

# Accelerating Monte Carlo Molecular Simulations Using Novel Extrapolation Schemes Combined with Fast Database Generation on Massively Parallel Machines

Sahar Amir, Ahmad Kadoura, Amgad Salama and Shuyu Sun

King Abdullah University of Science and Technology (KAUST)

Thuwal, KSA

{sahar.amir & ahmad.kadoura & amgad.salama & shuyu.sun}@kaust.edu.sa

**Abstract** – In this paper we introduce efficient thermodynamically consistent techniques to extrapolate and interpolate normalized Canonical NVT ensemble averages for Lennard-Jones (L-J) fluids at different thermodynamic conditions from expensively simulated data points. This leads to a significant speed up in generating intensive data. Preliminary results show promising applicability in oil and gas modelling. To investigate the efficiency of these methods, two databases corresponding to different combinations of normalized density and temperature values are generated. One contains 175 Markov chains each with 10,000,000 MC cycle results and the other contains 3000 Markov chains with 61,000,000 MC cycle results. The relative error of the thermodynamic extrapolation and thermodynamic interpolation schemes were investigated with respect to classical interpolation and extrapolation.

**Index Terms** – Molecular Simulation, Monte Carlo, Canonical Ensemble, Lennard-Jones Fluid, Extrapolation, and Interpolation.

## I. INTRODUCTION

Accurate determination of thermodynamic properties of fluid has always been a challenging task. Therefore, there is a strong motivation to try to estimate such properties based on understanding the behaviour of a smaller number of particles at the conditions of interest and finding a way to relate it to the entire domain under study. However, this technique requires long computation times, which makes its coupling with simulators unattractive. An interesting idea has been introduced suggests that, if one could obtain thermodynamic properties of a fluid at given conditions based on a single Markov chain simulation it may be possible to extrapolate these properties to obtain others at different conditions [1-4]. For instance, in histogram reweighting, histograms are scaled to estimate the new properties [1]. Thermodynamic scaling can be used on these histograms too in the non-physical umbrella sampling known as temperature scaling, density scaling or temperature-density scaling [2-4]. In contrast to some of the previously reported methods where unphysical non-Boltzmann distributions were utilized and some tuned parameters were required for scaling [1-4], our proposed thermodynamic schemes maintain physically meaningful distributions and do not contain any tuned parameters in the processes.

## II. THERMODYNAMIC EXTRAPOLATION

Thermodynamic extrapolation can be applied with varying temperatures and/or varying densities:

### A. Reweighting

For varying temperatures, the thermodynamic extrapolation will be done in the  $\beta = 1/T$  direction. Reweighting the original sampled chain is used to compute the value to be estimated and apply the effects of the change in temperature. The energy can be estimated using the following equation:

$$R_{\text{final}} = \left( \frac{\sum \left( (U_{k \beta_0} - U_{\text{tail}}) * e^{-(\Delta \beta) U_{k \beta_0}} \right)}{\sum e^{-(\Delta \beta) U_{k \beta_0}}} \right) + U_{\text{tail}}$$

$U_{k \beta_0}$  is the energy of the nearest MC chain to the data point to which the extrapolation will be done. The same approach can be used to find the pressure with respective differences in the colored parts of the equation where in addition to  $P_{\text{tail}}$  the pressure ideal is considered too ( $P_{\text{ideal}} = \rho/\beta$ ).

