ORIGINAL PAPER

Partial nearest neighbor PDF for multi-component material systems

Uduzei F. Edgal

Abstract   We develop general expressions for partial nearest neighbor probability density functions (PNNPDF's) for equilibrium multi-component systems, valid for arbitrary partial densities, temperatures, and interaction potentials; thus providing an alternative means of describing structure at the microscopic (atomic scale) level for multi-component material systems. This paper thus complements an earlier paper (U.F. Edgal and D.L. Huber, J. Phys. Chem. B **108**, 13777–13788 2004) in the analytical investigation of the classical statistical thermodynamics of multi-component systems. The connection between PNNPDF's and the commonly employed partial m-body distribution functions is detailed. Results for PNNPDF's and partial m-body distribution functions applicable for the poisson-distributed multi-component system and the low density binary mixture of hard spheres are provided. The statistical geometry of the systems is further studied through a brief investigation of particle clustering. A major hallmark of the above investigation involves the several multiple integrals and multiple sums encountered, that were quite formidable to perform, even in the absence of particle interactions.

Keywords   Partial nearest neighbor probability density function · Partial m-tuplet distribution function · General multi-component point process · Ordinary multi-component point process · Statistical geometry · Singlet cluster probability

U. F. Edgal (✉)
Department of Engineering Technology, Old Dominion University, 5115 Hampton Blvd., Norfolk, VA 23529, USA
e-mail: uedgal@odu.edu

1 Introduction

This paper addresses the microstructure of equilibrium multi-component systems providing results applicable for a variety of materials. Microstructure (or structure at the atomic or "smallest" scale of relevance) is of fundamental importance in the fields of condensed matter physics/chemistry and materials science/engineering, as it is strongly connected with various other characterizations of material systems and processes. For instance, translation symmetry at the atomic scale leads to the Bloch theorem, Brillouin zone analysis and thus the well defined electron–phonon band structures and associated properties of crystals [1]. Structure at the micro- and nano-scales are studied for self-assembling systems for the development of micro-electromechanical systems (MEMS) and nano-electromechanical systems (NEMS) with desired properties [2]. This has led to the fabrication of high quality MEMS and NEMS sensors and actuators, self-assembling ferromagnetic particles for high density magnetic recording [3], etc. The microstructure of particles in superconductors is known to affect the super-conducting transition temperature [1,4]; various 2-D and 3-D systems are known to exhibit micro-structural precursor at the point of phase transition [1,5]; etc. Also, it was shown in some recent papers by the author [6] that accurate knowledge of the distributions governing microstructure of equilibrium material systems implies accurate knowledge of the free energy and other statistical thermodynamic properties of equilibrium material systems (classical and quantum). Hence this paper complements the earlier paper of Edgal and Huber [6b] in the analytical investigation of the classical statistical thermodynamics of multi-component material systems. Up until now, the most detailed theories and most precise experiments on the condensed state of matter have been carried out in perfect crystals. However, most naturally occurring substances and technologically important materials are not a collection of solids with perfect symmetry. Hence it is of paramount importance that powerful analytic schemes be developed to adequately treat a vast array of complex materials from liquids, glasses, and liquid crystals to quasi-crystals, alloys, polymers, and much more.

The present paper employs the distribution of the configuration of nearest neighbors of some arbitrarily chosen point within a given equilibrium material system of interest to describe the microstructure of the system. The distribution is referred to as the nearest neighbor probability density function (NNPDF) for pure systems; while for multi-component systems, they are referred to as partial NNPDF's (PNNPDF's). Although NNPDF's have long been known in the mathematics literature [7], it is only in recent times that this method for describing structure in arbitrary systems has begun to gain significant recognition in the physical sciences. In 1956, Reiss [8] used the first nearest neighbor distribution function (ordinarily used for single component systems) to analyze a two-component system after suitable modifications. The distribution functions allowed for the formulation of the degree of "pairing" between oppositely charged ions in a material medium. (See Wiley [8c] for a broad review and other references employing similar model). Several decades later, Edgal and Huber [6b] developed PNNPDF's (up to an arbitrary m-th nearest neighbors) for a poisson non-interaction multi-component fluid system. NNPDF's have also been developed for single component systems [9], but have usually been restricted either to poisson distributed systems, or the NNPDF's have involved only "first" nearest neighbors

where interactions are involved. In 1991, generalization of NNPDF's involving distributions for nearest neighbors up to an arbitrary m-th nearest neighbor was provided by Edgal [10] for arbitrary densities, temperatures, and interaction potentials. Truskett et al. [10b] also derived related distribution functions, and they provide exact results in the one dimensional system of identical hard rods. In the present paper, further development is provided by deriving similar equilibrium expressions, but this time for PNNPDF's in mixed systems for arbitrary partial densities, temperatures and interaction potentials.

In random material media, it is also usual to study the statistical geometry of such media by determining statistical distributions for various geometrical elements that may be defined within such media [11]. The distributions are usually related to various other properties of the media. For instance the distribution of specific void volume (or void fraction) and specific interface area between void and solid, allows for the determination of fluid permeability for viscous flow through porous media. By centering simple geometrical elements such as spheres at locations of particles distributed within space, it is possible to determine the probability that a given sphere belongs to a cluster of one sphere (singlet cluster), two spheres (doublet cluster), three spheres (triplet cluster), etc. Such cluster probabilities have been used to study continuum percolation [12] behavior of random media. Cluster probabilities have also been used by Edgal and Wiley [12] to study the nature of impurity levels in semiconductors (whereby the scaled hydrogenic model is said to apply to clusters of one impurity atom, while clusters of two or more impurity atoms are said to lead to "deep centers" in semiconductors—The sphere associated with each impurity atom being of one "Bohr" radius).

Because coordinates of nearest neighbors are fundamental geometrical elements which may be used to construct a variety of other geometrical elements in random media, the present paper may be said to address, though at a fundamental level, the statistical geometry of material systems. Furthermore, the paper also investigates the probability distribution of singlet clusters in material media for multi-component non-interacting, as well as, interacting particle systems. Clearly, from all of the above discussions, we find that the results for the emergent concepts of NNPDF's and PNNPDF's (as well as other results of this paper) should soon begin to find rapid use in a variety of problems in the physical sciences. In Sect. 2 of this paper, the partial radial distribution function (also the partial pair or two-body correlation function) as well as higher order partial m-body (or m-tuplet) distribution functions, which are more commonly used for microstructure investigations of multi-component systems in the literature [13], are introduced. The relationship of partial m-body distribution functions with partial NNPDF's are also provided. It is noted that while accurate expressions for partial m-body distribution functions have hitherto been elusive, the present paper provides expressions for PNNPDF's expected to lead to more accurate analytical investigation of microstructure. In section III, the derivation for general expressions for PNNPDF's in equilibrium multi-component systems is provided. In Sect. 4, application to two kinds of systems (the poisson distributed and weakly interacting multi-component systems), are considered. PNNPDF's, partial m-tuplet distribution functions, and singlet cluster probability distribution are provided for the two kinds of systems. In this section, it is evident that though partial m-tuplet distribution functions can be expressed in terms of PNNPDF's, accurate investigation of PNNPDF's does not necessarily

lead readily to accurate investigation of partial m-tuplet distribution functions. This being because highly formidable multiple sums and multiple integrals are encountered (even in the absence of interaction). Section 5 provides additional discussion on microstructure and applications of the results of this paper.

2 PNNPDF's and partial m-tuplet distribution functions

We begin this section with a general review of partial m-tuplet distribution functions, after which their relationship to partial NNPDF's are given. When particles are said to occupy the locations $\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots$, where their radial portions are ordered respectively according to $0 < r_1 < r_2 < r_3 < \cdots$, no particle may occupy locations $\vec{r}$ defined by $0 < r < r_1$; $r_1 < r < r_2$; etc. in the case of partial NNPDF's, while the said locations may be occupied by particles in the case of partial m-tuplet distribution functions. We consider a multi-component system with n constituents in a volume V with the ith constituent having N_i particles. The total number of particles is $N_1 + N_2 + \cdots + N_n = N$. Partial densities involve $\rho_i = \frac{N_i}{V}$.

We first consider the general multi-component point process (GMPP) in which an origin $\left(\vec{r}_0\right)$ is arbitrarily located within the multi-component system. $\vec{r}_0$ does not necessarily coincide with the location of a particle. In the case of the ordinary multi-component point process (OMPP), $\vec{r}_0$ necessarily coincides with the location of a particle. We start by defining an N-tuplet distribution function for all N particles in the system. The distribution is written as $F_N\left(\vec{r}'_{11}, \ldots, \vec{r}'_{1N_1}; \vec{r}'_{21}, \ldots, \vec{r}'_{2N_2}; \ldots; \vec{r}'_{n1}, \ldots, \vec{r}'_{nN_n}\right)$ which defines the marginal probability density for the event that a particle of species i occupies location $\vec{r}'_{ij}$ for j ranging from 1 to N_i, and i ranging from 1 to n. The coordinates $\vec{r}'_{11}, \ldots, \vec{r}'_{nN_n}$ are not necessarily radially ordered as discussed above. The first index in the coordinates refer to particle species. The normalization for the above distribution is written as

$$\int_V \cdots \int_V F_N\left(\vec{r}'_{11}, \ldots, \vec{r}'_{nN_n}\right) d\vec{r}'_{11} \ldots d\vec{r}'_{nN_n} = 1 \qquad (1)$$

Since there are N_i of the coordinates that refer to the location of species i particles, a factor of $\prod_{i=1}^n (N_i!)$ would have been introduced to avoid accounting for configurations which differ by mere permutation of identical particles (of the same species) more than once. (For the single component system, a factor of N! would have had to be used). By implication therefore, Eq. 1 distinguishes between identical particles, and this is what provides a special property for m-tuplet distribution functions.

From Eq. 1, $F_N\left(\vec{r}'_{11}, \ldots, \vec{r}'_{nN_n}\right)$ will need to be vanishingly small in the thermodynamic limit; hence it is usually preferable to work with the new distribution $f_N\left(\vec{r}'_{11}, \ldots, \vec{r}'_{nN_n}\right)$ defined by

$$F_N\left(\vec{r}'_{11}, \ldots, \vec{r}'_{nN_n}\right) = \frac{1}{V^N} f_N\left(\vec{r}'_{11}, \ldots, \vec{r}'_{nN_n}\right) \qquad (2)$$

Also, $f_N \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{nN_n} \right)$ is a better function to deal with, as it is dimensionless. The reduced partial m-tuplet distribution function $f_{p_1+\cdots+p_n} \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1p_1}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{np_n} \right)$ is obtained by integrating over some of the coordinate variables [13]

$$\frac{1}{V^{N-(s_1+\cdots+s_n)-(p_1+\cdots+p_n)}}$$

$$\times \int_V \cdots \int_V f_{N-(s_1+\cdots+s_n)} \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1(N_1-s_1)}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{n(N_n-s_n)} \right)$$

$$\times \prod_{t_1=p_1+1}^{N_1-s_1} d\vec{r}\,'_{1t_1} \cdots \prod_{t_n=p_n+1}^{N_n-s_n} d\vec{r}\,'_{nt_n} = f_{p_1+\cdots+p_n} \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1p_1}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{np_n} \right)$$

$$(3)$$

The normalization of the reduced partial m-tuplet distribution function readily follows as:

$$\frac{1}{V^{p_1+\cdots+p_n}} \int_V \cdots \int_V f_{p_1+\cdots+p_n} \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1p_1}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{np_n} \right)$$

$$\times \prod_{t_1=1}^{p_1} d\vec{r}\,'_{1t_1} \cdots \prod_{t_n=1}^{p_n} d\vec{r}\,'_{nt_n} = 1 \qquad (4)$$

By virtue of distinguishability of all particles,

$$\frac{1}{V^{p_1+\cdots+p_n}} f_{p_1+\cdots+p_n} \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1p_1}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{np_n} \right)$$

is the probability density function for the event that some *specific set* of particles, p_i of which are of species i, are located at $\vec{r}\,'_{i1}, \ldots, \vec{r}\,'_{ip_i}$ (for $i = 1, \ldots, n$). It is usual to define a new event which does not distinguish among identical particles. The new event specifies that the p_i coordinates occupied by species i particles may be occupied by any of the N_i particles of species i. We write the probability density for this event as:

$$\frac{1}{V^{p_1+\cdots+p_n}} f'_{p_1, p_2, \ldots, p_n} \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1p_1}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{np_n} \right);$$

and this is rewritten as $f_{p_1, p_2, \ldots, p_n} \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1p_1}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{np_n} \right)$ in conformity with usual literature usage [13]. The number of distinct ways of choosing p_i particles of species i from N_i identical particles for $i = 1, \ldots, $ n is,

$$\frac{N_1!}{(N_1 - p_1)! p_1!} \frac{N_2!}{(N_2 - p_2)! p_2!} \cdots \frac{N_n!}{(N_n - p_n)! p_n!}.$$

The number of ways of placing the p_i chosen particles (species i) into the locations $\vec{r}'_{i1}, \ldots, \vec{r}'_{ip_i}$ is $p_i!$ (where i $= 1, \ldots, n$). Hence the above distribution functions are related as:

$$
f_{p_1,\ldots,p_n} \left(\vec{r}'_{11}, \ldots, \vec{r}'_{1p_1}; \ldots; \vec{r}'_{n1}, \ldots, \vec{r}'_{np_n} \right)
$$
$$
= \frac{1}{V^{p_1+\cdots+p_n}} \frac{N_1! \cdots N_n! \, (p_1! \cdots p_n!)}{(N_1 - p_1)! p_1! \cdots (N_n - p_n)! p_n!}
$$
$$
\times f_{p_1+\cdots+p_n} \left(\vec{r}'_{11}, \ldots, \vec{r}'_{1p_1}; \ldots; \vec{r}'_{n1}, \ldots, \vec{r}'_{np_n} \right) \tag{5}
$$

Hence it is true that:

$$
\int_V \cdots \int_V f_{p_1,\ldots,p_n} \left(\vec{r}'_{11}, \ldots, \vec{r}'_{1p_1}; \ldots; \vec{r}'_{n1}, \ldots, \vec{r}'_{np_n} \right)
$$
$$
\times d\vec{r}'_{11} \ldots d\vec{r}'_{1p_1} \ldots d\vec{r}'_{n1} \ldots d\vec{r}'_{np_n} = \frac{N_1! \cdots N_n!}{(N_1 - p_1)! \cdots (N_n - p_n)!}
$$

Changing notations, we rewrite the above equation as

$$
\int_V \cdots \int_V f_{m_1,\ldots,m_n} \left(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m} \right) d\vec{r}'_{k_1 1} \ldots d\vec{r}'_{k_m m}
$$
$$
= \frac{N_1! \cdots N_n!}{(N_1 - m_1)! \cdots (N_n - m_n)!} \tag{6}
$$

where m_i of the species indices $(k_1, \ldots, k_m)$ refer to species i (for i $= 1, \ldots, n$) and $(m_1 + \cdots + m_n) = m$. Clearly, Eq. 6 does not give the usual value of unity expected of a normalization of a probability density function.

