Population annealing: Massively parallel simulations in statistical physics

Martin Weigel

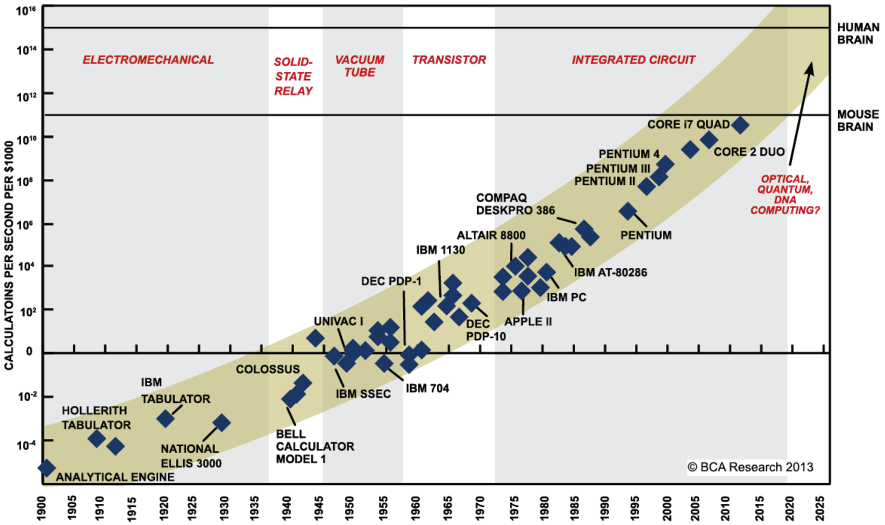
Applied Mathematics Research Centre, Coventry University, Coventry, United Kingdom

with Michal Borovský (Coventry & Kosice), Lev Barash, Lev Shchur (Landau Institute), and Wolfhard Janke (Leipzig)

17th International NTZ-Workshop on New Developments in Computational Physics
Universität Leipzig, November 24, 2016

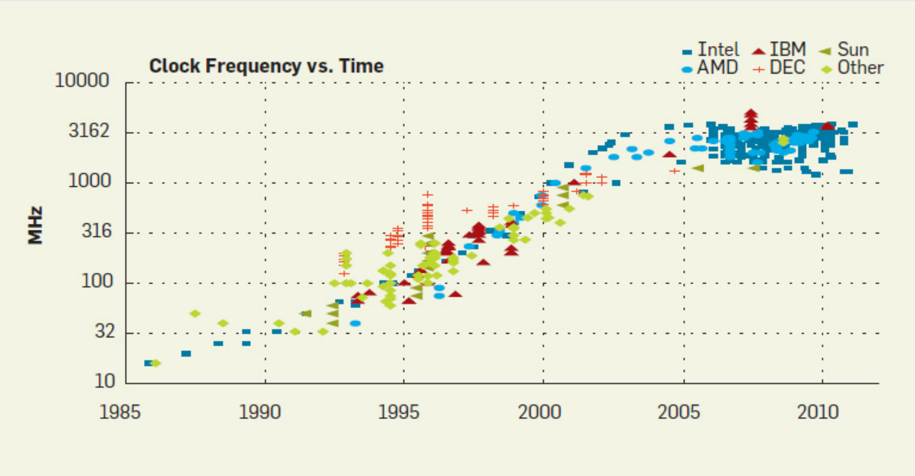
Parallel Computing and Monte Carlo

Moore's law

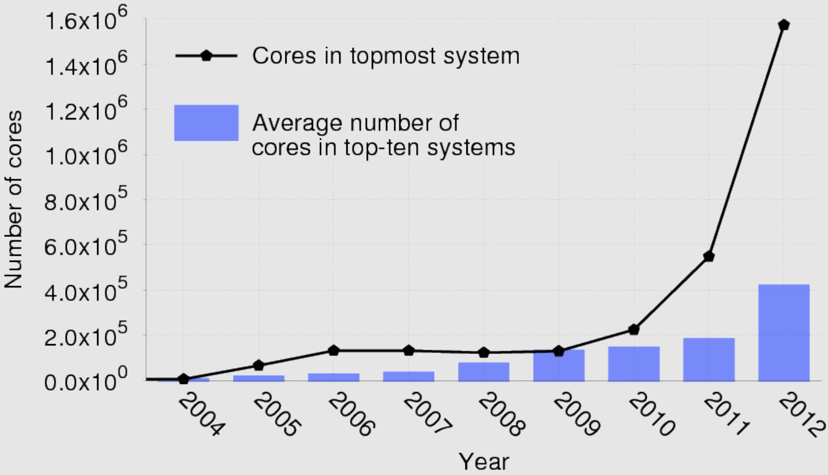


SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, *THE VIKING PRESS*, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

Moore's law


Clock Frequency vs. Time
Intel
IBM
Sun
AMD
DEC
Other
MHz
10000
3162
1000
316
100
32
10
1985
1990
1995
2000
2005
2010

Moore's law



1.6x10^6
1.4x10^6
1.2x10^6
1.0x10^6
8.0x10^5
6.0x10^5
4.0x10^5
2.0x10^5
0.0x10^0
Number of cores
Cores in topmost system
Average number of cores in top-ten systems
2004
2005
2006
2007
2008
2009
2010
2011
2012
Year

Monte Carlo simulations

Most successful approach is importance sampling through Markov chains, an inherently sequential process.

Monte Carlo simulations

Most successful approach is importance sampling through Markov chains, an inherently sequential process.

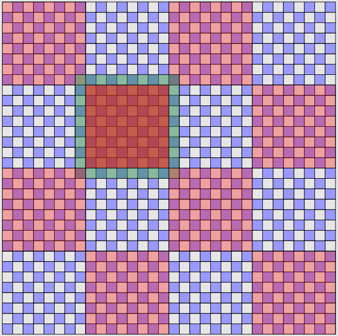
What to do it the era of parallel computing?

Monte Carlo simulations

Most successful approach is importance sampling through Markov chains, an inherently sequential process.

What to do it the era of parallel computing?

- use domain decompositions

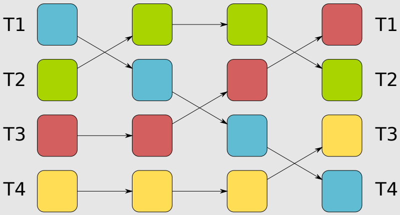


Monte Carlo simulations

Most successful approach is importance sampling through Markov chains, an inherently sequential process.

What to do it the era of parallel computing?

- use domain decompositions
- moderately parallel variants such as parallel tempering (Swendsen and Wang, 1986; Geyer, 1991; Hukushima and Nemoto, 1996)

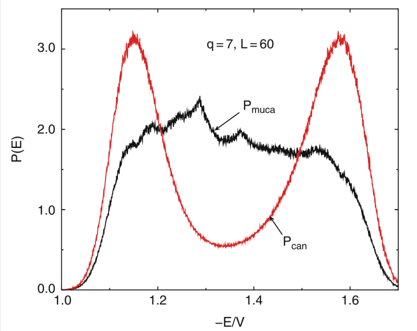


Monte Carlo simulations

Most successful approach is importance sampling through Markov chains, an inherently sequential process.

What to do it the era of parallel computing?

- use domain decompositions
- moderately parallel variants such as parallel tempering (Swendsen and Wang, 1986; Geyer, 1991; Hukushima and Nemoto, 1996)
- parallel multicanonical (Zierenberg et al., 2013) and Wang-Landau simulations (Vogel et al., 2013)

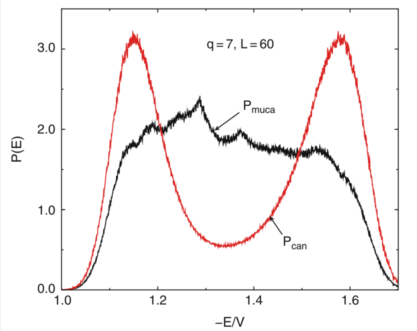


Monte Carlo simulations

Most successful approach is importance sampling through Markov chains, an inherently sequential process.

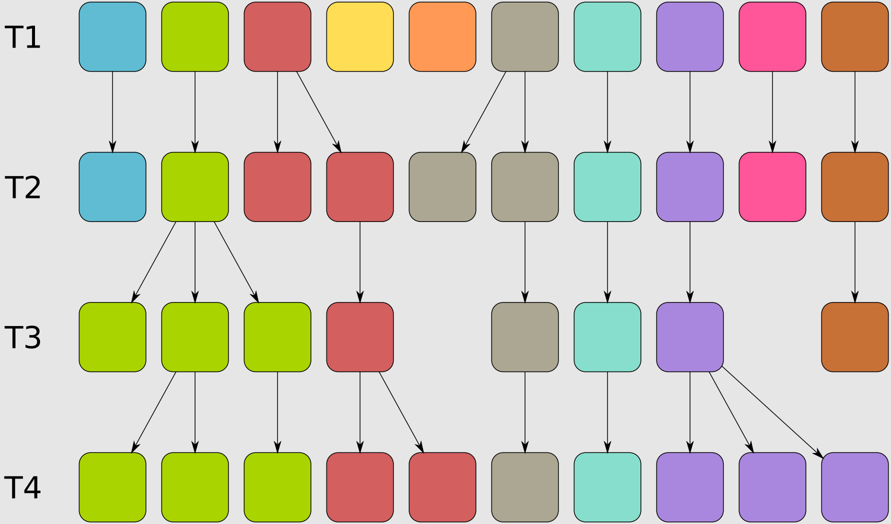
What to do it the era of parallel computing?

- use domain decompositions
- moderately parallel variants such as parallel tempering (Swendsen and Wang, 1986; Geyer, 1991; Hukushima and Nemoto, 1996)
- parallel multicanonical (Zierenberg et al., 2013) and Wang-Landau simulations (Vogel et al., 2013)



Can make use of a few dozen to a few hundred cores, but what to do with 10^6 cores?

Population annealing



Population Annealing

Population annealing

Population annealing algorithm `(Hukushima + Iba, 2003; Machta, 2010):`

1. Set up an equilibrium ensemble of R independent copies of the system at inverse temperature β_0. Typically $\beta_0 = 0$, where this can be easily achieved.

Population annealing

Population annealing algorithm (Hukushima + Iba, 2003; Machta, 2010):

1. Set up an equilibrium ensemble of R independent copies of the system at inverse temperature β_0. Typically $\beta_0 = 0$, where this can be easily achieved.
2. To create an approximately equilibrated sample at $\beta_i > \beta_{i-1}$, resample configurations with their relative Boltzmann weight $\exp[-(\beta_i - \beta_{i-1})E_j]/Q$, where $Q = \sum \exp(-(\beta_i - \beta_{i-1})E_j)$.

Population annealing

Population annealing algorithm (Hukushima + Iba, 2003; Machta, 2010):

1. Set up an equilibrium ensemble of R independent copies of the system at inverse temperature β_0. Typically $\beta_0 = 0$, where this can be easily achieved.
2. To create an approximately equilibrated sample at $\beta_i > \beta_{i-1}$, resample configurations with their relative Boltzmann weight $\exp[-(\beta_i - \beta_{i-1})E_j]/Q$, where $Q = \sum \exp(-(\beta_i - \beta_{i-1})E_j)$.

