Enabling Biomolecular Simulations on Peta-Scale Computers

Elena Akhmatskaya and Ross Nobes
Fujitsu Laboratories of Europe
Hayes Park Central, Hayes UB4 8FE, United Kingdom

Akira Asato and Yoshie Inada
Peta-Scale Computing Research Center, Fujitsu Laboratories
4-1-1 Kamikodanaka, Nakahara-ku,

rotein folding: How do proteins fold?

Misfolding: Alzheimer's disease, cystic fibrosis, BSE, emphysema, arkinson's disease, Huntington's disease, cancer and cancer relate yndromes

rotein design: Predict sequence given structure

umanised antibodies, novel enzymes, biosensors

igand binding

ational drug design

rotein structure prediction: Predict structure given sequence

Structural genomics, pharmaceutical industry

he key to effective cancer treatment is to design drugs that can ei tabilise the normally folded structure or disrupt the pathway that o a misfolded protein

...y is Biomolecular Simulation So Hard?

...iological molecules are complex

...Example:

each amino acid residue: 3 possible conformations

100-residue protein $3^{100} = 10^{47}$ configurations

...iological processes are "slow"

...n extreme case of multi-scale simulation

fast bond vibrations and slow folding processes in a single

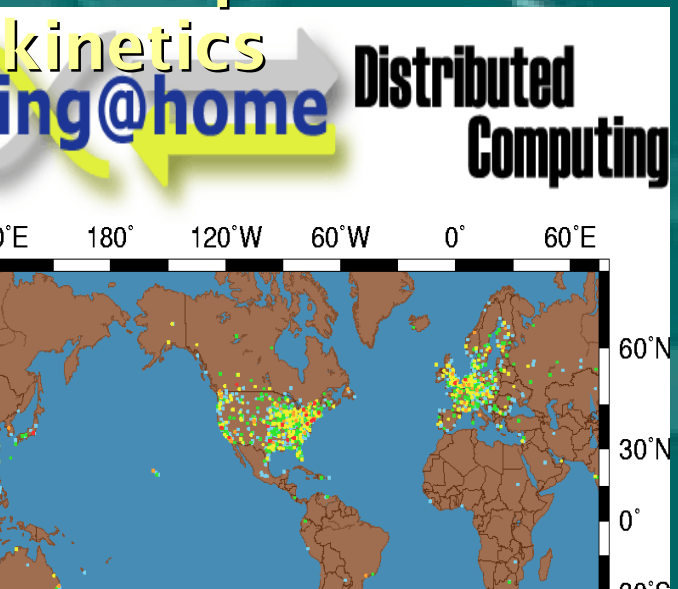
's Blue Gene Project

nounced in 1999 with goal of building peta-
s computer to address the grand challenge of
tein folding

ue Matter: software system to run MD on highly
allel machines

· 2004: study of beta-hairpin folding (5239
atoms)

· Independent short MD runs to simulate



Folding@Home

- Initiated at Stanford University as a worldw
grid computing project

· 2001: the longest ever simulation for 3
microseconds of beta-hairpin (177 ato
was performed on ~ 5000 CPUs

hallenge in simulations of physical systems: proper sampling of
ble conformations over accessible simulation times

ling is important for: • Collecting statistics • Finding the globa
num

ficient sampling is the greatest *source of inaccuracy* in computatio
gy

lar sampling methods

MD *Molecular Dynamics*

MC *Monte Carlo*

HMC *Hybrid Monte Carlo*

resent

SHMC *Targeted Shadowing Hybrid Monte Carlo*

Developed in Fujitsu Laboratories of Europe in collaboratio
with Prof. S. Reich (University of Potsdam)

E. Akhmatskaya, S. Reich, "The Targeted Shadowing Hybrid Monte Carlo (T
Method",

Ideas Behind TSHMC

TSHMC combines *the best features* of three popular methods: molecular dynamics, Monte Carlo and dissipative particle dynamic (DPD)

TSHMC uses *high-order* approximations to *modified Hamiltonians* improve sampling

TSHMC provides a *flexible optional momentum update*:

- *complete update*: for sampling only
- *partial update*: to reproduce stochastic Langevin dynamics
- *selective partial update*: for application to mesoscale dynamics simula

Due to its great flexibility, TSHMC can be *easily combined with* one/several *enhanced sampling methods* (e.g. replica exchange, ju walking, etc.) to escape from energy barriers and sample even wid

Perform?

