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 Please provide an abstract of no more than 250 words. Your abstract should explain the main contributions of your article, and should not contain any material that is not included in the main text.

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

Outline
Separate documents for:

1. Equilibrium methods (self-diffusivity, viscosity) for liquids
2. Non-equilibrium methods (self-diffusivity, viscosity) for liquids
3. Methods for thermal conductivity
4. Methods for ionic conductivity

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-EMD (NEMD) methods. The equilibrium methods involve post-processing

Table 1. Equilibrium molecular dynamics equations.

Property	γ	ξ	Equation 1	Equation 2
Self-diffusion	D	x	$\int_0^\infty dt\langle v_{x,i}(t)v_{x,i}(0)\rangle$	$\frac{1}{2t}\langle(x_i(t)-x_i(0))^2\rangle$
Shear viscosity	η	xv_y	$\beta V \int_0^\infty dt\langle P_{x,y}(t)P_{x,y}(0)\rangle$	$\frac{m^2\beta}{2t}\langle(x_i(t)v_{y,i}(t)-x_i(0)v_{y,i}(0))^2\rangle$
Thermal conductivity	λ	xE	$\frac{k\beta^2}{V}\int_0^\infty dt\langle S_x(t)S_x(0)\rangle$	$\frac{k\beta^2}{2Vt}\left\langle\left(x_i(t)(E_i(t)-\langle E\rangle)-x_i(0)(E_i(0)-\langle E\rangle)\right)^2\right\rangle$

$$P_{x,y}(t)=\frac{1}{V}\sum_{i=1}^N\left(\frac{P_{x,i}(t)P_{y,i}(t)}{m}+x_i(t)f_{y,i}(t)\right)$$
$$S_x(t)=\frac{d}{dt}\sum_{i=1}^N\left(x_i(t)(E_i(t)-\langle E\rangle)\right).$$

of a standard molecular dynamics trajectory while NEMD methods require modifications of the underlying equations of motion and/or boundary conditions of the system. Many codes such as LAMMPS and GROMACS 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 a related topic, "…it will give nonsense unless you know exactly what you are doing".

Equilibrium Molecular Dynamics (EMD) Routes to 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 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 "Einstein" formula for the transport coefficient. Thus an equivalent expression for γ is

$$\gamma = \frac{\langle(\xi(t)-\xi(0))^2\rangle}{2t} \tag{2}$$

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 [Could cite Allen and Tildesley I suppose].

An important implicit assumption in the above equations 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.

Transport properties have to be estimated from the long-time integral of Equation 1 or the slope of Equation 2. Both methods involve some "judgement" on the part of the user and results can vary depending on where the slope is taken and for how long the integral is carried out. Some recent work has suggested some guidelines for how to compute an objective estimate of the viscosity using Equation 1. Similar approaches for estimating other transport properties from Equations 1 or

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.

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 ($\propto \gamma$). In the case of self-diffusivity, we recommend the Einstein (MSD) approach. By contrast, for viscosity we typically recommend Green-Kubo, although for some systems the Einstein approach may be preferable.

Checklist for Equilibrium Einstein Approach for Self-Diffusivity

The Einstein approach is the most commonly used method for computing self-diffusivity. It is robust and we recommend this as the best approach.

1. Ensemble - for a liquid solution, it is safest to run in the microcanonical (NVE) ensemble.

   (a) Generally you want a self-diffusivity at a specified T and P. One needs to then equilibrate a system by first running an appropriate NPT simulation until equilibrium is well sampled (link to equilibration document). Using the average density computed from the NPT run, set the volume and run another NVT equilibration run. The last configuration from this run can then be used as input to start the NVE production run. The average pressure and temperature will need to be computed and should be close to (but not exactly the same) as the input 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 below).

   (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. System size correction must be applied / finite size effects accounted for. See for example *Yeh and Hummer* (*2004*); *Moultos et al.* (*2016*). This requires that the viscosity be calculated first or multiple simulations run with varying box sizes / number of molecules in order to estimate the infinite size limit of the self-diffusivity.

3. Simulation length needed depends on number of molecules for which transport properties are desired. Fewer molecules = 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).

4. Ensure energy conservation (via adjusting time step, constraint tolerances, etc) (Link to document about initializing NVE in the right "ballpark.")

5. Output frequency - how frequently do you write coordinates? Should be frequently enough so that you have around 1000 data points over the entire MSD.

6. 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.

