

# Algebra and geometry in the study of enzymatic cascades

Alicia Dickenstein

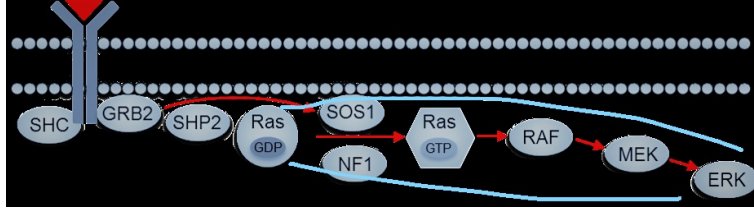
**Abstract** In recent years, techniques from computational and real algebraic geometry have been successfully used to address mathematical challenges in systems biology. The algebraic theory of chemical reaction systems aims to understand their dynamic behavior by taking advantage of the inherent algebraic structure in the kinetic equations, and does not need a priori determination of the parameters, which can be theoretically or practically impossible. This chapter gives a brief introduction to general results based on the network structure. In particular, we describe a general framework for biological systems, called MESSI systems, that describe Modifications of type Enzyme-Substrate or Swap with Intermediates and include many post-translational modification networks. We also outline recent methods to address the important question of multistationarity, in particular in the study of enzymatic cascades, and we point out some of the mathematical questions that arise from this application.

## 1 Introduction

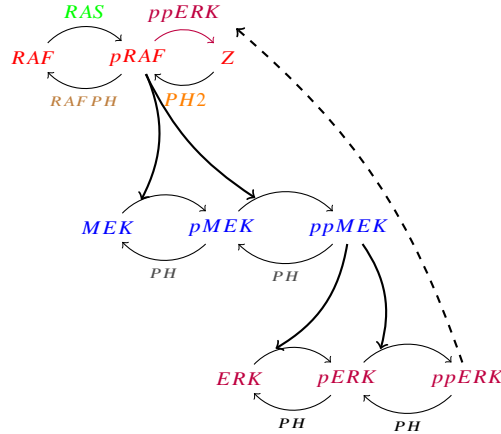
We start by introducing the cartoon mechanisms of two enzymatic signal pathways depicted in research articles. The important RAS signaling pathway in Figure 1 includes an extracellular ligand and a transmembrane receptor, which trigger a cascade of protein-protein interactions and enzymatic reactions, then integrated into key biological responses controlling cell proliferation, differentiation or death. When this pathway is altered, it can drive to unhealthy cell proliferation [40]. Figure 2 presents a more precise description of the last part of the enzymatic cascade.

---

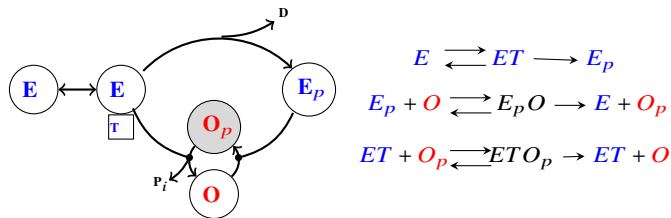
Dept. of Mathematics, FCEN, University of Buenos Aires and IMAS, UBA-CONICET, Argentina,  
e-mail: alidick@dm.uba.ar



**Fig. 1** The RAS signaling pathway, starting in the membrane of the cell



**Fig. 2** Part of the RAS signaling pathway inside the cell, possibly with retroactivity



**Fig. 3** EnvZ-OmpR bacterial model

Figure 3 depicts an osmolarity regulation network in bacteria, which is implemented in part by the EnvZ/OmpR two-component system [48]. The sensor kinase EnvZ (denoted by E in the diagram) autophosphorylates on a histidine residue ( $E_p$ ) and catalyzes the transfer of the phosphate group to the aspartate residue of the response regulator OmpR (O), which then acts as an effector. In this mechanism, when EnvZ is bounded to ATP (ET) also catalyzes hydrolysis of the phosphorylated OmpR-P ( $O_p$ ), which is a transcription factor that regulates the expression of various protein pores. This unusual design keeps the limit concentration of OmpR-P at a value that is independent of the positive initial concentrations.

When we first look at these biological mechanisms, it does not seem evident that algebra and geometry can be used to analyze them. But we will argue in this chapter that this is indeed the case and that we can contribute with these mathematical tools to the understanding of questions in Systems Biology.

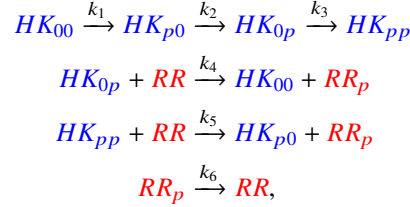
In particular, in the realm of biochemical reaction networks, that is, chemical reaction networks in biochemistry, the usual mass-action kinetics modeling of the evolution of the concentrations of the different chemical species along time (as RAS, RAF, MEK, ERK, E, O, etc. above) yields an autonomous system of polynomial ordinary differential equations  $\frac{dx}{dt} = f_{\kappa}(x)$  in the unknown vector of concentrations  $x$  of the species as functions of time, for each choice of the (real positive) reaction rate constants  $\kappa$  (see Definition 1). In fact, these equations are associated to a labeled directed graph  $G$  of reactions. The monomial terms come from the labels of the nodes of  $G$  by complexes in the given species, the coefficients depend on the (positive) reaction rate constants  $\kappa$  that label the edges of  $G$ , and the total production of each reaction (which is the difference of the labels of the target and source nodes). The real polynomials  $f_{\kappa}(x)$  carry a combinatorial structure inherited from  $G$  and we will also think of  $\kappa$  as *parameters* and consider the *family* of differential systems parametrized by them. Chemical Reaction Network Theory (CNRT) was initiated by Horn and Jackson and subsequently by Feinberg and his students and collaborators [22] and has seen a great development over the last years, when new combinatorial and algebro-geometric techniques have been introduced. We refer the reader to the survey article [16] for basic definitions, results and further references, and we review here some advances developed after that article was published.

In Section 4 we recall the notion of MESSI systems we introduced in [41]. Many post-translational modification networks are MESSI networks. For example: the motifs in [23], sequential distributive multisite networks [51], sequential processive multisite phosphorylation networks [12], phosphorylation cascades or the bacterial EnvZ/OmpR network from [48] in Figure 3. Our work is inspired by and extends some results in several previous articles [24, 39, 42, 28, 29, 31, 47, 50]. MESSI is an acronym for Modifications of type Enzyme-Substrate or Swap with Intermediates (see Definition 2). Networks with an underlying MESSI structure include many post-translational modification networks, as well as all linear systems arising from mass-action kinetics (a.k.a. Laplacian dynamics [38]). We summarize some results and algorithms based on this structure to predict conservation relations, persistence, the capacity for multistationarity, and the description of regions of multistationarity. Once the network has the capacity for multistationarity, the next main question is how to predict parameters of, if possible, regions in parameter space which give rise to multistationary systems, which are called *multistationarity regions*. In Section 5 we comment on several recent approaches to study multistationarity in chemical reaction networks. Section 6 mentions the mostly unexplored question of the a priori determination of the occurrence of oscillations in chemical reaction networks, in particular, in enzymatic networks. We end the paper with two main open questions.

## 2 Basics of mass-action kinetics

In this section we set the basic terminology and the mathematical concepts mentioned in the introduction. In particular, we discuss the notion of multistationarity.

Two-component signal transduction systems enable bacteria to sense, respond, and adapt to a wide range of environments, stressors, and growth conditions. Before giving the precise Definition 1, we instantiate mass-action kinetics in a biological example of a simple two-component mechanism. It relies on *phosphotransfer* reactions. Upon receiving a signal, the *hybrid histidine kinase* HK can self-phosphorylate. This is a hybrid histidine kinase with two phosphorylatable domains. We denote the phosphorylation state of each site by  $p$ , if the site is phosphorylated, and 0, if it is not; the four possible forms are  $HK_{00}$ ,  $HK_{p0}$ ,  $HK_{0p}$ ,  $HK_{pp}$ . The *response regulator protein* is denoted by  $RR$  when it is unphosphorylated and  $RR_p$  denotes the phosphorylated form. Given a vector of *reaction rate constants*  $k = (k_1, \dots, k_6) \in \mathbb{R}_{>0}^6$ , the (directed) graph of reaction equals:



where each of the ten nodes corresponds to a *complex* on the six *chemical species*, that we number in the following order:  $HK_{00}$ ,  $HK_{p0}$ ,  $HK_{0p}$ ,  $HK_{pp}$ ,  $RR$ ,  $RR_p$ . Mass-action kinetics specifies how the respective *concentrations*  $x_1, \dots, x_6$  of these six species evolve with time. The basic principle in this modeling is derived from the idea that the rate of an elementary reaction is proportional to the probability of collision of the reactants, which under an independence assumption equals the product of their concentrations. We derive the following autonomous polynomial dynamical system  $\frac{dx_i}{dt} = f_i(x)$ ,  $i = 1, \dots, 6$ :

$$\begin{aligned}
 \frac{dx_1}{dt} &= -k_1 x_1 + k_4 x_3 x_5, & \frac{dx_2}{dt} &= k_1 x_1 - k_2 x_2 + k_5 x_4 x_5, \\
 \frac{dx_3}{dt} &= k_2 x_2 - k_3 x_3 - k_4 x_3 x_5, & \frac{dx_4}{dt} &= k_3 x_3 - k_5 x_4 x_5, \\
 \frac{dx_5}{dt} &= -k_4 x_3 x_5 - k_5 x_4 x_5 + k_6 x_6, & \frac{dx_6}{dt} &= k_4 x_3 x_5 + k_5 x_4 x_5 - k_6 x_6.
 \end{aligned}$$

It is straightforward to check that the following linear dependencies hold and generate all the linear dependencies among  $f_1, \dots, f_6$ :

$$f_1 + f_2 + f_3 + f_4 = f_5 + f_6 = 0,$$

from which we deduce two linear *conservation relations*:

$$x_1 + x_2 + x_3 + x_4 = T_1, \quad x_5 + x_6 = T_2.$$

Thus, trajectories lie in a 4-plane in 6-space. The *total conservation constants*  $T_1, T_2$  are determined by the initial conditions  $(x_1(0), \dots, x_6(0))$ .

