Lecture: Computational Systems Biology
Universität des Saarlandes, SS 2012

09 Stochastic modelling

Dr. Jürgen Pahle

26.6.2012

Recap

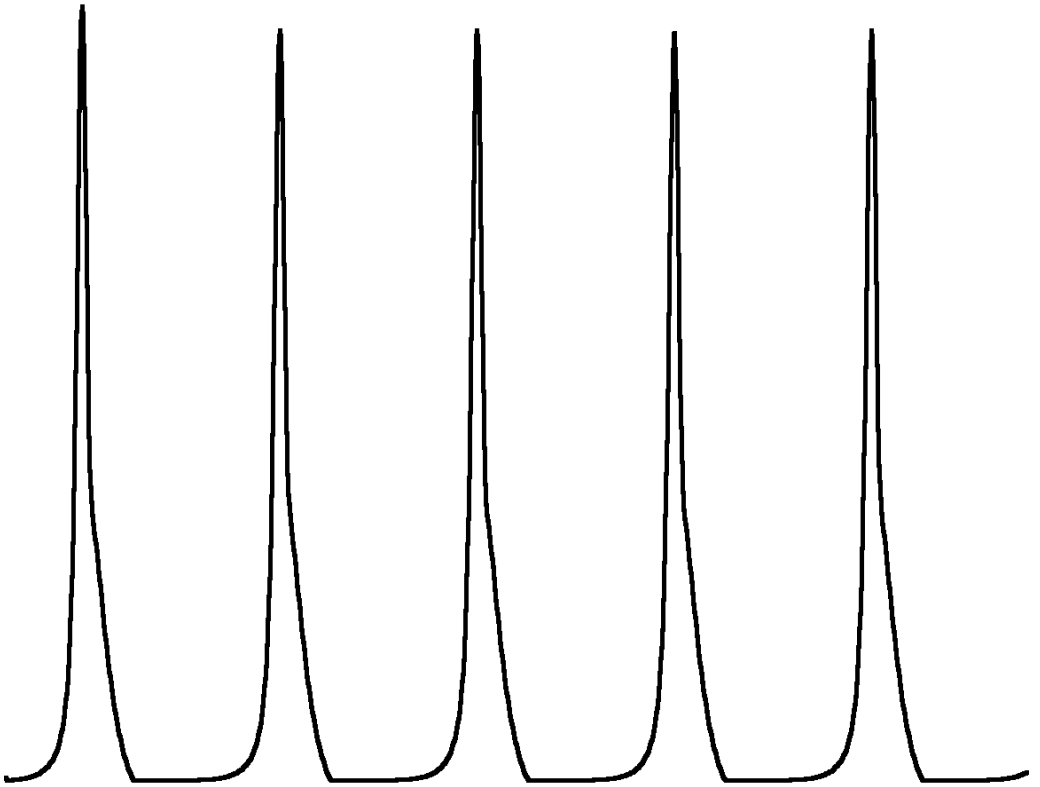
- General workflow of modelling biochemical networks
- Different types of biochemical networks → different experimental data and computational analysis methods
- Main types of systems:
  - Metabolism, e.g. glycolysis (catabolic) or amino acid synthesis (anabolic)
  - Signal transduction pathways, e.g. MAPK cascades, NF-κB or Calcium signalling
  - Gene expression networks, e.g. cell cycle

Mathematical models

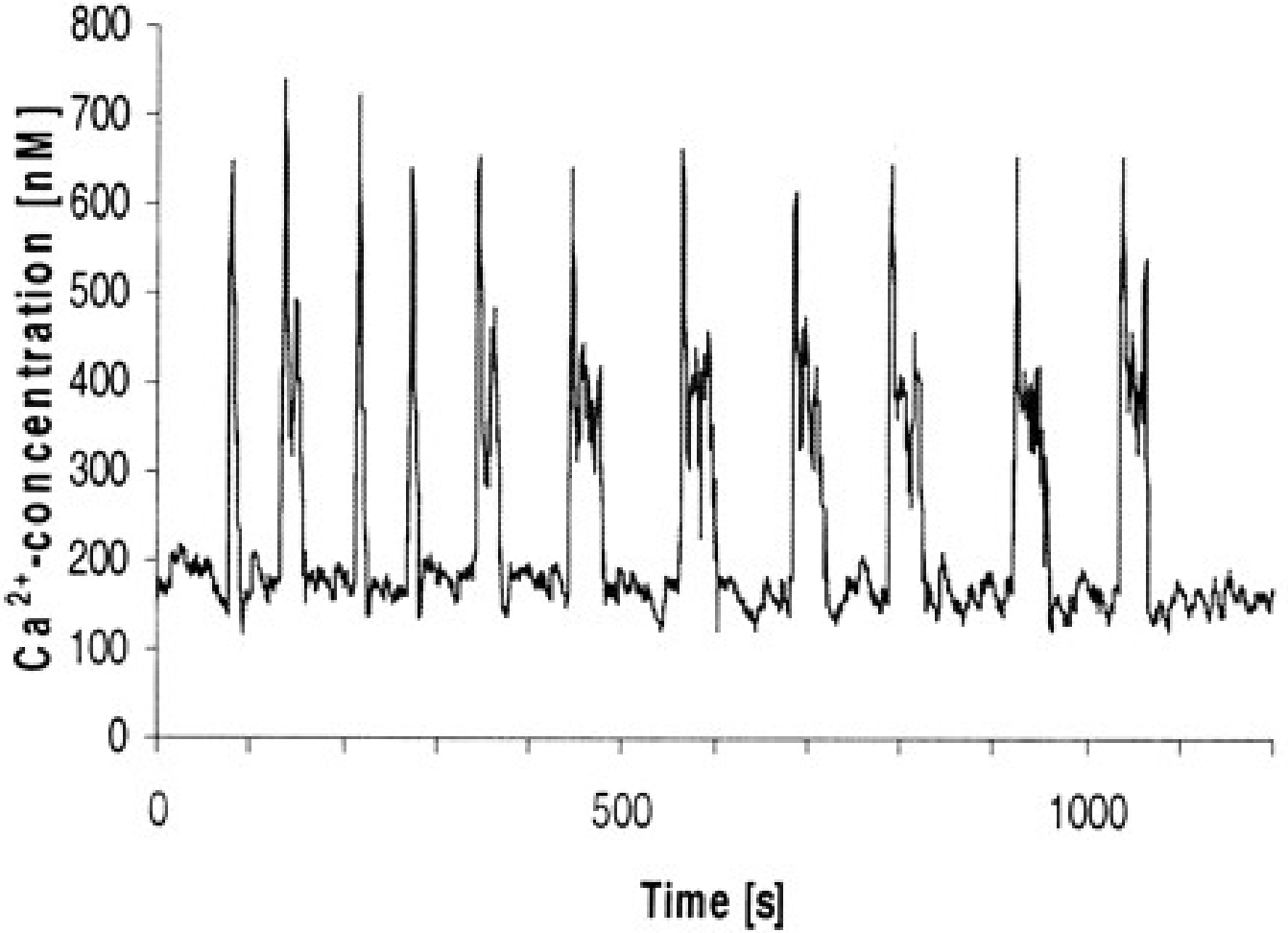
- Different levels of detail:
  - **microscopic** models: only a few particles and the corresponding forces are simulated (molecular dynamics, ligand binding), computationally expensive !!!
  - **mesoscopic** models: single particles are distinguishable, but acting forces and positions of the particles are neglected
  - **macroscopic** models: particles of one type are grouped together, only the particle numbers (or the concentrations) are considered, systems are assumed homogeneous
- Macroscopic models:
  - **deterministic** models: ordinary differential equation systems
  - **stochastic** models: the system is modeled as a random process
  - **hybrid** models: mix of deterministic and stochastic elements

