

Estimation of Van der Waals parameters of a Lennard-Jones fluid via Molecular Dynamics Simulations

Alessandro Lovo, mat. 1236048

May 19, 2020

1 Introduction

The aim of this report is to use Molecular Dynamics (MD) simulations of a fluid subject to Lennard-Jones (LJ) potential to derive its equation of state. Then, by fitting it with the Van der Waals (VdW) equation of state an estimate of the VdW parameters will be obtained.

1.1 Lennard-Jones and Van der Waals model

A LJ fluid is a set of particle subject only to the pair-wise interaction potential in eq 1 where r is the distance between the centers of the two particles considered. The LJ parameters σ and ϵ , together with the mass of the particles m , provide typical scales of the system and are thus used to work in reduced units (here denoted with a $*$). For example if t is the unit of time, the corresponding reduced unit would be $t^* = t\sqrt{\epsilon/m\sigma^2}$.

$$U(r) = 4\epsilon \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right) \quad (1)$$

This potential models the short range attraction due to the interaction between temporary dipoles (London forces) and the hard Pauli repulsion between the electron clouds.

By some statistical mechanics mean field computations one finds that a fluid of N particles in a volume V at pressure P and temperature T interacting via the LJ potential follows the VdW equation of state (eq 2) where N_A is Avogadro's number and k_B the Boltzmann constant. By defining the volume per particle $v = V/N$ and going to reduced units one gets a nicer form of the equation of state.

$$\left(P + \frac{a}{N_A^2} \frac{N^2}{V^2} \right) \left(V - \frac{b}{N_A} N \right) = N k_B T \quad \rightarrow \quad \left(P^* + \frac{a^*}{v^{*2}} \right) (v^* - b^*) = T^* \quad (2)$$

The VdW parameters a, b are connected to the LJ parameters ϵ, σ through

$$\begin{cases} a = N_A^2 \frac{8\pi\epsilon\sigma^3}{3} \\ b = N_A \frac{2\pi\sigma^3}{3} \end{cases} \quad \rightarrow \quad \begin{cases} a^* = \frac{1}{N_A^2 \epsilon \sigma^3} a = \frac{8\pi}{3} \approx 8.38 \\ b^* = \frac{1}{N_A \sigma^3} b = \frac{2\pi}{3} \approx 2.09 \end{cases} \quad (3)$$

and the VdW equation of state displays a phase transition with a critical point at

$$T_c^* = \frac{8a^*}{27b^{*2}}, \quad P_c^* = \frac{a^*}{27b^{*2}}, \quad v_c^* = 3b^*$$

but the model itself is not able to correctly describe the region where the liquid and vapour phases coexist (fig 1). To describe that region one should use the Maxwell 'equal-area' rule to obtain a

substance	$\epsilon/k_B[\text{K}]$	$\sigma[\text{pm}]$	$m[a.u.]$	$a[\text{Jm}^3\text{mol}^{-2} \times 10^{-5}]$	$b[\text{m}^3\text{mol}^{-1}]$	$T_c[\text{K}]$
Ar	120	341	40	0.136	3.20	150.8
N ₂	95.2	370	28	0.137	3.87	126.2
O ₂	118	358	32	0.138	3.19	154.6
CO ₂	189	449	44	0.366	4.29	304.2

Table 1: LJ and VdW experimental parameters in real units [2] [3]

substance	a^*	b^*	T_c^*
Ar	5.69	1.34	1.26
N ₂	5.67	1.27	1.33
O ₂	5.10	1.15	1.31
CO ₂	4.27	0.79	1.61

Table 2: VdW experimental parameters in reduced units

constant pressure for each isotherm in the coexistence zone [1], and this complicates a lot the model. For the sake of this report of that region will be simply avoided.

