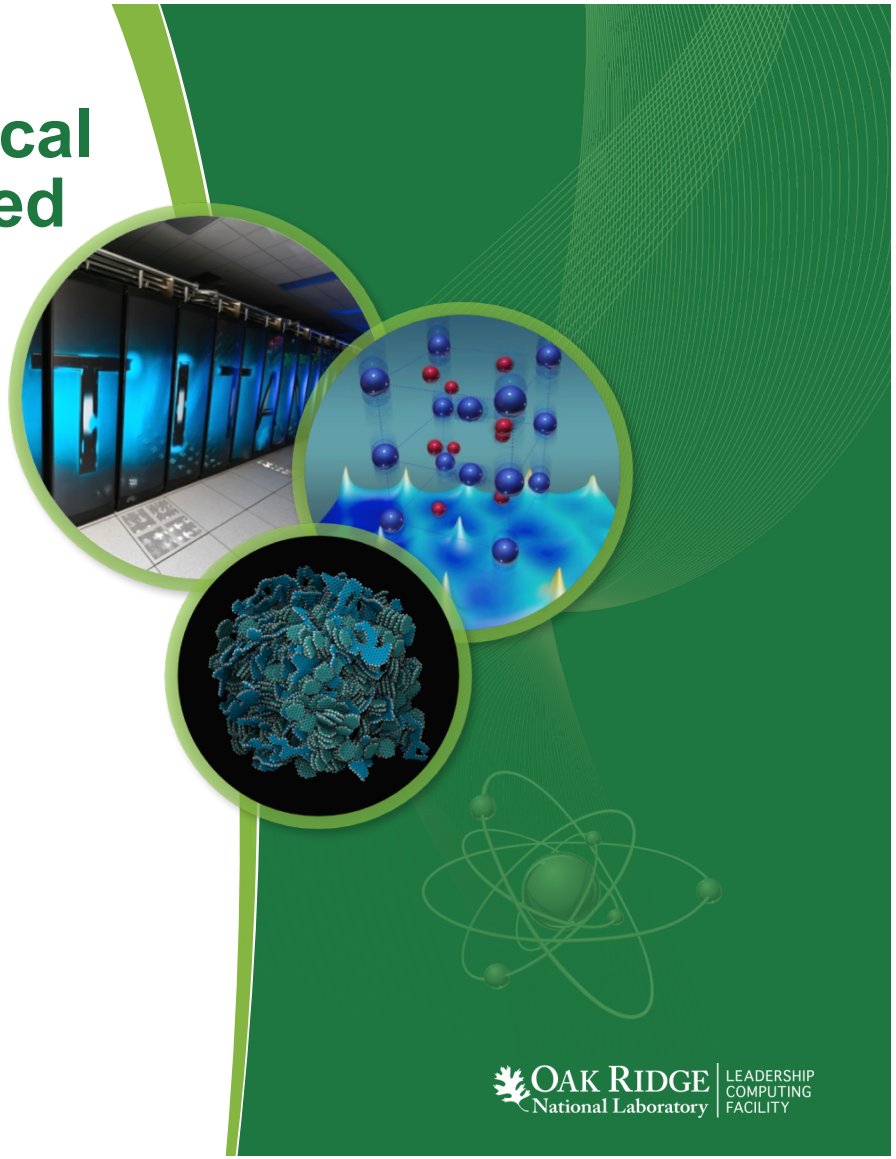
Scalable and efficient multicanonical algorithms for first-principles based Monte Carlo simulations

Ying Wai Li

National Center for Computational Sciences

Oak Ridge National Laboratory, USA



IPAM Workshop on Optimization and Optimal Control for Complex Energy and Property Landscapes

UCLA, Los Angeles October 2-6, 2017

ORNL is managed by UT-Battelle
for the US Department of Energy

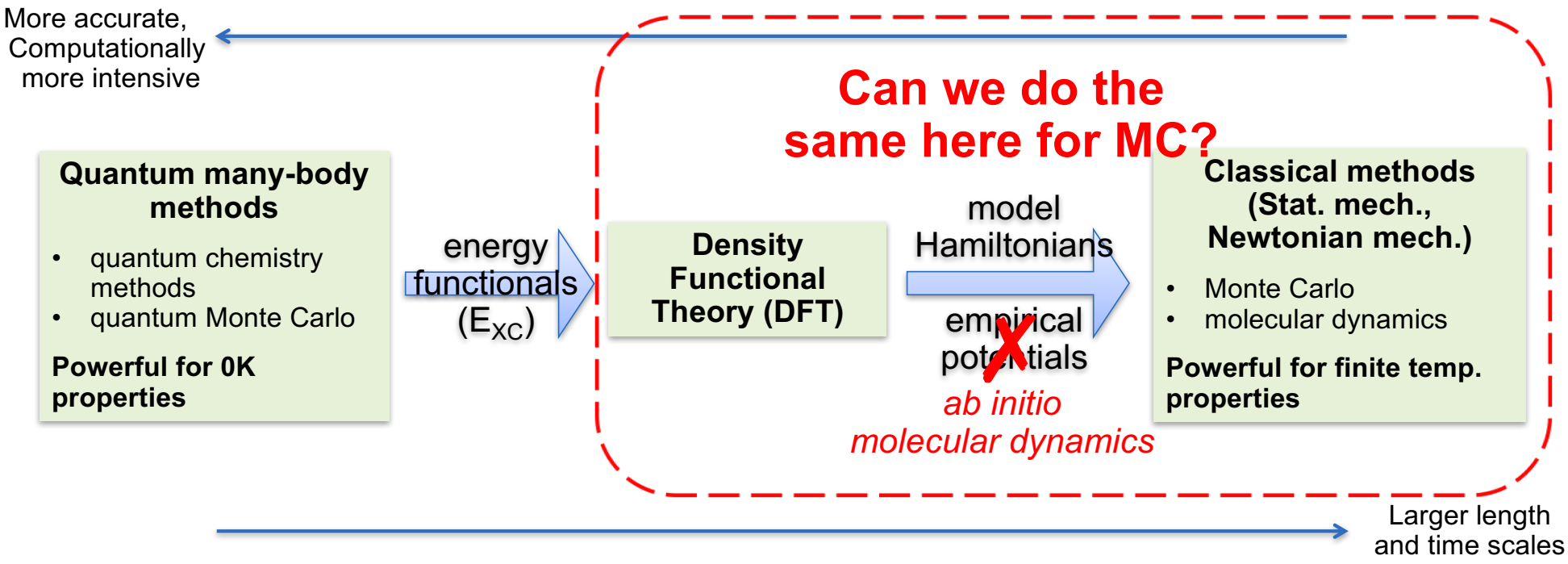
OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

Overview

- Motivation for first-principles based statistical mechanics
- Background of classical Monte Carlo (MC) methods
  - Metropolis sampling
  - Wang-Landau (WL) sampling
- Limitations of the original first principles based WL sampling
- Strategies to improve and speed-up:
  1. **Parallelization:**
     Replica-Exchange Wang-Landau sampling (REWL)
  2. **Improved algorithm:**
     Histogram-free multicanonical sampling

Motivation for first-principles based statistical mechanics

- Study finite temperature, equilibrium properties of materials from first principles
- Bridging length and time scales of computer simulations
- Goal: sample energy landscapes thoroughly to construct phase diagrams



Equilibrium statistical mechanics at a glance

Canonical ensemble:

- System with a heat bath at fixed temperature (T), number of particles (N), volume (V)
- Partition function:

$$Z(T) = \sum_{\text{all states}} e^{-E/k_B T} \equiv \sum_{E} g(E) e^{-E/k_B T}$$

Canonical distribution:

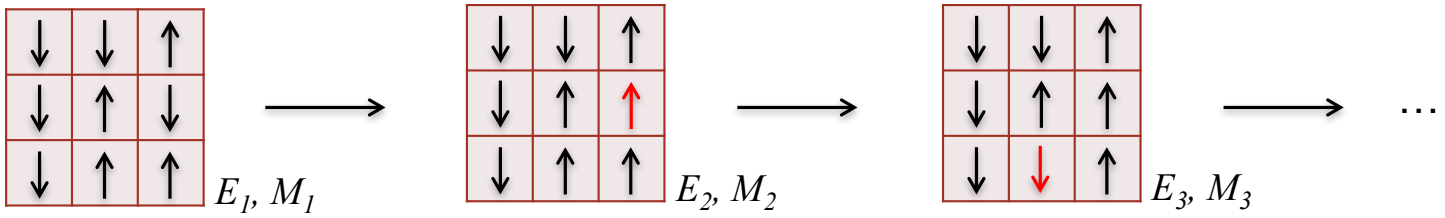
$$P(E,T) = \frac{1}{Z(T)} g(E) e^{-E/k_B T}$$

- Thermodynamic observables:
  - Average energy: $\langle E \rangle_T = \frac{1}{Z_T} \sum_E E g(E) e^{-E/k_B T}$
  - Specific heat: $C_V(T) = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2}$
  - Free energy: $F(T) = -k_B T \ln(Z_T)$

