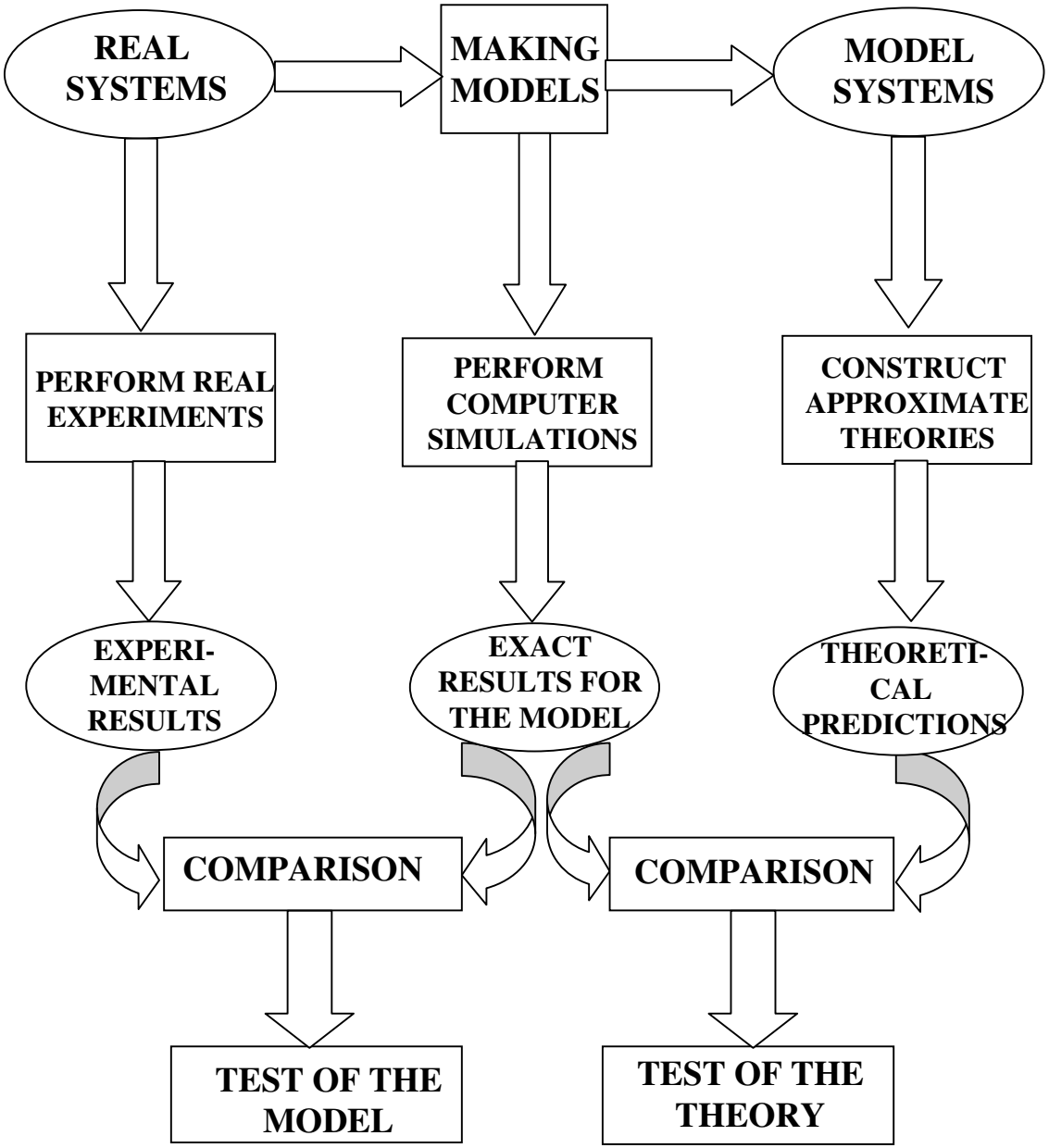


Allen, Tildesley: *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987)

The Reverse Monte Carlo (RMC) method

Classical simulation methods (Monte Carlo, molecular dynamics):

(pair)potential ⟶ molecular level structure (pair correlation)

principal limit: the used potential function can only be validated **afterwards**
(good agreement with experimental datal)
strictly, this justification is only valid for the given system in the given thermodynamic state

Question: is it possible to construct a simulation method that can reproduce at least a few properties of the system of interest automatically, without a use of any potential function

Reverse Monte Carlo (RMC) is such a method

(McGreevy and Pusztai, *Mol. Simul*. **1,** 359 (1988).

The main principles of the Reverse Monte Carlo method:

- It reproduces several experimental quantities of the system that only depend on the $\underline{q}^N$ position coordinates of the particles

- These quantities are reproduced within the (estimated) experiomental error

-Then there is a question how well we can reproduce the other $\underline{q}^N$ -dependent properties of the system (problem of the uniqueness of the method)

- Avoiding the use of any kind of potential functions results in the loss of the thermodynamic information, RMC can only provide structural information

Reverse Monte Carlo simulation

- N particles in a box of volume V, random moves, periodic boundary conditions

- Acceptance of the moves: a new criterion has to be introduced
  the aim is to well reproduce the measured quantities

-Assume that we have n different experimental functions:

$$f_1^{\exp}(x), f_2^{\exp}(x), f_3^{\exp}(x), \ldots, f_n^{\exp}(x), \text{ where } x = x(\underline{q}^N)$$

each function is measured at the same m values of the independent parameter x:

$$f_i^{exp}(x_j), \text{ where } 1 \le i \le n \text{ and } 1 \le j \le m$$

- It is assumed that the measurements are only affected by random and not by systematic error. Then the error of each measured point follows a Gaussian distribution. The (random) experimental error of the $f_i^{\exp}(x_j)$ measured point is denoted by $\sigma_i(x_j)$.

