

# VI2

# Stochastische Effekte

16. Januar 2014

# Übersicht

Stochastische Effekte

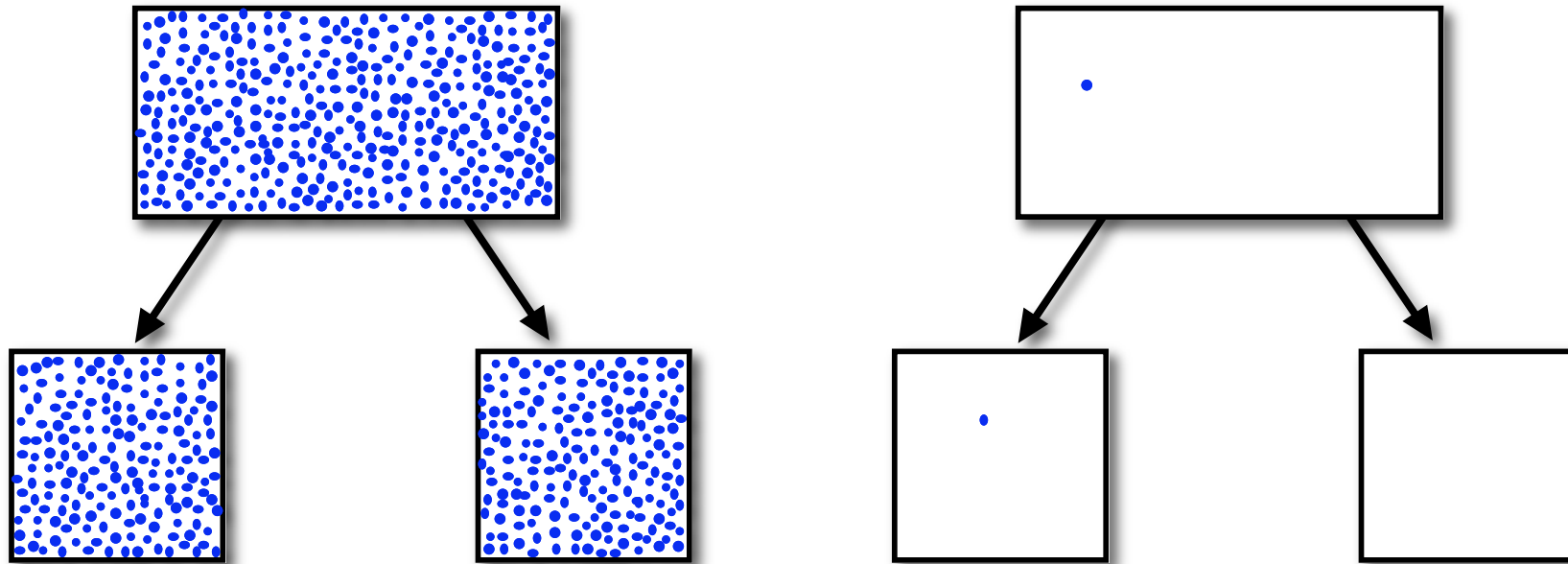
=> was sind das?

=> warum wichtig?

=> wie behandeln?

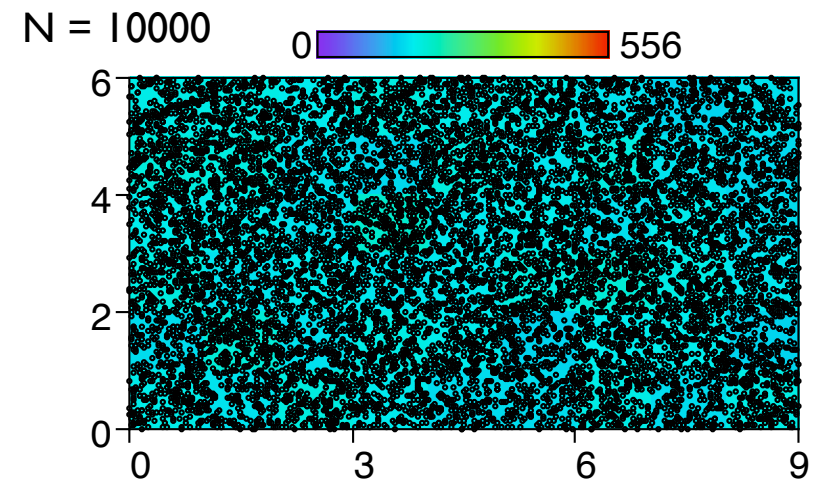
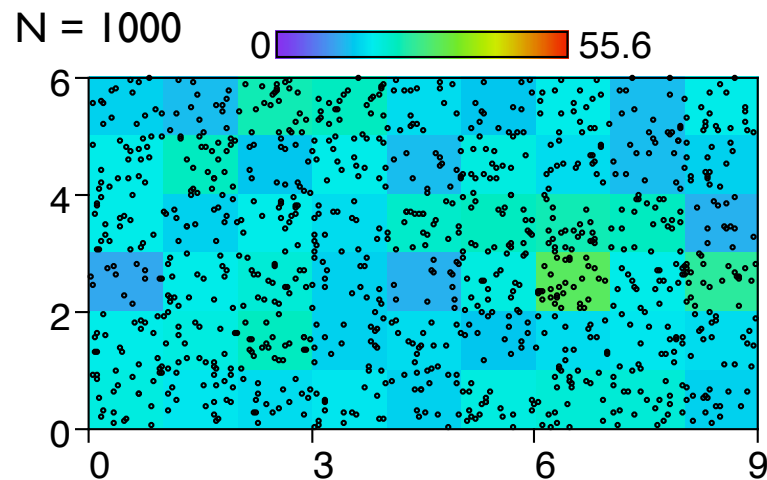
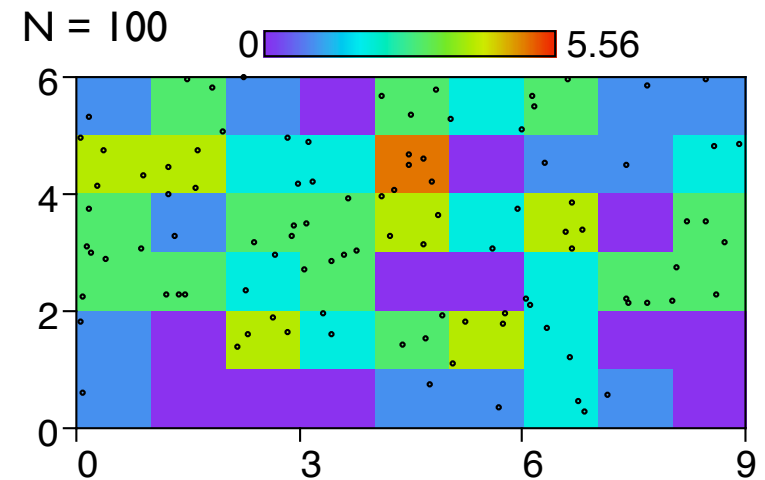
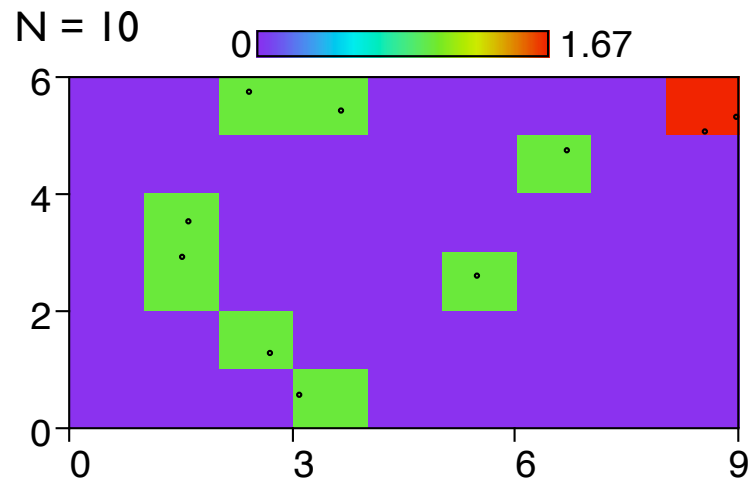
Gillespie-Algorithmus  $\Leftrightarrow$  direkte Implementierung

# Stochastische Effekte



$$\text{Dichte} = \frac{\text{ununterscheidbare Teilchen}}{\text{Volumen}}$$

# Dichtefluktuationen



# Poisson-Verteilung

Betrachte Kontinuum  $w$  mit im Mittel  $\lambda$  Ereignissen pro Einheitsintervall  $\Delta w$

Annahmen:

- i) Seltenheit:  $\ll 1$  Ereignisse in  $[w, w+\Delta w]$ , höchstens ein Ereignis
- ii) Proportionalität:  $\langle N \rangle = \lambda \Delta w$
- iii) Geschichtslosigkeit (Markov-Prozess)

Wahrscheinlichkeit, dass  $k$  Ereignisse pro Einheitszeitintervall auftreten:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

Mittelwert:  $\langle k \rangle = \sum k p_k = \lambda$

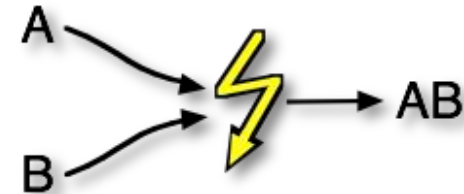
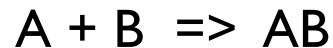
Varianz:  $\sigma^2 = \sum p_k (k - \langle k \rangle)^2 = \lambda$

Relative Streuung (Fehler):  $\frac{\Delta k}{k} = \frac{\sigma}{\langle k \rangle} = \frac{1}{\sqrt{\lambda}}$

|                       |     |      |       |
|-----------------------|-----|------|-------|
| Mittlere Teilchenzahl | 100 | 1000 | 1 Mol |
| relative Unsicherheit | 10% | 3%   | 1e-12 |

# Reaktionen im Teilchenbild

Assoziation:



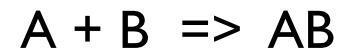
Kontinuierliche Ratengleichung:  $\frac{d[AB]}{dt} = k[A][B]$

Anzahl neuer AB in  $V$  während  $\Delta t$ :

$$\begin{aligned}\Delta N_{AB} &= \frac{d[AB]}{dt} V \Delta t \\ &= k_{AB} \frac{N_A}{V} \frac{N_B}{V} V \Delta t \\ &= \frac{k_{AB} \Delta t}{V} N_A N_B \\ &= P_{AB} N_A N_B\end{aligned}$$

Reaktionsrate  $k_{AB} \Rightarrow$  Reaktionswahrscheinlichkeit  $P_{AB}$

# Direkte Implementierung



```
Continuous_AB.py

# continuous association of A and B

# parameter
tEnd = 5.0
dt = 0.01
volume = 100.0

# rate and probability
kAB = 1.0
prob = kAB * dt / volume

# initial conditions: particle numbers
A = 1000
B = 1000
AB = 0

# convert to densities
A = A/volume
B = B/volume
AB = AB/volume

# main loop
t = 0.0
print t, "\t", A, "\t", B, "\t", AB

while(t<tEnd):
    dAB = dt * kAB * A * B

    AB += dAB
    A -= dAB
    B -= dAB

    # increment time and output
    t += dt
    print t, "\t", A, "\t", B, "\t", AB
```

```
Stochastic_AB.py

# Stochastic association of A + B => AB

import random

# parameter
tEnd = 5.0
dt = 0.01
volume = 100.0

# rate and probability
kAB = 1.0
prob = kAB * dt / volume

# initial conditions
A = 1000
B = 1000
AB = 0

# main loop
t = 0.0
print t, "\t", A/volume, "\t", B/volume, "\t", AB/volume

while(t<tEnd):
    dAB = 0
    # check for every pair A, B
    for ia in xrange(A):
        for ib in xrange(B):
            r = random.random()
            if (r < prob):
                dAB+=1

    AB += dAB
    A -= dAB
    B -= dAB

    # increment time and output
    t += dt
    print t, "\t", A/volume, "\t", B/volume, "\t", AB/volume
```

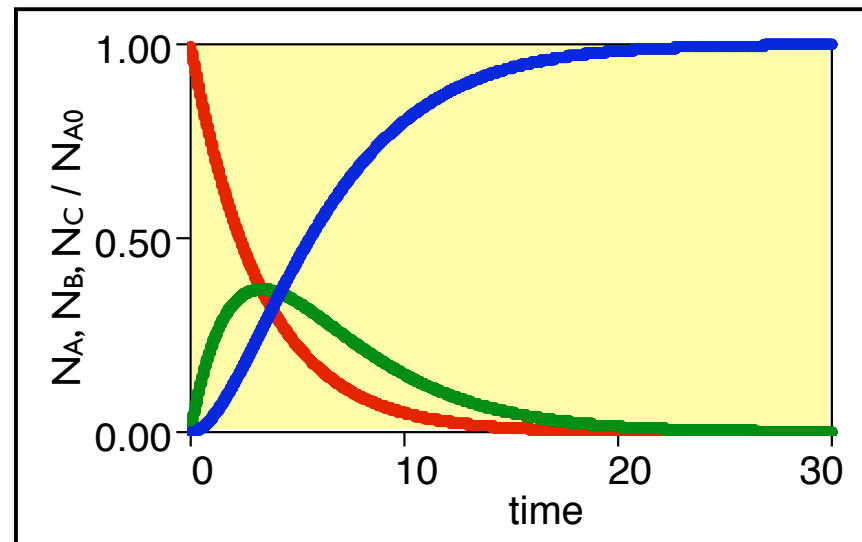
Achtung: didaktische Implementierungen!

