

An overview and discussion of correlation functions

Clark grp mtg 3-11/2021

First - Distribution Functions

- Statistics/Probability Theory -

The cumulative distribution function of a real-valued **random variable** X is the function given by^{[2]:p. 77}

$$F_X(x) = \mathbb{P}(X \leq x) \quad (\text{Eq.1})$$

where the right-hand side represents the **probability** that the random variable X takes on a value less than or equal to x .

Discrete vs. continuous variables within molecular simulations:

Discrete: an observation of a phenomena → in a snapshot do we observe x or y? its either a yes or a no
(discrete data is counted data)

Continuous: a variable that has an infinite number of values between any two values (continuous data is measured data)

- Time should be continuous, but we sample time in discrete intervals within our simulations...so this is an example of how discretize a continuous variable into a discrete variable
- Averaging of a discrete variable can create a continuous representation and thus turn it into a continuous variable, or fitting discrete variables to a function can create a continuous representation, or interpolation of discrete data can turn it continuous, also lots of measurements can make it pretty continuous

Most of what we do involves **discrete** variables:

The probability distribution of a discrete random variable X is a list of each possible value of X together with the probability that X takes that value in one trial of the experiment.

The probabilities in the probability distribution of a random variable X must satisfy the following two conditions:

- Each probability $P(x)$ must be between 0 and 1:

$$0 \leq P(x) \leq 1. \quad (4.2.1)$$

- The sum of all the possible probabilities is 1:

$$\sum P(x) = 1. \quad (4.2.2)$$

Most of what we do involves **discrete** variables:

However, if we have many many measurements we can apply tools for continuous variables:

The probability density function of a continuous random variable can be determined from the cumulative distribution function by differentiating^[3] using the [Fundamental Theorem of Calculus](#); i.e. given $F(x)$,

$$f(x) = \frac{dF(x)}{dx}$$

as long as the derivative exists.

The CDF of a [continuous random variable](#) X can be expressed as the integral of its probability density function f_X as follows:^[2]p. 86

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

In the case of a random variable X which has distribution having a discrete component at a value b ,

$$P(X = b) = F_X(b) - \lim_{x \rightarrow b^-} F_X(x).$$

If F_X is continuous at b , this equals zero and there is no discrete component at b .

What if there are multiple variables?

- If we are interested in how two random variables track with one another, or are *correlated*, we get a correlation function (a special type of distribution function).
 - Often contingent on spatial or temporal distance (indicating dependencies based upon distance or time)
 - If it is the same variable measured at two different points → autocorrelation function
 - Correlation functions of different random variables are often called cross-correlation functions

Correlation functions

- If we are interested in how two random variables track with one another, or are *correlated*, we get a correlation function (a special type of distribution function).
 - Often contingent on spatial or temporal distance (indicating dependencies based upon distance or time)
 - If it is the same variable measured at two different points → autocorrelation function
 - Correlation functions of different random variables are often called cross-correlation functions

Correlation functions

- The ensemble you are in is important – we are often dealing with the thermal/canonical average of two random variables

The most common definition of a correlation function is the [canonical ensemble](#) (thermal) average of the scalar product of two random variables, s_1 and s_2 , at positions R and $R + r$ and times t and $t + \tau$:

$$C(r, \tau) = \langle s_1(R, t) \cdot s_2(R + r, t + \tau) \rangle - \langle s_1(R, t) \rangle \langle s_2(R + r, t + \tau) \rangle.$$

Here the brackets, $\langle \dots \rangle$, indicate the above-mentioned thermal average. It is a matter of convention whether one subtracts the uncorrelated average product of s_1 and s_2 , $\langle s_1(R, t) \rangle \langle s_2(R + r, t + \tau) \rangle$ from the correlated product, $\langle s_1(R, t) \cdot s_2(R + r, t + \tau) \rangle$, with the convention differing among fields.

You are the master of your destiny here – You decide the variables of relevance to your specific scientific questions

Correlation functions

- The ensemble you are in is important – we are often dealing with the thermal/canonical average of two random variables

The most common definition of a correlation function is the **canonical ensemble** (thermal) average of the scalar product of two random variables, s_1 and s_2 , at positions R and $R + r$ and times t and $t + \tau$:

$$C(r, \tau) = \langle \mathbf{s}_1(R, t) \cdot \mathbf{s}_2(R + r, t + \tau) \rangle - \langle \mathbf{s}_1(R, t) \rangle \langle \mathbf{s}_2(R + r, t + \tau) \rangle.$$

Here the brackets, $\langle \dots \rangle$, indicate the above-mentioned thermal average. It is a matter of convention whether one subtracts the uncorrelated average product of s_1 and s_2 , $\langle \mathbf{s}_1(R, t) \rangle \langle \mathbf{s}_2(R + r, t + \tau) \rangle$ from the correlated product, $\langle \mathbf{s}_1(R, t) \cdot \mathbf{s}_2(R + r, t + \tau) \rangle$, with the convention differing among fields.

Lets say that S1 and S2 are the same variable...autocorrelation function, and lets say it is the position of a specific atom type, like O-O in water --> position-position autocorrelation function for O-atoms and this is **time-dependent radial distribution function** for O-atoms

-if it is between two different atom types, O- and H-atoms, then we have a position-position cross-correlation function, also called a **time-dependent radial distribution function** or pair distribution function (the language is sloppy here)

When you include time, you are examining the dynamical properties of the system

So, what does a time-correlation function tell us? Qualitatively, a TCF describes how long a given property of a system persists until it is averaged out by microscopic motions and interactions with its surroundings. It describes how and when a statistical relationship has vanished.

Time-dependent RDFs

- $G(r, t)$ characterizes the diffusion of pairs in time, can provide insight into reactivity also
- Ex of a position-position cross-correlation function for studying reactivity

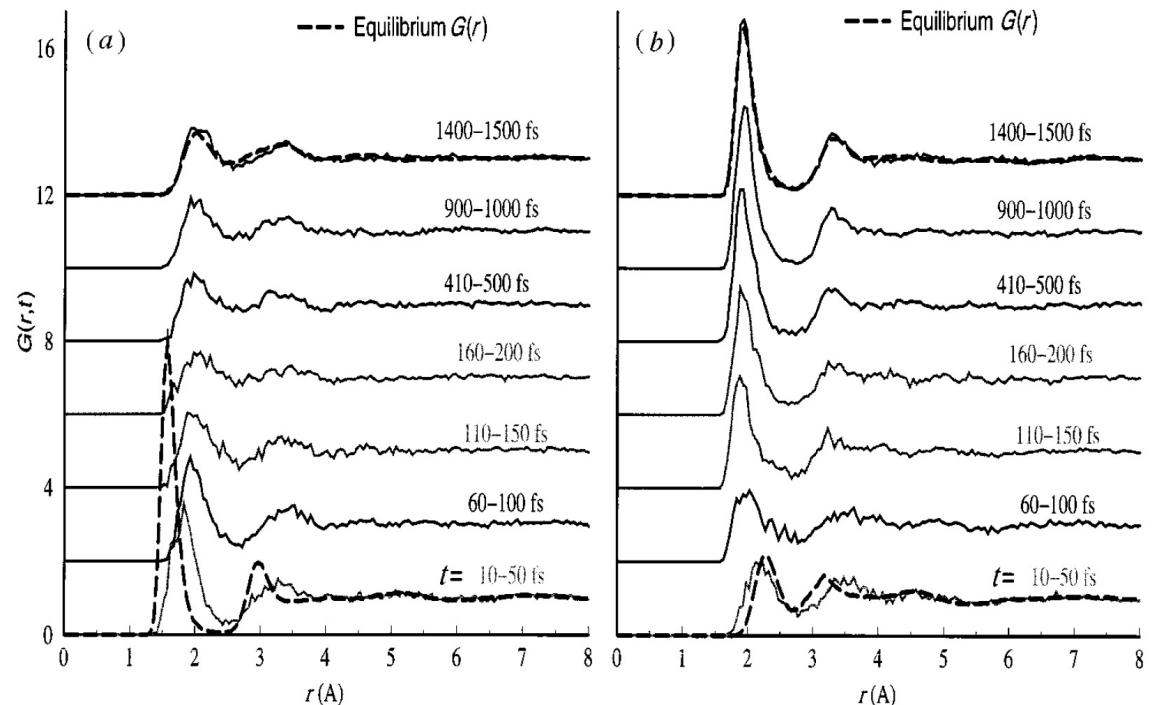


Figure 2. Time-dependent radial distribution function ($G(r, t)$) for the charge shifting of $\text{ClO}^- \rightarrow \text{Cl}^- \text{O}$ in H_2O . (a) $G(r, t)$ for oxygen of $[\text{ClO}]^-$ and hydrogen of H_2O . (b) $G(r, t)$ for chlorine of $[\text{ClO}]^-$ and hydrogen of H_2O . Dashed lines at the top and bottom of each of the sets show the equilibrium radial distribution functions of ClO^- and $\text{Cl}^- \text{O}$ respectively.

Time dependent correlation functions (1 variable)



Contributed by [Andrei Tokmakoff](#)
Professor (Chemistry) at [University of Chicago](#)

Returning to the microscopic fluctuations of a molecular variable A , there seems to be little information in observing the trajectory for a variable characterizing the time-dependent behavior of an individual molecule. However, this dynamics is not entirely random, since they are a consequence of time-dependent interactions with the environment. We can provide a statistical description of the characteristic time scales and amplitudes to these changes by comparing the value of A at time t with the value of A at time t' later.

