Wang-Landau Monte Carlo simulation

Aleš Vítek
IT4I, VP3

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

PART 1

Classical Monte Carlo Methods

Outline

1. Statistical thermodynamics, ensembles
2. Numerical evaluation of integrals, crude Monte Carlo (MC)
3. MC methods for statistical thermodynamics, Markov chains
4. Variants of MC methods

Statistical thermodynamics

microscopic description

(positions, momenta, interactions)

Statistical thermodynamics

microscopic description

(positions, momenta, interactions)

macroscopic description

(pressure, density, temperature)

Statistical thermodynamics

microscopic description

(positions, momenta, interactions)

macroscopic description

(pressure, density, temperature)

Statistical Thermodynamics

Statistical thermodynamics

time averages → Molecular Dynamics methods

$$B = \bar{b} \equiv \lim_{\tau \to +\infty} \frac{1}{\tau} \int_{t_0}^{t_0+\tau} b(\vec{r}_K(t), \vec{p}_K(t))\, \mathrm{d}t$$

Statistical thermodynamics

ensemble averages → Monte Carlo methods

$$B\left(A_1,\ldots,A_n\right)=\langle b\rangle\equiv\frac{\int b(\vec{r}_k,\vec{p}_k)\rho(\vec{r}_k,\vec{p}_k;A_1,\ldots,A_n)\,\mathrm{d}\Gamma}{\int\rho(\vec{r}_k,\vec{p}_k;A_1,\ldots,A_n)\,\mathrm{d}\Gamma}$$

Statistical thermodynamics

Statistical (Gibbs) ensembles

- ✓ **microstate (μ-state): positions and momenta**
- ✓ **macrostate (m-state): macroscopic variables (A_1, …)**
- ✓ **1 m-state = many μ-states → Gibbs ensemble**
- ✓ **probability distribution in system's phase space,** $\rho(\vec{r}_K, \vec{p}_K; A_1, \ldots, A_n)$

Statistical thermodynamics

Examples of statistical (Gibbs) ensembles

- ✓ **micro-canonical:** N, V, E
- ✓ **canonical:** N, V, T
- ✓ **grand-canonical:** μ, V, T
- ✓ **isobaric-isothermic:** N, P, T
- ✓ **etc.**

Statistical thermodynamics

Canonical Gibbs ensemble

$$\rho\left(\vec{r}_K, \vec{p}_K; N, V, T\right) \propto \exp\left[-\frac{H_N(\vec{r}_K, \vec{p}_K; V)}{k_B T}\right]$$

$$H_N(\vec{r}_K, \vec{p}_K; V) = \sum_{K=1}^{N} \frac{\vec{p}_K^{\,2}}{2M_K} + W(\vec{r}_K; V)$$

Statistical thermodynamics

Statistical thermodynamics as a mathematical problem

$$B\left(A_1,\ldots,A_n\right)=\left\langle b\right\rangle \equiv \frac{\int b(\vec{r}_K,\vec{p}_K)\rho(\vec{r}_K,\vec{p}_K;A_1,\ldots,A_n)\,\mathrm{d}\Gamma}{\int \rho(\vec{r}_K,\vec{p}_K;A_1,\ldots,A_n)\,\mathrm{d}\Gamma}$$

$$\mathrm{d}\Gamma=\frac{1}{N!\mathrm{h}^{3N}}\mathrm{d}^3\vec{r}_1\,\mathrm{d}^3\vec{p}_1\ldots\mathrm{d}^3\vec{r}_N\,\mathrm{d}^3\vec{p}_N$$

Statistical thermodynamics

Statistical thermodynamics as a mathematical problem

$$B\left(A_1,\ldots,A_n\right) = B_{\text{kin}} + \frac{\int b_{\text{con}}(\vec{r}_K)\rho_{\text{con}}(\vec{r}_K;A_1,\ldots,A_n)\,\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N}{\int \rho_{\text{con}}(\vec{r}_K;A_1,\ldots,A_n)\,\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N}$$

