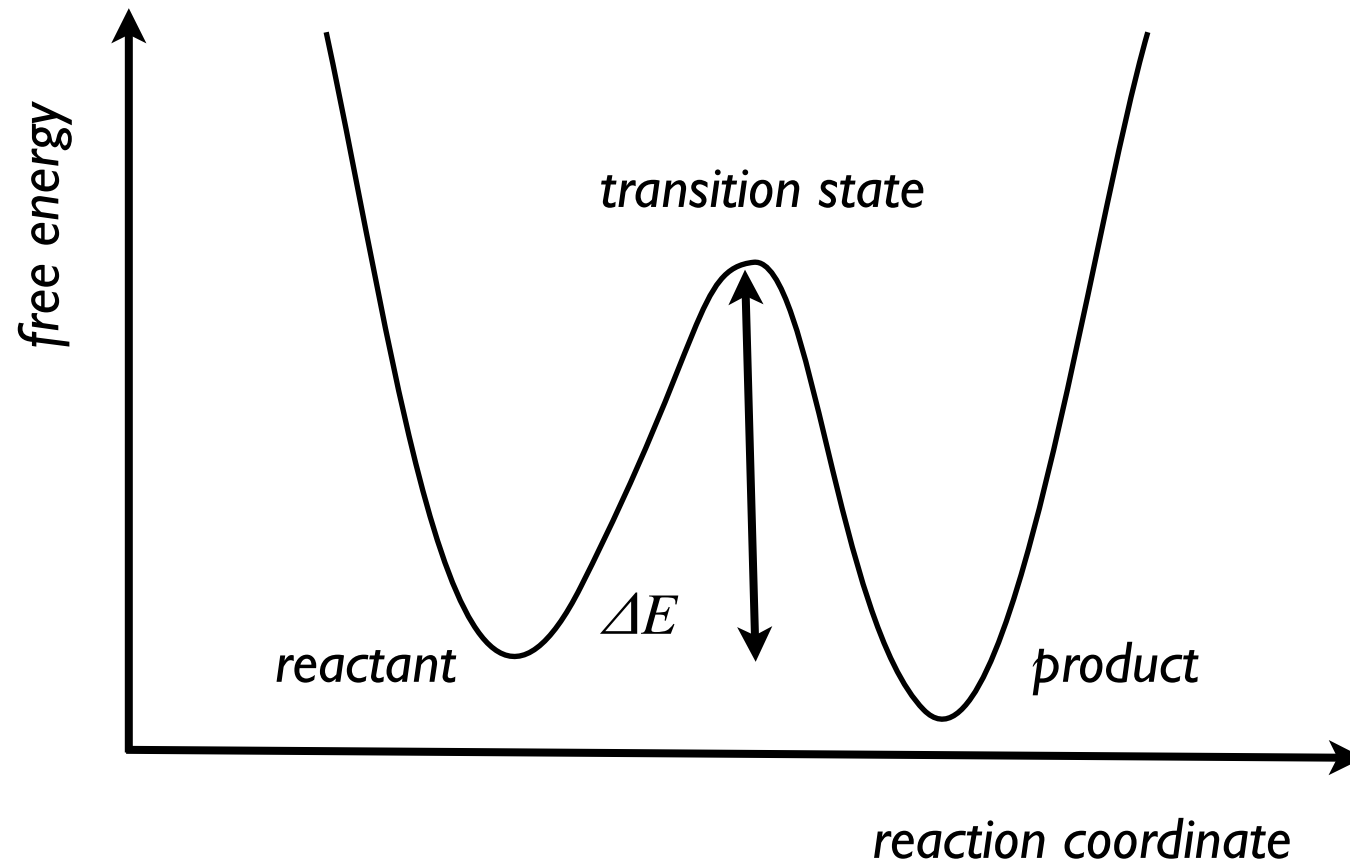




# Transition Path Theory for the Modeling and Simulation of Reactive Processes

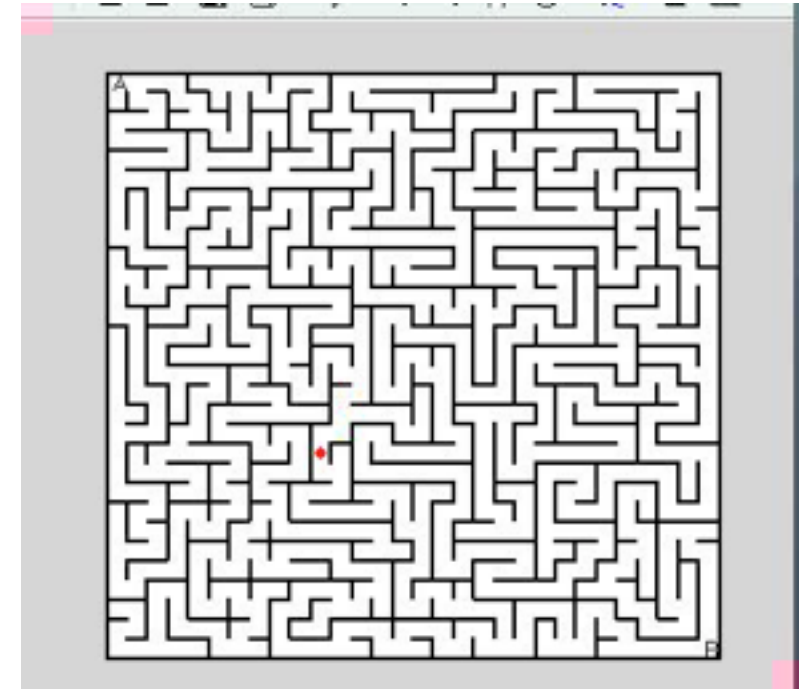
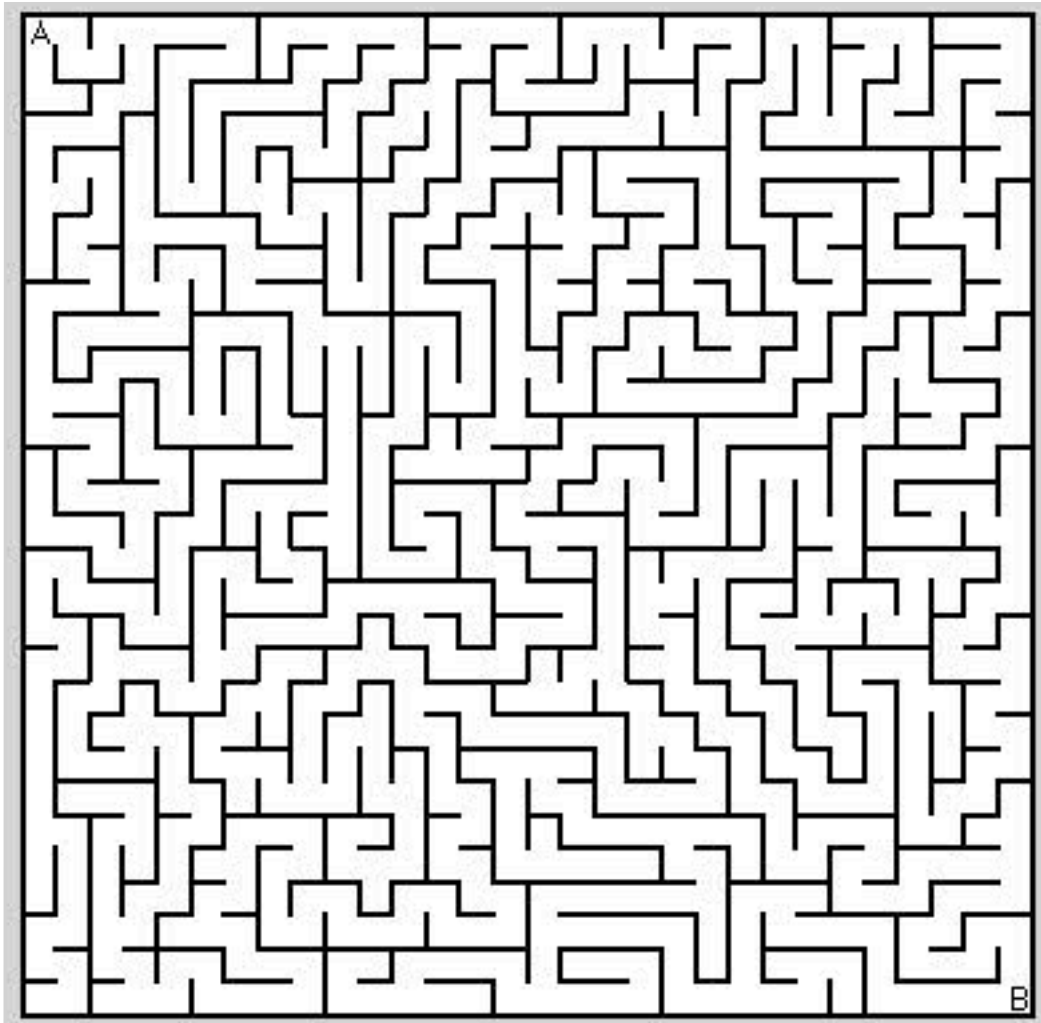
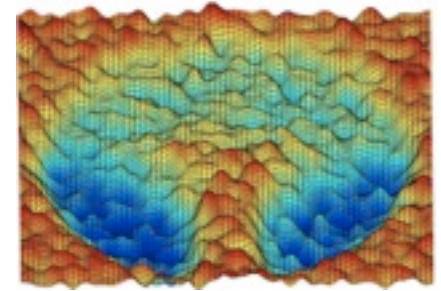
Eric Vanden-Eijnden  
Courant Institute

*Transition state theory (TST) picture of rare events:*



*Entropic (i.e. volume) effects matter, presence of dead-ends, dynamical traps, etc.*

Example: a maze



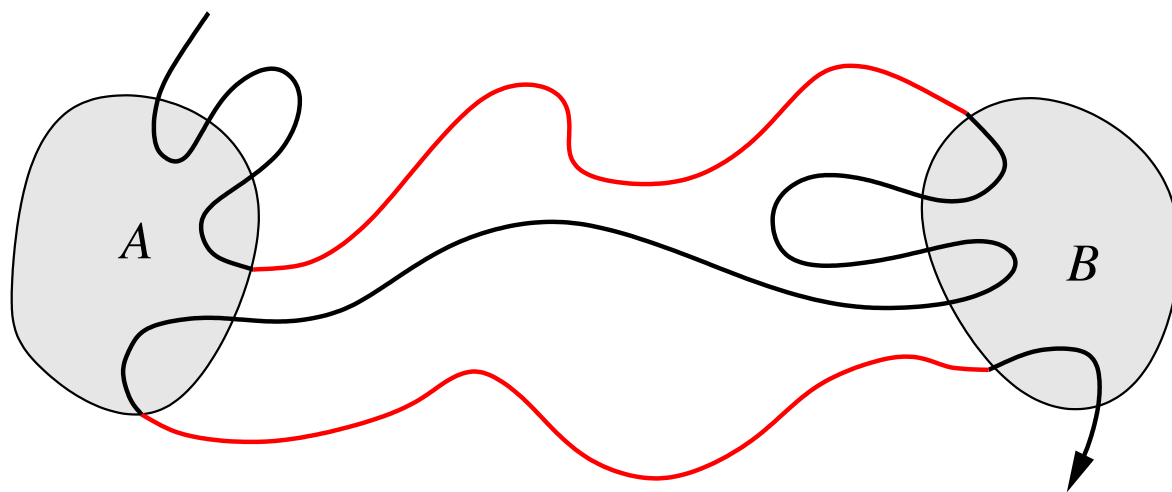
*Hard to understand by simple inspection even if the trajectory is given.*



Framework to understand general reactions: *Transition Path Theory* (E,V.-E.)

Key concept: reactive trajectories, i.e. those trajectories by which the reaction occurs.

Conceptually, these reactive trajectories can be obtained by pruning a long ergodic trajectory which oscillates between *A* and *B*.

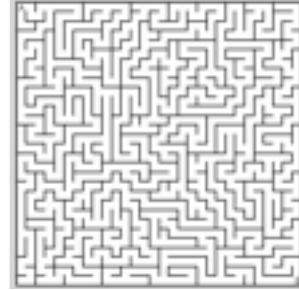


*Understanding the mechanism of the reaction*  
= *characterizing the statistical mechanics properties of the reactive trajectories*  
(i.e. the red pieces in the figure)

*Nothing special required about A and B at this point.*

Discrete set-up:  $p_{ij}$  = probability that  $x(t+1) = j$  given that  $x(t) = i$

Detailed balance:  $\pi_i p_{ij} = \pi_j p_{ji}$  ( $\pi_i$  = equilibrium distribution)



Two key questions:

What is the equilibrium probability  $\pi_i^R$  to find the trajectory at state  $i$  and that it be reactive?

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

What is the probability current of reactive trajectories from state  $i$  to state  $j$ ?

$$f_{ij}^R = \max\{f_{ij} - f_{ji}, 0\} \quad \text{where} \quad f_{ij} = (1 - q_i)\pi_i p_{ij} q_j$$

where  $q_i$  is the committor function (aka *pfold*) which gives the probability that the trajectory starting from  $i$  will reach next the product rather than the reactant.

*The committor function is the reaction coordinate because it permits (along with the equilibrium probability) to express all the statistical properties of the reactive trajectories.*

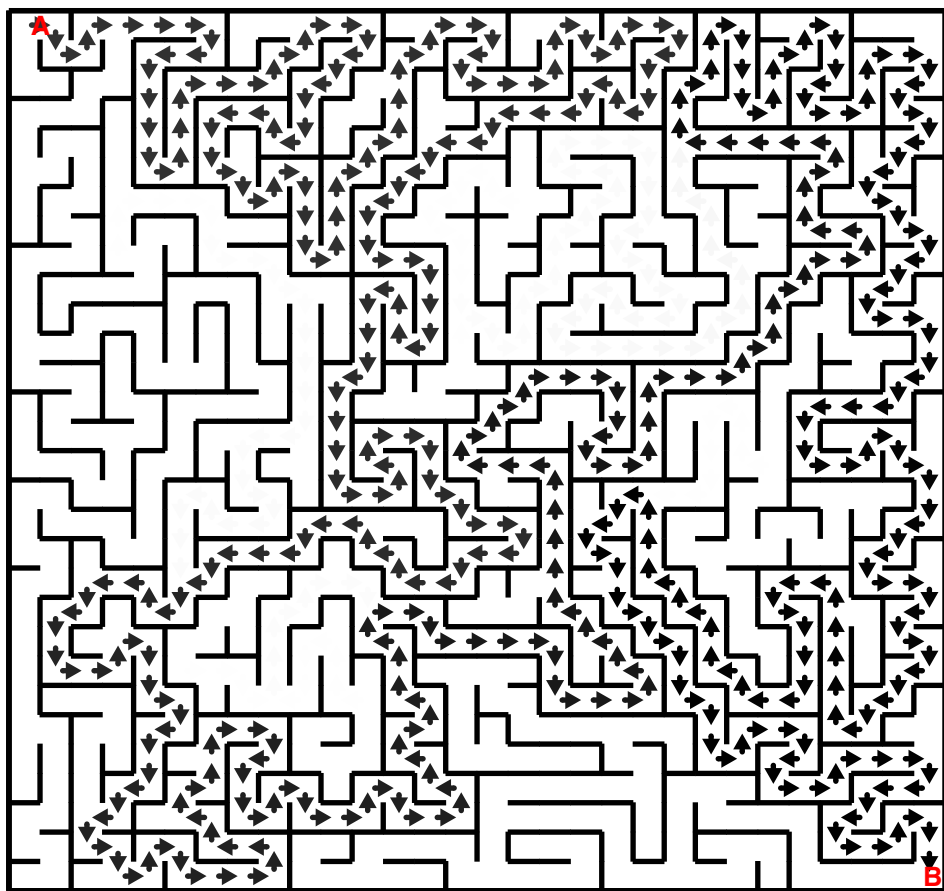
*The probability current, in particular, links concepts of reaction coordinate to that of transition pathway.*

