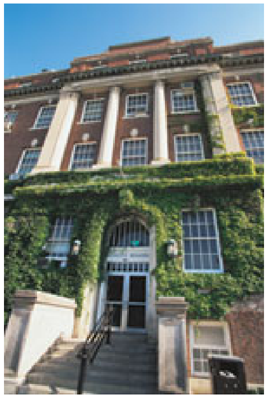
Coarse-graining in Soft Materials
A Beginner's Guide to Fundamental Challenges, Limitations, & Opportunities

Shekhar Garde
Department of Chemical & Biological Engineering
Rensselaer Polytechnic Institute, Troy, NY.


JNCASR Bengaluru, 12/06/2009

My bio?

Born in Kolhapur, Maharashtra. Went to school till the 12th grade there.

1988-1992 – B. Chem. Eng. UDCT, Bombay

1992-1997 – Ph.D., Chemical Engineering, Univ. Delaware

1997-1999 – Director’s Fellow, Los Alamos National Labs

(other than learning the art of biomolecular simulations, spent most of my time hiking, rock climbing, hiking the Grand-Canyon many different ways, etc.)

1999- present Rensselaer Polytechnic Institute

Elaine and Jack S. Parker Endowed Chair in Engineering,

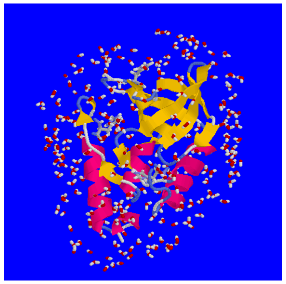
Head of the Chemical & Biological Engineering

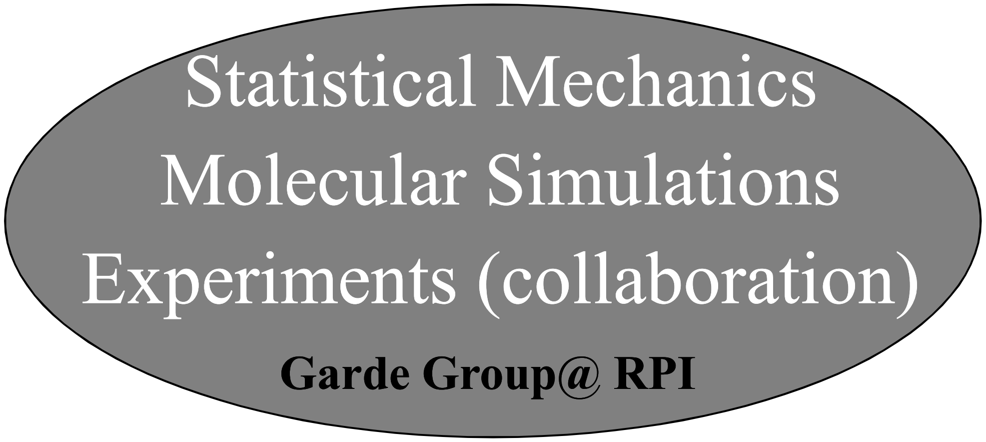
Biological Systems

- Protein Folding/stability
- Peptide Folding/structure
- Bioinformatics
- Biomaterials (enzymes in novel media)
- Mechanistic Studies (intein splicing)

Water & aqueous solutions

- Liquid state theory
- Water structure near solutes, ions, interfaces
- Water-mediated interactions



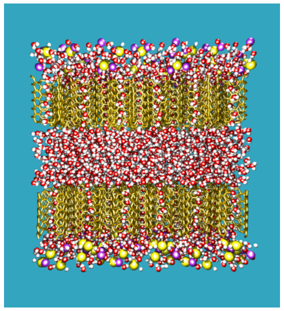


Statistical Mechanics

Molecular Simulations

Experiments (collaboration)

Garde Group@ RPI



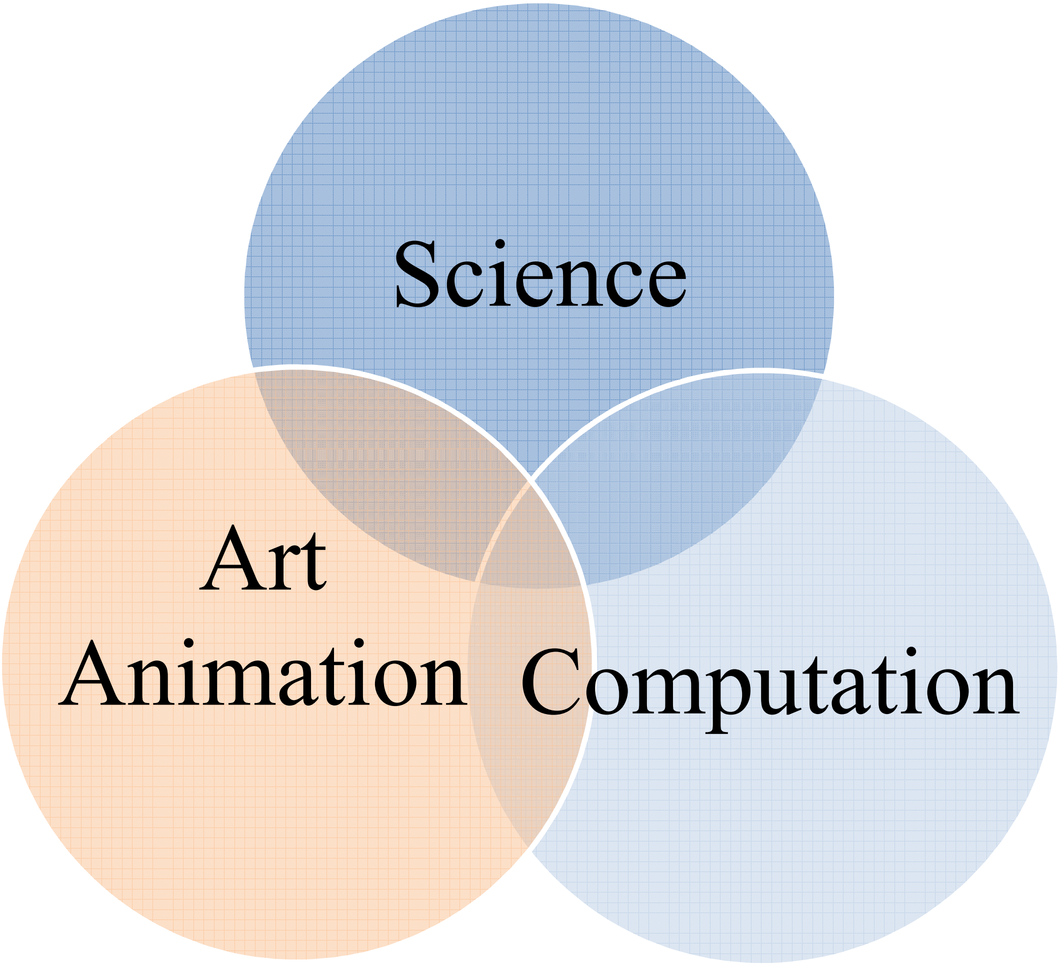
Polymers/multi-scale modeling

- Coarse-graining strategies
- Phase behavior

Nanosystems

- Water flow through CNTs
- Selective partitioning into CNTs
- Friction at the nanoscale

Molecularium

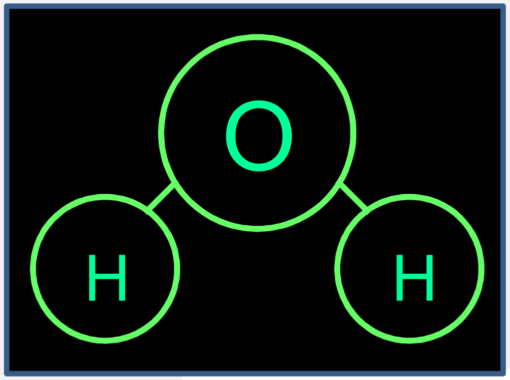


Science

Art
Animation

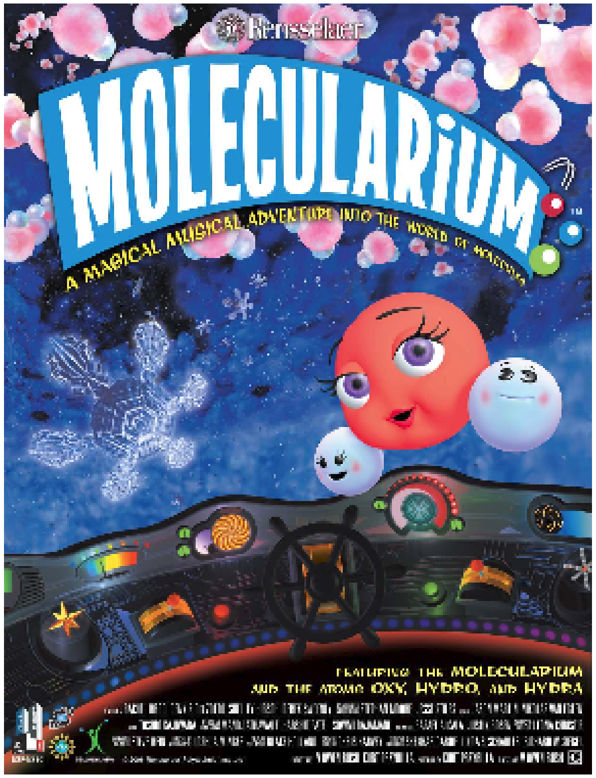
Computation

Molecularium



Molecularium…its evolution

- Fundamental scientific concepts
- Narrative: exciting story & characters
- Collaboration



Two Complex biological systems

Ahiri and Kaveri

A beginner's guide

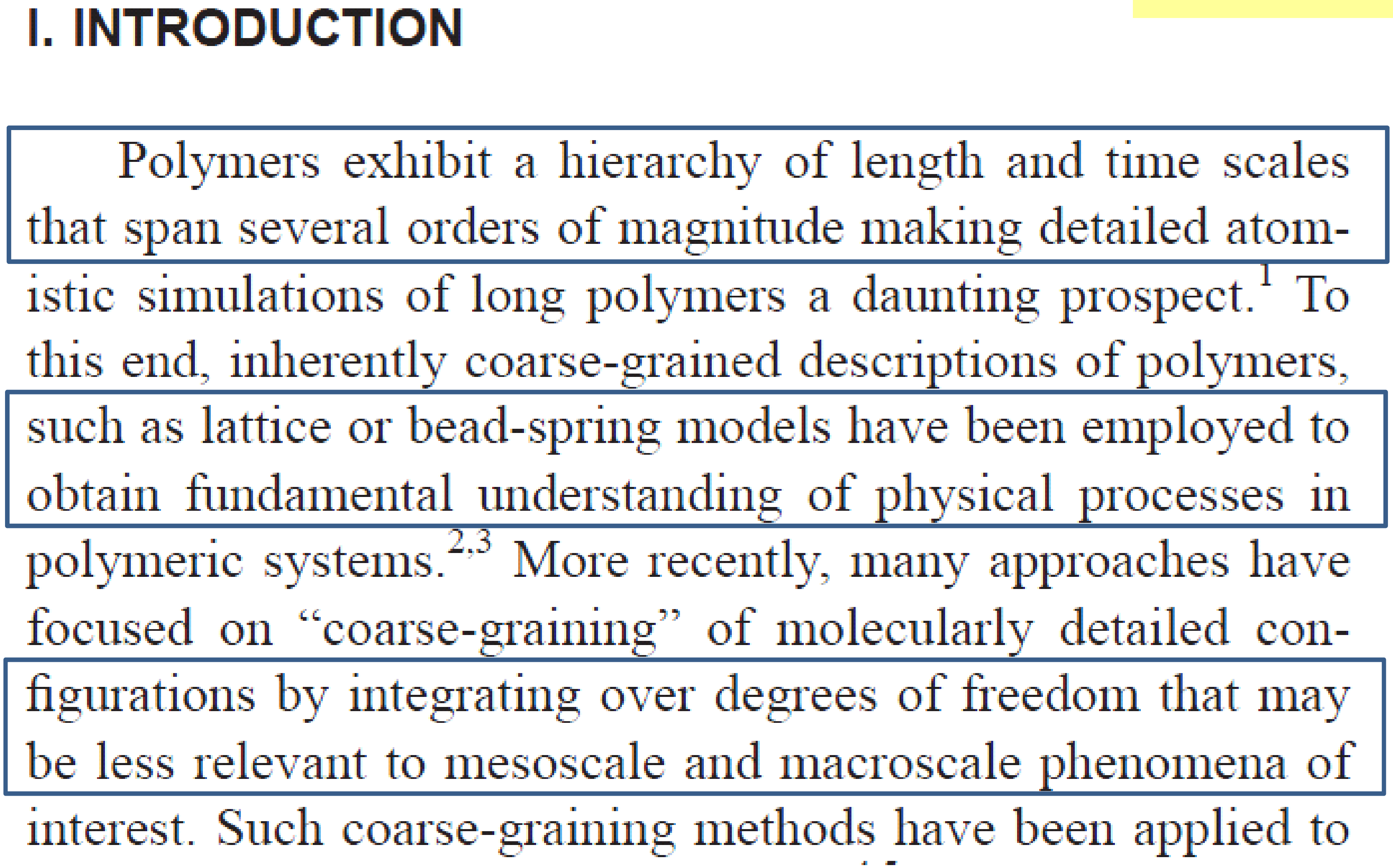
- Refer's to me!
- Not for experts.
- Focus will be on basic concepts.

