldsymbol{a} \,, \tag{40}$$

where we have introduced the vector of driving affinities $\boldsymbol{a} \in \mathbf{R}^R$ containing the collection of all the $a_i$'s. One sees directly from the expression of Eq. (40) how the chemostatting procedure might lead to nonequilibrium behavior. For generic values of the parameters $\{a_i\}$, $\boldsymbol{a} \notin \mathrm{Im}\, \mathbb{S}^\top$ and the potential condition discussed in Sec. II C breaks down.

Let us now consider a single chemostat with external parameter $a$ that is put in contact with a noninteracting CRN. The extended stoichiometric matrix in this case reads:

$$\mathbb{S}_{\rm res} = \left(\begin{array}{c|c} 1 & \\ \mathbf{0} & \mathbb{S} \end{array}\right) , \tag{41}$$

and one sees that, due to the first column $(1, 0 \cdots 0)^\top$, the conservation law $\boldsymbol{\ell}_0$ is no longer a left nullvector of the stoichiometric matrix. Intuitively, the effect of the chemostat is to break down internal mass conservation. This is expected since the internal mass is no longer conserved in the open system. Also, by construction, the left nullvectors of $\mathbb{S}_{\rm res}$ are left nullvectors of $\mathbb{S}$ so that $\mathbb{S}_{\rm res}$ is now full row-rank with $\mathrm{rank}\, \mathbb{S}_{\rm res} = N$. In practice, a single chemostat can not drive the system out of equilibrium since it does not affect the Wegscheider condition, hence the reversibility of the dynamics. *Proof.* First, notice that the driving affinity takes the form $\boldsymbol{a} = (a, 0 \ldots 0)^\top$. By construction, $\boldsymbol{\ell}_0 \cdot \mathbb{S}_{\rm res} = (1, 0 \ldots 0)^\top$ which constitutes a generating vector for the space where $\boldsymbol{a}$ lives. It follows that $\boldsymbol{a} = a\, \boldsymbol{\ell}_0 \cdot \mathbb{S}_{\rm res}$ and the affinity becomes:

$$\boldsymbol{A} = -\mathbb{S}_{\rm res}^\top \left[\boldsymbol{\mu}(t) - a\, \boldsymbol{\ell}_0\right] . \tag{42}$$

[6] The chemostat acts as a reservoir of molecules. Ideal chemostats exchange matter with the system while remaining in equilibrium.

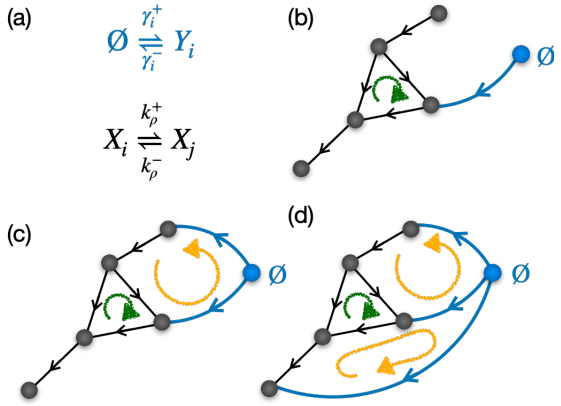



FIG. 5. Example of the graphical representation of chemostatting for unimolecular reactions. **(a)** Species labeled $Y_i$ are subjected to chemostatting reactions (in blue) with rates $\gamma_i^\pm$, describing the contact with reservoirs. **(b)** Chemostatting one species corresponds to the simple addition of one edge in the graph which connects the node $\emptyset$ to the corresponding chemostatted species/node. In this case, the only cycle present in the graph is the internal one (in green). **(c)** Chemostatting a second species results in the emergence of a new cycle, depicted in yellow. **(d)** The same holds true when chemostatting a third species, and so on. Such emergent cycles (in yellow) are not constrained by the Wegscheider condition on the rates and they can drive the system out of equilibrium.

The total affinity $\boldsymbol{A} \in \mathrm{Im}\,\mathbb{S}_{\rm res}^\top$ is therefore still conservative and the dynamics still reversible □

From Eq. (42), the equilibrium point $\boldsymbol{A}^* = 0$ is given by $\boldsymbol{\mu}^{\rm eq} = a\ \boldsymbol{\ell}_0$. As for the closed equilibrium, $\boldsymbol{\mu}^{\rm eq}$ is expressed in terms of the law of internal mass conservation $\boldsymbol{\ell}_0$ with the driving parameter $a$ playing the role of an effective initial condition. Indeed, one can think of $a$ as the initial value of chemical potential inside the reservoir which is not allowed to relax on the relevant timescales of the dynamics. The chemostatting of a single species thus results in a global adaptation of the chemical potential of all the species to the one of the reservoir, which is a consequence of the internal mass conservation law $\boldsymbol{\ell}_0$. Notice that the previous proof extends to an arbitrary number of chemostats provided that they all share the same value of driving parameter $a$. Whenever the driving force is parameterized by a single parameter $a$, the single conservation law $\boldsymbol{\ell}_0$ of the closed system is sufficient to form a basis for $\boldsymbol{a}$. As a consequence, in order to drive a CRN out of equilibrium, at least two chemostats are needed which are held at different values of driving parameter.

Some procedures of chemostatting in the context of CRNs [10, 12, 13] differ from the present treatment as they assume the chemostats to fully fix the concentrations of the chemostatted species. Accordingly, the $Y_i$'s are treated as external parameters and the evolution is restricted to the $X_i$'s as controlled by the sub-matrix $\mathbb{S}_X$. The present approach is advantageous in that it allows one to extend the graphical representation introduced in Sec. II A-II C to noninteracting open systems. Apparently, in presence of chemostats, the graphical representation is lost since $\mathbb{S}_{\rm res}$ (just like $\mathbb{S}_X$) does not take the form of an incidence matrix. An open question is therefore whether it is possible to understand the role played by chemostats graph-theoretically. To answer it, we propose the following: augment the original graph by an extra node $\emptyset$ and, for each chemostatting reaction in Eq. (35), introduce a new edge linking the chemostatted species $y_i$ to the node $\emptyset$ (see Fig. 5). By construction, the incidence matrix of the augmented graph is endowed with an extra line:

$$\left(\begin{array}{c|c} -1\cdots-1 & 0\ldots0 \\ \hline \multicolumn{2}{c}{\mathbb{S}_{\rm res}} \end{array}\right) \tag{43}$$

which bookkeeps the exchange of particles between the chemostatting node $\emptyset$ and the system. As shown in Fig. 5d, adding more than one chemostatting edges results in the appearance of new cycles in the graph. Those 'emergent' cycles are associated to cycle affinities that do not necessarily satisfy the Wegscheider condition, and thus play the role of driving the system out of equilibrium. Again, one sees that at least two chemostats are needed to introduce an emergent cycle. Following the procedure outlined in Sec. II A one can build the $\mathbb{T}$ matrix from the augmented graph associated with the incidence matrix of Eq. (43) and find the corresponding basis for cycles and cocycles. Now the set of cycles $\{\boldsymbol{c}^\alpha\}$ includes both the internal and the emergent cycles, which are graphically identified. Notice that the cycles identified from $\mathbb{T}$ form a basis for $\mathrm{Ker}\,\mathbb{S}_{\rm res}$, despite the fact that the latter is not an incidence matrix. *Proof.* By construction, any right nullvector of (43) is also a right nullvector of $\mathbb{S}_{\rm res}$. Furthermore, the extra line in the matrix of Eq. (43) is not linear independent from the rows of $\mathbb{S}_{\rm res}$ and it can be obtained by taking the negative sum of all the rows of $\mathbb{S}_{\rm res}$, which is full-row rank as recalled previously. It follows that the kernel of the two matrices must be the same □ As a consequence, the number of emergent cycles is given by:

$$\#\,\text{emergent cycles} = \dim \mathrm{Ker}\,\mathbb{S}_{\rm res} - \dim \mathrm{Ker}\,\mathbb{S}\,. \tag{44}$$

This relation shows that the number of emergent cycles, defined above from the analysis of the graph associated to the augmented stoichiometric matrix, Eq. (43), happens to be in accordance with the algebraic result of Ref. [12] (for generic interacting CRNs), which showed that the number of independent chemical cycles induced by chemostatting is given by the r.h.s. of Eq. (44).

## III. INTERACTING REACTION NETWORKS

### A. Motivations

In the previous section, we studied the case of noninteracting networks where each reaction, Eq. (9), involves the transformation of a single molecule at a time. While often employed in the effective descriptions of molecular machines [54, 55] or gene expression [19, 56], these networks are highly atypical for real biochemical scenarios where interactions play a fundamental role. For example, a large branch of inorganic chemistry is based on heterogeneous catalysis, where interactions among different chemical species increase

the efficiency of reaction pathways [57]. Likewise, in intracellular processes and in prebiotic scenarios, autocatalytic interactions are at the core of the capability of living systems to self-replicate [23]. Among others, metabolic cycles (e.g. the glycolysis network in the introduction) are the archetype of biochemical networks where interactions are essential.

In the chemical setting, an interaction involves the collision of two (or more) molecules which activates a change in their internal state. The following kinetic scheme involves two different types of such interactions:

$$X_1 + X_2 \overset{1}{\rightleftharpoons} X_3 \qquad 2X_3 \overset{2}{\rightleftharpoons} X_1 \,, \tag{45}$$

respectively an interspecific interaction between species $X_1$ and $X_2$ and an intraspecific interaction between two identical molecules of species $X_3$. In this simple example, one foresees the dramatic consequences for the geometrical picture discussed in Sec. II. A suitable representation of Eq. (45) is a hypergraph [24] where the species as nodes are connected by hyperedges involving more than two nodes (Fig. 6). For such a hypergraph, there is no obvious way to identify a spanning tree nor to split the system into disconnected components. Although it is still possible to identify $\mathrm{Ker}\,\mathbb{S}$ and $\mathrm{Im}\,\mathbb{S}^\top$ algebraically, a procedure to construct a meaningful basis for cycles and cocycles is lacking [12]. In addition, the stoichiometric matrix $\mathbb{S}$ is no longer an incidence matrix, and in full generality it will exhibit several conservation laws, i.e. left nullvectors [10]. As a consequence, the identification of the independent degrees of freedom among the pool of $N$ species is non-trivial [58], and the result of chemostatting highly depends on the chemostatting procedure.

In this part of the paper, we study a generalized notion of cycles and cocycles for generic interacting CRNs, i.e. in the absence of the graph-theoretical picture. We prove that, using the reduced row echelon form of matrices provided by linear algebra, a cycle and cocycle basis can be built which mirrors and generalizes many of the aspects discussed in Secs. II A-II B. In particular, it allows one to directly extend the force and current decompositions, Eqs. (19) and (26), to arbitrarily complex CRNs. We also introduce an integration operator which generalizes the one in Eq. (23) – without relying on a notion of spanning tree – and which allows for the integration of conservative affinities along hypergraphs. A geometrical interpretation of cocycles in hypergraphs is discussed. Finally, in the spirit of Ref. [59], these results are applied in the linear-response regime to unveil hidden symmetries in the response of complex CRNs.

### B. Construction of cycles and cocycles based on row reduction

Let us consider a generic interacting CRN whose topology is encoded in the stoichiometric matrix $\mathbb{S}$ with rank $M$. Algebraically, the rank of $\mathbb{S}$ quantifies the number of independent species and independent reactions in the network, which are the same by rank-nullity theorem. Accordingly, one can choose $M$ out of the $R$ reactions to be independent and, following the same convention as in Sec. II B, reorder the columns

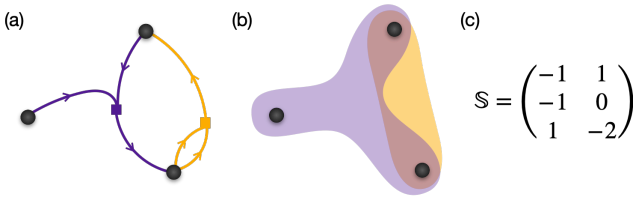



FIG. 6. (a) Hypergraph representation of the kinetic scheme Eq. (45) in the main text. The hypergraph is composed of $N = 3$ nodes corresponding to the species $\{X_1, X_2, X_3\}$ and two hyperedges corresponding to reaction 1 (purple) and reaction 2 (orange). The hyperedges differ from simple edges as they present branches which can connect to different nodes (purple) or to the same node (orange). One sees how the geometrical representation of cycles and cocycles breaks down in this case. Although a loop-resembling path can be graphically identified, it does not correspond to a cycle. At the same time the removal of the orange hyperedge does not result in a splitting of the hypergraph into disconnected components. (b) The first reaction is non-pairwise as it involves all three species, while the second reaction is still pairwise-like but with a stoichiometric coefficient $\nu_3^{+2} \neq 1$. (c) As a consequence, the stoichiometric matrix $\mathbb{S}$ is no longer an incidence matrix.

of $\mathbb{S}$ in a way such that they are placed first. Their label index is then $1 \leq \gamma \leq M$. These are the independent reactions which, in the case of noninteracting networks, constituted the cochords defining the spanning tree $T_{\mathcal{G}}$ (see Sec. II A). The remaining $\alpha$-labeled reactions are in number $R - M = \dim \mathrm{Ker}\,\mathbb{S}$, so that $M + 1 \leq \alpha \leq R$. Contrary to noninteracting networks, the conservation laws are generally more than one, being in number $N - M = \dim \mathrm{Ker}\,\mathbb{S}^\top \geq 0$. By Eq. (6), each of them is associated to a physical quantity which is conserved.

We now show how the algebraic row reduction of $\mathbb{S}$ allows one to identify: (*i*) a choice of $N - M$ conserved quantities and (*ii*) a generalization of the cycle and cocycle bases for the set of interacting reactions described by the (non-incidence) matrix $\mathbb{S}$. Using for instance the Gauss–Jordan elimination, a standard procedure in linear algebra, the stoichiometric matrix $\mathbb{S}$ is reduced to

$$-\mathbb{G}\,\mathbb{S} = \begin{pmatrix} \mathbb{0} & \mathbb{0} \\ \mathbb{1}_M & \mathbb{T} \end{pmatrix} . \tag{46}$$

Here the $N \times N$ matrix $\mathbb{G}$ is invertible and encodes the elementary operations performing the Gauss–Jordan elimination (see also Appendix A for another – explicit – construction of $\mathbb{G}$). Upon a permutation of rows, one recognizes in the r.h.s. of Eq. (46) the canonical reduced row echelon form [60], where the $M$ pivot elements constitute the bottom-left identity matrix $\mathbb{1}_M$ [7]. We adopt the convention of Eq. (46) instead of the canonical one for consistency with the geometrical analysis

[7] In the reduced row echelon form, the left part need not always be an identity matrix. The pivots, i.e. the columns containing a leading one and zeroes in all the other entries are generally scattered in the matrix. In our case, since we placed $M$ independent columns first in $\mathbb{S}$ (by permutation), the reduced row echelon form presents an identity matrix as in Eq. (46). In fact, row reduction provides another way of permuting the columns of $\mathbb{S}$: if $\mathbb{S}$ is not already having its $M$ first columns independent, the row echelon form provides $M$ pivots, their position being that of $M$ independent columns [60], which can then be placed first.

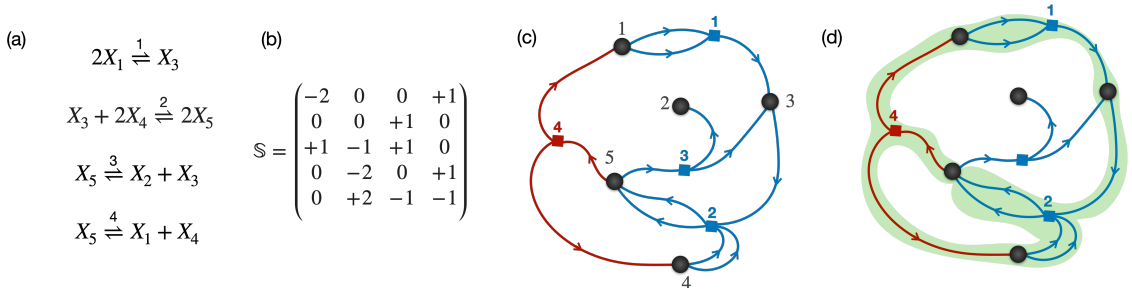



FIG. 7. (a) Example of an interacting CRN involving four reactions and five species $\{X_1, X_2, X_3, X_4, X_5\}$. (b) The corresponding stoichiometric matrix $\mathbb{S}$ is not longer an incidence matrix: its columns contain more than two entries and the values of the stoichiometric coefficients are in general different from $\pm 1$. As reported in the main text the rank $\mathbb{S} = 3$ and the matrix has two left nullvectors and one cycles, respectively $\boldsymbol{\ell}_1 = (1, 2, 1, 2, 0)$, $\boldsymbol{\ell}_2 = (0, 0, 1, 1, 1)$ and $\boldsymbol{c} = (1/2, 1/2, 0, 1)^\top$. (c) The hypergraphical representation of the reaction scheme in terms of nodes (species) and hyperedges (reactions). Following the convention described in the main text the independent reactions (cochords in blue) are numbered first and the dependent reaction (chord in red) associated to the cycle $\boldsymbol{c}$ is numbered last. (d) The reactions (1, 2 and 4) involved in the cycle $\boldsymbol{c}$ are highlighted in green. Upon performing each reaction $\rho$ a (fractional) number of time $c_\rho$, the state of the system is left unchanged.

introduced for noninteracting CRNs, as detailed below. The first $N - M \geq 0$ rows filled with zeroes reflect the fact that $\mathbb{S}$ is not necessarily full row-rank, due to the possible existence of conservation laws. We stress that the reduced row echelon form, hence the matrix $\mathbb{T}$ in Eq. (46), is unique; it does not depend on the specific form of $\mathbb{G}$ (which is not unique). Applying the matrix $\mathbb{G}$ to Eq. (6) one obtains:

$$\frac{d}{dt}(\mathbb{G}\,\boldsymbol{x}(t))_i = 0\,, \quad \text{for } \; 0 \leq i \leq N - M\,. \tag{47}$$

Thus, the first $N - M$ elements of $\mathbb{G}\boldsymbol{x}$ are (a choice of) conserved quantities for the evolution of concentrations.

The first $M$ columns are independent by our convention. The remaining $R - M$ columns of $\mathbb{S}$ can thus be obtained as linear combinations of the independent ones. This is precisely what the matrix $\mathbb{T}$ in Eq. (46) encodes: each dependent column $\mathbb{S}_\alpha$ is given by $\mathbb{S}_\alpha = \sum_\gamma \mathbb{S}_\gamma \mathbb{T}_{\alpha\gamma}$. Notably, we can use the notion of independence/dependence among reactions to restore the terms chords and cochords even in the absence of a spanning tree. In particular, we name chords (resp. cochords) the set of dependent (resp. independent) reactions.

Let us now use the invertibility of $\mathbb{G}$ and take the transpose of Eq. (46), so that:

$$\mathbb{S}^\top = -\begin{pmatrix} \mathbb{0} & \mathbb{1}_M \\ \mathbb{0} & \mathbb{T}^\top \end{pmatrix} (\mathbb{G}^{-1})^\top\,. \tag{48}$$

Since $\mathbb{G}$ is full-rank, the image of $\mathbb{S}^\top$ is spanned by the $M$ non-zero columns of the reduced row echelon form. This explains the choice of the same notation $\mathbb{T}$ as for the matrix in Eq. (14) which was used to construct the cycles and cocycles for simple graphs. In that case, the matrix $\mathbb{T}$ was built from the spanning tree (see Sec. II A) while here it is obtained by algebraic means. As a consequence, the entries of the new matrix $\mathbb{T}$ are no longer restricted to $\{0, \pm 1\}$ as in Eq. (14), but may take fractional entries (see Appendix A). In both cases, $\mathbb{T}$ allows one to define a basis for $\mathrm{Im}\,\mathbb{S}^\top$, which in the previous case was identified as the space spanned by the cocycles $\boldsymbol{c}^\gamma$'s . Accordingly, we interpret the column vectors in $(\mathbb{1}_M \;\; \mathbb{T})^\top$ as a family of generalized cocycles $\{\boldsymbol{c}^\gamma\}$. Following in the analogy, it is natural to use matrix $\mathbb{T}$ in Eq. (46) to construct a basis for the kernel of $\mathbb{S}$. In particular, we define a family of generalized cycles $\{\boldsymbol{c}^\alpha\}$ as the column vectors in $(-\mathbb{T}^\top \;\; \mathbb{1}_{R-M})^\top$. As in Sec. II B, the rank-nullity theorem ensures that these vectors constitute a basis for $\mathrm{Ker}\,\mathbb{S}$. In fact, it is possible to show that any basis of $\mathrm{Ker}\,\mathbb{S}$ can be reduced to that form, with $\mathbb{T}$ uniquely defined by Eq. (46) (see Appendix A).

We have shown here that the row reduction of $\mathbb{S}$ allows one to identify a basis for the kernel and the coimage of $\mathbb{S}$ which we connect to the previously defined cycles and cocycles. In fact, if one uses this algebraic procedure in the case of noninteracting CRNs, one recovers the expressions for the cycles and cocycles in Eq. (13) with $\mathbb{T}$ given by Eq. (14) – provided the graph is oriented using the convention depicted in Fig. 3. In this sense, the newly defined vectors $\boldsymbol{c}^\alpha$'s and $\boldsymbol{c}^\gamma$'s are a genuine generalization of the cycles and cocycle and we thus use the same terminology to designate them. Notably, the orthogonality relations in Eq. (15) and (18) apply directly to the new sets $\{\boldsymbol{c}^\gamma\}$ and $\{\boldsymbol{c}^\alpha\}$, which opens to the possibility of interpreting them as geometrical objects on hypergraphs, as discussed in the next section. As a consequence, the decompositions of currents and affinities Eqs. (19) and (26) discussed in Sec. II C, and based on the orthogonality relations, directly apply to interacting CRNs, together with the expressions (20) and (31) of the driving affinities and of the transient currents (when a stationary state exists).

We conclude with a remark. The vectors $\boldsymbol{c}^\alpha$'s and the $\boldsymbol{c}^\gamma$'s are not the only bases of $\mathrm{Ker}\,\mathbb{S}$ and $\mathrm{Im}\,\mathbb{S}^\top$ (for instance the first $M$ columns of $\mathbb{S}$ span the coimage of $\mathbb{S}$). The interest of the definition of cycles and cocycles we put forward is that it allows for a physical decomposition of the affinity akin to the Helmholtz–Hodge decomposition, and that it can be used to build a geometrical representation of the forces and currents on hypergraphs for general CRNs, as we present now. (See also Appendix D for a definition of oblique projectors based on the $\boldsymbol{c}^\alpha$'s and $\boldsymbol{c}^\gamma$'s that generalizes to arbitrary CRNs those defined in Ref. [14] for noninteracting CRNs.)

