Developing track- and event reconstruction using deep learning with quantifiable accuracy for the PANDA at FAIR experiment

Sverker Holmgren, Michael Papenbrock, Karin Schönning

1 Introduction

PANDA (proton-antiProton ANnihilations in DArmstadt) at FAIR (Facility for Antiproton and Ion Research) (see Figure 1) is a *Big Science* facility currently under construction in Darmstadt, Germany. It has more than 450 members in 19 countries. PANDA will be part of a new generation of experiments that relies entirely on software filtering for online data selection. With an event rate of up to 20 MHz and a corresponding raw data rate of up to 200 GB/s, this poses a particular challenge, both in terms of identifying efficient implementations based on currently available high performance computing architectures, but also in terms of how to analyze, validate, and improve the accuracy of the models involved.

These tasks require a merging of knowledge of physics, the nature of computational hardware and numerical behaviour, and the mathematics of deep artificial neural networks.

PANDA at FAIR will shed light on some of the most challenging problems in curiosity-driven contemporary physics. One is the strong interaction - one of four fundamental forces in Nature. Though generating 99% of the visible mass of the Universe, it is not yet understood from first principles. Another challenge is to understand why there is so much more matter in the universe than antimatter. Equal amounts should have emerged from the Big Bang, but something put the Universe out of the equilibrium. This is referred to as *baryogenesis*. In Uppsala, we use *hyperons* as a diagnostic tool to study these questions. Hyperons are similar to protons, but have one or more of their light quarks replaced by a heavier *strange* or *charm* quark. Hyperons provide a new angle on the strong interaction as well as baryogenesis. However, hyperons also come with challenges, in particular due to their complex decay topology, and have therefore not been objected to the thorough investigations we have seen in *e.g.* the case of the proton. The PANDA at FAIR experiment is perfectly suited to study hyperons thanks to the large production probability in antiproton-proton collisions.

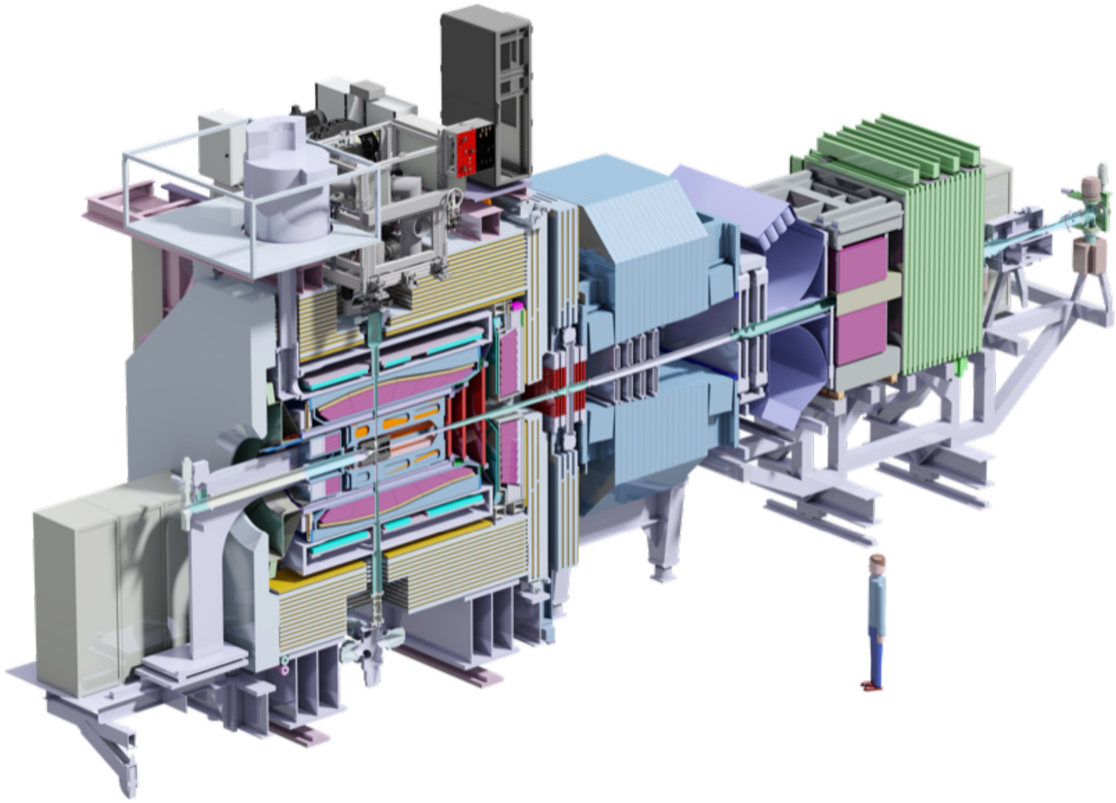

Figure 1: The PANDA detector.

2 Project aim

The goal of this project is to develop reconstruction algorithms for hyperon events based on neural networks that can run in a high performance environment. Due to their distinct topology, hyperon events require dedicated reconstruction algorithms. Hence, the project can be broken down into several areas.

Track and event reconstruction As tracks and noise create hit signals in the various subdetectors, a number of tasks have to be completed in order to extract meaningful information. Hits have to be clustered into tracks, which then have to be fitted and grouped into events. This space is commonly occupied by more traditional algorithms such as Kalman filter techniques, cellular automata, or Hough transformations. Preliminary tests with neural networks on the other hand have shown significant potential for these tasks while their performance is expected to scale better with detector occupancy.

Hardware acceleration It has been shown that the architecture of Graphics Processing Units (GPUs) lends itself very well to accelerating neural networks. Corresponding software libraries to utilize the power of these devices are widely available today. The project will explore this and other acceleration architectures for accommodating the real-time high performance needs of the experiment.

Improving accuracy in training The basic design of multi-layer non-linear artificial neural networks have been available for decades. The recent surge in deep learning is sometimes solely attributed to advances in computational hardware, but the truth is far more nuanced. The proper application of deep learning will require a suitable network structure, that not only can properly model the problem at hand, but that can also be trained to do so within reasonable time. For *e.g.* image analysis of photos and machine translation, best practice has been established, but this has frequently been based on a combination of empirical observation and mathematical understanding. We will develop suitable network structures, loss functions, and training schemes for our specific application.

By building on and developing existing analysis of optimization schemes from scientific computing, we hope that our results will have relevance for other deep learning applications, and also for theory for other very high-dimensional numerical optimization problems.

Controlling uncertainty The non-linear nature and general structure of a deep learning network can make it sort of a "black box", where it is hard to understand what factors contribute to the accuracy of the model, and also what the effect is of perturbations in the input data.

Nonetheless, there are nascent techniques for validating and quantifying the uncertainty in trained networks. We will use such techniques, and also integrate and extend uncertainty quantification methods from other scientific computing problems. A specific subproblem in our application is the work needed to ensure that the model does not overfit to differences between simulated and real-world experimental data.

3 Financing

The student would have Sverker Holmgren as the main supervisor and be located at the Division of Scientific Computing, with assistant supervisor Michael Papenbrock from the Department for Physics and Astronomy. The division of Scientific Computing will support the project with 25% of faculty funding, and appropriate supervisor resources, as well as general support services. The physics partners contribute with 25% from faculty funding, supervisor resources and some travel funding for PANDA meetings and workshops.