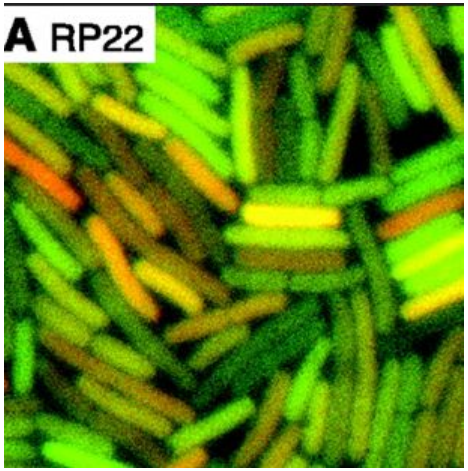
Stochastic modeling of biochemical and genetic networks

(adapting notes from my student Brian Ingalls)

Variability in biology

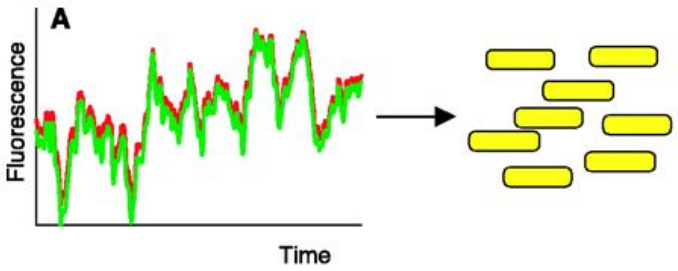
- Development:
  genetically identical but phenotypically different

- Gene expression:
  genetically identical but phenotypically different

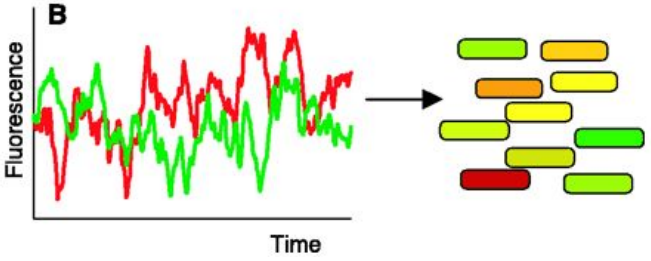


Intrinsic and Extrinsic noise

Elowitz *et. al,* Science, 2002

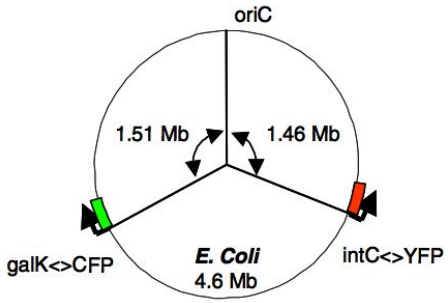


Extrinsic noise

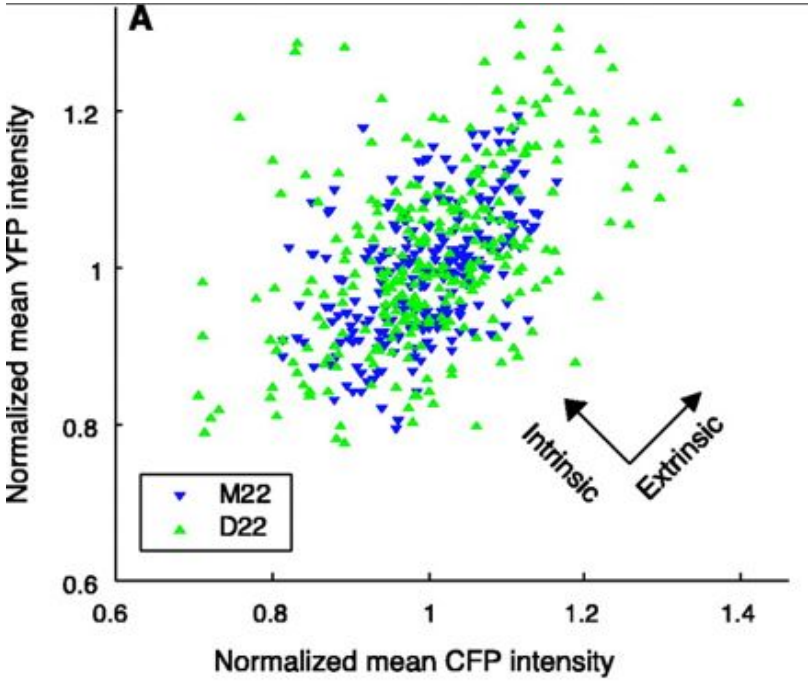


Intrinsic and extrinsic noise

Experimental design:
2 reporter genes in
`equivalent' positions in the genome



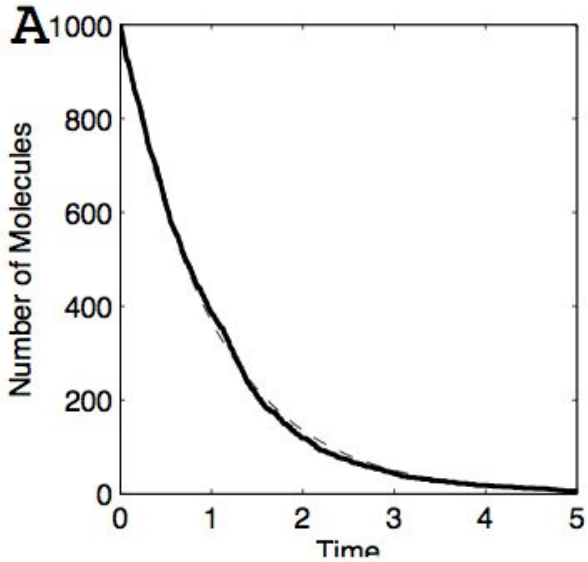
Results



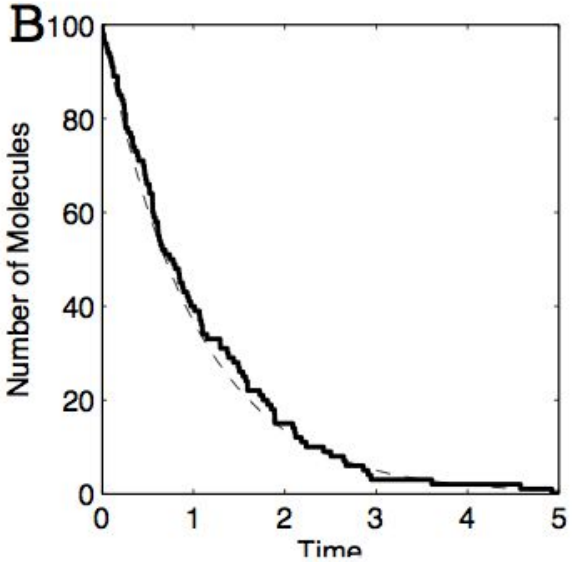
Variability in Molecular Biology

- All reaction events are driven by thermal agitation ('noise')
- Over large numbers of events, 'noise' is averaged out
- Variability can be a nuisance, or can be useful (e.g. persistence)
- At the cellular level, we distinguish
  – *intrinsic noise* (thermal agitation)
  – *extrinsic noise* (variability in localization, environmental factors)

