Modélisation probabiliste en biologie cellulaire et moléculaire

Thèse sous la direction de

M. Adimy, M.C. Mackey & L. Pujo-Menjouet

Romain Yvinec

Institut Camille Jordan - Université Claude Bernard Lyon 1

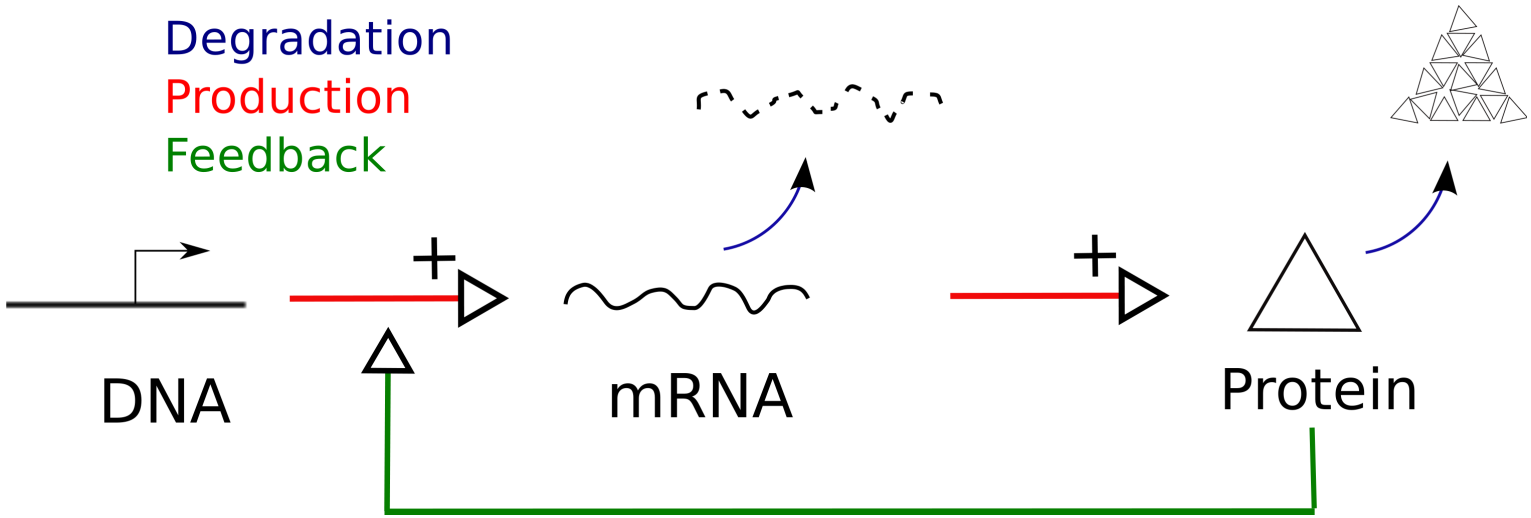
Vendredi 05 octobre 2012

Outline

Outline

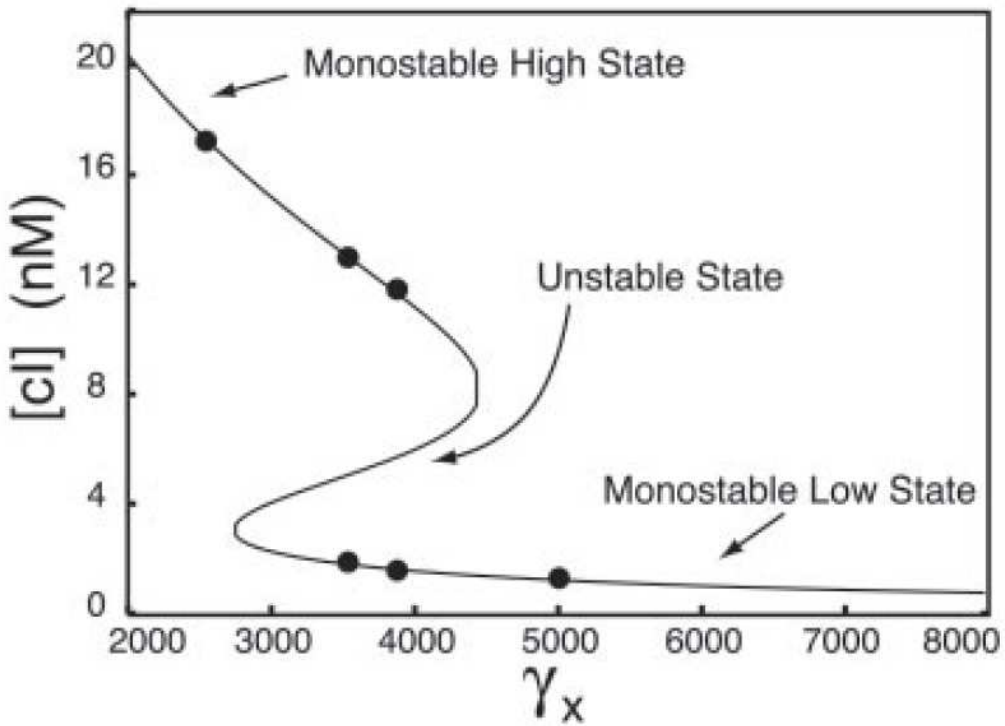
Central Dogma

- Expression of a gene through transcription/translation processes.



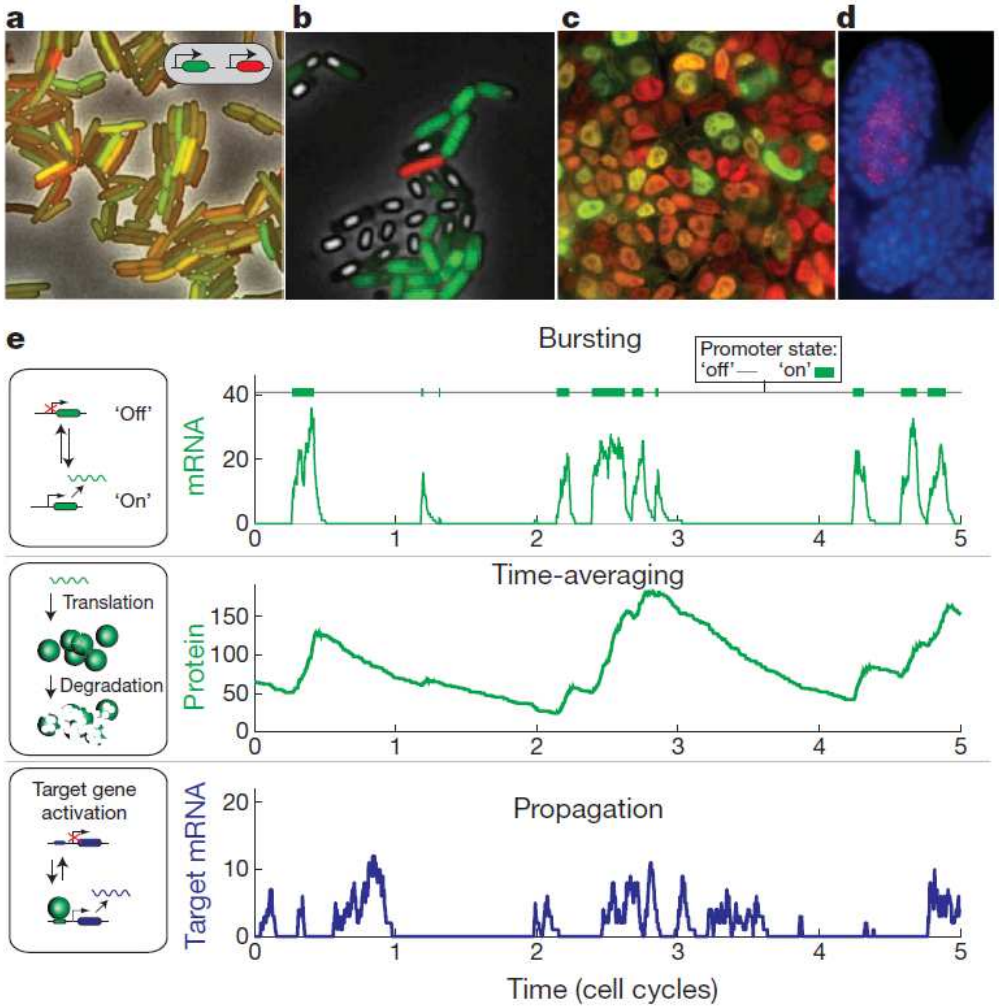
- Non-linear Feedback regulation.

- Bifurcation analysis in Ordinary Differential Equation.
- Application to synthetic biology.



[Goodwin, 1965],[Hasty et al., 2001].

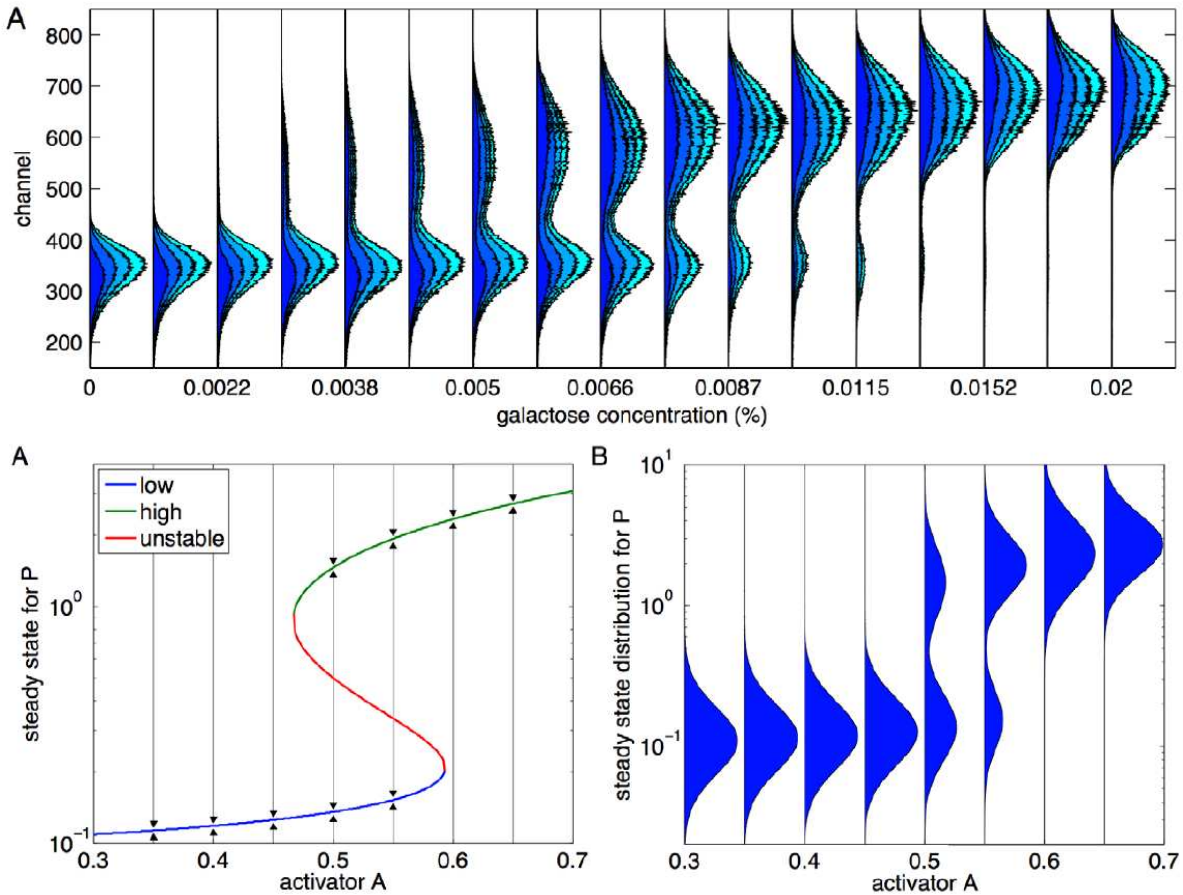
Stochasticity in molecular biology



[Eldar and Elowitz, 2010].

