Determination of the diffusion coefficient of liquid He at high temperature by first principles molecular dynamics

Saurabh Sivakumar

ABSTRACT

In this study an AIMD simulation using the Bussi – Donadio – Parinello thermostat for the liquid helium system of 125 atoms is performed using the open-source quantum simulation toolbox Qbox for various temperatures ranging from 1000 – 15000 K. The diffusion coefficient is then calculated by using the Einstein's relation and the mean squared displacement. The diffusion coefficient increases as a function of temperature and subsequently the Arrhenius relation for the diffusion is studied. Error measures for the quantities calculated are provided as a statistical measure, the standard error. Other possible sources of error that may arise in the simulation are noted and methods to rectify the same are suggested.

INTRODUCTION

Computer simulation methods, such as Monte Carlo or Molecular Dynamics (MD), are powerful computational techniques that are often used to provide detailed and essentially exact information on classical many-body problems. [1] MD is basically the numerical solution of Newton's equations of motions for several molecules in a system. This allows for the calculation of both the equilibrium thermodynamic and dynamical properties of a system to be computed. [2] However, any computational resource is finite, which limits the time and length scales accessible by computer simulations. One way to partially bridge the gap between the microscopic size of the simulated system and the macroscopic reality is to introduce periodic boundary conditions (PBC). [1] [2]

Molecular systems typically consist of many particles, it is impossible to determine the properties of such complex systems analytically; MD simulation circumvents this problem by using numerical methods. However, long MD simulations are mathematically ill-conditioned, generating cumulative errors in numerical integration that can be minimized with proper selection of algorithms and parameters, but not eliminated entirely. [3] The base idea behind MD is that we must empirically generate the interacting potential between atoms. The origins of potentials (nuclei and electrons) are completely excluded in this method. Due to this reason, electronic and magnetic properties cannot be obtained. Once the potential is known and the initial positions and velocities are given, the time-evolutions of the atoms are then computed. [4] [5]

When we deal with the translational or rotational motion of smaller atoms such as Helium, we must consider the quantum effects in addition to the traditional classical MD approach. Furthermore, In the standard MD approach, the instantaneous force on each atom is calculated as a gradient of a prescribed interatomic potential function (force field), which is generally a function of the atomic coordinates that you can equally well regard as the coordinates of their nuclei. A first-principles based approach, such as ab-initio molecular dynamics (AIMD), where the forces are calculated on-the-fly from accurate electronic structure calculations, is thus more

useful here. [6, 2] However, the increased accuracy and predictive power of AIMD simulations comes at significant computational cost. For this reason, density functional theory (DFT) is to date the by far the most employed electronic structure theory, but it is important to note that AIMD is a general concept that in principle can be used in conjunction with any electronic structure method. [7] [8] In AIMD the interatomic forces are computed using the electronic structure method, in this case DFT. With the advent of AIMD, where the forces are computed on-the-fly by accurate electronic structure calculations, the scope of either method has been greatly extended. The pioneering work in this domain was done by Roberto Car and Michele Parrinello by extending the Born – Oppenheimer approximation. [9]

In this paper an AIMD simulation is run for 125 Helium atoms for a range of temperatures from 1000 to 15000K. The simulation is run in the canonical ensemble where the amount of substance (N), volume (V) and temperature (T) are conserved. In NVT, the energy of endothermic and exothermic processes is exchanged with a thermostat. A variety of thermostat algorithms are available to add and remove energy from the boundaries of an MD simulation in a realistic way, approximating the canonical ensemble. Popular methods to control temperature include velocity rescaling, the Nosé–Hoover thermostat, Nosé–Hoover chains, the Bussi-Donadio-Parrinello thermostat, the Berendsen thermostat, the Andersen thermostat and Langevin dynamics. [10] [11] [12] [13] The Bussi-Donadio-Parrinello thermostat is what will be used in the simulation. It is a thermostat that was designed to rectify the 'flying ice cube effect' present in previous thermostats such as the Berendsen Thermostat although they both use velocity rescaling to fix the temperatures.

