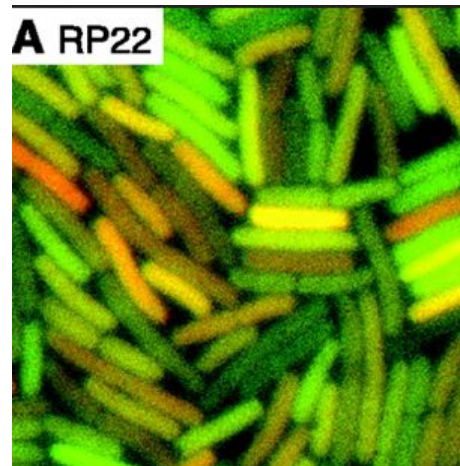


Stochastic modeling of biochemical and genetic networks (adapting notes from my student Brian Ingalls)

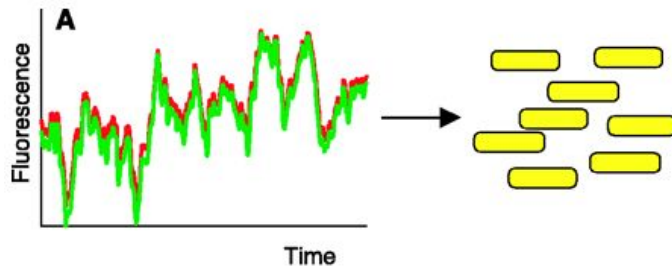
Variability in biology

- **Development:**
genetically identical
but phenotypically
different
- **Gene expression:**
genetically identical
but phenotypically
different