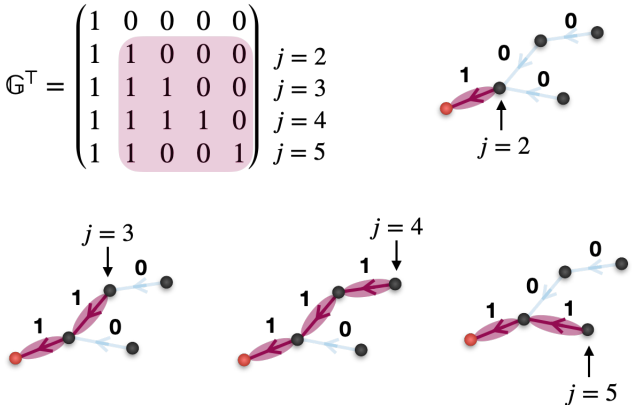



FIG. 8. One reads along the line of the bottom-right $(N-1)\times(N-1)$ block of the matrix $\mathbb{G}^\top$, defined by the transpose of Eq. (22), the integration paths along the spanning tree which are used in Eq. (23). Those correspond to the unique escape routes connecting any node $j \neq$ root to the root. Here we report $\mathbb{G}^\top$ for the example in Fig. 1 where the corresponding block is highlighted in purple. Notice that every node $j \neq$ root is the source $s(\gamma)$ of exactly one cochord in the spanning tree (see Fig. 3a). Hence, for a fixed $j \neq$ root, the sum $\sum_\gamma \mathbb{G}^\top_{js(\gamma)}$ runs over all the cochords which connect the node $j$ to the root. For the example of spanning tree in Fig. 3a we depict graphically the escape routes for the various nodes $2 \leq j \leq N = 5$.

## C. Geometry of hypergraphs

### 1. *Cycles with weights*

The geometricity of cycles and cocycles is rooted in the orthogonality relations Eqs. (15) and (18). Those are the expression of the one-to-one correspondence between cocycles and independent reactions on the one side, and cycles and dependent reactions, on the other side. In the previous section we pointed out that for interacting CRNs the entries of the cycles vectors $\boldsymbol{c}^\alpha$ may be fractional. Contrary to the case of noninteracting CRNs, cycles are decorated with weights given by the entries of the $\mathbb{T}$ matrix in Eq. (46). Intuitively, these physical weights express the (fractional) number of times that each reaction must be performed along a cycle in order to leave the state of the system invariant. For illustrative purposes, we report in Fig. 7 the example of an interacting CRN with five chemical species and four reactions. In this case the stoichiometric matrix has rank $M = 3$ and exhibits $N - M = 2$ conservation laws and a number of cycles $R - M = 1$. In this case, one obtains $\mathbb{T} = (-1/2, -1/2, 0)^\top$ and $\boldsymbol{c} = (1/2, 1/2, 0, 1)^\top$ (we drop the index $\alpha = 1$ for simplicity). One sees already for this simple example that identifying the cycle graphically is not easy due to the topological complexity of the hypergraph (Fig. 7c). Nevertheless, the cycles are an important feature of the dynamics, since they are a basis of the non-equilibrium stationary currents (when a steady state is reached). A key aspect is that they are described using the same matrix $\mathbb{T}$ as for the cocycles. As we will see, such duality between cycles and cocycles leads to interesting consequences.

### 2. *Integrating on the hypergraph*

We now ask the question of what is the geometrical meaning of the weights in the matrix $\mathbb{T}$, underlying both cycles and cocycles. First let us recall the integration matrix $\mathbb{G}^\top$ previously introduced for simple noninteracting CRNs. It was explicitly constructed by fixing a spanning tree $T_\mathcal{G}$ and one root [see Eq. (22) and Appendix A] such that the $(N-1)\times(N-1)$ bottom-right block contains the set of paths on the spanning tree along which we integrate the conservative affinities $\boldsymbol{A}$ to define the potential $\boldsymbol{V}$ [Eq. (23)]. From a purely graphical viewpoint, each path can be interpreted as the unique 'escape route' in $T_\mathcal{G}$ along which a unit 'charge' placed on a given node is expelled through the root leaving no trace along the way (see Fig. 8). Thus, following this geometrical view, we may re-express matrix $\mathbb{G}^\top$ as

$$\mathbb{G}^\top = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \vdots & \multicolumn{3}{c}{\multirow{2}{*}{escape routes}} \\ 1 & & & \end{array}\right), \tag{49}$$

where the escape routes constitute the bottom-right $(N-1)\times(N-1)$ submatrix of $\mathbb{G}^\top$. For graphs, the escape routes involve a succession of adjacent edges, irrespective of the connectivity of each node, which is reflected in the entries of $\mathbb{G}^\top$ being 0 or 1. This is no longer the case in hypergraphs due to the presence of branching in the hyperedges. Notice that in Eq. (49) the root is naturally associated to the conservation law $\boldsymbol{\ell}_0$ which appears as the first column of $\mathbb{G}^\top$. Thus for the case of interacting CRNs it is natural to generalize the structure in Eq. (49) by picking a root for each of the (now possibly multiple) conservation laws of $\mathbb{S}$. Doing so, one obtain a set of roots, each one associated to a given conservation law, and we may ask what are the corresponding escape routes on the hypergraph, i.e. the 'hyper-paths' along which a charge placed on any node is expelled through the roots leaving no trace. In the absence of a spanning tree we lack a graphical procedure to find such escape routes; nevertheless in Appendix A we show that, given a suited set of roots [8], the escape routes can be obtained algebraically and are uniquely defined. By construction, they involve the $M$ independent reactions (the cochords) in analogy to the escape routes defined from the spanning tree in simple graphs.

Accordingly, we introduce a generalized matrix $\mathbb{G}^\top$ such

[8] In the case of a simple noninteracting CRN, the choice of the root is fully arbitrary since all the $N$ species in the system are equally constrained by the left nullvector $\boldsymbol{\ell}_0 \in \mathrm{Ker}\,\mathbb{S}^\top$. It is no longer the case for an interacting CRNs for which the conservation laws are shaped by the interactions and will typically involve subsets of species. Then, given a conservation law one chooses a root among the subset of species that are constrained by that conservation law. The procedure is repeated for every conservation law. One species cannot be picked twice as a root.

$$\mathbb{G}^\top = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & -1 & 0 & 0 \\ 1 & 1 & -1/2 & 1/2 & 1 \\ 2 & 1 & -1 & 0 & 1 \end{pmatrix} \begin{matrix} \\ \\ j=3 \\ j=4 \\ j=5 \end{matrix}$$

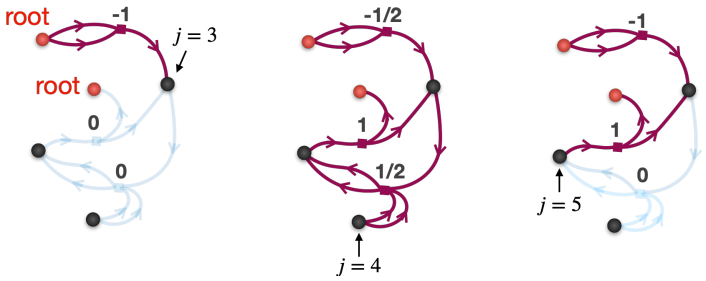



FIG. 9. The escape routes are read as lines of the bottom-right ($M \times M$) block in the $\mathbb{G}^\top$ matrix. For interacting CRNs, there is no simple graphical procedure to fill the entries of $\mathbb{G}^\top$. Nevertheless, an algebraic algorithm to identify the entries of $\mathbb{G}^\top$ is described in Appendix A. Here we report the $\mathbb{G}^\top$ for the example of Fig. 7. The numbering of the nodes follows the convention discussed in the main text: the two roots are labelled as node 1 and 2 and the rest of the nodes are numbered last (see Fig. 7). In this case the rank $\mathbb{S} = M = 3$ hence we highlight in purple the bottom-right ($3 \times 3$) block. We also represent the escape routes graphically together with the entries of the corresponding line. By construction, the escape routes are constrained to live on the independent reactions (cochords); for this reason we have removed from the hypergraph reaction 4 which is the reaction associated to the cycle $\boldsymbol{c}$ (the red chord in Fig. 7).

that

$$\mathbb{G}^\top = \left( \begin{array}{c|c} \text{csv} & \begin{matrix} 0 \cdots\cdots 0 \\ 0 \cdots\cdots 0 \end{matrix} \\ \hline \text{laws} & \begin{matrix} \\ \text{escape} \\ \text{routes} \end{matrix} \end{array} \right) \begin{matrix} \updownarrow N-M \\ \updownarrow M \end{matrix} \qquad (50)$$
$$\mathbb{G}^\top_M$$

where the conservation laws, spanning $\mathrm{Ker}\,\mathbb{S}^\top$, constitute the first $N - M$ columns, the top-right block is padded with zeroes and the $M \times M$ bottom-right square matrix contain the escape routes from each node $\notin$ {roots}. In Fig. 9 we represent such escape routes for the example given in Fig. 7. As compared with the case of simple graphs, each escape route is now a 'multi-path', i.e. a combination of the hyperedges, and is decorated with weights which tell how many times each cochord reaction is applied for the unit charge to vanish through the roots. As such, they constitute the generalization to hypergraphs of the simple escape routes identified from the spanning tree in graphs. Notably, the matrix in Eq. (50) realizes the row reduction of Eq. (46); see Appendix A for a proof. It corresponds to a geometrically-informed choice for the matrix of the row-reduction (hence the choice of notation).

Let us now consider a generic conservative force $\boldsymbol{F} \in \mathrm{Im}\,\mathbb{S}^\top$, so that

$$\boldsymbol{F} = \sum_\gamma F_\gamma \boldsymbol{c}^\gamma = -\mathbb{S}^\top \boldsymbol{V} \qquad (51)$$

where we have expressed it as (minus) the discrete gradient operator $\mathbb{S}^\top$ acting on a potential $\boldsymbol{V} \in \mathbf{R}^N$ defined on the nodes. By construction, we can invert Eq. (51) using the matrix $\mathbb{G}^\top$. Physically it means that the matrix $\mathbb{G}^\top$ allows one to integrate any conservative force $\boldsymbol{F}$ on the hypergraph in order to find the corresponding potential landscape $\boldsymbol{V}$. In particular, denoting $\mathbb{G}^\top_M$ the $N \times M$ right-block obtained by excluding the conservation laws from $\mathbb{G}^\top$ [see Eq. (50)] one finds:

$$\boldsymbol{V} = \mathbb{G}^\top_M \left( F_\gamma \right) \updownarrow \mathbb{R} \qquad (52)$$

Notice that first $N - M$ rows of $\mathbb{G}^\top_M$ are padded with zeroes, which corresponds to fix the potential $V_{\text{root}} = 0 \ \forall$ root. As we discussed in Sec. II D if $F_\gamma = A^c_\gamma$ [the cocycles affinities in Eq. (19)], the integration procedure in Eq. (52) yields the chemical potential of thermodynamics $\boldsymbol{V} = \boldsymbol{\mu}$. Alternatively, if $F_\gamma = \log(k^+_\gamma / k^-_\gamma)$, the integration in Eq. (52) leads to the standard chemical potential $\boldsymbol{V} = \boldsymbol{\mu}^\ominus$. Those equalities of potentials hold up to linear combinations of the conservation laws which can always be added to $\boldsymbol{V}$ leaving $\boldsymbol{F}$ unchanged.

### 3. *Cocycles on the hypergraph*

In Sec. II A we identified the cocycle $\boldsymbol{c}^\gamma$ with the boundary of a source island $\boldsymbol{v}^\gamma$ on the graph $\mathcal{G}$. Upon removing the cocycle from $\mathcal{G}$, the source island is fully disconnected from the target island, containing the root.

The binary splitting of $\mathcal{G}$ associated to every cocycle no longer holds for interacting CRNs: the removal of a cocycle from the hypergraph does not necessarily split it into disconnected components (see for instance Fig. 10). It is therefore legitimate to ask whether a geometrical interpretation of cocycles exists at all for the hypergraph.

We have already seen how the matrix $\mathbb{G}^\top_M$ allows one to transform a set of conservative forces defined on the cochords into a set of potentials defined on the nodes (with fixed potential $V_{\text{root}} = 0$ for the arbitrarily chosen roots). Also, it directly relates to the family of cocycles since, from Eq. (46), we have $-(\mathbb{S}^\top \mathbb{G}^\top_M)_\gamma = \boldsymbol{c}^\gamma$. Accordingly, we can introduce the potentials $\boldsymbol{v}^\gamma = (\mathbb{G}^\top_M)_\gamma$ such that $\boldsymbol{c}^\gamma = -\mathbb{S}^\top \boldsymbol{v}^\gamma$. Namely, each potential $\boldsymbol{v}^\gamma$, upon differentiation with $\mathbb{S}^\top$, generates a force that is the cocycle $\boldsymbol{c}^\gamma$. In the case of noninteracting CRNs the source of the cocycle $\boldsymbol{c}^\gamma$ was easily identified as the collection of nodes $i$ such that $v^\gamma_i = 1$. By analogy, for interacting CRNs, we define the source of $\boldsymbol{c}^\gamma$ from the set of nodes $i$ such that

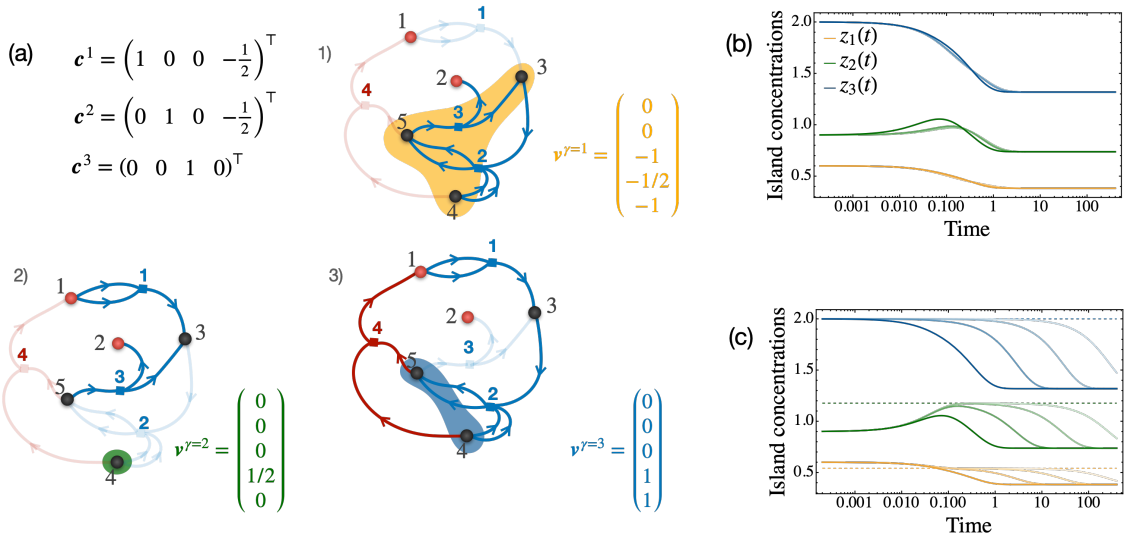



FIG. 10. **(a)** We report the cocycles $\boldsymbol{c}^\gamma$'s, with $1 \leq \gamma \leq 3$, for the example of interacting CRN introduced in Fig. 7. Removing a cocycle results in the emergence of a new conservation law $\boldsymbol{v}^\gamma$ which can be read from the columns of $\mathbb{G}_M^\top$. We interpret the nodes belonging to the new conservation law as the source island of the corresponding cocycle, where each node has a certain weight (altitude). For instance, by removing the cocycle $\boldsymbol{c}^1$, one identifies from $\boldsymbol{v}^{\gamma=1} = (\mathbb{G}_M^\top)_{\gamma=1}$ a source island (orange) containing nodes 3, 4 and 5 with different (negative) weights. The conservation laws are reported for every cocycle, $\boldsymbol{c}^1$ to $\boldsymbol{c}^3$. We stress that, in contrast to the case of simple graphs, the source islands are generally not disconnected from the roots, as in the case of the orange and blue islands. Let us focus on cocycle $\boldsymbol{c}^3$, which consists of the sole reaction 3. We compare the dynamical relaxation of the island concentrations $z_\gamma = \boldsymbol{v}^\gamma \cdot \boldsymbol{x}$ upon suppressing (to various degrees) reaction 2 **(b)** and reaction 3 **(c)**. Eq. (6) is solved numerically with initial condition $x_i(0) = 1 \; \forall i$. The rates are chosen in accordance to Wegscheider criterion (31) with $k_\rho^\pm \sim O(1) \; \forall \rho$. In **(c)** the rates of reaction 3 are suppressed by a factor $\varepsilon = 10^{-1}, 10^{-2}, 10^{-3}$ (different shadows) such that $k^\pm_{\rho=3} \mapsto \varepsilon k^\pm_{\rho=3}$ and $k^+_\rho / k^-_{\rho=3} = \text{const}$. The dashed lines correspond to the equilibrium steady-state value obtained for $\varepsilon = 0$, i.e. full suppression of reaction 3. As anticipated, in this case $z_3$ becomes a constant of motion.

$v_i^\gamma \neq 0$. Because the entries of $\boldsymbol{v}^\gamma$ are no longer restricted to zeroes and ones, the corresponding island on the hypergraph now has a geography. Namely, each node is given an altitude which quantifies the impact of the corresponding node onto the outward cocycle current $J^e_\gamma$ [Eq. (27)]. Assume now that one is able to cancel the latter current (through a protocol; see next paragraph for an example), i.e. to effectively 'remove' the cocycle $\boldsymbol{c}^\gamma$ from the hypergraph. Then one gets:

$$0 = J^e_\gamma = \boldsymbol{c}^\gamma \cdot \boldsymbol{J} = -\boldsymbol{v}^\gamma \cdot \mathbb{S}\boldsymbol{J} = -\partial_t \left(\boldsymbol{v}^\gamma \cdot \boldsymbol{x}\right) , \tag{53}$$

where we used the definition of $\boldsymbol{v}^\gamma$ and the evolution Eq. (6). Accordingly, the vector $\boldsymbol{v}^\gamma$ can be seen as a new conservation law which emerges when removing $\boldsymbol{c}^\gamma$, the corresponding conserved quantity being the (weighted) concentration of the source island $z_\gamma = \boldsymbol{v}^\gamma \cdot \boldsymbol{x}(t) = \text{const}$. Notice that the weights contained in $\boldsymbol{v}^\gamma$ can be negative, which reflects the fact that, at the level of the full hypergraph (see Fig. 10), the source island is still connected to the target island (no new component necessarily arises). This means that, microscopically, exchanges of molecules between the source and the target islands are still possible after the removal of the cocycle. For this reason, it is not possible to interpret the cocycle as the geometrical boundary of the source island on the hypergraph; nonetheless, from an algebraic viewpoint, the consequence of the removal of a cocycle is still the emergence of a new conservation laws in the dynamics.

### 4. *Evolution of islands*

Finally, let us connect the geometrical pictures of the islands, identified by the columns of $\mathbb{G}_M^\top$, with dynamics. Applying the $\mathbb{G}$ matrix to Eq. (6) one finds that each island concentration $z_\gamma$ evolves by the corresponding cocycle current $J^e_\gamma$ [Eq. (26)] such that:

$$\frac{d}{dt} z_\gamma(t) = J^e_\gamma(t) \quad \forall \gamma \; . \tag{54}$$

Equation (54) can be seen as an integrated continuity equation, where $z_\gamma(t) = (\mathbb{G}_M \, \boldsymbol{x}(t))_\gamma$ is the sum of the concentrations of the nodes of the island, weighted by the components of $\boldsymbol{v}^\gamma$ (which can be seen as the elevation map of the island). Islands thus constitute a geometrical coarse-graining of the $N$ species into $M$ independent degrees of freedom $z_\gamma$. In this sense the variables $z_\gamma$'s are macroscopic (integrated) and describe the relaxation of the system to its steady state (since $J^e_\gamma(t \to \infty) = 0 \; \forall \gamma$, when a steady state is reached).

We end this section by putting forward a possible application of this formalism to control, in the chemical setting. In chemistry, molecular inhibitors are often employed to delay, slow or prevent chemical reactions. Like in an inverse catalysis, the inhibitor acts by suppressing the reaction rate constants $k^\pm_\rho \to 0$ of a target reaction $\rho$, thus introducing a slow timescale at the kinetic level. In complex CRNs it is not clear *a priori* what is the effect of suppressing a reaction on (*i*) the macroscopic relaxation timescale and (*ii*) the steady-state

concentrations reached in the long time limit. It depends on a number of factors including the initial condition, the distribution of reaction rate constants and the topology of the network. Nevertheless, some remarks follow directly from our algebraic approach. In particular, we have identified the cocycles as the relaxation modes of the dynamics, whose removal leads to the emergence of new conservation laws (left zero modes of $\mathbb{S}$). As such, we expect them to have a strong impact on the timescales of relaxation. In Fig. 10 we show the finite time relaxation of the island concentrations $z_\gamma$'s [Eq. (54)] for the example in Fig. 7 when a cocycle reaction is suppressed, compared to the case when a non-cocycle reaction is suppressed. As anticipated, the behavior strongly differs. The removal of a reaction that is not a cocycle affects minimally the finite-time dynamics, leaving unchanged the characteristic relaxing time and the equilibrium steady state. On the contrary, upon decreasing the reaction rates of a cocycle, the dynamics develops a *plateau* which corresponds to a new timescale controlled by the inhibitor. In the limit of complete suppression of the cocycle the system relaxes to a new equilibrium state, which is a sign of the emergence of a new conservation law.

### D. Connection between stochastic dynamics and thermodynamics

In this part, we explain how the underlying stochastic dynamics allows one to make the connection to the thermodynamic properties of CRNs. We start by closed systems (that follow a reversible stochastic dynamics) and consider then the case of open systems, driven out of equilibrium by chemostatting reactions (that necessarily obey local detailed balance, as we show constructively).

#### 1. *Closed systems*

We assume first that the system is closed and has at least one mass-like conservation law. Interestingly, this is related to Gordan's theorem (see e.g. [23, 61]), which states that the two following conditions are mutually exclusive: (*i*) There exists a left nullvector $\boldsymbol{\ell}$ of $\mathbb{S}$, with all components positive ($\ell_i \geq 0$). Such vector $\boldsymbol{\ell}$ is a mass-like conservation law. (*ii*) There exists a vector $\boldsymbol{j}$ such that $\mathbb{S}\boldsymbol{j} > 0$. We exclude case (*ii*) because, from the rate equation (6), it implies that the set of reactions can create matter, which cannot occur for a closed system – this would contradict Lavoisier's principle [62]. (Notice that chemostatted autocatalytic reactions fall in case (*ii*) [23].)

