Maximum probability bassins for the analysis of atomic densities from molecular dynamics simulations

The structure of liquids is usually analyzed by studying the reduced 2-particles density [1], representing the probability of finding two particles with positions in a given volume of the three-dimensional space, irrespective of the positions of all other particles. The calculation of this function for a homogeneous system allows to construct density maps to identify the regions of space where it is more likely to find one particle in the neighborhood of a given fixed particle (chosen as origin of the reference frame). However, sometimes these distributions of probability are not able to fully characterize the structure of the liquid. Indeed, complete information is contained in the n-particles densities, with $n > 2$, but their representation in the three-dimensional space is difficult. The purpose of our work is to introduce a method to investigate the structure of a classical liquid by identifying important structures of the n-particle densities.

The aim of the visit was the discussion about the basic formulation of the problem and about the information one can obtain by performing the proposed analysis.

A similar problem arises in electronic structure calculations for the analysis of the N-electrons wavefunction of a molecular system. We propose to apply the methods introduced for the analysis of electronic structures to quantitatively characterize the partitioning of space around a given particle in the liquid. Our starting point is the topological analysis, so-called Bader analysis, of the 2-particles density, from a molecular dynamic simulation of a classical liquid. Our proposal is then to combine Bader analysis [2, 3] with a domain optimization procedure [4] in order to find the regions of space where it is more likely to find one and only one particle in the neighborhood of a fixed particle. This analysis would allow for a detailed characterization of the solvation structure in three-dimensional space and the problem will be treated as a n particles problem because our approach takes into account the positions of all particles around a chosen one. Bader analysis is a technique used for electronic structure calculations to partition the space in domains which are associated to the maxima of the electronic density. Similar domains can be identified from the 2-particles density computed for the classical liquid. The domains are then optimized in order to maximize the probability of finding one and only one particle in each domain.

During our visit, we had the opportunity of discussing about the rigorous formulation of the optimization problem in terms of functional derivatives, about the algorithm used for such an optimization and about the physical meaning of the domains. The probability of finding one and only one particle in a certain region Ω of space, with all particles outside that region, depends on a *characteristic function* $\chi_\Omega(\mathbf{r})$, which determines if the particle in the position $\mathbf{r}$ is inside the domain

Ω or not. The introduction of the characteristic function allows to sample the probability by using standard molecular dynamic simulations. Calculating the functional derivative of the probability $P[\chi_\Omega]$, with respect to the characteristic function, gives a condition which has to be satisfied, at the border, by the optimal domain. Numerically, this problem has been treated on the basis of the shape and topological derivatives, referring once again to functional analysis. The level set method [5, 6] is thus introduced, by defining the level set function which identifies the border of the domain. The shape and topological derivatives of the functional (probability, in our case) give a prescription on how the level set function, and consequently the border, has to be varied in order to converge to the maximum of the functional. During this visit, it was then possible to express rigorously the optimization problem which allowed us to design an efficient optimization algorithm. Preliminary results on liquid water have been obtained in order to test the validity of this analysis and the used numerical approach.

Some questions are still opened, concerning the physical meaning of the calculated domains and some useful quantities which is worth computing. We applied this method, in a preliminary version, to the identification of the domains in the first solvation shell of water. The calculation of the four largest domains, identifying the tetrahedral structure of water around a water molecule, is consistent with what could be expected. The definition of localization bassins however, paves the way for a study of the molecular rearrangments in the solvation shell. Furthermore, if we study the domains a bit further from the central molecule (interstitial spaces of the first solvation shell), the structure of water becomes less symmetric and we expect to obtain non-trivial information about the localization of water. When the domains are identified, we can calculate the average occupation of each domain, the fluctuation of the number of particles or the time-correlation function of the occupation.

The results of this preliminary analysis will be presented during a conference and we plan to write an article once final results will be obtained.

References

[1] J.P. Hansen and I.R. McDonald. *Theory of simple liquids*, Elsevier Ltd., II edition (1986)

[2] G. Henkelman, A. Arnaldsson and H. Jónsson. *Comput. Mat. Sc.* **36** 354 (2006)

[3] W. Tang, E. Sanville and G. Henkelman. *J. Phys.: Condens. Matter* **21** 084204 (2009)

[4] A. Scemama, M. Caffarel and A. Savin. *J. Comput. Chem.* **28** 442 (2007)

[5] S. Osher and F. Santosa. *J. Comput. Phys.* **171** 272 (2001)

[6] S. Osher and J.A. Sethian. *J. Comput. Phys.* **79** 12 (1988)