Computing chemical potentials of solutions from structure factors

Bingqing Cheng

1Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria

(Dated: May 10, 2022)

The chemical potential of a component in a solution is defined as the free energy change as the amount of the component changes. Computing this fundamental thermodynamic property from atomistic simulations is notoriously difficult, because of the convergence issues in free energy methods and finite size effects. This paper presents the S0 method, which can be used to obtain chemical potentials from static structure factors computed from equilibrium molecular dynamics simulations under the isothermal-isobaric ensemble. The S0 method is demonstrated on the systems of binary Lennard-Jones particles, urea–water mixtures, a NaCl aqueous solution, and a high-pressure carbon-hydrogen mixture.

The chemical potential of a solute as a function of its concentration is pivotal for understanding many important physical and biological processes such as osmosis, the solvation of organic molecules, the behavior of electrolytes, precipitation of crystals from solutions, and phase equilibria of mixtures.

Despite the massive importance, computing the chemical potentials of solutions from atomistic simulations is notoriously difficult and has not become a routine calculation. This is because that the existing ways [1-9] have many caveats and are often restricted to a subset of systems. For example, methods based on Monte Carlo particle insertion and removal [10,11] can have numerical convergence issues [5] and do not work for dense fluids. Thermodynamic integration or overlapping distribution method from the real solution to a reference fluid [1] involves running multiple simulations with different coupling constants as well as along different thermodynamic paths, may have singularity problems at the end points of the integration [7,8], and is difficult to use for molecules with complex topology. The direct coexistence [6] needs long equilibration time, and only works at conditions close to the solubility limit. Kirkwood-Buff integrals [12] have severe finite size effects [13].

Here we propose a generic, easy-to-use method (the S0 method) for computing chemical potentials of solutes at different concentrations in solutions. This method is based on the thermodynamic relationship between the composition fluctuations to derivatives of the chemical potentials to the concentration. In practice, only the static structure factors computed from equilibrium isothermal-isobaric (NPT) molecular dynamics (MD) simulations at different solute concentrations are needed. We benchmark the method on the systems of binary Lennard-Jones particles, urea–water, NaCl aqueous solution, and a high-pressure carbon-hydrogen mixture.

**Theory** Consider a large binary particle reservoir with fixed numbers of type A and type B particles in the NPT ensemble (illustrated in Fig. 1). The Gibbs free energy for this reservoir can be expressed as

\[ G = \mu_A N_A + \mu_B N_B, \]

where \( \mu_A \) and \( \mu_B \) are the chemical potentials of type A and type B particles, respectively, and \( N_A \) and \( N_B \) are their numbers.

Inside the NPT ensemble, a fixed volume \( V \), which is large but much smaller than the reservoir, with permeable boundary can be regarded a grand canonical ensemble

\[ N_A, N_B, P, T \]

\[ \mu_A, \mu_B, V, T \]

where

\[ (\muVT) \]

The corresponding grand potential is

\[ \Omega = -k_B T \ln \sum_{n_A=0}^{N_A} \sum_{n_B=0}^{N_B} \exp \left[ \frac{n_A \mu_A}{k_B T} \right] \exp \left[ \frac{n_B \mu_B}{k_B T} \right] Q(n_A, n_B, V, T), \]

(2)

where \( Q \) is the grand partition function of the \( V \)-region with the numbers of the two types of particles \( n_A \) and \( n_B \). By taking the derivatives of \( \Omega \) with respect to the chemical potentials, one gets the particle number fluctuations \( \Delta n_A = n_A - \langle n_A \rangle \) and \( \Delta n_B = n_B - \langle n_B \rangle \):

\[ \frac{\langle \Delta n_A \Delta n_B \rangle}{k_B T} = -\frac{\partial^2 \Omega}{\partial \mu_A \partial \mu_B} = \frac{\partial \langle n_A \rangle}{\partial \mu_B}. \]

(3)

where \( \langle \ldots \rangle \) indicate the expectation value of an observable in the grand canonical ensemble. As derived in the seminal paper of Kirkwood-Buff [12], such equilibrium fluctuations can be used obtain the derivatives of the \( \mu \) to the at constant \( (P,T) \) conditions, e.g.

\[ \left( \frac{\partial \mu_A}{\partial c_A} \right)_{T,P} = \frac{k_B T}{c_A} \frac{1}{\langle \Delta n_A^2 \rangle/n_A - \langle \Delta n_A \Delta n_B \rangle/n_B}, \]

(4)

where \( c_A = \langle n_A \rangle/V \) equals to the concentration of the type A particles in the NPT reservoir. Ref. [12] then relates the particle number fluctuations to the Kirkwood-Buff integrals of radial distribution functions. Instead, here we take a different route and evaluate \( \langle \Delta n_A \Delta n_B \rangle \) using static structure factors detailed below.

