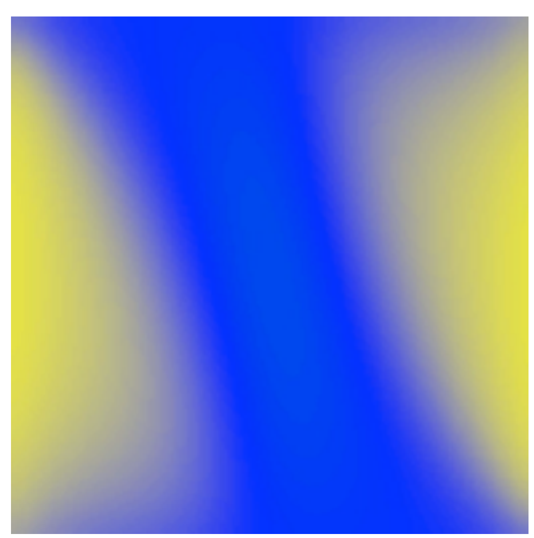
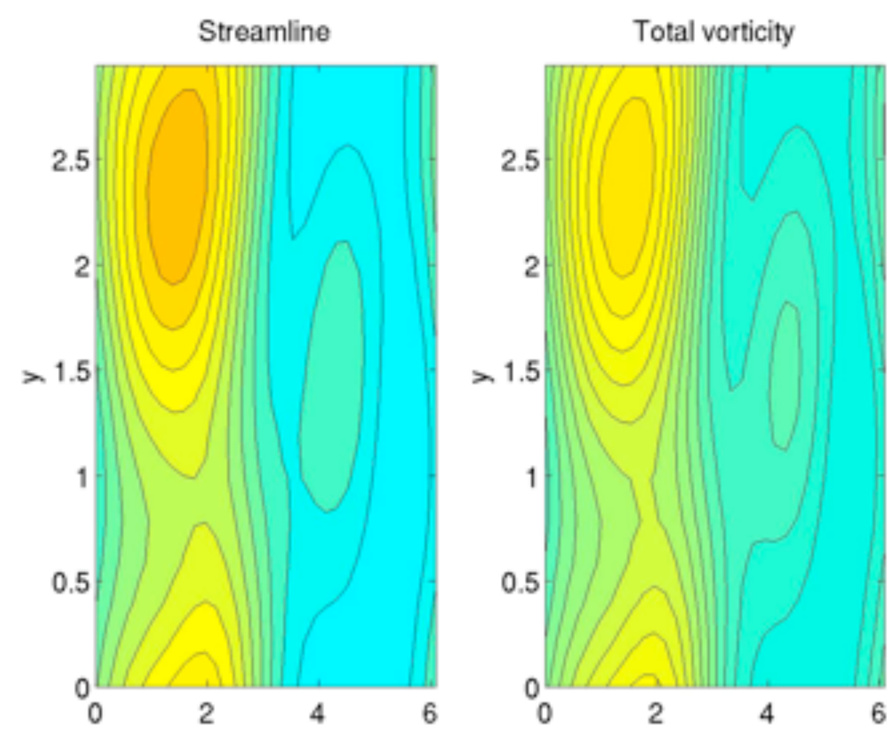
# A tour on the computational side of Large Deviation Theory

Eric Vanden-Eijnden
Courant Institute

- Rare events matter  - challenge for modeling and computations.
- Rare events pathways are often predictable - large deviation theory (LDT).
  Numerical aspects: minimum action method & string method.
- LDT-based importance sampling strategies - UQ and reliability Q, filtering, etc.
- Beyond LDT - when entropy matters.

*It is easier to go down a hill than up but the view is much better at the top - H. W. Beecher*





# Unlikely (or infrequent) events matter

- Some events are rare (i.e. infrequent) but have dramatic consequences - massive earthquakes, giant hurricanes, pandemics, etc.

- Rare events can be less dramatic but important nonetheless - for example, electronic components or engineering devices used in automobile, aerospace and medical are required to be extremely reliable, with very small probability of failure.

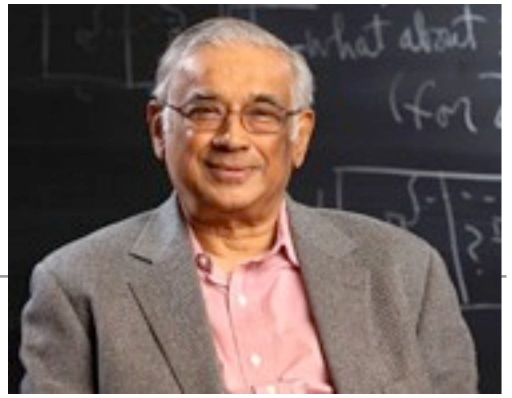
- Rare events always happen eventually given enough time and may be the most interesting/important aspect of the system's dynamics - e.g. conformational changes of macromolecules, kinetic phase transitions, thermally induced magnetization reversal in micromagnets, regime changes in climate, etc.

  - *Metastability.*

# Unlikely (or infrequent) events matter


- Realistic models (SODEs, SPDEs, Markov jump processes, etc.) are often too complex to be amenable to analytical solution.
- Direct numerical simulations of these models very challenging or even impossible.
- Requires new computational approaches based e.g. on large deviation theory (LDT).
- Requires to go beyond LDT in certain situations - entropic effects, ruggedness, etc.

# Large Deviation Theory

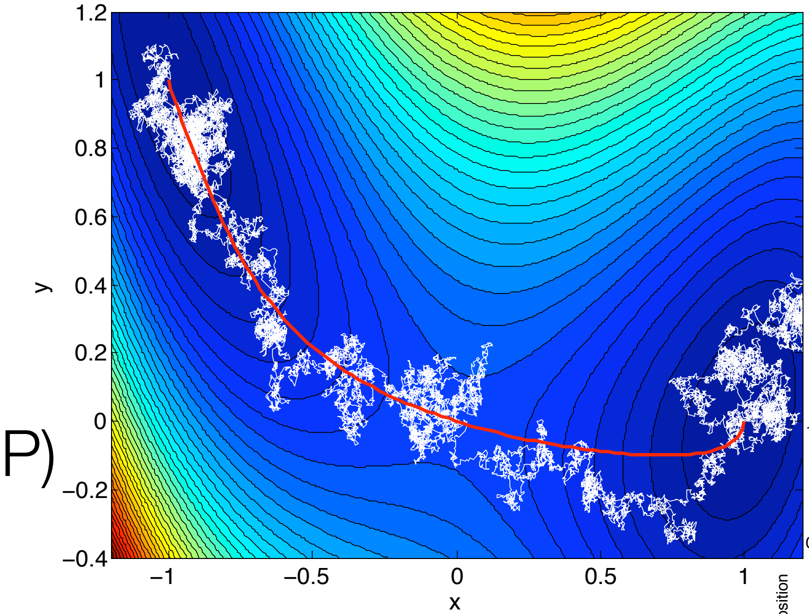


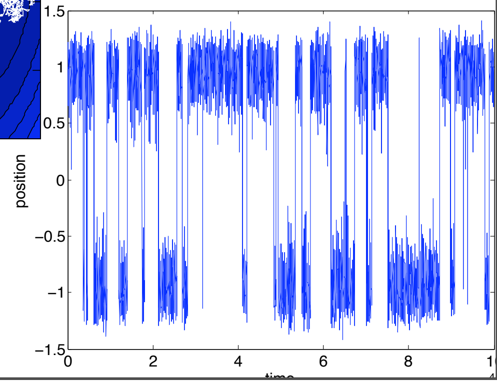
- The way rare events occur is often predictable: the probability of the rare event is dominated by the most likely (i.e. least unlikely) scenario for it to happen - essence of *large deviation theory.*

*Happy families are all alike; every unhappy family is unhappy in its own way -Tolstoi.*

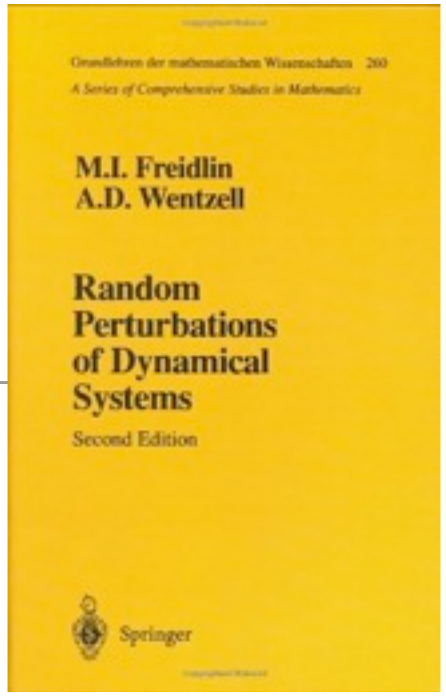
- Calculation of the path of maximum likelihood (PML) reduces to a deterministic optimization problem.

- For example, in gradient systems (i.e. systems navigating over an energy landscape), rare events are associated with barrier crossing events and follow the minimum energy path (MEP) connecting two minima of the energy potential.



# Freidlin-Wentzell Approach to LDT




Key object: *action functional*

$$S_T(\phi) = \frac{1}{2}\int_0^T \left|\sigma^{-1}(\phi(t))(\dot{\phi}(t) - b(\phi(t)))\right|^2 dt$$

associated with S(P)DE

$$dX^\varepsilon(t) = b(X^\varepsilon(t))dt + \sqrt{\varepsilon}\,\sigma(X^\varepsilon(t))dW(t)$$

*Then:* Probability that $\{X^\varepsilon(t)\}_{t\in[0,T]}$ be close to a given path $\{\phi(t)\}_{t\in[0,T]}$ is roughly

$$\mathbb{P}\left\{\sup_{0\le t\le T}|X^\varepsilon(t) - \phi(t)| < \delta\right\} \asymp \exp(-\varepsilon^{-1}S_T(\phi))$$

.

Reduce estimation of probability to a *minimization problem*, e.g.

$$\mathbb{P}\{X^\varepsilon(T) \in A | X^\varepsilon(0) = x\} \asymp \exp(-\varepsilon^{-1}\inf S_T(\phi)),$$

where the infinum is taken over all $\phi(\cdot)$ such that $\phi(0) = x$ and $\phi(T) \in A$.

# The role of the Quasi-Potential (QP)

Looking at long time-intervals, $T \asymp \exp(\varepsilon^{-1}C)$ for some $C > 0$, the *quasi-potential* $V : H^2 \mapsto \mathbb{R}$ is the key quantity:

$$V(x,y) = \inf_T \inf_\phi \{S_T(\phi) : \phi(0) = x, \phi(T) = y\}$$

Interpretation:

$$V(x,y) = -\lim_{T\to\infty} \lim_{\delta\to 0} \lim_{\varepsilon\to 0} \varepsilon^{-1} \log \mathbb{P}\{\tau_y^\varepsilon(x) \leq T\}$$

where $\tau_y^\varepsilon(x)$ is the first is the *first entrance time* in a $\delta$-neighborhood of $y$:

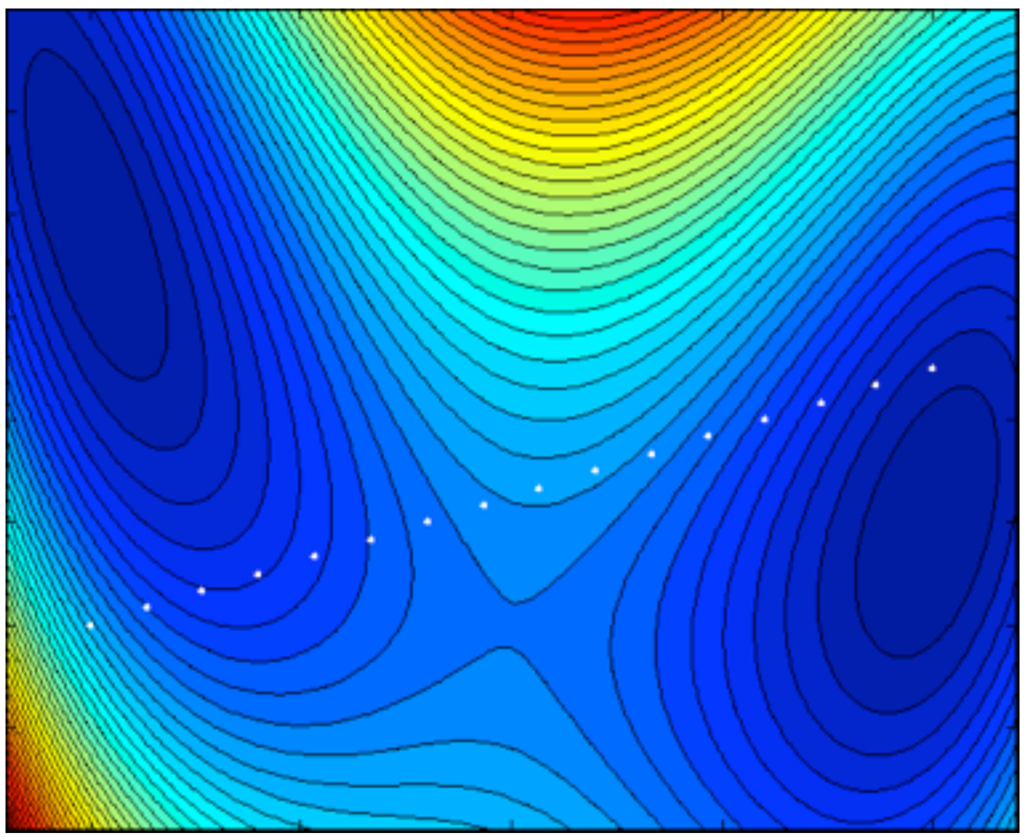
$$\tau_y^\varepsilon(x) = \inf\{t : X^\varepsilon(t) \in B_\delta(y), X^\varepsilon(0) = x\}$$

*Long time intervals most relevant since the effects of the noise is unavoidable on these timescales.*

# Computational side of LDT

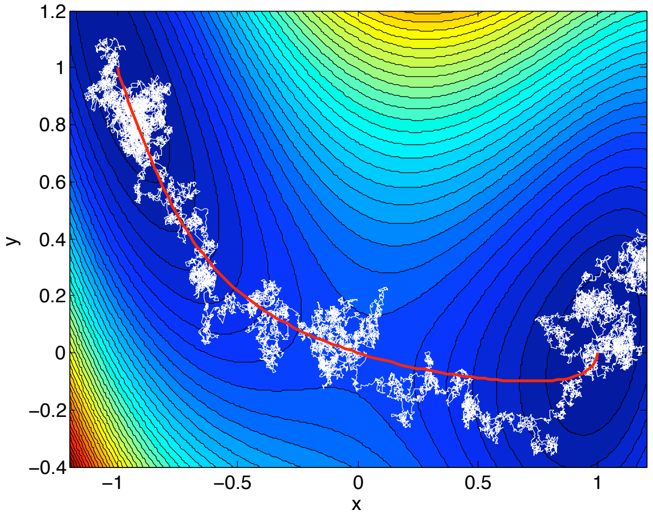
*Over every mountain there is a path, although it may not be seen from the valley -T. Roethke*

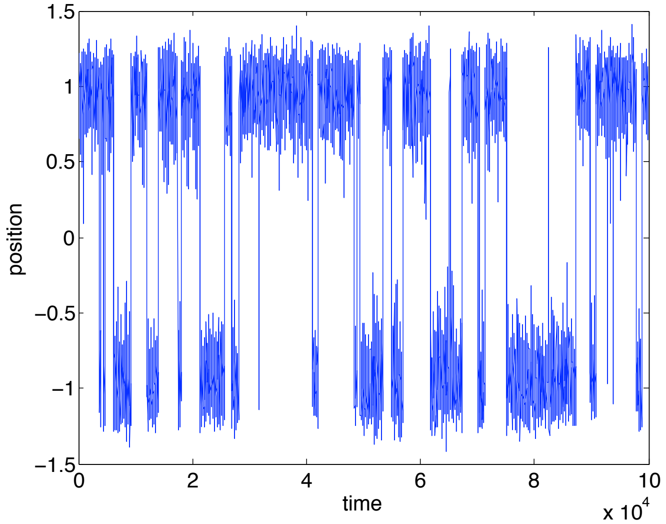
- Numerical tools (string method, minimum action method - MAM, etc.) have been developed to identify the paths by which rare events are most likely to occur.