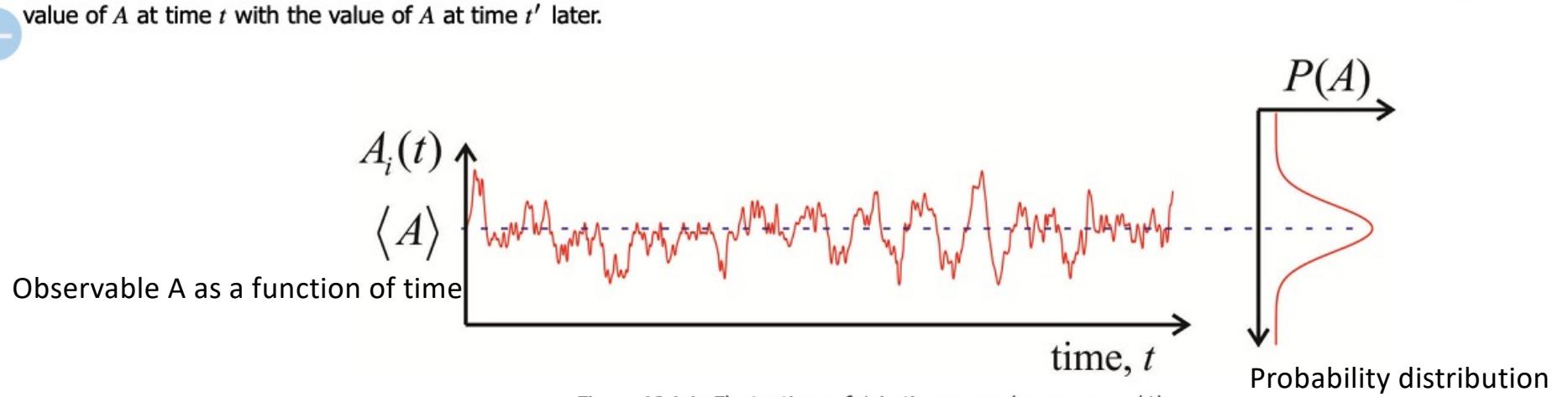


Figure 10.1.1: Fluctuations of A in time around an average $\langle A \rangle$.

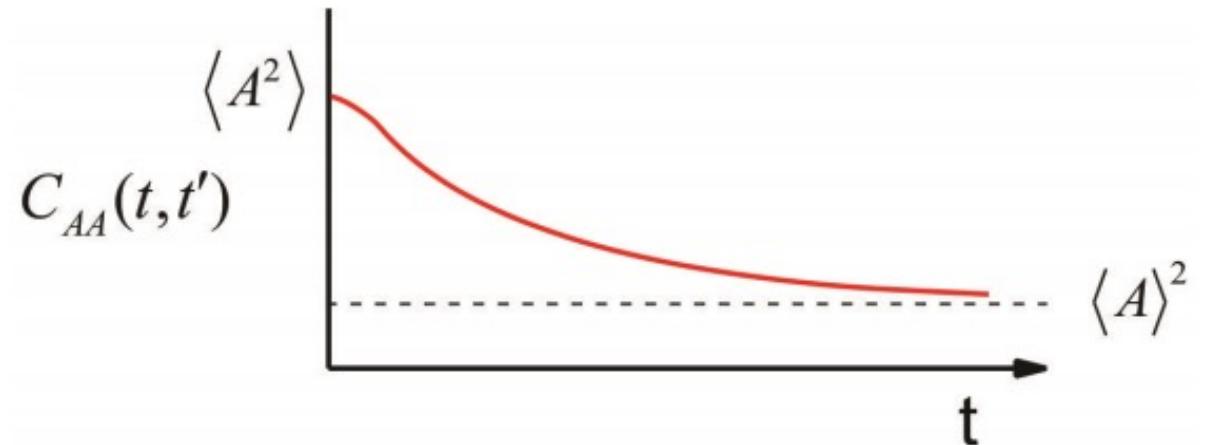
We define a time-correlation function (TCF) as a time-dependent quantity, $A(t)$, multiplied by that quantity at some later time, $A(t')$, and averaged over an equilibrium ensemble:

$$C_{AA}(t, t') \equiv \langle A(t)A(t') \rangle_{eq} \quad (10.1.1)$$

The classical form of the correlation function is evaluated as

$$C_{AA}(t, t') = \int d\mathbf{p} \int d\mathbf{q} A(\mathbf{p}, \mathbf{q}; t) A(\mathbf{p}, \mathbf{q}; t') \rho_{eq}(\mathbf{p}, \mathbf{q}) \quad (10.1.2)$$

A typical correlation function for random fluctuations at thermal equilibrium in the variable A might look like

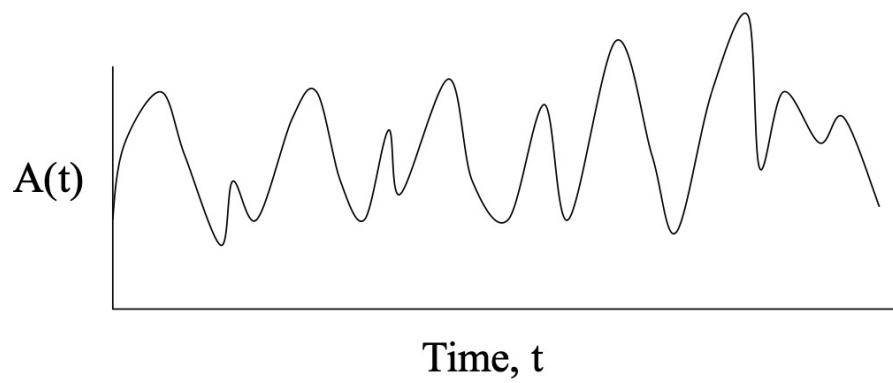


We define a time-correlation function (TCF) as a time-dependent quantity, $A(t)$, multiplied by that quantity at some later time, $A(t')$, and averaged over an equilibrium ensemble:

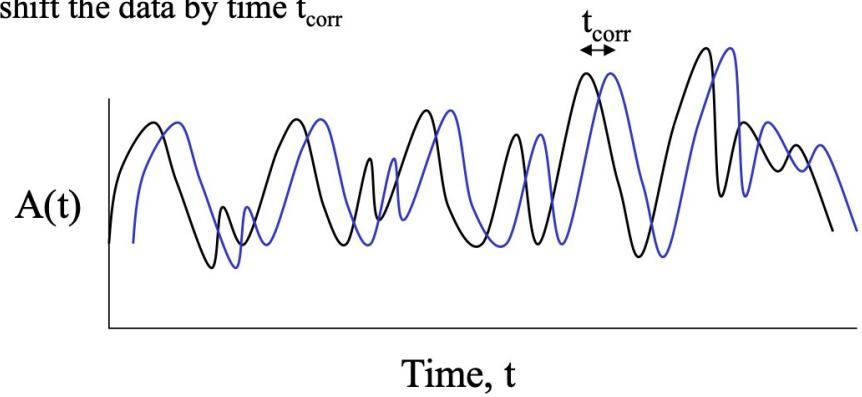
$$C_{AA}(t, t') \equiv \langle A(t)A(t') \rangle_{eq} \quad (10.1.1)$$

The classical form of the correlation function is evaluated as

$$C_{AA}(t, t') = \int d\mathbf{p} \int d\mathbf{q} A(\mathbf{p}, \mathbf{q}; t) A(\mathbf{p}, \mathbf{q}; t') \rho_{eq}(\mathbf{p}, \mathbf{q}) \quad (10.1.2)$$



Let us shift the data by time t_{corr}



The time correlation function is an overlap integral!!!!

C is not a function of time – (you are integrating over time) it is a function of the correlation time!!!

It is described by a number of properties:

- When evaluated at $t = t'$, we obtain the maximum amplitude, the mean square value of A , which is positive for an autocorrelation function and independent of time.

$$C_{AA}(t, t) = \langle A(t)A(t) \rangle = \langle A^2 \rangle \geq 0 \quad (10.1.9)$$

- For long time separations, as thermal fluctuations act to randomize the system, the values of A become uncorrelated

$$\lim_{t \rightarrow \infty} C_{AA}(t, t') = \langle A(t) \rangle \langle A(t') \rangle = \langle A \rangle^2 \quad (10.1.10)$$

- Since it is an equilibrium quantity, correlation functions are stationary. That means they do not depend on the absolute point of observation (t and t'), but rather the time interval between observations. A stationary random process means that the reference point can be shifted by an arbitrary value T

$$C_{AA}(t, t') = C_{AA}(t + T, t' + T) \quad (10.1.11)$$

So, choosing $T = -t'$ and defining the time interval $\tau \equiv t - t'$, we see that only τ matters

$$C_{AA}(t, t') = C_{AA}(t - t', 0) = C_{AA}(\tau) \quad (10.1.12)$$

Implicit in this statement is an understanding that we take the time-average value of A to be equal to the equilibrium ensemble average value of A , i.e., the system is **ergodic**. So, the correlation of fluctuations can be expressed as either a time-average over a trajectory of one molecule

$$\overline{A(t)A(0)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T d\tau A_i(t + \tau) A_i(\tau) \quad (10.1.13)$$

or an equilibrium ensemble average

$$\langle A(t)A(0) \rangle = \sum_n \frac{e^{-\beta E_n}}{Z} \langle n | A(t)A(0) | n \rangle \quad (10.1.14)$$

4. Classical correlation functions are real and even in time:

$$\langle A(t)A(t') \rangle = \langle A(t')A(t) \rangle \quad (10.1.15)$$

$$C_{AA}(\tau) = C_{AA}(-\tau) \quad \text{Time reversiblity} \quad (10.1.16)$$

5. When we observe fluctuations about an average (Figure 10.1.1), we often redefine the correlation function in terms of the deviation from average

$$\delta A \equiv A - \langle A \rangle \quad (10.1.17)$$

and

$$C_{\delta A \delta A}(t) = \langle \delta A(t)\delta A(0) \rangle = C_{AA}(t) - \langle A \rangle^2 \quad (10.1.18)$$

Now we see that the long time limit when correlation is lost

$$\lim_{t \rightarrow \infty} C_{\delta A \delta A}(t) = 0 \quad (10.1.19)$$

and the zero time value is just the variance

$$C_{\delta A \delta A}(0) = \langle \delta A^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2 \quad (10.1.20)$$

