

# Best Practices for Transport Properties : v1.4

Edward J Maginn<sup>1\*,†§</sup>, Daniel R. Roe<sup>2†¶</sup>, J. Richard Elliott<sup>3†\*\*</sup>, Richard A. Messerly<sup>4†††</sup>, Daniel Midname Carlson<sup>5†‡‡</sup>

<sup>1</sup>The University of Notre Dame; <sup>2</sup>Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health; <sup>3</sup>The University of Akron;

<sup>4</sup>Thermodynamics Research Center, National Institute of Standards and Technology;

<sup>5</sup>Chemical Engineering Department, Brigham Young University

*This LiveCoMS document is maintained online on GitHub at <https://github.com/ejmaginn/TransportCheckList>; to provide feedback, suggestions, or help improve it, please visit the GitHub repository and participate via the issue tracker.*

*Contribution of the National Institute of Standards and Technology, not subject to US copyright.*

*This version dated April 16, 2018*

## Abstract

The ability to predict transport properties (i.e. diffusivity, viscosity, conductivity) is one of the primary benefits of molecular simulation. Although most studies focus on the accuracy of the simulation output compared to experimental data, such a comparison primarily tests the adequacy of the force field (i.e. the model). By contrast, the reliability of different simulation methodologies for predicting transport properties is the focus of this manuscript. Unfortunately, obtaining reproducible estimates of transport properties from molecular simulation is not as straightforward as static properties. Therefore, this manuscript discusses the best practices that should be followed to ensure that the simulation output is reliable, i.e. is a valid representation of the force field implemented.

There are two classes by which transport properties are predicted: equilibrium molecular dynamics (EMD) and non-equilibrium molecular dynamics (NEMD). This manuscript presents the best practices for EMD, leaving NEMD for a future publication. As self-diffusivity and shear viscosity are the most prevalent transport properties found in the literature, the discussion will also be limited to these properties with the expectation that future publications will discuss best practices for thermal conductivity, ionic conductivity, and transport diffusivity.

## \*For correspondence:

[Maginn's email](#) (EM); [Roe's email](#) (DR); [Elliott's email](#) (JRE); [richard.messerly@nist.gov](mailto:richard.messerly@nist.gov) (RAM); [Carlson's email](#) (DC)

<sup>†</sup>These authors contributed equally to this work

<sup>‡</sup>These authors also contributed to this work

## 1 Introduction

Transport properties describe the rates at which mass, momentum, heat or charge move through a given substance. They involve mean squared displacements (MSDs) of molecules as the system evolves dynamically. In general, these prop-

erties can be computed by equilibrium molecular dynamics (EMD) or by non-equilibrium molecular dynamics (NEMD) methods. The EMD methods involve post-processing of a standard molecular dynamics (MD) trajectory while NEMD methods require modifications of the underlying equations of motion and/or boundary conditions of the system. Many

codes such as LAMMPS ([13]) and GROMACS ([10]) have analysis tools that automatically estimate transport properties from an EMD or NEMD simulation, but there are often insufficient checks as to whether the actual underlying simulations are adequate for making these estimates. For this reason, following best practices is imperative to ensure that meaningful predictions are obtained. The purpose of this document is to improve the quality of published results and to reduce the time required for a novice in the field to obtain meaningful and reliable results.

In addition to the present manuscript, we highly recommend reviewing this list of existing resources:

1. Text books:
  - (a) Ref. [5], pages 58-64, 204-208, and 240-256
  - (b) Ref. [9], pages 87-90 and 509-523
  - (c) Ref. [14], pages 374-382
2. Class notes
  - (a) Ref. [1]
  - (b) Ref. [4]
  - (c) Ref. [3]
  - (d) Ref. [2]
3. Published articles
  - (a) Ref. [11]
  - (b) Ref. [7]
  - (c) Ref. [22]
  - (d) Ref. [19], pages 13139-13140
4. Software manuals
  - (a) Ref. [10]
  - (b) Ref. [13]

Most text books and class notes provide a thorough discussion of EMD/NEMD theory with little discussion of practical considerations. Review articles tend to focus on the numerical advantages and disadvantages of different methods but assume that the reader already understands the subtleties of implementing each method. Furthermore, although software manuals describe some of the theory and implementation of these methods in their respective environments, the documentation is typically insufficient for someone not familiar with best practices for estimating transport properties. This document supplements the existing literature by providing a succinct checklist and common pitfalls.

## 2 Equilibrium Molecular Dynamics (EMD) for Estimating Transport Properties

It is most convenient to consider compiling the transport properties as an implicit part of any equilibrium MD simulation. The added computational overhead is relatively small,

especially for the self-diffusivity. The main caveat is that longer simulations than normal may be required to achieve reasonable averages.

The general formula for computing a transport property via an EMD simulation is given as

$$\gamma = \int_0^\infty dt \langle \dot{\xi}(t) \dot{\xi}(0) \rangle \quad (1)$$

where  $\gamma$  is the transport coefficient (within a multiplicative constant) and  $\xi$  is the perturbation in the Hamiltonian associated with the particular transport property under consideration and  $\dot{\xi}$  signifies a time derivative. Integrals of the form given by Equation 1 are known as “Green-Kubo” integrals. It is easy to show that an integrated form of Equation 1 results in an equivalent expression for  $\gamma$  known as the “Einstein” formula

$$\gamma = \lim_{t \rightarrow \infty} \frac{\langle (\xi(t) - \xi(0))^2 \rangle}{2t} = \frac{1}{2} \lim_{t \rightarrow \infty} \frac{d}{dt} \langle (\xi(t) - \xi(0))^2 \rangle \quad (2)$$

where the derivative form is often preferred.

For self-diffusivity,  $\xi$  is the Cartesian atom position and the time correlation function,  $\dot{\xi}$ , in Equation 1 is of the molecular velocities. For the shear viscosity, the integral in Equation 1 is of the time correlation of the off-diagonal elements of the stress tensor. For the thermal conductivity the integral is over the energy current, and for the electrical conductivity the integral is over the electric current. Table 1 provides the relevant equations for self-diffusivity ( $D$ ) and shear viscosity ( $\eta$ ), as these properties are the focus of this work.

Although both Equation 1 (Green-Kubo) and Equation 2 (Einstein) are theoretically rigorous, in practice one method is often preferred depending on the property being estimated. In the case of self-diffusivity, we recommend the Einstein (MSD) approach. By contrast, for shear viscosity we typically recommend Green-Kubo, although for some systems the Einstein approach may be preferable. As the simulation set-up and computational cost are essentially the same for the Green-Kubo and Einstein approaches, the primary difference is the post-simulation data analysis required. Precision and reproducibility of the estimated value are key factors for selecting between the Green-Kubo or Einstein methods. For this reason, we emphasize the importance of proper and clearly communicated data analysis and rigorous uncertainty quantification.

## 3 Checklist

This section provides an overview of the checklist items for each property ( $D$  and  $\eta$ ) and method (Green-Kubo and Einstein). Detailed discussions for each checklist item are found in Sections 4-6.