String method and MAM evolve curves while controlling their parametrization until they converge to the MLP.

Applicable to both gradient systems (where MLPs are MEPs) and non-gradient systems.




- These tools have been applied to problems from biochemistry, material sciences, atmosphere-ocean sciences, etc.

# Geometric rephrasing of QP definition

(M. Heymann & E. V.-E., 2008)

$$V(x,y) \asymp \inf_T \inf\{S_T(\varphi) : \varphi(0) = x, \varphi(T) = y\}$$
$$= \tfrac{1}{2} \inf \left\{ \int_\Gamma \|b\|_a \sin^2(2\eta) ds : \Gamma \text{ joins } x \text{ to } y \right\}$$

where $\|u\|_a^2 = \langle u, u\rangle_a$ with $\langle u, v\rangle_a = \langle a^{-1}(x)u, v\rangle$ if $\langle \cdot, \cdot\rangle$ was the original inner-product, and $\eta$ is the local value of the angle between $\Gamma$ and $b$.

Parametrize $\Gamma = \{\varphi(s) : s \in [0,1]\}$ so that:

$$V(x,y) = \inf \left\{ \int_0^1 \left( \|\varphi'\|_a \|b(\varphi)\|_a - \langle \varphi', b(\varphi)\rangle_a \right) ds : \varphi(0) = x, \varphi(1) = y \right\}$$
$$= \inf_{\substack{\varphi(\cdot) \\ \varphi(0)=x \\ \varphi(1)=y}} \sup_{\substack{\theta(\cdot) \\ H(\varphi,\theta)=0}} \int_0^1 \langle \varphi', \theta\rangle ds$$

where $H(\varphi, \theta)$ is the Hamiltonian associated with the Lagrangian $L(\varphi, \varphi') = \|\varphi' - b(\varphi)\|_a^2$:

$$H(\varphi, \theta) = \langle b(\varphi), \theta\rangle + \tfrac{1}{2}\langle \theta, a(\varphi)\theta\rangle$$

# Geometric rephrasing of QP definition

(M. Heymann & E. V.-E., 2008)

The associated Euler-Lagrange equations can be written as

$$\begin{cases} \lambda\varphi' = H_\theta(\varphi,\theta) = b(\varphi) + \theta \\ \lambda\theta' = -H_\varphi(\varphi,\theta) = -\nabla b^T(\varphi)\theta \end{cases}$$

where $\lambda$ is determined such that $0 = H(\varphi,\theta)$ and we impose some (arbitray) constraint on the paramterization of $\Gamma$, e.g. $|\varphi'| = cst$.

Differentiating the equation for $\varphi$ and using the one for $\theta$, this system of equations can be written as

$$0 = \lambda^2\varphi'' + \lambda\lambda'\varphi' - \lambda H_{\theta\varphi}\varphi' + H_{\theta\theta}H_\varphi$$

which can be reduced to

$$\boxed{0 = \lambda^2\varphi'' - P\left(\lambda H_{\theta\varphi}\varphi' - H_{\theta\theta}H_\varphi\right), \qquad P = \mathrm{Id} - \frac{\varphi'\otimes\varphi'}{|\varphi'|^2} \qquad (\star)}$$

where we used $P\varphi' = 0$, $P\varphi'' = \varphi''$ (since $|\varphi'| = cst$) and

$$\theta = a^{-1}\left(\lambda\varphi' - b\right), \qquad \lambda = \frac{\|b\|_a}{\|\varphi'\|_a}$$

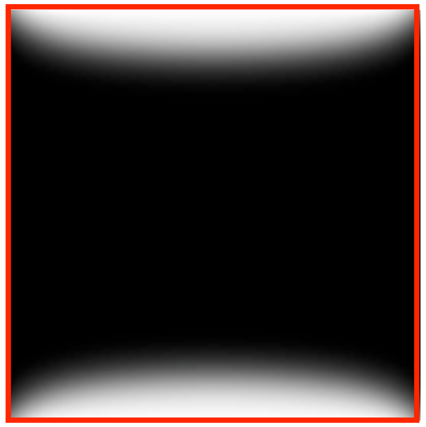
# Allen-Cahn equation in 2D - MEPs

Allen-Cahn energy for $u : [0,1]^2 \mapsto \mathbb{R}$:

$$E(u) = \int_{[0,1]^2} \left(\tfrac{1}{2}\delta|\nabla u|^2 + \tfrac{1}{4}\delta^{-1}(1-u^2)^2\right) dx$$

with Dirichlet boundary condition: $u = +1$ on the right and left edges of $[0,1]^2$, $u = -1$ on top and bottom ones.

*Simulations by W. Ren*

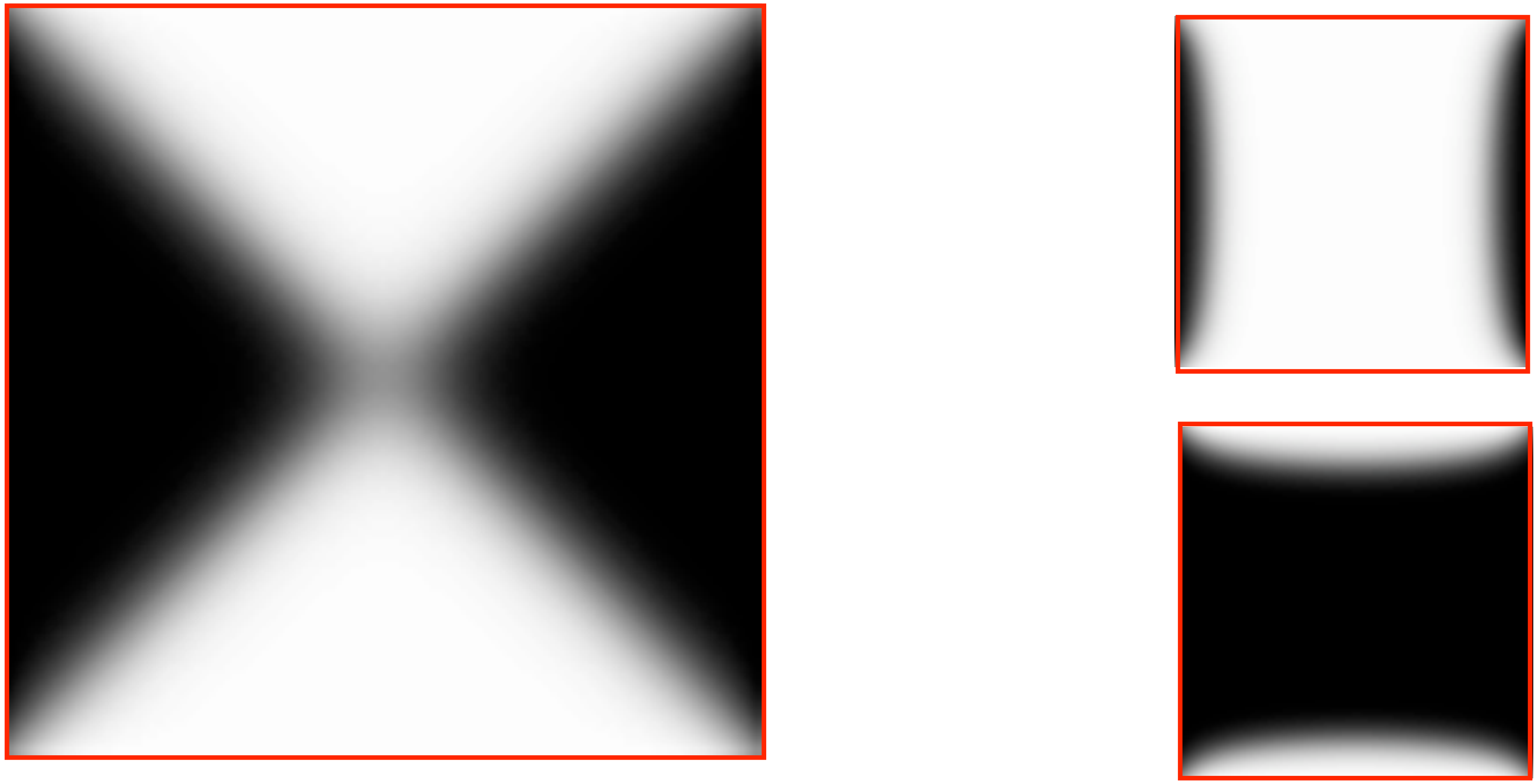


# Allen-Cahn equation in 2D - MEPs

Allen-Cahn energy for $u : [0,1]^2 \mapsto \mathbb{R}$:

$$E(u) = \int_{[0,1]^2} \left(\tfrac{1}{2}\delta|\nabla u|^2 + \tfrac{1}{4}\delta^{-1}(1-u^2)^2\right) dx$$

with Dirichlet boundary condition: $u = +1$ on the right and left edges of $[0,1]^2$, $u = -1$ on top and bottom ones.



*Simulations by W. Ren*

# Allen-Cahn equation in 2D - MLPs at finite T

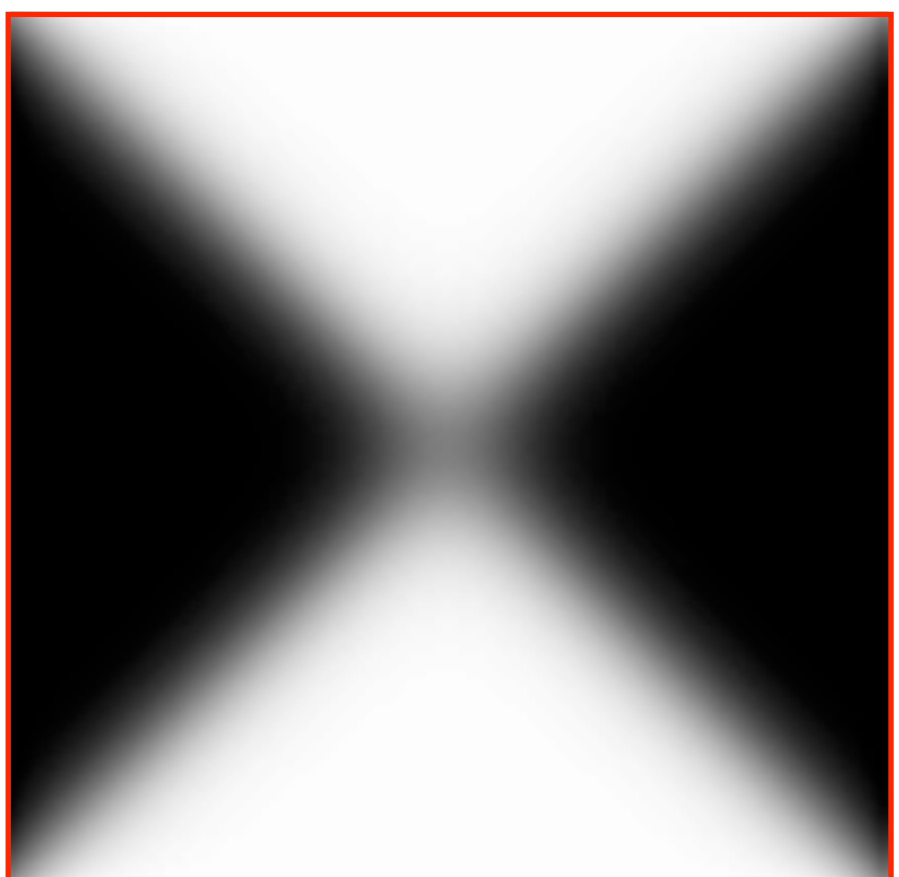
Dynamics:

$$u_t = \delta^2 \Delta u + u - u^3 + \sqrt{2\varepsilon}\,\eta(t,x)$$

Action:

$$S_T(\varphi) = \frac{1}{2}\int_0^T \int_\Omega |\varphi_t - \delta^2 \Delta\varphi - \varphi + \varphi^3|^2 dx dt$$

