Computational Physics and Master program in Computational Science at the University of Oslo

Department of Physics and Center for Computing in Science Education, UiO

Aug 15, 2023

Why should we focus on Computational Science and Data Science?

- By 2020, it is expected that one of every two jobs in the STEM fields will be in computing (Association for Computing Machinery, 2013). Note the year of 2013. Today we can safely say almost all jobs!
- Computation is an essential and cross-cutting element of all S(cience)T(echnology)E(ngineering)M(athema disciplines
- Computational science has developed into a discipline of its own right
- Computations and the understanding of large data sets will play an even larger role in basically all disciplines of STEM fields, Medicine, the Social Sciences, the Humanities and education
- Students at both undergraduate and graduate level are unprepared to use computational modeling, data science, and high performance computing – skills valued by a very broad range of employers.
- The 3rd Industrial Revolution will alter significantly the demands on the workforce. To adapt a highly-qualified workforce to coming challenges requires strong fundamental bases in STEM fields. Computational Science and Data Science can provide such a background at all stages.

Master program in Computational Science

The program is a collaboration between seven departments and classical disciplines:

- Institute of Theoretical Astrophysics
- Department of Biosciences
- Department of Chemistry
- Department of Geoscience
- Department of Informatics
- Department of Mathematics
- Department of Physics

The program is multidisciplinary and everbody who has completed undergraduate studies in science and engineering, with a sufficient quantitative background, is eligible. The language of instruction is Norwegian and/or English.

Vision for the future: Scientific Computing and Data Science

Scientific computing focuses on the development of predictive computer models of the world around us. As study of physical phenomena through experimentation has become impossible, impractical and/or expensive, computational modeling has become the primary tool for understanding—equal in stature to analysis and experiment. The discipline of scientific computing is the development of new methods that make challenging problems tractable on modern computing platforms, providing scientists and engineers with key windows into the world around us.

Data science focuses on the development of tools designed to find trends within datasets that help scientists who are challenged with massive amounts of data to assess key relations within those datasets. These key relations provide hooks that allow scientists to identify models which, in turn, facilitate making accurate predictions in complex systems. For example, a key data science goal on the biological side would be better care for patients (e.g., personalized medicine). Given a patient's genetic makeup, the proper data-driven model would identify the most effective treatment for that patient.

Aims of the program

A specific aim of this program is to develop your ability to pose and solve problems that combine insights from a specific discipline with mathematical tools and computational skills. This provides a unique combination of applied and theoretical knowledge and skills. These features are invaluable for the development of multi-disciplinary educational and research programs. The main focus is not to educate computer specialists, but to educate students with a solid understanding in basic science as well as an integrated knowledge on how to use essential methods from computational science. This requires an education that covers both the specific disciplines like physics, biology, geoscience, mathematics etc with a strong background in computational science.

A significant aspect of this program is the ability to offer new educational opportunities that are aligned with the needs of a 21st century workforce. Many companies are seeking individuals who have knowledge of both a specific discipline and computational modeling.

Overarching description of the CS program

In this program you learn to use the computer as a laboratory for solving problems in science and engineering. The program offers exciting thesis projects from many disciplines: biology and life science, chemistry, mathematics, informatics, physics, geophysics, mechanics, geology, computational finance, computational informatics, b ig data analysis, digital signal processing and image analysis – you select research field according to your interests.

A Master's degree from this program gives you a methodical training in planning, conducting, and reporting large research projects, often together with other students and university teachers. The projects emphasize finding practical solutions, developing an intuitive understanding of the science and the scientific methods needed to solve complicated problems, use of many tools, and not least developing own creativity and independent thinking. The thesis work is a scientific project where you learn to tackle a scientific problem in a professional manner. The program aims also at developing a deep understanding of the role of computing in solving modern scientific problems.

From this program you gain deep insights in the fundamnetal role computations play in our advancement of science and technology, as well as the role computations play in society.