(things that one should know but may not)

Challenges	Limitations	Opportunities
Mostly of conceptual kind.	What CG can't do for you.	Many exciting problems are waiting to be solved.

Why coarse-grain?

Ashbaugh,
Garde, et al.
JCP, 2005.

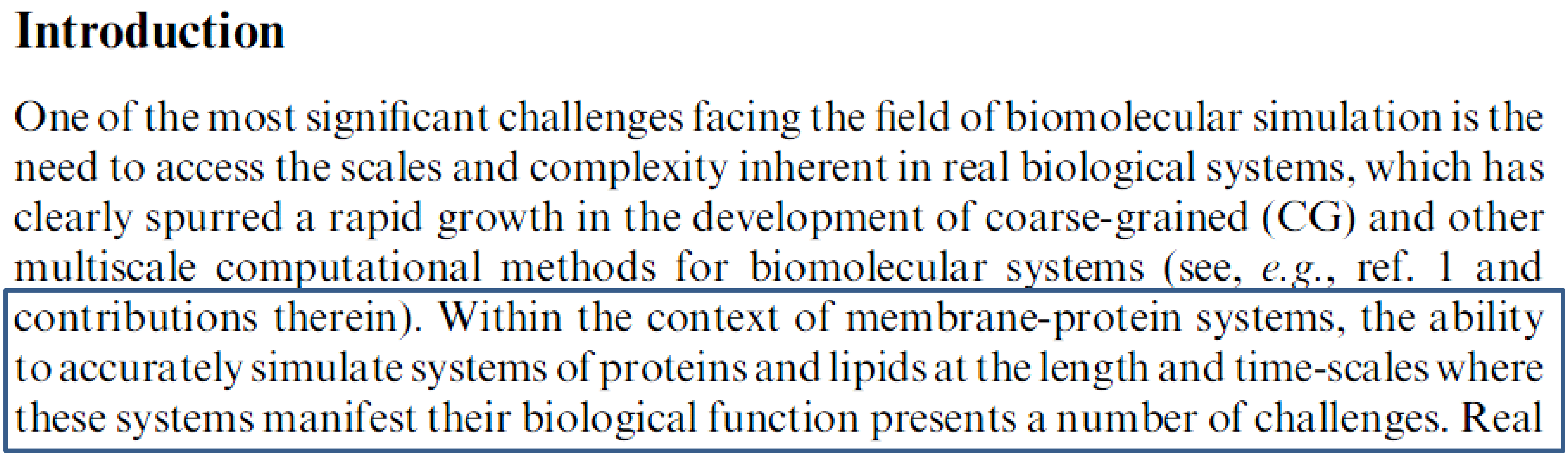


I. INTRODUCTION

Polymers exhibit a hierarchy of length and time scales that span several orders of magnitude making detailed atomistic simulations of long polymers a daunting prospect.[1] To this end, inherently coarse-grained descriptions of polymers, such as lattice or bead-spring models have been employed to obtain fundamental understanding of physical processes in polymeric systems.[2,3] More recently, many approaches have focused on "coarse-graining" of molecularly detailed configurations by integrating over degrees of freedom that may be less relevant to mesoscale and macroscale phenomena of interest. Such coarse-graining methods have been applied to

Why coarse-grain?

Faraday Discussion 144 (2010)



Introduction

One of the most significant challenges facing the field of biomolecular simulation is the need to access the scales and complexity inherent in real biological systems, which has clearly spurred a rapid growth in the development of coarse-grained (CG) and other multiscale computational methods for biomolecular systems (see, *e.g.*, ref. 1 and contributions therein). Within the context of membrane-protein systems, the ability to accurately simulate systems of proteins and lipids at the length and time-scales where these systems manifest their biological function presents a number of challenges. Real

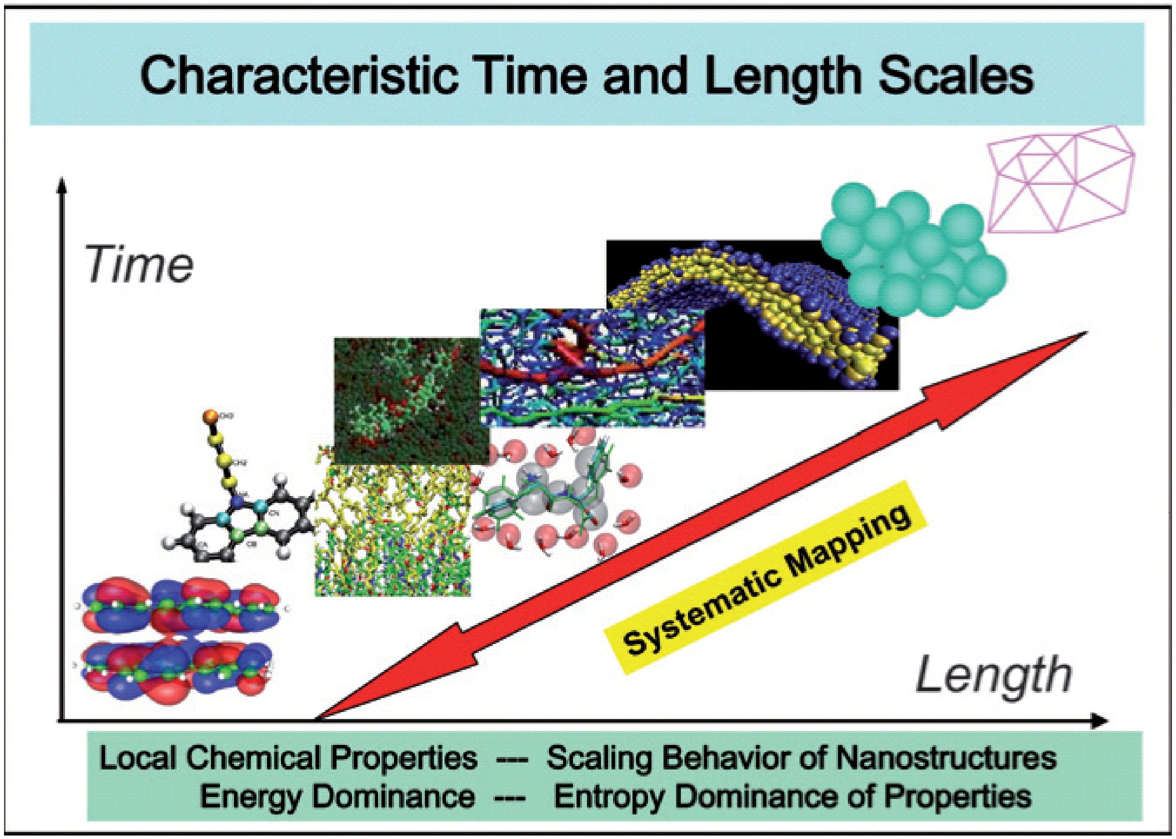
Greg Voth's paper in Faraday Disc.

Why coarse-grain?

Kremer, K., et al.
Faraday
Discussions.

1 Introduction

Material properties of soft matter systems are determined by processes and interactions on a wide range of length and time scales. While these mutually influence each other, it is not straight forward to provide quantitative information and understanding without taking this properly into account. Although this holds for many



Motivation

- Many lengthscales…
- Many timescales…
- Integrating over unimportant degrees of freedom

Multi-scale methods

- Coarse-graining (and fine-graining)
- Hybrid methods
- Variable adaptive resolution methods

- CG is not new. Has been around for centuries!?
  -- application of Newton’s laws to
  predict trajectories
  -- continuum treatments of matter
  (Navier Stokes Eqn.; Diffusion Eqn.)

- Lets think in the modern context
  -- CG not applied to gases
  (low density…easy computations)
  -- continuum treatments of matter
  (Navier Stokes Eqn.; Diffusion Eqn.)

Systems of Interest

- Gases? -- in a sense, doing detailed single molecule QM calculation and then applying stat mech ideas can be thought of as early multi-scale modeling. (These days we don't typically worry about CG of gases.)

- Liquids/soft-condensed matter

  CG is being developed and applied to obtain:

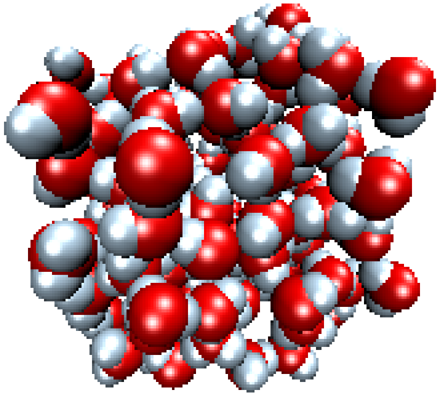
  - ➢ Structure
  - ➢ Thermodynamics
  - ➢ Dynamics

Structure

- RDFs, S(k)

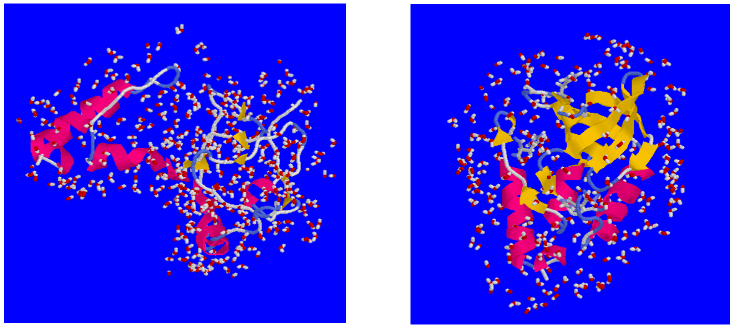
  atomic, molecular liquids, mixtures

  self-assembly in colloidal systems

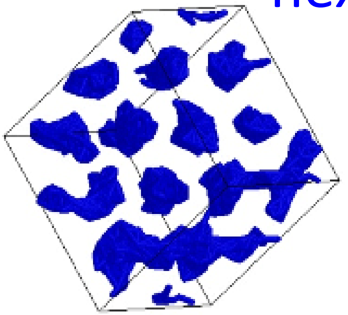


- Conformations:

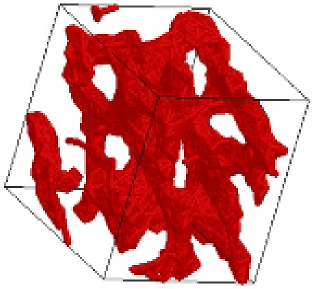
  proteins, polymers



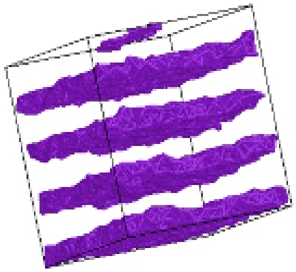
- Microstructure



hexagonal



bicontinuous



lamellar

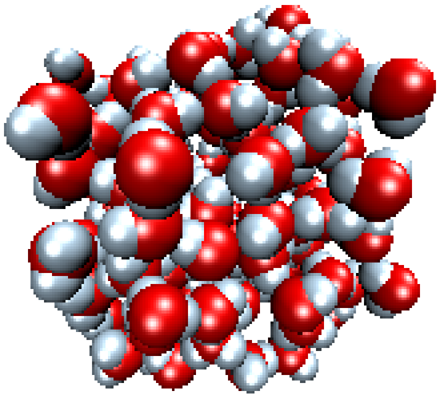
Does my CG approach give me the right structure?