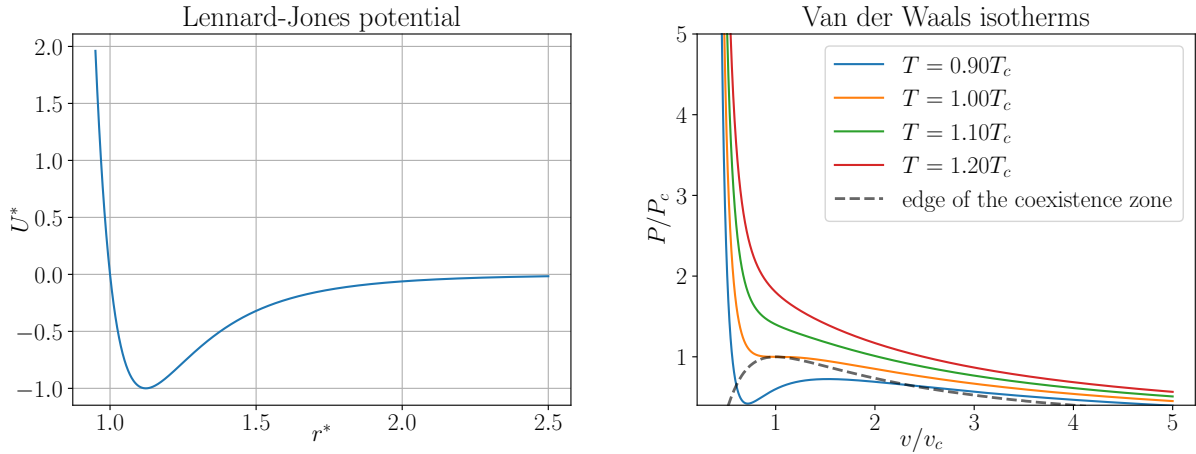


Figure 1: Lennard-Jones potential and Van der Waals isotherms with coexistence zone

1.2 Some experimental data

LJ and VdW models describe well systems where the only potential energy is given by London forces, i.e systems of small quasi-spherical non polar particles. Here are some of them (tabs 1, 2):

From this experimental parameters one can immediately see that the statistical mechanics predictions in eq 2 are not very accurate, and also, since the reduced parameters vary with the substance, there are other effects in the physical systems beyond the LJ - VdW approximation.

2 Choice of simulation parameters

All the following MD simulations are done using the programm LAMMPS [4] and the basic idea is to use it to obtain various triplets (v^*, P^*, T^*) out of the coexistence zone and then fit them with the VdW formula (eq 2).

Anyways, to obtain meaningful results one has to properly set the simulation parameters.

2.1 Simulation ensemble

Since the coexistence zone is avoided both NVT and NPT ensembles will in principle work, producing a unique value of respectively P,V vor every input value of respectively VT, PT. Anyways the NVT ensembles has some advantages. When fixing the volume of the system this has no error and won't fluctuate during the simultaion while on the other hand fixing the pressure in the NPT ensemble will imply the usage of a barostat that acts rescaling the volume of the system; thus both volume and pressure will fluctuate around the equilibrium value.

2.2 Size of the system and initial positions

When choosing the size of the simulation system one has to find a compromise between accuracy and computational cost. Looking at the literature for this kind of simulations $N \sim 10^3$ particles with periodic boundary conditions seems a good compromise [5]. For this specific case the initial configuration is an $11 \times 11 \times 11$ cubic lattice of particles with spacing $l^* = v^{*1/3}$ where v^* is the volume per particle set in the simulation. Also, looking at the experimental values of b^* , one can consider 1.4 as an a-priori minimum value for v^* .

2.3 Timestep

Since in this system there are no internal degrees of freedom (the particles are spheres interacting via LJ potential) a balanced choice for a timestep is around $dt = 10\text{fs}$ which means, if one considers Argon LJ parameters $dt^* \approx 0.0046$. Computing dt^* using other substances yields a value between 0.004 and 0.005. In the following $dt^* = 0.004$ will be used.

2.4 Thermostat typical time and sampling rate

When using Nosè-Hoover thermostat one has to set the desired temperature T_{imp} and the temperature damping time τ_T . Since, as will be shown, the thermodynamic properties of the system oscillate with a period proportional to τ_T , it is important to choose τ_T and the sampling frequency of thermodynamic properties together in order to avoid a biased results.

In the LAMMPS manual it is suggested to use $\tau_T \sim 100dt$ [4] so one can try some values around it and see what happens.

In fig 2 one can see the typical behavior of the temperature (other thermodynamical properties are very similar).