Metropolis sampling: Monte Carlo with canonical ensemble

A Markov chain that obeys canonical distribution at temperature T
(enabled by Boltzmann factor in acceptance + detailed balance)

1. Generate a new configuration
2. Calculate the change in energy, ΔE
3. Accept with probability: $P(E \rightarrow E + \Delta E) = \min\left\{1, e^{-\Delta E / k_B T}\right\}$
4. Calculate and accumulate physical observables of interest
5. Repeat steps 1-3 until a desired number of steps have been performed



Energy and physical observables are accumulated to calculate the ensemble averages

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, J. Chem. Phys. **21**, 1087 (1953)

Critical slowing down of Metropolis sampling

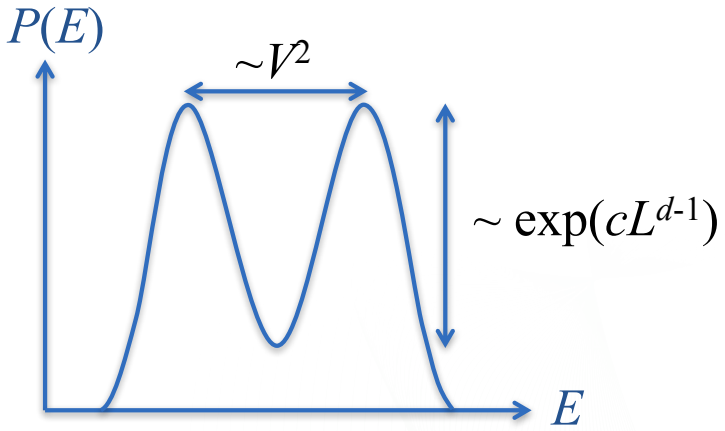
- Generates a canonical distribution $P(E)$ at temperature T (ideally)
- Known problem: critical slowing down near phase transition temperature

Round trip time between the two peaks:

$$\tau \sim V^2 \exp(cL^{d-1}), \quad V = L^d$$

L : characterizes system size

d : dimension



- Difficult to overcome energy barrier → trapped in metastable states
- Problem with sampling low temperature behavior

Solutions: replica exchange; generalized ensemble independent of T

Wang-Landau sampling: MC with generalized ensemble

- An iterative Monte Carlo method to estimate density of states in energy, $g(E)$
- A random walk in energy space - sampling weight modified "on-the-fly"

Histogram $H(E)$, density of states $g(E)$, modification factor f

1. Initialize: $H(E) = 0$, $g(E) = 1$, $f_0 = e^1$
2. Generate a trial configuration, accept with probability:

$$p(E_{old} \to E_{new}) = \min\left(\frac{g(E_{old})}{g(E_{new})}, 1\right) \qquad E = \begin{cases} E_{new}, & \text{if accepted} \\ E_{old}, & \text{otherwise} \end{cases}$$

3. Update $g(E) \to g(E) * f$, $H(E) \to H(E) + 1$
4. Repeat steps 2-3 until the histogram is "flat"; reset $H(E) = 0$, $f_{i+1} = f_i^{1/2}$
5. Repeat steps 2-4 until $f = f_{final} \sim \exp(10^{-8})$

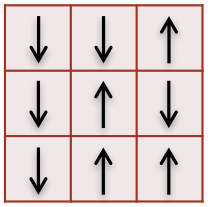
→ Final density of states, $g(E)$

Physical observables are calculated for all temperatures from a single simulation

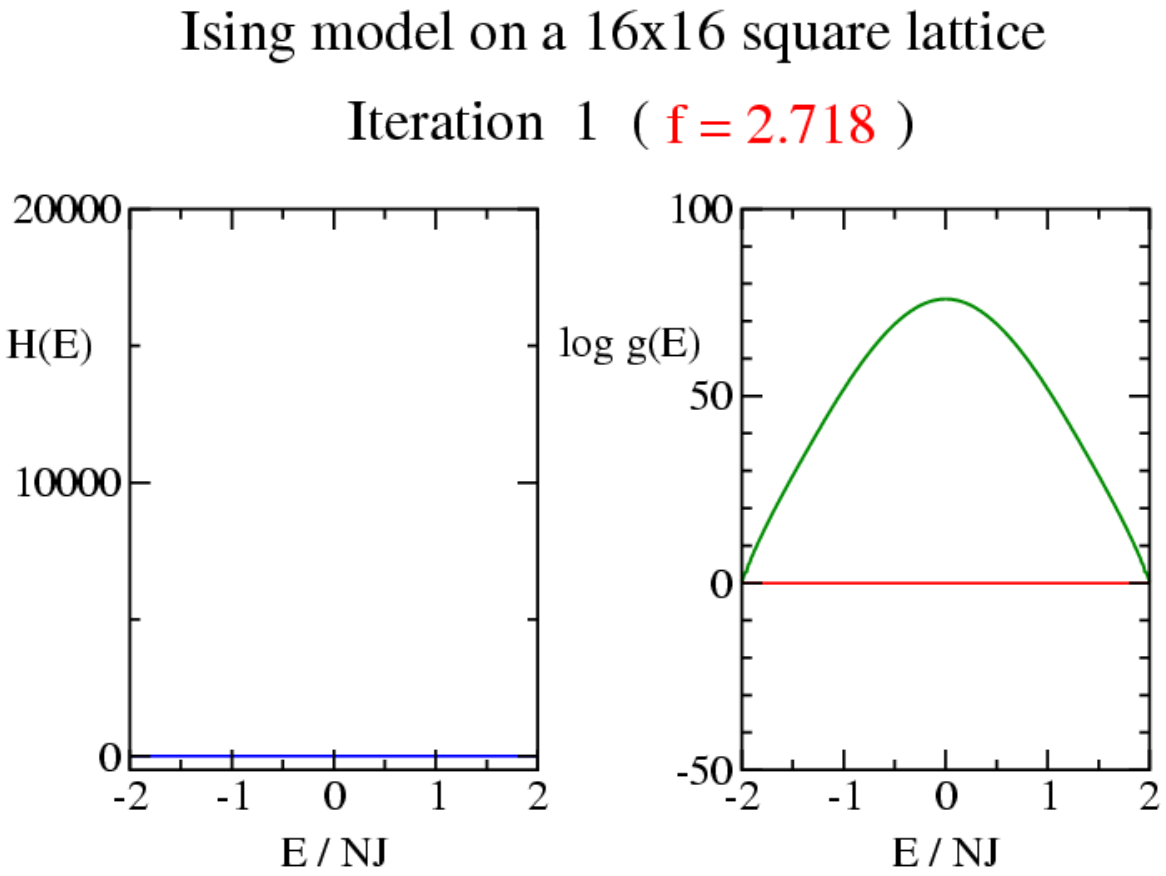
F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); Phys. Rev. E **64**, 056101 (2001)

How Wang-Landau sampling works?

Pedagogical example:



$$E = -J \sum_{<ij>} \sigma_i \sigma_j$$

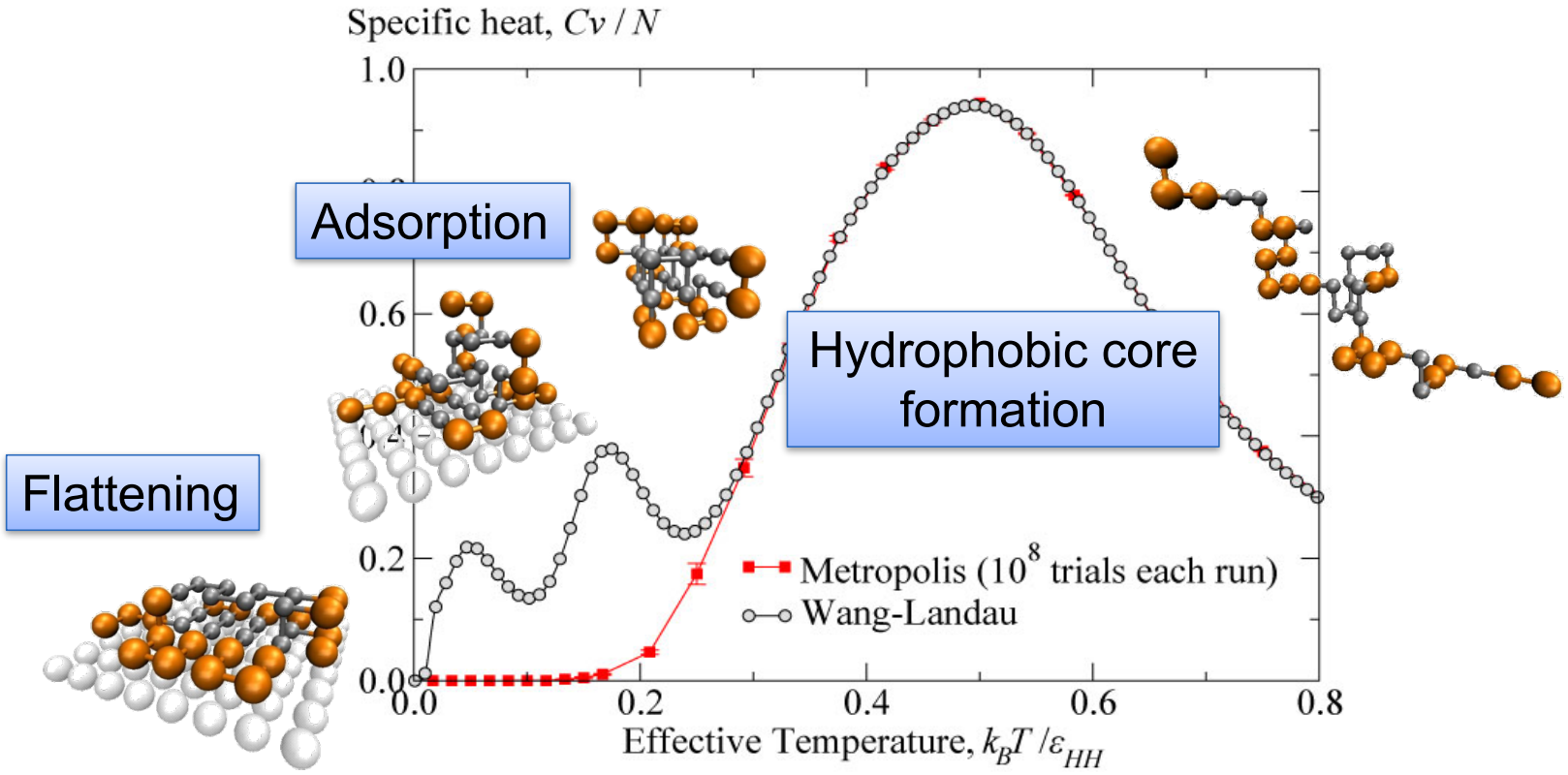


Animation Courtesy: D. P. Landau

F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); Phys. Rev. E **64**, 056101 (2001)

Wang-Landau sampling vs Metropolis algorithm

Hydrophobic-Polar model 36mer, $\varepsilon_{SH} = \varepsilon_{SP} = (1/12)\ \varepsilon_{HH}$ (weakly attractive surface)



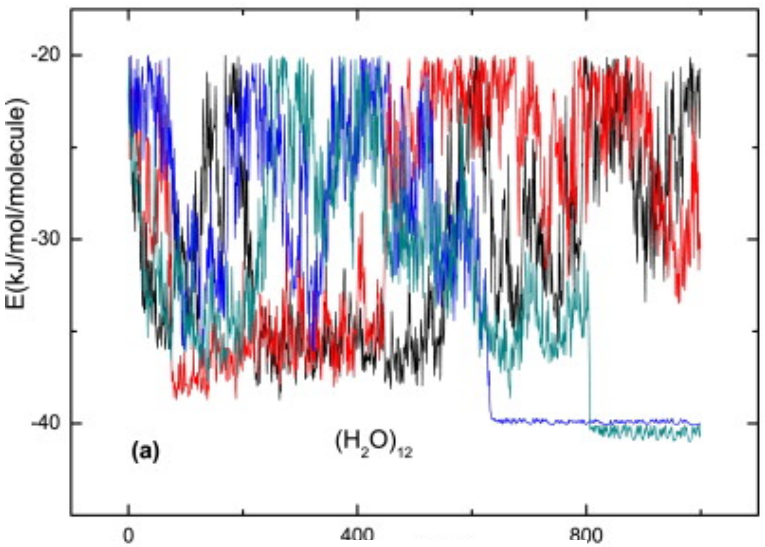
*Error bars smaller than the size of data points are not shown.

Y. W. Li, T. Wüst and D. P. Landau, Comput. Phys. Commun. **182,** 1896-1899 (2011); Phys. Rev. E **87**, 012706 (2013)
T. Wüst, Y. W. Li and D. P. Landau, J. Stat. Phys. **144**, 638 (2011)

Problem with previous parallel Wang-Landau scheme

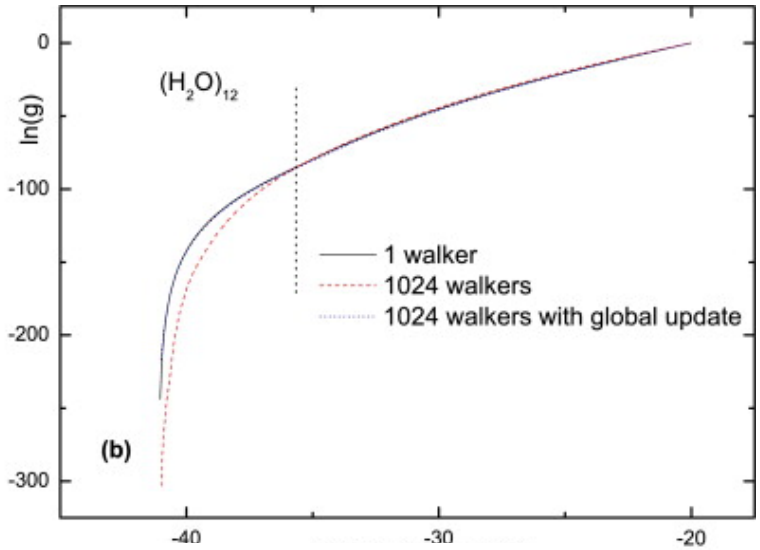
- Scaling limited by single “image” of the external package
- Using multiple MC walkers naïvely results in inaccurate low temperature statistics
- Bottleneck that hinders massive scaling for future machines

“time series”: footprints of walkers



Monte Carlo steps

density of states $g(E)$

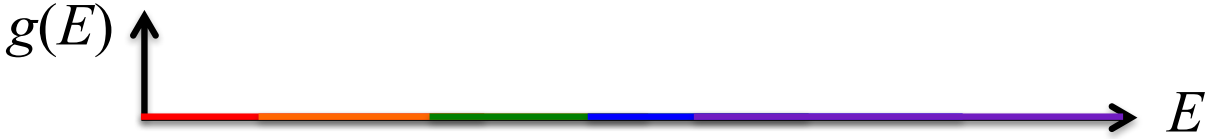


Energy (kJ / mol / molecule)

J. Yin and D. P. Landau, Comput. Phys. Commun. **183**, 1568-1573 (2012)

Parallel, scalable MC simulations: (Replica-Exchange Wang-Landau sampling, REWL)

Serial:

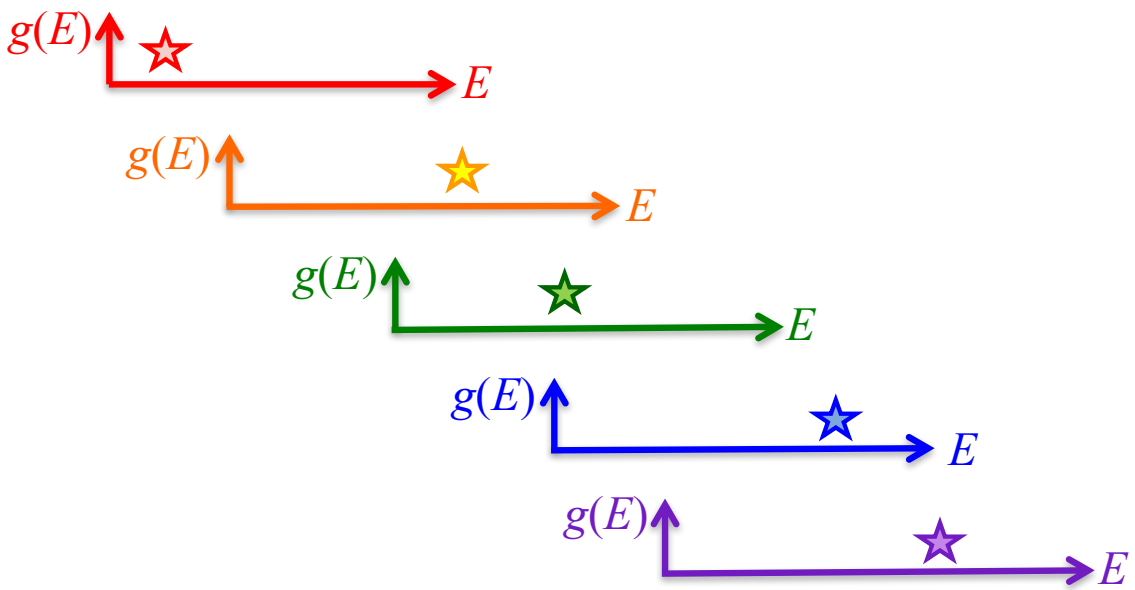


Parallel:

T. Vogel, Y. W. Li, T. Wüst and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013); Phys. Rev. E **90**, 023302 (2014)

Parallel, scalable MC simulations: (Replica-Exchange Wang-Landau sampling, REWL)

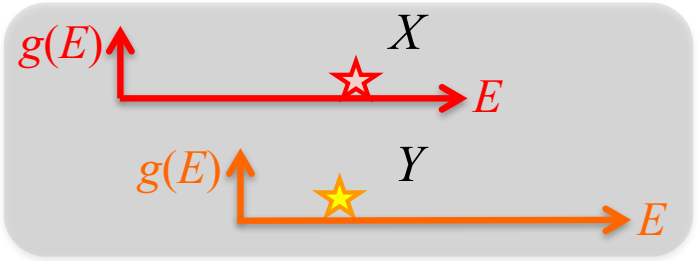
1. Split phase space into several smaller, overlapping windows
2. Ordinary Wang-Landau procedure within a window