Much more accurate measurements

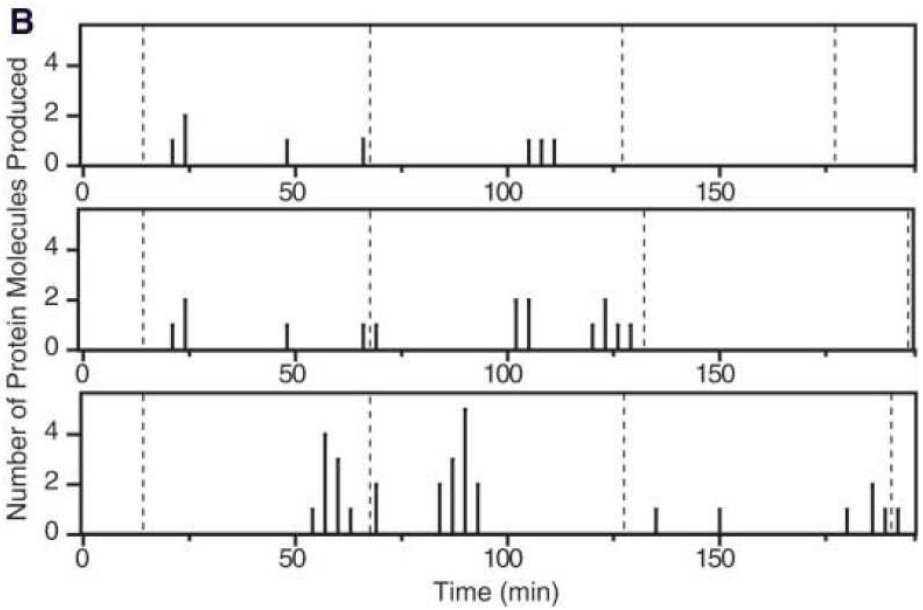
- Bifurcation can be studied on probability distributions.

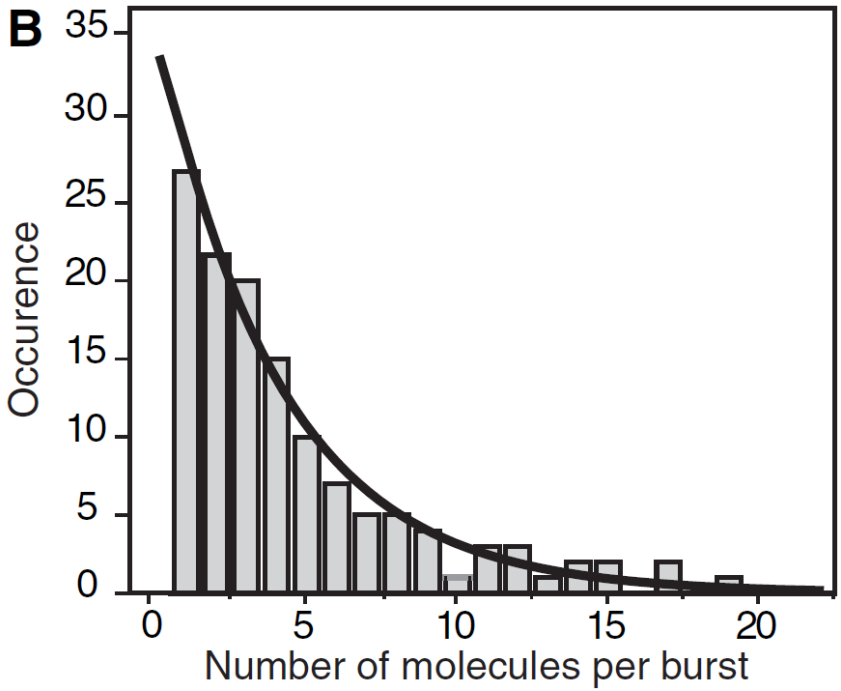


[Song et al., 2010].

Much more accurate measurements

- Trajectories can be analyzed on single cells.

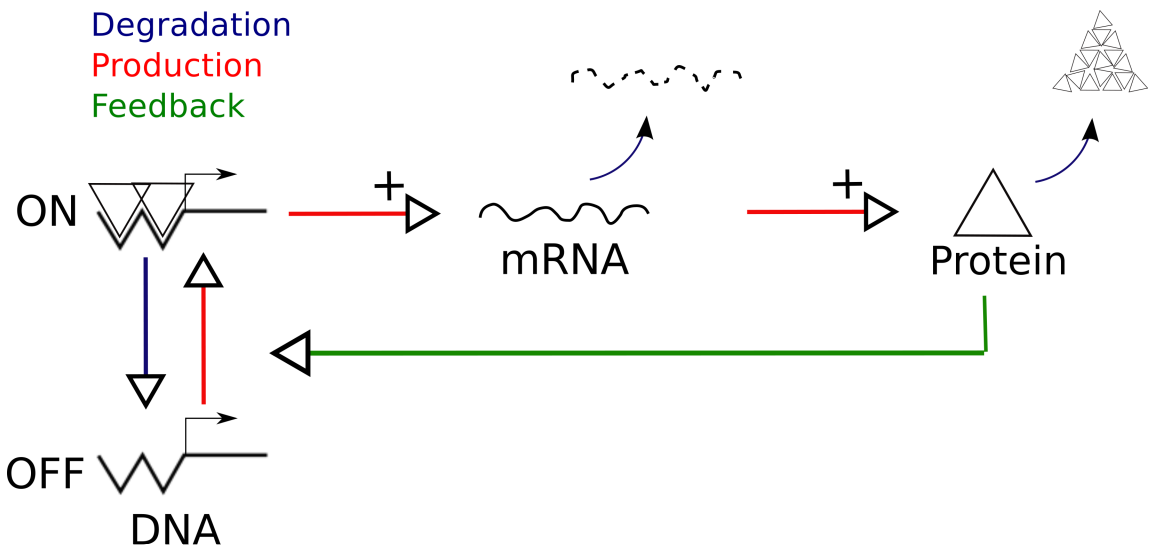




[Yu et al., 2006].

New Central dogma

- Take into account gene state switching. Interpretation as stochastic processes.



[Berg, 1978],
[Peccoud and Ycart, 1995],
[Kepler and Elston, 2001],
[Paulsson, 2005],
[Lipniacki et al., 2006],
[Paszek, 2007],

[Shahrezaei and Swain, 2008].

The bursting phenomena

Question 1) **When does the stochastic model predict burst phenomenon ?**

Question 2) **What can we say in such cases ?**

Outline

We consider the following 2d stochastic kinetic chemical reaction model (X='mRNA', Y='Protein')

$$
\begin{aligned}
\emptyset &\xrightarrow{\lambda_1(X,Y)} X, && \text{Production of } X \text{ at rate } \lambda_1(X,Y)\\
X &\xrightarrow{N\gamma_1(X,Y)} \emptyset, && \text{Destruction of } X \text{ at rate } N\gamma_1(X,Y)\\
\emptyset &\xrightarrow{N\lambda_2(X,Y)} Y, && \text{Production of } Y \text{ at rate } N\lambda_2(X,Y)\\
Y &\xrightarrow{\gamma_2(X,Y)} \emptyset, && \text{Destruction of } Y \text{ at rate } \gamma_2(X,Y)
\end{aligned}
$$

with $\gamma_1(0,Y) = \gamma_2(X,0) = 0$ to ensure non-negativity.

$$
\begin{aligned}
\mathbb{B}_N f(x,y) =& \lambda_1(x,y)\Big[f(x+1,y) - f(x,y)\Big]\\
&+ N\gamma_1(x,y)\Big[f(x-1,y) - f(x,y)\Big]\\
&+ N\lambda_2(x,y)\Big[f(x,y+1) - f(x,y)\Big]\\
&+ \gamma_2(x,y)\Big[f(x,y-1) - f(x,y)\Big].
\end{aligned}
$$

Theorem (R.Y.)

If

- The degradation function on X satisfies

$$\inf_{x\geq 1, y\geq 0} \gamma_1(x,y) = \underline{\gamma} > 0.$$

- The production rate of Y satisfies $\lambda_2(0,y) = 0$, for all $y \geq 0$.
- λ_1 and λ_2 are linearly bounded by $x+y$, and either λ_1 or λ_2 is bounded.

Then, for all $T > 0$, $(X^N(t), Y^N(t))_{t\geq 0}$ converges in $L^1(0,T)$ to $(0, Y(t))$, whose generator is given by

$$\begin{cases} \mathbb{B}_\infty \varphi(y) &= \lambda_1(0,y)\Big(\int_0^\infty P_t(\gamma_1(1,\cdot)\varphi(\cdot))(y)dt - \varphi(y)\Big) \\ &\qquad + \gamma_2(0,y)\Big[\varphi(y-1)-\varphi(y)\Big], \\ P_t g(y) &= \mathbb{E}\Big[g(Z(t,y)e^{-\int_0^t \gamma_1(1,Z(s,y))ds}\Big], \\ Ag(z) &= \lambda_2(1,z)\big(g(z+1)-g(z)\big). \end{cases}$$

Sketch of the proof

- We first show tightness and convergence of X based on

$$N\underline{\gamma}\mathbb{E}\Big[\int_0^t \mathbf{1}_{\{X^N(s)\geq 1\}} ds\Big] \leq \mathbb{E}\big[X^N(0)\big] + \mathbb{E}\Big[\int_0^t \lambda_1(X^N(s), Y^N(s)) ds\Big].$$

Sketch of the proof

- We first show tightness and convergence of X based on

$$N\underline{\gamma}\mathbb{E}\Big[\int_0^t \mathbf{1}_{\{X^N(s)\geq 1\}} ds\Big] \leq \mathbb{E}\big[X^N(0)\big] + \mathbb{E}\Big[\int_0^t \lambda_1(X^N(s), Y^N(s)) ds\Big].$$

- We identify the limiting martingale problem

Sketch of the proof

- We first show tightness and convergence of X based on

$$N\underline{\gamma}\mathbb{E}\Big[\int_0^t \mathbf{1}_{\{X^N(s)\geq 1\}} ds\Big] \leq \mathbb{E}\big[X^N(0)\big] + \mathbb{E}\Big[\int_0^t \lambda_1(X^N(s), Y^N(s)) ds\Big].$$

- We identify the limiting martingale problem

$$\lambda_2(x,y)\Big[f(x,y+1)-f(x,y)\Big] + \gamma_1(x,y)\Big[f(x-1,y)-f(x,y)\Big] = 0.$$

Sketch of the proof

- We first show tightness and convergence of X based on

$$N\underline{\gamma}\mathbb{E}\Big[\int_0^t \mathbf{1}_{\{X^N(s)\geq 1\}} ds\Big] \leq \mathbb{E}\big[X^N(0)\big] + \mathbb{E}\Big[\int_0^t \lambda_1(X^N(s), Y^N(s)) ds\Big].$$

- We identify the limiting martingale problem

$$A^x g(y) = \lambda_2(x,y)\Big[g(y+1) - g(y)\Big],$$

for any $x \geq 1$. and we introduce the semigroup P_t^x

$$P_t^x g(y) = \mathbb{E}\big[g(Z_t^{x,y}) e^{-\int_0^t \gamma_1(x, Z_s^{x,y}) ds}\big].$$

Now for any bounded function g, define $f(0,y) = g(y)$ and

$$f(x,y) = \int_0^\infty P_t^x(\gamma_1(x,.) f(x-1,.))(y) dt.$$

Then

$$\lambda_2(x,y)\Big[f(x,y+1) - f(x,y)\Big] + \gamma_1(x,y)\Big[f(x-1,y) - f(x,y)\Big] = 0.$$

- A similar proof for a (continuous state) PDMP model, of generator

$$\mathbb{B}f(x,y) = -N\gamma_1(x,y)\frac{\partial f}{\partial x} + (N\lambda_2(x,y) - \gamma_2(x,y))\frac{\partial f}{\partial y} + \lambda_1(x,y)\int_0^\infty (f(x+z,y) - f(x,y))h(z)dz.$$

- These proofs are based on a simple idea ([Debussche et al., 2011],[Kang and Kurtz, 2011]).
- Other proof : reduction on the Fokker-Planck equation.
- Different scalings lead to different models.

