

Stochastic Gene Expression in Prokaryotes: A Point Process Approach

Emanuele LEONCINI

INRIA Rocquencourt - INRA Jouy-en-Josas

ASMDA – Mataró

June 28th 2013



Central role of protein production

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

Central role of protein production

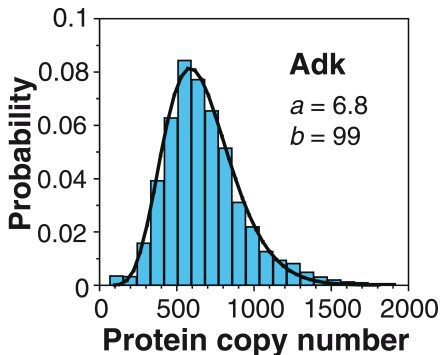
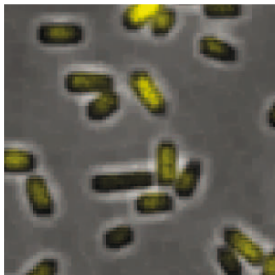
- Proteins are the core of biologic processes: *enzymes*, DNA replication machinery, ...
- ~ 50% of the bacteria dry weight
- ~ 3.5 millions of proteins in each cell
- ~ 2000 types of proteins produced at any time at any growth condition (volume growth)
- proteins ranging from few dozens up to 10^5

A highly consuming process:

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

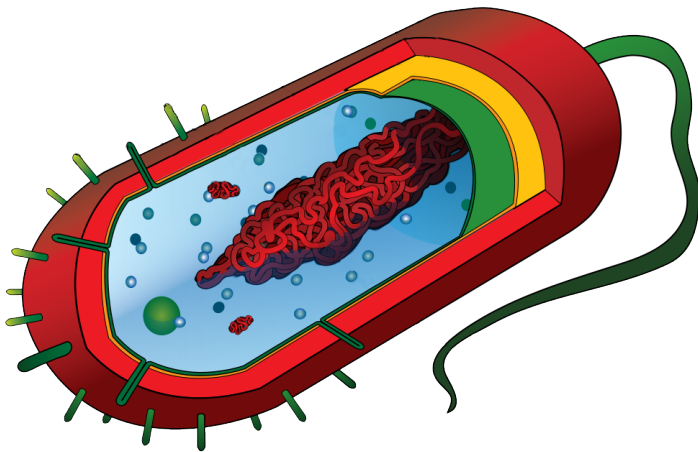
Stochasticity in protein production: experimental viewpoint

Adk cytoplasm protein¹



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

Stochasticity in protein production: structural



Stochasticity in bacteria

Sources of stochasticity:

- bacterial *cytoplasm*: disordered medium
- main cellular motility mechanism: diffusion in a stiff medium
- most cellular processes require the encounter of macromolecules (*Poisson* process)

Stochasticity in bacteria

Sources of stochasticity:

- bacterial *cytoplasm*: disordered medium
- main cellular motility mechanism: diffusion in a stiff medium
- most cellular processes require the encounter of macromolecules (*Poisson* process)

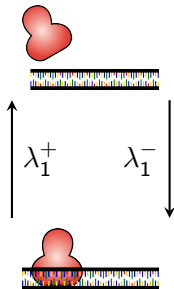
Protein production: inherently stochastic process.

