Marked Poisson Point Process approach: a non-Markovian description of cell stochastic processes

Emanuele LEONCINI

INRIA Rocquencourt - INRA Jouy-en-Josas

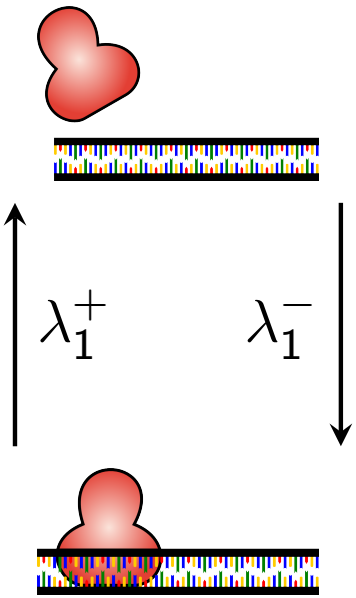
"In honour of Michael Mackey's 70th birthday"

June 3rd 2013

Stochastic model of Gene Expression in Prokaryotes

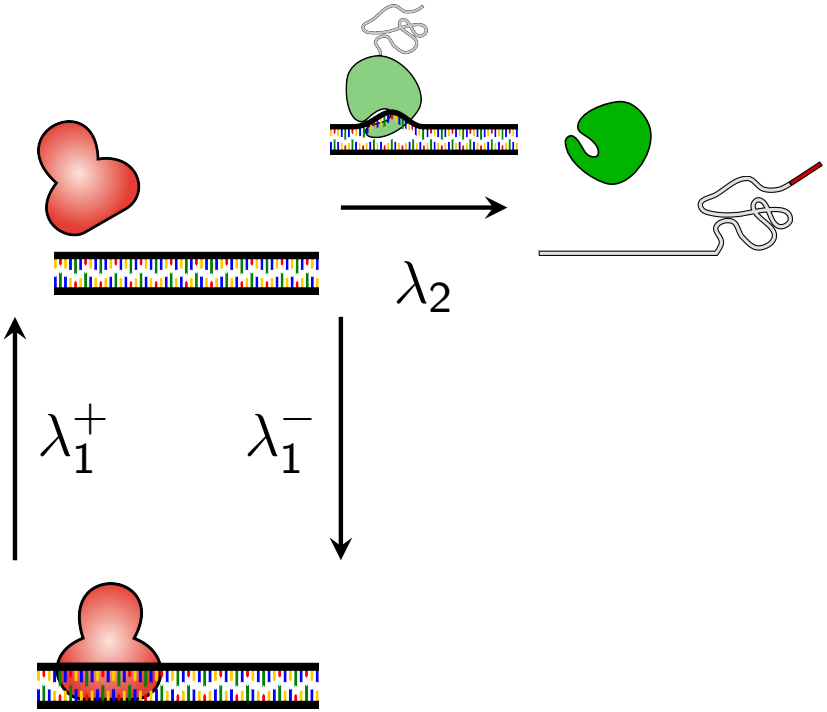
4-Step model: activation

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



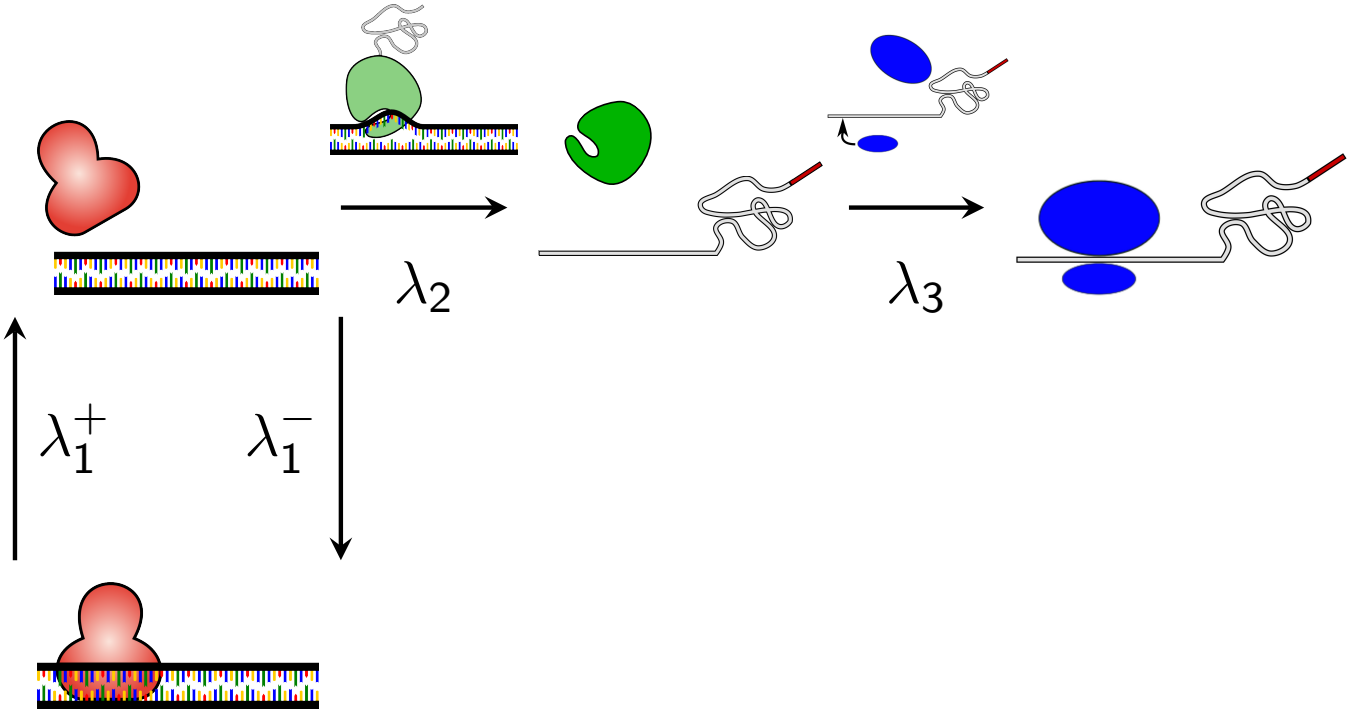
4-Step model: transcription

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



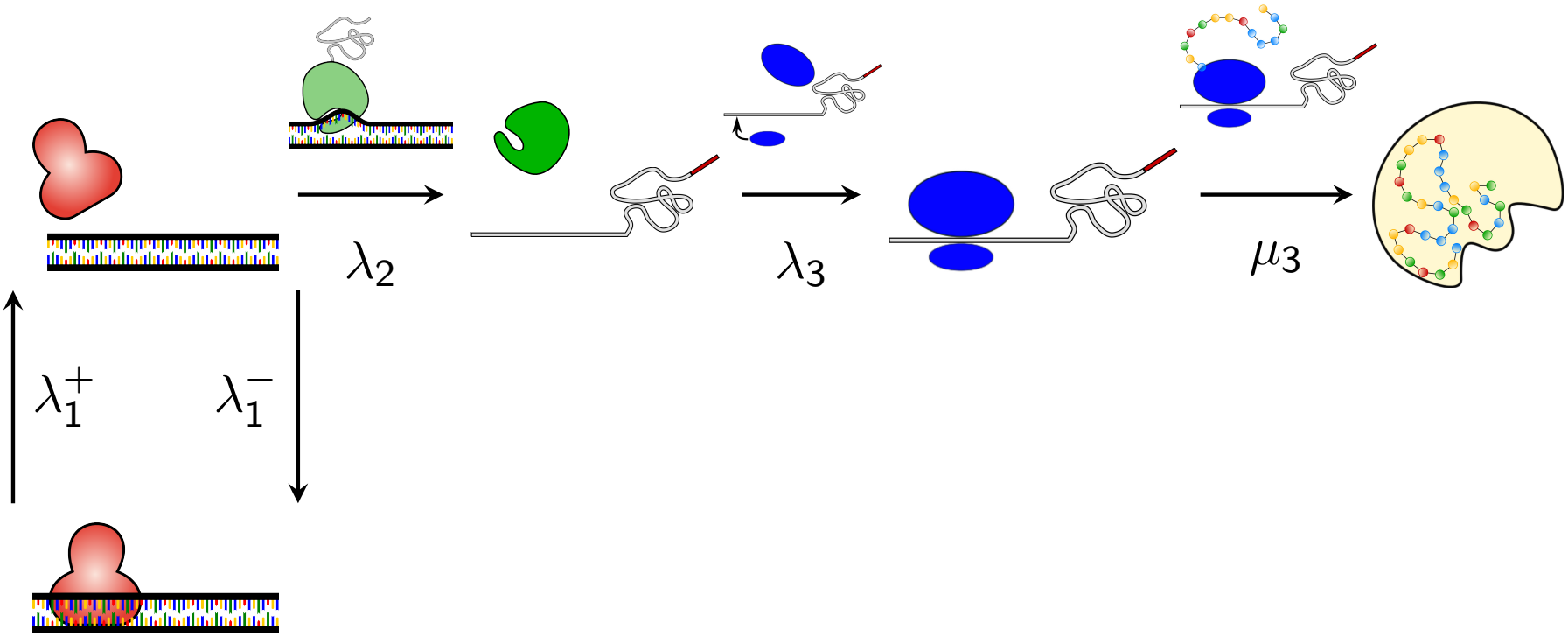
4-Step model: translation initiation

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



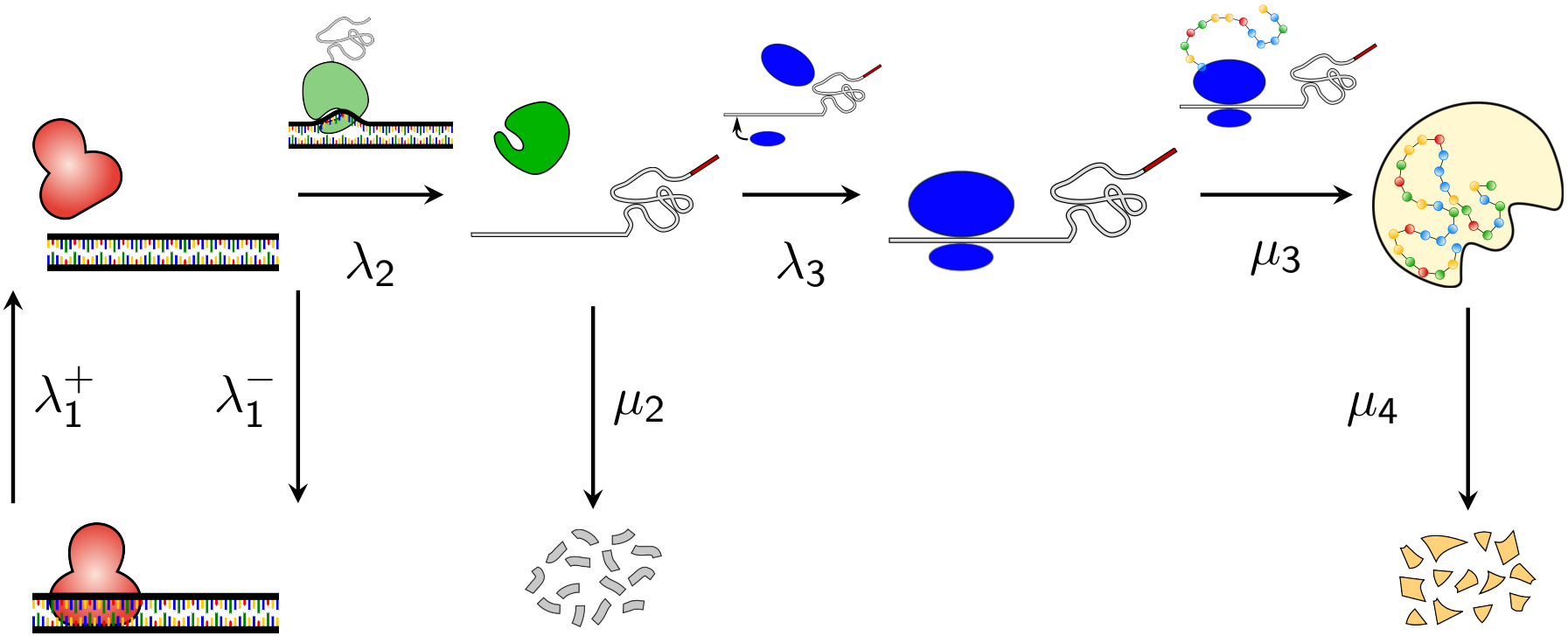
4-Step model: translation completion

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



4-Step model

$$
\begin{tikzcd}
Y(t) \in \{0,1\} \arrow[r] & M(t) \in \mathbb{N} \arrow[r] \arrow[d] & R(t) \in \mathbb{N} \arrow[r] & P(t) \in \mathbb{N} \arrow[d] \\