Calcium dynamics (simulated deterministically)

Spiking



Experimental time series

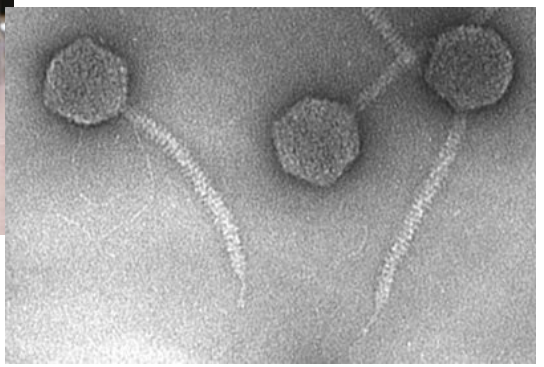


Bursting calcium conc. oscillations in single rat hepatocytes stimulated with ATP (1.2 μM). (taken from: Kummer et al. (2000), Biophys. J., 79 (3), pp. 1188-1195)

Reasons for stochastic modeling

- Small particle numbers on single cell level (e.g. signal transduction, gene expression)
  → discreteness of the system, random fluctuations
- Bi-stable systems:

Calico cat



λ phage

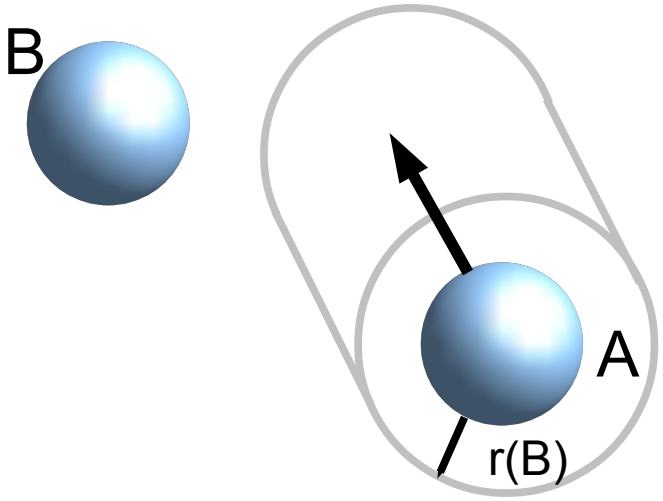
- Stochasticity as an important property of the system:
  noise-sustained oscillations, stochastic resonance, etc.
- Extinction of species
- Rare events

images: commons.wikimedia.org

ODE modelling

- Future (and history) of a system modelled with ODEs is uniquely defined (no two different trajectories can cross) → from one initial condition the system always reaches the same attractor (steady state, limit cycle, etc.)
- Based on continously-valued variables (concentrations)
- Neglects fluctuations in molecular numbers due to stochastic timing of reactive events

Basis of the Stochastic Approaches



$$a_\mu(x)\cdot dt = c_\mu \cdot h_\mu(x)\cdot dt$$

specific probabilistic reaction rate

product of

probability of collision
(~ average relative speed * collision cross-section area / volume) and

probability of reaction after collision
(collision energy larger than threshold)

number of different combinations of substrate particles

Chemical Master Equation (CME)

$$\frac{\partial P(x,t|x_0,t_0)}{\partial t} = \sum_{j=1}^{M} \Big[a_j(x-\nu_j) * P(x-\nu_j,t|x_0,t_0) - a_j(x) * P(x,t|x_0,t_0) \Big]$$

"probability flux" to x from other states

"probability flux" from x to other states

- ν_j is stoichiometric vector of reaction j
- More important for the simulation methods is the so-called Reaction Probability Density Function
  - When will the next reaction take place?
  - Which reaction will it be?

$$P(\tau,\mu) = \begin{cases} a_\mu \exp(-a_0 \tau) & if\ 0 \leqslant \tau < \infty \wedge \mu = 1,\ldots,M \\ 0 & otherwise \end{cases}$$

How to derive the Reaction Probability Density Function...

D.T. Gillespie (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**:403-434

Stochastic Simulation (Gillespie 1976)

1) Calculate probabilities for all reactions
2) Calculate stochastic time step t (exponentially distributed, sum of all reaction prob.)

$$\tau = \frac{1}{a_0} \ln(r_1)$$

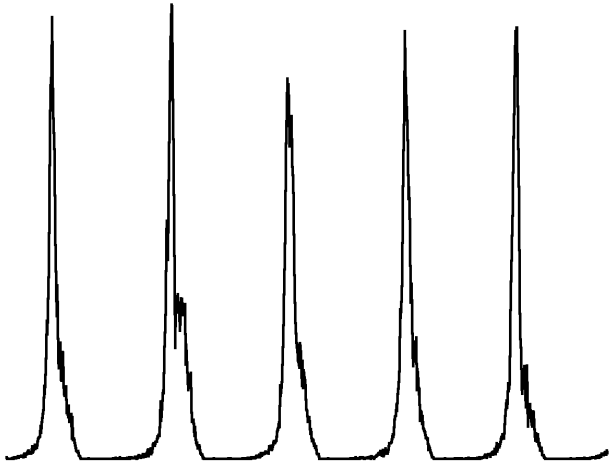
3) Monte Carlo Simulation: The reaction to be realized is chosen by “playing roulette”, discrete distribution

$$\sum_{\alpha=1}^{\mu-1} \frac{a_\alpha}{a_0} \leqslant r_2 \leqslant \sum_{\alpha=1}^{\mu} \frac{a_\alpha}{a_0}$$

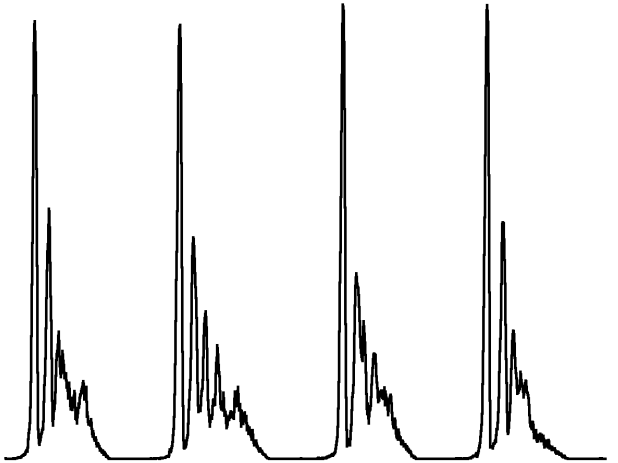
4) Instantiate the reaction: Change particle numbers according to stoichiometry

Calcium dynamics (simulated stochastically)