Outline

We look at the stochastic process

$$dx = -\gamma(x)dt + dN(\lambda(x), h(x,\cdot)),$$

whose generator is

$$Af = -\gamma(x)f'(x) + \lambda(x)\Big(\int_0^\infty f(x+y)h(x,y)dy - f(x)\Big),$$

and evolution equation on densities

$$\frac{\partial u(t,x)}{\partial t} = \frac{\partial \gamma(x)u(t,x)}{\partial x} - \lambda(x)u(t,x) + \int_0^x u(t,y)\lambda(y)h(y,x-y)dy,$$

and with $\int_0^\infty h(x,y)dy = 1$, for all x.

Probabilistic techniques

If jumps are independent of positions, *i.e.* $h(x,y) = h(y)$, we have :

Proposition

Suppose $x \mapsto \lambda(x)$ *is continuous on* $(0,\infty)$, $\lambda(0) > 0$, $\gamma(x) = \gamma x$, $\mathbb{E}\left[h\right] < \infty$, *and*

$$\lim_{x\to\infty} \frac{\lambda(x)\mathbb{E}\left[h\right]}{\gamma x} < 1,$$

then there exist $\beta < 1$, $B < \infty$ *and* π *(invariant measure) such that*

$$\|P(t,x,\cdot) - \pi\|_V \leq BV(x)\beta^t, \quad x \in E, \quad t > 0,$$

where $\|\mu\|_f = \sup_{|g|\leq f} \mid \mu(g) \mid$ *and* $V(x) = x + 1$.

Probabilistic techniques

If jumps are independent of positions, *i.e.* $h(x,y) = h(y)$, we have :

Proposition

Suppose $x \mapsto \lambda(x)$ *is continuous on* $(0,\infty)$, $\lambda(0) > 0$, $\gamma(x) = \gamma x$, $\mathbb{E}\left[h\right] < \infty$, *and*

$$\lim_{x\to\infty} \frac{\lambda(x)\mathbb{E}\left[h\right]}{\gamma x} < 1,$$

then there exist $\beta < 1$, $B < \infty$ *and* π *(invariant measure) such that*

$$\|P(t,x,\cdot) - \pi\|_V \leq BV(x)\beta^t, \quad x \in E, \quad t > 0,$$

where $\|\mu\|_f = \sup_{|g|\leq f} \mid \mu(g) \mid$ *and* $V(x) = x + 1$.

- $Ax = -\gamma x + \lambda(x)\Big(\int_0^\infty (x+y)h(y)dy - x\Big) = -(1 - \frac{\lambda(x)\mathbb{E}\left[h\right]}{\gamma x})\gamma x$

Semigroup techniques

$$\underbrace{\frac{\partial u(t,x)}{\partial t}}_{\frac{du}{dt}} = \underbrace{\frac{\partial \gamma(x)u(t,x)}{\partial x} - \lambda(x)u(t,x)}_{Au=(A_0-\lambda)u} + \underbrace{\int_0^x u(t,y)\lambda(y)h(y,x-y)dy}_{Bu=J(\lambda u)}$$

$$(A, D(A)) \Rightarrow S(t)u(x) = P_0(t)u(x)e^{-\int_0^t \lambda(\phi_r x)dr}$$

Let $C = A + B$. Denote the resolvent $R_s^S u = \int_0^\infty e^{-st} S(t)udt$.

Semigroup techniques

$$\underbrace{\frac{\partial u(t,x)}{\partial t}}_{\frac{du}{dt}} = \underbrace{\frac{\partial \gamma(x)u(t,x)}{\partial x} - \lambda(x)u(t,x)}_{Au=(A_0-\lambda)u} + \underbrace{\int_0^x u(t,y)\lambda(y)h(y,x-y)dy}_{Bu=J(\lambda u)}$$

$$(A, D(A)) \Rightarrow S(t)u(x) = P_0(t)u(x)e^{-\int_0^t \lambda(\phi_r x)dr}$$

Let $C = A + B$. Denote the resolvent $R_s^S u = \int_0^\infty e^{-st} S(t)udt$.

Theorem ([Tyran-Kamińska, 2009])
There is a minimal substochastic semigroup P generated by an extension of $(C, D(A))$, and which resolvent is given by

$$R_s^P u = \lim_{n\to\infty} R_s^S \sum_{k=0}^{n} (J(\lambda R_s^S))^k u,$$

and if $K = \lim_{\sigma\to 0} J(\lambda R_\sigma^S)$ has a unique invariant density, then so does for P (and P is stochastic).

- Under good conditions, K is the transition operator for the discrete Markov chain "post-jump", and has for kernel

$$\begin{cases} k(x,y) &= \displaystyle\int_0^x \mathbf{1}_{\{(0,y)\}}(z) h(z,y-z) \frac{\lambda(z)}{\gamma(z)} e^{Q(x)-Q(z)} dz, \\ Q(x) &= \displaystyle\int_x^{\overline{x}} \frac{\lambda(z)}{\gamma(z)} dz. \end{cases}$$

- Modulo integrability conditions, invariant density v^* for K and invariant density u^* for P are related through

$$\begin{cases} \gamma(x)u^*(x) = \displaystyle\int_0^x \overline{H}(z,x-z)\lambda(z)u^*(z)dz, \ \overline{H}(z,x) = \int_x^\infty h(z,y)dy, \\ v^*(x) = \displaystyle\int_0^x h(z,x-z)\lambda(z)u^*(z)dz, \\ u^*(x) = \displaystyle\frac{1}{\gamma(x)} \int_x^\infty e^{Q(y)-Q(x)} v^*(y)dy, \\ v^*(x) = \displaystyle\int_0^x h(z,x-z)\frac{\lambda(z)}{\gamma(z)} e^{-Q(z)} \int_z^\infty v^*(y) e^{Q(y)} dydz. \end{cases}$$

Condition for ergodicity in the exponential case

If jumps are independent of positions, *i.e.* $h(x,y) = h(y)$ and exponentially distributed, of mean b, *i.e.* $h(y) = \frac{1}{b}e^{-y/b}$, then

Theorem (M. Tyran-Kamińska, M. Mackey, R.Y.)

Under technical assumptions (for integrability), and if

$$\lim_{x\to\infty} \frac{\lambda(x)}{\gamma(x)} < \frac{1}{b},$$

$$Q(0) := \int_0^{\overline{x}} \frac{\lambda(z)}{\gamma(z)} dz = \infty,$$

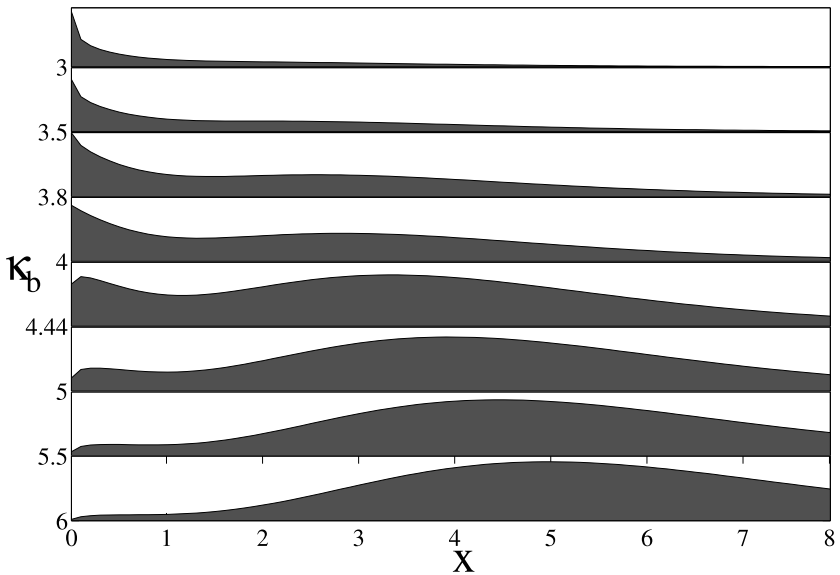
then P is ergodic with unique invariant density

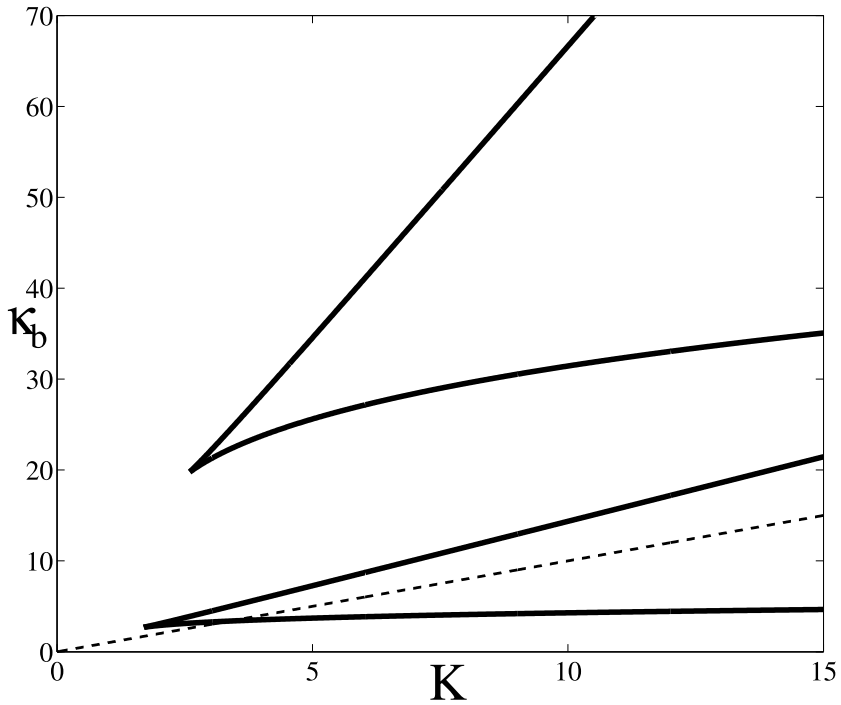
$$u^*(x) = \frac{1}{c\gamma(x)} e^{-x/b - Q(x)}.$$

Bifurcation

This analytical approach allows us to deduce that the number of modes of the stationary state is linked to the solution of

$$\lambda(x) = \frac{\gamma(x)}{b} + \gamma'(x).$$





Further results (not developed here)

- This can be used to find $\lambda(x)$ and b from observations of (u^*, γ).
- The convergence rate can be estimated from coupling techniques.

Further results (not developed here)

- This can be used to find $\lambda(x)$ and b from observations of (u^*, γ).
- The convergence rate can be estimated from coupling techniques.

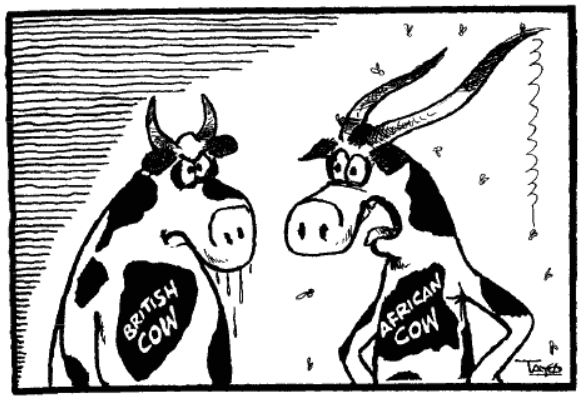
Perspectives

- Other jump size kernel h.
- Waiting time properties.
- Switch and bursting model.
- Include cell division and study population dynamics.
- Characterize oscillations in two-dimensional model.

