Course Information Sheet for entry in 2023-24: DPhil in Computational Discovery

About the course

The DPhil in Computational Discovery is a multidisciplinary programme spanning projects in Advanced Molecular Simulations, Machine Learning and Quantum Computing to develop new tools and methodologies for life sciences discovery.

This innovative course has been developed in close partnership between Oxford University and IBM Research. Each research project has been co-developed by Oxford academics working with IBM scientists. Students will have a named IBM supervisor/s and many opportunities for collaboration with IBM throughout the studentship.

The scientific focus of the programme is at the interface between Physical and Life Sciences. By bringing together advances in data and computing science with large complex sets of experimental data, more realistic and predictive computational models can be developed. These new tools and methodologies for computational discovery can drive advances in our understanding of fundamental cellular biology and drug discovery. Projects will span the emerging fields of Advanced Molecular Simulations, Machine Learning, and Quantum Computing while addressing both fundamental questions in each of these fields as well as at their interfaces.

Students will benefit from the interdisciplinary nature of the course cohort as well as the close interactions with IBM Scientists. After a very short induction period of one or two weeks, during which some basic training is provided, you will start a research project in your academic supervisor's laboratory.

Most laboratories have weekly meetings where members present and discuss their research results with other members of the laboratory. You will also regularly present your work in progress seminars, which are attended by other research groups working in related areas.

Whilst working on your research project you will participate in a comprehensive, flexible skills training programme which includes a range of workshops and seminars in transferable skills, generic research skills and specific research techniques. There are also numerous seminars and lectures by local and visiting scientists and you are provided with many opportunities to meet leading scientists.

Projects available for entry in 2023-24

You may identify up to three projects to be considered for in your application.

Project 1:

Robustness and Generalisation of Graph Machine Learning Models

Supervisors

Xiaowen Dong and Mihai Cucuringu

Summary

Can we predict whether a molecule can be a potent drug against certain bacteria from its chemical structure? Can we detect whether a piece of news on a social media site corresponds to misinformation from its spreading pattern? These are just some of the big questions facing today's society that can potentially be solved by an emerging area of research: graph machine learning (GML). A key open challenge in GML is the lack of understanding of model robustness and generalisation capability with respect to a perturbation of the graph data, which may come from natural noisy data characteristics (eg spurious correlations or irrelevant information) or malicious attempts (eg adversarial attacks).

In this project, building on top of recent preliminary results from Oxford and IBM, we propose to address this challenge through the lens of graph topology and spectral analysis, and study the implication of perturbations that action in both the graph spatial domain (eg add or delete graph edges) and the graph spectral domain (eg via modification of eigenvalues/eigenvectors of the graph Laplacian) for robustness and generalisation capability.

The expected outputs of this project include:

1. developing theoretical foundations for characterizing robustness and generalization of GML;
2. devising novel efficient algorithms for improving GML;
3. identifying suitable real-world application domains; *and*
4. contributing to joint publications at top AI/ML conferences and patents.

Project 2:

Accelerated Design of New Sustainable Battery Materials with Artificial Intelligence Methods

Supervisors

Saiful Islam

Summary

The provision of sustainable low-carbon energy is among the most urgent challenges of our time, and poses fundamental, exciting scientific questions. Materials performance lies at the heart of the development of green energy technologies, and computational methods now play a vital role in modelling the properties of energy materials. However, a full understanding of the atomistic processes within materials and across interfaces that control the performance of energy storage devices such as lithium-ion batteries remains incomplete. Emerging artificial intelligence (AI) and machine learning techniques are powerful tools offering innovative capabilities for studying new battery materials on length scales from individual atoms to tens of nanometres, promising quantum-mechanical accuracy and predictive power, whilst being many orders of magnitude faster than conventional methods.

The vision of this project is the innovative use of cutting-edge machine learning simulation techniques to probe the atomic-level operation of battery materials, thereby enabling a previously missing microscopic understanding and an accelerated design of new sustainable materials with enhanced performance. We will address electric vehicle application objectives of increasing the energy density and charge rates of battery electrodes and solid electrolytes, with a particular focus on how their macroscopic properties can be connected to the microscopic structure.

