# *Physics of self-assembly and signaling*

Tom Chou

UCLA

Depts. of Biomathematics & Mathematics

NSF DMS-1021818

# Physics of self-assembly and signaling

**I. Self-assembly:**

– evolution of yeast clusters (Mike Travisano)

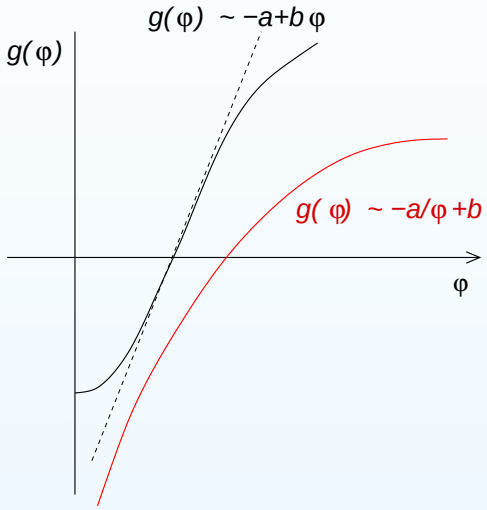
– cooperation and aggregation of daughters (Corina Tarnita)

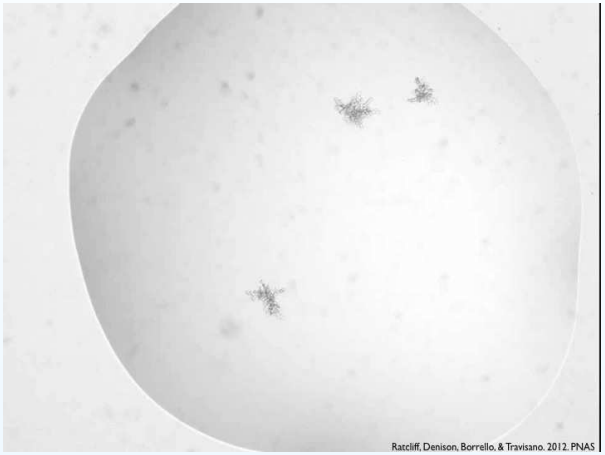
**II. Signaling, Quorum sensing:**

– quorum sensing (Avigdor Eldar, Hyun Youk)

– diffusion sensing (R. Redfield) and a toy model

# I. Self assembly $\Leftrightarrow$ yeast budding/aggregation



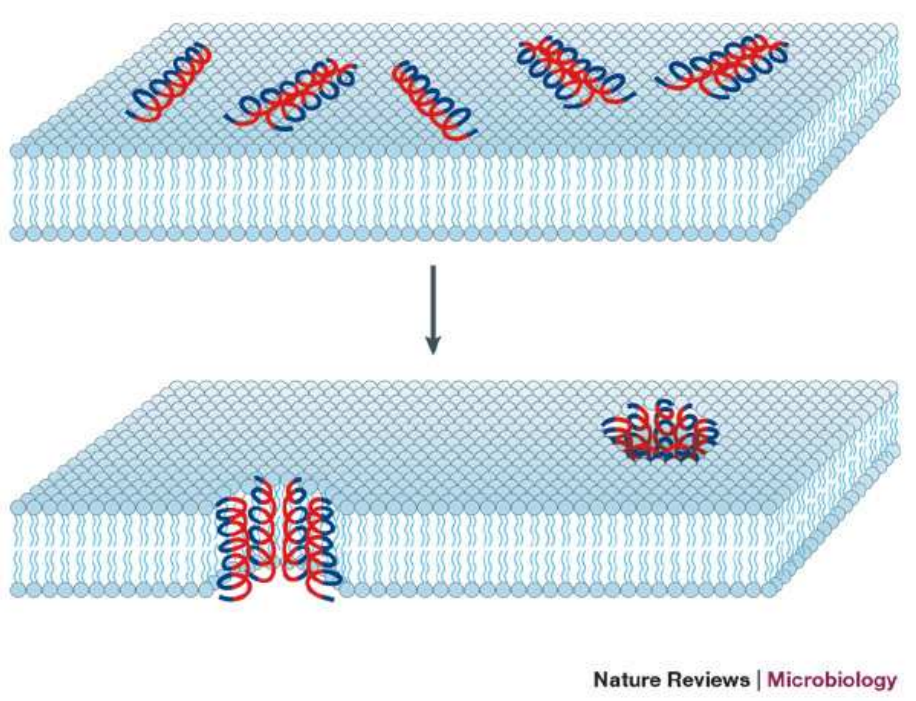

Growth law: $\dot{M}(t) = g(\varphi(t))M(t)$, Resources: $\varepsilon\dot{\varphi}(t) = \lambda - M\varphi$

Constraints: $g(\varphi) \approx -a + b\varphi$, $g(\varphi) \approx -a/\varphi + b$

Effective growth law: $\dot{M}(t) \approx b\lambda - aM(t)$, $\dot{M}(t) \approx b(1 - \frac{a}{b\lambda}M)M$

growth/self-assembly of clusters with total mass contraint

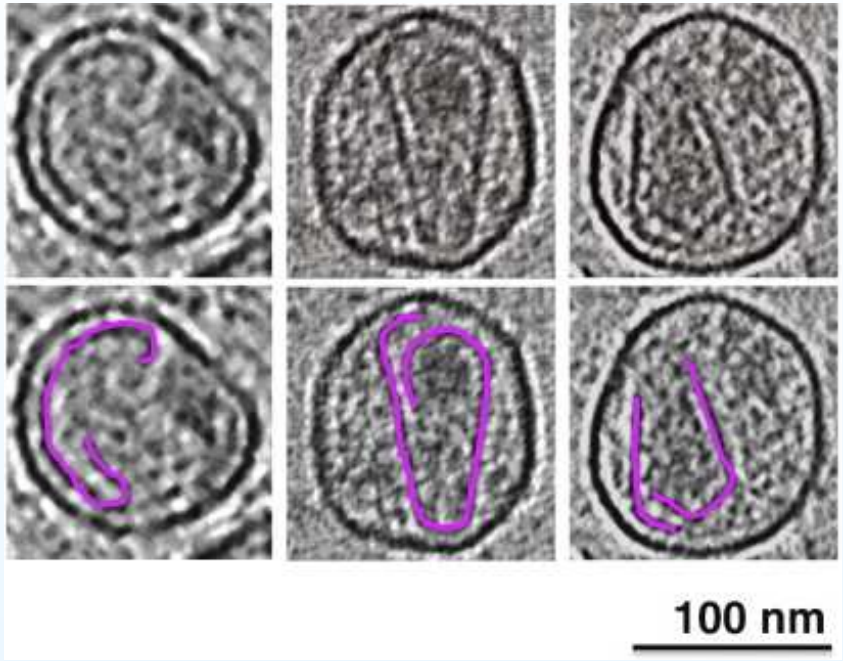
# I. Self-assembly in biology: peptide aggregation




Integral membrane or membrane-associated peptides self-assemble into antimicrobial pores.

maximum cluster size: $N \sim 4 - 8$

# I. Self-assembly in biology: virus assembly

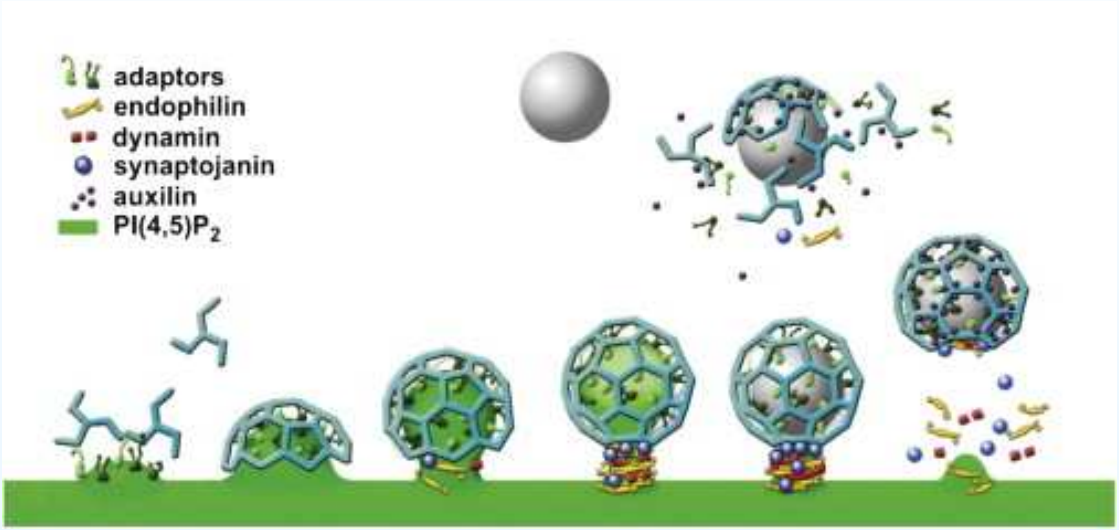


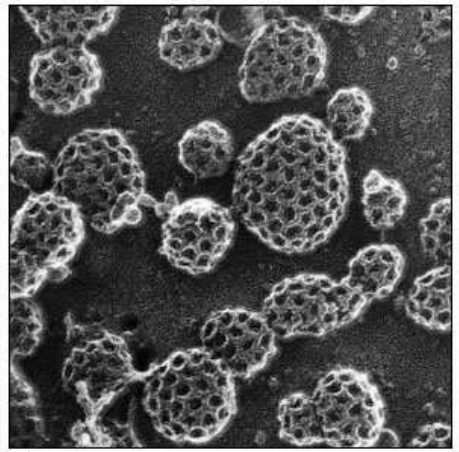

Z. Yu *et al.*, Unclosed HIV-1 Capsids Suggest a Curled Sheet Model of Assembly, J. Mol. Biol., **425**, 112, (2013).

D. C. Rapaport, Simulation of capsid assembly in closed volumes (cells), Phys. Biol. **7**, 045001 (2010)

maximum cluster size: $N \sim 100 - 1000$

# I. Self-assembly in biology: Clathrin-coated pits




Banerjee, Berezhkovskii, and Nossal, *Biophys. J.*, **102**, (2012).

maximum cluster size: $N \sim 25 - 50$

# I. What's been done:

- Mass-action equations for mean size distribution used in many contexts:

  Virus capsid assembly: Morozov, Bruinsma, Rudnick, *J. Chem. Phys.*, **131**, 155101, 2009

