Towards a global understanding of protein production in prokaryotes

Emanuele LEONCINI

INRIA Rocquencourt - INRA Jouy-en-Josas

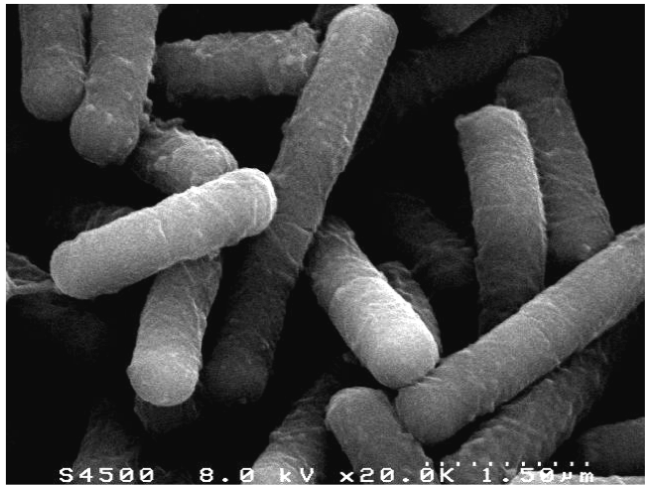
PhD Defense

December 17th 2013

Outline:

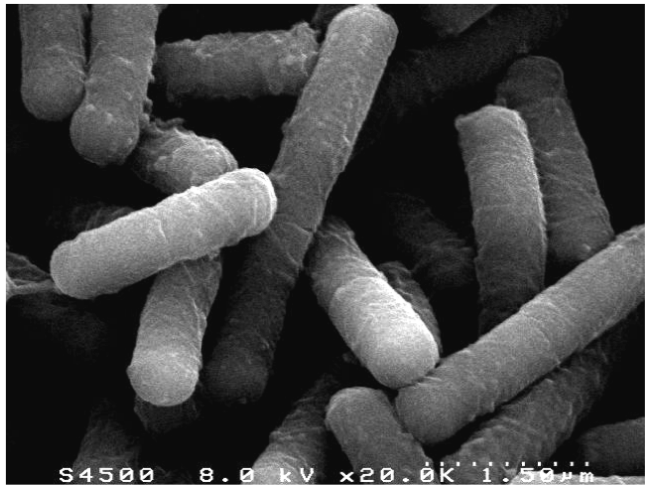
Central role of proteins in prokaryotes

- Proteins are the core of biologic processes: *enzymes*, DNA replication machinery, ...
- $\sim$ 50% of the bacteria dry weight
- $\sim$ 3.5 millions of proteins in each cell
- $\sim$ 2000 types of proteins produced at any time at any growth condition
- proteins ranging from few dozens up to 10^5



Central role of proteins in prokaryotes

- Proteins are the core of biologic processes: *enzymes*, DNA replication machinery, ...
- $\sim$ 50% of the bacteria dry weight
- $\sim$ 3.5 millions of proteins in each cell
- $\sim$ 2000 types of proteins produced at any time at any growth condition
- proteins ranging from few dozens up to 10^5

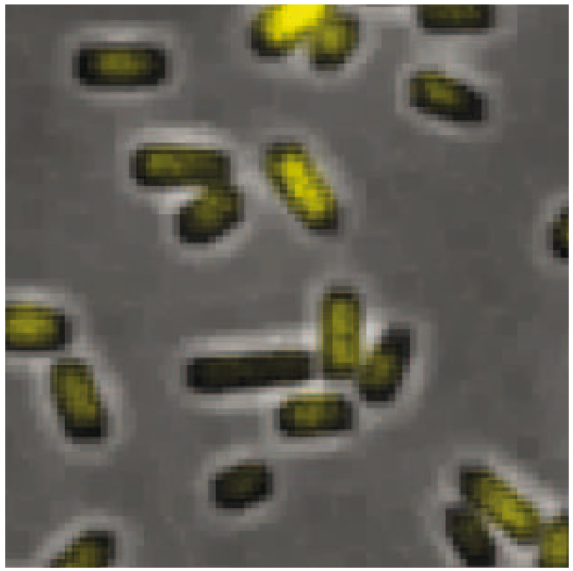


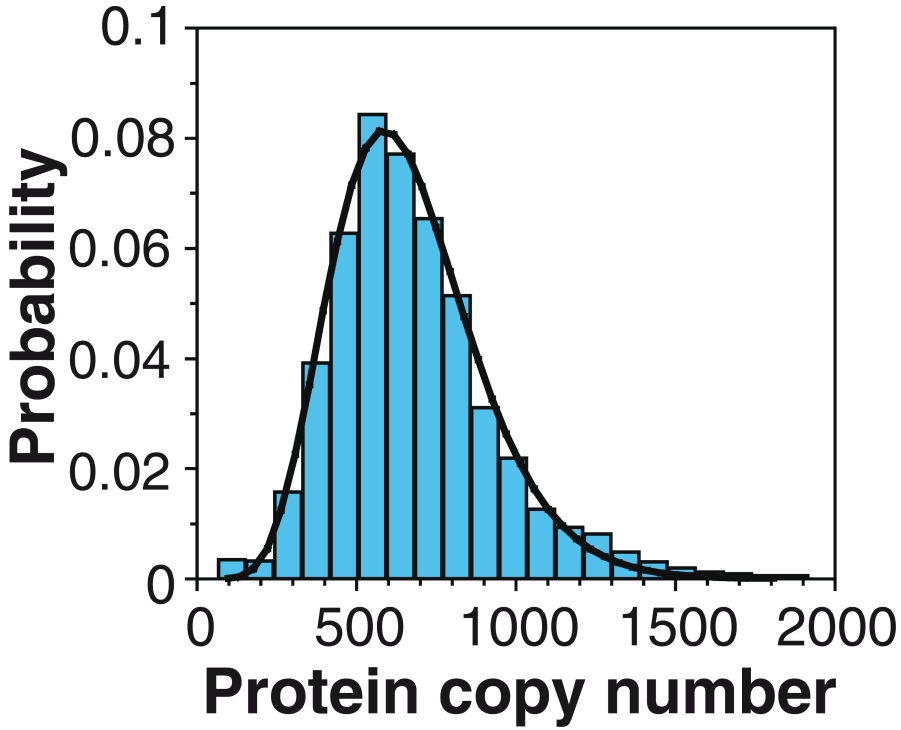
A highly consuming process:

- at each generation, the bacterium has to duplicate all proteins
- more than 85% of cell resources

Stochasticity in protein production: experimental viewpoint

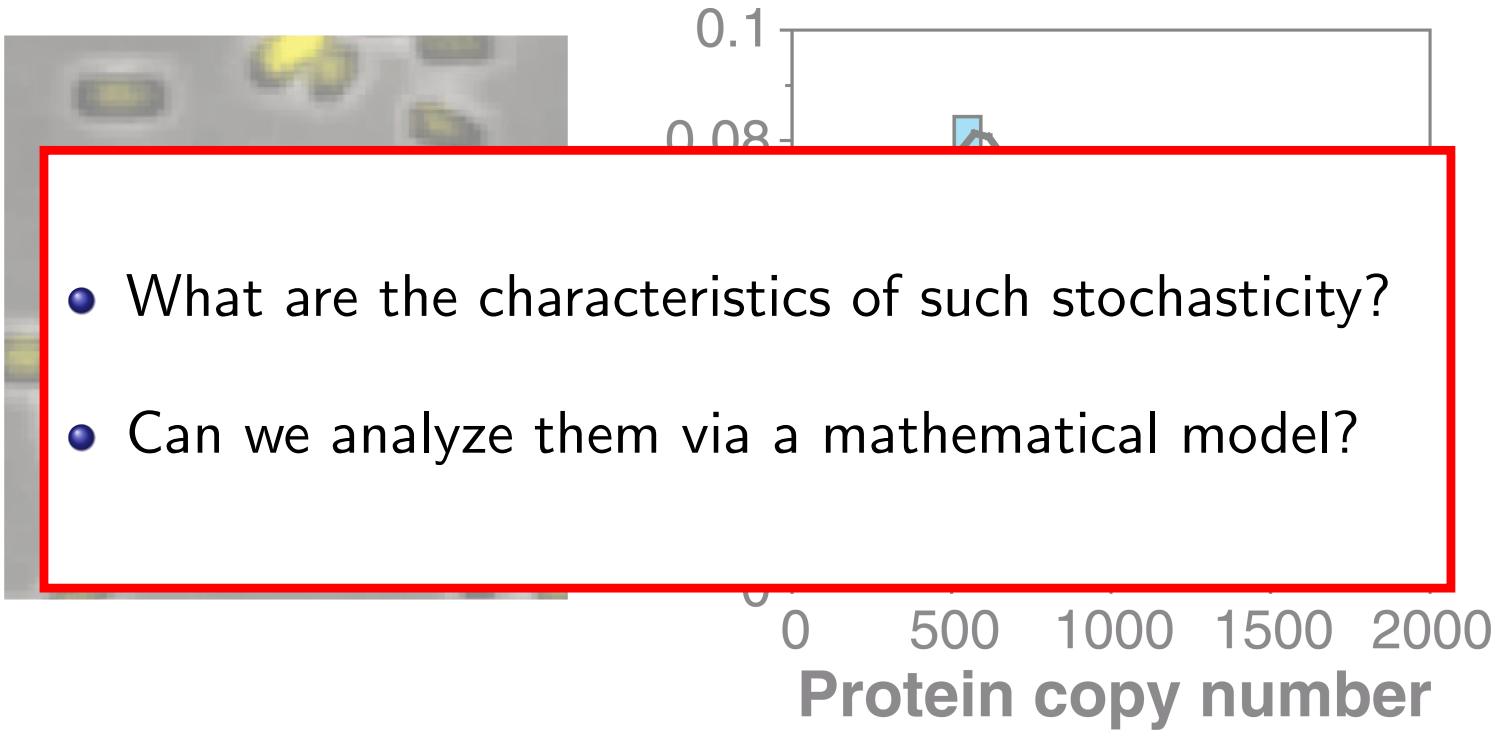
E. Coli – ADk cytoplasm protein[1]





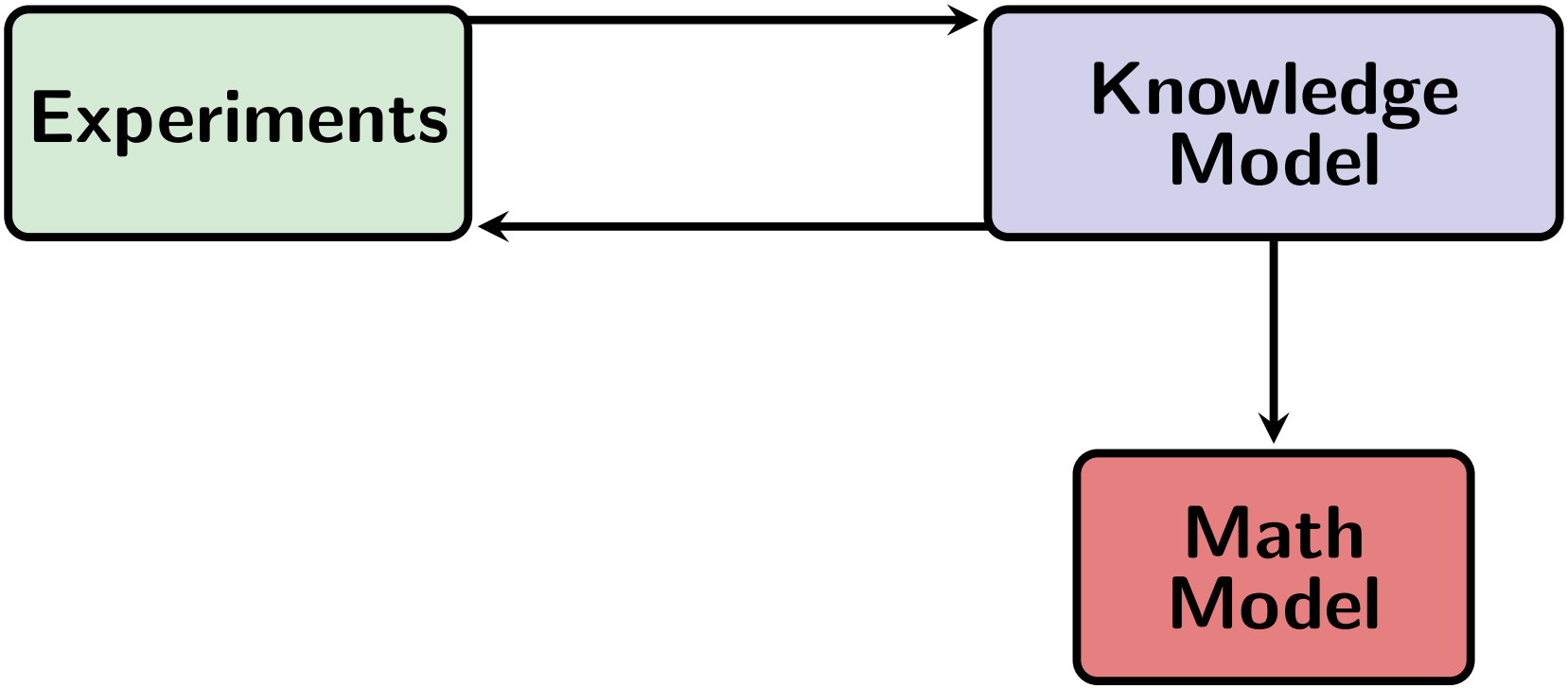
[1] Yuichi Taniguchi et al. *Science* (2010), pp. 533–538.

