[DRAFT]: Best Practices for Thermodynamic Property Prediction from Molecular Simulations

Bryce C. Manubay,[1,*] John D. Chodera,[2,†] and Michael R. Shirts[1,‡]

[1]*University of Colorado Boulder, Boulder, CO 80309*
[2]*Computational Biology Program, Sloan Kettering Institute,*
Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States
(Dated: January 11, 2018)

This document describes a collected set of best practices for computing various physical properties from molecular simulations of liquid mixtures.
Keywords: best practices; molecular dynamics simulation; physical property computation

Todo list

I. Preliminaries

Definitions

- V: Volume
- U: Total internal energy (including potential and kinetic, excluding external energy such as due to gravity, etc)
- S: Entropy
- N: Number of particles
- T: Temperature
- P: Pressure
- k_B: Boltzmann constant
- β: $(k_BT)^{-1}$
- M: Molar mass
- ρ: Density (M/V)
- H: Enthalpy
- G: Gibbs Free Energy (free enthalpy)
- A: Helmholtz Free Energy
- μ: Chemical potential
- D: Total dipole moment
- u: reduced energy
- f: reduced free energy

* bryce.manubay@colorado.edu
† john.chodera@choderalab.org
‡ Corresponding author; michael.shirts@colorado.edu

Macroscopically, the quantities V, U, N are constants (assuming the system is not perturbed in any way), as we assume that the fluctuations are essentially zero, and any uncertainty comes from our inability to measure that constant value precisely. For a mole of compound (about 18 mL for water), the relative uncertainty in any of these quantities is about 10^{-12}, far lower than any thermodynamics experiment can actually measure.

However, in a molecular simulation, these quantities are not necessarily constant. For example, in an NVT equilibrium simulation, U is allowed to vary. For a long enough simulation (assuming ergodicity, which can pretty much always be assumed with correctly implemented simulations and simple fluids), then the ensemble average value of U = $\langle U\rangle$ will converge to a constant value, and in the limit of large simulations/long time will converge to the macroscopic value U; at least, the macroscopic value of that given model, though perhaps not the U for the real system. In an NVT simulation, clearly V is constant. In an NPT simulation, however, V is a variable, and we must estimate what the macroscopic value would be using the ensemble average $\langle V\rangle$.

The quantities T, P, and μ are typically *set* as constant during the equilibrium simulations and experiments of interest here. More precisely, the system is in contact with a thermal bath with a fixed T (or in the case of NPT simulations, in contact with a thermal and mechanical bath), and we sample from the systems in equilibrium with this bath. There are a number of quantities that can be used to ESTIMATE constants such as T and P. For example, $\langle \frac{1}{3Nk_B}\sum_i m_i|v_i|^2\rangle$, where m is the mass of each particle and $|v_i|$ is the magnitude of the velocity of each particle, is an estimate of T (the temperature of the bath), and it's average will be equal to the T. But it is not the temperature. This quantity fluctuates, but T remains constant; otherwise the simulation could not be said to be at constant temperature.

Ensemble averages of some quantity X ($\langle X\rangle$) are assumed to be averages over the appropriate Boltzmann weighting, i.e. in the NVT ensemble with classical statistical mechanics, they would be $\int X(\vec{x},\vec{p})e^{-\beta U(\vec{x},\vec{p})}d\vec{x}d\vec{p}$. We note that in the limit of very large systems, $\langle X\rangle_{NPT} = \langle X\rangle_{NVT} = \langle X\rangle_{\mu VT}$.

Ensemble averages can be computed by one of two ways. First, they can be computed directly, by running a simulation that produces samples with the desired Boltzmann distribution. In that case ensemble averages can be computed as simple averages, $\langle V \rangle = \frac{1}{N}\sum_i V_i$, where the sum is over all observations. Uncertainties can be estimated in a number of different ways, but usually require estimating the number of uncorrelated samples. Secondly, they can be calculated as reweighted estimates from several different simulations, as $\langle V \rangle$ = $\frac{1}{\sum_i w_i} V_i w_i$ where w_i is a reweighting factor that can be derived from importance sampling theory.

