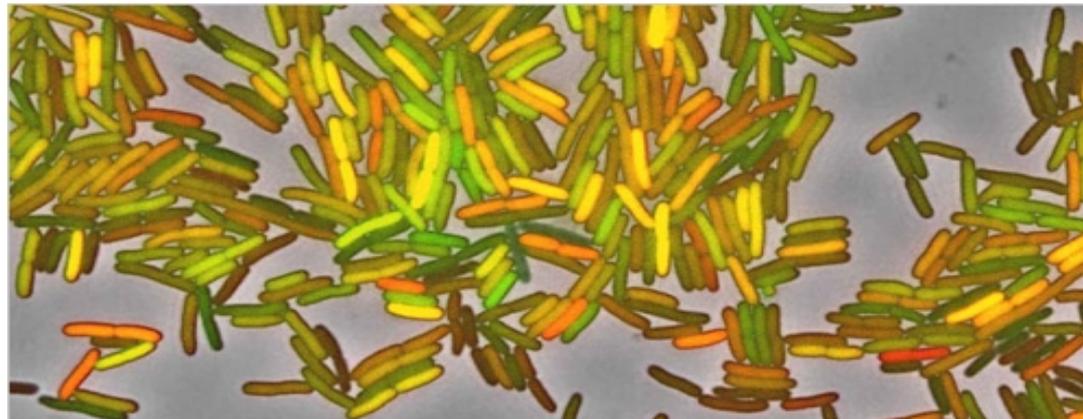
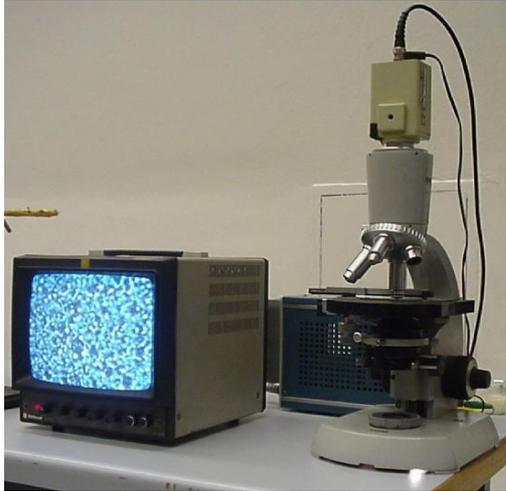
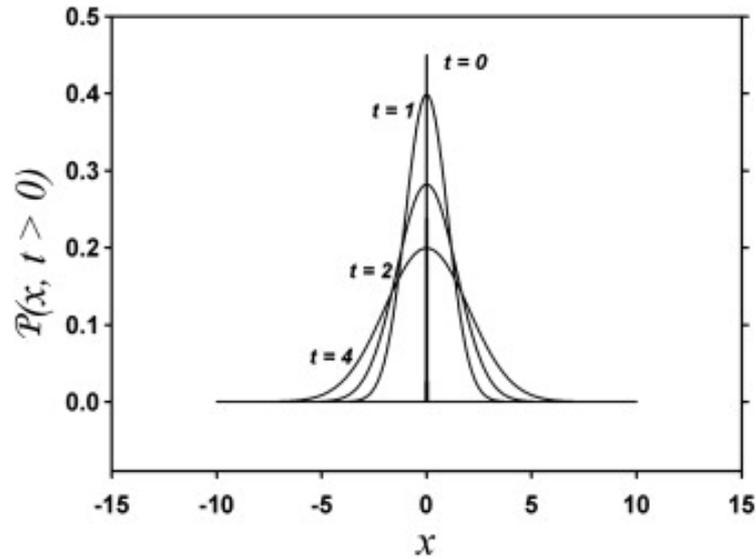


# Stochastische Simulationen

# Viele Vorgänge sind vom Zufall bestimmt



# Wie können wir komplizierte Zufallsvorgänge simulieren?



## Stochastische Prozesse

Wie simuliert man biochemische System stochastisch?

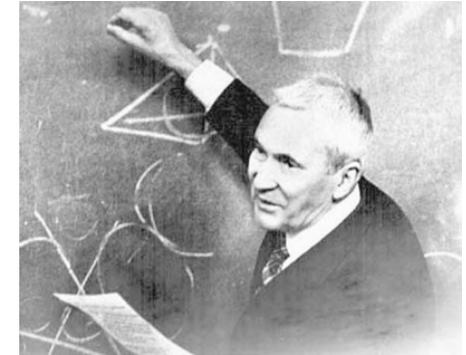
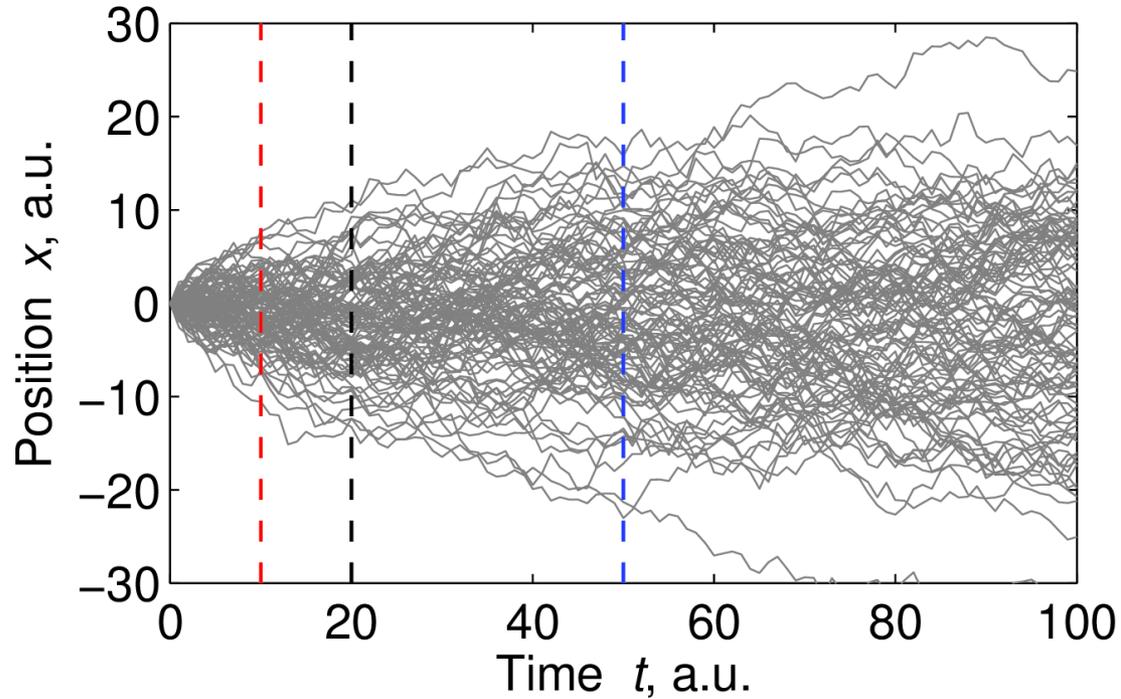
Schnelle Simulationsmethoden für große Systeme

## Stochastische Prozesse

Wie simuliert man biochemische System stochastisch?

Schnelle Simulationsmethoden für große Systeme

# Stochastische Prozesse beschreiben Geschichten und ihre Wahrscheinlichkeiten



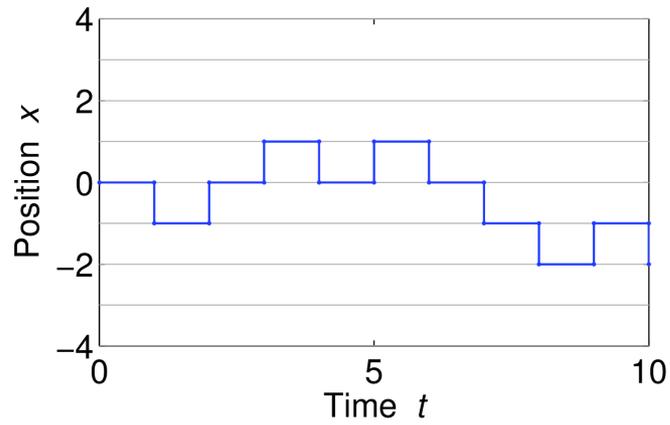
Andrej N. Kolmogorow

## Stochastischer Prozess

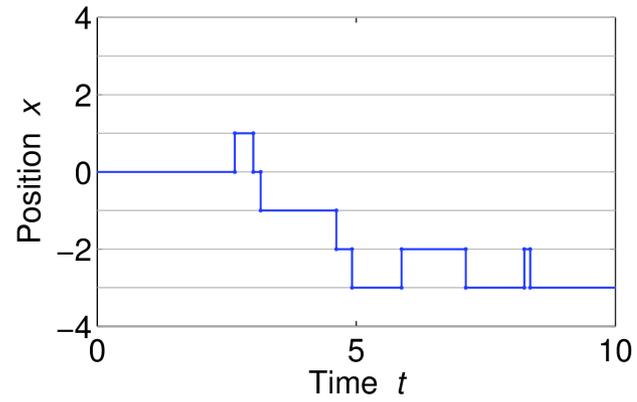
Eine Menge alternativer Geschichten (Zeitverläufe) versehen mit Wahrscheinlichkeiten

