Speeding up Wang-Landau sampling of lattice model proteins using GPUs

Ying Wai Li, Thomas Wüst and David P. Landau

Center for Simulational Physics, The University of Georgia, USA

Acknowledgment:

Pan-American Advanced Studies Institute (PASI) program of NSF

Outline

HP lattice protein model

Monte Carlo methods

Statistical Physics

Wang-Landau sampling

Thermodynamics

How do GPUs help?

Two ways for physics simulations

Deterministic

- molecular dynamics

Stochastic

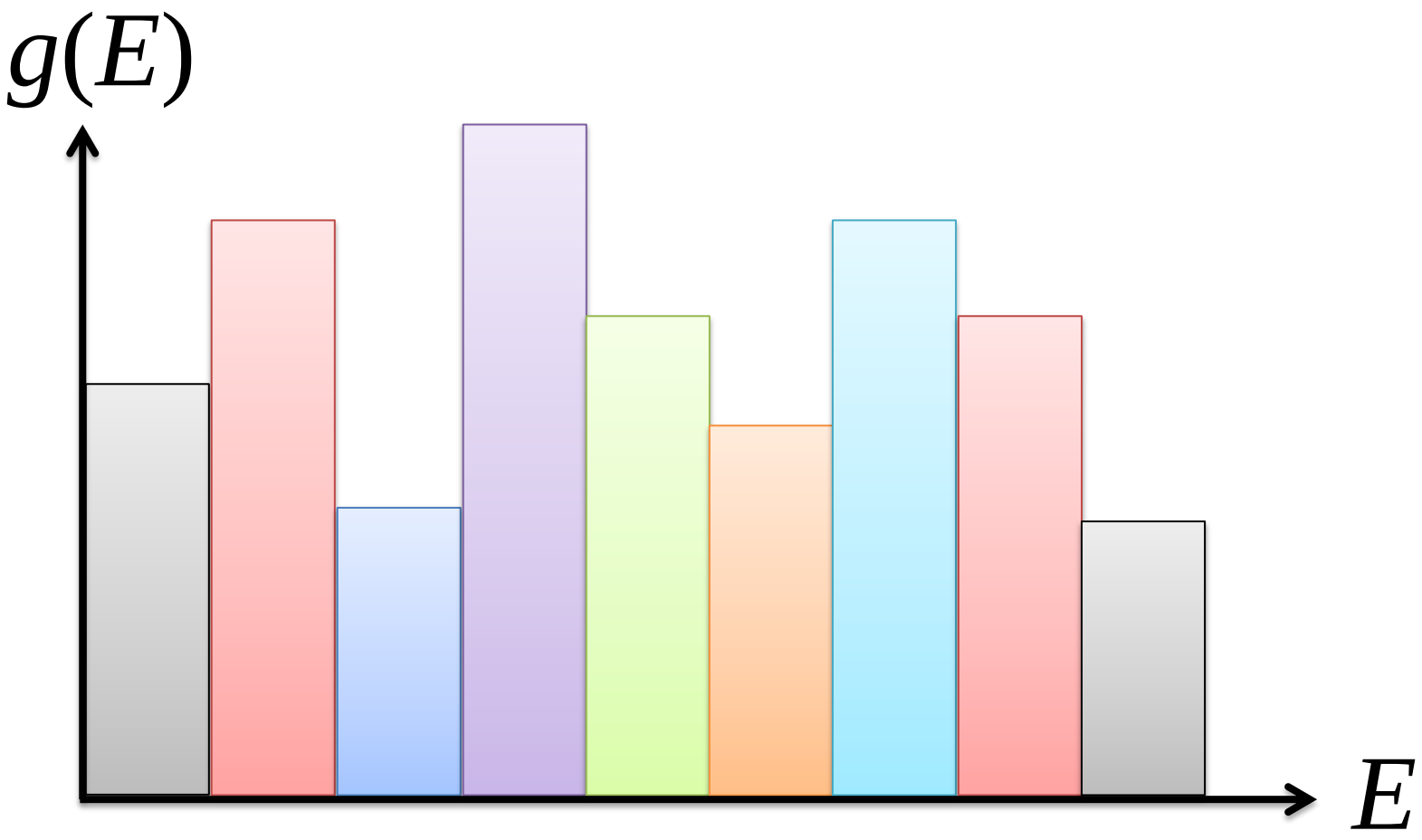
- Monte Carlo

Wang-Landau sampling

- A random walk in energy space to obtain $g(E)$
- Physical observables are available for all temperatures from a single simulation

Density of states in energy, $g(E)$

Proportion of configurations having energy E



Definitions of thermodynamic quantities

Partition function:

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

Average energy:

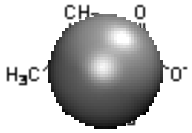
$$\langle E \rangle_T = \frac{1}{Z_T} \sum_E E g(E) e^{-E/k_B T}$$

Specific heat:

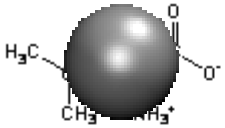
$$C_V(T) = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2}$$

Simplifying proteins into models

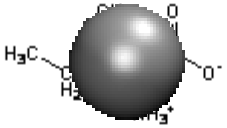
Amino acids with hydrophobic side chains



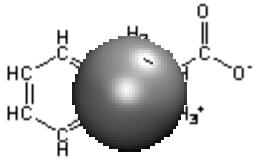
Valine
Val
V



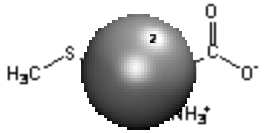
Leucine
Leu
L



Isoleucine
Ile
I

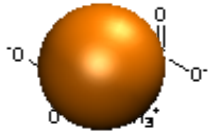


Phenylalanine
Phe
F

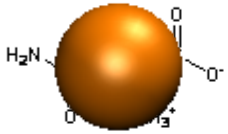


Methionine
Met
M

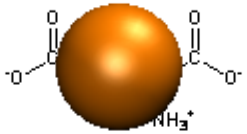
Amino acids with hydrophilic side chains



Aspartic acid
Asp
D

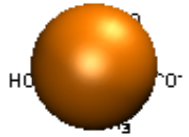


Asparagine
Asn
N

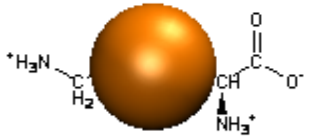


Glutamic acid
Glu
E

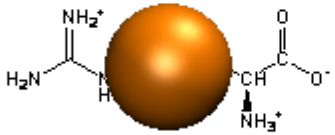
Glutamine
Gln
Q



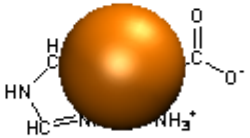
Serine
Ser
S



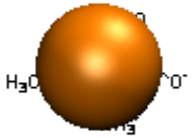
Lysine
Lys
K



Arginine
Arg
R

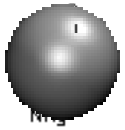


Histidine
His
H

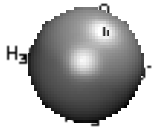


Threonine
Thr
T

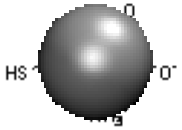
Amino acids with intermediate side chains



