Stochastic simulations
Application to molecular networks

Didier Gonze

Unité de Chronobiologie Théorique
Service de Chimie Physique - CP 231
Université Libre de Bruxelles
Belgium
Overview

- **Introduction: theory and simulation methods**
  - Definitions (intrinsic vs extrinsic noise, robustness,...)
  - Deterministic vs stochastic approaches
  - Master equation, birth-and-death processes
  - Gillespie and Langevin approaches
  - Application to simple systems

- **Literature overview**
  - Measuring the noise, intrinsic vs extrinsic noise
  - Determining the sources of noise
  - Assessing the robustness of biological systems

- **Application to circadian clocks**
  - Molecular bases of circadian clocks
  - Robustness of circadian rhythms to noise
Deterministic vs stochastic approaches

A

Deterministic

Concentration

Time

B

Stochastic

Concentration

Time

C

Deterministic

Concentration

Time

D

Stochastic

Concentration

Time
Deterministic vs stochastic approaches

Ordinary differential equations

\[
\frac{dX}{dt} = f_{\text{production}}(X) - f_{\text{consumption}}(X)
\]

Stochastic differential equations

\[
\frac{dX}{dt} = f_{\text{production}}(X) - f_{\text{consumption}}(X) + f_{\text{noise}}
\]

Discrete stochastic simulations

\[ P(\text{production}) \rightarrow X \rightarrow P(\text{consumption}) \]
Sources of noise

Intrinsic noise

Noise resulting from the probabilistic character of the (bio)chemical reactions. It is particularly important when the number of reacting molecules is low. It is inherent to the dynamics of any genetic or biochemical systems.

Extrinsic noise

Noise due to the random fluctuations in environmental parameters (such as cell-to-cell variation in temperature, pH, kinetics parameters, number of ribosomes,...).

Both Intrinsic and extrinsic noise lead to fluctuations in a single cell and results in cell-to-cell variability.
Noise in biology

Gene transcription

- RNAP
- Ribosome
- Protein
- mRNA

Repression

- Repressor
- RNAP
- Regulatory & RNAP binding sites (promoter)

Activation

- Activator
- RNAP
- Regulatory & RNAP binding sites (promoter)
Noise in biology

- Regulation and binding to DNA
- Transcription to mRNA
- Splicing of mRNA
- Transportation of mRNA to cytoplasm
- Translation to protein

- Conformation of the protein
- Post-translational changes of protein
- Protein complexes formation
- Proteins and mRNA degradation
- Transportation of proteins to nucleus
Noise in biology

Noise-producing steps in biology

Effects of noise

Georges Seurat
*Un dimanche après-midi à la Grande Jatte*

Fedoroff & Fontana (2002) *Science*
Effects of noise

Destructive effect of noise
- Imprecision in the timing of genetic events
- Imprecision in biological clocks
- Phenotypic variations

Constructive effect noise
- Noise-induced behaviors
- Stochastic resonance
- Stochastic focusing
Noise-induced phenotypic variations

Stochastic kinetic analysis of a developmental pathway bifurcation in phage-\( \lambda \) \textit{Escherichia coli} cell

Arkin, Ross, McAdams (1998) \textit{Genetics} 149: 1633-48

Fluctuations in rates of gene expression can produce highly erratic time patterns of protein production in individual cells and wide diversity in instantaneous protein concentrations across cell populations.

When two independently produced regulatory proteins acting at low cellular concentrations competitively control a switch point in a pathway, stochastic variations in their concentrations can produce probabilistic pathway selection, so that an initially homogeneous cell population partitions into distinct phenotypic subpopulations.
Imprecision in biological clocks

Circadian clocks limited by noise

For example, in a previously studied model that depends on a **time-delayed negative feedback**, reliable oscillations were found when reaction kinetics were approximated by continuous differential equations. However, when the **discrete nature of reaction events** is taken into account, the oscillations persist but with **periods and amplitudes** that **fluctuate widely in time**. Noise resistance should therefore be considered in any postulated molecular mechanism of circadian rhythms.
Noise-induced behaviors

- Noise-induced oscillations
- Noise-induced synchronization
- Noise-induced excitability
- Noise-induced bistability
- Noise-induced pattern formation

Noise-induced oscillations in an excitable system

Vilar et al, PNAS, 2002
Stochastic resonance is the phenomenon whereby the addition of an optimal level of noise to a weak information-carrying input to certain nonlinear systems can enhance the information content at their outputs.