# Diskrete und kontinuierliche stochastische Prozesse

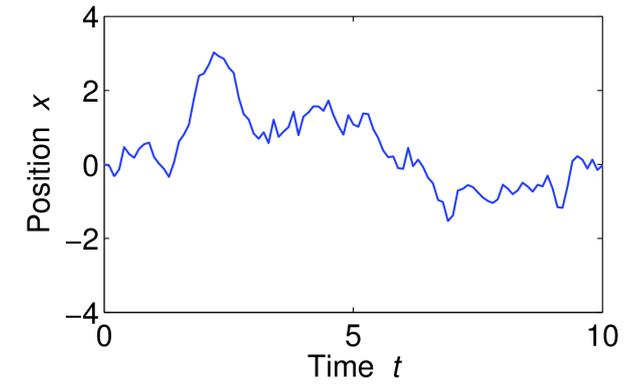
Diskrete Zeit



Kontinuierliche Zeit

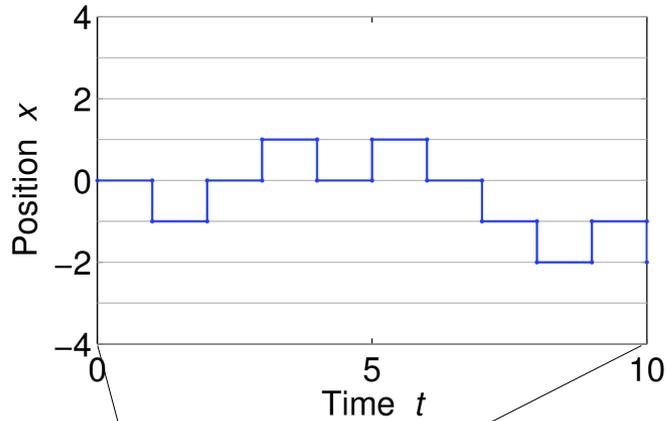


Kontinuierlicher Raum

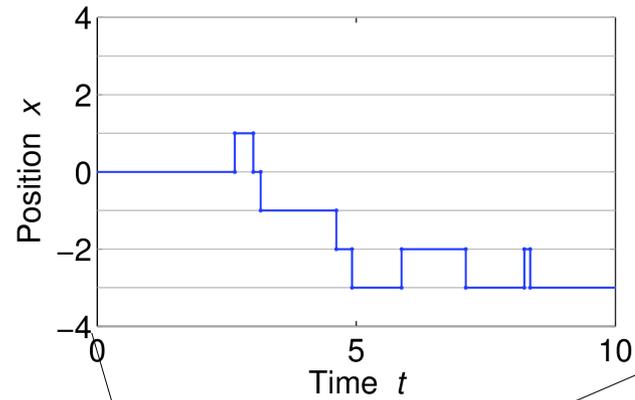


# Diskrete und kontinuierliche stochastische Prozesse

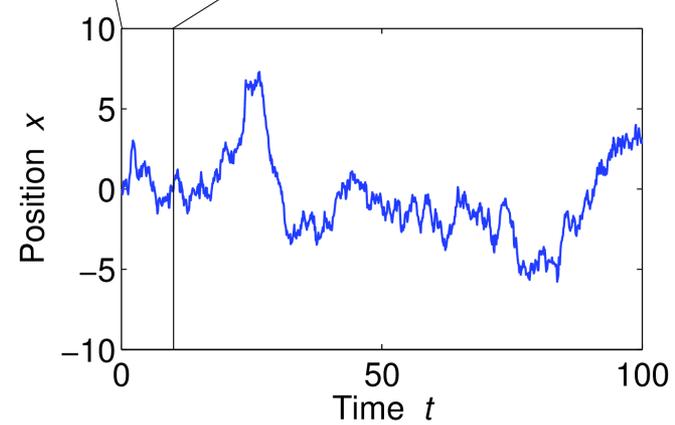
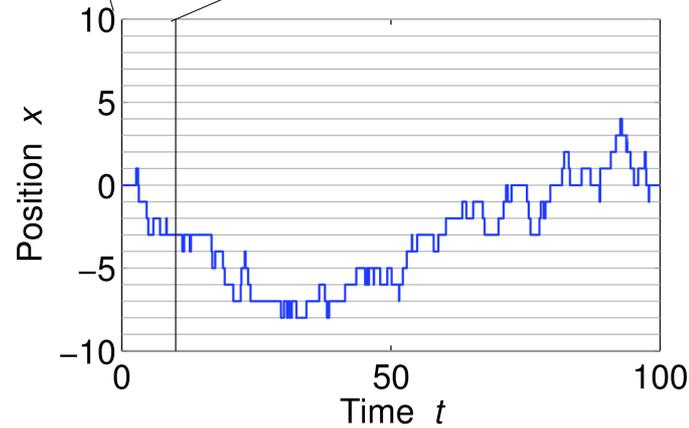
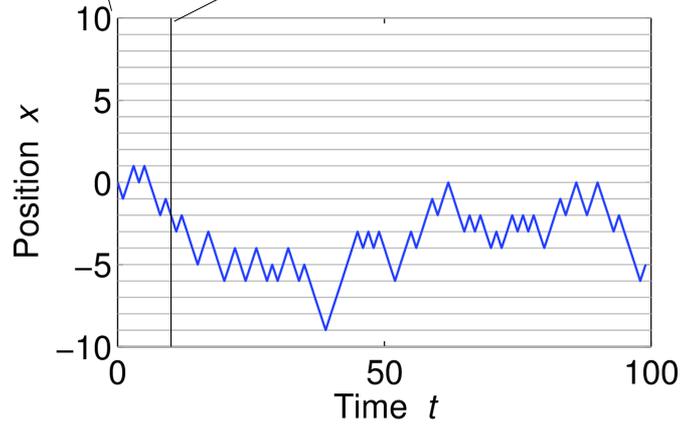
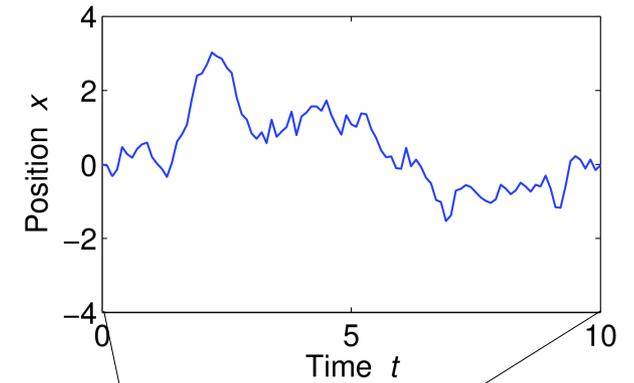
Diskrete Zeit



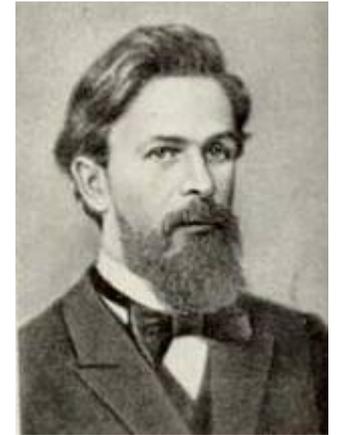
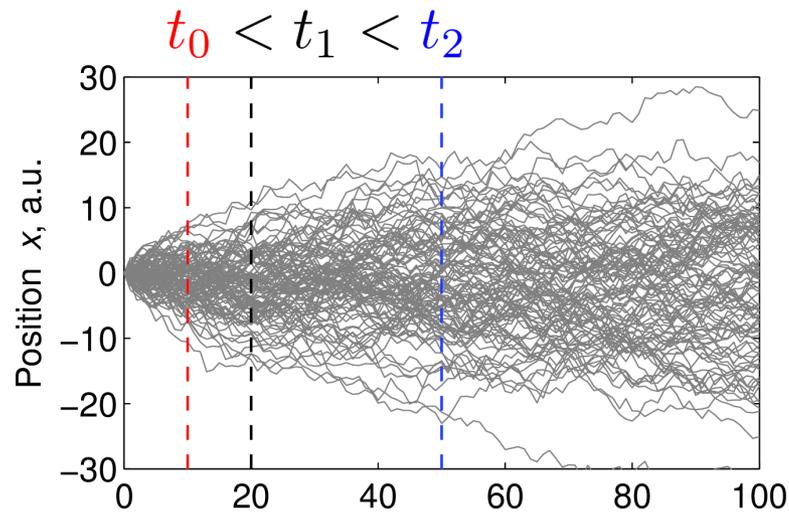
Kontinuierliche Zeit



Kontinuierlicher Raum



# Markow-Prozesse sind Prozesse "ohne Gedächtnis"



Andrej A. Markow

Übergangswahrscheinlichkeit  
(=bedingte Wahrscheinlichkeit)

$$P(x_2, t_2 | x_1, t_1) = \frac{P((x_2, t_2 | x_1, t_1))}{P(x_1, t_1)}$$

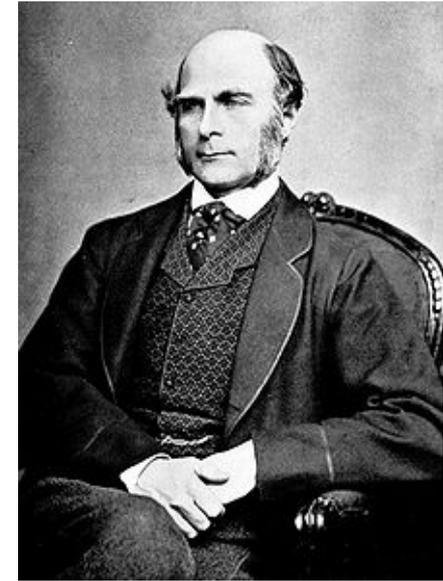
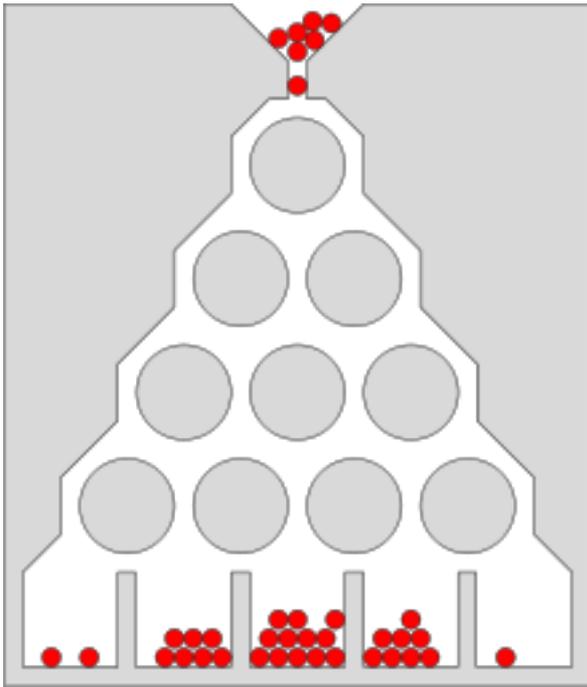
Unabhängigkeit von Vergangenheit:

$$P(x_2, t_2 | x_1, t_1) = P(x_2, t_2 | x_1, t_1, x_0, t_0)$$

Zukunft Gegenwart

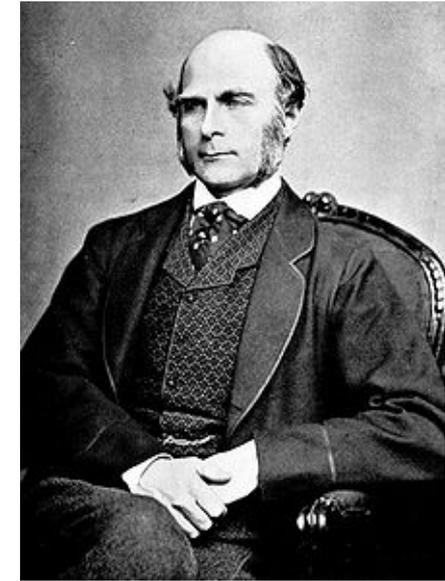
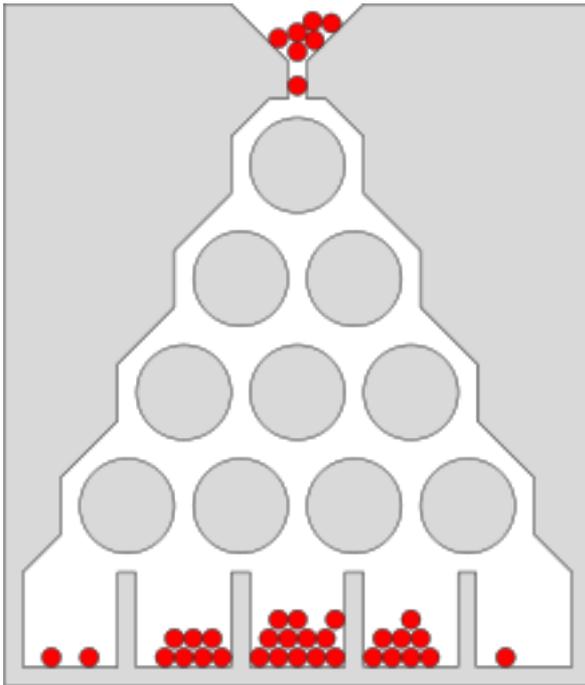
Vergangenheit

# Ein einfacher Markowprozess: der diskrete Random walk

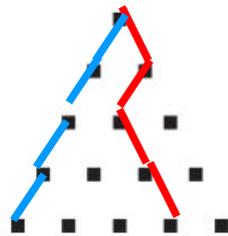


Sir Francis Galton  
(1822-1911)