## CHECKLIST FOR COMPUTING SELF-DIFFUSIVITY WITH EINSTEIN EQUILIBRIUM APPROACH

- ☐ **Simulation set-up.** No amount of data analysis can compensate for a poorly designed experiment. It is imperative that the simulation sufficiently samples the relevant region of phase space.
  - ☐ Sample from the correct ensemble. See Sec. 4.1.1.
  - ☐ Increase the information extracted from simulation results.
    - ☐ Perform multiple replicate simulations. See Sec. 4.1.2.
    - ☐ Ensure that simulations are sufficiently long. See Sec. 5.2.2.
    - ☐ Increase the output frequency. See Sec. 5.2.1.
  - ☐ Check for system size effects. See Sec. 5.1.2
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of transport properties.
  - ☐ Improve precision by averaging over:
    - ☐  $N$  molecules. See Sec. 4.2.1.
    - ☐ Three dimensions (xx, yy, zz). See Sec. 4.2.1 and 5.1.3.
    - ☐ Multiple replicate simulations. See Sec. 4.1.2.
  - ☐ Clearly communicate how  $D$  is obtained from Equation 2. See Secs. 4.2.2 and 5.2.3.
  - ☐ Report the uncertainty in  $D$ :
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis, i.e. variation in  $D$  with respect to the time cut-off, etc. See Sec. 5.2.3.
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Special topics.** Check if your system of interest requires some special considerations. See Sec. 5.4.

## CHECKLIST FOR COMPUTING SELF-DIFFUSIVITY WITH GREEN-KUBO EQUILIBRIUM APPROACH

- ☐ **Simulation set-up.** No amount of data analysis can compensate for a poorly designed experiment. It is imperative that the simulation sufficiently samples the relevant region of phase space.
  - ☐ Sample from the correct ensemble. See Sec. 4.1.1.
  - ☐ Increase the information extracted from simulation results.
    - ☐ Perform multiple replicate simulations. See Sec. 4.1.2.
    - ☐ Ensure that simulations are sufficiently long. See Sec. 5.3.2
    - ☐ Increase the output frequency. See Sec. 5.3.1.
  - ☐ Check for system size effects. See Sec. 5.1.2.
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of transport properties.
  - ☐ Improve precision by averaging over:
    - ☐  $N$  molecules. See Sec. 4.2.1.
    - ☐ Three dimensions (xx, yy, zz). See Sec. 4.2.1 and 5.1.3.
    - ☐ Multiple replicate simulations. See Sec. 4.1.2.
  - ☐ Clearly communicate how  $D$  is obtained from Equation 1. See Secs. 4.2.2 and 5.3.3.
  - ☐ Report the uncertainty in  $D$ :
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis, i.e. variation in  $D$  with respect to the time cut-off, etc. See Sec. 5.3.3.
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Special topics.** Check if your system of interest requires unique considerations. See Sec. 5.4.

## CHECKLIST FOR COMPUTING VISCOSITY WITH GREEN-KUBO EQUILIBRIUM APPROACH

- ☐ **Simulation set-up.** No amount of data analysis can compensate for a poorly designed experiment. It is imperative that the simulation sufficiently samples the relevant region of phase space.
  - ☐ Sample from the correct ensemble. See Sec. 4.1.1.
  - ☐ Increase the information extracted from simulation results.
    - ☐ Perform multiple replicate simulations. See Sec. 4.1.2.
    - ☐ Ensure that simulations are sufficiently long. See Sec. 6.1.1.
    - ☐ Increase the output frequency. See Sec. 6.1.2.
  - ☐ Check for system size effects. See Sec. 6.1.3.
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of transport properties.
  - ☐ Improve precision by averaging over multiple:
    - ☐ Pressure tensor elements (three off-diagonal or all six). See Sec. 4.2.1 and 5.1.3.
    - ☐ Replicate simulations. See Sec. 4.1.2 and 6.1.4.
  - ☐ Clearly communicate how  $\eta$  is obtained from Equation 1. See Secs. 4.2.2 and 6.2.1.
  - ☐ Report the uncertainty in  $\eta$ :
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis, i.e. variation in  $\eta$  with respect to the time cut-off, fitting model, etc. See Sec. 6.2.1.
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Special topics.** Check if your system of interest requires some special considerations. See Sec. 6.4.

## CHECKLIST FOR COMPUTING VISCOSITY WITH EINSTEIN EQUILIBRIUM APPROACH

- ☐ **Simulation set-up.** No amount of data analysis can compensate for a poorly designed experiment. It is imperative that the simulation sufficiently samples the relevant region of phase space.
  - ☐ Sample from the correct ensemble. See Sec. 4.1.1.
  - ☐ Increase the information extracted from simulation results.
    - ☐ Perform multiple replicate simulations. See Sec. 4.1.2.
    - ☐ Ensure that simulations are sufficiently long. See Sec. 6.1.1.
    - ☐ Increase the output frequency. See Sec. 6.1.2.
  - ☐ Check for system size effects. See Sec. 6.1.3.
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of transport properties.
  - ☐ Improve precision by averaging over multiple:
    - ☐ Pressure tensor elements (three off-diagonal or all six). See Sec. 4.2.1 and 5.1.3.
    - ☐ Replicate simulations. See Sec. 4.1.2 and 6.1.4.
  - ☐ Clearly communicate how  $\eta$  is obtained from Equation 2. See Secs. 4.2.2 and 6.3.1.
  - ☐ Report the uncertainty in  $\eta$ :
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis, i.e. variation in  $\eta$  with respect to the time cut-off, fitting model, etc. See Sec. 6.3.1.
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Special topics.** Check if your system of interest requires some special considerations. See Sec. 6.4.

**Table 1.** Equilibrium molecular dynamics equations.

Property	$\gamma$	$\xi$	Green-Kubo (Equation 1)	Einstein (Equation 2)
Self-diffusivity	$D$	$r$	$\frac{1}{3} \int_0^\infty dt \langle \frac{1}{N} \sum_{i=1}^N v_{\alpha,i}(t) v_{\alpha,i}(0) \rangle$	$\frac{1}{6} \lim_{t \rightarrow \infty} \frac{d}{dt} \langle \frac{1}{N} \sum_{i=1}^N  r_i(t) - r_i(0) ^2 \rangle$
Shear viscosity	$\eta$	$r_\alpha v_\beta$	$\frac{V}{k_B T} \int_0^\infty dt \langle P_{\alpha,\beta}(t) P_{\alpha,\beta}(0) \rangle$	$\frac{V}{2k_B T} \lim_{t \rightarrow \infty} \frac{d}{dt} \langle (P_{\alpha,\beta}(t) - P_{\alpha,\beta}(0))^2 \rangle$
$P_{\alpha,\beta}(t) = \frac{1}{V} \sum_{i=1}^N (m v_{\alpha,i}(t) v_{\beta,i}(t) + r_{\alpha,i}(t) f_{\beta,i}(t))$ , $\alpha \neq \beta$ $\alpha, \beta = x, y$ , or $z$ Cartesian coordinates				

## 4 General transport checklist items

### 4.1 General transport: Simulation set-up

#### 4.1.1 Correct Ensemble

For a liquid solution, it is safest to run in the microcanonical (NVE, constant number of molecules, volume, energy) ensemble. However, it is most common to desire  $D$  and  $\eta$  at a specified temperature ( $T$ ) and pressure ( $P$ ). This requires performing a series of simulations in different ensembles:

1. NPT ensemble at desired  $T$  and  $P$  until equilibrium is well sampled
2. NVT ensemble where the volume is set such that the density is the average density computed from the NPT run
3. NVE ensemble where the final configuration of the NVT run is used as the initial configuration