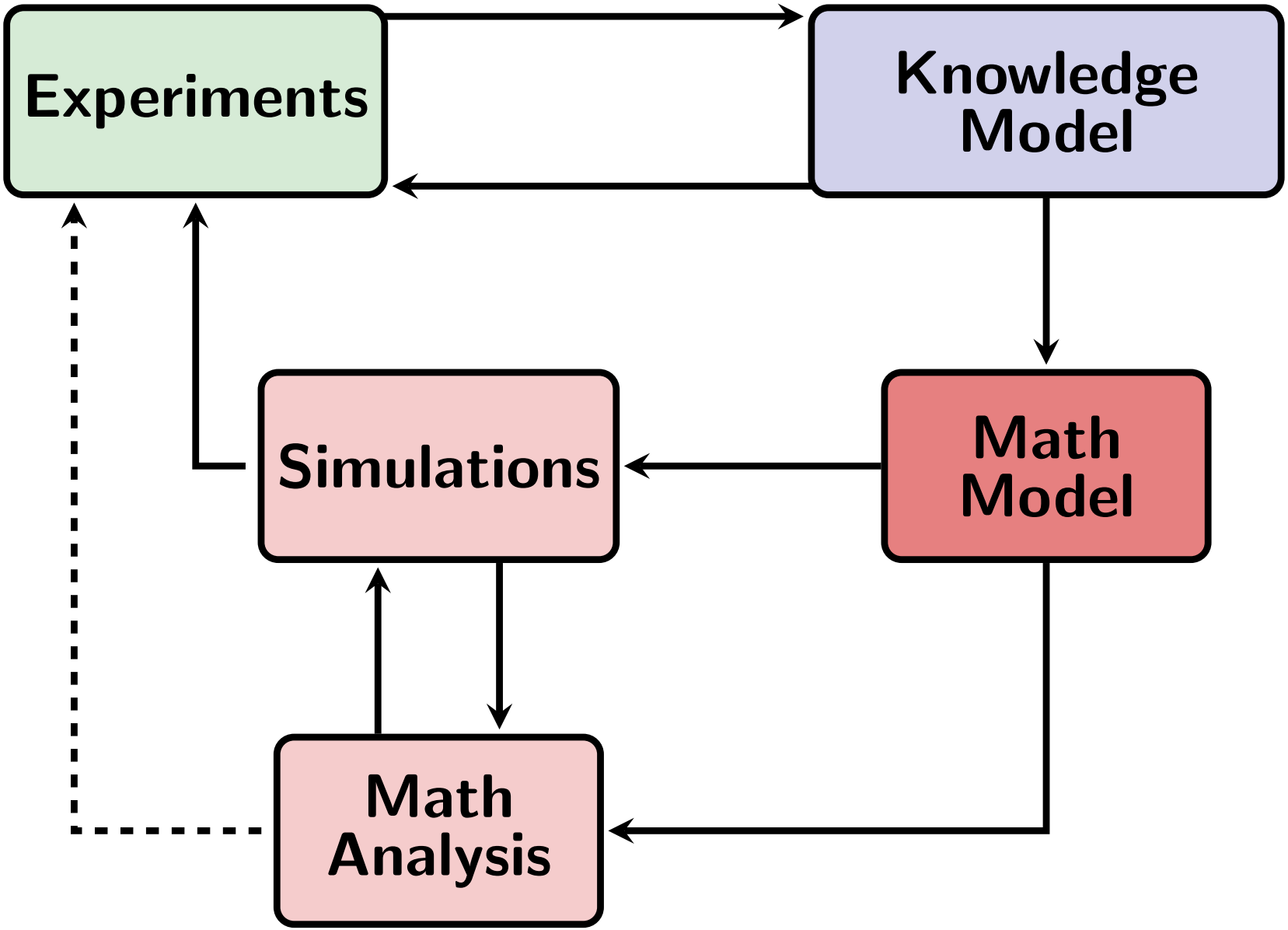
Stochasticity in protein production: experimental viewpoint

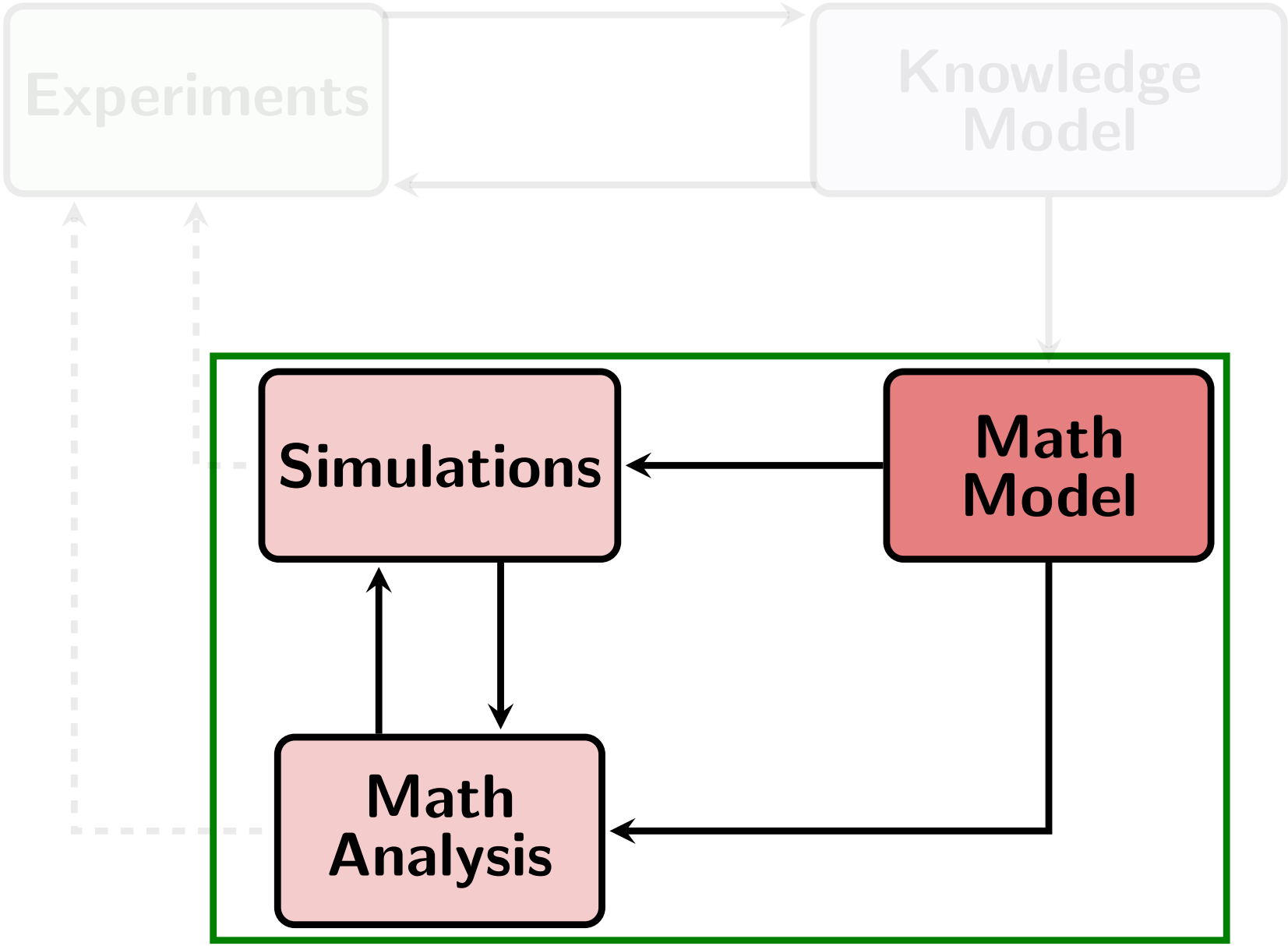


- What are the characteristics of such stochasticity?
- Can we analyze them via a mathematical model?


Experiments
Knowledge
Model


Experiments
Knowledge Model
Math Model

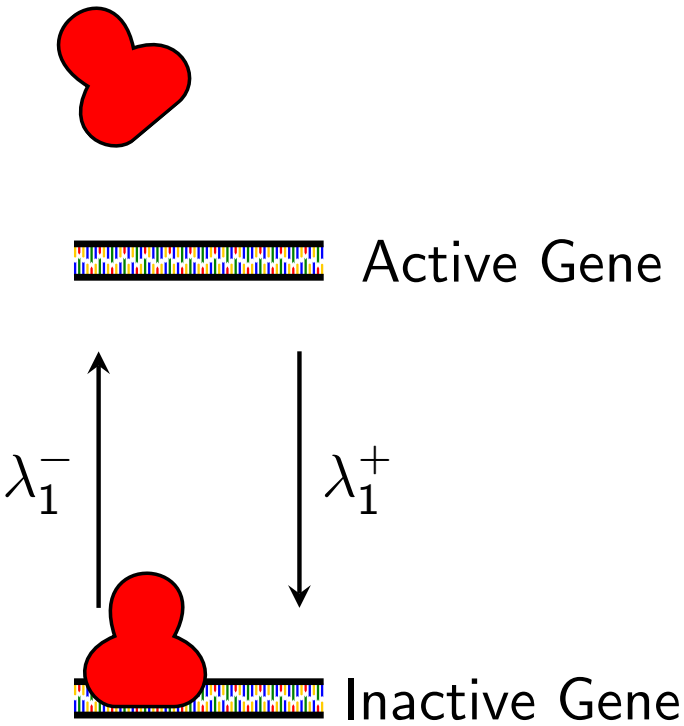

Experiments
Knowledge Model
Simulations
Math Model
Math Analysis


Experiments
Knowledge
Model
Simulations
Math
Model
Math
Analysis

Part I

Stochastic model of Gene Expression: Single-Protein

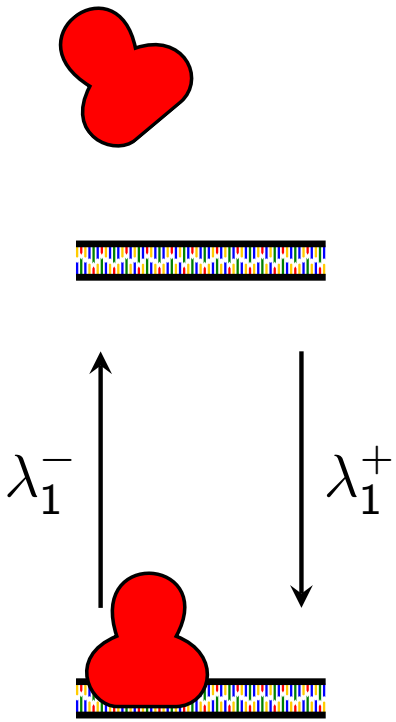
4-Stage model: activation



$Y(t) \in \{0, 1\}$

Gene status

4-Stage model: activation



$Y(t) \in \{0, 1\}$

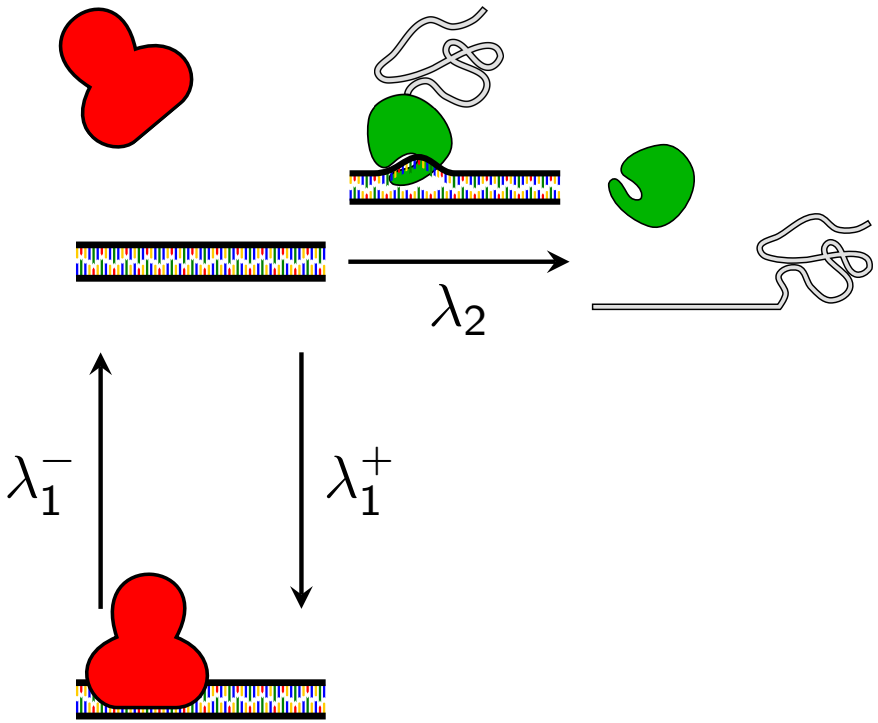
Gene status

Exponential assumption

- diffusion in a stiff medium
- thermal noise
- biochemical reactions (in general): encounter of macromolecules