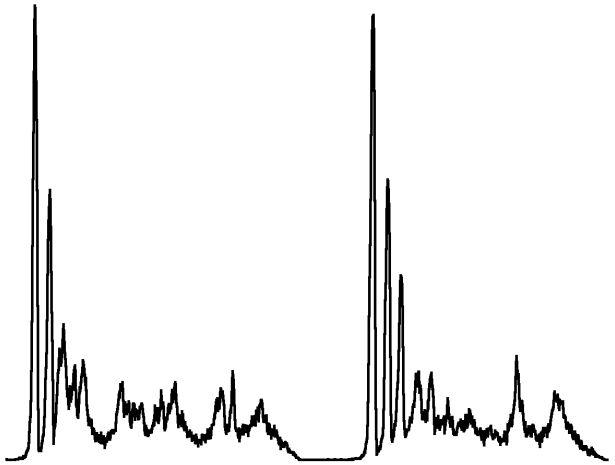
spiking



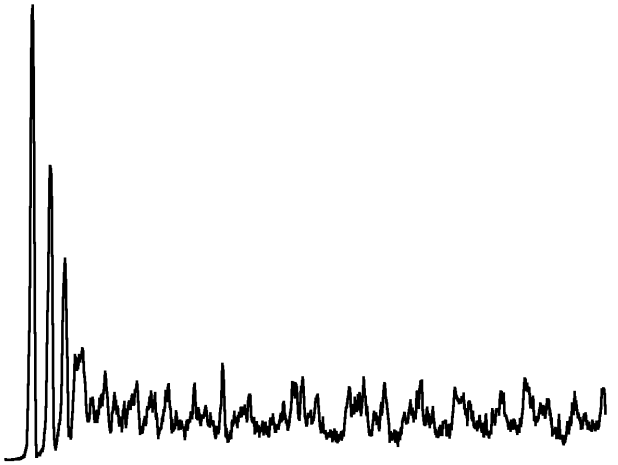
bursting



irregular/chaotic



overstimulation



"Exact" stochastic simulation methods

- **Direct Method** (Gillespie 1976)
  Reaction time and type of next reaction event are computed separately
- **First Reaction Method** (Gillespie 1976)
  Calculation of putative reaction times for all reactions, realization of the reaction with the shortest reaction time
- **Next Reaction Method** (Gibson und Bruck 2000)
  Extension of the First Reaction Method, makes clever use of data structures (priority queue, dependency graph),
  Complexity reduction, only 1 random number per iteration
- **Optimized Direct Method** (Cao et al. 2004)
  Efficient implementation of the Direct Method
  Dependency graph and sorting of the reactions
- **Constant-time Method** (Slepoy et al. 2008)
  Reaction selection in constant time

Problem of the Stochastic Approach

Each single reaction event has to be calculated →
Run time depends on particle numbers →

Simulation of bigger systems is **very slow**

Example: Simulation of Ca^{2+}-oscillations
Simulated time: 1000 s
Particle numbers: ~ 10M
PC 1.8 GHz

→ Run time ~ 5 days

other issues: reversible reactions, higher-order kinetics

$$Y_s^{(1)} \simeq 1000.7 \pm 0.9 \quad \text{and} \quad \Delta_s^2 \simeq 762 \pm 35 \qquad (37)$$

where again the ± uncertainties represent 95% confidence limits. Our result for Δ_s^2 is obviously consistent with the theoretically predicted value of 750.

The computer programs used to obtain the statistical estimates in (34) and (37) are easy to write, especially since they require no periodic write-outs for subsequent plotting. However, in performing this type of calculation, care must be taken to make each run long enough to wipe out effects due to the choice of the initial state. Moreover, as is typical of Monte Carlo estimates of averages, if one wishes to reduce the uncertainties by $1/f$, one must make f^2 times as many runs. These two features can conspire to make such calculations rather time consuming on a computer, and hence rather expensive; for example, it cost about $160 to obtain the figures quoted in (37). The chief advantage of the method is that it is universally applicable, and does not require complicated or specialized analytical techniques. The method is therefore best reserved for cases in which an analytical calculation either cannot be made at all, or else is of questionable validity because of various approximations employed.

IVB. *The Lotka Reactions.* In 1920, Lotka observed that the set of coupled, autocatalytic reactions[20]

$$\overline{X} + Y_1 \xrightarrow{c_1} 2Y_1 \qquad (38a)$$

$$Y_1 + Y_2 \xrightarrow{c_2} 2Y_2$$

$$Y_2 \xrightarrow{c_3} Z$$

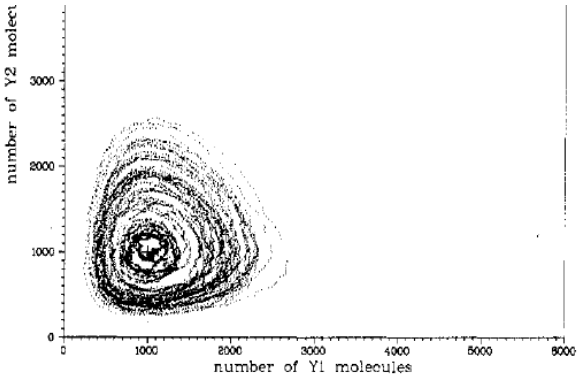


Figure 12. Results of a 1.5×10^6 reaction stochastic simulation run of the Lotka reactions *including* reaction 38d, with $c_1 = 0.0002$, $c_2 = 0.01$, $c_3 = c_4 = 10$, and $X = 10^5$ (constant), corresponding to the deterministic steady-state $Y_1 = Y_2 = 1000$. (a) A 200 rpd plot of Y_1 vs. t for $0 \leq t \leq 30$. (b) A 200 rpd plot of Y_2 vs. Y_1 for $0 \leq t \leq 30$.

of Y_2 through natural causes.

A steady state system would be characterized by the condition

$$dY_1/dt = dY_2/dt = 0$$

and it is easy to show from (39) that this condition is satisfied when

$$Y_1 = Y_{1s} \equiv c_3/c_2 \quad \text{and} \quad Y_2 = Y_{2s} \equiv c_1 X/c_2 \qquad (40)$$

That is, if we start out at $t = 0$ with $Y_1 = Y_{1s}$ and $Y_2 = Y_{2s}$, then the deterministic reaction-rate equations predict that this situation will persist indefinitely. However, as we shall

The simulation run from which the plots in Figures 8a–c were constructed contained a total of 1×10^6 reactions. The cost of compiling and executing the simulation program for this run on the Univac 1110 computer was about $40, and the subsequent plotting programs averaged about $9 per graph.

Gillespie (1977) Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81(25):2340-61

Other problems

- Model must not contain reversible reactions
  - In deterministic models forward and reverse reactions can cancel each other out
  - Stochastic models consider each single reaction event
- Higher-order kinetics?
  - Stochastic models are based on mass action kinetics
  - Only simple enzyme kinetics, such as Michealis-Menten has been shown to be valid in stochastic models under the same conditions

Other problems (cont.)

- Stochastic formalism lacks behind in terms of analysis methods (stochastic bifurcation analysis, stochastic MCA, stochastic parameter fitting, etc.)