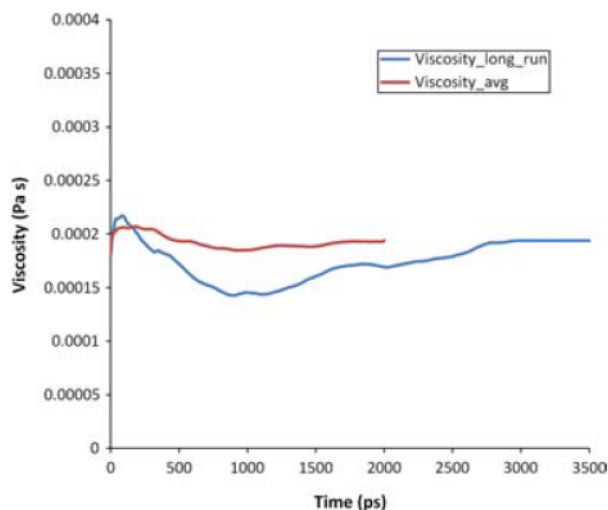
The average pressure and temperature for the NVE production run are computed and should be close to (but not exactly the same) as the input  $P$  and  $T$  to the original NPT run. These average pressures and temperatures must be reported along with the self-diffusivity and viscosity.

Note that, although the best practice is to use the NVE ensemble (Steps 1-3), it is common to see values reported using the NPT (just Step 1) or NVT (Steps 1-2) ensemble. We strongly discourage the use of the NPT ensemble alone. By contrast, the NVT ensemble has been implemented successfully and is quite common, especially for viscosity. For example, Fanourgakis et al. reported that the NVT and NVE ensembles provide nearly identical results for viscosity [8]. Therefore, we recommend using either the NVT or NVE ensemble with NVE being preferred.

#### 4.1.2 Replicate simulations

To smooth noise in Green-Kubo integral or Einstein slope, we recommend performing independent replicate trajectories (i.e. different initial configurations or random seed to initialize velocities). The primary advantage of performing replicates as opposed to one longer simulation is the computational speed-up. Figure 1, borrowed from Ref. [20], demonstrates that an average of 10 replicate simulations of 2 ns length converges

to the same value as a single 4 ns simulation. Since these replicates can be performed in parallel the computational time is reduced by a factor of two, in this example.



**Figure 1.** Green-Kubo viscosity plot. Copied from Figure 2 of Ref. [20]. Red curve represents the average viscosity over 10 independent 2 ns trajectories whereas the blue curve is obtained from a single 4 ns simulation. For further details, see Ref. [20].

Also, the uncertainty is inversely proportional to the square root of the number of replicates (see Figure 7 of Ref. [25] and Figure 8 of Ref. [16]), increasing the number of replicates is a simple, fast, and direct way to reduce the uncertainty. For example, note in Figure 1 the fluctuations in  $\eta$  are much smaller for the average of 10 replicates compared to that of a single longer simulation. As fluctuations in  $\eta$  are typically much larger than  $D$ , more replicate simulations are required for estimating viscosity (see Sec. 6.1.4).

In addition, replicate simulations are useful if a single simulation does not adequately sample phase space, i.e. is trapped in a local minimum or has slow dynamics. Furthermore, replicates can provide rigorous estimates of uncertainty (see Sec. 4.2.3).

Note that, although the best practice is to start each independent replicate at the NPT step, it is common to use the same density (NVT step) for each replicate. This approach is

acceptable assuming that the author provide the corresponding uncertainty in  $P$  (see Sec. 4.1).

## 4.2 General transport: Post-simulation analysis

### 4.2.1 Improved precision

In practice, several tricks-of-the-trade are employed to reduce fluctuations and, thereby, the standard deviation. For self-diffusivity, it is a standard practice to average the mean-square-displacement or velocity autocorrelation function over all  $N$  molecules (see Table 1). For shear viscosity, it is not possible to average over the number of particles because viscosity is a collective property that depends on the pressure/stress tensor of the system. For this reason, diffusivity estimates are much more precise than viscosity estimates, and additional tactics are typically employed to improve the viscosity precision, namely, large amounts of replicate simulations.

For self-diffusivity, it is also standard practice to average the  $x$ ,  $y$ , and  $z$  displacements or velocities. For viscosity, the recommended practice is to use multiple components from the pressure/stress tensor. For example, although early studies only implemented a single off-diagonal component (typically  $xy$ ), the common practice in recent studies is to use all three off-diagonal ( $xy$ ,  $yz$ ,  $zx$ ) and sometimes three additional modified diagonal terms of the pressure/stress tensor (see Sec. 6.1.4).

Finally, for both self-diffusivity and shear viscosity it is common to average over multiple time origins ( $t_0$ ). It is important that the difference between subsequent  $t_0$  values ( $\delta t_0$ ) be longer than the correlation time so that the different time intervals are independent.

### 4.2.2 Clear communication

Transport properties are estimated by numerical integration of Equation 1 or calculating the slope of Equation 2 with respect to time. Both methods involve some judgment on the part of the user and results can vary depending on where the slope is taken (Einstein approach) and for how long the integral is carried out (Green-Kubo approach). Some recent work has suggested some guidelines for how to compute an objective estimate of the viscosity using the Green-Kubo approach [25]. Similar methods for estimating other transport properties from Equations 1 or 2 should be possible to develop. As no single best practice can be recommended for the region over which the slope or integral is calculated, it is important to justify how this decision was made. Furthermore, it is critical to quantify the degree of variability in the estimated property that arises from varying the time interval included in the data analysis.

### 4.2.3 Uncertainty quantification

Replicates can provide a rigorous uncertainty assessment. We recommend bootstrapping the uncertainties by randomly sampling which replicates are included in the data analysis procedure:

1. Randomly select (with replacement) a set of replicate simulations
2. Calculate the relevant average quantity from this random set, i.e.  $\langle \dot{\xi}(t)\dot{\xi}(0) \rangle$  for Green-Kubo or  $\langle (\xi(t) - \xi(0))^2 \rangle$  for Einstein
3. Compute transport property ( $\gamma$ ) from Equations 1 or 2
4. Repeat steps 1-3 thousands of times
5. Generate distribution of the estimated values for  $D$  or  $\eta$
6. Compute uncertainty by integrating distribution at desired confidence level

## 4.3 General transport: Common pitfalls

When simulating in the NVE ensemble, it is imperative that the integrator conserve energy. The most common method to check for energy conservation is to systematically adjust the time step. We also recommend checking the constraint tolerances.

An important implicit assumption in Equations 1 and 2 is that the time over which these expressions are evaluated is much larger than the correlation time of the variable  $\xi$ . This assumption is often satisfied easily for simple liquids, where relaxation times are fast, but becomes problematical for systems with sluggish dynamics. In such cases, we recommend that RECOMMENDATION.

Obtaining reliable results with reasonable uncertainties can require simulations that are much longer than the longest relaxation times in the system, which are often unknown at the start of a simulation. Therefore, insufficient simulation time is a common pitfall in estimating transport properties. To avoid this pitfall, we recommend performing a series of progressively longer simulations to determine if the estimated values deviate significantly with increasing simulation time.