*Unconstrained reversal*



*Time-constrained reversal*

# Allen-Cahn equation in 2D - MLPs at finite T

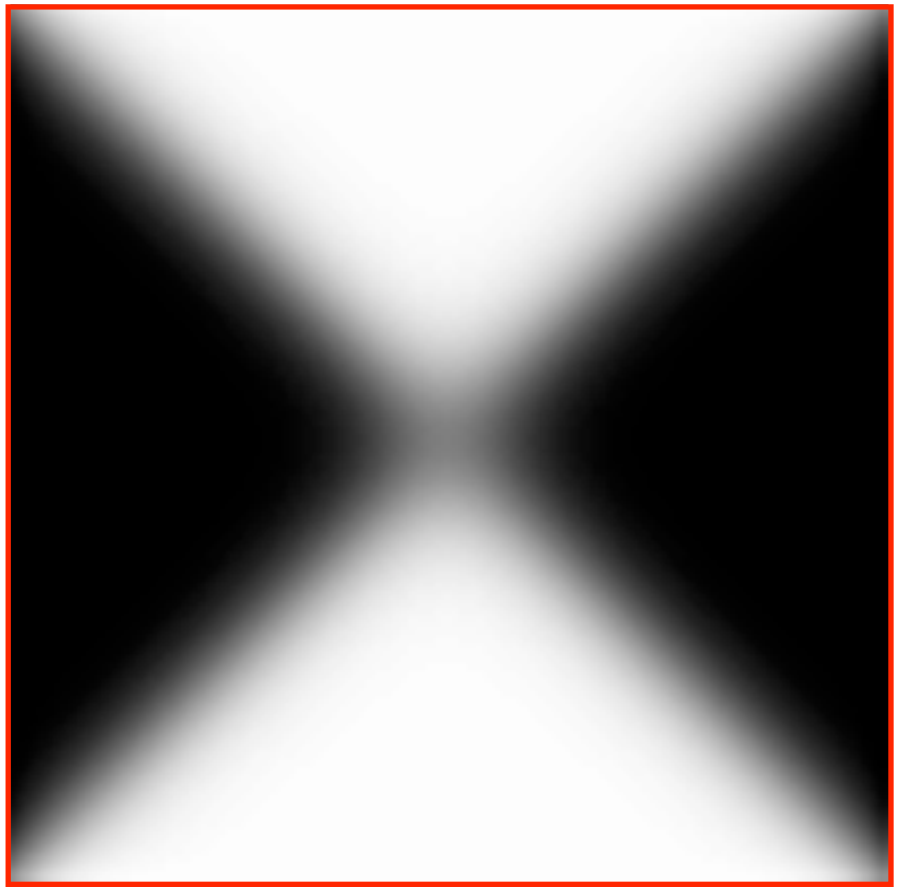
Dynamics:

$$u_t = \delta^2 \Delta u + u - u^3 + \sqrt{2\varepsilon}\,\eta(t,x)$$

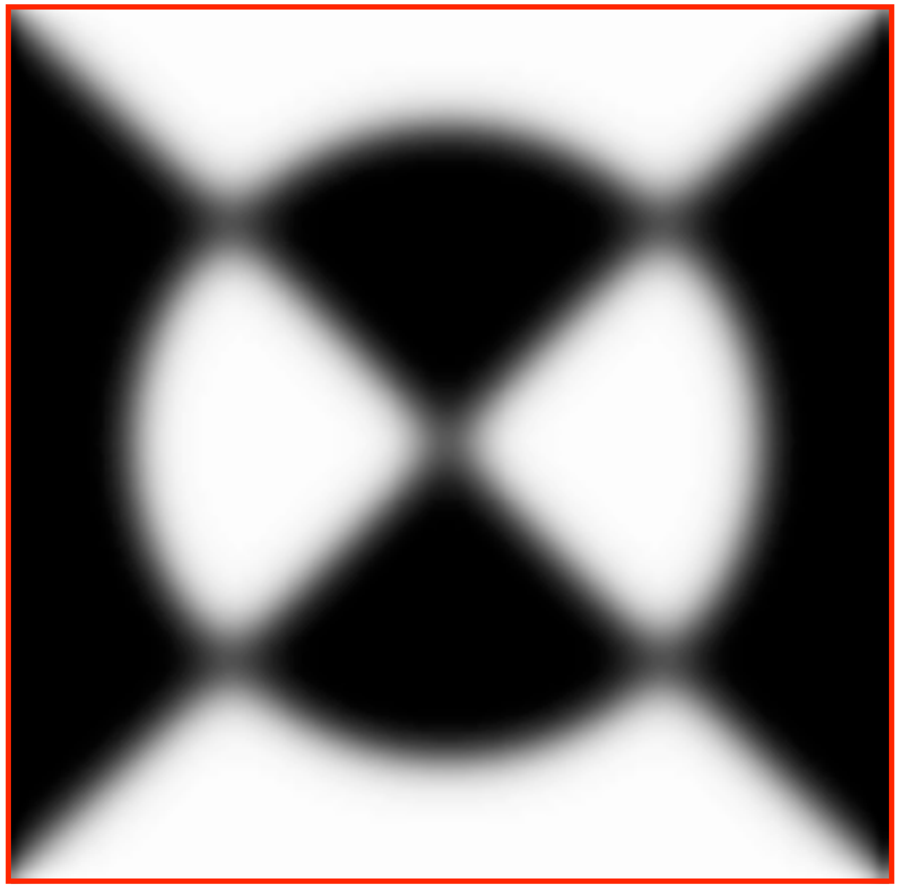
Action:

$$S_T(\varphi) = \frac{1}{2}\int_0^T \int_\Omega |\varphi_t - \delta^2 \Delta\varphi - \varphi + \varphi^3|^2 dx dt$$

*Unconstrained reversal*



*Time-constrained reversal*



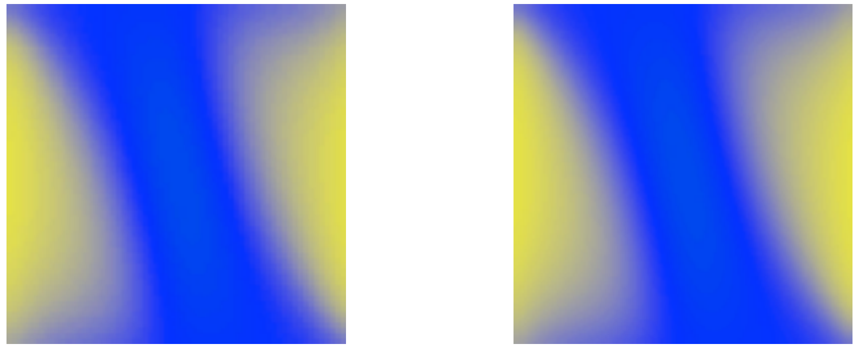
# Thermally induced magnetization reversal

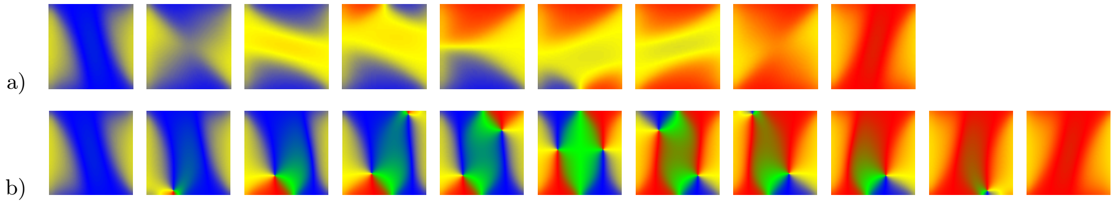
(W. E, W. Ren & E. V.-E., 2003: A. Kent, D. Stein & E. V.-E., 2009 ...)

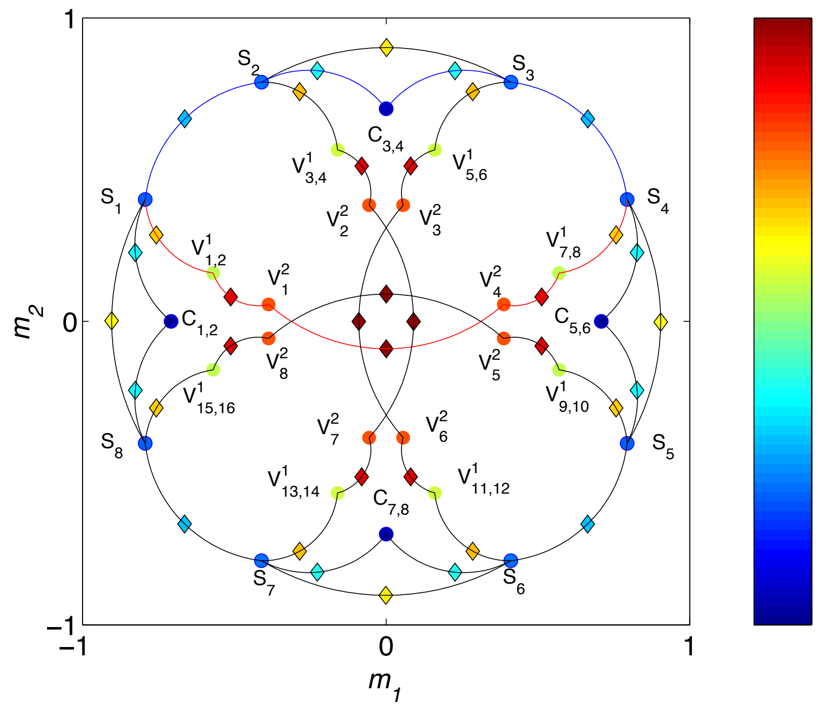
Main building blocks in Magnetoelectronics (used e.g. as storage devices, etc.)

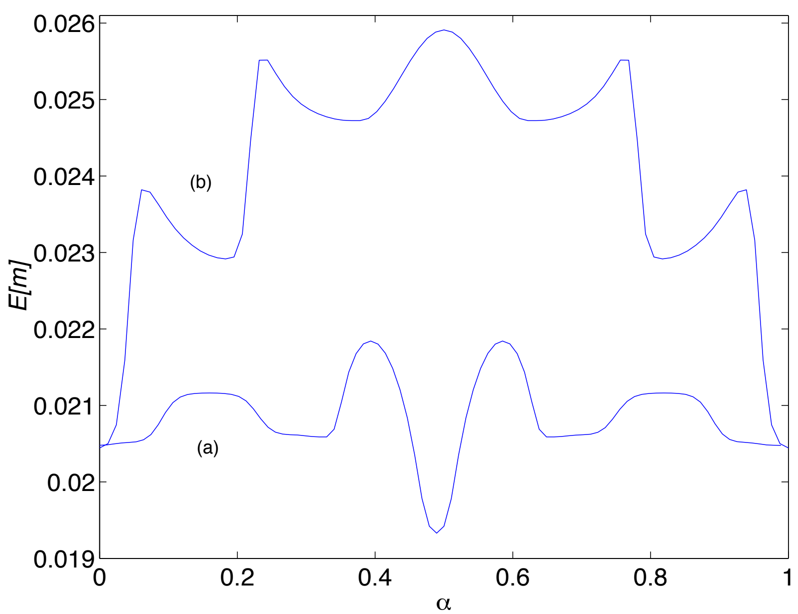
Small elements at superparamagnetic limit, where thermal effects become important and limit data retention time by magnetization reversal.

Reversal complex due to non-uniformity in space.





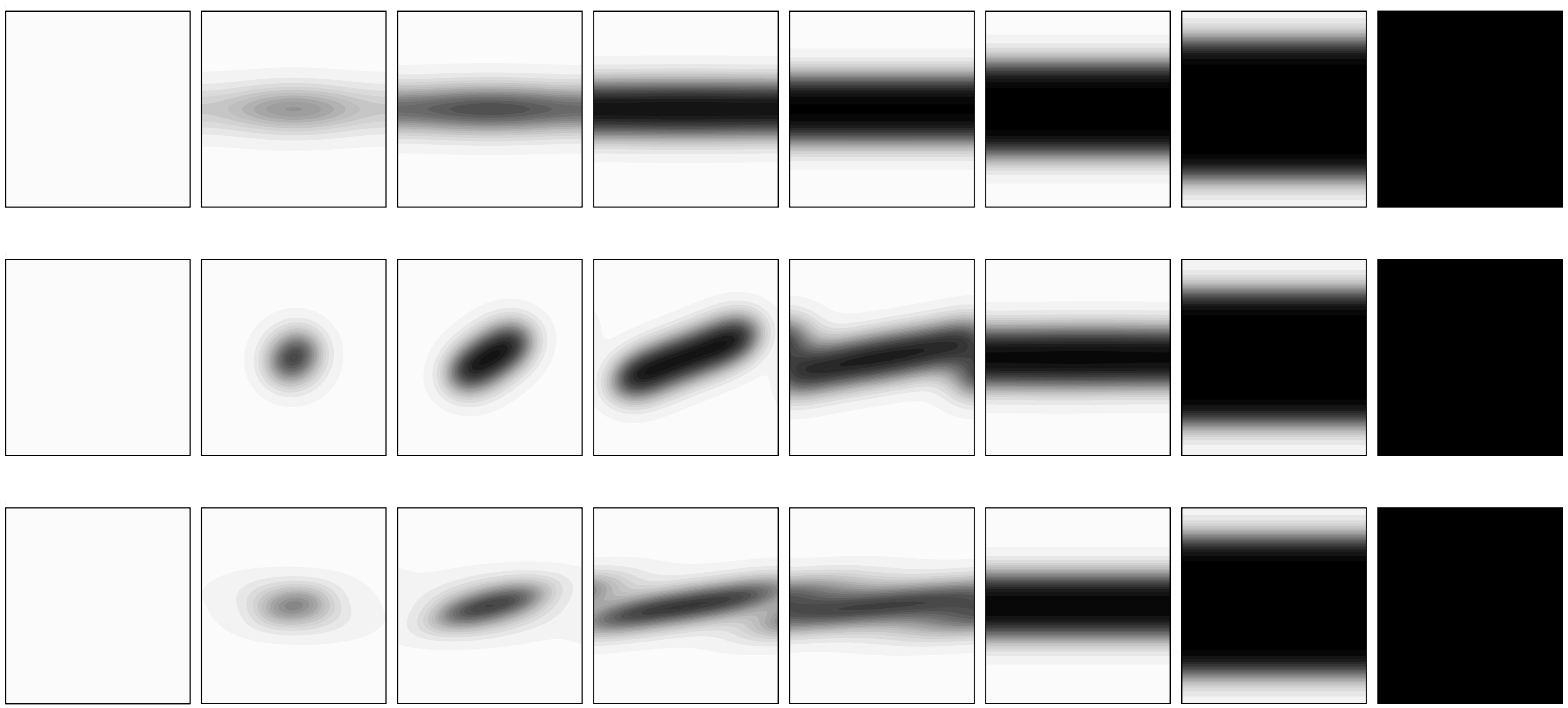




Dynamics can be reduced to a Markov jump process on energy map, whose nodes are the energy minima and whose edges are the minimum energy paths