"If an analytical solution is required, then the deterministic approach will always be easier [..]"
Gillespie, 1976

Approximate stochastic methods

- **Mesoscopic Approach** (Morton-Firth 1998 and others)
  Single particles are distinguished, but their position and velocity are neglected, multi-state particles possible
- **PW-DMC** (Resat 2001)
  Reactions with high probability are allowed to fire multiple times, grouping of reaction events
- **Stochastic differential equations** (SDE)
  Differential equations with noise-term on the right hand side

Rationale for the τ-Leap method

- Each single reaction event in the system has to be calculated in the exact stochastic simulation methods
  $\rightarrow$ **huge computational effort** for bigger systems in terms of particle numbers, because the number of reaction events (and the computation time) per unit time is roughly proportional to the number of particles present

- Can we speed up the simulation, if we group several reaction events?
  (we will loose exactness but, maybe, we can get a good approximation)

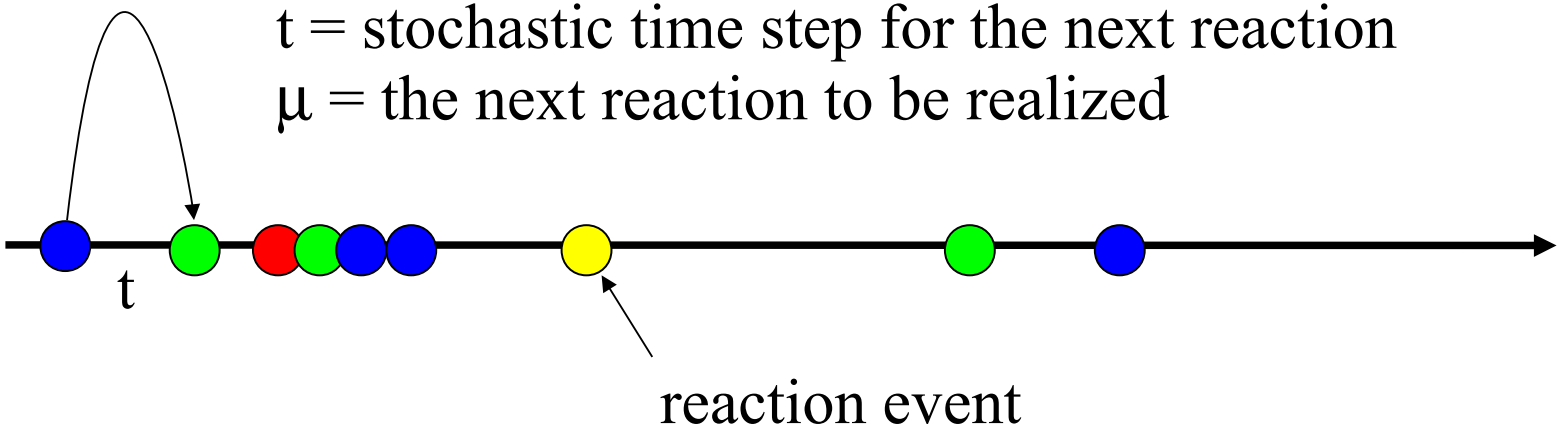
The τ-Leap Method

Direct Method

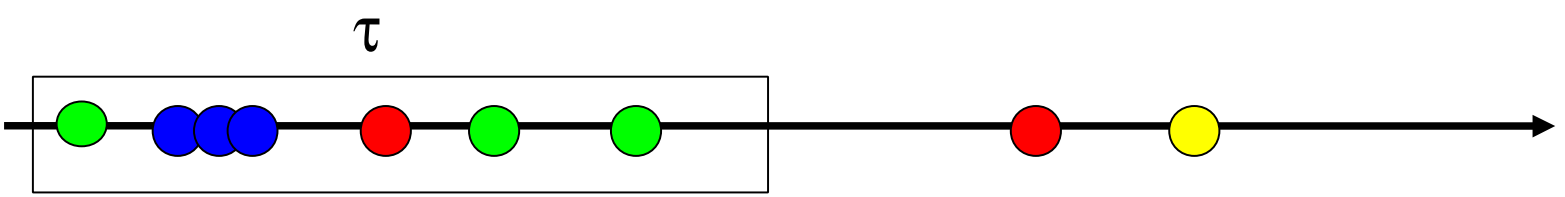
in each step calculate:

t = stochastic time step for the next reaction

μ = the next reaction to be realized



τ-Leap Method



for τ-leap calculate for each reaction i:

k_i = number of reaction events of reaction i

within the step of length τ (Poisson distribution)

Discussion of the τ-Leap method

- τ has to be:
  - small enough $\rightarrow$ no change in the propensities during leap
  - large enough $\rightarrow$ leap over many single reaction events to get a speed-up
- Open question: How to determine the leap value?

$$\text{run time} \leftrightarrow \text{accuracy}$$

- Simple τ-choosing strategy, estimated midpoint method, and many over variants
- Other problem: How to avoid negative particle numbers during simulation?

Connections

Stochastic simulation algorithms

Leap-condition: a_μ do not change during leap τ

⬇

Tau-Leap algorithm

τ is “macroscopically infinitesimal” (many reactions fire)

⬇

Chemical Langevin Equation

In the limit noise term vanishes

⬇

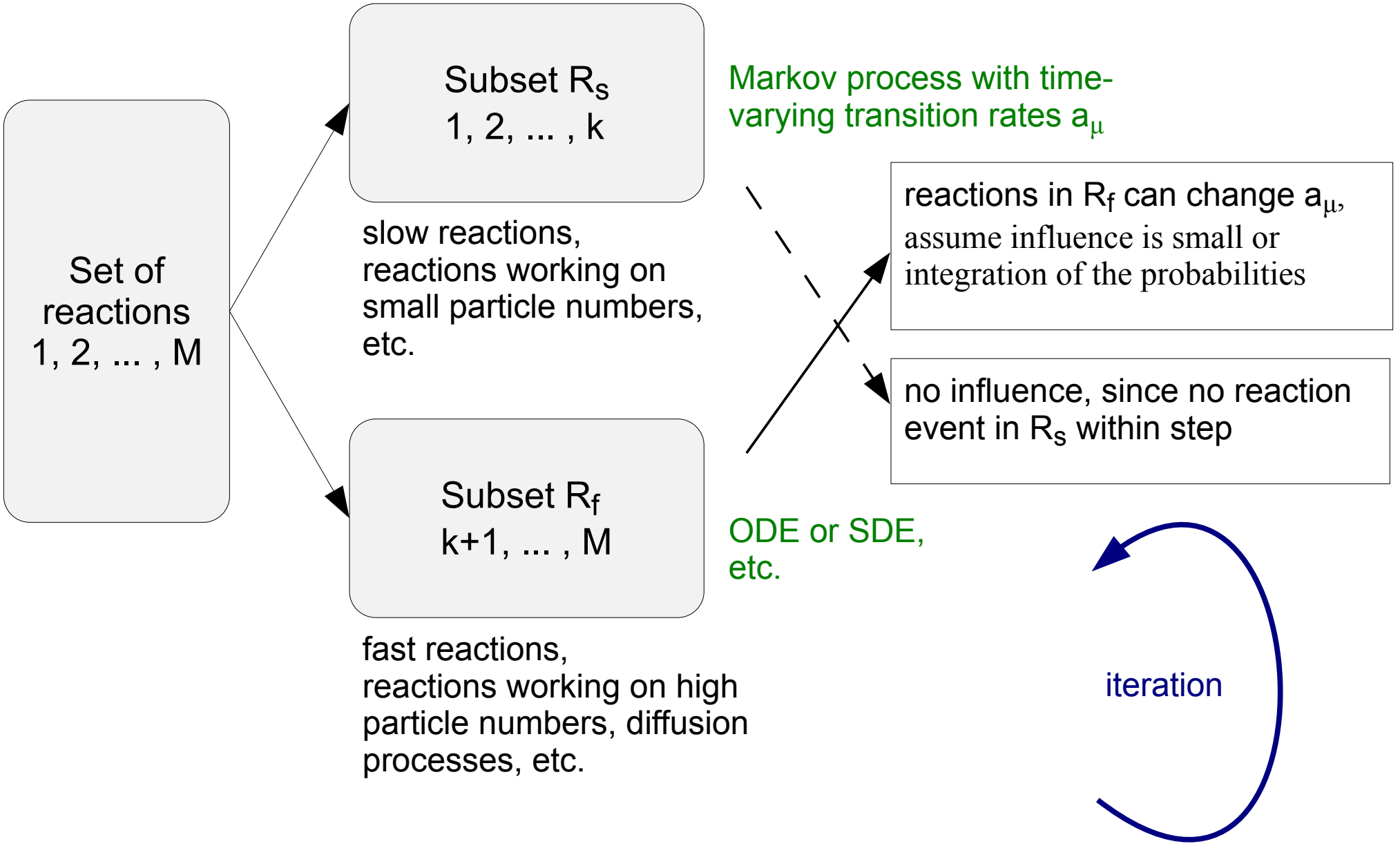
Euler update scheme (ODEs)

