Molecular Dynamics of Liquid Argon

Daniel Felipe Forero Sánchez*
(Dated: December 11, 2019)

I. INTRODUCTION

Molecular dynamics simulations have been fundamental tools in science since they allow for the detailed analysis of various properties of the system that are impossible to probe experimentally. However, to properly simulate a physical system is a complex problem on its own, owing to the size of such systems. Molecular dynamics treats the system as a collection of particles evolving according to the Lennard-Jones potential [4]; it can be then regarded as a simulation from "first principles". To track the position and velocities of 10^{23} particles is, in practice, impossible. In such a limit one often uses emergent models such as fluid dynamics. Nonetheless, it is possible to properly model a small physical system with MD and make it an accurate reproduction of the real world [8]

II. THEORY

A. The Lennard-Jones potential

The Lennard-Jones potential [4] is commonly used to define the weak interatomic interactions in fluids. It is defined as

$$U(r_{ij}) = 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \tag{1}$$

where $r_{ij} \equiv \|\mathbf{r}_i - \mathbf{r}_j\|$ is the distance between two particles at positions $\mathbf{r}_{i,j}$. It depends on the constants ϵ and σ, which define the energy and length scale of the system, respectively.

B. The Radial distribution function

The *Radial distribution function* (RDF) is defined as

$$g(r) = \frac{\rho(r)}{\rho_0}, \tag{2}$$

where $\rho_0 = N/V$ is the constant average density (total number of particles over total volume of the system) and $\rho(r)$ is the density of particles at a distance r. The function $g(r)$ is interpreted as the probability of finding a pair of particles separated by a given distance r. The RDF is also known as the pair-correlation function. These correlators are ubiquitous in physics and the RDF has analogous quantities in other fields such as cosmology (see cosmological two-point correlation function $\xi(s)$[7]). This is not surprising since they encode relevant statistical information about the system (see section II C).

C. The Structure factor

The *Structure factor* $S(k)$ is a fundamental quantity in the study of matter, since it can be experimentally measured through scattering experiments[9]. More importantly, it encodes relevant physical information given that it is closely related to physical quantities such as the scattering potential.
Furthermore, $S(k)$ is directly related to $g(r)$ through a Fourier transform:

$$S(k) = 1 + 4\pi\rho_0 \int_{\mathbb{R}^+} r^2[g(r) - 1]\frac{\sin kr}{kr}dr. \tag{3}$$

We expect then, that the peaks in $S(k)$ are located at $k \sim 2\pi/s$ with s the typical separation of particles in the system.
Following the analogy with cosmology, the structure factor is analogous to the cosmological power spectrum $P(k)$ measured, for example, in the CMB[7]. It should be noted that the properties $g(r)$ and $S(k)$ are static in the sense that they are independent of time.

D. The Diffusion coefficient

The diffusion equation

$$\dot{C}(\mathbf{r}, t) = D\nabla^2 C(\mathbf{r}, t), \tag{4}$$

describes the evolution of the concentration, C, of a substance. The parameter D is the *diffusion coefficient* and defines how fast the substance will diffuse in the media. The kernel of equation 4 is a Gaussian distribution with $\sigma_r^2 = 6Dt$ and $\langle r\rangle = 0$, so to compute the diffusion coefficient we compute the second moment of the distribution $C(\mathbf{r})$, $\langle r^2\rangle$. This yields the Einstein relation[1]

$$\frac{\partial}{\partial t}\langle r^2\rangle = 6D. \tag{5}$$

It is then possible to measure the diffusion coefficient from the *mean square displacement* (MSD) via the rela-

* Section de Physique, LASTRO, EPFL

tions 6, 7.

$$D = \lim_{t\to\infty} \frac{1}{6t} \langle \|\Delta \mathbf{r}(t)\|^2 \rangle \tag{6}$$

$$\langle \|\Delta \mathbf{r}(t)\|^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{r}_i(t) - \mathbf{r}_i(0)\|^2 \tag{7}$$

We can alternatively use the *velocity autocorrelation function* (VACF)[5]

$$D = -\frac{1}{3} \int_{\mathbb{R}^+} \mathrm{d}\tau \, \langle \mathbf{v}(\tau) \cdot \mathbf{v}(0) \rangle, \tag{8}$$

$$\langle \mathbf{v}(\tau) \cdot \mathbf{v}(0) \rangle = \sum_{i=1}^{N} \mathbf{v}(\tau) \cdot \mathbf{v}(0) \tag{9}$$