Canonical ensemble

$$\boxed{\rho_{\text{con}}\left(\vec{r}_K;N,V,T\right) \propto \exp\left[-\frac{W_N(\vec{r}_K;V)}{k_{\text{B}}T}\right]}$$

Numerical evaluation of integrals

1D – <u>rectangular</u>, trapezoid, Simpsonian, …

$$\int_a^b f(x)\,\mathrm{d}x \approx \frac{b-a}{n}\sum_{j=1}^{n} f\left(a+\frac{b-a}{n}\cdot j\right) = \frac{b-a}{n}\sum_{j=1}^{n} f(x_j)$$

Numerical evaluation of integrals

1D – <u>rectangular</u>, trapezoid, Simpsonian, ...

$$\int_a^b f(x)\,\mathrm{d}x \approx \frac{b-a}{n}\sum_{j=1}^{n} f\left(a+\frac{b-a}{n}\cdot j\right) = \frac{b-a}{n}\sum_{j=1}^{n} f(x_j)$$

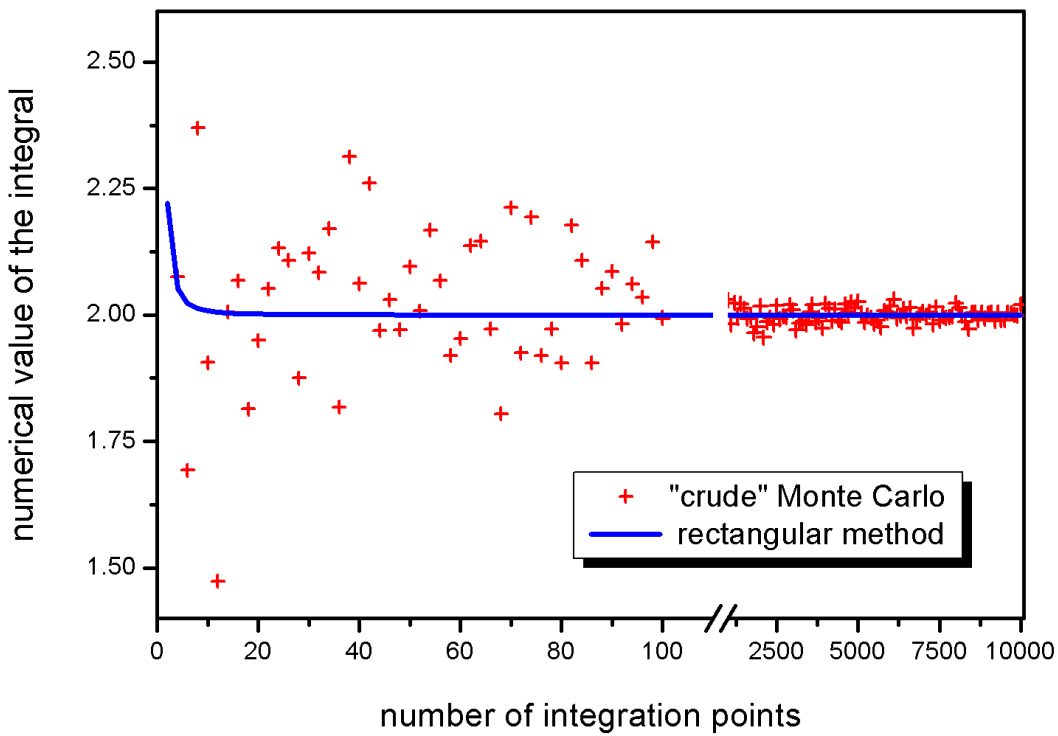
1D – <u>crude Monte Carlo</u>

$$\int_a^b f(x)\,\mathrm{d}x \approx \frac{b-a}{n}\sum_{j=1}^{n} f(x_j)$$

x_j randomly distributed over [a,b]