 & \varnothing & & \varnothing
\end{tikzcd}
$$



4-Step model

$$
\begin{tikzcd}
Y(t) \in \{0,1\} \arrow[r, "\lambda_2"'] & M(t) \in \mathbb{N} \arrow[r, "\lambda_3"'] \arrow[d, "\mu_2"] & R(t) \in \mathbb{N} \arrow[r, "\mu_3"'] & P(t) \in \mathbb{N} \arrow[d, "\mu_4"] \\
& \varnothing & & \varnothing
\end{tikzcd}
$$

Goal: characterize mean and variance of the number of proteins P at equilibrium

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

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

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

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:

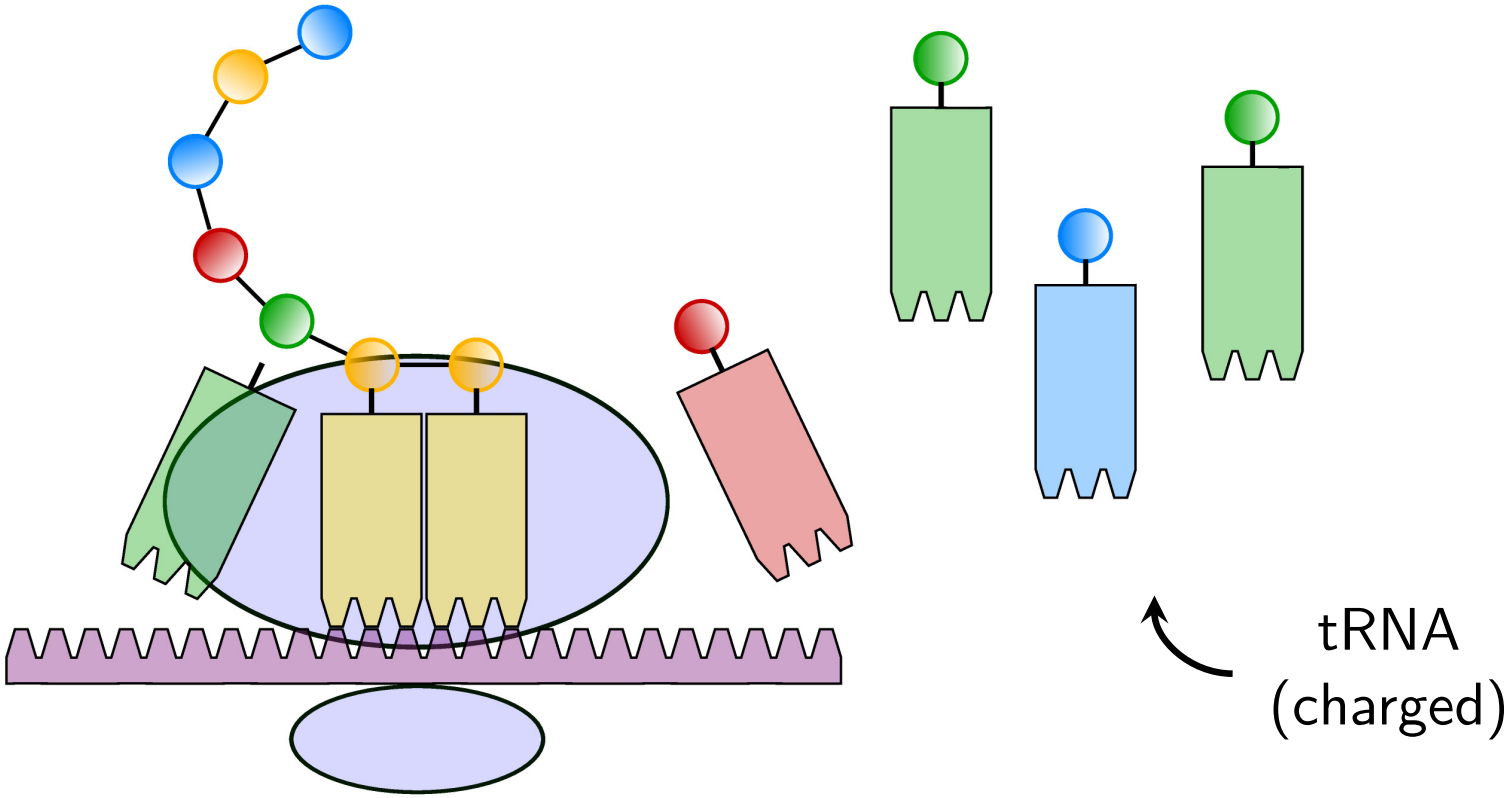
Not each described process has an exponentially distributed duration.

$$
\begin{tikzcd}
Y(t) \in \{0,1\} \arrow[r, "?"'] & M(t) \in \mathbb{N} \arrow[r, "\lambda_3"'] \arrow[d, "\mu_2"] & R(t) \in \mathbb{N} \arrow[r, "?"'] & P(t) \in \mathbb{N} \arrow[d, "?"] \\
& \varnothing & & \varnothing
\end{tikzcd}
$$