  P. L. Krapivsky, E. Ben-Naim, and S. Redner, *Statistical Physics of Irreversible Processes*, CUP, 2010

- Extensive work on asymptotic analysis of mass-action, Becker-Döring eqs:

  P.-E. Jabin and B. Niethammer, *J. Differential Equations*, **191**, 518-543, 2003

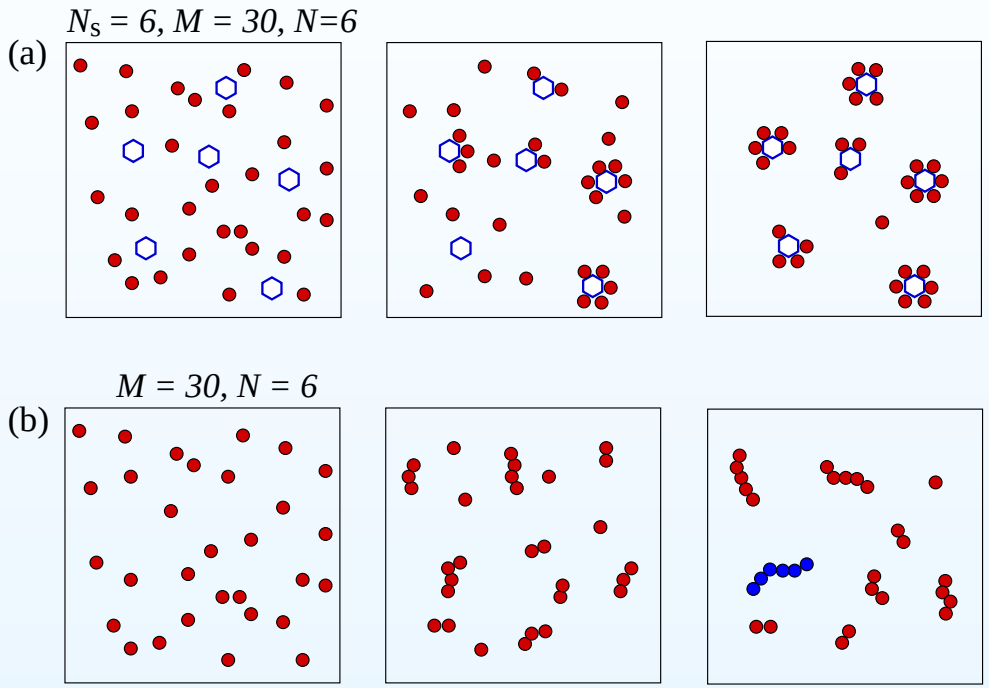
  J. A. D. Wattis and J. R. King, *J. Phys. A: Math. Gen.*, **31**, 7169-7189, 1998

- Very little on full stochastic treatments of nucleation:

  J. S. Bhatt and I. J. Ford, *J. Chem. Phys.*, **118**, 3166-3176, 2003

  F. Schweitzer *et al.*, *Physica A*, **150**, 261-279, 1988

# I. Heterogeneous vs. homogeneous nucleation




$N_s$ = number of seeds

$M$ = total mass

$N$ = maximum cluster size

(a) Heterogeneous nucleation: growth on preexisting seeds (hexagons)

(b) Homogeneous nucleation: spontaneous dimer formation

D'Orsogna, Lakatos, Chou, *J. Chem. Phys.*, **136**, (2012)

Chou & D'Orsogna, *PRE*, **84**, (2011).

# I. Heterogeneous nucleation: mass action

Assume fixed number of seeds $N_\mathrm{s}$, $p_k = p$, $q_k = q$, and $q/p \equiv \varepsilon \ll 1$. Mass-action equations for concentration $c_k(t)$ of seeds with $k$ bound "ligands":

$$\begin{aligned}
\dot{c}_0 &= -m(t)c_0 + \varepsilon c_1, \\
\dot{c}_k &= -m(t)c_k - \varepsilon c_k + m(t)c_{k-1} + \varepsilon c_{k+1}, \\
\dot{c}_N &= -\varepsilon c_N + m(t)c_{N-1},
\end{aligned}$$

Constraints: $m(t) \equiv M - \sum_{k=1}^{N} k c_k(t)$, $N_\mathrm{s} = \sum_{j=0}^{N} c_j(t)$,

Initial conditions: $c_k(t=0) = N_\mathrm{s}\delta_{k,0}$, $m(t=0) = M$.

Many asymptotic results known

# I. Heterogeneous nucleation: Master eqn

Full probability for $m$ free monomers, $n_0$ free seeds, $n_1$ seeds bound to one monomer, $n_i$ seeds bound to $i$ monomers:

$$\dot{P}(m,\{n\};t) = -\Lambda(m,\{n\})P(m,\{n\};t) + \sum_{i=0}^{N-1}(m+1)(n_i+1)W_i^+W_{i+1}^-P(m,\{n\};t) + \varepsilon\sum_{i=1}^{N}(n_i+1)W_{i-1}^-W_i^+P(m,\{n\};t)$$

where $\Lambda \equiv m\sum_{i=0}^{N-1}(n_i+\varepsilon n_{i+1})$ and

$$W_i^+W_{i+1}^-P(m,\{n\};t) = P(m+1,n_0,n_1,\ldots,n_i+1,n_{i+1}-1,\ldots,n_N;t)$$

# I. Heterogeneous nucleation: Master eqn

Constraints:

$$M = m + \sum_{k=1}^{N} k n_k \quad \text{and} \quad N_{\rm s} = \sum_{k=0}^{N} n_k .$$

Initial condition:

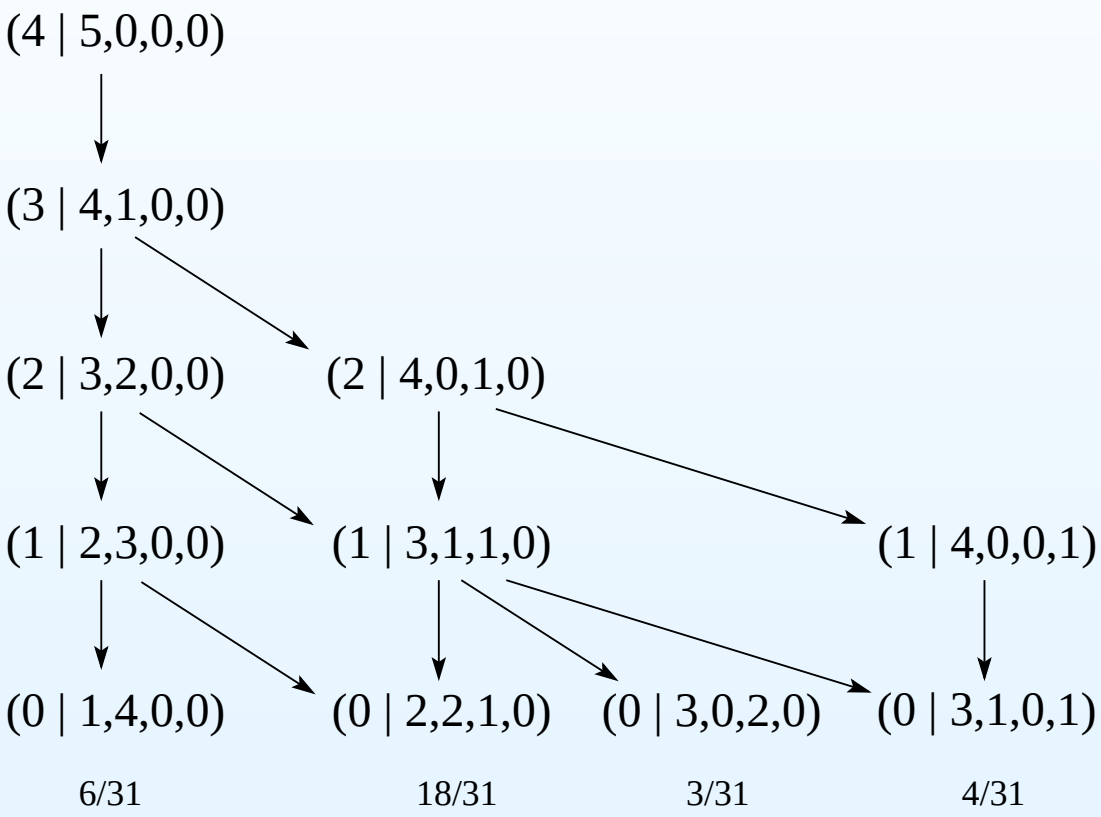
$$P(m, \{n\}; t = 0) = \delta_{m,M} \delta_{n_0, N_{\rm s}} \delta_{n_{i>1}, 0}$$

Mean values:

$$\langle n_k(t) \rangle = \sum_{m, \{n\}} n_k P(m, \{n\}; t)$$

# Heterogeneous nucleation: mass-action vs. ME

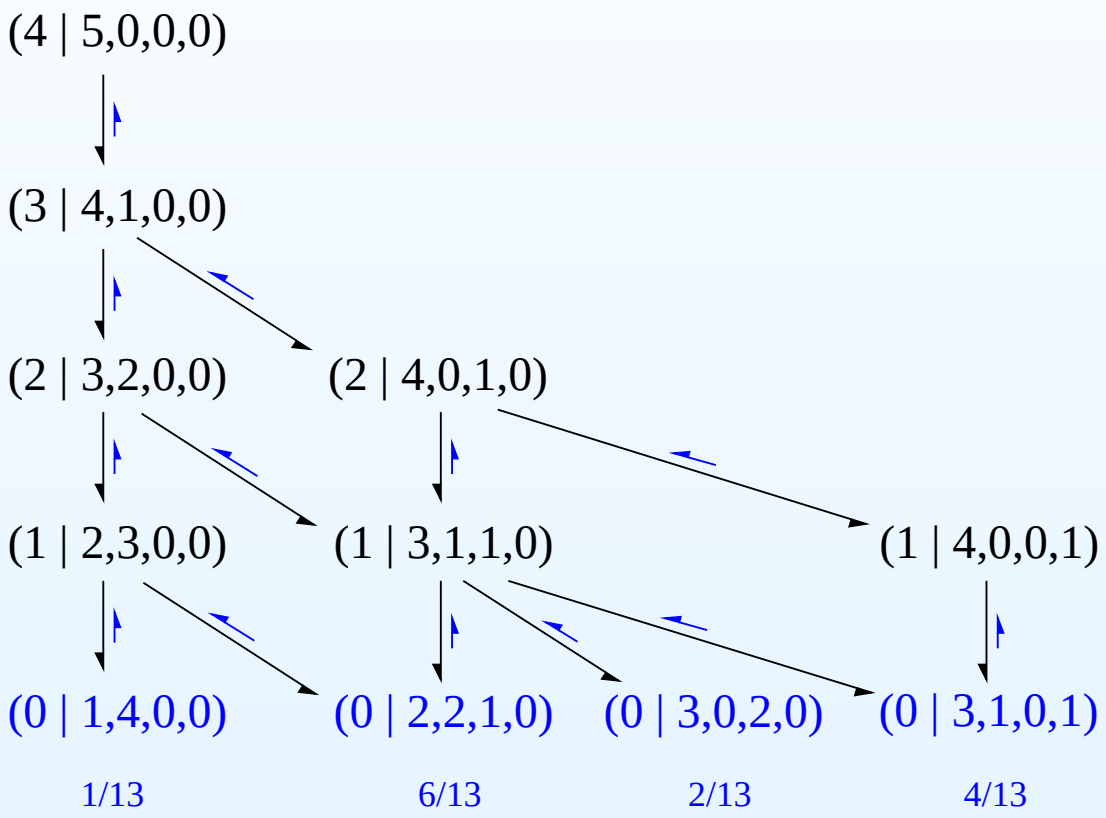
$\varepsilon = 0$, irreversible kinetics $\Rightarrow$ quenched cluster distribution:




- Exact, explicit relations derived for $\langle n_k^* \rangle$.