Reverse Monte Carlo simulation

- The $f(x)$ functions are also evaluated on the simulated configurations at each point (i.e., for each i, j pair) – it can be done since they only depend on $\underline{q}^N$
  The simulated functions are denoted by $f_i^{\mathrm{RMC}}(x_j)$

- **<u>Aim:</u> the simulated $f_i^{\mathrm{RMC}}(x_j)$ values should follow a Gaussian distribution with a width of $\sigma_i(x_j)$ around the experimental $f_i^{\mathrm{exp}}(x_j)$ poins**

  - The acceptance criterion has to be determined accordingly

Reminder – the Metropolis Monte Carlo algorithm

-We want to calculate the ensemble average of the general quantity M.
On the canonical (N,V,T) ensemble

$$< M > = \frac{\int_q M(\underline{q}^N) \exp\left(-\beta U(\underline{q}^N)\right) d\underline{q}^N}{\int_q \exp\left(-\beta U(\underline{q}^N)\right) d\underline{q}^N}$$

- The ensemble of the possible microscopic states is replaced by the few sampled microscopic states, and the value of $<M>$ is calculated on this sample

-**Problem of the representativity of the sample** Solution: **weighed sampling**
Let a given microscopic state be sampled with the probability (weight) of $w(\underline{q}^N)$
Then:

$$< M > = \frac{\sum_q M(\underline{q}^N) \exp\left(-\beta U(\underline{q}^N)\right) / w(\underline{q}^N)}{\sum_q \exp\left(-\beta U(\underline{q}^N)\right) / w(\underline{q}^N)}$$

Reminder – the Metropolis Monte Carlo algorithm

$$< M > = \frac{\sum_q M(\underline{q}^N) \exp\left(-\beta U(\underline{q}^N)\right) / w(\underline{q}^N)}{\sum_q \exp\left(-\beta U(\underline{q}^N)\right) / w(\underline{q}^N)}$$

Let us use the following weighing: $w(\underline{q}^N) = \exp\left(-\beta U(\underline{q}^N)\right)$

Then $$< M > = \frac{\sum_{i=1}^{k} M(\underline{q}_i^N)}{k}$$

Instead of sampling with equal probability and weighing with the Boltzmann factor in the averaging, the sampling is done with a Boltzmann-weighing and unweighed averages are calculated.

Importance sampling

Question: **How can we realize a Boltzmann-weighed sampling?**

Boltzmann-weighed sampling

(Metropolis, Rosenbluth, Rosenbluth, Teller, Teller, *J. Chem. Phys.* **21,** 1087 (1953)

- Put N particles in a basic (cubic, rectangular, prism-shaped, etc.) box of volume V

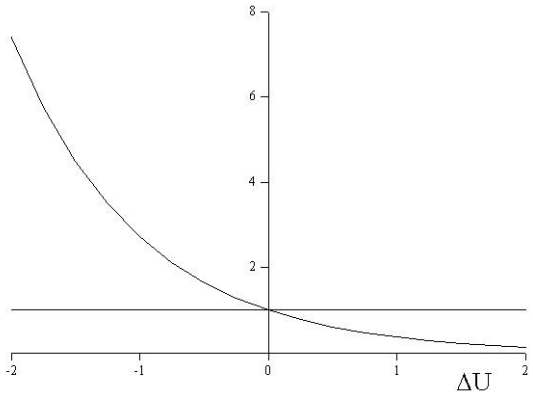
- Make random changes in the configuration (random move of a randomly chosen particle)

- Calculate the configurational energy $U(\underline{q}^N)$

- Decide whether the new configuration is accepted or rejected:

  •if $\Delta U = U_{new}\text{-}U_{old} \leq 0$ ⟶ always accepted

  • if $\Delta U = U_{new}\text{-}U_{old} > 0$ ⟶ accepted by the probability of $\exp(-\Delta U/k_B T)$
  rejected by the probability of $1\text{-}\exp(-\Delta U/k_B T)$



$$P_{acc} = \min[1, \exp(-\beta \Delta U)]$$

Sampling is started after the equilibrium is reached

Demonstrate that the above procedure indeed leads to Boltzmann-weighed sampling

P_{ij}: the probability that we bring the system from the microscopic state i to j

We do it by changing some of the particle's coordinates: $q_i' = q_i + \alpha\xi_i$

α: the maximum allowed change of the given coordinate

ξ_i: a random number of uniform distribution from the [-1,1] interval (usually one single particle is moved within a cube of the edge length 2α, centered at the original position of the particle)

For such moves the $P_{ij} = P_{ji}$ equality holds.

Let the i and j microscopic states such that $U_i(\underline{q}^N) > U_j(\underline{q}^N)$

Let us consider a canonical ensemble of the system studied, i.e., a sufficiently large number of systems corresponding to the (N,V,T) macroscopic parameters

Let the number of systems in states i and j be denoted by ν_i and ν_j, respectively.

Let us make a Monte Carlo move in each of thse systems. Now the number of systems that are

- brought to state j from state i is $\nu_i P_{ij}$
- brought to state i from state j is $\nu_j P_{ji} \exp\left(-\beta\left(U_i(\underline{q}^N) - U_j(\underline{q}^N)\right)\right)$

The change of the number of systems in state j (on the expense of those in state i) is:

$$\Delta_{ji} = \nu_i P_{ij} - \nu_j P_{ji} \exp\left(-\beta\left(U_i(\underline{q}^N) - U_j(\underline{q}^N)\right)\right) =$$

$P_{ji} = P_{ij}$

$$= P_{ij}\left(\nu_i - \nu_j \exp\left(-\beta\left(U_i(\underline{q}^N) - U_j(\underline{q}^N)\right)\right)\right)$$

