

# **Inverted CERN School of Computing 2026**

## **Report of Contributions**

Contribution ID: 2

Type: **not specified**

# Modern I/O and Storage Systems for Scientific Computing (Lecture 1)

*Thursday, 19 March 2026 10:00 (1 hour)*

## Abstract

I/O bottlenecks are often the hidden performance killers in HEP data analysis, yet many scientists lack a deep understanding of how storage systems actually work. This lecture demystifies the complete I/O stack (from application code to physical storage) and introduces both classical and modern techniques for optimal performance. We'll explore fundamental concepts (buffered I/O, direct I/O, mmap), diagnose real bottlenecks using Linux tools, and cover network I/O patterns critical for Grid jobs. The session culminates with modern asynchronous I/O APIs like `io_uring` that are revolutionizing high-performance computing. Participants will leave with practical skills to optimize their analysis workflows and a solid foundation for future learning.

## Topics

### Fundamentals and Diagnosis

- I/O stack and core concepts
- Diagnosing I/O performance problems
- Optimization strategies for local storage

### Network I/O, Frameworks, and Modern APIs

- Network I/O for Grid Computing
- Modern Asynchronous I/O APIs
- Quick reference for common HEP scenarios

### Hands-On Session (examples)

- Profile and diagnose I/O bottlenecks in analysis code
- Optimize ROOT analysis workflow
- Build simple async storage system with `io_uring`

## Number of lecture hours

2

## Number of exercise hours

1

## Attended school

tCSC 2024 Heterogenous Architectures (Belgrade)

## **Comments**

**Author:** DEL MONTE, Gianmaria (CERN)

**Presenter:** DEL MONTE, Gianmaria (CERN)

Contribution ID: 3

Type: **not specified**

## Authentication and Authorization for Global Research Infrastructure (Exercises)

*Thursday, 19 March 2026 16:40 (1 hour)*

In today's interconnected world, authentication and authorization have become critical components of modern research infrastructure. This lecture will begin by exploring traditional methods, such as certificate-based authentication, before progressing to how initiatives like **WLCG**, **IRIS** (UK computing resource federation) and **SKA SRCNet** (Square Kilometre Array - SKA Regional Centres Network) leverage industry-standard, token-based **Authentication and Authorization Infrastructure** (AAI) to enable large-scale research federations.

Next, we will examine how modernised certificate-based PKI (Public Key Infrastructure) can complement token-based systems, creating a secure and user-friendly approach for accessing research resources. This hybrid model ensures both flexibility and robustness in managing identity and access across global collaborators.

The practical exercise will give attendees hands-on experience deploying a token issuer and a certificate authority using Kubernetes Helm charts. Participants will:

1. Obtain a token from the issuer and analyse its content to understand the underlying architecture.
2. Issue a certificate and explore how token information translates into certificate issued via automated certificated issuer and how the 2 technologies can be bridged.
3. Use the issued certificate to access research resources in a simulated environment, demonstrating secure and streamlined access workflows.

By the end of the session, attendees will gain both conceptual understanding and practical skills for implementing modern AAI solutions in global research infrastructures.

### Number of lecture hours

1

### Number of exercise hours

1

### Attended school

CSC 2024 (Hamburg)

### Comments

Can travel to CERN for in-person attendance.

I also attended the tCSC 2025 Abingdon but I can't multi-select.

**Author:** CHUNG, Donald (STFC)

**Presenter:** CHUNG, Donald (STFC)

Contribution ID: 5

Type: **not specified**

# Git in Practice: Techniques for Collaborative Development (Lecture)

*Thursday, 19 March 2026 11:30 (1 hour)*

## Git in Practice: Techniques for Collaborative Development

This lecture will focus on **Git good practices and advanced commands** to help developers work more efficiently and collaboratively in shared repositories. The session will go beyond the simple *pull-commit-push* workflow and explore practical techniques for maintaining clean, reliable version histories in multi-developer environments.

### The lecture will cover:

- **Collaborating with Git:** How Git supports teamwork and version control in modern development, with examples of effective branching strategies and workflows.
- **Branching and merging strategies:** Understanding feature branches, Git workflows, and trunk-based development to manage change efficiently.
- **Advanced Git commands:** Practical use of git rebase, git merge, git reset, git bisect, and git cherry-pick. Discussing when and how to use them safely to simplify history and resolve issues.
- **Conflict resolution:** Techniques for diagnosing and resolving merge conflicts cleanly and confidently.
- **Git hooks and CI/CD integration:** Leveraging Git hooks to automate quality checks, enforce standards, and integrate testing into CI/CD pipelines.
- **Merge requests and code reviews:** Best practices for creating, reviewing, and approving merge requests, and giving constructive feedback to maintain high code quality and consistency.

### The accompanying exercise session would involve:

While this lecture does not naturally lend itself to a traditional exercise session, it could include a **collaborative GitHub repository activity** where participants apply these techniques in a shared environment to simulate a larger development team. Alternatively, a less collaborative exercise could involve **improving the commit history** of a sample repository, using techniques such as rebasing or squashing to create a cleaner, more linear history.

## Number of lecture hours

1

## Number of exercise hours

1

## Attended school

CSC 2025 (Lund)

## Comments

I would love to be able to travel to CERN from my home organisation to attend the school.

**Author:** SALMON, Lizzie

**Presenter:** SALMON, Lizzie

Contribution ID: 6

Type: **not specified**

# How performance analysis of real software works: the case of MadGraph (Lecture 1)

*Monday, 16 March 2026 09:15 (1 hour)*

The growing computational demands of high-precision simulations for the High-Luminosity LHC (HL-LHC) era require a careful optimization of Monte Carlo event generators.

In these lectures, we will explore different profiling techniques and their application to MadGraph5\_aMC@NLO, one of the most widely used software in high-energy physics to simulate particle collisions:

- We will begin with an overview of state-of-the-art profiling methodologies and tools for performance analysis on modern architectures.
- These techniques will then be applied to MadGraph, with a particular focus on Next-to-Leading Order (NLO) computations, which represent a major computational challenge.
- By studying the performance characteristics of the various components of the code, we will try to identify areas for optimization and discuss strategies to exploit hardware acceleration, including data-parallel and heterogeneous computing paradigms.

---

The goal of the lectures is to provide both conceptual and practical insights into how systematic profiling can guide software optimization, enabling efficient use of next-generation computing resources.

## Number of lecture hours

2

## Number of exercise hours

1

## Attended school

tCSC 2025 Heterogenous Architectures (Split)

## Comments

**Author:** MASSARO, Daniele (CERN)**Presenter:** MASSARO, Daniele (CERN)

Contribution ID: 7

Type: **not specified**

## Parallel computing with RISC-V vectors (lecture)

*Monday, 16 March 2026 15:00 (1 hour)*

For faster computing it is necessary to utilize data parallelism. The tools available for programmers for parallel computing depend on hardware and CPUs architecture. The RISC-V architecture provides a vector processor style Single Instruction, Multiple Data (SIMD) instruction.

The proposed 1 hour lecture will include:

- Concept of parallel data processing. Classification of various approaches.
- Overview of an array processor and vector processor approach. How these theoretical approaches are used for vectorization in CPUs and GPUs.
- Description of how SIMD is implemented in x86 architecture; its features, advantages and problems.
- Description of how RISC-V "V" extension is implementing scalable vectorization; its features, problems and advantages over x86 architecture approach.
- Examples of vectorizations done by compiler and necessary manual vectorizations.

### Number of lecture hours

1

### Number of exercise hours

0 (no exercises)

### Attended school

tCSC 2024 Heterogenous Architectures (Belgrade)

### Comments

**Author:** GALIAMOV, Airat (Promwad d.o.o.)

**Presenter:** GALIAMOV, Airat (Promwad d.o.o.)

Contribution ID: 8

Type: **not specified**

## Graphical Record Linkage with Big Data

One of the core databases in the European Central Bank is the RIAD (Register of Institutions and Affiliates Data) register, a massive repository of 16 million counterparties, covering companies, institutions, funds and much more.