Given a numbering of the species as above, we usually identify a complex on these species with a nonnegative integer vector. For example, the complex  $y = X_3 + X_5$  is identified with the vector  $e_3 + e_5 = (0, 0, 1, 0, 1, 0) \in \mathbb{Z}_{\geq 0}^6$ . The general definition is as follows.

### Definition of chemical reaction networks and mass-action kinetics

**Definition 1** A *chemical reaction network* (on a finite set of  $s$  species, which we assume ordered) is a finite labeled directed graph  $G = (V, E, (\kappa_{ij})_{(i,j) \in E}, (y_i)_{i=1, \dots, m})$ , whose vertices  $V$  are labeled by complexes  $y_1, \dots, y_m \in \mathbb{Z}_{\geq 0}^s$  and whose edges  $(i, j) \in E$  are labeled by positive real numbers  $i \xrightarrow{\kappa_{ij}} j$ . We will also say that  $G$  is a network.

Mass-action kinetics specified by the network  $G$  gives the following autonomous system of ordinary differential equations in the *concentrations*  $x = (x_1, x_2, \dots, x_s)$  of the species as functions of time:

$$\frac{dx}{dt} = \sum_{(i,j) \in E} \kappa_{ij} x^{y_i} (y_j - y_i) = f_\kappa(x). \quad (1)$$

Here,  $\frac{dx}{dt}$  and  $y_j - y_i$  are column vectors.

Note that the coordinates  $f_1, \dots, f_s$  of  $f_\kappa$  are polynomials in  $\mathbb{R}[x_1, \dots, x_s]$  (note that to ease the notation we omit the dependence of  $f_i$  on  $\kappa$ ). Many systems occurring in population dynamics, for example the oscillatory Lotka-Volterra equations, can be viewed as arising from a chemical reaction network as in (1), but for instance not the “chaotic” Lorenz equations. A simple characterization of autonomous dynamical systems arising from chemical reaction networks under mass-action kinetics has been given by Hárs and Tóth. We refer to the book [20], which also contains an introduction to the stochastic modeling of chemical kinetics.

Another direct consequence of the form of the equations in (1) is that for any trajectory  $x(t)$ , the vector  $\frac{dx}{dt}$  lies for all  $t$  (in any interval  $I$  containing 0 where it is defined) in the so called *stoichiometric subspace*  $S$ , which is the linear subspace generated by the differences  $\{y_j - y_i \mid (i, j) \in E\}$ . Using the shape of the polynomials  $f_i$  it can be seen that the positive orthant  $\mathbb{R}_{>0}^s$  and its closure  $\mathbb{R}_{\geq 0}^s$  are forward-invariant for the dynamics. Then, any trajectory  $x(t)$  starting at a nonnegative point  $x(0)$  lies for all  $t \in I \cap \mathbb{R}_{>0}$  in the closed polyhedron  $(x(0) + S) \cap \mathbb{R}_{\geq 0}^s$ , which is called a *stoichiometric compatibility class*, or for short, an *S-class*.

Denote by  $q$  the codimension of  $S$ . Given a basis  $\ell_1, \dots, \ell_q$  of linear forms in the dual of  $S$ , let  $T_i = \ell_i(x(0))$ ,  $i = 0, \dots, q$ . The equations  $\ell_1(x) = T_1, \dots, \ell_q(x) = T_q$  of  $x(0) + S = S_T$  give linear *conservation relations* and, as above, the constant coefficient  $T_i$  of such a linear equation is called a *total conservation constant*.

### The steady state variety and the notion of multistationarity

The *steady state variety*  $V_\kappa(f)$  of the kinetic system (1) equals the nonnegative real zeros of  $f_1, \dots, f_s$ :

$$V_\kappa(f) = \{x \in \mathbb{R}_{\geq 0}^s : f_1(x) = \dots = f_s(x) = 0\}. \quad (2)$$

An element of  $V_\kappa(f)$  is called a *steady state* of the system and corresponds to a constant trajectory in the nonnegative orthant. We say that system (1) exhibits *multistationarity* if there exist at least two *positive* steady states with the same total conservation constants, that is, in the same  $S$ -class. This is an important property for chemical reaction networks modeling biological processes, since the occurrence of multistationarity allows for different *responses* of the cell under the same total conservation constants, depending on the initial conditions.

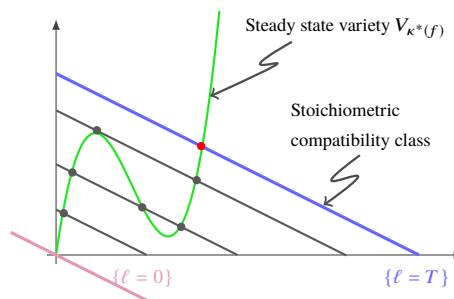
In fact, our point of view will be the following. The underlying reaction network  $(V, E, (y_i)_{i=1, \dots, m})$  defines a *family* of autonomous polynomial dynamical systems depending on the positive parameters  $\kappa \in \mathbb{R}_{> 0}^{\#E}$ . We say that it has the *capacity for multistationarity* if there is a choice of reaction rate constants  $\kappa = (\kappa_{ij})_{(i,j) \in E}$  and total conservation constants  $T = (T_1, \dots, T_q)$  for which the intersection of the steady state variety  $V_\kappa(f)$  with the *positive* points of linear variety  $S_T$  consists of more than one point (that is: there exist parameters  $\kappa$  and  $T$  such that there are at least two points in the positive orthant lying in the intersection of the steady state variety  $V_\kappa(f)$  with the  $S$ -class defined by  $T$ ).

---

There are many results to decide the capacity for multistationarity of a given chemical reaction network, starting with [14]. Most of them have been summarized in Theorem 1.4 of [39]. In fact, these results give in general necessary and sufficient conditions for the stronger condition that the map  $f_\kappa$  is injective on the positive points of all  $S$ -classes. There are several implementations of different algorithms, starting with the pioneering algorithm implemented by Feinberg and his group in the Chemical Reaction Network Toolbox. The link to the corresponding webpage together with links to other algorithms can be found at [https://reaction-networks.net/wiki/Mathematics\\_of\\_Reaction\\_Networks](https://reaction-networks.net/wiki/Mathematics_of_Reaction_Networks). We recall some of the tools to address this question in Sections 4 and 5.

In Figure 4, there is a range of values of  $T$  for which there are three positive steady states on the corresponding translate  $S_T$  of  $S$  (i.e., in an  $S$ -class) for a fixed value  $\kappa^*$  of positive rate constants. So, the chemical reaction network has the capacity for multistationarity and  $\kappa^*$  is a choice of *multistationarity parameter*.

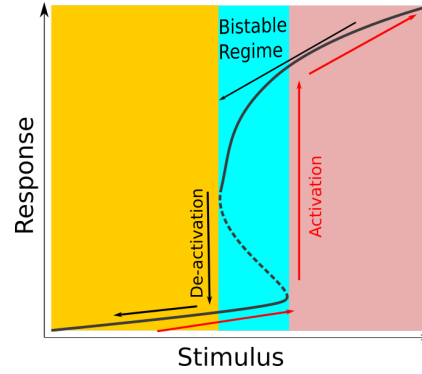
**Fig. 4** The green curve represents the steady state variety  $V_{\kappa^*}(f)$ . The subspace  $S = \{\ell = 0\}$  is a line. The number of points of intersection of the translates  $S_T = \{\ell = T\}$  of  $S$  with  $V_{\kappa^*}(f)$  in the positive orthant depends on the total conservation constant  $T$ .



