

Available online at www.sciencedirect.com**SciVerse ScienceDirect**

Procedia Computer Science 18 (2013) 2147 – 2156

Procedia
Computer Science

2013 International Conference on Computational Science

An Efficient Method of Reweighting and Reconstructing Monte Carlo Molecular Simulation Data for Extrapolation to Different Temperature and Density Conditions

Shuyu Sun, Ahmad Kadoura and Amgad Salama

Computational Transport Phenomena Laboratory (CTPL), Division of Physical Sciences and Engineering (PSE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia

Abstract

This paper introduces an efficient technique to generate new molecular simulation Markov chains for different temperature and density conditions, which allow for rapid extrapolation of canonical ensemble averages at a range of temperatures and densities different from the original conditions where a single simulation is conducted. Obtained information from the original simulation are reweighted and even reconstructed in order to extrapolate our knowledge to the new conditions. Our technique allows not only the extrapolation to a new temperature or density, but also the double extrapolation to both new temperature and density. The method was implemented for Lennard-Jones fluid with structureless particles in single-gas phase region. Extrapolation behaviors as functions of extrapolation ranges were studied. Limits of extrapolation ranges showed a remarkable capability especially along isochors where only reweighting is required. Various factors that could affect the limits of extrapolation ranges were investigated and compared. In particular, these limits were shown to be sensitive to the number of particles used and starting point where the simulation was originally conducted.

© 2013 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/4.0/).

Selection and peer review under responsibility of the organizers of the 2013 International Conference on Computational Science

Keywords: Molecular simulation; Monte Carlo simulation; Boltzmann distribution; MC chain reweighting; MC chain reconstruction

1. Introduction

Monte Carlo (MC) molecular simulation is capable of producing a huge set of useful information about thermodynamic systems. No wonder, various methods have been developed to benefit from this technique. For instance, histogram reweighting [1] was introduced where usually a histogram is created per simulation at given thermodynamic conditions. These histograms or combined ones into a single collective distribution are scaled in order to estimate properties at new different conditions. Histogram reweighting technique has been implemented by several groups in various ways and different ensembles to serve different applications [2-6].

* Corresponding author. Tel.: +966-544-700-084.

E-mail address: shuyu.sun@kaust.edu.sa

In addition to histogram reweighting, two types of thermodynamics scaling MC were established based on non-physical or umbrella sampling [7]. First one is known as temperature-scaling (TSMC) [8] where properties can be predicted at different temperatures than simulated ones. The other method proposed is density-scaling MC (DSMC) [9]. In DSMC, a single simulation can provide free energies to get ensemble averages for a range of densities along isotherms. Valleau has implemented DSMC to study columbic phase transition [10] and subcritical Lennard-Jonesium [11]. Years later, same author has developed temperature-and-density-scaling (TDSMC) for both canonical [12] and isothermal-isobaric ensemble [13]. Thermodynamics scaling concept was also coupled with Gibbs ensemble (TSGE) by Kiyohara et al [14].

In this work, a set of MC simulations in canonical NVT ensemble is run for a range of temperatures and densities. MC chains obtained by each simulation was saved to create a data bank for the whole set. Later on, sampled MC chains were used to extrapolate ensemble averages such as energy and pressure. Lennard-Jones (LJ) fluid in single-phase region was investigated.

In contrast to some of the reported methods in the literature where unphysical non-Boltzmann distributions are utilized and some tuned parameters are required for reconstruction [10, 11], our proposed extrapolation schemes retain physically meaningful distributions and do not contain any tuned parameters in the reweighting and reconstructing processes. In addition, the presented schemes are simple and self-consistent. Numerical examples to be illustrated indicate that the proposed method is implementation friendly and CPU efficient.

Following sections will include theoretical framework of the method followed by numerical examples. These examples include pressure and energy extrapolation along isochors, isotherms and paths where both temperature and density are changing. Working extrapolation ranges were investigated too in terms of number of particles and simulation thermodynamic conditions.

2. Theory

Statistical thermodynamics suggests that the macroscopic value of an observable quantity (A), can be expressed as an average or expected value of a corresponding function among coordinates (x) and momenta (p) of N particles in the system.

$$\langle A \rangle = \frac{\int e^{-\beta H} A(\mathbf{p}^N, \mathbf{x}^N) d^3 p_1 d^3 p_2 \dots d^3 p_N d^3 x_1 d^3 x_2 \dots d^3 x_N}{\int e^{-\beta H} d^3 p_1 d^3 p_2 \dots d^3 p_N d^3 x_1 d^3 x_2 \dots d^3 x_N} \quad (1)$$

Apparently, evaluating such an integral using traditional quadrature technique is prohibitively cumbersome which clearly suggests that better techniques are necessarily needed. A naive way to evaluate such an integral may be by calculating the integrand at a large number of points of the independent variables randomly distributed over the domain of interest, which in the limits when the number of such points approaches infinity reduces to the correct average. However, such a methodology is not robust in evaluating integrals as those given in (1) because most of the calculations are spent on points where the Boltzmann factor is negligibly small. This has been because the value of the integrand is measured at a predetermined set of points. In other words, the choice of quadrature points does not depend on the value of the integrand and therefore many points might be chosen in areas where the integrand is vanishingly small. However, if it is possible to sample the random points such that much are chosen in the region where Boltzmann factor is large and fewer points otherwise, it might be possible to construct an efficient algorithm.

