LECTURE 7: NOISE AND STOCHASTIC SIMULATION

Introduction to cellular system modelling

Daniel Georgiev

Summer 2015

OUTLINE

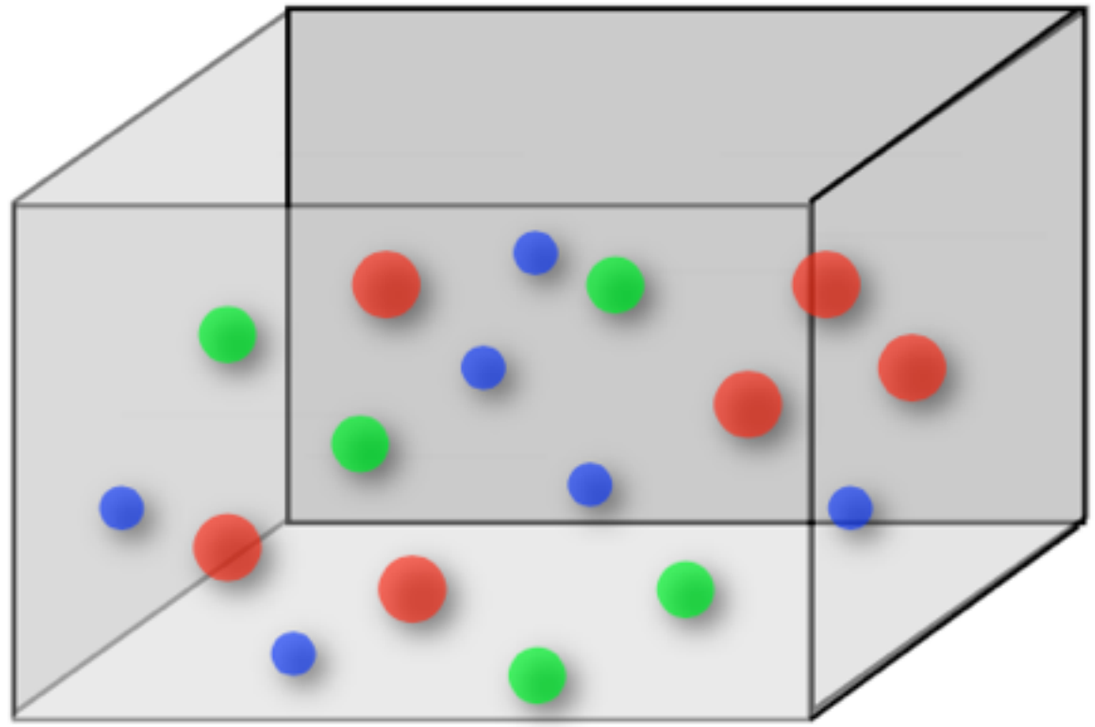
- Markov chain formulation
- RNA production and degradation
- Stationary distribution
- Mean and variance
- Gene expression mean and cv
- Bursting
- Deterministic approximation
- Stochastic simulation algorithms
- First reaction algorithm
- Direct method
- RuleBender
- Factors affecting expression noise
- Intrinsic vs extrinsic noise

RECALL:

Formulation of Stochastic Chemical Kinetics

Gillespie, Physical A, 1992

Reaction volume=Ω



Key Assumptions

(**Well-Mixed**) The probability of finding any molecule in a region $d\Omega$ is given by $\frac{d\Omega}{\Omega}$.

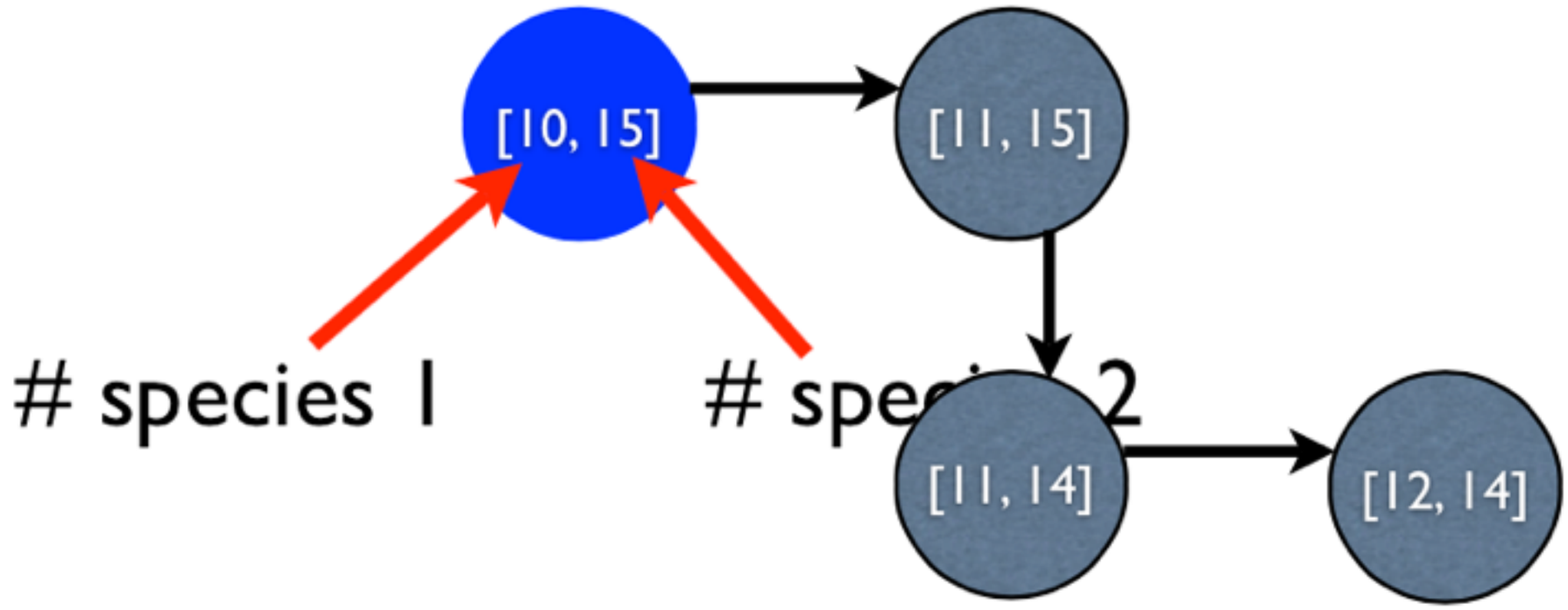
(**Thermal Equilibrium**) The molecules move due to the thermal energy. The reaction volume is at a constant temperature T. The velocity of a molecule is determined according to a Boltzman distribution:

$$f_{v_x}(v) = f_{v_y}(v) = f_{v_z}(v) = \sqrt{\frac{m}{2\pi k_B T}}\, e^{-\frac{m}{2k_B T}v^2}$$

SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

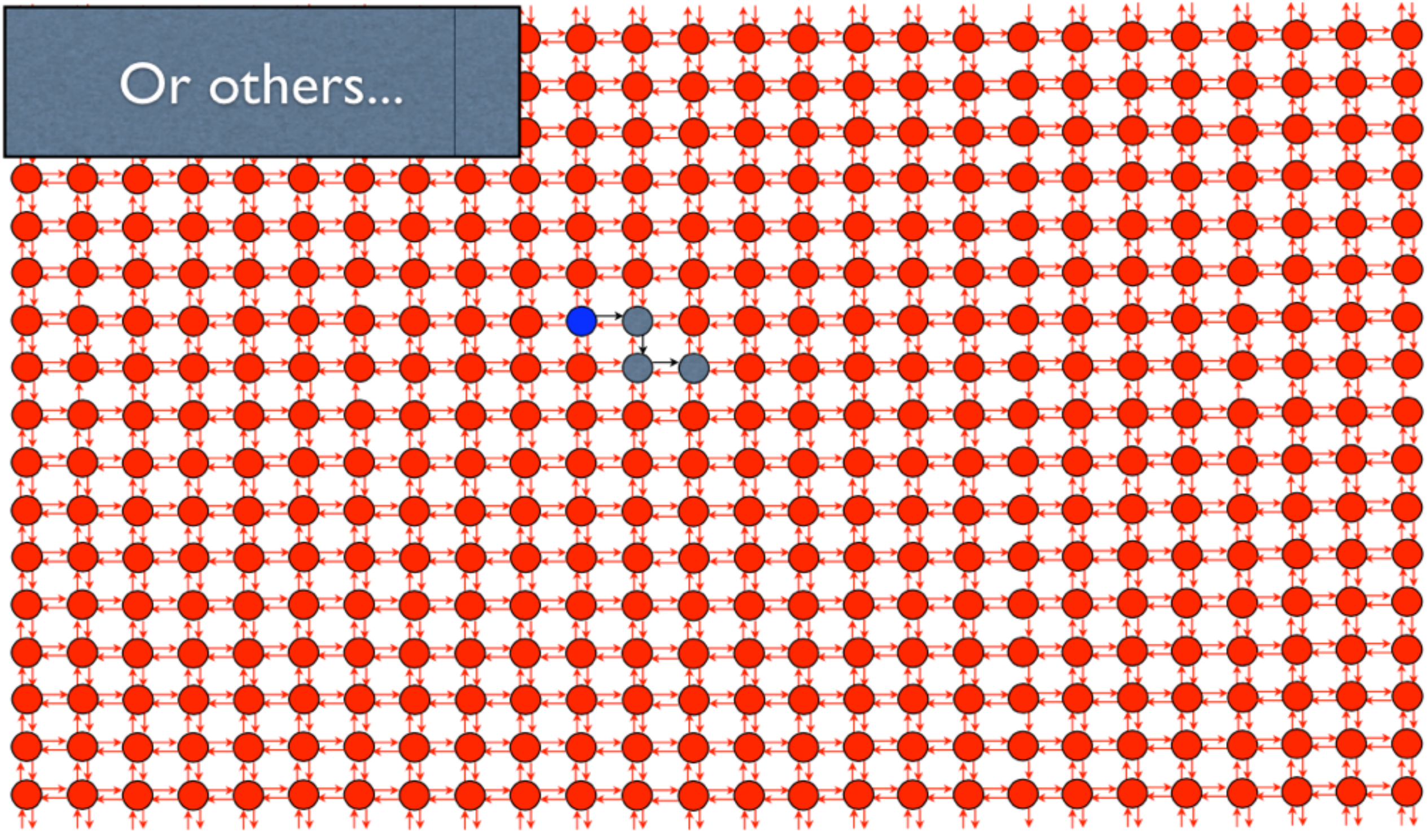
MARKOV CHAIN DESCRIPTION OF CHEMICAL KINETICS

- At any time, the state of the system is *defined* by its integer population vector: $\mathbf{x} \in \mathbb{Z}^N$
- Reactions are transitions from one state to another:



SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

COMBINATORIALLY COMPOSED STATE SPACE



SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

THE REACTION TIMES HAVE AN EXPONENTIAL DISTRIBUTION

Probability reaction will occur in $[t, t+\Delta t)$: $w\Delta t + \mathcal{O}(\Delta t)^2$

Probability reaction *will not occur* in $[t, t+\Delta t)$ $1 - w\Delta t + \mathcal{O}(\Delta t)^2$

Probability a reaction *will not occur* in two such time intervals $[t, t+2\Delta t)$: $\left(1 - w\Delta t + \mathcal{O}(\Delta t)^2\right)^2 = 1 - 2w\Delta t + \mathcal{O}(\Delta t)^2$

Suppose that $\tau = K\Delta t$, then the probability that *no reaction will occur* in the interval $[t, t+\tau)$ is

$$\left(1 - w\frac{\tau}{K} + \mathcal{O}(K^{-2})\right)^K$$

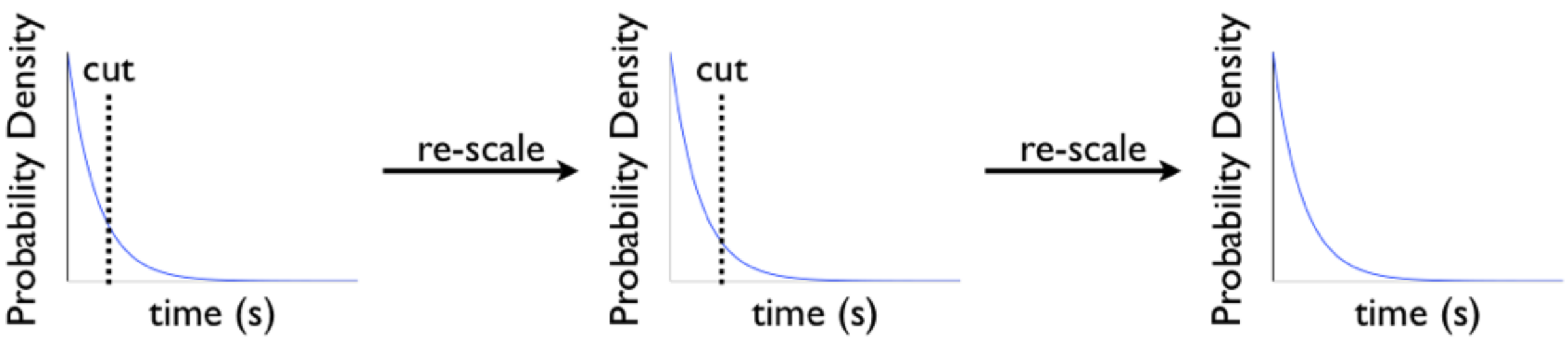
Taking the limit as K goes to infinity yields that the probability that *no reaction will occur* in the interval $[t, t+\tau)$ is

$$\lim_{k\to\infty} \left(1 - w\frac{\tau}{K} + \mathcal{O}(K^{-2})\right)^K = \exp(-w\tau)$$

SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

EXPONENTIAL DISTRIBUTION IS FORGETFUL

- We have assumed that the system is fully described by the population vectors.
- If no reaction occurs, then nothing will have changed.
- Waiting times must be *memoryless* random variables.



- No matter where we cut and scale the distribution, it must always looks the same.

The exponential is the *only*
continuous r.v. with this property.

SLIDE CONTRIBUTED BY MUSTAFA KHAMMASH, ETH

UNDERLYING CHEMICAL MASTER EQUATION

$$P(n, t+dt) = P(n-1, t) \cdot kdt$$ Prob.$\{N(t) = n-1$ and mRNA created in [t,t+dt)$\}$

$$+\ P(n+1, t) \cdot (n+1)\gamma dt$$ Prob.$\{N(t) = n+1$ and mRNA degraded in [t,t+dt)$\}$

$$+\ P(n, t) \cdot (1 - kdt)(1 - n\gamma dt)$$ Prob.$\{N(t) = n$ and mRNA not created nor degraded in [t,t+dt)$\}$

$$P(n, t+dt) - P(n, t) = P(n-1, t)kdt + P(n+1, t)(n+1)\gamma dt - P(n, t)(k + n\gamma)dt + O(dt^2)$$