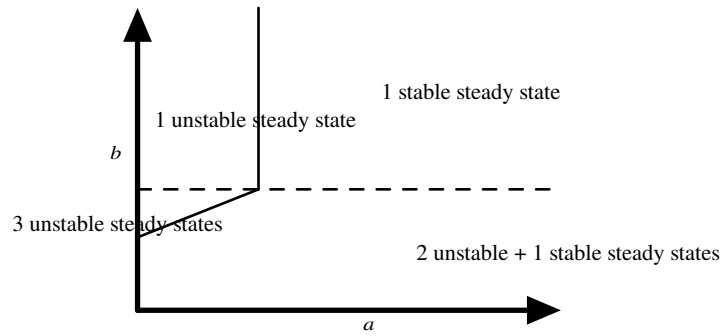
We feature two kinds of multistationarity pictures from the literature. One way to find the special values rendering these figures is by measurements in experiments or by exhaustive (and lucky) simulations of the trajectories taking sample values in the space of parameters and initial conditions. Instead, one can try to develop algebro-geometric tools to analyze the mathematical models arising from biochemical reaction networks, with the goal of making predictions from the structure of the networks.

Figure 5 corresponds to a 2-site sequential phosphorylation and dephosphorylation that we describe in Section 3 below. This network has 15 parameters: 12 reaction constants and 3 total conservation constants. In the picture, all the reaction rate constants and two of the total conservation constants have been specialized and only the total conservation constant  $E_{\text{tot}}$  of one enzyme is varying. This is considered to be the *input* variable (or *stimulus*) and it is represented on the  $x$ -axis. The number of chemical species is equal to 9, but only one of the phosphorylated substrates  $s^*$  at steady state is represented, which is considered the *response* of the system. It happens that in this case any positive value of  $s^*$  is one coordinate of a positive steady state and different steady states in the same  $S$ -class have different  $s^*$  coordinates. The steady state  $s^*$ -coordinate represented on the  $y$ -axis. For small or big values of  $E_{\text{tot}}$ , only one value of  $s^*$  is possible, so this is a monostationary regime. In the middle zone, there are three steady states, two *stable* and one *unstable*, so this is the bistable regime (stability of steady states is determined by the negativity of the real part of the eigenvalues of the Jacobian). This figure corresponds to a two dimensional very particular “slice” of points originally in  $24 = 15 + 9$  variables, where 14 variables have been specialized and 8 variables are not shown.

Figure 6 represents a two dimensional “slice”, but in *parameter space*, of another mechanism that we do not specify, but in which only two of the parameters ( $a, b$ ) are allowed to vary. For each of the values of ( $a, b$ ) outside the line segments separating the regions, there are either one or three positive steady states, which could be stable or unstable. In fact, in most biochemical networks these curves separating the regions are *far from being line segments*; they are high order algebraic hypersurfaces that separate different semialgebraic regions where the qualitative dynamics is the same, in a high dimensional parameter space. Moreover, regions with interesting behaviour could be small.



**Fig. 5** Only one parameter is allowed to vary



**Fig. 6** Only two of the parameters are allowed to vary

The separating hypersurfaces related to the question of multistationarity are described by the union of the discriminant associated to the equations describing  $V_\kappa(f)$  and  $S_T$  with respect to the  $x$  variables (which vanishes whenever there is a point where the intersection of the steady state variety and the  $S$ -class is non-transversal), and the union for any  $i \in \{1, \dots, s\}$  of the resultant describing the fact that there is a common point with  $x_i = 0$ . In each chamber (connected component) of the complement of the union of these algebraic varieties, the number of real roots is the same and moreover, for each of the real roots it holds that the sign of each of the coordinates does not change as the parameters are moved, and thus the number of real roots with a fixed sign (for instance, positive roots) is constant along the chamber. We refer the



reader to the book [26] for the notions of discriminant and resultant, which are in general not linear. These polynomials in the parameters can be computed effectively -in theory- via different computational algebraic geometry methods of elimination of variables, but standard computations are not feasible when there are many variables. Even if one can compute these equations, it is a very complicated task to describe then all the possible chambers in the complement of its zero locus, or at least to find one representative in each chamber. There are implementations by M. Safey El Din, which work very well in small examples using his package RAGlib [46].

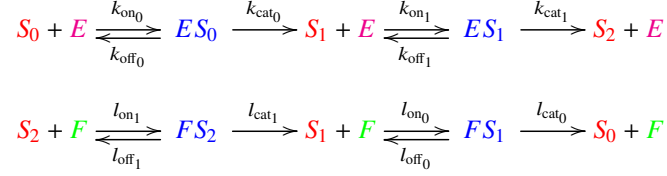
### 3 Two important families of enzymatic networks

In this section, we introduce common enzymatic mechanisms that will help us exemplify and clarify the concepts we will introduce in Section 4.

#### Sequential phosphorylations

The multisite  $n$ -phosphorylation system describes the site phosphorylation of a protein (with  $n$  sites where a phosphate group can be absorbed or emitted) by a pair of enzymes (a kinase and a phosphatase) in a sequential and distributive mechanism. The Nobel Prize in Physiology or Medicine was awarded in 1992 to Edmond Fischer and Edwin Krebs “for their discoveries concerning reversible protein phosphorylation as a biological regulatory mechanism.” The kinase and the phosphatase speed up the transformation of other proteins without being incorporated in the final products of the process, which is crucial in the regulation of metabolism in the body. Multi-site phosphorylation plays important regulatory roles in cell cycle regulation and inflammation pathways, and is implicated in multiple disorders, including Alzheimer disease. Because of the important role played by these systems in signal transduction networks inside the cell, there is a body of work on the mathematics of phosphorylation systems (which belong to the more general class of post-translational modification systems). We refer the reader to the papers [33, 42, 49] and the references therein.

We now describe the special case of a sequential phosphorylation/dephosphorylation with  $n = 2$  sites, which is also known as the dual futile cycle. There are nine species: three substrates (the unphosphorylated substrate  $S_0$ , the substrate with one and two phosphorylated sites  $S_1$  and  $S_2$ ), two enzymes (the kinase  $E$  and the phosphatase  $F$ ), and four intermediate species ( $ES_0$ ,  $ES_1$ ,  $FS_2$  and  $FS_1$ ). We give to the twelve rate constants the usual names in the literature [51].



We number the species and their concentrations as follows:  $x_1, x_2, x_3$  denote the respective concentrations of  $S_0, S_1, S_2$ ;  $y_1, y_2, y_3, y_4$  denote the respective concentrations of the intermediate species  $ES_0, ES_1, FS_2, FS_1$ ,  $x_4$  is the concentration of the kinase  $E$ , and  $x_5$  the concentration of the phosphatase  $F$ . The associated system of ODE's defined in (1) equals in this case:

$$\begin{array}{ll}
\frac{dx_1}{dt} = -k_{\text{on}_0}x_1x_4 + k_{\text{off}_0}y_1 + l_{\text{cat}_0}y_4 & \frac{dx_4}{dt} = -k_{\text{on}_0}x_1x_4 - k_{\text{on}_1}x_2x_4 + (k_{\text{off}_0} + k_{\text{cat}_0})y_1 \\
\frac{dx_2}{dt} = -k_{\text{on}_1}x_2x_4 + k_{\text{cat}_0}y_1 + k_{\text{off}_1}y_2 & \quad + (k_{\text{off}_1} + k_{\text{cat}_1})y_2 \\
\quad - l_{\text{on}_0}x_2x_5 + l_{\text{cat}_1}y_3 + l_{\text{off}_0}y_4 & \frac{dx_5}{dt} = -l_{\text{on}_0}x_2x_5 - l_{\text{on}_1}x_3x_5 + (l_{\text{off}_1} + l_{\text{cat}_1})y_3 \\
\frac{dx_3}{dt} = k_{\text{cat}_1}y_2 - l_{\text{on}_1}x_3x_5 + l_{\text{off}_1}y_3 & \quad + (l_{\text{off}_0} + l_{\text{cat}_0})y_4 \\
\frac{dy_1}{dt} = k_{\text{on}_0}x_1x_4 - (k_{\text{off}_0} + k_{\text{cat}_0})y_1 & \frac{dy_3}{dt} = l_{\text{on}_1}x_3x_5 - (l_{\text{off}_1} + l_{\text{cat}_1})y_3 \\
\frac{dy_2}{dt} = k_{\text{on}_1}x_2x_4 - (k_{\text{off}_1} + k_{\text{cat}_1})y_2 & \frac{dy_4}{dt} = l_{\text{on}_0}x_2x_5 - (l_{\text{off}_0} + l_{\text{cat}_0})y_4.
\end{array}$$

There are 3 independent linear conservation laws, for instance:

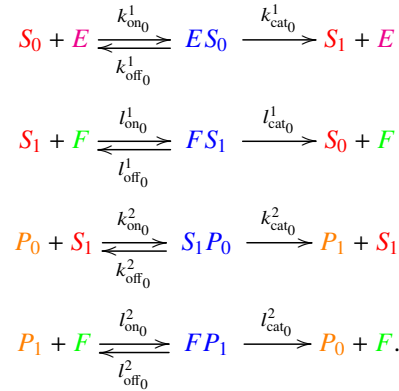
$$\begin{array}{l}
x_1 + x_2 + x_3 + y_1 + y_2 + y_3 + y_4 = S_{\text{tot}} \\
x_4 + y_1 + y_2 = E_{\text{tot}} \\
x_5 + y_3 + y_4 = F_{\text{tot}},
\end{array}$$

where  $S_{\text{tot}}, E_{\text{tot}}, F_{\text{tot}}$  are positive real numbers for any choice of initial condition in the positive orthant. As we pointed out in Section 1, there are  $12 + 3 = 15$  parameters. The  $n$ -site sequential mechanism is similar, with  $3n + 3$  variables,  $6n$  reaction rate constants and always 3 total conservation constants, so a total of  $6n + 3$  parameters.