The program opens up for flexible backgrounds

While discipline-based master's programs tend to introduce very strict requirements for specific courses, we believe in adapting a computational thesis topic to the student's background, thereby opening up for students with a wide range of bachelor's degrees. A very heterogeneous student community is thought to be a strength and unique feature of this program.

Thesis directions

- Computational Science: Applied Mathematics and Risk Analysis
- Computational Science: Astrophysics

- Computational Science: Bioinformatics
- Computational Science: Biology
- Computational Science: Chemistry
- Computational Science: Geoscience
- Computational Science: Imaging and Biomedical Computing
- Computational Science: Materials science
- Computational Science: Mechanics
- Computational Science: Physics

The thesis projects will be tailored to your needs, wishes and scientific background. The projects can easily incorporate topics from more than one discipline. Many of the master thesis projects have advisors from more than one department.

Career prospects

Candidates who are capable of modeling and understanding complicated systems in natural science, are in short supply in society. The computational methods and approaches to scientific problems students learn when working on their thesis projects are very similar to the methods they will use in later stages of their careers. To handle large numerical projects demands structured thinking and good analytical skills and a thorough understanding of the problems to be solved. This knowledge makes the students unique on the labor market.

Career opportunities are many, from research institutes, universities and university colleges and a multitude of companies. Examples include IBM, Hydro, Statoil, and Telenor. The program gives an excellent background for further studies, with a PhD as one possible goal.

The program has also a strong international element which allows students to gain important experience from international collaborations in science, with the opportunity to spend parts of the time spent on thesis work at research institutions abroad.

The Center for Computing in Science Education (CCSE)

The goal of CCSE is to integrate computing as a natural tool in basic educations, to make the education research near and to prepare students for an interdisciplinary workplace.

The center will function as our coordintaing unit for common scientific and social activities.

Location: fourth floor, eastern wing of the Physics building.

The Center for Computational and Data Science (dScience)

This newly establised center is taylored to innovative research on Computational Science and Data Science and coordinates research and educational activities in the these fields. Look up the above weblink for more information.

AI/ML and some statements you may have heard (and what do they mean?)

1. Fei-Fei Li on ImageNet: **map out the entire world of objects** (The data that transformed AI research)

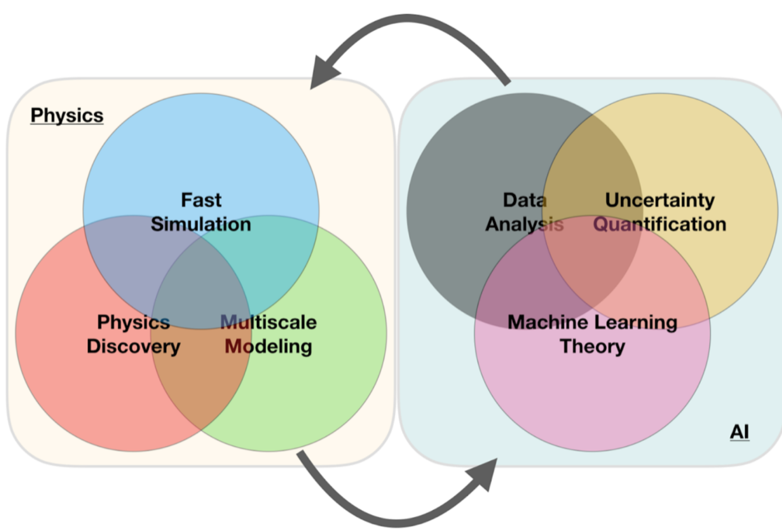
2. Russell and Norvig in their popular textbook: **relevant to any intellectual task; it is truly a universal field** (Artificial Intelligence, A modern approach)

3. Woody Bledsoe puts it more bluntly: **in the long run, AI is the only science** (quoted in Pamilla McCorduck, Machines who think)

If you wish to have a critical read on AI/ML from a societal point of view, see Kate Crawford's recent text Atlas of AI

Here: with AI/ML we intend a collection of machine learning methods with an emphasis on statistical learning and data analysis