# Ein einfacher Markowprozess: der diskrete Random walk



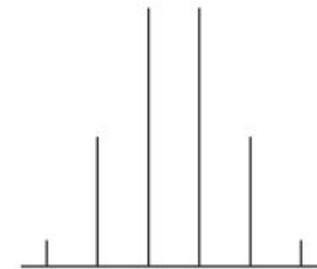
Sir Francis Galton  
(1822-1911)



$$P_{\text{Weg}} = \left(\frac{1}{2}\right)^t$$

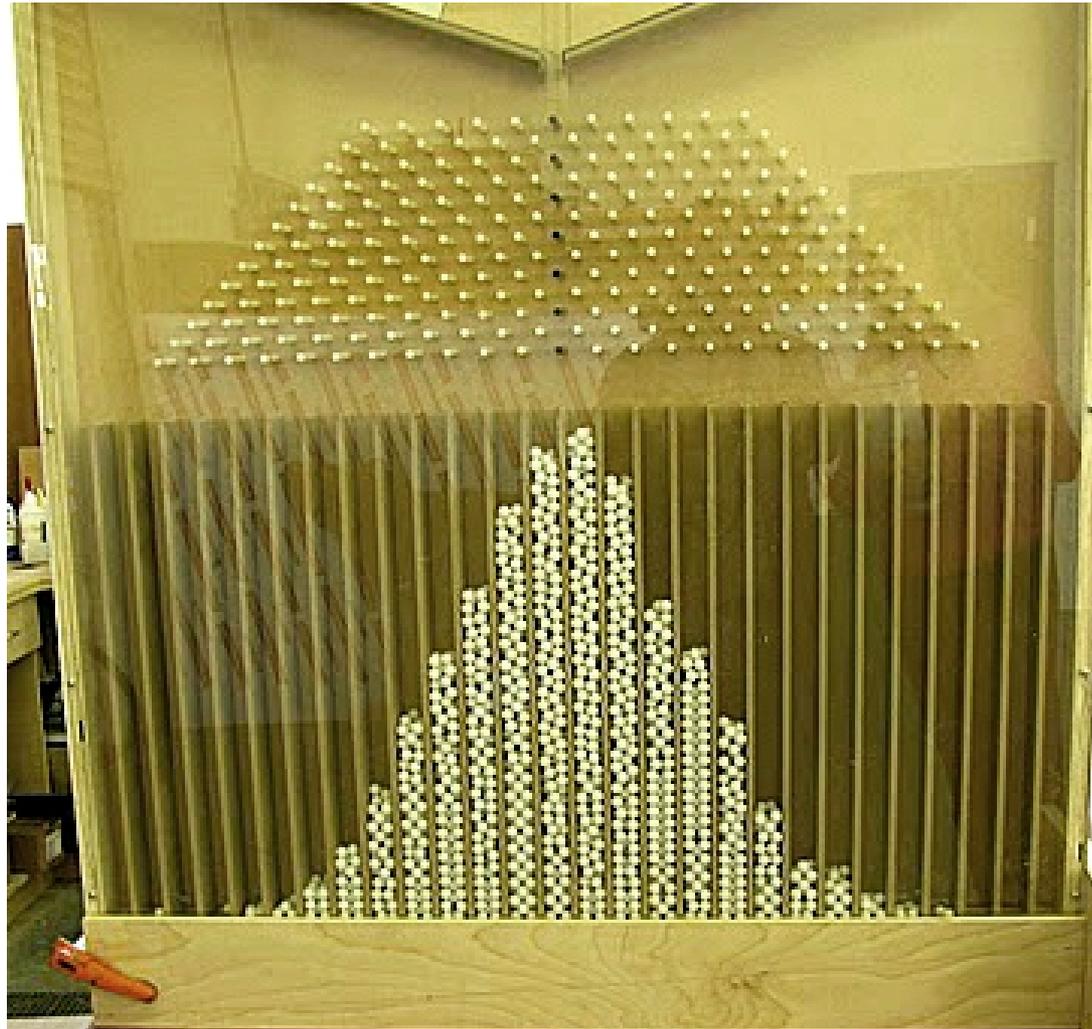
			1			
		1	1			
	1	2	1			
	1	3	3	1		
1	4	6	4	1		
1	5	10	10	5	1	

$$N = \binom{t}{x}$$



$$P(x, t) = \binom{t}{x} \left(\frac{1}{2}\right)^t$$

# Ein einfacher Markowprozess: der diskrete Random walk



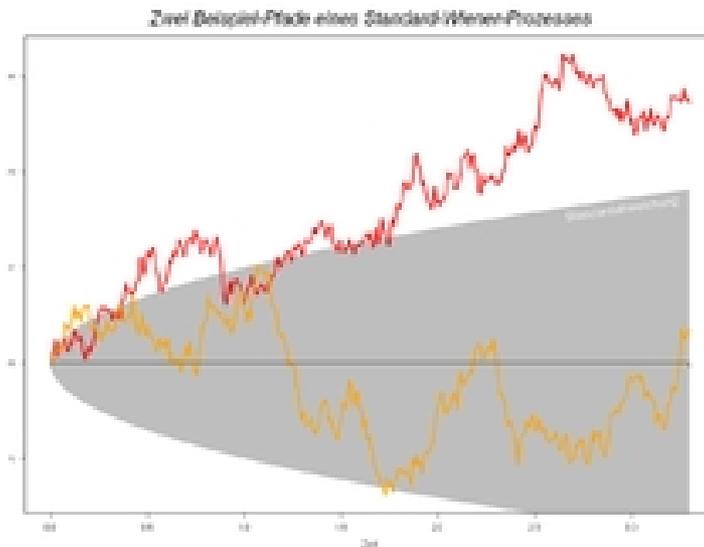
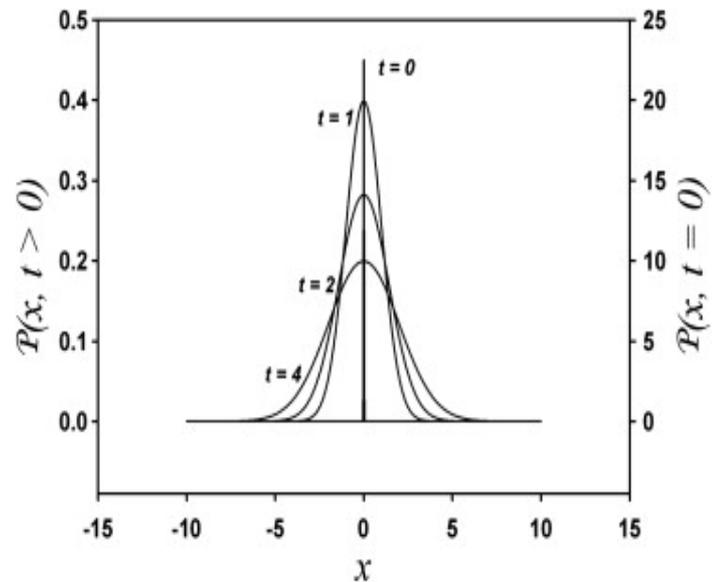
# Der Wienerprozess beschreibt die Brownsche Bewegung



Albert Einstein  
(1879-1955)



Norbert Wiener  
(1894-1964)

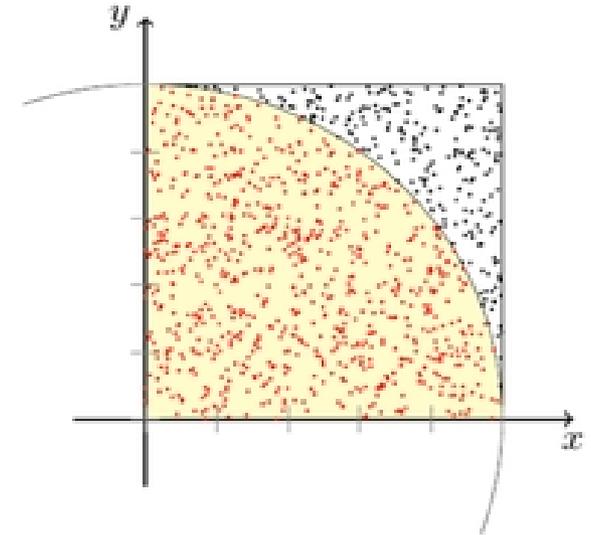


# Monte-Carlo-Methoden erlauben die Berechnung von Integralen und Wahrscheinlichkeiten

## Flächenbestimmung durch Monte-Carlo-Sampling

Relative Fläche  $\pi/4 \approx [\text{Punkte im Viertelkreis}] / [\text{Punkte im Quadrat}]$

- Punkte aus Pseudozufallszahlen
- “Wahrscheinlichkeitsexperiment” → Streuung der Ergebnisse
- Schätzwerte sind binomialverteilt!

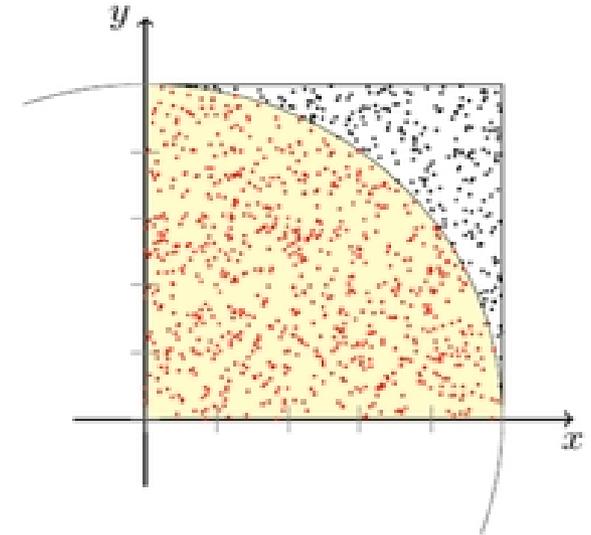


# Monte-Carlo-Methoden erlauben die Berechnung von Integralen und Wahrscheinlichkeiten

## Flächenbestimmung durch Monte-Carlo-Sampling

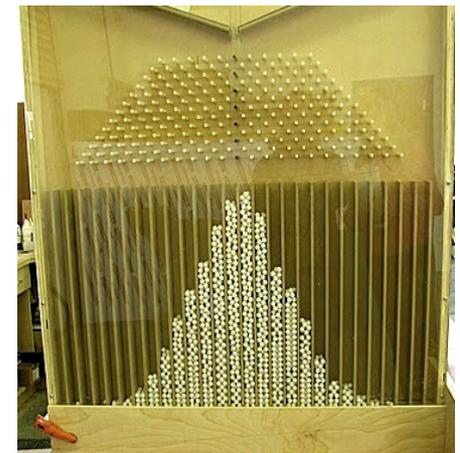
Relative Fläche  $\pi/4 \approx [\text{Punkte im Viertelkreis}] / [\text{Punkte im Quadrat}]$

- Punkte aus Pseudozufallszahlen
- “Wahrscheinlichkeitsexperiment” → Streuung der Ergebnisse
- Schätzwerte sind binomialverteilt!



## Monte-Carlo-Simulationen in der Stochastik

- Verteilung wird genähert durch große Menge von “Geschichten”
- Statistik über Prozess? → Statistik über Realisierungen!
- Sehr verbreitete Methode (Börsenkurse, Meteorologie, Bayes-Statistik)



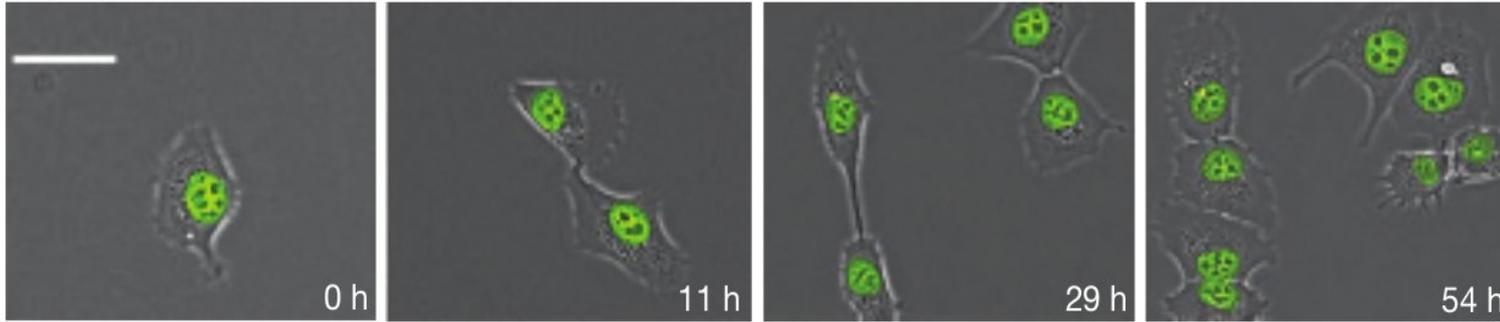
## Stochastische Prozesse

Wie simuliert man biochemische Systeme stochastisch?

