Sampling Rare Events in Hybrid Statistical Physics – Quantum Mechanical Studies of Hydrophobic Interactions in Liquids: A GPU Implementation

Petre Lameski[1], Ljupčo Pejov[2], Leonid Djinevski[3], Dimitar Trajanov[1], Sonja Filiposka[1]

[1] Ss. Cyril and Methodious University, Faculty of Computer Science and Engineering,
Rugjer Boshkovikj 16, 1000 Skopje, Macedonia
`{petre.lameski,dimitar.trajanov,`
`sonja.filiposka}@finki.ukim.mk`
[2] Ss. Cyril and Methodious University, Institute of Chemistry,
Arhimedova 5, 1001 Skopje, Macedonia
`ljupcop@pmf.ukim.mk`
[2] FON University,
Av. Kiro Gligorov 5, 1000 Skopje, Macedonia,
`leonid.djinevski@fon.edu.mk`

Abstract. A method for sampling rare events throughout hybrid statistical physics quantum mechanical studies of hydrophobic interactions in liquid phases was analyzed and implemented on GPU architecture. The rare events considered in the present case are related to formation of the so-called dangling X-H bonds of small polar solvent molecules around notably non-polar solute molecules. The approach is based on detecting various populations of polar solvent molecules which reside in the first solvation shell around the non-polar solute and noting the differences between such populations. In particular, the search for structurally different populations of solvent molecules around the solute is started by detecting the differences in vibrational spectroscopic characteristics of the X-H stretching frequencies, which could be computed by various approaches, such as, e.g. within the Stark shift approximation. The results show x25 overall speedup of the GPU implementation compared to the sequential implementation.

Keywords: Rare Events, Hydrophobic Interactions, Statistical Physics Simulations, GPU, CUDA.

1 Introduction

Hydrophobic intermolecular interactions have become increasingly popular in the last decade. The relevance of these noncovalent forces arises from the fact that they occur in realistic systems, which are particularly relevant to biomedical sciences. For example, most of the biologically relevant molecular systems are characterized with a notable hydrophobic part, while practically all biochemical processes occur in aqueous medium, which is a physiological medium in biological system. Thus, the interactions between water as a "solvent" and the "hydrophobic" part of "biomolecules" serve as prototypical and at the same time realistic examples of hydrophobic intermolecular interactions. While these non-covalent interactions are often considered as "non-bonding forces", such statement is not fundamentally true. Hydrophobic interactions are simply a particular class of non-covalent interactions, which can result in small interaction or dissociation energies, but at the same time characteristic mutual arrangements of atoms. In the present study, we address a particularly interesting issue related to typical "rare events" which occur in aqueous condensed media in which biomolecular systems are realistically embedded. As a typical example, we consider the phenomenon of occurrence of "dangling" X-H bonds from the solvent, directed towards the hydrophobic part of a typical biomolecule. Such directional properties of the X-H bond in question are quite unexpected from the viewpoint of "chemical intuition", but have been recently confirmed to exist in realistic systems by combination of certain experimental and computational techniques [Silvi, B., Wieczorek, R., Latajka, Z., Alikhani, M. E., Dkhissi A., Bouteiller, Y.: Critical analysis of the calculated frequency shifts of hydrogen-bonded complexes, J. Chem. Phys., 111, 6671-6678 (1999). Tomlinson-Phillips, J.; Davis, J.; Ben-Amotz, D.; Spångberg, D.; Pejov, Lj.; Hermansson, K.: Structure and dynamics of Water Dangling OH Bonds in hydrophobic Hydration shells. Comparison of Simulation and Experiment, J. Phys. Chem. A, 115, 6177-6183 (2011)]. We actually present a general generic algorithm for detection of certain population of solvent molecules that have these certain characteristic, on the basis of molecular dynamics simulation trajectories and predicted spectroscopic properties of the X-H oscillators in question. The algorithm has also been implemented for GPU computational architecture. The algorithm has also been implemented for parallel computational architecture. More particularly the implementation harnesses the tremendous computational power of the GPU. Much success of significant speedup has been done in many research areas [2]. In this paper we are presenting additional prove of efficient utilization of GPU technology in the area of chemistry.

This paper is organized in 6 sections. In Section 2 we present an overview of the proposed module, followed by its parallel implementation described in Section 3. The used methodologies for testing of the parallel implementation are presented in Section 4, while the obtained results were presented in Section 5. The conclusion is given in Section 6.