To simplify our discussion of reweighting, we use some additional notation. We define the reduced potential $u = \beta U(\vec{x})$ in the canonical (NVT) ensemble, $u = \beta U + \beta PV$ in the isobaric-isothermal (NPT) ensemble, and $u = \beta U - \beta N\mu$ in the grand canonical ensemble (similar potentials can be defined in other ensembles). We then define $f = \int e^{-u} dx$, where the integral is over all of the DOF of the system (x for NVT, x, V for NPT, and x, N for μVT). For NPT, we then have $f = \beta G$, and for NVT we have $f = \beta A$, while for μVT we have $f = -\beta\langle P \rangle V$.

A. Observables by reweighting

To calculate expectations at one set of parameters generated with parameters that give rise to a different set of probability distributions, we start with the definition of an ensemble average given a probability distribution $p_i(x)$.

$$\langle X \rangle_i = \int X(x) p_i(x) dx \tag{1}$$

We then multiply and divide by $p_j(x)$, to get

$$\langle X \rangle_i = \int X(x) p_i(x) \frac{p_j(x)}{p_j(x)} dx = \int X(x) p_j(x) \frac{p_i(x)}{p_j(x)} dx \tag{2}$$

We then note that this last integral can be estimated by the Monte Carlo estimate

$$\langle X \rangle_i = \int X(x) p_j(x) \frac{p_i(x)}{p_j(x)} dx = \frac{1}{N}\sum_{n=1}^{N} X(x_n) \frac{p_i(x_n)}{p_j(x_n)} \tag{3}$$

Where the x_k are sampled from probability distribution $p_j(x)$

We now define the mixture distribution of K other distributions as: $p_m(x) = \frac{1}{N}\sum_{i=1}^{N} N_k p_k(x)$, where $N = \sum_k N_k$. We can construct a sample from the mixture distribution by simply pooling all the samples from k individual simulations. The formula for calculating ensemble averages in a distribution $p_i(x)$ from samples from the mixture distribution is:

$$\langle X \rangle_i = \sum_{n=1}^{N} X(x_n) \frac{p_i(x_n)}{\sum_{k=1}^{N_k} p_k(x_n)} \tag{4}$$

In the case of Boltzmann averages, then $p_i(x) = e^{f_i - u_i(x)}$, where the reduced free energy f is unknown. Reweighting from the mixture distribution becomes.

$$\langle X \rangle_i = \sum_{n=1}^{N} X(x_n) \frac{e^{f_i - u_i(x_n)}}{\sum_{k=1}^{N_k} e^{f_k - u_k(x)}} \tag{5}$$

which can be seen to be the same formula as the MBAR formula for expectations. The free energies can be obtained by setting $X = 1$, and looking at the K equations obtained by reweighting to the K different distributions.

B. Observables by derivatives

Finite differences at different temperatures and pressures can be calculated by including states with different reduced potentials. For example, $u_j(x) = \beta_i U(x) + \beta_i (P_i + \Delta P)V$, or $u_j = \frac{1}{k_B(T_i+\Delta T)} U(x) + \frac{1}{k_B(T_i+\Delta T)} P_i V$. However, the relationship between f and G can be problematic when looking at differences in free energy with respect to temperature, because $G_2 - G_1 = \beta_2 f_2 - \beta_1 f_1$. We can in general write:

$$\Delta G_{ij}(T) = k_B T \left(\Delta f_{ij}(T) - \Delta f_{ij}(T_{ref})\right) + \frac{T}{T_{ref}} \Delta G_{ij}(T_{ref}) \tag{6}$$

, where $\Delta G_{ij}(T_{ref})$ is known at some temperature.