One of our main task is to make sure that every entity refers to a different counterparty and no duplicates are generated. While not every record is a duplicate, every new entry must be checked against a huge dataset to ensure data integrity. Legacy string-matching algorithms and pre-defined checks based on the reported attributes served us well so far, but are becoming somehow obsolete in the face of modern scale and semantic complexity.

Key challenges in this domain include semantic ambiguity and logical consistency. For instance, how do you teach a system that “Ferrari S.p.A” and “Ferrari F.Lli Lunelli S.P.A (In Breve Ferrari S.P.A.)” may sound similar, but one is a famous car producer while the other one produces wines and they are very distinct companies? And how do you then make it realize that “HP” and “Hewlett-Packard” refer to the same company despite looking different? Furthermore, how do you prevent “chain reactions” of bad matches, where a single error bridges two distinct corporate groups into one large incorrect cluster? A modern approach to address these is Deep Entity Matching (DEM) combined with Graph Theory.

In this lecture, we will explore how to move from simple text comparison to a “Two-Pass” Neural Architecture. We will cover the use of efficient blocking strategies to narrow the search space, Cross-Encoders for high-precision semantic checks, and the application of Graph Topology to detect and cut illogical links in the data. The aim of this session is to demonstrate how High-Performance Computing can be used to solve critical data infrastructure problems in a wide variety of contexts.

### Comments

### Number of lecture hours

1

### Number of exercise hours

0 (no exercises)

### Attended school

tCSC 2025 Security (Abingdon)

**Author:** Mr GARCÍA FERNÁNDEZ, Isidro Javier (European Central Bank)

**Presenter:** Mr GARCÍA FERNÁNDEZ, Isidro Javier (European Central Bank)

Contribution ID: 9

Type: **not specified**

# Neural Scaling Laws (Lecture 1)

*Monday, 16 March 2026 10:15 (1 hour)*

See attached PDF for full citations and bibliography.

As the historic race to build AI infrastructure pushes projected hyperscaler spending near 7 trillion dollars by 2030, one of the primary justifications for such staggering capital deployment is the predictability offered by neural scaling laws. This presentation examines the scale-invariant nature of deep learning, defined as a power law of the form  $y = ax^b$ , where scaling inputs yields predictable gains across orders of magnitude.

The presentation first gives an overview of the field's trajectory by listing seminal work. A central focus is the resolution of the debate between Kaplan et al. and Hoffmann et al. regarding the optimal allocation of compute. We continue by challenging the Chinchilla paradigm through the lens of inference economics, showing that for widely deployed models, the optimal frontier shifts toward smaller, "over-trained" architectures to minimize inference-time costs. Furthermore, the presentation opens a discussion on the sociopolitical implications of these scaling laws. As the capital requirements for large model training skyrocket, we mention the risks of power centralization within a few major tech companies. In response, we explore emerging research on decentralized model training, investigating how scaling laws adapt to distributed compute environments and whether decentralized infrastructure can offer a counterweight to the current centralization of intelligence.

The presentation will include plots and equations from the aforementioned academic work and follow a structure similar to the presentations of Pearce and Hashimoto.

The presentation is mainly designed as a lecture.

It can also include a practical exercise involving fitting scaling law curves to provided empirical data of large model training, depending on availability.

## Number of lecture hours

2

## Number of exercise hours

1

## Attended school

CSC 2025 (Lund)

## Comments

The presentation could potentially fit into the ML track or the accelerated computing track.

**Author:** AILLET, Albert

**Presenter:** AILLET, Albert

Contribution ID: 10

Type: **not specified**

# Dimensionality Reduction: Linear Transformations and Autoencoders

*Tuesday, 17 March 2026 15:00 (1 hour)*

## The Need for Dimensionality Reduction

High-dimensional datasets often contain redundancy, noise, and complex internal structures, making direct interpretation and processing challenging. Dimensionality reduction methods offer effective solutions providing a compressed representation by projecting data into lower-dimensions while preserving their underlying structures. A set of powerful techniques are available:

- . **SVD-based Linear Transformations:** Methods like Principal Component Analysis (PCA) learn projection vectors based on data variance. These linear models perform well in near-linear sub-spaces but struggle with complex, nonlinear data manifolds.

- . **Deep Learning Approaches:** In contrast, deep learning methods such as AutoEncoders (AEs) excel at learning nonlinear mappings and extracting complex latent-space representations from data.

The compressed representation is an essential component of modern AI-driven analysis and feature abstraction.

## Lecture Overview

This lecture introduces and motivates the need for dimensionality reduction, reviews key methods, and addresses their importance in data-intensive workflows. Through conceptual learning and practical examples, participants will gain the knowledge and motivation to apply these methods in their own research problems. A practical aspect of the lecture will also cover applying these methods in post-processing and downstream tasks such as clustering, classification, and noise removal through data reconstruction.

## Lecture Outline:

### 1- Introduction to Dimension Reduction

- . Basic ideas
- . How Dimension Reduction works

### 2- Concepts and Examples

- . Principal Component Analysis
- . AutoEncoders: Fully-connected AE, Convolutional AE

### 3- Applications

- . Compression and Denoising
- . Classification and Clustering
- . Tips: How to best select the appropriate models for specific problems

## Number of lecture hours

1

## Number of exercise hours

0 (no exercises)

## Attended school

tCSC 2025 Machine Learning (Malmö)

## **Comments**

**Author:** Dr KHODABAKHSH, Athar (Helmholtz-Zentrum Berlin)

**Presenter:** Dr KHODABAKHSH, Athar (Helmholtz-Zentrum Berlin)

Contribution ID: 11

Type: **not specified**

# Keeping Things Stable: How Controllers Tackle Complexity in Modern Physics (Lecture 1)

*Tuesday, 17 March 2026 11:30 (1 hour)*

## Lecture

What do the LHC beamline, the International Space Stations orbit, and cryogenic detector temperatures have in common? They are *inherently unstable* and require precise, real-time control. Modern physics experiments and technologies depend on such controllers, and as systems grow more complex, controllers must not only *stabilize* but also *optimize* for performance, robustness and **computing efficiency**.

Every control system relies on a feedback loop: sensors measure states, a controller computes corrective actions, and actuators apply them—all under tight real-time constraints. At the LHC, for example, tiny beamline deviations must be detected and corrected through rapid calculations that keep the beam on track. Such high-precision, multi-input/multi-output systems push classical approaches like **PID** and **state-space controls** to their limit, especially when dealing with strong nonlinearities or frequent disturbances. That for, this lecture also introduces **ML-based strategies**, especially **reinforcement learning** (RL), which excels when system dynamics are complex, nonlinear, or difficult to model analytically. RL is especially promising for highly nonlinear or poorly understood dynamics, where traditional modeling becomes infeasible. Also RL can adapt in large or shifting parameter spaces and discover strategies beyond manual tuning.

The talk gives a concise overview of **modern control theory** in physics: from PID controllers and state-space methods (including LQR and MPC) to current challenges in controller synthesis and computing.

By the end, participants will understand the main families of controllers, their strengths, and when each method is the right tool for demanding real-world physics applications.

## Hands-on Exercise

In the exercise participants will design a PID controller to stabilize a classical benchmark system: the **inverted pendulum**. In a simulation (provided as a python notebook) they must choose which PID components (P, I, D) are necessary and tune them to balance overshoot, steady-state error, and stability—making core control concepts visually and intuitively clear. The participants will work in teams of two, allowing them to exchange ideas, reflect on their insights and finally find parameters for their pendulum. Participants who finish early may **explore the advanced methods** mentioned early, like the ML-based control.

Teams will then participate in a **competition**: their tuned parameters will be applied to a real inverted pendulum (the speaker will provide). Performance is judged by upright time and deviation from the setpoint, with a **prize** for the winning team (in agreement with the CSC team).

To conclude, the pendulum will also be demonstrated with LQR and ML-based controllers, enabling a direct comparison of classical, state-space, and learning-based control. The session ends with an open discussion of insights and practical lessons.