III. METHODS

A. Grid Integration

This kind of integration algorithms rely on space discretization of space and time coordinates, that is, they work on a grid. While there are a great variety of similar integration methods, each one having advantages and disadvantages, we focus on the one used by the `md1` code [2]. This particular implementation in `FORTRAN90`, uses the *velocity Verlet* integration algorithm [10]. This second order method has the following update rules:

$$r_{n+1} = r_n + h v_n + h^2 \frac{f(t_n)}{2m}, \tag{10}$$

$$v_{n+1} = v_n + h \frac{f(t_{n+1}) + f(t_n)}{2m}, \tag{11}$$

where r_n and v_n define, respectively, the position and velocity at time $t_n \equiv nh$, h is the time step, f is the force $m\ddot{r}_n = f(t_n)$ and m, the mass.
This algorithm was used to integrate the equations of motion of N particles in the simulation.

B. Monte-Carlo

Monte Carlo (MC) simulations are based on the idea that one can numerically approximate any integral by randomly sampling the integrand and computing the average of the values obtained. This is,

$$\int_0^L f(x)\mathrm{d}x = L \left[\frac{1}{L} \int_0^L f(x)\mathrm{d}x \right] = L \langle f \rangle_{\text{uniform}}. \tag{12}$$

In a more general manner, we could write the above for any probability distribution $\omega(\mathbf{x})$ and any dimension D such that $\mathbf{x} \in \mathbb{R}^D$:

$$I = \int f(\mathbf{x})\omega(\mathbf{x})\mathrm{d}\mathbf{x} = \langle f \rangle_\omega, \tag{13}$$

where $\omega(\mathbf{x}) = \frac{\exp(-\beta H(\mathbf{x}))}{\mathcal{Z}(\beta)}$, where $\mathcal{Z}(\beta)$ is the partition function defined in the usual manner, $H(\mathbf{x})$ is some function that defines the probability density function (PDF) $\omega(\mathbf{x})$, along with the parameter β. The integral, I, can then be approximated by

$$I \approx \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i), \tag{14}$$

where the $\mathbf{x}_i$ are random numbers (vectors) that follow the given PDF (or even a similar one).
Physically speaking, the PDF to be sampled is the one that dictates how probable a state $\mathbf{x}$ is, based on its energy H at a temperature β^{-1}. The Monte Carlo simulation will then randomly probe the phase space for the system and allow the estimation of important physical quantities, which are typically, averages (such as $E = \langle H \rangle$, the average energy of the system.).

1. *Blocking Analysis*

To generate the random numbers $\mathbf{x}_i \sim \omega(\mathbf{x})$ it is usually convenient to use correlated sampling. The Metropolis-Hastings [3, 6] algorithm allows for obtaining random numbers that follow any desired PDF, with the caveat that the resulting chain of numbers will be correlated. This is an issue if one wants to estimate the error in the computation of the integrals, since the usual formula for the error of a mean of N numbers x,

$$\sigma_{\langle x \rangle} = \frac{\sigma_x}{\sqrt{N}}, \tag{15}$$

is only valid for an independent chain of numbers.
Blocking analysis is a tool to estimate the uncertainty in the integral computation for a correlated chain of numbers. To do this, we transform the initial chain of N numbers into a chain of $N/2$ numbers via averaging pairwise, This operation is commonly called an average-pool. Such a transformation is done recursively M times, such that at the M-th iteration, the length of the chain is $N(M) = N/2^M$. We can compute the standard deviation at each iteration as usual

$$\sigma_x^2(M) = \frac{1}{N(M)} \sum_{i=1}^{N(M)} (x_{M,i} - \langle x \rangle)^2, \tag{16}$$

where $x_{M,i}$ denotes the i-th value in the M-th chain.
Now, the error in the average evolves with M as

$$\sigma_{\langle x \rangle}(M) = \frac{\sigma_x(M)}{\sqrt{N(M)}}. \tag{17}$$

When $\sigma_{\langle x \rangle}(M)$ reaches a plateau, we see that it is no longer dependent of M and is indeed the correct estimation of the error in the average.