If $\Delta_{ji} \geq 0$, then (because of $P_{ij} \geq 0$):

$$\nu_i - \nu_j \exp\left(-\beta\left(U_i(\underline{q}^N) - U_j(\underline{q}^N)\right)\right) \geq 0$$

$$\frac{\nu_i}{\nu_j} \geq \exp\left(-\beta\left(U_i(\underline{q}^N) - U_j(\underline{q}^N)\right)\right)$$

$$\frac{\nu_i}{\nu_j} \geq \frac{\exp\left(-\beta\left(U_i(\underline{q}^N)\right)\right)}{\exp\left(-\beta\left(U_j(\underline{q}^N)\right)\right)}$$

Similarly, if $\Delta_{ji} < 0$, then

$$\frac{\nu_i}{\nu_j} < \frac{\exp\left(-\beta\left(U_i(\underline{q}^N)\right)\right)}{\exp\left(-\beta\left(U_j(\underline{q}^N)\right)\right)}$$

Ensemble average: $\langle M \rangle = \dfrac{\sum M(\underline{q}^N) \exp(-\beta U(\underline{q}^N)}{\sum \exp(-\beta U(\underline{q}^N)}$

weighed sampling

$w(\underline{q}^N) = \exp(-\beta U(\underline{q}^N)$

random moves

$P_{acc} = \min[1, \exp(-\beta \underline{\Delta U(q^N)})]$

Reverse Monte Carlo

Now the sample configurations should follow a Gaussian distribution around each experimental $f_i^{exp}(x_j)$ values, instead of following a Boltzmann distribution of $U(\underline{q}^N)$

Strategy for designing the RMC method:

- let us find the weight of the individual configurations in the ensemble average
- let this weight be given in an exponential form

Reverse Monte Carlo

Let us use the following notation: $\delta_{ij} = \left| f_i^{\exp}(x_j) - f_i^{RMC}(x_j) \right|$

The probability that the simulated value deviates from the experimental one by δ_{ij} is:

$$P(\delta_{ij}) = \frac{1}{\sqrt{2\pi}\,\sigma_i(x_j)} \exp\left(\frac{-\delta_{ij}^2}{2\sigma_i^2(x_j)} \right)$$

The probability of a given configuration for being sampled should then be:

$$P(\underline{q}^N) = \prod_{i=1}^{n} \prod_{j=1}^{m} P\left(\delta_{ij}(\underline{q}^N)\right) = \underbrace{\left(\frac{1}{\sqrt{2\pi}\,\overline{\sigma}} \right)^{nm}}_{C} \exp\left(\underbrace{-\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\delta_{ij}^2(\underline{q}^N)}{\sigma_i^2(x_j)}}_{\chi^2} \right)$$

where $\overline{\sigma} = \sqrt[nm]{\prod_{i=1}^{n} \prod_{j=1}^{m} \sigma_i(x_j)}$

Therefore $P(\underline{q}^N) = C \exp\left(\frac{-\chi^2(\underline{q}^N)}{2} \right)$

The probability of a given configuration for being sampled is

$$P(\underline{q}^N) = C \exp\left(\frac{-\chi^2(\underline{q}^N)}{2}\right)$$

where $$\chi^2 = \sum \sum \frac{\left[f_i^{\exp}(x_j) - f_i^{RMC}(x_j)\right]^2}{\sigma_i^2(x_j)}$$

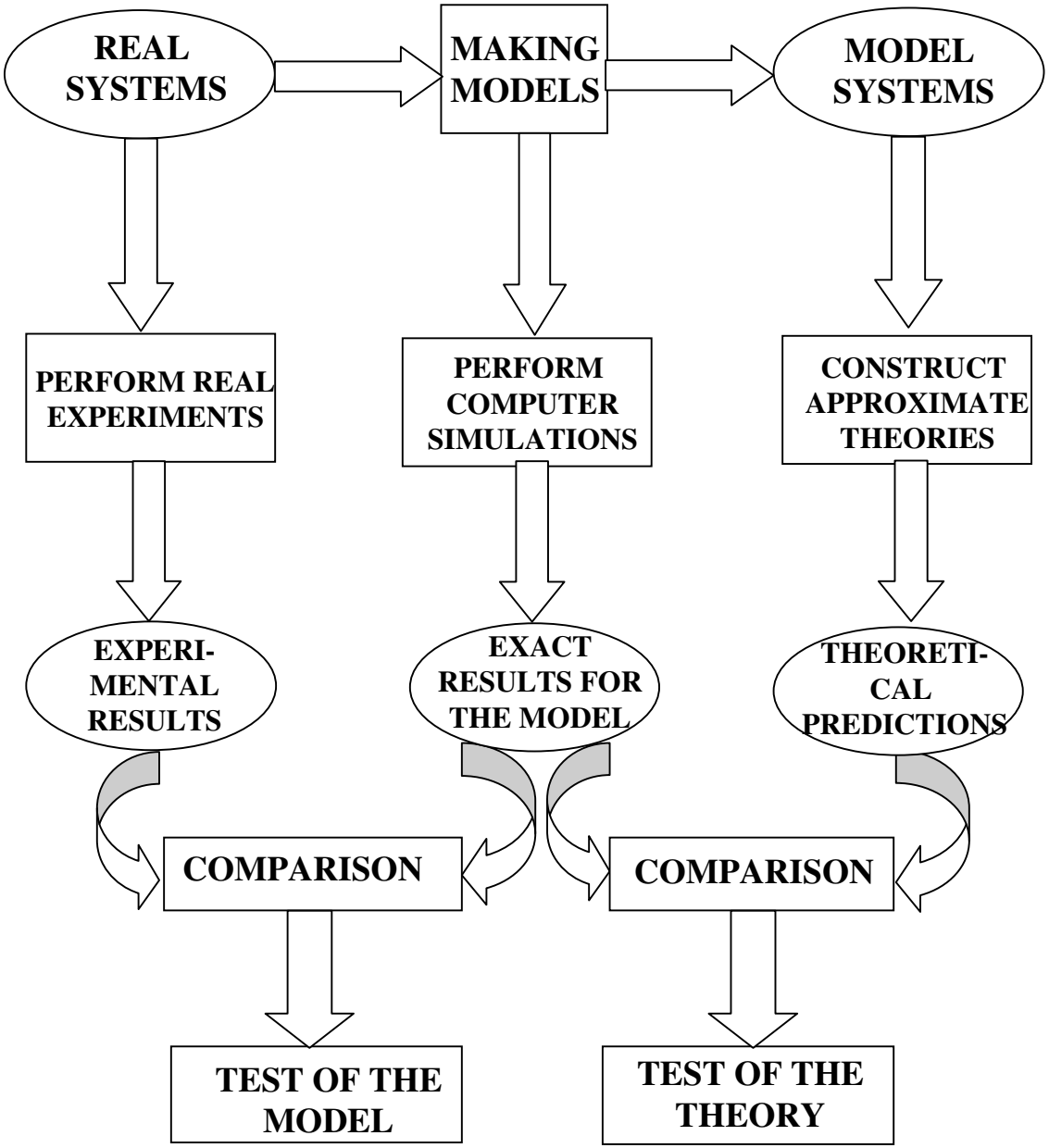
Therefore the acceptance criterion of the random moves is