The project will involve the creation of accurate fitting databases and machine-learning-based interatomic potentials to model the underlying atomistic behaviour of novel battery electrodes and solid electrolytes. No equivalent concerted AI-modelling project on battery materials that inter-links such different expertise is being undertaken within any current IBM-Oxford Studentship project.

Project 3:

Quantum Relativistic Simulations for Dense Plasma Systems

Supervisor

Gianluca Gregori

Summary

The equations governing quantum mechanics have been known for nearly 100 years, but even today the task of reconciling them with the equations of classical motions remains unsolved. The problem is usually expressed in terms of the trajectory (or path) that a particle follows. While in ordinary Newtonian theory a particle moves along a well-defined path, this is not true anymore in orthodox (or the Copenhagen interpretation of) quantum mechanics. This is because it is impossible to simultaneously define the position and momentum of the particle. There is, however, an alternative possibility, called Bohmian mechanics or the de Broglie-Bohm interpretation, whereby the particle follows a definite path.

The Bohmian mechanics approach also offers another advantage. Since, as pointed out by de Broglie, the inclusion of quantum effects are entirely equivalent to the change of the spacetime metric, all relativistic effects in the many-body particle interactions can be included by rewriting and solving the equations of motion in this modified metric. This framework will be important for studying high-temperature dense plasmas as those found in the interior of stars but also in inertial fusion experiments. Moreover, as the proposed approach treats both electrons and ions on an equal footing,

the plasma response to perturbations inclusive of both electron and ion dynamics can be fully calculated.

The student's primary task will consist in developing this new computational approach within a Molecular Dynamics framework. The student will initially focus on the numerical implementation and then apply that to realistic systems – ie for the calculation of transport coefficients in inertial fusion experiments. We expect this project to produce important results that are of interest not only to the fusion community, but the broader community of researchers working in high energy density physics, planetary science and extreme materials.

Project 4:

Effective Transport Coefficients in Extreme Dynamic Material

Supervisor

Gianluca Gregori

Summary

Characterizing and quantifying mass, momentum, and energy transport in materials under extreme conditions is vital in many areas of research, ranging from inertial confinement fusion to the behaviour of matter in the interiors of giant planets and stars. With temperatures of a few electron volts (eV) and densities comparable to solids, warm dense matter (WDM) forms a key constituent of planetary interiors as well as cooler stellar objects such as brown dwarfs and the crust of neutron stars. Transport properties are difficult to model in WDM. Our goal here is to develop an experimental and numerical framework that can be used to measure effective transport in WDM and then use the experimental data to construct a suitable representation via symbolic regression or a trained neural network.

Our proposed work utilises recent advances in diagnostics and in machine learning. For the experiments, we intend to use X-ray photon correlation spectroscopy (XPCS) in novel ways to extract effective transport coefficients in dynamic laser-driven materials. Here we want to propose a novel machine learning approach to address the complex micro-physics of material strength properties and to identify their emergent behaviour via closed mathematical expressions. This is done by using a Graph Neural Network (GNN) to represent the discrete description of the underlying continuum system and then applying deep learning techniques to obtain a representation of the material properties as a function of the state variables (density, temperature, etc).

Our long-term goal is the development of augmented methods to ultimately improve the design and verification integrated modelling of WDM systems, in the sense that fluid simulations using these effective transport coefficients may now be able to capture the relevant physical processes at all scales. We expect this project to produce important results that are of interest not only to the fusion community, but the broader community of researchers working in high energy density physics, planetary science and extreme materials.

Project 5:

ML-Driven Fragment-Based Drug Development Using Data From High-Throughput X-Ray Crystallography and Biophysical Measurements

Supervisors

Charlotte Deane and Frank Von Delft

Summary

An approach will be developed that integrates machine learning with experimental measurements for the rapid design of bioactive compounds that are suitable as chemical probes. Chemical probes are often used as the starting point for drug discovery campaigns and can help elucidate the mechanism of molecule-target interactions. Machine learning techniques will be developed that use structural data as input to suggest potent, synthetically tractable molecules to make within this workflow. Techniques that better deconvolve signal from noise in biophysical assays will also be investigated.