exponentially distributed duration

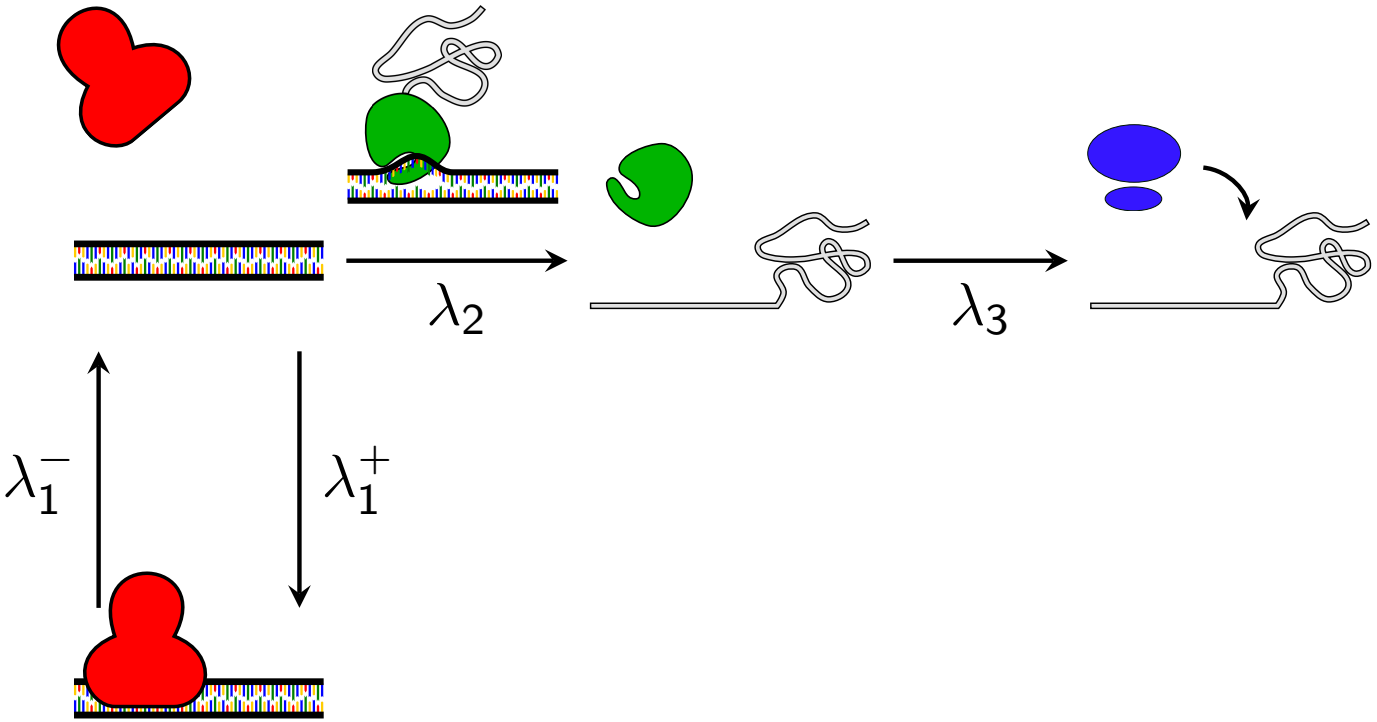
4-Stage model: transcription



$$Y(t) \in \{0,1\} \longrightarrow M(t) \in \mathbb{N}$$

Messenger

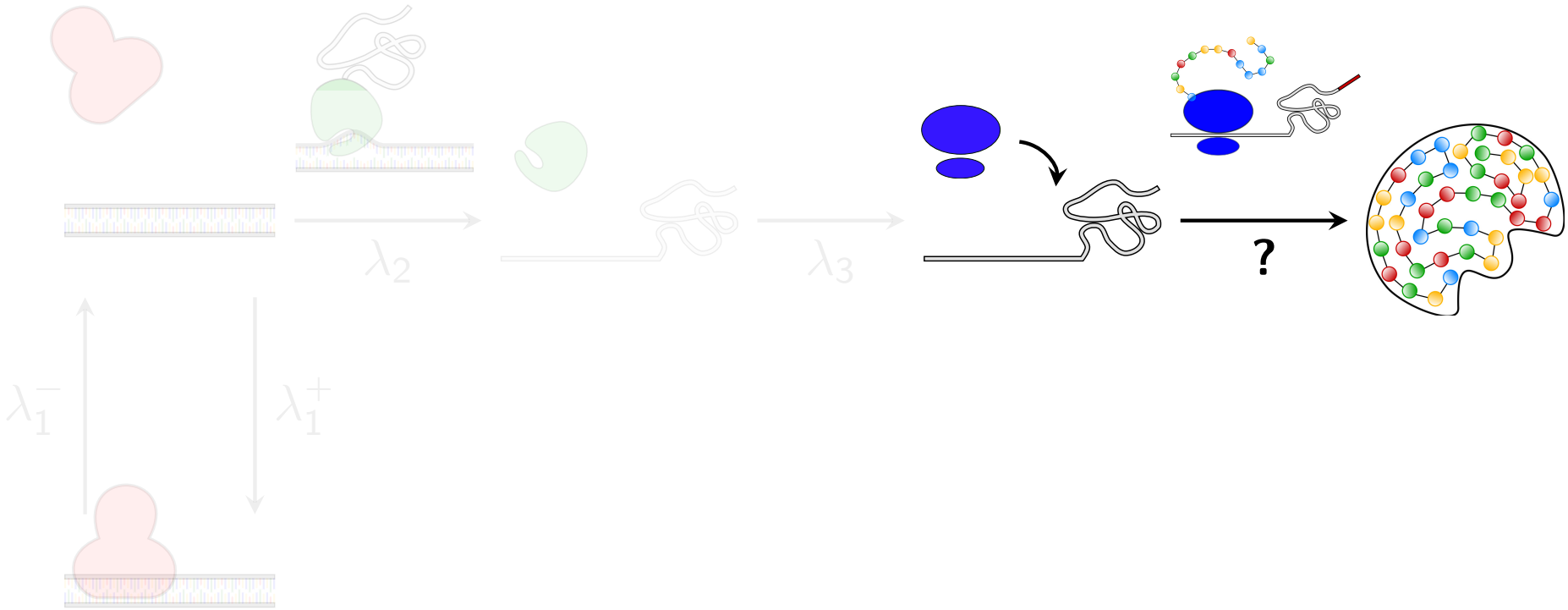
4-Stage model: translation initiation

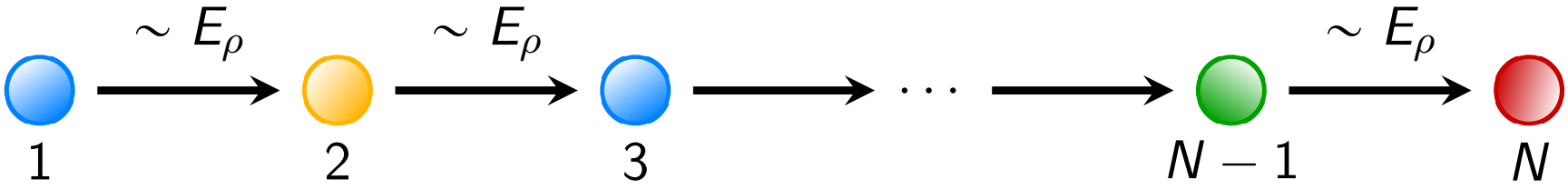


$$Y(t) \in \{0,1\} \longrightarrow M(t) \in \mathbb{N} \longrightarrow R(t) \in \mathbb{N}$$

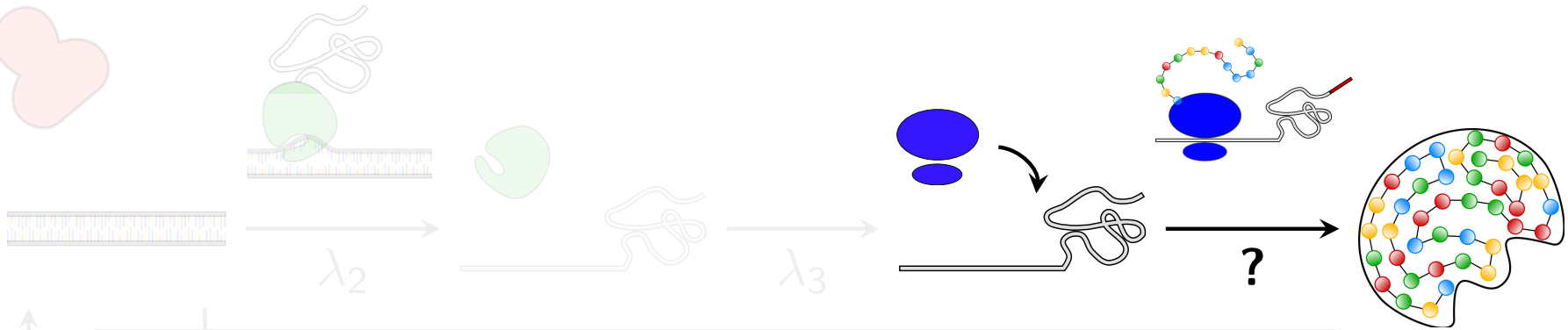
Ribosome

4-Stage model: translation completion



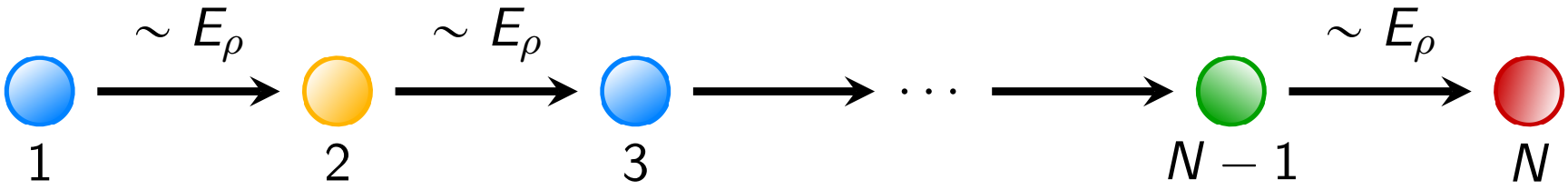


4-Stage model: translation completion

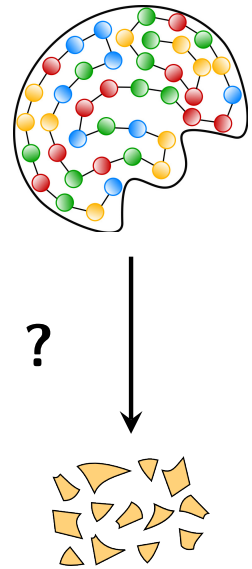


Protein Elongation

- exponentially distributed elementary steps
- Erlang distributed elongation $T^{(\text{el.})} \sim \text{Erl}(N, \rho)$ ($N \approx 400$ in average)



4-Stage model



Protein decay

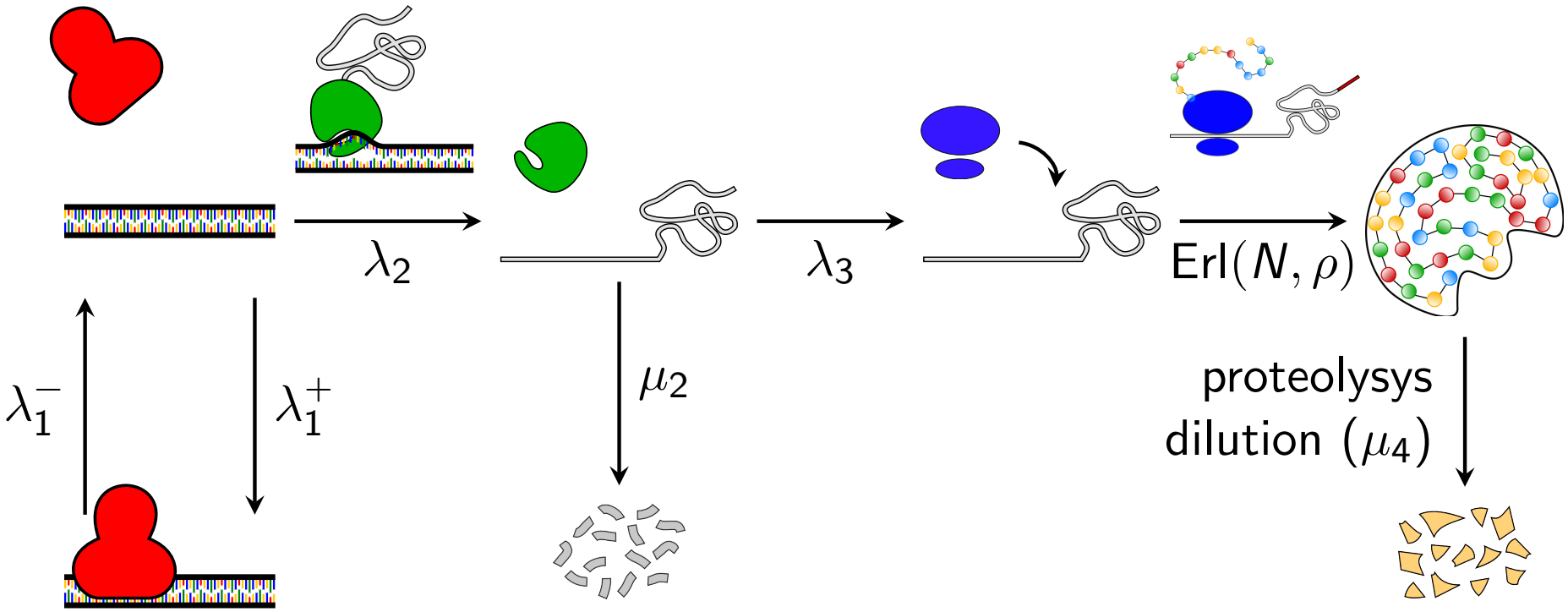
Proteolysis:

- stochastic & discrete;
- exponentially distributed duration;

Dilution:

- deterministic & continuous;
- main decay mechanism;

4-Stage model



$$Y(t) \in \{0,1\} \xrightarrow{\lambda_2} M(t) \in \mathbb{N} \xrightarrow{\lambda_3} R(t) \in \mathbb{N} \xrightarrow{\mathrm{Erl}(N,\rho)} P(t) \in \mathbb{N}$$

$$M(t) \xrightarrow{\mu_2} \varnothing \qquad P(t) \xrightarrow{\text{lysys/dil.}} \varnothing$$

4-Stage model

Protein production system:

- fixed number of proteins required
- more than 85% of resources

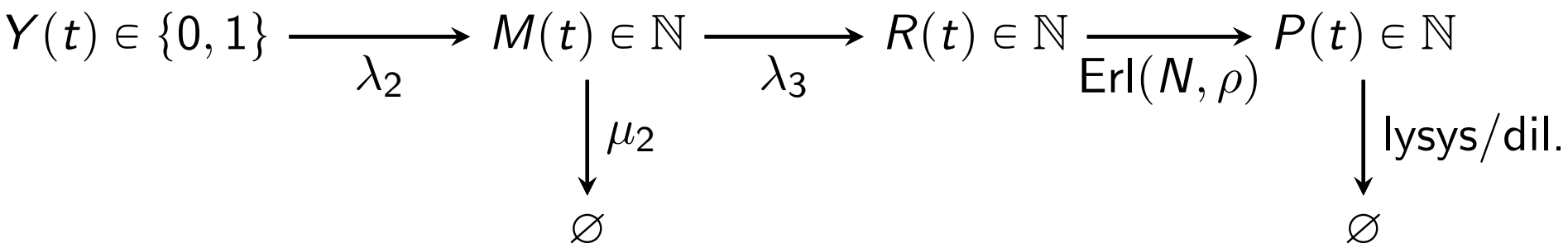
→ variance is a **key parameter**

$$Y(t) \in \{0,1\} \xrightarrow[\lambda_2]{} M(t) \in \mathbb{N} \xrightarrow[\lambda_3]{} R(t) \in \mathbb{N} \xrightarrow[\mathrm{Erl}(N,\rho)]{} P(t) \in \mathbb{N}$$

$$M(t) \in \mathbb{N} \downarrow \mu_2 \quad \varnothing$$

$$P(t) \in \mathbb{N} \downarrow \text{lysys/dil.} \quad \varnothing$$

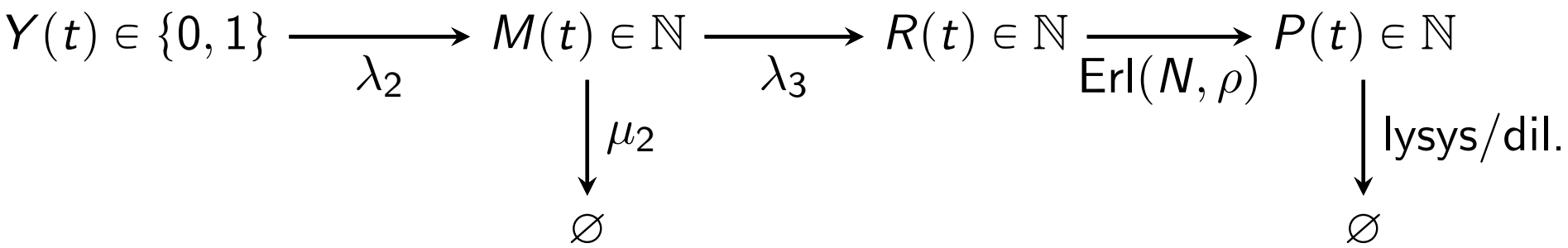
4-Stage model



Goals:

- characterize **mean** and **variance** of proteins P at equilibrium
- identify the role of the different parameters

4-Stage model



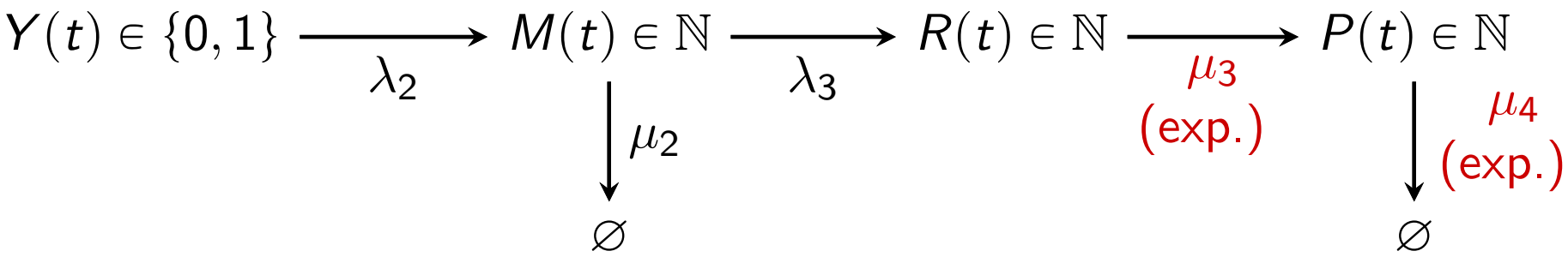
Goals:

- characterize **mean** and **variance** of proteins P at equilibrium
- identify the role of the different parameters

Why is it difficult to achieve these goals?

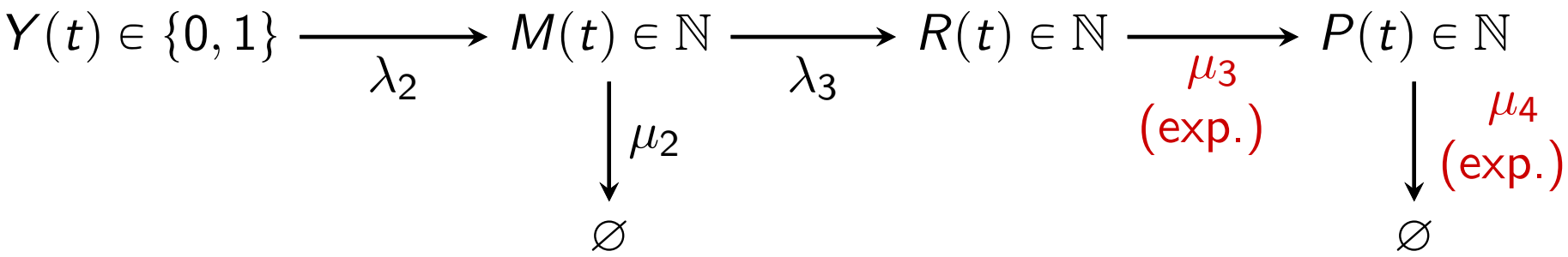
The need for a new framework to overcome the limitations of the classic approach.

Limits of classic approach



- *Assumption:* each step has exponentially distributed duration
- *Markovian* description of protein production

Limits of classic approach



- *Assumption:* each step has exponentially distributed duration
- *Markovian* description of protein production

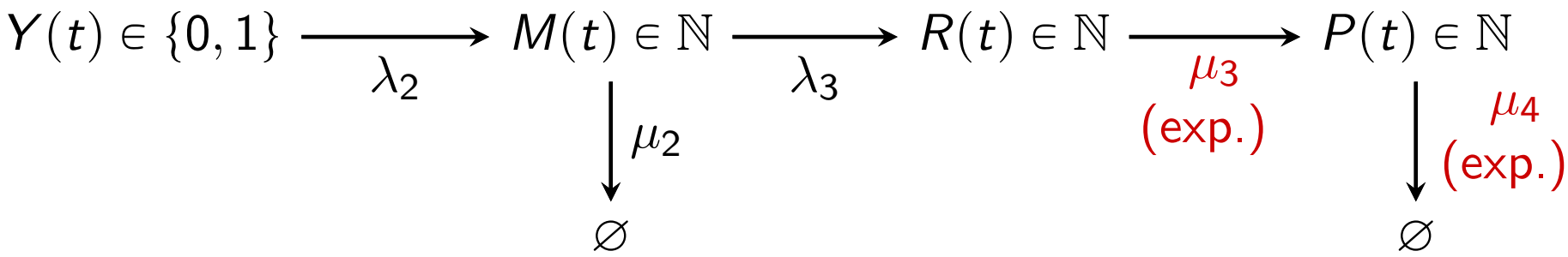
30 years of gene expression:

Berg (1978), Rigney (1979), Swain (2002), Paulsson (2005), ...

Original motivation for Math Models:

Lack of experimental data

Limits of classic approach



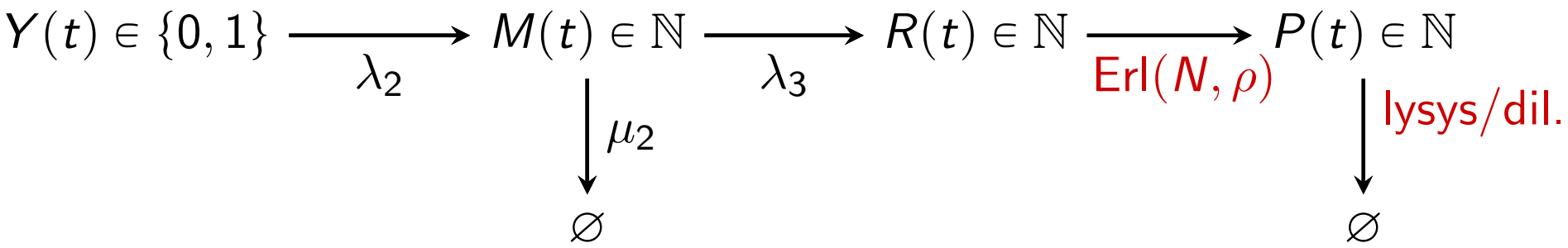
- *Assumption:* each step has exponentially distributed duration
- *Markovian* description of protein production

Results:

- *reference model* for biologists
- quantitative characterization of protein fluctuations

$$\mathrm{var}(P) = \mathbb{E}[P]\left[1 + \frac{\lambda_3\mu_3(\mu_2+\mu_3+\mu_4)}{(\mu_2+\mu_3)(\mu_2+\mu_4)(\mu_3+\mu_4)}\right]$$

Limits of classic approach

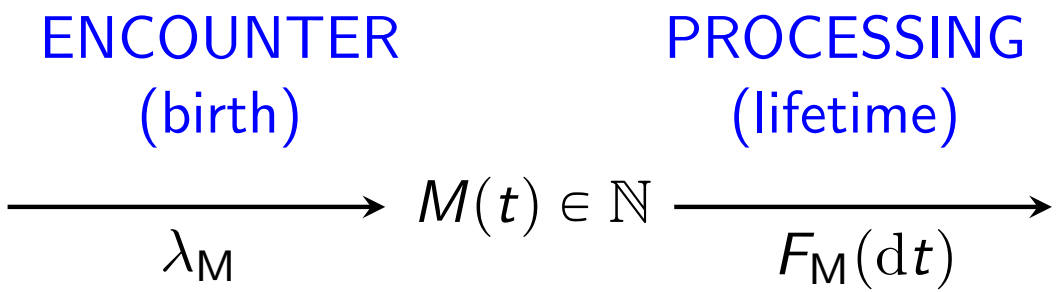


Limitations

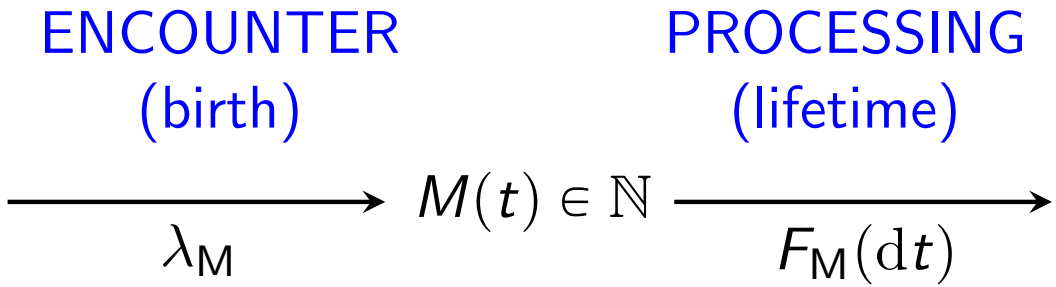
Classic framework cannot be used in non-Markovian description

How to include non exponential steps?
Marked Poisson Point Process (MPPP) framework

Explanatory model



Explanatory model



Assumptions:

- births (s_n) follow a **Poisson process** of parameter λ_M
- lifetimes (σ_n) with distribution $F_M(\mathrm{d}t)$ (**mark**)

Explanatory model

ENCOUNTER (birth) PROCESSING (lifetime)

$$\xrightarrow[\lambda_M]{} M(t) \in \mathbb{N} \xrightarrow[F_M(\mathrm{d}t)]{}$$

Assumptions:

- births (s_n) follow a **Poisson process** of parameter λ_M
- lifetimes (σ_n) with distribution $F_M(\mathrm{d}t)$ (**mark**)

$\mathcal{M} = (s_n, \sigma_n)$ **marked Poisson point process** on $\mathbb{R} \times \mathbb{R}_+$ with intensity $\mu_{\mathcal{M}} = \lambda_M \, \mathrm{d}u \otimes F_M(\mathrm{d}v)$

Explanatory model

ENCOUNTER (birth) PROCESSING (lifetime)

$$\xrightarrow[\lambda_M]{} M(t) \in \mathbb{N} \xrightarrow[F_M(\mathrm{d}t)]{}$$

At equilibrium, $M(\infty)$ is a functional of Poisson process $\mathcal{M}$:

$$M(\infty) = \mathcal{M}\left(\mathbb{1}_{\{u \leqslant 0 \leqslant u+v\}}\right) = \int_{\mathbb{R}\times\mathbb{R}_+} \mathbb{1}_{\{u \leqslant 0 \leqslant u+v\}} \mathcal{M}(\mathrm{d}u, \mathrm{d}v)$$

where $\mathcal{M}(f) = \sum_n f\left(s_n, \sigma_n\right) = \int_{\mathbb{R}\times\mathbb{R}_+} f(x,y)\mathcal{M}(\mathrm{d}x, \mathrm{d}y)$

Explanatory model

ENCOUNTER (birth) PROCESSING (lifetime)

$$\xrightarrow[\lambda_M]{} M(t) \in \mathbb{N} \xrightarrow[F_M(\mathrm{d}t)]{}$$

At equilibrium, $M(\infty)$ is a functional of Poisson process $\mathcal{M}$:

$$M(\infty) = \mathcal{M}\left(\mathbb{1}_{\{u \leqslant 0 \leqslant u+v\}}\right) = \int_{\mathbb{R}\times\mathbb{R}_+} \mathbb{1}_{\{u \leqslant 0 \leqslant u+v\}} \mathcal{M}(\mathrm{d}u, \mathrm{d}v)$$

where $\mathcal{M}(f) = \sum_n f\left(s_n, \sigma_n\right) = \int_{\mathbb{R}\times\mathbb{R}_+} f(x,y)\mathcal{M}(\mathrm{d}x, \mathrm{d}y)$

For any nice $f : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$, the **Laplace transform** of $\mathcal{M}$ is

$$\mathcal{L}_{\mathcal{M}}(f) = \mathbb{E}\left[e^{-\mathcal{M}(f)}\right] = \exp\left(-\int\left(1 - e^{-f(u,v)}\right)\lambda_M \,\mathrm{d}u F_M(\mathrm{d}v)\right)$$

Explanatory model

ENCOUNTER (birth) PROCESSING (lifetime)

$$\xrightarrow[\lambda_M]{} M(t) \in \mathbb{N} \xrightarrow[F_M(\mathrm{d}t)]{}$$

At equilibrium, $M(\infty)$ is a functional of Poisson process $\mathcal{M}$:

$$M(\infty) = \mathcal{M}\left(\mathbb{1}_{\{u \leqslant 0 \leqslant u+v\}}\right) = \int_{\mathbb{R}\times\mathbb{R}_+} \mathbb{1}_{\{u \leqslant 0 \leqslant u+v\}} \mathcal{M}(\mathrm{d}u, \mathrm{d}v)$$

where $\mathcal{M}(f) = \sum_n f\left(s_n, \sigma_n\right) = \int_{\mathbb{R}\times\mathbb{R}_+} f(x,y)\mathcal{M}(\mathrm{d}x, \mathrm{d}y)$