Schnelle Simulationsmethoden für große Systeme

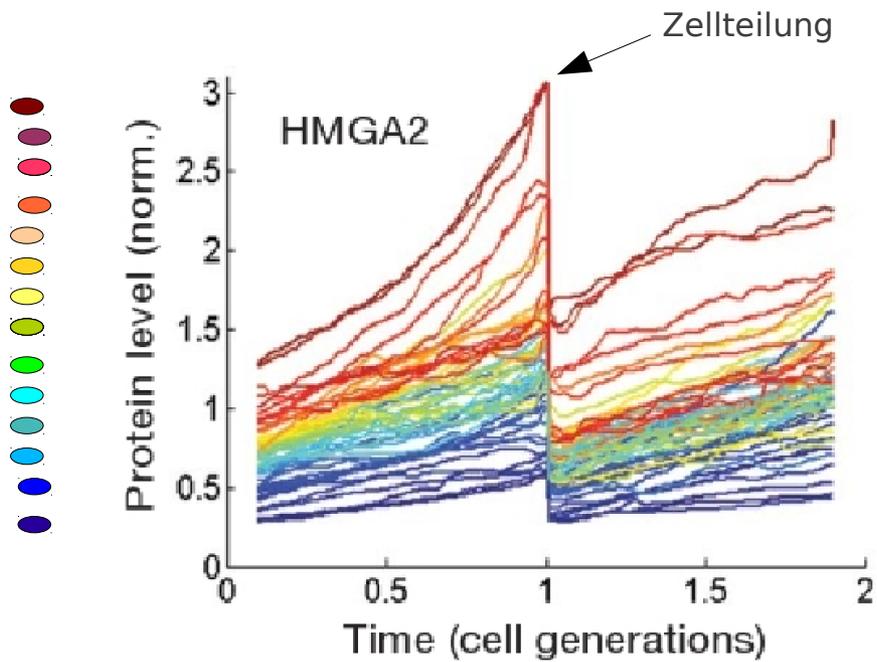
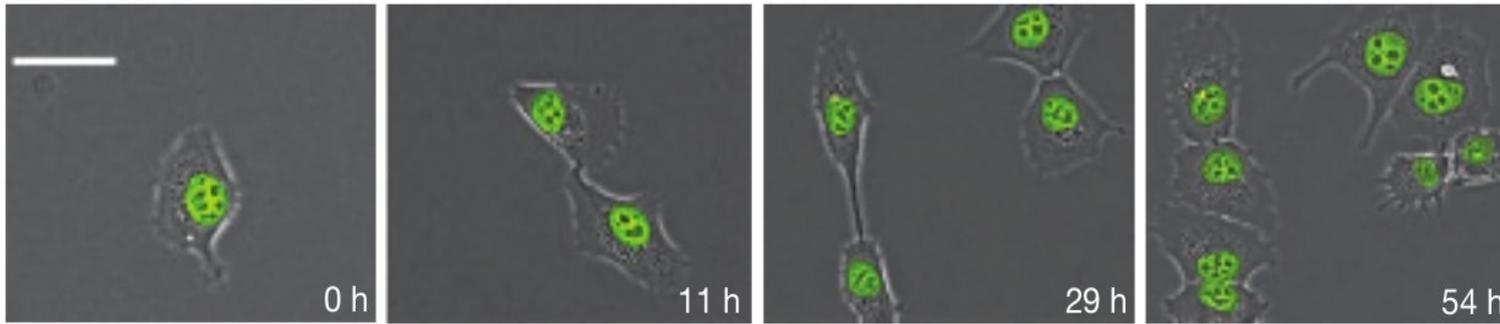
# Wie fluktuieren die Proteinmengen in Zellen?

Topoisomerase 1 in Krebszellen (YFP-markiert)



# Wie fluktuieren die Proteinmengen in Zellen?

Topoisomerase 1 in Krebszellen (YFP-markiert)

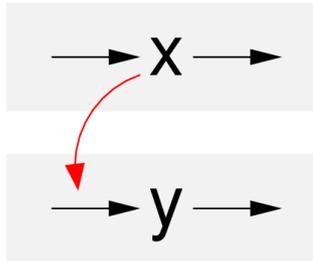


# Ratengleichungen beschreiben eine deterministische Dynamik

## Annahme:

Systemzustand (Konzentrationen) entwickelt sich kontinuierlich und deterministisch  
Das heißt: Momentanzustand legt zukünftiges Verhalten fest.

“Genaue” Raten liefern deterministische Ratengleichungen



## Ratengleichungen:

$$\frac{dc_x}{dt} = v_0 - \mu_x c_x$$

$$\frac{dc_y}{dt} = k_y c_x - \mu_y c_y$$

## Allgemeine Ratengleichung

(gewöhnliche Differentialgleichung):

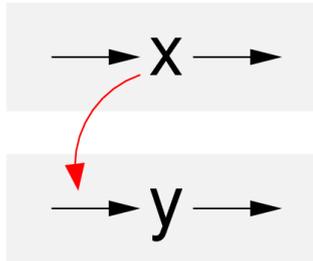
$$\frac{dc_i(t)}{dt} = \sum_l n_{il} v_l(\mathbf{c})$$

# Ratengleichungen beschreiben eine deterministische Dynamik

## Annahme:

Systemzustand (Konzentrationen) entwickelt sich kontinuierlich und deterministisch  
Das heißt: Momentanzustand legt zukünftiges Verhalten fest.

“Genaue” Raten liefern deterministische Ratengleichungen



## Ratengleichungen:

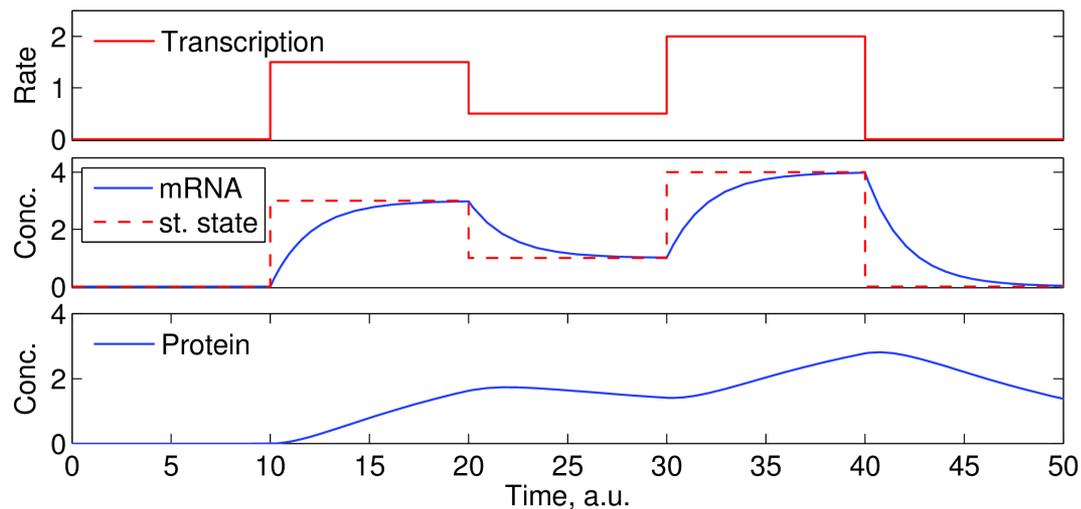
$$\frac{dc_x}{dt} = v_0 - \mu_x c_x$$

$$\frac{dc_y}{dt} = k_y c_x - \mu_y c_y$$

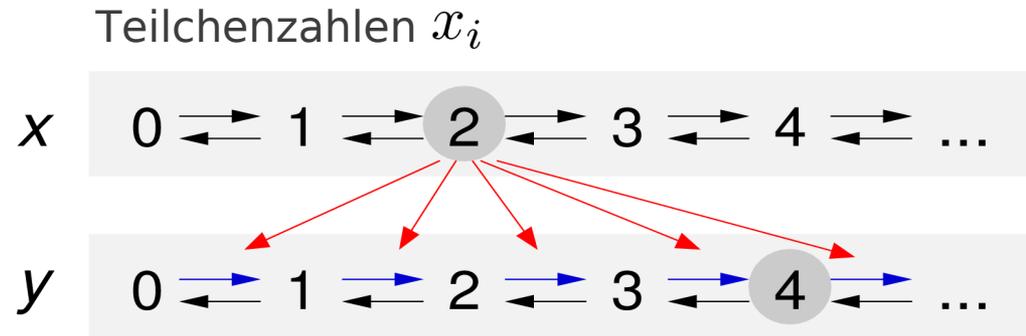
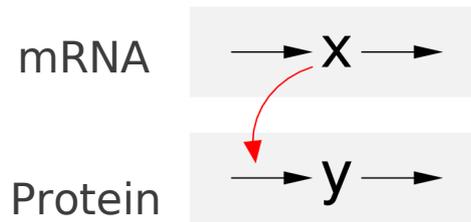
## Allgemeine Ratengleichung

(gewöhnliche Differentialgleichung):

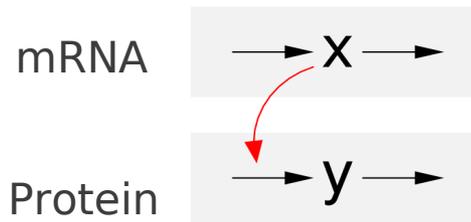
$$\frac{dc_i(t)}{dt} = \sum_l n_{il} v_l(\mathbf{c})$$



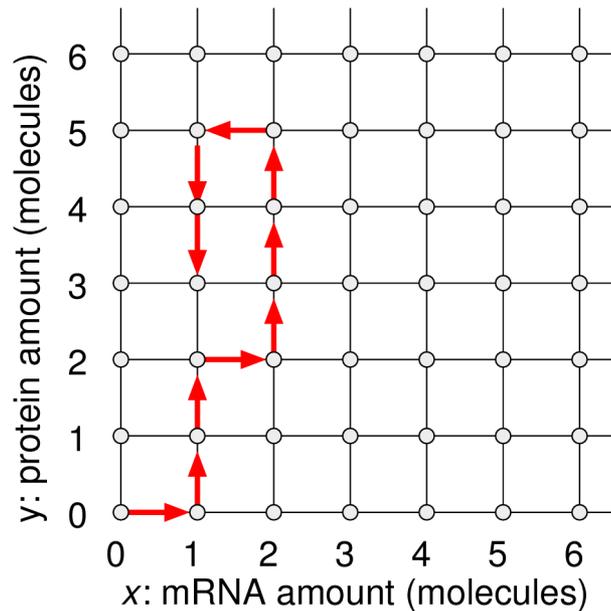
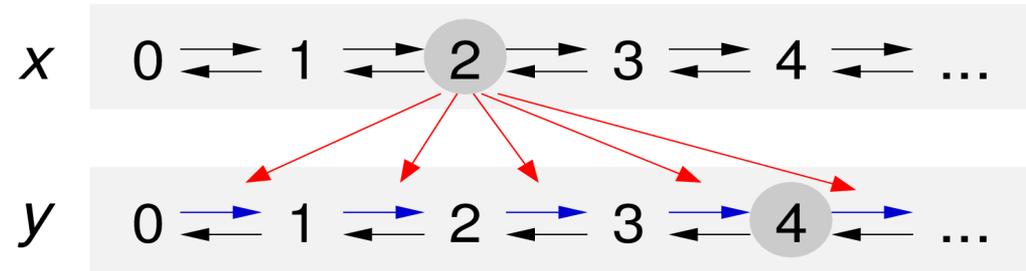
# Zufallsprozesse lassen die Teilchenzahlen schwanken