# I. Heterogeneous nucleation: mass-action vs. ME

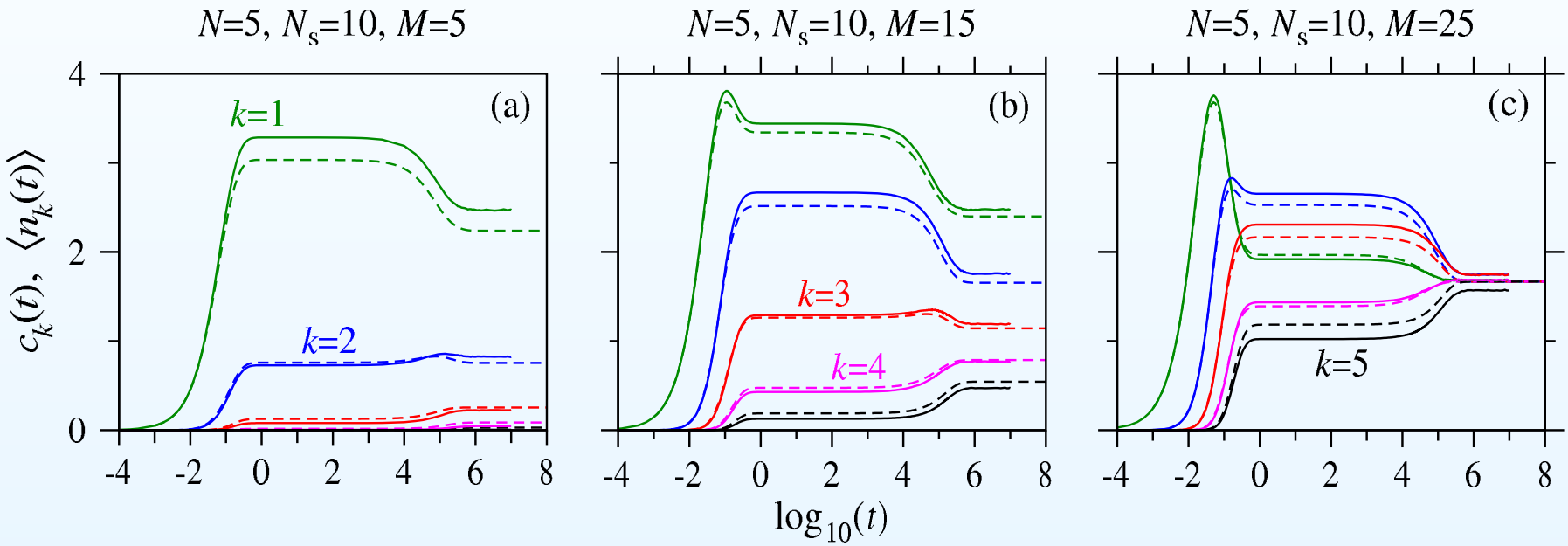
$\varepsilon > 0 \Rightarrow$ equilibrium distribution after $t \sim 1/\varepsilon$:




- Recursion relations derived for $\langle n_k^{\mathrm{eq}} \rangle$

# I. Heterogeneous nucleation: mass-action vs. ME

$\sigma \equiv M/(N_s N)$:



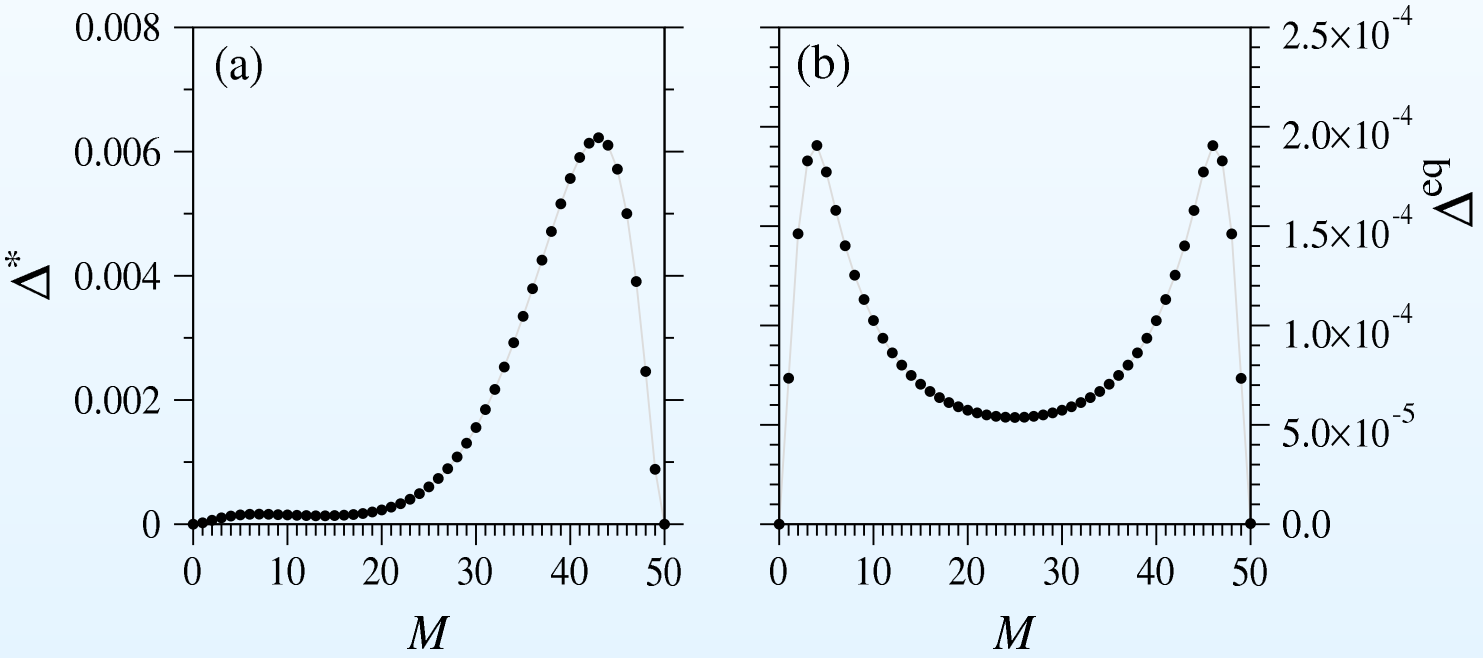

$\sigma = 0.1$ $\sigma = 0.3$ $\sigma = 0.5$

- Mass-action models provide reasonably accurate means

# I. Heterogeneous nucleation: small errors

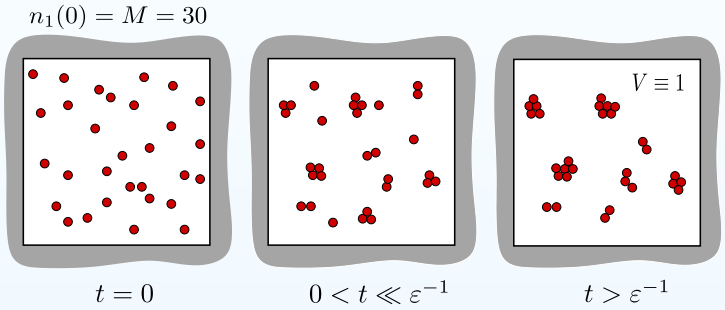
Define a standard error between mass-action and discrete model:

$$\Delta(t) \equiv \frac{1}{N+1} \sum_{k=0}^{N} \left| \frac{\langle n_k(t) \rangle}{N_s} - \frac{c_k(t)}{N_s} \right|^2 .$$




Errors typically small....what about homogeneous nucleation?