One of the powerful techniques that have been applied to such problem has been the important sampling-incorporated Monte Carlo algorithm [15]. In this technique a random walk is designed along the region of space where the integrand is significant by rejecting the moves, which takes the particle away from the region of predefined acceptance criteria.

The energy of intermolecular interactions in each configuration is evaluated and the Boltzmann probability is used to either accept or reject the new configuration. The MC simulation can be performed in various

ensembles by imposing different constrained variables. When MC cycles are sufficiently long, all possible states are sampled by performing different set of trial moves.

In canonical ensemble (NVT ensemble), simulation is conducted at a predefined temperature (T) and number density (ρ). An N number of particles, in this case structureless particles, are placed in a cell with volume (V) corresponding to the specified density under periodic boundary condition. Throughout the simulation, only translational moves in random direction and magnitude are allowed to the particles such that the number of particles, the volume and the temperature are all preserved.

At the temperature and density conditions where NVT-MC simulations were run, repulsive short interaction term ($1/r^{12}$) and attractive long range term ($1/r^6$) belonging to sampled configuration were collected. For the methodology convenience, the normalized and reduced distances were sampled rather than the absolute ones appearing in equation (2) below. In other words, r was normalized by σ (value of r at which $U(r) = 0$) to obtain r^* while energy U was normalized by potential well depth (ε) to get U^* . Furthermore, normalized distances were reduced by the normalized length of the simulation box L^* ($s^* = r^*/L^*$) to represent energy in the final form as shown in (4).

$$U_{ij} = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2)$$

$$U_{ij}^* = 4 \left[\left(\frac{1}{r_{ij}^*} \right)^{12} - \left(\frac{1}{r_{ij}^*} \right)^6 \right] \quad (3)$$

$$U_{ij}^* = 4 \left[\left(\frac{1}{L^* s_{ij}^*} \right)^{12} - \left(\frac{1}{L^* s_{ij}^*} \right)^6 \right] \quad (4)$$

Therefore, the normalized energy of the system and pressure at each configuration can be evaluated based on the sum of pair wise interactions as follows by (5) and (6).

$$U^* = 4 \sum_{i < j} \left[\left(\frac{1}{L^* s_{ij}^*} \right)^{12} - \left(\frac{1}{L^* s_{ij}^*} \right)^6 \right] + N u_{tail}^* \quad (5)$$

$$P^* = \frac{\rho^*}{\beta^*} + \frac{8\rho^*}{N} \sum_{i < j} \left[2 \left(\frac{1}{L^* s_{ij}^*} \right)^{12} - \left(\frac{1}{L^* s_{ij}^*} \right)^6 \right] + p_{tail}^* \quad (6)$$

β^* is the normalized Boltzmann factor and equals to reciprocal of normalized temperature ($T^* = T/(\varepsilon/k_B)$, where k_B is Boltzmann's constant) while normalized number density $\rho^* = \rho\sigma^3$. Tail correction functions for energy and pressure when dealing with LJ fluids can be respectively expressed in terms of the reduced cutoff radius (s_c^*) as

$$u_{tail}^* = \frac{8}{3} \pi \rho^* \left[\frac{1}{3} \left(\frac{1}{L^* s_c^*} \right)^9 - \left(\frac{1}{L^* s_c^*} \right)^3 \right] \quad (7)$$

$$p_{tail}^* = \frac{16}{3} \pi \rho^{*2} \left[\frac{2}{3} \left(\frac{1}{L^* s_c^*} \right)^9 - \left(\frac{1}{L^* s_c^*} \right)^3 \right] \tag{8}$$

2.1. Extrapolation along Isochores

Starting with an MC chain simulated at a certain temperature and density, ensemble averages of quantity A such as energy or pressure could be predicted at a series of neighboring points with different temperatures but same density. This is possible by reweighing the original sampled chain. Reweighting factors (w_k) are related to the temperature or Boltzmann factor (β) difference between the reference case and targeted one. Noting that, w_k is proposed such that the probability distribution function is retained. Equation (10) gives an example how energy could be extrapolated.