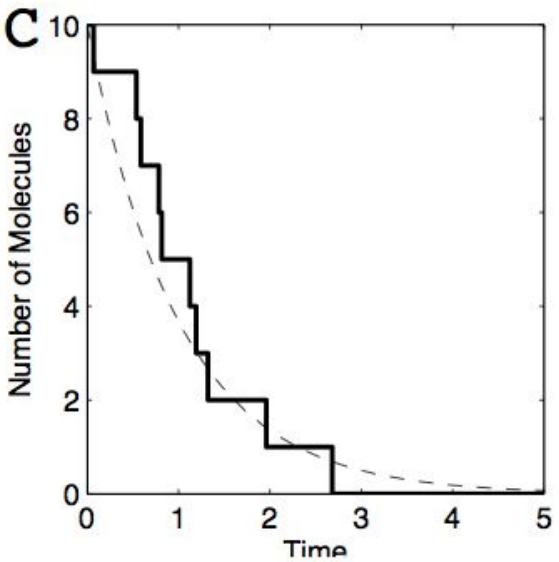
Deterministic behaviour vs. random (stochastic) behaviour



1000 molecules



100 molecules



10 molecules

$$A \rightarrow 0$$

da/dt = -k a deterministic limit as # molecules large

A discrete modeling framework

- Spatial homogeneity assumed
- System State: number of molecules of each species, collected in vector **N**
- Reaction k characterized by
  - *stoichiometry* vector $\mathbf{s}_k$
  - reaction *propensity* a_k

reaction propensities are basically the reaction rates

[ignoring volume, which is part of the propensities]

Example

$$R_1: \; A + B \xrightarrow{k_1} C \qquad\qquad R_2: \; C \xrightarrow{k_2} \emptyset$$

State: $\mathbf{N} = \begin{bmatrix} N_A \\ N_B \\ N_C \end{bmatrix}$

Stoichiometry:

$$\mathbf{s}_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \begin{matrix} \leftarrow A \\ \leftarrow B \\ \leftarrow C \end{matrix} \qquad \mathbf{s}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \begin{matrix} \leftarrow A \\ \leftarrow B \\ \leftarrow C \end{matrix}$$

Propensity: (mass-action)

$$a_1(\mathbf{N}) = k_1 N_A N_B \qquad a_2(\mathbf{N}) = k_2 N_C$$

A technical complication

for a homodimer type of reaction such as

$$A + A \rightarrow B$$

we cannot write $k * a^2$,

because the reaction cannot take place if $a = 1$!

the correct formula has: $k * a\,(a-1)/2$

i.e. we need to consider all possible pairs

when numbers are large, $a-1 \approx a$, so $k * a\,(a-1)/2 \approx (k/2)a^2$

and the ½ can be absorbed; hence we don't worry about this

see my notes for more, including effects of volume (also not trivial)

Example

$$R_1: \; A + B \xrightarrow{k_1} C \qquad R_2: \; C \xrightarrow{k_2} 0$$

Stoichiometry:

$$\mathbf{s}_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \begin{matrix} \leftarrow A \\ \leftarrow B \\ \leftarrow C \end{matrix} \qquad \mathbf{s}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \begin{matrix} \leftarrow A \\ \leftarrow B \\ \leftarrow C \end{matrix}$$

State:

$$\mathbf{N} = \begin{bmatrix} N_A \\ N_B \\ N_C \end{bmatrix} = \begin{bmatrix} 11 \\ 6 \\ 3 \end{bmatrix}$$

when reaction 1 “fires”:

$$\mathbf{N} \rightarrow \mathbf{N} + \mathbf{s}_1 = \begin{bmatrix} 11 \\ 6 \\ 3 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 5 \\ 4 \end{bmatrix}$$

Stochastic Dynamic Model: ("Markov") stochastic process

- Cannot confidently characterize state of the system at time t (as in **N**=**N**(t))
- Instead, describe *probability* of the state taking a particular value at time t:

$P(\mathbf{N}, t) =$ Probability that the system is in state **N** at time t

e.g.

$$P\left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, 2\right) = 0.03$$

Stochastic dynamic model

[recall, in queueing theory: Poisson arrivals, exponential waiting times]

Assumption: for *small* time-increments *dt*:

(1) at most one reaction can occur during intervals of length *dt*

(2) probability that reaction R_k occurs in the interval *[t, t+dt]* is:

$$a_k(\mathbf{N}(t))\,dt$$

(3) thus, probability that *no reactions occur* during *[t, t+dt]* is:

$$1 - \sum_k a_k(\mathbf{N}(t))dt$$

Sum over all reactions, assumed independent

(actually, even if two reactions occur, it won't make any difference, because we'll get terms $(dt)^2$ that will go to zero when forming the ODE next)

Stochastic dynamic model: probability balance

$$P(\mathbf{N}, t+dt) = P(\mathbf{N}, t) \cdot \underbrace{\left(1 - \sum_k a_k(\mathbf{N})\, dt\right)}_{\text{probability of no reactions firing}}$$

$$+ \sum_k \underbrace{P(\mathbf{N} - \mathbf{s}_k, t)\, a_k(\mathbf{N} - \mathbf{s}_k)\, dt.}_{\text{probability of reaction } R_k \text{ occurring while in state } \mathbf{N} - \mathbf{s}_k}$$

to be in a state N at time $t+dt$, either
the state was N at time t, and no reactions happened, or
some reaction happened, which means that:
the state was N-s_k at time t (for each possible reaction k, one such term, and I am assuming at most one could happen), which is an event with probability equal to (being there) x (prob of transition from there)

