Evaluating Diffusion Coefficients in LAMMPS

David Keffer
Department of Materials Science & Engineering
University of Tennessee, Knoxville

date begun: March 28, 2016
date last updated: February 27, 2018

Table of Contents

I. Purpose of Document

The purpose of this document is to provide a practical introduction to the evaluation of diffusion coefficients in LAMMPS. The notes begin with some formal theory and conclude with practical implementation.

II. Types of Diffusion Coefficients

Diffusion occurs in various systems. Diffusion coefficients therefore take on different meanings depending upon the type of system they are describing. In this document, we examine two types of diffusion: (1) self-diffusion and (2) Fickian diffusion. The diffusion coefficients describing these processes are not the same but they are related.

II.A. Self-Diffusion

We often think of diffusion as a response to a concentration gradient. Remember that concentration of component A, c_A, is the product of the inverse of the molar volume, $\overline{V}$, and the mole fraction, x_A.

$$c_A = \frac{1}{\overline{V}} x_A \tag{1}$$

Thus a concentration gradient can be due to either a gradient in the molar volume (equivalently the density) or a gradient in composition or both. In a system at equilibrium, there is no concentration gradient. In a pure component system, the concept of a concentration gradient only exists if there is an external force, for example a pressure gradient, giving rise to a density gradient. In a multicomponent system at equilibrium, there is neither a composition gradient nor a density gradient.

Yet in a system at equilibrium, either single component or multicomponent, atoms and/or molecules (collected in the term particles) still experience Brownian motion. This motion is called self-diffusion and is described by a self-diffusion coefficient, D_{self}. Each particle experiences a random motion that allows it to move in space without any corresponding change in the averaged density or composition profile. In a single component system, there is one self-diffusion coefficient, D_{self}. In a multicomponent system, each species has its own self-diffusion coefficient, $D_{self,A}$, $D_{self,B}$, etc.

The self-diffusion coefficient measured in molecular dynamics (MD) simulations is most like the isotopic tracer diffusion coefficient measured using, for example, Pulse-Field Gradient (PFG) Nuclear Magnetic Resonance (NMR). In this case, there is no gradient. While the self-diffusivity of only the isotope active to NMR is measured, it is generally assumed (especially for elements heavier than hydrogen) that the self-diffusivity of all isotopes is the same.

Self-diffusion coefficients are relatively easy to get from an MD simulation. In part this is because, self-diffusion coefficients rely on a single-particle correlation function, as shown below.

This means we can average our result over all particles of a given species to increase the statistical reliability of the result.

II.A. Fickian (or transport) Diffusion

In a multicomponent case, we often think of Fick's law as the definition of diffusion, just as we think of Fourier's law as the definition of heat conduction. Thus, Fickian diffusion describes a mixing process in which a non-uniform distribution of species becomes more uniform. From the point of view of thermodynamics, this is an entropy-generating process. The underlying driving force for this mixing process is again the Brownian motion of the particles, but the conventions by which diffusion coefficients are introduced to describe this process are completely different from self-diffusion.

Often in an undergraduate course, Fick's law is presented as

$$J_A = -D\nabla c_A \qquad \text{(meaningless, generic version of Fick's law)} \qquad (2)$$

where J_A is a diffusive flux and D is a diffusion coefficient. Without further explanation, this is a meaningless and generic version of Fick's law. It convey an idea but cannot be unambiguously used in a quantitative sense. Because MD simulations are intended to deliver quantitative results, we need to be much more careful than this.

Clearly, a Fickian diffusivity is not defined for a single component system. More subtly, multiple diffusion coefficients exist for even a binary system, D_{AB}, the diffusion coefficient of A relative to B, and D_{BA}, the diffusion coefficient of B relative to A. Unless one is careful, these diffusion coefficients are not the same. We will discuss this further below.

The Fickian diffusion coefficient measured in MD simulations describes are the same kind of diffusion coefficients generated through such undergraduate experiments as the gaseous diffusion measured by Winkelmann's method (see for example, http://utkstair.org/clausius/docs/che310/index.html), in which acetone diffuses in air from a vapor/liquid interface, or the salt diffusion measurement (also at http://utkstair.org/clausius/docs/che310/index.html), in which NaCl diffuses from a concentrated solution into a dilute solution.