$$P_{acc} = \min\left[1, \exp\left(-\Delta\chi^2/2\right)\right]$$

The RMC calculation is technically rather similar to Metropolis Monte Carlo

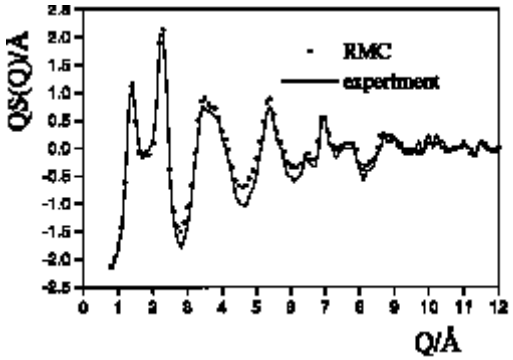
- *technically* the difference is small:
  - the role of βU is now played by $\chi^2/2$
  - the role of the temperature is now played by σ
  - the sample configurations are following Gaussian distributions around the experimental data rather than Boltzmann distribution of the energy

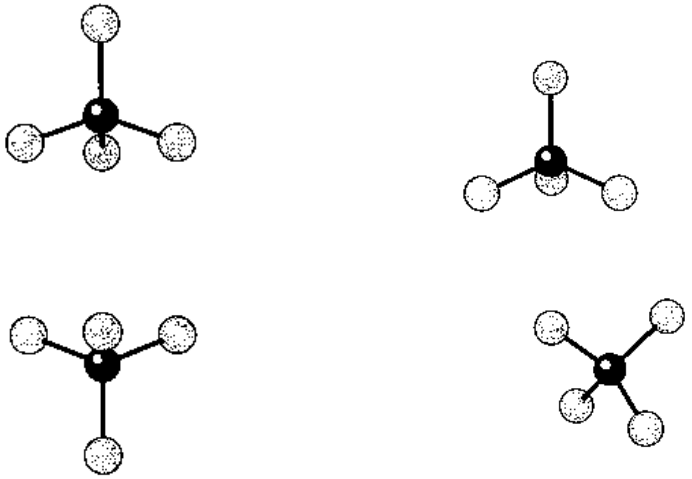
- *principally* the difference is enormous:
  - RMC is, in fact, not a simulation method, but a method for the evaluation of the experimental data



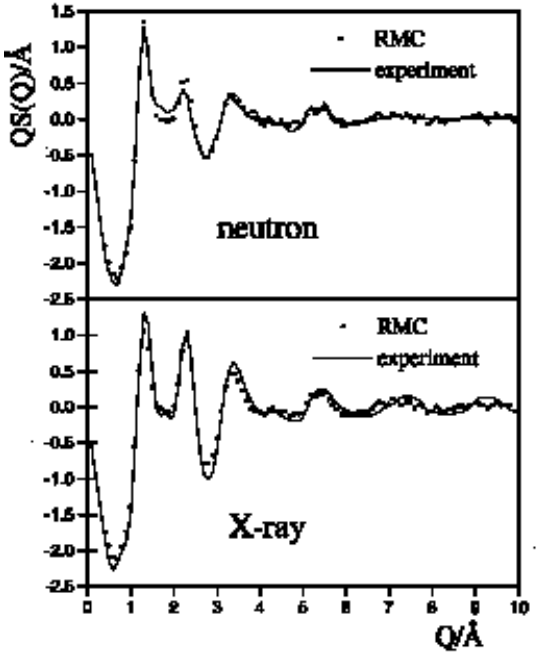
Allen, Tildesley: *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987)