# Beispiel: Reaktions-Kette



Raten:  $\frac{dA}{dt} = -k_1 A$        $\frac{dB}{dt} = k_1 A - k_2 B$        $\frac{dC}{dt} = k_2 B$

Zeitentwicklung aus den kontinuierlichen Raten:



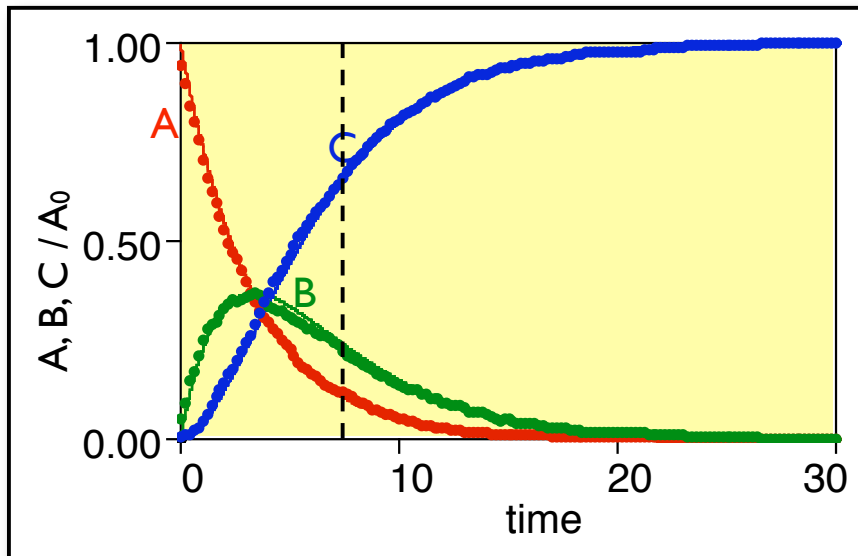
$$k_1 = k_2 = 0.3$$

# Stochastische Simulation

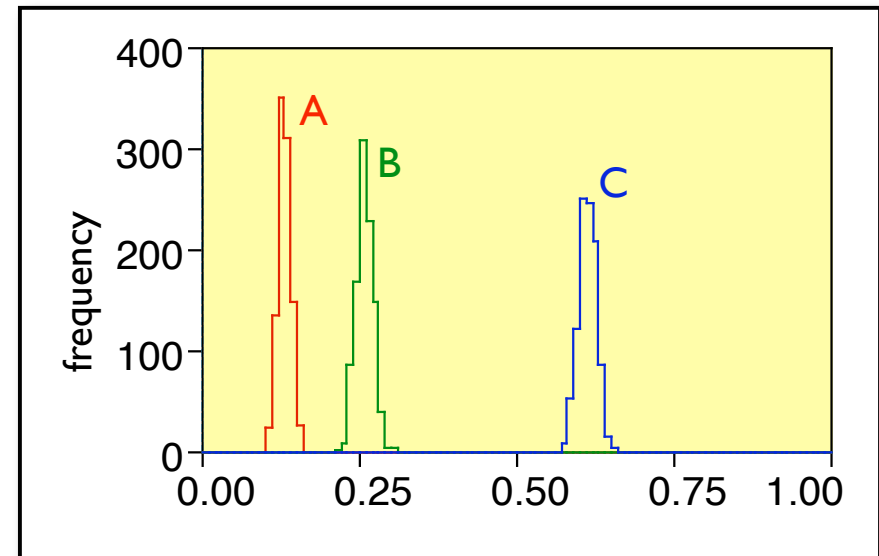


$A_0 = 1000$  Teilchen bei  $t = 0$

$t = 7$



$k_1 = k_2 = 0.3$

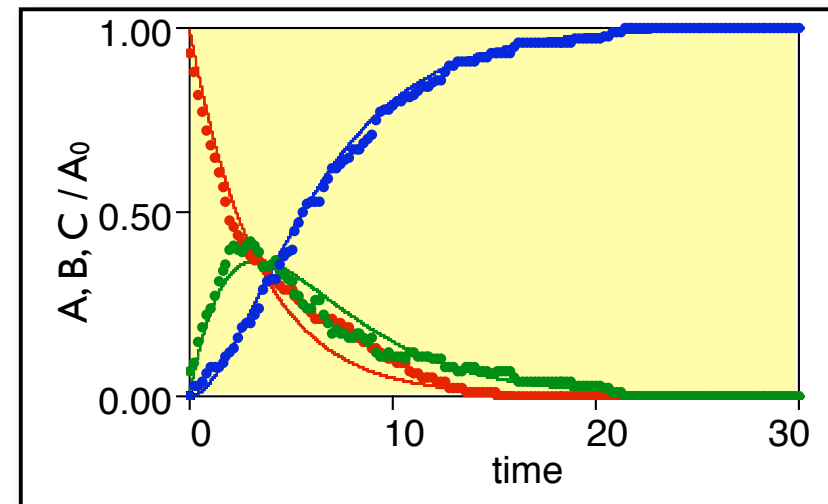
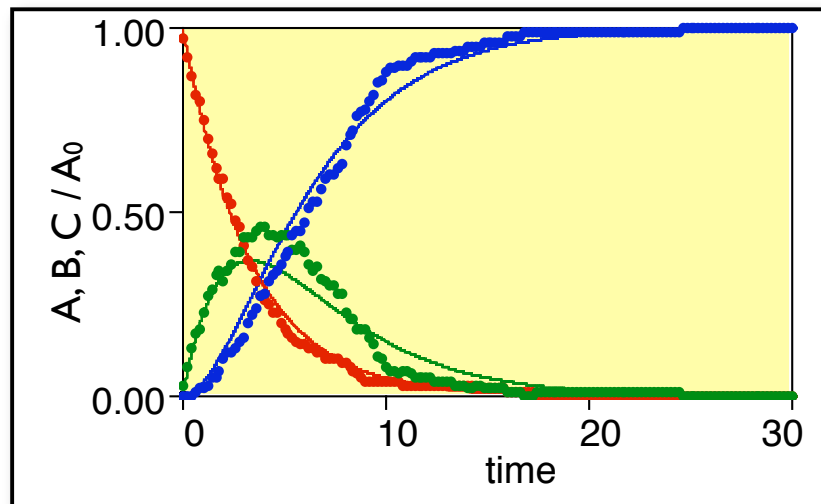
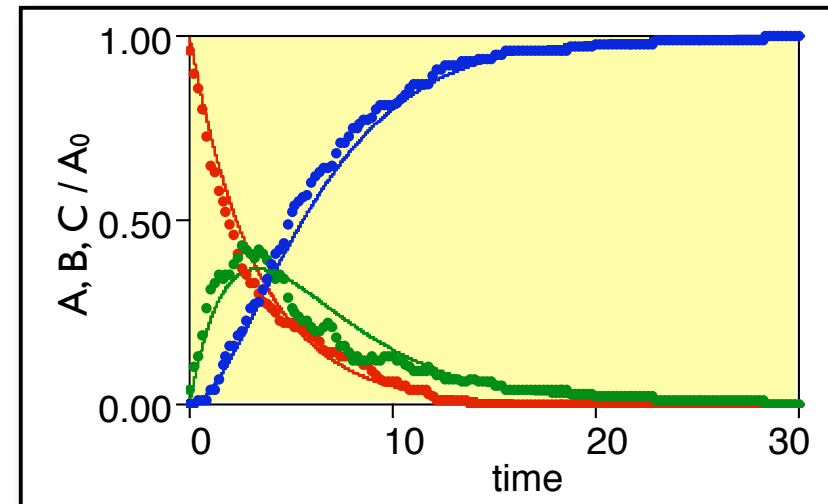
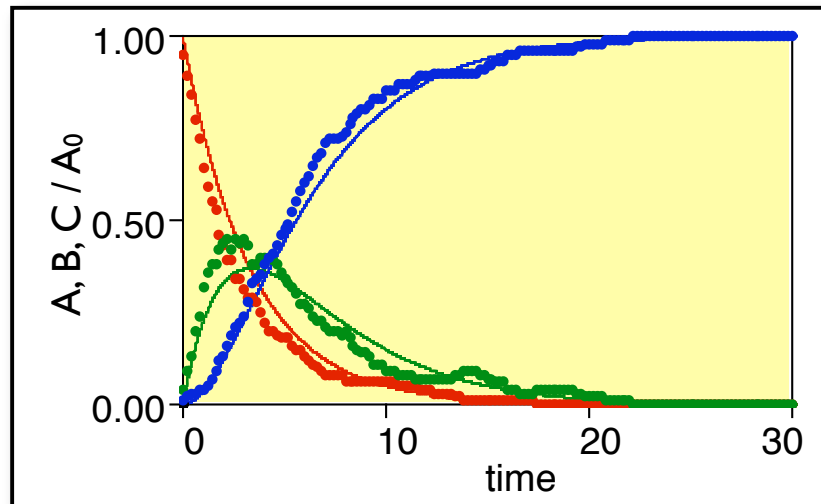


Werte bei  $t = 7$  (1000 Läufe)