At this point, we provide vivid additional comments for further insight into partial m-tuplet distribution functions. From the above discussion,

$$
\frac{1}{V^{p_1+\cdots+p_n}} f_{p_1+\cdots+p_n} \left(\vec{r}'_{11}, \ldots, \vec{r}'_{1p_1}; \ldots; \vec{r}'_{n1}, \ldots, \vec{r}'_{np_n} \right) \prod_{t_1=1}^{p_1} d\vec{r}'_{1t_1} \cdots \prod_{t_n=1}^{p_n} d\vec{r}'_{nt_n}
$$

provides the probability that some specific set of particles, p_i of which are of species i (for i ranging from 1 to n) are each located in volume elements $d\vec{r}'_{11}, \ldots, d\vec{r}'_{1p_1}; \ldots;$ $d\vec{r}'_{n1}, \ldots, d\vec{r}'_{np_n}$ situated at $\vec{r}'_{11}, \ldots, \vec{r}'_{1p_1}; \ldots; \vec{r}'_{n1}, \ldots, \vec{r}'_{np_n}$ respectively. Hence by introducing the factor of

$$
\frac{(N_1! \cdots N_n!)}{(N_1 - p_1)! \cdots (N_n - p_n)!},
$$

the new probability function

$$\frac{1}{V^{p_1+\cdots+p_n}} \frac{N_1! \cdots N_n!}{(N_1-p_1)! \cdots (N_n-p_n)!}$$

$$\times f_{p_1+\cdots+p_n} \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1p_1}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{np_n} \right) \prod_{t_1=1}^{p_1} d\vec{r}\,'_{1t_1} \cdots \prod_{t_n=1}^{p_n} d\vec{r}\,'_{nt_n}$$

accounts for the permutation of identical particles between the $(p_1 + \cdots + p_n)$ particles and the rest of the system's particles. This new probability function therefore gives the appropriate probability for the occupation of the volume elements $d\vec{r}\,'_{11}, \ldots, d\vec{r}\,'_{1p_1}; \ldots; d\vec{r}\,'_{n1}, \ldots, d\vec{r}\,'_{np_n}$ whereby particles of the same species in the overall system are considered indistinguishable. (For similar developments, see for instance the derivation and discussions surrounding Eq. 29.7 and the other equations in sections 29 and 40 of Hill [13]). However, events corresponding to different values of the coordinates $\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1p_1}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{np_n}$ (as we carry out integration of the new probability function over the variables) are *not* mutually exclusive (while they are mutually exclusive if we were integrating the old probability function

$$\frac{1}{V^{p_1+\cdots+p_n}} f_{p_1+\cdots+p_n} \left(\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1p_1}; \ldots; \vec{r}\,'_{n1}, \ldots, \vec{r}\,'_{np_n} \right) \prod_{t_1=1}^{p_1} d\vec{r}\,'_{1t_1} \cdots \prod_{t_n=1}^{p_n} d\vec{r}\,'_{nt_n}$$

where identical particles are assumed distinguishable). This is because, as we integrate the new probability function, much of the configurations for the overall system, considered as belonging to the event whereby the $(p_1 + \cdots + p_n)$ particles assume a given configuration, are also considered to belong to a different event corresponding to a different configuration of the same $(p_1 + \cdots + p_n)$ particles. This is so, since other particles in the system that are identical to the $(p_1 + \cdots + p_n)$ particles can take up old coordinate values that the $(p_1 + \cdots + p_n)$ particles "earlier" took in the integration process, while the $(p_1 + \cdots + p_n)$ particles themselves also take up the "earlier" coordinate values of other particles (within the system) that are identical to the $(p_1 + \cdots + p_n)$ particles. The above concept is probably more easily studied employing the distribution for the non-interacting multi-component system and/or employing small N values (e.g., N = 2, 3, ...) with few number of species types (e.g., n = 1, 2, ...). Later on, the above developments will help us to relate partial m-tuplet distribution functions to PNNPDF's. Also, the above discussion shows why terminologies such as partial m-tuplet distribution function, partial m-particle average density function, partial m-body correlation function, etc. (rather than the term partial probability density function) are often used in the literature [13] for $f_{m_1,\ldots,m_n}(\vec{r}\,'_{k_11}, \ldots, \vec{r}\,'_{k_mm})$.

Conditional m-tuplet distribution functions can be formulated in terms of $f_{m_1+\cdots+m_n}$ $(\vec{r}\,'_{k_11}, \ldots, \vec{r}\,'_{k_mm})$ or $f_{m_1,\ldots,m_n}(\vec{r}\,'_{k_11}, \ldots, \vec{r}\,'_{k_mm})$ (note the change of notation). For instance, the conditional m-tuplet distribution function (when identical particles are *distinguishable*) for the event that some specific m_1 particles of species 1 are located at $\vec{r}\,'_{11}, \ldots, \vec{r}\,'_{1m_1}$; some specific m_2 particles of species 2 are located at $\vec{r}\,'_{2(m_1+1)}, \ldots,$

$\vec{r}'_{2(m_1+m_2)}; \ldots;$ some specific m_n particles of species n are located at $\vec{r}'_{n(m_1+\cdots+m_{n-1}+1)}$,
$\ldots, \vec{r}'_{n(m_1+\cdots+m_n)};$ given that some other specific p_1 particles of species 1 are located at $\vec{r}'_{1(m_1+\cdots+m_n+1)}, \ldots, \vec{r}'_{1(m_1+\cdots+m_n+p_1)};$ some other specific p_2 particles of species 2 are located at $\vec{r}'_{2(m_1+\cdots+m_n+p_1+1)}, \ldots, \vec{r}'_{2(m_1+\cdots+m_n+p_1+p_2)}; \ldots;$ some other specific p_n particles of species n are located at $\vec{r}'_{n(m_1+\cdots+m_n+p_1+\cdots+p_{n-1}+1)}, \ldots,$
$\vec{r}'_{n(m_1+\cdots+m_n+p_1+\cdots+p_n)};$ is given as

$$
f_{m_1+\cdots+m_n}\left(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m} \middle| \vec{r}'_{k_{m+1}(m+1)}, \ldots, \vec{r}'_{k_{m+p}(m+p)}\right)
$$
$$
= \frac{f_{(m_1+p_1)+\cdots+(m_n+p_n)}\left(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_{m+p}(m+p)}\right)}{f_{p_1+\cdots+p_n}(\vec{r}'_{k_{m+1}(m+1)}, \ldots, \vec{r}'_{k_{m+p}(m+p)})}
$$

(where $p = p_1 + \cdots + p_n$). $\frac{1}{V^m} f_{m_1+\cdots+m_n}(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m})$ is a well formulated (appropriately normalized) probability density function (unlike $f_{m_1,\ldots,m_n}(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m})$), hence it is easy to see that the above conditional distribution function normalizes adequately as:

$$
\int_V \cdots \int_V \frac{1}{V^m} f_{m_1+\cdots+m_n}\left(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m} \middle| \vec{r}'_{k_{m+1}(m+1)}, \ldots, \vec{r}'_{k_{m+p}(m+p)}\right)
$$
$$
\times d\vec{r}'_{k_1 1} \ldots d\vec{r}'_{k_m m} = 1
$$

Similarly, we also have the case when identical particles are *indistinguishable*, the conditional m-tuplet distribution function is written as:

$$
f_{m_1,\ldots,m_n}\left(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m} \middle| \vec{r}'_{k_{m+1}(m+1)}, \ldots, \vec{r}'_{k_{m+p}(m+p)}\right)
$$
$$
= \frac{f_{m_1+p_1,\cdots,m_n+p_n}\left(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_{m+p}(m+p)}\right)}{f_{p_1,\ldots,p_n}(\vec{r}'_{k_{m+1}(m+1)}, \ldots, \vec{r}'_{k_{m+p}(m+p)})} \tag{7}
$$

Using the relationship of Eq. 5, we have that the conditional distribution of Eq. 7 normalizes (also unusually) as:

$$
\int_V \cdots \int_V f_{m_1,\ldots,m_n}\left(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m} \middle| \vec{r}'_{k_{m+1}(m+1)}, \ldots, \vec{r}'_{k_{m+p}(m+p)}\right) d\vec{r}'_{k_1 1} \ldots d\vec{r}'_{k_m m}
$$
$$
= \frac{(N_1 - p_1)! \cdots (N_n - p_n)!}{(N_1 - (m_1 + p_1))! \cdots (N_n - (m_n + p_n))!}
$$

For the OMPP process, the distribution $f_{m_1,\ldots,m_n}(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m})$ may be replaced by the conditional m-tuplet distribution function $f_{m_1,\ldots,m_n}\left(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m} \middle| \vec{r}'_{i0}\right)$.

The particle at location $\vec{r}'_{i0}$ is said to be located at the origin and is of species i. Hence,

$$f_{m_1,\ldots,m_{i-1},m_i+1,m_{i+1},\ldots,m_n}(\vec{r}'_{i0}, \vec{r}'_{k_11}, \ldots, \vec{r}'_{k_mm}) = f_1(\vec{r}'_{i0})$$
$$\times f_{m_1,\ldots,m_n}\left(\vec{r}'_{k_11}, \ldots, \vec{r}'_{k_mm}\,\middle|\,\vec{r}'_{i0}\right)$$

(Clearly, $f_1(\vec{r}'_{i0})$ is the special case of $f_{m_1,\ldots,m_{i-1},m_i+1,m_{i+1},\ldots,m_n}(\vec{r}'_{i0}, \vec{r}'_{k_11}, \ldots,$ $\vec{r}'_{k_mm})$ where $m_k = 0$ for all k from 1 to n). For a homogeneous system, we have that $f_1(\vec{r}'_{i0}) = \frac{N_i}{V} = \rho_i$ (understanding of course that the probability density for finding a particle of given species anywhere in volume V should be constant, thus the integrand in Eq. 4 for instance should be a constant of unity for $p_1 + \cdots + p_n = 1$, and this result may then be employed in Eq. 5).

Considering the GMPP process again, we have that the PNNPDF for the event that the first nearest neighbor of the origin is of species k_1, at location $\vec{r}_{k_11}, \ldots,$ the mth nearest neighbor of the origin is of species k_m at location $\vec{r}_{k_mm}$(where the radial portion of the coordinates are ordered according to $r_{k_11} < r_{k_22} < \cdots < r_{k_mm}$) is written as: $g_{1,\ldots,m}(\vec{r}_{k_11}, \ldots, \vec{r}_{k_mm})$. This pdf assumes identical particles are indistinguishable. The indices $k_1, \ldots, k_m$ may each assume any of the species values $1, 2, \ldots, n$ as before. It is also assumed m_i of the nearest neighbors are of species i; and $m_1 + \cdots + m_n = m$. An alternative way to write the above PNNPDF reflecting the number of nearest neighbors of given species within the total of m nearest neighbors is: $g_{1,\ldots,m}^{m_1,\ldots,m_n}(\vec{r}_{k_11}, \ldots, \vec{r}_{k_mm})$. In Fig. 1 is shown an arbitrarily situated origin and its m nearest neighbors. In all developments in this paper, it is assumed $m \ll N_i(i = 1, \ldots, n)$. Also, under the criterion of indistinguishability of identical particles, the marginal pdf for the event that the s_1th nearest neighbor is of species k_1 at location $\vec{r}_{k_11}; \ldots$; the s_qth nearest neighbor is of species k_q at location $\vec{r}_{k_qq}$, is readily obtained by appropriately integrating $g_{1,\ldots,m}^{(l_m)}(\vec{r}_{k_11}, \ldots, \vec{r}_{k_mm})$ for a select set of (l_m) indices, and adding up. The marginal pdf is written as $g_{s_1,\ldots,s_q}(\vec{r}_{k_11}, \ldots, \vec{r}_{k_qq})$ or $g_{s_1,\ldots,s_q}^{t_1,\ldots,t_n}(\vec{r}_{k_11}, \ldots, \vec{r}_{k_qq})$ where t_1 of the indices $(k_1, \ldots, k_q)$ refer to species 1; $\ldots$; t_n of the indices $(k_1, \ldots, k_q)$ refer to species n. Clearly, we must have $s_1 < s_2 < \cdots < s_q \leq m$. The upper index (l_m) in the PNNPDF $g_{1,\ldots,m}^{(l_m)}(\vec{r}_{k_11}, \ldots, \vec{r}_{k_mm})$ indicates a specific choice of values of $k_1, \ldots, k_m$ of which there are a maximum of n^m of such choices. (Note that if m were greater than, or of the order of N_i, the maximum of such choices would have been awkward to determine. Such difficulties are avoided since we restrict ourselves to the condition $s_q, m \ll N_i$).

Next we determine the select set of (l_m) indices for the PNNPDF's $g_{1,\ldots,m}^{(l_m)}(\vec{r}_{k_11}, \ldots,$ $\vec{r}_{k_mm})$ used to construct the marginal pdf $g_{s_1,\ldots,s_q}(\vec{r}_{k_11}, \ldots, \vec{r}_{k_qq})$. Obviously, the set of distributions $g_{1,\ldots,m}^{(l_m)}(\vec{r}_{k_11}, \ldots, \vec{r}_{k_mm})$ we are to select from must be such that their s_1th nearest neighbor is of species k_1 at location $\vec{r}_{k_11}$ (similar to that of the marginal pdf in question); $\ldots$; their s_qth nearest neighbor is of species k_q at location $\vec{r}_{k_qq}$ (similar to that of the marginal pdf in question). Now, there are $(s_1 - 1)$
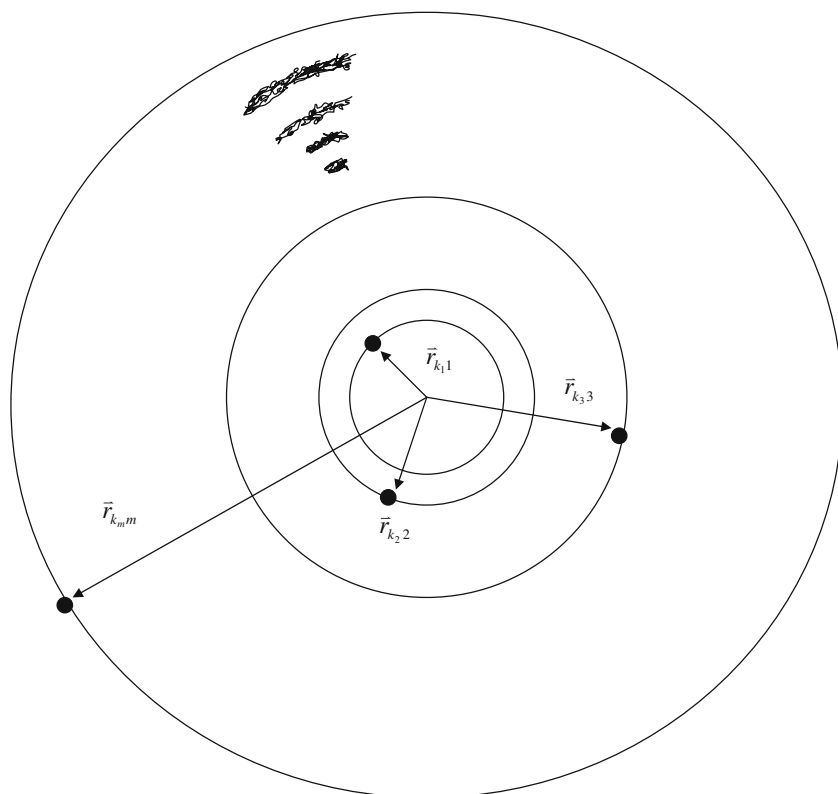
Fig. 1 The arbitrarily situated origin (center of concentric circles) and its m nearest neighbors in the n-component system. $\vec{r}_{k_i i}$ is the location of the origin's ith nearest neighbor (species k_i) where k_i may be 1, 2, …, n

neighbors nearer to the origin than the s_1th nearest neighbor. Hence there are n^{s_1-1} choices of species that can be made for the nearer neighbors. Similarly, there are $(s_2 - s_1 - 1)$ neighbors between the s_1th and s_2th nearest neighbors. Hence there are $n^{s_2-s_1-1}$ choices of species that can be made for such "in-between" choices. A similar thing can also be said for the $(s_3 - s_2 - 1)$ neighbors between the s_2th and s_3th nearest neighbors; etc. Hence the select set of (l_m) indices indicated above will total (in number), $n^{\{(s_1-1)+(s_2-s_1-1)+\cdots+(s_q-s_{q-1}-1)+(m-s_q)\}} = n^{m-q}$. Similar to $g_{1,\ldots,m}^{(l_m)}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$, we can also write the above marginal pdf with an upper index as $g_{s_1,\ldots s_q}^{(l_q)}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_q q})$, where (l_q) indicates a specific choice of $(k_1, \ldots, k_q)$ of which there are a maximum of n^q of such choices.