---

## Lecture Overview

- Why control theory matters in modern physics
- Real-time precision and computational challenges

- PID control: concepts, tuning, applications
- State-space control: LQR and MPC
- ML-driven control: reinforcement learning
- Summary of major controllers and their requirements

### **Practical Session Overview**

- PID control via inverted-pendulum simulation
- Competition: best real-world controller tuning
- Demonstration: PID vs. LQR vs. RL
- Final discussion and conclusions

### **Number of lecture hours**

2

### **Number of exercise hours**

1

### **Attended school**

CSC 2025 (Lund)

### **Comments**

I am based at my Home Institute in Vienna, but I prefer to hold the lecture in-person at CERN.

If another participant from the field of ML (at best RL) wants to support me with the exercise, I am very open for it.

**Author:** BOHMER, Alexander (Vienna University of Technology (AT))

**Presenter:** BOHMER, Alexander (Vienna University of Technology (AT))

Contribution ID: 12

Type: **not specified**

# Speed Without The Pain: Accelerating Python With Numba (Lecture 1)

*Tuesday, 17 March 2026 09:00 (1 hour)*

## Relevance to iCSC Attendees:

In scientific computing, we often find ourselves at a crossroads when starting a new project: do you reach for a fast, compiled language and sacrifice some development time and readability for performance, or do you reach for good old easy-to-write Python, and prepare to make a few coffees while your simulation or analysis runs? Libraries like Numpy and Scipy do a great job of bridging this divide by providing highly performant compiled backends with simple-to-use Pythonic interfaces. But what do you do when your specific workflow doesn't conform neatly to one of their vectorised functions? When we find ourselves waiting on slow, Python-only for loops, is it time to brush off our C++ skills?

This is where Numba comes in! With a few simple-to-understand decorators, not only can you just-in-time compile your plain Python code to achieve near-compiled performance, but you can also easily implement parallelism into your code –in fact, so easily that it might feel a little suspicious! The best part is that it's able to do all of this without compromising the very things that make Python such an inviting choice in the first place: high readability, fast development, and flexibility.

In these sessions, participants will learn about the difference between compiled and interpreted languages and where that stereotypical Python slowness comes from. We'll go over common use cases to show when you should use something like Numpy, and when you might want Numba instead, whilst giving a practical and hands-on overview of how you can start writing Numba code to speed up your serial (and soon-to-be-parallel) Python code!

By the end, participants will no longer be at a crossroads: they'll be able to achieve speed without the pain!

## Session Details:

Python is ubiquitous in scientific computing due to its ease-of-use and its wide support for valuable data analysis, machine learning, and numerical libraries. However, as an interpreted language, plain Python often falls short when it comes to performance, especially in cases where you can't leverage highly optimised libraries such as Numpy. These sessions aim to give attendees an overview of where these differences in performance come from, what solutions are available, and in particular give practical information on how to apply Numba to speed up their code.

Numba has been chosen due to its high performance yield for its relatively small increase in code complexity, relying mostly on decorators and very small syntax changes for parallelising for loops –this is especially useful when trying to optimise Python code which has already been written, post hoc. It is also highly portable, and is simply installed as a pip module and used alongside the normal Python interpreter, making it very accessible to install and run.

The lectures would cover:

- The difference between compiled and interpreted languages
- How slow pure Python really is, and why it's so slow
- What options are available for speeding up Python code (Numpy, interpreters, JIT, etc..)
- When these different options are suitable (in particular, why libraries like Numpy should typically be your first choice due to their high levels of optimisation for the problems they address)
- When JIT compilation may be the most appropriate solution
- How to make use of the Numba library for optimising Python code with JIT compilation:

- What Numba is capable of, and what it isn't capable of
- How to use Numba decorators to flag functions for JIT compilation
- How to use Numba parallel ranges for easy implementation of parallelism in for loops
- (Optionally) How Numba can be used to write CUDA kernels in plain Python

The exercises would cover:

- Timing and benchmarking code in Python
- Optimising some common vector operations with Numpy
- Optimising an algorithm which does not port well to Numpy using Numba JIT compilation
- Parallelising this algorithm using Numba "prange" (CPU-only, threaded computation)

### **Number of lecture hours**

2

### **Number of exercise hours**

1

### **Attended school**

CSC 2025 (Lund)

### **Comments**

If possible, I'd definitely like to travel to CERN and deliver the lectures and exercises in person.

**Author:** JONATHAN, Kiran

**Presenter:** JONATHAN, Kiran

Contribution ID: 13

Type: **not specified**

## **Cybersecurity in scientific computing: integrity, trust, and security principles in research environments (Lecture 1)**

*Wednesday, 18 March 2026 09:00 (1 hour)*

The class focuses on cybersecurity as a key element of scientific computing, with particular emphasis on technical integrity, trust, and reproducibility of results in large, distributed research environments such as CERN. It focuses on security principles, architecture, and concepts, showing how a security by design approach supports high quality in modern computing workflows. The principles discussed are cross-domain and applicable to fields that use intensive data pipelines and rely on international collaboration.

The planned session combines a high-level view of system architecture with specific technical mechanisms for protecting code, data, and pipelines. It includes an overview of threats specific to scientific infrastructure, the most common mistakes made by researchers, and practical principles for working securely with e.g. containers, HPC clusters, Grid/Cloud environments, analytical notebooks, and code repositories. Particular emphasis will be placed on data integrity issues, dependency security, architectural aspects of security in complex research projects, and the development of general best practices.

The course is not a general introduction to cybersecurity. It is assumed that participants have basic knowledge, acquired, for example, during previous editions of CERN main or security schools. However, preparatory materials may be made available to ensure that the topic is fully accessible to all participants.

### **Number of lecture hours**

2

### **Number of exercise hours**

2

### **Attended school**

CSC 2024 (Hamburg)

### **Comments**

I am open to come to CERN in person either give the lecture online.

**Author:** Mrs KOWALCZUK, Aleksandra (Accenture/WSB Merito Universities/University of Warsaw)

**Presenter:** Mrs KOWALCZUK, Aleksandra (Accenture/WSB Merito Universities/University of Warsaw)

Contribution ID: 14

Type: **not specified**

# Hell of Learning Pathologies and Stairway to Hyperparameters - a practical guide to neural networks tuning (Lecture)

*Wednesday, 18 March 2026 14:00 (1 hour)*

You have your data. After many trials, you built your neural net. You trained it, and you are pretty glad about the value of the chosen loss function. And yet... when it comes to final evaluation - it simply doesn't work. Now what?

Welcome to Failure Modes and Learning Pathologies Hell! In this lecture, I will discuss the common problems that may arise when dealing with certain types of Artificial Neural Networks (aka ML models), and also the established methods to prevent them.

That's not everything, because alongside these problems we will see the common ways to approach them. I will discuss the subtle differences between 1. the loss function that we rely on during training and 2. evaluation metrics. And finally, I would like to introduce the audience to "Stairway to Hyperparameters" - a guide on how to search the (almost) perfect architecture for their objectives with Hyperparameters Optimization.

Lecture content (just short description):

1. Highway to Hell - Introduction to Failure Modes and Learning Pathologies for various neural nets:
  - Identity Mapping, Autoencoder Collapse, Overfitting, Majority Prediction, etc.
2. Stabilization methods:
  - Regularization, dropout, early stops, batch normalization, gradient clipping, etc.
3. Proper assessment of model performance:
  - Loss function vs evaluation metrics - when they differ and why they matter?
4. Stairway to Hyperparameters
  - Hyperparameters Optimization - a crashcourse
  - Available tools and packages
  - Practical approach of "dos and dont's" when preparing the HPO tuning
  - When does HPO end? Can we tune the tuner?