T. Vogel, Y. W. Li, T. Wüst and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013); Phys. Rev. E **90**, 023302 (2014)

Parallel, scalable MC simulations: (Replica-Exchange Wang-Landau sampling, REWL)

1. Split phase space into several smaller, overlapping windows
2. Ordinary Wang-Landau procedure within a window
3. Replica-exchange between neighboring windows at intervals



$g_i(E)$

$g_j(E)$

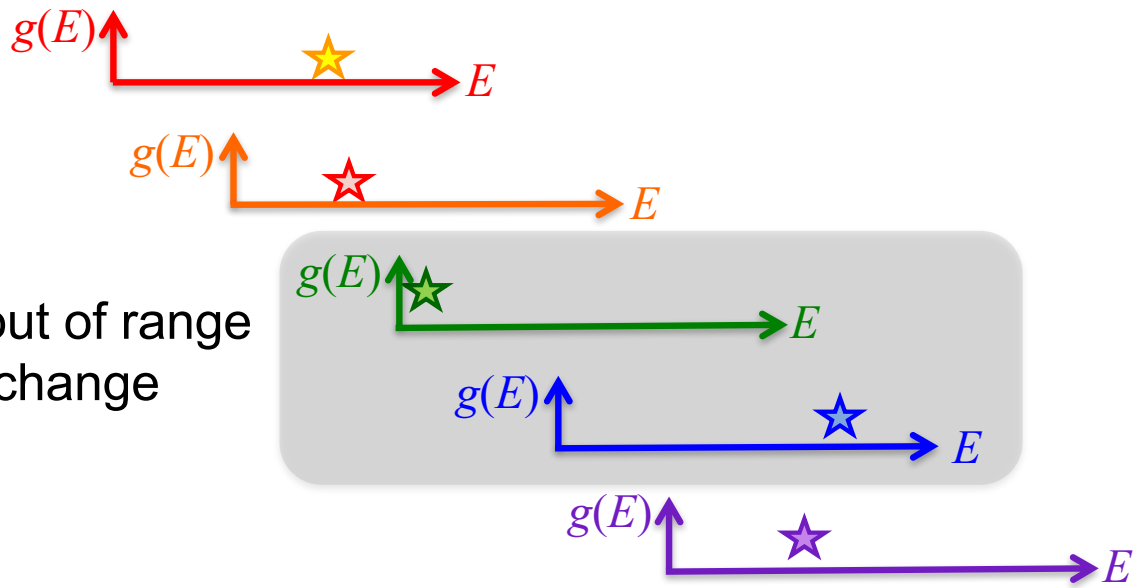
Probability of exchanging configurations X and Y:

$$P(X \leftrightarrow Y) = \min\left[1, \frac{g_i(E_X)g_j(E_Y)}{g_i(E_Y)g_j(E_X)}\right]$$

T. Vogel, Y. W. Li, T. Wüst and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013); Phys. Rev. E **90**, 023302 (2014)

Parallel, scalable MC simulations: (Replica-Exchange Wang-Landau sampling, REWL)

1. Split phase space into several smaller, overlapping windows
2. Ordinary Wang-Landau procedure within a window
3. Replica-exchange between neighboring windows at intervals

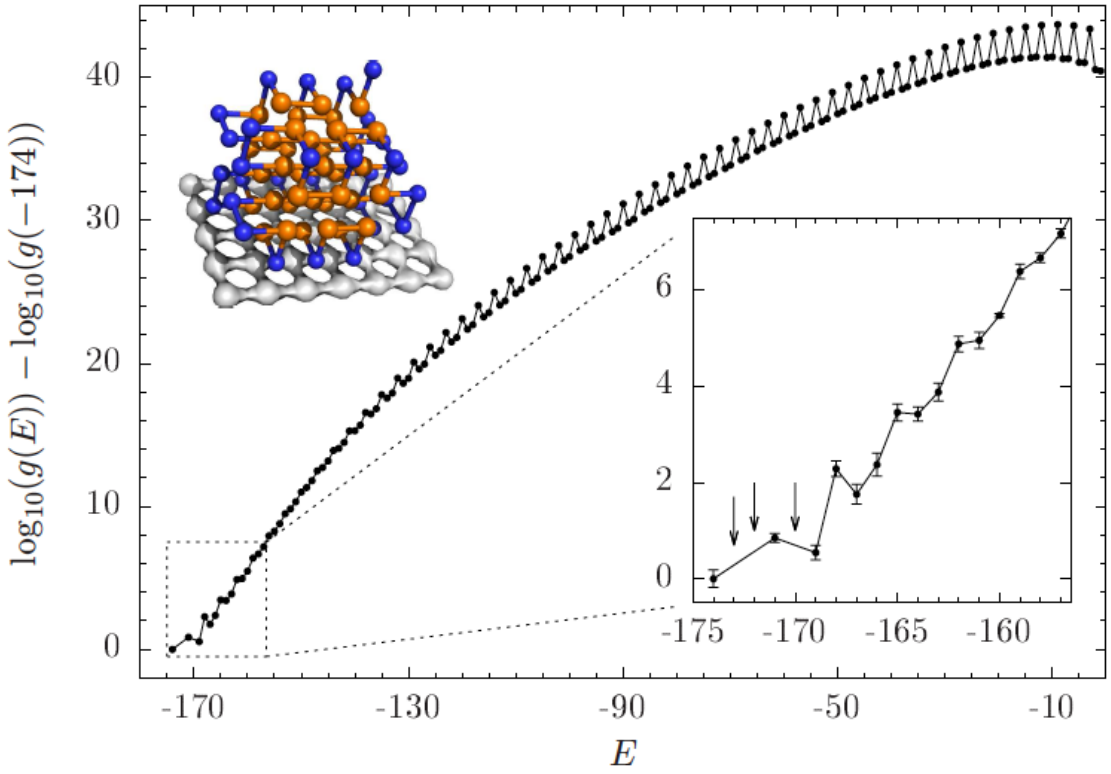


Energy out of range
→ no exchange

T. Vogel, Y. W. Li, T. Wüst and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013); Phys. Rev. E **90**, 023302 (2014)

Robustness in determining $g(E)$

$g(E)$ for HP 67mer adsorption: not accessible by serial WL sampling



h = 9

m = 1

o = 75%

$N^{\text{serial}} \sim 10^{11}$

$N^{\text{parallel}} \sim 10^{10}$

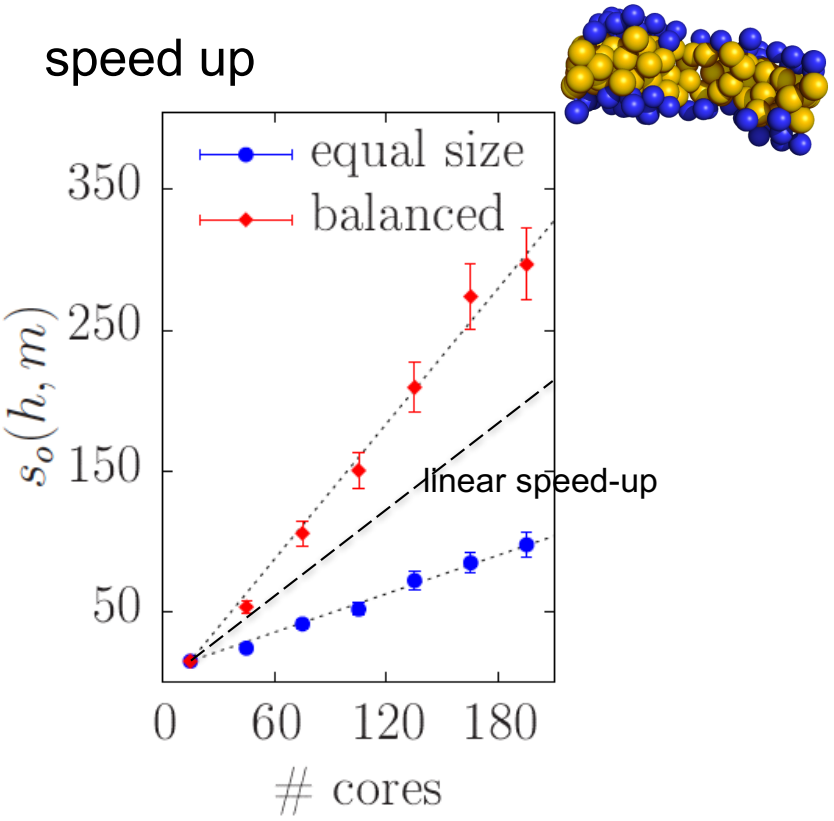
9 processors

speed up $\approx$ 20 !

