Structural quantities

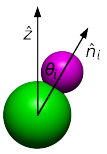
Correlation functions

- radial distribution function (RDF, also pair correlation/distribution function), $g(r)$ = probability of finding a particle at distance r (from another particle), normalized to ideal gas
- structure factor (diffraction → Fourier transform of $g(r)$)
- angular correlation function – good for small nonspherical molecules
- time autocorrelation functions

Order parameters

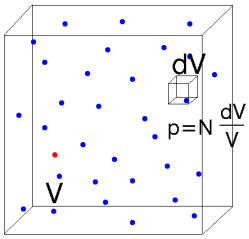
- Ordering in the z-direction: $\langle \sum_i \cos\theta_i \rangle$: ↑ vs. ↓
- For symmetric molecules: $\left\langle \sum_i \left(\frac{3}{2}\cos^2\theta_i - \frac{1}{2}\right)\right\rangle$: ↕ vs. ↔
- Nematic liquid crystal – the "director" is not known:

$$Q_{\alpha\beta} = \left\langle \sum_i \left(\frac{3}{2}\hat{n}_{\alpha,i}\hat{n}_{\beta,i} - \frac{1}{2}\delta_{\alpha\beta}\right)\right\rangle,$$ eigenvector of the max. eigenvalue = director

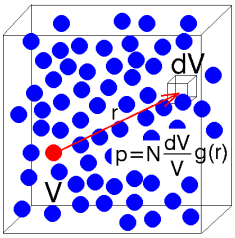


- Formulas detecting local order (e.g., tetrahedral around water molecules as Steinhardt's parameters based on spherical harmonics), onset of crystallization can be detected . . .

Structure of fluids – correlation functions



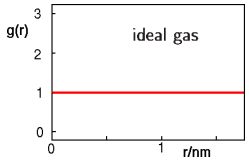
randomly distributed molecules (ideal gas)

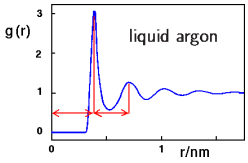


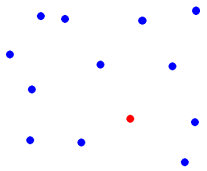
liquid

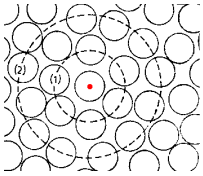
$g(r)$ = pair correlation function = radial distribution function = probability density of finding a particle r apart from another particle, normalized so that for randomly distributed particles (ideal gas) it is 1

Structure of fluids – correlation functions

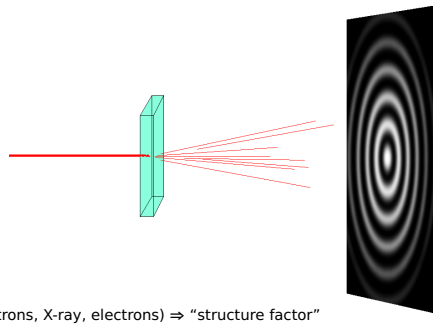






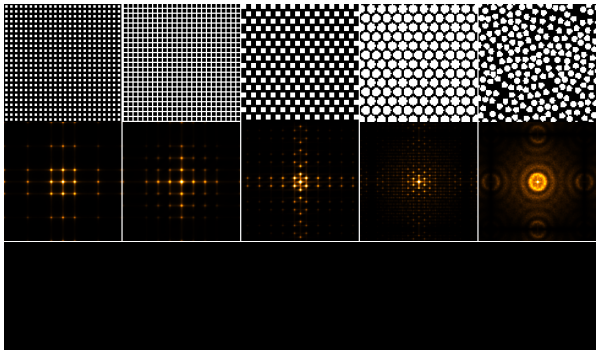


How to obtain structure – experiment

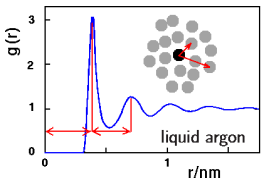


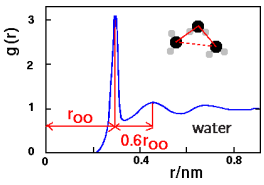
- Diffraction (neutrons, X-ray, electrons) ⇒ "structure factor"
- inverse Fourier transform ⇒ RDF