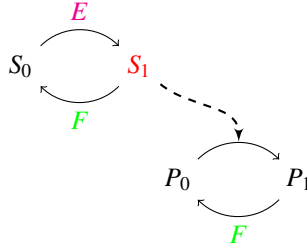
### Phosphorylation cascades

We have already encountered a coarse diagram of an enzymatic cascade in Figure 2. MAP kinase cascades are important signal transduction systems in molecular biology for which there is also a body of mathematical work, see for instance [35, 40] and the references therein. These cascades correspond to a network of enzymatic reactions arranged in layers, where usually in each of them there is a futile cycle of sequential phosphorylations and such that the fully phosphorylated substrate serves as an enzyme for the next layer.

The simplest case of a cascade with the capacity of multistationarity [23] consists of a cascade with two layers and a single phosphorylation/dephosphorylation at each layer, with one phosphatase. It corresponds to the a labeled digraph, with 9 variables and 18 parameters, where each single phosphorylation follows the same mechanism as in our previous example, with an intermediate species. The nine species are the substrates  $S_0, S_1$  in the first layer, the substrates  $P_0, P_1$  in the second layer, four intermediate complexes, a kinase  $E$  and the *same* phosphatase  $F$  to dephosphorylate the substrates in both layers. The forward enzyme in the second layer is the phosphorylated substrate  $S_1$  from the first layer.



This mechanism is usually depicted as follows, hiding the reaction rate constants and the intermediate species:



In this case, there are 4 linearly independent conservation relations. Denoting with small letters the concentration of each of the species, these conservation relations can be chosen as follows, as predicted in Theorem 3.2 in [41] (see (4) below):

$$\begin{array}{l}
 s_0 + s_1 + e s_0 + f s_1 + s_1 p_0 = S_{\text{tot}} \\
 p_0 + p_1 + s_1 p_0 + f p_1 = P_{\text{tot}} \\
 e + e s_0 = E_{\text{tot}} \\
 f + f s_1 + f p_1 = F_{\text{tot}},
 \end{array}$$

where  $S_{\text{tot}}, P_{\text{tot}}, E_{\text{tot}}, F_{\text{tot}}$  are positive real numbers for any choice of initial condition in the positive orthant.

We can also consider cascades with any number  $n$  of layers. In this case, the number of variables, the number of reaction rate constants and the number of independent linear conservation relations (as well as the number of linear conservation constants) grow linearly with  $n$ .

## 4 MESSI systems

In this section we recall the notion of MESSI networks from [41], to describe a common structure underlying the four examples above in their different variants as well as many “popular” biological networks, that consist of Modifications of type Enzyme-Substrate or Swap with Intermediates. The occurrence of this structure allows us to prove general results for quite different mechanisms. The basic ingredient of a MESSI structure is a partition of the set of species, which reflects the different chemical behaviors. This grouping of the chemical species into disjoint subsets is in accordance with the intuitive partition of the species according to their function that biochemists have. We will denote the disjoint union of sets with the symbol  $\sqcup$ .

### Definition of a MESSI system

**Definition 2** A MESSI network is a chemical reaction network satisfying the following properties. First of all, there exists a partition of the set  $\mathcal{S}$  of species

$$\mathcal{S} = \mathcal{S}^{(0)} \sqcup \mathcal{S}^{(1)} \sqcup \mathcal{S}^{(2)} \sqcup \dots \sqcup \mathcal{S}^{(m)}, \quad (3)$$

where  $m \geq 1$ ,  $\mathcal{S}^{(0)}$  is the subset of intermediate species and could be empty, and all  $\mathcal{S}^{(i)}$  with  $i \geq 1$  are nonempty subsets, formed by what we call core species. We require that the complexes and reactions satisfy the following conditions. An intermediate species can only be part of a monomolecular complex consisting only of this species (called an intermediate complex). Non-intermediate complexes are called core complexes and consist of one or otherwise two chemical species belonging to different subsets of the partition. Denote by  $y \rightarrow_{\circ} y'$  the existence of an edge from complex  $y$  to complex  $y'$  or a directed path of reactions from  $y$  to  $y'$  through intermediate complexes. We require that for any intermediate complex  $y_0$ , there exist core complexes  $y, y'$  such that  $y \rightarrow_{\circ} y_0 \rightarrow_{\circ} y'$ . If there are two monomolecular core complexes  $y \rightarrow_{\circ} y'$ , then both should consist of a species in the same  $\mathcal{S}^{(\alpha)}$ . We further ask that if there is a reaction between a monomolecular and a bimolecular complex, the monomolecular complex is an intermediate, and that if  $y, y'$  are bimolecular core complexes such that  $y \rightarrow_{\circ} y'$ , then there exist two different core subsets  $\mathcal{S}^{(\alpha)}, \mathcal{S}^{(\beta)}$  in the partition, such that both  $y$  and  $y'$  consist of a species in each of them.

When endowed with mass-action kinetics, a MESSI network gives rise to a MESSI system of polynomial autonomous ODE's.

### All the networks we mentioned are MESSI

All the networks we mentioned in the text (plus many other common biochemical networks) can be endowed with the structure of a MESSI system. We gave different colors to the different subsets in a possible partition of the species.

For instance, in the cascade depicted in Figure 2 in the Introduction, the intermediate species (complexes) are not displayed, but we presented with different colors a possible partition of the core species that defines a MESSI structure. In the network depicted in Figure 3 the partition into a subset of intermediate species (in black), and two subsets of core species (in red and blue) also defines a MESSI structure.

In the two-component system in Section 2, we could take  $\mathcal{S}^{(0)} = \emptyset$ ,  $\mathcal{S}^{(1)} = \{\text{HK}_{00}, \text{HK}_{p0}, \text{HK}_{0p}, \text{HK}_{pp}\}$ , and  $\mathcal{S}^{(2)} = \{\text{RR}, \text{RR}_p\}$ .

In the example of the sequential phosphorylation in Section 3, we could take  $\mathcal{S}^{(0)} = \{ES_0, ES_1, FS_2, FS_1\}$ ,  $\mathcal{S}^{(1)} = \{S_0, S_1, S_2\}$ ;  $\mathcal{S}^{(2)} = \{E\}$ , and  $\mathcal{S}^{(3)} = F$ . It can be checked that all conditions are satisfied. Note that if we consider the coarser partition with the same set of intermediate species  $\mathcal{S}^{(0)}$ , the same set  $\mathcal{S}^{(1)}$  of core species, and just one other set  $\{E, F\}$  of core species, we also have a MESSI structure. In fact, there is in general a *poset* of possible partitions (and in other examples there could be non-comparable partitions).

On the other side, in the example of the cascade in Section 3, we can partition the set of nine species as follows to define a MESSI structure in the 2-layer cascade:  $\mathcal{S}^{(0)}$  consists of the four intermediate species  $\{ES_0, FS_1, S_1P_0, FP_1\}$ , plus the core subsets  $\mathcal{S}^{(1)} = \{S_0, S_1\}$ ,  $\mathcal{S}^{(2)} = \{P_0, P_1\}$ ,  $\mathcal{S}^{(3)} = \{E\}$ , and  $\mathcal{S}^{(4)} = \{F\}$ .

### Conservation laws

The first general results about MESSI systems is that we can describe enough (explicit) conservation linear relations with positive coefficients. Given a partition (3) of the set  $\mathcal{S}$  of variables into one intermediate subset and  $m \geq 1$  nonempty core subsets defining a MESSI structure in a given network  $G$ , note that the associated autonomous polynomial dynamical system defined in (1) is linear in the variables of each  $\mathcal{S}^{(i)}$  union the subset  $\text{Int}_i$  consisting of those intermediate species  $y'$  for which there exists a core complex  $y$  containing one species of  $\mathcal{S}^{(i)}$  such that  $y \rightarrow_{\circ} y'$  (for any fixed  $i = 1, \dots, m$ ). The union of these subsets  $\text{Int}_i$  equals  $\mathcal{S}^{(0)}$ , but they are in general not disjoint, because if in the recent notation  $y$  also contains a species in another  $\mathcal{S}^{(j)}$ , then  $y'$  also belongs to  $\text{Int}_j$ . These intersections account for several important properties of the systems.