Stochastic model of Gene Expression in Prokaryotes

4-Step model: activation

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

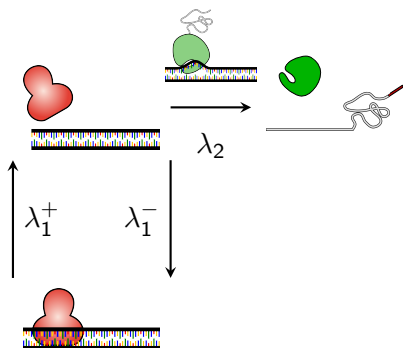
Gene status



4-Step model: transcription

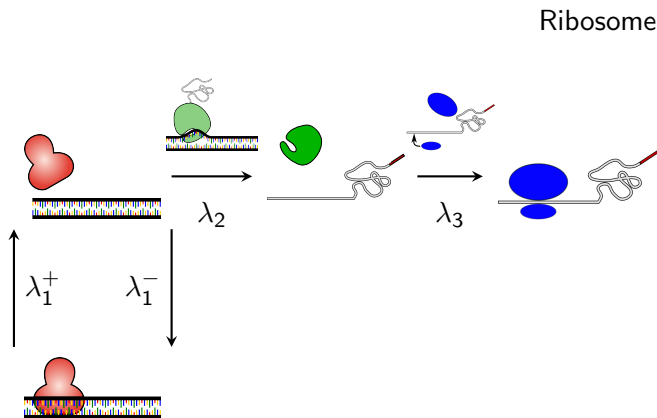
$$Y(t) \in \{0, 1\} \longrightarrow M(t) \in \mathbb{N}$$