6. The characteristic time scale of a random process is the correlation time, τ_c . This characterizes the time scale for TCF to decay to zero. We can obtain τ_c from

$$\tau_c = \frac{1}{\langle \delta A^2 \rangle} \int_0^\infty dt \langle \delta A(t) \delta A(0) \rangle \quad (10.1.21)$$

which should be apparent if you have an exponential form

$$C(t) = C(0) \exp(-t/\tau_c). \quad (10.1.22)$$

Example 10.1.2: Dipole Moment Correlation Function

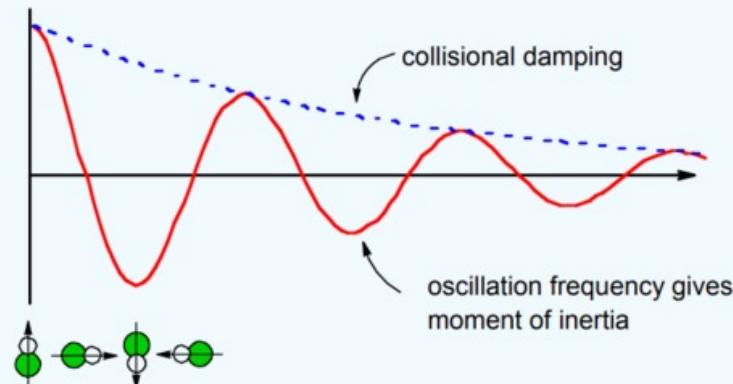
Now consider the correlation function for the dipole moment of a polar diatomic molecule in a dilute gas, $\bar{\mu}$. For a rigid rotating object, we can decompose the dipole into a magnitude and a direction unit vector:

$$\bar{\mu}_i = \mu_0 \cdot \hat{u}.$$

We know that $\langle \hat{u} \rangle = 0$ since all orientations of the gas phase molecules are equally probable. The correlation function is

$$\begin{aligned} C_{\mu\mu}(t) &= \langle \bar{\mu}(t)\bar{\mu}(0) \rangle \\ &= \langle \mu_0^2 \rangle \langle \hat{u}(t) \cdot \hat{u}(0) \rangle \end{aligned}$$

This correlation function projects the time-dependent orientation of the molecule onto the initial orientation. Free inertial rotational motion will lead to oscillations in the correlation function as the dipole spins. The oscillations in this correlation function can be related to the speed of rotation and thereby the molecule's moment of inertia (discussed below). Any apparent damping in this correlation function would reflect the thermal distribution of angular velocities. In practice a real gas would also have the collisional damping effects described in Example 10.1.1 superimposed on this relaxation process.



Time integrals of time correlation functions yield interesting properties!

Transport coefficients like the diffusion coefficient D , the viscosity coefficient η , and the thermal conductivity coefficient λ all can be expressed as time integrals of appropriate time correlation functions. This formulation is known as the Green-Kubo formulas, the results of linear response theory in statistical mechanics. To illustrate the main features of the Green-Kubo formalism we will treat explicitly the most straightforward case, that of diffusion.

We begin with the expression for the mean squared displacement of a typical particle in a fluid at equilibrium, $\langle \Delta^2 r(t) \rangle = \langle [\underline{R}(t) - \underline{R}(0)]^2 \rangle$. Using the simple relation

$$\underline{R}(t) - \underline{R}(0) = \int_0^t dt' \underline{v}(t') \quad (3.1)$$

we can write for the mean squared displacement function

$$\langle \Delta^2 r(t) \rangle = \int_0^t dt' \int_0^t dt'' \langle \underline{v}(t') \cdot \underline{v}(t'') \rangle \quad (3.2)$$

What appears in the integrand is the velocity autocorrelation function, involving the velocities of the particle at two different times, t' and t'' . For a fluid system in equilibrium one can invoke the property of time translation invariance. That is to say the correlation function depends only the time difference,

That is to say the correlation function depends only the time difference,

$$\langle \underline{v}(t') \cdot \underline{v}(t'') \rangle = \langle \underline{v}(t' - t'') \cdot \underline{v}(0) \rangle \quad (3.3)$$

This is essentially a statement that one can shift the time origin by any amount without affecting the correlation function. The double integral in (3.2) extends over a square region as shown in Fig. 3.1(a). To take advantage of (3.3) we transform the two integrals so that one of the variables of integration is the time difference $t' - t''$. Then the other time variable would not appear in (3.3), which is unspecified, and the integration over this variable can be carried out without knowing the velocity autocorrelation function.. We therefore introduce a change of variable, $\tau = t' - t''$ and $d\tau = -dt''$. Eq. (3.2) becomes

$$\langle \Delta^2 r(t) \rangle = \int_0^t dt' \int_{t'-t}^{t'} d\tau \alpha(\tau) \quad (3.4)$$

$$\langle \Delta^2 r(t) \rangle = \int_0^t dt' \int_{t'-t}^{t'} d\tau \alpha(\tau) \quad (3.4)$$

with $\alpha(t) = \langle \underline{v}(t) \cdot \underline{v}(0) \rangle$ being the velocity autocorrelation function. After the coordinate transformation, the region of integration now has the shape of a parallelogram, see Fig. 3.1(b). The order of integration in (3.4) is over τ first and then t' , or covering the integration region first with a horizontal strip, extending from $t'-t$ to t' , and then moving the strip from $t' = 0$ to $t' = t$.

$$\langle \Delta^2 r(t) \rangle = \int_0^t dt' \int_{t'-t}^{t'} d\tau \alpha(\tau) \quad (3.4)$$

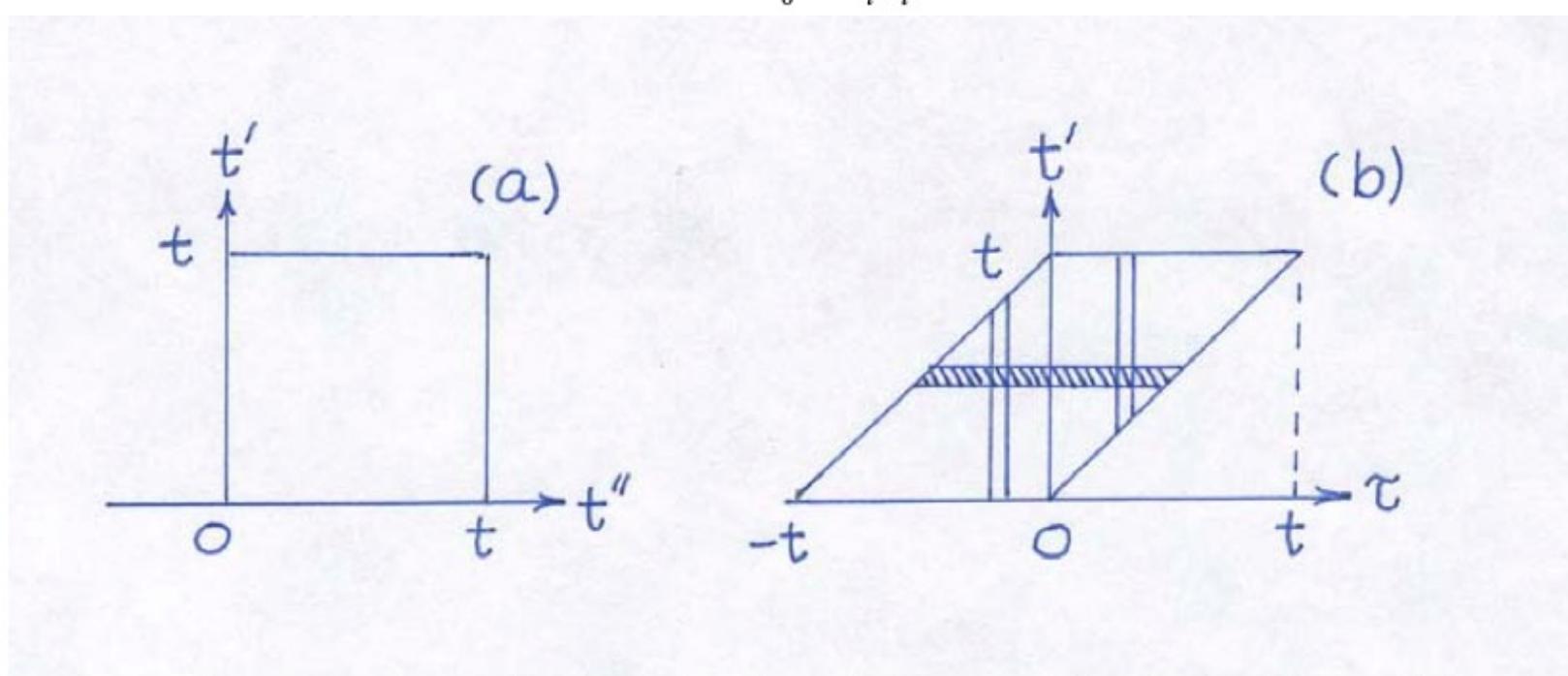


Fig. 3.1. The region of integration for the double integral in (3.2), (a), and for the double integral in (3.4), (b). Exchanging the order of integration, as (3.5), means that one integrates first over vertical strips as opposed to integrating over horizontal strips originally.

