Supporting Information

Multiscale Coarse-graining and Structural Correlations: Connections to Liquid State Theory

W.G. Noid, Jhih-Wei Chu, Gary S. Ayton, and Gregory A. Voth

Center for Biophysical Modeling and Simulation and Department of Chemistry, University of Utah, 315 S. 1400 E. Rm 2020, Salt lake City, Utah 84112-0850

1 Introduction

The present document is provided as a supporting information section for the manuscript "Multiscale Coarse-graining and Structural Correlations: Connections to Liquid State Theory" submitted by Noid, Chu, Ayton, and Voth in 2006. This section is essentially a summary of the notes by the author in deriving the results presented in the manuscript. All of the results contained within these notes are presented and briefly derived within the manuscript and its appendix. However, these notes are attached because they provide additional detail into the relevant derivations and results that will hopefully clarify the present work and render it both more accessible and more useful. Although, the notes have been 'tidied up' to a large extent, they are by no means to be considered a polished manuscript. Certain sections of these notes feature slightly different notation from that in the published manuscript, but it is hoped that the differences will not be distracting or confusing. Additional side comments and remarks on generalizing the theory are also included in these notes.

A central result of this work is the derivation of the "normal" force-matching equations for a multi-component system. The derivation elucidates the role of structural correlations in determining a coarse-grained force-field and demonstrates that the force-matching method incorporates critical structural correlations into the coarse-grained force field. Moreover, the work clarifies the general relationship between coarse-graining and the "inverse problem" of simple liquid theory. A generalized YBG equation for coarse-grained systems has been derived. It is proved that for homogeneous isotropic systems evolving under a central pair-wise decomposable force-field this YBG equation may be reduced to a form that is equivalent to the force-matching equations.

This supporting information section provides a detailed presentation on the relationship between the force-matching method and simple liquid theory that was not possible within the length constraints of the published manuscript. The first section provides definitions and derives simple identities that are useful in the present work. The following section derives the "normal" force-matching equations for a multi-component system. The presentation emphasizes the relationship between these equations and the system structure. The final section generalizes the derivation of the YBG equation for a multi-component CG system. These equations are reduced to a one-dimensional integral equation that is equivalent to the FM equations.

2 Preliminaries

The present section provides definitions and derivations of results that will be useful for considering the FM and YBG equations in the following sections. The relevant distribution functions are defined and simplified. The properties of delta functions in spherical coordinates are reviewed and are employed to derive useful expressions for the derivative of the radial distribution function as well as an average of the three-particle distribution function.

2.1 Definitions and symmetries of distribution functions

Cartesian coordinates of particles averaged over are represented by $\underline{x}$. in the following all cg particles of a given type are considered indistinguishable. The labels i_α, j_β, k_γ, will be used to represent arbitrary but distinct particles, and in general will also be referred to with $\underline{r}_1$, $\underline{r}_2$, and $\underline{r}_3$, respectively.

The two- and three- particle density functions are defined:

$$\rho^{(2)}_{\alpha\beta}(\underline{r}_1,\underline{r}_2) = C^{(2)}_{\alpha\beta}\left\langle \delta(\underline{r}_1-\underline{x}_{1_\alpha})\delta(\underline{r}_2-\underline{x}_{2_\beta})\right\rangle \tag{1}$$

$$\rho^{(3)}_{\alpha\beta\gamma}(\underline{r}_1,\underline{r}_2,\underline{r}_3) = C^{(3)}_{\alpha\beta\gamma}\left\langle \delta(\underline{r}_1-\underline{x}_{1_\alpha})\delta(\underline{r}_2-\underline{x}_{2_\beta})\delta(\underline{r}_3-\underline{x}_{3_\gamma})\right\rangle \tag{2}$$

where $C^{(n)}$ is a numerical constant equal to the number of distinct permutations of particles that are equivalent. For example, $C^{(2)}_{\alpha\beta} = \sum\sum_{i_\alpha \neq j_\beta} 1 = (1-\delta_{\alpha\beta})N_\alpha N_\beta + \delta_{\alpha\beta}N_\alpha(N_\alpha-1)$.

In the absence of an external field a given system may be homogeneous (i.e. the one-particle density is a constant independent of location s.t. $\rho^{(1)}_\alpha(\underline{r}_\alpha)=\rho_\alpha$) and isotropic (i.e. the two-particle distribution function is depends only upon the distance between two particles). The two- and three-particle correlation functions are then defined:

$$g^{(2)}_{\alpha\beta}(\underline{r}_1,\underline{r}_2) = \frac{1}{\rho_\alpha\rho_\beta}\rho^{(2)}_{\alpha\beta}(\underline{r}_1,\underline{r}_2) = \frac{C^{(2)}_{\alpha\beta}}{\rho_\alpha\rho_\beta}\left\langle \delta(\underline{r}_1-\underline{x}_{i_\alpha})\delta(\underline{r}_2-\underline{x}_{j_\beta})\right\rangle \tag{3}$$

$$g^{(3)}_{\alpha\beta\gamma}(\underline{r}_1,\underline{r}_2,\underline{r}_3) = \frac{1}{\rho_\alpha\rho_\beta\rho_\gamma}\rho^{(3)}_{\alpha\beta\gamma}(\underline{r}_1,\underline{r}_2,\underline{r}_3) = \frac{C^{(3)}_{\alpha\beta\gamma}}{\rho_\alpha\rho_\beta\rho_\gamma}\left\langle \delta(\underline{r}_1-\underline{x}_{i_\alpha})\delta(\underline{r}_2-\underline{x}_{j_\beta})\delta(\underline{r}_3-\underline{x}_{k_\gamma})\right\rangle . \tag{4}$$

This assumed homogeneity and isotropy may be subtly broken by the periodic boundary conditions for a unit cell. Moreover, this symmetry is also typically broken in complex interfacial systems. Although, the FM equations are independent of such assumptions, the relationship between FM and simple liquid theory is most straight-forward when such conditions are met. In the following analysis these assumptions will be explicitly stated and subsequently employed.

As a result of the assumed homogeneity and isotropy, the two-particle correlation function, $g^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})$, depends only upon the interparticle distance: $r_{i_\alpha j_\beta} = |\underline{r}_{i_\alpha} - \underline{r}_{j_\beta}|$. In other words, because the system may be translated or rotated without changing the system, the distribution function, $g^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})$ is equal to its rotational and translational average, allowing one to define the standard radial distribution function (rdf):

$$g_{\alpha\beta}(r_{21}) = \hat{\mathrm{P}}_{\mathrm{RT}}\, g^{(2)}_{\alpha\beta}(\underline{r}_1,\underline{r}_1+\underline{r}_{21}) = \int_{\Omega_{21}}\frac{1}{4\pi}d\Omega_{21}\int_{V_0}\frac{1}{V_0}d\underline{r}_1\; g^{(2)}_N(\underline{r}_1,\underline{r}_1+\underline{r}_{21}). \tag{5}$$

The preceding equation defines the projection operator $\hat{\mathrm{P}}_{\mathrm{RT}}$ which will be useful in the following.

$$\hat{\mathrm{P}}_{\mathrm{RT}} = \int_{\Omega_{21}}\frac{1}{4\pi}d\Omega_{21}\int_{V_0}\frac{1}{V_0}d\underline{r}_1 = \frac{1}{4\pi V_0}\int_0^\pi d\theta\sin\theta\int_0^{2\pi}d\phi\int_{V_0}d\underline{R} \tag{6}$$

The second equality suppresses the indices from the first equation that are to be implicitly understood.

The preceding definitions and analysis rely *only* upon the translational and rotational invariance of the system. It does not assume any particular form for any interaction. Moreover it applies equally well to cg sites α, β embedded within the atomistic system as long as there exists a canonical transformation that identifies the coarse-grained sites as being explicit Cartesian degrees of freedom in the underlying Hamiltonian. This partitioning is typically possible and is certainly possible for the center of mass definition of cg sites. Probably the partitioning of coordinates may be re-expressed in terms of mapping operators and delta functions, though, this is not considered here. The cg degrees of freedom are considered to be described by Cartesian coordinates. Probably a generalisation to non-Cartesian coordinates is not difficult.

2.2 Delta functions and spherical polar coordinates

The previous subsection reviewed some definitions and properties of the two and three- particle distribution functions. The analysis demonstrated that for a system with translational and rotational symmetry, the two-particle distribution function depends only upon the interparticle distance. This relationship holds for cg sites as well as atomic coordinates. A projection (averaging) operator was defined which reduced the two-particle correlation function (depending upon six variables) to the rdf (depending upon only the interparticle distance). To demonstrate the equivalence of this projection to the conventional definition of the rdf, it is necessary to integrate Dirac delta functions in spherical polar coordinates. Since this mathematical operation is perhaps slightly unfamiliar and will play a role in the following analysis, this subsection reviews certain properties of delta functions in polar coordinates and applies these properties to derive the rdf from the two-particle distribution function.

