Stochastic simulation and analysis of biochemical networks

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Biophysik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Herr Dipl.-Inform. Jürgen Pahle
geboren am 28.04.1976 in Landau/Pfalz

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Christian Limberg

Gutachter/innen:

1. Prof. Dr. Thomas Höfer
2. Prof. Dr. Ursula Kummer
3. Prof. Dr. Hermann-Georg Holzhütter

Tag der mündlichen Prüfung: 31. März 2008

Abstract

Stochastic effects in biochemical networks can affect the functioning of these systems significantly. Signaling pathways, such as calcium signal transduction, are particularly prone to random fluctuations. Thus, an important question is how this influences the information transfer in these pathways.

First, a comprehensive overview and systematic classification of stochastic simulation methods is given as methodical basis for the thesis. Here, the focus is on approximate and hybrid approaches. Also, the hybrid solver in the software system Copasi is described whose implementation was part of this PhD work.

Then, in most cases, the dynamic behavior of biochemical systems shows a transition from stochastic to deterministic behavior with increasing particle numbers. This transition is studied in calcium signaling as well as other test systems. It turns out that the onset of stochastic effects is very dependent on the sensitivity of the specific system quantified by its divergence. Systems with high divergence show stochastic effects even with high particle numbers and vice versa.

Finally, the influence of noise on the performance of signaling pathways is investigated. Simulated and experimentally measured calcium time series are stochastically coupled to an intracellular target enzyme activation process. Then, the information transfer under different cellular conditions is estimated with the information-theoretic quantity transfer entropy. The amount of information that can be transferred increases with rising particle numbers. However, this increase is very dependent on the current dynamical mode of the system, such as spiking, bursting or irregular oscillations.

The methods developed in this thesis, such as the use of the divergence as an indicator for the transition from stochastic to deterministic behavior or the stochastic coupling and information-theoretic analysis using transfer entropy, are valuable tools for the analysis of biochemical systems.

Keywords:
stochastic simulation, biochemical networks, calcium signaling, information transfer

Zusammenfassung

Stochastische Effekte können einen großen Einfluss auf die Funktionsweise von biochemischen Netzwerken haben. Vor allem Signalwege, z.B. Calciumsignaltransduktion, sind anfällig gegenüber zufälligen Schwankungen. Daher stellt sich die wichtige Frage, wie dadurch der Informationstransfer in diesen Systemen beeinträchtigt wird.

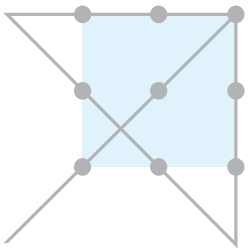
Zunächst werden eine Reihe von stochastischen Simulationsmethoden diskutiert und systematisch klassifiziert. Dies dient als methodische Grundlage der ganzen Dissertation. Der Schwerpunkt liegt hier auf approximativen und hybriden Ansätzen, einschließlich der Hybridmethode des Softwaresystems Copasi, deren Implementierung Teil dieser Arbeit war.

Die Dynamik biochemischer Systeme zeigt in den meisten Fällen einen Übergang von stochastischem zu deterministischem Verhalten mit steigender Partikelzahl. Dieser Übergang wird für Calciumsignaltransduktion und andere Systeme untersucht. Es zeigt sich, dass das Auftreten stochastischer Effekte stark von der Sensitivität des Systems abhängt. Ein Maß dafür ist die Divergenz. Systeme mit hoher Divergenz zeigen noch mit hohen Teilchenzahlen stochastische Effekte und umgekehrt.

Schließlich wird der Einfluss von zufälligen Fluktuationen auf die Leistungsfähigkeit von Signalpfaden erforscht. Dazu werden simulierte sowie experimentell gemessene Calcium-Zeitreihen stochastisch an die Aktivierung eines Zielenzyms gekoppelt. Das Schätzen des informationstheoretischen Maßes Transferentropie unter unterschiedlichen zellulären Bedingungen dient zur Abschätzung des Informationstransfers. Dieser nimmt mit steigender Partikelzahl zu, ist jedoch sehr abhängig von der momentanen Dynamik (z.B. spikende, burstende oder irreguläre Oszillationen).

Die hier entwickelten Methoden, wie der Gebrauch der Divergenz als Indikator für den stoch./det. Übergang oder die stochastische Kopplung und informationstheoretische Analyse mittels Transferentropie, sind wertvolle Werkzeuge für die Analyse von biochemischen Systemen.

Schlagwörter:
stochastische Simulation, biochemische Netzwerke,
Calciumsignaltransduktion, Informationstransfer



EML

Research

To M.P. and W.P.

Contents

List of Figures

List of Tables

Chapter 1

General introduction

Γνῶθι σεαυτόν
(Know yourself!)

Inscription on the
Temple of Apollo at Delphi

1.1 Historical overview

Ever since the Scottish botanist Robert Brown in 1827 observed the irregular movement of pollen particles suspended in water, people were fascinated by random processes. At first, it was believed that some kind of "vital force" was the cause of this so-called Brownian motion. However, this could be ruled out, since dust particles exhibit a similar jittery movement. Instead, it became clear that molecular fluctuations inevitably emerge in any system, animated or not, if it has a temperature above absolute zero.

In 1905 Albert Einstein's well-known article "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen" [Ein05] provided the mathematical framework for the study of such stochastic processes, which was successively extended and refined by many people, e.g., Paul Langevin, Itō Kiyoshi and Norbert Wiener.

Interestingly, the theory of stochastic processes found some of its first applications in the field of financial mathematics. Nevertheless, also chemical kinetic systems were studied in a stochastic manner quite early. Some important contributions were made by Max Delbrück (1940) [Del40], Anthony F. Bartholomay (1957) [Bar57; Bar58], Donald A. McQuarrie (1963) [McQ63; McQ67] and others. Lacking computing power for numerical calculations, these works dealt primarily with analytical solutions of the so-called

Chemical Master Equation, a differential equation for the probability distribution of the system states (e.g. set of species' particle numbers). Since in general this equation is infinitely dimensional, it can only be solved for the simplest systems, such as uni-molecular or bi-molecular equilibrium reactions [Lau00].

It is possible, though, to calculate single instances of the stochastic process, i.e. individual time courses of the species particle numbers that are governed by the Chemical Master Equation. In 1976 and 1977 Daniel T. Gillespie devised two simple algorithms for that in his seminal articles [Gil76; Gil77]. In particular during the last years [ARM98; GTG01], this stochastic simulation of biochemical networks has attracted renewed interest.

Already before that, however, it had become increasingly apparent that the goal to understand the huge biochemical network in living cells will neither be achieved by genomics approaches nor by biochemical experiments alone. This is mainly due to the vast degree of interaction and the resulting complexity of this network. Therefore, computational methods, such as the numerical simulation of biochemical processes, are essential in order to come closer to an understanding of the cell. The integration of these computational methods with corresponding experimental techniques has become known as *Systems Biology* [Kit02][1].

Simulations in systems biology, having their seeds in the quantitative modeling of enzyme kinetics, are largely based on numerical integration of ordinary differential equations (ODE). This approach proved very successful, in particular for metabolic systems [GH64; HS96]. Not only are there a variety of efficient simulation methods, e.g. [Pet83; Ros63], but also many powerful analysis techniques, e.g. [Fel97; Sav76].

In spite of the wide-spread use of the ODE approach, two of its limitations have come to the fore recently:

- The system is assumed to be homogeneous (within each reaction compartment). Therefore, intracellular *spatial organization* is largely neglected.

- The system is assumed to evolve deterministically. *Random fluctuations* in molecular numbers, particularly of small-numbered subsystems, are not considered.

Both spatial organization and stochastic effects can affect the dynamical behavior of biochemical systems dramatically [TRR91; RWA02], and taking these into account can be crucial for the understanding of the underlying

[1] For a well-written review of some of systems biology's origins, see [Buc02].

mechanisms. Also, diffusion and noise seem to be closely connected [Bha04]. Spatial effects can be investigated using, e.g., partial differential equations or spatial stochastic methods [TAT05]. Given the increasing availability of sophisticated measuring techniques which provide sub-cellular resolution (e.g. [Pet07]), we anticipate that spatial modeling will be of high importance in the future. Nevertheless, in the present thesis we will concentrate on stochastic effects in spatially homogeneous systems.

As mentioned, these stochastic effects can, in principle, be studied by using Gillespie's algorithm (provided the system is well-mixed). As this algorithm, however, is bound to tediously calculate each individual reaction event in the system, it is clear that its computational demands can be enormous for realistic systems.

The lack of sufficient computing power for stochastic simulations surely was one of the reasons that more than twenty years elapsed between Gillespie's original article in 1976 and the (re-)ignition of considerable interest in stochastic methods around 2000. Another reason might have been the development of appropriate single-cell measuring techniques with satisfactory temporal resolution that allows the study of fluctuations beyond the bulk behavior.

However, because the stochastic simulation algorithm is computationally expensive, a multitude of approximate approaches have been proposed recently that try to accelerate stochastic simulations by sacrificing an acceptable amount of exactness, e.g. [Gil01; RWD01; RA03; BT04].

Also the question arises, if time-consuming stochastic simulation is required for a specific system (or parts of it) at all. In the past, this question has often been answered by an educated guess depending on whether or not the modeler expected stochastic effects to occur. The most important criterion for that decision has been the number of particles in the system, because the relative amplitude of fluctuations, in many cases, scale with the inverse square-root of the particle numbers [RWA02]. The deterministic interpretation represents the limiting behavior in the case of infinitely high particle numbers and volumes, but constant concentrations [Kur72].

Nevertheless, answering this question in a more general manner turned out to be difficult, and relatively few studies address this issue [Fal03a; GHG04].

Among the biochemical systems especially signaling pathways, as well as gene expression, are prone to stochastic phenomena because they often involve species with low particle numbers and, in addition, possess highly non-linear kinetics. For instance, calcium signaling [BBL98] is a particularly well-studied system in that respect. Many articles have been published which deal with stochasticity in this system [GTG01; SJ02; Fal03b; Fal04;

GTMG03; PM03a; PGDM08].

In 1986, Woods et al. [WCC86] were the first to observe that, upon stimulation by an agonist, the cytosolic calcium level in single rat hepatocytes display an intriguing oscillatory behavior. This is highly interesting because different properties of these calcium oscillations are believed to confer specificity. This means that intracellular target enzymes can be activated differentially depending on how the oscillations of the ubiquitous second-messenger Ca^{2+} are modulated.

Therefore, a number of models for calcium signaling have been developed (see Schuster et al. [SMH02], for a review). Also, the decoding of calcium signals was investigated experimentally [HRGST95; Ber97; DS98; DXL98; LLW$^+$98; OM98] as well as theoretically [PLvzM$^+$98; GBD00; RJ03; DHD03; LK03; LOK04; SPH04; SKM05; MPS05; MPS06; SPH08].

Since signaling pathways have to cope with fluctuations in small particle numbers, an important question is how this affects information transfer. The most general framework to address this question seems to be information theory [CT91], a discipline which was established in the 1940s by C.E. Shannon [Sha48]. Information theory deals with the performance of communication systems under the influence of noise in a statistical manner.

Information theory has been applied in the field of neurophysiology [BT99] and many others. However, studies using it to investigate signaling pathways are rare. For instance, in Palus et al. [PSvzM$^+$98] the calcium response of hamster insulin secreting cells (HIT) under pharmacological stimulation is studied using coarse-grained entropy rates. Prank et al. [PSL$^+$98; PGB00; KGP05] analyzed the information transfer in a calcium oscillations model which has been stimulated by a generated hormone signal.

1.2 Research objectives

Stochastic effects in biochemical networks can affect the functioning of these systems significantly.

In most cases, one can observe a transition between a stochastic and a quasi-deterministic regime with increasing particle numbers. However, when this transition really occurs is very system dependent. Since reliable indicators for the onset of stochastic effects were mostly missing, one of the aims of the present thesis is to investigate this transition in different systems and to find more objective and general criteria.

In particular signaling pathways are influenced by molecular fluctuations. This immediately raises questions such as: How can their performance under the influence of noise be quantified? How does information transfer change

in the transition range between stochastic and deterministic behavior and under different cellular conditions? To address these questions we will use an information-theoretic setting. This facilitates the direct comparison of experimental and simulated data.

In contrast to earlier studies by Prank et al. [PSL+98; PGB00; KGP05] which investigated the information transfer between a generated hormone signal (band-limited Gaussian noise) and a model of calcium oscillations stimulated by it, we will concentrate on the information transfer between calcium and intracellular target enzymes. Also, we will use a model of calcium oscillations to generate biologically plausible input signals. Finally, a stochastic coupling between the calcium system and the enzyme system will be employed, instead of deterministic coupling as in [PLvzM+98; LOK04]. This is essential in order to study the dependence of the information transfer on particle numbers.

The basis for the study of stochastic effects are, of course, appropriate stochastic simulation methods. Whereas the numerical integration of ODEs is widely known, this is still not true for stochastic approaches. Therefore, it seems in order to firstly review the different methods suggested in the literature. Here, the focus will be on approximate and hybrid methods, since they attempt to bridge the important intermediate range between stochastic and deterministic behavior. This includes the hybrid method in the software system COPASI which was implemented during this PhD work.

1.3 Outline of the thesis

This thesis is structured as follows:

Chapter 2: First, a number of reasons for the use of stochastic modeling and simulation are discussed. Then, a comprehensive overview and systematic classification of the various simulation methods is given as a basis for the following chapters.

Chapter 3 provides the biological background on calcium signaling and details about calcium models. Also, the experiments on rat hepatocytes, including the materials used, are described there.

Chapter 4: This chapter focuses on stochastic simulation in the software system COPASI, in particular, the hybrid approach, whose implementation was part of this PhD work, and a number of practical issues.

Chapter 5 deals with the transition from stochastic to quasi-deterministic behavior in biochemical systems. In this context, particularly the role of the so-called divergence is investigated.

Chapter 6: Here, the information transfer in stochastic signaling pathways, namely between calcium and intracellular target enzymes, is studied using the information-theoretic quantity transfer entropy. The main question is: How does this information transfer change under the influence of different levels of noise and different dynamical conditions.

Chapter 7 concludes the thesis and contains a brief outlook and suggestions for future projects.

Chapter 2

Review of stochastic simulation methods

> I have come to believe that one's knowledge of any dynamical system is deficient unless one knows a valid way to numerically simulate that system on a computer.
>
> Daniel T. Gillespie

The simulation of (bio-)chemical systems is most often done using deterministic methods, e.g. numerical integration of ordinary differential equations[1]. However, in some cases a stochastic treatment is necessary. Since the seminal work by Gillespie in 1976 a multitude of different stochastic simulation algorithms have been proposed. After a brief introduction of deterministic simulation and a discussion of why and when stochastic simulation is required, a comprehensive review of the different stochastic approaches is given with an emphasis on hybrid methods. This chapter provides the methodical basis for the present thesis. This work was part of the BMBF-project HepatoSys (1st funding period 2004–2006) [Hep].

[1] It should be mentioned here that we are concerned with spatially homogeneous models only. However, there are also a variety of modeling approaches and simulation methods for spatially heterogeneous systems (e.g. partial differential equations). Two examples of spatial and stochastic simulation are MCell [MCe] and SmartCell [Sma].

2.1 Deterministic simulation

The majority of biochemical simulations are based on deterministic methods, e.g. using numerical integration of ordinary differential equations (ODEs) [HS96]. Here, the concentrations of the participating chemical species are taken as continuously valued components of the system's state. Their change over time is described by a differential equations system. Deterministic in this context means that random fluctuations in molecule numbers are ignored. Given an initial state the future of the system is determined in a unique way.

We consider a biochemical network which contains N reacting species S_i $(i = 1, \ldots, N)$ participating in M reactions R_μ $(\mu = 1, \ldots, M)$. Let $X_i(t)$ be the number of particles of species S_i present in the system at time t. If the particles are confined to the constant volume V, we can refer to the value $X_i(t)/V$ as the concentration $[X_i](t)$ of species S_i at time t. At this point it is worth noting that, strictly speaking, the concentration of species can only take discrete values. The conventional assumption of continuous concentrations is only valid in the limit of high particle numbers and volumes.

For each species involved in a reaction there is a stoichiometric coefficient $v_{i\mu}$, which indicates the proportions in which the substrates and products are consumed and produced respectively. Associated with each reaction R_μ is a kinetic function f_μ which defines the velocity of that reaction, i.e. the number of R_μ reaction events (or "firings") per unit time. The kinetic function usually takes the substrate particle numbers or concentrations as arguments. Other variables in the kinetic function, such as the concentration of enzymes catalyzing the reaction or the temperature and pH value of the environment, are also possible because they too can influence the reaction velocity. In the case of enzyme concentrations those additional arguments are often called modifiers[2].

The basis of the deterministic approach and, at the same time, the simplest example of a kinetic function is the mass-action kinetics, where the reaction velocity is proportional to the product of the substrate concentrations to the power of their respective molecularity:

$$R_\mu : \alpha S_1 + \beta S_2 + \ldots \rightarrow \text{products} \tag{2.1}$$

with

$$f_\mu([X_1], [X_2], \ldots) = k \cdot [X_1]^\alpha \cdot [X_2]^\beta \cdot \ldots \tag{2.2}$$

[2]This use of the term "modifier" should not be confused with the usual meaning of an regulator of enzyme activity.

When several elementary reactions are lumped together into one complex reaction, for instance in enzyme kinetics, more complex kinetic functions (e.g. Michaelis-Menten/Briggs-Haldane kinetics) are used.

A biochemical differential equations system comprises one equation for each variable, i.e. concentration of species in the system. The right hand sides of these equations contain one summand for each reaction, where the species appears as substrate or product. Each summand in turn is the product of the kinetic function of that reaction and the corresponding stoichiometric coefficient. We get:

$$\frac{\mathrm{d}[X_i]}{\mathrm{d}t} = \sum_{\mu=1}^{M} v_{i\mu} \cdot f_\mu \qquad (i = 1, \ldots, N). \tag{2.3}$$

For a well-defined initial value problem, one must provide initial values $[X_i](t = 0)$ in addition to the equations.

In all but the most simple cases it is not possible to find a closed-form solution for such systems and one must rely on numerical integration methods that iteratively compute the solution starting from the initial values.

Differential equations corresponding to biochemical systems are almost always stiff meaning that the changes of variables in the system occur on disparate scales which can be several orders of magnitude apart. Such stiff systems require the use of special stiff ODE solvers such as the Rosenbrock method [PTVF92] or LSODE [Hin83]. The use of simpler methods, such as the popular fourth-order Runge-Kutta method, is not recommended, since here the stiffness either causes numerical problems or slows down the solver considerably.

The most prominent advantage of deterministic simulation methods is that many efficient algorithms are available, for stiff and non-stiff systems. The time course can be calculated very rapidly.

Another important benefit of the deterministic formalism is that there exists a rich body of research on the analysis of these models. This includes bifurcation analysis, e.g. numerical continuation methods [XPP], Metabolic Control Analysis [HS96] and others.

The deterministic approach has not only been employed in most of the studies concerned with the dynamical behavior of biochemical systems so far, but there also exist many good software tools. Among them are programs for the numerical integration of ODEs in general, such as Berkeley Madonna or XPPAut [Mad; XPP], as well as programs tailored to the simulation of biochemical networks, such as COPASI ([Cop], also cf. the list of SBML-compliant software at [SBM]).

2.2 Why stochastic simulation?

The deterministic approach has proven very successful. However, despite of its widespread use it has a number of drawbacks:

- Small particle numbers in cells (e.g. in signaling pathways) lead to random fluctuations which can change the dynamic behavior considerably [QSE02; MA97].
- Bi- or multi-stable systems can not be described adequately [ARM98].
- Stochasticity itself can be an important property of the system. This includes "the generation of errors in DNA replication leading to mutation and evolution, noise-driven divergence of cell fates, noise-induced amplification of signals, and maintenance of the quantitative individuality of cells" [RWA02, p. 231].

For very small particle numbers (e.g. in gene expression) the continuous description using concentrations breaks down.

> "In this case we cannot apply continuous state models even as an approximation. [...] Discrete state space, but deterministic, models are also out of question, since fluctuations cannot be neglected even in the *zeroth approximation*, because they are not superimposed upon the phenomenon, but they represent the phenomenon itself." [ÉT89, p. 7]

Inherent stochastic fluctuations in molecule numbers can change the dynamic behavior of biochemical systems significantly and can even give rise to a qualitatively different behavior of the whole system [SPA05; SA06]. One example is the production of proteins in random bursts [MA97] (related studies can be found in [KE01; MA99]). Another example is the lysis/lysogeny switch in λ phage-infected *Escherichia coli* cells [ARM98].

This is not restricted to systems containing low-numbered species, but can also happen when the "system operates near an instability point of a deterministic model. In this case small fluctuations may be amplified and produce observable, even macroscopic effects. It may also happen that the deterministic model of a system is structurally stable while the stochastic model is not, or vice versa" [ÉT89, p. 7]. We will see an example of that in Chap. 5.

Cells must act in a coordinated way despite ubiquitous molecular noise. Therefore, specific cellular systems to cope with or even exploit stochastic effects are likely to be very important and deserve further study. Stochastic

modeling provides the appropriate framework to investigate, for example, the robustness of cells against random perturbations [RBH94; BL99; Hal99; Kut01; GHG02b; GHG02a; VKBL02; GHLG03; GTMG03; GHG04; Kit04; BFLM05] or the constructive effects of noise [ARM98; Hum00; LKS+01; PM03a].

A well-written review of stochastic effects and noise in cellular systems can be found in Rao et al. 2002 [RWA02].

Finally, noise can even help with elucidating the functioning of biochemical systems. "Fluctuations can be a source of information. According to the so-called fluctuation-dissipation theorem the dissipative processes leading to equilibrium are interconnected with the fluctuations around the equilibrium point. Using the spirit of this theorem the kinetic rate constants can be calculated from equilibrium measurements" [ÉT89, p. 7]. Examples where noise is exploited to study the underlying cellular processes include Elowitz et al. 2002 [ELSS02] and Ozbudak et al. 2002 [OTK+02]. Here, extrinsic and intrinsic noise sources are distinguished in gene expression. In Hallett and Pettit 1997 [HP97] the authors investigate the increase of cytosolic calcium concentration upon stimulation. They fitted a stochastic model to experimental data and concluded that a process with six subsequent steps is most likely to lead to the observed behavior.

In all the above cases stochastic fluctuations should be accounted for and a stochastic modeling framework which takes into account the discreteness of the system and correctly reproduces the fluctuations is required.

However, stochastic modeling is mathematically more involved than the conventional deterministic approach using ODEs. For example, it is not trivial to correctly incorporate lumped reactions using higher-order kinetics which is commonplace in the deterministic framework (see Sec. 2.4.2).

Also, stochastic simulations usually are much more computationally demanding than deterministic methods. Exact methods take time proportional to the number of single reaction events occurring during the simulation time and this is dependent on the actual particle numbers.

One other drawback of stochastic modeling is that it still lacks behind in terms of appropriate analysis methods such as stochastic bifurcation analysis, stochastic Metabolic Control Analysis (MCA) [WBH04] or stochastic sensitivity analysis [GCPD05].

Therefore, Érdi and Tóth remark [ÉT89, p. 91]: "We have to make clear that the formulation of the theory of stochastic kinetics does not reduce the importance of deterministic kinetics, since for great classes of phenomena the stochastic model is only slightly *better* than the deterministic approach, while the mathematics of the stochastic model is much more complicated."

2.3 Stochastic simulation: When?

Now, the practical question arises whether or not stochastic simulation should be preferred over numerical integration of ODEs for a specific system one wants to analyze. In the past, most often rules of thumb based on the rough particle numbers in the system have been used to answer this question. Rao states in [RWA02, p. 234]:

> "How is a modeler to choose between modeling approaches — an implicit or explicit treatment of noise, a continuous or discrete representation of molecules? When simulating processes that involve only a few molecules, discrete stochastic models are superior to continuous models. However, in many processes, there are many copies of some species and few of others. In these circumstances, it is not always clear which approach is better. [...] As multiscale approaches for simulating stochastic processes are desperately lacking, personal proclivities currently dictate the choice of approach, as modelling and simulation are, at this stage, more art than science."

Based on the asymptotic equivalence of the Master Equation and the chemical Langevin equation (see Subsection 2.4.2) one could argue "[...] that the relative magnitude of the molecular fluctuations scales roughly as the inverse square root of the number of reacting molecules" [RWA02, p. 233] and this backs up using heuristics based on particle numbers.

However, giving a general threshold for the particle numbers, above which it is safe to employ deterministic methods, is surely impossible. The emergence of stochastic effects is very model-dependent and this has to be checked in each individual case. In addition, it has been established that even consistency in the mean does not hold in general [ÉT89, p. 160]. To assist the modeler with this decision we investigate the transition from stochastic to quasi-deterministic behavior with increasing particle numbers in more detail in Chap. 5.

Nevertheless, definite indicators necessitating stochastic modeling are a) when the particle numbers are in a range where the concept of continuous concentrations is no longer appropriate, or b) when phenomena associated with stochasticity itself are the object of research.

Finally, it should also be mentioned that increasingly bigger and more heterogeneous models, e.g. combining gene expression, signaling and metabolic pathways, might require different formalisms for the different subsystems, eventually leading to hybrid modeling.

2.4 Stochastic simulation methods

The stochastic approach is based on the notion of a stochastic reaction constant c_μ for each reaction R_μ in the system[3] (see Fig. 2.1). $c_\mu \mathrm{d}t$ is the average probability that a specific combination of substrate particles in the system will react in the next infinitesimal time step $\mathrm{d}t$ according to reaction R_μ. Multiplied with the number of possible R_μ substrate particle combinations h_μ (see Gillespie 1976 [Gil76], for details) we get a propensity a_μ for reaction R_μ as $a_\mu = h_\mu c_\mu$ with:

$$a_\mu \mathrm{d}t = \text{the average probability that a reaction } R_\mu \text{ will occur in the system in the next infinitesimal time step } \mathrm{d}t. \tag{2.4}$$

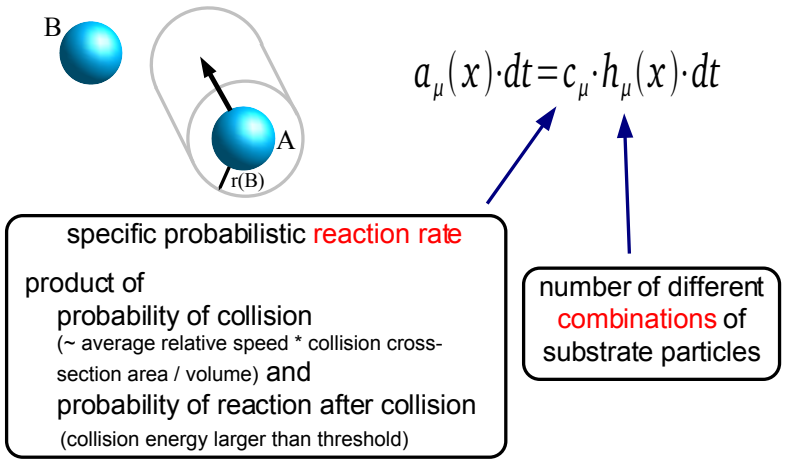


Figure 2.1: Basis of the stochastic approach: reaction propensity a_μ. Particles A and B (with A moving relative to B) can react when B is localized within the "collision volume" swept out by A during time $\mathrm{d}t$ [Gil77].

a_μ can usually be derived easily from the conventional deterministic reaction rate. In general, we have $c_\mu = \frac{k_\mu \prod(l_{\mu_i}!)}{V^{K_\mu - 1}}$ and $h_\mu = \prod_{i=1}^{L_\mu} \binom{X_i}{l_{\mu_i}}$, if the reaction R_μ has K_μ reactants, i.e., substrate molecules, taking part. There are L_μ different types of reactants and for each type, l_{μ_i} is the stoichiometric number, i.e., the number of identical reactant molecules. Thus $K_\mu = \sum l_{\mu_i}$. The numbers X_i refer to the total particle numbers present in volume V. Finally, the rate k_μ includes any complex factors that might arise from the kinetics of the reaction.

[3] One condition for this constant to be well-defined is that many more non-reactive molecular collisions occur than reactive ones.

It turns out that within this framework the biochemical system can be identified with a (homogeneous) jump-type Markov process for which the so-called Chemical Master Equation can be formulated [Gil76; McQ67]

$$\frac{\partial P(x,t|x_0,t_0)}{\partial t} = \sum_{\mu=1}^{M}[a_\mu(x-v_\mu)P(x-v_\mu,t|x_0,t_0) - a_\mu(x)P(x,t|x_0,t_0)] \quad (2.5)$$

which describes the time evolution of the system state probability distribution $P(x,t)$, given that the initial state at point in time t_0 is x_0 (x denotes the vector of species particle numbers X_i and v_μ is the stoichiometric vector of reaction R_μ). In general, this differential-difference equation is infinitely-dimensional. It can be solved either analytically or numerically only in the simplest cases [Lau00]. Therefore, one resorts to stochastic simulation methods that, using Monte Carlo calculation schemes, generate instances of the underlying stochastic process governed by the Master Equation. However, since each run yields only one single trajectory, several runs have to be computed to be able to calculate statistics of $P(x,t)$.

Stochastic simulation algorithms can roughly be divided into exact, approximate or hybrid methods depending on whether or not they introduce approximations or combine different approaches into one calculation scheme.

In the following we will give an overview of a number of different stochastic approaches proposed in the literature. Other reviews on stochastic and hybrid simulation methods can be found in Turner 2004 [TSB04] or Meng 2004 [MSD04]. Also, see Appendix A for a list of biochemical software tools with stochastic simulation methods implemented.

2.4.1 Exact stochastic methods

The so-called exact stochastic methods correctly account for inherent stochastic fluctuations and correlations. In addition, the discrete nature of the system is considered. Hence, they remain valid for very small particle numbers.

However, since they explicitly simulate each reaction event in the system, they have time complexity approximately proportional to the overall number of particles $\sum X_i$ present in the system. Therefore, they are prohibitively slow on large systems.

Gillespie [Gil76] proposed two simple stochastic simulation algorithms, namely the Direct Method and the First Reaction Method. They are based on the so-called Reaction Probability Density Function:

$$P(\tau, \mu|x, t) = a_\mu(x) \exp\left(- \sum_{\mu=1}^{M} a_\mu(x)\tau \right). \tag{2.6}$$

This function determines the probability $P(\tau, \mu|x, t)\mathrm{d}\tau$ that starting in state x at time-point t the next reaction in the system will occur in the time interval $[t + \tau, t + \tau + \mathrm{d}\tau]$ and will be of type R_μ. The reaction times are exponentially distributed and this defines a homogeneous Poisson process.

By iteratively drawing random numbers according to that density function and updating the system state according to the chosen reaction's stoichiometry the system can be simulated over time, one reaction event after the other. The Direct Method and the First Reaction Method, as well as the Next Reaction Method and Optimized Direct Method described below, are mathematically equivalent but differ in how they calculate samples of $P(\tau, \mu|x, t)$.

Direct Method (Gillespie 1976/77)

The Direct Method [Gil76; Gil77] uses two random numbers per step to separately compute a) the stochastic time step τ and b) the type of the next reaction R_μ. The algorithm proceeds as follows:

1. First, the sum of all propensities for the M possible individual reactions is calculated:
$$a_0 = \sum_{\mu=1}^{M} a_\mu \tag{2.7}$$

2. The stochastic time step is calculated:
$$\tau = -\frac{1}{a_0} \ln r_1 \tag{2.8}$$
Here r_1 is denoting a uniformly distributed random number in the range $]0, 1]$.

3. Finally, the reaction taking place is determined. For this purpose, a second uniformly distributed random number r_2 is generated and the reaction μ chosen according to the following criteria:
$$\sum_{\alpha=1}^{\mu-1} \frac{a_\alpha}{a_0} \leq r_2 \leq \sum_{\alpha=1}^{\mu} \frac{a_\alpha}{a_0} \tag{2.9}$$

4. The corresponding reaction R_μ is realized, i.e., the number of the participating molecules is increased or decreased according to the stoichiometry, and the time is incremented by τ. The whole process (1-4) is repeated as many times as necessary to reach the desired simulation time.

First Reaction Method (Gillespie 1976/77)

The First Reaction Method [Gil76; Gil77] uses M random numbers in each step to compute putative reaction times τ_μ for each of the M reactions R_μ in the system. The τ_μ are exponentially distributed with parameter a_μ. The reaction with the smallest reaction time $\tau_{\min} = \min(\tau_1, \ldots, \tau_M)$, is executed and all putative reaction times are recalculated before the next step. Because of the wasteful use of random numbers and redundant recalculations the First Reaction Method is computationally inefficient. However, we found that the First Reaction Method can be numerically advantageous because, depending on the machine-dependent resolution of floating-point numbers, the Direct Method cannot calculate very seldom reaction events amongst very frequent reactions in the system.

Next Reaction Method (Gibson and Bruck 2000)

Gibson and Bruck [GB00] reduced the computational complexity of the First Reaction Method by the intelligent use of data structures:

- A so-called dependency graph stores dependencies between the reactions → redundant recalculations of a_μ are avoided.
- Absolute putative reaction times τ_μ (since the beginning of simulation) are used instead of relative ones (since the last reaction event). Random numbers are "recycled" during a reaction time update according to $\tau_\mu = (a_{\mu,\text{old}}/a_\mu(x))(\tau_{\mu,\text{old}} - t) + t$.
- An indexed priority queue contains all reactions sorted according to their putative reaction times → the next reaction can always be found in the root of the tree with constant complexity.

Even though this so-called Next Reaction Method is mathematically equivalent to the Direct Method, recalculations of a_μ are minimized, and, asymptotically, only one random number per step is needed. Therefore it performs well, in particular, in the case of many reacting species and reactions.

