SCIENCES SORBONNE UNIVERSITÉ

Ph.D. in
Nested Metasampling for Material Science

INSP INSTITUT DES NANOSCIENCES DE PARIS

INSP/ SU, Campus P. et M. Curie (Jussieu)
Director: Massimiliano Marangolo

Keywords: Nuclear quantum effects, DFT, Bayesian methods, partition function, machine learning

Scientific description: We propose an interdisciplinary project involving different fields such as Bayesian data analysis, statistical physics and applied mathematics for the efficient exploration and minimization of complex scalar functions. The project is centered on the Nested Sampling optimization algorithm based on Bayesian statistics to explore the relevant parameter space recursively effectively, obtain the topology thereof, and assess the quality and reliability of the obtained optimum within a pre-determined statistical error. The method, which was proposed 20 years ago, has been mainly applied to data analysis; recently, we contributed to assess its application to quantum statistical physics by simplifying the calculation of the quantum partition function Q. The evaluation of Q is crucial, as all thermodynamic properties can be derived from it. For a single system configuration, distinct energy models are available, each corresponding to a given accuracy and computational complexity. The bottleneck is the speedy evaluation of the energies for as large as possible samples of complex systems. To date, the most used theory in material science is the Density Functional Theory (DFT), which allows at most the simulation of a few hundreds of atoms for a few picoseconds, at a viable computational cost. Moreover, the DFT energy function has no analytic expression. An alternative are machine-learning generated energy models. These can provide excellent results at a much lower cost (up to 1000 smaller than the DFT) but are strongly dependent on the training process as, while they interpolate well, extrapolations will generate physically unacceptable errors. Finally, analytical interaction models, based on physical principles and fitted to theoretical and experimental data, can be much more efficient, but are usually limited to a narrow range of conditions. Finding the correct analytical expression and optimal set of parameters is a rather lengthy process requiring both expertise and intuition. This results in a maze of different expressions adapted to specific situations. In this respect, nested sampling can tell us to what extent a given model is reliable and whether adding another parameter to it makes it better or not. This will provide a highly valuable mathematical guideline, based on the quantitative assignement of probability to the different models and/or hypotheses, in this quest that will be complemented by additional mathematical methods for the reduction of pertinent parameters. Moreover, because the parameters result from highly non-linear fitting, nested sampling (or, better, metasampling as we work in the parameter space) can describe the topology of the whole parameter space, including secondary minima.

Techniques/methods in use: The candidate He/She will benefit from a very open scientific and interdisciplinary environment: the project brings together physicists with distinct expertise (Fabio Finocchi for the theory of solid state and Martino Trassinelli for Bayesian methods for atomic spectroscopy at INSP), in close collaboration with Julien Salomon at INRIA for his mathematical skills in model selection algorithms. We also expect to benefit from the SCAI environment, exchanging with researchers in applied mathematics and statistical learning.

Applicant skills: Quantum and statistical physics, applied mathematics, good programming skills.

Ph.D. supervisor(s): Martino Trassinelli, trassinelli@insp.jussieu.fr, tel.: 01 44 27 62 30
Fabio Finocchi, fabio@insp.jussieu.fr, tel.: 01 44 27 51 16

Ph.D. location: INSP, Sorbonne Université, 4 place Jussieu, 75005 Paris

Possibility for a Doctoral thesis: Yes (financed by the Sorbonne Center of Artificial Intelligence)