A simple perspective on the interface between ML and Physics



Scientific Machine Learning

An important and emerging field is what has been dubbed as scientific ML, see the article by Deiana et al Applications and Techniques for Fast Machine Learning in Science, arXiv:2110.13041

The authors discuss applications and techniques for fast machine learning (ML) in science – the concept of integrating power ML methods into the real-time experimental data processing loop to accelerate scientific discovery. The report covers three main areas

1. applications for fast ML across a number of scientific domains;
2. techniques for training and implementing performant and resource-efficient ML algorithms;
3. and computing architectures, platforms, and technologies for deploying these algorithms.

Machine Learning software

Machine learning is an extremely rich field, in spite of its young age. The increases we have seen during the last three decades in computational capabilities have been followed by developments of methods and techniques for analyzing and handling large date sets, relying heavily on statistics, computer science and mathematics. The field is rather new and developing rapidly.

Popular software packages written in Python for ML are

- Scikit-learn,
- Tensorflow,
- PyTorch
- Keras,

and more. These are all freely available at their respective GitHub sites. They encompass communities of developers in the thousands or more. And the number of code developers and contributors keeps increasing.

Lots of room for creativity

Not all the algorithms and methods can be given a rigorous mathematical justification, opening up thereby for experimenting and trial and error and thereby exciting new developments.

A solid command of linear algebra, multivariate theory, probability theory, statistical data analysis, optimization algorithms, understanding errors and Monte Carlo methods is important in order to understand many of the various algorithms and methods.

Job market, a personal statement: A familiarity with ML is almost becoming a prerequisite for many of the most exciting employment opportunities. And add quantum computing and there you are!

Types of machine learning

The approaches to machine learning are many, but are often split into two main categories. In *supervised learning* we know the answer to a problem, and let the computer deduce the logic behind it. On the other hand, *unsupervised learning* is a method for finding patterns and relationship in data sets without any prior knowledge of the system. Some authours also operate with a third category, namely *reinforcement learning*. This is a paradigm of learning inspired by behavioural psychology, where learning is achieved by trial-and-error, solely from rewards and punishment.

Another way to categorize machine learning tasks is to consider the desired output of a system. Some of the most common tasks are:

- Classification: Outputs are divided into two or more classes. The goal is to produce a model that assigns inputs into one of these classes. An example is to identify digits based on pictures of hand-written ones. Classification is typically supervised learning.
- Regression: Finding a functional relationship between an input data set and a reference data set. The goal is to construct a function that maps input data to continuous output values.
- Clustering: Data are divided into groups with certain common traits, without knowing the different groups beforehand. It is thus a form of unsupervised learning.

Examples

The large amount of degrees of freedom pertain to both theory and experiment in the physical sciences. With increasingly complicated experiments that produce large amounts data, automated classification of events becomes increasingly important. Here, deep learning methods offer a plethora of interesting research avenues.

- Reconstruction of particle trajectories or classification of events are typical examples where ML methods are being used. However, since these data can often be extremely noisy, the precision necessary for discovery in physics requires algorithmic improvements. Research along such directions, interfacing nuclear and particle physics with AI/ML is expected to play a significant role in physics discoveries related to new facilities. The treatment of corrupted data in imaging and image processing is also a relevant topic.

- Design of detectors represents an important area of applications for ML/AI methods in nuclear physics.

And more

- An important application of AI/ML methods is to improve the estimation of bias or uncertainty due to the introduction of or lack of physical constraints in various theoretical models.
- In theory, we expect to use AI/ML algorithms and methods to improve our knowledge about correlations of physical model parameters in data for quantum many-body systems. Deep learning methods show great promise in circumventing the exploding dimensionalities encountered in quantum mechanical many-body studies.
- Merging a frequentist approach (the standard path in ML theory) with a Bayesian approach, has the potential to infer better probabilitity distributions and error estimates. As an example, methods for fast Monte-Carlo-based Bayesian computation of nuclear density functionals show great promise in providing a better understanding
- Machine Learning and Quantum Computing is a very interesting avenue to explore. See for example talk of Sofia Vallecorsa.