The BDP thermostat rescales velocities using a coefficient α which is sampled from a Gamma distribution. The thermostat reproduces the fluctuations of the kinetic energy of the particles in NVT. The total linear momentum and, for non-periodic systems, the angular momentum are conserved. The formalism can also be trivially extended to thermalize independently different parts of the system. The performance of the BDP thermostat is independent from the choice of the time constant of the thermostat for a sufficiently long simulation. [11] [14]

Here the objective is to use the BDP thermostat in tandem with AIMD methods to determine the diffusion coefficient/diffusivity of the Helium system. Diffusion coefficient is a proportionality constant between the molar flux due to molecular diffusion and the gradient in the concentration of the species (or the driving force for diffusion). Diffusivity in the macroscale is encountered in Fick's law and numerous other equations of physical chemistry. [15, 16] Einstein defined the diffusion coefficient in his work on the Brownian motion as the following in terms of the mobility of the particle, the Boltzmann constant, and the temperature.

$$D = \mu k_b T$$

The formula contains several terms that are tough to determine from an MD a simulation. Therefore, we use an alternative approach by computing the mean square displacement of the particles over the course of a simulation. In statistical mechanics, the mean squared displacement (MSD) is a measure of the deviation of the position of a particle with respect to a reference position over time. [17] The MSD at time is an ensemble average and thus can be

computed easily from a simulation. In this paper the diffusivity is calculated using the Einstein relation for a Brownian particle in 3 dimensions. [2]

$$< r^2 > = \ 6Dt$$

Using the above formula, the diffusivity of liquid Helium can be computed by finding the MSD from an AIMD simulation run for a sufficient time. Liquid helium is a physical state of helium at low temperatures and standard atmospheric pressures. Liquid helium may show superfluidity. The temperature required to produce liquid helium is low since the attractions between the helium atoms are relatively weak. This is so because helium is a noble gas, and the interatomic attractions are reduced even more by the effects of quantum mechanics. These are significant in helium because of its low atomic mass of about four atomic mass units. The zero-point energy of liquid helium is less if its atoms are less confined by their neighbors. Important early work on the characteristics of liquid helium was done by Lev Landau, later extended by physicist Richard Feynman. [18]

METHODS

An AIMD simulation of 125 Helium atoms in a 10 Bohr unit cell is set up using a LDA exchange correlation functional with a plane wave energy cutoff of 45 Rydberg. A time step of 30 atomic units is chosen for the MD. Using these settings and the BDP thermostat, AIMD simulations are run for a timescale of ~0.015 nanoseconds (ns) for the temperatures **1000 K, 2000 K, 4000 K, 5000 K, 7500 K, 10000 K, 12500 K and 15000 K.** The simulations are run using the open-source software Qbox on the UC Davis CSIF computers. [19]. The initial velocities for the simulations are sampled from a Boltzmann distribution. Since we use a velocity rescaling thermostat, it is usually a better idea to choose the initial velocities from a distribution close to the target temperature.

Typical AIMD simulations are run for a timescale of 0.1 – 1 ns. In this work, attempts were made to run the simulation for 120000 MD steps (3600000 a.u time ~ 0.9 ns), however system limitations and frequent crashes led to a limited timescale of 0.015 ns being considered. The simulations for different temperatures terminated at different timescales, thus for consistency and the sake of brevity a uniform timescale is chosen (Refer Appendix). This shorter timescale meant the thermostat did not reach the equilibrium temperature. (Not a large concern in the calculation of the diffusion coefficient however). Using the simulation data, a utilities script provided with the code is used to compute the mean squared displacement for each step of the MD simulation.

Subsequently the MSD data is plotted vs the time and then fit using a linear line in the region where it shows a linear behavior. The fitting is done using the python package scipy and subsequently the error metrics and the measures of the fit are computed.

RESULTS

Figure 1 describes the evolution of the MSD with time for the 8 temperatures considered.