Lets think more about structure

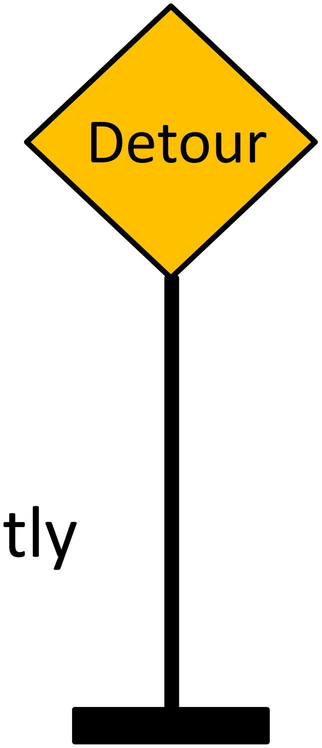
- RDFs, S(k)

  atomic, molecular liquids, mixtures

  self-assembly in colloidal systems



- What are $g^{(2)}$, $g^{(3)}$, …, $g^{(N)}$ correlation functions?



2 Silly questions:

- Does a fluid interacting with pair-wise interactions have $g^{(3)}$?
- Can the information about $g^{(3)}$ be derived exactly from $g^{(2)}$? What about the other way?

Simple terminology

- Hamiltonian

$$E = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right]$$

- Pair-wise additivity
- Manybody interactions
- Spherically symmetric vs anisotropic interactions

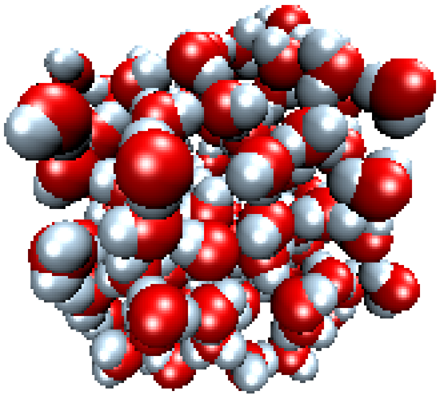
(and how might these come about)

Lets think more about structure

- RDFs, S(k)

  atomic, molecular liquids, mixtures

  self-assembly in colloidal systems



- What are $g^{(2)}$, $g^{(3)}$, …, $g^{(N)}$ correlation functions?

2 Silly questions:

- Does a fluid interacting with pair-wise interactions have $g^{(3)}$?
- Can the information about $g^{(3)}$ be derived exactly from $g^{(2)}$? What about the other way?

Thermodynamics

U = internal energy

$H = U + PV$

$A = U - TS$

$G = H - TS$

- C_p, C_v, compressibility, thermal expansion, EOSv
- ΔG of hydration, folding, binding/association
- $\Delta\Delta G$ wrt pH, salt, mutation

Does my CG approach give me the right thermodynamics?

Humor me with a bit more Thermodynamics!

Do some simple thermomath here and remind about the following question.

Does my CG approach give me the right thermodynamics?

Does my CG approach give me the right thermodynamics?

Yes, if

U(S,V)

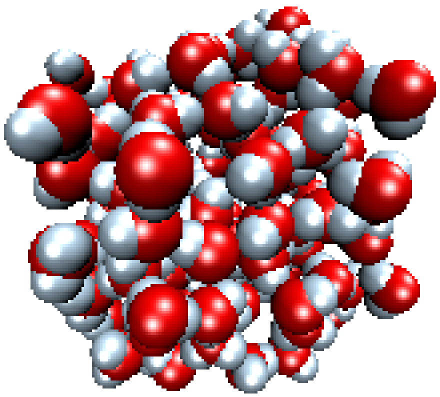
H(S,P)

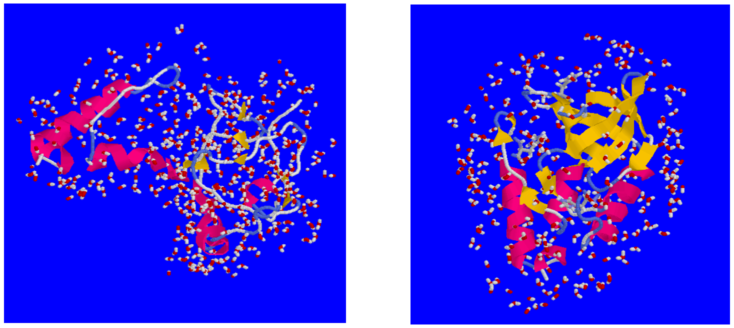
A(T,V)

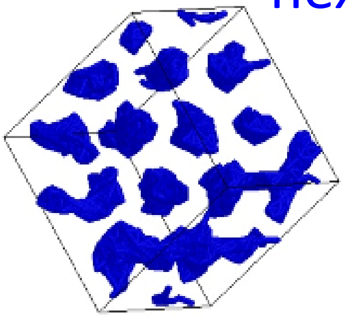
G(T,P) are right.

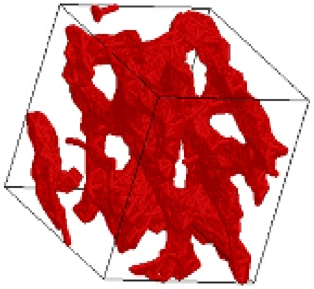
Dynamics

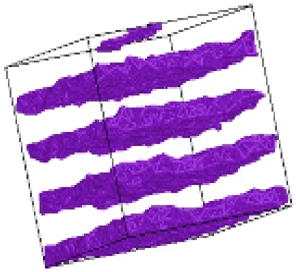
- Time correlation functions
- Relaxation times (for evolution of…)
- First passage times
- Slow moving vs fast moving
- Diffusivity
- Viscosity






hexagonal


bicontinuous


lamellar

Does my CG approach give me the right dynamics?

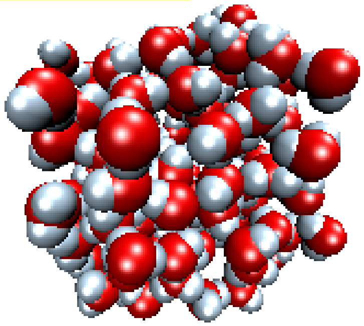
Lets go back to structure

Does my CG approach give me the right structure?

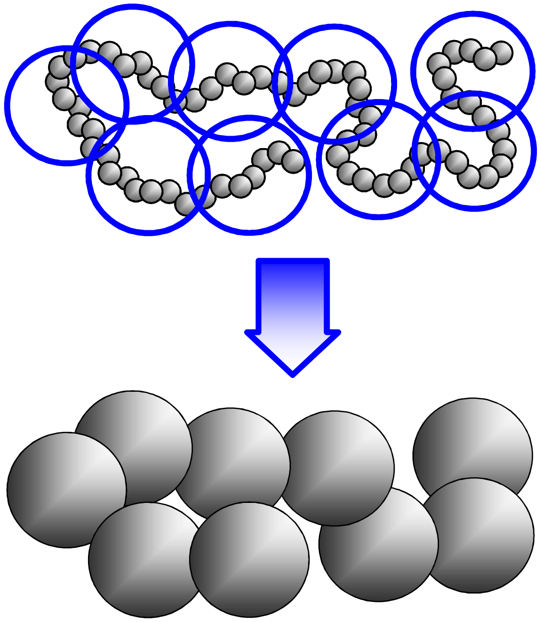
- RDFs, S(k)

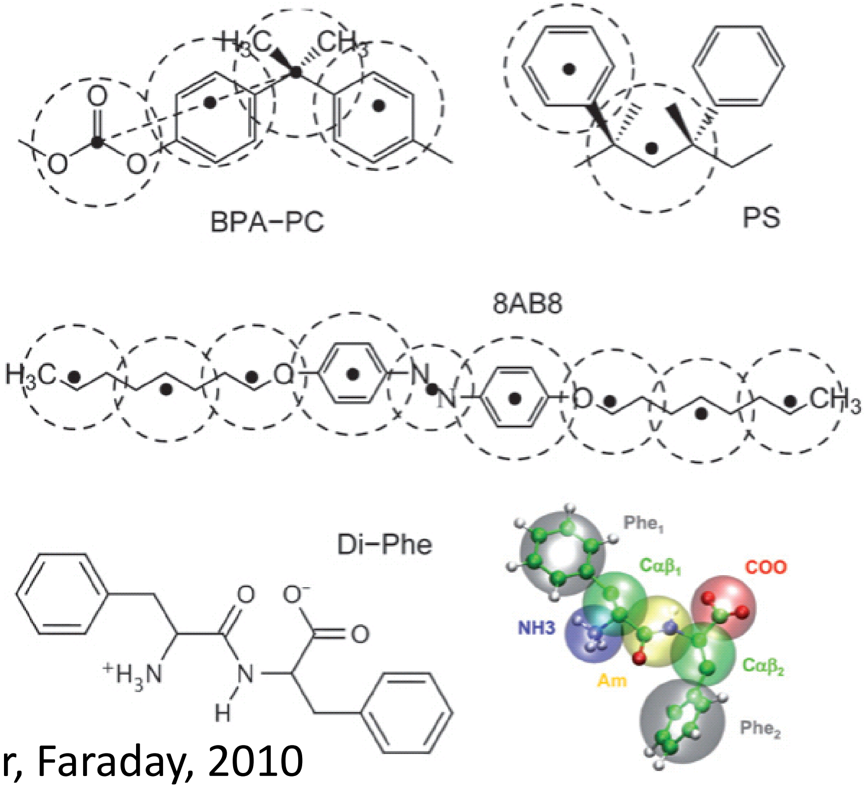
  atomic, molecular liquids, mixtures

  self-assembly in colloidal systems

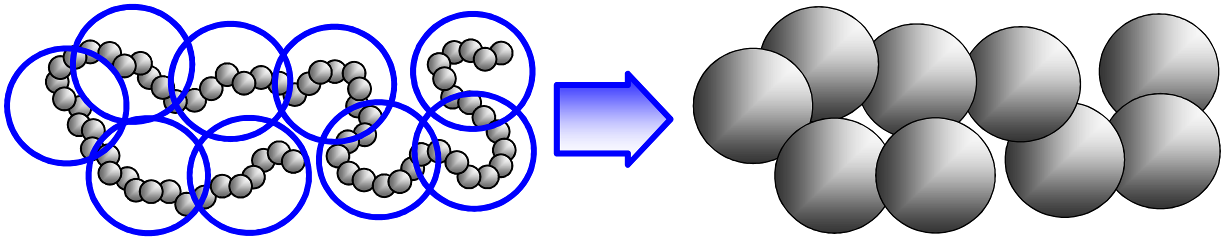


- A typical CG approach



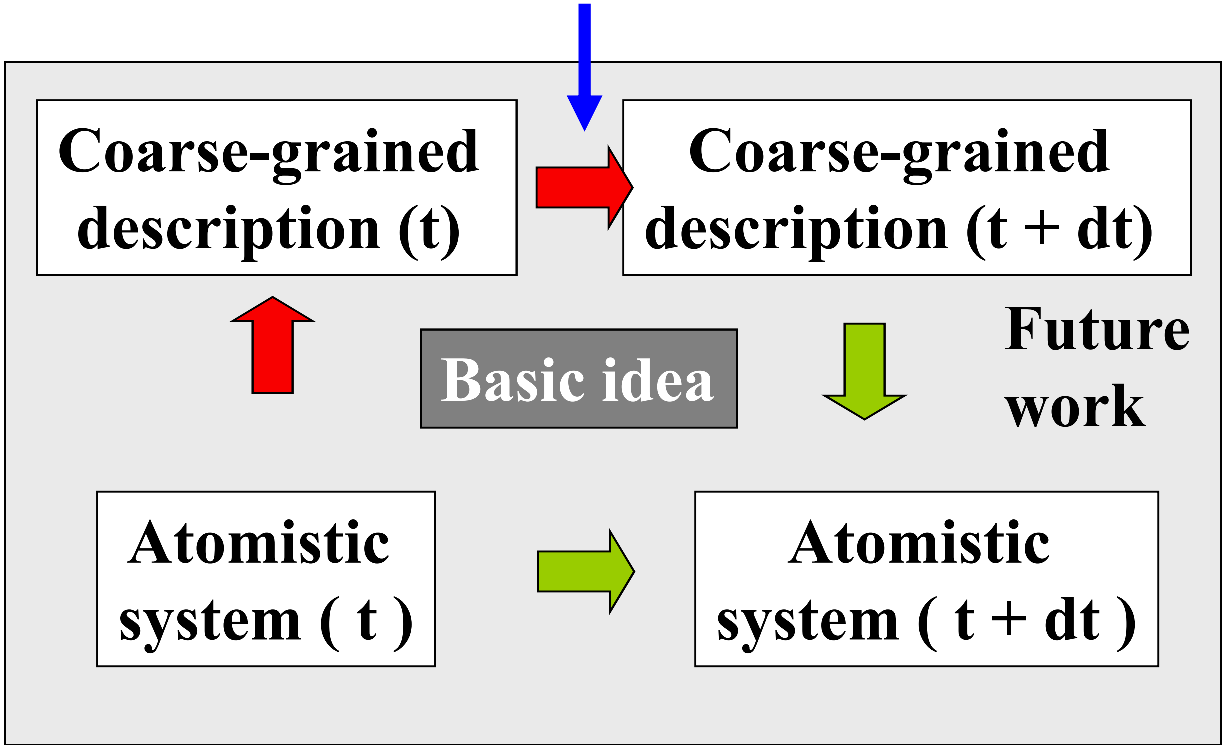


Kremer, Faraday, 2010



- CG unit selected heuristically (systematically?). No automated approach as far as I know. What is done next?