$$A(s^*, L_o^*)_{\beta_n^*} = \frac{\sum_{k=1}^{Nsamp} w_k A_k(s_k^*, L_o^*)_{\beta_o^*}}{\sum_{k=1}^{Nsamp} w_k} \quad \text{with} \quad w_k = e^{-\Delta\beta^* U_k^*(s_k^*, L_o^*)_{\beta_o^*}} \tag{9}$$

$$U^*(s^*, L_o^*)_{\beta_n^*} = \frac{4 \sum_{k=1}^{Nsamp} e^{-\Delta\beta^* U_k^*(s_k^*, L_o^*)_{\beta_o^*}} \sum_{i<j} \left[\left(\frac{1}{L_o^* s_{ij}^*} \right)^{12} - \left(\frac{1}{L_o^* s_{ij}^*} \right)^6 \right]}{\sum_{k=1}^{Nsamp} e^{-\Delta\beta^* U_k^*(s_k^*, L_o^*)_{\beta_o^*}}} + Nu_{tail}^*(s_c^*, L_o^*)_{\beta_n^*} \tag{10}$$

2.2. Extrapolation along Isotherms

Extrapolation along isotherms where density differs between reference state and targeted ones requires reconstruction of the MC chain itself in addition to reweighing. In this specific case, the scaled LJ model becomes important. Due to density change both weighting and scaling factors of the sampled properties are changed.

$$A(s^*, L_n^*)_{\beta_o^*} = \frac{\sum_{k=1}^{Nsamp} w_k A_k(s_k^*, L_n^*)_{\beta_o^*}}{\sum_{k=1}^{Nsamp} w_k} \quad \text{with} \quad w_k = e^{-\beta_o^* (U_k^*(s_k^*, L_n^*)_{\beta_o^*} - U_k^*(s_k^*, L_o^*)_{\beta_o^*})} \tag{11}$$

$$U^*(s^*, L_n^*)_{\beta_o^*} = \frac{4 \sum_{k=1}^{Nsamp} e^{-\beta_o^* (U_k^*(s_k^*, L_n^*)_{\beta_o^*} - U_k^*(s_k^*, L_o^*)_{\beta_o^*})} \sum_{i<j} \left[\left(\frac{1}{L_n^* s_{ij}^*} \right)^{12} - \left(\frac{1}{L_n^* s_{ij}^*} \right)^6 \right]}{\sum_{k=1}^{Nsamp} e^{-\beta_o^* (U_k^*(s_k^*, L_n^*)_{\beta_o^*} - U_k^*(s_k^*, L_o^*)_{\beta_o^*})}} + Nu_{tail}^*(s_c^*, L_n^*)_{\beta_o^*} \tag{12}$$

2.3. Combined Temperature and Density Change

Similar to the previous case, when combined change occur both reconstruction and reweighting are required. Moreover, weighting factors will be slightly different as they include both β^* and L^* change.

$$A(\mathbf{s}_k^*, L_n^*)_{\beta_n^*} = \frac{\sum_{k=1}^{Nsamp} w_k A_k(\mathbf{s}_k^*, L_n^*)_{\beta_o^*}}{\sum_{k=1}^{Nsamp} w_k} \quad \text{with} \quad w_k = e^{-(\beta_n^* U_k^*(\mathbf{s}_k^*, L_n^*)_{\beta_o^*} - \beta_o^* U_k^*(\mathbf{s}_k^*, L_o^*)_{\beta_o^*})} \quad (13)$$

3. Results and Discussion

Two data sets were created to test the validity of this method. Using these sets, extrapolation limits were investigated under the effect of different particles number as well as simulation temperatures and densities. Originally and for both sets, analysis was conducted over normalized properties in single-phase region with β^* ranging from 0.1 till 0.7 incremented by 0.1. At each β^* 17 simulations were run with different ρ^* values ranging from 0.10 to 0.90 with 0.05 increments. Cutoff radius was taken as a fraction of 0.49 of the length of the box ($r_c^* = 0.49L^*$).

A total of 61,440,000 MC cycles were performed per simulation while 10% of these cycles were discarded to account for stabilization. A sampling distance of 1024 cycles was adopted leading to a final MC chain of length of 54000 entries. The only difference between the two sets was in number of particles, first one was conducted using $N = 512$ while the other one with $N = 256$. Initially, particles were uniformly distributed in the simulation box.

3.1. Extrapolation with $N = 512$

Normalized energy and pressure have been both estimated along two isochors (Fig. 1) by extrapolating data from series of 7 simulations conducted at β^* ranging from 0.1 to 0.7. From each MC chain simulated, 5 neighboring points in both directions were predicted with an increment of $\Delta\beta^* = 0.01$. At $\rho^* = 0.5$, slight discontinuities appear (Fig.1 Left) suggesting an effect of original thermodynamic conditions on extrapolation ranges.