Now, it should be noted that the PNNPDF's $g_{1,\ldots,m}^{(l_m)}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ and $g_{s_1,\ldots,s_q}^{(l_q)}$ $(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_q q})$ cannot each normalize to unity for a specific choice of species for neighbors. See Edgal and Huber [6b] for further discussion. All possible choices of species for neighbors must be considered for normalization. Hence the stipulation for normalization must be written as:

$$\sum_{(l_m)} \int_{\Omega_s} \left[\int_0^R \int_0^{r_{k_m m}} \cdots \int_0^{r_{k_2 2}} g_{1,\ldots,m}^{(l_m)} (\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m}) dr_{k_1 1} \ldots dr_{k_m m} \right] d\Omega_s$$

$$= \sum_{(l_q)} \int_{\Omega_q} \left[\int_0^R \int_0^{r_{k_q q}} \cdots \int_0^{r_{k_2 2}} g_{s_1,\ldots,s_q}^{(l_q)} (\vec{r}_{k_1 1}, \ldots \vec{r}_{k_q q}) dr_{k_1 1} \ldots dr_{k_q q} \right] d\Omega_q = 1 \quad (8)$$

R is the largest radial value attainable and this depends on the location of the origin within volume V. The symbols Ω_s and Ω_q denote the total space for the angular parts of the variables $\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m}$ and $\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_q q}$ respectively. The sum over (l_m) involves n^m terms, while the sum over (l_q) involves n^q terms.

The partial m-tuplet distribution $f_{m_1,\ldots,m_n}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ may also in turn be constructed from the PNNPDF $g_{s_1,\ldots,s_m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ as follows. The coordinates $\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m}$ in $f_{m_1,\ldots,m_n}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ are now radially ordered ($r_{k_1 1} < r_{k_2 2} < \cdots < r_{k_m m}$) with m_1 of them involving particles of species 1; ...; m_n of them involving particles of species n (and $m_1 + \cdots + m_n = m$) as before. It is assumed $s_m \ll N_i$ even though s_m may be large and may in fact be considered to tend to ∞ sometimes. Results obtained by this condition are in no way limited since large s_m implies very distant neighbors where the behavior of m-tuplet distribution functions become "featureless," assuming a character reflective of the average overall background of the system. This is commonly depicted as a constant value (at large radial distances) of the "radial" correlation function [13], which is the radial distribution function (the case of m = 1 for the OMPP m-tuplet distribution function) divided by the corresponding function for the non-interaction case.

Now, earlier discussions above (see also for instance sections 29 and 40 of Hill [13]) show that $f_{m_1,\ldots,m_n}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m}) d\vec{r}_{k_1 1} \ldots d\vec{r}_{k_m m}$ is an actual probability for the occupation of volume elements $d\vec{r}_{k_1 1} \ldots d\vec{r}_{k_m m}$ when identical particles are considered indistinguishable for the GMPP process. But a similar probability is readily formulated as a sum employing the PNNPDF's $g_{s_1,\ldots,s_m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ for different values of $s_1, \ldots, s_m$. Hence PNNPDF's and partial m-tuplet distribution functions are readily related as follows:

$$f_{m_1,\ldots,m_n} \left(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m} \right) d\vec{r}_{k_1 1} \ldots d\vec{r}_{k_m m}$$
$$= \sum_{s_1} \cdots \sum_{s_m} g_{s_1,\ldots,s_m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m}) d\vec{r}_{k_1 1} \ldots d\vec{r}_{k_m m} \quad (9)$$

In the multiple sum, s_m is required to vary from m to some large number (which can actually be set to infinity—see below for further discussion), s_{m-1} is required to vary from $(m - 1)$ to $(s_m - 1)$, ..., s_1 is required to vary from 1 to $(s_2 - 1)$. It is assumed that the largest value we need to consider for s_m is usually $\ll N_i$ (though formally it should be N). This is in line with earlier restrictions stated for our formulation of the pdf $g_{s_1,\ldots,s_m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$. If the largest value were $\gtrsim N_i$, the species types presumed for neighbors (ie. the s_1th, ..., s_mth nearest neighbors, and other neighbors

in-between the said nearest neighbors including the origin) in the distributions, would become difficult to reconcile with the values of $s_1, \ldots, s_m$ (in the sums) in general. (Such difficulty is however absent in the single component case, where the maximum value of s_m can easily be rightfully taken as N). In practical situations requiring the sum of Eq. 9 to be determined, it is important to note that $g_{s_1,\ldots,s_m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ may be expected to usually get rapidly small beyond some large values of $s_1, \ldots, s_m$. In which case, we do not need to contend with terms where one or more of $s_1, \ldots, s_m$ is $\gtrsim N_i$. Hence only a finite (and possibly small) number of terms (where the maximum of s_m is $\ll N_i$) will usually need to be considered in the sum of Eq. 9. Moreover, for large distances where large values of $s_1, \ldots, s_m$ are involved, such regions do not require further scrutiny as they are known to involve featureless distributions as discussed above. It should also be noted that because the coordinates now obey the relationship $r_{k_1 1} < r_{k_2 2} < \cdots < r_{k_m m}$, the distribution

$$f_{m_1,\ldots,m_n}\left(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m}\right) \text{ normalizes to} \frac{N_1! \cdots N_n!}{(N_1 - m_1)! \cdots (N_n - m_n)!(m_1 + \cdots + m_n)!}.$$

If the PNNPDF for the OMPP process may be written as $g^i_{1,\ldots,m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ (where a particle of species i is located at the origin), it is easy to see that the OMPP partial m-tuplet distribution function $f_{m_1,\ldots,m_n}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m} | \vec{r}_{i0})$ may be obtained from PNNPDF's of the form $g^i_{s_1,\ldots,s_m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ using a multiple sum expression similar to Eq. 9.

We remark in passing that for the multi-component poisson point process (non-interacting particle system), the constancy of $f_{m_1+\cdots+m_n}(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m})$ means $f_{m_1+\cdots+m_n}(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m}) = 1$ (by Eq. 4), and thus from (5),

$$f_{m_1,\ldots,m_n}\left(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m}\right) = \frac{1}{V^m} \frac{N_1! \cdots N_n!}{(N_1 - m_1)! \cdots (N_n - m_n)!}$$
$$\approx \rho_1^{m_1} \rho_2^{m_2} \cdots \rho_n^{m_n} \quad \text{(for m} \ll N_i) \tag{10}$$

In Sect. 4, the above simple result is re-derived using Eq. 9 as one way to show the authenticity of Eq. 9. Being partial distributions, the sum over the above partial m-tuplet distribution functions, i.e. $\sum_{(l_m)} f_{m_1,\ldots,m_n}(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m})$, where (l_m) indicates different choices of the species $(k_1, \ldots, k_m)$, of which there is a maximum of n^m such choices, yields the m-tuplet distribution for the corresponding single component system (i.e. $f_m(\vec{r}'_1, \ldots, \vec{r}'_m)$ where particles of different species are not distinguished) involving N particles in volume V (whereby $N = N_1 + \cdots + N_n$). This is readily shown for the non-interacting case as

$$\sum_{(l_m)} f_{m_1,\ldots,m_n}(\vec{r}'_{k_1 1}, \ldots, \vec{r}'_{k_m m}) = \sum_{(l_m)} \rho_1^{m_1} \cdots \rho_n^{m_n} = (\rho_1 + \cdots + \rho_n)^m$$
$$= \rho^m = f_m(\vec{r}'_1, \ldots, \vec{r}'_m).$$

The second equality is easily tested for small values of m and n.

Finally, we remark that while PNNPDF's provide detailed information about microstructure of local neighborhoods (which can be extended as far as desired) within material systems, partial m-tuplet distributions generally provide only "average" information on the microstructure of such local neighborhoods as they are formulated to deal with a fixed number of neighbor particles irrespective of the extent of the local neighborhood under investigation. PNNPDF's on the other hand are required to feature a number of nearest neighbor particles that grows in proportion to the size of neighborhood being investigated.

3 PNNPDF's for interacting particles

With the connection between PNNPDF's and partial m-tuplet distribution functions established, we may next proceed to determine general forms for PNNPDF's for interacting multi-component systems. In Edgal [10], the joint PDF for the coordinates of all $(N-1)$ neighbors of the particle at the origin (arbitrarily situated within volume V) in an ordinary point process for non-interacting single component poisson fluid was given. The joint PDF is easily modified to obtain that for an OMPP process as:

$$g^i_{1,\ldots,N-1}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_{N-1}(N-1)}) = (N_1! \cdots N_{i-1}!(N_i - 1)!N_{i+1}! \cdots N_n!)$$
$$\times \prod_{l=1}^{N-1} h(\vec{r}_{k_l l}) f_{k_l l} \qquad (11)$$

where $\vec{r}_{k_l l}$ involves the radial ($r_{k_l l}$), azimuthal ($\phi_{k_l l}$), and theta ($\theta_{k_l l}$) coordinates, and the $\vec{r}_{k_l l'}s$ are radially ordered (as discussed in Sect. 2). $h(\vec{r}_{k_l l}) = \frac{1}{V}$, $f_{k_l l} = r^2_{k_l l} \sin \theta_{k_l l}$. The superscript i on the left hand side of Eq. 11 indicates a particle of species i is at the origin. For the GMPP process, Eq. 11 is easily rewritten as:

$$g_{1,\ldots,N}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_N N}) = (N_1! \cdots N_n!) \prod_{l=1}^{N} h(\vec{r}_{k_l l}) f_{k_l l} \qquad (12)$$

The pdf for the GMPP process for instance is easily written in the presence of arbitrary interactions, under equilibrium conditions as

$$g_{1,\ldots,N}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_N N}) \propto (N_1! \cdots N_n!) e^{-\beta U_N} \prod_{l=1}^{N} h(\vec{r}_{k_l l}) f_{k_l l} \qquad (13)$$

$\beta = 1/k_B T$, k_B is Boltzmann's constant, T is temperature, and U_N is total potential energy for given configuration of the system. U_N in its most general form can be written as a sum of k-body potentials involving k particles, where k may range from 1 to N, and the k-body potentials generally varies for different configurations and different particle species involved. The potential energy of the m-nearest neighbors (in the GMPP process), assuming these are isolated from the remaining $(N - m)$ particles (called "atmospheric" particles) in the system, may be written as $U_{m_1,\ldots,m_n}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$

or simply $U_{m_1,...,m_n}$. It is assumed the m-nearest neighbors involve m_1 species 1 particles, ..., m_n species n particles (where $m_1 + \cdots + m_n = m$). Similarly, the potential energy of the atmospheric particles (assuming isolation from the m nearest neighbors) may be written as $U_{N_1-m_1,...,N_n-m_n}(\vec{r}_{k_{m+1}(m+1)}, ..., \vec{r}_{k_N N})$ or simply $U_{N_1-m_1,...,N_n-m_n}$. We may therefore write:

$$U_N = U_{m_1,...,m_n} + U_{N_1-m_1,...,N_n-m_n} + U_{\text{bdy}} \qquad (14)$$

where U_{bdy} is the additional energy which results when the mutual interaction between the above two sets of particles is turned on. Considering the condition of normalization, and integrating the joint PDF of Eq. 13 over all allowed positions of the coordinates except the first m coordinates, we obtain the m-nearest neighbor joint PDF as:

$$g_{1,...,m}(\vec{r}_{k_1 1}, ..., \vec{r}_{k_m m}) = \frac{\exp(A/kT)}{(N_1 - m_1)! \cdots (N_n - m_n)!} \exp\left(-\beta U_{m_1,...,m_n}\right)$$

$$\times \int_{\hat{V}} \cdots \int_{\hat{V}} \exp\left[-\beta(U_{bdy} + U_{N_1-m_1,...,N_n-m_n})\right]$$

$$\times \prod_{l=m+1}^{N} d^3 \hat{r}_{k_l l} \qquad (15)$$

A is the configurational contribution to the free energy of N_1 (species 1) particles; ...; N_n (species n) particles in the space of volume V. (Note that A is said to be properly formulated in accordance with the correct Boltzmann's counting for a multicomponent system). $\hat{V}$ is the volume $(V - v_m)$ of the region outside the sphere of radius r_m containing the m-nearest neighbors (see Fig. 1). $v_m = \left(\frac{4}{3}\right) \pi r_m^3$. Note that boundary effects of volume V are ignored. The coordinates in the integral of Eq. 15 have been written as $\hat{\vec{r}}_{k_l l}$ to indicate their radial parts are not ordered. To account for the resulting permutations of the coordinates of particles of the same species, the factor of $\frac{1}{(N_1-m_1)! \cdots (N_n-m_n)!}$ has been introduced. Eq. 15 is exact, and the integral provides what may be termed an "atmospheric volume effect." The integral and the factor $\frac{1}{(N_1-m_1)! \cdots (N_n-m_n)!}$ is known to constitute the configurational partition function of $N - (m_1 + \cdots + m_n)$ particles (involving $(N_1 - m_1)$ species 1 particles; ...; $(N_m - m_n)$ species n particles) in volume $\hat{V}$ with the m-nearest neighbors acting as external sources of fields; and this function may be written as:

$$Z_s^e(N_1 - m_1, ..., N_n - m_n, \hat{V})$$

$$= Z_s(N_1 - m_1, ..., N_n - m_n, \hat{V}) \times \left[\frac{1}{(N_1 - m_1)! \ldots (N_n - m_n)!} \right. \qquad (16)$$

$$\times \int_{\hat{V}} \cdots \int_{\hat{V}} \frac{e^{-\beta U_{bdy}} e^{-\beta U_{N_1-m_1,...,N_n-m_n}}}{Z(N_1 - m_1, ..., N_n - m_n, \hat{V})^{\hat{V}} S_{N_1-m_1,...,N_n-m_n}} \left. \prod_{l=m+1}^{N} d^3 \hat{r}_{k_l l} \right]$$

where $Z_s(N_1 - m_1, \ldots, N_n - m_n, \hat{V}) = Z(N_1 - m_1, \ldots, N_n - m_n, \hat{V})$ $\times {}^{\hat{V}}S_{N_1-m_1,\ldots,N_n-m_n}$. The quantities Z_s and Z are similarly defined as the configurational partition function of $(N_1 - m_1)$ species 1 particles; … ; $(N_n - m_n)$ species n particles (in the absence of external forces) in a volume of size $\hat{V}$. The shapes of the volumes employed in both cases, however, differ essentially in the fact that the volume employed in calculating Z_s contains a "void" of size v_m (which contains the m-nearest neighbors), while the volume employed for calculating Z is of some 'standard" shape (which is actually the shape of volume V) that does not contain the said void. Clearly, Z may be more readily calculable than Z_s. The function ${}^{\hat{V}}S_{N_1-m_1,\ldots,N_n-m_n}$ accounts for the difference in shape between the above two volumes. The quantity

$$\frac{e^{-\beta U_{N_1-m_1,\ldots,N_n-m_n}}}{Z(N_1 - m_1, \ldots, N_n - m_n, \hat{V}) {}^{\hat{V}}S_{N_1-m_1,\ldots,N_n-m_n}}$$

is the probability for a given configuration of the atmospheric particles (in the absence of external forces) when they are restricted within volume $\hat{V}$. The quantity in square brackets in Eq. 16 therefore yields the average of $e^{-\beta U_{bdy}}$ (i.e., $<e^{-\beta U_{bdy}}>$), in the phase space of size $Z_s(N_1 - m_1, \ldots, N_n - m_n, \hat{V})$. The joint PDF of Eq. 15 may therefore be rewritten as:

$$g_{1,\ldots,m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m}) = \lfloor \exp(A/kT) Z(N_1 - m_1, \ldots, N_n - m_n, V) \rfloor$$
$$\times \left[{}^{\hat{V}}S_{N_1-m_1,\ldots,N_n-m_n} \left\langle e^{-\beta U_{bdy}} \right\rangle Z(N_1 - m_1, \ldots, N_n - m_n, \hat{V}) / \right.$$
$$\left. \times Z(N_1 - m_1, \ldots, N_n - m_n, V) \right]$$
$$\times \left[\exp(-\beta U_{m_1,\ldots,m_n}) \right] \tag{17}$$

Following the same arguments on shape and surface (or boundary) effects as provided in Edgal [10], we declare that effect (on the system free energy or partition function) due to volume size changes is much stronger than effect due to the accompanying shape change. Hence writing the shape effect factor ${}^{\hat{V}}S_{N_1-m_1,\ldots,N_n-m_n}$ as $e^{-\beta S}$, and assuming changes in interaction energy and changes in volume size produce effects that are at par, it must be true that $\left| U_{m_1,\ldots,m_n} \right| \gg |S|$ for m large enough. (This is to say shape effects are "naturally" or generally very weak effects). Also, writing the boundary effect factor $\langle e^{-\beta U_{bdy}} \rangle$ as $e^{-\beta \bar{E}}$, we have that for large enough m, when the surface-to-volume ratio (of the volume v_m) is sufficiently small (often enough), it must be true that $\left| U_{m_1,\ldots,m_n} \right| \gg \left| U_{bdy} \right| \sim \left| \bar{E} \right|$. Now, it is expected that for m large enough, the volume v_m is large and its size fluctuation (δv_m) is relatively small (i.e., $\delta v_m \ll v_m$) within the phase space for nearest neighbor configurations. This will lead to fluctuation in the shape effect term (S), written as S'; and because shape effects are "naturally" weak, we expect $\left| S' \right| \ll |S|$. Now, in the case of boundary effects, because the process of averaging is known to usually smoothen out fluctuations substantially, we have that fluctuations in v_m should not only lead to relatively small fluctuations in U_{bdy} (written as bdy U'_{bdy}), but should also lead to small fluctuations in $\langle e^{-\beta U_{bdy}} \rangle$ or $e^{-\beta E}$, which should in turn lead to small fluctuations in $\bar{E}$ (written as

E'); i.e., $\left|U_{bdy}\right|, \left|\bar{E}\right| \gg \left|U'_{bdy}\right|, \left|E'\right|)$. Hence an important implication of the above discussion is that while a large m-value (possibly several hundreds or thousands) is required to make $\left|U_{m_1,\ldots,m_n}\right| \gg \left|U_{bdy}\right|, \left|\bar{E}\right|, \left|S\right|$, a relatively much smaller m value (possibly a few tens or so) is required to make $\left|U_{m_1,\ldots,m_n}\right| \gg \left|U'_{bdy}\right|, \left|E'\right|, \left|S'\right|$. This last condition is what is actually needed for surface and shape effects to be considered relatively weak, and are referred to as "*reduced*" surface and shape effects. The larger non fluctuating terms $U_{bdy}, \bar{E}, S$ are not important as later developments easily show that they may be featured in the normalization constant of the PNNPDF.

Another proposition of no less importance argued in Edgal [10], which is also adopted in this paper, is as follows. If numerical errors (such as round off errors) made in computing functions such as $e^{-\beta U_{m_1,\ldots,m_n}}$, as well as errors made in the general process for evaluating the system's partition function (Z), may be expressed as a multiplicative factor involving the exponential e^{ERR}, the scheme employed for evaluating Z, if "adequate," must be such that m may be chosen large enough so that $\left|U_{m_1,\ldots,m_n}\right| \gg ERR/\beta$. Hence if the second term in square brackets in Eq. 17 may be written as $\left[Z(N_1 - m_1, \ldots, N_n - m_n, \hat{V})/Z(N_1 - m_1, \ldots, N_n - m_n, V)\right]$, the third term in square brackets may be modified to give $\left[\exp(-\beta(U_{m_1,\ldots,m_n} + E' - S' - ERR/\beta))\right]$, while the first term in square brackets may be said to contain configuration independent quantities (including quantities involving $\bar{E}, S$) and thus written collectively as the normalization constant $h_{k_1 1,\ldots,k_m m}$.

Clearly, since neglect of some or all of the quantities $E', S', \mathrm{ERR}/\beta$ in the modified third term of Eq. 17 may be regarded as equivalent to a mere small relative error in the computation of say the quantity $U_{m_1,\ldots,m_n}$ (such as round off errors), one may readily expect that the resulting approximate PNNPDF should be sufficiently accurate, with the accuracy increasing as m increases. In developments soon to follow, it will be seen that the modified second term in square brackets of Eq. 17 is an exponential function whose argument, apart from being much larger than the quantity $(\left|ERR\right|/\beta) + \left|E'\right| + \left|S'\right|$ (as may be inferred from earlier discussions), varies vary rapidly with $r_{k_m m}$. Hence it may be expected that the magnitude of the variation in this argument (for small changes in $r_{k_m m}$) will be much larger than the quantity $(\left|ERR\right|/\beta) + \left|E'\right| + \left|S'\right|$ everywhere. Also, since $\left|U_{m_1,\ldots,m_n}\right| \gg \left|ERR\right|/\beta + \left|E'\right| + \left|S'\right|$, the magnitude of the variation in $U_{m_1,\ldots,m_n}$ for slight changes in "spatial extent" or "spatial form" of nearest neighbor configurations, will often be much larger than $((\left|ERR\right|/\beta) + \left|E'\right| + \left|S'\right|)$. Spatial extent of the m-nearest neighbors may be expressed as $r_{k_m m}$, while their spatial form may be expressed as: $y_{k_1 1}, \ldots, y_{k_m m}, \theta_{k_1 1}, \ldots, \theta_{k_m m}, \phi_{k_1 1}, \ldots, \phi_{k_m m}$ (where $y_{k_i i} = \frac{r_{k_i i}}{r_{k_m m}}$, and thus $y_{k_m m} = 1$). Different spatial extents for given spatial form implies a uniform contraction or expansion of the configuration of nearest neighbors.

Now, we may express the modified second term in square brackets of Eq. 17 as an exponential as indicated above. If we write the argument of this exponential as "Arg," then since Arg varies rapidly (at least with $r_{k_m m}$) as earlier indicated, it must be true that the function $E = Arg - \beta U_{m_1,\ldots,m_n}$ at any given point (in the space of near neighbor configuration) will usually be equal to the value of the function $\hat{E} = E - \beta(E' - S' - ERR/\beta)$ evaluated at some nearby point (s) and vice versa. (Only at a set of points of small "measure" near the "edge" of the space of near neighbor configurations, may

one find that the above relationship may not hold). Hence the outline features of the graphs of E and $\hat{E}$ are expected to be essentially identical if details within localities of "small" extent may be ignored. Similarly, one may expect the graphs of functions of E and $\hat{E}$ (i.e. f(E) and f($\hat{E}$)) to be essentially identical in outline features. Hence the exact PNNPDF (which is a function of $\hat{E}$) and the approximate PNNPDF (in which E', S', ERR/β are ignored) are expected to be identical in outline features. Hence the use of the approximate PNNPDF in constructing nearest neighbor configurations is expected to lead to "*slight*" deformations in their extent and form. Hence, if for instance the mean of a function of the configuration of nearest neighbors (which may be expressed as a function of $U_{m_1,\ldots,m_n}$; i.e., f($U_{m_1,\ldots,m_n}$)), is to be determined with the approximate PNNPDF, such mean may be expected to be written as f($\bar{U}_{m_1,\ldots,m_n} + U'$), where f($\bar{U}_{m_1,\ldots,m_n}$) is the exact mean, and $\left| U' \right|$ is of the order of $(\left| ERR \right| /\beta) + \left| E' \right| + \left| S' \right|$. In the poisson non-interaction case where $U_{m_1,\ldots,m_n} = 0$, we have that $E' = S' = 0$. In this case, effect of numerical approximation errors in evaluating various terms of the PNNPDF can be more exactly studied.

In the presence of interaction, Edgal [10] suggested that we may write the configurational partition function as

$$Z(N_1, \ldots, N_n, V) = \frac{1}{N_1! \ldots N_n!} (\varepsilon V)^N$$

where ε is a dimensionless quantity which is some function of $N_1, \ldots, N_n$, V. We may write:

$$\varepsilon(N_1 - m_1, \ldots, N_n - m_n, \hat{V}) = \varepsilon(N_1, \ldots, N_n, V) - m_1 \frac{\partial \varepsilon(N_1, \ldots, N_n, V)}{\partial N_1} - \cdots$$
$$-m_n \frac{\partial \varepsilon(N_1, \ldots, N_n, V)}{\partial N_n} - v_m \frac{\partial \varepsilon(N_1, \ldots, N_n, V)}{\partial V}$$

(since $m_1 \ll N_1, \ldots, m_n \ll N_n$, and $v_m \ll V$). Noting that ε may be written as a function of the partial densities $\rho_1, \ldots, \rho_n$, we have that:

$$\varepsilon(N_1 - m_1, \ldots, N_n - m_n, \hat{V})$$
$$= \varepsilon(\rho_1, \ldots, \rho_n) - \frac{m_1}{V} \frac{\partial \varepsilon(\rho_1, \ldots, \rho_n)}{\partial \rho_1} - \cdots \cdots - \frac{m_n}{V} \frac{\partial \varepsilon(\rho_1, \ldots, \rho_n)}{\partial \rho_n}$$
$$- v_m \left(\frac{\partial \varepsilon(\rho_1, \ldots, \rho_n)}{\partial \rho_1} \frac{\partial \rho_1}{\partial V} + \cdots \cdots + \frac{\partial \varepsilon(\rho_1, \ldots, \rho_n)}{\partial \rho_n} \frac{\partial \rho_n}{\partial V} \right)$$
$$= \varepsilon - \frac{(m_1 - \rho_1 v_m)}{V} \frac{\partial \varepsilon}{\partial \rho_1} - \cdots - \frac{(m_n - \rho_n v_m)}{V} \frac{\partial \varepsilon}{\partial \rho_n}$$

Hence we may write:

$$Z(N_1 - m_1, \ldots, N_n - m_n, \hat{V}) = \frac{\hat{V}^{N-(m_1+\cdots+m_n)}}{(N_1 - m_1)! \ldots (N_n - m_n)!} \left(\varepsilon - \frac{(m_1 - \rho_1 v_m)}{V} \frac{\partial \varepsilon}{\partial \rho_1} \right.$$
$$\left. - \cdots - \frac{(m_n - \rho_n v_m)}{V} \frac{\partial \varepsilon}{\partial \rho_n} \right)^{N-(m_1+\cdots+m_n)}$$

Hence in the thermodynamic limit, and considering that such terms as $\frac{\rho m_i}{\varepsilon} \frac{\partial \varepsilon}{\partial \rho_i}$ are configuration independent, we write the GMPP process PNNPDF (of Eq. 17) in the presence of arbitrary inter-particle interaction at arbitrary partial densities and temperatures as:

$$g_{1,\ldots,m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$$
$$= h_{k_1 1, \ldots, k_m m} \exp \left[-\frac{4}{3} \pi r_{k_m m}^3 \rho \left(1 - \frac{\rho_1}{\varepsilon} \frac{\partial \varepsilon}{\partial \rho_1} - \cdots - \frac{\rho_n}{\varepsilon} \frac{\partial \varepsilon}{\partial \rho_n} \right) \right] e^{-\beta U_{m_1,\ldots,m_n}}$$
$$(18)$$

(Note that terms of order $(|ERR|/\beta) + |E'| + |S'|$ in the arguments of the exponentials of the modified terms in square brackets of Eq. 17 have been ignored). Equation 18 is new, and interestingly enough, it is quite accurate sufficiently for large m, where from previous arguments, m is considered large enough when it is a few tens or so. This remains true even when the number of species types (n) is large (i.e., n>m for instance), since the arguments on surface and shape effects, continue to remain true.