Outline

Prion Diseases

- Creutzfeldt-Jakob : First human prion disease described (1929).
- Mad cow disease, Scrapie.
- Kuru (New Guinea)



We may be skinny, but at least we're not MAD !

Prion Diseases

Epidemiology

- All prion diseases are transmissible, *i.e.* infectious.
- Some prion diseases are sporadic, they appear spontaneously, without cause.
- Some prion diseases are genetic.

Symptoms

- Affect the structure of the brain ;
- Convulsion, Dementia, Loss of balance ;
- Always fatal.

Outline

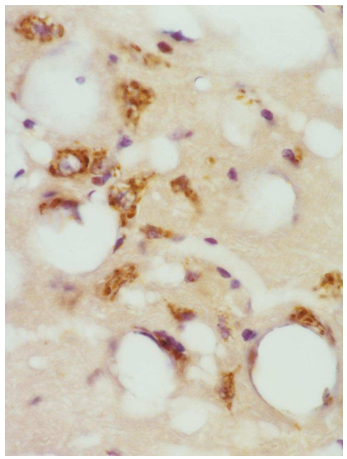
What is prion ?

A protein

- A protein called PRION is the cause of this disease
- It is neither a bacteria, nor an viroid like agent !
- Stanley Prusiner was awarded Nobel price in Physiology and Medicine in 1997 for his discovery.

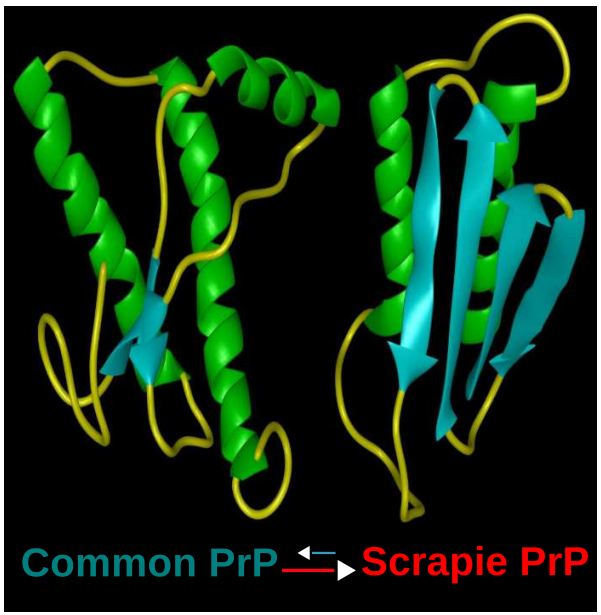
Histopathology

- Accumulation of a protein in the amyloid form.
- Spongiosis.



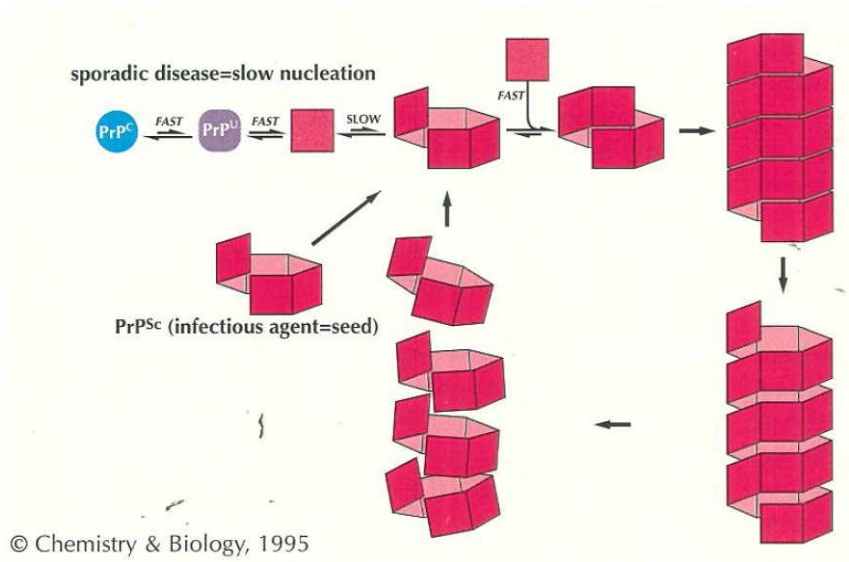
Nucleation-Polymerization

Conformation change



Normal $\Leftrightarrow$ Misfolded

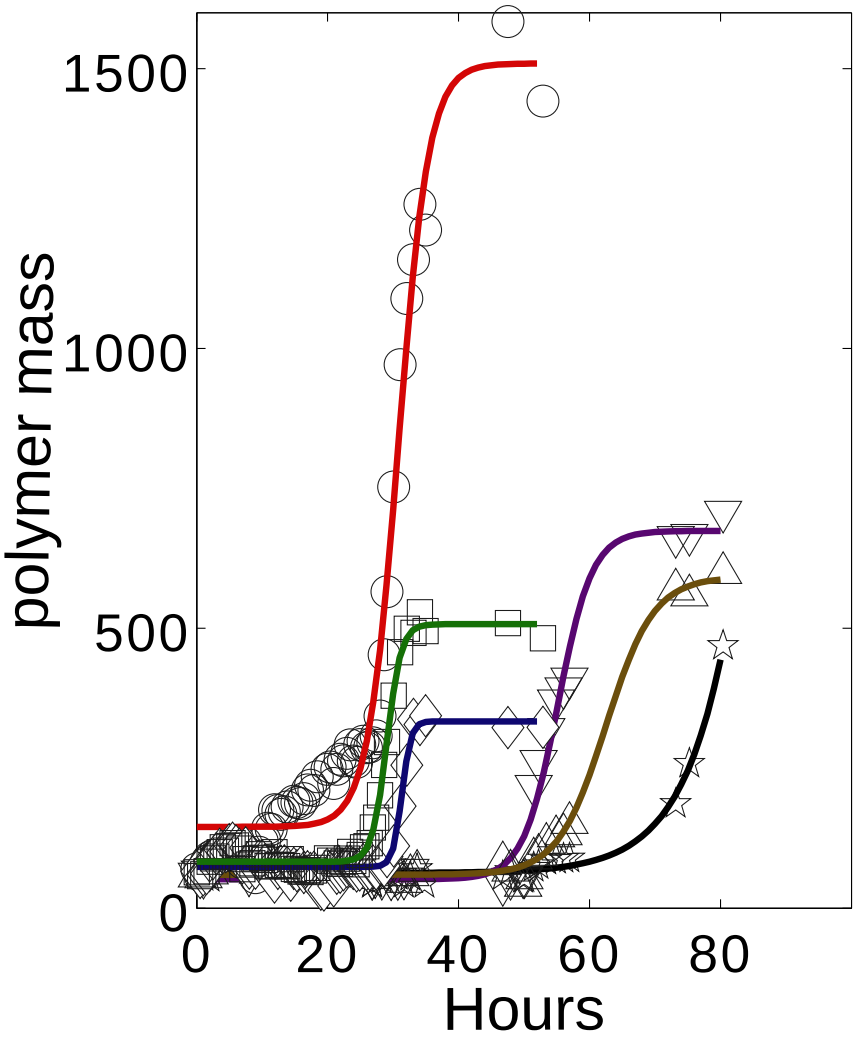
Lansbury's model



[Lansbury and Caughey, 1995].

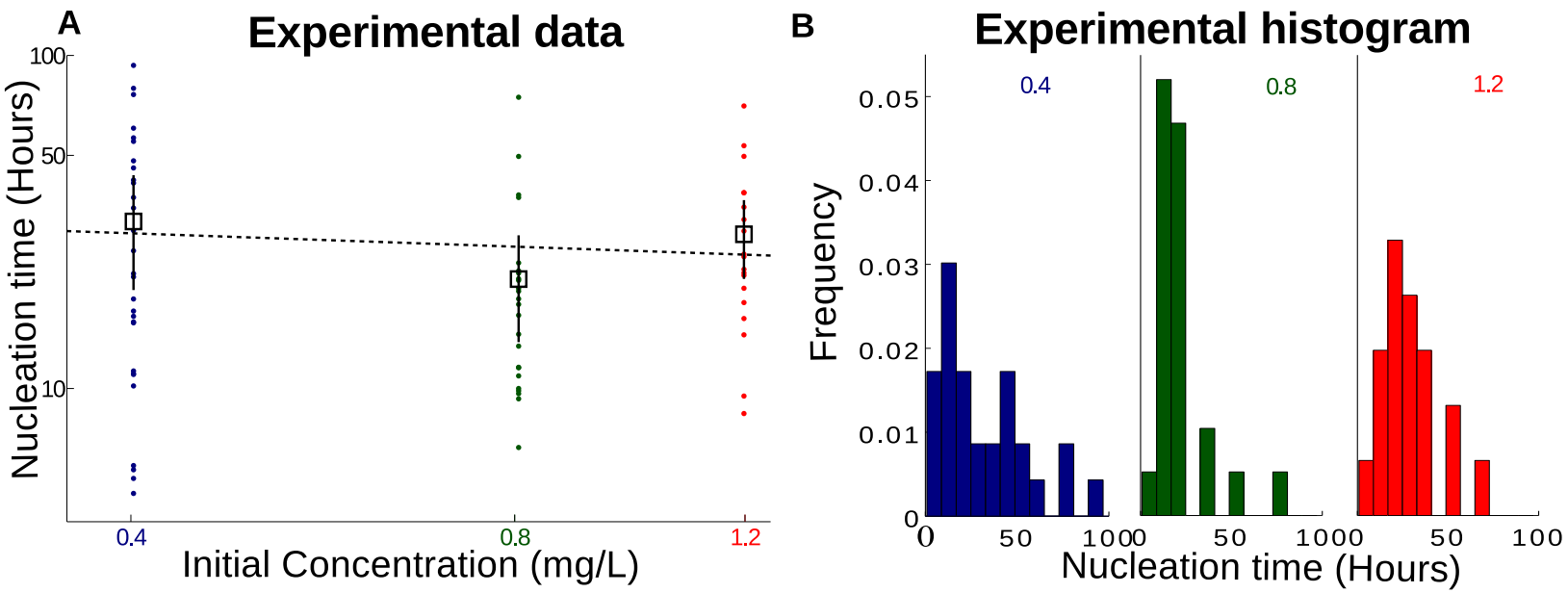
Outline

- *In vitro* spontaneous polymerization experiments.
- Time series of polymer mass.



Statistics of nucleation time

- Relation between the nucleation time and the initial concentration in log plots.



- Full distribution of the nucleation time.

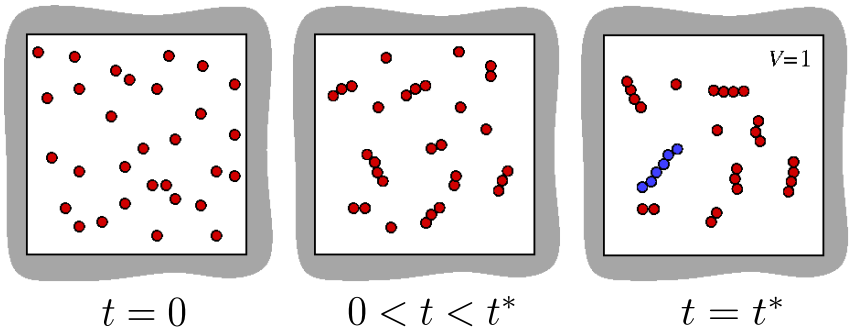
Questions

- Can a probabilistic model reproduce the observed variability ?
- Can it help to identify parameters ?
- Can a model include different strain structures ?