### B. Reconstruction

For varying densities, the thermodynamic extrapolation will be done in the  $\rho$  direction. Reconstructing and reweighting the original nearest MC chain can be utilized to estimate the targeted values. In this case, both of the weighting and the scaling factors of the sampled properties change. The energy and pressure can be extrapolated using these steps [5]: 1) Find the energy MC chain for the nearest data point from which the estimation will be performed, which is  $U_{k \beta_0}$ , find the MC chains saved for the values  $\text{Dist}_{12} = \left(\frac{1}{L_0 * r}\right)^{12}$  and  $\text{Dist}_6 = \left(\frac{1}{L_0 * r}\right)^6$ . 2) Find the old and new volume of the system using  $v_{\text{old}} = \frac{N}{\rho_{\text{old}}}$  and  $v_{\text{new}} = \frac{N}{\rho_{\text{new}}}$  and find the L values using both densities. 3) Find the L ratio using:  $L_{\text{ratio}} = L_{\text{old}}/L_{\text{new}}$ . 4) Then, use the following formulas:

$$U_{\text{tail}} = N \left(\frac{8}{3}\right) \pi \rho_{\text{new}} \left( \left(\frac{1}{3}\right) \left(\frac{1}{rc * L_{\text{new}}}\right)^9 - \left(\frac{1}{rc * L_{\text{new}}}\right)^3 \right)$$

$$U_{\text{resd}} = 4 (L_{\text{ratio}}^{12} * \text{Dist}_{12} - L_{\text{ratio}}^6 * \text{Dist}_6)$$

$$R_{\text{final}} = \left( \frac{\sum ((U_{\text{vir}} + U_{\text{tail}}) \times e^{-(\beta * U_{\text{new}} - \beta * U_{\text{old}})})}{\sum e^{-(\beta * U_{\text{new}} - \beta * U_{\text{old}})}} \right) + U_{\text{tail}}$$

The same approach can be used to find the pressure with respective differences. The equations used in calculating the pressure are:

$$P_{\text{tail}} = \left(\frac{16}{3}\right) \pi \rho_{\text{new}}^2 \left( \left(\frac{2}{3}\right) \left(\frac{1}{rc * L_{\text{new}}}\right)^9 - \left(\frac{1}{rc * L_{\text{new}}}\right)^3 \right)$$

$$P_{\text{vir}} = 8 (2 * L_{\text{ratio}}^{12} * \text{Dist}_{12} - L_{\text{ratio}}^6 * \text{Dist}_6) * \left(\frac{\rho_{\text{new}}}{N}\right)$$

### C. Reconstruction and Reweighting

For combined change in temperature and density values, a computation similar to the reconstruction is done. However, the weighting factors will be slightly different because they include both  $\beta$  and  $\rho$  change.

$$F_{\beta_n \rho_n} = \frac{\sum w_k (F_k)_{\text{rec}}}{\sum w_k}, \quad F = P, E$$

$$w_k = e^{-\Delta(\beta U_k)}_{\beta_0 \rho_0}$$

### III. THERMODYNAMIC INTERPOLATION

For the thermodynamic interpolation, four MC chains are extrapolated thermodynamically to find four values for the point under study and used with the relative weights. Many ways can be used to find the relative weights for each chain. Here, we adopted the relative weight to depend on the magnitude of opposite area and applied a power-law  $\alpha$  to each one:

$$B_i = A_i^\alpha, \quad A_i = \Delta x \Delta y, \quad \alpha > 0$$

$$w_{\text{Total}} = \sum_{i=1}^4 (B_i), \quad w_i = \frac{B_i}{w_{\text{Total}}}$$

The intuition behind using an  $\alpha$  power applies to any function of two or more arguments. By rising the weight powers in a way that they sum to one at the end, we are extenuating the nearest ones and the further ones decay faster than when no power is used, then the following sum is formed:

$$F_{\beta_n \rho_n} = \sum_{i=1}^4 w_i (F_{\beta_i \rho_i})_{\text{Thermo\_Ex}}, \quad F = P, E$$

$$w_k = \frac{A_k^\alpha}{\sum_{i=1}^4 (A_i^\alpha)}$$

experiments indicated that the best value for  $\alpha$  need to be investigated for each MC chain database table once then it could be used for the entire table after that. The reason for that has not been investigated yet. For example, in a database with  $\Delta\beta = 0.1$  and  $\Delta\rho = 0.04$  the absolute difference between the actual result from MC simulation and the estimated one using the thermodynamic interpolation is shown in Fig. 1. For this example, the best value found with the smallest error for energy in the experiments was at  $\alpha = 1.1$ .