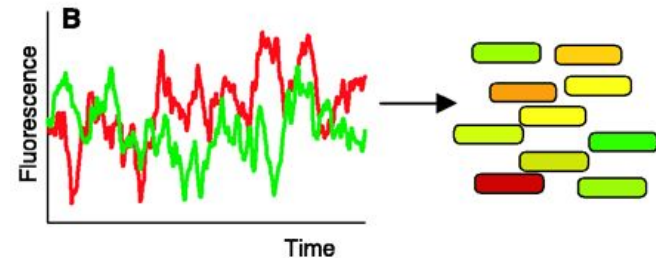


Intrinsic and Extrinsic noise

Elowitz *et. al*, Science, 2002

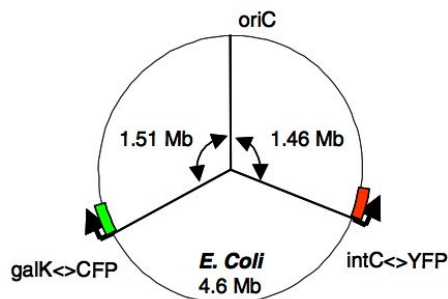


Extrinsic noise

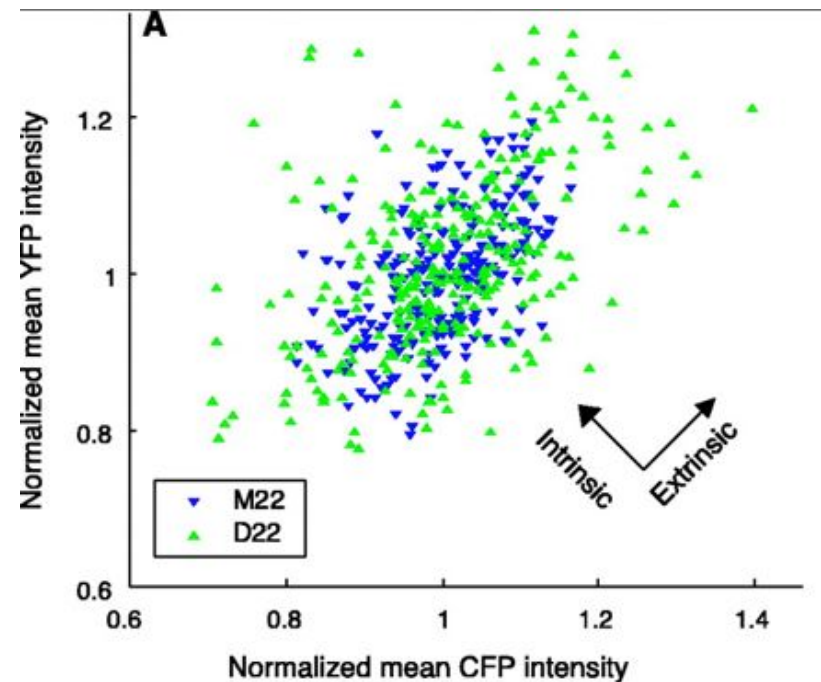


Intrinsic and extrinsic noise

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



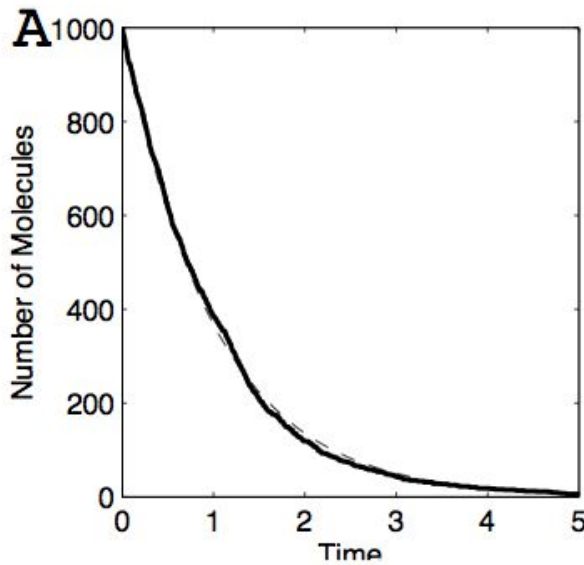
Results



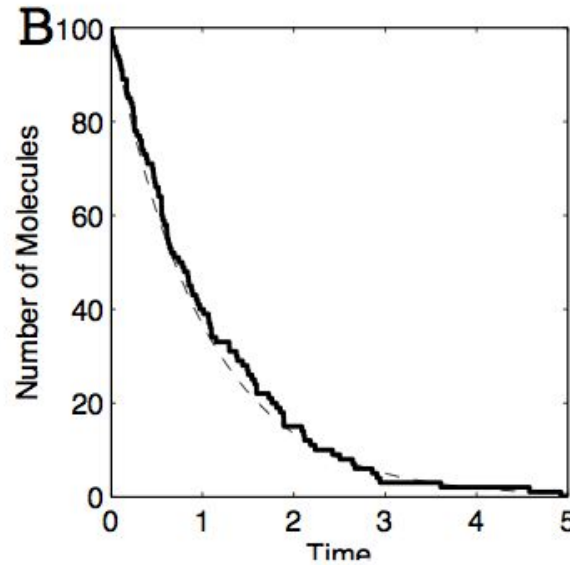
Variability in Molecular Biology

- 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)

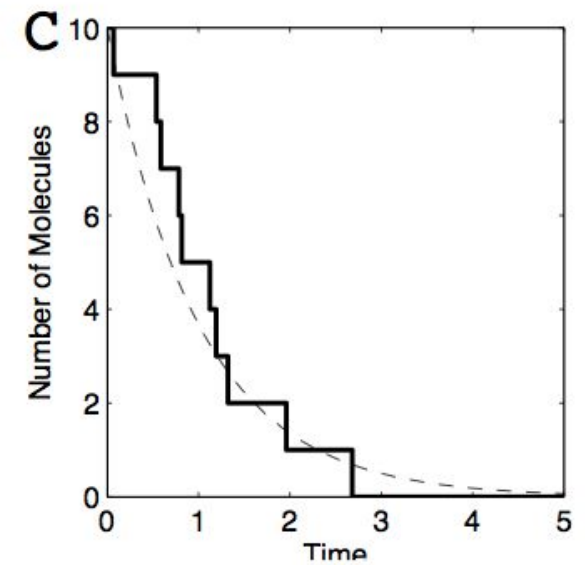
Deterministic behaviour vs. random (stochastic) behaviour



1000 molecules



100 molecules



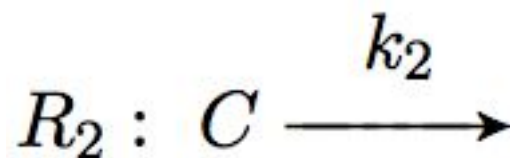
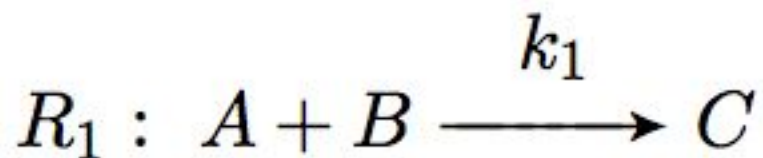
10 molecules

$A \rightarrow$

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

Example



State:

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

Stoichiometry:

$$\mathbf{s}_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \quad \begin{array}{l} \leftarrow A \\ \leftarrow B \\ \leftarrow C \end{array}$$

$$\mathbf{s}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \quad \begin{array}{l} \leftarrow A \\ \leftarrow B \\ \leftarrow C \end{array}$$

Propensity:

$$a_1(\mathbf{N}) = k_1 N_A N_B$$

$$a_2(\mathbf{N}) = k_2 N_C$$

A technical complication

for a homodimer type of reaction such as



we cannot write $k * A^2$,

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

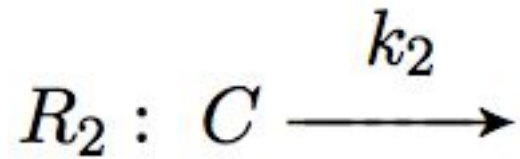
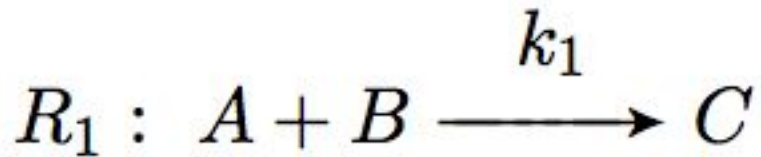
the correct formula has: $k * A (A-1)/2$

i.e. we need to consider all possible pairs

when numbers are large, $A-1 \approx A$ and the 2 can be absorbed;
hence we don't worry about this

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

Example



Stoichiometry:

$$\mathbf{s}_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \begin{array}{l} \leftarrow A \\ \leftarrow B \\ \leftarrow C \end{array}$$

$$\mathbf{s}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \begin{array}{l} \leftarrow A \\ \leftarrow B \\ \leftarrow C \end{array}$$

State:

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

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: stochastic process

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

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

e.g.