Figure 1: Top: Graph showing the MSD (Bohr^2) on the y axis vs the Time in atomic units for the temperatures 1000 K and 2000 K. **Bottom:** Graph showing the MSD (Bohr^2) on the y axis vs the Time in atomic units for the temperatures 4000 K, 5000 K, 7500 K, 10000 K, 12500 K and 15000 K.

Subsequently, a linear fit is performed on the data where they show linear behavior, and the results are noted. The associated graphs along with the fits are shown in **Figure 2.** The diffusion coefficient is calculated as the slope of the fitted line divided by 6 in accordance with Einstein's relation. The regions for the fit are chosen by the fact that these represent the most linear path although for higher temperatures the relation between MSD and Time is generally linear, a smaller window is used to avoid noise.

Tempera ture Kelvins	Range for the fit in MD timesteps (Region of linear behavior) (Bohr)	Slope of the fitted line (Bohr^2/ a.u)	Standard error of the fitted slope	R^2 of the fit	95% confidence interval of slope	Diffusion Coefficient (cm^2/s)	Standard Error of the diffusion coefficient
1000	[0,400]	0.0003104	6.39813e-06	0.855	+/- 1.25412e-05	**1.1319e-05**	1.0663e-06
2000	[0,400]	0.0004513	4.75589e-06	0.957	+/- 9.32219e-06	**1.6455e-05**	7.9264e-07
4000	[0,10000]	0.0005024	3.07103e-07	0.996	+/- 6.01964e-07	**1.8319e-05**	5.1183e-08
5000	[0,10000]	0.0007562	5.59324e-07	0.994	+/- 1.09635e-06	**2.7574e-05**	9.3220e-08
7500	[0,10000]	0.0017835	1.46414e-06	0.993	+/- 2.86992e-06	**6.5029e-05**	2.4402e-07
10000	[0,8000]	0.0031749	2.60185e-06	0.994	+/- 5.09998e-06	**1.1576e-04**	4.3364e-07
12500	[0,8000]	0.0046435	1.66226e-06	0.998	+/- 3.25826e-06	**1.6931e-04**	2.7704e-07
15000	[0,8000]	0.0050534	3.63602e-06	0.995	+/- 7.12709e-06	**1.8425e-04**	6.0600e-07

As we can see from the R^2 values of the fit, the fits are accurate in the region where they are being fit to. The data is linear in these limits, so the estimates of the diffusion coefficients will be accurate, and this is reflected in the low values of standard errors for the estimates. Subsequently when we plot the diffusion coefficients vs the temperature along with the error bars, we find the bars to be so small that they are barely visible for most cases. **(Figure 3)**