Suppose we interchange the order of integration, and use instead a vertical strip which extends from $t' = 0$ to $t' = t + \tau$ when $-t < \tau < 0$, and another strip extending from $t' = \tau$ to $t' = t$ when $0 < \tau < t$. Thus,

$$\langle \Delta^2 r(t) \rangle = \int_{-t}^0 d\tau \alpha(\tau) \int_0^{t+\tau} dt' + \int_0^t d\tau \alpha(\tau) \int_\tau^t dt'$$

$\downarrow \qquad \qquad \qquad \downarrow$
 $t + \tau \qquad \qquad t - \tau$

(3.5)

In the first integral we can change τ to $-\tau$, and make use of the fact that $\alpha(\tau) = \alpha(-\tau)$, a property of classical time correlation functions. Then the two integrals are the same, so that (3.5) becomes

$$\langle [\underline{R}(t) - \underline{R}(0)]^2 \rangle = 2 \int_0^t dt' (t - t') \langle \underline{v}(t') \cdot \underline{v}(0) \rangle$$

$$\begin{aligned}
<[\underline{R}(t) - \underline{R}(0)]^2> &= 2 \int_0^t dt' (t-t') <\underline{v}(t') \cdot \underline{v}(0)> \\
&= 6v_o^2 \int_0^t dt' (t-t') \psi(t')
\end{aligned} \tag{3.6}$$

where $\psi(t) = <\underline{v}(t) \cdot \underline{v}(0)> / <\underline{v}(0) \cdot \underline{v}(0)>$ is the normalized velocity autocorrelation function, $<\underline{v}(0) \cdot \underline{v}(0)> = 3v_o^2$, and $v_o^2 = k_B T / m$, v_o being the thermal speed.

Eq.(3.6) is the desired relation between the mean squared displacement and the velocity autocorrelation function. We know that the self-diffusion coefficient D is related to the mean squared displacement according to

$$D = \frac{1}{6} \left[\frac{<[\underline{R}(t) - \underline{R}(0)]^2>}{t} \right]_{t \rightarrow \infty} \tag{3.7}$$

Then combining this with (3.6) we obtain

$$D = v_o^2 \int_0^\infty dt \psi(t) \quad (3.8)$$

Eq.(3.8) is one of the Green-Kubo formulas for the transport coefficient of a fluid. It relates a transport coefficient to a time integral of a time correlation function, in this the relation is between the self-diffusion coefficient and the velocity autocorrelation function. Analogous relations exist for the shear viscosity η and the thermal conductivity λ , and the corresponding time correlation functions, the transverse component of the stress-stress correlation and the heat flux correlation function, respectively.

The significance of the Green-Kubo formulas is that one can use them to calculate the transport coefficients which describe the relaxation or dissipation response of a fluid (to an external perturbation) in terms of time correlation functions which describe the thermal fluctuations in the fluid at equilibrium. This connection between the dissipation in the fluid out of equilibrium and the fluctuations in the fluid at equilibrium is a general feature of linear response theory in statistical mechanics; it is expressed by the so-called fluctuation-dissipation theorem. One should keep in mind that this equivalence holds only in the linear response regime, meaning that the perturbation to the fluid is sufficiently small that one need only to consider the first term in the deviation from equilibrium, the linear response.

TCFs from a simulation trajectory

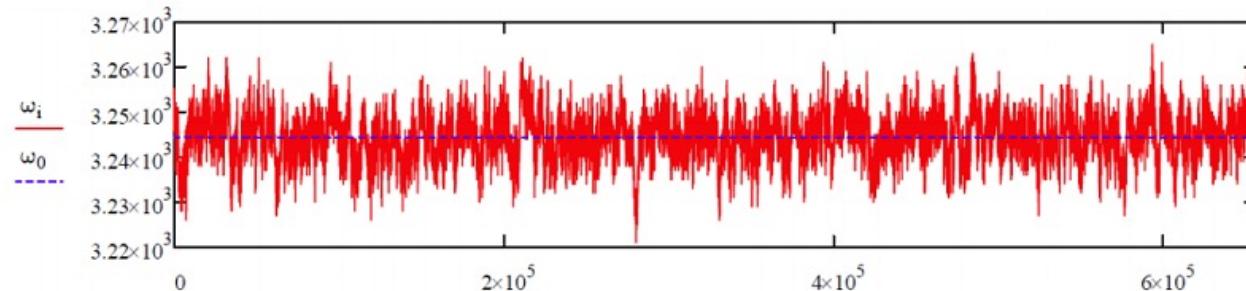
In practice classical correlation functions in molecular dynamics simulations or single molecule experiments are determined from a time-average over a long trajectory at discretely sampled data points. Let's evaluate C_{AA} for a discrete and finite trajectory in which we are given a series of N observations of the dynamical variable A at equally separated time points t_i . The separation between time points is $t_{i+1} - t_i = \Delta t$, and the length of the trajectory is $T = N\Delta t$. Then we have ,

$$C_{AA} = \frac{1}{T} \sum_{i,j=1}^N \Delta t A(t_i) A(t_j) = \frac{1}{N} \sum_{i,j=1}^N A_i A_j \quad (10.2.1)$$

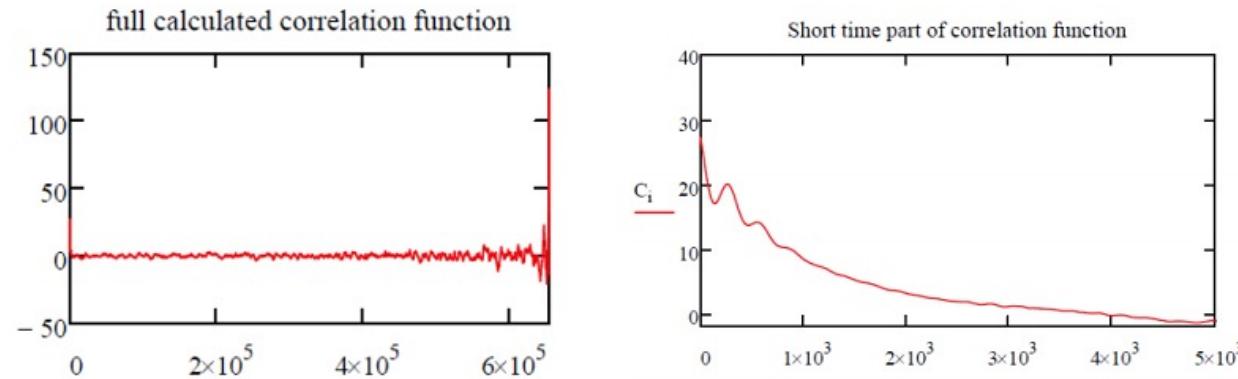
where $A_i = A(t_i)$. To make this more useful we want to express it as the time interval between points $\tau = t_j - t_i = (j - i)\Delta t$, and average over all possible pairwise products of A separated by τ . Defining a new count integer $n = j - i$, we can express the delay as $\tau = n\Delta t$. For a finite data set there are a different number of observations to average over at each time interval (n). We have the most pairwise products— N to be precise—when the time points are equal ($t_i=t_j$). We only have one data pair for the maximum delay $\tau = T$. Therefore, the number of pairwise products for a given delay τ is $N - n$. So we can write Equation 10.2.1 as

$$C_{AA}(\tau) = C(n) = \frac{1}{N-n} \sum_{i=1}^{N-n} A_{i+n} A_i \quad (10.2.2)$$

Note that this expression will only be calculated for positive values of n , for which $t_j \geq t_i$. As an example consider the following calculation for fluctuations in a vibrational frequency $\omega(t)$, which consists of 32000 consecutive frequencies in units of cm^{-1} for points separated by 10 femtoseconds, and has a mean value of $\omega_0 = 3244 cm^{-1}$. This trajectory illustrates that there are fast fluctuations on femtosecond time scales, but the behavior is seemingly random on 100 picosecond time scales



After determining the variation from the mean $\delta\omega(t_i) = \omega(t_i) - \omega_0$, the frequency correlation function is determined from Equation 10.2.2, with the substitution $\delta\omega(t_i) \rightarrow A_i$.



We can see that the correlation function reveals no frequency correlation on the time scale of 10^4 – 10^5 fs, however a decay of the correlation function is observed for short delays signifying the loss of memory in the fluctuating frequency on the 10^3 fs time scale. From Equation ???, we find that the correlation time is $\tau_C = 785 fs$.

Power-spectrum of atomic velocities/vibrational density of states from velocity autocorrelation function

The vibrational density of states is estimated from the Fourier transform of the autocorrelation function of atomic velocities:

$$VDOS(f) = \int \frac{\langle \vec{v}(0)\vec{v}(t) \rangle}{\langle \vec{v}(0)\vec{v}(0) \rangle} \exp(i2\pi ft) dt \quad (1)$$

IR and Raman from ab initio MD

MD-based vibrational spectrum:

$$P(\omega) = \int \langle \dot{\mathbf{A}}(\tau) \dot{\mathbf{A}}(t + \tau) \rangle e^{-i\omega t} dt \quad (1)$$

- If \mathbf{A} = all positions: total, unweighted spectrum
- If \mathbf{A} = molecular dipole moment μ : IR
- If \mathbf{A} = molecular polarizability α : Raman

$$\mu = -2e \sum_i \mathbf{r}_i + e \sum_j Z_j \mathbf{R}_j$$

with \mathbf{r}_i : Wannier centers; \mathbf{R}_j : nuclear positions

Wannier functions = molecular orbitals when basis set = plane waves

$$\text{IR: } A(\omega) \propto \int \langle \dot{\mu}(\tau) \dot{\mu}(t + \tau) \rangle_\tau e^{-i\omega t} dt$$