Proposition

$$\mathbb{E}\left[M\right] = \lambda_M \mathbb{E}\left[\sigma\right], \qquad \sigma \sim F_M(\mathrm{d}v)$$
$$\mathrm{var}(M) = \mathbb{E}\left[M\right]$$

Proteins

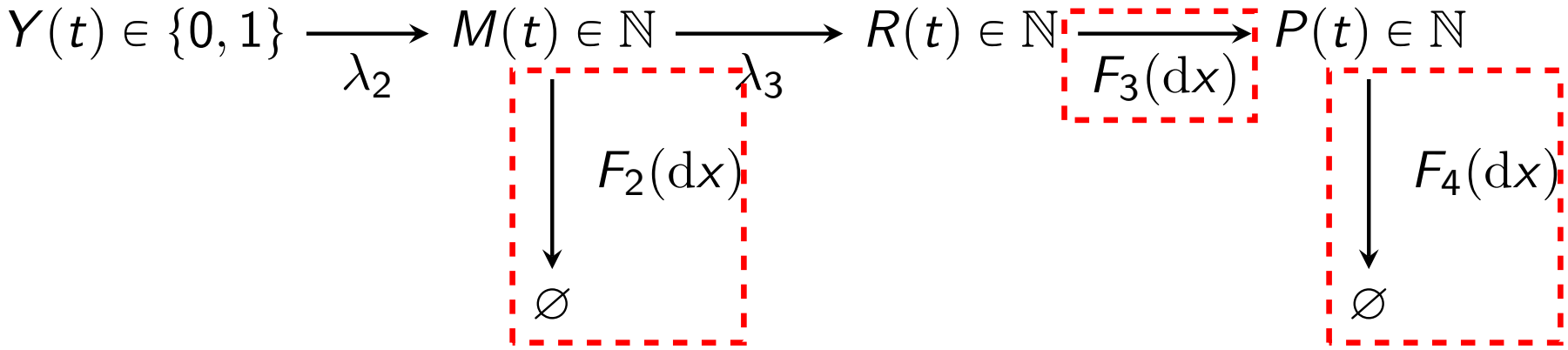
In the case of proteins

- $\mathcal{P} = \left(s_n, \sigma_n, \mathcal{N}_{\tilde{\lambda}}^{\sigma_n}\right)$ is a Poisson process with intensity measure $\mu_{\mathcal{P}}$
- the **mark** is now $\left(\sigma_n, \mathcal{N}_{\tilde{\lambda}}^{\sigma_n}\right)$
- $\mathcal{N}_{\tilde{\lambda}}^{\sigma_n}$ is a Poisson process on $\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$ associated to a mRNA born in s_n

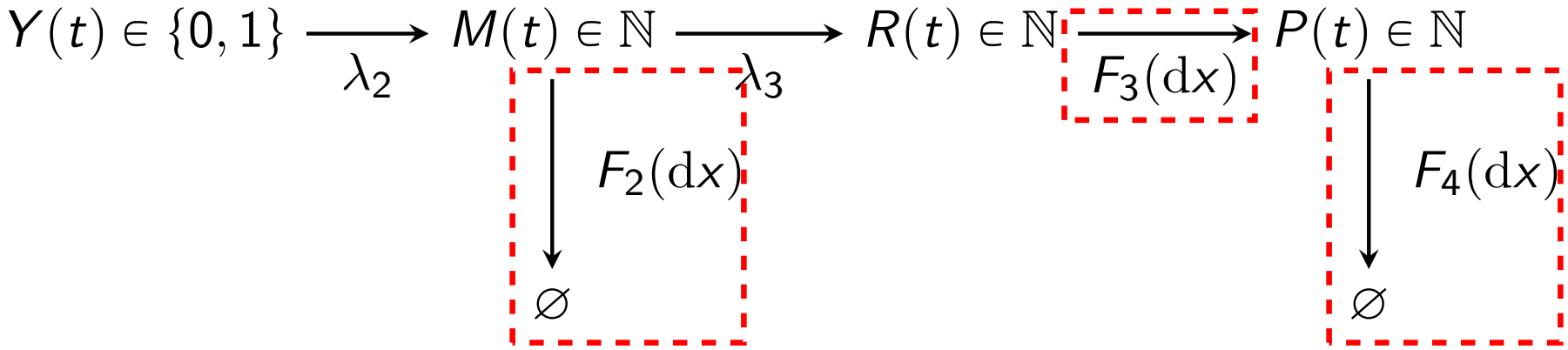
same approach but more complicated

MPPP & Gene Expression

General formulas of protein statistics



General formulas of protein statistics



Mean:

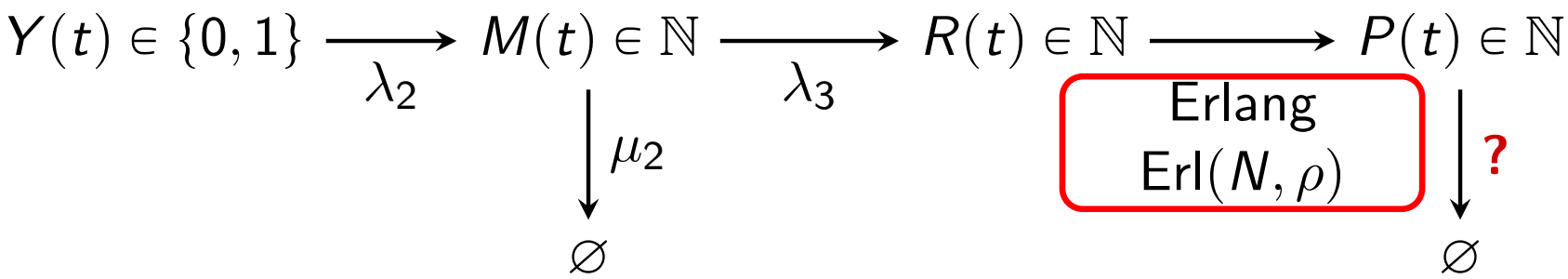
$$\mathbb{E}(P) = \lambda_2 \lambda_3 \mathbb{E}(F_2)\mathbb{E}(F_4)$$

Variance (always active gene):

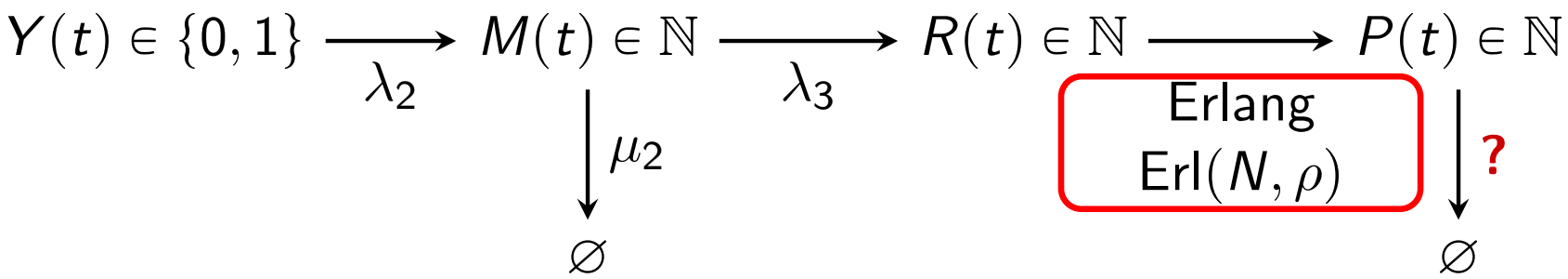
$$\mathrm{var}(P) = \mathbb{E}(P) + \lambda_2 \lambda_3^2 \int_{\mathbb{R}\times\mathbb{R}_+} \mathbb{1}_{\{s\leqslant 0\}} \left[\int_{\mathbb{R}_+^2} \mathbb{1}_{\{-(u+s+t)\leqslant y\leqslant -(u+s)\}} \overline{F}_4(u)\, \mathrm{d}u F_3(\mathrm{d}y) \right]^2 \mathrm{d}s F_2(\mathrm{d}t)$$

Application

Application



Protein decay: Proteolysis & Dilution



Protein decay:

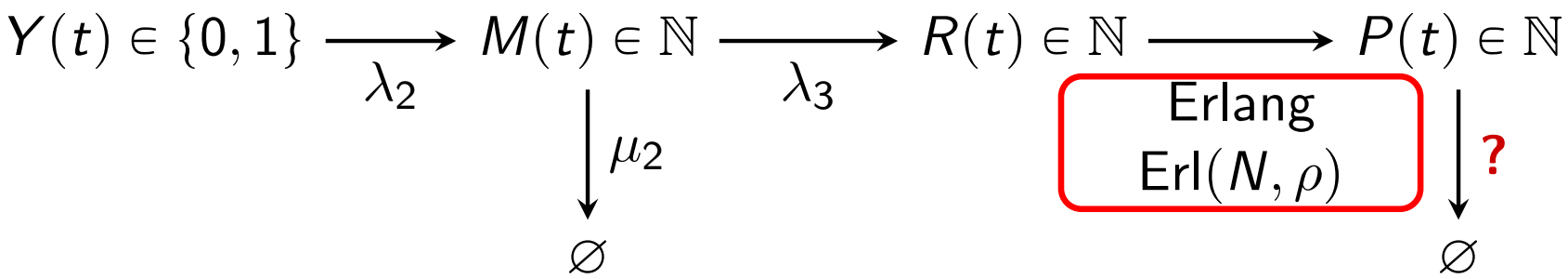
Proteolysis

1. stochastic and discrete nature (*exponential*);
2. used classically to describe protein decay

Dilution

1. main phenomenon of decay (in general);
2. deterministic and continuous nature;
3. understudied phenomenon in models;

Protein decay: Proteolysis & Dilution

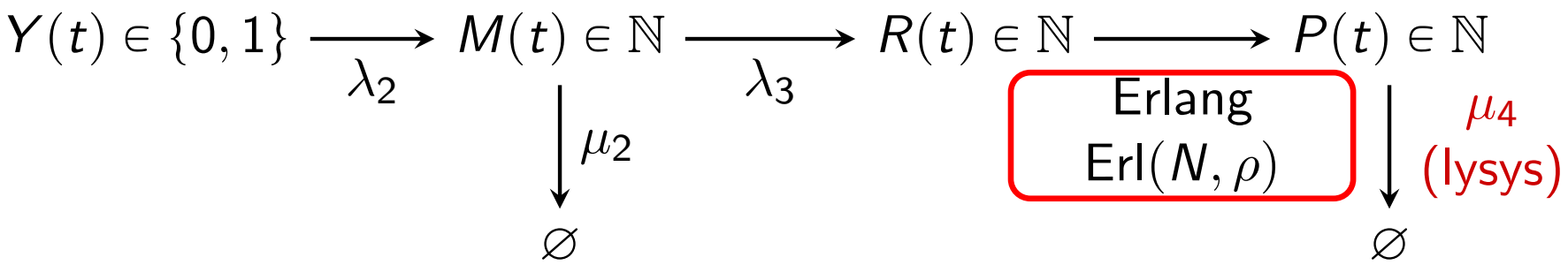


Protein decay:

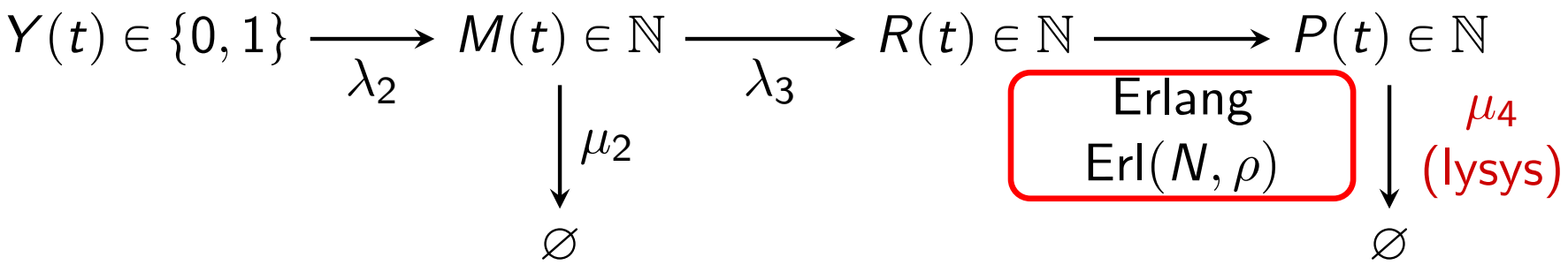
MPPP framework can describe both *proteolysis* and *dilution*

- *Proteolysis:* exponentially distributed of parameter μ_4
- *Dilution:* deterministic, with growth rate parameter ν

Application: protein elongation & proteolysis



Application: protein elongation & proteolysis

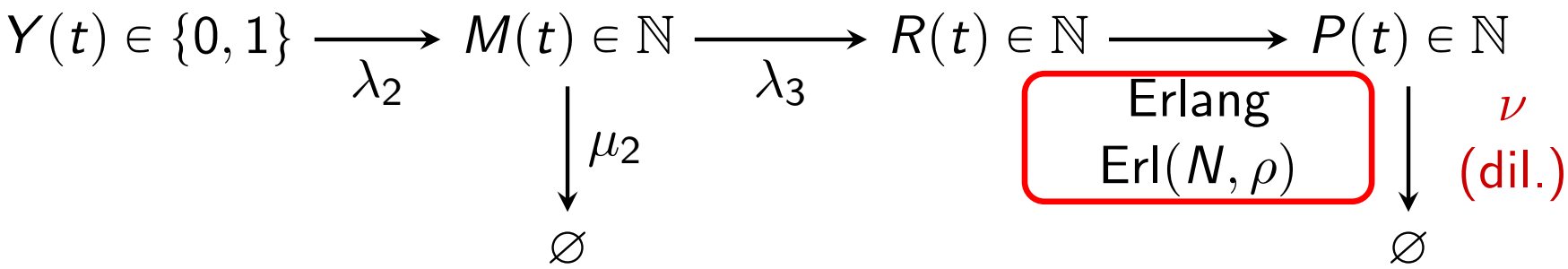


Protein variance: Erlang elongation & proteolysis

$$\mathrm{var}^{(\mathrm{Erl})}_{\mathrm{LYSIS}}(P) = \mathbb{E}[P]\left[1 + \frac{2\lambda_3\mu_2}{\mu_2^2 - \mu_4^2}\frac{\rho^{2N}}{(N-1)!} \times \left(\frac{\mu_2}{\left(\rho^2 - \mu_4^2\right)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2 - \mu_4^2}\right)\mathrm{d}s - \frac{\mu_4}{\left(\rho^2 - \mu_2^2\right)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2 - \mu_2^2}\right)\mathrm{d}s\right)\right]$$