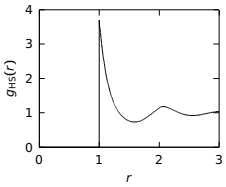
How to obtain structure



Argon, hard spheres, water







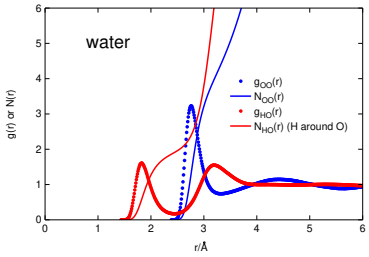
- The structure of simple fluid (argon, HS) is organized by shells.
- The structure of water is determined by the tetrahedral geometry of hydrogen bonds.
- After several molecular diameters, the correlations decay to zero.

Running coordination number

Also "cumulative radial distribution function" $\rho = N/V$ = number density

$$N(r) = 4\pi\rho \int_0^r g(r')r'^2 \mathrm{d}r'$$

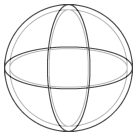
For $r_{\min}$ = first minimum on the RDF curve, $N(r_{\min})$ = "coordination number" = averaged number of molecules in the first shell



RDF from simulations – simple

Histogram of the number of particle pairs, h_i, so that

$$r \in [r_i - \Delta r/2, r_i + \Delta r/2) \quad \text{alternatively: } \mathcal{I}_i = [r_i, r_i + \Delta r)$$



The volume of the shell

$$\Delta V_i = \frac{4\pi}{3}\left[\left(r_i + \frac{\Delta r}{2}\right)^3 - \left(r_i - \frac{\Delta r}{2}\right)^3\right]$$

Mean number of molecules around a selected particle in case of uniformly distributed molecules (ideal gas, $\rho = N/V$):

$$\rho \Delta V_i$$

Sum over all particles (1/2 to count each pair just once):

$$h_i^{\text{id. gas}} = \frac{N}{2}\rho\Delta V_i$$

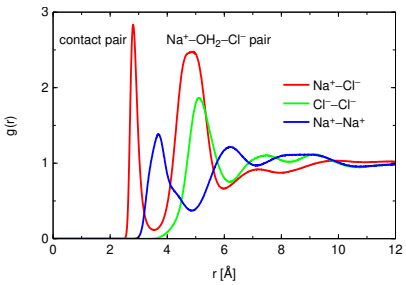
Radial distribution function:

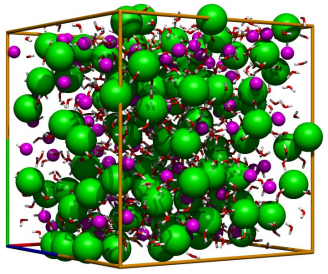
$$g(r_i) = \frac{\langle h_i \rangle}{h_i^{\text{id. gas}}} = \frac{2\langle h_i \rangle}{N\rho\Delta V_i}$$

Another example – brine

show/ssbrine.sh

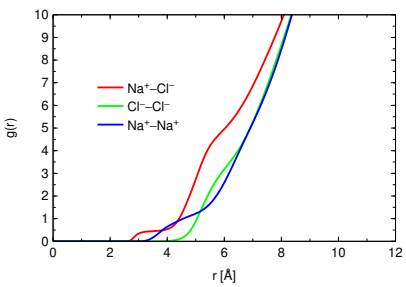
RDFs of supersaturated (10.5 mol/kg) solution of NaCl:

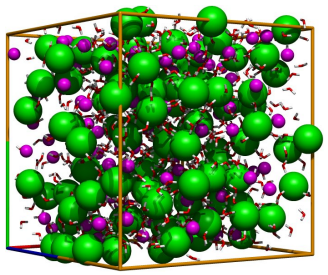




Another example – brine

CNs of supersaturated (10.5 mol/kg) solution of NaCl:





SIMOLANT – installation (Windows)

How to install SIMOLANT (Windows):

