Best Practices for Transport Properties : v1.3

Edward Midname Maginn[1*,†§], **Daniel Midname Roe**[2†¶], **J. Richard Elliott**[3†**], **Richard Alma Messerly**[4†††], **Sunny Midname Hwang**[5†‡‡]

[1]Maginn's institution; [2]Roe's institution; [3]Elliott's institution; [4]National Institute of Standards and Technology; [5]Hwang's institution

***For correspondence:**
Maginn'semail (EM); Roe'semail (DR); Elliott'semail (JRE); richard.messerly@nist.gov (RAM); Hwang'semail (SH)

[†]These authors contributed equally to this work
[‡]These authors also contributed to this work

Present address: [§]Maginn's Department, Institute, Country; [¶]Roe's Department, Institute, Country; [**]Elliott's Department, Institute, Country; [††]Thermodynamics Research Center, NIST, USA; [‡‡]Hwang's Department, Institute, Country

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.

List of people to contact: Peter Cummings, Richard Rowley, Joachim Gross, Raj Khare, Richard Sadus, Ioannis Economou, Jadran Vrabec, Daniel Carlson, Chris Iocavella (any other Richards we can come up with)

Outline
General outline of equilibrium methods of self-diffusivity and viscosity for liquids:

1. Introduction
2. Discussion of different methods within EMD (Green-Kubo, Einstein)
3. General simulation set-up
4. Diffusion
    (a) Brief discussion of why we recommend Einstein over Green-Kubo?
    (b) Simulation setup that is specific to Einstein/diffusion
    (c) Data analysis specific to Einstein/diffusion
    (d) Common pitfalls for Einstein/diffusion
5. Viscosity
    (a) Brief discussion of why we recommend Green-Kubo over Einstein?
    (b) Simulation setup that is specific to Green-Kubo/viscosity

(c) Data analysis specific to Green-Kubo/viscosity
(d) Common pitfalls for Green-Kubo/viscosity

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 properties 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 (***LAMMPS*** (***2017***)) and GROMACS (***GROMACS*** (***2016***)) 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. As Berk Hess has said in a forum post on using the GROMACS "energy -vis" option for calculating viscosity, "...it will give nonsense unless you know exactly what you are doing" ***Hes*** (***????***). 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) ***Allen and Tildesley*** (***1987***), pages 58-64, 204-208, and 240-256
   (b) ***Frenkel and Smit*** (***2002***), pages 87-90 and 509-523
   (c) ***Leach*** (***2001***), pages 374-382
2. Class notes
   (a) http://paros.princeton.edu/cbe422/md2.pdf
   (b) http://paros.princeton.edu/cbe520/Transport.pdf
   (c) https://engineering.ucsb.edu/ shell/che210d/Computing_properties.pdf
   (d) https://www3.nd.edu/ ed/notes.pdf
   (e) http://www.eng.buffalo.edu/ kofke/ce530/Lectures/Lecture12.ppt.pdf
3. Published articles
   (a) ***Hess*** (***2002***)
   (b) ***Chen et al.*** (***2009***)
   (c) ***Ungerer et al.*** (***2007***)
   (d) ***Nieto-Draghi et al.*** (***2015***), pages 13139-13140
4. Software manuals
   (a) ***GROMACS*** (***2016***)
   (b) ***LAMMPS*** (***2017***)

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.

Table 1. Equilibrium molecular dynamics equations.