In the case of the OMPP process, the above developments remain essentially the same except for minor changes which include changing the notation for PNNPDF to $g_{1,\ldots,m}^i(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ for the case where the particle at the origin is of species i. Also, N_i is written as $N_i - 1$, and the potential energy $U_{m_1,\ldots,m_n}$ is changed to $U_{i,m_1,\ldots,m_n}$ where the additional subscript of i indicates the particle (species i) at the origin also contributes to the potential energy. [The configurational contribution to the free energy A in Eq. 17 becomes that for N_1 (species 1) particles, $\ldots$, N_{i-1} (species $i-1$) particles, $N_i - 1$ (species i) particles, N_{i+1} (species $i+1$) particles, $\ldots$, N_n (species n) particles in a space of volume V, causing some modification of the normalization factor $h_{k_1 1,\ldots,k_m m}$].

Since the PNNPDF of Eq. 18 is accurately determined for large m ($m \gtrsim 10$), hence a fairly elaborate integration scheme must be embarked upon in the attempt to obtain for instance the normalization constant or the joint PDF of fewer variables in the set $\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m}$. This is further exacerbated (especially in determining the normalization constant) by the fact that PNNPDF's do not normalize by a straightforward integration, but according to the scheme given in Eq. 8. Several constants of the form $h_{k_1 1,\ldots,k_m m}$ are involved (of which there are n^m in number), and are referred to as "partial" normalization constants. This introduces a new level of difficulty in handling multi-component systems, which is otherwise absent in single component systems. This difficulty is however obviated for practical calculations by simulating "fairly" large multi-component systems, where m must not only be larger than n, but must be large enough such that it may be said to be accurate to employ $\alpha_1 m$ (species 1) particles, $\alpha_2 m$ (species 2) particles, $\ldots$, $\alpha_n m$ (species n) particles in the simulation, where $\alpha_1 m, \alpha_2 m, \ldots, \alpha_n m$ are each considerably larger than unity. $\alpha_1, \ldots, \alpha_n$ are the molar

fractions $\frac{N_1}{N}, \ldots, \frac{N_n}{N}$ respectively of the various species in the system. The larger the system considered, the more accurate it is to ignore fluctuations in the number of particles of different species in volume v_m. In fact, in many situations, we may expect it is sufficient to choose the number of particles of species i as $\sim \alpha_i m$ rather than the exact value of $\alpha_i m$. For such large systems, distributions are highly sharply peaked, and the region of phase space being ignored by not considering the variability in the number of particles of different species within volume v_m is sufficiently small and safely ignored. As the large system is therefore made to perform a "random walk" through phase space (such as by the "metropolis algorithm"), the ordering of particles (by species), according to distance from the origin, will change several times with the appropriate "weighting." (This actually amounts to considering simultaneously, the relevant set of PNNPDF's, which constitutes a relatively "small" set, with each PNNPDF having the same set of m_i values but different partial normalization constants). By integrating PNNPDF's for such large systems to obtain PNNPDF's of smaller m values, we may then expect to obtain accurate expressions for accurate study of Eq. 18 (including studies of local partial density fluctuations for different species).

In the foregoing discussion, coordinates which describe the spatial configuration of particles has been of interest. However, coordinates associated with the internal degrees of freedom (including particle momenta or velocities) may also be readily incorporated into the formulations.

4 Applications to weakly interacting multi-component systems

Two types of multi-component systems shall be treated in this section. The first shall be the poisson non-interacting multi-component particle (PNMP) system; and the second shall be the low density interacting gas of hard particle mixtures (HPM). For each kind of system, PNNPDF's, partial m-tuplet distribution functions, and the singlet cluster probability distribution (see Sect. 1) are derived.

For the PNMP system, we have that $U_{m_1,\ldots,m_n} = 0$ and Edgal and Huber [6b] indicates $\varepsilon = 1$. Hence the PNNPDF of Eq. 18 yields:

$$
g_{1,\ldots,m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m}) \prod_{j=1}^{m} d\vec{r}_{k_j j}
$$

$$
= h_{k_1 1,\ldots,k_m m} \exp\left[-\frac{4}{3}\pi r_{k_m m}^3 \rho\right] r_{k_1 1}^2 \sin\theta_{k_1 1} \cdots r_{k_m m}^2
$$

$$
\times \sin\theta_{k_m m} \prod_{j=1}^{m} dr_{k_j j} d\theta_{k_j j} d\phi_{k_j j} \tag{19}
$$

This is exactly Eq. 14 of Edgal and Huber [6b] with the normalization constant $h_{k_1 1,\ldots,k_m m} = \rho_{k_1} \cdots \rho_{k_m}$. In Eq. 19, integrating $r_{k_1 1}$ from 0 to $r_{k_2 2}$, integrating $r_{k_2 2}$ from 0 to $r_{k_3 3}, \ldots$, integrating $r_{k_{s_1-1}(s_1-1)}$ from 0 to $r_{k_{s_1} s_1}$ implies we are integrating out $r_{k_1 1}, r_{k_2 2}, \ldots, r_{k_{s_1-1}(s_1-1)}$, and this replaces the term of

$$r_{k_1 1}^2 \cdots r_{k_{s_1-1}(s_1-1)}^2 \prod_{j=1}^{s_1-1} dr_{k_j j}$$

in Eq. 19 with the factor $\frac{1}{3^{s_1-1}(s_1-1)!} r_{k_{s_1} s_1}^{3(s_1-1)}$. Next we integrate $r_{k_{s_1+1}(s_1+1)}$ from $r_{k_{s_1} s_1}$ to $r_{k_{s_1+2}(s_1+2)}$, we integrate $r_{k_{s_1+2}(s_1+2)}$ from $r_{k_{s_1} s_1}$ to $r_{k_{s_1+3}(s_1+3)}$, $\ldots$,we integrate $r_{k_{s_2-1}(s_2-1)}$ from $r_{k_{s_1} s_1}$ to $r_{k_{s_2} s_2}$, and this replaces the term of

$$r_{k_{s_1+1}(s_1+1)}^2 \cdots r_{k_{s_2-1}(s_2-1)}^2 \prod_{j=s_1+1}^{s_2-1} dr_{k_j j}$$

in Eq. (19) with

$$\frac{1}{(s_2 - s_1 - 1)!} \frac{1}{3^{s_2-s_1-1}} (r_{k_{s_2} s_2}^3 - r_{k_{s_1} s_1}^3)^{s_2-s_1-1}.$$

Observe that all integrations begin from $r_{k_{s_1} s_1}$. This is because we are keeping the variable $r_{k_{s_1} s_1}$ fixed. This process is also said to integrate out $r_{k_{s_1+1}(s_1+1)}, \ldots,$ $r_{k_{s_2-1}(s_2-1)}$. We similarly integrate out $r_{k_{s_2+1}(s_2+1)}, \ldots, r_{k_{s_3-1}(s_3-1)}$, with the understanding that the variable $r_{k_{s_2} s_2}$ is kept fixed. This process is continued until we finally integrate out $r_{k_{s_q-1+1}(s_{q-1}+1)}, \ldots, r_{k_{s_q-1}(s_q-1)}$ with the understanding that the variable $r_{k_{s_q-1} s_{q-1}}$ is kept fixed, resulting in replacement of the term of

$$r_{k_{s_{q-1}+1}(s_{q-1}+1)}^2 \cdots r_{k_{s_q-1}(s_q-1)}^2 \prod_{j=s_{q-1}+1}^{s_q-1} dr_{k_j j}$$

in Eq. 19 with

$$\frac{1}{(s_q - s_{q-1} - 1)!} \frac{1}{3^{s_q-s_{q-1}-1}} (r_{k_{s_q} s_q}^3 - r_{k_{s_{q-1}} s_{q-1}}^3)^{s_q-s_{q-1}-1}.$$

We equate s_q to m, and thus the right hand side of Eq. 19 becomes

$$h_{k_1 1, \ldots, k_{s_q} s_q} \frac{(r_{k_{s_1} s_1}^3 - r_{k_{s_0} s_0}^3)^{s_1-s_0-1} r_{k_{s_1} s_1}^2}{3^{s_1-s_0-1}(s_1 - s_0 - 1)!} \cdots \frac{(r_{k_{s_q} s_q}^3 - r_{k_{s_{q-1}} s_{q-1}}^3)^{s_q-s_{q-1}-1} r_{k_{s_q} s_q}^2}{3^{s_q-s_{q-1}-1}(s_q - s_{q-1} - 1)!}$$

$$\times \exp\left(-\frac{4}{3}\pi r_{k_{s_q} s_q}^3 \rho\right) \prod_{j=1}^{q} dr_{k_{s_j} s_j} \prod_{l=1}^{s_q} \sin\theta_{k_l l} d\theta_{k_l l} d\phi_{k_l l}$$

(It is assumed $s_0 = 0$ and $r_{k_{s_0} s_0} = 0$). Integrating out the $\theta_{k_l l'} s$ and $\phi_{k_l l'} s$ (except $\theta_{k_{s_1} s_1}, \ldots, \theta_{k_{s_q} s_q}$, and $\phi_{k_{s_1} s_1}, \ldots, \phi_{k_{s_q} s_q}$), then results in Eq. 19 becoming

$$
\int_{\Omega_{s_q-q}} \left[\int_0^{r_{k_{s_q} s_q}} \cdots \int_0^{r_{k_{s_q-1}+2(s_q-1+2)}} \cdots \int_0^{r_{k_{s_2} s_2}} \cdots \int_0^{r_{k_{s_1}+2(s_1+2)}} \int_0^{r_{k_{s_1} s_1}} \cdots \int_0^{r_{k_2 2}} \right.
$$

$$
\left. g_{1,\ldots,s_q} \left(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_{s_q} s_q} \right) \prod_{i=1}^{s_1-1} dr_{k_i i} \prod_{j=s_1+1}^{s_2-1} dr_{k_j j} \cdots \prod_{l=s_{q-1}+1}^{s_q-1} dr_{k_l l} \right] d\Omega_{s_q-q}
$$

$$
= \left[\prod_{l=s_0+1}^{s_1-1} (4\pi\rho_{k_l}) \cdots \cdots \prod_{l=s_{q-1}+1}^{s_q-1} (4\pi\rho_{k_l}) \right] \prod_{l=1}^{q} \left(4\pi\rho_{k_{s_l}} \right)
$$

$$
\times \prod_{j=1}^{q} \left(\frac{(r_{k_{s_j} s_j}^3 - r_{k_{s_{j-1}} s_{j-1}}^3)^{s_j-s_{j-1}-1} r_{k_{s_j} s_j}^2}{4\pi 3^{s_j-s_{j-1}-1} (s_j - s_{j-1} - 1)!} \right)
$$

$$
\times e^{-\frac{4}{3}\pi r_{k_{s_q} s_q}^3 \rho} \prod_{j=1}^{q} \sin\theta_{k_{s_j} s_j} dr_{k_{s_j} s_j} d\theta_{k_{s_j} s_j} d\phi_{k_{s_j} s_j} \tag{20}
$$

Ω_{s_q-q} stands for the domain of all the angular coordinates that are integrated out. Now, the number of coordinates that have been integrated out in Eq. 19 from among the set of coordinates $\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_{s_q} s_q}$ is (s_q-q). Each of these coordinates is associated with any of the n species of particles. Hence there are n^{s_q-q} possible set of distinct species assignments that can be made to the coordinates that have been integrated out. Except for the factor of $h_{k_1 1,\ldots,k_m m}$ in Eq. 19, the functional dependence on these coordinates are all identical (and are independent of species type). Hence except for the first term in square brackets, all terms on the right hand side of Eq. 20 remain the same as we consider different assignment of species to the coordinates. If we add up all the first terms (in square brackets) that we get as we consider all the n^{s_q-q} possible set of distinct species assignments, we get $(4\pi)^{s_q-q}(\rho_1 + \cdots + \rho_n)^{s_q-q} = (4\pi\rho)^{s_q-q}$. This result is easily proved if we consider small values of n and (s_q-q) (for instance $1, 2, \ldots$). Hence we conclude that the marginal pdf for the event that the s_1th nearest neighbor is of species k_1 at $\vec{r}_{k_1 1}, \ldots$, the s_qth nearest neighbor is of species k_q at $\vec{r}_{k_q q}$ is:

$$
g_{s_1,\ldots,s_q} \left(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_q q} \right) \prod_{j=1}^{q} d\vec{r}_{k_j j}
$$

$$
= (4\pi\rho)^{s_q-q} \prod_{j=1}^{q} \left(\frac{(r_{k_j j}^3 - r_{k_{j-1}(j-1)}^3)^{s_j-s_{j-1}-1} r_{k_j j}^2 \rho_{k_{s_j}}}{3^{s_j-s_{j-1}-1} (s_j - s_{j-1} - 1)!} \right)
$$

$$
\times \exp\left(-\frac{4}{3}\pi r_{k_q q}^3 \rho \right) \prod_{j=1}^{q} (\sin\theta_{k_j j} dr_{k_j j} d\theta_{k_j j} d\phi_{k_j j}) \tag{21}
$$

Observe that the coordinate $\vec{r}_{k_{s_i} s_i}$ has been written as $\vec{r}_{k_i i}$ to emphasize the location of the s_ith nearest neighbor, rather than the fact that there are $(s_i - 1)$ other neighbors nearer to the origin (See Eq. 9). We now address the normalizability of the PNNPDF's of Eq. 19 and Eq. 21 under the scheme of Eq. 8. In the case of Eq. 21 we can first integrate out the angular variables to obtain a factor of $(4\pi)^q$. Next, we can interchange the sums and integrals specified in Eq. 8, if functional dependencies on variables of integration are identical for each term of the sum. But under the sum, we have that all terms in Eq. 21 are identical except the factor of $\rho_{k_{s_j}}$. Hence we obtain an additional factor given as $\sum_{(j_q)} \left(\rho_{k_{s_1}} \cdots \rho_{k_{s_q}} \right)$, and this sums to $(\rho_1 + \cdots + \rho_n)^q = \rho^q$. We are therefore left with integration over the radial coordinates of the r.h.s. of Eq. 21 which has now become (after the summation) the expression:

$$f = (4\pi\rho)^{s_q} \prod_{j=1}^{q} \left(\frac{(r_{k_j j}^3 - r_{k_{j-1}(j-1)}^3)^{s_j - s_{j-1} - 1} r_{k_j j}^2}{3^{s_j - s_{j-1} - 1}(s_j - s_{j-1} - 1)!} \right) \exp\left(-\frac{4}{3}\pi r_{k_q q}^3 \rho \right) \prod_{j=1}^{q} dr_{k_j j}$$

(22)

$(r_{k_0 0} = 0)$. We integrate out all radial coordinates by typically varying $r_{k_i i}$ from 0 to $r_{k_{(i+1)}(i+1)}$, while the last radial coordinate $r_{k_q q}$ is varied from 0 to ∞. This integration is an extremely formidable one to perform for general q; but it is readily shown to yield the result of unity (as required) for the first several values of q (e.g., $q = 1, 2, \ldots$). For instance, for $q = 1$, we get:

$$\int_0^{\infty} (4\pi\rho)^{s_1} \left(\frac{r_{k_1 1}^{3(s_1 - 1)} r_{k_1 1}^2}{3^{s_1 - 1}(s_1 - 1)!} \right) \exp\left(-\frac{4}{3}\pi r_{k_1 1}^3 \rho \right) dr_{k_1 1} = 1.$$

For $q = 2$, we get:

$$\int_0^{\infty} \int_0^{r_{k_2 2}} (4\pi\rho)^{s_2} \left(\frac{r_{k_1 1}^{3(s_1 - 1)} r_{k_1 1}^2}{3^{s_1 - 1}(s_1 - 1)!} \right) \left(\frac{(r_{k_2 2}^3 - r_{k_1 1}^3)^{s_2 - s_1 - 1} r_{k_2 2}^2}{3^{s_2 - s_1 - 1}(s_2 - s_1 - 1)!} \right)$$

$$\times \exp\left(-\frac{4}{3}\pi r_{k_2 2}^3 \rho \right) dr_{k_1 1} dr_{k_2 2}$$

$$= \int_0^{\infty} (4\pi\rho)^{s_2} \left(\frac{r_{k_2 2}^{3(s_2 - 1)}}{3^{s_1 - 1}(s_1 - 1)!} \right) \left(\frac{r_{k_2 2}^2}{3^{s_2 - s_1 - 1}(s_2 - s_1 - 1)!} \right)$$

$$\times \left(\int_0^1 y^{s_1 - 1}(1 - y)^{s_2 - s_1 - 1} dy \right) \exp\left(-\frac{4}{3}\pi r_{k_2 2}^3 \rho \right) dr_{k_2 2}$$

$$= \int_0^{\infty} (4\pi\rho)^{s_2} \left(\frac{r_{k_2 2}^{3(s_2 - 1)} r_{k_2 2}^2}{3^{s_2 - 1}(s_2 - 1)!} \right) \exp\left(-\frac{4}{3}\pi r_{k_2 2}^3 \rho \right) dr_{k_2 2} = 1$$

Hence we easily conclude that the PNNPDF of Eq. 21 normalizes adequately. But Eq. 19 is a special case of Eq. 21 where $s_1 = 1, \ldots, s_q = q$ (and q is set equal to m). Hence Eq. 19 also normalizes adequately as has been earlier shown in Edgal and Huber [6b].