Numerical evaluation of integrals

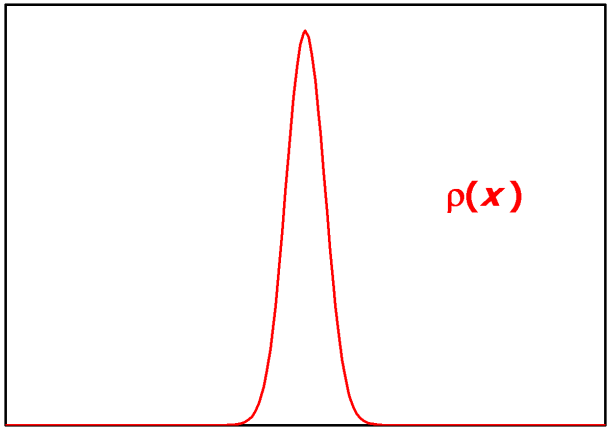
Example $\int_0^{\pi} \sin x \, dx$



Numerical evaluation of integrals

Localized functions

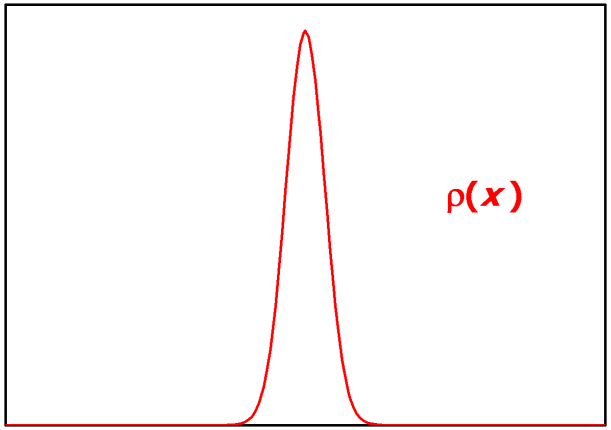
$$\int_a^b f(x)\rho(x)\mathrm{d}x$$



Numerical evaluation of integrals

<u>Localized functions</u>

$$\int_a^b f(x)\rho(x)\mathrm{d}x$$

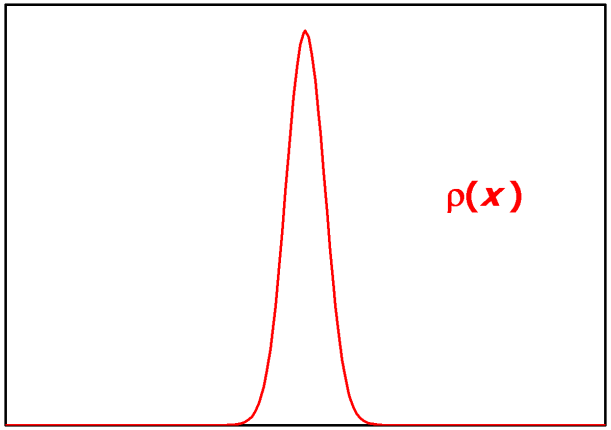


<u>Solution</u>: Sampling from $\rho(x)$.

Numerical evaluation of integrals

Localized functions

$$\int_a^b f(x)\rho(x)\mathrm{d}x$$



Solution: Sampling from $\rho(x)$.

How? Markov chains.

Markov chains

<u>Simplified example – finite number of microstates</u>

- ✓ microstates: $s_1, \ldots, s_n$
- ✓ probability distribution: $\rho_1 = \rho(s_1), \ldots, \rho_n = \rho(s_n)$
- ✓ equilibrium PD: $\rho_1^{(e)}, \ldots, \rho_n^{(e)}$

Markov chains

Simplified example – finite number of microstates

- ✓ Markov chain: $s^{(1)} \rightarrow s^{(2)} \rightarrow \ldots \rightarrow s^{(N)} \rightarrow \ldots$