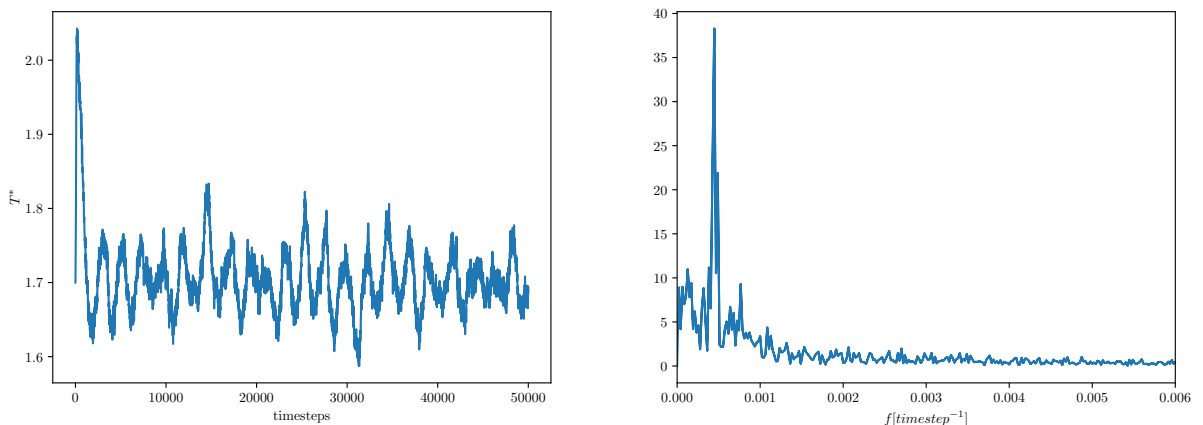


Figure 2: Trend of the temperature as the simulation proceeds (sampling every 10 timesteps) and its frequency spectrum. The peak corresponds to a period of 2238 timesteps. This simulation is run at $v^* = 5.8$, $T^* = 1.7$, $\tau_T = 500dt$.

A quite remarkable fact is that the period of the oscillation τ_{osc} does not depend either on v^* or T^* ,

but only on τ_T . If we plot τ_{osc} and the rms amplitude of the oscillation of the temperature δ_T as τ_T varies we obtain the results in fig 3. I.e. $\tau_{osc} \approx 4.67\tau_T$ while the amplitude of the oscillation does not vary a lot with τ_T .

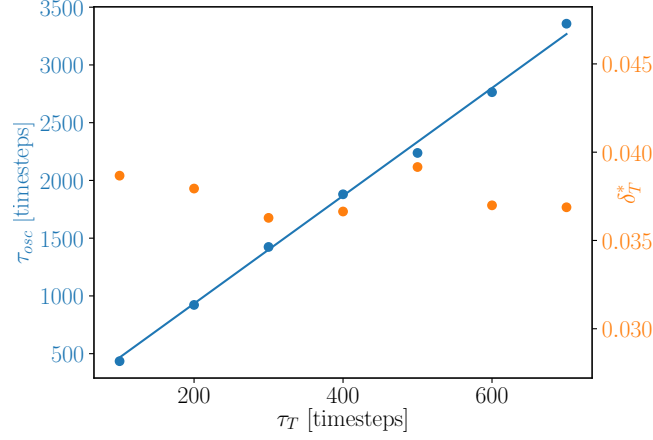


Figure 3: Behavior of τ_{osc} and δ_T as τ_T varies

In order to avoid aliasing effects the sampling rate should be at least $2/\tau_{osc}$, but anyways, since the behavior of the properties of the system is quite noisy (peaks at higher frequency in fig 2) it is better to have an even finer sampling and then make averages. For this reason a reasonable sampling rate could be $10/\tau_{osc}$. Since $\tau_{osc} \propto \tau_T$, in order to have a reasonable number of snapshots of the system without running the simulation for too many timesteps, it is better to choose a rather small value of τ_T . Henceforth $\tau_T = 200dt$ and consequently the system will be sampled every 100 timesteps ($\frac{1}{10}\tau_{osc} = 93.4dt$).

2.5 Block averaging and lenght of the simulation

The aim of each simulation is to obtain a triplet of values (T^*, v^*, P^*) with associated errors (apart for the volume that is fixed and hence has no errors). But since snapshots are correlated with each other one cannot do a simple average of all of them. Instead one should do *block averages*. This means that if the sampled data are a set of N values of a given property $\{f_i\}_{i=1}^N$ one creates a second dataset $\{f_j^{(M)}\}_{j=1}^{N^{(M)}}$ by grouping together M consecutive datapoints such as:

$$f_j^{(M)} = \frac{1}{M} \sum_{i=Mj}^{M(j+1)-1} f_i, \quad N^{(M)} = \frac{N}{M}, \quad \sigma(\overline{f^{(M)}}) = \sqrt{\frac{\text{var}(\{f_j^{(M)}\}_{j=1}^{N^{(M)}})}{N^{(M)}}} \quad (4)$$

Then by monitoring how the standard deviation of the mean $\sigma(\overline{f^{(M)}})$ behaves as a function of M one can choose the proper value of M (fig 4).