Messenger



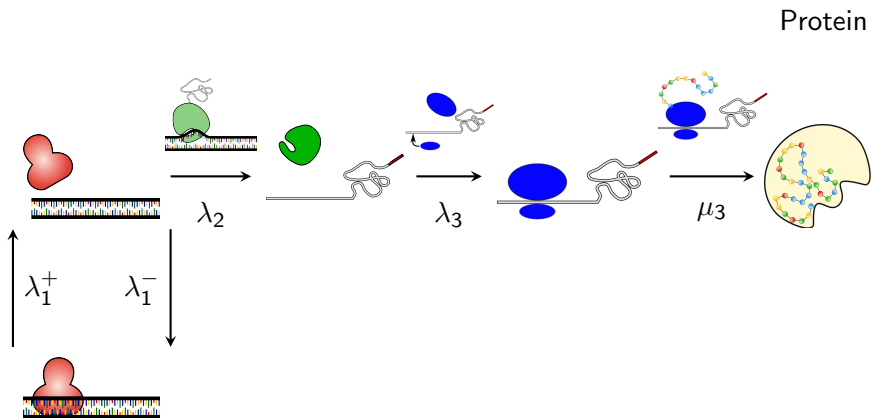
4-Step model: translation initiation

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

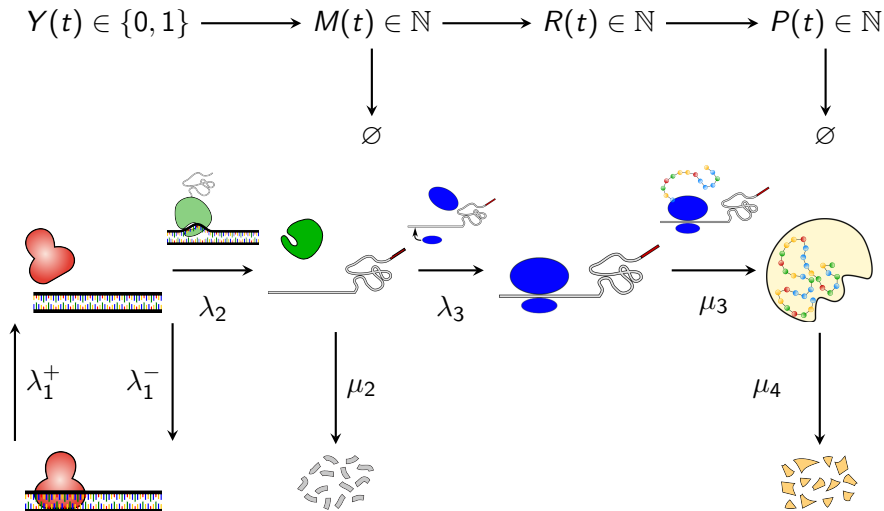


4-Step model: translation completion

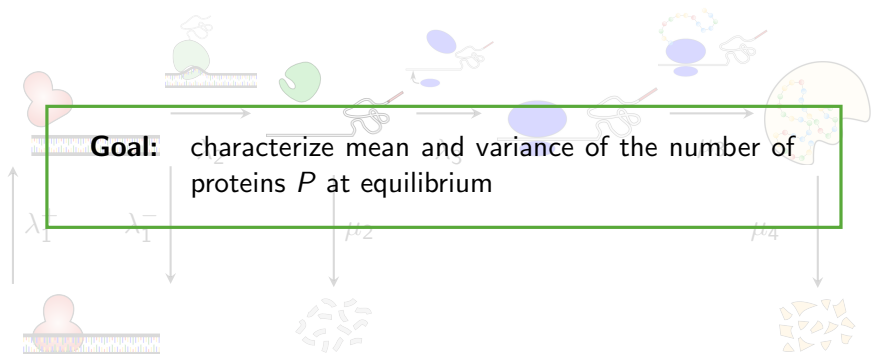
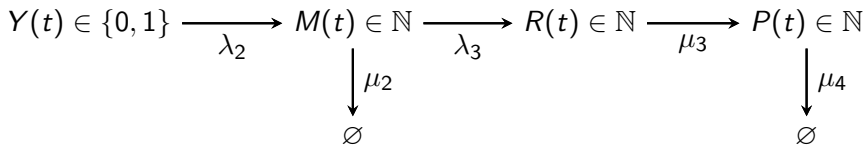
$$Y(t) \in \{0, 1\} \longrightarrow M(t) \in \mathbb{N} \longrightarrow R(t) \in \mathbb{N} \longrightarrow P(t) \in \mathbb{N}$$



4-Step model



4-Step model



Classic assumptions

Properties

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

Classic assumptions

Properties

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

Tools

- Markov processes
- Fokker-Plank equations



explicit analytic formulas of mean and variance as function of the main parameters

Results:

- models used by biologists
- quantitative characterisation of protein fluctuations

$$\mathbb{E}[P] = \frac{\lambda_2 \lambda_3}{\mu_2 \mu_4}$$

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

if $\delta_+ \stackrel{\text{def}}{=} \frac{\lambda_1^+}{\lambda_1^+ + \lambda_1^-} = 1$ (active gene)

Classic approach:

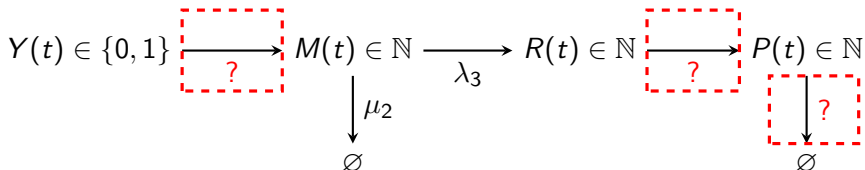
Exponential assumption:

Not each described process has an exponentially distributed duration.

Classic approach:

Exponential assumption:

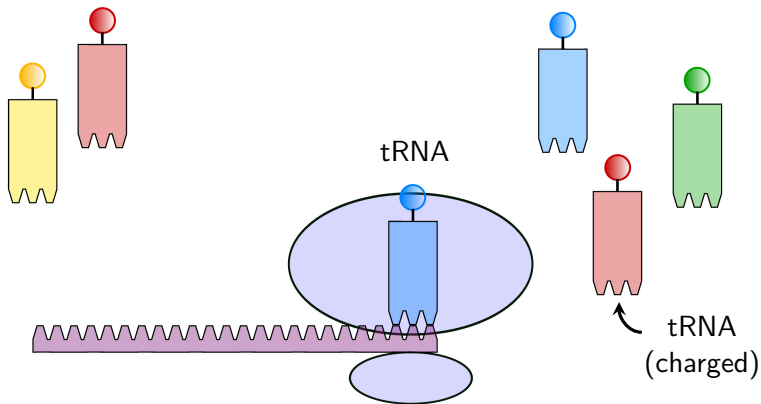
Not each described process has an exponentially distributed duration.



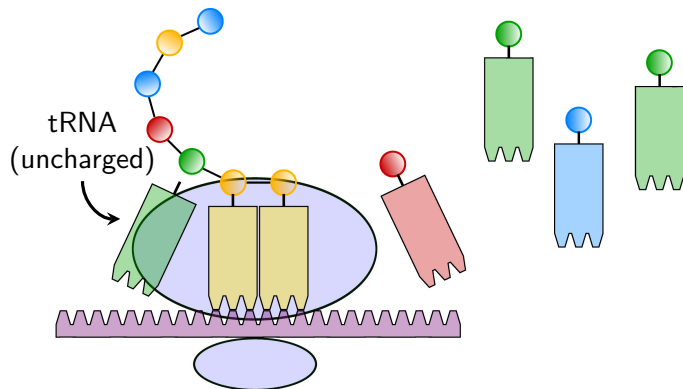
The duration of the following processes is not exponential