# Non-gradient effects - example: AC with advection

$$u_t = \delta^2 \Delta u + u - u^3 + c \sin(y) u_x + \sqrt{2\varepsilon}\, \eta(t, x)$$

in 2D with periodic BC



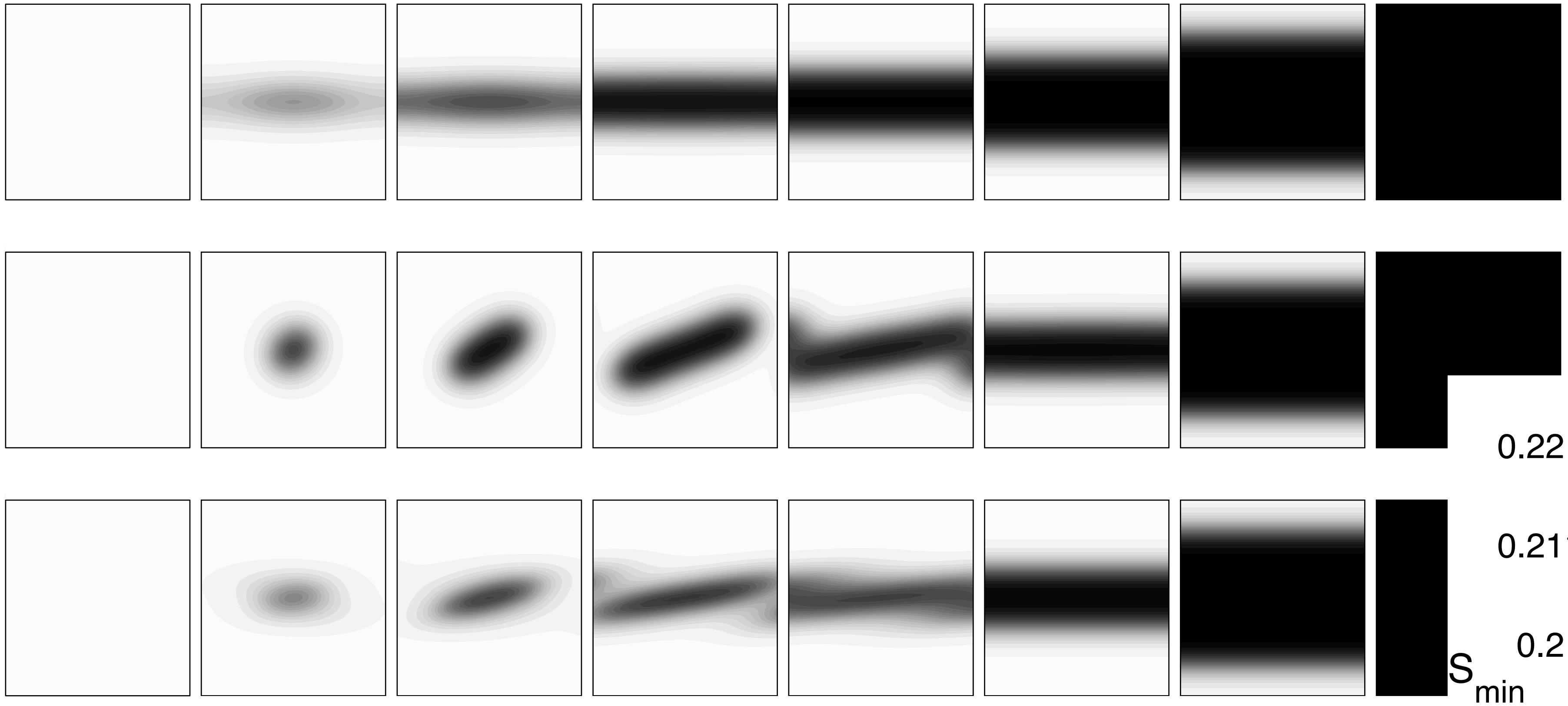
MLPs for different shear strengths

*Simulations by M. Heymann and M. Tao*

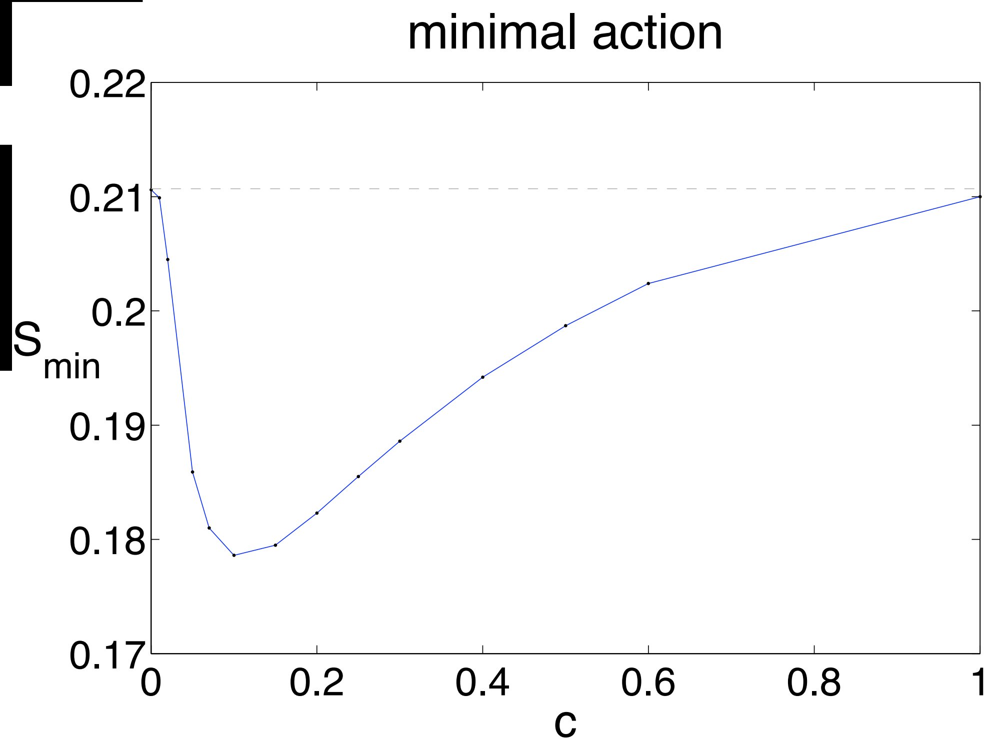
# Non-gradient effects - example: AC with advection

$$u_t = \delta^2 \Delta u + u - u^3 + c\sin(y)\, u_x + \sqrt{2\varepsilon}\,\eta(t,x)$$

in 2D with periodic BC



MLPs for different shear strengths

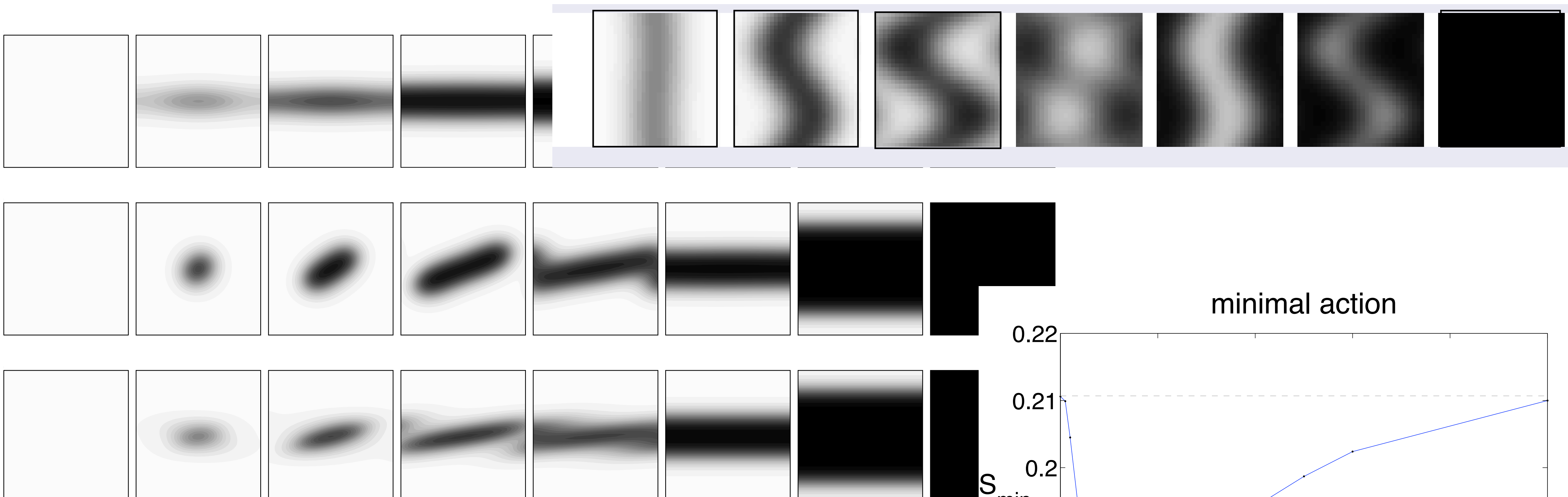



*Simulations by M. Heymann and M. Tao*

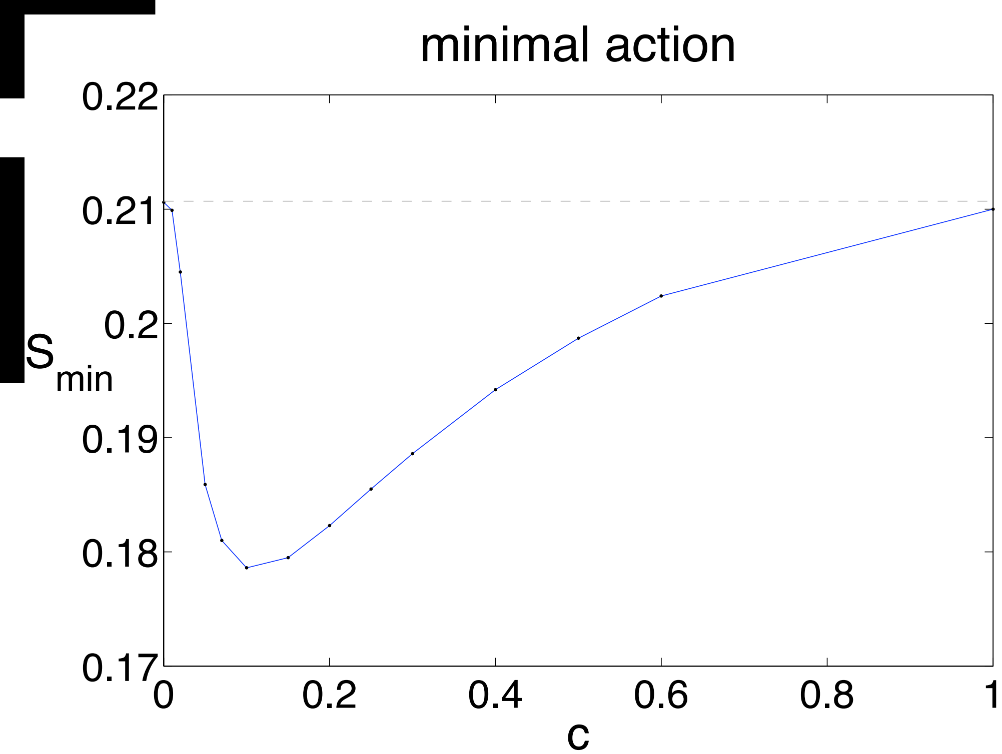
# Non-gradient effects - example: AC with advection

$$u_t = \delta^2 \Delta\, u + \; u - u^3 + \; c\sin(y)\, u_x + \; \sqrt{2\varepsilon}\,\eta(t,x)$$

in 2D with periodic BC

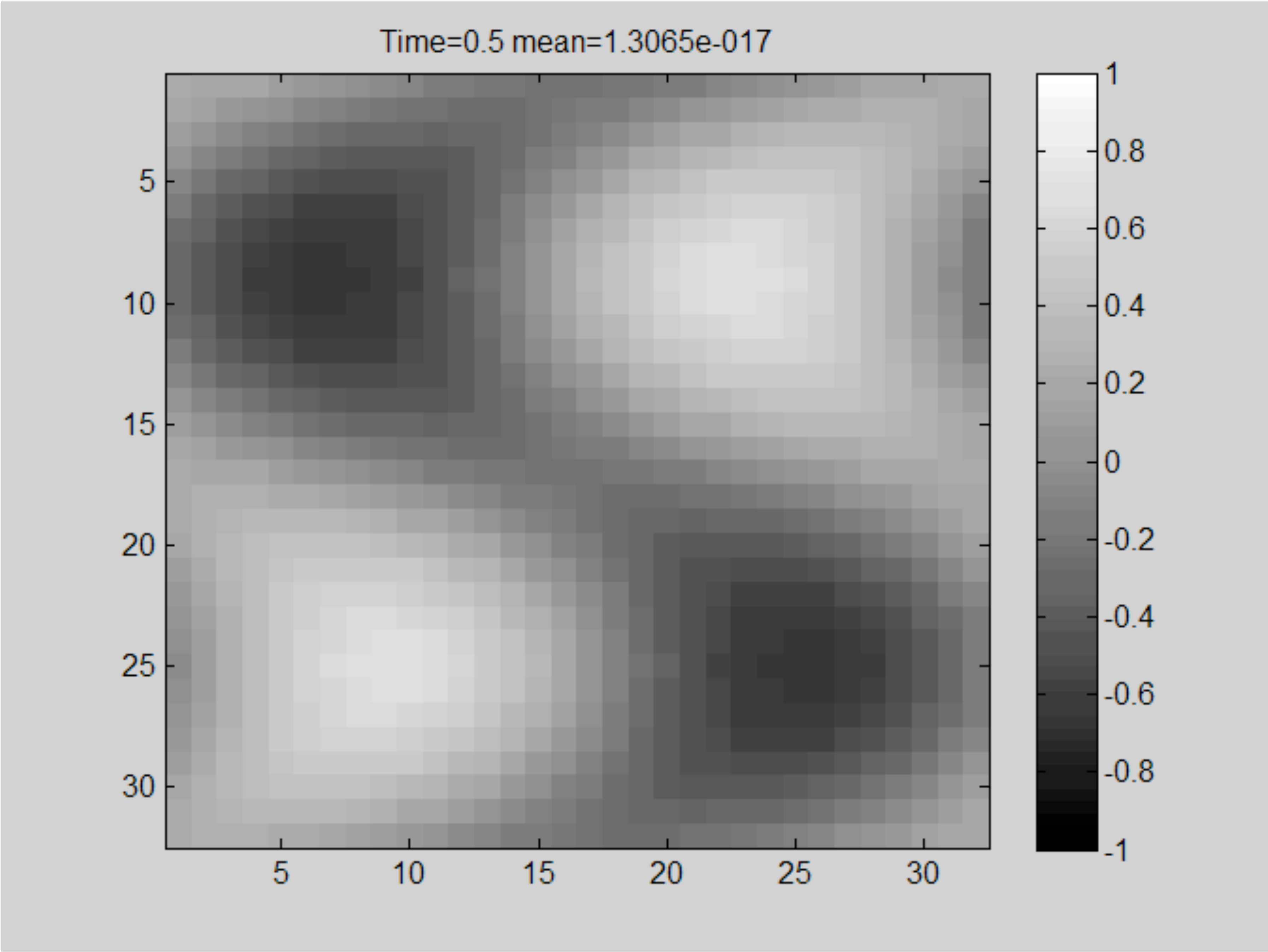


MLPs for different shear strengths




*Simulations by M. Heymann and M. Tao*

# Unstable limit cycle along PML



# Application to fluid dynamics problems

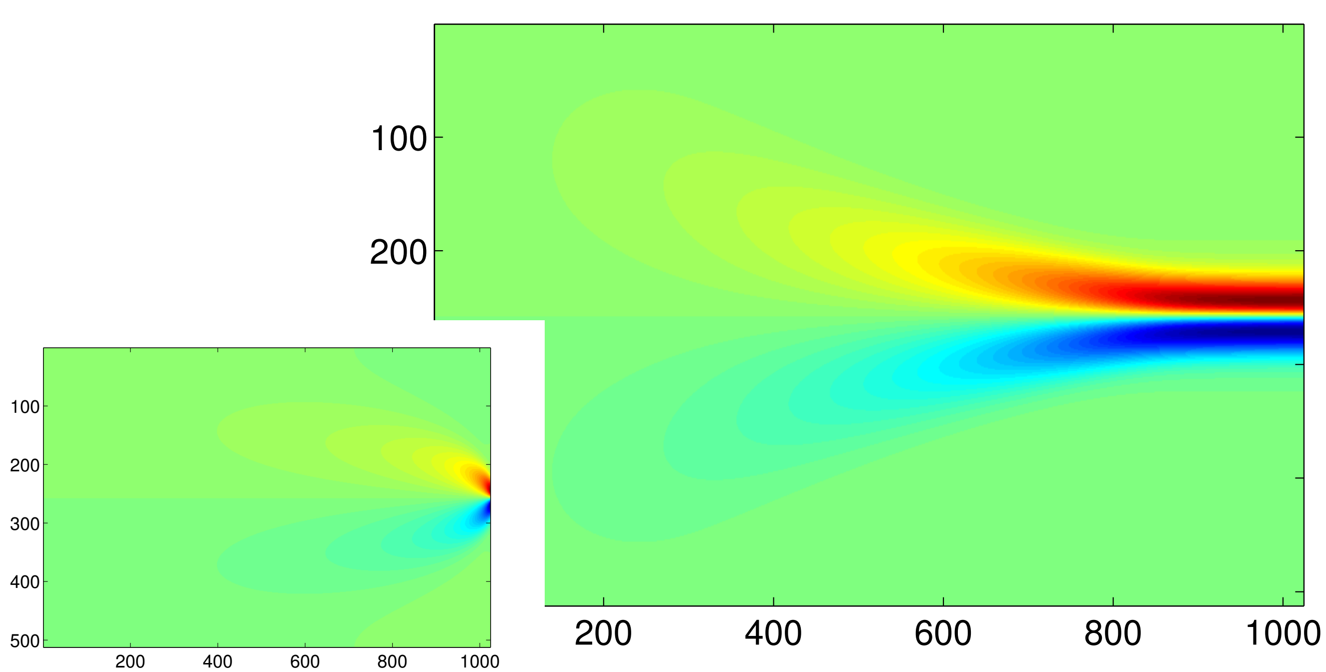
▷ Random Burgers equation (with T. Grafke, R. Grauer and T. Schäfer)