MD

Fast exploration of phase space

TSHMC

HMC

MC

Good control over statistical error

TSHMC

HMC

MD

Applicable to biomolecules

TSHMC

MC

Massively parallel computation

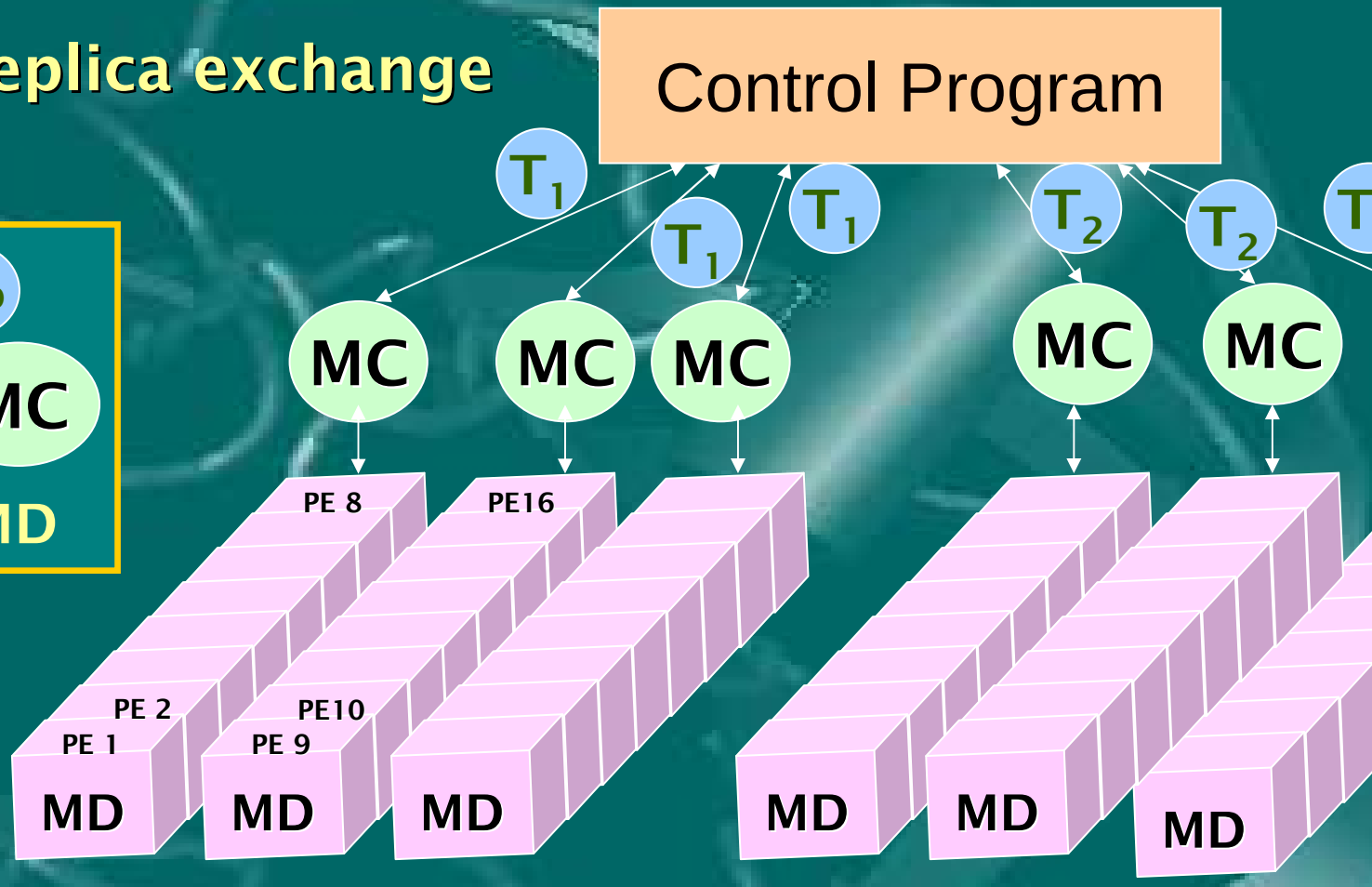
HMC

TSHMC

Enhanced Sampling Methods

ous schemes for parallelising TSHMC in combination with enhance
pling methods on massively parallel computers are feasible

