

Automatic model generation*

Tibor Nagy^{1,2} , János Tóth^{2,3} , Tamás Ladics⁴

¹Institute of Materials and Environmental Chemistry,
Research Centre for Natural Sciences,
Hungarian Academy of Sciences, Budapest, Hungary

²Laboratory for Chemical Kinetics,
Eötvös Loránd University, Budapest, Hungary

³Department of Mathematical Analysis
Budapest University of Technology and Economics,
Budapest, Hungary

⁴Department of Science and Engineering,
John von Neumann University, Kecskemét, Hungary

Correspondence Tibor Nagy,
Institute of Material and Environmental Chemistry,
Research Centre for Natural Sciences,
Hungarian Academy of Sciences,
1117 Budapest, Magyar tudósok körútja 2, Hungary

Email : nagy.tibor@ttk.mta.hu

Funding information

National Research Development and Innovation Office,
SNN 125739.

April 3, 2019

Abstract

*Dedicated to Professors László Lovász and József Pálinkás.

The goal of the paper is to automatize the selection of mechanisms which are able to describe a set of measurements. In order to do so first we construct a set of possible mechanism fulfilling chemically reasonable requirements with a given number of species and reaction steps. Then we try to fit all the mechanisms, and offer the best fitting one to the chemist for further analysis. The method can also be used to a kind of lumping: to reproduce the results of a big mechanism using a smaller one, with less number of species. We show two applications: one on an artificial example and another one on a small real life data.

KEYWORDS reaction networks, inverse problem of reaction kinetics, automatically generated models, reduced models, model parametrization, fitting rate coefficients

1 INTRODUCTION

Our goal is to help solve an inverse problem of reaction kinetics: we try to build "reaction mechanism templates", candidate set of reactions which are able to serve as models for given (measured or simulated with a detailed model) concentration vs. time curves. This we do by constructing a large set of reaction mechanisms then by discarding those which do not fulfil some chemically relevant restrictions. To put it in another way, we provide a set of possible models automatically for the chemist, and say, which of them fits best to the measurements or the simulation measurements of a detailed model. The approach of the Ref. [14] is quite similar, but (almost) all the details are different. Their goal is to construct possible cellular reaction networks.

Thus, given are: concentrations as functions of time, and our task is to find a *model*, a *reaction mechanism*, i.e. a set of reaction steps (also called complex chemical reaction, reaction network, or simply reaction), endowed with mass action type kinetics and with appropriate reaction rate coefficients from the set of *all* possible models. Concentrations used as reference are either determined from measurements or from simulations of a detailed model. Here we show how the method works by testing mainly on *simulated* data, and also a set of real experimental data: measurements on the salicylic acid transport under different conditions..

A related problem is to find *good initial estimates* of the parameters^[13], in order to accelerate parameter fitting, which needs to be done for a large

number of candidate models.

The structure of our paper is as follows. Section 2 offers a possible scenario to construct candidate sets of reactions which are to describe the concentration vs. time functions. Section 3 shows the applicability of the method on two examples. In the first one, a specific model has been chosen from the set of models constructed according to Section 2; reference data are generated by its simulation. In the second one we use experimental data on salicylic acid transport from the stomach to the intestine with or without addenda measured at different pH values. Beyond the original model (the classical irreversible consecutive reaction) other mechanisms are also tried. The last Section is about possible extensions and formulates some open problems, mainly in combinatorics.

2 A SET OF CHEMICALLY REASONABLE RESTRICTIONS

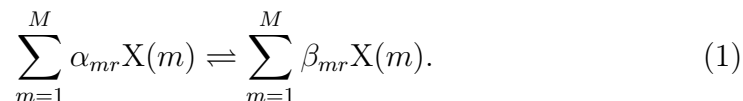
Basic concepts of formal reaction kinetics are presented here shortly, for the formal and detailed expansion we propose the use of e.g. Refs. [22] or [5].

2.1 Species

First, we have to fix the number of species M . In our illustrating examples it will usually be 2 or 3; in applications this may be equal to the number of measured concentrations.

2.2 Complexes

In formal reaction kinetics the linear combinations of species with stoichiometric coefficients on the sides of the reaction arrows are called *complexes*, a slightly unfortunate name, because this word is used in chemistry with a completely different meaning. Their number is usually denoted by N . The stoichiometric coefficients of the m^{th} species on the left side of the r^{th} reaction arrow is denoted by α_{mr} , those on the right side by β_{mr} , thus the general form of our mechanism is



Mass conservation In systems where chemical reactions significantly modify the mole fraction of species (e.g. the changes are larger than 1%), mass conservation has to be strictly fulfilled. This is the case in combustion models, and we shall accept this view here. Whereas atmospheric chemistry models usually contain steps from which major components of air (e.g. N_2 , O_2 , CO_2 in tropospheric chemistry models) are simply omitted if formed as negligible change in their concentration is induced by the step. The number of steps is less if only mass conserving reactions are allowed.

As a consequence of mass conservation $M = 1$ is immediately excluded, because e.g. the reaction step $\text{X} \rightleftharpoons 2\text{X}$ is not allowed. Also excluded is the empty complex 0, as it immediately leads to mass destruction or mass creation. Still, the empty complex may either be useful to describe the sticking of a species to the wall or leave the reactor in any other way, or to express an inflow or outflow in formal reactions, e.g. $0 \longrightarrow \text{X}$ or $\text{X} + \text{Y} \longrightarrow 0$.