The diffusion coefficient increases in magnitude as the value of temperature increases. This is the expected behavior as at higher temperatures the atoms in the simulation tend to move a lot more leading to higher MSD and subsequently higher Diffusivity. [20]
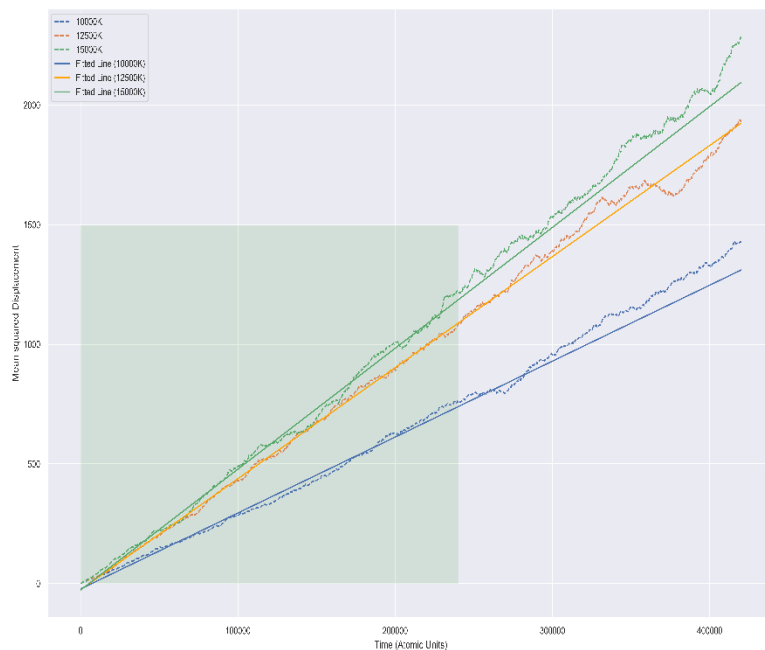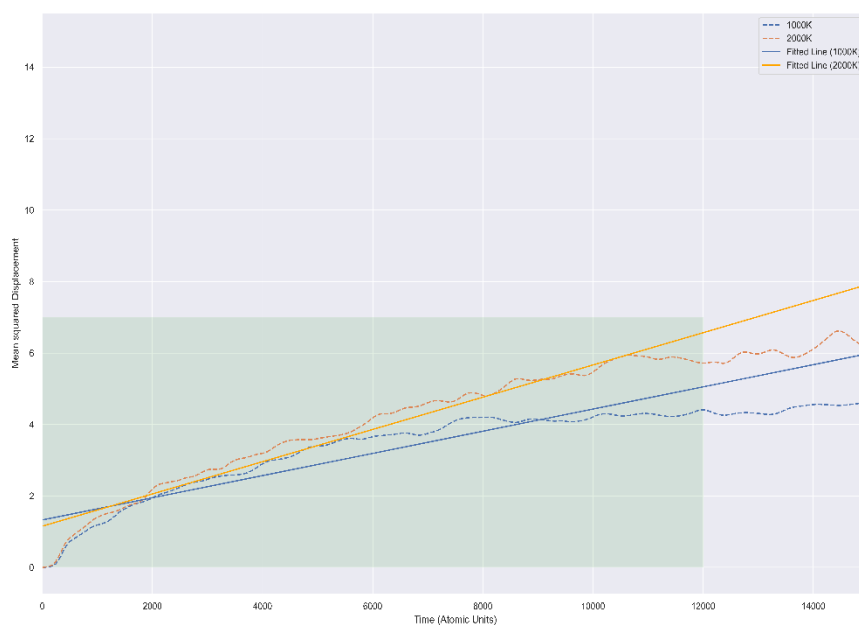
Figure 2: MSD (Bohr^2) vs Time (a.u) along with the fitted line and the shaded green area represents the region of the graphs used for the fitting. **Top:** For temperatures 1000 K and 2000 K (smaller timestep than the full for visualization). **Bottom Left:** For temperatures 4000 K, 5000 K and 7500 K. **Bottom Right:** For temperatures 10000 K, 12500 K and 15000 K