- Google 🔍 simolant or links:
  - https://github.com/kolafaj/SIMOLANT → executables
  - http://old.vscht.cz/fch/software/simolant
- Download `simolant-win.zip`
- Create a folder and unpack `simolant-win.zip` into it
  Do not run directly from "folder" `simolant-win.zip`
  - help would not work
  - you would not find the created files
- Run `simolant.exe`

Hints:

- The calculated data are exported to protocols `simolant.txt` and `simolant.csv` with decimal dot, the CSV separator is comma.
  If you like decimal comma (e.g., for localized excel) and CSV separator semicolon, select comma in panel "Measure" (so that it gets green: comma).[a]
- If you restart SIMOLANT, the old versions of `simolant.txt` and `simolant.csv` are overwritten by the new files. To avoid this, you may change the base name `simolant` in Menu: File → Protocol name..

[a] SIMOLANT does not detect localization and never will.

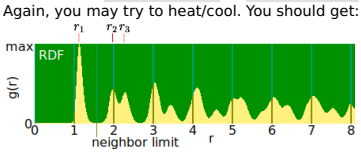
SIMOLANT: Observe RDF yourself

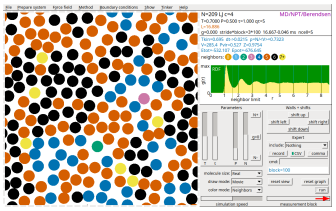
`simolant -l10 -N209 -Pblock=100`

- Menu: Prepare system → Periodic liquid

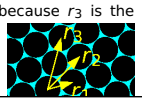
The RDF graph is scaled from 0 to maximum.

- For smoother results, move slider "measurement block" to maximum (esp. for gas).
- Increase/decrease temperature a bit (slider "T"). You may change pressure, too (slider "P").
- Set the temperature to maximum ($T = 5$): only a small first peak should remain.
- Decrease temperature below 0.4 to crystallize.
- To obtain a nearly perfect crystal without defects, select Menu: Prepare system → Periodic crystal. Again, you may try to heat/cool. You should get:





- Apparently $r_3/r_1 = 2$ because r_3 is the second neighbor ●●●
- Similarly, $r_2/r_1 = \sqrt{3}$



Radial distribution function in the *NVT* ensemble

+

3D (e.g., in periodic b.c.), good for a homogeneous and isotropic liquid:

$$g(r) \equiv g(r_{12}) \overset{\text{isotr.}}{=} g(\vec{r}_{12}) \overset{\text{homog.}}{=} g(\vec{r}_1, \vec{r}_2) = \frac{N(N-1)}{\rho^2 Q_{\text{NVT}}} \int \dots \int \exp[-\beta U(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)]\, \mathrm{d}\vec{r}_3 \dots \mathrm{d}\vec{r}_N$$

Equivalently: $g(\vec{r}) = \left(1 - \frac{1}{N}\right) V \langle \delta(\vec{r}_{12} - \vec{r}) \rangle$

For a mixture: $g_{ij}(\vec{r}) = V \langle \delta(\vec{r}_{12} - \vec{r}) \rangle$

Normalization and thermodynamic limit (fluid): $\lim_{N\to\infty, r\to\infty} g(r) = 1$

Using n-particle distribution functions:

$$g(\vec{r}_{12}) = \frac{\rho_2(\vec{r}_1, \vec{r}_2)}{\rho_1(\vec{r}_1)\rho_2(\vec{r}_2)},$$

where $\rho_1 = \rho$ in a homogeneous fluid.