Since with MBAR, one can make the differences as small as one would like (you don't have to actually carry out a simulation at those points), we can use the simplest formulas: central difference for first derivatives:

$$\frac{\partial A}{\partial x} \approx \frac{1}{2\Delta x}\left(A(x+\Delta x) - A(x - \Delta x)\right)$$

And for 2nd derivatives:

$$\frac{\partial^2 A}{\partial x^2} \approx \frac{1}{\Delta x^2}\left(A(x+\Delta x) - 2A(x) + A(x - \Delta x)\right)$$

Thus, only properties at two additional points need to be evaluated to calculate both first and 2nd derivatives.

For mixed partial derivatives, we can write (in the central difference approximation).

$$\frac{\partial^2 A}{\partial x \partial y} \approx \frac{A(x+\Delta x, y+\Delta y) - A(x+\Delta x, y-\Delta y) - A(x-\Delta x, y+}{4\Delta x \Delta y}$$

It may first appear that these finite difference calculations will propagate significant error as they subtract similar numbers. However, MBAR calculates the covariance matrix between $\langle A \rangle$, $\langle A(x+\Delta x) \rangle$, and $\langle A(x-\Delta x) \rangle$, meaning in practice the uncertainty is far lower than would be expected by standard error propagation of uncorrelated observables. We should therefore then use parameters that are as close as possible.

Note that if the finite differences are re-evaluated using reweighting approaches, it is important that the simulation used generates the correct Boltzmann distribution. If not, reweighted observables will be incorrect, and the results of the finite difference approach will have significant error.

The following document details calculation of various mechanical observables by both direct methods pulled from literature sources and the use of reweighting techniques. Corrections in certain observables are also summarized where suggested by previous authors.

II. Single Phase Properties

A. Pure Solvent Properties

1. Molar volume / Density

We suggest, that molar volume is usually a better quantity to use than density. Molar volume contains the same information as density, since we assume the molar masses are invariant under parameterization. The molar volume is simply the volume divided by the number of moles, and in an NPT simulation, the volume is what one directly calculates. Any additional error propagation is complicated by putting the density in the denominator. Starting with the equation used to calculate the density experimentally:

$$\rho = \frac{M}{V} \tag{7}$$

a. Direct calculation The molar volume is simply $\frac{\langle V \rangle}{n}$, where n is the number of modes. We replace the average with the ensemble estimate (calculated either directly, or with reweighting) to obtain:

$$\rho = \frac{M}{\langle V \rangle} \tag{8}$$

b. Derivative Estimate From the differential definition of the Gibbs free energy $dG = VdP - SdT + \sum_i \mu_i dN_i$ that V can be calculated from the Gibbs free energy as:

$$\langle V \rangle = \left(\frac{\partial G}{\partial P}\right)_{T,N} \tag{9}$$

The density can therefore be estimated from the Gibbs free energy.

$$\rho = \frac{M}{\left(\frac{\partial G}{\partial P}\right)_{T,N}} \tag{10}$$

The derivative can be estimated using a central difference numerical method utilizing Gibbs free energies reweighted to different pressures.

$$\left(\frac{\partial G}{\partial P}\right)_{T,N} \approx \frac{G_{P+\Delta P} - G_{P-\Delta P}}{2\Delta p} \tag{11}$$

The density can then finally be estimated.

$$\rho \approx \frac{M}{\frac{G(P+\Delta P) - G_(P-\Delta P)}{2\Delta P}} \tag{12}$$

This can be calculated from the reduced free energy f if desired by simply substituting:

$$\rho \approx \frac{\beta M}{\frac{f(P+\Delta P) - f_(P-\Delta P)}{2\Delta P}} \tag{13}$$

Intuitively, one would imagine that equation 12 would be a worse estimate of density given that the calculations involved have more room for error than direct simulations. That being said, this method should prove invaluable when estimating densities of unsampled states using MBAR.