=> Fluktuationen

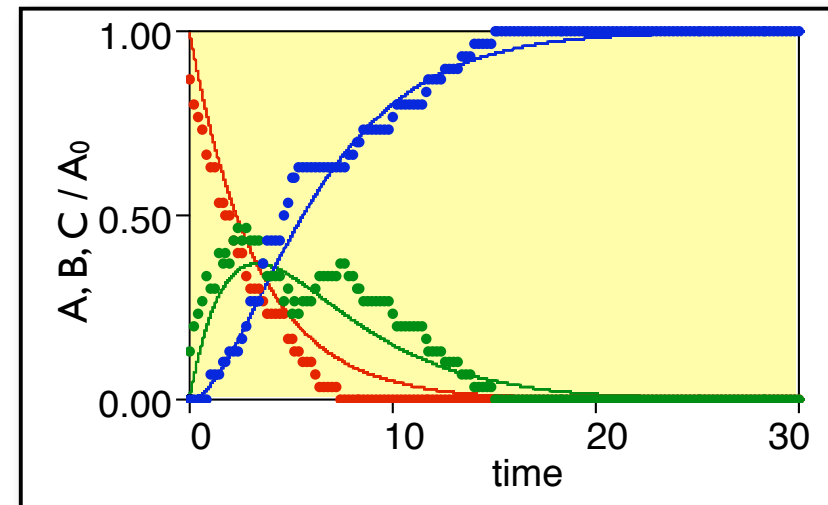
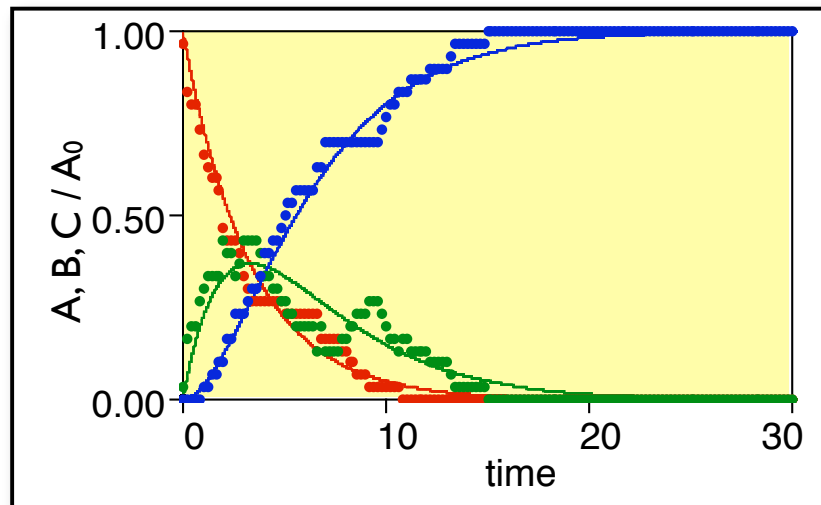
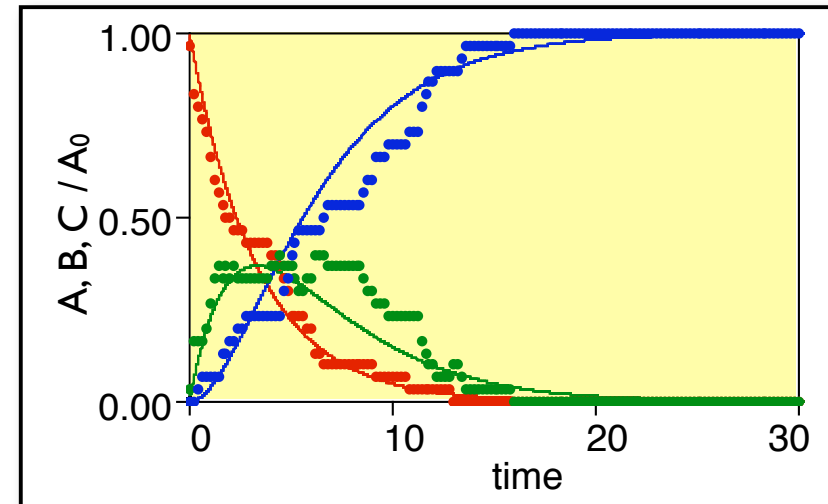
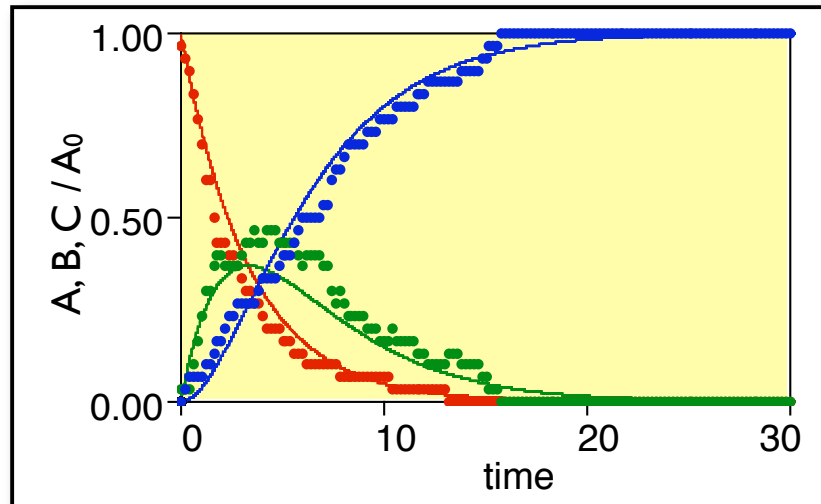
# Weniger Teilchen => Mehr Rauschen

$A_0 = 100$



# Noch weniger Teilchen

$$A_0 = 30$$



# Varianz vs. Teilchenanzahl

Poisson:

relative Abweichung  $\propto 1/\sqrt{N}$

1000 Simulationsläufe,  
Werte sichern bei  $t = 7$ .

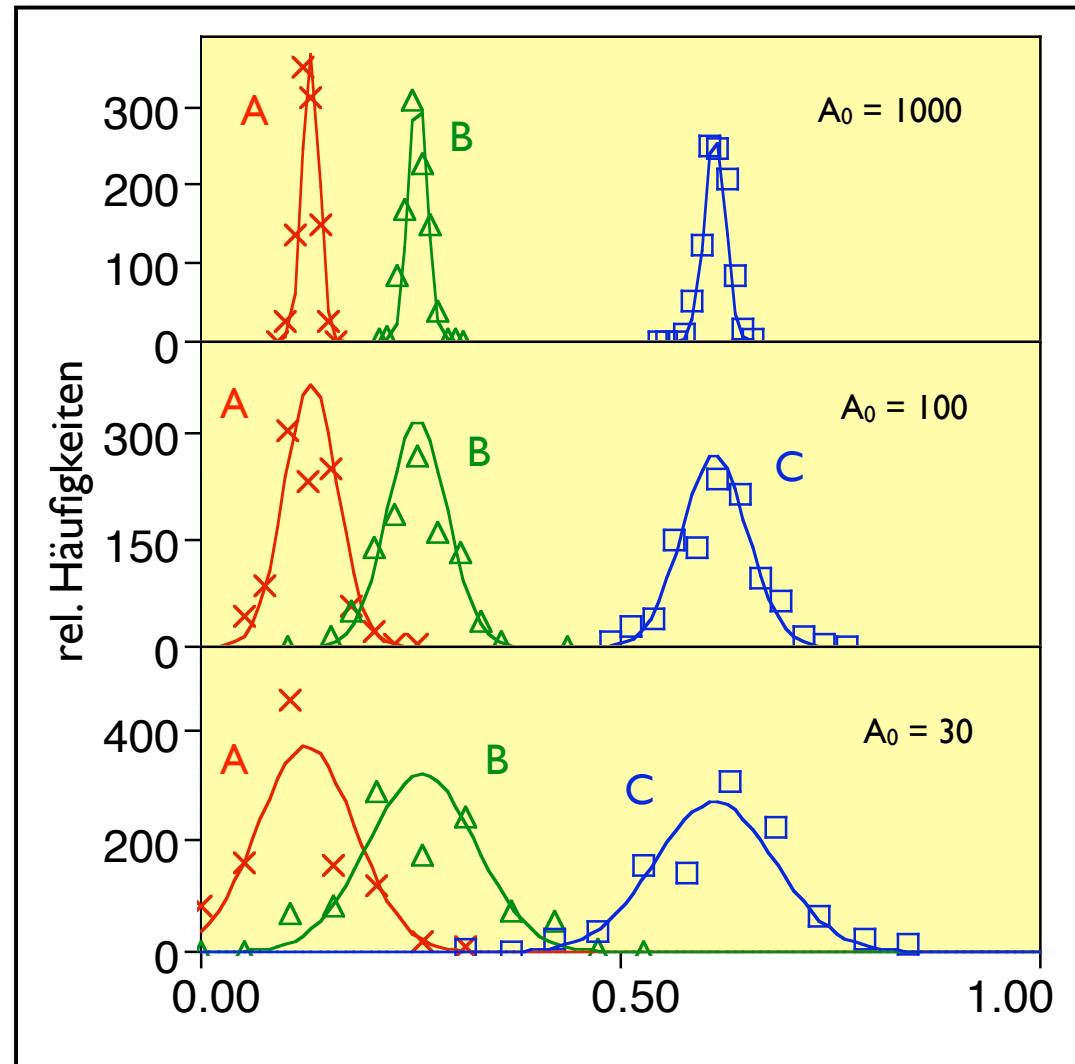
Fit der Verteilung mit Gaussvert.  
(Normalverteilung)

$$g(x) = \exp \left[ -\frac{(x - \langle x \rangle)^2}{w/\sqrt{A_0}} \right]$$

$$\langle A \rangle = 0.13, \quad w_A = 0.45$$

$$\langle B \rangle = 0.26, \quad w_B = 0.55$$

$$\langle C \rangle = 0.61, \quad w_C = 0.45$$



# Gillespie-Algorithmus

2340

## Exact Stochastic Simulation of Coupled Chemical Reactions

Daniel T. Gillespie\*

Research Department, Naval Weapons Center, China Lake, California 93555 (Received May 12, 1977)

Publication costs assisted by the Naval Weapons Center

There are two formalisms for mathematically describing the time behavior of a spatially homogeneous chemical system: The *deterministic* approach regards the time evolution as a continuous, wholly predictable process which is governed by a set of coupled, ordinary differential equations (the "reaction-rate equations"); the *stochastic* approach regards the time evolution as a kind of random-walk process which is governed by a single differential-difference equation (the "master equation"). Fairly simple kinetic theory arguments show that the stochastic formulation of chemical kinetics has a firmer physical basis than the deterministic formulation, but unfortunately the stochastic master equation is often mathematically intractable. There is, however, a way to make exact numerical calculations within the framework of the stochastic formulation without having to deal with the master equation directly. It is a relatively simple digital computer algorithm which uses a rigorously derived Monte Carlo procedure to numerically simulate the time evolution of the given chemical system. Like the master equation, this "stochastic simulation algorithm" correctly accounts for the inherent fluctuations and correlations that are necessarily ignored in the deterministic formulation. In addition, unlike most procedures for numerically solving the deterministic reaction-rate equations, this algorithm never approximates infinitesimal time increments  $dt$  by finite time steps  $\Delta t$ . The feasibility and utility of the simulation algorithm are demonstrated by applying it to several well-known model chemical systems, including the Lotka model, the Brusselator, and the Oregonator.

naive Integration  
=> shortcomings

Gillespie  
=> Idee an der Tafel  
=> Parallelen zu den Poisson-Annahmen

D. Gillespie, J. Phys. Chem. **81** (1977) 2340–2361

# Das Problem

Zerfallsreaktion:  $A \Rightarrow \emptyset$

Wahrscheinlichkeit für ein Ereignis in  $(t, t+\Delta t)$  bei  $A(t)$  Molekülen  $= A(t) k \Delta t$

Naiver Algorithmus:

```
A = A0
For every timestep:
    get random number  $r \in [0, 1)$ 
    if  $r \leq A * k * dt$ :
        A = A-1
```

Es geht, aber:  $A(t) * k * dt \ll 1$  — Genauigkeit  
 $\Rightarrow$  meistens passiert nichts

Gillespie überlegte sich: "Wann ist das nächste Ereignis?"

# Exponentiell verteilte Wartezeiten

**Idee:** wann findet die **nächste Reaktion** statt?

Mit  $A(t)$  Molekülen:

$g(A(t), s)$  = Wahrscheinlichkeit, dass keine Reaktion in  $(t, t+s)$  stattfindet  
+ Poisson-Annahmen

=> DGL für Verteilung der Wartezeiten:

$$g(A(t), s) = \exp[-A(t) k s] = e^{-s/s_0}$$

mittlere Wartezeit:  $s_0 = \frac{1}{k A(t)}$

# Gillespie – Schritt I

**Idee:** Finde heraus, **wann** die **nächste Reaktion** stattfinden wird!  
(Zwischen den diskreten Reaktionen passiert nichts ...)

Es seien  $A(t)$  Moleküle zum Zeitpunkt  $t$  im System

$f(A(t), s)$  = W'keit. dass für  $A(t)$  Moleküle die nächste Reaktion im Intervall  $(t+s, t+s+ds)$  stattfindet, mit  $ds \Rightarrow 0$

$g(A(t), s)$  = W'keit. dass für  $A(t)$  Moleküle in  $(t, t+s)$  keine Reaktion passiert.

