Engineering Tripos Part IIB, 4G5: Materials and Molecules: Modelling, Simulation and Machine Learning, 2023-24

Module Leader

Prof Gabor Csanyi [1]

Lecturer

Prof G Csanyi

Timing and Structure

Michaelmas term. ~12 lectures + Coursework. Assessment: 100% coursework.

Aims

The aims of the course are to:

- Introduce the concepts of computer simulation of material and molecular properties on the atomic scale;
- Teach the basic techniques of molecular dynamics and data analysis
- Provide hands-on experience with some widely used software packages (ASE, Ovito, etc)

Objectives

As specific objectives, by the end of the course students should be able to:

- Understand the principles of how microscopic simulation can be used to calculate material properties;
- Know what the fundamental capabilities and limitations of molecular simulation are;
- Carry out simple molecular simulation using a software package and measure observables, analyse results.

Content

In the last few decades computer simulations have emerged as a new scientific methodology – sandwiched between mathematical theories and experiment – with applications across the sciences and engineering. Because the parameters can be carefully controlled, these "theoretical experiment" provide powerful ways to develop fundamental understanding of the connection between microscopic models of the interactions between atoms and molecules and observable properties of many-particle systems.

The course starts with a swift walk-through of fundamental modelling concepts, ranging from quantum mechanics and statistical mechanics to the practicalities of numerical simulation, multiple length and time scales, and error control.

The second section is about specific models for materials and molecules which facilitate calculation of basic properties of matter, allowing both a deeper understanding of experimental observations and also first principles prediction of new phenomena.

The final section is on modern many-parameter models (aka "machine learning") and an introduction to how this allows breaking previously established limitations of numerical approaches, both for direct first principles dynamical simulations as well as using statistical "data mining" methods.

There are links with **4A9 (Molecular Thermodynamics)** and it would be interesting for some students to take both courses. 4G5 is more practical and much of it is about realistic models for specific systems, while 4A9 is more theoretical and statistical.

Specific topics are listed below. Each bullet is slightly more than a lecture's worth of material.

- Introduction?
  Overview of the course: (i) survey of fundamental modelling questions ; (ii) examples of the kinds of problems the course will address: phase diagrams, molecular structure and mechanical response, data mining for molecular properties; (iii) computational frameworks and tools, python packages, computational resources. ?
- Bottom up vs top down modelling? ?
  First principles simulation, prediction vs understanding, limitations (both conceptual and practical). Hierarchy of approximation, starting with the Quantum Mechanical models such as the Schrödinger Equation. Links to statistical mechanics, thermodynamical concepts at the roots of simulation techniques. ?
- Practical techniques
  Numerical simulation of ensembles: temperature, pressure, entropy, trajectories, correlation times, molecular dynamics and Monte Carlo techniques. Error estimation.?
- Empirical force fields and interatomic potentials? ?
  Simple organic bonding force fields for molecules, and Embedded Atom Models (EAM) for metals, mathematical relations between them and possible directions for increasing complexity and power of description.
- Free energy as a fundamental target of molecular simulation, links to experimental observables, both in terms of static and dynamic properties, statistical distributions, single molecule experiments.
- Machine learning for molecules: fundamentals?
  Molecular descriptors, uniqueness, symmetry, information compression. 3D structural descriptions, graph models, string representations.
- Review of regression tools: linear models, kernel regression, Gaussian processes, nonlinear regression (artificial neural networks)? ?

- *Computer Project intro: fundamentals of atomistic simulation*
- *Computer Project I: the mechanics of rubber, very large deformability*
- *Computer Project II: predicting organic crystal structures, Aspirin*
- *Computer Project III: machine learning for molecular properties, solubility of drugs*

Coursework

Assessment is by 100% Coursework, which consists of reports on Computer Project 1 and **one** out of the remaining two computer projects (**excluding** the intro project : fundamentals). Reports are due 1 Dec 2021.

Coursework	Format	Due date
		& marks
[Coursework activity #1 Report / Final] Computer project I	Individual Report anonymously marked	1 December 2 [30/60]

Coursework	Format	Due date & marks
[Coursework activity #2 Report / Final] One of (i) Computer project II, or (ii) Computer project III	Individual Report anonymously marked	1 December 2 [30/60]

Booklists

Please refer to the Booklist for Part IIB Courses for references to this module, this can be found on the associated Moodle course.

Also:

- Understanding Molecular Simulation, From Algorithms to Applications, D. Frenkel and B. Smit, (Academic Press).
- Computer Simulation of Liquids, M. P. Allen and D. J. Tildesley (Clarendon Press).
- Introduction to Modern Statistical Mechanics, D. Chandler (Oxford University Press).
- Molecular Modelling, Principles and Applications, A. R. Leach (Longman).

Examination Guidelines

Please refer to <u>Form & conduct of the examinations</u> [2].

Last modified: 30/05/2023 15:32

Source URL (modified on 30-05-23): https://teaching24-25.eng.cam.ac.uk/content/engineering-tripos-part-iib-4g5-materials-and-molecules-modelling-simulation-and-machine-3

Links
[1] mailto:gc121@cam.ac.uk
[2] https://teaching24-25.eng.cam.ac.uk/content/form-conduct-examinations