### IV. RELATIVE ERRORS

Many experiments were done in this research to find the best scheme according to the least relative error. Some of these experiments are shown here in TABLE 1. Estim indicates that an extrapolation or interpolation for the energy or pressure will be done in that variable direction of the database.

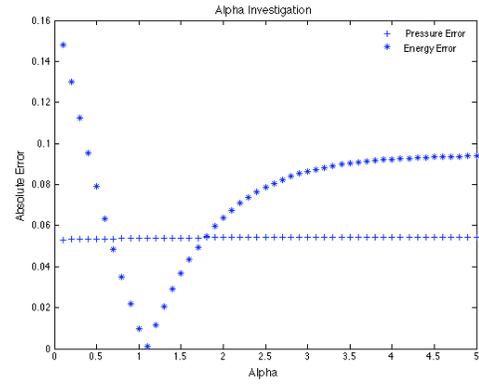


Fig. 1 investigating  $\alpha$  effect on the accuracy of the function

TABLE 1  
RELATIVE ERROR EXAMPLES

| $\beta$    | $\rho$     | Method used                 | Relative pressure error | Relative energy error |
|------------|------------|-----------------------------|-------------------------|-----------------------|
| 0.06 Estim | 0.2        | Classical Extrapolation     | 26.04%                  | 24.83%                |
|            |            | Classical Interpolation     | 0.52%                   | 0.11%                 |
|            |            | Thermodynamic Extrapolation | 4.60%                   | 12.58%                |
|            |            | Thermodynamic Interpolation | 2.16%                   | 7.07%                 |
| 0.12       | 0.22 Estim | Classical Extrapolation     | 22.72%                  | 6.87%                 |
|            |            | Classical Interpolation     | 33.20%                  | 7.00%                 |
|            |            | Thermodynamic Extrapolation | 29.39%                  | 6.43%                 |
|            |            | Thermodynamic Interpolation | 15.72%                  | 3.24%                 |
| 0.06 Estim | 0.22 Estim | Classical Extrapolation     | 30.97%                  | 22.93%                |
|            |            | Classical Interpolation     | 31.4%                   | 7.05%                 |
|            |            | Thermodynamic Extrapolation | 21.79%                  | 3.96%                 |
|            |            | Thermodynamic Interpolation | 24.42%                  | 0.0019%               |

### V. CONCLUSION

According to all of the previously done tests in this research, we can order the schemes according to the least relative error in this order: 1) Thermodynamic interpolation. 2) Thermodynamic extrapolation. 3) Classical interpolation. 4) Classical extrapolation. Thermodynamic interpolation has the least relative error. Still there are penalties to be paid for the less relative error where many flops are used to do the computations on each MC chain. Further discussion on this matter will be presented in latter work on this topic.

### VI. REFERENCES

- [1] Conrad, P. and J, De Pablo. *Comparison of histogram reweighting techniques for a flexible water model*. Fluid phase equilibria, 1998. **150**: p. 51-61
- [2] Torrie, G. and J. Valleau. *Monte Carlo study of a phase separating liquid mixture by umbrella sampling*. The Journal of chemical physics, 1977. **66**: p. 1402-1408
- [3] Valleau, J.P. *Density scaling Monte Carlo study of subcritical Lennard Jonesium*. The Journal of Chemical Physics, 1993. **99**: p. 4718-4720
- [4] Valleau, J.P. *Temperature-and-density-scaling Monte Carlo: isothermal isobaric thermodynamics of Lennard Jonesium*. Molecular Simulation, 2005. **31**: p. 255-275
- [5] Sun, S. and A. Kadoura, A. Salama. *An Efficient Method of Reweighting and Reconstructing Monte Carlo Molecular Simulation Data for Extrapolation to Different Temperature and Density Conditions*. Procedia Computer Science, 2013. **18**: p. 2147-2156.