*NB: The committor function  $q_i$  satisfies a closed equation:*

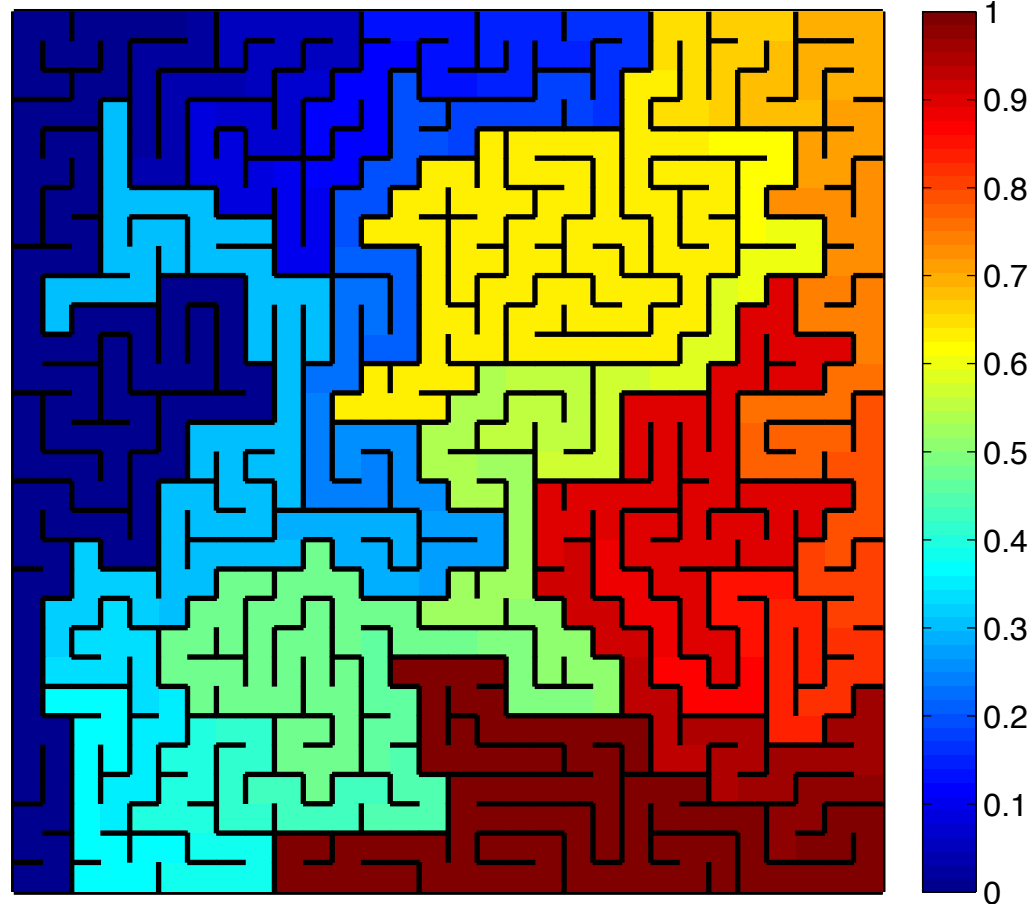
$$\begin{cases} \sum_j p_{ij} q_j = q_i & \text{if } i \notin A \cup B \\ q_i = 0 & \text{if } i \in A = \text{reactant} \\ q_i = 1 & \text{if } i \in B = \text{product} \end{cases}$$

## The maze example:

Effective current



Committor

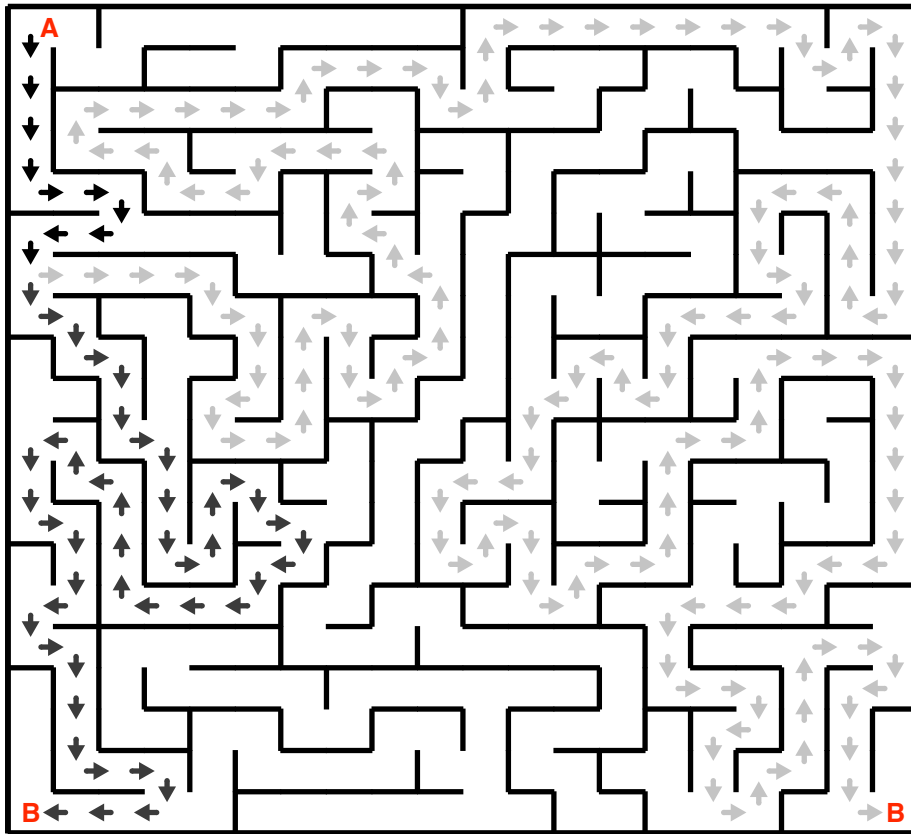


*Current concentrates on the productive path across the maze (which is unique in this example but is NOT a reactive trajectory)*

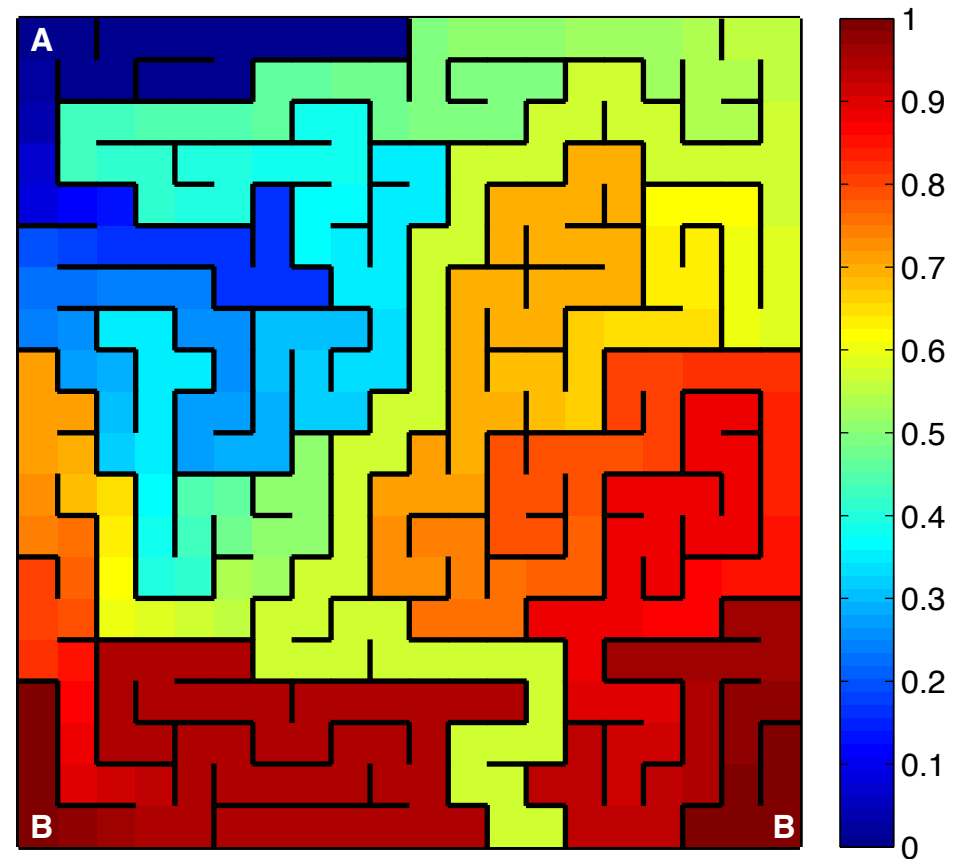
*Committor function foliates the maze and separate deadends.*

## Another maze with one entrance, two exits

Effective current

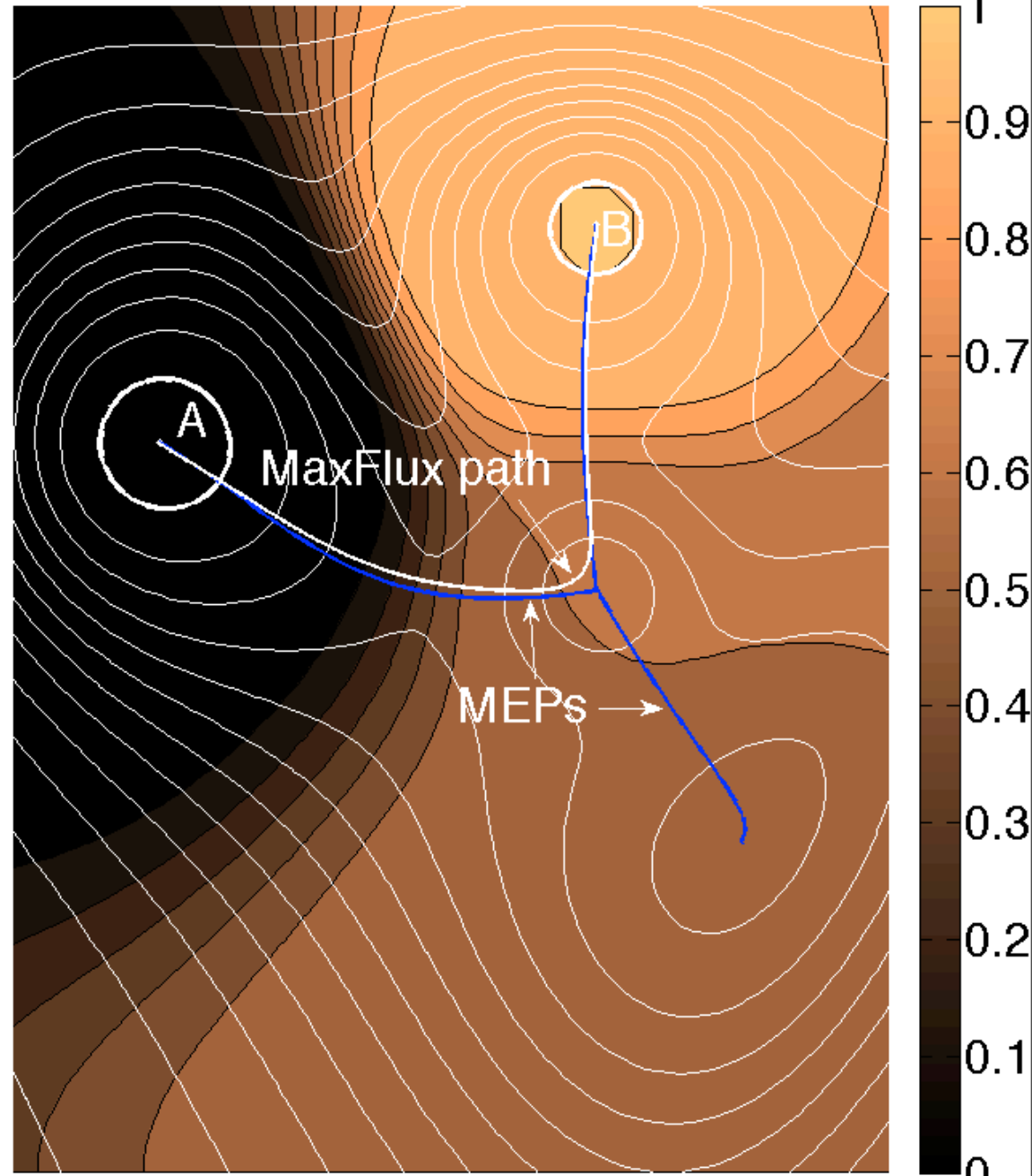
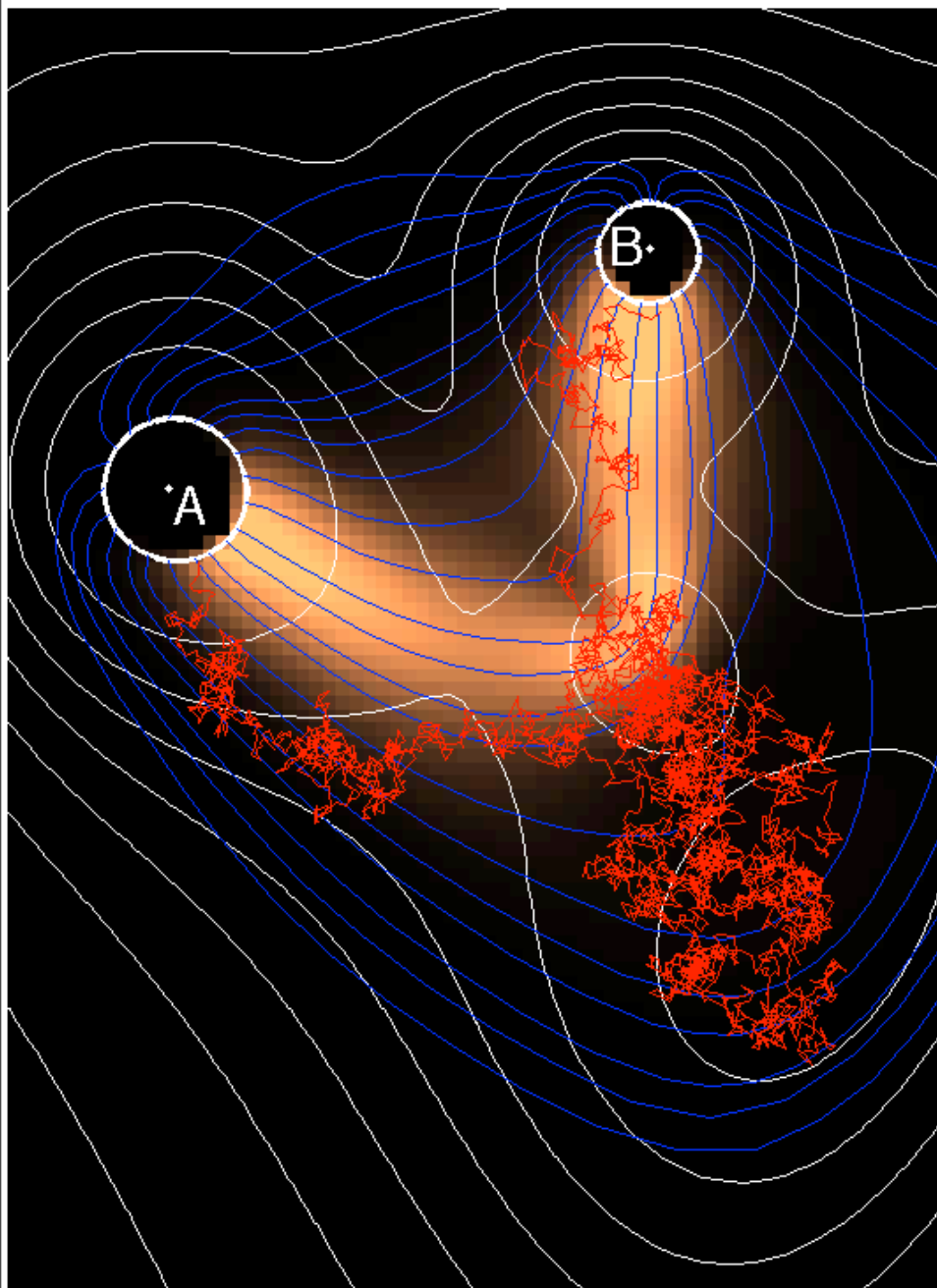