In order to extend our understanding to reweighting limitations, Fig. 2 was generated. On left part, U^* was estimated starting from same series of points as in Fig. 1 but with $\rho^* = 0.9$ and $\Delta\beta^* = 0.02$. Weighting factors damps sharply away from original point (Fig. 2 Right) leading to loss of information of the residual part of energy. As the tail correction functions are not reweighted and independent of β^* , U^* is represented by tail corrections only. In following section more extensive analysis to extrapolation limits is provided.

The same data set with $N = 512$ was used to test the reconstruction approach proposed by this paper. Similar to previous case U^* and P^* were both reproduced along three different isotherms (Fig. 3). This time, only 2 neighboring points from each direction were estimated with $\Delta\rho^* = 0.01$ between one point and another. Discontinuities at some parts of the isotherms support the previous conclusion that extrapolation limits are affected by the original thermodynamic conditions.

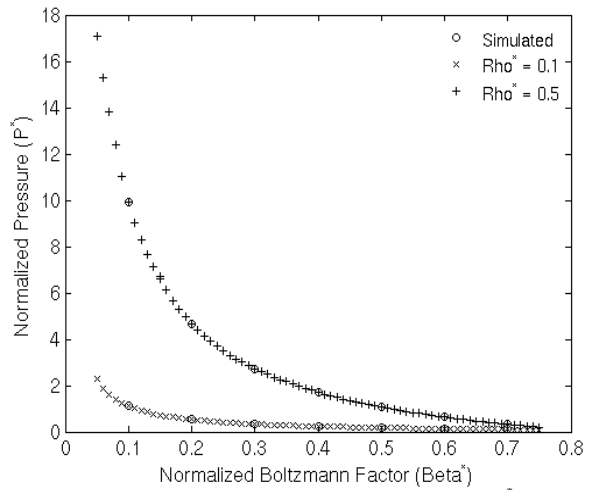
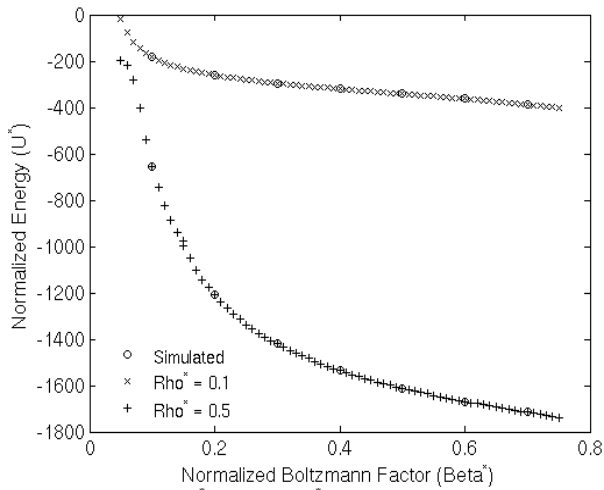


Fig. 1. Estimated U^* (Left) and P^* (Right) values along two isochors, by reweighting series of MC chains simulated at different β^* values; open circles refer to simulated values and origins of extrapolation.

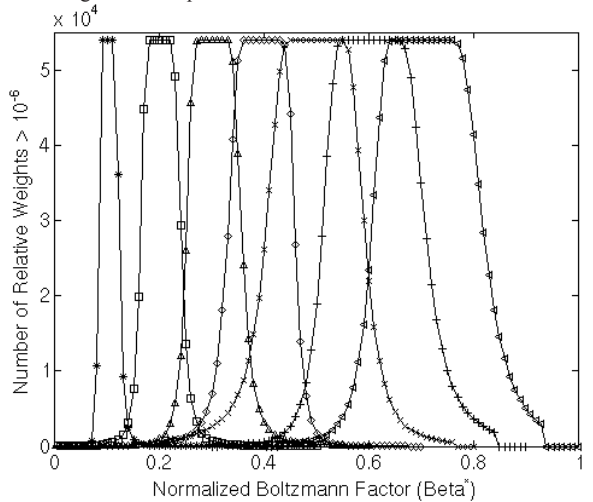
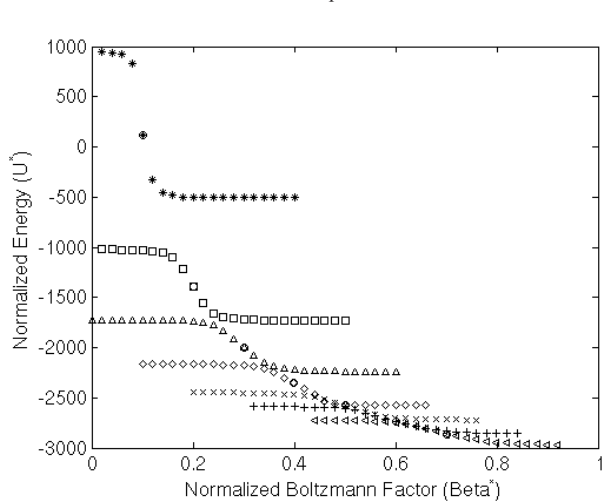