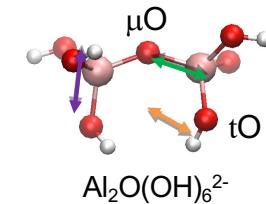
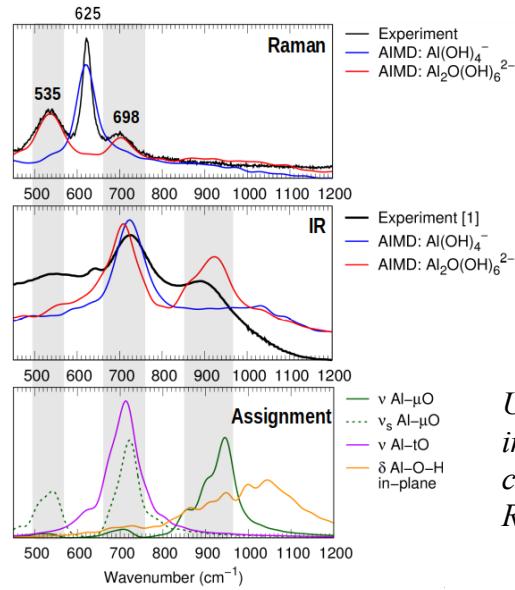
$$\text{Raman: } I_{||}(\omega) \propto \frac{(\omega_{in} - \omega)^4}{\omega} \frac{1}{1 - \exp\left(-\frac{\hbar\omega}{k_B T}\right)} \int \langle \dot{\alpha}_{xx}(\tau) \dot{\alpha}_{xx}(t + \tau) \rangle_\tau e^{-i\omega t} dt$$

- Orthogonal component obtained using α_{xy}
- α obtained from the induced dipole and the applied electric field via $\mu_{\text{ind}} = \alpha E$

Practically with CP2K + TRAVIS

1. Localization of the Wannier centers from the trajectory without E gives the dipoles and their time derivatives are used to calculate the IR via TRAVIS
2. Three other localizations are done in the presence of a electric field either along x, y or z, applied within CP2K. This gives the induced dipoles along the 3 dimensions, TRAVIS calculates the spatial average and the time derivative of this average is used to compute the Raman

Example



Using bond lengths or angles as A in (1) instead of dipoles or polarizabilities, we can help the assignment of the IR and Raman spectra

Time correlation function calculation in internal-GROMACS

gmx dipoles

Synopsis

```
gmx dipoles [-en [<.edr>]] [-f [<.xtc/.trr/...>]] [-s [<.tpr>]]
[-n [<.ndx>]] [-o [<.xvg>]] [-eps [<.xvg>]] [-a [<.xvg>]]
[-d [<.xvg>]] [-c [<.xvg>]] [-g [<.xvg>]]
[-adip [<.xvg>]] [-dip3d [<.xvg>]] [-cos [<.xvg>]]
[-cmap [<.xpm>]] [-slab [<.xvg>]] [-b <time>] [-e <time>]
[-dt <time>] [-nojw] [-xvg <enum>] [-mu <real>]
[-mumax <real>] [-epsilonRF <real>] [-skip <int>]
[-temp <real>] [-corr <enum>] [-[no]pairs] [-[no]quad]
[-ncos <int>] [-axis <string>] [-s1 <int>]
[-gkatom <int>] [-gkatom2 <int>] [-rcmax <real>]
[-[no]phi] [-nlevels <int>] [-ndegrees <int>]
[-acflen <int>] [-[no]normalize] [-P <enum>]
[-fitfn <enum>] [-beginfit <real>] [-endfit <real>]
```

Description

`gmx dipoles` computes the total dipole plus fluctuations of a simulation system. From this you can compute e.g. the dielectric constant for low-dielectric media. For molecules with a net charge, the net charge is subtracted at center of mass of the molecule.

The file `Mtot.xvg` contains the total dipole moment of a frame, the components as well as the norm of the vector. The file `aver.xvg` contains $\langle|\mu|^2\rangle$ and $|\langle\mu\rangle|^2$ during the simulation. The file `dipdist.xvg` contains the distribution of dipole moments during the simulation. The value of `-mumax` is used as the highest value in the distribution graph.

Furthermore, the dipole autocorrelation function will be computed when option `-corr` is used. The output file name is given with the `-c` option. The correlation functions can be averaged over all molecules (`mol`), plotted per molecule separately (`molsep`) or it can be computed over the total dipole moment of the simulation box (`total`).

Option `-g` produces a plot of the distance dependent Kirkwood G-factor, as well as the average cosine of the angle between the dipoles as a function of the distance. The plot also includes gOO and hOO according to Nymand & Linse, J. Chem. Phys. 112 (2000) pp 6386-6395. In the same plot, we also include the energy per scale computed by taking the inner product of the dipoles divided by the distance to the third power.

EXAMPLES

```
gmx dipoles -corr mol -P 1 -o dip_sqr -mu 2.273 -mumax 5.0
```

This will calculate the autocorrelation function of the molecular dipoles using a first order Legendre polynomial of the angle of the dipole vector and itself a time t later. For this calculation 1001 frames will be used. Further, the dielectric constant will be calculated using an `-epsilonRF` of infinity (default), temperature of 300 K (default) and an average dipole moment of 2.273 (SPC). For the distribution function a maximum of 5.0 will be used.

Time correlation function calculation in internal-GROMACS

gmx hbond

Synopsis

- The lifetime of the H-bonds is calculated from the average over all autocorrelation functions of the existence functions (either 0 or 1) of all H-bonds:

$$C(\tau) = \langle s_i(t) s_i(t + \tau) \rangle$$

- with $s_i(t) = \{0, 1\}$ for H-bond i at time t . The integral of $C(\tau)$ gives a rough estimate of the average H-bond lifetime τ_{HB} :

$$\tau_{HB} = \int_0^{\infty} C(\tau) d\tau$$

Description

`gmx hbond` computes and analyzes hydrogen bonds. Hydrogen bonds are determined based on cutoffs for the angle Hydrogen - Donor - Acceptor (zero is extended) and the distance Donor - Acceptor (or Hydrogen - Acceptor using `-node`). OH and NH groups are regarded as donors, O is an acceptor always, N is an acceptor by default, but this can be switched using `-nitacc`. Dummy hydrogen atoms are assumed to be connected to the first preceding non-hydrogen atom.

You need to specify two groups for analysis, which must be either identical or non-overlapping. All hydrogen bonds between the two groups are analyzed.

If you set `-shell`, you will be asked for an additional index group which should contain exactly one atom. In this case, only hydrogen bonds between atoms within the shell distance from the one atom are considered.

With option `-ac`, rate constants for hydrogen bonding can be derived with the model of Luzar and Chandler (Nature 394, 1996; J. Chem. Phys. 113:23, 2000) or that of Markovitz and Agmon (J. Chem. Phys 129, 2008). If contact kinetics are analyzed by using the `-contact` option, then $n(t)$ can be defined as either all pairs that are not within contact distance r at time t (corresponding to leaving the `-r2` option at the default value 0) or all pairs that are within distance $r2$ (corresponding to setting a second cut-off value with option `-r2`). See mentioned literature for more details and definitions.

Output:

- `-num`: number of hydrogen bonds as a function of time.
- `-ac`: average over all autocorrelations of the existence functions (either 0 or 1) of all hydrogen bonds.
- `-dist`: distance distribution of all hydrogen bonds.
- `-ang`: angle distribution of all hydrogen bonds.
- `-hx`: the number of $n-n+i$ hydrogen bonds as a function of time where n and $n+i$ stand for residue numbers and i ranges from 0 to 6. This includes the $n-n+3$, $n-n+4$ and $n-n+5$ hydrogen bonds associated with helices in proteins.
- `-hbn`: all selected groups, donors, hydrogens and acceptors for selected groups, all hydrogen bonded atoms from all groups and all solvent atoms involved in insertion.
- `-hbm`: existence matrix for all hydrogen bonds over all frames, this also contains information on solvent insertion into hydrogen bonds. Ordering is identical to that in `-hbn` index file.
- `-dan`: write out the number of donors and acceptors analyzed for each timeframe. This is especially useful when using `-shell`.
- `-nhbdist`: compute the number of HBonds per hydrogen in order to compare results to Raman Spectroscopy.

Note: options `-ac`, `-life`, `-hbn` and `-hbm` require an amount of memory proportional to the total numbers of donors times the total number of acceptors in the selected group(s).

Many important caveats associated with HB lifetimes

are tested and discussed below.

3.2.1. Continuous Hydrogen Bonds. If one decides that a HB must exist continuously (the uninterrupted HB²⁴), one can compute a distribution of lifetimes $P(t)$ by making a histogram of the number of HBs that existed continuously from time 0 to t . The probability of breaking a hydrogen bond is constant and independent of the history of the bond, and hence the tail of the $P(t)$ curve should fall off exponentially;^{19,25} this is indeed what we find. The lifetime distribution can be converted into an autocorrelation function by

$$C(t) = 1 - \int_0^t P(\tau) d\tau \quad (1)$$

An overall HB lifetime τ_1 can now be associated with the integral of eq 1

$$\tau_1 = \int_0^\infty C(t) dt \quad (2)$$

Results for the different systems and different frequency of saving Δt coordinates are given in Table 1.

J. Phys. Chem. B **2006**, *110*, 4393–4398

Thermodynamics of Hydrogen Bonding in Hydrophilic and Hydrophobic Media

David van der Spoel,^{*†} Paul J. van Maaren,[†] Per Larsson,[‡] and Nicușor Timmeau[†]

Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-751 24 Uppsala, Sweden, and Stockholm Bioinformatics Center, Stockholm University, SE-10691 Stockholm, Sweden

Received: December 12, 2005

3.2.2. Interrupted Hydrogen Bonds. If we allow HBs to break and re-form, we can analyze HB lifetimes by defining a binary function $h(t)$, which is 1 when a H bond is present and 0 otherwise.²⁵ The autocorrelation function $c_h(t)$ of $h(t)$ was computed and averaged over similar types of HBs. In the terminology of Luzar,²⁴ this is the “intermittent HB correlation function”, which by design is insensitive to the saving frequency Δt , except for the very shortest times ($0 \leq t \leq \Delta t$). Since the simulation systems are finite, the $c_h(t)$ do not go to zero and the $c_h(t)$ have to be scaled to zero at long times.¹⁹ In the remainder of this paper we will use this definition for the analysis of kinetics and thermodynamics of hydrogen bonding.