Optimized Direct Method (Cao et al. 2004)

Cao et al. [CLP04] demonstrated that a clever implementation of the Direct Method using a dependency graph and sorting of reactions according to their propensities can be even faster than Gibson and Bruck's Next Reaction Method.

2.4.2 Approximate stochastic methods

The huge computational effort needed for exact stochastic simulation entailed a lively search for approximate simulation methods that sacrifice an acceptable amount of accuracy in order to speed up the simulation. The proposed methods often involve a grouping of reaction events (PW-DMC, τ-Leap Method), i.e. they permit more than one reaction event per step.

One approximate stochastic method, StochSim, does not primarily aim at speeding up the simulation and shall be discussed first.

StochSim, Mesoscopic Approach (1998)

In the mesoscopic approach by Morton-Firth (now Firth) [MF98; MFB98; LS01] single particles are distinguished and represented by separate software objects. Though, their positions and velocities in the reaction volume are disregarded. This characteristic makes StochSim mesoscopic, residing on a middle conceptual level between microscopic molecular dynamics and macroscopic approaches considering only particle numbers.

In each step two particles (or one particle and one pseudo-particle for mono-molecular reactions) are randomly chosen. Whether a reaction event takes place between the two, is then determined using a random number and a look-up table.

This procedure can be even more time-consuming than the Direct Method when many non-reactive selections are chosen. The advantages are that multi-state particles are possible (with different reaction rates for each configuration) and that the life-cycle of single particles can be traced.

The approach has also been extended to cover simple 2D-spatial modeling with stationary particles [Shi02].

Probability-Weighted Dynamic Monte Carlo (2001)

The Probability-Weighted Dynamic Monte Carlo Method [RWD01] (PW-DMC method) is a rather ad-hoc approach in which reactions with high probability are allowed to fire multiple times. Instead of considering single fast reaction events separately, several events are grouped together and

simulated as if they were one event. This reduces the effective propensity a_μ of fast reactions and can reduce computation times, but has the major drawback that fluctuations can be misdescribed.

A related approach was already used in Gibson & Bruck 2000 [GB00]. There, a large fixed number of subsequent first-order reaction steps (transcription, translation) was lumped together. With exponentially distributed reaction times for each single step the whole process can be described using a gamma distribution.

τ-Leap Method (2001)

In 2001 Gillespie [Gil01] developed an approximate stochastic simulation method named τ-Leap Method to accelerate the simulation procedure. This method avoids the meticulous reconstruction of every individual reaction event. Instead, it leaps along the time axis in steps of length τ containing many single reaction events. τ has to fulfill the so-called Leap Condition: It must be small enough that no significant change in the propensities a_μ during $[t, t+\tau]$ occurs. Then, the reaction channels effectively decouple and the number of firings K_μ during time τ, starting from state x at time t, can be approximated by Poisson distributed random variables:

$$K_\mu(\tau; x, t) = \mathcal{P}(a_\mu(x), \tau) \tag{2.10}$$

with

$$\text{Prob}\{\mathcal{P}(a_\mu, \tau) = k\} = \frac{(a_\mu \tau)^k}{k!} e^{-a_\mu \tau}. \tag{2.11}$$

In each step a Poisson random number k_μ is drawn for each reaction R_μ and the system state is updated according to:

$$x(t+\tau) = x(t) + \sum_{\mu=1}^{M} k_\mu v_\mu. \tag{2.12}$$

Due to the drawing of Poisson random numbers, each τ-leap is more expensive than one step of the Direct or First Reaction Method. However, since many single reaction events can be leaped over when τ is large enough, the simulation can be much faster after all. Fig. 2.2 illustrates the procedure.

The most important question remaining is how to choose an appropriate τ-value. Determining this value involves a trade-off between accuracy of the simulation and computation time. In the original article, Gillespie proposed a simple τ-choosing strategy. This has been improved upon in [GP03] and [CGP06].

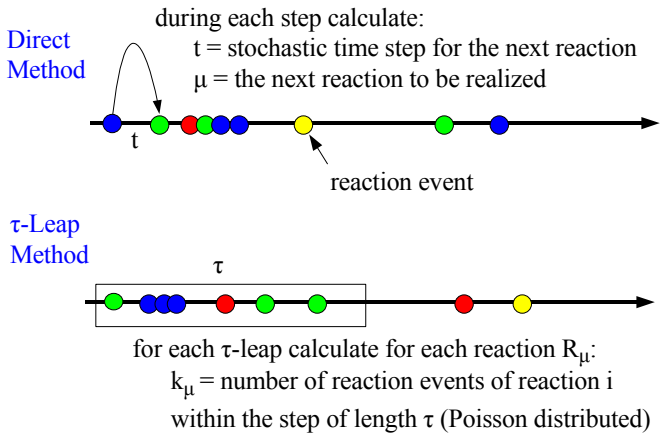


Figure 2.2: Schema of the τ-Leap Method [Gil01].

A number of variants and extensions of τ-Leaping have been developed recently that tackle some of the issues of the τ-Leap Method from 2001:

- Rathinam et al. [RPCG03] draw on the analogy to numerical integration of ODEs and describe an "Implicit τ-Leaping Method" in order to deal with stiffness in the system.

- Tian and Burrage 2004 [TB04] and Chatterjee et al. [CVK05] suggested to use (bounded) binomial random numbers instead of Poisson distributed ones. This so-called "Binomial τ-Leaping" avoids one of the problems of the τ-Leap Method, namely the generation of negative particle numbers in some cases.

- Another approach to avoid negative particle numbers has been developed by Cao et al. [CGP05c]. In this "Modified τ-Leaping Method" exact stochastic simulation, allowing only one reaction event per step, is used for those particle numbers that are critically low.

- Burrage and Tian [BT04] constructed "Poisson Runge-Kutta Methods (PRK)" to increase the efficiency of τ-Leaping.

- The "R-Leaping" in [ACK06] is a variant of the k_α-Leaping method described in [Gil01].

Langevin Method

If the value of τ can be chosen big enough such that every reaction channel on average produces a very large number of firings ($\langle \mathcal{P}(a_\mu(x), \tau) \rangle = a_\mu(x)\tau \gg 1$) while still satisfying the Leap Condition, the Poisson random variables can be approximated by normal random variables $\mathcal{N}_\mu$ [Gil00]:

$$\begin{aligned} K_\mu(\tau; x, t) &= \mathcal{P}_\mu(a_\mu(x), \tau) \\ &\approx \mathcal{N}_\mu(a_\mu(x)\tau, a_\mu(x)\tau) \qquad (2.13) \\ &= a_\mu(x)\tau + \sqrt{a_\mu(x)\tau} \cdot \mathcal{N}_\mu(0, 1) \qquad (2.14) \end{aligned}$$

In this case, the τ value is called to be "macroscopically infinitesimal". The Poisson random variables for the system state update can be replaced by normal random variables which are easier to calculate. Conceptually, the procedure is now equivalent to the (chemical) Langevin equation, a stochastic differential equation (SDE):

$$x(t + \tau) = x(t) + \sum_{\mu=1}^{M} v_\mu a_\mu(x)\tau + \sum_{\mu=1}^{M} v_\mu \sqrt{a_\mu(x)\tau} \cdot n_\mu \qquad (2.15)$$

with n_μ unit normal random variables.

In the limiting case $a_\mu(x)\tau \to \infty$, the last (noise) term becomes negligibly small compared with the second (see Gillespie 2000 [Gil00]) yielding the Euler update method for the numerical integration of ODEs. The Langevin Method therefore illustrates how the stochastic simulation algorithms (SSA) are connected to the deterministic method through a series of approximations (SSA $\to$ τ-Leap Method $\to$ Langevin Method $\to$ deterministic reaction rate equations).

Complex stochastic kinetics

The stochastic formalism is based on irreversible elementary reactions. How to correctly handle higher-order kinetics, where a number of elementary reactions have been lumped into one complex reaction is still subject to research.

Bundschuh et al. 2003 [BHJ03] observed that a naive direct stochastic simulation of higher-order kinetics[4] can lead to a failure of correctly describing the fluctuations and even the mean of particle numbers. Nevertheless,

[4]The direct stochastic simulation of higher-order kinetics is done by extracting the

in many cases the use of higher-order terms in stochastic simulation is indeed justified as is discussed in Rao and Arkin 2003 [RA03] and Cao et al. [CGP05b] for the stochastic quasi-steady-state approximation (QSSA) and Michaelis-Menten kinetics, and in Sec. 2.4.4 for the quasi-equilibrium approximation (QEA). See also Sec. 4.3, for a discussion of the implementation in COPASI.

2.4.3 Hybrid methods

On the one hand, deterministic simulation methods are very effective in simulating biochemical systems with high numbers of molecules. However, they completely neglect internal fluctuations which primarily occur when only few molecules are present in the system. On the other hand, stochastic simulation methods reproduce those random fluctuations correctly but can only do that efficiently for systems containing relatively few molecules.

The basic idea of hybrid simulation methods is to combine the advantages of complementary simulation approaches: The whole system is subdivided into appropriate parts and different simulation methods operate on these parts at the same time. Fig. 2.3 shows a schematic view of this procedure. Here, we have two subsystems containing the fast and slow reactions respectively. Fast reactions often involve high-numbered species, e.g. in metabolism. Slow reactions or reactions involving low-numbered species can frequently be found in signal transduction or gene expression systems. The two subsystems are simulated iteratively by using adequate simulation methods, for instance, numerical integration of ODEs and stochastic simulation respectively.

Mathematically, this corresponds to a partitioning of the Chemical Master Equation (for a rigorous derivation of the partitioning process and mathematical details of the approximations employed, see Haseltine and Rawlings 2002 and 2005 [HR02; HR05] and the references given in the respective subsections below).

In between two reaction events in the slow/discrete subset of reactions, the fast subset evolves due to the action of the fast reactions only. This means that, during that time, its behavior can be approximated using ODEs, SDEs or approximate stochastic simulation methods independently of the slow subset.

---

mass-action part of the kinetic function. The remainder including the influence of the modifiers constitutes the stochastic reaction rate c_μ. It is easy to see that this approach runs into problems whenever the modifiers themselves are fast changing variables because only the c_μ value at the beginning of each step is considered.

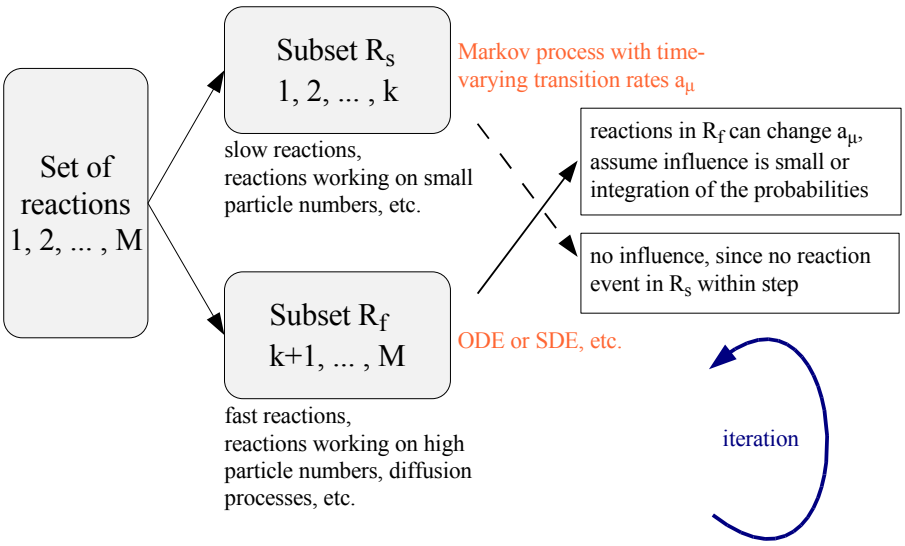


Figure 2.3: Schematic view of a hybrid simulation method.

However, the reaction propensities a_μ of slow reactions are, in general, also dependent on species whose concentrations are changed by fast reactions. Taking this into account, the slow subset no longer constitutes a homogeneous Poisson process. Instead, it has to be described by a Master Equation with time-varying transition propensities $a_\mu(t)$. Gillespie [Gil92] derives the correct Reaction Probability Density Function for this case:

$$P(\tau, \mu | x, t) = a_\mu(t+\tau) \exp\left(- \int_t^{t+\tau} \sum_{1..M} a_\mu(t) dt \right). \qquad (2.16)$$

A number of hybrid algorithms are based on this technique, for instance [Ben04; Neo04; ACT+05; SSK06; GCPS06]. But, they differ in how exactly the non-homogeneous Poisson process is sampled (and also the partitioning strategies used).

Fast reactions involving high-numbered species (e.g. metabolism) can be simulated efficiently with hybrid algorithms. These reactions, in particular, would slow down a pure stochastic simulation considerably. This potential speed-up and that random fluctuations are considered where necessary are the main advantages of hybrid approaches. Nevertheless, there are still some open questions:

- Synchronization of the subnetworks? (Provision for time-varying a_μ in the slow subnet. Conversion between continuous concentrations and discrete particle numbers.)
- Reliable criteria for the partitioning?.
- Dynamic repartitioning → additional computational overhead.
- How to handle fast reactions with small particle numbers involved?

Dynamic/automated (re-)partitioning is not only important when the system state varies considerably over time. It also makes the methods much more user-friendly by avoiding tedious modeler intervention. In 2002 Rao et al. noted in this respect: "[...] hybrid models involving continuous and discrete representations. [...] these approaches require direct intervention by the modeller — a cumbersome and sometimes impossible task. The long-term goal is to develop algorithms that do this both automatically and adaptively" [RWA02, p. 234]. However, since dynamic partitioning causes computational overhead, the decision in favor or against it is always associated with a trade-off between user-friendliness/accuracy and simulation speed.

Likewise, on the one hand, taking into account time-varying $a_\mu(t)$ in the slow/discrete subsystem requires algorithms that are mathematically more involved. On the other hand, the increase in accuracy associated with it renders additional updates of the propensities $a_\mu(t)$ unnecessary so that larger steps can be taken.

Hybrid modeling and simulation might become even more important in the future because of the emergence of ever more complex and heterogeneous models. Kitano remarked [Kit02, p. 209]: "Although some processes can be modelled by either stochastic computation or differential equations alone, many require a combination of both methods. But some biochemical processes take place within a millisecond whereas others can take hours or days. Additionally, biological processes often involve the interaction of different types of process, such as biochemical networks coupled to protein transport, chromosome dynamics, cell migration or morphological changes in tissues."

In the following sections, we will briefly characterize each hybrid method in turn[5] and systematically classify them along different dimensions, namely:

1. which methods are combined,
2. whether the partitioning is dynamic/automated or user-defined,
3. which partitioning policy is used and

[5]These subsections can safely be skipped on a first reading.

4. if time-varying $a_\mu(t)$ in the slow/discrete subsystem(s) are considered or not ($\rightarrow$ synchronization).

Table 2.1 summarizes the results in a concise form.

Alur et al. 2001

In [ABI+01] the authors describe an agent-based framework for the modeling and simulation of hybrid models using the specification language CHARON [AGH+00]. The different agent instances correspond to certain components of the whole system. A combination of ODEs and mode-switching is used. An agent changes from discrete (Direct Method) to continuous (numerical integration of ODEs) behavior according to the current particle number in the corresponding component. Though the partitioning is dynamic, it is rather heuristic. Also, possibly time-varying propensities $a_\mu(t)$ within one step in the discrete components are not considered.

Haseltine and Rawlings 2002

Haseltine and Rawlings [HR02] investigate the reaction-wise partitioning of the Chemical Master Equation into a slow and fast reaction subset. Even though they mention the possibility of dynamic partitioning, only static partitioning is used according to a heuristic for the separation of reaction velocities. The behavior of the fast subsystem is approximated using numerical integration of ODEs or SDEs whereas the Direct Method is used for the slow subnetwork. Haseltine and Rawlings discuss the handling of time-varying a_μ in the slow subnetwork but apply a constant a_μ approximation instead. Because of this, they introduce a "probability of no reaction" which facilitates the scaling of the average step-size and, thus, can limit the approximation error.

Pahle et al. 2002

This hybrid method combines ODEs with the Next Reaction Method. It dynamically partitions the system on the basis of particle numbers using a hysteresis scheme but does not consider time-varying propensities a_μ in the slow subsystem.

This method was developed in my diploma thesis [Pah02] and integrated into the software system COPASI [Cop] during my PhD work. For a detailed description of the algorithm and some implementation issues, see Sec. 4.2.

Hybrid approaches	Methods combined	Dynamic partitioning	Partitioning criteria	Time-varying a_μ
Alur 2001 [ABI+01]	Direct Method / ODE	√	particle no.	
Haseltine 2002 [HR02]	Direct Method / ODE or SDE		heuristics	√/×
Pahle 2002 [Pah02]	Next Reaction Method / ODE	√	particle no.	
Adalsteinsson 2004 [AME04]	Direct Method / ODE		user-defined	
Bentele 2004 [Ben04]	Next Reaction Method / SDE	√	relat. fluct., particle no.	√
Burrage 2004 [BTB04]	Direct M. / τ-Leap / SDE	√	propensities, particle no.	
Kiehl 2004 [KMS04]	Direct Method / ODE		user-defined	√/×
Neogi 2004 [Neo04]	Stoch. Sim. / ODE	√	particle no.	√
Puchalka 2004 [PK04]	Next Reaction M. / τ-Leap Method	√	substrate no., relat. prop.	
Takahashi 2004 [TKHT04]	Next Reaction Method / ODE		user-defined	
Vasudeva 2004 [VB04]	Direct Method / ODE	√	propensities, particle no.	
Alfonsi 2005 [ACT+05]	Next Reaction Method / SDE	√	propensities	√
Salis 2005 [SK05a; SSK06]	Next Reaction Method / SDE	√	propensities, particle no.	√
Griffith 2006 [GCPS06]	Direct Method / ODE	√	propensities, particle no.	√
Harris 2006 [HC06]	Direct M. / τ-Leap / Langevin M. / ODE	√	propensities	
Wagner 2006 [WMP06]	First Reaction M. / discr. Gauss / ODE	√	error criterion	

Table 2.1: Overview and systematic classification of the different hybrid simulation methods. √/× denotes that the authors describe how this feature can be realized in principle but do not demonstrate it in their examples or implementation.

BioNetS (Adalsteinsson et al. 2004)

The BioNetS software [Bio; AME04] includes a hybrid simulation method that uses discrete stochastic simulation and the chemical Langevin equation on discrete and continuous reaction subsets respectively. The user has to specify which of the species should be discrete and which should be continuous. This leads to a static partitioning of the system where reactions exclusively involving continuous species are marked continuous.

The discrete subsystem is propagated using two different methods (dt is a user-defined time step and M_d the number of discrete reactions):

- If $(\sum_{\mu=1}^{M_d} a_\mu)dt \leq \epsilon$, a fixed-time step version of the stochastic simulation is used. Two random numbers are used to determine a) if a discrete reaction occurs, and b) which reaction it will be.

- Otherwise Gillespie's Direct Method is used to determine a stochastic reaction time τ and the type of the next discrete reaction.

The continuous subset is then either simulated over time dt or τ using a semi-implicit Langevin equation solver which completes that step. In this algorithm the influence of the continuous subset on the discrete subset of reactions is neglected.

Bentele et al. 2004

The method proposed in [Ben04; BE04] involves a generalized system state that includes both the particle numbers and a vector of additional variables, one for each reaction in the discrete/slow subsystem. These variables are used to sample the non-homogeneous Poisson process representing the slow subsystem as follows: For each slow reaction μ they give the probability for no reaction event $P_\mu^0(t)$ (since the last occurrence). They are numerically integrated along with the fast system ($\frac{\mathrm{d}P_\mu^0(t)}{\mathrm{d}t} = -P_\mu^0(t) \cdot a_\mu(t)$) as it is propagated over time. Each zero-crossing of one of the residuals is detected and the corresponding slow reaction event is realized. The fast subsystem is either described by an ODE or a chemical Langevin equation. In the latter case it is simulated using the Euler-Maruyama Method for the numerical integration of stochastic differential equations (SDE). The method is able to dynamically partition the system depending on the level of relative fluctuations in particle numbers and absolute particle numbers.

Burrage et al. 2004

Instead of partitioning the reaction system into just two subsystems the hybrid algorithm in [BTB04] even uses a 3-level fragmentation with the Direct Method, the τ-Leap Method and a numerical solver for SDEs (Euler-Maruyama) as the respective simulation methods. However, no theoretical basis for the 3-level partitioning scheme is given and the user has to define thresholds for a heuristic dynamic partitioning scheme based on particle numbers and propensities a_μ. There is no consideration of time-varying a_μ in the slower subsystems.

Kiehl et al. 2004

This hybrid method (see [KMS04] for details) is based on a static user-defined partitioning of the system according to simple heuristics. The behavior of the fast subset of reactions is approximated by numerical integration of ODEs (Runge-Kutta Method) and the Direct Method is used on the slow subset.

This approach considers time-varying propensities a_μ within one step in the discrete/slow subset due to the influence of the continuous/fast subset. However, first a tentative step size is estimated using the a_μ values at the beginning of the step only. Then, the correct step size is determined while numerically integrating the fast subset over the previously estimated interval. If the correct step size turns out to be smaller than the previously estimated one the algorithm backs up.

For the example calculations on the λ phage system in [KMS04], calculation of time-varying a_μ is turned off and the propensities in the slow subset are assumed to be approximately constant during one time step.

Neogi 2004

In [Neo04] a combination of numerical integration of ODEs with exact stochastic simulation is described. The system is dynamically partitioned depending on the particle numbers (using the reaction rates has only been envisaged). To account for time-varying a_μ in the discrete subsystem "thinning" is used. That means the corresponding non-homogeneous Poisson process is sampled by first calculating the next discrete reaction event using a maximal rate. Then the correct reaction rate is determined and the reaction event is only realized if a generated random number is smaller than the ratio (correct rate)/(maximal rate).

Puchalka and Kierzek 2004

Puchalka and Kierzek's "Maximal Timestep Method" [PK04] uses a combination of the Next Reaction Method [GB00] and Gillespie's τ-Leap Method [Gil01]. The system is dynamically partitioned using two criteria for reactions to be assigned to the fast subset:

$$\min\{X_i : S_i \text{ is substrate of } R_\mu\} > n \tag{2.17}$$

and

$$a_\mu(x)/a_0(x) > r \tag{2.18}$$

n and r are two user-defined parameters. Also, the maximal leap size is bounded by another user-defined parameter κ. Using the τ-Leap Method on the fast subsystem has the drawback that, compared with methods employing numerical integration of ODEs [Neo04; SK05a; ACT+05; GCPS06], here it is not clear how to correctly consider the influence of the fast subsystem on the slow subsystem, i.e. time-varying a_μ during one step. However, since one of the conditions of the τ-Leap Method is precisely that the propensities a_μ do not change considerably during one leap, this influence can be neglected.

Takahashi et al. 2004

The simulation approach described in [TKHT04] uses a meta-algorithm that knows about the dependencies between so-called "steppers" (modules implementing different simulation algorithms). The steppers represent the different subsystems. Their interaction is driven by the meta-algorithm using a scheduling scheme. Three subclasses of steppers are provided, including "DifferentialSteppers" implementing continuous numerical integration of ODEs (Runge-Kutta Methods) and "DiscreteSteppers" implementing discrete stochastic simulation (Next Reaction Method). The state variables reside in a global vector and are changed by the steppers as the simulation proceeds.

The method by Takahashi et al. requires intensive modeler intervention, because the system needs to be partitioned and the appropriate steppers have to be set up by hand prior to the simulation. The partitioning is static and does not change. Also, in their test models used the faster subsystems are uni-directionally dependent on the slower subsystems. That means that time-varying $a_\mu(t)$ are not considered and the influence of the fast subsystems on the slower ones is neglected.

This algorithm is integrated into the E-Cell environment [E-C].

Vasudeva and Bhalla 2004

In [VB04] a combination of the (exponential) Euler Method for the numerical integration of ODEs and an approximate fixed time step stochastic simulation is proposed.

This approach uses no explicit partitioning of the system. Rather reactions are numerically integrated whenever

1. their propensities a_μ are sufficiently high and

2. none of the dependent particle numbers are below a predefined limit.

Otherwise their reaction events are determined stochastically. This leads to an implicit, but dynamic, partitioning of the system. The threshold values are user-defined.

Since it is difficult to extend that simple scheme to more sophisticated numerical ODE solvers using adaptive-time steps the restriction to the Euler Method is a severe limitation in many practical cases.

A GENESIS/Kinetikit implementation is available online [Kin].

Alfonsi et al. 2005

In Alfonsi et al. [ACT+05] three different hybrid calculation schemes are described: a Direct Hybrid Method, a First Reaction Hybrid Method and a Next Reaction Hybrid Method. They combine the respective stochastic simulation methods with numerical integration of ODEs (Runge-Kutta method). A residual for the combined reaction propensity of the slow subnet (Direct Hybrid Method) or one residual for each slow reaction (First Reaction and Next Reaction Hybrid Methods) are numerically integrated along with the fast subsystem. Zero-crossings of the residuals indicate slow reaction events. The system is dynamically partitioned using the reaction propensities as criterion.

Salis and Kaznessis 2005

The algorithm by Salis 2005 [SK05a] ([SSK06] describes the implementation of the method on a mainframe computer) is similar to the next reaction hybrid method in [ACT+05]. The differences are that the dynamic partitioning is based on two criteria, the reaction propensity and particle numbers of participating species, and that an additional approximation allows the firing of multiple slow reactions within one step.

Griffith et al. 2006

Griffith et al. [GCPS06] combine ODEs with stochastic simulation. They try to improve on previous partitioning schemes by setting up the following two conditions, which depend on two user-defined parameters γ and Λ:

$$X_i > \gamma \cdot |v_{i\mu}| \qquad \forall i : S_i = \{\text{reactant or product of } R_\mu\} \tag{2.19}$$

and

$$a_\mu(X) \geq \Lambda \cdot a_{\max} \tag{2.20}$$

with $a_{\max}$ the rate of the fastest discrete reaction.

The first condition ensures that reactions are put in the stochastically simulated subset whenever low numbers of substrate or product particles necessitates a discrete treatment. The second condition is a realization of a criterion for partitioning, which was mentioned already in the article by Haseltine 2002 [HR02], namely, that there must be a large enough separation between the reaction rates in the slow and the fast subsystem respectively to make up for the computational overhead caused by dynamic partitioning and switching of reactions. Only those reactions where a continuous treatment is profitable, i.e. they are much faster than the slow reactions and keeping them in the slow subsystem would lead to many time-consuming discrete reaction events, are switched to the fast subsystem.

A nice property of this approach is that there is no absolute threshold. Rather the threshold for repartitioning can change dynamically according to the current slow subsystem's reaction rates.

Also, the implementation takes into account time-varying propensities by numerically integrating the sum of the propensities of the slow reactions together with the state of the fast subsystem until time τ where τ fulfills $\int_t^{t+\tau} \sum_{1..M} a_\mu(t)dt = r_1$, with r_1 a unity mean exponentially distributed random number. Then using a second (uniformly distributed) random number r_2 one of the slow reactions is chosen as in Gillespie's Direct Method, the reaction is realized and the numerical integration continues until the next discrete event.

For the numerical integration, a fifth-order Runge-Kutta method or the stiff/non-stiff solver CVODE [CH96] are used.

Harris and Clancy 2006

The so-called "Partitioned Leaping Method" by Harris and Clancy [HC06] extends the τ-Leaping approach by integrating exact stochastic simulation, τ-Leaping, the Langevin Method (SDE) and the explicit Euler method (ODE) into one algorithmic framework. On a single reaction basis the reactions are classified, depending on their propensities, as belonging to one of the four regimes and propagated accordingly. Hence, the dynamic partitioning takes place in an implicit manner.

Wagner et al. 2006

The hybrid method described in [WMP06] is, like the one by Burrage et al. 2004, a 3-level method dividing the reactions into three subset. The calculation scheme combines a modified First Reaction Method, an approximate method using discrete Gaussian distributions and an ODE solver. The partitioning is done dynamically on the basis of a single tolerance parameter that describes the maximal error per approximation. Time-varying $a_\mu(t)$ in the slower reaction subsets are neglected. Even though the use of a single tolerance parameter is very attractive from a user's point of view, this method involves a number of internal approximations that are difficult to justify.

2.4.4 Stochastic quasi-equilibrium approximation

Several authors proposed the use of quasi-equilibrium approximations (QEA) in stochastic simulations,e.g. [CGP05a; CGP05d; ELVE05; SK05b; Gou05; HR05; SV05] (see also [GPC07; ELVE07] for a discussion).

For instance, Cao et al. [CGP05d] devised the "Slow-Scale Simulation Method" in which the system is subdivided reaction- and species-wise into a fast and a slow subsystem. A "virtual fast process" is defined which consists of the fast species evolving under only the fast reactions, i.e. all slow reactions are turned off. This virtual fast process is Markovian. Also, it needs to be stable in a way such that its relaxation time is much faster than the expected time to the next slow reaction. Then, on the slow time scale, it can be approximated using its asymptotic solution. Finally, the system is simulated considering only the slow reaction events explicitly. For this, the Direct Method can be used with modified $\overline{a}_\mu$ ("slow-scale propensity functions) in order to take into account the influence of the (asymptotic) virtual fast process. In each step, values for the fast species are determined by drawing random samples from the asymptotic probability distribution. Cao et al. give two simple examples for calculating the slow-scale propensities

$\overline{a}_\mu$ in their article, but this can be difficult for more complex models.

The multiscale stochastic solver by Samant 2007 [SOV07] is also based on stochastic QEA. This quite involved method uses the Computational Singular Perturbation (CSP) method by Lam, Goussis et al. [Lam93; LG94] for the partitioning of the system and the Modified τ-Leaping Method [CGP05c] for simulation.

Stochastic quasi-equilibrium approximations can be regarded as residing in the middle between exact stochastic simulation methods and hybrid approaches. The system is partitioned into subsets of reactions or species as in hybrid simulation methods. However, only the slow/discrete subset is explicitly simulated (stochastically).

2.5 Conclusions

Stochastic modeling and simulation is important whenever fluctuation phenomena play a role either as a destructive or constructive element. We will see in Chap. 5 that this is not restricted to systems containing only few particles but can also happen in larger systems that are sensitive or operate near bifurcation points.

Despite of the multitude of proposed stochastic methods, there seems to be none that fits all problems. For smaller models in terms of particle numbers and whenever a correct treatment of the fluctuations is required, one of the exact methods should be used. They are relatively straightforward to implement and there also exists a number of corresponding software tools, for instance COPASI or DIZZY [Cop; Diz], which makes it easy to use them without the need of programming. Since they are mathematically equivalent it is a matter of taste which one of them to employ. For bigger systems though, they are prohibitively computationally expensive.

For an accelerated stochastic simulation, one of the approximate methods could be considered. However, some of them are ad-hoc procedures tailored to specific problems rather than general stochastic solvers. The StochSim algorithm might be interesting, if one wants to model multistate molecules or to trace the life-cycle of single particles. The τ-Leap Method or one of its variants seems to be a promising approach, but it suffers from the need to find a correct τ value. This problem is difficult, in particular when stiffness comes into play. All approximate methods should be used with care. Their assumptions have to be thoroughly checked in each case. Otherwise they can misdescribe the fluctuations (e.g. the PW-DMC method and different forms of lumping) in some cases. With the exception of the τ-Leap Method (implemented in [Cel; Diz; E-C; Sma] general software tools are mostly missing. In

addition, these approximate methods are rather cumbersome to implement and often need intervention by the modeler.

The most promising direction is the development of hybrid methods because they directly tackle the pressing problem of stiffness in the system. They appear to be flexible enough to allow for general stochastic solvers in the future even for very big and heterogeneous models. However, for the time being an established type of partitioning (reaction-wise and/or species-wise) and, above all, reliable criteria for an automatic and adaptive partitioning during the simulation are still missing. Hybrid algorithms are the most challenging methods to implement. They also still need much user intervention unless automatic partitioning is implemented. Only a few software tools exist, which permit hybrid simulation, e.g. [HSG+06; AME04; SSK06], but this is expected to change in the future.

See Appendix A for a list of software systems which support stochastic simulation of biochemical systems.

Chapter 3

Calcium signal transduction

Mmm milk.
Good for the bones,
good for the kids.
I pity the fool
who ain't got no calcium in his diet!

B.A., Robot Chicken (1.16)

Since calcium signal transduction is a repeatedly occurring model system throughout the following chapters a brief introduction shall be given here of how dynamic calcium behavior is believed to arise, and how it could be decoded by target enzymes in cells. One simple model of calcium oscillations which will be used later on is described in detail. Finally, the experimental setup for measuring time-varying calcium levels in single rat hepatocytes is specified.

3.1 Signal transduction via Ca^{2+} ions

Calcium ions serve as important and versatile second messengers. Calcium signaling is ubiquitous and controls a variety of cell functions from fertilization, secretion, enzyme activation and gene expression to cell death [BBL98; Car02], in many excitable and non-excitable cell types.

The cytosolic concentration of Ca^{2+} in most cells is kept at low values ($\sim 0.1-0.2\,\mu M$) by calcium ATPases that pump calcium ions from the cytosol out of the cell, SERCA (sarco-/endoplasmic reticulum calcium ATPases), that fill intracellular calcium stores, and mitochondrial calcium pumps. After the activation of specific membrane-bound G-protein coupled receptors by

agonists like nucleotides (e.g. ATP, UTP) or hormones (vasopressin, acetylcholin, angiotensin II, etc.) receptor-associated G_α subunits change GDP for GTP and get activated. This, in turn, leads to the activation of phospholipase C (PLC) in the cell membrane. Activated PLC catalyzes the hydrolysis of phosphatidylinositol-4,5-bisphosphate to diacylglycerol (DAG) and inositol-1,4,5-trisphospate (IP_3) (These two signaling branches have a downstream cross-link, when DAG and Ca^{2+} together activate protein kinase C). IP_3 diffuses through the cytosol and binds to specific IP_3-dependent receptors in the endoplasmic membrane, which leads to an influx of calcium ions into the cytosol. This fast process has an autocatalytic character (Ca^{2+}-induced Ca^{2+} release, CICR) and causes a sharp rise of cytosolic calcium concentration up to $0.5 - 1.0\,\mu M$. Also, calcium ions from the extracellular space enter the cell which seems to be needed for sustained oscillations in some cell types. Eventually, calcium is re-sequestered into intracellular compartments or pumped back out of the cell again and the cytosolic concentration drops to its basal level. Fig.3.1 illustrates this process in a schematic manner.