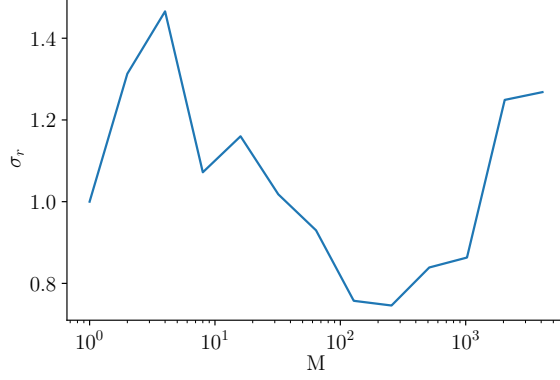


Figure 4: Relative standard deviation of the mean of the temperature $\sigma_r = \sigma(\overline{T^{(M)}})/\sigma(\overline{T^{(1)}})$ as a function of the block averaging parameter. With $M \sim 10$ the average blocks are temporally spaced by $\sim \tau_{osc}$. For $M > 10^3$ the standard deviation is computed on a too small set of values so it diverges. This data are extracted from a simulation lasting 10^6 timesteps.

From fig 4 one can see that the standard deviation of the mean temperature tends to stabilize around $M \sim 10^2$ (notice the logarithmic scale on the x axis), meaning that the average blocks are no more significantly correlated with each other and so the variance of the mean acquires the correct meaning. With this choice of M and considering that the system is sampled every 100 timesteps, at the end of the block averaging process one has a data point every 10^4 timesteps. So, in order to be able to compute a meaningful average and variance one needs at least 10-20 points, meaning that the simulation should proceed for at least $(1 - 2) \cdot 10^5$ timesteps ($2 \cdot 10^5$ will be used).

2.6 Equilibration of the system

The starting condition for the simulation are out of the thermodynamic equilibrium, which is reached in the first part of the simulation (equilibration). This portion of data should not be used for the computation of the averages of thermodynamic properties of the system. The equilibration is already visible in fig 2 as the first big spike of the oscillation of the temperature, and to make this equilibration process shorter the initial velocities of the particles are sampled from a gaussian distribution such as the average kinetic energy corresponds to $k_B T_{imp}$. But for how long does it affect the data? To better see it one can plot the behavior of T^* and P^* with different block averaging parameters M (fig 5).

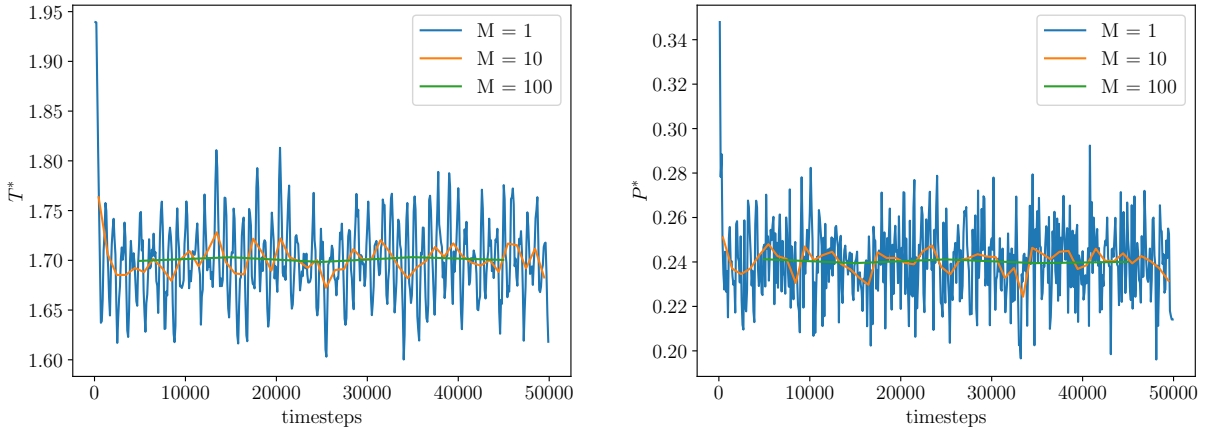


Figure 5: Trend of the temperature and pressure in the first part of the simulation plotted with a different degree of block averaging.

As can be seen from fig 5 the initial drift in temperature and pressure becomes less and less visible as M grows, and actually for $M = 100$ (that is the value used for the analysis) there seems to be no trend at all. To be conservative one can anyways discard the first two datapoints, i.e. consider the first 20000 timesteps as equilibration of the system.

One could say that it is better to do some more sophisticated analysis in order to determine more precisely the end of the equilibration phase of the simulation, but actually this will make virtually no difference in the final results since after the block averaging process the datapoints are widely spaced (10000 timesteps).

2.7 Cutoff radius for the LJ potential

Henceforth, for simplicity of notation we refer to average properties by neglecting the overbar.