4. Calculate estimates for observable quantities $\mathcal{O}$ as population averages $\sum_j \mathcal{O}_j/R$.

Population annealing

Population annealing algorithm (Hukushima + Iba, 2003; Machta, 2010):

1. Set up an equilibrium ensemble of R independent copies of the system at inverse temperature β_0. Typically $\beta_0 = 0$, where this can be easily achieved.
2. To create an approximately equilibrated sample at $\beta_i > \beta_{i-1}$, resample configurations with their relative Boltzmann weight $\exp[-(\beta_i - \beta_{i-1})E_j]/Q$, where $Q = \sum \exp(-(\beta_i - \beta_{i-1})E_j)$.

4. Calculate estimates for observable quantities $\mathcal{O}$ as population averages $\sum_j \mathcal{O}_j/R$.
5. Goto step 2 until target temperature is reached.

Population annealing

Population annealing algorithm (Hukushima + Iba, 2003; Machta, 2010):

1. Set up an equilibrium ensemble of R independent copies of the system at inverse temperature β_0. Typically $\beta_0 = 0$, where this can be easily achieved.
2. To create an approximately equilibrated sample at $\beta_i > \beta_{i-1}$, resample configurations with their relative Boltzmann weight $\exp[-(\beta_i - \beta_{i-1})E_j]/Q$, where $Q = \sum \exp(-(\beta_i - \beta_{i-1})E_j)$.

4. Calculate estimates for observable quantities $\mathcal{O}$ as population averages $\sum_j \mathcal{O}_j/R$.
5. Goto step 2 until target temperature is reached.

This is a correct *sequential Monte Carlo algorithm*, but it is not very efficient.

Population annealing

Population annealing algorithm (Hukushima + Iba, 2003; Machta, 2010):

1. Set up an equilibrium ensemble of R independent copies of the system at inverse temperature β_0. Typically $\beta_0 = 0$, where this can be easily achieved.
2. To create an approximately equilibrated sample at $\beta_i > \beta_{i-1}$, resample configurations with their relative Boltzmann weight $\exp[-(\beta_i - \beta_{i-1})E_j]/Q$, where $Q = \sum \exp(-(\beta_i - \beta_{i-1})E_j)$.
3. Update each copy (replica) by θ rounds of an MCMC algorithm at inverse temperature β_i.
4. Calculate estimates for observable quantities $\mathcal{O}$ as population averages $\sum_j \mathcal{O}_j/R$.
5. Goto step 2 until target temperature is reached.

This is a correct *sequential Monte Carlo algorithm*, but it is not very efficient.
To improve it, all configurations undergo evolution with a standard Markov chain Monte Carlo (MCMC) algorithm ('single spin flips').

Benchmark: the 2D Ising model

Check results for the fruit fly of statistical mechanics, the 2D Ising model.

Benchmark: the 2D Ising model

Check results for the fruit fly of statistical mechanics, the 2D Ising model.

Hamiltonian

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} s_i s_j, \quad s_i = \pm 1$$

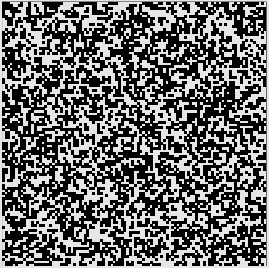
Benchmark: the 2D Ising model

Check results for the fruit fly of statistical mechanics, the 2D Ising model.

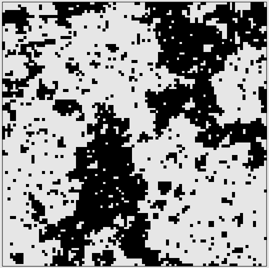
Hamiltonian

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} s_i s_j, \quad s_i = \pm 1$$

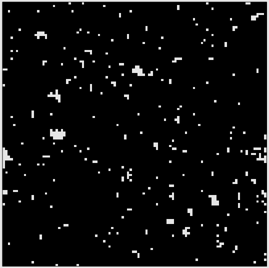
$T \gg T_c$



$T \approx T_c$

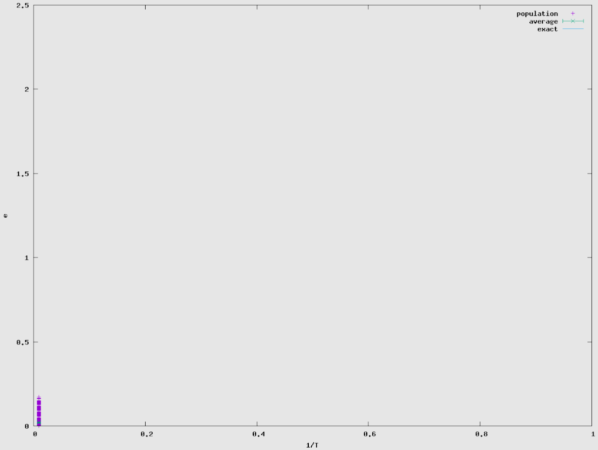


$T \ll T_c$



Population annealing

A sequential annealing of the population from infinite temperature, $\beta = 0$, down to $\beta = 1$.

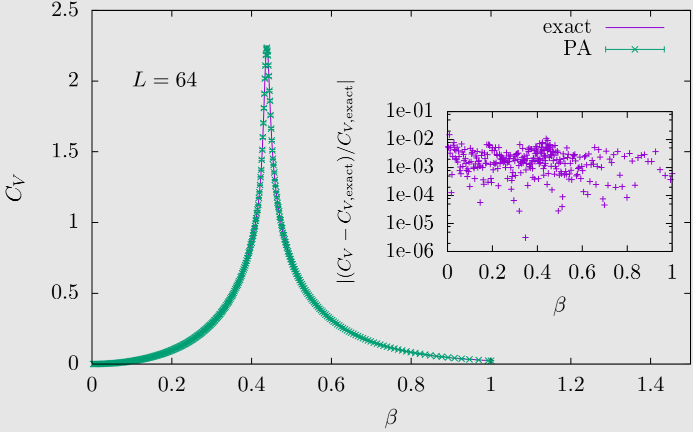


Correct results?

Exact results are available for finite lattices for the internal energy, specific heat and free energy (Ferdinand + Fisher, 1969).

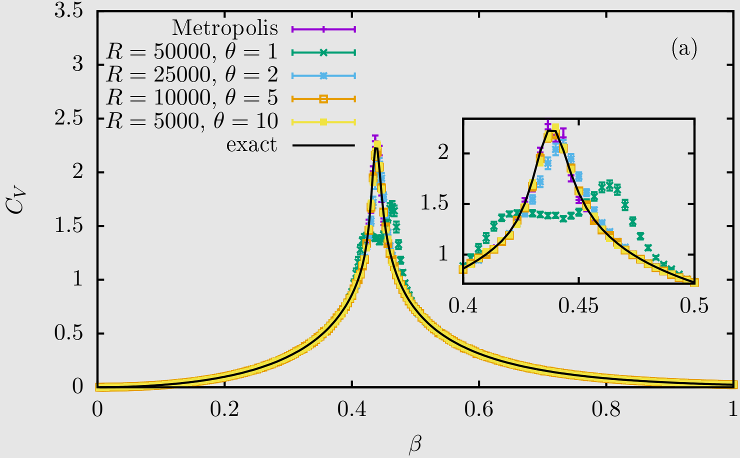
Correct results?

Exact results are available for finite lattices for the internal energy, specific heat and free energy (Ferdinand + Fisher, 1969).

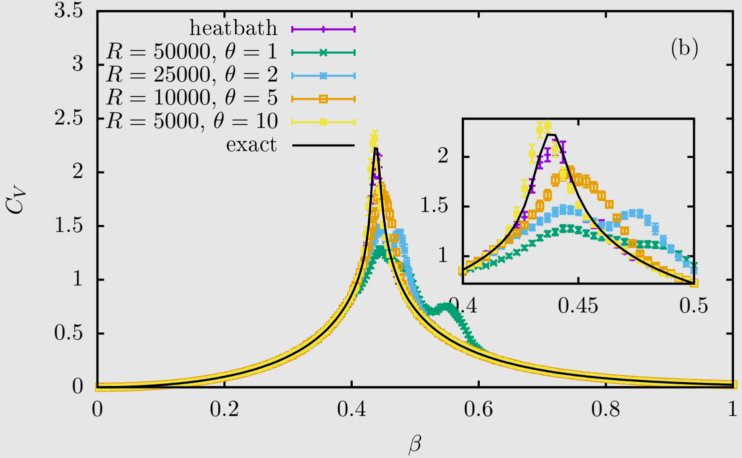


$R = 50\,000$, $\theta = 10$

Not always

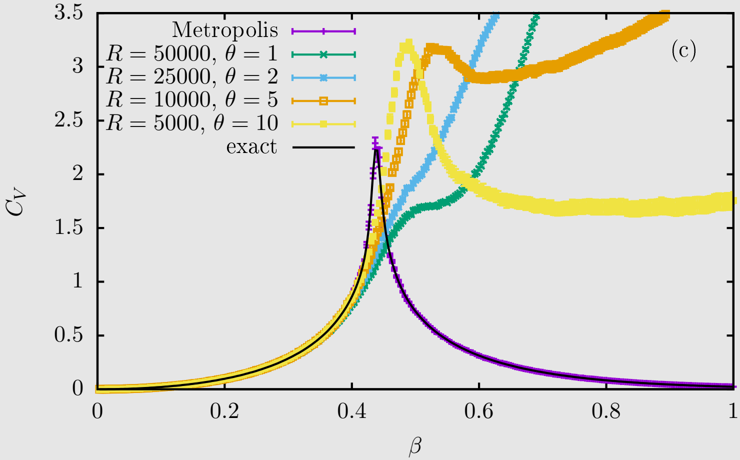

3.5
3
2.5
2
1.5
1
0.5
0
C_V
0
0.2
0.4
0.6
0.8
1
β
Metropolis
$R = 50000,\ \theta = 1$
$R = 25000,\ \theta = 2$
$R = 10000,\ \theta = 5$
$R = 5000,\ \theta = 10$
exact
(a)
2
1.5
1
0.4
0.45
0.5

Not always



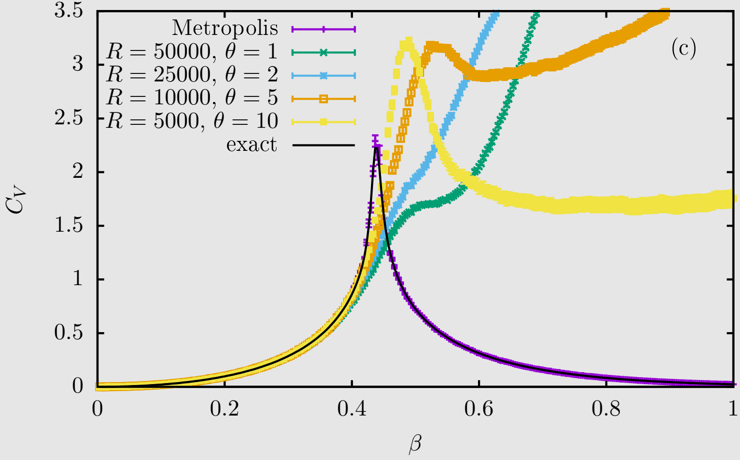
heatbath
$R = 50000,\ \theta = 1$
$R = 25000,\ \theta = 2$
$R = 10000,\ \theta = 5$
$R = 5000,\ \theta = 10$
exact
(b)
C_V
β
0
0.2
0.4
0.6
0.8
1
0.5
1.5
2
2.5
3
3.5
0.45

Not always



3.5
3
2.5
2
1.5
1
0.5
0
C_V
0
0.2
0.4
0.6
0.8
1
β
Metropolis
$R = 50000,\ \theta = 1$
$R = 25000,\ \theta = 2$
$R = 10000,\ \theta = 5$
$R = 5000,\ \theta = 10$
exact
(c)

Not always



Need to understand dependence on parameters, R, θ, $\Delta\beta$.