Hybrid Simulation Methods

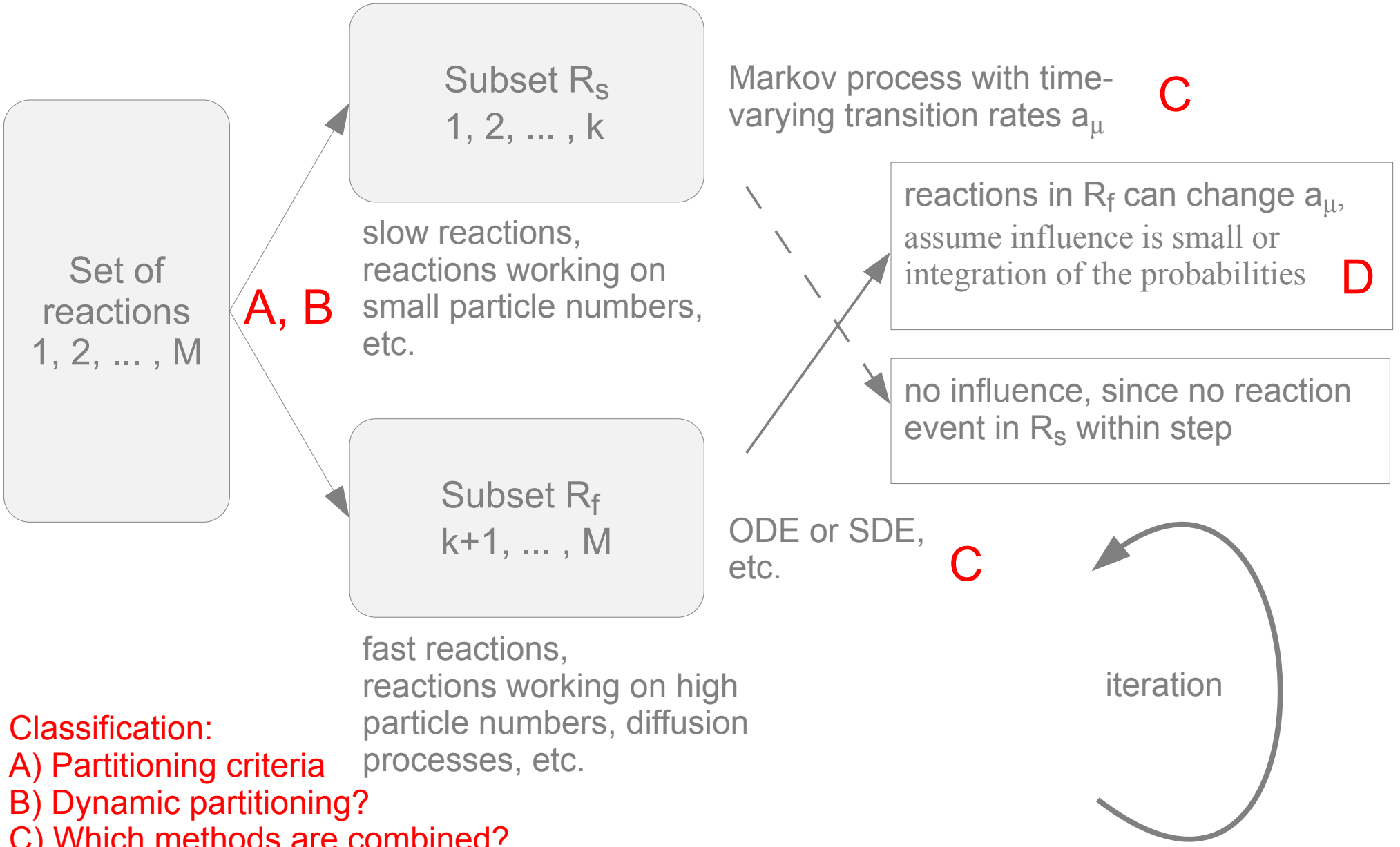
- Deterministic and stochastic approaches are complementary
- Idea: Divide the network into several parts and use the appropriate simulation method on each subnetwork
- Repartitioning if needed

- Problem: Synchronization

Hybrid Algorithm Schema



Hybrid Algorithm Schema



Classification:
A) Partitioning criteria
B) Dynamic partitioning?
C) Which methods are combined?
D) Provision for time-varying propensities a_μ

Hybrid Algorithms 1/2

Hybrid approaches	*Methods integrated*	*Dynamic Partitioning*	*Partitioning criteria*	*Variable prob. during stoch. step*
Alur (2001)	Direct Method / ODE	✓	particle no.	✘
Haseltine (2002)	Direct Method / SDE	✘	heuristics	✓/✘
Pahle (2002)	Next Reaction Method / ODE	✓	particle no.	✘
Adalsteinsson (2004)	Direct Method / ODE	✘	user-defined	✘
Bentele (2004)	Next Reaction Method / SDE	✓	relat. fluct. and particle no.	✓
Burrage (2004)	Direct Method. / tau-Leaping / SDE	✓	propensities and particle no.	✘
Kiehl (2004)	Direct Method / ODE	✘	user-defined	✓/✘
Neogi (2004)	Stoch. Sim. / ODE	✓	particle no.	✓
Puchalka (2004)	Next Reaction Method / tau-Leaping	✓	substrate no. and relat. prop.	✘

Hybrid Algorithms 2/2

Hybrid approaches	*Methods integrated*	*Dynamic Partitioning*	*Partitioning criteria*	*Variable prob. during stoch. step*
Takahashi (2004)	Next Reaction Method / ODE	✘	user-defined	✘
Vasudeva (2004)	Direct Method / ODE	✓	propensities and particle no.	✘
Alfonsi (2005)	Next Reaction Method / SDE	✓	propensities	✓
Salis (2005)	Next Reaction Method / SDE	✓	propensities and particle no.	✓
Griffith (2006)	Direct Method / ODE	✓	propensities and particle no.	✓
Harris (2006)	Direct M. / tau-Leap / Langevin / ODE	✓	propensities	✘
Wagner (2006)	First Reaction M. / discr. Gauss / ODE	✓	error criterion	✘

Propensity calculations

- Reaction Probability Density Function
  - When will the next reaction take place?
  - Which reaction will it be?

$$P(\tau,\mu|x,t)=\begin{cases} a_\mu(x)\cdot\exp(-a_0(x)\tau) & if\ 0\leqslant\tau<\infty\wedge\mu=1,\ldots,M \\ 0 & otherwise \end{cases}$$

- Time-varying propensities a_μ

$$P(\tau,\mu|x,t)=a_\mu(t+\tau)\cdot\exp(-\int_t^{t+\tau}a_0(t)\,dt)$$

How to derive Reaction Probability Density Function with time-varying propensities...

