

**CPIB SUMMER SCHOOL 2011:
INTRODUCTION TO BIOLOGICAL
MODELLING**

*Lecture 3.2
Stochastic Models*

Nick Monk

Deterministic vs. Stochastic Models

Deterministic models (differential equations):

- the state is represented by **continuous variables**,
- reactions/interactions are represented as **continuous processes** (production, decay, movement,...).

This is appropriate when the numbers of each species and the frequency of reaction events are “sufficiently large”.

This is not always the case.

e.g. transcription (1 or 2 copies of each gene); mRNA-protein networks in prokaryotes; low-density populations, ...

Reaction Rates vs. Events

In deterministic models (differential equations) each process (e.g. transcription, degradation) has a **rate** (which is typically a function of the state).

e.g. Rate of transcription = $\frac{A[TF]}{h + [TF]}$

This gives a measure of how frequently each type of reaction is expected to occur (for a given state).

Alternatively, we can develop models centred on **individual reaction events** (e.g. the production of a single mRNA molecule, or the binding of two protein molecules).

Reaction Rates vs. Events

In a typical biological system, reaction events do not occur at regular intervals.

e.g. in a “well-mixed” solution, molecular motion can be represented as Brownian motion; bimolecular reactions occur when two reactant molecules approach closely enough (and in the correct orientation).

e.g. transcription depends on the assembly of a polymerase complex and then sequential addition of nucleotides.

The timing of these events is essentially **stochastic** rather than deterministic.

Deterministic vs. Stochastic Models

Deterministic	Stochastic
Ordinary differential equations	Continuous time Markov chains
Concentrations of molecules	Numbers of molecules
Future is “predictable” given present knowledge	Includes randomness; every simulation is different
Wide range of techniques available for analysis	Not as many techniques for analysis; often rely on simulations
Good for large number of individuals; qualitative analysis	Better for simulating dynamics with small numbers of molecules
Represents population average	Represents population variability

Stochastic kinetics

- Assume homogeneity:
 - $P(\text{molecule in volume } \delta V)$ is equal for each δV on the timescale of the chemical reactions that change the state.
 - In other words, we assume that the “reaction mixture” (i.e. the inside of the cell) is well-mixed. This may be questionable.

• Then the state of a cell can be represented by a vector \mathbf{n} , where

$$\mathbf{n} = [n_1, n_2, \dots, n_N],$$

and n_i is the (integer) number of molecules of species i in the cell ($i = 1, 2, \dots, N$). Note that we could equally well use concentrations rather than numbers.

The stoichiometric matrix

- Elementary reaction j changes the state: $n_i \rightarrow n_i + v_{ij}$, where v_{ij} is the **stoichiometric matrix**.
- The state vector \mathbf{n} changes by \mathbf{v}_j (the state change vector) when reaction j occurs:

$$\mathbf{n} \rightarrow \mathbf{n} + \mathbf{v}_j, \quad \mathbf{v}_j = (v_{1j}, v_{2j}, \dots, v_{Nj})$$
- Note that the elements of the state change vector are integers (v_{ij} is just the change in the number of molecules of type i when reaction j occurs).

The stoichiometric matrix

For example, consider transcription of a gene, translation of the mRNA, reversible dimerisation of the resulting protein, and degradation of each species.

State: $\mathbf{n} = [M, P, D] = [\text{mRNA}, \text{protein}, \text{dimer}]$

Reactions:	State change vectors:
1. Transcription: $\emptyset \rightarrow M$	$\mathbf{v}_1 = [1, 0, 0]$
2. Translation: $\emptyset \rightarrow P$	$\mathbf{v}_2 = [0, 1, 0]$
3. Dimerisation: $P + P \rightarrow D$	$\mathbf{v}_3 = [0, -2, 1]$
4. Dissociation: $D \rightarrow P + P$	$\mathbf{v}_4 = [0, 2, -1]$
5. M degradation: $M \rightarrow \emptyset$	$\mathbf{v}_5 = [-1, 0, 0]$
6. P degradation: $P \rightarrow \emptyset$	$\mathbf{v}_6 = [0, -1, 0]$
7. D degradation: $D \rightarrow \emptyset$	$\mathbf{v}_7 = [0, 0, -1]$