Assuming that a particle is not at a 'singular' coordinate (one that is multiply-covered by the spherical polar coordinates - more on this momentarily), the Dirac delta function may be expressed:

$$\delta(\underline{r} - \underline{r}') = \frac{1}{r^2 \sin\theta}\delta(r - r')\delta(\theta - \theta')\delta(\phi - \phi'). \tag{7}$$

The prefactor cancels the Jacobian arising from the transformation to spherical polar coordinates and ensures that the integral of the delta function over all space is unity. (This is a fundamental definition of the delta function in three-space.) Singular points are points in space that are multiply-covered by the transformation to spherical polar coordinates, e.g., $\theta = 0$ corresponds to the z-axis, irrespective of ϕ. Treating singular points is entirely analogous, although the normalisation factor has to be corrected for counting points multiple times in the mapping from Cartesian to polar coordinates.

The following identity will be useful for reducing the Yvon-Born-Green (YBG) equation to the FM equations. For all points other than $\underline{r}' = \underline{0}$, the following relation holds:

$$\int d\Omega A(\underline{r})\delta(\underline{r} - \underline{r}') = \frac{1}{r^2}A_{SP}(r, \Omega')\delta(r - r'). \tag{8}$$

Here Ω' denotes the angular degrees of freedom necessary to define $\underline{r}'$. Defining non-singular points requires $\Omega' = \{\theta', \phi'\}$ while points on the z-axis are defined by $\Omega' = \{\theta'\}$ alone. The expression $A_{SP}(r, \Omega')$ denotes the spherical polar representation of the function $A(\underline{r})$ evaluated s.t. the angular components of the function correspond to $\underline{r}'$. This of course may be further integrated s.t.

$$\int dr\ r^2 \int d\Omega\ A(\underline{r})\ \delta(\underline{r} - \underline{r}') = A_{SP}(r', \Omega') = A(\underline{r}'). \tag{9}$$

With this identity in hand, it is straight-forward to show that the above definition for the rdf agrees with

the standard definition. (Again assuming all particles of a given type are identical.)

$$g_{\alpha\beta}(r) = \left(\frac{C^{(2)}_{\alpha\beta}}{4\pi\rho_\alpha\rho_\beta V_0} \right) \frac{1}{r^2} \left\langle \delta(r_{i_\alpha j_\beta} - r) \right\rangle = \left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0} \right) \frac{1}{r^2} \left\langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha} \delta(r_{i_\alpha j_\beta} - r) \right\rangle \tag{10}$$

In the first equality, i_α and j_β label arbitrary but distinct particles of type α and β respectively. In the second equality, the constant $C^{(2)}_{\alpha\beta}$ has been expanded into an equivalent number of terms in the summation, since all the particles of a given type are identical.

Employing the identity (8) for the delta function in spherical polar coordinates, one may similarly prove that:

$$\begin{aligned} &\hat{\mathrm{P}}_{\mathrm{RT}} \int d\Omega_{31} \, (\underline{u}_{21} \cdot \underline{u}_{31}) \, g^{(3)}_{\alpha\beta\gamma}(\underline{r}_1, \underline{r}_2, \underline{r}_3) \\ &= \left(\frac{1}{4\pi V_0} \right) \left(\frac{C^{(3)}_{\alpha\beta\gamma}}{\rho_\alpha\rho_\beta\rho_\gamma} \right) \left(\frac{1}{r_{21} r_{31}} \right)^2 \left\langle \left(\underline{u}_{j_\beta i_\alpha} \cdot \underline{u}_{k_\gamma i_\alpha} \right) \delta\left(r_{k_\gamma i_\alpha} - r_{31} \right) \delta\left(r_{j_\beta i_\alpha} - r_{21} \right) \right\rangle \end{aligned} \tag{11}$$

2.3 The derivative of the rdf: relating to the forces in fm

Previously it has been demonstrated numerically and proved mathematically that the force term in the FM procedure may be related to the derivative of the radial distribution for a one-component system. The generalisation of this result for a multi-component system is straight-forward and presented here. From the outset it should be emphasised that the only assumption necessary for this result is that the total force on the CG sites may be related to the appropriate gradient of an N-particle distribution function. In particular, no assumptions are required regarding the form of the N-particle potential function.

2.3.1 Definitions and useful identities

A thermodynamic (ensemble) average of a function of one variable (or many variables), $A(r)$, is defined according to a distribution function $\rho(\underline{x}^N) = e^{-\beta \mathcal{V}_N(\underline{x}^N)}/Z_N$:

$$\langle A(r) \rangle \equiv \int d\underline{x}^N \rho(\underline{x}^N) A(r, \underline{x}^N). \tag{12}$$

It is assumed that the N-particle distribution function may be used to define a generalised 'force' on the cg sites:

$$\underline{\mathcal{F}}_{i_\alpha}(\underline{x}^N) = - \left(\frac{\partial}{\partial \underline{x}_{i_\alpha}} \mathcal{V}_N(\underline{x}^N) \right)_{\underline{x}^*_{i_\alpha}}. \tag{13}$$

The subscript of the parenthesis indicates that all other degrees of freedom (Cartesian coordinates of the other particles) have been fixed in performing the partial differentiation. It is important to note that this does most emphatically *not* assume any functional form for the cg site interactions, either pairwise, central, or otherwise. This equality only presumes that there exists an N-particle potential energy function that underlies the structure of the cg sites. It should be noted that this 'generalized force' may be effected by the particular distribution function for different ensembles. Nevertheless it will be assumed in the following that $\underline{\mathcal{F}}_i = \underline{F}_{i,\mathrm{AA}}$. This assumption should be investigated further. There may be some details involving the properties of the canonical partitioning of cg and residual degrees of freedom for non-Cartesian coordinates that have not yet been adequately considered also.

As for the one-component system, it will be convenient to define 'sum and difference' coordinates. Cartesian coordinates are denoted by $\underline{x}$, so the definitions relate the Cartesian coordinates to the mean position, $\underline{R}_{ij}$, and the relative positions, $\underline{r}_{ij}$, describing the two cg particles.

$$\underline{r}_{ij} = \underline{x}_i - \underline{x}_j \qquad \underline{x}_i = \underline{R}_{ij} + \frac{1}{2}\underline{r}_{ij} \tag{14}$$

$$\underline{R}_{ij} = \frac{1}{2}(\underline{x}_i + \underline{x}_j) \qquad \underline{x}_j = \underline{R}_{ij} - \frac{1}{2}\underline{r}_{ij} \tag{15}$$

In these equations and in the remainder of this section i implies i_α and j implies j_β, although the explicit dependence on particle type will be supressed. The derivation follows as before and unncessary indices will be suppressed for convenience of the author.

The Jacobian of this transformation is unity: $|\det \partial(\underline{r}_{ij}, \underline{R}_{ij})/\partial(\underline{x}_i, \underline{x}_j)| = 1$. It will be convenient to represent $\underline{r}_{ij}$ in spherical polar coordinates, $(r_{ij}, \theta_{ij}, \phi_{ij})$. The volume element then becomes $d\underline{r}_{ij} = r_{ij}^2 dr_{ij} d\Omega_{ij}$, where $d\Omega_{ij} = \sin\theta_{ij} d\theta_{ij} d\phi_{ij}$. The unit difference vector is represented:

$$\underline{u}_{ij} = \underline{u}_{ij}(\Omega_{ij}) = \underline{e}_x \sin\theta_{ij}\cos\phi_{ij} + \underline{e}_y \sin\theta_{ij}\sin\phi_{ij} + \underline{e}_z \cos\theta_{ij}. \tag{16}$$

This allows the transformation to be written in a convenient way:

$$\underline{x}_i(\underline{R}_{ij}, r_{ij}, \Omega_{ij}) = \underline{R}_{ij} + \frac{1}{2} r_{ij}\underline{u}_{ij}(\Omega_{ij}) \qquad \left(\frac{\partial \underline{x}_i(\underline{R}_{ij}, r_{ij}, \Omega_{ij})}{\partial r_{ij}}\right)_{\underline{R}_{ij}, \Omega_{ij}} = \frac{1}{2}\underline{u}_{ij} \tag{17}$$

$$\underline{x}_j(\underline{R}_{ij}, r_{ij}, \Omega_{ij}) = \underline{R}_{ij} - \frac{1}{2} r_{ij}\underline{u}_{ij}(\Omega_{ij}) \qquad \left(\frac{\partial \underline{x}_j(\underline{R}_{ij}, r_{ij}, \Omega_{ij})}{\partial r_{ij}}\right)_{\underline{R}_{ij}, \Omega_{ij}} = -\frac{1}{2}\underline{u}_{ij}. \tag{18}$$

These transformations enable one to evaluate the desired partial derivative of the density function:

$$\left(\frac{\partial \rho(\underline{x}^N)}{\partial r_{ij}}\right)_{\underline{R}_{ij}, \Omega_{ij}, \underline{x}^{(N-2)}} = \left(\frac{\partial \rho(\underline{x}^N)}{\partial \underline{x}_i}\right)_{\underline{x}_i^*} \cdot \left(\frac{\partial \underline{x}_i}{\partial r_{ij}}\right)_{\underline{R}_{ij}, \Omega_{ij}} + \left(\frac{\partial \rho(\underline{x}^N)}{\partial \underline{x}_j}\right)_{\underline{x}_j^*} \cdot \left(\frac{\partial \underline{x}_j}{\partial r_{ij}}\right)_{\underline{R}_{ij}, \Omega_{ij}} \tag{19}$$

$$= \frac{1}{2}\beta\rho(\underline{x}^N)\left((\underline{\mathcal{F}}_i - \underline{\mathcal{F}}_j)\cdot \underline{u}_{ij}\right). \tag{20}$$

The subscript $\underline{R}_{ij}, \Omega_{ij}, \underline{x}^{(N-2)}$ indicates that the mean coordinate and the orientation between particles i_α and j_β are fixed, as well as the Cartesian coordinates of the other $N-2$ particles in the system. On the rhs the subscript $\underline{x}_i^*$ implies that all Cartesian coordinates other than i_α are held fixed.