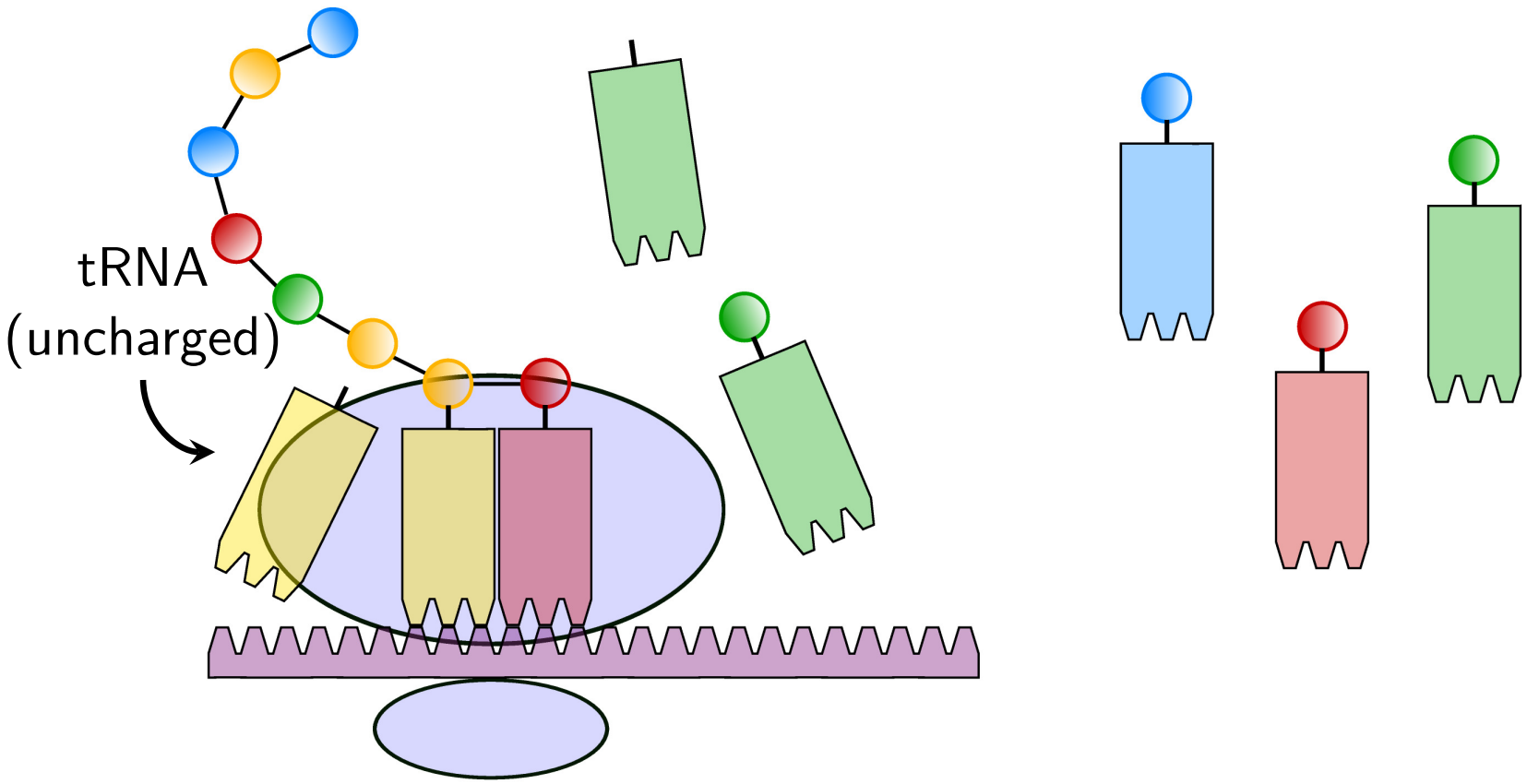
The duration of the following processes is not exponential

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

Protein chain elongation

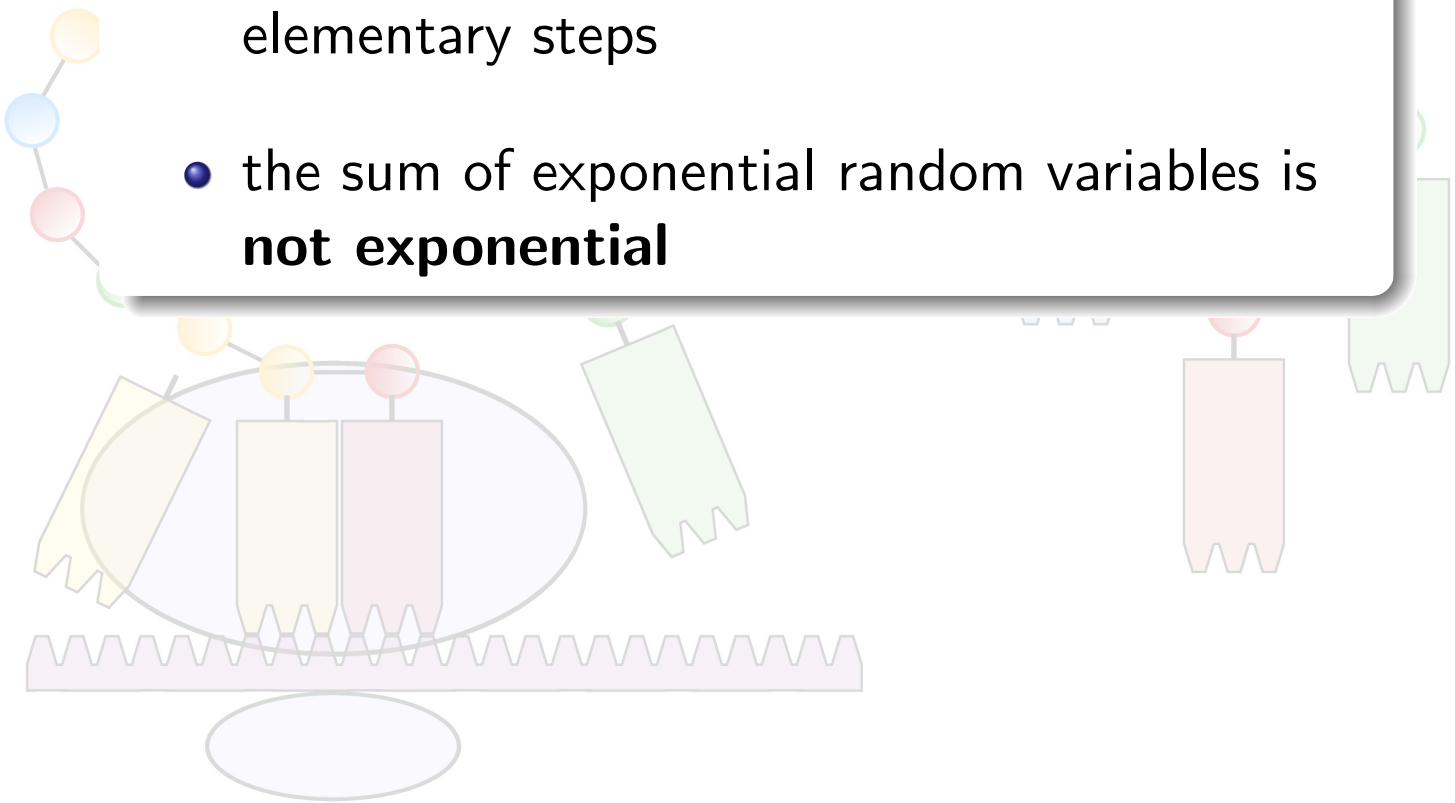


Protein chain elongation



Protein chain elongation

- **Elongation:** exponentially distributed elementary steps
- the sum of exponential random variables is **not exponential**



Protein chain elongation

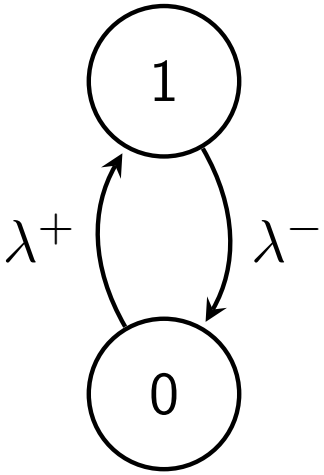
- **Elongation:** exponentially distributed elementary steps
- the sum of exponential random variables is **not exponential**