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

Stochastic dynamic model

- Assumption: there is a small time-increment dt such that
- at most one reaction can occur during any time interval of length dt
 - probability that reaction R_k occurs in a time interval $[t, t+dt]$ is $a_k(\mathbf{N}(t)) dt$
 - probability that *no reactions occur* during a time interval $[t, t+dt]$ is therefore:

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

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

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

Stochastic dynamic model: probability balance

Example: single species A

State: $N = N_A$

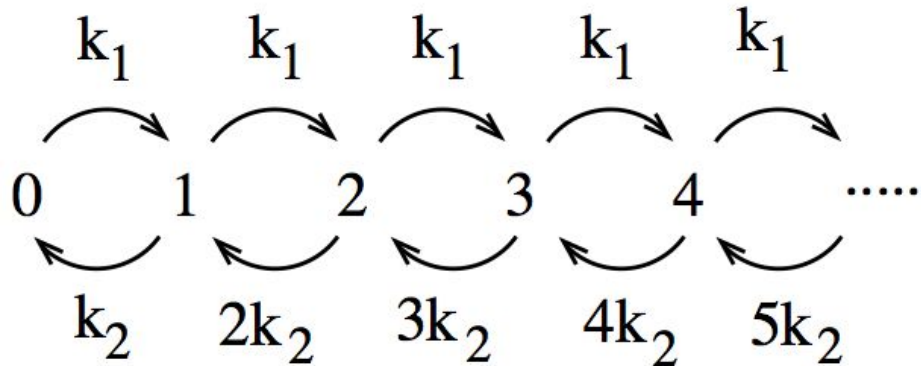
Reactions: $R_1 : \xrightarrow{k_1} A$ $R_2 : A \xrightarrow{k_2}$

– 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:



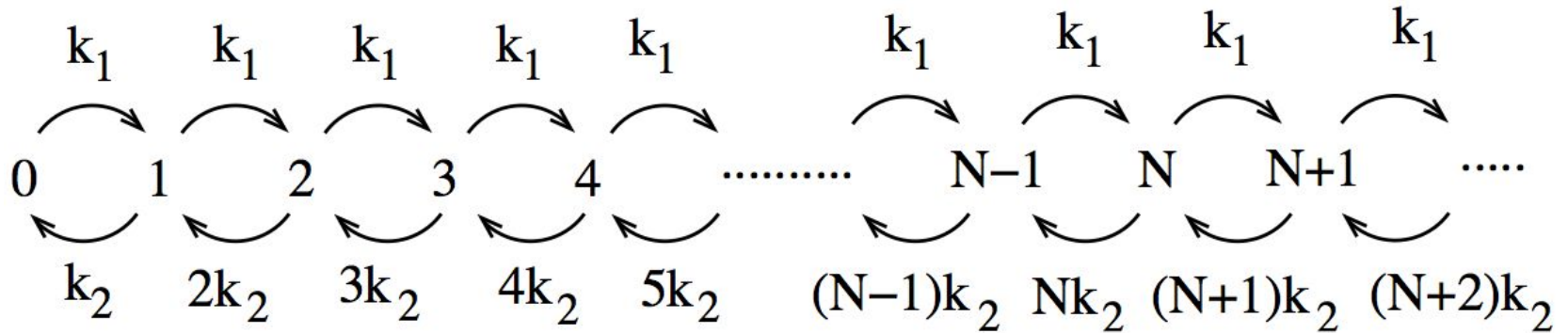
Probability balance:

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

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

Stochastic dynamic model: probability balance

Example: state transitions:



Probability balance:

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

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

$$P(2, t + dt) = P(2, t) [1 - (k_1 + 2k_2) dt] + P(1, t) \cdot k_1 dt + P(3, t) \cdot 3k_2 dt$$

$$P(3, t + dt) = P(3, t) [1 - (k_1 + 3k_2) dt] + P(2, t) \cdot k_1 dt + P(4, t) \cdot 4k_2 dt$$

\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

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:

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

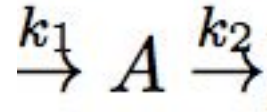
Differential equation for probability distribution:

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

The Chemical Master Equation (CME)

The Chemical Master Equation

Example system:



Master equation:

$$\frac{d}{dt}P(0,t) = -P(0,t)k_1 + P(1,t)k_2$$

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

$$\frac{d}{dt}P(2,t) = -P(2,t)(k_1 + 2k_2) + P(1,t)k_1 + P(3,t)3k_2$$

$$\vdots$$

$$\frac{d}{dt}P(N,t) = -P(N,t)(k_1 + Nk_2) + P(N-1,t)k_1 + P(N+1,t)(N+1)k_2$$

$$\vdots$$

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

Master equation: example

Network:



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

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

Master equation:

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

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

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

Master equation: example

Network:



Master equation:

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

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

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

$$0 = -P^{ss}(2,0)2k_1 + P^{ss}(1,1)k_2$$

Steady state: $0 = -P^{ss}(1,1)k_2 - P^{ss}(1,1)k_1 + P^{ss}(2,0)2k_1 + P^{ss}(0,2)2k_2$

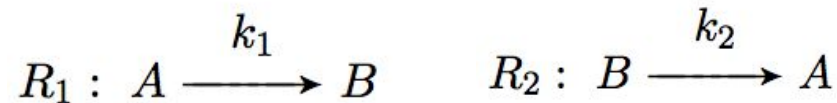
$$0 = -P^{ss}(0,2)2k_2 + P^{ss}(1,1)k_1$$

