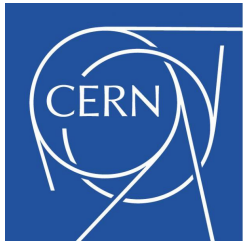


Real-time analyses - the LHCb case

•••

Rosen Matev
on behalf of the LHCb collaboration

Joint WLCG and HSF workshop
26-29 March 2018
Napoli, Italy

Real-time analysis

In this talk “real-time” means online and high-throughput / fast. Deadline for processing is determined by buffer size.

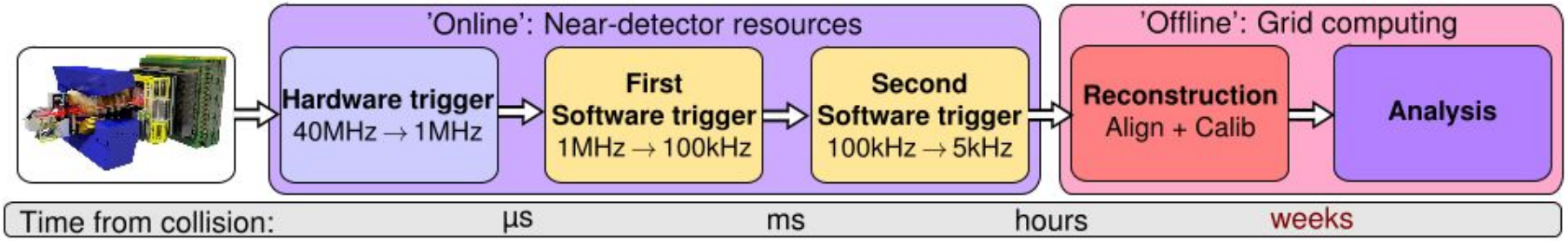
5. **Real-time analysis**: The typical goal of a physics analysis is to combine the products of the reconstruction algorithms (tracks, clusters, jets...) into complex objects (hadrons, gauge bosons, new physics candidates...) which can then be used to infer some fundamental properties of nature (CP asymmetry, lepton universality, Higgs couplings...). We define as real-time analysis any physics analysis step that goes beyond object reconstruction and is performed online within the trigger system, in certain cases using simplified algorithms to fit within the trigger system constraints. Real-time analysis techniques are so far quite experiment-specific. Techniques may include the selection and classification of all objects crucial to the calibration of the detector performance, evaluation of backgrounds, as well as physics searches that are otherwise impossible given limitations of data samples passing the trigger and saved for offline work.

HSF-CWP-2017-16

Why online analysis?

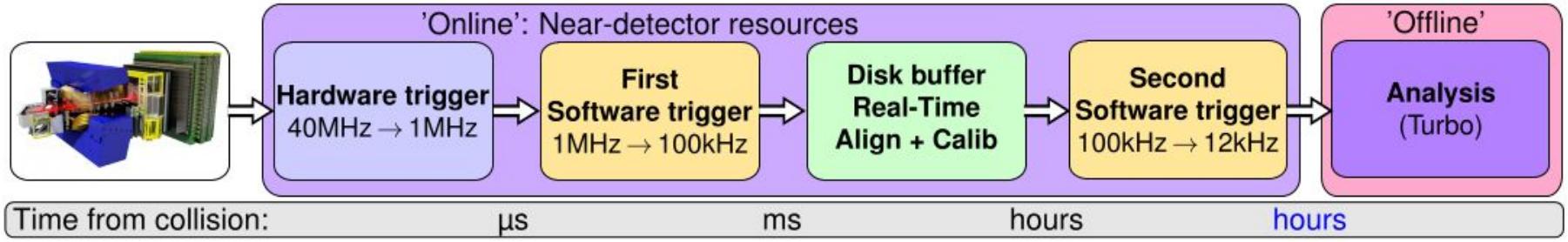
- Signal rates are already significant
  - LHCb operates at L = 4×10^{32} $cm^{-2}s^{-1}$ in Run II
  - 45 kHz of bb, ~1 MHz of cc
- Offline storage is expensive
- A good trigger throws away more background than signal. What's better than the final analysis?
- Do as much as possible online:
  - Offline-quality reconstruction
  - Offline-quality selections