# I. Homogeneous nucleation: mass-action




$c_k(t)$ : conc. of size-$k$ clusters

Becker-Döring mass-action eqs:

$$\dot{c}_1 = -c_1^2 - c_1 \sum_{j=2}^{N-1} c_j + 2\varepsilon c_2 + \varepsilon \sum_{j=3}^{N} c_j$$

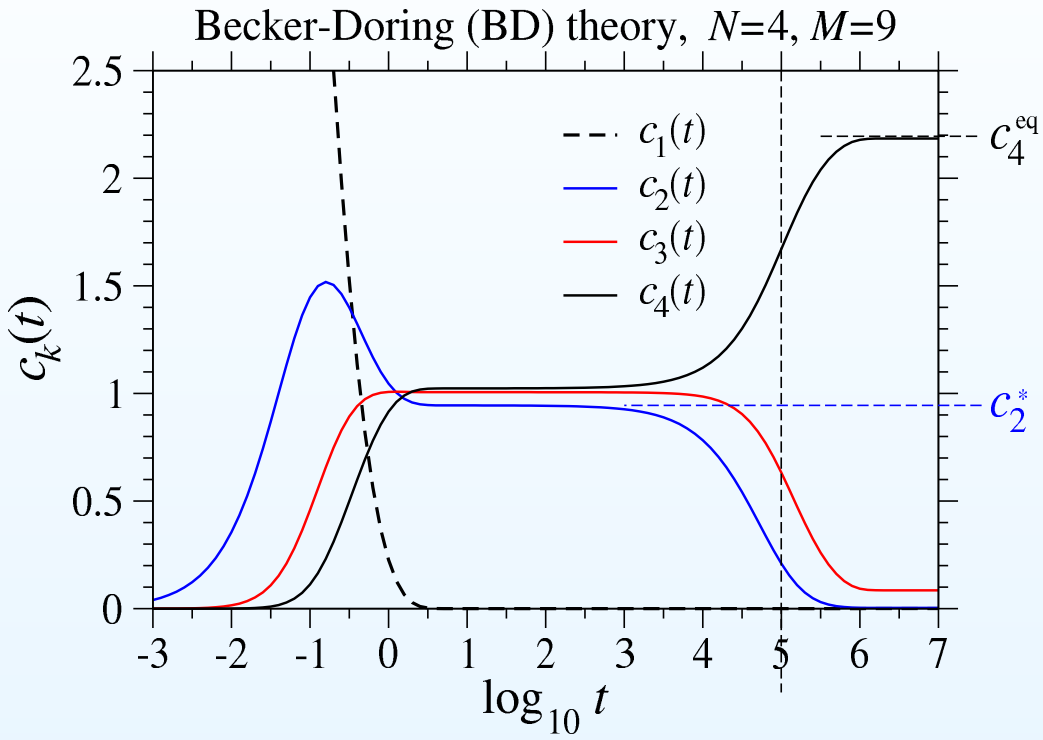
$$\dot{c}_2 = -c_1 c_2 + \tfrac{1}{2} c_1^2 - \varepsilon c_2 + \varepsilon c_3$$

$$\dot{c}_k = -c_1 c_k + c_1 c_{k-1} - \varepsilon c_k + \varepsilon c_{k+1}$$

$$\dot{c}_N = c_1 c_{N-1} - \varepsilon c_N,$$

detachment rate/attachment rate $= q/p \equiv \varepsilon \ll 1$

# I. Homogeneous nucleation: mass-action




Numerics: $N = 4$, $M = 9$, $\varepsilon = 10^{-5}$

Coarsening time $t_{\rm c} \sim 1/\varepsilon$ : $0 \to c_k^* \to c_k^{\rm eq}$

Largest cluster dominates: $c_k^{\rm eq} \approx \frac{\varepsilon}{2}\left(\frac{2M}{\varepsilon N}\right)^{k/N} + O(\varepsilon^{1-k/N-1/N})$

# I. Homogeneous nucleation: Master Equation

- Consider finite size using discrete Master Equation.

- Define Prob. $n_1$ monomers, $n_2$ dimers,...at $t$: $P(n_1, n_2, n_3, \ldots, n_N; t)$

$$\dot{P}(\{n\};t) = -\Lambda(\{n\})P(\{n\};t) + \tfrac{1}{2}(n_1+2)(n_1+1)W_1^+W_1^+W_2^-P(\{n\};t)$$
$$+\varepsilon(n_2+1)W_2^+W_1^-W_1^-P(\{n\};t)$$
$$+\sum_{i=2}^{N-1}(n_1+1)(n_i+1)W_1^+W_i^+W_{i+1}^-P(\{n\};t)$$
$$+\varepsilon\sum_{i=3}^{N}(n_i+1)W_1^-W_{i-1}^-W_i^+P(\{n\};t),$$

Total rate out of configuration $\{n\}$ is

$$\Lambda(\{n\}) = \frac{1}{2}n_1(n_1-1) + \sum_{i=2}^{N-1} n_1 n_i + \varepsilon\sum_{i=2}^{N} n_i,$$

# I. Homogeneous nucleation: Master Equation

Raising and lower operators $W_i^{\pm}$. For example,

$$W_1^+ W_i^+ W_{i+1}^- P(\{n\};t) = P(n_1+1,\ldots,n_i+1,n_{i+1}-1,\ldots;t).$$

represents detachment of a monomer from a $(i+1)$-cluster, forming an additional free monomer and a $(i)$-cluster

Initial condition: $P(\{n\};t=0) = \delta_{n_1,M}\delta_{n_2,0}\cdots\delta_{n_N,0}$

Mass conservation: $M = \sum_{j=1}^{N} jn_j$

With definition of mean: $\langle n_k(t)\rangle \equiv \sum'_{\{n_j\}} n_k P(\{n\};t)$, derive moment hierarchy:

# I. Homogeneous nucleation: moment hierarchy

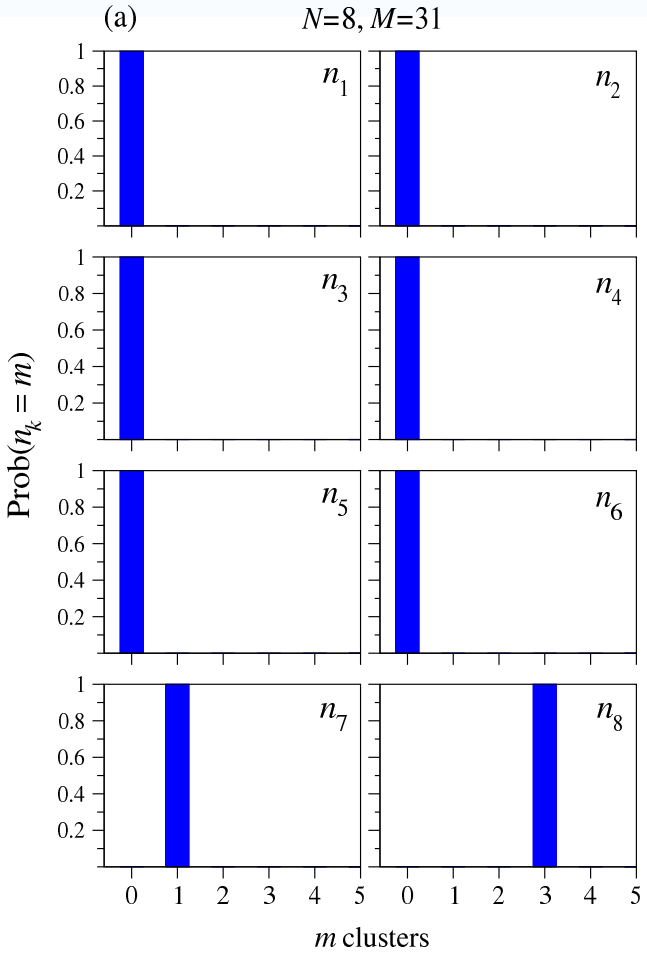
$$\begin{aligned}
\langle \dot{n}_1(t) \rangle &= -2 \left\langle \frac{n_1(n_1-1)}{2} \right\rangle - \sum_{j=2}^{N-1} \langle n_1 n_j \rangle + 2\varepsilon \langle n_2(t) \rangle + \varepsilon \sum_{j=3}^{N} \langle n_j \rangle \\
\langle \dot{n}_2(t) \rangle &= -\langle n_1 n_2 \rangle + \left\langle \frac{n_1(n_1-1)}{2} \right\rangle + \varepsilon \langle n_3 \rangle - \varepsilon \langle n_2 \rangle \\
\langle \dot{n}_k(t) \rangle &= -\langle n_1 n_k \rangle + \langle n_1 n_{k-1} \rangle - \varepsilon \langle n_k \rangle + \varepsilon \langle n_{k+1} \rangle \\
\langle \dot{n}_N(t) \rangle &= \langle n_1 n_{N-1} \rangle - \varepsilon \langle n_N \rangle .
\end{aligned}$$

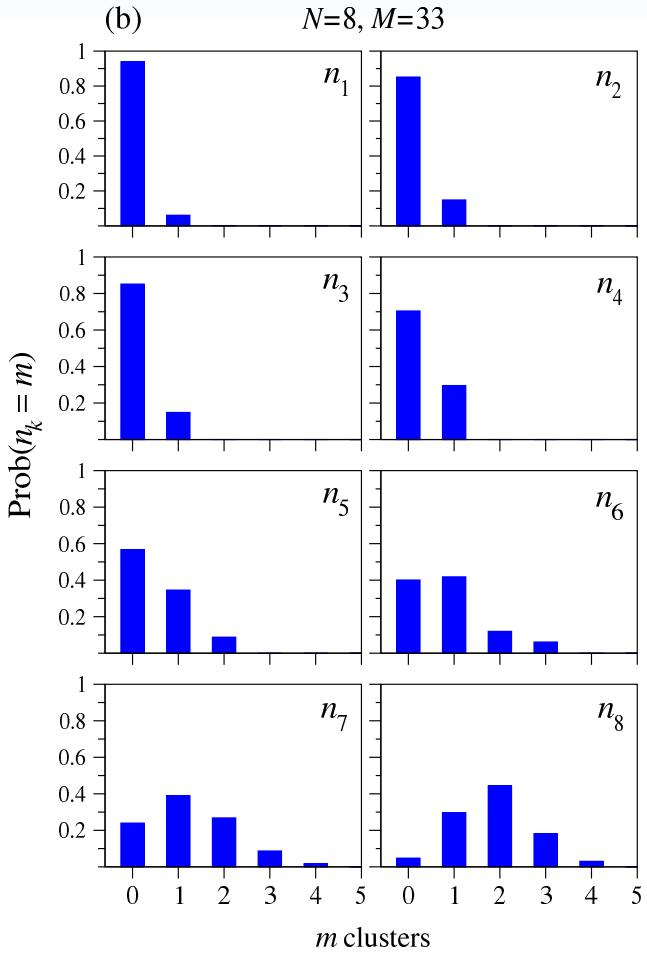
Reduces to Becker-Döring when: $\langle n_i n_j \rangle = \langle n_i \rangle \langle n_j \rangle \Rightarrow c_i c_j ,$