When computing the forces between the particles, in principle one should consider the interaction between all possible pairs of particles. But since the LJ potential is a short range interaction one can compute only the interaction between *near enough* particles thus considerably speeding up the simulation. In practice the LJ potential is cut at a radius r_c such as particles further from each other than r_c do not interact. Then to ensure the continuity of the potential, it is rigidly shifted. If $H(x)$ is the Heaviside step function one can write the following.

$$U^*(r^*) = 4 \left(r^{*-12} - r^{*-6} \right), \quad U_{\text{cut}}^*(r^*) = U^*(r^*) H(r^* - r_c^*)$$

$$U_{\text{cut,shifted}}^*(r^*) = (U^*(r^*) - U^*(r_c^*)) H(r^* - r_c^*)$$

But cutting the potential implies neglecting the contribution to the potential energy and the pressure of the farthest particles. This contributions are called *tail corrections* and can be approximated by eq 5 [6], where the last step is good within 1% if $r_c^* > 2$. Notice that shifting the potential does not affect the pressure since this depends only on derivatives of the potential; also, adding the tail correction has no computational cost, since there is an analytic form for P_{tail}^* .

By trying different values of r_c^* at different densities one obtains the plots in fig 6, where it is clear that adding the tail correction makes the the pressure less dependent on the choice of r_c^* . In particular one can say that a reasonable value for r_c^* is 3.0

$$U_{\text{tail}}^* = 8\pi \frac{1}{v^*} \left(\frac{1}{9r_c^{*9}} - \frac{1}{r_c^{*3}} \right) \approx -\frac{8\pi}{3} \frac{1}{v^*} \frac{1}{r_c^{*3}} \quad (5)$$

$$P_{\text{tail}}^* = 16\pi \frac{1}{v^{*2}} \left(\frac{2}{9r_c^{*9}} - \frac{1}{r_c^{*3}} \right) \approx -\frac{16\pi}{3} \frac{1}{v^{*2}} \frac{1}{r_c^{*3}} \quad (6)$$

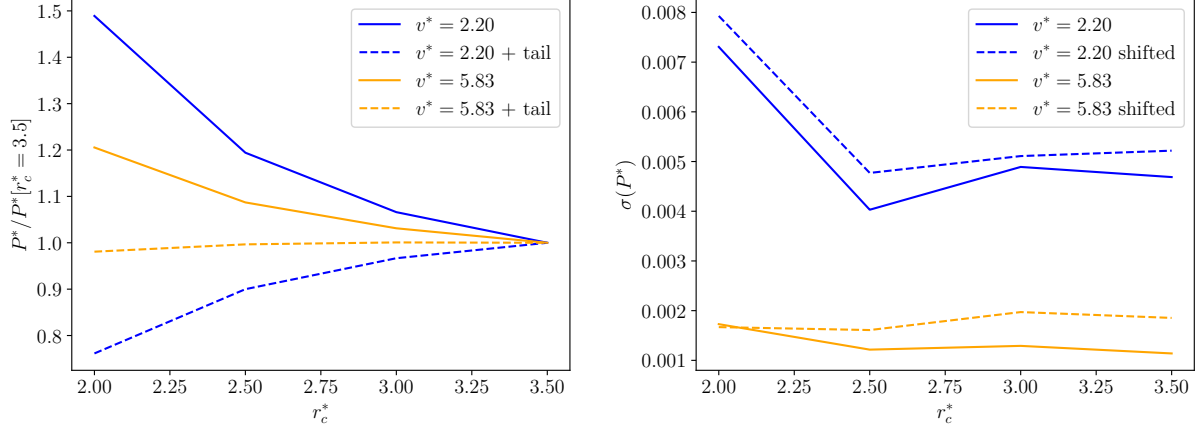


Figure 6: Behavior of the relative average pressure and its standard deviation as a function of r_c . The first with and without the tail correction, the second with and without the shift of the potential.

2.8 Neighbor list

The aim of cutting the LJ potential is to reduce the computational complexity of the simulation from $O(N^2)$ to $O(N)$. This is achieved via a *neighbor list* that for each particles contains the particles that are nearer than $r_c + \delta$ where δ is called *skin* and accounts for the fact that particles move, allowing the list to be updated every $t_{\text{update}} > dt$. From fig 7 one can see that the pressure oscillates less at $\delta^* = 0.3$, which is also the LAMMPS default value, while the average value of the pressure does not depend considerably on δ nor on t_{update} . Considering this to speed up the computations one can choose $t_{\text{update}} = 20dt$.

For completeness, LAMMPS provides other methods of updating the neighbor list, but for this kind of system they do not give any computational advantage.