  $s^{(K)}$ is from $\{s_1, \ldots, s_n\}$

- ✓ transition probabilities: $\pi_{ik} = \pi_{i \rightarrow k}$
  (stochastic matrix)

$$0 \leq \pi_{ik} \leq 1$$

$$\sum_{k=1}^{n} \pi_{ik} = 1$$

Markov chains

Simplified example – finite number of microstates

- ✓ PD function evolution:

$$\rho_k^{(N+1)} = \sum_{j=1}^{n} \rho_i^{(N)} \pi_{ik} \qquad \vec{\rho}^{(N+1)} = \vec{\rho}^{(N)} \overleftrightarrow{\pi}$$

- ✓ equilibrium PD function:

$$\rho_k^{(\mathrm{e})} = \sum_{j=1}^{n} \rho_i^{(\mathrm{e})} \pi_{ik} \qquad \vec{\rho}^{(\mathrm{e})} = \vec{\rho}^{(\mathrm{e})} \cdot \overleftrightarrow{\pi}$$

Markov chains

Simplified example – finite number of microstates

Theorem:

Let $s^{(1)}$, $s^{(2)}$, …, $s^{(N)}$, … be an infinite Markov chain of microstates generated using a stochastic matrix π_{ik} . Then the generated microstates are distributed over the state space of the system in congruence with the equilibrium PD function of the used stochastic matrix.

Remarks:

The Markov chain does not represent any physical time evolution.

From the statistical thermodynamics point of view, only the equilibrium PD function is physically relevant. (Not, *e. g.*, the stochastic matrix.)

Markov chains

Construction of an appropriate stochastic matrix

$$\rho_k^{(e)} = \sum_{j=1}^{n} \rho_j^{(e)} \pi_{jk}, \qquad \sum_{k=1}^{n} \pi_{jk} = 1$$

Mathematically

- ✓ **2n linear algebraic equations for n^2 variables**
- ✓ **many (infinite number of) solutions**

Markov chains

Construction of an appropriate stochastic matrix

$$\rho_k^{(e)} = \sum_{j=1}^{n} \rho_j^{(e)} \pi_{jk}, \quad \sum_{k=1}^{n} \pi_{jk} = 1$$

Mathematically

- ✓ **2n linear algebraic equations for n^2 variables**
- ✓ **many (infinite number of) solutions**

Detailed balance

$$\rho_k^{(e)} \pi_{kj} = \rho_j^{(e)} \pi_{jk}$$

Markov chains

Decomposition of the stochastic matrix

$$\pi_{ik} = p_{ik} a_{ik}$$

p_{ik} ... probability of **proposing** a particular transition

a_{ik} ... probability of **accepting** the proposed transition

Metropolis solution [1]

$$\rho_k^{(e)} p_{ki} a_{ki} = \rho_i^{(e)} p_{ik} a_{ik} \quad \Rightarrow \quad a_{ik} = \min\left\{1, \frac{\rho_k^{(e)}}{\rho_i^{(e)}} \frac{p_{ki}}{p_{ik}}\right\} \quad \overset{p_{ki}=p_{ik}}{\Rightarrow} \quad a_{ik} = \min\left\{1, \frac{\rho_k^{(e)}}{\rho_i^{(e)}}\right\}$$

[1] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087.

Markov chains

Decomposition of the stochastic matrix

$$\pi_{ik} = p_{ik} a_{ik}$$

p_{ik} ... probability of **proposing** a particular transition

a_{ik} ... probability of **accepting** the proposed transition

Metropolis solution [1]

$$\rho_k^{(e)} p_{ki} a_{ki} = \rho_i^{(e)} p_{ik} a_{ik} \quad \Rightarrow \quad a_{ik} = \min\left\{1, \frac{\rho_k^{(e)}}{\rho_i^{(e)}} \frac{p_{ki}}{p_{ik}}\right\} \quad \overset{p_{ki}=p_{ik}}{\Rightarrow} \quad a_{ik} = \min\left\{1, \frac{\rho_k^{(e)}}{\rho_i^{(e)}}\right\}$$

Notice that the PD function need not be normalized!