Here, we show that stochastic resonance enhances the normal feeding behaviour of paddle fish (Polyodon spathula) which use passive electroreceptors to detect electrical signals from planktonic prey (Daphnia).
Robustness is a property that allows a system to maintain its functions despite external and internal noise.

It is commonly believed that robust traits have been selected by evolution.

Autoregulation (negative feedback loops) in gene circuits provide stability, thereby limiting the range over which the concentrations of network components fluctuate.
Among these topologies the experimentally established chemotaxis network of *Escherichia coli* has the smallest sufficiently robust network structure, allowing accurate chemotactic response for almost all individuals within a population.
Theory of stochastic systems
Let's consider a single species (X) involved in a single reaction:

\[ n \ X + \ldots \rightarrow p \ X + \ldots \]

Deterministic description of its time evolution (ODE):

\[ \frac{dX}{dt} = \eta v \quad \text{with} \quad \eta = p - n \]

\( \eta \) = stoechiometric coefficient

\( v \) = reaction rate:

\[ v = kX^n \]
Deterministic formulation

Let's now consider a several species \((X_i)\) involved in a couple of reactions:

\[
\begin{align*}
  n_{11}X_1 + n_{21}X_2 + \ldots & \rightarrow p_{11}X_1 + p_{21}X_2 + \ldots \\
  n_{12}X_1 + n_{22}X_2 + \ldots & \rightarrow p_{12}X_1 + p_{22}X_2 + \ldots \\
  \ldots \\
  n_{1R}X_1 + n_{2R}X_2 + \ldots & \rightarrow p_{1R}X_1 + p_{2R}X_2 + \ldots 
\end{align*}
\]

Deterministic description of their time evolution (ODE):

\[
\frac{dX_i}{dt} = \sum_{r=1}^{R} \eta_{ir}v_r = \eta_{i1}v_1 + \eta_{i2}v_2 + \ldots + \eta_{iR}v_R
\]

\(v_r\) = rate of the different reactions \((r = 1, 2, \ldots R)\).

\(\eta_{ir} = p_{ir} - n_{ir}\) = stoechiometric coefficient of compound \(X_i\) in reaction \(r\).
Stochastic formulation

Stochastic description (in terms of the probabilities):

\[ P(X, t + dt) = P(X, t)P(\text{no change over } dt) + \sum_{r=1}^{R} P(X - \eta_r, t)P(\text{state change over } dt) \]

\[ P(\text{no change over } dt) = 1 - \sum_{r=1}^{R} w_r(X)dt \]

\[ P(\text{state change over } dt) = w_r(X - \eta_r)dt \]

\[
\lim_{dt \to 0} \frac{P(X, t + dt) - P(X, t)}{dt} = \frac{\partial P(X, t)}{\partial t} \\
\frac{\partial P(X, t)}{\partial t} = \sum_{r=1}^{R} \left( w_r(X - \eta_r)P(X - \eta_r, t) - w_r(X)P(X, t) \right)
\]

Chemical master equation
Comparison of the different formalisms

Deterministic description

Stochastic description (1 possible realization)

Stochastic description (10 possible realizations)

Stochastic description (probability distribution)
Stochastic formulation: birth-and-death

Birth-and-death process (single species):

\[
\begin{align*}
\overset{k_b}{\longrightarrow} & \quad X & \overset{k_d}{\longrightarrow} \\
\end{align*}
\]