The Master Equation

- Let the **rate** of reaction j be $r_j(\mathbf{n})$.
 In other words, the probability that reaction j occurs in a small time interval Δt is $r_j(\mathbf{n}) \Delta t$.
- Note that this probability depends on the network state.
- Note also that it is assumed that the reaction occurs instantaneously.
- Let $P(\mathbf{n}, t)$ be the probability that the network is in state \mathbf{n} at time t . The **master equation** describes how this probability changes in time, given a set of elementary reactions.

The Master Equation

- P_{arrive} = Prob. that network arrives in state \mathbf{n} in time $[t, t+\Delta t]$.
- P_{leave} = Prob. that network leaves state \mathbf{n} in time $[t, t+\Delta t]$.
- P_{stay} = Prob. that network stays in state \mathbf{n} in time $[t, t+\Delta t]$.

$$P_{\text{arrive}} = \Delta t \sum_{j=1}^M r_j(\mathbf{n} - \mathbf{v}_j) P(\mathbf{n} - \mathbf{v}_j, t),$$

$$P_{\text{leave}} = \Delta t \sum_{j=1}^M r_j(\mathbf{n}) P(\mathbf{n}, t),$$

$$P_{\text{stay}} = \left[1 - \Delta t \sum_{j=1}^M r_j(\mathbf{n}) \right] P(\mathbf{n}, t).$$

The Master Equation

$$P(\mathbf{n}, t + \Delta t) = P_{\text{stay}} + P_{\text{arrive}}$$

$$= \left[1 - \Delta t \sum_{j=1}^M r_j(\mathbf{n}) \right] P(\mathbf{n}, t) + \Delta t \sum_{j=1}^M r_j(\mathbf{n} - \mathbf{v}_j) P(\mathbf{n} - \mathbf{v}_j, t)$$

$$P(\mathbf{n}, t + \Delta t) - P(\mathbf{n}, t) = \Delta t \sum_{j=1}^M \left[r_j(\mathbf{n} - \mathbf{v}_j) P(\mathbf{n} - \mathbf{v}_j) - r_j(\mathbf{n}) P(\mathbf{n}) \right].$$

In the limit $\Delta t \rightarrow 0$, we get

$$\frac{dP(\mathbf{n}, t)}{dt} = \sum_{j=1}^M \left[r_j(\mathbf{n} - \mathbf{v}_j) P(\mathbf{n} - \mathbf{v}_j) - r_j(\mathbf{n}) P(\mathbf{n}) \right].$$

Simulating paths of the Master Equation

- The master equation encodes a continuous time, discrete state Markov process.
- The master equation is hard to solve, so often rely on simulating "paths" (i.e. stochastic realisations of the evolution of the state via elementary reactions that occur with probabilities in accordance with the master equation).
- First determine the probabilities of reactions occurring.
- Determine from this the next reaction time (drawn from an exponential distribution).
- Choose the reaction type to implement, according to fractional rates.

Simulating paths of the Master Equation

- Prob that some reaction will occur in time $\Delta t = \Delta t \sum_{j=1}^M r_j(\mathbf{n}) \equiv \Delta t r_0$.
- Let $p(t)$ = prob that the system has not left state \mathbf{n} at time t given that it was at \mathbf{n} at $t = 0$.
- Then $p(t + \Delta t) - p(t) = -r_0 p(t) \Delta t$
- In the limit: $\frac{dp}{dt} = -r_0 p(t) \Rightarrow p(t) = \exp(-r_0 t)$.
- Prob that a reaction has occurred at time t is $F(t) = 1 - p(t) = 1 - \exp(-r_0 t)$.