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

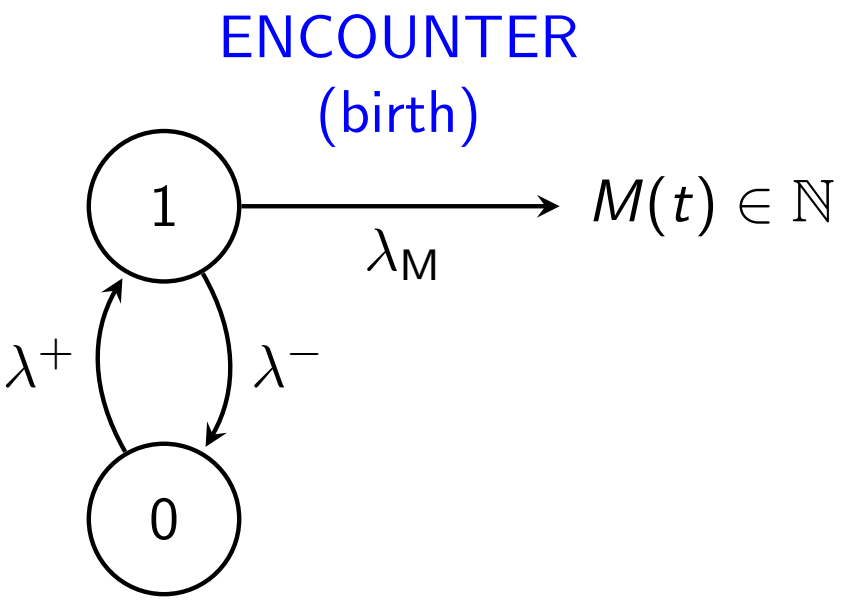
➡ elongation time described by normal random variable with var $=$ (mean elongation time)$/\sqrt{N}$