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. 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. How do you fit the MSD (what time interval do you use?) Short time is ballistic trajectory; very long time you get noise, so you need to fit to the "middle" of the MSD. We need to define protocols for how to objectively define this. You should compute the uncertainty in the fit of the slope. Report how the line was fit and associated variables. Is there literature on this? We need to come up with a recommendation for how to do this objectively and consistently.

9. Handling potential truncation: shifted force, shifted potential, cutoff, long range corrections. We need to come up with recommendations on this.
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. Multiple time origins used for MSD (block averaging) - is there a best practice?
12. 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

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 things apply as with the Einstein approach. Key differences:

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 "cutoff" needs to be determined when you decide the integral has converged. Need objective measures for determining this.
4. Averaging: independent simulations should be run, and each integral can be averaged together to obtain a smoothed integral.

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." Chem. Rev. 2015, 115, 13093-13164 provides a useful review of the status quo. EMD may deviate 2x more than NEMD from experimental data; hydrogen bonding throws in complications that may require empirical corrections.

1. Ensemble: NPT is not recommended. NVE is ideal; NVT has been used with success.
2. Finite size effects: one paper suggests it is not significant above a threshold size. More work needs to be done to verify this. Recommend that users look for system size effects and report whether they observe it or not.

3. Simulation length: overall you need about 10X more data to compute viscosity than diffusivity, since viscosity is a collective property.
4. Output frequency should be high (every 5-10 fs); this needs to be checked for the particular system
5. GK method works for fluids with relatively low viscosity (less than 50 cP). Higher viscosity systems are extremely difficult to compute with GK.
6. Sufficient simulation time for "4-5 molecular rotations" on average. Averaging over multiple simulations with analytic fitting of integral provides a good way of smoothing noise and provides an objective means of determining the viscosity (footnote 1).
7. 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.

Equilibrium Einstein Approach for Viscosity

Non-Equilibrium Molecular Dynamics (EMD) Routes to Transport Properties

NEMD methods require the use of a modified Hamiltonian to drive the system away from equilibrium. By monitoring the response of the system in the limit of a small perturbation, the transport coefficient associated with the perturbation can be calculated [I would cite Evans and Morriss here]. The basic idea behind the technique is that a system will respond in a linear fashion to a small perturbation. The following linear response theory equation is applicable in this limit

$$J = -L\nabla X \qquad (3)$$

where J is the response, X is the perturbation and L is the transport coefficient. For example, eq 3 takes the following form for the viscosity

$$j_y(p_x) = -\eta \frac{\partial v_x}{\partial y} \qquad (4)$$

where $j_y(p_x)$ is the momentum flux, $\frac{\partial v_x}{\partial y}$ is the velocity gradient or shear rate and η is the shear viscosity. The most widely used NEMD approach for viscosity calculations is the so-called "SLLOD" algorithm [cite] in which a shear rate is imposed on the system and the resulting stress is computed. The shear viscosity is found at a given shear rate from the ratio of the off-diagonal components of the stress tensor to the shear rate. This gives a shear-rate dependent viscosity, which must be extrapolated to a zero shear rate in order to estimate the Newtonian shear viscosity. This points out two issues with getting accurate Newtonian shear viscosities from NEMD: 1) Ensuring the off-diagonal stress tensor average is converged for a given shear rates and 2) Using a reliable method for extrapolating the shear-dependent viscosity to a zero shear rate.

As noted above, it is possible to obtain transport coefficients using equilibrium methods via Green-Kubo (eqn 1) and Einstein (eqn 2) approaches as well as non-equilibrium (eqn 3) methods. Which approach a user chooses depends on the property and, to some extent, the preference of the user. Each property and method requires a slightly different checklist, and so we have chosen to break these lists out into the property computed and whether equilibrium or non-equilibrium methods were used.

Non-Equilibrium SLLOD Approach for Viscosity

In this method, a shear is imposed on the fluid through sliding Lees-Edwards boundary conditions. The resulting stress is computed and related to the shear viscosity by the standard Newtonian viscosity relation.

1. Ensemble: Due to the sliding boundary conditions, energy is added to the system and so a thermostat must be used to remove this energy. Thus NVT must be used.