Theorem 3.2 in [41] asserts that given a partition of  $\mathcal{S} = \{x_1, \dots, x_s\}$  defining a MESSI structure as in (3), the following linear forms  $\ell_1, \dots, \ell_m$  belong to the dual of the stoichiometric subspace  $S$ :

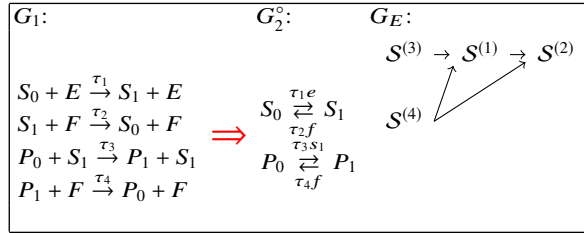
$$\ell_i(x) = \sum_{x_j \in \mathcal{S}^{(i)}} x_j + \sum_{x_j \in \text{Int}_i} x_j, \quad i = 1, \dots, m. \quad (4)$$

We refer the reader to Section 3 in [41] for conditions ensuring that these are a basis of conservation relations (and examples where this is not the case). We conclude that all MESSI systems are *conservative*. Thus, all  $\mathcal{S}$ -classes are compact, and all trajectories are bounded and defined for any positive time. In fact, given a MESSI network, if  $x$  is a trajectory of the associated mass-action kinetics dynamical system  $\dot{x}(t) = f(x(t))$ , for all  $t$  in an open interval containing  $\mathbb{R}_{\geq 0}$  with  $x(0) \in \mathbb{R}_{>0}^{\mathcal{S}}$ , let  $(T_1, \dots, T_m) = (\ell_1(x(0)), \dots, \ell_m(x(0)))$ . Then, we have that for any  $t \geq 0$  it holds that  $\ell_i(x(t)) = T_i$  for any  $i$ . Then, all the coefficients of the linear form  $\ell = \sum_{i=1}^m \ell_i$  are positive and  $\ell(x(t)) = \sum_{i=1}^m T_i > 0$ .

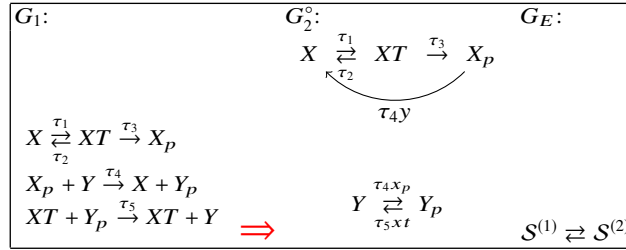
### The associated digraphs

In order to state some other general results for MESSI networks, we introduce three associated digraphs  $G_1, G_2, G_E$  associated with a given MESSI network  $G$  with a vector of rate constants  $k$ . We refer the reader to Section 3 in [41] for complete definitions, explanations and examples.

We eliminate all intermediate species to define  $G_1$ , which naturally inherits a MESSI structure: the species of  $G_1$  are the core species of  $G$ , its complexes are the core complexes of  $G$  and there is an edge between two core complexes  $y, y'$  precisely when  $y \rightarrow_{\circ} y'$  in  $G$ . The rate constants of  $G_1$  are rational functions  $\tau(\kappa)$  with nonzero denominator over all positive  $\kappa$ , in such a way that when viewed with mass action kinetics gives raise to a system of the form  $\dot{x}' = f^1(x')$ , the steady state variety  $V_{\tau(\kappa)(f^1)}$  of the system defined by  $G_1$  is a projection of the steady state variety  $V_{\kappa}(f)$  of the original system. They have been explicitly defined in display (15) of the Supplementary Material in [24], see displays (5.3) and (5.8) in [6]. To define the digraph  $G_2$ , we first consider for any  $i = 1, \dots, m$  the linear network obtained by “hiding” in the rate constants the concentration of all species  $x_j \notin \mathcal{S}^{(i)}$ . For instance, an edge  $X_j + X_k \rightarrow X_{j_1} + X_{k_1}$  with  $X_j, X_{j_1} \in \mathcal{S}^{(i_1)}, X_k, X_{k_1} \in \mathcal{S}^{(i_2)}$ , with rate constant  $c$ , gives raise to the following two edges in  $G_2$ : the edge  $X_j \rightarrow X_{j_1}$  with rate constant  $c x_k$ , and the edge  $X_k \rightarrow X_{k_1}$ , with rate constant  $c x_j$ . Note that we get this way a multidigraph  $MG_2$  with possibly repeated edges and loops. We then denote by  $G_2$  the digraph derived from  $MG_2$  after collapsing multiple edges into a single edge, with label equal to the sum of the labels of the different edges. The nodes in each connected component of  $G_2$  correspond to the species in one of the subsets  $\mathcal{S}^{(i)}$  of the partition if and only if this partition is minimal (in the poset of partitions of  $\mathcal{S}$  defining a MESSI structure on  $G$ ). The digraph  $G_2$  is linear (each node is labeled with a monomolecular complex with a single species) and again, if we formally associate to it mass-action kinetics, its steady state variety coincides with that of  $G_1$ . Finally, we denote by  $G_2^{\circ}$  the multidigraph obtained from  $G_2$  after deleting all loops. On the other side, the nodes of the digraph  $G_E$  are the subsets  $\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(m)}$  and



**Fig. 7** The graphs  $G_1$ ,  $G_2^{\circ}$  and  $G_E$  for the phosphorylation cascade in Section 3



**Fig. 8** The graphs  $G_1$ ,  $G_2^{\circ}$  and  $G_E$  for the EnvZ/OmpR two-component network in the Introduction

there is an edge from  $S^{(i_1)}$  to  $S^{(i_2)}$  with a label containing as a factor the concentration of any species in  $S^{(i_1)}$ .

The graphs  $G_1, G_2^{\circ}$  and  $G_E$  associated to two of the networks in the previous sections are depicted in Figures 7 and 8.

### Persistence

A chemical reaction system (1) is persistent if any trajectory starting from a point with positive coordinates stays at a positive distance from any point in the boundary, or informally, if no species which is present can tend to be eliminated in the course of the reaction. A steady state lying in the boundary of the nonnegative orthant (that is, with some coordinates equal to zero) is called *relevant* if it lies in the intersection of the boundary of the nonnegative orthant with a stoichiometric compatibility class through a point in  $\mathbb{R}_{>0}^s$ . As MESSI systems are conservative, Theorem 2 in [1] proves that a MESSI system is persistent when there are no *relevant boundary steady states*.

Given a MESSI network  $G$ , we identify the following hypotheses:

- (A) The associated digraph  $G_2$  is weakly reversible.
- (B) The associated digraph  $G_E$  has no directed cycles.

Hypothesis (A) means that for any pair of nodes in the same connected component, there is a directed path from one to the other. For instance, in the two examples

considered in Figures 7 and 8, hypothesis (A) is verified. Hypothesis (B) is also verified in the case of the cascade network, but not in the EnvZ/OmpR two-component network. However, even if they sound restrictive, there is a big range of signaling pathways that satisfy both hypotheses.

Theorem 3.15 in [41] asserts that a MESSI network  $G$  which satisfies hypotheses (A) and (B) does not have relevant boundary steady states, and is thus persistent. Moreover, as MESSI systems are conservative, a version of Brouwer’s fixed point theorem ensures the existence of a non-negative steady state in each  $S$ -class. So, the absence of relevant boundary steady states implies the existence of a positive steady state in each  $S$ -class.

---

### Explicit rational parametrizations

We want to describe the intersection  $V_\kappa(f) \cap S_T$  in the positive orthant. The steady state variety is defined in principle by  $s$  polynomial equations. Assume the dimension of  $S$  (and thus of  $S_T$  for any  $T$ ) equals  $s - q$  and can thus be defined by  $q$  linear equations. This implies that there are (at most)  $s - q$  linearly independent polynomials among  $f_1, \dots, f_s$ . A finite number of common solutions is expected, but this might not be true.

One way to simplify the computation of the intersection is the following. As  $S_T$  are linear varieties, they can be parametrized by  $s - q$  parameters. One could then parametrize  $S_T$  solving for  $q$  variables in terms of the other ones and then replace this in the equations of the steady state variety. This reduces the number of variables to from  $s$  to  $s - q$ , but the polynomials  $f_1, \dots, f_s$  are particular, with a monomial structure that comes from  $G$  and we would in general destroy the sparsity.

Denote by  $V_{>0,\kappa}(f) = V_\kappa(f) \cap \mathbb{R}_{>0}^s$ . One could then try to parametrize  $V_{>0,\kappa}(f)$  but general algebraic varieties do not have rational parametrizations. This is a very uncommon property for general algebraic varieties. However, rational parametrizations do exist for the positive points of the steady state variety in certain enzymatic biochemical networks, as proved by Thomson and Gunawardena in [50]. We extended this result for many other networks of biological interest which are MESSI. Theorem 4.1 in [41] proves the existence of an explicit and algorithmically constructible rational parametrization of  $V_{>0,\kappa}(f)$  for any MESSI network  $G$  satisfying conditions (A) and (B) above. Moreover, if the partition is minimal with  $m$  subsets of core species, we have that  $\dim V_{>0,\kappa}(f) = m = s - \dim S$ .