[1] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087.

Metropolis algorithm

… for evaluating configuration integrals of statistical thermodynamics (canonical ensemble)

$$B = B_{\text{kin}} + \frac{\int b_{\text{con}}(\vec{r}_K)\,\mathrm{e}^{-\frac{W(\vec{r}_K)}{k_{\mathrm{B}}T}}\,\mathrm{d}\vec{r}_1 \ldots \mathrm{d}\vec{r}_N}{\int \mathrm{e}^{-\frac{W(\vec{r}_K)}{k_{\mathrm{B}}T}}\,\mathrm{d}\vec{r}_1 \ldots \mathrm{d}\vec{r}_N}$$

and, consequently,

$$\rho_i^{(\mathrm{e})} \propto \mathrm{e}^{-\frac{W(\vec{r}_K^{(i)})}{k_{\mathrm{B}}T}} \equiv \mathrm{e}^{-\frac{W^{(i)}}{k_{\mathrm{B}}T}} \quad \Rightarrow \quad \frac{\rho_k^{(\mathrm{e})}}{\rho_i^{(\mathrm{e})}} = \mathrm{e}^{-\frac{\left(W^{(k)} - W^{(i)}\right)}{k_{\mathrm{B}}T}} = \mathrm{e}^{-\frac{\Delta W^{(i \to k)}}{k_{\mathrm{B}}T}}$$

Metropolis algorithm

- ✓ from the configuration generated in the preceding step, $r^{(i)}$, generate a new one using an algorithm obeying that the probability for old → new is the same as for new → old (usually isotropic translation and/or rotation)
- ✓ if $W^{(\text{new})} < W^{(i)}$ then $r^{(i+1)} = r^{(\text{new})}$
- ✓ if $W^{(\text{new})} > W^{(i)}$ then
  - ✓ $r^{(i+1)} = r^{(\text{new})}$ with probability $e^{-\Delta W^{(i\to\text{new})}/k_B T}$
  - ✓ $r^{(i+1)} = r^{(i)}$ with probability $1 - e^{-\Delta W^{(i\to\text{new})}/k_B T}$
- ✓ repeat until "infinity" (convergence)

Metropolis algorithm

Then

$$\boxed{\frac{\int b_{\mathrm{con}}(\vec{r}_K)\,\mathrm{e}^{-\frac{W(\vec{r}_K)}{k_{\mathrm{B}}T}}\,\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N}{\int \mathrm{e}^{-\frac{W(\vec{r}_K)}{k_{\mathrm{B}}T}}\,\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N} = \lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n} b_{\mathrm{con}}(\vec{r}_K^{\,(i)})}$$

Metropolis algorithm

Then

$$\boxed{\frac{\int b_{\mathrm{con}}(\vec{r}_K)\,\mathrm{e}^{-\frac{W(\vec{r}_K)}{k_{\mathrm{B}}T}}\,\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N}{\int \mathrm{e}^{-\frac{W(\vec{r}_K)}{k_{\mathrm{B}}T}}\,\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N} = \lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n} b_{\mathrm{con}}(\vec{r}_K^{\,(i)})}$$

Canonical (or NVT or isochoric-isothermic)

