The entropy multiparticle-correlation expansion for a mixture of spherical and elongated particles

Santi Prestipino[1] and Paolo V Giaquinta

Università degli Studi di Messina and Istituto Nazionale per la Fisica della Materia (INFM), Dipartimento di Fisica, Contrada Papardo, 98166 Messina, Italy

E-mail: Santi.Prestipino@unime.it and Paolo.Giaquinta@unime.it

Abstract. We derive the multiparticle-correlation expansion of the excess entropy of classical particles in the canonical ensemble using a new approach that elucidates the rationale behind each term in the expansion. This formula provides the theoretical framework for an entropy-based ordering criterion that has already been tested for a variety of model fluids and thermodynamic phenomena. In view of future investigations of the phase diagram of colloidal mixtures, we detail in this paper the case of a two-component system of spherical and rod-like particles and discuss the symmetries underlying both the self and distinct pair-distribution functions under various geometrical constraints.

---

[1] Author to whom any correspondence should be addressed.

JSTAT (2004) P09008

Contents

1. Introduction

One-phase criteria have often been introduced to estimate the location of the phase boundaries of liquid and solid phases [1]. These empirical rules are usually rather specific. Nevertheless, they can be quite useful when it is not easy or straightforward to evaluate the free energies of the competing phases. Among such rules, one formerly proposed by Giaquinta and co-workers [2] has proved to be a rather general and flexible tool that can be successfully applied to a variety of models under different structural and thermodynamic conditions. This rule actually qualifies as an ordering criterion that can be readily implemented on the basis of the properties of the more disordered phase. The theoretical framework is provided by the multiparticle-correlation expansion (MPCE) of the excess entropy of a classical fluid that was originally derived by H S Green in the canonical ensemble [3], and was later extended by Nettleton and M S Green to an open system [4]. According to this formula, the entropy can be written as an infinite series whose nth term is the contribution associated with density correlations involving n-particle multiplets. In short, the criterion states that the overall contribution to the entropy of a fluid arising from the spatial correlations involving *more than two* particles—a quantity called 'residual multi-particle entropy' (RMPE)—changes sign concurrently with the local emergence of any new kind of structural organization in the system. The zero-RMPE criterion has been tested against such diverse thermodynamic phenomena as freezing [2, 5, 6], fluid–fluid phase separation in hard-sphere mixtures [7], mesophase formation in model liquid crystals [8, 9], the Kosterlitz and Thouless metal–insulator transition in a two-dimensional Coulomb lattice gas [10], and, more recently, the density-maximum anomaly in liquid water [11].

    In a previous paper, we gave an entirely new proof of the entropy MPCE in the canonical ensemble by exploiting a simple combinatorial identity [12]. The proof applies to both continuous and lattice systems. In this paper we reconsider this derivation from a different perspective which discloses the significance of each term in the expansion. Moreover, we outline an iterative method for building up the expansion term by term. This method is then extended to systems composed of two different species of particles.

Indeed, the present physical motivation for developing this formalism, besides the self-standing interest in the formal derivation of an entropy MPCE for binary mixtures, is that of demonstrating the effectiveness and reliability of the zero-RMPE criterion also in the case of a colloidal mixture of spherical and rod-like particles. In particular, we plan to check whether the miscible low-density phase of such a model undergoes a kind of structural instability—of the type indicated by the vanishing of the RMPE—towards a more ordered phase, be it *lamellar* (for low concentrations of the spheres) or *immiscible* (the bulk-phase-separated system), for values of the total packing fraction close to those independently ascertained by experiment and numerical simulation [13, 14]. Previous studies have already shown that the RMPE of parallel spherocylinders vanishes at about the same density where the smectic order sets in [8]. We are confident that the effects on the phase behaviour that are associated with the addition of a small quantity of spheres will be correctly accounted for by the RMPE.

While deferring to a forthcoming paper the discussion of the numerical simulation study of a mixture of hard spheres and spherocylinders, we anticipate here the analysis of the symmetries owned by the three pair distribution functions that enter the calculation of the two-body entropy.

2. One-component systems

We start by discussing the entropy MPCE of one-component systems in the canonical ensemble. This choice is by no means restrictive since, as Baranyai and Evans first pointed out, it is always possible to take advantage of the canonical sum rules of the correlation functions and then rearrange the entropy MPCE in an ensemble-invariant form [15].

Let $\mathbf{R}^N = \{\mathbf{R}_1, \ldots, \mathbf{R}_N\}$ and $\mathbf{P}^N = \{\mathbf{P}_1, \ldots, \mathbf{P}_N\}$ be the set of all particle coordinates and momenta, respectively. The canonical partition function of the system can always be split into an ideal and an excess part, $Z_N = Z_N^{\mathrm{id}} Z_N^{\mathrm{ex}}$, where

$$Z_N^{\mathrm{id}} = \frac{1}{N!} \left(\frac{V}{\Lambda^3}\right)^N \qquad \text{and} \qquad Z_N^{\mathrm{ex}} = \frac{1}{V^N} \int \mathrm{d}^3 R_1 \cdots \mathrm{d}^3 R_N \, \mathrm{e}^{-\beta V_N(\mathbf{R}^N)}. \tag{2.1}$$

In equation (2.1), $\beta = 1/(k_{\mathrm{B}}T)$, $\Lambda = h/\sqrt{2\pi m k_{\mathrm{B}}T}$, V is the volume, and $V_N(\mathbf{R}^N)$ is an arbitrary potential-energy function (in the most general case, V_N is a sum of n-body terms with $n = 1, 2, \ldots, N$). The excess entropy $S_N^{\mathrm{ex}} \equiv S_N - S_N^{\mathrm{id}}$ reads

$$\frac{S_N^{\mathrm{ex}}}{k_{\mathrm{B}}} = - \int \frac{\mathrm{d}^3 R_1 \cdots \mathrm{d}^3 R_N}{V^N} \frac{\mathrm{e}^{-\beta V_N(\mathbf{R}^N)}}{Z_N^{\mathrm{ex}}} \ln \frac{\mathrm{e}^{-\beta V_N(\mathbf{R}^N)}}{Z_N^{\mathrm{ex}}}. \tag{2.2}$$

Upon defining a set of N normalized distribution functions (DFs) as

$$P^{(N)}(\mathbf{R}^N) = \frac{\mathrm{e}^{-\beta V_N(\mathbf{R}^N)}}{Z_N^{\mathrm{ex}}};$$

$$P^{(n)}(\mathbf{R}^n) = \int \frac{\mathrm{d}^3 R_{n+1} \cdots \mathrm{d}^3 R_N}{V^{N-n}} \frac{\mathrm{e}^{-\beta V_N(\mathbf{R}^N)}}{Z_N^{\mathrm{ex}}} \qquad (n = 1, \ldots, N-1), \tag{2.3}$$

with the properties

$$\int \frac{\mathrm{d}^3 R_1 \cdots \mathrm{d}^3 R_n}{V^n} P^{(n)}(\mathbf{R}^n) = 1 \qquad \text{and} \qquad \int \frac{\mathrm{d}^3 R_{n+1}}{V} P^{(n+1)}(\mathbf{R}^{n+1}) = P^{(n)}(\mathbf{R}^n), \tag{2.4}$$

the ordinary n-body DFs can be expressed as

$$
\begin{aligned}
f^{(n)}(\mathbf{r}^n) &\equiv \left\langle \sideset{}{'}\sum_{i_1\ldots i_n} \delta^3(\mathbf{R}_{i_1} - \mathbf{r}_1) \cdots \delta^3(\mathbf{R}_{i_n} - \mathbf{r}_n) \right\rangle \\
&= \frac{N!}{(N-n)!} \left\langle \delta^3(\mathbf{R}_1 - \mathbf{r}_1) \cdots \delta^3(\mathbf{R}_n - \mathbf{r}_n) \right\rangle \\
&= \frac{N!}{(N-n)!} \frac{P^{(n)}(\mathbf{r}^n)}{V^n},
\end{aligned} \tag{2.5}
$$

where the sum is carried out over all n-tuples of distinct particle labels. The reduced n-body DFs (for $n = 2, \ldots, N$) read

$$
\begin{aligned}
g^{(n)}(\mathbf{r}^n) &\equiv \frac{f^{(n)}(\mathbf{r}^n)}{f^{(1)}(\mathbf{r}_1) \cdots f^{(1)}(\mathbf{r}_n)} \\
&= \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{n-1}{N}\right) \frac{P^{(n)}(\mathbf{r}^n)}{P^{(1)}(\mathbf{r}_1) \cdots P^{(1)}(\mathbf{r}_n)} \\
&\equiv \prod_{a=1}^{n} \left(1 - \frac{a-1}{N}\right) \tilde{P}^{(n)}(\mathbf{r}^n),
\end{aligned} \tag{2.6}
$$

and verify the property

$$
\int \frac{\mathrm{d}^3 R_{n+1}}{V} P^{(1)}(\mathbf{R}_{n+1}) g^{(n+1)}(\mathbf{R}^{n+1}) = \left(1 - \frac{n}{N}\right) g^{(n)}(\mathbf{R}^n), \tag{2.7}
$$

which also holds for $n = 1$ if $g^{(1)} \equiv 1$. We note that $P^{(1)} = 1$ and $f^{(1)} = N/V$ if no one-body term is present in V_N, i.e., if no external potential acts on the particles (for the ideal gas, $P^{(n)} = 1$ for all n). From now on, we shall adopt the shorthand notations $P_{12\ldots n} = P^{(n)}(\mathbf{R}^n)$ and $\tilde{P}_{12\ldots n} = \tilde{P}^{(n)}(\mathbf{R}^n)$. Moreover, any integral of the form $V^{-n} \int \mathrm{d}^3 R_1 \cdots \mathrm{d}^3 R_n (\cdots)$ is hereafter denoted as $\int(\cdots)$.