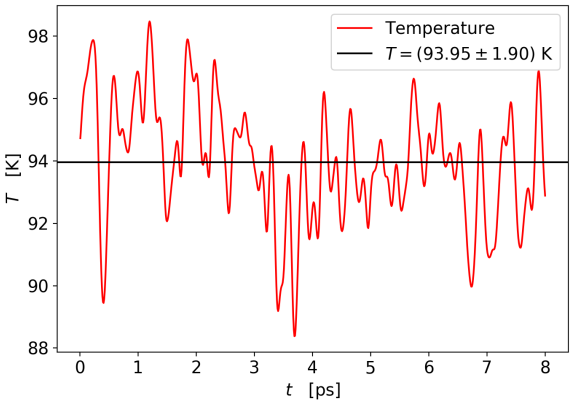

Figure 1. Temperature of the stabilized system in the Grid integration (GI) simulation.

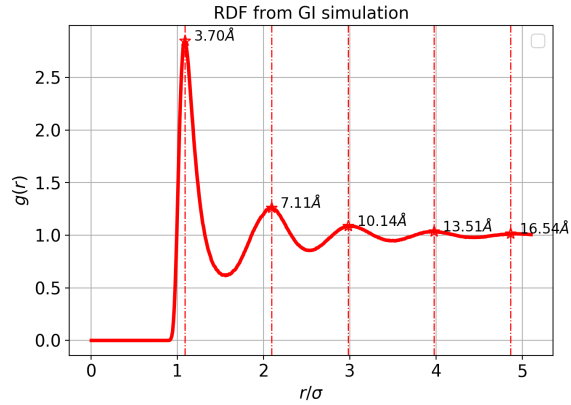

Figure 2. Radial distribution function for liquid Argon at $T \approx 94.4\,\mathrm{K}$ from the GI simulation. The labels on the peaks show the separations r at which it is more likely to find a pair of atoms.

IV. THE SIMULATIONS

A. Liquid Argon

Rahman[8] studied the properties of a liquid Argon system consisting of $N = 864$ atoms in a box of length $L = 34.8\,\text{Å} = 10.229\sigma$. The system was at temperature $T = 94.4\,\mathrm{K}$ and had a density $\rho_0 = 1.374\,\mathrm{g\,cm^{-3}}$. In this simulation, Argon atoms of mass $M = 6.69 \times 10^{-26}\,\mathrm{kg}$ interact through the potential shown in equation 1 with parameters $\epsilon/k_b = 120\,\mathrm{K}$ and $\sigma = 3.4\,\text{Å}$. Additionally, the interaction range was cut-off at $R = 2.25\sigma$, while periodic boundary conditions in all directions were assumed. Finally, the time step is defined $h = 1 \times 10^{-2}\,\mathrm{ps}$. For the Monte Carlo simulation the cut-off radius was changed to $R = 5\sigma$. The physical parameters were kept constant.

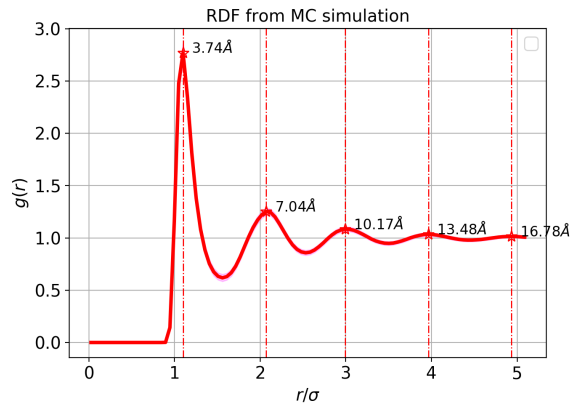

Figure 3. Radial distribution function for liquid Argon at $T \approx 94.4\,\mathrm{K}$ from the MC simulation. The labels on the peaks show the separations r at which it is more likely to find a pair of atoms.

V. RESULTS AND DISCUSSION

As a check of the simulation temperature, we study the temperature of the system, shown in figure 1. The average temperature of this simulation is then $T = (93.95 \pm 1.90)\,\mathrm{K}$.

A. Radial distribution function

The RDF was directly sampled by the `md1` code and the results are shown in figure 2. The peaks of the curve were extracted and are also shown (in physical units). The pattern is typical of a liquid, which has no definite crystal structure[9]. The results are comparable to those shown in Rahman [8], where the first 3 peaks were extracted. The relative errors for these, between the two measurements are 0.10%, 1.58%, and 2.59% respectively. Additionally, this simulation allows for the computation of two extra peaks as reported in figure 2.