LHCb data flow in Run I



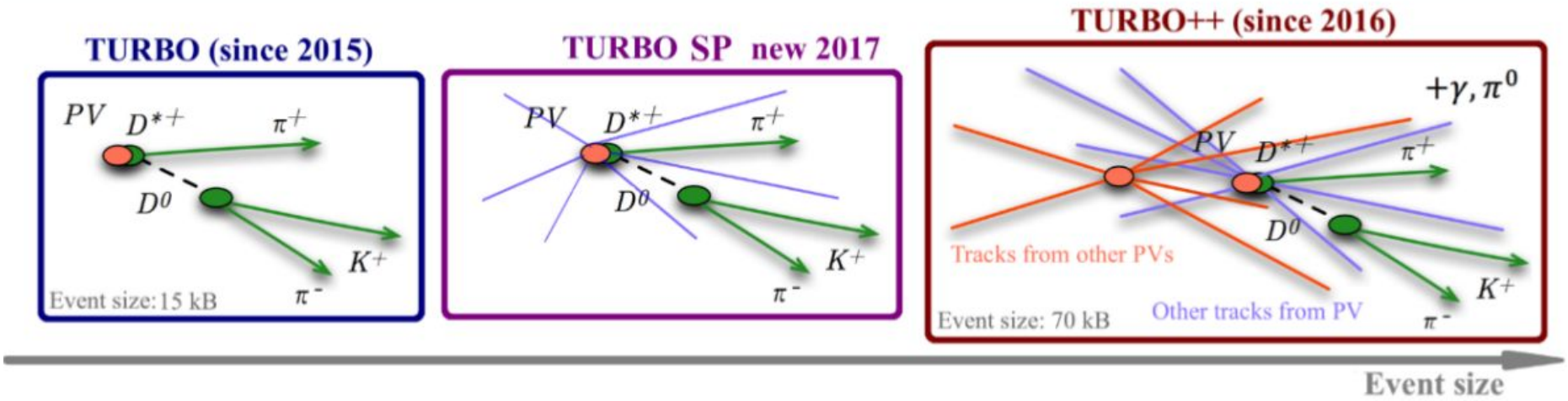
- Main B physics trigger: inclusive, topology-based MVA
- O(1000) exclusive selections made offline
- Does not scale for charm: reduce rate in another way

LHCb data flow in Run II



- HLT2 is now asynchronous, uses aligned and calibrated detector
- Reconstructed objects in HLT identical to those produced offline
- Selections of arbitrary complexity on the entire event possible
- Exclusive online selections now match what was previously only possible offline
- Can we do more with an offline storage capacity of 700 MB/s?