The above definition for the rdf, eqn (10), is used:

$$g_{\alpha\beta}(r) = \left(\frac{C^{(2)}_{\alpha\beta}}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\left\langle \delta(r_{i_\alpha j_\beta} - r)\right\rangle = \left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\left\langle \sum_{i_\alpha}\sum_{j_\beta \neq i_\alpha} \delta(r_{i_\alpha j_\beta} - r)\right\rangle$$

where we have also defined an 'unnormalised' distribution function:

$$G_{ij}(r) = \langle \delta(r - r_{ij})\rangle \tag{21}$$

s.t.

$$g_{\alpha\beta}(r) = \left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\sum_{i_\alpha}\sum_{j_\beta \neq i_\alpha} G_{ij}(r). \tag{22}$$

2.3.2 Result

The most general result is

$$\frac{dg_{\alpha\beta}(r)}{dr} = \frac{1}{2k_BT}\left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\left\langle \sum_{i_\alpha}\sum_{j_\beta \neq i_\alpha}\left(\underline{u}_{ij}\cdot(\underline{\mathcal{F}}_i - \underline{\mathcal{F}}_j)\right)\delta(r-r_{ij})\right\rangle. \tag{23}$$

For the specific case $\alpha = \beta$, the result slightly simplifies to:

$$\frac{dg_{\alpha\alpha}(r)}{dr} = \frac{1}{k_BT}\left(\frac{1}{4\pi\rho_\alpha^2 V_0}\right)\frac{1}{r^2}\left\langle \sum_{i_\alpha}\sum_{j_\alpha \neq i_\alpha}\delta(r-r_{ij})\left(\underline{\mathcal{F}}_i\cdot\underline{u}_{ij}\right)\right\rangle. \tag{24}$$

2.3.3 Proof

The following slightly adapts the proof for a one-component system. Consequently it must be appreciated that $\underline{x}^N$ corresponds to the Cartesian coordinates of all N atomic coordinates, including both cg and residual degrees of freedom. In particular $\underline{x}_i$ and $\underline{x}_j$ correspond to the Cartesian coordinates of the i_α and j_β cg sites, respectively.

$$\begin{aligned}
\frac{dG_{ij}(r)}{dr} &= \int d\underline{x}^N \rho(\underline{x}^N)\frac{d}{dr}\delta(r-r_{ij}) && (25)\\
&= (-1)\left(\prod_k \int d\underline{x}_k\right)\int d\underline{x}_i \int d\underline{x}_j \rho(\underline{x}^N)\frac{d}{dr_{ij}}\delta(r-r_{ij}) && (26)\\
&= (-1)\left(\prod_k \int d\underline{x}_k\right)\Phi_{ij}(r,\underline{x}^{(N-2)}) && (27)
\end{aligned}$$

$$\begin{aligned}
\Phi_{ij}(r,\underline{x}^{(N-2)}) &= \int d\underline{x}_i \int d\underline{x}_j \rho(\underline{x}^N)\frac{d}{dr_{ij}}\delta(r-r_{ij}) && (28)\\
&= \int d\underline{R}_{ij}\iiint r_{ij}^2 dr_{ij} d\Omega_{ij}\ \rho(\underline{x}^N)\left(\frac{\partial}{\partial r_{ij}}\delta(r-r_{ij})\right)_{\underline{R}_{ij},\Omega_{ij}} && (29)\\
&= \Xi - T_1 - T_2 && (30)
\end{aligned}$$

After performing integration by parts over the domain of integration, D, (the boundary terms, Ξ, should vanish because presumably the delta function vanishes on the boundary of the system), and making use of the identities above:

$$\begin{aligned}
\Xi &= \int d\underline{R}_{ij}\iint d\Omega_{ij}\ r_{ij}^2\rho(\underline{x}^N)\delta(r-r_{ij})\Big|_{\partial D} = 0 && (31)\\
T_1 &= \int d\underline{R}_{ij}\iiint dr_{ij}d\Omega_{ij}2r_{ij}\rho(\underline{x}^N)\delta(r-r_{ij}) && (32)\\
&= \frac{2}{r}\int d\underline{x}_i\int d\underline{x}_j\rho(\underline{x}^N)\delta(r-r_{ij}) && (33)\\
T_2 &= \int d\underline{R}_{ij}\iiint dr_{ij}d\Omega_{ij}\ r_{ij}^2\delta(r-r_{ij})\left(\frac{\partial\rho(\underline{x}^N)}{\partial r_{ij}}\right)_{\underline{R}_{ij},\Omega_{ij},\underline{x}^{(N-2)}} && (34)\\
&= \frac{1}{2}\beta\int d\underline{x}_i\int d\underline{x}_j\rho(\underline{x}^N)\delta(r-r_{ij})\left((\underline{\mathcal{F}}_i-\underline{\mathcal{F}}_j)\cdot\underline{u}_{ij}\right). && (35)
\end{aligned}$$

From these results it follows that:

$$\frac{dG_{ij}(r)}{dr} = \frac{2}{r}G_{ij}(r) + \frac{1}{2}\beta\left\langle \delta(r-r_{ij})\left((\underline{\mathcal{F}}_i - \underline{\mathcal{F}}_j)\cdot\underline{u}_{ij}\right)\right\rangle. \tag{36}$$

Finally

$$\begin{aligned}
\frac{dg_{\alpha\beta}(r)}{dr} &= \left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{-2}{r^3}\sum_i^{N_0}\sum_{j\neq i} G_{ij}(r) + \left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\sum_i^{N_0}\sum_{j\neq i}\frac{dG_{ij}(r)}{dr} && (37)\\
&= \frac{\beta}{2}\left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\sum_i^{N_0}\sum_{j\neq i}\left\langle \delta(r-r_{ij})\left((\underline{\mathcal{F}}_i - \underline{\mathcal{F}}_j)\cdot\underline{u}_{ij}\right)\right\rangle && (38)\\
&= \frac{1}{2k_BT}\left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\left\langle \sum_i^{N_0}\sum_{j\neq i}\delta(r-r_{ij})\left((\underline{\mathcal{F}}_i - \underline{\mathcal{F}}_j)\cdot\underline{u}_{ij}\right)\right\rangle. && (39)
\end{aligned}$$

Again, it is to be remembered that i and j indicate i_α and j_β, respectively. In the special case that $\alpha = \beta$, then the result reduces since the sum over i and j are identical:

$$\frac{dg_{\alpha\alpha}(r)}{dr} = \frac{1}{k_BT}\left(\frac{1}{4\pi\rho_\alpha^2 V_0}\right)\frac{1}{r^2}\left\langle \sum_i^{N_0}\sum_j^{N_0}\delta(r-r_{ij})\left(\underline{\mathcal{F}}_i\cdot\underline{u}_{ij}\right)\right\rangle. \tag{40}$$

3 Multi-component force-matching (fm) equations

In this section the fm equations are derived for a multi-component system starting from the residual/objective function. The equations are derived under the assumption that the $\alpha\beta$ interaction is equivalent to the $\beta\alpha$ interaction and that the fm force field is both pairwise-decomposable and central. Discrete delta functions are used as a basis. Using the results from section 2, it is demonstrated that the atomistic forces may be eliminated from the fm equations. Finally the resulting matrix equation is passed into the continuum limit to determine an N-component integral equation.