Bias and statistical error

Population annealing and MCMC

The general behavior of Markov chain Monte Carlo is well understood: if balance and ergodicity are fulfilled, it relaxes into equilibrium exponentially,

$$A_t \sim \langle A \rangle (1 + a e^{-t/t_{\mathrm{relax}}}).$$

Population annealing and MCMC

The general behavior of Markov chain Monte Carlo is well understood: if balance and ergodicity are fulfilled, it relaxes into equilibrium exponentially,

$$A_t \sim \langle A \rangle (1 + a e^{-t/t_{\mathrm{relax}}}).$$

In equilibrium, the Markov property results in correlations of successive measurements (in contrast to simple sampling):

$$C_{\Delta t} = \langle A_t A_{t+\Delta t} \rangle - \langle A_t \rangle \langle A_{t+\Delta t} \rangle \sim e^{-\Delta t/\tau_{\mathrm{exp}}}.$$

Population annealing and MCMC

The general behavior of Markov chain Monte Carlo is well understood: if balance and ergodicity are fulfilled, it relaxes into equilibrium exponentially,

$$A_t \sim \langle A \rangle (1 + a e^{-t/t_{\mathrm{relax}}}).$$

In equilibrium, the Markov property results in correlations of successive measurements (in contrast to simple sampling):

$$C_{\Delta t} = \langle A_t A_{t+\Delta t} \rangle - \langle A_t \rangle \langle A_{t+\Delta t} \rangle \sim e^{-\Delta t/\tau_{\mathrm{exp}}}.$$

Hence, while for N independent measurements the standard deviation $\sigma(\bar{A})$ of the mean $\bar{A} = \sum_t A_t$ ("error bar") is given by

$$\sigma^2_{\mathrm{uncorr}}(\bar{A}) = \frac{\sigma^2(A)}{N},$$

Population annealing and MCMC

The general behavior of Markov chain Monte Carlo is well understood: if balance and ergodicity are fulfilled, it relaxes into equilibrium exponentially,

$$A_t \sim \langle A \rangle (1 + a e^{-t/t_{\mathrm{relax}}}).$$

In equilibrium, the Markov property results in correlations of successive measurements (in contrast to simple sampling):

$$C_{\Delta t} = \langle A_t A_{t+\Delta t} \rangle - \langle A_t \rangle \langle A_{t+\Delta t} \rangle \sim e^{-\Delta t/\tau_{\mathrm{exp}}}.$$

Hence, while for N independent measurements the standard deviation $\sigma(\bar{A})$ of the mean $\bar{A} = \sum_t A_t$ ("error bar") is given by

$$\sigma^2_{\mathrm{uncorr}}(\bar{A}) = \frac{\sigma^2(A)}{N},$$

in the presence of correlations we find instead

$$\sigma^2(\bar{A}) = \frac{\sigma^2(A)}{N_{\mathrm{eff}}}, \quad N_{\mathrm{eff}} = N/2\tau_{\mathrm{int}},$$

Population annealing and MCMC

The general behavior of Markov chain Monte Carlo is well understood: if balance and ergodicity are fulfilled, it relaxes into equilibrium exponentially,

$$A_t \sim \langle A\rangle(1 + ae^{-t/t_{\mathrm{relax}}}).$$

In equilibrium, the Markov property results in correlations of successive measurements (in contrast to simple sampling):

$$C_{\Delta t} = \langle A_t A_{t+\Delta t}\rangle - \langle A_t\rangle\langle A_{t+\Delta t}\rangle \sim e^{-\Delta t/\tau_{\mathrm{exp}}}.$$

Hence, while for N independent measurements the standard deviation $\sigma(\bar{A})$ of the mean $\bar{A} = \sum_t A_t$ ("error bar") is given by

$$\sigma^2_{\mathrm{uncorr}}(\bar{A}) = \frac{\sigma^2(A)}{N},$$

in the presence of correlations we find instead

$$\sigma^2(\bar{A}) = \frac{\sigma^2(A)}{N_{\mathrm{eff}}}, \quad N_{\mathrm{eff}} = N/2\tau_{\mathrm{int}},$$

where

$$\tau_{\mathrm{int}} = \frac{1}{2} + \sum_{\Delta t=1}^{N-1} \frac{C_{\Delta t}}{C_0}\left(1 - \frac{\Delta t}{N}\right).$$

Bias and statistical errors

For MCMC we know that the bias decays exponentially $e^{-t/t_{\mathrm{relax}}}$, while the statistical errors fall off proportional to $1/\sqrt{N}$.

Bias and statistical errors

For MCMC we know that the bias decays exponentially $e^{-t/t_{\mathrm{relax}}}$, while the statistical errors fall off proportional to $1/\sqrt{N}$.

What about population annealing? This is not MCMC, but sequential MC. On the other hand, the spin flips ("equilibrating subroutine") are again MCMC, so the total method is a hybrid of MCMC and sequential MC.

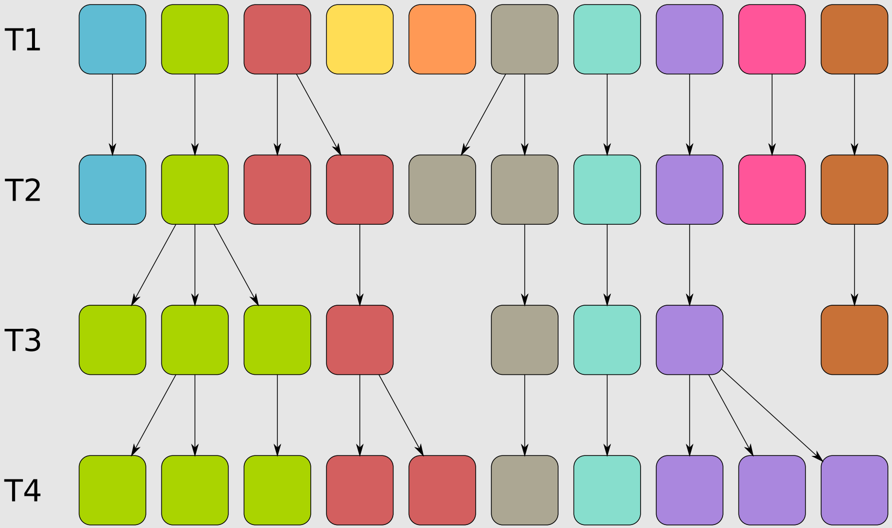
Bias and statistical errors

For MCMC we know that the bias decays exponentially $e^{-t/t_{\mathrm{relax}}}$, while the statistical errors fall off proportional to $1/\sqrt{N}$.

What about population annealing? This is not MCMC, but sequential MC. On the other hand, the spin flips ("equilibrating subroutine") are again MCMC, so the total method is a hybrid of MCMC and sequential MC.

Statistical errors
Population at each temperature forms generation of an ancestor tree. Members of different *families* are statistically independent.

Population annealing



Bias and statistical errors

For MCMC we know that the bias decays exponentially $e^{-t/t_{\mathrm{relax}}}$, while the statistical errors fall off proportional to $1/\sqrt{N}$.

What about population annealing? This is not MCMC, but sequential MC. On the other hand, the spin flips ("equilibrating subroutine") are again MCMC, so the total method is a hybrid of MCMC and sequential MC.

Statistical errors
Population at each temperature forms generation of an ancestor tree. Members of different *families* are statistically independent.

One might study the family statistics to understand correlations (Wang, Machta + Katzgraber, 2014/15):

Bias and statistical errors

For MCMC we know that the bias decays exponentially $e^{-t/t_{\mathrm{relax}}}$, while the statistical errors fall off proportional to $1/\sqrt{N}$.

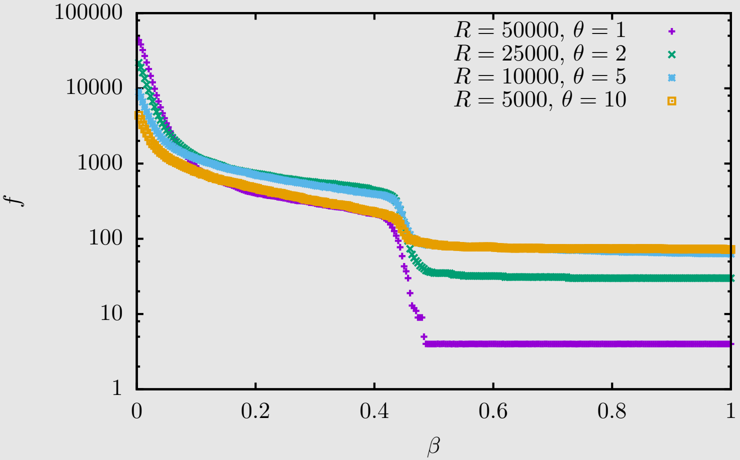
What about population annealing? This is not MCMC, but sequential MC. On the other hand, the spin flips ("equilibrating subroutine") are again MCMC, so the total method is a hybrid of MCMC and sequential MC.

Statistical errors
Population at each temperature forms generation of an ancestor tree. Members of different *families* are statistically independent.

