Finite-size effects and thermodynamic accuracy in many-particle systems

Benedikt M. Reible,[1] Julian F. Hille,[2] Carsten Hartmann,[3] and Luigi Delle Site [2,*]

[1]*University of Leipzig, Institute of Theoretical Physics, Brüderstr. 16, 04103 Leipzig, Germany*
[2]*Freie Universität Berlin, Institute of Mathematics, Arnimallee 6, 14195 Berlin, Germany*
[3]*Brandenburgische Technische Universität, Institute of Mathematics, Konrad-Wachsmann-Allee 1, 03046 Cottbus, Germany*

Finite-size effects arise when a sample of particles is not sufficient to provide a statistically satisfactory description of the bulk environment of a physical system. As a consequence, a reliable estimate of finite-size effects in many-particle systems is key to judge the validity of a theoretical model or the accuracy of a numerical simulation. In this context, we propose the use of a theorem on the free-energy cost for separating a system into smaller independent subsystems [J. Stat. Mech.: Theory Exp. (2017) 083201; Lett. Math. Phys. **112**, 97 (2022)] to estimate the relevance of finite-size effects in thermodynamic quantities from computer simulations. The key aspect of this study is that for two-body potentials, as mostly occurring in physics, the method requires only two-body distribution functions and the particle number density. The calculation of the involved physical quantities can be done numerically on a three-dimensional grid. In some cases even analytical estimates are possible and as an example the uniform interacting electron gas in the ground state is considered; we derive an approximating scaling law for the finite-size effects.

I. INTRODUCTION

Many-particle systems are one of the most popular topics in modern research, and it crosses different fields and disciplines: condensed matter, chemical physics, materials science, biophysics, to name a few. Theoretical models and modern simulation techniques are developing steadily so that realistic systems can be treated at high microscopic accuracy. A key problem in modeling and simulation of such systems is that artifacts can be induced by the limited size of a system and the limited simulation time allowed by current computational technology. The size of the system and simulation time are often small compared with an optimal or realistic system's size and an optimal or realistic timescale. The optimal system's size and the optimal timescale must include, as much as possible, the essential characteristics of the system in question so that the corresponding results can be generalized to larger systems and longer timescales (see, e.g., Refs. [1, 2] and references therein).

In this work we will not treat timescales and instead restrict ourselves to the finite-size aspect. The problem of dealing with systems whose current computationally affordable size is not sufficient for a realistic representation of the system is of course well known in the community, and several approaches have been developed to amend for it. For example, in

molecular simulation the use of periodic boundary conditions partly alleviates the problem, although if the unit cell is too small to provide an accurate statistical description of the bulk environment, then one has the paradox that the infinite replicas are actually modeling, for example, a liquid made of physically unrealistic droplets. Other methods that account for the finite-size effects, e.g., from the field of electronic structure calculations (see, e.g., Refs. [3–5]), or from molecular dynamics (see, e.g., Refs. [6,7]), are based on the extrapolation of the large-scale behavior from very small systems. This work adopts a complementary point of view: Our approach is based on previous work by some of the authors that provides upper and lower bounds of the free energy of a many-particle system when the system is divided into separated noninteracting subsystems. In essence, the theorem expresses the largest and smallest free energy cost possible to construct an interface so that the system can be represented as the union of smaller independent subsystems [8,9].

The key point of the concept proposed here is that if the largest free-energy cost possible, given by the theorem, is small compared with some characteristic energy that is supposed to describe bulk properties, then the current system size must be sufficient for representing the statistical or thermodynamic properties of the bulk environment since interface effects represent a negligible thermodynamic contribution. The theorem is valid for classical [8] and quantum [9] systems. We will show in this work that, when dealing with two-body potentials, the theorem can be easily applied to realistic systems with the only requirement that the particle density and the particle-particle radial distribution function are available. Such quantities are routinely available in simulations or even offered by experiments. As a showcase, we will treat the uniform gas of interacting electrons because in

---

this case it is technically simple to show in a clear way the essence of the idea proposed.

II. TWO-SIDED BOGOLIUBOV INEQUALITY, TWO-BODY POTENTIALS, AND RADIAL DISTRIBUTION FUNCTION

In previous work we have defined the upper and lower bound of the free-energy cost ΔF to divide a system of particles into smaller independent subsystems in the classical [8] and quantum [9] case. The system can be partitioned in several noninteracting subsystems, but for simplicity we will consider here the simple (although already rather relevant) case that a system is divided into two (equal) noninteracting subsystems; the extension to more subsystems is then straightforward. The theorem can be summarized as

$$\mathbf{E}_f[U] \leqslant \Delta F \leqslant \mathbf{E}_{f_1,f_2}[U], \tag{1}$$

where $f(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is the position probability density function of N particles confined in a volume $\Omega \subset \mathbb{R}^3$ with Hamiltonian H. For the classical case we have $f = \frac{e^{-\beta H(\mathbf{r}_1,\ldots,\mathbf{r}_N)}}{Z}$ with $Z = \frac{1}{h^{3N}N!}\int_{\Omega^N}\exp[-\beta H(\mathbf{r}_1, \ldots, \mathbf{r}_N)]d\mathbf{r}_1 \cdots d\mathbf{r}_N$, with $\beta = 1/k_BT$, k_B being the Boltzmann constant, T the temperature, and h the Planck constant. When the system has