Selected references

- Mehta et al. and Physics Reports (2019).
- Machine Learning and the Physical Sciences by Carleo et al
- Goodfellow, Bengio and Courville, Deep Learning
- My favorite book, Brunton and Kutz
- Mathematics for Machine Learning Book by A. Aldo Faisal, Cheng Soon Ong, and Marc Peter Deisenroth

What are the basic ingredients?

Almost every problem in ML and data science starts with the same ingredients:

- The dataset $\mathbf{x}$ (could be some observable quantity of the system we are studying)
- A model which is a function of a set of parameters α that relates to the dataset, say a likelihood function $p(\mathbf{x}|\alpha)$ or just a simple model $f(\alpha)$
- A so-called **loss/cost/risk** function $\mathcal{C}(\mathbf{x}, f(\alpha))$ which allows us to decide how well our model represents the dataset.

We seek to minimize the function $\mathcal{C}(\mathbf{x}, f(\alpha))$ by finding the parameter values which minimize $\mathcal{C}$. This leads to various minimization algorithms. It may surprise many, but at the heart of all machine learning algortihms there is an optimization problem.

Quantum Science and Technologies

Quantum Computing requirements.

1. be scalable
2. have qubits that can be entangled
3. have reliable initializations protocols to a standard state
4. have a set of universal quantum gates to control the quantum evolution
5. have a coherence time much longer than the gate operation time
6. have a reliable read-out mechanism for measuring the qubit states
7. and many more

Candidate systems

1. Superconducting Josephon junctions
2. Single photons
3. Trapped ions and atoms
4. Nuclear Magnetic Resonance
5. Quantum dots, expt at MSU
6. **Point Defects in semiconductors, experiments at UiO, center for Materials Science**
7. more

Education

1. Build up a series of courses in QIS, inspiration QuSTEAM (Quantum Information Science, Technology, Engineering, Arts and Mathematics) initiative from USA
2. Bachelor program in Computational Physics and Quantum Technologies
   (a) study direction/option in **quantum technologies**
   (b) study direction/option in **Artificial Intelligence and Machine Learning**
3. Master of Science program in Computational and Data Science
   (a) UiO has already MSc programs in CS and DS
4. PhD program in CS and DS
   (a) with directions in **quantum technologies**
   (b) with directions in **Artificial Intelligence and Machine Learning**

Courses, Prototype

Topics in a Bachelor of Science/Master of Science.

1. Information Systems
2. From Classical Information theory to Quantum Information theory
3. Classical vs. Quantum Logic
4. Classical and Quantum Laboratory
5. Discipline-Based Quantum Mechanics
6. Quantum Software
7. Quantum Hardware
8. more

Important Issues to think of

1. Lots of conceptual learning: superposition, entanglement, QIS applications, etc.
2. Coding is indispensable. That is why this should be a part of a CS/DS program
3. Teamwork, project management, and communication are important and highly valued
4. Engagement with industry: guest lectures, virtual tours, co-ops, and/or internships.
5. Diversity needs to be a priority
6. Mentorship should begin the moment students enroll.

Observations

1. Students do not really know what QIS is.
2. There is conflation of “Quantum Information Science” with “Quantum computing”.
3. Students perceive that a graduate degree is necessary to work in QIS. A BSc will help.

Future Needs/Problems (US observations mostly but transfer most likely to Europe as well)

1. There are already (USA) great needs for specialized people (Ph. D. s, postdocs), but also needs of people with a broad overview of what is possible in QIS.
2. There are not enough potential employees in QIS (USA). It is a supply gap, not a skills gap.
3. A BSc with specialization is a good place to start
4. It is tremendously important to get everyone speaking the same language. Facility with the vernacular of quantum mechanics is a big plus.
5. There is a huge list of areas where technical expertise may be important. But employers are often more concerned with attributes like project management, working well in a team, interest in the field, and adaptability than in specific technical skills.