Fickian diffusion coefficients are much more difficult to obtain reliably from MD simulation than are self-diffusion coefficients. This is because, Fickian diffusion coefficients rely on a many-particle correlation function, as shown below. This means we follow the center of mass of each species and cannot rely on particle-averaging (there is only one center-of-mass per species) to increase the statistical reliability of the result.

III. Functionality of Diffusion Coefficients

Like any thermodynamic or transport property, the self-diffusion coefficient is a function of thermodynamic state, for example, temperature, *T*, density, ρ, and composition, $\underline{x}$. Since the diffusivity is frequently most sensitive to temperature, often its dependence on the other two variables, density (or its conjugate variable, pressure) and composition, is ignored. Ignoring the dependence of the diffusivity on pressure and composition is an approximation.

Transport properties can in principle also be functions of the non-equilibrium fields, such as the concentration gradient. (Shear-thinning polymers provide an example of a transport property, the shear viscosity, which is a function of the non-equilibrium field strength, the shear rate.) Typically, one makes the assumption that a diffusion coefficient is not a function of field strength. There is no theory behind this, only empirical evidence. In essence, we obtain reasonable agreement with experiment when we truncate the Taylor series describing the diffusive flux as a function of the concentration gradient at the linear term.

IV. Derivation of Fickian Diffusion Coefficients

For this part of the lecture, refer to the following article.

Keffer, D.J., Gao, C.Y., Edwards, B.J., "On the Relationship between Fickian Diffusivities at the Continuum and Molecular Levels", *J. Phys. Chem B.* **109** 2005 pp. 5279-5288, doi: 10.1021/jp0446635.

V. Expressions for the Self-diffusion coefficients from MD

The derivation for the expression of a self-diffusion comes from a Green-Kubo integral. See for example Chapters 7 and 8 (especially Table 8.1) of Hansen & McDonald. For the case of the self-diffusivity, the argument of the Green-Kubo integral is the velocity auto correlation function (VACF). The diffusivity can equivalently be obtained from the mean square displacement (MSD). This is the more common approach, although both methods are formally equivalent. In Chapter 7, Haile has a simple and clear derivation of the MSD form from the VACF form. The details of these citations are on the references page of the course website.

In brief, the self-diffusion coefficient in the α direction can be obtained from the integration of the VACF.

$$D_{\alpha,self} = \int_0^\infty d\tau \left\langle v_{i,\alpha}(t+\tau) v_{i,\alpha}(t) \right\rangle \qquad (3)$$

In this expression, the angled brackets indicate an ensemble average, which is an average over both all particles $i = 1$ to N and an average over all time origins, t. This latter fact means that every time step in the MD simulation can be used as a time origin in the calculation of the VACF. We will discuss the practical implementation of this in the next section.

The self-diffusion coefficient can be also obtained from the mean square displacement (MSD).

$$D_{\alpha,self} = \frac{1}{2}\lim_{\tau\to\infty}\frac{\left\langle \left[r_{i,\alpha}(t+\tau) - r_{i,\alpha}(t)\right]^2\right\rangle}{\tau} \tag{4}$$

In this expression, often called Einstein's relation for the diffusivity, the angled brackets again indicate an ensemble average. The positions used in this calculation cannot have had periodic boundary conditions (PBCs) applied to them. If they have, then the trajectories have to be unfolded.

In expressions for both the VACF and MSD, infinity appears. Of course, one doesn't go to infinity observation times. At some point the effects become negligible. We discuss the impact of the choice of the infinite time limit in several examples below.

Regardless of the whether one uses the MSD or the VACF to compute the diffusivity, the average self-diffusion coefficient in an isotropic system in three-dimensions is given by

$$D_{self} = \frac{1}{3}\sum_{\alpha=1}^{3} D_{\alpha,self} \tag{5}$$

VI. Practical Determination of Self-diffusion coefficients from MD

For this part of the lecture, refer to the following online notes.

An Introduction to Self-Diffusion Coefficients

located at http://utkstair.org/clausius/docs/che548/pdf/selfD.pdf .

This includes code in FORTRAN and MATLAB to generate self-diffusion coefficients from the mean square displacement.