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

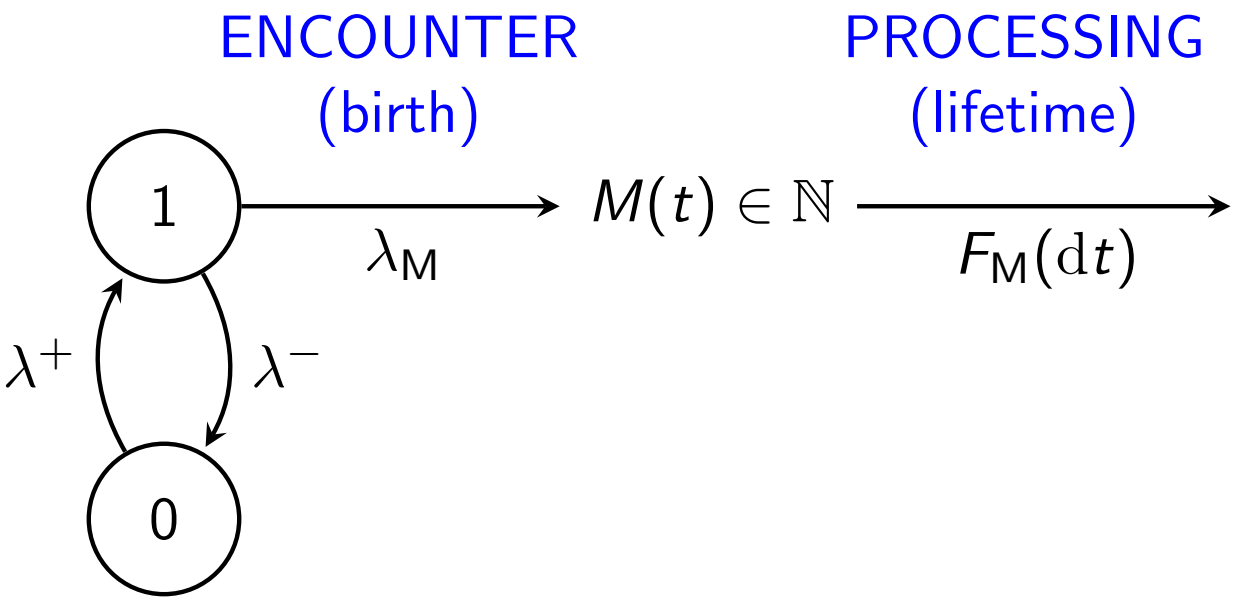
Explanatory model



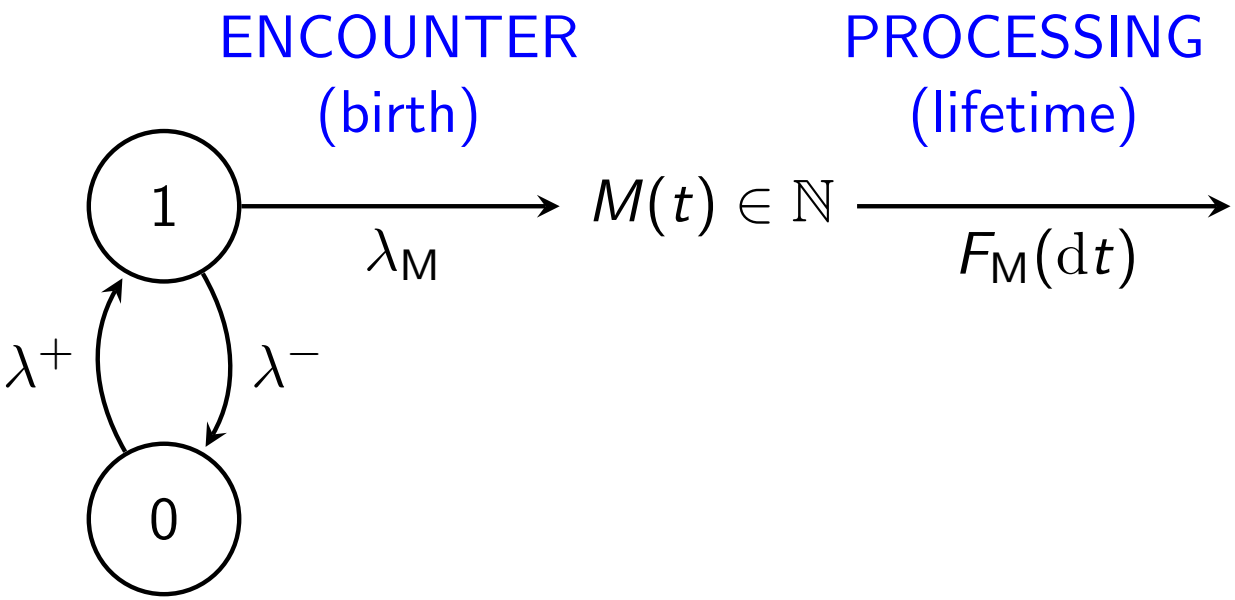
Explanatory model



Explanatory model



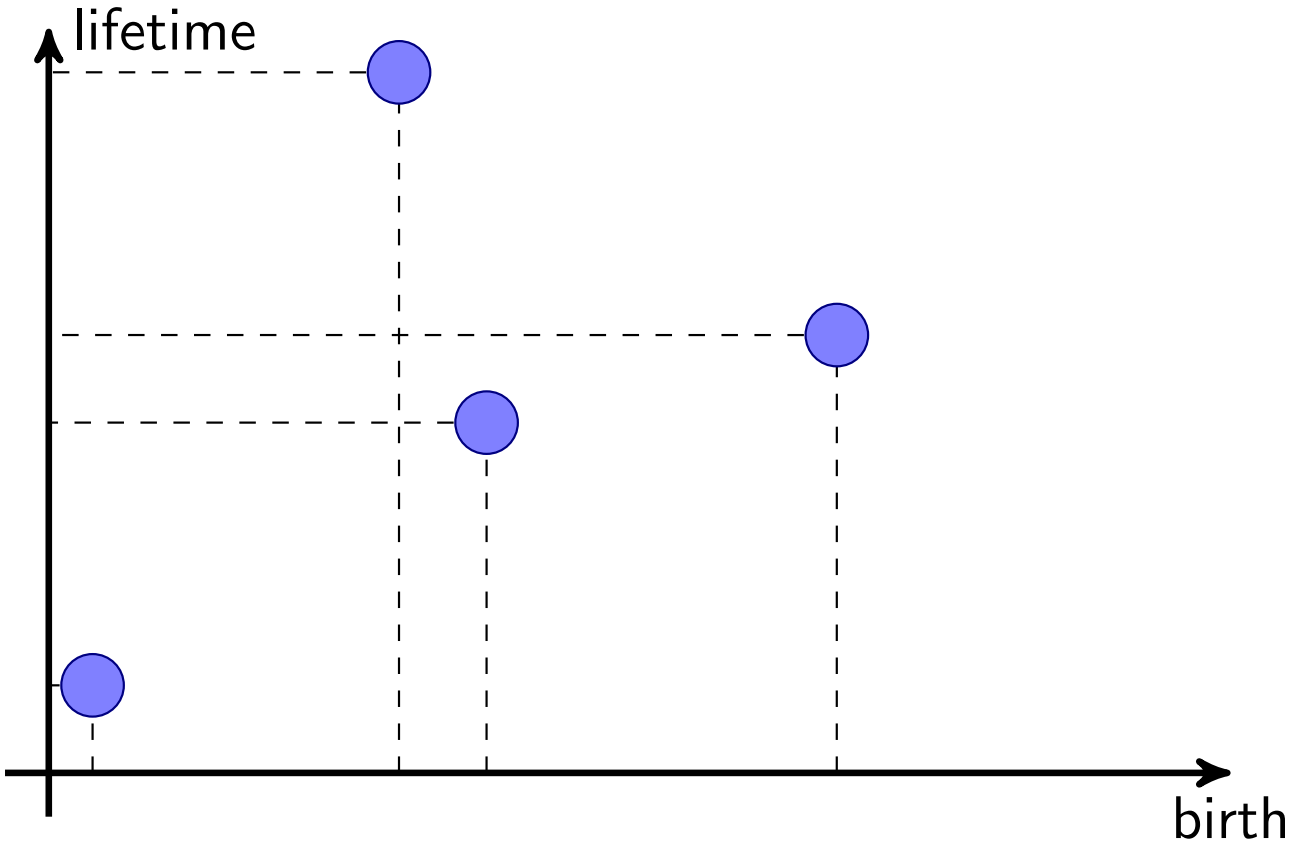
Explanatory model



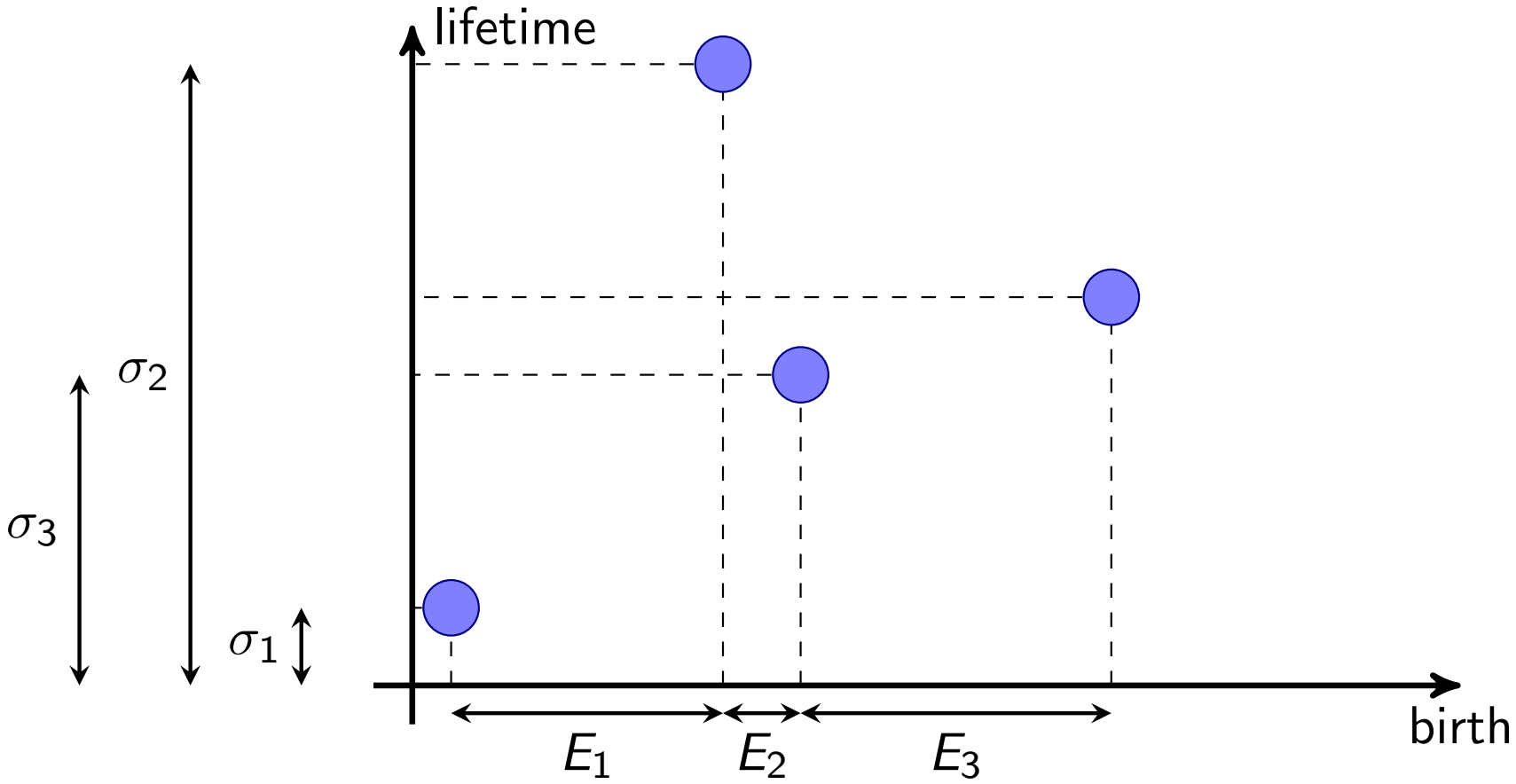
Assumptions:

- $Y(t) \in \{0, 1\}$ 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 $F_{\mathrm{M}}(\mathrm{d}t)$

Explanatory model



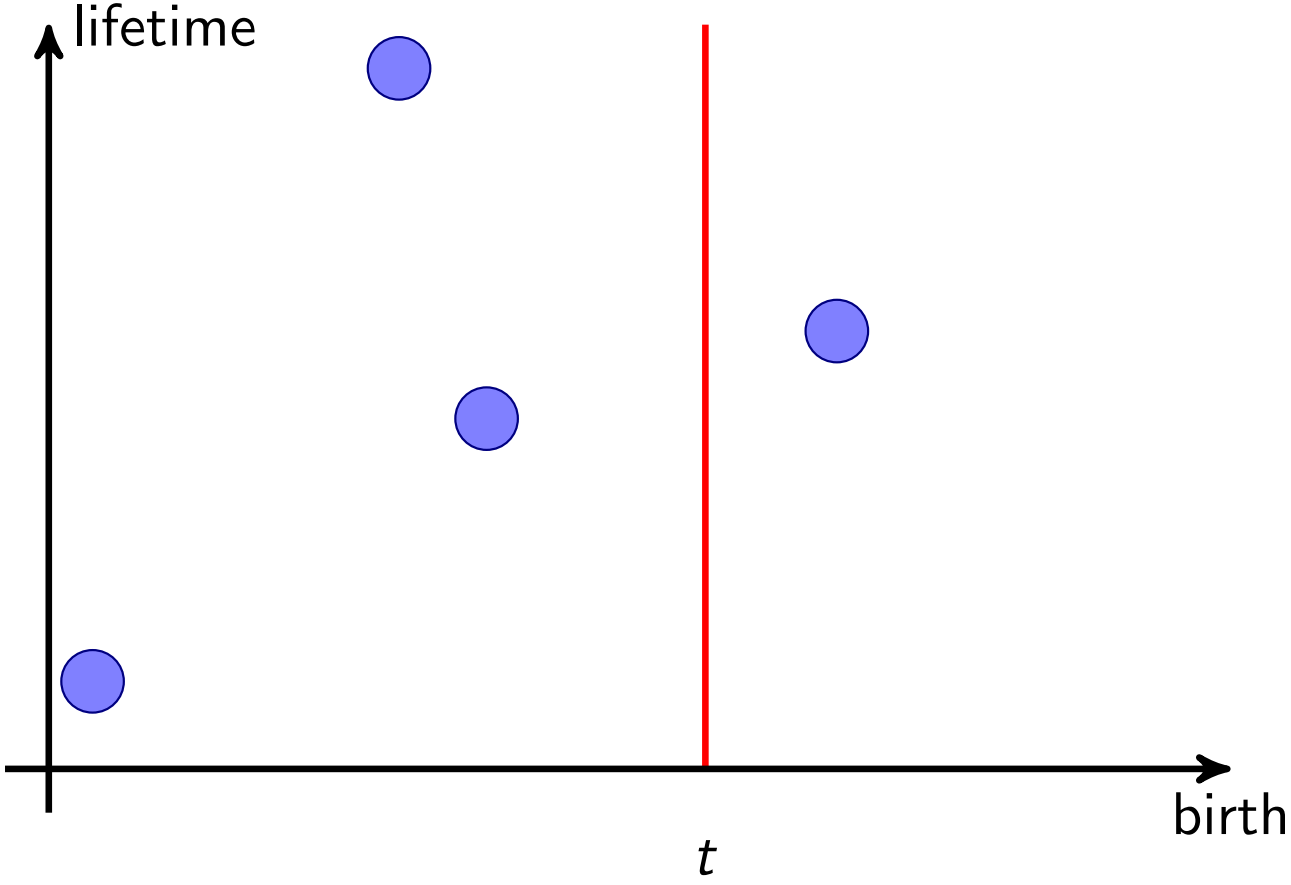
Explanatory model



E_i exponential random variables of parameter λ_M.