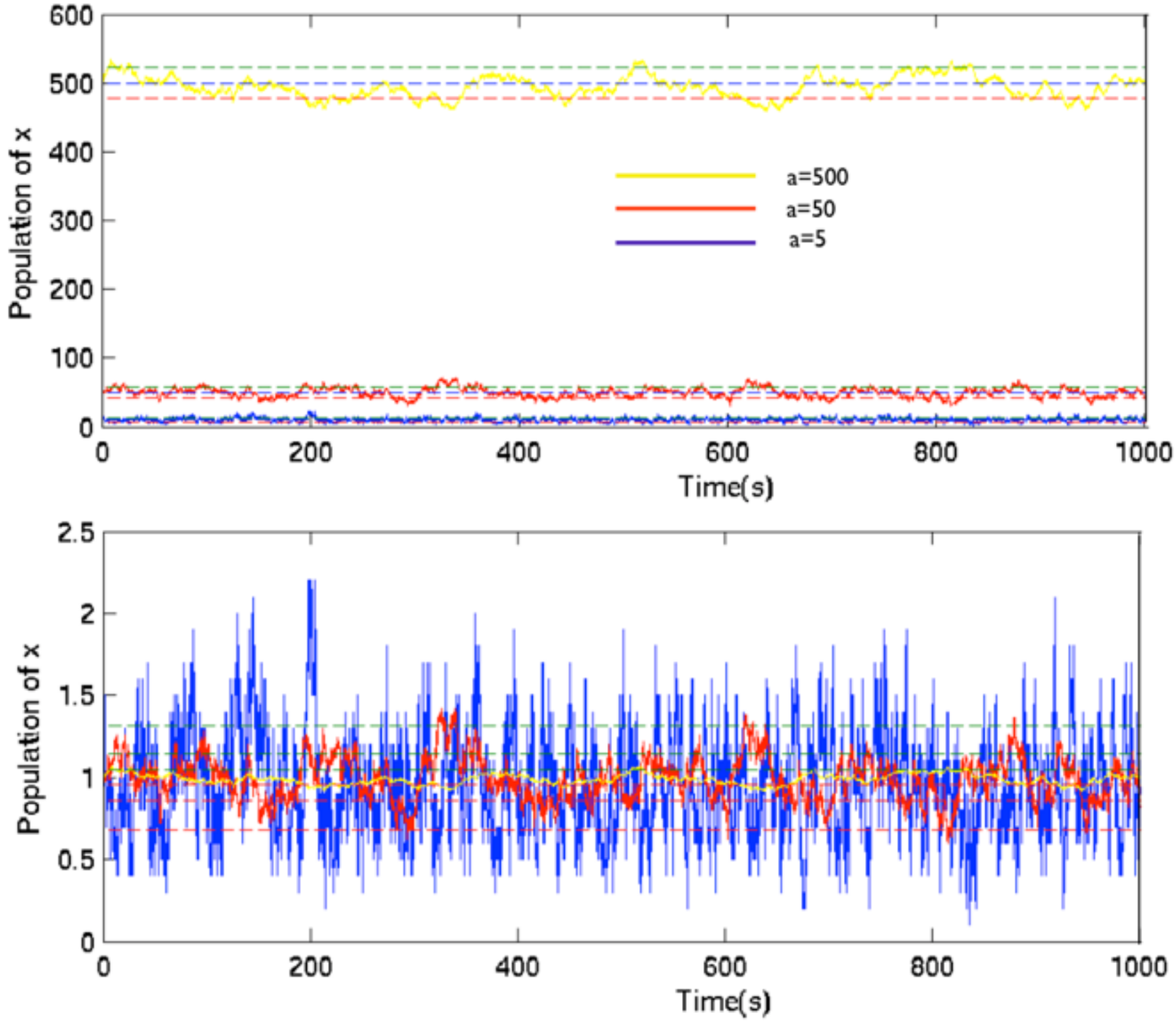
Dividing by dt and taking the limit as $dt \to 0$

The Chemical Master Equation

$$\frac{d}{dt}P(n, t) = kP(n-1, t) + (n+1)\gamma P(n+1, t) - (k + n\gamma)P(n, t)$$

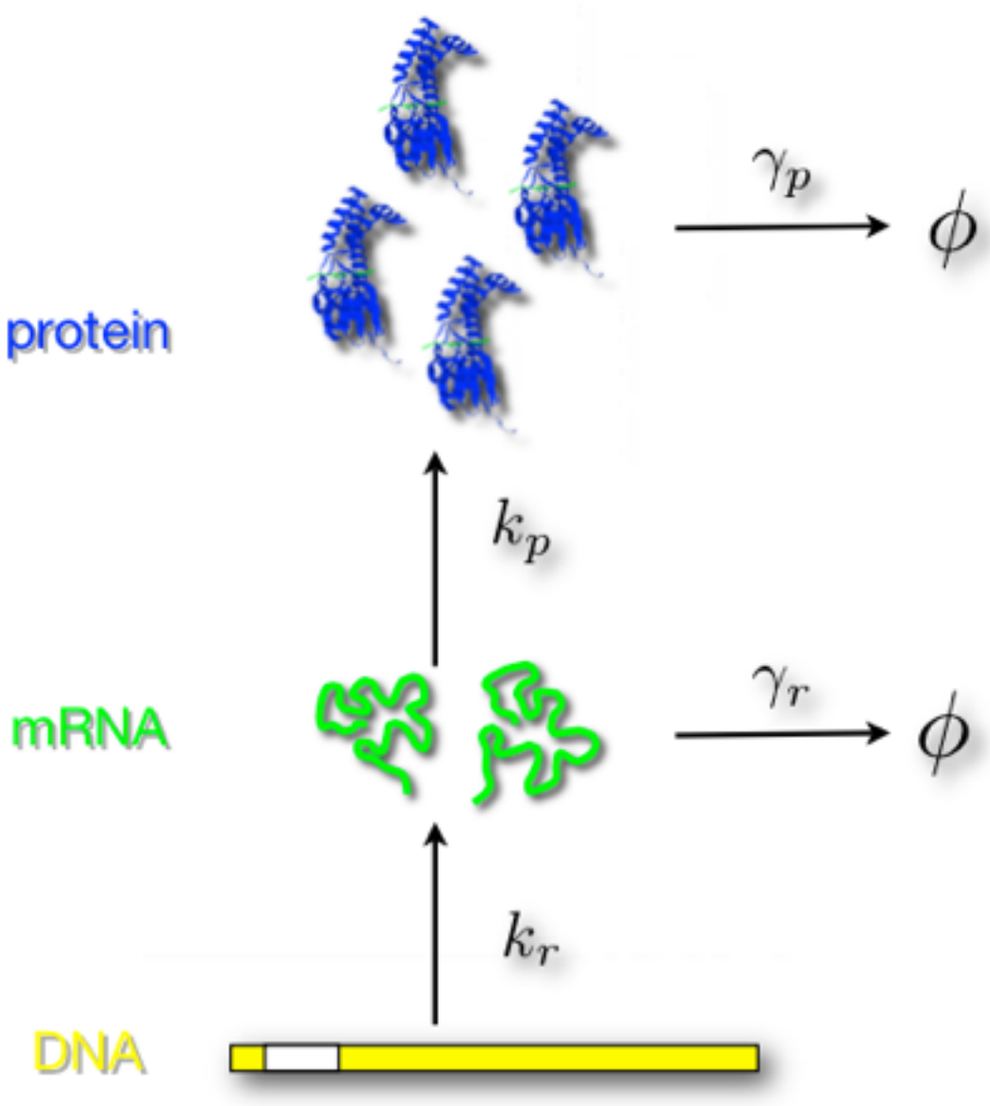
SLIDE CONTRIBUTED BY MUSTAFA KHAMMASH, ETH

STEADY STATE BEHAVIOUR FOR DIFFERENT MEANS



SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

TRANSCRIPTION + TRANSLATION



Reactants

$X_1(t)$ is # of mRNA; $X_2(t)$ is # of protein

Reactions

$$R_1 : \phi \xrightarrow{k_r} mRNA$$

$$R_2 : mRNA \xrightarrow{\gamma_r} \phi$$

$$R_3 : mRNA \xrightarrow{k_p} protein + mRNA$$

$$R_4 : protein \xrightarrow{\gamma_p} \phi$$

Stoichiometry and Propensity

$$S = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

$$w(X) = \begin{bmatrix} k_r \\ \gamma_r X_1 \\ k_p X_1 \\ \gamma_p X_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 \\ \gamma_r & 0 \\ k_p & 0 \\ 0 & \gamma_p \end{bmatrix}}_{W} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \underbrace{\begin{bmatrix} k_r \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{w_0}$$

SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

Steady-State Moments

$$A = SW = \begin{bmatrix} -\gamma_r & 0 \\ k_p & -\gamma_p \end{bmatrix}, \qquad Sw_0 = \begin{bmatrix} k_r \\ 0 \end{bmatrix}$$

$$\bar{X} = -A^{-1} Sw_0 = \begin{bmatrix} \frac{k_r}{\gamma_r} \\ \\ \frac{k_p k_r}{\gamma_p \gamma_r} \end{bmatrix}$$

Steady-State Covariance

$$BB^T = S\ diag(W\bar{X} + w_0)S^T = \begin{bmatrix} 2k_r & 0 \\ 0 & \frac{2k_p k_r}{\gamma_r} \end{bmatrix}$$

The steady-state covariances equation

$$A\bar{\Sigma} + \bar{\Sigma}A^T + BB^T = 0 \qquad \text{Lyapunov Equation}$$

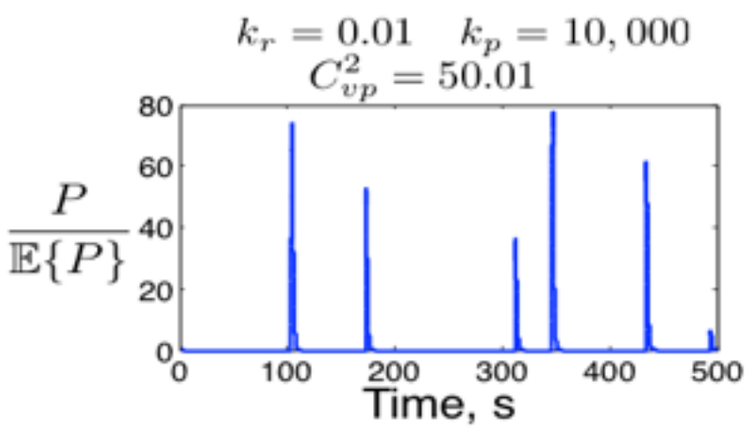
can be solved algebraically for $\bar{\Sigma}$.