been divided into two independent subsystems of n and $N - n$ particles confined in volumes Ω_1 and Ω_2, respectively ($\Omega = \Omega_1 \cup \Omega_2$), we have $f_1(\mathbf{r}_1, \ldots, \mathbf{r}_n) = e^{-\beta H_1(\mathbf{r}_1,\ldots,\mathbf{r}_n)}/Z_1$ and $f_2(\mathbf{r}_{n+1}, \ldots, \mathbf{r}_N) = e^{-\beta H_2(\mathbf{r}_{n+1},\ldots,\mathbf{r}_N)}/Z_2$ as the probability density functions of n and $N - n$ particles, respectively, and $H = H_1 + H_2 + U$, where U is the interaction potential between particles across the surface of separation. Similarly, for the quantum case $\mathbf{E}_f[U]$ and $\mathbf{E}_{f_1,f_2}[U]$ are calculated as $\text{Tr}(\Gamma U)$ and $\text{Tr}(\Gamma_1\Gamma_2 U)$, respectively, with Γ and $\Gamma_1\Gamma_2$ being the density matrix of the system and the density matrix of the two (noninteracting) separated subsystems; here Γ_1 and Γ_2 commute because they describe two noninteracting subsystems, and thus they are independent of each other (and act on different domains). In essence, in the quantum case, $f(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ and $f_1(\mathbf{r}_1, \ldots, \mathbf{r}_n) \otimes f_2(\mathbf{r}_{n+1}, \ldots, \mathbf{r}_N)$ are the diagonal terms of the density matrices Γ and $\Gamma_1 \otimes \Gamma_2$, respectively (see Appendix A). If one considers two-body potentials, $U(\mathbf{r}_i - \mathbf{r}_j)$, as those mostly occurring in particle-based representations of matter, the formulas for $\mathbf{E}_f[U]$ in both the classical and the quantum case are highly simplified and reduced to a two-body integral. Assuming that interactions mediated by U are only between particles in different domains Ω_1 and Ω_2, but not within either of them, we have

$$\mathbf{E}_f[U] = 2\sum_{i,j=1}^{N}\left(\int_{\Omega_1}\int_{\Omega_2}\left[\int_{\Omega^{N-2}}f(\mathbf{r}_1, \ldots, \mathbf{r}_N)\,d\mathbf{r}_1\cdots d\mathbf{r}_{i-1}d\mathbf{r}_{i+1}\cdots d\mathbf{r}_{j-1}d\mathbf{r}_{j+1}\cdots d\mathbf{r}_N\right]U(\mathbf{r}_i - \mathbf{r}_j)d\mathbf{r}_jd\mathbf{r}_i\right), \tag{2}$$

with $\mathbf{r}_i \in \Omega_1$ for all $i = 1, \ldots, N$, $\mathbf{r}_j \in \Omega_2$ for all $j = 1, \ldots, N$. In essence the integral above expresses the interaction of two particles, both spanning the entire domain, with the constrain that they are not in the same subdomain at the same time. The assumption $\Omega = \Omega_1 \cup \Omega_2$ implies that $\Omega \times \Omega = (\Omega_1 \times \Omega_1) \cup (\Omega_1 \times \Omega_2) \cup (\Omega_2 \times \Omega_1) \cup (\Omega_2 \times \Omega_2)$ and the integral of the total energy can be decomposed accordingly. Since the interdomain interactions are symmetric, the contributions to the free energy coming from $\Omega_1 \times \Omega_2$ and $\Omega_2 \times \Omega_1$ will be the same and (in case of indistinguishable particles) they will give the cross-interaction integral of Eq. (2).[1]

We define

$$\int_{\Omega^{N-2}}f(\mathbf{r}_1, \ldots, \mathbf{r}_N)\,d\mathbf{r}_1\cdots d\mathbf{r}_{i-1}d\mathbf{r}_{i+1}\cdots d\mathbf{r}_{j-1}d\mathbf{r}_{j+1}\cdots d\mathbf{r}_N =: \hat{g}(\mathbf{r}_i, \mathbf{r}_j). \tag{3}$$

Considering all particles indistinguishable, and taking the partitioning of all possible pairs obtained from the N particles into account, the quantity $\frac{N!}{(N-2)!\rho^2}\hat{g}(\mathbf{r}_i, \mathbf{r}_j) = \frac{N(N-1)}{\rho^2}\hat{g}(\mathbf{r}_i, \mathbf{r}_j) =: g(\mathbf{r}_i, \mathbf{r}_j)$, with $\rho = N/V$ being the average number density, is defined as the normalized two-body correlation function (see, e.g., Ch. 4.6 of Ref. [10]). For an isotropic system, the two-body correlation function corresponds to the particle-particle radial distribution function $g(|\mathbf{r}_i - \mathbf{r}_j|)$, since a privileged direction does not exist in an isotropic system. Under the hypothesis that we treat isotropic systems, the idea above can be put into practice because the radial distribution function is routinely available from simulations and/or experiments. Following the definition of Eq. (3), one has

$$\mathbf{E}_f[U] = 2\frac{\rho^2}{N(N-1)}\sum_{i,j}\int_{\Omega_1}\int_{\Omega_2}U(\mathbf{r}_i - \mathbf{r}_j)g(\mathbf{r}_i, \mathbf{r}_j)\,d\mathbf{r}_jd\mathbf{r}_i. \tag{4}$$

For the representation of the system as two noninteracting subsystems with n and $N - n$ particles, respectively, one finds that $f(\mathbf{r}_1, \ldots, \mathbf{r}_N) = f_1(\mathbf{r}_1, \ldots, \mathbf{r}_n)f_2(\mathbf{r}_{n+1}, \ldots, \mathbf{r}_N)$, thus one obtains

$$\mathbf{E}_{f_1,f_2}[U] = \sum_{i=1}^{n}\left\{\sum_{j=n+1}^{N}\int_{\Omega_1}\int_{\Omega_2}\left(\int_{\Omega_1^{n-1}}f_1(\mathbf{r}_1, \ldots, \mathbf{r}_n)\,d\mathbf{r}_1\cdots d\mathbf{r}_{i-1}d\mathbf{r}_{i+1}\cdots d\mathbf{r}_n\right.\right.$$

$$\left.\left.\times \int_{\Omega_2^{N-n-1}}f_2(\mathbf{r}_{n+1}, \ldots, \mathbf{r}_N)\,d\mathbf{r}_{n+1}\cdots d\mathbf{r}_{j-1}d\mathbf{r}_{j+1}\cdots d\mathbf{r}_N\right)U(\mathbf{r}_i - \mathbf{r}_j)\mathbf{1}_{\{|x_i-x_j|\geqslant\sigma\}}\,d\mathbf{r}_jd\mathbf{r}_i\right\}, \tag{5}$$