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

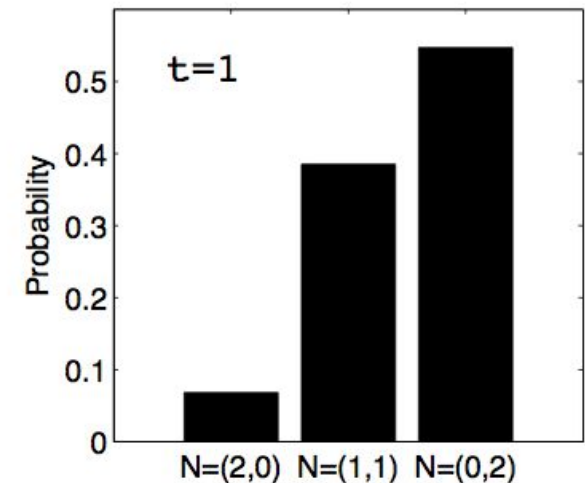
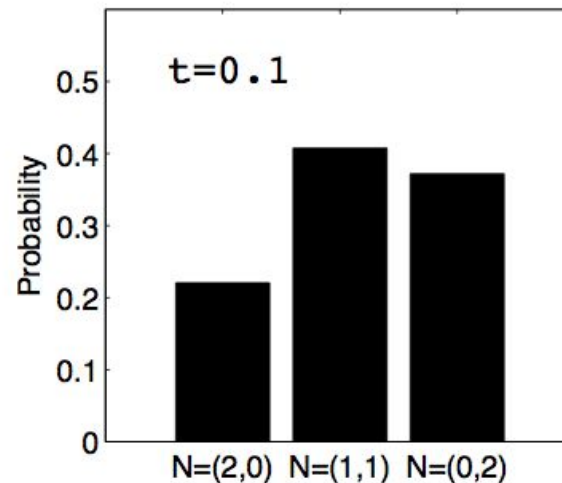
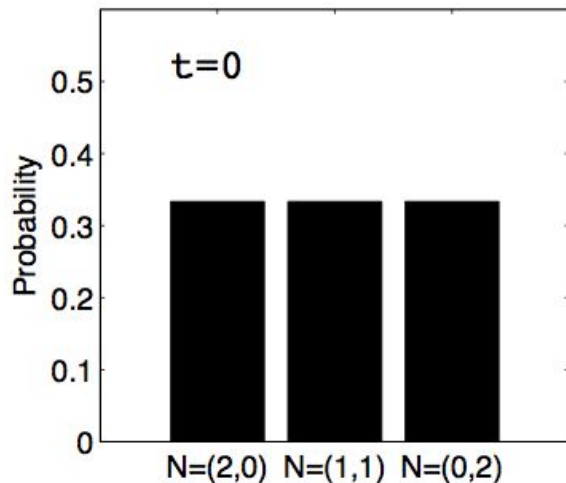
can prove binomial - in fact, can write ss distro in “closed form” for **WR deficiency zero networks!**

Master equation: example

Network:



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



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

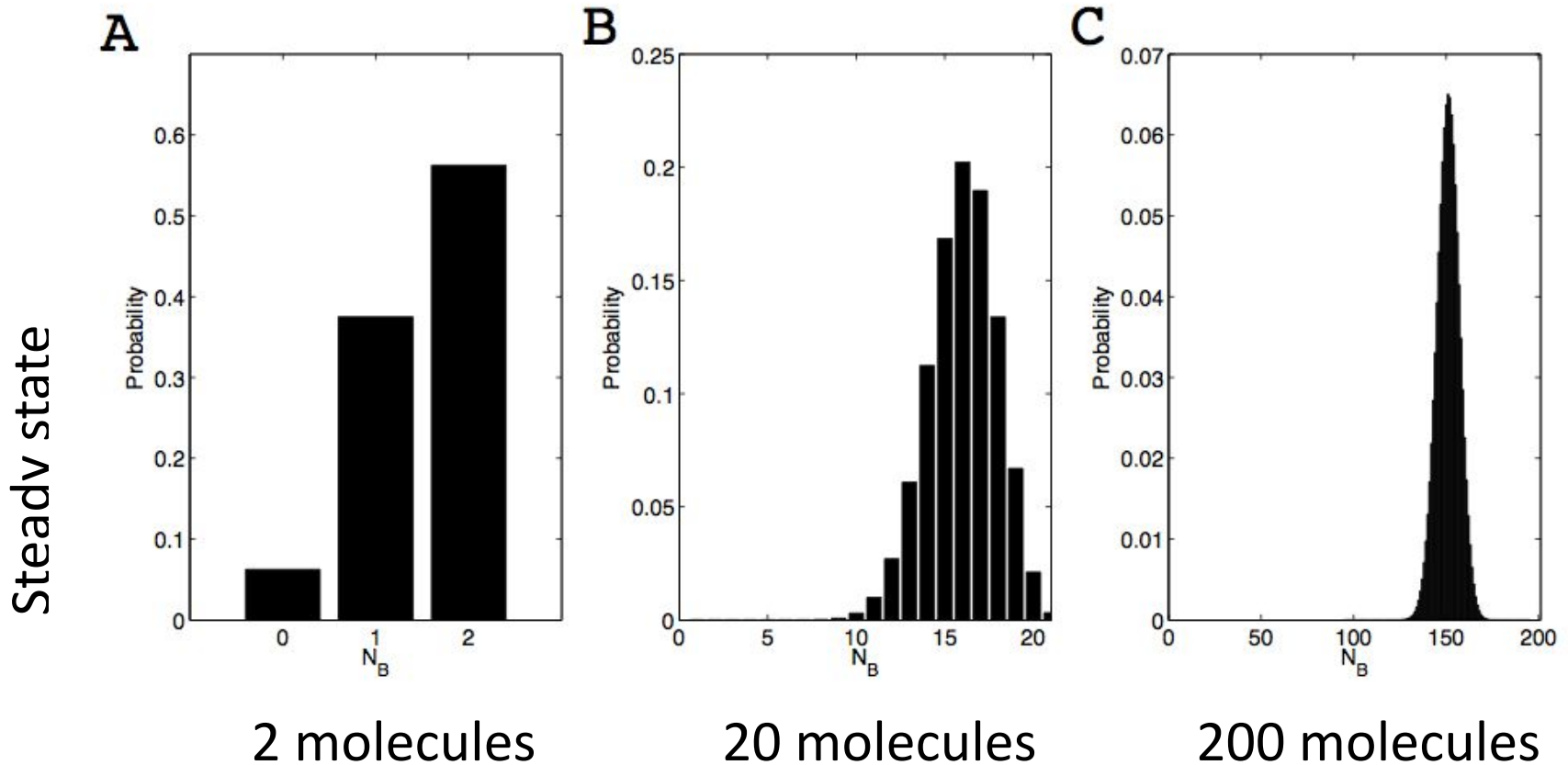
$E(B) = 0 \cdot P^{ss}(2,0) + 1 \cdot P^{ss}(1,1) + 2 \cdot P^{ss}(0,2) = 2 N k_1 / (k_1 + k_2) = \frac{3}{4} * N = 1.5$ when $N = 2$