$$\langle n_1 - 1 \rangle \approx \langle n_1 \rangle \Rightarrow c_1 .$$

# I. Kinetic Monte-Carlo simulations

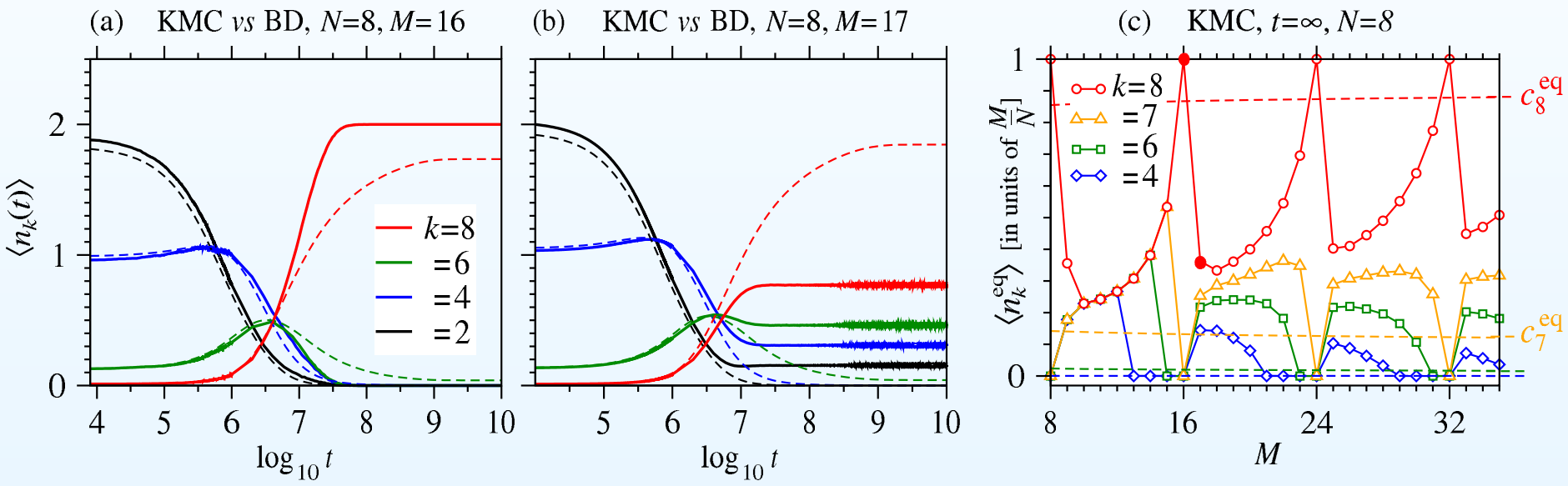
Use KMC to simulate full *distribution* of $n_k$. At equilibrium:



# I. Kinetic Monte-Carlo simulations

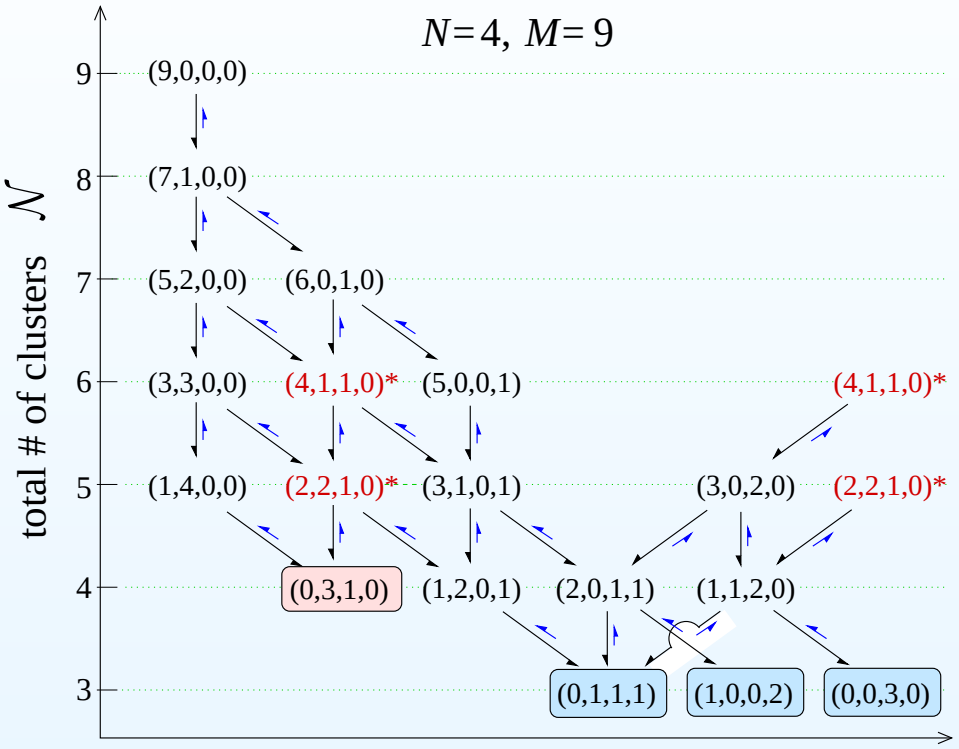
Compare BD ($c_k$) with KMC ($\langle n_k \rangle$) for $N = 8$ and $M = 16, 17$:




- $c_k$ does not compare well with $\langle n_k \rangle$, especially at equilibrium

# I. Homogeneous nucleation: exact equilibrium solution

How to solve for $\langle n_k^{\rm eq} \rangle$?




Enumerate states with minimum number of clusters:

- For the $N = 4, M = 9$, $\mathcal{N}_{\rm min} = 3$.
- Identify connections with $\mathcal{N}_{\rm min} + 1 = 4$, apply detailed balance:

# I. Homogeneous nucleation: exact equilibrium solution

Define

$$M = \sigma N - j \;\; (0 \leq j \leq N-1).$$

$\sigma - 1$: largest integer divisor

$j$: incommensurability or remainder

Exact solution

$$\begin{aligned} \langle n_N^{\rm eq} \rangle &= \frac{\sigma(\sigma-1)}{(\sigma+j-1)} + O(\varepsilon) \\ \langle n_{N-k}^{\rm eq} \rangle &= \frac{\sigma(\sigma-1)j(j-1)\ldots(j-k+1)}{(\sigma+j-1)(\sigma+j-2)\ldots(\sigma+j-k-1)} + O(\varepsilon) \end{aligned}$$

valid for $0 \leq j < N-1$ and all $k$.

# I. Homogeneous nucleation: exact equilibrium solution

For $j = N - 1$ (one additional monomer)

$$\langle n_1^{\rm eq} \rangle = \frac{2(N-1)!}{D(\sigma, N-1)} + O(\varepsilon)$$

$$\langle n_{N-k}^{\rm eq} \rangle = \frac{\prod_{\ell=1}^{k}(N-\ell)\prod_{i=1}^{N-k-1}(\sigma-2+i)}{D(\sigma, N-1)} + O(\varepsilon)$$

$$\langle n_N^{\rm eq} \rangle = (\sigma-1)\frac{D(\sigma-1, N-1)}{D(\sigma, N-1)} + O(\varepsilon),$$

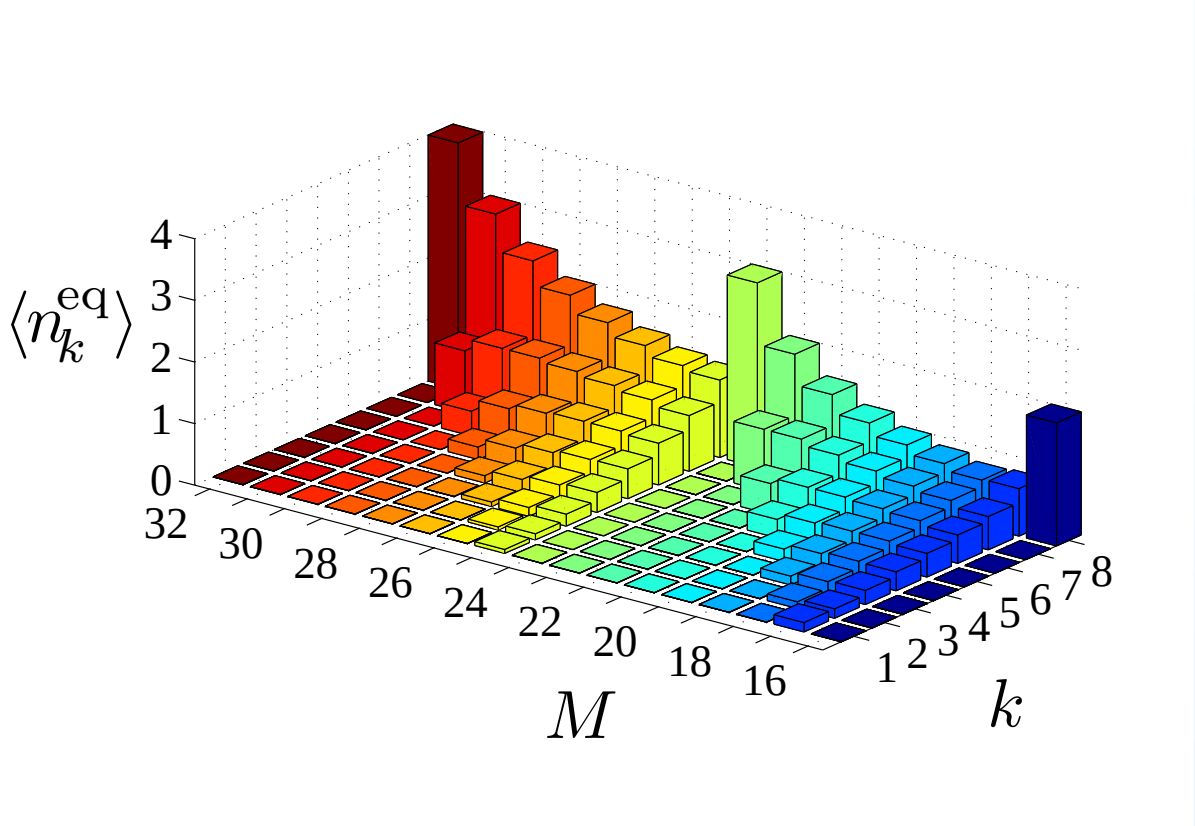
where

$$D(\sigma, j) = j! + \prod_{\ell=1}^{j-1}(\sigma+\ell).$$

Results verify KMC simulations in $\varepsilon \to 0^+$ limit

# I. Homogeneous nucleation: exact equilibrium solution

$\langle n_k^{\mathrm{eq}} \rangle$ as a function of total mass $M$: cluster dispersal




Mass in largest cluster only when $M$ divisible by $N$.

# I. Exact equilibrium solution: cluster dispersal

Mechanism of size-incommensurate cluster dispersal:

For $M = 17, N = 8$, extra monomer yields $(1, 0, 0, 0, 0, 0, 0, 2)$.