State transitions

Master equation for a birth-and-death process

\[
\frac{\partial P(X,t)}{\partial t} = k_b P(X-1,t) + k_d (X+1) P(X+1,t) - k_b P(X,t) - k_d X P(X,t)
\]
Stochastic formulation: birth-and-death

Birth-and-death process (multiple species):

\[
\begin{align*}
&\xrightarrow{k_{b1}} X_1 \xrightarrow{k_{d1}} \\
&\xrightarrow{k_{b2}} X_2 \xrightarrow{k_{d2}} \\
&\xrightarrow{k_{b3}} X_3 \xrightarrow{k_{d3}} \\
&\ldots
\end{align*}
\]

Master equation for a birth-and-death process

\[
\frac{\partial P(\{X_i\}, t)}{\partial t} = \sum_{r=1}^{R} [k_{br}(\{X_i - \eta_{ir}\})P(\{X_j \neq i, X_i - \eta_{ir}\}, t)]
\]

\[
+k_{dr}(\{X_i + \eta_{ir}\})P(\{X_j \neq i, X_i + \eta_{ir}\}, t) - k_{br}(\{X_i\})P(\{X_i\}, t) - k_{dr}(\{X_i\})P(\{X_i\}, t)
\]
Stochastic formulation: examples

\[ X + Y \xrightarrow{k} Z \]

\[ w(X, Y) = kXY \]

\[
\frac{\partial P(X, Y, Z, t)}{\partial t} = w(X + 1, Y + 1)P(X + 1, Y + 1, Z - 1) \\
- w(X, Y)P(X, Y, Z) \\
= k(X + 1)(Y + 1)P(X + 1, Y + 1, Z - 1) \\
- kXYP(X, Y, Z)
\]
Stochastic formulation: examples

\[ A + X \xrightarrow{k} 2X \]

\[ w(A, X) = kAX \]

\[
\frac{\partial P(A, X, t)}{\partial t} = w(A + 1, X - 1)P(A + 1, X - 1) \\
- w(A, X)P(A, X) \\
= k(A + 1)(X - 1)P(A + 1, X - 1) \\
- kAXP(A, X)
\]
Stochastic formulation: examples

\[ 2X \xrightarrow{k} E \]

\[ w(X) = \frac{k}{2} X(X - 1) \]

\[
\frac{\partial P(X, E, t)}{\partial t} = w(X + 2)P(X + 2, E - 1) - w(X)P(X, E) - \frac{k}{2}(X + 1)(X + 2)P(X + 2, E - 1) - \frac{k}{2}(X - 1)(X)P(X, E)
\]
Stochastic formulation: Fokker-Planck

Fokker-Planck equation

\[
\frac{\partial P(\mathbf{X}, t)}{\partial t} = -\sum_i \left( \frac{\partial}{\partial X_i} F_i(\mathbf{X}) P(\mathbf{X}, t) \right) + \sum_{i,j} \left( \frac{\partial^2}{\partial X_i X_j} G_{i,j}(\mathbf{X}) P(\mathbf{X}, t) \right)
\]

Drift term

Diffusion term

\[
F_i(\mathbf{X}) = \sum_{r=1}^{R} \eta_r w_r(\mathbf{X})
\]

\[
G_{i,j}(\mathbf{X}) = \sum_{r=1}^{R} \eta_r \eta_r^T w_r(\mathbf{X})
\]
This is a nice theory, but...

\[ A \Leftrightarrow B \Leftrightarrow C \]

For $N = 200$ there are more than $1000000$ possible molecular combinations!

We can not solve the master equation by hand.

We need to perform simulations (using computers).
Numerical simulation

The Gillespie algorithm

Direct simulation of the master equation

\[
P(\text{production}) \rightarrow X \rightarrow P(\text{consumption})
\]

The Langevin approach

Stochastic differential equation

\[
\frac{dX}{dt} = f_{\text{production}}(X) - f_{\text{consumption}}(X) + f_{\text{noise}}
\]
Gillespie algorithm

The Gillespie algorithm

A reaction rate $w_i$ is associated to each reaction step. These probabilities are related to the kinetics constants.

