
MDANSE - theoretical background

July 16, 2015

Description

This note describes the theoretical background of the analyses in *MDANSE*. *MDANSE* (**M**olecular **D**ynamics **A**nalysis for **N**eutron **S**cattering **E**xperiments) is a python application for analyzing molecular dynamics simulation data.

Contents

1	Dynamics Analyses	2
1.1	Mean-Square Displacement	2
1.2	Root Mean-Square Deviation	4
1.3	Radius of gyration	5
1.4	Angular Correlation	5
1.5	Velocity Autocorrelation Function	6
1.6	Density Of States	7
1.7	Global Motion Filtered Trajectory	7
1.8	Rigid-Body Trajectory	8
1.9	Center Of Mass Trajectory	10
1.10	Order Parameter	10
2	Scattering analyses	11
2.1	Introduction	12
2.2	Dynamic Coherent Structure Factor	14
2.3	Dynamic Incoherent Structure Factor	15
2.4	Dynamic Incoherent Structure Factor (Gaussian Approximation)	16
2.5	Elastic Incoherent Structure Factor	17
2.6	Static Coherent Structure Factor	19
3	Structure analyses	19
3.1	Pair Distribution Function	19
3.2	Coordination number	20
3.3	Spatial Density	20
	References	22
	Appendices	25
	Acronyms	27

1 Dynamics Analyses

The following analyses are described:

- Mean-Square Displacement
- Root Mean-Square Displacement
- Radius Of Gyration
- Velocity AutoCorrelation Function
- Density Of States
- Global Motion Trajectory
- Center Of Mass Trajectory
- Rigid-Body Trajectory
- Center Of Mass Trajectory

1.1 Mean-Square Displacement

Theory and implementation

Molecules in liquids and gases do not stay in the same place, but move constantly. This process is called diffusion and it happens quite naturally in fluids at equilibrium. During this process, the motion of an individual molecule does not follow a simple path [48]. As it travels, the molecule undergoes some collisions with other molecules which prevent it from following a straight line. If the path is examined in close detail, it will be seen to be a good approximation to a random walk. Mathematically, a random walk is a series of steps where each step is taken in a completely random direction from the one before. This kind of path was famously analysed by Albert Einstein in a study of Brownian motion. He showed that the **Mean-Square Displacement (MSD)** of a particle following a random walk is proportional to the time elapsed. This relationship can be written as

$$\langle r^2 \rangle = 6Dt + C \quad (1)$$

where $\langle r^2 \rangle$ is the **MSD** and t is the time. D and C are constants. The constant D defines the so-called diffusion coefficient.

The figure 1 shows an example of a **MSD** analysis performed on a waterbox of 768 water molecules.

To get the diffusion coefficient out of this plot, the slope of the linear part of the plot should be calculated.

Defining,

$$\mathbf{d}_\alpha(t, t_0) \doteq \mathbf{R}_\alpha(t_0 + t) - \mathbf{R}_\alpha(t_0). \quad (2)$$

the **MSD** of particle α can be defined as:

$$\Delta_\alpha^2(t) = \left\langle \mathbf{d}_\alpha^2(t, t_0) \right\rangle_{t_0} \quad (3)$$

where $\mathbf{R}_\alpha(t_0)$ and $\mathbf{R}_\alpha(t_0 + t)$ are respectively the position of particle α at times t_0 and $t_0 + t$. One can introduce a **MSD** with respect to a given axis \mathbf{n} :

$$\Delta_\alpha^2(t, t_0; \mathbf{n}) \doteq \left\langle d_\alpha^2(t, \tau; \mathbf{n}) \right\rangle_{t_0} \quad (4)$$

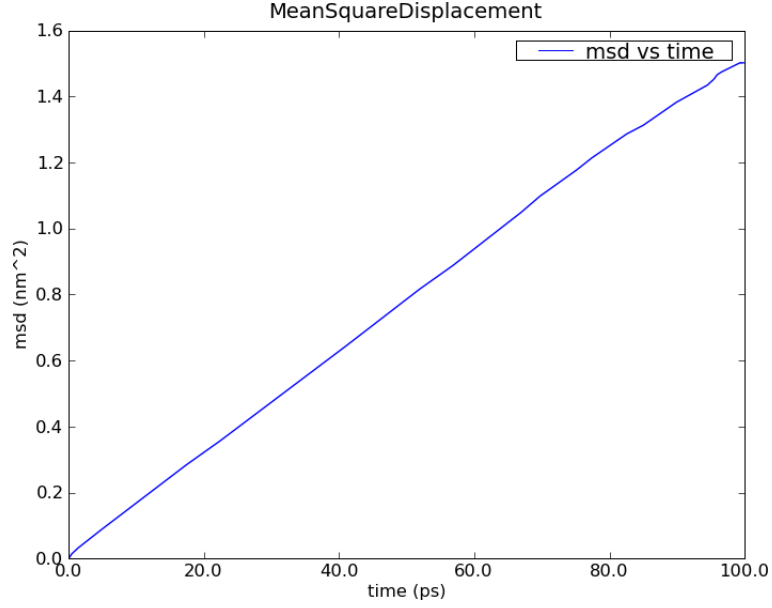


Figure 1: **MSD** calculated for a 100 ps MD simulation of 256 water molecules using NPT condition at 1 bar and 300 K.

with

$$d_{\alpha}(t, t_0; \mathbf{n}) \doteq \mathbf{n} \cdot \mathbf{d}_{\alpha}(t, t_0). \quad (5)$$

The calculation of **MSD** is the standard way to obtain diffusion coefficients from **Molecular Dynamics (MD)** simulations. Assuming Einstein-diffusion in the long time limit one has for isotropic systems

$$D_{\alpha} = \lim_{t \rightarrow \infty} \frac{1}{6t} \Delta_{\alpha}^2(t). \quad (6)$$

There exists also a well-known relation between the **MSD** and the velocity autocorrelation function. Writing $\mathbf{d}_{\alpha}(t) = \int_0^t d\tau \mathbf{v}_{\alpha}(\tau)$ in Eq. (3) one can show (see e.g. [50]) that

$$\Delta_{\alpha}^2(t) = 6 \int_0^t d\tau (t - \tau) C_{vv;\alpha\alpha}(\tau). \quad (7)$$

Using now the definition (6) of the diffusion coefficient one obtains the relation

$$D_{\alpha} = \int_0^t d\tau C_{vv;\alpha\alpha}(\tau). \quad (8)$$

With Eq. (28) this can also be written as

$$D_{\alpha} = \pi \tilde{C}_{vv;\alpha\alpha}(0). \quad (9)$$

Computationally, the **MSD** is calculated using the **Fast Correlation Algorithm (FCA)** [51]. In this framework, in the discrete case, the mean-square displacement of a particle is given by

$$\Delta^2(m) = \frac{1}{N_t - m} \sum_{k=0}^{N_t - m - 1} [\mathbf{r}(k + m) - \mathbf{r}(k)]^2, \quad m = 0 \dots N_t - 1, \quad (10)$$

where $\mathbf{r}(k)$ is the particle trajectory and N_t is the the number of frames of the trajectory. We define now the auxiliary function

$$S(m) \doteq \sum_{k=0}^{N_t-m-1} [\mathbf{r}(k+m) - \mathbf{r}(k)]^2, \quad m = 0 \dots N_t - 1, \quad (11)$$

which is splitted as follow:

$$S(m) = S_{AA+BB}(m) - 2S_{AB}(m), \quad (12)$$

$$S_{AA+BB}(m) = \sum_{k=0}^{N_t-m-1} [\mathbf{r}^2(k+m) + \mathbf{r}^2(k)], \quad (13)$$

$$S_{AB}(m) = \sum_{k=0}^{N_t-m-1} \mathbf{r}(k) \cdot \mathbf{r}(k+m). \quad (14)$$

The function $S_{AB}(m)$ can be computed using the **FCA** method described in Section 3.3. For $S_{AA+BB}(m)$ the following recursion relation holds:

$$S_{AA+BB}(m) = S_{AA+BB}(m-1) - \mathbf{r}^2(m-1) - \mathbf{r}^2(N_t-m), \quad (15)$$

$$S_{AA+BB}(0) = \sum_{k=0}^{N_t-1} \mathbf{r}^2(k). \quad (16)$$

This allows one to construct the following efficient scheme for the computation of the **MSD**:

1. Compute $DSQ(k) = \mathbf{r}^2(k)$, $k = 0 \dots N_t - 1$; $DSQ(-1) = DSQ(N_t) = 0$.
2. Compute $SUMSQ = 2 \cdot \sum_{k=0}^{N_t-1} DSQ(k)$.
3. Compute $S_{AB}(m)$ using the **Fast Fourier Transform (FFT)** method.
4. Compute $MSD(m)$ in the following loop:

$$\begin{aligned} SUMSQ &\leftarrow SUMSQ - DSQ(m-1) - DSQ(N_t-m) \\ MSD(m) &\leftarrow (SUMSQ - 2 \cdot S_{AB}(m)) / (N_t - m) \\ m &\text{ running from } 0 \text{ to } N_t - 1 \end{aligned}$$

It should be noted that the efficiency of this algorithm is the same as for the **FCA** computation of time correlation functions since the number of operations in step (1), (2), and (4) grows linearly with N_t .

1.2 Root Mean-Square Deviation

Theory and implementation

The **Root Mean-Square Deviation (RMSD)** is maybe the most popular estimator of structural similarity. It is a numerical measure of the difference between two structures that can be defined as:

$$RMSD(t) = \sqrt{\frac{\sum_{\alpha=1}^{N_{\alpha}} (\mathbf{r}_{\alpha}(t) - \mathbf{r}_{\alpha}(t_{ref}))^2}{N_{\alpha}}} \quad (17)$$

where N_{α} is the number of atoms of the system, and $\mathbf{r}_{\alpha}(t)$ and $\mathbf{r}_{\alpha}(t_{ref})$ are respectively the position of atom α at time t and t_{ref} where t_{ref} is a reference time usually choosen as the first step of the simulation. Typically, **RMSD** is used to quantify the structural evolution of the

system during the simulation. It can provide precious information about the system especially if it reached equilibrium or conversely if major structural changes occurred during the simulation.

In **Molecular Dynamics Analysis for Neutron Scattering Experiments (MDANSE)**, *RMSD* is computed using the discretized version of equation 17:

$$RMSD(n \cdot \Delta t) = \sqrt{\frac{\sum_{\alpha=1}^{N_{\alpha}} (\mathbf{r}_{\alpha}(t) - \mathbf{r}_{ref}(t))^2}{N_{\alpha}}}, \quad n = 0 \dots N_t - 1. \quad (18)$$

where N_t is the number of frames and Δt is the time step.

