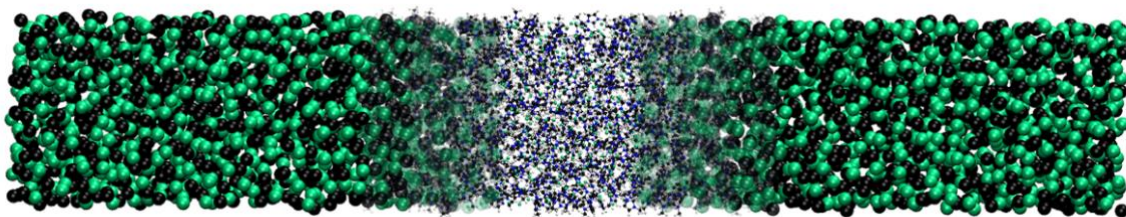




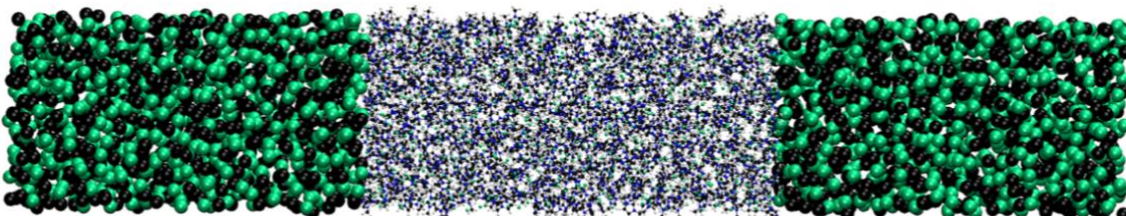
E-CAM Case Study

The development of the GC-AdResS scheme:

from smooth coupling



to a direct interface (abrupt)



With Dr. Christian Krekeler, Freie Universität Berlin

Abstract

GC-AdResS is a technique that speeds up computations without loss of accuracy for key system properties by dividing the simulation box into two or more regions having different levels of resolution, for instance a high resolution region where the molecules of the system are treated at an atomistic level of detail, and other regions where molecules are treated at a coarse grained level, and transition regions where a weighted average of the two resolutions is used. The goal of the E-CAM [GC-AdResS pilot project](#) was to eliminate the need of a transition region so as to significantly improve performance, and to allow much greater flexibility. For example, the low resolution region can be a particle reservoir (ranging in detail from coarse grained to ideal gas particles) and a high resolution atomistic region with no transition region, as was needed hitherto. The only requirement is that the two regions can exchange particles, and that a corresponding “thermodynamic” force is computed self-consistently, which it turns out is very simple to implement.





Could you please describe the goals of your pilot project

The goal of the pilot project (in a nutshell) was to establish GC-AdResS (Grand Canonical Adaptive Resolution Scheme) as a standard method, and to make it easier to use with any MD engine. The key idea of the adaptive resolution technique (AdResS)¹ is to partition the simulation box in three regions, namely: a region at atomistic resolution (AT region), a region at coarse-grained resolution (CG region), and, as an interface between AT and CG, a hybrid region (Δ region). Such an algorithm can be adapted to various situations, e.g. it is possible to use path-integral simulations inside the AT region or to study the aggregation of large molecules and the role of the solvent. It is also possible to thermalize only part of the simulation box and study the dynamics in the AT region. Even large polymer chains can be studied.

Despite the fact that this algorithm has been proven to be computationally robust for several highly challenging applications (see e.g. refs. in ²), in its original implementation the 3 regions mentioned above are coupled via a force term which in the Δ region is switching smoothly between the CG forces and the AT forces. The problem with that technical realization, is that this interpolation has to be done within the evaluation of the force kernel, which is normally the most optimized part of an MD package. As such, the performance of the MD code is significantly reduced.