Moreover, we identify conditions that ensure that this parametrization is monomial, or equivalently, that  $V_{>0,\kappa}(f)$  can be cut out by binomial equations (that is, polynomials with two terms) and, in this case, we give explicit binomials in Theorem 4.8 in [41] for what we call  $s$ -toric MESSI systems. Again, the conditions seem to be very restrictive, but there are plenty of interesting signaling pathways that



satisfy them; for instance the  $n$ -site phosphorylation networks and many enzymatic cascades, as the ones we presented in Section 3. In the case of the  $n$ -sequential phosphorylation network (which has  $3n + 3$  variables) we can parametrize the positive steady state variety with 3 parameters for any value of  $n$ . To compute the intersection  $V_\kappa(f) \cap S_T$  (which equals  $V_{>0,\kappa}(f) \cap S_T$  due to the absence of relevant boundary steady states, as we pointed out before), we can write 3 of the variables in terms of the remaining  $3n$  variables from the 3 conservation relations and replace them into  $3n$  linearly independent  $f_i$  (which exist in this case). We could substitute the parametrization into the conservation relations and thus get 3 equations in 3 variables. This is what makes the  $n$ -site amenable to computations even if in principle the number of variables tends to infinity with  $n$ . Note that if instead we plug in a parametrization of  $S_T$  into the equations of the steady state variety, we get a system, that besides losing sparsity, consists of  $3n$  equations in  $3n$  variables.

Recognizing the existence of a MESSI structure on a given network, checking the hypotheses in all our results and finding the rational parametrization are algorithmic and only depend on the structure and not on the particular parameters.

---

### Deciding multistationarity

The important biological mechanism of  $n$  sequential phospho-dephosphorylations has the capacity for multistationarity for  $n = 2$ , that is, there can be up to 3 positive steady states in  $V_\kappa(f) \cap S_T$  (for particular choices of the rate constants  $\kappa$  and positive linear conservation constants  $T$ ). This system has been first studied by L. Wang and E. Sontag in [51]. They proved that the maximal possible number of positive steady states is  $2n - 1$  and identified parameters for which there are  $n + 1$  positive steady states for  $n$  even (and  $n$  for  $n$  odd). Note that  $n + 1 = 2n - 1$  for  $n = 2$ . It has been proved in [36] that the upper bound  $2n - 1$  is attained for  $n = 3, 4$ , and it is probable that  $2n - 1$  is a sharp upper bound, but this has not been proven yet for  $n \geq 5$ . See also [33, 34, 35] for a discussion of other dynamical features (stability and oscillations).

In fact, the steady states of most popular MESSI systems (including all those recalled above) present an  $s$ -toric structure, and we gave in this case a characterization of the capacity for multistationarity, which lead to an algorithm based on tools from oriented matroid theory. The main ideas in this approach, which go back to [14] and several other papers, including articles in other applied areas, are collected and clarified in the paper [39]. We give below a simple version of the multistationarity results in Section 5 in [41], which is valid for other biochemical reaction networks for which the positive steady states can be defined by binomials in a parametric way and satisfying certain conditions (that we can ensure from the structure of the network, see e.g. Proposition 5.6 in [41]). In particular, these binomials are of the form  $p_\kappa = a(\kappa)x^\alpha - b(\kappa)x^\beta$ , with  $\alpha, \beta \in \mathbb{Z}_{\geq 0}^s$ , and  $a, b$  polynomial functions on the vector of rate constants  $\kappa \in \mathbb{R}_{>0}^r$  taking positive values over  $\mathbb{R}_{>0}^r$ .

Given such a binomial  $p_\kappa$ , consider the vector  $v_{p_\kappa} = a - b \in \mathbb{Z}^s$  (note that  $v_{p_\kappa} = -v_{-p_\kappa}$ , so indeed  $v_{p_\kappa}$  are integer vectors defined up to sign). Also, given a matrix  $M$  of size  $m_1 \times m_2$  of rank  $m_1$ , a subset  $J$  of indices of cardinality  $m_1$  determines a maximal minor of  $M$ , which we denote by  $M_J$ .

**Deciding mono/multistationarity:** Let  $G$  be a chemical reaction network. Denote by  $S^\perp$  a matrix whose rows define the dual of the stoichiometric subspace  $S$  with  $\text{rank}(S^\perp) = d$ . Assume that  $V_{>0,\kappa}(f)$  is cut out by  $s-d$  binomials  $p_{j,\kappa}, j = 1, \dots, s-d$ , with exponents  $v_{p_{j,\kappa}}$  which form the columns of a matrix  $B$ . Assume moreover that  $\text{rank}(B) = s-d$ . Then, the following statements are equivalent

1. Monostationarity: There is at most a single positive solution in  $V_{>0,\kappa}(F) \cap S_T$ , for any  $S$ -class intersecting the positive orthant, for any  $\kappa \in \mathbb{R}_{>0}^r$ .
2. For all subsets  $J \subseteq \{1, \dots, s\}$  of cardinality  $d$ , the product

$$(-1)^{\sum_{j \in J} j} \det(S_J^\perp) \det(B_{\{1, \dots, s\} \setminus J})$$

either is zero or has the same sign as all other nonzero products, and at least one such product is nonzero.

The previous result can be turned into an algorithm to decide if a network has the capacity for multistationarity, together with an algorithm to produce vectors of rate constants  $k$  for which multistationarity occurs (in case the network is not monostationary).

## 5 Other approaches to the question of multistationarity

The reader might have noticed that within a reasonable extension for a survey, we cannot properly define and explain all concepts. This section will then be only a pointer to some recent papers addressing the question of multistationarity, besides the articles and tools we have mentioned before. We also refer the reader to the recent survey [13] and the references therein.

Craciun, Helton and Williams applied in [15] the homotopy invariance of degree to determine the number of equilibria of biochemical reaction networks and how this number depends on parameters in the model. Conradi, Feliu, Mincheva and Wiuf give in [8] *necessary and sufficient conditions* for the multistationarity of networks having a *positive rational parametrization*, in terms of the reaction rate constants, also based on degree theory. This approach is very interesting since they can describe open multistationarity regions in rate constant space. However, it does not describe particular stoichiometric compatibility classes for which there is multistationarity, as it is also the case with the methods based on signs as the result we described about mono/multistationarity. The reason is that all these approaches are related (in more explicit or hidden ways) to properties of a Jacobian, for instance of an appropriate choice of the polynomials  $f_1, \dots, f_s$  and linear functions  $\ell_1 -$

$T_1, \dots, \ell_q - T_q$  giving equations for  $S_T$  with respect to the  $x$  variables, and so the linear conservation constants  $T_1, \dots, T_q$  do not appear. In [18] we considered extensions and simplifications of this approach via critical functions, for networks with special structure, in particular for special MESSI networks which are commonly used in modeling enzymatic pathways. We also propose a method based on the existence of triangular forms, relying on techniques from computational algebra.

Sadeghimanesh and Feliu provide in [45] a new determinant criterion to decide whether a network is multistationary, when the network obtained by removing intermediates has a binomial steady state ideal. In this case, they characterize the *multistationarity structure of the network*, i.e. which subsets of complexes are responsible for multistationarity. In particular, they compute the multistationarity structure of the  $n$ -site sequential distributive phosphorylation cycle for any  $n$ .

Together with Bihan and Giaroli, we incorporated in [6] a new tool from real algebraic geometry based on the article [7] by Bihan, Santos, and Spaenlehauer. The basic idea is the following. Given a sparse polynomial system, that is, with exponents in a specified finite set of integer points  $A$ , if it is possible to find  $p$  decorated simplices in a *regular* subdivision of  $A$ , then it is possible to scale the coefficients of the given system in an explicit way to get *at least*  $p$  nondegenerate positive real roots. This gives a lower bound on the number of positive roots. The hypotheses of regularity of the subdivision means that it comes from a lifting of the points in  $A$  after considering the projection of the domains of linearity of the lower convex hull of the lifted points. This is what gives the necessary compatibility to find a common open set in the space of coefficients where the  $p$  positive solutions can be jointly continued. The meaning that a simplex is decorated is the following. Let  $\{a_0, \dots, a_d\} \subset A$  denote the set of vertices of a maximal dimensional simplex in dimension  $d$ . Given (Laurent) polynomials  $g_1, \dots, g_d$  with support  $A$ , consider their subsums of monomials corresponding only to these exponents. So one gets a system with  $d$  polynomials in  $d$  variables and  $d + 1$  monomials of the form:

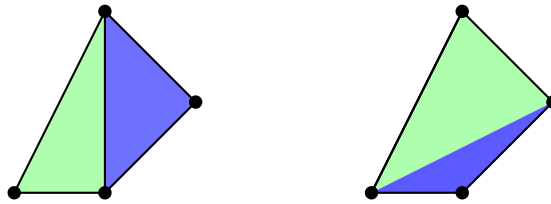
$$\sum_{j=0}^d c_j^i x^{a_j} = 0, i = 1, \dots, d.$$

This system has at most one positive root and it does have a (nondegenerate) positive root exactly when the following linear system does:

$$c_0^i + \sum_{j=1}^d c_j^i x_j = 0, i = 1, \dots, d.$$

This condition is equivalent to an alternance of signs of the minors of the  $d \times (d + 1)$  real matrix with coefficients  $c_{ij}$ . The simplex is said to be decorated by a choice of coefficients of the input polynomials when this is the case. It is interesting to note that, differently from the case of complex roots with nonzero coordinates, it is not always true that the lower bound in the case of positive solutions matches the maximum number of positive real roots for any regular subdivision. A simple

example is the following. Assume  $A = \{(0, 0), (1, 0), (1, 2), (2, 1)\}$  are the vertices of a parallelogram of Euclidean volume 2 in the plane. A sparse polynomial system ( $g_1 = g_2 = 0$ ) with this support can have  $2 \cdot 2 = 4$  isolated complex solutions with nonzero coordinates by Kouchnirenko's theorem and 3 positive solutions (and this number can be attained, see [5] and the references therein). But it is clear that the support can only have three regular subdivisions: either nothing is subdivided or we get any of the two subdivisions depicted in Figure 9, so the maximum lower bound  $p$  that one can obtain is 2. Nevertheless, this is up to now the only systematic way to find conditions on jointly on all the parameters that ensure the existence of several positive steady states, as for instance degree considerations are eventually based on parity considerations. But the best advantage of this approach is that it allows us to describe *multistationarity regions* in the space of all parameters, both reaction rate constants and linear conservation constants. Remark however that our conditions are only sufficient.