## Number of lecture hours

1

## Number of exercise hours

1

## Attended school

tCSC 2025 Machine Learning (Malmö)

## Comments

**Author:** SZUMEGA, Jaroslaw (CERN, Mines Paris - PSL)

**Presenter:** SZUMEGA, Jaroslaw (CERN, Mines Paris - PSL)

Contribution ID: 16

Type: **not specified**

# Self-Attention as a Stack of Finite-Dimensional Ising Models: A Statistical-Physics View with GPU-Accelerated Prototype (Lecture)

*Wednesday, 18 March 2026 10:00 (1 hour)***Overview:**

Transformer architectures are now central to scientific and industrial machine-learning workflows. Through this lecture, a self-attention head will be recast as a finite-dimensional Ising-like spin system, where token embeddings correspond to vector spins and attention weights emerge from learned spin-spin couplings and external fields. Transformers will be explained using the statistical physics basis for the AI-machine learning system and then we will explore how such Ising-based formulations behave when implemented as actual attention layers running on modern GPU hardware.

Starting from a theoretical framework, we interpret self-attention as a composition of Ising Hamiltonians and translate it into a concrete, differentiable attention module. This module can replace or augment standard dot-product attention in practice. We make no claims of immediate performance advantages over already established Transformer implementations.

The session shall describe how token embeddings, couplings, and external fields are represented in code. It will cover how the corresponding mean-field or iterative update equations are implemented and explain how these choices affect numerical stability, expressivity, and training dynamics. These are compared to canonical implementations.

Finally, outline and demonstrate a prototype implementation. This treats Ising-inspired attention dynamics as a time-stepping scheme. The scheme used will be a HPC explicit method simulation solver that leverages GPU acceleration.

The goal is not to propose a production-ready alternative to current libraries. Instead, to show how existing HPC toolchains for stencil-based simulations can be reused. This lets us explore dynamics, scaling, and numerical behaviour of spin-based attention models while targeting non-trivial sequence lengths. Ising attention offers a physically-grounded alternative. It simplifies scaling and masking. I aim to show how it naturally maps to HPC patterns. The aim of the session is to talk about building and measuring it in practice.

**Content:**

Lecture Hour (1 hour):

0. Introduction to the statistical-physics interpretation of self-attention and the mapping from token embeddings to vector spins.
1. Overview of the Ising Hamiltonian formulation.
2. Comparison of Ising-based attention dynamics with standard dot-product attention.
3. Description of the prototype implementation: time-stepping scheme at HPC explicit method simulation solver, leveraging GPU acceleration.
4. Demonstration of attention “phase diagrams,” clustering phenomena, and numerical stability considerations.

**Exercise Session (1 hour):**

1. Writing a custom attention function based on the Ising framework
2. Exploring scaling behavior with sequence length and inspecting numerical stability
3. Comparing attention weight distributions between standard and Ising-based formulations through visualization
4. Hands-on GPU implementation of basic Ising attention components

**Number of lecture hours**

1

**Number of exercise hours**

1

**Attended school**

tCSC 2025 Machine Learning (Malmö)

**Comments**

Home institute changed, currently year-1 master student at KTH Royal Insitute of technology, Stockholm, Sweden.

**Author:** JOSHI, Lehar Laxmi

**Presenter:** JOSHI, Lehar Laxmi

Contribution ID: 17

Type: **not specified**

# Welcome to the Inverted School of Computing

*Monday, 16 March 2026 09:00 (15 minutes)*

## Comments

## Attended school

## Number of exercise hours

## Number of lecture hours

**Presenter:** PACE, Alberto (CERN)

Contribution ID: **18**

Type: **not specified**

## **Closing remarks**

*Thursday, 19 March 2026 15:00 (15 minutes)*

**Number of lecture hours**

**Number of exercise hours**

**Attended school**

**Comments**

**Presenter:** PACE, Alberto (CERN)

Contribution ID: 19

Type: **not specified**

# Modern I/O and Storage Systems for Scientific Computing (Lecture 2)

*Thursday, 19 March 2026 14:00 (1 hour)*

## Abstract

I/O bottlenecks are often the hidden performance killers in HEP data analysis, yet many scientists lack a deep understanding of how storage systems actually work. This lecture demystifies the complete I/O stack (from application code to physical storage) and introduces both classical and modern techniques for optimal performance. We'll explore fundamental concepts (buffered I/O, direct I/O, mmap), diagnose real bottlenecks using Linux tools, and cover network I/O patterns critical for Grid jobs. The session culminates with modern asynchronous I/O APIs like `io_uring` that are revolutionizing high-performance computing. Participants will leave with practical skills to optimize their analysis workflows and a solid foundation for future learning.

## Topics

### Fundamentals and Diagnosis

- I/O stack and core concepts
- Diagnosing I/O performance problems
- Optimization strategies for local storage

### Network I/O, Frameworks, and Modern APIs

- Network I/O for Grid Computing
- Modern Asynchronous I/O APIs
- Quick reference for common HEP scenarios

### Hands-On Session (examples)

- Profile and diagnose I/O bottlenecks in analysis code
- Optimize ROOT analysis workflow
- Build simple async storage system with `io_uring`

## Comments

### Attended school

tCSC 2024 Heterogenous Architectures (Belgrade)

### Number of exercise hours

1

**Number of lecture hours**

2

**Author:** DEL MONTE, Gianmaria (CERN)

**Presenter:** DEL MONTE, Gianmaria (CERN)

Contribution ID: 20

Type: **not specified**

# Modern I/O and Storage Systems for Scientific Computing (Exercices)

*Thursday, 19 March 2026 15:40 (1 hour)*

## Abstract

I/O bottlenecks are often the hidden performance killers in HEP data analysis, yet many scientists lack a deep understanding of how storage systems actually work. This lecture demystifies the complete I/O stack (from application code to physical storage) and introduces both classical and modern techniques for optimal performance. We'll explore fundamental concepts (buffered I/O, direct I/O, mmap), diagnose real bottlenecks using Linux tools, and cover network I/O patterns critical for Grid jobs. The session culminates with modern asynchronous I/O APIs like `io_uring` that are revolutionizing high-performance computing. Participants will leave with practical skills to optimize their analysis workflows and a solid foundation for future learning.

## Topics

### Fundamentals and Diagnosis

- I/O stack and core concepts
- Diagnosing I/O performance problems
- Optimization strategies for local storage

### Network I/O, Frameworks, and Modern APIs

- Network I/O for Grid Computing
- Modern Asynchronous I/O APIs
- Quick reference for common HEP scenarios

### Hands-On Session (examples)

- Profile and diagnose I/O bottlenecks in analysis code
- Optimize ROOT analysis workflow
- Build simple async storage system with `io_uring`

## Comments

## Attended school

tCSC 2024 Heterogenous Architectures (Belgrade)

## Number of exercise hours

1

**Number of lecture hours**

2

**Author:** DEL MONTE, Gianmaria (CERN)

**Presenter:** DEL MONTE, Gianmaria (CERN)

Contribution ID: 21

Type: **not specified**

## Authentication and Authorization for Global Research Infrastructure (Lecture)

*Thursday, 19 March 2026 09:00 (1 hour)*

In today's interconnected world, authentication and authorization have become critical components of modern research infrastructure. This lecture will begin by exploring traditional methods, such as certificate-based authentication, before progressing to how initiatives like **WLCG**, **IRIS** (UK computing resource federation) and **SKA SRCNet** (Square Kilometre Array - SKA Regional Centres Network) leverage industry-standard, token-based **Authentication and Authorization Infrastructure** (AAI) to enable large-scale research federations.

Next, we will examine how modernised certificate-based PKI (Public Key Infrastructure) can complement token-based systems, creating a secure and user-friendly approach for accessing research resources. This hybrid model ensures both flexibility and robustness in managing identity and access across global collaborators.

The practical exercise will give attendees hands-on experience deploying a token issuer and a certificate authority using Kubernetes Helm charts. Participants will:

1. Obtain a token from the issuer and analyse its content to understand the underlying architecture.
2. Issue a certificate and explore how token information translates into certificate issued via automated certificated issuer and how the 2 technologies can be bridged.
3. Use the issued certificate to access research resources in a simulated environment, demonstrating secure and streamlined access workflows.

By the end of the session, attendees will gain both conceptual understanding and practical skills for implementing modern AAI solutions in global research infrastructures.