2. Molar Enthalpy

This section is on the relation of enthalpy to Gibbs free energy (should we need it). This is not an experimental quantity, but will be helpful in calculating related properties of interest. The enthalpy, H, can be found from the Gibbs free energy, G, by the Gibbs-Helmholtz relation:

a. Direct simulation estimate H can be calculated as directly as:

$$H = \langle U \rangle + P \langle V \rangle$$

, where U is the total internal energy (including both kinetic and potential energy).

b. Derivative estimate

$$H = -T^2 \left(\frac{\partial\left(\frac{G}{T}\right)}{\partial T}\right)_{P,N} \tag{14}$$

Transforming the derivative in the Gibbs-Helmholtz relation to be in terms of β instead of T yields:

$$H = -T^2 \left(\frac{\partial\left(\frac{G}{T}\right)}{\partial \beta}\frac{\partial \beta}{\partial T}\right)_{P,N} \tag{15}$$

Recall that $\beta = \frac{1}{k_B T}$, therefore $\frac{\partial \beta}{\partial T} = -\frac{1}{k_B T^2}$, or alternately $-k_B \beta^2$. Substituting these values into the enthalpy equation gives:

$$H = -T^2 \left(-\frac{1}{k_B T^2}\right)\left(\frac{\partial\left(\frac{G}{T}\right)}{\partial \beta}\right)_{P,N} \tag{16}$$

$$= \frac{1}{k_B}\left(\frac{\partial\left(\frac{G}{T}\right)}{\beta}\right)_{P,N} = \frac{\partial f}{\partial \beta}_{P,N} \tag{17}$$

3. Heat Capacity

There are two different heat capacities: the isochoric heat capacity $C_V = \left(\frac{\partial U}{\partial T}\right)_V$, and isobaric heat capacity $C_P = \left(\frac{\partial H}{\partial T}\right)_P$. The are calculated in similar ways.

a. Direct simulation estimate There is no direct way to estimate either heat capacity as an expectation of any observable.

b. Derivative estimates The definition of the isobaric heat capacity is:

$$C_P = \left(\frac{\partial H}{\partial T}\right)_{P,N} \tag{18}$$

This can be rewritten in terms of β finite differences (carried out at fixed P and N).

$$C_P = \left(\frac{\partial H}{\partial \beta}\frac{\partial \beta}{\partial T}\right) \tag{19}$$

$$= -k_B\beta^2\left(\frac{\partial H}{\partial \beta}\right) \tag{20}$$

And can also then be rewritten compactly as:

$$C_P = -k_B\beta^2\left(\frac{\partial^2 f}{\partial \beta^2}\right) \tag{21}$$

Which can then also be rewritten in terms of temperature derivatives of the reduced free energy alone as:

$$C_P = \frac{\partial H}{\partial T} = \frac{\partial}{\partial T}\left(-k_B T^2\left(\frac{\partial f}{\partial T}\right)\right) \tag{22}$$

$$= = -2k_B T\left(\frac{\partial f}{\partial T}\right) - k_B T^2\left(\frac{\partial^2 f}{\partial T^2}\right) \tag{23}$$

Which is somewhat more complicated and therefore less desired than the β derivative expression.

These derivatives could be computed by finite differences approach with two separate simulations using a central difference approch, or by finite difference with MBAR.

c. Fluctuation estimate The enthalpy fluctuation formula can also be used to calculate C_P[1].

$$C_P = \frac{\mathrm{Var}()}{H} N k_B T^2 \tag{24}$$

The form is equivalent for isochoric heat capacity, but with derivatives at constant volume rather than pressure.

4. Isothermal Compressibility

The definition of isothermal compressibility is:

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{25}$$

a. First Derivative Thus, it can be estimated by the finite difference of $\langle V \rangle$

$$\kappa_T = -\frac{1}{2V(T,P)^2}\left(\langle V(P+\Delta P, T)\rangle - \langle V(P-\Delta P, T)\rangle\right) \tag{26}$$