T. Vogel, Y. W. Li, T. Wüst and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013); Phys. Rev. E **90**, 023302 (2014)

Excellent strong and weak scaling

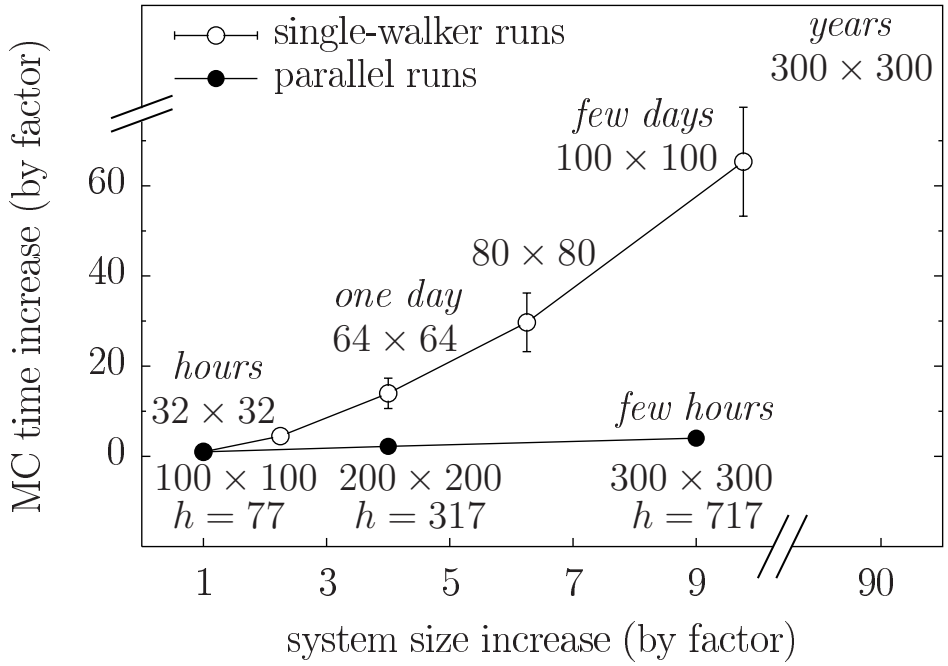
Lipid bilayer model



2D 10-state Potts model

$$E = -J \sum_{<ij>} \delta_{\sigma_i \sigma_j}$$

8	3	5
9	7	2
1	7	10



T. Vogel, Y. W. Li, T. Wüst and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013); Phys. Rev. E **90**, 023302 (2014)

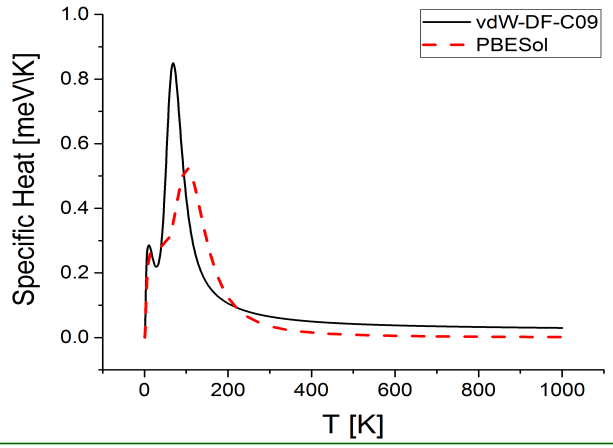
Applications of WL-DFT on Materials Science

Study of ferroelectric phase transition

(with S. F. Yuk, V. R. Cooper et al.)

- Transition temperature T_C depends on XC functionals and finite size effects
- Parametrize Landau-Ginsburg model with PBEsol and vdW-DF-C09 for a $1\times1\times1$ $PbTiO_3$ unit cell

(preliminary results)



Caloric materials Consortium (CaloriCool)

(with K. Odbadrakh et al.)

- To establish DFT-based caloric materials discovery framework
- Calculate caloric properties and relevant phase transformation data
- Sample magnetic states of FeRh with respect to thermal lattice expansion, structural excitations, etc.
- Theoretical predictions from OWL will provide data for informatics-based data mining and verification against existing data

Work in progress…

S.F. Yuk, K.C. Pitike, S.M. Nakhmanson, M. Eisenbach, Y. W. Li and V. R. Cooper, Scientific Reports **7**, 43482 (2017)

Motivation 1: Need for alternative representations of the simulated density of states

- For systems with continuous observables (such as the spin systems), the discreteness introduced by binning causes an artificial source of error for histogram methods

  → Is it possible to avoid binning?

- Provided some physical insights of the analytical form of the density of states of a system, can one obtain the coefficients directly through computer simulations?

Motivation 2: Need for improved Monte Carlo efficiency

From performing first-principles based statistical mechanics using LSMS

- Exploited all possible parallelization techniques
  - multiple MC random walkers to scale up
  - domain decomposition by LSMS / DFT code
  - GPU / accelerators for compute-intensive linear algebra
- **Yet, computational cost is still unaffordable**
  - *ab initio* energy evaluation is expensive
    (e.g. LSMS generates 1.5 energies/sec on full Titan @ ORNL, including GPUs, for 1024 iron atoms)
  - reliable, one-dimensional phase space construction is also expensive (at least 10^6 – 10^8 MC steps)

= 1 Titan week …!

Back to basic:
Is there a systematic way to reduce computational effort to achieve the same results?

LSMS:
Locally **S**elf-consistent **M**ultiple **S**cattering, a linear scaling, KKR density functional theory (DFT) method

WL-LSMS:
An ORNL application (Gordon Bell winner 2009) that combines WL and LSMS for studying Curie T of iron clusters

M. Eisenbach *et. al.*, SC'09: Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis, ACM (2009).

Histogram-free multicanonical sampling

- An iterative Monte Carlo method to estimate a basis expansion of the density of states, $g(E)$
- Determine and modify sampling weight at intervals (multicanonical sampling)

1. Assume an analytical form for the density of states and the correction:

$$\ln g(E) = \sum_{i=0}^{N} g_i \phi_i(E) \qquad \ln c(E) = \sum_{i=0}^{N} c_i \phi_i(E)$$

$\phi_i(E)$: a chosen basis set

The number of terms, N, and the coefficients $\{c_i\}$ (and hence $\{g_i\}$) are to be found iteratively

2. Initial guess: $g^0(E) = 1$
3. Generate a series of k energies as data set $\mathcal{D} = \{E_1, E_2, E_3, \ldots, E_k\}$, with acceptance probability:

$$p(E_{\text{old}} \to E_{\text{new}}) = \min\left(\frac{g(E_{\text{old}})}{g(E_{\text{new}})}, 1\right) \qquad E = \begin{cases} E_{new}, & \text{if accepted} \\ E_{old}, & \text{otherwise} \end{cases}$$

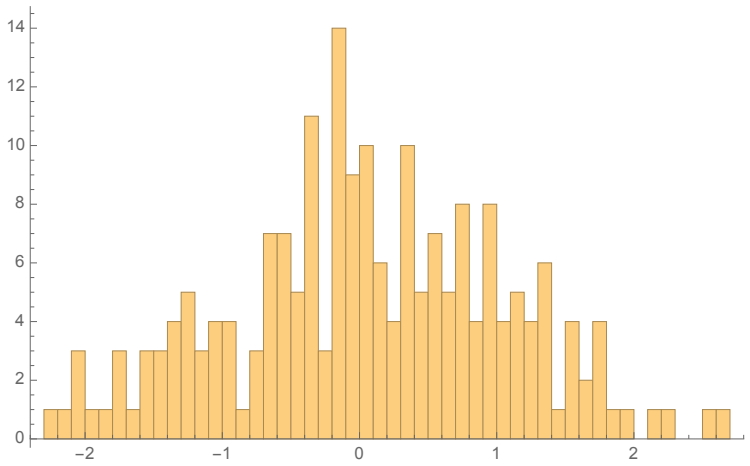
4. At intervals, use $\mathcal{D}$ to find the correction $\ln c(E)$
5. Update $g(E)$ with correction $c(E)$: $\ln g(E) \to \ln g(E) + \ln c(E)$. Discard $\mathcal{D}$.
6. Repeat steps 3-5 until $\ln c(E) \to 0$