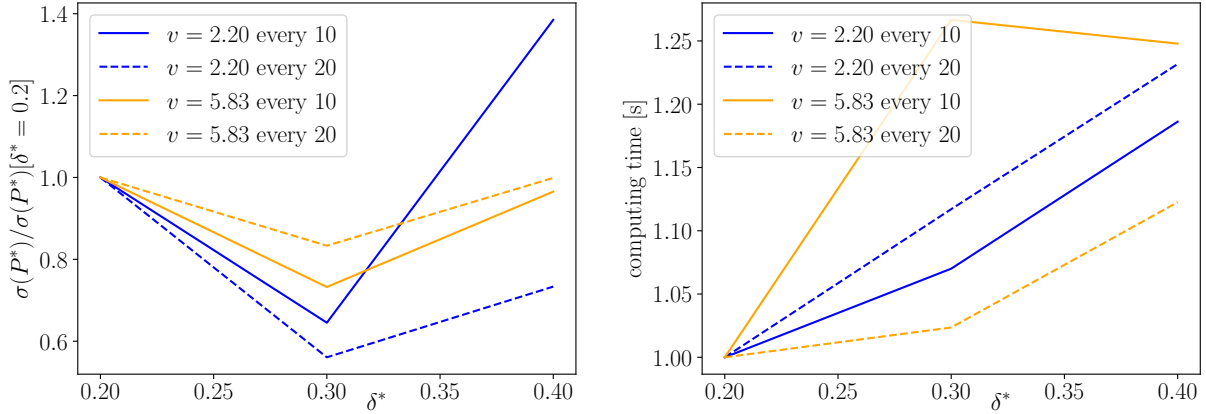


Figure 7: Behavior of the standard deviation of the pressure and the computing time of the simulation as a function of the skin δ^* and at different t_{update} . 'every 10' means $t_{\text{update}} = 10dt$.

2.9 List of simulation parameters

parameter	symbol	set value
ensemble		NVT
number of particles	N	1331
timestep	dt^*	0.004
thermostat typical time	τ_T	$200dt$
sampling time		$100dt$
block averaging parameter	M	100
length of the simulation		$200000dt$
equilibration		first 20000 timesteps
datapoints for computing the final average		18
approximated form for the LJ potential		$U_{\text{cut,shifted}}^*$
cutoff radius	r_c^*	3.0
neighbor list skin	δ^*	0.3
neighbor list updating time	t_{update}	$20dt$

Table 3: List of simulation parameters

3 Estimation of the VdW parameters

3.1 Supercritical isotherms

For supercritical isotherms it is possible to acquire the whole curve, but in order to optimize the efficiency of the simulations it is better to choose a proper spacing of the volume (fig 8). For this reason choosing to uniformly space $1/v^*$ seems the best choice.

On the other hand since at a fixed v^* P^* has a linear dependence on T^* , a uniform spacing of T^* seems reasonable.

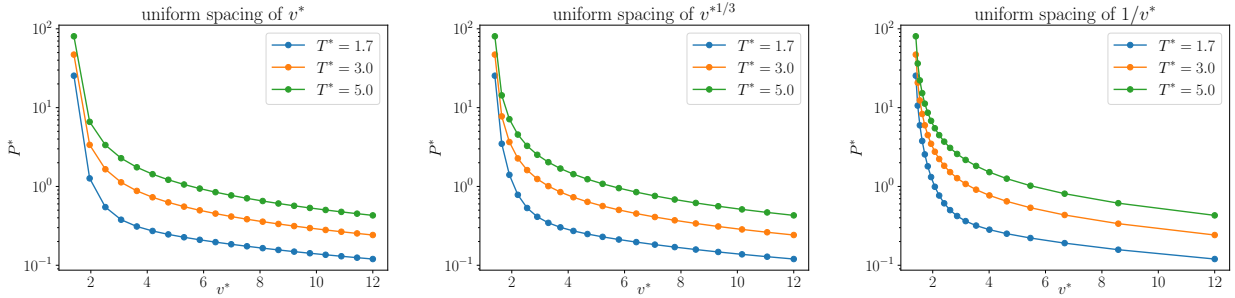


Figure 8: Points sampled from the theoretical VdW curve using Argon parameters

3.2 Subcritical isotherms

Since the VdW parameters depend on temperature it is quite tricky to know a priori if an isotherm is actually below the critical temperature. Even trickier is to avoid the phase coexistence zone. Thus this has to be done a posteriori, i.e simulate some points to roughly avoid the coexistence zone, then remove data where $\frac{\partial P^*}{\partial v^*} > 0$.

For this isotherms points are chosen by uniformly spacing the volume separately in the liquid and gas phase.