2 Computational Details and Algorithms

The basic idea of the developed algorithm is based on the following sequence. First, a molecular dynamic (MD) simulation of the system of interest is carried out, usually imposing periodic boundary conditions. Of course, MD simulation could be a classical or *ab intio* one, as well as Car-Parinello MD or dynamical simulation based on the atom-centered density matrix propagation scheme (ADMP). Further on, the X-H stretching frequencies of the solvent molecules (O-H in cases when solvent is water) are predicted. This can be done using various approaches. Perhaps the simplest one is to use the electric field Stark shift approximation. This approximation is actually based on the correlation between the strength of the local electric field at the X-H oscillator site and the experimental vibrational frequencies in a particular medium [1], [3], [5]. The O-H (or X-H in a general case) stretching frequency value is a first indicator of the involvement of that particular X-H bond in certain "unusual" (or "strange") intermolecular interactions within the condensed phase (liquid in the presently studied case). For example, due to the so-called "wall-effects", which manifests as a steepening of the X-H stretching potential, involvement of an O-H oscillator of a particular water molecule in a hydrophobic interaction often leads to upshift (*i.e.* blue shift) of the O-H stretching frequency of that particular oscillator.

However, the opposite is not true by default. In other words, O-H stretching frequency blue shift does not necessarily mean that the particular O-H oscillator is involved in a hydrophobic interaction, *i.e.* e.g. is a "dangling" O-H bond (out of the hydrogen bonding network formed by the solvent molecules). Formation of dangling O-H (or X-H in a general case) bond is certainly a "rare event" within the condensed phase, so one expects that a population of water (or solvent in a general case) molecules should be a minority one. Further, upon detection of potentially "interesting" X-H oscillators, the distances between X and H atoms from those particular oscillators and the particular centers of the molecule characterized by a significant hydrophobic segment (the "biomolecule") are computed and the corresponding distribution histograms are generated. The same distance distribution histograms are also generated for water molecules characterized by quite "ordinary" O-H oscillators. Certain differences of the peaks and widths of such distributions may serve as a straightforward indicator of unusual hydrophobic interactions, or rare events such as formations of "dangling" O-H bonds towards the hydrophobic biomolecule. Of course, afterwards one may visually inspect the geometric features of the particular solvent molecules that exhibit "unusual" frequency and bonding feature and directly detect the rarely occurring in-liquid structural features. The algorithm workflow is schematically shown in Fig. 1.

3 Parallel Implementation

In this section we present the parallel implementation developed with the CUDA programming model [6]. The parallelization of the described algorithm in Section 2 was performed for the three of the following segments:

1. Parallelization of the Euclidian distance calculation between each atom of the neopentane and each atom of molecule of water.
2. Parallelization of the minimum distance calculation between the nearest atoms of each molecule of water with the neopentane atoms
3. Parallelization of the histogram of distances calculation

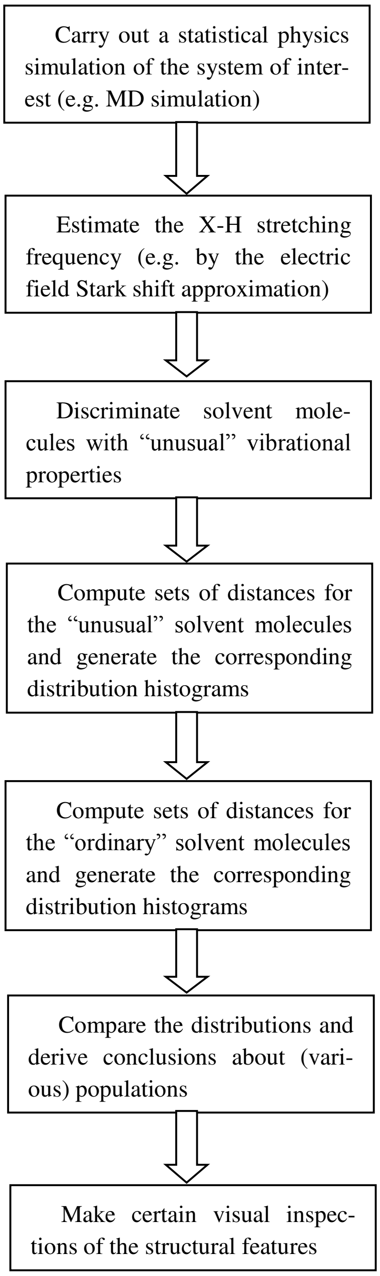


Fig. 1. A schematic presentation of the algorithm workflow.