## 5 Self-Diffusivity

We recommend the Einstein approach for computing self-diffusivity as it is robust and the most commonly used method. However, we also recommend validating that the Green-Kubo method provides similar estimates. Although systematic deviations are often observed between the two methods, if the analysis is done properly the values should agree within their statistical uncertainties [12, 15, 17]. Section 5.1 discusses self-diffusivity checklist items that apply to both the Einstein and Green-Kubo approaches. Sections 5.2 and 5.3 discuss checklist items that are specific to either the Einstein or Green-Kubo approaches, respectively, for estimating the self-diffusivity



constant. Section 5.4 provides a brief discussion of some topics that are relevant in certain applications.

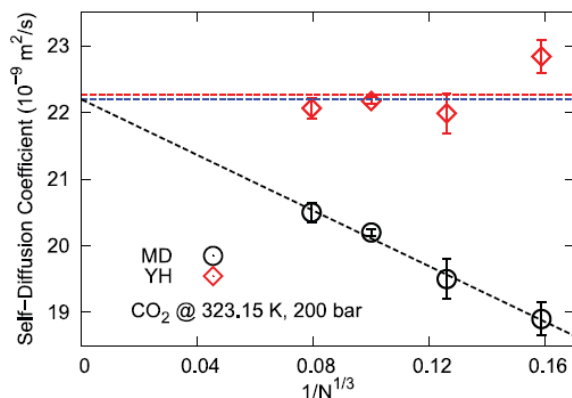
## 5.1 Self-Diffusivity: General

### 5.1.1 Unwrapped coordinates

Use “unwrapped coordinates” of molecule center of mass to determine mean squared displacement; can also track all atomic coordinates and ensure consistency with center of mass.

### 5.1.2 Finite size effects

Finite size effects tend to be significant and need to be accounted for, i.e. system size corrections must be applied [18, 24]. For example, see Figure 2 from Ref. [18]. Some correction approaches require that the viscosity be calculated first or that multiple simulations are run with varying box sizes / number of molecules in order to estimate the infinite size limit of the self-diffusivity.



**Figure 2.** System size dependence of self-diffusivity obtained with Einstein approach. Copied from Figure 1 of Ref. [18]. Blue dashed lines are obtained by extrapolating the MD results to the infinite system size, i.e.  $N^{-1/3} \rightarrow 0$ . Red diamonds are the values of  $D$  after correcting for finite size effects. The red dashed line is an average of these corrected values of  $D$ . For further details, see Ref. [18].

### 5.1.3 Improved precision

Compute the diffusion coefficient separately in each dimension, i.e.  $D_{xx}$ ,  $D_{yy}$ , and  $D_{zz}$ . For a homogeneous system,  $D_{xx}$ ,  $D_{yy}$ , and  $D_{zz}$  should be equal and provides a useful validation of simulation quality. The variation in these values can be used for a rough estimate of the statistical uncertainty, although more rigorous methods for uncertainty estimation are recommended (see Sec. 4.2.3).

## 5.2 Self-Diffusivity: Einstein

### 5.2.1 Output frequency

Output frequency of positions should be sufficient to have around 1000 data points over the entire MSD.

### 5.2.2 Simulation length

Simulation length needed depends on number of molecules for which transport properties are desired. Fewer molecules requires more simulation time and vice versa. Regardless, the simulation must be long enough so that the molecules are in the diffusive regime. We recommend computing the slope from a log-log plot of MSD with respect to time, which should be approximately 1 in the diffusive regime (see Figure 3). Another heuristic is whether the MSD is sufficiently large, i.e. larger than the square of the radius of gyration of the molecule at the low end and larger than the square of half the box length at the high end.

### 5.2.3 Data analysis

In order to obtain reliable estimates of  $D$ , it is important to consider how the linear regression is performed for the MSD with respect to time (Equation 2). Specifically, the time interval that is included in the regression can have a significant impact on the predicted value of  $D$ . We recommend that only the “middle” of the MSD be used in the fit. Short time must be excluded as it follows a ballistic trajectory, while very long time is excluded due to the increased noise. Currently, we are unaware of an objective approach for defining the “middle” region. Until such an approach exists, we recommend that the author reports how the region was selected and how much variability in  $D$  can be attributed to the choice of this region. In addition, the uncertainty in the fit of the slope should be reported. A typical plot, borrowed from Ref. [12], is provided in Figure 3, where the linear regressions at long time are included.

## 5.3 Self-Diffusivity: Green-Kubo

### 5.3.1 Output frequency

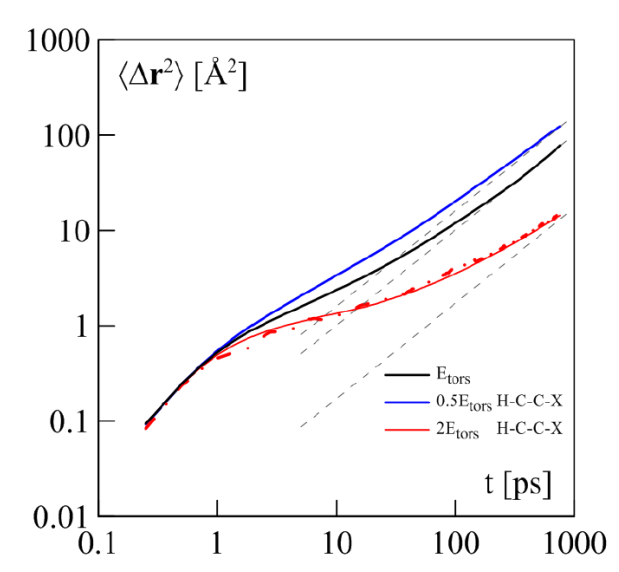
Need to write velocities instead of positions, and the frequency should be much higher because the integral of the velocity autocorrelation function (VACF) decays rapidly. Recommend writing every 5 fs.

### 5.3.2 Simulation length

Simulations should be long enough that the Green-Kubo integral has reached a plateau. Note that the plateau time is not the same as the required simulation time, since multiple time origins ( $t_0$ ) are used to compute the Green-Kubo integral.

### 5.3.3 Data analysis

Integrate the VACF numerically, providing details of how this is done.



**Figure 3.** Log-log plot of MSD with respect to time. Copied from Figure 2 of Ref. [12]. The gray dashed lines are the long-range asymptotes of the MSD, as determined by the authors. For further details, see Ref. [12].

Plot the running integral vs time. The data are best at short time and noise takes over at long times. Like with the MSD, a cut-off needs to be determined when you decide the integral has converged. It is important to report how sensitive the estimate is to this cut-off time.

#### 5.4 Self-Diffusivity: Special topics

For systems that require anisotropic pressure control (e.g. membranes, etc.), use of a barostat/thermostat that maintains the correct isothermal/isobaric ensemble (e.g. extended system, Langevin piston) is required.

Calculating diffusion in membrane systems with periodic boundary conditions require some additional consideration, e.g. Saffman-Delbruck model [6, 23].

The standard non-bonded long-range cut-off corrections are not straightforward problematic when computing diffusivity in a heterogeneous system.

## 6 Viscosity

Although the popularity of NEMD methods for predicting viscosity has increased in recent years, Ref. [7] demonstrate that EMD methods can be of equal accuracy and reliability to NEMD as long as best practices are followed, i.e. proper system set-up and thorough data analysis. That being said, EMD works best for fluids with relatively low viscosity, i.e. typically less than 20 cP although EMD has been successfully implemented for systems near 50 cP. Higher viscosity systems are extremely difficult to compute with EMD and so NEMD methods are often preferred in this case.