Fig. 2. Estimated U^* (Left) for long extrapolation range in comparison to simulated U^* (o) at $\rho^* = 0.9$ to show reweighting limitations; number of relative weighting factors (Right) which are greater than 10^{-6} damps as targeted β^* becomes further

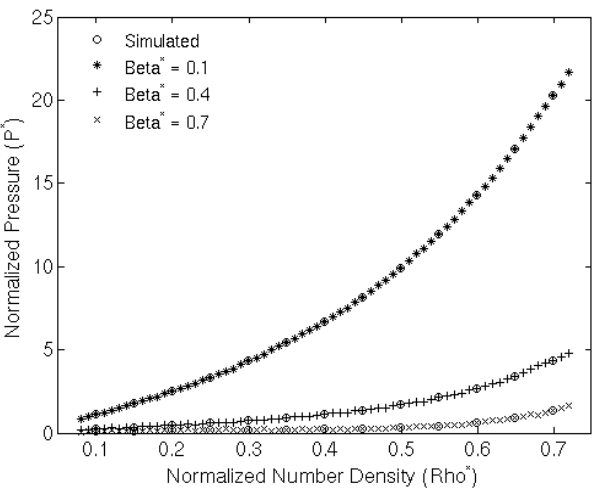
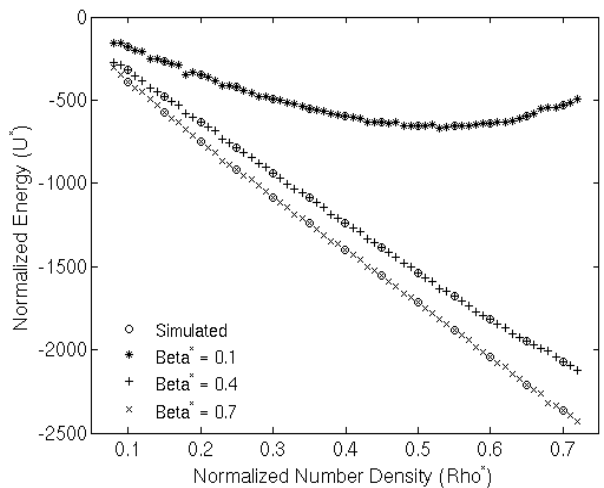


Fig. 3. Estimated U^* (Left) and P^* (Right) values along three isotherms, by reweighting and reconstructing series of MC chains simulated at different ρ^* values; open circles refer to simulated values and origins of extrapolation.

3.2. Extrapolation range limits

Effect of number of particles on extrapolation ranges was investigated with the help of the second set of experiments. As a criterion, the maximum limit in both directions for extrapolation starting from a certain simulation point is the last point where at least 80% of the relative weights corresponding to its chain are greater than 10^{-6} . As a result, extrapolation limits along isochors (Fig. 4) and isotherms (Fig. 5) were generated.

By referring to Fig. 4 and Fig. 5 several deductions could be made. First of all and as expected, prediction power along isochors, where merely reweighting is needed, is much more significant than isotherms where both reweighting and reconstruction are required. Moreover, prediction intervals increased significantly when less number of particles ($N = 256$) was used. Finally, a general trend is realized where ranges extend along isochors at lower ρ^* and higher β^* . On the other hand, along isotherms an opposite trend is detected.

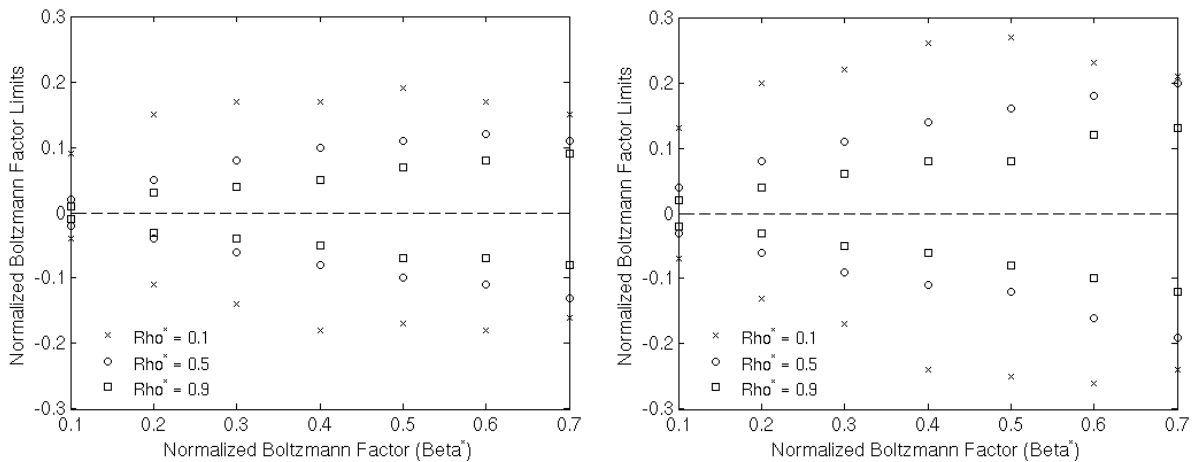


Fig. 4. Extrapolation range limits comparison along three different isochors between $N = 512$ (Left) and $N = 256$ (Right). Dashed line is reference line