Hydrogen-bond kinetics in liquid water

Alenka Luzar & David Chandler

Department of Chemistry, University of California, Berkeley, California 94720-1460, USA

Towards a unified description of the hydrogen bond network of liquid water dynamics based approach

Cite as: J. Chem. Phys. **141**, 214107 (2014); <https://doi.org/10.1063/1.4902538>

Submitted: 29 July 2014 . Accepted: 13 November 2014 . Published Online: 03 December

Abdullah Ozkanlar, Tiecheng Zhou, and Aurora E. Clark

3.4. Thermodynamics of Hydrogen Bond Breaking. If we assume that the process of HB breakage can be described as an Eyring process, we can relate τ_{HB} (eq 2) to the Gibbs energy of activation ΔG^\ddagger

Single mechanism of HB breakage

$$\tau_{\text{HB}} = \frac{h}{k_B T} \exp\left(\frac{\Delta G^\ddagger}{k_B T}\right) \quad (6)$$

where h is Planck's constant, k_B is Boltzmann's constant, and T is the temperature. We can then derive the activation enthalpy of HB breaking ΔH^\ddagger from the Van't Hoff equation

$$\Delta H^\ddagger = \frac{\partial(\Delta G^\ddagger/T)}{\partial(1/T)} \quad (7)$$

and hence the entropy of activation $T\Delta S^\ddagger = \Delta H^\ddagger - \Delta G^\ddagger$. The thermodynamical parameters are plotted in Figure 3 (data in

Time correlation function calculation in internal-GROMACS

gmx rotacf

Synopsis

```
gmx rotacf [-f [<.xtc/.trr/...>]] [-s [<.tpr>]] [-n [<.ndx>]]  
[-o [<.xvg>]] [-b <time>] [-e <time>] [-dt <time>]  
[-[no]w] [-xvg <enum>] [-[no]id] [-[no]aver]  
[-acflen <int>] [-[no]normalize] [-P <enum>]  
[-fitfn <enum>] [-beginfit <real>] [-endfit <real>]
```

Description

`gmx rotacf` calculates the rotational correlation function for molecules. Atom triplets (i,j,k) must be given in the index file, defining two vectors ij and jk. The rotational ACF is calculated as the autocorrelation function of the vector $n = ij \times jk$, i.e. the cross product of the two vectors. Since three atoms span a plane, the order of the three atoms does not matter. Optionally, by invoking the `-d` switch, you can calculate the rotational correlation function for linear molecules by specifying atom pairs (i,j) in the index file.

EXAMPLES

```
gmx rotacf -P 1 -nparm 2 -fft -n index -o rotacf-x-P1 -fa expfit-x-P1 -beginfit 2.5 -endfit 20.0
```

This will calculate the rotational correlation function using a first order Legendre polynomial of the angle of a vector defined by the index file. The correlation function will be fitted from 2.5 ps until 20.0 ps to a two-parameter exponential.

- The time correlation functions must work with molecules for each frame with same index
- Usually calculating for all molecules in the total simulation box
- For molecules within specific regions, the above TCF can be combined with the dynamic selections from GROMACS internal program: ***gmx select***

Time correlation function calculation in internal-GROMACS

gmx select

Synopsis

```
gmx select [-f [<.xtc/.trr/...>]] [-s [<.top/.pro/...>]] [-n [<.ndx>]]
[-os [<.xvg>]] [-oc [<.xvg>]] [-oi [<.dat>]]
[-on [<.ndx>]] [-om [<.xvg>]] [-of [<.xvg>]]
[-od [<.dat>]] [-ot [<.dat>]] [-b <time>] [-e <time>]
[-dt <time>] [-tu <time>] [-fgroup <selection>]
[-xvg <enum>] [-[no]mpbc] [-[no]pbc] [-sf <file>]
[-selpos <enum>] [-seltype <enum>] [-select <selection>]
[-[no]norm] [-[no]cfnorm] [-resnr <enum>]
[-pdbsatoms <enum>] [-[no]cumlt]
```

Description

gmx select writes out basic data about dynamic selections. It can be used for some simple analyses, or the output can be combined with output from other programs and/or external analysis programs to calculate more complex things. For detailed help on the selection syntax, please use gmx help selections.

Any combination of the output options is possible, but note that -os only operates on the first selection. Also note that if you provide no output options, no output is produced.

With -os, calculates the number of positions in each selection for each frame. With -norm, the output is between 0 and 1 and describes the fraction from the maximum number of positions (e.g., for selection 'resname RA and x < 5' the maximum number of positions is the number of atoms in RA residues). With -cfnorm, the output is divided by the fraction covered by the selection. -norm and -cfnorm can be specified independently of one another.

With -oc, the fraction covered by each selection is written out as a function of time.

With -on, the selected atoms/residues/molecules are written out as a function of time. In the output, the first column contains the frame time, the second contains the number of positions, followed by the atom/residue/molecule numbers. If more than one selection is specified, the size of the second group immediately follows the last number of the first group and so on.

With -om, the selected atoms are written as a index file compatible with make_ndx and the analyzing tools. Each selection is written as a selection group and for dynamic selections a group is written for each frame.

For residue numbers, the output of -oc can be controlled with -resnn: number (default) prints the residue numbers as they appear in the input file, while index prints unique numbers assigned to the residues in the order they appear in the input file, starting with 1. The former is more intuitive, but if the input contains multiple residues with the same number, the output can be less useful.

With -om, a mask is printed for the first selection as a function of time. Each line in the output corresponds to one frame, and contains either 0/1 for each atom/residue/molecule possibly selected. 1 stands for the atom/residue/molecule being selected for the current frame, 0 for not selected.

With -of, the occupancy fraction of each position (i.e., the fraction of frames where the position is selected) is printed.

With -opdb, a PDB file is written out where the occupancy column is filled with the occupancy fraction of each atom in the selection. The coordinates in the PDB file will be those from the input topology. -pdbsatoms can be used to control which atoms appear in the output PDB file with all all atoms are present, with maxsel all atoms possibly selected by the selection are present, and with selected only atoms that are selected at least in one frame are present.

With -ot, a histogram is produced that shows the number of selected positions as a function of the time the position was continuously selected. -cumlt can be used to control whether subintervals of longer intervals are included in the histogram.

-os, -of, and -ot only make sense with dynamic selections

Examples: select water molecules within the 1st solvation shell of ions or some distances

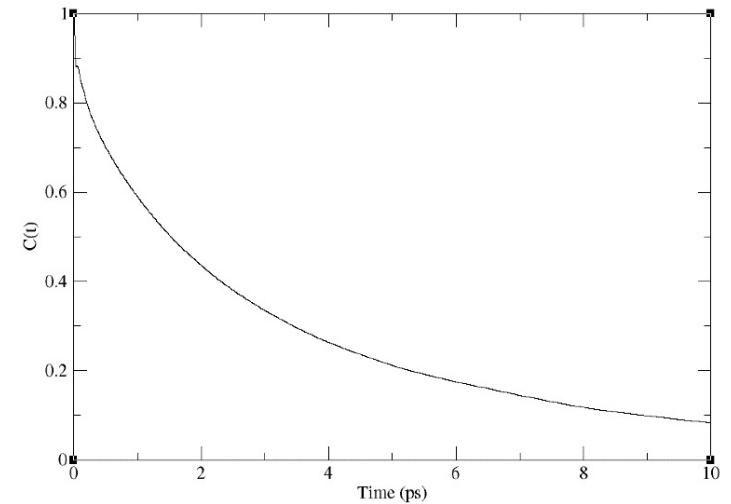
```
gmx select -f -n -os -oc -oi -om -on output.ndx -selrpos mol_com
> "1st shell" resname SOL and name OW and within 0.3 of resname Li
> "bulk sol" resname SOL and name OW and z > 10.0 and z < 12.0
```

Examples: rot-acf

(same as gmx hbond lifetime within selections)

- time distribution of hbond number, hbond lifetime, residence time
- same calculation procedure of the dipole relaxation time in my dielectric paper
- not sure if this can be called correlations since the index/ids within the dynamic selections cannot be guaranteed

Rotational Correlation Function



Correlations in Space only

All of the above considerations can be applied to correlations in space instead of time $G(\mathbf{r}) = \langle A(\mathbf{r}_0)A(\mathbf{r}_0+\mathbf{r}) \rangle$.

Molecular Distribution Functions

Probabilities. Nth order MDF

$$Z_N = \int_V e^{-\beta U(x_1, y_1, z_1, \dots, x_N, y_N, z_N)} dx_1, \dots, dz_N$$

Specific probability density of finding...

$$P^N(r_1, \dots, r_N) dr_1 \dots dr_N = \frac{e^{-\beta U_N} dr_1 \dots dr_N}{Z_N}$$

$$P^{(n)}(r_1, \dots, r_n) = \frac{\int \dots \int e^{-\beta U_N} dr_{n+1} \dots dr_N}{Z_N} \quad \text{"any"}$$

Generic probability density of finding...

$$\rho^{(n)}(r_1, \dots, r_n) = \frac{N!}{(N-n)!} P^{(n)}(r_1, \dots, r_n)$$

Nth order MDF

n will define the order of the distribution function

- Singlet Distribution Function
- Pair Distribution Function
- .
- .
- .