Property	γ	ξ	Green-Kubo (Equation 1)	Einstein (Equation 2)
Self-diffusivity	D	r	$\frac{1}{3}\int_0^\infty dt \langle \frac{1}{N}\sum_{i=1}^{N} v_{\alpha,i}(t)v_{\alpha,i}(0)\rangle$	$\frac{1}{6}\lim_{t\to\infty}\frac{d}{dt}\langle\frac{1}{N}\sum_{i=1}^{N}\lvert r_i(t)-r_i(0)\rvert^2\rangle$
Shear viscosity	η	$r_\alpha v_\beta$	$\frac{V}{k_bT}\int_0^\infty dt\langle P_{\alpha,\beta}(t)P_{\alpha,\beta}(0)\rangle$	$\frac{V}{2k_bT}\lim_{t\to\infty}\frac{d}{dt}\langle(P_{\alpha,\beta}(t)-P_{\alpha,\beta}(0))^2\rangle$

$P_{\alpha,\beta}(t) = \frac{1}{V}\sum_{i=1}^{N}\left(mv_{\alpha,i}(t)v_{\beta,i}(t)+r_{\alpha,i}(t)f_{\beta,i}(t)\right), \alpha \neq \beta$
α, $\beta = x, y$, or z Cartesian coordinates

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

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

where γ is the transport coefficient (within a multiplicative constant) and ξ 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 γ known as the "Einstein" formula

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

where the derivative form is often preferred.

For self-diffusivity, ξ 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. The relevant equations for self-diffusivity and shear viscosity are provided in Table 1.

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. 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.

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 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. 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 (δt_0) be longer than the correlation time so that the different time intervals are independent.

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 ξ. This assumption is often satisfied easily for simple liquids, where relaxation times are fast, but becomes problematical for systems with sluggish dynamics. 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.

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. 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 data analysis and uncertainty quantification.

Checklist for Equilibrium Einstein Approach for 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 ***Kondratyuk et al.*** (***2016***); ***Liu et al.*** (***2012***); ***Mondello and Grest*** (***1997***).

1. Ensemble - for a liquid solution, it is safest to run in the microcanonical (NVE, constant number of molecules, volume, energy) ensemble.
   (a) Generally you want a self-diffusivity at a specified temperature (T) and pressure (P). This requires performing a series of simulations in different ensembles:
      i. NPT ensemble at desired T and P until equilibrium is well sampled (link to equilibration document)
      ii. NVT ensemble where the volume is set such that the density is the average density computed from the NPT run
      iii. 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. The user should generate multiple starting states that can be used to determine error estimation (see item 10).
   (b) 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.
2. Ensure energy conservation (via adjusting time step, constraint tolerances, etc) (Link to document about initializing NVE in the right "ballpark.")
3. Output frequency should be sufficient to have around 1000 data points over the entire MSD. Typical recommendations are every NEED RECOMMENDATION

4. 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. Finite size effects tend to be significant and need to be accounted for, i.e. system size corrections must be applied. For example, see ***Yeh and Hummer*** (***2004***); ***Moultos et al.*** (***2016***). 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.
6. Multiple time origins used for MSD (block averaging) - is there a best practice?
7. 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 item 10).
8. 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. What about when the middle region is subdiffusive like in Figure 1?
9. 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. One way of checking this is if the slope of a plot of ln(MSD) vs ln (t) = 1. Other heuristics are: is the MSD sufficiently large (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).
10. Compute statistical uncertainty by running independent replicates (using multiple starting states from NPT run at desired temperature) and taking the standard deviation (what do the uncertainty people say?) How many replicates do we recommend? Link to Sampling/Uncertainty doc. Use Zwanzig/Szabo four-time correlation for MSD averaging. . . .
11. Calculating diffusion in membrane systems with PBC require some additional consideration, use of Saffman-Delbruck model: see http://pubs.acs.org/doi/abs/10.1021/acs.jpcb.6b09111, also http://dx.doi.org/10.1063/1.4932980
12. Handling potential truncation: shifted force, shifted potential, cutoff, long range corrections. We need to come up with recommendations on this.

FYI: Richard Elliot I developed a database for self-diffusivity that covered all experimental data from the literature as far as he could find. It is included as supporting information in Ind. Eng. Chem. Res. 2010, 49, 3411–3423. This paper provides a generalized correlation of the quantity rho*D (g/cm-s) of n-alkanes at all molecular weights, temperatures, and densities below the entanglement threshold. This semi-empirical correlation is used as the basis for correlating non-alkanes as well. Accuracy diminished for associating compounds, but experimental data were relatively few in number for associating compounds.