This project aims to generate an algorithmic formalism that achieves rapid design of bioactive compounds suitable as chemical probes. We will develop a machine learning approach that iteratively

integrates experimental data from low-cost robotic organic synthesis, high-throughput crystallography (XChem), and rapid sensor-based biophysical measurements (Grating-Coupled Interferometry). The engine will be able to suggest new molecules that are potent, synthetically tractable and have good pharmacological properties.

This approach builds on methodological discoveries made in the successful COVID Moonshot initiative, an open science consortium that Professor von Delft co-founded, which delivered preclinical candidates against SARS-CoV-2 main protease from fragment hits in 18 months with <£1m.

This project will address the two interrelated questions that must be answered for these proof-of-concept successes to become an effective platform for probe discovery. First, how can machine learning take structural biology data as input to suggest new molecules to make? Secondly, how can we deconvolve signal amid noise in biophysical assays when the input is a crude reaction mixture?

Project 6:

Taking the Structure of Proteins Into Account: Predicting if Infections are Resistant to β-lactam Antibiotics Using Graph-Based Convolutional Neural Networks

Supervisor

Phil Fowler

Summary

This project will use machine learning to predict whether novel protein mutations confer resistance to specific antibiotics. Antimicrobial Resistance is a great concern to modern medicine. The Fowler group has previously used simple machine learning and physical simulation to address this problem. We aim to use graph neural network techniques developed by the IBM team to create advanced deep learning models and then train, validate and test the models on the large datasets we possess.

Antimicrobial Resistance (AMR) is a growing threat to modern medicine; not only would infectious disease claim many more lives than it already does but it would adversely impact many treatments that rely on prophylaxis, eg many anti-cancer therapies.

Due to the much higher degree of genetic variation and to maximise the information derived from the clinical samples, the student will work closely with IBM Research to create and then train graph-based convolutional neural networks that capture the topology of the protein and the protein in complex with ligands in the connectivity of the hidden layers. Crucially, such models can be trained from gene sequences rather than mutations and so can innately deal with the greater genetic variation. These techniques incorporate structure-based approaches within a deep learning framework and have been successfully validated and applied to the binding of proteins to small molecules and biologics.

This work will also utilize open source software. New developments are intended to be contributed to open source efforts.

The project would suit a student wishing to learn and apply machine learning to an important biomedical problem who wants to work in a fast-paced and highly interdisciplinary environment. It could therefore suit students from a broad range of backgrounds.

Project 7:

Defining Computation and Connectivity in Neuronal Population Activity Underlying Motor Learning

Supervisor

Andrew Sharott

Summary

Neural network structure constrains the activity dynamics of the brain. Specifically, learning of movements guided by the outcome of previous actions leads to adaptations in the motor cortical network and its activity. To understand these mechanisms on the cellular level would require simultaneous recordings from hundreds of local neurons at millisecond timescale in vivo during learning of a skilled movement. We have successfully established an approach to simultaneously record thousands of neurons across motor regions in mice, using recently developed high-density

electrode silicon-probes in combination with machine-learning based kinematic analysis and cell-type specific optogenetic modulation.

Motivated by recent work that links structure of population activity to the underlying synaptic connectivity (Dahmen et al., 2022) and our experience in cortical microcircuits (Peng et al., 2019, 2022), we aim to identify core changes in neuronal microcircuits that underlie motor learning and execution. We will develop novel approaches to extract activity signatures reflecting plastic changes on the local synaptic level and model how these constrain the overall dimensionality of neuronal population activity. The results will provide a microcircuit level understanding of learning in motor circuits and lay the groundwork to study neural network architecture in high-density electrophysiological recordings.

Project 8:

Optimising Therapy for Brain Disorders Through AI-Refined Deep Brain Stimulation

Supervisor

Hayrie Cagnan

Summary

Brain stimulation is extensively used to modulate neural activity in order to alter behaviour. In recent years, closed-loop stimulation techniques have gained increasing traction to sense a biomarker such as elevated neural activity patterns, and deliver stimulation in time with such events. Closed-loop stimulation techniques are used both for establishing a causal link between behaviour and neural activity, and also to treat various neurological and psychiatric conditions.