mple: TSHMC + single replica exchange

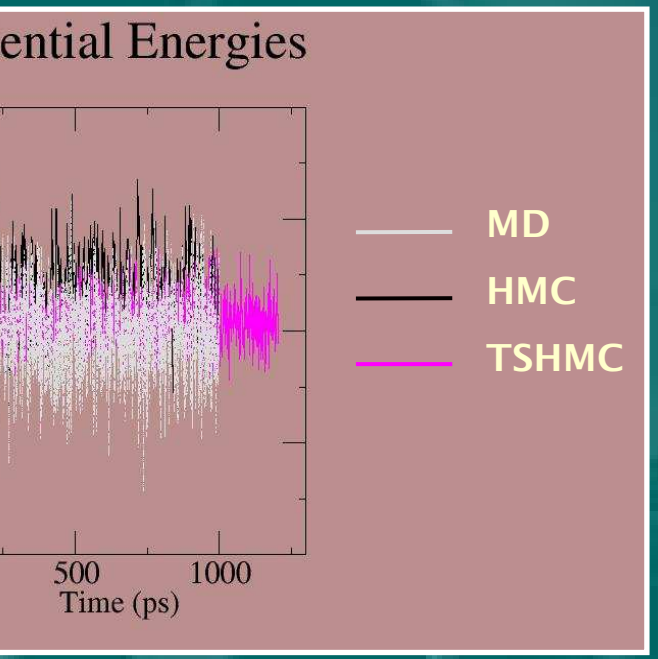


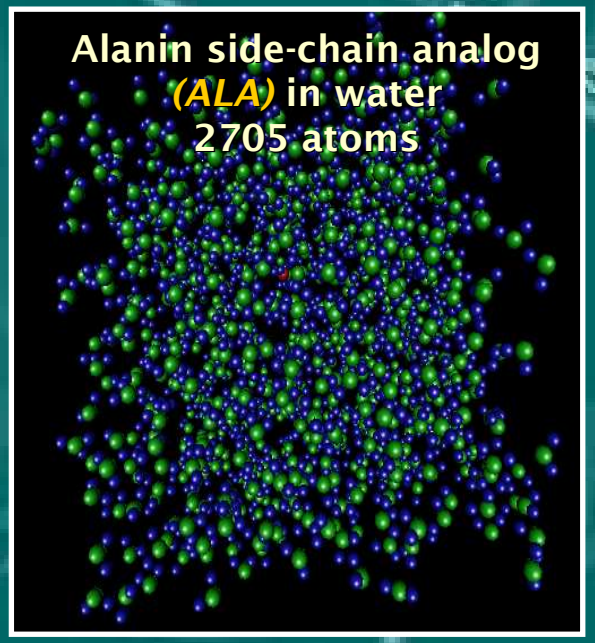
degree of parallelism

- Minimal communication and synchronisation is required
- Parallelisation strategies are flexible: adjustable to hardware
- Benefits from a large number of processors

potential, solvation energies of ALA predicted by MD, HMC and TSHMC are in goo
ment

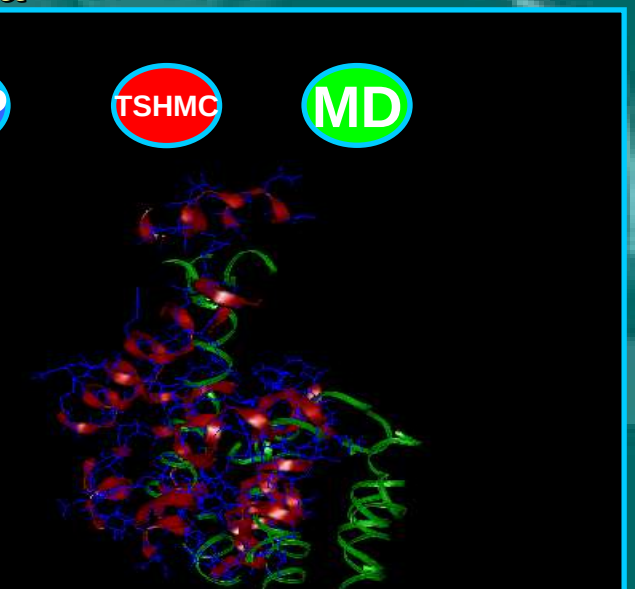
C demonstrates much *smoother energy profile* due to high-order energy conserva