Equilibrium Green-Kubo Approach for Self-Diffusivity

Many of the same best practices for the Einstein approach apply to the Green-Kubo method. Specifically, items 1, 2, 4, 5, 6, 7, and 12 from the previous Section are also applicable to the Green-Kubo approach. Some key differences are:

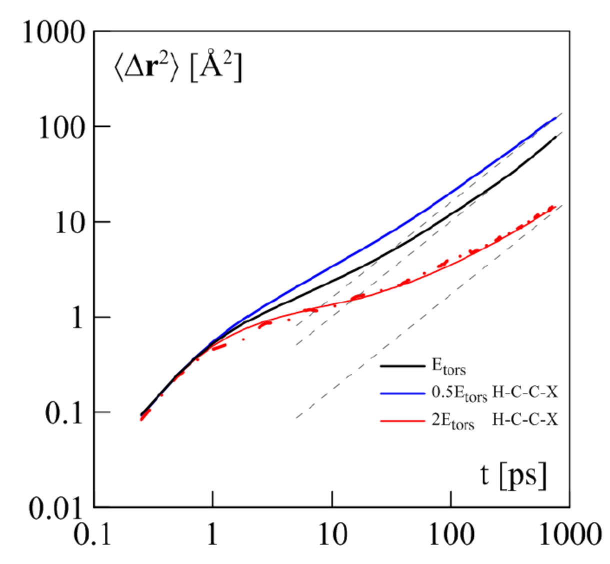

Figure 1

1. Need to write velocities instead of positions, and the frequency should be much higher because the integral of the velocity autocorrelation function decays rapidly. Recommend writing every 5 fs.
2. Integrate the VACF numerically, providing details of how this is done.
3. 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.
4. Independent simulations should be run and each integral can be averaged together to obtain a smoothed integral. With these replicate simulations, we recommend that the uncertainty be determined using a bootstrap methodology:
   (a) Randomly selecting (with replacement) from the set of replicate simulations
   (b) Calculating the average integral from this random set
   (c) Estimating the self-diffusivity as described in item 3
   (d) Repeat steps a)-c) thousands of times
   (e) Generate distribution of the estimated values of self-diffusivity
   (f) Compute uncertainty by integrating distribution at desired confidence level

Checklist for Equilibrium Green-Kubo Approach for Viscosity

Similar to self-diffusivity, EMD for viscosity is straightforward but its reliability compared to experimental data has not been evaluated with a comprehensive database. Many more experimental data are available for viscosity than for self-diffusivity. Anecdotal studies with small databases show encouraging results, but deviations from experiment can range from 5-35% even when results are said to be "good." Nieto-Draghi et al. provide a useful review of the status quo ***Nieto-Draghi et al.*** (***2015***). EMD may deviate 2x more than NEMD from experimental data; hydrogen bonding throws in complications that may require empirical corrections.

We recommend the Green-Kubo approach for predicting viscosity. This 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 reproducability.

Although the popularity of NEMD methods for predicting viscosity has increased in recent years, ***Chen et al.*** (***2009***) demonstrate that EMD methods can be of equal accuracy and reliability to NEMD as long as best practices (i.e. thorough data analysis) are implemented. That being said, EMD works best for fluids with relatively low viscosity (less than 50 cP). Higher viscosity systems are extremely difficult to compute with EMD and so NEMD methods are often preferred in this case.