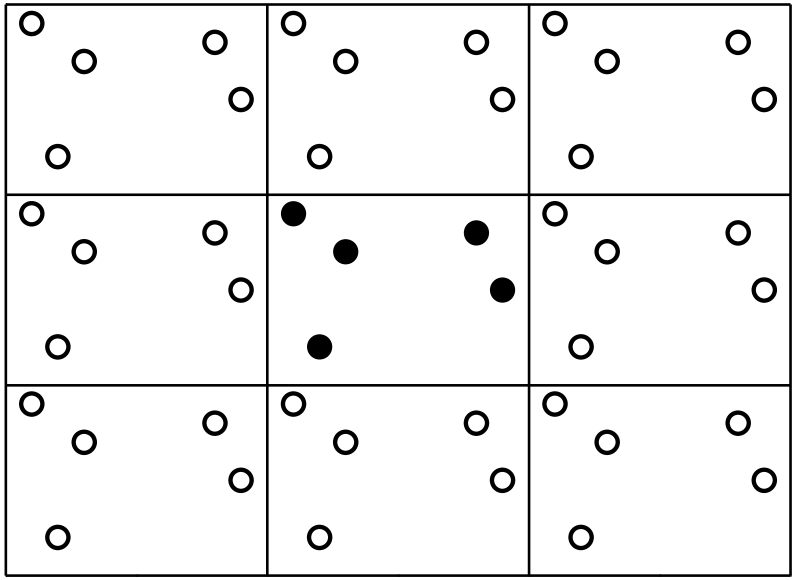
Monte Carlo

A general sketch of an MC simulation

- ✓ **propose an initial configuration**
- ✓ **employ the Metropolis algorithm to generate as many as possible further configurations**
- ✓ **collect necessary data along the MC path to evaluate ensemble averages**
- ✓ **calculate the averages as simple arithmetic means of the collected data**

Troublesomes

Periodic boundary conditions - example



Literature

M. LEWERENZ, *Monte Carlo Methods: Overview and Basics*, in NIC Series Volume 10: *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, NIC Juelich 2002 (http://www.fz-juelich.de/nic-series/volume10/volume10.html)

M. P. ALLEN, P. J. TILDESLEY, *Computer Simulation of Liquids*, Oxford University Press, Oxford 1987.

D. P. LANDAU, K. BINDER, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, Cambridge 2000 and 2005.

Appendix

Simulated annealing [1] and basin hopping [2]

MC simulation at a high temperature

$T \rightarrow 0$ K **simulated annealing**

gradient optimization **basin hopping**

equilibrium structure for $T = 0$ K

[1] S. Kirckpatrick, C. D. Gellat, M. P. Vecchi, Science 220 (1983) 671; [2] D. J. Wales, H. A. Scheraga, Science 285 (1999) 1368.

Example 2

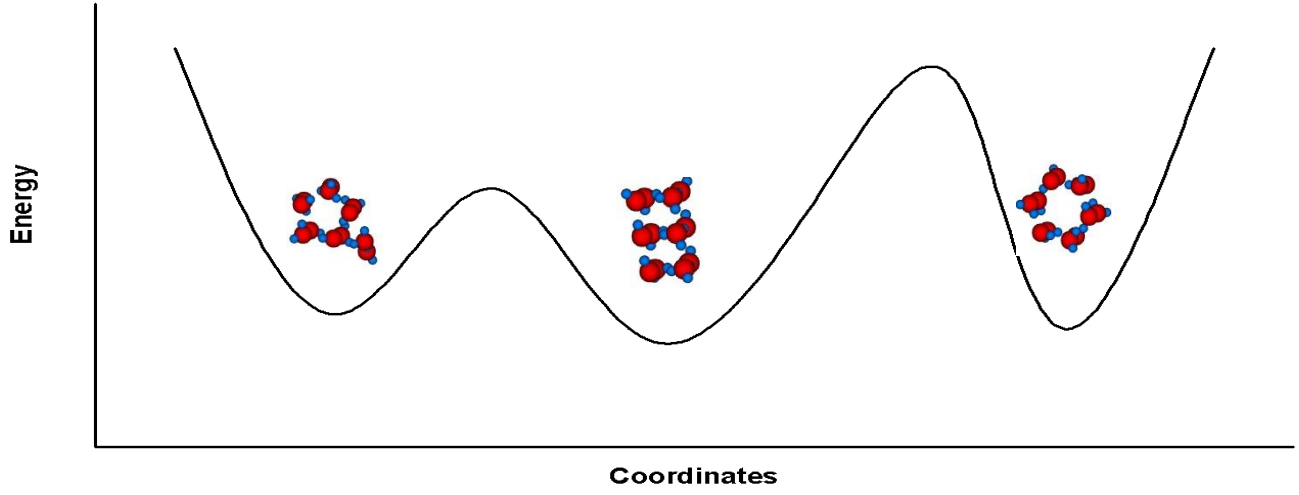
Simulated annealing of small water clusters.