Short complexes In most cases reaction steps containing complexes longer than 2, are not accepted; here we accept this view. In other words, we only allow *short complexes*, complexes of the form in which no more than two species take part. In the case when $M = 2$ we have thus $\text{X}, \text{Y}, 2\text{X}, 2\text{Y}, \text{X} + \text{Y}$ (zero complex excluded from the beginning). Their number in the general case, as it can be immediately seen, is

$$\bar{N}(M) = M + M + \binom{M}{2} = \frac{M(M+3)}{2}. \quad (2)$$

Let us mention in passing that the concept of short complexes proved to be really useful when systematically investigating the dynamic behaviour of small reaction mechanisms^[12]. However, our present restriction does not mean that termolecular reactions are unimportant. In gas phase reactions like



are quite common, see e.g. <https://engineering.columbia.edu/news/michael-burke-termolecular-reactions>. To take into consideration these one should also include $3\text{X}, 2\text{X} + \text{Y}, \dots, \text{X} + \text{Y} + \text{Z}$, the total number of which is

$$M + M(M-1) + \binom{M}{3} = \frac{M(M+1)(M+2)}{6} = \binom{M+2}{3}.$$

2.3 Reaction steps

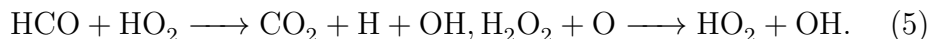
Reversibility All elementary reactions are strictly reversible due to microscopic reversibility. However, in some cases the rate of the forward or backward reaction can be negligible for the concentrations and the conditions (temperature and pressure) arising during in practical applications, thus this direction can be omitted and the kept direction can be considered as an irreversible step without inducing significant error.

Accordingly, in widely accepted combustion models irreversible steps are used, nevertheless, during model construction we assume that the reaction steps are reversible and the omission one of the directions can be investigated a posteriori.

- Hydrogen combustion:^[10]



- Carbon monoxide combustion:^[10]

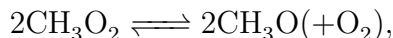


Further examples e.g. in methanol combustion (^[1]) can also be found.

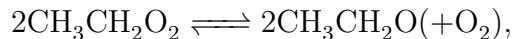
$\text{X} \rightleftharpoons \text{X}$, $2\text{X} \rightleftharpoons 2\text{X}$, $\text{X} + \text{Y} \rightleftharpoons \text{X} + \text{Y}$ Certainly there is no reason to include reaction steps which do not affect concentrations (i.e. make no macroscopic change) despite that they are taking place microscopically.

Mass conservation We discard steps like $\text{X} \rightleftharpoons 2\text{X}$, $\text{X} \rightleftharpoons \text{X} + \text{Y}$, as they obviously violate the law of mass conservation. However, these kinds of steps may be quite useful in model construction in chemical kinetics. E.g. the Lotka–Volterra reaction which was proposed for *approximately* describing oscillations in cold flames^[7], contains steps like $\text{X} \longrightarrow 2\text{X}$ and $\text{Y} \longrightarrow 0$.

$2\text{X} \rightleftharpoons 2\text{Y}$ In models one can rarely see such a step. However, studying the literature one can find similar steps: the self-reaction of peroxy radicals, like



or



Species	X, Y	$M = 2$
Complexes	X, Y, 2X, 2Y, X + Y	$\overline{N}(2) = 5$
Reaction steps	$X \rightleftharpoons Y, X \rightleftharpoons 2Y, Y \rightleftharpoons 2X,$ $2X \rightleftharpoons X + Y, 2Y \rightleftharpoons X + Y$	$\overline{R}(2) = 5$

Table 1: Species, complexes and reaction steps in the case $M = 2$

see Refs. [11] and [16]. One can omit the forming O_2 in atmospheric (e.g. tropospheric) chemistry as it is a major constituent of air and its concentration will be negligibly affected by this and similar transformations.

Both steps are of the form $2X \rightleftharpoons 2Y (+Z)$ (and are mass conserving—if O_2 is taken into consideration—as they fulfill the law of atomic balance).

Reactions passing all the criteria

$M = 2$ The species, complexes and reaction steps are shown in the Table below.

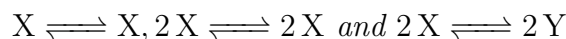
$M = 3$ The species, complexes and reaction steps are shown in the Table below.

Now the interesting (from the combinatorial point of view) question arises: What is the number of reactions fulfilling the requirements formulated above? Beyond combinatorics, formulae to give those numbers is of practical relevance too: it gives us a hint if it is possible to deal with all the systems constructed in this way within a tolerable time. We enumerated the steps from $M = 2$ up to $M = 20$ and have found the following cardinalities: 5, 24, 69, 155, 300, 525, 854 How to learn if there is a certain regularity in the sequence? The best way is to go to the Online Encyclopedia of Integer Sequences (<https://oeis.org/>) initiated by Neil James Alexander Sloane, and ask if it contains our sequence. In this case the answer was yes, and the formula $(M - 1)M(M^2 + 7M + 2)/8$ is provided to give the number of reaction steps of the given type. From the strict mathematical point of view now we had this statement as a conjecture, but it can rigorously be proved, as well.

Species	X, Y, Z	$M = 3$
Complexes	X, Y, Z, 2X, 2Y, 2Z, X + Y, Y + Z, Z + X	$\overline{N}(3) = 9$
Reaction steps	X \rightleftharpoons Y, X \rightleftharpoons Z, Y \rightleftharpoons Z, X \rightleftharpoons 2Y, X \rightleftharpoons 2Z, Y \rightleftharpoons 2X, Y \rightleftharpoons 2Z, Z \rightleftharpoons 2X, Z \rightleftharpoons 2Y, X \rightleftharpoons Y + Z, Y \rightleftharpoons X + Z, Z \rightleftharpoons X + Y, 2X \rightleftharpoons X + Y, 2X \rightleftharpoons X + Z, 2X \rightleftharpoons Y + Z, 2Y \rightleftharpoons X + Y, 2Y \rightleftharpoons X + Z, 2Y \rightleftharpoons Y + Z, 2Z \rightleftharpoons X + Y, 2Z \rightleftharpoons X + Z, 2Z \rightleftharpoons Y + Z, X + Y \rightleftharpoons X + Z, X + Y \rightleftharpoons Y + Z, X + Z \rightleftharpoons Y + Z	$\overline{R}(3) = 24$

Table 2: Species, complexes and reaction steps in the case $M = 3$

Statement 1 *Suppose the number of species is $M \in \mathbb{N}$. Then, the number of reversible, mass conserving reaction steps excluding steps of the form*



is

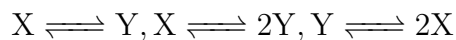
$$\overline{R}(M) := (M - 1)M(M^2 + 7M + 2)/8, M = 2, 3, \dots \quad (6)$$

Proof. The table below shows how many steps of the different types we have. Below we will have other sequences of numbers to be studied.