1. Ensemble: Similar recommendation to that for self-diffusivity, namely, direct simulation of NPT ensemble is not recommended. However, although it is ideal to simulate in the NVE ensemble, NVT has been used with success. For example, Fanourgakis et al. reported that the NVT and NVE ensembles provide nearly identical results ***Fanourgakis et al.*** (***2012***). Therefore, we recommend either the NVT or NVE ensemble with NVE being preferred.
2. 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.
3. Output frequency should be high (every 5-10 fs); this needs to be checked for the particular system.
4. Finite size effects: Figures 2-3 from ***Moultos et al.*** (***2016***) and ***Zhang et al.*** (***2015***), 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.
5. To improve statistical averaging, we recommend using all six of the symmetrized traceless stress tensor terms. Figure 8, taken from ***Chen et al.*** (***2009***), demonstrates that the viscosity is nearly identical when using the three off-diagonal terms or when using six terms.
6. To smooth noise in Green-Kubo integral, 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 5, taken from Payal et al. ***Payal et al.*** (***2012***), 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. However, Figure 4, taken from ***Zhang et al.*** (***2015***), 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.
7. Replicates can provide a more rigorous uncertainty assessment. We recommend bootstrapping the uncertainties by randomly sampling which replicates are included in the average and data analysis procedure (see Green-Kubo for self-diffusivity).
8. The number of replicates used in literature varies widely. ***Payal et al.*** (***2012***) somewhat arbitrarily used 10 replicates whereas ***Zhang et al.*** (***2015***) performed a systematic 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. Furthermore, because the uncertainty is inversely proportional to the square root of the number of replicates (see Figure 7 of ***Zhang et al.*** (***2015***) and Figure 8 of ***Ma et al.*** (***2017***)), increasing the number of replicates is a simple, fast, and direct way to reduce the uncertainty.
9. Report how the viscosity was estimated from the "running integral". There are three common methods:
   (a) 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 ps) is typically the region of choice, see ***Fanourgakis et al.*** (***2012***); ***Chen et al.*** (***2009***). 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 time interval were modified.

(b) 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 $t \to \infty$. 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

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

where $C, \omega, \tau_f, \tau_s, \beta_f, \beta_s$ (and sometimes $f(0)$) are fitting parameters. ω is the frequency of rapid pressure oscillations, τ_f and β_f are the time constant and exponent of fast relaxation in a stretched-exponential approximation, τ_s and β_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 ***GROMACS*** (***2016***). This method has been implemented successfully by ***Fanourgakis et al.*** (***2012***) where Figure 7, from ***Fanourgakis et al.*** (***2012***), 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.

(c) The method we recommend is to fit an analytic function directly to the "running integral". For example, ***Rey-Castro and Vega*** (***2006***) and ***Zhang et al.*** (***2015***) recommended fitting the "running integral" to a double-exponential function

$$\eta(t) = A\alpha\tau_1\left(1-\exp(-t/\tau_1)\right) + A(1-\alpha)\tau_2\left(1-\exp(-t/\tau_2)\right) \tag{4}$$

where A, α, τ_1, and τ_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. ***Zhang et al.*** (***2015***) recommend 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, we recommend that the author quantifies the uncertainty in the estimated viscosity due to the value of b, the weighting exponent. For example, ***Zhang et al.*** (***2015***) compared two different values of b in Figures 7, 12, and 13 (a fixed value of $b = 2$ is used by ***Rey-Castro and Vega*** (***2006***)). ***Zhang et al.*** (***2015***) also suggest that to improve the fit a cut-off time be implemented. They provide a heurestic 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 demonstrate the cut-off time dependence. For example, plots of the estimated viscosity with respect to the weighting exponent and cut-off time, such as those shown in Figure 6, provide a quantitative measure of confidence in the viscosity value.

10. Force fields: systematic consideration of the intra- vs. inter- molecular potential models; UA, vs. AUA vs. EA differences may be significant. For charged systems, polarizable force fields might be needed to get accurate results. Some studies have suggested that united-atom models are not capable of accurately reproducing viscosity and, therefore, anisotropic-united-atom or all-atom models are needed ***Allen and Tildesley*** (***1987***); ***Payal et al.*** (***2012***); ***Mondello and Grest*** (***1997***). However, other studies have shown that with the appropriate tuning of the united-atom force field parameters viscosity can be accurately predicted without significant deprecation of other properties ***Gordon*** (***2006***). ***Ungerer et al.*** (***2007***) discusses different test cases (i.e. state points, compound structures) where united-atom or anisotropic-united-atom models are adequate and inadequate for predicting viscosity.
11. ***GROMACS*** (***2016***) 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."