---

[1]Since Cartesian products such as $\Omega_1 \times \Omega_2$ are sets of ordered pairs, $\Omega_1 \times \Omega_2$ and $\Omega_2 \times \Omega_1$ can never be the same.

where $\mathbf{1}_{\{x \in A\}}$ denotes the indicator function of a set A that is equal to one if $x \in A$ and zero otherwise, and $|x_i - x_j|$ denotes the Cartesian distance between two particles along the direction perpendicular to the surface (in the yz plane) that separates the system in subsystems.

Following a suggestion by Lebowitz and Lieb [11,12], the condition $|x_i - x_j| \geqslant \sigma$ corresponds to a short-distance cutoff in the particle-particle interactions across the interface, equivalent to the creation of a corridor at the interface that divides the large system in disjoint subsystems. Such a corridor (or cutoff) avoids possible issues with singular potentials, such as Coulomb or Lennard-Jones potentials. In fact, since the subsystems are independent, particles belonging to different domains can come arbitrarily close to each other. For our practical application, the corridor shall be small enough so that its size is negligible compared with the size of the system. At this point, note that one has

$$\int_{\Omega_1^{n-1}} f_1(\mathbf{r}_1, \dots, \mathbf{r}_n) \, d\mathbf{r}_1 \cdots d\mathbf{r}_{i-1} d\mathbf{r}_{i+1} \cdots d\mathbf{r}_n = \hat{\rho}_1(\mathbf{r}_i) = \frac{\rho_1(\mathbf{r}_i)}{n},$$

with $\int_{\Omega_1} \rho_1(\mathbf{r}) d\mathbf{r} = n$, and similarly

$$\int_{\Omega_2^{N-n-1}} f_2(\mathbf{r}_{n+1}, \dots, \mathbf{r}_N) \, d\mathbf{r}_{n+1} \cdots d\mathbf{r}_{j-1} d\mathbf{r}_{j+1} \cdots d\mathbf{r}_N = \hat{\rho}_2(\mathbf{r}_j) = \frac{\rho_2(\mathbf{r}_j)}{N-n},$$

with $\int_{\Omega_2} \rho_2(\mathbf{r}) d\mathbf{r} = N - n$. The relevant consequence is that it is enough to know the respective average three-dimensional particle density $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$ of the two subsystems, so that

$$\mathbf{E}_{f_1, f_2}[U] = \frac{1}{n(N-n)} \sum_{i=1}^{n} \left\{ \sum_{j=n+1}^{N} \left(\int_{\Omega_1} \int_{\Omega_2} \rho_1(\mathbf{r}_i) \rho_2(\mathbf{r}_j) U(\mathbf{r}_i - \mathbf{r}_j) \mathbf{1}_{\{|x_i - x_j| \geqslant \sigma\}} \, d\mathbf{r}_j d\mathbf{r}_i \right) \right\}. \quad (6)$$

The expressions of Eq. (4) and Eq. (6) can be simplified in case of particles which are indistinguishable. In this case one has $\sum_{ij} = N(N-1)$, $\sum_{i=1}^{n} \sum_{j=n+1}^{N} = n(N-n)$, $\mathbf{r}_i = \mathbf{r}$, $\mathbf{r}_j = \mathbf{r}'$, and thus

$$\mathbf{E}_f[U] = 2\rho^2 \int_{\Omega_1} \int_{\Omega_2} U(\mathbf{r} - \mathbf{r}') g(\mathbf{r}, \mathbf{r}') \, d\mathbf{r}' d\mathbf{r}, \quad (7)$$

with $\mathbf{r} \in \Omega_1$ and $\mathbf{r}' \in \Omega_2$, and

$$\mathbf{E}_{f_1, f_2}[U] = \int_{\Omega_1} \int_{\Omega_2} \rho_1(\mathbf{r}) \rho_2(\mathbf{r}') U(\mathbf{r} - \mathbf{r}') \mathbf{1}_{\{|x - x'| \geqslant \sigma\}} \, d\mathbf{r}' d\mathbf{r} \quad (8)$$

again with $\mathbf{r} \in \Omega_1$ and $\mathbf{r}' \in \Omega_2$.

To obtain a criterion to quantify the thermodynamic accuracy that one can reach at a given size of the system, we define a quality parameter which measures the estimate of ΔF as introduced above, relative to some characteristic reference energy of the system:

$$q := \frac{|\Delta F|}{|E_{\text{ref}}|}. \quad (9)$$

E_{ref} can be chosen as the total energy of the system or, for example for liquid water, the total molecule-molecule hydrogen bond energy of the system.