Dann: 
$$f(A(t), s) ds = g(A(t), s) A(t+s) k ds$$

Keine Reaktion während  $(t, t+s)$ :

$$f(A(t), s) ds = g(A(t), s) \underbrace{A(t) k ds}_{\text{W'keit. für Reaktion in } (t+s, t+s+ds)}$$

# Wahrscheinlichkeit für (keine Reaktion)

Wir brauchen nun  $g(A(t), s)$

Verlängere  $g(A(t), s)$  ein wenig:

$$g(A(t), s + ds) = g(A(t), s) [1 - A(t + s) k ds]$$

noch etwas  $A(t+s) = A(t)$  und sortiere um:

$$\lim_{ds \rightarrow 0} \frac{g(A(t), s + ds) - g(A(t), s)}{ds} = \frac{dg(A(t), s)}{ds} = -A(t) k g(A(t), s)$$

Mit  $g(A, 0) = 1$  ("keine Reaktion während Zeit = 0")

=> Verteilung der Wartezeiten zwischen diskreten Reaktionsereignissen:

$$g(A(t), s) = \exp[-A(t) k s]$$

Lebensdauer = mittlere Wartezeit:  $s_0 = \frac{1}{kA(t)}$

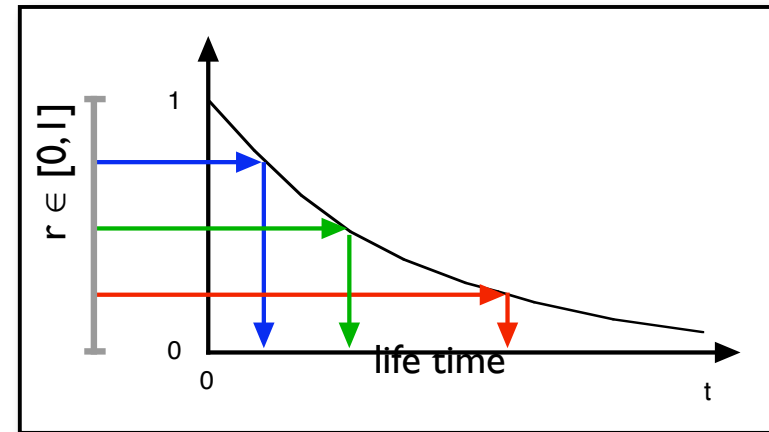
# Gillespie-Algorithmus: I Spezies

Exponentielle Verteilung der Wartezeiten:

$$g(A(t), s) = \exp[-A(t)ks]$$

Mit  $r = \exp[-A(t)ks]$

$$s = \frac{1}{kA(t)} \ln \left[ \frac{1}{r} \right] = \frac{1}{\alpha_0} \ln \left[ \frac{1}{r} \right]$$



=> Gillespie-Algorithmus für Reaktion  $A \Rightarrow \emptyset$ :

```
A = A0
While(A > 0):
    get random number r ∈ [0, 1)
    t = t + s(r)
    A = A-1
```

keine Abfrage!!!  
einfache Sprung  
in der Zeit

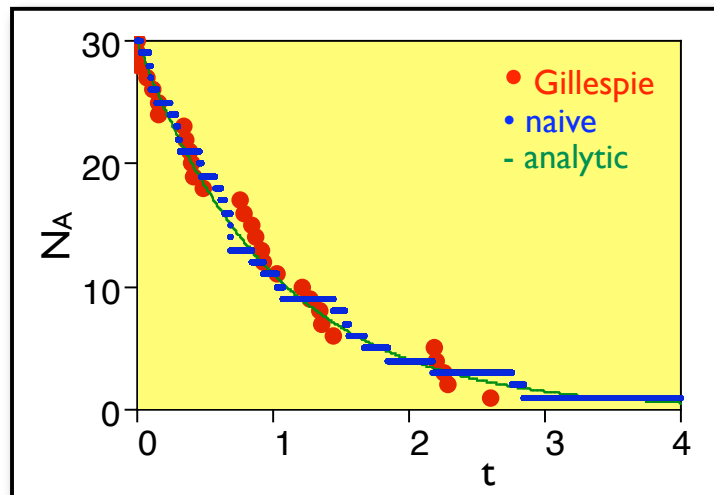
# Gillespie- vs. direkte Integration

## Naiv:

"Was ist die Wahrscheinlichkeit, dass im nächsten  $\Delta t$  ein Ereignis stattfindet?"

=> kleine feste Zeitschritte

=> Näherung erster Ordnung

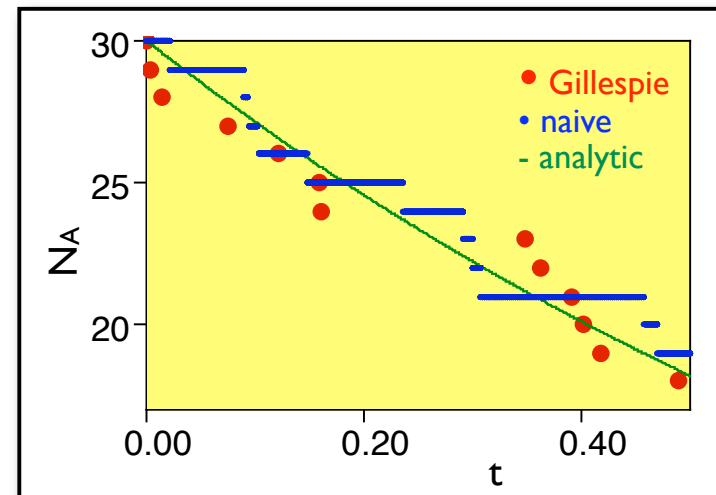


## Gillespie:

"Wie lange dauert es bis zum nächsten Ereignis?"

=> variable Zeitschritte

=> exakt



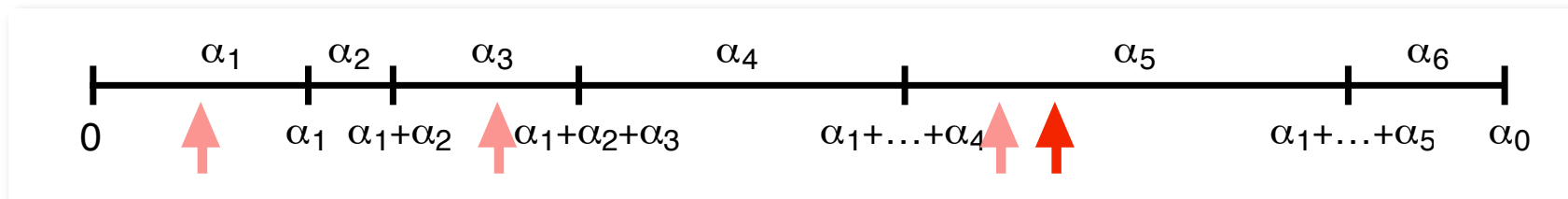
# Gillespie – viele Spezies

Für beliebige Anzahl an möglichen Reaktionen:

(i) Wahrscheinlichkeiten für jede einzelne Reaktion:  $\alpha_i$   $i = 1, \dots, N$   
 gesamte W'keit  $\alpha_0 = \sum \alpha_i$

(ii) Wartezeit bis zum nächsten beliebigen Ereignis:  $s = \frac{1}{\alpha_0} \ln \left[ \frac{1}{r_1} \right]$

(iii) Wähle eine der Reaktionen aus gemäß:  $\sum_{i=1}^{j-1} \alpha_i \leq \alpha_0 r_2 < \sum_{i=1}^j \alpha_i$



(iv) Update der Teilchenzahlen

# Ein Beispiel mit zwei Spezies

Reaktionen:



Ratengleichungen:

$$\frac{dA}{dt} = k_3 - 2A^2k_1 - ABk_2 \quad \frac{dB}{dt} = k_4 - ABk_2$$

Gleichgewichtszustand:  $A_{ss} = \sqrt{\frac{k_3 - k_4}{2k_1}} \quad B_{ss} = \frac{k_4}{k_2 A}$

mit  $k_1 = 10^{-3} \text{ s}^{-1} \quad k_2 = 10^{-2} \text{ s}^{-1} \quad k_3 = 1.2 \text{ s}^{-1} \quad k_4 = 1 \text{ s}^{-1}$

$$\Rightarrow A_{ss} = 10, \quad B_{ss} = 10$$

# Gillespie-Algorithmus

Erban, Chapman, Maini, arXiv:0704.1908v2 [q-bio.SC]

- (a4) Generate two random numbers  $r_1, r_2$  uniformly distributed in  $(0, 1)$ .
- (b4) Compute the propensity functions of each reaction by  $\alpha_1 = A(t)(A(t)-1)k_1$ ,  $\alpha_2 = A(t)B(t)k_2$ ,  $\alpha_3 = k_3$  and  $\alpha_4 = k_4$ . Compute  $\alpha_0 = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$ .
- (c4) Compute the time when the next chemical reaction takes place as  $t + \tau$  where

$$\tau = \frac{1}{\alpha_0} \ln \left[ \frac{1}{r_1} \right]. \quad (2.29)$$

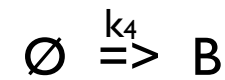
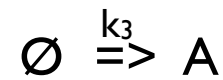
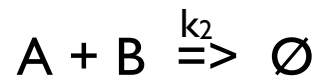
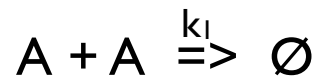
- (d4) Compute the number of molecules at time  $t + \tau$  by

$$A(t + \tau) = \begin{cases} A(t) - 2 & \text{if } 0 \leq r_2 < \alpha_1/\alpha_0; \\ A(t) - 1 & \text{if } \alpha_1/\alpha_0 \leq r_2 < (\alpha_1 + \alpha_2)/\alpha_0; \\ A(t) + 1 & \text{if } (\alpha_1 + \alpha_2)/\alpha_0 \leq r_2 < (\alpha_1 + \alpha_2 + \alpha_3)/\alpha_0; \\ A(t) & \text{if } (\alpha_1 + \alpha_2 + \alpha_3)/\alpha_0 \leq r_2 < 1; \end{cases} \quad (2.30)$$

$$B(t + \tau) = \begin{cases} B(t) & \text{if } 0 \leq r_2 < \alpha_1/\alpha_0; \\ B(t) - 1 & \text{if } \alpha_1/\alpha_0 \leq r_2 < (\alpha_1 + \alpha_2)/\alpha_0; \\ B(t) & \text{if } (\alpha_1 + \alpha_2)/\alpha_0 \leq r_2 < (\alpha_1 + \alpha_2 + \alpha_3)/\alpha_0; \\ B(t) + 1 & \text{if } (\alpha_1 + \alpha_2 + \alpha_3)/\alpha_0 \leq r_2 < 1; \end{cases} \quad (2.31)$$

Then continue with step (a4) for time  $t + \tau$ .

# Stochastisch vs. Deterministisch



Erban, Chapman, Maini, arXiv:0704.1908v2 [q-bio.SC]

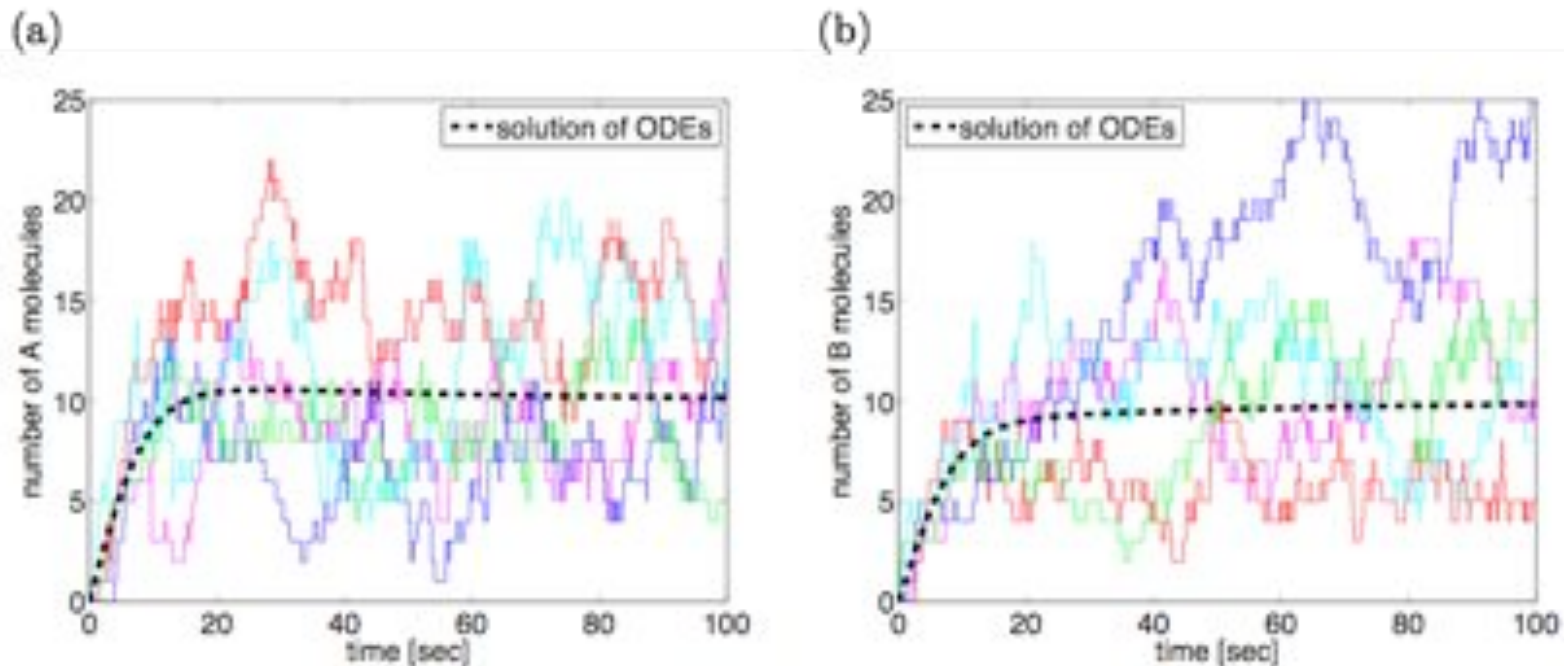
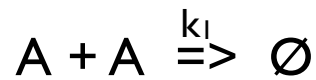
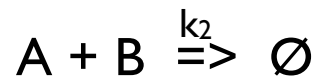


FIG. 2.3. Five realizations of SSA (a4)–(d4). Number of molecules of chemical species *A* (left panel) and *B* (right panel) are plotted as functions of time as solid lines. Different colours correspond to different realizations. The solution of (2.33)–(2.34) is given by the dashed line. We use  $A(0) = 0$ ,  $B(0) = 0$ ,  $k_1 = 10^{-3} \text{ sec}^{-1}$ ,  $k_2 = 10^{-2} \text{ sec}^{-1}$ ,  $k_3 = 1.2 \text{ sec}^{-1}$  and  $k_4 = 1 \text{ sec}^{-1}$ .