The instantaneous density field of particle number inside

![Fig. 1. An illustration of the isothermal-isobaric (NPT) and the grand canonical (\muVT) ensembles. The \muVT ensemble inside is only used for the theoretical derivation, and the whole NPT ensemble is used to compute the structure factors.](image-url)
the NPT ensemble is
\[ \rho_A(r, t) = \sum_{i=1}^{N_A} \delta(r_i(t) - r), \]
where \( r_i(t) \) is the position of atom \( i \) of type \( A \) at time \( t \), and the average number density is
\[ \rho_A^{(1)}(r) = \langle \rho_A(r, t) \rangle_{NPT}, \]
where \( r_i(t) \) is the position of atom \( i \) of type \( A \) at time \( t \), and for isotropic systems \( \rho_A^{(1)}(r) = c_A \). Note that \( \langle \ldots \rangle_{NPT} \) is the NPT ensemble average, while \( \langle \ldots \rangle \) without subscript indicates \( \muVT \) average. To consider the two-body correlations between the density at different points in space, \( r' \) and \( r'' \):
\[ \rho_{AB}^{(2)}(r', r'') = \langle \rho_A(r', t) \rho_B(r'', t) \rangle_{NPT}. \]
These density correlation functions from the NPT ensemble encodes the particle fluctuations inside the \( \muVT \) ensemble of volume \( V \), because
\[ \int_V d\mathbf{r} \rho_A^{(1)}(r) = \langle n_A \rangle, \]  
\[ \int_V d\mathbf{r} \int_V d\mathbf{r}' \rho_{AB}^{(2)}(r', r'') = \langle n_A n_B \rangle, \]
\[ \int_V d\mathbf{r} \int_V d\mathbf{r}'' \left( \rho_{AB}^{(2)}(r', r'') - \rho_A^{(1)}(r') \rho_B^{(1)}(r'') \right) = \langle \Delta n_A \Delta n_B \rangle. \]
One can do a Fourier expansion of the instantaneous density field in space \( r \) inside the \( \muVT \) ensemble with volume \( V \), e.g.
\[ \tilde{\rho}_A(k, t) = \int_V d\mathbf{r} \rho_A(r, t) \exp(ik \cdot r) = \sum_{i=1}^{N_A} \exp(ik \cdot r_i(t)). \]
As the density field is a real function, \( \tilde{\rho}_A(-k, t) = \tilde{\rho}_A^*(k, t) \), where the latter is the complex conjugate.

The static structure factor between two types of particles (A–A, A–B, or B–B) is defined as
\[ S_{AB}(k) = \frac{1}{\langle n_A \rangle \langle n_B \rangle} \langle \tilde{\rho}_A(k, t) \tilde{\rho}_B(-k, t) \rangle. \]
It is easy to verify that the \( S_{AB}(k) \) is the Fourier expansion of \( \rho_{AB}^{(2)}(r', r'') \), i.e.
\[ S_{AB}(k) = \frac{1}{\langle n_A \rangle \langle n_B \rangle} \int_V d\mathbf{r}' \exp(ik \cdot \mathbf{r}') \int_V d\mathbf{r}'' \exp(ik \cdot \mathbf{r}'') \langle \rho_A(r', t) \rho_B(r'', t) \rangle. \]
Combining Eqn. (10) with (13), the structure factor is related to the particle number fluctuations via
\[ S_{AB}^0 \equiv \lim_{k \to 0} S_{AB}(k) = \frac{\langle \Delta n_A \Delta n_B \rangle}{\sqrt{\langle n_A \rangle \langle n_B \rangle}}. \]
Furthermore, the Kirkwood–Buff integral (KBI) [12] between components \( A \) and \( B \) is and related to the structure factor by
\[ G_{AB} = \frac{1}{\sqrt{S_{AB}}} \left( S_{AB}^0 - \delta_{AB} \right). \]
Plugging Eqn. (14) into Eqn. (4), one obtains
\[ \left( \frac{\partial \mu_A}{\partial c_A} \right)_{T, P} = \frac{k_B T}{c_A} \left[ \frac{1}{S_{AA}^0 - S_{AB}^0 \sqrt{c_A/c_B}} \right]. \]
Importantly, although the Fourier expansions above are performed inside the volume \( V \) of the \( \muVT \) ensemble, for isotropic liquid with translational invariance, \( S_{AB}(k) \) should be the same in any parts of the larger NPT ensemble—which includes the whole volume of the NPT ensemble. To obtain \( S_{AB}(k) \) from NPT simulations of finite fluid systems with periodic boundary conditions, one can only let one dimension of the simulation box fluctuate with barostat, while only collecting \( S(k) \) for \( k \) along the plane of the two fixed dimensions. Even better, one can also perform the Fourier expansion on the scaled coordinates, and obtain the static structure factors using
\[ S_{AB}(k) = \frac{1}{\sqrt{N_A N_B}} \sum_{i=1}^{N_A} \exp(ik \cdot \mathbf{r}_{iA}(t)) \sum_{i=1}^{N_B} \exp(-ik \cdot \mathbf{r}_{iB}(t)) \]
where \( \mathbf{r}(t) = t(t)/NPT / l(t) \) and \( l(t) \) is the instantaneous dimension of the supercell. The scaling procedure is rigorous at the thermodynamic limit, where the NPT and the NVT ensembles are equivalent.