1.3 Radius of gyration

Theory and implementation

Radius Of Gyration (ROG) is the name of several related measures of the size of an object, a surface, or an ensemble of points. It is calculated as the Root Mean Square Distance between the system and a reference that can be either the center of gravity of the system either a given axis. In **MDANSE**, the reference is chosen to be the center of gravity of the system under study. Mathematically, it can be defined as:

$$ROG(t) = \sqrt{\frac{\sum_{\alpha=1}^{N_{\alpha}} (\mathbf{r}_{\alpha}(t) - \mathbf{r}_{cms}(t))^2}{N_{\alpha}}} \quad (19)$$

where N_{α} is the number of atoms of the system, and $\mathbf{r}_{\alpha}(t)$ and $\mathbf{r}_{cms}(t)$ are respectively the position of atom α and the center of mass of the system at time t .

ROG describes the overall spread of the molecule and as such is a good measure for the molecule compactness. For example, it can be useful when monitoring folding process.

In **MDANSE**, **ROG** is computed using the discretized version of equation 19:

$$ROG(n \cdot \Delta t) = \sqrt{\frac{\sum_{\alpha=1}^{N_{\alpha}} (\mathbf{r}_{\alpha}(t) - \mathbf{r}_{cms}(t))^2}{N_{\alpha}}}, \quad n = 0 \dots N_t - 1. \quad (20)$$

where N_t is the number of frames and Δt is the time step.

1.4 Angular Correlation

Theory and implementation

The angular correlation analysis computes the autocorrelation of a set of vectors describing the extent of a molecule in three orthogonal directions. This kind of analysis can be useful when trying to highlight the fact that a molecule is constrained in a given direction.

For a given triplet of non-colinear atoms $g=(\mathbf{a1}, \mathbf{a2}, \mathbf{a3})$, one can derive an orthonormal set of three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ using the following scheme:

- $\mathbf{v}_1 = \frac{\mathbf{n}_1 + \mathbf{n}_2}{\|\mathbf{n}_1 + \mathbf{n}_2\|}$ where \mathbf{n}_1 and \mathbf{n}_2 are respectively the normalized vectors along $(\mathbf{a1}, \mathbf{a2})$ and $(\mathbf{a1}, \mathbf{a3})$ directions.
- \mathbf{v}_2 is defined as the clockwise normal vector orthogonal to \mathbf{v}_1 that belongs to the plane defined by $\mathbf{a1}, \mathbf{a2}$ and $\mathbf{a3}$ atoms
- $\vec{v}_3 = \vec{v}_1 \times \vec{v}_2$

Thus, one can define the following autocorrelation functions for the vectors \mathbf{v}_1 , \mathbf{v}_2 and \mathbf{v}_3 defined on triplet t :

$$AC_{g,i}(t) = \langle \mathbf{v}_{t,i}(0) \cdot \mathbf{v}_{t,i}(t) \rangle, \quad i = 1, 2, 3 \quad (21)$$

And the angular correlation averaged over all triplets is:

$$AC_i(t) = \sum_{g=1}^{N_{\text{triplets}}} AC_{g,i}(t), \quad i = 1, 2, 3 \quad (22)$$

where N_{triplets} is the number of selected triplets.

1.5 Velocity Autocorrelation Function

Theory and implementation

The **Velocity AutoCorrelation Function (VACF)** is another interesting property describing the dynamics of a molecular system. Indeed, it reveals the underlying nature of the forces acting on the system.

In a molecular system that would be made of non interacting particles, the velocities would be constant at any time triggering the **VACF** to be a constant value. Now, if we think about a system with small interactions such as in a gas-phase, the magnitude and direction of the velocity of a particle will change gradually over time due to its collision with the other particles of the molecular system. In such a system, the **VACF** will be represented by a decaying exponential.

In the case of solid phase, the interaction are much stronger and, as a results, the atoms are bound to a given position from which they will move backwards and forwards oscillating between positive and negative values of their velocity. The oscillations will not be of equal magnitude however, but will decay in time, because there are still perturbative forces acting on the atoms to disrupt the perfection of their oscillatory motion. So, in that case the **VACF** will look like a damped harmonic motion.

Finally, in the case of liquid phase, the atoms have more freedom than in solid phase and because of the diffusion process, the oscillatory motion seen in solid phase will be cancelled quite rapidly depending on the density of the system. So, the **VACF** will just have one very damped oscillation before decaying to zero. This decaying time can be considered as the average time for a collision between two atoms to occur before they diffuse away.

Mathematically, the **VACF** of atom α in an atomic or molecular system is usually defined as

$$C_{vv;\alpha\alpha}(t) \doteq \frac{1}{3} \langle \mathbf{v}_\alpha(t_0) \cdot \mathbf{v}_\alpha(t_0 + t) \rangle_{t_0}. \quad (23)$$

In some cases, e.g. for non-isotropic systems, it is useful to define **VACF** along a given axis,

$$C_{vv;\alpha\alpha}(t; \mathbf{n}) \doteq \langle v_\alpha(t_0; \mathbf{n}) v_\alpha(t_0 + t; \mathbf{n}) \rangle_{t_0}, \quad (24)$$

where $v_\alpha(t; \mathbf{n})$ is given by

$$v_\alpha(t; \mathbf{n}) \doteq \mathbf{n} \cdot \mathbf{v}_\alpha(t). \quad (25)$$

The vector \mathbf{n} is a unit vector defining a space-fixed axis.

The **VACF** of the particles in a many body system can be related to the incoherent dynamic structure factor by the relation:

$$\lim_{q \rightarrow 0} \frac{\omega^2}{q^2} \mathcal{S}(\mathbf{q}, \omega) = G(\omega), \quad (26)$$

where $G(\omega)$ is the **Density Of States (DOS)**. For an isotropic system it reads

$$G(\omega) = \sum_{\alpha} b_{\alpha,inc}^2 \tilde{C}_{vv;\alpha\alpha}(\omega), \quad (27)$$

$$\tilde{C}_{vv;\alpha\alpha}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dt \exp[-i\omega t] C_{vv;\alpha\alpha}(t). \quad (28)$$

For non-isotropic systems relation (26) holds if the **DOS** is computed from the atomic velocity autocorrelation functions $C_{vv;\alpha\alpha}(t; \mathbf{n}_q)$, where \mathbf{n}_q is the unit vector in the direction of \mathbf{q} .

1.6 Density Of States

Theory and implementation

MDANSE calculates the power spectrum of the **VACF**, which in case of the mass-weighted **VACF** defines the phonon discrete **DOS**, (see Section 1.5) defined as:

$$DOS(n \cdot \Delta\nu) \doteq \sum_{\alpha} \omega_{\alpha} \tilde{C}_{vv;\alpha\alpha}(n \cdot \Delta\nu), \quad n = 0 \dots N_t - 1. \quad (29)$$

N_t is the total number of time steps and $\Delta\nu = 1/(2N_t\Delta t)$ is the frequency step. $DOS(n \cdot \Delta\nu)$ can be computed either for the isotropic case or with respect to a user-defined axis. The spectrum $DOS(n \cdot \Delta\nu)$ is computed from the *unnormalized VACF*, such that $DOS(0)$ gives an approximate value for the diffusion constant $D = \sum_{\alpha} D_{\alpha}$ (see Eqs. 8 and 9). $DOS(n \cdot \Delta\nu)$ is smoothed by applying a Gaussian window in the time domain [76] (see Section 3.3). Its width in the time domain is $\sigma_t = \alpha/T$, where T is the length of the simulation. We remark that the diffusion constant obtained from **DOS** is biased due to the spectral smoothing procedure since the **VACF** is weighted by this window Gaussian function. **MDANSE** computes the density of states starting from both atomic velocities and atomic coordinates. In this case the velocities are computed by numerical differentiation of the coordinate trajectories correcting first for possible jumps due to periodic boundary conditions.

1.7 Global Motion Filtered Trajectory

Theory and implementation

It is often of interest to separate global motion from internal motion, both for quantitative analysis and for visualization by animated display. Obviously, this can be done under the hypothesis that global and internal motions are decoupled within the length and timescales of the analysis. **MDANSE** can create **Global Motion Filtered Trajectory (GMFT)** by filtering out global motions (made of the three translational and rotational degrees of freedom), either on the whole system or on an user-defined subset, by fitting it to a reference structure (usually the first frame of the **MD**). Global motion filtering uses a straightforward algorithm:

- for the first frame, find the linear transformation such that the coordinate origin becomes the center of mass of the system and its principal axes of inertia are parallel to the three coordinates axes (also called principal axes transformation),
- this provides a reference configuration \mathcal{C}_{ref} ,
- for any other frames f , finds and applies the linear transformation that minimizes the RMS distance between frame f and \mathcal{C}_{ref} .

The result is stored in a new trajectory file that contains only internal motions. This analysis can be useful in case where diffusive motions are not of interest or simply not accessible to the experiment (time resolution, powder analysis ...).

1.8 Rigid-Body Trajectory

Theory and implementation

To analyze the dynamics of complex molecular systems it is often desirable to consider the overall motion of molecules or molecular subunits. We will call this motion rigid-body motion in the following. Rigid-body motions are fully determined by the dynamics of the centroid, which may be the center-of-mass, and the dynamics of the angular coordinates describing the orientation of the rigid body. The angular coordinates are the appropriate variables to compute angular correlation functions of molecular systems in space and time. In most cases, however, these variables are not directly available from *MD* simulations since *MD* algorithms typically work in cartesian coordinates. Molecules are either treated as flexible, or, if they are treated as rigid, constraints are taken into account in the framework of cartesian coordinates [54]. In *MDANSE*, **Rigid-Body Trajectory (RBT)** can be defined from a *MD* trajectory by fitting rigid reference structures, defining a (sub)molecule, to the corresponding structure in each time frame of the trajectory. Here ‘fit’ means the optimal superposition of the structures in a least-squares sense. We will describe now how rigid body motions, i.e. global translations and rotations of molecules or subunits of complex molecules, can be extracted from a *MD* trajectory. A more detailed presentation is given in [55]. We define an optimal rigid-body trajectory in the following way: for each time frame of the trajectory the atomic positions of a rigid reference structure, defined by the three cartesian components of its centroid (e.g. the center of mass) and three angles, are as close as possible to the atomic positions of the corresponding structure in the *MD* configuration. Here ‘as close as possible’ means as close as possible in a least-squares sense.