Or by the finite differences evaluation of:

$$\kappa_T = -\frac{\left(\frac{\partial^2 G}{\partial P^2}\right)_{T,N}}{\left(\frac{\partial G}{\partial P}\right)_{T,N}} = -\frac{\left(\frac{\partial^2 f}{\partial P^2}\right)_{T,N}}{\left(\frac{\partial f}{\partial P}\right)_{T,N}} \tag{27}$$

b. variance formula : κ_T can also be estimated from the ensemble average and fluctuation of volume (in the NPT ensemble) or particle number (in the μVT ensemble)[2]:

$$\kappa_T = \beta\frac{\mathrm{Var}(V)}{\langle V\rangle_{NTP}} = V\beta\frac{\mathrm{Var}(N)}{\langle N\rangle_{\mu VT}} \tag{28}$$

5. Speed of Sound

The definition of the speed of sound is[3]:

$$c^2 = \left(\frac{\partial P}{\partial \rho}\right)_S = -\frac{V^2}{M}\left(\frac{\partial P}{\partial V}\right)_S \tag{29}$$

$$c^2 = \frac{V^2}{\beta M}\left[\frac{\left(\frac{\gamma_V}{k_B}\right)^2}{\frac{C_V}{k_B}} + \frac{\beta}{V\kappa_T}\right] \tag{30}$$

Where:

$$\gamma_V = \left(\frac{\partial P}{\partial T}\right)_V \tag{31}$$

γ_V is known as the isochoric pressure coefficient. κ_T is the same isothermal compressibility from equation 20

An alternate derivation, applying the triple product rule to $\left(\frac{\partial P}{\partial V}\right)_S$ yields the following.

$$\left(\frac{\partial P}{\partial V}\right)_S = \frac{\left(\frac{\partial S}{\partial V}\right)_P}{\left(\frac{\partial S}{\partial P}\right)_V} \tag{32}$$

$$\left(\frac{\partial S}{\partial V}\right)_P = \left(\frac{\partial S}{\partial T}\right)_P \left(\frac{\partial T}{\partial V}\right)_P = \frac{C_P}{T}\left(\frac{\partial T}{\partial V}\right)_P = \frac{C_P}{TV\alpha} \tag{33}$$

Where $\alpha = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P = \left(\frac{\partial \ln V}{\partial T}\right)_P$ is the coefficient of thermal expansion. The second term in our triple product rule expansion, $\left(\frac{\partial S}{\partial P}\right)_V$, can be expressed as follows:

$$\left(\frac{\partial S}{\partial P}\right)_V = \left(\frac{\partial S}{\partial T}\right)_V \left(\frac{\partial T}{\partial P}\right)_V = \frac{C_V}{T}\left(\frac{\partial T}{\partial P}\right)_V = \frac{C_V}{T\gamma_V} \tag{34}$$

Thus our derivation yields:

$$\left(\frac{\partial P}{\partial V}\right)_S = \frac{C_P\gamma_V}{C_V V\alpha} \tag{35}$$

Horn et al set out several ways for calculating α[1]:

a. Analytical derivative of density with respect to temperature

$$\alpha = -\frac{d\ln\langle\rho\rangle}{dT} \tag{36}$$

b. Numerical derivative of density over range of T of interest The same finite differences approach as shown for isothermal compressibility can be applied here, thus:

$$\alpha = -\frac{d\ln\langle\rho\rangle}{dT} = -\frac{1}{2\rho(T,P)}\left(\ln\langle\rho(P,T+\Delta T)\rangle - \ln\langle V(P,T-\Delta T)\rangle\right) \tag{37}$$

c. Using the enthalpy-volume fluctuation formula

$$\alpha = \frac{\langle VH\rangle - \langle V\rangle\langle H\rangle}{k_B\langle T\rangle^2\langle V\rangle} \tag{38}$$

Finite differences approximations and/or analytical derivation can also be used to calculate γ_V or by note of the relation:

$$\gamma_V = -\frac{\alpha}{\kappa_T} \tag{39}$$

6. Dielectric Constant

This equation was provided by a literature reference authored by CJ Fennell[4] and is the standard for calculating the dielectric constant. Below, $\epsilon(0)$ is the zero frequency dielectric constant, V is the system volume and D is the total system dipole moment.

$$\epsilon(0) = 1 + \frac{4\pi}{3k_BT\langle V\rangle}(\langle D^2\rangle - \langle D\rangle^2) \tag{40}$$

B. Binary Mixture Properties

1. Mass Density, Speed of Sound and Dielectric Constant

The methods for these calculations are the same for a multicomponent system.