Now, we show how to build up the MPCE of the entropy term by term. The strategy is to consider a progressively increasing number of particles in the system. For $N = 1$, the (adimensional) excess entropy is just $S_1^{\mathrm{ex}}/k_{\mathrm{B}} = -\int P_1 \ln P_1$, which leads to a first-order approximation to the excess entropy of a N-particle system, $S_N^{\mathrm{ex}}/k_{\mathrm{B}} \approx S_N^{(1)}/k_{\mathrm{B}} \equiv -N \int P_1 \ln P_1$ (namely, each particle contributes to the entropy independently of the other particles).

Next, we move to a system of two particles only, and write its excess entropy as $S_2^{(1)}$ plus a remainder $k_{\mathrm{B}} R_2$ that is equal to

$$
R_2 \equiv \frac{S_2^{\mathrm{ex}} - S_2^{(1)}}{k_{\mathrm{B}}} = -\int P_{12} \ln P_{12} + 2 \int P_1 \ln P_1 = -\int P_{12} \ln \tilde{P}_{12}. \tag{2.8}
$$

Equation (2.8) leads to a second-order approximation for S_N^{ex} under the hypothesis that each distinct pair of particles contributes an equal two-body residual term to the entropy:

$$
\frac{S_N^{(2)}}{k_{\mathrm{B}}} = -N \int P_1 \ln P_1 - \binom{N}{2} \int P_{12} \ln \tilde{P}_{12}. \tag{2.9}
$$

Note that the approximation (2.9) is exact for $N = 2$, i.e., $S_2^{(2)} = S_2^{\mathrm{ex}}$.

When there are three particles in the system, the excess entropy is the sum of $S_3^{(2)}$ and a remainder $k_B R_3$:

$$R_3 \equiv \frac{S_3^{\mathrm{ex}} - S_3^{(2)}}{k_B} = -\int P_{123} \ln \tilde{P}_{123} + \binom{3}{2} \int P_{12} \ln \tilde{P}_{12}, \tag{2.10}$$

which suggests a third-order approximation for S_N^{ex} in the form

$$\frac{S_N^{(3)}}{k_B} = -N \int P_1 \ln P_1 - \binom{N}{2} \int P_{12} \ln \tilde{P}_{12} - \binom{N}{3} \left[\int P_{123} \ln \tilde{P}_{123} - \binom{3}{2} \int P_{12} \ln \tilde{P}_{12} \right]. \tag{2.11}$$

Again, note that $S_3^{(3)} = S_3^{\mathrm{ex}}$. Equation (2.11) reproduces the first three terms in the rhs of equation (5.9) of [12], and one may suspect that the remaining terms will be similarly discovered by arguing for $N = 4, 5, \ldots$ as we did for $N = 1, 2, 3$. The related proof can be obtained by induction over N. In fact, we know our target [12]:

$$\frac{S_N^{\mathrm{ex}}}{k_B} = -N \int P_1 \ln P_1 - \int P_{12\ldots N} \ln \tilde{P}_{12\ldots N}$$

$$= -N \int P_1 \ln P_1 - \sum_{n=2}^{N} \binom{N}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} \int P_{1\ldots a} \ln \tilde{P}_{1\ldots a}. \tag{2.12}$$

After taking $I_a \equiv \int P_{1\ldots a} \ln \tilde{P}_{1\ldots a}$ and $J_a \equiv \int P_{1\ldots a} \ln P_{1\ldots a}$ (for $a = 1, \ldots, N$), we first show that equation (2.12) is correct or, equivalently, that

$$I_N = \sum_{n=2}^{N} \binom{N}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} I_a. \tag{2.13}$$

To this aim, it is enough to observe that the coefficient of I_a in the sum (2.13) is (cf equation (5.10) of [12])

$$\sum_{n=a}^{N} (-1)^{n-a} \binom{N}{n}\binom{n}{a} = \begin{cases} 0, & \text{for } 0 \le a < N \\ 1, & \text{for } a = N. \end{cases} \tag{2.14}$$

Equation (2.12) can be built up term by term using the same procedure as sketched above for $N = 1, 2, 3$. Upon defining for a given M (with $2 \le M \le N - 1$)

$$\frac{S_N^{(M)}}{k_B} = -N J_1 - \sum_{n=2}^{M} \binom{N}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} I_a, \tag{2.15}$$

which is tantamount to truncating the sum (2.12) over n after the Mth term, we set $S_{M+1}^{\mathrm{ex}} \equiv S_{M+1}^{(M)} + k_B R_{M+1}$, where

$$R_{M+1} = -\underbrace{(J_{M+1} - (M+1)J_1)}_{I_{M+1}} + \sum_{n=2}^{M} \binom{M+1}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} I_a$$

$$= -\sum_{n=2}^{M+1} \binom{M+1}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} I_a + \sum_{n=2}^{M} \binom{M+1}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} I_a$$

$$= -\sum_{a=2}^{M+1} (-1)^{M+1-a} \binom{M+1}{a} I_a. \tag{2.16}$$

This result allows one to define a higher-order approximation to the entropy as

$$
\begin{aligned}
\frac{S_N^{(M+1)}}{k_{\mathrm{B}}} &\equiv \frac{S_N^{(M)}}{k_{\mathrm{B}}} + \binom{N}{M+1} R_{M+1} \\
&= -NJ_1 - \sum_{n=2}^{M} \binom{N}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} I_a - \binom{N}{M+1} \\
&\quad \times \sum_{a=2}^{M+1} (-1)^{M+1-a} \binom{M+1}{a} I_a \\
&= -NJ_1 - \sum_{n=2}^{M+1} \binom{N}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} I_a,
\end{aligned}
\tag{2.17}
$$

which is exactly the same as (2.15), but for the quantity $M + 1$ which replaces M.

The MPCE of the entropy remains formally the same if the particles possess further degrees of freedom besides those pertaining to the centre of mass. For instance, in the case of liquid-crystal molecules (i.e., elongated particles with cylindrical symmetry), there are two more degrees of freedom for each particle since two angular coordinates are needed to specify the orientation of the molecule in three-dimensional space. If the body z axis is taken to coincide with the molecular axis, we can represent the direction of a molecule by resorting to the Euler angles θ and ϕ (see figure 1). The third angle, ψ, describes a rotation around the molecular axis and, as such, it is not relevant for the configuration of the molecule. Let $\xi = (\mathbf{R}, \theta, \phi)$ be the five-dimensional vector of coordinates of an individual molecule. The interaction potential is then a function of ξ^N.

One of the simplest reference models for a liquid crystal is a system of hard spherocylinders or, equivalently, rods that cannot approach each other beyond a given distance σ (the spherocylinder diameter). Such particles show up–down symmetry. In this specific case, the potential $V_N(\xi^N)$ will also reflect this symmetry, in that it must be invariant upon interchanging (θ, ϕ) with $(\pi - \theta, \pi + \phi)$.