Optimal superposition. We consider a given time frame in which the atomic positions of a (sub)molecule are given by \mathbf{x}_α , $\alpha = 1 \dots N$. The corresponding positions in the reference structure are denoted as $\mathbf{x}_\alpha^{(0)}$, $\alpha = 1 \dots N$. For both the given structure and the reference structure we introduce the yet undetermined centroids \mathbf{X} and $\mathbf{X}^{(0)}$, respectively, and define the deviation

$$\Delta_\alpha \doteq \mathbf{D}(\mathbf{q}) [\mathbf{x}_\alpha^{(0)} - \mathbf{X}^{(0)}] - [\mathbf{x}_\alpha - \mathbf{X}]. \quad (30)$$

Here $\mathbf{D}(\mathbf{q})$ is a rotation matrix which depends on also yet undetermined angular coordinates which we chose to be *quaternion parameters*, abbreviated as vector $\mathbf{q} = (q_0, q_1, q_2, q_3)$. The quaternion parameters fulfill the normalization condition $\mathbf{q} \cdot \mathbf{q} = 1$ [56]. The target function to be minimized is now defined as

$$m(\mathbf{q}; \mathbf{X}, \mathbf{X}^{(0)}) = \sum_{\alpha} \omega_{\alpha} |\Delta_{\alpha}|^2. \quad (31)$$

where ω_{α} are atomic weights (see Section ??). The minimization with respect to the centroids is decoupled from the minimization with respect to the quaternion parameters and yields

$$\mathbf{X} = \sum_{\alpha} \omega_{\alpha} \mathbf{x}_{\alpha}, \quad (32)$$

$$\mathbf{X}^{(0)} = \sum_{\alpha} \omega_{\alpha} \mathbf{x}_{\alpha}^{(0)}. \quad (33)$$

We are now left with a minimization problem for the rotational part which can be written as

$$m(\mathbf{q}) = \sum_{\alpha} \omega_{\alpha} [\mathbf{D}(\mathbf{q}) \mathbf{r}_{\alpha}^{(0)} - \mathbf{r}_{\alpha}]^2 \stackrel{!}{=} Min. \quad (34)$$

The relative position vectors

$$\mathbf{r}_{\alpha} = \mathbf{x}_{\alpha} - \mathbf{X}, \quad (35)$$

$$\mathbf{r}_{\alpha}^{(0)} = \mathbf{x}_{\alpha}^{(0)} - \mathbf{X}^{(0)}, \quad (36)$$

are fixed and the rotation matrix reads [56]

$$\mathbf{D}(\mathbf{q}) = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(-q_0q_3 + q_1q_2) & 2(q_0q_2 + q_1q_3) \\ 2(q_0q_3 + q_1q_2) & q_0^2 + q_2^2 - q_1^2 - q_3^2 & 2(-q_0q_1 + q_2q_3) \\ 2(-q_0q_2 + q_1q_3) & 2(q_0q_1 + q_2q_3) & q_0^2 + q_3^2 - q_1^2 - q_2^2 \end{pmatrix}. \quad (37)$$

Quaternions and rotations. The rotational minimization problem can be elegantly solved by using quaternion algebra. Quaternions are so-called hypercomplex numbers, having a real unit, $\mathbf{1}$, and three imaginary units, \mathbf{I} , \mathbf{J} , and \mathbf{K} . Since $\mathbf{IJ} = \mathbf{K}$ (cyclic), quaternion multiplication is not commutative. A possible matrix representation of an arbitrary quaternion,

$$\mathbf{A} = a_0 \cdot \mathbf{1} + a_1 \cdot \mathbf{I} + a_2 \cdot \mathbf{J} + a_3 \cdot \mathbf{K}, \quad (38)$$

reads

$$\mathbf{A} = \begin{pmatrix} a_0 & -a_1 & -a_2 & -a_3 \\ a_1 & a_0 & -a_3 & a_2 \\ a_2 & a_3 & a_0 & -a_1 \\ a_3 & -a_2 & a_1 & a_0 \end{pmatrix}. \quad (39)$$

The components a_ν are real numbers. Similarly as normal complex numbers allow one to represent rotations in a plane, quaternions allow one to represent rotations in space. Consider the quaternion representation of a vector \mathbf{r} , which is given by

$$\mathbf{R} = x \cdot \mathbf{I} + y \cdot \mathbf{J} + z \cdot \mathbf{K}, \quad (40)$$

and perform the operation

$$\mathbf{R}' = \mathbf{Q}\mathbf{R}\mathbf{Q}^T, \quad (41)$$

where \mathbf{Q} is a normalized quaternion,

$$\|\mathbf{Q}\|^2 \doteq q_0^2 + q_1^2 + q_2^2 + q_3^2 = \frac{1}{4} \text{tr}\{\mathbf{Q}^T \mathbf{Q}\} = 1. \quad (42)$$

The symbol *tr* stands for ‘trace’. We note that a normalized quaternion is represented by an *orthogonal* 4×4 matrix. \mathbf{R}' may then be written as

$$\mathbf{R}' = x' \cdot \mathbf{I} + y' \cdot \mathbf{J} + z' \cdot \mathbf{K}, \quad (43)$$

where the components x', y', z' , abbreviated as \mathbf{r}' , are given by

$$\mathbf{r}' = \mathbf{D}(\mathbf{q})\mathbf{r}. \quad (44)$$

The matrix $\mathbf{D}(\mathbf{q})$ is the rotation matrix defined in (37).

Solution of the minimization problem. In quaternion algebra, the rotational minimization problem may now be phrased as follows:

$$m(\mathbf{q}) = \sum_{\alpha} \omega_{\alpha} \|\mathbf{Q}\mathbf{R}_{\alpha}^{(0)}\mathbf{Q}^T - \mathbf{R}_{\alpha}\|^2 \stackrel{!}{=} \text{Min}. \quad (45)$$

Since the matrix \mathbf{Q} representing a normalized quaternion is orthogonal this may also be written as

$$m(\mathbf{q}) = \sum_{\alpha} \omega_{\alpha} \|\mathbf{Q}\mathbf{R}_{\alpha}^{(0)} - \mathbf{R}_{\alpha}\mathbf{Q}\|^2 \stackrel{!}{=} \text{Min}. \quad (46)$$

This follows from the simple fact that $\|\mathbf{A}\| = \|\mathbf{A}\mathbf{Q}\|$, if \mathbf{Q} is normalized. Eq. (46) shows that the target function to be minimized can be written as a simple quadratic form in the quaternion parameters [55],

$$m(\mathbf{q}) = \mathbf{q} \cdot \mathbf{M}\mathbf{q}, \quad (47)$$

$$\mathbf{M} = \sum_{\alpha} \omega_{\alpha} \mathbf{M}_{\alpha}. \quad (48)$$

The matrices \mathbf{M}_{α} are positive semi-definite matrices depending on the positions \mathbf{r}_{α} and $\mathbf{r}_{\alpha}^{(0)}$:

$$\left. \begin{aligned} M_{\alpha,11} &= x_{\alpha}^2 + y_{\alpha}^2 + z_{\alpha}^2 + x_{0\alpha}^2 + y_{0\alpha}^2 + z_{0\alpha}^2 - 2x_{\alpha}x_{0\alpha} - 2y_{\alpha}y_{0\alpha} - 2z_{\alpha}z_{0\alpha} \\ M_{\alpha,12} &= 2(y_{\alpha}z_{0\alpha} - z_{\alpha}y_{0\alpha}) \\ M_{\alpha,13} &= 2(-x_{\alpha}z_{0\alpha} + z_{\alpha}x_{0\alpha}) \\ M_{\alpha,14} &= 2(x_{\alpha}y_{0\alpha} - y_{\alpha}x_{0\alpha}) \\ M_{\alpha,22} &= x_{\alpha}^2 + y_{\alpha}^2 + z_{\alpha}^2 + x_{0\alpha}^2 + y_{0\alpha}^2 + z_{0\alpha}^2 - 2x_{\alpha}x_{0\alpha} + 2y_{\alpha}y_{0\alpha} + 2z_{\alpha}z_{0\alpha} \\ M_{\alpha,23} &= -2(x_{\alpha}y_{0\alpha} + y_{\alpha}x_{0\alpha}) \\ M_{\alpha,24} &= -2(x_{\alpha}z_{0\alpha} + z_{\alpha}x_{0\alpha}) \\ M_{\alpha,33} &= x_{\alpha}^2 + y_{\alpha}^2 + z_{\alpha}^2 + x_{0\alpha}^2 + y_{0\alpha}^2 + z_{0\alpha}^2 + 2x_{\alpha}x_{0\alpha} - 2y_{\alpha}y_{0\alpha} + 2z_{\alpha}z_{0\alpha} \\ M_{\alpha,44} &= -2(y_{\alpha}z_{0\alpha} + z_{\alpha}y_{0\alpha}) \\ M_{\alpha,44} &= x_{\alpha}^2 + y_{\alpha}^2 + z_{\alpha}^2 + x_{0\alpha}^2 + y_{0\alpha}^2 + z_{0\alpha}^2 + 2x_{\alpha}x_{0\alpha} + 2y_{\alpha}y_{0\alpha} - 2z_{\alpha}z_{0\alpha} \end{aligned} \right\} \quad (49)$$

The rotational fit is now reduced to the problem of finding the minimum of a quadratic form with the constraint that the quaternion to be determined must be normalized. Using the method of Lagrange multipliers to account for the normalization constraint we have

$$m'(\mathbf{q}, \lambda) = \mathbf{q} \cdot \mathbf{M}\mathbf{q} - \lambda(\mathbf{q} \cdot \mathbf{q} - 1) \stackrel{!}{=} \text{Min}. \quad (50)$$

This leads immediately to the eigenvalue problem

$$\mathbf{M}\mathbf{q} = \lambda\mathbf{q}, \quad (51)$$

$$\mathbf{q} \cdot \mathbf{q} = 1. \quad (52)$$

Now any normalized eigenvector \mathbf{q} fulfills the relation $\lambda = \mathbf{q} \cdot \mathbf{M}\mathbf{q} \equiv m(\mathbf{q})$. Therefore the eigenvector belonging to the smallest eigenvalue, λ_{min} , is the desired solution. At the same time λ_{min} gives the average error per atom.

The result of *RBT* analysis is stored in a new trajectory file that contains only *RBT* motions.

