

ML & experimental particle physics

Johannes Albrecht (TU Dortmund & Lamarr)

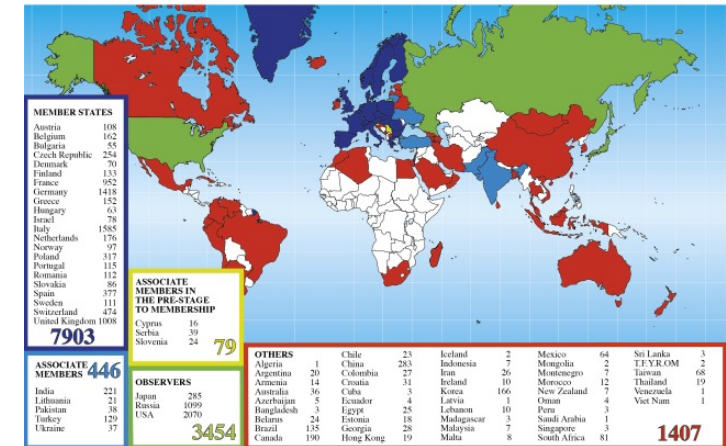
20.03.2024



Many references to B. Mitreska and Q. Führung, Lamarr meeting 8.11.23

- Particle Physics: Experimental situation very interesting
 - Entire community joined forces at one accelerator center at a time
 - extremely focussed field
 - We have 10 years of LHC data and will collect another 20
 - huge dataset to exploit optimally

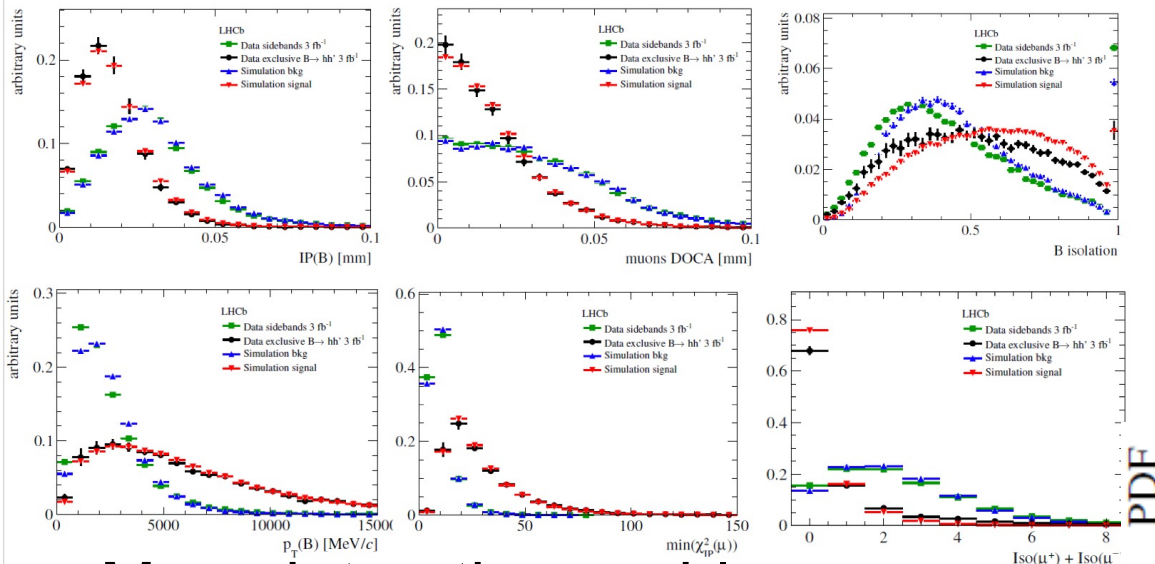
Distribution of All CERN Users by Location of Institute on 24 January 2018



- Modern Particle Physics is (in a part) data science
 - Process 4TB/s of data in „real time“
 - Many hypothesis based approaches, often optimal, sometimes not?
 - Better access to data in any way has large potential
- Particle physicists are not (entirely) naive
 - No quick solutions, joint projects needed and started within Lamarr, more collaborations welcome