The rotational kinetic energy K_{rot} of a massive rod, written in canonical coordinates, reads

$$
K_{\mathrm{rot}} = \frac{P_\theta^2}{2I} + \frac{P_\phi^2}{2I \sin^2 \theta},
\tag{2.18}
$$

where I is the moment of inertia relative to any axis perpendicular to the rod and passing through its centre. It then easily follows that the ideal and excess partition functions are

$$
Z_N^{\mathrm{id}} = \frac{1}{N!} \left(\frac{4\pi V}{\lambda^2 \Lambda^3} \right)^N \quad \text{and} \quad Z_N^{\mathrm{ex}} = \frac{1}{(4\pi V)^N} \int \mathrm{d}^5 \xi_1 \cdots \mathrm{d}^5 \xi_N \mathrm{e}^{-\beta V_N(\xi^N)},
\tag{2.19}
$$

where $\lambda = h/\sqrt{2\pi I k_{\mathrm{B}} T}$ and $\mathrm{d}^5 \xi = \sin\theta \, \mathrm{d}^3 R \, \mathrm{d}\theta \, \mathrm{d}\phi$. The factor $\sin\theta$ in the volume element originates from the Gaussian integral over P_ϕ. Because of this factor, the delta function of argument ξ should be interpreted as follows:

$$
\delta^5(\xi - \xi_1) \equiv \delta^3(\mathbf{R} - \mathbf{R}_1) \frac{\delta(\theta - \theta_1)}{\sin\theta} \delta(\phi - \phi_1).
\tag{2.20}
$$

With this proviso, the formal definition of the DFs, given by equation (2.5), as well as the overall appearance of the entropy formula (see equation (2.12)), remains unchanged
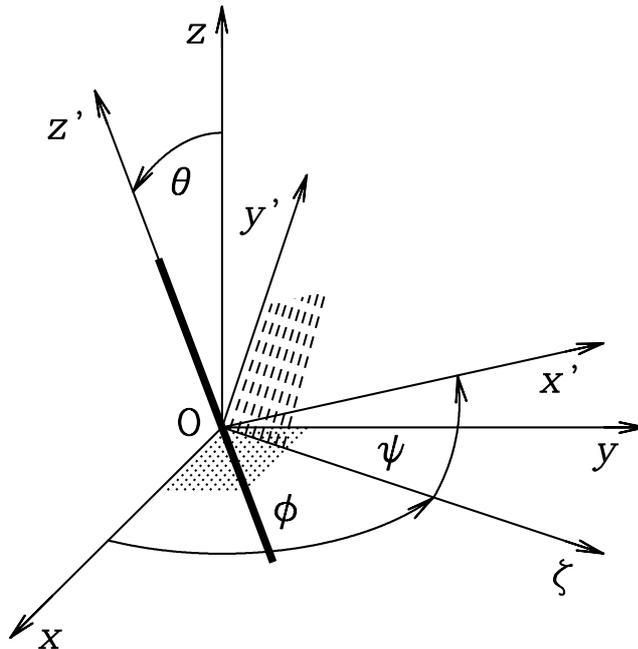
Figure 1. The Euler angles $\theta, \phi,$ and ψ. The axes x, y, and z form the laboratory reference frame, whereas x', y', and z' are parallel to the body set of axes. ζ, called the nodal line, is the straight line perpendicular to z and z'. It can also be viewed as the intersection between the xy and $x'y'$ planes. The z' axis is chosen so as to coincide with the symmetry axis of the molecule (represented in the picture as a rod). The Euler angles are in the ranges $0 \leq \theta < \pi$, $0 \leq \phi < 2\pi$, and $0 \leq \psi < 2\pi$.

provided that we now interpret $\int(\cdots)$ as $(4\pi V)^{-n} \int \mathrm{d}^5\xi_1 \cdots \mathrm{d}^5\xi_n (\cdots)$ (the value of n is always implicit in the form of the integrand).

Let us now consider the properties of the reduced pair distribution function (PDF) in relation to the symmetries of the system. Assume that no external field is present and that the molecular interaction includes, besides the hard-core repulsion, at most a pair term whose strength depends on the distance between the centres of mass. In such a case, the PDF $g^{(2)}(\xi_1, \xi_2)$ will only depend on the relative position of the two molecules. In a reference frame Σ_1 where molecule 1 is placed at the origin and oriented along the z axis, the position of molecule 2 is thoroughly described by the three (spherical) coordinates of its centre of mass ($r_{12}, \vartheta_{12},$ and φ_{12}) plus two more angles, θ_{12} and ϕ_{12}, specifying the direction of its axis. However, the above description is redundant since the orientation of, say, the x axis of Σ_1 is still arbitrary and we can always arrange things in such a way that $\varphi_{12} = 0$. Hence, $g^{(2)}$ will depend on four variables only, namely one distance (r_{12}) and three angles ($\vartheta_{12}, \theta_{12},$ and ϕ_{12}).

An even simpler situation is that of an artificially constrained nematic fluid, namely a system of elongated particles whose axes are kept parallel to each other while their centres of mass are free to move. In this case, the angles θ_{12} and ϕ_{12} are no longer necessary, with the result that the PDF depends just on r_{12} and ϑ_{12} (or, equivalently, on $\rho_{12} = r_{12}\sin\vartheta_{12}$ and $z_{12} = r_{12}\cos\vartheta_{12}$).

Another interesting case is that of an inhomogeneous system of rod-like molecules confined by an impenetrable wall. This model is useful for investigating the wetting properties of a nematic fluid and the onset and growth of a smectic layer at the wall. If the strength of the wall–particle attraction depends just on the distance z of the molecular centre of mass from the wall, the number density $f^{(1)}(\xi)$ turns out to be a function of $\chi = (z, \theta, \phi)$ (i.e., the axis direction is relevant even when the wall does not exert any attraction), while the PDF $g^{(2)}(\xi_1, \xi_2)$ depends on χ_1 and χ_2.

3. Two-component systems

In this section, we generalize the MPCE of the entropy to binary systems composed of two different kinds of particles. We shall proceed in two steps. We shall first use the iterative method outlined in the previous section as a guidance for conjecturing the complete formula from its first few terms. Then, we shall give a formal proof of this formula by induction over the total number of particles.

The canonical partition function of a two-component system with $N = N_1 + N_2$ classical point particles generally reads $Z_{N_1,N_2} = Z_{N_1}^{\mathrm{id},1} Z_{N_2}^{\mathrm{id},2} Z_{N_1,N_2}^{\mathrm{ex}}$, where the excess part has the form

$$Z_{N_1,N_2}^{\mathrm{ex}} = \frac{1}{V^N} \int \mathrm{d}^{3N_1} R \, \mathrm{d}^{3N_2} Q \, \mathrm{e}^{-\beta V_{N_1,N_2}(\mathbf{R}^{N_1}, \mathbf{Q}^{N_2})}. \tag{3.1}$$

In equation (3.1), the potential-energy function is arbitrary. The excess entropy is given by the integral

$$\frac{S_{N_1,N_2}^{\mathrm{ex}}}{k_{\mathrm{B}}} = -\frac{1}{V^N} \int \mathrm{d}^{3N_1} R \, \mathrm{d}^{3N_2} Q \, \frac{\mathrm{e}^{-\beta V_{N_1,N_2}(\mathbf{R}^{N_1}, \mathbf{Q}^{N_2})}}{Z_{N_1,N_2}^{\mathrm{ex}}} \ln \frac{\mathrm{e}^{-\beta V_{N_1,N_2}(\mathbf{R}^{N_1}, \mathbf{Q}^{N_2})}}{Z_{N_1,N_2}^{\mathrm{ex}}}. \tag{3.2}$$

Upon defining

$$P^{(N_1,N_2)}(\mathbf{R}^{N_1}, \mathbf{Q}^{N_2}) = \frac{\mathrm{e}^{-\beta V_{N_1,N_2}(\mathbf{R}^{N_1}, \mathbf{Q}^{N_2})}}{Z_{N_1,N_2}^{\mathrm{ex}}};$$

$$P^{(n_1,n_2)}(\mathbf{R}^{n_1}, \mathbf{Q}^{n_2}) = \int \frac{\mathrm{d}^3 R_{n_1+1} \cdots \mathrm{d}^3 R_{N_1}}{V^{N_1-n_1}} \frac{\mathrm{d}^3 Q_{n_2+1} \cdots \mathrm{d}^3 Q_{N_2}}{V^{N_2-n_2}} \frac{\mathrm{e}^{-\beta V_{N_1,N_2}(\mathbf{R}^{N_1}, \mathbf{Q}^{N_2})}}{Z_{N_1,N_2}^{\mathrm{ex}}}$$

$$(\text{with } n_1 \leq N_1, n_2 \leq N_2, n_1 + n_2 = 1, \ldots, N-1), \tag{3.3}$$

we have the following properties:

$$\int \frac{\mathrm{d}^3 R_1 \cdots \mathrm{d}^3 R_{n_1}}{V^{n_1}} \frac{\mathrm{d}^3 Q_1 \cdots \mathrm{d}^3 Q_{n_2}}{V^{n_2}} P^{(n_1,n_2)}(\mathbf{R}^{n_1}, \mathbf{Q}^{n_2}) = 1;$$

$$\int \frac{\mathrm{d}^3 R_{n_1+1}}{V} P^{(n_1+1,n_2)}(\mathbf{R}^{n_1+1}, \mathbf{Q}^{n_2}) = P^{(n_1,n_2)}(\mathbf{R}^{n_1}, \mathbf{Q}^{n_2}); \tag{3.4}$$

$$\int \frac{\mathrm{d}^3 Q_{n_2+1}}{V} P^{(n_1,n_2+1)}(\mathbf{R}^{n_1}, \mathbf{Q}^{n_2+1}) = P^{(n_1,n_2)}(\mathbf{R}^{n_1}, \mathbf{Q}^{n_2}).$$

When external fields are absent, $P^{(1,0)} = 1$ and $P^{(0,1)} = 1$. For a binary ideal mixture (i.e., $V_{N_1,N_2} = 0$), the P functions are all equal to 1.