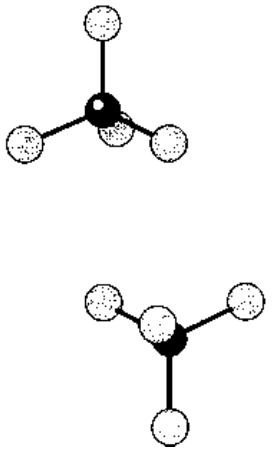
amorphous CCl_4





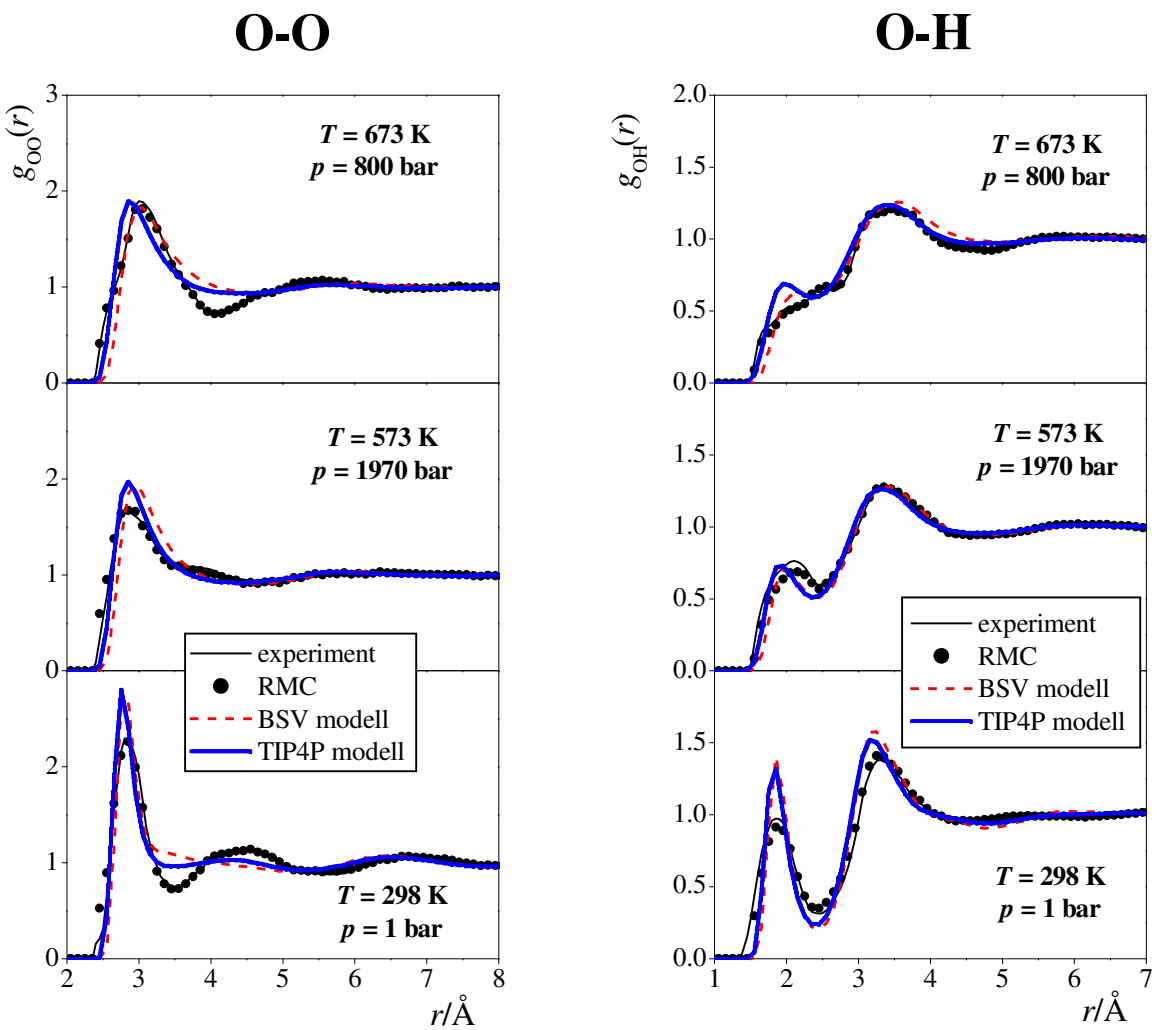
liquid CCl_4





PJ, *J. Chem. Phys.* **107,** 7433 (1997)

water at high pressures and high temperatures



PJ, Brodholt, Bruni, Ricci, Soper, Vallauri, *J. Chem. Phys.* **108,** 8528 (1998)

Technical questions about RMC

- Data of what kind of experiments should be used in RMC?

- X-ray diffraction
- neutron scattering
- light scattering
- EXAFS, NMR, etc,
  any method which results in data depending
  solely on the atomic positions

- Shall we fit *S*(*Q*)'s or *G*(*r*)'s?

-generally *S*(*Q*), because

(i) this is the direct result of the experiment
(conversion to *G*(*r*) can be done by RMC)

(ii) the error of *G*(*r*) at the different *r* points is unknown

The data set to be fitted should be as close to the primer experimental results as possible

- Shall we fit the measured total, or the separated partial functions ?

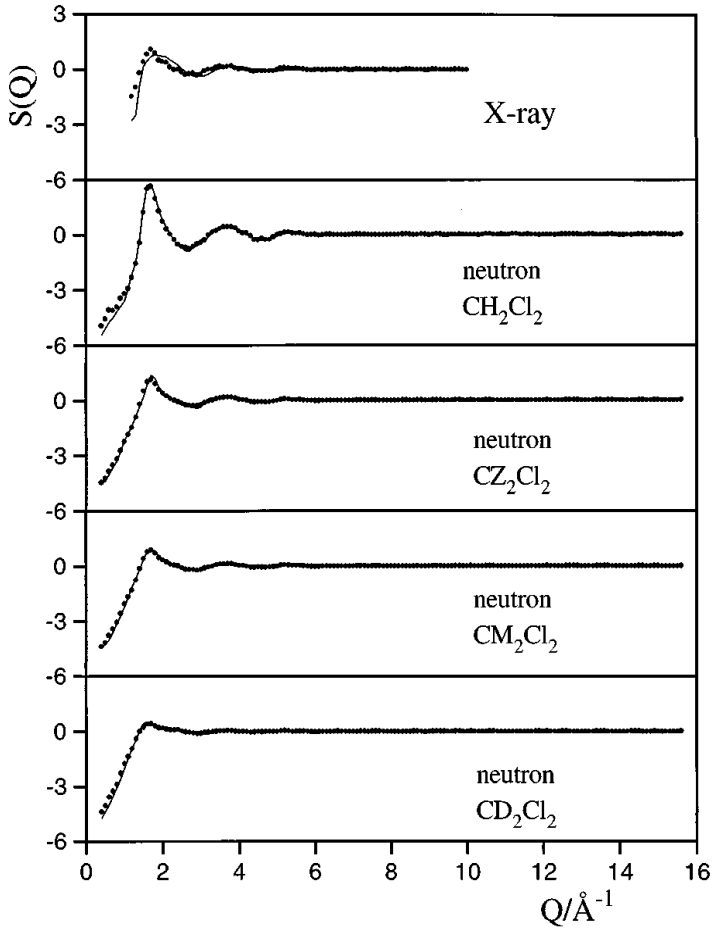
- again, the general rule is to make as little processing on the primer data as possible before being fed to the RMC procedure
  RMC can do the separation of the partials
  (But the question is always there how reliable this separation is if there are less independent data sets than partials!)