The stochastic simulation algorithm (SSA)

- Prob that a reaction has occurred at time t is $F(t) = 1 - p(t) = 1 - \exp(-r_0 t)$.
- Therefore, the system residence time in state \mathbf{n} (i.e. the time to the next reaction) is exponentially distributed with average = $1/r_0$.
- The probability that the first reaction to occur is of type j is its relative contribution to the total rate, i.e. $r_j(\mathbf{n})/r_0$.
- Algorithm (Doob, 1945; Gillespie, 1976,77):
 - Pick the next reaction time from an exponential distribution,
 - Choose a reaction type according to the fractional rates.

The stochastic simulation algorithm (SSA)

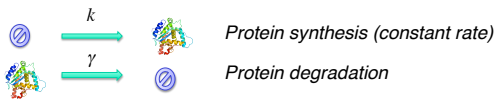
- Algorithm (Doob, 1945; Gillespie, 1976,77):
 - Pick the next reaction time from an exponential distribution,
 - Choose a reaction type according to the fractional rates.
- In practice:
 - Draw two random numbers s_1 and s_2 from the uniform distribution in the unit interval.
 - Reaction time = $\frac{1}{r_0(\mathbf{n})} \ln\left(\frac{1}{s_1}\right)$
 - Reaction type j is the smallest integer such that $\sum_{j=1}^i r_j(\mathbf{n}) > s_2 r_0(\mathbf{n})$.

The stochastic simulation algorithm (SSA)

1. Specify reaction stoichiometry (state update vectors ν_j), reaction propensities and simulation time.
2. Specify initial state $\mathbf{n} = \mathbf{n}_0$ and time $t = t_0$.
3. Calculate the sum of all reaction propensities [$= r_0(\mathbf{n})$].
4. Generate two random numbers s_1 and s_2 from the uniform distribution in the unit interval.
 - Time to next reaction $\theta = \frac{1}{r_0(\mathbf{n})} \ln\left(\frac{1}{s_1}\right)$
 - Reaction type j is the smallest integer such that $\sum_{j=1}^j r_j(\mathbf{n}) > s_2 r_0(\mathbf{n})$.
5. Update time $t \rightarrow t + \theta$.
6. Update state $\mathbf{n} \rightarrow \mathbf{n} + \nu_j$ according to reaction j .
7. Recalculate propensities.
8. Go to 3.

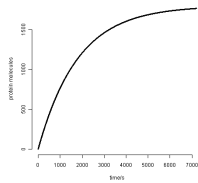
Simple Example

- A constitutively expressed protein
- Two types of reaction:



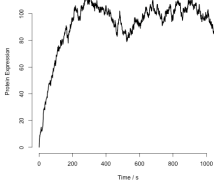
ODE Model:

$$\frac{dP}{dt} = k - \gamma P$$



- Parameters:
- $k = 1 \text{ Ms}^{-1}$
 - $\gamma = 1/1800 \text{ s}^{-1}$
 - $P_0 = 0$

Stochastic Simulation



- Parameters:
 - $k = 1 \text{ s}^{-1}$
 - $\gamma = 0.01 \text{ s}^{-1}$
 - $P_0 = 0$
- Random fluctuations in numbers of molecules

Each Simulation Run is Different

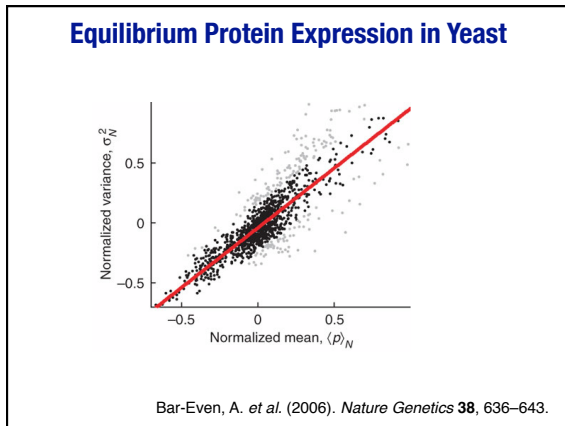
- Two simulations show different time courses because of stochasticity
- Statistical properties are the same
- Average of many stochastic simulations can look like the deterministic model