- *protein elongation*
- *mRNA elongation*
- *deterministic protein dilution (vs. "classic" stochastic proteolysis)*

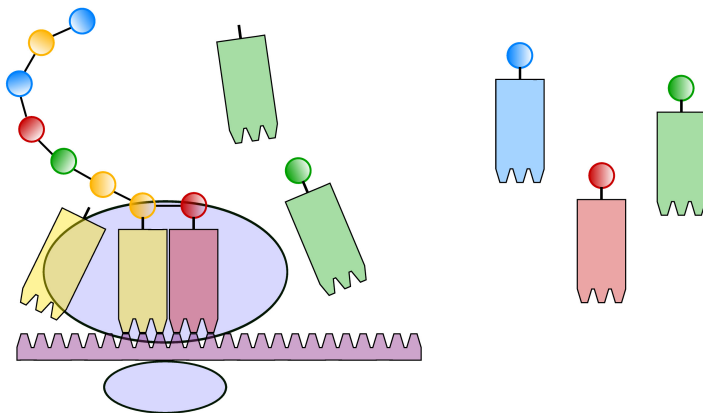
Protein chain elongation



Protein chain elongation

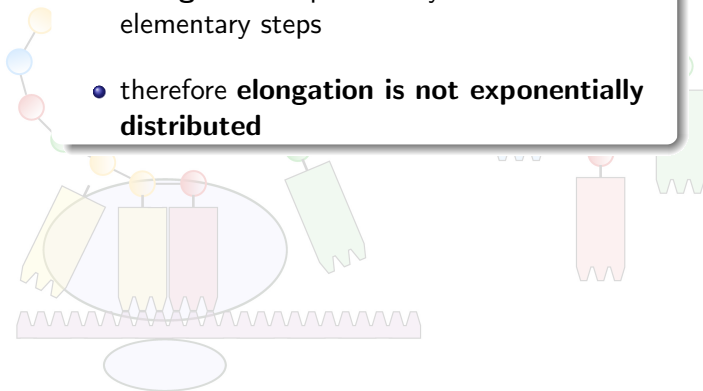


Protein chain elongation



Protein chain elongation

- **Elongation:** exponentially distributed elementary steps
- therefore **elongation is not exponentially distributed**



Protein chain elongation

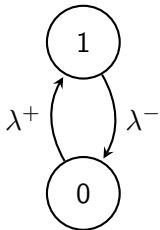
- **Elongation:** exponentially distributed elementary steps
- therefore **elongation is not exponentially distributed**

Large number of steps ($N \approx 400$ a.a.)

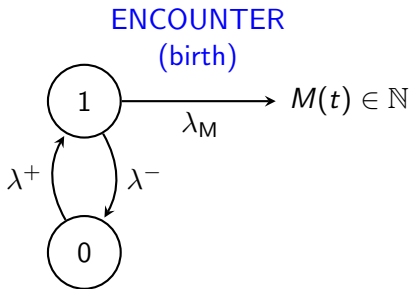
→ elongation time described by normal random variable with $\text{var} = (\text{mean elongation time})/\sqrt{N}$