We then assume that the CRN is 'in equilibrium', from the thermodynamics or from the stochastic viewpoints, which are equivalent. Namely, from the thermodynamics viewpoint, we ask that the reaction rate constants $k_\rho^\pm$'s satisfy Wegscheider's condition, Eq. (31), which is equivalent to the existence of a standard chemical potential $\boldsymbol{\mu}^\ominus$ such that

$$\frac{k_\rho^+}{k_\rho^-} = \exp\left[-(\mathbb{S}^\top \boldsymbol{\mu}^\ominus)_\rho\right]. \tag{55}$$

[In general $\mathbb{S}$ is not an incidence matrix, so that this condition is more involved than in the noninteracting case of Eq. (33)]. In turn, this is equivalent to requiring that the affinity $\boldsymbol{A}$ is conservative, i.e., $\boldsymbol{A} \in \mathrm{Im}\,\mathbb{S}^\top$ (the proofs of every equivalence discussed here are gathered in Appendix B). For graphs, this is KVL, and we will keep the same terminology for generic CRNs. In chemistry, this is called thermodynamic feasibility, and in fact Eq. (55) is rather natural if one pictures the chemical reaction $\rho$ as a transition between molecular conformations in the landscape of possible chemical combinations of atoms: the r.h.s. of Eq. (55) is the ratio of Kramers transition rates in such landscape. This brings us to the stochastic side: the thermodynamic conditions that we just discussed are equivalent to requiring the reversibility of the underlying stochastic process. Considering the dynamics of the population $\boldsymbol{n}$ of species, it means that every transition associated to a reaction $\rho$ in Eq. (3) verifies the detailed balance condition

$$\frac{W_\rho^+(\boldsymbol{n})}{W_\rho^-(\boldsymbol{n}+\mathbb{S}_\rho)} = \frac{\mathcal{P}^{\mathrm{eq}}(\boldsymbol{n}+\mathbb{S}_\rho)}{\mathcal{P}^{\mathrm{eq}}(\boldsymbol{n})} \quad \forall \boldsymbol{n}\,. \tag{56}$$

Here $W_\rho^\pm(\boldsymbol{n}) = W(\{n_i \mapsto n_i \pm \mathbb{S}_{i\rho}\})$ denotes the transition rate at population level, and the equilibrium distribution $\mathcal{P}^{\mathrm{eq}}(\boldsymbol{n})$ is a product-form Poisson-like law (constrained by the conservation laws) of parameters $\boldsymbol{x}^{\mathrm{eq}}$ (see Appendix B for explicit expressions). The vector $\boldsymbol{x}^{\mathrm{eq}}$ represents the average value of the species concentrations in the long-time limit and depends on the initial condition. As detailed in Appendix B, the affinity $\boldsymbol{A}$ cancels when evaluated in $\boldsymbol{x}^{\mathrm{eq}}$ so that defining $x_i^{\mathrm{eq}} = \exp[\mu_i^{\mathrm{eq}} - \mu_i^\ominus]$ and using the definition of affinities in Eq. (7), we obtain that the corresponding 'equilibrium chemical potential' $\boldsymbol{\mu}^{\mathrm{eq}}$ belongs to the left nullspace of $\mathbb{S}$, which means that $\boldsymbol{\mu}^{\mathrm{eq}}$ is a linear combination of the conservation laws – this is the reason why we assumed above that at least one exists (see also Ref. [63] for insights on the role of conservation laws).

Introducing the quasipotential $\Phi(\boldsymbol{n}) = -\frac{1}{\Omega}\log \mathcal{P}^{\mathrm{eq}}(\boldsymbol{n})$ associated to the equilibrium law, Eq. (56) rewrites

$$\frac{W_\rho^+(\boldsymbol{n})}{W_\rho^-(\boldsymbol{n}+\mathbb{S}_\rho)} = \exp\Big\{-\Omega\left[\Phi(\boldsymbol{n}+\mathbb{S}_\rho) - \Phi(\boldsymbol{n})\right]\Big\}. \tag{57}$$

At large volume $\Omega \gg 1$ with fixed $\boldsymbol{x} = \boldsymbol{n}/\Omega$, one finds that $\Phi(\boldsymbol{n}) \to \phi(\boldsymbol{x})$ with

$$\phi(\boldsymbol{x}) = \sum_i \left(x_i \log x_i - x_i - x_i \log x_i^{\mathrm{eq}} - x_i\right). \tag{58}$$

Here one recognizes that $\phi(\boldsymbol{x})$ is the free energy density, that is, the difference between the energy density and the entropy density. Notice that $\phi(\boldsymbol{x})$ is minimum (and cancels) at $\boldsymbol{x} = \boldsymbol{x}^{\mathrm{eq}}$. Defining (for any $\boldsymbol{n}$ and $\boldsymbol{x} = \boldsymbol{n}/\Omega$) $x_i = \exp(\mu_i - \mu_i^\ominus)$, one can expand Eq. (57) for $\Omega \gg 1$, which yields the expression of the entropy production $\Sigma_\rho$ of reaction $\rho$:

$$\Sigma_\rho(\boldsymbol{x}) = \log \frac{W_\rho^+(\boldsymbol{n})}{W_\rho^-(\boldsymbol{n}+\mathbb{S}_\rho)} = -(\mathbb{S}^\top \boldsymbol{\nabla}\phi(\boldsymbol{x}))_\rho \tag{59}$$

$$(\boldsymbol{\nabla}\phi(\boldsymbol{x}))_i = \log x_i - \mu_i^{\mathrm{eq}} + \mu_i^\ominus = \mu_i - \mu_i^{\mathrm{eq}}\,. \tag{60}$$

We keep here the usual denomination of entropy production for this ratio of rates at population level, but we remark already that it plays the same role as the affinity at the deterministic level of reaction rate constants [see Eq. (7)]. We observe that the constant $\mu_i^{\rm eq}$ in Eq. (60) plays no role in Eq. (59) since $\boldsymbol{\mu}^{\rm eq} \in \mathrm{Ker}\,\mathbb{S}^\top$ [9]. We can thus write for the vector $\boldsymbol{\Sigma}$ of the $\Sigma_\rho$'s:

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = -\mathbb{S}^\top \boldsymbol{\mu}(\boldsymbol{x})\,. \tag{61}$$

Here $\boldsymbol{\Sigma}(\boldsymbol{x})$ comes from the stochastic dynamics [the l.h.s. of Eq. (56)] while the associated chemical potential $\boldsymbol{\mu}(\boldsymbol{x})$ comes from the equilibrium distribution [the r.h.s. of Eq. (56)], and is thus of thermodynamic nature. We stress that Eq. (61) can be read for any occupation state of the system. If evaluated for $\boldsymbol{x}(t)$ solution of the rate equation (6), we see that $\boldsymbol{\Sigma}(\boldsymbol{x}(t))$ becomes equal to the affinity $\boldsymbol{A}(\boldsymbol{x}(t))$ and goes to zero at large time, as expected in equilibrium.

### 2. *Open systems*

To drive the CRN out of equilibrium, one adds chemostatting reactions of the form of Eq. (35) and splits species in a set $Y$ of chemostatted species and a set $X$ of non-chemostatted ones, as was done in Sec. II E in the noninteracting case. As well known (see e.g. Refs. [10, 64]), each chemostatted species either breaks a conservation law or induces an emergent cycle. The number of emergent cycles can be found algebraically by considering the extended stoichiometric matrix $\mathbb{S}_{\rm res}$ [defined as in Eq. (37)], and is given by Eq. (44).

The rates $\gamma_i^\pm$ of the chemostatting reaction of species $Y_i$ allow one to define a driving parameter $a_i$ through Eq. (36). Interestingly, in such chemostatting settings, one sees explicitly that the underlying stochastic dynamics satisfies 'local detailed balance' [65, 66] (see Ref. [67] for a review and Refs. [63, 68] for the case of CRNs). Namely, defining the vector of the entropy production rates of individual reactions as in Eq. (59) (including chemostatting reactions), one now finds at large $\Omega$

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = -\mathbb{S}_{\rm res}^\top \boldsymbol{\mu}(\boldsymbol{x}) + \boldsymbol{a}\,, \tag{62}$$

where the computation is done directly from the expression of the ratio of transition rates. Eq. (62) expresses that local detailed balance holds, with $\boldsymbol{a}$ playing the role of a chemical drive. As before for Eq. (61), this equation holds for any occupation $\boldsymbol{n}$ (at large $\Omega$) through $\boldsymbol{x} = \boldsymbol{n}/\Omega$; it makes the link between stochastic aspects (on the l.h.s.) and thermodynamic quantities (the r.h.s. being expressed as a function of chemical potential and drive). The case $\boldsymbol{a} = \boldsymbol{0}$ is the simplest case where equilibrium is recovered, since no chemostatting occurs then. In fact, the condition for the chemostatting not to drive the system out of equilibrium is $\boldsymbol{a} \in \mathrm{Im}\,\mathbb{S}_{\rm res}^\top$ – in which case the emergent cycles induce no current. This is expected, since the affinity $\boldsymbol{A}$ takes exactly the form of Eq. (61) [see Eq. (40), which also holds for interacting CRNs]. We refer the reader to Refs. [68–70] for relations to the second law.

[9] In fact, if one evaluates Eq. (57) starting from the l.h.s. instead of the r.hs. of Eq. (56), one arrives directly at Eq. (61).

## E. Linear response for CRNs

We now apply our framework to study the response of interacting networks to small perturbations out of equilibrium stationary states. As discussed in Sec. II D, the Wegscheider condition for closed CRNs ensures the existence of an equilibrium steady state $\boldsymbol{x}^{\rm eq}$, fixed by the initial conditions, and characterized by the vanishing of all the currents and affinities, $J_\rho^{\rm eq} = A_\rho^{\rm eq} = 0\ \forall \rho$. Close to this equilibrium state, we can linearize Eq. (8) and re-express it in matrix form:

$$\boldsymbol{J} \simeq \mathbb{\Lambda} \boldsymbol{A}\,, \tag{63}$$

where $\mathbb{\Lambda}$ is the $R \times R$ diagonal matrix of linear mobility defined by the diagonal entries $(\mathbb{\Lambda})_{\rho\rho} = \lambda_\rho^+(\boldsymbol{x}^{\rm eq}) = \lambda_\rho^-(\boldsymbol{x}^{\rm eq}) = \lambda_\rho^{\rm eq}$. Despite its familiar form of a linear phenomenological relation [47], Eq. (63) is little informative on the system's response. It describes the local response of each current $J_\rho$ to a small perturbation of the corresponding affinity $A_\rho$ without taking into account the cross-couplings between chemical reactions. In this sense, the Onsager reciprocal relations [71] are trivially satisfied by Eq. (63) being $\mathbb{\Lambda}$ a diagonal matrix. Furthermore, Eq. (63) is blind to the underlying network topology: we know that only $M$ out of the $R$ reactions (and the corresponding currents) in the network are linearly independent due to cycles. A natural question is thus how cross-couplings among reactions emerge in this context and how they relate to the network structure. The decompositions introduced in Sec. II C will provide a natural framework to address these aspects.

Eq. (63) is valid whenever the affinity is small, but one may further assume that the system is (*i*) closed, with affinity $\boldsymbol{A} \ll 1$ remaining small and conservative while relaxing to zero; (*ii*) open, with an external source (e.g. a chemostat) providing a constant non-conservative contribution to the total affinity $\boldsymbol{A} \ll 1$. In the first case, the system exhibits a transient relaxation towards $\boldsymbol{x}^{\rm eq}$. In the second case, the system reaches a nonequilibrium steady state $\boldsymbol{x}^*$, close to $\boldsymbol{x}^{\rm eq}$, with positive entropy production. We shall treat these two cases separately before revealing the connections between them, in the spirit of the Einstein relation between diffusivity and mobility.

### 1. *Transient response*

In the previous section, we have seen how the finite-time relaxation of the system is fully captured by the $M$ cocycle currents $J_\gamma^e$ in Eq. (54) (and this even outside the linear-response regime, when a steady state is reached). By substituting the decomposition Eq. (26) in the rate equation Eq. (6), one directly sees that the currents $J_\alpha^c$ do not contribute to the time evolution of $\boldsymbol{x}(t)$, since $\boldsymbol{c}^\alpha \in \mathrm{Ker}\,\mathbb{S}$. Accordingly, we can plug Eq. (63) into the definition of $J_\gamma^e$ and get

$$J_\gamma^e = \boldsymbol{c}^\gamma \cdot \boldsymbol{J} = \boldsymbol{c}^\gamma \cdot \mathbb{\Lambda}\boldsymbol{A} = \sum_{\gamma'} \underbrace{\boldsymbol{c}^{\gamma\top} \mathbb{\Lambda}\, \boldsymbol{c}^{\gamma'}}_{(\mathbb{L}_Q)_{\gamma\gamma'}} A_{\gamma'}^c\,, \tag{64}$$

where in the last step we used the decomposition in Eq. (19) together with the condition of reversibility Eq. (21). Equa-

tion (64) describes the linear relation between transient currents $J^e_\gamma$ and conservative affinities $A^c_\gamma$. They vanish together in the long time limit, as $\boldsymbol{x}(t) \to \boldsymbol{x}^{\rm eq}$. Accordingly, we identify the matrix $\mathbb{L}_Q$ in Eq. (64) as an $M \times M$ relaxation matrix. It is symmetric and positive-defined in accordance with Onsager reciprocal relations.

Since we are assuming a small perturbation around equilibrium, we introduce the distances from equilibrium for the concentration $\boldsymbol{x}(t)$ and the chemical potential $\boldsymbol{\mu}(t)$ as:

$$\delta\boldsymbol{x}(t) \;=\; \boldsymbol{x}(t) - \boldsymbol{x}^{\rm eq} \tag{65}$$

$$\delta\boldsymbol{\mu}(t) \;=\; \boldsymbol{\mu}(t) - \boldsymbol{\mu}^{\rm eq}\,, \tag{66}$$

so that $\delta\boldsymbol{x}\,,\ \delta\boldsymbol{\mu} \xrightarrow{t\to\infty} 0$. The affinity is assumed to be conservative, $\boldsymbol{A} = -\mathbb{S}^\top\boldsymbol{\mu}$, thus:

$$\boldsymbol{A} = -\mathbb{S}^\top\boldsymbol{\mu} = -\mathbb{S}^\top\delta\boldsymbol{\mu} = -\mathbb{S}^\top\left(\mathbb{X}^{\rm eq}\right)^{-1}\delta\boldsymbol{x} \tag{67}$$

where in the second equality we have introduced a diagonal matrix $\mathbb{X}^{\rm eq}$ whose entries are given by $(\mathbb{X}^{\rm eq})_i = x_i^{\rm eq}$. Also, by applying the matrix $\mathbb{G}$ to $\delta\boldsymbol{x}(t)$ one gets:

$$\mathbb{G}\,\delta\boldsymbol{x}(t) = \begin{pmatrix} \boldsymbol{0} \\ \hline \delta\boldsymbol{z}(t) \end{pmatrix} \begin{matrix} \updownarrow N-M \\ \updownarrow M \end{matrix}\,, \tag{68}$$

where $\delta\boldsymbol{z}(t)$ is the vector containing the distance to equilibrium for the $z_\gamma$ variables, $\delta z_\gamma = z_\gamma(t) - z_\gamma^{\rm eq}$. The first $N-M$ zeroes in Eq. (68) correspond to the conservation laws $\boldsymbol{\ell}$ (by definition, $\boldsymbol{\ell}\cdot\delta\boldsymbol{x}(t) = 0$). The relation in Eq. (68) can be inverted using the structure of the row reduction [see Eq. (A17) in Appendix A], and one gets:

$$\delta\boldsymbol{x}(t) = \mathbb{S}_M\,\delta\boldsymbol{z}(t)\,, \tag{69}$$

where we recall that $\mathbb{S}_M$ is the matrix consisting of the $M$ first columns of $\mathbb{S}$. We replace Eq. (69) into Eq. (67) to obtain an expression for the affinity as a function of the reduced set of variables $\delta z_\gamma$. In particular, for the $M$ cocycle affinities $A^c_\gamma$ we find:

$$A^c_\gamma = -\sum_{\gamma'}\Big(\underbrace{\mathbb{S}_M^\top\left(\mathbb{X}^{\rm eq}\right)^{-1}\mathbb{S}_M}_{\mathbb{H}_Q}\Big)_{\gamma,\gamma'}\delta z_{\gamma'}\,, \quad \forall\gamma\,. \tag{70}$$

Notice that the $M \times M$ matrix $\mathbb{H}_Q$ defined in this relation is symmetric and positive-defined, in accordance to the conservative nature of $A^c_\gamma$. Finally, combining Eqs. (54), (64) and (70) we obtain the linear evolution of $\delta\boldsymbol{z}$:

$$\frac{d}{dt}\delta\boldsymbol{z}(t) = -\mathbb{B}\,\delta\boldsymbol{z}(t)\,, \tag{71}$$

where $\mathbb{B} = \mathbb{L}_Q\mathbb{H}_Q$ is the stability matrix whose spectrum controls the relaxation to the equilibrium state and is strictly positive, $\mathrm{Sp}\,\mathbb{B} = \mathrm{Sp}\,\mathbb{L}_Q\mathbb{H}_Q > 0$. As a consequence, the system relaxes monotonically to the equilibrium steady state, which in the theory of dynamical systems is called a stable node. Interestingly, the matrices $\mathbb{L}_Q$ and $\mathbb{H}_Q$, which appear naturally in our deterministic framework, bear a physical meaning in the underlying stochastic dynamics. It is known that Gaussian temporal fluctuations around equilibrium are well described by the (linearized) chemical Langevin equation [72]. In Appendix C, we show that the Onsager matrix $\mathbb{L}_Q$ appears to be the covariance matrix of the Gaussian noise entering the Langevin description, where, in the large but finite $\Omega$ asymptotics, $\delta\boldsymbol{z}(t)$ becomes a stochastic process. The matrix $\mathbb{H}_Q$ appears as the Hessian matrix associated to the quadratic quasipotential $\varphi(\delta\boldsymbol{z}) = \frac{1}{2}\delta\boldsymbol{z}^\top\mathbb{H}_Q\,\delta\boldsymbol{z}$ from which the conservative force $-\mathbb{L}_Q\nabla\Phi$ of the Langevin equation is obtained [see Eq. (C12)] and that describes the equilibrium Gaussian distribution $\propto \exp\left[-\Omega\,\varphi(\delta\boldsymbol{z})\right]$ of the deviation $\delta\boldsymbol{z}$ around its average value $\boldsymbol{0}$.

### 2. *Steady-state response*

For an open system, relaxation to equilibrium is impeded by the continuous supply of external currents, as described in Sec. II E. Then, the overall affinity is non-conservative and takes the explicit form given by Eq. (40) with the non-vanishing circulations $A^e_\alpha$ determined by the chemostatting parameters $\boldsymbol{a}$:

$$A^e_\alpha = \boldsymbol{c}^\alpha\cdot\boldsymbol{A} = \boldsymbol{c}^\alpha\cdot\boldsymbol{a} \neq 0 \quad \forall\alpha\,. \tag{72}$$

Consider for simplicity the case of a time-independent chemostatting, $\boldsymbol{a} \ll 1$, so that the system reaches a nonequilibrium steady state $\boldsymbol{x}^*$ linearly close to the equilibrium state, $\delta\boldsymbol{x}^* = \boldsymbol{x}^* - \boldsymbol{x}^{\rm eq} \ll 1$. In the linear regime, we can replace Eq. (63) into Eq. (72) so that

$$A^e_\alpha = \boldsymbol{c}^\alpha\cdot\mathbb{A}^{-1}\boldsymbol{J} \overset{t\to\infty}{=} \sum_{\alpha'}\underbrace{\boldsymbol{c}^{\alpha\top}\mathbb{A}^{-1}\boldsymbol{c}^{\alpha'}}_{(\mathbb{L}_P)_{\alpha\alpha'}}J^{c,*}_{\alpha'}\,, \tag{73}$$

where in the last equality we used the current decomposition Eq. (26) under the steady state condition, i.e. when only the cyclic currents survive $J^{c,*}_\alpha = \lim_{t\to\infty}J^c_\alpha(t)$. The matrix $\mathbb{L}_P$ in Eq. (73) describes the linear relation between the non-conservative affinities maintaining the system out of equilibrium and the non-zero currents characterizing the steady state. As such, it corresponds to the Onsager matrix of the steady-state response and one verifies that it is symmetric and positive-defined. Notice that the response in Eq. (73) corresponds to the one initially studied by Schnakenberg [13]. In his analysis, Schnakenberg emphasizes the thermodynamic significance of cycles. Indeed, we see that the steady-state response is fully determined by a number of currents and affinities given by the number of cycles in the underlying topology.

Notably, the dimensions of the Onsager matrices controlling transient response $\mathbb{L}_Q$ and steady-state response $\mathbb{L}_P$ are not the same: they are fixed by the number of cocycles and cycles, respectively. In both cases, we see how the off-diagonal contributions to the response emerge once we restrict the analysis to the subset of physically relevant currents and affinities. Formally, this is done by projecting the total currents and affinities on the subspaces defined by the cycles and the cocycles (see Appendix D). A natural question is how the two

Onsager matrices $\mathbb{L}_P$ and $\mathbb{L}_Q$ are related, given that (*i*) they describe respectively relaxation to equilibrium and response to a small drive (that one thus expects to be related fluctuation-dissipation); (*ii*) they live in (complementary) spaces of different dimensions. We address this question in the following section.

### 3. *Hidden fluctuation-dissipation symmetries*

Following the same convention for the ordering of the reactions as in Sec. II B, we subdivide the diagonal matrix $\mathbb{\Lambda}$ as:

$$\mathbb{\Lambda} = \begin{pmatrix} \mathbb{\Lambda}_M & \mathbb{0} \\ \mathbb{0} & \mathbb{\Lambda}_{R-M} \end{pmatrix}, \tag{74}$$

where the upper diagonal block $\mathbb{\Lambda}_M$ corresponds to the $M$ cochord reactions and the lower diagonal block $\mathbb{\Lambda}_{R-M}$ to the $R-M$ chord reactions. Doing so, from Eqs. (64) and (73), the Onsager matrices $\mathbb{L}_Q$ and $\mathbb{L}_P$ explicitly read

$$\mathbb{L}_Q = \mathbb{\Lambda}_M + \mathbb{T}\mathbb{\Lambda}_{R-M}\mathbb{T}^\top \tag{75}$$
$$\mathbb{L}_P = \mathbb{\Lambda}_{R-M}^{-1} + \mathbb{T}^\top\mathbb{\Lambda}_M^{-1}\mathbb{T}\,, \tag{76}$$

and no apparent connection can be envisaged between them for generic $\mathbb{\Lambda}$.

In order to unveil such connection, we perform the following diagonal transformation for the variables:

$$\hat{\boldsymbol{J}} = \mathbb{\Lambda}^{-1/2}\boldsymbol{J}\,, \quad \hat{\boldsymbol{A}} = \mathbb{\Lambda}^{+1/2}\boldsymbol{A}\,. \tag{77}$$

One sees from Eq. (63) that such change of variable corresponds to a rescaling of the linear-regime current/affinity relation, such that $\hat{J} = \hat{A}$. Moreover, it preserves the orthogonality structure between the potential condition $\boldsymbol{A} \in \mathrm{Im}\,\mathbb{S}^\top$ and the stationary condition $\boldsymbol{J} \in \mathrm{Ker}\,\mathbb{S}$ discussed in Sec. II C: for the new variables, these conditions become

$$\hat{A} \in \mathrm{Im}\,(\mathbb{S}\mathbb{\Lambda}^{1/2})^\top \qquad \text{for conservative affinities} \tag{78}$$
$$\hat{J} \in \mathrm{Ker}\,\mathbb{S}\mathbb{\Lambda}^{1/2} \qquad \text{for stationary currents} \tag{79}$$

involving complementary orthogonal subspaces,

$$\mathrm{Im}\,(\mathbb{S}\mathbb{\Lambda}^{1/2})^\top \perp \mathrm{Ker}\,\mathbb{S}\mathbb{\Lambda}^{1/2}\,. \tag{80}$$

The matrix $\mathbb{\Lambda}$ being invertible, one easily verifies that $\{\mathbb{\Lambda}^{1/2}\boldsymbol{c}^\gamma\}$ constitutes a basis for the subspace in Eq. (78) while $\{\mathbb{\Lambda}^{-1/2}\boldsymbol{c}^\alpha\}$ forms a basis for the subspace in Eq. (79). Accordingly, we can introduce rescaled cocycles and cycles defined as $\hat{c}^\gamma = \mathbb{\Lambda}^{1/2}\boldsymbol{c}^\gamma\mathbb{\Lambda}_\gamma^{-1/2}$ and $\hat{c}^\alpha = \mathbb{\Lambda}^{-1/2}\boldsymbol{c}^\alpha\mathbb{\Lambda}_\alpha^{1/2}$.