2. Activity Coefficient

The definition of chemical potential in a pure substance is:

$$\mu(T,P) = \left(\frac{\partial G}{\partial N}\right)_{T,P} \tag{41}$$

which is a function of only temperature and pressure.

Then the definition of the chemical potential μ_i of compound i in a mixture is:

$$\mu_i(T,P,\vec{N}) = \left(\frac{\partial G}{\partial N_i}\right)_{T,P,N_{j\neq i}} \tag{42}$$

N_i refers to a molecule of component i and $N_{j\neq i}$ refers to all molecules other than component i, with $\vec{N}$ the vector of all component numbers. Since μ_i is intensive, this is equivalently a function of the vector of mole fractions $\vec{x}_i$ instead of simply of N_i.

For an ideal solution, the chemical potential μ_i can be related to the pure chemical potential by

$$\mu_i(T,P,\vec{x}_i) = \mu(T,P) + k_BT\ln(\gamma_i) \tag{43}$$

By analogy to this form, we can say

$$\mu_i(T,P,\vec{x}_i) = \mu(T,P) + k_BT\ln(x_i\gamma_i) \tag{44}$$

Where γ_i is the activity coefficient of component i, and is a function of T,P,and $\vec{x}_i$. Rearrangement of the previous equation yields:

$$\gamma_i = \frac{e^{\left(\frac{\mu_i(T,P,\vec{x}_i)-\mu(T,P)}{k_B T}\right)}}{x_i} \tag{45}$$

Although chemical potentials cannot be directly calculated from simulation, chemical potential residuals can. We can calculate the difference $\mu_i(T,P,\vec{x}_i) - \mu(T,P)$ by calculating $\Delta\mu(T,P)_{liquid} - \Delta\mu(T,P)_{gas}$ using a standard alchemical simulation of the pure substance, followed by the calculation of $\mu_i(T,P,\vec{x}_i)_{liquid} - \Delta\mu(T,P,\vec{x}_i)_{gas}$, and assuming that $\Delta\mu(T,P,\vec{x}_i)_{gas} = \Delta\mu(T,P)_{gas}$. Note: there are a few subtleties here relating to the $\ln x_i$ factor, but it appears that with alchemical simulations with only one particle that is allowed to change, this will cancel out (need to follow up).

Several of these alchemical simulation methods for calculating activity coefficients have been pioneered by Andrew Paluch [**?**]. A method detailing the calculation of infinite dilution activity coefficients γ_i^{inf} for binary a mixture follows directly:

$$\begin{aligned}\ln\gamma_2^{\infty}(T,P,x_2=0) &= \beta\mu_2^{res,\infty}(T,P,N_1,N_2=1)\\ &\quad + \ln\left[\frac{RT}{V_1(T,P)}\right] - \ln f_2^0(T,P)\end{aligned} \tag{46}$$

Where $\beta\mu_2^{res,\infty}$ is the dimensionless residual chemical potential of component 2 at infinite dilution. The residual is defined here as the difference between the liquid and ideal gas state. $V_1(T,P)$ is the molar volume of component 1 at T and P. $\ln f_2^0(T,P)$ is the natural logarithm of the pure liquid fugacity of component 2 and is defined as:

$$\ln f_2^0(T,P) = \beta\mu_2^{res}(T,P) + \ln\left[\frac{RT}{V_2(T,P)}\right] \tag{47}$$

Paluch et al. use a multistage free energy perturbation approach utilizing MBAR in order to calculate the residual chemical potentials (recall that the chemical potential is the partial molar Gibbs free energy and dimensionless Gibbs free energy differences between multiple states are readily computed with MBAR). The idea is to connect two states of interest. In the case of a pure liquid, connecting a system of pure liquid molecules with $N-1$ interacting molecules and one fully decoupled molecule to a system of N fully interacting molecules. The coupling/decoupling process is detailed by Paluch et al [**?**], but involves a linear alchemical switching function where LJ and electronic interactions are slowly turned on for the decoupled molecule until they are fully on. The free energy of this coupling is calculated by sampling summing the free energy changes along this path.