The DF of order (n_1, n_2) is defined as

$$f^{(n_1,n_2)}(\mathbf{r}^{n_1}, \mathbf{q}^{n_2}) = \frac{N_1!}{(N_1-n_1)!} \frac{N_2!}{(N_2-n_2)!} \frac{P^{(n_1,n_2)}(\mathbf{r}^{n_1}, \mathbf{q}^{n_2})}{V^{n_1+n_2}}. \tag{3.5}$$

While the definition of the *self* reduced DFs $g^{(n_1,0)}$ (for $n_1 \geq 1$) and $g^{(0,n_2)}$ (for $n_2 \geq 1$) is strictly analogous to that given for a one-component system, the *distinct* reduced DFs are defined, for $n_1, n_2 \geq 1$, as

$$g^{(n_1,n_2)}(\mathbf{r}^{n_1}, \mathbf{q}^{n_2}) = \prod_{a=1}^{n_1}\left(1 - \frac{a-1}{N_1}\right) \prod_{b=1}^{n_2}\left(1 - \frac{b-1}{N_2}\right) \tilde{P}^{(n_1,n_2)}(\mathbf{r}^{n_1}, \mathbf{q}^{n_2}), \tag{3.6}$$

where

$$\tilde{P}^{(n_1,n_2)}(\mathbf{r}^{n_1}, \mathbf{q}^{n_2}) \equiv \frac{P^{(n_1,n_2)}(\mathbf{r}^{n_1}, \mathbf{q}^{n_2})}{P^{(1,0)}(\mathbf{r}_1) \cdots P^{(1,0)}(\mathbf{r}_{n_1}) P^{(0,1)}(\mathbf{q}_1) \cdots P^{(0,1)}(\mathbf{q}_{n_2})}. \tag{3.7}$$

For $n_1 + n_2 \geq 1$, a property analogous to that expressed in equation (2.7) holds:

$$\int \frac{\mathrm{d}^3 r_{n_1+1}}{V} P^{(1,0)}(\mathbf{r}_{n_1+1}) g^{(n_1+1,n_2)}(\mathbf{r}^{n_1+1}, \mathbf{q}^{n_2}) = \left(1 - \frac{n_1}{N_1}\right) g^{(n_1,n_2)}(\mathbf{r}^{n_1}, \mathbf{q}^{n_2}), \tag{3.8}$$

plus a similar identity involving $g^{(n_1,n_2+1)}$.

From now on, a short notation is adopted where P_{ab} stands for $P^{(a,b)}(\mathbf{r}^a, \mathbf{q}^b)$ and any integral of the form $V^{-n_1-n_2}\int \mathrm{d}^{3n_1}r\, \mathrm{d}^{3n_2}q(\cdots)$ is simply denoted as $\int(\cdots)$. Moreover, for future convenience, we take $I_{ab} \equiv \int P_{ab} \ln \tilde{P}_{ab}$ and $J_{ab} \equiv \int P_{ab} \ln P_{ab}$ (with $a \leq N_1, b \leq N_2, a+b \geq 1$). For instance, the excess entropy (3.2) can be rewritten as

$$\begin{aligned} \frac{S_{N_1,N_2}^{\mathrm{ex}}}{k_{\mathrm{B}}} &= -\int P_{N_1 N_2} \ln P_{N_1 N_2} \\ &= -N_1 \int P_{10} \ln P_{10} - N_2 \int P_{01} \ln P_{01} - \int P_{N_1 N_2} \ln \tilde{P}_{N_1 N_2}, \end{aligned} \tag{3.9}$$

or $J_{N_1 N_2} = N_1 J_{10} + N_2 J_{01} + I_{N_1 N_2}$.

We now move on to determine an MPCE for the entropy of a mixture. To this aim, we consider increasing particle numbers in the system, starting from $N_1 + N_2 = 1$ (i.e., only one particle, either of type 1 or 2, is present). By reasoning in the usual way, we immediately obtain a first-order approximation to the excess entropy in the form

$$\frac{S_{N_1,N_2}^{(1)}}{k_{\mathrm{B}}} = -N_1 \int P_{10} \ln P_{10} - N_2 \int P_{01} \ln P_{01}. \tag{3.10}$$

Then, we analyse the three cases with $N_1 + N_2 = 2$. For $N_1 = 2$ and $N_2 = 0$ (or the other way round), things are the same as for a one-component system, i.e., $S_{2,0}^{\mathrm{ex}} = S_{2,0}^{(1)} + k_{\mathrm{B}} R_{2,0}$, with $R_{2,0} = -\int P_{20} \ln \tilde{P}_{20}$. Instead, for $N_1 = N_2 = 1$,

$$R_{1,1} \equiv \frac{S_{1,1}^{\mathrm{ex}} - S_{1,1}^{(1)}}{k_{\mathrm{B}}} = -\int P_{11} \ln P_{11} + \int P_{10} \ln P_{10} + \int P_{01} \ln P_{01} = -\int P_{11} \ln \tilde{P}_{11}, \tag{3.11}$$

thus leading to the following second-order approximation for S_{N_1,N_2}^{ex}:

$$\frac{S_{N_1,N_2}^{(2)}}{k_{\text{B}}} = -N_1 \int P_{10} \ln P_{10} - N_2 \int P_{01} \ln P_{01}$$
$$- \binom{N_1}{2} \int P_{20} \ln \tilde{P}_{20} - N_1 N_2 \int P_{11} \ln \tilde{P}_{11} - \binom{N_2}{2} \int P_{02} \ln \tilde{P}_{02}. \qquad (3.12)$$

For $N_1 + N_2 = 3$, we just reproduce below the expression of $S_{N_1,N_2}^{(3)}$, which results from carefully considering the implied four cases:

$$\frac{S_{N_1,N_2}^{(3)}}{k_{\text{B}}} = -N_1 \int P_{10} \ln P_{10} - N_2 \int P_{01} \ln P_{01} - \binom{N_1}{2} \int P_{20} \ln \tilde{P}_{20} - N_1 N_2 \int P_{11} \ln \tilde{P}_{11}$$
$$- \binom{N_2}{2} \int P_{02} \ln \tilde{P}_{02} - \binom{N_1}{3} \left[\int P_{30} \ln \tilde{P}_{30} - \binom{3}{2} \int P_{20} \ln \tilde{P}_{20} \right]$$
$$- \binom{N_1}{2} N_2 \left[\int P_{21} \ln \tilde{P}_{21} - \int P_{20} \ln \tilde{P}_{20} - 2 \int P_{11} \ln \tilde{P}_{11} \right]$$
$$- N_1 \binom{N_2}{2} \left[\int P_{12} \ln \tilde{P}_{12} - 2 \int P_{11} \ln \tilde{P}_{11} - \int P_{02} \ln \tilde{P}_{02} \right]$$
$$- \binom{N_2}{3} \left[\int P_{03} \ln \tilde{P}_{03} - \binom{3}{2} \int P_{02} \ln \tilde{P}_{02} \right]. \qquad (3.13)$$

On the basis of the above structure, we conjecture the following general formula:

$$\frac{S_{N_1,N_2}^{\text{ex}}}{k_{\text{B}}} = -N_1 \int P_{10} \ln P_{10} - N_2 \int P_{01} \ln P_{01} - \sum_{\substack{n_1+n_2 \geq 2 \\ n_1 \leq N_1, n_2 \leq N_2}} \binom{N_1}{n_1}\binom{N_2}{n_2}$$
$$\times \sum_{\substack{a+b \geq 2 \\ a \leq n_1, b \leq n_2}} (-1)^{n_1+n_2-a-b} \binom{n_1}{a}\binom{n_2}{b} \int P_{ab} \ln \tilde{P}_{ab}, \qquad (3.14)$$

which was preliminarily checked against the form of $S_{N_1,N_2}^{(4)}$ as independently got by our iterative method. Note that the double sum in equation (3.14) can also be arranged in such a way that all the terms with the same value of $n \equiv n_1 + n_2$ are gathered together:

$$I_{N_1 N_2} = \sum_{n=2}^{N_1+N_2} \sum_{n_1=\max\{n-N_2,0\}}^{\min\{n,N_1\}} \binom{N_1}{n_1}\binom{N_2}{n-n_1} \sum_{\substack{a+b \geq 2 \\ a \leq n_1, b \leq n-n_1}} (-1)^{n-a-b} \binom{n_1}{a}\binom{n-n_1}{b} I_{ab}.$$
$$(3.15)$$

First of all, we prove that equation (3.14) is an exact identity. Indeed, for given a and b (with $a \leq N_1$, $b \leq N_2$, $a+b \geq 2$), the coefficient of I_{ab} on the rhs of (3.14) is

$$- \sum_{\substack{n_1+n_2 \geq 2 \\ a \leq n_1 \leq N_1, b \leq n_2 \leq N_2}} \binom{N_1}{n_1}\binom{N_2}{n_2} \times (-1)^{n_1+n_2-a-b} \binom{n_1}{a}\binom{n_2}{b}$$
$$= - \sum_{n_1=a}^{N_1} (-1)^{n_1-a} \binom{N_1}{n_1}\binom{n_1}{a} \times \sum_{n_2=b}^{N_2} (-1)^{n_2-b} \binom{N_2}{n_2}\binom{n_2}{b}. \qquad (3.16)$$

On account of equation (2.14), the number (3.16) is always zero but for the case $(a,b) = (N_1, N_2)$, where it equals -1. Hence, equation (3.14) is correct.