Outline

Reversible aggregation model

$$C_i + C_1 \underset{q}{\overset{p}{\rightleftharpoons}} C_{i+1}$$



where

$$C_i = \sharp\{\text{molecules of size i}\}.$$

The nucleation time is given by a waiting time problem,

$$T_{lag} = \inf\{t \geq 0 : C_N(t) = 1\},$$

with initial condition $C_1(0) = M$, $C_i(0) = 0$, $i \geq 2$.
N is the nucleus size.

Constant monomer formulation

If we suppose

$$C_1(t) \equiv M$$

We can solve exactly the probability distributions (Poisson) and we deduce

$$S(t) := \mathbb{P}\big\{C_N(s) = 0, s \leq t\big\} = \mathbb{P}\big\{C_N(t) = 0\big\} = e^{-c_N(t)},$$

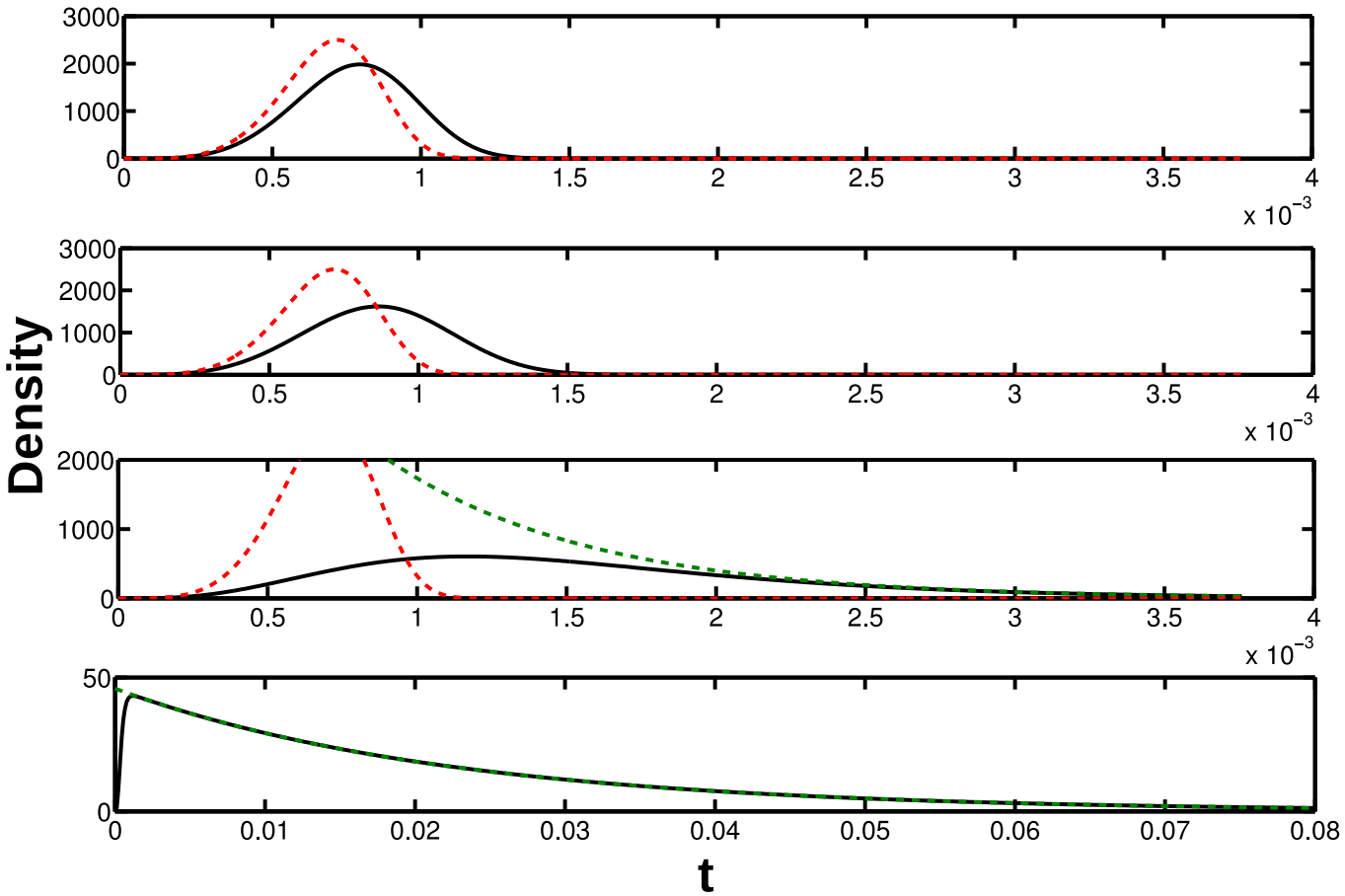
where $(c_i)_{i=2..N}$ are solution of the linear deterministic system (c_n is absorbing) :

$$\begin{cases} \dot{c}_2 &= pM(\frac{1}{2}M - c_2) - q(c_2 - c_3), \\ \dot{c}_i &= pM(c_{i-1} - c_i) - q(c_i - c_{i+1}), \quad 3 \leq i \leq N-2, \\ \dot{c}_{N-1} &= pM(c_{N-2} - c_{N-1}) - qc_{N-1}, \\ \dot{c}_N &= pMc_{N-1}. \end{cases}$$

Nucleation time distribution ($C_1(t) \equiv M$)

$M \to \infty$: Weibull $\frac{M^N}{2(N-2)!} t^{N-2} \exp(-\frac{M^N}{2(N-1)!} t^{N-1})$

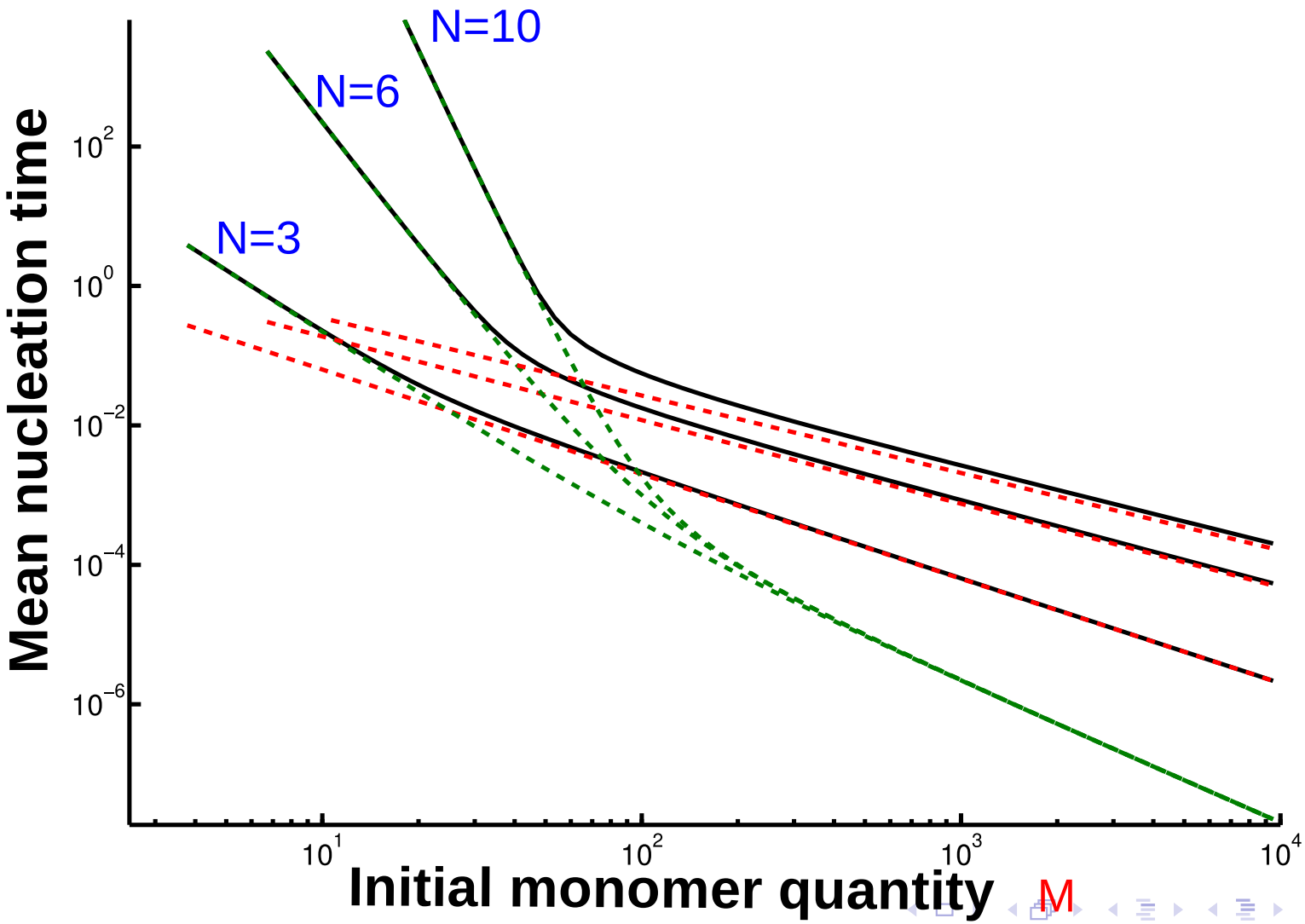
$q \to \infty$: exponential $\frac{M^N}{2q^{N-2}} \exp(-\frac{M^N}{2q^{N-2}} t)$



$N = 6$,
$M = 1000$,

$$q = \begin{cases} 10^2, \\ 10^3, \\ 4.10^3, \\ 10^4. \end{cases}$$

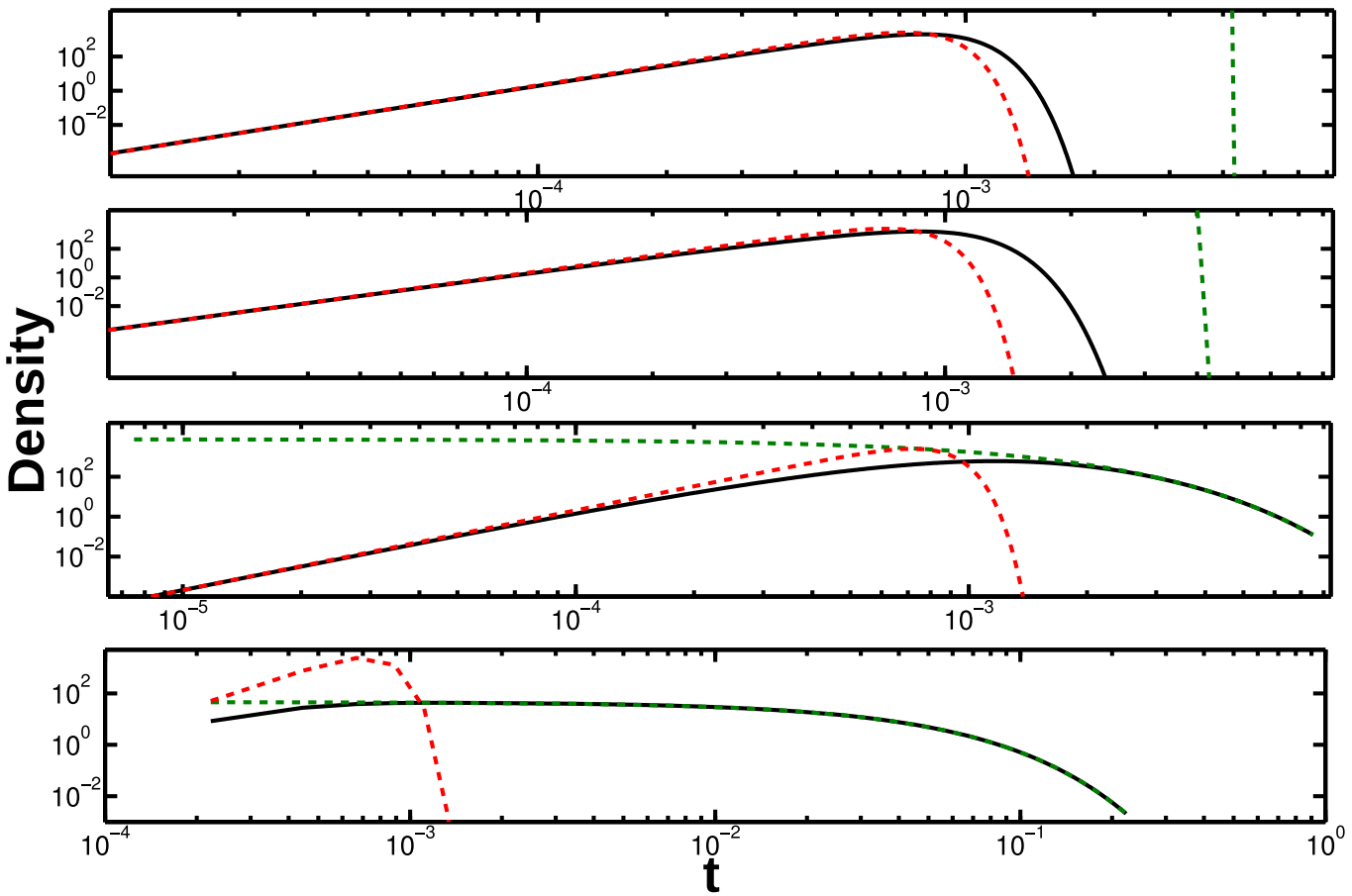
Mean nucleation time versus initial monomer quantity in log scale ($C_1(t) \equiv M$)