- Typical: many variables, all with some separation power

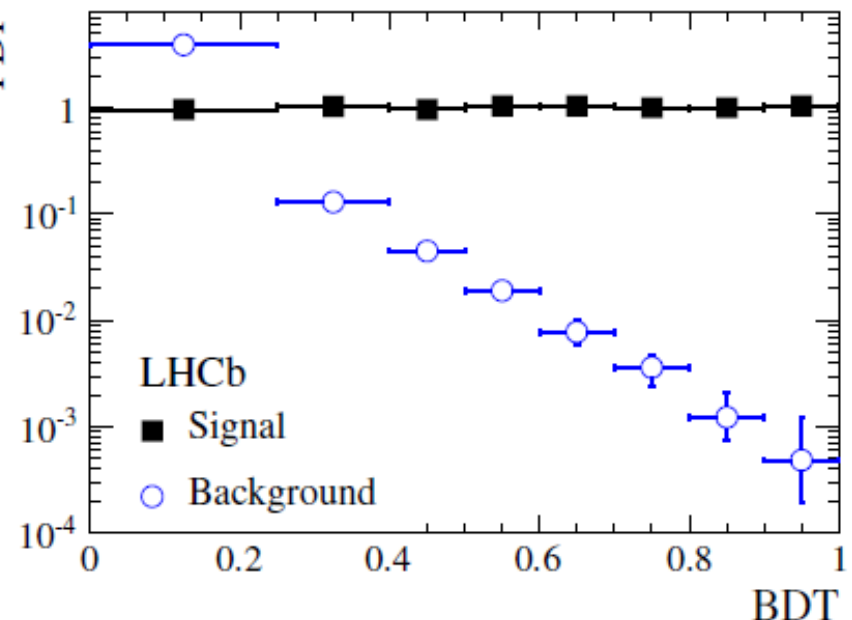
- Total of 12 variables, including:

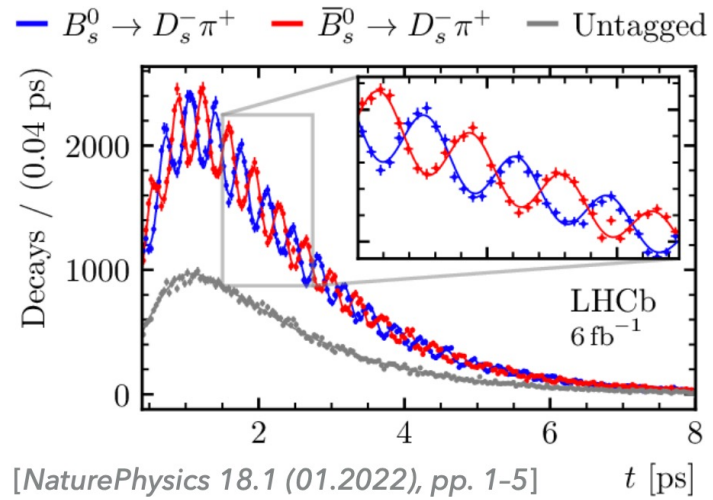


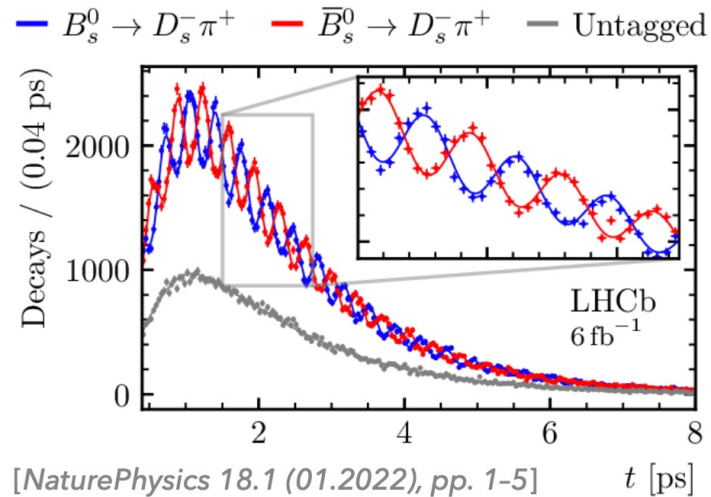
- Combine variables in classifier (BDT, NN, ..)
- Problem: simulation & signal proxy
- Put some thought in Labelling ?