Stochastic dynamic model: probability balance

Example: single species A

State: $N = N_A$

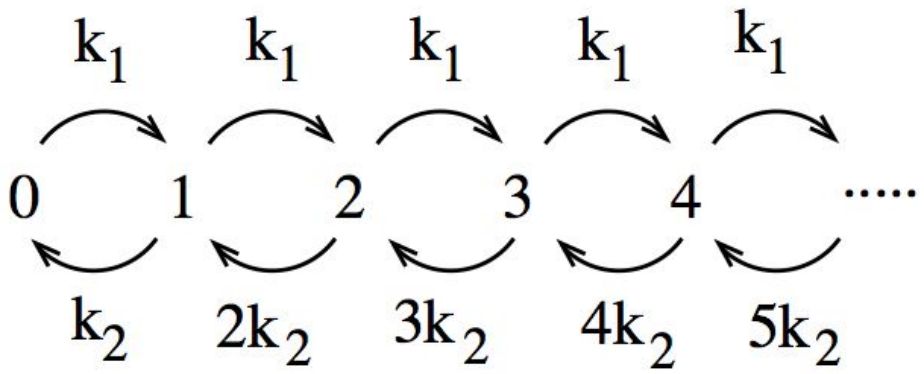
Reactions: $R_1 : 0 \xrightarrow{k_1} A$ $R_2 : A \xrightarrow{k_2} 0$

– stoichiometry $s_1 = [1]$ $s_2 = [-1]$

– propensity $a_1 = k_1$ $a_2 = k_2 N_A$

Stochastic dynamic model: probability balance

Example: state transitions:



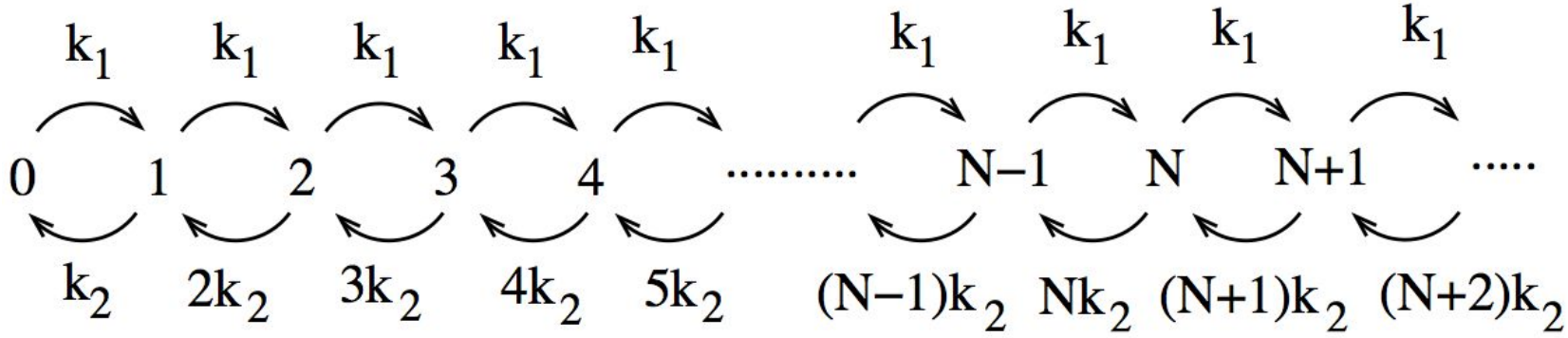
Probability balance:

$$P(0,t+dt) = P(0,t)\,[1-k_1 dt] + P(1,t)\cdot k_2\,dt$$
$$P(1,t+dt) = P(1,t)\,[1-(k_1+k_2)dt] + P(0,t)\cdot k_1\,dt + P(2,t)\cdot 2k_2\,dt$$

(note the four arrows going into or out of node “1”)

Stochastic dynamic model: probability balance

Example: state transitions:



Probability balance:

$$
\begin{aligned}
P(0,t+dt) &= P(0,t)\,[1-k_1 dt] + P(1,t)\cdot k_2\,dt \\
P(1,t+dt) &= P(1,t)\,[1-(k_1+k_2)dt] + P(0,t)\cdot k_1\,dt + P(2,t)\cdot 2k_2\,dt \\
P(2,t+dt) &= P(2,t)\,[1-(k_1+2k_2)dt] + P(1,t)\cdot k_1\,dt + P(3,t)\cdot 3k_2\,dt \\
P(3,t+dt) &= P(3,t)\,[1-(k_1+3k_2)dt] + P(2,t)\cdot k_1\,dt + P(4,t)\cdot 4k_2\,dt \\
&\vdots \\
P(N,t+dt) &= P(N,t)\,[1-(k_1+Nk_2)dt] + P(N-1,t)\cdot k_1\,dt + P(N+1,t)\cdot (N+1)k_2\,dt \\
&\vdots
\end{aligned}
$$

Differential equation for probability distribution:

Probability balance:

$$P(\mathbf{N}, t+dt) = P(\mathbf{N}, t) \cdot \underbrace{\left(1 - \sum_k a_k(\mathbf{N})\, dt\right)}_{\text{probability of no reactions firing}} + \sum_k \underbrace{P(\mathbf{N} - \mathbf{s}_k, t)\, a_k(\mathbf{N} - \mathbf{s}_k)\, dt.}_{\text{probability of reaction } R_k \text{ occurring while in state } \mathbf{N} - \mathbf{s}_k}$$

Subtracting $P(\mathbf{N}, t)$ from both sides:

$$P(\mathbf{N}, t+dt) - P(\mathbf{N}, t) = -P(\mathbf{N}, t) \left(\sum_k a_k(\mathbf{N})\, dt\right) + \sum_k P(\mathbf{N} - \mathbf{s}_k, t)\, a_k(\mathbf{N} - \mathbf{s}_k)\, dt$$

Dividing by *dt*:

$$\frac{P(\mathbf{N}, t+dt) - P(\mathbf{N}, t)}{dt} = -P(\mathbf{N}, t) \left(\sum_k a_k(\mathbf{N})\right) + \sum_k P(\mathbf{N} - \mathbf{s}_k, t)\, a_k(\mathbf{N} - \mathbf{s}_k)$$

Limit as *dt* tends to zero:

$$\frac{d}{dt} P(\mathbf{N}, t) = -P(\mathbf{N}, t) \left(\sum_k a_k(\mathbf{N})\right) + \sum_k P(\mathbf{N} - \mathbf{s}_k, t) a_k(\mathbf{N} - \mathbf{s}_k)$$