The Monte Carlo simulation is also able to directly sample the RDF at run-time. The number of equilibration and production cycles was set to 100 000 and the number of steps between property sampling to 5. Figure 3 shows the resulting curve. Notice that the curve is indeed very similar to figure 2. The peaks are also shown. These are compared with Ref. [8], resulting in percent errors of 1.04%, 0.61% and 2.23% respectively for the first 3 (available) peaks. Regarding the differences between the grid integration and the MC approaches, the RDFs show sub 1% differences for the first 5 peaks shown, while the last one presents a 1.5% percent error. This shows the equivalence of both techniques for the calculation of $g(r)$.

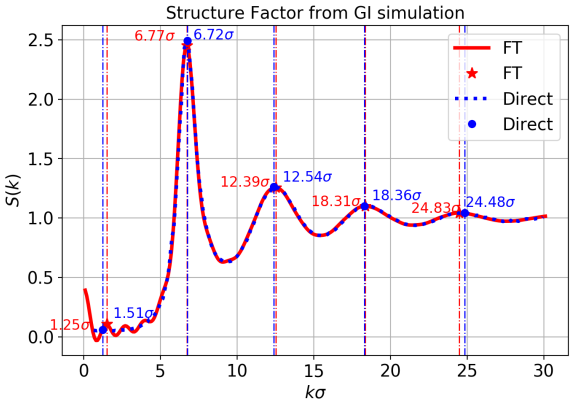

Figure 4. Structure factor of liquid Argon at $T \approx 94.4\,\mathrm{K}$. The values of the peak positions are shown. The values left of the peaks correspond to the FT of $g(r)$ (solid curve), while the values to the right are the peak positions for the directly-sampled curve (dotted line). The first peak is a numerical artifact, we ignore it in this analysis.

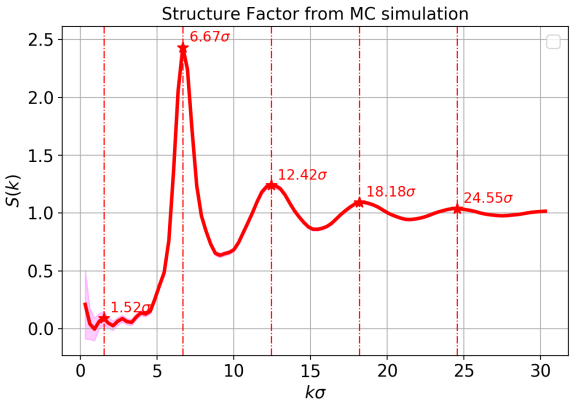

Figure 5. Structure factor of liquid Argon at $T \approx 94.4\,\mathrm{K}$. The values of the peak positions are shown. The shaded region corresponds to the standard deviation of the tries used while the solid line is the mean. The first peak is a numerical artifact, we ignore it in this analysis.

B. Structure factor

The structure factor can either be directly sampled from the velocity distribution or computed as the Fourier transform (FT) of $g(r)$ as shown in equation 3. Both these approaches were taken and results are shown in figure 4. The peak positions are extracted for both cases and are shown in the same figure. The first peak shown has no physical meaning and is more prominent on the FT, given the inaccuracies that arise from this computation. We then focus on the subsequent maxima. The values at the left of each peak in figure 4 show the peak positions for the FT computation, while the values on the right show it for the direct sampling case.
The values reported in Rahman [8] are shown in table I

$(k\sigma)_{\mathrm{R}}$	% $e_{\mathrm{R-FT}}$	% $e_{\mathrm{R-D}}$	% $e_{\mathrm{R-MC}}$	% $e_{\mathrm{D-MC}}$	% $e_{\mathrm{FT-MC}}$
6.80	1.15	0.40	2.00	1.59	0.84
12.50	0.33	0.88	0.61	0.27	0.95
18.50	0.76	1.03	1.75	0.71	0.99
24.80	1.30	0.13	1.04	1.17	0.26

Table I. Values reported for the simulation in reference [8] along with the relative percent errors , %e, among the different measurements. "R" represents Rahman [8], "FT" stands for Fourier transform, "D" for direct sampling and "MC" for MC direct sampling.