Comparison to Deterministic Simulation

- Both models have same mean and rise to that mean
- Stochastic model includes fluctuations about mean
- In this case, the mean is as given by the deterministic model

Equilibrium Distribution

- Noise associated with protein synthesis is (approximately) Poisson
- Key property: **variance of equilibrium distribution is proportional to mean**

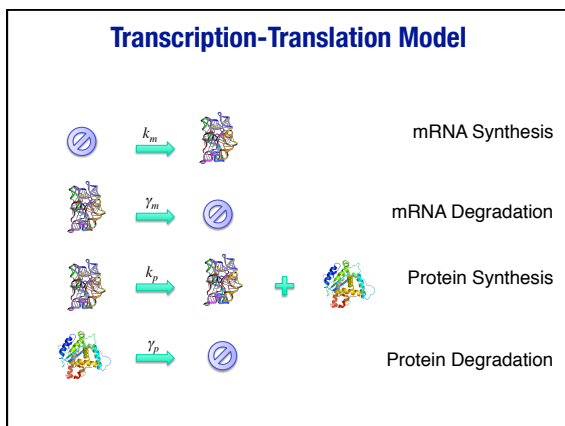


How Important is Poisson Noise?

Percentage error in mean (coefficient of variability) proportional to $1/\sqrt{\text{mean}}$.

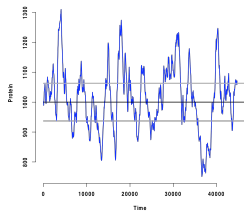
Stochasticity is particularly important in systems with low numbers of reactants

Number of molecules	Molar (bacterial cell)	Processes	c.v.
1	1nM	Genetic	100%
10	10nM	↕	32%
100	100nM		10%
1000	1μM		3%
10 ⁶	1mM		0.1%



Example Simulation: Supra-Poisson Variability

Protein noise can be greater than Poisson if synthesis is much faster than degradation.



Supra-Poisson variability

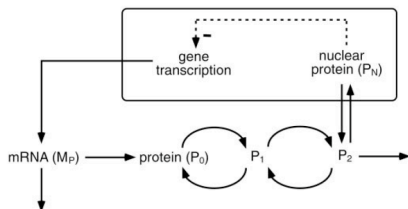
- Mean (mRNA) = 10.1
- Var (mRNA) = 11.0
- Mean (protein) = 1015
- Var (protein) = 11033

Examples of Behaviour

- As system volume gets large, mean of stochastic model can behave like deterministic model
- But individual realizations can be quite different!
 - Oscillations in stochastic model not seen in deterministic model
 - Mean of stochastic system different from deterministic model
 - Stochastic switching between (quasi) steady states

A simple model for circadian oscillations

Single gene negative feedback with protein phosphorylation and nucleo-cytoplasmic movement of protein (“clock protein”).



Gonze, Halloy, Goldbeter. *Proc. Natl. Acad. Sci. USA* **99**, 673–678 (2002).

A simple model for circadian oscillations

$$\frac{dM_P}{dt} = v_1 \frac{K_I^n}{K_I^n + P_N^n} - v_w \frac{M_P}{K_m + M_P}$$

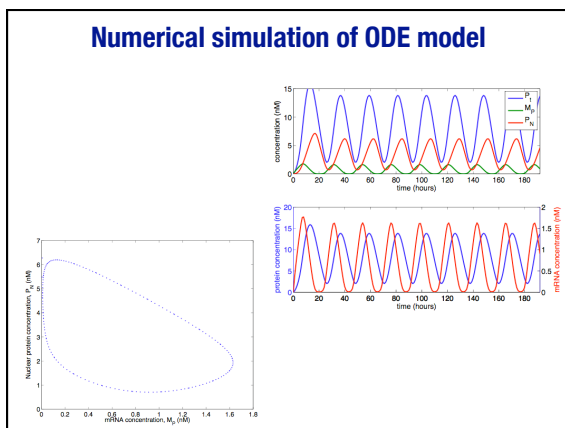
$$\frac{dP_0}{dt} = k_1 M_P - v_1 \frac{P_0}{K_1 + P_0} + v_2 \frac{P_1}{K_2 + P_1}$$