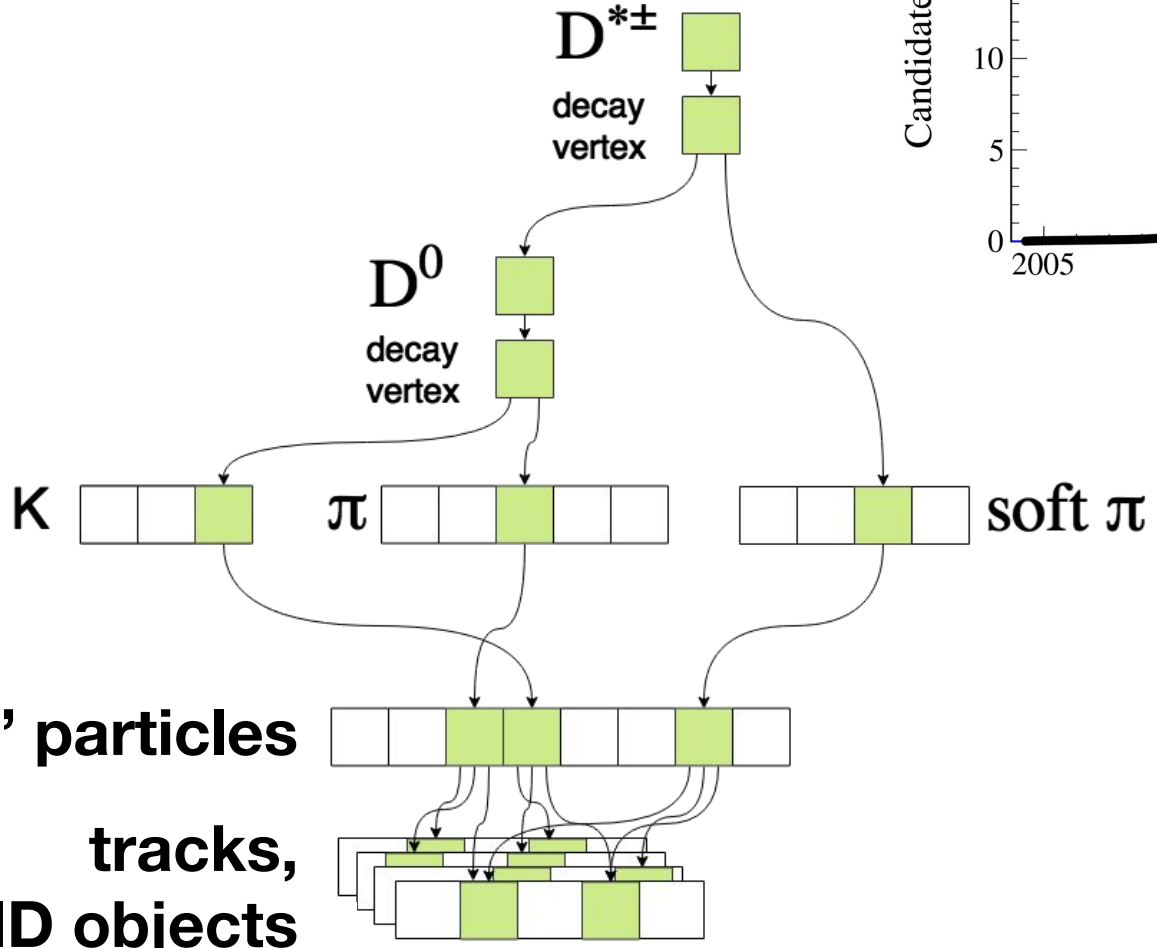
Turbo

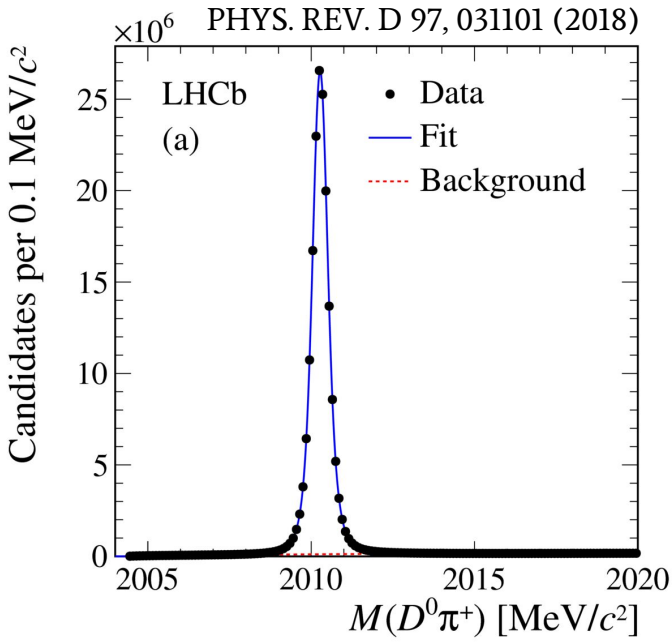


- Turbo is the LHCb paradigm for reduced event format data
- High degree of flexibility: Save only as much of the event as is needed
  - Keep all reconstructed objects, drop the raw event: 70kB
  - Keep only objects used to trigger: 15kB
  - 'Selective Persistence' anything in between