1.9 Center Of Mass Trajectory

Theory and implementation

The **Center Of Mass Trajectory** (*COMT*) analysis consists in deriving the trajectory of the respective centers of mass of a set of groups of atoms. In order to produce a visualizable trajectory, *MDANSE* assigns the centers of mass to pseudo-hydrogen atoms whose mass is equal to the mass of their associated group. Thus, the produced trajectory can be reused for other analysis. In that sense, *COMT* analysis is a practical way to reduce noticeably the dimensionality of a system.

1.10 Order Parameter

Theory and implementation

Adequate and accurate cross comparison of the NMR and *MD* simulation data is of crucial

importance in versatile studies conformational dynamics of proteins. NMR relaxation spectroscopy has proven to be a unique approach for a site-specific investigation of both global tumbling and internal motions of proteins. The molecular motions modulate the magnetic interactions between the nuclear spins and lead for each nuclear spin to a relaxation behavior which reflects its environment. Since its first applications to the study of protein dynamics, a wide variety of experiments has been proposed to investigate backbone as well as side chain dynamics. Among them, the heteronuclear relaxation measurement of amide backbone ^{15}N nuclei is one of the most widespread techniques. The relationship between microscopic motions and measured spin relaxation rates is given by Redfield's theory [70]. Under the hypothesis that ^{15}N relaxation occurs through dipole-dipole interactions with the directly bonded ^1H atom and chemical shift anisotropy (CSA), and assuming that the tensor describing the CSA is axially symmetric with its axis parallel to the N-H bond, the relaxation rates of the ^{15}N nuclei are determined by a time correlation function,

$$C_{ii}(t) = \langle P_2(\mu_i(0) \cdot \mu_i(t)) \rangle \quad (53)$$

which describes the dynamics of a unit vector $\mu_i(t)$ pointing along the ^{15}N - ^1H bond of the residue i in the laboratory frame. Here $P_2(\cdot)$ is the second order Legendre polynomial. The Redfield theory shows that relaxation measurements probe the relaxation dynamics of a selected nuclear spin only at a few frequencies. Moreover, only a limited number of independent observables are accessible. Hence, to relate relaxation data to protein dynamics one has to postulate either a dynamical model for molecular motions or a functional form for $C_{ii}(t)$, yet depending on a limited number of adjustable parameters. Usually, the tumbling motion of proteins in solution is assumed isotropic and uncorrelated with the internal motions, such that:

$$C_{ii}(t) = C^G(t) \cdot C_{ii}^I(t) \quad (54)$$

where $C^G(t)$ and $C_{ii}^I(t)$ denote the global and the internal time correlation function, respectively. Within the so-called model free approach [71, 72] the internal correlation function is modeled by an exponential,

$$C_{ii}^I(t) = S_i^2 + (1 - S_i^2) \exp\left(-\frac{t}{\tau_{eff,i}}\right) \quad (55)$$

Here the asymptotic value $S_i^2 = C_{ii}(+\infty)$ is the so-called generalized order parameter, which indicates the degree of spatial restriction of the internal motions of a bond vector, while the characteristic time $\tau_{eff,i}$ is an effective correlation time, setting the time scale of the internal relaxation processes. S_i^2 can adopt values ranging from 0 (completely disordered) to 1 (fully ordered). So, S_i^2 is the appropriate indicator of protein backbone motions in computationally feasible timescales as it describes the spatial aspects of the reorientational motion of N-H peptidic bonds vector.

When performing Order Parameter analysis, **MDANSE** computes for each residue i both $C_{ii}(t)$ and S_i^2 . It also computes a correlation function averaged over all the selected bonds defined as:

$$C^I(t) = \sum_{i=1}^{N_{bonds}} C_{ii}^I(t) \quad (56)$$

where N_{bonds} is the number of selected bonds for the analysis.

2 Scattering analyses

The following analyses are described:

- Dynamic Coherent Structure Factor
- Dynamic Incoherent Structure Factor
- Dynamic Incoherent Structure Factor (Gaussian Approximation)
- Elastic Incoherent Structure Factor
- Static Coherent Structure Factor
- Smoothed Static Coherent Structure Factor

Before introducing each of these analysis, a brief introduction about the scattering theory within the classical framework will be given.

2.1 Introduction

The quantity of interest in neutron scattering experiments with thermal neutrons is the *dynamic structure factor*, $\mathcal{S}(\mathbf{q}, \omega)$, which is closely related to the double differential cross-section [7], $d^2\sigma/d\Omega dE$. The double differential cross section is defined as the number of neutrons which are scattered per unit time into the solid angle interval $[\Omega, \Omega + d\Omega]$ and into the energy interval $[E, E + dE]$. It is normalized to $d\Omega$, dE , and the flux of the incoming neutrons,

$$\frac{d^2\sigma}{d\Omega dE} = N \cdot \frac{k}{k_0} \mathcal{S}(\mathbf{q}, \omega). \quad (57)$$

Here N is the number of atoms, and $k \equiv |\mathbf{k}|$ and $k_0 \equiv |\mathbf{k}_0|$ are the wave numbers of scattered and incident neutrons, respectively. They are related to the corresponding neutron energies by $E = \hbar^2 k^2 / 2m$ and $E_0 = \hbar^2 k_0^2 / 2m$, where m is the neutron mass. The arguments of the dynamic structure factor, \mathbf{q} and ω , are the momentum and energy transfer in units of \hbar , respectively:

$$\mathbf{q} = \frac{\mathbf{k}_0 - \mathbf{k}}{\hbar}, \quad (58)$$

$$\omega = \frac{E_0 - E}{\hbar}. \quad (59)$$

The modulus of the momentum transfer can be expressed in the scattering angle θ , the energy transfer, and the energy of the incident neutrons:

$$q = \sqrt{2 - \frac{\hbar\omega}{E_0} - 2 \cos \theta \sqrt{2 - \frac{\hbar\omega}{E_0}}}. \quad (60)$$

The dynamic structure factor contains information about the structure and dynamics of the scattering system [67]. It can be written as

$$\mathcal{S}(\mathbf{q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dt \exp[-i\omega t] \mathcal{F}(\mathbf{q}, t). \quad (61)$$

$\mathcal{F}(\mathbf{q}, t)$ is called the *intermediate scattering function* and is defined as

$$\mathcal{F}(\mathbf{q}, t) = \sum_{\alpha, \beta} \Gamma_{\alpha\beta} \langle \exp[-i\mathbf{q} \cdot \hat{\mathbf{R}}_{\alpha}(0)] \exp[i\mathbf{q} \cdot \hat{\mathbf{R}}_{\beta}(t)] \rangle, \quad (62)$$

$$\Gamma_{\alpha\beta} = \frac{1}{N} \left[\overline{b_{\alpha} b_{\beta}} + \delta_{\alpha\beta} (\overline{b_{\alpha}^2} - \overline{b_{\alpha}}^2) \right]. \quad (63)$$

The operators $\hat{\mathbf{R}}_\alpha(t)$ in Eq. (62) are the position operators of the nuclei in the sample. The brackets $\langle \dots \rangle$ denote a quantum thermal average and the time dependence of the position operators is defined by the Heisenberg picture. The quantities b_α are the scattering lengths of the nuclei which depend on the isotope and the relative orientation of the spin of the neutron and the spin of the scattering nucleus. If the spins of the nuclei and the neutron are not prepared in a special orientation one can assume a random relative orientation and that spin and position of the nuclei are uncorrelated. The symbol $\overline{\dots}$ appearing in $\Gamma_{\alpha\beta}$ denotes an average over isotopes and relative spin orientations of neutron and nucleus.

Usually one splits the intermediate scattering function and the dynamic structure factor into their *coherent* and *incoherent* parts which describe collective and single particle motions, respectively. Defining

$$b_{\alpha,coh} \doteq \overline{b_\alpha}, \quad (64)$$

$$b_{\alpha,inc} \doteq \sqrt{\overline{b_\alpha^2} - \overline{b_\alpha}^2}, \quad (65)$$

the coherent and incoherent intermediate scattering functions can be cast in the form

$$\mathcal{F}_{coh}(\mathbf{q}, t) = \frac{1}{N} \sum_{\alpha, \beta} b_{\alpha,coh} b_{\beta,coh} \langle \exp[-i\mathbf{q} \cdot \hat{\mathbf{R}}_\alpha(0)] \exp[i\mathbf{q} \cdot \hat{\mathbf{R}}_\beta(t)] \rangle, \quad (66)$$

$$\mathcal{F}_{inc}(\mathbf{q}, t) = \frac{1}{N} \sum_{\alpha} b_{\alpha,inc}^2 \langle \exp[-i\mathbf{q} \cdot \hat{\mathbf{R}}_\alpha(0)] \exp[i\mathbf{q} \cdot \hat{\mathbf{R}}_\alpha(t)] \rangle. \quad (67)$$

Rewriting these formulas, **MDANSE** introduces the partial terms as:

$$\mathcal{F}_{coh}(\mathbf{q}, t) = \sum_{I, J \geq I}^{N_{species}} \sqrt{n_I n_J \omega_{I,coh} \omega_{J,coh}} \mathcal{F}_{IJ,coh}(\mathbf{q}, t), \quad (68)$$

$$\mathcal{F}_{inc}(\mathbf{q}, t) = \sum_{I=1}^{N_{species}} n_I \omega_{I,inc} \mathcal{F}_{I,inc}(\mathbf{q}, t) \quad (69)$$

where:

$$\mathcal{F}_{IJ,coh}(\mathbf{q}, t) = \frac{1}{\sqrt{n_I n_J}} \sum_{\alpha}^{n_I} \sum_{\beta}^{n_J} \langle \exp[-i\mathbf{q} \cdot \hat{\mathbf{R}}_\alpha(t_0)] \exp[i\mathbf{q} \cdot \hat{\mathbf{R}}_\beta(t_0 + t)] \rangle_{t_0}, \quad (70)$$

$$\mathcal{F}_{I,inc}(\mathbf{q}, t) = \frac{1}{n_I} \sum_{\alpha=1}^{n_I} \langle \exp[-i\mathbf{q} \cdot \hat{\mathbf{R}}_\alpha(t_0)] \exp[i\mathbf{q} \cdot \hat{\mathbf{R}}_\alpha(t_0 + t)] \rangle_{t_0}. \quad (71)$$

where n_I , n_J , $N_{species}$, $\omega_{I,coh,inc}$ and $\omega_{J,coh,inc}$ are defined in Section ??.

The corresponding dynamic structure factors are obtained by performing the Fourier transformation defined in Eq. 61.