The new cycles and cocycles still satisfy the orthogonality relations Eqs. (15) and (18) and constitute a basis in $\mathbf{R}^R$, namely:

$$\{\hat{\boldsymbol{c}}^\gamma, \hat{\boldsymbol{c}}^\alpha\} = \begin{pmatrix} \mathbb{1}_M & -\hat{\mathbb{T}} \\ \hat{\mathbb{T}}^\top & \mathbb{1}_{R-M} \end{pmatrix}, \tag{81}$$

where $\hat{\mathbb{T}} = \mathbb{\Lambda}_M^{-1/2}\mathbb{T}\mathbb{\Lambda}_{R-M}^{1/2}$. As a consequence, the decompositions for the affinity Eq. (19) and the current Eq. (26) readily generalize to the new representation:

$$\hat{\boldsymbol{A}} = \sum_\gamma \hat{A}_\gamma^c\hat{\boldsymbol{c}}^\gamma + \sum_\alpha \hat{A}_\alpha^e\boldsymbol{e}^\alpha \tag{82}$$
$$\hat{\boldsymbol{J}} = \sum_\gamma \hat{J}_\gamma^e\boldsymbol{e}^\gamma + \sum_\alpha \hat{J}_\alpha^c\hat{\boldsymbol{c}}^\alpha \tag{83}$$

with macroscopic components defined as:

$$\hat{J}_\gamma^e = \hat{c}^\gamma\cdot\hat{\boldsymbol{J}}\,, \qquad \hat{J}_\alpha^c = e^\alpha\cdot\hat{\boldsymbol{J}}\,, \tag{84}$$
$$\hat{A}_\gamma^c = e^\gamma\cdot\hat{\boldsymbol{A}}\,, \qquad \hat{A}_\alpha^e = \hat{c}^\alpha\cdot\hat{\boldsymbol{A}}\,. \tag{85}$$

Finally, in the linear regime, the Onsager matrices $\hat{\mathbb{L}}_Q$ and $\hat{\mathbb{L}}_P$ such that $\hat{J}_\gamma^e = \sum_{\gamma'}(\hat{\mathbb{L}}_Q)_{\gamma\gamma'}\hat{A}_{\gamma'}^c$ and $\hat{A}_\alpha^e = \sum_{\alpha'}(\hat{\mathbb{L}}_P)_{\alpha,\alpha'}\hat{J}_{\alpha'}^{c,*}$ are obtained following the same procedure as before and read:

$$\hat{\mathbb{L}}_Q = \mathbb{1}_M + \hat{\mathbb{T}}\hat{\mathbb{T}}^\top = \mathbb{\Lambda}_M^{-1/2}\mathbb{L}_Q\mathbb{\Lambda}_M^{-1/2} \tag{86}$$
$$\hat{\mathbb{L}}_P = \mathbb{1}_{R-M} + \hat{\mathbb{T}}^\top\hat{\mathbb{T}} = \mathbb{\Lambda}_{R-M}^{1/2}\mathbb{L}_P\mathbb{\Lambda}_{R-M}^{1/2}\,. \tag{87}$$

Interestingly, the two matrices $\hat{\mathbb{T}}\hat{\mathbb{T}}^\top$ and $\hat{\mathbb{T}}^\top\hat{\mathbb{T}}$ share the same non-zero eigenvalues, meaning that the Onsager matrices $\hat{\mathbb{L}}_Q$ and $\hat{\mathbb{L}}_P$ also have the same spectra up to the multiplicity of eigenvalue $\lambda = 1$.

*Proof.* Let us consider an eigenvector $\boldsymbol{w}$ and the corresponding eigenvalue $\lambda \neq 0$ of the matrix $\hat{\mathbb{T}}^\top\hat{\mathbb{T}}$ so that

$$\exists \boldsymbol{w} : \hat{\mathbb{T}}^\top\hat{\mathbb{T}}\boldsymbol{w} = \lambda\boldsymbol{w}\,. \tag{88}$$

By multiplying by $\hat{\mathbb{T}}$ on the left one gets $\hat{\mathbb{T}}\hat{\mathbb{T}}^\top\hat{\mathbb{T}}\boldsymbol{w} = \lambda\hat{\mathbb{T}}\boldsymbol{w}$. If $\hat{\mathbb{T}}\boldsymbol{w}$ is different from zero, then $\lambda$ is also an eigenvalue of the matrix $\hat{\mathbb{T}}\hat{\mathbb{T}}^\top$. *Ad absurdum* let us assume that $\hat{\mathbb{T}}\boldsymbol{w} = 0$. From Eq. (88) we see that this implies $\lambda\boldsymbol{w} = 0$ which is against the original assumption ($\lambda \neq 0$). As a consequence, for any non-zero eigenvalues $\lambda$,

$$\lambda \in \mathrm{Sp}\,\hat{\mathbb{T}}^\top\hat{\mathbb{T}} \;\Leftrightarrow\; \lambda \in \mathrm{Sp}\,\hat{\mathbb{T}}\hat{\mathbb{T}}^\top \qquad \square \tag{89}$$

The diagonal transformation in Eq. (77) has revealed a hidden symmetry in the spectra of the Onsager matrices of complex CRNs. It links the transient relaxation of the system produced by a spontaneous (or imposed) fluctuation to the stationary response of the system to a drive. As such it generalizes the Einstein relation between mobility and diffusivity to non-trivial topologies. In this context, the Onsager matrix $\mathbb{L}_P$ plays the role of a 'mobility' matrix and the Onsager matrix $\mathbb{L}_Q$ that of a 'diffusivity' matrix (see Appendix C). We stress that this symmetry is highly nontrivial: the two matrices have different dimensions due to the existence of conservation laws and cycles.

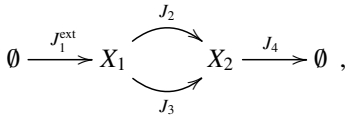
## F. Application: linear-regime thermodynamically feasible reconstruction of metabolic networks

We cast the problem mentioned in the Introduction by working out a couple of simple examples before giving a straightforward reconstruction algorithm.

Typically, in metabolomic datasets some species are injected in the cell by an external reaction of the kind $\emptyset \to Y$, according to the chemostatting procedure described in Sec. II E. Instead, there is no further chemostat associated with internal cycles, which then have vanishing affinity. Notice that, with respect to the chemostatting procedure previously described, the problem of reconstruction requires just a small shift of picture, where instead of separating internal and external species we separate internal and external reactions. A subset of external currents are thus treated as known (fixed) parameters and the problem consists in finding a value for the internal currents which grants thermodynamic feasibility (KVL) and conservation of chemical species (KCL).

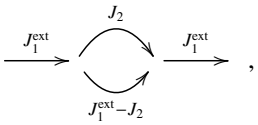
### 1. *Noninteracting network, one internal cycle.*

As the simplest possible case consider the network

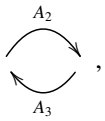



where for the sake of ease we momentarily loosen the assumption that there are no multiple reactions with the same stoichiometry. We assume that $J_1^{\rm ext}$ is known, and that all other currents have to be reconstructed. Notice that $J_4$ is also an external current, which we could consider to fix, but then we should keep in mind that by current conservation $J_4 = J_1^{\rm ext}$. In general, not any external current can be independently fixed, so some care is needed in choosing the boundary data.

We now want to impose dynamic and thermodynamic constraints. KCL [Eq. (29)] at the nodes of the network clearly implies




where current $J_2$ now acts as a free parameter. As far as KCL is involved, this parameter could take any real value. We will now implement thermodynamic constraints to reduce the span of $J_2$. Given the constitutive equation Eq. (7), and identifying the internal cycle




KVL [Eq. (21)] yields

$$\lambda_2^+\lambda_3^- = \lambda_2^-\lambda_3^+ , \tag{90}$$

where the $\lambda_\rho^\pm$'s are evaluated at the steady state $\lambda_\rho^\pm = \lambda_\rho^\pm(\boldsymbol{x}^*)$. Let us now rewrite this in terms of the external current $J_1^{\rm ext}$, of the internal current $J_2$ (that we take as free parameter), and of the velocities $\lambda_2^-$ and $\lambda_3^-$ (chosen arbitrarily). After a little work we obtain

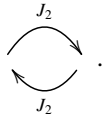
$$J_2 = \frac{1}{1+\lambda_3^-/\lambda_2^-}\, J_1^{\rm ext}. \tag{91}$$

In reconstruction problems the actual values of $\lambda_2^-$ and $\lambda_3^-$ are usually not known. However, by the fact that they are positive, this latter equation implies

$$0 \le J_2 \le J_1^{\rm ext}. \tag{92}$$

Notice that for positive $J_1^{\rm ext}$ both currents $J_2$ and $J_3$ have to flow left-to-right, which makes physical sense: one would not expect a river that bifurcates around an island to have upward flows along one of its branches!

However, reconstruction procedures that just implement KCL may fail to impose this constraint, thus producing a thermodynamically infeasible cycle. In particular, for vanishing $J_1^{\rm ext}$, one could have a perpetuum mobile:




Other reconstruction procedures based on linear optimization are instead bound to hit the boundary of the space of solutions, yielding the trivial solutions $J_2 = 0$ or $J_1^{\rm ext} - J_2 = 0$, which basically remove internal cycles and do not fully allow to explore meaningful possibilities.

The same solution as Eq. (91) can be found by the linear-regime approach. In view of Eq. (63), KVL prescribes

$$0 = A_2 - A_3 \approx \frac{J_2}{\lambda_2^{\rm eq}} - \frac{J_1^{\rm ext} - J_2}{\lambda_3^{\rm eq}} , \tag{93}$$

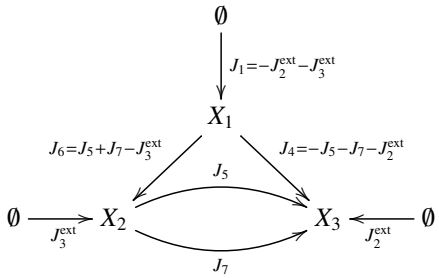
leading to

$$J_2 = \frac{1}{1+\lambda_3^{\rm eq}/\lambda_2^{\rm eq}}\, J_1^{\rm ext}. \tag{94}$$

This, in fact, is almost identical to Eq. (91), but for the fact that it is written in terms of equilibrium values of the velocities. We will discuss the difference between $\lambda$'s and $\lambda^{\rm eq}$'s later on. For the sake of our analysis, notice that, provided that the velocities are just some positive quantities, Eqs. (91) and (94) impose the exact same constraint. Linearization here has no actual consequence.

### 2. *Noninteracting network, two cycles.*

Let us now consider:



where we already implemented KCL in terms of the external currents $J_2^{\rm ext}$, $J_3^{\rm ext}$ and of the internal currents $J_5$, $J_7$.

KVL on the two internal cycles instead prescribes:

$$\lambda_5^+\lambda_7^- = \lambda_5^-\lambda_7^+ \tag{95}$$
$$\lambda_4^+\lambda_5^-\lambda_6^- = \lambda_4^-\lambda_5^+\lambda_6^+ . \tag{96}$$

Selecting $\lambda_4^-, \lambda_5^-, \lambda_6^-$ as free parameters, after some work from the first we find the linear equation $J_7 = J_5\,\lambda_7^-/\lambda_5^-$, and letting $\beta = 1 + \lambda_7^-/\lambda_5^-$ from the second we obtain the quadratic equation

$$\frac{\beta}{\lambda_5^-\lambda_6^-}J_5^2 + \left(\frac{1}{\lambda_5^-} + \frac{\beta}{\lambda_4^-} + \frac{\beta}{\lambda_6^-} - \frac{J_3^{\rm ext}}{\lambda_5^-\lambda_6^-}\right)J_5 + \frac{J_2^{\rm ext}}{\lambda_4^-} - \frac{J_3^{\rm ext}}{\lambda_6^-} = 0. \tag{97}$$

Once again, notice that when there is no external current $J_2^{\rm ext} = J_3^{\rm ext} = 0$ we get $J_5 = J_7 = 0$, no perpetuum mobile. Otherwise, for given values of $J_2^{\rm ext}, J_3^{\rm ext}$ one can use this equation to explore the possible values of the internal currents in terms of arbitrarily chosen parameters $\lambda_4^-, \lambda_5^-, \lambda_6^-$.

This quadratic problem is already becoming complicated (and it is easy to foresee that for more complicated topologies this will give rise to higher-order polynomial systems). Given that we are interested in some bulk characterization of the landscape, a full solution may be an overshooting. Therefore, like in the previous example, let us proceed by linearization of KVL:

$$0 = A_5 - A_7 \approx \frac{J_5}{\lambda_5^{\rm eq}} - \frac{J_7}{\lambda_7^{\rm eq}} \tag{98}$$
$$0 = A_4 - A_5 - A_6 \approx \frac{J_4}{\lambda_4^{\rm eq}} - \frac{J_5}{\lambda_5^{\rm eq}} - \frac{J_6}{\lambda_6^{\rm eq}}. \tag{99}$$

The first easily gives $J_7 = J_5\,\lambda_7^{\rm eq}/\lambda_5^{\rm eq}$, and letting $\beta^{\rm eq} = 1 + \lambda_7^{\rm eq}/\lambda_5^{\rm eq}$ the second yields

$$\left(\frac{1}{\lambda_5^{\rm eq}} + \frac{\beta^{\rm eq}}{\lambda_4^{\rm eq}} + \frac{\beta^{\rm eq}}{\lambda_6^{\rm eq}}\right)J_5 + \frac{J_2^{\rm ext}}{\lambda_4^{\rm eq}} - \frac{J_3^{\rm ext}}{\lambda_6^{\rm eq}} = 0. \tag{100}$$

But now this is a simple linear equation that, given the external currents, provides a reconstruction for any given choice of positive real $\lambda$'s. Notice that this equation can also be obtained from Eq. (97) by disregarding terms of order $J^2$. Thus the reconstruction is thermodynamically feasible.

### 3. *Considerations and problem setting.*

The key takeaways of these examples are the following. Both ways, direct solution of Kirchhoff's laws and linearization grant thermodynamic feasibility of reconstruction. The first is consistent with the basic tenets of reaction rate theory but increasingly complicated with topology. However, given that metabolic reconstruction is a very underdetermined problem and that we are more interested in spanning a space of viable possibilities rather than into specific solutions, we can resort to linearization to obtain a broad bulk of feasible reconstructions.

Given the stoichiometric matrix $\mathbb{S}$ of a metabolic network, the first step is to split it in terms of internal reactions $\mathbb{S}_X$, that we want to reconstruct, and external reactions $\mathbb{S}_Y$, for which there exist data or that we want to control. Then, given some input currents $\boldsymbol{J}_Y$ for the external reactions, we want to produce a reconstructed current $\boldsymbol{J}_X$ that satisfies KCL and KVL. We base our solution on the linear-regime assumption.

### 4. *Linear-regime reconstruction algorithm.*

0) Input $\mathbb{S}$;

1) Split it as $\mathbb{S} = (\mathbb{S}_Y, \mathbb{S}_X)$ in terms of internal reactions $\mathbb{S}_X$ and external reactions $\mathbb{S}_Y$. In the following we refer to $R$ as the number of internal reactions and $M$ as the rank of $\mathbb{S}_X$;

2) Input the external currents $\boldsymbol{J}_Y$;

3) For all left nullvectors $\boldsymbol{\ell}_X$ of $\mathbb{S}_X$ check that $\boldsymbol{\ell}_X \cdot \mathbb{S}_Y \boldsymbol{J}_Y = 0$, else revise the input currents or reduce the number of external reactions and go back to 1);

4) Row-reduce the composite matrix $(\mathbb{S}_Y, \mathbb{S}_X)$ to obtain the composite matrix $(\mathbb{1}, \mathbb{G}\mathbb{S}_X)$;

5) Reorder reactions in such way that $\mathbb{G}\mathbb{S}_X$ takes the form Eq. (46); Reorder $\mathbb{S}$ accordingly;

6) Input $R$ real positive parameters $\lambda_\rho^{\rm eq}$;

7) Let $\mathbb{\Lambda} = \mathrm{diag}\,\{\lambda_\rho^{\rm eq}\}_\rho$; Let $\mathrm{diag}\,(\mathbb{\Lambda}_M, \mathbb{\Lambda}_{R-M}) = \mathbb{\Lambda}$; Let $\mathbb{L}_P$ as per Eq. (76);

8) Let

$$\boldsymbol{J}_X = \begin{pmatrix} (\mathbb{1}_M - \mathbb{T}\mathbb{L}_P^{-1}\mathbb{T}^\top\mathbb{\Lambda}_M^{-1})\mathbb{G}_M\mathbb{S}_Y\boldsymbol{J}_Y \\ \mathbb{L}_P^{-1}\mathbb{T}^\top\mathbb{G}_M\mathbb{S}_Y\boldsymbol{J}_Y \end{pmatrix}. \tag{101}$$

In Appendix E we explain the mathematical rationale behind this algorithm and prove that it satisfies both KCL and KVL. In the conclusions we comment on a possible biochemical interpretation of the coefficients $\lambda_\rho^{\rm eq}$.

## IV. OUTLOOK

Interacting mass-action CRNs present a host of behaviors coming from the multiplicity of fixed points [73, 74] and ranging from non-linear oscillations [75] to chaos [76]. In our work, we focused on the stationary state and on relaxation properties, but the decomposition of currents and affinities that we identified could be useful tools to study such non-stationary phenomena. Also, naturally, the geometrical tools that we identified could help to study the role played by deficiency [22, 77, 78] in dissipation and noise [51, 79]. The symmetries in linear response could be compared to the recent approach of Ref. [80]. The separation of timescales

that we identified in Sec. III C 4 through the evolution of the population of islands associated to cocycles bears strong ties with the control of chemical kinetics through catalysts and inhibitors, following for instance the recent methods presented in Ref. [19]. Also, autocatalysis plays an essential role in biochemical processes and it is only recently that the classifications of CRNs leading to this type of self-replication have been identified [23]. Such classification could be investigated in the light of our geometrical tools. Another geometrical approach was recently proposed in Ref. [81] where a notion of Hessian geometry of CRNs is constructed; it would be worth to identify the link between such approach and the notions of cycles and cocycles we have put forward. In the same way, geometrical decompositions were identified in Markov processes [82] and in field theories [28], which may be related to the one present in this paper. And of course much could be gained from going beyond the well-stirred limit by considering extended systems where spatial inhomogeneities play a role.

On the graph-theoretical side, a known duality exists between vertices and faces for planar graphs, which, in our language exchanges the roles of cycles and cocycles. For nonplanar graphs, the concept of matroid [83] allows one to treat abstract independence sets based on circuits and to generalize dualities. For noninteracting CRNs, such duality thus implies a mapping between stationary currents (supported by cycles) and transient ones (supported by cocycles). It would be interesting to investigate the consequences of such a mapping. Our definition of cycles and cocycles of the hypergraph associated to a generic CRN opens natural questions: can such duality be extended to a class of hypergraphs, and what could we learn from it? Also, cycles were recently shown to control several aspects of fluctuations and large deviations in the graph associated to Markov jump processes [84–86]. Such results could be extended to dynamics on hypergraphs using the tools we have put forward.

Regarding the reconstruction algorithm of Sec. III F, the main open questions are about how to further constrain solutions with empirical data or reasonable target functions, and whether the linear-regime assumption is consistent with physiological conditions. About this latter, further analysis is needed to characterize the difference between the linear-regime landscape and the algebraic variety of solutions of the nonlinear KVL. About the former, the main virtue of our proposal is that the coefficients $\lambda_\rho^{\rm eq}$'s are independent one of each other and can take any real value, while previous reconstruction efforts had to deal with non-convex spaces of parameters where optimization algorithms could get stuck into subspaces or at boundaries. Let us argue that these parameters also make biochemical sense, by going back to their linear-response meaning. By the fluctuation-dissipation paradigm, the coefficients $\lambda_\rho^{\rm eq}$'s quantify the spontaneous activity of a system at equilibrium, that is, in the absence of external currents. In theory, one would have to realize the sole reaction $\rho$ *in vitro* and measure its activity. In practice, given that a single reaction's activity can be associated to the expression of the enzyme that catalyzes it, we propose that the $\lambda_\rho^{\rm eq}$ could be roughly proportional to the abundance of the corresponding enzyme, for which there could be available data.

In this paper, we treated mass-action CRNs, but, being most of the results of a purely topological nature, most of them apply to more generic reaction kinetic laws (such as effective enzymatic models) with the only requirement that there exist conjugate currents and forces such that $\boldsymbol{J} > 0$ if and only if $\boldsymbol{A} > 0$, and $\boldsymbol{J} = 0$ if and only if $\boldsymbol{A} = 0$. What is special about mass-action kinetics is that the cycle affinities $A_\alpha^e$ do not depend explicitly on the populations, and therefore are constants of motion.

In this respect, an interesting direction to explore is that of reaction networks where species are not chemical, but biological. There, no notion of thermodynamic feasibility imposes that the affinities of internal cycles have to be zero, but the decompositions of affinities and currents that we propose still apply to this case. In such context, migration from regional pools of species can also play the role of chemostatting. Ecological and evolutionary models are known to present a variety of phenomena such as strong space-time fluctuations [87], chaos [88] or sensitivity to noise [89]. Systems modeled by (generalized) Lotka–Voltera equations [87, 88, 90] are particularly amenable to the tools we propose, as, at the population level, they can be put in correspondence with CRNs [91]. We thus hope that the geometrical concepts we have identified can be helpful to study such problems.

## ACKNOWLEDGMENTS

SDC and VL acknowledge support by the ANR-18-CE30-0028-01 Grant LABS and by the EverEvol CNRS MITI project. The research was supported by the National Research Fund Luxembourg (project CORE ThermoComp R-AGR-3425-10) and by the European Research Council, project NanoThermo (ERC-2015-CoG Agreement No. 681456).  SDC is grateful to Emanuele Penocchio and Gianmaria Falasco for helpful discussions on chemical reaction networks. MP thanks Daniele De Martino, Alexander Skupin and Susan Ghaderi for useful insights and discussions on metabolic reconstruction issues. We are grateful to Pierre Recho for stimulating discussions that inspired this work.