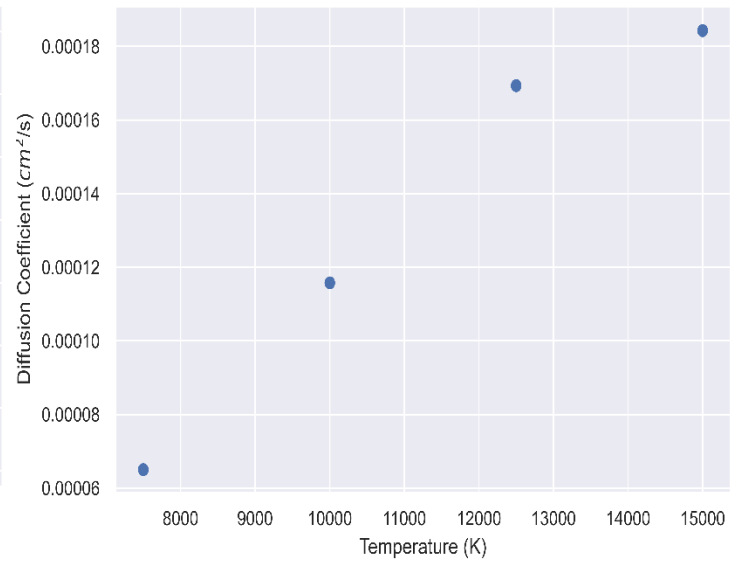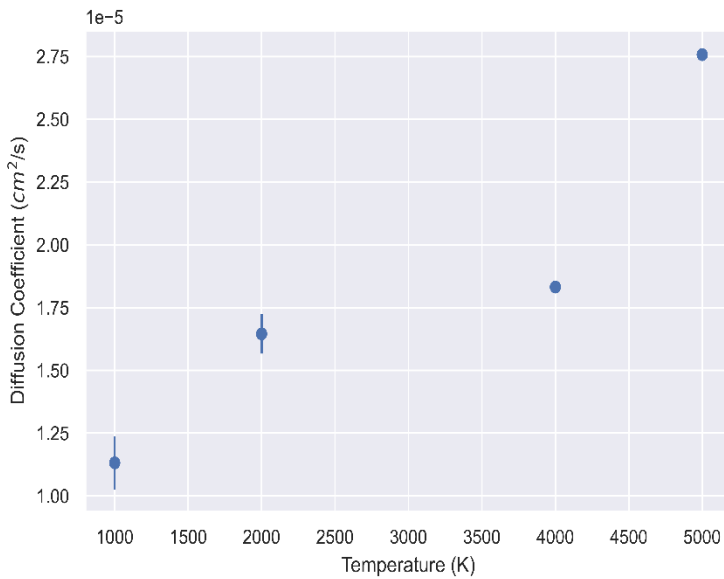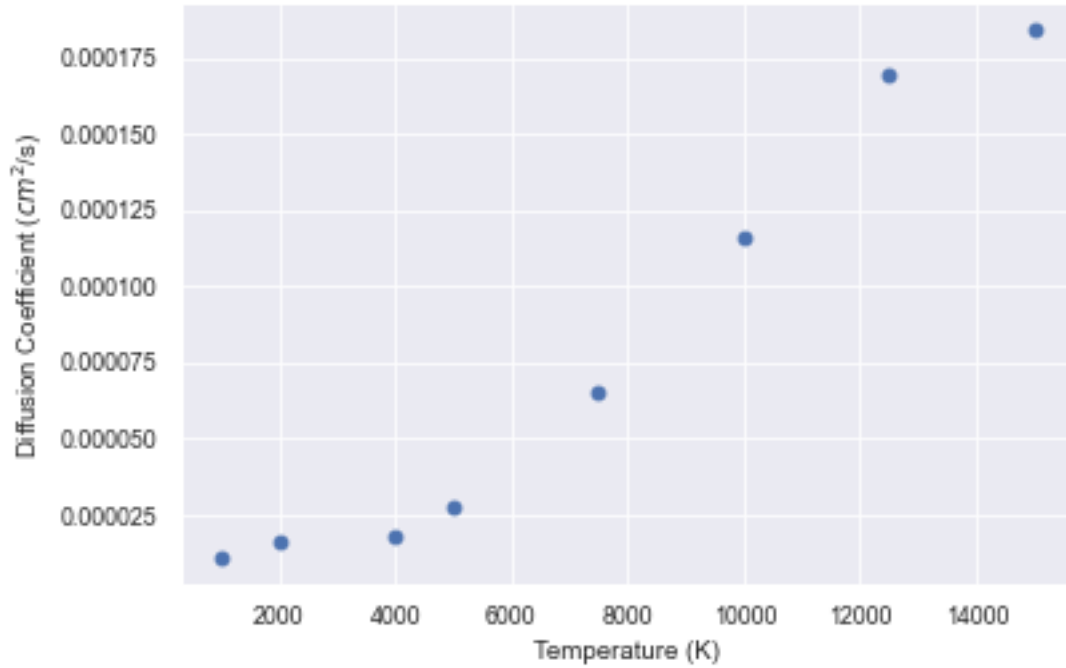
Figure 3: Diffusion Coefficient (cm^2/s) vs Temperature (K). The error bars for some cases are not visible are they are very small and cannot be represented on the scale. **Top:** The diffusion coefficients are plotted for all temperatures without any error bars **Bottom Left:** For temperatures 1000 K, 2000 K,4000K and 5000 K. with the error bars. **Bottom Right:** For temperatures 7500 K ,10000 K, 12500 K and 15000 K