Molecular Distribution Functions

Singlet Distribution Function

$$\int_V \rho^{(1)}(r_1) dr_1$$

"any" $\int_V \rho^{(1)}(r_1) dr_1 = N \int_V P^{(1)}(r_1) dr_1 = N$

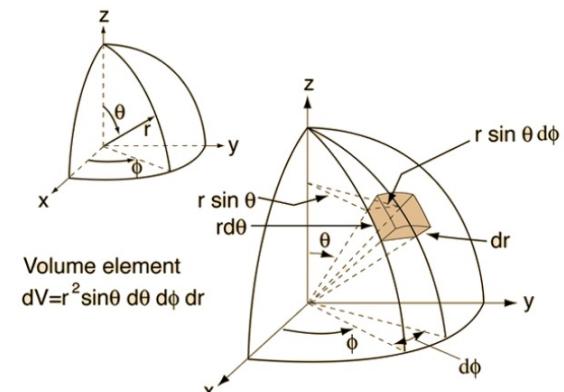
Individual probability of each particle

In homogeneous fluids we expect
 $\rho^{(1)} r_1$ to be constant at any point r_1

$$\rho^{(1)}(r_1) = \text{constant}$$

$$\text{const} * \int_V dr_1 = N$$

$$\boxed{\rho^{(1)}(r_1) = N/V = \rho}$$



Molecular Distribution Functions

Pair Distribution Function. Pair correlation function

Pair Distribution Function $\rho^{(2)}(r_1, r_2)dr_1 dr_2 = \frac{N!}{(N-n)!} P^{(n)}(r_1, r_2)dr_1 dr_2$

$$\rho^{(2)}(r_1, r_2) = \rho^{(1)}(r_1) d\rho^{(2)}(r_2) \quad \text{If independent probabilities}$$

$$\rho^{(n)}(r_1, \dots, r_n) = \rho^n g^{(n)}(r_1, \dots, r_n)$$

Nth order Correlation Function

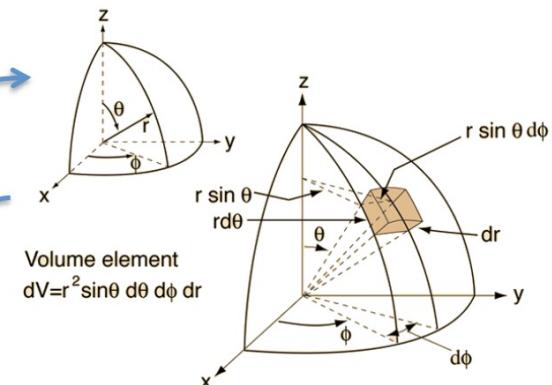
$$g^{(n)}(r_1, \dots, r_n) = \frac{V^n}{N^n} \frac{N!}{(N-n)!} \frac{\int \dots \int e^{-\beta U_N} dr_{n+1} \dots dr_N}{Z_N}$$

Pair Correlation Function

$$g^{(2)}(r_1, r_2) = \frac{V^2}{N^2} N(N-1) \frac{4\pi \int \dots \int e^{-\beta U_N} r^2 dr}{Z_N}$$

Pair Distribution Function

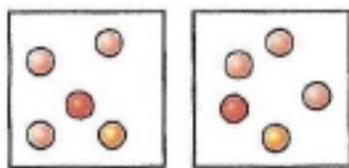
$$\rho^{(2)}(r_1, r_2) = \rho^2 g^{(2)}(r_1, r_2) 4\pi R^2 dR$$



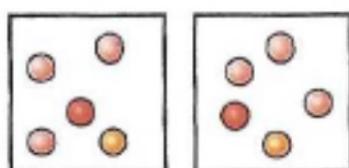
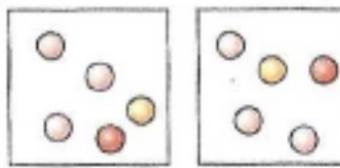
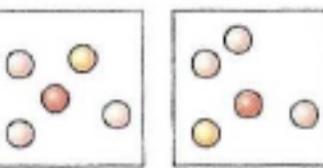
Molecular Distribution Functions

Pair Distribution Function

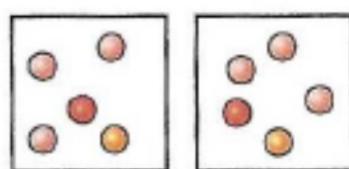
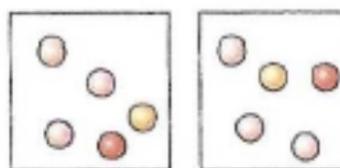
$\bullet = 1$ $\circ = 2$



(1) Integrate over positions of molecules 3–5.



(2) Integrate over direction of 2 from 1.



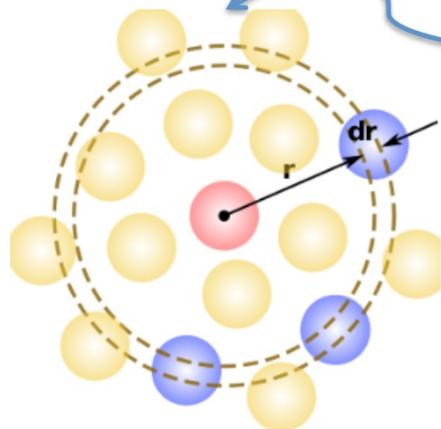
(3) Integrate over location of 1.

Then we can change 2 for any other N-1 molecules and we will have the average number of molecules at a distance R_a and R_b from any randomly selected molecule (this time we called molecule 1)

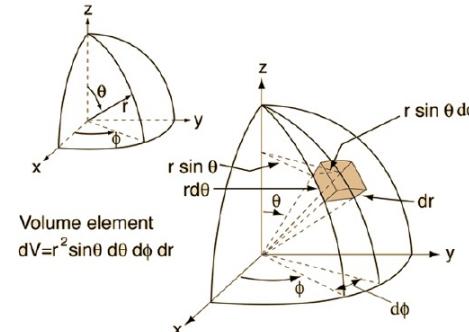
Pair Correlation Function

Fix molecule 1 and calculate $g(R)$ where
 $R = |r_1 - r_2|$

$$\rho g(r_1, r_2) dV = \rho g(R) 4\pi R^2 dR$$



Radial distribution function of the fluid



Probability of observing a second molecule in dV given that there is a molecule at the origin r .

Is normalized to $N-1$

- Conveys direct information on the mode of packing of the molecules (often referred to as the "structure") in the liquid
- Serves as a bridge between thermodynamic properties on the one hand, and molecular properties on the other

Thermodynamic quantities from $g(r)$

Structural characterization: Probability densities $\rho_N^{(1)}$ and $\rho_N^{(2)}$

$$\int \rho_N^{(1)}(\mathbf{r}^1) d\mathbf{r}^1 = N \quad \int \int \rho_N^{(2)}(\mathbf{r}^2) d\mathbf{r}^2 = N(N-1)$$

Relationship btw
probability density and
correlation function, and
probability density and
partition function

i.e. one can find N particles and $N(N-1)$ pairs of particles in the total volume, respectively.

$$\rho_N^{(1)}(\mathbf{r}) = \langle \sum_{i=1}^N \delta(\mathbf{r} - \mathbf{r}_i) \rangle. \quad \rho_N^{(2)}(\mathbf{r}, \mathbf{r}') = \langle \sum_{i=1}^N \sum_{j=1, j \neq i}^N \delta(\mathbf{r} - \mathbf{r}_i) \delta(\mathbf{r}' - \mathbf{r}_j) \rangle.$$

Note that these are $n=1$ and $n=2$ cases of a general correlation function $g^{(n)}$ given by :

-we know that partition
function is related to
Energy and if so, then
probability density and
correlation function are
related to energy

$$\begin{aligned} g^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n) &= \frac{1}{\rho^n} \rho^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n) \\ &= \frac{V^n N!}{Z_N N^n (N-n)!} \int d\mathbf{r}_{n+1} \cdots d\mathbf{r}_N e^{-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)} \end{aligned}$$

Z_N the partition function, V the volume, N the nb of particles

So that one can write for e.g. n=2:

$$g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{V^2(N-1)}{NZ_N} \int d\mathbf{r}_3 \cdots d\mathbf{r}_N e^{-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)}$$

which defines the **radial distribution (rdf)** $g_N^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ function, also given by

$$g_N^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \rho_N^{(2)}(\mathbf{r}_1, \mathbf{r}_2) / (\rho_N^{(1)}(\mathbf{r}_1)\rho_N^{(1)}(\mathbf{r}_2)).$$

For an homogeneous isotropic system, one has $\rho_N^{(1)}=\rho$. Dependence of the rdf only on relative distance between particles:

$$g_N^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = g(|\mathbf{r}_1 - \mathbf{r}_2|) = \frac{\rho_N^{(2)}(|\mathbf{r}_1 - \mathbf{r}_2|)}{\rho^2}$$

So that $pg(r)$ is the conditional probability to find another particle at a distance r away from the origin.

First, one needs to remember that the energy is related to the partition function $Q(N, V, \beta)$ as:

$$E = -\frac{\partial}{\partial \beta} \ln Q(N, V, \beta)$$

where :

$$Q(V, \beta, N) = \frac{Z_N}{\lambda^{3N} N!} = \frac{1}{\lambda^{3N} N!} \int_0^V dr^N e^{-\beta U(r^N)}$$

λ comes from the integration of the momentum p in the phase space given by:

$$\lambda = \left[\frac{\beta h^2}{2\pi m} \right]^{1/2}$$

so that:

$$\ln Q(N, V, \beta) = \ln Z_N - 3N \ln \lambda(\beta) - \ln N!$$

$$E = \frac{3N}{\lambda} \frac{\partial \lambda}{\partial \beta} - \frac{1}{Z_N} \frac{\partial Z_N}{\partial \beta}$$

$$E = \frac{3N}{\lambda} \frac{\partial \lambda}{\partial \beta} - \frac{1}{Z_N} \frac{\partial Z_N}{\partial \beta} \quad \text{and} \quad \frac{\partial \lambda}{\partial \beta} = \frac{1}{2\beta} \lambda$$

$$\begin{aligned} E &= \frac{3}{2} N k T + \frac{1}{Z_N} \int d\mathbf{r}_1 \cdots d\mathbf{r}_N U(\mathbf{r}_1, \dots, \mathbf{r}_N) e^{\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)} \\ &= \frac{3}{2} N k T + \langle U \rangle \end{aligned}$$