We next prove, by induction, that equation (3.14) also follows from our iterative method. We define, for any M such that $2 \leq M \leq N - 1$,

$$\frac{S_{N_1,N_2}^{(M)}}{k_{\rm B}} = -N_1 J_{10} - N_2 J_{01} - \sum_{n=2}^{M} \sum_{n_1=\max\{n-N_2,0\}}^{\min\{n,N_1\}} \binom{N_1}{n_1}\binom{N_2}{n-n_1}$$

$$\times \sum_{\substack{a+b\geq 2 \\ a\leq n_1, b\leq n-n_1}} (-1)^{n-a-b}\binom{n_1}{a}\binom{n-n_1}{b} I_{ab}. \tag{3.17}$$

Moreover, for each pair (M_1, M_2) satisfying $M_1 + M_2 = M + 1$, let us put $S_{M_1,M_2}^{\rm ex} = S_{M_1,M_2}^{(M)} + k_{\rm B} R_{M_1,M_2}$, where

$$R_{M_1,M_2} = -(\underbrace{J_{M_1 M_2} - M_1 J_{10} - M_2 J_{01}}_{I_{M_1 M_2}}) + \sum_{n=2}^{M} \sum_{n_1=\max\{n-M_2,0\}}^{\min\{n,M_1\}} \binom{M_1}{n_1}\binom{M_2}{n-n_1}$$

$$\times \sum_{\substack{a+b\geq 2 \\ a\leq n_1, b\leq n-n_1}} (-1)^{n-a-b}\binom{n_1}{a}\binom{n-n_1}{b} I_{ab}$$

$$= -\sum_{n=2}^{M+1} \sum_{n_1=\max\{n-M_2,0\}}^{\min\{n,M_1\}} \binom{M_1}{n_1}\binom{M_2}{n-n_1} \sum_{\substack{a+b\geq 2 \\ a\leq n_1, b\leq n-n_1}} (-1)^{n-a-b}\binom{n_1}{a}\binom{n-n_1}{b} I_{ab}$$

$$+ \sum_{n=2}^{M} \sum_{n_1=\max\{n-M_2,0\}}^{\min\{n,M_1\}} \binom{M_1}{n_1}\binom{M_2}{n-n_1} \sum_{\substack{a+b\geq 2 \\ a\leq n_1, b\leq n-n_1}} (-1)^{n-a-b}\binom{n_1}{a}$$

$$\times \binom{n-n_1}{b} I_{ab} = - \sum_{n_1=M+1-M_2}^{M_1} \binom{M_1}{n_1}\binom{M_2}{M+1-n_1}$$

$$\times \sum_{\substack{a+b\geq 2 \\ a\leq n_1, b\leq M+1-n_1}} (-1)^{M+1-a-b}\binom{n_1}{a}\binom{M+1-n_1}{b} I_{ab}$$

$$= - \sum_{\substack{a+b\geq 2 \\ a\leq M_1, b\leq M_2}} (-1)^{M+1-a-b}\binom{M_1}{a}\binom{M_2}{b} I_{ab}. \tag{3.18}$$

In the third line of the above equation, the identity (3.15) was used for $I_{M_1 M_2}$. In the end, the same structure of $S_{N_1,N_2}^{(M)}$ emerges for $S_{N_1,N_2}^{(M+1)}$ (i.e., equation (3.17) where $M + 1$ replaces M), as long as we define, in close analogy with the one-component case,

$$\frac{S_{N_1,N_2}^{(M+1)}}{k_{\rm B}} = \frac{S_{N_1,N_2}^{(M)}}{k_{\rm B}} + \sum_{M_1=\max\{M+1-N_2,0\}}^{\min\{M+1,N_1\}} \binom{N_1}{M_1}\binom{N_2}{M+1-M_1} R_{M_1,M+1-M_1}. \tag{3.19}$$

This concludes our proof.

Equations (2.12) and (3.14) express the canonical MPCE of the excess entropy in compact form for pure and mixed systems, respectively. In the literature, the expressions quoted for the first few terms of this expansion are far more involved because they are usually written in terms of the *reduced* DFs. We have already shown in [12] that the familiar form of the entropy formula for one-component systems in the canonical ensemble emerges when the Ps are eliminated in favour of the gs through the reverse of equation (2.6). In doing so, a constant term can be extracted from the correlation integrals which, if absorbed into the ideal-gas part of the entropy, makes the latter equivalent to the entropy of the *infinite-sized* ideal-gas system (for the sake of clarity, this argument is reformulated in appendix A).

Before providing a similar demonstration for mixed systems, let us see what happens to $S_{N_1,N_2}^{(2)}$. Upon repeatedly using equations (3.4) and (3.8), we obtain (with $\rho_1 = N_1/V$ and $\rho_2 = N_2/V$)

$$
\begin{aligned}
\frac{S_{N_1,N_2}^{(2)}}{k_{\mathrm{B}}} =\ & -\rho_1 \int \mathrm{d}^3 r\, P_{10}(\mathbf{r}) \ln P_{10}(\mathbf{r}) - \rho_2 \int \mathrm{d}^3 r\, P_{01}(\mathbf{r}) \ln P_{01}(\mathbf{r}) - \binom{N_1}{2} \ln \frac{N_1}{N_1 - 1} \\
& - \tfrac{1}{2}\rho_1^2 \int \mathrm{d}^3 r\, \mathrm{d}^3 r'\, P_{10}(\mathbf{r}) P_{10}(\mathbf{r}') g_{20}(\mathbf{r},\mathbf{r}') \ln g_{20}(\mathbf{r},\mathbf{r}') \\
& - \rho_1 \rho_2 \int \mathrm{d}^3 r\, \mathrm{d}^3 r'\, P_{10}(\mathbf{r}) P_{01}(\mathbf{r}') g_{11}(\mathbf{r},\mathbf{r}') \ln g_{11}(\mathbf{r},\mathbf{r}') - \binom{N_2}{2} \ln \frac{N_2}{N_2 - 1} \\
& - \tfrac{1}{2}\rho_2^2 \int \mathrm{d}^3 r\, \mathrm{d}^3 r'\, P_{01}(\mathbf{r}) P_{01}(\mathbf{r}') g_{02}(\mathbf{r},\mathbf{r}') \ln g_{02}(\mathbf{r},\mathbf{r}'),
\end{aligned}
\tag{3.20}
$$

where all terms, including the two constants, are *extensive*, namely, each term scales in the thermodynamic limit linearly with either N or V.

However, in a closed system the asymptotic value of the two-body *self* reduced DFs differs at large distances from 1 for $\mathcal{O}(N^{-1})$ terms [16]. This makes the numerical evaluation of the integrals in (3.20) particularly harmful for a small system (i.e., sensitive to its boundary). To (partially) cure this problem, we can take advantage of the canonical sum rules for the DFs (see equation (3.8)), adding (and subtracting) to each integral precisely the extensive term that makes the integrand of order N^{-2} at infinity:

$$
\begin{aligned}
\frac{S_{N_1,N_2}^{(2)}}{k_{\mathrm{B}}} =\ & -\rho_1 \int \mathrm{d}^3 r\, P_{10}(\mathbf{r}) \ln P_{10}(\mathbf{r}) - \rho_2 \int \mathrm{d}^3 r\, P_{01}(\mathbf{r}) \ln P_{01}(\mathbf{r}) \\
& - \tfrac{1}{2}\rho_1^2 \int \mathrm{d}^3 r\, \mathrm{d}^3 r'\, P_{10}(\mathbf{r}) P_{10}(\mathbf{r}') [g_{20}(\mathbf{r},\mathbf{r}') \ln g_{20}(\mathbf{r},\mathbf{r}') - g_{20}(\mathbf{r},\mathbf{r}') + 1] \\
& - \rho_1 \rho_2 \int \mathrm{d}^3 r\, \mathrm{d}^3 r'\, P_{10}(\mathbf{r}) P_{01}(\mathbf{r}') [g_{11}(\mathbf{r},\mathbf{r}') \ln g_{11}(\mathbf{r},\mathbf{r}') - g_{11}(\mathbf{r},\mathbf{r}') + 1] \\
& - \tfrac{1}{2}\rho_2^2 \int \mathrm{d}^3 r\, \mathrm{d}^3 r'\, P_{01}(\mathbf{r}) P_{01}(\mathbf{r}') [g_{02}(\mathbf{r},\mathbf{r}') \ln g_{02}(\mathbf{r},\mathbf{r}') - g_{02}(\mathbf{r},\mathbf{r}') + 1] \\
& - N_1 \left(\frac{N_1 - 1}{2} \ln \frac{N_1}{N_1 - 1} - \frac{1}{2} \right) - N_2 \left(\frac{N_2 - 1}{2} \ln \frac{N_2}{N_2 - 1} - \frac{1}{2} \right).
\end{aligned}
\tag{3.21}
$$

As a result, (i) the contribution from the boundary now grows like $V \times V^{2/3} N^{-2} \propto N^{-1/3}$; (ii) the overall constant term outside of the integrals is of $\mathcal{O}(1)$; and (iii) the new integrals now conform to those in the grand-canonical-ensemble expansion.