One might study the family statistics to understand correlations `(Wang, Machta + Katzgraber, 2014/15)`:

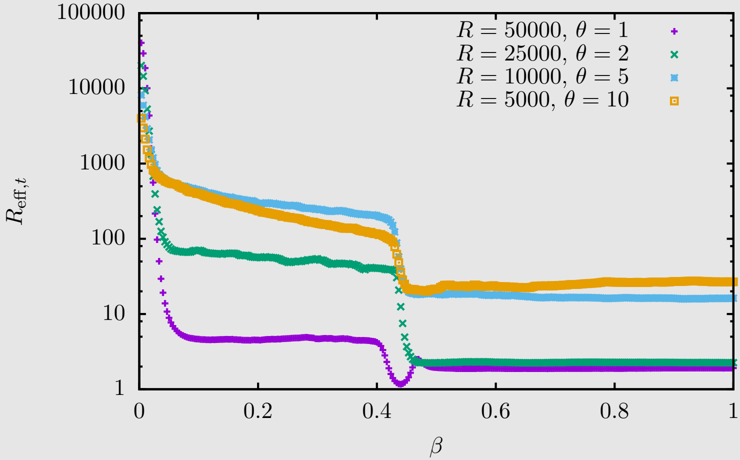
- Number of families: f
- Effective number of independent configurations: $N_{\mathrm{eff},t} = \left[\sum_i \mathfrak{n}_i^2\right]^{-1}$.
- Independence from family entropy: $N_{\mathrm{eff},s} = \exp\left(-\sum_i \mathfrak{n}_i \ln \mathfrak{n}_i\right)$

Families



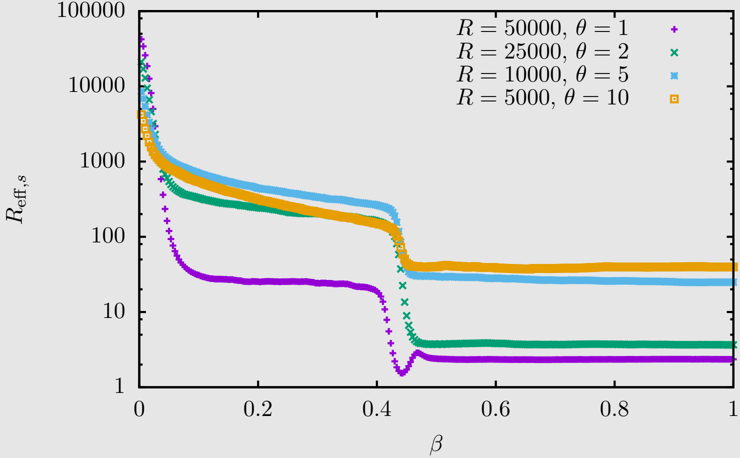
100000
10000
1000
100
10
1
0
0.2
0.4
0.6
0.8
1
f
β
$R = 50000,\ \theta = 1$
$R = 25000,\ \theta = 2$
$R = 10000,\ \theta = 5$
$R = 5000,\ \theta = 10$

Families

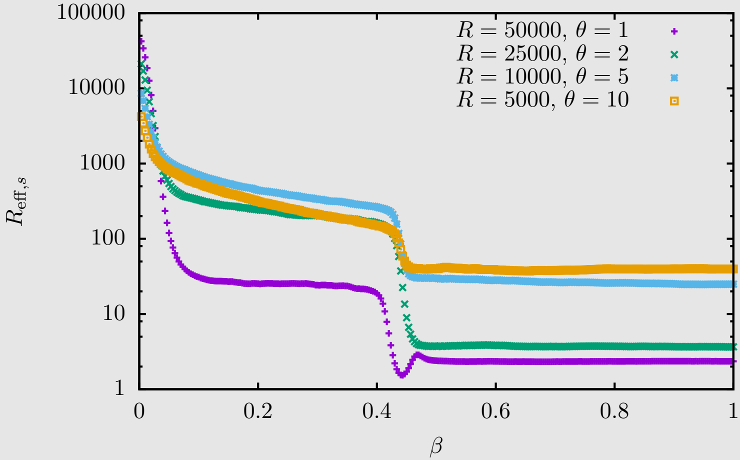


$R = 50000,\ \theta = 1$
$R = 25000,\ \theta = 2$
$R = 10000,\ \theta = 5$
$R = 5000,\ \theta = 10$
100000
10000
1000
100
10
1
0
0.2
0.4
0.6
0.8
1
$R_{\mathrm{eff},t}$
β

Families

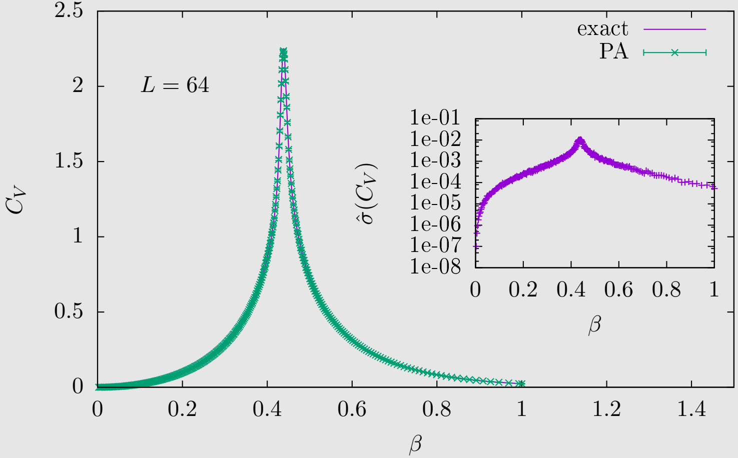

100000
10000
1000
100
10
1
$R_{\mathrm{eff},s}$
0
0.2
0.4
0.6
0.8
1
β
$R = 50000,\ \theta = 1$
$R = 25000,\ \theta = 2$
$R = 10000,\ \theta = 5$
$R = 5000,\ \theta = 10$

Families



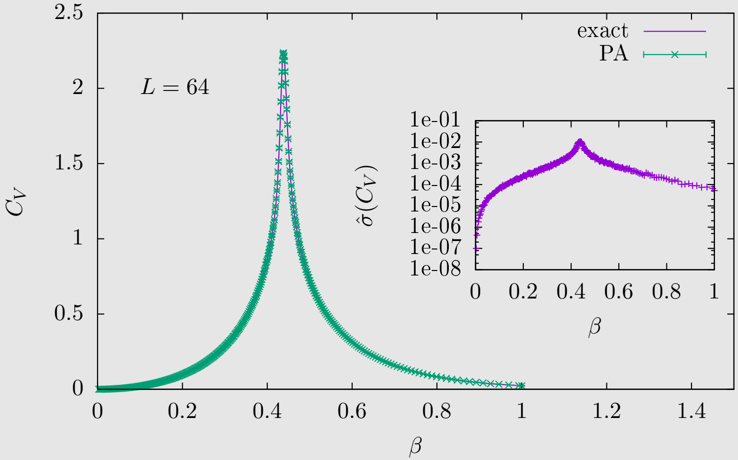
But statistical errors behave differently.

Families



But statistical errors behave differently.

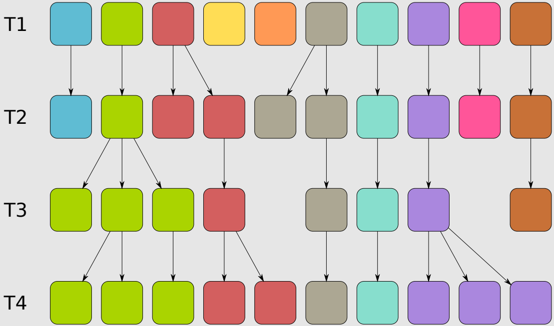
Families



But statistical errors behave differently. Families do not take spin flips into account!

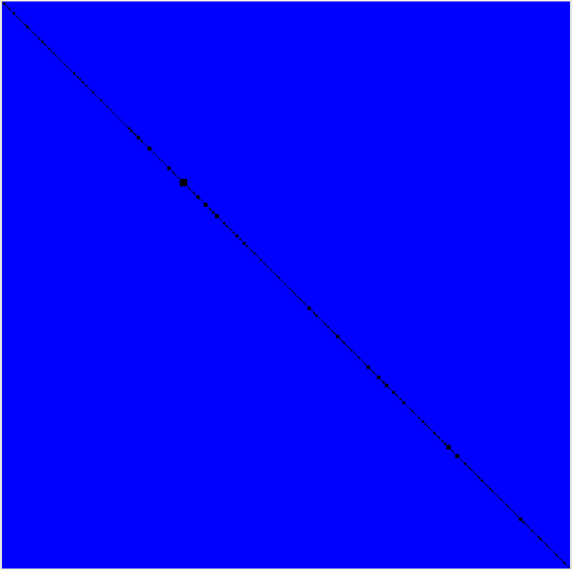
Correlations and tree distances

Correlations in population annealing are through the resampling, configurations with common ancestors are correlated. The closer the lowest common ancestor (LCA), the stronger the correlation.



Correlations and tree distances

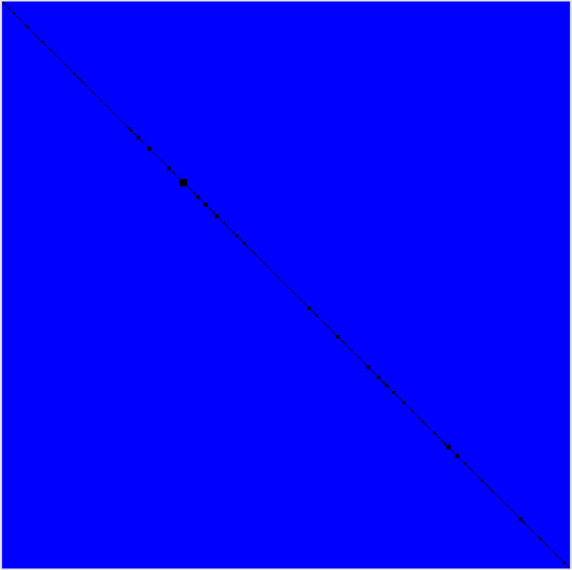
Correlations in population annealing are through the resampling, configurations with common ancestors are correlated. The closer the lowest common ancestor (LCA), the stronger the correlation.



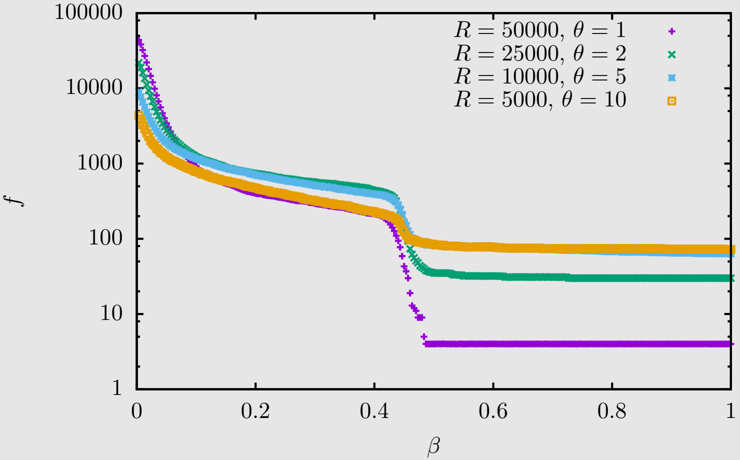
Correlations and tree distances

Correlations in population annealing are through the resampling, configurations with common ancestors are correlated. The closer the lowest common ancestor (LCA), the stronger the correlation.

Resampling correlates replicas, spin flips decorrelate them again.



Families



100000
10000
1000
100
10
1
f
0
0.2
0.4
0.6
0.8
1
β
$R = 50000,\ \theta = 1$
$R = 25000,\ \theta = 2$
$R = 10000,\ \theta = 5$
$R = 5000,\ \theta = 10$

Error analysis

Correlations decay with the distance in replica space $|i-j|$, so we can use methods of time series analysis (binning) to extract the effective number of independent samples.