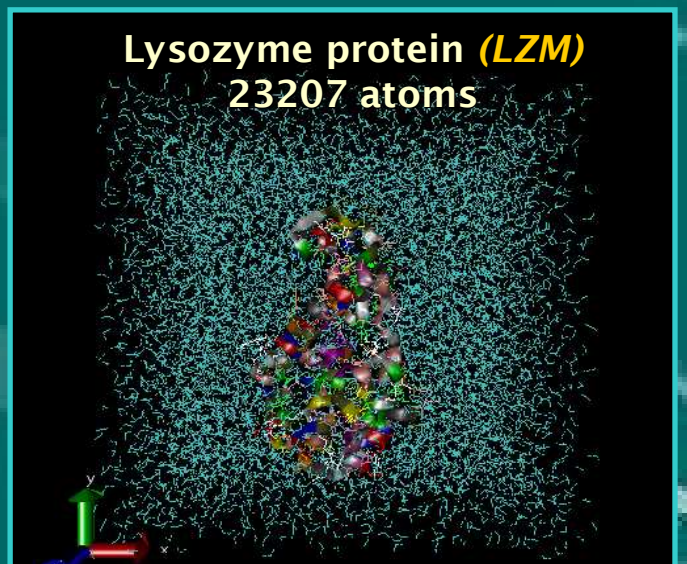


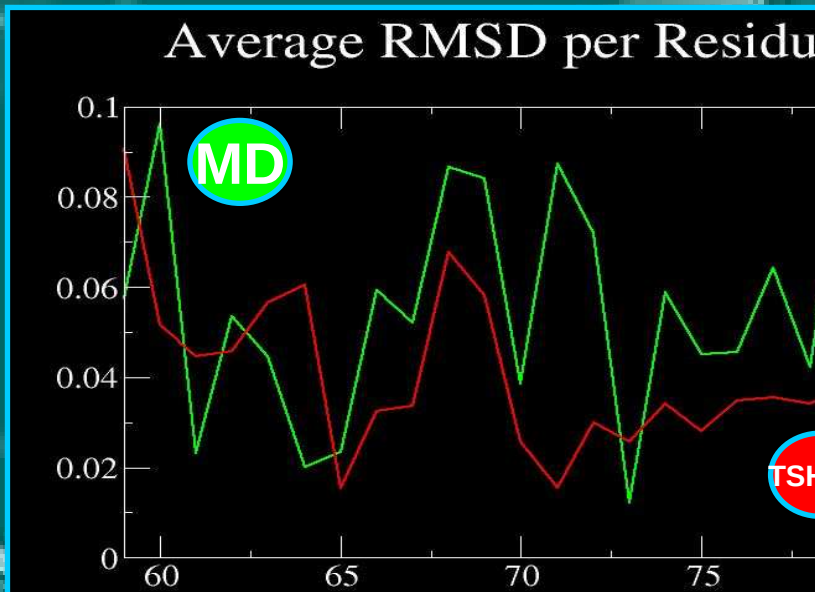


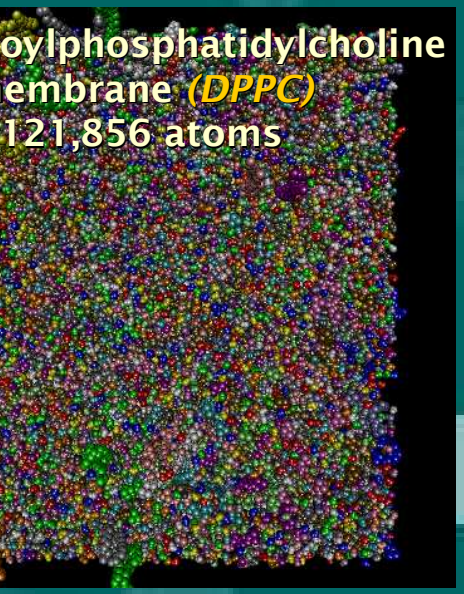
	MD	HMC	TSHM
Solvation energy, kcal/mol	1.62	1.60	1.67

ural properties of LZM from TSHMC are in better agreement with experiment than a
a







