Monte Carlo Methods for Rough Energy Landscapes

Jon Machta
University of Massachusetts Amherst,
Santa Fe Institute

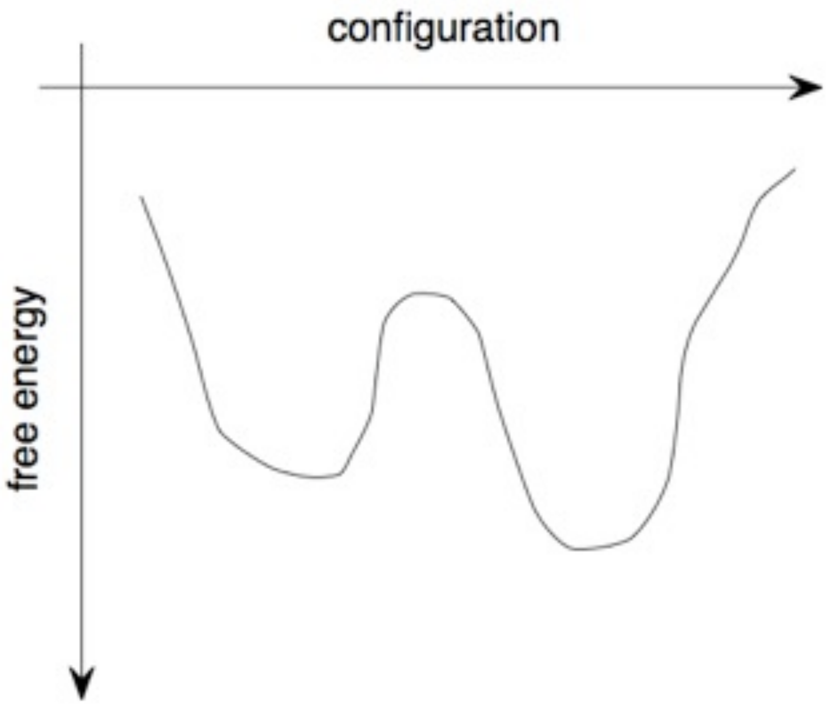
SFI, August 9, 2013

Collaborators

- Stephan Burkhardt
- Chris Pond
- Jingran Li
- Richard Ellis

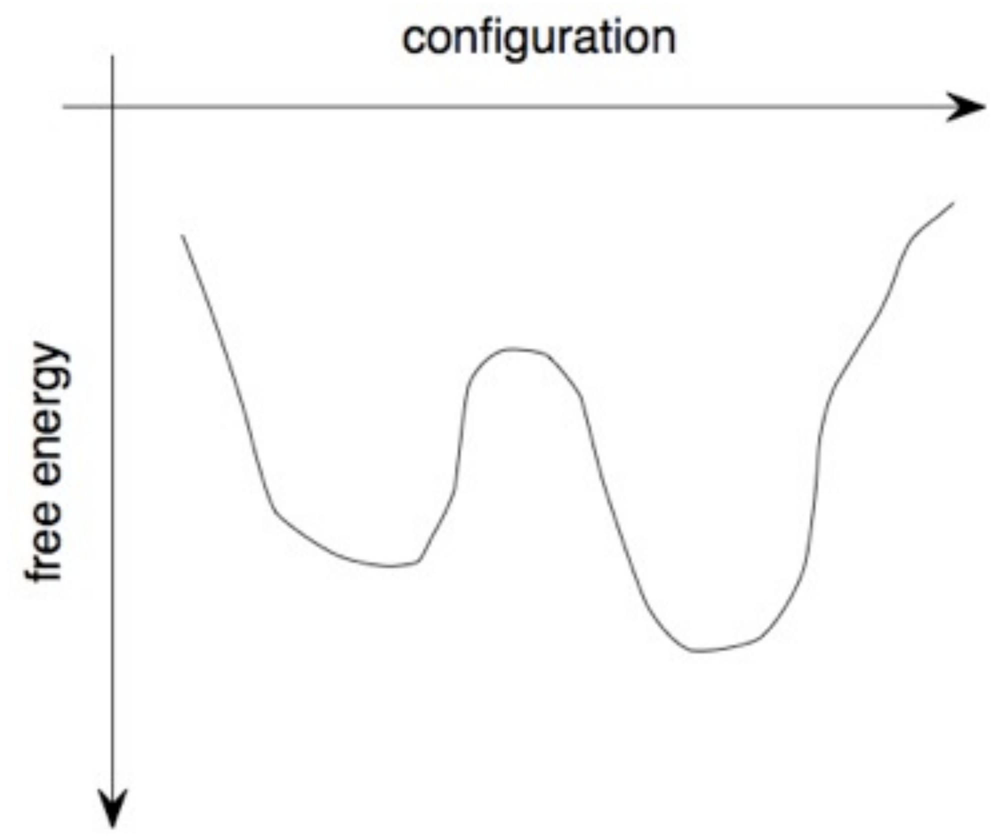
Motivation

- How to sample equilibrium states of systems with rugged free energy landscapes, e.g. spin glasses, configurational glasses, proteins, combinatorial optimization problems.



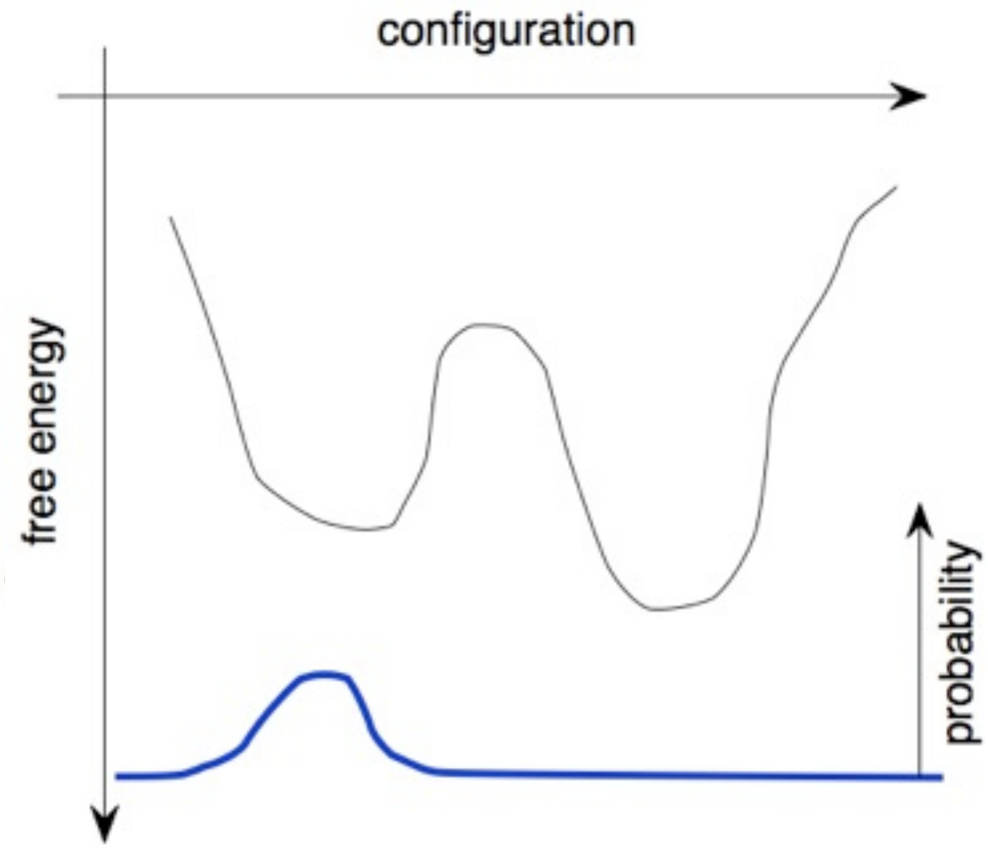
Problem

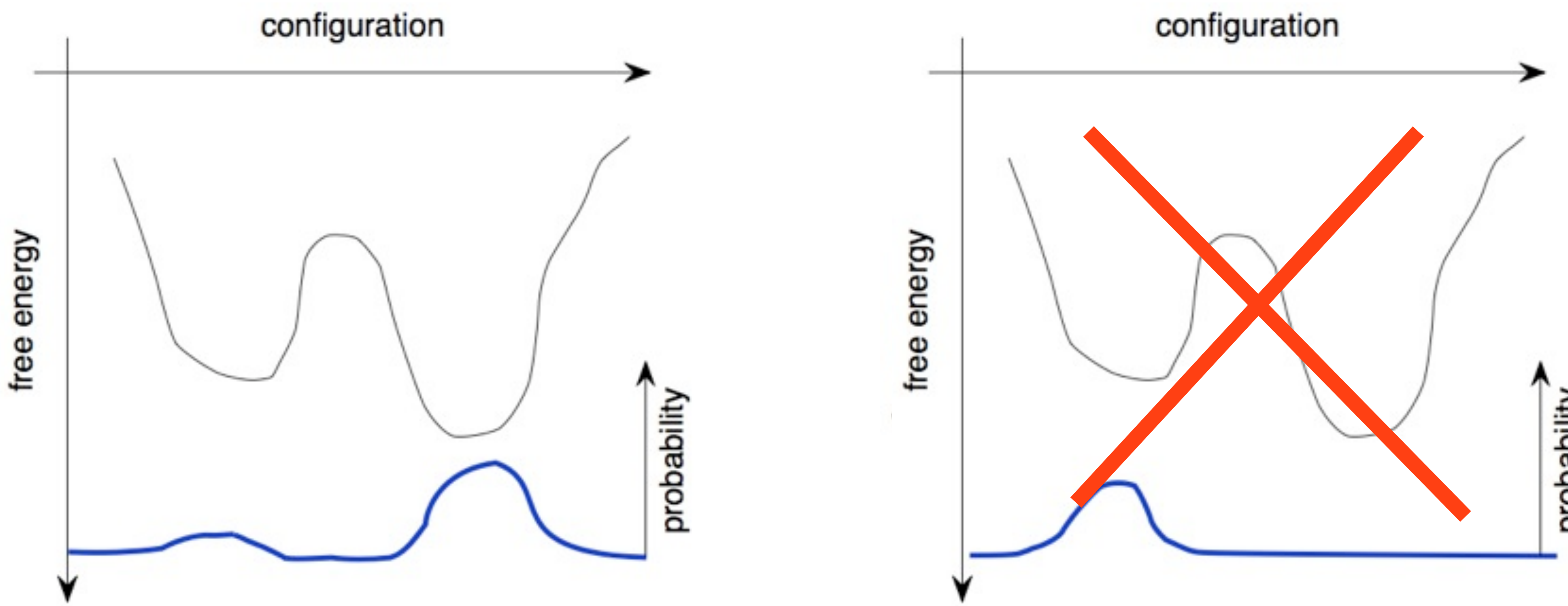
- Markov chain Monte Carlo (MCMC) at a single temperature such as the Metropolis algorithm gets stuck in local minima.