## Appendix A: Integration and differentiation on the networks of chemical reactions

### 1. Noninteracting CRNs: integration on spanning trees

Consider a set of unimolecular reactions, as in Sec. II, and assume that the corresponding graph $\mathcal{G}$ is connected. The stoichiometric matrix satisfies $(\mathbb{S}^\top \boldsymbol{V})_\rho = V_{t(\rho)} - V_{s(\rho)}$ so that $\mathbb{S}^\top$ can be seen as a gradient operator, which transforms a potential $\boldsymbol{V}$ defined on every species/node $i$ into a (chemical) force field between the source $s(\rho)$ and target $t(\rho)$ of every reaction/edge $\rho$. In this Appendix, we aim at building an explicit 'integrator': that is, if a conservative force $\boldsymbol{A}$ belongs to $\mathrm{Im}\,\mathbb{S}^\top$

we want to build a potential $\boldsymbol{V}$ such that $\boldsymbol{A} = -\mathbb{S}^\top \boldsymbol{V}$. This will be done by defining an integrator matrix $\mathbb{G}^\top$ from the entries of $\mathbb{S}$. Then, we present how these two matrices are related.

We remark in advance that $\boldsymbol{V}$ is not unique: if $\boldsymbol{V}$ and $\boldsymbol{V}'$ yield the same $\boldsymbol{A}$, we have that $\mathbb{S}^\top(\boldsymbol{V}' - \boldsymbol{V}) = 0$ so that $\boldsymbol{V}' - \boldsymbol{V} \in \mathrm{Ker}\,\mathbb{S}^\top$, meaning that the two potentials are equal up to a global constant, since $\mathrm{Ker}\,\mathbb{S}^\top$ is spanned by $\boldsymbol{\ell}_0 = (1, \cdots, 1)$. This is similar to what happens in the continuum when integrating a function: a primitive is defined « up to a constant ».

Since the labeling and the orientation of reactions is arbitrary, we can redefine them to our best convenience. To do so, we first select arbitrarily one of the nodes, that will play the role of the 'root' of the graph. Then, we fix a spanning tree $T_{\mathcal{G}}$ (see Fig. 2), that is, a set of $M$ independent reactions. This allows one to fix the orientations: the edges in $T_{\mathcal{G}}$ are directed toward the root while the other edges (the chords) are oriented arbitrarily. We then turn to labeling, starting by the nodes. The root is node 1, and we label the other nodes incrementally from the root along $T_{\mathcal{G}}$ as follows (see Fig. 1): at every branching point of $T_{\mathcal{G}}$, we pick one of the branches and continue the numbering of species incrementally, until we reach a 'leaf' (i.e. a node of the graph without further edge). We then come back to the last branching point and continue the procedure until every remaining node is exhausted. Second, we label the edges. From node 2, a single edge points towards the root, which we label as edge 1. Recursively, the edge pointing out of node $\gamma + 1$ (for $1 \le \gamma < M$) is labeled as edge $\gamma$. This exhausts the $M = N - 1$ cochords, labeled from 1 to $M$. The remaining $R - M$ chords (equal, in number, to the number of cycles) are labeled arbitrarily from $M + 1$ to $R$. Doing so, for the first $M$ columns of $\mathbb{S}$ (indexed by $1 \le \gamma \le M$) we have $s(\gamma) = \gamma + 1$ and $t(\gamma) \le \gamma$. This implies that the stoichiometric matrix takes the form

$$\mathbb{S} = \left( \underbrace{\begin{array}{cccc} 1 & 1 & & \\ -1 & & & 1 \\ & -1 & & \\ & & \ddots & \\ & (0) & & -1 \end{array}}_{\mathbb{S}_M} \middle| \; \mathbb{S}_{\mathrm{dep}} \right) \tag{A1}$$

where: (*i*) on the $N \times M$ block $\mathbb{S}_M$, the matrix $-\mathbb{1}_M$ lies on the lower diagonal and, on every column $\gamma$, there is a single entry 1 on line $t(\gamma) \le \gamma$; and (*ii*) the last $R - M$ columns correspond to the chords, which are the dependent reactions. This means there exists a $M \times (R - M)$ matrix $\mathbb{T}$ such that $\mathbb{S}_{\mathrm{dep}} = \mathbb{S}_M \mathbb{T}$, encoding the fact that every column of $\mathbb{S}_{\mathrm{dep}}$ can be expressed as a linear combination of the $M$ independent columns of $\mathbb{S}_M$. In fact, this encodes a graphical property: every chord is part of a cycle (and every cycle has exactly one chord, see Fig. 2), and the algebraic dependency we just explain encodes that the chord reaction can be obtained by applying all the cochord reactions of the cycle (with the adequate orientation). Notice last that the first line of $\mathbb{S}$ contains only positive entries, since by our convention the root only has entering edges.

We define an $N \times N$ matrix $\mathbb{G}$ from Eq. (22) and we recall that $\mathcal{U}(j)$ is the set of nodes (including $j$) that are upstream of $j$ on $T_{\mathcal{G}}$. From a potential $\boldsymbol{V}$ defined on the nodes (and imposed to verify $V_{\mathrm{root}} = 0$), we define a set of forces

$$A_\gamma = V_{s(\gamma)} - V_{t(\gamma)} = -(\mathbb{S}^\top \boldsymbol{V})_\gamma \tag{A2}$$

for every cochord. Because the line $i$ of $\mathbb{G}^\top$ contains 1 for every node located in between the root and node $i$ (i.e. for every node $j$ such that $i \in \mathcal{U}(j)$) we have, by telescoping sum

$$V_i = V_i - V_{\mathrm{root}} = \sum_\gamma A_\gamma \, \delta_{i \in \mathcal{U}(s(\gamma))} = \sum_\gamma (\mathbb{G}^\top)_{i,s(\gamma)} \, A_\gamma \,. \tag{A3}$$

Notice that, for every node $j \neq$ root there is exactly one cochord $\gamma$ on the spanning tree such that $j = s(\gamma)$. This allows one to express the sum in Eq. (A3) as a integration along the unique path of cochords $\gamma$ connecting node $i$ to the root along the spanning tree. Equations (A2) and (A3) express a one-to-one relation between a set of $M$ forces $A_\gamma$ defined on the cochords, and a set of $N = M + 1$ potentials $V_i$ (including $V_{\mathrm{root}} = 0$) defined on the nodes. These two equations thus encode the differentiation and the integration of a conservative force on a graph: indeed if $\boldsymbol{A} \in \mathrm{Im}\,\mathbb{S}^\top$ is a 'gradient', the potential $\boldsymbol{V}$ defined through the 'integral' (A3) of $M$ components of $\boldsymbol{A}$ generates the full vector $\boldsymbol{A}$ through $\boldsymbol{A} = -\mathbb{S}^\top \boldsymbol{V}$. As remarked above, such potential $\boldsymbol{V}$ is unique up to a constant, and the condition $V_{\mathrm{root}} = 0$ fixes $\boldsymbol{V}$ uniquely (independently of the choice of the spanning tree).

We now identify an algebraic relation between $\mathbb{G}$ and $\mathbb{S}$. To do so, one defines for every node $j \neq$ root (i.e. $2 \le j \le N$) a unit 'charge' as the potential $\boldsymbol{\mathcal{V}}^j$, a vector of entries $\mathcal{V}^j_i = \delta_{ij}$. The corresponding forces $\boldsymbol{\mathcal{A}}^j$ defined on the cochords from Eq. (A2) have entries

$$\mathcal{A}^j_\gamma = \begin{cases} -1, & \text{if } t(\gamma) = j \\ 1, & \text{if } s(\gamma) = j \\ 0, & \text{otherwise,} \end{cases} \qquad \text{i.e.,} \quad \mathcal{A}^j_\gamma = -(\mathbb{S}^\top_M{}')_{\gamma j} \tag{A4}$$

where $\mathbb{S}^\top_M{}'$ is the $M \times M$ matrix constituted of the $M$ last columns of $\mathbb{S}^\top_M$ [the transpose of the matrix defined in Eq. (A1)]. By direct application of Eq. (A3), we see that the unit potential $\boldsymbol{\mathcal{V}}^j$ is obtained from the force $\boldsymbol{\mathcal{A}}^j$ as

$$\mathcal{V}^j_i = \delta_{ij} = \sum_\gamma (\mathbb{G}^\top)_{i,s(\gamma)} \, \mathcal{A}^j_\gamma \,. \tag{A5}$$

We now interpret this relation algebraically. Since $s(\gamma) = \gamma + 1$, we define $\mathbb{G}_M$ as the $M \times M$ submatrix of $\mathbb{G}$ deprived from its first line and column (i.e. $\mathbb{G}_M$ is the black submatrix of $\mathbb{G}$ in Fig. 3b). Then the identity (A5) yields, from Eq. (A4)

$$-\mathbb{G}^\top_M \, \mathbb{S}^\top_M{}' = \mathbb{1}_M \,. \tag{A6}$$

See Refs. [92, 93] for similar results in the context of incidence matrix inversion and Ref. [19] for applications in chemistry.

Before going on, let us take a closer look at this relation. Because $\mathbb{G}^\top_M$ is lower triangular with only 1's on the diagonal, it is invertible and its inverse is given by $-\mathbb{S}^\top_M{}'$. We thus read Eq. (A6) as follows: the subset of $M$ independent reactions between $M$ independent species are described by the

lines of matrix $\mathbb{S}_M^{\top\prime}$, which constitutes an invertible 'core' of the full (and transposed) stoichiometric matrix $\mathbb{S}^\top$. What we have done so far to arrive at Eq. (A6) is to realize that $\mathbb{S}_M^{\top\prime}$ defines a set of forces on the cochords whose integration along the path from root to any node $j \neq$ root gives the 'unit charge' potential $\boldsymbol{\mathcal{V}}^j$ defined above – which is quite natural from the graph perspective. This provides an electrostatic picture of the incidence matrix of $\mathcal{G}$. To proceed, one now remarks that Eq. (A6) implies

$$-\mathbb{S}_M^{\top\prime}\,\mathbb{G}_M^\top = \mathbb{1}_M\,, \tag{A7}$$

which is algebraically trivial but not obvious from the underlying graph-theoretical viewpoint. Yet, completing the matrices, it implies that

$$-\mathbb{S}_M^\top\mathbb{G}^\top = -\left(\begin{array}{c|c} (\mathbb{S}_M^\top)_1 & \mathbb{S}_M^{\top\prime} \end{array}\right)\left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \vdots & & \mathbb{G}_M^\top & \\ 1 & & & \end{array}\right) = \left(\begin{array}{c|c} 0 & \\ \vdots & \mathbb{1}_M \\ 0 & \end{array}\right) \tag{A8}$$

where $(\mathbb{S}_M^\top)_1$ is the first column of $\mathbb{S}_M^\top$ and where we used the fact that $\boldsymbol{\ell}_0^\top$ is a right nullvector of $\mathbb{S}^\top$ (and thus of $\mathbb{S}_M^\top$). Transposing this relation and using the structure of $\mathbb{S}$ given by Eq. (A1) with $\mathbb{S}_{\rm dep} = \mathbb{S}_M\mathbb{T}$ we obtain

$$-\mathbb{G}\,\mathbb{S} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbb{1}_M & \mathbb{T} \end{pmatrix}, \tag{A9}$$

where the first line is full of zeroes. This is Eq. (25) announced in the main text. Physically, it encodes the fact that $\mathbb{S}^\top$, seen as a gradient operator, can be inverted on the cochords by the matrix $-\mathbb{G}^\top$; and that, if cochord forces are conservative, $A_\gamma = -(\mathbb{S}^\top \boldsymbol{V})_\gamma = -(\mathbb{S}_M^\top \boldsymbol{V})_\gamma$, then the corresponding chord forces write as

$$A_\alpha = -(\mathbb{S}^\top \boldsymbol{V})_\alpha = -(\mathbb{S}_{\rm dep}^\top \boldsymbol{V})_\alpha = -(\mathbb{T}^\top\,\mathbb{S}_M^\top \boldsymbol{V})_\alpha \tag{A10}$$

$$= \sum_\gamma \mathbb{T}_{\gamma\alpha} A_\gamma \tag{A11}$$

i.e., they are expressed as a linear combination of the $A_\gamma$'s. Mathematically, Eq. (A9) encodes the row-reduction of $\mathbb{S}$ in an echelon form (see e.g. Ref. [60]). Such identity is at the basis of our geometrical analysis of complex CRNs in Sec. III, and of the algebraic analysis presented in the next paragraph.

### 2. Interacting CRNs: integration on multi-paths

To generalize the construction presented in the previous paragraph, we now follow a complementary path. The matrix $\mathbb{S}^\top$ can still be understood as a (weighted) discrete gradient: every line $\rho$ of $\mathbb{S}^\top$ tells for a given reaction $\rho$ how the products (resp. reactants) contribute positively (resp. negatively) to the affinity $A_\rho$. Our aim is to explain how to 'invert' that gradient and to define an integration that allows one to reconstitute explicitly a potential $\boldsymbol{V}$ such that $\boldsymbol{A} = -\mathbb{S}^\top\boldsymbol{V}$ if $\boldsymbol{A} \in \mathrm{Im}\,\mathbb{S}^\top$ is a conservative affinity. For complex CRNs and their corresponding hypergraph (see Sec. III and Fig. 6), there exists no such notion as a spanning tree, and the topological construction of the previous paragraph (which consists in integrating from a chosen root to a node along the spanning tree) cannot be generalized.

Here we follow a mirror procedure, starting from algebra to build a geometrical picture. Denoting by $M$ the rank of $\mathbb{S}$, we re-order the reactions and species so that the first $M$ reactions are independent, and the last $M$ species are independent. Namely, the first $M$ columns of $\mathbb{S}$ are linearly independent, and the same holds for its last $M$ lines. The row-reduction of $\mathbb{S}$ in echelon form (see e.g. Ref. [60]) ensures that there exists an invertible $N \times N$ matrix $\mathbb{G}$ such that

$$-\mathbb{G}\,\mathbb{S} = \underbrace{\begin{pmatrix} \mathbb{0} & \mathbb{0} \\ \mathbb{1}_M & \mathbb{T} \end{pmatrix}}_{R-M} \begin{matrix} \updownarrow N-M \\ \updownarrow M \end{matrix} \tag{A12}$$

(notice that the ordering conventions we are using makes that the lines of 0's are placed first compared to the canonical row reduction). The matrix $\mathbb{G}$ is not unique and its entries can be found by the Gauss–Jordan elimination procedure through elementary line and columns operations [60]. This ensures that the entries of $\mathbb{G}$ can be taken as rational when, as in our case, $\mathbb{S}$ has integer entries.

To learn about geometry, it is convenient not to rely on Gauss–Jordan elimination and instead to build the matrix $\mathbb{G}$ in a fully explicit manner. We first fix a basis of $\mathrm{Ker}\,\mathbb{S}^\top$ as $N-M$ column vectors representing conservation laws (whose entries are taken as rational). Then, postulating the following form:

$$\mathbb{G}^\top = \underbrace{\left(\begin{array}{c|c} \text{csv} & \begin{matrix} 0 \cdots\cdots 0 \\ 0 \cdots\cdots 0 \end{matrix} \\ \hline \text{laws} & \begin{matrix} \text{escape routes} \\ (\mathbb{G}_M^{\top\prime}) \end{matrix} \end{array}\right)}_{\mathbb{G}_M^\top} \begin{matrix} \updownarrow N-M \\ \updownarrow M \end{matrix} \tag{A13}$$

we want to show that there exists a $M \times M$ matrix $\mathbb{G}_M^{\top\prime}$ which ensures that the key relation

$$-\mathbb{S}^\top\,\mathbb{G}^\top = \begin{pmatrix} \mathbb{0} & \mathbb{1}_M \\ \mathbb{0} & \mathbb{T}^\top \end{pmatrix} \tag{A14}$$

is satisfied.

To do so, we notice that the $N-M$ columns of conservation laws in Eq. (A13) ensure the $N-M$ first columns of 0's in Eq. (A14). Then, by our ordering conventions, the first $M$ columns of $\mathbb{S}$ are independent and form an $N \times M$ matrix $\mathbb{S}_M$, so that we can organize $\mathbb{S}^\top$ as

$$\mathbb{S}^\top = \left(\begin{array}{c} \mathbb{S}_M^\top \\ \hline \mathbb{S}_{\rm dep}^\top \end{array}\right) \begin{matrix} \updownarrow M \\ \updownarrow R-M \end{matrix}\,. \tag{A15}$$

There, $\mathbb{S}_{\text{dep}}$ are the $R-M$ last columns of $\mathbb{S}$, which are depending on the first $M$ ones; this means that they can be expressed as a linear combination of those, i.e. that there exists a $M\times(R-M)$ matrix $\mathbb{T}$ such that $\mathbb{S}_{\text{dep}} = \mathbb{S}_M\mathbb{T}$. With this property, we see that the proof of Eq. (A14) reduces to showing that

$$\mathbb{S}_M^{\top\prime}\mathbb{G}_M^{\top\prime} = \mathbb{1}_M \tag{A16}$$

where $\mathbb{S}_M^{\top\prime}$ is the $M\times M$ matrix consisting of the last $M$ columns of $\mathbb{S}_M^\top$, and $G_M^{\top\prime}$ is the $M\times M$ matrix consisting of the last $M$ lines of $\mathbb{G}_M^\top$ [defined in Eq. (A13)]. Physically, $\mathbb{S}_M^{\top\prime}$ represents a 'core' set of $M$ independent reactions between $M$ independent species. Crucially, it is an invertible matrix, since the last $M$ species are independent [10]. This implies that one can *define* $\mathbb{G}_M^{\top\prime}$ as the inverse of $\mathbb{S}_M^{\top\prime}$, which ensures the relation (A16) to be satisfied. Since $\mathbb{S}_M^{\top\prime}$ has integer entries, we obtain that $\mathbb{G}$ has rational entries (as is also the case when defining $\mathbb{G}$ through Gauss–Jordan elimination).

As we just described, this shows that the form of $\mathbb{G}^\top$ given in Eq. (A13) allows for the row-reduction of $\mathbb{S}$ as in Eq. (A12), with the 'escape routes' in Eq. (A13) being precisely given by the $M\times M$ matrix $\mathbb{G}_M^{\top\prime}$ defined from Eq. (A16). Before showing that the elements of that matrix play the geometrical role of escape routes indeed, it remains to prove that the row-reducing matrix $\mathbb{G}$ defined in Eq. (A13) is invertible. This is done by exhibiting its inverse: one checks with Eqs. (A15) and (A16) that

$$\mathbb{G}^{-1} = \left(\begin{array}{c|c} \mathbb{1}_{N-M} & \\ \mathbb{0} & \mathbb{S}_M \end{array}\right) \tag{A17}$$

is the inverse of $\mathbb{G}$, provided the conservation laws in Eq. (A13) are organized (as columns) as:

$$\begin{pmatrix}\text{csv}\\ \text{laws}\end{pmatrix} = \left(\begin{array}{c} \mathbb{1}_{N-M} \\ \hline -\mathbb{U}^\top \end{array}\right). \tag{A18}$$

Up to now, the specific choice of basis for the conservation laws was left undetermined, and this form fixes it. Its existence is shown *ad absurdum*.

*Proof*: Consider an arbitrary choice of basis for the $N-M$ conservation laws, and split it as follows:

$$\begin{pmatrix}\text{csv}\\ \text{laws}\end{pmatrix} = \left(\begin{array}{c} \mathbb{C}_1 \\ \hline \mathbb{C}_2 \end{array}\right) \begin{matrix} \updownarrow N-M \\ \updownarrow M \end{matrix}. \tag{A19}$$

[10] The proof goes as follows: if $\mathbb{S}_M^{\top\prime}$ is not invertible, there exists a vector $z\neq 0$ such that $\mathbb{S}_M^{\top\prime}z = 0$. Then, defining $\hat{z}\neq 0$ as $N-M$ lines of 0 followed by the $M$ components of $z$, one has $\mathbb{S}_M^\top\hat{z} = 0$ and also $\mathbb{S}_{\text{dep}}^\top\hat{z} = \mathbb{T}^\top\mathbb{S}_M^\top\hat{z} = 0$. From Eq. (A15), this implies $\mathbb{S}^\top\hat{z} = 0$, but this is absurd, since this represents a linear dependency between the last $M$ columns of $\mathbb{S}^\top$ (i.e. between the last $M$ species) – which are independent.

Correspondingly, split the $N$ lines of the stoichiometric matrix as

$$\mathbb{S} = \left(\begin{array}{c} \mathbb{S}_1 \\ \hline \mathbb{S}_2 \end{array}\right) \begin{matrix} \updownarrow N-M \\ \updownarrow M \end{matrix}, \tag{A20}$$

where by hypothesis the $M$ lines of $\mathbb{S}_2$ are independent while the $N-M$ lines of $\mathbb{S}_1$ depend on those of $\mathbb{S}_2$, meaning that there exists a $(N-M)\times M$ matrix $\mathbb{U}$ such that

$$\mathbb{S}_1 = \mathbb{U}\,\mathbb{S}_2\,. \tag{A21}$$

This identity and the decompositions above imply from the definition of conservation laws (they span $\operatorname{Ker}\mathbb{S}^\top$) that

$$\mathbb{S}_2^\top\mathbb{U}^\top\mathbb{C}_1 + \mathbb{S}_2^\top\mathbb{C}_2 = 0\,. \tag{A22}$$

Let us now show that $\mathbb{C}_1$ is invertible. *Ad absurdum*, if this is not the case, there exists a vector $\boldsymbol{x}\neq 0$ such that $\mathbb{C}_1\boldsymbol{x} = 0$. From Eq. (A22) this implies $\mathbb{S}_2^\top\mathbb{C}_2\boldsymbol{x} = 0$, and since the $M$ columns of $\mathbb{S}_2^\top$ are independent, we have also $\mathbb{C}_2\boldsymbol{x} = 0$. From Eq. (A19), we then read

$$\begin{pmatrix}\text{csv}\\ \text{laws}\end{pmatrix}\boldsymbol{x} = 0 \tag{A23}$$

which is absurd, since the column vectors of the matrix of conservation laws are independent. Hence $\mathbb{C}_1$ is an invertible matrix. Multiplying Eq. (A19) by $\mathbb{C}_1^{-1}$ on the right, we see that the conservation laws can be organized as in Eq. (A18), as announced. [Notice that matrix $\mathbb{U}$ in Eqs. (A18) and (A21) are the same, as seen from Eq. (A22).] □

To summarize, the stoichiometric matrix can be row-reduced in echelon form as in Eq. (A12), with an invertible matrix $\mathbb{G}$ taking the form Eq. (A13) and whose inverse takes the explicit form Eq. (A17) provided the columns of conservation laws in Eq. (A13) are organized as in Eq. (A18). We now depict how these algebraic results can be translated in geometrical terms.