The distance calculation between each atom of the water and neopentane molecule can be viewed as an atomic operation. The parallelization is strait forward, since each distance can be calculated by a separate thread without any data dependencies. Of all distances between each atom in a water molecule and each atom in the neopentane molecule we are interested in the minimal distance. The minimal distance reduction is implemented with the standard binary tree reduction [4].

The total number of needed Euclidian distance calculations is 500 molecules of water times 3 atoms for each molecule times 17 atoms of the neopentane molecule which makes a total of 25500 comparisons or 51 comparisons for each molecule of water in the simulation. We calculate the eucledian distance between each atom in each molecule, but for the interaction, we need the closest distance between the closest atom of water with the closest atom of neopentane. To do this we need to find the minimum of the 51 distances found between each of the atoms. To find the min distance we can use the binary tree minimal distance reduction.

To achieve the greatest speedup for the minimal distance reduction, we use the shared memory of each block to save the distance and the minimum index of the nearest points (nearest atoms between two molecules). Each block consists of 64 threads so we use for each atom only 51 thread and the rest of the threads are neglected and an arbitrary very large distance is written as their calculated distance. This allows us to achieve fastest reduction. In this way we calculate the minimum distance between each water molecule and the neopentane molecule. The parallelization of the histogram calculation of the minimum distances is done with a modified standard approach given in [8]. We calculate the histogram of minimum distances between the molecules of water and the molecule of neopentane. The histogram of distances is given in the expected range of distances between 0 and the maximum detected distance in the dataset. The number of bins used for the histogram is 64 which gives best performance and good enough resolution for the distribution of distances.

4 Testing Methodology

In this section we illustrate the testing methodology that was used. The sequential and the parallel implementations were executed on the same mashine. The GPU device in the system is consisted of one Nvidia GeForce GTX 680, while the installed CPU on the same system is Intel Core i7 3770 CPU 3.40GHz, and 32GB of RAM @ 16GHz. The Nvidia GeForce GTX 680 has 1536 cores, 192.2GB/s memory bandwidth, and 2GB of GDDR5 standard memory capacity. The implementation was compiled with the NVIDIA compiler *nvcc* from the CUDA 5.0 toolkit, and run on Ubuntu 12.04 LTS operating system. The sequential implementation was written in C++, using *gcc* standard libraries.

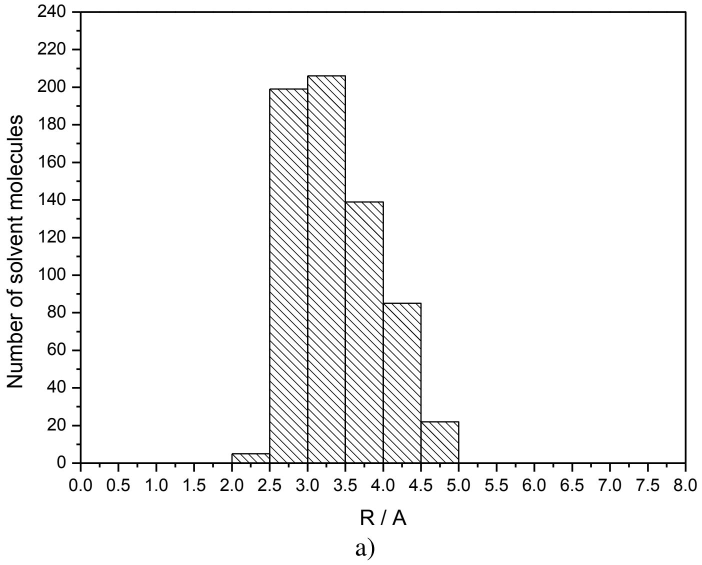
The testing methodology is based on an experiment for our algorithm. Since the GPU device is a Kepler architecture [7], we configure the L1 cache memory with 16, 32 and 48KB, and the cache line size with 32 and 128B, in order to obtain the best results.

5 Results and Discussion

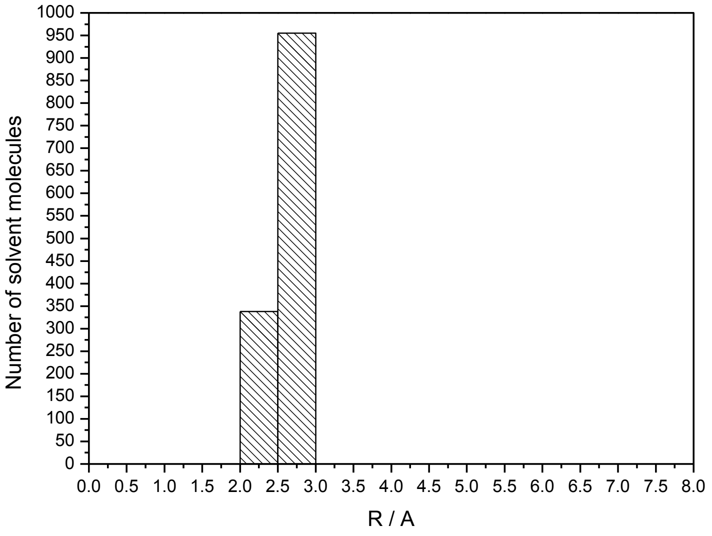
This section presents the results that were obtained for the experiment defined in Section 4.