Problem

- Markov chain Monte Carlo (MCMC) at a single temperature such as the Metropolis algorithm gets stuck in local minima.




configuration
free energy
probability
configuration
free energy
probability

Outline

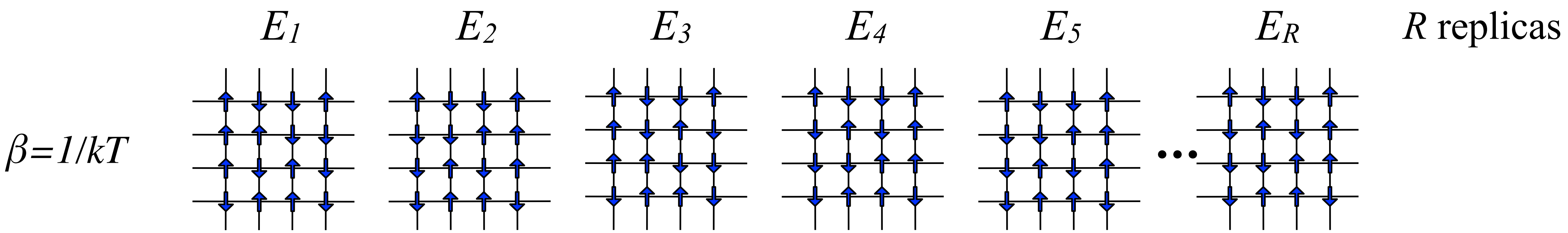
- **Population Annealing**
- Parallel Tempering
- PA vs PT
- Conclusions

Population Annealing

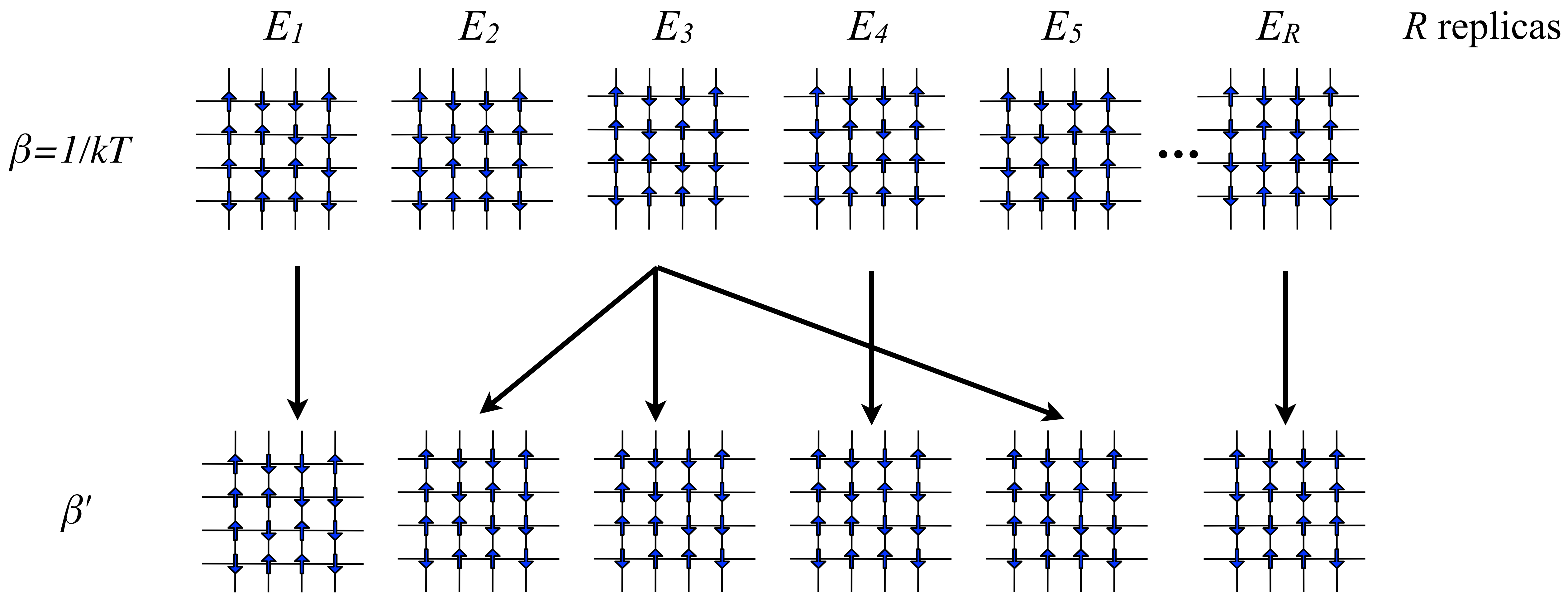
K. Hukushima and Y. Iba, in *THE MONTE CARLO METHOD IN THE PHYSICAL SCIENCES: Celebrating the 50th Anniversary of the Metropolis Algorithm*, edited by J. E. Gubernatis (AIP, 2003), vol. 690, pp. 200–206.

- Modification of simulated annealing for equilibrium sampling.
- A population of replicas is cooled according to an annealing schedule. Each replica is acted on by a MCMC (e.g. Metropolis) at the current temperature.
- During each temperature step, the population is differentially resampled according to Boltzmann weights to maintain equilibrium.

Population Annealing

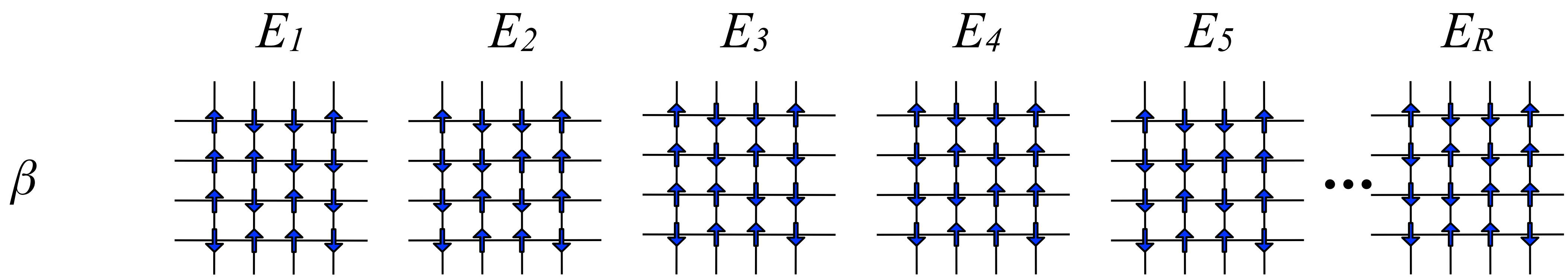


Population Annealing



Population annealing = simulated annealing with differential reproduction (resampling) of replicas

Temperature Step

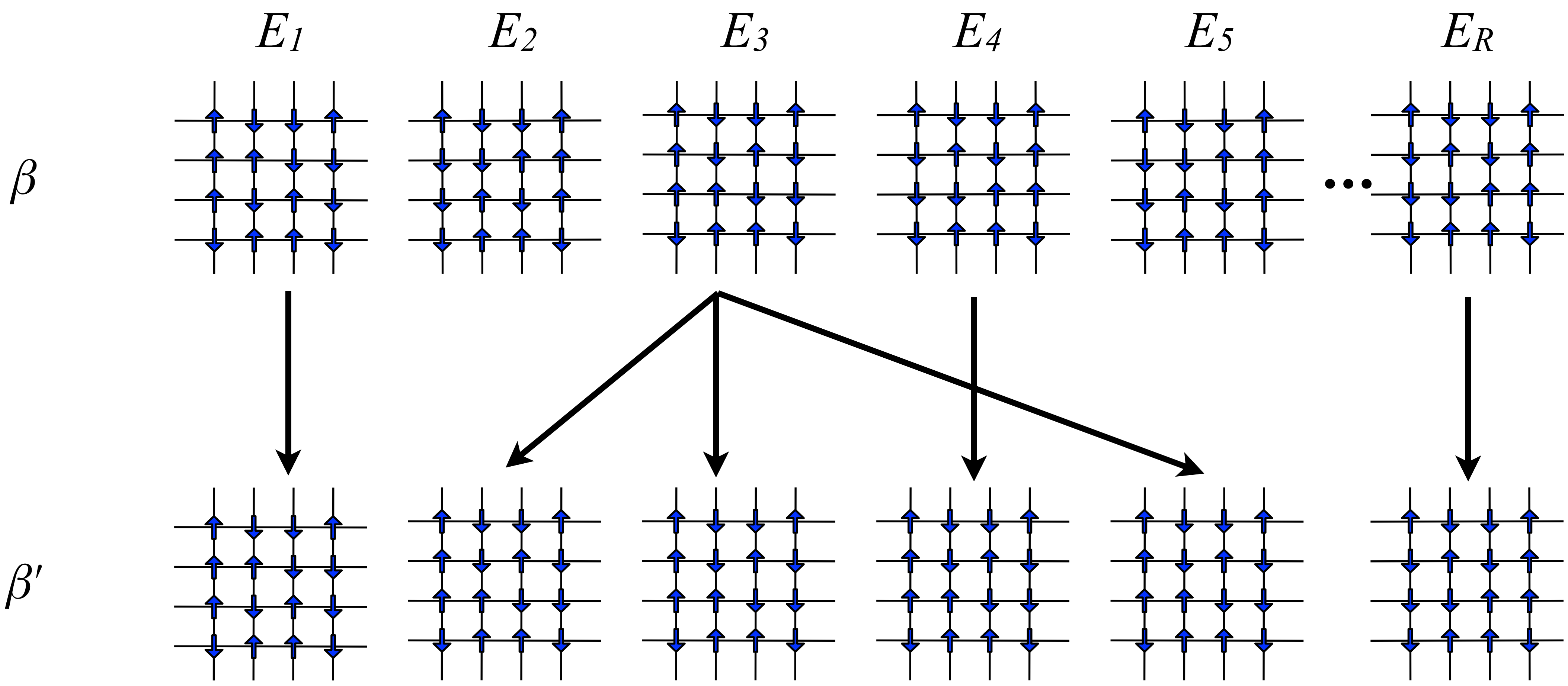


$$\tau_j(\beta,\beta') = \frac{\exp\left[-(\beta'-\beta)E_j\right]}{Q(\beta,\beta')}$$

$$Q(\beta,\beta') = \frac{\sum_{j=1}^{R}\exp\left[-(\beta'-\beta)E_j\right]}{R}$$

Replica j is reproduced n_j times where n_j is an integer random variate with mean τ_j.