Marked Poisson Point Process (MPPP): new description of gene expression

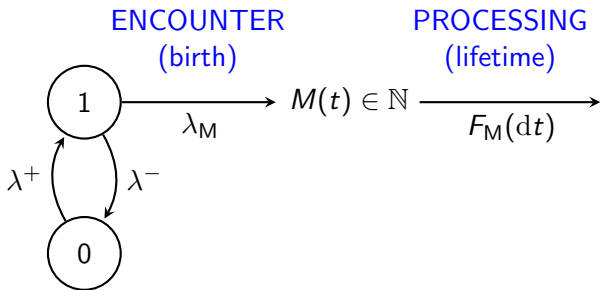
Explanatory model



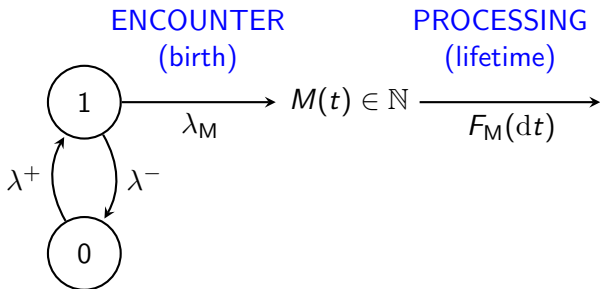
Explanatory model



Explanatory model



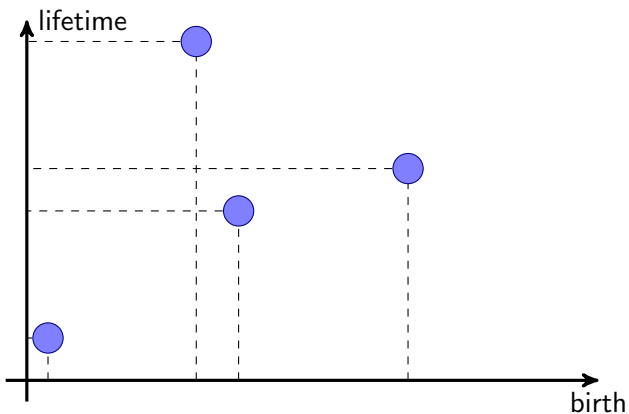
Explanatory model



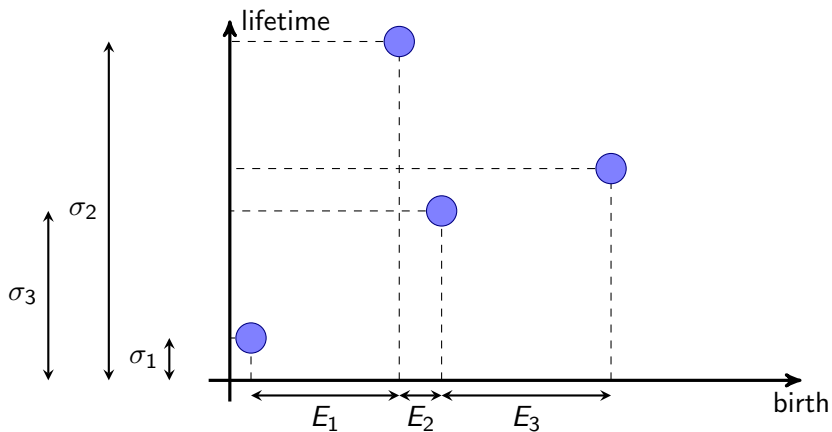
Assumptions:

- $Y(t) \in \{0, 1\}$ exponentially distributed switches with rates λ^+ , λ^-
- births (s_n) follows a **Poisson process** of parameter λ_M
- time (σ_n) to process the $M(t)$ has (general) distribution $F_M(dt)$ (**mark**)

Explanatory model



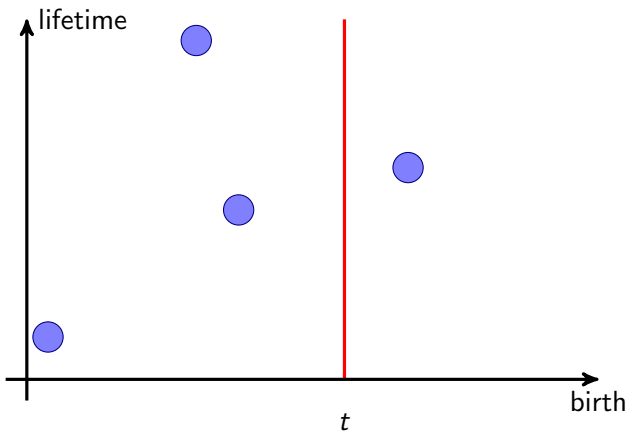
Explanatory model



E_i exponential random variables of parameter λ_M .

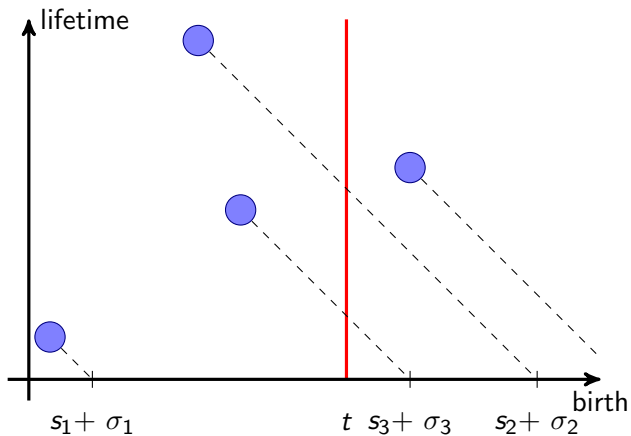
Explanatory model

How many $M(t)$ alive at time t ?