In practice, Eq. (1) allows us to determine the largest relative error by

$$q_{\max} := \frac{\max\{|\mathbf{E}_f[U]|, |\mathbf{E}_{f_1, f_2}[U]|\}}{|E_{\text{ref}}|}. \quad (10)$$

It follows that the thermodynamic accuracy of the model, chosen to represent the bulk properties of a system, can be measured by $q_{\max}$. In practice, the criterion can be summarized as follows: a small $q_{\max}$ implies that the separation or interface energy is negligible compared with the bulk reference energy, the consequence is that the bulk description of

the system is sufficient; thus we call a small value of $q_{\max}$ high accuracy; conversely a high value of $q_{\max}$ represents low accuracy. The reference to thermodynamic accuracy is due to the fact that $\max\{|\mathbf{E}_f[U]|, |\mathbf{E}_{f_1, f_2}[U]|\}$ represents the largest free-energy cost possible in the creation of a separating interface. Since the knowledge of the free energy of a system in a given thermodynamic state implies the knowledge of all thermodynamic quantities, the estimate of the free-energy cost related to the degree of representation of the bulk implicitly measures the thermodynamic accuracy of the system.

III. APPLICATION TO A UNIFORM ELECTRON GAS IN THE GROUND STATE

The homogeneous electron gas represents a theoretical and computational reference for understanding and modeling the relevant physics of Coulomb electron-electron interactions in more complex systems [13]. In this section, we will show how the idea presented before can be applied to the homogeneous electron gas. It must be said that being a popular test system, several approaches have been (and are being) developed to account for the finite-size effects in the calculations (see, for example, Refs. [3–5]). In this context, our idea is not in competition with the other methods but actually complements them since we look at the problem from a different perspective and thus provide further information. While the method we propose can be applied to a large class of classical and quantum systems, the choice of the homogeneous electron gas, besides its popularity, is also due to the simplifications for the calculations. In fact, as it will be clear later on, several quantities are analytically or semi-analytically available, and thus the situation is optimal for a demonstrative study.

Coming to the implementation of Eqs. (7) and (8) in our current system, the first important aspect is that the potential used for simulating the uniform electron gas is the two-body Coulomb interaction between particles with negative unit charge: $\frac{1}{2}\sum_i \sum_j 1/r_{ij}$. In addition, to assure neutrality, an artificial positively charged background is introduced (jellium model, see, e.g., Ref. [14]), and the corresponding interactions are two-body-like so that the basic requirements for the application of our procedure is fulfilled; atomic units will be used for every quantity of this study. Taking into account that particles are indistinguishable and have a uniform density, one finds

$$\mathbf{E}_f[U] = \rho^2 \int_{\Omega_1} \left\{ \int_{\Omega_2} \left[\frac{g(|\mathbf{r} - \mathbf{r}'|) - 1}{|\mathbf{r} - \mathbf{r}'|} \right] d\mathbf{r}' \right\} d\mathbf{r} \qquad (11)$$

and

$$\mathbf{E}_{f_1,f_2}[U] = 0. \qquad (12)$$

Explicit calculations are reported in Appendix B. The next natural question is how one can apply the criterion of q_{max} of Eq. (10) for a quantitative estimate of the overall thermodynamic accuracy of the model, given the system size. The reference energy with respect to which one can define in this context the relative weight of $\mathbf{E}_f[U]$ and $\mathbf{E}_{f_1,f_2}[U]$ is the total average potential energy of the jellium model of N electrons [15]:

$$E_{\text{ref}} = 2\pi N \rho \int_0^{R_0} [g(r) - 1] r \, dr, \qquad (13)$$

with R_0 defined as the smallest value for which $g(r) \simeq 1$ for all $r \geqslant R_0$. The expression of Eq. (13) is nothing but the average Coulomb potential per particle of the jellium model, multiplied, in this case, by the number of particles of the specific system (size) considered, N. One may even simplify it further and consider a precise analytic approximation [16] of the average Coulomb energy per particle corresponding to Eq. (13):

$$\frac{E_{\text{ref}}}{N} = -\frac{0.916}{r_s} + a \ln \left(1 + \frac{b}{r_s} + \frac{b}{r_s^2} \right), \qquad (14)$$

where $r_s = (\frac{3}{4\pi\rho})^{1/3}$ is the Wigner-Seitz radius and $a = \frac{\ln 2 - 1}{4\pi^2} = -0.0076$, $b = 27.42$.

It follows that for the specific case of the interacting electron gas, one has

$$q_{max} = \frac{|\mathbf{E}_f[U]|}{|E_{\text{ref}}|}. \qquad (15)$$

The accuracy of the model system in representing bulk properties can be controlled by choosing a system's size such that q_{max} is below a certain threshold (chosen as a reference of acceptable accuracy).

Data for ρ and $g(|\mathbf{r} - \mathbf{r}'|)$ are available in literature (see, e.g., Refs. [15]), thus we deal with integrals in three (one particle integral) and six (two-particle integral) dimensions which can be accurately calculated, for example, on a uniform three-dimensional grid. In the current case one can reach a reasonable estimate of q_{max} also by simple analytic arguments, as will be shown later on.

It must be noticed that here we consider the gas in the ground state, thus $T = 0$. This means that we actually deal with the case where $\Delta F = \Delta E$. However, note that for $T > 0$ the density and the radial distribution function are available [17], thus following the same procedure of the example treated here, the calculation of $\mathbf{E}_f[U]$ and $\mathbf{E}_{f_1,f_2}[U]$ would provide the estimate of the largest ΔF. This means that one can access, via routinely available quantities, the error in free energy to assess the relevance of finite-size effects on the thermodynamic accuracy of a given system also for the case of finite temperatures.