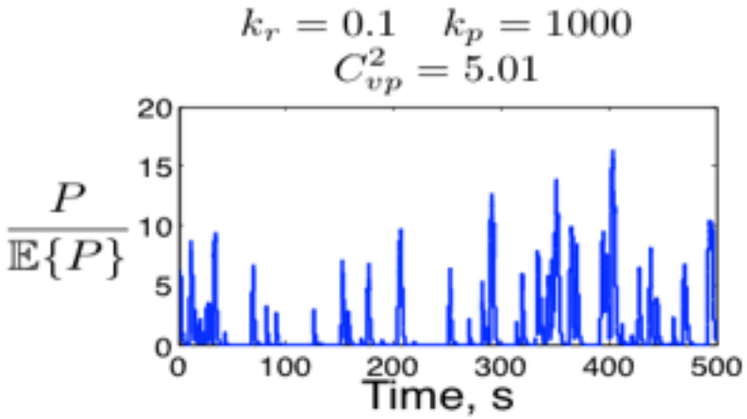
$$\bar{\Sigma} = \begin{bmatrix} \frac{k_r}{\gamma_r} & \frac{k_p k_r}{\gamma_r(\gamma_r+\gamma_p)} \\ \\ \frac{k_p k_r}{\gamma_r(\gamma_r+\gamma_p)} & \frac{k_p k_r}{\gamma_p \gamma_r}(1 + \frac{k_p}{\gamma_r+\gamma_p}) \end{bmatrix}$$

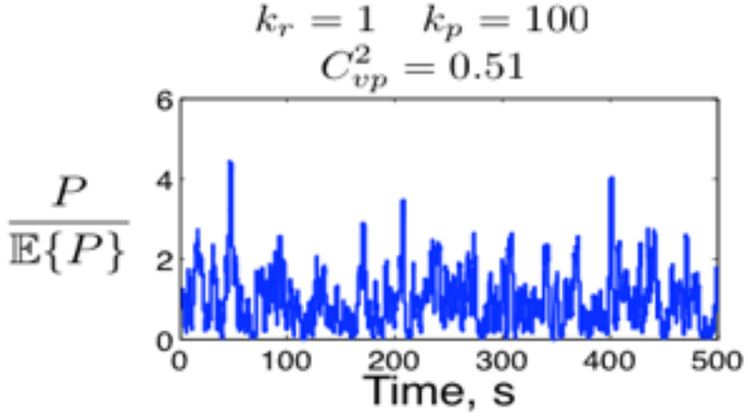
SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

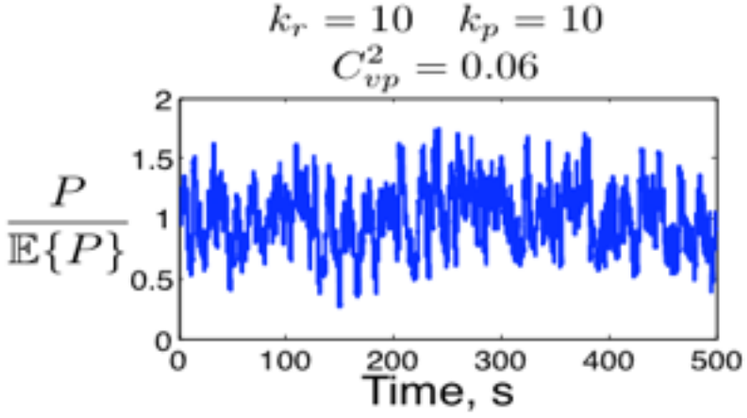
LARGE MEAN ~= SMALL FLUCTUATIONS (BURSTING)

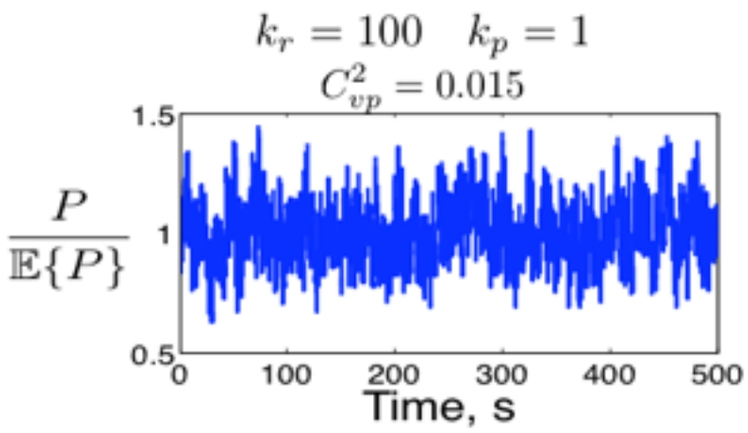
$$\mathbb{E}\{P\} = 100, \quad \gamma_r = \gamma_p = 1$$

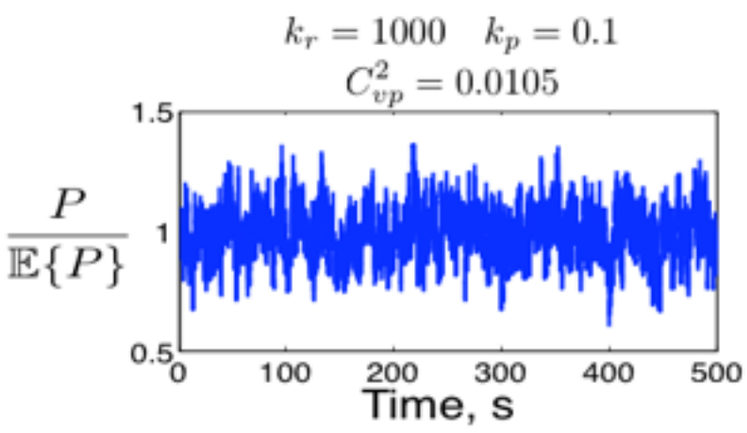












SLIDE CONTRIBUTED BY MUSTAFA KHAMMASH, ETH

STOCHASTIC AND DETERMINISTIC MODELS

Define $X^{\Omega}(t) = \frac{X(t)}{\Omega}$.

Question: How does $X^{\Omega}(t)$ relate to $\Phi(t)$?

Fact: Let $\Phi(t)$ be the deterministic solution to the reaction rate equations

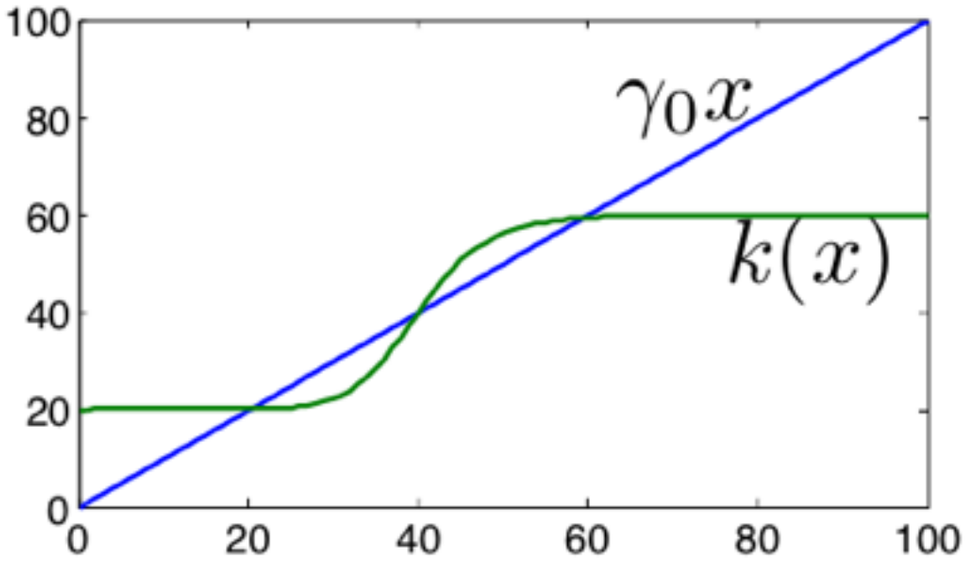
$$\frac{d\Phi}{dt} = Sf(\Phi), \;\; \Phi(0) = \Phi_0.$$