Turbo illustrated

Turbo

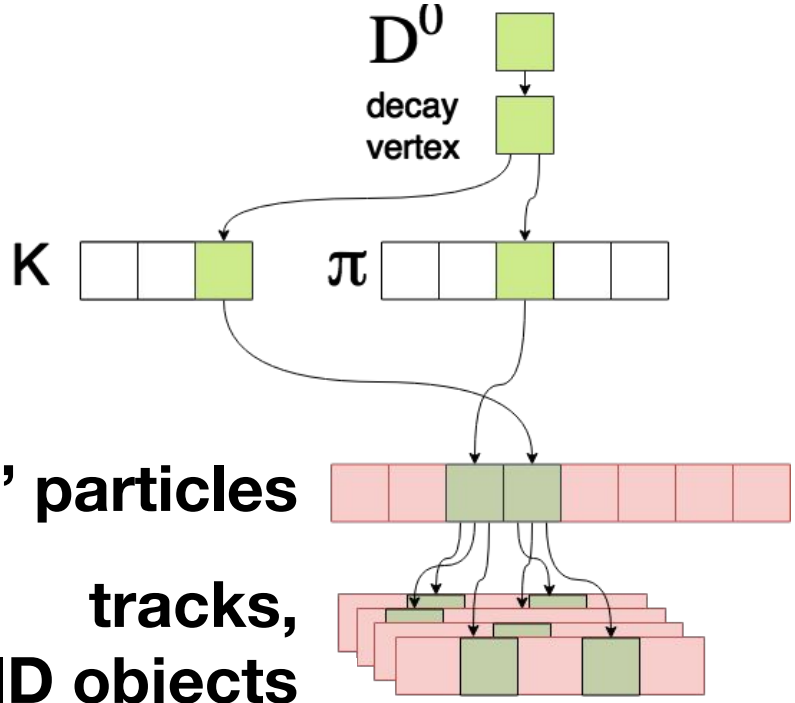




Turbo illustrated

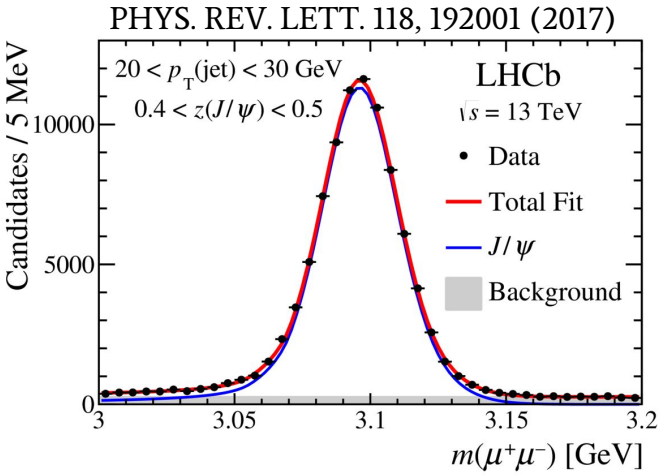
Turbo

Turbo++



“proto” particles

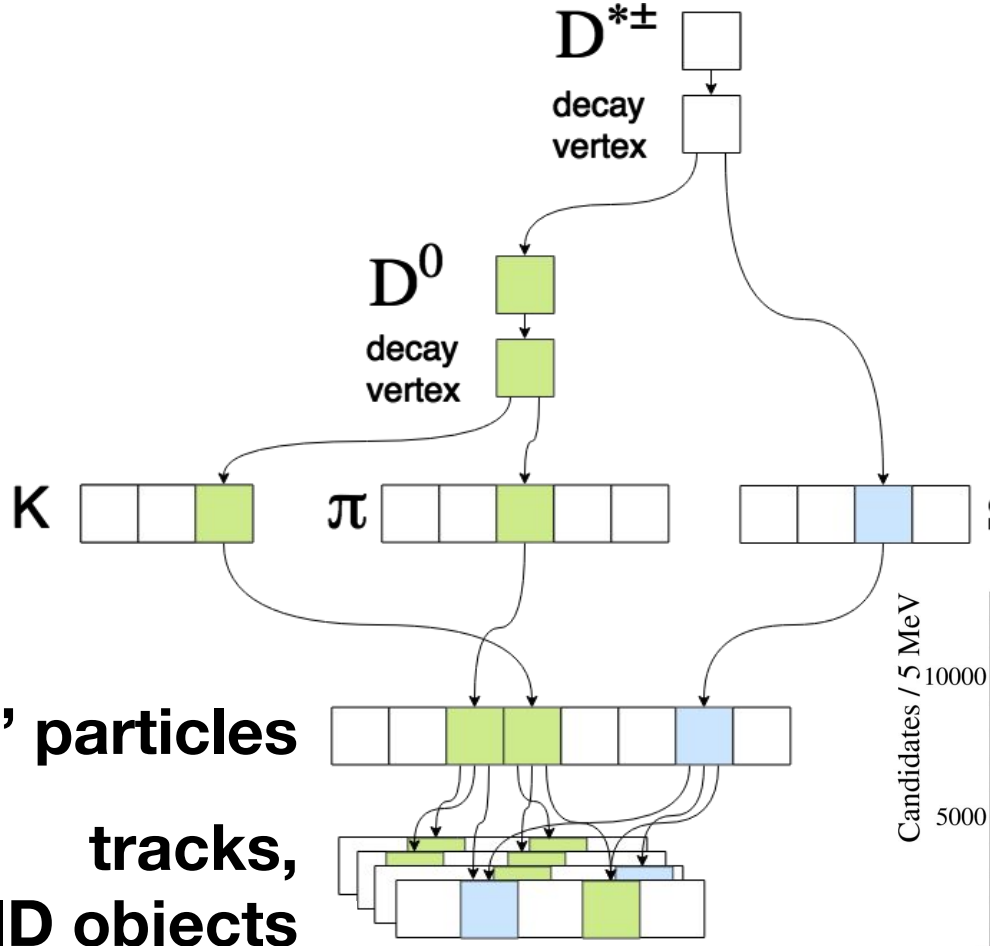
tracks,
PID objects



Turbo illustrated

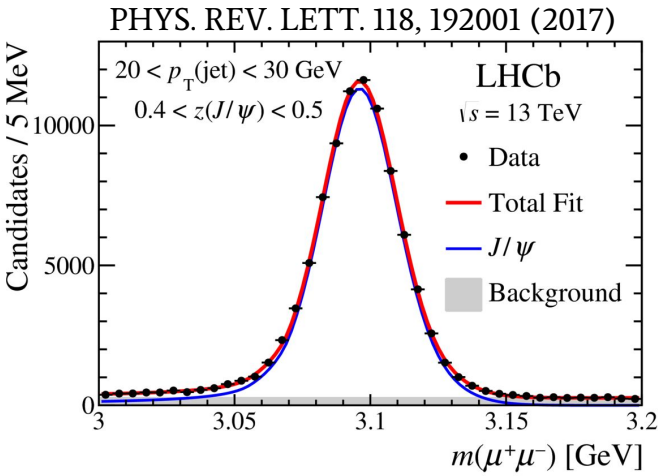
Turbo

TurboSP



“proto” particles