The recommended EMD approach for predicting viscosity is Green-Kubo. We should note that Hess claims that the Einstein relation is more convenient than Green-Kubo for viscosity because “inaccuracies in the long time correlations can be ignored by only considering integral over shorter times.” However, several advances have been implemented with the Green-Kubo approach since 2002 (when Ref. [11] was published). The Green-Kubo approach now appears to be the most popular EMD method found in the literature. More importantly, less arbitrary data analysis methods exist that improve the reliability and reproducibility (see Sec. 6.2.1). Section 6.1 discusses self-diffusivity checklist items that apply to both the Einstein and Green-Kubo approaches. Sections 6.2 and 6.3 discuss checklist items specific to the Green-Kubo and Einstein approaches, respectively, for estimating viscosity. Section 6.4 provides a brief discussion of some topics that are relevant in certain applications.

### 6.1 Viscosity: General

#### 6.1.1 Simulation length

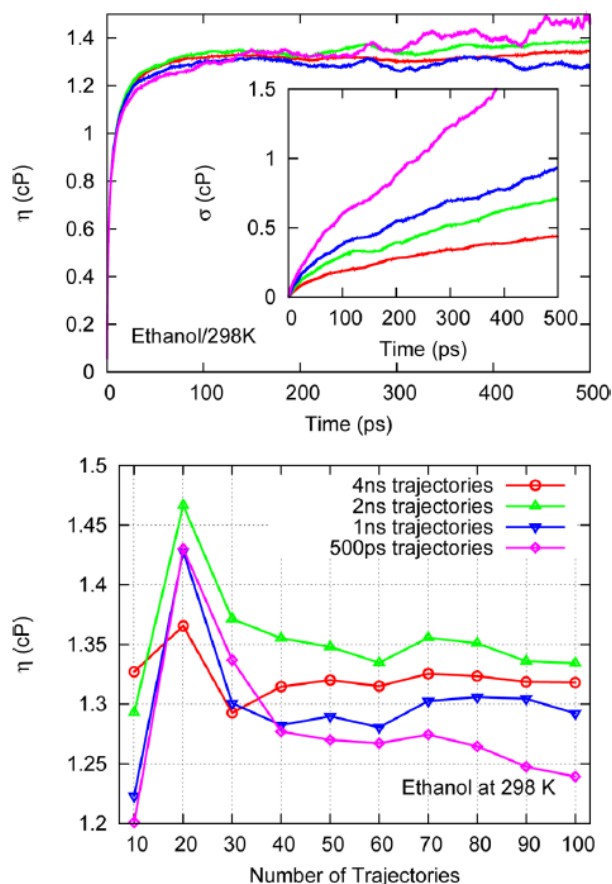
Overall you need about 10X more data to compute viscosity than diffusivity, since viscosity is a collective property. Also requires sufficient simulation time for “4-5 molecular rotations” on average.

Figure 4, borrowed from Ref. [25], demonstrates that if the length of each independent trajectory is too short the viscosity will not converge to the correct value, regardless of how many replicates are used. Specifically, the average viscosity obtained from 100 replicates of 500 ps appears to diverge from the 1, 2, and 4 ns simulation results, suggesting that 500 ps is not sufficiently long for this system.

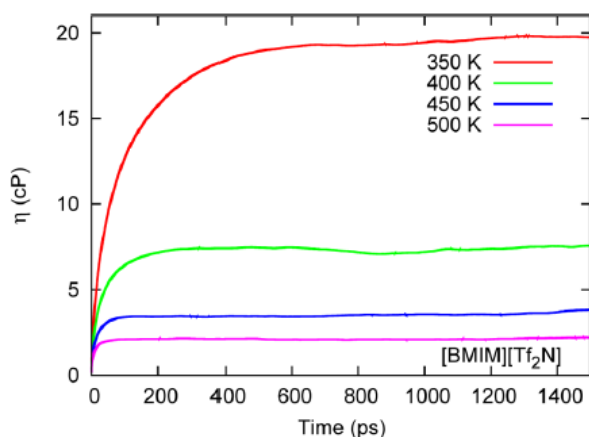
It is important not to confuse the Green-Kubo integration time (the abscissa for the top panel of Figure 4) with the simulation length (the different color lines in both panels of Figure 4). Recall that the Green-Kubo integral (plotted in the top panel) is evaluated using multiple time origins ( $t_0$ ), so the Green-Kubo integral contains more independent trajectories for the 4 ns line than the 500 ps line. Therefore, the time at which the Green-Kubo integral reaches a plateau (around 100 ps in the top panel of Figure 4) is not the same as the required simulation time. Although for sufficient independent trajectories, the required simulation time should typically be around an order of magnitude greater than the plateau time.

Figure 5, borrowed from Ref. [25], demonstrates that the plateau time increases with increasing viscosity, where an order of magnitude increase in viscosity corresponds to approximately an order of magnitude increase in the plateau time. In order to account for the increase in the plateau time, higher viscosity fluids require longer overall simulation times.





**Figure 4.** Viscosity dependence on simulation length. Copied from Figure 8 of Ref. [25]. Different lines and symbols correspond to different simulation length, i.e. trajectory time. For further details, see Ref. [25].



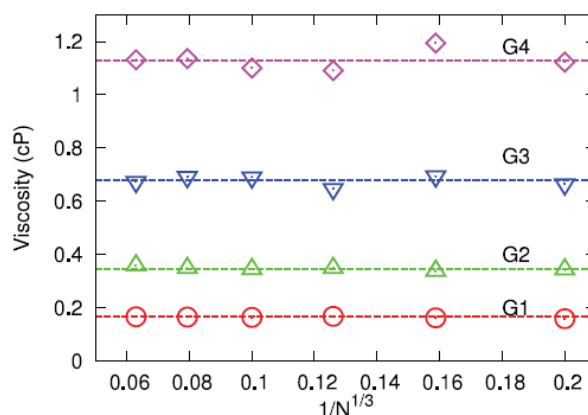
**Figure 5.** Plateau time dependence on viscosity. Copied from Figure 10 of Ref. [25]. Different lines correspond to different temperatures and, thus, different viscosities. For further details, see Ref. [25].

### 6.1.2 Output frequency

Output frequency should be high (every 5-10 fs); this needs to be checked for the particular system.

### 6.1.3 Finite size effects

Figures 6-7 from Refs. [18] and [25], respectively, suggest that finite size effects are not significant for systems with as few as 125 and 500 molecules, respectively. More work needs to be done to verify this. We recommend that users look for system size effects by plotting the viscosity with respect to  $N^{-1/3}$ , where  $N$  is the number of molecules. The range of  $N$  should span an order of magnitude or, if this is computationally intractable, at least a factor of two. The author should report any dependence observed for viscosity with respect to system size. If a linear trend is observed with respect to  $N^{-1/3}$ , the infinite system size viscosity can be extrapolated as the intercept from a linear regression. The author should report the uncertainty associated with this linear fit and extrapolation.