3.1 Multi-component residual

The fm objective for a multi-component system compares the total force on coarse-grained (cg) site, i_α,computed in atomistic simulations, $\underline{F}^I_{i,\mathrm{AA}}$, with the force 'predicted', or used for coarse grain (CG) simulations, $\underline{\Phi}^{I,CG}_{i_\alpha}$ in the cg representation of the atomistic configuration:

$$\Psi = \frac{1}{3N_I(\sum N_\alpha)}\sum_I^{N_I}\sum_\alpha^{N_T}\sum_{i_\alpha}^{N_\alpha}\left|\underline{F}^{I,\mathrm{AA}}_{i_\alpha} - \underline{\Phi}^I_{i_\alpha}\right|^2. \tag{41}$$

In the residual equation (41):

capital letters	I	indicates one particular md configuration of N_I sampled
greek letters	α, β	indicate one particular type of cg sites of N_T present
latin letters	i_α, j_β	indicate one particular atom of a given cg type of N_α present

The two fundamental assumptions underlying the whole fm procedure are that:

1. The total force on each cg site is *assumed* to be derived from a *pairwise-decomposable* force field:

$$\underline{\Phi}^I_{i_\alpha} = \sum_\beta^{N_T} \sum_{j_\beta \neq i_\alpha}^{N_\beta} \underline{\Phi}^I_{i_\alpha j_\beta}(\underline{r}_{i_\alpha}, \underline{r}_{j_\beta}). \tag{42}$$

2. Moreover, this force field is *assumed* to be central

$$\underline{\Phi}^I_{i_\alpha j_\beta}(\underline{r}_{i_\alpha}, \underline{r}_{j_\beta}) = \underline{\Phi}^I_{i_\alpha j_\beta}(r_{i_\alpha j_\beta}). \tag{43}$$

The fm force field defined by these assumptions may be represented in any complete basis set. Discrete delta functions are particularly convenient for this purpose, especially as they provide a direct link to the continuum description. However, in principle the choice of basis is arbitrary and should not alter the resulting conclusions. The motivation for using these discrete delta functions comes from the work of G.S. Ayton. The discrete delta functions are dimensionless $[\delta_D] = 1$ and are related to Ayton's mesoscopic delta function, which have dimensions $[\delta_M] = 1/[\Delta r]$, according to $\delta_D = \Delta r \delta_M$. Using the discrete delta function basis, the fm force field may then be represented:

$$\underline{\Phi}^I_{i_\alpha j_\beta}(r_{i_\alpha j_\beta}) = \sum_d \underline{u}^I_{i_\alpha j_\beta} \phi^{\alpha\beta}_d \delta_D(r^I_{i_\alpha j_\beta} - r_d) \tag{44}$$

where r_d defines the grid points in the fm force table.

From this it can be seen that

$$\underline{\Phi}^I_{i_\alpha} = \sum_\beta \sum_d \phi^{\alpha\beta}_d \underline{\mathcal{G}}^I_{i_\alpha \beta; d}, \tag{45}$$

where

$$\underline{\mathcal{G}}^I_{i_\alpha \beta; d} = \sum_{j_\beta \neq i_\alpha} \underline{u}^I_{i_\alpha j_\beta} \delta_D(r^I_{i_\alpha j_\beta} - r_d) \tag{46}$$

is the sum of the unit vector to i_α from all j_β on a sphere of radius r_d. Then taking care that $\phi^{\alpha\beta}_d = \phi^{\beta\alpha}_d$,

$$\frac{\partial \underline{\Phi}^I_{k_\gamma}}{\partial \phi^{\alpha\beta}_d} = (1 - \delta_{\alpha\beta})\delta_{\gamma\alpha} \underline{\mathcal{G}}^I_{k_\alpha \beta; d} + \delta_{\gamma\beta} \underline{\mathcal{G}}^I_{k_\beta \alpha; d}. \tag{47}$$

The total force on the k_γ cg site depends upon $\phi^{\alpha\beta}_d$ only if either $\gamma = \alpha$ or if $\gamma = \beta$. The first term in the expression arises from forces on k_γ from j_β when $\gamma = \alpha$, while the second arises from forces on k_γ from i_α when $\gamma = \beta$. The first factor prevents over-counting in the situation that $\alpha = \beta$.

Minimizing the residual with respect to $\phi^{\alpha\beta}_d$ one obtains the result:

$$\frac{\partial \Psi}{\partial \phi^{\alpha\beta}_d} = \frac{-2}{3N_I(\sum N_\alpha)} \sum_I \sum_\gamma \sum_{k_\gamma} \left[\underline{F}^I_{k_\gamma} \cdot \frac{\partial \underline{\Phi}^I_{k_\gamma}}{\partial \phi^{\alpha\beta}_d} - \underline{\Phi}^I_{k_\gamma} \cdot \frac{\partial \underline{\Phi}^I_{k_\gamma}}{\partial \phi^{\alpha\beta}_d} \right] \tag{48}$$

$$= \frac{-2}{3N_I(\sum N_\alpha)} \sum_I \sum_\gamma \sum_{k_\gamma} \left[\left(\underline{F}^I_{k_\gamma} - \underline{\Phi}^I_{k_\gamma} \right) \cdot \frac{\partial \underline{\Phi}^I_{k_\gamma}}{\partial \phi^{\alpha\beta}_d} \right] = 0. \tag{49}$$

The linear least squares problem defined by eqn (41) has here been decomposed into a system of linear algebraic equations. In the form presented in the second line the equations may be considered to be a

system of equations $\underline{F}^I_{k_\gamma} = \underline{\Phi}^I_{k_\gamma}$ where $\underline{\Phi}^I_{k_\gamma}$ depends linearly upon the force field according to eqn (45) which may be made "over-determined" by increasing the number of configurations sampled, I, s.t. that there are more equations than force table elements . Izvekov and Voth have implemented this scheme with an additional approximation in trying to solve the associated over-determined system of equations via block-avergaging.

However, while the approach of Izvekov and Voth may be more numerically amenable to solution because of issues with the conditional number of the matrix equation (see Chu et al.), one may alternatively consider the "normal" solution to the system of eqns (49). In fact, the "normal" equations are more amenable to theoretical analysis. (These equations are referred to as "normal" because the resulting matrix equation requires inverting a symmetric, i.e. normal, matrix for which such an inverse exists.) Then *assuming* that the sum of configurations I may be considered equivalent to an ensemble average, s.t.

$$\frac{1}{N_I}\sum_I A^I = \langle A \rangle = \text{Tr } \rho(\underline{x}^N) A(\underline{x}^N), \tag{50}$$

one obtains

$$\left\langle \sum_\gamma \sum_{k_\gamma} \underline{\Phi}^I_{k_\gamma} \cdot \frac{\partial \underline{\Phi}^I_{k_\gamma}}{\partial \phi_d^{\alpha\beta}} \right\rangle = \left\langle \sum_\gamma \sum_{k_\gamma} \underline{F}^I_{k_\gamma} \cdot \frac{\partial \underline{\Phi}^I_{k_\gamma}}{\partial \phi_d^{\alpha\beta}} \right\rangle, \tag{51}$$

in which all force information from atomistic md simulations is included in the average on the right hand side.

Upon employing eqn (47), the normal fm equations may be expressed:

$$\sum_\gamma \sum_{d'} \left[(1-\delta_{\alpha\beta}) \phi^{\alpha\gamma}_{d'} G^{dd'}_{\alpha;\beta\gamma} + \phi^{\beta\gamma}_{d'} G^{dd'}_{\beta;\alpha\gamma} \right] = (1-\delta_{\alpha\beta}) b^d_{\alpha\beta} + b^d_{\beta\alpha}, \tag{52}$$

where

$$b^d_{\alpha\beta} = \left\langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha} \left(\underline{F}_{i_\alpha} \cdot \underline{u}_{i_\alpha j_\beta} \right) \delta_D \left(r_{i_\alpha j_\beta} - r_d \right) \right\rangle \tag{53}$$

$$\begin{aligned} G^{dd'}_{\alpha;\beta\gamma} &= \left\langle \sum_{i_\alpha} \left(\underline{\mathcal{G}}^d_{i_\alpha \beta} \cdot \underline{\mathcal{G}}^{d'}_{i_\alpha \gamma} \right) \right\rangle \\ &= \left\langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha} \sum_{k_\gamma \neq i_\alpha} \left(\underline{u}_{i_\alpha j_\beta} \cdot \underline{u}_{i_\alpha k_\gamma} \right) \delta_D \left(r_{i_\alpha j_\beta} - r_d \right) \delta_D \left(r_{i_\alpha k_\gamma} - r_{d'} \right) \right\rangle. \end{aligned} \tag{54}$$

A couple of comments are in order regarding eqns (52)-(54). All information regarding the forces observed in atomistic md simulations has been repackaged in $b^d_{\alpha\beta}$, which is an average over the total instantaneous force on each α cg site projected onto the unit vector connecting *all* β particles to *each* α particle at a fixed distance d. In this way $b^d_{\alpha\beta}$ correlates instantaneous fluctuations in the density of β particles at a fixed distance from each α particle to the *total* force on each α cg site.