Representing data without a histogram

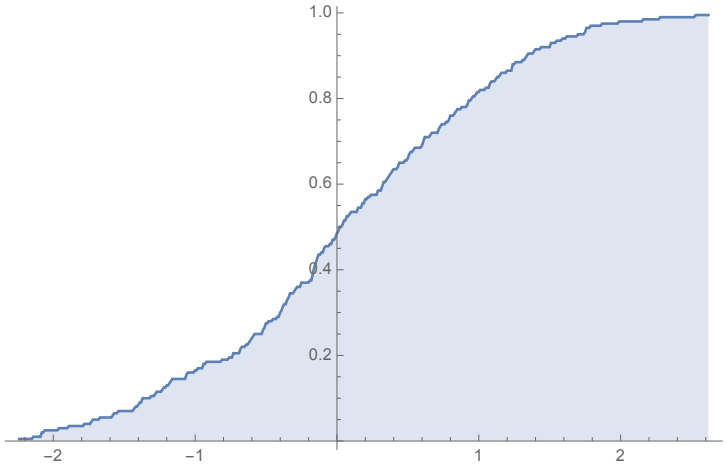
- Generate k energy measurements and stored as a data set:
  $\mathcal{D} = \{E_1, E_2, E_3, \ldots, E_k\}, \qquad E_1 < \ldots < E_k$
- Construct cumulative distribution function, CDF:

$$CDF(E) = \int_{-\infty}^{E} H(E')\,dE'$$

Histogram, $H(E)$



CDF, $CDF(E)$



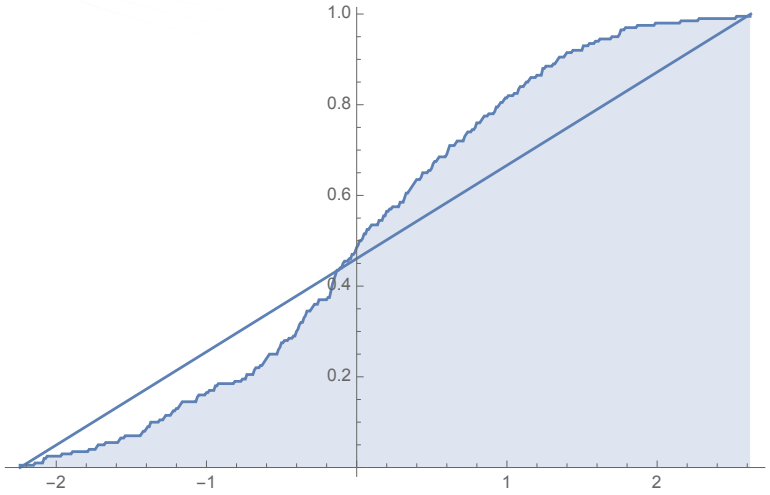
B. A. Berg and R. C. Harris, Comput. Phys. Commun. **179**, 443-448 (2009)

Representing data without a histogram

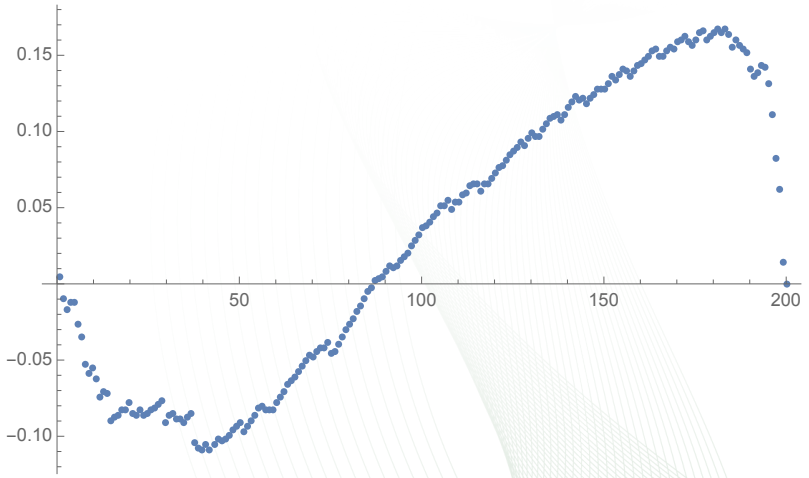
At convergence, a flat histogram is resulted
→ a straight line of constant slope for the CDF

- Rewrite CDF with a straight line and remainder $R(E)$: $$CDF(E) = \frac{E - E_{\min}}{E_{\max} - E_{\min}} + R(E)$$

$CDF(E)$



Remainder, $R(E)$



Finding correction $c(E)$ from $R(E)$

- $R(E)$ is the deviation of the empirical CDF from the uniform distribution (flat histogram)
- It provides a means of calculating the correction to $g(E)$, i.e., the coefficients c_i in $\ln c(E)$
- Recall that:

$$CDF(E) = \int_{-\infty}^{E} H(E')dE' \qquad CDF(E) = \frac{E - E_{\min}}{E_{\max} - E_{\min}} + R(E)$$

$$\rightarrow \quad H(E) = \frac{dCDF(E)}{dE} = \frac{1}{E_{\max} - E_{\min}} + \underbrace{\frac{dR(E)}{dE}}_{\ln c(E)}$$

- Finally, update $\ln g(E)$ → $\ln g(E) + \ln c(E)$

Y. W. Li and M. Eisenbach, in Proceedings of the Platform for Advanced Scientific Computing Conference (PASC '17). Association for Computing Machinery (ACM), New York, NY, USA, Article 10, 7 pages (2017).

Approximating remainder $R(E)$

- Expand $R(E)$ into a series of orthonormal basis set:

$$R(E) = \sum_{i=0}^{N} r_i \psi_i(E)$$

- The coefficients r_i are then found by:

$$r_i = \boldsymbol{N} \int_{E_{min}}^{E_{max}} R(E)\psi_i(E)dE$$

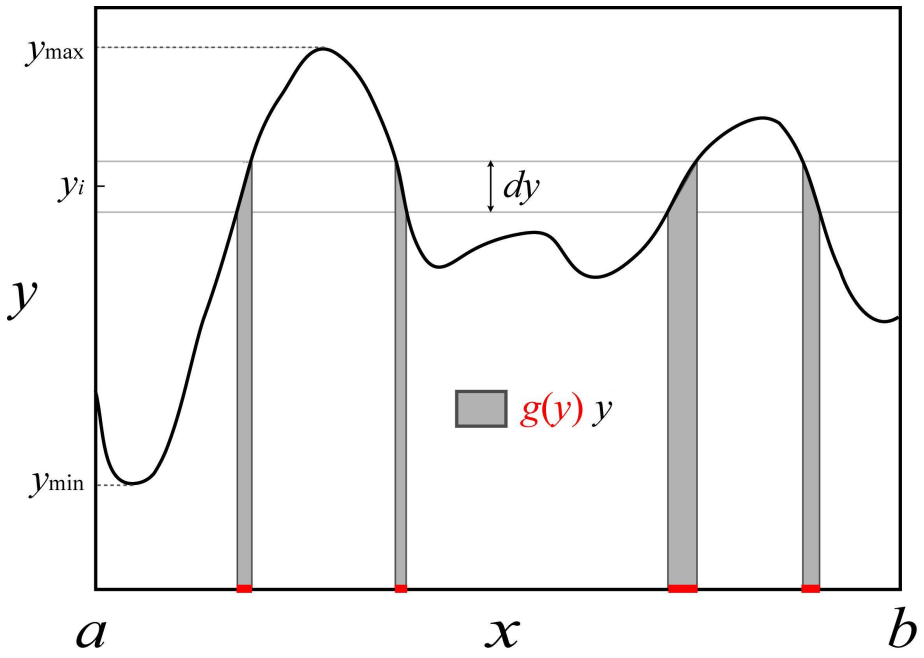
  where $\boldsymbol{N}$ is some constants dependent on choice of basis set

- Start from $N = 0$. Perform a statistical test to obtain the probability p that the empirical remainder comes from $R(E)$.
- If not, N → N+1. Repeat until p reaches a predefined value, say, 0.5.

Y. W. Li and M. Eisenbach, in Proceedings of the Platform for Advanced Scientific Computing Conference (PASC '17). Association for Computing Machinery (ACM), New York, NY, USA, Article 10, 7 pages (2017).

Test case: numerical integration by Wang-Landau sampling[1]

$$I = \int_a^b y(x)dx = \sum_{y_i=y_{\min}}^{y_{\max}} g(y_i)y_i = \int_{y_{\min}}^{y_{\max}} g(y)ydy$$



- A “model” with continuous “phase space”
- A stringent test with exact answer available

Attention:
This scheme is NOT meant to be an efficient numerical integration algorithm. For an improved version, see [2].

[1] Y. W. Li, T. Wüst, D. P. Landau and H. Q. Lin, Comput. Phys. Commun. **177**, 524 (2007)
[2] W. Atisattapong and P. Marupanthorn, Comput. Phys. Commun., in press (2017)

Integration for $\int_{-2}^{2} x^2 dx = 5.3333\ldots$

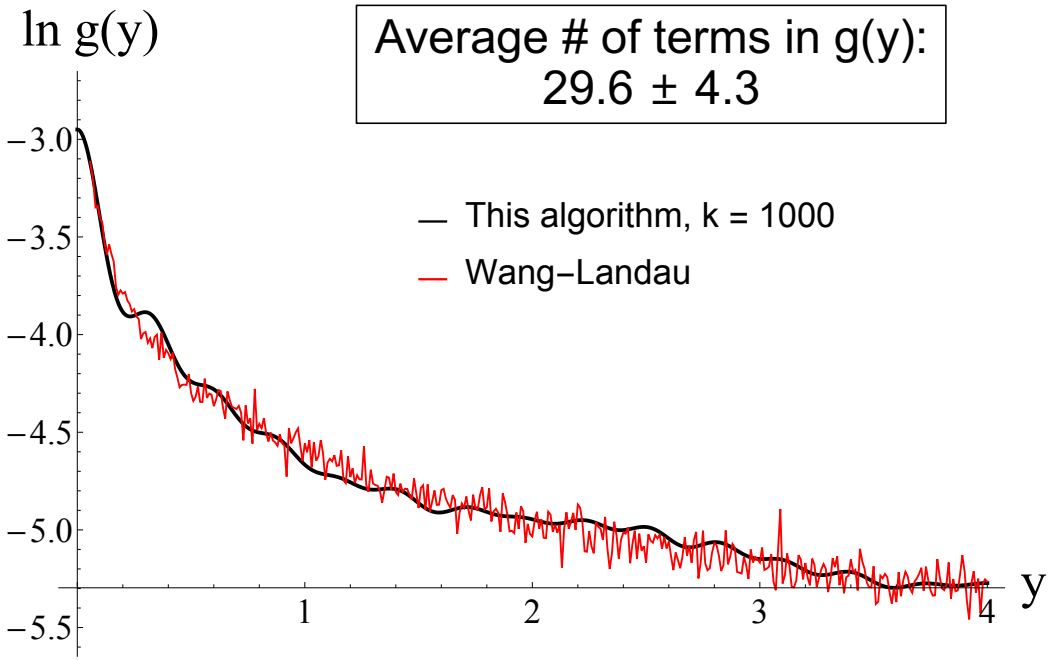
Exact g(y): $g(y) = \dfrac{2(2-\sqrt{y})}{y}$