An important quantity describing *structural* properties of liquids is the *static structure factor*, which is defined as

$$\mathcal{S}(\mathbf{q}) \doteq \int_{-\infty}^{+\infty} d\omega \mathcal{S}_{coh}(\mathbf{q}, \omega) = \mathcal{F}_{coh}(\mathbf{q}, 0). \quad (72)$$

In the classical framework the intermediate scattering functions are interpreted as classical time correlation functions. The position operators are replaced by time-dependent vector functions and quantum thermal averages are replaced by classical *ensemble averages*. It is well known that this procedure leads to a loss of the universal detailed balance relation,

$$\mathcal{S}(\mathbf{q}, \omega) = \exp[\beta \hbar \omega] \mathcal{S}(-\mathbf{q}, -\omega), \quad (73)$$

and also to a loss of all odd moments

$$\langle \omega^{2n+1} \rangle \doteq \int_{-\infty}^{+\infty} d\omega \omega^{2n+1} \mathcal{S}(\mathbf{q}, \omega), \quad n = 1, 2, \dots \quad (74)$$

The odd moments vanish since the classical dynamic structure factor is even in ω , assuming invariance of the scattering process with respect to reflections in space. The first moment is also universal. For an atomic liquid, containing only one sort of atoms, it reads

$$\langle \omega \rangle = \frac{\hbar q^2}{2M}, \quad (75)$$

where M is the mass of the atoms. Formula (75) shows that the first moment is given by the average kinetic energy (in units of \hbar) of a particle which receives a momentum transfer $\hbar \mathbf{q}$. Therefore $\langle \omega \rangle$ is called the *recoil moment*. A number of ‘recipes’ has been suggested to correct classical dynamic structure factors for detailed balance and to describe recoil effects in an approximate way. The most popular one has been suggested by Schofield [68]

$$\mathcal{S}(\mathbf{q}, \omega) \approx \exp\left[\frac{\beta \hbar \omega}{2}\right] \mathcal{S}_d(\mathbf{q}, \omega). \quad (76)$$

One can easily verify that the resulting dynamic structure factor fulfills the relation of detailed balance. Formally, the correction (76) is correct to first order in \hbar . Therefore it cannot be used for large q -values which correspond to large momentum transfers $\hbar q$. This is actually true for all correction methods which have suggested so far. For more details we refer to Ref. [9].

2.2 Dynamic Coherent Structure Factor

Theory and implementation

Please refer to Section 2.1 for more details about the theoretical background related to the dynamic coherent structure factor. In this analysis, *MDANSE* proceeds in two steps. First, it computes the partial and total intermediate coherent scattering function using equation 68. Then, the partial and total dynamic coherent structure factors are obtained by performing the Fourier Transformation, defined in Eq. 61, respectively on the total and partial intermediate coherent scattering functions.

MDANSE computes the coherent intermediate scattering function on a rectangular grid of equidistantly spaced points along the time-and the q -axis, respectively:

$$\mathcal{F}_{\text{coh}}(q_m, k \cdot \Delta t) \doteq \sum_{I=1, J \geq I}^{N_{\text{species}}} \sqrt{n_I n_J \omega_{I, \text{com}} \omega_{J, \text{com}}} \overline{\langle \rho_I(-\mathbf{q}, 0) \rho_J(\mathbf{q}, k \cdot \Delta t) \rangle}^q, \quad k = 0 \dots N_t - 1, \quad m = 0 \dots N_q - 1. \quad (77)$$

where N_t is the number of time steps in the coordinate time series, N_q is a user-defined number of q -shells, N_{species} is the number of selected species, n_I the number of atoms of species I , ω_I the weight for specie I (see Section ?? for more details) and $\rho_I(\mathbf{q}, k \cdot \Delta t)$ is the Fourier transformed particle density for specie I defined as,

$$\rho_I(\mathbf{q}, k \cdot \Delta t) = \sum_{\alpha}^{n_I} \exp[i\mathbf{q} \cdot \mathbf{R}_{\alpha}(k \cdot \Delta t)]. \quad (78)$$

The symbol $\overline{\dots}^q$ in (77) denotes an average over q -vectors having *approximately* the same modulus $q_m = q_{\text{min}} + m \cdot \Delta q$. The particle density must not change if jumps in the particle trajectories due to periodic boundary conditions occur. In addition the *average* particle density,

N/V , must not change. This can be achieved by choosing q -vectors on a lattice which is reciprocal to the lattice defined by the **MD** box. Let $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ be the basis vectors which span the **MD** cell. Any position vector in the **MD** cell can be written as

$$\mathbf{R} = x'\mathbf{b}_1 + y'\mathbf{b}_2 + z'\mathbf{b}_3, \quad (79)$$

with x', y', z' having values between 0 and 1. The primes indicate that the coordinates are box coordinates. A jump due to periodic boundary conditions causes x', y', z' to jump by ± 1 . The set of dual basis vectors $\mathbf{b}^1, \mathbf{b}^2, \mathbf{b}^3$ is defined by the relation

$$\mathbf{b}_i \mathbf{b}^j = \delta_i^j. \quad (80)$$

If the q -vectors are now chosen as

$$\mathbf{q} = 2\pi (k\mathbf{b}^1 + l\mathbf{b}^2 + m\mathbf{b}^3), \quad (81)$$

where k, l, m are integer numbers, jumps in the particle trajectories produce phase changes of multiples of 2π in the Fourier transformed particle density, i.e. leave it unchanged. One can define a grid of q -shells or a grid of q -vectors along a given direction or on a given plane, giving in addition a *tolerance* for q . **MDANSE** looks then for q -vectors of the form (87) whose moduli deviate within the prescribed tolerance from the equidistant q -grid. From these q -vectors only a maximum number per grid-point (called generically q -shell also in the anisotropic case) is kept.

The q -vectors can be generated isotropically, anisotropically or along user-defined directions.

The $\sqrt{\omega_I}$ may be negative if they represent normalized coherent scattering lengths, i.e.

$$\sqrt{\omega_I} = \frac{b_{I,\text{coh}}}{\sqrt{\sum_{I=1}^{N_{\text{species}}} n_I b_{I,\text{coh}}^2}}. \quad (82)$$

Negative coherent scattering lengths occur in hydrogenous materials since $b_{\text{coh},H}$ is negative [7]. The density-density correlation is computed via the **FCA** technique described in Section 3.3.

2.3 Dynamic Incoherent Structure Factor

Theory and implementation

Please refer to Section 2.1 for more details about the theoretical background related to the dynamic incoherent structure factor. In this analysis, **MDANSE** proceeds in two steps. First, it computes the partial and total intermediate incoherent scattering function $\mathcal{F}_{\text{inc}}(\mathbf{q}, t)$ using equation 69. Then, the partial and total dynamic incoherent structure factors are obtained by performing the Fourier Transformation, defined in Eq.61, respectively on the total and partial intermediate incoherent scattering function.

MDANSE computes the incoherent intermediate scattering function on a rectangular grid of equidistantly spaced points along the time-and the q -axis, respectively:

$$\mathcal{F}_{\text{inc}}(q_m, k \cdot \Delta t) \doteq \sum_{I=1}^{N_{\text{species}}} n_I \omega_{I,\text{inc}} F_{I,\text{inc}}(q_m, k \cdot \Delta t), \quad k = 0 \dots N_t - 1, \quad m = 0 \dots N_q - 1. \quad (83)$$

where N_t is the number of time steps in the coordinate time series, N_q is a user-defined number of q -shells, N_{species} is the number of selected species, n_I the number of atoms of species I , $\omega_{I,\text{inc}}$ the weight for specie I (see Section ?? for more details) and $F_{I,\text{inc}}(q_m, k \cdot \Delta t)$ is defined as:

$$F_{I,\text{inc},\alpha}(q_m, k \cdot \Delta t) = \sum_{\alpha=1}^{n_I} \overline{\langle \exp[-i\mathbf{q} \cdot \mathbf{R}_\alpha(0)] \exp[i\mathbf{q} \cdot \mathbf{R}_\alpha(t)] \rangle}^q. \quad (84)$$

The symbol $\overline{\dots}^q$ in (84) denotes an average over q -vectors having *approximately* the same modulus $q_m = q_{min} + m \cdot \Delta q$. The particle density must not change if jumps in the particle trajectories due to periodic boundary conditions occur. In addition the *average* particle density, N/V , must not change. This can be achieved by choosing q -vectors on a lattice which is reciprocal to the lattice defined by the **MD** box. Let $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ be the basis vectors which span the **MD** cell. Any position vector in the **MD** cell can be written as

$$\mathbf{R} = x'\mathbf{b}_1 + y'\mathbf{b}_2 + z'\mathbf{b}_3, \quad (85)$$

with x', y', z' having values between 0 and 1. The primes indicate that the coordinates are box coordinates. A jump due to periodic boundary conditions causes x', y', z' to jump by ± 1 . The set of dual basis vectors $\mathbf{b}^1, \mathbf{b}^2, \mathbf{b}^3$ is defined by the relation

$$\mathbf{b}_i \mathbf{b}^j = \delta_i^j. \quad (86)$$

If the q -vectors are now chosen as

$$\mathbf{q} = 2\pi \left(k\mathbf{b}^1 + l\mathbf{b}^2 + m\mathbf{b}^3 \right), \quad (87)$$

where k, l, m are integer numbers, jumps in the particle trajectories produce phase changes of multiples of 2π in the Fourier transformed particle density, i.e. leave it unchanged. One can define a grid of q -shells or a grid of q -vectors along a given direction or on a given plane, giving in addition a *tolerance* for q . **MDANSE** looks then for q -vectors of the form (87) whose moduli deviate within the prescribed tolerance from the equidistant q -grid. From these q -vectors only a maximum number per grid-point (called generically q -shell also in the anisotropic case) is kept.

The q -vectors can be generated isotropically, anisotropically or along user-defined directions.

The correlation functions defined in 84 are computed via the **FCA** technique described in Section 3.3. Although the efficient **FCA** technique is used to compute the atomic time correlation functions, the program may consume a considerable amount of CPU-time since the number of time correlation functions to be computed equals the number of atoms times the total number of q -vectors. This analysis is actually one of the most time-consuming among all the analysis available in **MDANSE**.