Our new approach is rather direct, we divide the simulation box in two regions (CG and AT). We still need the external force, the thermodynamic force, and the region where we apply this force we define as Δ region (results in that region are artificial and cannot be used, which is similar to the old GC-AdResS implementation). However, due to that approach everything else, the definition of the regions, the external force, a “capping” force to avoid non-physical configurations at the interface, can be set up outside the force kernel of any MD code. This increases the performance of the code and makes it simpler to implement in new codes and md engines.

Why is this important scientifically, and from an industry perspective in particular ?

The new developed approach is a proof of concept. It shows a roadmap on how to implement AdResS in any MD code available, as all the tools to use it are already present in the MD package³.

Furthermore, the removal of AdResS from the central force calculation inside the MD code, means that the AdResS-enhanced version automatically takes full advantage of further optimizations of the latter. Thus, it is independent of the hardware architecture, and does not hinder the code’s performance and scalability. The reduction of the simulation time is of course, inherently useful in order to provide cheaper calculations.

¹ M. Praprotnik, L. Delle Site, and K. Kremer, J. Chem. Phys. 123, 224106 (2005); M. Praprotnik, L. Delle Site, and K. Kremer, Annu. Rev. Phys. Chem. 59, 545 (2008); M. Praprotnik, L. Delle Site, and K. Kremer, Phys. Rev. E 73, 066701 (2006)

² M. Praprotnik, S. Matysiak, L. Delle Site, K. Kremer, and C. Clementi, J. Phys.: Condens. Matter 19, 292201 (2007); S. Matysiak, C. Clementi, M. Praprotnik, K. Kremer and L. Delle Site, J. Chem. Phys. 128, 024503 (2008); B. P. Lambeth, C. Junghans, K. Kremer, C. Clementi, and L. Delle Site, J. Chem. Phys. 133, 221101 (2010); A. Poma and L. Delle Site, Phys. Rev. Lett. 104, 250201 (2010)

³ C. Krekeler, A. Agarwal, C. Junghans, M. Praprotnik, L. Delle Site, J. Chem. Phys. 149, 024104 (2018)





Then, it is also important to mention that this concept is technically universal. It can be used for example to couple Quantum Mechanics with Molecular Dynamics and Molecular Dynamics to continuum. As another byproduct, we found that the CG potentials are not necessary and can be replaced by non-interacting ideal gas particles. The only requirement that the reservoir has to fulfil is to be at the same density as the AT region (for more details see Ref. ⁴). In simulations using the previous GC-AdResS formulation, the CG potentials had to be set up at the same thermodynamic state point and reproduce structural properties of the atomistic system. Now, we get the same results by using non-interacting particles, which means that we have simplified a previously demanding GC-AdResS method to its bare essentials.⁹

The industrial significance of this project are: (a) it allows in principle larger systems to be simulated efficiently with critical atomistic details to be handled at a much lower computational cost; and, (b) it facilitates the simulation of problems in physics, chemistry and biology as open-systems thereby capturing making them more realistic and relevant.

Can you elaborate a little bit more on your approach and explain to what extent it is built on what was there before and to what extent there are new things?

In the old GC-AdResS implementation (in GROMACS) the 3 neighbour list groups (AT, CG and Δ) were all treated separately, and for each, a different kernel was called. The kernel for the hybrid region is a generic force kernel and not the highly optimized machine code force kernel, normally used in GROMACS. There the atomistic and coarse grained forces of the particles in the hybrid region were calculated and weighted via a smooth weighting function according to its position in the hybrid region. Thus, a smooth transition was guaranteed and particles could slowly adjust to the new environment (no particles overlap). That kernel was a major bottleneck and interfering at that stage destroyed the performance of the MD code.

From a technical point of view, we also removed the neighbor list of the hybrid region and the generic AdResS force kernel, but kept the external thermodynamic force. And we move the force “capping” (i.e. the rescaling of the force if it exceeds a certain threshold) from the force kernel to the stochastic dynamics integrator routine. With these changes we get a considerably more efficient AdResS, still within the theoretical background of GC-AdResS but with one major bottleneck less.