Let $X^{\Omega}(t)$ be the stochastic representation of the same chemical systems with $X^{\Omega}(0) = \Phi_0$. Then for every $t \geq 0$:

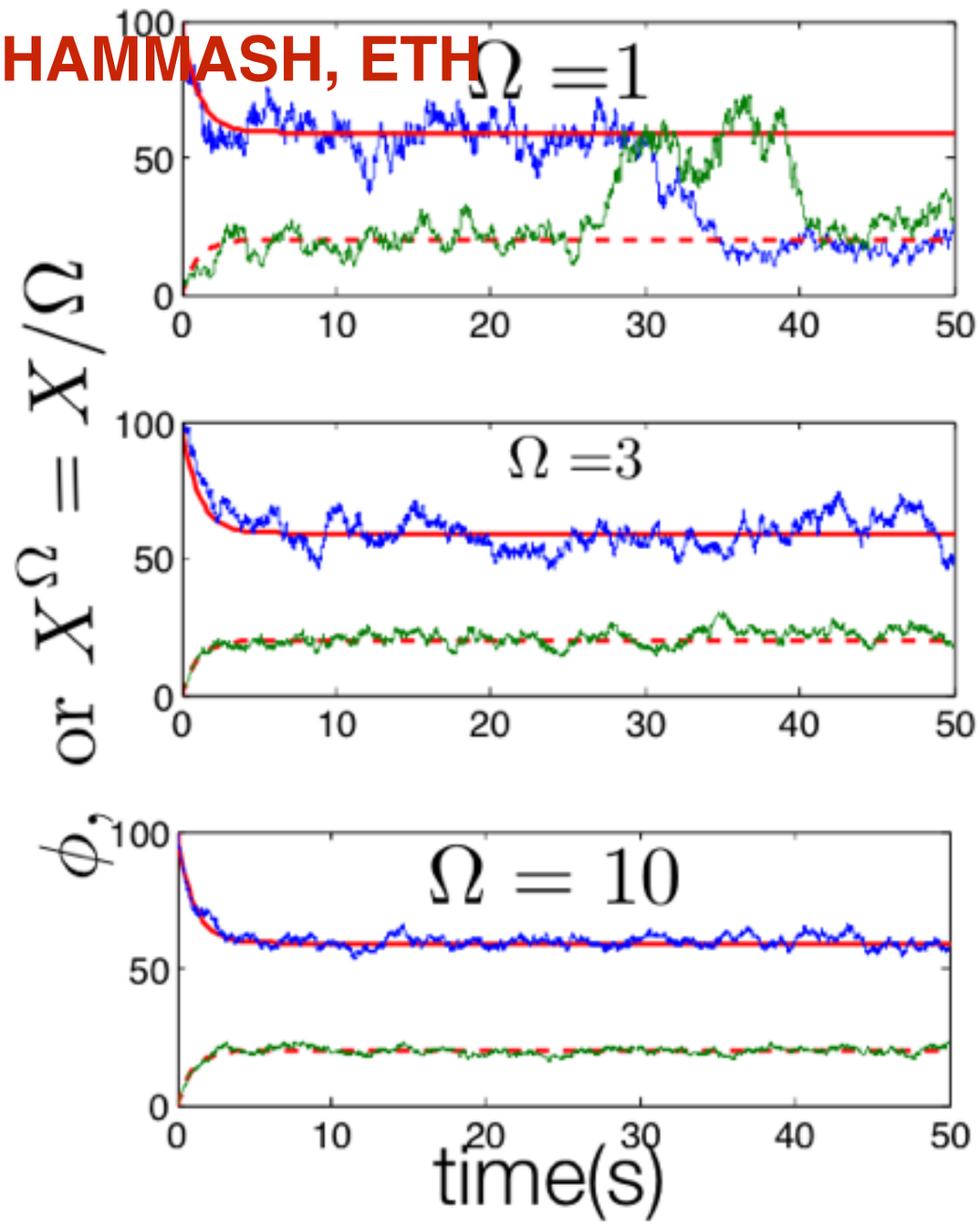
$$\lim_{\Omega \to \infty} \sup_{s \leq t} \left| X^{\Omega}(s) - \Phi(s) \right| = 0 \;\; a.s.$$

SLIDE CONTRIBUTED BY MUSTAFA KHAMMASH, ETH

x produced with rate $k(x)$
and degraded with rate $\gamma_0 x$.



ϕ, or $X^{\Omega} = X/\Omega$



$$w_1(\phi) = \gamma_0 \phi$$

$$w_2(\phi) = \left(20 + 40\frac{\phi^{10}}{40^{10} + \phi^{10}}\right)$$

Deterministic

$$w_1(X) = \Omega\gamma_0 X/\Omega = \gamma_0 X$$

$$w_2(X) = \Omega\left(20 + 40\frac{(X/\Omega)^{10}}{40^{10} + (X/\Omega)^{10}}\right)$$

Stochastic

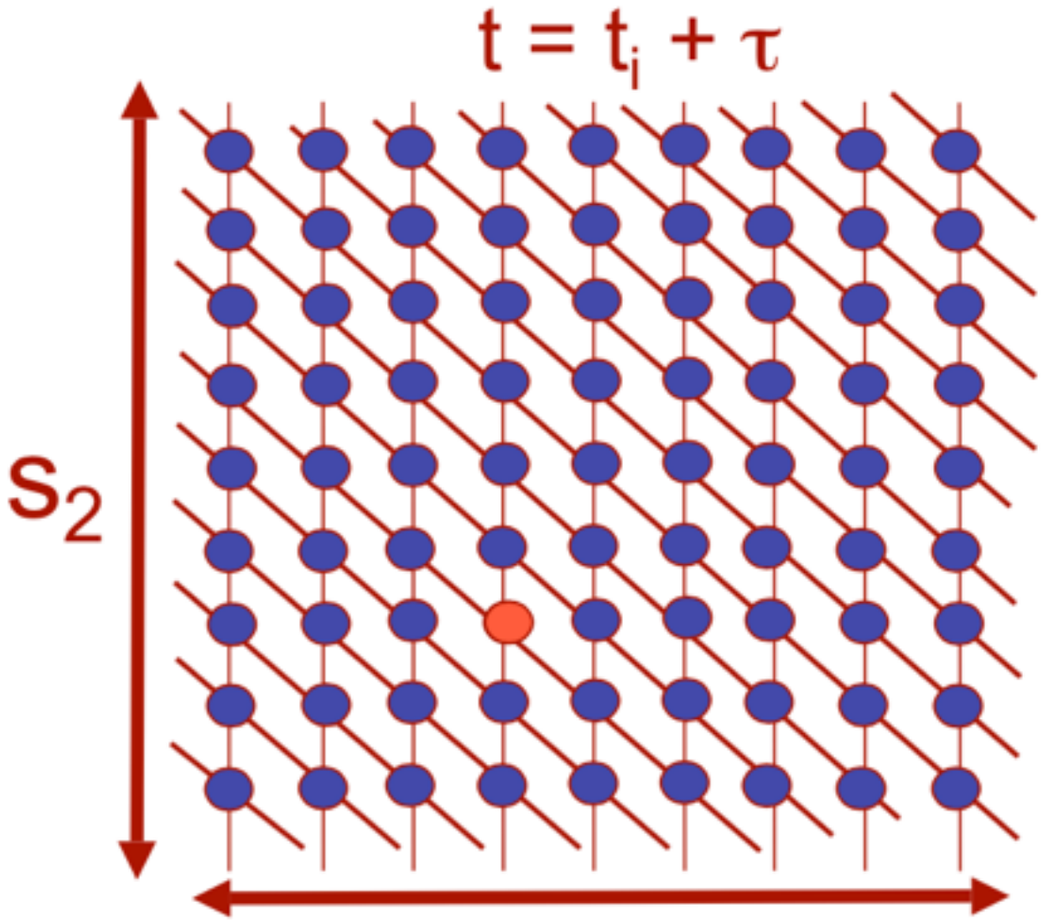
SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

MONTE CARLO SIMULATION OF MARKOV CHAINS

- Stochastic Simulation Algorithm
  - D.T. Gillespie, J. Phys. Chem. A **81**, 2340 (1977)
  - M. Gibson and J. Bruck, J. Phys. Chem. **104**, 1876 (2000)

SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

FIRST REACTION METHOD



Step 1. Generate the time of the next reaction of each type.