### Comments

Can travel to CERN for in-person attendance.

I also attended the tCSC 2025 Abingdon but I can't multi-select.

### Attended school

CSC 2024 (Hamburg)

### Number of exercise hours

1

### Number of lecture hours

1

**Author:** CHUNG, Donald (STFC)

**Presenter:** CHUNG, Donald (STFC)

Contribution ID: 22

Type: **not specified**

## **Cybersecurity in scientific computing: integrity, trust, and security principles in research environments (Lecture 2)**

*Wednesday, 18 March 2026 11:30 (1 hour)*

The class focuses on cybersecurity as a key element of scientific computing, with particular emphasis on technical integrity, trust, and reproducibility of results in large, distributed research environments such as CERN. It focuses on security principles, architecture, and concepts, showing how a security by design approach supports high quality in modern computing workflows. The principles discussed are cross-domain and applicable to fields that use intensive data pipelines and rely on international collaboration.

The planned session combines a high-level view of system architecture with specific technical mechanisms for protecting code, data, and pipelines. It includes an overview of threats specific to scientific infrastructure, the most common mistakes made by researchers, and practical principles for working securely with e.g. containers, HPC clusters, Grid/Cloud environments, analytical notebooks, and code repositories. Particular emphasis will be placed on data integrity issues, dependency security, architectural aspects of security in complex research projects, and the development of general best practices.

The course is not a general introduction to cybersecurity. It is assumed that participants have basic knowledge, acquired, for example, during previous editions of CERN main or security schools. However, preparatory materials may be made available to ensure that the topic is fully accessible to all participants.

### **Comments**

I am open to come to CERN in person either give the lecture online.

### **Attended school**

CSC 2024 (Hamburg)

### **Number of exercise hours**

2

### **Number of lecture hours**

2

**Author:** Mrs KOWALCZUK, Aleksandra (Accenture/WSB Merito Universities/University of Warsaw)

**Presenter:** Mrs KOWALCZUK, Aleksandra (Accenture/WSB Merito Universities/University of Warsaw)

Contribution ID: 23

Type: **not specified**

## **Cybersecurity in scientific computing: integrity, trust, and security principles in research environments (Exercises 1)**

*Wednesday, 18 March 2026 15:30 (1 hour)*

The class focuses on cybersecurity as a key element of scientific computing, with particular emphasis on technical integrity, trust, and reproducibility of results in large, distributed research environments such as CERN. It focuses on security principles, architecture, and concepts, showing how a security by design approach supports high quality in modern computing workflows. The principles discussed are cross-domain and applicable to fields that use intensive data pipelines and rely on international collaboration.

The planned session combines a high-level view of system architecture with specific technical mechanisms for protecting code, data, and pipelines. It includes an overview of threats specific to scientific infrastructure, the most common mistakes made by researchers, and practical principles for working securely with e.g. containers, HPC clusters, Grid/Cloud environments, analytical notebooks, and code repositories. Particular emphasis will be placed on data integrity issues, dependency security, architectural aspects of security in complex research projects, and the development of general best practices.

The course is not a general introduction to cybersecurity. It is assumed that participants have basic knowledge, acquired, for example, during previous editions of CERN main or security schools. However, preparatory materials may be made available to ensure that the topic is fully accessible to all participants.

### **Number of lecture hours**

2

### **Number of exercise hours**

2

### **Attended school**

CSC 2024 (Hamburg)

### **Comments**

I am open to come to CERN in person either give the lecture online.

**Author:** Mrs KOWALCZUK, Aleksandra (Accenture/WSB Merito Universities/University of Warsaw)

**Presenter:** Mrs KOWALCZUK, Aleksandra (Accenture/WSB Merito Universities/University of Warsaw)

Contribution ID: 24

Type: **not specified**

## **Cybersecurity in scientific computing: integrity, trust, and security principles in research environments (Excercises 2)**

*Wednesday, 18 March 2026 16:30 (1 hour)*

The class focuses on cybersecurity as a key element of scientific computing, with particular emphasis on technical integrity, trust, and reproducibility of results in large, distributed research environments such as CERN. It focuses on security principles, architecture, and concepts, showing how a security by design approach supports high quality in modern computing workflows. The principles discussed are cross-domain and applicable to fields that use intensive data pipelines and rely on international collaboration.

The planned session combines a high-level view of system architecture with specific technical mechanisms for protecting code, data, and pipelines. It includes an overview of threats specific to scientific infrastructure, the most common mistakes made by researchers, and practical principles for working securely with e.g. containers, HPC clusters, Grid/Cloud environments, analytical notebooks, and code repositories. Particular emphasis will be placed on data integrity issues, dependency security, architectural aspects of security in complex research projects, and the development of general best practices.

The course is not a general introduction to cybersecurity. It is assumed that participants have basic knowledge, acquired, for example, during previous editions of CERN main or security schools. However, preparatory materials may be made available to ensure that the topic is fully accessible to all participants.

### **Comments**

I am open to come to CERN in person either give the lecture online.

### **Attended school**

CSC 2024 (Hamburg)

### **Number of exercise hours**

2

### **Number of lecture hours**

2

**Author:** Mrs KOWALCZUK, Aleksandra (Accenture/WSB Merito Universities/University of Warsaw)

**Presenter:** Mrs KOWALCZUK, Aleksandra (Accenture/WSB Merito Universities/University of Warsaw)

Contribution ID: 25

Type: **not specified**

## Git in Practice: Techniques for Collaborative Development (Exercises)

*Thursday, 19 March 2026 16:40 (1 hour)*

### Git in Practice: Techniques for Collaborative Development

This lecture will focus on **Git good practices and advanced commands** to help developers work more efficiently and collaboratively in shared repositories. The session will go beyond the simple *pull-commit-push* workflow and explore practical techniques for maintaining clean, reliable version histories in multi-developer environments.

#### The lecture will cover:

- **Collaborating with Git:** How Git supports teamwork and version control in modern development, with examples of effective branching strategies and workflows.
- **Branching and merging strategies:** Understanding feature branches, Git workflows, and trunk-based development to manage change efficiently.
- **Advanced Git commands:** Practical use of git rebase, git merge, git reset, git bisect, and git cherry-pick. Discussing when and how to use them safely to simplify history and resolve issues.
- **Conflict resolution:** Techniques for diagnosing and resolving merge conflicts cleanly and confidently.
- **Git hooks and CI/CD integration:** Leveraging Git hooks to automate quality checks, enforce standards, and integrate testing into CI/CD pipelines.
- **Merge requests and code reviews:** Best practices for creating, reviewing, and approving merge requests, and giving constructive feedback to maintain high code quality and consistency.

#### The accompanying exercise session would involve:

While this lecture does not naturally lend itself to a traditional exercise session, it could include a **collaborative GitHub repository activity** where participants apply these techniques in a shared environment to simulate a larger development team. Alternatively, a less collaborative exercise could involve **improving the commit history** of a sample repository, using techniques such as rebasing or squashing to create a cleaner, more linear history.

### Comments

I would love to be able to travel to CERN from my home organisation to attend the school.

### Attended school

CSC 2025 (Lund)

### Number of exercise hours

1

**Number of lecture hours**

1

**Author:** SALMON, Lizzie

**Presenter:** SALMON, Lizzie

Contribution ID: 26

Type: **not specified**

# Hell of Learning Pathologies and Stairway to Hyperparameters - a practical guide to neural networks tuning (Excercises)

*Wednesday, 18 March 2026 16:30 (1 hour)*

You have your data. After many trials, you built your neural net. You trained it, and you are pretty glad about the value of the chosen loss function. And yet... when it comes to final evaluation - it simply doesn't work. Now what?

Welcome to Failure Modes and Learning Pathologies Hell! In this lecture, I will discuss the common problems that may arise when dealing with certain types of Artificial Neural Networks (aka ML models), and also the established methods to prevent them.