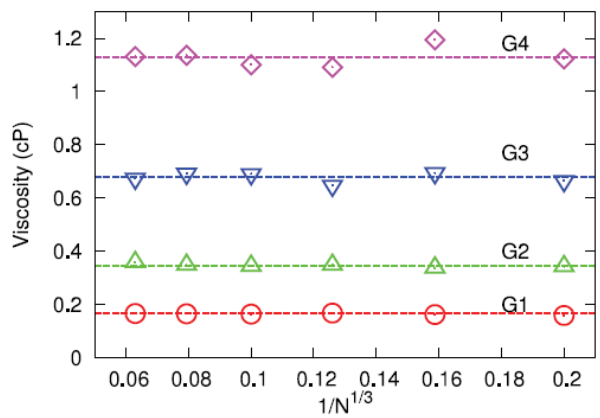

Figure 2

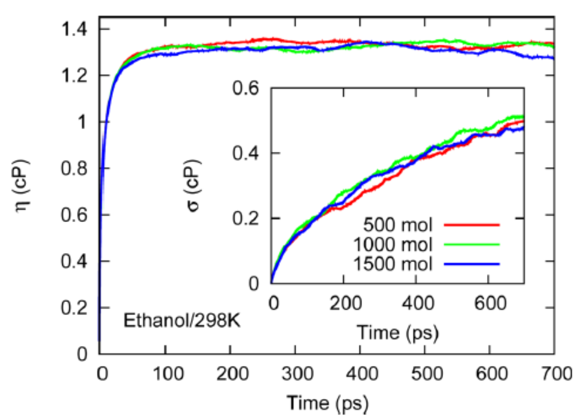

Figure 3

Equilibrium Einstein Approach for Viscosity

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

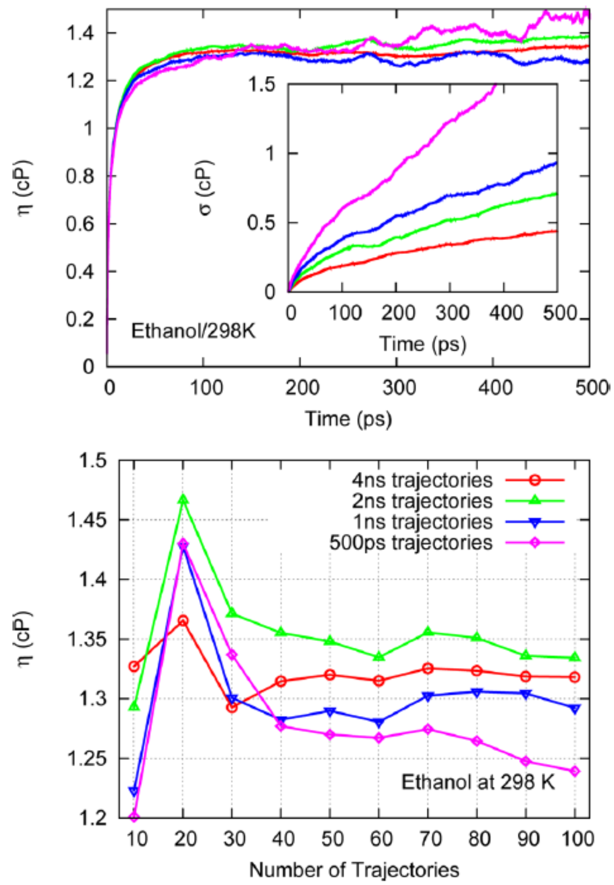

Figure 4

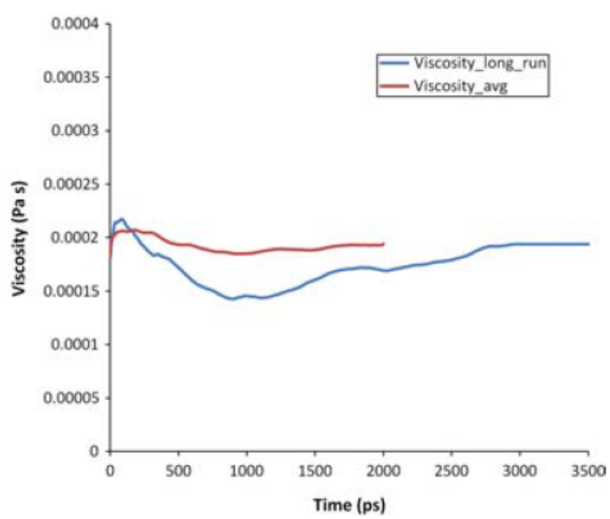

Figure 5

---

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.

Similar to the Green-Kubo approach for viscosity, the key to obtaining precise estimates with the Einstein approach is to average multiple replicate simulations. The viscosity with respect to time is

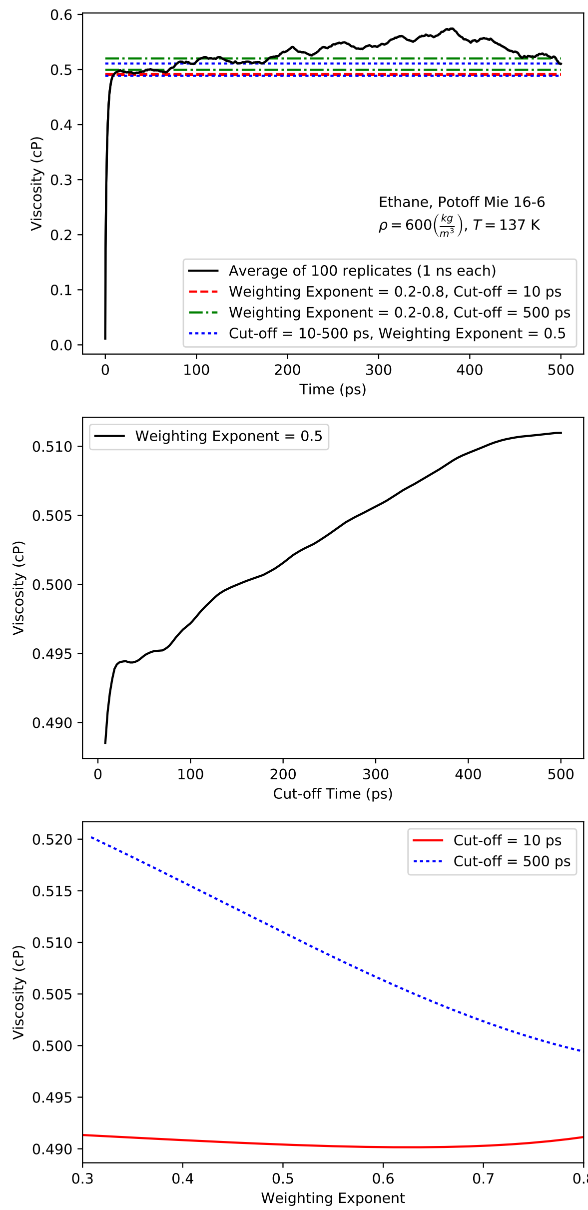

Figure 6. Uncertainty quantification of cut-off time and weighting exponent for method proposed by Zhang et al. Results are shown for a system of 400 united-atom ethane molecules, simulated with the ***Potoff and Bernard-Brunel*** (***2009***) model at saturated liquid density for a temperature of 137 K. Top panel plots the Green-Kubo running integral as average of 100 replicates, each of 1 ns duration. The dashed lines represent different estimates of uncertainty attributed to the cut-off time or the weighting exponent. Middle panel plots the estimated viscosity as a function of cut-off time for a fixed weighting exponent. Bottom panel plots the estimated viscosity as a function of the weighting exponent for two different cut-off times. For longer cut-off times, the viscosity depends strongly on the weighting exponent (around $\pm 2\%$) while shorter times are much less dependent (around $\pm 0.1\%$). For the recommended weighting exponent, the cut-off time can cause the viscosity to vary by around $\pm 2\%$. Note that these results may depend strongly on the state point and number of molecules.