$$\frac{dP_1}{dt} = v_1 \frac{P_0}{K_1 + P_0} - v_2 \frac{P_1}{K_2 + P_1} - v_3 \frac{P_1}{K_3 + P_1} + v_4 \frac{P_2}{K_4 + P_2}$$

$$\frac{dP_2}{dt} = v_3 \frac{P_1}{K_3 + P_1} - v_4 \frac{P_2}{K_4 + P_2} - v_d \frac{P_2}{K_d + P_2} - k_2 P_2 + k_2 P_N$$

$$\frac{dP_N}{dt} = k_1 P_2 - k_2 P_N$$

Note the nonlinear functional forms for degradation of mRNA and P_2 .

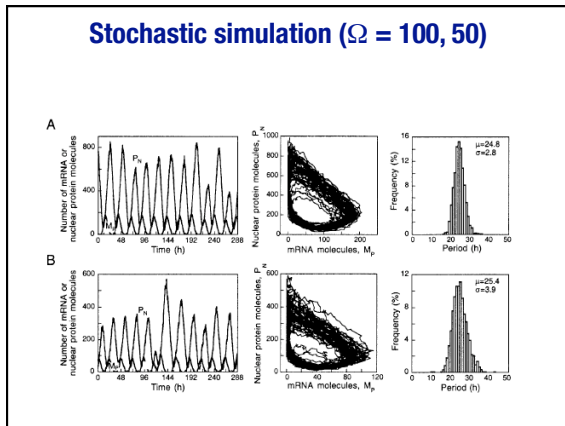


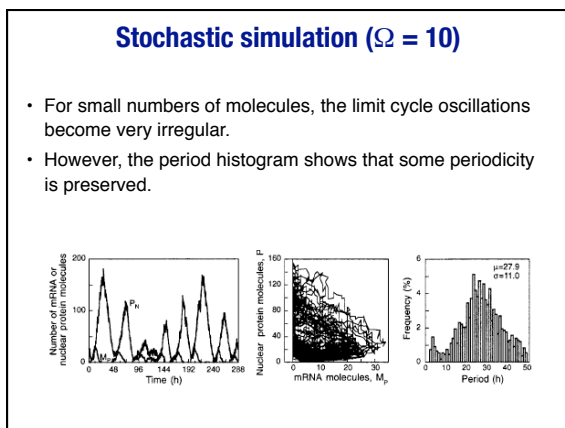
Elementary reactions: transcription

Need to represent binding of transcription factor P_N to multiple sites on the DNA. These are the elementary reactions that we need for the SSA.

Consider 4 binding sites. Transcription is blocked only if all 4 sites are bound by P_N .

Reaction number	Reaction step	Probability of reaction
1	$G + P_N \xrightarrow{a_1} GP_N$	$w_1 = a_1 \times G \times P_N / \Omega$
2	$GP_N \xrightarrow{d_1} G + P_N$	$w_2 = d_1 \times GP_N$
3	$GP_N + P_N \xrightarrow{a_2} GP_{N2}$	$w_3 = a_2 \times GP_N \times P_N / \Omega$
4	$GP_{N2} \xrightarrow{d_2} GP_N + P_N$	$w_4 = d_2 \times GP_{N2}$
5	$GP_{N2} + P_N \xrightarrow{a_3} GP_{N3}$	$w_5 = a_3 \times GP_{N2} \times P_N / \Omega$
6	$GP_{N3} \xrightarrow{d_3} GP_{N2} + P_N$	$w_6 = d_3 \times GP_{N3}$
7	$GP_{N3} + P_N \xrightarrow{a_4} GP_{N4}$	$w_7 = a_4 \times GP_{N3} \times P_N / \Omega$
8	$GP_{N4} \xrightarrow{d_4} GP_{N3} + P_N$	$w_8 = d_4 \times GP_{N4}$
9	$[G, GP_N, GP_{N2}, GP_{N3}] \xrightarrow{v_1} M_P$	$w_9 = v_1 \times (G + GP_N + GP_{N2} + GP_{N3})$