2. The computed viscosity will be shear rate dependent; typical shear rates are much higher than what is obtainable experimentally, so a series of simulations must be run at different shear rates. Extrapolation to zero shear rate is done to estimate the Newtonian viscosity. There is no formal theoretical method for making this extrapolation, but the Carreau model *Hieber and Chiang* (*1992*); *Kioupis and Maginn* (*2000*), has often been used.
3. User selects a series of shear rates to impose; some trial and error is required to ensure that the rate is high enough to get a stress "signal" but low enough so that extrapolation to zero shear can be obtained. Rule of thumb: the inverse shear rate where shear thinning starts roughly corresponds to the rotational correlation function time constant of the longest molecular axis.
4. Parametric studies should be done varying the thermostat time constant to ensure proper temperature profiles.
5. It is possible to implement this method with "real walls", mimicking an experimental viscosity measurement. In this case, the walls are thermostatted and a nature temperature profile evolves.

Reverse Non-Equilibrium Approach for Viscosity

This method is like the opposite of SLLOD; the stress is imposed by swapping velocities of molecules between layers or slabs, and the resulting shear profile is measured. The basics of the method are shown in Fig. 1. This is much easier to compute, but requires more computational infrastructure. The method *Müller-Plathe* (*1999*) has been implemented in LAMMPS *LAM* (*????*) and critically compared to EMD and standard NEMD methods for simple Lennard-Jones systems *Tenney and Maginn* (*2010*).

Thermal Conductivity

Molecular simulation methods for thermal conductivity are fairly underdeveloped, perhaps owing to the radiative heat transfer issue.

Checklist for calculating ionic conductivity. This is a hard property to compute. Once again, NEMD and EMD methods need to be treated separately. NEMD methods (imposing electrical field and measuring fluxes) might be

For thermal conductivity of solids, beware of isotope effects, radiation, electrons,... other deviations from classical mechanics.

Ionic Conductivity

Probably recommend EMD approaches similar to GK viscosity. NEMD methods are perhaps easiest to apply for this property.

Transport Diffusivity

Unlike self-diffusivity, this is a collective property like viscosity. Can be computed with EMD and NEMD methods. Need to modify these checklists for this property.

1. Darken correction
2. DCV GCMD
3. External Field NEMD

Other things...

In the case of self-diffusivity, it is straightforward to obtain reliable estimates through the Stokes-Einstein relation. In the case of viscosity, EMD signal-to-noise ratio is not as favorable and Green-Kubo relations may be preferable. One way of enhancing signal-to-noise is through NEMD, but care must be taken when extrapolating to zero shear. All fluids are shear-thinning if the shear rate is

high enough, and NEMD methods tend to exert remarkably high shear rates. Thermal conductivity is simpler than the other two in the sense that it approaches an asymptote in the long chain limit, whereas the scaling changes with molecular weight for diffusivity and viscosity when one surpasses the entanglement threshold (roughly 1500 amu for olefin polymers). Unfortunately, molecular dynamics alone cannot accurately characterize the thermal conductivity at low density. Apparently radiative heat transfer may be important in these cases, with deviations around 30% from MD results. MD results for spheres are consistent with the Chapman-Enskog relations in the low density limit for all three properties.

Acknowledgments

References

LAMMPS Users' Workshop;. http://lammps.sandia.gov/workshops/Feb10/Craig_Tenney/rnemd-lammps2010.pdf, http://lammps.sandia.gov/tutorials/sor13/SoR_09-Common_NonEquilibrium_Calculations.pdf.

Hieber CA, Chiang HH. Shear-rate-dependence modeling of polymer melt viscosity. Polymer Engineering & Science. 1992; 32(14):931–938. http://dx.doi.org/10.1002/pen.760321404, doi: 10.1002/pen.760321404.

Kioupis LI, Maginn EJ. Impact of Molecular Architecture on the High-Pressure Rheology of Hydrocarbon Fluids. The Journal of Physical Chemistry B. 2000; 104(32):7774–7783. http://dx.doi.org/10.1021/jp000966x, doi: 10.1021/jp000966x.

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_2, 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.

Müller-Plathe F. Reversing the perturbation in nonequilibrium molecular dynamics: An easy way to calculate the shear viscosity of fluids. Phys Rev E. 1999 May; 59:4894–4898. https://link.aps.org/doi/10.1103/PhysRevE.59.4894, doi: 10.1103/PhysRevE.59.4894.

Tenney CM, Maginn EJ. Limitations and recommendations for the calculation of shear viscosity using reverse nonequilibrium molecular dynamics. The Journal of Chemical Physics. 2010; 132(1):014103. http://dx.doi.org/10.1063/1.3276454, doi: 10.1063/1.3276454.

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.