Committor



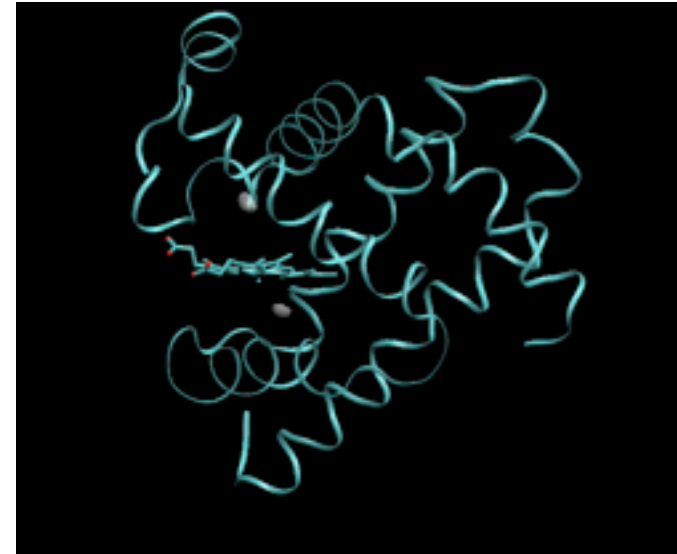
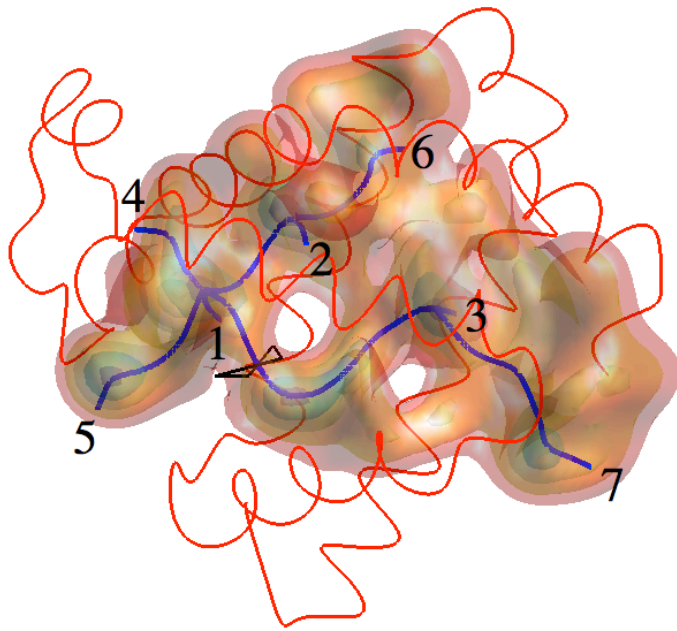
*Automatically pick the most likely of the two exits.*

## Can be generalized to continuous dynamics



*Here too, current prunes out deadends and dynamical traps (e.g. small barriers).*

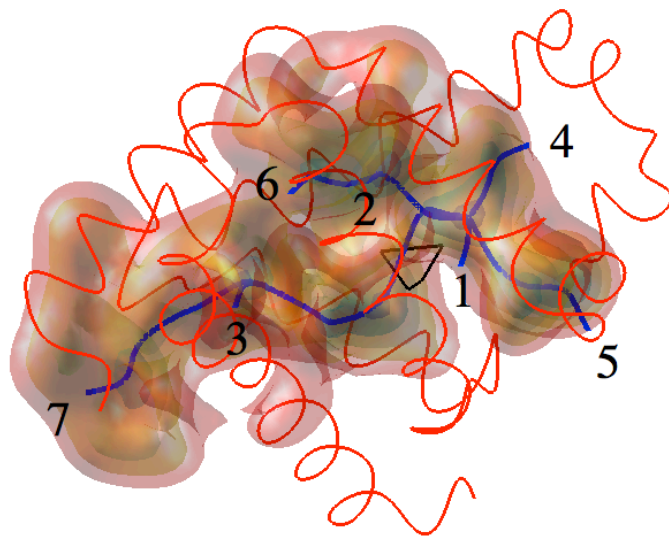
## Finding most likely exit pathways in Myoglobin (with Masha Cameron)



Compute current of reactive trajectories using the free energy (FE) landscape obtained by single sweep and TAMD (cf JACS paper with L. Maragliano, G. Cottone and G. Ciccotti).

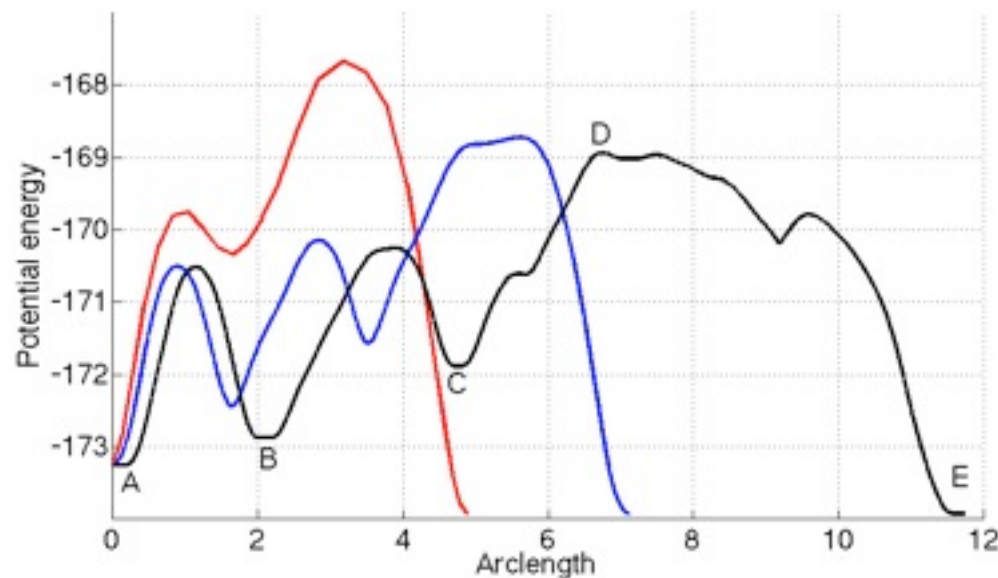
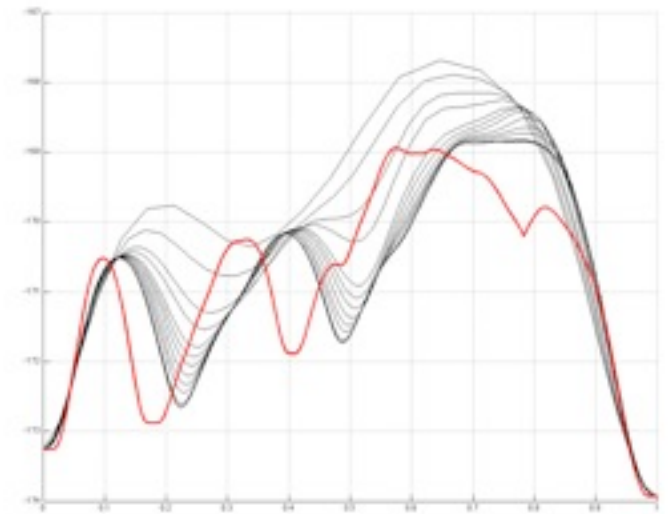
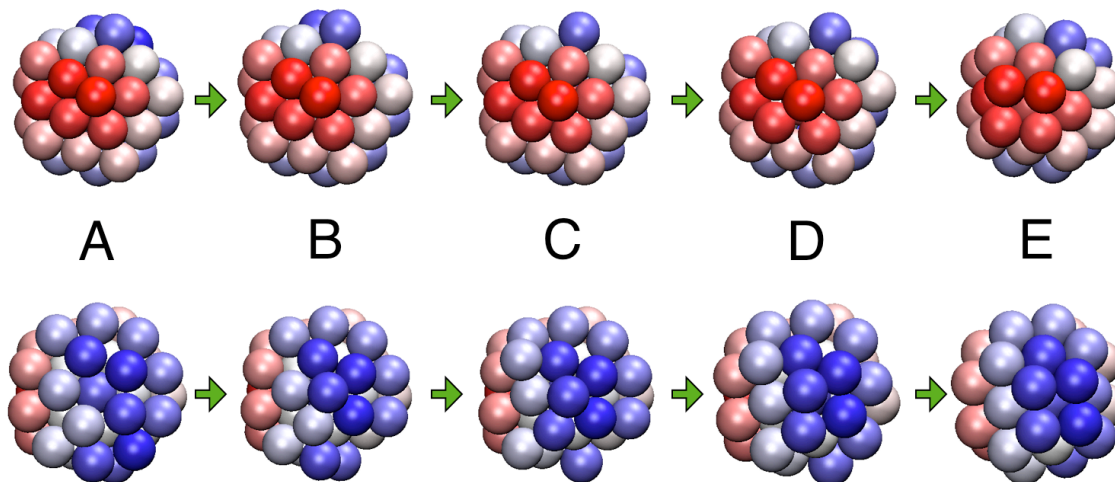
Identify lines of current carrying most of the flux - number in the figures indicates ordering.

Here A and B are the region where the FE is respectively below and above certain values.





# Finding most likely transition pathways in LJ38 (with Masha Cameron)

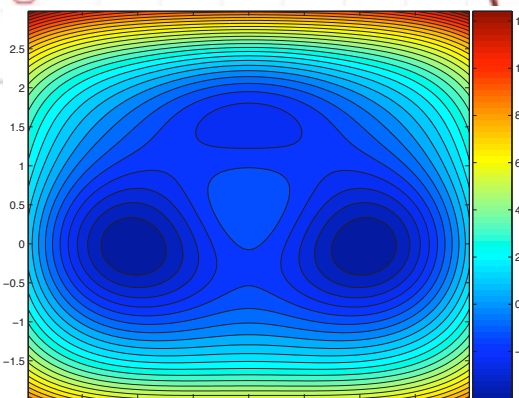
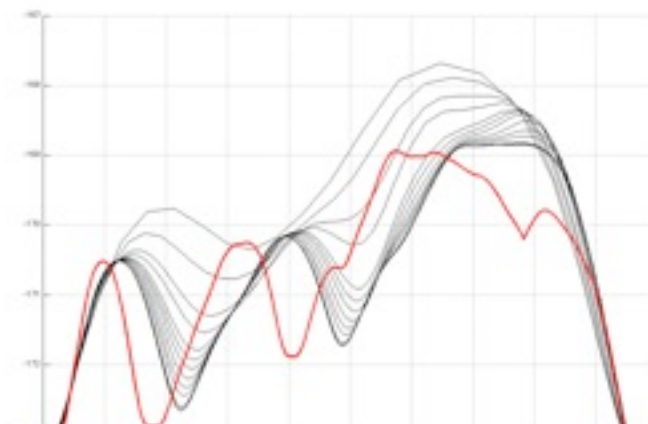
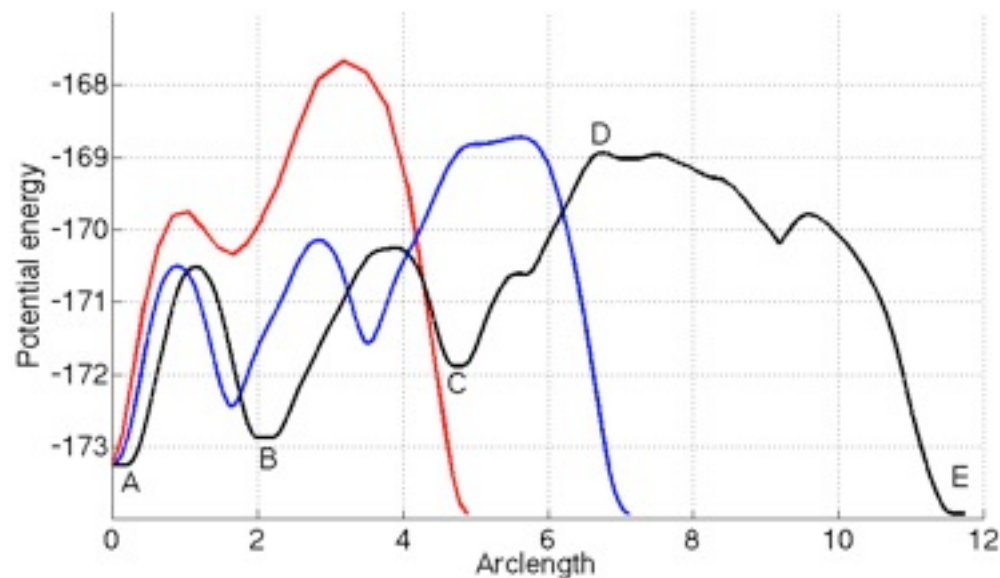
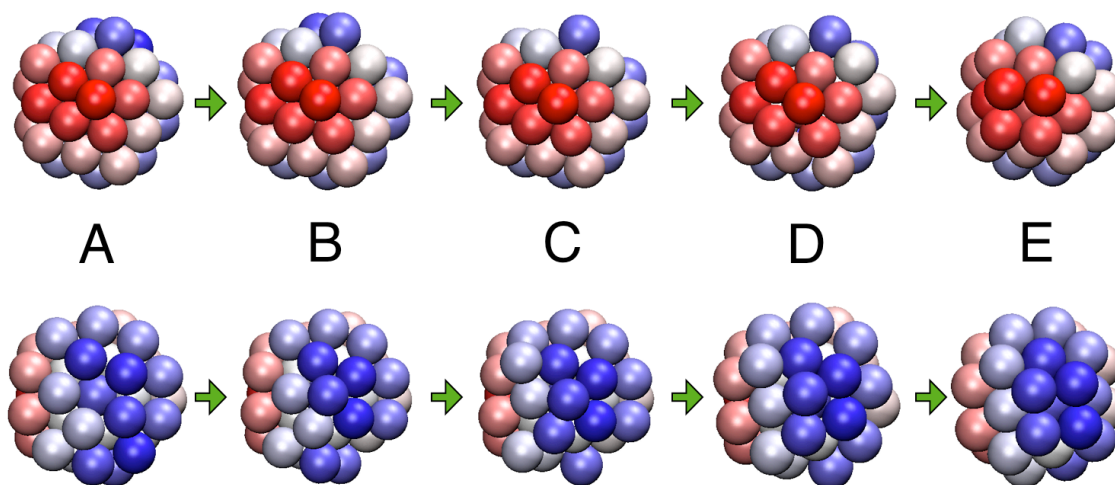


Compute directly line of MaxFlux.

Obtained by quenching temperature all the way down to  $k_B T = 0.12$ .