VII. Two Simple Checks

There are two very simple checks that should always be done after computing a diffusion coefficient from an MD simulation.

First, you should look at the maximum value of the MSD used in your calculation. Take the square root of that maximum. This root-mean-square (RMS) displacement has units of length and reveals the average distance between the starting and ending point of the simulation. If this distance is very small, you cannot report a reasonable diffusivity. For example, if your RMS displacement is on the order of 1 Å, then this means that the atom really hasn't moved at all. This is important. Certainly, it can be reported that during the course of the simulation, the RMS displacement was small. However, a diffusivity generated from such a small displacement is likely meaningless. Reasonable diffusivities are generated when a particle is able to explore a statistically varied environment.

An example of where the RMS displacement is reported to demonstrate inter-cage motion a nanoporous material in this is in the following article. (See tables SI-12 and SI-13 in the supplementary information.

Xiong, R., Odbadrakh, K., Michalkova, A., Luna, J.P., Petrova, T., Keffer, D.J., Nicholson, D.M., Fuentes-Cabrera, M.A., Lewis, J.P., Leszczynski, J., "Evaluation of Functionalized Isoreticular Metal Organic Frameworks (IRMOFs) as Smart Nanoporous Preconcentrators of RDX", *Sensor Actuat B-Chem* **148** 2010 pp. 459–468, doi: 10.1016/j.snb.2010.05.064.

Second, you calculate the exponent relating the MSD to the observation time in order to demonstrate that you have obtained a result in the infinite-time-limit required by Einstein's relation for the diffusivity. It is not enough to look at the plot. Often the sub-diffusive behavior looks "pretty linear" to the eye. This simple check can provide irrefutable evidence that you have simulated sufficiently long.

An example of where the exponent is reported is in the following article.

Wang, Q., Keffer, D.J., Petrovan, S., Thomas, J.B., "Molecular Dynamics Simulation of Polyethylene Terephthalate Oligomers",*J. Phys. Chem. B.* **114**(2) 2010 pp. 786–795, doi: 10.1021/jp909762j.

A second example is in Table SI-10 of the supplementary information associated with Xiong et al. from 2010 in *Sensor Actuat B-Chem*.

Numerical values of the self-diffusivity are reported in Table 1. The MSDs are plotted on a log-log scale in Figure 11. The exponents are reported in the legend.

VIII. Temperature Dependence of Self-Diffusion Coefficients

Diffusion, especially in solids, is often considered as an activated process. Determining the diffusion coefficient at several temperatures allows one to construct a conventional Arrhenius plot and extract the activation energy for transport, in exactly the same manner as is done when the diffusion coefficients are measured experimentally.

An example of such an Arrhenius plot appears in Figure 7 of the supplementary information associated with Xiong et al. from 2010 in *Sensor Actuat B-Chem*.

IX. Composition Dependence of Self-Diffusion Coefficients

For this part of the lecture, refer to the following article.

Keffer, D.J., Adhangale, P., "The composition dependence of self and transport diffusivities from molecular dynamics simulations", *Chem. Eng. J.* **100** (1-3) 2004 pp. 51-69, doi: 10.1016/j.cej.2003.11.028

X. Statistically Reliable Fickian Diffusion Coefficients

For this part of the lecture, refer to the following article.

Keffer, D.J., Edwards, B.J., Adhangale, P., "Determination of statistically reliable transport diffusivities from molecular dynamics simulation", *J. Non-Newtonian Fluid Mech.* **120** (1-3) 2004 pp. 41-53, doi: 10.1016/j.jnnfm.2004.01.014.

XI. Deviations from Ordinary Diffusion

XI.A. Single-file Motion

There are some highly confined systems, in which the concept of a self-diffusion coefficient does not apply. For example, particles confined to single-file motion do not give rise to dynamic behavior that can be described by a self-diffusion coefficient. Specifically, in ordinary diffusion the MSD is related to the observation time to the first power. In single-file motion, the MSD is related to the observation time to the one-half power.

For this part of the lecture, refer to the following article.

Keffer, H.T., McCormick, A.V., D., Davis, "Unidirectional and single-file diffusion in $AlPO_4$-5: molecular dynamics investigations", *Mol. Phys.* **87**(2) 1996 pp. 367-387, doi: 10.1080/00268979600100241.