The time until the next reaction is a random variable of exponential distribution:

$$P_{\tau_\mu}(t) = w_\mu(\mathbf{x})\mathrm{e}^{-w_\mu(\mathbf{x})t}$$

To generate each next reaction time, generate r_1 from a uniform distribution on (0,1) and use the equation:

$$\tau_\mu = \frac{1}{w_\mu(\mathbf{x})}\log\frac{1}{r_\mu}$$

Step 2. Decide which reaction has occurred.

This is simply the reaction with the smallest τ_μ:

$$k = \arg\left\{\min_{\mu\in\{0,\ldots,M\}} \tau_\mu\right\}$$

Step 3. Update current Time (t=t+τ_k) and State (**x** = **x**+$\mathbf{s}_\mathbf{k}$).

In the FRM each reaction requires M rv's.

SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

DISTRIBUTION OF MINIMUM OF EXPONENTIAL RVs

Let $\{\tau_1, \tau_2, \ldots, \tau_M\}$ be a set of exponentially distributed random variables: $\tau_\mu \in \mathrm{EXP}\,(w_\mu)$

The minimum of $\{\tau_\mu\}$ is an exponentially distributed random variable given by:

$$\min_{\mu \in \{0,\ldots,M\}} \tau_\mu \in \mathrm{EXP}\,(|\mathbf{w}|_1)$$

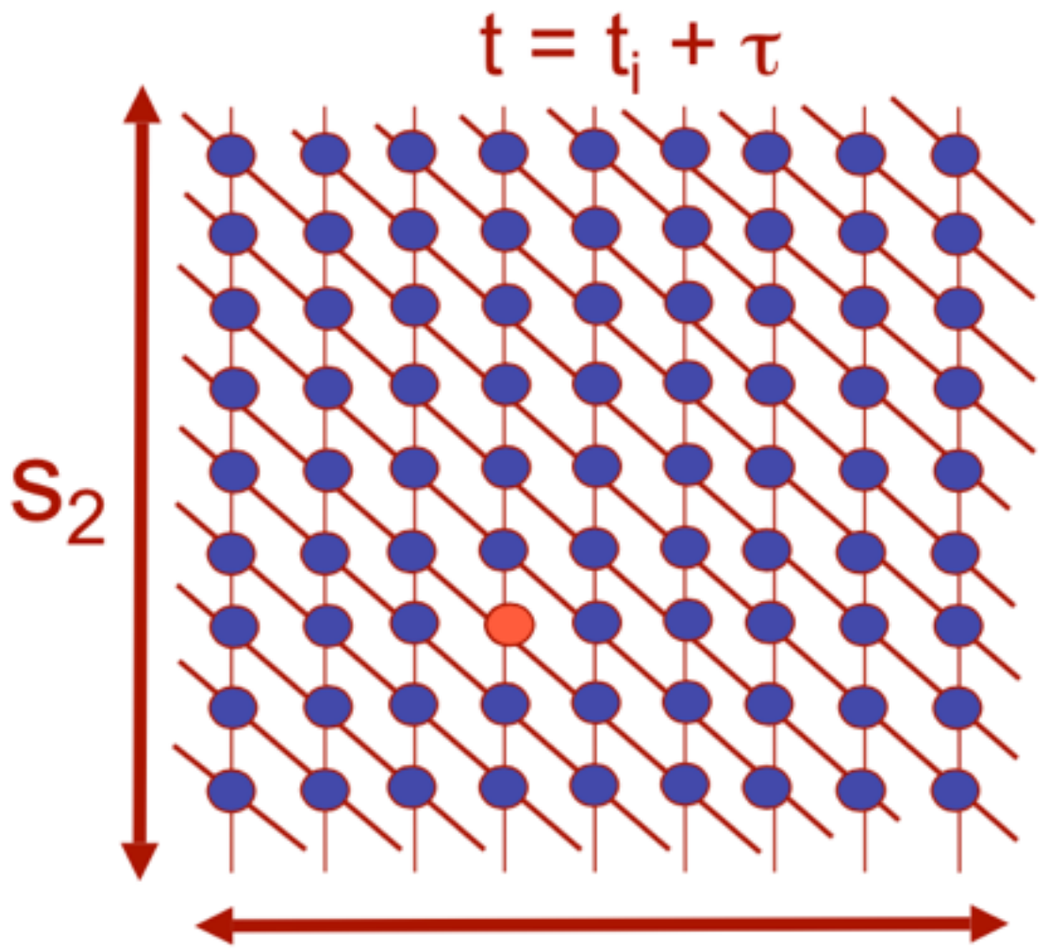
The argument, *k*, of this distribution is also a random variable with distribution:

$$P(k = \mu) = \frac{w_\mu}{|\mathbf{w}|_1}$$

In the DM we only generate 2 rv's.

SLIDE CONTRIBUTED BY BRIAN MUNSKY, UC Boulder

DIRECT METHOD (Gillespie Algorithm)



Step 1. Generate the time of the next reaction.

The time until the next reaction is a random variable of exponential distribution:

$$P_\tau(t) = |\mathbf{w}(\mathbf{x})|_1 \mathrm{e}^{-|\mathbf{w}(\mathbf{x})|_1 t}$$

To generate the next reaction time, generate r_1 from a uniform distribution on (0,1) and use the equation:

$$\tau = \frac{1}{|\mathbf{w}|_1} \log \frac{1}{r_1}$$

Step 2. Decide which reaction has occurred.

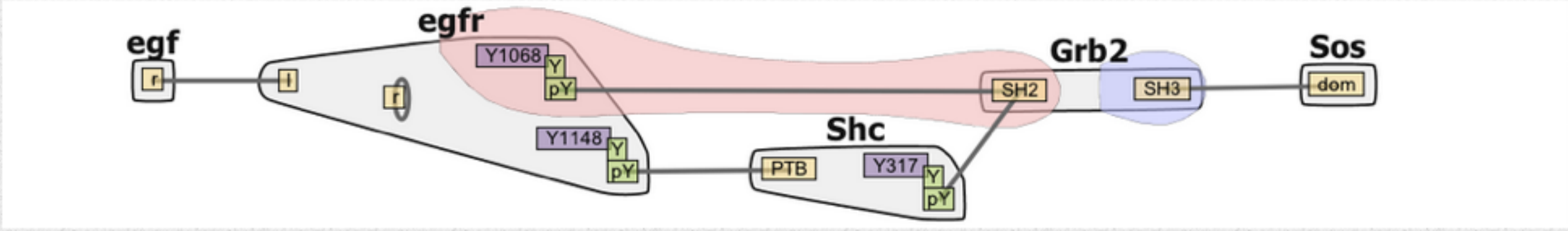
To obtain a realization of which reaction will occur, generate a second uniform random number, r_2, and find the smallest *k* such that:

$$\sum_{\mu=1}^{k-1} w_\mu(\mathbf{x}) \leq r_2 \, |\mathbf{w}|_1 \leq \sum_{\mu=1}^{k} w_\mu(\mathbf{x})$$

Step 3. Update current Time ($t=t+\tau$) and State ($\mathbf{x} = \mathbf{x}+s_k$).