That's not everything, because alongside these problems we will see the common ways to approach them. I will discuss the subtle differences between 1. the loss function that we rely on during training and 2. evaluation metrics. And finally, I would like to introduce the audience to "Stairway to Hyperparameters" - a guide on how to search the (almost) perfect architecture for their objectives with Hyperparameters Optimization.

Lecture content (just short description):

1. Highway to Hell - Introduction to Failure Modes and Learning Pathologies for various neural nets:
  - Identity Mapping, Autoencoder Collapse, Overfitting, Majority Prediction, etc.
2. Stabilization methods:
  - Regularization, dropout, early stops, batch normalization, gradient clipping, etc.
3. Proper assessment of model performance:
  - Loss function vs evaluation metrics - when they differ and why they matter?
4. Stairway to Hyperparameters
  - Hyperparameters Optimization - a crashcourse
  - Available tools and packages
  - Practical approach of "dos and dont's" when preparing the HPO tuning
  - When does HPO end? Can we tune the tuner?

## Comments

## Attended school

tCSC 2025 Machine Learning (Malmö)

## Number of exercise hours

1

## Number of lecture hours

1

**Author:** SZUMEGA, Jaroslaw (CERN, Mines Paris - PSL)

**Presenter:** SZUMEGA, Jaroslaw (CERN, Mines Paris - PSL)

Contribution ID: 27

Type: **not specified**

## How performance analysis of real software works: the case of MadGraph (Lecture 2)

*Monday, 16 March 2026 14:00 (1 hour)*

The growing computational demands of high-precision simulations for the High-Luminosity LHC (HL-LHC) era require a careful optimization of Monte Carlo event generators.

In these lectures, we will explore different profiling techniques and their application to MadGraph5\_aMC@NLO, one of the most widely used software in high-energy physics to simulate particle collisions:

- We will begin with an overview of state-of-the-art profiling methodologies and tools for performance analysis on modern architectures.
- These techniques will then be applied to MadGraph, with a particular focus on Next-to-Leading Order (NLO) computations, which represent a major computational challenge.
- By studying the performance characteristics of the various components of the code, we will try to identify areas for optimization and discuss strategies to exploit hardware acceleration, including data-parallel and heterogeneous computing paradigms.

---

The goal of the lectures is to provide both conceptual and practical insights into how systematic profiling can guide software optimization, enabling efficient use of next-generation computing resources.

### Comments

### Attended school

tCSC 2025 Heterogenous Architectures (Split)

### Number of exercise hours

1

### Number of lecture hours

2

**Author:** MASSARO, Daniele (CERN)

**Presenter:** MASSARO, Daniele (CERN)

Contribution ID: 28

Type: **not specified**

## How performance analysis of real software works: the case of MadGraph (Exercices)

*Monday, 16 March 2026 16:30 (1 hour)*

The growing computational demands of high-precision simulations for the High-Luminosity LHC (HL-LHC) era require a careful optimization of Monte Carlo event generators.

In these lectures, we will explore different profiling techniques and their application to MadGraph5\_aMC@NLO, one of the most widely used software in high-energy physics to simulate particle collisions:

- We will begin with an overview of state-of-the-art profiling methodologies and tools for performance analysis on modern architectures.
- These techniques will then be applied to MadGraph, with a particular focus on Next-to-Leading Order (NLO) computations, which represent a major computational challenge.
- By studying the performance characteristics of the various components of the code, we will try to identify areas for optimization and discuss strategies to exploit hardware acceleration, including data-parallel and heterogeneous computing paradigms.

---

The goal of the lectures is to provide both conceptual and practical insights into how systematic profiling can guide software optimization, enabling efficient use of next-generation computing resources.

### Number of lecture hours

2

### Number of exercise hours

1

### Attended school

tCSC 2025 Heterogenous Architectures (Split)

### Comments

**Author:** MASSARO, Daniele (CERN)**Presenter:** MASSARO, Daniele (CERN)

Contribution ID: 29

Type: **not specified**

# Keeping Things Stable: How Controllers Tackle Complexity in Modern Physics (Lecture 1)

*Tuesday, 17 March 2026 14:00 (1 hour)*

## Lecture

What do the LHC beamline, the International Space Stations orbit, and cryogenic detector temperatures have in common? They are *inherently unstable* and require precise, real-time control. Modern physics experiments and technologies depend on such controllers, and as systems grow more complex, controllers must not only *stabilize* but also *optimize* for performance, robustness and **computing efficiency**.

Every control system relies on a feedback loop: sensors measure states, a controller computes corrective actions, and actuators apply them—all under tight real-time constraints. At the LHC, for example, tiny beamline deviations must be detected and corrected through rapid calculations that keep the beam on track. Such high-precision, multi-input/multi-output systems push classical approaches like **PID** and **state-space controls** to their limit, especially when dealing with strong nonlinearities or frequent disturbances. That for, this lecture also introduces **ML-based strategies**, especially **reinforcement learning** (RL), which excels when system dynamics are complex, nonlinear, or difficult to model analytically. RL is especially promising for highly nonlinear or poorly understood dynamics, where traditional modeling becomes infeasible. Also RL can adapt in large or shifting parameter spaces and discover strategies beyond manual tuning.

The talk gives a concise overview of **modern control theory** in physics: from PID controllers and state-space methods (including LQR and MPC) to current challenges in controller synthesis and computing.

By the end, participants will understand the main families of controllers, their strengths, and when each method is the right tool for demanding real-world physics applications.

## Hands-on Exercise

In the exercise participants will design a PID controller to stabilize a classical benchmark system: the **inverted pendulum**. In a simulation (provided as a python notebook) they must choose which PID components (P, I, D) are necessary and tune them to balance overshoot, steady-state error, and stability—making core control concepts visually and intuitively clear. The participants will work in teams of two, allowing them to exchange ideas, reflect on their insights and finally find parameters for their pendulum. Participants who finish early may **explore the advanced methods** mentioned early, like the ML-based control.

Teams will then participate in a **competition**: their tuned parameters will be applied to a real inverted pendulum (the speaker will provide). Performance is judged by upright time and deviation from the setpoint, with a **prize** for the winning team (in agreement with the CSC team).

To conclude, the pendulum will also be demonstrated with LQR and ML-based controllers, enabling a direct comparison of classical, state-space, and learning-based control. The session ends with an open discussion of insights and practical lessons.

---

## Lecture Overview

- Why control theory matters in modern physics
- Real-time precision and computational challenges

- PID control: concepts, tuning, applications
- State-space control: LQR and MPC
- ML-driven control: reinforcement learning
- Summary of major controllers and their requirements

### **Practical Session Overview**

- PID control via inverted-pendulum simulation
- Competition: best real-world controller tuning
- Demonstration: PID vs. LQR vs. RL
- Final discussion and conclusions

### **Comments**

I am based at my Home Institute in Vienna, but I prefer to hold the lecture in-person at CERN.

If another participant from the field of ML (at best RL) wants to support me with the exercise, I am very open for it.

### **Attended school**

CSC 2025 (Lund)

### **Number of exercise hours**

1

### **Number of lecture hours**

2

**Author:** BOHMER, Alexander (Vienna University of Technology (AT))

**Presenter:** BOHMER, Alexander (Vienna University of Technology (AT))

Contribution ID: 30

Type: **not specified**

# Keeping Things Stable: How Controllers Tackle Complexity in Modern Physics (Exercices)

*Tuesday, 17 March 2026 17:30 (1 hour)*

## Lecture

What do the LHC beamline, the International Space Stations orbit, and cryogenic detector temperatures have in common? They are *inherently unstable* and require precise, real-time control. Modern physics experiments and technologies depend on such controllers, and as systems grow more complex, controllers must not only *stabilize* but also *optimize* for performance, robustness and **computing efficiency**.