Next, we determine the m-tuplet distribution function from PNNPDF's for the PNMP system. This is done by substituting Eq. 21 into Eq. 9 (assuming m = q) to get:

$$
f_{m_1,\ldots,m_n}\left(\vec{r}_{k_1 1},\ldots,\vec{r}_{k_m m}\right)\prod_{j=1}^{m} d\vec{r}_{k_j j}
$$

$$
= \sum_{s_1}\cdots\sum_{s_m}(4\pi\rho)^{s_m - m}\prod_{j=1}^{m}\left(\frac{(r_{k_j j}^3 - r_{k_{j-1}(j-1)}^3)^{s_j - s_{j-1} - 1}r_{k_j j}^2\rho_{k_{s_j}}}{3^{s_j - s_{j-1} - 1}(s_j - s_{j-1} - 1)!}\right)
$$

$$
\times \exp\left(-\frac{4}{3}\pi r_{k_m m}^3\rho\right)\prod_{j=1}^{m}(\sin\theta_{k_j j}dr_{k_j j}d\theta_{k_j j}d\phi_{k_j j}) \tag{23}
$$

This summation is also a formidable one to appropriately set up and perform for general n and m. However, it may readily be done for the first several values of n and m. For n = 1 (single component case) we have that Eq. 23 becomes:

$$
f_m\left(\vec{r}_1,\ldots,\vec{r}_m\right)\prod_{j=1}^{m}d\vec{r}_j = \sum_{s_1=1}^{s_2-1}\sum_{s_2=2}^{s_3-1}\cdots\sum_{s_{m-1}=m-1}^{s_m-1}\sum_{s_m=m}^{M}(4\pi\rho)^{s_m - m}
$$

$$
\times\prod_{j=1}^{m}\left(\frac{(r_j^3 - r_{(j-1)}^3)^{s_j - s_{j-1} - 1}r_j^2\rho}{3^{s_j - s_{j-1} - 1}(s_j - s_{j-1} - 1)!}\right)
$$

$$
\times \exp\left(-\frac{4}{3}\pi r_m^3\rho\right)\prod_{j=1}^{m}(\sin\theta_j dr_j d\theta_j d\phi_j) \tag{24}
$$

(The upper summation index M is later made to tend to ∞). Performing the multiple sum of Eq. 24 is quite deceptively formidable. The case of m = 1 has already been successfully done in Edgal [10a]. For m = 2, Eq. 24 is rewritten as:

$$
f_2\left(\vec{r}_1,\vec{r}_2\right) = \sum_{s_1=1}^{s_2-1}\sum_{s_2=2}^{M}(4\pi\rho)^{s_2-2}\prod_{j=1}^{2}\left(\frac{(r_j^3 - r_{(j-1)}^3)^{s_j - s_{j-1} - 1}\rho}{3^{s_j - s_{j-1} - 1}(s_j - s_{j-1} - 1)!}\right)
$$

$$
\times \exp\left(-\frac{4}{3}\pi r_2^3\rho\right) \tag{25}
$$

The sum in Eq. 25 is performed by fixing values of s_1, then performing the sum over s_2. (This is easily understood by preparing a table of values of s_1, s_2 pairs). For instance, for $s_1 = 1$, the sum over s_2 becomes:

$$
e^{-\frac{4}{3}\pi r_2^3\rho}\rho\sum_{s_2=2}^{M}\frac{(4\pi\rho)^{s_2-2}(r_2^3 - r_1^3)^{s_2-2}\rho}{3^{s_2-2}(s_2-2)!} = \rho^2 e^{-\frac{4}{3}\pi r_2^3\rho}\sum_{s_2=0}^{M-2}\frac{\left(\frac{4}{3}\pi\rho(r_2^3 - r_1^3)\right)^{s_2}}{s_2!}
$$

For $s_1 = 2$, the sum over s_2 becomes:

$$e^{-\frac{4}{3}\pi r_2^3 \rho} \frac{r_1^3 \rho}{(3)1!} \rho(4\pi\rho) \sum_{s_2=3}^{M} \frac{(4\pi\rho)^{s_2-3}(r_2^3 - r_1^3)^{s_2-3}}{3^{s_2-3}(s_2-3)!}$$

$$= \frac{r_1^3}{(3)1!} \rho^2 (4\pi\rho) e^{-\frac{4}{3}\pi r_2^3 \rho} \sum_{s_2=0}^{M-3} \frac{\left(\frac{4}{3}\pi\rho(r_2^3 - r_1^3)\right)^{s_2}}{s_2!}$$

etc. Hence the multiple sum of Eq. 25 is finally rewritten as:

$$f_2\left(\vec{r}_1, \vec{r}_2\right) = \rho^2 e^{-\frac{4}{3}\pi r_2^3 \rho} \sum_{s_1=0}^{M-2} \frac{\left(\frac{4}{3}\pi\rho r_1^3\right)^{s_1}}{s_1!} \sum_{s_2=0}^{M-s_1-2} \frac{\left(\frac{4}{3}\pi\rho(r_2^3 - r_1^3)\right)^{s_2}}{s_2!} \qquad (26)$$

Results for multiple sums are not common in books on mathematical tables and formulas; however, we notice that the two arguments in the sums are terms for typical exponential series. If convergence is not a problem, we can truncate the sums and yet obtain highly accurate results for Eq. 26. For instance, the first sum can be truncated at $S_1 = M_1$ where $M_1 \ll M$, but both $M_1, M \to \infty$. Hence Eq. 26 is rewritten as:

$$f_2\left(\vec{r}_1, \vec{r}_2\right) = \lim_{\substack{M_1, M \to \infty \\ (M_1 \ll M)}} \rho^2 e^{-\frac{4}{3}\pi r_2^3 \rho} \sum_{s_1=0}^{M_1} \frac{\left(\frac{4}{3}\pi\rho r_1^3\right)^{s_1}}{s_1!} \sum_{s_2=0}^{M-s_1-2} \frac{\left(\frac{4}{3}\pi\rho(r_2^3 - r_1^3)\right)^{s_2}}{s_2!}$$

$$\to \lim_{M_1 \to \infty} \rho^2 e^{-\frac{4}{3}\pi\rho r_2^3} \sum_{s_1=0}^{M_1} \frac{\left(\frac{4}{3}\pi\rho r_1^3\right)^{s_1}}{s_1!} e^{\frac{4}{3}\pi\rho(r_2^3 - r_1^3)}$$

$$\to \rho^2 e^{-\frac{4}{3}\pi\rho r_2^3} e^{\frac{4}{3}\pi\rho(r_2^3 - r_1^3)} e^{\frac{4}{3}\pi\rho r_1^3} = \rho^2 \qquad (27)$$

This is the correct result for m-tuplet distribution function for a PNMP system for n = 1, m = 2 (see Eq. 10). For n = 2 (binary multi-component system) and m = 1, there are two kinds of m-tuplet distribution functions involved; and these are $f_{1,0}(\vec{r}_{11})$ and $f_{0,1}(\vec{r}_{21})$ involving species 1 and species 2 respectively. Using Eq. 23, we then have

$$f_{1,0}(\vec{r}_{11})d\vec{r}_{11} = \sum_{s_1=1}^{M} (4\pi\rho)^{s_1-1} \frac{\left(r_{11}^3\right)^{s_1-1} r_{11}^2 \rho_1}{3^{s_1-1}(s_1-1)!} e^{-\frac{4}{3}\pi\rho r_{11}^3} \left(\sin\theta_{11} dr_{11} d\theta_{11} d\phi_{11}\right)$$

$$(28)$$

For $f_{0,1}(\vec{r}_{21})$, we simply replace in Eq. 28, $\rho_1, r_{11}, \theta_{11}, \phi_{11}, \vec{r}_{11}$ by $\rho_2, r_{21}, \theta_{21}, \phi_{21},$ $\vec{r}_{21}$ respectively. The sum of Eq. 28 readily yields the correct result of ρ_1 for $f_{1,0}(\vec{r}_{11})$ (while that for $f_{0,1}(\vec{r}_{21})$ yields the correct result of ρ_2).

For n = 2, m = 2, four kinds of m-tuplet distribution functions are involved, viz: $f_{2,0}(\vec{r}_{11}, \vec{r}_{12})$, $f_{0,2}(\vec{r}_{21}, \vec{r}_{22})$, $f_{1,1}(\vec{r}_{11}, \vec{r}_{22})$, $f_{1,1}(\vec{r}_{21}, \vec{r}_{12})$. From Eq. 23, we have

for instance:

$$f_{2,0}(\vec{r}_{11}, \vec{r}_{12}) \prod_{j=1}^{2} d\vec{r}_{1j} = \sum_{s_1=1}^{s_2-1} \sum_{s_2=2}^{M} (4\pi\rho)^{s_2-2} \prod_{j=1}^{2} \left(\frac{(r_{1j}^3 - r_{1(j-1)}^3)^{s_j-s_{j-1}-1} r_{1j}^2 \rho_1}{3^{s_j-s_{j-1}-1}(s_j - s_{j-1} - 1)!} \right)$$

$$\times \exp\left(-\frac{4}{3}\pi r_{12}^3 \rho\right) \prod_{j=1}^{2} (\sin\theta_{1j} dr_{1j} d\theta_{1j} d\phi_{1j}) \qquad (29)$$

Using the same manipulation employed to evaluate Eq. 25, we easily get $f_{2,0}(\vec{r}_{11}, \vec{r}_{12}) = \rho_1^2$. Similarly $f_{0,2}(\vec{r}_{21}, \vec{r}_{22})$ is also obtained as ρ_2^2. On the other hand $f_{1,1}(\vec{r}_{11}, \vec{r}_{22})$ is formulated using Eq. 23 as:

$$f_{1,1}(\vec{r}_{11}, \vec{r}_{22}) d\vec{r}_{11} d\vec{r}_{22} = \sum_{s_1=1}^{s_2-1} \sum_{s_2=2}^{M} (4\pi\rho)^{s_2-2}$$

$$\times \prod_{j=1}^{2} \left(\frac{(r_{jj}^3 - r_{(j-1)(j-1)}^3)^{s_j-s_{j-1}-1} r_{jj}^2 \rho_j}{3^{s_j-s_{j-1}-1}(s_j - s_{j-1} - 1)!} \right)$$

$$\times \exp\left(-\frac{4}{3}\pi r_{22}^3 \rho\right) \prod_{j=1}^{2} (\sin\theta_{jj} dr_{jj} d\theta_{jj} d\phi_{jj}) \quad (30)$$

Hence, manipulating this equation similar to the manipulation employed for Eq. 25 yields the correct value of $\rho_1\rho_2$ for $f_{1,1}(\vec{r}_{11}, \vec{r}_{22})$ (and a similar result is also obtained for $f_{1,1}(\vec{r}_{21}, \vec{r}_{12})$)

Finally for the PNMP system, we consider singlet clustering. The basic model involved assumes particles of different species are introduced into a medium according to a random poisson point process. Their configuration (depicting that of an equilibrium PNMP system) is then "frozen" into place, and then the particles assume a spherical structure (with different particle species having different spherical radii). The spheres of the particles will overlap if in close enough proximity to one another. The above system is reminiscent of monovalent donor and acceptor impurities introduced into a semiconductor material by "diffusion" or "ion-implantation" process. After impurities are frozen into their locations, donor "electrons" (acceptor "holes") are assumed to occupy a spherical region described about each donor (acceptor) particle acting as centers of the spheres. The radius of each sphere being one "Bohr" radius, and differing for different impurity type. Non-overlapping spheres are termed singlet clusters and may be used to determine to first order the density of impurity levels in semiconductors; while overlapping spheres of the same or different species, form complexes said to generally lead to "deep centers" in semiconductors. (see Edgal and Wiley) [12]. For simplicity, we consider a 2 species situation where the radius of the spheres are a_1, a_2 (for species 1 and 2 respectively) with $a_1 < a_2$. To determine the singlet probability of each species, we consider the following mutually exclusive events. We use the OMPP distribution for m nearest neighbors. Later in the analysis we obtain results in the limit as m tends to infinity. For the case where species 1 is at the origin, we

consider the first event which involves the fact that a particle of species 2 is the first nearest neighbor, while all the other (m−1) nearest neighbors may be of any species type. The second event assumes a particle of species 1 is the first nearest neighbor, a particle of species 2 is the second nearest neighbor, while all the other (m−2) nearest neighbors may be of any species type. The third event assumes particles of species 1 constitute the first and second nearest neighbors, while a particle of species 2 is the 3rd nearest neighbor, and all the other (m−3) nearest neighbors may be of any species type. This continues similarly until the m-th event which assumes particles of species 1 constitute the 1st, 2nd,…, (m−1)th nearest neighbors, while a particle of species 2 is the m-th nearest neighbor. The PNNPDF appropriate for the first event is $g^1(\vec{r}_{21})$; the PNNPDF appropriate for the second event is $g^1(\vec{r}_{11}, \vec{r}_{22})$; the PNNPDF appropriate for the third event is $g^1(\vec{r}_{11}, \vec{r}_{12}, \vec{r}_{23})$; … ; the PNNPDF appropriate for the m-th event is $g^1(\vec{r}_{11}, \vec{r}_{12}, \dots, \vec{r}_{1(m-1)}, \vec{r}_{2m})$. Hence the singlet probability for species 1 is obtained from the mutually exclusive events as:

$$
\begin{aligned}
C_1(1, \rho_1, \rho_2) = \lim_{m\to\infty} & \left[\int_{a_1+a_2}^{\infty} g^1(r_{21})\, dr_{21} + \int_{a_1+a_2}^{\infty} \int_{2a_1}^{r_{22}} g^1(r_{11}, r_{22})\, dr_{11} dr_{22} \right. \\
& + \int_{a_1+a_2}^{\infty} \int_{2a_1}^{r_{23}} \int_{2a_1}^{r_{12}} g^1(\vec{r}_{11}, \vec{r}_{12}, \vec{r}_{23}) dr_{11} dr_{12} dr_{23} \\
& + \cdots + \int_{a_1+a_2}^{\infty} \int_{2a_1}^{r_{1(m-1)}} \cdots \int_{2a_1}^{r_{13}} \int_{2a_1}^{r_{12}} \\
& \left. \times g^1(r_{11}, r_{12}, \dots, r_{1(m-1)}, r_{2m})\, dr_{11} dr_{12} \cdots dr_{1(m-1)} dr_{2m} \right] \\
= \lim_{m\to\infty} & \left[4\pi\rho_2 \int_{a_1+a_2}^{\infty} r_{21}^2 e^{-\frac{4}{3}\pi\rho r_{21}^3}\, dr_{21} \right. \\
& + (4\pi)^2 \rho_1\rho_2 \int_{a_1+a_2}^{\infty} \int_{2a_1}^{r_{22}} r_{11}^2 r_{22}^2 e^{-\frac{4}{3}\pi\rho r_{22}^3}\, dr_{11} dr_{22} \\
& + \cdots + (4\pi)^m \rho_1^{(m-1)} \rho_2 \int_{a_1+a_2}^{\infty} \int_{2a_1}^{r_{2m}} \cdots \int_{2a_1}^{r_{13}} \int_{2a_1}^{r_{12}} \\
& \left. \times \left(r_{11}^2 r_{12}^2 \cdots r_{1(m-1)}^2 r_{2m}^2 \right) e^{-\frac{4}{3}\pi\rho r_{2m}^3}\, dr_{11} \cdots dr_{1(m-1)} dr_{2m} \right]
\end{aligned}
$$

$$
\begin{aligned}
= \lim_{m \to \infty} \Bigg[&\left\{ \alpha_1^0 \alpha_2 e^{-\hat{\eta}_{12}} \right\} + \left\{ \alpha_1^1 \alpha_2 (1 + (\hat{\eta}_{12} - \hat{\eta}_1)) e^{-\hat{\eta}_{12}} \right\} \\
&+ \left\{ \alpha_1^2 \alpha_2 (1 + (\hat{\eta}_{12} - \hat{\eta}_1) + \frac{1}{2!}(\hat{\eta}_{12} - \hat{\eta}_1)^2) e^{-\hat{\eta}_{12}} \right\} \\
&+ \left\{ \alpha_1^3 \alpha_2 (1 + (\hat{\eta}_{12} - \hat{\eta}_1) + \frac{1}{2!}(\hat{\eta}_{12} - \hat{\eta}_1)^2 + \frac{1}{3!}(\hat{\eta}_{12} - \hat{\eta}_1)^3) e^{-\hat{\eta}_{12}} \right\} \\
&+ \cdots + \cdots + \left\{ \alpha_1^m \alpha_2 (1 + (\hat{\eta}_{12} - \hat{\eta}_1) + \frac{1}{2!}(\hat{\eta}_{12} - \hat{\eta}_1)^2 \right. \\
&\left. + \cdots + \frac{1}{m!}(\hat{\eta}_{12} - \hat{\eta}_1)^m) e^{-\hat{\eta}_{12}} \right\} \Bigg]
\end{aligned}
\tag{31}
$$

where $\hat{\eta}_1 = \frac{4}{3}\pi\rho(2a_1)^3$, $\hat{\eta}_{12} = \frac{4}{3}\pi\rho(a_1 + a_2)^3$. The terms in the above expression can be regrouped as follows:

$$
\begin{aligned}
C_1(1, \rho_1, \rho_2) = \alpha_2 e^{-\hat{\eta}_{12}} \Bigg[&\frac{\alpha_1^0}{0!}(\hat{\eta}_{12} - \hat{\eta}_1)^0 (1 + \alpha_1^1 + \alpha_1^2 + \alpha_1^3 + \cdots) \\
&+ \frac{\alpha_1^1}{1!}(\hat{\eta}_{12} - \hat{\eta}_1)^1 (1 + \alpha_1^1 + \alpha_1^2 + \alpha_1^3 + \cdots) \\
&+ \frac{\alpha_1^2}{2!}(\hat{\eta}_{12} - \hat{\eta}_1)^2 (1 + \alpha_1^1 + \alpha_1^2 + \alpha_1^3 + \cdots) + \cdots \Bigg] \\
= \alpha_2 e^{-\hat{\eta}_{12}} &(1 + \alpha_1^1 + \alpha_1^2 + \alpha_1^3 + \cdots) \sum_{i=0}^{\infty} \frac{\alpha_1^i (\hat{\eta}_{12} - \hat{\eta}_1)^i}{i!} \\
= e^{-\hat{\eta}_{12}} e^{\alpha_1(\hat{\eta}_{12} - \hat{\eta}_1)} &= e^{-\frac{4}{3}\pi\rho_2(a_1+a_2)^3} e^{-\frac{4}{3}\pi\rho_1(2a_1)^3}
\end{aligned}
\tag{32}
$$

This is the result expected. Note that the above final result for $C_1(1, \rho_1, \rho_2)$, could also have been obtained by more direct (and easier) method using the well known poisson distribution (for the occupation of volume elements by some fixed number of particles). Such method is possible because of the *independence* of the distribution of each particle in the system. However, such independence is absent in the interaction problem; hence the above method has more general applicability. In a similar fashion, the singlet probability for species 2 can be obtained by using the same mutually exclusive events as above. Noting that a particle of species 2 (radius a_2) is now at the origin, we obtain the same result except that we replace $(a_1 + a_2)$ by $2a_2$, and $2a_1$ by $(a_1 + a_2)$. Hence we get:

$$
C_2(1, \rho_1, \rho_2) = e^{-\frac{4}{3}\pi\rho_2(2a_2)^3} e^{-\frac{4}{3}\pi\rho_1(a_1+a_2)^3}
\tag{33}
$$

In Fig. 2, we show a 2-component system involving randomly distributed spheres (different components having different radii).

Finally, we consider the low density HPM system, involving a binary mixture of hard spheres. In Edgal and Huber [6b], results are given for this system for the parameter
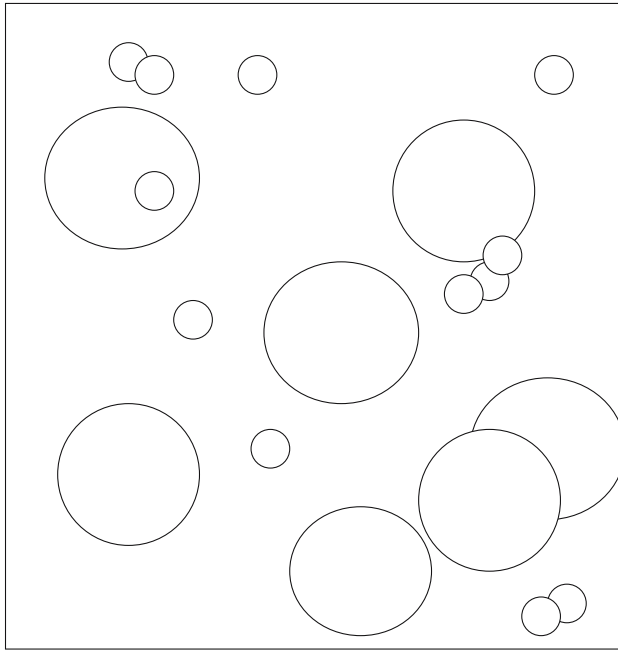
Fig. 2 Poisson distributed two-component overlapping spheres (showing four singlet clusters of the smaller spheres, and three singlet clusters of the larger spheres)

ε. Hence we easily obtain the PNNPDF for the GMPP process in this case (using Eq. 18) as:

$$
\begin{aligned}
g_{1,\ldots,m}(\vec{r}_{k_11},\ldots,\vec{r}_{k_mm}) &= h_{k_11,\ldots,k_mm} \\
&\times \exp\left[-\frac{4}{3}\pi r_{k_mm}^3\rho\left(1+\rho\left(\frac{2}{3}\pi(2b_1)^3\alpha_1^2+\frac{2}{3}\pi(2b_2)^3\alpha_2^2+\frac{4}{3}\pi(b_1+b_2)^3\alpha_1\alpha_2\right)\right)\right] \\
&\times F\left(\vec{r}_{k_11},\ldots,\vec{r}_{k_mm}\right)
\end{aligned}
\tag{34}
$$

b_1, b_2 are the radii of the hard spheres, $F\left(\vec{r}_{k_11},\ldots,\vec{r}_{k_mm}\right)$ is an indicator function used to replace the second exponential in Eq. 18. This function is unity when no two of the m-nearest neighbor hard spheres overlap, otherwise it is zero.

In principle, all n^m PNNPDF's (each having an unknown partial normalization constant) are needed to fully carry on a successful investigation of multi-component systems. However, as earlier argued, we can also use Eq. 34 with a large enough m for a successful investigation (where $m_1 \sim \alpha_1 m$, $m_2 \sim \alpha_2 m$, ..., $m_n \sim \alpha_n m$). A difficulty with this nonetheless, is the highly formidable multiple integrals and multiple sums that will be encountered. Hence, invariably, the interaction multi-component problem is best dealt with by computer simulation, except in cases where special interactions or special situations are considered. In the very dilute limit of the binary hard sphere gas, we assume we may approximate the indicator function (in Eq. 34) as unity "almost always," and this will enable us readily perform integrations on the PNNPDF's as in

the poisson case. Hence Eq. 34 can be integrated (as was done for the PNMP case) to yield the following marginal PDF:

$$
g_{s_1,\ldots,s_q}\left(\vec{r}_{k_1 1},\ldots,\vec{r}_{k_q q}\right)\prod_{j=1}^{q}d\vec{r}_{k_j j} \approx h_{k_1 s_1,\ldots,k_q s_q}\,(4\pi\rho)^{s_q-q}
$$

$$
\times \prod_{j=1}^{q}\left(\frac{(r_{k_j j}^3 - r_{k_{j-1}(j-1)}^3)^{s_j-s_{j-1}-1}r_{k_j j}^2 \rho_{k_{s_j}}}{3^{s_j-s_{j-1}-1}(s_j - s_{j-1} - 1)!}\right)
$$

$$
\times \exp\left[-\frac{4}{3}\pi r_{k_q q}^3 \rho\left(1 + \rho\left(\frac{2}{3}\pi(2b_1)^3\alpha_1^2\right.\right.\right.
$$

$$
\left.\left.\left.+\ \frac{2}{3}\pi(2b_2)^3\alpha_2^2 + \frac{4}{3}\pi(b_1+b_2)^3\alpha_1\alpha_2\right)\right)\right]F(\vec{r}_{k_1 1},\ldots,\vec{r}_{k_q q})
$$

$$
\times \prod_{j=1}^{q}(\sin\theta_{k_j j}dr_{k_j j}d\theta_{k_j j}d\phi_{k_j j}) \tag{35}
$$

Observe that the indicator function has been "artificially" re-introduced (after integration). Nevertheless, the approximation on the indicator function (during integration) "washes" out some of the features of the marginal pdf; causing for instance the marginal pdf to have largely similar features for both the GMPP and OMPP processes.

Performing a multiple sum on Eq. 35 similar to that of Eq. 23 is extremely difficult even when we consider small m, n values. Hence the present development in the formulation of m-tuplet distribution functions for the interaction problem shall involve arguments which largely depend on what is already understood about m-tuplet distribution functions. First we note that the term of $(4\pi\rho)^{s_q-q}$ in Eq. 35 may be seen as the normalization constant of the pdf $g_{s_1,\ldots,s_q}\left(\vec{r}_{k_1 1},\ldots,\vec{r}_{k_q q}\right)$ for the non interaction case, where $h_{k_1 s_1,\ldots,k_q s_q} = 1$. In the interaction case $h_{k_1 s_1,\ldots,k_q s_q}$ will vary with q and the species values $k_1,\ldots,k_q$. We can write $h_{k_1 s_1,\ldots,k_q s_q}(4\pi\rho)^{s_q-q} = (4\pi\rho c)^{s_q-q}$ where c is also a function of q and $k_1,\ldots,k_q$. The values $h_{k_1 s_1,\ldots,k_q s_q}$ and c can be studied in some fashion. (In the non interaction case, c = 1). Now, the indicator function causes the magnitude of the p.d.f. of Eq. 35 to be smaller than that of the non interaction case close to the origin. Also, the exponential of Eq. 35 behaves like that of the non interaction case, but with a higher effective density $\tilde{\rho} = \rho(1 + 4\eta_1\alpha_1 + 4\eta_2\alpha_2 + \eta_{12}\alpha_1\alpha_2)$, where $\eta_1 = \frac{4}{3}\pi b_1^3\rho_1$, $\eta_2 = \frac{4}{3}\pi b_2^3\rho_2$, $\eta_{12} = \frac{4}{3}\pi(b_1+b_2)^3\rho$. Thus the pdf of Eq. 35 decreases more sharply than that of the noninteraction case far from the origin. Hence Eq. 35 provides a more sharply peaked pdf (with a narrower width) than that of the non interaction case, and thus we must have $c, h_{k_1 s_1,\ldots,k_q s_q} > 1$ for appropriate normalization. The larger the value of q, the larger the volume of influence of the indicator function, and thus the tendency to generally consider larger values of $r_{k_q q}$. But large values of $r_{k_q q}$ imply stronger decay of the exponential function in Eq. 35. Hence the pdf of Eq. 35 gets narrower (and more sharply peaked) relative to the counterpart pdf for the non interaction case for larger q. This means $h_{k_1 s_1,\ldots,k_q s_q}$ is expected to increase monotonically with q. However, this does not prove whether c has to increase or decrease with q, because the larger the value of s_q, the smaller c gets (even though

we must still have c > 1), implying the presence of two competing effects on the value of c. If we apply the multiple sum of Eq. 23 to the pdf of Eq. 35, several s_q values are involved, and for simplicity, some average value of c (written as $\bar{c}$), applicable for all s_q values involved, may be used. We also have that $\bar{c} > 1$, and larger q will tend to increase $\bar{c}$, while the competing effect requires that the larger the number of terms in the multiple sum, the larger the maximum s_q value involved, and thus this tends to decrease $\bar{c}$ even though we must always have $\bar{c} > 1$. There are limits on s_q (and invariably on q also) as we consider different distances ($r_{k_q q}$) from the origin. For instance, if we consider a single component system with radius of hard core of particles being b_0, $r_{k_q q}$ will be written as r_q, and for $0 < r_q < b_0$, the maximum that s_q can be is necessarily 1. As r_q increases form b_0 to $2b_0$, the maximum that s_q can be increases "stepwise" from 2 to 13. (The value of 13 being obtained by considering the number of spheres that can be arranged as compactly as possible around a central sphere according to the hexagonal closed packed structure) [1], etc. Hence we see that $\bar{c}$ is necessarily a function of $r_{k_q q}$ or r_q (in general) written as $\bar{c}(r_{k_q q})$ or $\bar{c}(r_q)$, and this function can be studied "piece-wise" considering different ranges of $r_{k_q q}$ or r_q as above. In later developments, we shall use information already known about m-tuplet distribution functions, to determine the general behavior of $\bar{c}(r_{k_q q})$ or $\bar{c}(r_q)$. (Note that $\bar{c}$ may be seen as a "reasonable" forge-factor which not only accommodates the competing effects mentioned above, but also accounts to a good degree, for the approximation made on the indicator function during integration of Eq. 34 to get Eq. 35).