How do we derive physically consistent particle-particle interaction potentials?



Coarse-grained description (t)

Coarse-grained description (t + dt)

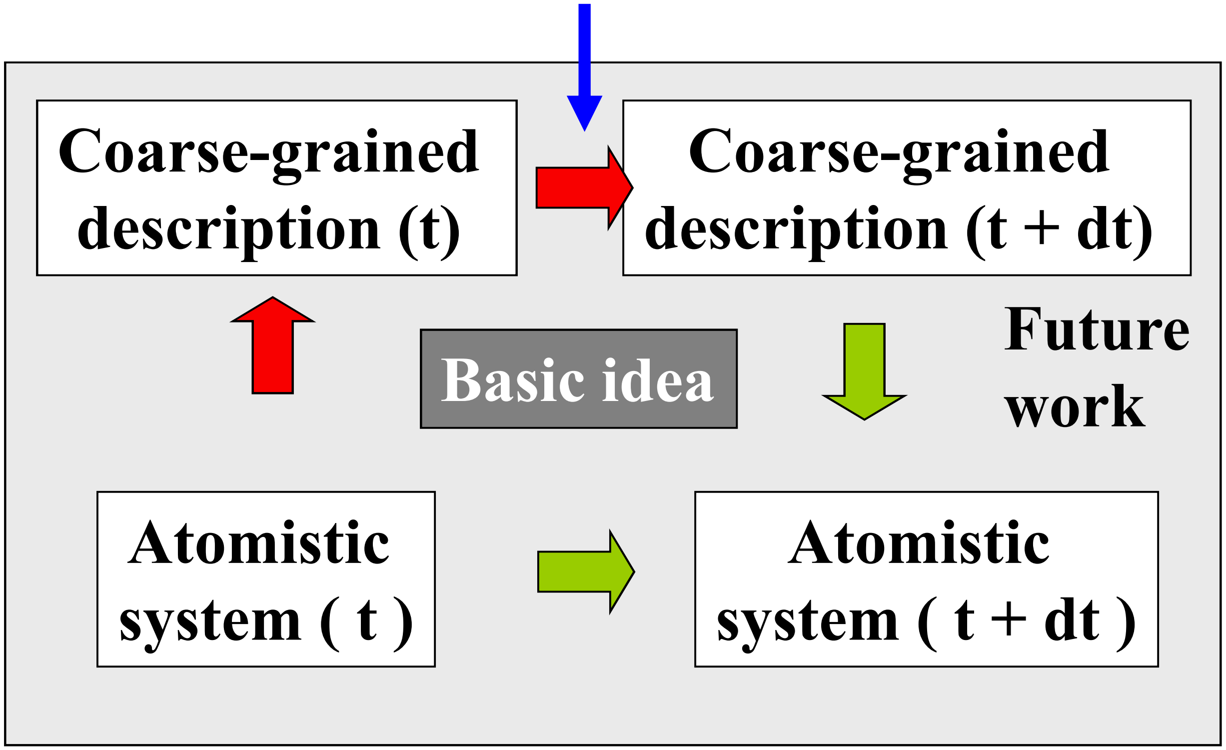
Basic idea

Future work

Atomistic system (t)

Atomistic system (t + dt)

How do we derive ~~physically~~ consistent particle-particle interaction potentials?



Coarse-grained description (t)

Coarse-grained description (t + dt)

Basic idea

Future work

Atomistic system (t)

Atomistic system (t + dt)

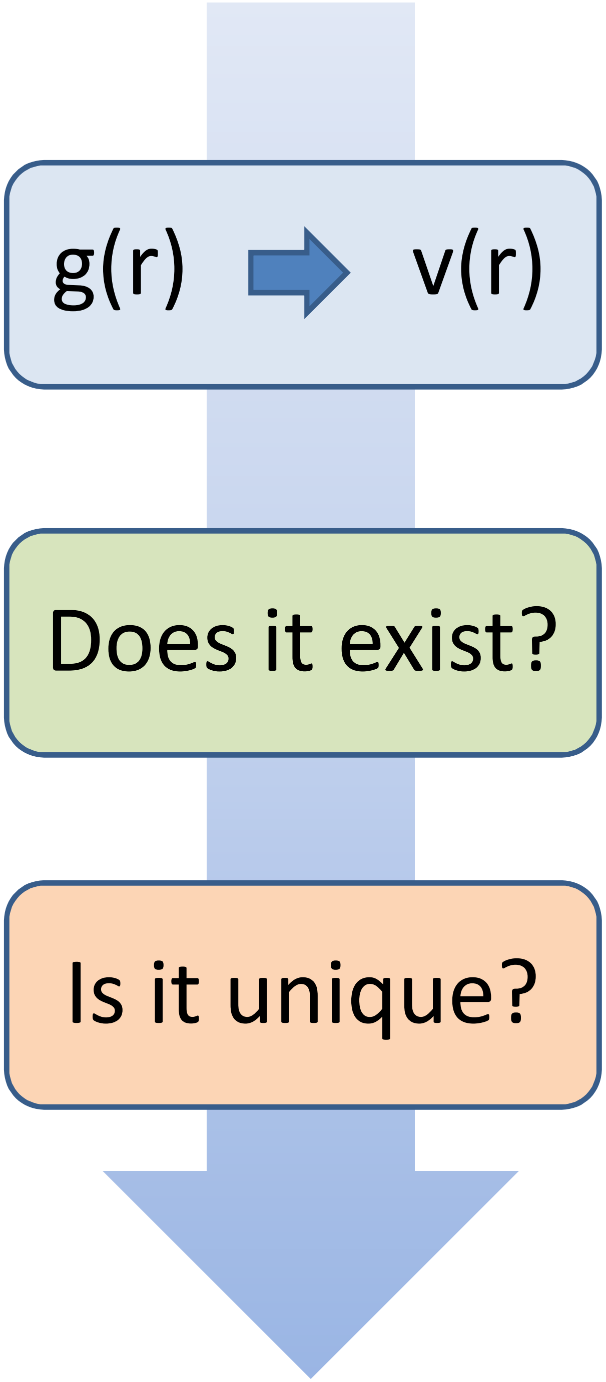
- Boltzmann = Inverse Boltzmann
- Reverse Monte Carlo (RMC or IMC)
- Force matching algorithm (Voth)

- Inverse Boltzmann
- Reverse Monte Carlo

Give T, density, and g(r) find v(r)

structure → potential

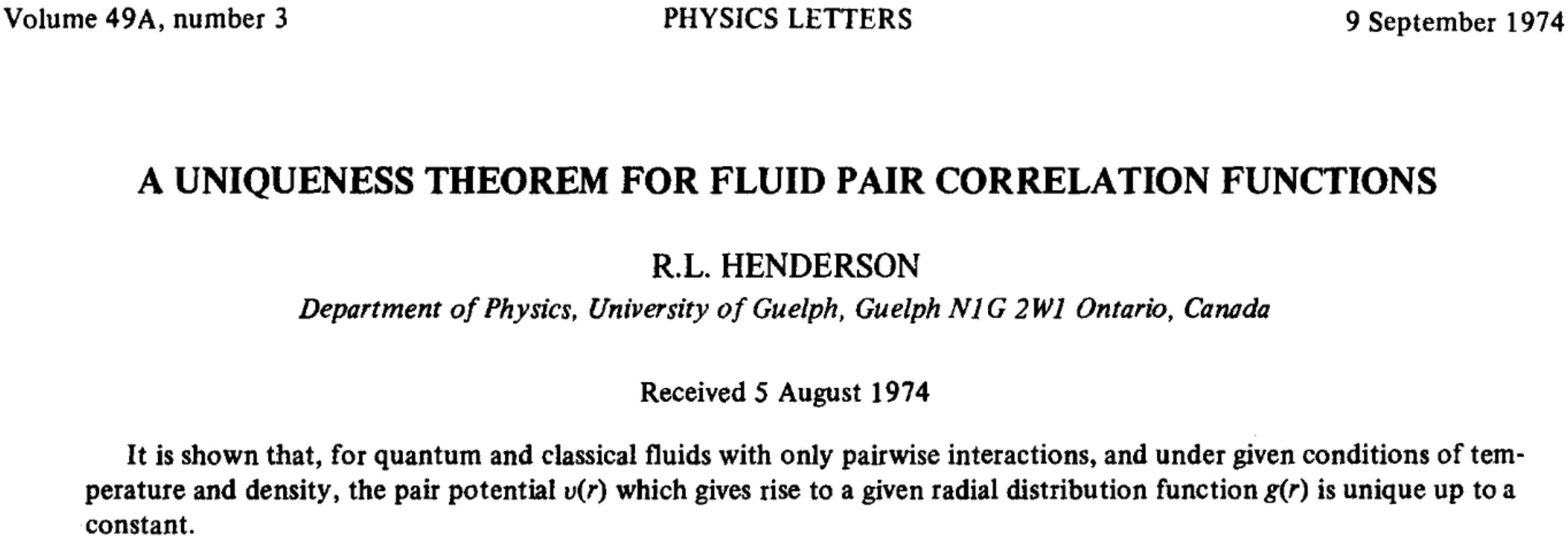
g(r) → v(r)



$g(r) \rightarrow v(r)$

Does it exist?

Is it unique?



Volume 49A, number 3 PHYSICS LETTERS 9 September 1974

A UNIQUENESS THEOREM FOR FLUID PAIR CORRELATION FUNCTIONS

R.L. HENDERSON
Department of Physics, University of Guelph, Guelph N1G 2W1 Ontario, Canada

Received 5 August 1974

It is shown that, for quantum and classical fluids with only pairwise interactions, and under given conditions of temperature and density, the pair potential $v(r)$ which gives rise to a given radial distribution function $g(r)$ is unique up to a constant.

It is shown that, for quantum and classical fluids with only pairwise interactions, and under given conditions of temperature and density, the pair potential v(r) which gives rise to a given radial distribution function g(r) is unique up to a constant.