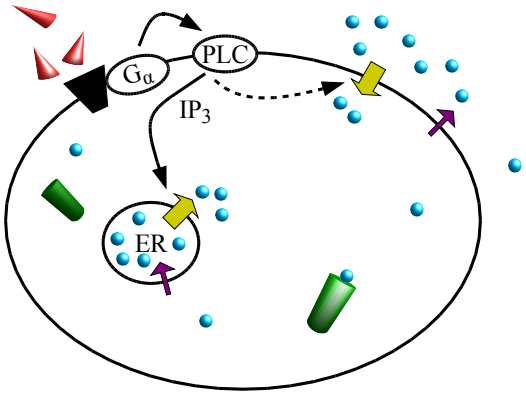


Figure 3.1: Schematic view of calcium signal transduction. Ca^{2+} ions (blue balls), agonists (red cones), G-protein coupled receptors (black), target enzymes (green cylinders). See text for a detailed description.

An intriguing fact is that, even in non-excitable cells like hepatocytes, the concentration of cytosolic calcium can display regular (spiking) or more complex (bursting) oscillations or prolonged elevated levels [WCC86] after stimulation by an agonist and depending on the nature of this agonist. Some examples of different calcium dynamics can be seen in Fig. 3.2. This oscillatory

behavior is not only believed to save the cell from the toxic effects of sustained high cytosolic calcium levels and from desensitization, but has also been shown to increase the efficiency of calcium signaling [DXL98]. In addition to these temporal patterns of calcium dynamics, interesting spatio-temporal patterns (e.g. calcium puffs and waves) have been described [BBL98; Fal03a]. However, we will concern ourselves in this study with temporal patterns only.

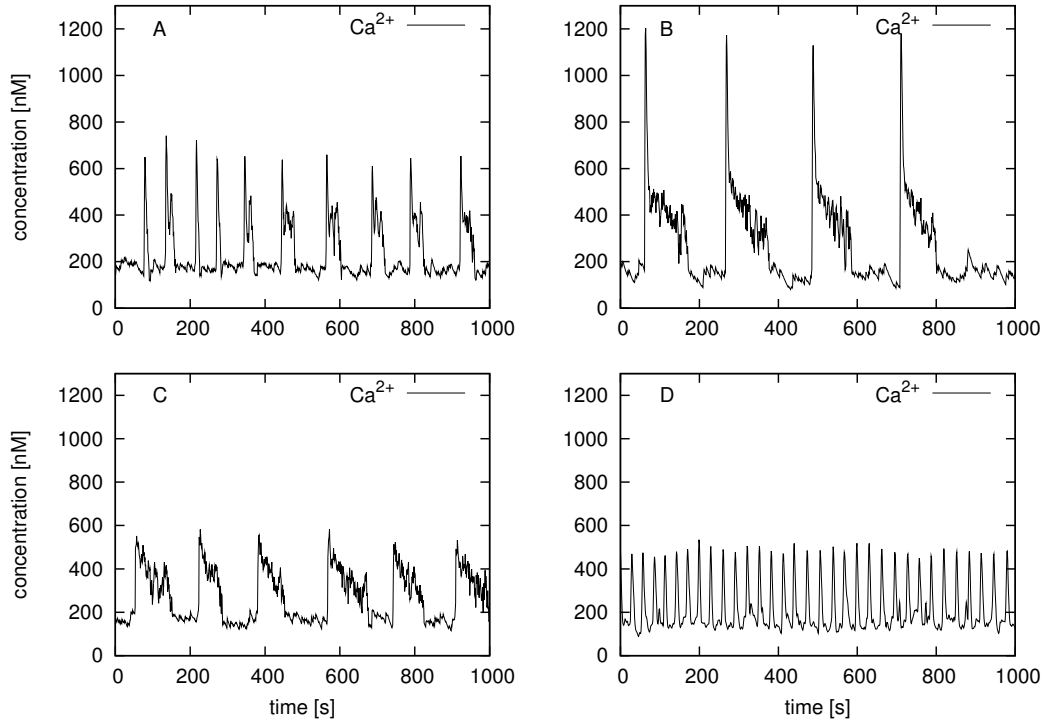


Figure 3.2: Different calcium dynamics in single rat hepatocytes. (A) mix of calcium spikes and bursts (stimulated with 1.2 µM ATP). (B) bursting with distinct primary spikes and smaller secondary spikes (stimulated with 1 µM ATP). (C) strong bursting (stimulated with 1.5 µM ATP). (D) high-frequency spiking (stimulated with 2 µM phenylephrin). Data kindly provided by Anne K. Green and C. Jane Dixon.

3.2 Models of calcium signaling

Due to both its importance for the functioning of many cell types and its interesting dynamics [KS01], calcium signal transduction has attracted numerous theoretical studies. Many different models of calcium signaling have been proposed, ranging from simple one-pool models [SS91] to more elaborate

ones [LOK04] incorporating many different processes. For reviews on calcium models, see Schuster et al. 2002 [SMH02] or Dupont et al. 2000 [DSC$^+$00].

Most of these models focus on the simulation of simple periodic oscillations (spiking). Only few models are able to display periodic bursting oscillations [SL95; HDG99; KOD$^+$00; MHBH00], let alone aperiodic bursting oscillations [KOD$^+$00; MHBH00] in non-excitable cells.

We will use one of these models, a relatively simple receptor-operated model developed in [KOD$^+$00]. This model, even though phenomenological and not aiming at representing all known physiological details, is able to display simple and complex behavior, depending on the kinetics of the receptor complex and thus depending on the agonist-specific receptor, as occurs in real cells (cf. Table 3.1 for a list of reactions and corresponding kinetic functions of the core model).

reactions			kinetics
	$\rightarrow$	G_α	$\nu_1 = k_1$
	$\xrightarrow{G_\alpha}$	G_α	$\nu_2 = k_2 \cdot [G_\alpha]$
G_α	$\xrightarrow{PLC}$		$\nu_3 = \frac{k_3 \cdot [G_\alpha] \cdot [PLC]}{K_4 + [G_\alpha]}$
G_α	$\xrightarrow{Ca^{2+}}$		$\nu_4 = \frac{k_5 \cdot [G_\alpha] \cdot [Ca^{2+}]}{K_6 + [G_\alpha]}$
	$\xrightarrow{G_\alpha}$	PLC	$\nu_5 = k_7 \cdot [G_\alpha]$
PLC	$\rightarrow$		$\nu_6 = \frac{k_8 \cdot [PLC]}{K_9 + [PLC]}$
	$\xrightarrow{G_\alpha}$	Ca^{2+}	$\nu_7 = k_{10} \cdot [G_\alpha]$
Ca^{2+}	$\rightarrow$		$\nu_8 = \frac{k_{11} \cdot [Ca^{2+}]}{K_{12} + [Ca^{2+}]}$

Table 3.1: Model of calcium oscillations [KOD$^+$00]. Parameters: $k_1 = .212$, $k_3 = 1.52$, $K_4 = .19$, $k_5 = 4.88$, $K_6 = 1.18$, $k_7 = 1.24$, $k_8 = 32.24$, $K_9 = 29.09$, $k_{10} = 13.58$, $k_{11} = 153$, $K_{12} = .16$. k_2 is bifurcation parameter and is set to different values in simulations depending on the desired behavior.

Here, G_α denotes the active subunit of the G-protein, PLC the activated form of PLC, and Ca^{2+} cytosolic calcium. G_α is activated upon binding of an agonist (included in k_2) and this process is autocatalytic. There is also a small term (k_1) for the spontaneous activation of G_α. It is inactivated via two processes, one being activated by Ca^{2+} (via phosphokinase C) and

one by PLC. PLC is activated by G_α and inactivated by a simple enzymatic reaction. Finally, G_α also triggers the increase of calcium concentration in the cytosol and calcium is removed by an active transport mechanism.

3.3 Decoding of the Ca^{2+} signal

As we have seen, the rise of calcium concentration in the cytosol upon stimulation by an agonist is not uniform and, in most cases, the calcium concentration oscillates in response to receptor stimulation.

Because the second messenger Ca^{2+} carries information coming from different sources (different hormones, neurotransmitters, etc.) to a variety of cellular targets like transcription factors (e.g. NF-κB, Oct/OAP or NF-AT [DXL98; LLW$^+$98]) or proteins (e.g. calmodulin, CaM kinase II [DS98], glycogen phosphorylase, PKC [OM98]), it is believed that the different signals are encoded by means of amplitude-, frequency-, duration-, timing- and/or shape-modulation of these calcium oscillations (see Fig. 3.3). The FM mode of calcium signals [Ber97] has been studied experimentally [DS98; LLW$^+$98; DXL98; OM98] and theoretically [PLvzM$^+$98; GBD00; DHD03; SPH04; SPH08] and can be considered as an established fact. With increasing stimulus strength the frequency of calcium oscillations usually rises and cellular targets can be activated differentially by different frequencies. Recent studies showed that, depending on the waveform of bursting calcium oscillations, calcium-binding proteins can be activated differentially at the same time making use of the cooperative nature of calcium binding [LOK04; SKM05].

Stimulation of hepatocytes with, e.g., vasopressin results in spiking calcium oscillations [WCC86]. When stimulated with ATP, bursting oscillations are observed [DWCC90]. These differences in dynamic behavior offer an explanation for the differences in physiological response, which occur when different stimuli are applied.

A variety of proteins are influenced by intracellular Ca^{2+} levels (calmodulin, CaM kinases, glycogen phosphorylase B, PKC, calcineurin; see [Cel96] for a review). The most prominent class in that respect is the family of EF hand proteins including the ubiquitous calmodulin. Ca^{2+} activation of calmodulin and similar proteins happens in a cooperative manner. Calmodulin has four binding sites with high affinity ($K_d \approx 0.1 - 1\,\mu M$) for Ca^{2+}. That is the reason why in numerical studies [PLvzM$^+$98; LOK04; SKM05] often a Hill term of 4^{th} order is used.

In addition to this cooperative activation Prank et al. [PLvzM$^+$98] added an autophosphorylation term, that leads to a memory effect of activation.

In Schuster et al. [SKM05] not only a Ca^{2+}-activated enzyme is studied, but also one, whose regulation includes an inhibitory term resulting in a biphasic activation profile.

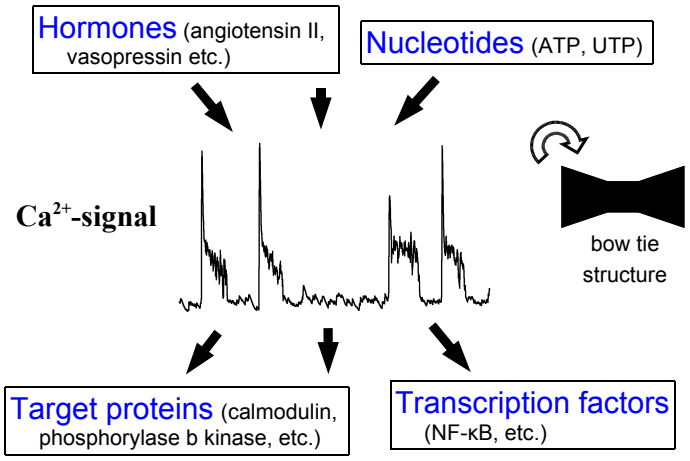


Figure 3.3: The bow tie structure of calcium signaling [SKM05].

3.4 Experiments

Single hepatocytes were isolated from fed male Wistar-strain rats (150 – 250 g) by collagenase perfusion as described previously [DCG95]. Briefly, the hepatic portal vein was cannulated and an initial Ca^{2+}-free perfusion was followed by perfusion with collagenase (0.04 % w/v) and Ca^{2+} (3.8 mM) for 15 min. The perfusion rate was 30 ml/min throughout. The cells were harvested and incubated at 37 °C at low density (10^3 cells/ml) in 2 % type IX agarose in William's medium E (WME). Single hepatocytes were prepared for microinjection with the bioluminescent Ca^{2+} indicator aequorin as described previously [CL91]. The injected cell was transferred to a perfusable cup held at 37 °C, positioned under a cooled, low-noise photomultiplier, and continuously superfused with WME, to which agonists were added. Photon counts were sampled every 50 ms by computer. At the end of an experiment, the total aequorin content of each cell was determined by discharging the aequorin by lysing the cell. The data were normalized retrospectively by computer, by calculating the photon counts per second divided by the total

counts remaining. The computed fractional rate of aequorin consumption could then be plotted as $[Ca^{2+}]_i$ using in vitro calibration data and exponential smoothing with time constants: for resting $[Ca^{2+}]_i$, 12 s; for transients, 1 s.

3.4.1 Materials

Aequorin was provided by Prof. O. Shimomura (Marine Biological Laboratory, Woods Hole, MA, U.S.A). Collagenase was obtained from Roche Diagnostics (Lewes, U.K.) and WME from Invitrogen (Paisley, U.K.). Agarose and agonists were purchased from Sigma-Aldrich (Poole, U.K.).

Chapter 4

Stochastic simulation in Copasi

> Beware of bugs in the above code;
> I have only proved it correct,
> not tried it.
>
> Donald E. Knuth

The software system Copasi [Cop], the Complex Pathway Simulator, is a comprehensive and user-friendly tool for the studying of biochemical networks. It provides a variety of different simulation and analysis methods. One very convenient feature is that deterministic and stochastic simulation can be used on the same model. In the following Copasi's stochastic simulation methods will be presented, in particular the hybrid solver whose implementation was part of this PhD work. Also, some issues concerning the switch between deterministic and stochastic methods will be discussed. Copasi is the result of the collective work by many people, in particular Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Natalia Simus, Irina Surovtsova, Mudita Singhal, Liang Xu, Pedro Mendes and Ursula Kummer. Parts of this chapter have been published in Bioinformatics in 2006 [HSG+06] and in [SSPK06; SGP+06].

4.1 The software system Copasi

Copasi is a comprehensive biochemical simulation and analysis tool with the following features:

- User-friendly graphical user interface (GUI) and command line version for batch processing.
- Deterministic (LSODA [Pet83]), stochastic and hybrid simulation.

- Analysis: moieties, steady states, metabolic control, sensitivities, elementary modes, Lyapunov exponents.
- Parameter scans, optimization, parameter estimation.
- File formats supported:
  SBML [SBM], COPASI's XML format, GEPASI file format, BERKELEY MADONNA [Mad], XPPAUT [XPP], C source code.
- Available for different platforms:
  LINUX, MS WINDOWS, MAC OS X and SOLARIS SPARC.

COPASI is being developed as a joint project between the Kummer group at the University of Heidelberg, Germany (formerly at the EML Research gGmbH, Heidelberg) and the Mendes groups at the Virginia Bioinformatics Institute, USA and at the University of Manchester, UK. It is free for non-commercial use and can be downloaded from `http://www.copasi.org`.

The main stochastic solver implemented in COPASI uses the Next Reaction Method by Gibson and Bruck (cf. Chap. 2.4.1). In addition to the hybrid solver which is described in the following section, we also integrated the τ-Leap Method (see Sec. 2.4.2) as a module. However, due to the experimental nature of this method this feature is not yet activated in the publicly available version.

COPASI is also capable of computing the divergence of the system in order to support the user in making a decision whether or not stochastic simulation is necessary (see Chap. 5, for details).

4.2 Hybrid simulation method

Our hybrid method combines the stochastic simulation algorithm by Gibson and Bruck (Next Reaction Method; Sec. 2.4.1) with different algorithms for the numerical integration of ODEs (4^{th} order Runge-Kutta, LSODA — see [Pet83]). The biochemical network is dynamically partitioned into a deterministic and a stochastic subnet depending on the current particle numbers in the system. The user can define limits for when a particle number should be considered low or high. The stochastic subnet contains reactions involving low numbered species as substrate or product. The remaining reactions, i.e., all those that only affect high numbered species, form the deterministic subnet. The two subnets are then simulated in parallel using the stochastic and deterministic solver, respectively. The reaction probabilities in the stochastic subnet are approximated as constant between two stochastic reaction events.

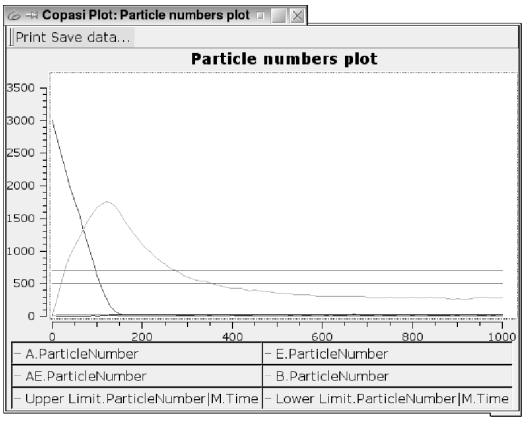


Figure 4.1: Result of the hybrid simulation of the system in Eq. (4.1) in a time interval 10^3 s, for a volume $5 \cdot 10^{-19}$ ml, initial concentrations $[A](t_0) = 10\,\text{mMol/ml}$, $[E](t_0) = 0.1\,\text{mMol/ml}$, $[AE](t_0) = 0\,\text{mMol/ml}$, $[B](t_0) = 0\,\text{mMol/ml}$ and parameters $k_1 = 0.01\,\text{mMol/(ml·s)}$, $k_2 = 20\,\text{ml/(mMol·s)}$, $k_3 = 10\,\text{s}^{-1}$, $k_4 = 1\,\text{s}^{-1}$, $k_5 = 0.01\,\text{s}^{-1}$. The lower and the upper limit for the particle numbers are 500 and 700, correspondingly.

As an example, we investigate a simple open biochemical system where A reacts to B catalyzed by E. The corresponding reaction system is the following:

$$\begin{aligned} \rightarrow A \\ A + E \rightleftharpoons AE \\ AE \rightarrow B + E \\ B \rightarrow \end{aligned} \tag{4.1}$$

This corresponds to the following systems of equations:

$$\begin{aligned} [A]' &= k_1 - k_2 \cdot [A] \cdot [E] + k_3 \cdot [AE] \\ [E]' &= -k_2 \cdot [A] \cdot [E] + k_3 \cdot [AE] + k_4 \cdot [AE] \\ [AE]' &= k_2 \cdot [A] \cdot [E] - k_3 \cdot [AE] - k_4 \cdot [AE] \\ [B]' &= k_4 \cdot [AE] - k_5 \cdot [B] \end{aligned} \tag{4.2}$$

In Fig. 4.1 we simulated this system using our hybrid algorithm and low particle numbers.

Whenever the particle numbers drop below the user-defined limits the corresponding reactions are simulated stochastically and random fluctuations are captured. Otherwise the faster numerical integration is used for reactions with high-numbered species.

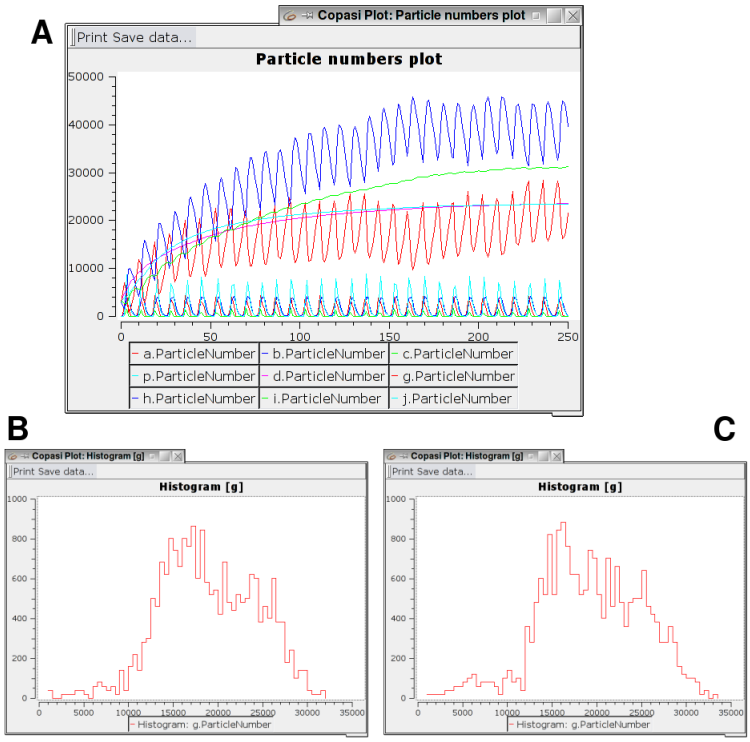


Figure 4.2: Hybrid simulation of the calcium oscillation system in [KKP+05] (see Chap. 3), comprising species a (G_α), b (PLC) and c (Ca^{2+}), coupled to a linear pathway of reactions (species d, g, h, i, j) via a calcium protein buffer complex p, which activates the reaction from g to h. All other steps in the linear pathway have Henri-Michaelis-Menten kinetics. Shown are the particle numbers over time (A) and a comparison of the hybrid (B) (lower and upper particle number limits 9900 and 10100, respectively) and pure stochastic (C) simulations in terms of the particle number histograms of species g.

Our hybrid method, which was developed and described in [Pah02], is similar to the one by Haseltine and Rawlings (2002) [HR02], though in fact they both were implemented independently during the same time. The main difference is the dynamic partitioning with user-defined limits for the particle numbers and the hysteresis-like repartitioning scheme.

The dynamical partitioning is vital, e.g. for oscillating systems, but the speedup is very model-dependent. Because of the computational overhead for partitioning the system the hybrid method can occasionally take longer than pure stochastic methods. In addition, low-numbered species that take part in fast reactions slow the simulation down by forcing the fast reactions to be simulated stochastically. We use two different user-defined limits for the particle numbers (lower and upper limit) and a hysteresis-like updating scheme (metabolites with particle numbers between those limits do not change their status). This avoids unnecessary and time-consuming reaction swaps if the particle numbers are fluctuating in a medium range.

We settled on the simple partitioning criterion using particle numbers for three reasons. First, the amplitude of relative fluctuations of particle numbers are high in low-numbered species. Single reaction events can have a significant impact here. Reactions involving those species should therefore be handled stochastically. Second, most of the computational effort of stochastic simulation algorithms is spent on fast reactions. In order to speed up the simulation, fast reactions should be taken out of the stochastic subsystem and simulated deterministically. The higher the number of substrate particles, the faster the reaction will be. This is true for mass action kinetics and for some parts of the phase space of enzyme kinetics. Third, if only reactions involving high-numbered species are simulated deterministically, the relative changes in particle numbers are minimal. For this reason, the change in reaction probabilities in the stochastic subnet caused by the fast subnet during one step can be neglected.

Our algorithm therefore approximates the influence of the deterministic subnet on the stochastic subnet during one step as constant, that means that the reaction propensities in the stochastic subnet are constant during this time interval.

This hybrid algorithm is able to simulate models faster than pure stochastic methods, while still taking into account random effects in the stochastic subnetwork. If the limits for the particle numbers are set to zero, the whole network will be simulated deterministically. With increasing limits the calculation eventually converges to an exact stochastic simulation of the system. We tested our implementation in this limit successfully on the Discrete Stochastic Model Test Suite [Stoa]. If the particle number limits are in an intermediate range, between two repartitionings of the system the simulation

proceeds similar to the method described and validated in [HR02]. In addition, we compared our hybrid solver to an exact stochastic solver with respect to distributions of species particle numbers in different test systems. One example is shown in Fig. 4.2 where the calcium model described in Chap. 3.1 extended by a linear pathway of reactions has been used. Panels B and C show the resulting probability distributions of particle numbers.

However, we want to stress that due to the still heuristic partitioning criterion it is possible for the user to choose the limits such that the result of the hybrid solver deviates from the exact stochastic result as possible with other hybrid methods.

Nevertheless, the user-defined limits for the partitioning of the system allows to study the sensitivity/robustness of individual subnetworks w.r.t. noise. By changing this limit, one can observe if the resulting trajectory is dramatically changing, e.g. if additional subnetworks are simulated stochastically. If this is the case, these subnetworks exhibit a pronounced sensitivity towards intrinsic noise. Thus, dissection of the whole system w.r.t. noise-sensitive and noise-robust subnetworks is possible.

ODEs describing biochemical networks are often stiff. In our hybrid method we therefore use the LSODA algorithm (Petzold, 1983 [Pet83]), which is adequate for the numerical integration of the deterministic subnetwork in the presence of stiffness. We also implemented a hybrid solver that uses a 4^{th} order Runge-Kutta method for cases when one is certain that the system is never stiff. Because the hybrid calculation requires many separate ODE integrations in small time intervals, the use of a simple one-step solver, such as Runge-Kutta, can be faster since it lacks the computational overhead of predictor corrector methods.

There exist several mathematically equivalent algorithms for the stochastic simulation of biochemical networks (see Sec. 2.4.1). In our hybrid solver we chose the method by Gibson and Bruck for the simulation of the stochastic part of the network as it was the most convenient to integrate into our hybrid calculation scheme.

4.3 Using complex kinetics in stochastic simulations

When stochastically simulating a reaction network which has been described by a set of ODEs all reaction rates have to be transferred to a corresponding reaction probability. This is rather simple and straightforward in the case of mass action kinetics [Gil76]. However, enzyme kinetic rate laws represent a

lumping of terms each corresponding to an elementary mass action reaction; an important question is whether it is justifiable to use such a rate expression with stochastic simulations. Several authors [vGK01; RA03; CGP05a; CGP05b] have shown that as long as the initial assumptions for the assumed kinetics hold (e.g. excess substrate, fast reversible enzyme-substrate-complex formation, etc.), it is indeed justified to assume the enzymatic reaction to constitute one single step with the respective rate law.

Basically, the rate law consists of a mass action part and a kinetic part [Hof95]. The kinetic part depends on reactant amounts and other factors, so it is not constant, but it could be assumed to freeze and become constant for the single reaction event that is computed in each step of the algorithm. This rate then has to be computed anew for the next iteration.

Taking our example from above (Eq.(4.2)), the same system could also be described using the Michaelis-Menten form as long as the formation of AE is fast and reversible compared to product release and as long as there is a substantial surplus of substrate compared to enzyme. In this case the equations lump to the following ones:

$$\begin{aligned} [A]' &= k_1 - V_{max} \cdot \frac{[A]}{K_M + [A]} \\ [B]' &= -k_5 \cdot [B] + V_{max} \cdot \frac{[A]}{K_M + [A]} \end{aligned} \tag{4.3}$$

In Fig. 4.3 we compare time series of the stochastically simulated elaborate system with the lumped system. We made sure that the assumptions for lumping the system hold. As can be easily seen, both trajectories correspond to each other.

4.4 Reversible reactions in stochastic simulations

In order to perform stochastic simulations, reversible reactions have to be handled as separate irreversible forward and backward reactions. In deterministic simulations forward and backward reaction rates can cancel each other out; in stochastic simulations each single reaction event has to be considered separately. COPASI provides a feature that, at the modeler's request, converts all reversible reactions to the corresponding individual forward and backward reactions in order to allow for a stochastic simulation of the model. The tool adjusts the reaction scheme and model description automatically. However, due to the difficulty in dissecting an arbitrary reversible kinetics into two irreversible kinetic functions, fully automatic conversion is only

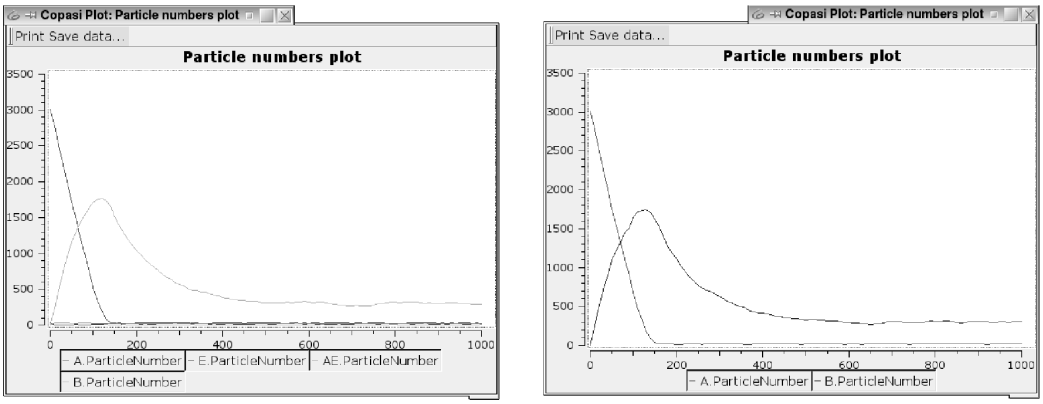


Figure 4.3: Results of stochastic simulations of the detailed (Eq. (4.1)) and the lumped (Eq. (4.3)) system in the time interval 10^3 s, for the volume $5 \cdot 10^{-19}$ ml, initial concentrations $[A](t_0) = 10\,\text{mMol/ml}$, $[E](t_0) = 0.1\,\text{mMol/ml}$, $[AE](t_0) = 0\,\text{mMol/ml}$, $[B](t_0) = 0\,\text{mMol/ml}$ and parameters $k_1 = 0.01\,\text{mMol/(ml·s)}$, $k_2 = 20\,\text{ml/(mMol·s)}$, $k_3 = 10\,\text{s}^{-1}$, $k_4 = 1\,\text{s}^{-1}$, $k_5 = 0.01\,\text{s}^{-1}$, $V_{max} = 0.1\,\text{mMol/(ml·s)}$, $K_m = 0.1\,\text{mMol/ml}$.

guaranteed for mass action kinetics. For more complex kinetics, COPASI uses heuristics to separate the kinetic function into two terms that are meaningful for the forward and backward reactions. If that fails, the user will have to manually adjust the kinetics after the conversion.

In Fig. 4.4 we show the conversion of the reversible mass-action kinetics assumed for the formation of AE in the above example to the individual forward and backward reaction for use in the stochastic simulation.

Acknowledgments

The following people contributed significantly to COPASI by programming, testing and useful discussions (in alphabetical order): Jens Christian Brasen, Diogo Camacho, Carel van Gend, Ankur Gupta, Yonqun He, Wenjun Hu, Katrin Hübner, Adaoha Ihekwaba, Aejaaz Kamal, Mrinmyee Kulkarni, Ann Zahle Larsen, Sarah Lilienthal, Rohan Luktuke, Bharat Mehrotra, Femke Mensonides, Saroj Mohapatra, Ursula Rost, Gaurav Singh, Anurag Srivastava, Iulian Stoleriu, Wei Sun, Sameer Tupe, Ralph Voigt, Andreas Weidemann, Katja Wegner, Jürgen Zobeley. We thank the BMBF, the Klaus Tschira Foundation and the Virginia Bioinformatics Institute for generous funding of this project.

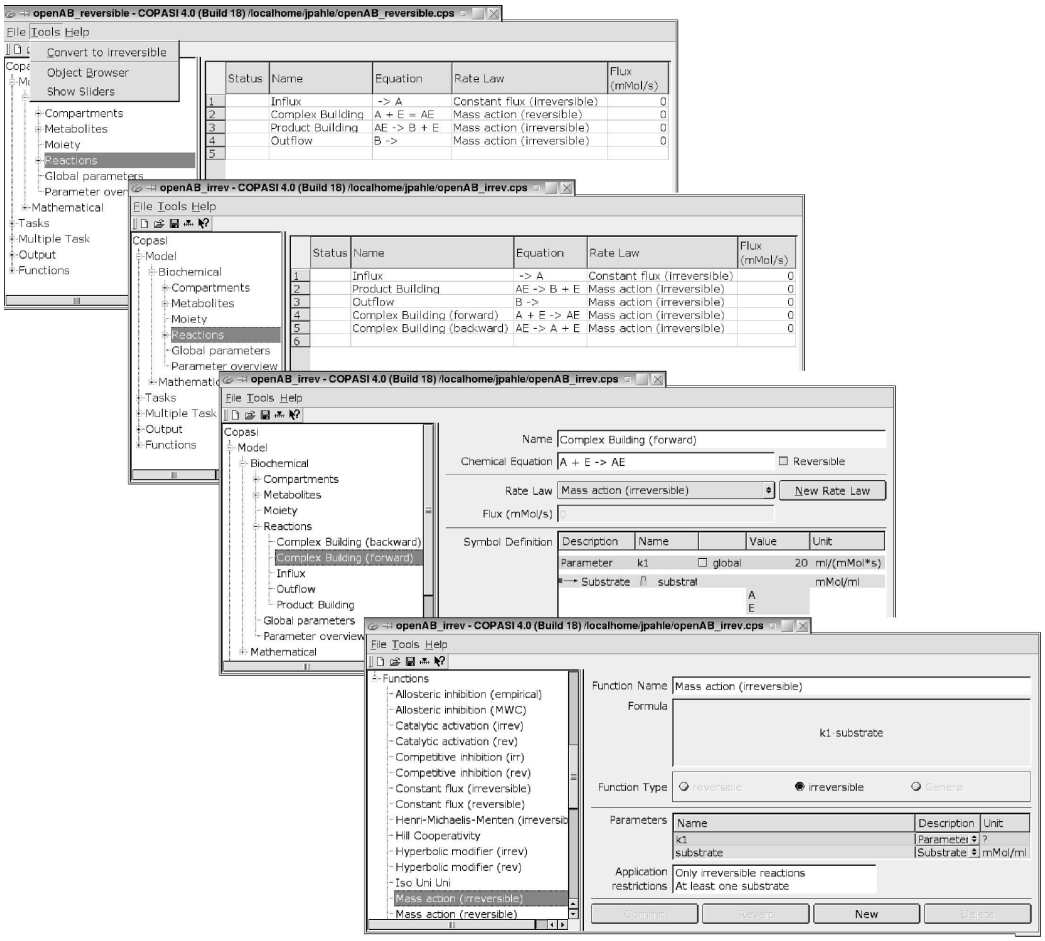


Figure 4.4: Application of the Copasi menu tool "Convert to irreversible" to our model system (Eq. 4.1).