DISCUSSION OF RESULTS

The results clearly indicate that the values of the diffusion coefficient increase with an increase in temperature. Since the fits are made in regions of great linear correlation the values of the slopes are estimated with reasonable accuracy. Subsequently a fit is made of the log of the diffusion coefficient vs 1000/T in an attempt to model the Arrhenius relation for diffusion which is given below. The constant D_0 is the maximal diffusion coefficient at infinite temperature, E_A is the activation energy for diffusion and R is the universal gas constant. [21] [22]

$$D = D_0 \, exp\left(\frac{-E_A}{RT}\right)$$

As we can see the above equation can also be expressed in terms of the natural logarithm as

$$log \, D = log \, D_0 - \frac{E_A}{R}\frac{1}{T}$$

When plotting the log of the diffusion coefficient vs 1000/T and fitting a linear line, we can find the constant D_0 and the slope as $-\frac{E_A}{R}$.
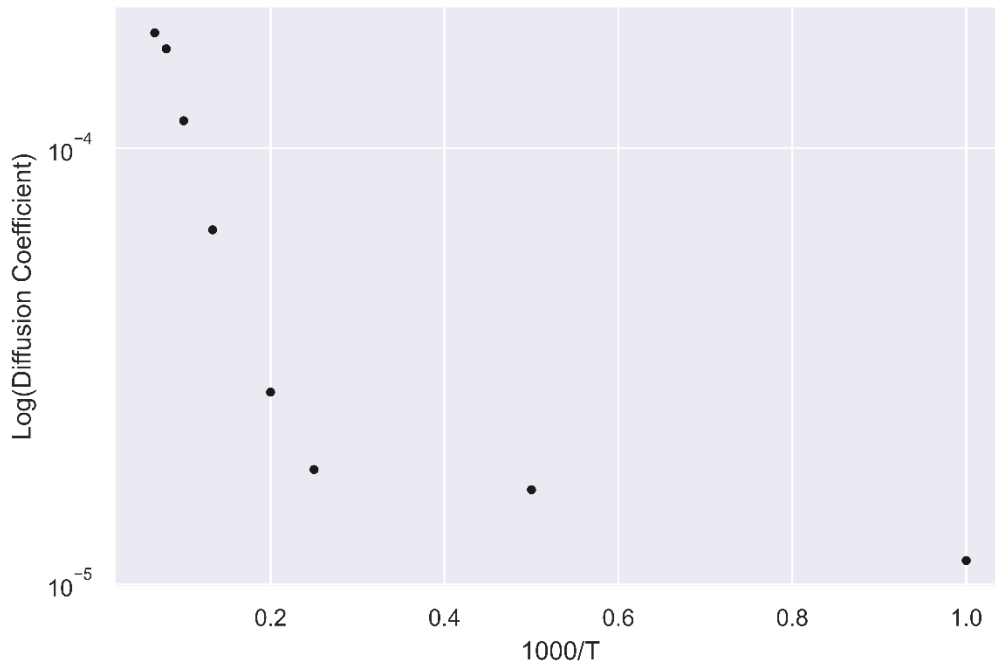

Figure 4: Log (Diffusion coefficient) vs 1000/T

The behavior is linear in the range for T > 2000 K based on the 8 values of temperatures used to make the plot. Subsequently the system would not represent a great fit using the Arrhenius equation in the lower temperature regime. At higher temperatures this might be a good fit. This is in accordance with what we observed with the MSD vs time plots. For the lower 2 temperatures of 1000 K and 2000 K the behavior of the MSD with respect to the time is like that of a solid rather than a liquid. It may be fair to assume that at these temperatures and simulation conditions Helium behaves as a solid rather than a liquid. [23] [24]

Using a linear fit to the above data we obtain the constant D_0 the maximal diffusion coefficient at infinite temperature as 1.0001 cm^2/s and the slope $-\frac{E_A}{R}$ as -0.14227. The coefficient of determination of the fit, the measure of how good the fit is quite poor however at 0.41 which is what we expected after noting the behavior of the system as described.