Differential equation for probability distribution:

$$\frac{d}{dt}P(\mathbf{N},t) = -P(\mathbf{N},t)\left(\sum_k a_k(\mathbf{N})\right) + \sum_k P(\mathbf{N}-\mathbf{s}_k,t)a_k(\mathbf{N}-\mathbf{s}_k)$$

$$= \sum_k \left(\underbrace{-P(\mathbf{N},t)a_k(\mathbf{N})}_{\text{flow out of state } \mathbf{N}} + \underbrace{P(\mathbf{N}-\mathbf{s}_k,t)a_k(\mathbf{N}-\mathbf{s}_k)}_{\text{flow into state } \mathbf{N}} \right)$$

this is the **Chemical Master Equation** (CME)

The **Chemical Master Equation**

Example system: $0 \xrightarrow{k_1} A \xrightarrow{k_2} 0$

Master equation:

$$
\begin{aligned}
\frac{d}{dt}P(0,t) &= -P(0,t)\,k_1 + P(1,t)\,k_2 \\
\frac{d}{dt}P(1,t) &= -P(1,t)\,(k_1+k_2) + P(0,t)k_1 + P(2,t)\,2k_2 \\
\frac{d}{dt}P(2,t) &= -P(2,t)\,(k_1+2k_2) + P(1,t)k_1 + P(3,t)\,3k_2 \\
&\vdots \\
\frac{d}{dt}P(N,t) &= -P(N,t)\,(k_1+Nk_2) + P(N-1,t)\,k_1 + P(N+1,t)\,(N+1)k_2 \\
&\vdots
\end{aligned}
$$

The master equation is a system of (typically infinitely many) differential equations

Master equation: example

Network:

$$R_1: \; A \xrightarrow{k_1} B \qquad R_2: \; B \xrightarrow{k_2} A$$

State: $\mathbf{N} = (N_A, N_B)$

Two molecules: $\mathbf{N} = (2,0), (1,1), (0,2)$

Master equation (conservation law makes it a *finite* set!):

$$\begin{aligned}
\frac{d}{dt}P((2,0),t) &= -P((2,0),t)\,2k_1 + P((1,1),t)\,k_2 \\
\frac{d}{dt}P((1,1),t) &= -P((1,1),t)\,k_2 - P((1,1),t)\,k_1 + P((2,0),t)2k_1 + P((0,2),t)\,2k_2 \\
\frac{d}{dt}P((0,2),t) &= -P((0,2),t)\,2k_2 + P((1,1),t)\,k_1
\end{aligned}$$

Master equation: example

Network:

$$R_1: \; A \xrightarrow{k_1} B \qquad R_2: \; B \xrightarrow{k_2} A$$

Master equation:

$$\begin{aligned}
\frac{d}{dt}P((2,0),t) &= -P((2,0),t)\,2k_1 + P((1,1),t)\,k_2 \\
\frac{d}{dt}P((1,1),t) &= -P((1,1),t)\,k_2 - P((1,1),t)\,k_1 + P((2,0),t)2k_1 + P((0,2),t)\,2k_2 \\
\frac{d}{dt}P((0,2),t) &= -P((0,2),t)\,2k_2 + P((1,1),t)\,k_1
\end{aligned}$$

Steady state:

$$\begin{aligned}
0 &= -P^{ss}(2,0)\,2k_1 + P^{ss}(1,1)\,k_2 \\
0 &= -P^{ss}(1,1)\,k_2 - P^{ss}(1,1)\,k_1 + P^{ss}(2,0)\,2k_1 + P^{ss}(0,2)\,2k_2 \\
0 &= -P^{ss}(0,2)\,2k_2 + P^{ss}(1,1)\,k_1
\end{aligned}$$

$$P^{ss}(2,0) = \frac{k_2^2}{(k_1+k_2)^2}, \qquad P^{ss}(1,1) = \frac{2k_1k_2}{(k_1+k_2)^2}, \qquad P^{ss}(0,2) = \frac{k_1^2}{(k_1+k_2)^2}$$

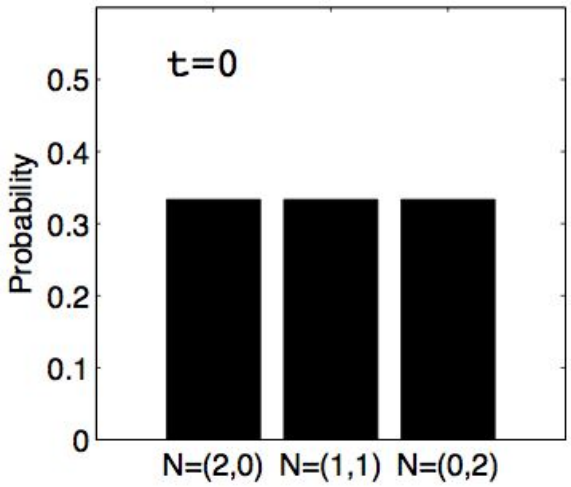
can prove binomial - in fact, can write ss distro in "closed form" for WR deficiency zero networks!
(Poisson if no conservations)

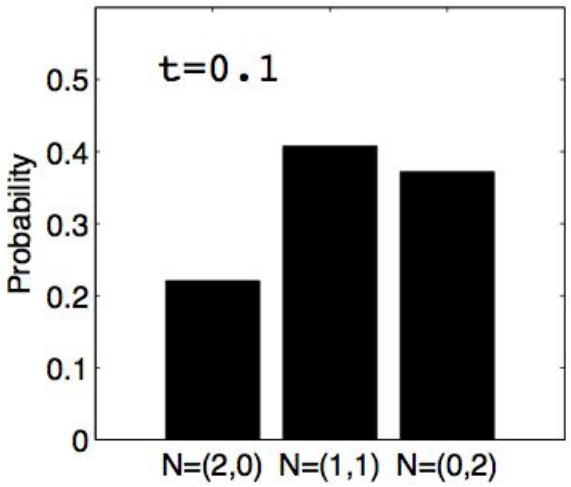
Master equation: example

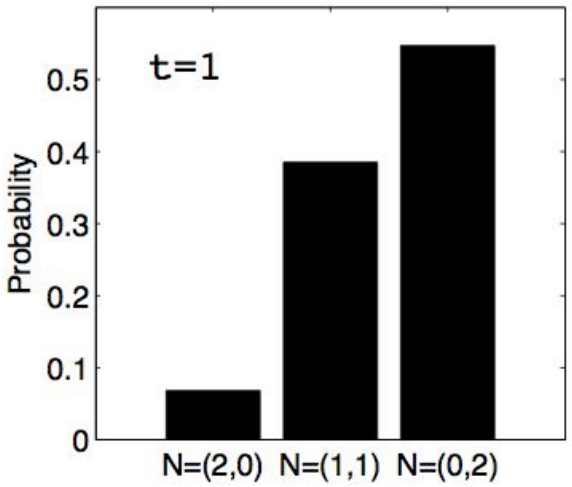
Network:

$$R_1: A \xrightarrow{k_1} B \qquad R_2: B \xrightarrow{k_2} A$$

Dynamics: two molecules ($k_1 = 3$, $k_2 = 1$)







$$P^{ss}(2,0) = \frac{k_2^2}{(k_1+k_2)^2}, \qquad P^{ss}(1,1) = \frac{2k_1k_2}{(k_1+k_2)^2}, \qquad P^{ss}(0,2) = \frac{k_1^2}{(k_1+k_2)^2}$$

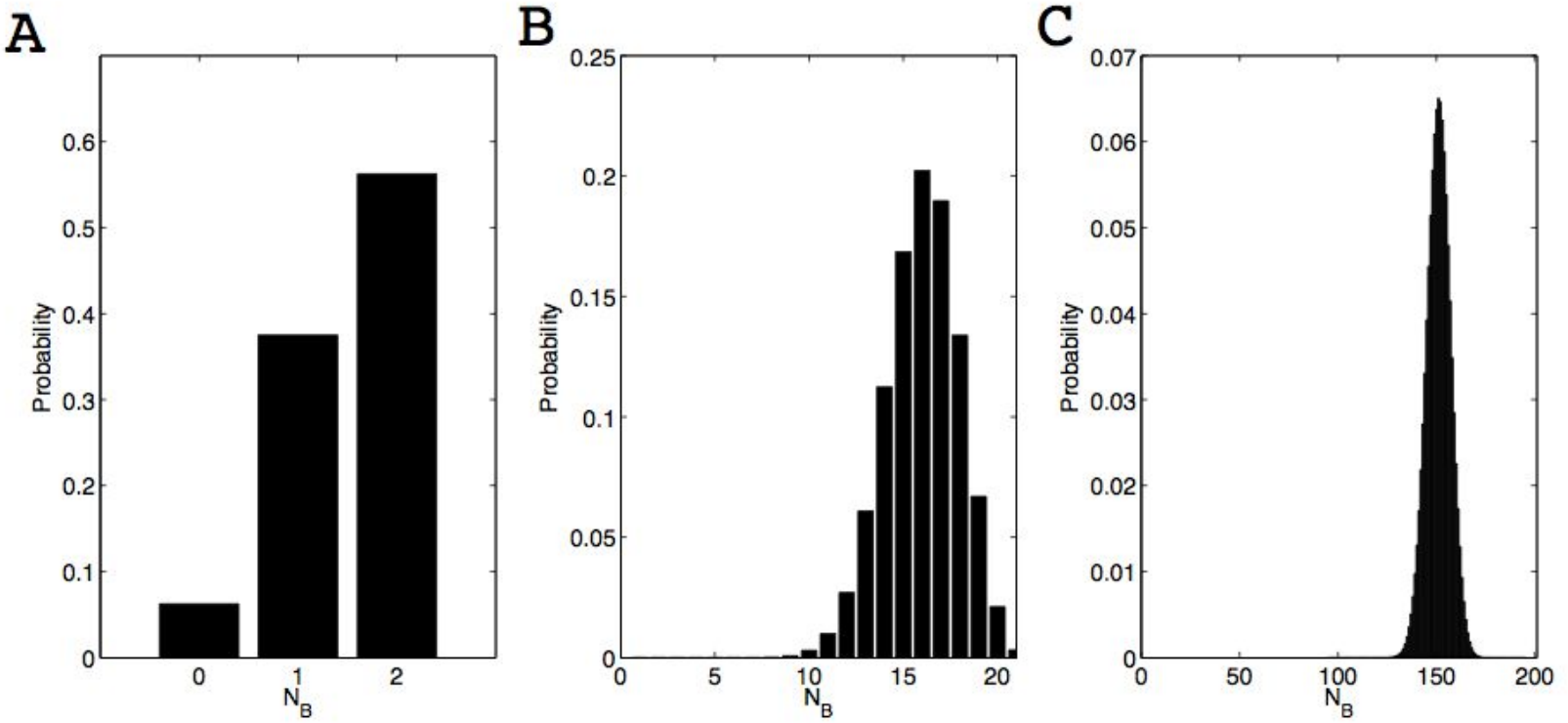
$E(B) = 0.P^{ss}(2,0) + 1.P^{ss}(0,1) + 2.P^{ss}(0,2) = 2\ N\ k_1 / (k_1 + k_2) = ¾*N = 1.5$ when $N = 2$
which is what we get from the ODE by setting $k_1*A = k_2*B$ and $A=N-B$
so in this case (generally true for 0th and first order reactions), mean of stochastic = deterministic
(and also for transients, not just ss)

Master equation: example

Network:

$$R_1: A \xrightarrow{k_1} B \qquad R_2: B \xrightarrow{k_2} A$$

Steadv state



2 molecules 20 molecules 200 molecules

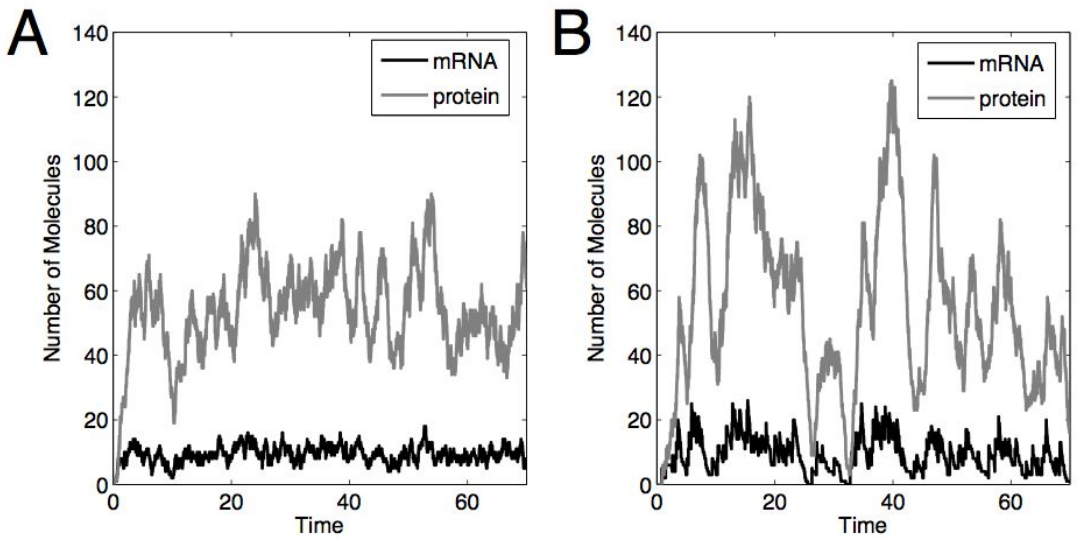
¾*N = ¾*200 = 150, note how distribution peaks at 150 in case of large # molecules!