Temperature Step

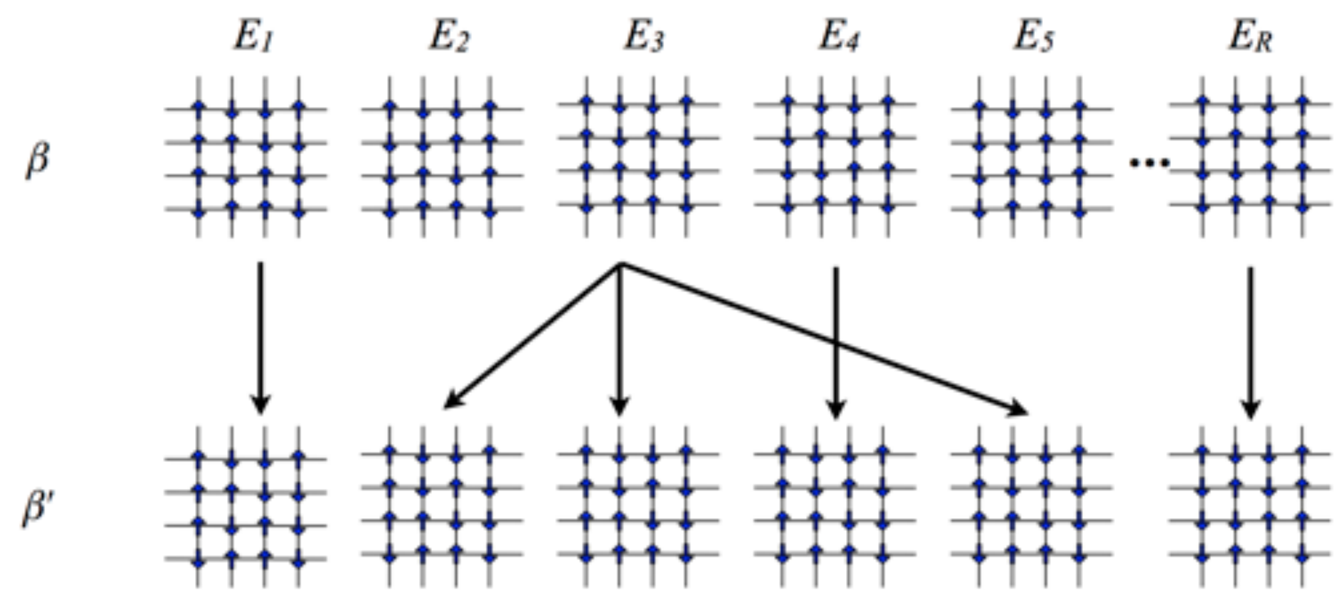


$$\tau_j(\beta, \beta') = \frac{\exp\left[-(\beta' - \beta)E_j\right]}{Q(\beta, \beta')}$$

$$Q(\beta, \beta') = \frac{\sum_{j=1}^{R} \exp\left[-(\beta' - \beta)E_j\right]}{R}$$

Replica j is reproduced n_j times where n_j is an integer random variate with mean τ_j.

Resampling



e.g. $n_1 = 1, \; n_2 = 0, \; n_3 = 3, \; n_4 = 1$

$$\tau_j(\beta, \beta') = \frac{\exp\left[-(\beta' - \beta)E_j\right]}{Q(\beta, \beta')}$$

$$Q(\beta, \beta') = \frac{\sum_{j=1}^{R} \exp\left[-(\beta' - \beta)E_j\right]}{R}$$

Replica j is reproduced n_j times where n_j is an integer random variate with mean τ_j.

Nearest integer resampling

$$\mathrm{Prob}(n = \lfloor \tau \rfloor) = \tau - \lceil \tau \rceil$$

$$\mathrm{Prob}(n = \lceil \tau \rceil) = \tau - \lfloor \tau \rfloor$$

$$\mathrm{Prob}(n - \tau > 1) = 0$$

Minimizes variance of n_j.
Maximizes number of ancestors with descendants.
Population size fluctuates.

Population Annealing is related to...

- Simulated annealing
- Sequential Monte Carlo
  - See e.g. "Sequential Monte Carlo Methods in Practice", A. Doucet, et. al. (2001)
  - aka Particle Filters
  - Nested Sampling, Skilling
- Go with the Winners, Grassberger (2002)
- Diffusion (quantum) Monte Carlo
- Nonequilibrium Equality for Free Energy Differences, Jarzynski (1997)
- Histogram Re-weighting, Swendsen and Ferrenberg (1988)

Convergence to Equilibrium

Suppose the population at β is a correct, iid sample:

$$Y_i = \exp\left[-(\beta' - \beta)E_i + \lambda A_i\right]$$

JM, R. Ellis, JSP 144, 544 (2011)

Convergence to Equilibrium

Suppose the population at β is a correct, iid sample:

$$Y_i = \exp\left[-(\beta' - \beta)E_i + \lambda A_i\right]$$

$$I(\beta, \beta', \lambda) = \mathbf{E}\log(\frac{1}{R}\sum_{j=1}^{R} Y_j)$$

JM, R. Ellis, JSP 144, 544 (2011)

Convergence to Equilibrium

Suppose the population at β is a correct, iid sample:

$$Y_i = \exp\left[-(\beta' - \beta)E_i + \lambda A_i\right]$$

$$I(\beta, \beta', \lambda) = \mathbf{E} \log\left(\frac{1}{R}\sum_{j=1}^{R} Y_j\right)$$

$$\langle A \rangle'_R = \left.\frac{dI(\beta, \beta', \lambda)}{d\lambda}\right|_{\lambda=0}$$

JM, R. Ellis, JSP 144, 544 (2011)

Convergence to Equilibrium

Suppose the population at β is a correct, iid sample:

$$Y_i = \exp\left[-(\beta' - \beta)E_i + \lambda A_i\right]$$

$$I(\beta, \beta', \lambda) = \mathbf{E} \log\left(\frac{1}{R} \sum_{j=1}^{R} Y_j\right) = \log(\mathbf{E}Y) - \frac{1}{2R} \frac{\mathbf{Var}Y}{(\mathbf{E}Y)^2} + \mathrm{O}\left(\frac{1}{R^{3/2}}\right)$$

$$\langle A \rangle'_R = \left.\frac{dI(\beta, \beta', \lambda)}{d\lambda}\right|_{\lambda=0}$$

JM, R. Ellis, JSP 144, 544 (2011)

Convergence to Equilibrium

Suppose the population at β is a correct, iid sample:

$$Y_i = \exp\left[-(\beta' - \beta)E_i + \lambda A_i\right]$$

$$I(\beta, \beta', \lambda) = \mathbf{E} \log(\frac{1}{R}\sum_{j=1}^{R} Y_j) = \log(\mathbf{E}Y) - \frac{1}{2R}\frac{\mathbf{Var}Y}{(\mathbf{E}Y)^2} + \mathrm{O}\left(\frac{1}{R^{3/2}}\right)$$

$$\langle A \rangle'_R = \left.\frac{dI(\beta, \beta', \lambda)}{d\lambda}\right|_{\lambda=0} = \mathbf{E}' A + \mathrm{O}\left(\frac{1}{R}\right)$$

JM, R. Ellis, JSP 144, 544 (2011)

Convergence to Equilibrium

Suppose the population at β is a correct, iid sample:

$$Y_i = \exp\left[-(\beta' - \beta)E_i + \lambda A_i\right]$$

$$I(\beta, \beta', \lambda) = \mathbf{E}\log(\frac{1}{R}\sum_{j=1}^{R} Y_j) = \log(\mathbf{E}Y) - \frac{1}{2R}\frac{\mathbf{Var}Y}{(\mathbf{E}Y)^2} + \mathrm{O}\left(\frac{1}{R^{3/2}}\right)$$

$$\langle A\rangle'_R = \left.\frac{dI(\beta, \beta', \lambda)}{d\lambda}\right|_{\lambda=0} = \mathbf{E}'A + \mathrm{O}\left(\frac{1}{R}\right)$$

A bias inversely proportional to the population size results from the re-weighting to β'

JM, R. Ellis, JSP 144, 544 (2011)

Convergence to Equilibrium

Resampling complicates the analysis...

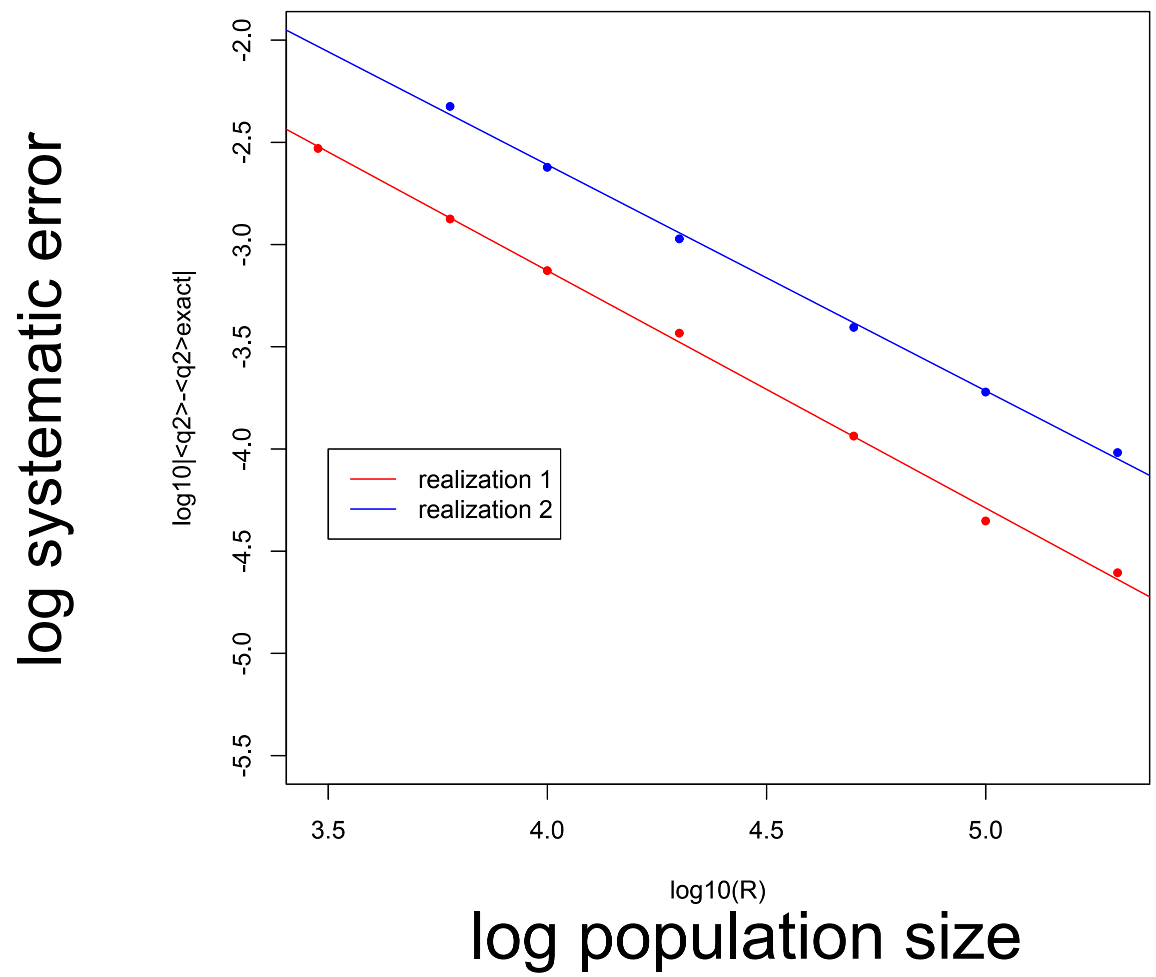
Conjecture: Fixing other parameters of population annealing, observable converge to their equilibrium values inversely in population size.