Initial number of molecules of each species are specified.

The time interval is computed stochastically according the reaction rates.

At each time interval, the reaction that occurs is chosen randomly according to the probabilities $w_i$ and both the number of molecules and the reaction rates are updated.
**Gillespie algorithm**

**Principle of the Gillespie algorithm**

Probability that reaction \( r \) occurs

\[
P_r = \frac{w_r}{\sum_{i=1}^{R} w_i}
\]

Reaction \( r \) occurs if

\[
P_{r-1} < z_1 \leq P_{r-1} + P_r
\]

Time step to the next reaction

\[
\Delta t = \frac{1}{\sum_{i=1}^{R} w_i} \ln \frac{1}{z_2}
\]

---

**Gillespie algorithm**

\[A \xrightarrow{w_1} B\]

\[B + C \xrightarrow{w_2} D\]

\[D \xrightarrow{w_3} E + F\]

...
In practice...

1. Calculate the transition probability \( w_i \) which are functions of the kinetics parameters \( k_r \) and the variables \( X_i \).

2. Generate \( z_1 \) and \( z_2 \) and calculate the reaction that occurs as well as the time till this reaction occurs.

3. Increase \( t \) by \( \Delta t \) and adjust \( X \) to take into account the occurrence of the reaction that just occurred.
A key parameter in this approach is the **system size** $\Omega$. This parameter has the unit of a volume and is used to convert **concentration** $x$ into a **number of molecules** $X$:

$$X = \Omega x$$

For a given concentration (defined by the deterministic model), bigger is the system size ($\Omega$), larger is the number of molecules. Therefore, $\Omega$ allows us to control directly the number of molecules present in the system (hence the noise).

Typically, $\Omega$ appears in the reaction steps involving two (or more) molecular species because these reactions require the collision between two (or more) molecules and their rate thus depends on the number of molecules present in the system.

\[
\begin{align*}
A + B & \rightarrow C \\
v & = A B / \Omega \\
2A & \rightarrow D \\
v & = A (A-1) / 2 \Omega
\end{align*}
\]
Gillespie algorithm: improvements & extensions

Next Reaction Method (Gibson & Bruck, 2000)

Gibson & Bruck’s algorithm avoids calculation that is repeated in every iteration of the computation. This adaptation improves the time performance while maintaining exactness of the algorithm.

Tau-Leap Method (Gillespie, 2001)

Instead of which reaction occurs at which time step, the Tau-Leap algorithm estimated how many of each reaction occur in a certain time interval. We gain a substantial computation time, but this method is approximative and its accuracy depends on the time interval chosen.

Delay Stochastic Simulation (Bratsun et al., 2005)

Bratsun et al. have extended the Gillespie algorithm to account for the delay in the kinetics. This adaptation can therefore be used to simulate the stochastic model corresponding to delay differential equations.
Langevin stochastic equation

Langevin stochastic differential equation

\[ \frac{dX}{dt} = f(X) + g(X)\xi(t) \]

If the noise is white (uncorrelated), we have:

\[ <\xi(t)> = 0 \]
\[ <\xi(t)\xi(t')> = D\delta(t-t') \]

\(D\) measures the strength of the fluctuations.
Gillespie vs Langevin modeling

Gillespie approach

\[ w_1 = \nu_m \frac{X}{K_M + X} \]

\[ w_2 = \nu_d X \]

\[ \frac{dX}{dt} = \nu_m \frac{X}{K_M + X} - \nu_d X \]

Langevin approach

\[ \frac{dX}{dt} = \nu_m \frac{X}{K_M + X} - \nu_d X + \xi(t) \]
Gillespie vs Langevin modeling

Gillespie

\[ \Omega \rightarrow \]