For general (N_1, N_2), the constant terms amount to

$$\sum_{\substack{n_1+n_2 \geq 2 \\ n_1 \leq N_1, n_2 \leq N_2}} \binom{N_1}{n_1}\binom{N_2}{n_2} \sum_{\substack{a+b \geq 2 \\ a \leq n_1, b \leq n_2}} (-1)^{n_1+n_2-a-b} \binom{n_1}{a}\binom{n_2}{b}$$

$$\times \ln \left[\prod_{i_1=0}^{\max\{0,a-1\}} \left(1 - \frac{i_1}{N_1}\right) \prod_{i_2=0}^{\max\{0,b-1\}} \left(1 - \frac{i_2}{N_2}\right) \right]. \tag{3.22}$$

For fixed a and b (with $a+b \geq 2$) the prefactor of each logarithm is the opposite of (3.16). Hence, the sum (3.22) equals $\ln(N_1! N_1^{-N_1}) + \ln(N_2! N_2^{-N_2})$, leading to an alternative expression for the entropy MPCE:

$$\frac{S_{N_1,N_2}}{k_B} = N_1 \left[\frac{3}{2} - \ln(\rho_1 \Lambda_1^3) \right] + N_2 \left[\frac{3}{2} - \ln(\rho_2 \Lambda_2^3) \right]$$

$$- N_1 \int P_{10} \ln P_{10} - N_2 \int P_{01} \ln P_{01} - \sum_{\substack{n_1+n_2 \geq 2 \\ n_1 \leq N_1, n_2 \leq N_2}} \binom{N_1}{n_1}\binom{N_2}{n_2}$$

$$\times \sum_{\substack{a+b \geq 2 \\ a \leq n_1, b \leq n_2}} (-1)^{n_1+n_2-a-b} \binom{n_1}{a}\binom{n_2}{b} \int P_{ab} \ln g_{ab}. \tag{3.23}$$

We argue that, using the sum rules (3.8), it is always possible to 'complete' the integrals $\int P_{ab} \ln g_{ab}$ in such a way that the integrand behaves like N^{-a-b} at infinity. In a way analogous to the one-component case, this is accomplished by adding and subtracting a suitable constant (which is $-N_1/2 - N_2/2$ to second order). If ensemble invariance must hold for the entropy MPCE, the constant needed is necessarily the opposite of the number that, added to

$$\binom{N_1}{n_1}\binom{N_2}{n_2} \sum_{\substack{a+b \geq 2 \\ a \leq n_1, b \leq n_2}} (-1)^{n_1+n_2-a-b} \binom{n_1}{a}\binom{n_2}{b}$$

$$\times \ln \left[\prod_{i_1=0}^{\max\{0,a-1\}} \left(1 - \frac{i_1}{N_1}\right) \prod_{i_2=0}^{\max\{0,b-1\}} \left(1 - \frac{i_2}{N_2}\right) \right], \tag{3.24}$$

gives back a $\mathcal{O}(1)$ quantity as an outcome.

If either n_1 or n_2 is zero, we are led to the one-component case, which is treated in appendix A. Otherwise, when both n_1 and n_2 are non-zero, the sum in (3.24) yields

$$\sum_{\substack{a+b \geq 2 \\ a \leq n_1, b \leq n_2}} (-1)^{n_1+n_2-a-b} \binom{n_1}{a}\binom{n_2}{b} \left\{ \sum_{i_1=0}^{\max\{0,a-1\}} \ln\left(1 - \frac{i_1}{N_1}\right) + \sum_{i_2=0}^{\max\{0,b-1\}} \ln\left(1 - \frac{i_2}{N_2}\right) \right\}$$

$$= \sum_{i_1=0}^{n_1-1} \ln\left(1 - \frac{i_1}{N_1}\right) \sum_{\substack{a+b \geq 2 \\ i_1+1 \leq a \leq n_1, b \leq n_2}} (-1)^{n_1+n_2-a-b} \binom{n_1}{a}\binom{n_2}{b}$$

$$+ \sum_{i_2=0}^{n_2-1} \ln\left(1 - \frac{i_2}{N_2}\right) \sum_{\substack{a+b\geq 2 \\ a\leq n_1, i_2+1\leq b\leq n_2}} (-1)^{n_1+n_2-a-b} \binom{n_1}{a}\binom{n_2}{b}$$

$$= \underbrace{\sum_{b=0}^{n_2}(-1)^{n_2-b}\binom{n_2}{b}}_{0} \sum_{i_1=1}^{n_1-1}\left[\ln\left(1-\frac{i_1}{N_1}\right)\sum_{a=i_1+1}^{n_1}(-1)^{n_1-a}\binom{n_1}{a}\right]$$

$$+ \underbrace{\sum_{a=0}^{n_1}(-1)^{n_1-a}\binom{n_1}{a}}_{0} \sum_{i_2=1}^{n_2-1}\left[\ln\left(1-\frac{i_2}{N_2}\right)\sum_{b=i_2+1}^{n_2}(-1)^{n_2-b}\binom{n_2}{b}\right] = 0. \quad (3.25)$$

Hence, in order to produce an ensemble-invariant formula, a null integral correction must be added to the canonical correlation integrals associated with the distinct DFs ($n_1 = n_2 = 1$ was such a case; see equations (3.20) and (3.21)).

We reproduce hereafter the entropy MPCE in its final form:

$$\frac{S_{N_1,N_2}}{k_B} = N_1\left[\frac{5}{2} - \ln(\rho_1\Lambda_1^3)\right] + N_2\left[\frac{5}{2} - \ln(\rho_2\Lambda_2^3)\right] - N_1\int P_{10}\ln P_{10} - N_2\int P_{01}\ln P_{01}$$

$$- \sum_{\substack{n_1+n_2\geq 2 \\ n_1\leq N_1, n_2\leq N_2}} \binom{N_1}{n_1}\binom{N_2}{n_2} \sum_{\substack{a+b\geq 2 \\ a\leq n_1, b\leq n_2}} (-1)^{n_1+n_2-a-b}\binom{n_1}{a}\binom{n_2}{b}$$

$$\times \int P_{ab}\ln g_{ab} - N_1 - N_2. \quad (3.26)$$

In the rhs of the above formula, the first four terms are the only ones which survive in the absence of any interaction between particles and, as such, constitute the ideal contribution to the entropy. Instead, the last two terms arise from the resummation of the numbers (A.7) for each species. When suitably absorbed into the integrals $\int P_{a0}\ln g_{a0}$ and $\int P_{0b}\ln g_{0b}$, these terms eventually make the canonical-ensemble entropy expansion look like the grand-canonical one. In particular, for a large ($N_1, N_2 \gg 1$) and *homogeneous* binary system, the entropy expansion starts as follows:

$$\frac{S_{N_1,N_2}}{k_B} = N_1\left[\frac{5}{2} - \ln(\rho_1\Lambda_1^3)\right] + N_2\left[\frac{5}{2} - \ln(\rho_2\Lambda_2^3)\right]$$

$$- \tfrac{1}{2}\rho_1^2 \int d^3r\, d^3r'[g_{20}(\mathbf{r},\mathbf{r}')\ln g_{20}(\mathbf{r},\mathbf{r}') - g_{20}(\mathbf{r},\mathbf{r}') + 1]$$

$$- \rho_1\rho_2 \int d^3r\, d^3r'\,[g_{11}(\mathbf{r},\mathbf{r}')\ln g_{11}(\mathbf{r},\mathbf{r}') - g_{11}(\mathbf{r},\mathbf{r}') + 1]$$

$$- \tfrac{1}{2}\rho_2^2 \int d^3r\, d^3r'\,[g_{02}(\mathbf{r},\mathbf{r}')\ln g_{02}(\mathbf{r},\mathbf{r}') - g_{02}(\mathbf{r},\mathbf{r}') + 1] + \cdots. \quad (3.27)$$

If the above expansion is truncated after its pair-correlation terms, this formula provides a low-density approximation for the entropy of the mixture.