D.T Gillespie (2002) *Markov Processes: An Introduction for Physical Scientists*. Academic, New York

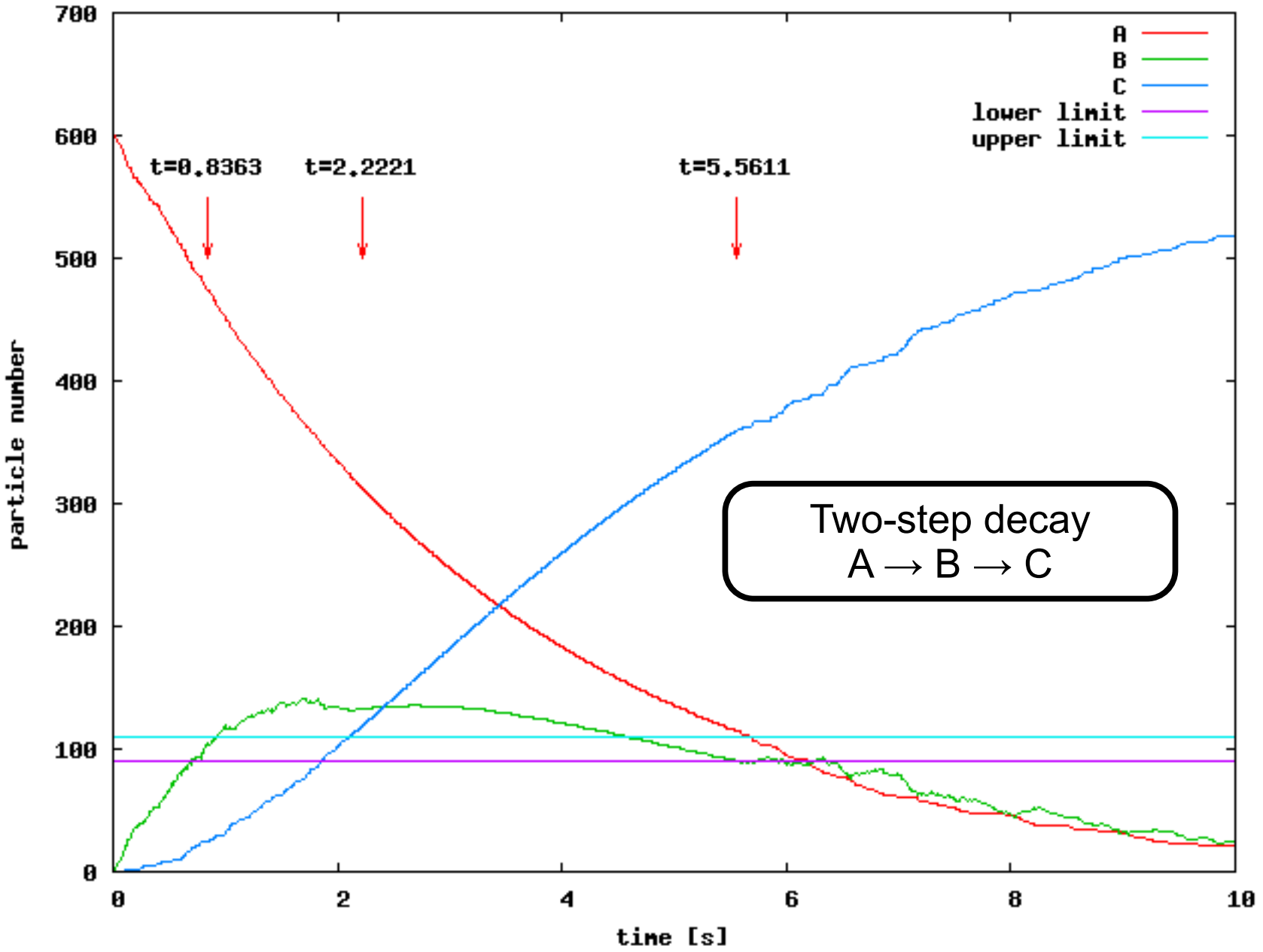
$$\int_{t}^{t+\tau} a_0^{slow}(t_1)\,dt_1 + \log(r_1) = 0$$

$$P(\mu|\tau) = \frac{a_\mu(t+\tau)}{a_0^{slow}(t+\tau)}$$

Hybrid Simulation Methods

- **Advantages**:
  - Efficient simulation of fast reaction events, which would slow down stochastic simulations
  - Random fluctuations are considered if needed
- **Problems**:
  - Synchronisation
  - Reliable criteria for partitioning?
  - Repartitioning → computational overhead
  - Implementation much more complicated than the Gillespie algorithm

Hybrid Simulation with COPASI



Partitioning of the Master Equation

- Haseltine & Rawlings (2005) On the origins of approximations for stochastic chemical kinetics. *J. Chem. Phys.* **123**:164115
- Rao & Arkin (2003) Stochastic chemical kinetics and the quasi-steady state assumption: Application to the Gillespie algorithm *J. Chem. Phys.* **118**(11):4999
- Cao, Gillespie and Petzold (2005) The slow-scale stochastic simulation algorithm *J. Chem Phys.* **122**:014116
- Haseltine & Rawlings (2002) Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.* **117**(15):6959

J. Pahle (2009) Biochemical simulations: stochastic, approximate stochastic and hybrid approaches.
Briefings in Bioinformatics **10**(1):53-64, doi:10.1093/bib/bbn050

BRIEFINGS IN BIOINFORMATICS. VOL 10. NO 1. 53–64 doi:10.1093/bib/bbn050
Advance Access publication January 16, 2009

Biochemical simulations: stochastic, approximate stochastic and hybrid approaches

Jürgen Pahle

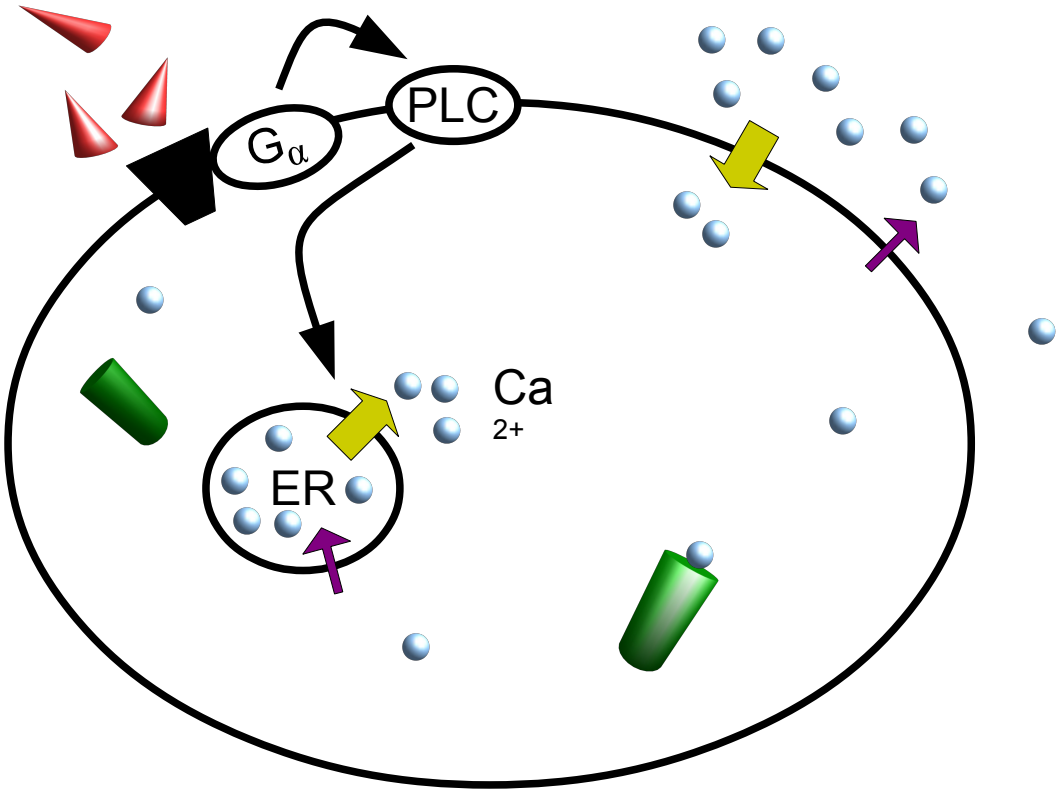
Submitted: 1st July 2008; Received (in revised form): 13th October 2008