2.4 Reaction mechanisms

A set of reaction steps is usually called a (kinetic reaction) *mechanism*. In formal reaction kinetics alternative names (complex chemical) *reaction*, or *reaction network* are also used.

The actual number of species is M We may start from three species and arrive at a mechanism with two species, as e.g.



Type of step	$M =$	2	3	4
$X \rightleftharpoons Y$	$\frac{M(M-1)}{2}$	1	3	6
$X \rightleftharpoons 2Y$	$M(M-1)$	2	6	12
$2X \rightleftharpoons X+Y$	$M(M-1)$	2	6	12
$X+Y \rightleftharpoons Z$	$\frac{M(M-1)}{2}(M-2)$	0	3	12
$X+Y \rightleftharpoons 2Z$	$\frac{M(M-1)}{2}(M-2)$	0	3	12
$X+Y \rightleftharpoons X+Z$	$\frac{M(M-1)}{2}(M-2)$	0	3	12
$X+Y \rightleftharpoons Z+A$	$\frac{1}{2} \frac{M(M-1)}{2} \frac{(M-2)(M-3)}{2}$	0	0	3
Total	$\bar{R}(M)$	5	24	69

Table 3: Number of reaction steps of different type

Number of steps	1	2	3	4	5	6	≥ 6
$M = 3$	9	246	1994	10611	42501	134596	$\bar{R}(3)$
$M \leq 3$	24	276	2024	10626	42504	134596	$\left(\bar{R}(3)\right)_R$

Table 4: Number of mechanisms containing different number of reaction steps

may be obtained as one of the mechanisms with three species, but it actually contains only two species. It would be desirable to exclude such cases.

Suppose we have $M = 3$ species, then the number of mechanisms $\bar{R}(3)$ with not more than three species, and also the number of mechanisms $\tilde{R}(3)$ with exactly three species are shown in the Table below.

In the case when we have at most three species it may happen that we only have 2, thus the mechanisms with exactly 3 species is $\tilde{R}(3) = \left(\bar{R}(3)\right)_R - \left(\bar{R}(2)\right)_R \binom{3}{2}$ because of the three species one can select any two in $\binom{3}{2}$ different ways. As $\bar{R}(2) = 5$, last (and following) elements in the two rows of the table coincide, as one can only have 5 reaction steps with two species, see Table 1. Let us formulate the corresponding—obvious—general statement.

Statement 2 *Suppose the number of species is $M \in \mathbb{N}$. Then, the number of reversible, mass conserving mechanisms excluding steps of the form $X \rightleftharpoons X$, $2X \rightleftharpoons 2X$ and $2X \rightleftharpoons 2Y$ which do contain M species and consist of*

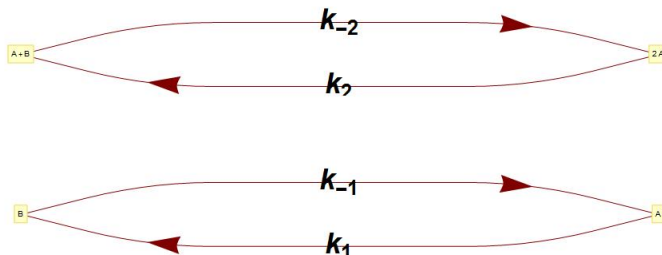


Figure 1: The Wegscheider mechanism

R reactions is

$$\tilde{R}(M) := \binom{\bar{R}(M)}{R} - \binom{\bar{R}(M-1)}{R} \binom{M}{M-1}. \quad (7)$$

Consequently, in case $R > \bar{R}(M-1)$ one only has the first term.

Mass conservation Even if the steps are mass conserving the mechanism may not be as the example $\{X \rightleftharpoons Y, Y \rightleftharpoons 2X\}$ shows. To check this property is not a trivial problem, we do this using our program `ReactionKinetics` described in Chapter 4 of Ref [22], where the reader can also find relevant references, as well.

Detailed balancing Following Section 7.8 of Ref. [22] we review the history of detailed balancing shortly.

After such men as Maxwell and Boltzmann, and before Einstein (see the references here: https://en.wikipedia.org/wiki/Detailed_balance), at the beginning of the twentieth century, it was Wegscheider^[26] who constructed the reaction mechanism in Fig. 1 to show that in some cases the existence of a positive stationary state alone does not imply the equality of all the individual forward and backward reaction rates in equilibrium: A relation (in this case $k_{-1}k_2 = k_{-2}k_1$) should hold between the reaction rate coefficients to ensure this. Equalities of this kind will be called (and later exactly defined) as *spanning forest conditions* below. Let us emphasize that violation of this equality between the reaction rate coefficients does not exclude the existence of a positive stationary state; it can be shown to exist and unique for all values of the reaction rate coefficients. (Problem 7.12 of Ref. [22] proves both statements.)

