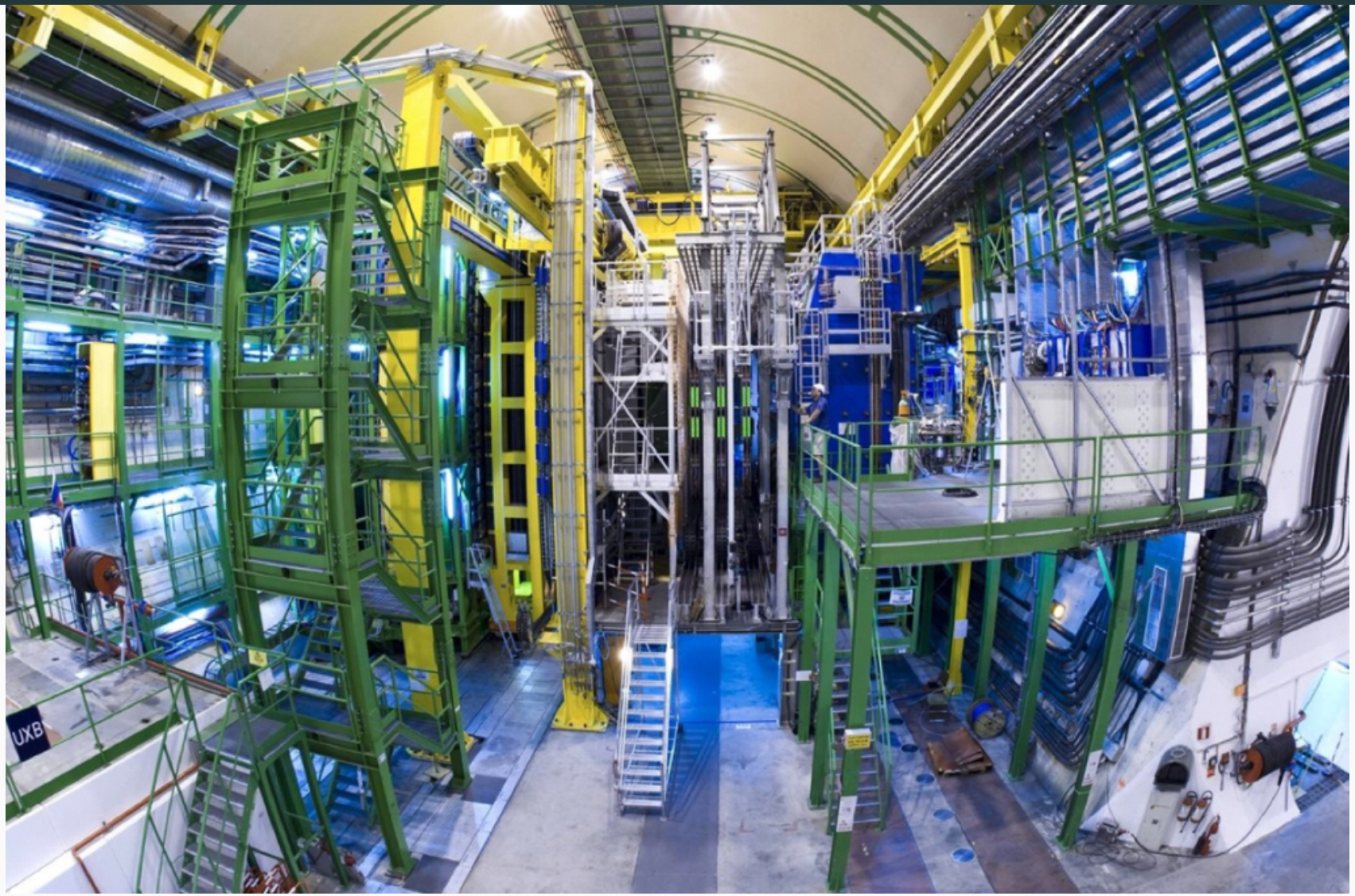
LHCb Machine Learning challenges

Niko Neufeld[1], Stefan Roiser[1], Andrey Ustyuzhanin[1,2]

April 29, 2016

[1]CERN LHCb, [2]Yandex School of Data Analysis

Overview



Detector Operation & Data Quality

The first frontier of data taking

- Data Quality Monitoring
- Trigger operation anomaly detection
- Online farm hardware failure monitoring and prediction

Triggers Optimization

Triggers are event selection procedures that should filter out uninteresting events

- Topological trigger (finds decays of certain topologies)
- ~100 triggers
  - Quota management for trigger lines
  - Efficiency analysis

Tracking

Tracking - identification of particle track by hit patterns

- usage of advanced algorithms and computational architectures for efficient track finding

Grid Data Storage Optimization

Events are stored in the LHCb grid for a longer term

- Access pattern analysis
- Event indexing
- Job scheduling

Data Preprocessing

Before releasing the collected data it is pre-processed by several algorithms that use machine learning already

- Particle identification
- Flavour tagging
- Jet tagging

Data Analysis

Once data is stored in the grid, it should be analysed as accurately as possible

- Event selection
  - Quality metrics definition/selection
  - Algorithm training and optimization
- Training on Data vs MC
- Can data processing happen automatically right as it is taken?
  - *turbo* stream (`https://cds.cern.ch/record/2011573`)
  - many more steps to go...
- Collabroative data analysis workflow

Simulation

Simulation is the largest user of CPU time.

- Generative models

Data Published & Outreach

Flavours of physics Kaggle challenge:

`https://kaggle.com/c/flavours-of-physics`

Heavy Flavour Data Minig workshop, Feb'16

`https://indico.cern.ch/event/433556/`

Thank you!