XI.B. Sub-Diffusive Behavior in Intermediate Time-scales of Confined Systems

Confinement certainly restricts diffusion. At very long time scales, confinement can reduce the diffusion coefficient significantly, but the diffusion remains ordinary, that is the MSD is related to the observation time to the first power. However, in intermediate time scales, one may observe sub-diffusive behavior, in which the MSD is related to the observation time to a power less than one. This behavior is real but represents only dynamic behavior corresponding to a limited observation time. If left alone for a sufficiently long time, eventually the diffusive behavior will become ordinary. (Certainly, it may be the case that for a particular application, the time scale of interest is one that corresponds to the non-ordinary, sub-diffusive behavior.

For this part of the lecture, refer to the following article.

Calvo-Muñoz, E.M., Esai Selvan, M., Xiong, R., Ojha, M., Keffer, D.J., Nicholson, D.M., Egami, T., "Applications of a General Random Walk Theory for Confined Diffusion", *Phys. Rev. E* **83**(1) 2011 article # 011120, doi: 10.1103/PhysRevE.83.011120.

XI.C. So-Called Super-Diffusion

From time to time, reports appear in the literature, claiming to observe so-called super-diffusion, where the exponent relating MSD is related to the observation time is greater than 1. Let us remind the student that in ordinary diffusion the MSD is related to the observation time to the first power. In convection, the MSD is related to the observation time to the second power. *To my knowledge, there is no physical mechanism that can generate super-diffusion. Therefore, all reports of super-diffusion are likely erroneous, resulting from a lack of rigor in defining the frame of reference with which diffusion is measured.*

This error is easy to understand, if the mass flux measured contains an element of diffusion and an element of convection, the resulting mean square displacement will fall between 1 and 2. See Section IV above for the derivation of diffusivities with a proper frame of reference. This same section provides an example to compute self-diffusivities relative to center-of-mass motion.

XII. Built in LAMMPS Functionality

LAMMPS has built in MSD and VACF functions. These can be invoked with a combination of compute and fix commands as shown in the `DIFFUSE` subdirectory in the LAMMPS `examples` directory.

There is one important caveat to know about the implementation provided in these examples. *When done this way, the MSD and VACF are not ensemble averages because they only use the starting point of the simulation as a time origin.* In general, any step in the simulation, can be used as a time origin to compute the ensemble average. Therefore, this script throws away a vast majority of statistically significant data. I don't recommend using this. Instead, I recommend using a post-processing script to calculate the diffusivity.

XIII. Calculating the Self-Diffusion Coefficient in LAMMPS

On the examples section of the course website, there is a complete example providing a LAMMPS input file for a binary Lennard-Jones fluid. This example also includes a FORTRAN code to calculate the self-diffusivity of each component in the simulation as well as the exponent relating the MSD to the observation time. Here we mention a few points.

The sample input file generates the trajectory in a custom format designated by the dump command:

```
dump    1       all custom 100 output.lammpstrj id type xu yu zu
```

This format dictates unscaled and unwrapped coordinates. This is the format assumed by the FORTRAN code, selfd_from_trj_v01.f.

After the lammps simulation is run, the self-diffusion code can be compiled with the command

`source script1` or `source ./script1` depending upon your path settings.

This command compiles the code and moves the executable file, a.out, into the same directory as the output from the simulation. Modify input file, selfd_lammps.txt, to match simulation specifications. As downloaded, the lammps input file and the self-diffusion input file are consistent.

Run with command

`a.out` or `./a.out` depending upon your path settings

The output is located in the following files.

- selfd_lammps_001.txt (contains copy of input file)
- selfd_lammps_out_001.txt (contains screen output)
- selfd_lammps_out_Ar_001.txt (contains component 1 self-diffusivities, exponents and MSD)
- selfd_lammps_out_B_001.txt (contains component 2 self-diffusivities, exponents and MSD)

Four values are reported for self-diffusivities, exponents and MSD, for the x, y and z dimensions and for total. The "self-diffusivities" are fit to the portion of the curve specified by the fractions deltminfrac and deltmaxfrac. The lower limit should be set to avoid short times before the long-time limit required by the Einstein relation is reached. The upper limit should be set to avoid long times where the statistical averaging is poor.