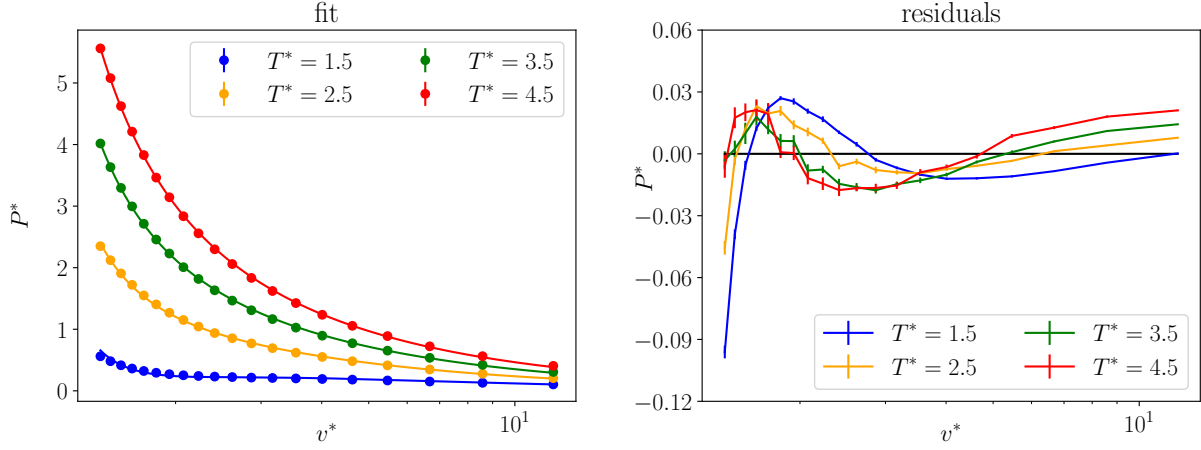


Figure 9: Fit and residuals of some of the supercritical isotherms.

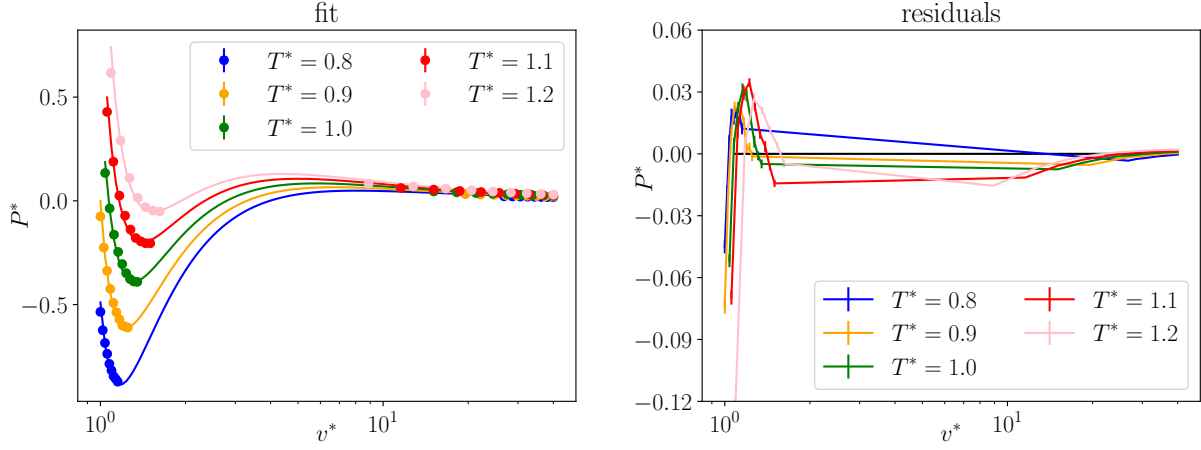


Figure 10: Fit and residuals of the subcritical isotherms. The coexistence zone has not been completely avoided, as there are points in the liquid phase with a lower pressure than the ones in the gas phase. These points have not been discarded since it would have meant to discard almost all the liquid phase.

3.3 Results

By fitting the isotherms with the VdW formula (eq 2) one obtains the results in figs 9, 10 and tab 4. The fact that there is a clear trend in the residuals, that the values of a^* , b^* and T_c^* are a decreasing function of temperature and the high values of the reduced χ^2 suggest that, even though the VdW model is quite good at describing the equation of state of the system, there are non negligible higher order terms.

Another possibility instead of fitting each isotherm, is to fit all data points with a surface. If one neglects the error on temperature (which is actually smaller than the error on pressure), one gets the results in fig 11, where the surface has VdW parameters

$$a^* = 3.592 \pm 0.001, \quad b^* = 0.7733 \pm 0.0001, \quad T_c^* = 1.3764 \pm 0.0003, \quad \chi^2/\text{Ndof} = 3223$$

The very high value of the reduced χ^2 is not a surprise given the behavior of the VdW parameters obtained by fitting individual isotherms.