If $\varepsilon > 0$, many other (seven) 3-cluster states are also sampled:

$(0, 1, 0, 0, 0, 0, 1, 1)$ $(0, 0, 0, 1, 0, 1, 1, 0)$ $(0, 0, 1, 0, 0, 0, 2, 0)$ $(0, 0, 0, 0, 1, 2, 0, 0)$

$(0, 0, 1, 0, 0, 1, 0, 1)$ $(0, 0, 0, 0, 2, 0, 1, 0)$

$(0, 0, 0, 1, 1, 0, 0, 1)$

These $8$ states have similar weights $\Rightarrow$

flat mean cluster size distribution.

# I. Validity of mean-field, mass-action models

| equilibrium cluster numbers ($\varepsilon \ll 1$) | $\frac{M}{N} \to 0$ | $\frac{M}{N}$ finite | $\frac{M}{N} \gg N$ |
|---|---|---|---|
| BD ($N = \infty$) | MFT$^*$ | $\times$ | $\times$ |
| BD (finite $N$) | MFT$^*$ | MFT | MFT$\dagger$ |
| discrete model | exact$^*$ | exact | exact$\dagger$ |

- Results depend on how $N, M \to \infty$ limits are taken.
- Results indicated by $^*$ or $\dagger$ match in the $\varepsilon \to 0^+$ limit.

# I. Table of results

| | **heterogeneous** | **homogeneous** |
|---|---|---|
| **metastable** $\varepsilon = 0$ | exact formula | numerical recursion relation |
| **equilibrium** $0 < \varepsilon << 1$ | numerical recursion relation | explicit asymptotic result |

- extension to nonconstant $p_k$, $q_k$?
- grand canonical calculation for $\varepsilon \not\to 0^+$?

# I. Time to first maximal cluster formation

How long does it take to form first complete cluster?

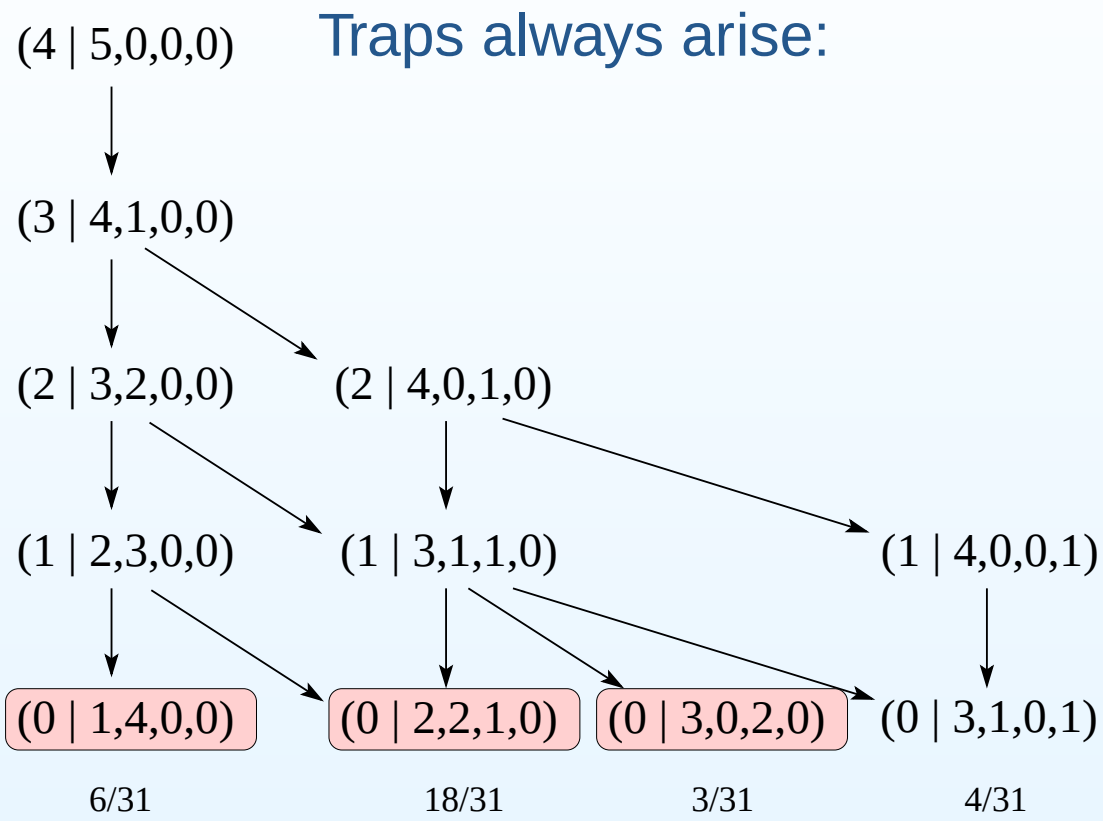
Define $S(n_1, n_2, \ldots, n_{N-1}, t)$ as the probability that starting from the state $\{n_1, n_2, \ldots, n_{N-1}\}$, that $n_N = 0$ at time $t$. $\mathbf{S}$ obeys Backward equation:

$$\dot{\mathbf{S}}(t) = \mathbf{M}\mathbf{S}(t)$$

where $\mathbf{S}(t)$ is the vector of survival probabilities, and $\mathbf{M}$ is a transition matrix.

mean time: $\langle T(\{n_i\})\rangle = \int_0^\infty S(\{n_i\}, t)\mathrm{d}t \equiv \tilde{S}(\{n_i\}, s = 0).$

# I. First passage times: heterogenous nucleation



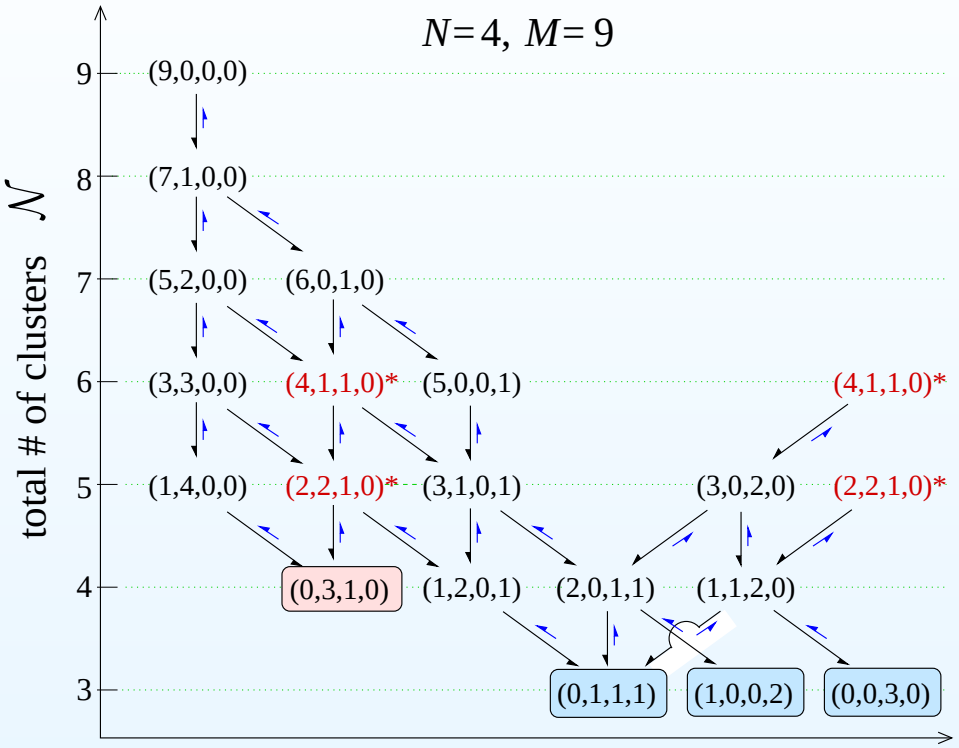

Traps always arise:

For $\varepsilon \equiv q/p \ll 1$, cluster formation may not occur and $\langle T \rangle \sim \frac{1}{\varepsilon}$.

For $\varepsilon \equiv q/p \gg 1$, cluster formation is a rare event: $\langle T \rangle \sim \varepsilon$.

# I. First passage times: homogeneous nucleation

## Traps almost always arise:



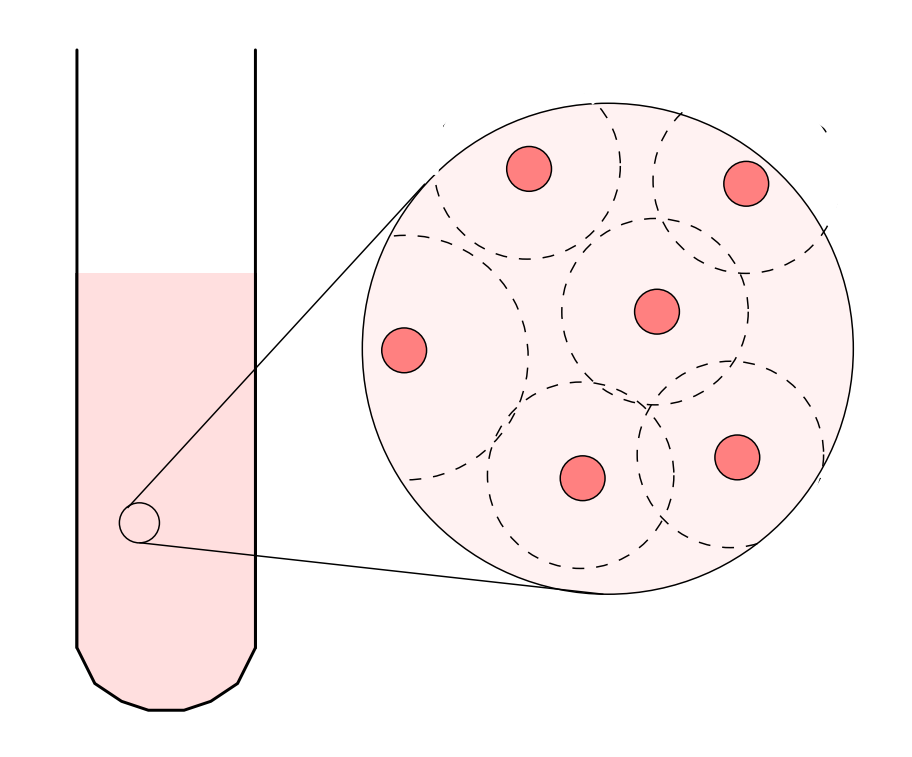

For $\varepsilon \equiv q/p \ll 1$, cluster formation may not occur and $\langle T \rangle \sim \frac{1}{\varepsilon}$.