- Many interesting problems:

- Simulation: enormous resources, event generation (learning possible?), detector response, etc
- Smart use of entire event?
- Hypothesis based analysis
→ learned Standard Model ?

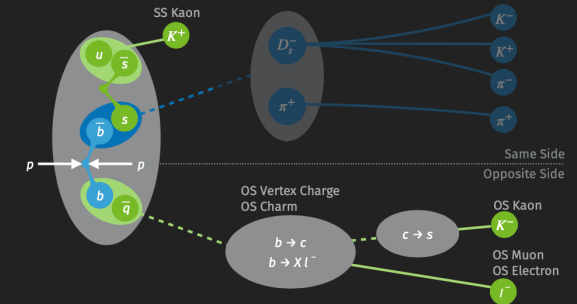






Flavour Tagging algorithms

- ▶ Currently dedicated track-based algorithms for each process
 - ▶ SSK (for B_s^0), $SS\pi$ (for B^0), SSp (for B^0), OSK, OS μ , OS e , OS c , OS-VtxCharge
 - ▶ Rectangular selection specific to process (based on PID and event topology)
- ▶ Two output values:
 - ▶ Tag decision $d \in \{-1, 0, 1\}$: prediction of initial flavour based on track charge
 - ▶ Mistag estimate $\eta \in [0, 0.5]$: MVA-based probability of wrong tag decision



Q. Fühling

Hybrid-ML: High energy physics and ML

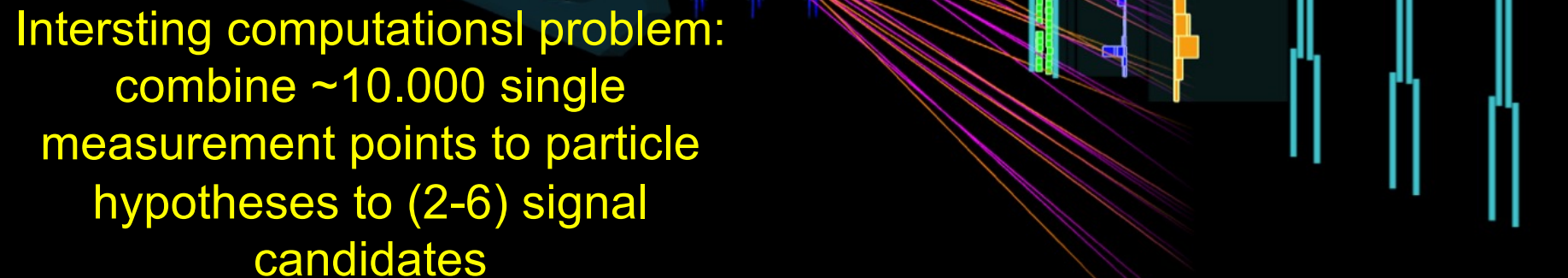
08.11.2023

- Limits performance of analysis, currently, we loose factor 20
- Complex, hypothesis based analysis → can we do better?
- Inclusive tagger in development (since many years ...)
- Differences: simulation-data known, but complex
 - Often recovered by multivariate reweighting
 - Alternative approaches to be studied (e.g. Domain Adaptation)
- Work in progress (Fühling, Bunse, Popp)

The LHCb logo, featuring the letters 'LHCb' in a stylized, overlapping font.