The functional form of the regression used to calculate self-diffusivities is

$$y = m \cdot x + b$$

$$MSD = m \cdot \tau + b$$

You should only perform a regression over the linear portion of the curve, which corresponds to the infinite-time limit in which the Einstein relation is valid.

The code returns two sets of exponents called “exponents” and “corrected exponents”. The exponents are calculated to provide a measure of the extent to which the MSDs are in the long-time limit. These exponents should be one, designating ordinary diffusion in the long-time limit.

The functional form of the regression used to calculate "exponents" is

$$MSD = m \cdot \tau^n$$

$$ln(MSD) = ln(m) + n \cdot ln(\tau)$$

The slope of this log-log plot provides the exponent relating the MSD to the observation time. This is an approximation which ignores the y-intercept, *b*, in the self-diffusivity fit.

The "corrected exponents" remove this error. The functional form of the regression used to calculate "corrected exponents" is

$$MSD = b + m \cdot \tau^n$$

$$ln(MSD - b) = ln(m) + n \cdot ln(\tau)$$

The difference between the uncorrected and corrected exponents is significant. For the total self-diffusivity, the uncorrected value of the exponent is 1.108 while the corrected value of the exponent is 1.000.

There are two warning related to this code. First, this is a serial code. For large systems it may take a long time. Second, this code reads in all frames to compute the auto-correlation to its maximum value. For large systems, it may run out of memory.

Also included in the example zip file is an excel file that generates the same (to machine precision) self-diffusivity and exponents for component 1 as the FORTRAN code.

Several plots from that Excel file are shown below.