# Zufallsprozesse lassen die Teilchenzahlen schwanken



Teilchenzahlen  $x_i$

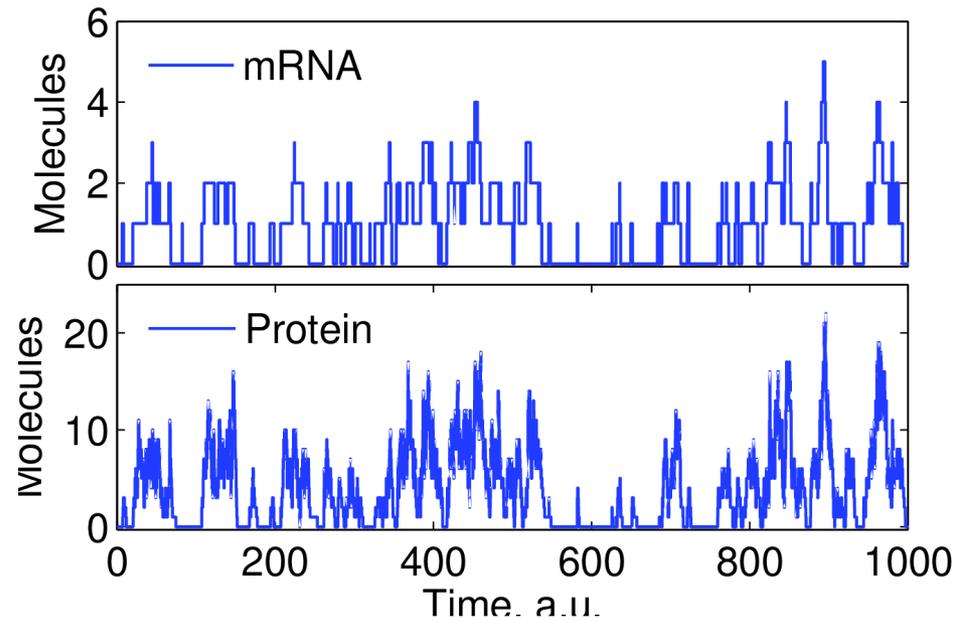
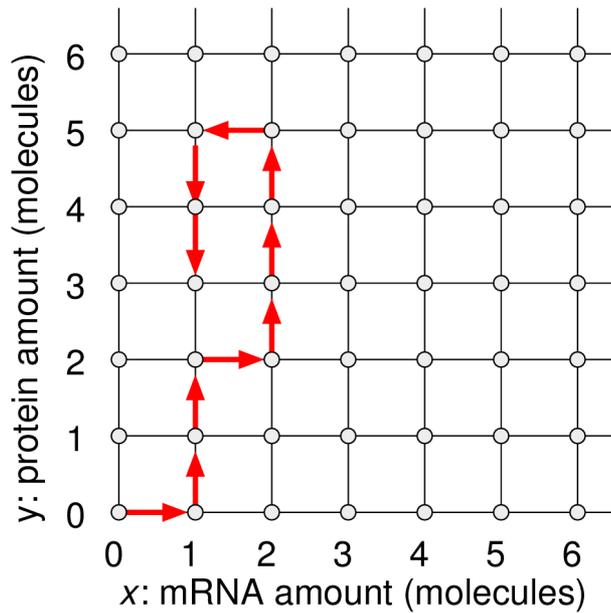
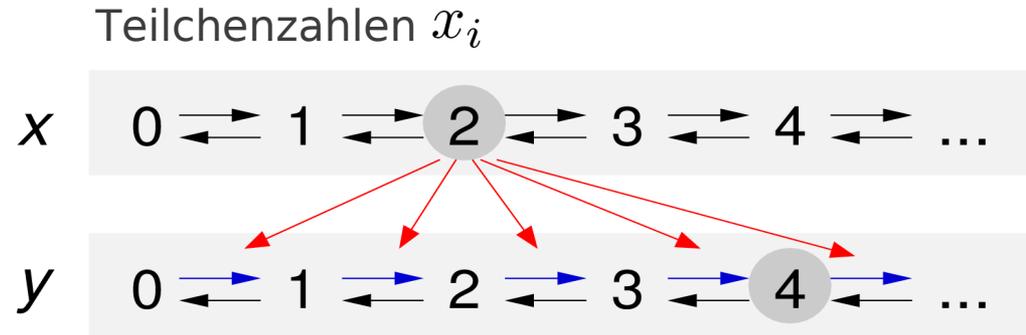
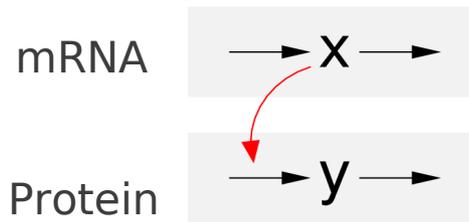


Reaktionsraten:

- Transkription  $a_1 = \omega_1$
- Abbau mRNA  $a_2 = \omega_2 x$
- Translation  $a_3 = \omega_3 x$
- Abbau Protein  $a_4 = \omega_4 y$

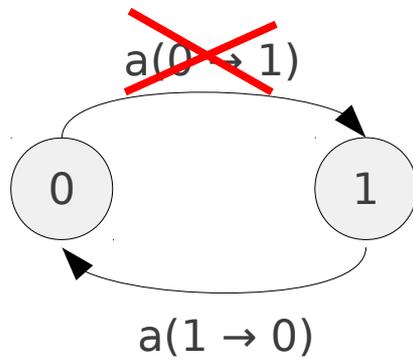
mit festen Einzelraten  $\omega_l$  pro Molekül

# Zufallsprozesse lassen die Teilchenzahlen schwanken



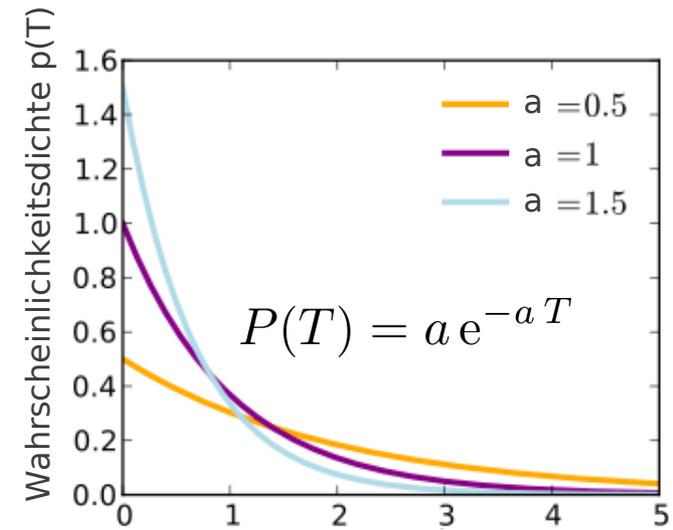
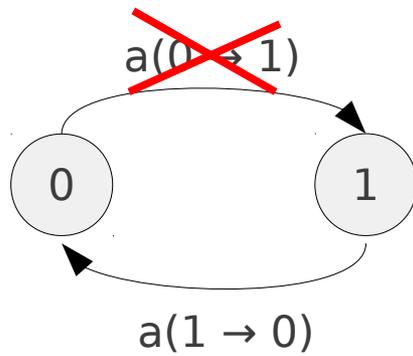
# Wie lange dauert es bis zur nächsten Reaktion?

Zweizustandssystem



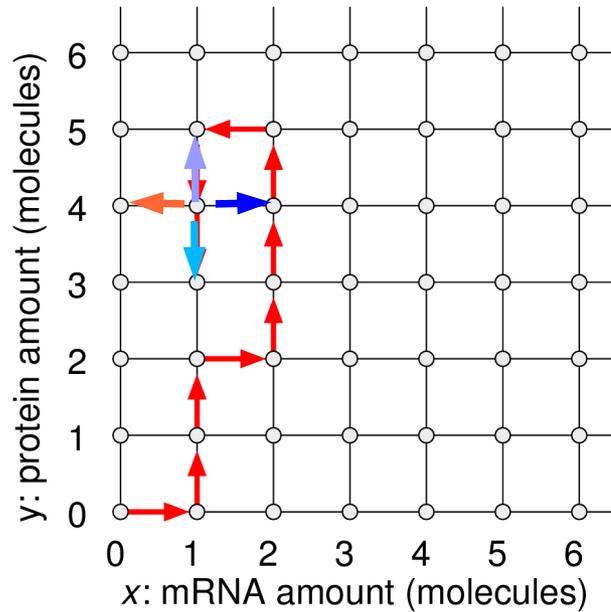
# Wie lange dauert es bis zur nächsten Reaktion?

Zweizustandssystem





# Wie bestimmt man die nächste zufällige Reaktion?

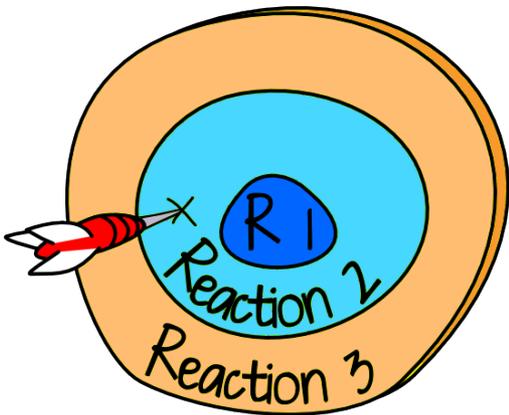


1. **Gesamtübergangsrate** ergibt sich aus Summe  $a = \sum_l a_l$  der einzelnen Übergangsraten

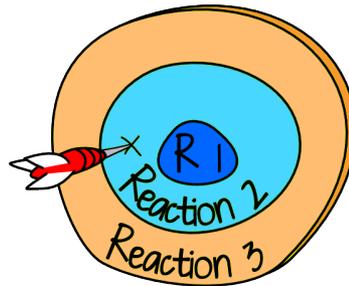
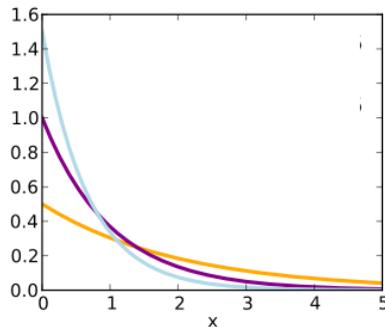
→ Wartezeit  $T$  exponentialverteilt mit Parameter  $a$

2. **Welcher Übergang** findet tatsächlich statt?

Einzelwahrscheinlichkeiten  $P(l) = \frac{a_l}{a}$



# Der “Gillespie”-Algorithmus



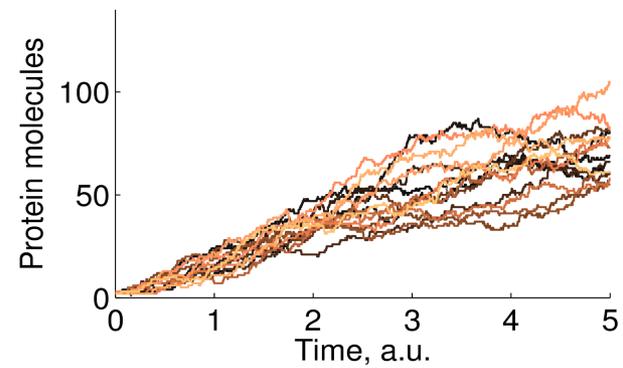
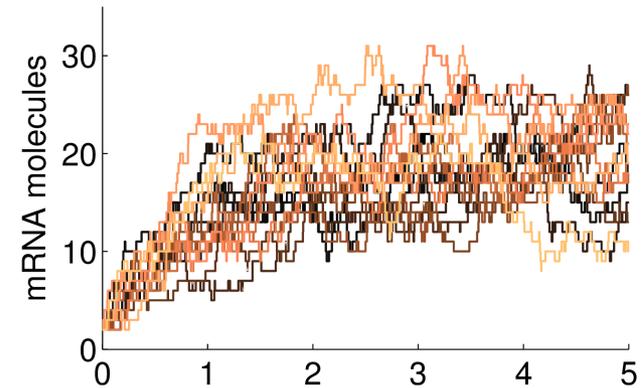
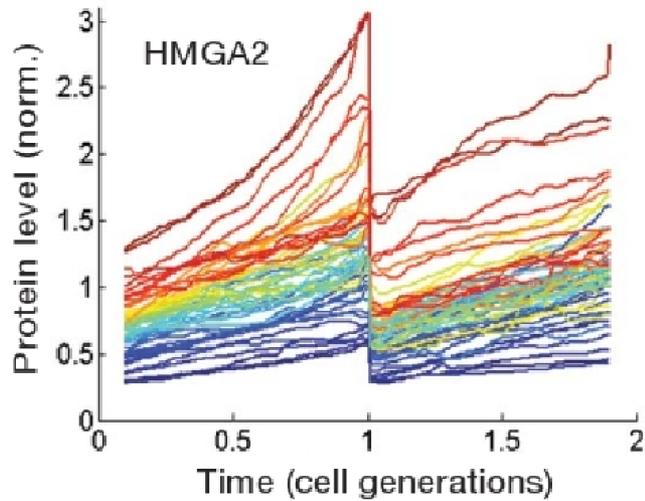
Daniel T. Gillespie