Chapter 5

Transition from stochastic to deterministic behavior in biochemical systems

"[...] at the edge of chaos"

Christopher Langton

In the following chapter we investigate the transition from stochastic to quasi-deterministic behavior in biochemical systems, i.e. how many copies of participating molecules are needed to show a deterministic response. This is important for both the understanding of robustness in cellular systems and for decisions about which simulation method is appropriate. In our main test system, namely signal transduction via calcium, we observe that the transition occurs within a range of particle numbers which roughly corresponds to the number of receptors and channels in the cell, and depends heavily on the attractive properties of the phase space of the respective systems dynamics. The same dependence was found in other test systems, such as the peroxidase-oxidase reaction and MAP kinase cascades.

This work was done in collaboration with Ursula Kummer, Borut Krajnc, Anne K. Green, C. Jane Dixon, and Marko Marhl. It has in parts been published in the Biophysical Journal in 2005 [KKP+05]. A subsequent manuscript is in preparation.

5.1 Introduction

Improved experimental technology has led to the possibility of studying increasingly large biochemical systems *in vivo*. However, the experimental results are often very complex, which is of course due to the underlying complexity of the biochemistry in the living cell itself. This has resulted in the more and more heavy use of computational means to support experimental investigations. Simulation and modeling are now being employed regularly to understand the dynamic properties of a biochemical network. One problem of computational investigations is that the choice of, e.g., the simulation method relies on rather heuristic, if any, rules. However, the more intensive use of these methods asks for reliable and analytical decisions.

Simulations of biochemical systems have mostly been performed by integrating ordinary differential equations (ODEs) or stochastic algorithms. When using ODEs one computes continuous concentrations of the participating species. The integration is very fast, but of course it is only suitable when the participating molecule numbers are high enough to be approximated as concentrations. For low particle numbers, stochastic algorithms that compute discrete particle numbers are more accurate, but also computationally expensive. The decision regarding which of these methods to employ to get a realistic result and at the same time to use the fastest possible method for this goal has commonly been made using intuition because there are no reliable and rational rules.

To compensate for some of the computational expenses of the stochastic methodologies, approximate stochastic methods and hybrid methods have been developed recently (see Chap. 2 for details). The hybrid methods need to partition the system into a deterministic and a stochastic subsystem. Again, this is so far mostly done rather heuristically by considering the velocity of reactions or the particle numbers of involved species.

This heuristics is partially justified because there are already a lot of heuristics and simplifications involved when setting up the model itself. One example of this is the inclusion or negligence of spatial dimensions in the model. If space is considered as well, the system can be described by ODEs, partial differential equations, or the respective stochastic algorithm. However, even though space doubtlessly plays a very important role in the functioning of the cell, many models are built assuming homogeneity of the system. This is due to multiple reasons. First of all, even modern experimental technology still prevents the observation of spatially localized concentration changes in the cell for many species. Therefore, spatially resolved experimental data are still rare. Second, many questions concerning, e.g., biochemical mechanisms in small cells like the leukocytes or hepatocytes discussed be-

low can be answered to some extent with the homogeneity assumption (e.g., [OKKP03]) saving computational time. Still, neglecting the spatial dimension of the system is almost always a severe simplification.

Nevertheless, the simplifications and assumptions made while setting up a model are usually thought through and actively done by the scientist who is studying the respective biochemical system. However, the choice of the suitable simulation method is often more passively done because explicit knowledge about when which method is the appropriate one is largely missing. Gonze et al. related the appropriateness of deterministic simulations to the rate constants in a model of the circadian rhythm [GHG04]. However, a generalization of this result for any model is hard to infer.

Therefore, we think that it is of general interest to find a rational basis to actively decide for or against a specific simulation method. This basis should allow the scientist to select the best methodology for his/her specific model with all its assumptions and simplifications.

We therefore studied the transition between stochastic and deterministic behavior primarily in a common model system, namely calcium oscillations, to find a measure that supports this decision process. The findings should not only be applicable for this specific system, and we will show some supporting results on different test systems as well.

Calcium ions act as second messengers in a variety of cell types [BBL98]. They influence cellular functions such as excitability, contraction, metabolism, or exocytosis directly via the modification of enzymatic functions or gene expression [BBL98]. Calcium ions are therefore an integral part of the information-processing machinery in living organisms.

Due to its central importance, the function of calcium as second messenger has been studied intensively, e.g., in hepatocytes. In this cell type, the principal chain of events occurring during calcium signal transduction is rather well known. A detailed description can be found in Chap. 3.1.

The number of receptors and ion channels in the cell can be very low (in the range of 10^3–10^5 per cell), which leads to the question of whether the deterministic approaches used for modeling and simulating this system are valid and to what degree they are valid.

Stochastic simulations of calcium oscillations have been performed in the case of spiking oscillations (e.g., [KW93; PAB$^+$98]). However, in these cases, no detailed comparison to deterministic simulations has been done. In the case of bursting calcium oscillations, no simulations on discrete particle basis of a system displaying deterministic bursting have been reported at all. However, Falcke et al. showed that bursting behavior can arise during the stochastic simulation of spiking [Fal03b; Fal04]. Falcke and others also studied under which conditions a deterministic description of calcium concentrations

based on channel kinetics is appropriate [Fal04; Fal03a; SJ02]. Knowledge like this is necessary to decide which simulation method should be used for a particular system and its particular behavior.

With this in mind, we have studied the stochastic simulation of spiking and bursting calcium oscillations and the transition from stochastic to deterministic behavior. We present experimental data on bursting calcium oscillations that exemplify the need to perform stochastic simulations. For the computational side we used tools developed recently to automatically convert the corresponding differential equations to the stochastic discrete equations and to perform the simulations. We observed a transition at particle numbers in the range of actual particle numbers in the cell.

The transition became apparent, when we compared the results obtained by stochastic, discrete simulations according to Gillespie [Gil76] and the numerical integration of ODEs. For high particle numbers the resulting simulations were basically the same. However, gradually lowering the number of particles, some significant differences between both computational approaches emerged. Therefore, we defined a transition range as the approximate number of particles at which significant differences between stochastic and deterministic simulation start to occur (the solutions do not match anymore). Minute fluctuations of the trajectory are not considered. It is of special interest to analyze whether this transition range depends on the complexity of calcium oscillations. Our results show that the transition range indeed changes with changing dynamics of the system. Thus, the transition range cannot be generally determined for a system being valid for all parameter values, but is dependent on the individual dynamics of a certain parameter set. However, it is not the degree of complexity (e.g., complex periodic versus simple periodic behavior) that determines the transition range. Our results show that it is rather the attractive property of the respective phase space that plays a more important role than the complexity of Ca^{2+} oscillations. The attractive properties of the phase space have been quantified by the sum of Lyapunov exponents (the divergence). Our results indicate that at lower divergence the transition from stochastic to deterministic behavior occurs at lower particle numbers, which means that the system is well characterized by ODEs at realistic particle numbers. At higher divergence values the transition occurs at significantly higher particle numbers, which indicates the need to employ stochastic modeling. These findings are in accordance with the experimental observation that apparently stochastic behavior is more common in bursting calcium oscillations during high agonist doses, which corresponds to a high divergence value in the corresponding model. The results were also verified with other models and should apply for many types of biochemical models.

5.2 Materials and methods

5.2.1 Computations

Deterministic simulations were performed by numerically integrating ODEs with the Rosenbrock [Ros63] and LSODE [Pet83] algorithms.

For the stochastic simulations, we used the stochastic algorithm developed by Gillespie [Gil76] (Direct Method), which is described in detail in 2.4.1.

On the basis of this algorithm, software was implemented, which is able to automatically convert a system of differential equations into the corresponding stochastic system and to perform the stochastic simulations (e.g., STODE, which is freely available at [Stoe] or COPASI [Cop].

5.2.2 Calculation of Lyapunov exponents and the divergence

Lyapunov exponents are an important concept for quantifying the sensitivity of dynamic systems against perturbations of initial states (see [KS97], for details). They describe the exponential divergence or convergence of nearby trajectories on an attractor and positive exponents can indicate chaotic behavior. A number of algorithms have been proposed for the calculation of these exponents, e.g. Rosenstein 1993 [RCD93], Wolf 1985 [WSSV85] and Kantz 1997 [KS97]. The algorithm by Wolf has been implemented into COPASI [Cop; HSG+06]. Implementations of the algorithms by Rosenstein and Kantz can be found, for instance, in TISEAN [HKS99].

The sum of all Lyapunov exponents equals the so-called average divergence of the system. This divergence, on the other hand, is identical to the trace of the Jacobian matrix. By average we mean an average over a specific trajectory that is long enough to be regarded as representative of a certain dynamic regime of the system. Also, usually, a transient has been cut off.

The computation of the divergence value alone is much easier than the calculation of the complete spectrum of Lyapunov exponents [WSSV85]. One can numerically integrate the trace of the Jacobian along with the system's state variables. Finally, this integral is divided by the simulation time to yield the average value. This method was employed in this study and it is also implemented in COPASI.

5.2.3 Experimental

The experimental setup used for measuring calcium concentrations in single rat hepatocytes over time is described in detail in Sec. 3.4.

5.3 Results

Using ATP as the agonist for the activation of hepatocytes results in bursting calcium oscillations for a wide range of concentrations. An example for a low dose (1.2 μM) of ATP is shown in Fig. 5.1. Each main spike is followed by a series of secondary oscillations. The overall oscillation is by no means periodic and the number of secondary oscillations varies. Increasing the agonist concentration successively results in bursting oscillations, with increasing amounts and length of secondary oscillations on average (Fig. 5.2). These bursts are irregular in their nature, meaning that the amplitude of the secondary oscillations are not simply decreasing with time.

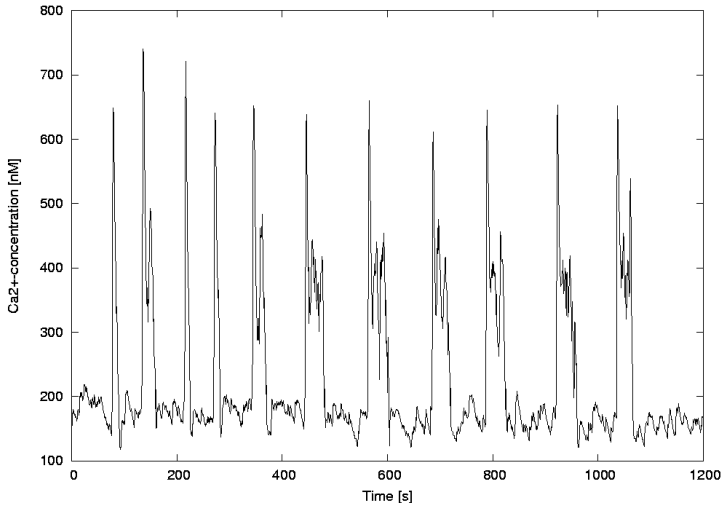


Figure 5.1: Experimentally measured calcium concentration in hepatocytes with 1.2 μM ATP added.

Modeling these bursting oscillations in hepatocytes has so far never been able to account for the long stretches of secondary oscillations seen in these time series and only models generating chaotic bursting have been able to account for some of the nonperiodicities visible. Prolonged secondary oscillations might carry important information for the cell because, e.g., a very prolonged elevated level of calcium concentration in the cell can be responsible for apoptosis [BBL98].

For a detailed computational analysis of calcium oscillations in hepatocytes, we restrict ourselves firsthand to using a core model developed by Kummer et al. [KOD$^+$00] (see Chap. 3.2 for a complete description). This model captures the basic dynamic characteristics of the complete model.

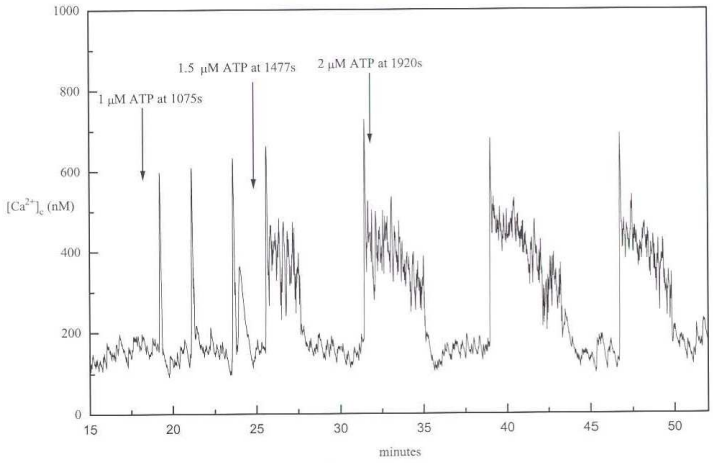


Figure 5.2: Experimentally measured calcium concentration in hepatocytes with increasing amounts of ATP added as indicated.

Later on, we will see that our findings also hold true for more detailed, physiological models.

Again, it has to be emphasized that this model is a simplified picture and does not include all the processes that are known to occur in the context of calcium signal transduction. Especially, one variable, namely IP_3 has been eliminated completely. For a more detailed and more realistic model, see, e.g., Larsen et al. [LOK04]. However, the basic dynamical characteristics are captured in this model (as was shown in Kummer et al. [KOD$^+$00]) and therefore we use it to study the transition from stochastic to deterministic behavior in dependence on the system dynamics.

The bifurcation diagram of the model is shown in Fig. 5.3. At smaller values of k_2 the system behavior is characterized by simple periodic spiking Ca^{2+} oscillations. By increasing the value of k_2 periodic bursting Ca^{2+} oscillations appear, and a period adding route leads to a very small chaotic regime around $k_2 = 2.9259$. Beyond the chaotic regime there is again a small periodic regime before the system settles into a steady state.

In Fig. 5.4 we show an example of deterministically simulated periodic bursting oscillations for $k_2 = 2.85$.

We simulated the same time series as obtained by the ODEs (e.g., Fig. 5.4) on particle basis with the stochastic algorithm described above. We varied the number of participating particles to study the transition from deterministic to stochastic behavior with decreasing particle numbers. In the following, we will emphasize the number of calcium ions in the system. However, we want to point out that the number of particles of the other participating

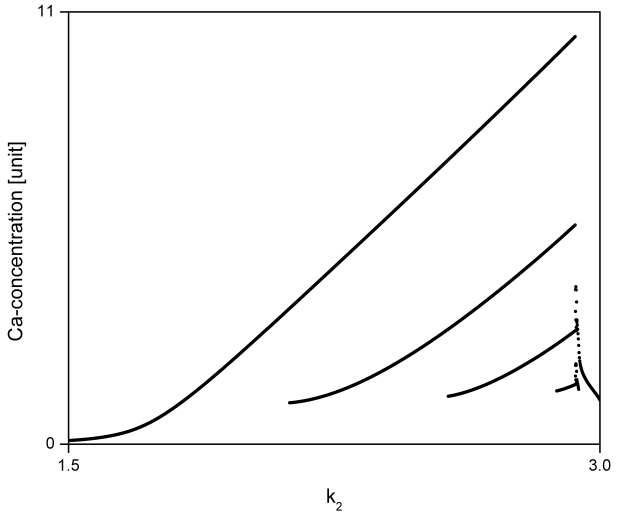


Figure 5.3: Bifurcation diagram of the core model for bursting calcium oscillations (Table 3.1). Parameters are: $k_1 = 0.212, k_3 = 1.52, K_4 = 0.19, k_5 = 4.88, K_6 = 1.18, k_7 = 1.24, k_8 = 32.24, K_9 = 29.09, k_{10} = 13.58, k_{11} = 153, K_{12} = 0.16$. Initial conditions are: $[\mathrm{G}_\alpha] = 0.01, \mathrm{PLC} = 0.01, \mathrm{Ca}^{2+} = 0.01$.

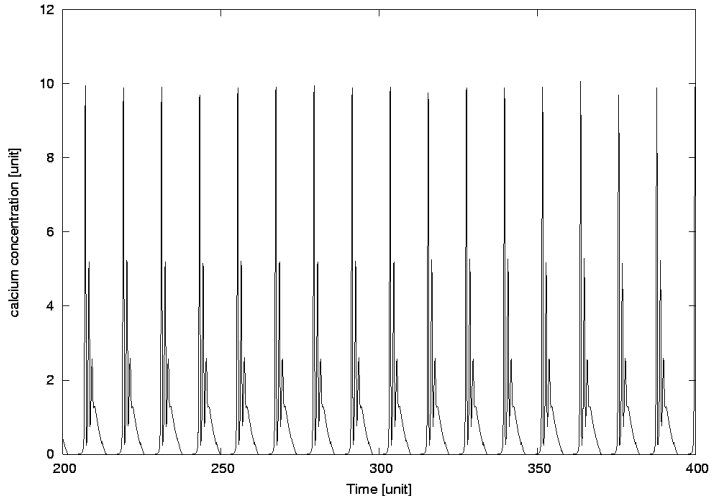


Figure 5.4: Deterministic simulation of periodic bursting of calcium concentration. Parameters as in Fig. 5.1; $k_2 = 2.85$, divergence = -401.9.

species is in the same range or higher (depending on the parameters) in this simple model system. Therefore, we focus on the species with the lowest particle numbers. This was achieved by changing the volume of the system and leaving the concentration constant. Computationally, this is equivalent to letting the volume constant and changing the particle number in this volume plus adjusting the kinetic parameters such that the same qualitative systems behavior will arise. Otherwise, just changing the particle number in the same constant volume will of course result in completely different behavior.

The results of the stochastic simulations for $k_2 = 2.85$ are presented in Figs. 5.5 and 5.6 for different particle numbers.

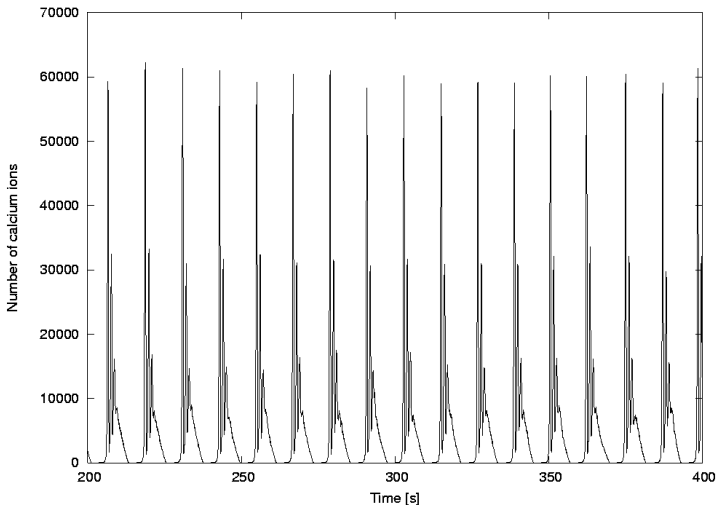


Figure 5.5: Stochastic simulation of bursting calcium oscillations close to the deterministic limit. Parameters as in Fig. 5.4.

Fig. 5.5 shows that for large particle numbers the particle-based simulations approach the deterministic limit, which is in accordance with theory. The question arises, however, of how to determine the transition between stochastic and deterministic behavior. This is usually a continuous convergence and it is difficult to exactly determine the transition. Therefore, it is reasonable to introduce a transition range as the approximate number of particles at which differences between stochastic and deterministic behavior become negligible. To estimate the particle number in the transition range between stochastic and deterministic behavior, we studied the use of standard approaches for estimating differences between noisy signals and the respective deterministic signal. All of these standard approaches like autocorrelation functions, signal/noise ratio, or interspike interval histograms (ISIH) face

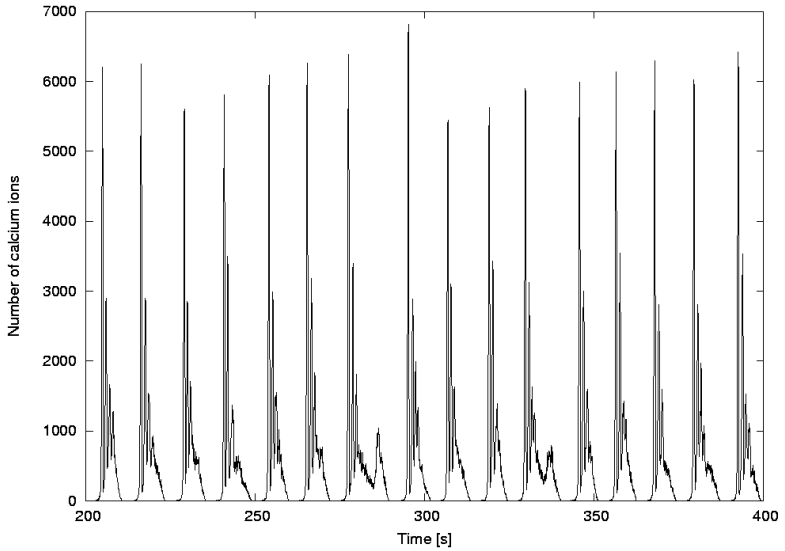


Figure 5.6: Stochastic simulation of bursting calcium oscillations with lower particle numbers compared to Fig. 5.5. Parameters as in Fig. 5.4.

strong limitations in this case. The reason is that in many cases the stochastic simulation does not result simply in a noisy version of the deterministic limit (as shown below). It is rather apparent that the stochasticity of the system often results in completely different dynamics compared to the deterministic solution. However, the global character of the solution (the attractor) is underrepresented when considering the above-mentioned standard approaches. Thus, on one hand, a noisy limit cycle will result, e.g., in a very different ISIH compared to the deterministic solution even if the coarse limit cycle is the same. On the other hand, a comparison between different attractors will of course also result in a very different ISIH. Thus, it is almost impossible to differentiate between a solution that displays a completely different attractor and a solution that still displays the global attractive properties of the deterministic solution, but has added noise. However, a scientist modeling a system will most certainly choose the faster deterministic simulation, if the global picture of the simulation is accurate. A better method would be to use a similarity measure of the global attractors resulting from the different simulations as such. Few approaches for such a similarity measure are described in literature so far (e.g., [Kan94; Kou01]). These need extensive sets of data that are hard to create in stochastic simulations due to the computational expense. Therefore, for this study, we restrict ourselves to matching the solutions in a graphical and/or visual way. However, we want to include and develop such global similarity measures in future studies.

For spiking ($k_2 = 2.0$) and periodic bursting oscillations ($k_2 = 2.85$) with ten-thousands of particles, only small fluctuations in the amplitude are observed. However, decreasing the particle number to thousands leads to a system already showing significant stochastic influence (Fig. 5.6). Stochastic influences are big variations in the amplitude and period as well as prolonged secondary oscillations during bursting behavior like those seen in the experimental investigations. The transition from deterministic to stochastic behavior occurs in this case in the range of tens of thousands of particles.

For chaotic bursting Ca^{2+} oscillations at $k_2 = 2.9259$ deterministic-like behavior was observed only down to a number of particles in the range of hundreds of thousands. Decreasing the particle numbers down to tens of thousands already showed significant stochastic influences, e.g., a phase space that corresponds more to a noisy limit cycle rather than to a chaotic attractor, i.e., hardly any amplitude variations. Decreasing the particle numbers even further leads to additional prolonged secondary oscillations. Of course, there is no possibility to simply match the deterministic and the stochastic simulation in this case like done above. Therefore, and to get an estimate, we relied on visual inspection taking, e.g., prolonged secondary oscillations as signs for stochasticity. These signs ceased to appear in the range of hundreds of thousands of particles in the system in the parameter regime where chaos is displayed in the deterministic limit.

For the steady state at $k_2 = 3.0$, we observe that even higher numbers of particles are needed to approach the deterministic limit (Figs. 5.7 and 5.8). The deterministic limit is not reached with particle numbers in the high hundred-thousands, which is well above the physiological range. Moreover, for lower particle numbers, qualitative behavior is observed that again displays most of the characteristics of the complex periodic regime (Fig. 5.8).

The above findings are summarized in Table 5.1. For different values of k_2, which correspond to different behaviors of the system, the number of particles is estimated at which the transition between stochastic and deterministic behavior appears.

k_2	Number of particles	Behavior
2.0	Ten-thousands	Periodic spiking
2.85	Ten-thousands	Periodic bursting
2.9259	Hundred-thousands	Chaos
2.99	Millions	Regular oscillations
3.0	Greater than millions	Steady state

Table 5.1: Transition ranges in dependence on k_2.

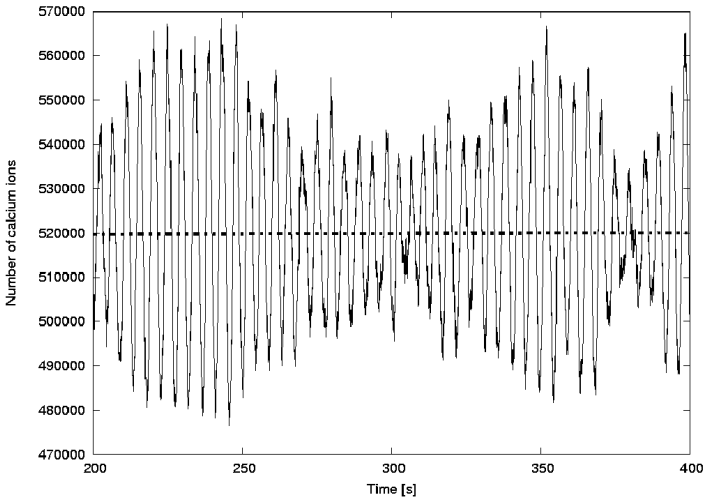


Figure 5.7: Stochastic simulation of calcium behavior corresponding to parameters for which the deterministic solution is a steady state ($k_2 = 3.0$) with particle numbers far above physiological concentrations. The dashed line indicates the deterministic steady state.

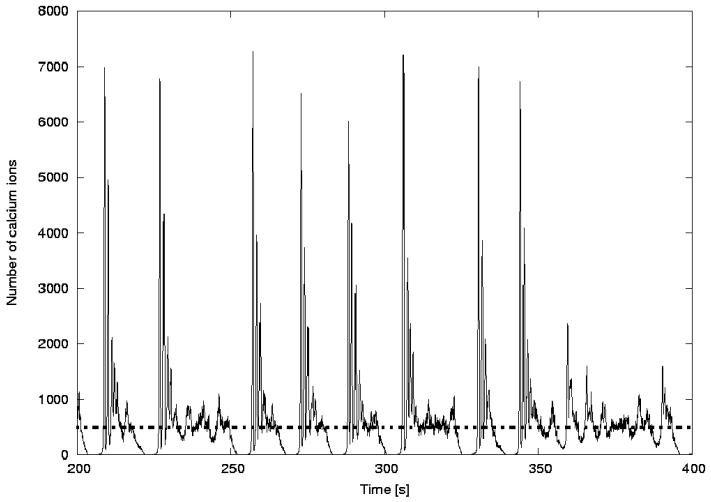


Figure 5.8: Stochastic simulation of calcium behavior corresponding to parameters for which the deterministic solution is a steady state ($k_2 = 3.0$) with lower particle numbers compared to Fig. 5.7. The dashed line indicates the deterministic steady state.

To explain the results presented in Table 5.1 we estimate the attractive properties of the phase space for different values of k_2. The hypothesis is that the stochastic influences could be more pronounced in case of weaker attractive properties of the phase space. Hence, in a weaker attractive phase space a higher number of particles would be needed to reach the deterministic limit. We use the sum of Lyapunov exponents (the divergence) for estimating the attractive properties of the phase space. By varying the parameter value of k_2 and corresponding to the different types of oscillations described above, different values of divergence (Fig. 5.9) were computed. Fig. 5.9 shows that the value of divergence approaches zero with increasing values of k_2. By comparing Fig. 5.9 with Table 5.1, we observe that the sensitivity of the system to stochastic influences increases with increasing divergence. This means that for a dynamic state representative of a highly negative divergence value the system is well described by deterministic methods even for relatively low (thousands) particle numbers whereas higher particle numbers are needed to approach the deterministic limit if the divergence of the system is close to zero.

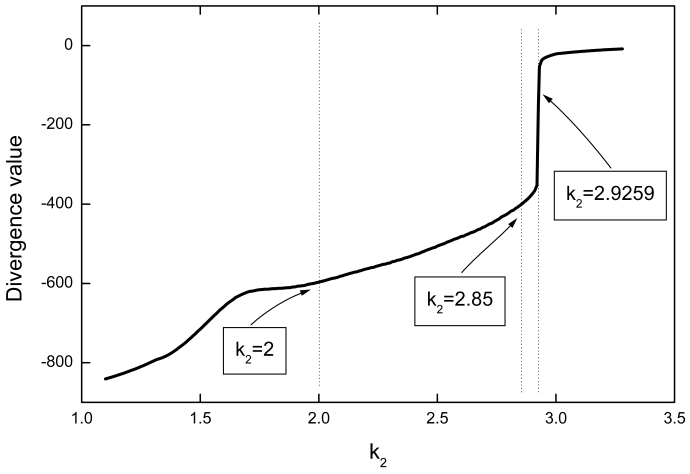


Figure 5.9: Divergence value of the core model for bursting calcium oscillations. Parameters as in Fig. 5.3.

To verify that these findings are not restricted to our small core model, we additionally analyzed a completely different model of Ca^{2+} oscillations proposed by Shen and Larter [SL95]. Likewise, a model of the peroxidase-oxidase reaction [OLK03] was studied (see Fig. 5.10).

This system describes the oxidation of NADH catalyzed by peroxidases.

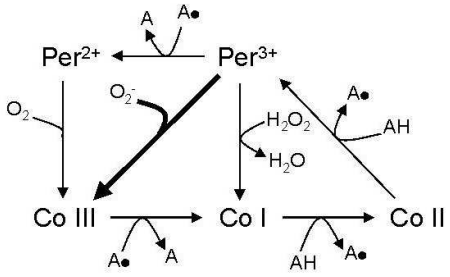


Figure 5.10: Schematic view of the peroxidase-oxidase reaction model described in [OLK03].

We observed again that the number of particles required to obtain results matching the corresponding deterministic solutions is directly related to the divergence of the attractors, as already described above. A divergence value close to zero implies that the attractor is weak and can easily be altered by the stochastic fluctuations. In Fig. 5.11 two different dynamical regimes with low and high divergence value are shown as examples. The corresponding time series show a transition between stochastic and deterministic behavior at different particle numbers.

We also tested a model for mitogen-activated protein kinase cascades [Kho00]. One of the important parameters of that model is v_1, the stimulus strength. We scanned the divergence of that model in dependence of v_1. The results are shown in Fig. 5.12. Here, the divergence decreases with increasing parameter value for $v_1 \in [0.5, 3]$. For the default value ($v_1 = 2.5$) we observed the transition roughly between 100 and 1000 particles of MAPK-PP. For $v_1 = 1$ we expected the system to be more sensitive to random fluctuations due to the higher divergence. Indeed here, the transition takes place approximately between 1000 and 10000 particles.

Finally, we studied the influence of calcium buffers in the cell. For this purpose, we included a simple linear equation for the binding and release of calcium to protein buffers with the latter being present in large quantities compared to calcium and therefore assumed to have a constant concentration. Thus, the equations for calcium concentration reads:

$$[\mathrm{Ca}^{2+}]' = k_{10}[\mathrm{G}_\alpha] - k_{11}\frac{[\mathrm{Ca}^{2+}]}{([\mathrm{Ca}^{2+}] + K_{12})} - k_{13}[\mathrm{Ca}^{2+}] + k_{14}[\mathrm{P}], \qquad (5.1)$$

with P representing the calcium concentration bound to protein buffers.

The inclusion of this simple term leads to a decrease in divergence, because the partial derivation of the equation describing the evolution of the

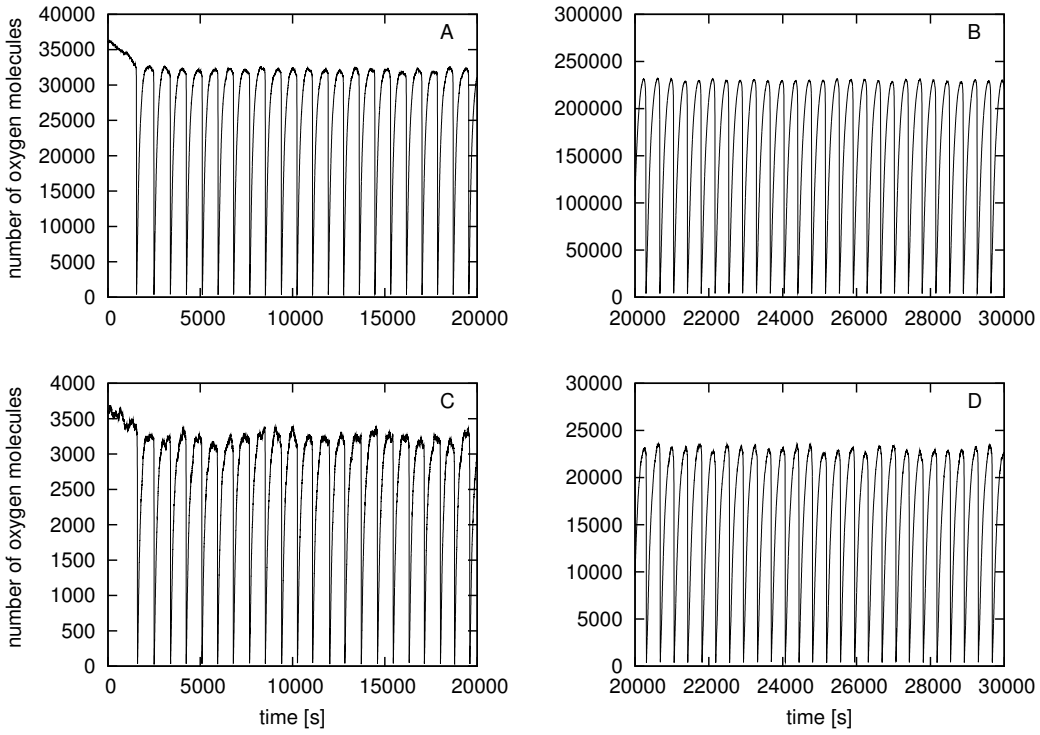


Figure 5.11: Transition from stochastic to deterministic behavior in the peroxidase-oxidase system [OLK03]. Time series of the numbers of oxygen molecules. Panels A and C: low divergence (bifurcation par. $k_{12} = 0.04$) $\rightarrow$ transition roughly between thousands and ten-thousands of particles. Panels B and D: high divergence (bifurcation par. $k_{12} = 0.08$) $\rightarrow$ transition roughly between ten-thousands and hundred-thousands particles.