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:



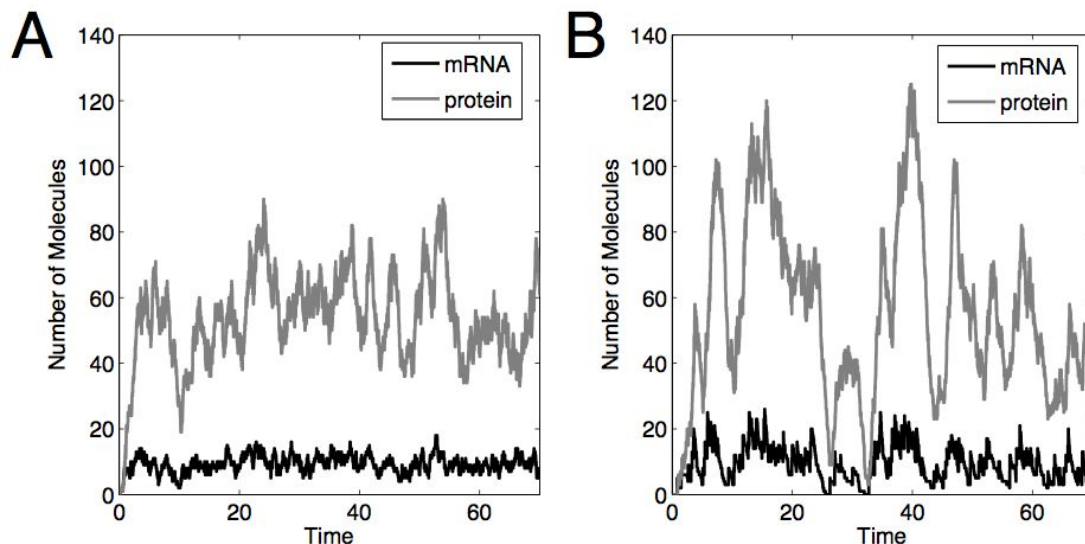
$\frac{3}{4} * N = \frac{3}{4} * 200 = 150$, note how distribution peaks at 150 in case of large # molecules!

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)
 - 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.



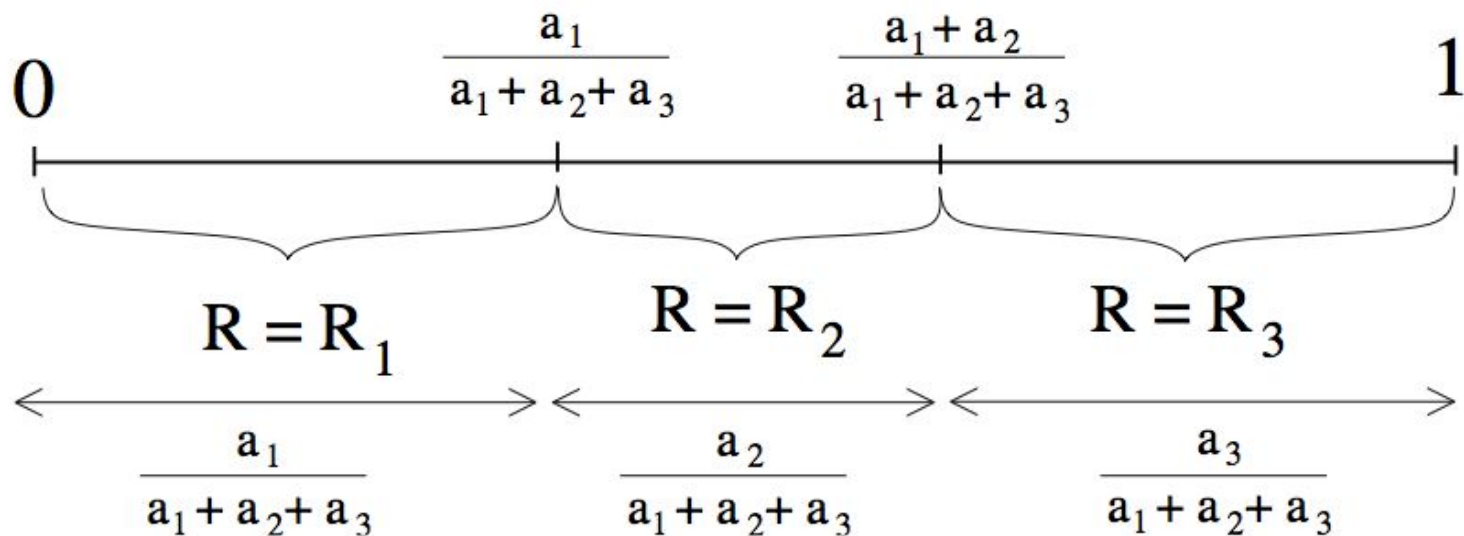
Stochastic Simulation Algorithm

- Depends on the properties of 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