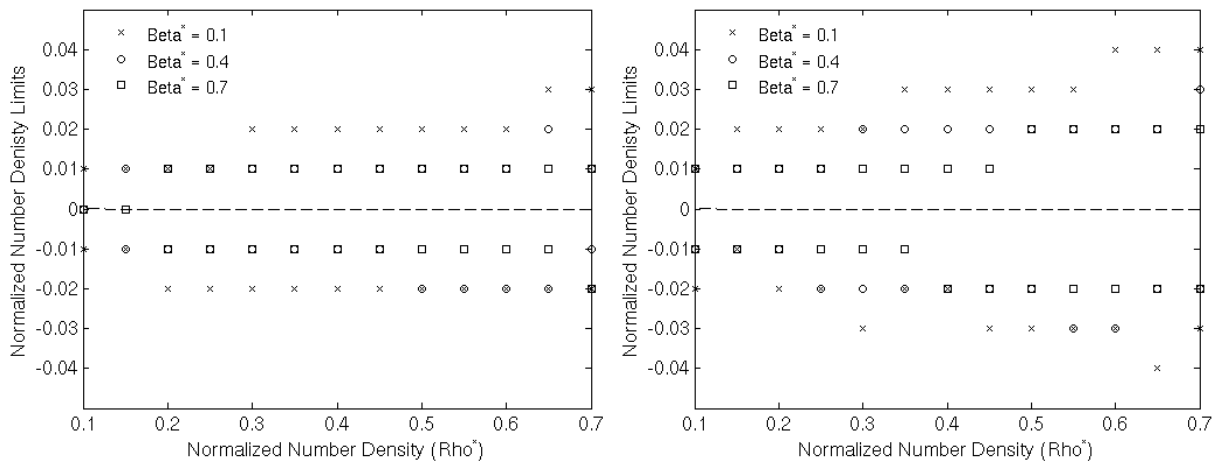


Fig. 5. Extrapolation range limits comparison along three different isotherms between $N = 512$ (Left) and $N = 256$ (Right). Dashed line is reference line

3.3. Absolute extrapolation range limits

Extrapolation limits along same three isochors and three isotherms (Fig. 4 and Fig. 5 respectively) are given an absolute sense by representing them in absolute units rather than normalized ones. For that purpose, $\epsilon/k_B = 147 K$ and $\sigma = 3.723 \text{ \AA}$ methane LJ model parameters [4] were used. Results show stunning extrapolation limits

where prediction of ensemble averages along 51.49 Kg/m³ isochor is possible up to 4900 K temperature and down to 850 K starting from simulation point at $T = 1470$ K (Fig. 6 Right).

In the case of extrapolation along isotherms, the maximum range with respect to considered data in this paper was attained at high densities along high temperature isotherms (Fig. 7). In both systems with $N = 512$ and $N = 256$, the maximum limit was about 20 Kg/m³ in both directions.

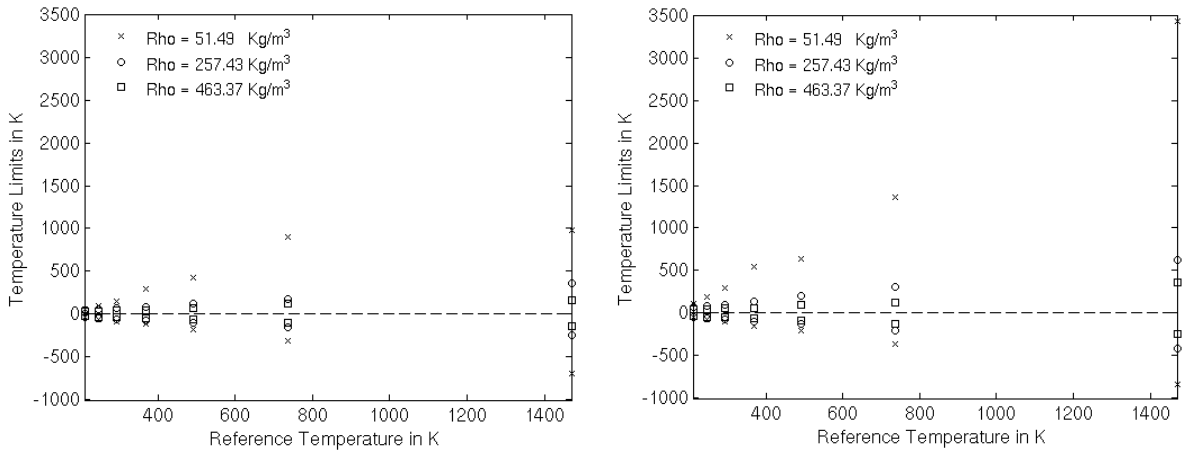


Fig. 6. Absolute extrapolation range limits comparison along three different isochors between $N = 512$ (Left) and $N = 256$ (Right). Dashed line is reference line