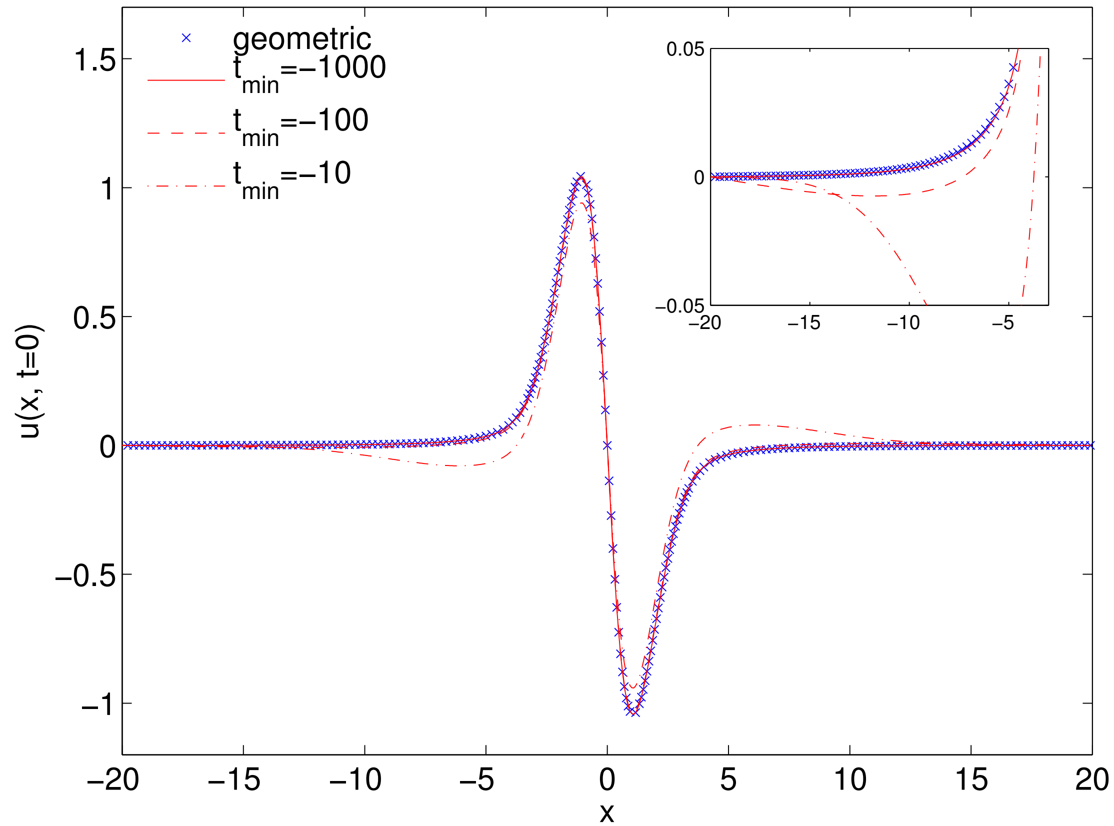
Key question: statistics of high gradients (responsible for dissipation)

$$u_t + u u_x = \nu u_{xx} + \eta(x,t), \qquad \langle \eta(x,t)\eta(x',t')\rangle = \chi(x-x')\delta(t-t')$$

Hamilton's equation associated with LD action (cf. Balkovsky *et al.;* Chernykh & Stepanov):

$$u_t + u u_x - \nu u_{xx} = \int \chi(x-x')p(x')dx'$$
$$p_t + u p_x + \nu p_{xx} = 0.$$



# Application to fluid dynamics problems

▷ 2D barotropic QG equation (with W. Ren & X. Yang)

Dissipation by Eckman drag - fixes the noise by FDT

$$d\mu = Z^{-1}\exp(-\varepsilon^{-1}Q)\delta(E-E_0)\,dUd\psi,$$

$$\frac{\partial q}{\partial t} + \nabla^\perp\psi\cdot\nabla q + U(t)\frac{\partial q}{\partial x} + \beta\frac{\partial\psi}{\partial x} = -\gamma q + \sqrt{2\varepsilon\gamma}\,\eta(t,x,y) - \gamma\psi\lambda(t),$$
$$q = \Delta\psi + h,$$
$$\frac{\mathrm{d}U}{\mathrm{d}t} = \fint h\frac{\partial\psi}{\partial x}\,\mathrm{d}\boldsymbol{x} - \tilde\gamma\beta + \sqrt{2\varepsilon\tilde\gamma}\,\xi(t) + \tilde\gamma U\lambda(t),$$

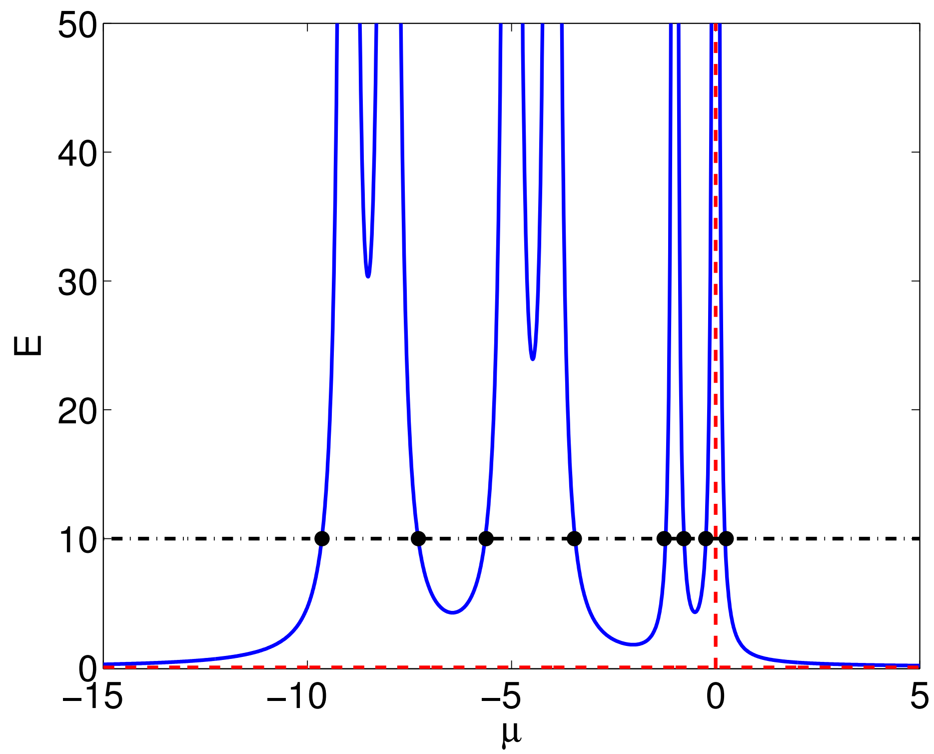
Energy $\quad E = \frac{1}{2}U^2 + \frac{1}{2}\fint |\nabla\psi|^2\,\mathrm{d}x\,\mathrm{d}y,$

Enstrophy $\quad Q = \beta U + \frac{1}{2}\fint |q|^2\,\mathrm{d}x\,\mathrm{d}y.$

Most likely states of equilibrium distribution minimize enstrophy at $E = cst$:

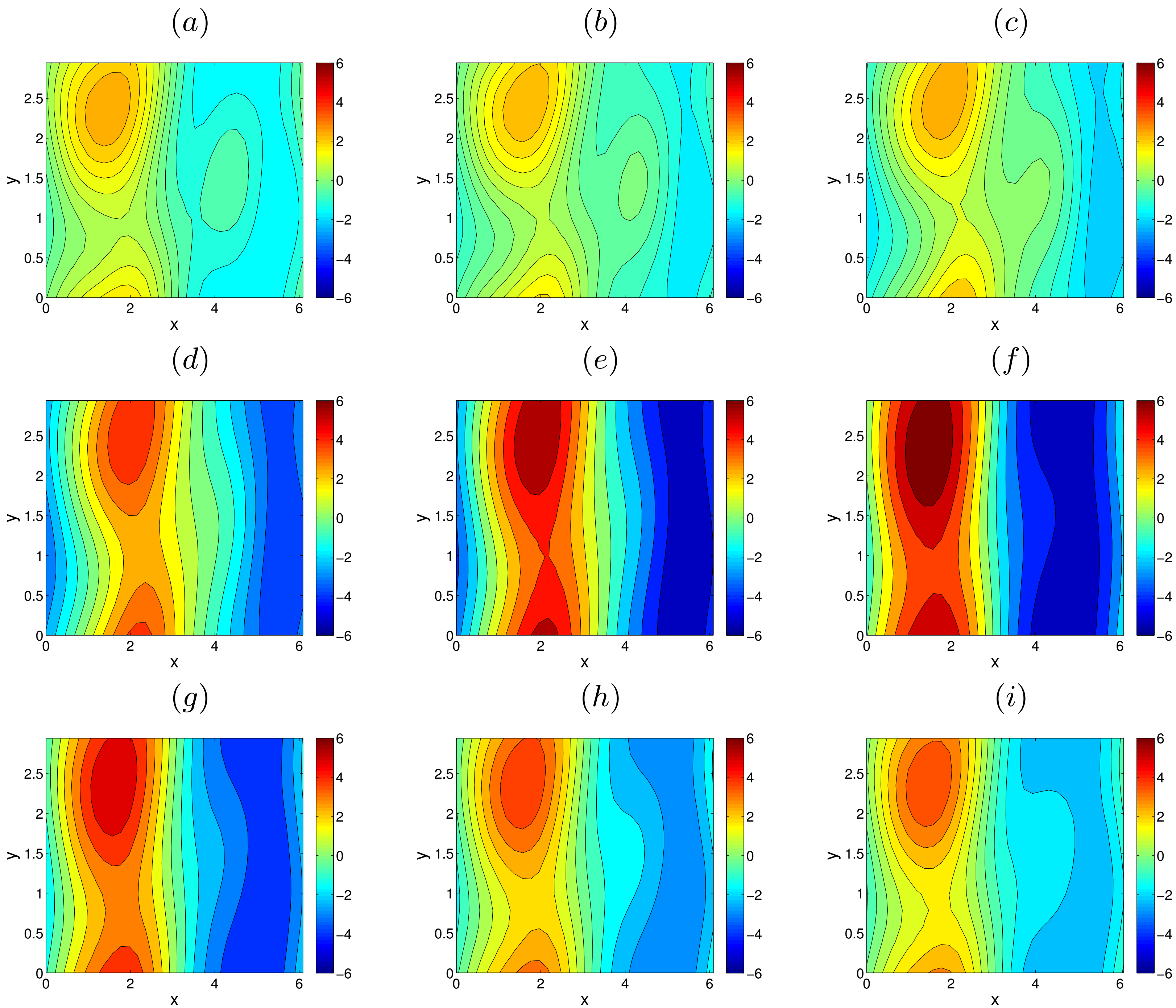
$$\mu U_c = -\beta, \quad \mu\Delta\psi_c = \Delta(\Delta\psi_c + h).$$

For $E$ large enough, there are more than one solution, out of which two are stable in absence of noise.



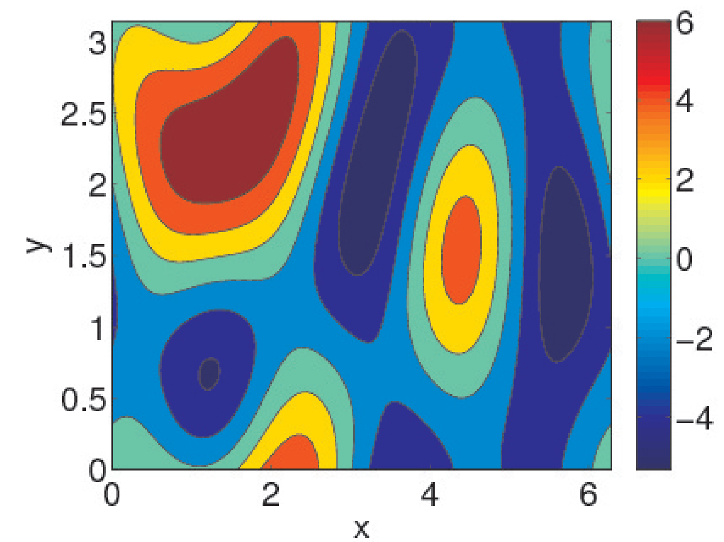
# Application to fluid dynamics problems