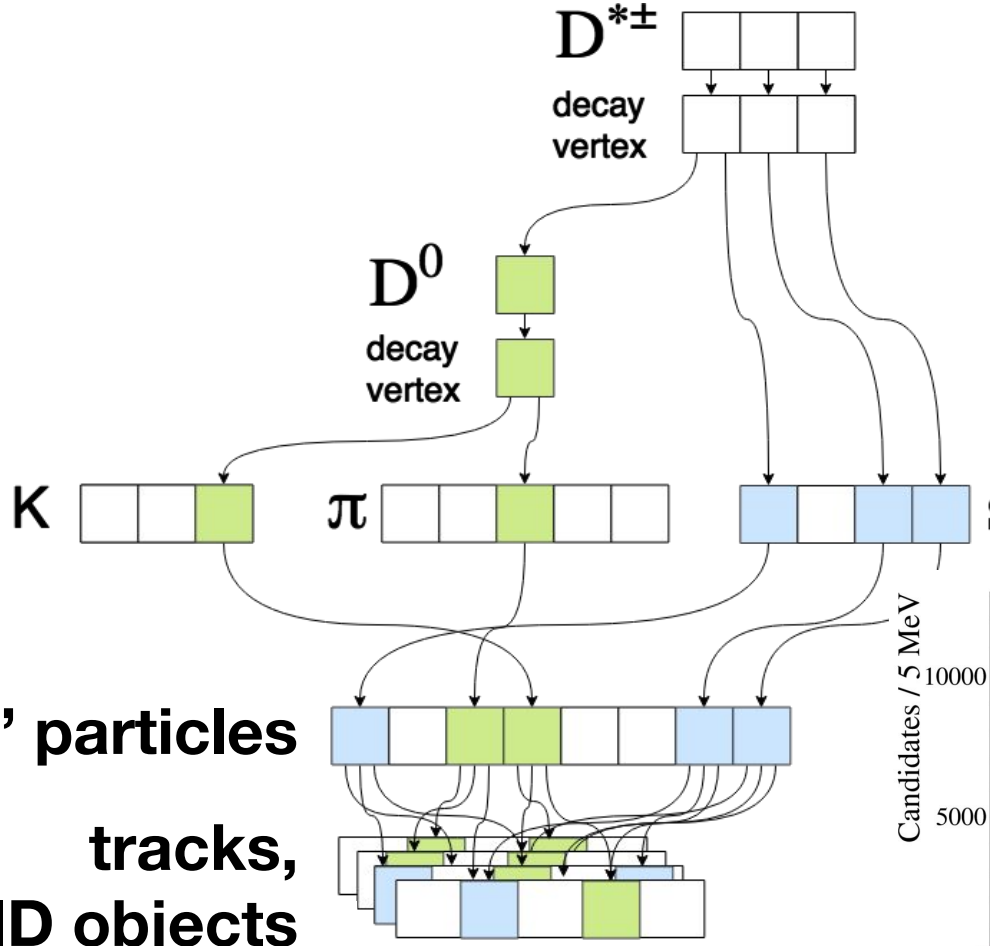
tracks,
PID objects



Turbo illustrated

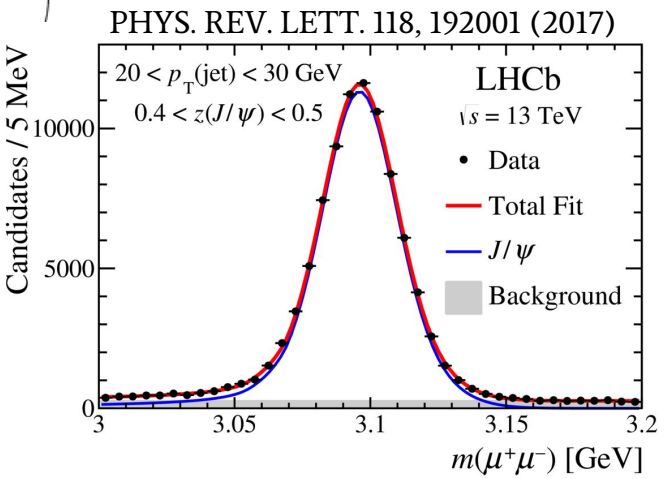
Turbo

TurboSP



“proto” particles

tracks,
PID objects

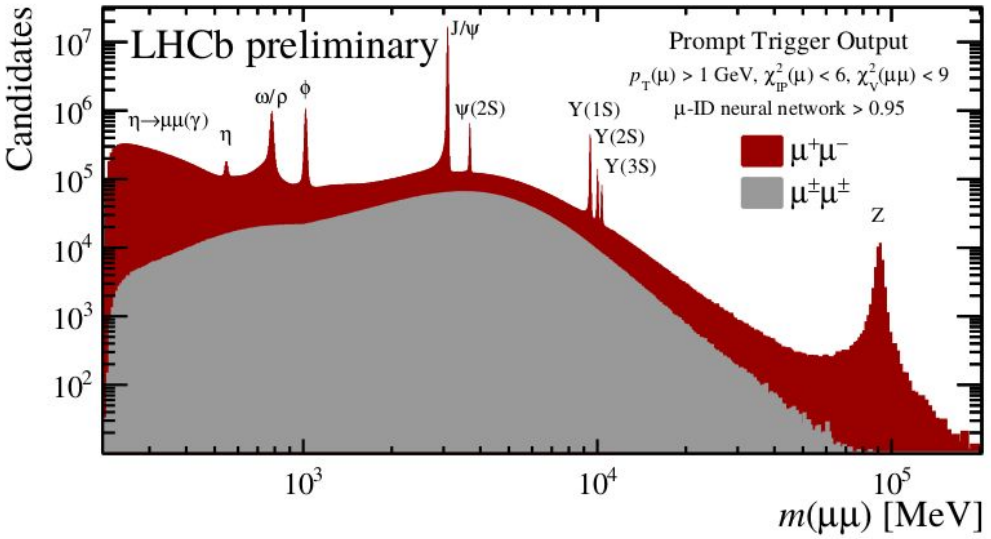


Turbo in Run II

- In 2017 528 trigger lines at HLT2, half are Turbo