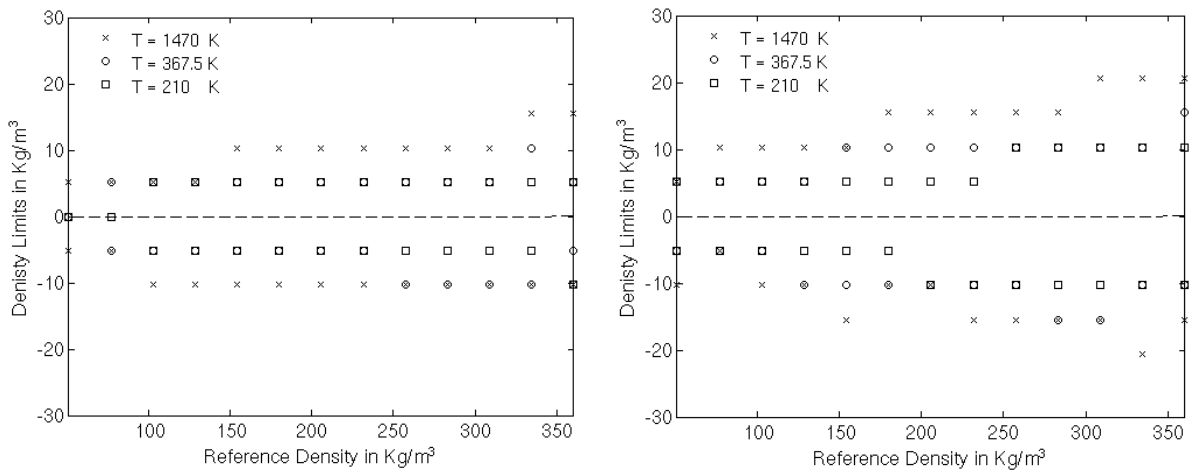


Fig. 7. Absolute extrapolation range limits comparison along three different isotherms between $N = 512$ (Left) and $N = 256$ (Right). Dashed line is reference line

3.4. Extrapolation along paths with changing both β^* and ρ^*

Based on the extrapolation analysis, chains generated by simulations with $N = 256$ were chosen to take this method a step further by extrapolating to points residing different temperatures and densities from the original one. Three isotherms of U^* and P^* (Fig. 8) were created starting with simulated data at the mid isotherm ($\beta^* = 0.4$).

MC chains collected at 10 different ρ^* ranging from 0.15 up to 0.6 at $\beta^* = 0.4$ were reweighted and reconstructed as stated earlier in equation (16). Even though the extrapolation was successful still many discontinuities in generated isotherms appeared. This problem can be explained by the extrapolation limits discussed above; hence we believe that, this problem could be simply solved by creating a finer data bank. Half the distance along β^* and ρ^* increments will be enough to smoothen all the produced curves.