Building on our recent work (West et al 2022, Cagnan et al 2017), this PhD project aims to formalise stimulation parametrisation by using theoretical models of brain circuits in combination with state of the art machine learning approaches. Specifically, we will train artificial neural networks to classify discrete brain states of interest and optimise stimulation parameters to achieve precise manipulation of activity propagating across brain circuits. The successful development of such an approach would provide a powerful framework to guide next generation stimulation strategies both for usage in basic science and clinical applications.

Project 11:

Foundations of Stochastic Gradient Descent (and Generalization)

Supervisor

Patrick Rebeschini

Summary

Stochastic gradient descent is one of the most widely used algorithmic paradigms in modern machine learning. Despite its popularity, there are many open questions related to its generalization capabilities. For instance, while there is preliminary evidence that early-stopped gradient descent applied to over-parameterized models is robust with respect to label mispecifications, a complete theory that can account for this phenomenon is currently lacking.

The goal of this project is to rigorously investigate the robustness properties of early-stopped gradient descent from a theoretical point of view in simplified settings involving linear models, and to establish novel connections of such a methodology with the field of distributionally-robust optimization. The project will combine tools from the study of random structures in high-dimensional probability (eg concentration inequalities, theory of optimal transport) with the general framework of gradient and mirror descent methods from optimization and online learning (eg regularization).

Project 12:

Modelling Proton Delocalization in Hydrogen-Bond Networks with Quantum Simulators

Supervisor

Tristan Farrow

Summary

A new way of studying biochemical structures and extending computational models is through quantum information and quantum simulators/computers. An example of a widely relevant use-case for quantum simulators in biochemistry and pharmacology involves the study of tautomerization – when protons transfer between molecular sites resulting in different configurations of a chemical compound. Tautomerization plays an important role in canonical biochemical reactions by providing a pathway that can enable catalytic enzyme reactions, determine the tautomeric forms of photoproteins in luminescence or the aromatization of molecules. The molecular structures and their dynamics can be modelled as open quantum systems for improved accuracy. Traditional quantum chemistry assumes that protons are fixed and presently lacks the tools to model proton delocalization and decoherence processes to describe the dynamics of these open systems. Quantum simulators have the potential to model systems more accurately and to offer new ways of thinking and understanding biochemical processes.

Project 13:

Accelerated Modelling of Reaction Pathways using Machine Learning for Carbon Capture Materials

Supervisor

Fernanda Duarte

Summary

The climate crisis, due to anthropogenic emissions, is likely the single largest issue facing the planet in the 21st Century. Materials to aid in curbing, and ultimately removing, carbon emissions are vital to prevent unsustainable climate change. To meet this challenge, novel materials and new capabilities are required. Accurate and reliable predictions of material capabilities are a critical part in material design. Computational methods play an increasing role in providing these capabilities prior to laboratory work. This proposal seeks to address the issue of accurately, reliably, and efficiently computing reaction pathways and using such methods to improve carbon capture materials.

There are numerous carbon capture materials proposed, but only a handful are commercially operating to date. Additionally, for these materials to be used in a cyclic manner, the carbon dioxide must be removed, and the material reformed in an economical and sustainable manner.

The project will develop accurate and efficient Machine Learning (ML) models to elucidate reaction pathways and evaluate novel designs. Where data exist in the open literature, we will use this, but we will also seek to generate our own data specific to the domain at hand, which currently lacks such data sets. The overall aim of this project will be to develop ML models applicable to describing chemical reaction pathways. We will show these models operating on well-known and understood systems prior to using them to improve the current generation of carbon capture materials.

Project 14:

Using Carbon Dioxide to Make Plastics and Materials – Circular Carbon Economies

Supervisor

Charlotte Williams

Summary

The project combines a long-standing expertise in both IBM and Oxford Chemistry, Williams Research group, into ensuring the next generation polymers and plastics maximise carbon dioxide recycling and minimize negative environmental impacts. This will be achieved through investigation of impacts throughout the lifecycle – from ensuring the monomers used are bio-derived or even waste carbon dioxide, to delivering efficient manufacturing processes, to designing polymers with the highest performances to minimize need for additives to ensuring all materials are recyclable after use.