For $\varepsilon \equiv q/p \gg 1$, cluster formation is a rare event $\langle T \rangle \sim \varepsilon$.

# I. Summary

- Derived fully stochastic, discrete models for heterogeneous and homogeneous nucleation
- Found exact results for heterogeneous nucleation ($\varepsilon = 0$) and homogeneous nucleation ($\varepsilon \to 0^+$).
- Qualitative difference between mass-action and discrete problem arises in homogeneous self-assembly.
- Mean cluster size distribution broadens if size and mass are incommensurate, even if $N, M \to \infty$, as long as $M/N$ finite
- First nucleation times $T$ analyzed. *Mean* times typcially diverge due to “traps”.

# II. Signaling and quorum sensing



- Self-signaling: autoinducer (AI) concentration is highest near self
- What are the constraints on detection of global AI concentration?

# II. Signaling and quorum sensing: baseline conc.

Lowest order homogenization $\Rightarrow$ uniform concentration:

$$\begin{aligned}
\frac{\mathrm{d}\bar{n}_{\mathrm{b}}(t)}{\mathrm{d}t} &= g\left(1-\frac{\bar{n}_{\mathrm{b}}}{K}\right)\bar{n}_{\mathrm{b}}(t) \\
\frac{\mathrm{d}\bar{C}(t)}{\mathrm{d}t} &= -(\mu+k)\bar{C}+4\pi a^2 J_1 \bar{n}_{\mathrm{b}}(t) \\
\frac{\mathrm{d}\bar{C}_{\mathrm{a}}(t)}{\mathrm{d}t} &= -\mu_{\mathrm{a}}\bar{C}_{\mathrm{a}}+k\bar{C},
\end{aligned}$$

Activation $C \to C_{\mathrm{a}}$ can represent general delay.

In the the $k \to \infty$ and $\mu_{\mathrm{a}} \to 0$ limit:

$$\bar{C}_{\mathrm{a}}(t) \approx \frac{4\pi a^2 J_1 K}{g} \ln\left(\frac{n_0}{K}(e^{gt}-1)+1\right),$$

# II. Signaling and quorum sensing: local concentration

$$D\nabla^2\delta C(\mathbf{r}) = (\mu + k)\delta C$$

$$D_{\rm a}\nabla^2\delta C_{\rm a}(\mathbf{r}) = \mu_{\rm a}\delta C_{\rm a}(\mathbf{r}) - k\delta C(\mathbf{r}),$$

$$\delta C_{\rm a}(r = a) = \frac{k}{\sqrt{(\mu_{\rm a} + k)D_{\rm a}}}\left(\frac{J_1 a^2}{D + a\sqrt{(\mu + k)D}}\right)$$

Define: $F(t) \equiv \dfrac{\bar{C}_{\rm a}(t)}{\bar{C}_{\rm a}(t) + \delta C_{\rm a}(r = a)}$

Global detection efficient only when $F(t) \approx 1$

# II. Signaling and quorum sensing: $F$ ratio

Using simplest forms for $\bar{C}_{\rm a}(t)$ and $\delta C_{\rm a}(r=a)$ and assume $D \approx D_{\rm a}$:

$$F(t) \approx \frac{4\pi a D K \ln(\frac{n_0}{K}(e^{gt}-1)+1)}{g + 4\pi a D K \ln(\frac{n_0}{K}(e^{gt}-1)+1)} \xrightarrow[gt\ll 1]{} \frac{4\pi a D n_0 t}{1+4\pi a D n_0 t},$$

For $gt \gg 1$,

$$F(gt \gg 1) \approx \begin{cases} \dfrac{4\pi a D n_0 e^{gt}}{g + 4\pi a D n_0 e^{gt}} & \dfrac{n_0 e^{gt}}{K} \ll 1 \\ \dfrac{4\pi a D K t}{1+4\pi a D K t} & \dfrac{n_0 e^{gt}}{K} \gg 1 \end{cases}$$

# II. Signaling and quorum sensing

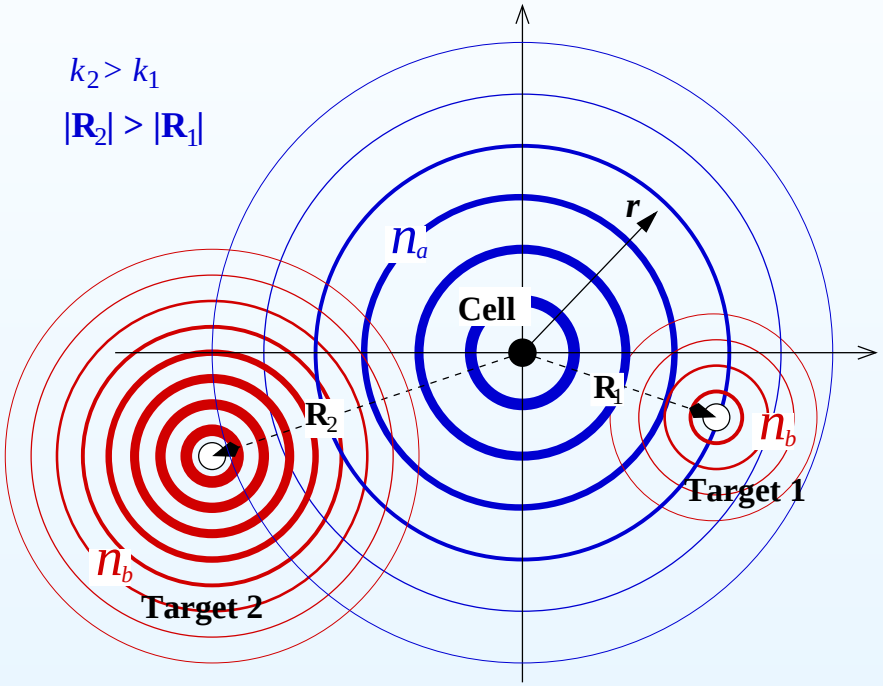
- $F(t)$ depends on parameters/conditions ($t, n_0, \mu, k, D, g, K$).
- True quorum sensing only if threshold concentration $C_{\rm a}^* \gg \delta C_{\rm a}$.

**Diffusion sensing hypothesis** (R. Redfield, 2002, Trends Microbiol.):

Bacterial release cheap AI. If it is also sensed, then AI is not degraded/transported away, and cell can release more expensive enzymes (e.g. proteases). This process would rely on total $C_{\rm a}$ sensed regardless of source. $F(t)$.

# II. Toy model: Diffusion sensing and chemolocation

S. Nowak, B. Chakrabarti, T. Chou, A. Gopinathan, *Physical Biol.*, 2010.




– "Protease" ($n_{\mathrm{a}}(\mathbf{r}, t)$) released by motile cell simulates targets

– Targets release chemoattractant ($n_{\mathrm{b}}(\mathbf{r}, t)$).

– To which target does the cell migrate?

# II. Diffusion sensing and chemolocation

simple moving source model:

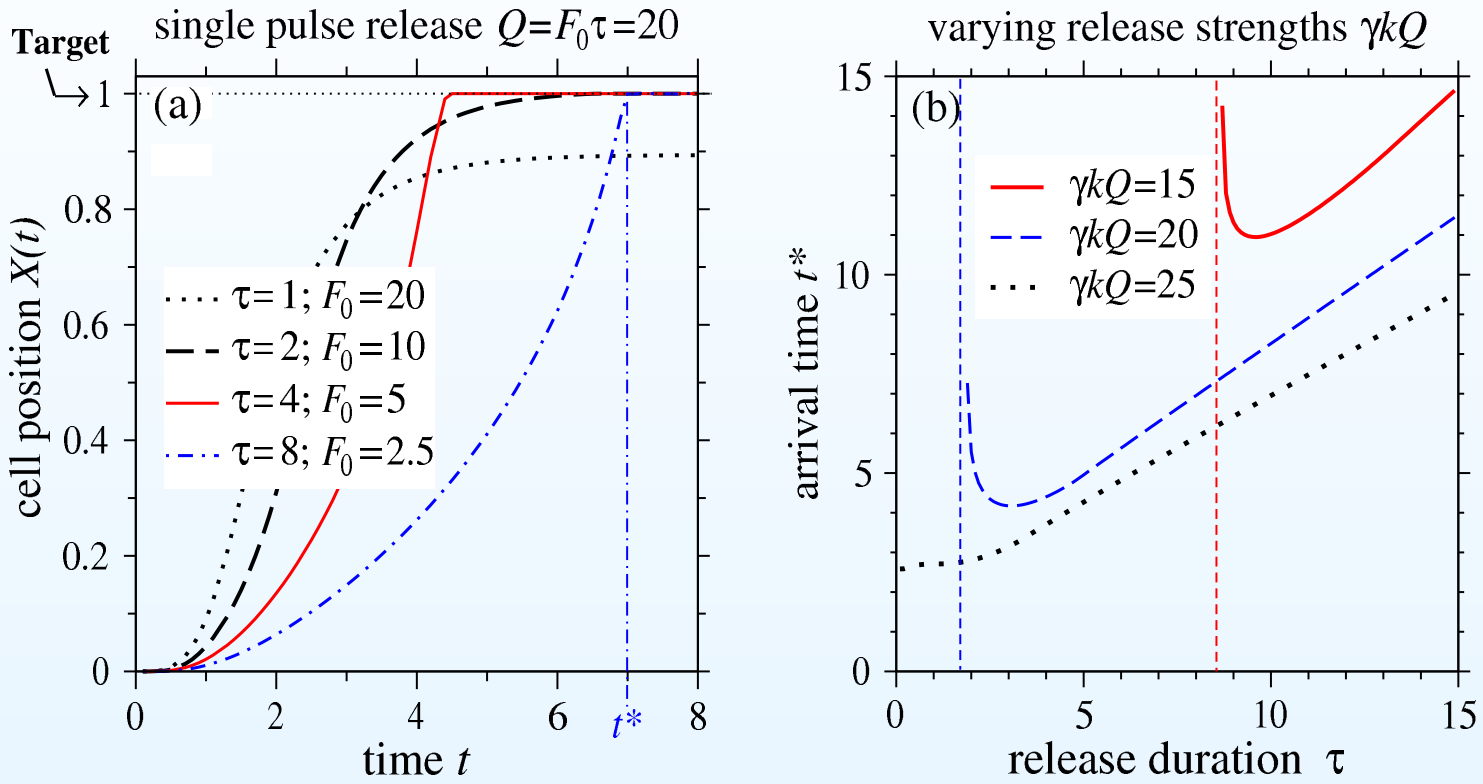
$$\dot{n}_a(\mathbf{r},t) = D_a\nabla^2 n_a - \mu_a n_a + F(t)\delta(\mathbf{r}-\mathbf{R}(t)),$$

$$\dot{n}_b(\mathbf{r},t) = D_b\nabla^2 n_b - \mu_b n_b + \sum_j \delta(\mathbf{r}-\mathbf{R}_j)K_j\left[n_a(\mathbf{r},t-\tau),t\right],$$

$$\dot{\mathbf{R}}(t) = \nabla U\left[n_b(\mathbf{r},t)\right]_{\mathbf{r}=\mathbf{R}(t)},$$

