Efficient simulation techniques for solution dynamics and precipitation processes

Fabio Pietrucci (Sorbonne Université, IMPMC) & Sara Bonella (CECAM-EPFL Lausanne)

Scientific context

Computer simulations of materials, nanostructures, liquids and biological systems play a prominent role, advocated as the 3rd pillar of science together with theory and experiments, in modern research in physics and chemistry. However, predictive models of realistic systems entail a computational cost that is usually large, and often prohibitive, due to the need to include from thousands to millions of atoms, simulation times well beyond nanoseconds, and interatomic forces consistent with quantum mechanics. For these reasons, there is a large and growing interest in reduced models able to drastically reduce the number or numerical cost of degrees of freedom to be explicitly included in the equations of motion. Among available strategies we recall approximate force fields (often belonging to the machine-learning family) that circumvent Schrödinger's equation to compute forces, possibly coarse-graining groups of atoms as single entities, or implicit solvent methods, replacing the solvent in a solution with effective solute-solute interactions. The present proposal addresses timely and still open scientific questions in the latter field, from the original viewpoint of the comparison between constrained dynamics and stochastic dynamics, targeting ambitious and relevant phenomena and aiming at developing new theoretical and computational approaches, while providing advanced training in statistical mechanics and molecular dynamics, and taking full advantage of a Germaine de Stael grant and of interactions with the ISCD-funded MAESTRO interdisciplinary research team.

Theoretical and computational problem

Compared to the simpler effort of describing stationary statistical properties, the construction of reduced dynamical models of complex atomistic systems is a challenging task: one needs to design approximations of the equations of motion that drastically reduce the number of degrees of freedom or the cost of their evolution, while at retaining an accurate description of dynamical and physico-chemical observables. Reducing the dimensionality / complexity of the mathematical problem can be achieved via strategies in the Born-Oppenheimer or Langevin families, for example. The first approach was followed by one of the proponents (Sara Bonella) [Coretti2018]. In Born-Oppenheimer approaches (both for ab initio molecular dynamics (MD) and realistic models of polarization described by classical force fields), the dynamics of fast degrees of freedom (such as electrons) is simplified by imposing that they relax to the minimum of their energy for each configuration of the slow degrees of freedom. The MaZe approach enforces this condition via an appropriate set of constraints applying a modified SHAKE algorithm [Bonella2020, Coretti2020,Girardier2021]. The fully Lagrangian formulation of the problem ensures time-reversible, symplectic evolution for the system while enabling use of a time step appropriate for the slow degrees of freedom. The second approach was recently exploited by the second proponent (Fabio Pietrucci) in the context of the MAESTRO collaboration between physicists, chemists and mathematicians, financed by the ISCD institute, with the developùent of likelihood-based Langevin parameter estimation algorithms for MD projected on a few degrees of freedom [Vroylandt2022]. In both cases new promising approaches have been developed and proposed to the scientific community, but intensive work is today needed to advance the methodologies until they can be applied to challenging problems.

These methodological advances will target an important application: the study of precipitation, i.e., the nucleation of an ordered or disordered solid phase from an initial dispersed solution. This problem encompasses a broad range of timely and challenging problems, from the synthesis of nanoparticles and advanced materials, to drug crystallization in pharmaceutical industry, to amyloid formation in neurodegenerative diseases [Baftizadeh2013], to bone formation [Debroise2022], etc. Such processes combine two big challenges: the simulation of a large number of solute molecules dispersed in a larger number of solvent ones, and the sampling of rare barrier-crossing events. The first problem will be significantly alleviated by new techniques - Langevin or MaZe based - developed within the present PhD project, while the second can benefit from innovative approaches (generalized coordinates and enhanced sampling techniques) recently introduced by the proponents in the context of homogeneous and heterogeneous nucleation [Pipolo17,Fitzner17]. By developing numerical schemes that avoid or simplify the explicit simulation of solvent dynamics, we will drastically reduce the computational cost while retaining a statistically accurate prediction of relevant structural, thermodynamic and kinetic observables.

Objectives and work program

We identified the following research tasks, progressing from method exploration and refinement on benchmark systems to developments targeting complex problems that can hardly be tackled with present means. In addition to its intrinsic interest, the parallel MaZe and Langevin developments also mitigate risks of failure to achieve our goals:

1. Using two-component Lennard-Jones fluids (heavy solute in light solvent) we will investigate under which hypothesis and up to what extent the MaZe framework is able to yield structural and dynamical observables in agreement with easily achievable brute-force MD. We will initially consider a stable disperse solution. We will address the relationship between solute-solute MaZe probability distributions and molecular dynamics (MD) potential of mean force, and the nature and correct description of thermal fluctuations of average properties. Next, we will move to supersatured conditions where the disperse phase is metastable, undergoing nucleation and precipitation on a timescale dependent on conditions.