IV. SCALING LAW FOR THE THERMODYNAMIC ACCURACY OF THE UNIFORM GAS OF INTERACTING ELECTRONS IN THE GROUND STATE

In first approximation, one can estimate the trend of q_{max} as a function of the (cubic) box size L at different particle densities for the uniform gas of interacting electrons in the ground state as follows. The potential energy per electron is a known quantity [see Eq. (13)]: $\epsilon_N = 2\pi\rho \int_0^{R_0} [g(r) - 1] r \, dr$, thus the total potential energy of the system is just $\mathcal{E}_N = N\epsilon_N$ in a system of N particles. From $\mathcal{E}_N$, one can estimate the energy per interaction $\mathcal{E}_{N,I}$, knowing that the total number of interactions over the whole domain is $N(N-1)/2$. It follows that $\mathcal{E}_{N,I} = \frac{2}{N(N-1)} \mathcal{E}_N = \frac{2}{N-1} \epsilon_N$, and, assuming for simplicity $N \gg 1$, one can use $\mathcal{E}_{N,I} = \frac{2}{N} \epsilon_N$. At this point, notice that interactions across the interface region of $\Omega_1 \cup \Omega_2$ will occur only for $|\mathbf{r} - \mathbf{r}'| \leqslant R_0$, with $\mathbf{r} \in \Omega_1$ and $\mathbf{r}' \in \Omega_2$. This observation implies that the maximum distance of interaction along the direction perpendicular to the surface that ideally represents Ω as $\Omega_1 \cup \Omega_2$ is $|x - x'| = R_0$. In turn, the volume of Ω which contains the particles characterized by the actual cross-interactions between the subsystems is $\Delta = L^2 R_0$.

The number of particles in this volume is $N_\Delta = L^2 R_0 \rho = N \frac{R_0}{L}$. Given that we have uniform density, the number of cross interactions is $\frac{N_\Delta}{2} \frac{N_\Delta}{2}$, that is, half of the interaction sites are in Ω_1 and the corresponding other half of interaction sites is in Ω_2 (that is, interactions among sites in the same region are not counted). It follows that the total potential energy of cross interactions is $\mathcal{E}_{\text{cross}} = \frac{N_\Delta^2}{4} \mathcal{E}_{N,I} = \frac{N_\Delta^2}{2N} \epsilon_N$. Therefore, the maximum of the quality parameter can be estimated by

$$q_{max} \approx \frac{|\mathcal{E}_{\text{cross}}|}{|\mathcal{E}_N|} = \frac{\left| \frac{N_\Delta^2}{2N} \epsilon_N \right|}{N|\epsilon_N|} = \frac{N_\Delta^2}{2N^2},$$

from which one can infer the scaling law

$$q_{max} \approx \frac{1}{2} \left(\frac{R_0}{L} \right)^2. \qquad (16)$$

The relation of Eq. (16) provides a quick estimate of the quality of the thermodynamic accuracy that the system would have at the chosen simulation size. For example, taking data from Ref. [15] one has $R_0 \approx 2.5 r_s$ for all $r_s \in [0.8, 10]$ (in atomic units). Since $\frac{N}{L^3} = \frac{3}{4\pi r_s^3}$, one finds $N = \frac{3}{4\pi} (\frac{L}{r_s})^3$. For $L = 10 r_s$ it follows that $q_{max} \approx 0.03$ which corresponds to $N \approx 240$, that is, 240 electrons are sufficient to have an overall

TABLE I. Thermodynamic accuracy according to q_{max} as a function of the size of the system in terms of number of particles.

N	q_{max}
10	$\approx 26\%$
50	$\approx 9\%$
75	$\approx 7\%$
100	$\approx 6\%$
250	$\approx 3\%$
500	$\approx 2\%$
1000	$\approx 1\%$

accuracy of about 3.0%, that can be considered a sufficiently high accuracy. It should be noted that in Eq. (16) the particle density of the system is encoded in R_0 since one has $R_0 = R_0(r_s)$; however the fact that $R_0 \approx 2.5 r_s$, for the whole range of r_s that we treat here, makes the scaling law universal, in that it does not depend on r_s.

Table I reports the accuracy, according to the criterion of q_{max}, for different sizes of the system. The trend of convergence of the system to a sufficient bulk size based on q_{max} agrees well with the trend, determined from electronic structure calculations, of convergence of the total energy as a function of the system size [18]. Moreover, in a very detailed Monte Carlo study of the finite-size effects for one-component plasma [19,20], it is reported that, in a first approximation, the scaling law for the excess internal energy per particle follows an $N^{-2/3}$ trend in the low-coupling regime. This is the same trend that was found here, as $q_{max} \approx \frac{\alpha}{L^2}$ and $L \approx N^{1/3}$, with α a certain constant. However, in the same study it has been found that the scaling law could be even of an N^{-1} form or with some sizable corrections to the $N^{-2/3}$ form; a characteristic that could not be predicted by our current estimate. This example makes clear the role that our results play; that is, our approach is certainly valuable for giving a reliable magnitude order of finite-size effects without the need of a numerical simulation, however specific and precise details of the size dependence (e.g., correlation lengths) must be provided by accurate simulations as those reported above. Nevertheless, it can be said further that, in principle, the criterion based on q_{max} goes even beyond the criteria based on the convergence of a structure [e.g., $g(r)$] or convergence of a total energy [18], it actually involves the thermodynamics of the system. In fact, it expresses the thermodynamic concept of the response of the system to a perturbation, in this case: the building of an interface. A large interface energy, compared with a global quantity of reference, implies that the bulk characteristics (at the given size) is not sufficient for distributing the perturbation and yet keeping the same physical characteristics of the overall system as before. As a consequence, one would not have the bulk response as it would be in a sufficiently large system. For example, in the simulation of magnetic surfactant systems, the size of the system must be such that one can investigate whether surface-active properties are independent of the (bulk) paramagnetic response of the fluid [21]. As a consequence, the definition of a "*sufficiently well-described bulk*" involves the cost of building an interface as expressed by the criterion of q_{max}.