2.4 Dynamic Incoherent Structure Factor (Gaussian Approximation)

Theory and implementation

The **MSD** can be related to the incoherent intermediate scattering function via the cumulant expansion [49, 50]

$$\mathcal{F}_{\text{inc}}^g(\mathbf{q}, t) = \sum_{I=1}^{N_{\text{species}}} n_I \omega_{I,\text{inc}} \mathcal{F}_{I,\text{inc}}^g(\mathbf{q}, t) \quad (88)$$

where N_{species} is the number of selected species, n_I the number of atoms of species I , $\omega_{I,\text{inc}}$ the weight for specie I (see Section ?? for more details) and

$$\mathcal{F}_{I,\text{inc}}^g(\mathbf{q}, t) = \frac{1}{n_I} \sum_{\alpha}^{n_I} \exp[-q^2 \rho_{\alpha,1}(t) + q^4 \rho_{\alpha,2}(t) \mp \dots]. \quad (89)$$

The cumulants $\rho_{\alpha,k}(t)$ are defined as

$$\rho_{\alpha,1}(t) = \frac{1}{2!} \langle d_{\alpha}^2(t; \mathbf{n}_q) \rangle \quad (90)$$

$$\rho_{\alpha,2}(t) = \frac{1}{4!} \left[\langle d_{\alpha}^4(t; \mathbf{n}_q) \rangle - 3 \langle d_{\alpha}^2(t; \mathbf{n}_q) \rangle^2 \right] \quad (91)$$

\vdots

The vector \mathbf{n}_q is the unit vector in the direction of \mathbf{q} . In the Gaussian approximation the above expansion is truncated after the q^2 -term. For certain model systems like the ideal gas, the harmonic oscillator, and a particle undergoing Einstein diffusion, this is exact. For these systems the incoherent intermediate scattering function is completely determined by the *MSD*.

MDANSE allows one to compute the total and partial incoherent intermediate scattering function in the *Gaussian approximation* by discretizing equation 88:

$$\mathcal{F}_{\text{inc}}^g(q_m, k \cdot \Delta t) \doteq \sum_{I=1}^{N_{\text{species}}} n_I \omega_{I,\text{inc}} F_{I,\text{inc}}^g(q_m, k \cdot \Delta t), \quad k = 0 \dots N_t - 1, \quad m = 0 \dots N_q - 1. \quad (92)$$

with for each specie the following expression for the intermediate scattering function:

$$F_{I,\alpha,\text{inc}}^g(q_m, k \cdot \Delta t) = \frac{1}{n_I} \sum_{\alpha}^{n_I} \exp \left[-\frac{(q_m)^2}{6} \Delta_{\alpha}^2(k \cdot \Delta t) \right] \quad \text{isotropic system}, \quad (93)$$

$$F_{I,\alpha,\text{inc}}^g(q_m, k \cdot \Delta t) = \frac{1}{n_I} \sum_{\alpha}^{n_I} \exp \left[-\frac{(q_m)^2}{2} \Delta_{\alpha}^2(k \cdot \Delta t; \mathbf{n}) \right] \quad \text{non-isotropic system}. \quad (94)$$

N_t is the total number of time steps in the coordinate time series and N_q is a user-defined number of q -shells. The (q, t) -grid is the same as for the calculation of the intermediate incoherent scattering function (see Section 2.3). The quantities $\Delta_{\alpha}^2(t)$ and $\Delta_{\alpha}^2(t; \mathbf{n})$ are the mean-square displacements, defined in Equations (3) and (4), respectively. They are computed by using the algorithm described in Section 1.1. *MDANSE* corrects the atomic input trajectories for jumps due to periodic boundary conditions. It should be noted that the computation of the intermediate scattering function in the Gaussian approximation is much ‘cheaper’ than the computation of the full intermediate scattering function, $\mathcal{F}_{\text{inc}}(q, t)$, since no averaging over different q -vectors needs to be performed. It is sufficient to compute a single mean-square displacement per atom.

2.5 Elastic Incoherent Structure Factor

Theory and implementation

The **Elastic Incoherent Structure Factor** (*EISF*) is defined as the limit of the incoherent intermediate scattering function for infinite time,

$$EISF(\mathbf{q}) \doteq \lim_{t \rightarrow \infty} \mathcal{F}_{\text{inc}}(\mathbf{q}, t). \quad (95)$$

Using the above definition of the EISF one can decompose the incoherent intermediate scattering function as follows:

$$\mathcal{F}_{\text{inc}}(\mathbf{q}, t) = EISF(\mathbf{q}) + \mathcal{F}'_{\text{inc}}(\mathbf{q}, t), \quad (96)$$

where $\mathcal{F}'_{\text{inc}}(\mathbf{q}, t)$ decays to zero for infinite time. Taking now the Fourier transform it follows immediately that

$$\mathcal{S}_{\text{inc}}(\mathbf{q}, \omega) = EISF(\mathbf{q})\delta(\omega) + \mathcal{S}'_{\text{inc}}(\mathbf{q}, \omega). \quad (97)$$

The *EISF* appears as the amplitude of the *elastic* line in the neutron scattering spectrum. Elastic scattering is only present for systems in which the atomic motion is confined in space, as for solids. To understand which information is contained in the *EISF* we consider for simplicity a system where only one sort of atoms is visible to the neutrons. To a very good approximation this is the case for all systems containing a large amount of hydrogen atoms, as biological

systems. Incoherent scattering from hydrogen dominates by far all other contributions. Using the definition of the van Hove self-correlation function $G_s(\mathbf{r}, t)$ [7],

$$b_{\text{inc}}^2 G_s(\mathbf{r}, t) \doteq \frac{1}{2\pi^3} \int d^3q \exp[-i\mathbf{q} \cdot \mathbf{r}] \mathcal{F}_{\text{inc}}(\mathbf{q}, t), \quad (98)$$

which can be interpreted as the conditional probability to find a tagged particle at the position \mathbf{r} at time t , given it started at $\mathbf{r} = \mathbf{0}$, one can write:

$$EISF(\mathbf{q}) = b_{\text{inc}}^2 \int d^3r \exp[i\mathbf{q} \cdot \mathbf{r}] G_s(\mathbf{r}, t = \infty). \quad (99)$$

The *EISF* gives the sampling distribution of the points in space in the limit of infinite time. In a real experiment this means times longer than the time which is observable with a given instrument. The *EISF* vanishes for all systems in which the particles can access an infinite volume since $G_s(\mathbf{r}, t)$ approaches $1/V$ for large times. This is the case for molecules in liquids and gases.

For computational purposes it is convenient to use the following representation of the *EISF* [14]:

$$EISF(\mathbf{q}) = \sum_{I=1}^{N_{\text{species}}} n_I \omega_{I,\text{inc}} EISF_I(q) \quad (100)$$

where N_{species} is the number of selected species, n_I the number of atoms of species I , $\omega_{I,\text{inc}}$ the weight for specie I (see Section ?? for more details) and for each specie the following expression for the elastic incoherent scattering function is

$$EISF_I(\mathbf{q}) = \frac{1}{n_I} \sum_{\alpha}^{n_I} \langle |\exp[i\mathbf{q} \cdot \mathbf{R}_{\alpha}]|^2 \rangle. \quad (101)$$

This expression is derived from definition (95) of the *EISF* and expression (69) for the intermediate scattering function, using that for infinite time the relation

$$\langle \exp[-i\mathbf{q} \cdot \mathbf{R}_{\alpha}(0)] \exp[i\mathbf{q} \cdot \mathbf{R}_{\alpha}(t)] \rangle = \langle |\exp[i\mathbf{q} \cdot \mathbf{R}_{\alpha}]|^2 \rangle \quad (102)$$

holds. In this way the computation of the *EISF* is reduced to the computation of a static thermal average. We remark at this point that the length of the *MD* trajectory from which the *EISF* is computed should be long enough to allow for a representative sampling of the conformational space.

MDANSE allows one to compute the elastic incoherent structure factor on a grid of equidistantly spaced points along the q -axis:

$$EISF(q_m) \doteq \sum_{I=1}^{N_{\text{species}}} n_I \omega_I EISF_I(q_m), m = 0 \dots N_q - 1. \quad (103)$$

where N_q is a user-defined number of q -shells, the values for q_m are defined as $q_m = q_{\text{min}} + m \cdot \Delta q$, and for each specie the following expression for the elastic incoherent scattering function is:

$$EISF_I(q_m) = \frac{1}{n_I} \sum_{\alpha}^{n_I} \overline{\langle |\exp[i\mathbf{q} \cdot \mathbf{R}_{\alpha}]|^2 \rangle}^q. \quad (104)$$

Here the symbol $\overline{\dots}^q$ denotes an average over the q -vectors having the same modulus q_m . The program corrects the atomic input trajectories for jumps due to periodic boundary conditions.

2.6 Static Coherent Structure Factor

Theory and implementation

This analysis is a shortcut to obtain the static coherent structure factor defined as $S(q) = \mathcal{F}_{\text{coh}}(q, t = 0)$. It uses exactly the same procedure as the one defined in Section 2.2.

3 Structure analyses

The following analyses are described:

- Pair Distribution Function
- Coordination Number
- Spatial Density

3.1 Pair Distribution Function

Theory and implementation

The **Pair Distribution Function (PDF)** is an example of a pair correlation function, which describes how, on average, the atoms in a system are radially packed around each other. This proves to be a particularly effective way of describing the average structure of disordered molecular systems such as liquids. Also in systems like liquids, where there is continual movement of the atoms and a single snapshot of the system shows only the instantaneous disorder, it is extremely useful to be able to deal with the average structure.

The **PDF** is useful in other ways. For example, it is something that can be deduced experimentally from x-ray or neutron diffraction studies, thus providing a direct comparison between experiment and simulation. It can also be used in conjunction with the interatomic pair potential function to calculate the internal energy of the system, usually quite accurately.

Mathematically, the **PDF** can be computed using the following formula:

$$PDF(r) = \sum_{I=1, J \geq I}^{N_{\text{species}}} n_I n_J \omega_I \omega_J g_{IJ}(r) \quad (105)$$

where N_{species} is the number of selected species, n_I and n_J are respectively the numbers of atoms of species I and J , ω_I and ω_J respectively the weights for species I and J (see Section ?? for more details) and $PDF_{\alpha\beta}(r)$ is the partial **PDF** for I and J species that can be defined as:

$$PDF_{IJ}(r) = \frac{\langle \sum_{\alpha=1}^{n_I} n_{\alpha J}(r) \rangle}{n_I \rho_J 4\pi r^2 dr} \quad (106)$$

where ρ_J is the density of atom of specie J and $n_{\alpha J}(r)$ is the mean number of atoms of specie J in a shell of width dr at distance r of the atom α of specie I .

From the computation of **PDF**, two related quantities are computed in **MDANSE**, the **Radial Distribution Function (RDF)** defined as:

$$RDF(r) = 4\pi r^2 \rho_0 PDF(r) \quad (107)$$

and the **Total Correlation Function (TCF)** defined as:

$$TCF(r) = 4\pi r \rho_0 (PDF(r) - 1.0) \quad (108)$$

where ρ_0 is the average atomic density defined as:

$$\rho_0 = \frac{N}{V} \quad (109)$$

where N is the total number of atoms of the system and V the volume of the simulation box.