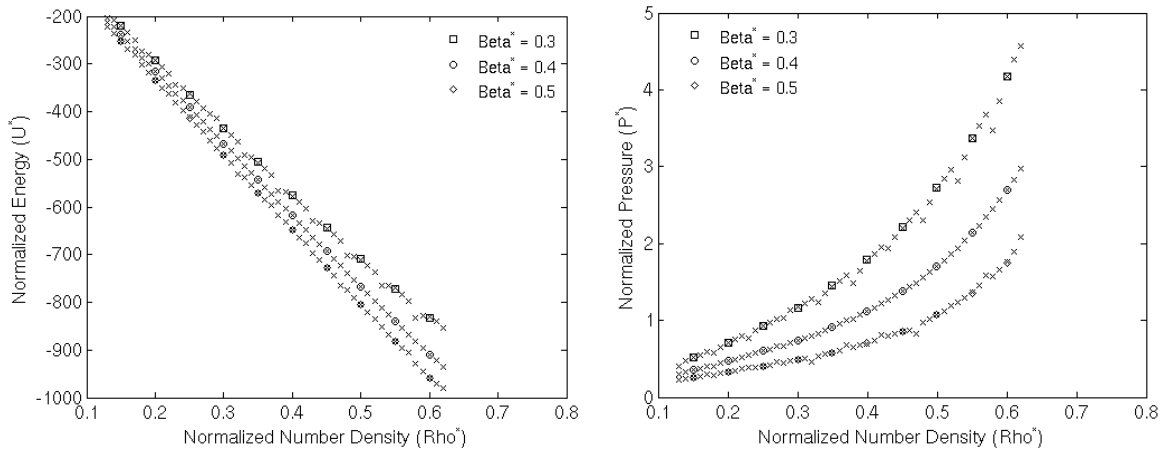


Fig. 8. Estimated U^* (Left) and P^* (Right) values along three isotherms, by reweighting and reconstructing series of MC chains simulated at different ρ^* and β^* ; open circles refer to simulated values and origins of extrapolation along $\beta^* = 0.4$ while open squares and diamonds refer to simulated data at the targeted isotherms ($\beta^* = 0.3$ and $\beta^* = 0.5$) respectively for the sake of comparison.

4. Conclusion

In conclusion, the introduced extrapolation scheme proved its validity and effectiveness by reproducing U^* and P^* along wide range of temperatures and densities. It has been shown that extrapolating along isochors is more powerful than isotherms, as it includes reweighting of the original MC chain solely without any reconstruction. Limits of extrapolation ranges had been investigated concluding that, simulation thermodynamic conditions have a significant effect on these limits. Therefore, one needs to be careful while choosing prediction intervals. Moreover, less number of particles was capable to extend these limits significantly but attention must be paid not to lose correct statistical representation of the system as N decreases.

Work is in progress, to create a finer data bank to use this proposed approach in order to investigate extrapolation ranges limits in more detailed fashion. Preprocessing of the data bank such as reducing storage requirement and improving data quality are also of interest. In addition, application of these schemes to other ensembles as well as including non-Van der Waals interactions (e.g., Columbic interactions) will be considered. Further work is to use this technique in reducing multi-sites LJ models into single site ones by optimizing corresponding LJ parameters.

Acknowledgements

The work presented in this paper has been supported in part by 1) The project entitled "Study of Sulfur Solubility using Thermodynamics Model and Quantum Chemistry", funded by King Abdulaziz City for Science and Technology (KACST), the Saudi Arabian national science agency, and 2) The project entitled "Simulation of Subsurface Geochemical Transport and Carbon Sequestration", funded by the GRP-AEA Program at KAUST.

References

- [1] Ferrenberg AM, Swendsen RH. New Monte Carlo technique for studying phase transitions. *Physical Review Letters*. 1989;63:1658-.
- [2] Kiyohara K, Gubbins KE, Panagiotopoulos AZ. Phase coexistence properties of polarizable Stockmayer fluids. arXiv preprint physics/9610022. 1996.
- [3] Conrad P, De Pablo J. Comparison of histogram reweighting techniques for a flexible water model. *Fluid phase equilibria*. 1998;150:51-61.

- [4] Errington JR, Panagiotopoulos AZ. Phase equilibria of the modified Buckingham exponential-6 potential from Hamiltonian scaling grand canonical Monte Carlo. *The Journal of chemical physics*. 1998;**109**:1093.
- [5] Ghazi SM, Kanhere D. Thermodynamic Properties of Ga₂₇Si₃ Cluster Using Density Functional Molecular Dynamics. *The Journal of Physical Chemistry A*. 2011;**116**:11-7.
- [6] Menzl G, Köfinger J, Dellago C. Phase Transition and Interpore Correlations of Water in Nanopore Membranes. *Physical Review Letters*. 2012;**109**:20602.
- [7] Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*. 1977;**23**:187-99.
- [8] Torrie G, Valleau J. Monte Carlo study of a phase - separating liquid mixture by umbrella sampling. *The Journal of chemical physics*. 1977;**66**:1402-8.
- [9] Valleau J. Density-scaling: a new Monte Carlo technique in statistical mechanics. *Journal of Computational Physics*. 1991;**96**:193-216.
- [10] Valleau JP. The Coulombic phase transition: Density - scaling Monte Carlo. *The Journal of chemical physics*. 1991;**95**:584-9.
- [11] Valleau JP. Density - scaling Monte Carlo study of subcritical Lennard - Jonesium. *The Journal of chemical physics*. 1993;**99**:4718.
- [12] Valleau JP. Temperature-and-density-scaling Monte Carlo: methodology and the canonical thermodynamics of Lennard-Jonesium. *Molecular Simulation*. 2005;**31**:223-53.
- [13] Valleau JP. Temperature-and-density-scaling Monte Carlo: isothermal–isobaric thermodynamics of Lennard-Jonesium. *Molecular Simulation*. 2005;**31**:255-75.
- [14] Kiyohara K. Thermodynamic scaling Gibbs ensemble Monte Carlo: a new method for determination of phase coexistence properties of fluids. *Molecular Physics*. 1996;**89**:965-74.
- [15] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *The Journal of chemical physics*. 1953;**21**:1087.