In order to compute the average energy, one needs to compute the average potential U . Assume the case of a pairwise potential, so that:

$$U(\mathbf{r}_1, \dots, \mathbf{r}_N) = \frac{1}{2} \sum_{i,j, i \neq j} u(|\mathbf{r}_i - \mathbf{r}_j|) \equiv U_{\text{pair}}(\mathbf{r}_1, \dots, \mathbf{r}_N)$$

i.e. U is a sum of terms depending only between 2 particles, and thus contains $N(N-1)$ terms. Then, we can write:

$$\begin{aligned} \langle U \rangle &= \frac{1}{2Z_N} \sum_{i,j, i \neq j} \int d\mathbf{r}_1 \cdots d\mathbf{r}_N u(|\mathbf{r}_i - \mathbf{r}_j|) e^{-\beta U_{\text{pair}}(\mathbf{r}_1, \dots, \mathbf{r}_N)} \\ &= \frac{N(N-1)}{2Z_N} \int d\mathbf{r}_1 \cdots d\mathbf{r}_N u(|\mathbf{r}_1 - \mathbf{r}_2|) e^{-\beta U_{\text{pair}}(\mathbf{r}_1, \dots, \mathbf{r}_N)} \end{aligned}$$

(i.e. all terms in the first line are the exact same integrals, just with different labels)

$$\langle U \rangle = \frac{1}{2} \int d\mathbf{r}_1 d\mathbf{r}_2 u(|\mathbf{r}_1 - \mathbf{r}_2|) \left[\frac{N(N-1)}{Z_N} \int d\mathbf{r}_3 \cdots d\mathbf{r}_N e^{-\beta U_{\text{pair}}(\mathbf{r}_1, \dots, \mathbf{r}_N)} \right]$$

which contains a two-body probability distribution function $\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ so that $\langle U \rangle$ can be rewritten as :

$$\begin{aligned} \langle U \rangle &= \frac{1}{2} \int d\mathbf{r}_1 d\mathbf{r}_2 u(|\mathbf{r}_1 - \mathbf{r}_2|) \left[\frac{N(N-1)}{Z_N} \int d\mathbf{r}_3 \cdots d\mathbf{r}_N e^{-\beta U_{\text{pair}}(\mathbf{r}_1, \dots, \mathbf{r}_N)} \right] \\ &= \frac{1}{2} \int d\mathbf{r}_1 d\mathbf{r}_2 u(|\mathbf{r}_1 - \mathbf{r}_2|) \rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \\ &= \frac{N^2}{2V^2} \int d\mathbf{r}_1 d\mathbf{r}_2 u(|\mathbf{r}_1 - \mathbf{r}_2|) g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \end{aligned}$$

where we remind (slide 2) that :

$$g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{V^2(N-1)}{NZ_N} \int d\mathbf{r}_3 \cdots d\mathbf{r}_N e^{-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)}$$

One thus arrives to the result:

$$\langle U \rangle = \frac{N^2}{2V} \int_0^\infty dr 4\pi r^2 u(r) g(r)$$

$$E = \frac{3}{2} N k T + \frac{N}{2} 4\pi \rho \int_0^\infty dr r^2 u(r) g(r)$$

Thermodynamic quantities from the computed g(r): Pressure

We use the Maxwell relation:

$$P = \frac{1}{\beta} \frac{\partial \ln Q}{\partial V} = \frac{1}{\beta Z_N} \frac{\partial Z_N}{\partial \beta}$$

The volume dependence can be made explicit by changing the variables: $s_i = V^{-1/3} r_i$

so that we have from the partition function Z_N the desired derivative from V:

$$\begin{aligned} Z_N &= V^N \int d\mathbf{s}_1 \cdots d\mathbf{s}_N e^{-\beta U(V^{1/3}\mathbf{s}_1, \dots, V^{1/3}\mathbf{s}_N)} \\ \frac{\partial Z_N}{\partial V} &= \frac{N}{V} Z_N - \beta V^N \int d\mathbf{s}_1 \cdots d\mathbf{s}_N \frac{1}{3V} \sum_{i=1}^N \mathbf{r}_i \cdot \frac{\partial U}{\partial \mathbf{r}_i} e^{-\beta U(V^{1/3}\mathbf{s}_1, \dots, V^{1/3}\mathbf{s}_N)} \\ &= \frac{N}{V} Z_N + \beta \int d\mathbf{r}_1 \cdots d\mathbf{r}_N \frac{1}{3V} \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i e^{-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)} \end{aligned}$$

And:

$$\frac{1}{Z_N} \frac{\partial Z_N}{\partial V} = \frac{N}{V} + \frac{\beta}{3V} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle$$

which involves the force \mathbf{F}_i acting on a particle i

$$\frac{1}{Z_N} \frac{\partial Z_N}{\partial V} = \frac{N}{V} + \frac{\beta}{3V} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle$$

Again, for pairwise interactions, \mathbf{F}_i can be written as : $\sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i = \sum_{i=1}^N \sum_{j=1, j \neq i}^N \mathbf{r}_i \cdot \mathbf{F}_{ij}$
and since one has (interchanging i-j summations):

$$\sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i = \frac{1}{2} \left[\sum_{i,j, i \neq j} \mathbf{r}_i \cdot \mathbf{F}_{ij} + \sum_{i,j, i \neq j} \mathbf{r}_j \cdot \mathbf{F}_{ji} \right]$$

Using Newton's third law, we furthermore have : $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$

$$\sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i = \frac{1}{2} \left[\sum_{i,j, i \neq j} \mathbf{r}_i \cdot \mathbf{F}_{ij} - \sum_{i,j, i \neq j} \mathbf{r}_j \cdot \mathbf{F}_{ij} \right] = \frac{1}{2} \sum_{i,j, i \neq j} (\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{F}_{ij} \equiv \frac{1}{2} \sum_{i,j, i \neq j} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij}$$

and the Ensemble average is given by:

$$\begin{aligned} \frac{\beta}{3V} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle &= \frac{\beta}{6V} \left\langle \sum \mathbf{r}_{ij} \cdot \mathbf{F}_{ij} \right\rangle \\ &= \frac{\beta}{6VZ_N} \int d\mathbf{r}_1 \cdots d\mathbf{r}_N \sum_{i,j, i \neq j} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij} e^{-\beta U_{pair}(\mathbf{r}_1, \dots, \mathbf{r}_N)} \end{aligned}$$

We do the same as for the energy:

$$\begin{aligned}
\frac{\beta}{3V} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle &= \frac{\beta N(N-1)}{6VZ_N} \int d\mathbf{r}_1 \cdots d\mathbf{r}_N \mathbf{r}_{12} \cdot \mathbf{F}_{12} e^{-\beta U_{pair}(\mathbf{r}_1, \dots, \mathbf{r}_N)} \\
&= \frac{\beta}{6V} \int d\mathbf{r}_1 d\mathbf{r}_2 \mathbf{r}_{12} \cdot \mathbf{F}_{12} \left[\frac{N(N-1)}{Z_N} \int d\mathbf{r}_3 \cdots d\mathbf{r}_N e^{-\beta U_{pair}(\mathbf{r}_1, \dots, \mathbf{r}_N)} \right] \\
&= \frac{\beta}{6V} \int d\mathbf{r}_1 d\mathbf{r}_2 \mathbf{r}_{12} \cdot \mathbf{F}_{12} \rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \\
&= \frac{\beta N^2}{6V^3} \int d\mathbf{r}_1 d\mathbf{r}_2 \mathbf{r}_{12} \cdot \mathbf{F}_{12} g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \\
&= -\frac{\beta N^2}{6V^3} \int d\mathbf{r}_1 d\mathbf{r}_2 u'(r_{12}) r_{12} g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = -\frac{\beta N^2}{6V^2} \int_0^\infty dr 4\pi r^3 u'(r) g(r)
\end{aligned}$$

and for a pair potential U_{pair} : $\mathbf{F}_{12} = -\frac{\partial U_{pair}}{\partial \mathbf{r}_{12}} = -u'(|\mathbf{r}_1 - \mathbf{r}_2|) \frac{(\mathbf{r}_1 - \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} = -u'(r_{12}) \frac{\mathbf{r}_{12}}{r_{12}}$

And finally:

$$\boxed{\frac{P}{kT} = \rho - \frac{\rho^2}{6kT} \int_0^\infty dr 4\pi r^3 u'(r) g(r)}$$

The pressure can be computed from the derivative of the pair potential and from $g(r)$.

Readings on time-correlation functions

1. Berne, B. J., Time-Dependent Properties of Condensed Media. In *Physical Chemistry: An Advanced Treatise*, Vol. VIIIB, Henderson, D., Ed. Academic Press: New York, 1971.
2. Berne, B. J.; Pecora, R., *Dynamic Light Scattering*. R. E. Krieger Publishing Co.: Malabar, FL, 1990.
3. Chandler, D., *Introduction to Modern Statistical Mechanics*. Oxford University Press: New York, 1987.
4. Mazenko, G., *Nonequilibrium Statistical Mechanics*. Wiley-VCH: Weinheim, 2006.
5. McHale, J. L., *Molecular Spectroscopy*. 1st ed.; Prentice Hall: Upper Saddle River, NJ, 1999.
6. McQuarrie, D. A., *Statistical Mechanics*. Harper & Row: New York, 1976; Ch. 21.
7. Schatz, G. C.; Ratner, M. A., *Quantum Mechanics in Chemistry*. Dover Publications: Mineola, NY, 2002; Ch. 6.
8. Wang, C. H., *Spectroscopy of Condensed Media: Dynamics of Molecular Interactions*. Academic Press: Orlando, 1985. 9. Zwanzig, R., *Nonequilibrium Statistical Mechanics*. Oxford University Press: New York, 2001.

Advance in Chemical Physics, Volume XVII

Edited by I. Prigogine, Stuart A. Rice

Copyright © 1970, by John Wiley & Sons, Inc.

ON THE CALCULATION OF TIME CORRELATION FUNCTIONS

B. J. BERNE

and

G. D. HARP

CONTENTS