$Q(N, s)$ is the complementary cumulative distribution function of $\mathrm{Erl}(N, 1)$, N is the number of amino acids and ρ is the rate of elongation of each amino acid

Application: protein elongation & dilution



Protein variance: Erlang elongation & dilution

$$\mathrm{var}_{\mathrm{DIL.}}^{(\mathrm{Erl})}(P) = \mathbb{E}[P]\left[\frac{1}{2} + \frac{2\lambda_3\mu_2}{\mu_2^2-\nu^2}\frac{\rho^{2N}}{(N-1)!} \times \left(\frac{\mu_2}{(\rho^2-\nu^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\nu^2}\right)\mathrm{d}s - \frac{\nu}{(\rho^2-\mu_2^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\mu_2^2}\right)\mathrm{d}s\right)\right]$$

$Q(N,s)$ is the complementary cumulative distribution function of $\mathrm{Erl}(N,1)$ and N is the number of amino acids

Proteolysis decay $[\mu_4 = \nu]$

$$\mathrm{var}_{\mathrm{LYSIS}}^{(\mathrm{Erl})}(P) = \mathbb{E}[P]\left[1 + \frac{2\lambda_3\mu_2}{\mu_2^2 - \nu^2}\frac{\rho^{2N}}{(N-1)!} \times \left(\frac{\mu_2}{(\rho^2-\nu^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\nu^2}\right)\mathrm{d}s - \frac{\nu}{(\rho^2-\mu_2^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\mu_2^2}\right)\mathrm{d}s\right)\right]$$

Dilution decay

$$\mathrm{var}_{\mathrm{DIL.}}^{(\mathrm{Erl})}(P) = \mathbb{E}[P]\left[\frac{1}{2} + \frac{2\lambda_3\mu_2}{\mu_2^2 - \nu^2}\frac{\rho^{2N}}{(N-1)!} \times \left(\frac{\mu_2}{(\rho^2-\nu^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\nu^2}\right)\mathrm{d}s - \frac{\nu}{(\rho^2-\mu_2^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\mu_2^2}\right)\mathrm{d}s\right)\right]$$

Proteolysis decay $[\mu_4 = \nu]$

$$\mathrm{var}^{(\mathrm{Erl})}_{\mathrm{LYSIS}}(P) = \mathbb{E}[P]\left[1 + \frac{2\lambda_3\mu_2}{\mu_2^2 - \nu^2}\frac{\rho^{2N}}{(N-1)!} \times \left(\frac{\mu_2}{(\rho^2-\nu^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\nu^2}\right)\mathrm{d}s - \frac{\nu}{(\rho^2-\mu_2^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\mu_2^2}\right)\mathrm{d}s\right)\right]$$

True for any distribution

This result is general since it holds true for any choice of distributions

Dilution decay

$$\mathrm{var}^{(\mathrm{Erl})}_{\mathrm{DIL.}}(P) = \mathbb{E}[P]\left[\frac{1}{2} + \frac{2\lambda_3\mu_2}{\mu_2^2 - \nu^2}\frac{\rho^{2N}}{(N-1)!} \times \left(\frac{\mu_2}{(\rho^2-\nu^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\nu^2}\right)\mathrm{d}s - \frac{\nu}{(\rho^2-\mu_2^2)^N}\int_{\mathbb{R}_+} s^{N-1}e^{-s}Q\left(N, \frac{s}{\rho^2-\mu_2^2}\right)\mathrm{d}s\right)\right]$$

Summary

MPPP Approach:

- more appropriate description of gene expression
- mix of different probability distributions
- mix of deterministic/stochastic processes
- analytic formulas of mean and variance
- impact of dilution/proteolysis

. . . but

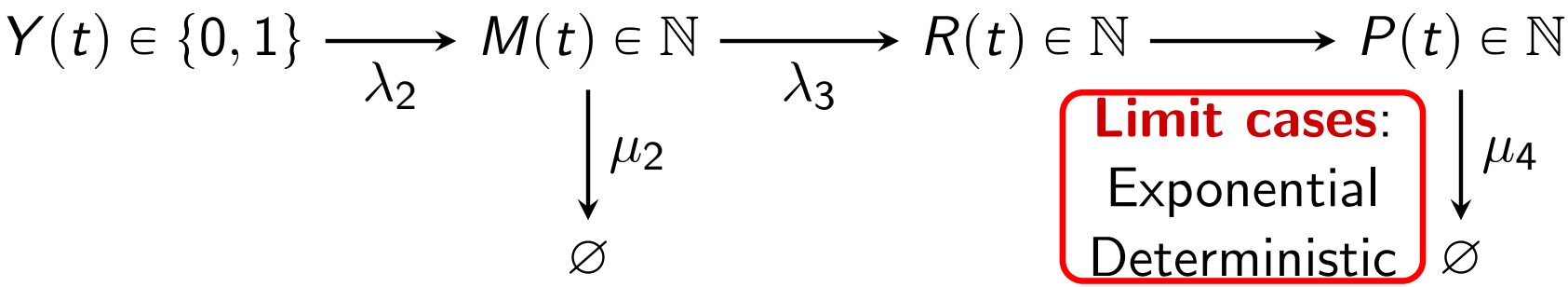
Despite the obtained formulas can be still studied numerically, we have lost a bit of intuition of the formulas derived in the classic approach, i.e.

$$\mathrm{var}(P) = \mathbb{E}[P]\left[1 + \frac{\lambda_3\mu_3(\mu_2+\mu_3+\mu_4)}{(\mu_2+\mu_3)(\mu_2+\mu_4)(\mu_3+\mu_4)}\right]$$

Idea:

study two specific cases leading to variance formulas "explicit" with respect to model parameters

Protein elongation: "limit cases"



"Limit cases":

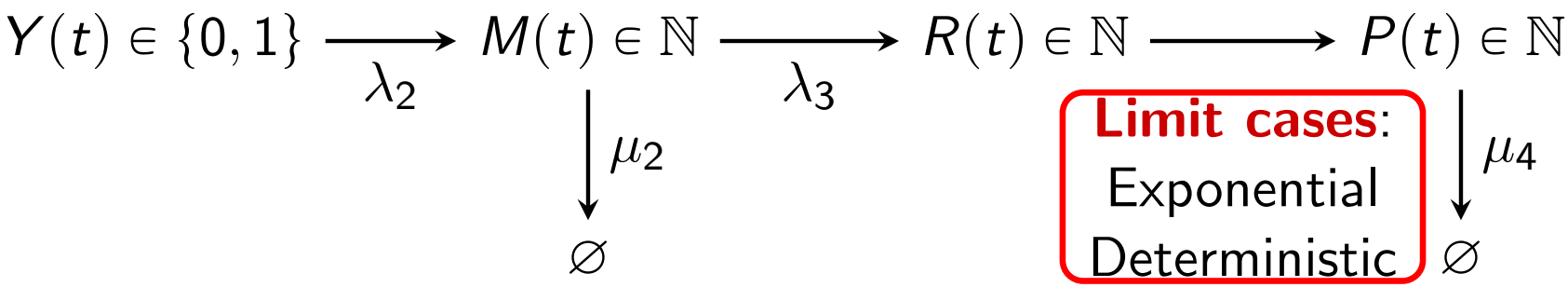
Exponential:

- classic assumption
- high fluctuations on elongation step

Deterministic:

- no fluctuations of elongation duration
- closer to reality (on average $\approx$ 400 a.a.)

Protein elongation: "limit cases"



Limit cases

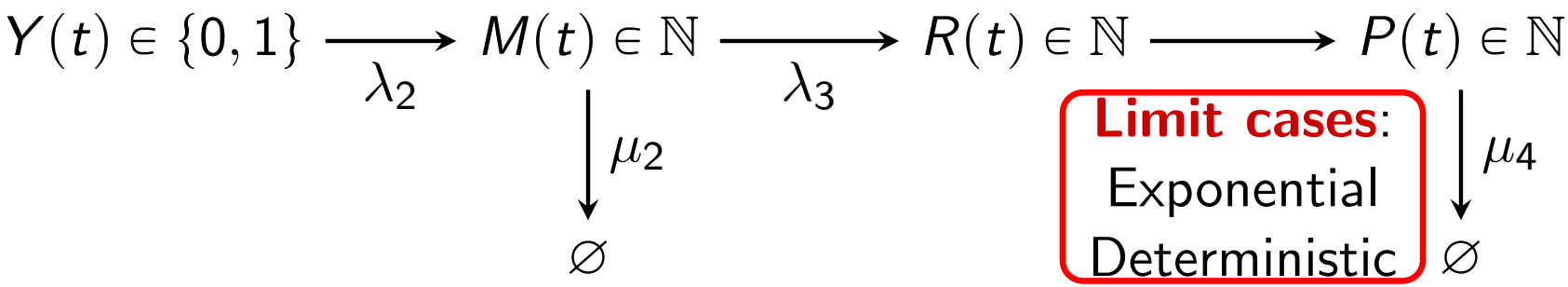
Exponential elongation:

$$\mathrm{var}_{\mathrm{E}}(P) = \mathbb{E}(P)\left[1 + \frac{\lambda_3\mu_3(\mu_2+\mu_3+\mu_4)}{(\mu_2+\mu_3)(\mu_2+\mu_4)(\mu_3+\mu_4)} + \right.$$
$$\left. + \frac{\lambda_2\lambda_3(1-\delta_+)\mu_3\mu_4^2}{(\Lambda+\mu_2)(\mu_4^2-\mu_3^2)} \times \left(\frac{\Lambda+\mu_2+\mu_3}{\mu_3(\mu_2+\mu_3)(\Lambda+\mu_3)} - \frac{\Lambda+\mu_2+\mu_4}{\mu_4(\mu_2+\mu_4)(\Lambda+\mu_4)}\right)\right],$$

Deterministic elongation:

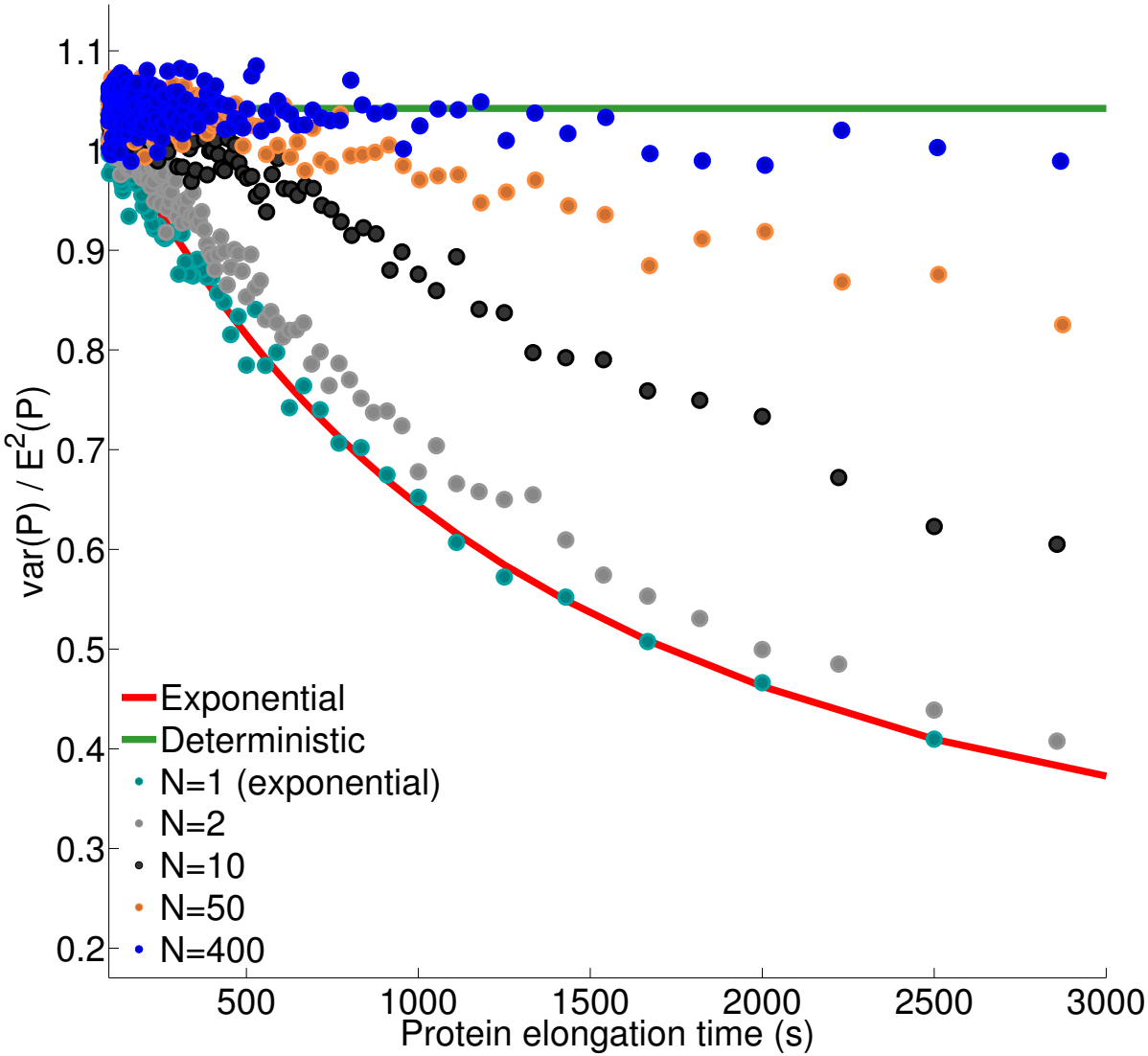
$$\mathrm{var}_{\mathrm{D}}(P) = \mathbb{E}(P)\left[1 + \frac{\lambda_3}{\mu_2+\mu_4} + \frac{\lambda_2\lambda_3(1-\delta_+)(\Lambda+\mu_2+\mu_4)}{(\mu_2+\mu_4)(\Lambda+\mu_2)(\Lambda+\mu_4)}\right]$$