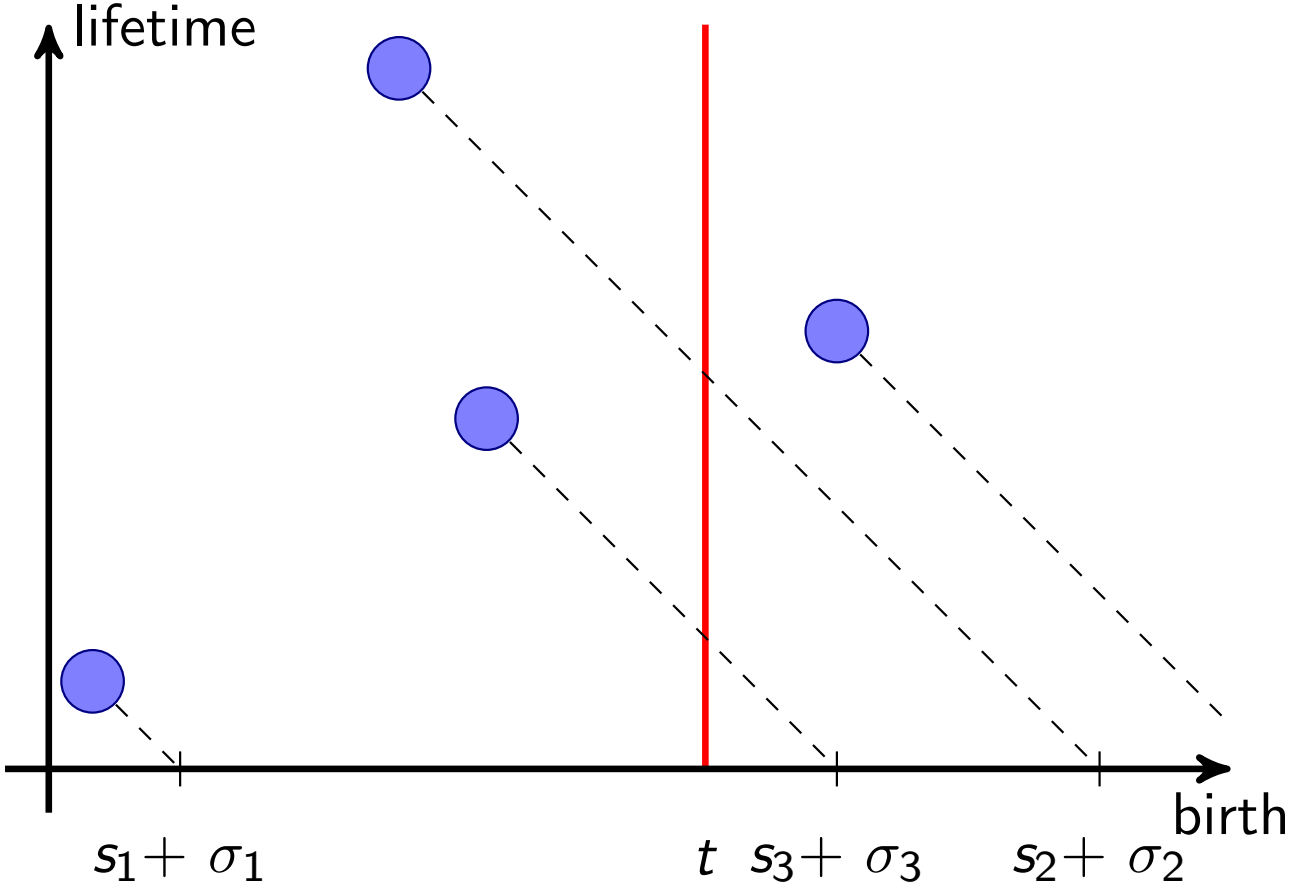
Explanatory model

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



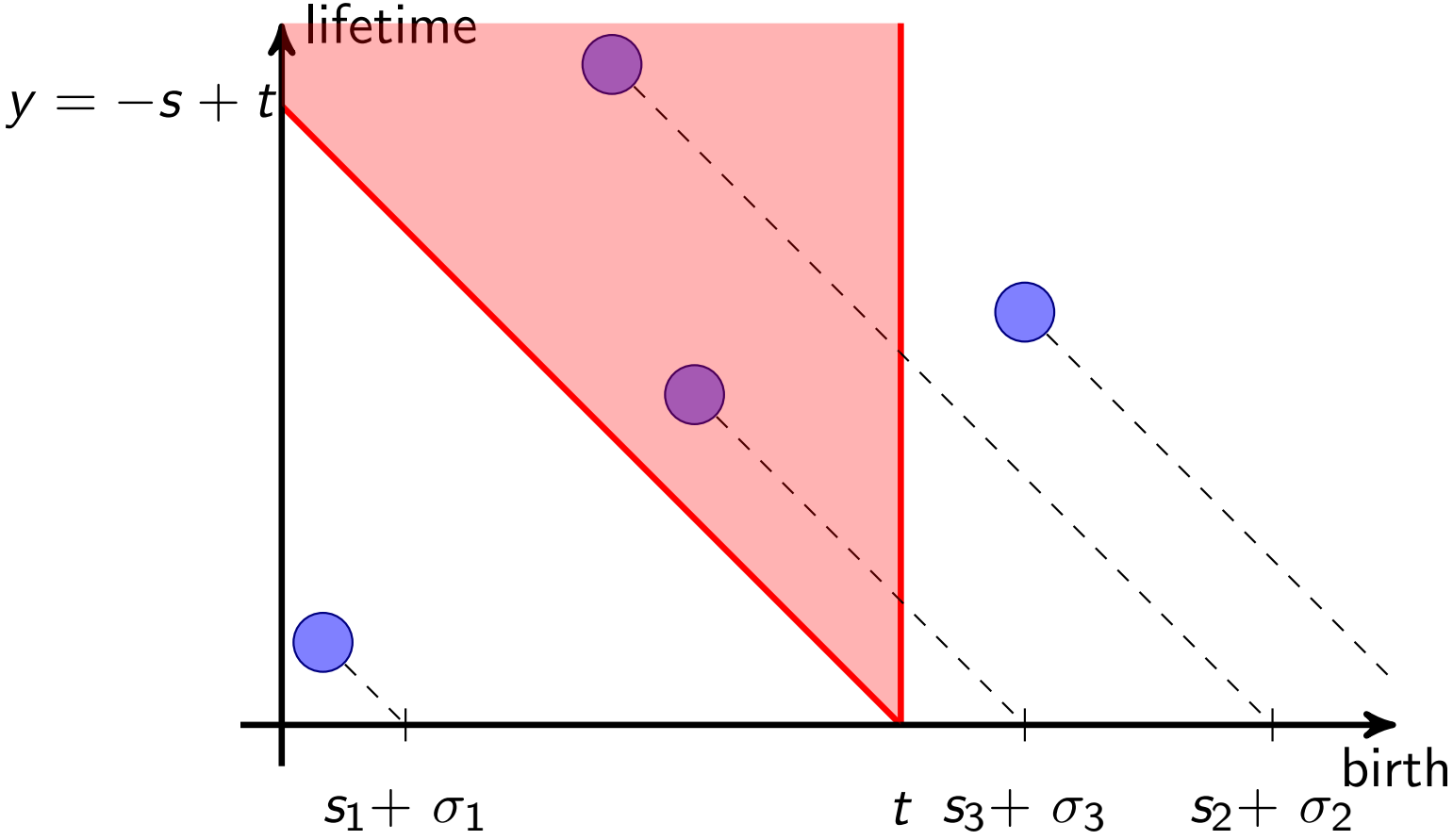
Explanatory model

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



Explanatory model

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



Explanatory model: general results

At equilibrium:

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

Explanatory model: general results

At equilibrium:

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

Proposition

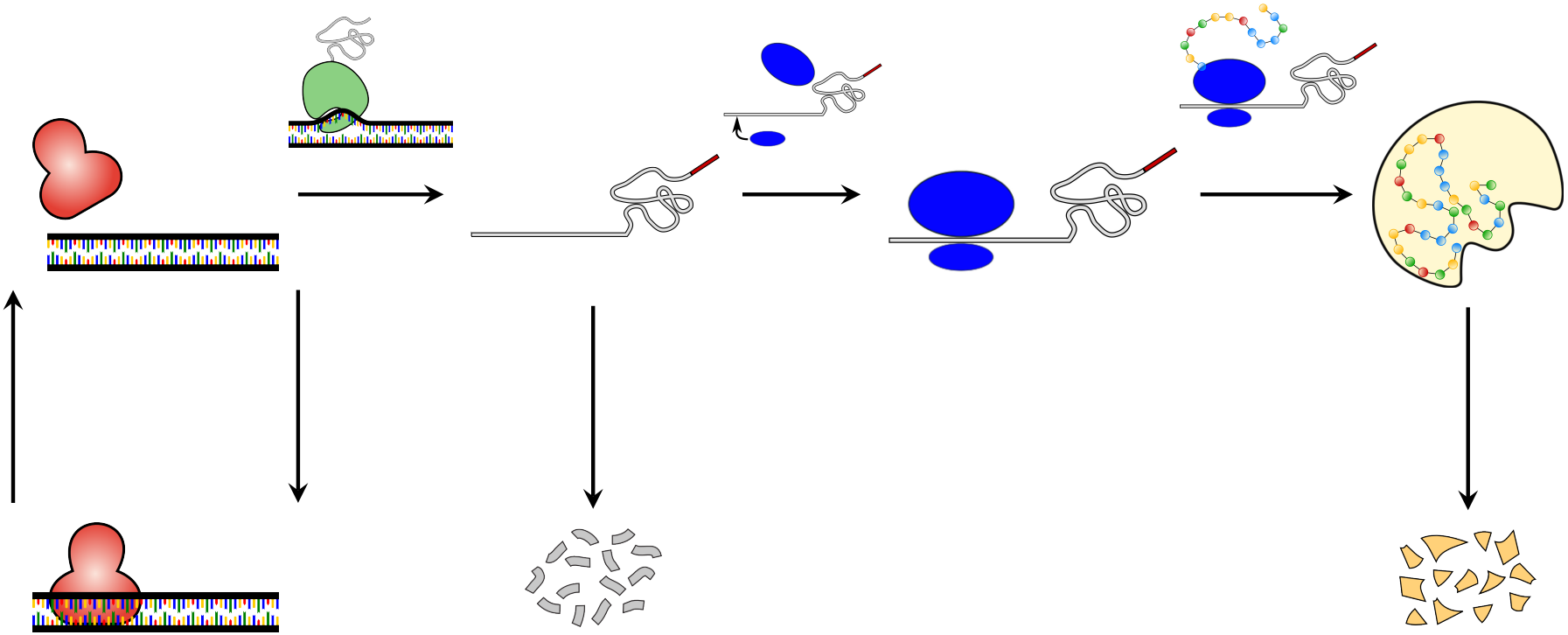
$$\mathbb{E}\left[M\right] = \delta_+ \lambda_{\mathrm{M}} \mathbb{E}\left[\sigma\right]$$

$$\mathrm{var}(M) = \mathbb{E}\left[M\right] + 2\lambda_{\mathrm{M}}^2 \delta_+ (1-\delta_+) \cdot \\ \cdot \int_0^{+\infty} \int_{-u}^{0} e^{-(\lambda^+ + \lambda^-)v} (1 - F_{\mathrm{M}}(u))(1 - F_{\mathrm{M}}(u+v))\, \mathrm{d}u\, \mathrm{d}v$$

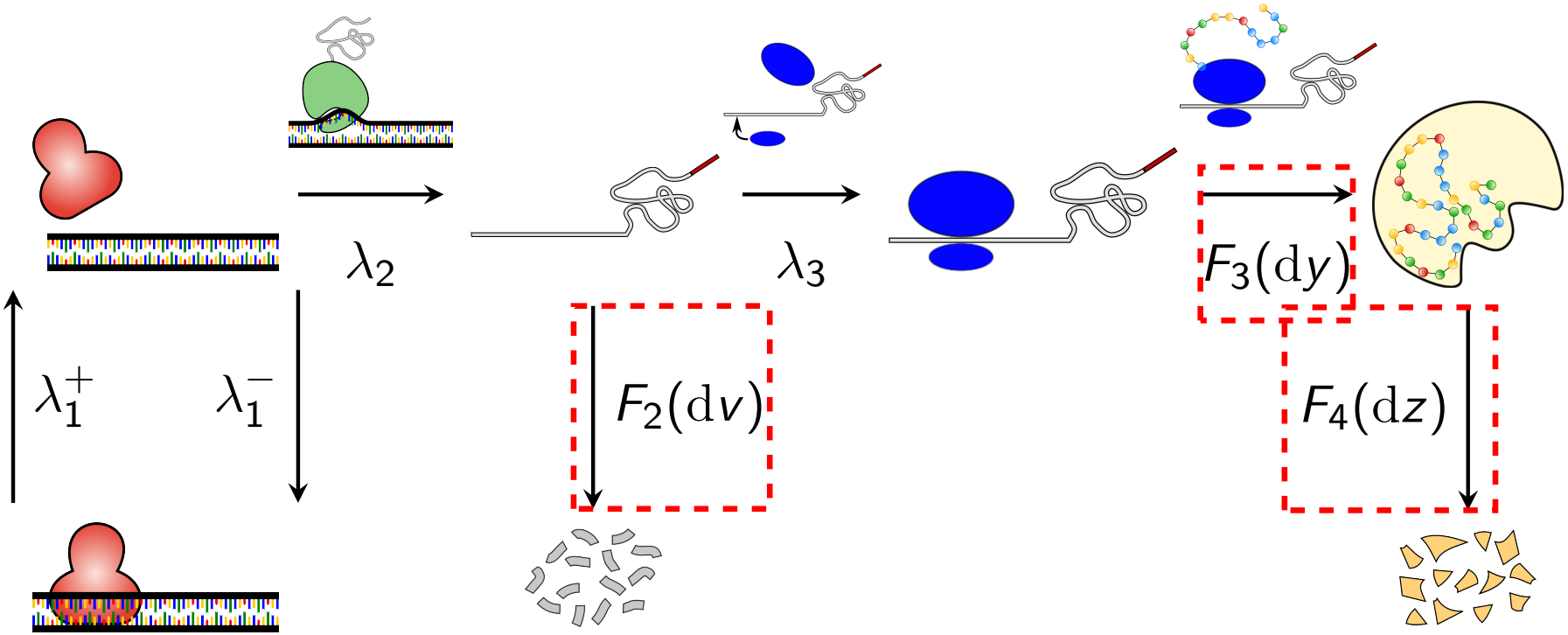
where $\delta_+ = \frac{\lambda^+}{\lambda^+ + \lambda^-}$

MPPP Applications & Model Extensions

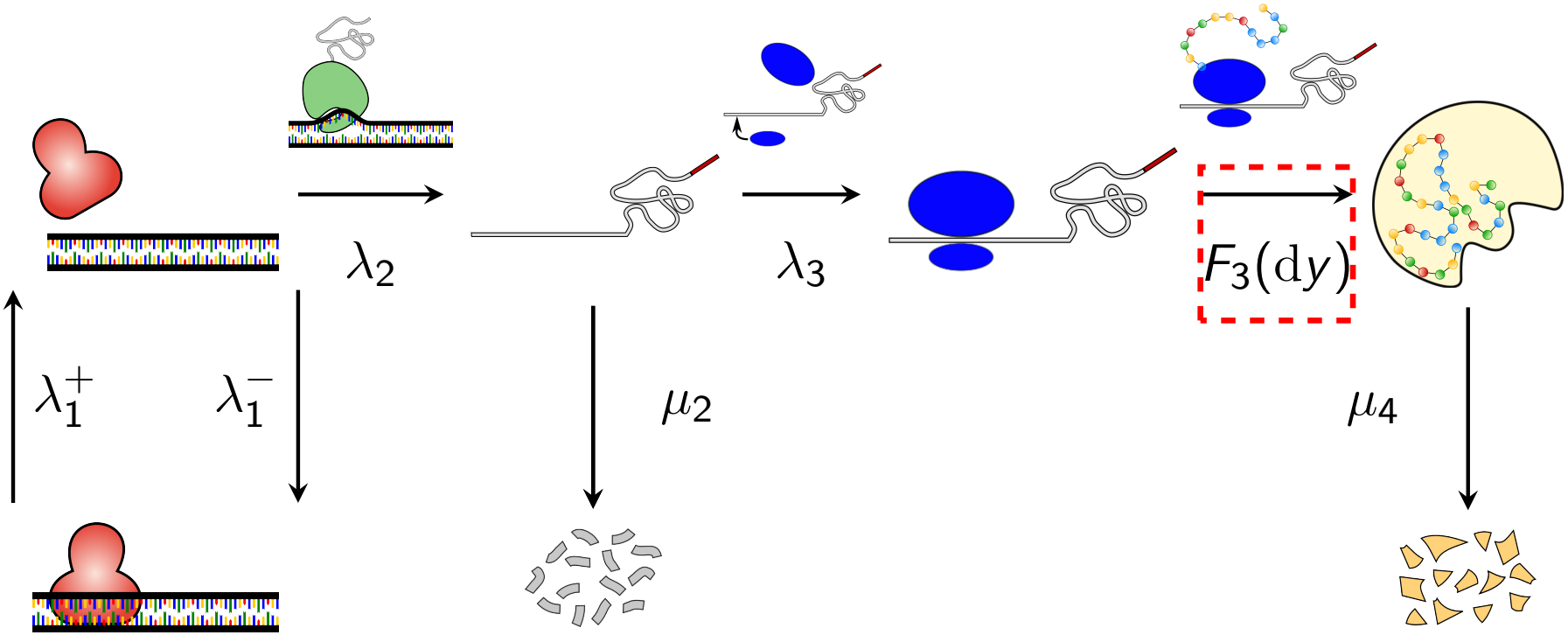
$$\begin{tikzcd}
Y(t) \in \{0,1\} \arrow[r] & M(t) \in \mathbb{N} \arrow[r] \arrow[d] & R(t) \in \mathbb{N} \arrow[r] & P(t) \in \mathbb{N} \arrow[d] \\
& \varnothing & & \varnothing
\end{tikzcd}$$



$$
\begin{tikzcd}
Y(t) \in \{0,1\} \arrow[r, "\lambda_2"'] & M(t) \in \mathbb{N} \arrow[r, "\lambda_3"'] \arrow[d, "F_2(\mathrm{d}v)"] & R(t) \in \mathbb{N} \arrow[r, "F_3(\mathrm{d}y)"'] & P(t) \in \mathbb{N} \arrow[d, "F_4(\mathrm{d}z)"] \\
& \varnothing & & \varnothing
\end{tikzcd}
$$



$$
\begin{tikzcd}
Y(t) \in \{0,1\} \arrow[r, "\lambda_2"'] & M(t) \in \mathbb{N} \arrow[r, "\lambda_3"'] \arrow[d, "\mu_2"] & R(t) \in \mathbb{N} \arrow[r, "F_3(\mathrm{d}y)"'] & P(t) \in \mathbb{N} \arrow[d, "\mu_4"] \\
 & \varnothing & & \varnothing
\end{tikzcd}
$$