Master equation

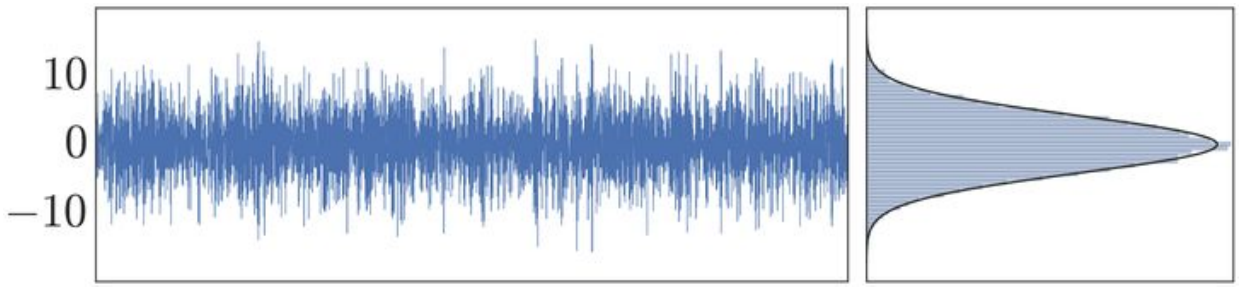
- Approaches the deterministic (differential equation) description at large molecule numbers
- Difficult to analyse. Approximation approaches:
  – moment closure / “fluctuation-dissipation”
  – chemical Langevin equation (stochastic differential equation) - intermediate regimes
  – various finite-dimensional truncations
- Alternative approach: simulation of an ensemble of sample paths:

Gillespie's Stochastic Simulation Algorithm

- Monte Carlo (probabilistic approach)
- Generates individual *sample paths*
- Sample statistics from an *ensemble* of sample paths indicate probability distribution.
- Ergodicity: compute along a single sample path



example [to discuss later] of gene expression model: simple (A) vs bursting (B)



make histograms by "horizontal slices"

Stochastic Simulation Algorithm

- Depends on two *random variables* (variables characterized by probability distributions):

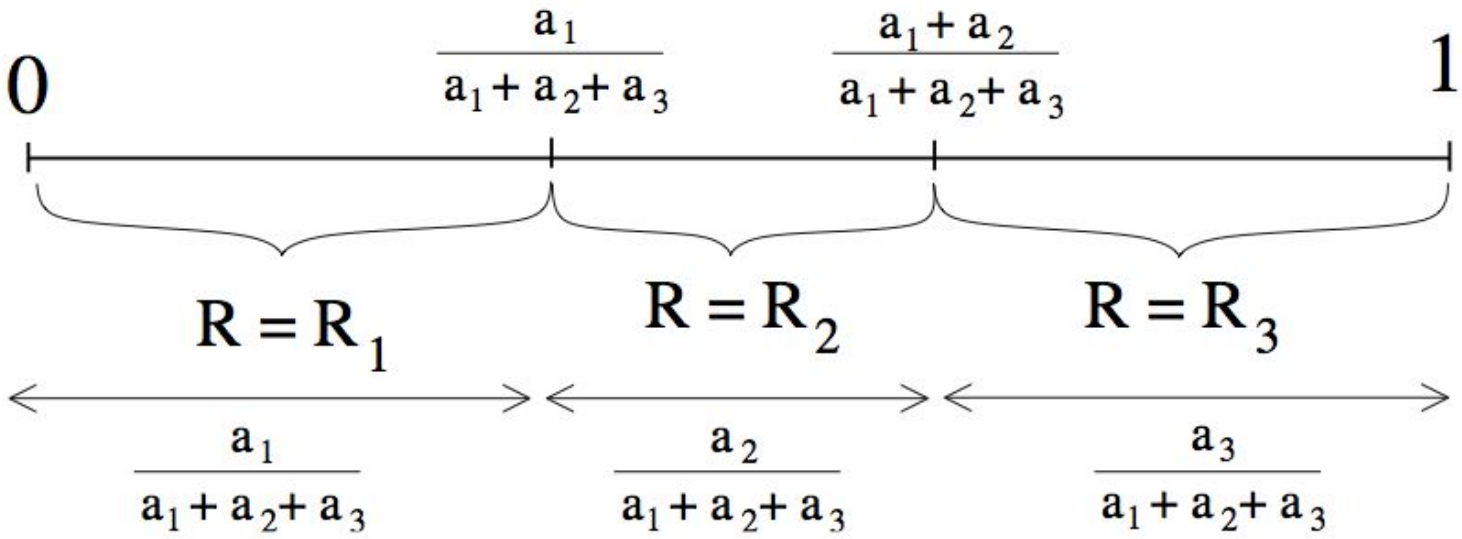
  – *time until next reaction* event *T*
  – *which reaction* occurs next *R*

Determining the next reaction

- Example: Reactions R_1, R_2, R_3,
  propensities a_1, a_2, a_3

$$P(R = R_1) = \frac{a_1}{a_1 + a_2 + a_3} \qquad P(R = R_3) = \frac{a_3}{a_1 + a_2 + a_3}$$

$$P(R = R_2) = \frac{a_2}{a_1 + a_2 + a_3}$$



Determining the next reaction

- Example: Reactions R_1, R_2, R_3,
  propensities a_1, a_2, a_3

Sampling algorithm: *u* drawn from a uniform distribution over [0,1]

$$\text{if} \quad 0 \leq u \leq \frac{a_1}{a_1+a_2+a_3}, \quad \text{then we set } R = R_1$$

$$\text{if} \quad \frac{a_1}{a_1+a_2+a_3} < u \leq \frac{a_1+a_2}{a_1+a_2+a_3}, \quad \text{then we set } R = R_2$$

$$\text{if} \quad \frac{a_1+a_2}{a_1+a_2+a_3} < u \leq \frac{a_1+a_2+a_3}{a_1+a_2+a_3} = 1, \quad \text{then we set } R = R_3.$$