Protein elongation: "limit cases"



Counter-intuitive theoretical result: $\mathrm{var}_{\mathsf{D}}(P) \geqslant \mathrm{var}_{\mathsf{E}}(P)$

Erlang elongation: simulations

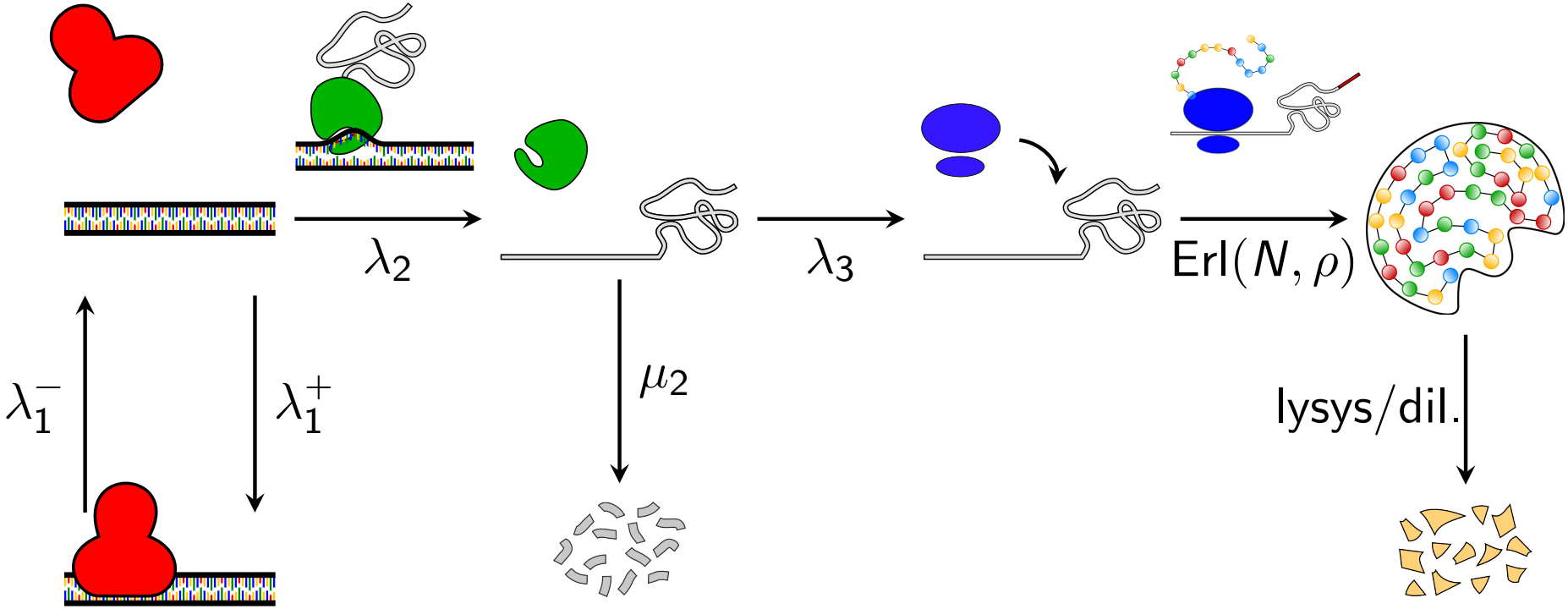


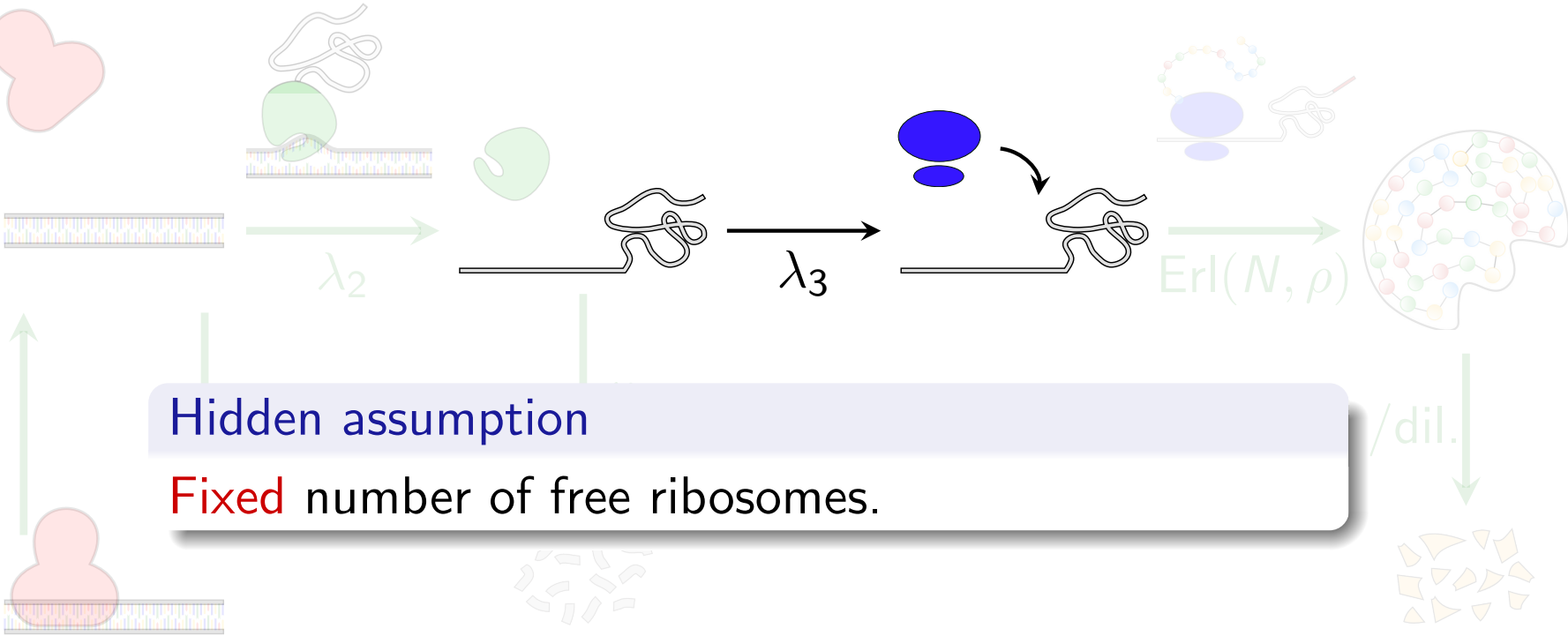
Conclusions

- analysis and proof of the correct assumptions of gene expression
- analytic formulas for any distribution
- two "explicit" formulas (estimation of impact of different choices)
- counter-intuitive: $\mathrm{var}_{\mathsf{DET}}(P) \geqslant \mathrm{var}_{\mathsf{EXP}}(P)$
- classic models underestimate protein variance
- (MPPP) appropriate math tool to describe cell stochastic processes (**Encounter + Processing**)

Part II

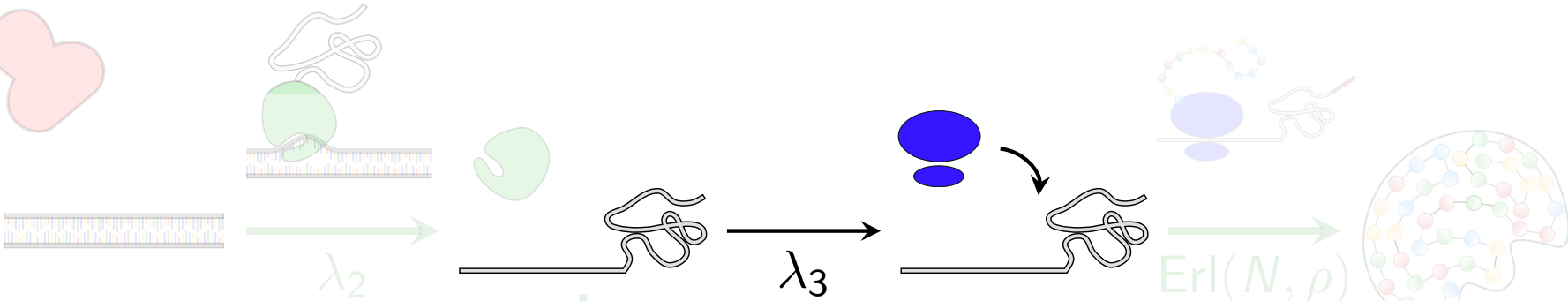
Stochastic model of Gene Expression: Multi-Protein


λ_2
λ_3
Erl(N, ρ)
λ_1^-
λ_1^+
μ_2
lysys/dil.



Hidden assumption

Fixed number of free ribosomes.



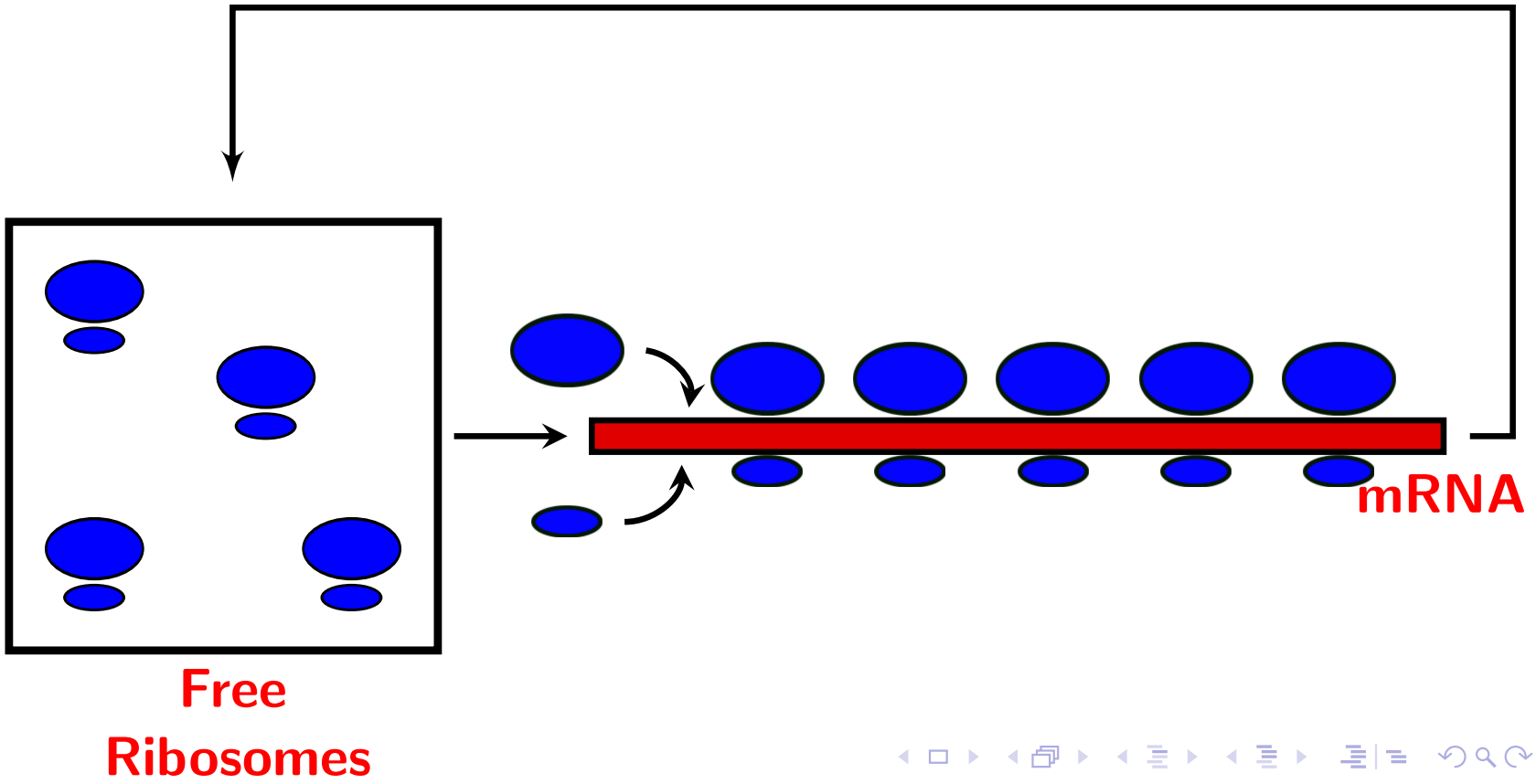
Hidden assumption

Fixed number of free ribosomes.