T^*	a^*	b^*	T_c^*	χ^2/Ndof
0.8	3.938 ± 0.005	0.7680 ± 0.0004	1.5192 ± 0.0010	542
0.9	3.863 ± 0.003	0.7670 ± 0.0003	1.4925 ± 0.0007	658
1.0	3.863 ± 0.004	0.7733 ± 0.0004	1.4801 ± 0.0008	811
1.1	3.854 ± 0.003	0.7776 ± 0.0004	1.4685 ± 0.0006	968
1.2	3.932 ± 0.004	0.7932 ± 0.0005	1.4688 ± 0.0007	1539
1.5	4.435 ± 0.003	0.8863 ± 0.0006	1.4826 ± 0.0005	366
2.0	4.025 ± 0.005	0.8449 ± 0.0006	1.4115 ± 0.0008	150
2.5	3.589 ± 0.005	0.8085 ± 0.0006	1.315 ± 0.001	140
3.0	3.051 ± 0.006	0.7710 ± 0.0006	1.173 ± 0.002	446
3.5	2.707 ± 0.008	0.7521 ± 0.0007	1.067 ± 0.002	319
4.0	2.198 ± 0.009	0.7224 ± 0.0008	0.901 ± 0.003	558
4.5	2.01 ± 0.01	0.7173 ± 0.0008	0.829 ± 0.003	296

Table 4: Results of the fit on individual isotherms. For the subcritical isotherms the reduced χ^2 improves a lot fitting separately the liquid and gas phase, but this way one obtains unphysical results ($b^* < 0$).

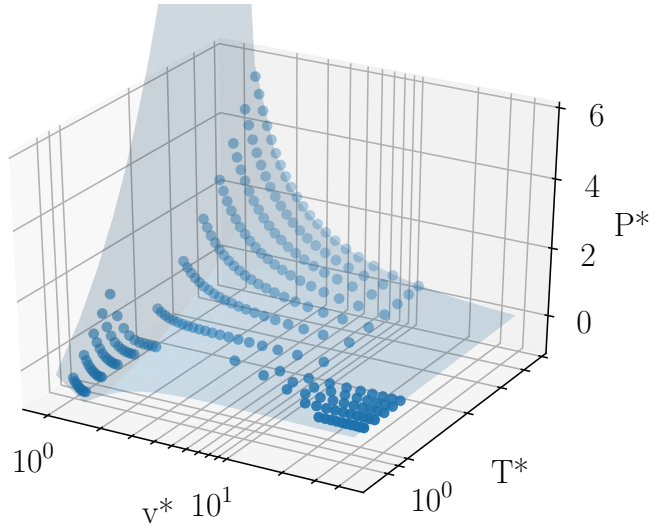


Figure 11: All valid data points from the simulations and the surface fitting them. Since the plot has a logarithmic scale on each axis, points with $P^* < 0$ are not shown.

4 Conclusion

By studying the effect of the tuning of the simulation parameters it has been possible to choose reasonable values of them when running simulations. The simulated LJ system roughly follows the VdW model but there are other effects at play which imply a dependence of the VdW parameters on the temperature. For this reason the average value of those parameters obtained by fitting all the data is meaningful only in the range of temperature considered in these simulations.

This is in fact a good result because it means that MD simulations (like real systems) go beyond the VdW approximation.

References

- [1] J. Lekner *Parametric solution of the Van der Waals liquid-vapor coexistence curve* Victoria University, Wellington, New Zeland, 3 April 1981. https://www.researchgate.net/publication/243489706_Parametric_solution_of_the_van_der_Waals_liquid-vapor_coexistence_curve
- [2] *Van der Waals constants for real Gases* http://www2.ucdsb.on.ca/tiss/stretton/database/van_der_waals_constants.html
- [3] Wikipedia *Critical Point (thermodynamics)* [https://en.wikipedia.org/wiki/Critical_point_\(thermodynamics\)](https://en.wikipedia.org/wiki/Critical_point_(thermodynamics))
- [4] LAMMPS <https://lammps.sandia.gov>
- [5] Y. Kataoka & Y. Yamada *Van der Waals type equation of state for Lennard-Jones fluid and the fluctuation of the potential energy by molecular dynamics simulations*, 17 August 2011 <http://dx.doi.org/10.1080/08927022.2011.551881>
- [6] *Simulations of bulk phases* https://sites.engineering.ucsb.edu/~shell/che210d/Simulations_of_bulk_phases.pdf