# of particles around one chosen particle in NVT "over whole system" (imagine periodic b.c.):

$$\int_V \rho g(\vec{r}) \mathrm{d}\vec{r} = N - 1$$

Ideal gas: $g(\vec{r}) = 1 - \frac{1}{N}$

Extension (short-range potentials): $\lim_{r\to\infty} g(r) = 1 - \frac{\kappa_T}{N\kappa_T^{\text{id. gas}}} + \mathcal{O}(1/N^2)$, where κ_T = compressibility

Calculation of RDF in simulations – exact

+

Histogram of the count of pairs of particles, h_i, so that $r \in \mathcal{I}_i$

$$\mathcal{I}_i = [r_i - \Delta r/2, r_i + \Delta r/2), \quad \text{optionally } \mathcal{I}_i = [r_i, r_i + \Delta r)$$

$$r_i = i\Delta r, \quad i = 1, \dots, i_{\max}$$

$$\begin{aligned}
\langle h_i \rangle &= \frac{1}{Q_{\text{NVT}}} \sum_{j<k} \int_{r_{jk} \in \mathcal{I}_i} \exp[-\beta U(\vec{r}^N)]\, \mathrm{d}\vec{r}^N \\
&= \frac{1}{Q_{\text{NVT}}} \binom{N}{2} V \int_{r_{12} \in \mathcal{I}_i} \left\{ \int \exp[-\beta U(\vec{r}^N)]\, \mathrm{d}\vec{r}_3 \dots \mathrm{d}\vec{r}_N \right\} \mathrm{d}\vec{r}_{12} \\
&= \frac{N}{2} \rho \int_{\mathcal{I}_i} g(r) \mathrm{d}\vec{r} \\
&\approx \frac{N^2}{2V} g(r_i) \Delta V_i
\end{aligned}$$

The formula again:

$$g(r_i) = \frac{2\langle h_i \rangle}{N \rho \Delta V_i}$$

Correlation function and thermodynamics

+

For simple fluid (point particles interacting by a pair spherically symmetric potential):

Residual internal energy:

$$\langle U \rangle = \frac{1}{Q_{\text{NVT}}} \int \sum_{i<j} u_{ij}(r_{ij}) \mathrm{e}^{-\beta U} \mathrm{d}\vec{r}_1 \dots \mathrm{d}\vec{r}_N$$

$$\langle U \rangle = \binom{N}{2} \frac{V}{Q_{\text{NVT}}} \int \mathrm{e}^{-\beta U} 4\pi r_{12}^2 \mathrm{d}r_{12} u(r_{12}) \mathrm{d}\vec{r}_3 \dots \mathrm{d}\vec{r}_N$$

$$\langle U \rangle = \frac{N}{2} \rho \int u(r) g(r) \mathrm{d}\vec{r} = 2N\pi\rho \int u(r) g(r) r^2 \mathrm{d}r$$

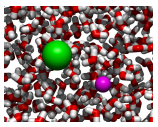
Pressure:

$$\frac{\beta P}{\rho} = 1 - \frac{2\pi}{3} \beta \rho \int g(r) u'(r) r^3 \mathrm{d}r$$

Case study: RDF and PMF in infinite dilution. 1. Direct simulation

The potential of mean force (PMF) between two particles (ions) in a solution is

$$U_{\text{PMF}}(r) = -RT \ln g(r).$$

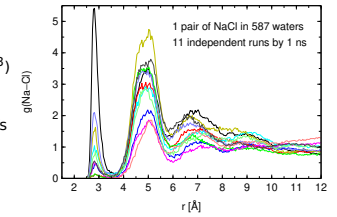


Calculating $g(r)$ or $U_{\text{PMF}}(r)$ in **infinite dilution** is demanding:

- poor statistics (only 1 pair of ions),
- possible barrier between contact and water-separated ion pairs.

Example: NaCl in water

- Force field: Joung–Cheatham NaCl + SPC/E water
- 1 pair of Na^+Cl^- in 587 water molecules (box 26 Å^3)
- $T = 300$ K, $\rho = 1002.5$ kg m^{-3}
- mass repartitioning: equalization factor 0.8, $h = 2$ fs
- 1 measurement every 10 fs
- RDF range $c = 12$ Å



RDF+PMF. 2. Why is it so inaccurate?