## Stochastischer Simulationsalgorithmus (“direkte Methode”)

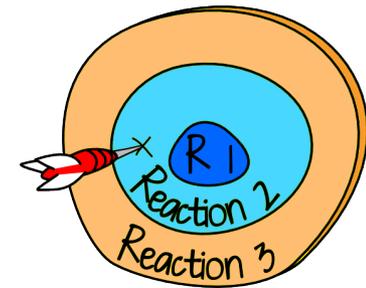
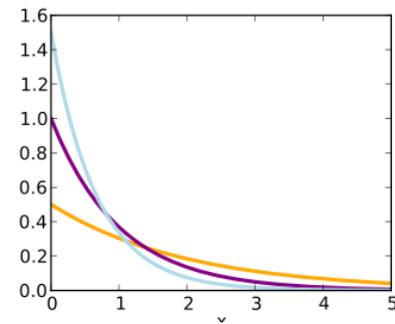
1. Wähle den Zustand für Zeit  $t = 0$
2. Simuliere das nächste Reaktionsereignis
  - (a) Bestimme alle möglichen Reaktionen  $l$  und berechne ihre Raten  $a_l$  und deren Summe  $a$ . Daraus ergibt sich die typische Wartezeit  $T_* = 1/a$ .
  - (b) Berechne die Wartezeit  $T$  bis zum nächsten Reaktionsereignis: ziehe die Zufallszahl  $T$  aus einer Exponentialverteilung mit Konstante  $a$  (d.h. typische Wartezeit  $T_* = 1/a$ ).
  - (c) Entscheide, welcher der möglichen Übergänge geschieht: Zufallsentscheidung mit Wahrscheinlichkeiten  $P(l) = a_l/a$
3. Weiter bei (2.) bis gewünschte Simulationsdauer erreicht ist

# Das Ziel ist erreicht: wir können Proteinmengen simulieren

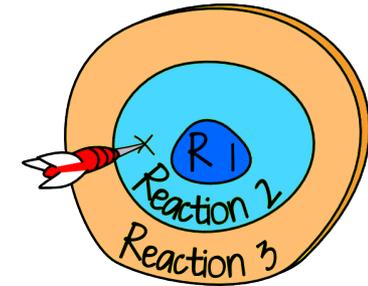
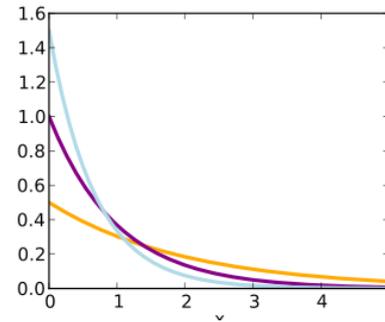
Simulation mit Gillespie-Algorithmus



# Vor- und Nachteile des Gillespie-Algorithmus



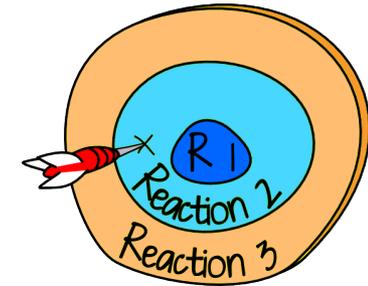
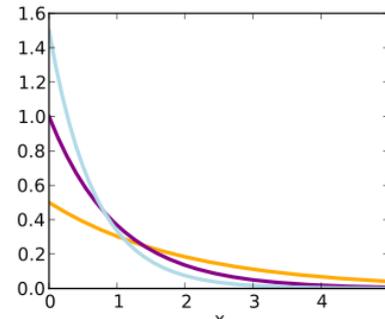
# Vor- und Nachteile des Gillespie-Algorithmus



## Vorteile

- Einfaches und anschauliches Prinzip
- Sehr flexibel, anwendbar für andere chemische Reaktionssysteme und beliebige (diskrete) Modelle
- Gegenüber Mastergleichung: nur aktueller Zustand wird betrachtet (auch bei Berechnung der Übergangsraten)
- Sammeln von Geschichten (kumulativ: Ergebnisse werden besser ..)
- Variable Zeitschritte: lange Wartezeiten erhöhen nicht die Rechenzeit

# Vor- und Nachteile des Gillespie-Algorithmus



## Vorteile

- Einfaches und anschauliches Prinzip
- Sehr flexibel, anwendbar für andere chemische Reaktionssysteme und beliebige (diskrete) Modelle
- Gegenüber Mastergleichung: nur aktueller Zustand wird betrachtet (auch bei Berechnung der Übergangsraten)
- Sammeln von Geschichten (kumulativ: Ergebnisse werden besser ..)
- Variable Zeitschritte: lange Wartezeiten erhöhen nicht die Rechenzeit



## Nachteile

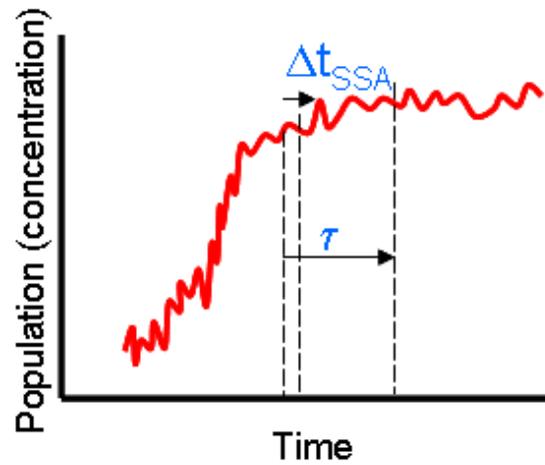
- Annahme: Außenwelt konstant während Wartezeit
- Algorithmus ist langsam, falls viele Ereignisse pro Zeit stattfinden, d.h. bei großen Teilchenzahlen
- Der Bezug zur Ratengleichung bleibt ungeklärt

## Stochastische Prozesse

Wie simuliert man biochemische System stochastisch?

Schnelle Simulationsmethoden für große Systeme

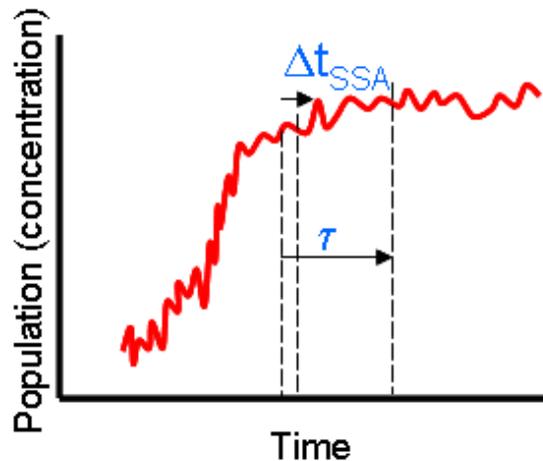
# Modelle mit vielen Teilchen erfordern schnelle Rechenverfahren



## **Problem:**

Die direkte Methode ist sehr langsam bei großen Reaktionsraten (d.h. Teilchenzahlen).

# Modelle mit vielen Teilchen erfordern schnelle Rechenverfahren



## Problem:

Die direkte Methode ist sehr langsam bei großen Reaktionsraten (d.h. Teilchenzahlen).

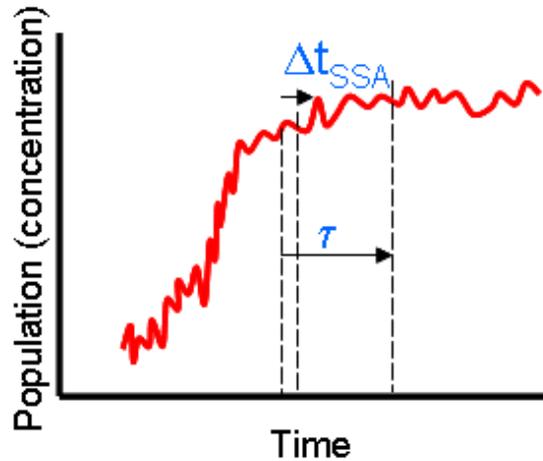
**Idee:** Einzelereignisse nicht einzeln simulieren, sondern im Zeitintervall  $\tau$  zählen!

**Annahme:** feste Reaktionsrate  $a$  im ganzen Zeitintervall

Mittlere Reaktionszahl  $\bar{r} = \text{Rate } a * \text{Intervallauer } \tau$ .

Tatsächliche Reaktionszahlen  $n$  sind poissonverteilt: 
$$P(n) = \frac{\bar{r}^n}{n!} \cdot e^{-\bar{r}}$$

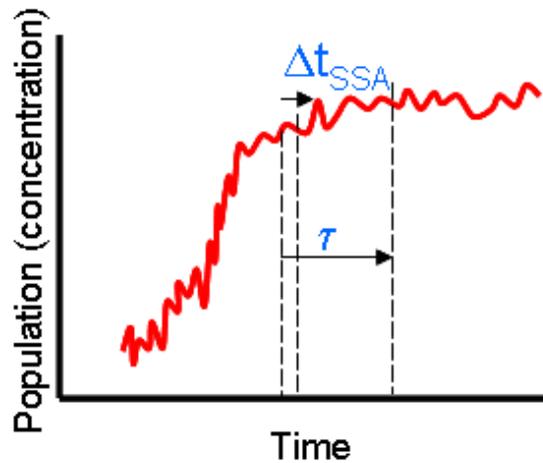
# Die $\tau$ -Sprung-Methode zählt Reaktionen in Zeitintervallen



**$\tau$ -Sprung-Methode (“ $\tau$  leaping”)** für diskrete Zeitintervalle der Länge  $\tau$

1. Wähle Anfangszustand (Molekülzahlen)
2. Für jedes Zeitintervall (Länge  $\tau$ ): berechne aktuelle Rate  $a_i$  für jede Reaktion
3. Für Reaktion  $i$  ist die Zahl der Ereignisse poissonverteilt  $\bar{r}_i$
4. Würfle poissonverteilte Reaktionszahlen und berechne die neuen Molekülzahlen
5. Weiter bei (2.), bis gewünschte Simulationslänge erreicht ist

# Schwankende Reaktionszahlen wirken als “chemisches Rauschen”



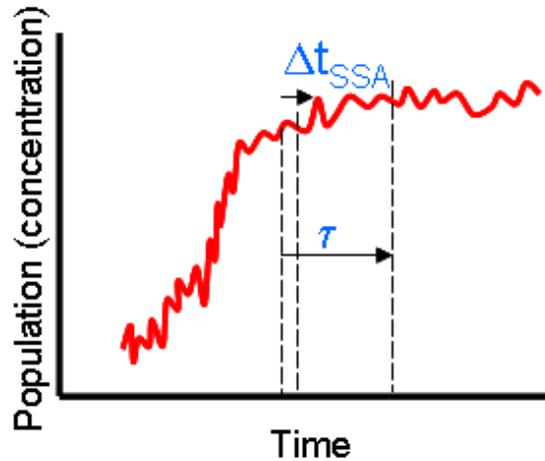
## Viele Reaktionsereignisse:

Zahl  $R_l$  im Intervall  $\tau$  ist ungefähr normalverteilt:

$$R_l = \bar{r}_l + \sqrt{\bar{r}_l} \eta_l = a_l \tau + \sqrt{a_l \tau} \eta_l$$