- How many independent experimental data sets do we need ?

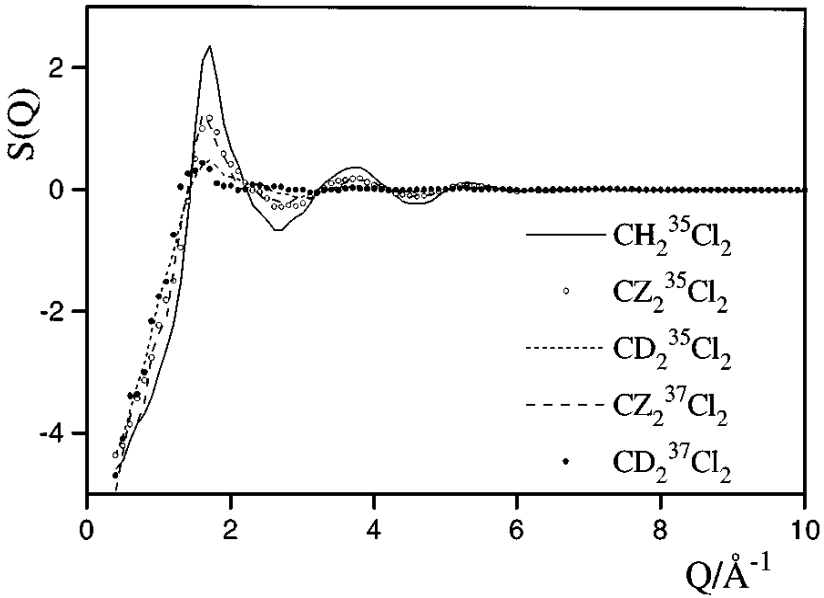
- as much as possible – the more experimental information we have on the system the more reliable is our model
  Results of new experiments can be predicted!

Liquid CH_2Cl_2

Fitted $S(Q)$ functions



Predicted $S(Q)$ functions



PJ, *J. Chem. Phys.* **107,** 562 (1997)

Water

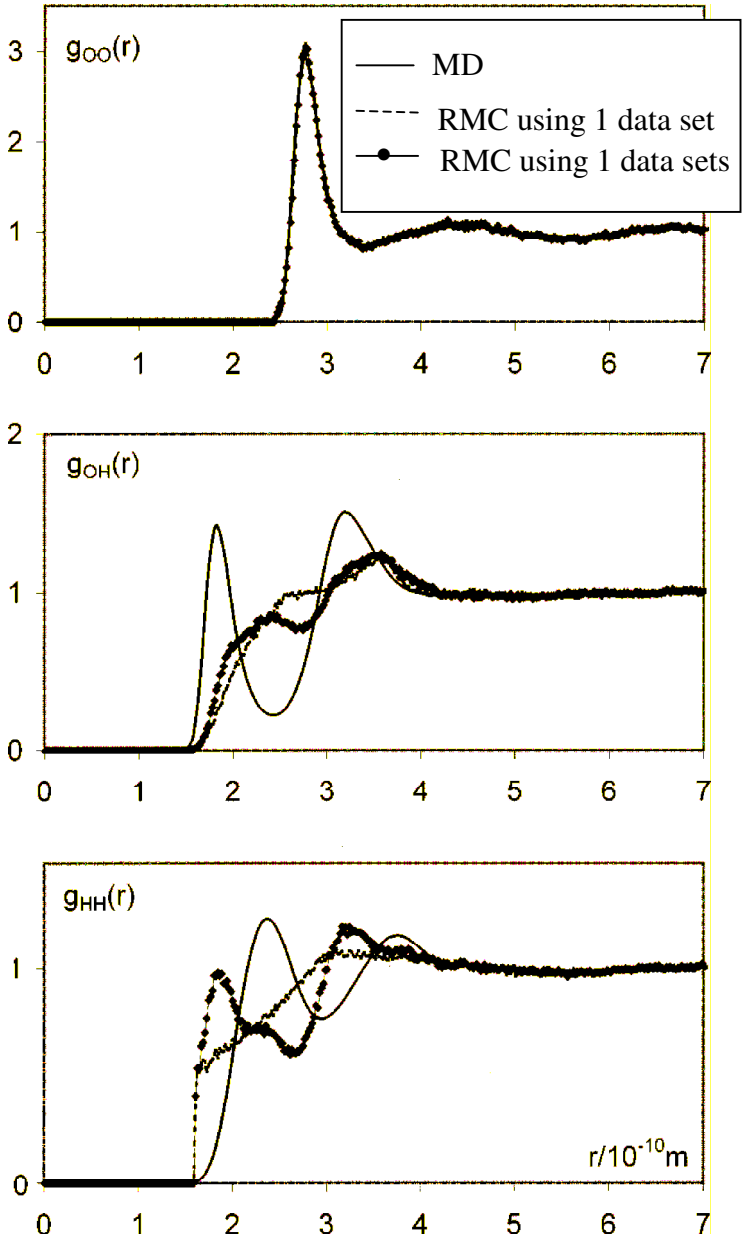
- normal MD simulation (TIP4P water)

- RMC, fitting the X-ray $S(Q)$ of the MD data

- RMC fitting the X-ray $S(Q)$ and neutron $S(Q)$ of D_2O of the MD data

The pair correlation functions cannot be reproduced if the number of used data sets is less than the number of partials.

Reason: highly ordered local structure of water



Tóth, Baranyai, *J. Chem. Phys.* **114,** 2027 (2001)

- Shall we fit the measured total, or the separated partial functions ?