Barrier crossing:

- $g_{\min} \approx 0.1 \Rightarrow$ 10× typical time (in water ≈ 1 ps) needed,
- it is not so bad, but worsens the time scale caused by diffusion

Diffusion:

- Experimental molar conductivities [S m^2 mol^{-1}]: $\lambda_\oplus = 0.00501, \lambda_\ominus = 0.00763$
- Approximate mutual conductivity (not so important): $\lambda = \sqrt{\lambda_\oplus^2 + \lambda_\ominus^2}$
- (Mutual) diffusion coefficient: $D = \lambda RT/F^2 = 2.4 \cdot 10^{-9}$ m^2 s^{-1}
- Distance to travel: $d = c/2$ (c = cutoff = 12 Å)
- Typical diffusion time (1D approx.): $t = d^2/2/D = 70$ ps

Both phenomena together cause poor convergence in time scale $t \times 10 \sim 1$ ns

RDF+PMF. 3. Metadynamics $h(r) = 1$

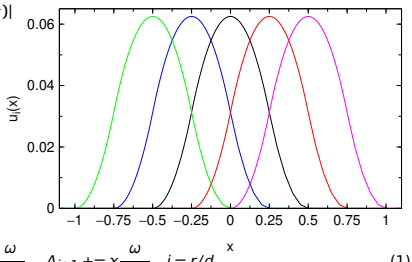
Generalized coordinate: $r = |\vec{r}(\text{Na}^+) - \vec{r}(\text{Cl}^-)|$

The bias potential is

$$U_{\text{bias}}(r) = \frac{1}{2d^2} \sum_{i=0}^{n_{\max}} A_i u(r - id),$$

where $d = 0.25$ Å and

$$u(x) = \begin{cases} 0 & \text{for } x < -2d, \\ (2d + x)^2/2 & \text{for } x \in [-2d, -d], \\ d^2 - x^2/2 & \text{for } x \in [-d, +d], \\ (2d - x)^2/2 & \text{for } x \in [+d, +2d], \\ 0 & \text{for } x > +2d \end{cases}$$

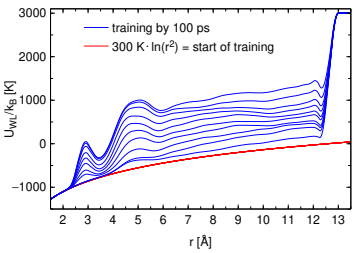


Wang–Landau update: $A_i \mathrel{+}= (d - x) \cdot \frac{\omega}{h(r)}, \quad A_{i+1} \mathrel{+}= x \frac{\omega}{h(r)}, \quad i = r/d$ (1)

- Starting from a basin $U = 0$ inside wall $A_{n_{\max}} = 3000$ K
- ω selected so that the basin fills to 40% of $A_{n_{\max}}$
- Simple version: $h(r) = 1$ (the density of Cartesian points will be included later)
- Opposite possibility: $h(r) = 4\pi r^2$ (see below)

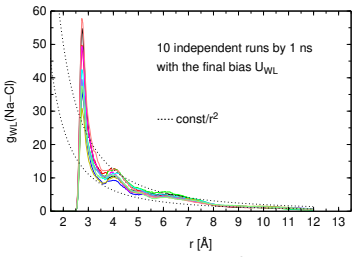
RDF+PMF. 4. Metadynamics $h(r) = 1$ raw results

Training $U_{\text{WL}} \approx U_{\text{PMF}} + \text{const}$, total time = 1 ns



- The convergence is slow.
- Does not matter, faster runs with bias will follow.

Raw radial distribution function with the bias



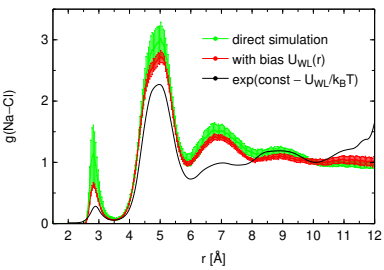
- The difference from const/r^2 is due to discretization, peaks appear for large dU_{PMF}/dr.
- Finer grid would be problematic close to cation–anion contact.

RDF+PMF. 5. Metadynamics $h(r) = 1$ final results

- — 10 independent $g(r)$'s with the bias, multiplied by $\exp(-U_{\text{WL}}/k_{\text{B}}T)$ and shifted to $g(10\text{ Å}) = 1$
- — direct simulation
- RDF grid 0.04 Å, standard errors from 11 or 10 blocks
- Both methods are based on 11 ns trajectories (comparable times)

Summary:

- Wang–Landau is more accurate at short separations.
- For long r, the accuracy is comparable.
- Sampled uniformly: half samples $r < c/2$, half samples $r > c/2$.
- Short separations are oversampled!