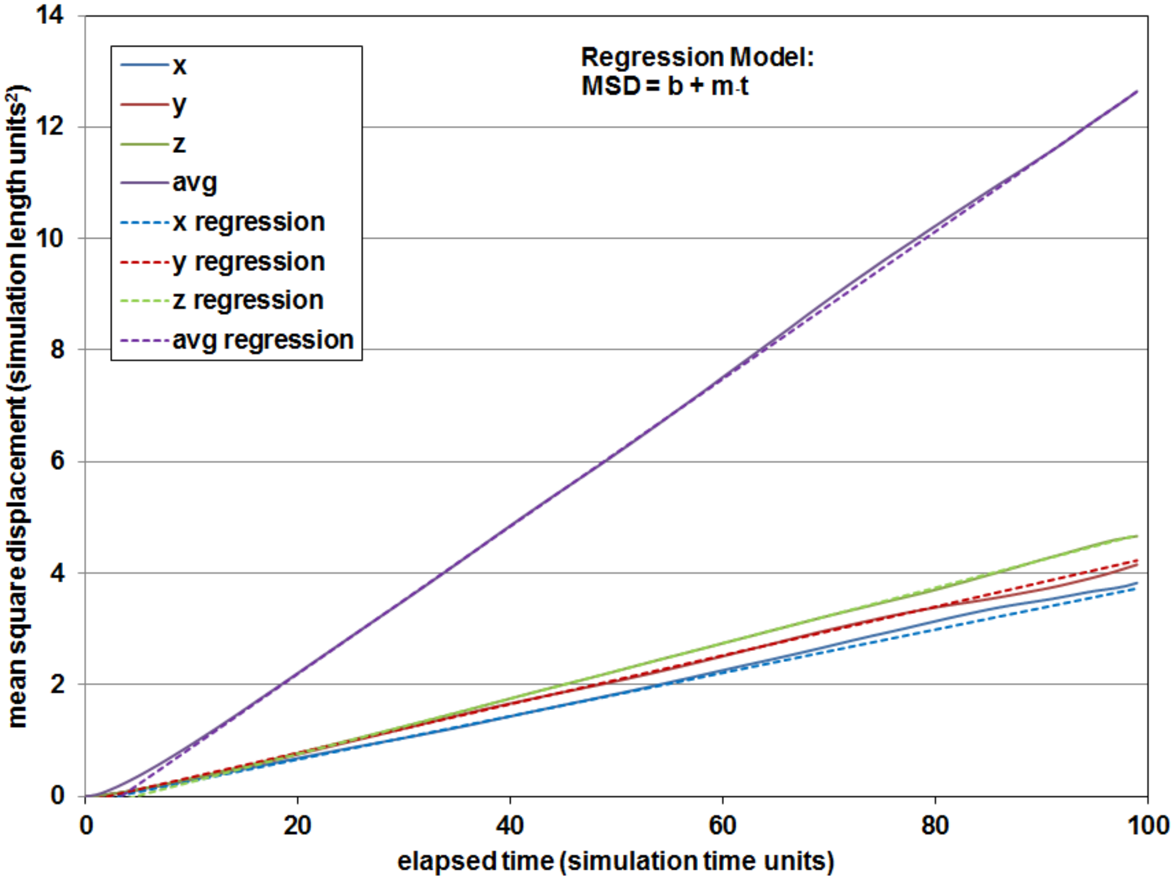

Figure 1. Mean square displacement vs. observation time from a LAMMPS simulation. The regressions are fit from 0.25 to 0.5 of the simulation duration.

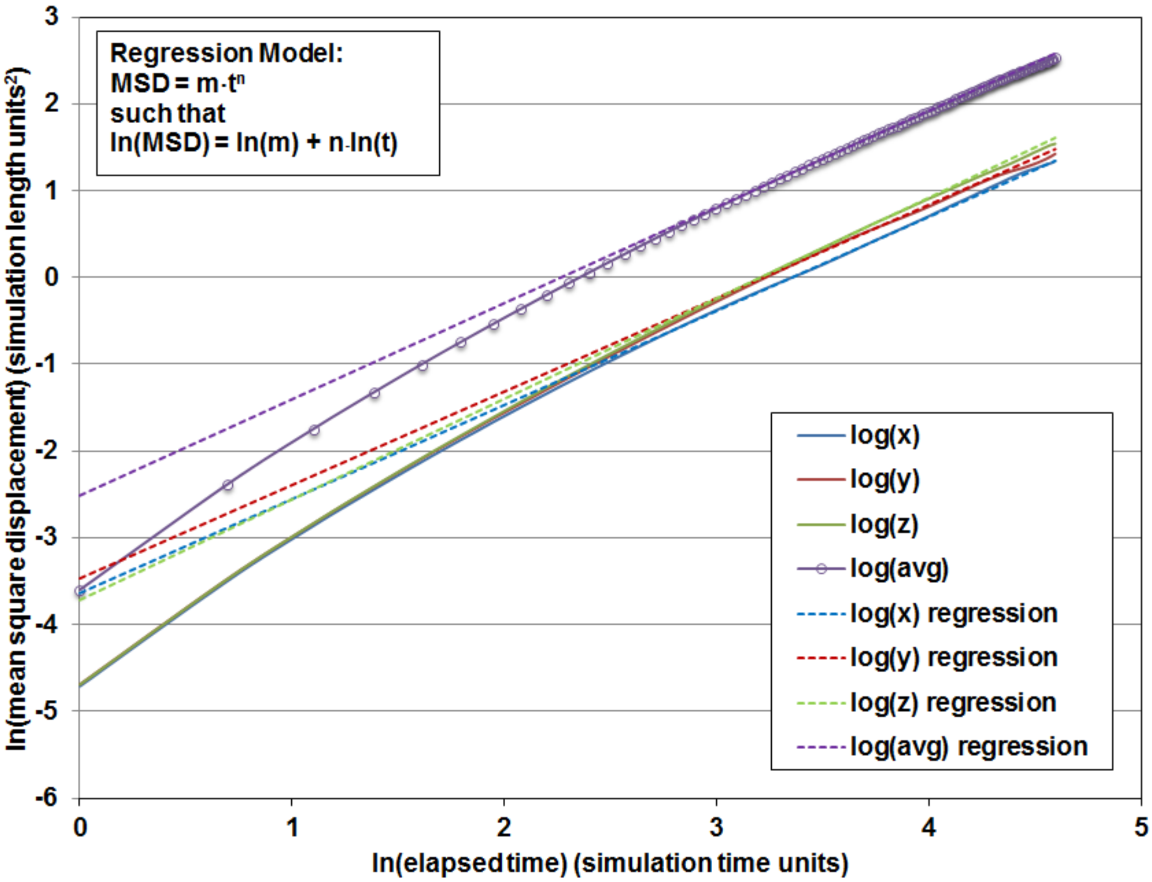

Figure 3. Mean square displacement vs. observation time from a LAMMPS simulation on a log-log plot. The regressions are fit from 0.25 to 0.5 of the simulation duration. On the total MSD, data plots are shown to emphasize that bulk of the data is located to the right on the log-log plot. This curve corresponds to the uncorrected expression, $MSD = b + m \cdot \tau^n$.