Nucleation time distribution in log scale ($C_1(t) \equiv M$)

$M \to \infty$: Weibull $\frac{M^N}{2(N-2)!} t^{N-2} \exp(-\frac{M^N}{2(N-1)!} t^{N-1})$

$q \to \infty$: exponential $\frac{M^N}{2q^{N-2}} \exp(-\frac{M^N}{2q^{N-2}} t)$



$N = 6$,
$M = 1000$,

$$q = \begin{cases} 10^2, \\ 10^3, \\ 4.10^3, \\ 10^4. \end{cases}$$

Mass conservative formulation

We now suppose

$$\sum_{i=1}^{N} iC_i(t) \equiv M.$$

Kolmogorov backward equations $\Rightarrow$ Linear system on

$$S(t, \{C^0\}) = \mathbb{P}\{C_N(t) = 0 \mid C_i(0) = C_i^0\}$$

Problem : Dimension of the system

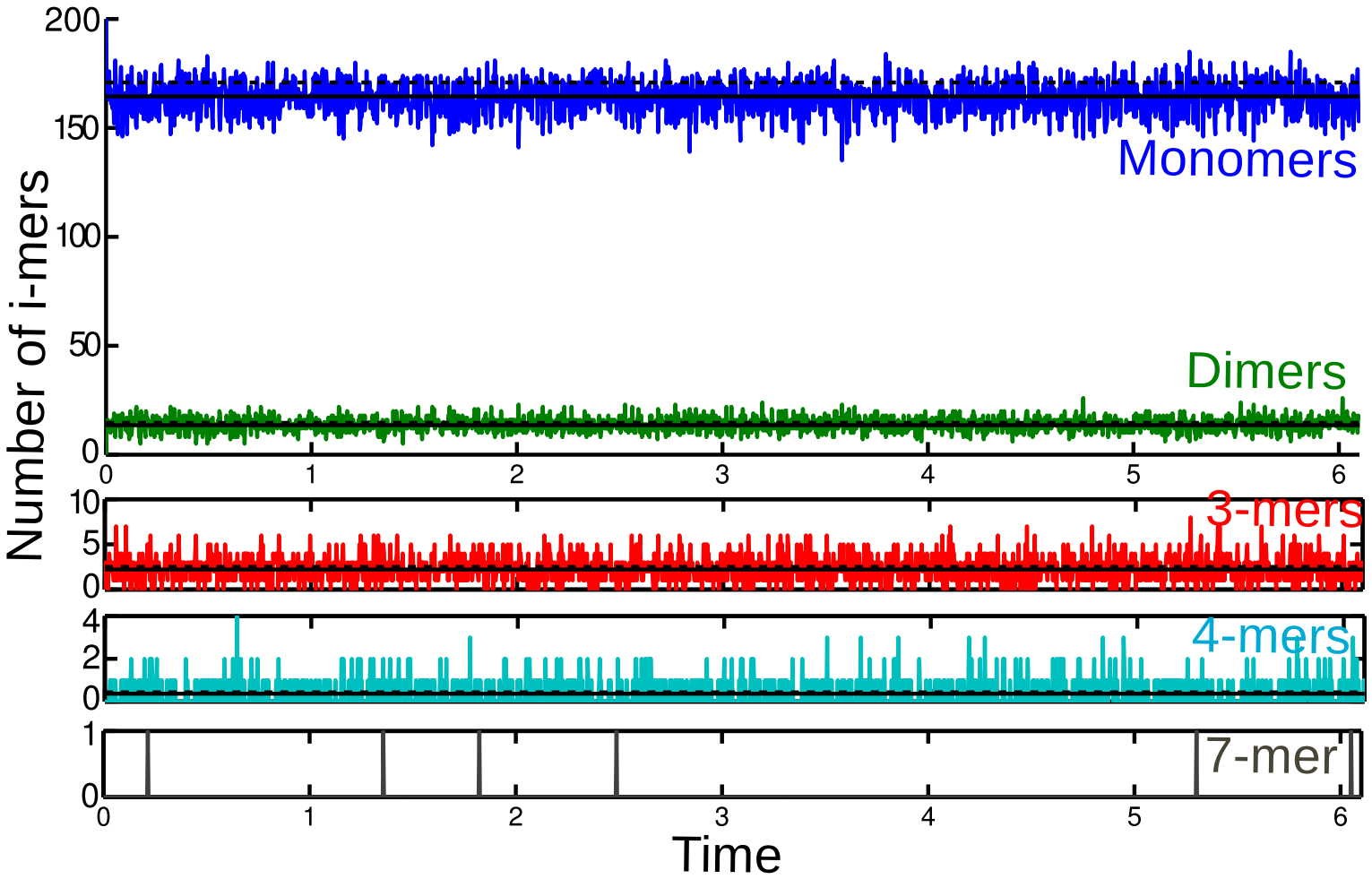
$$\sharp\{\text{configuration } \{C^0\}, \sum_{i=1}^{N} iC_i^0 = M, C_N^0 = 0\} \approx \frac{M^N}{N!}$$

In general, we look for approximate solution for extreme parameter values : $q \gg M$ and $q \ll M$. We use

- known deterministic solution ;
- time scale separation ;
- scaling laws ;
- phase space dimension reduction ;
- linear model ;
- numerical simulation.

Trajectories in the unfavorable case $q \gg M$

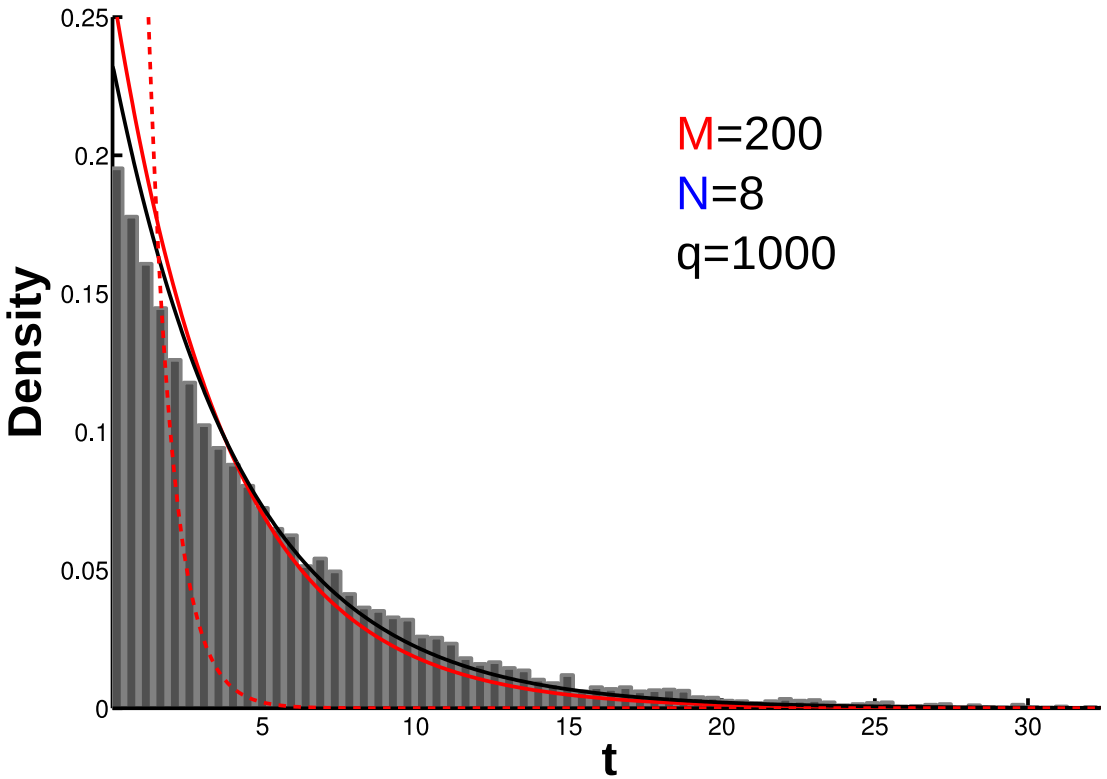
Pre-equilibrium hypothesis (ex : $M = 200, N = 8, q = 1000$).



Nucleation time distribution in the unfavorable case $q \gg M$: exponential law

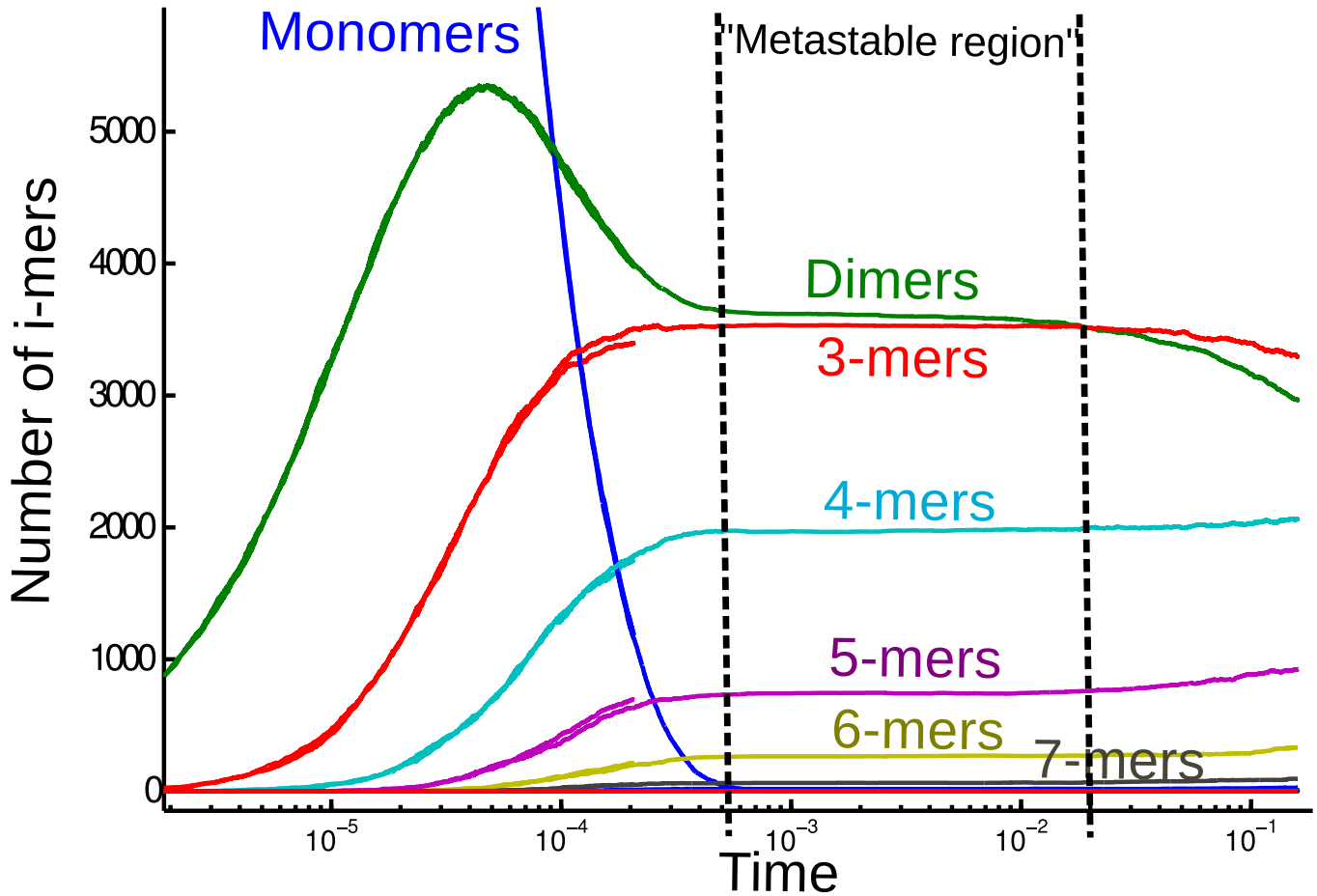
$T_{lag} \sim$ exponential law, of parameter