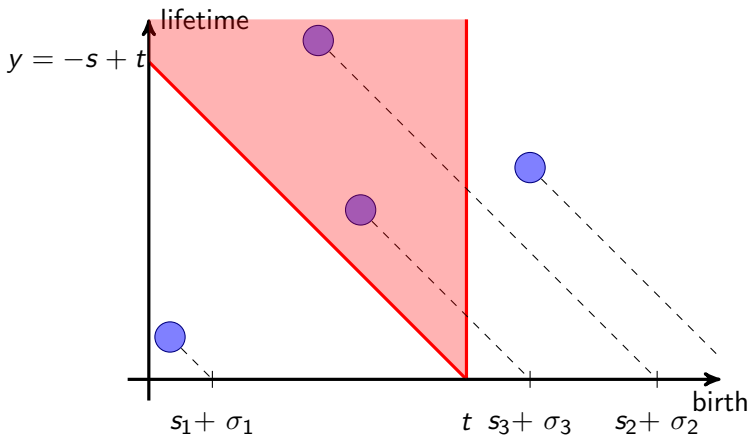
Explanatory model

How many $M(t)$ alive at time t ?



Explanatory model

How many $M(t)$ alive at time t ?



Explanatory model: general results

At equilibrium:

$$M = \mathcal{N}_M (\mathbb{1}_{\{u \leq 0 \leq u+v\}}) = \int_{\mathbb{R} \times \mathbb{R}_+} \mathbb{1}_{\{u \leq 0 \leq u+v\}} \mathcal{N}_{\lambda_M} (du, dv)$$

Explanatory model: general results

At equilibrium:

$$M = \mathcal{N}_M (\mathbb{1}_{\{u \leq 0 \leq u+v\}}) = \int_{\mathbb{R} \times \mathbb{R}_+} \mathbb{1}_{\{u \leq 0 \leq u+v\}} \mathcal{N}_{\lambda_M}(du, dv)$$

The Laplace transform of a Marked Poisson point process is

$$\mathcal{L}_{\mathcal{N}_{\lambda_M}}(f) = \mathbb{E} \left[e^{-\mathcal{N}_M(f)} \right] = \exp \left(- \int \left(1 - e^{-f(x,y)} \right) \lambda_M dx F_M(dy) \right)$$

where $\mathcal{N}_{\lambda_M}(f) = \sum_n f(s_n, \sigma_n)$.

Explanatory model: general results

At equilibrium:

$$M = \mathcal{N}_M (\mathbb{1}_{\{u \leq 0 \leq u+v\}}) = \int_{\mathbb{R} \times \mathbb{R}_+} \mathbb{1}_{\{u \leq 0 \leq u+v\}} \mathcal{N}_{\lambda_M}(du, dv)$$

Proposition

$$\mathbb{E}[M] = \delta_+ \lambda_M \mathbb{E}[\sigma]$$

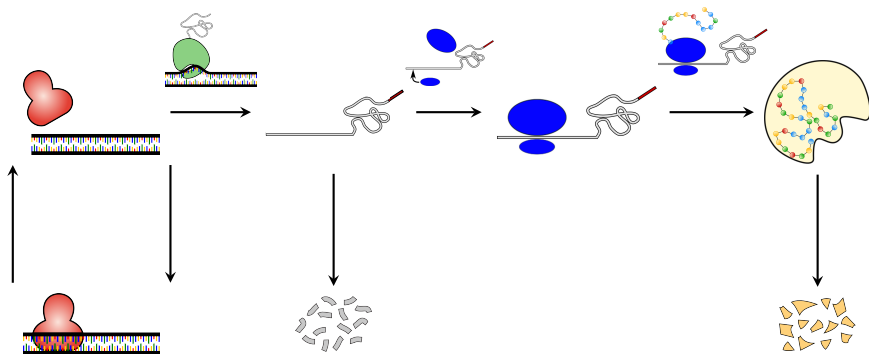
$$\text{var}(M) = \mathbb{E}[M] + 2\lambda_M^2 \delta_+ (1 - \delta_+).$$

