Master's Internship Chemistry/Physics/Data Sciences • Feb. 2026

Département de Chimie, Laboratoire de Physique de l'ENS et Centre de Sciences des Données, École Normale Supérieure – PSL, Paris

Generative AI for reactive molecular configuration exploration

Context Machine-learned interatomic potentials (MLIPs) are transforming molecular simulations by providing highly accurate descriptions of molecular interactions, derived from quantum calculations, at a fraction of the computational cost (10.1016/j.cossms.2025.101214). While this technology has already revolutionized materials science and shows great promise for biomolecular studies, significant challenges remain in modeling chemical reactions. In particular, constructing chemically diverse and reliable datasets for a given reaction, or family of reactions, remains difficult.

Such datasets must include transition-state structures, which are inherently unstable and require precise identification of the reaction coordinate (RC). This challenge is further complicated when reactions involve solvent molecules. Identifying and efficiently sampling along the RC is therefore a central difficulty. Current studies often rely on molecular dynamics (MD) within active learning frameworks. In the ENS theoretical chemistry group, we have developed such a tool (10.1039/D4DD00209A), which has already led to significant advances (10.1038/s41557-024-01593-y, 10.1073/pnas.2322040121). However, these approaches are limited by slow decorrelation and possible instabilities when the MLIP is not fully converged.

Internship objective The internship will focus on optimizing data generation for the active learning of MLIPs, particularly for reactive events, by leveraging generative models such as Normalizing Flows (NFs) or score-based approaches, which can efficiently produce diverse molecular configurations. Unlike MD-based sampling, generative approaches promise much greater computational efficiency. However, training an accurate MLIP for systems with thousands of degrees of freedom remains a major challenge.

Proposed strategy To address this, we will build on recent proof-of-concept work from the ENS physics/data sciences group (10.1021/acs.jctc.4c00700, 10.1021/acs.jctc.4c00506), and follow hybrid strategy that combines: (i) the generation of a CV representation involving the solute and, possibly, a small shell of solvent atoms using an NF, (ii) the relaxation of the remaining solvent atoms using a pre-trained, solvent-specific MLIP. The solvent-specific MLIP will be trained beforehand on simulations of pure solvent. To handle the additional difficulty of learning the distribution of rare events across the transition region, the NF will be conditioned on the value of a low-dimensional RC. The internship will be organized in a bottom-up strategy, starting with multidimensional model potentials, then progressively moving toward realistic chemical systems.

Techniques/Methods The candidate will gain strong experience in generative models, molecular dynamics simulations and the construction of machine-learned interatomic potentials for simulations. Tools: Generative models; Molecular dynamics simulations; MLIP architectures; programming tools.

Candidate profile Master's student in physics, chemistry, computer sciences, applied maths, or a related field, with a strong interest for interdisciplinary research. Programming skills in Python are essential, and previous knowledge of deep learning is an advantage.

Research environment Research will take place with Guillaume Stirnemann and Damien Laage (Chemistry department) and Marylou Gabrié (LPENS and Center for Data Sciences). ENS is located in the very stimulating research environment of the Latin Quarter, at the heart of Paris. Our groups have extensive experience in, respectively, applying MLIP approaches for chemical reactivity studies and the development of generative model approaches to reliably sample physical and chemical phase space.

Extension into a PhD can be considered upon mutual agreement (funding opportunities are already considered).

Contact information Interested candidates should contact Guillaume Stirnemann (guillaume.stirnemann@ens.psl.eu), Damien Laage (damien.laage@ens.psl.eu) and Marylou Gabrié (marylou.gabrie@ens.psl.eu) **as soon as possible**, together with a curriculum vitae and contact information for one or two references.