In *MDANSE*, the *PDF*, the *RDF* and the *TCF* are further splitted into an intra- and inter-molecular parts which added together give respectively the total *PDF*, *RDF* and *TCF*.

3.2 Coordination number

Theory and implementation

In chemistry, the **Coordination Number (CN)** is the total number of neighbours of a central atom in a molecule or ion. The definition used in *MDANSE* is somewhat different and can be seen as an extension of as the former definition. Indeed, in *MDANSE*, the *CN* is not defined over one defined central atom but around the centers of gravity of a set of group of atoms. So, if only one group made of only atom is selected for the analysis, then, the definition is the same as the original definition. In that context, the *CN* is defined as:

$$n(r, r + dr) = \frac{1}{N_G} \sum_{g=1}^{N_G} \sum_{I=1}^{N_{species}} n_{gI}(r, r + dr) \quad (110)$$

where N_G is the number of groups of atoms, $N_{species}$ is the number of species found in the system and $n_{gI}(r)$ is the *CN* defined for specie I defined as the number of atoms of species I found in a shell of width dr at a distance r of the center of gravity of the group of atom g .

MDANSE allows one to compute the *CN* on a set of equidistantly spaced distances at different times:

$$CN(r_m) \doteq \frac{1}{N_{frames}} \frac{1}{N_G} \sum_{f=1}^{N_{frames}} \sum_{g=1}^{N_G} \sum_{I=1}^{N_{species}} CN_{gI}(r_m, t_f), \quad m = 0 \dots N_r - 1, \quad n = 0 \dots N_{frames} - 1. \quad (111)$$

where N_r and N_{frames} are respectively the number of distances and times at which the *CN* is evaluated and

$$CN_{gI}(r_m, t_f) = n_{gI}(r_m, t_f), \quad (112)$$

is the number of atoms of specie I found within $[r_m, r_m + dr]$ at frame f from the center of gravity of group g .

From these expression, several remarks can be done. Firstly, the Eqs 111 and 112 can be restricted to intramolecular and intermolecular distances only. Secondly, these equations can be averaged over the selected frames providing a time averaged intra and intermolecular *CN*. Finally, the same equations (time-dependent and time-averaged) can be integrated over r to provide a cumulative *CN*. *MDANSE* computes all these variations.

The concept of *CN* is useful for structure-related analysis. It can reveal for instance some packing effects that may have occurred during the simulation.

3.3 Spatial Density

Theory and implementation

The **Spatial Density (SD)** can be seen as an generalization of the pair distribution function. Indeed, pair distribution functions are defined as orientationally averaged distribution functions. Although these correlation functions reflects many key features of the short-range order in molecular systems, it should be realized that an average spatial assembly of non-spherical particles

can not be uniquely characterized from these one-dimensional functions. So, structural models postulated for the molecular ordering in nonsimple systems based only on one-dimensional *PDF* will always be somewhat ambiguous. The goal of *SD* analysis is to provide greater clarity in the structural analysis of molecular systems by utilizing distribution function which span both the radial and angular coordinates of the separation vector. This can provide useful information about the average local structure in a complex system.

MDANSE allows one to compute the *SD* in spherical coordinates on a set of concentric shells surrounding the centers of mass of selected triplets of atoms using the formula:

$$SD(r_l, \theta_m, \phi_n) \doteq \frac{1}{N_{triplets} N_{groups}} \sum_{t=1}^{N_{triplets}} \sum_{g=1}^{N_{groups}} \langle n_{tg}(r_l, \theta_m, \phi_n) \rangle, \quad l = 0 \dots N_r - 1, m = 0 \dots N_\theta - 1, n = 0 \dots N_\phi - 1 \quad (113)$$

where $N_{triplets}$ and N_{groups} are respectively the number of triplets and groups, r_l , θ_m and ϕ_n are the spherical coordinates at which the *SD* is evaluated, N_r , N_θ and N_ϕ are respectively the number of discrete r , θ and ϕ values and $n_{tg}(r_l, \theta_m, \phi_n)$ is the number of group of atoms of type g whose centers of mass is found to be in the volume element defined by $[r, r + dr]$, $[\theta, \theta + d\theta]$ and $[\phi, \phi + d\phi]$ in the spherical coordinates basis centered on the center of mass of triplet t .

So technically, *MDANSE* proceeds more or less on the following way:

- defines the center of mass c_i^t $i = 1, 2 \dots N_{triplets}$ for each triplet of atoms,
- defines the center of mass c_i^g $i = 1, 2 \dots N_{groups}$ for each group of atoms,
- constructs an oriented orthonormal basis \mathcal{R}_i^t $i = 1, 2 \dots N_{triplets}$ centered on each c^t , this basis is defined from the three vectors \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{v}_3 ,
 - $\mathbf{v}_1 = \frac{\mathbf{n}_1 + \mathbf{n}_2}{\|\mathbf{n}_1 + \mathbf{n}_2\|}$ where \mathbf{n}_1 and \mathbf{n}_2 are respectively the normalized vectors in $(\mathbf{a1}, \mathbf{a2})$ and $(\mathbf{a1}, \mathbf{a3})$ directions where $(\mathbf{a1}, \mathbf{a2}, \mathbf{a3})$ are the three atoms of the triplet t ,
 - \mathbf{v}_2 is defined as the clockwise normal vector orthogonal to \mathbf{v}_1 that belongs to the plane defined by $\mathbf{a1}$, $\mathbf{a2}$ and $\mathbf{a3}$ atoms,
 - $\vec{v}_3 = \vec{v}_1 \times \vec{v}_2$
- expresses the cartesian coordinates of each c^g in each \mathcal{R}^t ,
- transforms these coordinates in spherical coordinates,
- discretizes the spherical coordinates in r_l , θ_m and ϕ_n ,
- does $n_{tg}(r_l, \theta_m, \phi_n) = n_{tg}(r_l, \theta_m, \phi_n) + 1$.

References

- [1] Rog, T.; Murzin, K.; Hinsén, K.; Kneller, G.R. *J. Comp. Chem.* **2002**, *24*, 657-667.
- [2] Kneller, G.R.; Keiner, V.; Kneller, M.; Schiller, M. *Comp. Phys. Comm.* **1995**, *91*, 191-214.
- [3] Hinsén, K. *J. Comp. Chem.* **2000**, *21*, 79-85.
- [4] <http://sourcesup.cru.fr/projects/mmtk/>
- [5] <http://www.unidata.ucar.edu/software/netcdf/>
- [6] Rew, R.; Davis, G.; Emmerson, S.; Davies, H. NetcdfUser'sGuide for c, an interface for Data Access, version 3. <http://www.unidata.ucar.edu/packages/netcdf/guidec>, 1997.
- [7] Lovesey, S.W. *Theory of Neutron Scattering from Condensed Matter, Vol. 1*; Clarendon Press, Oxford, 1984.
- [8] Bee, M. *Quasielastic Neutron Scattering: Principles and Applications in Solid State Chemistry, Biology, and Materials Science*; Adam Hilger, Bristol, 1988.
- [9] Kneller, G.R. *Molecular Physics* **1994**, *83*(1), 63-87.
- [10] Kneller, G.R.; Geiger, A. *Molecular Physics* **1989**, *68*(2), 487-498.
- [11] Kneller, G.R.; Geiger, A. *Molecular Physics* **1990**, *70*(3), 465-483.
- [12] Boehm H.J.; Ahlrichs, R. *Mol. Phys.* **1985**, *54*(6), 1261-1274.
- [13] Anderson, J.; Ullo, J.J.; Yip, S. *J. Chem. Phys.* **1987**, *86*(7), 4078-4089.
- [14] Kneller, G.R.; Doster, W.; Settles, M.; Cusack, S.; Smith, J.C. *J. Chem. Phys.*, **1992**, *97*(12), 8864-8879.
- [15] Dianoux, A.J.; Kneller, G.R.; Sauvageol, J.L.; Smith, J.C. *J. Chem. Phys.*, **1993**, *99*(7), 5586-5596.
- [16] van Rossum, G. The Python web site, <http://www.python.org/>.
- [17] <http://www.python.org/download/>
- [18] <http://numpy.scipy.org/>
- [19] <http://pyro.sourceforge.net/>
- [20] Hinsén, K. ScientificPython. <http://dirac.cnrs-orleans.fr>
- [21] <http://sourcesup.cru.fr/projects/scientific-py/>
- [22] OpenSource web site. <http://www.opensource.org>, 1997.
- [23] <http://www.cosc.canterbury.ac.nz/greg.ewing/python/Pyrex/>
- [24] <http://wiki.python.org/moin/TkInter>
- [25] <http://matplotlib.sourceforge.net/>
- [26] <http://www.ambermd.org>

- [27] <http://www.charmm.org/>
- [28] http://www.cse.scitech.ac.uk/ccg/software/DL_POLY
- [29] <http://www.accelrys.com/products/materials-studio>
- [30] <http://www.accelrys.com/products/materials-studio/modules/discover.html>
- [31] <http://www.accelrys.com/products/materials-studio/modules/forcite.html>
- [32] <http://www.ks.uiuc.edu/Research/namd>
- [33] <http://www.ks.uiuc.edu/Research/vmd>
- [34] <http://cms.mpi.univie.ac.at/vasp>
- [35] <http://cns-online.org/v1.21>
- [36] <http://www.unidata.ucar.edu/software/netcdf/ncdump-man-1.html>
- [37] <http://www.unidata.ucar.edu/software/netcdf/ngen-man-1.html>
- [38] Kneller, G.R.; Calligari, P. *Acta Crystallographica* **2006**, *D62* 302-311.
- [39] <http://sourcesup.cru.fr/projects/nmoldyn/>
- [40] <http://dirac.cnrs-orleans.fr/plone/software/nmoldyn>
- [41] <http://www.cecill.info/>
- [42] <http://sourceforge.net/projects/pywin32/>
- [43] <http://www.ucar.edu/>
- [44] <http://my.unidata.ucar.edu/content/software/netcdf/software.html>
- [45] <http://www.rcsb.org/pdb/>
- [46] <http://docs.python.org/library/configparser.html>
- [47] http://dirac.cnrs-orleans.fr/Manuals/MMTK/MMTK_37.html
- [48] <http://www.compsoc.man.ac.uk/lucky/Democritus/>
- [49] Rahman, A.; Singwi, K.S.; Sjölander, A. *Phys. Rev.* **1962**, *126*, 986-996.
- [50] Boon, J.P.; Yip, S. *Molecular Hydrodynamics*; McGraw-Hill, New York, 1980.
- [51] Kneller, G.R. Technical Report Jül 2215, Forschungszentrum Jülich (ISSN 0366-0885), ZB Forschungszentrum Jülich, D-52425 Jülich, Germany.
- [52] Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions*; Dover, New York, 1972.
- [53] Brooks, B.R.; Janezic, D.; Karplus, M. *J. Comp. Chem.* **1995**, *16*(12), 1522-1542.
- [54] Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C *J. Comput. Phys.* **1977**, *23*, 327-341.
- [55] Kneller, G.R. *Molecular Simulation* **1991**, *7*, 113-119.