This would also explain why the error in measuring the diffusion coefficients in the lower temperatures of 1000 K and 2000 K was higher than the other cases. If the system indeed exhibits solid properties in the regime, these are cases where the Einstein's relation is not a good approximation and more accurate measurements, and other relations may be needed. This generally arises due to the characteristic vibrational frequency that is present in all solids. The relation between the diffusion coefficient D and the velocity autocorrelation function can be modelled. This expression is known as the Green Kubo expression and is a significant result in statistical mechanics and is often seen as a more general result for determining the diffusion coefficient.

Now the possible sources of error in the simulation we conducted are discussed in detail. Firstly, as mentioned above the relative runtime of the simulation is fairly small in the macroscopic scale (~0.015 ns). A simulation must be modelled to a macroscopic timescale of 1 – 10 ns typically to calculate macroscopic properties. This could potentially be a source of error in the simulation here. Furthermore, the BDP thermostat is invariant to changes in the time constant, τ only when run for larger timescales such as 1 ns. [11]. Since this time is not achieved in this simulation the time constant for the thermostat needs to be optimized for every temperature.

The consequences of not strictly following these procedures would have led to errors in the simulation. The results are evident when we look at the figure showing the variation in temperature over time in the Appendix. There is a lot of variation, and the thermostat hasn't reached an equilibrium yet. Other possible errors could have arisen from the value of the time step being considered. A time step of 30 a.u may be too large for the system being considered. There is also the case of the inherent error associated with any linear fit. The graphs are fit with the method of least squares which itself deals with a minimization procedure to obtain the slopes and intercepts. The method always has inherent errors, and this is reflected in the standard errors of the slope obtained. Although the values in this case are very small due to the data.

A few other possible sources that could have arisen in the simulation come from the fact that the simulation uses a fixed cutoff energy provided. In an ideal scenario the cutoff energy must be optimized for the system and calculations will be needed to run for that as well. Additionally, since the BDP thermostat is a velocity rescaling thermostat, better results could possibly be obtained by running with an alternate thermostat such as the Nose – Hoover. [25] A Nose-Hoover thermostat

operates by tethering the simulated system to a virtual mass, using one or more virtual chains. Temperature of the simulated system is controlled by importing or exporting energy to and from the simulated system, using the tethered virtual mass. This thermostat guarantees a Maxwell-Boltzmann velocity ensemble for a sufficiently large (and long) sample size. Lastly a change that will improve the results but at the cost of greater runtime is using a different and better exchange correlation functional. The local density approximation (LDA) functional used in the simulation is not necessarily the best and more advanced functionals such as GGA or Hybrid or meta -GGA functionals will lead to more accurate results. [7]

CONCLUSION

The simulations were conducted according to the given requirements and the results were studied accordingly. A linear fit was made for the various cases of temperatures and the diffusion coefficient was measured using Einstein's relation for 3D Brownian motion. The salient points to note are that the Diffusion coefficient increases by an order of magnitude with an increase in the temperature. The Arrhenius relation for the data was studied and the results discussed. The possible sources of error in the simulations were discussed and solutions to minimize them were suggested.

REFERENCES

[1] T. D. Kuhne, "Ab-Initio Molecular Dynamics," WIREs Comput. Mol. Sci, pp. 391 - 405, 2014.

[2] D. Frenkel and B. Smit, Understanding molecular simulation From Algorithms to Applications, San Diego: Academic Press, 2002.

[3] M. Tuckerman, B. J. Berne and G. J. Martyna, "Reversible multiple time scale molecular dynamics," The Journal of Chemical Physics, vol. 97, no. 3, 1992.

[4] D. Beeman, "Some multistep methods for use in molecular dynamics calculations," Journal of Computational Physics, vol. 20, pp. 130-139, 1976.