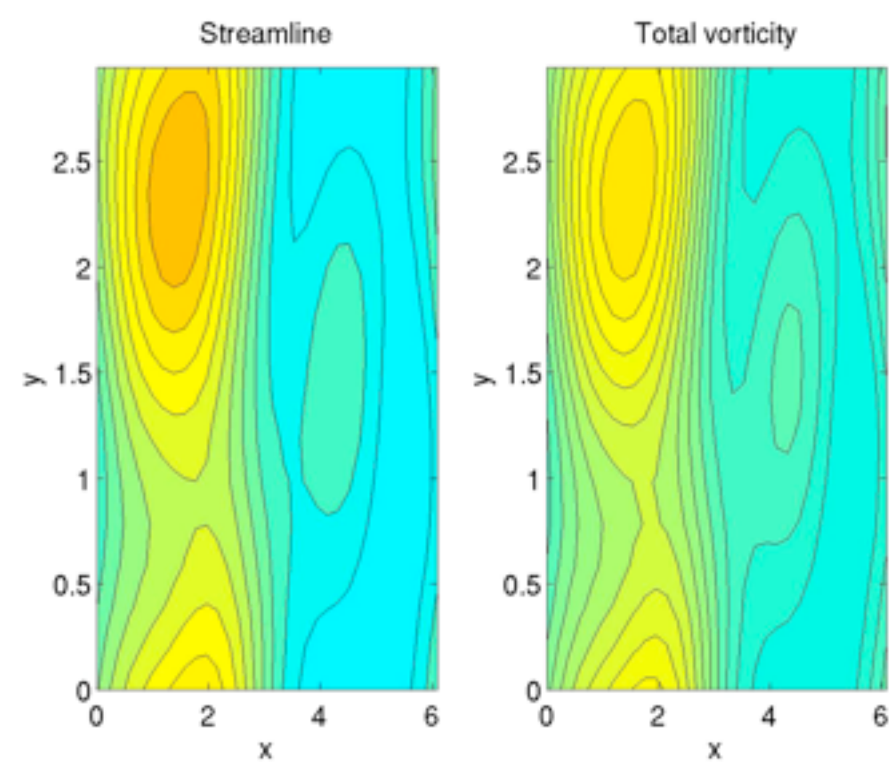
▷ 2D barotropic QG equation




Snapshots of streamfunction along PML

16 modes topography with periodic BC:



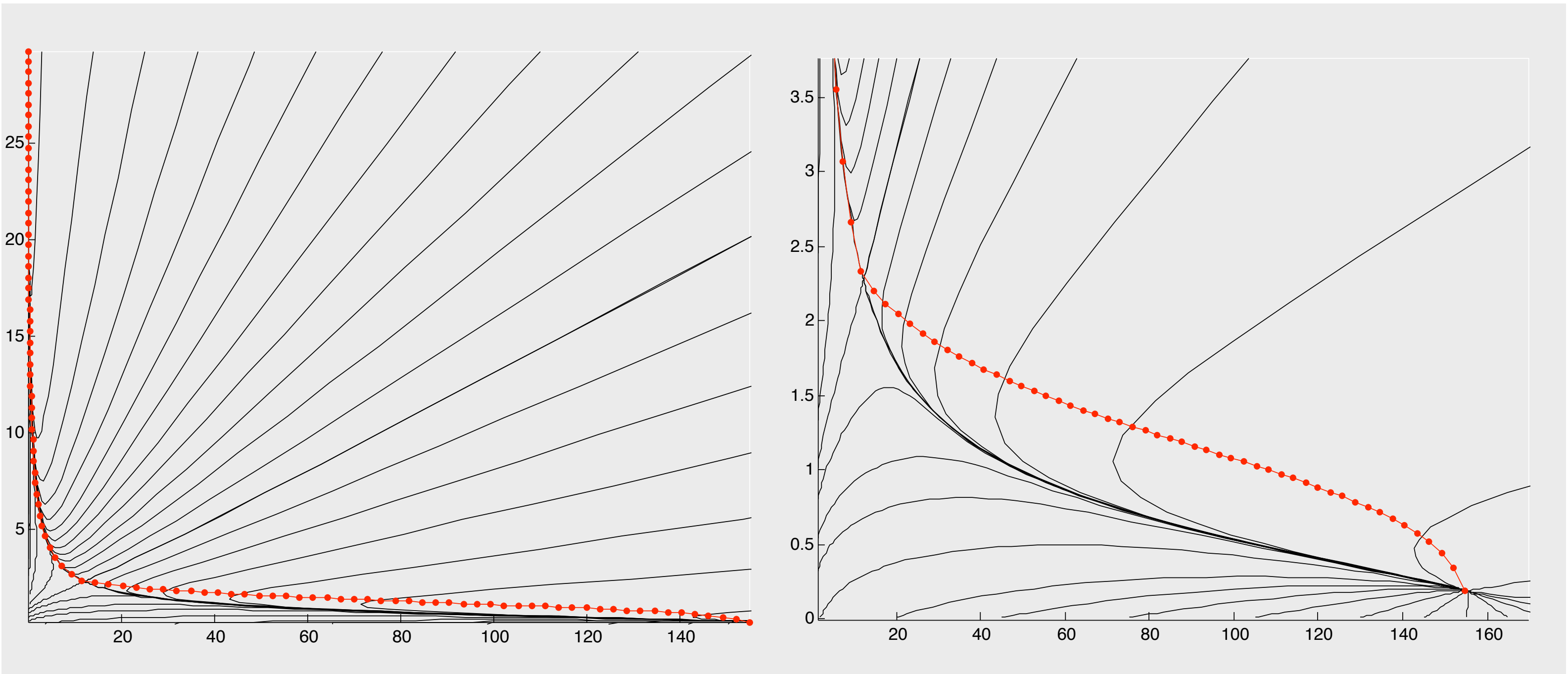
# Markov Jump Processes

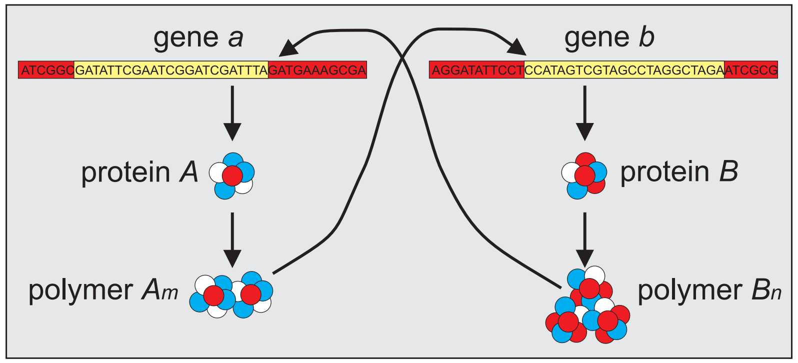
- *M* species involved in *N* reactions with rates depending on system state
- Reactions are fast but lead to small changes:

$$(Lf)(x) = \sum_{j=1}^{N} \varepsilon^{-1} a_j(x) \left( f(x + \varepsilon \nu_j) - f(x) \right)$$

- Hamiltonian of LDT: $$H(x,\theta) = \sum_{j=1}^{N} a_j(x) \left( e^{\langle \theta, \nu_j \rangle} - 1 \right)$$

▷ Application to genetic toggle switch (Roma *et al.*)






Expression of gene *a* inhibits expression of gene *b* and vice-versa

# Beyond LDT - when entropy matters

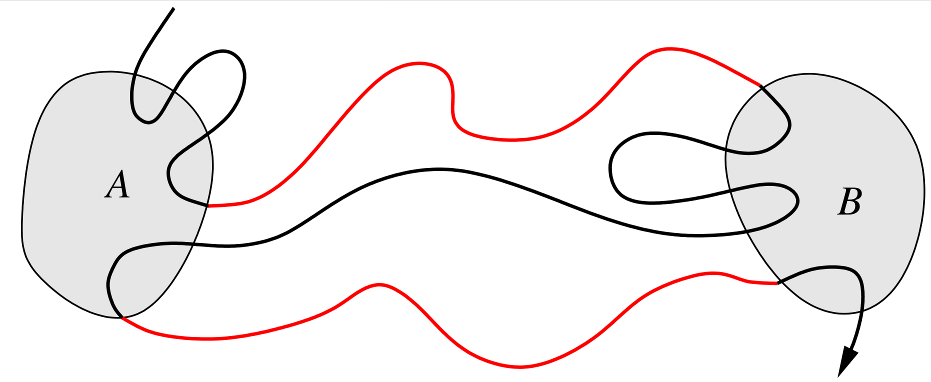
- LDT can fail if entropic effects matter
  - many alternative paths for the event,
    with lower probability individually, but large one globally.

- These situations require a more general approach to rare event analysis.

# Transition Path Theory

(Weinan E & V.-E. 2006, ...
Schütte, Metzner & V.-E. 2009,
Berezhkovskii, Hummer & Szabo 2009)

- Main idea: focus on `reactive' trajectories associated with a given transition (reaction)




- Generator = transition rate matrix $L_{i,j}$ $\quad i, j \in S = \{1, 2, \ldots, N\}$

- Microscopic reversibility (detailed balance)

$$\mu_i L_{i,j} = \mu_j L_{j,i}$$

- Probability distribution of reactive trajectories

$$\mu_i^R = \mu_i q_i (1 - q_i)$$

- Probability current of reactive trajectories

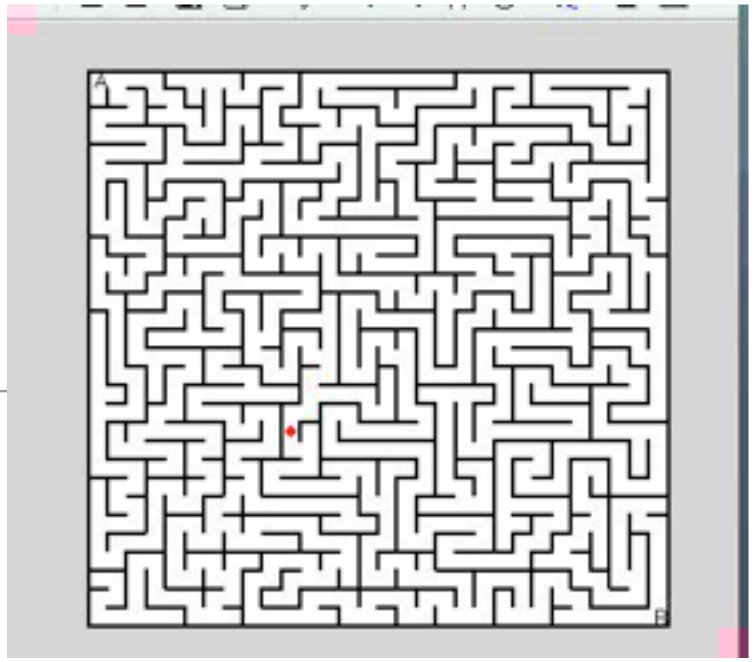
*Generator of loop erased reactive paths*

$$f_{i,j}^R = \mu_i L_{i,j} (q_j - q_i)_+ \qquad L_{i,j}^R = L_{i,j} (q_j - q_i)_+$$

- Committor function:

$$q_i = \mathbb{E}^i (\tau_B < \tau_A)$$

# The example of a maze

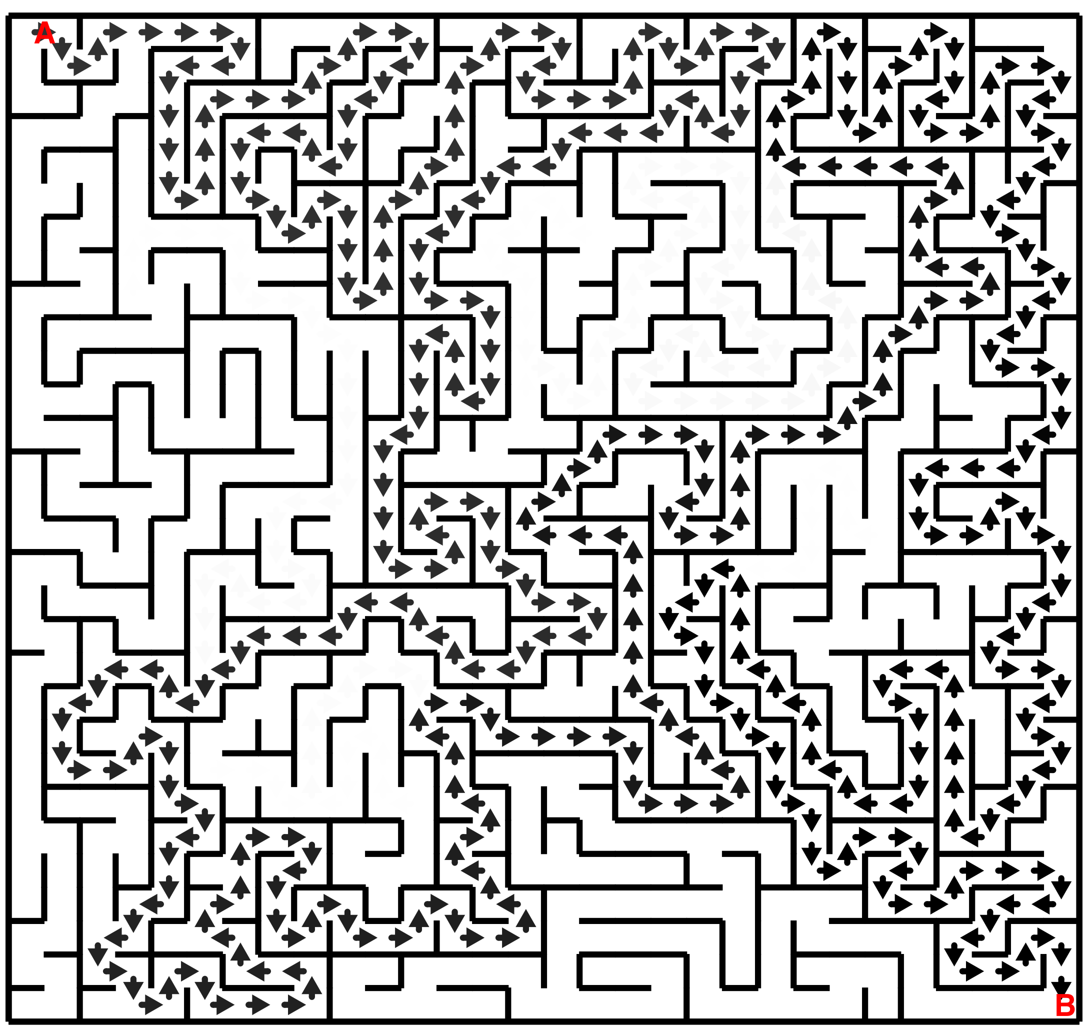


No energy landscape, no clear transition state - not a mountain pass problem.