**Fig. 9** The two proper subdivisions of a circuit

We refer the reader to Section 3 in [27] for a simple example explaining the technical results in [6]. These tools allowed us to find in that article precise multistationarity regions in enzyme cascades with any number  $n$  of layers of Goldbeter-Koshland loops (with a single phosphorylation/dephosphorylation in each layer), which are multistationary as soon as the two first phosphatases are the same. Interestingly, the number of variables is of the order of  $4n$  and the dimension of the stoichiometric subspace  $S$  is of the order of  $2n$ , so it is cut out by roughly  $2n$  linear equations and parametrized by a similar number of variables. So, even taking advantage of the parametrizations of the steady state variety and a translate  $S_T$  of  $S$ , we need to deal with of the order of  $2n$  equations in  $2n$  variables. When the two layers with the same phosphatase are the last ones, it is possible to find particular multistationarity reaction rate constants for the cascade following the approach in [4]. Other papers based on the study of extrapolation of multistationarity from that of simpler subnetworks are for instance [9, 37].

In ongoing work with Giaroli, Pérez Millán and Rickster [17], we are able to use this setting to give a precise region in the space of all parameters for which the  $n$ -sequential phospho/dephosphorylation mechanism can have  $n + 1$  for  $n$  even (and  $n$  for  $n$  odd) positive steady states, assuming that only  $\frac{1}{4}$  of the intermediate complexes are part of the reactions. In another recent work Conradi, Iosif, and Kahle [10] also use tools from polyhedral geometry. They show that for reaction networks whose

positive steady states can be cut out by binomials, multistationarity is scale invariant in the space of linear conservation constants (that is, if there is multistationarity for some value of the linear concentration constants, then there is multistationarity on the entire ray containing this value (possibly for different reaction rate constants). They consider the chamber decomposition in linear conservation constant space, which allows them to show that for values of these constants in one of the five chambers the 2-site sequential phosphorylation network cannot be multistationary.

Other approaches use numeric or symbolic methods to detect points in different chambers of the complement of the discriminant and the resultants that we mentioned before, see for instance [30, 32]. The general mathematical problem is the search of positive roots of sparse polynomial systems; see for instance [21] where these techniques have been applied to a geometric problem.

### Stability and convergence

The important question of deciding stability of a given steady state  $x^*$  of a chemical reaction network with fixed constants  $k^*$  can be formalized via Routh-Hurwitz theorem by means of the satisfiability of certain polynomial inequalities which correspond to minors of the Jacobian matrix at the point  $x^*$ , as a pattern of signs of these minors corresponds to all eigenvalues of the Jacobian having negative real part. However, this is a difficult question if the point  $x^*$  is given implicitly and if one tries to trace these inequalities as the parameters vary. So, only in few cases there is a complete analysis (see for instance [33]).

Another important question is to ensure convergence of the trajectories. Note that if a trajectory defined on the whole positive real line converges for  $t \rightarrow +\infty$  to a point  $p$ , then  $p$  is a steady state. A first question is to decide global convergence in the presence of a single steady state in each  $S$ -class. We refer the reader to the results (and the references) in [19] for diverse architectures of processive multisite phosphorylation networks, which are based on previous work by Angeli, De Leener and Sontag [2].

## 6 Oscillations

Another important biological feature is the possible occurrence of oscillations. Oscillations have been observed experimentally in signaling networks formed by phosphorylation and dephosphorylation, which seems to be the main mechanism in the 24-hour period in eukaryotic circadian clocks (see for instance [43, 11] and the references therein). Despite the many articles studying sequential phospho/dephosphorylation networks, it is not currently known whether in the 2-site sequential mechanism there could be trajectories which oscillate.

Instead, Suwanmajo and Krishnan showed recently in [49] that oscillations occur intrinsically in the the dual-site phosphorylation and dephosphorylation network, in

which the mechanism for phosphorylation is processive while the one for dephosphorylation is distributive (or vice-versa), arising from a Hopf bifurcation. We also refer to the interesting paper [44], where the authors propose a systematic analysis of the long-term dynamics of phosphorylation systems. They describe bistability and oscillations when the network has nonzero levels of reaction processivity. Processivity means that the intermediate complex does not dissociate into substrate plus enzyme after a phospho/dephosphorylation, but only after two or more. Conradi, Mincheva, and Shiu showed in [11] for the mixed mechanism in [49] that in the three-dimensional space of linear conservation constants, the border between the existence of a stable or an unstable steady state is defined by the vanishing of a single Hurwitz determinant, which consists generically of simple Hopf bifurcations. Besides the Routh-Hurwitz criterion, their analysis relies on an algebraic Hopf-bifurcation criterion due to Yang and a monomial parametrization of the positive steady state variety. It would be very interesting to extend these kind of analyses to other mechanisms, in particular, in other phosphorylation networks.

Rendall and Hell studied in [34, 35] the existence of parameters for which Hopf bifurcations occur and generate periodic orbits in the case of (MAP kinase) cascades. They also explain how geometric singular perturbation theory allows to generalize results from simple models to more complex ones. Also Banaji presents in [3] some results on how oscillation is inherited by chemical reaction networks (CRNs) when they are built in natural ways from smaller oscillatory networks, showing a particularly nice result for fully open networks (where for any species  $X$ , there are reactions  $0 \rightarrow X$  and  $X \rightarrow 0$ ), also based on regular and singular perturbation theory. We also mention the pioneering work of Karin Gatermann introducing algebraic and combinatorial techniques for the search of Hopf bifurcations [25].

## 7 Mathematical challenges

In this section we enumerate some of the main open questions in this area. They involve difficult mathematical questions and moreover, systems of biological interest usually have a big number of variables and parameters.

1. Give general precise bounds for the number of positive solutions of (parametric families of) sparse polynomial systems and apply them to find the number of positive steady states: (a) develop tools to obtain better *lower bounds* for the number of positive steady states; (b) develop tools to get good *upper bounds* for the number of positive steady states. Moreover, find regions in parameter space with the predicted number of positive steady states, or at least where lower/upper bounds apply.
2. Predict or preclude oscillations from structure: how do (sustained) oscillations arise in phosphorylation networks? can we find “atoms of oscillation”? Moreover, describe “regions of oscillation” in parameter space.

## Conclusion

We can use algebro-geometric notions and methods to analyze system biology models. Algebraic and combinatorial methods allow us to predict (some) qualitative dynamic behaviours of our models from the structure of the network, without simulations and without measuring all the parameters a priori. We do have several promising results, but in many cases they tend to be too complex to be understood or computed. Answers to the above questions would require to develop a combination of tools from dynamical systems, real algebraic geometry, computational and numerical algebraic geometry, differential algebra, and biochemistry!

**Acknowledgements** I am very grateful to the Committee for Women in Mathematics of the International Mathematical Union for the organization of this meeting and for the many activities they take in charge to ensure development and recognition of female mathematicians around the world. I am particularly grateful to Carolina Araujo for all her work to make (WM)<sup>2</sup> a great success. I also thank Magalí Giaroli, Mercedes Pérez Millán and Enrique Tobis for their help with the figures.