Note that $G^{d'd}_{\alpha;\gamma\beta}$ is symmetric with respect to the exchange $d', \gamma \rightarrow d, \beta$ and contains contributions from both two- and three- particle correlations since, although $i_\alpha \neq j_\beta$ and $i_\alpha \neq k_\gamma$ the possibility $j_\beta = k_\gamma$ has not been excluded from the sum. In fact extracting the contributions from two particle correlations, eqn (54) may be re-expressed:

$$G^{dd'}_{\alpha;\beta\gamma} = \delta_{\beta\gamma}\delta_{d'd} \left(G^{(2)} \right)^d_{\alpha\beta} + \left(G^{(3)} \right)^{dd'}_{\alpha;\beta\gamma}, \tag{55}$$

where the two- and three- particle contributions have been explicitly separated

$$\left(G^{(2)}\right)^{d}_{\alpha\beta} = \left\langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha} \delta_D\left(r_{i_\alpha j_\beta} - r_d\right)\right\rangle \tag{56}$$

$$\left(G^{(3)}\right)^{dd'}_{\alpha;\beta\gamma} = \left\langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha,\, k_\gamma} \sum_{k_\gamma \neq i_\alpha j_\beta} \left(\underline{u}_{i_\alpha j_\beta} \cdot \underline{u}_{i_\alpha k_\gamma}\right) \delta_D\left(r_{i_\alpha j_\beta} - r_d\right) \delta_D\left(r_{i_\alpha k_\gamma} - r_{d'}\right)\right\rangle. \tag{57}$$

Because the two-particle term $\left(G^{(2)}\right)^{d}_{\alpha\beta}$ requires $j_\beta = k_\gamma$, this term can only arise when $\beta = \gamma$ and only contributes to the $d' = d$ term of $G^{dd'}_{\alpha;\beta\gamma}$. Clearly this term can be related to the $\alpha\beta$ radial distribution function.

The three-particle correlation function explicitly describes the effects of a third particle γ on the $\alpha\beta$ distribution. Fixing a central particle it is an average of the $\cos\theta_{ik}$ describing the angle between three particles where the β particle is confined to a sphere of radius r_d and the γ particle is confined to a sphere of radius $r_{d'}$ around the central α particle. (Note that θ_{ik} is defined in terms of three particles - i_α, j_β, and k_γ. The vector $\underline{r}_{i_\alpha j_\beta}$ may be used to define the z-axis for the polar coordinates representing the $\underline{r}_{i_\alpha k_\gamma}$ vector.) When coarse-graining on the molecular level, excluded volume effects prevent the β particle from overlapping the γ particle and consequently there is a depletion zone at very small θ_{ik} which is never sampled in the md simulations. At the same time the angle $\theta_{ik} \approx \pi$ is sampled by the md simulations and the constrained average of this dot product is negative when $d = d'$. This effect will also be relatively pronounced for $d \approx d'$. In addition to a depletion effect from exluded volume there will also be an enhancement effect arising from the solvation shell structure. The dot product involved in eqn (57) may at first appear somewhat artificial. A little analysis reveals that this form arises directly from the assumption that the pair interaction is defined along the vector between two particles. Consequently a third particle can only influence the interaction between two other particles along the vector between the two particles. Moreover, symmetry implies that $\left\langle \sum_{k_\gamma \neq i_\alpha j_\beta} \underline{u}_{i_\alpha k_\gamma} \delta_D\left(r_{i_\alpha j_\beta} - r_d\right) \delta_D\left(r_{i_\alpha k_\gamma} - r_{d'}\right)\right\rangle$ must lie along $\underline{u}_{i_\alpha j_\beta}$. This may be a useful measure of the validity of the central force field approximation. However, although the form of eqn (57) may appear somewhat arbitrary or unique to force-matching, it will be shortly shown that the same functional form appears quite naturally from the YBG equation under the assumption that the underlying site interaction is central.

Employing eqn (55), eqn (52) may be further simplified:

$$(1-\delta_{\alpha\beta})\left[b^{d}_{\alpha\beta} - \phi^{\alpha\beta}_{d}\left(G^{(2)}\right)^{d}_{\alpha\beta}\right] + \left[b^{d}_{\beta\alpha} - \phi^{\beta\alpha}_{d}\left(G^{(2)}\right)^{d}_{\beta\alpha}\right] = (1-\delta_{\alpha\beta})\sum_{\gamma}\sum_{d'} \phi^{\alpha\gamma}_{d'}\left(G^{(3)}\right)^{dd'}_{\alpha;\beta\gamma} + \sum_{\gamma}\sum_{d'} \phi^{\beta\gamma}_{d'}\left(G^{(3)}\right)^{dd'}_{\beta;\alpha\gamma}. \tag{58}$$

(The notation may not be ideal.)

Equation (58) may be considered the final statement of the "normal" fm equations for a multi-component system and requires some discussion. As before all information regarding forces from md simulations are contained within the terms, $b^{d}_{\alpha\beta}$, defined in eqn (53). Clearly $\left(G^{(2)}\right)^{d}_{\alpha\beta} = \left(G^{(2)}\right)^{d}_{\beta\alpha}$ and moreover, eqn (58) was derived under the assumption that $\phi^{\alpha\beta}_{d} = \phi^{\beta\alpha}_{d}$, although that is not explicitly clear from the equation. Note that the lhs depends only upon information regarding two particles fixed at a distance r_d: the total force given a distance between two particles, $b^{d}_{\alpha\beta}$, the pair fm force $\phi^{\alpha\beta}_{d}$, and the two-particle density function, $\left(G^{(2)}\right)^{d}_{\alpha\beta}$. Conversely the rhs depends only upon the force arising from the third particle k_γ at all distances

$r_{d'}$ and contains information about the density of this third particle given the distance between the first two. On both sides of the equation the first term considers the total force on a central particle, i_α, while the second term considers the total force on a central particle j_β. When $\alpha = \beta$ the two terms are identical on both sides of the equation.

3.2 Simplification

In the present section, using the results from section (2), the "normal" fm equations expressed in equation (58) will be re-expressed in a form that is independent of the atomistic force information and that also depends upon the radial distribution function.

In subsection (2.2) it was shown that for any system with translational and rotational symmetry the two-particle correlation is equivalent to the rdf defined in equation (10):

$$g_{\alpha\beta}(r) = \left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\left\langle \sum_{i_\alpha}\sum_{j_\beta \neq i_\alpha} \delta(r_{i_\alpha j_\beta} - r)\right\rangle. \tag{rdf}$$

Moreover in subsection (2.3) it was shown that for the specific case $\alpha = \beta$

$$\frac{dg_{\alpha\alpha}(r)}{dr} = \frac{1}{k_B T}\left(\frac{1}{4\pi\rho_\alpha^2 V_0}\right)\frac{1}{r^2}\left\langle \sum_{i_\alpha}\sum_{j_\alpha \neq i_\alpha} \delta_D(r - r_{ij})\left(\underline{\mathcal{F}}_i \cdot \underline{u}_{ij}\right)\right\rangle, \tag{dgdr-aa}$$

while for $\alpha \neq \beta$

$$\frac{dg_{\alpha\beta}(r)}{dr} = \frac{1}{2k_B T}\left(\frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\right)\frac{1}{r^2}\left\langle \sum_{i_\alpha}\sum_{j_\beta \neq i_\alpha} \left(\underline{u}_{ij}\cdot(\underline{\mathcal{F}}_i - \underline{\mathcal{F}}_j)\right)\delta_D(r - r_{ij})\right\rangle. \tag{dgdr-ab}$$

Recall that the above expressions are based on the assumption that $\underline{\mathcal{F}}_i = -\partial\mathcal{V}_N(\underline{x}^N)/\partial\underline{x}_i$, where $\mathcal{V}_N(\underline{x}^N) = -k_B T\log\rho(\underline{x}^N)$ and $\rho(\underline{x}^N)$ is the N particle distribution function underlying the rdf.

Consider the fm eqn (58) first for the case $\alpha = \beta$.

$$\left[b_{\alpha\alpha}^d - \phi_d^{\alpha\alpha}\left(G^{(2)}\right)_{\alpha\alpha}^d\right] = \sum_\gamma\sum_{d'}\phi_{d'}^{\alpha\gamma}\left(G^{(3)}\right)_{\alpha;\alpha\gamma}^{dd'} \tag{59}$$

These equations are identical to those obtained earlier for the one-component case, with the exception that the third particle, k_γ, is not necessarily of type α. Multiplying both sides by the expression, $1/\left(4\pi r_d^2\rho_\alpha\rho_\alpha V_0\right)$ and recalling the definitions in eqns (53), (56), and (57), one obtains:

$$\left[k_B T\left.\frac{dg_{\alpha\alpha}}{dr}\right|_d - \left.\phi_d^{\alpha\alpha}g_{\alpha\alpha}\right|_d\right] = \sum_\gamma\sum_{d'}\phi_d^{\alpha\gamma}\left(M^{(\mathrm{FM})}\right)_{\alpha;\alpha\gamma}^{dd'}, \tag{60}$$

where the lhs contains discrete representations of the rdf and its derivative. The rhs has been expressed in terms of a force-matching three-body kernel:

$$\begin{aligned}\left(M^{(\mathrm{FM})}\right)_{\alpha;\alpha\gamma}^{dd'} &= \frac{1}{4\pi\rho_\alpha^2 V_0}\frac{1}{r_d^2}\left(G^{(3)}\right)_{\alpha;\alpha\gamma}^{dd'} \\ &= \frac{1}{4\pi\rho_\alpha^2 V_0}\frac{1}{r_d^2}\left\langle \sum_{i_\alpha}\sum_{j_\alpha \neq i_\alpha,\, k_\gamma}\sum_{k_\gamma \neq i_\alpha, j_\alpha}\left(\underline{u}_{i_\alpha j_\alpha}\cdot\underline{u}_{i_\alpha k_\gamma}\right)\delta_D\left(r_{i_\alpha j_\alpha} - r_d\right)\delta_D\left(r_{i_\alpha k_\gamma} - r_{d'}\right)\right\rangle\end{aligned} \tag{61}$$