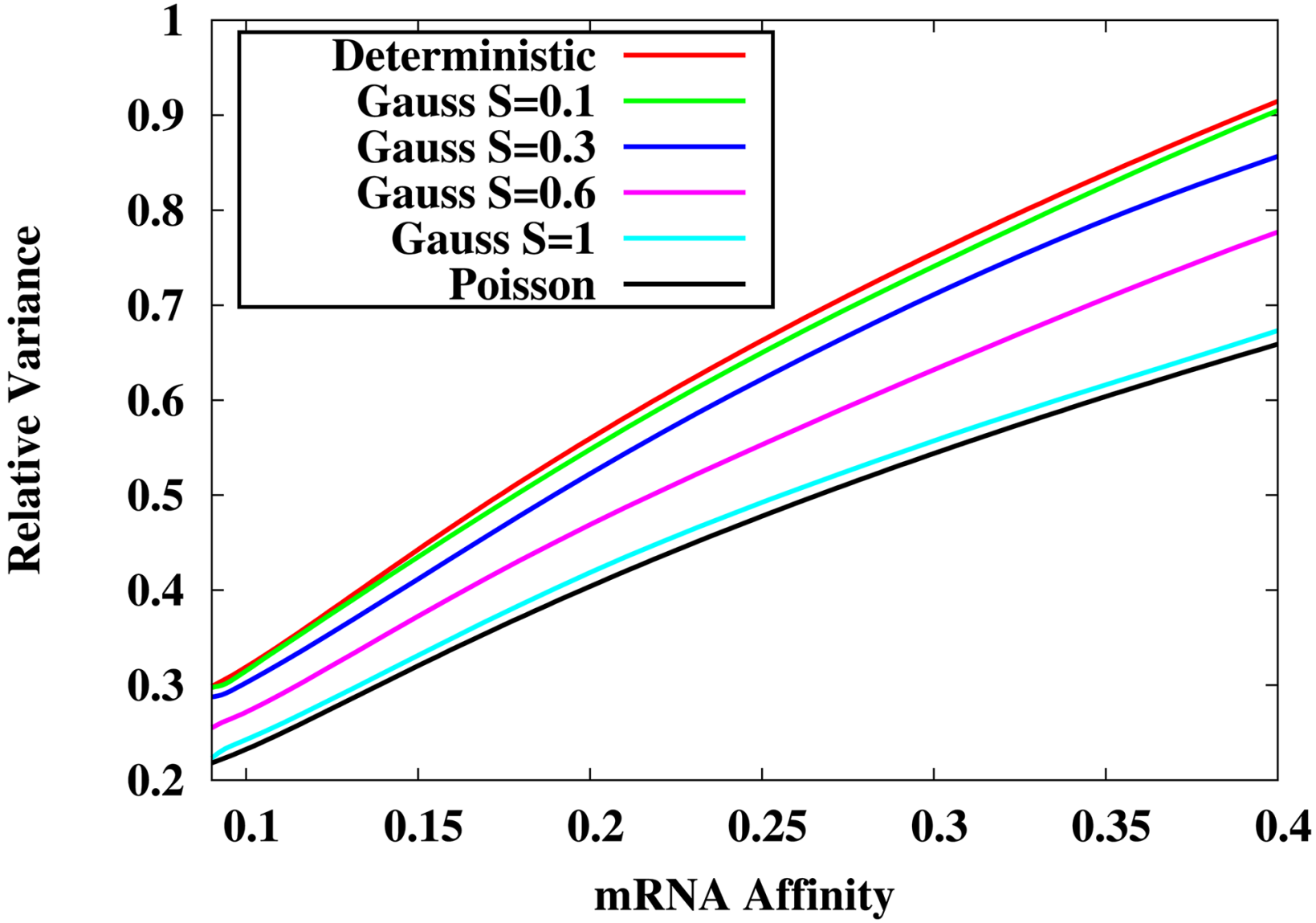


$$
\begin{tikzcd}
Y(t) \in \{0,1\} \arrow[r, "\lambda_2"'] & M(t) \in \mathbb{N} \arrow[r, "\lambda_3"'] \arrow[d, "\mu_2"] & R(t) \in \mathbb{N} \arrow[r, "F_3(\mathrm{d}y)"'] & P(t) \in \mathbb{N} \arrow[d, "\mu_4"] \\
& \varnothing & & \varnothing
\end{tikzcd}
$$

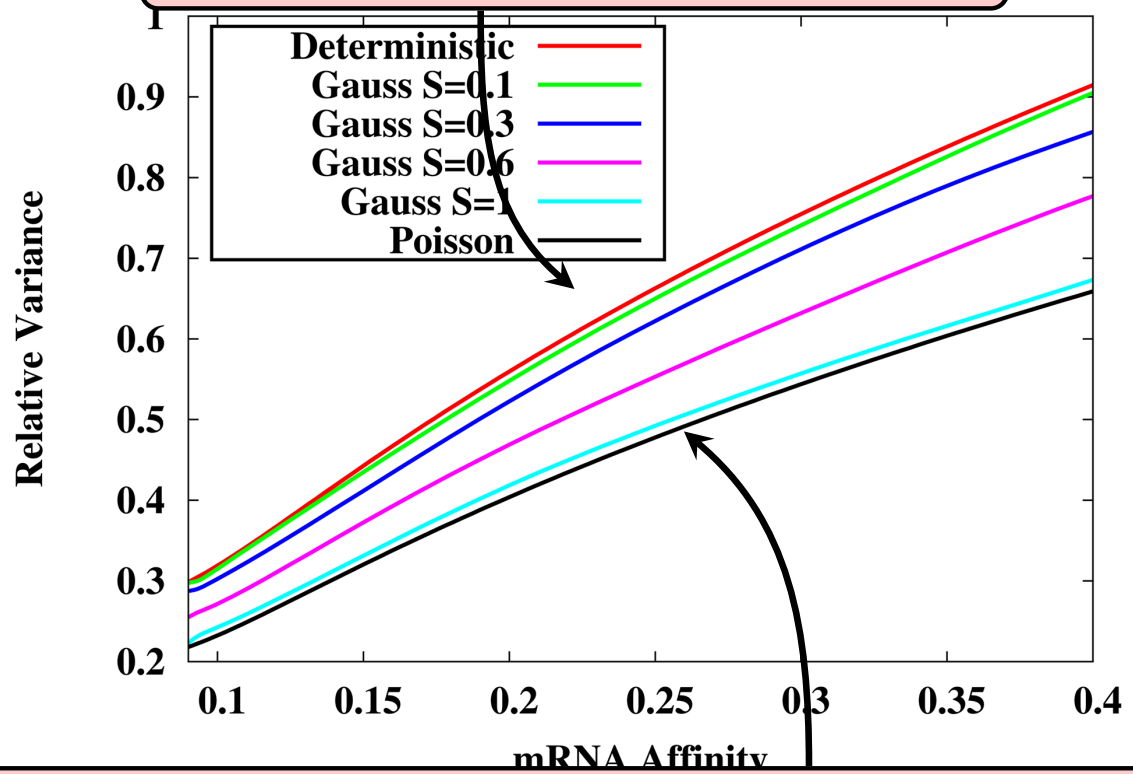
Choices for $F_3(\mathrm{d}y)$

Exponential	➡	explicit close formula depending on model parameters
Normal	➡	analytic formula
Deterministic	➡	explicit close formula depending on model parameters (limit case)

Deterministic vs Exponential



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



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

Conclusions

- (MPPP) appropriate math tool to describe cell stochastic processes (**Exponential + General Distribution**)

Conclusions

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

Conclusions

- (MPPP) appropriate math tool to describe cell stochastic processes (**Exponential + General Distribution**)
- 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 (**Exponential + General Distribution**)
- 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 (**Exponential + General Distribution**)
- 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: $\mathrm{var}_{\mathsf{DET}}(P) > \mathrm{var}_{\mathsf{EXP}}(P)$

Thanks.