Population annealing and MCMC

The general behavior of Markov chain Monte Carlo is well understood: if balance and ergodicity are fulfilled, it relaxes into equilibrium exponentially,

$$A_t \sim \langle A \rangle (1 + a e^{-t/t_{\mathrm{relax}}}).$$

In equilibrium, the Markov property results in correlations of successive measurements (in contrast to simple sampling):

$$C_{\Delta t} = \langle A_t A_{t+\Delta t} \rangle - \langle A_t \rangle \langle A_{t+\Delta t} \rangle \sim e^{-\Delta t/\tau_{\mathrm{exp}}}.$$

Hence, while for N independent measurements the standard deviation $\sigma(\bar{A})$ of the mean $\bar{A} = \sum_t A_t$ ("error bar") is given by

$$\sigma^2_{\mathrm{uncorr}}(\bar{A}) = \frac{\sigma^2(A)}{N},$$

in the presence of correlations we find instead

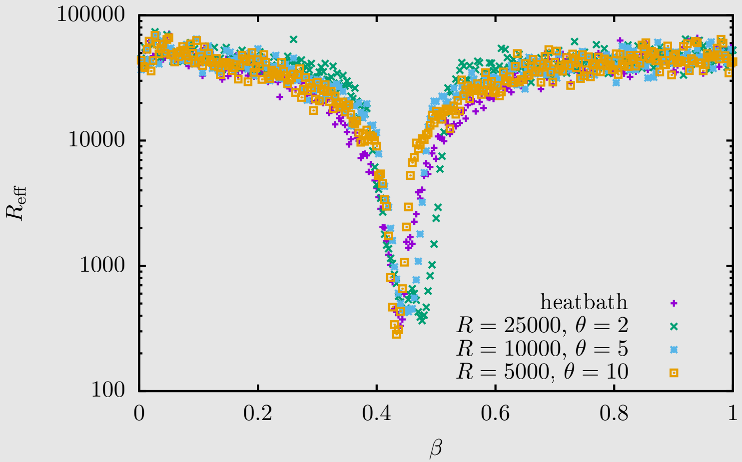
$$\sigma^2(\bar{A}) = \frac{\sigma^2(A)}{N_{\mathrm{eff}}}, \quad N_{\mathrm{eff}} = N/2\tau_{\mathrm{int}},$$

where

$$\tau_{\mathrm{int}} = \frac{1}{2} + \sum_{\Delta t=1}^{N-1} \frac{C_{\Delta t}}{C_0} \left(1 - \frac{\Delta t}{N}\right).$$

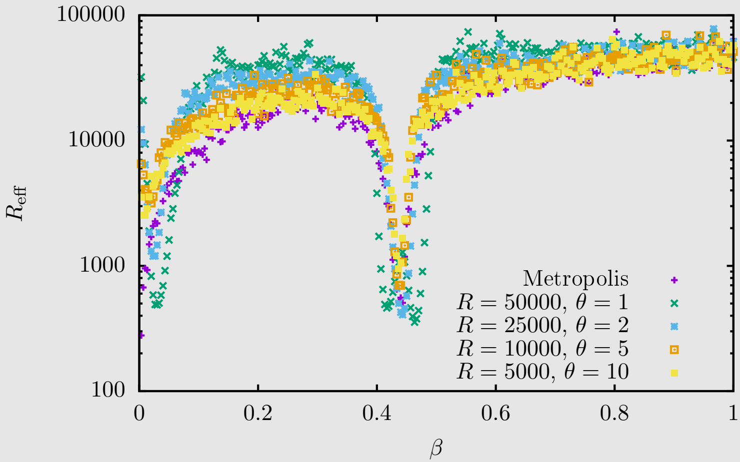
Error analysis

Correlations decay with the distance in replica space $|i - j|$, so we can use methods of time series analysis (binning) to extract the effective number of independent samples.



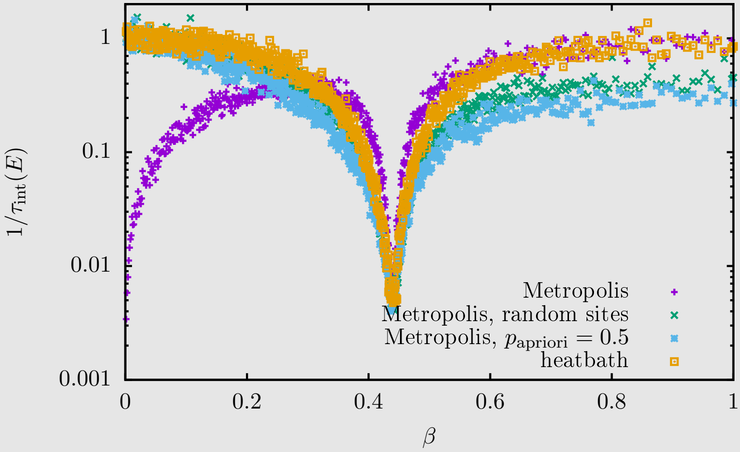
Error analysis

Correlations decay with the distance in replica space $|i - j|$, so we can use methods of time series analysis (binning) to extract the effective number of independent samples.



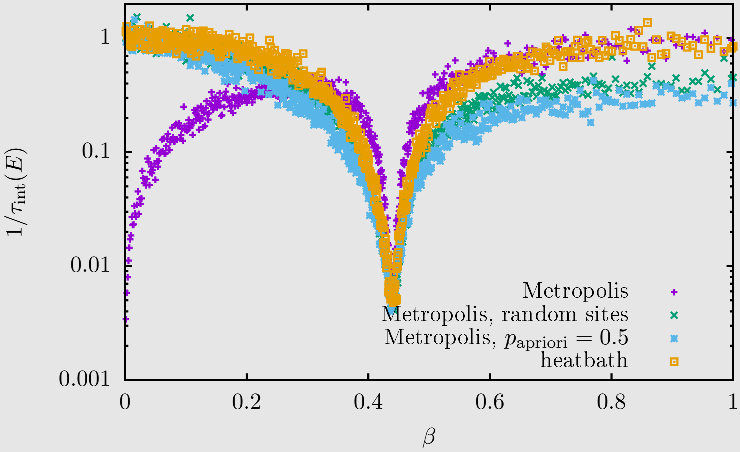
An aside: MCMC and ergodicity

Autocorrelation times for variants of Metropolis and heatbath dynamics.



An aside: MCMC and ergodicity

Autocorrelation times for variants of Metropolis and heatbath dynamics.



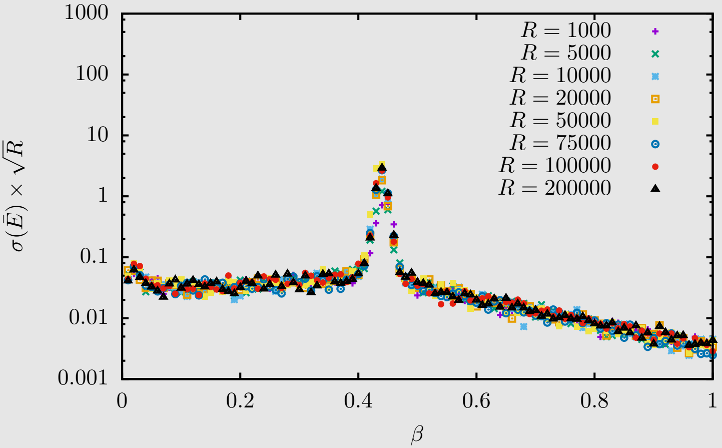
Sequential Metropolis update is not ergodic for $\beta \to 0$!

Statistical errors

Statistical errors decrease $\propto 1/\sqrt{R}$.

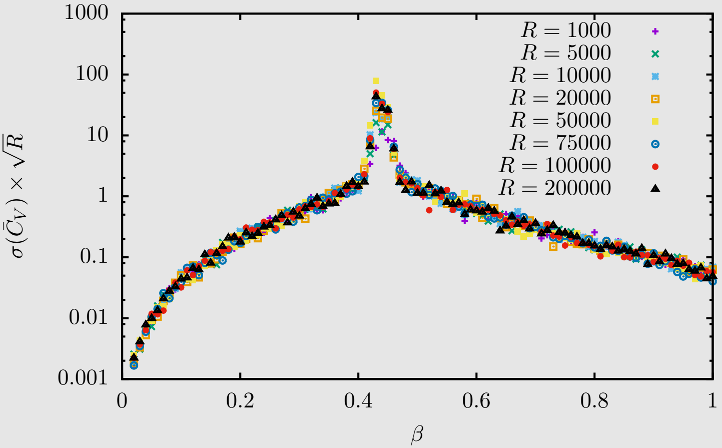
Statistical errors

Statistical errors decrease $\propto 1/\sqrt{R}$.



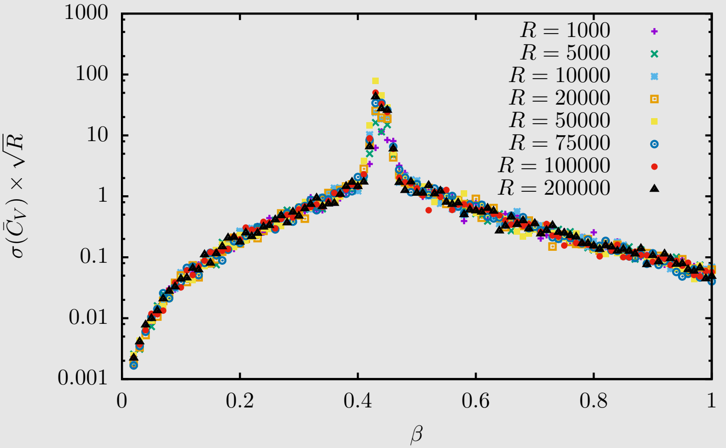
Statistical errors

Statistical errors decrease $\propto 1/\sqrt{R}$.



Statistical errors

Statistical errors decrease $\propto 1/\sqrt{R}$.



Also, we can show that the effective population size $R_{\mathrm{eff}} = R[1 - \exp(-\theta/\tau)]$ and $R_{\mathrm{eff}} \sim 1/\Delta\beta$.

Bias: no resampling

Consider a situation *without resampling*.

Bias: no resampling

Consider a situation *without resampling*. If the population is in equilibrium at β, after a step to $\beta + \Delta\beta$ it relaxes as

$$E(t) = \langle E \rangle_{\beta+\Delta\beta} - [\langle E \rangle_{\beta+\Delta\beta} - \langle E \rangle_{\beta}]e^{-t/\tau_{\mathrm{exp}}}$$

Bias: no resampling

Consider a situation *without resampling*. If the population is in equilibrium at β, after a step to $\beta+\Delta\beta$ it relaxes as

$$E(t) = \langle E\rangle_{\beta+\Delta\beta} - [\langle E\rangle_{\beta+\Delta\beta} - \langle E\rangle_{\beta}]e^{-t/\tau_{\mathrm{exp}}}$$

and

$$\langle E\rangle_{\beta+\Delta\beta} - \langle E\rangle_{\beta} \approx \frac{\partial\langle E\rangle}{\partial\beta}\Delta\beta = E'\Delta\beta = -\beta^2 V C_V \Delta\beta,$$

Bias: no resampling

Consider a situation *without resampling*. If the population is in equilibrium at β, after a step to $\beta + \Delta\beta$ it relaxes as

$$E(t) = \langle E \rangle_{\beta+\Delta\beta} - [\langle E \rangle_{\beta+\Delta\beta} - \langle E \rangle_{\beta}] e^{-t/\tau_{\rm exp}}$$

and

$$\langle E \rangle_{\beta+\Delta\beta} - \langle E \rangle_{\beta} \approx \frac{\partial \langle E \rangle}{\partial \beta} \Delta\beta = E' \Delta\beta = -\beta^2 V C_V \Delta\beta,$$

Hence the remaining bias after θ rounds of spin flips is

$$\Delta E = E(\theta) - \langle E \rangle_{\beta+\Delta\beta} \approx \beta^2 V C_V \Delta\beta e^{-\theta/\tau_{\rm exp}}.$$