Consider the $\alpha \neq \beta$ case:

$$\left(b^d_{\alpha\beta} + b^d_{\beta\alpha}\right) - 2\phi^{\alpha\beta}_d \left(G^{(2)}\right)^d_{\alpha\beta} = \sum_\gamma \sum_{d'} \phi^{\alpha\gamma}_{d'} \left(G^{(3)}\right)^{dd'}_{\alpha;\beta\gamma} + \sum_\gamma \sum_{d'} \phi^{\beta\gamma}_{d'} \left(G^{(3)}\right)^{dd'}_{\beta;\alpha\gamma}. \tag{62}$$

Multiplying both sides by $1/\left(4\pi r_d^2 \rho_\alpha \rho_\beta V_0\right)$ and employing the relations for the rdf and its derivative, this result may be simplified:

$$2\left[k_B T \left.\frac{dg_{\alpha\beta}}{dr}\right|_d - \phi^{\alpha\beta}_d \left. g_{\alpha\beta}\right|_d\right] = \sum_\gamma \sum_{d'} \phi^{\alpha\gamma}_{d'} \left(M^{(\mathrm{FM})}\right)^{dd'}_{\alpha;\beta\gamma} + \sum_\gamma \sum_{d'} \phi^{\beta\gamma}_{d'} \left(M^{(\mathrm{FM})}\right)^{dd'}_{\beta;\alpha\gamma} \tag{63}$$

where

$$\begin{aligned}
\left(M^{(\mathrm{FM})}\right)^{dd'}_{\alpha;\beta\gamma} &= \frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\frac{1}{r_d^2}\left(G^{(3)}\right)^{dd'}_{\alpha;\beta\gamma} \\
&= \frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\frac{1}{r_d^2}\left\langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha, k_\gamma} \sum_{k_\gamma \neq i_\alpha, j_\beta} \left(\underline{u}_{i_\alpha j_\beta} \cdot \underline{u}_{i_\alpha k_\gamma}\right) \delta_D\left(r_{i_\alpha j_\beta} - r_d\right) \delta_D\left(r_{i_\alpha k_\gamma} - r_{d'}\right)\right\rangle
\end{aligned} \tag{64}$$

3.3 Continuum limit

A particularly convenient feature of the discrete delta function basis is that it allows a straight-forward method for passing from a discrete to a continuum representation. This involves changing the discrete delta functions to dirac delta functions, $\delta_D(r_d - r') = \Delta r' \delta_{r_d,r'}/\Delta r' \to dr'\delta(r - r')$, and sums over d' into integrals over r', $\sum_{d'} \to \int$. The factor $\delta_{r_d,r'}$ denotes a Kronecker delta corresponding to whether r_d and r' are in the same bin of the force table. This transformation is performed here and the discrete algebraic equations derived for force matching are transformed into a linear one-dimensional integral equation. It should be noted that since each side of equation (63) involves one more factor of $\Delta r'$ than there are summations/integrals, in going to the continuum limit, both sides should be divided by $\Delta r'$.

Comparing the results for the n-component fm equations in either case, it can be seen that:

$$\left[k_B T \left.\frac{dg_{\alpha\beta}}{dr}\right|_d - \phi^{\alpha\beta}_d \left. g_{\alpha\beta}\right|_d\right] = \sum_\gamma \sum_{d'} \frac{1}{2}\left[\phi^{\alpha\gamma}_{d'} \left(M^{(\mathrm{FM})}\right)^{dd'}_{\alpha;\beta\gamma} + \phi^{\beta\gamma}_{d'} \left(M^{(\mathrm{FM})}\right)^{dd'}_{\beta;\alpha\gamma}\right], \tag{65}$$

where

$$\left(M^{(\mathrm{FM})}\right)^{dd'}_{\alpha;\beta\gamma} = \frac{1}{4\pi\rho_\alpha^2 V_0}\frac{1}{r_d^2}\left\langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha, k_\gamma} \sum_{k_\gamma \neq i_\alpha, j_\beta} \left(\underline{u}_{i_\alpha j_\beta} \cdot \underline{u}_{i_\alpha k_\gamma}\right) \delta_D\left(r_{i_\alpha j_\beta} - r_d\right) \delta_D\left(r_{i_\alpha k_\gamma} - r_{d'}\right)\right\rangle. \tag{66}$$

In the continuum limit these equations become

$$\left[k_B T \frac{d}{dr} - \phi^{\alpha\beta}(r)\right] g_{\alpha\beta}(r) = \sum_\gamma \int dr' \frac{1}{2}\left[\phi^{\alpha\gamma}(r') M^{(\mathrm{FM})}_{\alpha;\beta\gamma}(r, r') + \phi^{\beta\gamma}(r') M^{(\mathrm{FM})}_{\beta;\alpha\gamma}(r, r')\right], \tag{67}$$

where

$$M^{(\mathrm{FM})}_{\alpha;\beta\gamma}(r, r') = \frac{1}{4\pi\rho_\alpha\rho_\beta V_0}\frac{1}{r^2}\left\langle \sum_{i_\alpha} \sum_{j_\beta \neq i_\alpha, k_\gamma} \sum_{k_\gamma \neq i_\alpha, j_\beta} \left(\underline{u}_{i_\alpha j_\beta} \cdot \underline{u}_{i_\alpha k_\gamma}\right) \delta_D\left(r_{i_\alpha j_\beta} - r\right) \delta_D\left(r_{i_\alpha k_\gamma} - r'\right)\right\rangle. \tag{68}$$

4 Multi-component ybg equation for cg systems

In the present section the theory of the Yvon-Born-Green equation is generalized to consider the distribution functions for systems described by multiple types of coarse- grained sites. In the first subsection, the generalized BBGKY equation for n=2 is derived and then reduced to the YBG equation under the assumptions that the system is homogeneous and that the potential energy is pair-wise decomposable. In the following subsection, this result is then further simplified under the assumptions that the system is homogeneous and that all interactions between coarse-grained sites are directed along the vector between sites and only depend upon the inter-site distance. With these assumptions, the generalized YBG equation may be simplified to a form that is equivalent to the FM equations.

4.1 Deriving the Yvon-Born-Green equation for a multi-component cg system

The derivation of the Yvon-Born-Green (YBG) equations starts from the assumption of a Hamiltonian with a pairwise-decomposable potential energy describing the CG system. In particular, from this perspective there is no consideration of the residual degrees of freedom. It is simply assumed that the observed cg structure arose from a CG Hamiltonian. In particular, this Hamiltonian is assumed to take the general form:

$$H_N^{(\mathrm{CG})} = \sum_\alpha \sum_{i_\alpha} \frac{|\underline{p}_{i_\alpha}|^2}{2m_\alpha} + V_N^{(CG)}(\underline{r}^{N_{CG}}) + \sum_\alpha \sum_{i_\alpha} U_\alpha(\underline{r}_{i_\alpha}). \tag{69}$$

The kinetic energy term can probably be generalised to allow for momentum-coupling, though i have not worked it out. What may be important is that there is no momentum-coordinate coupling. The CG potential is represented as a general N-body potential, but it will ultimately be necessary to assume a pairwise decomposable force field. An external field has been added to the Hamiltonian for generality in deriving the YBG equation. This adds little complexity to the derivation. However, in order to reduce the YBG equation to the FM equations it will be necessary in the end to assume that $U_\alpha = 0$.

The N-particle phase space distribution function describing the CG sites, $f^{(N)}(\underline{r}^N, \underline{p}^N)$, evolves according to the Liouville equation, which may be expressed:

$$\begin{aligned} \frac{\partial f^{(N)}}{\partial t} &= \sum_\gamma \sum_{k_\gamma} \left[-\left(\underline{F}_{k_\gamma} + \underline{X}_{k_\gamma} \right) \cdot \frac{\partial}{\partial \underline{p}_{k_\gamma}} - \frac{1}{m_\gamma} \underline{p}_{k_\gamma} \cdot \frac{\partial}{\partial \underline{r}_{k_\gamma}} \right] f^{(N)} && (70) \\ &= (-1) \sum_\gamma \sum_{k_\gamma} \left[\hat{\mathcal{F}}_{k_\gamma} + \hat{\mathcal{X}}_{k_\gamma} + \hat{\mathcal{W}}_{k_\gamma} \right] f^{(N)}, && (71) \end{aligned}$$

where $\hat{\mathcal{F}}_{k_\gamma} = \underline{F}_{k_\gamma} \cdot \partial / \partial \underline{p}_{k_\gamma}$, $\hat{\mathcal{X}}_{k_\gamma} = \underline{X}_{k_\gamma} \cdot \partial / \partial \underline{p}_{k_\gamma}$, and $\hat{\mathcal{W}}_{k_\gamma} = (1/m_\gamma) \underline{p}_{k_\gamma} \cdot \partial / \partial \underline{r}_{k_\gamma}$. $\underline{X}_{k_\gamma}$ is the force on CG particle k_γ from the external field.