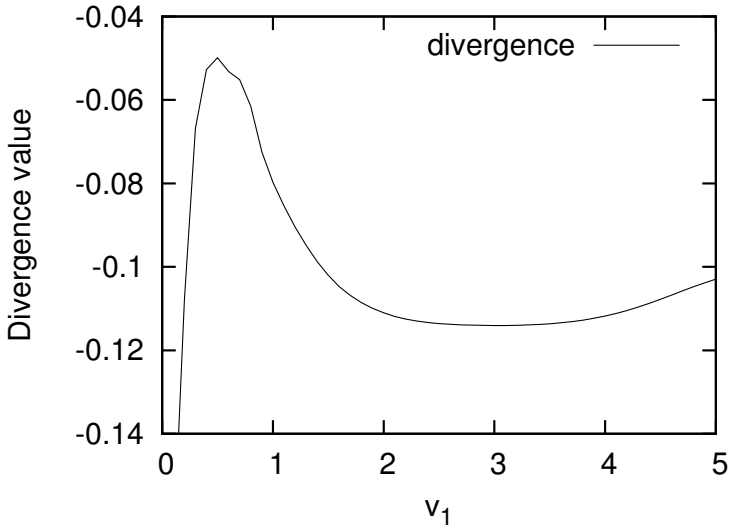


Figure 5.12: Scan of the divergence (y-axis) of the MAP kinase cascade model in [Kho00] in dependence of the stimulus strength v_1 (x-axis).

calcium concentration becomes more negative and this feeds into the sum of the Lyapunov exponents. This means that according to our hypothesis the sensitivity toward stochasticity should decrease as well. Indeed, the high frequency stochastic fluctuations are diminished by the presence of the buffer. However, we have found that buffering can change the system dynamics in such a way that the divergence is not always uniformly decreased globally but can in fact even rise with increasing buffering speed (see Fig. 5.13). This results in different behavior in stochastic and deterministic simulations (e.g. bursting oscillations and steady state, respectively) even if the particle numbers are large and the system is buffered.

Most of the calcium in cells is bound to protein buffers. But, even with $\sim 80\,\%$ of the calcium being bound to these buffers, the systems behavior is still strongly influenced whenever the divergence of the system is large (e.g., for $k_2 = 3$). Fig. 5.14 shows that in this case the high-frequent part of the noise in the system is filtered out by the participating buffer (compared to Fig. 5.6). Nevertheless, the system is still not running into a steady state as it would be when calculated deterministically, but rather it shows complex bursting oscillations.

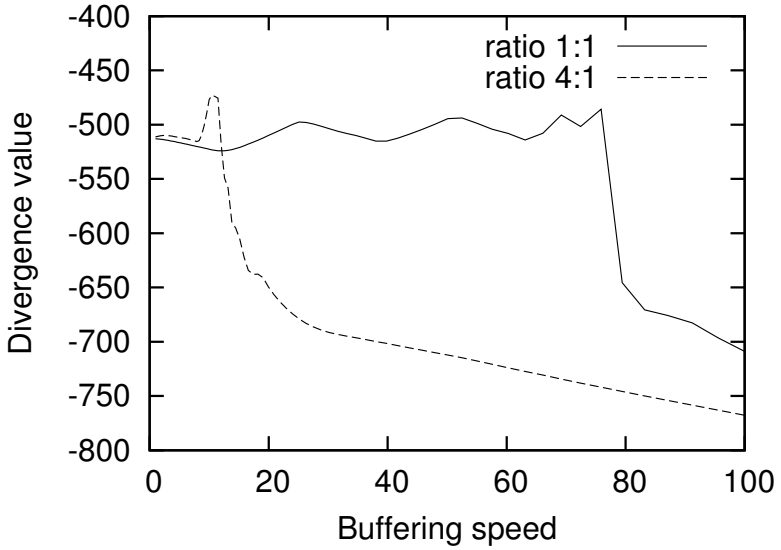


Figure 5.13: Calculation of the divergence value (y-axis) for different protein buffering speeds (x-axis). The buffering speed is the sum of the calcium binding and the calcium release rates. The ratio of binding rate to release rate is 1:1 or 4:1, leading to about 50 % or 80 % of the total calcium bound to the buffer (k_2 equals 2.5).

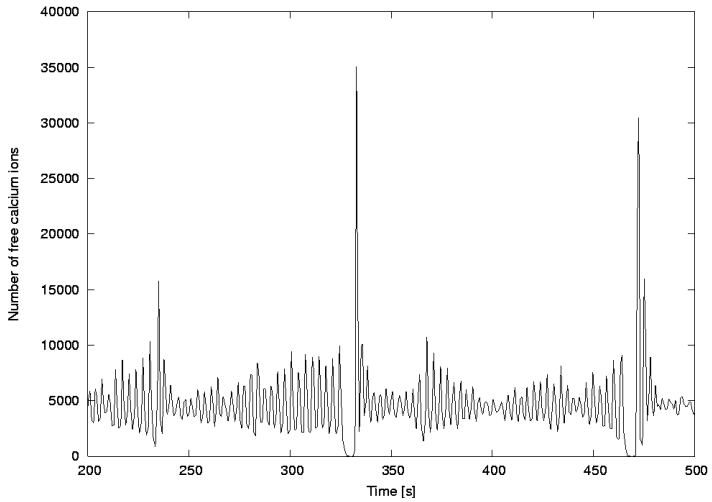


Figure 5.14: Stochastic simulation of calcium behavior corresponding to the core model with parameters as in Fig. 5.3, including binding of calcium ions to protein buffers (Eq. (5.1); $k_{13} = 10, k_{14} = 1$). Please note that the total calcium ion concentration is by far higher because $\sim 80\,\%$ are bound to protein buffers in this case.

5.4 Discussion

We studied the transition from deterministic to stochastic behavior in simulations of Ca^{2+} oscillations on a particle basis. We mainly used the model developed by Kummer et al. [KOD$^+$00] (see Chap. 3 for details) but also other systems, such as the peroxidase-oxidase reaction and MAP kinase cascades. Here, we studied in detail the dependency of the transition on the system properties. We observed that the transition from stochastic to deterministic behavior depends heavily on the attractive properties of the corresponding attractors in phase space, quantified by the divergence. We conclude that the divergence plays a more important role in determining the transition range from stochastic to deterministic behavior than the complexity of the Ca^{2+} oscillations. The transition occurs at higher particle numbers if the corresponding value of the divergence is close to zero compared to the particle numbers needed when the system has a highly negative divergence. Comparing the ranges of particle numbers sensitive to stochastic influences, we observe that oscillations characterized by a divergence close to zero show a 10-100-fold larger sensitivity compared to the oscillations with highly negative divergence in the presented case study.

The real particle numbers in calcium signal transduction roughly correspond to the transition number in the cases with low divergence, namely simple periodic and complex periodic oscillations. This is especially true for a model in which calcium buffers are included. However, the number of particles needed to reach the deterministic limit in cases with high divergence values (chaotic, regular oscillations, steady state) is far above the concentrations of receptors, channels, and calcium ions in the real cell, even if buffers are included in the model. This is also in accordance with the experimental observation that at high agonist concentrations that correspond to high divergence values in our model, more apparent stochastic influences are visible. Therefore, one can argue on the one hand, that the stochastic influence during simple periodic and complex periodic behavior should not be tremendous, because the real particle numbers are not well below the transition range. On the other hand, pronounced stochastic effects should be present in the real system for high agonist concentrations (corresponding to a high value of k_2). However, because the studied model is rather qualitative in its nature, more studies with more realistic models are needed to clarify this point in sufficient detail. The important issue here is that the transition from stochastic to deterministic behavior for certain systems dynamics in general occurs clearly above physiological concentrations and the resulting stochasticity in the system might be of physiological importance. This is especially true for physiological effects that result from the prolonged

secondary oscillations of the bursting calcium oscillations as described above. Interestingly, such prolonged secondary oscillations have often been observed experimentally (e.g., [GCD95]). If the elevated level of calcium concentration is sustained for considerable time, it will result in different biochemical responses in the cell, e.g., in cell death [NBO92].

Our findings, showing that transition from stochastic to deterministic behavior occurs at higher particle numbers if the corresponding value of the divergence is close to zero, can also be explained intuitively. If the contractive properties of an attractor in phase space are weak, then the attractor can be more easily deformed, if perturbed continuously, which is the case when studying stochastic simulations. Recently, it has been shown that Ca^{2+} oscillations are more flexible in response to external forcing if the divergence takes values close to zero [MS03; PM03b]. Moreover, it has been shown that the flexibility of Ca^{2+} oscillations does not significantly depend on the type of Ca^{2+} oscillations. Therefore, we argue that in the case of determining the transition from stochastic to deterministic behavior the divergence plays a major role.

In the studied systems, no noise-induced chaos has been found as reported in a number of cases (for a review, see Gao et al. [GCHL99]). However, it has been observed in earlier studies that adding noise to a periodic bursting calcium oscillation could result in deterministic chaotic oscillations [KBO00].

Our results show that it is not sufficient to decide in favor of or against the stochastic simulation of a system on the basis of knowing the number of particles for a certain model in general, but it rather demands taking into account the specific dynamics of the model and the attractive properties of a particular oscillatory regime. Moreover, relatively large concentrations (corresponding to nanomolar and millimolar), which often are simulated deterministically, already show a pronounced sensitivity toward stochasticity. Therefore, a careful analysis of this sensitivity should preclude a decision for a certain simulation method in the case of simulating calcium oscillations. Because our results are very general in their nature, this holds for other simulations of biochemical systems as well. Calculating the divergence of the system as one measure for the decision in favor of or against a specific simulation methodology could be easily automated. The software system COPASI, for instance, contains this feature to aid the user in his/her decision process. Moreover, calculating the divergence on the basis of deterministic simulations is computationally fast compared to many trials of stochastic simulations that would be needed to just try out which method is more appropriate. It is also possible to compute the sensitivity of the divergence with respect to different parameters of the system, which gives a more general view on how robust the decision for or against a specific simulation method is when parameters are

changed. However, we also would like to point out, that the absolute values of the divergence might be insufficient as a basis for the decision process, if a system contains, e.g., very positive and very negative Lyapunov exponents at the same time (which was not the case in the studied examples). In this case, a weighting of these individual components might be necessary, which is a topic of our future research. In addition, bistable systems require also a special treatment. Such systems will display both stable solutions when different runs of stochastic simulations are performed whereas, e.g., the initial conditions have to be changed in the deterministic approach to gain the same kind of information. However, the appearance of the individual solution is again subject to similar criteria as described above. Moreover, in the case of a stable steady-state solution with no proximity to any other type of solution (e.g., oscillations), there are cases where the amplitude of the noise due to a stochastic simulation around this steady state stays the same, independent of the divergence of the system (as in the simple system $A \rightleftharpoons B$ with influx of A and efflux of B, if all rates are altered such that their ratio stays the same). However, due to the attractive properties of the respective steady state, which is again dependent on the divergence, the individual trajectory is able to stay away from the steady state much longer if the divergence is high compared to a system with rates corresponding to a low divergence. Thus, if simulating a short time span representing a real world example, the stochastic simulation of the system with low divergence will quickly fluctuate around the steady state whereas the system with high divergence might deviate from the steady state for the whole time. Therefore, the computation of the divergence again adds to the knowledge in differentiating between the two simulation methods.

Finally, we want to emphasize that inclusion of a spatial dimension will be an important issue in the future. Particle numbers in small discrete volumes will be even lower than considering the particle numbers for the whole cell. We think that at least for systems described by diffusively coupled ODEs, our findings will still be applicable, but this will be a matter of future investigations.

Acknowledgments

We acknowledge Carel van Gend for fruitful discussions. Ursula Kummer and Jürgen Pahle thank the Klaus Tschira Foundation and the German Federal Ministry of Education and Research for funding. Anne K. Green and C. Jane Dixon thank the Wellcome Trust for funding.

Chapter 6

Information transfer in signaling pathways: a study using coupled simulated and experimental data

> Statistics are like bikinis.
> What they reveal is suggestive,
> but what they conceal is vital.
>
> Aaron Levenstein

Signaling pathways are particularly prone to random fluctuations. Thus, an important question is how this influences the information transfer in these pathways.

The stochastic interpretation of biochemical networks lends itself readily to ideas coming from information theory. Information theory deals with the uncertainty of events in a statistical way. The decrease in uncertainty is equated with the information gained by an observation. In this chapter we will show how the information-theoretic measure transfer entropy can be applied to signaling pathways, namely calcium signaling in order to quantify the information transferred from calcium to target enzymes under different cellular conditions. This study was done in collaboration with Anne K. Green, C. Jane Dixon and Ursula Kummer and part of the material contained in this chapter has been published in BMC Bioinformatics [PGDK08].

6.1 Information theory

Even though information-theoretic ideas can be traced back at least to the twenties of the past century (Harry Nyquist, Ralph Hartley), the establishment of information theory as a discipline is most often attributed to Claude Elwood Shannon (1916-2001). His seminal work "A Mathematical Theory of Communication" from 1948 [Sha48] deals with the problem of reconstructing a message, which has been disturbed during transmission over a noisy communication channel: "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point" [Sha48, p. 379]. This idea, initially only applied to engineering problems associated with data compression and communication, has, due to its general nature, also proven very fruitful in other fields such as linguistics and (neuro-)biology. Other important contributions to the development of information theory were made by, e.g., Norbert Wiener, David A. Huffman, Solomon Kullback, Richard Leibler and many others.

First, we will give a brief introduction into some basic concepts of information theory, which are relevant for our study. A recommendable reference book for further study is "Elements of information theory" by T.M. Cover and J.A. Thomas [CT91] (see also [Ash90; Mac03]).

The information content of an event A (in an appropriately defined probability space) can be quantified by its degree of uncertainty expressed as the negative logarithm of its probability ($-\log Pr\{A\}$). This is a strictly decreasing function on $(0, 1)$ with values from ∞ to 0 and it formalizes the intuitive notion that rare events are more interesting than frequent ones. A certain event ($Pr\{A\} = 1.0$) has no information content. For notational convenience, the information content of an impossible event B with $Pr\{B\} = 0$ is usually defined to be zero.

Now, the so-called information entropy $H(X)$, or Shannon entropy, is a functional of the probability mass/density function p of a discrete/continuous[1] random variable X (with values x_i):

$$H(X) = -\sum_i p(x_i) \log_b p(x_i) \tag{6.1}$$

[1] Shannon entropy and the other information-theoretic quantities can also be defined for continuous random variables. Because of the limited space we will only show the discrete versions here.

where the base b of the logarithm can be chosen at will and only determines the units of information entropy. In the following, we will use $\log_2$ for binary digits [bits]. One bit is the amount of information needed for answering one yes/no question with equally probable answers.

The Shannon entropy can be interpreted as the averaged information content over the set of possible outcomes or symbols. It also gives the number of bits needed to optimally encode independent draws of the random variable. One important property is that the information entropy is highest ($= \log N$, with N the number of different outcomes) when all outcomes are equally probable. Intuitively, the uncertainty is maximal here because the outcome cannot be predicted at all in that case.

Shannon entropy should not be confused with information. Information is actually associated to a (potential) decrease in entropy due to an observation. In other words, information measures the sharpening of our knowledge (in terms of the difference in uncertainty about a system before and after an observation). Assuming there is no measuring error, we are certain about the outcome after a measurement. In that case the uncertainty drops to zero as a result of the observation and the information gained equals the system's Shannon entropy before the measurement.

There is also a conditional version of the information entropy which will become important for our study of the dependence of two stochastic processes:

$$H(X|Y) = -\sum_{i,j} p(x_i, y_j) \log p(x_i|y_j). \tag{6.2}$$

It quantifies the remaining uncertainty about the random variable X when Y is known already.

If we do not know the true probability distribution p of a random variable, and thus its true entropy, we can just assume it to have the probability distribution q. The exceeding entropy, or, equally, the excess number of bits we would need if we used q for encoding, is given by the so-called Kullback-Leibler divergence $D_{KL}(p\|q)$ [KL51]:

$$D_{KL}(p\|q) = \sum_i p(x_i) \log \frac{p(x_i)}{q(x_i)}. \tag{6.3}$$

Due to its asymmetry the Kullback-Leibler divergence is not a true metric. However, it is an important concept for comparing probability distributions. The well-known mutual information $I(X,Y)$ often used to quantify the dependence between two random variables is based on the Kullback-Leibler divergence:

$$I(X,Y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \tag{6.4}$$

Here, it is assumed that the two random variables X and Y are independent, i.e. that their joint probability distribution $p(x_i, y_j)$ equals the product of the two marginal probability distributions $p(x_i) \cdot p(y_j)$. The mutual information quantifies the distance between this assumed joint probability distribution and the true joint distribution.

In other words, it gives the amount of information shared by the two random variables. It is zero if the two random variables are statistically independent and maximal ($= H(X)$) if one random variable can be calculated from the other one. It can easily be seen that the mutual information is symmetric $I(X,Y) = I(Y,X)$.

The following summarizes some important relationships between entropies and the mutual information:

$$H(Y|X) = H(X,Y) - H(X) \tag{6.5}$$

with $H(X,Y) = -\sum_{i,j} p(x_i, y_j) \log p(x_i, y_j)$

$$\begin{aligned} I(X,Y) = I(Y,X) &= H(X) - H(X|Y) && (6.6)\\ &= H(X) - (H(X,Y) - H(Y)) && (6.7)\\ &= H(X) + H(Y) - H(X,Y). && (6.8) \end{aligned}$$

In the case of stochastic processes I and J we have one random variable for each process and each point in time. Random variables of different time points are generally not independent. This dynamical information is usually neglected [Sch00] when the mutual information is employed. Instead, it is assumed that the processes are stationary and that their probability distributions do not change over time.

6.1.1 Transfer entropy

In biochemical systems the dynamics plays an important role. The reaction network is identified with a jump Markov process (see Sec. 2.4) and future

states are, of course, dependent on the current state. Therefore, in the following we will use a quantity which takes this dynamical information into account:

The so-called transfer entropy $T_{J\to I}$ is an information-theoretic [WS49] measure proposed by Schreiber 2000 [Sch00] to quantify the dependence of one stochastic process on a second one. Its definition for discrete systems I and J reads as follows:

$$T_{J\to I} = \sum_n p(i_{n+1}, i_n^{(k)}, j_n^{(l)}) \log \frac{p(i_{n+1}|i_n^{(k)}, j_n^{(l)})}{p(i_{n+1}|i_n^{(k)})} \tag{6.9}$$

with $i_n(j_n)$ the system state of process $I(J)$ at time point n and $i_n^{(k)} = (i_n, \ldots, i_{n-k-1})$.

The transfer entropy has Kullback-Leibler divergence form and measures the deviation of process I from its Markov process behavior of order k due to the interaction with process J. The degree of deviation is identified with the information transferred. The most notable properties of the transfer entropy are that it is directional (asymmetric) and dynamic, meaning that the history of the process is taken into account.

6.2 Introduction

The topology of signaling cascades has been studied in quite some detail. However, how information is processed exactly is still relatively unknown. Since quite diverse information has to be transported by one and the same signaling cascade (e.g. in case of different agonists), it is clear that the underlying mechanism is more complex than a simple binary switch which relies on the mere presence or absence of a particular species. Therefore, finding means to analyze the information transferred will help in deciphering how information is processed exactly in the cell.

The simulation of complex biochemical networks has become very important to gain insight into the dynamic behavior of cellular processes [EB01; ÉT89; HS96]. Signaling pathways, in particular, often evade intuitive, and therefore rather static, explanations because of their highly nonlinear dynamics and many cross-links. However, despite the emergence of sophisticated high-throughput and *in vivo* imaging techniques, there is still a lack of high-quality single-cell multivariate data.

Such data would be very helpful in elucidating the nuts and bolts of many signaling mechanisms. In this study we use calcium signaling as an example which represents one of the most versatile second-messenger pathways (see Chap. 3 for details). Since a range of different agonists such as hormones (e.g. vasopressin) or nucleotides (e.g. ATP) trigger calcium responses and, on the other hand, a range of different targets (e.g. Ca^{2+} dependent proteins such as calmodulin, CaM kinase II, protein kinase C, phosphorylase kinase or transcription factors e.g. NF-AT or NF-κB) exist in the cell [Cel96], specific information is likely to be encoded in the calcium signal and decoded again later on. It has been proposed that information might be encoded in the amplitude, frequency, duration, waveform or timing of calcium oscillations and the search for this *calcium code* has attracted a number of experimental and theoretical studies (for a review, see [LK03]).

On the experimental side, mainly the frequency decoding of spiking calcium oscillations has been examined. De Koninck and Schulman 1998 [DS98] demonstrated the sensitivity of immobilized CaM kinase II to Ca^{2+} oscillation frequency by *in vitro* rapid superfusion. Li 1998 et al. [LLW$^+$98] found that NF-AT is activated optimally at a Ca^{2+} oscillation frequency of about 1/min and Dolmetsch et al. 1998 [DXL98] studied the differential regulation of T-cell NF-AT and NF-κB by Ca^{2+} oscillations of different frequencies. The interesting work of Oancea and Meyer 1998 [OM98] describes the activation of protein kinase C γ (PKCγ) by DAG combined with high-frequency Ca^{2+} spikes, which points to a joint code of calcium and DAG in that case.

Most theoretical studies also limit themselves to the spiking mode of calcium oscillations. Dupont et al. 2003 [DHD03] could successfully reproduce the findings of [DS98] in a model. Gall et al. 2000 [GBD00] examined the activation of liver glycogen phosphorylase by modeling a de-/phosphorylation cycle. Salazar et al. 2004 [SPH04] studied the activation of target proteins by Ca^{2+} oscillations in terms of efficiency, speed and specificity. The same group identified in [SPH08] three dimensionless parameters (effective activation rate, relative oscillation frequency and the duty ratio of the oscillations) that determine the mean activity of target proteins. Marhl et al. [MPS05; MPS06] investigated the decoding of time-limited calcium oscillations by downstream proteins.

Recently, the bursting mode of Ca^{2+} oscillations has been investigated by Larsen et al. 2004 [LOK04] and Schuster et al. 2005 [SKM05]. Using a simple model of calcium oscillations [LOK04] and artificially generated calcium bursts [SKM05] respectively to drive protein activation, these studies showed that specific information can be encoded in the waveform of bursting oscillations and thus that different proteins can be activated differentially at the same time. Rozi and Jia [RJ03] studied the activation of glycogen

phosphorylase by spiking as well as bursting calcium oscillations.

Even though information-theoretic measures [WS49] are in widespread use for physiological data, e.g. [IRW+05], and neural information transfer, e.g. [BT99], their application to biochemical systems is restricted to only relatively few studies. For instance, Prank et al. 1998 [PSL+98] and Kropp et al. 2005 [KGP05] studied the encoding of hormonal signals in intracellular calcium signals using the so-called coding fraction and mutual information. The authors drive a deterministic model of calcium spiking with a specific form of generated noise and estimate the amount of information transferred. In [PLvzM+98] the same group couples a deterministic model of CaM kinase activation to experimentally measured data from HIT (hamster insulin secreting tumor) cells, but here the results are not analyzed in an information-theoretic manner.

We propose to use the information-theoretic measure transfer entropy [Sch00] to estimate the information transferred by spiking or bursting calcium oscillations under different conditions. Transfer entropy has advantages over conventional measures such as (time-lagged) correlations, in that it detects all statistical dependencies (linear and non-linear), it is asymmetric, i.e. it distinguishes between information source and target, and it considers shared information due to a common history of the source and target by using conditioned transition probabilities. Transfer entropy has been applied to physiological data [IRW+05; KTO+06], financial time series [MK02], geological data [MWY07] and others [NST+06; LS06], but, so far, not to biochemical data.

We used both simulated and experimentally measured time series for the estimation of transfer entropy. The simulated data were generated by a stochastic version of the simple calcium oscillations model in [KOD+00] extended by a stochastically simulated activation of target protein. We set up a framework for stochastic simulation of the calcium system, stochastic coupling of the enzyme activation process and estimation of the transfer entropy using kernel density estimation methods. We used this framework to investigate calcium information transfer in systems with different levels of activation and particle numbers.

Since multivariate experimental data is scarce, we devised a method, inspired by hybrid deterministic/stochastic simulation techniques, which allows the stochastic coupling of the enzyme activation process to arbitrary univariate calcium time series. We took experimental data from single-cell measurements on rat hepatocytes and coupled the activation of the stochastically simulated enzyme to them in order to get bivariate data. Finally, we used these semi-experimental data as input for the estimation of the information transfer. Fig. 6.1 shows this procedure in a schematic view.

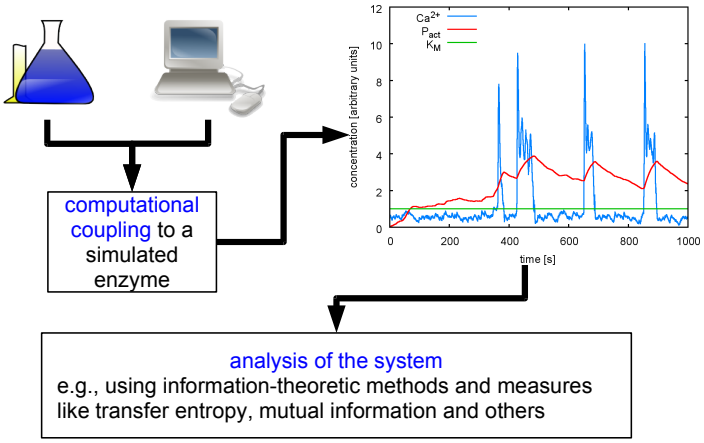


Figure 6.1: Schematic view on the coupling process and subsequent analysis.

6.3 Results

In this study we used a simple receptor-operated model [KOD+00] (cf. Sec. 6.6 for details on the model) with three variables (G_α, PLC, cytosolic Ca^{2+}) to generate calcium time series. This model was simulated stochastically by Gillespie's algorithm [Gil76] (see Sec. 2.4.1). Fig. 6.2 shows simulated time series of the Ca^{2+} concentration under different cellular activation levels. The model is able to display understimulation (data not shown), spiking (panel A), bursting (panel B) and irregular behavior (panel C) as well as overstimulation (panel D). Spiking and bursting behavior is observed experimentally when hepatocytes are stimulated with vasopressin and ATP, respectively.

The concentration of the active form of a simulated Ca^{2+}-dependent enzyme, which was stochastically coupled to the calcium data, is also shown. We implemented a stochastic coupling scheme to be able to couple the simulated enzyme to arbitrary, simulated or experimental, calcium time series. This method is described in detail in Sec. 6.6.2.

The coupling of the simulated enzyme to experimental data leads to semi-experimental time-series, one of which is shown in Fig. 6.3. Here an experimentally measured time series of the Ca^{2+} concentration in a single rat hepatocyte (see Sec. 6.6.4 for further details) was computationally coupled to a simulated target enzyme according to Eq. (6.10). The hepatocyte was stimulated with ATP, which led to a bursting mode of calcium oscillations. The

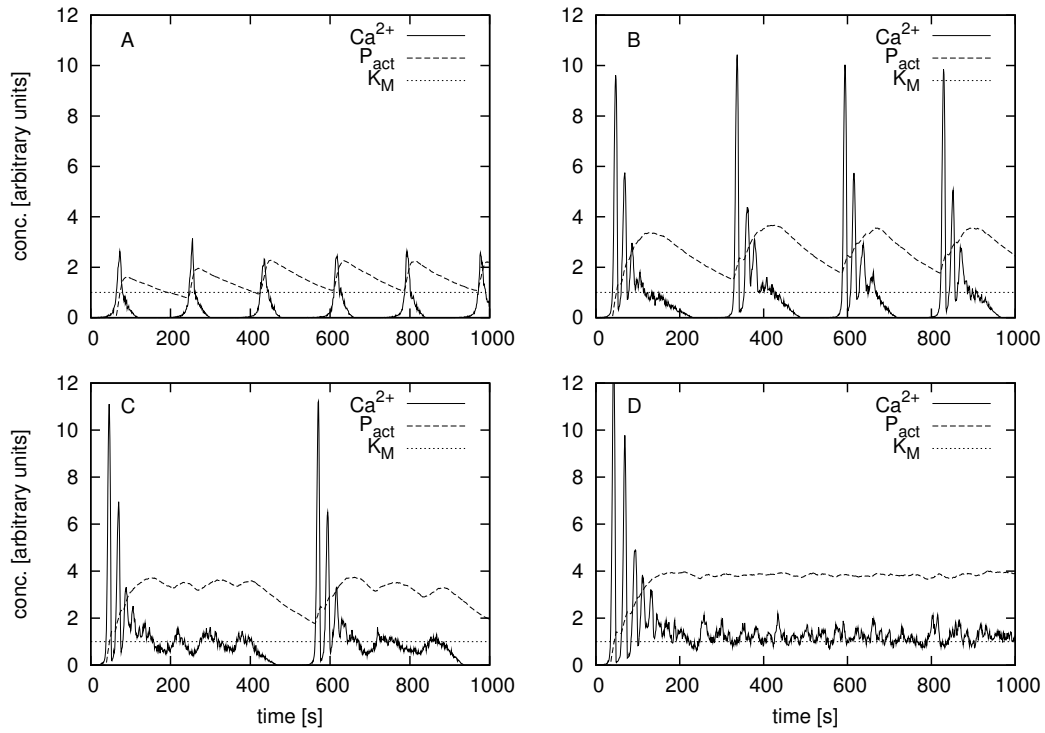


Figure 6.2: Coupling of the stochastically simulated activation of an enzyme to simulated calcium time series with different dynamical behavior according to Eq. (6.10). From top left to bottom right we see a spiking (A), bursting (B), irregular (C) behavior and overstimulation (D). k_2 values 2, 2.85, 2.99 and 3.2, respectively, and volume 10^{-12} [arbitrary units]. x-axis: time [s]. y-axis: concentration of Ca^{2+} and the active form of the enzyme P_{act} and the enzyme's K_M value.

integrating character of the enzyme, which was shown elsewhere [LOK04] to permit frequency decoding of the calcium oscillations, can easily be seen.

Using these simulated and semi-experimental time series we investigated the information transferred from the calcium signal to the enzyme by estimating the transfer entropy (see Sec. 6.6.3). In Fig. 6.4 an example of a scan over a range of bandwidths ϵ for the kernel density estimation is shown. The calcium system has been simulated in the bursting mode ($k_2 = 2.85$) and with different values for the volume leading to different particle numbers. We used time courses of length 10000 s, sampled every second, after a transient of 10000 s has been cut off. For the density estimation we used a rectangular kernel and set the length of the Theiler window to 20 and the minimal number of neighbors to 5. As shown in the diagram, the estimates are biased towards zero for $\epsilon \to 0$. For small ϵ values more and more sam-

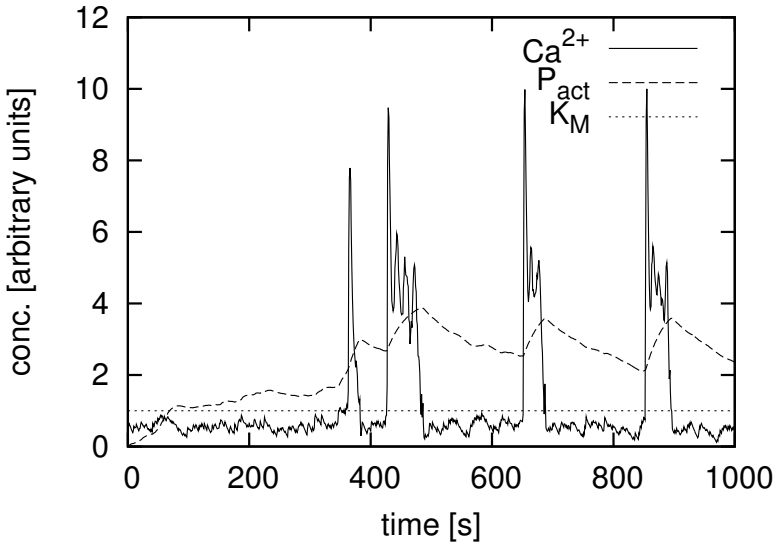


Figure 6.3: Coupling of the stochastically simulated activation of an enzyme to an experimentally measured calcium time series according to Eq. (6.10). Here the hepatocyte was stimulated using ATP (1.5 µM). x-axis: time [s]. y-axis: concentration of Ca^{2+} and the active form of the enzyme P_{act} and the simulated enzyme's K_M value (reaction volume of the simulated enzyme 10^{-10} [arbitrary units]).

ples are missing enough neighbors within the kernel bandwidth and those "lonely samples" are excluded from the estimation. For $\epsilon \to \infty$ the kernel eventually covers the whole attractor, which also results in a value of zero for the transfer entropy. In between, there is a plateau-like range, where the estimate is almost independent of the ϵ value and which is supposed to be the best estimate of the true information transfer. We plotted the corresponding maxima in the diagram (horizontal lines). In the following we will always use those maximal values of the ϵ scans as estimates of the transfer entropy (see Discussion in Sec. 6.4).