- 25% of the trigger rate is Turbo but it counts for only 10% of the bandwidth
- 20% effective increase in rate (for fixed bandwidth)
- Many analyses would not be possible without Turbo

CERN-EP-2017-248
LHCb-PAPER-2017-038
October 5, 2017

Search for dark photons produced in 13 TeV *pp* collisions

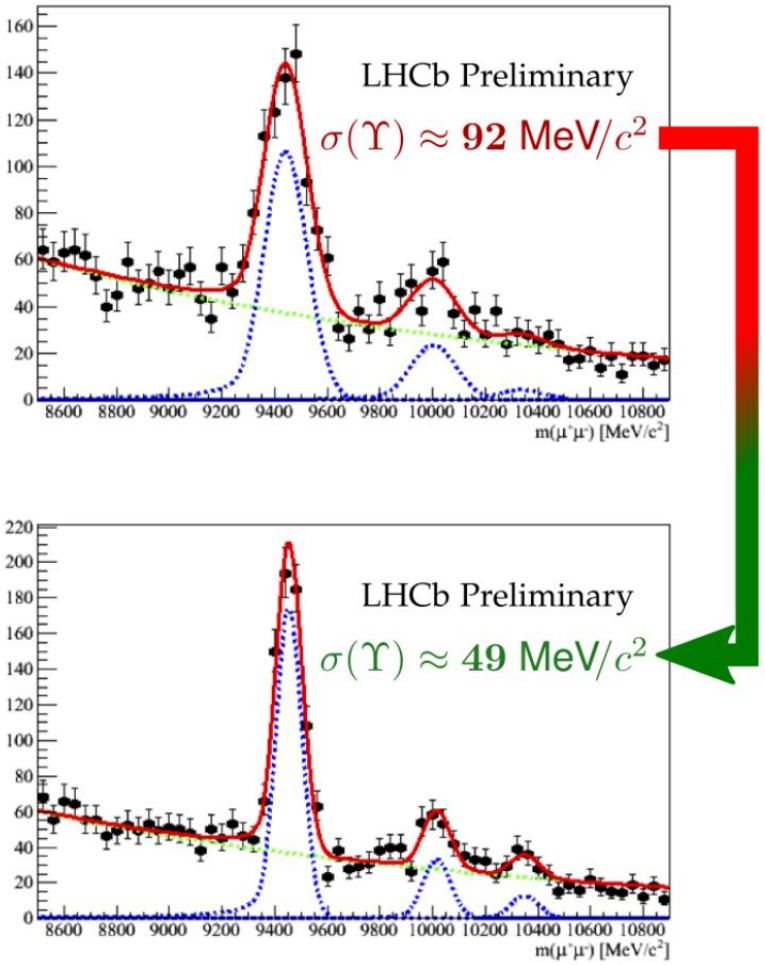


But there's no turning back...

- Throwing away most of the event means care must be taken
  - Integration testing and robust procedures are critical
- Turbo relies on never needing to reprocess
  - Online calibration and alignment and
  - Online data quality monitoring are even more important
- ...but wait these are also real-time analyses!