Consider an operator

$$\hat{I}_{\alpha\beta} = C_{\alpha\beta}^{(2)} \int \left(d\Gamma_{i_\alpha} d\Gamma_{j_\beta} \right)^*, \tag{72}$$

which is an integral over the phase space of all particles *except* i_α and j_β and which is multiplied by the combinatorial factor defined earlier. This operator is defined such that $\hat{I}_{\alpha\beta} f^{(N)}(\underline{r}^N, \underline{p}^N) = f_{\alpha\beta}^{(2)}(\Gamma_{i_\alpha}, \Gamma_{j_\beta})$. Then applying this operator to equation (70), one obtains

$$\left[\frac{\partial}{\partial t} + \left(\hat{\mathcal{W}}_{i_\alpha} + \hat{\mathcal{W}}_{j_\beta} \right) + \left(\hat{\mathcal{X}}_{i_\alpha} + \hat{\mathcal{X}}_{j_\beta} \right) \right] f_{\alpha\beta}^{(2)}(\Gamma_{i_\alpha}, \Gamma_{j_\beta}) = -\hat{I}_{\alpha\beta} \left(\hat{\mathcal{F}}_{i_\alpha} + \hat{\mathcal{F}}_{j_\beta} \right) f^{(N)}(\Gamma^N). \tag{73}$$

Note that i_α and j_β describe arbitrary but distinct i, j. All particles of a given type are indistinguishable but the labels α, β are significant. The lhs of this equality is trivial to derive but the rhs is slightly less trivial to show and is a result of the identity:

$$\int d\Gamma_{k_\gamma} \hat{\mathcal{F}}_{k_\gamma} f^{(N)} = 0, \tag{74}$$

which is true only because (i) $f^{(N)} = 0$ for $|\underline{p}_{k_\gamma}| \to \pm\infty$ and (ii) it has been *assumed* that $\partial \underline{F}_{i_\alpha}(\underline{r}^{(N)})/\partial \underline{p}_{j_\beta} = 0$ for all i_α, j_β.

Now if it is *assumed* that the force field is pairwise additive, i.e. $\underline{F}_{i_\alpha} = \sum_\beta \sum_{j_\beta} \underline{F}_{i_\alpha j_\beta}(\underline{r}_{i_\alpha}, \underline{r}_{j_\beta})$. then it follows that

$$\hat{I}_{\alpha\beta}\left(\hat{\mathcal{F}}_{i_\alpha} f^{(N)}(\Gamma^{(N)})\right) = \hat{\mathcal{F}}_{i_\alpha j_\beta} f^{(2)}(\Gamma_{i_\alpha}, \Gamma_{j_\beta}) + \sum_\gamma \int d\Gamma_{k_\gamma} \hat{\mathcal{F}}_{i_\alpha k_\gamma} f^{(3)}(\Gamma_{i_\alpha}, \Gamma_{j_\beta}, \Gamma_{k_\gamma}), \tag{75}$$

where the $\alpha\beta$ superscript indicates the type of interaction. Employing this identity in the kinetic equation (73) provides the result:

$$\begin{aligned}\left[\frac{\partial}{\partial t} + \left(\hat{\mathcal{W}}_{i_\alpha} + \hat{\mathcal{W}}_{j_\beta}\right) + \left(\hat{\mathcal{X}}_{i_\alpha} + \hat{\mathcal{X}}_{j_\beta}\right) + \left(\hat{\mathcal{F}}_{i_\alpha j_\beta} + \hat{\mathcal{F}}_{j_\beta i_\alpha}\right)\right] f^{(2)}_{\alpha\beta}(\Gamma_{i_\alpha}, \Gamma_{j_\beta}) \\ = -\sum_\gamma \int d\Gamma_{k_\gamma} \left(\hat{\mathcal{F}}_{i_\alpha k_\gamma} + \hat{\mathcal{F}}_{j_\beta k_\gamma}\right) f^{(3)}_{\alpha\beta\gamma}(\Gamma_{i_\alpha}, \Gamma_{j_\beta}, \Gamma_{k_\gamma}).\end{aligned} \tag{76}$$

This equation is a generalisation of the BBGKY hierarchy of kinetic equations for $n = 2$ to a multicomponent system. As before α, β, γ are all arbitrary and may or may not be distinct. The labels i, j, k are also arbitrary but are assumed to be distinct, i.e. $j_\beta \neq i_\alpha, k_\gamma \neq i_\alpha, j_\beta$.

At this point, it is further *assumed* that the two-particle phase space distribution functions are at equilibrium. Consequently

$$\frac{\partial}{\partial t} f^{(2)}_{\alpha\beta}(\Gamma_{i_\alpha}, \Gamma_{j_\beta}) = 0 \tag{77}$$

and, moreover, because the kinetic energy term is both separable and diagonal, the n-particle distribution function assumes the form:

$$f^{(n)}_{\alpha\beta}(\Gamma^{(n)}) = \mathcal{P}^{(n)}(\underline{p}^{(n)}) \rho^{(n)}(\underline{r}^{(n)}), \tag{78}$$

where $\rho^{(n)}(\underline{r}^{(n)})$ is the n-particle distribution function defined earlier and

$$\mathcal{P}^{(n)}(\underline{p}^{(n)}) = \left(\frac{1}{2\pi m k_B T}\right)^{3n/2} \exp\left[-\beta \sum_{i=1}^{n} |\underline{p}_i|^2 / 2m_i\right]. \tag{79}$$

(- more precisely the mass factor in the prefactor should be more carefully defined -) s.t.

$$\frac{\partial}{\partial \underline{p}_{i_\alpha}} \mathcal{P}^{(2)}_{\alpha\beta}(\underline{p}_{i_\alpha}, \underline{p}_{j_\beta}) = -\frac{\beta}{m_\alpha} \underline{p}_{i_\alpha} \mathcal{P}^{(2)}_{\alpha\beta}(\underline{p}_{i_\alpha}, \underline{p}_{j_\beta}) \tag{80}$$

and moreover

$$\int d\underline{p}_{k_\gamma} \mathcal{P}^{(3)}_{\alpha\beta\gamma}(\underline{p}_{i_\alpha}, \underline{p}_{j_\beta}, \underline{p}_{k_\gamma}) = \mathcal{P}^{(2)}_{\alpha\beta}(\underline{p}_{i_\alpha}, \underline{p}_{j_\beta}). \tag{81}$$

With these results the kinetic term may be cancelled from both sides of equation (76) which may then be re-organised:

$$\frac{\underline{p}_{i_\alpha}}{m_\alpha}\cdot\left[\left(\frac{\partial}{\partial\underline{r}_{i_\alpha}}-\beta\left(\underline{X}_{i_\alpha}+\underline{F}_{i_\alpha j_\beta}\right)\right)\rho^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})-\sum_\gamma\beta\int d\underline{r}_{k_\gamma}\underline{F}_{i_\alpha k_\gamma}\rho^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma})\right]$$
$$+\frac{\underline{p}_{j_\beta}}{m_\beta}\cdot\left[\left(\frac{\partial}{\partial\underline{r}_{j_\beta}}-\beta\left(\underline{X}_{j_\beta}+\underline{F}_{j_\beta i_\alpha}\right)\right)\rho^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})-\sum_\gamma\beta\int d\underline{r}_{k_\gamma}\underline{F}_{j_\beta k_\gamma}\rho^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma})\right]=0. \quad (82)$$

Since this equation holds $\forall\underline{p}_{i_\alpha},\underline{p}_{j_\beta}$, the one equality equation (82) implies two independent sets of equations:

$$\left(k_BT\frac{\partial}{\partial\underline{r}_{i_\alpha}}-\left(\underline{X}_{i_\alpha}+\underline{F}_{i_\alpha j_\beta}\right)\right)\rho^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})=\sum_\gamma\int d\underline{r}_{k_\gamma}\underline{F}_{i_\alpha k_\gamma}\rho^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma}) \quad (83)$$

$$\left(k_BT\frac{\partial}{\partial\underline{r}_{j_\beta}}-\left(\underline{X}_{j_\beta}+\underline{F}_{j_\beta i_\alpha}\right)\right)\rho^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})=\sum_\gamma\int d\underline{r}_{k_\gamma}\underline{F}_{j_\beta k_\gamma}\rho^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma}). \quad (84)$$

This result is a generalisation of the YBG equation for $n=2$ to a multi-component system. The only assumptions employed in deriving this equality are that there exists a CG Hamiltonian of the form given in equation (69) defining an equilibrium distribution function which gave rise to the density functions $\rho^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})$ and $\rho^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma})$ and that moreover the CG potential is pairwise decomposable.