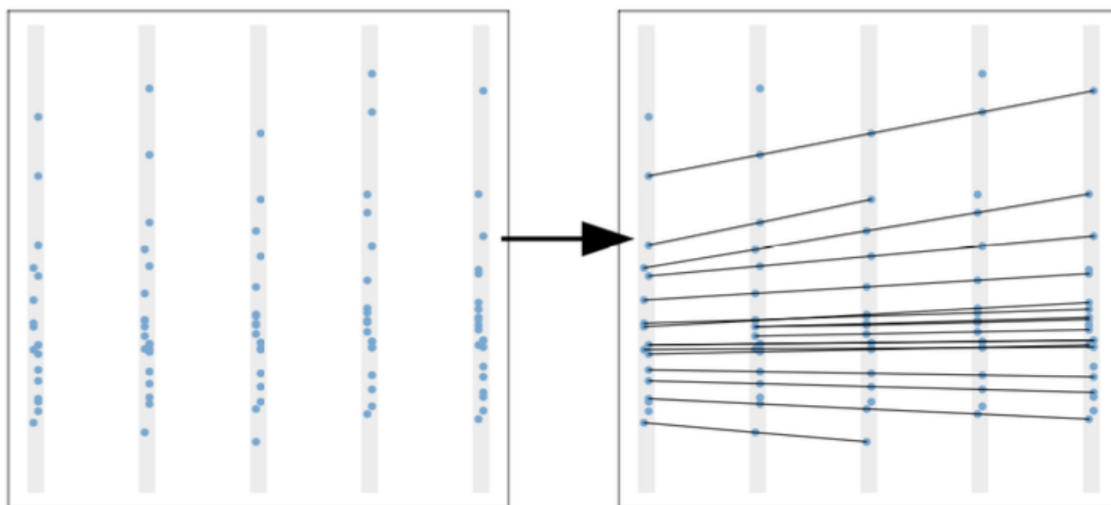
„Event“, rate: 30MHz

Event 58049711
Run 153460
Wed, 03 Jun 2015 12:05:39

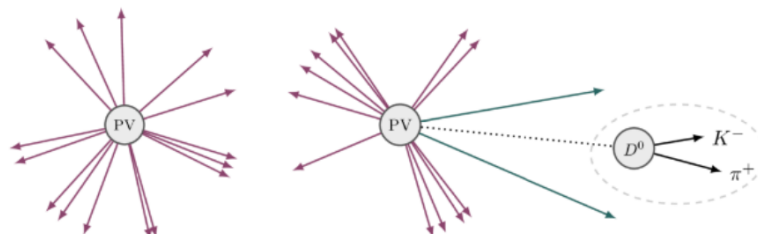
A visualization of a detector event. On the left, a beam of particles is shown as a cone of colored lines (purple, orange, yellow) entering a detector structure. The lines spread out and hit various detector layers, which are represented by vertical bars of different colors (green, blue, orange, cyan). The resulting data points are shown as vertical lines and histograms, illustrating the complex task of identifying particles from a large number of measurement points.

Interesting computational problem:
combine ~10.000 single
measurement points to particle
hypotheses to (2-6) signal
candidates

- Correctly match the hits in detector planes
- p-p collisions are busy with producing many tracks



- We need reconstructed tracks to form particles



Many problems where expert knowledge might be able to help:

- Pattern recognition
- Parameter estimate
- Ring finding
- Anomaly testing
- System architecture, hybrid optimization
(Volume: 400 nvidia A5000 GPUs + O(70k) Intel E5-2630 equivalent physical cores, 30PB disc, 1-2M network cost)
- Fast and efficient candidate selection
~signal fraction up to $1:10^{10}$
- ...



European Research Council
Established by the European Commission

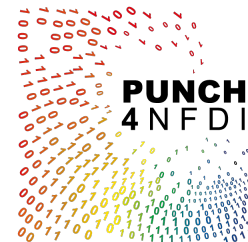


Bundesministerium
für Bildung
und Forschung

SFB 876 Verfügbarkeit von
Information durch Analyse unter
Ressourcenbeschränkung



FSP
Erforschung
von Universum
und Materie



There is still some space left...