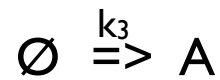
# Stationärer Zustand



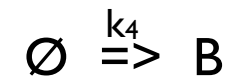
$$k_1 = 10^{-3} \text{ s}^{-1}$$



$$k_2 = 10^{-2} \text{ s}^{-1}$$



$$k_3 = 1.2 \text{ s}^{-1}$$



$$k_4 = 1 \text{ s}^{-1}$$

kontinuierliches Modell:

$$A_{ss} = 10, \quad B_{ss} = 10$$

$\Leftrightarrow$

Aus langen Gillespie-Simulationen:

$$A_{ss} = 9.6, \quad B_{ss} = 12.2$$

Erban, Chapman, Maini, arXiv:0704.1908v2

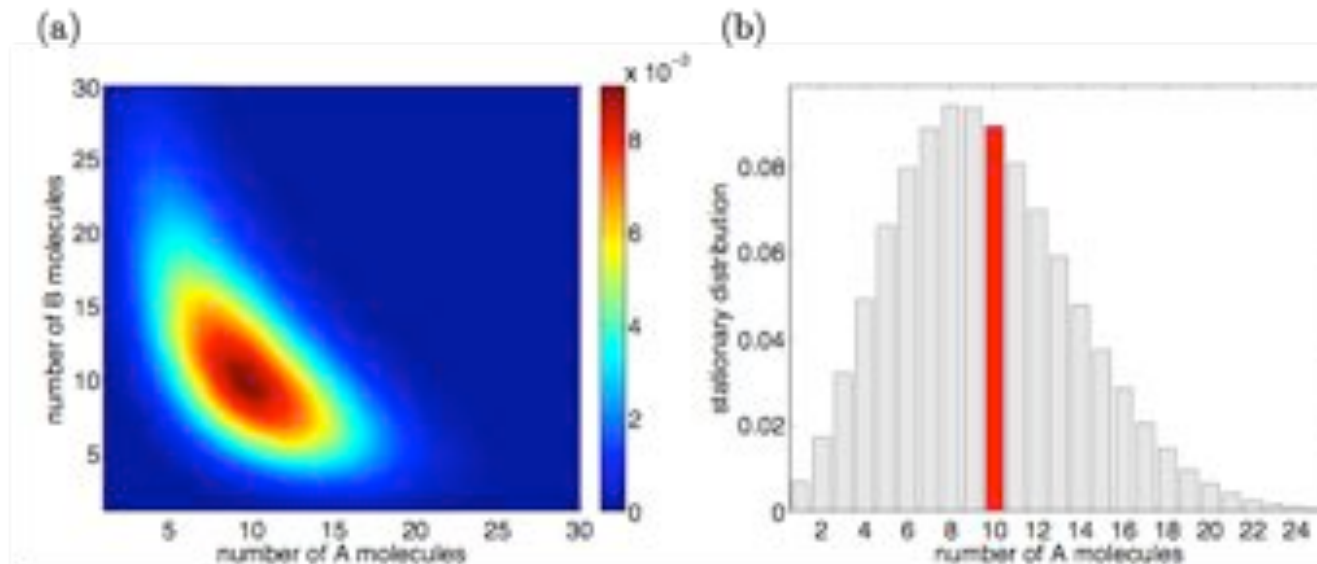
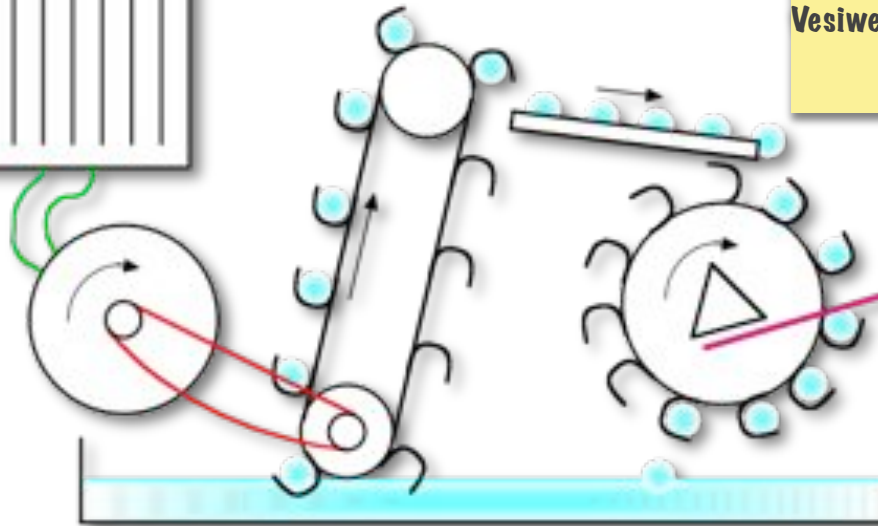
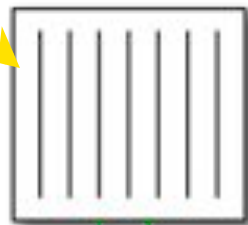


FIG. 2.4. (a) Stationary distribution  $\phi(n, m)$  obtained by long time simulation of (a4)–(d4) for  $k_1 = 10^{-3} \text{ sec}^{-1}$ ,  $k_2 = 10^{-2} \text{ sec}^{-1}$ ,  $k_3 = 1.2 \text{ sec}^{-1}$  and  $k_4 = 1 \text{ sec}^{-1}$ . (b) Stationary distribution of  $A$  obtained by (2.35).

# Photosynthese ist...

...die Umwandlung von Lichtenergie in chemische Energie  
(einer der wichtigsten Prozesse weltweit)



biologischer Überblick

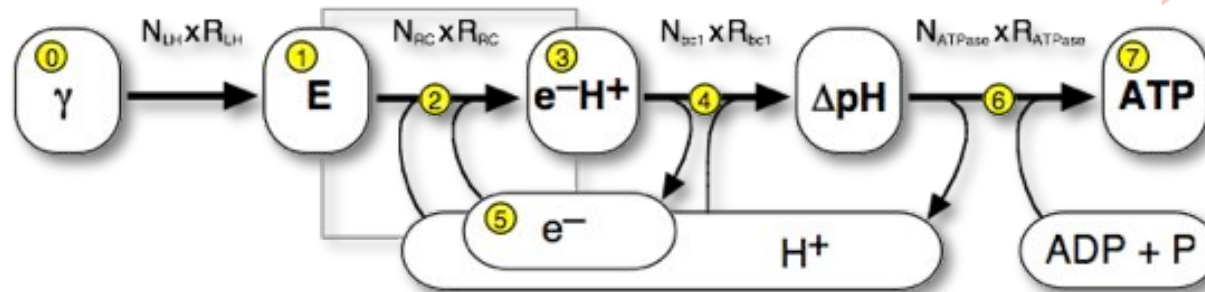
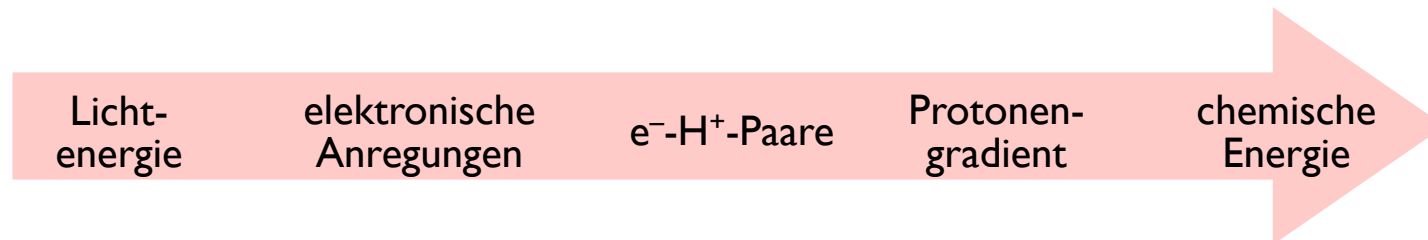
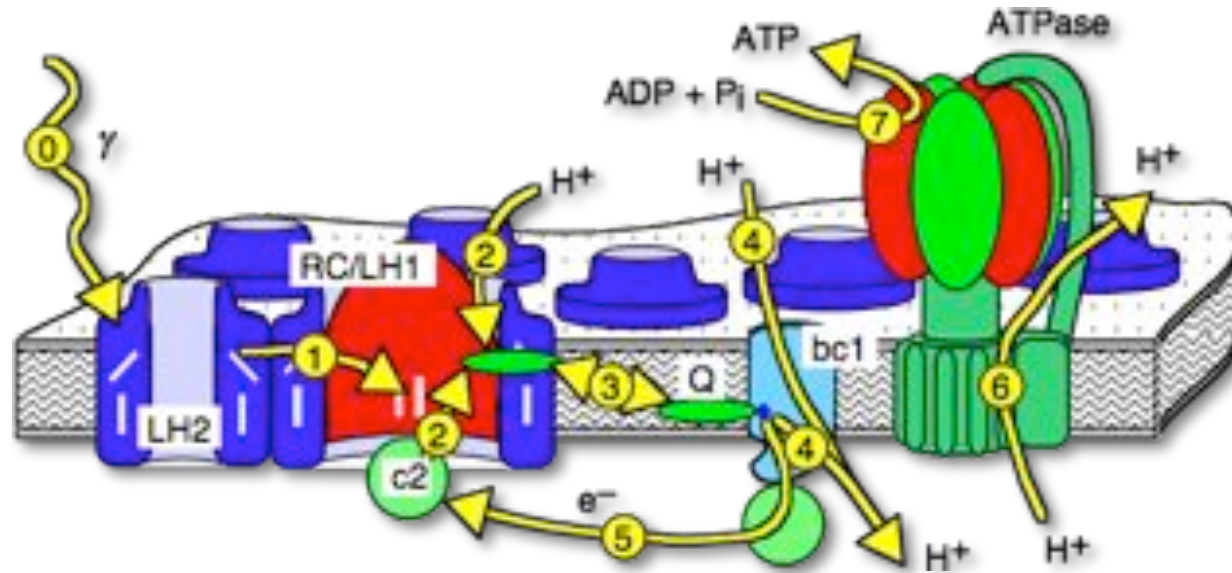
Stochastische Simulation aus einzelnen Reaktionen

=> Unterschiede zu Gillespie?