4.2 Simplifying the ybg equation

Equations (83) and (84) are equivalent. In fact in the case of a central force field, they both reduce to identical one-dimensional integral equations. Therefore w.l.o.g equation (83) will be analysed.

In the absence of an external field $\underline{X}=\underline{0}$, equation (83) reduces to

$$\left(k_BT\frac{\partial}{\partial\underline{r}_{i_\alpha}}-\underline{F}_{i_\alpha j_\beta}\right)g^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})=\sum_\gamma\rho_\gamma\int d\underline{r}_{k_\gamma}\underline{F}_{i_\alpha k_\gamma}g^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma}). \quad (85)$$

As discussed above, in the absence of an external field (and under the assumptions of homogeneity and isotropy) the two-particle distribution function depends only upon the inter-particle distance: $g^{(2)}_N(\underline{r}_1,\underline{r}_2)=\psi_{g2}(|\underline{r}_2-\underline{r}_1|)=\psi_{g2}(r_{21})$. Similarly, by symmetry it can be seen that the three particle distribution function depends only upon $r_{21},r_{31},\theta_{31}$, since r_{32} may be determined from r_{31},θ_{31}, where θ_{31} is the angle formed between particles 2,1, and 3, with 1 at the vertex: $g^{(3)}_N(\underline{r}_1,\underline{r}_2,\underline{r}_3)=\psi_{g3}(r_{21},r_{31},\theta_{31})$. These symmetries imply that:

$$\left(\frac{\partial}{\partial\underline{r}_{i_\alpha}}g^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})\right)=\left(\frac{\partial}{\partial\underline{r}_{i_\alpha j_\beta}}\psi_{g2}(r_{i_\alpha j_\beta})\right)=\underline{u}_{i_\alpha j_\beta}\frac{d}{dr_{i_\alpha j_\beta}}g^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta}). \quad (86)$$

To further simplify the YBG equation to an equation in one variable, it is *assumed* that the force depends only upon the interparticle distance, $r^{\alpha\beta}_{ij}$:

$$\underline{F}_{i_\alpha j_\beta}=\underline{u}_{i_\alpha j_\beta}f^{\alpha\beta}(r_{i_\alpha j_\beta}). \quad (87)$$

With this key assumption, the YBG equation (83) may be reduced to

$$\underline{u}_{i_\alpha j_\beta}\left(k_BT\frac{d}{dr_{i_\alpha j_\beta}}-f^{\alpha\beta}(r_{i_\alpha j_\beta})\right)g^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})=\sum_\gamma\rho_\gamma\int d\underline{r}_{k_\gamma i_\alpha}\underline{u}_{i_\alpha k_\gamma}f^{\alpha\gamma}(r_{i_\alpha k_\gamma})g^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma}). \quad (88)$$

Thus the dot product arises quite naturally in the YBG theory, just as in the FM equations: the effect of a third particle on the interaction between two other particles can only exert influence to the force along the vector between the first two particles. From homogeneity/symmetry, effects along any other vector must vanish in the average.

$$\left(k_BT\frac{d}{dr_{i_\alpha j_\beta}}-f^{\alpha\beta}(r_{i_\alpha j_\beta})\right)g^{(2)}_{\alpha\beta}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta})=\sum_\gamma\rho_\gamma\int d\underline{r}_{k_\gamma i_\alpha}\;\; f^{\alpha\gamma}(r_{k_\gamma i_\alpha})\left(\underline{u}_{i_\alpha j_\beta}\cdot\underline{u}_{i_\alpha k_\gamma}\right)g^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma}). \tag{89}$$

Note that although the YBG equation appears to depend explicitly upon $\underline{r}_{i_\alpha},\underline{r}_{j_\beta}$, it is clear from the above that the left, and thus also the right, hand side of the equation depends only upon $r_{i_\alpha j_\beta}$. Applying the projection operator, $\hat{\mathrm{P}}_{\mathrm{RT}}$, defined earlier and taking advantage of the definition of the rdf:

$$\begin{aligned}\left(k_BT\frac{d}{dr_{i_\alpha j_\beta}}-f^{\alpha\beta}(r_{i_\alpha j_\beta})\right)g_{\alpha\beta}(r_{i_\alpha j_\beta}) &= \sum_\gamma\rho_\gamma\int d\underline{r}_{k_\gamma i_\alpha}\;\; f^{\alpha\gamma}(r_{k_\gamma i_\alpha})\hat{\mathrm{P}}_{\mathrm{RT}}\left[\left(\underline{u}_{i_\alpha j_\beta}\cdot\underline{u}_{i_\alpha k_\gamma}\right),g^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma})\right] &(90)\\ &= \sum_\gamma\int dr_{k_\gamma i_\alpha}\;\; f(r_{k_\gamma i_\alpha})\mathrm{M}^{(YBG)}_{\alpha;\beta\gamma}(r_{i_\alpha j_\beta},r_{k_\gamma i_\alpha}), &(91)\end{aligned}$$

where

$$\begin{aligned}\mathrm{M}^{(YBG)}_{\alpha;\beta\gamma}(r_{i_\alpha j_\beta},r_{i_\alpha k_\gamma}) &= \rho_\gamma\left(r_{k_\gamma i_\alpha}\right)^2\int d\Omega_{k_\gamma i_\alpha}\hat{\mathrm{P}}_{\mathrm{RT}}\left[\left(\underline{u}_{i_\alpha j_\beta}\cdot\underline{u}_{i_\alpha k_\gamma}\right)g^{(3)}_{\alpha\beta\gamma}(\underline{r}_{i_\alpha},\underline{r}_{j_\beta},\underline{r}_{k_\gamma})\right] &(92)\\ &= \left(\frac{C^{(3)}_{\alpha\beta\gamma}}{4\pi V_0\rho_\alpha\rho_\beta}\right)\left(\frac{1}{r_{i_\alpha j_\beta}}\right)^2\left\langle\left(\underline{u}'_{i_\alpha j_\beta}\cdot\underline{u}'_{i_\alpha k_\gamma}\right)\delta\left(r'_{k_\gamma i_\alpha}-r_{k_\gamma i_\alpha}\right)\delta\left(r'_{i_\alpha j_\beta}-r_{i_\alpha j_\beta}\right)\right\rangle &(93)\\ &= \mathrm{M}^{(FM)}_{\alpha;\beta\gamma}(r_{i_\alpha j_\beta},r_{i_\alpha k_\gamma}), &(94)\\ & &(95)\end{aligned}$$

i.e.

$$\mathrm{M}^{(YBG)}_{\alpha;\beta\gamma}(r_{12},r_{13})=\mathrm{M}^{(FM)}_{\alpha;\beta\gamma}(r_{12},r_{13})=\mathrm{M}_{\alpha;\beta\gamma}(r_{12},r_{13}). \tag{96}$$

It is critical to appreciate that primed quantities are measured in md frames and averaged over according to the ensemble average defined by the angular brackets.

Upon performing a similar analysis on the equation for the $\beta\alpha$ interaction and appreciating the symmetry of the pair potential and rdf, one obtains two equations:

$$\begin{aligned}\left(k_BT\frac{d}{dr_{12}}-f^{\alpha\beta}(r_{12})\right)g_{\alpha\beta}(r_{12}) &= \sum_\gamma\int dr_{13}\;\; f^{\alpha\gamma}(r_{13})\mathrm{M}_{\alpha;\beta\gamma}(r_{12},r_{13}), &(97)\\ \left(k_BT\frac{d}{dr_{12}}-f^{\alpha\beta}(r_{12})\right)g_{\alpha\beta}(r_{12}) &= \sum_\gamma\int dr_{23}\;\; f^{\beta\gamma}(r_{23})\mathrm{M}_{\beta;\alpha\gamma}(r_{12},r_{23}). &(98)\end{aligned}$$

Obviously the lhs of both equations are identical. This is just a result of the symmetry arising from assuming a central potential. It is less obvious that the two expressions on the rhs of this system are identical in general, although it is obvious that $rhs_1=rhs_2$ when $\alpha=\beta$. Nevertheless, the two rhs's must be identical as well.

Consequently the two equalities may be averaged without loss of generality giving rise to the final expression which is identical to the multi-component FM equations:

$$\left[k_B T \frac{d}{dr} - \phi^{\alpha\beta}(r)\right] g_{\alpha\beta}(r) = \sum_{\gamma} \int dr' \frac{1}{2} \left[f^{\alpha\gamma}(r') \mathrm{M}_{\alpha;\beta\gamma}(r, r') + f^{\beta\gamma}(r') \mathrm{M}_{\beta;\alpha\gamma}(r, r')\right]. \tag{99}$$