Bias: no resampling

Consider a situation *without resampling*. If the population is in equilibrium at β, after a step to $\beta + \Delta\beta$ it relaxes as

$$E(t) = \langle E\rangle_{\beta+\Delta\beta} - [\langle E\rangle_{\beta+\Delta\beta} - \langle E\rangle_{\beta}]e^{-t/\tau_{\mathrm{exp}}}$$

and

$$\langle E\rangle_{\beta+\Delta\beta} - \langle E\rangle_{\beta} \approx \frac{\partial\langle E\rangle}{\partial\beta}\Delta\beta = E'\Delta\beta = -\beta^2 VC_V\Delta\beta,$$

Hence the remaining bias after θ rounds of spin flips is

$$\Delta E = E(\theta) - \langle E\rangle_{\beta+\Delta\beta} \approx \beta^2 VC_V\Delta\beta e^{-\theta/\tau_{\mathrm{exp}}}.$$

Taking biases from previous steps into account, one finds

$$\Delta E(\beta) \approx \sum_{j=1}^{n(\beta)} E'(\beta_0 + j\Delta\beta)\Delta\beta \, \exp\left[-\theta \sum_{k=j}^{n(\beta)} 1/\tau_k\right].$$

Bias: no resampling

Consider a situation *without resampling*. If the population is in equilibrium at β, after a step to $\beta + \Delta\beta$ it relaxes as

$$E(t) = \langle E \rangle_{\beta+\Delta\beta} - [\langle E \rangle_{\beta+\Delta\beta} - \langle E \rangle_{\beta}] e^{-t/\tau_{\mathrm{exp}}}$$

and

$$\langle E \rangle_{\beta+\Delta\beta} - \langle E \rangle_{\beta} \approx \frac{\partial \langle E \rangle}{\partial \beta} \Delta\beta = E' \Delta\beta = -\beta^2 V C_V \Delta\beta,$$

Hence the remaining bias after θ rounds of spin flips is

$$\Delta E = E(\theta) - \langle E \rangle_{\beta+\Delta\beta} \approx \beta^2 V C_V \Delta\beta e^{-\theta/\tau_{\mathrm{exp}}}.$$

Taking biases from previous steps into account, one finds

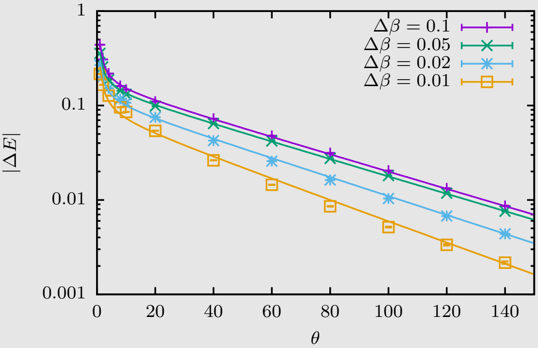
$$\Delta E(\beta) \approx \sum_{j=1}^{n(\beta)} E'(\beta_0 + j\Delta\beta)\Delta\beta \, \exp\left[-\theta \sum_{k=j}^{n(\beta)} 1/\tau_k\right].$$

For the simple case where all $\tau_i = \tau$ are equal and E' is independent of β, one finds

$$\Delta E(\beta) \approx E' \Delta\beta \frac{e^{-\theta/\tau}}{1 - e^{-\theta/\tau}} \left[1 - e^{-\frac{\theta\beta}{\tau\Delta\beta}}\right]$$

Bias: no resampling (cont'd)

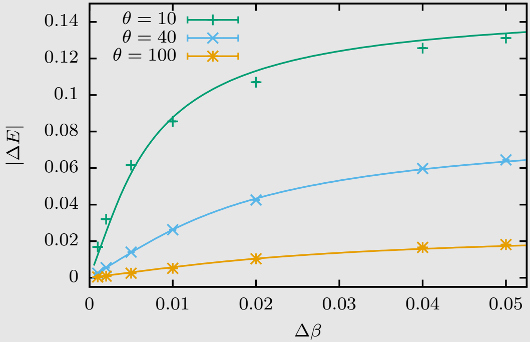
This is borne out very well by actual simulations.



$$\Delta E(\beta) \approx E' \Delta\beta \frac{e^{-\theta/\tau}}{1 - e^{-\theta/\tau}} \left[1 - e^{-\frac{\theta\beta}{\tau\Delta\beta}}\right]$$

Bias: no resampling (cont'd)

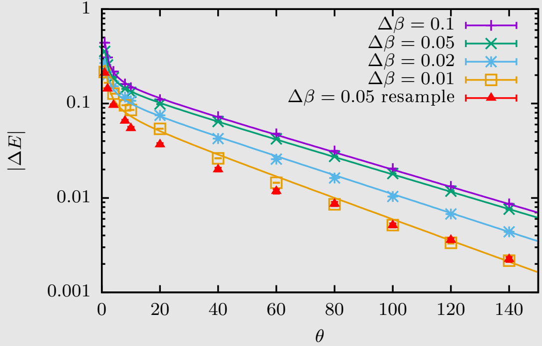
This is borne out very well by actual simulations.



$$\Delta E(\beta) \approx E' \Delta\beta \frac{e^{-\theta/\tau}}{1 - e^{-\theta/\tau}} \left[1 - e^{-\frac{\theta\beta}{\tau\Delta\beta}} \right]$$

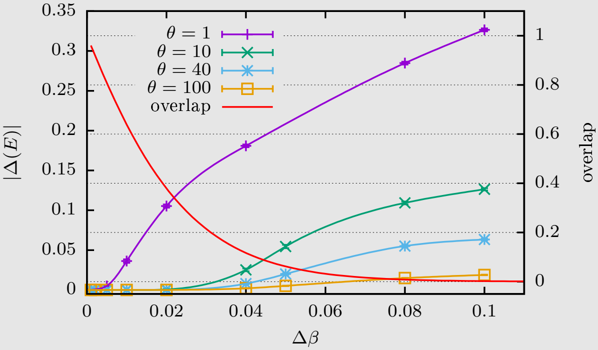
Bias: with resampling

When turning on resampling, the dependence on θ is essentially unchanged.



Bias: with resampling

When turning on resampling, the dependence on θ is essentially unchanged.



Bias is strongly reduced by resampling as soon as $\Delta\beta$ is such that the histogram overlap is $\gtrsim 0.1$.

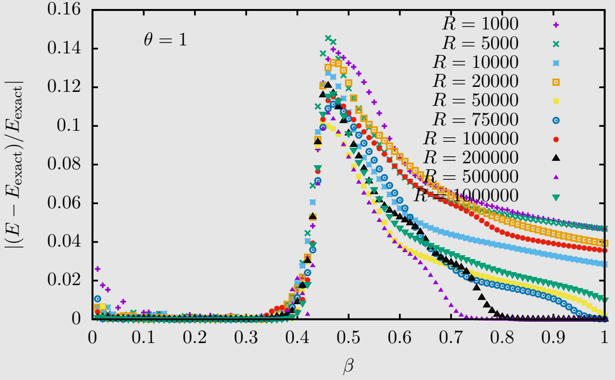
We can show analytically that additional resampling bias is $\propto \Delta\beta$.

Bias

Bias in population size R was argued to be $\sim 1/R$ (Machta + Ellis, 2011; Wang, Machta + Katzgraber, 2015). But consider

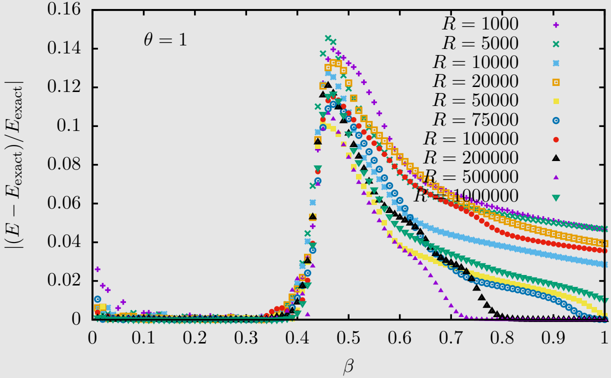
Bias

Bias in population size R was argued to be $\sim 1/R$ (Machta + Ellis, 2011; Wang, Machta + Katzgraber, 2015). But consider



Bias

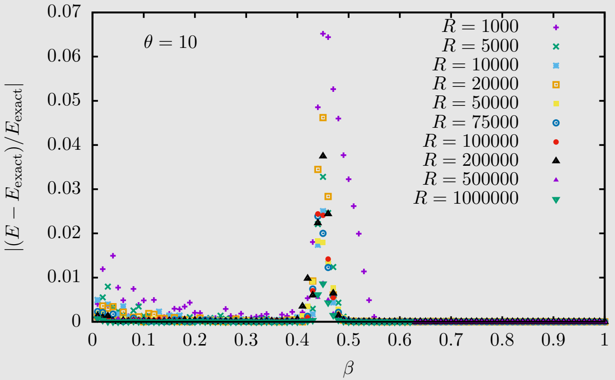
Bias in population size R was argued to be $\sim 1/R$ (Machta + Ellis, 2011; Wang, Machta + Katzgraber, 2015). But consider



Decay is slower than $1/R$ and, in fact, compatible with a logarithmically slow convergence. The relevant decay is proportional to $1/R_{\mathrm{eff}}$, and this is only proportional to R for $\theta \gtrsim \tau$.

Bias

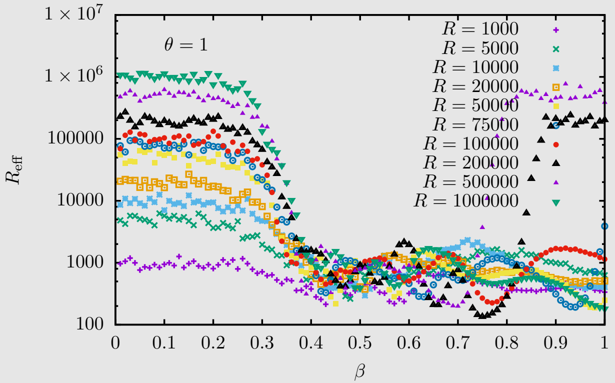
Bias in population size R was argued to be $\sim 1/R$ (Machta + Ellis, 2011; Wang, Machta + Katzgraber, 2015). But consider



Decay is slower than $1/R$ and, in fact, compatible with a logarithmically slow convergence. The relevant decay is proportional to $1/R_{\text{eff}}$, and this is only proportional to R for $\theta \gtrsim \tau$.

Bias

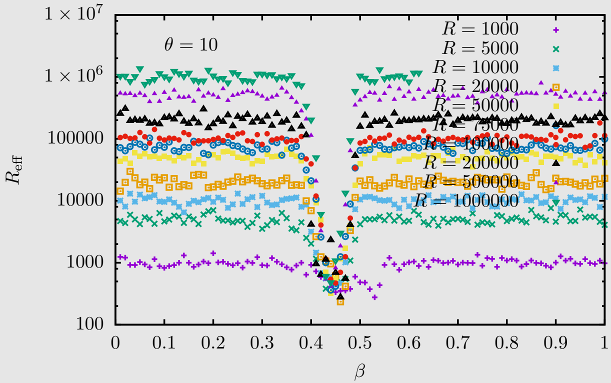
Bias in population size R was argued to be $\sim 1/R$ (Machta + Ellis, 2011; Wang, Machta + Katzgraber, 2015). But consider



Decay is slower than $1/R$ and, in fact, compatible with a logarithmically slow convergence. The relevant decay is proportional to $1/R_{\mathrm{eff}}$, and this is only proportional to R for $\theta \gtrsim \tau$.