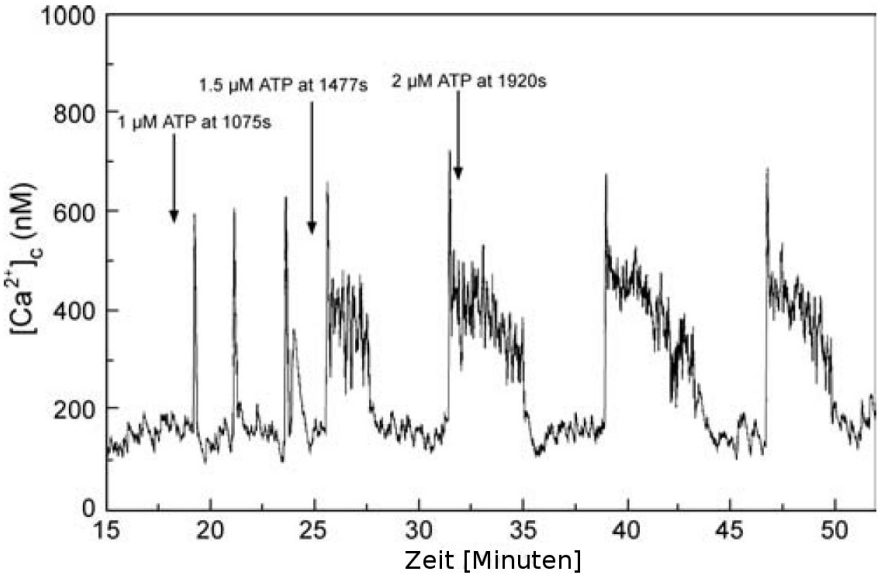
Abstract

Computer simulations have become an invaluable tool to study the sometimes counterintuitive temporal dynamics of (bio-)chemical systems. In particular, stochastic simulation methods have attracted increasing interest recently. In contrast to the well-known deterministic approach based on ordinary differential equations, they can capture effects that occur due to the underlying discreteness of the systems and random fluctuations in molecular numbers. Numerous stochastic, approximate stochastic and hybrid simulation methods have been proposed in the literature. In this article, they are systematically reviewed in order to guide the researcher and help her find the appropriate method for a specific problem.

Keywords: *stochastic simulation; biochemical systems; approximate stochastic simulation; hybrid simulation methods; systems biology*

Signal transduction via Ca^{2+} ions

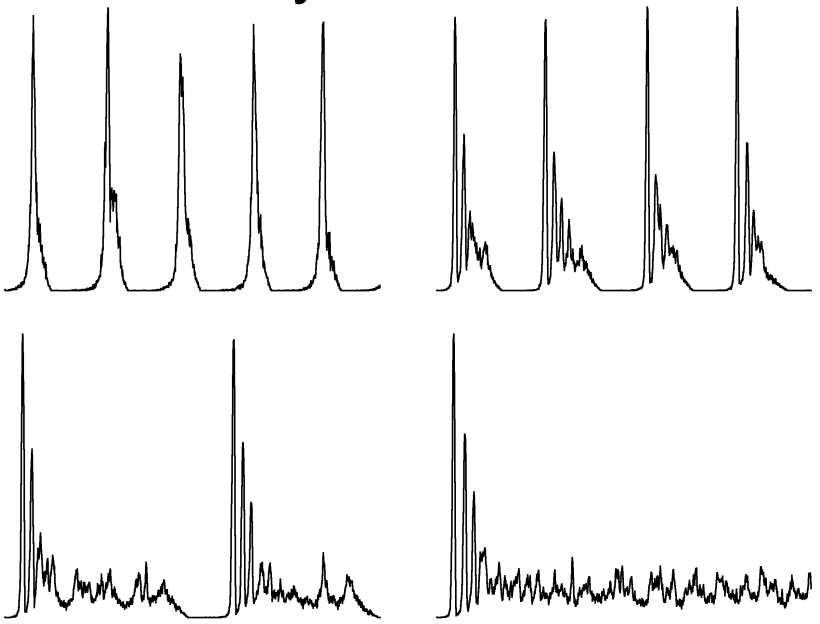




Ca^{2+} oscillations model:
U. Kummer et al. (2000) *Biophys. J.* **79**:1188

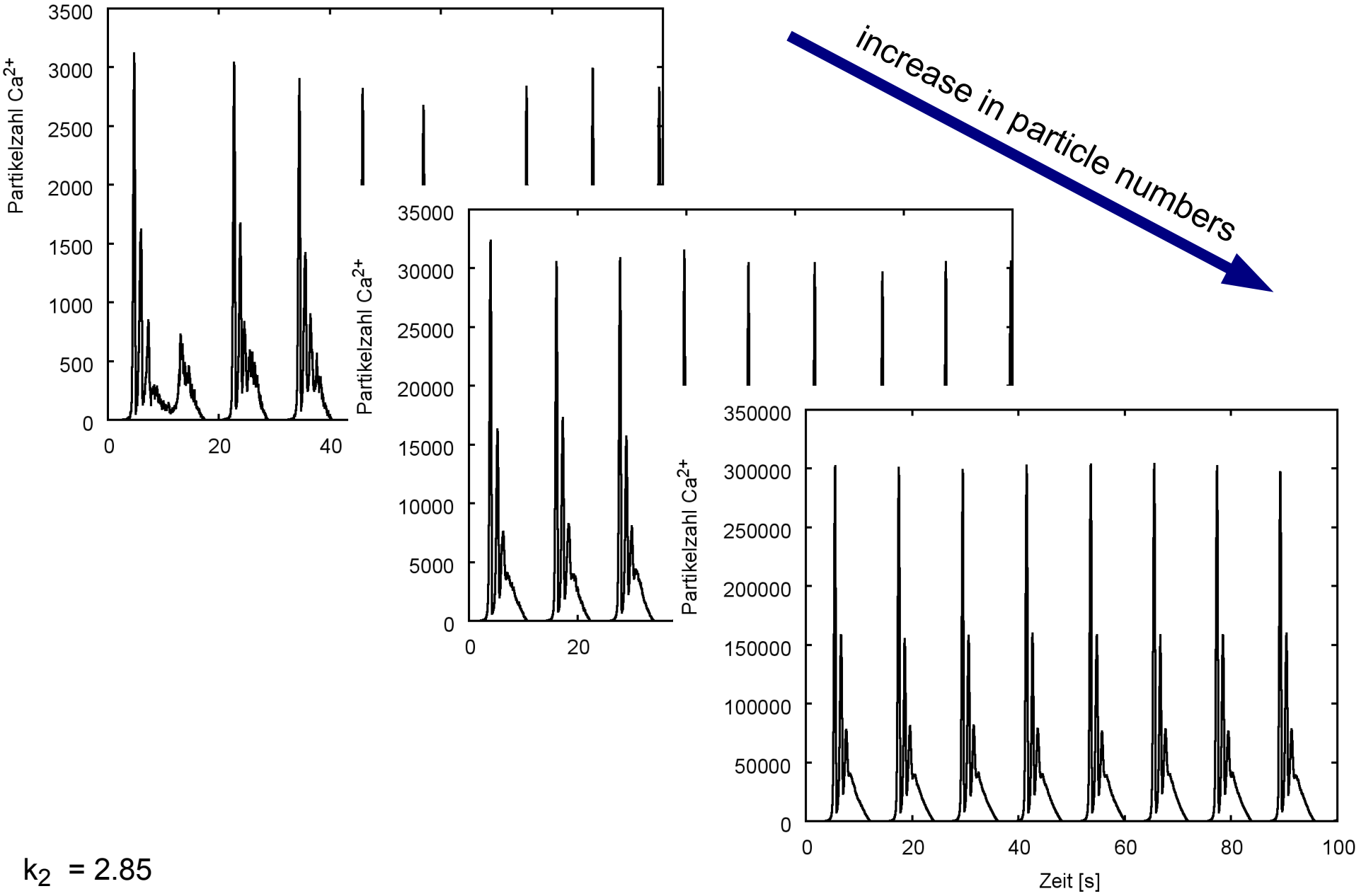
Bifurcation parameter k_2 corresponds to the strengh of stimulation, e.g. the concentration of ATP in the experiment