The uniqueness theorem follows from an inequality for the Helmholtz free energy. We consider two systems under identical conditions of temperature, volume and density, and described respectively by Hamiltonians H_1 and H_2. Then, in both classical and quantum statistics, the corresponding free energies F_1 and F_2 are related by [5]

$$F_2 \leqslant F_1 + \langle H_2 - H_1 \rangle_1 \ , \tag{3}$$

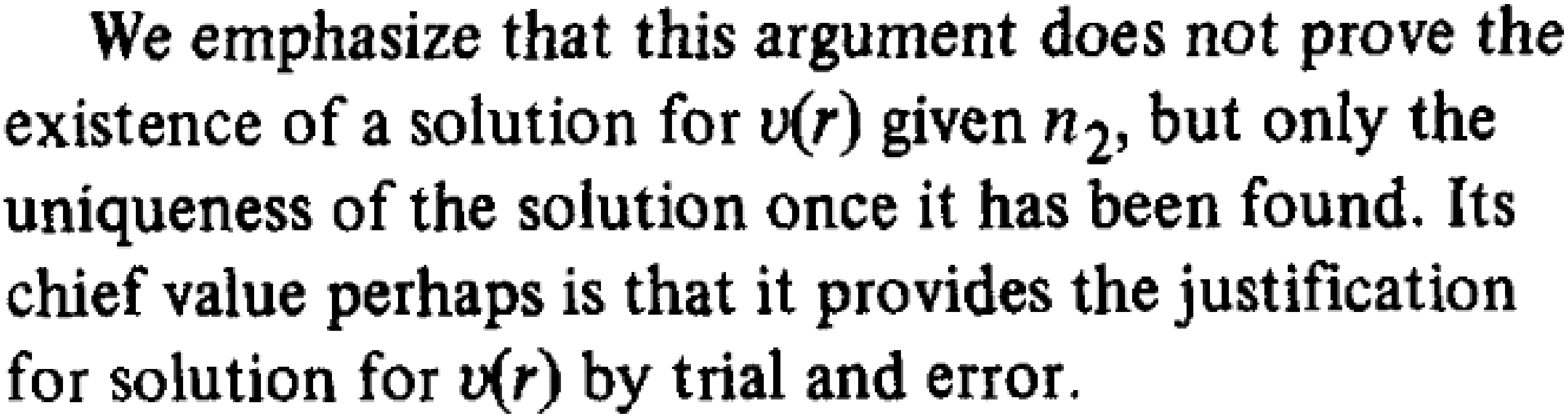
where $\langle \ \rangle_1$ denotes averaging with the distribution appropriate to H_1. The important point in the following argument is that the *equality* holds in (3), at finite temperatures, if and only if $H_2 - H_1$ is independent of all coordinates and momenta*.

$$F < F' + \tfrac{1}{2} \int \mathrm{d}\boldsymbol{r}\, \mathrm{d}\boldsymbol{r}' \, [v(\boldsymbol{r}-\boldsymbol{r}') - v'(\boldsymbol{r}-\boldsymbol{r}')] \, n_2'(\boldsymbol{r},\boldsymbol{r}') \, , \quad (4)$$

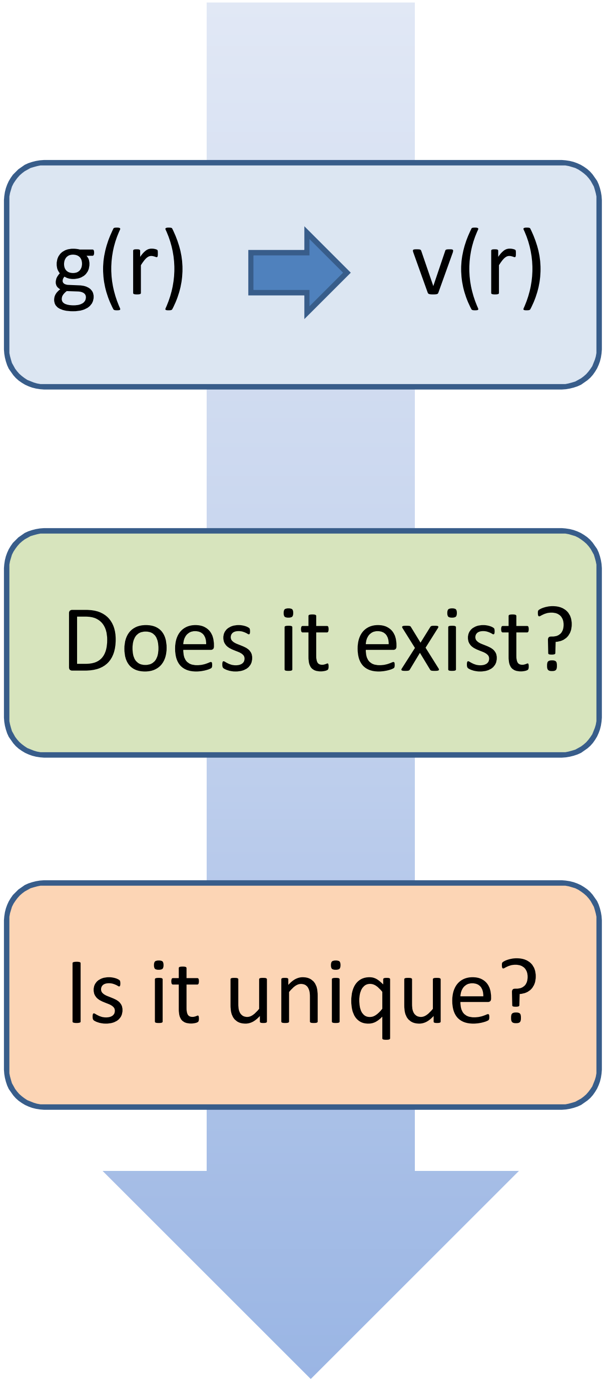
where use has been made of (2). This argument may be repeated with primed and unprimed systems interchanged, giving

$$F' < F + \tfrac{1}{2} \int \mathrm{d}\boldsymbol{r}\, \mathrm{d}\boldsymbol{r}' \, [v'(\boldsymbol{r}-\boldsymbol{r}') - v(\boldsymbol{r}-\boldsymbol{r}')] \, n_2(\boldsymbol{r},\boldsymbol{r}') \, . \quad (5)$$

If we now suppose that $n_2(\boldsymbol{r},\boldsymbol{r}') = n_2'(\boldsymbol{r},\boldsymbol{r}')$ for all $\boldsymbol{r}$ and $\boldsymbol{r}'$, addition of (4) and (5) yields $0 < 0$. Thus, two pair potentials which differ by more than a constant cannot give rise to the same n_2, and this is equivalent to the stated result. We note that this argument is independent of whether quantum or classical statistics is used.



We emphasize that this argument does not prove the existence of a solution for $v(r)$ given n_2, but only the uniqueness of the solution once it has been found. Its chief value perhaps is that it provides the justification for solution for $v(r)$ by trial and error.



g(r) ⇨ v(r)

Does it exist?

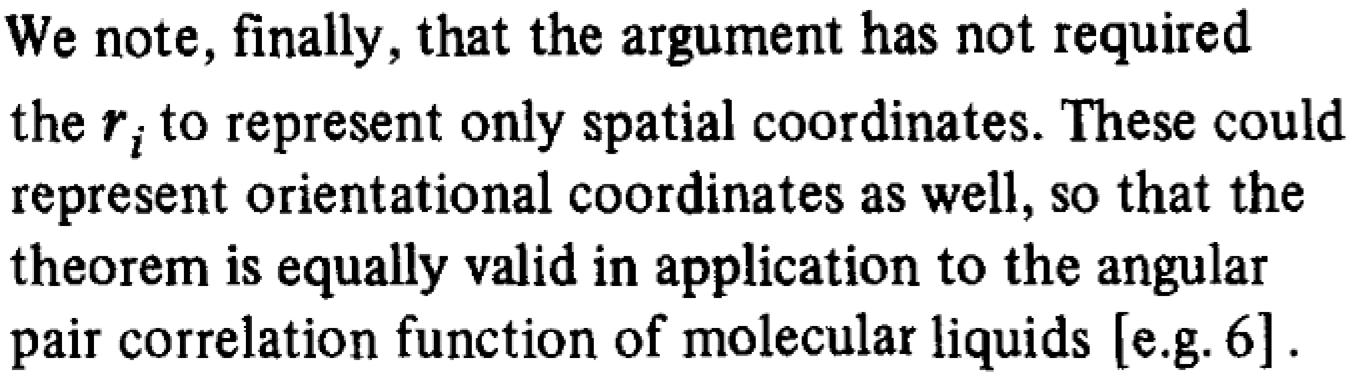
Is it unique?

"I don't know. It may or may not."

"I guarantee that if exists, it is the only one!

(well... within a constant) ."

J. R. Henderson

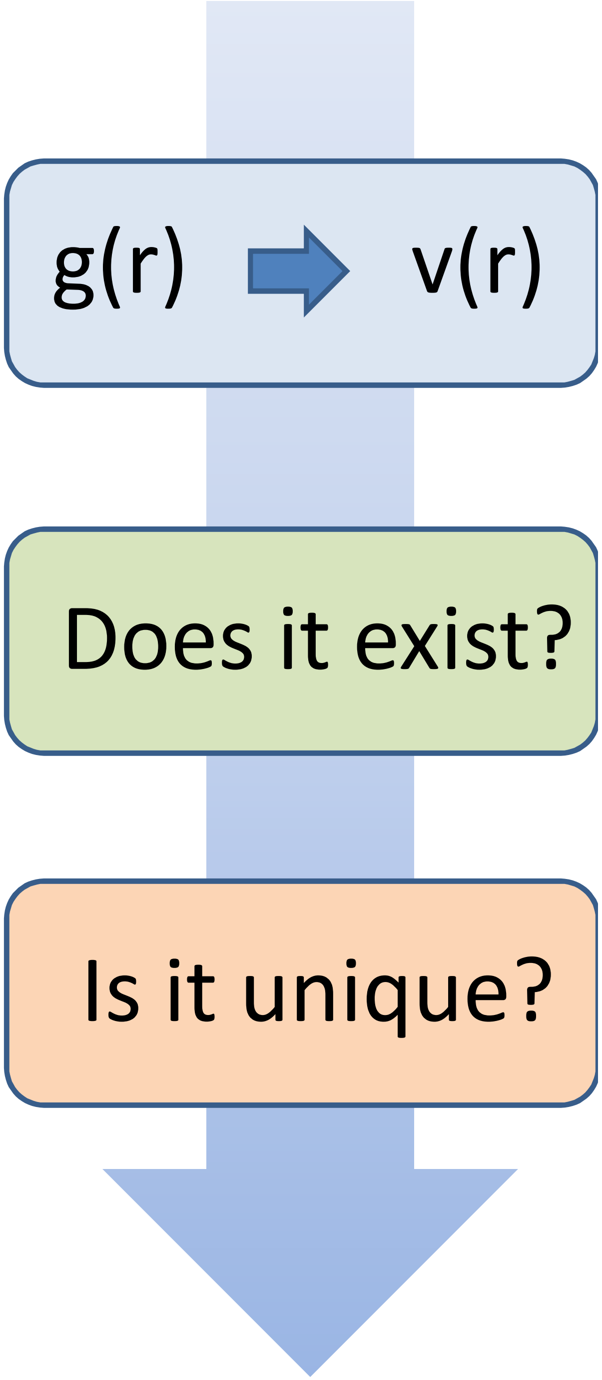

We note, finally, that the argument has not required the $\boldsymbol{r}_i$ to represent only spatial coordinates. These could represent orientational coordinates as well, so that the theorem is equally valid in application to the angular pair correlation function of molecular liquids [e.g. 6] .

“Works for non-spherical potentials as well.”

J. R. Henderson

Lets try this with a test case:

A simple LJ fluid



$g(r) \Rightarrow v(r)$

Does it exist?

YES

Is it unique?

YES

5614 *Ind. Eng. Chem. Res.* **2006,** *45,* 5614−5618

Do Inverse Monte Carlo Algorithms Yield Thermodynamically Consistent Interaction Potentials?