V. CONCLUSIONS

We have proposed a procedure to estimate the relevance of the finite size of a many-particle system in providing a statistically or thermodynamically satisfactory description of the bulk environment which characterizes a substance. Such a procedure is based on a rigorous theorem, and we have shown that for two-body potentials it can be written in terms of simple two-particle integrals that require only the knowledge of the particle density and the radial distribution function. The theorem provides the upper and lower bound to the free-energy cost to construct an interface that divides a system in independent subsystems. The criterion proposed here uses this theorem for the determination of the largest free-energy cost and relates it to a characteristic energy of the system. If such a free-energy cost is large, then it implies that the statistical description of the bulk environment is not sufficient. At an intuitive physical level the criterion is similar to a surface or volume energy ratio which often has been empirically used to define proper statistical or thermodynamic subsystems (e.g., in the grand canonical case, see Refs. [22–24] and references therein). In the current case there is no empiricism, the theorem is rigorous, and the corresponding estimate provides valid information about the sought-for size of the system. The information provided by this procedure is complementary to the other approaches that account for the finite-size effects by extrapolating from small systems. In fact, the analysis that we provide carries a direct, particle-based, thermodynamic estimate of the validity of the statistical mechanics description of the bulk, at the given system size. As a showcase, we have presented the treatment of the uniform interacting electron gas in the ground state. The procedure of application to this system is technically rather simple and allows us to use analytic formulas, however the corresponding numerical procedure of integration on a three-dimensional grid can be straightforwardly extended to other systems, provided that the particle density and the radial distribution functions of the system are known. Beyond the example of the uniform electron gas at $T > 0$, as discussed in the text, one may apply the idea to classical systems as for example liquid water, at a given average density and temperature, modeled with classical molecular models [25], to semiclassical systems such as parahydrogen at low temperature [26–28], to quantum systems such as ultracold atomic Fermi gases [29], to name a few. It must be underlined that the application of the method to the systems cited above may imply the need of additional physical assumptions and implies an increase of technical complexity, arising from the numerical integration on a grid. For example, in the case of the most complex system cited above (in the sense of molecular structure), that is liquid water, a particular care is needed in the analytic derivation of each integral and on its calculation on a grid. As discussed before, for interaction potentials that present a singularity as the interparticle distance goes to zero, the need of a short-distance cut-off arises. Its optimal choice would require well founded assumptions regarding the specific physics of the system. Once the integral is defined then one must deal with the computational complexity of the calculation; in fact the choice of an optimal grid size for the numerical integration will require a careful study about the convergence of each integral as a function

of the grid size. Such a study is mandatory to avoid that numerical inaccuracy leads to misleading physical results. In conclusion, the use of the proposed method, keeping in mind the warnings above, would be rather useful in the estimate of the optimal size of a simulation box. Specifically, it would provide a thermodynamic information complementary to the information based on other approaches for the estimate of finite-size effects.

APPENDIX A: TWO-PARTICLE DISTRIBUTION FUNCTION FROM THE DENSITY MATRIX

In terms of a density matrix $\hat{\rho}$, the average of a physical quantity A is written as $\mathbf{E}_{\hat{\rho}}[A] := \hat{\rho}A = \sum_{i=1}^{\infty} a_i \langle \psi_i, A\psi_i \rangle$, where $(a_i)_{i \in \mathbb{N}} \subset [0, 1]$ is a sequence such that $\sum_{i=1}^{\infty} a_i = 1$, and $(\psi_i)_{i \in \mathbb{N}}$ is an orthonormal basis of the underlying Hilbert space. In the specific case of this paper, it holds $A = A(\mathbf{r}_k, \mathbf{r}_l)$, that is, A is a two-body potential. In such a case, the expectation value of A in the state ψ_i is given by

$$
\langle \psi_i, A\psi_i \rangle = \int_{\mathbb{R}^{3N}} A(\mathbf{r}_k, \mathbf{r}_l) \, |\psi_i(\mathbf{r}_1, \dots, \mathbf{r}_{k-1}, \mathbf{r}_k, \mathbf{r}_{k+1}, \dots, \mathbf{r}_{l-1}, \mathbf{r}_l, \mathbf{r}_{l+1}, \dots, \mathbf{r}_N)|^2 \, d\mathbf{r}_1 \cdots d\mathbf{r}_N
$$

$$
= \int_{\mathbb{R}^6} A(\mathbf{r}_k, \mathbf{r}_l) \Bigg[\int_{\mathbb{R}^{3(N-2)}} |\psi_i(\mathbf{r}_1, \dots, \mathbf{r}_{k-1}, \mathbf{r}_k, \mathbf{r}_{k+1}, \dots, \mathbf{r}_{l-1}, \mathbf{r}_l, \mathbf{r}_{l+1}, \dots, \mathbf{r}_N)|^2
$$

$$
\times \, d\mathbf{r}_1 \cdots d\mathbf{r}_{k-1} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{l-1} d\mathbf{r}_{l+1} \cdots d\mathbf{r}_N \Bigg] d\mathbf{r}_k d\mathbf{r}_l.
$$

Now for all $i \in \mathbb{N}$, we define

$$
g_i(\mathbf{r}_k, \mathbf{r}_l) := \int_{\mathbb{R}^{3(N-2)}} |\psi_i(\mathbf{r}_1, \dots, \mathbf{r}_{k-1}, \mathbf{r}_k, \mathbf{r}_{k+1}, \dots, \mathbf{r}_{l-1}, \mathbf{r}_l, \mathbf{r}_{l+1}, \dots, \mathbf{r}_N)|^2
$$

$$
\times \, d\mathbf{r}_1 \cdots d\mathbf{r}_{k-1} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{l-1} d\mathbf{r}_{l+1} \cdots d\mathbf{r}_N
$$

as the two-body distribution of the state i. Thus, one obtains for the average of A:

$$
\mathbf{E}_{\hat{\rho}}[A] = \sum_{i=1}^{\infty} a_i \langle \psi_i, A\psi_i \rangle = \int_{\mathbb{R}^6} A(\mathbf{r}_k, \mathbf{r}_l) \left(\sum_{i=1}^{\infty} a_i g_i(\mathbf{r}_k, \mathbf{r}_l) \right) d\mathbf{r}_k d\mathbf{r}_l.
$$