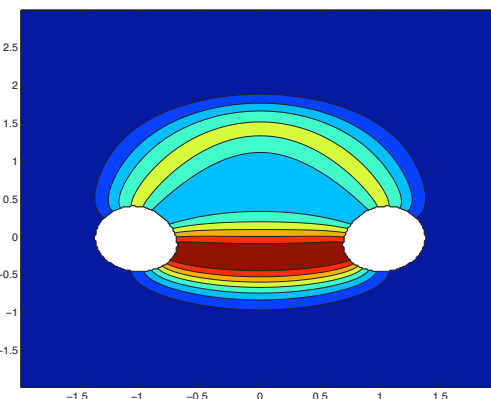


# Finding most likely transition pathways in LJ38 (with Masha Cameron)



Compute

Obtained  
temperat  
 $k_B T = 0$



## *Folding pathway of the PinWW domain*

F. Noe, Ch. Schuette E.V.-E., L. Reich, and T. Weikl , PNAS (2009)

*A total of 180 MD simulations were started, 100 from near-native conformations and 80 from different denatured conformations and run for 115 ns each at a temperature of 360 K. The simulations were conducted with the GROMACS program by using explicit SPC solvent and the GROMOS96 force field.*

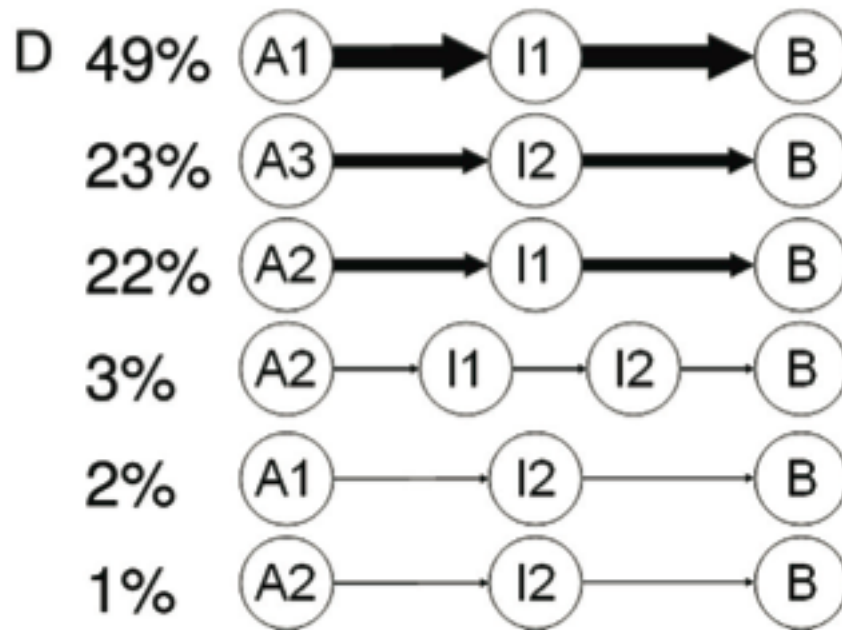
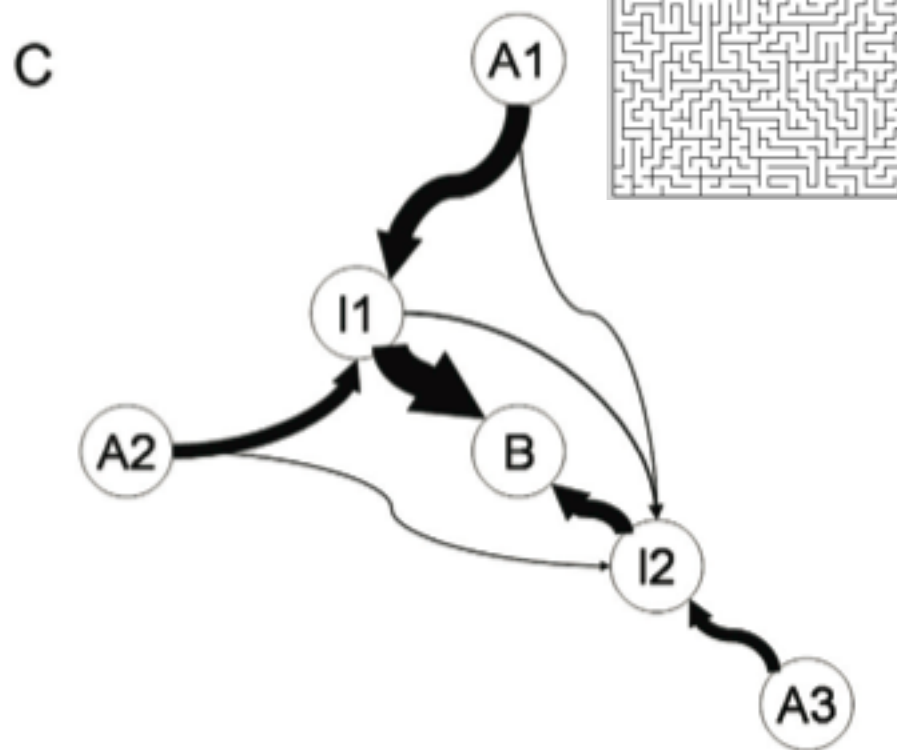
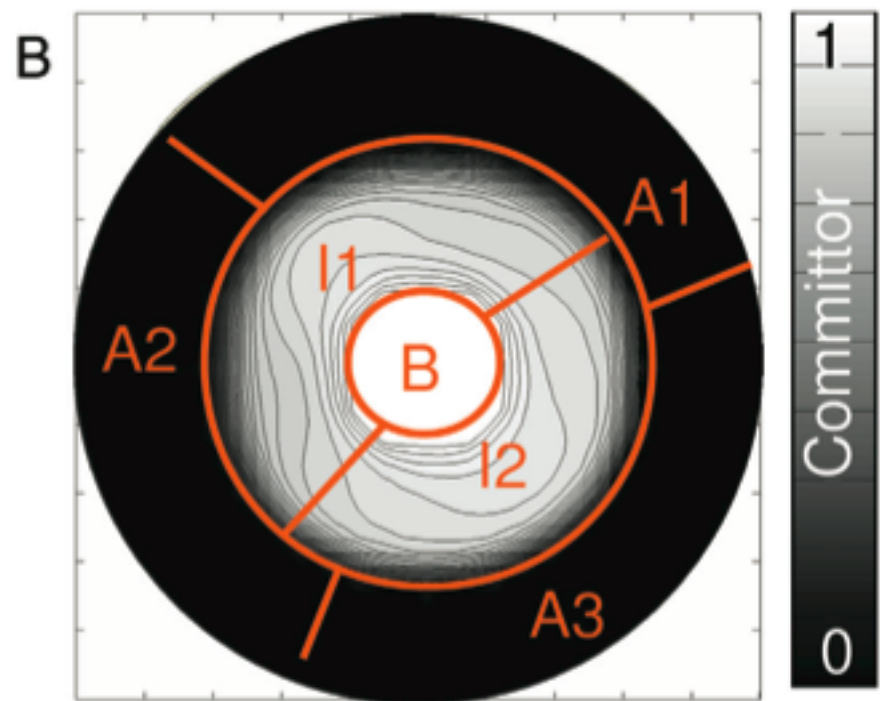
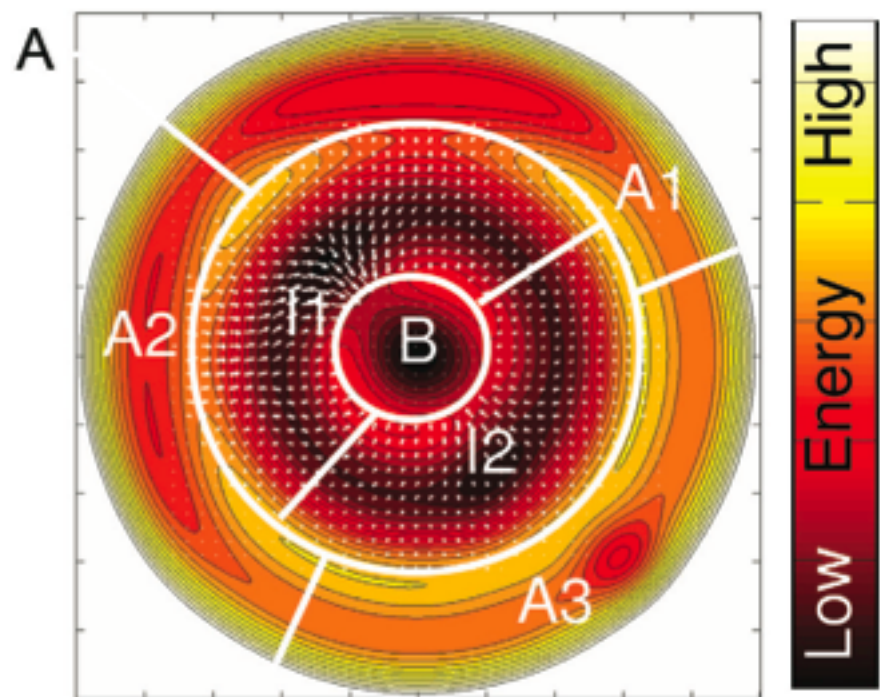
*The simulated structures were aligned onto the native structure and then clustered finely into 1734 kinetically connected and well-populated clusters.*

*Based on the MD data a stochastic matrix  $p_{ij}$  was constructed by likelihood maximization method to model the transitions between these clusters.*

*Protein like a maze with 1734 positions.*

*This Markov State Model (MSM) was then analyzed using TPT.*

# Cartoon picture of folding: motion in a maze

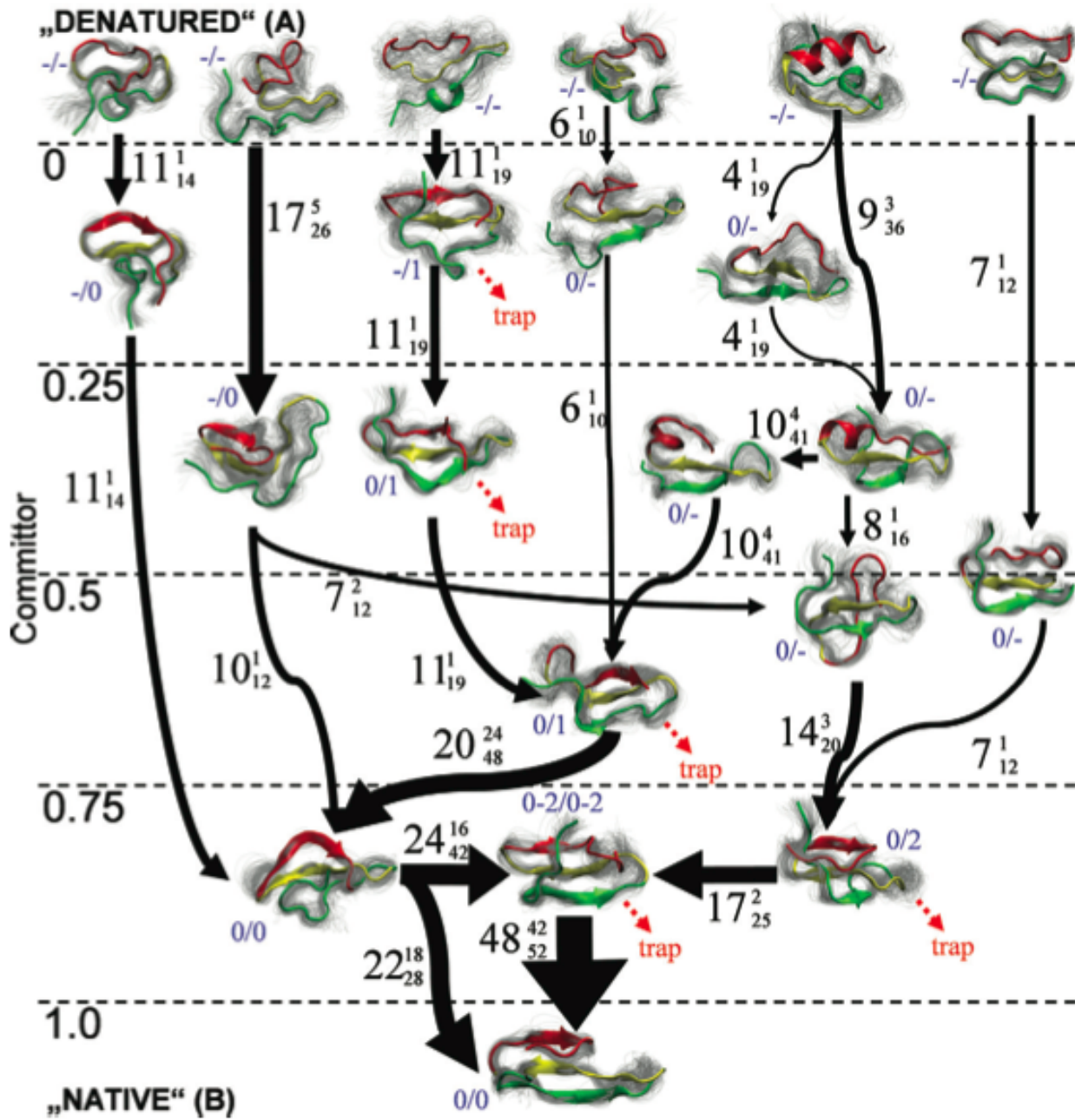




# Folding flux between macro-states

Folded set B defined as the set of clusters with average backbone root mean square difference to the X-ray structure of less than 0.3 nm.

Denatured set A defined as the set of all clusters with little  $\beta$ -structure (having a mean of  $<3$  h-bonds in hairpin 1, which has 6 h-bonds in the native state, and  $<1$  h-bonds in hairpin 2, which has 3 h-bonds in the native state)

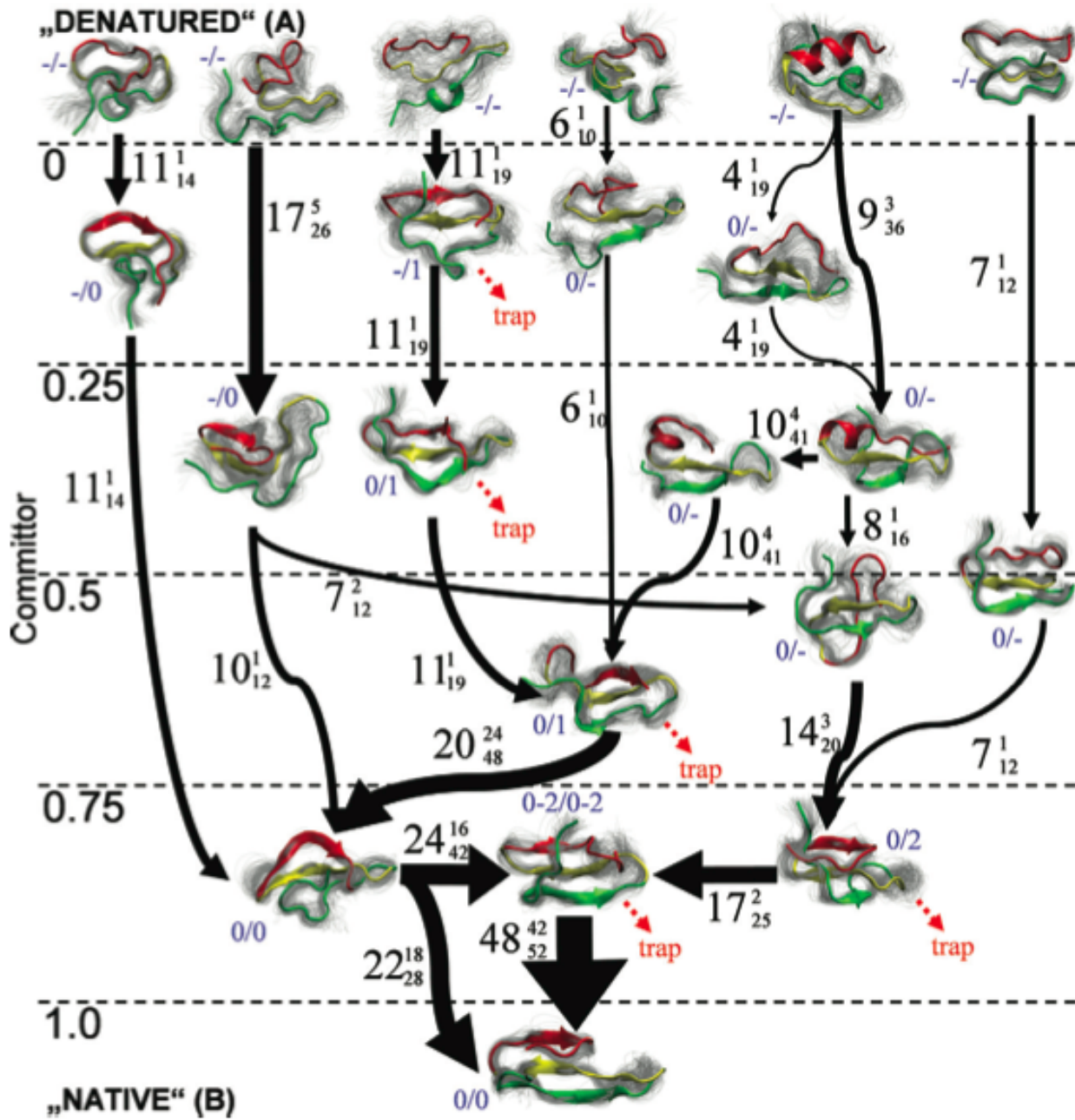


**Fig. 3.** PinWW folding flux. (Left) The network of the 70% most relevant folding pathways for PinWW. The numbers on the left indicate the committer probabilities, the thickness of the arrow indicates the flux of folding trajectories between each pair of conformations. For each conformation, a representative mean structure is shown in color along with an overlay of equilibrium-distributed structures from that conformation indicating the structural flexibility (gray cloud). The numbers next to the arrows give the normalized net flux (large number) and the 80% confidence-interval limits (small numbers) in percentages. The blue numbers next to the structures indicate whether the first/second hairpin has the native register (0), is register-shifted by one or two residues (1,2) or is not formed at all (-). (Lower Right) Register-shifted trap states that do not carry significant folding flux but reduce the folding speed by nearly a factor of 2.