To determine \( \lim_{k \to 0} S_{AB}(k) \) from MD simulations of finite system sizes, one can compute \( S(k) \) at small \( k \) under the NPT ensemble, and extrapolate to \( k \to 0 \) case using the Ornstein–Zernike form [13]:
\[ S_{AB}(k) = \frac{S_{AB}^0}{1 + k^2 S_{AA}^0}. \]
Finally, to compute chemical potentials, one can run multiple equilibrium NPT simulations with different concentrations and then obtain \( \mu_A(c_A) = \langle c_A \rangle_{NPT} \) using numerical integration with respect to \( \ln c_A \):
\[ \mu_A(c_A) = \mu_A^0 + k_B T \left[ \ln \left( \frac{c_A}{c_A^0} \right) + \int_{\ln c_A^0}^{\ln c_A} d\ln(c_A) \left[ \frac{S_{AA}^0}{S_{AB}^0 \sqrt{c_A/c_B}} - 1 \right] \right], \]
referenced to the chemical potential \( \mu_A^0 \) at a standard molar concentration \( c_A^0 \). One can conveniently select this reference to be the pure state. Strictly speaking, Eqn. (19) provides the relative chemical potential instead of the absolute value, but only the former is a physical observable. The first
two terms on the right hand side of Eqn. (19) is the ideal-mixture chemical potential $\mu^{id}$, and the third term is the excess chemical potential $\mu^{ex} = k_B T \ln (\gamma_A)$ where $\gamma_A$ is the activity coefficient of the solute A. The ratio $\gamma_A^2 = S_A^0 - S_{AB}^0 V c_A / c_B$ by $\gamma_A = 1 + d \ln (\gamma_A)/d \ln c_A$.

In addition, although one can evaluate $\mu_A(c_A)$ and $\mu_B(c_B)$ separately using Eqn. (19) from the same simulations, one can obtain $\mu_A$ from $\mu_B$ employing the Gibbs-Duhem equation under constant P,T conditions [15]:

$$N_A d\mu_A + N_B d\mu_B = 0. \quad (20)$$

To automatically satisfy the Gibbs-Duhem with the S0 method, one can use

$$\left( \frac{\partial \mu_A}{\partial \ln (\chi_A)} \right)_{T,P} = -\frac{k_B T}{\chi_B S_{AA}^0 + \chi_A S_{AB}^0 - 2 \sqrt{\chi_A \chi_B} S_{AB}^0},$$

where $\chi_A$ and $\chi_B = 1 - \chi_A$ are the molar fraction of A and B.

It is worthwhile discussing the finite size effects in the current approach, and in particular, the difference with the KBI method. Typically, when using KBI, one either starts from the pair correlation functions, or one collects $n_A$ and $n_B$ inside a fixed volume $V$ during NPT simulations [13][16]. However, the open boundary of $V$ imposes very large finite size effects [13], and even for $\mu_V T$ systems with hundreds of thousands of atoms, a significant fraction are lying on the boundary [17]. Without finite size corrections, the KBI approach is hardly applicable [13][16]. In contrast, computing $S(k)$ in an NPT ensemble avoids such boundary effects. In addition, the effects coming from using a finite wavelength $(2\pi/k)$ is partly corrected by the physically-inspired extrapolation (Eqn. (18)).