alongside the percent errors for all the approaches tested in the present work. Notice that the errors are relatively small, as in the case of $g(r)$. Furthermore, the average error among all peaks is lowest for the direct sampling case vs. Rahman [8] (0.61 %), while the highest is for the MC vs. the same reference (1.3 %). The last two columns of table I show the percent errors of the peaks extracted from $S(k)$ sampled from the MC simulation compared to the two techniques used in the GI one. Evidently, the errors are small and, in average, below 1%. It is then safe to state that the approaches are all equivalent, as expected.
It is worth noting that in figures 3 and 5, the shaded region corresponding to the "tries" or the different extractions of the functions at different steps is almost invisible. Indeed, in the case of $S(k)$, it is only visible in the region of $k\sigma$ close to 0, where the function is dominated by artifacts of the Fourier transform (peaks related to the inverse length of the simulation box). This shows this properties are in fact, static.

C. Diffusion Coefficient

The diffusion coefficient can be computed in two different ways as described in section II D. The first approach is to use de Einstein relation (eq. 6) and compute D from the slope of the MSD. The reported value is the mean of the fitted values to 19 different measurements with their respective standard deviation, $D_{\mathrm{MSD}} = (2.56 \pm 0.24) \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$. The MSD fit was done starting from $t = 0.3\,\mathrm{ps}$, from where we assumed the ballistic regime had passed, to a relation MSD(t) $= \langle \|\Delta \mathbf{r}(t)\|^2 \rangle = mt + c$. Notice that errors in the fit parameters m, c are negligible, thus not reported here in favor of the statistical error in D_{MSD}.
As a second approach, the diffusion coefficient is also calculated from the integral of the VACF. The integral is done numerically with a simple trapezoidal-rule algorithm. The reported value is again the mean of 19 repetitions alongside the respective standard deviation, $D_{\mathrm{VACF}} = (2.57 \pm 0.30) \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$. The VACF is shown in figure 7.
Evidently, both measured values lie within a single standard deviation from one another. To properly test the agreement between the values, we perform a (different

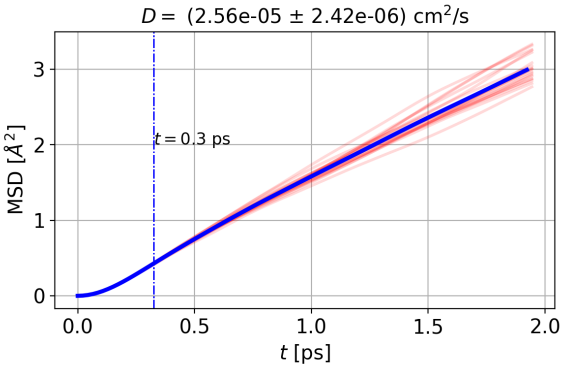

Figure 6. Mean Square Displacement (MSD) obtained from the simulations. The faint (red) lines correspond to each one of the 19 different simulations while the solid (blue) line shows the mean. The vertical line at $t = 0.3$ ps shows the point from which the curves were fitted.

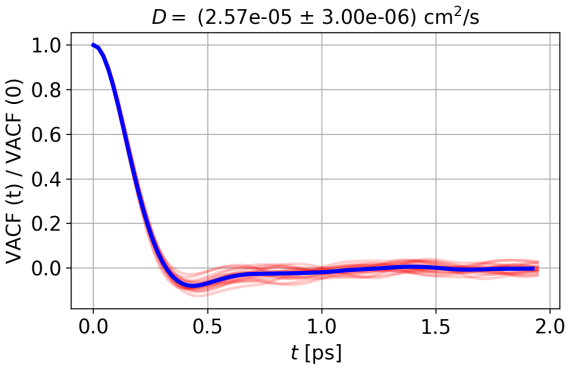

Figure 7. Normalized VACF plot for the GI simulation. Solid curve shows the mean value of all tries (the 19 simulatons) and the reported value is its integral. Shaded region (faint curves) show the different tries. The reported uncertainty in D_{VACF} is given by the standard deviation of the integrals of these curves.

variance) t-test with the null hypothesis that the mean values obtained from the slope of the MSD and the integral of the VACF are identical. Such a test yields a $p = 0.91 > 0.05$, so we conclude that we can't reject the null hypothesis hence the averages are identical.
We can also compare with the results in [8]. In this case, we get a 5.31% error in the VACF measurement while in the MSD measurement a value of 4.97% is obtained. In this case we notice that both our values lie below the result in Ref. [8].