Quality of the density of states:

- much smoother than Wang-Landau sampling
- still some low frequency fluctuations

Computational cost:

- Histogram-free MUCA: O(**10^5**) MC steps
- Wang-Landau sampling: O(**10^6**) MC steps



Average # of terms in g(y): 29.6 ± 4.3

(comparing only a single run to show statistical noise)

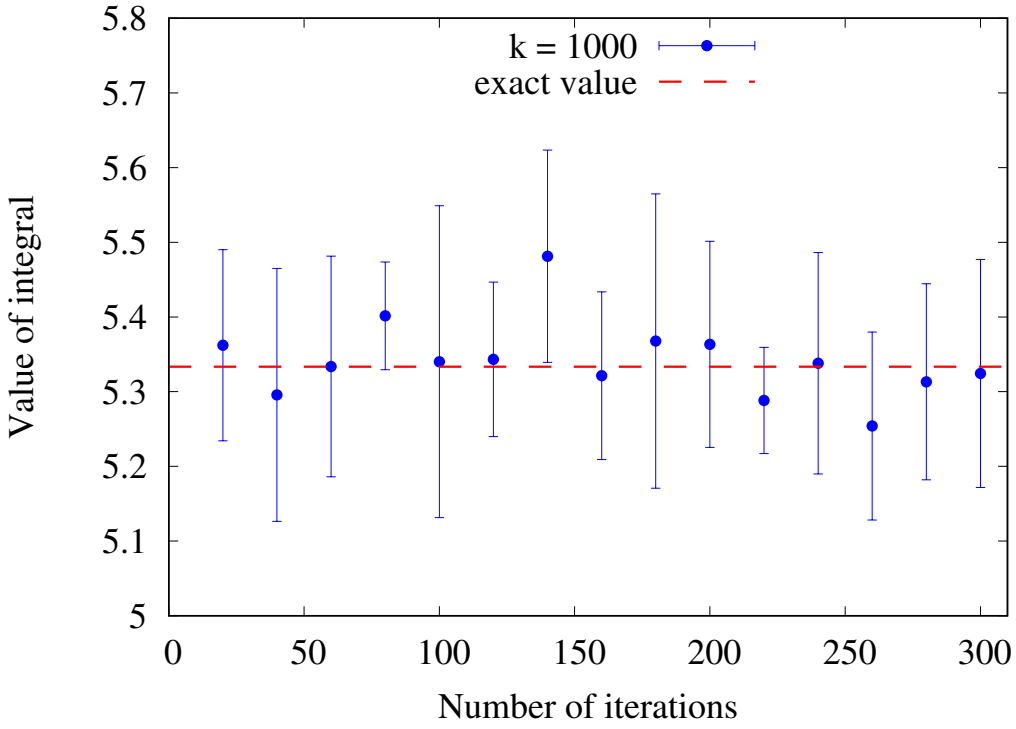
A 10x speed-up!

Y. W. Li and M. Eisenbach, in Proceedings of the Platform for Advanced Scientific Computing Conference (PASC '17). Association for Computing Machinery (ACM), New York, NY, USA, Article 10, 7 pages (2017).

Integration for $\int_{-2}^{2} x^2 dx = 5.3333\ldots$

"Simulation" details:

- $\mathcal{D} = \{E_1, E_2, E_3, \ldots, E_k\}$, $k = 1000$
- Basis set to fit the remainder, $\psi_i(E)$: Fourier sine series
- Basis set for $\ln g(E)$ and $\ln c(E)$, $\phi_i(E)$: Fourier cosine series
- Statistical fit test: Kolmogorov-Smirnov test, p = 0.5



(error bars are obtained from 5 independent runs)

Y. W. Li and M. Eisenbach, in Proceedings of the Platform for Advanced Scientific Computing Conference (PASC '17). Association for Computing Machinery (ACM), New York, NY, USA, Article 10, 7 pages (2017).

Observations

- Speed-up comes from two new features:
  - (major) moving random walkers *intentionally* to achieve uniform sampling of energy space
  - (minor) by assuming an analytical form for the density of states / phase space, the number of parameters to optimize is reduced from O(100-1000) to O(1-100)

- The >10 times reduction in MC steps is a considerable speedup for expensive energy evaluations, e.g. density functional theory calculations

- Generation of data points (energy evaluations) is independent of each other → readily parallelizable

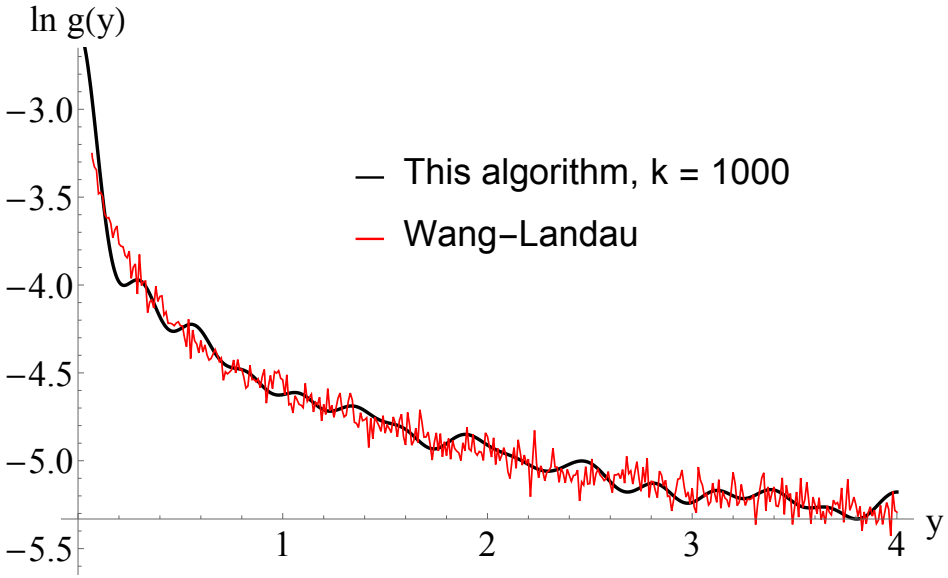
- Accuracy also depends on basis functions
  - … but a good choice is not trivial!
  - requires human insights / lots of experiments to select a suitable set of basis functions

Can we better control convergence?

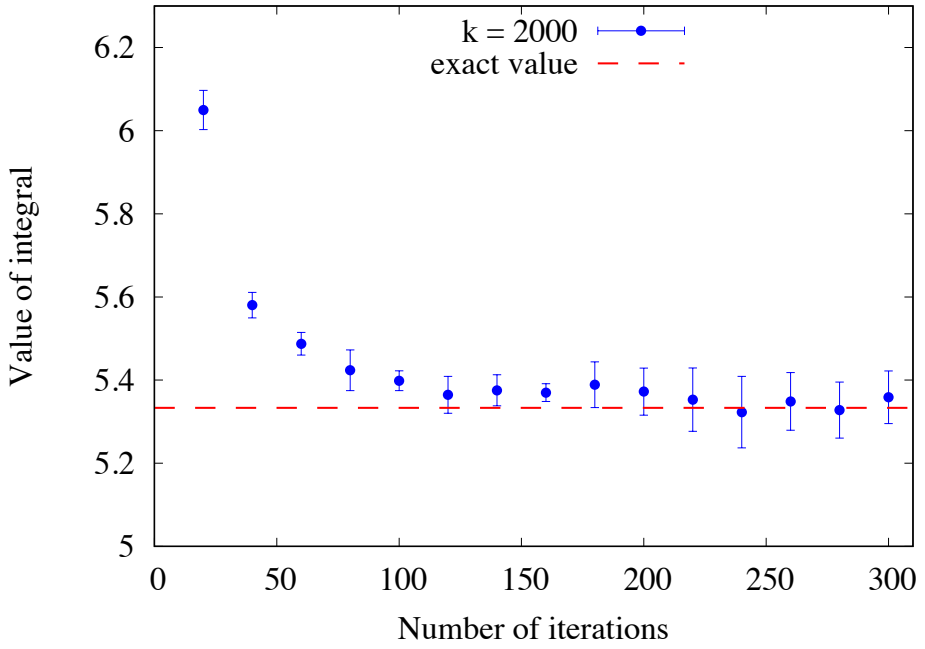
1. Construct remainder using random permutations of basis set (instead of starting from N = 0)

$$R(E) = \sum_{i=0}^{N} r_i \psi_i(E)$$

2. Use a milder correction factor: $\ln c(E) = \frac{1}{4}\frac{dR(E)}{dE}$



(comparing only a single run to show statistical noise)



(error bars are obtained from 5 independent runs)

Can we better control convergence (quantitatively)?