Bias

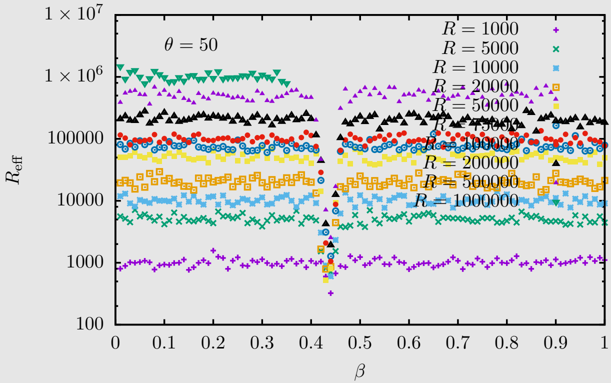
Bias in population size R was argued to be $\sim 1/R$ (Machta + Ellis, 2011; Wang, Machta + Katzgraber, 2015). But consider



Decay is slower than $1/R$ and, in fact, compatible with a logarithmically slow convergence. The relevant decay is proportional to $1/R_{\mathrm{eff}}$, and this is only proportional to R for $\theta \gtrsim \tau$.

Bias

Bias in population size R was argued to be $\sim 1/R$ (Machta + Ellis, 2011; Wang, Machta + Katzgraber, 2015). But consider



Decay is slower than $1/R$ and, in fact, compatible with a logarithmically slow convergence. The relevant decay is proportional to $1/R_{\mathrm{eff}}$, and this is only proportional to R for $\theta \gtrsim \tau$.

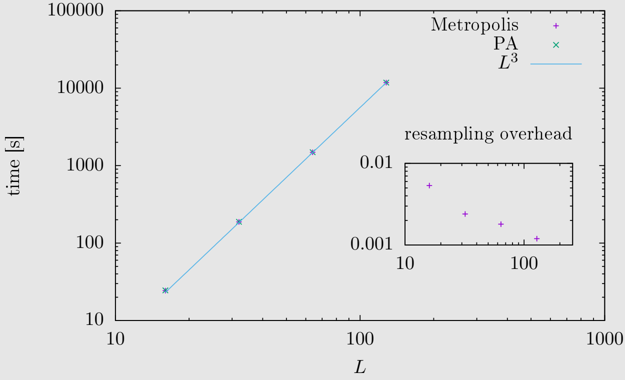
How good is it?

Computational efficiency

The actual computational overhead of population annealing over the underlying spin-flip dynamics is negligible.

Computational efficiency

The actual computational overhead of population annealing over the underlying spin-flip dynamics is negligible.

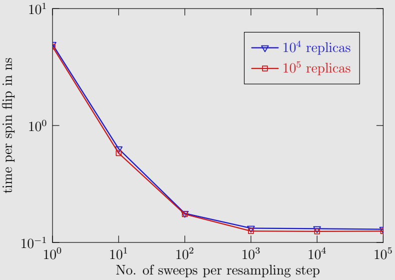


Massively parallel approach

The approach is naturally suitable for an implementation on massively parallel hardware such as GPUs.

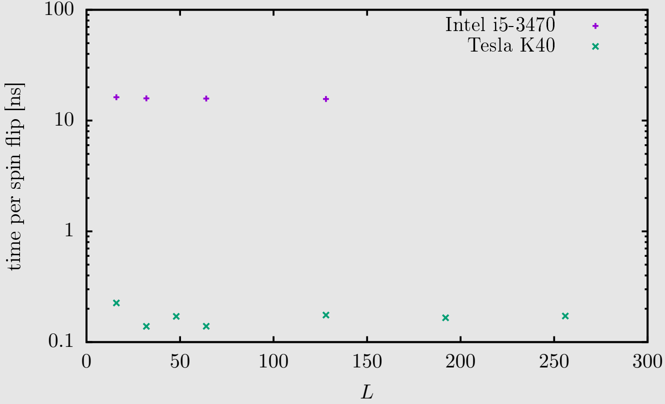
Massively parallel approach

The approach is naturally suitable for an implementation on massively parallel hardware such as GPUs.



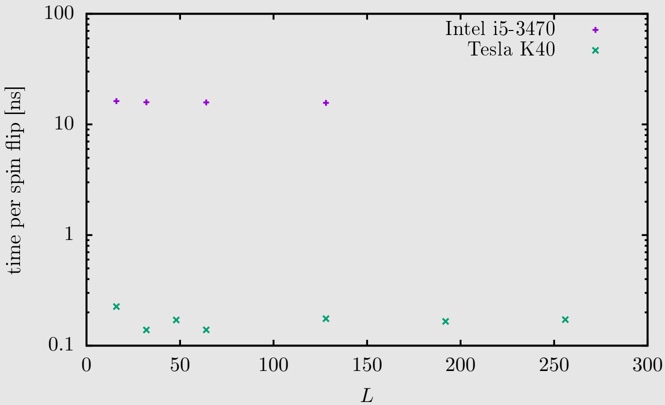
Massively parallel approach

The approach is naturally suitable for an implementation on massively parallel hardware such as GPUs.



Massively parallel approach

The approach is naturally suitable for an implementation on massively parallel hardware such as GPUs.



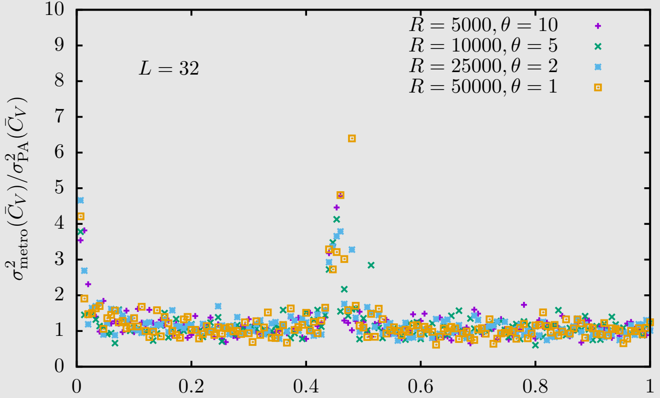
Approximately 100-fold speed-up! (Plus additional factor of 2–15 with multi-spin coding.)

Performance

Compare variance of averages measured in PA vs. those in an MCMC temperature sweep.

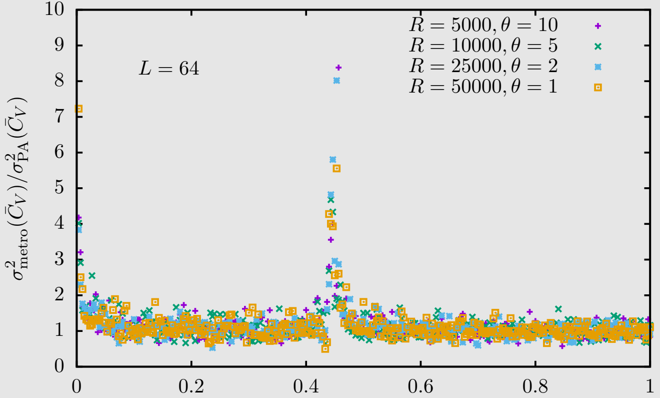
Performance

Compare variance of averages measured in PA vs. those in an MCMC temperature sweep.



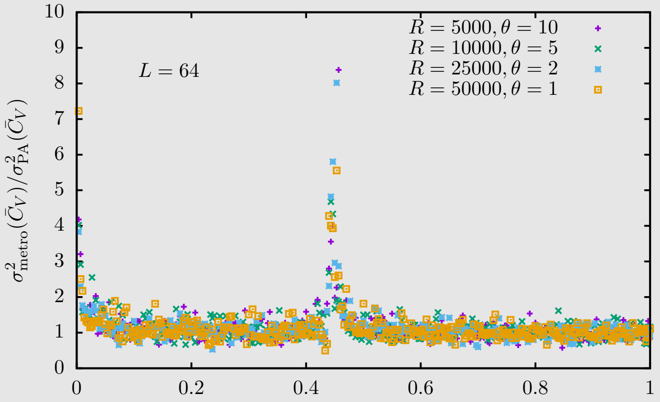
Performance

Compare variance of averages measured in PA vs. those in an MCMC temperature sweep.



Performance

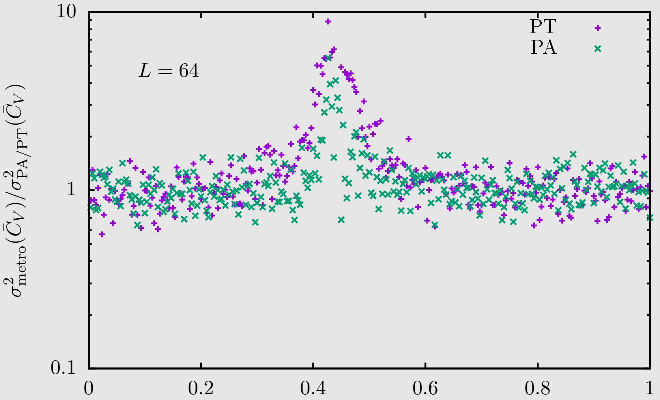
Compare variance of averages measured in PA vs. those in an MCMC temperature sweep.



How about parallel tempering?

Performance

Compare variance of averages measured in PA vs. those in an MCMC temperature sweep.



How about parallel tempering?

Let's make it even better

Improvements

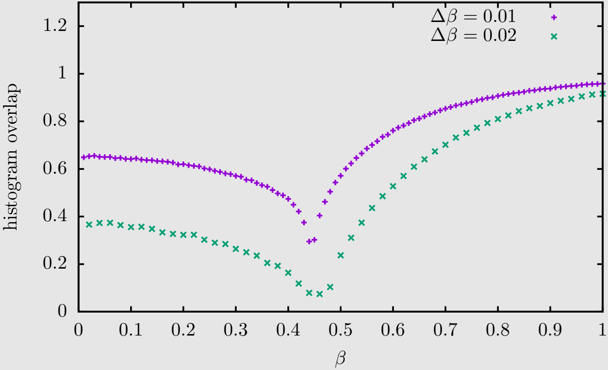
Three natural extensions that improve the algorithm significantly:

1. **Adaptive temperature steps:** Efficiency and bias of the resampling depends on histogram overlap.

Improvements

Three natural extensions that improve the algorithm significantly:

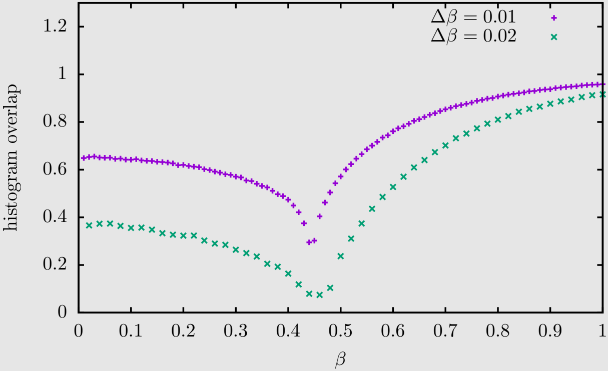
1. **Adaptive temperature steps:** Efficiency and bias of the resampling depends on histogram overlap.