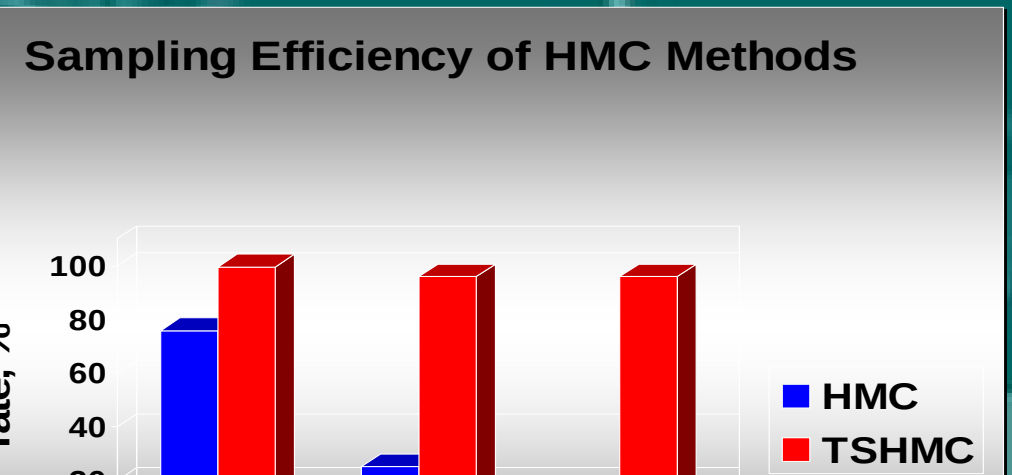


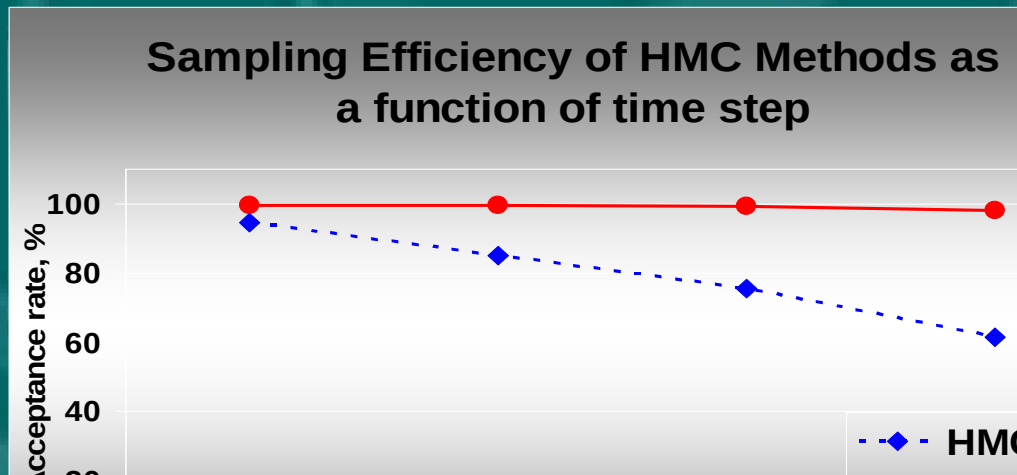
Comparison of *integrated autocorrelation functions for D* simulation produced by MD, HMC and TSHMC shows that *TSHMC is at least ~5 times more efficient than MD* and ~ times more efficient than HMC in terms of generating independent samples

HMC demonstrates *stable high mpling efficiency* for all simulated stems

mpared with HMC, *TSHMC increases number of sampled configurations* a factor of up to 40

Sampling efficiency of TSHMC is h even for large time steps whereas efficiency of HMC decreases quick with increasing time step





Conclusions

We propose a new simulation methodology base
on
a hybrid of Monte Carlo, molecular dynamics an
dissipative particle dynamics and called Targete
Shadowing Hybrid Monte Carlo (TSHMC)

TSHMC offers a rigorous, flexible and efficient
approach to conformational sampling in
biomolecular simulation

TSHMC is extremely well suited to massively
parallel computation and can be run efficiently
next generation peta-scale computers