The interpretation of Fig. 3 of the matrix $\mathbb{G}$ for graphs can be generalized to hypergraphs, without relying on the notion of spanning tree. To do so, one defines an 'escape' protocol as follows. The $N-M$ dependent species are labeled as 'roots'. For each of the $M$ independent species, we place a unit charge in its corresponding node $i$, and we ask how many times each of the $M$ independent reactions have to be applied (possibly a fractional and/or negative number of times) in order to expel completely the charge from $i$ through the set of roots. How each reaction acts on the charges is governed by the stoichiometry of Eq. (3). Because the matrix $\mathbb{S}_M^{\top\prime}$ precisely represents the action of the independent reactions on the independent species, we see that Eq. (A16), rewritten as

$$\mathbb{G}_M^{\top\prime}\mathbb{S}_M^{\top\prime} = \mathbb{1}_M\,, \tag{A24}$$

tells that the *line* entries of $\mathbb{G}_M^{\top\prime}$ precisely solve the escape problem. Indeed, since $\sum_k (\mathbb{G}_M^{\top\prime})_{ik}(\mathbb{S}_M^{\top\prime})_{kj} = \delta_{ij}$ we see that applying each independent reaction $k$ (with $1 \le k \le M$) a number $(\mathbb{G}_M^{\top\prime})_{ik}$ of times will expel a unit charge from the node $i$, while leaving empty the rest of the nodes $\notin \{\text{root}_\beta\}$ (see Fig. 9). We thus see how the algebraically trivial passage from Eq. (A16) to Eq. (A24) allows one to build a geometrical interpretation of the lines of the row-reducing matrix $\mathbb{G}^\top$. In Sec. III C 3, we explain how this leads to a generalization of the notion of cocycle from graphs to hypergraphs.

There also exists a geometrical interpretation of Eq. (A16) itself. Here instead, one places a unit force on a cochord (i.e. independent reaction) $\gamma$ and one asks which charge have to be set on the independent nodes (which are not roots) so as to produce this force. The solution is now given by the *columns* of $\mathbb{G}_M^{\top\prime}$. Indeed, we read from $\sum_k (\mathbb{S}_M^{\top\prime})_{\gamma k}(\mathbb{G}_M^{\top\prime})_{k\gamma'} = \delta_{\gamma\gamma'}$ that column $\gamma$ of $\mathbb{G}_M^{\top\prime}$ gives the set of charges on the set of independent species $k$ that generates a unit force on the cochord $\gamma$ (and 0 on the other cochords $\gamma' \neq \gamma$). Notice that on the chords, the generated force is not necessarily equal to 0 (it is in fact given by the entries of $\mathbb{T}$ which is used to define cocycles, see Sec. III C 3). For every $\gamma$, this set of charges can be seen as the 'elevation map' of an 'island' associated to $\boldsymbol{c}^\gamma$. Such elevation map is a potential landscape that generates a force given by the entries of cocycle $\boldsymbol{c}^\gamma$ (see Fig. 10 for an example).

As a last remark, we explain how, following an argument similar to that leading to the form of Eq. (A18) for the conservation laws, one can find a basis of the chemical cycles (that is, of $\operatorname{Ker}\mathbb{S}$) such that

$$\begin{pmatrix} \text{cycles} \end{pmatrix} = \begin{pmatrix} -\mathbb{T} \\ \hline \mathbb{1}_{R-M} \end{pmatrix}, \tag{A25}$$

where $\mathbb{T}$ is the matrix that expresses the dependency $\mathbb{S}_{\text{dep}} = \mathbb{S}_M\mathbb{T}$ of the dependent reactions of $\mathbb{S}_{\text{dep}}$ as a function of the independent ones of $\mathbb{S}_M$ (see the decomposition in Eq. (A15)).
*Proof*: We start from a basis of $\operatorname{Ker}\mathbb{S}$ composed of $R - M$ column vectors written as

$$\begin{pmatrix} \mathfrak{C}_2 \\ \hline \mathfrak{C}_1 \end{pmatrix} \begin{matrix} \updownarrow M \\ \updownarrow R-M \end{matrix}. \tag{A26}$$

The definition of cycles tells that $\mathbb{S}_M(\mathbb{T}\mathfrak{C}_1 + \mathfrak{C}_2) = 0$. The same *ad absurdum* argument as above tells that $\mathfrak{C}_1$ is invertible so that we can multiply the basis (A26) on the right by $\mathfrak{C}_1^{-1}$, and still keep a matrix whose columns are a basis for the cycles. It takes the announced form in Eq. (A25) since the relation above implies $\mathbb{T}\mathfrak{C}_1 + \mathfrak{C}_2 = 0$, as the columns of $\mathbb{S}_M$ are independent □

This proof is the algebraic counterpart, for generic CRNs, of the property that the same matrix $\mathbb{T}$ controls the dependency between reactions and the organization of cycles – a property we also obtained in Sec. II B from graph theory for unimolecular reactions.

## Appendix B: Reversibility and the Wegscheider–Kolmogorov condition

Consider the set of $R$ complex reactions (3) between $N$ species $\boldsymbol{X}$, described by the $N \times R$ stoichiometric matrix $\mathbb{S}$. These reactions describe, at the microscopic level, a stochastic population process on the numbers $\boldsymbol{n}$ of each chemical species $\boldsymbol{X}$; and at the macroscopic level, a continuity Equation (6) for the evolution of the concentrations $\boldsymbol{x}(t)$, which involves the current $\boldsymbol{J}$ of Eq. (5), expressed in terms of the affinities defined in Eq. (7). In this Appendix, we present in a unified manner the equivalence between the so-called Wegscheider condition [49] [equivalent to Kolmogorov's criterion [46] in the language of Markov chains (see e.g. [94])] and varied notions of reversibility, both at the microscopic population level and the macroscopic concentrations level.

We stress that the rate constants $k_\rho^\pm$ of the reactions (3) are *macroscopic* in the sense that they enter in the deterministic description Eq. (6) of the real-valued concentrations $\boldsymbol{x}$, and do not depend on the system's size. They differ from the microscopic rates of the individual reactions which, at the molecular level, scale with the system's size as

$$\kappa_\rho^\pm = \Omega \frac{k_\rho^\pm}{\boldsymbol{\Omega}^{\nu^{\pm\rho}}} \qquad \forall \rho, \tag{B1}$$

where $\Omega$ is the system's volume and $\boldsymbol{\Omega}^{\nu^{\pm\rho}} = \Omega^{\sum_i \nu_i^{\pm\rho}}$. This corresponds to the fact that, at fixed number of molecules, reactions involving (the collision of) several species are rarer as $\Omega$ gets larger (see e.g. §7.5.3 in [95]). Notice that when we discuss the population dynamics stochastic process, the vector $\boldsymbol{x}$ denotes the rational-valued vector $\boldsymbol{n}/\Omega$ representing the discrete concentrations of species. See also Appendix C for a discussion on the large-$\Omega$ asymptotics.

We recall (see Sec. III) that we can fix a basis of $R - M$ cycles $\boldsymbol{c}^\alpha$ which span the right nullspace $\operatorname{Ker}\mathbb{S}$ of $\mathbb{S}$, of dimension $R - M$. The following properties are equivalent:

I. Wegscheider's condition:

$$\forall \alpha, \ \prod_\rho \left(\frac{k_\rho^+}{k_\rho^-}\right)^{c_\rho^\alpha} = 1\,, \tag{B2}$$

i.e. the product of macroscopic transition rates of every cycle is the same in both directions along the cycle.

I'. Kolmogorov's condition:

$$\forall \alpha, \ \prod_\rho \left(\frac{\kappa_\rho^+}{\kappa_\rho^-}\right)^{c_\rho^\alpha} = 1\,, \tag{B3}$$

i.e. the product of microscopic transition rates of every cycle is the same in both directions along the cycle.

II. Existence of the standard chemical potential $\boldsymbol{\mu}^{\ominus}$:

$$\exists\, \boldsymbol{\mu}^{\ominus}: \quad \forall \rho,\ \frac{k_\rho^+}{k_\rho^-} = \exp\left[-\left(\mathbb{S}^\top \boldsymbol{\mu}^{\ominus}\right)_\rho\right]. \tag{B4}$$

For noninteracting CRNs, this is detailed balance, see Eq. (33).

III. Existence of concentrations canceling affinities:

$$\exists\, \boldsymbol{x}^{\mathrm{eq}}: \quad \forall \rho,\ A_\rho(\boldsymbol{x}^{\mathrm{eq}}) = 0\,. \tag{B5}$$

IV. Existence of concentrations canceling currents:

$$\exists\, \boldsymbol{x}^{\mathrm{eq}}: \quad \forall \rho,\ J_\rho(\boldsymbol{x}^{\mathrm{eq}}) = 0\,. \tag{B6}$$

V. Reversible constrained product Poisson law at the population level:

$$\exists\, \boldsymbol{x}^{\mathrm{eq}}: \quad |\mathcal{P}^{\mathrm{eq}}\rangle \propto \sum_{\boldsymbol{n}} \frac{(\Omega \boldsymbol{x}^{\mathrm{eq}})^{\boldsymbol{n}}}{\boldsymbol{n}!}\, \delta(\boldsymbol{\ell}(\boldsymbol{n}) - \boldsymbol{\mathcal{L}})\, |\boldsymbol{n}\rangle \tag{B7}$$

is an equilibrium steady state of the microscopic dynamics of occupation numbers. We use the Doi–Peliti ket notation $|\cdot\rangle$ for occupation states (see the proof, Sec. B). The components of $\boldsymbol{\ell}(\boldsymbol{n})$ are the conserved quantities, their values being the components of $\boldsymbol{\mathcal{L}}$ (fixed by initial condition). Vector notations are used ($\boldsymbol{n}! = \prod_{i=1}^N n_i!$, etc.).

VI. Microscopic reversibility: the stochastic dynamics of occupation numbers $\boldsymbol{n}$ verifies detailed balance.

VII. Gradient condition on affinities:

$$\forall t,\ \boldsymbol{A}(\boldsymbol{x}(t)) \in \operatorname{Im} \mathbb{S}^\top. \tag{B8}$$

*Remark*: in the proofs, we will often make use of the following writings of the affinity of a reaction $\rho$, that comes from Eq. (7)

$$A_\rho(\boldsymbol{x}) = \log \frac{k_\rho^+}{k_\rho^-} - \overbrace{\left(\mathbb{S}^\top \log \boldsymbol{x}\right)_\rho}^{\in \operatorname{Im} \mathbb{S}^\top}. \tag{B9}$$

### 1. Proof of I⇔I'

For every cycle $\boldsymbol{c}^\alpha$, we have

$$\prod_\rho \left(\frac{\kappa_\rho^+}{\kappa_\rho^-}\right)^{c_\rho^\alpha} = \prod_\rho \left(\frac{k_\rho^+}{k_\rho^-}\right)^{c_\rho^\alpha} \boldsymbol{\Omega}^{(\boldsymbol{\nu}^{-\rho} - \boldsymbol{\nu}^{+\rho})\, c_\rho^\alpha} = \prod_\rho \left(\frac{k_\rho^+}{k_\rho^-}\right)^{c_\rho^\alpha} \tag{B10}$$

since $\prod_\rho \boldsymbol{\Omega}^{(\boldsymbol{\nu}^{-\rho} - \boldsymbol{\nu}^{+\rho})\, c_\rho^\alpha} = \Omega^{\sum_{i\rho} \mathbb{S}_{i\rho} c_\rho^\alpha} = \Omega^{\sum_i (\mathbb{S}\boldsymbol{c}^\alpha)_i}$ and $\mathbb{S}\boldsymbol{c}^\alpha = 0$ by definition. The conditions (B2) and (B3) are thus the same □

### 2. Proof of I⇔II

If II holds, then for any cycle $\boldsymbol{c}^\alpha$, since $\mathbb{S}\,\boldsymbol{c}^\alpha = 0$, one has

$$\prod_\rho \left(\frac{k_\rho^+}{k_\rho^-}\right)^{c_\rho^\alpha} = \exp\left(-\boldsymbol{c}^\alpha \cdot\ \mathbb{S}^\top \boldsymbol{\mu}^{\ominus}\right) = \exp\left(-\boldsymbol{\mu}^{\ominus} \cdot\ \mathbb{S}\boldsymbol{c}^\alpha\right) = 1 \tag{B11}$$

which yields I. Conversely, if I holds, for any cycle $\boldsymbol{c} \in \operatorname{Ker} \mathbb{S}$ we have

$$\boldsymbol{A} \cdot \boldsymbol{c} = \underbrace{\log \prod_\rho \left(\frac{k_\rho^+}{k_\rho^-}\right)^{c_\rho}}_{=0 \text{ from Eq. (B2)}} + \underbrace{\log \boldsymbol{x}^{-\mathbb{S}\boldsymbol{c}}}_{=0} = 0\,, \tag{B12}$$

which implies that $\boldsymbol{A} \in (\operatorname{Ker} \mathbb{S})^\perp = \operatorname{Im} \mathbb{S}^\top$; combining then with Eq. (B9) we obtain $\log \frac{\boldsymbol{k}^+}{\boldsymbol{k}^-} \in \operatorname{Im} \mathbb{S}^\top$, which is precisely Eq. (B4) □

In practice, if the rates verify Wegscheider's condition (B2), identifying a standard corresponding chemical potential $\boldsymbol{\mu}^{\ominus}$ can be done using the hypergraph integration procedure described in Sec. III C 2.

### 3. Proof of II⇔III

If II holds, then

$$\forall \rho,\ A_\rho(\boldsymbol{x}^{\mathrm{eq}}) = 0$$

$$\Leftrightarrow\ \forall \rho,\ \left(\mathbb{S}^\top \log \boldsymbol{x}\right)_\rho = \log \frac{k_\rho^+}{k_\rho^-} \overset{\text{(B4)}}{=} -\left(\mathbb{S}^\top \boldsymbol{\mu}^{\ominus}\right)_\rho \tag{B13}$$

$$\Leftrightarrow\ \boldsymbol{\mu}^{\ominus} - \log \boldsymbol{x}^{\mathrm{eq}} \in \operatorname{Ker} \mathbb{S}^\top \tag{B14}$$

but $\operatorname{Ker} \mathbb{S}^\top$ is never an empty set, so that we can find $\boldsymbol{x}^{\mathrm{eq}}$ canceling all affinities, which is III. Conversely, if III holds, Eq. (B9) implies that $\log \frac{\boldsymbol{k}^+}{\boldsymbol{k}^-} \in \operatorname{Im} \mathbb{S}^\top$, which is precisely II □

### 4. Proof of III⇔IV

It is obvious from the expression (5) of the currents as function of the affinities. Notice that interestingly, this means that for complex CRNs, stochastic reversibility is equivalent to the existence of a fixed point with zero macroscopic current for the deterministic dynamics of Eq. (6).

### 5. Proof of III⇔V

We use Fock space notations for occupation vectors $|\boldsymbol{n}\rangle$ and the Doi–Peliti operators [96, 97] to represent the reactions at the microscopic level of occupation numbers (see [98–100] for reviews). We attach an annihilation operator $a_i$ and a creation one $a_i^\dagger$ to every species $i$. For a single species, they act as $a|n\rangle = n|n-1\rangle$, $a^\dagger|n\rangle = |n+1\rangle$, while for several species they only act on their attached species. The number operator $\hat{n}_i = a_i^\dagger a_i$ is diagonal and $\hat{n}_i|\boldsymbol{n}\rangle = n_i|\boldsymbol{n}\rangle$. The action of the creation/annihilation operators on (arbitrarily normalized) unconstrained Poisson laws is well known and easily checked:

$$a_i \sum_{\boldsymbol{n}} \frac{\boldsymbol{x}^{\boldsymbol{n}}}{\boldsymbol{n}!}|\boldsymbol{n}\rangle = x_i \sum_{\boldsymbol{n}} \frac{\boldsymbol{x}^{\boldsymbol{n}}}{\boldsymbol{n}!}|\boldsymbol{n}\rangle \tag{B15}$$

$$a_i^\dagger \sum_{\boldsymbol{n}} \frac{\boldsymbol{x}^{\boldsymbol{n}}}{\boldsymbol{n}!}|\boldsymbol{n}\rangle = \frac{\hat{n}_i}{x_i} \sum_{\boldsymbol{n}} \frac{\boldsymbol{x}^{\boldsymbol{n}}}{\boldsymbol{n}!}|\boldsymbol{n}\rangle\,. \tag{B16}$$

When constraints are present inside the Poisson law as on the r.h.s. of Eq. (B7), similar replacement rules $a_i \mapsto x_i$ and $a_i^\dagger \mapsto \frac{\hat{n}_i}{x_i}$ hold as in Eqs. (B15)-(B16), provided the operators on the l.h.s. leave the conserved quantities $\boldsymbol{\ell}(\boldsymbol{n})$ unchanged.

In the Doi–Peliti approach, the Markov dynamics in the population space is represented as a linear operator $\mathbb{W}$ acting on the probability vector $|P(t)\rangle = \sum_{\boldsymbol{n}} P(\boldsymbol{n},t)|\boldsymbol{n}\rangle$. For the $R$ reactions of the form (3), we decompose $\mathbb{W} = \sum_\rho \mathbb{W}_\rho$ with

$$\mathbb{W}_\rho = \kappa_\rho^+\Big[(\boldsymbol{a}^\dagger)^{\boldsymbol{\nu}^{-\rho}}\boldsymbol{a}^{\boldsymbol{\nu}^{+\rho}} - \hat{\boldsymbol{n}}^{\boldsymbol{\nu}^{+\rho}}\Big] + \kappa_\rho^-\Big[(\boldsymbol{a}^\dagger)^{\boldsymbol{\nu}^{+\rho}}\boldsymbol{a}^{\boldsymbol{\nu}^{-\rho}} - \hat{\boldsymbol{n}}^{\boldsymbol{\nu}^{-\rho}}\Big] \quad \text{(B17)}$$

where the microscopic rates $\kappa_\rho^\pm$ are defined in Eq. (B1). Every reaction respects the conservation laws of $\mathbb{S}$ so that one can apply the replacement rules mentioned above to compute the action of $\mathbb{W}_\rho$ on the vector $|\mathcal{P}^{\rm eq}\rangle$ defined in Eq. (B7). It's a matter of simple algebra, using the definition of affinity of Eqs. (5)-(7), to find

$$\begin{aligned}&\mathbb{W}_\rho|\mathcal{P}^{\rm eq}\rangle \quad \text{(B18)}\\ &= -\Omega\left[k_\rho^-\Big(1-\mathrm{e}^{A_\rho}\Big)\Big(\frac{\hat{\boldsymbol{n}}}{\Omega}\Big)^{\boldsymbol{\nu}^{-\rho}} + k_\rho^+\Big(1-\mathrm{e}^{-A_\rho}\Big)\Big(\frac{\hat{\boldsymbol{n}}}{\Omega}\Big)^{\boldsymbol{\nu}^{+\rho}}\right]|\mathcal{P}^{\rm eq}\rangle\end{aligned}$$

where $A_\rho = A_\rho(\boldsymbol{x}^{\rm eq})$. Let us now come to the proof of the equivalence III⇔V. If III holds, then we have the existence of a vector of concentrations $\boldsymbol{x}^{\rm eq}$ which cancels every affinity, see Eq. (B5). From the identity (B18), we find that $\mathbb{W}_\rho|\mathcal{P}^{\rm eq}\rangle = 0$, where $|\mathcal{P}^{\rm eq}\rangle$ defined in Eq. (B7) is evaluated on the $\boldsymbol{x}^{\rm eq}$ we just found (whose components are thus promoted from being average concentrations to being a parameters of a constrained product Poisson law). This proves that $|\mathcal{P}^{\rm eq}\rangle$ is a steady state of $\mathbb{W}$. To check explicitly that it verifies detailed balance, one introduces a diagonal operator $\hat{\mathcal{P}}^{\rm eq}$ whose components along the diagonal are those of the vector $|\mathcal{P}^{\rm eq}\rangle$ in Eq. (B7). Detailed balance is then equivalent to checking that $\mathbb{W}\hat{\mathcal{P}}^{\rm eq} = \hat{\mathcal{P}}^{\rm eq}\mathbb{W}^\top$. Using the identities $a_i\,\hat{\mathcal{P}}^{\rm eq} = \Omega x_i^{\rm eq}\,\hat{\mathcal{P}}^{\rm eq}(a_i^\dagger)^\top$ and $a_i^\dagger\,\hat{\mathcal{P}}^{\rm eq} = (\Omega x_i^{\rm eq})^{-1}\hat{\mathcal{P}}^{\rm eq}a_i^\top$ [11], one finds

$$\begin{aligned}&\mathbb{W}_\rho\hat{\mathcal{P}}^{\rm eq} - \hat{\mathcal{P}}^{\rm eq}\mathbb{W}_\rho^\top \quad \text{(B19)}\\ &= \hat{\mathcal{P}}^{\rm eq}\Big\{\kappa_\rho^-\Big(e^{A_\rho}-1\Big)(a^{\dagger\boldsymbol{\nu}^{+\rho}}a^{\boldsymbol{\nu}^{-\rho}})^\top + \kappa_\rho^+\Big(e^{-A_\rho}-1\Big)(a^{\dagger\boldsymbol{\nu}^{-\rho}}a^{\boldsymbol{\nu}^{+\rho}})^\top\Big\}\end{aligned}$$

where $A_\rho$ denotes $A_\rho(\boldsymbol{x}^{\rm eq})$. If III holds, then from Eq. (B5) we obtain $\mathbb{W}_\rho\hat{\mathcal{P}}_\rho^{\rm eq} = \hat{\mathcal{P}}^{\rm eq}\mathbb{W}_\rho^\top$ $(\forall\rho)$; hence, summing over $\rho$, detailed balance indeed holds. Conversely, if V holds, there exists a vector $\boldsymbol{x}^{\rm eq}$ such that $\mathbb{W}\hat{\mathcal{P}}^{\rm eq} = \hat{\mathcal{P}}^{\rm eq}\mathbb{W}^\top$ and Eq. (B19) yields

$$\begin{aligned}\sum_\rho \Big\{&\kappa_\rho^-\Big(e^{A_\rho(\boldsymbol{x}^{\rm eq})}-1\Big)a^{\dagger\boldsymbol{\nu}^{+\rho}}a^{\boldsymbol{\nu}^{-\rho}}\\ &+\kappa_\rho^+\Big(e^{-A_\rho(\boldsymbol{x}^{\rm eq})}-1\Big)a^{\dagger\boldsymbol{\nu}^{-\rho}}a^{\boldsymbol{\nu}^{+\rho}}\Big\} = 0\,, \quad \text{(B20)}\end{aligned}$$

since $\hat{\mathcal{P}}^{\rm eq}$ is an invertible operator. Consider now a given reaction $\rho$, applied in the direction where it transforms $|\boldsymbol{n}\rangle$ into $|\boldsymbol{n}-\boldsymbol{\nu}^{+\rho}+\boldsymbol{\nu}^{-\rho}\rangle$. Since by hypothesis reaction $\rho$ is the only one performing that transformation, we obtain by taking the scalar product of Eq. (B20) between $\langle\boldsymbol{n}-\boldsymbol{\nu}^{+\rho}+\boldsymbol{\nu}^{-\rho}|$ and $|\boldsymbol{n}\rangle$ that $e^{-A_\rho(\boldsymbol{x}^{\rm eq})}-1=0$. We thus see that necessarily $A_\rho(\boldsymbol{x}^{\rm eq})=0\,(\forall\rho)$ which is precisely III □

[11] These identities are verified by direct computation. We stress that in the canonical scalar product, the transpose operator does not allow one to switch between $a$ and $a^\dagger$. In fact from $\langle n|(a^\dagger)^\top = \langle n+1|$ and $\langle n|a^\top = \langle n-1|n$, one sees that $a^\top = \hat{n}\,a^\dagger$ and $(a^\dagger)^\top = a\frac{1}{\hat{n}}$, where $\hat{n}$ is the number operator $\hat{n}|n\rangle = n|n\rangle$.