Improvements

Three natural extensions that improve the algorithm significantly:

1. **Adaptive temperature steps:** Efficiency and bias of the resampling depends on histogram overlap.

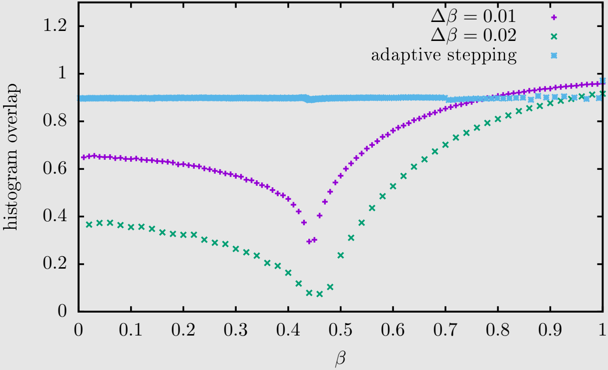


$\Rightarrow$ choose temperature step adaptively on the fly to ensure fixed overlap of neighboring energy histograms (as estimated from populations).

Improvements

Three natural extensions that improve the algorithm significantly:

1. **Adaptive temperature steps:** Efficiency and bias of the resampling depends on histogram overlap.



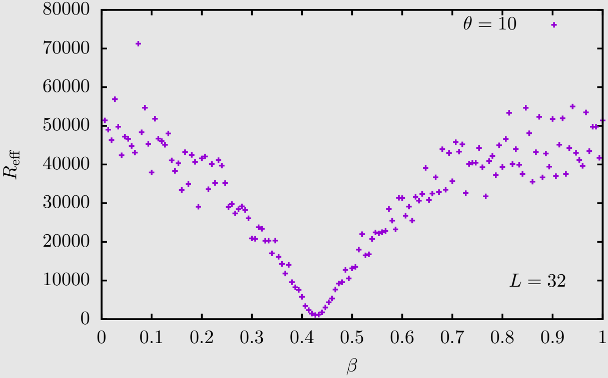
$\Rightarrow$ choose temperature step adaptively on the fly to ensure fixed overlap of neighboring energy histograms (as estimated from populations).

Improvements

2. **Adaptive time steps:** Number of independent replicas R_{eff} crucially determines bias as well as statistical errors.

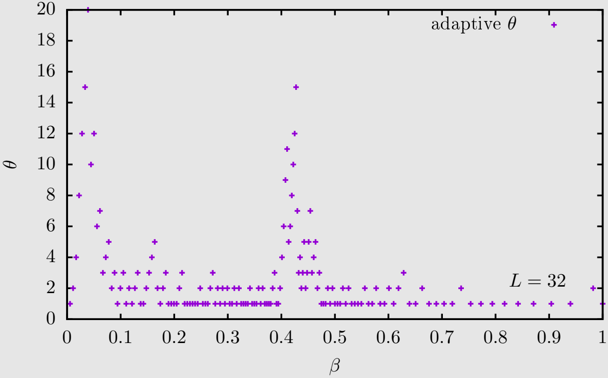
Improvements

2. **Adaptive time steps:** Number of independent replicas R_{eff} crucially determines bias as well as statistical errors.



Improvements

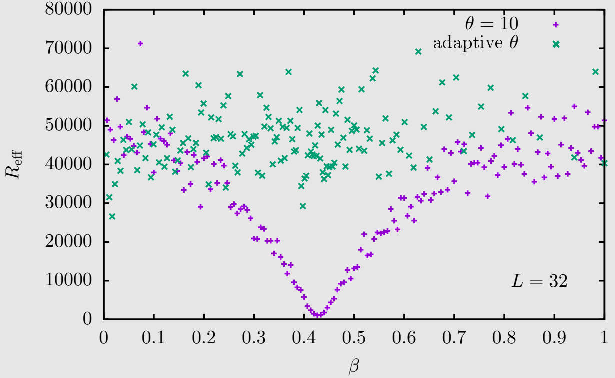
2. **Adaptive time steps:** Number of independent replicas R_{eff} crucially determines bias as well as statistical errors.



$\Rightarrow$ choose $\theta \propto R/R_{\text{eff}}$ to effectively decorrelate configurations.

Improvements

2. **Adaptive time steps:** Number of independent replicas R_{eff} crucially determines bias as well as statistical errors.



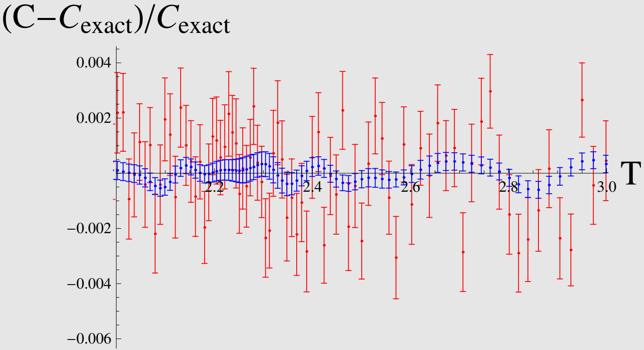
$\Rightarrow$ choose $\theta \propto R/R_{\text{eff}}$ to effectively decorrelate configurations.

Improvements

3. **Multi-histogram analysis:** Information from neighboring temperatures is also relevant.

Improvements

3. **Multi-histogram analysis:** Information from neighboring temperatures is also relevant.



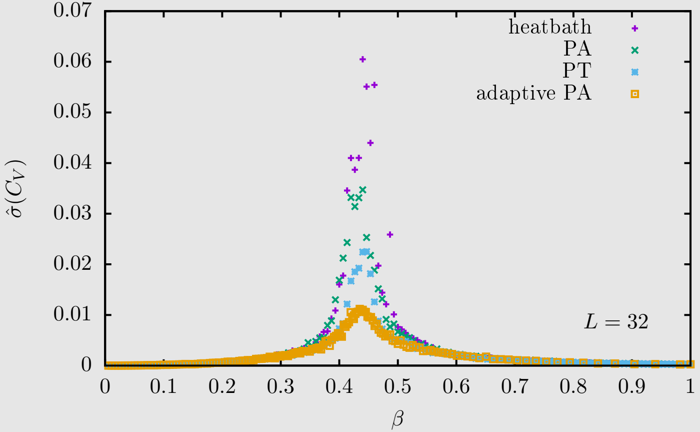
This also allows to estimate the density of states. Iterations as in the Ferrenberg/Swendsen scheme are not required.

Comparison

Adaptive scheme performs significantly better than original one.

Comparison

Adaptive scheme performs significantly better than original one.



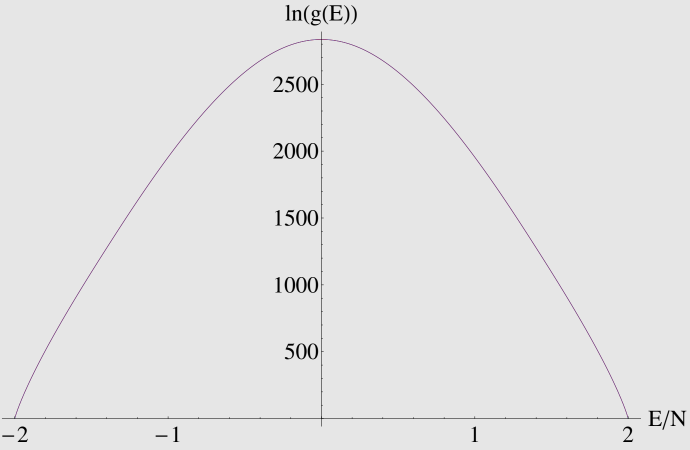
Applications?

Sampling the density of states

Something that we normally think can only be done with multicanonical or Wang-Landau simulations.

Sampling the density of states

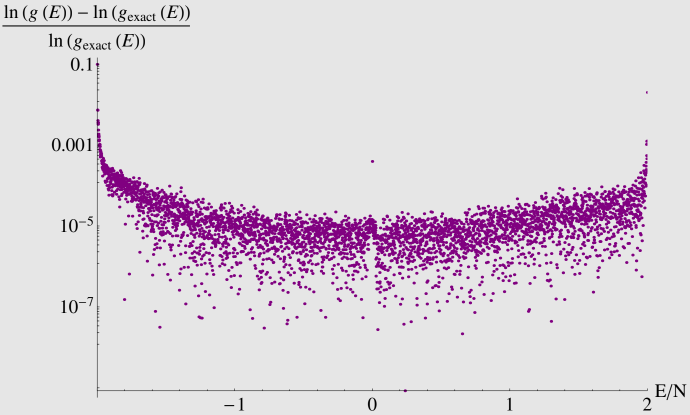
Something that we normally think can only be done with multicanonical or Wang-Landau simulations.



$L = 64$

Sampling the density of states

Something that we normally think can only be done with multicanonical or Wang-Landau simulations.



$L = 64$

Spin glasses

Population annealing has already been used quite successfully to simulate the 3D Edwards-Anderson model (Wang, Machta + Katzgraber, 2014; 2015).

Spin glasses

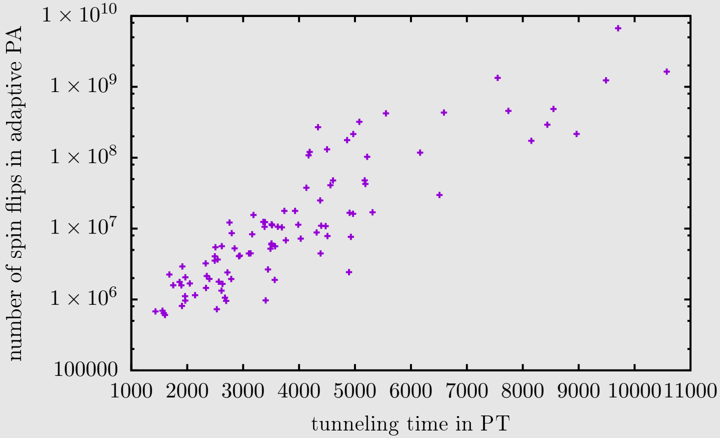
Population annealing has already been used quite successfully to simulate the 3D Edwards-Anderson model (Wang, Machta + Katzgraber, 2014; 2015).

With the adaptive scheme, this should even work significantly better: tailored for systems with a wide distribution of hardness of instances.

Spin glasses

Population annealing has already been used quite successfully to simulate the 3D Edwards-Anderson model (`Wang, Machta + Katzgraber, 2014; 2015`).

With the adaptive scheme, this should even work significantly better: tailored for systems with a wide distribution of hardness of instances.



Off-lattice systems and polymers

Conclusions

Main points:

- naturally suited for massively parallel architectures
- can estimate free energies and density of states with high precision
- can be easily turned into a fully self-adaptive algorithm

Technical insights:

- raw family numbers are not so useful
- can calculate statistical errors from one simulation
- bias is asymptotically

$$\Delta A \propto \frac{\Delta\beta}{R_{\mathrm{eff}}} \exp(-\theta/\tau_{\mathrm{eff}})$$

- hence bias decays more slowly with computational effort R than for MCMC, but this does not matter in most cases as statistical errors $\propto 1/\sqrt{R}$ dominate
- advantage over PT: ballistic movement through temperature space