$$< C_1 C_{N-1} >_{t\to\infty} (M) \approx c_1(t \to \infty) c_{N-1}(t \to \infty) \approx \frac{M^N}{2q^{N-2}}$$



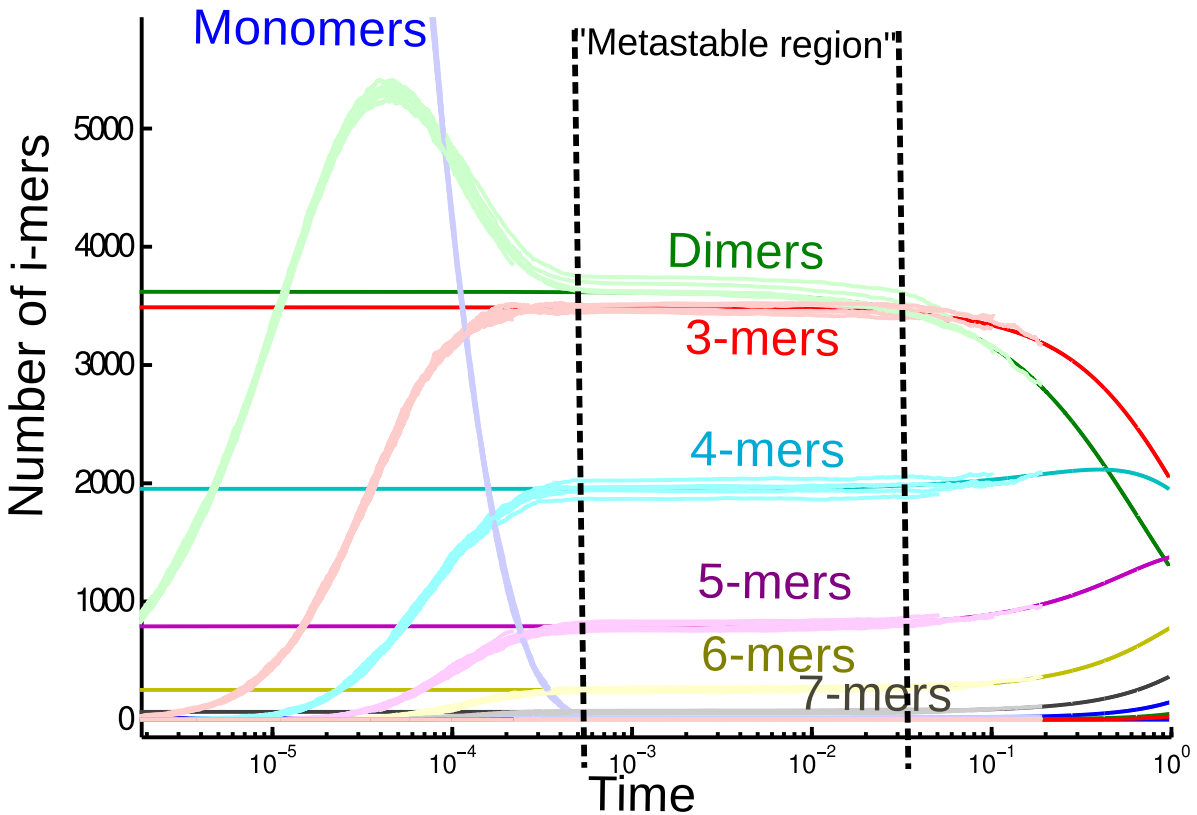
Trajectories in the favorable case $M >> q$ and N large

"Metastable" trajectory (ex : $M = 30000, N = 10, q = 1$).



Trajectories in the favorable case $M >> q$ and N large

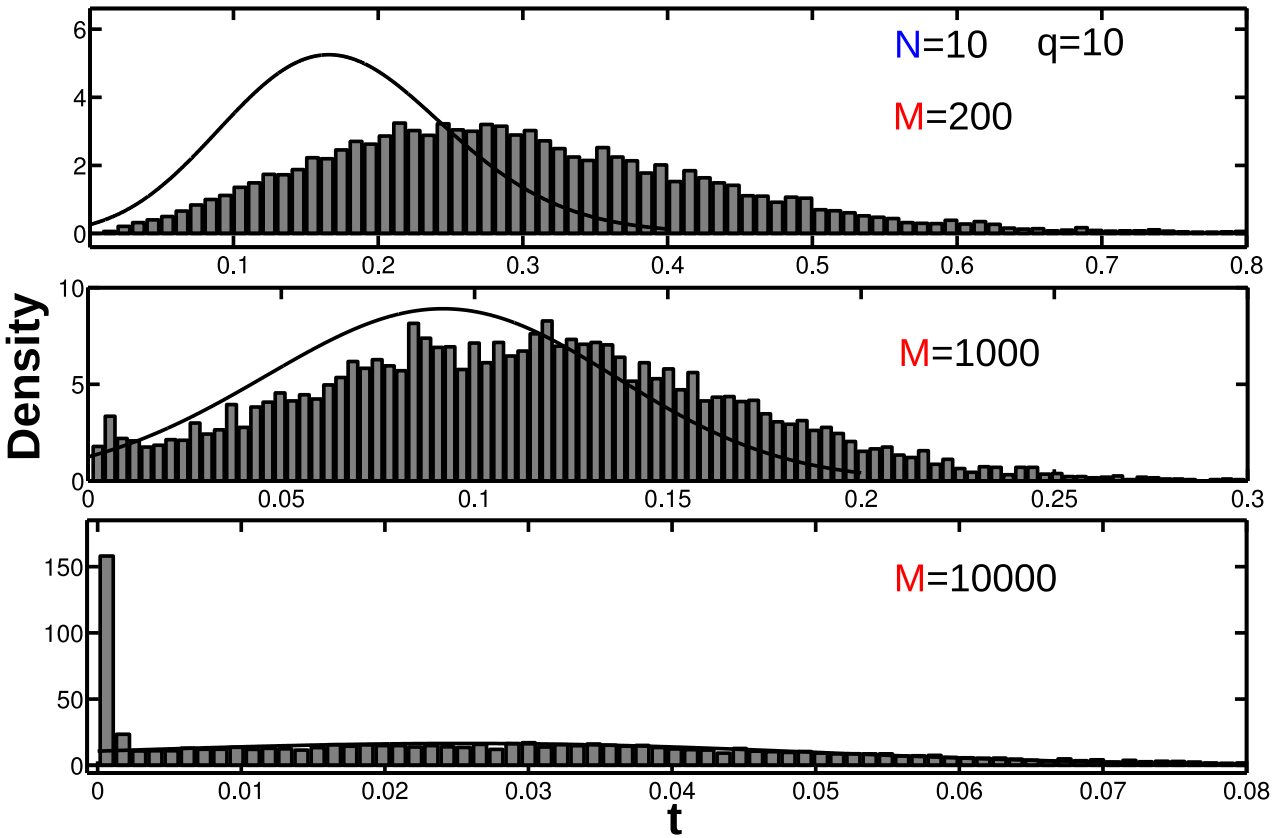
"Metastable" trajectory. Known phenomenon for the deterministic model ([Penrose, 1989],[Wattis, 2006])



1. irreversible aggregation (up to c_i^*)
2. slow "diffusion" with constant monomer $C_1(t) \equiv c_1^*$
3. convergence to equilibrium

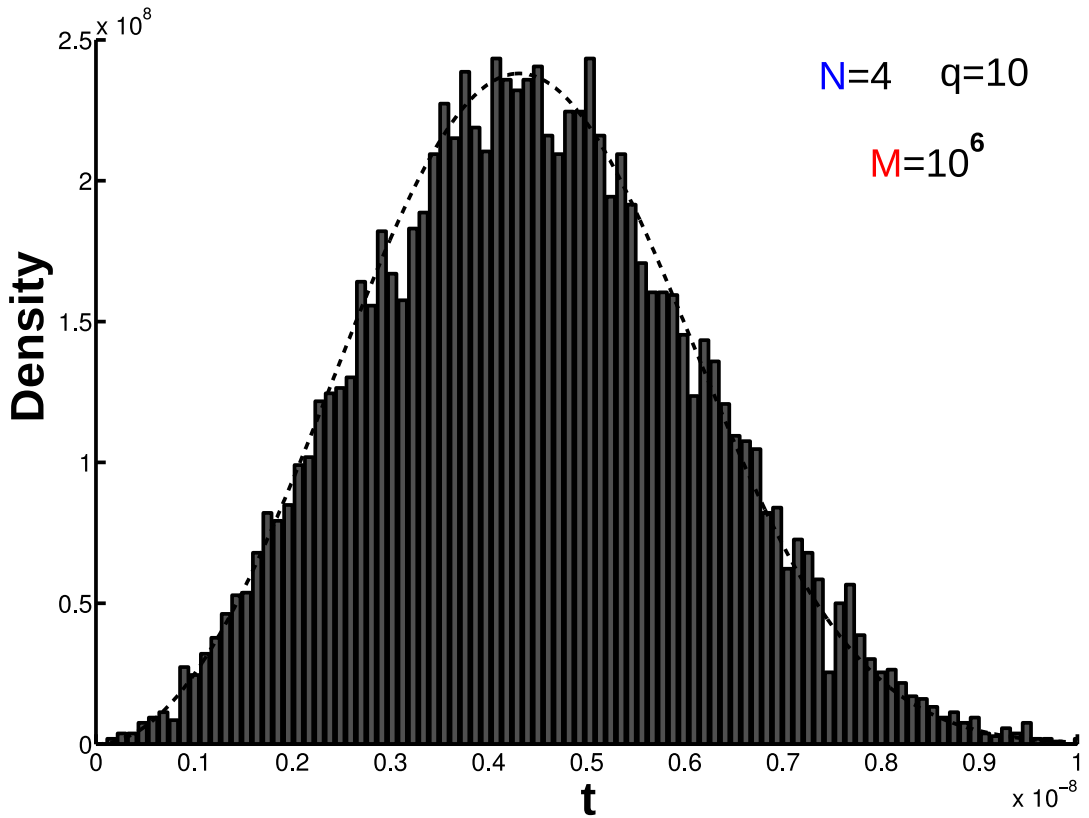
Nucleation time distribution in the favorable case $M >> q$ and N large, $c_N^* < 1$: bimodal distribution

$c_N^* < 1$: Linear model with $C_1 \equiv c_1^*$, $C_i(0) = c_i^*$ (solid line).