Representing transcription in SSA

- The SSA assumes that all reactions are elementary (i.e. instantaneous mass-action kinetics).
- Regulated transcription is not an elementary reaction; how can we represent it?
 1. Model multiple transcription factor binding events as elementary reactions (but can we parameterise this?)
 2. Relax the requirement for only elementary reactions and use a nonlinear propensity function.

Representing transcription in SSA

- We have already seen an example of representing multiple transcription factor bindings:

$$\text{transcription rate} = \frac{K^n}{K^n + P^n}$$

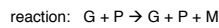
Reaction number	Reaction step	Probability of reaction
1	$G + P_N \xrightarrow{a_1} GP_N$	$w_1 = a_1 \times G \times P_N / \Omega$
2	$GP_N \xrightarrow{d_1} G + P_N$	$w_2 = d_1 \times GP_N$
3	$GP_N + P_N \xrightarrow{a_2} GP_N P_N$	$w_3 = a_2 \times GP_N \times P_N / \Omega$
4	$GP_N P_N \xrightarrow{d_2} GP_N + P_N$	$w_4 = d_2 \times GP_N P_N$
5	$GP_N P_N + P_N \xrightarrow{a_3} GP_N P_N P_N$	$w_5 = a_3 \times GP_N P_N \times P_N / \Omega$
6	$GP_N P_N P_N \xrightarrow{d_3} GP_N P_N + P_N$	$w_6 = d_3 \times GP_N P_N P_N$
7	$GP_N P_N P_N + P_N \xrightarrow{a_4} GP_N P_N P_N P_N$	$w_7 = a_4 \times GP_N P_N P_N \times P_N / \Omega$
8	$GP_N P_N P_N P_N \xrightarrow{d_4} GP_N P_N P_N + P_N$	$w_8 = d_4 \times GP_N P_N P_N P_N$
9	$[G, GP_N, GP_N P_N, GP_N P_N P_N] \xrightarrow{v} M_P$	$w_9 = v \times (G + GP_N + GP_N P_N + GP_N P_N P_N)$

Gonze, Halloy, Goldbeter. *Proc. Natl. Acad. Sci. USA* **99**, 673–678 (2002).

Representing transcription in SSA

- Alternatively, we can represent transcription as a single reaction with a nonlinear propensity:

$$\text{transcription rate} = v \frac{K^n}{K^n + P^n}$$



$$\text{propensity} = v \frac{K^n}{K^n + N_p^n}$$

where N_p is the number of molecules of transcription factor.

Simpler Oscillator Model (after Goodwin)

- Single gene negative feedback loop with dimerisation

$$\frac{dm}{dt} = \frac{k_1}{1 + (p_2/p_0)^n} - d_1 m$$

$$\frac{dp_1}{dt} = k_2 m - 2k_3 p_1^2 + 2k_4 p_2 - d_2 p_1$$

$$\frac{dp_2}{dt} = k_3 p_1^2 - k_4 p_2 - d_3 p_2$$

Simpler Oscillator Model (after Goodwin)

	Reaction	Rate
1. transcription	$\emptyset \rightarrow M$	$\frac{k_1}{1 + (p_1/p_0)^n}$
2. M degradation	$M \rightarrow \emptyset$	$d_m m$
3. translation	$\emptyset \rightarrow P_1$	$k_p m$
4. P_1 dimerisation	$P_1 + P_1 \rightarrow P_2$	$k_d p_1(p_1 - 1)$
5. P_1 degradation	$P_1 \rightarrow \emptyset$	$d_p p_1$
6. P_2 dissociation	$P_2 \rightarrow P_1 + P_1$	$k_s p_2$
7. P_2 degradation	$P_2 \rightarrow \emptyset$	$d_p p_2$