Desired features:

1. a quantitative measure to tell whether the estimated analytical form is a good fit to the data
2. a quantitative measure to tell when to terminate the simulation
3. Increasing the number of data points should improve the answer

Solutions:

1. ~~Before: Kolmogorov-Smirnov test~~
   Now: Define a distance measure

$$d_k = \frac{1}{k}\sum_{i=1}^{k}(CDF_{\text{fitted}}(E_i) - CDF_{\text{data}}(E_i))^2$$

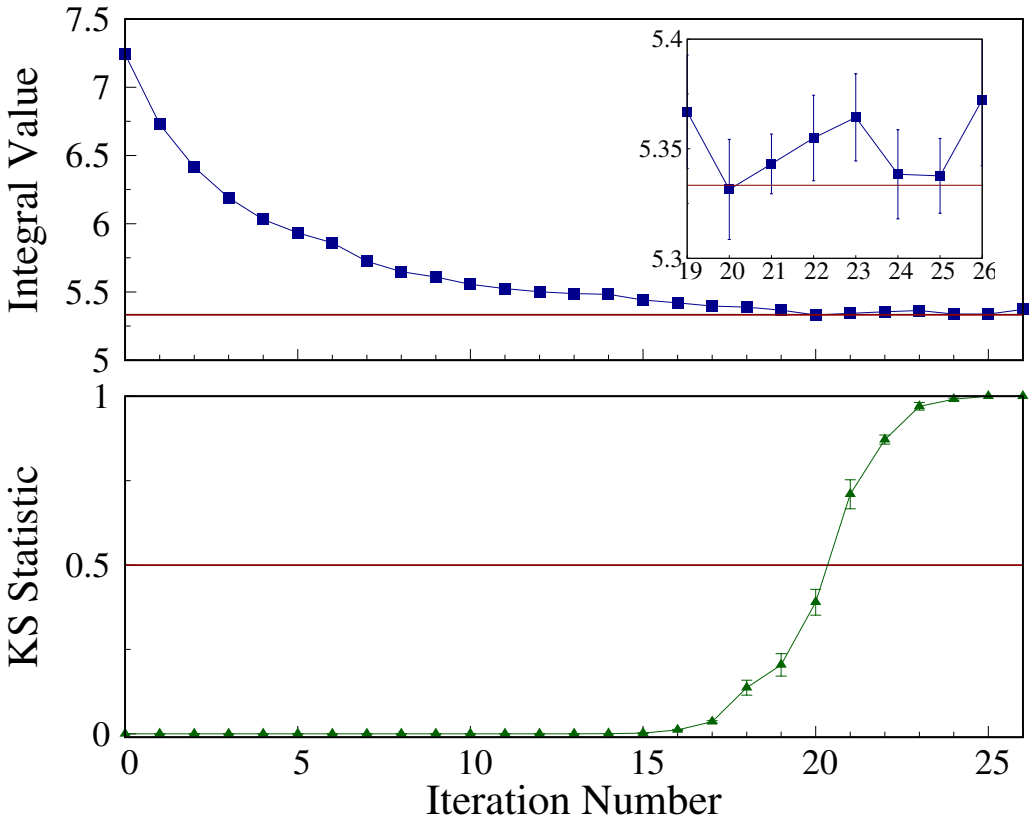
   to cut off the number of terms in analytical fit
2. Kolmogorov-Smirnov test to terminate the simulation
3. This is automatically fulfilled when overfitting does not occur

Integration for $\int_{-2}^{2} x^2 dx = 5.3333\ldots$

Estimated I: 5.34307 ± 0.01363

"Simulation" details:

- $\mathcal{D} = \{E_1, E_2, E_3, \ldots, E_k\}$, $k = 10{,}000$
- Slowly quench the accuracy by decreasing d_k to d_k / 1.5 at each iteration
- Statistical fit test condition to terminate the simulation: Kolmogorov-Smirnov test, p = 0.5



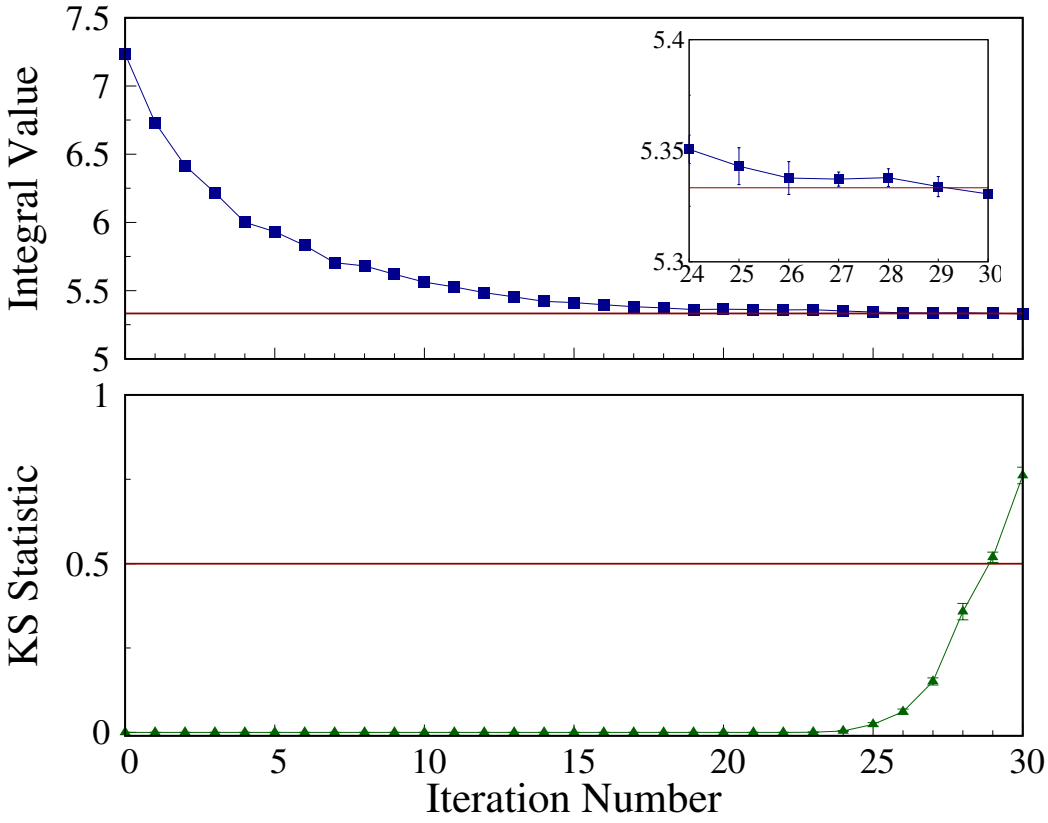
(error bars are obtained from 5 independent runs)

Integration for $\int_{-2}^{2} x^2 dx = 5.3333\ldots$

Estimated I: 5.33386 ± 0.0045

"Simulation" details:

- $\mathcal{D} = \{E_1, E_2, E_3, \ldots, E_k\}$, $k = 100{,}000$
- Slowly quench the accuracy by decreasing d_k to d_k / 1.5 at each iteration
- Statistical fit test condition to terminate the simulation: Kolmogorov-Smirnov test, p = 0.5



(error bars are obtained from 5 independent runs)

Summary

- Large-scale Monte Carlo simulations for sampling energy space evenly
- The need to improve **quality** and **efficiency** of MC simulations motivated by first-principles based statistical mechanics
- Two new MC algorithms proposed to address the problems of
  - **Scalability**: Replica-exchange Wang-Landau sampling
  - **Efficiency**: Histogram-free multicanonical sampling

Outlook

- Application to real physical systems under way…
  (goal: capable of obtaining results comparable to experiemental data)
- Improvements on performance and accuracy of the histogram-free multicanonical sampling method

Acknowledgements

Collaborators:

- Dr. Markus Eisenbach, ORNL
- Prof. Thomas Vogel, U. of North Georgia
- Dr. Thomas Wüst, ETH (Switzerland)
- Prof. David P. Landau, U. of Georgia
- Dr. Valentino R. Cooper, ORNL
- Dr. Simuck F. Yuk, ORNL
- Dr. K. Odbadrakh, U. of Tennessee
- Mr. Alfred C. Farris, U. of Georgia

Funding and computer resources:

- Oak Ridge Leadership Computing Facility (OLCF), ORNL, USA
- Laboratory-Directed Research and Development (LDRD) Seed Funding, ORNL, USA
- The INCITE program for computer resources, Department of Energy, USA