It follows that the two-body distribution in a mixed state (i.e., in a superposition of different states represented by the density matrix) is nothing but $g(\mathbf{r}_k, \mathbf{r}_l) = \sum_i a_i g_i(\mathbf{r}_k, \mathbf{r}_l)$, i.e., the (non-normalized) $g(\mathbf{r}_i, \mathbf{r}_j)$ as previously defined. For the gas of interacting electrons in the ground state, the only term required (by the specific choice of the problem) is the ground state, thus

$$
g(\mathbf{r}_k, \mathbf{r}_l) = \int_{\mathbb{R}^{3(N-2)}} |\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)|^2 \, d\mathbf{r}_1 \cdots d\mathbf{r}_{k-1} d\mathbf{r}_{k+1} \cdots d\mathbf{r}_{l-1} d\mathbf{r}_{l+1} \cdots d\mathbf{r}_N.
$$

APPENDIX B: DETAILS ABOUT THE EXPLICIT CALCULATION OF $\mathbf{E}_f[U]$

For the total electron-electron interactions one has

$$
\mathbf{E}_f[V_{ee}] = \frac{1}{2} \rho^2 \int_\Omega \left\{ \int_\Omega \left[\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right] g(|\mathbf{r} - \mathbf{r}'|) \, d\mathbf{r}' \right\} d\mathbf{r}. \tag{B1}
$$

The positively charged background is smeared uniformly over the whole simulation box with the same density as the electrons and interacts electrostatically with itself,

$$
V_b = \frac{1}{2} \int_\Omega \int_\Omega \frac{n(\mathbf{R})n(\mathbf{R}')}{|\mathbf{R} - \mathbf{R}'|} \, d\mathbf{R} d\mathbf{R}', \tag{B2}
$$

and with the electrons,

$$
V_{eb} = - \int_\Omega \int_\Omega \frac{\rho(\mathbf{r})n(\mathbf{R})}{|\mathbf{r} - \mathbf{R}|} \, d\mathbf{r} d\mathbf{R}. \tag{B3}
$$

Note that $n(\mathbf{R}) = \rho(\mathbf{r}) = \rho = $ constant, that is, the charge distribution of the positive charges is equivalent to the charge distribution of the negative charges. This implies that one can substitute $\mathbf{R}$ with $\mathbf{r}$ and $\mathbf{R}'$ with $\mathbf{r}'$ in Eq. (B2) and obtain

$$
V_b = \frac{1}{2} \rho^2 \int_\Omega \int_\Omega \frac{1}{|\mathbf{r} - \mathbf{r}'|} \, d\mathbf{r} d\mathbf{r}', \tag{B4}
$$

at the same time in Eq. (B3), one can substitute $\mathbf{R}$ with $\mathbf{r}'$ and obtain

$$V_{eb} = -\rho^2 \int_\Omega \int_\Omega \frac{1}{|\mathbf{r} - \mathbf{r}'|} \, d\mathbf{r} d\mathbf{r}'. \tag{B5}$$

$$\mathbf{E}_f[V_{ee} + V_{eb} + V_b] = \frac{1}{2}\rho^2 \int_{\Omega_1} \left\{ \int_{\Omega_2} \left[\frac{g(|\mathbf{r} - \mathbf{r}'|) - 1}{|\mathbf{r} - \mathbf{r}'|} \right] d\mathbf{r}' \right\} d\mathbf{r} + \frac{1}{2}\rho^2 \int_{\Omega_2} \left\{ \int_{\Omega_1} \left[\frac{g(|\mathbf{r} - \mathbf{r}'|) - 1}{|\mathbf{r} - \mathbf{r}'|} \right] d\mathbf{r}' \right\} d\mathbf{r}$$

$$+ \frac{1}{2}\rho^2 \int_{\Omega_1} \left\{ \int_{\Omega_1} \left[\frac{g(|\mathbf{r} - \mathbf{r}'|) - 1}{|\mathbf{r} - \mathbf{r}'|} \right] d\mathbf{r}' \right\} d\mathbf{r} + \frac{1}{2}\rho^2 \int_{\Omega_2} \left\{ \int_{\Omega_2} \left[\frac{g(|\mathbf{r} - \mathbf{r}'|) - 1}{|\mathbf{r} - \mathbf{r}'|} \right] d\mathbf{r}' \right\} d\mathbf{r}. \tag{B6}$$

It follows that

$$\mathbf{E}_f[U_{ee}] = \frac{1}{2}\rho^2 \int_{\Omega_1} \left\{ \int_{\Omega_2} \left[\frac{g(|\mathbf{r} - \mathbf{r}'|) - 1}{|\mathbf{r} - \mathbf{r}'|} \right] d\mathbf{r}' \right\} d\mathbf{r} + \frac{1}{2}\rho^2 \int_{\Omega_2} \left\{ \int_{\Omega_1} \left[\frac{g(|\mathbf{r} - \mathbf{r}'|) - 1}{|\mathbf{r} - \mathbf{r}'|} \right] d\mathbf{r}' \right\} d\mathbf{r}. \tag{B7}$$

Note that $\mathbf{r}$ and $\mathbf{r}'$ are interchangeable, thus

$$\mathbf{E}_f[U_{ee}] = \rho^2 \int_{\Omega_1} \int_{\Omega_2} \left[\frac{g(|\mathbf{r} - \mathbf{r}'|) - 1}{|\mathbf{r} - \mathbf{r}'|} \right] d\mathbf{r}' d\mathbf{r}. \tag{B8}$$

$\mathbf{E}_{f_1,f_2}[U_{ee}] = 0$ can be easily derived from the results above, however the simple physical argument is that the interaction across the surface of separation is nothing else than the interaction between two neutral domains.