Determining the time to the next reaction

- T can shown to be an *exponential random variable*
- Cumulative distribution function:

$$P(0 \leq T \leq t) = 1 - e^{-at}$$

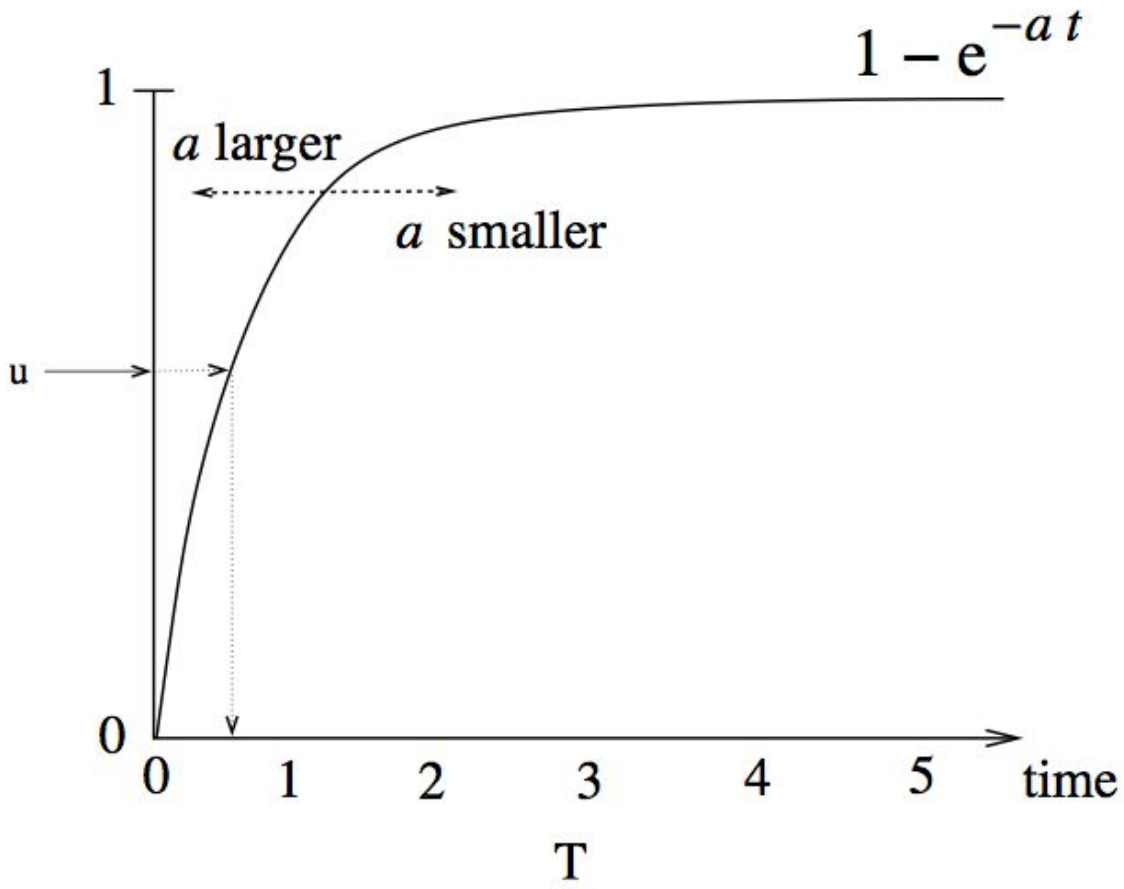
where *a* is the sum of the propensities:

$$a = a_1 + a_2 + a_3.$$

(think of Poisson random arrivals)

Time to the next reaction

Sampling algorithm: *u* drawn from a uniform distribution over [0,1]