Sandeep Jain, Shekhar Garde, and Sanat K. Kumar*

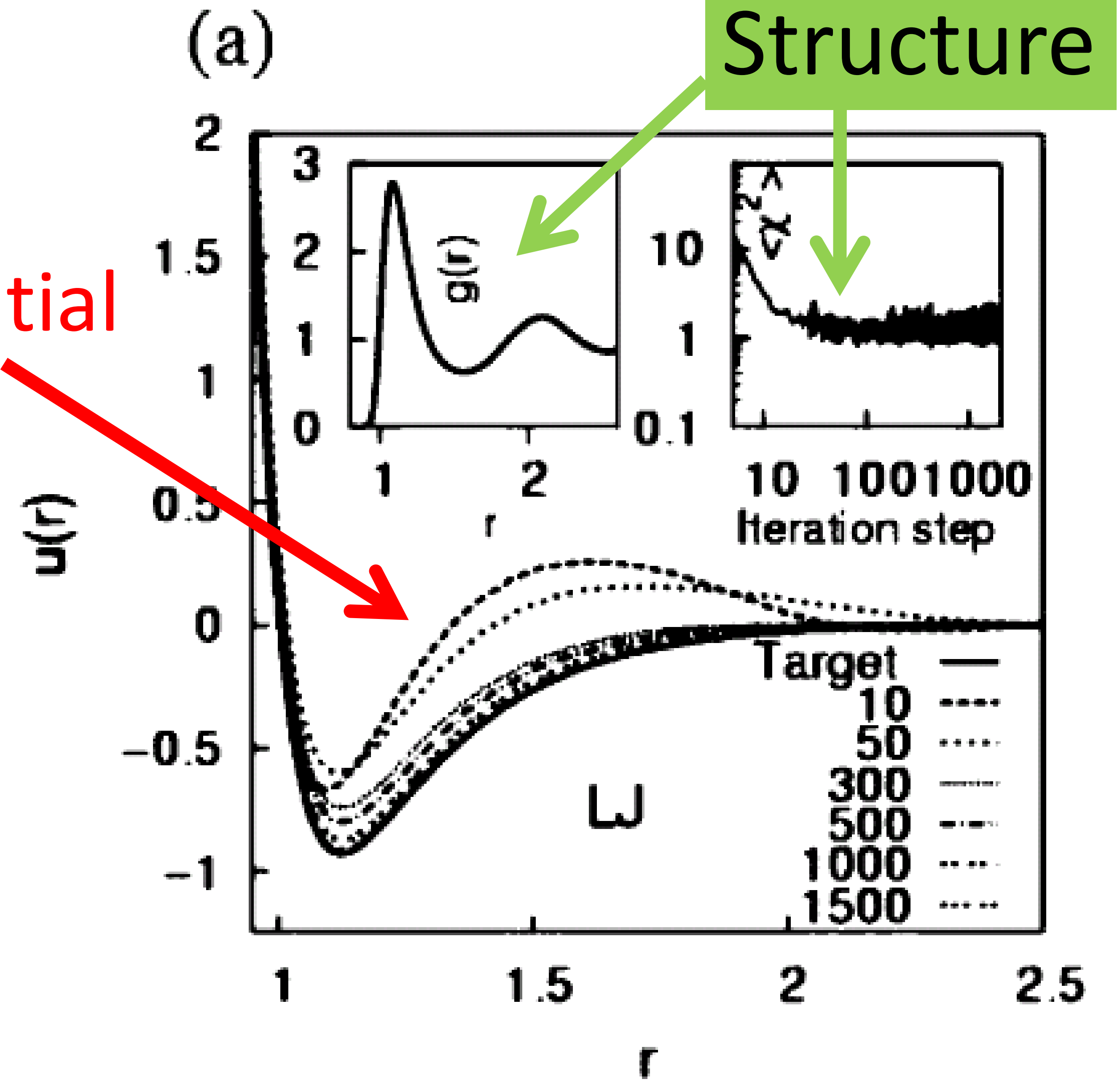
The Howard P. Isermann Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180

We numerically verify a statistical mechanics theorem which shows that there is a one-to-one equivalence between the structure of a liquid (i.e., the pair correlation function) and its pairwise additive intermolecular potential. Specifically, we show for three systems interacting with simple spherically symmetric pairwise additive potentials that inverse Monte Carlo (IMC) simulations can obtain the underlying potentials by only using the target pair correlation functions. The convergence of potentials obtained by the standard IMC procedure is, however, extremely slow. Interestingly, we find that the repulsive part of the potential converges rapidly, consistent with the well-accepted notion that it essentially determines the structure of condensed liquids. We show that additional information about the system, such as thermodynamic properties (e.g., average energy and or pressure) can be included in a modified IMC procedure. Because internal energy and pressure are primarily sensitive to the attractive part of the potential, the convergence to the true potential is improved by an order of magnitude. Although the improved convergence is a technical advance, no new information is obtained on the final converged potential by this approach, as expected by the Henderson theorem.

Boltzmann Or RMC or IMC

$$u_{i+1}(r) = u_i(r) + fk_{\mathrm{B}}T \ln\left(\frac{g_i(r)}{g_{\mathrm{target}}(r)}\right) \qquad (1)$$

$$\chi_j{}^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{g_i(r_j) - g_{\mathrm{target}}(r_j)}{\sigma(r_j)}\right)^2 \qquad (2)$$



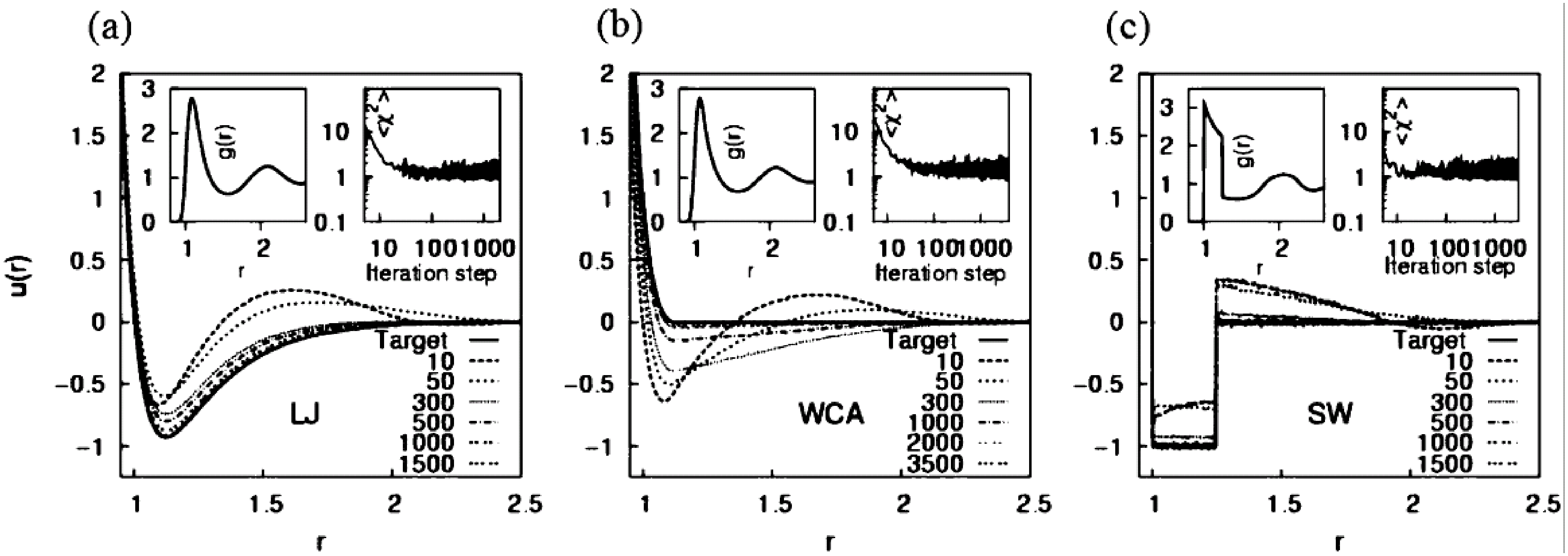
Structure

potential

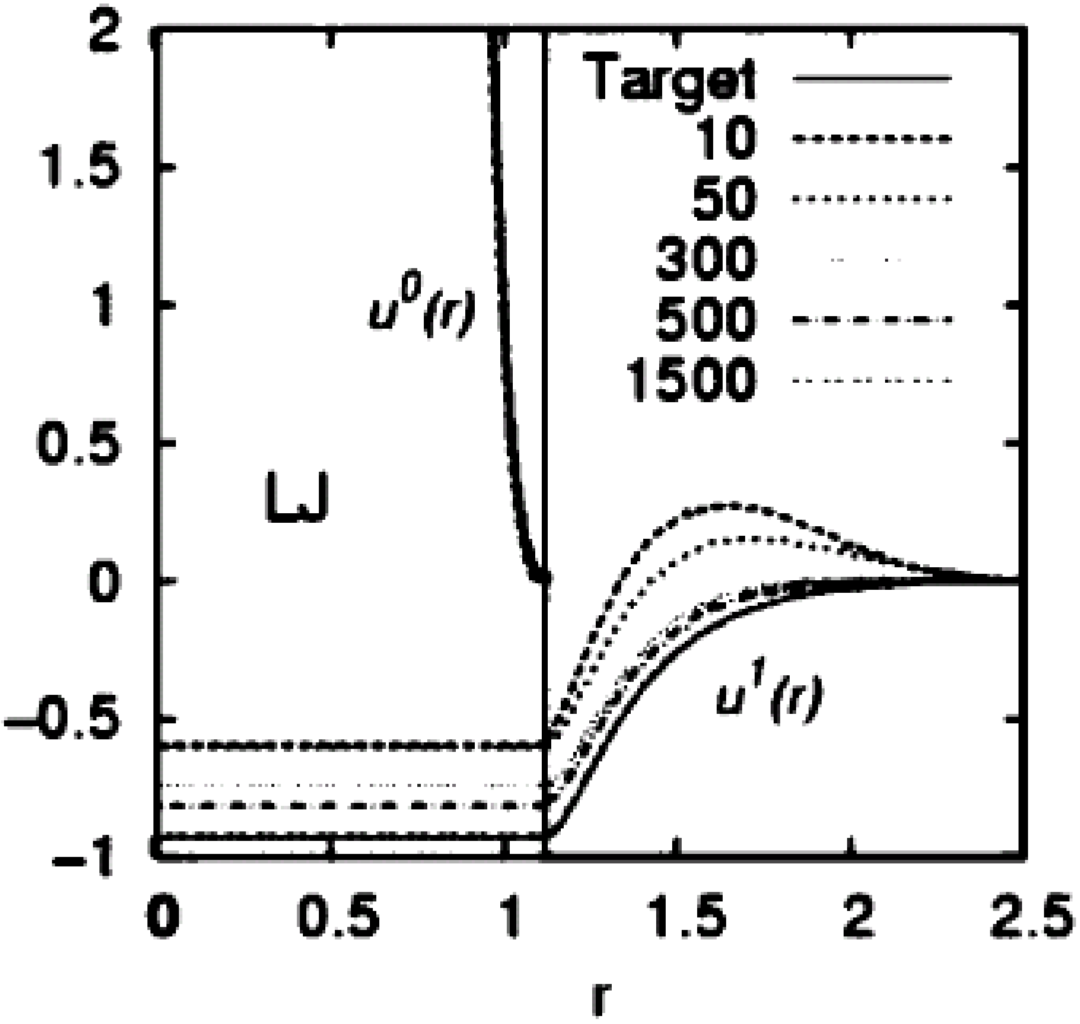
LJ fluid

WCA

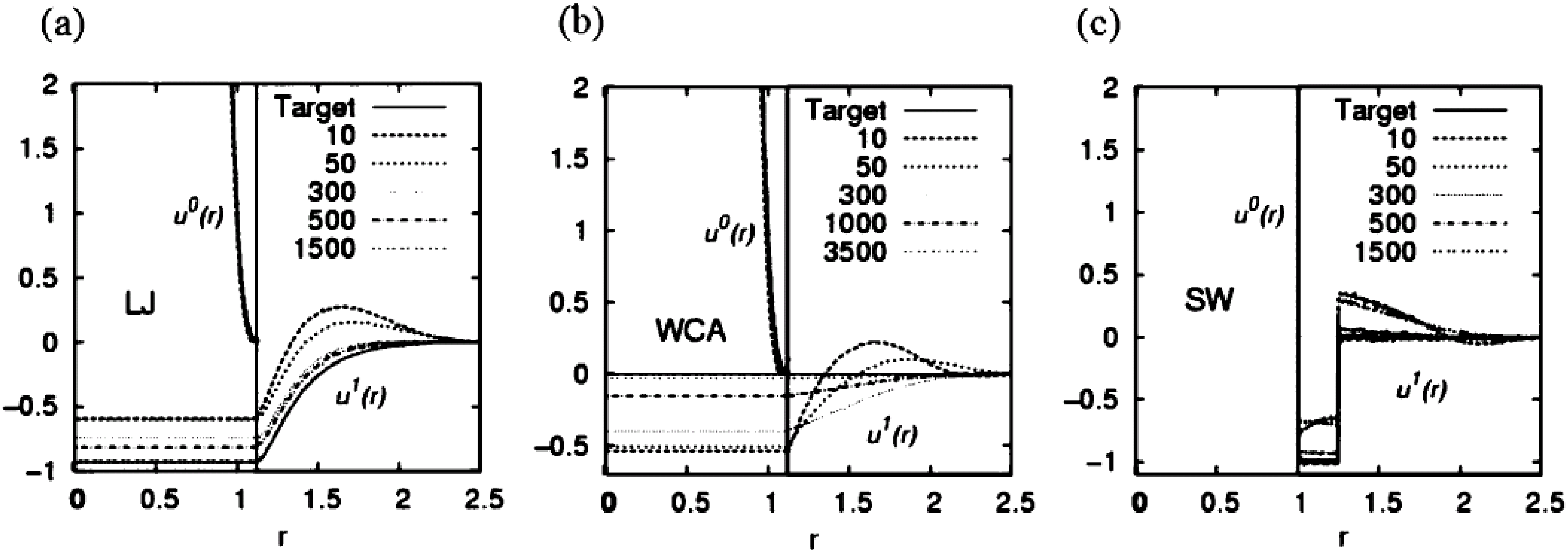
Sq. well



Repulsive part converges fast



Repulsive part converges fast



Adding terms to fix energy, pressure, etc.