Stoichiometric matrix: $v = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & -1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & -1 \end{pmatrix}$

Simulations: Stochastic Resonance

$k_1 = 2, k_2 = 0.5, k_3 = 0.0001, k_4 = 0.001, d_1 = 0.03, d_2 = d_3 = 0.02, p_0 = 50$

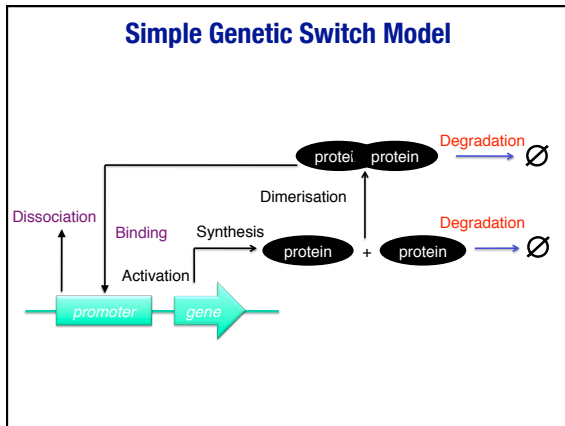
$n = 7$

$n = 10$

Stochastic model can show sustained oscillations even when the ODE model has damped oscillations

Simulations

Even when the ODE model has sustained oscillations ($n = 10$) the stochastic oscillations can have a significantly greater amplitude

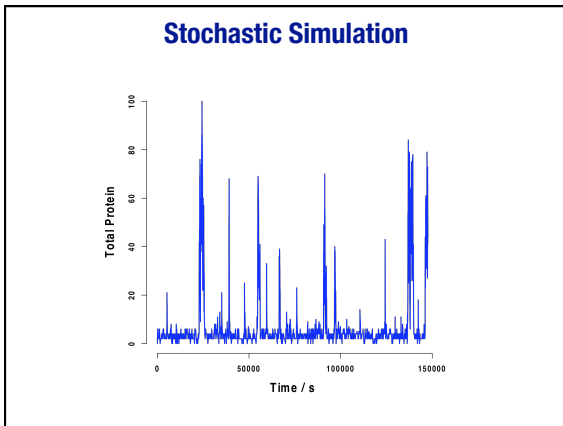


Elementary Reactions

Process	Reaction	Parameter
Basal synthesis	$DNA \rightarrow DNA + Prot$	k_0
Dimerization	$Prot + Prot \rightarrow Dimer$	k_{dim}
Degradation	$Prot \rightarrow \emptyset$	μ
Degradation	$Dimer \rightarrow \emptyset$	μ
Binding	$DNA + Dimer \rightarrow Complex$	k_{ass}
Dissociation	$Complex \rightarrow DNA + Dimer$	k_{dis}
Activated synthesis	$Complex \rightarrow Complex + Prot$	k_1

Realistic Parameters (1 μm cell)

Parameter	Value	Comment
k_0	$0.03 s^{-1}$	Molecules per cell
k_{dim}	$0.0001 s^{-1}$	Diffusion limited
μ	$0.01 s^{-1}$	Rapidly turned over protein (monomer and dimer)
k_{ass}	$0.01 s^{-1}$	Diffusion limited: faster because of 1D diffusion
k_{dis}	$0.08 s^{-1}$	Binding site affinity is 8nM
k_1	$1 s^{-1}$	Transcription, translation and mRNA stability