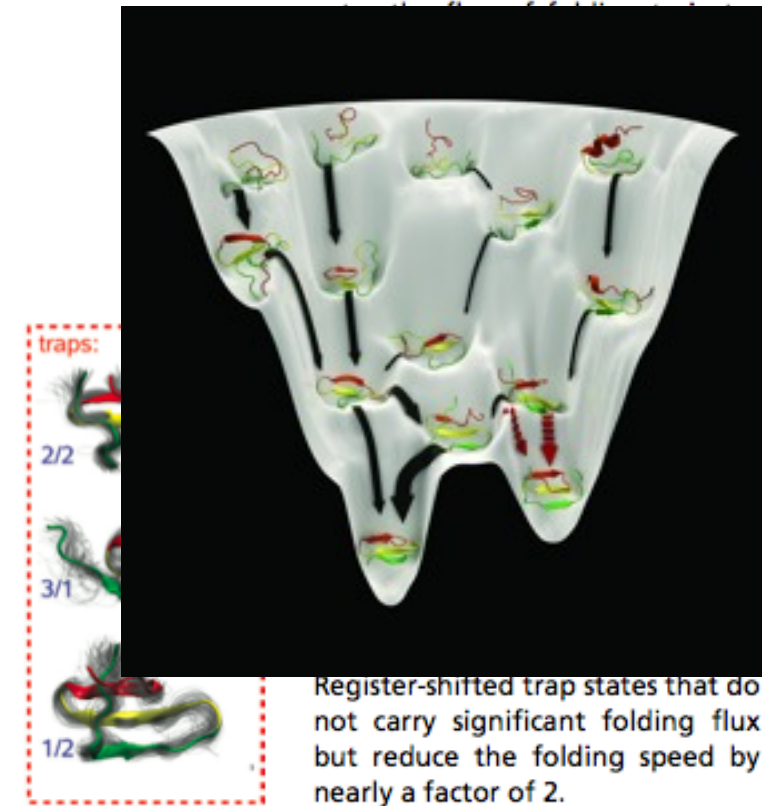
# Folding flux between macro-states

Folded set B defined as the set of clusters with average backbone root mean square difference to the X-ray structure of less than 0.3 nm.

Denatured set A defined as the set of all clusters with little  $\beta$ -structure (having a mean of  $<3$  h-bonds in hairpin 1, which has 6 h-bonds in the native state, and  $<1$  h-bonds in hairpin 2, which has 3 h-bonds in the native state)

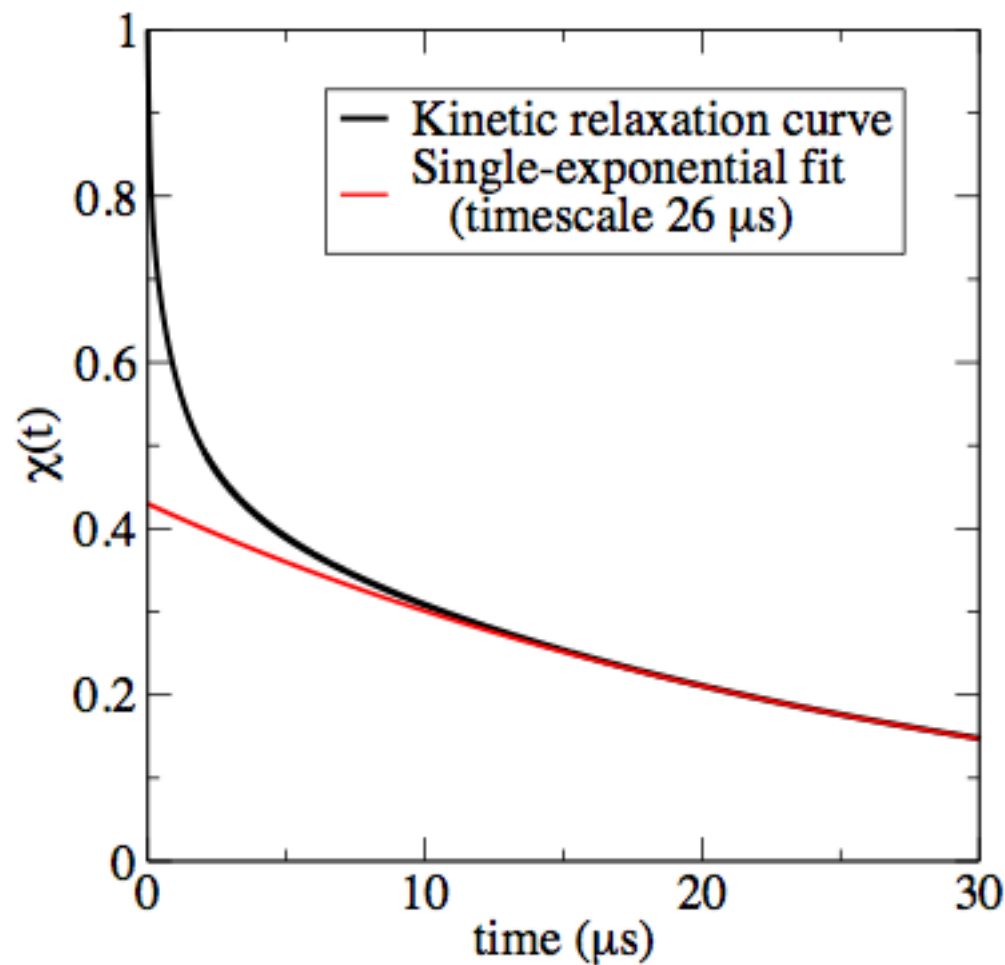
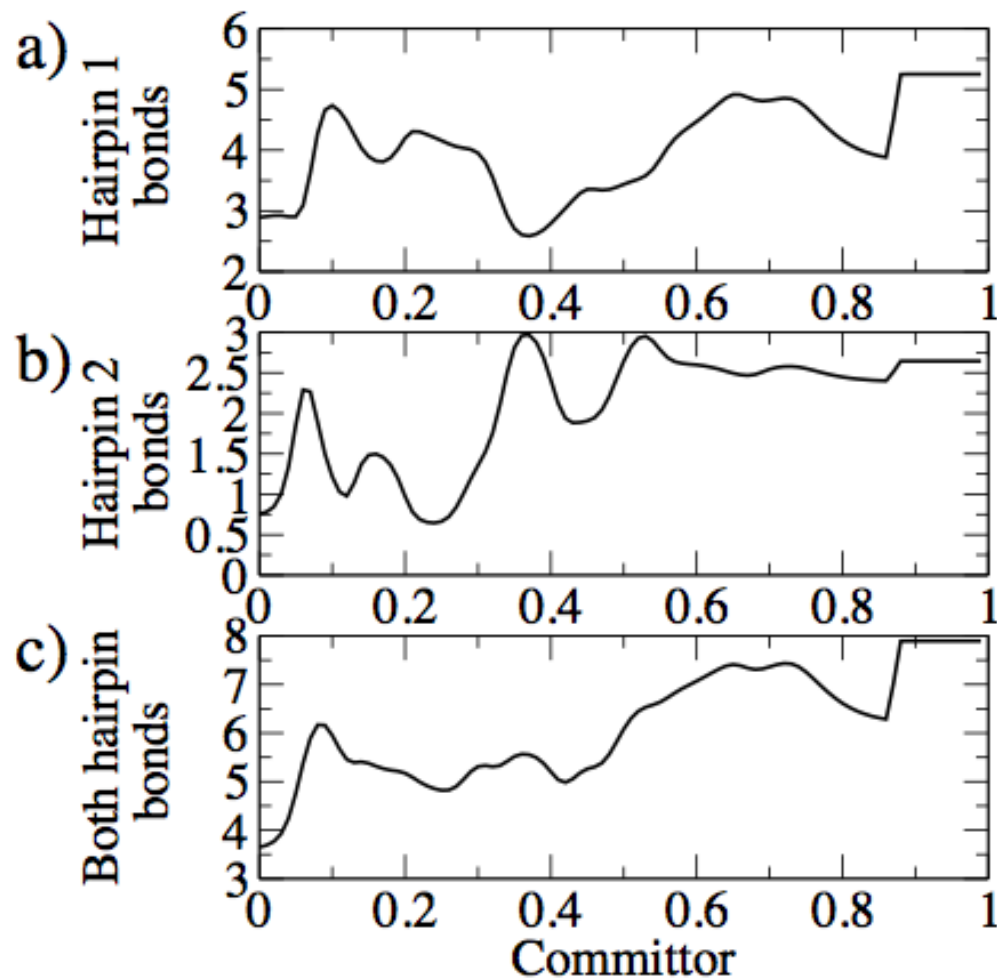


**Fig. 3.** PinWW folding flux. (Left) The network of the 70% most relevant folding pathways for PinWW. The numbers on the left indicate the committer probabilities, the thickness of the arrow indi-



Register-shifted trap states that do not carry significant folding flux but reduce the folding speed by nearly a factor of 2.

Number of native contacts is NOT the right reaction coordinate.



Experimental value  $13 \mu\text{s}$



**Basic idea:** evolve a curve rather than a point in your dynamical system.

$$\dot{x} = b(x) \quad \Rightarrow \quad \dot{\phi} = b(\phi) + \lambda \phi'$$

where  $b(x)$  is the velocity field, and the new term  $\lambda \phi'$  is added to control the parametrization of the curve:

$$\gamma = \{ \phi(s) : s \in [0, 1], |\phi'| = cst. \}$$

In practice, discretize the curve into  $N$  images  $\phi_i$ ,  $i = 1, \dots, N$ , and evolve them using a time-splitting algorithm:

1. Evolve every image *independently* using the original dynamics for a lag-time  $\Delta t$

$$\dot{\phi}_i = b(\phi_i) \quad i = 1, \dots, N$$

2. Interpolate a curve thru the images, and redistribute these images to maintain the desired parametrization, e.g.

$$|\phi_{i+1} - \phi_i| = cst. \quad i = 1, \dots, N - 1$$

Example: motion by steepest descent in potential energy with end-points free.

In this case, the string method identify the Minimum Energy Path (MEP).

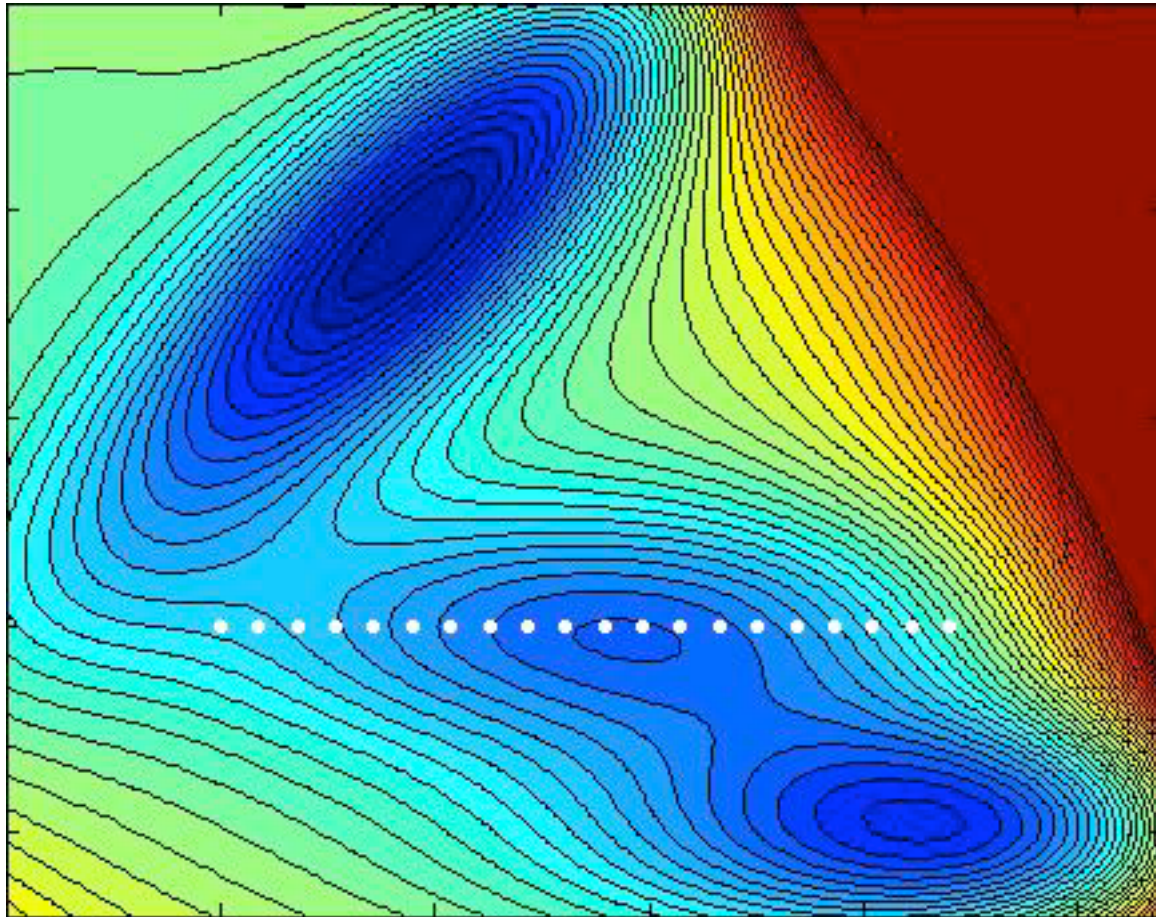
$$\dot{x} = -\nabla V(x)$$

*Similar to NEB but without projection of the force nor use of springs.*



Example: motion by steepest descent in potential energy with end-points free.

In this case, the string method identify the Minimum Energy Path (MEP).