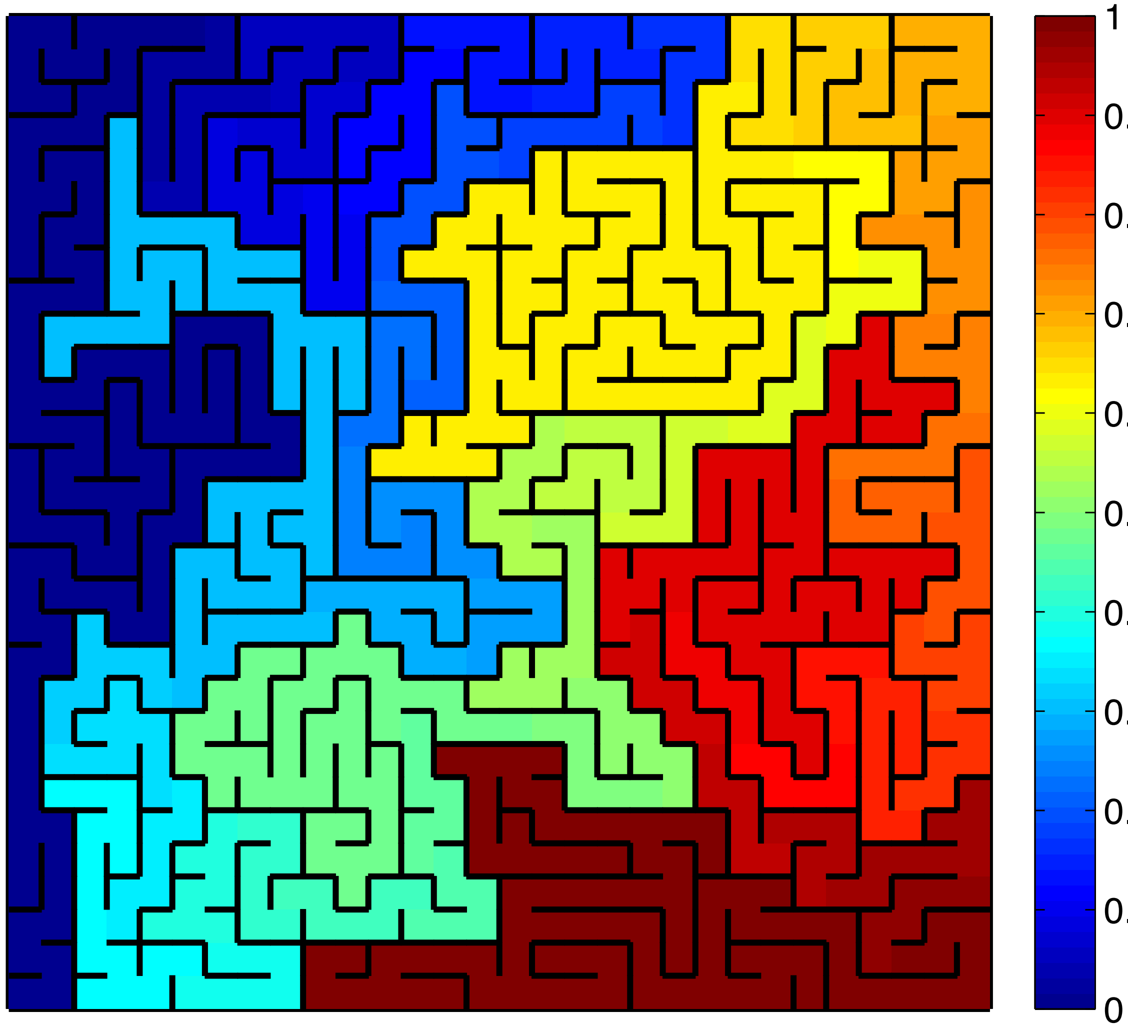
Many different reactive paths from entrance to exit - the walkers wanders in the many deadends of the maze and this affects dramatically the rate of the `reaction’.

Yet the current of these reactive trajectories concentrates on a single path.

Effective current




Committor



# Reorganization of LJ cluster after self-assembly

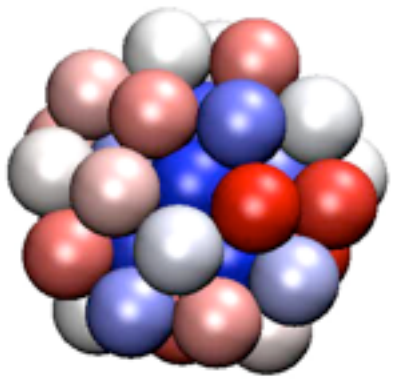
*with Masha Cameron*

- Double funnel landscape - ground state is not accessible directly by self-assembly and requires dynamical reorganization;
- Can be described by a MJP using the network calculated by Wales;
    - ~1e4 local minima of potential;
    - ~1e4 saddle points between these minima;
    - Network of orbits (minimum energy paths);
    - Transition matrix ;

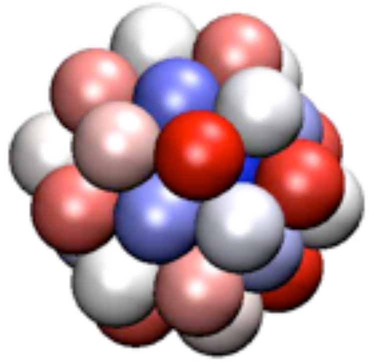
$$L_{i,j} = \nu e^{-\Delta V_{i,j}/\varepsilon}$$

- Spectral analysis difficult (large network, many small eigenvalues).
- Can be analyzed by TPT.
- LDT only applicable at extremely low temperatures. At physical temperature, system remains strongly metastable, but pathways and rate of transition are different than those predicted by LDT (entropic effects).

icosahedron

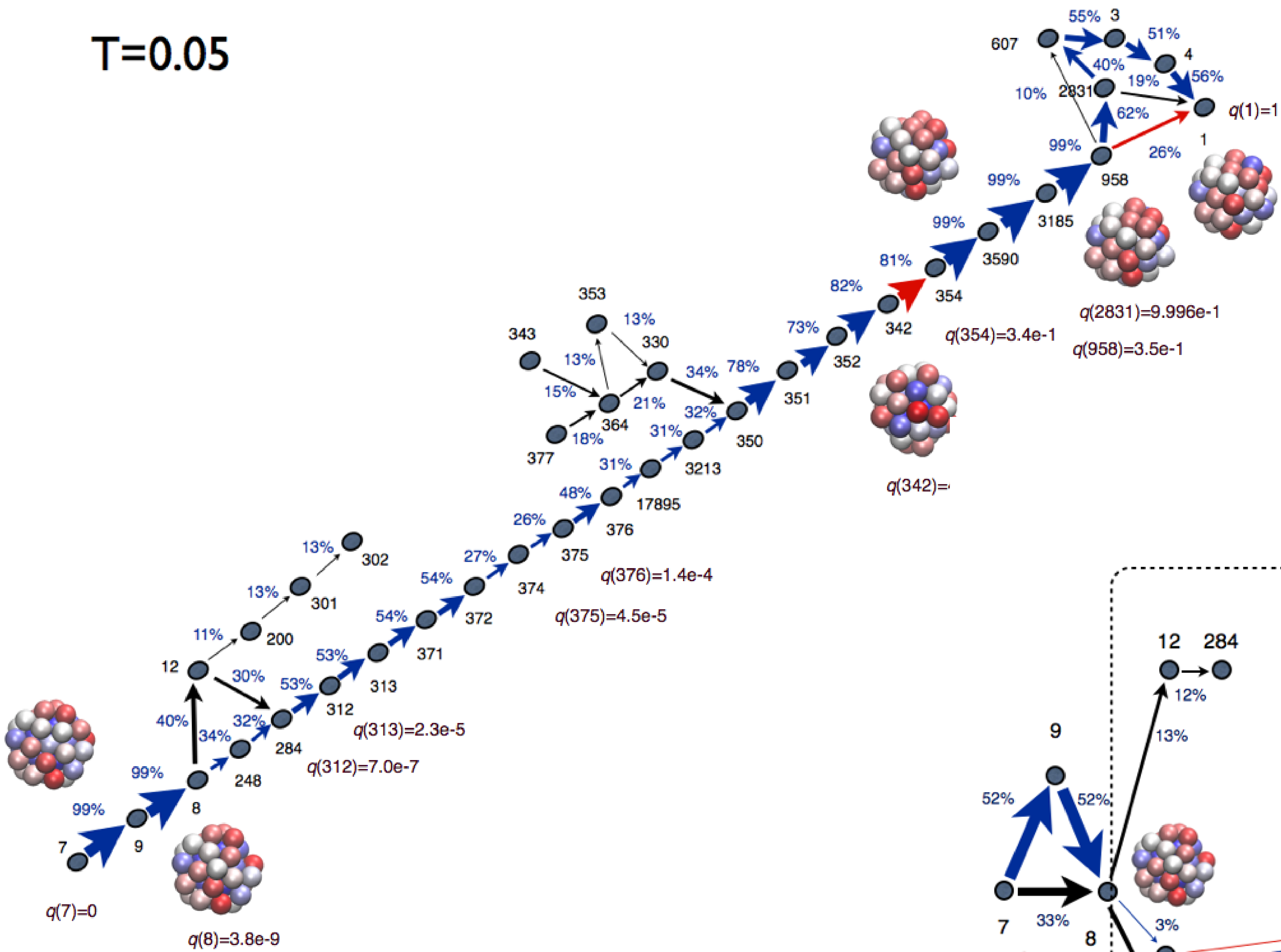


octahedron



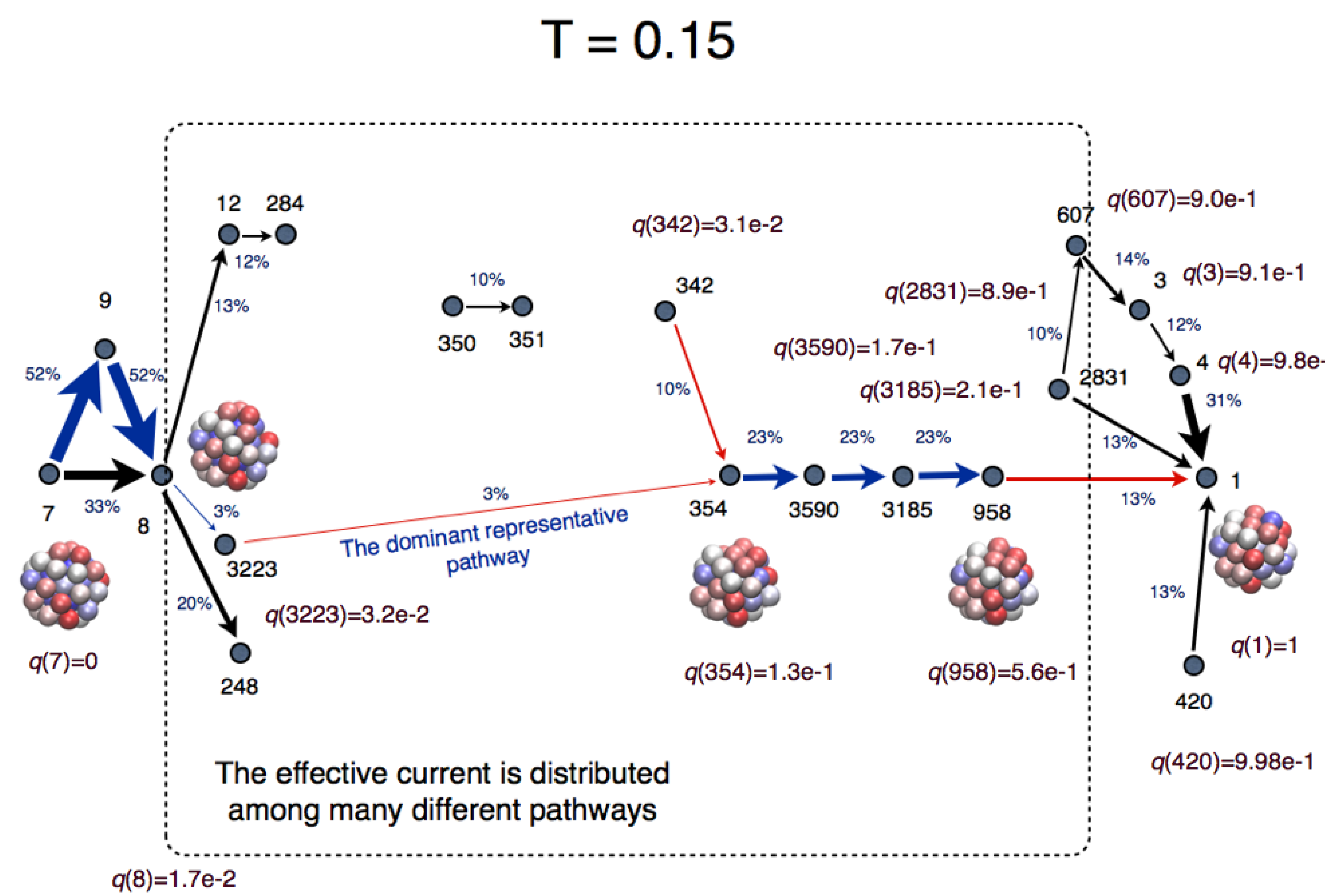
# Reorganization of LJ cluster after self-assembly

*with Masha Cameron*




Number of transition pathways increase with T

They become shorter and go thru higher barriers



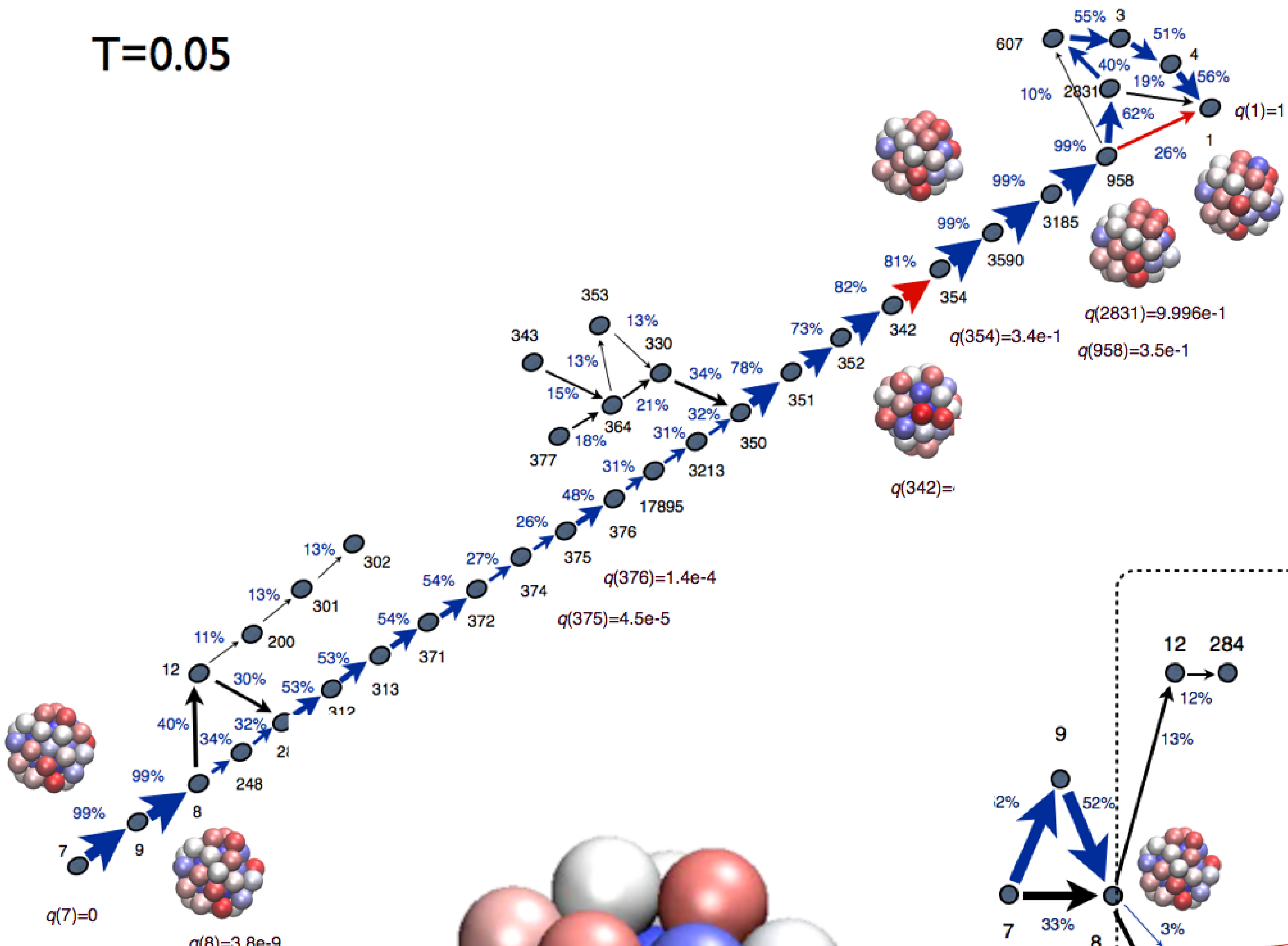
# Reorganization of LJ cluster after self-assembly