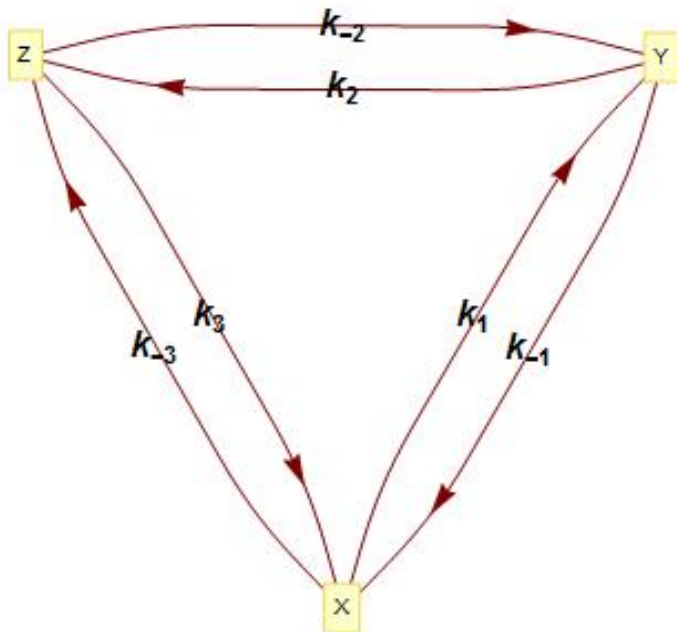


Figure 2: The reversible triangle reaction

Here we mention that Figs. 1–3. show the Feinberg–Horn–Jackson graph of the mechanism which is a directed graph with the complexes as vertices and with the reaction step arrows as directed edges. The number of complexes is denoted by N , the number of connected components of the Feinberg–Horn–Jackson graph is L , whereas the number of independent reaction steps (the rank of γ) is S .

A similar statement holds for the reversible triangle reaction in Fig. 2. The necessary and sufficient condition for the existence of such a positive stationary state for which all the reaction steps have the same rate in the forward and backward direction (a detailed balanced stationary state) is now that the product of the reaction rate coefficients is the same if taken in either direction: $k_1 k_2 k_3 = k_{-1} k_{-2} k_{-3}$. Equalities of this kind will be called as *circuit conditions* below. Again, violation of this equality does not exclude the existence of a positive stationary state; it can again be shown to exist and be unique for all values of the reaction rate coefficients. (Problem 7.11

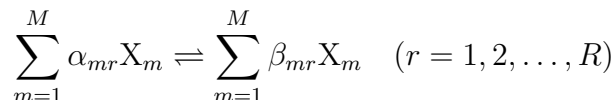
of the reference mentioned shows both statements.)

These examples are qualitatively different from, e.g., the simple one-step mechanism $X + Y \xrightleftharpoons[k_{-1}]{k_1} Z$ which has the same stationary reaction rate in both directions no matter what the values of the reaction rate coefficients are, see Problem 7.10. To put it another way, this mechanism is *unconditionally detailed balanced* while the previous examples (being detailed balanced only if some equalities are fulfilled) were *conditionally detailed balanced*. To show a less trivial example, have a look at Fig. 3.

A quarter of a century after Wegscheider, the authors Fowler and Milne^[6] formulated in a very vague form a general principle called the *principle of detailed balance* stating that in real thermodynamic equilibrium, all the sub-processes should be in dynamic equilibrium separately in such a way that they do not stop but proceed with the same velocity in both directions. Obviously, this also means that time is reversible at equilibrium; that is why this property may also be called *microscopic reversibility*, although it may be appropriate to reserve this expression for a similar property of the stochastic model (see Chap. 10 of Ref. [22]). A relatively complete summary of the early developments was given by Tolman^[20]. The modern formulation of the principle accepted by IUPAC^[8] essentially means the same (given that the principle of charity is applied when reading): “The principle of microscopic reversibility at equilibrium states that, in a system at equilibrium, any molecular process and the reverse of that process occur, on the average, at the same rate.”

Now we give a precise formulation of the concept in such a way that at a detailed balanced stationary point (which can only exist in a reversible reaction), all pairs of reaction–antireaction step pairs proceed with the same rate in both directions.

The reaction we study here consists of reversible pairs of reaction steps like



and their usual induced kinetic differential equations assuming mass action type kinetics (and disregarding the change of temperature, pressure and re-

action volume) is

$$\dot{c}_m = \sum_{r=1}^P (\beta_{mr} - \alpha_{mr}) (k_r^+ \prod_{p=1}^M c_p^{\alpha_{pr}} - k_r^- \prod_{p=1}^M c_p^{\beta_{pr}}),$$

where $c_m(t) := [X_m](t)$ is the concentration of species X_m . (Note that here R denotes the half of the number of reaction steps, or the number of reversible pairs; we reserve the notation P for the total number of reaction steps. Thus, in the case of reversible steps $P = 2R$.) Shortly,

$$\dot{\mathbf{c}} = \boldsymbol{\gamma}(\mathbf{k}^+ \odot \mathbf{c}^\alpha - \mathbf{k}^- \odot \mathbf{c}^\beta)$$

with $\boldsymbol{\alpha} := (\alpha_{mr})$, $\boldsymbol{\beta} := (\beta_{mr})$, $\boldsymbol{\gamma} := \boldsymbol{\beta} - \boldsymbol{\alpha}$, $\mathbf{c}^\alpha := \prod_{m=1}^M c_m^{\alpha_{mr}}$, and with the componentwise (or Schur) product \odot of vectors. Here the positive numbers k_r^\pm are the *reaction rate coefficients*, the vectors formed from them are \mathbf{k}^\pm . The reaction is detailed balanced at the positive stationary concentration \mathbf{c}_* if all the steps proceed with the same rate in both directions, or, to put it another way

$$\text{DB: } \boldsymbol{\gamma}(\mathbf{k}^+ \odot \mathbf{c}_*^\alpha - \mathbf{k}^- \odot \mathbf{c}_*^\beta) = \mathbf{0} \text{ implies}$$

$$\mathbf{k}^+ \odot \mathbf{c}_*^\alpha = \mathbf{k}^- \odot \mathbf{c}_*^\beta \text{ or } \boldsymbol{\gamma}^\top \log(\mathbf{c}_*) = \log(\mathbf{K}),$$

where $\mathbf{K} := \frac{\mathbf{k}^+}{\mathbf{k}^-}$.