We also tested our estimation process by using surrogate data (constrained realizations, see Schreiber 2000 [SS00] for details). We estimated the transfer entropy of time series in which the temporal order of the calcium signal was destroyed by shuffling the samples (data not shown). This removed all dependencies, while the marginal probability distributions were preserved. Indeed, here the estimated transfer entropy showed values near zero ($\sim 0.02 - 0.07$).

To investigate how the information transfer changes with varying particle numbers in the system, we simulated the calcium model using a range of different volumes. Systems with low volumes, corresponding to low particle

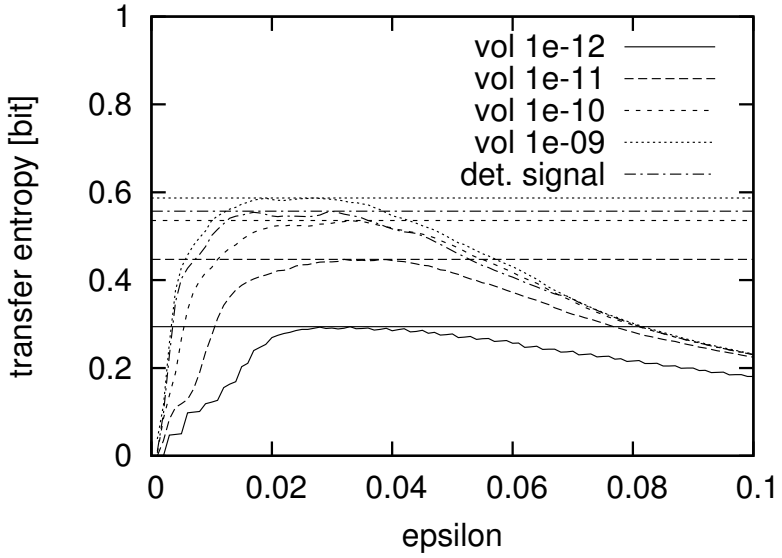


Figure 6.4: Scan of the estimated transfer entropies from Ca^{2+} to active protein P_{act} in the stochastically simulated system ($k_2 = 2.85$, bursting). x-axis: ϵ values. y-axis: estimates of the transfer entropy in simulated systems of volumes 10^{-12} to 10^{-9} [arbitrary units] respectively. Also, the estimating process was applied to a deterministically simulated calcium signal (det. signal). In this case, the reaction volume of the simulated enzyme was 10^{-10}.

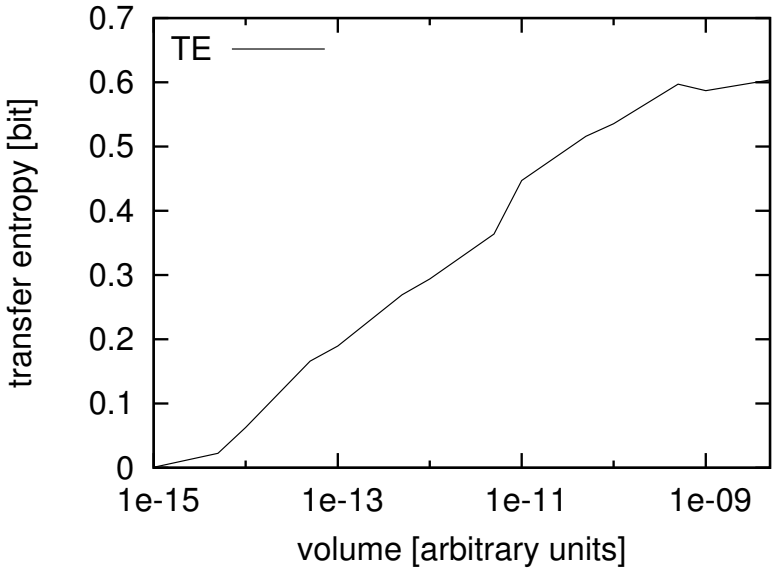


Figure 6.5: Maximum values of the estimated transfer entropy for different volumes in systems with $k_2 = 2.85$ (bursting). x-axis: volume. y-axis: estimates of the transfer entropy.

numbers, usually display strong random fluctuations, which could hamper the information transfer. Therefore our hypothesis was that a minimal number of particles are needed to allow for the faithful transfer of a certain amount of information. In fact, this is the case. Fig. 6.5 shows a scan of the transfer entropy of simulated systems in the bursting mode ($k_2 = 2.85$) with different volumes. Here the information transfer increases with increasing volume (and particle numbers) until it seems to flatten out at about 0.6 bit/sample for volumes greater than $5 \cdot 10^{-10}$ [arbitrary units]. Interestingly, this corresponds to the particle numbers where the simulations display quasi-deterministic behavior (see Chap. 5). With even higher volumes the system should eventually converge to the deterministic limit. In this case, also the coupling would be quasi-deterministic and the estimation of the transfer entropy should diverge (see discussion in Sec. 6.4). Therefore, regimes where the transfer entropy does not increase uniformly with increasing volume deserve further study, since this would be a helpful indicator that the transition to quasi-deterministic behavior is not uniform (see Chap. 5). However, the huge computational cost prevented us from testing whether or not the apparent flattening is statistically significant in this case.

We also investigated the information transfer when the calcium system is in different dynamical modes (cf. Fig. 6.2). Fig. 6.6 shows a scan of transfer entropy estimates for different volumes (between 10^{-13} and $5 \cdot 10^{-9}$) where

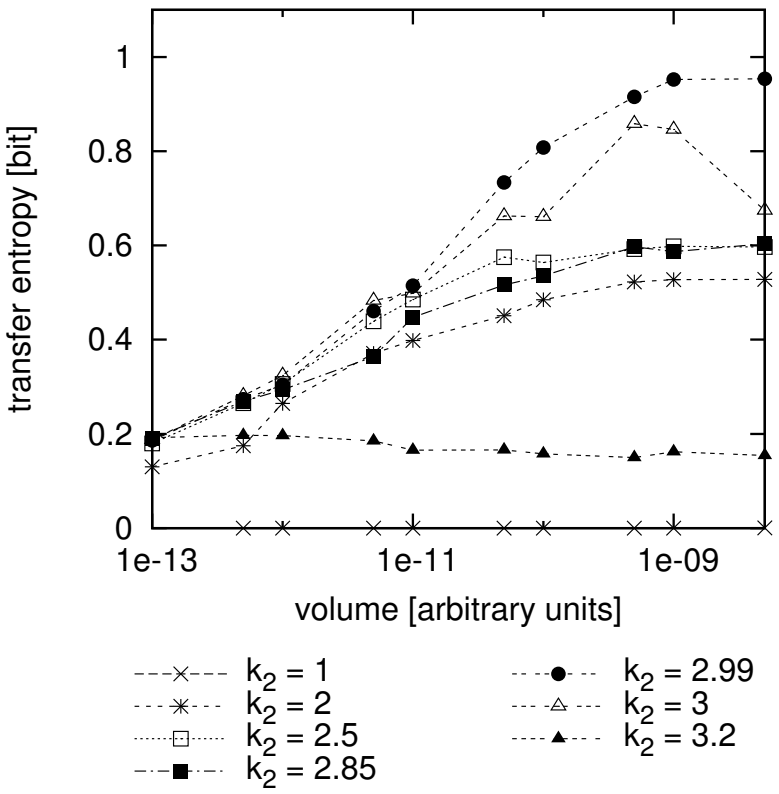


Figure 6.6: Maximum values of the estimated transfer entropy for different volumes and different k_2 values corresponding to different dynamic modes (1 understimulation, 2 spiking, 2.5, 2.85 bursting, 2.99 irregular/elevated oscillations, 3, 3.2 overstimulation) in the simulated system. x-axis: volume. y-axis: estimates of the transfer entropy.

we varied the value of the bifurcation parameter k_2 to get different dynamics, such as understimulation ($k_2 = 1$), spiking ($k_2 = 2$), bursting ($k_2 = 2.5, 2.85$), irregular behavior/elevated oscillations ($k_2 = 2.99$) and overstimulation ($k_2 = 3, 3.2$).

In the case of under- or overstimulation ($k_2 = 1$ or $k_2 = 3.2$), the system is in a (noisy) steady state and this results in low values for the transfer entropy. For $k_2 = 1$ the calcium concentration is near its resting level, which is far below the K_M value of the enzyme. No enzyme gets activated and no information can be transferred. For $k_2 = 3.2$ the calcium steady state lies above the enzyme's K_M value and the amount of active enzyme reaches its maximum. In contrast to understimulation, here the information transfer is not exactly zero, even though it takes low values of ~ 0.2. The reason for this is that now the noisy steady state is near the K_M value of the enzyme

and it can pick up some random fluctuations in calcium concentration.

If the system is in an oscillatory mode, such as spiking ($k_2 = 2$) or bursting ($k_2 = 2.5, 2.85$), the transfer entropy increases with increasing volume until it seems to flatten out for volumes above $5 \cdot 10^{-10}$, as shown above.

An interesting effect can be observed for $k_2 = 2.99$ and $k_2 = 3$, where the deterministic limits of the calcium dynamics are elevated oscillations and an elevated steady state, respectively. However, the stochastic system shows irregular behavior with small volumes. For high volumes, oscillations are observed even for $k_2 = 3$. For both parameter values, the generally very high level of transfer entropy is due to the position of the center of their oscillations. It is near the K_M value of the enzyme, so that the enzyme is responsive even to minute variations in the calcium level. Interestingly, for $k_2 = 3$ the transfer entropy shows a maximum at the volumes 10^{-9} and $5 \cdot 10^{-10}$. An explanation for this effect is that for the higher volume $5 \cdot 10^{-9}$, the system is already near the deterministic limit, which is just a rather uninteresting elevated steady state with relatively low information transfer. On the other hand, for smaller volumes, the information transfer gets degraded because of increasing stochastic fluctuations. Those fluctuations are especially pronounced in this parameter range, because the sensitivity of the system (measured by the divergence) is high (see Chap. 5 for details). Those two opposed trends lead to a maximum in a range where the system is still oscillatory, but not yet too noisy.

If we look at the estimates for a volume of $5 \cdot 10^{-9}$ only (the biggest systems considered in this study), there is a slight increase in estimated transfer entropy from spiking to increasingly complex bursting oscillations (see Table 6.1). The transfer entropy is very high for elevated oscillations near the enzyme's K_M value and it drops to a very low value in the case of an elevated steady state, e.g. overstimulation. Intuitively, one would think that the information transfer should be correlated to the complexity (spiking, bursting or irregular oscillations) of the calcium oscillations, since more complex input signals can potentially carry more information. However, this can only be hinted at from our experiments. One should be wary not to over-interpret the absolute numbers, since we found them very much dependent on the estimation process used. Also, they are subject to statistical fluctuations. Furthermore, the enzyme is most sensitive for calcium levels near its K_M value. For the input signal to generate a high information transfer, it is important to meet that range. The transfer entropy nicely detects this for the oscillatory regime with $k_2 = 2.99$ and high volumes, where the oscillations exactly meet the K_M value. Here the estimated transfer entropy is high, even though the dynamics is apparently less complex than in the bursting mode.

To compare simulations with experimental data we coupled an experimen-

k_2	Dynamic behavior	TE
1	Understimulation	0.00
2	Spiking	0.52
2.5	Bursting	0.59
2.85	Bursting	0.60
2.99	Elevated oscillations	0.95
3.2	Overstimulation	0.15

Table 6.1: Maximum values of the transfer entropy (TE) for different stimulation strengths k_2 and their respective dynamic regime in the simulated system with volume $5 \cdot 10^{-9}$.

tally measured calcium time series from a single hepatocyte to the stochastic model of enzyme activation. In this case the cell was stimulated using 1.5 µM ATP and showed bursting behavior (see Fig. 6.7, inset). We monitored the calcium concentration over a time period of 3904 s (one sample per second). The reaction volume of the simulated enzyme was set to 10^{-10} [arbitrary units]. For the kernel density estimation, we used a Theiler window of length 20 and reduced the minimal number of neighbors to 2 because of the smaller number of samples available. Fig. 6.7 shows a scan of the transfer entropy estimates from this semi-experimental time series over a range of ϵ values. The estimated transfer entropy has a maximum at about 0.35 bit.

For a direct comparison, we calculated 10 stochastically simulated calcium time series of length 3904 s showing bursting behavior ($k_2 = 2.85$). One of them can be seen in the inset of Fig. 6.8. We then coupled these time series to the same enzyme process and estimated the transfer entropy using the same set of parameters as before. We plotted the results of the 10 different simulations plus the mean value in Fig. 6.8. The mean of the estimated transfer entropies has a maximum of about 0.57 bit. The variance of the estimated values is biggest in the plateau region with a maximum in standard deviation of approximately 0.03.

The significantly higher transfer entropy values of the simulated system can partly be explained by the existence of two episodes in the experimental data without bursts (The calcium-mobilizing agonist was absent from the experimental medium for the duration of these two episodes). We removed these episodes and repeated the estimation which yielded a transfer entropy maximum of roughly 0.39 bit. An explanation for the remaining discrepancy is that the simulated bursts have a considerably longer duration than the bursts in real hepatocytes. Therefore, the calcium signal spends more time within the sensitive region of the enzyme (near the K_M value) which clearly

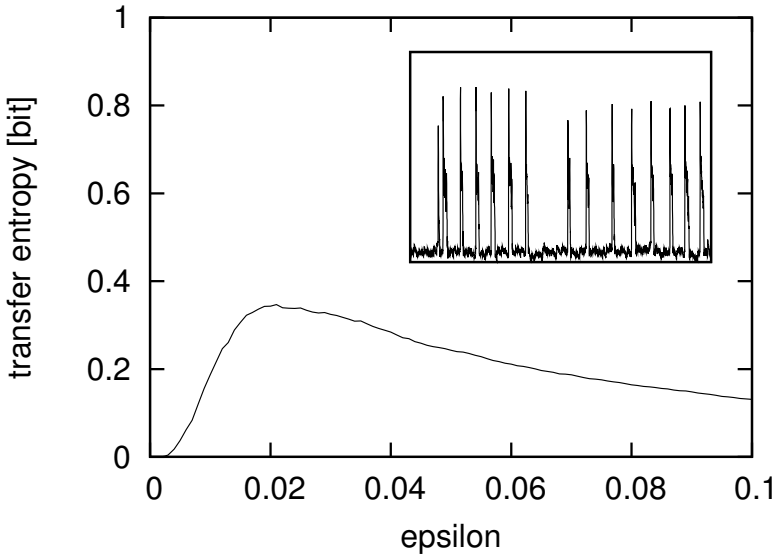


Figure 6.7: Scan of the estimated transfer entropy from an experimentally measured Ca^{2+} time series (3904 s) to the simulated target enzyme P_{act} according to Eq. (6.10). Here the hepatocyte was stimulated using ATP (1.5 µM). x-axis: ϵ values. y-axis: estimates of the transfer entropy (reaction volume of the simulated enzyme 10^{-10} [arbitrary units]).

increases information transfer.

6.4 Discussion

In the following we will motivate the choice of several technical elements as well as discuss their strengths and limitations.

6.4.1 Stochastic coupling procedure

Stochastic fluctuations in cellular systems are not just random noise, but can even change the dynamics of the system [RWA02] as was seen, for instance, in our simulations for parameter values near bifurcation points ($k_2 = 2.99$ and small volumes). Therefore it is important to consider random effects (and the effects of the system size on those fluctuations) when modeling systems with relatively low particle numbers, e.g. signal transduction pathways.

It should be noted here that, even in those cases where stochastic effects do not change the dynamics significantly, deterministic coupling of a biochemical reaction system to experimental data, as, e.g., in [PLvzM+98], is not appropriate for our purposes. The estimation of transfer entropy di-

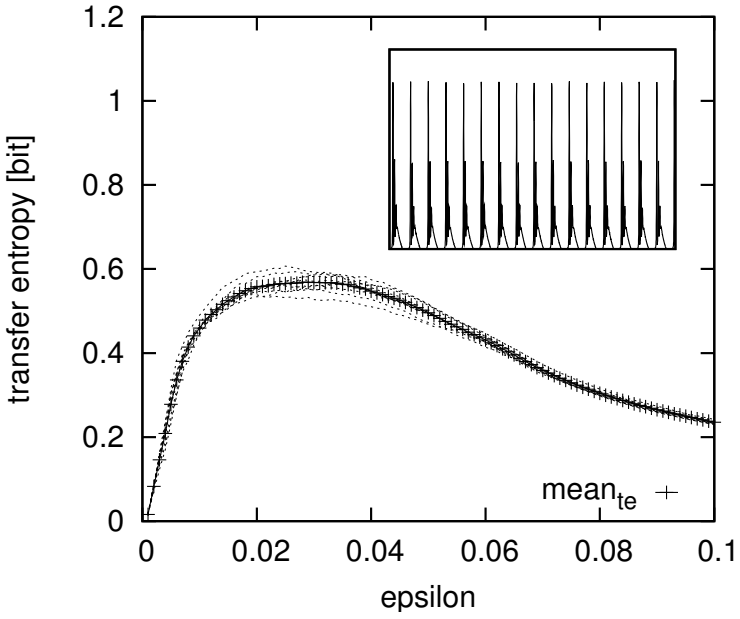


Figure 6.8: Scan of the estimated transfer entropies from Ca^{2+} to active enzyme P_{act} in the stochastically simulated system with $k_2 = 2.85$ and 10^{-10} volume. Shown are the estimates for ten different time series of length 3904 s each and their mean value. x-axis: ϵ values. y-axis: estimates of the transfer entropy.

verges for coarse-grained continuous systems and increasing resolution if the coupling between the processes is deterministic [Sch00]. Therefore our stochastic coupling scheme of the simulated enzyme to calcium time series is absolutely essential for this study.

Since the experimentally measured calcium concentration is only known at a discrete set of points in time and therefore we assumed it to be constant between two samplings, the coupling of a simulated enzyme to those time series can only be an approximation. However, it is apparent that, in the limit of a sampling frequency of the given time series near the frequency of reaction events in the system and a measurement resolution in the range of single particles, our method converges to the mathematically exact solution. For nearly every practical case, neither the number of samples nor the resolution will satisfy these theoretical conditions. To make sure that this fact did not compromise our results, we compared simulated data where the enzyme was only coupled to a calcium time series with data that was calculated by exact stochastic simulation of the whole system, i.e. calcium dynamics plus enzyme activation, and where no approximation was involved (data not shown). For the parameter values and sampling times we used, our results were not changed considerably by the approximate coupling.

One shortcoming of the stochastic coupling procedure described here is that it is a one-way process. Obviously, the input calcium time series is fixed and can not be changed during the process and so possible feedback of the target enzyme on the calcium system, e.g. calcium buffering by proteins or feedback via protein kinase C, has to be neglected.

6.4.2 Choice of model parameters

The volume of a hepatocyte is about 2 pL [SAB+07]. Assuming that the cytosol, where the free Ca^{2+} is located, takes up about half of the total volume of the cell and that, in the case of bursting, the calcium level peaks around 1 µM, this results in a particle number of about 600 000. This particle number roughly corresponds to a volume of 10^{-10} in the arbitrary units of the calcium model used. Therefore our results lie well in the range of physiologically meaningful parameters. Also the parameters of the simulated enzyme have been chosen to be, at least, biologically plausible. Most of the time calcium binding to enzymes occurs cooperatively, as e.g. with calmodulin. Calmodulin has four binding sites with high affinity ($K_d \approx 0.1 - 1\,\mu M$) for Ca^{2+}. For this reason, we, like the authors of other numerical studies [PLvzM+98; LOK04; SKM05], employ a Hill term of 4^{th} order. The K_M value of the simulated enzyme lies between the calcium resting level and the amplitude of secondary peaks, in the case of bursting oscillations.

The reason for choosing this core calcium model instead of a more detailed one was that, even though it is very simple in terms of size and kinetic functions, it can show both spiking and bursting behavior in addition to (elevated) steady states, thereby mimicking the dynamics of real cells after stimulation by different agonists (see [KOD+00] for details). Most other models cannot show bursting oscillations. It also was relatively easy to implement and fast to simulate stochastically. Nevertheless, the generation of some of the time series with high particle numbers required computation times in the range of several days. In fact, the purpose of this study was not to analyze this specific calcium model and therefore the approach presented here is not restricted to that model. It should also be mentioned that our framework can easily be applied to arbitrary enzyme regulation mechanisms, provided that they allow stochastic simulation of the Gillespie type, i.e. a propensity can be assigned to every possible event in the system.

One problem of the calcium model we used is that the amplitudes of the oscillations vary for different dynamic modes (see Fig. 6.2), whereas in real hepatocytes the amplitudes of calcium oscillations have been reported to be independent of the type of oscillations. Also the duration of bursts is longer than in experiments which, we believe, led to the discrepancy in transfer entropy between simulated and experimental data (see Figs. 6.7 and 6.8). To mitigate this issues we plan to use more realistic calcium models with more constant oscillation amplitudes, e.g. [LOK04], in the future.

6.4.3 Estimation of transfer entropy

Often, (time-lagged) correlations are used to quantify the coherence of two observables. However, correlations can only indicate linear relations, not non-linear ones. Therefore mutual information has been developed which is sensitive to all statistical dependencies [SKD+02]. Unfortunately, this measure is still (like correlations) symmetric and cannot distinguish between information sources and targets.

The transfer entropy, on the other hand, is explicitly asymmetric because it uses conditioned transition probabilities. As stated by Schreiber [Sch00, p. 461], "*transfer entropy* is able to distinguish effectively driving and responding elements and to detect asymmetry in the interaction of subsystems." In addition, the use of transition probabilities makes it a dynamic measure, meaning that it can account for the history of the processes. This, together with its ability to consider linear and non-linear dependencies, renders it appropriate for use on non-linear signal transduction systems.

We found that a major issue with this measure is that it is not trivial to be estimated from time series in a reliable way and that the estimation is

quite data-intensive.

One crucial point is that the processes have to be ergodic to allow for the estimation of the probability densities from one time series alone. Also they must be Markovian. In other words, their histories of length k and l (see Sec. 6.6.3), which are taken into account, must be longer than possible correlation times. This is very important, because the transfer entropy detects the deviation from this Markov property. One simple example where this condition would not be fulfilled is when we just reversed the direction and estimated the transfer entropy from the enzyme signal to the calcium signal ($P_{act} \rightarrow Ca^{2+}$). We saw already that in our setting there can be no feedback from the enzyme to the calcium system and thus no information can be transferred this way. Because the transfer entropy is a directional measure and can distinguish between information transferred in one and the other direction, one would naively think that it should equal zero (plus statistical fluctuations) here. This, however, is not the case, because the calcium signal alone is not Markovian. In fact, in the model it is coupled to G_α and PLC and their influence will lead to a transfer entropy which is not zero. There are two possible solutions to this issue: a) consider the whole system (Ca^{2+}, G_α and PLC) or condition on all coupled subsystems, or b) take into account a long enough history for the processes in which all relevant information is already embedded.

In all practical applications of the transfer entropy, especially with purely experimental data, one has to fix the lengths of the two signal histories (k and l) with care. Since the characteristic time-scale of auto-dependencies in measured data is not known a priori, they can not be regarded as stemming from an order-one Markov process. Therefore, one should estimate the transfer entropy using different values for the order of the underlying processes, and longer histories should be preferred. However, the often very limited amount of data renders this avenue infeasible in many cases, since kernel estimation would have to be applied to distribution functions in four and more dimensions. One possible resort here would be coarse-graining of the time series and the use of the discrete version of the transfer entropy. In the present study we restricted ourselves to the order-one case, the reason being that, in our case, the coupled protein is actually described by a Markov process of order one and is not dependent on previous values. Therefore, a history length of 1 ($k = l = 1$) suffices.

Kernel density estimation is known to be very dependent on the choice of a correct kernel bandwidth ϵ. Rules of thumb exist for the optimal bandwidth of (univariate) Gaussian kernels (see [Sil86]) which, however, are said to often lead to oversmoothing. Little has been done for multivariate kernels however. Instead of just using one bandwidth, we scanned the estimated

transfer entropy over a range of different ϵ values and checked for the range of bandwidths where the estimates are independent of ϵ, e.g. a plateau is visible in the scans (Fig. 6.4). If there is a definite plateau, its values are simultaneously the maximal values of the scan. Due to this and because the estimated transfer entropy was observed to underestimate the true value [KS02], we chose to take the maxima of the scans as estimates of the transfer entropy.

The calcium signal and the enzyme signal have different ranges of values. Therefore we normalized the time series to have mean 0.0 and standard deviation 1.0 prior to the estimation, which allowed us to use the same ϵ in both spaces. This is justified, because the (continuous) transfer entropy is independent of coordinate transformations [KS02].

To improve our calculations, we used a Theiler window approach and excluded all estimates where only less than a required minimal number of neighbors could be found. This avoided spurious effects caused by temporal correlations and dampened statistical fluctuations, respectively. In this study we mainly used rectangular kernels. However, we also tried Gaussian kernels (data not shown), which did not change our results considerably.

Transfer entropy is an averaged measure, i.e. it describes the information transfer over the whole observation interval. We observed that periods in the experimental calcium time series without bursts (Fig. 6.7) decreased the overall transfer entropy. Therefore, if the processes under study are expected to show some kind of locking or unlocking episodes, which we would dub *statistical locking*, the measure would have to be calculated on smaller (disjoint or overlapping) windows in order to see possible changes over time. Care has to be taken, though, that the windows are big enough to get a sound statistical basis for the estimation.

We want to stress that the absolute values of our transfer entropy estimates are, of course, dependent on the parameters of the estimation procedure. In particular, the minimum number of neighbors needed for a sample to be considered plays a major role here. Setting this number to values greater than 1 helps to diminish statistical fluctuations, but can create a bias towards zero if there are not enough samples available. Therefore, one should be cautious when interpreting these values and should not mix results coming from different estimation procedures without justification. We only compared estimates where the estimation parameters, the type of kernel and the length of the input time series were the same. We attributed the discrepancy in estimated transfer entropy of simulated and experimentally measured data to lacking realism of the simple calcium oscillations model used. Hence, we note here that transfer entropy could very well be employed as a measure of realism for signaling pathway models. We envisage its use

in biochemical modeling where models are optimized so as to have the same information transfer as observed in experiment.

Also, regarding the rates of information transfer we estimated in this study, one should be cautious. Even though they can provide a useful basis for hypotheses on the functioning of cellular signal transduction, it is not known what fraction of the information that can maximally be transferred is actually used by downstream cellular processes. Because it is not yet clear what features of the calcium signal really carry relevant information, we used an information-theoretic approach. It potentially measures all the information from the calcium signal that can be found in the protein signal. In addition, this model-free approach facilitates direct comparisons between simulated and experimentally measured data. Nevertheless, specific information transferred from calcium to cellular processes could, in principle, be estimated by extending the simple model to include these processes under consideration and estimating the transfer entropy directly between calcium and the observables of these processes. This includes the detection, analysis and quantification of possible cross-talk between different signaling pathways.

6.4.4 A general framework

It should be mentioned that there are many potential variants and extensions of the estimation algorithm (simple or adaptive histograms, adaptive kernel density estimation, likelihood estimators and others [Sil86]), which we could not cover here. However, regardless of the algorithm used, the basic strength of the information-theoretic approach is that it is model-free. This allows the direct comparison of simulated and experimental data.

6.5 Conclusions

In this study we combined methodologies from different fields in order to elucidate the cellular information transfer via Ca^{2+} signaling. The main ingredients we used are:

- Modeling and simulation of calcium signal transduction, in particular stochastic approaches.
- Stochastic coupling of a Ca^{2+}-dependent protein to experimental and simulated data.
- The so-called transfer entropy introduced by Schreiber [Sch00] and its estimation using kernel density estimation techniques.

We developed and implemented a framework for the analysis of both simulated and experimentally measured Ca^{2+} time series with the information-theoretic measure transfer entropy. This involved the stochastic coupling of a simulated enzyme to arbitrary calcium time series and the estimation of the transfer entropy of these bivariate data using kernel density estimation methods.

This study presents the first application of transfer entropy to biochemical signaling pathways.

We investigated the information transfer from calcium to the target enzyme under different conditions, namely different particle numbers (by varying the volume) and different calcium dynamics (corresponding to different stimuli). We found that, most of the time, information transfer increases with increasing particle numbers in the system. However, this increase is different in systems with different dynamic modes (spiking, bursting, etc.). More complex dynamic modes (bursting or irregular oscillations) tend to result in higher values of the transfer entropy. We observed that the input signal has to lie in the sensitive range, e.g. near the K_M value of the enzyme, for the information transfer to be efficient. We also estimated the transfer entropy based on experimental data from hepatocytes. The values of these estimates are significantly lower than those from comparable simulated data. The major reason for this seems to be the unphysiologically long duration of simulated bursts. Further study is needed to investigate that in detail.

Even though the estimation of transfer entropy from time series is tricky and there are still some unsolved issues, it is a promising tool not only for the quantification of information transfer in biochemical networks, but also, for instance, to distinguish between different stochastic time series where a pure visual investigation is difficult. The direct comparison of two stochastic trajectories is difficult: Not the actual trajectory is important, but the features of it, that are essential for the correct functioning of the cell. In the case of calcium signaling, they are the ones that can be decoded by downstream elements.

Each dynamic state exhibits its own sensitivity to random fluctuations (see Chap. 5) and this should be reflected in the faster degradation of information transfer if the sensitivity is high. Therefore, one possible application of this approach could be the detection of the transition between stochastic and quasi-deterministic behavior, in cases where it is difficult to be identified by visual inspection alone. We saw one example of that already in the case of $k_2 = 2.99$ (see Sec. 6.3), where the stochastic behavior is qualitatively different from the deterministic limit and where the transfer entropy could detect this transition. Another application could be information theory-based model fitting where models are optimized so as to have the same information

transfer as observed experimentally.

It is worth mentioning that our framework is not at all limited to calcium signaling. Stochastic coupling and/or estimation of transfer entropy from biochemical data can be easily applied to other biochemical models and pathways. Therefore, we think, that our approach could also be a valuable tool for the analysis of other signaling pathways.

Our approach can also be extended in a number of ways. On the technical side, for example, the estimation of transfer entropy from limited data sets should be improved. This could include the automatic determination of an optimal kernel bandwidth, the use of different kernels or alternative probability density estimation methods.

On the biological side, we plan to investigate the type and amount of information carried by the different properties of the calcium signal (amplitude, frequency, duration, shape, timing), because it is not yet clear which of those are really used in cells. Thus we hope that the transfer entropy can give valuable hints for further theoretical and experimental studies. Furthermore, we want to use our framework to study different enzyme regulation mechanisms and to analyze other signaling pathways including their possible cross-talks.

6.6 Methods

6.6.1 Model

In this study we used a simple receptor-operated model [KOD$^+$00] with three variables (G_α, PLC, cytosolic Ca^{2+}) to generate calcium time series (see Sec. 3.2).

This model was simulated stochastically by Gillespie's algorithm [Gil76] (see Sec. 2.4.1 for details). Because the original model has arbitrary units, we scaled it in time (by 1/20) to have roughly the same frequency as observed experimentally. This scaling corresponds to a division of the rate parameters $k_1, k_2, k_3, k_5, k_7, k_8, k_{10}$ and k_{11} by 20.

The parameter k_2 represents the stimulation strength and serves as bifurcation parameter to vary the dynamic behavior of the model. In addition, we changed the numbers of particles present by varying the volume of the system.

Coupled to this simple signal generating model is a model for calcium binding to a protein. In the following we will use a slight modification of the regulation mechanism described in [LOK04].

$$\frac{\mathrm{d}[\mathrm{P_{act}}]}{\mathrm{d}t} = \frac{k_{\mathrm{act}} \cdot [\mathrm{Ca^{2+}}]^p}{K_M^p + [\mathrm{Ca^{2+}}]^p} \cdot [\mathrm{P_{inact}}] - k_{\mathrm{inact}} \cdot [\mathrm{P_{act}}] \tag{6.10}$$

with $[\mathrm{P_{tot}}] = [\mathrm{P_{inact}}] + [\mathrm{P_{act}}] = \text{constant}$.

Activation of the inactive form of protein $\mathrm{P_{inact}}$ to its active form $\mathrm{P_{act}}$ by $\mathrm{Ca^{2+}}$ is modeled by a Hill term of order four while deactivation follows mass action kinetics (Eq. (6.10)). The parameters were set to $k_{\mathrm{act}} = 0.025$, $k_{\mathrm{inact}} = 0.005$, $K_M = 1.0$, $[\mathrm{P_{tot}}] = 5.0$ and $p = 4$.

6.6.2 Stochastic coupling

As mentioned, we employed Gillespie's algorithm for the stochastic simulation of the model system.

However, since we not only want to analyze simulated calcium dynamics, but also intend to couple measured calcium time series to our enzyme activation model, we have to take that influence into account. The coupled calcium system exerts an influence on the reaction propensities a_μ in the protein model and thus they can no longer be considered constant between two reaction events. Mathematically, this is equivalent to changing the homogeneous Poisson process into an inhomogeneous one and therefore the pure stochastic simulation methods cannot be used in this case.