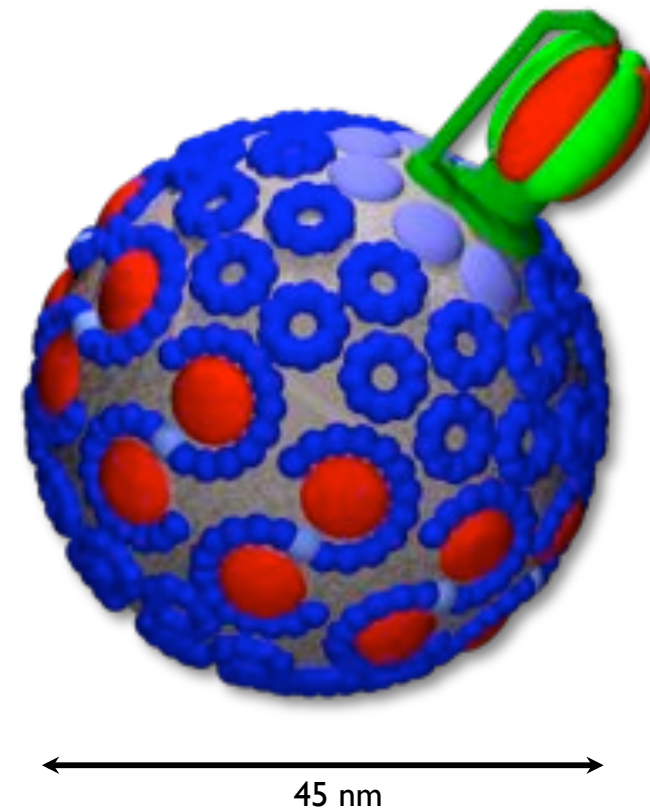
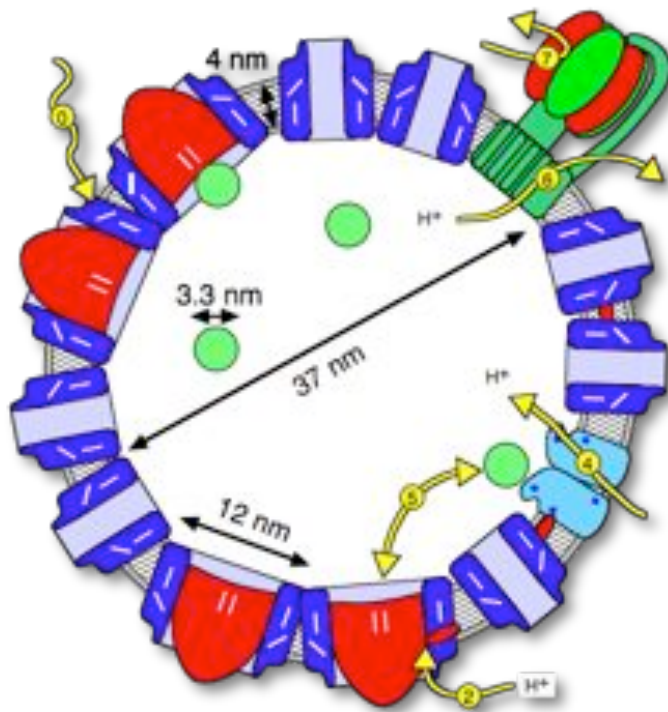
=> Reihenfolge von Näherung und Simulation

Vesiweb

# Photosynthese in *Rb. sphaeroides*



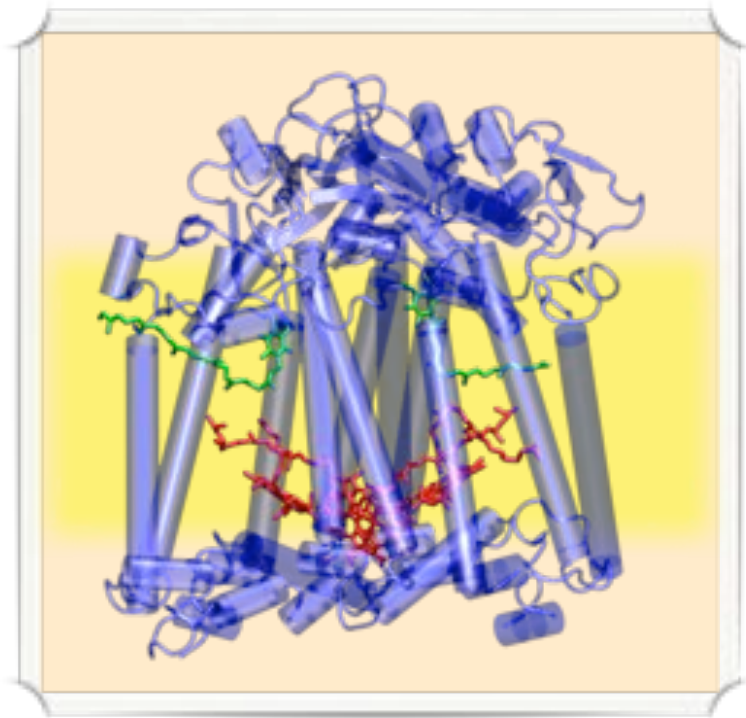
# Chromatophoren-Vesikel



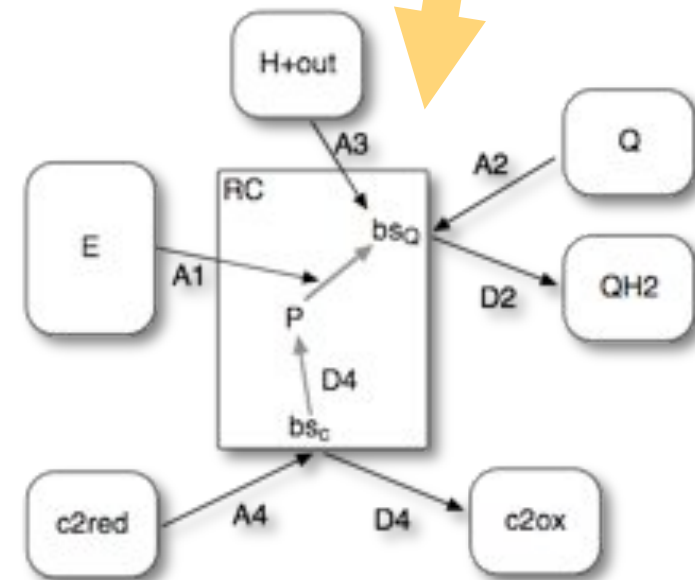
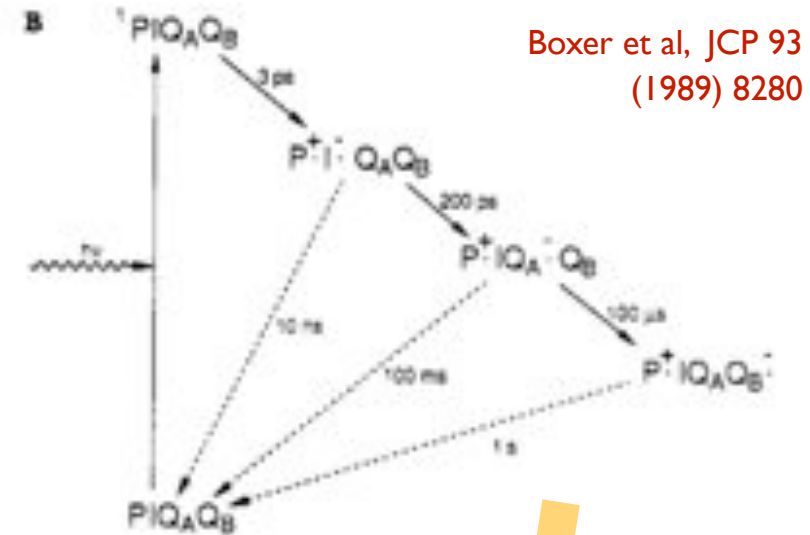
- => einfach: 4 Proteine + 2 Transporter +  $H^+$ ; Kristallstrukturen u. Reaktionen
- => klein: <30 Proteine + LHCs; alle Prozesse im Vesikel
- => abgeschlossen: definierte Randbedingungen für die Simulation
- => praktisch: Anregung mit Licht, spektroskopische Messungen

# Modellierung der Proteine

RC: Photon => Ladungstrennung



1AIJ.pdb



# Stochastische Reaktionen

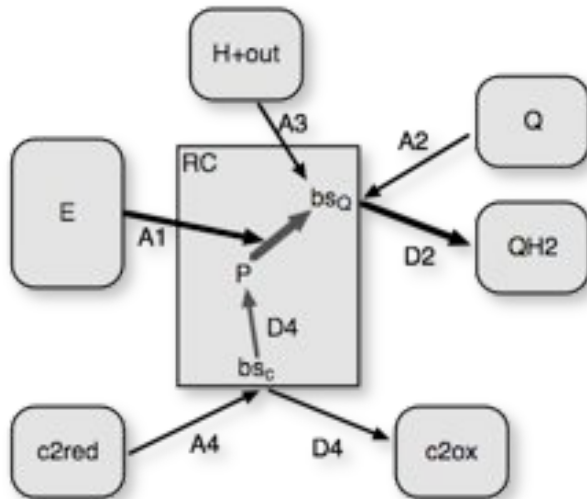
**Wenn BS frei ist => Assoziation möglich:  $BS + X \Rightarrow BS:X$**

1) sind alle Bedingungen erfüllt?

2) chemische Reaktionskinetik:

Reaktionsrate: 
$$\frac{d[BS:X]}{dt} = k_{on} [BS] [X]$$

Bindungs-WV.keit pro BS: 
$$P_{on} = k_{on} [X] \Delta t$$



```
for each timestep Δt:  
  for each reaction:  
    conditions fulfilled?  
    determine probability:  
    perform reaction
```

Geyer, Lauck, Helms, J Biotech 129 (2007) 212

# Bsp: Elektronentransfer im RC

```
// R1: transfers an electron to the Quinone
//      using the energy from an exciton{      if
(bs_Q && (e_P == 1) && (e_Q == 0) &&
      ((He_Q == 0) || (He_Q == 1)))
{
    if (LHPoolp->take_out(LH_kon))
    {
        e_P = 0;          e_Q = 1;
writeInternals();      }
    }
}
```

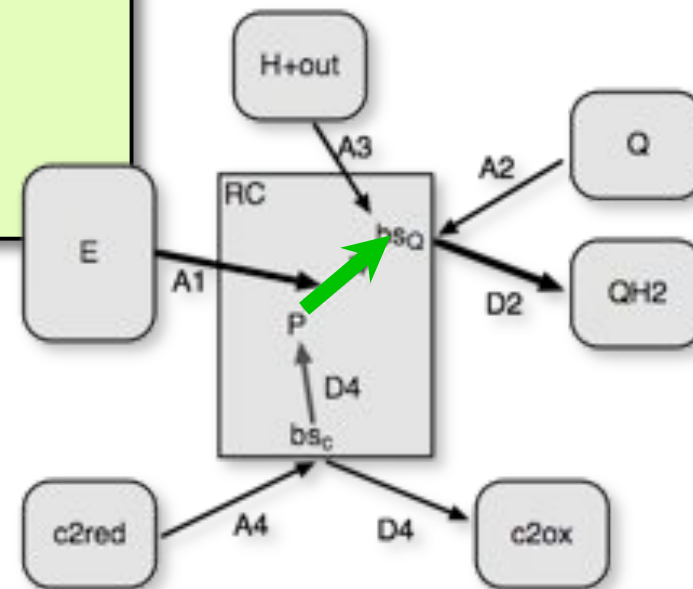
Bedingungen?

Wahrscheinlichkeit?

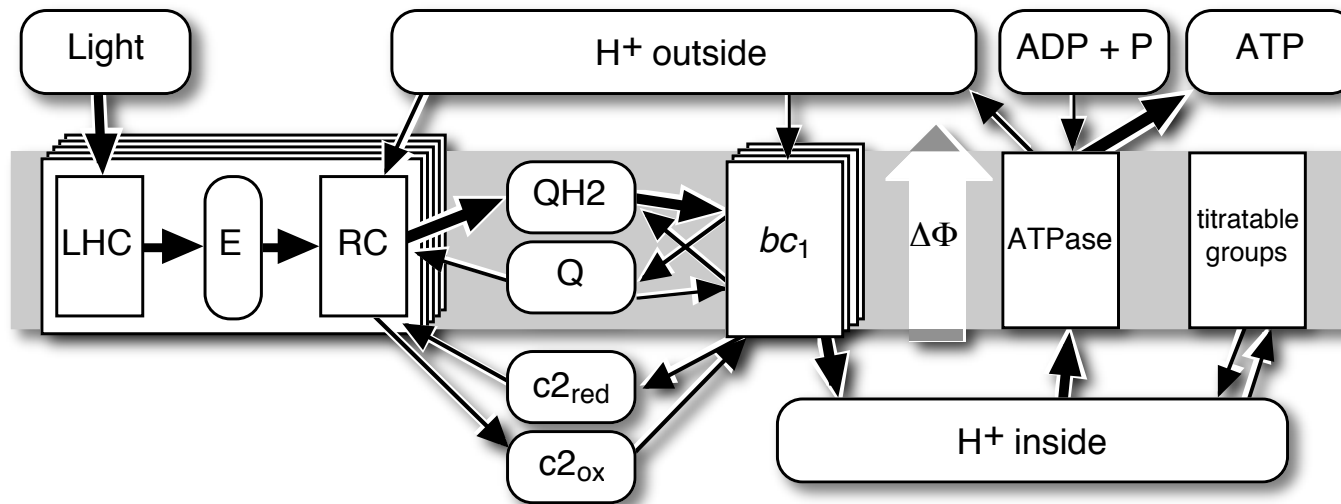
Reaktion!