$$P(R = R_1) = \frac{a_1}{a_1 + a_2 + a_3} \quad P(R = R_3) = \frac{a_3}{a_1 + a_2 + a_3}$$
$$P(R = R_2) = \frac{a_2}{a_1 + a_2 + a_3}$$



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 $[0,1]$

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

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

if $\frac{a_1 + a_2}{a_1 + a_2 + a_3} < u \leq \frac{a_1 + a_2 + a_3}{a_1 + a_2 + a_3} = 1$, 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:

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

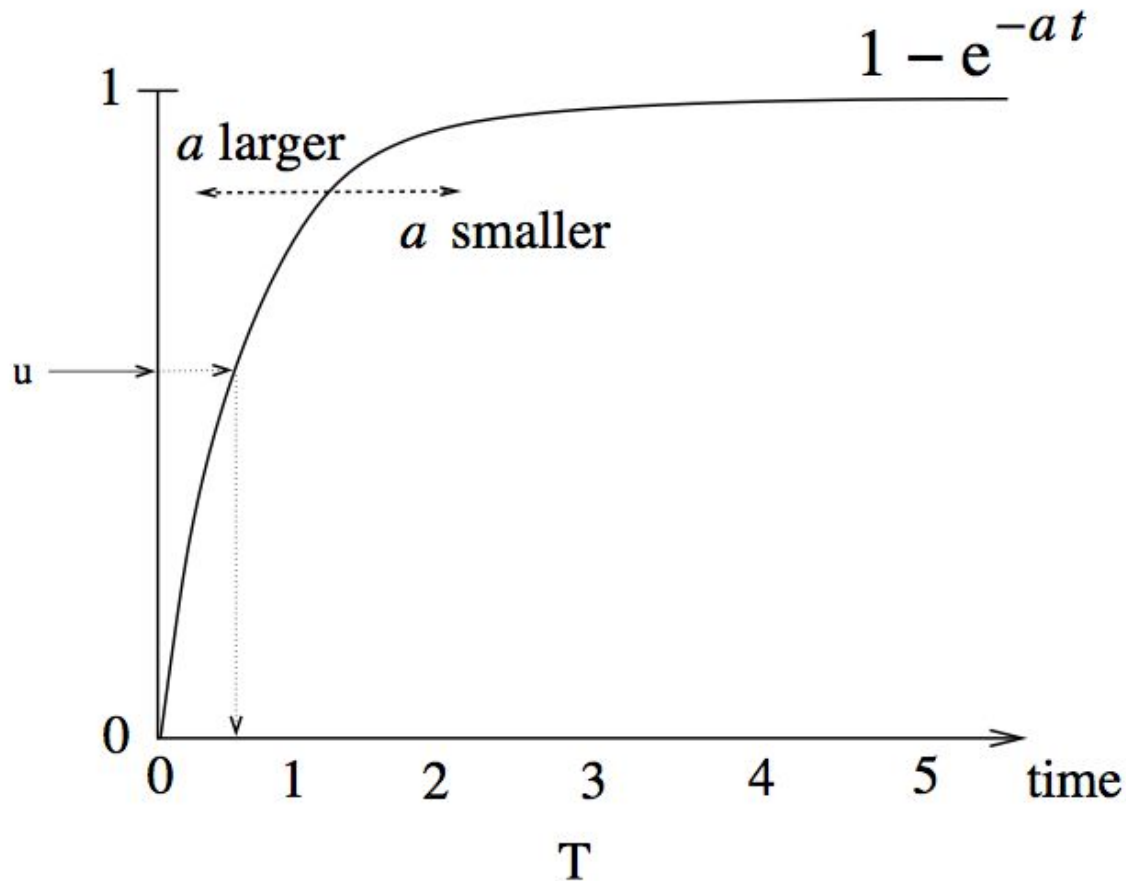
where a is the sum of the propensities:

$$a = a_1 + a_2 + a_3$$

(think of Poisson random arrivals)

Time to the next reaction

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

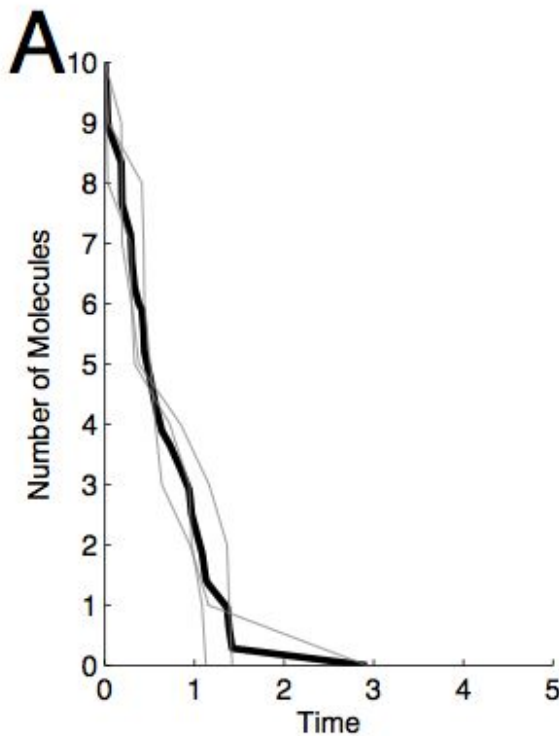
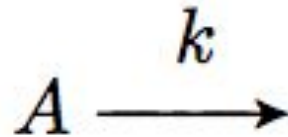