### 6. Proof of V⇔VI

Obviously, V implies VI. Conversely, if the microscopic occupation-number dynamics verifies detailed balance, let us show that Kolmogorov's condition (B3) is verified, which will prove that I' and hence V holds (as we already showed). Consider a basis of cycles $\boldsymbol{c}^\alpha$ of the stoichiometric matrix $\mathbb{S}$. They can be taken to have (positive or negative) integer entries. Then, a given $\boldsymbol{c}^\alpha$ corresponds to a succession of reactions, where each reaction $\rho$ is used $\boldsymbol{c}_\rho^\alpha$ times. In fact, depending on the precise order in which these reactions are applied, the algebraic cycle $\boldsymbol{c}^\alpha$ corresponds to many possible cycles at the population level. We now consider a given $\boldsymbol{c}^\alpha$ and choose arbitrary such an ordering. It leaves any configuration of the occupations invariant (and the same is true if the cycle is applied in reverse order). Detailed balance at the level of occupations implies that the product of transition rates of the cycle and its reverse are the same, at the level of population rates. Using the Doi–Peliti formalism, we express such product of rates, starting from configuration $\boldsymbol{n}$, as follows:

$$\prod_\rho^{\leftarrow}(\kappa_\rho^+)^{c_\rho^\alpha}\langle\boldsymbol{n}|(\boldsymbol{a}^\dagger)^{c_\rho^\alpha\boldsymbol{\nu}^{-\rho}}\boldsymbol{a}^{c_\rho^\alpha\boldsymbol{\nu}^{+\rho}}|\boldsymbol{n}\rangle\,. \quad \text{(B21)}$$

The rate for reaction $\rho$ is raised to the power $c_\rho^\alpha$, and the arrow on the product sign indicates that the operators of the first reaction (in the considered ordering) are placed right of the ones of the next reaction, down to the last reaction involved. For the reversed reaction, the product of rates is

$$\prod_\rho^{\rightarrow}(\kappa_\rho^-)^{c_\rho^\alpha}\langle\boldsymbol{n}|(\boldsymbol{a}^\dagger)^{c_\rho^\alpha\boldsymbol{\nu}^{+\rho}}\boldsymbol{a}^{c_\rho^\alpha\boldsymbol{\nu}^{-\rho}}|\boldsymbol{n}\rangle\,. \quad \text{(B22)}$$

Noticing now the identity $\langle\boldsymbol{n}|(\boldsymbol{a}^\dagger)^{c_\rho^\alpha\boldsymbol{\nu}^{-\rho}}\boldsymbol{a}^{c_\rho^\alpha\boldsymbol{\nu}^{+\rho}}|\boldsymbol{n}\rangle = \langle\boldsymbol{n}|(\boldsymbol{a}^\dagger)^{c_\rho^\alpha\boldsymbol{\nu}^{+\rho}}\boldsymbol{a}^{c_\rho^\alpha\boldsymbol{\nu}^{-\rho}}|\boldsymbol{n}\rangle$ [12], the equality of Eqs. (B21) and (B22) yields that for all $\alpha$, $\prod_\rho(\kappa_\rho^+)^{c_\rho^\alpha} = \prod_\rho(\kappa_\rho^-)^{c_\rho^\alpha}$ which is the announced Eq. (B3) □

[12] This identity is shown by taking the transpose of $(\boldsymbol{a}^\dagger)^{c_\rho^\alpha\boldsymbol{\nu}^{-\rho}}\boldsymbol{a}^{c_\rho^\alpha\boldsymbol{\nu}^{+\rho}}$ and performing the similarity transformations $(\hat{\boldsymbol{n}}!)^{-1}a_i\hat{\boldsymbol{n}}! = a_i\hat{n}_i$ and $(\hat{\boldsymbol{n}}!)^{-1}a_i^\dagger\hat{\boldsymbol{n}}! = (\hat{n}_i)^{-1}a_i^\dagger$ for every species $i$ involved.

### 7. Proof of I⇔VII

Every implication in I⇒VII⇒II is immediate, using the identity (B9). Then we already showed that II⇒I; so that finally we have both I⇒VII and VII⇒I □

Notice that, interestingly, in the implication VII⇒II inferred from Eq. (B9), we deduce a property valid independently of

$\boldsymbol{x}(t)$ (namely, $\log \frac{k^+}{k^-} \in \mathrm{Im}\,\mathbb{S}^\top$) from a property depending on $\boldsymbol{x}(t)$ (namely, $\boldsymbol{A}(\boldsymbol{x}(t)) \in \mathrm{Im}\,\mathbb{S}^\top$).

## Appendix C: Effective Fokker–Planck and Langevin dynamics close to an equilibrium point

Consider an arbitrary function $f(\boldsymbol{n})$ of the population state, i.e. the number of particles $\boldsymbol{n}$ for each chemical species $\boldsymbol{X}$. The master equation on the probability distribution $P(\boldsymbol{n},t)$ in the population space, for the chemical reactions (3), is equivalent to the following evolution equation for the average $\langle f\rangle = \sum_{\boldsymbol{n}} P(\boldsymbol{n},t) f(\boldsymbol{n})$:

$$\partial_t\langle f\rangle = \sum_{\boldsymbol{n},\rho} f(\boldsymbol{n})\Big\{W_\rho^+(\boldsymbol{n}-\mathbb{S}_\rho)P(\boldsymbol{n}-\mathbb{S}_\rho,t) - W_\rho^+(\boldsymbol{n})P(\boldsymbol{n},t) + W_\rho^-(\boldsymbol{n}+\mathbb{S}_\rho)P(\boldsymbol{n}+\mathbb{S}_\rho,t) - W_\rho^-(\boldsymbol{n})P(\boldsymbol{n},t)\Big\} \tag{C1}$$

$$= \sum_{\boldsymbol{n},\rho} P(\boldsymbol{n},t)\Big\{\Big[f(\boldsymbol{n}+\mathbb{S}_\rho)-f(\boldsymbol{n})\Big]W_\rho^+(\boldsymbol{n}) + \Big[f(\boldsymbol{n}-\mathbb{S}_\rho)-f(\boldsymbol{n})\Big]W_\rho^-(\boldsymbol{n})\Big\} \tag{C2}$$

where $\mathbb{S}_\rho$ designates the column vector of $\mathbb{S}$ describing reaction $\rho$, and $W_\rho^\pm(\boldsymbol{n}) = W(\{n_i \mapsto n_i \pm \mathbb{S}_{i\rho}\})$ are the transition rates at the species population level. In full generality, the transition rates are given by the product of the 'molecular' reaction rates $\kappa_\rho^\pm$ and the number of reactants $\boldsymbol{n}^{[\nu^{\pm\rho}]} = \boldsymbol{n}!/(\boldsymbol{n}-\nu^{\pm\rho})!$, so that

$$W_\rho^\pm(\boldsymbol{n}) = \kappa_\rho^\pm\, \boldsymbol{n}^{[\nu^{\pm\rho}]} = \Omega\, k_\rho^\pm \frac{\boldsymbol{n}^{[\nu^{\pm\rho}]}}{\boldsymbol{\Omega}^{\nu^{\pm\rho}}}\,, \tag{C3}$$

where in the second equality we used Eq. (B1) to make the dependence of the molecular rates on system size explicit. Notably, the extensivity of the rates $\kappa_\rho^\pm$ depends on the stoichiometry of the corresponding reactions $\nu^{\pm\rho}$. This corresponds to the fact that collisions between particles, which are required for multiple-species reactions to occur, get rarer when $\Omega$ increases at fixed $\boldsymbol{n}$. Intuitively, this is even more so when the number of involved species is larger (see e.g. [15, 95, 101, 102]). The relevance of such scaling is for instance seen as follows: Using these rates in Eq. (C2) for $f(\boldsymbol{n}) = n_i$, one recovers the macroscopic rate Equation (6) as $\Omega \to \infty$, with $\boldsymbol{x} = \boldsymbol{n}/\Omega$ fixed (in the large-$\Omega$ limit where the average of concentrations product becomes the products of the average).

In the large-size asymptotics $\Omega \to \infty$, we expand Eq. (C2) for the rescaled function of the concentrations $\bar f(\boldsymbol{x}) = f(\Omega\boldsymbol{x})$ and for the probability density $\bar P(\boldsymbol{x},t)$, and define $\bar W_\rho^+(\boldsymbol{x}) = k_\rho^\pm \boldsymbol{x}^{\nu^{\pm\rho}}$, to get

$$\partial_t\langle \bar f\rangle = \int \mathrm{d}^N\boldsymbol{x} \sum_\rho \bar P(\boldsymbol{x},t)\Big\{\Big[\bar W_\rho^+(\boldsymbol{x}) - \bar W_\rho^-(\boldsymbol{x})\Big]\sum_i \mathbb{S}_{i\rho}\partial_i \bar f(\boldsymbol{x}) + \frac{1}{2\Omega}\Big[\bar W_\rho^+(\boldsymbol{x}) + \bar W_\rho^-(\boldsymbol{x})\Big]\sum_{i,j}\mathbb{S}_{i\rho}\mathbb{S}_{j\rho}\partial_{ij}\bar f(\boldsymbol{x})\Big\}, \tag{C4}$$

where we neglected terms of order $\Omega^{-2}$ and higher. We recognize the first square bracket to be the macroscopic current $J_\rho(\boldsymbol{x})$, see Eq. (5). The coefficient of $\partial_{ij}\bar f$ is proportional to the symmetric matrix $\mathbb{D}(\boldsymbol{x})$ of components

$$\mathbb{D}_{ij}(\boldsymbol{x}) = \sum_\rho \frac{1}{2}\mathbb{S}_{i\rho}(\bar W_\rho^+(\boldsymbol{x}) + \bar W_\rho^-(\boldsymbol{x}))(\mathbb{S}^\top)_{\rho j} \tag{C5}$$

so that, overall, Eq. (C4) becomes

$$\partial_t\langle \bar f\rangle = \Big\langle \sum_i (\mathbb{S}\boldsymbol{J}(\boldsymbol{x}))_i \partial_i \bar f(\boldsymbol{x}) + \frac{1}{\Omega}\sum_{ij}\mathbb{D}_{ij}(\boldsymbol{x})\partial_{ij}\bar f(\boldsymbol{x})\Big\rangle. \tag{C6}$$

Formally, the evolution equation (C6) for the average of $\bar f(\boldsymbol{x})$ is the same as that of a Fokker–Planck equation corresponding to the Langevin equation

$$\partial_t \boldsymbol{x}(t) = \mathbb{S}\boldsymbol{J}(\boldsymbol{x}(t)) + \boldsymbol{\eta}(\boldsymbol{x}(t),t) \tag{C7}$$

with $\boldsymbol{\eta}(\boldsymbol{x},t)$ a Gaussian white noise of zero average and covariance $\langle \eta_i(\boldsymbol{x},t)\eta_j(\boldsymbol{x},t')\rangle = \frac{1}{\Omega}\mathbb{D}_{ij}(\boldsymbol{x})\,\delta(t'-t)$ (notice that the time discretization of such multiplicative noise has no importance in the small-noise regime $\Omega \to \infty$ we are considering). However, such a formal treatment has the problem in that it discards possible scaling with $\Omega$ of the probability density $\bar P(\boldsymbol{x},t)$ itself (and consistently of $\bar f(\boldsymbol{x})$), which would invalidate the large-$\Omega$ expansion and truncation. This was noticed within a large variety of contexts in the literature [72, 95, 103–105]. A regime where the above expansion is necessarily valid is that of $\boldsymbol{x}(t)$ close to a stationary point $\boldsymbol{x}^*$, i.e. $\boldsymbol{x}(t) = \boldsymbol{x}^* + \delta\boldsymbol{x}(t)$ with $\mathbb{S}\boldsymbol{J}(\boldsymbol{x}^*) = 0$ and $\delta\boldsymbol{x}(t) = O(\Omega^{-1/2})$. Then, the Langevin equation (C7) reduces to:

$$\partial_t \delta\boldsymbol{x}(t) = \mathbb{S}\boldsymbol{J}(\boldsymbol{x}^* + \delta\boldsymbol{x}(t)) + \boldsymbol{\eta}(t)\,, \tag{C8}$$

where $J(\boldsymbol{x}^* + \delta\boldsymbol{x}(t))$ is understood as truncated to first order in $\delta\boldsymbol{x}(t)$ (i.e. the Langevin equation is linear) and the centered Gaussian noise $\boldsymbol{\eta}(t)$ is now additive with correlations

$$\langle \eta_i(t)\eta_j(t')\rangle = \frac{1}{\Omega}\mathbb{D}^*_{ij}\,\delta(t'-t) \tag{C9}$$

where the matrix $\mathbb{D}^*$ is obtained from Eq. (C5) and reads

$$\mathbb{D}^* = \mathbb{D}(\boldsymbol{x}^*) = \mathbb{S}\mathbb{\Lambda}^*\mathbb{S}^\top. \tag{C10}$$

Here, $\mathbb{\Lambda}^*$ the $R\times R$ diagonal matrix with the entries of the vector $\frac{1}{2}(k_\rho^+ \boldsymbol{x}^{*\nu^{+\rho}} + k_\rho^- \boldsymbol{x}^{*\nu^{-\rho}})$.

In general, in irreversible dynamics, the drift of this Langevin equation is not simply related to the noise covariance matrix $\mathbb{D}^*$ of Eq. (C10). Focusing now on conservative affinities as in Sec. III E, the dynamics is reversible and there exists an equilibrium stationary point $\boldsymbol{x}^* = \boldsymbol{x}^{\mathrm{eq}}$ (see Appendix B) that cancels the current and the affinity vectors. Then using Eqs. (63) and (67) and remarking that $\mathbb{\Lambda}^* = \mathbb{\Lambda}$, one has $\boldsymbol{J}(\boldsymbol{x}^{\mathrm{eq}} + \delta\boldsymbol{x}) = \mathbb{\Lambda}\boldsymbol{A}(\boldsymbol{x}^{\mathrm{eq}} + \delta\boldsymbol{x}) = -\mathbb{\Lambda}\mathbb{S}^\top(\mathbb{X}^{\mathrm{eq}})^{-1}\delta\boldsymbol{x}$ and thus from Eq. (C8)

$$\partial_t \delta\boldsymbol{x}(t) = -\mathbb{D}\,(\mathbb{X}^{\mathrm{eq}})^{-1}\delta\boldsymbol{x}(t) + \boldsymbol{\eta}(t) \tag{C11}$$

with $\mathbb{D} = \mathbb{D}^* = \mathbb{D}(\boldsymbol{x}^{\mathrm{eq}})$. Hence, the symmetric matrix $\mathbb{D} = \mathbb{S}\mathbb{\Lambda}\mathbb{S}^\top$ read from Eq. (C10) plays at the same time the role of

the noise amplitude and the prefactor of the potential gradient in the Langevin equation (C11), which is an incarnation of the Onsager reciprocity [47, 71].

We now connect the previous analysis to the core of the paper. We first note that the rank of $\mathbb{D}$ is $M$ and not $N$. This means that in general some directions of the noise present a zero amplitude. This corresponds to the fact that the degrees of freedom $\boldsymbol{x}(t)$ representing the instantaneous concentrations at time $t$ present one or several conservation law(s), both at the deterministic level the rate equation (3) and at the stochastic level. In Sec. III E, we identified $M$ independent degrees of freedom $\delta\boldsymbol{z}(t)$, defined in Eq. (68), at the deterministic level. Using the same procedure at the stochastic level, we define a stochastic process $\delta\boldsymbol{z}(t)$ from $\delta\boldsymbol{x}(t)$ [that satisfies Eq. (C11)]. Now, the noise that governs the evolution of $\delta\boldsymbol{z}(t)$ is non-singular. Indeed, multiplying Eq. (C11) by $\mathbb{G}$, using Eq. (69) and taking the last $M$ components, we find by direct computation:

$$\partial_t \delta\boldsymbol{z}(t) = -\mathbb{L}_Q\, \mathbb{H}_Q\, \delta\boldsymbol{z}(t) + \tilde{\boldsymbol{\eta}}(t)\,, \tag{C12}$$

where the (now non-singular) centered Gaussian white noise $\tilde{\boldsymbol{\eta}}(t) \in \mathbf{R}^M$ has correlations $\langle \tilde{\eta}_i(t)\tilde{\eta}_j(t')\rangle = \frac{1}{\Omega}(\mathbb{L}_Q)_{ij}\delta(t'-t)$. In these expressions, $\mathbb{L}_Q$ and $\mathbb{H}_Q$ are the $M\times M$ matrix defined in Eq. (64) and (70) respectively. As expected, the deterministic drift of Eq. (C12) is the same as the one derived at the deterministic level [see Eq. (71)]. At the stochastic level, for $\delta\boldsymbol{z}(t)$, the matrix $\mathbb{L}_Q$ plays at the same time the role of a relaxation response matrix close to an equilibrium point and the correlation matrix of the noise that describes the small Gaussian fluctuations close to that point. Accordingly, to the linearized Langevin equation (C12) one can associate the Gaussian stationary probability density $\bar{P}(\delta\boldsymbol{z}) \propto \exp\left[-\frac{\Omega}{2}\delta\boldsymbol{z}^\top \mathbb{H}_Q\, \delta\boldsymbol{z}\right]$, where we thus identify the matrix $\mathbb{H}_Q$ as the Hessian matrix of the equilibrium quasipotential. This concludes our illustration that the two matrices $\mathbb{L}_Q$ and $\mathbb{H}_Q$ that appeared in Sec. III E in the analysis of the deterministic relaxation close to an equilibrium point in fact also play a role at the Gaussian stochastic level.

## Appendix D: Cycles, cocycles and oblique projectors

In this Appendix we show how the decompositions in Eq. (19) and Eq. (26) can be reformulated in terms of complementary oblique projectors. We follow the line of [14] where the formalism was first introduced and discussed for graphs (corresponding to unimolecular reactions). From the families of cocycles and cycles introduced in the main text, we define two $R\times R$ matrices as:

$$\mathbb{Q}^\top = \left(\begin{array}{c|c} \overbrace{\mathbb{1}_M}^{\boldsymbol{c}^\gamma} & \mathbb{0} \\ \mathbb{T}^\top & \mathbb{0} \end{array}\right), \quad \mathbb{P} = \left(\begin{array}{c|c} \mathbb{0} & \overbrace{-\mathbb{T}}^{\boldsymbol{c}^\alpha} \\ \mathbb{0} & \mathbb{1}_{R-M} \end{array}\right) \tag{D1}$$

By construction, their images correspond to the spaces spanned by the $\boldsymbol{c}^\gamma$'s and the $\boldsymbol{c}^\alpha$'s, namely $\mathrm{Im}\,\mathbb{Q}^\top = \mathrm{Im}\,\mathbb{S}^\top$ and $\mathrm{Im}\,\mathbb{P} = \mathrm{Ker}\,\mathbb{S}$. Taking the transpose of Eq. (D1) one obtains two more matrices,

$$\mathbb{Q} = \left(\begin{array}{c|c} \overbrace{\mathbb{1}_M}^{\boldsymbol{e}^\gamma} & \mathbb{T} \\ \mathbb{0} & \mathbb{0} \end{array}\right), \quad \mathbb{P}^\top = \left(\begin{array}{c|c} \mathbb{0} & \overbrace{\mathbb{0}}^{\boldsymbol{e}^\alpha} \\ -\mathbb{T}^\top & \mathbb{1}_{R-M} \end{array}\right) \tag{D2}$$

whose images are now spanned by the $\boldsymbol{e}^\gamma$'s and $\boldsymbol{e}^\alpha$'s, i.e. $\mathrm{Im}\,\mathbb{Q} = \mathrm{Span}\,(\boldsymbol{e}^\gamma)$ and $\mathrm{Im}\,\mathbb{P}^\top = \mathrm{Span}\,(\boldsymbol{e}^\alpha)$.

We recall that a square matrix $\mathbb{A}$ is a projector if and only if it is idempotent $\mathbb{A}^2 = \mathbb{A}$. It can be directly checked that this property holds for $\mathbb{Q}^\top$ and $\mathbb{P}$, as well as for $\mathbb{Q}$ and $\mathbb{P}^\top$ making them oblique projectors with $\mathbb{Q}^\top \neq \mathbb{Q}$ and $\mathbb{P} \neq \mathbb{P}^\top$ as soon as $\mathbb{T}$ is present. In particular, they form pairs of complementary oblique projectors such that $\mathbb{P} + \mathbb{Q} = \mathbb{P}^\top + \mathbb{Q}^\top = \mathbb{1}_R$ and $\mathbb{Q}\mathbb{P} = \mathbb{Q}^\top\mathbb{P}^\top = 0$.

As a consequence, we may re-express the decompositions in the main text as:

$$\boldsymbol{A} = \sum_\gamma A^c_\gamma\, \boldsymbol{c}^\gamma + \sum_\alpha A^e_\alpha\, \boldsymbol{e}^\alpha = \mathbb{Q}^\top \boldsymbol{A} + \mathbb{P}^\top \boldsymbol{A} \tag{D3}$$

$$\boldsymbol{J} = \sum_\gamma J^e_\gamma\, \boldsymbol{e}^\gamma + \sum_\alpha J^c_\alpha\, \boldsymbol{c}^\alpha = \mathbb{Q}\boldsymbol{J} + \mathbb{P}\boldsymbol{J}\,. \tag{D4}$$

These expressions are analogous the ones reported in [14] with the main difference being that here the operators are not derived from the spanning tree of a graph but from the family of $\{\boldsymbol{c}^\gamma\}$ and $\{\boldsymbol{c}^\alpha\}$ constructed in Sec. III B using the reduced row echelon form of $\mathbb{S}$. Thus, the construction we put forward in Sec. III and in this Appendix generalizes the oblique projector method of Ref. [14] for the decomposition of currents and affinities from unimolecular CRNs (and graphs) to arbitrary CRNs (and their associated hypergraphs).