[5] L. Vertlet, "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules," American Physical Society, vol. 159, 1967.

[6] D. Marx and J. Hutter, Ab Initio Molecular Dynamics:Basic Theory and Advance Methods, Cambridge: Cambridge University Press, 2009.

[7] D. S. Sholl and J. A. Steckel, Density functional theory: A practical Introduction, Wiley, 2011.

[8] R. M. Martin, Electronic Structure: Basic Theory and Practical Methods, Cambridge: Cambridge University Press, 2020.

[9] R. Car and M. Parrinello, "Unified Approach for Molecular Dynamics and Density-Functional Theory," PHYSICAL REVIEW LETTERS, vol. 55, no. 22, 1985.

[10] S. Nosé, "A unified formulation of the constant temperature molecular-dynamics methods," Journal of Chemical Physics, vol. 81, no. 1, pp. 511-519, 1984.

[11] G. Bussi, D. Donadio and M. Parrinello, "Canonical sampling through velocity-rescaling," J. Chem. Phys, vol. 126, 2007.

[12] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, " Molecular-Dynamics with Coupling to an External Bath," Journal of Chemical Physics, vol. 81, no. 8, 1984.

[13] H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," The Journal of Chemical Physics, vol. 72, no. 4, 1980.

[14] W. G.Hoovera and B. LeeHolian, "Kinetic moments method for the canonical ensemble distribution," Physics Letters A, vol. 211, no. 5, pp. 253-257, 1996.

[15] E. L. Cussler, Diffusion: Mass Transfer in Fluid Systems, New York: Cambridge University Press, 1997.

[16] U. M. B. Marconi, A. Puglisi, L. Rondoni and A. Vulpiani, "Fluctuation-Dissipation: Response Theory in Statistical Physics," Physics Report, vol. 461, p. 111, 2008.

[17] J. R. Welty, C. E. Wicks, R. E. Wilson and G. Rorrer, Fundamentals of Momentum, Heat, and Mass Transfer, Wiley, 2014.

[18] J. Wilks, The Properties of Liquid and Solid Helium, Oxford: Clarendon Press, 1967.

[19] F. Gygi, "Architecture of Qbox: A scalable first-principles molecular dynamics code," IBM Journal of Research and Development (, vol. 52, no. 1.2, 2008.

[20] K. A. Dill and S. Bromberg, Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology, Garland Science, 2003.

[21] K. J. Laidler, Chemical Kinetics, Harper & Row, 1987.

[22] L. Shcherbak, "Empirical Correlations between the Arrhenius' Parameters of Impurities' Diffusion Coefficients in CdTe Crystals," Journal of Phase Equilibria and Diffusion , 2015.

[23] R. Levine, Molecular Reaction Dynamics, Cambridge University Press, 2005.

[24] MIT OCW, "ocw.mit.edu," October 2003. [Online]. Available: https://ocw.mit.edu/courses/nuclear-engineering/22-103-microscopic-theory-of-transport-fall-2003/lecture-notes/lec2.pdf. [Accessed 12 3 2021].

[25] J. Ruiz-Franco, L. Rovigatti and E. Zaccarelli, "On the effect of the thermostat in non-equilibrium molecular dynamics simulations," THE EUROPEAN PHYSICAL JOURNAL E, vol. 41, p. 80, 2018.

APPENDIX

The data and the associated graphs are attached in the zip file along with the report. The data contains information for the temperature and the mean square displacement for various time steps. As noted above, the MD simulations for various cases ran until certain timesteps were achieved before they crashed. The data is thus not uniform in the MD steps and 14000 was used as it represents 0.015 ns and the temperature of 10000 K ran for the lowest MD steps of ~14200. Some other simulations like the one for 1000 K (~35000 MD steps) ran for larger MD steps, however unfortunately they could not be used as the values for the other temperatures do not reflect the same level of detail. The data for these other simulations however is still present in the submission.

Additionally, below a plot of the temperature vs the time is shown for perusal.