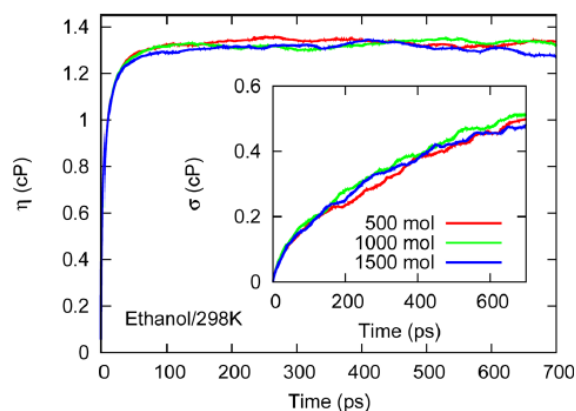


**Figure 6.** Finite size effects for viscosity obtained with Green-Kubo approach. Copied from Figure 3 of Ref. [18]. Different symbols correspond to different types of glymes (Gi). Dashed lines are average value for each glyme from various system sizes ( $N$ ). For further details, see Ref. [18].

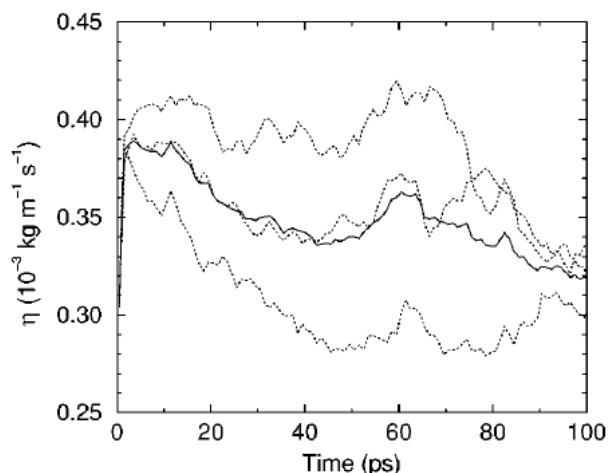
### 6.1.4 Improved precision

To improve statistical averaging, it is common to include multiple terms from the pressure tensor. For example, Figure 8, borrowed from Ref. [11], demonstrates the improvement of averaging the three off-diagonal elements of the pressure tensor, compared to a single off-diagonal element. We recommend using all six of the symmetrized traceless stress tensor terms. Although we are not aware of any studies that rigorously quantify the improved precision for using all six terms, Figure 9, borrowed from Ref. [7], demonstrates that the average viscosity is nearly identical when using the three off-diagonal terms or when using six terms.

The key to improving precision of viscosity estimates is to perform several replicate simulations. The number of replicates used in literature varies widely. [20] somewhat arbitrarily used 10 replicates whereas [25] performed a systematic



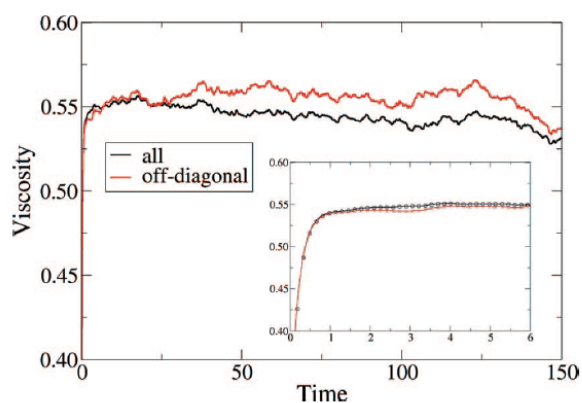
**Figure 7.** Finite size effects for viscosity obtained with Green-Kubo approach. Copied from Figure 9 of Ref. [25]. Different colors correspond to different number of molecules. For further details, see Ref. [25].



**Figure 8.** Green-Kubo viscosity plot. Copied from Figure 5 of Ref. [11]. Dashed lines represent a single off-diagonal element of the pressure tensor while solid line is the average of the three off-diagonal elements. For further details, see Ref. [11].

investigation of the minimal number of replicates required for convergence. They observed that a value of 30-40 replicates was statistically equivalent to 100 replicates for their system. However, the necessary number of replicates depends on the system. Specifically, the compound, the temperature, the number of molecules, and the simulation time all influence the optimal number of replicates. We recommend that the author plot how  $\eta$  varies with respect to the number of replicates for a range of 10-30 replicates, to determine if additional simulations are needed.

## 6.2 Viscosity: Green-Kubo



**Figure 9.** Green-Kubo viscosity plot. Copied from Figure 1 of Ref. [7]. Red line is obtained by averaging the three off-diagonal elements while the black line is obtained from all six pressure tensor elements. For further details, see Ref. [7].

### 6.2.1 Data analysis

It is imperative to report how the viscosity was estimated from Equation 1. There are three common methods: average over specified time interval, fit autocorrelation function to a model, or fit “running integral” to a model. We recommend the latter methodology but discuss each below.

A slightly ambiguous but common practice is to report an average that is obtained over a specified time interval. Due to large fluctuations at long times, the initial plateau at short times (around 10–100 ps) is typically the region of choice, see Refs. [7, 8]. However, it is important to explain how this time interval was selected (i.e. visual inspection, test of convergence, magnitude of fluctuations, etc.) and to quantify how much the estimated viscosity changes if the time interval were modified.

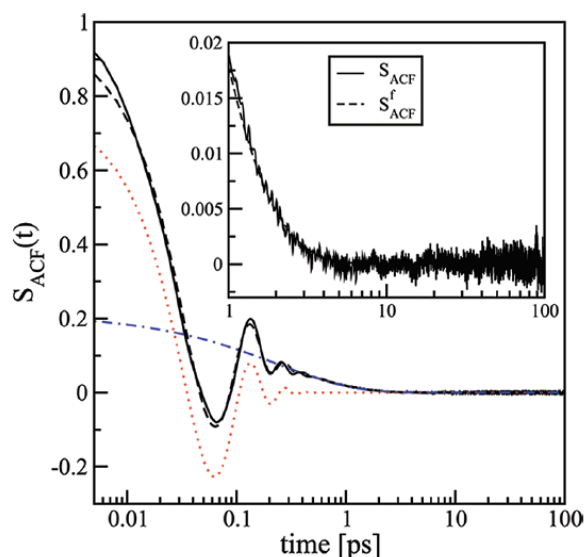
An alternative method is to fit a model to the autocorrelation function before calculating the “running integral.” The integral of the model fit can then be evaluated in the limit as  $t \rightarrow \infty$ . This helps to overcome large fluctuations at long times and, thereby, reduces uncertainties. The primary difficulty is finding a model that can adequately match the autocorrelation function without introducing bias into the estimate of viscosity. A common function found in the literature is

$$f(t)/f(0) = (1 - C)\cos(\omega t) \exp(-t/\tau_f)^{\beta_f} + C \exp(-t/\tau_s)^{\beta_s} \quad (3)$$

where  $C, \omega, \tau_f, \tau_s, \beta_f, \beta_s$  (and sometimes  $f(0)$ ) are fitting parameters.  $\omega$  is the frequency of rapid pressure oscillations,  $\tau_f$  and  $\beta_f$  are the time constant and exponent of fast relaxation in a stretched-exponential approximation,  $\tau_s$  and  $\beta_s$  are constants for slow relaxation,  $C$  is the pre-factor that determines the weight between fast and slow relaxation,  $f(t)$  is the autocorrelation function at time  $t$ , and  $f(0)$  is the initial (time-zero)

autocorrelation function [10].