Protein = {BS; Reaktionen(Zustand)}

Geyer, Lauck, Helms, J Biotech 129 (2007) 212



# "Pools-and-Proteins"–Modell



## 40 aktive Proteine

- unabhängig voneinander
- stochastische Reaktionen mit je 1 Molekül
- Anzahl wie auf dem Vesikel

## 19 passive Pools

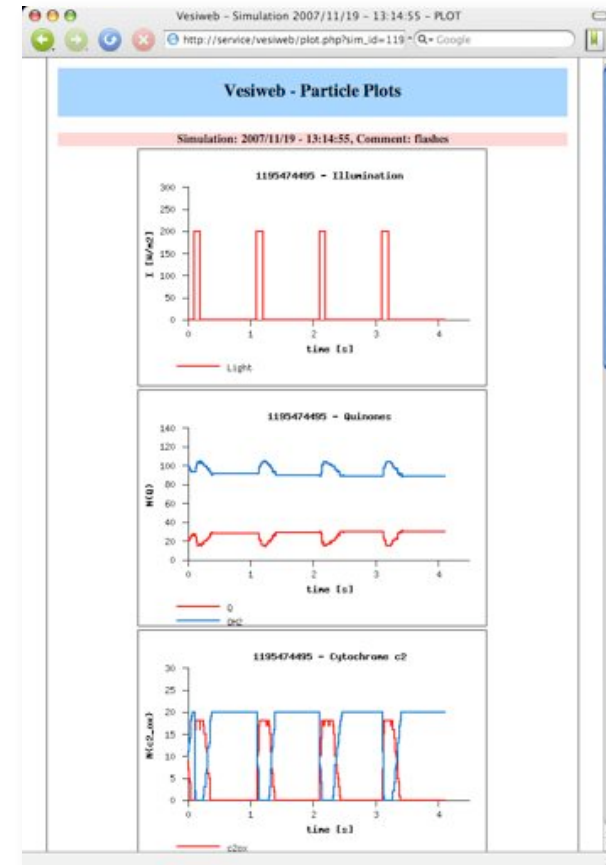
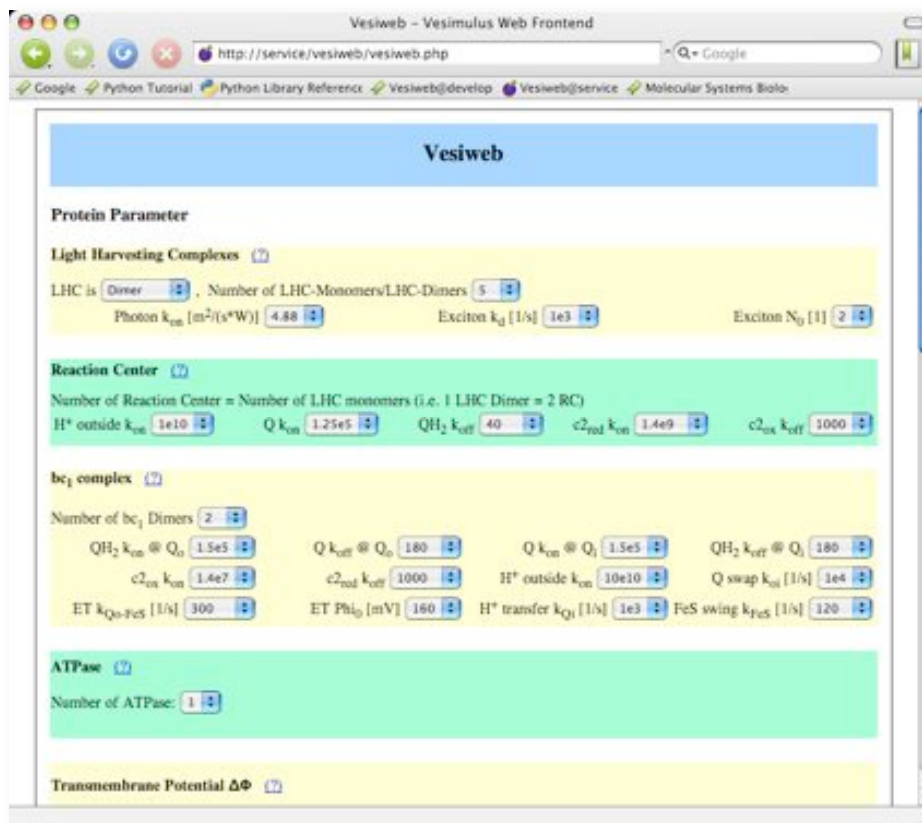
- ein Pool pro Metabolit
- hier: Diffusion ist schnell

Verbindungen definieren das biologische System  
=> Pfade als "emergent behavior"

Geyer, Lauck, Helms, J Biotech 129 (2007) 212

# Web Interface

Simulationen über Konfigurationsdatei oder web-interface @ [service.bioinformatik.uni-saarland.de/vesiweb](http://service.bioinformatik.uni-saarland.de/vesiweb)

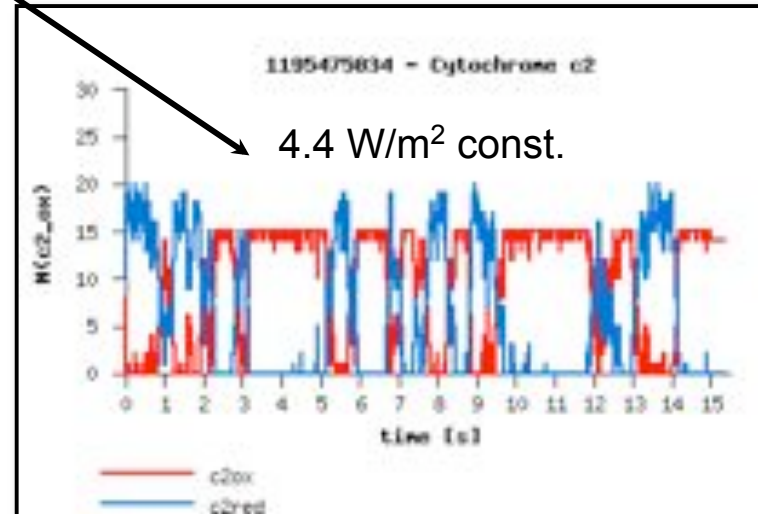
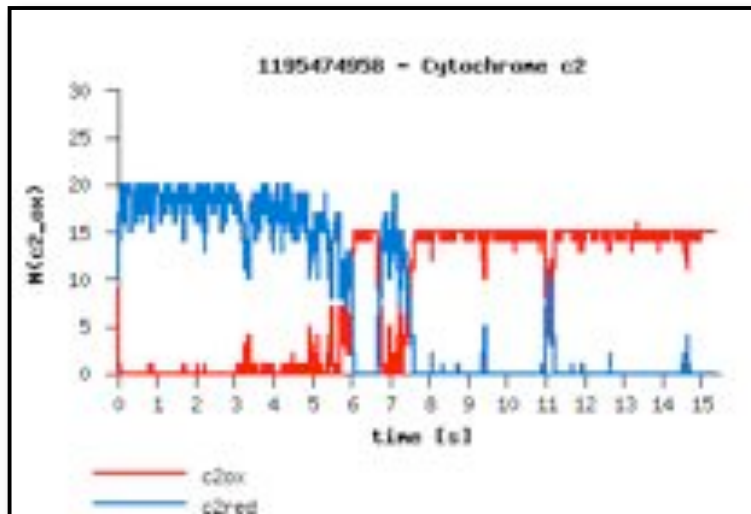
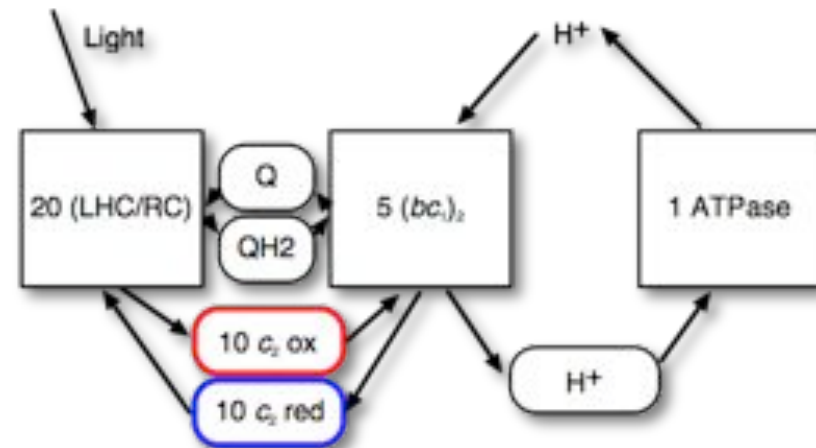
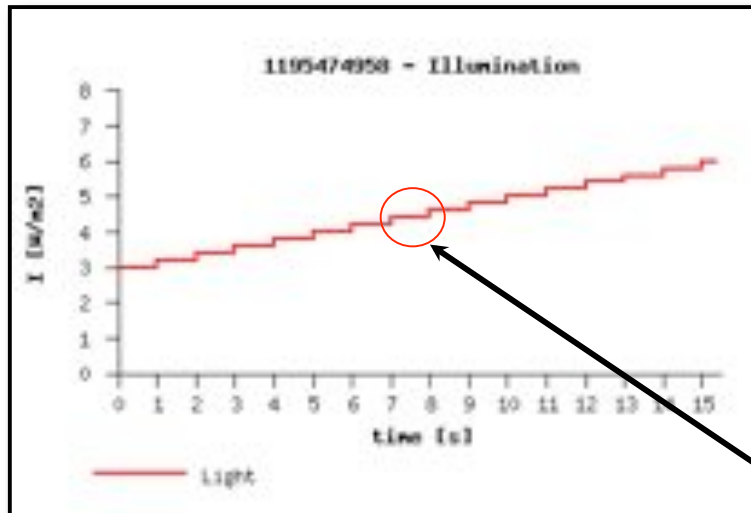


- Verstehen der Prozesse
- Modell-Verifikation + Parametrisierung gegen Experimente

[Florian Lauck. T.G., 2006]