Stochastic Simulation Algorithm

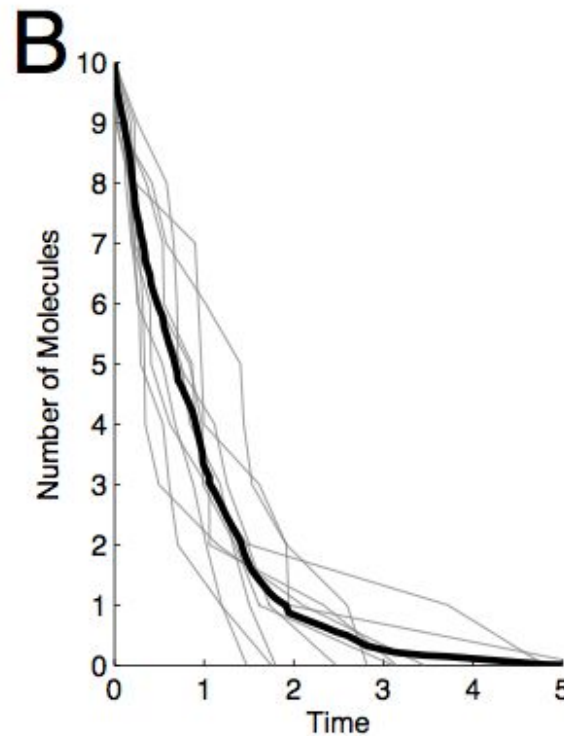
Stochastic Simulation Algorithm (SSA)

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 (Figure 7.43).
4. Draw a sample τ from the random variable T (Figure 7.44).
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.