- [56] Altmann, S.L. *Rotations, Quaternions, and Double Groups*; Clarendon Press, Oxford, 1986.
- [57] Kneller, G.R.; Hinsen, K. *J. Chem. Phys.* **2001**, *115*, 11097-11105.
- [58] Papoulis, A. *Probability, random variables, and Stochastic Processes*; 3rd ed; McGraw-Hill, New York, 1991.
- [59] Makhoul, J. *J. Proc. IEEE* **1975**, *63*, 561-580.
- [60] Burg, J. *Maximum Entropy Spectral Analysis*; PhD thesis, Stanford University, Stanford, CA, 1975.
- [61] Makhoul, J. *J. IEEE Transactions on Acoustics, Speech, and Signal Processing* **1977**, *25*, 423-428.
- [62] Lynden-Bell, R.M.; Stone, A.J. *Molecular Simulation* **1989**, *3*, 271-281.
- [63] Edmonds, A.R. *Angular momentum in Quantum Mechanics*; Princeton University Press, Princeton, New Jersey, 1957.
- [64] Brink, D.M.; Satchler, G.R. *Angular Momentum*; Clarendon Press, Oxford, 1968.
- [65] Rose, M.E. *Elementary Theory of Angular Momentum*; John Wiley, New York, 1957.
- [66] Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*; Oxford University Press, Oxford, 1987.
- [67] van Hove, L. *Phys. Rev.* **1954**, *95*(1), 249-262.
- [68] Schofield, P. *Phys. Rev. Letters* **1960**, *4*(5), 239-240.
- [69] Fischer, H.E.; Barnes, A.C.; Salmon, P.S. *Rep. Prog. Phys.* **2006**, *69*(1), 233-299.
- [70] Refield, A. *IBM J. Res. & Dev.* **1957**, *1*, 19-31.
- [71] Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559.
- [72] Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559-4570.
- [73] F. Zhang and R. Bruschweiler. *J. Am. Chem. Soc.* **2002**, *124*, 12654-12655.
- [74] Brigham, E.O. *The Fast Fourier Transform*; Prentice Hall, Englewood Cliffs (NJ) USA, 1974.
- [75] Papoulis, A. *Signal Analysis*, McGraw-Hill, Singapore, 1984.
- [76] Harris, F.J. *Proc. IEEE* **1978**, *66*(1), 51-83.
- [77] http://www.unidata.ucar.edu/software/netcdf/guide_12.html
- [78] <http://www.tcl.tk/software/tcltk/>

Appendices

The FCA algorithm Most of the quantities which can be extracted from *MD* simulations are time correlation functions. Correlation functions of discrete time series can be efficiently calculated by using the *FFT* [74]. The *FCA* allows the number of multiplications (complexity) to be reduced from $\propto N_t^2$ to $\propto N_t \log_2(N_t)$. In *MDANSE* all time correlation functions are computed using the *FCA* method which will be outlined in the following. We will also briefly comment on spectral smoothing of Fourier transformed correlation functions.

We consider two time series

$$a(k \cdot \Delta t), \quad b(k \cdot \Delta t), \quad k = 0 \dots N_t - 1, \quad (114)$$

of length $T = (N_t - 1) \cdot \Delta t$ which are to be correlated. In the following the shorthands $a(k)$ and $b(k)$ will be used. The discrete correlation function of $a(k)$ and $b(k)$ is defined as

$$c_{ab}(m) \doteq \begin{cases} \frac{1}{N_t - m} \sum_{k=0}^{N_t - m - 1} a^*(k) b(k + m), & m = 0 \dots N_t - 1, \\ \frac{1}{N_t - |m|} \sum_{k=|m|}^{N_t - 1} a^*(k) b(k - |m|), & m = -(N_t - 1) \dots -1. \end{cases} \quad (115)$$

The prefactors in front of the sums ensure the proper normalization of the individual channels, $m = -(N_t - 1) \dots N_t - 1$. The asterisk denotes a complex conjugate. According to (115), $c_{ab}(m)$ has $2N_t - 1$ data points and obeys the symmetry relation

$$c_{ab}(m) = c_{ba}^*(-m). \quad (116)$$

In case that $a(k)$ and $b(k)$ are identical, the corresponding correlation function $c_{aa}(m)$ is called an *autocorrelation* function. We define now the extended, periodic time series

$$A(k) = \begin{cases} a(k) & k = 0 \dots N_t - 1 \\ 0 & k = N_t \dots 2N_t - 1 \end{cases}, \quad (117)$$

$$B(k) = \begin{cases} b(k) & k = 0 \dots N_t - 1 \\ 0 & k = N_t \dots 2N_t - 1 \end{cases}, \quad (118)$$

which have the period $2N_t$,

$$A(k) = A(k + m \cdot 2N_t), \quad B(k) = B(k + m \cdot 2N_t), \quad m = 0, \pm 1, \pm 2, \dots \quad (119)$$

The discrete, cyclic correlation of $A(k)$ and $B(k)$ is defined as

$$S_{AB}(m) = \sum_{k=0}^{2N_t - 1} A^*(k) B(k + m). \quad (120)$$

It is easy to see that

$$c_{ab}(m) = \frac{1}{N_t - |m|} S_{AB}(m), \quad -(N_t - 1) \leq m \leq N_t - 1. \quad (121)$$

Using the correlation theorem of discrete periodic functions [74], $S_{AB}(m)$ can be written as

$$S_{AB}(m) = \frac{1}{2N_t} \sum_{n=0}^{2N_t - 1} \exp \left[2\pi i \left(\frac{mn}{2N_t} \right) \right] \tilde{A}^* \left(\frac{n}{2N_t} \right) \tilde{B} \left(\frac{n}{2N_t} \right) \quad (122)$$

where $\tilde{A}\left(\frac{n}{2N_t}\right)$ and $\tilde{B}\left(\frac{n}{2N_t}\right)$ are the discrete Fourier transforms of $A(k)$ and $B(k)$, respectively:

$$\tilde{A}\left(\frac{n}{2N_t}\right) = \sum_{k=0}^{2N_t-1} \exp\left[-2\pi i \left(\frac{nk}{2N_t}\right)\right] A(k), \quad (123)$$

$$\tilde{B}\left(\frac{n}{2N_t}\right) = \sum_{k=0}^{2N_t-1} \exp\left[-2\pi i \left(\frac{nk}{2N_t}\right)\right] B(k). \quad (124)$$

If the Fourier transforms of the signals $A(k)$ and $B(k)$ as well as the inverse transform in (122) are computed by *FFT*, $S_{AB}(m)$ can be computed by $\propto N_t \log_2(N_t)$ instead of $\propto N_t^2$ multiplications. It is sometimes said that the *FFT* method induces spurious correlations. We emphasize that this is only the case if the time series $a(k)$ and $b(k)$ are not properly extended, as indicated in Eqs. (117) and (118). The *FFT* method and the direct scheme (115) give, apart from round-off errors, *identical results*.

In many cases not only the computation of a correlation function is required, but also the computation of its Fourier spectrum. In principle one could use the product $\tilde{A}^*\left(\frac{n}{2N_t}\right) \tilde{B}\left(\frac{n}{2N_t}\right)$ which is already available as an intermediate step in the computation of $S_{AB}(m)$ according to (122). This would, however, not be a good estimate for the spectrum of $c_{ab}(m)$ [75]. In *MDANSE* all spectra are smoothed by applying a window in the time domain [75]:

$$P_{ab}\left(\frac{n}{2N_t}\right) = \Delta t \cdot \sum_{m=-(N_t-1)}^{N_t-1} \exp\left[-2\pi i \left(\frac{nm}{2N_t}\right)\right] W(m) \frac{1}{N - |m|} S_{AB}(m). \quad (125)$$

The time step Δt in front of the sum yields the proper normalization of the spectrum. In *MDANSE* a Gaussian window [76] is used:

$$W(m) = \exp\left[-\frac{1}{2} \left(\alpha \frac{|m|}{N_t - 1}\right)^2\right], \quad m = -(N_t - 1) \dots N_t - 1. \quad (126)$$

Its widths in the time and frequency domain are $\sigma_t = \alpha/T$ and $\sigma_\nu = 1/(2\pi\sigma_t)$, respectively. We recall that $T = (N_t - 1) \cdot \Delta t$ is the length of the simulation. σ_ν corresponds to the width of the resolution function of the Fourier spectrum.

Acronyms

<i>CN</i>	Coordination Number.	1, 20
<i>COMT</i>	Center Of Mass Trajectory.	1, 10
<i>DOS</i>	Density Of States.	1, 7
<i>EISF</i>	Elastic Incoherent Structure Factor.	1, 17, 18
<i>FCA</i>	Fast Correlation Algorithm.	1, 3, 4, 15, 16, 25
<i>FFT</i>	Fast Fourier Transform.	1, 4, 25, 26
<i>GMFT</i>	Global Motion Filtered Trajectory.	1, 7
<i>MDANSE</i>	Molecular Dynamics Analysis for Neutron Scattering Experiments.	1, 5, 7, 8, 10, 11, 13–21, 25, 26
<i>MD</i>	Molecular Dynamics.	1, 3, 7, 8, 10, 15, 16, 18, 25
<i>MSD</i>	Mean-Square Displacement.	1–4, 16, 17
<i>PDF</i>	Pair Distribution Function.	1, 19–21
<i>RBT</i>	Rigid-Body Trajectory.	1, 8, 10
<i>RDF</i>	Radial Distribution Function.	1, 19, 20
<i>RMSD</i>	Root Mean-Square Deviation.	1, 4, 5
<i>ROG</i>	Radius Of Gyration.	1, 5
<i>SD</i>	Spatial Density.	1, 20, 21
<i>TCF</i>	Total Correlation Function.	1, 19, 20
<i>VACF</i>	Velocity AutoCorrelation Function.	1, 6, 7