Langevin

\[ D \rightarrow \]
Spatial stochastic modeling
Spatial stochastic modeling


Reactions
Blue $\rightarrow$ 2 Blue
Blue + Red $\rightarrow$ 2 Red
Red $\rightarrow$ nothing
Michaelis-Menten

Reactional scheme

\[ E + S \overset{k_1}{\underset{k_{-1}}{\rightleftharpoons}} C \overset{k_2}{\rightarrow} E + P \]

Deterministic evolution equations

\[
\begin{align*}
\frac{dS}{dt} &= -k_1 ES + k_{-1} C \\
\frac{dE}{dt} &= -k_1 ES + k_{-1} C + k_2 C \\
\frac{dC}{dt} &= k_1 ES - k_{-1} C - k_2 C \\
\frac{dP}{dt} &= k_2 C
\end{align*}
\]
Michaelis-Menten

Reactional scheme

\[ E + S \xrightleftharpoons[k_{-1}]{k_1} C \rightarrow E + P \]

Stochastic transition table

<table>
<thead>
<tr>
<th>( r )</th>
<th>reaction</th>
<th>reaction rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( E + S \xrightarrow{k_1} C )</td>
<td>( w_1 = k_1 ES/\Omega )</td>
</tr>
<tr>
<td>2</td>
<td>( C \xrightarrow{k_{-1}} E + S )</td>
<td>( w_2 = k_{-1}C )</td>
</tr>
<tr>
<td>3</td>
<td>( C \xrightarrow{k_2} E + P )</td>
<td>( w_3 = k_2C )</td>
</tr>
</tbody>
</table>

Master equation

\[
\frac{\partial P(S, C; E; t)}{\partial t} = - (k_1SE + (k_{-1} + k_2)C)(P(S, C; t)) \\
+ k_1(S + 1)(E + 1)P(S + 1, C - 1; t) \\
+ k_{-1}(C + 1)P(S - 1, C + 1; t) \\
+ k_2(C + 1)P(S, C + 1; t)
\]
Michaelis-Menten

Deterministic simulation

Stochastic simulation
Michaelis-Menten

Quasi-steady state assumption

If $E \ll S_0$ then $dC/dt = 0$ quasi-steady state

$$C = \frac{E_T S}{k_1 + k_2} + S$$

Rate of production of $P$:

$$v = \frac{dP}{dt} = k_2 C = V_{max} \frac{S}{K_M + S}$$

$$V_{max} = k_2 E_T \text{ and } K_M = \frac{k_1 + k_2}{k_1}$$

Stochastic transition table:

<table>
<thead>
<tr>
<th>$r$</th>
<th>reaction</th>
<th>reaction rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$S \xrightarrow{v} P$</td>
<td>$w_1 = V_{max} \Omega \frac{S}{K_S \Omega + S}$</td>
</tr>
</tbody>
</table>
Gene expression

Reactional scheme

Thattai - van Oudenaarden model

\[
\begin{align*}
\text{[Gene (G)]} & \xrightarrow{k_1} \text{mRNA (R)} \\
\text{mRNA (R)} & \xrightarrow{k_2} \\
\text{[mRNA (R)]} & \xrightarrow{k_3} \text{Protein (P)} \\
\text{Protein (P)} & \xrightarrow{k_4}
\end{align*}
\]
Gene expression

mRNA

Poisson distribution (computed from the simulation results)

Theoretical Poisson distribution

Protein

non Poisson distribution (computed from the simulation results)

Theoretical Poisson distribution
As the number of molecules increases, the steady-state probability density function becomes sharper. The distribution is given by

\[ p(j) = \binom{n}{j} \frac{k_1^j k_2^{n-j}}{(k_1 + k_2)^n} \]
**Bruxellator**

**Reactional scheme**

<table>
<thead>
<tr>
<th>$r$</th>
<th>reaction</th>
<th>rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$A \xrightarrow{k_1} X$</td>
<td>$v_1 = k_1 A$</td>
</tr>
<tr>
<td>2</td>
<td>$B + X \xrightarrow{k_2} Y + C$</td>
<td>$v_2 = k_2 BX$</td>
</tr>
<tr>
<td>3</td>
<td>$2X + Y \xrightarrow{k_3} 3X$</td>
<td>$v_3 = k_3 X^2Y$</td>
</tr>
<tr>
<td>4</td>
<td>$X \xrightarrow{k_4} D$</td>
<td>$v_4 = k_4 X$</td>
</tr>
</tbody>
</table>