Research will exploit discoveries in the Williams group allowing carbon dioxide and bio-derived monomers to be transformed into polymers, plastics and elastics. The project focusses on efficient catalysis to make carbon dioxide derived thermoplastics for use in future electronics and electrical

applications, including as an insulator and in heat management systems, which show lower greenhouse gas emissions throughout their life cycles.

The project will involve a period of secondment and close collaboration with the IBM Almaden (San Francisco) team headed up by Jim Hedrick. The research at IBM focusses on using continuous flow methods in polymer synthesis to accelerate manufacturing and improve control over polymer properties. Polymer property and processing assessments will be made between the laboratories at the University of Oxford and IBM Almaden.

Project 15:

Developing Geospatial Foundation Models for Climate Model Evaluation and the Detection of Extreme Climate Events

Supervisor

Philip Stier

Summary

Foundation models are a general class of AI models trained (generally self-supervised) on a large set of multimodal data. The resulting foundation model can be fine-tuned to solve a wide array of downstream tasks. Despite the methodology is general and applicable to different domains and applications, current popular examples are mostly focused on natural language processing (eg GPT-3 for natural language and Dall-E for text-to-image tasks).

As foundation models are complex and trained on large datasets, they tend to exhibit an emergent property where a system's behaviour is implicitly induced rather than explicitly specified. This is especially advantageous for many applications in climate science where the underlying physical processes are sometimes too complex for a limited amount of data to capture, or high quality data for training models able to detect climate events of interest might be scarce. The aim of this PhD project is to develop new modular deep learning architectures for foundation models that allow one to deal with the multivariate nature of climate data and its spatio-temporal intermittence.

The project will explore transformer-based architectures to allow parallelization between modalities before the extracted data representation is recombined. Focusing on training efficiency and computation, the project may also investigate whether it is possible to understand the added value of bringing in an additional modality or sets of training samples. Ultimately, the foundation models developed during the project will be tested and compared to the regular paradigm (eg developing a bespoke model for each application), in downstream tasks. This might include earth observation for climate hazards (eg flood, wildfire, landslide, drought) and climate model evaluation against observations. If successful, these foundation models will be an extremely powerful tool that will enable more efficient and accurate climate impact assessment and earth observation.

Project 16:

Advancing Synthesis Prediction with Machine Learning - A Data Driven/Mechanistic Approach

Supervisors

Professor Fernanda Duarte and Dr Teodoro Laino

Summary

The project will apply the latest machine learning (ML) techniques to chemical applications, including the exploration of reaction pathways toward medicinally relevant scaffolds. The aim will be to develop interpretable ML algorithms that facilitate the prediction of synthetic routes and provide a mechanistic understanding of their outcome.

This project will enable the student to explore fundamental scientific questions at the interface of chemistry and machine learning and apply these insights to tackle timely real-world applications. It will also provide the opportunity to work with multi-disciplinary teams in academia and industry. The group of Professor Fernanda Duarte will provide world-leading expertise in reaction pathway modelling and automation, while the team at IBM Research will bring expertise in the development of computational chemistry software and AI techniques.

Applicants must have, or expect to obtain, a Master's (or equivalent) degree in Chemistry, Physics, Computer Science or a related subject. Previous experience in computational chemistry or machine learning would be an advantage. The successful candidate will be based at the University of Oxford and, as part of their project, will spend at least three months working with at IBM Research in Zurich, Switzerland.

Supervision

The allocation of graduate supervision is the responsibility of the Medical Sciences Graduate School and it is not always possible to accommodate the preferences of incoming graduate students to work with a particular member of staff. Under exceptional circumstances a supervisor may be found outside the department leading the course.

Your supervisor may appoint a senior member of the laboratory as your day-to-day supervisor. Further support is available from your college advisor.

The frequency of meetings with supervisors will depend on which department your DPhil is based. Commonly, within those departments based with the Medical Science Division, once a fortnight is typical.

Assessment

There are a number of key stages in the research programme. Within a month of starting, you will meet with your academic and IBM supervisors to finalise your project and agree on an initial programme of research. Within the first three months, you will complete an analysis of your training needs with your academic supervisor.

Students begin the DPhil in Computational Discovery programme as a probationary research student (PRS). Before the end of their fourth term students are required to apply for Transfer to DPhil Status.