Every control system relies on a feedback loop: sensors measure states, a controller computes corrective actions, and actuators apply them—all under tight real-time constraints. At the LHC, for example, tiny beamline deviations must be detected and corrected through rapid calculations that keep the beam on track. Such high-precision, multi-input/multi-output systems push classical approaches like **PID** and **state-space controls** to their limit, especially when dealing with strong nonlinearities or frequent disturbances. That for, this lecture also introduces **ML-based strategies**, especially **reinforcement learning** (RL), which excels when system dynamics are complex, nonlinear, or difficult to model analytically. RL is especially promising for highly nonlinear or poorly understood dynamics, where traditional modeling becomes infeasible. Also RL can adapt in large or shifting parameter spaces and discover strategies beyond manual tuning.

The talk gives a concise overview of **modern control theory** in physics: from PID controllers and state-space methods (including LQR and MPC) to current challenges in controller synthesis and computing.

By the end, participants will understand the main families of controllers, their strengths, and when each method is the right tool for demanding real-world physics applications.

## Hands-on Exercise

In the exercise participants will design a PID controller to stabilize a classical benchmark system: the **inverted pendulum**. In a simulation (provided as a python notebook) they must choose which PID components (P, I, D) are necessary and tune them to balance overshoot, steady-state error, and stability—making core control concepts visually and intuitively clear. The participants will work in teams of two, allowing them to exchange ideas, reflect on their insights and finally find parameters for their pendulum. Participants who finish early may **explore the advanced methods** mentioned early, like the ML-based control.

Teams will then participate in a **competition**: their tuned parameters will be applied to a real inverted pendulum (the speaker will provide). Performance is judged by upright time and deviation from the setpoint, with a **prize** for the winning team (in agreement with the CSC team).

To conclude, the pendulum will also be demonstrated with LQR and ML-based controllers, enabling a direct comparison of classical, state-space, and learning-based control. The session ends with an open discussion of insights and practical lessons.

---

## Lecture Overview

- Why control theory matters in modern physics
- Real-time precision and computational challenges

- PID control: concepts, tuning, applications
- State-space control: LQR and MPC
- ML-driven control: reinforcement learning
- Summary of major controllers and their requirements

**Practical Session Overview**

- PID control via inverted-pendulum simulation
- Competition: best real-world controller tuning
- Demonstration: PID vs. LQR vs. RL
- Final discussion and conclusions

**Number of lecture hours**

2

**Number of exercise hours**

1

**Attended school**

CSC 2025 (Lund)

**Comments**

I am based at my Home Institute in Vienna, but I prefer to hold the lecture in-person at CERN.

If another participant from the field of ML (at best RL) wants to support me with the exercise, I am very open for it.

**Author:** BOHMER, Alexander (Vienna University of Technology (AT))

**Presenter:** BOHMER, Alexander (Vienna University of Technology (AT))

Contribution ID: 32

Type: **not specified**

## Neural Scaling Laws (Lecture 2)

*Monday, 16 March 2026 11:40 (1 hour)*

See attached PDF for full citations and bibliography.

As the historic race to build AI infrastructure pushes projected hyperscaler spending near 7 trillion dollars by 2030, one of the primary justifications for such staggering capital deployment is the predictability offered by neural scaling laws. This presentation examines the scale-invariant nature of deep learning, defined as a power law of the form  $y = ax^b$ , where scaling inputs yields predictable gains across orders of magnitude.

The presentation first gives an overview of the field's trajectory by listing seminal work. A central focus is the resolution of the debate between Kaplan et al. and Hoffmann et al. regarding the optimal allocation of compute. We continue by challenging the Chinchilla paradigm through the lens of inference economics, showing that for widely deployed models, the optimal frontier shifts toward smaller, "over-trained" architectures to minimize inference-time costs. Furthermore, the presentation opens a discussion on the sociopolitical implications of these scaling laws. As the capital requirements for large model training skyrocket, we mention the risks of power centralization within a few major tech companies. In response, we explore emerging research on decentralized model training, investigating how scaling laws adapt to distributed compute environments and whether decentralized infrastructure can offer a counterweight to the current centralization of intelligence.

The presentation will include plots and equations from the aforementioned academic work and follow a structure similar to the presentations of Pearce and Hashimoto.

The presentation is mainly designed as a lecture.

It can also include a practical exercise involving fitting scaling law curves to provided empirical data of large model training, depending on availability.

### Comments

The presentation could potentially fit into the ML track or the accelerated computing track.

### Attended school

CSC 2025 (Lund)

### Number of exercise hours

1

### Number of lecture hours

2

**Author:** AILLET, Albert

**Presenter:** AILLET, Albert

Contribution ID: 33

Type: **not specified**

## Neural Scaling Laws (Exercices)

*Monday, 16 March 2026 16:30 (1 hour)*

See attached PDF for full citations and bibliography.

As the historic race to build AI infrastructure pushes projected hyperscaler spending near 7 trillion dollars by 2030, one of the primary justifications for such staggering capital deployment is the predictability offered by neural scaling laws. This presentation examines the scale-invariant nature of deep learning, defined as a power law of the form  $y = ax^b$ , where scaling inputs yields predictable gains across orders of magnitude.

The presentation first gives an overview of the field's trajectory by listing seminal work. A central focus is the resolution of the debate between Kaplan et al. and Hoffmann et al. regarding the optimal allocation of compute. We continue by challenging the Chinchilla paradigm through the lens of inference economics, showing that for widely deployed models, the optimal frontier shifts toward smaller, "over-trained" architectures to minimize inference-time costs. Furthermore, the presentation opens a discussion on the sociopolitical implications of these scaling laws. As the capital requirements for large model training skyrocket, we mention the risks of power centralization within a few major tech companies. In response, we explore emerging research on decentralized model training, investigating how scaling laws adapt to distributed compute environments and whether decentralized infrastructure can offer a counterweight to the current centralization of intelligence.

The presentation will include plots and equations from the aforementioned academic work and follow a structure similar to the presentations of Pearce and Hashimoto.

The presentation is mainly designed as a lecture.

It can also include a practical exercise involving fitting scaling law curves to provided empirical data of large model training, depending on availability.

### Comments

The presentation could potentially fit into the ML track or the accelerated computing track.

### Attended school

CSC 2025 (Lund)

### Number of exercise hours

1

### Number of lecture hours

2

**Author:** AILLET, Albert

**Presenter:** AILLET, Albert

Contribution ID: 34

Type: **not specified**

# Self-Attention as a Stack of Finite-Dimensional Ising Models: A Statistical-Physics View with GPU-Accelerated Prototype (Exercises)

*Wednesday, 18 March 2026 15:30 (1 hour)***Overview:**

Transformer architectures are now central to scientific and industrial machine-learning workflows. Through this lecture, a self-attention head will be recast as a finite-dimensional Ising-like spin system, where token embeddings correspond to vector spins and attention weights emerge from learned spin-spin couplings and external fields. Transformers will be explained using the statistical physics basis for the AI-machine learning system and then we will explore how such Ising-based formulations behave when implemented as actual attention layers running on modern GPU hardware.

Starting from a theoretical framework, we interpret self-attention as a composition of Ising Hamiltonians and translate it into a concrete, differentiable attention module. This module can replace or augment standard dot-product attention in practice. We make no claims of immediate performance advantages over already established Transformer implementations.

The session shall describe how token embeddings, couplings, and external fields are represented in code. It will cover how the corresponding mean-field or iterative update equations are implemented and explain how these choices affect numerical stability, expressivity, and training dynamics. These are compared to canonical implementations.

Finally, outline and demonstrate a prototype implementation. This treats Ising-inspired attention dynamics as a time-stepping scheme. The scheme used will be a HPC explicit method simulation solver that leverages GPU acceleration.

The goal is not to propose a production-ready alternative to current libraries. Instead, to show how existing HPC toolchains for stencil-based simulations can be reused. This lets us explore dynamics, scaling, and numerical behaviour of spin-based attention models while targeting non-trivial sequence lengths. Ising attention offers a physically-grounded alternative. It simplifies scaling and masking. I aim to show how it naturally maps to HPC patterns. The aim of the session is to talk about building and measuring it in practice.