**Deterministic evolution equations**

\[
\begin{align*}
\frac{dX}{dt} &= k_1 a - k_2 b X + k_3 X^2 Y - k_4 X \\
\frac{dY}{dt} &= k_2 b X - k_3 X^2 Y
\end{align*}
\]
Bruxellator

Stochastic transition table

<table>
<thead>
<tr>
<th>$r$</th>
<th>reaction</th>
<th>reaction rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$A \xrightarrow{k_1} X$</td>
<td>$w_1 = k_1 A$</td>
</tr>
<tr>
<td>2</td>
<td>$B + X \xrightarrow{k_2} Y + C$</td>
<td>$w_2 = k_2 BX/\Omega$</td>
</tr>
<tr>
<td>3</td>
<td>$2X + Y \xrightarrow{k_3} 3X$</td>
<td>$w_3 = k_3 X(X - 1)Y/2\Omega^2$</td>
</tr>
<tr>
<td>4</td>
<td>$X \xrightarrow{k_4} D$</td>
<td>$w_4 = k_4 X$</td>
</tr>
</tbody>
</table>

Master equation

$$\frac{\partial P(X, Y; t)}{\partial t} = - (k_1 A + k_2 BX + k_3 X^2 Y + k_4 X)P(X, Y; t) + k_1 A P(X - 1, Y; t) + k_2 B (X + 1) P(X + 1, Y - 1; t) + k_3 (X - 1)^2 (Y + 1) P(X - 1, Y + 1; t) + k_4 (X + 1) P(X + 1, Y; t)$$
Bruxellator

Deterministic simulation

Stochastic simulation, $\Omega = 1000$

Stochastic simulation, $\Omega = 100$
Quantification of the noise

- Histogram of periods

- Auto-correlation function

\[ C(\tau) = \frac{1}{T-\tau} \int_{0}^{T-\tau} x(t)x(t+\tau)dt \]

\[ C(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x(n)x(n+m) \]
Bruxellator
Bruxellator

The Lotka-Volterra predator-prey model is a pair of first-order, non-linear, differential equations describing the dynamics of biological systems in which two species interact, one as the prey and the other as the predator. The equations are as follows:

\[
\begin{align*}
\frac{dX}{dt} &= \alpha X - XY \\
\frac{dY}{dt} &= XY - \beta Y
\end{align*}
\]

Where:
- \(X\) represents the prey population,
- \(Y\) represents the predator population,
- \(\alpha\) is the prey growth rate,
- \(\beta\) is the predator mortality rate.

These equations model the growth and interaction of the two populations over time.
Lotka-Volterra

Deterministic simulation

Stochastic simulation
The two non-dimensional variables $x$ and $y$ are

\begin{align*}
x &= \text{voltage-like variable (activator)} - \text{slow variable} \\
y &= \text{recovery-like variable (inhibitor)} - \text{fast variable}
\end{align*}

The nonlinear function $f(x)$ (shaped like an inverted N, as shown in figure 2) is one of the nullclines of the deterministic system; a common choice for this function is

$$f(x) = x - ax^3$$

$D(t)$ is a white Gaussian noise with intensity $D$. 

The Fitzhugh-Nagumo model is an example of a two-dimensional excitable system. It was proposed as a simplification of the famous model by Hodgkin and Huxley to describe the response of an excitable nerve membrane to external current stimuli.
Fitzhugh-Nagumo

Deterministic

Stochastic

excitability

oscillations