Figure 10, from Ref. [8], demonstrates that Equation 3 can reliably fit the autocorrelation function for this system. However, small deviations in the model fit can lead to significant bias in the estimated viscosity. Similar to the methods discussed previously, it is important to quantify the variability in viscosity that arises from the model fit. For example, we recommend bootstrapping the uncertainties of the model fit. Furthermore, if a weighting function or cut-off time is implemented the impact of these parameters should be discussed.



**Figure 10.** Fit of autocorrelation function to Equation 3. Copied from Figure 1 of Ref. [8].  $S_{ACF}$  and  $S_{ACF}^f$  correspond to the raw autocorrelation function and the fit to Equation 3, respectively. The red dotted line and blue dashed-dotted line correspond to the fast and slow autocorrelation components, respectively, i.e. the first and second terms of Equation 3. For further details, see Ref. [8].

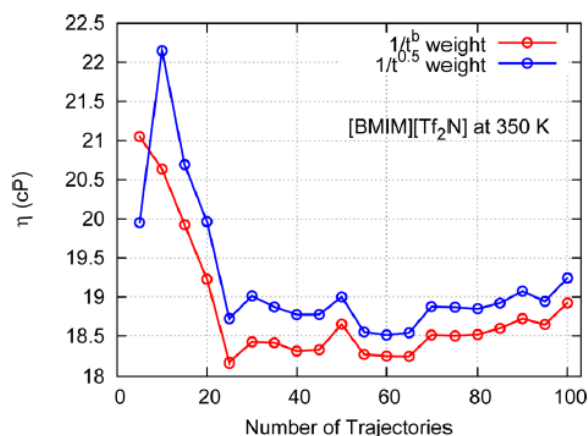
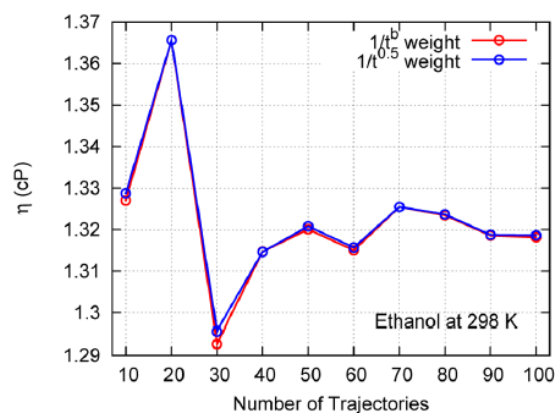
The method we recommend is to fit an analytic function directly to the “running integral”. For example, Refs. [21] and [25] recommended fitting the “running integral” to a double-exponential function

$$\eta(t) = A\alpha\tau_1(1 - \exp(-t/\tau_1)) + A(1 - \alpha)\tau_2(1 - \exp(-t/\tau_2)) \quad (4)$$

where  $A$ ,  $\alpha$ ,  $\tau_1$ , and  $\tau_2$  are fitting parameters. The primary advantage over the previous approach is that uncertainties in the model fit do not propagate through the integration. It is important to include a description of how the fit is performed, i.e. the objective function, weighting model, range of data included, etc.

Ref. [25] recommends that the data be weighted by the inverse of the standard deviation with respect to time. They fit the standard deviation to a weighting model of the form  $w \propto t^{-b}$ , where  $w$  is the weight,  $t$  is the time, and  $b$  is the weighting exponent. If such a model is utilized, the resulting estimate

of  $\eta$  may depend strongly on  $b$ , the weighting exponent. For example, Figure 11, borrowed from Ref. [25], compares  $\eta$  for two different values of  $b$  in the weighting model, namely, when  $b$  is a predetermined value of 0.5 and when  $b$  is fit to the standard deviation in the replicate averages. Note that Ref. [21] recommended a value of  $b = 2$ . Ref. [25] demonstrated that for  $b = 2$  the estimated value of  $\eta$  for [BMIM][Tf<sub>2</sub>N] at 350 K is approximately 11 cP (compared to  $\approx 19$  cP in the bottom panel of Figure 11). For these reasons, we recommend that the author quantifies the uncertainty in the estimated viscosity due to the value of  $b$ . Propagating the uncertainty in  $\eta$  from  $b$  can be accomplished by implementing a two-step bootstrap method. First, a distribution of  $b$  values are obtained by bootstrapping the fit of the weighting model to the standard deviation. Second, a distribution of  $\eta$  values are computed by fitting Equation 4 with each value of  $b$  from the distribution generated in the previous step.



**Figure 11.** Viscosity dependence on the exponent of the weighting model,  $b$ .  $b = 0.52$  for Ethanol at 298K, top panel, while  $b$  is between 0.60–0.73 for [BMIM][Tf<sub>2</sub>N] at 350 K, bottom panel. For further details, see Ref. [25].

Ref. [25] also suggests that to improve the fit a cut-off time be implemented. They provide a heuristic that the cut-off

time correspond to when the standard deviation is 40% the plateau value. Regardless of how the cut-off is determined, it is important to quantify the degree to which the estimated viscosity depends on this parameter. For example, Zhang et al. reported that the viscosity decreased by 0.8% and 6.1% when using a cut-off time corresponding to a standard deviation of 30% or 20% the plateau value, respectively. However, the magnitude of variability depends strongly on the system. We recommend that the author quantify the cut-off time dependence.

## 6.3 Viscosity: Einstein

### 6.3.1 Data analysis

Similar to the Einstein approach for self-diffusivity, in general, the initial time should be discarded. Since the Einstein relation is valid in the limit of infinite time, it is common to fit the slope at long time. However, it is also common to fit the slope to an intermediate time interval. We recommend that the author explain why the slope was calculated using a given time interval and how much variability is introduced if a different region is selected. For example, a useful measure of uncertainty would be the range of  $\eta$  values obtained by analyzing numerous (order of 100) different time regions.

Since the viscosity is estimated from the slope of the Einstein integral, the average of replicates can be performed in one of two ways. The first option is to calculate the viscosity (i.e. the slope) with respect to time for each replicate and then average the replicate viscosities. However, this approach results in large fluctuations and, therefore, large uncertainties.

The second, and recommended, method is to average the Einstein integral of the multiple replicates. The resulting Einstein integral is often linear over a large time interval if sufficient replicates are used. Subsequently, the slope is determined from this average Einstein integral. Fortunately, with sufficient replicate simulations the slope tends to be fairly constant over intermediate and long time intervals. The number of replicates needed has not been rigorously investigated as it has for the Green-Kubo approach. For this reason, we recommend creating a plot of viscosity with respect to number of replicates (see Figure 4) to determine when sufficient replicates have been simulated. It is our experience that the necessary number of replicates is similar to that for Green-Kubo. As recommended for Green-Kubo, we also recommend bootstrapping the uncertainty. This is done by randomly sampling which replicates are included in the average Einstein integral, calculating the viscosity from the slope, and producing a distribution of these viscosity values from thousands of different random sets of replicates.

## 6.4 Viscosity: Special topics

The GROMACS manual reports that viscosity “is very dependent on the treatment of the electrostatics. Using a (short) cut-off results in large noise on the off-diagonal pressure elements, which can increase the calculated viscosity by an order of magnitude.” [10, 11]

## 7 Conclusions

## 8 Acknowledgments

Funder and other information can be given here.