A successful transfer of status from PRS to DPhil status will require the submission of a transfer report. Students who are successful at transfer will also be expected to apply for and gain confirmation of DPhil status to show that their work continues to be on track. This will need to done within nine terms of admission.

Both milestones normally involve an interview with two assessors (other than your supervisor) and therefore provide important experience for the final oral examination.

You will be expected to submit an original thesis of up to 50,000 words after three or, at most, four years from the date of admission. To be successfully awarded a DPhil you will need to defend your thesis orally (viva voce) in front of two appointed examiners.

Changes to courses

The University will seek to deliver this course in accordance with the description set out above. However, there may be situations in which it is desirable or necessary for the University to make changes in course provision, either before or after registration. These may include significant changes made necessary by a pandemic (including Covid-19), epidemic or local health emergency. For further information, please see the University's Terms and Conditions (http://www.graduate.ox.ac.uk/terms) and our page on changes to courses (http://www.graduate.ox.ac.uk/coursechanges).

Expected length of course

	Full Time Only
Expected length	3 to 4 years

Costs

Annual fees for entry in 2023-24

Fee status	Annual Course fees
Home	£8,960
Overseas	£29,700

Information about course fees

Course fees are payable each year, for the duration of your fee liability (your fee liability is the length of time for which you are required to pay course fees). For courses lasting longer than one year, please be aware that fees will usually increase annually. Information about how much fees and other costs may increase is set out in the University's Terms and Conditions (http://www.graduate.ox.ac.uk/terms).

Course fees cover your teaching as well as other academic services and facilities provided to support your studies. Unless specified in the additional cost information (below), course fees do not cover your accommodation, residential costs or other living costs. They also don't cover any additional costs and charges that are outlined in the additional cost information.

Graduate students who have reached the end of their standard period of fee liability may be required to pay a termly University and/or a college continuation charge.

The University continuation charge, per term for entry in 2023-24 is £572, please be aware that this will increase annually. For part-time students, the termly charge will be half of the termly rate payable by full-time students.

If a college continuation charge applies (not applicable for non-matriculated courses) it is likely to be in the region of £100 to £600 per term. Please contact your college for more details.

Additional cost information

There are no compulsory elements of this course that entail additional costs beyond fees (or, after fee liability ends, continuation charges) and living costs. However, please note that, depending on your choice of research topic and the research required to complete it, you may incur additional expenses, such as travel expenses, research expenses, and field trips. You will need to meet these additional costs, although you may be able to apply for small grants from your department and/or college to help you cover some of these expenses.

Living costs

In addition to your course fees, you will need to ensure that you have adequate funds to support your living costs for the duration of your course.

The likely living costs for 2023-24 are published below. These costs are based on a single, full-time graduate student, with no dependants, living in Oxford. We provide the cost per month so you can multiply up by the number of months you expect to live in Oxford.

Likely living costs for one month

	Lower range	Upper range
Food	£300	£470
Accommodation	£715	£860
Personal items	£180	£305
Social activities	£40	£90
Study costs	£35	£80
Other	£20	£35
Total	£1,290	£1,840

Likely living costs for nine months

	Lower range	Upper range
Food	£2,700	£4,230
Accommodation	£6,435	£7,740
Personal items	£1,620	£2,745
Social activities	£360	£810
Study costs	£315	£720
Other	£180	£315
Total	£11,610	£16,560

Likely living costs for twelve months

	Lower range	Upper range
Food	£3,600	£5,640
Accommodation	£8,580	£10,320
Personal items	£2,160	£3,660
Social activities	£480	£1,080
Study costs	£420	£960
Other	£240	£420
Total	£15,480	£22,080

When planning your finances for any future years of study at Oxford beyond 2023-24, it is suggested that you allow for potential increases in living expenses of 5% or more each year – although this rate may vary significantly depending on how the national economic situation develops.

More information about how these figures have been calculated is available at www.graduate.ox.ac.uk/livingcosts.

Document accessibility

If you require an accessible version of the document please contact Graduate Admissions and Recruitment by email (graduate.admissions@admin.ox.ac.uk) or via the online form (http://www.graduate.ox.ac.uk/ask).