Quantifying Convergence to Equilibrium

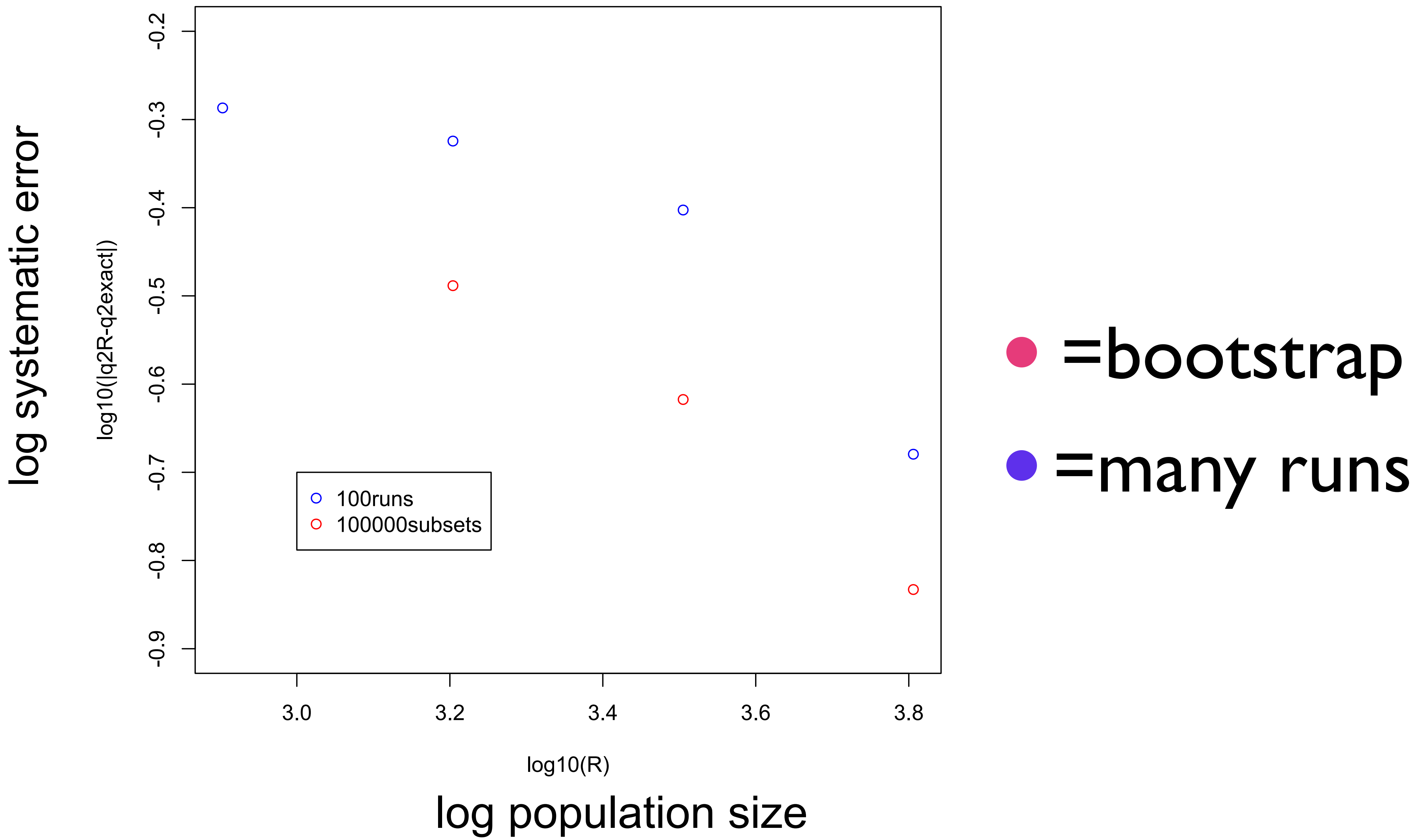
- Do many runs for each of many population sizes. Expensive.
- Is there an analog to the autocorrelation function and the exponential autocorrelation time?

Quantifying Convergence to Equilibrium

- Idea: resample small populations from a single run with a large population (bootstrap).
  - Choose a small population of parents and measure the observable in their descendants.
  - Repeat many times for different populations sizes.



- Problem: A small sample within a large population is not the same as an isolated small population because the sample within the large populations exists in a more competitive environment.



Direct Estimate of Free Energies

$$Q(\beta, \beta') = \frac{\sum_{j=1}^{R} \exp\left[-(\beta' - \beta)E_j\right]}{R}$$

$$-\beta_k F(\beta_k) = \sum_{\ell=k}^{K} \ln Q(\beta_{\ell+1}, \beta_\ell) + \beta_K F(\beta_K)$$

Derivation:

$$\begin{aligned} \frac{Z(\beta')}{Z(\beta)} &= \frac{\sum_\gamma e^{-\beta' E_\gamma}}{Z(\beta)} \\ &= \sum_\gamma e^{-(\beta'-\beta)E_\gamma} \left(\frac{e^{-\beta E_\gamma}}{Z(\beta)}\right) \\ &= \langle e^{-(\beta'-\beta)E_\gamma} \rangle_\beta \\ &\approx \frac{1}{R} \sum_{j=1}^{R} e^{-(\beta'-\beta)E_j} = Q(\beta, \beta'). \end{aligned}$$

Weighted Averages

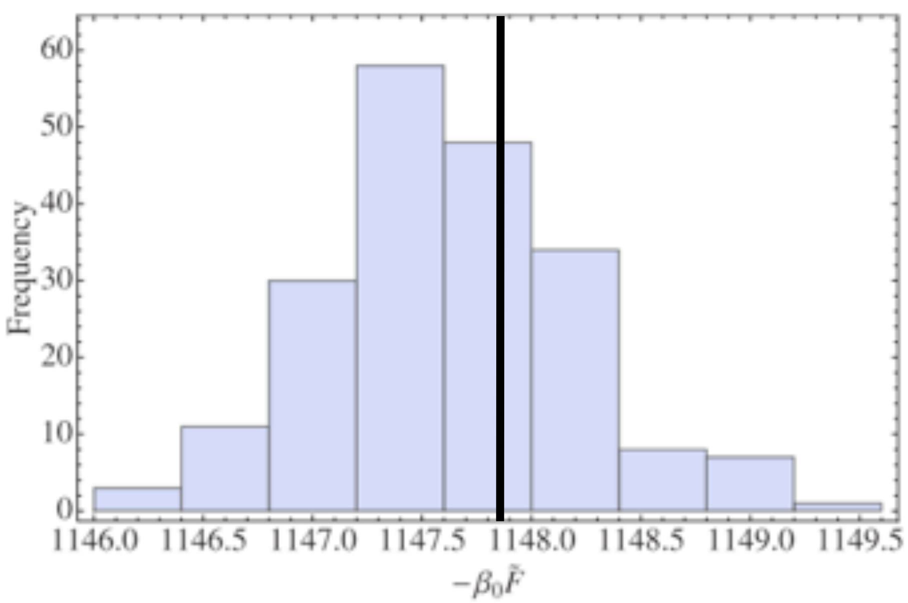
JM, PRE82,26704(2010)

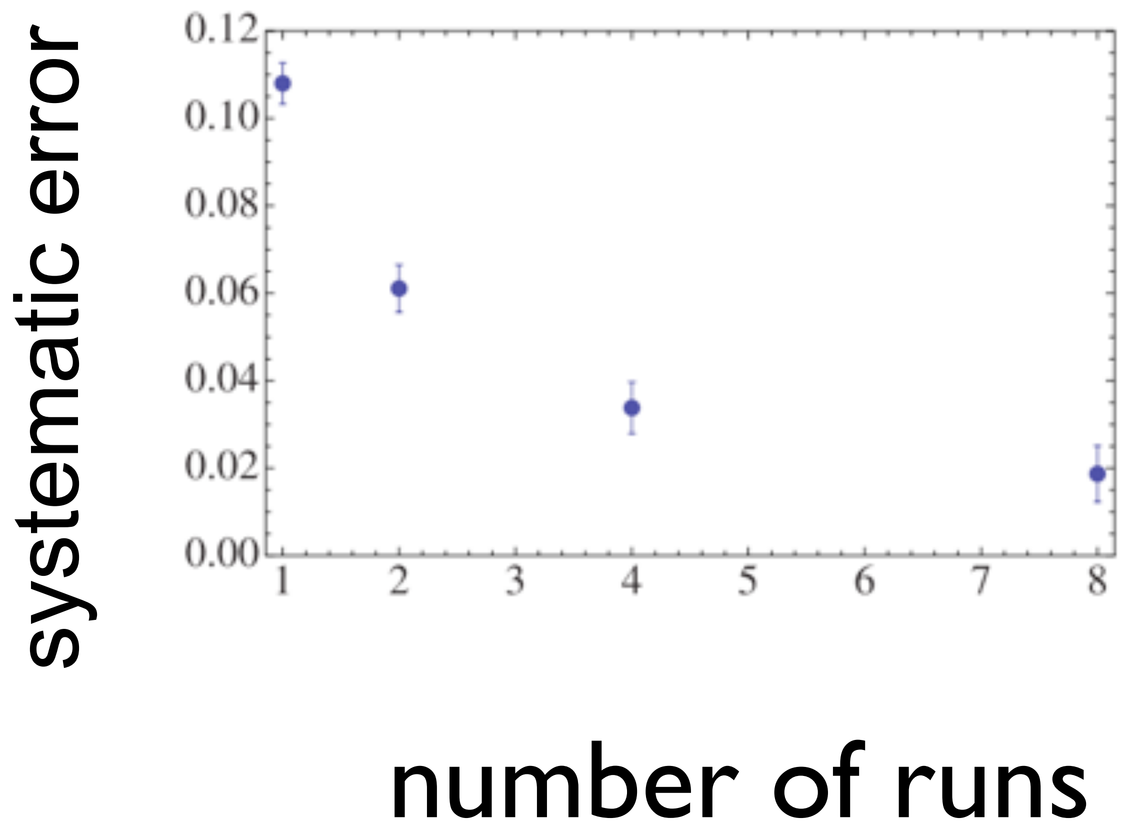
- Results using small population size are biased.
- Results from independent runs can be combined and biases reduced using *weighted* averages.
- Observables from each run weighted by the exponential of the free energy estimator for that run.

$$\overline{A}(\beta) = \sum_{m=1}^{M} \tilde{A}_m(\beta)\omega_m(\beta) \qquad \omega_m(\beta) = \frac{e^{-\beta \tilde{F}_m(\beta)}}{\sum_{i=1}^{M} e^{-\beta \tilde{F}_i(\beta)}}$$