- again, the general rule is to make as little processing on the primer data as possible before being fed to the RMC procedure
RMC can do the separation of the partials
(But the question is always there how reliable this separation is if there are less independent data sets than partials!)

- How many independent experimental data sets do we need ?

- as much as possible – the more experimental information we have on the system the more reliable is our model
Results of new experiments can be predicted!

- Shall we move atoms or molecules ?

- better to move atoms rather than molecules because the atoms are the scattering units (unless we can reliably subtract the intramolecular term from the scattering data)
But we have to keep the molecules together somehow!

- Can we extract the potential function from the RMC results?

- several attempts have been made
  (e.g., → start with a tentative $u(r)$,
  → do a MC simulation
  → alter $u(r)$ randomly
  → do a new simulation with the new $u(r)$
  → see if the random change in $u(r)$ led to an improvement in the reproduction of the experimental data
  → decide on the acceptance of the change in $u(r)$
  → repeate the procedure until sufficient convergence

  use a similar procedure but describe u(r) analytically (e.g., in the form of the Lennard-Jones potential) and alter only its parameters

  use EPSR

- we can extract a potential that reasonably reproduces the structure
  BUT it is not reliable from the thermodynamic point of view

- thermodynamic properties are much more sensitive to the details of the potential function than the structure

- Is the resulting structure unique ?

No

- If the potential is pairwise additive then $g(r)$ determines $u(r)$ up to a constant

Henderson, *Phys. Lett.* **49A,** 197 (1974)

Since $u(r)$ determines uniquely not only $g(r)$ but also $\rho^{(3)}(r_1,r_2,r_3)$, ..., $\rho^{(N)}(\mathrm{r}_1,\mathrm{r}_2,...,\mathrm{r}_N)$, in this case $g(r)$ should determine all the higher order correlation terms, and thus the RMC results should be unique

- generalization of the Henderson theorem:
if the potential contains 2, 3,..., and i-body terms but not higher order terms than i then the set of $\rho^{(2)}(r_1,r_2)$, $\rho^{(3)}(r_1,r_2,r_3)$, ..., $\rho^{(i)}(\mathrm{r}_1,\mathrm{r}_2,...,\mathrm{r}_i)$ correlation functions uniquely determine the potential, and hence the full structure (including higher order correlation terms)

Evans, *Mol. Simul.* **4,** 409 (1990)

- The potential in a real system is never purely pairwise additive, but only two-body correlation can be measured

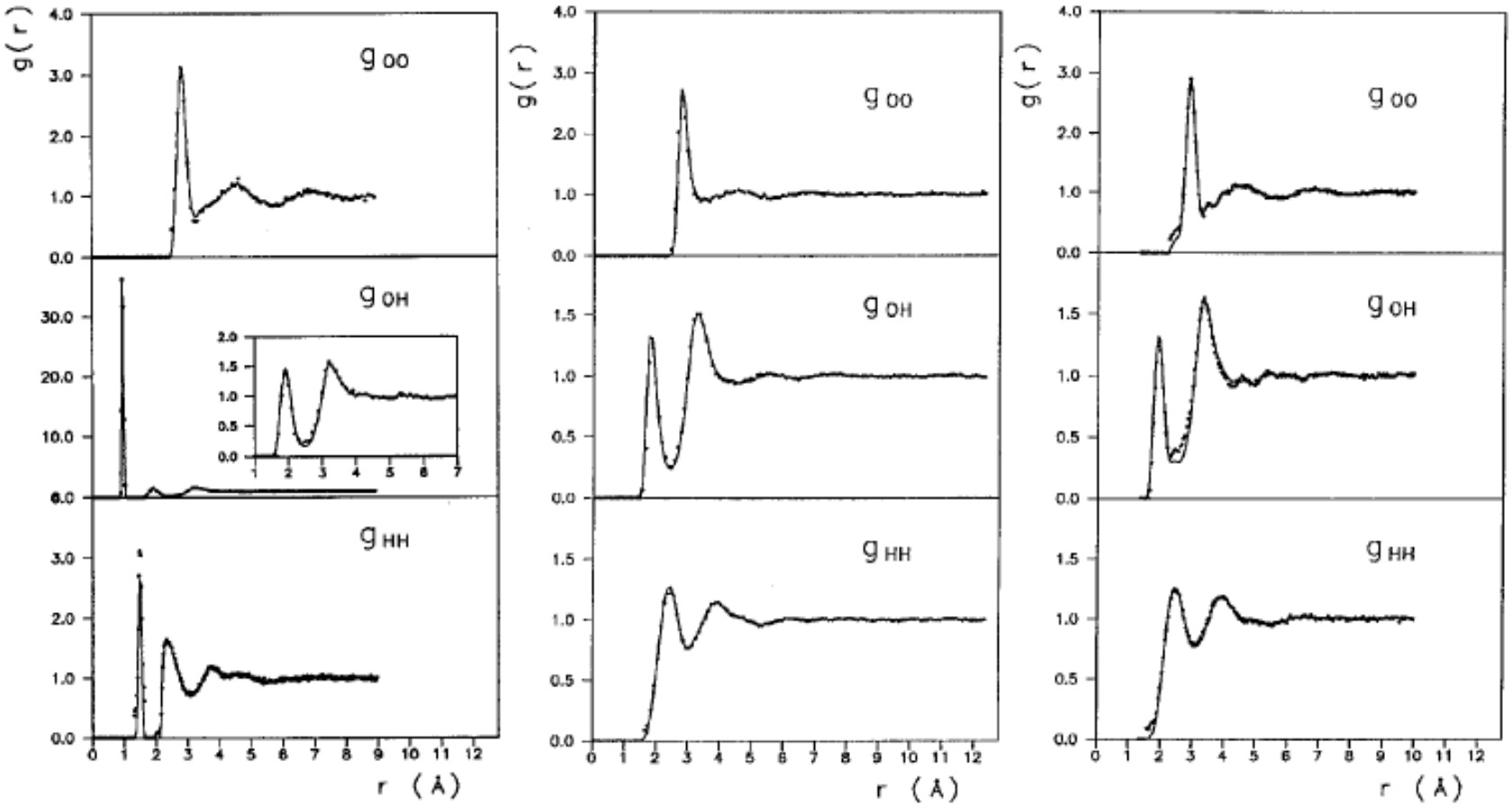
How important are the many-body forces in the system studied