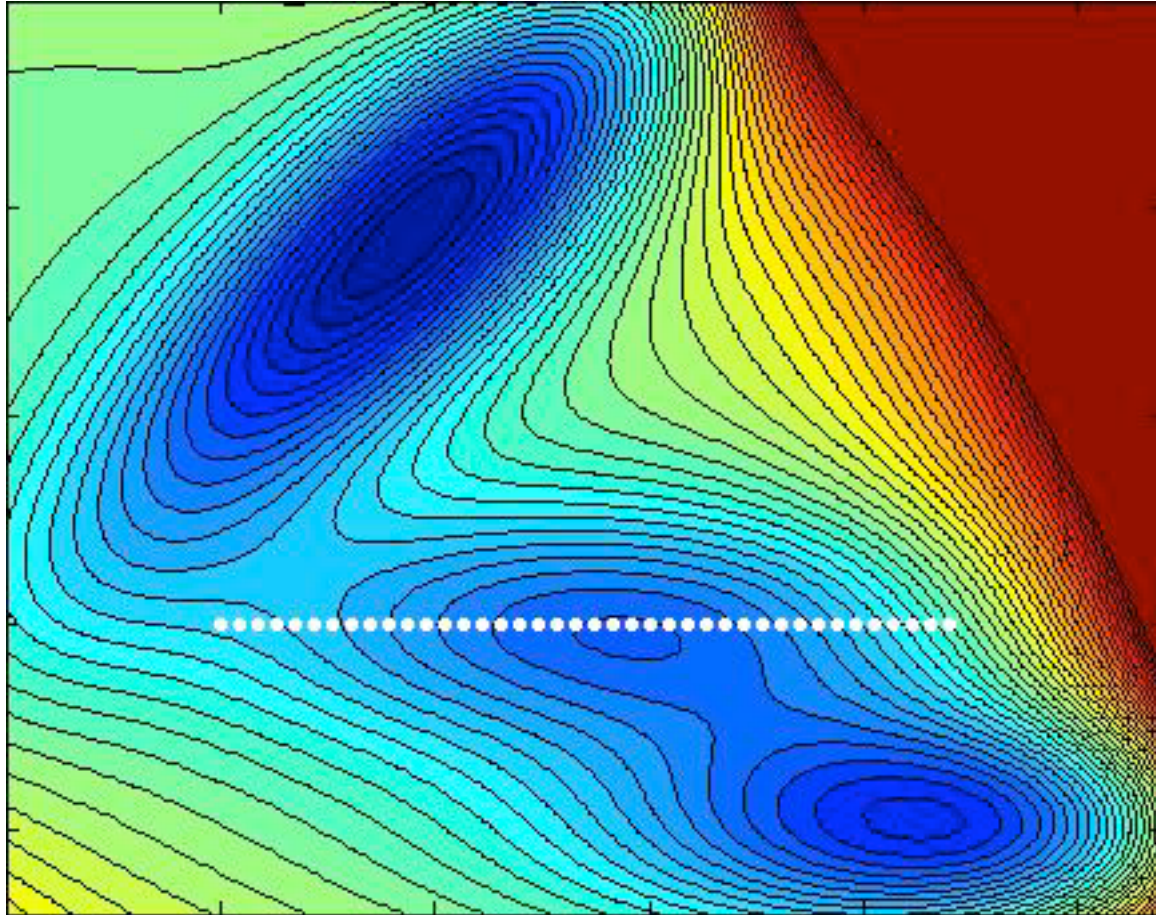
$$\dot{x} = -\nabla V(x)$$

*Similar to NEB but without projection of the force nor use of springs.*

Fairly robust e.g. to noise in the dynamics (more on this later)

$$\dot{x} = -\nabla V(x) + \sqrt{2k_B T} \eta$$

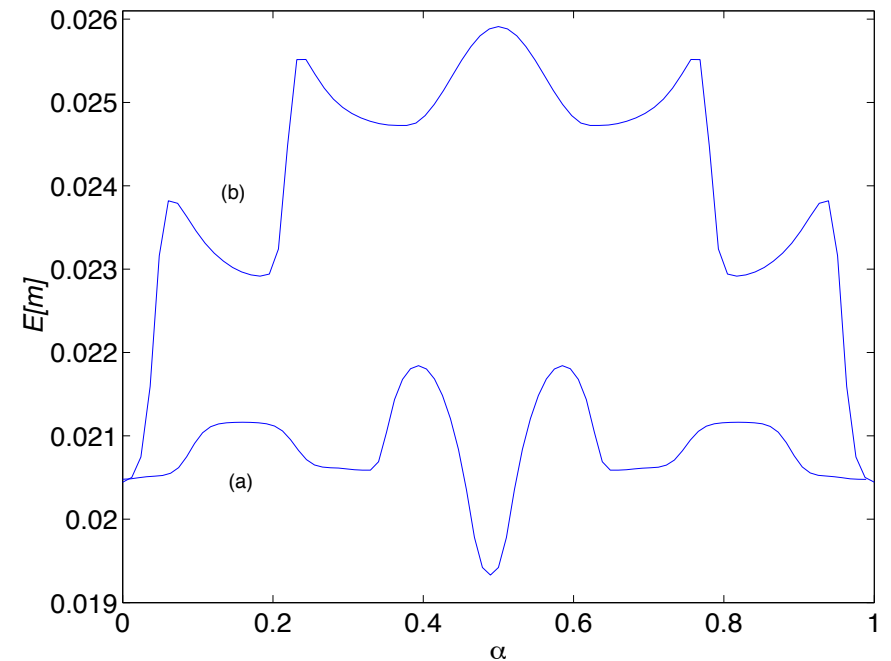
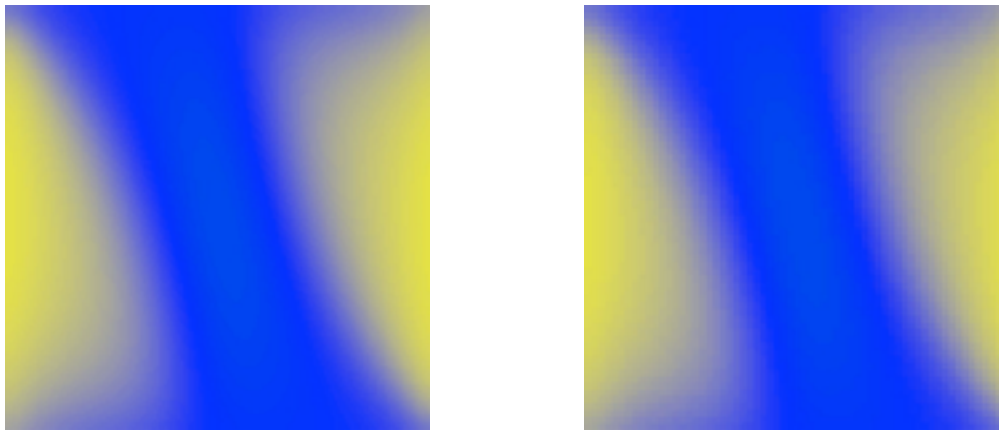
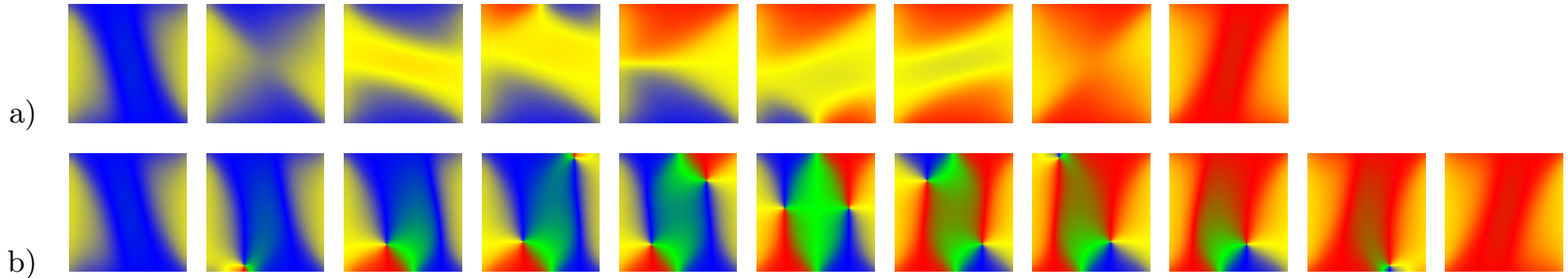
Fairly robust e.g. to noise in the dynamics (more on this later)



$$\dot{x} = -\nabla V(x) + \sqrt{2k_B T} \eta$$

When the velocity field  $b(x)$  is the gradient of a potential, the method identifies MEPs.

Example: magnetization reversal in thin sub-micron ferro magnetic elements



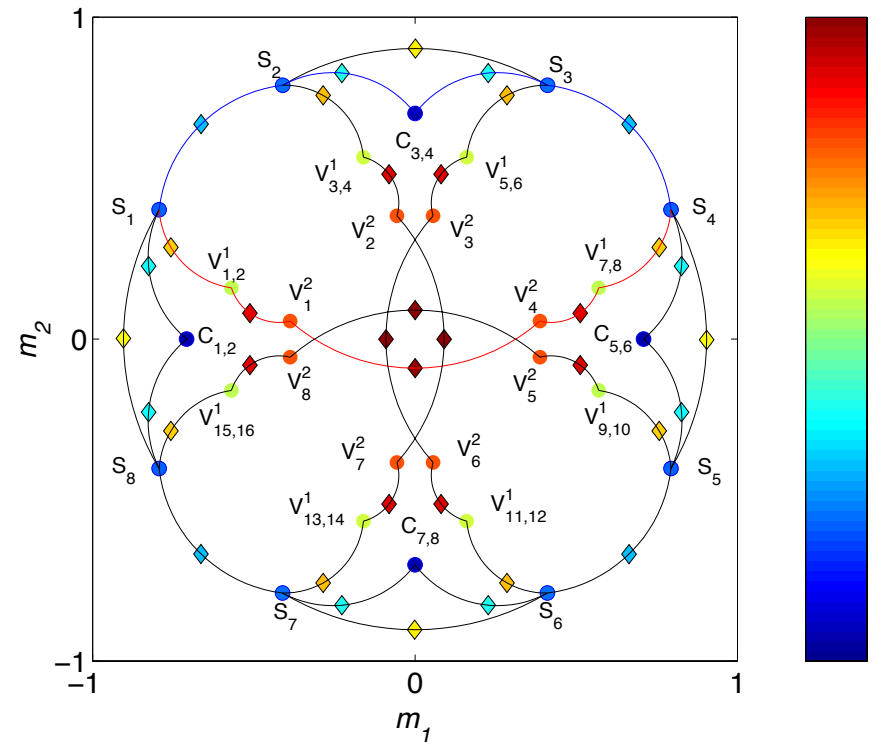
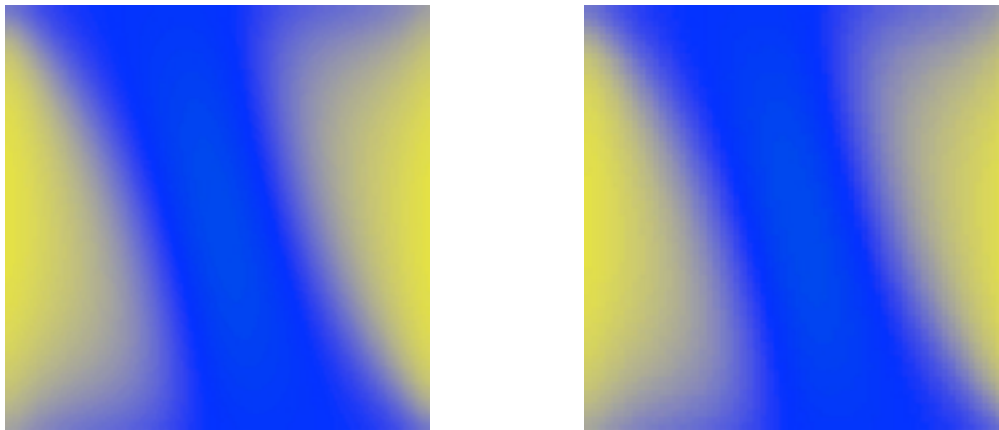
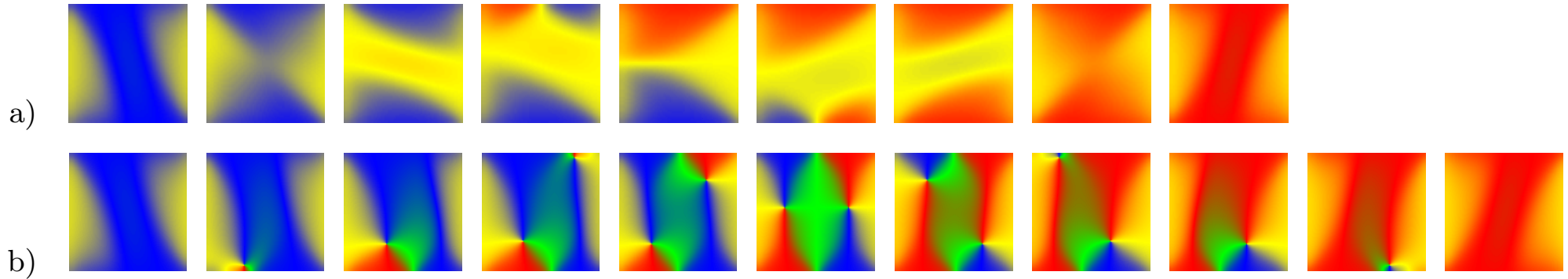
W. E. W. Ren & E. V.-E. *J. App. Phys.* **93**, 2275-2282 (2003);

G. D. Chaves-O'Flynn, D. Bedau, E. V.-E., A. D. Kent, and D. L. Stein, *IEEE Trans. Mag.* **46**, 2272-2274 (2010).



When the velocity field  $b(x)$  is the gradient of a potential, the method identifies MEPs.

Example: magnetization reversal in thin sub-micron ferro magnetic elements



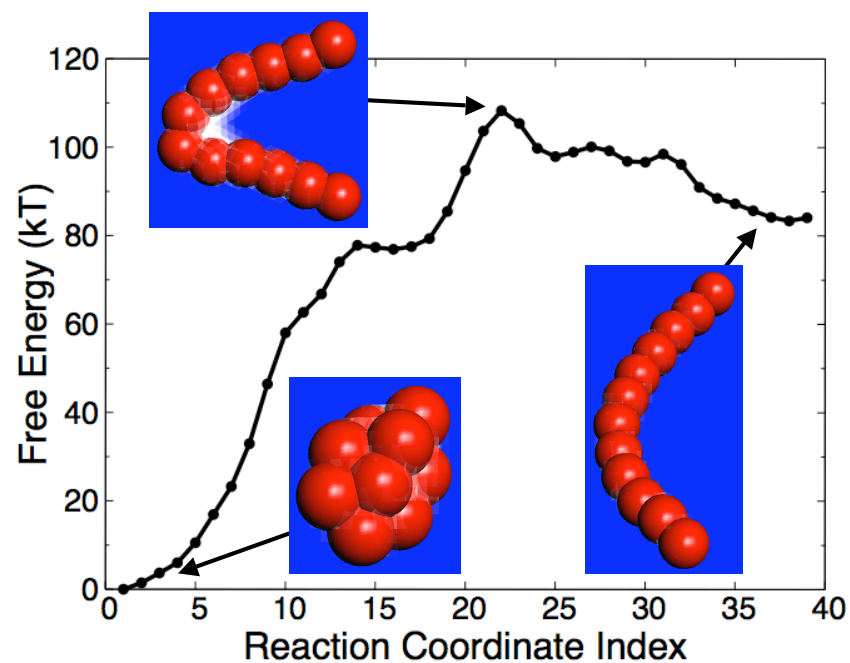
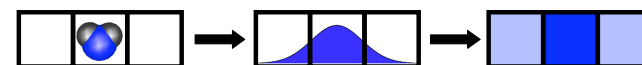
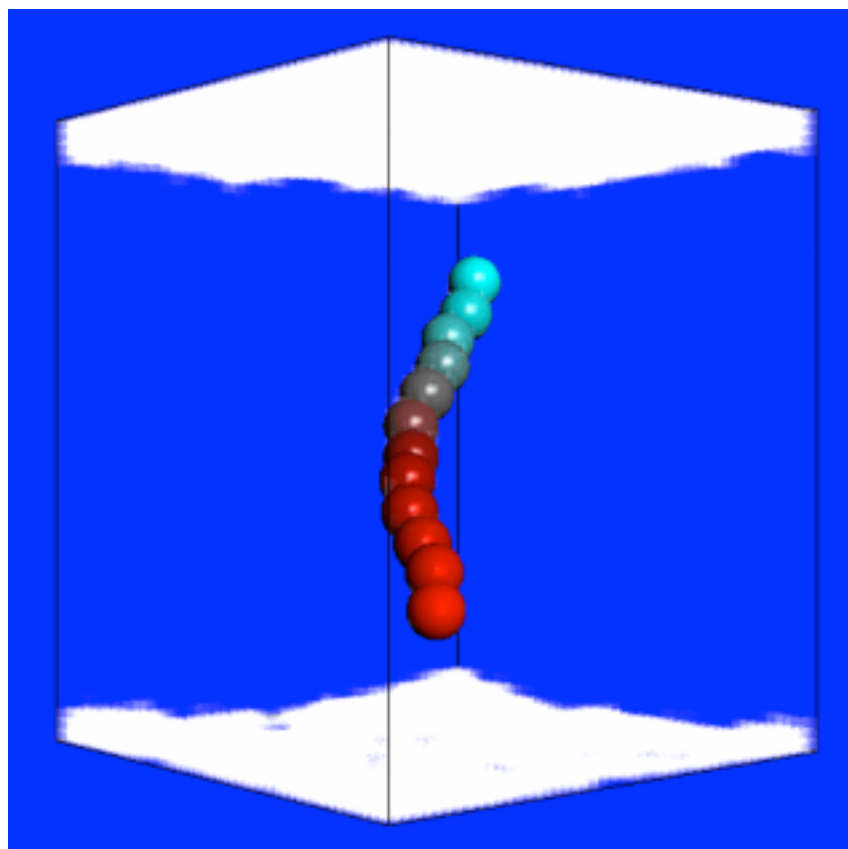
W. E. W. Ren & E. V.-E. *J. App. Phys.* **93**, 2275-2282 (2003);

G. D. Chaves-O'Flynn, D. Bedau, E. V.-E., A. D. Kent, and D. L. Stein, *IEEE Trans. Mag.* **46**, 2272-2274 (2010).

When the velocity field  $b(x)$  is the mean-force (i.e. the gradient of the free energy or potential of mean force associated with some collective variables), the method identifies Minimum Free Energy Paths - MFEPs.