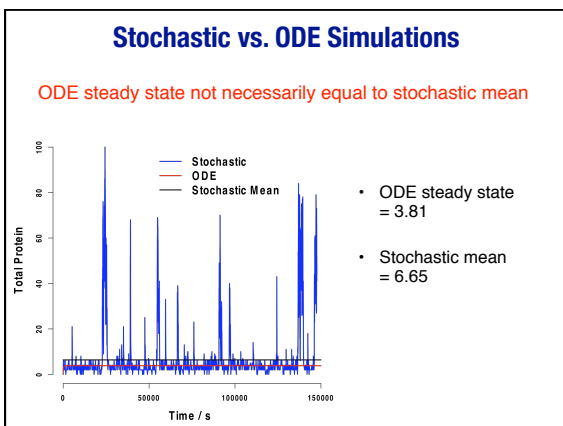
ODE Model

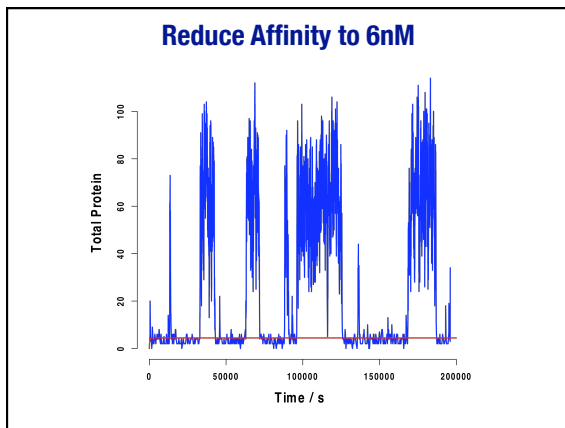
$$\frac{dP_1}{dt} = k_0 D + k_1 C - k_{dim} P_1^2 - \mu P_1$$

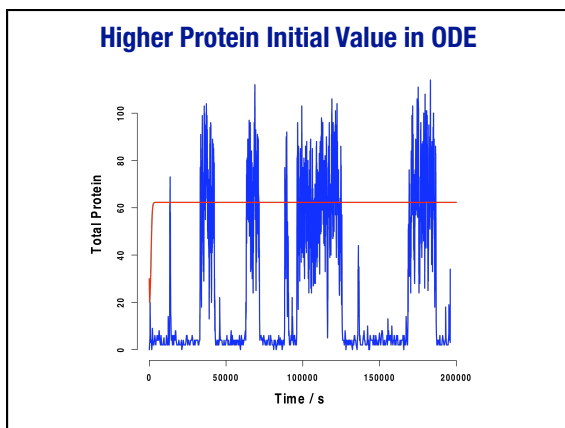
$$\frac{dP_2}{dt} = \frac{1}{2} k_{dim} P_1^2 - \mu P_2 - k_{ass} D P_2 + k_{dis} C$$

$$\frac{dC}{dt} = k_{ass} D P_2 - k_{dis} C$$

$$C + D = D_0$$

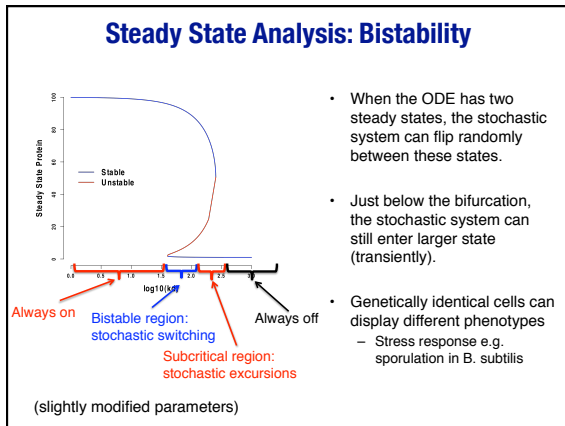






Stochastic Switching

- Stochastic model switches between
 - “on” state in which protein is produced under self-activation
 - “off” state in which protein is produced at basal level
- ODE simulation with a low (0) initial protein value goes to a low steady state
- ODE simulation with a higher initial protein value goes to a higher steady state



- When the ODE has two steady states, the stochastic system can flip randomly between these states.
- Just below the bifurcation, the stochastic system can still enter larger state (transiently).
- Genetically identical cells can display different phenotypes
 - Stress response e.g. sporulation in *B. subtilis*