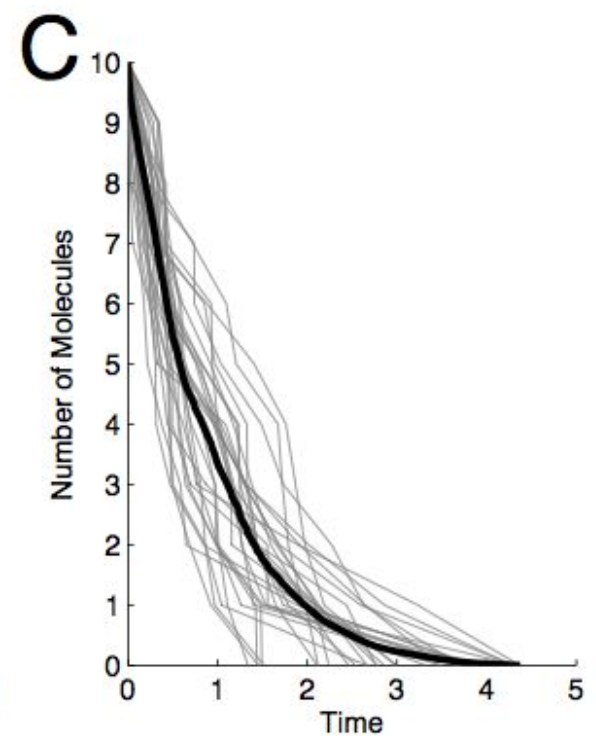
Ensemble average: constitutive decay



3 samples



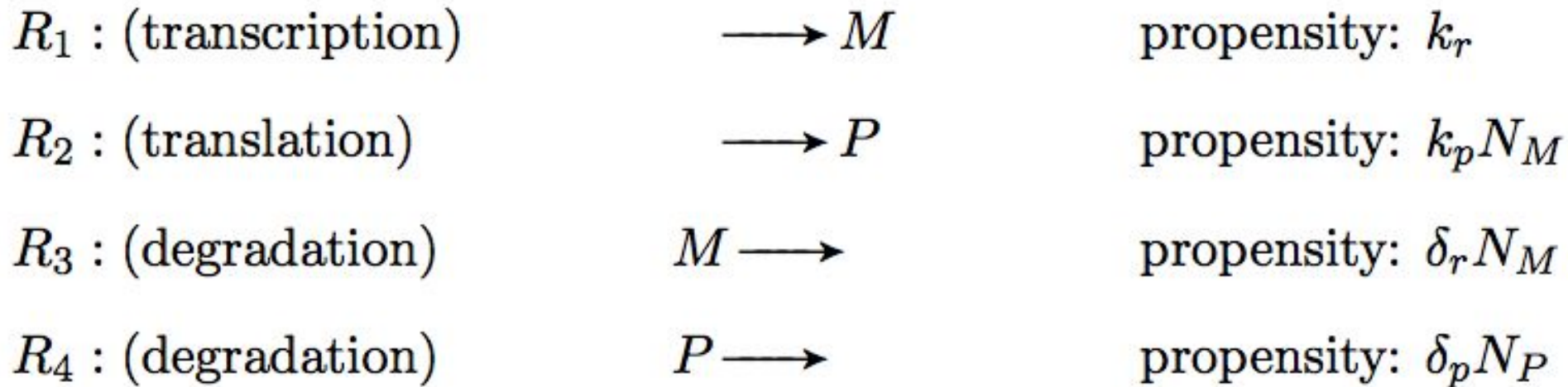
10 samples



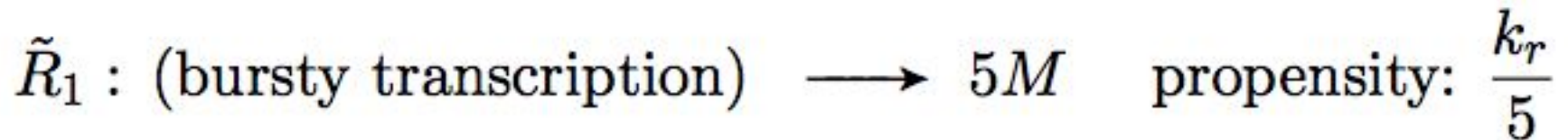
30 samples

Bursty transcription

Gene expression:

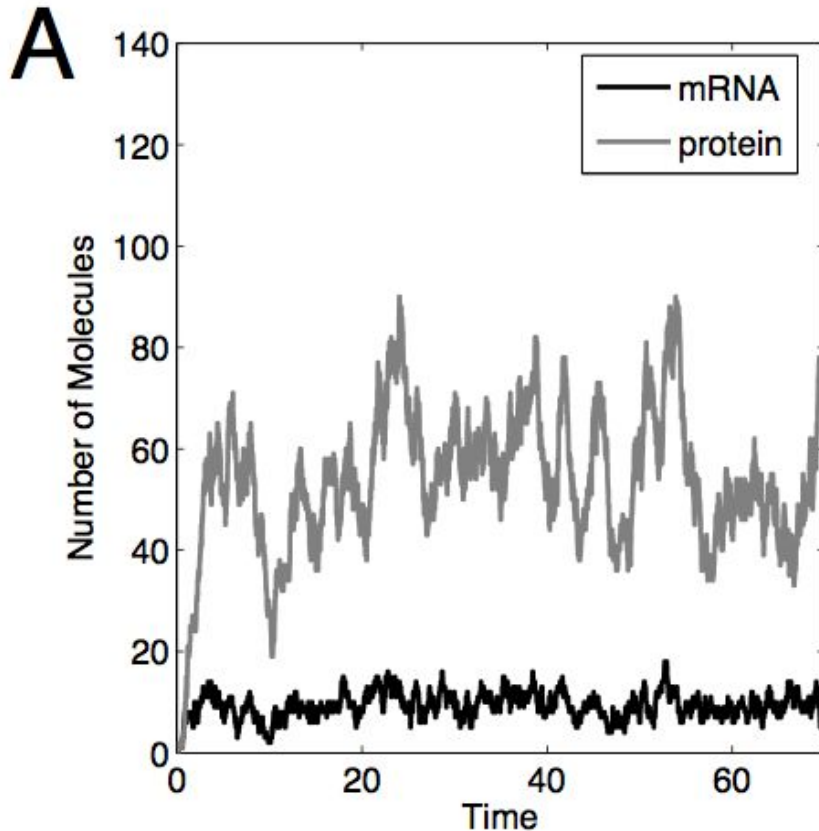


Bursty transcription:

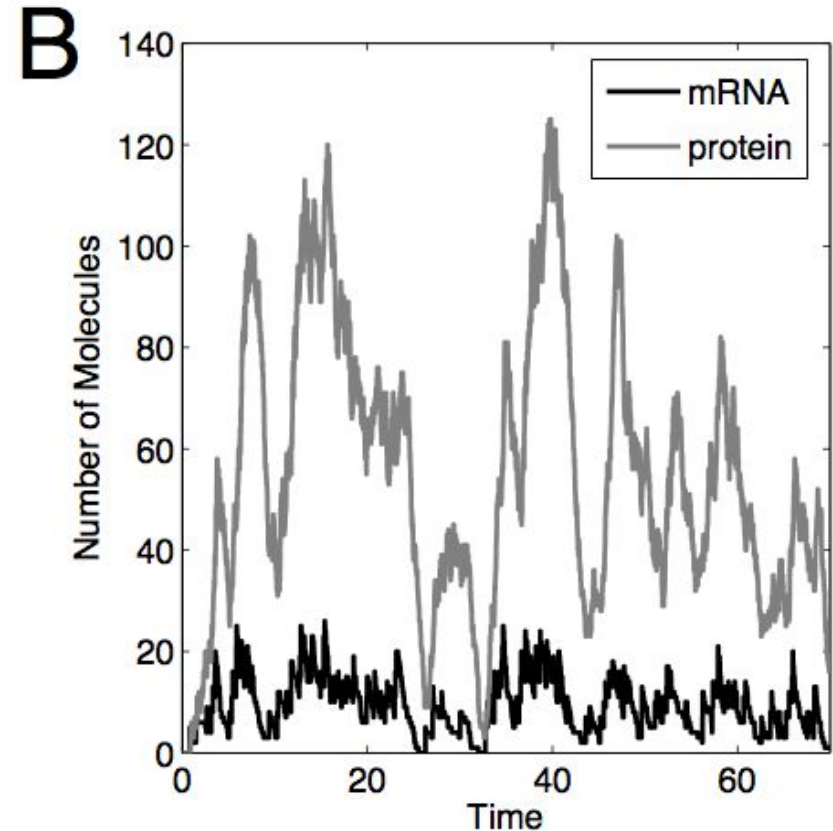


ODE system is the same. But stochastically they behave very differently!

Bursty transcription far noisier!



“standard”
transcription



bursty
transcription