3. *Excess Molar Properties*

The general definition of an excess molar property can be stated as follows:

$$y^E = y^M - \sum_i x_i y_i \tag{48}$$

Where y^E is the excess molar quantity, y^M is the mixture quantity, x_i is the mole fraction of component i in the mixture and y_i is the pure solvent quantity. In general, the simplest methods for calculating excess molar properties for binary mixtures will require three simulations. One simulation is run for each pure component and a third will be run for the specific mixture of interest. We note that only one set of pure simulations are needed to calculate excess properties at all compositions.

4. *Excess Molar Heat Capacity and Volume*

Excess molar heat capacities and volume will be calculated using the methods for the pure quantities in section I in combination with the general method for excess property calculation above.

5. *Excess Molar Enthalpy*

Excess molar enthalpy can be calculated using the general relation of molar enthalpy as it relates to Gibbs Free Energy from section I and the general method of excess molar property calculation above or by the following[5]:

$$H^E = \langle E^M\rangle + PV^E - \sum_i x_i\langle E_i\rangle \tag{49}$$

Where $\langle E\rangle$ denotes an ensemble average of total energy and V^E is calculated using the general method of excess molar properties.

C. Suggested Corrections

1. *Heat Capacity*

Horn et al suggest a number of vibrational corrections be applied to the calculation of C_P due to a number of approximations made during the simulation of the liquid [1]. The following terms were added as a correction.

$$\left(\frac{\partial E_{vib,l}}{\partial T}\right)_P = \left(\frac{\partial E^{QM}_{vib,l,intra}}{\partial T}\right)_P + \left(\frac{\partial E^{QM}_{vib,l,inter}}{\partial T}\right)_P - \left(\frac{\partial E^{CM}_{vib,l,inter}}{\partial T}\right)_P \quad (50)$$

Where:

$$\left(\frac{\partial E^{CM}_{vib}}{\partial T}\right)_P = k_B n_{vib} \quad (51)$$

$$\left(\frac{\partial E^{QM}_{vib}}{\partial T}\right)_P = \sum_{i=1}^{n_{vib}} \left(\frac{h^2 v_i^2 e^{\frac{hv_i}{k_B T}}}{k_B T^2 \left(e^{\frac{hv_i}{k_B T}} - 1\right)^2}\right) \quad (52)$$

Above, n_{vib} is the number of vibrational modes, h is Planck's constant and v_i is the vibrational frequency of mode i.

III. Properties Involving Change of Phase

A. Pure Solvent Properties

1. *Enthalpy of Vaporization*

The definition of the enthalpy of vaporization is[6]:

$$\Delta H_{vap} = H_{gas} - H_{liq} = U_{gas} - U_{liq} + P(V_{gas} - V_{liq}) \quad (53)$$

If we assume that $V_{gas} >> V_{liq}$ and that the gas is ideal:

$$\Delta H_{vap} = U_{gas,potential} - U_{liq,potential} + RT \quad (54)$$

B. Suggested Corrections

1. *Enthalpy of Vaporization*

An alternate, but similar, method for calculating the enthalpy of vaporization is recommended by Horn et al [1].

$$\Delta H_{vap} = -\frac{E_{liq,potential}}{N} + RT - PV_{liq} + C \quad (55)$$

In the above equation C is a correction factor for vibrational energies, polarizability, non-ideality of the gas and pressure. It can be calculated as follows.

$$\begin{aligned} C_{vib} &= C_{vib,intra} + C_{vib,inter} \\ &= (E_{vib,QM,gas,intra} - E_{vib,QM,liq,intra}) \\ &\quad + (E_{vib,QM,liq,inter} - E_{vib,CM,liq,inter}) \end{aligned} \quad (56)$$