Morally

$\lambda_3 = \tilde{\lambda}_3 \cdot$ #Free Ribosomes

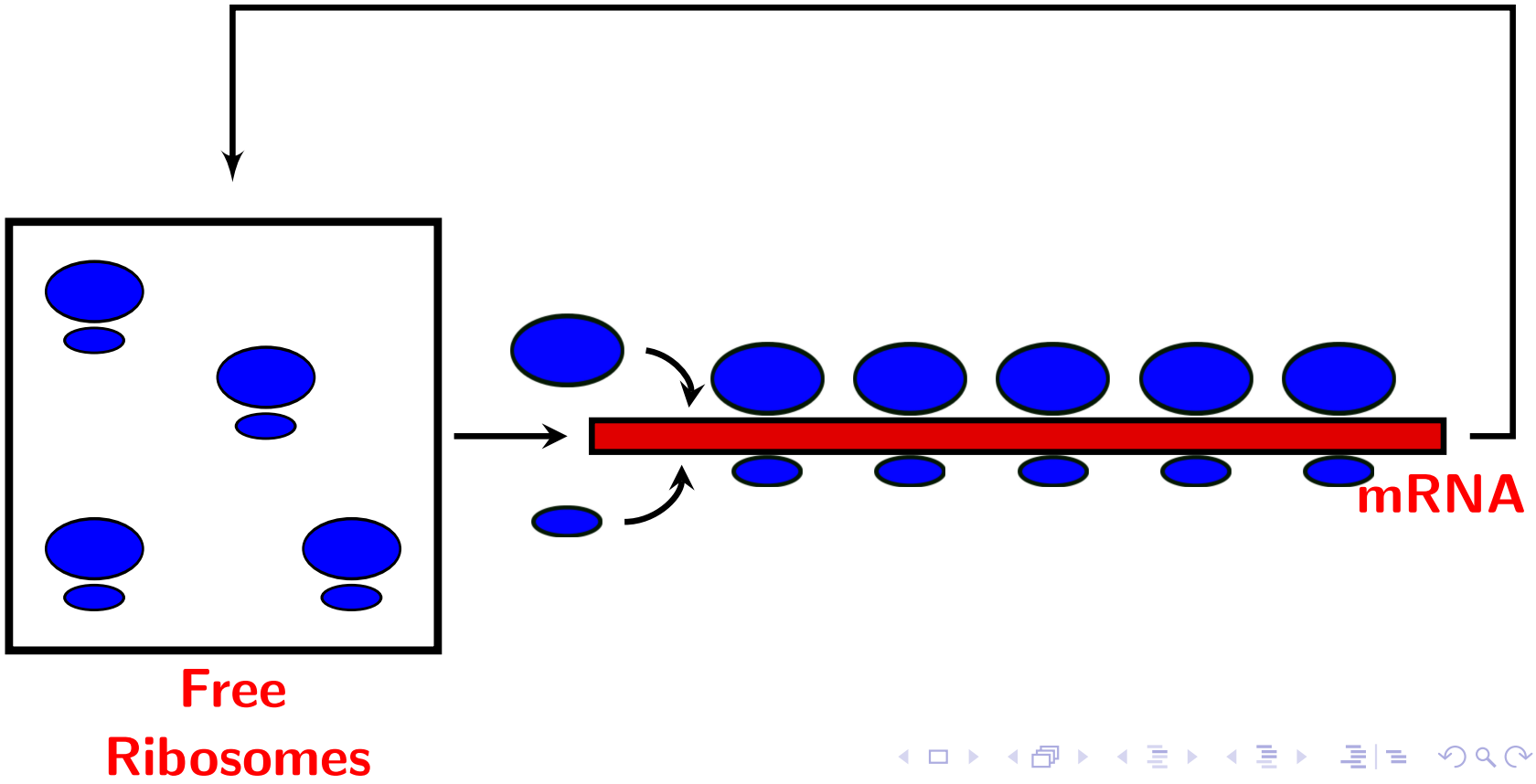
Free ribosomes



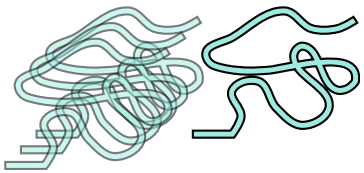
Free ribosomes

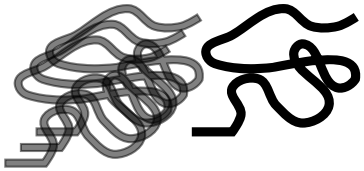
Free ribosomes

- few *free ribosomes* available
- random variable



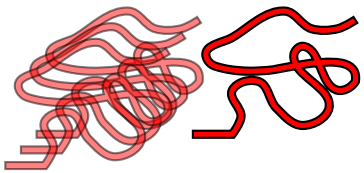
Free ribosomes



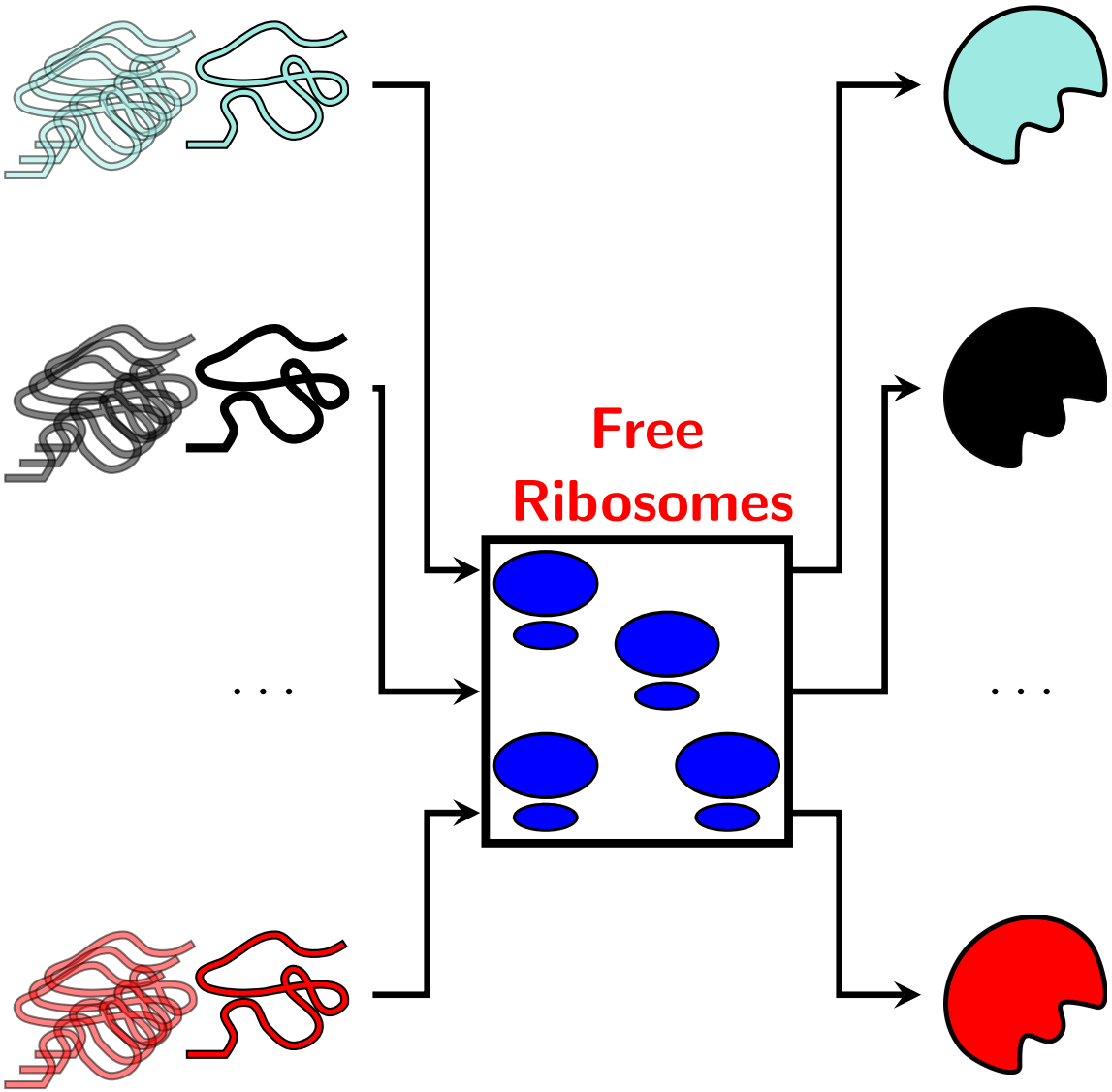


. . .

. . .



Free ribosomes



Ribosomes

- expensive
- limited number (N)

Numbers

- $N \approx 40.000$ total ribosomes
- $\geqslant 100.000$ ribosomes *required*

Objective:

Characterize the law of *free ribosomes* and the state of the system

A simple model:

- P values of concentrations of proteins
- N total (fixed) number of ribosomes
- K_p (fixed) number of mRNAs of class p
- C_p capacity of mRNAs of class p

A simple model:

- P values of concentrations of proteins
- N total (fixed) number of ribosomes
- K_p (fixed) number of mRNAs of class p
- C_p capacity of mRNAs of class p

Quantities of interest

- $X_{p,k}^N(t)$ number of ribosomes attached on the k^{th} messenger of class p (active ribosomes)
- $R(t) = N - \sum_{p,k} X_{p,k}(t)$ free ribosomes

for $p = 1, \ldots, P$, $k = 1, \ldots, K_p$

Reversible Markov process:

$$(\mathcal{X}(t)) = \{(X_{p,k}(t)) : 1 \leqslant p \leqslant P, 1 \leqslant k \leqslant K_p\}$$

Reversible Markov process:

$$(\mathcal{X}(t)) = \{(X_{p,k}(t)) : 1 \leqslant p \leqslant P, 1 \leqslant k \leqslant K_p\}$$

Proposition (Invariant distribution)

The invariant distribution π of $(\mathcal{X}(t))$ is given by

$$\pi(x) = \frac{1}{Z} \frac{1}{R(x)!} \prod_{p=1}^{P} \prod_{k=1}^{K_p} \rho_p^{x_{p,k}} \qquad x \in \mathcal{S}$$

where $\rho_p = \lambda_p/\mu_p$, $R(x)$ free ribosomes.

Reversible Markov process:

$$(\mathcal{X}(t)) = \{(X_{p,k}(t)) : 1 \leqslant p \leqslant P, 1 \leqslant k \leqslant K_p\}$$

Proposition (Invariant distribution)

The invariant distribution π of $(\mathcal{X}(t))$ is given by

$$\pi(x) = \frac{1}{Z} \frac{1}{R(x)!} \prod_{p=1}^{P} \prod_{k=1}^{K_p} \rho_p^{x_{p,k}} \qquad x \in \mathcal{S}$$

where $\rho_p = \lambda_p/\mu_p$, $R(x)$ free ribosomes.

- product form invariant distribution
- complicated normalization constant

Large scale model

Scaling

- N ribosomes (N large)
- $K_p^N \approx N\beta_p$, for $p = 1, \ldots, P$

Large scale model

Scaling

- N ribosomes (N large)
- $K_p^N \approx N\beta_p$, for $p = 1, \ldots, P$

Without constraints on ribosomes let G_p number of ribosomes attached at equilibrium on a mRNA of class (p, k)

Saturation Condition:

$$\sum_{p=1}^{P} K_p^N \mathbb{E}(G_p) > N$$

Large scale model

Scaling

- N ribosomes (N large)
- $\boxed{K_p^N \approx N\beta_p, \text{ for } p = 1, \ldots, P}$

Without constraints on ribosomes let G_p number of ribosomes attached at equilibrium on a mRNA of class (p, k)

Saturation Condition:

$$\boxed{\sum_{p=1}^{P} K_p^N \mathbb{E}(G_p) > N}$$

Large scale model

Scaling

- N ribosomes (N large)
- $K_p^N \approx N\beta_p$, for $p = 1, \ldots, P$

Without constraints on ribosomes let G_p number of ribosomes attached at equilibrium on a mRNA of class (p, k)

Saturation Condition:

$$\sum_{p=1}^{P} \beta_p \mathbb{E}(G_p) > 1$$

Large scale model

Ingredients

Limiting regime: $K_p^N \approx N\beta_p$

Saturation condition: $$\sum_{p=1}^{P} \beta_p \rho_p \frac{C_p \rho_p^{C_p+1} - (C_p+1)\rho_p^{C_p} + 1}{(1-\rho_p)(1-\rho_p^{C_p+1})} > 1$$

Large scale model

Ingredients

Limiting regime: $K_p^N \approx N\beta_p$

Saturation condition: $$\sum_{p=1}^{P} \beta_p \rho_p \frac{C_p \rho_p^{C_p+1} - (C_p+1)\rho_p^{C_p} + 1}{(1-\rho_p)(1-\rho_p^{C_p+1})} > 1$$

Theorem (Free ribosomes limiting regime)

At equilibrium, as $N \to \infty$, the nb. of free ribosomes is Poisson with parameter $\gamma(\underline{C})$, solution of equation

$$\sum_{p=1}^{P} \beta_p \rho_p \gamma \frac{C_p \rho_p^{C_p+1} \gamma^{C_p+1} - (C_p+1)\rho_p^{C_p} \gamma^{C_p} + 1}{(1-\rho_p\gamma)(1-\rho_p^{C_p+1}\gamma^{C_p+1})} = 1,$$

$\underline{C} = (C_p)$ and $\rho_p = \lambda_p/\mu_p$.

Large scale model

Ingredients

Limiting regime: $K_p^N \approx N\beta_p$

Saturation condition: $$\sum_{p=1}^{P} \beta_p \rho_p \frac{C_p \rho_p^{C_p+1} - (C_p+1)\rho_p^{C_p} + 1}{(1-\rho_p)(1-\rho_p^{C_p+1})} > 1$$

Theorem (Active ribosomes limiting regime)

At equilibrium, as $N \to \infty$,

- *the r.v. $(X_{p,k}^N)$ are independent*
- *the system $(X_{p,k}^N)$ behaves as if there were no constraints: λ_p is now replaced by $\lambda_p \gamma$*

Large scale model

Theorem (Free ribosomes limiting regime)

At equilibrium, as $N \to \infty$, the nb. of free ribosomes is Poisson with parameter $\gamma(\underline{C})$, solution y of equation

$$\sum_{p=1}^{P} \beta_p \rho_p \gamma \frac{C_p \rho_p^{C_p+1} \gamma^{C_p+1} - (C_p+1)\rho_p^{C_p} \gamma^{C_p} + 1}{(1-\rho_p\gamma)(1-\rho_p^{C_p+1}\gamma^{C_p+1})} = 1.$$

Proof.

- Mean-Field Limit Setting
- Change of Probability Measure
- Local Central Limit Theorem

Conclusions

Multi-protein model:

- *competition* for *common resources* (*free ribosomes*)
- Markov-model with limited number of ribosomes and limited mRNA capacity

Results:

- *free ribosomes* are Poisson distributed (asymptotic regime)
- *active ribosomes* are independent in the asymptotic regime
- interactions described by fixed-point equation

Perspectives

- model with varying numbers of messengers
- *polymerases*: similar approach, but... some peculiarity:
  - several types of polymerases (different *gamma factors*)
  - stand-by polymerases on DNA (*buffering?*)
  - polymerase diffusion transport mechanism (not only 3D diffusion?)
- extrinsic noise: better understanding of the role of ribosomes/polymerases on the protein production

Thanks.