We based the new implementation on the existing GC-AdResS code in GROMACS version 5.1.0 to 5.1.5. But it can be employed in any other MD packages such as for e.g. LAMMPS and Espresso++. The coupling to the continuum is also being done in the code of [HALMD](#).

Can you outline the software that you developed during your pilot project?

The major contribution is the abrupt AdResS code. This code contains the changes discussed above. The second larger contribution added to the abrupt coupling scheme, was the local thermostat version, where we thermalize only the hybrid and particle reservoir (CG) region. That means we have a NVE like environment in the AT region, thus we can also study the dynamics (i.e. kinetics) of a system. The remaining modules are tools which help the run AdResS and analyze the resulting data. Everything can

⁴ L. Delle Site, C. Krekeler, J. Whittaker, A. Agarwal, R. Klein and F. Höfling, Advanced Theory and Simulations 2, 1900014 (2019)





be found on E-CAM project deliverables D4.3⁵ and D4.4⁶, and on E-CAM Software Library for Meso- and Multi-scale Modelling simulations at <https://e-cam.readthedocs.io/en/latest/Meso-Multi-Scale-Modelling-Modules/index.html> .

Does your approach scale well or has potential to scale well on a massively parallel machine? And how much does it accelerate simulations of large systems compared with a pure atomistic system?

We considerably improved the performance compared to the old GC-AdResS implementation and also compared to full atomistic simulations (by a factor of 2.5 improvement). Further tests conducted by E-CAM (see E-CAM deliverable D7.6⁷) showed that there is a load imbalance issue on massive parallel machines, and that is why the performance improvement was not superior. The ALL (A Load Balancing Library) currently being developed by E-CAM partners in Juelich will probably resolve this, and further enhance performance .

Are your codes being applied to systems of interest?

We applied it to water, and to ionic liquids systems (Ref.⁸). Another study looked at the potential of mean force of the aggregation of micelles in water using GC-AdResS (Ref. ⁹). Furthermore, since our latest paper (Ref.¹⁰), several scientific groups have become interested in applying our innovation to their systems. In a DFG special research field project (CRC 1114) the method we developed is used as a basis to couple particle based simulations to a continuum. Two PhD students are employed to include AdResS in HAL-MD. This code runs on GPU's and can simulate continuum mechanics. Another PhD student (DFG funded) is currently developing a path integral extension, including application to several water models. Matej Praprotnik (National Institute of Chemistry, Ljubljana) is also switching to this setup.

Can people add functionalities to your codes and what would they need to do?

Of course, the idea of our approach is to provide a basic recipe on how to set up AdResS in any given code.

What do you think are the future prospects and likely impact of this project?

Scientific groups have picked up the method to look at e.g. the impact of long range vs. short range interactions. The advantage of this new system is that the reservoir does not have any potential or interaction with the atomistic. That means that the impact of structural properties due to the direct surrounding can be studied.

⁵ B. Duenweg, J. Castagna, S. Chiacchiera, H. Kobayashi, and C. Krekeler, "Meso- and multi-scale modelling E-CAM modules II", Mar. 2018. <https://doi.org/10.5281/zenodo.1210075>

⁶ B. Duenweg, J. Castagna, S. Chiacchiera, and C. Krekeler, "Meso- and multi-scale modelling E-CAM modules III", Jan. 2019. <https://doi.org/10.5281/zenodo.2555012>

⁷ A. O'Cais, J. Castagna, "E-CAM Software Porting and Benchmarking Data III", April 2019 <https://doi.org/10.5281/zenodo.2656216>

⁸ C. Krekeler, A. Agarwal, C. Junghans, M. Praprotnik, L. Delle Site, J. Chem. Phys. 149, 024104 (2018)

⁹ B. Shadrack Jabes, R. Klein and L. Delle Site, Advanced Theory and Simulations, 1, 1800025 (2018)

¹⁰ L. Delle Site, C. Krekeler, J. Whittaker, A. Agarwal, R. Klein and F. Höfling, Advanced Theory and Simulations 2, 1900014 (2019)