The QM and CM subscripts stand for quantum and classical mechanics, respectively.

$$C_{pol} = \frac{N}{2} \frac{(d_{gas} - d_{liq})^2}{\alpha_{p,gas}} \quad (57)$$

Where d_i is the dipole moment of a molecule in phase i and $\alpha_{p,gas}$ is the mean polarizability of a molecule in the gas phase.

$$C_{ni} = P_{vap}\left(B - T\frac{dB}{dT}\right) \quad (58)$$

Where B is the second virial coefficient.

$$C_x = \int_{P_{ext}}^{P_{vap}} \left[V(P_{ext})\left[1 - (P - P_{ext})\kappa_T\right] - TV\alpha\right] dP \quad (59)$$

Where P_{ext} is the external pressure and $V(P_{ext})$ is the volume at P_{ext}.

This is frequently done as a single simulation calculation by assuming the average intramolecular energy remains constant during the phase change, which is rigorously correct for something like a rigid water molecule (intramolecular energies are zero), but less true for something with structural rearrangement between gas and liquid phases.

As discussed by myself and MRS, we have decided to not initially begin the parameterization process using enthalpy of vaporization data. While force field parameterization is commonly done using said property we have ample reason to not follow classical practice. First of all, the enthalpy data is usually not collected at standard temperature and pressure, but at the saturation conditions of the liquid being vaporized [7]. This would require corrections to be made to get the property at STP (the process will be explained below) using fitted equations for heat capacity. Not only is this inconvenient, but it adds an unknown complexity when adjusting experimental uncertainties due to the added correction. Often times the uncertainties of these "experimental" enthalpies are unrecorded because they are estimated from fitted Antoine equation coefficients [7].

An additional issue is the necessity of having to use gas phase simulation data in order to validate a parameterization process meant for small organic liquids and their mixtures. Following an example of Wang et al. [8] we plan to instead use enthalpy of vaporization calculations as an unbiased means of testing the success of the parameterization. If the parameterization procedure is expanded to use enthalpy of vaporization, corrections can be made to the experimental data in order to get a value at STP using the following equation.

$$\Delta H_{vap}(T) = \Delta H_{vap}^{ref} + \int_{T_{ref}}^{T} (C_{P,gas} - C_{P,liq})\, dT \quad (60)$$

[1] H. et al., “Development of an improved four-site water model for biomolecular simulations: Tip4p-ew,” *Journal of Chemical Physics*, vol. 120, no. 20, pp. 9665–9678, 5 2004.

[2] D. et al., “Insights into protein compressibility from molecular dynamics simulations,” *The Journal of Physical Chemistry B*, vol. 105, no. 3, pp. 715–724, 12 2000.

[3] R. Lustig, “Direct molecular nvt simulation of the isobaric heat capacity, speed of sound and joule-thomson coefficient,” *Molecular Simulation*, vol. 37, no. 6, pp. 457–465, 5 2011.

[4] C. Fennell, “Simple liquid models with corrected dielectric constants,” *The Journal of Physical Chemistry B*, vol. 116, no. 23, pp. 6936–6944, 3 2012.

[5] D. et al., “Enthalpies of mixing predicted using molecular dynamics simulations and opls force field,” *Fluid Phase Equilibria*, vol. 289, no. 2, pp. 156–165, 3 2010.

[6] W. et al., “Application of molecular dynamincs simulations in molecular property prediction. 1. density and heat of vaporization,” *Journal of Chemical Theory and Computation*, vol. 7, no. 7, pp. 2151–2165, 5 2011.

[7] C. et al., “Enthalpies of vaporization of organic and organometallic compounds, 1880âĂŞ2002,” *Journal of Physical and Chemical Reference Data*, vol. 32, no. 2, pp. 519–878, 4 2003.

[8] W. et al., “Development of polarizable models for molecular mechanical calculations iv: van der waals parameterization,” *Journal of Physical Chemistry B*, vol. 116, no. 24, pp. 7088–7101, 6 2012.