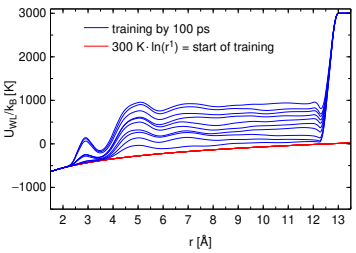


Can we do it better?

- Yes, let's try pre-bias $h(r) = r$ in (1)!
- $h(r) = r^2$ (correct Cartesian volume element) would oversample long separations

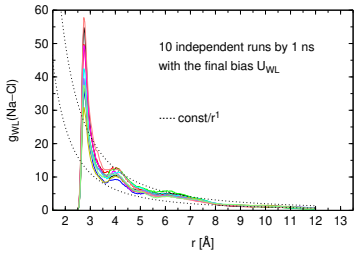
RDF+PMF. 6. Metadynamics $h(r) = r$ raw results

Training $U_{WL} \approx U_{PMF}$ + const, total time = 1 ns



- The convergence is slow at short separations.

Raw radial distribution function with the bias



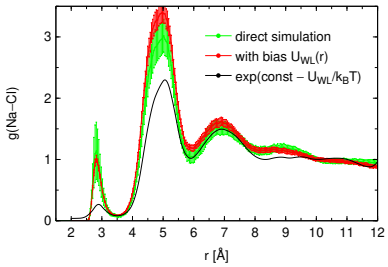
- The difference from const/r^2 is due to discretization, peaks appear for large dU_{PMF}/dr.
- Finer grid would be problematic close to cation–anion contact.

RDF+PMF. 7. Metadynamics $h(r) = r$ final results

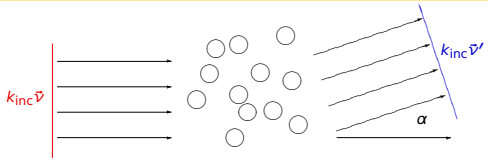
- — 10 independent $g(r)$'s with the bias, multiplied by $\exp(-U_{WL}/k_BT)$ and shifted to $g(10\,\text{Å}) = 1$
- — direct simulation
- RDF grid 0.04 Å, standard errors from 11 or 10 blocks
- both methods are based on 11 ns trajectories (comparable times)

Summary:

- Accuracy is spread more uniformly over distances.



Structure factor



- Weak (the same wave scatters only once)
- Distance of atom $\vec{r}_j$ from plane $\hat{v}\cdot\vec{r} = 0$ is $\hat{v}\cdot\vec{r}_j$ (for $|\hat{v}| = 1$).
- Distance from wave source (wave plane of eq. $\hat{v}\cdot\vec{r} = \text{const}$) via atom r_j to the detector ($\hat{v}'\cdot\vec{r} = \text{const}$) is $\hat{v}\cdot\vec{r}_j - \hat{v}'\cdot\vec{r}_j + \text{const}$
- Wave vector $\vec{k}_{inc} = k_{inc}\hat{v}$, $k_{inc} = 2\pi/\lambda$
- Formally the incident wave is (except phase):

Definition $k = 2\pi/\lambda$ is common in physics wheras $k = 1/\lambda$ in crystallography; then, factor 2π is in the exponential: $\exp[2\pi i\vec{k}\cdot\vec{r}]$.

$$\exp[\,i(k_{inc}\hat{v} - k_{inc}\hat{v}')\cdot\vec{r}_j] = \exp[\,i\vec{k}\cdot\vec{r}_j], \quad \text{where} \quad \vec{k} = k_{inc}\hat{v} - k_{inc}\hat{v}', \quad k = |\vec{k}| = 2\sin(\alpha/2)k_{inc} \approx \alpha k_{inc}$$