[1] T. Morishita and A. M. Ito, Traveling without dwelling: Extending the timescale accessible to molecular dynamics simulation, Phys. Rev. Res. **1**, 033032 (2019).

[2] J. J. Harris, G. A. Pantelopulos, and J. E. Straub, Finite-size effects and optimal system sizes in simulations of surfactant micelle self-assembly, J. Phys. Chem. B **125**, 5068 (2021).

[3] T. Dornheim, S. Groth, and M. Bonitz, The uniform electron gas at warm dense matter conditions, Phys. Rep. **744**, 1 (2018).

[4] M. Holzmann, R. C. Clay III, M. A. Morales, N. M. Tubman, D. M. Ceperley, and C. Pierleoni, Theory of finite size effects for electronic quantum Monte Carlo calculations of liquids and solids, Phys. Rev. B **94**, 035126 (2016).

[5] T. Dornheim and J. Vorberger, Overcoming finite-size effects in electronic structure simulations at extreme conditions, J. Chem. Phys. **154**, 144103 (2021).

[6] M. Heidari, K. Kremer, R. Potestio, and R. Cortes-Huerto, Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: Revisiting the spatial block analysis method, Entropy **20**, 222 (2018).

[7] F. Höfling and S. Dietrich, Finite-size corrections for the static structure factor of a liquid slab with open boundaries, J. Chem. Phys. **153**, 054119 (2020).

[8] L. Delle Site, G. Ciccotti, and C. Hartmann, Partitioning a macroscopic system into independent subsystems, J. Stat. Mech.: Theory Exp. (2017) 083201.

[9] B. Reible, C. Hartmann, and L. Delle Site, Two-sided Bogoliubov inequality for quantum systems to estimate finite size effects in quantum molecular simulations, Lett. Math. Phys. **112**, 97 (2022).

[10] M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford University Press, New York, 2010).

[11] E. Lieb and J. Lebowitz, Existence of Thermodynamics for Real Matter with Coulomb Forces, Phys. Rev. Lett. **22**, 631 (1969).

[12] E. H. Lieb and J. L. Lebowitz, The constitution of matter: Existence of thermodynamics for systems composed of electrons and nuclei, Adv. Math. **9**, 316 (1972).

[13] D. Pines and P. Nozieres, *Theory of Quantum Liquids* (Benjamin, New York, 1966).

[14] G. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, 2005).

[15] P. Gori-Giorgi, F. Sacchetti, and G. B. Bachelet, Analytic static structure factors and pair-correlation functions for the unpolarized homogeneous electron gas, Phys. Rev. B **61**, 7353 (2000).

[16] T. Chachiyo, Simple and accurate uniform electron gas correlation energy for the full range of densities, J. Chem. Phys. **145**, 021101 (2016).

[17] T. Sjostrom and J. Dufty, Uniform electron gas at finite temperatures, Phys. Rev. B **88**, 115123 (2013).

[18] O. Ciftja, Energy of a finite three-dimensional electron gas of spinless electrons, J. Phys. Chem. Solids **136**, 109135 (2020).

[19] J. M. Caillol, Thermodynamic limit of the excess internal energy of the fluid phase of a one-component plasma: A Monte Carlo study, J. Chem. Phys. **111**, 6538 (1999).

[20] J. M. Caillol and D. Gilles, An accurate equation of state for the one-component plasma in the low coupling regime, J. Phys. A: Math. Theor. **43**, 105501 (2010).

[21] D. M. Reed, A. Fortenberry, E. A. Wolfe, R. A. Stanhope, C. I. Daniel, C. M. Hern, A. E. Smith, and P. Scovazzo, Interfacial vs bulk phenomena effects on the surface tensions of aqueous magnetic surfactants in uniform magnetic fields, Langmuir **36**, 10074 (2020).

[22] A. Agarwal, J. Zhu, C. Hartmann, H. Wang, and L. Delle Site, Molecular dynamics in a grand ensemble: Bergmann–Lebowitz model and adaptive resolution simulation, New J. Phys. **17**, 083042 (2015).

[23] A. Agarwal and L. Delle Site, Path integral molecular dynamics within the grand canonical-like adaptive resolution technique: Simulation of liquid water, J. Chem. Phys. **143**, 094102 (2015).

[24] L. Delle Site and M. Praprotnik, Molecular systems with open boundaries: Theory and simulation, Phys. Rep. **693**, 1 (2017).

[25] S. Chatterjee, P. G. Debenedetti, F. H. Stillinger, and R. M. Lynden-Bell, A computational investigation of

thermodynamics, structure, dynamics and solvation behavior in modified water models, J. Chem. Phys. **128**, 124511 (2008).

[26] I. F. Silvera and V. V. Goldman, The isotropic intermolecular potential for H_2 and D_2 in the solid and gas phases, J. Chem. Phys. **69**, 4209 (1978).

[27] A. B. Poma and L. Delle Site, Adaptive resolution simulation of liquid para-hydrogen: Testing the robustness of the quantum-classical adaptive coupling, Phys. Chem. Chem. Phys. **13**, 10510 (2011).

[28] R. Potestio and L. Delle Site, Quantum locality and equilibrium properties in low-temperature parahydrogen: A multiscale simulation study, J. Chem. Phys. **136**, 054101 (2012).

[29] S. Giorgini, L. P. Pitaevskii, and S. Stringari, Theory of ultracold atomic Fermi gases, Rev. Mod. Phys. **80**, 1215 (2008).