Detailed balance may hold

- at any (positive) values of the reaction rate coefficients (unconditionally detailed balanced), or
- only if the values of the rate coefficients fulfil certain conditions—e.g. circuit or spanning tree conditions—(conditionally detailed balanced).

What are the necessary and sufficient conditions of this property? First we give an algebraic characterization that can be proved using Fredholm's alternative theorem.

Theorem 1 (See Refs. [22, 25]) *The reaction is detailed balanced, if and only if for all nonzero vector solutions to the system of linear equations $\boldsymbol{\gamma}\mathbf{a} = \mathbf{0}$ one has*

$$\mathbf{K}^{\mathbf{a}} = \mathbf{1}. \tag{8}$$

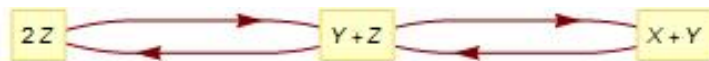


Figure 3: No matter what the values of the reaction rate coefficients are, this is detailed balanced.

As the elements of γ are integers and the vectors \mathbf{a} are solutions of a homogeneous linear equations, their coordinates can supposed to be integers.

Next we cite a pair of structural criteria showing what the reasons of detailed balancing are. To formulate this, we need a few concept and also a few formal definitions.

Definition 1 *The circuit conditions are that the product of reaction rate coefficients along any set of independent cycles is the same in both directions.*

Definition 2 *Let us take a spanning forest of the Feinberg–Horn–Jackson graph, and let the corresponding reaction step vectors be $\gamma_{.,u}$ ($u = 1, 2, \dots, N-L$). Then, $\sum_u^{N-L} a_u \gamma_{.,u} = \mathbf{0}$ has $N-L-S$ independent solutions. With these $\prod \frac{k_u^+}{k_u^-} = 1$ should hold: these are the spanning forest conditions.*

Note that the number of the edges of the spanning *tree* is L less than the number of its vertices, if again L is the number of the connected components of the Feinberg–Horn–Jackson graph.

Theorem 2 (Ref. [4]) *The mechanism is detailed balanced, if and only if the circuit conditions and the spanning forest conditions hold.*

An application of Feinberg’s theorem (and also the detailed description with examples of the concepts) can be found in Ref. [15].

Example 1 *An unconditionally detailed balanced reaction can be seen in Figure 3. The reason is that both structural conditions are empty:*

1. *It does not contain cycles.*
2. *Its deficiency ($:= N - L - S = 3 - 1 - 2$) is zero, thus no spanning forest conditions are to be considered.*

With the approach of applying Theorem 1 one sees that $\gamma \mathbf{a} = \mathbf{0}$ has no nonzero solutions, or the kernel of the linear map γ only contains the zero vector.

Example 2 A conditionally detailed balanced mechanism is shown in Figure 4. Here the spanning forest conditions (their number is $N-L-S = 5-2-1 = 2$) are as follows.

$$k_{-1}^2 k_3 = k_{-3} k_1^2, \quad k_{-1} k_2 = k_{-2} k_1. \quad (9)$$

We also have the circuit conditions:

$$k_{-4} k_{-2} k_3 = k_{-3} k_2 k_4. \quad (10)$$

Let us try to discuss unconditionally and conditionally detailed balanced reactions in a more systematic way.

If $\boxed{M = 2, R = 2}$, then the following three mass conserving mechanisms remain.



They are all conditionally detailed balanced. (One can easily show that with $M = 2$ and 2 reversible steps there are no unconditionally detailed balanced mechanisms.) They contain no circles, thus only the *spanning forest conditions* should hold, and these are shown in the second column above.

If $\boxed{M = 3, R = 2}$, there are 18 conditionally detailed balanced, 189 unconditionally detailed balanced mechanisms. Let us calculate the number of conditionally and unconditionally detailed balanced reactions for the cases $M = 3, 4$. Then, the following table results, where **MC** denotes the total number of mass conserving mechanisms, **UDB** is for unconditionally detailed balanced mechanisms, and **CDB** stands for conditionally detailed balanced mechanisms: **MC=UDB+CDB**

How to use detailed balance in our calculations? We have two choices. One can either fit using independent reaction rate coefficients and check detailed balance later, or use detailed balance as a condition when fitting. In our calculations we have used the second alternative. Let us mention that our

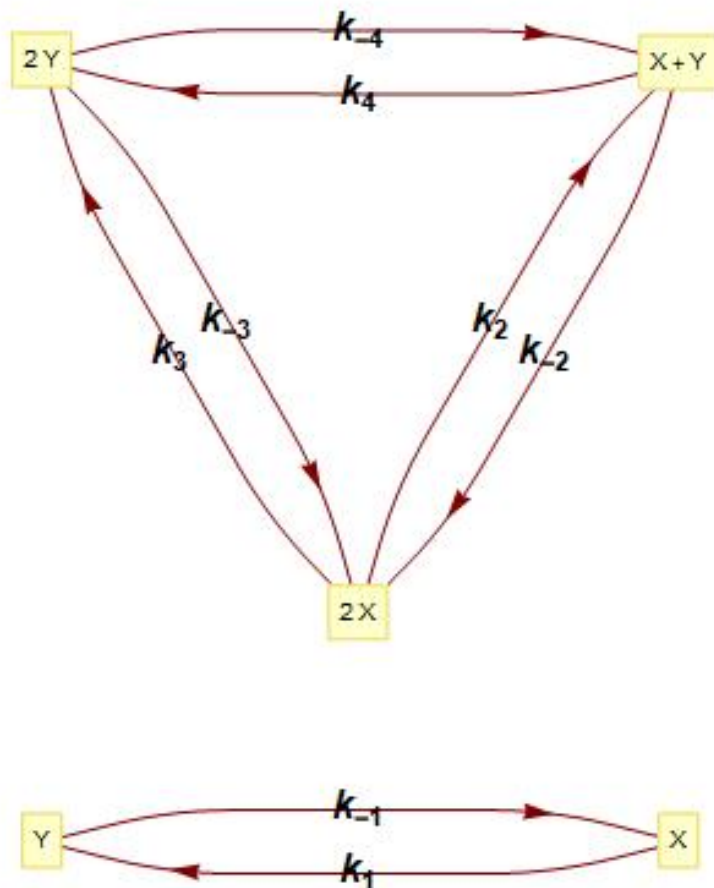


Figure 4: Detailed balanced, if the system of equalities (9) and (10) holds.