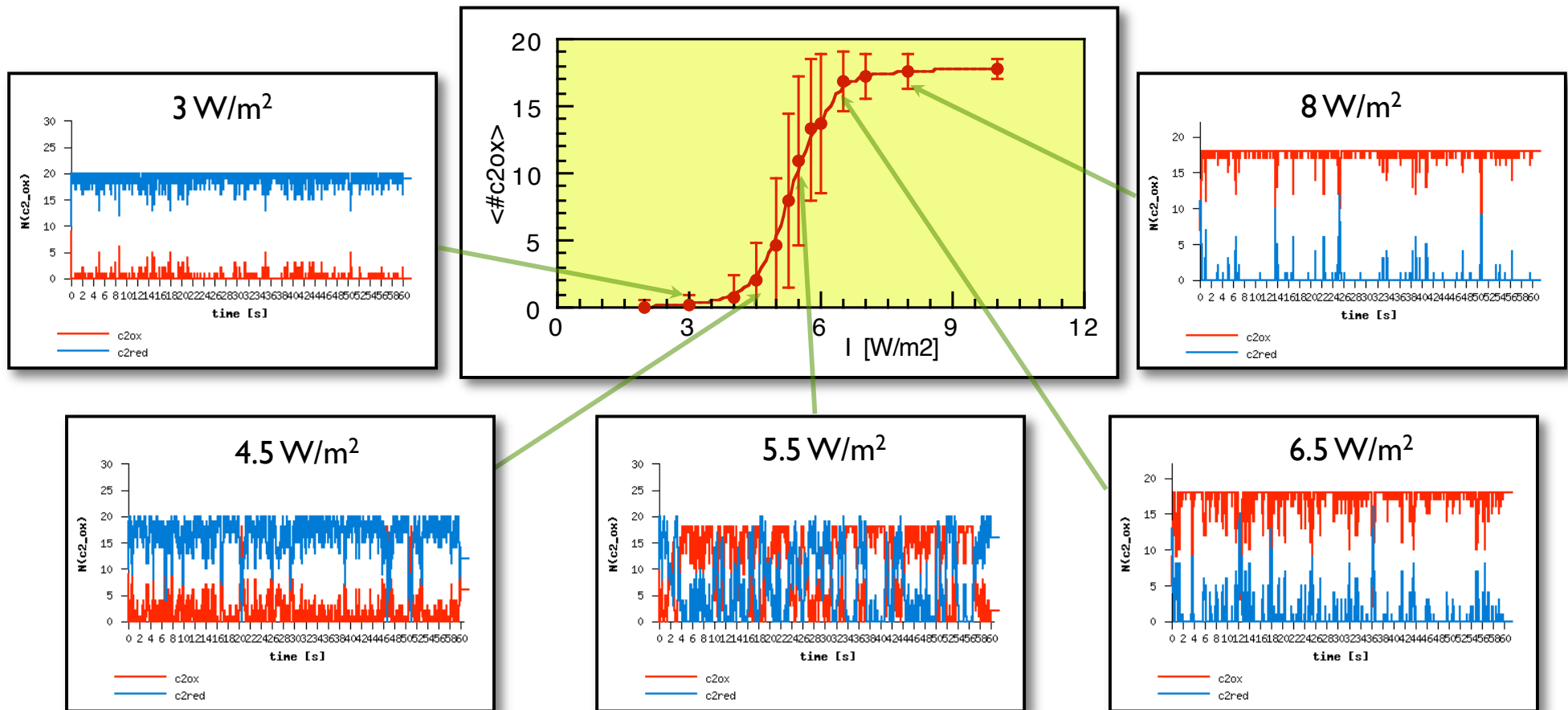
# Stochastische Effekte

Oxidationszustand des Cytochrom c-Pools bei kontinuierlicher Beleuchtung



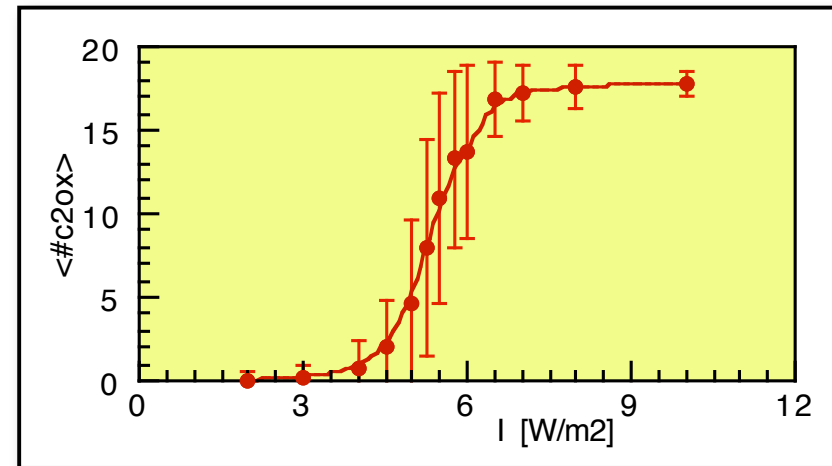
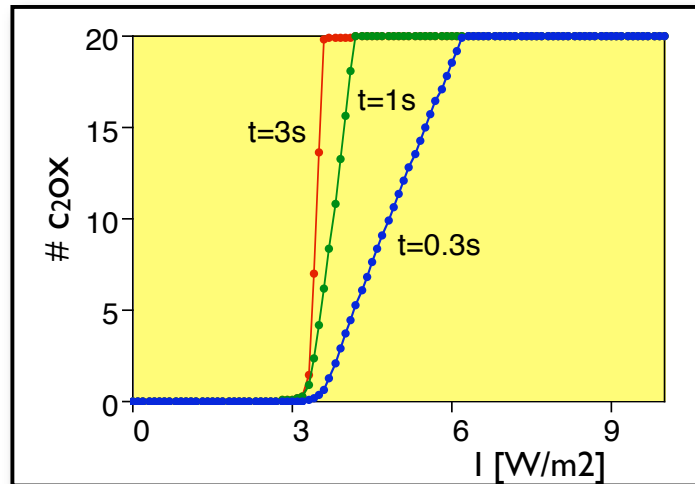
# Steady State $\Leftrightarrow$ Fluktuationen

60 Sek. bei konstanter Beleuchtung mit 10 RC/LHC-Dimeren und 4 bcI-Dimeren  
 $\Rightarrow$  Oxidationszustand des Cytochrom c-Pools



$\Rightarrow$  weicher Übergang mit starken Fluktuationen

# Deterministisch vs. Stochastisch



Gleichungen mitteln, dann simulieren  $\Leftrightarrow$  Simulieren, dann Ergebnisse mitteln

scharfer Übergang

$\Leftrightarrow$

weicher Übergang auch  
für lange Zeiten

nur numerische Unsicherheiten

$\Leftrightarrow$

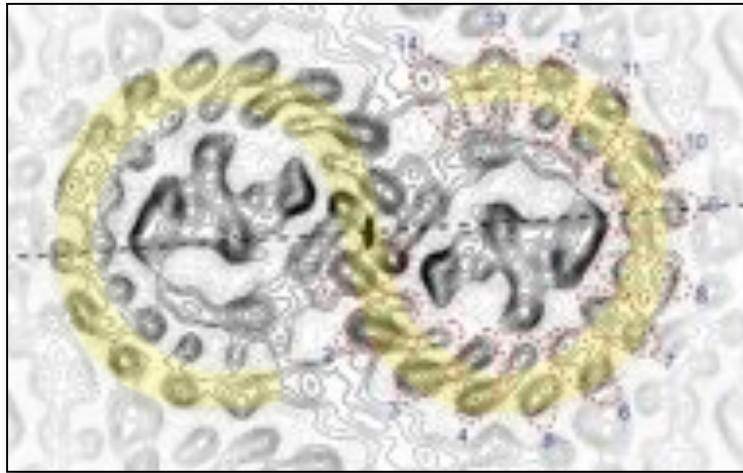
Fluktuationen  $\approx$  Signal

Reproduzierbare Werte

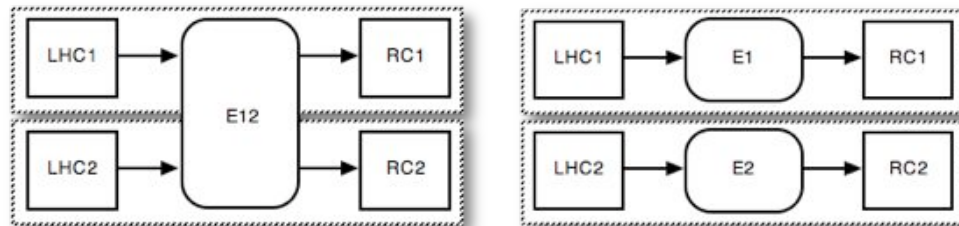
$\Leftrightarrow$

nur Mittelwert reproduzierbar

# Dimerisierung der LHI



Qian, Hunter, Bullough, JMB 349 (2005) 948

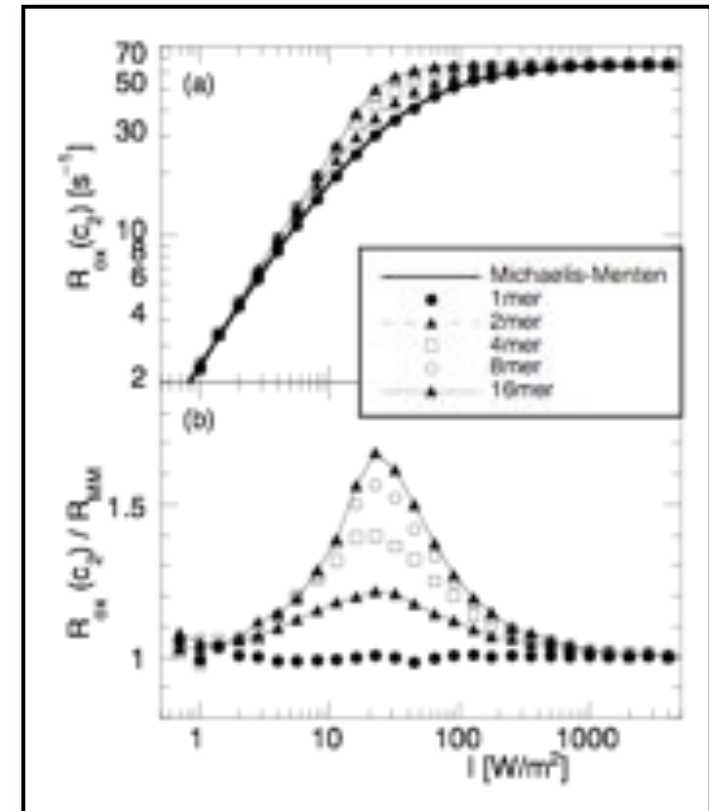


=> Effekt nur sichtbar bei unabhängigen, stochastisch modellierten Proteinen

zusätzlich: Dimer ist die effizientere Antenne

Geyer, Biophys J 93 (2007) 4374

I Antenne für 2 unabhängige RCs  
=> weniger Photonen werden abgewiesen  
=> Effizienz erhöht



Geyer, Lauck, Helms, J Biotech 129 (2007) 212

# Vesimulus: Warum nicht Gillespie?

## Gillespie:

hohe Anzahl von Spezies/Zuständen  
=> Komplexität!

Jeweils eine Reaktion

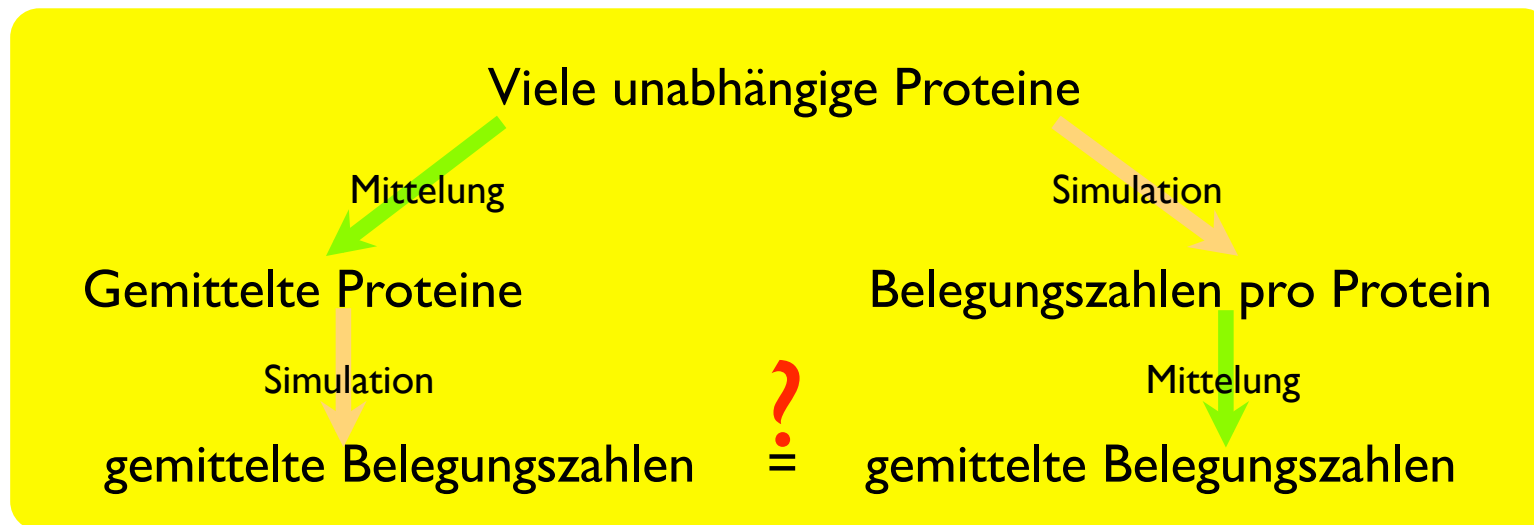
Numerisch effizient, exakt

## Pools-and-Proteins:

Belegung unabhängiger Bindungsstellen  
=> Zustände als Filter

Mehrere Reaktionen pro Zeitschritt

Natürliche Modellierung



# Zusammenfassung

**Stochastische Effekte** sind oft wichtig — wann?

Modellierung I: **Gillespie**

- numerisch effizient
- numerisch exakt
- biochemische Prozesse

=> stochastische Integration kontinuierlicher Ratengleichungen

Modellierung II: **Pools-and-Proteins**

- direkte Modellierung individueller Proteine
- Netzwerktopologie bleibt erhalten (auf allen Ebenen)

=> mikroskopisches Verhalten jedes einzelnen Proteins