## References

- [1] Athanassios Panagiotopoulos, The University of Princeton Class Notes;. Accessed: 2018-01-01. <http://paros.princeton.edu/cbe422/md2.pdf>; <http://paros.princeton.edu/cbe520/Transport.pdf>.
- [2] David Kofke, The University of Buffalo Class Notes;. Accessed: 2018-01-01. <http://www.eng.buffalo.edu/kofke/ce530/Lectures/Lecture12.ppt.pdf>.
- [3] Ed Maginn, The University of Notre Dame Class Notes;. Accessed: 2018-01-01. <https://www3.nd.edu/~ed/notes.pdf>.
- [4] Scott Shell, The University of California Santa Barbara Class Notes;. Accessed: 2018-01-01. [https://engineering.ucsb.edu/shell/che210d/Computing\\_properties.pdf](https://engineering.ucsb.edu/shell/che210d/Computing_properties.pdf).
- [5] Allen MP, Tildesley DJ. Computer simulation of liquids. Oxford England New York: Clarendon Press ; Oxford University Press; 1987.
- [6] Camley BA, Lerner MG, Pastor RW, Brown FLH. Strong influence of periodic boundary conditions on lateral diffusion in lipid bilayer membranes. The Journal of Chemical Physics. 2015; 143(24):243113. <https://doi.org/10.1063/1.4932980>, doi: 10.1063/1.4932980.
- [7] Chen T, Smit B, Bell AT. Are pressure fluctuation-based equilibrium methods really worse than nonequilibrium methods for calculating viscosities? The Journal of Chemical Physics. 2009; 131(24):246101. <https://doi.org/10.1063/1.3274802>, doi: 10.1063/1.3274802.
- [8] Fanourgakis GS, Medina JS, Prosmi R. Determining the Bulk Viscosity of Rigid Water Models. The Journal of Physical Chemistry A. 2012; 116(10):2564–2570. <http://dx.doi.org/10.1021/jp211952y>, doi: 10.1021/jp211952y, pMID: 22352421.
- [9] Frenkel D, Smit B. Understanding molecular simulation : from algorithms to applications. 2nd ed. Computational science series, San Diego: Academic Press; 2002.
- [10] GROMACS. GROMACS Reference Manual; 2016.
- [11] Hess B. Determining the shear viscosity of model liquids from molecular dynamics simulations. The Journal of Chemical Physics. 2002; 116(1):209–217. <http://aip.scitation.org/doi/abs/10.1063/1.1421362>, doi: 10.1063/1.1421362.

- [12] **Kondratyuk ND**, Norman GE, Stegailov VV. Rheology of liquid n-triacontane: Molecular dynamics simulation. *Journal of Physics: Conference Series*. 2016; 774(1):012039. <http://stacks.iop.org/1742-6596/774/i=1/a=012039>.
- [13] **LAMMPS**. LAMMPS Users Manual. Sandia National Laboratories; 2017.
- [14] **Leach AR**. Molecular modelling : principles and applications. 2 ed. Pearson Prentice Hall; 2001.
- [15] **Liu H**, Maginn E, Visser AE, Bridges NJ, Fox EB. Thermal and Transport Properties of Six Ionic Liquids: An Experimental and Molecular Dynamics Study. *Industrial & Engineering Chemistry Research*. 2012; 51(21):7242–7254. <http://dx.doi.org/10.1021/ie300222a>, doi: 10.1021/ie300222a.
- [16] **Ma J**, Zhang Z, Xiang Y, Cao F, Sun H. On the prediction of transport properties of ionic liquid using 1-n-butylmethylpyridinium tetrafluoroborate as an example. *Molecular Simulation*. 2017; 43(18):1502–1512. <https://doi.org/10.1080/08927022.2017.1321760>, doi: 10.1080/08927022.2017.1321760.
- [17] **Mondello M**, Grest GS. Viscosity calculations of n-alkanes by equilibrium molecular dynamics. *The Journal of Chemical Physics*. 1997; 106(22):9327–9336. <https://doi.org/10.1063/1.474002>, doi: 10.1063/1.474002.
- [18] **Moultos OA**, Zhang Y, Tsimpanogiannis IN, Economou IG, Maginn EJ. System-size corrections for self-diffusion coefficients calculated from molecular dynamics simulations: The case of CO<sub>2</sub>, n-alkanes, and poly(ethylene glycol) dimethyl ethers. *The Journal of Chemical Physics*. 2016; 145(7):074109. <http://dx.doi.org/10.1063/1.4960776>, doi: 10.1063/1.4960776.
- [19] **Nieto-Draghi C**, Fayet G, Creton B, Rozanska X, Rotureau P, de Hemptinne JC, Ungerer P, Rousseau B, Adamo C. A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes. *Chemical Reviews*. 2015; 115(24):13093–13164. <http://dx.doi.org/10.1021/acs.chemrev.5b00215>, doi: 10.1021/acs.chemrev.5b00215, pMID: 26624238.
- [20] **Payal RS**, Balasubramanian S, Rudra I, Tandon K, Mahlke I, Doyle D, Cracknell R. Shear viscosity of linear alkanes through molecular simulations: quantitative tests for n-decane and n-hexadecane. *Molecular Simulation*. 2012; 38(14-15):1234–1241. <https://doi.org/10.1080/08927022.2012.702423>, doi: 10.1080/08927022.2012.702423.
- [21] **Rey-Castro C**, Vega LF. Transport Properties of the Ionic Liquid 1-Ethyl-3-Methylimidazolium Chloride from Equilibrium Molecular Dynamics Simulation. The Effect of Temperature. *The Journal of Physical Chemistry B*. 2006; 110(29):14426–14435. <http://dx.doi.org/10.1021/jp062885s>, doi: 10.1021/jp062885s, pMID: 16854152.
- [22] **Ungerer P**, Nieto-Draghi C, Rousseau B, Ahunbay G, Lachet V. Molecular simulation of the thermophysical properties of fluids: From understanding toward quantitative predictions. *Journal of Molecular Liquids*. 2007; 134(1):71 – 89. <http://www.sciencedirect.com/science/article/pii/S016773220600331X>, doi: <https://doi.org/10.1016/j.molliq.2006.12.019>, eMLG/JMLG 2005 Special Issue.
- [23] **Venable RM**, Ingólfsson HI, Lerner MG, Perrin BS, Camley BA, Marrink SJ, Brown FLH, Pastor RW. Lipid and Peptide Diffusion in Bilayers: The Saffman–Delbrück Model and Periodic Boundary Conditions. *The Journal of Physical Chemistry B*. 2017; 121(15):3443–3457. <https://doi.org/10.1021/acs.jpcc.6b09111>, doi: 10.1021/acs.jpcc.6b09111, pMID: 27966982.
- [24] **Yeh IC**, Hummer G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *The Journal of Physical Chemistry B*. 2004; 108(40):15873–15879. <http://dx.doi.org/10.1021/jp0477147>, doi: 10.1021/jp0477147.
- [25] **Zhang Y**, Otani A, Maginn EJ. Reliable Viscosity Calculation from Equilibrium Molecular Dynamics Simulations: A Time Decomposition Method. *Journal of Chemical Theory and Computation*. 2015; 11(8):3537–3546. <http://dx.doi.org/10.1021/acs.jctc.5b00351>, doi: 10.1021/acs.jctc.5b00351, pMID: 26574439.