Systematic investigation of the problem of uniqueness

RMC fit of neutron data

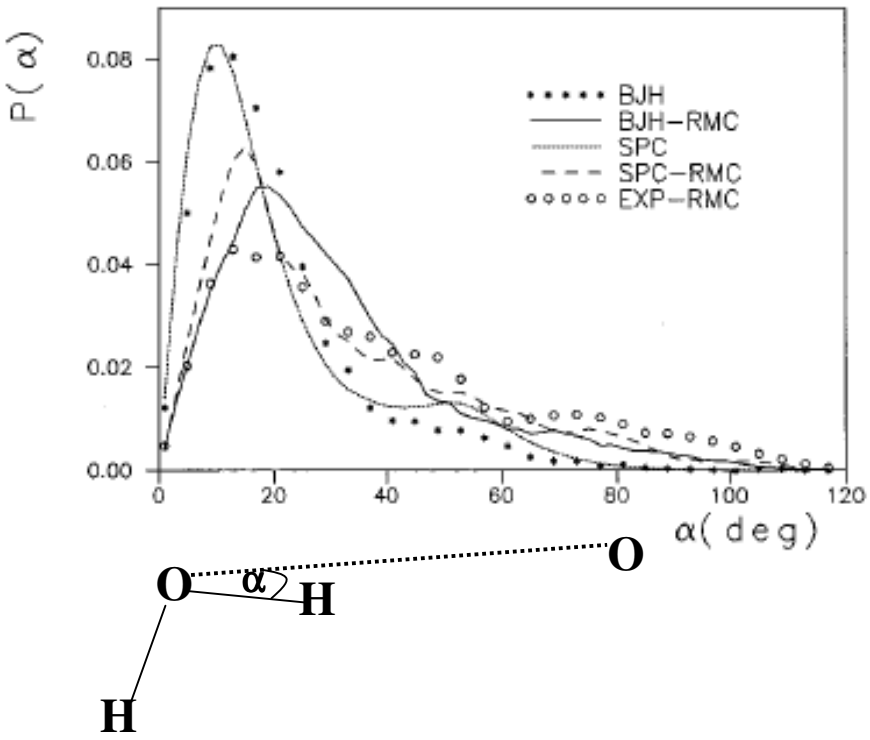
RMC fit of MD results with the (flexible) BJH model

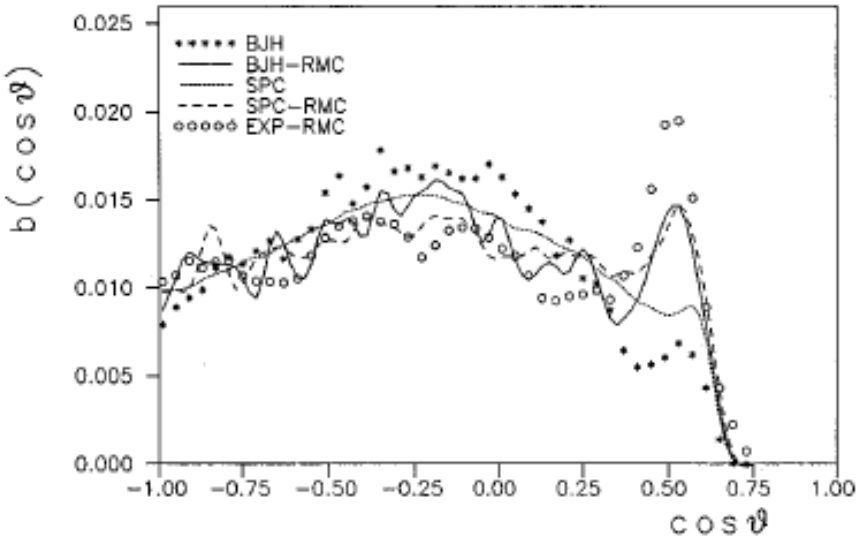
RMC fit of MD results with the (rigid) SPC model

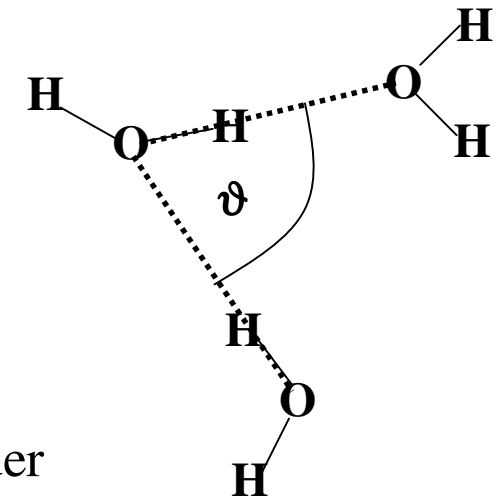


PJ, Bakó, Pálinkás, Radnai, Soper, *J. Chem. Phys.* **105,** 245 (1996)

Reproduction of angular distributions







- The original distributions are not reproduced accurately
- But the main features of the original data are reproduced
- The reproduced (RMC) data always correspond to lower order
- A level of uncertainty of the RMC data can be estimated

Summary – advantages and limitations of RMC

advantages

- produces 3D atomic configurations consistent with experimental data
- fits experimental data better than any conventional simulation
- random errors can be estimated
- systematic errors can be detected
- partials can be separated
- results of new experiments can be predicted

limitations

- thermodynamic information is lost
- reliability of the results is limited by the quality of the data
- not unique