## References

1. Angeli D., De Leenher P., and Sontag. E: A Petri net approach to the study of persistence in chemical reaction networks, *Mathematical Biosciences* 210, 598–618 (2007).
2. Angeli D., De Leenher, P., and Sontag, E.: Graph-theoretic characterizations of monotonicity of chemical networks in reaction coordinates. *J. Math. Biol.* 61, 581–616 (2010).
3. Banaji, M.: Inheritance of oscillation in chemical reaction networks. *Applied Mathematics and Computation* 325, 191–209 (2018).
4. Banaji, M., and Pantea, C.: The inheritance of nondegenerate multistationarity in chemical reaction networks. *SIAM J Appl Math* 78(2), 1105–1130 (2018).
5. Bihan F., and Dickenstein, A.: Descartes’ Rule of Signs for Polynomial Systems supported on Circuits, *Int. Math. Res. Notices* 22, 6867–6893 (2017).
6. Bihan F., Dickenstein, A., and Giaroli, M.: Lower bounds for positive roots and regions of multistationarity in chemical reaction networks. *arXiv:1807.05157* (2018).
7. Bihan, F., Santos, F., and Spaenlehauer, P-J.: A polyhedral method for sparse systems with many positive solutions. *SIAM J. Appl. Algebra Geometry* 2(4), 620–645 (2018).
8. Conradi, C., Feliu, E., Mincheva, M., and Wiuf, C.: Identifying parameter regions for multistationarity. *PLoS Comput. Biol.* 13(10), e1005751 (2017).
9. Conradi, C., Flockerzi, D., Raisch, J., and Stelling, J.: Subnetwork analysis reveals dynamic features of complex (bio)chemical networks *PNAS* 104 (49), 19175–19180 (2007).
10. Conradi, C., Iosif, A., Kahle, T.: Multistationarity in the space of total concentrations for systems that admit a monomial parametrization. *arXiv:1810.08152* (2018).
11. Conradi, C., Mincheva, M., Shiu, A: Emergence of oscillations in a mixed-mechanism phosphorylation system. *arXiv:1809.02886* (2018).
12. Conradi, C., and Shiu, A.: A global convergence result for processive multisite phosphorylation systems. *Bull. Math. Biol.* 77(1), 126–155 (2015).
13. Conradi C., and Pantea C.: Multistationarity in Biochemical Networks: Results, Analysis, and Examples. *Algebraic and Combinatorial Computational Biology*, Chapter 9, Eds. Robeva R. and Macaulay M., *Mathematics in Science and Computation*, Academic Press (2019).

14. Craciun, G., and Feinberg, M., Multiple equilibria in complex chemical reaction networks. I. The injectivity property. *SIAM J. Appl. Math.* 65(5), 1526–1546 (2005).
15. Craciun, G., Helton, J. W., and Williams, R.: Homotopy methods for counting reaction network equilibria. *Math. Biosci.* 216(2), 140–149 (2008).
16. Dickenstein, A.: Biochemical reaction networks: an invitation for algebraic geometers. *MCA 2013, Contemporary Mathematics* 656, 65–83 (2016).
17. Dickenstein, A., Giaroli, M., Pérez Millán, M., and Riechster, R.: Only  $\frac{1}{4}$  of the intermediates ensure  $n$  positive steady states in the  $n$ -site phosphorylation system. *Manuscript* (2019).
18. Dickenstein, A., Pérez Millán, M., Shiu, A., and Tang, X.: Multistationarity in Structured Reaction Networks. To appear: *Bull. Math. Biol.* (2019).
19. Eithun, M., and Shiu, A.: An all-encompassing global convergence result for processive multisite phosphorylation systems. *Math. Biosci.* 291, 1–9 (2017).
20. Érdi, P., and Tóth, J.: *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models.* Manchester University Press (1989).
21. Faugère, J.-C., Moroz, G., Rouillier, F., Safey El Din, M.: Classification of the Perspective-Three-Point problem, Discriminant variety and Real solving polynomial systems of inequalities. *ISSAC 2008 Proceedings*, D. Jeffrey (eds), Austria (Hagenberg) (2008).
22. Feinberg, M.: *Lectures on chemical reaction networks*, U. of Wisconsin (1979). available at: <http://www.che.eng.ohio-state.edu/~FEINBERG/LecturesOnReactionNetworks>.
23. Feliu, E., and Wiuf, C.: Enzyme-sharing as a cause of multi-stationarity in signalling systems. *J. R. Soc. Interface* 9 (71), 1224–1232 (2012).
24. Feliu, E., and Wiuf, C.: Simplifying biochemical models with intermediate species. *J. R. Soc. Interface* 10: 20130484 (2013).
25. Gatermann, K., Eiswirth, M., and Sensse, A.: Toric ideals and graph theory to analyze Hopf bifurcations in mass action systems, *J. Symb. Comput.* 40(6), 1361–1382 (2005).
26. Gelfand, I., Kapranov, M., and Zelevinsky, A.: *Discriminants, resultants and multidimensional determinants.* Birkhäuser Boston (1994).
27. Giaroli, M., Bihan, F., and Dickenstein, A.: Regions of multistationarity in cascades of Goldbeter-Koshland loops. *J. Math. Biol.* (2018). <https://doi.org/10.1007/s00285-018-1304-0>
28. Gnacadja, G.: Reachability, persistence, and constructive chemical reaction networks (part II): a formalism for species composition in chemical reaction network theory and application to persistence, *J. Math. Chem.* 49(10), 2137–2157 (2011).
29. Gnacadja, G.: Reachability, persistence, and constructive chemical reaction networks (part III): a mathematical formalism for binary enzymatic networks and application to persistence, *J. Math. Chem.* 49(10), 2158–2176 (2011).
30. Gross, E., Harrington, H. A., Rosen, Z., and Bernd Sturmfels, B.: Algebraic Systems Biology: A Case Study for the Wnt Pathway *Bull Math Biol.* 78(1), 21–51 (2016).
31. Gunawardena, J.: A linear framework for time-scale separation in nonlinear biochemical systems, *PLoS ONE* 7:e36321 (2012).
32. Harrington, H. A, Mehta, D., Byrne, H., and Hauenstein, J.: Decomposing the parameter space of biological networks via a numerical discriminant approach. *arXiv:1604.02623* (2016).
33. Hell, J., and Rendall, A. D.: A proof of bistability for the dual futile cycle. *Nonlin. Anal. RWA* 24, 175–189 (2015).
34. Hell, J., and Rendall, A. D.: Sustained oscillations in the MAP kinase cascade. *Math.. Biosci.* 282, 162–173 (2016).
35. Hell, J., and Rendall, A. D.: Dynamical Features of the MAP Kinase Cascade. In: Graw F., Matth´aus F., Pahle J. (eds) *Modeling Cellular Systems. Contributions in Mathematical and Computational Sciences*, vol 11. Springer, Cham (2017).
36. Holstein, K., Flockerzi, D., and Conradi, C.: Multistationarity in sequential distributed multisite phosphorylation networks. *Bull. Math. Biol.* 75(11), 2028–2058 (2013).



37. Joshi, B., and Shiu, A.: Which small reaction networks are multistationary? *SIAM J. Appl. Dyn. Syst.* 16(2), 802–833 (2017).
38. Mirzaev L., and Gunawardena J.: Laplacian dynamics on general graphs, *Bull. Math. Biol.* 75, 2118–2149 (2013).
39. Müller, S., Feliu, E., Regensburger, G., Conradi, C., Shiu, A., and Dickenstein, A.: Sign conditions for injectivity of generalized polynomial maps with appl. to chemical reaction networks and real algebraic geometry, *Found. Comput. Math.* 6(1), 69–97 (2016).
40. Patel, A.L., and Shvartsman, S.Y.: Outstanding questions in developmental ERK signaling. *Development* 145(14) (2018).
41. Pérez Millán, and Dickenstein, A.: The structure of MESSI biological systems. *SIAM J. Appl. Dyn. Syst.* 17(2), 1650–1682 (2018).
42. Pérez Millán, M., Dickenstein, A., Shiu, A., and Conradi, C.: Chemical reaction systems with toric steady states. *Bull. Math. Biol.* 74(5) 1027–1065 (2012).
43. Qiao, L., Nachbar, R. B., Kevrekidis, I. G., and Shvartsman, S. Y.: Bistability and oscillations in the Huang-Ferrell model of MAPK signalling. *PLoS Comp. Biol.* 3, 1819–1826 (2007).
44. Rubinstein, B., Mattingly, H., Berezhkovskii, A., and Shvartsman, S.: Long-term dynamics of multisite phosphorylation. *Mol. Biol. Cell* 27(14), 2331–2340 (2016).
45. Sadeghimanesh, A.H., and Feliu, E.: The multistationarity structure of networks with intermediates and a binomial core network. arXiv:1808.07548 (2018).
46. Safey El Din, M.: RAGlib library, available at: <https://www-polsys.lip6.fr/~safey/RAGLib/>.
47. M. Sáez, C. Wiuf, E. Feliu: Graphical reduction of reaction networks by linear elimination of species. *J. Math. Biol.* 74(1-2), 195–237 (2017).
48. Shinar, G., and Feinberg, M.: Structural sources of robustness in biochemical reaction networks, *Science* 327(5971) 1389–1391 (2010), .
49. Suwanmajo, T., and Krishnan, J.: Mixed mechanisms of multi-site phosphorylation *Journal of The Royal Society Interface* 12(107) (2015). <http://doi.org/10.1098/rsif.2014.1405>
50. Thomson, T., and Gunawardena, J.: The rational parameterisation theorem for multisite post-translational modification systems. *J. Theoret. Biol.* 261(4), 626–636 (2009).
51. Wang, L., Sontag, E.: On the number of steady states in a multiple futile cycle. *J. Math. Biol.* 57(1), 29–52 (2008).