Standard-  
normalverteilte  
Zufallszahl

# Schwankende Reaktionszahlen wirken als “chemisches Rauschen”



## Viele Reaktionsereignisse:

Zahl  $R_l$  im Intervall  $\tau$  ist ungefähr normalverteilt:

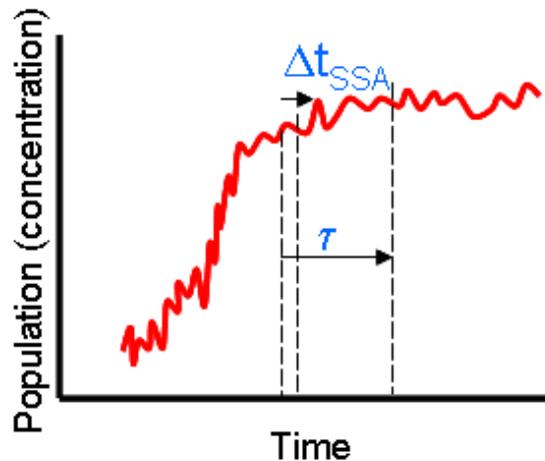
$$R_l = \bar{r}_l + \sqrt{\bar{r}_l} \eta_l = a_l \tau + \sqrt{a_l \tau} \eta_l$$

Standard-normalverteilte Zufallszahl

## Änderung der Teilchenzahl im Zeitintervall

$$x_i(t + \tau) - x_i(t) = \sum_l n_{il} R_l = \sum_l n_{il} [a_l \tau + \sqrt{a_l \tau} \eta_l]$$

# Schwankende Reaktionszahlen wirken als “chemisches Rauschen”



## Viele Reaktionsereignisse:

Zahl  $R_l$  im Intervall  $\tau$  ist ungefähr normalverteilt:

$$R_l = \bar{r}_l + \sqrt{\bar{r}_l} \eta_l = a_l \tau + \sqrt{a_l \tau} \eta_l$$

Standard-normalverteilte Zufallszahl

## Änderung der Teilchenzahl im Zeitintervall

$$x_i(t + \tau) - x_i(t) = \sum_l n_{il} R_l = \sum_l n_{il} [a_l \tau + \sqrt{a_l \tau} \eta_l]$$

## Änderung der Teilchenzahl pro Zeitintervall

$$\frac{x_i(t + \tau) - x_i(t)}{\tau} = \sum_l n_{il} a_l + \sum_l n_{il} \sqrt{a_l} \frac{\eta_l}{\sqrt{\tau}}$$

# Schwankende Reaktionszahlen wirken als “chemisches Rauschen”



Itō Kiyoshi (1915-2008)

## Viele Reaktionsereignisse:

Zahl  $R_l$  im Intervall  $\tau$  ist ungefähr normalverteilt:

$$R_l = \bar{r}_l + \sqrt{\bar{r}_l} \eta_l = a_l \tau + \sqrt{a_l \tau} \eta_l$$

Standard-  
normalverteilte  
Zufallszahl

## Änderung der Teilchenzahl im Zeitintervall

$$x_i(t + \tau) - x_i(t) = \sum_l n_{il} R_l = \sum_l n_{il} [a_l \tau + \sqrt{a_l \tau} \eta_l]$$

## Änderung der Teilchenzahl pro Zeitintervall

$$\frac{x_i(t + \tau) - x_i(t)}{\tau} = \sum_l n_{il} a_l + \sum_l n_{il} \sqrt{a_l} \frac{\eta_l}{\sqrt{\tau}}$$

Symbolische Schreibweise: stochastische Differentialgleichung

$$\frac{dx_i}{dt} = \underbrace{\sum_l n_{il} a_l(\mathbf{x})}_{\text{Ratengleichung}} + \underbrace{\sum_l \left[ n_{il} \sqrt{a_l(\mathbf{x})} \right] \xi_l(t)}_{\text{Rauschterm}}$$

Ratengleichung

Rauschterm

“weißes Rauschen”

Änderungen normalverteilt,  
unabhängig in jedem Zeitpunkt

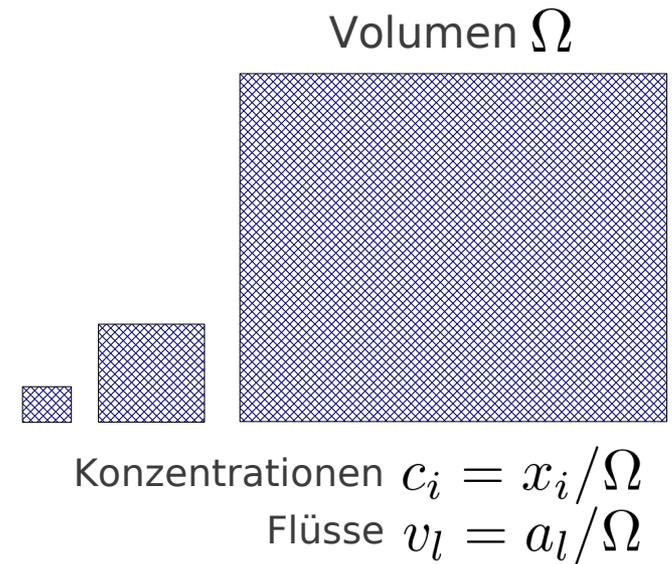
# In großen Systemen sinkt der Einfluss des chemischen Rauschens

Chemische Langevingleichung für Teilchenzahlen  $x_i$

$$\frac{dx_i}{dt} = \underbrace{\sum_l n_{il} a_l(\mathbf{x})}_{\text{Ratengleichung}} + \underbrace{\sum_l \left[ n_{il} \sqrt{a_l(\mathbf{x})} \right] \xi_l(t)}_{\text{Rauschterm}}$$

Ratengleichung

Rauschterm



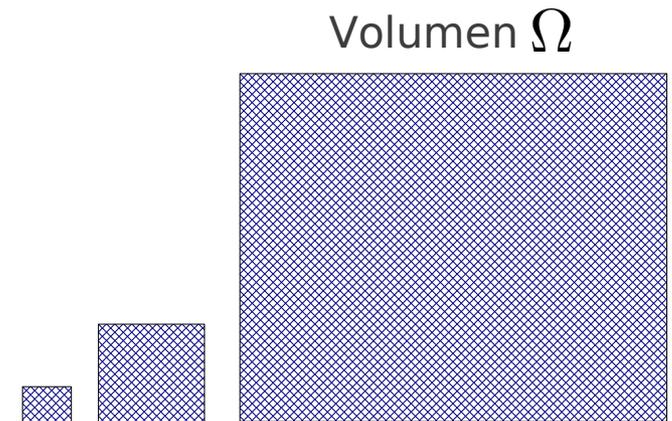
# In großen Systemen sinkt der Einfluss des chemischen Rauschens

Chemische Langevingleichung für Teilchenzahlen  $x_i$

$$\frac{dx_i}{dt} = \sum_l n_{il} a_l(\mathbf{x}) + \sum_l \left[ n_{il} \sqrt{a_l(\mathbf{x})} \right] \xi_l(t)$$

Ratengleichung

Rauschterm



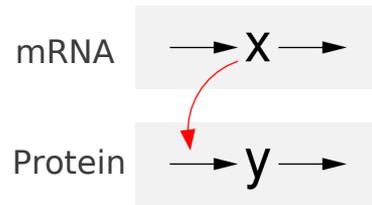
Chemische Langevingleichung für Konzentrationen  $c_i$

$$\frac{dc_i}{dt} = \sum_l n_{il} v_l(\mathbf{c}) + \frac{1}{\sqrt{\Omega}} \sum_l \left[ n_{il} \sqrt{v_l(\mathbf{c})} \right] \xi_l(t)$$

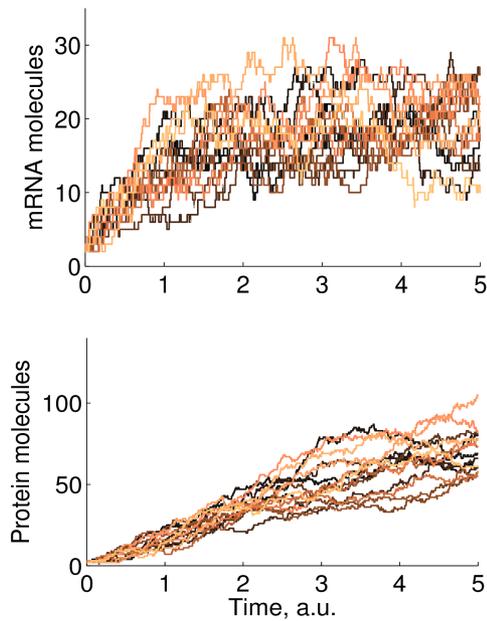
Ratengleichung

Rauschterm

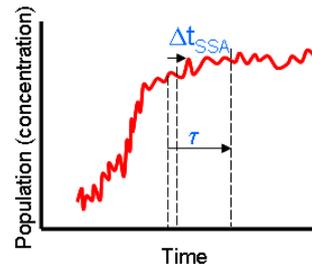
# Stochastische Modelle führen auf deterministische Ratengleichung



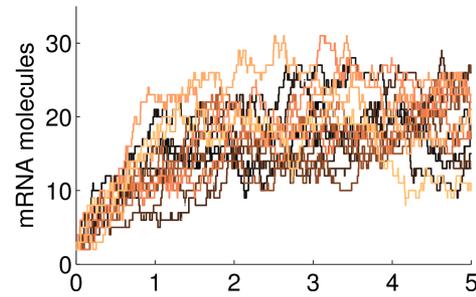
Direkte Methode



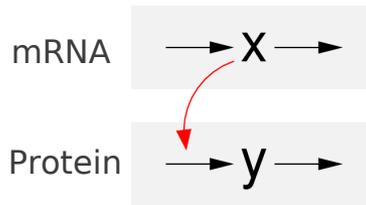
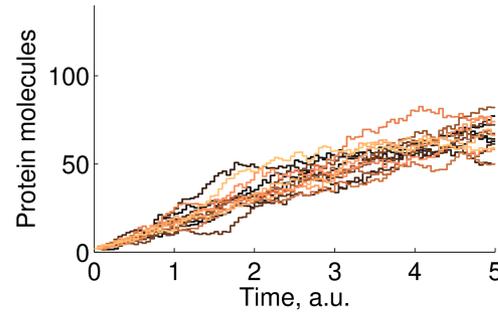
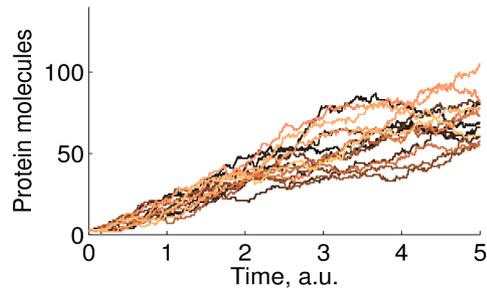
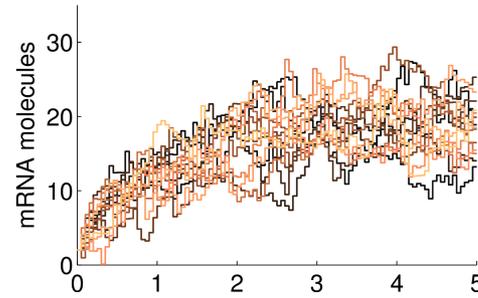
# Stochastische Modelle führen auf deterministische Ratengleichung



Direkte Methode



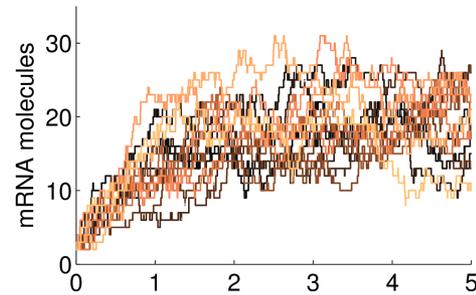
Langevingleichung



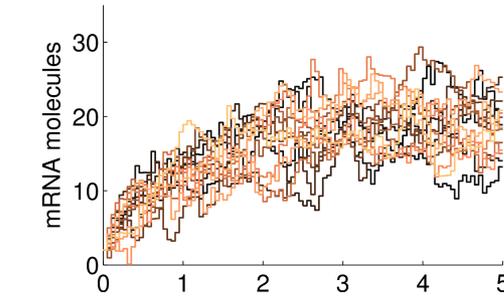
# Stochastische Modelle führen auf deterministische Ratengleichung