PART 2

Wang-Landau algorithm

Problem: low temperatures

- ✓Simulation at low temperatures – combination of integration and optimization
- ✓Molecular clusters, biomolecules:
- ✓Complicated energy landscapes
- ✓A lot of minims separated by large energy barriers



Mean value of variable B

$$B\left(A_1,\ldots,A_n\right) = B_{\text{kin}} + \frac{\int b_{\text{con}}(\vec{r}_K)\rho_{\text{con}}(\vec{r}_K;A_1,\ldots,A_n)\,\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N}{\int \rho_{\text{con}}(\vec{r}_K;A_1,\ldots,A_n)\,\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N}$$

IF $b_{\text{con}}(\vec{r}_K) \equiv b\left(W(\vec{r}_K)\right)$ **, then mean value of** B **can be expressed as a 1D integral**

$$B\left(A_1,\ldots,A_n\right) = B_{\text{kin}} + \frac{\int_{W_{\min}}^{\infty} b_{\text{con}}(W)\rho_{\text{con}}(W;A_1,\ldots,A_n)\,\Omega(W)\mathrm{d}W}{\int_{W_{\min}}^{\infty} \rho_{\text{con}}(W;A_1,\ldots,A_n)\,\Omega(W)\mathrm{d}W}$$

where $\Omega(W)$ **is classical density of states**

$$\Omega(W) = \lim_{\Delta W\to 0} \frac{\displaystyle\int_{\vec{r}_1,\ldots,\vec{r}_N;W(\vec{r}_1,\ldots,\vec{r}_N)\in\langle W,W+\Delta W\rangle} 1\;\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N}{\displaystyle\int_{\vec{r}_1,\ldots,\vec{r}_N} 1\;\mathrm{d}\vec{r}_1\ldots\mathrm{d}\vec{r}_N}$$

How to calculate $\Omega(W)$ **? Wang-Landau algorithm**

Calculation of Ω Wang-Landau algorithm – basic idea

- Completaly random walk in configuration space => energy histogram h(W)~Ω(W), but low energy configurations will not be visited

- Metropolis algorithm: configurations are sampled by $\rho(E_i(r)$, => energy histogram h(W)~Ω(W) x $\rho(W(r))$

- If we use $\rho(W_i(r)) = 1/(\Omega(W))$, then energy histogram will be constant function

Wang-Landau algorithm – calculation of density of states

- Start of simulation: $\Omega(W)$ is unknown, we start with constant $\Omega(W) = 1$ and from random initial configuration r
- Following configuration r_2 created by small random change of foregoing configuration r1 is accepted with probability

$$\min\left\{1, \frac{\rho\left(W_{new}\left(r_{new}\right)\right)}{\rho\left(W_{old}\left(r_{old}\right)\right)}\right\} = \left\{1, \frac{\Omega\left(W_{old}\right)}{\Omega\left(W_{new}\right)}\right\}$$

- Visiting of energy W_i: change of $\Omega(W_i) \rightarrow \Omega(W_i) \times k_0$, where modification factor $k_0 > 1$.
- *e. g.,* $k_0 = 2$, big k → big statistical errors, small k → long simulation
- After reaching of flat energy histogram $h(W)$, modification factor is changed e.g. by rule $k_{j+1} = \sqrt{k_j}$, energy histogram is deleted (density of states isn't deleted!) and we start a new simulation
- Computations is finished, when k reaches predeterminated value (e. g. 1,000 001)

Wang-Landau algorithm – accuracy of results depends on:

- Final value of modification factor k_{final}
- Softness of deals of energy
- Criterion of flatness of histogram
- Complexity of simulated system
- Details of implemented algorithm

Wangův-Landau algorithm: parallelization

- MPI, each core has one own walker, all cores share actual density of states and actual energy histogram