	$R = 1$	2	3
$M = 3$	24 $24=24+0$	276 $207=189+18$	2024 $602=0+602$
4	69 $69=69+0$	2346 $2064=2004+60$	52394 $30768=2478+5898$

Table 5: The number of not mass conserving and of mass conserving and either unconditionally detailed balanced or conditionally detailed balanced mechanisms

calculations heavily rely on the program package `ReactionKinetics` written in Wolfram language (Mathematica) and downloadable from `extras.springer.com` using the ISBN number 978-1-4939-8641-5. This package is aimed at helping the chemist to do many kinds of symbolic and numerical investigations of reaction mechanisms including solving the induced kinetic differential equations or simulating the usual stochastic model, but excluding parameter estimation. The codes written to the present paper will be provided to the reader upon request.

3 APPLICATIONS OF THE METHOD

Although to illustrate the method we use toy models, we think the results are promising. Two applications of the method will be shown.

1. Of the three reversible mechanisms constructed with $M = 2$ species and $R = 2$ pairs of reaction steps above we simulated data for one of them, and the program fitted all three candidate mechanisms and identified the best of them. As they are inherently conditionally detailed balanced, detailed balance was used as a constraint, a prescribed relationship between the reaction rate coefficients during fitting.
2. *Real experimental data:* reliable old measurements^[17] on the transport of salicylic acid with or without different additives under different pH values could be explained by the simple consecutive mechanism: $X \longrightarrow Y \longrightarrow Z$. Note that there are no chemical reactions involved, and the same species in different compartments are denoted as different species and their transport is described as formal reactions.

Now we tried to fit all the six models $X \longrightarrow Y \longrightarrow Z$, $X \longrightarrow Z \longrightarrow Y$, etc. generated by permuting the order of the compartments.

3.1 Simulated data in the $M = 2, R = 2$ case

We solved the deterministic model

$$\begin{aligned}x'(t) &= \frac{k_{-1}k_1}{k_2}x(t)y(t) - k_2x(t)^2 - k_1x(t) + k_{-1}y(t), \\y'(t) &= -\frac{k_{-1}k_1}{k_2}x(t)y(t) + k_2x(t)^2 + k_1x(t) - k_{-1}y(t)\end{aligned}$$

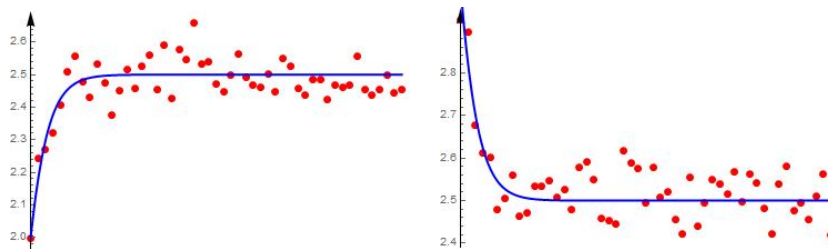


Figure 5: The original model: $X \xrightleftharpoons[k_{-1}]{k_1} Y, 2X \xrightleftharpoons[k_{-2}]{k_2} X + Y$

of the reaction $X \xrightleftharpoons[0.1]{0.1} Y, 2X \xrightleftharpoons[k_{-2}]{1} X + Y$ with the given reaction rate coefficients and k_{-2} calculated from the condition of detailed balance (11). The initial concentrations were $x(0) = 2, y(0) = 3$. Then, we took a sample from the concentrations at equidistant discrete times with a sampling step size 0.1 and added a 2% relative error. Next, we tried to fit the model of the reaction from which we started. We have chosen the initial estimates of the parameters to be 0.5, 0.5, 0.5. the estimated parameters became

$$k_1 = 0.0971753, k_{-1} = 0.0971771, k_2 = 0.999665.$$

Figure 5 suggests a good fitting.

In the second case everything remained unchanged except that the reaction to be fitted was $2X \rightleftharpoons X + Y, 2Y \rightleftharpoons X + Y$. Note that condition of detailed balance is different from the first case again, see (13). With the same data neither the numerical result, nor the fitting seems to be too bad (although slightly worse:

$$k_1 = 0.0741455, k_{-1} = 0.0741468, k_2 = 1.01966.$$

see also Figure 6.

Finally, the reaction to be fitted was $X \rightleftharpoons Y, 2Y \rightleftharpoons X + Y$. Note that condition of detailed balance is different in this case, see (12). Again, with the same data the fitting does not seem to be too bad, but the estimated parameter values are unacceptable:

$$k_1 = -0.357567, k_{-1} = -0.357585, k_2 = 1.16416.$$

see also Figure 7. (One can force the program to provide positive reaction rate coefficients only, but in the first two cases this was not needed.)