We consider a typical example where our algorithm is applied, in a generic way. We consider a highly non-polar organic molecule (a saturated hydrocarbon molecule) dissolved in water. A rare event that we seek for is the formation of a dangling O-H bond pointing towards the hydrocarbon molecule. After the MD simulation carried out by the standard high-quality force field (OPLS for the hydrocarbon, SPC for water), accounting for all specificities of this statistical physics simulation, we analyze the nearest-neighbor water molecules to the organic solute, i.e. the water molecules residing in the first hydration shell, as determined from the analysis of the corresponding radial distribution function. After carrying out the analysis as depicted in Figure 1, the distance distribution histograms for the ordinary and unusual water molecules are obtained. These are shown in Figure 2 and 3. As can be seen, the two distributions are notably different in both the shape, width and center, indicating an existence of two substantially different populations of water molecules around the solute. The smaller average distance exhibited in the second case is the origin of the wall-effect, leading to O-H stretching frequency blue shift.

From the computational point of view, the total execution time of the parallel implementation described above is approximately 25 times faster than the serial implementation. The first and the second parallel segment were 40 times faster than the sequential, while the parallel histogram segment was 30 time faster than the sequential.



a)



b)

Fig.1. The distance distribution histograms for the "ordinary" and "unusual" water molecules with the outlined algorithm.

6 Conclusion and Future work

The intention of this paper was to show a parallel implementation of our algorithm for execution on GPU processors. The output analysis of our implementation is presented and conclusion of existence of two substantially different populations of water molecules around the solute was provided. By using this parallel implementation, the necessary time for performing the analysis was significantly decreased. The algorithm is performing the calculation in value number range that is with-in the scope of the standard float data type. Using floats for the GPU processing, contributes towards more optimal implementation compared to using of doubles.

The results show an overall speedup of x25, which provides another justified example for adopting of high-end GPU devices for intensive computing.

As future work we plan to further our research in the area of data structures (in order to find more efficient data structure that can map the atoms of neopentane and water molecules). Additionally, we want to test out the execution of our implementation on a multi-GPU system, and in a distributed parallel system, utilizing both MPI and CUDA.

References

1. Buckingham, A.: Solvent effects in infra-red spectroscopy. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 248(1253), 169-182 (1958)

2. Che, S., Boyer, M., Meng, J., Tarjan, D., Sheaffer, J., Skadron, K.: A performance study of general-purpose applications on graphics processors using cuda. Journal of parallel and distributed computing 68(10), 1370-1380 (2008)
3. Corcelli, S., Skinner, J.: Infrared and raman line shapes of dilute hod in liquid h2o and d2o from 10 to 90 c. The Journal of Physical Chemistry A 109(28), 6154-6165 (2005)
4. Harris, M., et al.: Optimizing parallel reduction in cuda. NVIDIA Developer Technology 2, 45 (2007)
5. Hermansson, K., Lindgren, J., Probst, M.M.: Non-additivity of oh frequency shifts in ion-water systems. Chemical physics letters 233(4), 371-375 (1995)
6. Nvidia, C.: Compute unified device architecture programming guide (2007)
7. Nvidia, C.: Next generation cuda compute architecture: Kepler gk110 (2012)
8. Podlozhnyuk, V.: Histogram calculation in cuda. NVIDIA Corporation, White Paper (2007)
9. Silvi, B., Wieczorek, R., Latajka, Z., Alikhani, M., Dkhissi, A., Bouteiller, Y.: Critical analysis of the calculated frequency shifts of hydrogen-bonded complexes. The Journal of chemical physics 111(15), 6671-6678 (1999)
10. Tomlinson-Phillips, J., Davis, J., Ben-Amotz, D., Spangberg, D., Pejov, L., Hermansson, K.: Structure and dynamics of water dangling oh bonds in hydrophobic hydration shells. comparison of simulation and experiment. The Journal of Physical Chemistry A 115(23), 6177-6183 (2011)