Different dynamics:



Transition

(from stochastic to quasi-deterministic behaviour)



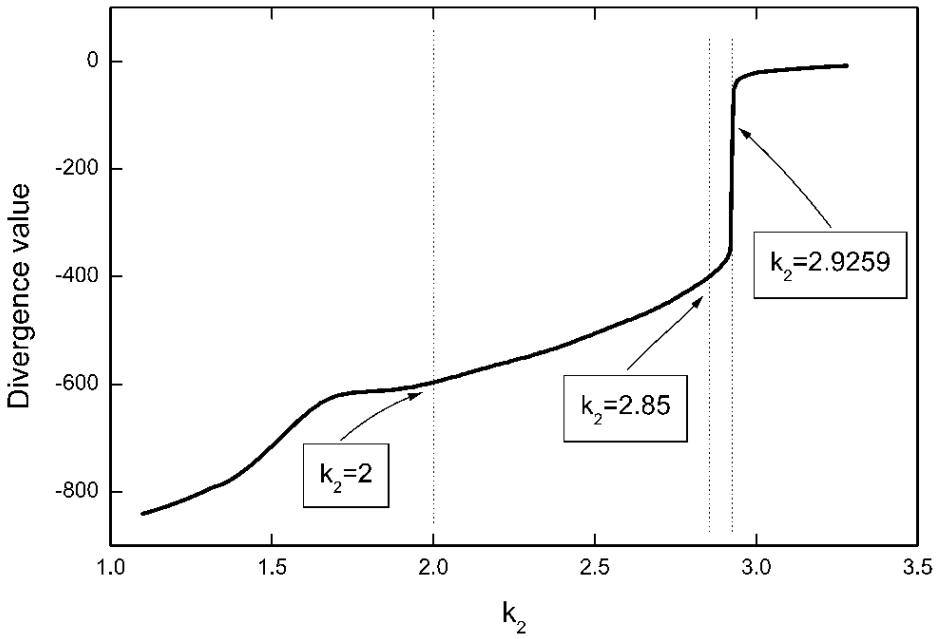
k_2 = 2.85

Divergence

- Motivation:
  - Relative stochastic effects decrease with higher particle numbers
  - The system shows a quasi-deterministic behavior
  - This transition is dependent on the specific system (and even the current dynamics!)
- Our studies show that the so-called **divergence** has a strong effect on when this transition occurs
  - high divergence → high particle numbers are needed
    and vice versa
- Divergence is the average of the sum of all Lyapunov exponents (exponential convergence or divergence of the trajectory)

Results

k_2	Particle number	Dynamics
2	ten thousands	periodic spiking
2.85	ten thousands	periodic bursting
2.9259	hundred thousands	chaos
2.99	millions	regular oscillations
3	> millions	steady state

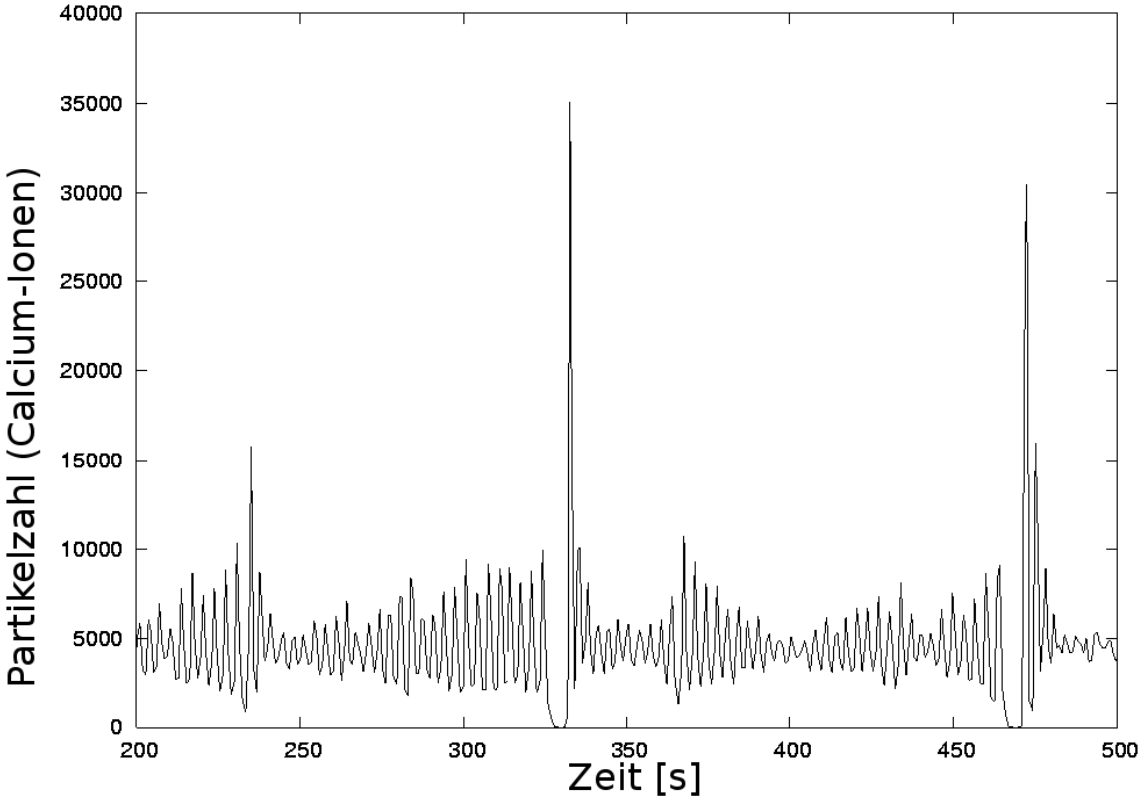


- The transition range is dependent on the sensitivity of the system. The sensitivity of the system is, in turn, dependent on the current dynamics
- A good indicator for this is not the "complexity" of the oscillations but the so-called "divergence" of the system (average sum of Lyapunov exponents)
- System can show qualitatively different behavior when simulated stochastically

Other test models

- Peroxidase-oxidase reaction (Olsen et al. 2003)
- MAP-Kinase cascade (Kholodenko 2000)
- Buffered calcium system

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



k_2 = 3.0