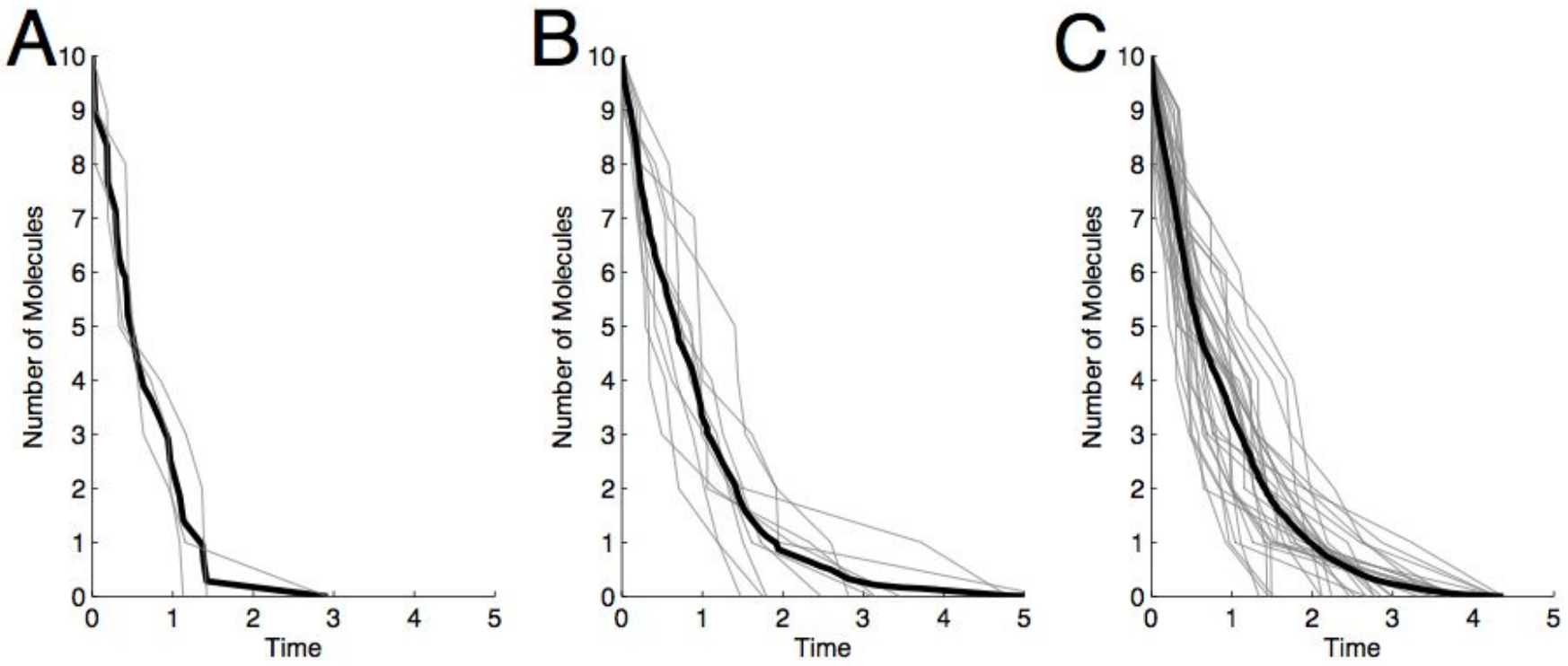
Stochastic Simulation Algorithm

Stochastic Simulation Algorithm (SSA)

1. Set the initial state $\mathbf{N}$. Initialize time t to zero.
2. Calculate the reaction propensities $a_k(\mathbf{N})$.
3. Draw a sample R_k from the random variable R
4. Draw a sample τ from the random variable T
5. Increment the simulation time $t \rightarrow t + \tau$ to account for the elapsed time.
6. Update the state vector $\mathbf{N} \rightarrow \mathbf{N} + \mathbf{s}_k$ to reflect the fact that reaction R_k has occurred.
7. Return to step 2.

Ensemble average: constitutive decay

$$A \xrightarrow{k} 0$$



3 samples 10 samples 30 samples

exponential decay in mean (= deterministic model)

Bursty transcription

Gene expression:

R_1 : (transcription) $\quad 0 \longrightarrow M \quad$ propensity: k_r

R_2 : (translation) $\quad 0 \longrightarrow P \quad$ propensity: $k_p N_M$

R_3 : (degradation) $\quad M \longrightarrow 0 \quad$ propensity: $\delta_r N_M$

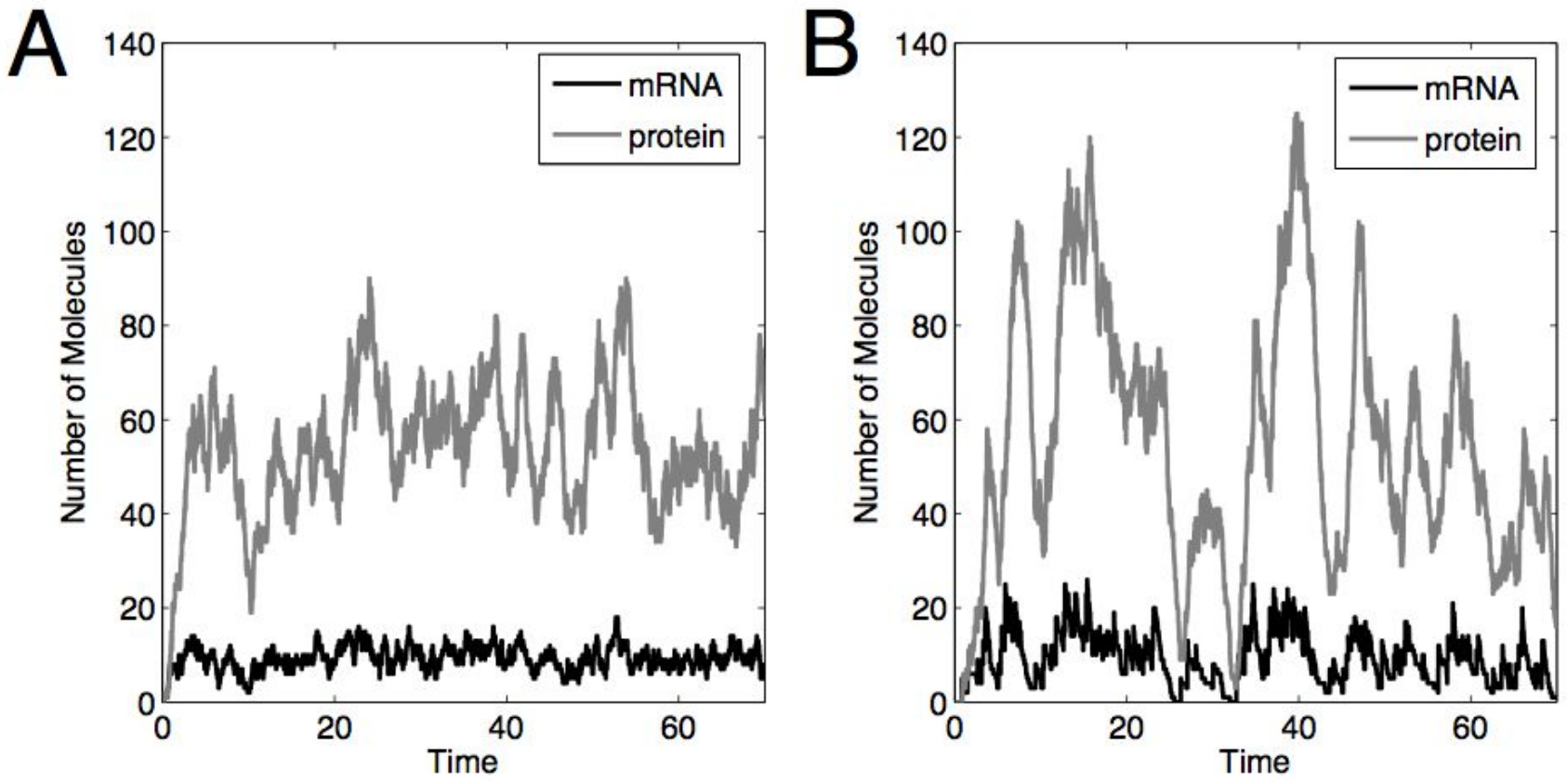
R_4 : (degradation) $\quad P \longrightarrow 0 \quad$ propensity: $\delta_p N_P$

Bursty transcription:

$\tilde{R}_1$: (bursty transcription) $\longrightarrow 5M \quad$ propensity: $\dfrac{k_r}{5}$

ODE system is the same. But stochastically they behave very differently!

Bursty transcription far noisier!
(theory to be developed next)



"standard" transcription

"bursty" transcription