$$\Delta(r) = \begin{cases} (a + br)(r - R_{\text{cut}})^{y} & r < R_{\text{cut}} \\ 0 & r \geq R_{\text{cut}} \end{cases} \qquad (3)$$

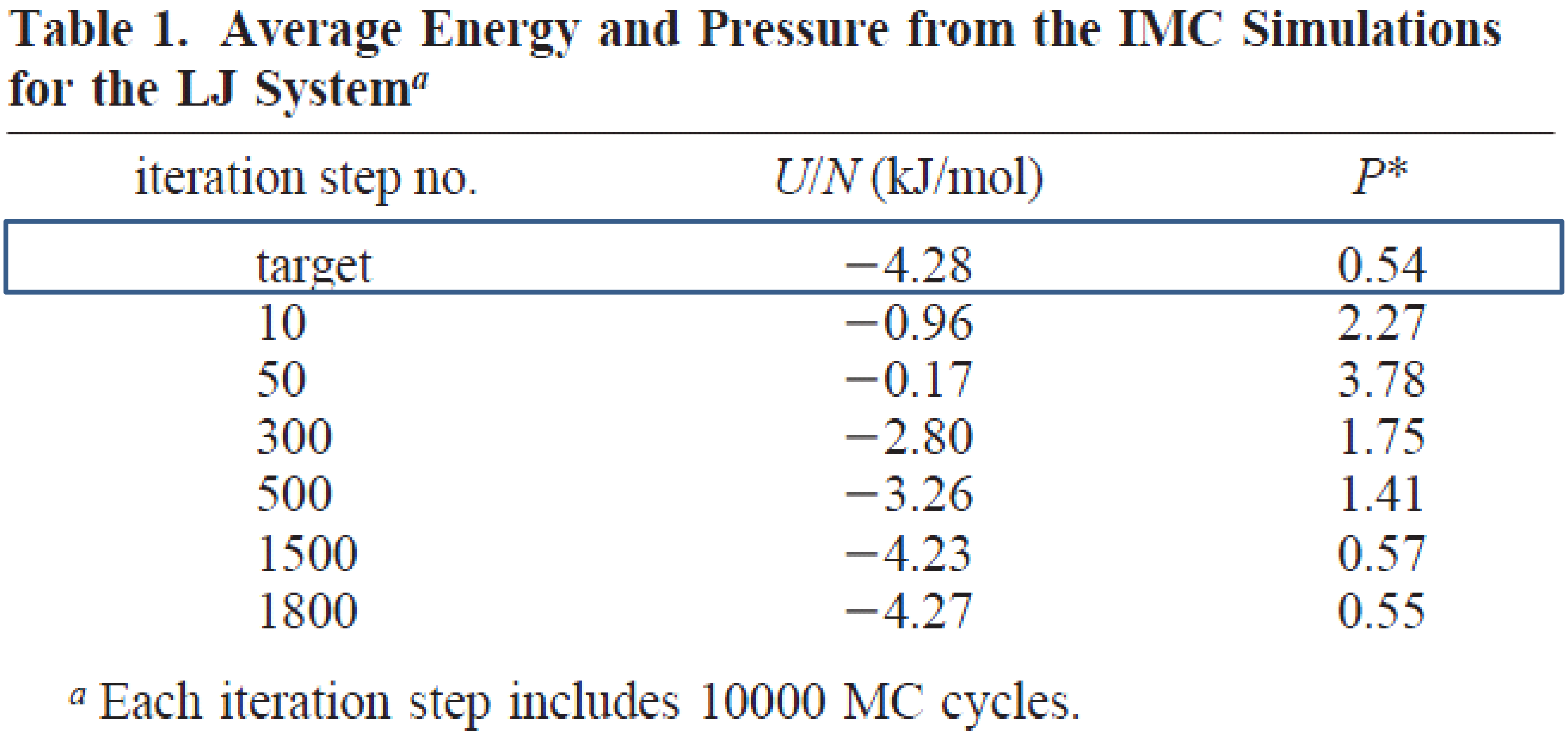
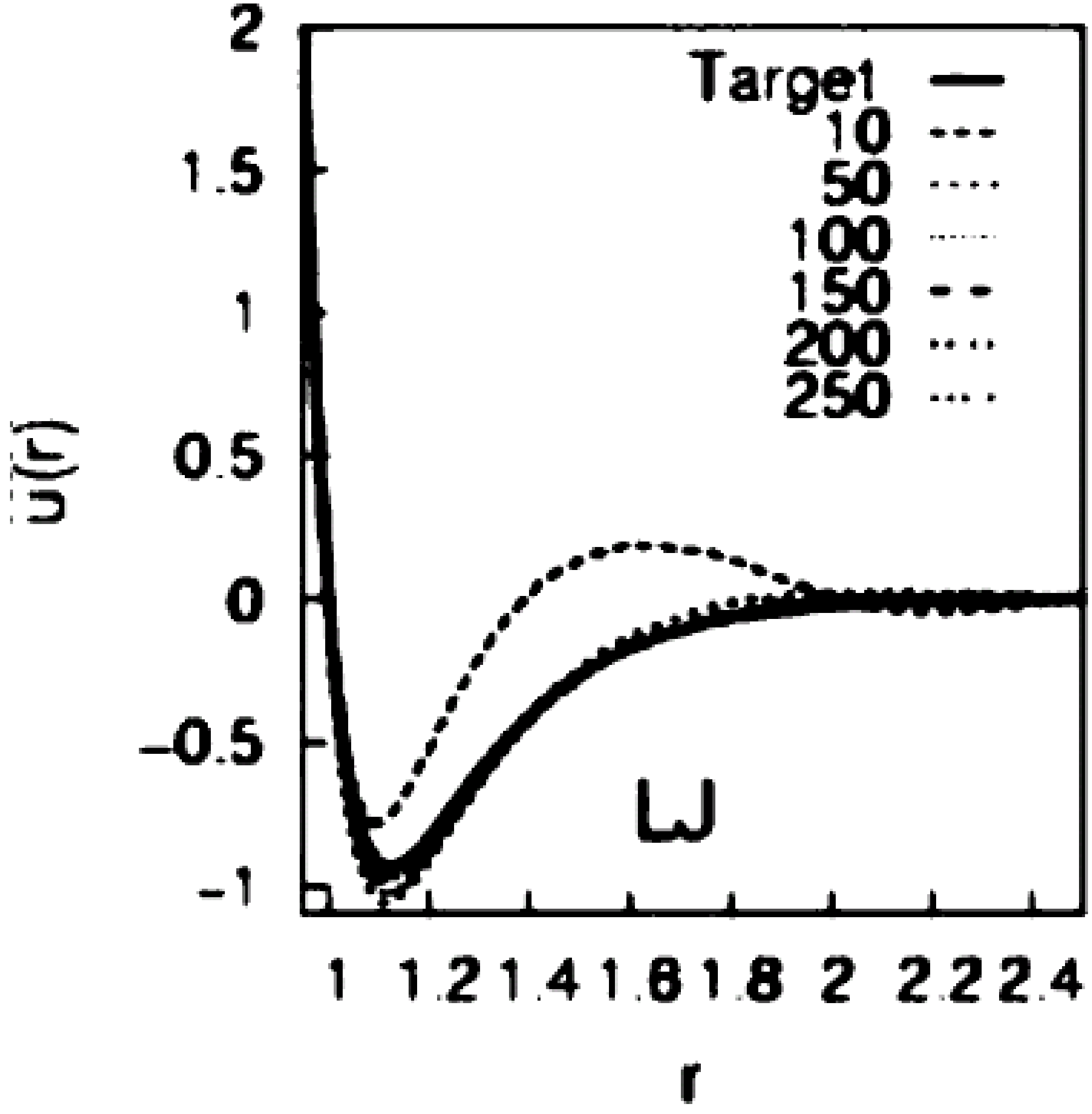


Table 1. Average Energy and Pressure from the IMC Simulations for the LJ System[a]

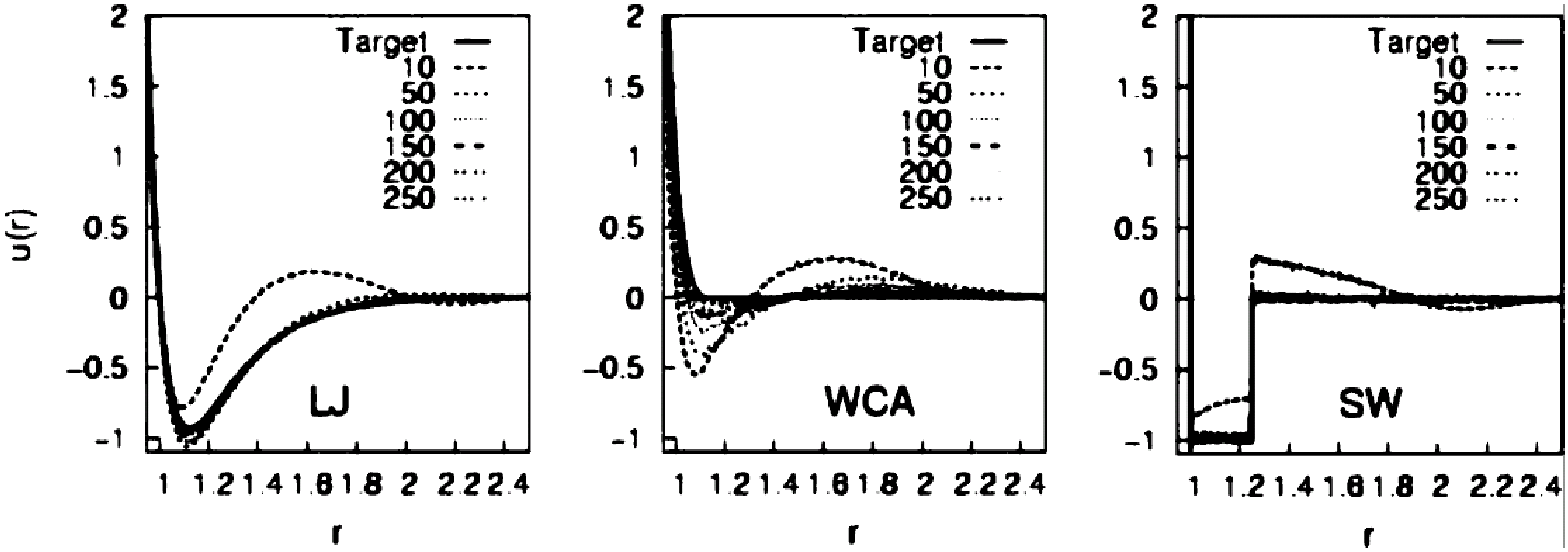
iteration step no.	U/N (kJ/mol)	P^*
target	−4.28	0.54
10	−0.96	2.27
50	−0.17	3.78
300	−2.80	1.75
500	−3.26	1.41
1500	−4.23	0.57
1800	−4.27	0.55

[a] Each iteration step includes 10000 MC cycles.