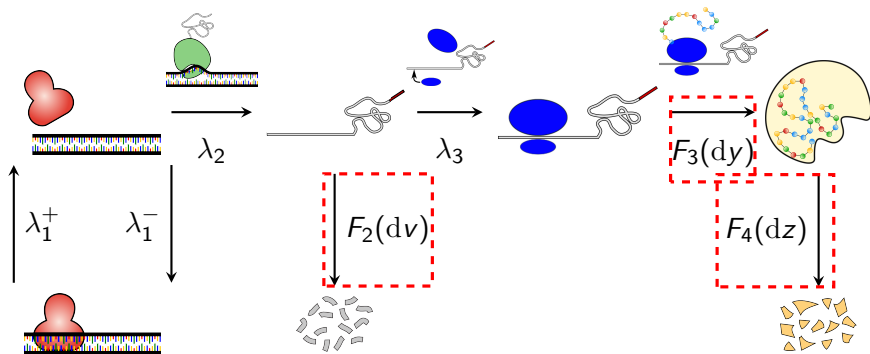
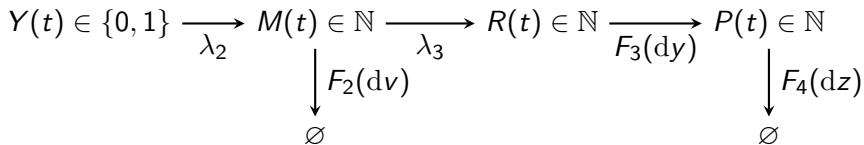
$$\cdot \int_0^{+\infty} \int_{-u}^0 e^{-(\lambda^+ + \lambda^-)v} (1 - F_M(u))(1 - F_M(u+v)) du dv$$

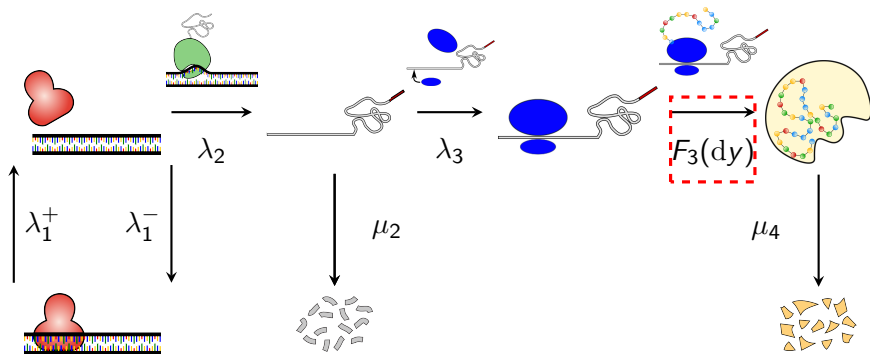
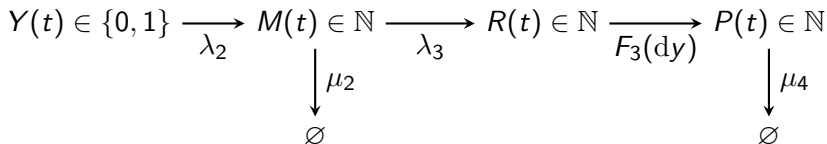
where $\delta_+ = \frac{\lambda^+}{\lambda^+ + \lambda^-}$