*L. Maragliano, A. Fischer, E.V.-E., G. Ciccotti, J. Chem. Phys. **125** 024106 (2006);  
L. Maragliano, E.V.-E., Chem. Phys. Lett. **446** 182-190 (2007);*

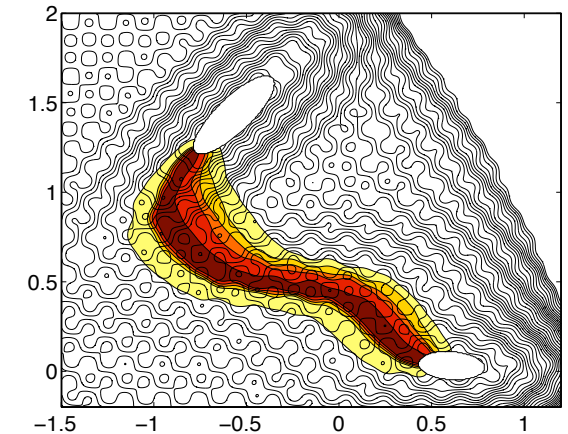
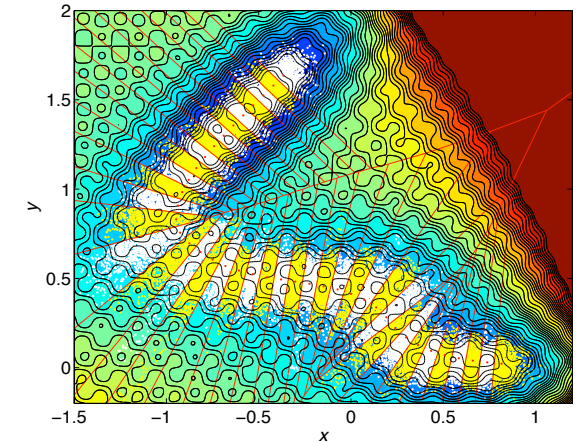
Example: hydrophobic collapse of polymeric chain



*T.F. Miller III, E.V.-E. & D. Chandler, Proc. Nat. Acad. Sci. USA **104**, 14459-14464 (2007)*

Other generalization: **Finite-temperature string method** to identify principal curves (i.e. the curve that is its own expectation)

W. E. W. Ren & E. V.-E. *J. Phys. Chem. B* **109**, 6688-6693 (2005)  
E. V.-E. & M. Venturoli, *J. Chem. Phys.* **130**, 194103 (2009)

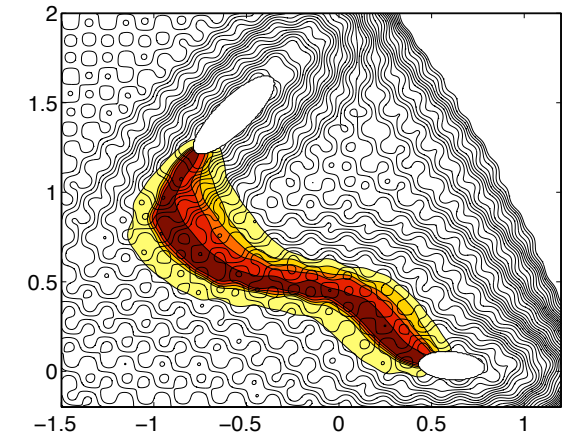
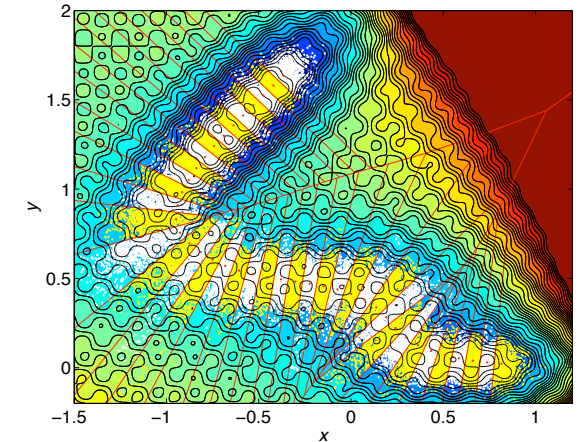
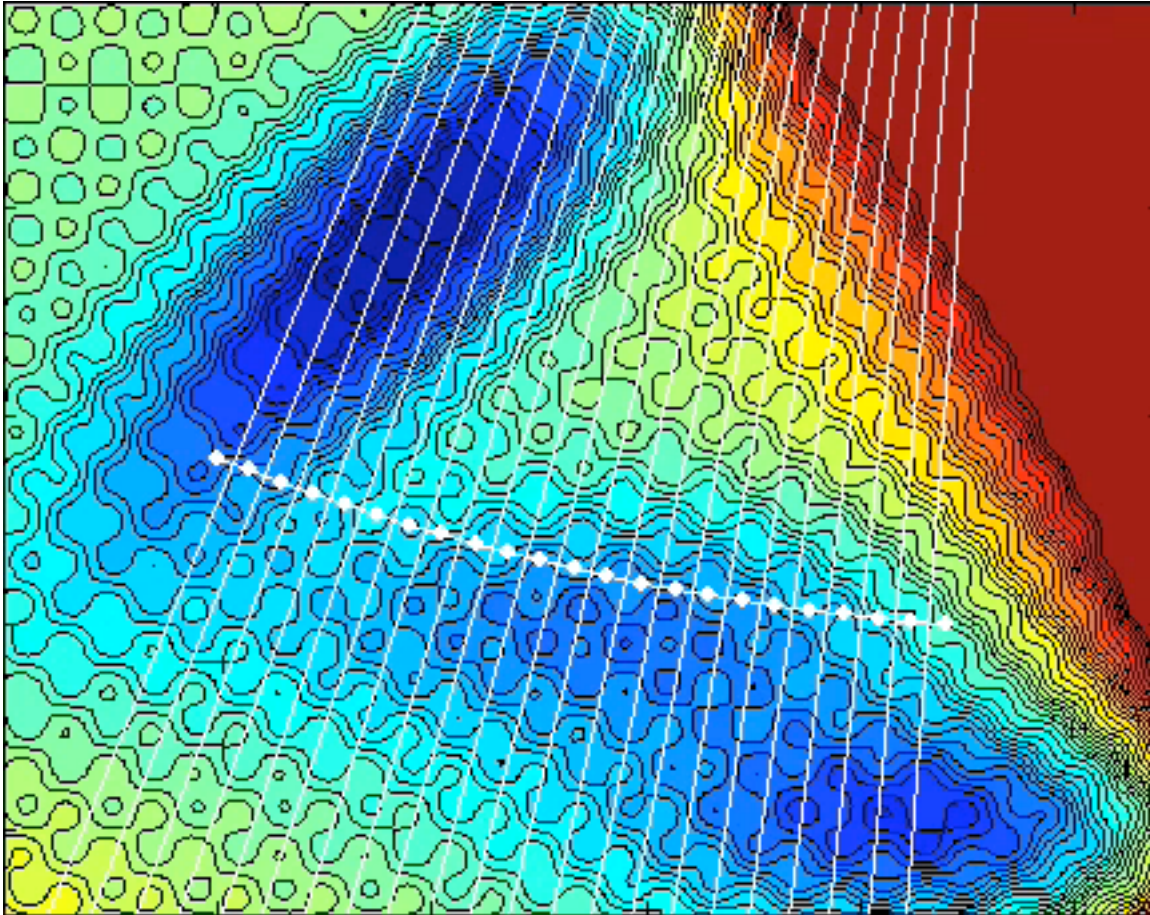


As 'velocity field', use a measure of the discrepancy between the image position and the expected value of the equilibrium distribution in the Voronoi cell associated with this image - that is, drag the former towards the latter.



Other generalization: **Finite-temperature string method** to identify principal curves (i.e. the curve that is its own expectation)

W. E. W. Ren & E. V.-E. *J. Phys. Chem. B* **109**, 6688-6693 (2005)  
E. V.-E. & M. Venturoli, *J. Chem. Phys.* **130**, 194103 (2009)



As 'velocity field', use a measure of the discrepancy between the image position and the expected value of the equilibrium distribution in the Voronoi cell associated with this image - that is, drag the former towards the latter.



## What if we play with the equation for the end points?

For example, in the direction parallel to the string, reverse the velocity (e.g. the gradient of the potential) of one end-point - this makes climb towards saddle point!

$$\dot{\phi} = -\nabla V(\phi) \quad \Rightarrow \quad \dot{\phi} = -\nabla V(\phi) + 2[\nabla V(\phi) \cdot \hat{\tau}] \hat{\tau}$$

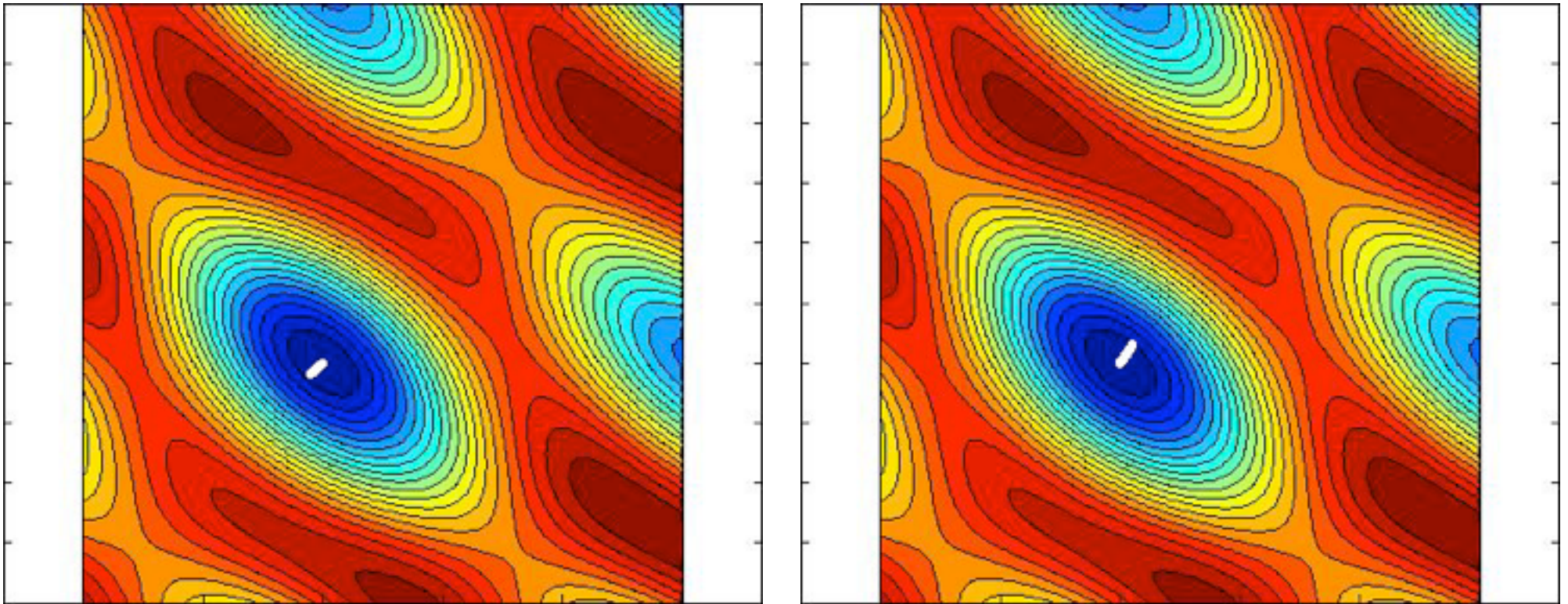
Simple alternative to dimmer method, ART, gentle ascent dynamics, etc.

Give better control: one always knows if the string is still in the right basin, and what the energy barrier is along it.

## What if we play with the equation for the end points?

For example, in the direction parallel to the string, reverse the velocity (e.g. the gradient of the potential) of one end-point - this makes climb towards saddle point!

$$\dot{\phi} = -\nabla V(\phi) \quad \Rightarrow \quad \dot{\phi} = -\nabla V(\phi) + 2[\nabla V(\phi) \cdot \hat{\tau}] \hat{\tau}$$



Simple alternative to dimmer method, ART, gentle ascent dynamics, etc.

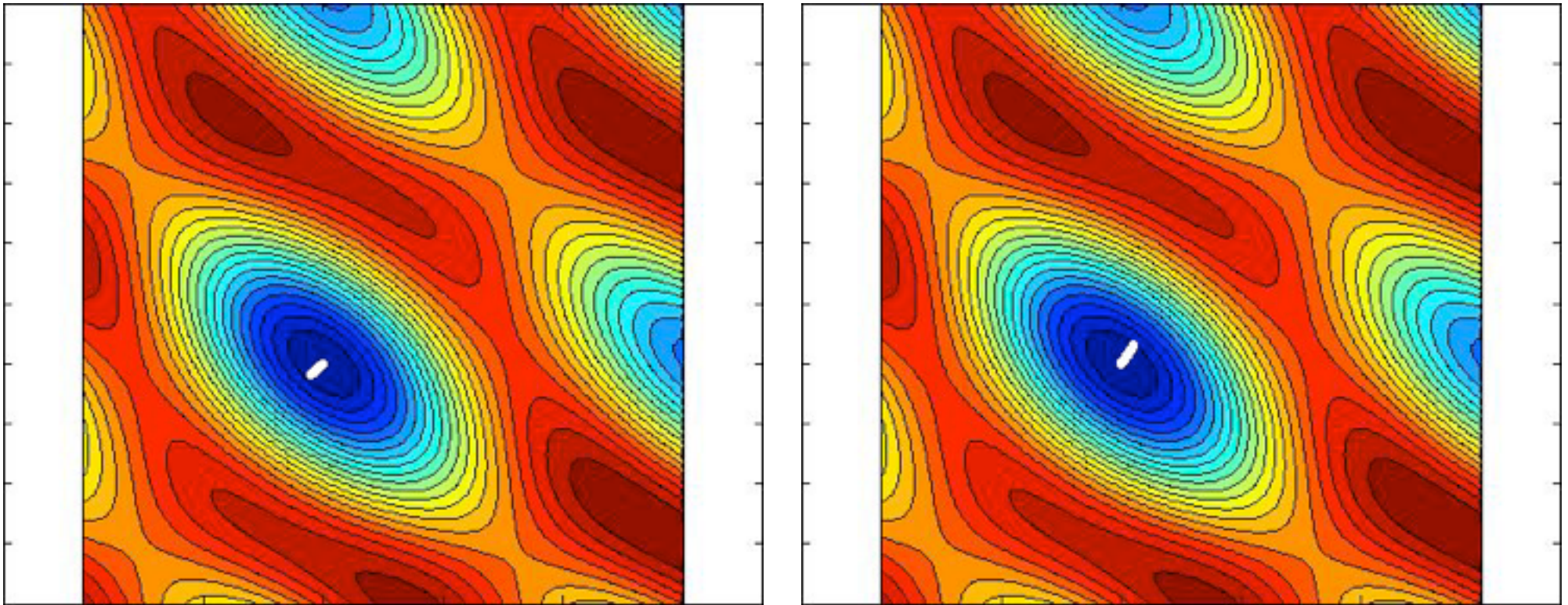
Give better control: one always knows if the string is still in the right basin, and what the energy barrier is along it.



## What if we play with the equation for the end points?

For example, in the direction parallel to the string, reverse the velocity (e.g. the gradient of the potential) of one end-point - this makes climb towards saddle point!