D. Blocking Analysis

In addition to the sampling of $g(r)$ and $S(k)$ from the MC simulation, the energy E and pressure P were obtained from the simulation. During testing, we observed

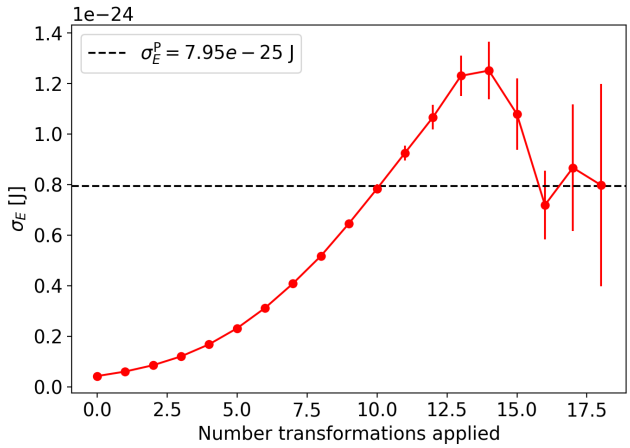

Figure 8. Results of blocking analysis for the error in the estimation of energy E of the system. Dashed lines show the estimated plateau value.

that after 7000 cycles, the energy was already stabilized, therefore, the next run was done with 1×10^6 cycles and the equilibration ones were skipped. With these results, 18 block-transformations were applied.
Results for the errors in the estimation of $E = \langle H \rangle$ is shown in figure 8. The analysis allows for the computation of the error in E from the plateau value of the curve. This value is taken to be the mean of the last three values in figure 8. The measured value of the energy is then $E = (-95\,743.47 \pm 7.95) \times 10^{-25}$ J.

Notice that up to pool 12, $\sigma_E(M)$ was dominated by the exponential factor 2^M coming from $N(M)$. However, later it decreases for the next few polls until it starts oscillating around the reported plateau value σ_E^{P}.

VI. CONCLUSIONS

We conclude that both simulations show equivalent results in terms of the physical properties of the system that were measured, this is, the static properties $g(r)$ and $S(k)$. Even though both simulations involve integration, the conceptual basis is different. While the GI simulation integrates the equation of motion in configuration space, the MC simulation explores the phase space and finds there the "physical" configurations, those with high canonical probability.
The features measured have been compared with the results reported in Rahman [8]. An excellent ($< 5\%$) agreement was found in all measurements, validating the present work's results and showing the equivalence between the approaches taken to analyze the system.
The GI simulation also allowed us to explore the diffusion coefficient and to compute it in two ways as described before. The values obtained $D_{\mathrm{VACF}} = (2.57 \pm 0.30) \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ and $D_{\mathrm{MSD}} = (2.56 \pm 0.24) \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ are in agreement as proven

by the t-test.

The error estimation in the MC case for the average energy E was successfully done with blocking analysis, from which a plateau value $\sigma_E^{\mathrm{P}} = 7.95 \times 10^{-25}$ J was extracted. We expect a longer simulation would result in a much clearer plateau, however the number of transformations M is very limited by the length of the initial chain N. A small increase in M implies an exponential increase in N, and thus, in the computing time needed.

[1] Einstein, A., Annalen der Physik **322**, 549 (1905).

[2] Ercolessi, F., "A molecular dynamics primer," Tech. Rep.

[3] Hastings, W. K., Biometrika **57**, 97 (1970).

[4] Jones, J. E., Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **106**, 463 (1924).

[5] Kubo, R., Journal of the Physical Society of Japan **12**, 570 (1957).

[6] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., Citation: J. Chem. Phys **21**, 1087 (1953).

[7] Peebles, P. J. E. P. J. E., *The large-scale structure of the universe* (Princeton University Press, 1980) p. 422.

[8] Rahman, A., Physical Review **136**, A405 (1964).

[9] Simon, S. H., Journal of Chemical Information and Modeling **53**, 1689 (2013), arXiv:arXiv:1011.1669v3.

[10] Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R., J. Chem. Phys **76**, 637 (1982).