The reaction probability density function for such systems with time-dependent a_μ reads (see Gillespie 1992 [Gil92]):

$$P(\tau, \mu) = a_\mu(t+\tau) \exp\left(- \int_t^{t+\tau} \sum_{1..M} a_\mu(t) dt \right). \tag{6.11}$$

One can sample this inhomogeneous Poisson process by integrating the differential reaction probabilities over time. Whenever a stopping criterion for one of the reactions is reached, the integration is interrupted and the corresponding reaction event is instantiated. This method has been used in hybrid stochastic/deterministic simulation methods (see Chap. 2), where the set of reactions is partitioned into a stochastically simulated and a deterministically simulated subset. During the simulation, the influence of the (fast) deterministic subset on the stochastic subset has to be considered.

When we couple a time series to a stochastically simulated system, we do not know the states of the system which produced the time series between

two samples. Therefore a reasonable presumption is to assume piece-wise constant particle numbers between two sample times. In this special case we can use, for instance, Gillespie's Direct Method, with the modification that recalculation of all the a_μ in the system, which are dependent on the coupled time series, is needed whenever a new sample is observed in the time series. This approximation is discussed in Sec. 6.4.

We implemented this simple method in C++-code dynamically linked to OCTAVE (Version 2.9.9 on Linux)[Oct]. Our implementation accepts time series with arbitrary sampling times, both evenly and unevenly sampled.

6.6.3 Kernel density estimation

For the estimation of the transfer entropy, usually either the time series is coarse-grained by histogram-based methods and the transfer entropy is estimated on the symbolic time series or kernel density estimation (Parzen window) [Sil86] methods are used.

In this study, we set the history lengths $k = l = 1$, which is justified in the discussion (Sec. 6.4). One should keep in mind though, that in the general case longer history lengths might be required. Setting those parameters correctly is crucial for a reliable estimation. If this is the case, probability densities in spaces of dimension > 3 must be estimated.

We implemented a kernel density estimation method for the transfer entropy (see [KS02]) in C++-code, which has been dynamically linked to OCTAVE (Version 2.9.9 on Linux)[Oct]. For the estimation of local probability densities, we mainly used a rectangular kernel with variable radius ϵ:

$$\hat{p}_{\{x_i\},\epsilon}(x_i) = \frac{1}{N\epsilon} \sum_{n=1}^{N} K\left(\frac{x_i - x_i(n)}{\epsilon}\right) \tag{6.12}$$
$$K(r) = \frac{1}{2}\Theta(1 - |r|)$$

with Θ the Heaviside function.

To avoid spurious effects caused by temporal correlations, we employed a Theiler window approach which excluded all neighbors that were too close in time. In addition, in order to dampen statistical fluctuations, only samples that had a user-defined minimal number of neighbors were considered. The kernel density estimation procedure was implemented using a 2-dimensional

box-assisted neighbor-searching algorithm [KS97], which resulted in a five to six fold speed-up compared to the naive implementation.

We scanned the transfer entropy of the simulated calcium model and the coupled enzyme for different volumes (between 10^{-13} and $5 \cdot 10^{-9}$ [arbitrary units]), corresponding to different particle numbers in the system (roughly between 600 and 30 000 000 during primary peaks), and for different dynamics, e.g. different values of the bifurcation parameter k_2 (1 understimulation, 2 spiking, 2.5, 2.85 bursting, 2.99 irregular, 3, 3.2 overstimulation). The same kernel bandwidth were used in the calcium and the protein concentration spaces, but the data was normalized to have mean 0.0 and standard deviation 1.0 prior to the estimation.

6.6.4 Experiments

The experimental setup and the materials used to measure the calcium concentration in single rat hepatocytes upon stimulation by different agonists is described in Sec. 3.4.

In addition to the procedure described there, we transformed these data to be roughly in the same range of values ($[0, 10]$) as that of the simulated data.

Acknowledgments

JP and UK would like to thank the Klaus Tschira Foundation and the BMBF for financial support. Also we would like to thank Sven Sahle for fruitful discussions and Katrin Hübner, Markus Müller and Jocelyn Faberman for helpful comments on the manuscript.

Chapter 7

General conclusions and outlook

> Prediction is very difficult,
> especially about the future.
>
> Niels Bohr

The underlying theme of the present thesis is the investigation of the border between stochastically and deterministically evolving biochemical systems. This intermediate regime is highly important due to a number of reasons:

First of all, biological systems seem to have evolved exactly towards that border [RWA02] in many cases. In Chap. 5, for instance, we observed that the number of receptors and channels in the Ca^{2+}-phosphoinositide pathway of hepatocytes are roughly in the range where the transition from stochastic to quasi-deterministic behavior occurs. Conceivable driving forces for this phenomenon include the trade-off between minimal maintenance costs ($\rightarrow$ small particle numbers) and reliable functioning of the cell ($\rightarrow$ high particle numbers)[1]. Signaling pathways, e.g., are dependent on a minimal number of particles to reach a certain rate of information transfer (see Chap. 6). Also, cells possibly attenuate [BFLM05] or produce noise [SPA05] in order to keep it on a certain intermediate level that, on the one hand, allows exploitation of stochastic effects and, on the other hand, avoids the detrimental aspects of molecular fluctuations.

Another argument for the study of this middle range is the emergence of sophisticated single-cell measuring techniques [ELSS02; $OTK^{+}02$] that extend the set of observable phenomena to include stochastic effects beyond the mere bulk behavior of biochemical networks.

[1] A related but deterministic approach is the "principle of flux minimization" in [Hol04].

Finally, a very practical reason for the relevance of this study is that the field of Systems Biology promised to shed some light on the systemic aspects of biochemical systems. This implies the formulation and analysis of increasingly big and heterogeneous models, as, for instance, in Klipp et al. 2005 [KNK+05], which integrate gene expression, metabolic and signaling pathways. The simulation of these models requires algorithms that are able to span the gap between stochastic and deterministic regimes, such as the approximate and hybrid methods reviewed in Chap. 2. Since already the consideration of growing volume and cell division or temperature changes, in principle, necessitates hybrid modeling, it is anticipated that this area will even be of more importance in the future. Very helpful in that respect could also be finding means to predict whether or not pronounced stochastic effects should be expected in a specific model system (see Chap. 5).

We explored this intermediate range between stochastic and deterministic behavior primarily from three different angles:

- The computational aspect (Chapters 2 and 4): What simulation approaches exist, especially approximate and hybrid methods, that are capable of bridging the gap between the different regimes? What are the potential hurdles and pitfalls?

- What determines the transition from stochastic to quasi-deterministic behavior (Chap. 5)? How can this transition range be located in specific models?

- Stochastic analysis (Chap. 6): How can the performance of signaling pathways be quantified (information theory) and how does this change under different cellular conditions, particularly system size?

These perspectives are intimately connected: Chap. 2 with an overview of stochastic simulation algorithms, of course, provides the basis for all subsequent chapters. The divergence (Chap. 5) as easily computable quantity can indicate the onset of stochastic effects. Therefore, it could be used to aid researchers looking for the most appropriate simulation method or as partitioning criterion in automated hybrid simulation methods (Chap. 2). Hybrid methods can also be used to elucidate the transition from stochastic to deterministic behavior. Which reactions are mostly responsible for stochastic behavior can be analyzed by varying the partitioning scheme. The stochastic coupling procedure in Chap. 6 that uses ideas from hybrid simulation represents another connection. Also, the transfer entropy (Chap. 6) for the quantification of information transfer seems to be useful to distinguish between apparently similar stochastic trajectories in a more objective manner.

We mostly used calcium signal transduction (Chap. 3) as test system because it is not only a well-studied and physiologically very relevant system but also prototypically exhibits all the properties we sought to investigate. However, we also showed the general applicability of our approaches, for instance in Chap. 5, by testing them on other systems like the peroxidase-oxidase reaction [OLK03] and mitogen-activated protein kinase cascades [Kho00].

This study comprises, to the best of our knowledge, the first application of transfer entropy to biochemical signaling pathways. Also, the use of the divergence value as an indicator for the stochastic/deterministic transition is an original contribution. The stochastic coupling procedure introduced in Chap. 6 extends the common deterministic coupling of time series and ODEs. It is essential for systems where randomness ought to be taken into account. We collected and analyzed the different stochastic methods that were scattered throughout the literature in the hope that our systematic presentation will help researchers find the most appropriate simulation algorithm for their specific model.

7.1 Outlook & suggestions for future projects

In the course of this study, naturally, more questions were raised than could currently be answered. A number of promising directions for future work were already discussed in the respective chapters. Nevertheless, we will briefly highlight some of the biggest challenges and give concrete suggestions for future studies:

- The visual inspection and comparison of stochastic and deterministic time series should be substituted by a more objective and automated procedure which would improve the precise detection of the transition range. This problem looks simple at first sight. However, keeping in mind that in the general case the dynamics can change qualitatively if stochastic fluctuations are taken into account it turns out to be a major challenge. Direct comparisons of the time series using squared error sums, inter-spike interval histograms etc. do not seem appropriate. One possible solution would be to compare the attractors of the deterministic solution and the different stochastic solutions instead. We already did some experiments using so-called cross-correlation integrals [Kan94] and the preliminary results look promising.

- The estimation of the transfer entropy from limited data sets should be improved. Two possible solutions for that would be a) to develop

better estimators or b) to boost the statistics. Ad a), an approach developed in [KSG04] for mutual information can possibly be extended to the transfer entropy. Ad b), multiple stochastic coupling of the enzyme process to the same experimental time series would sharpen the knowledge of the probability densities of the enzyme system (personal communication, Prof. Markus Müller, Universidad Autónoma del Estado de Morelos, Mexico).

Sophisticated estimation techniques that give reliable results even with a limited amount of data will be even more important when transfer entropy is to be applied to pure experimental data. Here, it is crucial to consider long enough processes' histories, since the exact order of the processes is not known a priori. This necessitates the estimation of probability densities in spaces of four and more dimensions.

- The transfer entropy is an averaged quantity. However, a sliding-window approach which would be required to detect changes in information transfer over time is difficult to implement because of the data-intensive estimation process. One solution to avoid this problem is the use of a "local version" of the transfer entropy (personal communication, Joe Lizier, CSIRO ICT Centre, Sidney, Australia).

- Finally, a useful future step would be to formalize the knowledge about the specific biochemical simulation methods. This knowledge could then be implemented into a modeling environment, such as Sycamore [Syc], thus guiding users intent on modeling and analyzing a specific biochemical system.

Bibliography

[ABD+04] Ander, M.; Beltrao, P.; Di Ventura, B.; Ferkinghoff-Borg, J.; Foglierini, M.; Kaplan, A.; Lemerle, C.; Tomás-Oliveira, I.; Serrano, L.: SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. In: *IEE Systems Biology*, volume 1(1):pp. 129–138, 2004. 129

[ABI+01] Alur, R.; Belta, C.; Ivančić, F.; Kumar, V.; Mintz, M.; Pappas, G.J.; Rubin, H.; Schug, J.: Hybrid modeling and simulation of biomolecular networks. In: *Lecture Notes in Computer Science, Proceedings of the 4th Int. Workshop on Hybrid Systems: Computation and Control*, volume 2034, pp. 19–32. 2001. 24, 25

[ACK06] Auger, A.; Chatelain, P.; Koumoutsakos, P.: *R*-leaping: Accelerating the stochastic simulation algorithm by reaction leaps. In: *J. Chem. Phys.*, volume 125(8):p. 084103, 2006. 19

[ACT+05] Alfonsi, A.; Cancès, E.; Turinici, G.; Di Ventura, B.; Huisinga, W.: Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. In: *ESAIM Proceedings*, volume 14, pp. 1–13. 2005. 22, 25, 28, 29

[AGH+00] Alur, R.; Grosu, R.; Hur, Y.; Kumar, V.; Lee, I.: Modular specifications of hybrid systems in CHARON. In: *Hybrid Systems: Computation and Control, Third International Workshop, LNCS*, volume 1790, pp. 6–19. 2000. 24

[AME04] Adalsteinsson, D.; McMillen, D.; Elston, T.C.: Biochemical network stochastic simulator (BioNetS): software for stochastic modeling of biochemical networks. In: *BMC Bioinformatics*, volume 5(1):p. 24, 2004. 25, 26, 33, 129

[ARM98] Arkin, A.; Ross, J.; McAdams, H.H.: Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *escherichia coli* cells. In: *Genetics*, volume 149:pp. 1633–1648, 1998. 2, 10, 11

[Ash90] Ash, R.B.: *Information Theory*. Dover Publications, New York, 1990. 74

[Bar57] Bartholomay, A.F.: *A stochastic approach to chemical reaction kinetics*. Ph.D. thesis, Harvard University, 1957. 1

[Bar58] Bartholomay, A.F.: Stochastic models for chemical reactions: I. Theory of the unimolecular reaction process. In: *Bull. Math. Biol.*, volume 20:pp. 175–190, 1958. doi:10.1007/BF02478297. 1

[BBL98] Berridge, M.J.; Bootman, M.D.; Lipp, P.: Calcium - a life and death signal. In: *Nature*, volume 395:pp. 645–648, 1998. 3, 35, 37, 55, 58

[BE04] Bentele, M.; Eils, R.: General stochastic hybrid method for the simulation of chemical reaction processes in cells. In: *Computational Methods in Systems Biology CMSB, LNCS*, pp. 248–251. Springer, Heidelberg, 2004. 26

[Ben04] Bentele, M.: *Stochastic Simulation and System Identification of large Signal Transduction Network in Cells*. Ph.D. thesis, Ruperto-Carola University of Heidelberg, 2004. 22, 25, 26

[Ber97] Berridge, M.J.: The AM and FM of calcium signalling. In: *Nature*, volume 386:pp. 759–760, 1997. 4, 39

[BFLM05] Brandman, O.; Ferrell Jr., J.E.; Li, R.; Meyer, T.: Interlinked fast and slow positive feedback loops drive reliable cell decisions. In: *Science*, volume 310:pp. 496–498, 2005. 11, 101

[Bha04] Bhalla, U.S.: Signaling in small subcellular volumes. I. Stochastic and diffusion effects on individual pathways. In: *Biophys. J.*, volume 87:pp. 733–744, 2004. 3

[BHJ03] Bundschuh, R.; Hayot, F.; Jayaprakash, C.: Fluctuations and slow variables in genetic networks. In: *Biophys. J.*, volume 84:pp. 1606–1615, 2003. 20

[Bio] BioNetS. [http://x.amath.unc.edu/BioNetS]. 26, 129

[BL99] Barkai, N.; Leibler, S.: Circadian clocks limited by noise. In: *Nature*, volume 403:pp. 267–268, 1999. 11

[BLSS91] Bartol Jr., T.M.; Land, B.R.; Salpeter, E.E.; Salpeter, M.M.: Monte Carlo simulation of miniature endplate current generation in the vertebrate neuromuscular junction. In: *Biophys. J.*, volume 59:pp. 1290–1307, 1991. 129

[BT99] Borst, A.; Theunissen, F.E.: Information theory and neural coding. In: *Nature Neuroscience*, volume 2(11):pp. 947–957, 1999. 4, 79

[BT04] Burrage, K.; Tian, T.: Poisson Runge-Kutta methods for chemical reaction systems. In: Lu, Y.Y.; Sun, W.; Tang, T., editors, *Scientific Computing and Application - Proc. Hong Kong Conf. Sci. Comput.*, pp. 82–96. Nova Publishers, 2004. 3, 19

[BTB04] Burrage, K.; Tian, T.; Burrage, P.: A multi-scaled approach for simulating chemical reaction systems. In: *Prog. Biophys. Mol. Biol.*, volume 85:pp. 217–234, 2004. 25, 27

[Buc02] Buchman, T.G.: The community of the self. In: *Nature*, volume 420:pp. 246–251, 2002. 2

[Car02] Carafoli, E.: Calcium signaling: A tale for all seasons. In: *PNAS*, volume 99(3):pp. 1115–1122, 2002. 35

[Cel] Cellware. [http://www.cellware.org]. 32, 129

[Cel96] Celio, M., editor: *Guidebook to the Calcium-Binding Proteins*. Oxford University Press, Oxford, 1996. 39, 78

[CGP05a] Cao, Y.; Gillespie, D.; .Petzold, L: Multiscale stochastic simulation algorithm with stochastic partial equlibrium assumption for chemically reacting systems. In: *J. Comput. Phys.*, volume 206(2):pp. 395–411, 2005. doi:10.1016/j.jcp.2004.12.014. 31, 49

[CGP05b] Cao, Y.; Gillespie, D.T.; Petzold, L.R.: Accelerated stochastic simulation of the stiff enzyme-substrate reaction. In: *J. Chem. Phys.*, volume 123:p. 144917, 2005. 21, 49

[CGP05c] Cao, Y.; Gillespie, D.T.; Petzold, L.R.: Avoiding negative populations in explicit Poisson tau-leaping. In: *J. Chem. Phys.*, volume 123(5):p. 054104, 2005. doi:10.1063/1.1992473. 19, 32

[CGP05d] Cao, Y.; Gillespie, D.T.; Petzold, L.R.: The slow-scale stochastic simulation algorithm. In: *J. Chem. Phys.*, volume 122(1):p. 014116, 2005. doi:10.1063/1.1824902. 31

[CGP06] Cao, Y.; Gillespie, D.T.; Petzold, L.R.: Efficient step size selection for the tau-leaping simulation method. In: *J. Chem. Phys.*, volume 124:p. 044109, 2006. doi:10.1063/1.2159468. 18

[CH96] Cohen, S.C.; Hindmarsh, A.C.: CVODE, a stiff/nonstiff ODE solver in C. In: *Computers in Physics*, volume 10(2):pp. 138–143, 1996. 30

[CL91] Cobbold, P.H.; Lee, J.A.C.: Aequorin measurements of cytoplasmic free calcium. In: McCormack, J.G.; Cobbold, P.H., editors, *Cellular Calcium: A Practical Approach*, pp. 55–81. I.R.L. Press, Oxford, 1991. 40

[CLP04] Cao, Y.; Li, H.; Petzold, L.: Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. In: *J. Chem. Phys.*, volume 121(9):pp. 4059–4067, 2004. 17

[Cop] Copasi. [`http://www.copasi.org`]. 9, 24, 32, 43, 57, 129

[CT91] Cover, T.M; Thomas, J.A.: *Elements of Information Theory*. John Wiley & Sons, New York; Chichester; Brisbane; Toronto; Singapore, 1991. 4, 74

[CVK05] Chatterjee, A.; Vlachos, D.G.; Katsoulakis, M.A.: Binomial distribution based τ-leap accelerated stochastic simulation. In: *J. Chem. Phys.*, volume 122(2):p. 024112, 2005. 19

[DCG95] Dixon, C.J.; Cobbold, P.H.; Green, A.K.: Actions of ADP, but not ATP, on cytosolic free Ca^{2+} in single rat hepatocytes mimicked by 2-methylthioATP. In: *Br. J. Pharmacol.*, volume 116:pp. 1979–1984, 1995. 40

[Del40] Delbrück, M.: Statistical fluctuations in autocatalytic reactions. In: *J. Chem. Phys*, volume 8:pp. 120–124, 1940. 1

[DHD03] Dupont, G.; Houart, G.; De Koninck, P.: Sensitivity of CaM kinase II to the frequency of Ca^{2+} oscillations: a simple model. In: *Cell Calcium*, volume 34:pp. 485–497, 2003. 4, 39, 78

[Diz] Dizzy. [`http://magnet.systemsbiology.net/software/Dizzy`]. 32, 129

[DMS+04] Dhar, P.; Meng, T.C.; Somani, S.; Ye, L.; Sairam, A.; Chitre, M.; Hao, Z.; Sakharkar, K.: Cellware - a multi-algorithmic software for computational systems biology. In: *Bioinformatics*, volume 20(8):pp. 1319–1321, 2004. 129

[DMS+05] Dhar, P.K.; Meng, T.C.; Somani, S.; Ye, L.; Sakharkar, K.; Krishnan, A.; Ridwan, A.B.M.; Wah, S.H.K.; Chitre, M.; Hao, Z.: Grid cellware: the first grid-enabled tool for modelling and simulating the cellular processes. In: *Bioinformatics*, volume 21(7):pp. 1284–1287, 2005. 129

[DS98] De Koninck, P.; Schulman, H.: Sensitivity of CaM kinase II to the frequency of Ca^{2+} oscillations. In: *Science*, volume 279:pp. 227–230, 1998. 4, 39, 78

[DSC+00] Dupont, G.; Swillens, S.; Clair, C.; Tordjmann, T.; Combettes, L.: Hierarchical organization of calcium signals in hepatocytes: from experiments to models. In: *Biochim. Biophys. Acta (BBA) - Molecular Cell Research*, volume 1498(2-3):pp. 134–152, 2000. 38

[DWCC90] Dixon, C.J.; Woods, N.M.; Cuthbertson, K.S.R.; Cobbold, P.H.: Evidence for two Ca^{2+}-mobilizing purinoreceptors on rat hepatocytes. In: *Biochem. J.*, volume 269:pp. 499–502, 1990. 39

[DXL98] Dolmetsch, R.E.; Xu, K.; Lewis, R.S.: Calcium oscillations increase the efficiency and specificity of gene expression. In: *Nature*, volume 392:pp. 933–936, 1998. 4, 37, 39, 78

[Dyn] Dynetica. [`http://www.duke.edu/~you/Dynetica_page.htm`]. 129

[E-C] E-Cell. [`http://www.e-cell.org`]. 28, 32, 129

[EB01] Endy, D.; Brent, R.: Modelling cellular behavior. In: *Nature*, volume 409:pp. 391–395, 2001. 77

[EE04] Elf, J.; Ehrenberg, M.: Spontaneous separation of bi-stable chemical systems into spatial domains of opposite phases. In: *Systems Biology*, volume 1(2):pp. 230–235, 2004. 129

[Ein05] Einstein, A.: Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen [in German]. In: *Ann. Phys.*, volume 17:pp. 549–560, 1905. 1

[ELSS02] Elowitz, M.B; Levine, A.J.; Siggia, E.D.; Swain, P.S.: Stochastic gene expression in a single cell. In: *Science*, volume 297:pp. 1183–1186, 2002. 11, 101

[ELVE05] E, W.; Liu, D.; Vanden-Eijnden, E.: Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. In: *J. Chem. Phys.*, volume 123:p. 194107, 2005. 31

[ELVE07] E, W.; Liu, D.; Vanden-Eijnden, E.: Response to "Comment on 'Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates' [J. Chem. Phys. 123, 194107 (2005)]". In: *J. Chem. Phys.*, volume 126:p. 137102, 2007. 31

[ÉT89] Érdi, P.; Tóth, J.: *Mathematical models of chemical reactions – Theory and applications of deterministic and stochastic models.* Princeton University Press, Princeton, NJ, 1989. 10, 11, 12, 77

[Fal03a] Falcke, M.: Deterministic and stochastic models of intracellular Ca^{2+} waves. In: *New Journal of Physics*, volume 5:pp. 96.1–96.28, 2003. 3, 37, 56

[Fal03b] Falcke, M.: On the role of stochastic channel behavior in intracellular Ca^{2+} dynamics. In: *Biophys. J.*, volume 84:pp. 42–56, 2003. 3, 55

[Fal04] Falcke, M.: Reading the patterns in living cells: the physics of Ca^{2+} signaling. In: *Adv. Phys.*, volume 53:pp. 255–440, 2004. 3, 55, 56

[Fel97] Fell, D.: *Understanding the Control of Metabolism.* Portland Press Ltd., London, 1997. 2

[GB00] Gibson, M.A.; Bruck, J.: Efficient exact stochastic simulation of chemical systems with many species and many channels. In: *J. Phys. Chem. A*, volume 104(9):pp. 1876–1889, 2000. 16, 18, 28

[GBD00] Gall, D.; Baus, E.; Dupont, G.: Activation of the liver glycogen phosphorylase by Ca^{2+} oscillations: a theoretical study. In: *J. Theor. Biol.*, volume 207:pp. 445–454, 2000. 4, 39, 78

[GCD95] Green, A.K.; Cobbold, P.H.; Dixon, C.J.: Cytosolic free Ca^{2+} oscillations induced by diadenosine 5',5"'-p1,p3-trisphosphate and diadenosine 5',5"'-p1,p4-tetraphosphate in single rat hepatocytes are indistinguishable from those induced by ADP and ATP respectively. In: *Biochem. J.*, volume 310:pp. 629–635, 1995. 71

[GCHL99] Gao, J.B.; Chen, C.C.; Hwang, S.K.; Liu, J.M.: Noise-induced chaos. In: *Int. J. Mod. Phys. B*, volume 13:pp. 3283–3305, 1999. 71

[GCPD05] Gunawan, R.; Cao, Y.; Petzold, L.; Doyle, F.J.: Sensitivity analysis of discrete stochastic systems. In: *Biophys. J.*, volume 88:pp. 2530–2540, 2005. 11

[GCPS06] Griffith, M.; Courtney, T.; Peccoud, J.; Sanders, W.H.: Dynamic partitioning for hybrid simulation of the bistable HIV-1 transactivation network. In: *Bioinformatics*, volume 22(22):pp. 2782–2789, 2006. 22, 25, 28, 30

[GH64] Garfinkel, D.; Hess, B.: Metabolic control mechanisms VII. A detailed computer model of the glycolytic pathway in ascites cells. In: *J. Biol. Chem.*, volume 239(4):pp. 971–983, 1964. 2

[GHG02a] Gonze, D.; Halloy, J.; Gaspard, P.: Biochemical clocks and molecular noise: Theoretical study of robustness factors. In: *J. Chem. Phys.*, volume 116(24):pp. 10997–11010, 2002. 11

[GHG02b] Gonze, D.; Halloy, J.; Goldbeter, A.: Robustness of circadian rhythms with respect to molecular noise. In: *PNAS*, volume 99:pp. 673–678, 2002. 11

[GHG04] Gonze, D.; Halloy, J.; Goldbeter, A.: Emergence of coherent oscillations in stochastic models for circadian rhythms. In: *Physica A*, volume 342:pp. 221–233, 2004. 3, 11, 55

[GHLG03] Gonze, D.; Halloy, J.; Leloup, J.-C.; Goldbeter, A.: Stochastic models for circadian rhythms: effect of molecular noise on periodic and chaotic behavior. In: *C. R. Biologies*, volume 326:pp. 189–203, 2003. 11

[Gil76] Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. In: *J. Comp. Phys.*, volume 22:pp. 403–434, 1976. 2, 13, 14, 15, 16, 48, 56, 57, 80, 96

[Gil77] Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. In: *J. Phys. Chem.*, volume 81(25):pp. 2340–2361, 1977. 2, 13, 15, 16

[Gil92] Gillespie, D.T.: *Marcov Processes – An Introduction for Physical Scientists*. Academic Press, Inc., San Diego, CA, 1992. 22, 97

[Gil00] Gillespie, D.T.: The chemical Langevin equation. In: *J. Chem. Phys.*, volume 113(1):pp. 297–306, 2000. 20

[Gil01] Gillespie, D.T.: Approximate accelerated stochastic simulation of chemically reacting systems. In: *J. Chem. Phys.*, volume 115(4):pp. 1716–1733, 2001. xiii, 3, 18, 19, 28

[Gou05] Goutsias, J.: Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. In: *J. Chem. Phys.*, volume 122:p. 184102, 2005. 31

[GP03] Gillespie, D.T.; Petzold, L.R.: Improved leap-size selection for accelerated stochastic simulation. In: *J. Chem. Phys.*, volume 119(16):pp. 8229–8234, 2003. 18

[GPC07] Gillespie, D.T.; Petzold, L.R.; Cao, Y.: Comment on "Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates" [J. Chem. Phys. 123, 194107 (2005)]. In: *J. Chem. Phys.*, volume 126:p. 137101, 2007. 31

[GTG01] Gracheva, M.E.; Toral, R.; Gunton, J.D.: Stochastic effects in intercellular calcium spiking in hepatocytes. In: *J. theor. Biol.*, volume 212:pp. 111–125, 2001. doi:10.1006/jtbi.2001.2362. 2, 3

[GTMG03] Gunton, J.D.; Toral, R.; Mirasso, C.; Gracheva, M.E.: The role of noise in some physical and biological systems. In: *Recent. Res. Devel. Applied Phys.*, volume 6:p. 497, 2003. 4, 11

[Hal99] Hallet, M.B.: The unpredictability of cellular behavior: trivial or of fundamental importance to cell biology. In: *Perspect. Biol. Med.*, volume 33(1):pp. 110–9, 1999. 11

[HC06] Harris, L.A.; Clancy, P.: A "partitioned leaping" approach for multiscale modeling of chemical reaction dynamics. In: *J. Chem. Phys.*, volume 125:p. 144107, 2006. doi:10.1063/1.2354085. 25, 31

[HDG99] Houart, G.; Dupont, G.; Goldbeter, A.: Bursting, chaos and birhythmicity originating from self-modulation of the inositol 1,4,5-trisphosphate signal in a model for intracellular Ca^{2+} oscillations. In: *Bull. Math. Biol.*, volume 61(3):pp. 507–530, 1999. 38

[Hep] HepatoSys. [`http://www.systembiologie.de`]. 7

[HFE05] Hattne, J.; Fange, D.; Elf, J.: Stochastic reaction-diffusion simulation with MesoRD. In: *Bioinformatics*, volume 21(12):pp. 2923–2924, 2005. 129

[Hin83] Hindmarsh, A.C.: ODEPACK, a systematized collection of ODE solvers. In: Stepleman, R.S.; et al., editors, *Scientific Computing, North-Holland, Amsterdam (vol. 1 of IMACS Transactions on Scientific Computation)*, pp. 55–64. North Holland, Amsterdam, 1983. 9

[HKS99] Hegger, R.; Kantz, H.; Schreiber, T.: Practical implementation of nonlinear time series methods: The TISEAN package. In: *Chaos*, volume 9(2):pp. 413–435, 1999. 57

[Hof95] Hofmeyr, J.-H.S.: Metabolic regulation: A control analytic perspective. In: *J. Bioenerg. Biomembr.*, volume 27(5):pp. 479–490, 1995. 49

[Hol04] Holzhütter, H.-G.: The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. In: *Eur. J. Biochem.*, volume 271:pp. 2905–2922, 2004. 101

[HP97] Hallett, M.B.; Pettit, E.J.: Stochastic events underlie Ca^{2+} signalling in neutrophils. In: *J. Theor. Biol.*, volume 186:pp. 1–6, 1997. 11

[HR02] Haseltine, E.L.; Rawlings, J.B.: Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. In: *J. Chem. Phys.*, volume 117(15):pp. 6959–6969, 2002. doi:10.1063/1.1505860. 21, 24, 25, 30, 47, 48

[HR05] Haseltine, E.L.; Rawlings, J.B.: On the origins of approximations for the stochastic chemical kinetics. In: *J. Chem. Phys.*, volume 123:p. 164115, 2005. 21, 31

[HRGST95] Hajnóczky, G.; Robb-Gaspers, L.D.; Seitz, M.B.; Thomas, A.P.: Decoding of cytosolic calcium oscillations in the mitochondria. In: *Cell*, volume 82(3):pp. 415–424, 1995. 4

[HS96] Heinrich, R.; Schuster, S.: *The Regulation of Cellular Systems*. Chapman and Hall, New York, NY, 1996. 2, 8, 9, 77

[HSG+06] Hoops, S.; Sahle, S.; Gauges, R.; Lee, C.; Pahle, J.; Simus, N.; Singhal, M.; Xu, L.; Mendes, P.; Kummer, U.: COPASI - a COmplex PAthway SImulator. In: *Bioinformatics*, volume 22(24):pp. 3067–3074, 2006. 33, 43, 57, 129

[Hum00] Hume, D.A.: Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. In: *Blood*, volume 96(7):pp. 2323–2328, 2000. 11

[IRW+05] Imas, O.A.; Ropella, K.M.; Ward, B.D.; Wood, J.D.; Hudetz, A.G.: Volatile anesthetics disrupt frontal-posterior recurrent information transfer at gamma frequencies in rat. In: *Neuroscience Lett.*, volume 387(3):pp. 145–50, 2005. 79

[Jar] Jarnac. [`http://sys-bio.org/software/jarnac.htm`]. 129

[Kan94] Kantz, H.: Quantifying the closeness of fractal measures. In: *Phys. Rev. E*, volume 49(6):pp. 5091–5097, 1994. 62, 103

[KBO00] Kummer, U.; Baier, G.; Olsen, L.F.: Robustness in a model for calcium signal transduction dynamics. In: Hofmeyr, J.S.; Rohwer, J.M.; Snoep, J.L., editors, *Proceedings of the BTK2000 - Animating the Cellular Map*, pp. 171–177. Stellenbosch University Press, Stellenbosch, South Africa, 2000. 71

[KE01] Kepler, T.B.; Elston, T.C.: Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. In: *Biophys. J.*, volume 81:pp. 3116–3136, 2001. 10

[KGP05] Kropp, M.; Gabbiani, F.; Prank, K.: Differential coding of humoral stimuli by timing and amplitude of intracellular calcium spike trains. In: *IEE Proc.-Syst. Biol.*, volume 152(4):pp. 263–268, 2005. 4, 5, 79

[Kho00] Kholodenko, B.N.: Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. In: *Eur. J. Biochem.*, volume 267:pp. 1583–1588, 2000. xiv, 66, 68, 103

[Kie02] Kierzek, A.M.: STOCKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm. In: *Bioinformatics*, volume 18(3):pp. 470–481, 2002. 129

[Kin] Kinetikit. [`http://www.ncbs.res.in/~bhalla/kkit`]. 29, 129

[Kit02] Kitano, H.: Computational systems biology. In: *Nature*, volume 420:pp. 206–210, 2002. 2, 23

[Kit04] Kitano, H.: Biological robustness. In: *Nat. Rev. Genet.*, volume 5(11):pp. 826–37, 2004. 11

[KKP+05] Kummer, U.; Krajnc, B.; Pahle, J.; Green, A.K.; Dixon, C.J.; Marhl, M.: Transition from stochastic to deterministic behavior in calcium oscillations. In: *Biophys. J.*, volume 89(3):pp. 1603–1611, 2005. xiii, 46, 53