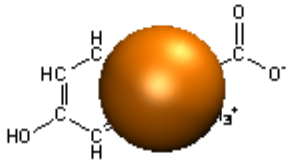
Glycine
Gly
G



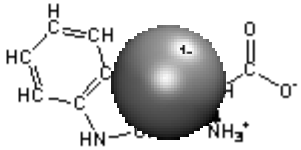
Alanine
Ala
A



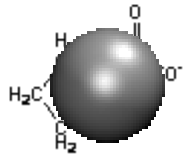
Cysteine
Cys
C



Tyrosine
Tyr
Y

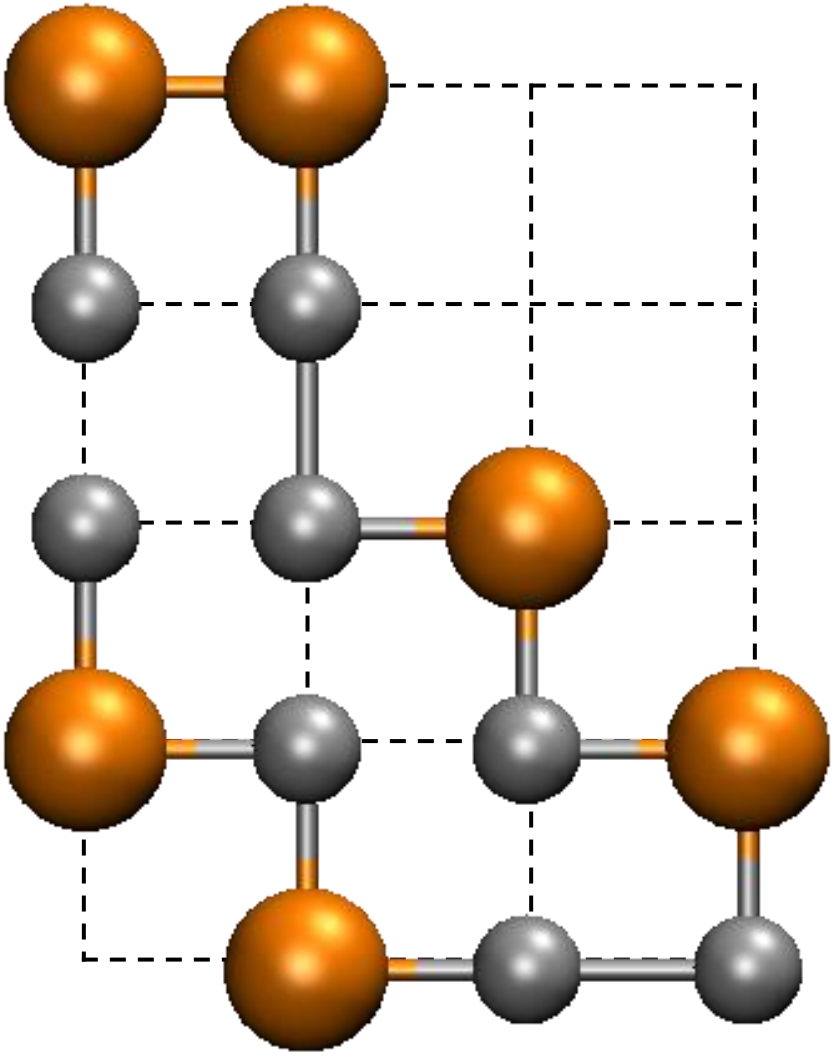


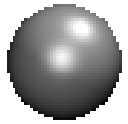
Tryptophan
Trp
W



Proline
Pro
P

The HP protein model





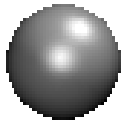
Hydrophobic

Polar

Attractions between:

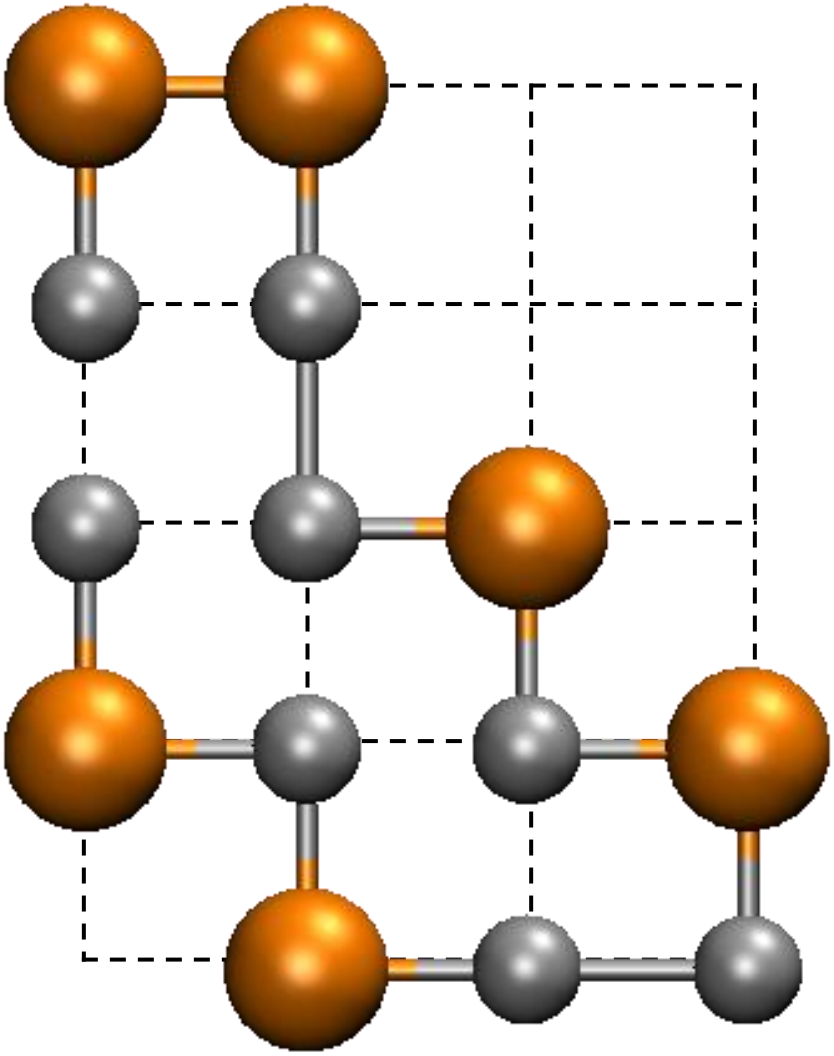
nearest-neighbor

non-bonded



K. A. Dill, Biochemistry **24**, 1501 (1985); K. F. Lau & K. A. Dill, Macromolecules **22**, 3896, (1989)

The HP protein model



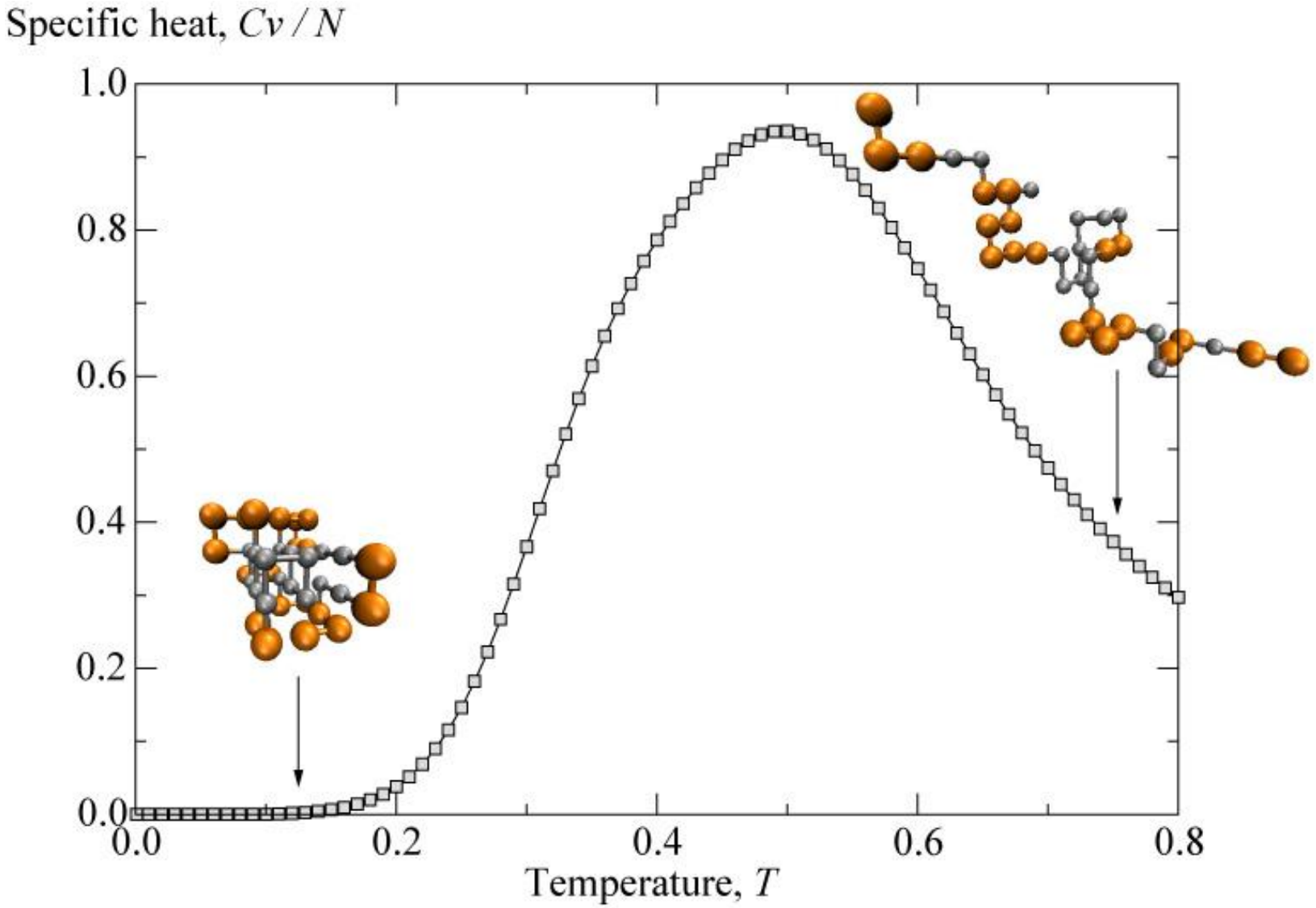
$$E = -n_{HH}\varepsilon_{HH}$$

n_{HH} : number of hydrophobic monomer-monomer pairs

ε_{HH} : attractive strength between H monomers

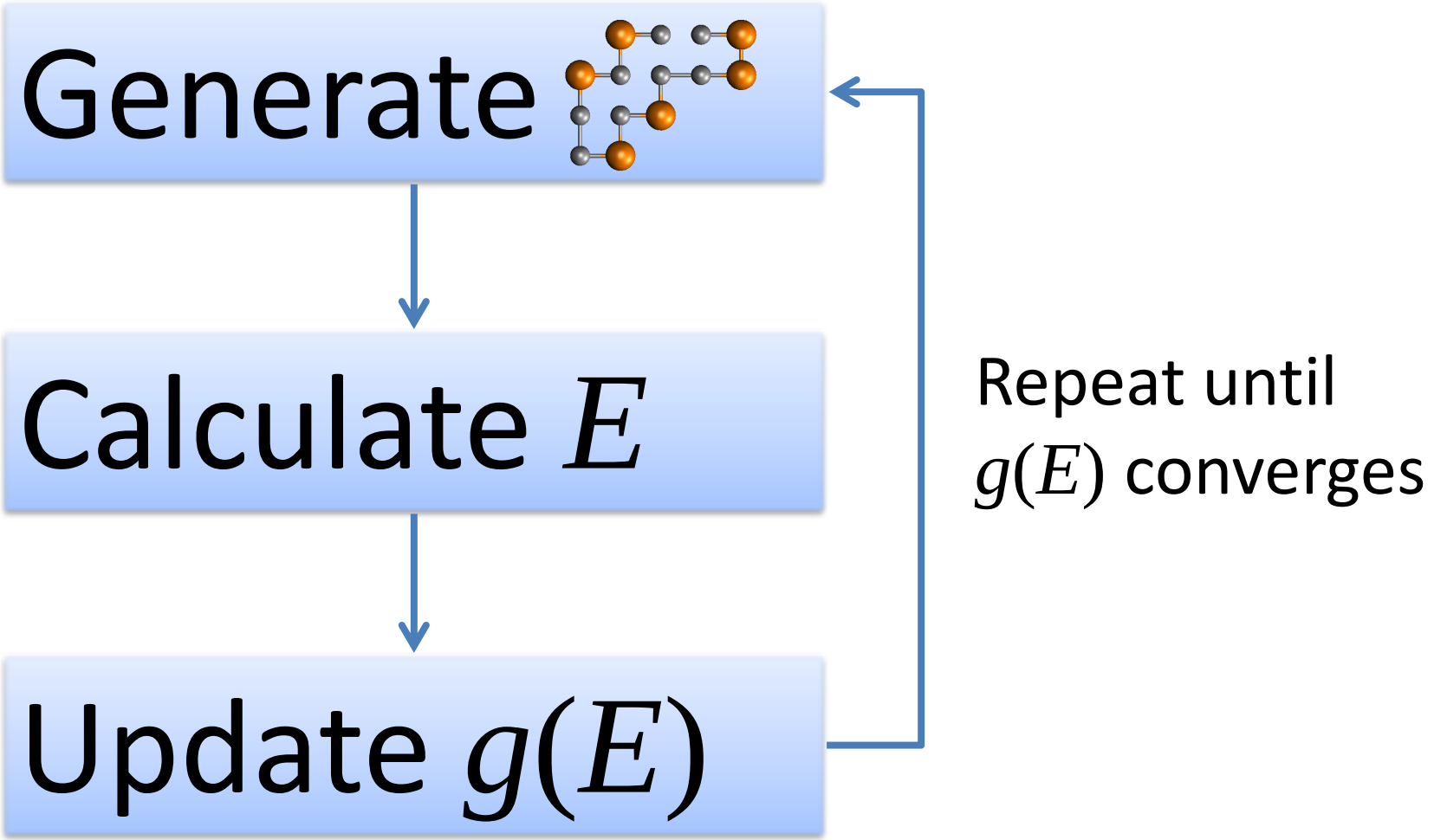
K. A. Dill, Biochemistry **24**, 1501 (1985); K. F. Lau & K. A. Dill, Macromolecules **22**, 3896, (1989)