**Content:**

Lecture Hour (1 hour):

0. Introduction to the statistical-physics interpretation of self-attention and the mapping from token embeddings to vector spins.
1. Overview of the Ising Hamiltonian formulation.
2. Comparison of Ising-based attention dynamics with standard dot-product attention.
3. Description of the prototype implementation: time-stepping scheme at HPC explicit method simulation solver, leveraging GPU acceleration.
4. Demonstration of attention “phase diagrams,” clustering phenomena, and numerical stability considerations.

**Exercise Session (1 hour):**

1. Writing a custom attention function based on the Ising framework
2. Exploring scaling behavior with sequence length and inspecting numerical stability
3. Comparing attention weight distributions between standard and Ising-based formulations through visualization
4. Hands-on GPU implementation of basic Ising attention components

## **Comments**

Home institute changed, currently year-1 master student at KTH Royal Institute of technology, Stockholm, Sweden.

## **Attended school**

tCSC 2025 Machine Learning (Malmö)

## **Number of exercise hours**

1

## **Number of lecture hours**

1

**Author:** JOSHI, Lehar Laxmi

**Presenter:** JOSHI, Lehar Laxmi

Contribution ID: 35

Type: **not specified**

## Speed Without The Pain: Accelerating Python With Numba (Lecture 2)

*Tuesday, 17 March 2026 10:00 (1 hour)*

### Relevance to iCSC Attendees:

In scientific computing, we often find ourselves at a crossroads when starting a new project: do you reach for a fast, compiled language and sacrifice some development time and readability for performance, or do you reach for good old easy-to-write Python, and prepare to make a few coffees while your simulation or analysis runs? Libraries like Numpy and Scipy do a great job of bridging this divide by providing highly performant compiled backends with simple-to-use Pythonic interfaces. But what do you do when your specific workflow doesn't conform neatly to one of their vectorised functions? When we find ourselves waiting on slow, Python-only for loops, is it time to brush off our C++ skills?

This is where Numba comes in! With a few simple-to-understand decorators, not only can you just-in-time compile your plain Python code to achieve near-compiled performance, but you can also easily implement parallelism into your code –in fact, so easily that it might feel a little suspicious! The best part is that it's able to do all of this without compromising the very things that make Python such an inviting choice in the first place: high readability, fast development, and flexibility.

In these sessions, participants will learn about the difference between compiled and interpreted languages and where that stereotypical Python slowness comes from. We'll go over common use cases to show when you should use something like Numpy, and when you might want Numba instead, whilst giving a practical and hands-on overview of how you can start writing Numba code to speed up your serial (and soon-to-be-parallel) Python code!

By the end, participants will no longer be at a crossroads: they'll be able to achieve speed without the pain!

### Session Details:

Python is ubiquitous in scientific computing due to its ease-of-use and its wide support for valuable data analysis, machine learning, and numerical libraries. However, as an interpreted language, plain Python often falls short when it comes to performance, especially in cases where you can't leverage highly optimised libraries such as Numpy. These sessions aim to give attendees an overview of where these differences in performance come from, what solutions are available, and in particular give practical information on how to apply Numba to speed up their code.

Numba has been chosen due to its high performance yield for its relatively small increase in code complexity, relying mostly on decorators and very small syntax changes for parallelising for loops –this is especially useful when trying to optimise Python code which has already been written, post hoc. It is also highly portable, and is simply installed as a pip module and used alongside the normal Python interpreter, making it very accessible to install and run.

The lectures would cover:

- The difference between compiled and interpreted languages
- How slow pure Python really is, and why it's so slow
- What options are available for speeding up Python code (Numpy, interpreters, JIT, etc..)
- When these different options are suitable (in particular, why libraries like Numpy should typically be your first choice due to their high levels of optimisation for the problems they address)
- When JIT compilation may be the most appropriate solution
- How to make use of the Numba library for optimising Python code with JIT compilation:

- What Numba is capable of, and what it isn't capable of
- How to use Numba decorators to flag functions for JIT compilation
- How to use Numba parallel ranges for easy implementation of parallelism in for loops
- (Optionally) How Numba can be used to write CUDA kernels in plain Python

The exercises would cover:

- Timing and benchmarking code in Python
- Optimising some common vector operations with Numpy
- Optimising an algorithm which does not port well to Numpy using Numba JIT compilation
- Parallelising this algorithm using Numba "prange" (CPU-only, threaded computation)

## Comments

If possible, I'd definitely like to travel to CERN and deliver the lectures and exercises in person.

## Attended school

CSC 2025 (Lund)

## Number of exercise hours

1

## Number of lecture hours

2

**Author:** JONATHAN, Kiran

**Presenter:** JONATHAN, Kiran

Contribution ID: 36

Type: **not specified**

## Speed Without The Pain: Accelerating Python With Numba (Exercises)

*Tuesday, 17 March 2026 16:30 (1 hour)*

### Relevance to iCSC Attendees:

In scientific computing, we often find ourselves at a crossroads when starting a new project: do you reach for a fast, compiled language and sacrifice some development time and readability for performance, or do you reach for good old easy-to-write Python, and prepare to make a few coffees while your simulation or analysis runs? Libraries like Numpy and Scipy do a great job of bridging this divide by providing highly performant compiled backends with simple-to-use Pythonic interfaces. But what do you do when your specific workflow doesn't conform neatly to one of their vectorised functions? When we find ourselves waiting on slow, Python-only for loops, is it time to brush off our C++ skills?

This is where Numba comes in! With a few simple-to-understand decorators, not only can you just-in-time compile your plain Python code to achieve near-compiled performance, but you can also easily implement parallelism into your code –in fact, so easily that it might feel a little suspicious! The best part is that it's able to do all of this without compromising the very things that make Python such an inviting choice in the first place: high readability, fast development, and flexibility.

In these sessions, participants will learn about the difference between compiled and interpreted languages and where that stereotypical Python slowness comes from. We'll go over common use cases to show when you should use something like Numpy, and when you might want Numba instead, whilst giving a practical and hands-on overview of how you can start writing Numba code to speed up your serial (and soon-to-be-parallel) Python code!

By the end, participants will no longer be at a crossroads: they'll be able to achieve speed without the pain!

### Session Details:

Python is ubiquitous in scientific computing due to its ease-of-use and its wide support for valuable data analysis, machine learning, and numerical libraries. However, as an interpreted language, plain Python often falls short when it comes to performance, especially in cases where you can't leverage highly optimised libraries such as Numpy. These sessions aim to give attendees an overview of where these differences in performance come from, what solutions are available, and in particular give practical information on how to apply Numba to speed up their code.

Numba has been chosen due to its high performance yield for its relatively small increase in code complexity, relying mostly on decorators and very small syntax changes for parallelising for loops –this is especially useful when trying to optimise Python code which has already been written, post hoc. It is also highly portable, and is simply installed as a pip module and used alongside the normal Python interpreter, making it very accessible to install and run.

The lectures would cover:

- The difference between compiled and interpreted languages
- How slow pure Python really is, and why it's so slow
- What options are available for speeding up Python code (Numpy, interpreters, JIT, etc..)
- When these different options are suitable (in particular, why libraries like Numpy should typically be your first choice due to their high levels of optimisation for the problems they address)
- When JIT compilation may be the most appropriate solution
- How to make use of the Numba library for optimising Python code with JIT compilation:

- What Numba is capable of, and what it isn't capable of
- How to use Numba decorators to flag functions for JIT compilation
- How to use Numba parallel ranges for easy implementation of parallelism in for loops
- (Optionally) How Numba can be used to write CUDA kernels in plain Python

The exercises would cover:

- Timing and benchmarking code in Python
- Optimising some common vector operations with Numpy
- Optimising an algorithm which does not port well to Numpy using Numba JIT compilation
- Parallellising this algorithm using Numba "prange" (CPU-only, threaded computation)

### **Number of lecture hours**

2

### **Number of exercise hours**

1

### **Attended school**

CSC 2025 (Lund)

### **Comments**

If possible, I'd definitely like to travel to CERN and deliver the lectures and exercises in person.

**Author:** JONATHAN, Kiran

**Presenter:** JONATHAN, Kiran