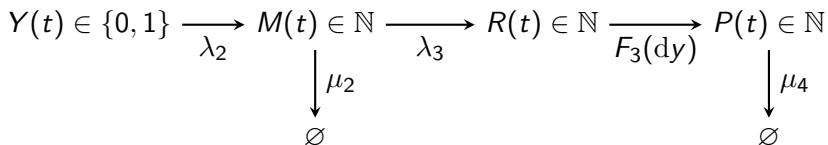
MPPP Applications & Model Extensions

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

 \downarrow
 \emptyset
 \downarrow
 \emptyset






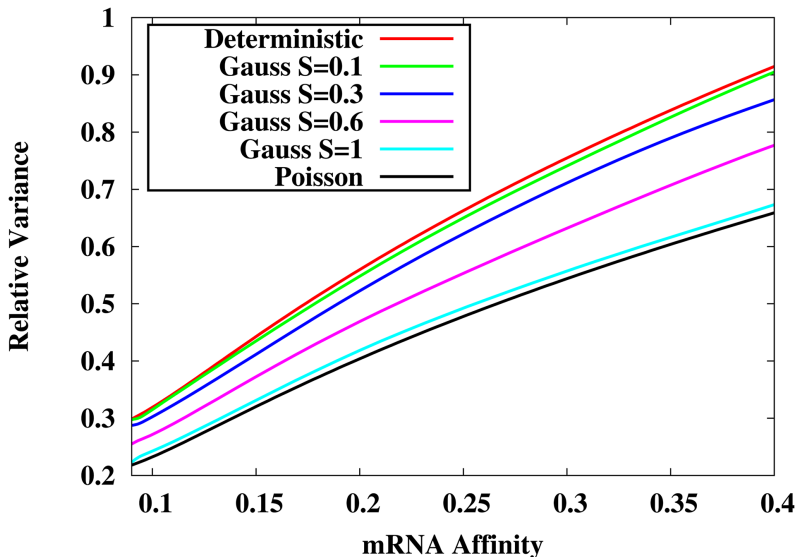


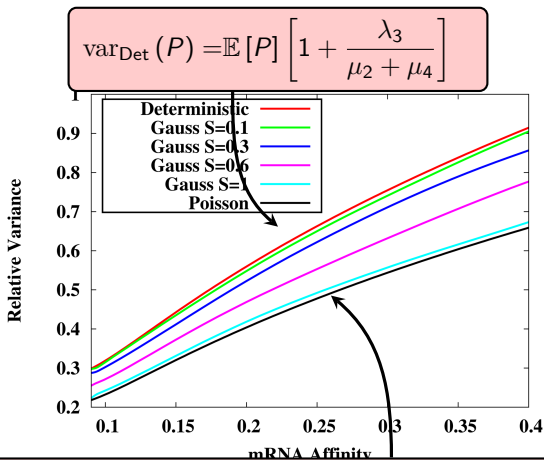
Choices for $F_3(dy)$

- Exponential \rightarrow explicit close formula depending on model parameters
- Normal \rightarrow analytic formula
- Deterministic \rightarrow explicit close formula depending on model parameters (limit case)



Deterministic vs Exponential





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

Conclusions

- (MPPP) appropriate math tool to describe cell stochastic processes
(**Encounter + Processing**)

Conclusions

- (MPPP) appropriate math tool to describe cell stochastic processes (**Encounter + Processing**)
- analysis and proof of the correct assumption for protein degradation (*proteolysis*/volume dilution)

Conclusions

- (MPPP) appropriate math tool to describe cell stochastic processes (**Encounter + Processing**)
- analysis and proof of the correct assumption for protein degradation (*proteolysis*/volume dilution)
- analytic form formula for any distribution
explicit formula depending on the model parameters for specific and interesting distributions

Conclusions

- (MPPP) appropriate math tool to describe cell stochastic processes (**Encounter + Processing**)
- analysis and proof of the correct assumption for protein degradation (*proteolysis*/volume dilution)
- analytic form formula for any distribution
explicit formula depending on the model parameters for specific and interesting distributions
- deterministic protein elongation might be an upper-bound for protein variance

Conclusions

- (MPPP) appropriate math tool to describe cell stochastic processes (**Encounter + Processing**)
- analysis and proof of the correct assumption for protein degradation (*proteolysis*/volume dilution)
- analytic form formula for any distribution
explicit formula depending on the model parameters for specific and interesting distributions
- deterministic protein elongation might be an upper-bound for protein variance
- counter-intuitive: $\text{var}_{\text{DET}}(P) > \text{var}_{\text{EXP}}(P)$

V. Fromion, E. Leoncini, and P. Robert. “Stochastic Gene Expression in Cells: A Point Process Approach”. In: *SIAM Journal on Applied Mathematics* 73.1 (2013), pp. 195–211

Thanks.