Specific heat as a signal of transitions



*Error bars are smaller than the size of data points.

Serial simulation flow chart



Parallelization

Algorithm:

Parallelize Wang-Landau sampling

Model:

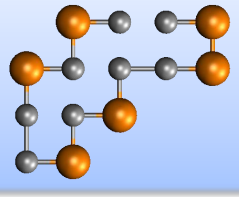
Energy calculation

(and other structural parameters)

~~Serial~~ simulation flow chart

ar al l e

l

Generate 

Calculate E

Update ~~$g(E)$~~

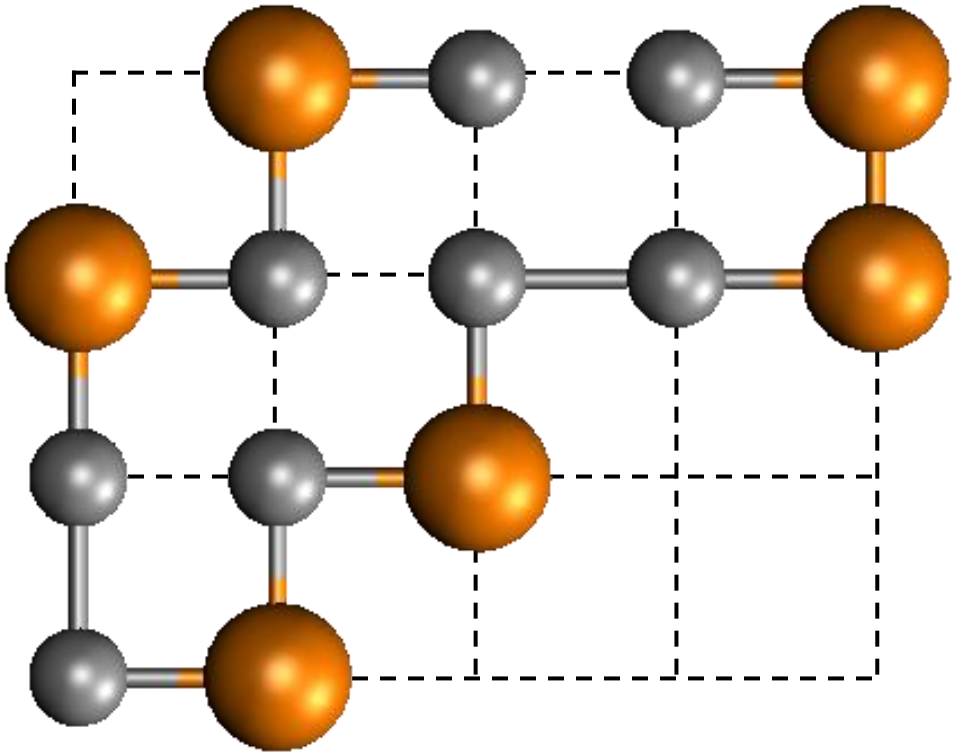
Repeat until $g(E)$ converges

use many $g(E)$ in a simulation

Parallelized energy calculation

To find n^i_{HH}:

- One thread per monomer
- Each thread examines 2×D nearest neighbors



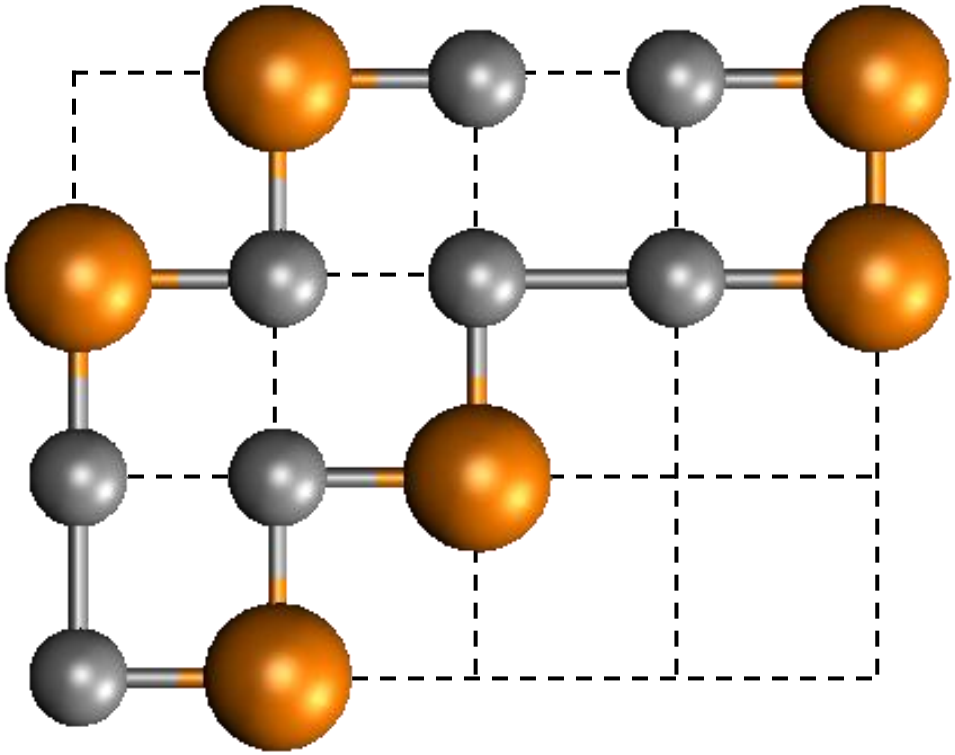
$$n_{HH} = \frac{1}{2}\sum_{i=1}^{N} n^i_{HH}$$

$$E = -n_{HH}\varepsilon_{HH}$$

Parallelized energy calculation

To find n^i_{HH}:

- 2×D threads per monomer
- Each thread examines only one nearest neighbor



$$n_{HH} = \frac{1}{2}\sum_{i=1}^{N} n^i_{HH}$$

$$E = -n_{HH}\varepsilon_{HH}$$

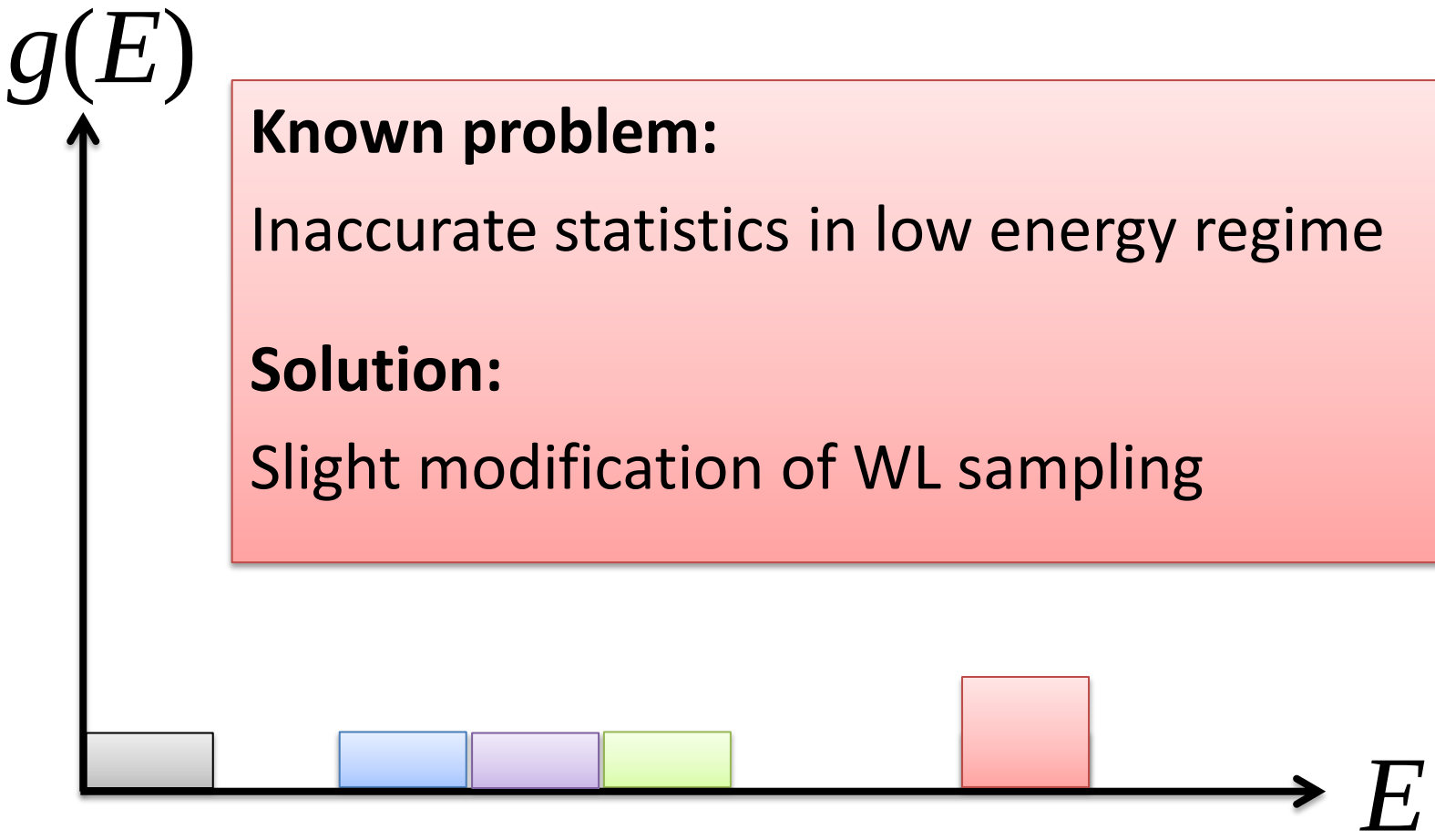
Parallelized Wang-Landau sampling (Scheme 1)

Multiple random walkers in single $g(E)$

J. Yin & D. P. Landau, J. Chem. Phys. **134**, 074501 (2011)

Parallelized Wang-Landau sampling (Scheme 1)

Multiple random walkers in single $g(E)$



Known problem:

Inaccurate statistics in low energy regime

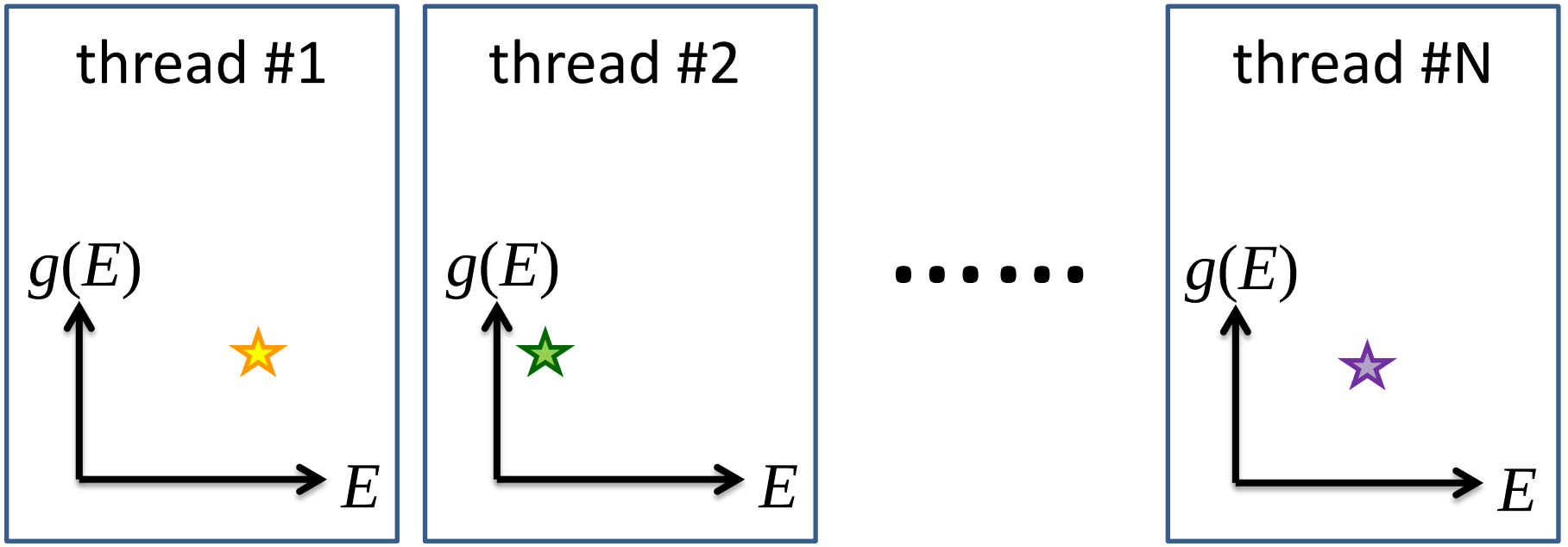
Solution:

Slight modification of WL sampling

J. Yin & D. P. Landau, J. Chem. Phys. **134**, 074501 (2011)

Parallelized Wang-Landau sampling (Scheme 2)

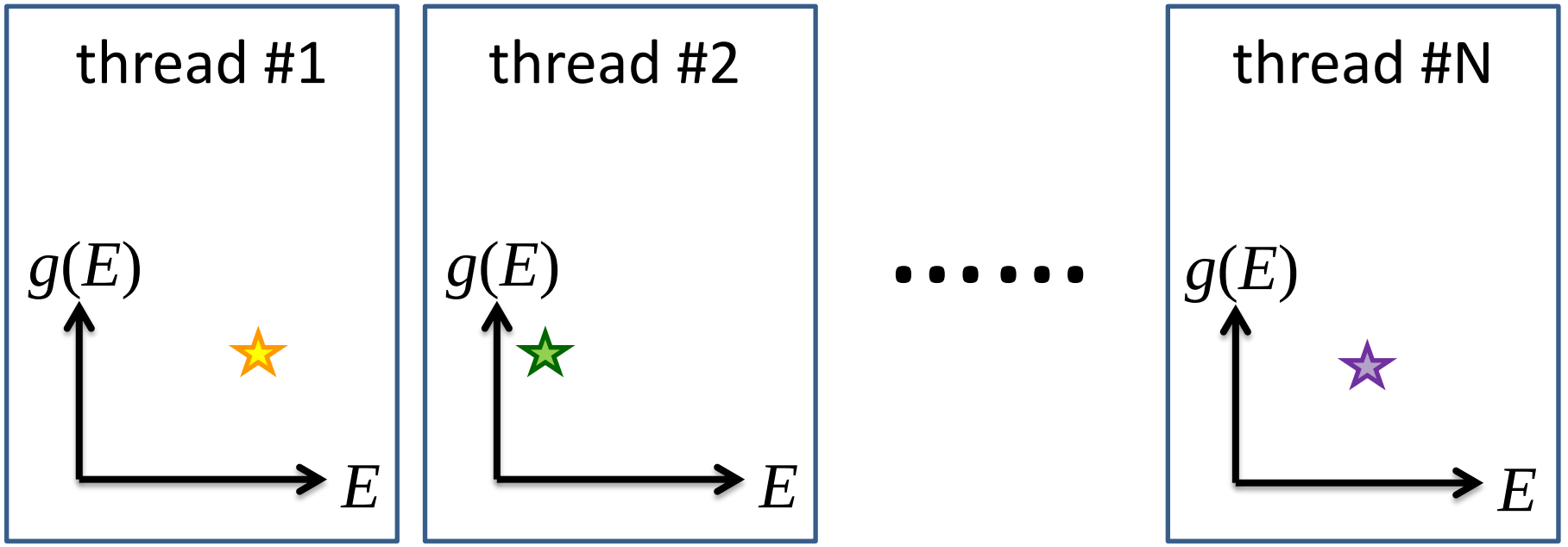
Multiple $g(E)$ in single simulation



In progress…

Parallelized Wang-Landau sampling (Scheme 2)

Multiple $g(E)$ in single simulation

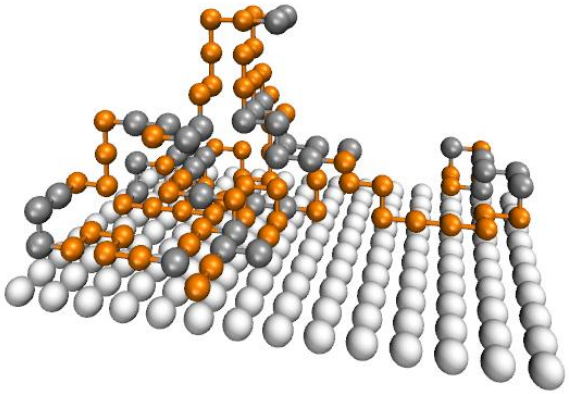


In progress…

More current (and future) work

Parallel calculations of structural parameters

$$R_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\vec{r}_i - \vec{r}_{cm})^2}$$



HP protein adsorption

Combination of parallel schemes ?

Parallel Monte Carlo trial moves ?