Structure factor

Waves scattered by all atoms interfere:

$$K = \sum_{j=1}^{N} \exp[\,i(k_{inc}\hat{v} - k_{inc}\hat{v}')\cdot\vec{r}_j] = \sum_{j=1}^{N} \exp[\,i\vec{k}\cdot\vec{r}_j], \qquad \text{signal} \propto \langle|K|^2\rangle$$

Definition of the structure factor (pure compound):

$$S(\vec{k}) = \left\langle \frac{1}{N}\left|\sum_{j=1}^{N}\exp(-i\vec{k}\cdot\vec{r}_j)\right|^2 \right\rangle$$

It is a function of vector $\vec{k}$, we may imagine it in the periodic b.c. in a cube of edge L:

$$\vec{k} = \frac{2\pi\vec{n}}{L},\ \vec{n} \in \mathbb{Z}^3$$

Similarly, a 3D RDF is (not exactly isotropic even for a liquid)

$$g(\vec{r}_{12}) = \frac{N(N-1)}{\rho^2 Q_{NVT}} \int_{L^3}\cdots\int_{L^3} \exp[-\beta U(\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_N)]\,\mathrm{d}\vec{r}_3\ldots\mathrm{d}\vec{r}_N$$

Structure factor and RDF

In a periodic box $V = L^3$; $\vec{k} = 2\pi\vec{n}/L$:

$$\begin{aligned} S(\vec{k}) &= \left\langle \frac{1}{N}\left|\sum_{j=1}^{N}\exp(-i\vec{k}\cdot\vec{r}_j)\right|^2 \right\rangle \\ &= \left\langle \frac{1}{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\exp(-i\vec{k}\cdot\vec{r}_j)\exp(i\vec{k}\cdot\vec{r}_l) \right\rangle = \left\langle \frac{1}{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\exp(i\vec{k}\cdot\vec{r}_{jl}) \right\rangle \\ &= 1 + \frac{1}{N}\left\langle \sum_{j<l}^{N}[\exp(-i\vec{k}\cdot\vec{r}_{jl}) + \exp(i\vec{k}\cdot\vec{r}_{jl})] \right\rangle = 1 + \frac{1}{N}\left\langle \binom{N}{2}[2\cos(\vec{k}\cdot\vec{r}_{jl})] \right\rangle \\ &= 1 + \rho\int_{L^3} g(\vec{r})\cos(\vec{k}\cdot\vec{r})\,\mathrm{d}\vec{r} \end{aligned}$$

The integral diverges as $V \to \infty \Rightarrow$ we subtract $0 = \int_V \cos(\vec{k}\cdot\vec{r})\mathrm{d}\vec{r}$, $\vec{k} = 2\pi\vec{n}/L$
(other method: convergence factor $\exp(-\alpha k^2)$, $\lim_{\alpha\to 0}$)

$$S(\vec{k}) = 1 + \rho\int_V [g(\vec{r}) - 1]\cos(\vec{k}\cdot\vec{r})\mathrm{d}\vec{r} \overset{V\to\infty}{=} 1 + \rho\int [g(\vec{r}) - 1]\cos(\vec{k}\cdot\vec{r})\mathrm{d}\vec{r}$$

Structure factor and RDF

Since $g(\vec{r}) = g(-\vec{r})$, we have

$$S(\vec{k}) - 1 = \rho\int [g(\vec{r}) - 1]\cos(\vec{k}\cdot\vec{r})\mathrm{d}\vec{r} = \rho\int [g(\vec{r}) - 1]\mathrm{e}^{-i\vec{k}\cdot\vec{r}}\mathrm{d}\vec{r}$$

and the 3D inverse Fourier transform is

$$g(\vec{k}) - 1 = \frac{1}{8\pi^3\rho}\int [S(\vec{k}) - 1]\mathrm{e}^{i\vec{k}\cdot\vec{r}}\mathrm{d}\vec{k}$$

Once again using $S(\vec{k}) = S(-\vec{k})$:

$$g(\vec{k}) - 1 = \frac{1}{8\pi^3\rho}\int [S(\vec{k}) - 1]\cos(\vec{k}\cdot\vec{r})\mathrm{d}\vec{k}$$