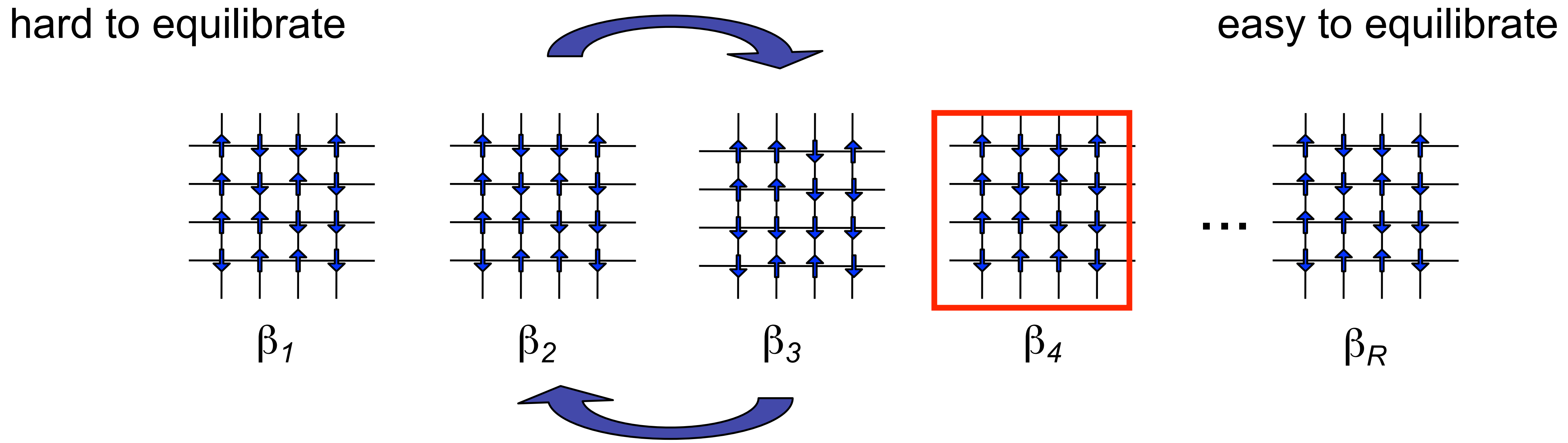
Weighted Averages

- Number of runs needed for unbiased results determined by the width of the distribution of the free energy estimator--an intrinsic measure of equilibration.





Parallel Tempering



•*R* replicas at inverse temperatures $\beta_1 > \beta_2 > \ldots > \beta_R$ (each with the same couplings).

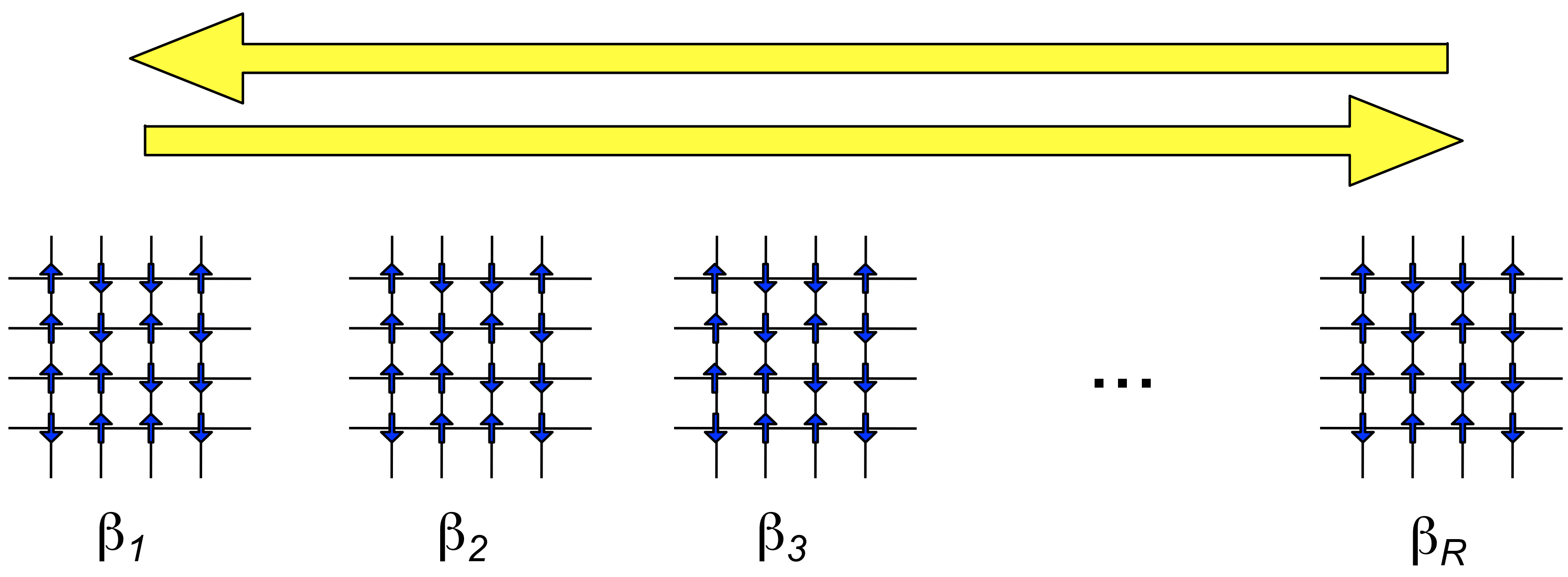
•MCMC (e.g. Metropolis) on each replica

•Exchange replicas with energies E and E' and temperatures β and β',with probability:

$$p_{\text{swap}} = \min\left[1, e^{(\beta-\beta')(E-E')}\right]$$

Intuition

- Mixing is accelerated by “round trips” from low to high temperature and back.



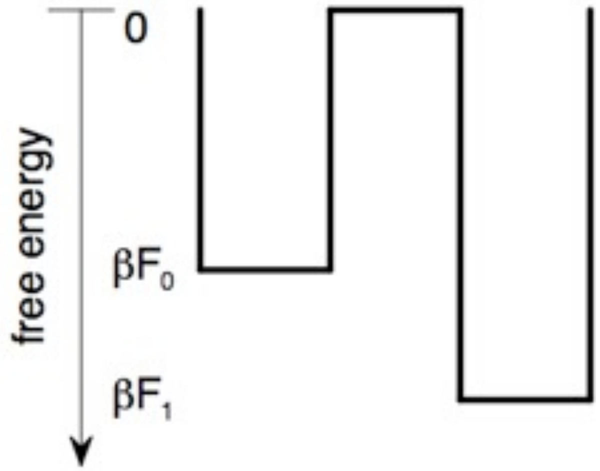
Population Annealing	Parallel Tempering
Sequential Monte Carlo	Markov Chain Monte Carlo
# Replicas ($R>>1$) *space*	#Sweeps ($t>>1$) *parallel time*
#Temperature steps (T) *parallel time*	#Replicas (R) *space*
work=RT	*work*=Rt
$(A(R)-A_{eq}) \sim R_0/R$	$(A(t)-A_{eq}) \sim e^{(-t/\tau)}$

$\approx$

Two-Well landscape

- Consider a toy free energy landscape with two free energy minima separated by a high barrier.

  – JM, PRE **80**, 056706 (2009)

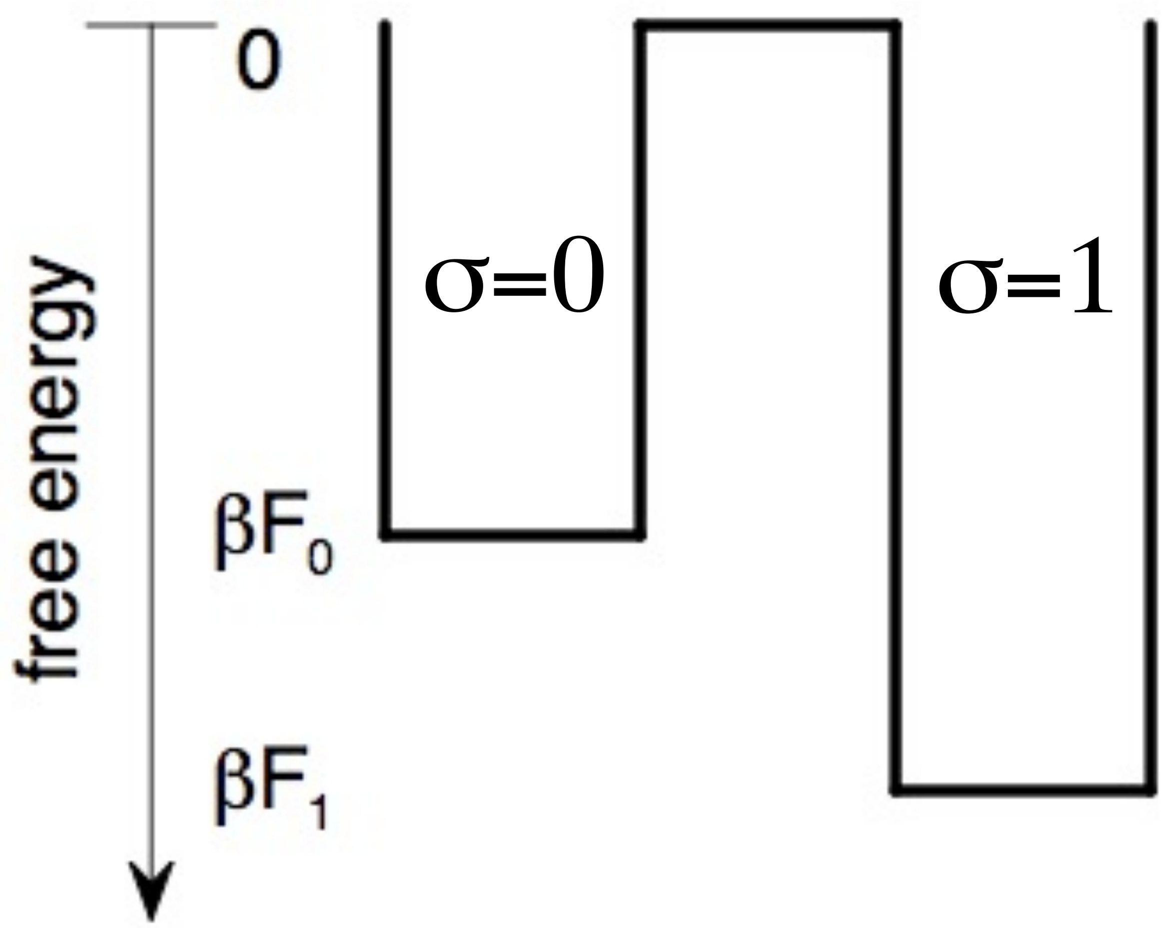


Two-Well Landscape

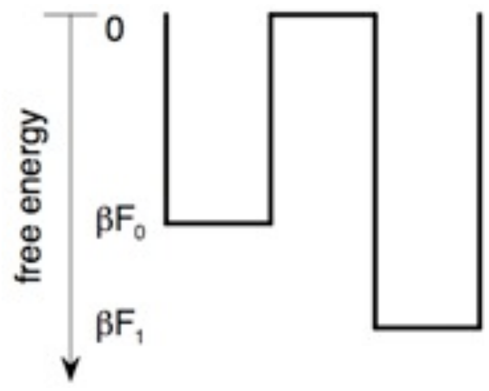
$$\beta F_\sigma(\beta) = -\frac{1}{2}(\beta - \beta_c)^2(K + H\sigma)$$

$$\beta \delta F = -\frac{1}{2}(\beta - \beta_c)^2 H$$

$$\text{Prob}\left[\sigma = +1\right] = \frac{1}{1 + e^{-\beta\delta F}}$$



Dynamics of the two-well model



- Assumptions:
  - Fast equilibration within each well.
  - No transitions between wells except at β_c where each well is equally probable.
  - Energy is normally distributed in each well; from thermodynamics:

$$\langle E \rangle = -(\beta - \beta_c)(K + H\sigma) \qquad \mathbf{Var}(E) = (K + H\sigma)$$