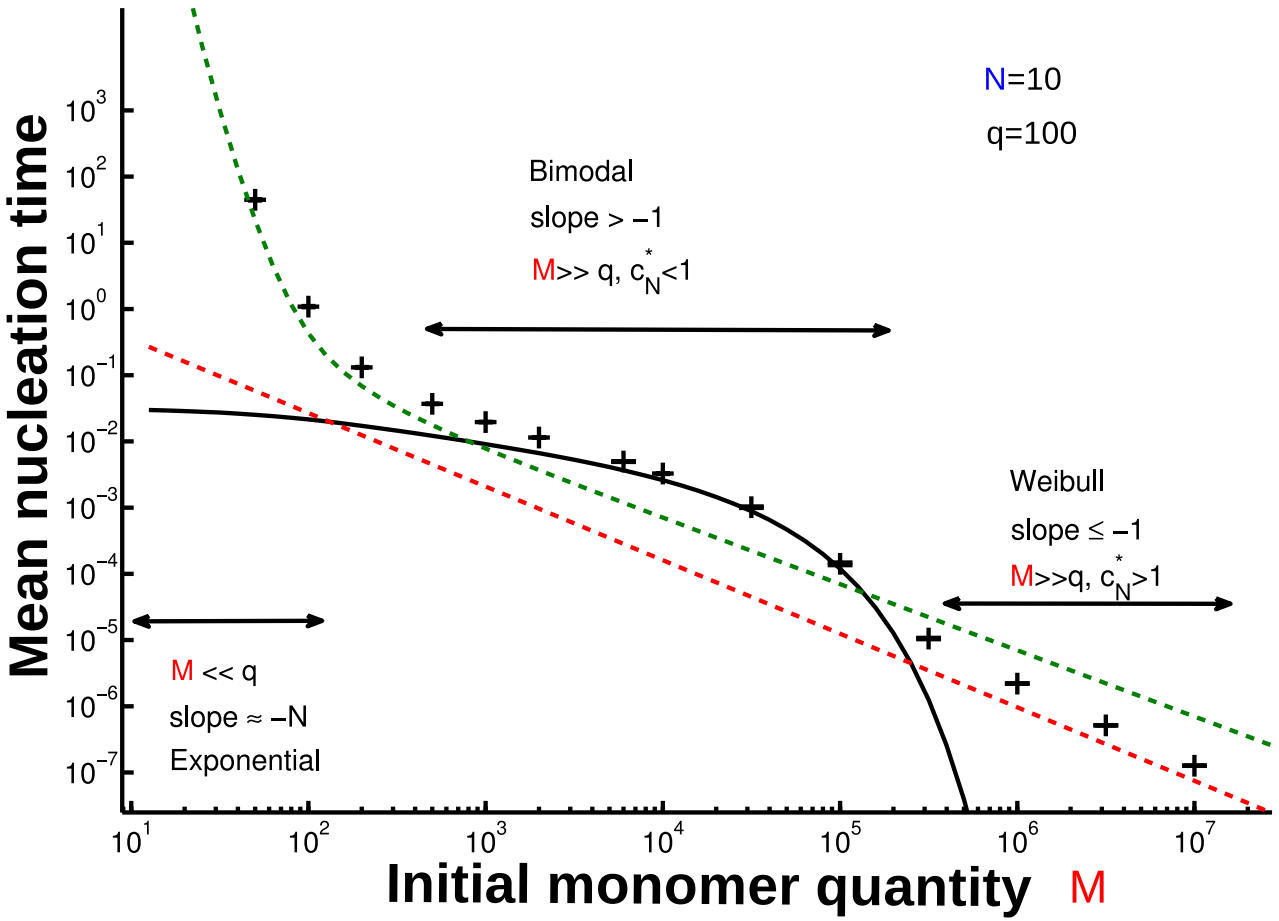


Nucleation time distribution in the favorable case $M >> q$ and N small, $c_N^* > 1$: Weibull law

$c_N^* > 1$: Linear model with $C_1 \equiv M$, Weibull law (dashed line).

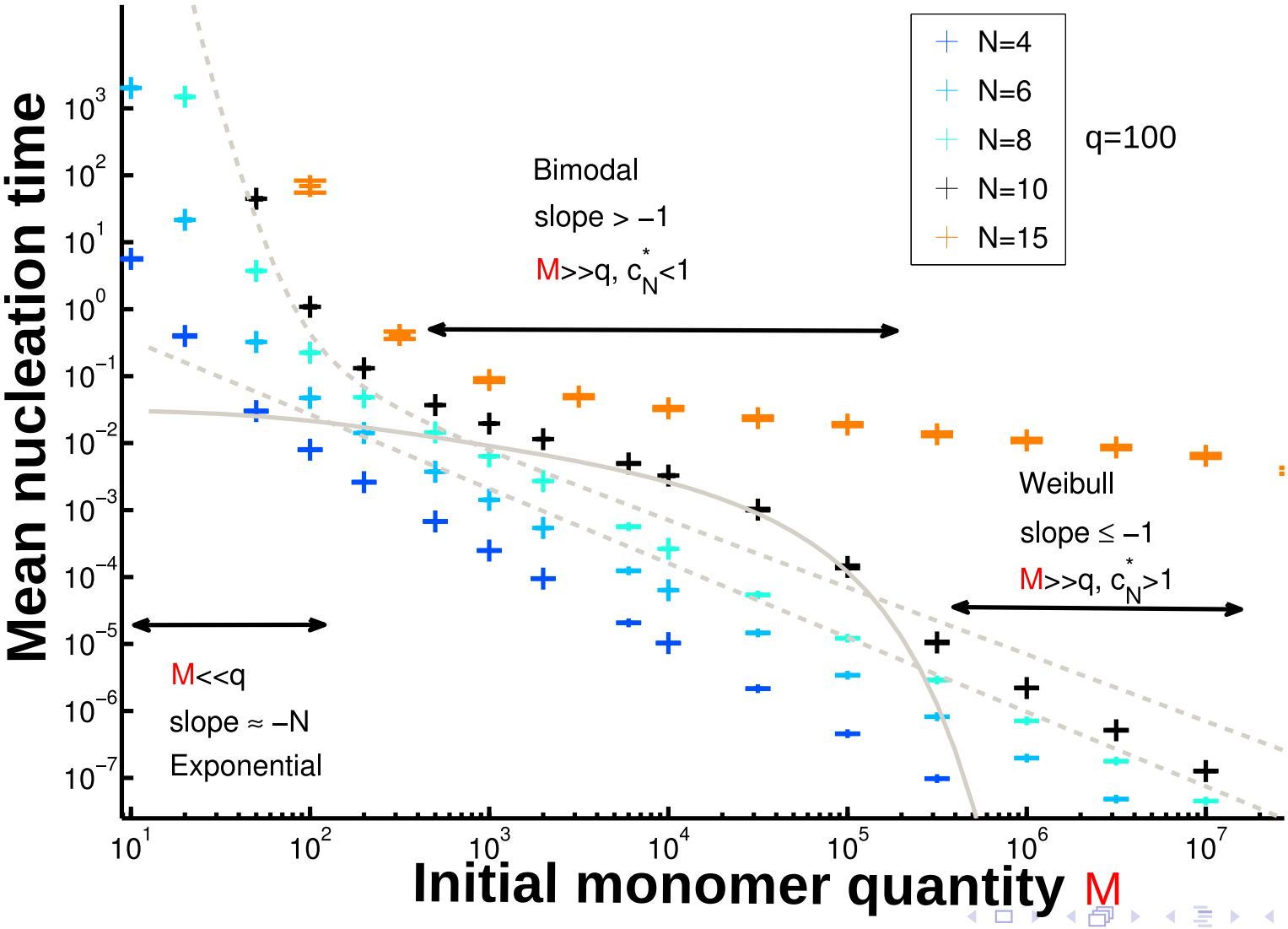


Mean nucleation time versus initial monomer quantity in log scale : 2 or 3 phases according nucleus size N



1. exponential
2. Linear model (starting from c_i^*)
3. Weibull

Mean nucleation time versus initial monomer quantity in log scale : 2 or 3 phases according nucleus size N



Conclusion/Perspectives

- Different behavior of the nucleation time
- Parameter Identifiability depending on parameter region

Conclusion/Perspectives

- Different behavior of the nucleation time
- Parameter Identifiability depending on parameter region

Perspectives

- Different nucleation regime $\Rightarrow$ Different polymerization regime
- Possibility to take into account different polymer structures
- Study the nucleation time for size-dependent rates


Berg, O. G. (1978).
A model for the statistical fluctuations of protein numbers in a microbial population.
J. Theor. Biol., 71(4) :587–603.


Debussche, A., Crudu, A., Muller, A., and Radulescu, O. (2011).
Convergence of stochastic gene networks to hybrid piecewise deterministic processes.
Ann. Appl. Probab. (to appear).


Eldar, A. and Elowitz, M. B. (2010).
Functional roles for noise in genetic circuits.
Nature, 467(7312) :167–73.


Goodwin, B. C. (1965).
Oscillatory behavior in enzymatic control processes.
Adv. Enzyme. Regul., 3 :425–437.


Hasty, J., McMillen, D., Isaacs, F., and Collins, J. J. (2001).

Computational studies of gene regulatory networks : in numero molecular biology.
Nat. Rev. Genet., 2(4) :268–279.


Kang, H.-W. and Kurtz, T. G. (2011).
Separation of time-scales and model reduction for stochastic reaction networks.
Ann. Appl. Probab. (to appear).


Kepler, T. and Elston, T. (2001).
Stochasticity in transcriptional regulation : Origins, consequences, and mathematical representations.
Biophys. J., 81 :3116–3136.


Lansbury, P. J. and Caughey, B. (1995).
The chemistry of scrapie infection : implications of the 'ice 9' metaphor.
Chem. Biol., 2(2) :1–5.


Lipniacki, T., Paszek, P., Marciniak-Czochra, A., Brasier, A. R., and Kimmel, M. (2006).
Transcriptional stochasticity in gene expression.
J. Theor. Biol., 238(2) :348–367.


Paszek, P. (2007).
Modeling stochasticity in gene regulation : characterization in the terms of the underlying distribution function.
B. Math. Biol., 69(5) :1567–601.


Paulsson, J. (2005).
Models of stochastic gene expression.
Phys. Life Rev., 2(2) :157–175.


Peccoud, J. and Ycart, B. (1995).
Markovian modelling of Gene Product Synthesis.
Theor. Popul. Biol., 48(2) :222–234.


Penrose, O. (1989).
Metastables states for the becker-döring cluster equations.

Commun. Math. Phys., 541 :515–541.


Shahrezaei, V. and Swain, P. (2008).
Analytic distributions for stochastic gene expression.
Proc. Natl. Acad. Sci., 105 :17256–17261.


Song, C., Phenix, H., Abedi, V., Scott, M., Ingalls, B. P., Perkins, T. J., and Kaern, M. (2010).
Estimating the Stochastic Bifurcation Structure of Cellular Networks.
PLoS Comput. Biol., 6 :e1000699/1–11.


Tyran-Kamińska, M. (2009).
Substochastic semigroups and densities of piecewise deterministic markov processes.
J. Math. Anal. Appl., 357(2) :385–402.


Wattis, J. A. D. (2006).

An introduction to mathematical models of coagulation-fragmentation processes : a discrete deterministic mean-field approach.
Physica D, 222(1-2) :1–20.

Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X. (2006).
Probing gene expression in live cells, one protein molecule at a time.
Science, 311 :1600–1603.