**Binary Lennard-Jones system** A binary mixture (A, B) of Lennard-Jones (LJ) fluids was simulated using a purely repulsive 6-12 LJ potential truncated and shifted with cutoff radius $2^{1/6} \sigma$. The potential parameters are $\sigma_{AA} = \sigma_{BB} = \sigma_{AB} = 1$, and $\epsilon_{AA} = 1.2$, $\epsilon_{BB} = 1.0$, with $\epsilon_{AB} = (\epsilon_{AA} + \epsilon_{BB})/2$. Constant temperature and pressure was enforced through a stochastic velocity rescaling thermostat and Nose-Hoover barostat. Three systems sizes, 4000, 23328 and 108000 total number of atoms, were considered in the range of mole fractions of A particles, $\chi_A = 0.05, \ldots, 1.0$. Simulations were carried out using LAMMPS [18] with a time step of $10^{-3}$. Total simulation time was $10^6$ simulation steps for the two larger systems, and $10^7$ for $N = 4000$. To compute $S(k)$, we collected a snapshot per 2000 steps of the trajectory.

Fig. 2A shows $S_{AB}(k)$ as a function of $k^2 = ||k||^2$ computed from NPT simulations at different system sizes for a selected mole fraction of particles. The dashed curves are fits to Eqn. (18) with a maximum cutoff in the wavevector $k_{cut} = 2 \pi / 0.2 \sigma^{-1}$, although we found the fitted values for $S^0$ is insensitive to this choice. Even with a small size of 4000 particles the estimate for $S^0_{AB}$ is converged. Such insensitivity to system size is again confirmed from Fig. 2B which shows the the extrapolated values of $S^0_{AB}$ at different molar fractions. As a benchmark, in Fig. 2C we compare the excess chemical potential $\mu^e_A$ and $\mu^e_B$, with the reference $c^0_A$ and $c^0_B$ set to the concentrations of pure A and pure B (Eqn. (19)), computed using the S0 method and Widom particle insertion [19]. Widom insertion simulations were performed the same system size (23328 atoms), $4 \times 10^6$ time step and one particle insertion each time step. Both methods captures the fact that $\mu^e$ for both A and B are larger when $\chi_A$ is higher, because A particles have a stronger repulsive core, but the S0 method has much better statistical efficiency.

**Urea in water** We analyzed the MD trajectories of urea/water mixtures from Ref. [20], simulated using the Kirkwood-Buff derived force field [21] and SPC/E water at 1 atm pressure and 300 K temperature. Four urea molar concentrations were considered (2, 4, 6, and 8 M), and the system size is 13000-16000 total number of urea and water molecules. The comparison of the derivative of the activity coefficient $\gamma_\text{urea} = 1 + d \ln (\gamma_\text{urea})/d \ln c_\text{urea}$ is in Fig. 3. The S0 results closely agree with the Ref. [13] that uses KBI with finite size corrections, and with Exp. 1 [21][22].

**NaCl aqueous solution** Simulations of NaCl water solutions at different molar concentrations (0.1-9.3M) were performed using LAMMPS [18] at 298.15 K and 1 bar.

![Figure 2](image-url)
on how adding hydrogen to the liquid carbon changes the chemical potential of carbon atoms. We used a machine learning potential developed in Ref. [26], and performed NPT simulations at different $N_C/N_H$ ratios at $T=5000$ K and $P=100$ GPa. The system sizes are between 11232 and 82944 total number of atoms.

In Fig. 5 we show the derivative of the activity coefficient ($\gamma_C^L$) and excess chemical potential $\mu_C^{ex}$ of carbon at different molar fraction. The S0 method predicts $\mu_C^{ex}$ that are in good agreement with the results obtained by coexistence simulations [26], but it can be used for the regime of low carbon concentration which becomes prohibitive using the coexistence approach. Interestingly, $\gamma_C$ shows nonmonotonic behavior with a minimum at $N_H/N_C = 2$, and maxima at $N_H/N_C = 1$ and 4. These magic numbers are probably related to the chemical bonds between C and H.