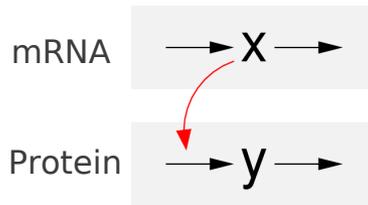
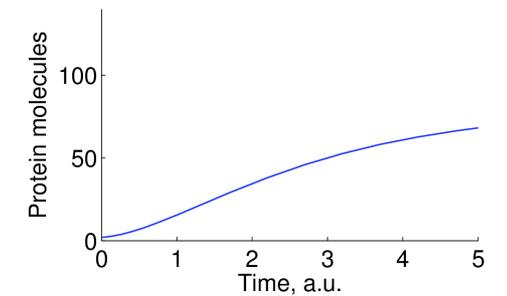
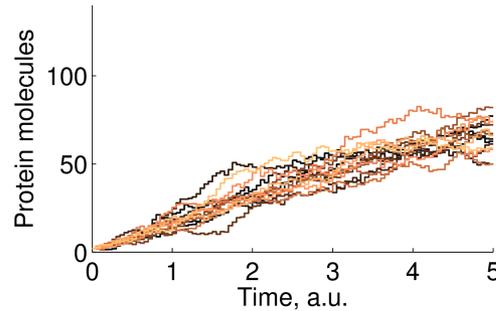
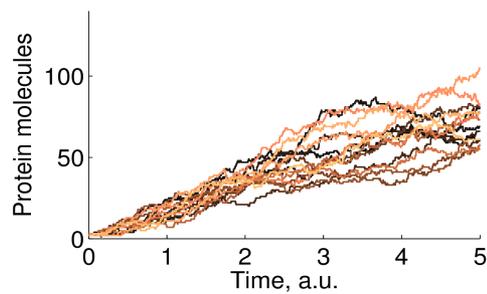
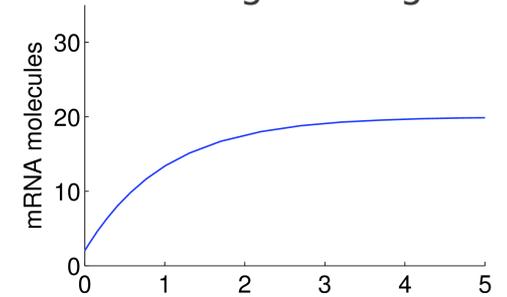
Direkte Methode



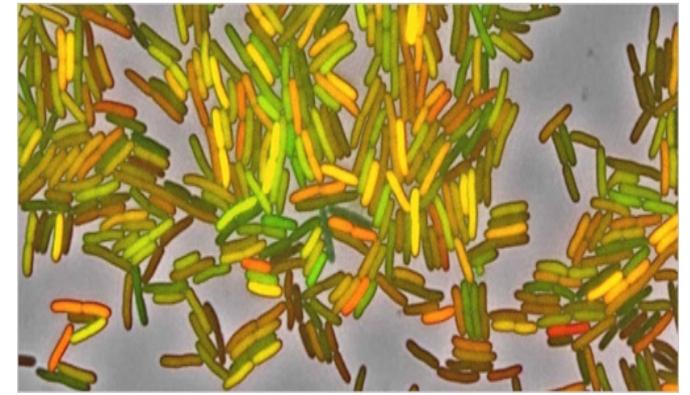
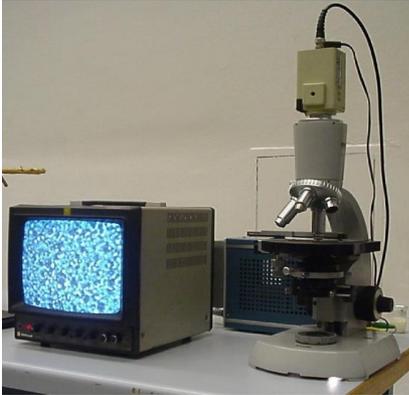
Langevingleichung



Ratengleichung



# Richtig oder falsch?



- Ein Zufallsprozess ist die Menge aller möglichen Geschichten eines Systems (?)
- In Markowprozessen ist die zukünftige Entwicklung durch die Gegenwart festgelegt (?)
- Stochastische Simulationen liefern dieselbe Information wie die Mastergleichung (?)
- Mikroskopische Prozesse lassen sich durch chemisches Rauschen annähern (?)
- Systeme mit großen Teilchenzahlen verhalten sich deterministisch (?)

# Literatur

## **Direkte Methode**

D. T. Gillespie (1977). *Exact stochastic simulation of coupled chemical reactions*.  
J. Phys. Chem., 81:2340-2361.

## **$\tau$ -Sprung-Methode**

D.T. Gillespie (2001). *Approximate accelerated stochastic simulation of chemically reacting systems*.  
J. Chem. Phys, 115(4):1716-1733.

## **Chemische Langevingleichung**

D.T. Gillespie (2000). *The chemical Langevin equation*.  
J. Chem. Phys., 113(1):297-306.

## **Lehrbuch:**

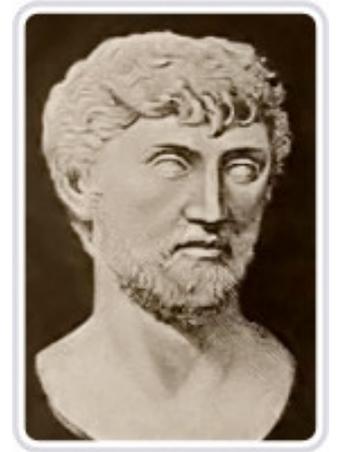
*Stochastische dynamische Systeme. Konzepte, numerische Methoden, Datenanalysen.*  
Josef Honerkamp, Wiley-VCH 1990

# Bewegung von Staubteilchen .. ein Hinweis auf Atome?

## “Sonnenstäubchen” (aus: De rerum natura, 2. Buch)

...

Wie in dem Sonnenstrahle die winzigen Körperchen wimmeln,  
Weil dergleichen Gewimmel beweist, auch in der Materie  
Gibt's ein unsichtbares, verborgenes Weben der Kräfte.  
Denn bei den Stäubchen erkennst du, wieviele die Richtung verändern,  
Trifft sie ein heimlicher Stoß, und wie sie sich rückwärts wenden,  
Hierhin und dorthin getrieben nach allen möglichen Seiten.  
Merke, die ganze Bewegung beginnt hier bei den Atomen.  
Denn es erhalten zuerst die Urelemente den Anstoß,  
Hierauf werden die Körper, die wenig Verbindungen haben  
Und in der Kraft sich am nächsten den Urelementen vergleichen,  
Durch unmerkbare Stöße von diesen dann weiter getrieben,  
Und sie führen dann selbst den Stoß auf die größeren weiter,  
So geht von dem Atom die Bewegung empor und sie endet  
Mählich bei unseren Sinnen, bis endlich auch das sich bewegt,  
Was wir im Lichte der Sonne mit Augen zu schauen vermögen,  
Ohne doch deutlich die Stöße zu sehn, die Bewegung erregen.



Titus Lucretius Carus  
(ca 97-55 v.C.)

Was ist Zufall?

Wie wird Zufall in der Physik erklärt?

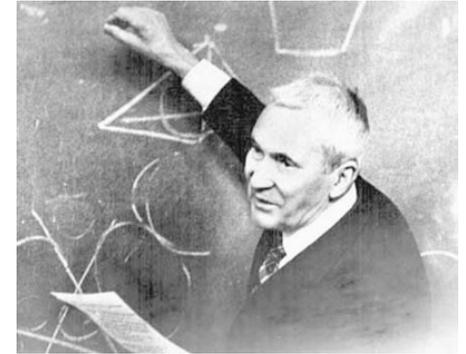
**Warum ist es überhaupt sinnvoll,  
deterministisch zu Denken?**

Gibt es "objektiven" Zufall in der Natur?

# Was ist Zufall?



Pierre-Simon Laplace



Andrej N. Kolmogorow

**Zufall:** Vorgänge, für die keine kausale Erklärung gegeben werden kann

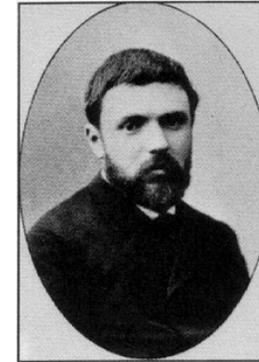
# Wie wird Zufall in der Physik erklärt?



Ludwig Boltzmann



W. Heisenberg, N. Bohr, W. Pauli



Henri Poincaré

## **Unbekannter Einfluss der Außenwelt**

Alle Systeme sind Störungen durch die Außenwelt ausgesetzt, die nicht genau bekannt sind (es sei denn, das Modell umfasst das gesamte Universum)

## **Unbekannte mikroskopische Freiheitsgrade**

Hängt das Systemverhalten von (mikroskopischen oder makroskopischen) Freiheitsgrade des Systems ab, die unbekannt sind, ist keine genaue (deterministische) Simulation möglich.

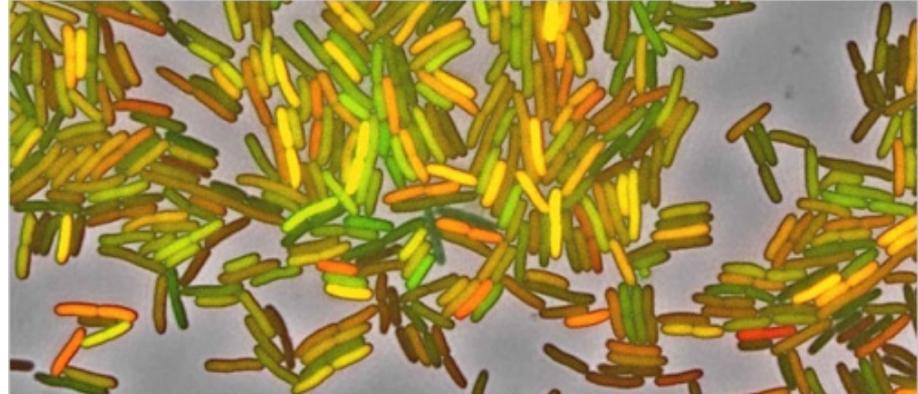
## **Quantenmechanische Unsicherheit**

Kopenhagener Deutung behauptet objektiven Zufall im Messprozess: quantenmechanische Größen sind nicht eindeutig vorherbestimmt und daher prinzipiell nicht vorhersagbar; ihre Wahrscheinlichkeiten folgen aus der Wellenfunktion.

## **Deterministisches Chaos**

In manchen Systemen können minimale Störungen des Anfangszustands oder der Dynamik das Systemverhalten drastisch ändern. Bei beschränkter Messgenauigkeit ist das Systemverhalten nicht genau vorhersagbar (höchstens Wahrscheinlichkeiten aus Systemdynamik).

# Ist die Natur zufällig oder ist Zufall subjektiv?



## **Aleatorische Wahrscheinlichkeit**

Quantenmechanische Unbestimmtheit sorgt für "objektiven" Zufall im Mikrokosmos, der sich dann in die Makroebene fortpflanzt.

Wahrscheinlichkeit = relative Häufigkeiten in gleichartigen Systemen

→ Frequentistischer Ansatz in der Statistik

## **Epistemische Wahrscheinlichkeit**

Modell beschreibt nicht die Welt "an sich", sondern was wir über die Welt wissen (können). Fehlt Information (z.B. über unbekannte Systemvariablen) bleiben Möglichkeiten offen.

Wahrscheinlichkeit = Maß für subjektive Unsicherheit

→ Bayesianischer Ansatz in der Statistik