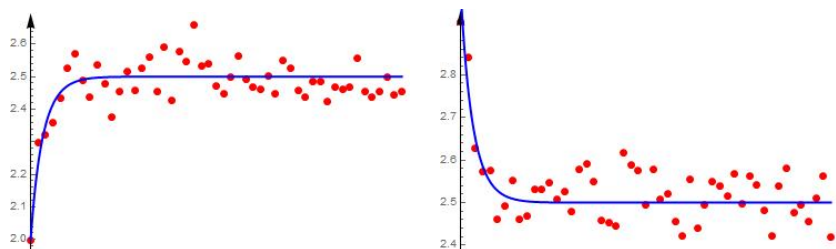


Figure 6: A slightly worse model: $2X \rightleftharpoons X + Y$, $2Y \rightleftharpoons X + Y$

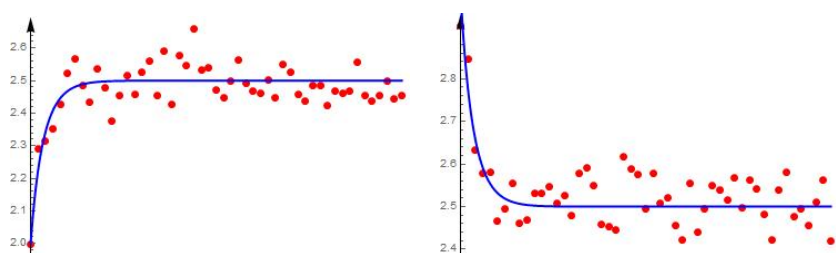


Figure 7: An unacceptable model: $X \rightleftharpoons Y$, $2Y \rightleftharpoons X + Y$

3.2 Salicylic acid transport with $M = 3, R = 2$ case

The concentrations of salicylic acid were measured in two compartments X (gastric fluid) and Z (intestine fluid), assuming that the process is described with the simple consecutive reaction $X \longrightarrow Y \longrightarrow Z$. The concentrations measured in the different compartments can be seen in the Table.

t_i (hours)	$x(t_i)$ (M/l)	t_i (hours)	$z(t_i)$ (M/l)
1	0.01579	0	0
2	0.01429	1	0.0003
3	0.01327	2	0.000614
4	0.01230	3	0.000917
5	0.01148	4	0.00143
6	0.01066	5	0.00201
7	0.00988	6	0.00269
8	0.00912	7	0.00338
9	0.00851	8	0.00402
10	0.00791	9	0.00473

Fitting the mechanism $X \xrightarrow{k_1} Y \xrightarrow{k_2} Z$ gives nice results: starting from the initial estimates (1, 1) for the reaction rate coefficients (more precisely,

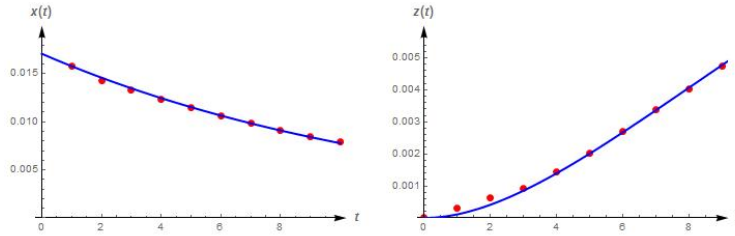


Figure 8: Fitting a bad model to real life data

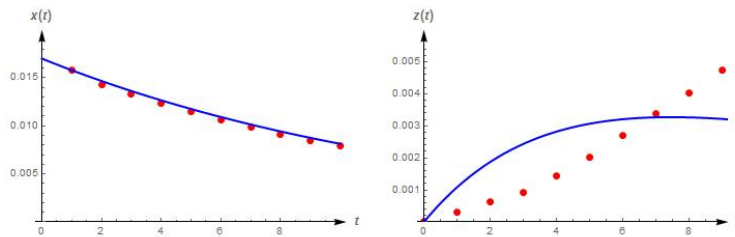


Figure 9: Fitting to real life data

transport coefficients) gives the results with small standard error $\{k_1 \rightarrow 0.0786482 \pm 0.000812339, k_2 \rightarrow 0.181337 \pm 0.00527615\}$, a "good" correlation matrix

$$\begin{bmatrix} 1. & -0.499719 \\ -0.499719 & 1. \end{bmatrix}.$$

The fitting can also be seen to be good, see Fig. 8.

However, if one wants to fit erroneously the model $X \rightarrow Z \rightarrow Y$, then the Z species fits badly, as seen in Fig. 9. Let us remark, that the standard errors are also larger than previously, which is not too interesting if one keeps in mind that the transport coefficients of the bad model are meaningless.

4 CONCLUSIONS, OUTLOOK

At the end we make a few remarks on the restrictions used above, and also, on the final use of the results obtained from the type of calculations we have proposed.

4.1 Restrictions

We have shown two sets of reactions, but fitting to the given problem one can formulate specific sets of restrictions taking into consideration the chemical nature of the problem and trying to reduce the number of possible candidate models.

Mass action type kinetics We have almost always used it: especially in connection with detailed balance.

Complex balancing It is quite natural to ask why do we not select *complex balanced reactions*, because this concept turned out to be more fundamental during the development of formal reaction kinetics^[5]. Because in the case of mass action type kinetics we have a nice (structural) necessary and sufficient condition (the reaction should be weakly reversible and should have a zero deficiency $\delta := N - L - S = 0$), this property can also be simply tested. (As far as the concept goes, but one may have combinatorial blow up again.)

Detailed balancing and complex balancing are not so far from each other as they seem to be^[3, 9, 24, 19].

Atomic structure In this paper we only used chemical species without atomic structures. From the knowledge of the atomic structures further restrictions are derived: it is natural to assume the law of atomic balance.

Combinatorial blow up. Still, it will work for systems with not so many species. Many things should be calculated once, it is only the fitting what takes time.

Our immediate goal is to extend and apply the method to multiple real life data sets.

4.2 Final use of the the results

Present the best fitting reaction using AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) etc. to the chemist for further investigations and interpretation. Here we do not want to spend too much time with the detailed statistical analysis of simulated data, see the following books on this topics: [2, 18, 23, 27].

The method presented here can also be applied as a special kind of lumping,^[21]. Suppose a big kinetic model is given and we construct such small models in

the way described above which are able to reproduce the measured or simulated data obtained from the big one.