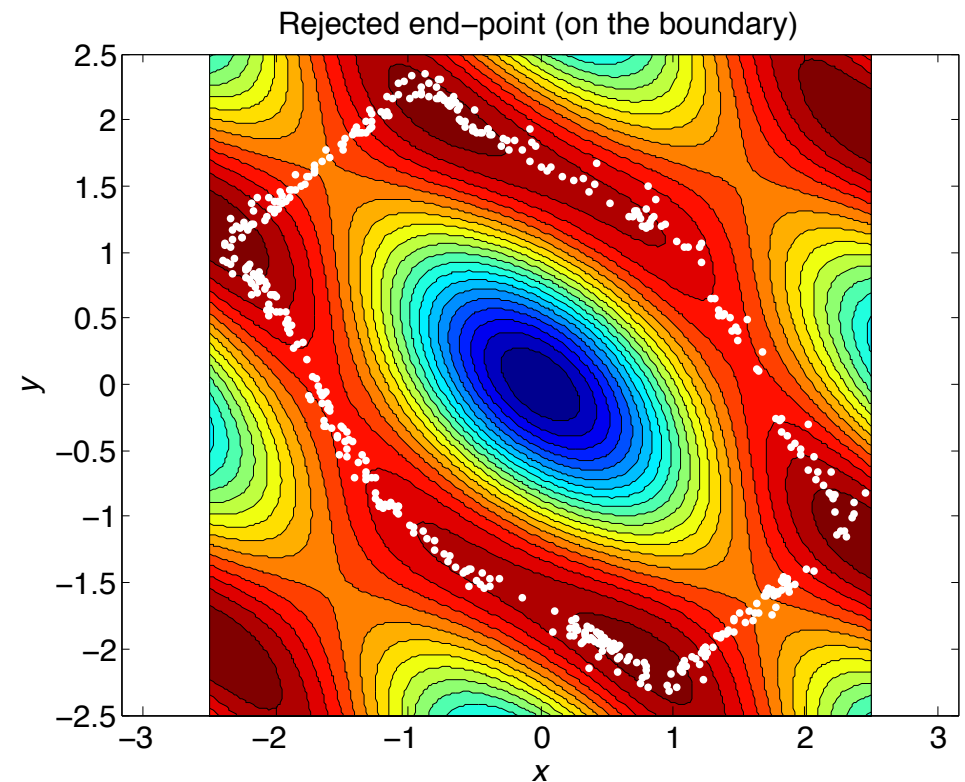
$$\dot{\phi} = -\nabla V(\phi) \quad \Rightarrow \quad \dot{\phi} = -\nabla V(\phi) + 2[\nabla V(\phi) \cdot \hat{\tau}] \hat{\tau}$$



Simple alternative to dimmer method, ART, gentle ascent dynamics, etc.

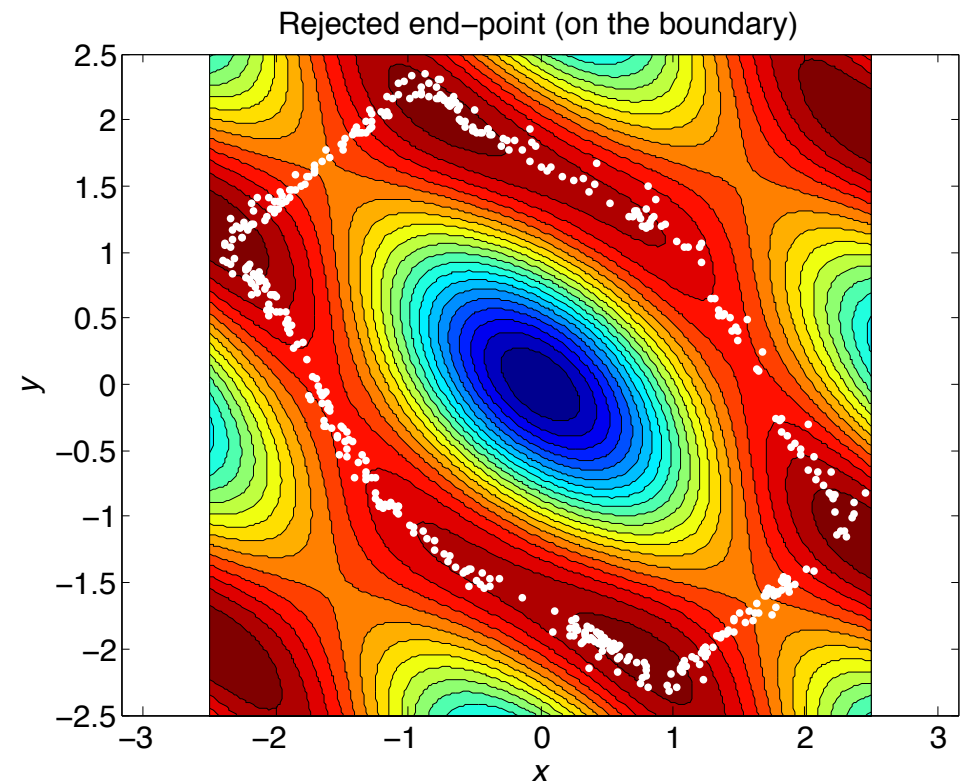
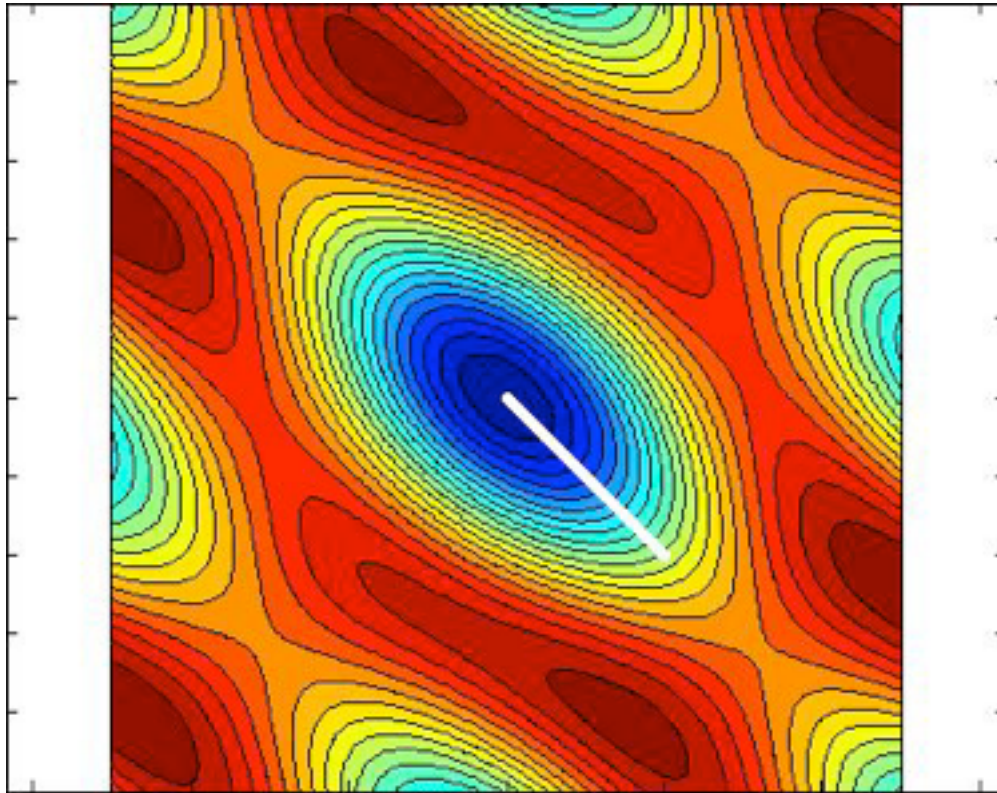
Give better control: one always knows if the string is still in the right basin, and what the energy barrier is along it.

Another idea: make the end-point perform a (biased) random walk, but reject any move such that the energy along the string is not monotonic - this keeps the string in the basin, and permits to explore it without leaving it!



*Method to find all the saddle points around a minimum.*

Another idea: make the end-point perform a (biased) random walk, but reject any move such that the energy along the string is not monotonic - this keeps the string in the basin, and permits to explore it without leaving it!



*Method to find all the saddle points around a minimum.*



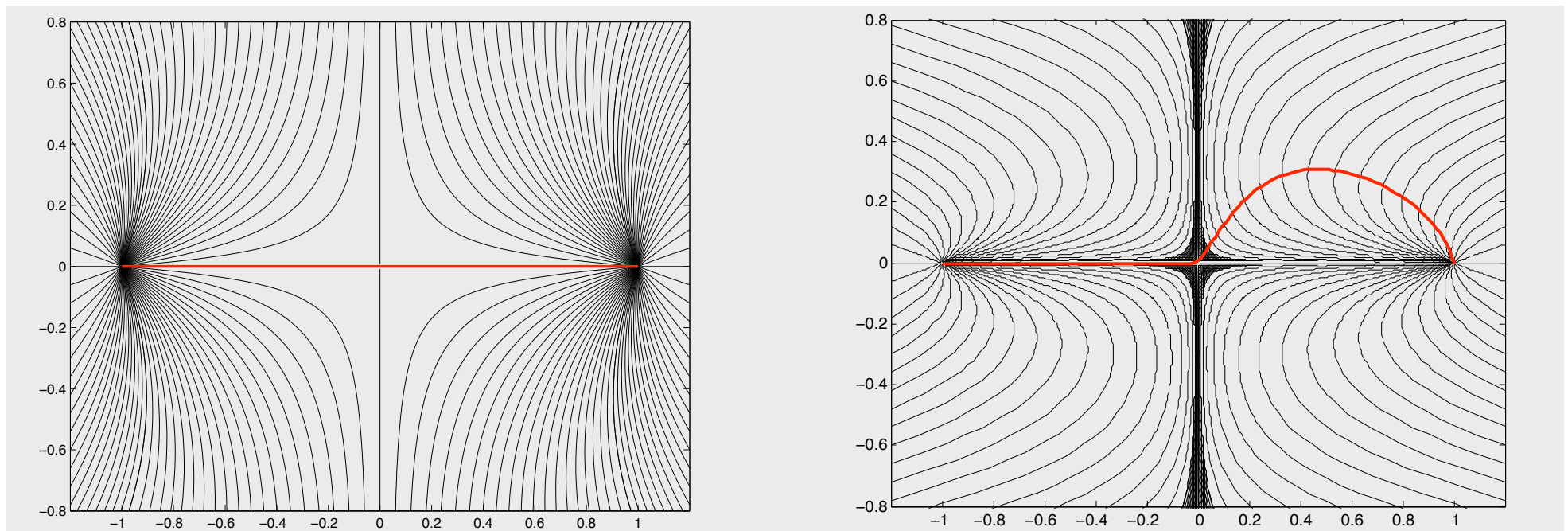
## What about non-gradient systems (i.e. activated processes arising out-of-equilibrium)?

Their pathway can be identified too in the low noise limit using the framework of Large Deviation Theory (LDT) as the Maximum Likelihood Path (MLP) which minimizes the LDT action.

Use the Minimum Action Method to identify this MLP.

W. E, W. Ren & E.V.-E., *Comm. Pure App. Math* **52**, 637-656 (2004);  
M. Heymann & E.V.-E., *Comm. Pure App. Math* **61**, 1052-1117 (2008).

Simple illustrative example due to Maier and Stein:

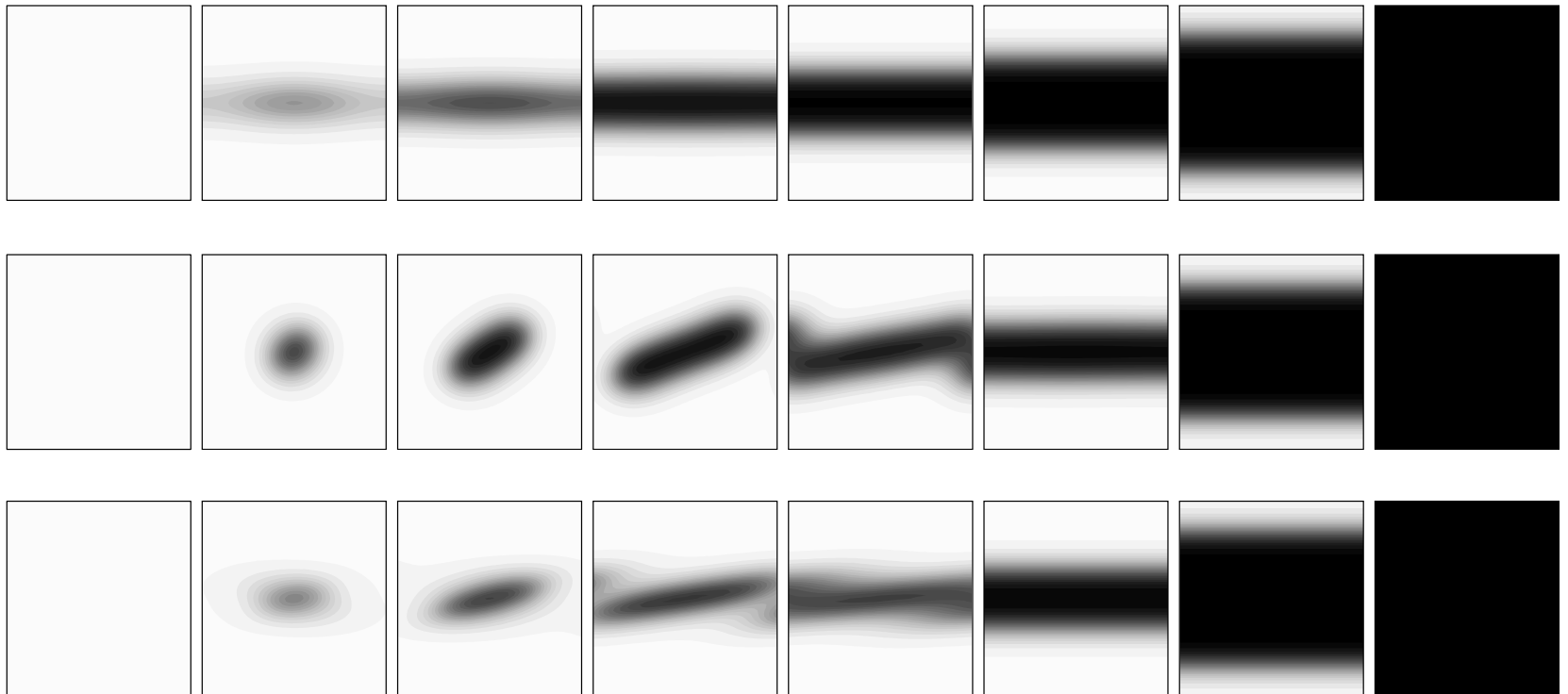


# More sophisticated example: phase-transition in the presence of a shear flow

*M. Heymann & E.V.-E. Phys. Rev. Lett. **100**, 140601 (2008)*

$$\dot{u} = \underbrace{\kappa\Delta u + u - u^3}_{-DE(u)} + c \sin(y) \partial_x u + \eta$$

$$E(u) = \int_{\Omega} \left( \frac{1}{2} \kappa |\nabla u|^2 + \frac{1}{4} (1 - u^2)^2 \right) dx$$

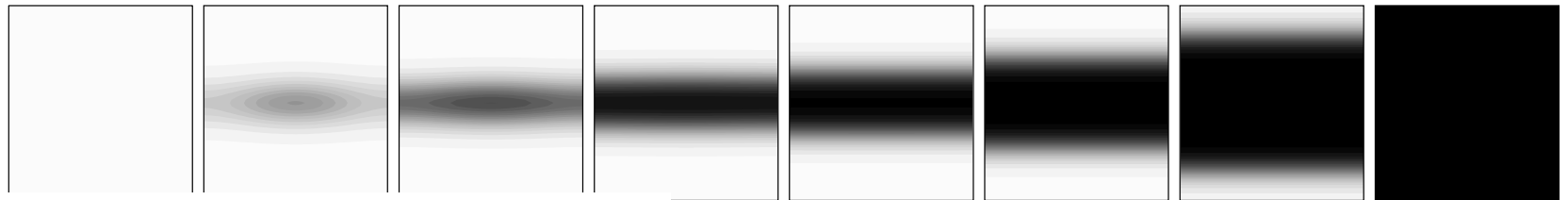


# More sophisticated example: phase-transition in the presence of a shear flow

M. Heymann & E.V.-E. *Phys. Rev. Lett.* **100**, 140601 (2008)

$$\dot{u} = \underbrace{\kappa \Delta u + u - u^3}_{-DE(u)} + c \sin(y) \partial_x u + \eta$$

$$E(u) = \int_{\Omega} \left( \frac{1}{2} \kappa |\nabla u|^2 + \frac{1}{4} (1 - u^2)^2 \right) dx$$



minimal action