RuleBender:

biological rule-based modeling, simulation, and visualization



Rule Based Modeling and Simulation with RuleBender

+

NFsim

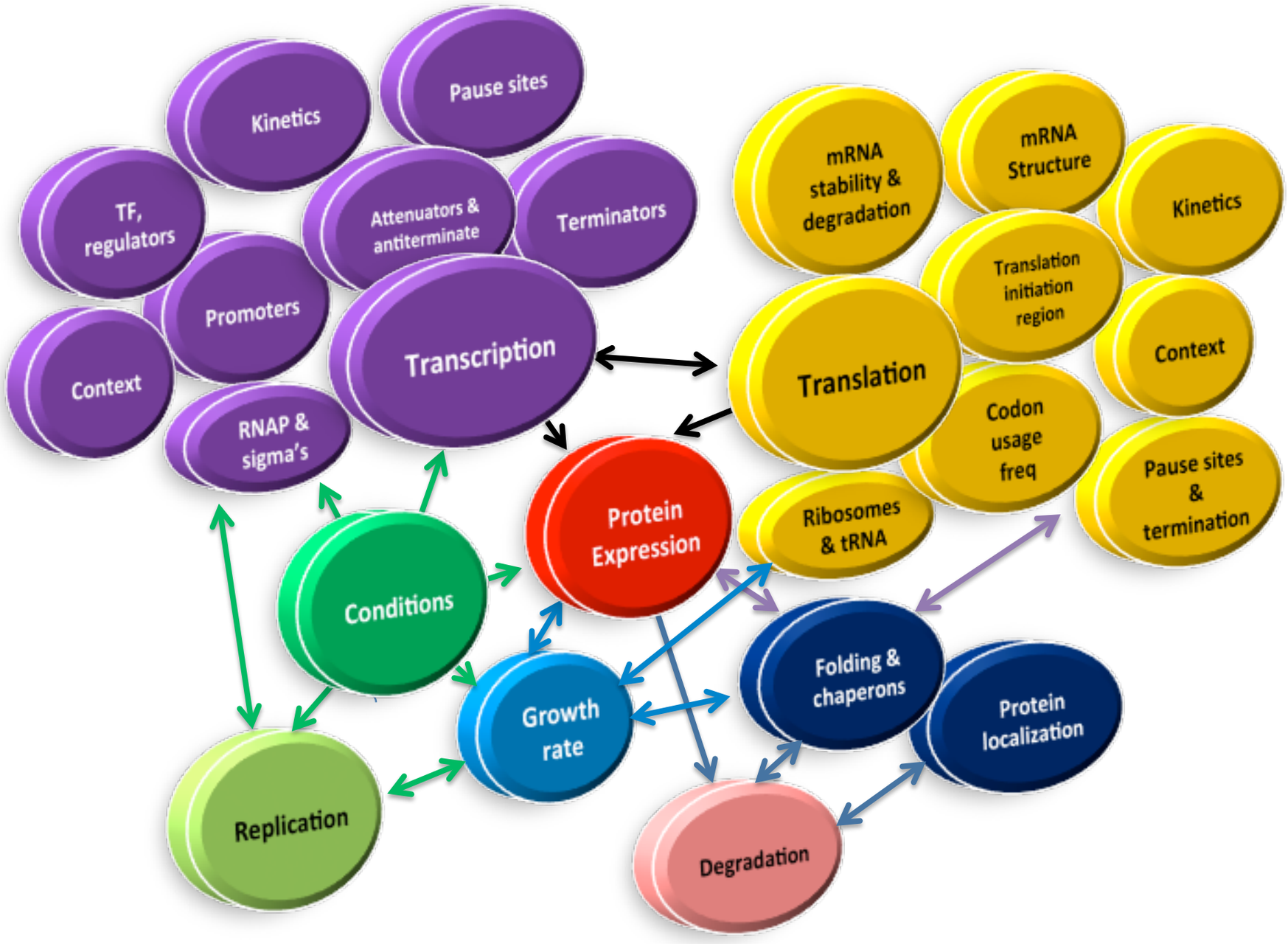
the network free stochastic simulator

=

EFFICIENT RULE BASED MODELS & FAST STOCHASTIC SIMULATIONS OF LARGE SYSTEMS

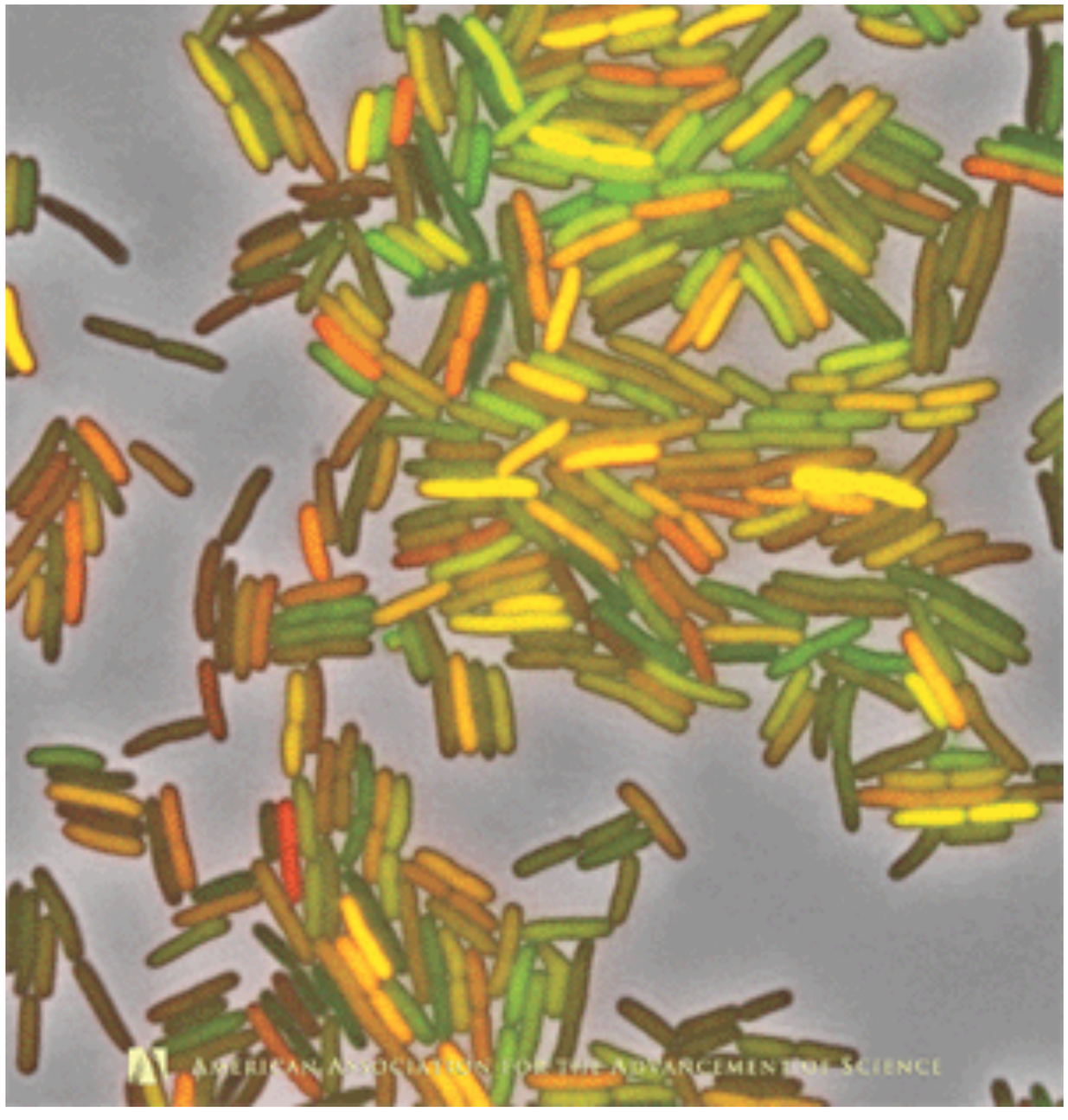
SLIDE CONTRIBUTED BY ADAM ARKIN AND RON WEISS, MIT/BERKELEY

SOURCES OF NOISE



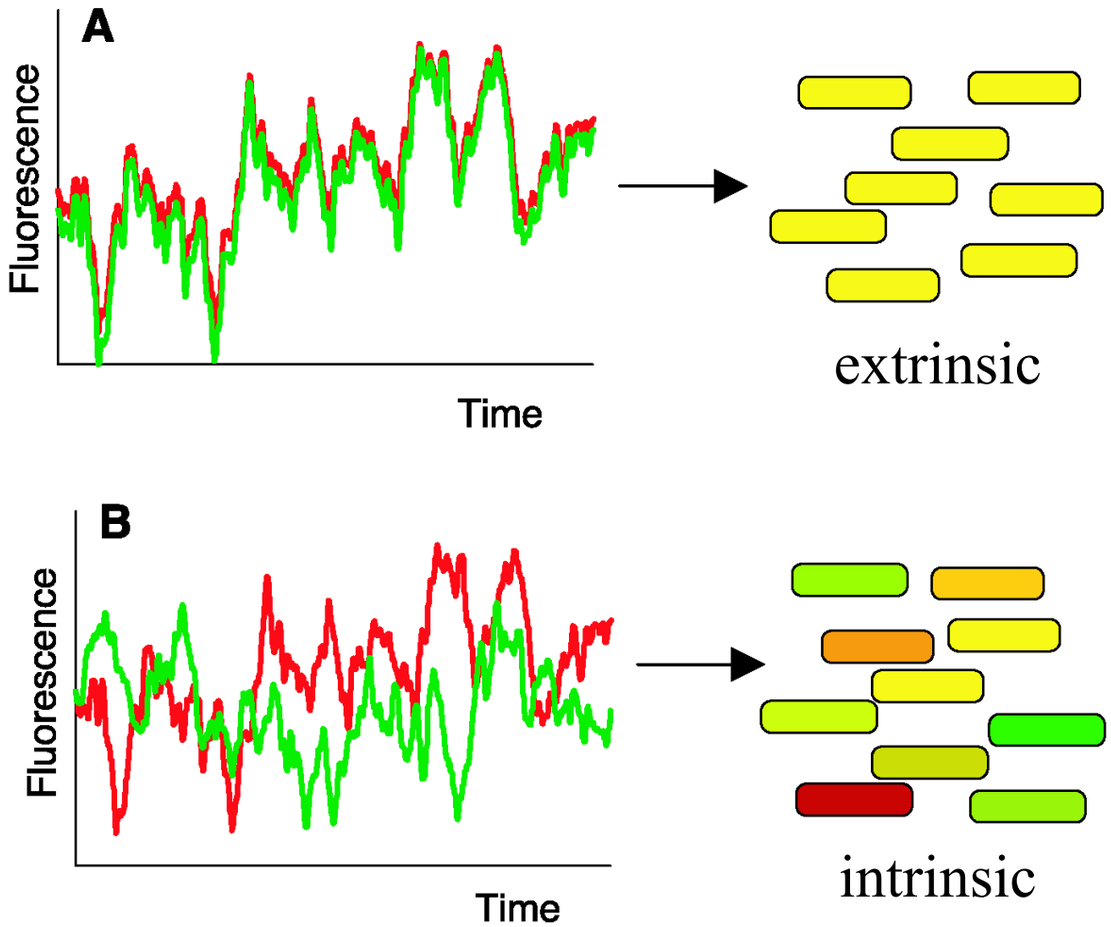
SLIDE CONTRIBUTED BY ADAM ARKIN AND RON WEISS, MIT/BERKELEY

SOURCES OF NOISE



SLIDE CONTRIBUTED BY ADAM ARKIN AND RON WEISS, MIT/BERKELEY

EXTRINSIC VS INTRINSIC NOISE



SLIDE CONTRIBUTED BY ADAM ARKIN AND RON WEISS, MIT/BERKELEY

QUANTIFICATION OF EXTRINSIC VS INTRINSIC NOISE

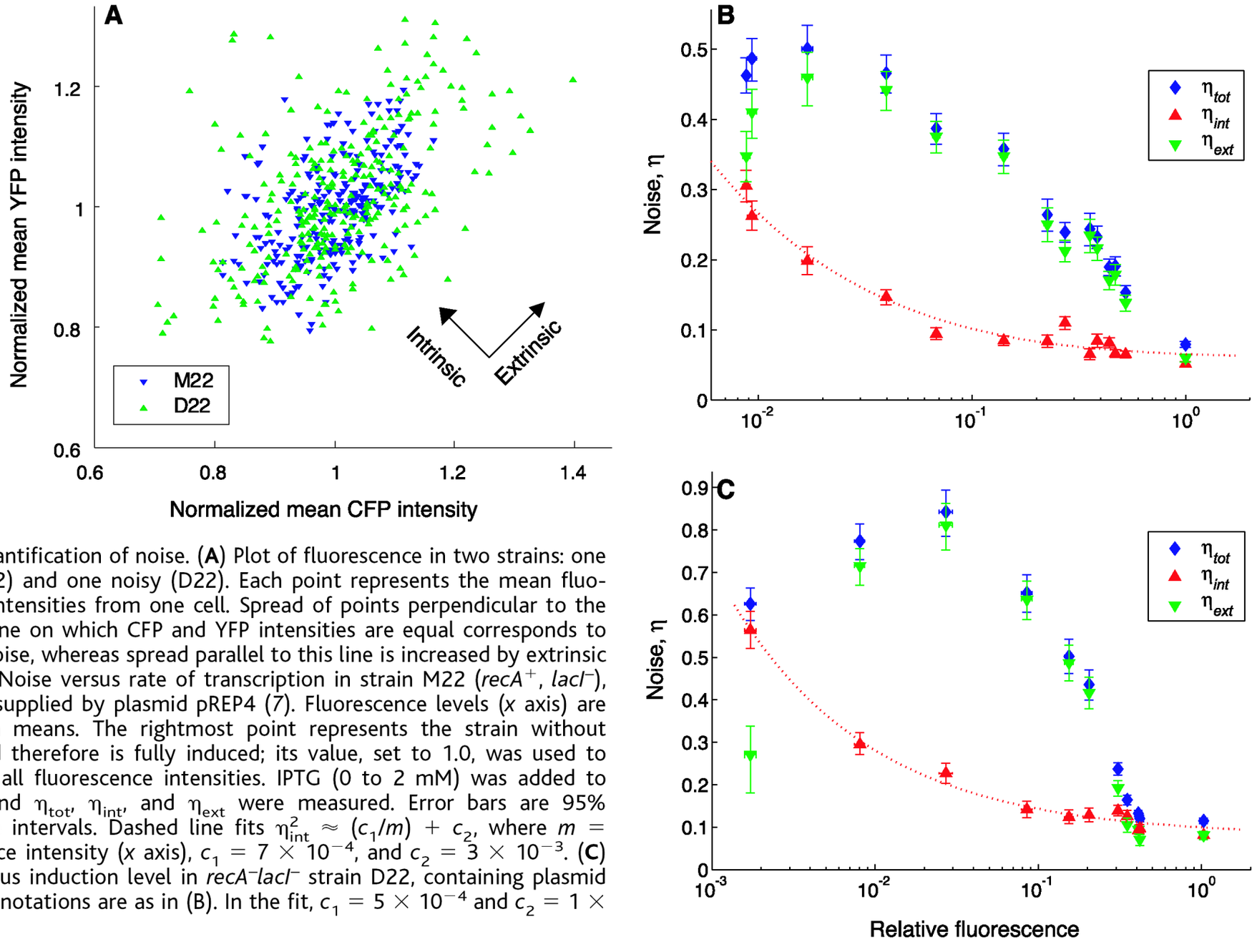


Fig. 3. Quantification of noise. (**A**) Plot of fluorescence in two strains: one quiet (M22) and one noisy (D22). Each point represents the mean fluorescence intensities from one cell. Spread of points perpendicular to the diagonal line on which CFP and YFP intensities are equal corresponds to intrinsic noise, whereas spread parallel to this line is increased by extrinsic noise. (**B**) Noise versus rate of transcription in strain M22 (*recA*$^+$, *lacI*$^-$), with LacI supplied by plasmid pREP4 (*7*). Fluorescence levels (*x* axis) are population means. The rightmost point represents the strain without pREP4 and therefore is fully induced; its value, set to 1.0, was used to normalize all fluorescence intensities. IPTG (0 to 2 mM) was added to cultures and η_{tot}, η_{int}, and η_{ext} were measured. Error bars are 95% confidence intervals. Dashed line fits $\eta^2_{int} \approx (c_1/m) + c_2$, where m = fluorescence intensity (*x* axis), $c_1 = 7 \times 10^{-4}$, and $c_2 = 3 \times 10^{-3}$. (**C**) Noise versus induction level in *recA*$^-$*lacI*$^-$ strain D22, containing plasmid pREP4. All notations are as in (B). In the fit, $c_1 = 5 \times 10^{-4}$ and $c_2 = 1 \times 10^{-2}$.