The MPCE of the entropy derived above applies to a binary mixture of interacting *point* particles. However, the formula does not change even if the particles possess further degrees of freedom. In fact, the same comments that we made at the end of section 2 still apply. A case of this sort is a mixture of spheres and spherocylinders. We shall analyse

14

in a future publication the phase diagram of this model system in terms of the RMPE. Here, we focus our attention on the general structure of the *distinct* two-body reduced DF, since the case of the sphere–sphere DF is obvious while the symmetries of the self DF for spherocylinders have been already discussed in section 2.

In order to find the maximum number of independent scalar variables that intervene in the calculation of $g_{11}(\xi, \mathbf{q})$, it is convenient to work in a reference system where the spherocylinder (species 1) is centred at the origin and lies along the z axis. In this reference frame, the position of the sphere (species 2) can be parameterized in terms of two variables only, namely the length r_{12} and colatitude ϑ_{12} of the vector joining the two centres of mass (as usual, $\varphi_{12} = 0$ for a convenient choice of the x axis). The same result is obviously obtained when viewing the situation from a reference system where the sphere is centred at the origin. In this case, the need for two further θ_{12} and ϕ_{12} variables (see section 2) is only apparent, since the z axis and the x axis can be chosen in such a way that such two variables vanish altogether. If the axes of the spherocylinders are frozen and parallel to each other, no further simplification occurs, and the distinct two-body DF is again a function of r_{12} and ϑ_{12}.

The numerical calculation of g_{11} for a homogeneous mixture is carried out as follows. After taking $\xi = (\mathbf{r}, \theta, \phi)$, $\mathbf{x} = \mathbf{q} - \mathbf{r}$, and invoking homogeneity, we have

$$\frac{\rho_1}{4\pi}\rho_2\, g_{11}(\xi, \mathbf{q}) = \frac{1}{4\pi V} \int \mathrm{d}^5\xi \left\langle \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \delta^5(\Xi_i - \xi)\delta^3(\mathbf{Q}_j - \mathbf{q}) \right\rangle$$

$$= \frac{1}{4\pi V} \left\langle \sum_{i,j} \int \mathrm{d}\theta\, \mathrm{d}\phi\, \delta(\Theta_i - \theta)\delta(\Phi_i - \phi) \int \mathrm{d}^3r\, \delta^3(\mathbf{R}_i - \mathbf{r})\delta^3(\mathbf{Q}_j - \mathbf{r} - \mathbf{x}) \right\rangle$$

$$= \frac{N_1}{4\pi V} \left\langle \sum_{j=1}^{N_2} \delta^3(\mathbf{Q}_j - (\mathbf{R}_1 + \mathbf{x})) \right\rangle. \tag{3.28}$$

Hence, $\rho_2\, g_{11}(\xi, \mathbf{q})$ is the average *density* of spheres in $\mathbf{R}_1 + \mathbf{x}$ when a spherocylinder is centred in $\mathbf{R}_1$. An explicit formula is

$$g_{11}(r_{12}, \vartheta_{12}) \simeq \frac{\Delta \mathcal{N}_2(r_{12}, \vartheta_{12})}{\rho_2 \cdot 2\pi r_{12}^2 \sin\vartheta_{12}\Delta r_{12}\Delta\vartheta_{12}}, \tag{3.29}$$

$\Delta\mathcal{N}_2(r_{12}, \vartheta_{12})$ being the number of spheres within a tiny spherical ring of volume $2\pi r_{12}^2 \sin\vartheta_{12}\Delta r_{12}\Delta\vartheta_{12}$, centred at the position specified by r_{12} and ϑ_{12}.

4. Conclusions

In this paper, we have outlined a constructive method for building up the entropy multiparticle-correlation expansion in the canonical ensemble, term by term, for both pure and mixed systems of classical particles. The aim of this effort is twofold:

(i) to unveil the hidden combinatorial structure behind the expansion;

(ii) to set the stage for an application of the entropy-based ordering criterion introduced by Giaquinta and co-workers to the phase diagram of a binary mixture of hard spheres and spherocylinders.

In this respect, we have discussed here the general symmetries owned by the three pair distribution functions. A detailed analysis of the phase diagram of the model in the framework provided by the zero-RMPE criterion will be the object of a forthcoming publication.

Appendix A. Ensemble invariance of the entropy MPCE

In this appendix, an argument appearing in [12] is reproduced for the reader's convenience. This argument deals with the overall constant term that appears in equation (2.12) when we eliminate the normalized DFs in favour of the reduced DFs through equation (2.6).

For one-component systems, this constant amounts to

$$\sum_{n=2}^{N} \binom{N}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} \ln \frac{(N-1)(N-2)\cdots(N-a+1)}{N^{a-1}}. \tag{A.1}$$

For each a value, the prefactor of the respective logarithm is just the number (2.14). Hence, the sum (A.1) equals $\ln(N! N^{-N})$, thus yielding a new form of the entropy MPCE:

$$\frac{S_N}{k_B} = N\left[\frac{3}{2} - \ln(\rho \Lambda^3)\right] - N \int P_1 \ln P_1 - \sum_{n=2}^{N} \binom{N}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} \int P_{1\ldots a} \ln g_{1\ldots a}, \tag{A.2}$$

where $\rho = N/V$.

We now show that any single term in the sum over n at (A.1) is extensive. In fact, we have first

$$\binom{N}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} \sum_{k=1}^{a-1} \ln\left(1 - \frac{k}{N}\right) = \binom{N}{n} \sum_{k=1}^{n-1} \left[\ln\left(1 - \frac{k}{N}\right) \sum_{a=k+1}^{n} (-1)^{n-a} \binom{n}{a}\right]. \tag{A.3}$$

We prove in appendix B that

$$\sum_{a=k+1}^{n} (-1)^{n-a} \binom{n}{a} = (-1)^{n-1-k} \binom{n-1}{k}. \tag{A.4}$$

Then,

$$\binom{N}{n} \sum_{a=2}^{n} (-1)^{n-a} \binom{n}{a} \ln \frac{(N-1)(N-2)\cdots(N-a+1)}{N^{a-1}}$$

$$= -\binom{N}{n} \sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} \ln\left(1 - \frac{k}{N}\right)$$

$$= \binom{N}{n} \left\{\frac{1}{N} \sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} k + \frac{1}{2N^2} \sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} k^2 + \ldots\right\}. \tag{A.5}$$

In appendix B, we also show that

$$\sum_{k=1}^{n-1}(-1)^{n-k}\binom{n-1}{k}k^d = \begin{cases} 0, & \text{for } d=1,\ldots,n-2 \\ -(n-1)!, & \text{for } d=n-1. \end{cases} \quad \text{(A.6)}$$

In conclusion, we obtain

$$\binom{N}{n}\sum_{a=2}^{n}(-1)^{n-a}\binom{n}{a}\ln\frac{(N-1)(N-2)\cdots(N-a+1)}{N^{a-1}} \sim -\frac{N}{n(n-1)}. \quad \text{(A.7)}$$

As discussed in the main text, the integrals in equation (A.2) cannot be easily computed numerically since, for finite N, the system boundary also contributes in a significant way. However, upon taking advantage of the canonical-ensemble sum rules for the reduced DFs (see equation (2.7)), it should always be possible to make every integrand in equation (A.2) sufficiently small at large distances. This is accomplished by adding (and subtracting) a quantity equal to the number in (A.7), with the result of leaving an overall $\mathcal{O}(1)$ number outside the integral. Furthermore, the new form of the integral can be made identical to the (so-called) fluctuation integral of the same order which appears in the grand-canonical-ensemble expansion.

We show this explicitly for the third-order term in the entropy expansion, which, when expressed in terms of the reduced DFs, reads

$$-\binom{N}{3}\ln\frac{(N-1)^2}{N(N-2)} - \frac{1}{6}\rho^3\int \mathrm{d}^3r_1\,\mathrm{d}^3r_2\,\mathrm{d}^3r_3\,P_1P_2P_3g_{123}\ln\frac{g_{123}}{g_{12}g_{13}g_{23}}. \quad \text{(A.8)}$$

In order to conform to the grand-canonical-ensemble expansion, we have to add the integral

$$-\tfrac{1}{6}\rho^3\int \mathrm{d}^3r_1\,\mathrm{d}^3r_2\,\mathrm{d}^3r_3\,P_1P_2P_3(-g_{123}+3g_{12}g_{13}-3g_{12}+1), \quad \text{(A.9)}$$

which, in view of the canonical-ensemble sum rules (2.4) and (2.7), is equal to $-N/6$. This number is exactly the same constant that must be subtracted from

$$-\binom{N}{3}\ln\frac{(N-1)^2}{N(N-2)} \quad \text{(A.10)}$$

in order to produce an $\mathcal{O}(1)$ constant. As a *caveat*, we note that

$$-\frac{1}{6}\rho^3\int \mathrm{d}^3r_1\,\mathrm{d}^3r_2\,\mathrm{d}^3r_3\,P_1P_2P_3[\kappa g_{123}+(1-2\kappa)g_{12}g_{13}+(\kappa-2)g_{12}+1] = -\frac{N}{6} \quad \text{(A.11)}$$

for any real κ, not simply -1.