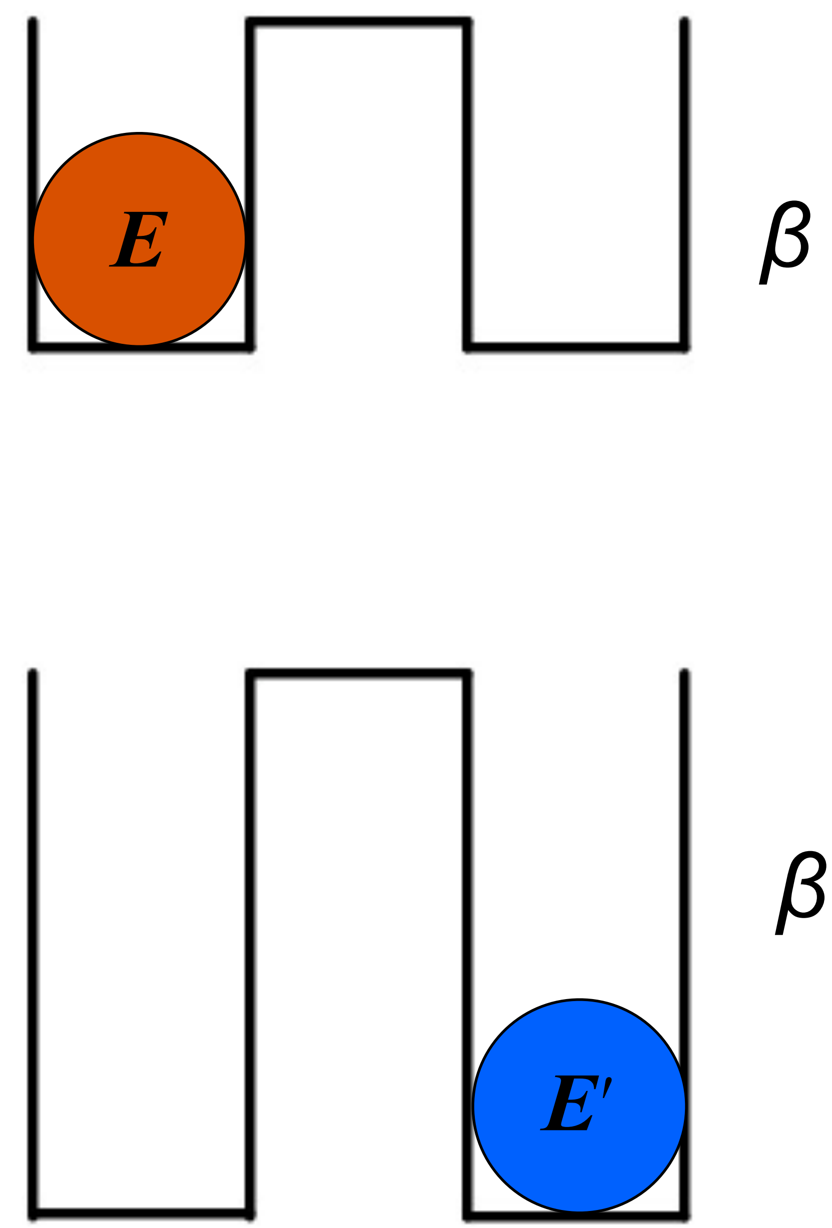
Replica exchange probabilities

- For the two-well model, replica exchange transition probabilities can be computed exactly. For symmetric wells ($H=0$):

$$p_{\text{swap}} = \min\left[1, e^{(\beta-\beta')(E-E')}\right]$$

$$\langle E \rangle = -(\beta - \beta_c)K \ , \ \mathbf{Var}(E) = K$$

$$p_{\text{swap}} = \frac{1}{2}\text{Erfc}\left((\beta - \beta')\sqrt{K}\right)$$



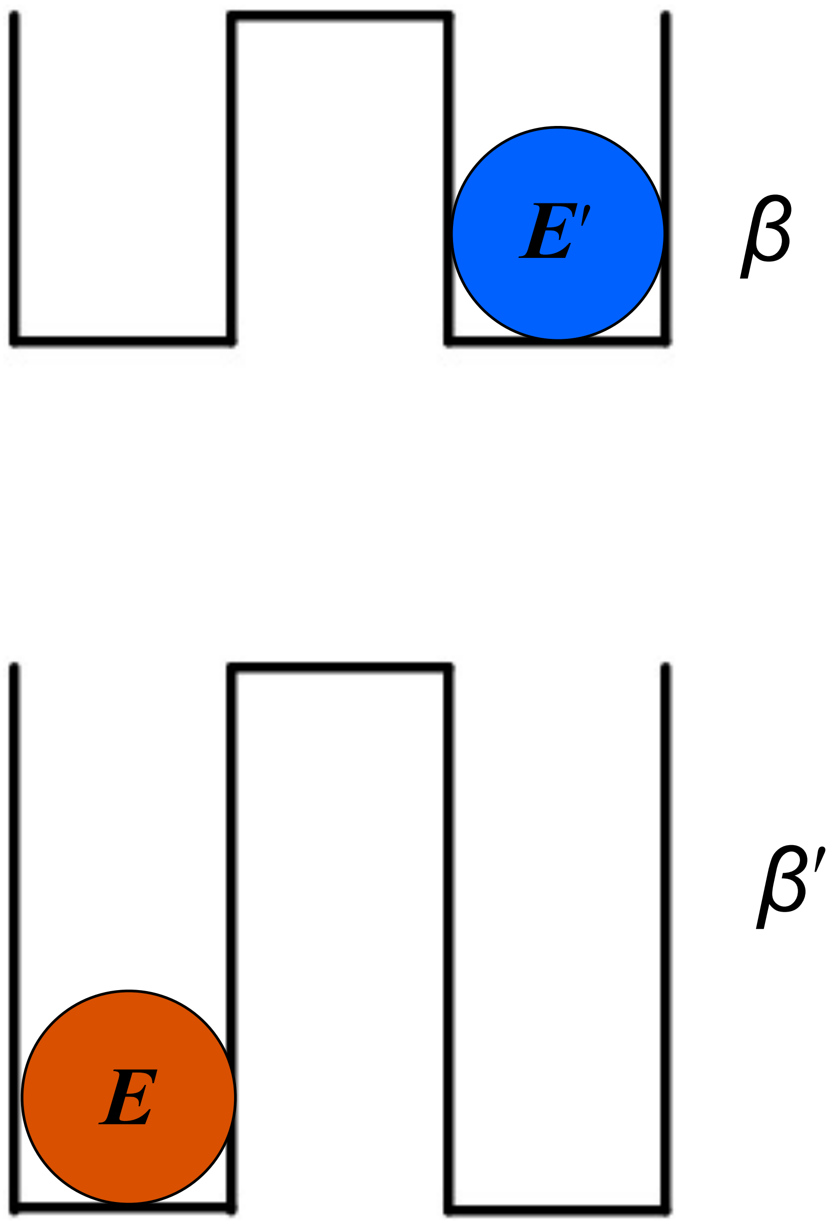
Replica exchange probabilities

- For the two-well model, replica exchange transition probabilities can be computed exactly. For symmetric wells ($H=0$):

$$p_{\text{swap}} = \min\left[1, e^{(\beta-\beta')(E-E')}\right]$$

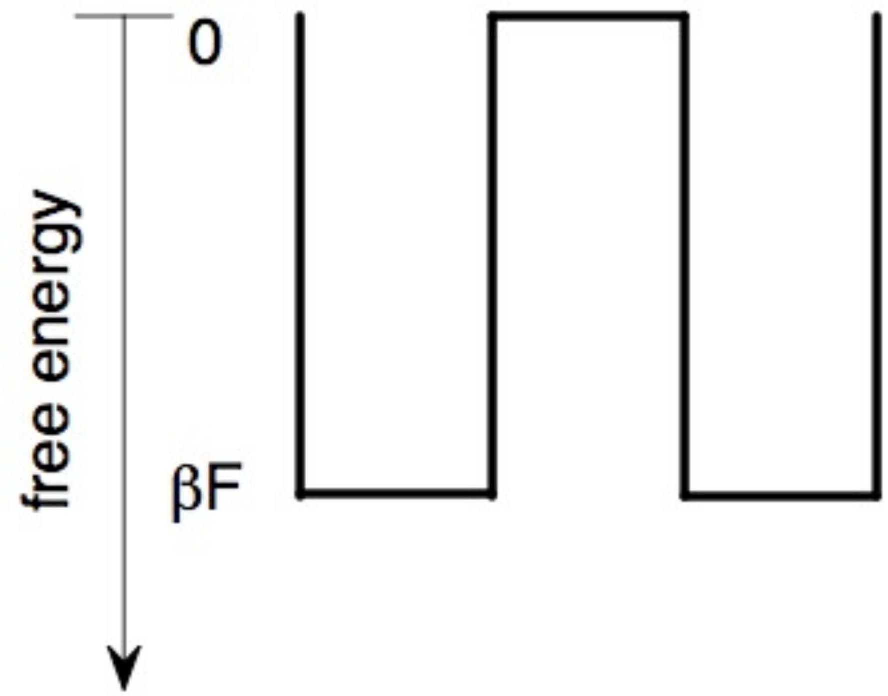
$$\langle E \rangle = -(\beta - \beta_c)K \ , \ \mathbf{Var}(E) = K$$

$$p_{\text{swap}} = \frac{1}{2}\text{Erfc}\left((\beta - \beta')\sqrt{K}\right)$$