estimated from the slope of the Einstein integral. Thus, 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

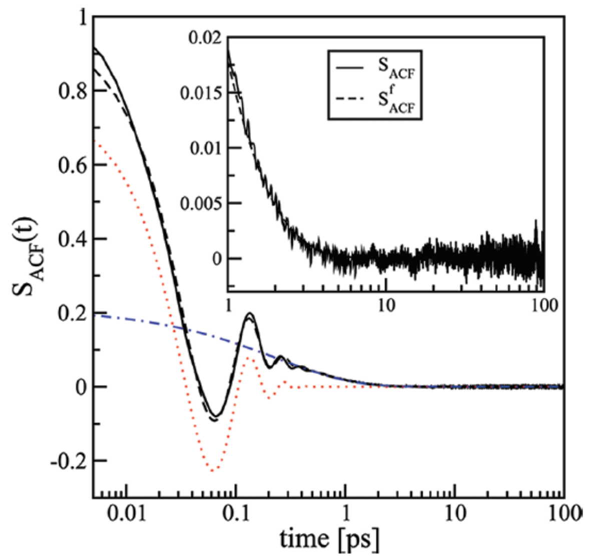

Figure 7

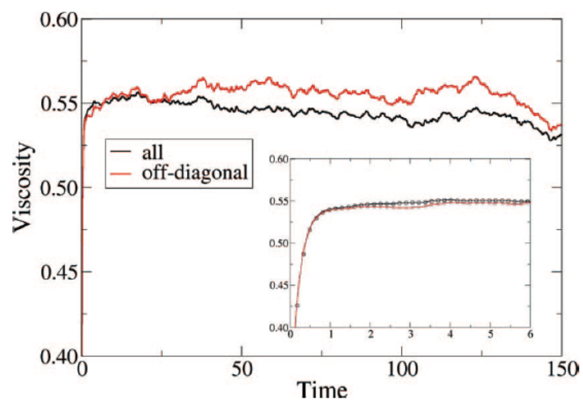

Figure 8

---

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.

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."

Acknowledgments

Funder and other information can be given here.

References

[gmx-users[how to calculate the viscosity of the bulk;. https://mailman-1.sys.kth.se/pipermail/gromacs.org_gmx-users/2009-February/039411.html.

Allen MP, Tildesley DJ. Computer simulation of liquids. Oxford England New York: Clarendon Press ; Oxford University Press; 1987.

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.

Fanourgakis GS, Medina JS, Prosmiti 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.

Frenkel D, Smit B. Understanding molecular simulation : from algorithms to applications. 2nd ed. Computational science series, San Diego: Academic Press; 2002.

Gordon PA. Development of intermolecular potentials for predicting transport properties of hydrocarbons. The Journal of Chemical Physics. 2006; 125(1):014504. https://doi.org/10.1063/1.2208359, doi: 10.1063/1.2208359.

GROMACS. GROMACS Reference Manual; 2016.

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.

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.

LAMMPS. LAMMPS Users Manual. Sandia National Laboratories; 2017.

Leach AR. Molecular modelling : principles and applications. 2 ed. Pearson Prentice Hall; 2001.

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.

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.

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.

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 CO2, 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.

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.

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.

Potoff JJ, Bernard-Brunel DA. Mie Potentials for Phase Equilibria Calculations: Applications to Alkanes and Perfluoroalkanes. Journal of Physical Chemistry B. 2009; 113(44):14725–14731. doi: 10.1021/jp9072137.

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.

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.

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.

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.