2. In a second phase, we will perform a critical comparison between MaZe-based and Langevin-based models (the latter developed within the ISCD-MAESTRO initiative), both from the physical and from the computational viewpoints. Compared to Langevin, MaZe dynamics lacks noise and dissipation forces calibrated via the fluctuation-dissipation relation, and its statistical mechanical properties (like time-correlation functions and first passage time statistics) will be investigated and the method refined to include temperature effects. We will gradually move from the benchmark Lennard-Jones system to a realistic water solution of fullerene molecules simulated with classical force-fields. The latter system is amenable to a Markovian Langevin description that retains thermodynamic and kinetic properties of the all-atom MD simulation, as proved in the two-fullerene case [Palacio-Rodriguez2022].

3. In the last phase, we will tackle the open challenge of simulating the precipitation of carbonates and phosphates from water solutions. Such processes play an important role in environmental, biological and biomedical sciences, and so far standard MD simulations proved very difficult. We will assess the capability of MaZe and Langevin algorithms, combined with enhanced-sampling, to efficiently simulate the structural, thermodynamic and kinetic aspects of these processes. Lack of a sizable solute-solvent mass gap, a non-trivial role of water and possible precipitate polymorphism

increase the complexity. We will compare with a recent study in our group on the nucleation of small calcium-phosphate nanoparticles using metadynamics and umbrella sampling based on in-house topological reaction coordinates (article in preparation).

Interest of the collaboration, complementarity of the team

The scientific goals of the project are perfectly suited to the complementary expertise of the two proposing teams. The supervisors are both recognized experts in rare events methods, kinetic models (Fabio Pietrucci), and the MaZe scheme (Sara Bonella). The proposed work program entails developments in these areas that can only advance through close collaboration, synergy and cross-fertilization. Frequent exchanges will be facilitated by a Franco-Suisse travel grant “Germaine de Staël” obtained by the proponents. The PhD candidate will also benefit from regular interactions with young researchers working on related topics in the French and Swiss team. Both supervisors are members of broader collaborations that will provide further training and networking opportunities. Sara Bonella is the Deputy Director of the CECAM HQ at EPFL, where she is a member of the MARVEL consortium for simulation of materials and has recently initiated a collaboration with CECAM director, I. Pagonabarraga, on multiscale methods. At Sorbonne, Fabio Pietrucci coordinates, together with A. Marco Saitta, the multidisciplinary MAESTRO team, financed by the ISCD institute, aimed at solving challenges in the modeling of metastable systems for energy applications, including ten permanent researchers (physicists, chemists, mathematicians and computer scientists), that recently welcomed Sara Bonella as a new member. For all these reasons, ISCD is the ideal environment for the successful development of the proposed project. The **profile of the PhD candidate** includes a master degree in physics or mathematics, with advanced knowledge of analytical mechanics, numerical physics and statistical physics.

References

- H. Vroylandt, L. Goudenège, P. Monmarché, F. Pietrucci, B. Rotenberg, *Proc Natl Acad Sci USA* (in press) arXiv:2110.04246
- S. Pipolo, M. Salanne, G. Ferlat, S. Klotz, A.M. Saitta, F. Pietrucci, *Phys Rev Lett* 119, 245701 (2017)
- M. Fitzner, G.C. Sosso, F. Pietrucci, S. Pipolo, A. Michaelides, *Nat Commun* 8, 2257 (2017)
- F. Baftizadeh, F. Pietrucci, X. Biarnes, and A. Laio, *Phys Rev Lett* 110, 168103 (2013)
- K. Palacio-Rodriguez, F. Pietrucci (submitted) arXiv:2106.05415
- A. Coretti, S. Bonella, G. Ciccotti, *J Chem Phys Communications* 149, 191102 (2018)
- S. Bonella, A. Coretti, R. Vuilleumier, G. Ciccotti, *Phys Chem Chem Phys*, 22, 10775 (2020)
- A. Coretti, L. Scalfi, C. Bacon, B. Rotenberg, R. Vuilleumier, G. Ciccotti, M. Salanne, S. Bonella, *J Chem Phys* 152, 194701 (2020)
- D. Girardier, A. Coretti, G. Ciccotti, S. Bonella, *Eur J Phys B*, 94, 158 (2021)