To conclude, we present the S0 method to compute chemical potentials of mixtures just from equilibrium MD simulations at the NPT ensemble, by simply computing static structure factors. We demonstrate the generality and robustness of the S0 method on diverse systems including a model LJ system, organic molecule in water, aqueous electrolyte, and a high-pressure solution. In principle, the S0 method is also applicable to larger molecules such as polymers, or large-molecule solvents. Compared with the previous methods such as particle insertion [16] [11], thermodynamic integration [12], the S0 method is more generally applicable and works particularly well for dense fluids with complex interactions and high solute concentrations. We envisage the S0 method will largely simplify the computation of the chemical potentials of complex solutions, and make them routine endeavors.

**Acknowledgements** I thank Daan Frenkel for providing feedback on an early draft and for stimulating discussions. I thank Debashish Mukherji and Robinson Cortes-Huerto for kindly sharing the trajectories for urea water mixtures. I thank Aleks Reinhardt for useful suggestions on the manuscript.

**Data availability statement** All PYTHON scripts and simulation input files generated for the study are in the SI repository (url to be inserted upon acceptance of the paper).
[1] E. Sanz and C. Vega, The Journal of Chemical Physics 126, 014507 (2007).
[2] A. S. Paluch, S. Jayaraman, J. K. Shah, and E. J. Maginn, The Journal of Chemical Physics 133, 124504 (2010).
[3] M. Lisal, W. R. Smith, and J. Kolafa, The Journal of Physical Chemistry B 109, 12956 (2005).
[4] F. Moucka, M. Lisal, J. Skvor, J. Jirsak, I. Nezbeda, and W. R. Smith, The Journal of Physical Chemistry B 115, 7849 (2011).
[5] C. Perego, F. Giberti, and M. Parrinello, The European Physical Journal Special Topics 225, 1621 (2016).
[6] I. S. Joung and T. E. Cheatham III, The Journal of Physical Chemistry 112, 9020 (2008).
[7] L. Li, T. Totton, and D. Frenkel, The Journal of Chemical Physics 146, 214110 (2017).
[8] L. Li, T. Totton, and D. Frenkel, The Journal of Chemical Physics 149, 054102 (2018).
[9] H. A. Vinutha and D. Frenkel, The Journal of Chemical Physics 154, 124502 (2021).
[10] M. P. Allen and D. J. Tildesley, *Computer Simulation in Chemical Physics*, Vol. 397 (Springer Science & Business Media, 2012) pp. 108–111.
[11] B. Smit and D. Frenkel, Molecular Physics 68, 951 (1989).
[12] J. G. Kirkwood and F. P. Buff, The Journal of Chemical Physics 19, 774 (1951).
[13] R. Cortes-Huerto, K. Kremer, and R. Potestio, The Journal of Chemical Physics 145, 144103 (2016).
[14] J.-L. Barrat and J.-P. Hansen, *Basic Concepts for Simple and Complex Liquids* (Cambridge University Press, 2003).
[15] F. Moucka, I. Nezbeda, and W. R. Smith, The Journal of Chemical Physics 139, 124505 (2013).
[16] V. Braten, O. Wilhelmsen, and S. K. Schnell, Journal of Chemical Information and Modeling 61, 840 (2021).
[17] M. Heidari, K. Kremer, R. Potestio, and R. Cortes-Huerto, Entropy 20, 222 (2018).
[18] S. Plimpton, J. Comput. Phys. 117, 1 (1995).
[19] D. Frenkel and B. Smit, *Understanding Molecular Simulation: from Algorithms to Applications*, Vol. 1 (Elsevier, 2001).
[20] T. E. de Oliveira, P. A. Netz, K. Kremer, C. Junghans, and D. Mukherji, The Journal of Chemical Physics 144, 174106 (2016).
[21] S. Weerasinghe and P. E. Smith, The Journal of Physical Chemistry B 107, 3891 (2003).
[22] R. Stokes, Australian Journal of Chemistry 20, 2087 (1967).
[23] O. Miyawaki, A. Saito, T. Matsuo, and K. Nakamura, Bioscience, Biotechnology, and Biochemistry 61, 466 (1997).
[24] Z. Mester and A. Z. Panagiotopoulos, The Journal of Chemical Physics 142, 044507 (2015).
[25] A. Benavides, J. Aragones, and C. Vega, The Journal of Chemical Physics 144, 124504 (2016).
[26] B. Cheng, M. Bethkenhagen, and S. Hamel, Unpublished (2022).