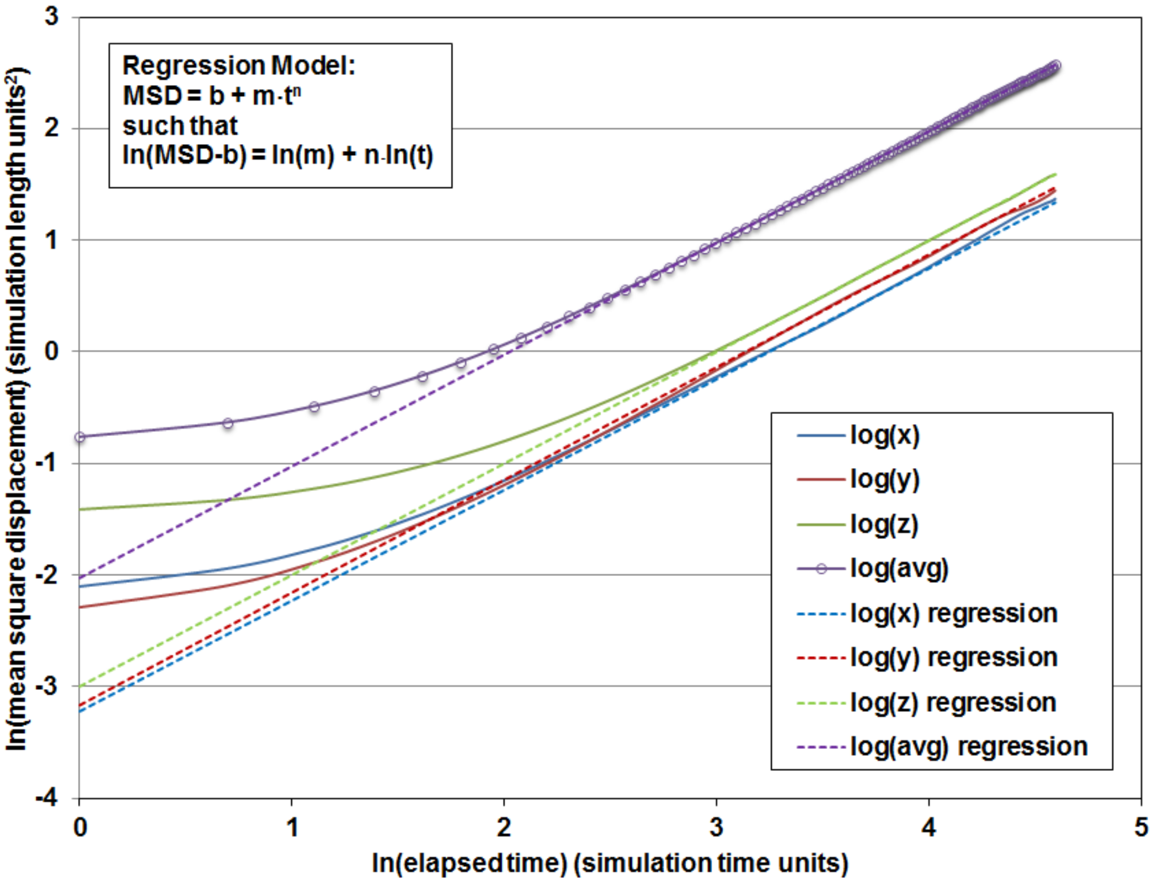

Figure 3. Mean square displacement vs. observation time from a LAMMPS simulation on a log-log plot. The regressions are fit from 0.25 to 0.5 of the simulation duration. On the total MSD, data plots are shown to emphasize that bulk of the data is located to the right on the log-log plot. This curve corresponds to the corrected expression, $MSD = m \cdot \tau^n$.