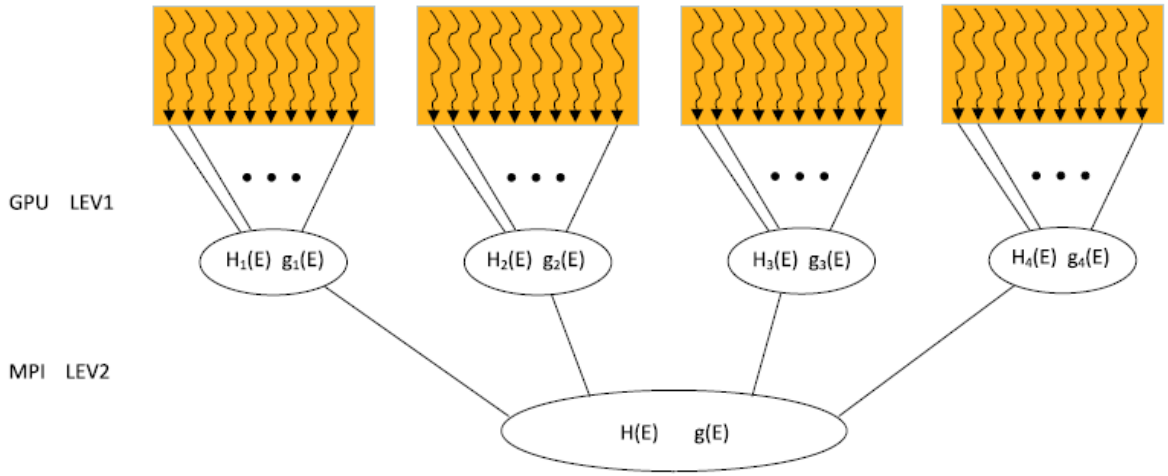


Fig. 1. Schematic plot of a four GPU implementation of the parallel Wang–Landau algorithm, where $H(E)$ is the energy histogram and $g(E)$ is the density of states. Each rectangle stands for a GPU card which contains thousands of threads.

W-L algorithm-example:

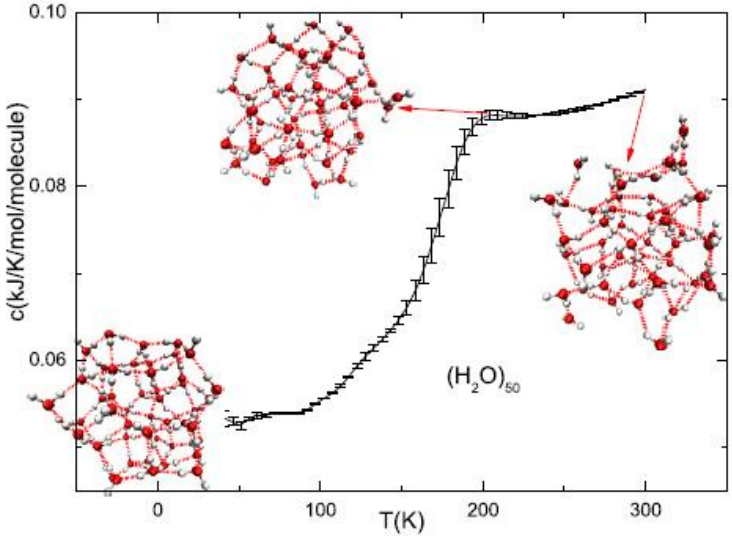


Fig. 6. Specific heat curve for $(H_2O)_{50}$. Two snapshots (with pointing arrow) are taken at the average energy for corresponding temperatures, respectively, and another snapshot (leftmost) is for the ground state structure after quenching. The red ball stands for an Oxygen atom, the white ball for a Hydrogen atom, and the dash line for the hydrogen bonding.

THE JOURNAL OF CHEMICAL PHYSICS **134**, 024109 (2012)

Thank you for your attention