PT for the two-well model

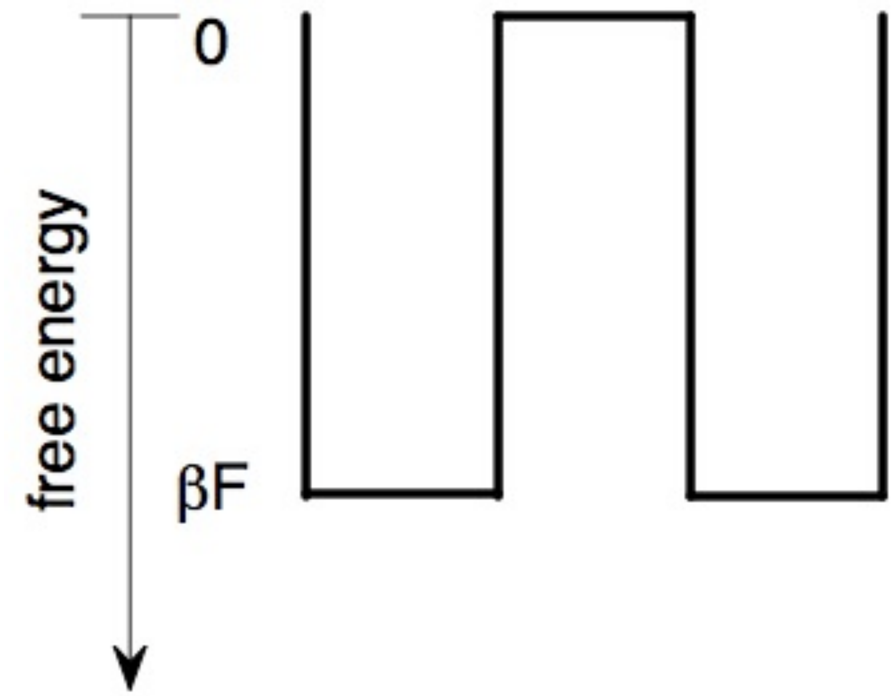
$$p_{\mathrm{swap}} = \frac{1}{2}\mathrm{Erfc}\left((\beta - \beta')\sqrt{K}\right)$$



PT for the two-well model

- *Diffusion* of replicas

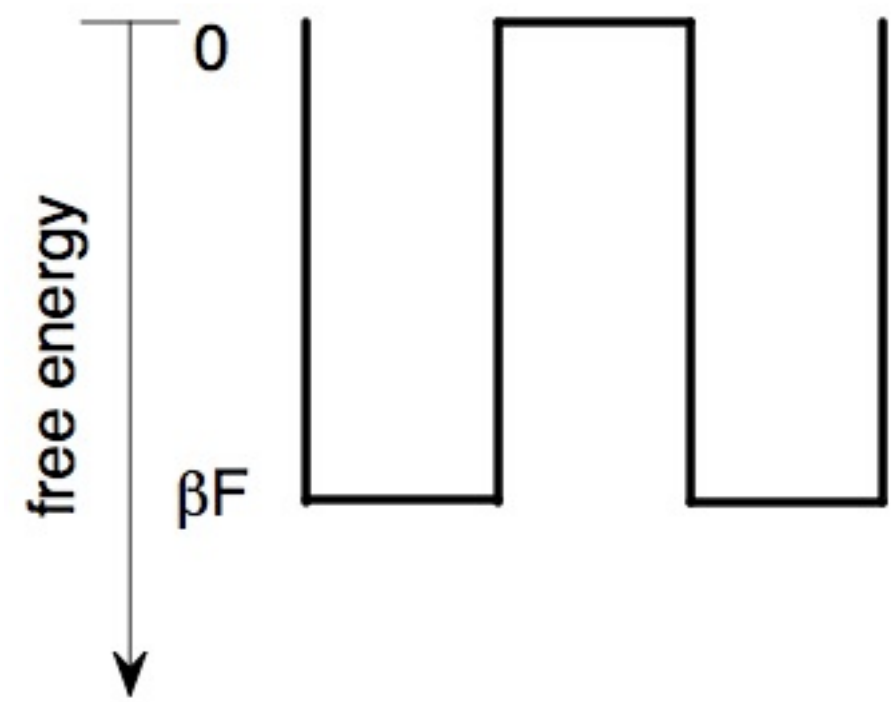
$$p_{\rm swap} = \frac{1}{2}\mathrm{Erfc}\left((\beta - \beta')\sqrt{K}\right)$$



PT for the two-well model

- *Diffusion* of replicas
- Equilibration time τ is proportional to the <u>mean first passage time</u> for diffusion from β_0 to β_c with R equally spaced replicas

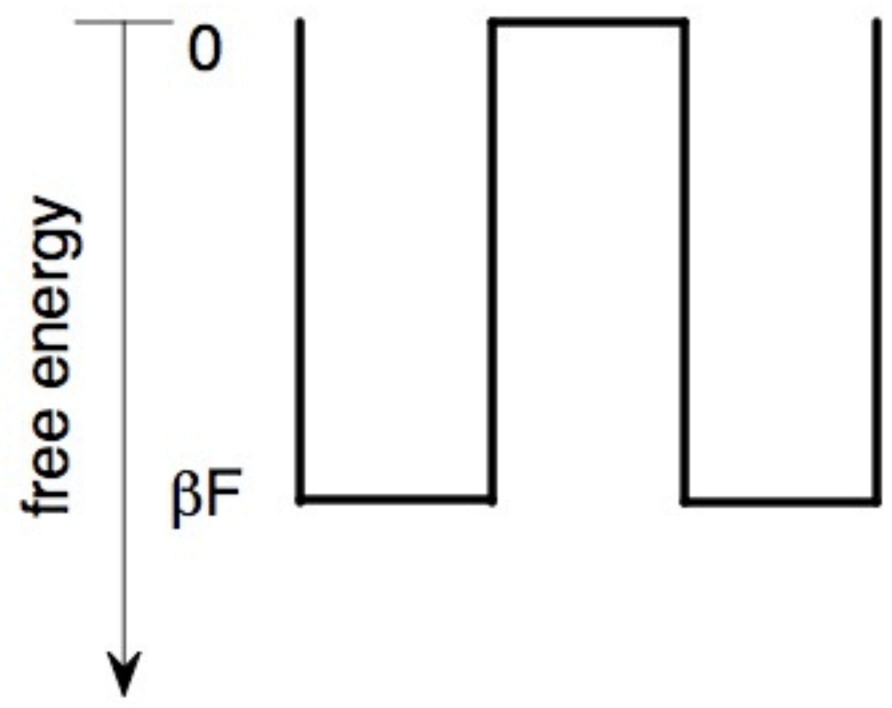
$$p_{\text{swap}} = \frac{1}{2}\text{Erfc}\left((\beta - \beta')\sqrt{K}\right)$$



PT for the two-well model

- *Diffusion* of replicas
- Equilibration time τ is proportional to the <u>mean first passage time</u> for diffusion from β_0 to β_c with R equally spaced replicas
- Optimum number of replicas balances diffusion time and replica exchange acceptance fraction.

$$p_{\mathrm{swap}} = \frac{1}{2}\mathrm{Erfc}\left((\beta - \beta')\sqrt{K}\right)$$

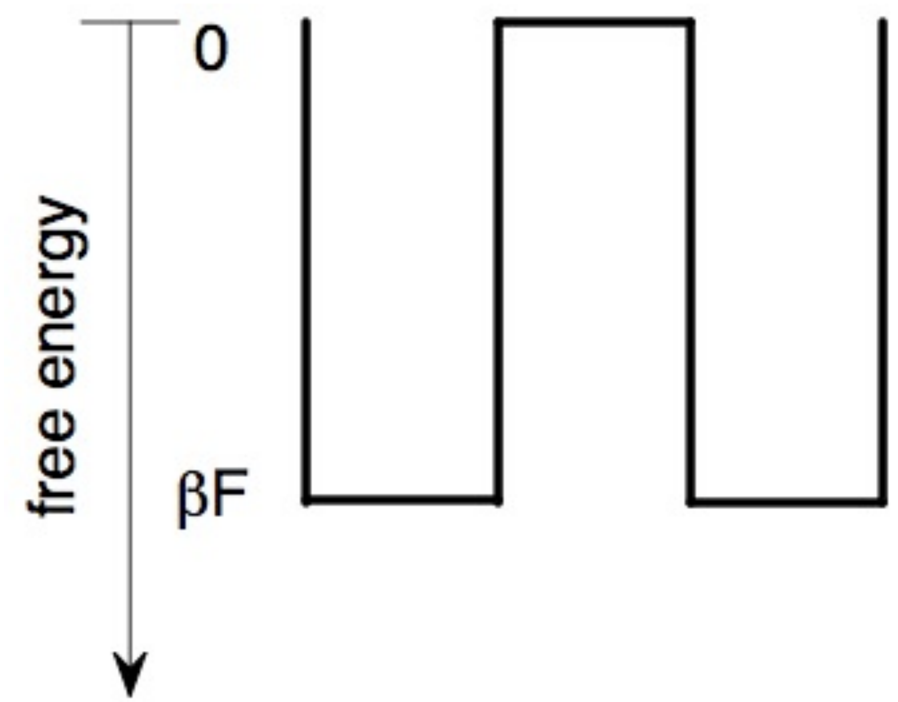


PT for the two-well model

- *Diffusion* of replicas

$$p_{\text{swap}} = \frac{1}{2}\text{Erfc}\left((\beta - \beta')\sqrt{K}\right)$$

- Equilibration time τ is proportional to the mean first passage time for diffusion from β_0 to β_c with R equally spaced replicas



- Optimum number of replicas balances diffusion time and replica exchange acceptance fraction.

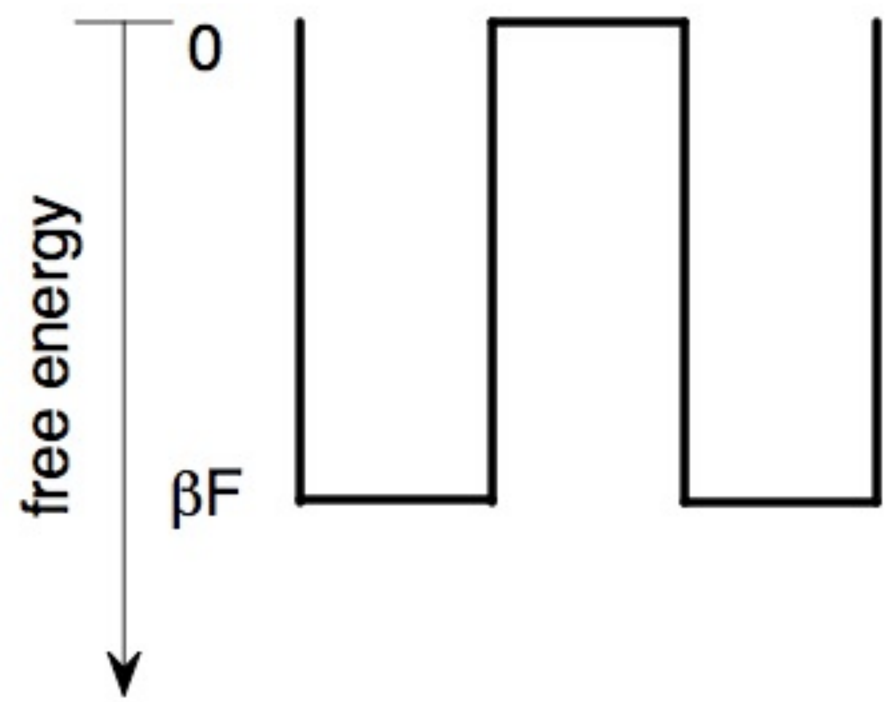
$$R_{\text{opt}} = 1 + 0.594(\beta_0 - \beta_c)\sqrt{K} \sim \sqrt{-\beta F}$$