Integral solutions/approximations possible for linear response $U[n_{\rm b}]$

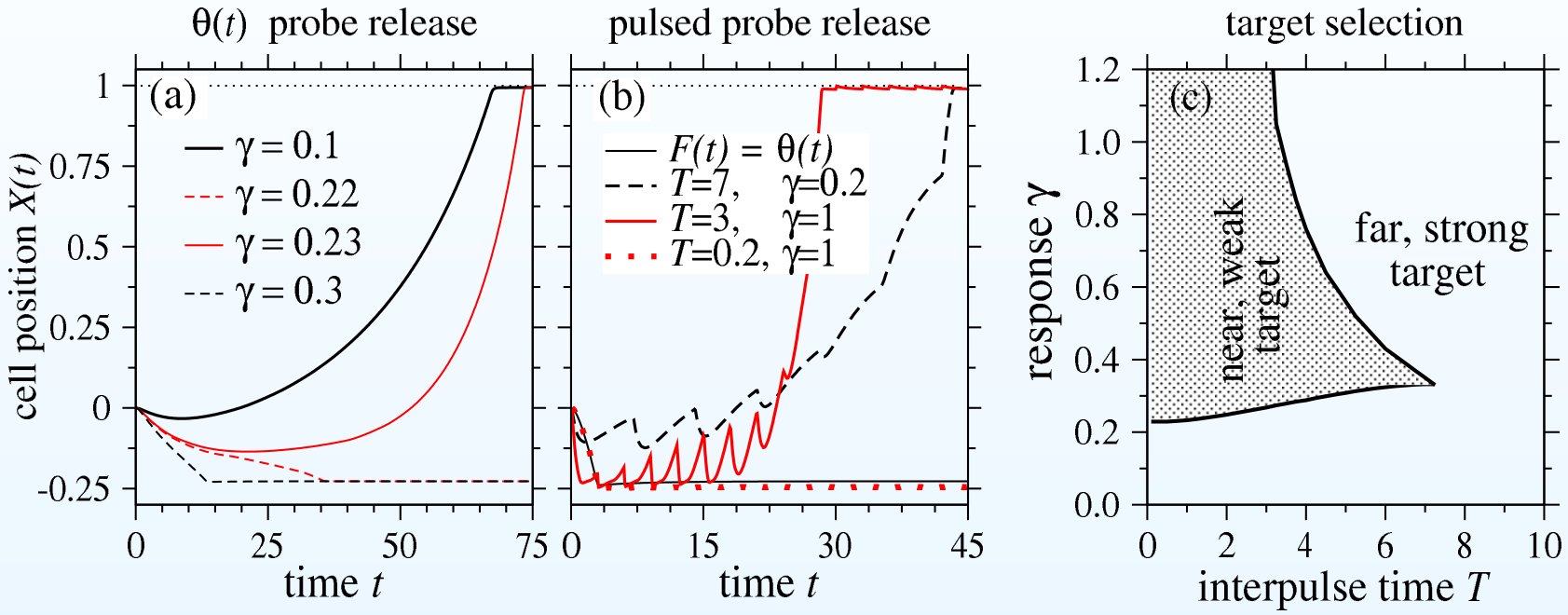
# II. Diffusion sensing and chemolocation




$$D = \gamma = k = \mu_A = \mu_B = 1$$

Motion during activator release $\Rightarrow$ optimal pulse duration

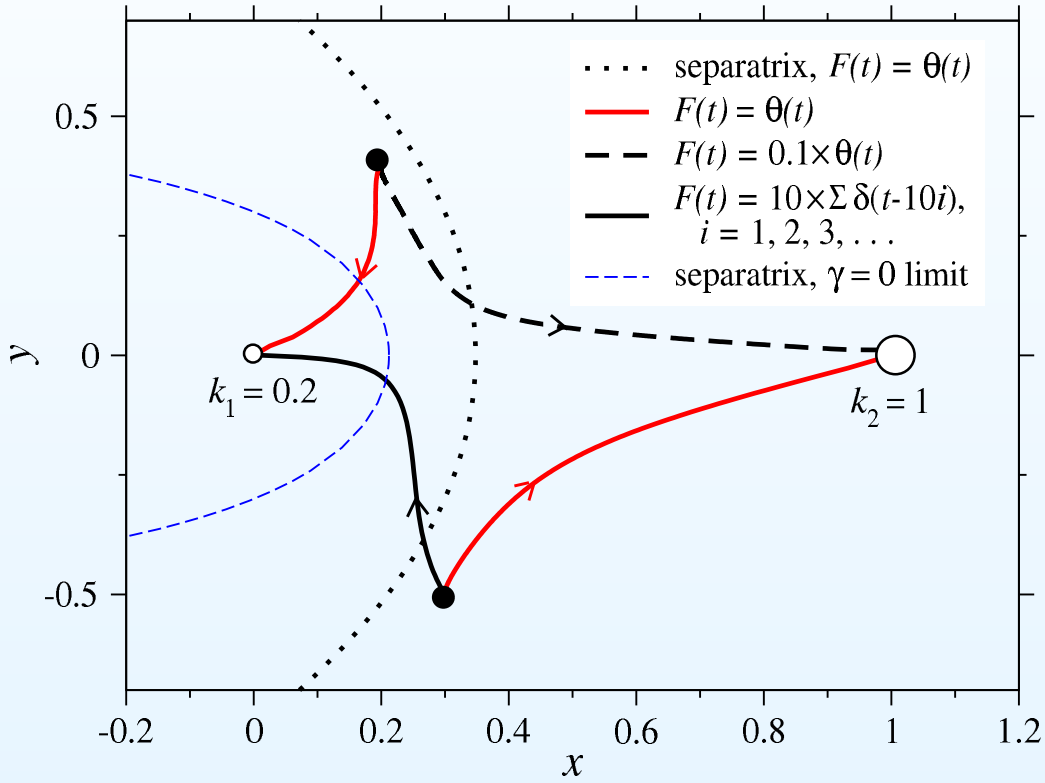
# II. Diffusion sensing and chemolocation




Different responses $\gamma$ and different interpulse periods allow taxis toward different targets

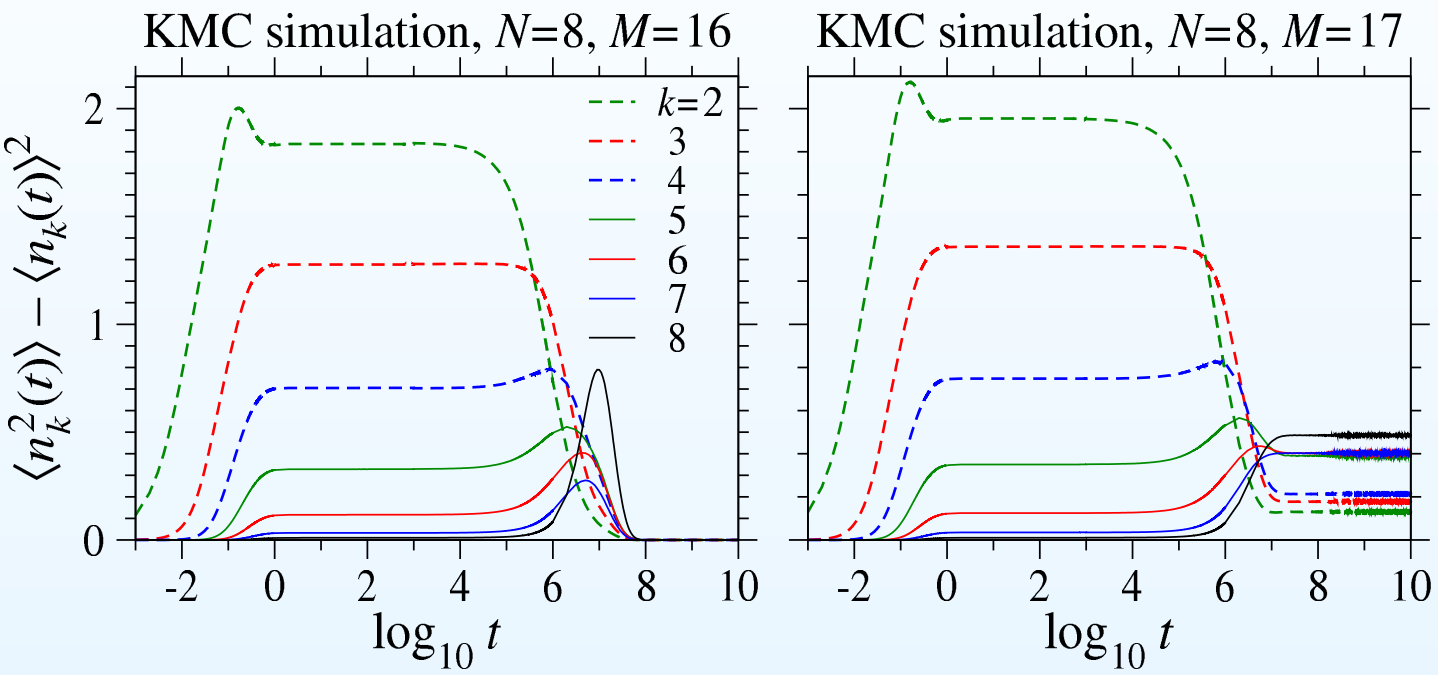
# II. Diffusion sensing and chemolocation

## Target selection separatrices:



# KMC simulations: variances

$N = 8, \varepsilon = 10^{-6}$:




- Variance in metastable regime is large, although mean is accurate
- Variance at equilibrium is low when commensurate, but appreciable when incommensurate