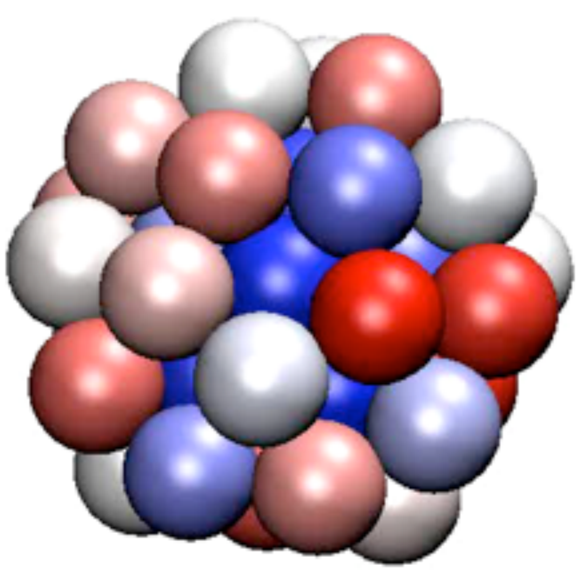
*with Masha Cameron*

Number of transition pathways increase with T

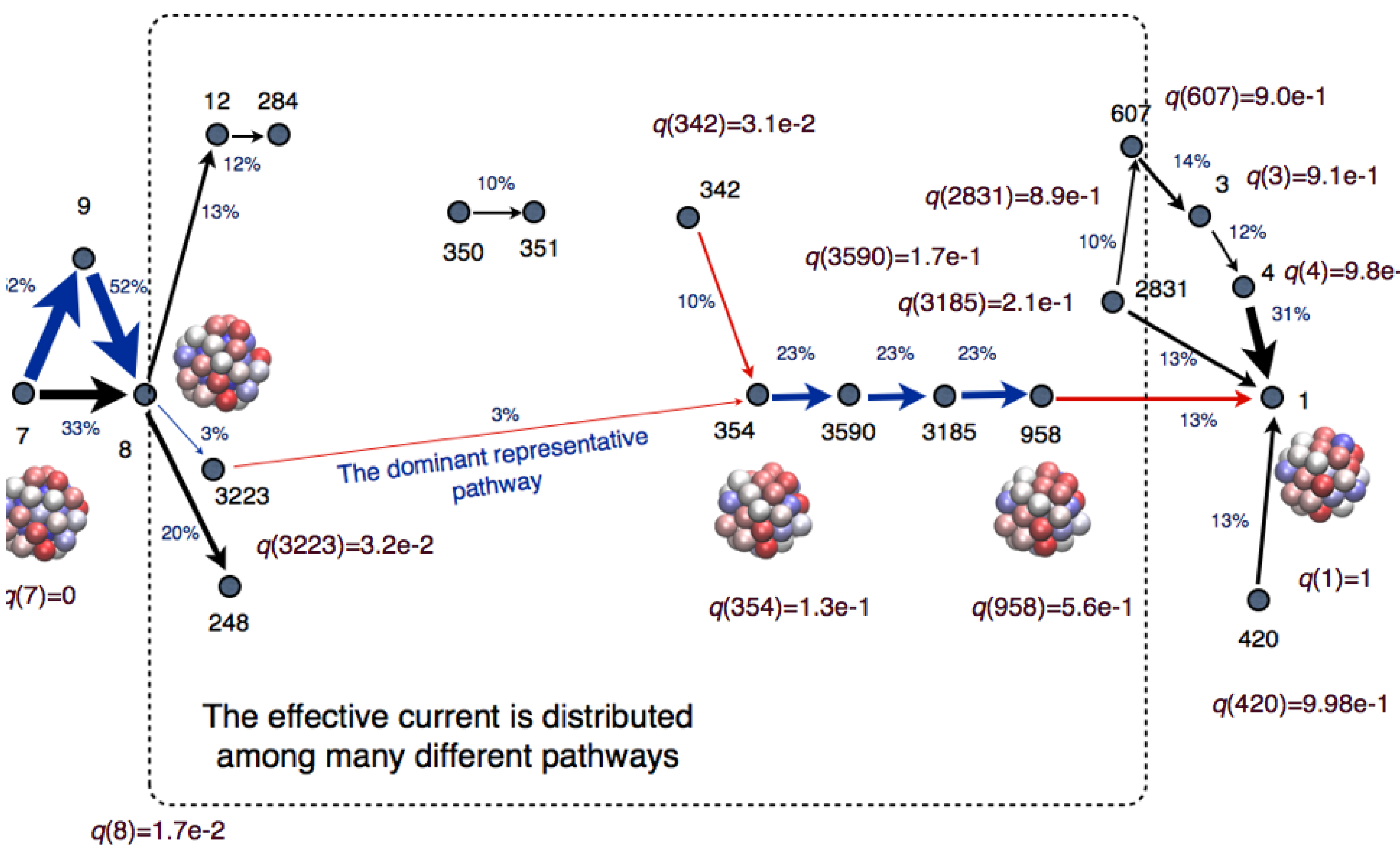
They become shorter and go thru higher barriers

**T=0.05**



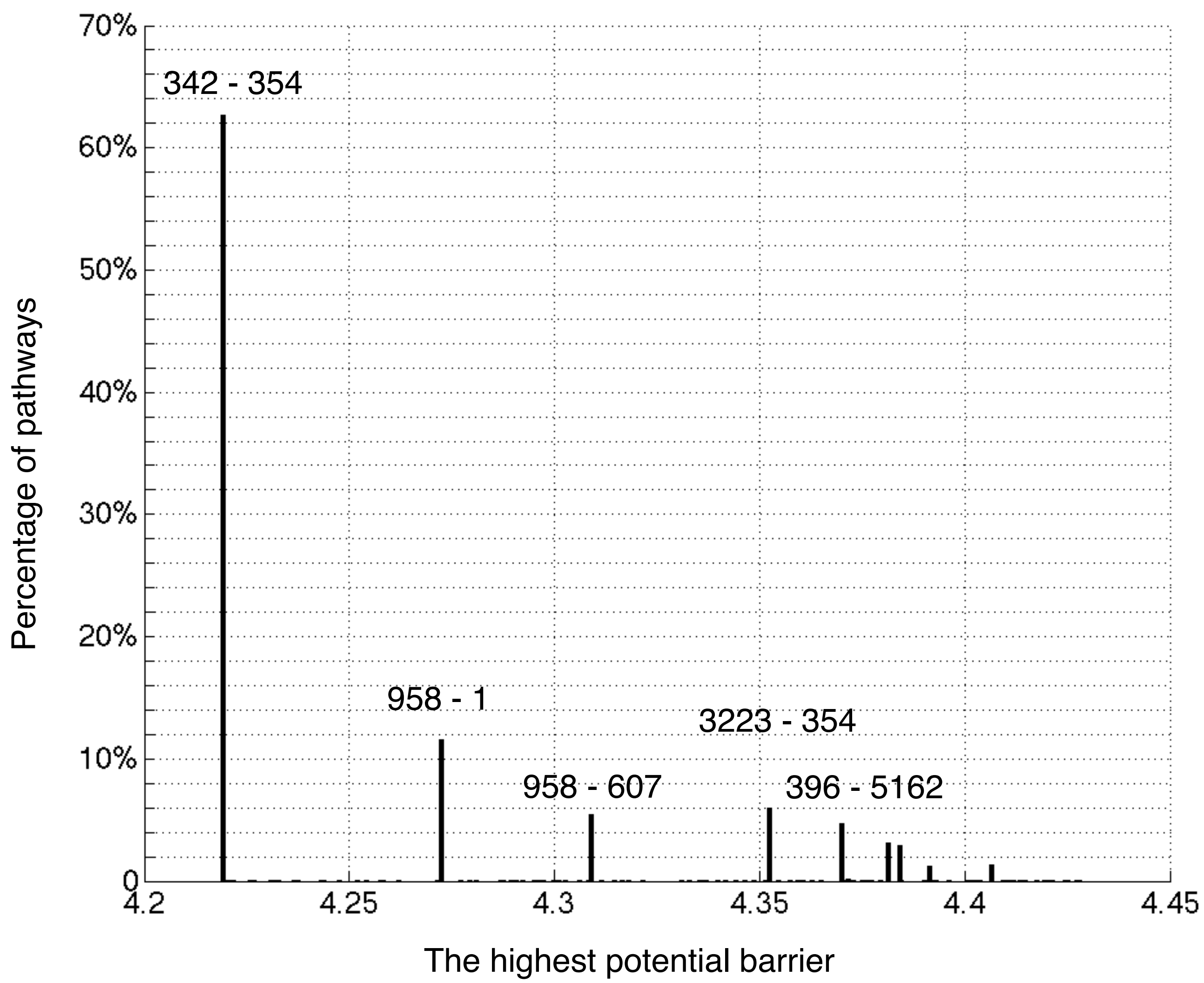
**T = 0.15**



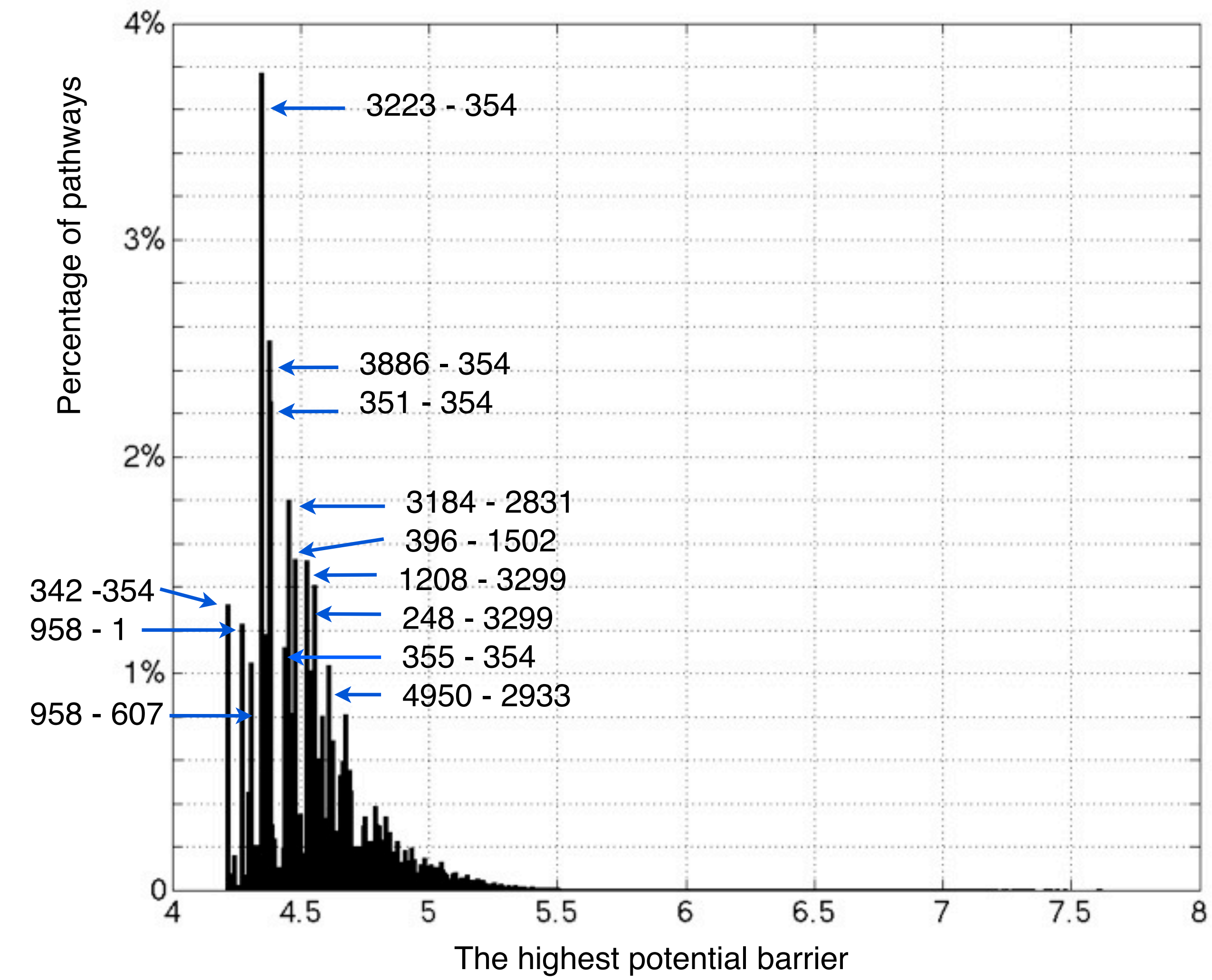
# Reorganization of LJ cluster after self-assembly

*with Masha Cameron*

T= 0.05




T= 0.15



# Conclusions

- LDT gives rough estimate of probability and path of maximum likelihood by which rare event occur.

- Can be integrated in importance sampling procedures to estimate the probability of the rare events, its rate of occurrence, etc.

- Can be used in data assimilation techniques - allows to optimally incorporate the information from noisy observations. Also offers the possibility to prevent or precipitate the occurrence of rare events in situations where the system's evolution can be acted upon.

- In situation where entropy matters and LDT becomes inapplicable, TPT is the right framework.

*Thanks to Weinan E, Weiqing Ren, Matthias Heymann, Masha Cameron, Jonathan Weare, ...*

*Thank you!*

# A few references


• W. E, W. Ren, and and E. V.-E. ``String method for the study of rare events,'' Phys. Rev. B. **66** (2002), 052301.

• W. E, W. Ren, and E. V.-E., ``Minimum action method for the study of rare events,'' Comm. Pure Applied Math. **52**, 637-656 (2004).

• M. Heymann and E. V.-E., ``The Geometric Minimum Action Method: A least action principle on the space of curves,'' Comm. Pure. App. Math. **61**(8), 1052-1117 (2008).

• W. E and E. V.-E., ``Toward a theory of transition paths,'' J. Stat. Phys. **123**, 503-523 (2006).

• W. E and E. V.-E., ``Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events,'' Annu. Rev. Phys. Chem. **61**, 391-420 (2010).