Carefully designed set of restrictions may decrease the number of candidate reactions, and increasing capacity of computers can do away with larger numbers, too.

Our calculations and data can be requested from the authors.

NOTATIONS

Classical notation	Notation in formal kinetics	Meaning Number of	Unit?
	L	linkage classes	
	M	species	
	N	complexes	
	P	reaction steps	
	R	reversible pairs of reactions	
	S	independent reactions	

Table 6: Notations

ACKNOWLEDGEMENTS

Irene Otero-Muras raised complex balance as possible property to use for restriction. Judit Zádor has supplied us with (real life chemical) references continuously. The present work has been supported by the National Research, Development and Innovation Office Hungary (SNN 125739).

ORCID

Tibor Nagy <https://orcid.org/0000-0002-1412-3007>

János Tóth <https://orcid.org/0000-0003-3065-5596>

Tamás Ladics <https://orcid.org/>

References

- [1] V. Aranda, J. M. Christensen, M. U. Alzueta, P. Glarborg, S. Gersen, Y. Gao, and P. Marshall. Experimental and kinetic modeling study of methanol ignition and oxidation at high pressure. *Int. J. Chem. Kinet.*, 45:283–294, 2013.
- [2] Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, New York, London, 1974.
- [3] A. Dickenstein and M. P. Millán. How far is complex balancing from detailed balancing? *Bull. Math. Biol.*, 73:811–828, 2011.
- [4] M. Feinberg. Necessary and sufficient conditions for detailed balancing in mass action systems of arbitrary complexity. *Chem. Eng. Sci.*, 44(9):1819–1827, 1989.
- [5] M. Feinberg. *Foundations of Chemical Reaction Network Theory*. Springer International Publishing, New York, 2019.
- [6] R. H. Fowler and E. A. Milne. A note on the principle of detailed balancing. *Proc. Natl. Acad. Sci. USA*, 11:400–402, 1925.
- [7] D. A. Frank-Kamenetskii. *Diffusion and Heat Transfer in Chemical Kinetics*. USSR Academy of Science Press, Moscow-Leningrad, 1947. In Russian.
- [8] V. Gold, K. L. Loening, A. D. McNaught, and P. Shemi. *IUPAC Compendium of Chemical Terminology*. Blackwell Science Oxford, second edition, 1997.
- [9] A. N. Gorban and G. S. Yablonsky. Extended detailed balance for systems with irreversible reactions. *Chem. Eng. Sci.*, 66(21):5388–5399, 2011.
- [10] D Healy, D M Kalitan, C J Aul, E L Petersen, G Bourque, and H J Curran. Oxidation of C1–C5 alkane quinary natural gas mixtures at high pressures. *Energy and Fuels*, 24(3):1521–1528, 2010.
- [11] O. Horie, J. N. Crowley, and G. K. Moortgat. Methylperoxy self-reaction: products and branching ratio between 223 and 333 K. *The Journal of Physical Chemistry*, 94(21):8198–8203, 1990.

- [12] F. Horn. Stability and complex balancing in mass-action systems with three short complexes. *Proc. R. Soc. Lond. A*, 334:331–342, 1973.
- [13] T. Ladics, J. Tóth, and T. Nagy. Automatic initial estimates. *Int. J. Chem. Kinet.*, 2019.
- [14] L. Lok and R. Brent. Automatic generation of cellular reaction networks with Molecuizer 1.0. *Nature biotechnology*, 23(1):131, 2005.
- [15] I. Nagy and J. Tóth. Microscopic reversibility or detailed balance in ion channel models. *J. Math. Chem.*, 50(5):1179–1199, 2012.
- [16] A. C. Noell, L. S. Alconcel, D. J. Robichaud, M. Okumura, and S. P. Sander. Near-infrared kinetic spectroscopy of the HO₂ and C₂H₅O₂ self-reactions and cross reactions. *The Journal of Physical Chemistry A*, 114(26):6983–6995, 2010.
- [17] I. Rácz, I. Gyarmati, and J. Tóth. Effect of hydrophilic and lipophilic surfactant materials on the salicylic-acid transport in a three compartment model. *Acta Pharmaceutica Hungarica*, 47:201–208, 1977. In Hungarian.
- [18] G. A. F. Seber and C. J. Wild. *Nonlinear Regression*. John Wiley & Sons, New York, 2003.
- [19] G. Szederkényi and K. M. Hangos. Finding complex balanced and detailed balanced realizations of chemical reaction networks. *J. Math. Chem.*, 49:1163–1179, 2011.
- [20] R. C. Tolman. The principle of microscopic reversibility. *Proc. Natl. Acad. Sci. USA*, 11:436–439, 1925.
- [21] J. Tóth, G. Li, H. Rabitz, and A. S. Tomlin. The effect of lumping and expanding on kinetic differential equations. *SIAM J. Appl. Math.*, 57:1531–1556, 1997.
- [22] J. Tóth, A. L. Nagy, and D. Papp. *Reaction Kinetics: Exercises, Programs and Theorems. Mathematica for Deterministic and Stochastic Kinetics*. Springer Nature, New York, 2018.
- [23] T. Turányi and A. S. Tomlin. *Analysis of Kinetic Reaction Mechanisms*. Springer, 2014.

- [24] A. van der Schaft, Sh. Rao, and B. Jayawardhana. Complex and detailed balancing of chemical reaction networks revisited. *Journal of Mathematical Chemistry*, 53(6):1445–1458, 2015.
- [25] M. O. Vlad and J. Ross. Thermodynamically based constraints for rate coefficients of large biochemical networks. *Systems Biology and Medicine*, 1(3):348–358, 2009.
- [26] R. Wegscheider. Über simultane Gleichgewichte und die Beziehungen zwischen Thermodynamik und Reaktionskinetik homogener Systeme. *Zsch. phys. Chemie*, 39:257–303, 1901/2.
- [27] Th. Weise. Global optimization algorithms-theory and application. *Self-published*, 2, 2009.