We conclude the section by pointing out a connection between the oblique projectors and the Onsager matrices of linear response (see Sec. III E). First, one can always define new projectors using a change of basis. In particular, we may define $\hat{\mathbb{Q}} = \mathbb{\Lambda}^{-1/2}\mathbb{Q}\mathbb{\Lambda}^{1/2}$ and $\hat{\mathbb{P}} = \mathbb{\Lambda}^{-1/2}\mathbb{P}\mathbb{\Lambda}^{1/2}$ which are still complementary oblique projectors. Then, one finds:

$$\mathbb{Q}\mathbb{\Lambda}\mathbb{Q}^\top = \left(\begin{array}{c|c} \mathbb{L}_Q & \mathbb{0} \\ \hline \mathbb{0} & \mathbb{0} \end{array}\right) \quad \hat{\mathbb{Q}}\hat{\mathbb{Q}}^\top = \left(\begin{array}{c|c} \hat{\mathbb{L}}_Q & \mathbb{0} \\ \hline \mathbb{0} & \mathbb{0} \end{array}\right) \tag{D5}$$

$$\mathbb{P}^\top\mathbb{\Lambda}^{-1}\mathbb{P} = \left(\begin{array}{c|c} \mathbb{0} & \mathbb{0} \\ \hline \mathbb{0} & \mathbb{L}_P \end{array}\right) \quad \hat{\mathbb{P}}^\top\hat{\mathbb{P}} = \left(\begin{array}{c|c} \mathbb{0} & \mathbb{0} \\ \hline \mathbb{0} & \hat{\mathbb{L}}_P \end{array}\right) \tag{D6}$$

Thus, in both representations, the Onsager matrices appear as the invertible cores of the symmetric $R\times R$ matrices constructed from the oblique projectors. In Ref. [14], the matrices in Eqs. (D5)-(D6) were shown to govern the different contributions to the entropy production in linear response; we thus have shown in Sec. III E that these matrix also control the macroscopic relation between currents and affinities in the linear-response regime for generic CRNs.

## Appendix E: Proof of reconstruction feasibility

We want to show that Eq. (101) is consistent with stationary KCL,

$$\mathbb{S}\boldsymbol{J} = \mathbb{S}\begin{pmatrix} \boldsymbol{J}_Y \\ \boldsymbol{J}_X \end{pmatrix} = 0\,, \tag{E1}$$

and with linear-regime KVL, obtained by plugging Eq. (63) into Eq. (21), namely

$$\mathbb{P}^\top \mathbb{A}^{-1} \boldsymbol{J}_X = 0\,, \tag{E2}$$

where we made use of the oblique projector $\mathbb{P}$ introduced in Appendix D. Splitting $\mathbb{S} = (\mathbb{S}_Y, \mathbb{S}_X)$ as external and internal reactions, find $\mathbb{T}$ by row reduction of $\mathbb{S}_X$, and consider

$$\mathbb{Q} = \begin{pmatrix} -\mathbb{G}_M \mathbb{S}_X \\ \mathbb{0} \end{pmatrix} = \begin{pmatrix} \mathbb{1}_M & \mathbb{T} \\ \mathbb{0} & \mathbb{0} \end{pmatrix} \tag{E3}$$

obtained by adding or removing sufficient zero rows to make it a square $R \times R$ matrix. The matrix $\mathbb{G}_M$ is defined in Eq. (50) in the main text. As explained in Appendix D, $\mathbb{Q}$ and $\mathbb{P}$ are complementary oblique projectors so that $\mathbb{P} = \mathbb{1} - \mathbb{Q}$ and $\mathbb{S}_X \mathbb{P} = \mathbb{0}$. Then solutions of the system Eqs. (E1)-(E2) can be found by exploiting the projector algebra. In particular expanding the identity

$$\boldsymbol{J}_X = \mathbb{P}\boldsymbol{J}_X + \mathbb{Q}\boldsymbol{J}_X \tag{E4}$$

we find

$$\mathbb{S}_X \mathbb{Q} \boldsymbol{J}_X = -\mathbb{S}_Y \boldsymbol{J}_Y \tag{E5}$$

$$\mathbb{P}^\top \mathbb{A}^{-1} \mathbb{P} \boldsymbol{J}_X + \mathbb{P}^\top \mathbb{A}^{-1} \mathbb{Q} \boldsymbol{J}_X = 0. \tag{E6}$$

where $\boldsymbol{J}_Y$ are the external currents. By applying the matrix $\mathbb{G}_M$ to the first, in view of Eq. (E3), we find

$$\mathbb{Q}\boldsymbol{J}_X = \begin{pmatrix} \mathbb{G}_M \mathbb{S}_Y \boldsymbol{J}_Y \\ \mathbb{0} \end{pmatrix}, \tag{E7}$$

where we used $\mathbb{Q}^2 = \mathbb{Q}$. Plugging this latter into the second we find:

$$\mathbb{P}^\top \mathbb{A}^{-1} \mathbb{P} \boldsymbol{J}_X = -\mathbb{P}^\top \mathbb{A}^{-1} \begin{pmatrix} \mathbb{G}_M \mathbb{S}_Y \boldsymbol{J}_Y \\ \mathbb{0} \end{pmatrix}. \tag{E8}$$

A solution $\boldsymbol{J}_X^+$ that is consistent with the above equation can be found by the Moore–Penrose pseudoinverse

$$\boldsymbol{J}_X^+ = -(\mathbb{P}^\top \mathbb{A}^{-1} \mathbb{P})^+ \mathbb{P}^\top \mathbb{A}^{-1} \begin{pmatrix} \mathbb{G}_M \mathbb{S}_Y \boldsymbol{J}_Y \\ \mathbb{0} \end{pmatrix}. \tag{E9}$$

Projecting once again $\mathbb{P}\boldsymbol{J}_X = \mathbb{P}\boldsymbol{J}_X^+$, using Eq. (E4), we finally find

$$\begin{aligned} \boldsymbol{J}_X &= \left[\mathbb{1}_M - \mathbb{P}(\mathbb{P}^\top \mathbb{A}^{-1} \mathbb{P})^+ \mathbb{P}^\top \mathbb{A}^{-1}\right] \begin{pmatrix} \mathbb{G}_M \mathbb{S}_Y \boldsymbol{J}_Y \\ \mathbb{0} \end{pmatrix} \\ &= \begin{pmatrix} (\mathbb{1}_M - \mathbb{T}\mathbb{L}_P^{-1} \mathbb{T}^\top \mathbb{A}_M^{-1}) \mathbb{G}_M \mathbb{S}_Y \boldsymbol{J}_Y \\ \mathbb{L}_P^{-1} \mathbb{T}^\top \mathbb{G}_M \mathbb{S}_Y \boldsymbol{J}_Y \end{pmatrix}, \end{aligned} \tag{E10}$$

where in the last expression we made explicit the projector-based solution in terms of known matrices. The fact that the above system is full-rank grants, *a posteriori*, that this solution is unique and correct.


[1] H. Risken, *The Fokker-Planck equation: methods of solution and applications*, 2nd ed., Springer series in synergetics (Springer-Verlag, New York, 1996).

[2] S. Scheidl, Mobility in a one-dimensional disorder potential, Zeitschrift für Physik B Condensed Matter **97**, 345 (1995).

[3] P. Le Doussal and V. M. Vinokur, Creep in one dimension and phenomenological theory of glass dynamics, Physica C: Superconductivity **254**, 63 (1995).

[4] S. Carnot, *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance.* (Bachelier, Paris, 1824).

[5] M. O. Magnasco, Molecular combustion motors, Phys. Rev. Lett. **72**, 2656 (1994).

[6] T. Hill, *Free Energy Transduction in Biology: the steady-state kinetic and Thermodynamic formalism.* (Elsevier Science, Saint Louis, 1977).

[7] B. L. Clarke, Stoichiometric network analysis, Cell Biophysics **12**, 237 (1988).

[8] T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics*, Vol. 1 (Springer, New York, NY, 1989).

[9] M. Feinberg, *Foundations of Chemical Reaction Network Theory*, Applied Mathematical Sciences (Springer International Publishing, 2019).

[10] R. Rao and M. Esposito, Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics, Phys. Rev. X **6**, 041064 (2016).

[11] D. D. Martino, F. Capuani, and A. D. Martino, Inferring metabolic phenotypes from the exometabolome through a thermodynamic variational principle, New Journal of Physics **16**, 115018 (2014).

[12] M. Polettini and M. Esposito, Irreversible thermodynamics of open chemical networks. I. emergent cycles and broken conservation laws, The Journal of Chemical Physics **141**, 024117 (2014).

[13] J. Schnakenberg, Network theory of microscopic and macroscopic behavior of master equation systems, Rev. Mod. Phys. **48**, 571 (1976).

[14] M. Polettini, Cycle/cocycle oblique projections on oriented graphs, Lett. Math.Phys. **105**, 86 (2015).

[15] D. Andrieux and P. Gaspard, Fluctuation theorem and onsager reciprocity relations, The Journal of Chemical Physics **121**, 6167 (2004).

[16] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Reports on Progress in Physics **75**, 126001 (2012).

[17] T. R. Gingrich, G. M. Rotskoff, and J. M. Horowitz, Inferring dissipation from current fluctuations, Journal of Physics A: Mathematical and Theoretical **50**, 184004 (2017).

[18] J. M. Horowitz and M. Esposito, Thermodynamics with continuous information flow, Phys. Rev. X **4**, 031015 (2014).

[19] E. Ilker, O. Güngör, B. Kuznets-Speck, J. Chiel, S. Deffner, and M. Hinczewski, Shortcuts in Stochastic Systems and Control of Biophysical Processes, Physical Review X **12**, 021048 (2022).

[20] N. Nakanishi, *Graph Theory and Feynman Integrals*, Mathematics and its applications : a series of monographs and texts (Gordon and Breach, 1971).

[21] N. Freitas, J.-C. Delvenne, and M. Esposito, Stochastic and quantum thermodynamics of driven rlc networks, Phys. Rev. X **10**, 031005 (2020).

[22] M. Feinberg, Chemical reaction network structure and the stability of complex isothermal reactors—I. the deficiency zero and deficiency one theorems, Chemical Engineering Science **42**, 2229 (1987).

[23] A. Blokhuis, D. Lacoste, and P. Nghe, Universal motifs and the diversity of autocatalytic systems, Proceedings of the National Academy of Sciences **117**, 25230 (2020).

[24] S. Klamt, U.-U. Haus, and F. Theis, Hypergraphs and cellular networks, PLOS Computational Biology **5**, 1 (2009).

[25] T. Carletti, D. Fanelli, and S. Nicoletti, Dynamical systems on hypergraphs, Journal of Physics: Complexity **1**, 035006 (2020).

[26] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Networks beyond pairwise interactions: Structure and dynamics, Physics Reports **874**, 1 (2020).

[27] S. Majhi, M. Perc, and D. Ghosh, Dynamics on higher-order networks: a review, Journal of The Royal Society Interface **19**, 20220043 (2022).

[28] J. O'Byrne, Non-Equilibrium Currents in Stochastic Field Theories: a Geometric Insight (2021), arXiv:2108.13535 [cond-mat.stat-mech].

[29] Y.-J. Yang and H. Qian, Bivectorial Nonequilibrium Thermodynamics: Cycle Affinity, Vorticity Potential, and Onsager's Principle, Journal of Statistical Physics **182**, 46 (2021).

[30] C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee, Current status and applications of genome-scale metabolic models, Genome biology **20**, 1 (2019).

[31] G. J. Baart and D. E. Martens, Genome-scale metabolic models: reconstruction and analysis, in *Neisseria meningitidis* (Springer, 2012) pp. 107–126.

[32] S. J. Jol, A. Kümmel, M. Terzer, J. Stelling, and M. Heinemann, System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes, PLoS computational biology **8**, e1002415 (2012).

[33] D. A. Beard, S.-d. Liang, and H. Qian, Energy balance for analysis of complex metabolic networks, Biophysical journal **83**, 79 (2002).

[34] H. Qian, D. A. Beard, and S.-d. Liang, Stoichiometric network theory for nonequilibrium biochemical systems, European Journal of Biochemistry **270**, 415 (2003).

[35] K. C. Soh and V. Hatzimanikatis, Network thermodynamics in the post-genomic era, Current opinion in microbiology **13**, 350 (2010).

[36] N. D. Price, I. Famili, D. A. Beard, and B. Ø. Palssaon, Extreme pathways and kirchhoff's second law, Biophysical journal **83**, 2879 (2002).

[37] R. Nigam and S. Liang, Thermodynamic feasibility of metabolic networks, in *Advances In Bioinformatics And Its Applications* (World Scientific, 2005) pp. 25–36.

[38] D. De Martino, F. Capuani, M. Mori, A. De Martino, and E. Marinari, Counting and correcting thermodynamically infeasible flux cycles in genome-scale metabolic networks, Metabolites **3**, 946 (2013).

[39] J. Saldida, A. P. Muntoni, D. de Martino, G. Hubmann, B. Niebel, A. M. Schmidt, A. Braunstein, A. Milias-Argeitis, and M. Heinemann, Unbiased metabolic flux inference through combined thermodynamic and 13c flux analysis, bioRxiv 10.1101/2020.06.29.177063 (2020).

[40] A. C. Müller and A. Bockmayr, Fast thermodynamically constrained flux variability analysis, Bioinformatics **29**, 903 (2013).

[41] H. P. Bonarius, G. Schmid, and J. Tramper, Flux analysis of underdetermined metabolic networks: the quest for the missing constraints, Trends in Biotechnology **15**, 308 (1997).

[42] A. Hoppe, S. Hoffmann, and H.-G. Holzhütter, Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks, BMC systems biology **1**, 1 (2007).

[43] B. Niebel, S. Leupold, and M. Heinemann, An upper limit on gibbs energy dissipation governs cellular metabolism, Nature Metabolism **1**, 125 (2019).

[44] H. S. Haraldsdóttir and R. M. Fleming, Identification of conserved moieties in metabolic networks by graph theoretical analysis of atom transition networks, PLOS Computational Biology **12**, e1004999 (2016).

[45] P. J. Olver and C. Shakiban, Vector spaces and bases, in *Applied Linear Algebra* (Springer International Publishing, Cham, 2018) pp. 75–128.

[46] A. Kolmogoroff, Zur Theorie der Markoffschen Ketten, Mathematische Annalen **112**, 155 (1936).

[47] S. de Groot and P. Mazur, *Non-equilibrium Thermodynamics*, Dover Books on Physics (Dover Publications, 1984).

[48] I. Prigogine, Thermodynamics of chemical transformations, in *Modern Thermodynamics* (John Wiley & Sons, Ltd, 2014) Chap. 9, pp. 231–264.

[49] R. Wegscheider, Über simultane Gleichgewichte und die Beziehungen zwischen Thermodynamik und Reactionskinetik homogener Systeme, Monatshefte für Chemie und verwandte Teile anderer Wissenschaften **22**, 849 (1901).

[50] C. Xue and N. Goldenfeld, Coevolution maintains diversity in the stochastic "kill the winner" model, Phys. Rev. Lett. **119**, 268101 (2017).

[51] E. Levien and P. C. Bressloff, Robustness of stochastic chemical reaction networks to extrinsic noise: The role of deficiency, Multiscale Modeling & Simulation **16**, 1519 (2018).

[52] A. J. McKane and T. J. Newman, Predator-prey cycles from resonant amplification of demographic stochasticity, Phys. Rev. Lett. **94**, 218102 (2005).

[53] D. Blackmond, "if pigs could fly" chemistry: A tutorial on the principle of microscopic reversibility, Angewandte Chemie International Edition **48**, 2648 (2009).

[54] A. I. Brown and D. A. Sivak, Theory of nonequilibrium free energy transduction by molecular machines, Chemical Reviews **120**, 434 (2020).

[55] J. A. Owen, T. R. Gingrich, and J. M. Horowitz, Universal thermodynamic bounds on nonequilibrium response with biochemical applications, Phys. Rev. X **10**, 011066 (2020).

[56] M. Assaf, E. Roberts, and Z. Luthey-Schulten, Determining the stability of genetic switches: Explicitly accounting for mrna noise, Phys. Rev. Lett. **106**, 248102 (2011).

[57] C. Vogt and B. M. Weckhuysen, The concept of active site in heterogeneous catalysis, Nat Rev Chem **6**, 89 (2022).

[58] H. Stöckel, Linear and nonlinear generalizations of onsager's reciprocity relations. treatment of an example of chemical reaction kinetics, Fortschritte der Physik/Progress of Physics **31**, 165 (1983).

[59] M. Polettini, Fluctuation-dissipation supersymmetry (2020), arXiv:2002.08839 [cond-mat.stat-mech].

[60] P. Petersen, Basic theory, in *Linear Algebra* (Springer New York, New York, NY, 2012) pp. 1–123.

[61] J. M. Borwein and A. S. Lewis, *Convex analysis and nonlinear optimization: theory and examples*, 2nd ed., CMS books in mathematics No. 3 (Springer, New York, 2006).

[62] A.-L. de Lavoisier, *Traité élémentaire de chimie (Elementary Treatise on Chemistry)* (Paris: Cuchet, 1789).

[63] R. Rao and M. Esposito, Conservation laws shape dissipation, New Journal of Physics **20**, 023007 (2018).

[64] R. Rao and M. Esposito, Conservation laws and work fluctuation relations in chemical reaction networks, The Journal of Chemical Physics **149**, 245101 (2018).

[65] S. Katz, J. L. Lebowitz, and H. Spohn, Phase transitions in stationary nonequilibrium states of model lattice systems, Physical Review B **28**, 1655 (1983).

[66] S. Katz, J. L. Lebowitz, and H. Spohn, Nonequilibrium steady states of stochastic lattice gas models of fast ionic conductors, Journal of Statistical Physics **34**, 497 (1984).

[67] C. Maes, Local detailed balance, SciPost Physics Lecture Notes , 032 (2021).

[68] N. Freitas and M. Esposito, Emergent second law for nonequilibrium steady states (2021).

[69] M. Esposito and C. Van den Broeck, Three faces of the second law. I. Master equation formulation, Physical Review E **82**, 011143 (2010).

[70] H. Ge, Mesoscopic kinetic basis of macroscopic chemical thermodynamics: A mathematical theory, Physical Review E **94**, 10.1103/PhysRevE.94.052150 (2016).

[71] L. Onsager, Reciprocal relations in irreversible processes. I., Phys. Rev. **37**, 405 (1931).

[72] J. M. Horowitz, Diffusion approximations to the chemical master equation only have a consistent stochastic thermodynamics at chemical equilibrium, The Journal of Chemical Physics **143**, 044111 (2015).

[73] G. Craciun and M. Feinberg, Multiple Equilibria in Complex Chemical Reaction Networks: II. The Species-Reaction Graph, SIAM Journal on Applied Mathematics **66**, 1321 (2006).

[74] G. Craciun and M. Feinberg, Multiple Equilibria in Complex Chemical Reaction Networks: Semiopen Mass Action Systems, SIAM Journal on Applied Mathematics **70**, 1859 (2010).

[75] M. Feinberg, Chemical Oscillations, Multiple Equilibria, and Reaction Network Structure, in *Dynamics and Modelling of Reactive Systems*, edited by W. E. Stewart, W. H. Ray, and C. C. Conley (Academic Press, 1980) pp. 59–130.

[76] I. R. Epstein and K. Showalter, Nonlinear chemical dynamics: Oscillations, patterns, and chaos, The Journal of Physical Chemistry **100**, 13132 (1996).

[77] D. F. Anderson, G. Craciun, and T. G. Kurtz, Product-Form Stationary Distributions for Deficiency Zero Chemical Reaction Networks, Bulletin of Mathematical Biology **72**, 1947 (2010).

[78] D. Cappelletti and C. Wiuf, Product-Form Poisson-Like Distributions and Complex Balanced Reaction Systems, SIAM Journal on Applied Mathematics **76**, 411 (2016).

[79] M. Polettini, A. Wachtel, and M. Esposito, Dissipation in noisy chemical networks: The role of deficiency, The Journal of Chemical Physics **143**, 184103 (2015).

[80] D. Forastiere, R. Rao, and M. Esposito, Linear stochastic thermodynamics (2022).

[81] T. J. Kobayashi, D. Loutchko, A. Kamimura, and Y. Sughiyama, Kinetic derivation of the Hessian geometric structure in chemical reaction networks, Physical Review Research **4**, 033066 (2022).

[82] Y.-J. Yang and Y.-C. Cheng, Potentials of continuous Markov processes and random perturbations, Journal of Physics A: Mathematical and Theoretical **54**, 195001 (2021).

[83] W. G. Faris, The fundamental theorem of calculus for a matroid, Journal of Mathematical Physics **53**, 063305 (2012).

[84] M. Polettini and M. Esposito, Transient fluctuation theorems for the currents and initial equilibrium ensembles, Journal of Statistical Mechanics: Theory and Experiment **2014**, P10033 (2014).

[85] P. Pietzonka, J. Guioth, and R. L. Jack, Cycle counts and affinities in stochastic models of nonequilibrium systems, Physical Review E **104**, 064137 (2021).

[86] P. E. Harunari, A. Garilli, and M. Polettini, The beat of a current (2022).

[87] F. Roy, M. Barbier, G. Biroli, and G. Bunin, Complex interactions can create persistent fluctuations in high-diversity ecosystems, PLOS Computational Biology **16**, e1007827 (2020).

[88] M. T. Pearce, A. Agarwala, and D. S. Fisher, Stabilization of extensive fine-scale diversity by ecologically driven spatiotemporal chaos, Proceedings of the National Academy of Sciences **117**, 14572 (2020).

[89] T. Butler and N. Goldenfeld, Robust ecological pattern formation induced by demographic noise, Physical Review E **80**, 030902 (2009).

[90] G. Bunin, Ecological communities with Lotka-Volterra dynamics, Physical Review E **95**, 042414 (2017).

[91] E. De Giuli and C. Scalliet, Dynamical mean-field theory: from ecosystems to reaction networks (2022).

[92] J. Resh, The Inverse of a Nonsingular Submatrix of an Incidence Matrix, IEEE Transactions on Circuit Theory **10**, 131 (1963), conference Name: IEEE Transactions on Circuit Theory.

[93] J. H. Bevis, F. J. Hall, and I. J. Katz, Integer generalized inverses of incidence matrices, Linear Algebra and its Applications **39**, 247 (1981).

[94] F. P. Kelly, *Reversibility and stochastic networks*, Wiley series in probability and mathematical statistics (Wiley, Chichester ; New York, 1979).

[95] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer Series in Synergetics Springer-Verlag, Berlin, 2009).

[96] M. Doi, Second quantization representation for classical many-particle system, J. Phys. A: Math. Gen. **9**, 1465 (1976).

[97] L. Peliti, Path integral approach to birth-death processes on a lattice, Journal de Physique **46**, 15 (1985).

[98] D. C. Mattis and M. L. Glasser, The uses of quantum field theory in diffusion-limited reactions, Rev. Mod. Phys. **70**, 979 (1998).

[99] J. Cardy, Reaction-diffusion processes, Lectures held at Warwick University (2006).

[100] K. J. Wiese, Coherent-state path integral versus coarse-grained effective stochastic equation of motion: From reaction diffusion to stochastic sandpiles, Physical Review E **93**, 042117 (2016).

[101] D. T. Gillespie, A rigorous derivation of the chemical master equation, Physica A: Statistical Mechanics and its Applications **188**, 404 (1992).

[102] K. Sekimoto, *Stochastic energetics*, Lecture notes in physics No. 799 (Springer, Heidelberg, 2010).

[103] N. G. v. Kampen, *Stochastic processes in physics and chemistry*, 3rd ed. (Elsevier, Amsterdam ; Boston, 2007).

[104] D. T. Gillespie, The chemical Langevin equation, The Journal of Chemical Physics **113**, 297 (2000).

[105] A. Ceccato and D. Frezzato, Remarks on the chemical Fokker-Planck and Langevin equations: Nonphysical currents at equilibrium, The Journal of Chemical Physics **148**, 064114 (2018).