Performing the multiple sum of Eq. 23 on Eq. 35 for the case m = 1 is not very informative, as it involves mere adjustments of the partial normalization constants and $\bar{c}$. For m = 2, n = 1 (single component hard sphere gas), we have in Eq. 34 and 35 that $\alpha_1 = 1$ and $\alpha_2 = 0$, and we write b_1 as b_0. The m-tuplet distribution $f_2(\vec{r}_1, \vec{r}_2)$ is obtained as a sum similar to Eq. 25 as:

$$f_2(\vec{r}_1, \vec{r}_2) \approx \sum_{s_1=1}^{s_2-1} \sum_{s_2=2}^{M} (4\pi\rho\bar{c}(r_2))^{s_2-2} \prod_{j=1}^{2} \left(\frac{(r_j^3 - r_{(j-1)}^3)^{s_j - s_{j-1} - 1} \rho}{3^{s_j - s_{j-1} - 1}(s_j - s_{j-1} - 1)!} \right)$$
$$\times \exp\left(-\frac{4}{3}\pi r_2^3 \rho(1 + 4\eta_0)\right) F(\vec{r}_1, \vec{r}_2)$$

($\eta_0 = \frac{4}{3}\pi b_0^3 \rho$). Observe that the indicator function does not introduce complications in performing the sum as it simply factors out of the multiple sum. This is unlike the case when we had to integrate Eq. 34 by first temporarily setting the indicator function to unity, and reintroducing it after integration (for simplification). Following the scheme that led to Eq. 27, we obtain:

$$f_2(\vec{r}_1, \vec{r}_2) = \rho^2 e^{\frac{4}{3}\pi\rho(\bar{c}(r_2)-1)r_2^3} e^{-\frac{4}{3}\pi r_2^3 4\eta_0\rho} F(\vec{r}_1, \vec{r}_2) \tag{36}$$

For the OMPP process, one particle is at the origin (i.e., $\vec{r}_1 = \vec{r}_0$), and we can write:

$$f_2(\vec{r}_0, \vec{r}) = f_1(\vec{r}_0) f_1(\vec{r} \mid \vec{r}_0)$$

Hence we obtain the OMPP one-particle distribution function, which is also the *radial distribution function* as:

$$f_1\left(r \mid \vec{r}_0\right) \approx \begin{cases} 4\pi\rho r^2 e^{-\frac{4}{3}\pi\rho r^3(1+4\eta_0-\bar{c}(r))} & (r > 2b_0) \\ 0 & (r < 2b_0) \end{cases} \tag{37}$$

(integration over angular variables has been carried out). For large enough r (say $r \gtrsim d$), the radial distribution function is known to behave similar to that of the poisson fluid, and thus we need to have $\bar{c}(r) \approx 1 + 4\eta_0$ for $r \gtrsim d$. For r < d, we may assume $\bar{c}(r)$ is roughly constant for simplicity. Using the normalization of the radial distribution function which yields N-1 (see Sect. 2), and using some typical value for d (see literature simulation results) [13], we may then be able to determine reasonable values for $\bar{c}$.

For m = 2, n = 2, we earlier indicated that four kinds of m-tuplet distribution functions are involved. Following the same type of developments as above, we easily arrive at the radial distribution functions for these cases as:

$$f_1\left(r_j \mid \vec{r}_{i0}\right) \approx \begin{cases} 4\pi\rho_j r_j^2 \exp(-\frac{4}{3}\pi\rho r_j^3(1 - \bar{c}_{ij}(r_j) \\ \quad +4\eta_1\alpha_1 + 4\eta_2\alpha_2 + \eta_{12}\alpha_1\alpha_2)) & (r_j > b_i + b_j) \\ 0 & (r_j < b_i + b_j) \end{cases} \tag{38}$$

(i, j = 1, 2). For large enough r_j, it is again true that we must have $\bar{c}_{ij}(r_j) \rightarrow 1 + 4\eta_1\alpha_1 + 4\eta_2\alpha_2 + \eta_{12}\alpha_1\alpha_2$.

In determining the singlet cluster probabilities, the hard cores of particles of radius b_0 (single component case) or b_1, b_2 (binary multi-component case), should not be confused with the spheres of radii a_1, a_2 associated with particles for clustering. The former spheres determine the configuration of particles, which are frozen into place, before clustering may be considered. Mutually exclusive events can also be used as before to formulate the singlet cluster probabilities. The PNNPDF's to be used are those for the OMPP process, and hence Eq. 34 may be used with the indicator function rewritten as $F(\vec{r}_{0i}, \vec{r}_{k_1 1}, \ldots \vec{r}_{k_m m})$. The PNNPDF's are said to be valid for small m values at low densities. The integrals of Eq. 31 however become extremely difficult to perform. In the single component case, where the hard core of particles has radius b_0, and only one type of overlapping spheres (radius a_0) for clustering are considered, the NNPDF to be used is:

$$g(r) = \begin{cases} hr^2 e^{-\frac{4}{3}\pi\rho r^3(1+4\eta_0)} & (r > 2b_0) \\ 0 & (r < 2b_0) \end{cases} \tag{39}$$

where the normalization constant $h = 4\pi\rho\,(1 + 4\eta_0)\,e^{8\eta_0(1+4\eta_0)}$
Hence:

$$C(1, \rho) = \int\limits_{2a_0}^{\infty} g(r)dr = \begin{cases} \exp\left[-(1 + 4\eta_0)(8\hat{\eta}_0 - 8\eta_0)\right] & (a_0 > b_0) \\ 1 & (a_0 < b_0) \end{cases}$$

where $\frac{4}{3}\pi\rho a_0^3 = \hat{\eta}_0$.

5 Conclusions

In recent times, much of science, engineering, and technology has gone "nano." It becomes ever more important therefore, to develop accurate or exact theoretical (and experimental) means of structurally characterizing a variety of systems at the sub-atomic, atomic, and molecular scales. The present paper addresses the popular method employing m-body distribution functions as well as the emerging technique which employs PNNPDF's, to describe microstructure of arbitrary equilibrium material systems. The importance of structure (at various scales) in the study of materials, and in other disciplines, cannot be over emphasized, and has been extensively discussed in Ref. [10]. More importantly, it has been demonstrated that adequate knowledge of the configuration of constituent particles in terms of PNNPDF's (in arbitrary equilibrium systems), implies adequate knowledge of the free energy of the system involved, and vice-versa [6].

Developments have shown that PNNPDF's are relatively much easier to derive with high accuracy than their counterpart m-body distribution functions. Even while m-body distribution functions may be related in principle to the more accurately derivable PNNPDF's, the computational difficulty in manipulating the functional relationship is still so formidable that m-body distributions seem practically impossible to determine accurately beyond the cases of "very small" m values. m-body distribution functions do not normalize (to unity) as other regular probability distributions. This was determined to be related to the fact that events represented by different particle configurations are not mutually exclusive in the formulation of m-body distribution functions. In the non interacting multi-component problem, several new and exact results were obtained including the verification of the relationship between m-body distribution functions and PNNPDF's. Multiple sums and multiple integrals encountered proved to be very formidable to perform even for this relatively simple case. For the system of binary mixture of hard spheres, the parameter ε is known accurately only to second order in the density ρ. Hence PNNPDF's and other results provided in this case are valid only in the low density regime. Results applicable at higher densities are possible if ε is known more accurately. The complementary work of Edgal and Huber [6b] outlines the numerical scheme by which ε may be determined more accurately for interacting systems.

Formulating accurate general expressions for PNNPDF's with quite general many-body inter-particle interaction potential has been possible because "shape" and "surface" effects which ordinarily would not have been easily accounted for are known

to be weak and hence ignorable. A particularly notable revelation in this respect was that surface and shape effects presented themselves in "reduced" forms, hence allowing the general expression for the PNNPDF $g_{1,...,m}(\vec{r}_{k_1 1}, \ldots, \vec{r}_{k_m m})$ to be easily obtained accurately for m not too large (say a few tens or so). Surface and shape effects (like computational errors) actually manifested themselves as small additive terms in the expression for the quantity $\hat{E}$ (cf Sect. 3). The smallness being judged in relative terms as compared to the magnitude of variation (over small distances in nearest neighbor configuration space) of $\hat{E}$ which is relatively quite rapid along at least one of the coordinate directions (the radial $r_{k_m m}$ direction). This then allowed the conclusion to be made that the graphs of the quantities $\hat{E}$ and E (see Sect. 3) are "close" in outline features; leading to the grander conclusion that the graph of the PNNPDF which included surface and shape effects must be quite similar in outline features to that which ignores surface and shape effects (provided details within small local neighborhoods are ignored). The added complexity of having to determine several partial normalization constants in the multicomponent problem (as opposed to the case of just one normalization constant for single component systems) was said to be circumvented by considering PNNPDF's involving "fairly" large m values. The resulting sharply peaked distribution in phase space then allows us employ only one system (where $m_1 \sim \alpha_1 m; \ldots, m_n \sim \alpha_n m$) that may be allowed to make a random walk through phase space (by standard numerical simulation schemes) effectively leading to consideration of only a relatively "small" number of partial normalization constants.

A limited amount of statistical geometry of random media has also been addressed, firstly because nearest neighbor coordinate variables (whose distributions are extensively studied) are considered fundamental in constructing more complex geometrical structures, and secondly because singlet clustering has also been treated. Finally, it is expected that employment of NNPDF's and PNNPDF's should soon rapidly become commonplace as their use is not only presently an emergent phenomenon, but also, results of this paper show that they may be used for accurate analytical investigation of equilibrium systems. Also, relationship (as provided in this paper) between PNNPDF's and partial m-body distribution functions, which are currently well known, should lead to ready familiarization with PNNPDF's.

References

1. (a) C. Kittel, *Quantum Theory of Solids*, 2nd edn. (John Wiley, New York, 1987); (b) M.P. Marder, *Condensed Matter Physics* (John Wiley and Sons, Inc., New York, 2000)
2. (a) F. Iker, N. Andre, T. Pardoen, J.P. Raskin, Electrochem. Solid State Lett. **8**(10), H87–H89 (2005); (b) V. Agache, E. Quevy, D. Collard, L. Buchaillot, Appl. Phys. Lett. **79**(23), 3869–3871 (2001)
3. P.A. Greaney, D.C. Chrzan, B.M. Clemens, W.D. Nix, Appl. Phys. Lett. **83**(7), 1364–1366 (2003)
4. (a) J.B. Goodenough, J. Phys.: Condens. Matter **15**, R257–R326 (2003); (b) X. Sike, Z. Zhibo, M. Ruzhang, Supercond. Sci. Technol. **2**, 122–124 (1989)
5. T.M. Truskett, S. Torquato, S. Sastry, P.G. Debenedetti, F.H. Stillinger, Phys. Rev. E. **58**(3), 3083–3088 (1998)

6. (a) U.F. Edgal, J. Chem. Phys. **94**(12), 8179–8190 (1991); (b) U.F. Edgal, D.L. Huber, J. Phys. Chem. B. **108**, 13777–13788 (2004); (c) U.F. Edgal, D.L. Huber, Physica A, **362**, 295–304 (2006)
7. (a) M.G. Kendall, P.A.P. Moran, *Geometrical Probability* (Charles Griffin and Company Ltd., London, 1963); (b) S.A. Roach, *The Theory of Random Clumping* (Methuen, London, 1968)
8. (a) H. Reiss, J. Chem. Phys. **25**, 400, (1956); *ibid,* **25**, 408 (1956); (b) H. Reiss, C.S. Fuller, F.J. Morin, Bell Syst. Tech. J. **35**, 535 (1956); (c) J.D. Wiley, J. Phys. Chem. Solids **32**, 2053 (1971)
9. (a) S. Torquato, B. Lu, J. Rubinstein, Phys. Rev. A. **41**(4), 2059–2075 (1990); (b) H. Reiss, R.V. Casberg, J. Chem. Phys. **61**, 1107 (1974); (c) S. Torquato, *Random Heterogeneous Materials: Microstructure and Macroscopic Properties* (Springer Verlag, New York, 2002)
10. (a) U.F. Edgal, J. Chem. Phys. **94**(12), 8191–8202 (1991); (b) T.M. Truskett, S. Torquato, P.G. Debenedetti, Phys. Rev. E **58**(6), 7369–7380 (1998)
11. (a) R.K. Bowles, D.S. Corti, Mol. Phys. **98**(7), 479–438 (2000); (b) H. Reiss, J. Phys. Chem. **96**, 4736–4747 (1992); (c) S. Torquato, G. Stell, J. Chem. Phys. **82**, 980 (1985); (d) S. Sastry, D.S. Corti, P.G. Debenedetti, F.H. Stillinger, Phys. Rev. E **56**(5), 5524–5532 (1997); (e) S. Sastry, P.G. Debenedetti, F.H. Stillinger, Phys. Rev. E **56**(5), 5533–5543 (1997)
12. (a) J. Quintanilla, S. Torquato, Phys. Rev. E **54**(5), 5331–5339 (1996); (b) U.F. Edgal, J.D. Wiley, Phys. Rev. B. **27**(8), 4997–5006 (1983)
13. (a) J.E. Meyer, M.G. Meyer, *Statistical Mechanics* (John Wiley, 1977); (b) S.K. Srivastava, N.H. March, *Condensed Matter: Disordered Solids* (World Scientific, Singapore, 1995); (c) T.L. Hill, *Statistical Mechanics: Principles and Selected Applications* (McGraw Hill, New York, 1956) (sections 29 and 40); (d) I.Z. Fisher, *Statistical Theory of Liquids* (The University of Chicago Press, Chicago, 1964); (e) J.S. Rowlinson, F. L. Swinton, *Liquids and Liquid Mixtures* (Butterworth Scientific, London, 1982)