The $k \to 0$ limit:

$$NVT: \quad S(0) = 0$$

$$\mu VT: \quad S(0) = 1 + \rho\int [g(\vec{r}) - 1]\mathrm{d}\vec{r} = k_BT\left(\frac{\partial\rho}{\partial p}\right)_T \quad \text{(compressibility equation)}$$

Isotropic structure factor

- Let's choose $\vec{k} = (0, 0, k)$ and spherical coordinates (ϕ, θ, r).
- Then $\vec{k}\cdot\vec{r} = kr\cos\theta$ and (g and S are odd functions)

$$\begin{aligned} S(k) - 1 &= \rho\int [g(\vec{r}) - 1]\cos(\vec{k}\cdot\vec{r})\mathrm{d}\vec{r} \qquad \text{(subst. } z = \cos\theta) \\ &= \rho\int_0^\infty r^2\mathrm{d}r\int_0^{2\pi}\mathrm{d}\phi\int_{-1}^{1}[g(r) - 1]\cos(krz)\mathrm{d}z \\ &= 2\pi\rho\int_0^\infty r^2[g(r) - 1]\mathrm{d}r\left[-\frac{\sin(krz)}{kr}\right]_{-1}^{1} \\ &= \frac{4\pi\rho}{k}\int_0^\infty r[g(r) - 1]\sin(kr)\mathrm{d}r \end{aligned}$$

In the same way we can evaluate the inverse Fourier transform:

$$\rho[g(r) - 1] = \frac{1}{8\pi^3}\int [S(\vec{k}) - 1]\cos(\vec{k}\cdot\vec{r})\,\mathrm{d}\vec{k} = \frac{1}{2\pi^2 r}\int_0^\infty \sin(kr)[S(k) - 1]\,k\mathrm{d}k$$

$$r[g(r) - 1] \leftrightarrow k[S(k) - 1] \quad \text{(sine transform)}$$

$S(k)$ from simulations

$S(\vec{k})$ in simulations can be calculated directly by the same code as the k-space part of the Ewald summation. If needed, we may sphericalize it:

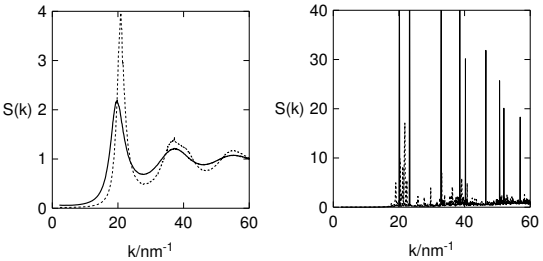
$$S(k) = \sum_{k=|\vec{k}|} S(\vec{k}) \Big/ \sum_{k=|\vec{k}|} 1$$

$S(k)$ can be obtained by the inverse Fourier transform of $g(r)$. The RDF $g(r)$ must have a long range and/or it has to be filled (by 1) for long distances.

Alternatively, 3D $S(\vec{k})$ can be obtained by the inverse Fourier transform from RDF $g(\vec{r})$ and vice versa.

- Simulation-based $S(k)$ for a mixture (where individual site-site RDFs need not be available) allows for a direct comparison with the experiment
- $S(k)$ and $S(\vec{k})$ may detect unwanted crystallization of a supercooled system

Structure factor



— liquid - - - glass ($T = 0$)

— crystal - - - crystallite
(both cooled to $T = 0$)

Structure factor for multiatomic systems

Structure factor for a "mixture" of atoms, b_j = coherent scattering length for point scattering (nuclear scattering of neutrons), in general integration over the scatterer density (form factor)

$$S(k) = 1 + N\frac{\langle|Q(\vec{k})|^2\rangle - \sum_j b_j^2}{\left(\sum_j b_j\right)^2}$$

$$Q(\vec{k}) = \sum_j b_j \exp[-2\pi i\vec{k}\cdot\vec{r}_j/L]$$

$$S = \sum_I\sum_J w_{IJ}S_{IJ}, \qquad w_{IJ} = \frac{N_Ib_IN_Jb_J}{(\sum_I N_Ib_I)^2}$$

N_I = number of atoms of type I ($\sum_I N_Ib_I = \sum_j b_j$)