Rapid convergence

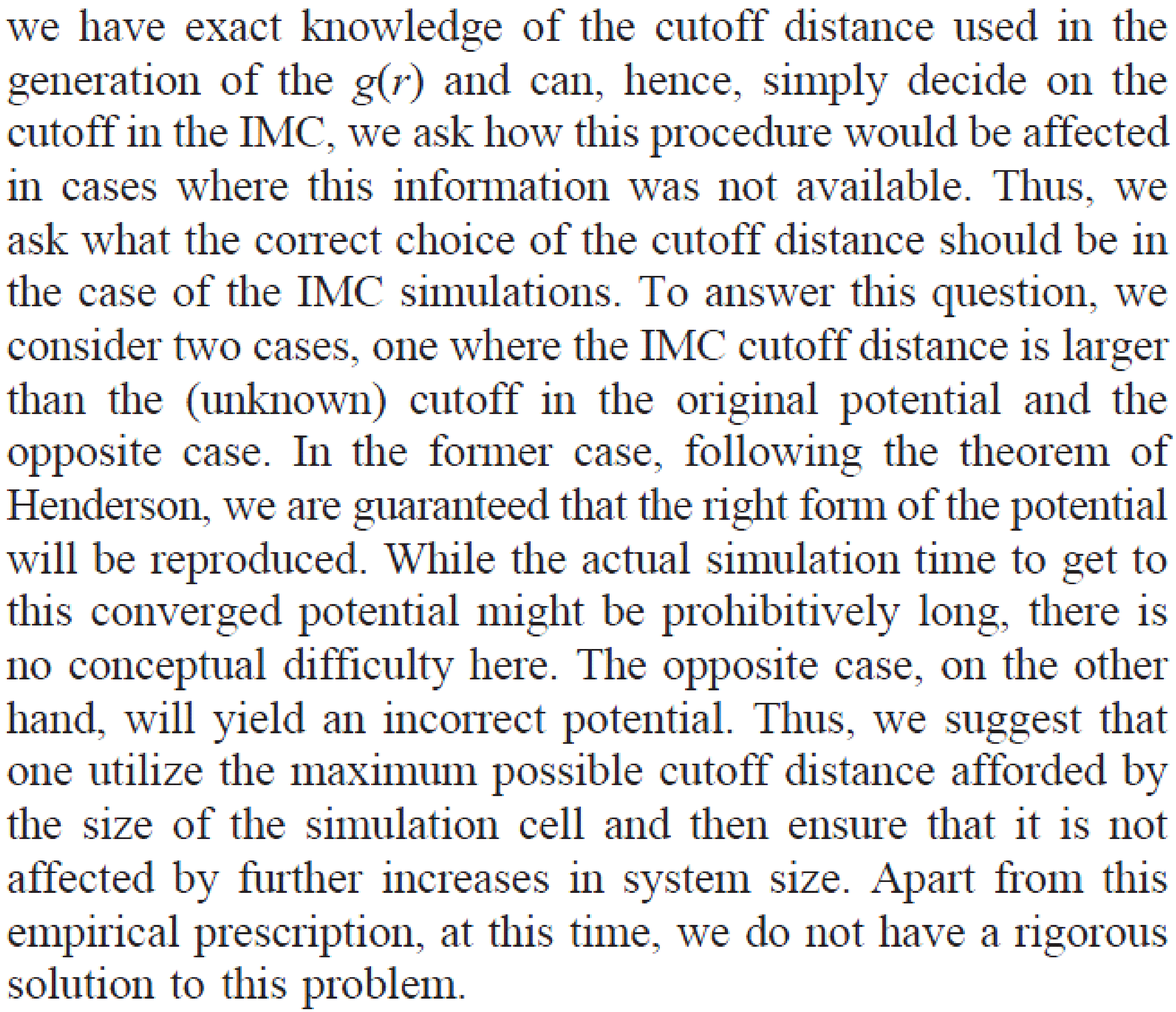


Rapid convergence



Why does this work?**

Cut off



we have exact knowledge of the cutoff distance used in the generation of the $g(r)$ and can, hence, simply decide on the cutoff in the IMC, we ask how this procedure would be affected in cases where this information was not available. Thus, we ask what the correct choice of the cutoff distance should be in the case of the IMC simulations. To answer this question, we consider two cases, one where the IMC cutoff distance is larger than the (unknown) cutoff in the original potential and the opposite case. In the former case, following the theorem of Henderson, we are guaranteed that the right form of the potential will be reproduced. While the actual simulation time to get to this converged potential might be prohibitively long, there is no conceptual difficulty here. The opposite case, on the other hand, will yield an incorrect potential. Thus, we suggest that one utilize the maximum possible cutoff distance afforded by the size of the simulation cell and then ensure that it is not affected by further increases in system size. Apart from this empirical prescription, at this time, we do not have a rigorous solution to this problem.

PHYSICAL REVIEW E VOLUME 52, NUMBER 4 OCTOBER 1995

Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach

Alexander P. Lyubartsev* and Aatto Laaksonen
Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, S-109 61, Stockholm, Sweden

(Received 15 March 1995; revised manuscript received 5 July 1995)

An approach is presented to solve the reverse problem of statistical mechanics: reconstruction of interaction potentials from radial distribution functions. The method consists of the iterative adjustment of the interaction potential to known radial distribution functions using a Monte Carlo simulation technique and statistical-mechanics relations to connect deviations of canonical averages with Hamiltonian parameters. The method is applied to calculate the effective interaction potentials between the ions in aqueous NaCl solutions at two different concentrations. The reference ion-ion radial distribution functions, calculated in separate molecular dynamics simulations with water molecules, are reproduced in Monte Carlo simulations, using the effective interaction potentials for the hydrated ions. Application of the present method should provide an effective and economical way to simulate equilibrium properties for very large molecular systems (e.g., polyelectrolytes) in the presence of hydrated ions, as well as to offer an approach to reduce a complexity in studies of various associated and aggregated systems in solution.

II. THEORETICAL BACKGROUND

Consider a system with a Hamiltonian (potential energy) given as

$$H\{q_i\} = \sum_{\alpha} K_\alpha S_\alpha\{q_i\} \ , \tag{1}$$

where $S_\alpha\{q_i\}$ are functions of particle coordinates q_i, and K_α are constants defining the interaction potential. The summation in Eq. (1) may also be replaced by an integral.

The Hamiltonian of a system with pair interactions can be given in the same fashion as Eq. (1):

$$\begin{aligned} H = \sum_{i,k} \Psi(|q_i - q_k|) &= \sum_{i,k} \int_0^\infty \Psi(r)\delta(r - |q_i - q_k|)dr \\ &= \int_0^\infty \Psi(r) \sum_{i,k} \delta(r - |q_i - q_k|)dr \ . \end{aligned} \tag{2}$$

In comparison with Eq. (1), the sum is now replaced by an integral, α by r, K_α by $\Psi(r)$, and $S_\alpha\{q_i\}$ by $\sum_{i,k}\delta(r - |q_i - q_k|)$.

In the vicinity of an arbitrary point in the space of Hamiltonians, $\{K_\alpha\}$, we can write

$$\Delta\langle S_\alpha\rangle = \sum_\gamma \frac{\partial\langle S_\alpha\rangle}{\partial K_\gamma}\Delta K_\gamma + O(\Delta K^2) \ , \tag{3}$$

where the derivatives $\partial\langle S_\alpha\rangle/\partial K_\gamma$ are expressed as

$$\frac{\partial\langle S_\alpha\rangle}{\partial K_\gamma} = \frac{\partial}{\partial K_\gamma}\left[\frac{\int dq\, S_\alpha(q)\exp\left[-\beta\sum_\lambda K_\lambda S_\lambda(q)\right]}{\int dq\, \exp\left[-\beta\sum_\lambda K_\lambda S_\lambda(q)\right]}\right]$$

$$= -\beta(\langle S_\alpha S_\gamma\rangle - \langle S_\alpha\rangle\langle S_\gamma\rangle) \ , \tag{4}$$

and q is the set of degrees of freedom of the reduced sys-

Conclusions

Does my CG approach give me the right thermodynamics?

Yes, if

U(S,V)

H(S,P)

A(T,V)

G(T,P) are right.

g(r) → v(r)

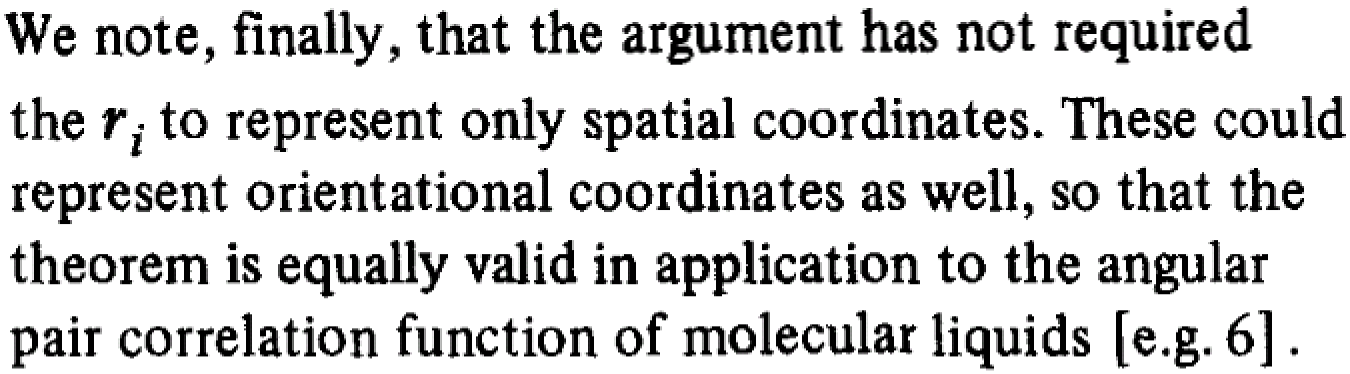
Does it exist?

Is it unique?

“I don’t know. It may or may not.”

“I guarantee that if exists, it is the only one! (well... within a constant) .”

J. R. Henderson



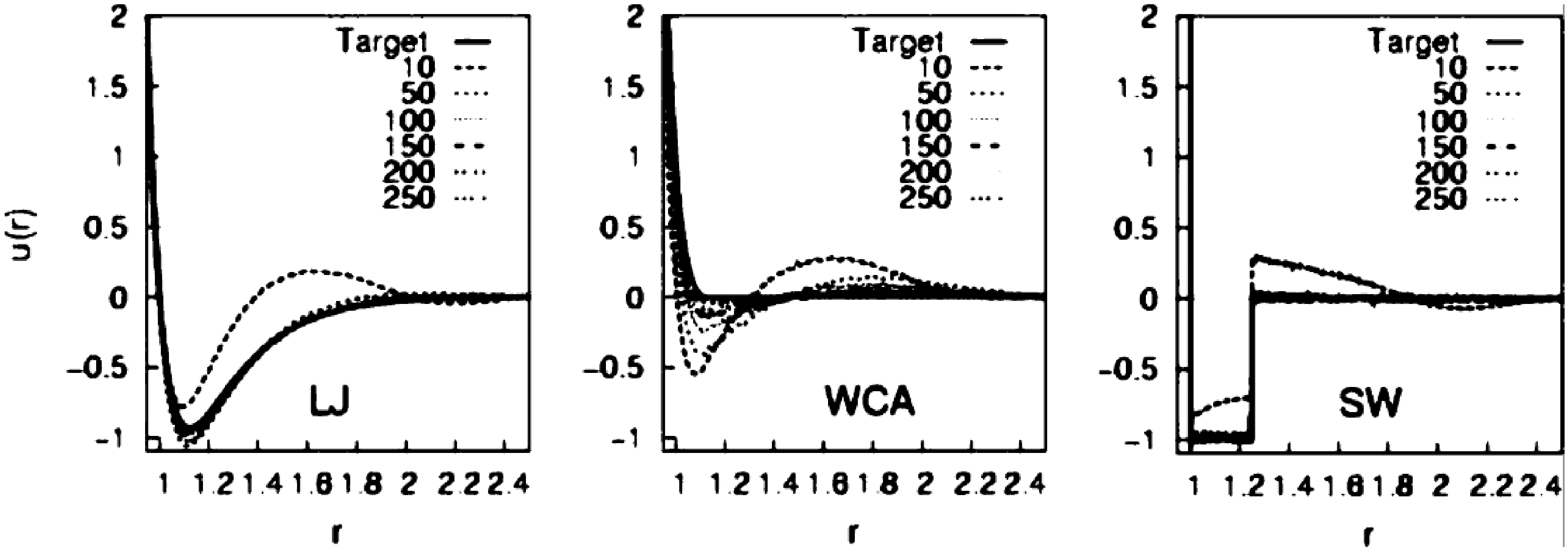
We note, finally, that the argument has not required the r_i to represent only spatial coordinates. These could represent orientational coordinates as well, so that the theorem is equally valid in application to the angular pair correlation function of molecular liquids [e.g. 6] .

"Works for non-spherical potentials as well."

J. R. Henderson

?

Rapid convergence



Why does this work?**

Questions?