PT for the two-well model

- *Diffusion* of replicas
- Equilibration time τ is proportional to the mean first passage time for diffusion from β_0 to β_c with R equally spaced replicas
- Optimum number of replicas balances diffusion time and replica exchange acceptance fraction.

$$p_{\text{swap}} = \frac{1}{2}\text{Erfc}\left((\beta - \beta')\sqrt{K}\right)$$

$$\tau \sim R^2 \sim \beta F$$



$$R_{\text{opt}} = 1 + 0.594(\beta_0 - \beta_c)\sqrt{K} \sim \sqrt{-\beta F}$$

PA for the two-well model

$$Y_i = \exp\left[-(\beta' - \beta)E_i + \lambda A_i\right]$$

$$I(\beta, \beta', \lambda) = \mathbf{E}\log\left(\frac{1}{R}\sum_{j=1}^{R} Y_j\right) = \log(\mathbf{E}Y) - \frac{1}{2R}\frac{\mathbf{Var}Y}{(\mathbf{E}Y)^2} + \mathrm{O}\left(\frac{1}{R^{3/2}}\right)$$

$$\langle A\rangle'_R = \left.\frac{dI(\beta, \beta', \lambda)}{d\lambda}\right|_{\lambda=0} = \mathbf{E}'A + \mathrm{O}\left(\frac{1}{R}\right)$$

$(A(R) - A_{eq}) \sim R_0/R$

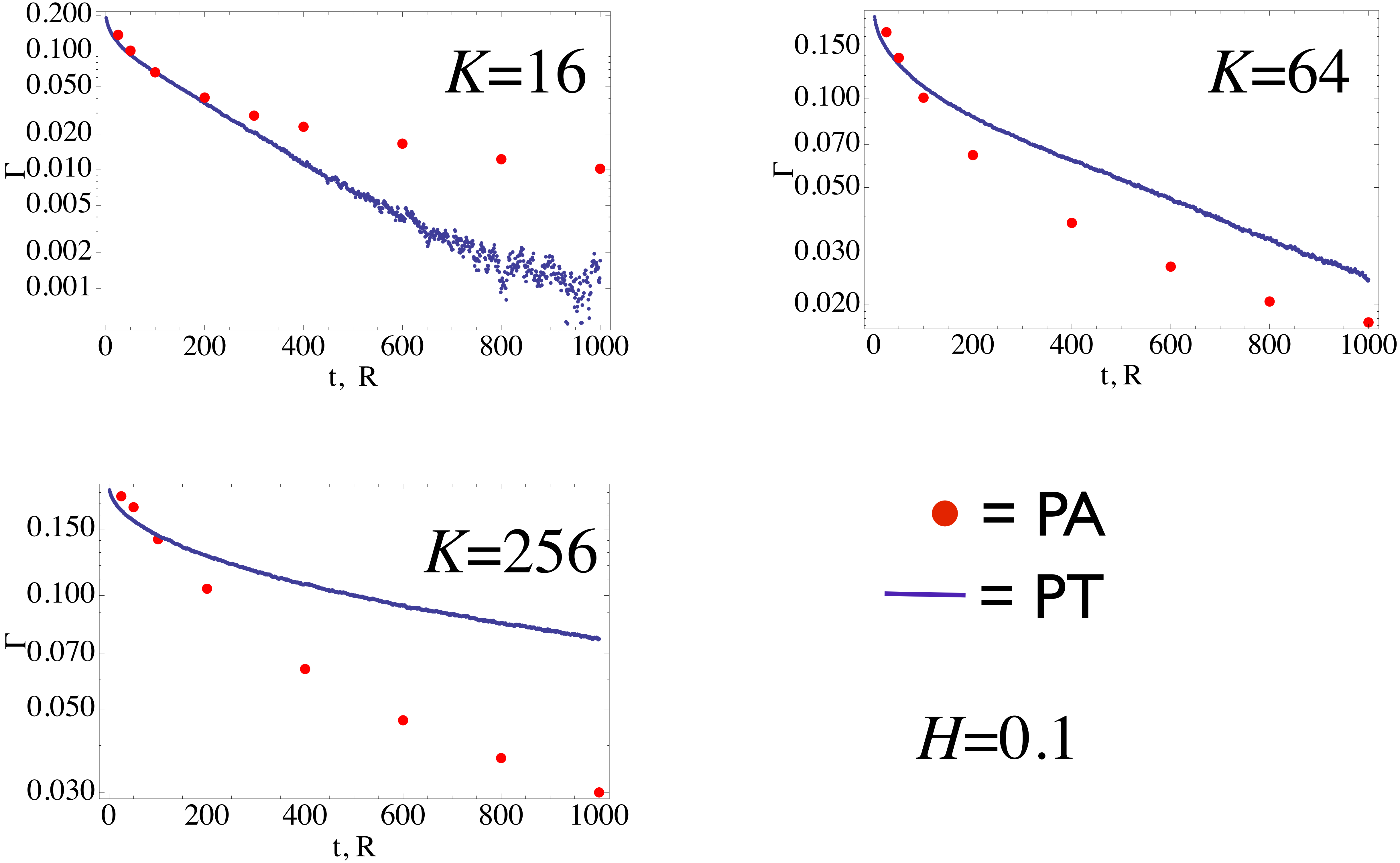
$T \sim K^{1/2}$

$R_0 \sim T^a \quad a = 1$?

JM, R. Ellis, JSP 144, 544 (2011)

Population Annealing	Parallel Tempering
# Replicas (R) *space*	#Sweeps (t) *parallel time*
#Temperature steps (T) *parallel time*	#Replicas (R) *space*
$(A(R) - A_{eq}) \sim R_0/R$	$(A(t) - A_{eq}) \sim e^{(-t/\tau)}$
$R_0 \sim T \quad T \sim K^{1/2}$	$\tau \sim K \quad R \sim K^{1/2}$
work to eq $\sim R_0 T \sim K$	work to eq $\sim \tau R \sim K^{3/2}$

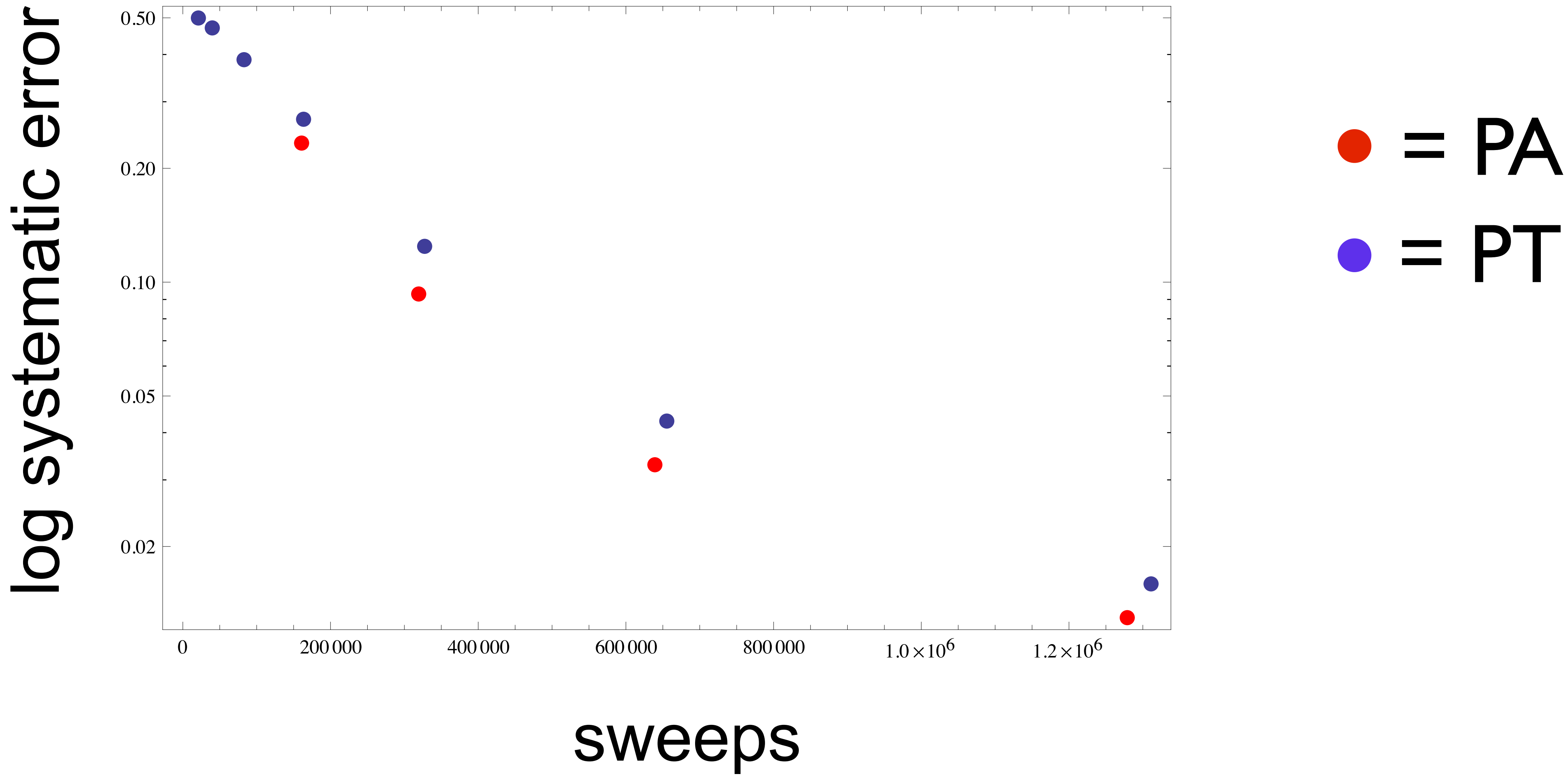
Compare PA and PT for the Two-Well Model



$$\Gamma = \langle \mathrm{Prob}\,[\sigma = +1] \rangle - \langle \mathrm{Prob}\,[\sigma = +1] \rangle_{\mathrm{eq}}$$

PA vs PT for 3D Spin Glass

- For a small handful of disorder realizations, doing many runs for different population sizes (for PA) or run lengths (for PT), the convergence to equilibrium is comparable as measured in sweeps.



Conclusions

- Both parallel tempering and population annealing overcome the exponential slowing associated with large free energy barriers.
- Population annealing is comparably efficient to parallel tempering and has several features to recommend it:
  - Massively parallel
  - Direct measurement of free energies
  - Perhaps more efficient?