Appendix B. Two combinatorial identities

In this appendix, the formulae (A.4) and (A.6) are proved by induction.

First, we prove that, for any $n\geq 2$ and $1\leq k\leq n-1$,

$$\sum_{a=k+1}^{n}(-1)^{n-a}\binom{n}{a} = (-1)^{(n-1)-k}\binom{n-1}{k}. \quad \text{(B.1)}$$

Equation (B.1) is also valid for $k = 0$ and any $n \geq 1$. We argue by induction over n. For $n = 2$ and $k = 1$, formula (B.1) is trivially correct. Then, assuming that the formula is correct for an arbitrary fixed n and all positive $k < n$, we calculate the lhs of equation (B.1) for $n + 1$ and any $1 \leq k \leq n - 1$ (the case $k = n$ is obvious):

$$
\begin{aligned}
\sum_{a=k+1}^{n+1} (-1)^{(n+1)-a} \binom{n+1}{a} &= -\sum_{a=k+1}^{n} (-1)^{n-a} \left[\binom{n}{a} + \binom{n}{a-1} \right] + 1 \\
&= -(-1)^{(n-1)-k} \binom{n-1}{k} + \sum_{a=k}^{n} (-1)^{n-a} \binom{n}{a} \\
&= (-1)^{n-k} \left[\binom{n-1}{k} + \binom{n-1}{k-1} \right] = (-1)^{n-k} \binom{n}{k},
\end{aligned}
\tag{B.2}
$$

which is just the rhs of equation (B.1), but for $n + 1$ which replaces n.

Next, we show that, for $n \geq 3$,

$$
\sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} k^d = \begin{cases} 0, & \text{for } d = 1, \ldots, n-2 \\ -(n-1)!, & \text{for } d = n-1, \end{cases}
\tag{B.3}
$$

while it is trivial to check that

$$
\sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} k^d = -1 \qquad (n = 2, d = 1).
\tag{B.4}
$$

Arguing inductively, let us suppose that equation (B.3) is valid for a given value n (checking this for $n = 3$ is immediate), and see what happens for $n + 1$:

$$
\begin{aligned}
\sum_{k=1}^{n} (-1)^{n+1-k} \binom{n}{k} k^d &= -\sum_{k=1}^{n-1} (-1)^{n-k} \binom{n}{k} k^d - n^d \\
&= -\sum_{k=1}^{n-1} (-1)^{n-k} \left[\binom{n-1}{k} + \binom{n-1}{k-1} \right] k^d - n^d \\
&= -\sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} k^d + \sum_{k=0}^{n-2} (-1)^{n-k} \binom{n-1}{k} (k+1)^d - n^d \\
&= -\sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} k^d + \sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} (k+1)^d + (-1)^n \\
&= -\sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} k^d + \sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} \sum_{m=0}^{d} \binom{d}{m} k^m + (-1)^n \\
&= -\sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} k^d + \sum_{m=0}^{d} \binom{d}{m} \left\{ \sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k} k^m \right\} + (-1)^n.
\end{aligned}
\tag{B.5}
$$

Focusing on this intermediate result, we distinguish three cases.

(i) For $1 \le d < n-1$, the first term is zero as is the second for $m > 0$. As for the rest,

$$\sum_{k=1}^{n-1}(-1)^{n-k}\binom{n-1}{k} + (-1)^n = -\sum_{k=1}^{n-1}(-1)^{(n-1)-k}\binom{n-1}{k} + (-1)^n$$

$$= -\sum_{k=0}^{n-1}(-1)^{(n-1)-k}\binom{n-1}{k} = -(1-1)^{n-1} = 0. \tag{B.6}$$

(ii) For $d = n-1$ instead, the first term is $(n-1)!$, while the second is zero for $0 < m < d$. The rest is equal to

$$(n-1)! + \sum_{k=1}^{n-1}(-1)^{n-k}\binom{n-1}{k}k^{n-1} + \sum_{k=1}^{n-1}(-1)^{n-k}\binom{n-1}{k} + (-1)^n$$

$$= \sum_{k=1}^{n-1}(-1)^{n-k}\binom{n-1}{k} + (-1)^n = 0, \tag{B.7}$$

like in the previous case.

(iii) Finally, and given all the above results, for $d = n$ we have

$$-\sum_{k=1}^{n-1}(-1)^{n-k}\binom{n-1}{k}k^n + \binom{n}{0}\sum_{k=1}^{n-1}(-1)^{n-k}\binom{n-1}{k}$$

$$+ \sum_{m=1}^{n-2}\binom{n}{m}\sum_{k=1}^{n-1}(-1)^{n-k}\binom{n-1}{k}k^m - \binom{n}{n-1}(n-1)!$$

$$+ \binom{n}{n}\sum_{k=1}^{n-1}(-1)^{n-k}\binom{n-1}{k}k^n + (-1)^n = -n!, \tag{B.8}$$

as we wanted to show.

References

[1] See for example Löwen H, 1994 *Phys. Rep.* **237** 249
   Monson P A and Kofke D A, 2000 *Adv. Chem. Phys.* **115** 113
[2] Giaquinta P V and Giunta G, 1992 *Physica* A **187** 145
   Giaquinta P V, Giunta G and Prestipino Giarritta S, 1992 *Phys. Rev.* A **45** R6966
[3] Green H S, 1952 *The Molecular Theory of Fluids* (Amsterdam: North-Holland)
[4] Nettleton R E and Green M S, 1958 *J. Chem. Phys.* **29** 1365
[5] Cheng A, Klein M L and Caccamo C, 1993 *Phys. Rev. Lett.* **71** 1200
   Lomba E, Lopez-Martin J L, Cataldo H M and Tejero C F, 1994 *Phys. Rev.* E **49** 5164
   Lomba E and Almarza N E, 1994 *J. Chem. Phys.* **100** 8367
[6] Saija F, Prestipino S and Giaquinta P V, 2000 *J. Chem. Phys.* **113** 2806
   Saija F, Prestipino S and Giaquinta P V, 2001 *J. Chem. Phys.* **115** 7586
   Saija F, Saitta A M and Giaquinta P V, 2003 *J. Chem. Phys.* **119** 3587
[7] Saija F, Pastore G and Giaquinta P V, 1998 *J. Phys. Chem.* B **102** 10368
   Saija F and Giaquinta P V, 2002 *J. Phys. Chem.* B **106** 2035
   Saija F and Giaquinta P V, 2002 *J. Chem. Phys.* **117** 5780

  [8] Costa D, Saija F and Giaquinta P V, 1998 *Chem. Phys. Lett.* **283** 86
      Costa D, Saija F and Giaquinta P V, 1999 *Chem. Phys. Lett.* **299** 252 (erratum)
      Costa D, Micali F, Saija F and Giaquinta P V, 2002 *J. Phys. Chem.* B **106** 12297
      Costa D, Saija F and Giaquinta P V, 2003 *J. Phys. Chem.* B **107** 9514
  [9] Cuetos A, Martinez-Haya B, Rull L F and Lago S, 2002 *J. Chem. Phys.* **117** 2934
      Cuetos A, Martinez-Haya B, Rull L F and Lago S, 2002 *J. Chem. Phys.* **117** 11405 (erratum)
 [10] Donato M G, Prestipino S and Giaquinta P V, 1999 *Eur. Phys. J.* B **11** 621
 [11] Saija F, Saitta A M and Giaquinta P V, 2003 *J. Chem. Phys.* **119** 3587
 [12] Prestipino S and Giaquinta P V, 1999 *J. Stat. Phys.* **96** 135
      Prestipino S and Giaquinta P V, 2000 *J. Stat. Phys.* **98** 507
 [13] Adams M, Dogic Z, Keller S L and Fraden S, 1998 *Nature* **393** 349
 [14] Dogic Z, Frenkel D and Fraden S, 2000 *Phys. Rev.* E **62** 3925
 [15] Baranyai A and Evans D J, 1989 *Phys. Rev.* A **40** 3817
 [16] Lebowitz J L and Percus J K, 1961 *Phys. Rev.* **122** 1675

JSTAT (2004) P09008