[KL51] Kullback, S.; Leibler, R.A.: On information and sufficiency. In: *The Annals of Mathematical Statistics*, volume 22(1):pp. 79–86, 1951. 75

[KMS04] Kiehl, T.R.; Mattheyses, R.M.; Simmons, M.K.: Hybrid simulation of cellular behavior. In: *Bioinformatics*, volume 20(3):pp. 316–322, 2004. 25, 27

[KNK+05] Klipp, E.; Nordlander, B.; Krüger, R.; Gennemark, P.; Hohmann, S.: Integrative model of the response of yeast to osmotic shock. In: *Nature Biotechnology*, volume 23(8):pp. 975–982, 2005. 102

[KOD+00] Kummer, U.; Olsen, L.F.; Dixon, C.J.; Green, A.K.; Bornberg-Bauer, E.; Baier, G.: Switching from simple to complex oscillations in calcium signaling. In: *Biophys. J.*, volume 79:pp. 1188–1195, 2000. xvii, 38, 58, 59, 70, 79, 80, 91, 96

[Kou01] Koussoulas, N.T.: Spectral moments and the analysis of chaotic systems. In: *Int. J. Bifur. & Chaos*, volume 11(7):pp. 2051–2059, 2001. 62

[KS97] Kantz, H.; Schreiber, T.: *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, UK, 1997. 57, 99

[KS01] Keener, J.; Sneyd, J.: *Mathematical Physiology*, chapter Calcium Dynamics. Springer, 2001. 37

[KS02] Kaiser, A.; Schreiber, T.: Information transfer in continuous processes. In: *Physica D*, volume 166(1-2):pp. 43–62, 2002. 93, 98

[KSG04] Kraskov, A.; Stögbauer, H.; Grassberger, P.: Estimating mutual information. In: *Phys. Rev. E*, volume 69:p. 066138, 2004. 104

[KTO+06] Katura, T.; Tanaka, N.; Obata, A.; Sato, H.; Maki, A.: Quantitative evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics. In: *Neuroimage*, volume 31(4):pp. 1592–600, 2006. 79

[Kur72] Kurtz, T.G.: The relationship between stochastic and deterministic models for chemical reactions. In: *J. Chem. Phys.*, volume 57(7):pp. 2976–2978, 1972. 3

[Kut01] Kuthan, H.: Self-organisation and orderly processes by individual protein complexes in the bacterial cell. In: *Progr. Biophys. Mol. Biol.*, volume 75:pp. 1–17, 2001. 11

[KW93] Kraus, M.; Wolf, B.: Cytosolic calcium oscillators: critical discussion and stochastic modeling. In: *Biol. Signals*, volume 2:pp. 1–15, 1993. 55

[Lam93] Lam, S.H.: Using CSP to understand complex chemical networks. In: *Combustion Science and Technology*, volume 89(5-6):pp. 375–404, 1993. 32

[Lau00] Laurenzi, I.J.: An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics. In: *J. Chem. Phys.*, volume 113(8):pp. 3315–3322, 2000. 2, 14

[LB05] Lok, L.; Brent, R.: Automatic generation of cellular reaction networks with Moleculizer 1.0. In: *Nature Biotechnology*, volume 23(1):pp. 131–136, 2005. 129

[LG94] Lam, S.H.; Goussis, D.A.: The CSP method of simplifying kinetics. In: *Int. J. Chem. Kin.*, volume 26(4):pp. 461–486, 1994. 32

[LK03] Larsen, A.Z.; Kummer, U.: Information processing in calcium signal transduction. In: Falcke, M.; Malchow, D., editors, *Understanding Calcium Dynamics*, volume 623, pp. 153–178. Springer Verlag, Berlin, 2003. 4, 78

[LKS+01] Läer, L.; Kloppstech, M.; Schöfl, C.; Sejnowski, T.J.; Brabant, G.; Prank, K.: Noise enhanced hormonal signal transduction through intracellular calcium oscillations. In: *Biophys. Chem.*, volume 91(2):pp. 157–166, 2001. 11

[LLW+98] Li, W.-H.; Llopis, J.; Whitney, M.; Zlokarnik, G.; Tsien, R.Y.: Cell-permeant caged $InsP_3$ ester shows that Ca^{2+} spike frequency can optimize gene expression. In: *Nature*, volume 392:pp. 936–941, 1998. 4, 39, 78

[LOK04] Larsen, A.Z.; Olsen, L.F.; Kummer, U.: On the encoding and decoding of calcium signals in hepatocytes. In: *Biophys. Chem.*, volume 107(1):pp. 83–99, 2004. 4, 5, 38, 39, 59, 78, 81, 90, 91, 96

[LS01] Le Novère, N.; Shimizu, T.S.: StochSim: modelling of stochastic biomolecular processes. In: *Bioinformatics*, volume 17(6):pp. 575–576, 2001. 17, 129

[LS06] Lungarella, M.; Sporns, O.: Mapping information flow in sensorimotor networks. In: *PLOS Comp. Biol.*, volume 2(10):p. e144, 2006. 79

[MA97] McAdams, H.H.; Arkin, A.: Stochastic mechanisms in gene expression. In: *Proc. Natl. Acad. Sci. USA*, volume 94:pp. 814–819, 1997. 10

[MA99] McAdams, H.H.; Arkin, A.: It's a noisy business! In: *Trends Genet.*, volume 15(2):pp. 65–69, 1999. 10

[Mac03] MacKay, D.J.C.: *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, UK, 2003. 74

[Mad] Berkeley Madonna. [http://www.berkeleymadonna.com]. 9, 44

[MCe] MCell. [http://www.mcell.cnl.salk.edu]. 7, 129

[McQ63] McQuarrie, D.A.: Kinetics of small systems. I. In: *J. Chem. Phys.*, volume 38(2):pp. 433–436, 1963. 1

[McQ67] McQuarrie, D.A.: Stochastic approach to chemical kinetics. In: *J. Appl. Prob.*, volume 4(3):pp. 413–478, 1967. 1, 14

[Mes] MesoRD. [http://mesord.sourceforge.net]. 129

[MF98] Morton-Firth, C.J.: *Stochastic Simulation of Cell Signaling Pathways*. Ph.D. thesis, University of Cambridge, UK, 1998. 17

[MFB98] Morton-Firth, C.J.; Bray, D.: Predicting temporal fluctuations in an intracellular signalling pathway. In: *J. Theor. Biol.*, volume 192:pp. 117–128, 1998. 17, 129

[MHBH00] Marhl, M.; Haberichter, T.; Brumen, M.; Heinrich, R.: Complex calcium oscillations and the role of mitochondria and cytosolic proteins. In: *BioSystems*, volume 57:pp. 75–86, 2000. 38

[MK02] Marschinski, R.; Kantz, H.: Analysing the information flow between financial time series - an improved estimator for transfer entropy. In: *Europ. Phys. J. B*, volume 30(2):pp. 275–281, 2002. 79

[Mol] Moleculizer. [http://www.molsci.org/~lok/moleculizer]. 129

[MPS05] Marhl, M.; Perc, M.; Schuster, S.: Selective regulation of cellular processes via protein cascades acting as band-pass filters for time-limited oscillations. In: *FEBS Letters*, volume 579:pp. 5461–5465, 2005. 4, 78

[MPS06] Marhl, M.; Perc, M.; Schuster, S.: A minimal model for decoding of time-limited Ca^{2+} oscillations. In: *Biophys. Chem.*, volume 120:pp. 161–167, 2006. 4, 78

[MS03] Marhl, M.; Schuster, S.: Under what condition signal transduction pathways are highly flexible in response to external forcing? A case study on calcium oscillations. In: *J. Theor. Biol.*, volume 224:pp. 491–501, 2003. 71

[MSD04] Meng, T.C.; Somani, S.; Dhar, P.: Modeling and simulation of biological systems with stochasticity. In: *In Silico Biology*, volume 4:p. 0024, 2004. 14

[MWY07] Materassi, M.; Wernik, A.; Yordanova, E.: Determining the verse of magnetic turbulent cascades in the earth's magnetospheric cusp via transfer entropy analysis: preliminary results. In: *Nonlin. Processes Geophys.*, volume 14:pp. 153–161, 2007. 79

[NBO92] Nicotera, P.; Bellomo, G.; Orrenius, S.: Calcium-mediated mechanisms in chemically induced cell death. In: *Annu. Rev. Pharmacol. Toxicol.*, volume 32:pp. 449–470, 1992. 71

[Neo04] Neogi, N.A.: Dynamic partitioning of large discrete event biological systems for hybrid simulation and analysis. In: *Hybrid Systems: Computation and Control, HSCC 2004, LNCS 2993*, pp. 463–476. Springer, Heidelberg, 2004. 22, 25, 27, 28

[NST+06] Nichols, J.M.; Seaver, M.; Trickey, S.T.; Salvino, L.W.; Pecora, D.L.: Detecting impact damage in experimental composite structures: an information-theoretic approach. In: *Smart Mater. Struct.*, volume 15(2):pp. 424–434, 2006. 79

[Oct] Octave. [`http://www.octave.org`]. 98

[OKKP03] Olsen, L.F.; Kummer, U.; Kindzelskii, A.L.; Petty, H.R.: A model of the oscillatory metabolism of activated neutrophils. In: *Biophys. J.*, volume 84(1):pp. 69–81, 2003. 55

[OLK03] Olsen, L.F.; Lunding, A.; Kummer, U.: Mechanism of melatonin-induced oscillations in the peroxidase-oxidase reaction. In: *Arch. Biochem. Biophys.*, volume 410(2):pp. 287–295, 2003. xiv, 65, 66, 67, 103

[OM98] Oancea, E.; Meyer, T.: Protein kinase C as a molecular machine for decoding calcium and diacylglycerol signals. In: *Cell*, volume 95(3):pp. 307–318, 1998. 4, 39, 78

[OTK+02] Ozbudak, E.M.; Thattai, M.; Kurtser, I.; Grossman, A.A.; van Oudenaarden, A.: Regulation of noise in the expression of a single gene. In: *Nat. Genet.*, volume 31:pp. 69–73, 2002. 11, 101

[PAB+98] Prank, K.; Ahlvers, U.; Baumgarte, F.; Musmann, H.G.; von zur Mühlen, A.; Schöffl, C.; Brabant, G.: Stochastic model of intracellular calcium spikes. In: Greiser, E.; Wischnewsky, M., editors, *Methoden der medizinischen Informatik, Biometrie und Epidemiologie in der modernen Informationsgesellschaft*, pp. 385–388. MMV, Medizin-Verlag, München, 1998. 55

[Pah02] Pahle, J.: *Eine Hybridmethode zur Simulation biochemischer Prozesse* [in German]. Diploma thesis, ILKD, Universität Karlsruhe (TH) and EML, Heidelberg, 2002. 24, 25, 47

[Pet83] Petzold, L.: Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. In: *SIAM J. Sci. Stat. Comput.*, volume 4:pp. 136–148, 1983. 2, 43, 44, 48, 57

[Pet07] Petty, H.: Fluorescence microscopy: Established and emerging methods, experimental strategies, and applications in immunology. In: *Microsc. Res. Tech.*, volume 70(8):pp. 687–709, 2007. 3

[PGB00] Prank, K.; Gabbiani, F.; Brabant, G.: Coding efficiency and information rates in transmembrane signaling. In: *BioSystems*, volume 55:pp. 15–22, 2000. 4, 5

[PGDK08] Pahle, J.; Green, A.K.; Dixon, C.J.; Kummer, U.: Information transfer in signaling pathways: a study using coupled simulated and experimental data. In: *BMC Bioinformatics*, volume 9:p. 139, 2008. doi:10.1186/1471-2105-9-139. 73

[PGDM08] Perc, M.; Green, A.K.; Dixon, C.J.; Marhl, M.: Establishing the stochastic nature of intracellular calcium oscillations from experimental data. In: *Biophys. Chem.*, volume 132:pp. 33–38, 2008. 4

[PK04] Puchalka, J.; Kierzek, A.M.: Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. In: *Biophys. J.*, volume 86(3):pp. 1357–1372, 2004. 25, 28

[PLvzM+98] Prank, K.; Läer, L.; von zur Mühlen, A.; Brabant, G.; Schöfl, C.: Decoding of intracellular calcium spike trains. In: *Europhys. Lett.*, volume 42(2):pp. 143–147, 1998. 4, 5, 39, 79, 88, 90

[PM03a] Perc, M.; Marhl, M.: Noise enhances robustness of intracellular Ca^{2+} oscillations. In: *Physics Letters A*, volume 316:pp. 304–310, 2003. 4, 11

[PM03b] Perc, M.; Marhl, M.: Sensitivity and flexibility of regular and chaotic calcium oscillations. In: *Biophys. Chem.*, volume 104:pp. 509–522, 2003. 71

[PSL+98] Prank, K.; Schöfl, C.; Läer, L.; Wagner, M.; von zur Mühlen, A.; Brabant, G.; Gabbiani, F.: Coding of time-varying hormonal signals in intracellular calcium spike trains. In: *Pac. Symp. Biocomput.*, pp. 633–44, 1998. 4, 5, 79

[PSvzM+98] Paluš, M.; Schöfl, C.; von zur Mühlen, A.; Brabant, G.; Prank, K.: Coarse-grained entropy rates quantify fast Ca^{2+} dynamics modulated by pharmacological stimulation. In: *Pac. Symp. Biocomput.*, pp. 645–56, 1998. 4

[PTVF92] Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P.: *Numerical Recipes in C: The Art of Scientific Computing, 2nd ed.* Cambridge University Press, Cambridge, 1992. 9

[QSE02] Qian, H.; Saffarin, S.; Elson, E.L.: Concentration fluctuations in a mesoscopic oscillating chemical reaction system. In: *Proc. Natl. Acad. Sci. USA*, volume 99(16):pp. 10376–10381, 2002. 10

[RA03] Rao, C.V.; Arkin, A.P.: Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. In: *J. Chem. Phys.*, volume 118(11):pp. 4999–5010, 2003. 3, 21, 49

[RBH94] Ross, I.L.; Browne, C.M.; Hume, D.A.: Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. In: *Immunol. Cell. Biol.*, volume 72(2):pp. 177–185, 1994. doi:10.1038/icb.1994.26. 11

[RCD93] Rosenstein, M.T.; Collins, J.J.; De Luca, C.J.: A practical method for calculating largest Lyapunov exponents from small data sets. In: *Physica D*, volume 65(1-2):pp. 117–134, 1993. 57

[RJ03] Rozi, A.; Jia, Y.: A theoretical study of effects of cytosolic Ca^{2+} oscillations on activation of glycogen phosphorylase. In: *Biophys. Chem.*, volume 106:pp. 193–202, 2003. 4, 78

[ROB05] Ramsey, S.; Orrell, D.; Bolouri, H.: Dizzy: stochastic simulation of large-scale genetic regulatory networks. In: *J. Bioinform. Comput. Biol.*, volume 3(2):pp. 415–36, 2005. 129

[Ros63] Rosenbrock, H.H.: Some general implicit processes for the numerical solution of differential equations. In: *The Computer Journal*, volume 5(4):pp. 329–330, 1963. doi:doi:10.1093/comjnl/5.4.329. 2, 57

[RPCG03] Rathinam, M.; Petzold, L.R.; Cao, Y.; Gillespie, D.T.: Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. In: *J. Chem. Phys.*, volume 119(24):pp. 12784–12794, 2003. doi:10.1063/1.1627296. 19

[RWA02] Rao, C.V.; Wolf, D.W.; Arkin, A.P.: Control, exploitation and tolerance of intracellular noise. In: *Nature*, volume 420:pp. 231–237, 2002. 2, 3, 10, 11, 12, 23, 88, 101

[RWD01] Resat, H.; Wiley, H.S.; Dixon, D.A.: Probability-weighted dynamic Monte Carlo method for reaction kinetics simulations. In: *J. Phys. Chem. B*, volume 105:pp. 11026–11034, 2001. doi:10.1021/jp011404w. 3, 17

[SA06] Samoilov, M.S.; Arkin, A.P.: Deviant effects in molecular reaction pathways. In: *Nature Biotechnology*, volume 24(10):pp. 1235–1240, 2006. 10

[SAB+07] Savage, V.M.; Allen, A.P.; Brown, J.H.; Gillooly, J.F.; Herman, A.B.; Woodruff, W.H.; West, G.B.: Scaling of number, size, and metabolic rate of cells with body size in mammals. In: *PNAS*, volume 104(11):pp. 4718–4723, 2007. 90

[Sav76] Savageau, M.A.: *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology.* Addison-Wesley, Reading, MA, 1976. 2

[SBM] SBML. [http://www.sbml.org]. 9, 44

[Sch00] Schreiber, T.: Measuring information transfer. In: *Phys. Rev. Lett.*, volume 85(2):pp. 461–4, 2000. 76, 77, 79, 90, 91, 94

[SGP+06] Sahle, S.; Gauges, R.; Pahle, J.; Simus, N.; Kummer, U.; Hoops, S.; Lee, C.; Singhal, M.; Xu, L.; Mendes, P.: Simulation of biochemical networks using COPASI: A complex pathway simulator. In: *Proc. of the Winter Simulation Conference WSC 2006, Monterey, California, USA, December 3-6, 2006*, pp. 1698–1706. WSC, 2006. 43

[Sha48] Shannon, C.E.: A mathematical theory of communication. In: *Bell System Technical Journal*, volume 27:pp. 379–423,623–656, 1948. 4, 74

[Shi02] Shimizu, T.S.: *Cell Signalling Pathways - A Computer-Based Study*. Ph.D. thesis, University of Cambridge, Cambridge, UK, 2002. 17

[Sil86] Silverman, B.W.: *Density estimation for statistics and data analysis*. Chapman & Hall/CRC, London, 1986. 92, 94, 98

[SJ02] Shuai, J.-W.; Jung, P.: Stochastic properties of Ca^{2+} release of inositol 1,4,5-trisphosphate receptor clusters. In: *Biophys. J.*, volume 83:pp. 87–97, 2002. 3, 56

[SK05a] Salis, H.; Kaznessis, Y.: Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. In: *J. Chem. Phys.*, volume 122(5):p. 054103, 2005. 25, 28, 29

[SK05b] Salis, H.; Kaznessis, Y.N.: An equation-free probabilistic steady-state approximation: Dynamic application to the stochastic simulation of biochemical reaction networks. In: *J. Chem. Phys.*, volume 123:p. 214106, 2005. doi:10.1063/1.2131050. 31

[SKD+02] Steuer, R.; Kurths, J.; Daub, C.O.; Weise, J.; Selbig, J.: The mutual information: Detecting and evaluating dependencies between variables. In: *Bioinformatics*, volume 18(Suppl. 2):pp. S231–S240, 2002. 91

[SKM05] Schuster, S.; Knoke, B.; Marhl, M.: Differential regulation of proteins by bursting calcium oscillations - a theoretical study.

In: *BioSystems*, volume 81:pp. 49–63, 2005. xiii, 4, 39, 40, 78, 90

[SL95] Shen, P.; Larter, R.: Chaos in intracellular Ca^{2+} oscillations in a new model for non-excitable cells. In: *Cell Calcium*, volume 17:pp. 225–232, 1995. 38, 65

[Sma] SmartCell. [`http://smartcell.embl.de/`]. 7, 32, 129

[SMH02] Schuster, S.; Marhl, M.; Höfer, T.: Modelling of simple and complex calcium oscillations - From single-cell responses to intercellular signalling. In: *Eur. J. Biochem.*, volume 269:pp. 1333–1355, 2002. 4, 38

[SOV07] Samant, A.; Ogunnaike, B.A.; Vlachos, D.G.: A hybrid multiscale Monte Carlo algorithm (HyMSMC) to cope with disparity in time scales and species populations in intracellular networks. In: *BMC Bioinformatics*, volume 8:p. 175, 2007. 32

[SPA05] Samoilov, M.; Plyasunov, S.; Arkin, A.P.: Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. In: *PNAS*, volume 102(7):pp. 2310–2315, 2005. 10, 101

[SPH04] Salazar, C.; Politi, A.; Höfer, T.: Decoding of calcium oscillations by phosphorylation cycles. In: Mamitsuka, H.; Smith, T.; Holzhütter, H.G.; Kanehisa, M.; DeLisi, C.; Heinrich, R.; Miyano, S., editors, *Proceedings of Fourth International Workshop on Bioinformatics and Systems Biology*, pp. 50–51. Kyoto, 2004. 4, 39, 78

[SPH08] Salazar, C.; Politi, A.Z.; Höfer, T.: Decoding of calcium oscillations by phosphorylation cycles: Analytic results. In: *Biophys. J.*, volume 94(4):p. 1203, 2008. doi:10.1529/biophysj.107.113084. 4, 39, 78

[SS91] Somogyi, R.; Stucki, J.W.: Hormone-induced calcium oscillations in liver cells can be explained by a simple one pool model. In: *J. Biol. Chem.*, volume 266(17):pp. 11068–11077, 1991. 37

[SS00] Schreiber, T.; Schmitz, A.: Surrogate time series. In: *Physica D*, volume 142(3-4):pp. 346–382, 2000. 82

[SSK06] Salis, H.; Sotiropoulus, V.; Kaznessis, Y.N.: Multiscale Hy3S: Hybrid stochastic simulation for supercomputers. In: *BMC Bioinformatics*, volume 7:p. 93, 2006. 22, 25, 29, 33

[SSPK06] Surovtsova, I.; Sahle, S.; Pahle, J.; Kummer, U.: Approaches to complexity reduction in a systems biology research environment (Sycamore). In: *Proc. of the Winter Simulation Conference WSC 2006, Monterey, California, USA, December 3-6, 2006*, pp. 1683–1689. WSC, 2006. 43

[Stoa] Stochastic Model Test Suite. [`http://www.calibayes.ncl.ac.uk/Resources/dsmts`]. 47

[Stob] Stochastirator. [`http://opnsrcbio.molsci.org/stochastirator/stoch-main.html`]. 129

[Stoc] StochSim. [`http://www.pdn.cam.ac.uk/groups/comp-cell/StochSim.html`]. 129

[Stod] Stocks. [`http://www.ibb.waw.pl/stocks/`]. 129

[Stoe] Stode. [`http://projects.villa-bosch.de/bcb/software/Carel`]. 57, 129

[SV05] Samant, A.; Vlachos, D.G.: Overcoming stiffness from stochastic simulation stemming from partial equilibrium: a multiscale Monte Carlo algorithm. In: *J. Chem. Phys.*, volume 123(14):p. 144114, 2005. doi:10.1063/1.2046628. 31

[Syc] Sycamore. [`http://sycamore.eml.org`]. 104

[TAT05] Takahashi, K.; Arjunan, S.N.V.; Tomita, M.: Space in systems biology of signaling pathways – towards intracellular molecular crowding in silico. In: *FEBS Letters*, volume 579:pp. 1783–1788, 2005. 3

[TB04] Tian, T.; Burrage, K.: Binomial leap methods for simulating stochastic chemical kinetics. In: *J. Chem. Phys.*, volume 121(21):pp. 10356–10364, 2004. 19

[THT+99] Tomita, M.; Hashimoto, K.; Takahashi, K.; Shimizu, T.S.; Matsuzaki, Y.; Miyoshi, F.; Saito, K.; Tanida, S.; Yugi, K.; Venter, J.C.; Hutchison III, C.A.: E-CELL: software environment for whole-cell simulation. In: *Bioinformatics*, volume 15(1):pp. 72–84, 1999. 129

[TKHT04] Takahashi, K.; Kaizu, K.; Hu, B.; Tomita, M.: A multi-algorithm, multi-timescale method for cell simulation. In: *Bioinformatics*, volume 20(4):pp. 538–546, 2004. 25, 28

[TRR91] Thomas, A.P.; Renard, D.C.; Rooney, T.A.: Spatial and temporal organization of calcium signalling in hepatocytes. In: *Cell Calcium*, volume 12:pp. 111–126, 1991. 2

[TSB04] Turner, T.E.; Schnell, S.; Burrage, K.: Stochastic approaches for modelling in vivo reactions. In: *Comp. Bio. Chem.*, volume 28:pp. 165–178, 2004. 14

[VB04] Vasudeva, K.; Bhalla, U.S.: Adaptive stochastic-deterministic chemical kinetic simulations. In: *Bioinformatics*, volume 20(1):pp. 78–84, 2004. 25, 29, 129

[vGK01] van Gend, C.; Kummer, U.: STODE - automatic stochastic simulation of systems described by differential equations. In: Yi; Hucka; Morohashi; Kitano, editors, *Proceedings of the Second International Conference on Systems Biology ICSB*, pp. 326–333. 2001. 49, 129

[VKBL02] Vilar, J.M.G.; Kueh, H.Y.; Barkai, N.; Leibler, S.: Mechanisms of noise-resistance in genetic oscillators. In: *PNAS*, volume 99(9):pp. 5988–5992, 2002. 11

[WBH04] Wang, L.; Birol, I.; Hatzimanikatis, V.: Metabolic control analysis under uncertainty: Framework development and case studies. In: *Biophys. J.*, volume 87:pp. 3750–3763, 2004. 11

[WCC86] Woods, N.M.; Cuthbertson, K.S.R.; Cobbold, P.H.: Repetitive transient rises in cytoplasmic free calcium in hormone-stimulated hepatocytes. In: *Nature*, volume 319:pp. 600–602, 1986. 4, 36, 39

[WMP06] Wagner, H.; Möller, M.; Prank, K.: COAST: Controllable approximative stochastic reaction algorithm. In: *J. Chem. Phys.*, volume 125(17):p. 174104, 2006. doi:10.1063/1.2361284. 25, 31

[WS49] Weaver, W.; Shannon, C.E.: *The Mathematical Theory of Communication*. University of Illinois Press, Urbana and Chicago, IL, 1949. 77, 79

[WSSV85] Wolf, A.; Swift, J.B.; Swinney, H.L.; Vastano, J.A.: Determining Lyapunov exponents from a time series. In: *Physica D*, volume 16:pp. 285–317, 1985. 57

[XPP] XPP-Aut. [`http://www.math.pitt.edu/~bard/xpp/xpp.html`]. 9, 44

[YHY03] You, L.; Hoonlor, A.; Yin, J.: Modeling biological systems using Dynetica - a simulator of dynamic networks. In: *Bioinformatics*, volume 19(3):pp. 435–436, 2003. 129

Appendix A

List of software systems for biochemical systems capable of stochastic simulation. The type of the simulation engines implemented and corresponding references are given.

BioNetS Gillespie, Langevin Eq., Hybrid Solver [Bio; AME04]

Cellware Gillespie, Gibson, τ-Leap (explicit), Langevin-type Eq. with Poisson noise-term [Cel; DMS$^+$04; DMS$^+$05]

Copasi LSODA, Gibson, Hybrid Solver [Cop; HSG$^+$06]

Dizzy Gillespie, Gibson, τ-Leap [Diz; ROB05]

Dynetica Gillespie [Dyn; YHY03]

E-Cell Gillespie, Gibson, τ-Leap (implicit and explicit), Langevin Eq., Hybrid Solver [E-C; THT$^+$99]

Jarnac Gillespie [Jar]

Kinetikit Hybrid Solver (Vasudeva/Bhalla) [Kin; VB04]

MCell Spatial stochastic simulation [MCe; BLSS91]

MesoRD Spatial stochastic simulation [Mes; HFE05; EE04]

Moleculizer Spatial stochastic simulation [Mol; LB05]

SmartCell Gillespie, τ-Leap, Spatial stochastic simulation [Sma; ABD$^+$04]

Stochastirator Gibson [Stob]

StochSim Stochastic mesoscopic approach [Stoc; LS01; MFB98]

Stocks Gillespie (lin. growing vol. & and cell division possible) [Stod; Kie02]

Stode Gillespie [Stoe; vGK01]

Acknowledgment

First and foremost, I would like to thank Prof. Ursula Kummer for her kind support and for providing me with the opportunity to pursue my own ideas.

I am also very grateful to Prof. Thomas Höfer for supporting my work.

The warmest thanks to Prof. Hermann-Georg Holzhütter for his willingness to act as referee.

I very much appreciated the great atmosphere in the working group. Thanks to all current and former group members, in particular, Ralph Gauges, Dr. Sven Sahle, Dr. Jürgen Zobeley, Dr. Ursula Rost, Dr. Katja Wegner, Dr. Femke Mensonides, Dr. Iulian Stoleriu, Dr. Carel van Gend, Dr. Natalia Simus, Dr. Irina Surovtsova, Tim Johann, Dr. Andreas Weidemann, and Jocelyn Faberman. Also, many thanks to all other colleagues at the EML Research gGmbH, Heidelberg.

Of course, I would like to thank my external collaborators, Prof. Pedro Mendes, Dr. Stefan Hoops, Prof. Marko Marhl, Borut Krajnc, Dr. C. Jane Dixon, and Prof. Anne K. Green.

I am most fortunate to have been given all the encouragement I needed by Katrin Hübner and my family. Thank you!

Finally, I gratefully acknowledge funding by the German Federal Ministry of Education and Research (BMBF) and the Klaus Tschira Foundation, Heidelberg.

Publications

Articles

J. Pahle, A.K. Green, C.J. Dixon, and U. Kummer. Information transfer in signaling pathways: a study using coupled simulated and experimental data. In: *BMC Bioinformatics*, volume 9:p. 139, 2008.

S. Sahle, R. Gauges, J. Pahle, N. Simus, U. Kummer, S. Hoops, C. Lee, M. Singhal, L. Xu, and P. Mendes. Simulation of biochemical networks using COPASI: A complex pathway simulator. In *Proc. of the Winter Simulation Conference WSC 2006, Monterey, California, USA, December 3-6, 2006*, pages 1698–1706. WSC, 2006.

I. Surovtsova, S. Sahle, J. Pahle, and U. Kummer. Approaches to complexity reduction in a systems biology research environment (Sycamore). In *Proc. of the Winter Simulation Conference WSC 2006, Monterey, California, USA, December 3-6, 2006*, pages 1683–1689. WSC, 2006.

S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI - a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.

U. Kummer, B. Krajnc, J. Pahle, A.K. Green, C.J. Dixon, and M. Marhl. Transition from stochastic to deterministic behavior in calcium oscillations. *Biophys. J.*, 89(3):1603–1611, 2005.

J. Pahle, I. Valova, G. Georgiev, and N. Georgieva. Oscillatory simulation of mitral and granule cell behavior in the olfactory bulb. In H.R. Arabnia, editor, *Proc. of the International Conference on Artificial Intelligence (IC-AI'2000)*, pages 483–489, Las Vegas, NV, USA, 2000. CSREA Press.

Coeditorship

U. Kummer, J. Pahle, I. Surovtsova, and J. Zobeley, editors. *Proceedings of the 4th Workshop on Computation of Biochemical Pathways and Genetic Networks*, Berlin, 2005. Logos-Verlag.

R. Gauges, U. Kummer, J. Pahle, and U. Rost, editors. *Proceedings of the 3rd Workshop on Computation of Biochemical Pathways and Genetic Networks*, Berlin, 2003. Logos-Verlag.

Invited talks

Heidelberger Life Science Lab, Butenschoen-Haus, Landau "Einführung in biochemische Simulationsmethoden", September 29, 2007

Biochemisches Kolloquium an der Universität zu Köln "Copasi – A Complex Pathway Simulator", October 23, 2006

Seminar am Berlin Center for Genome Based Bioinformatics "Biochemical Simulations: Which method to use?", November 30, 2005

Oberseminar Bioinformatik der Gruppen Backofen, Dittrich und Schuster an der Universität Jena "Stochastic and Hybrid Simulation Methods in Biochemistry", December 9, 2004

Hauptseminar Verkehrsökonometrie und -modellierung an der Technischen Universität Dresden "Stochastische und hybride Simulationsmethoden in der Biochemie", July 14, 2004

GK-120 Symposium Modelling and Simulation of Cellular Processes an der Freien Universität Berlin "Simulation of biochemical networks – deterministic, stochastic and hybrid approaches", April 12, 2003

Other talks and tutorials

ECCS '07 in Dresden (talk) "Information transfer in calcium signal transduction", October 1, 2007

ICSB 2006 in Yokohama (tutorials) "Modeling, simulating, and analyzing biochemical systems with Copasi" and "Advanced model analysis with Copasi", October 8, 2006

BMBF Platform Meeting Modeling & Bioinformatics (talk) in Berlin "Hybrid methods for the simulation of biochemical networks", January 27, 2005

Posters

J. Pahle, A. Politi, C. Salazar, R. Nitschke, U. Kummer and T. Höfer "Generation and decoding of hormonal calcium signals in hepatocytes", Status seminar HepatoSys, Heidelberg, Germany, October 15/16, 2007

J. Pahle and U. Kummer "Information transfer in calcium signal transduction". Biosystems Modeling Workshop, SAMSI, Research Triangle Park, NC, USA, March 5-7, 2007

J. Pahle and U. Kummer and S. Sahle "Hybrid simulation of biochemical reaction networks, Status seminar HepatoSys, Heidelberg, Germany, July 12, 2006

J. Pahle and S. Sahle and U. Kummer "Dynamic behavior of buffered calcium ions in stochastic and deterministic simulations" EMBO Workshop on Principles of Self-Organization in Living Matter, Heidelberg, Germany, June 2-5, 2006

J. Pahle and U. Kummer and S. Sahle "Hybrid simulation of biochemical reaction networks", HepatoSys Evaluation meeting, Berlin, Germany, November 29, 2005

J. Pahle and U. Kummer and B. Krajnc and M. Marhl "Transition from Stochastic to Deterministic Behavior in Biochemical Systems" ISMB/ECCB 2004, Glasgow, UK, July 31 - August 4, 2004

Berlin, December 17, 2007

Selbständigkeitserklärung

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst habe. Es wurden ausschließlich die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht.

Berlin, 17. Dezember, 2007

Jürgen Pahle