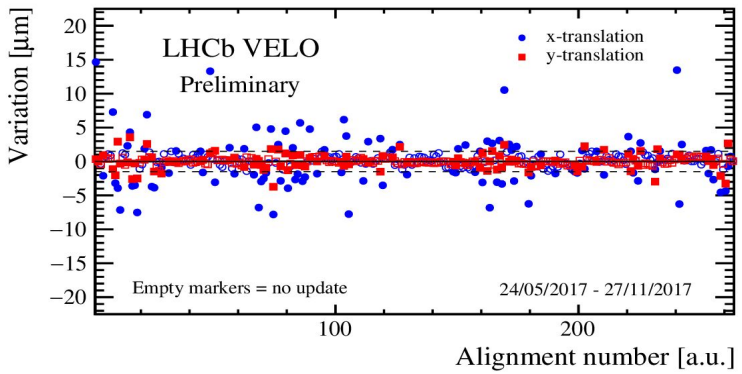
Real-time alignment and calibration

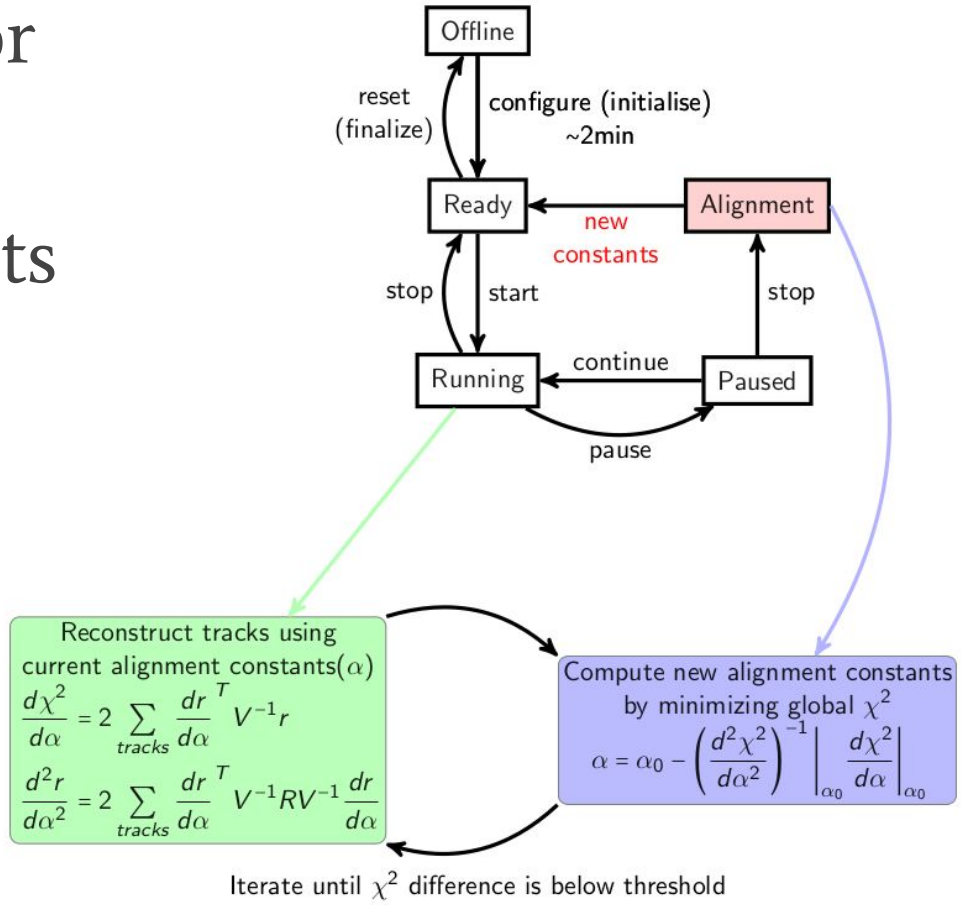
- Why push calibrations to online?
- Efficient & pure output requires full reconstruction at HLT2
  - Online selections == offline selections
  - Alignment and calibration of full detector needed
- While HLT1 is written to disk, alignment & calibration tasks run



General alignment procedure

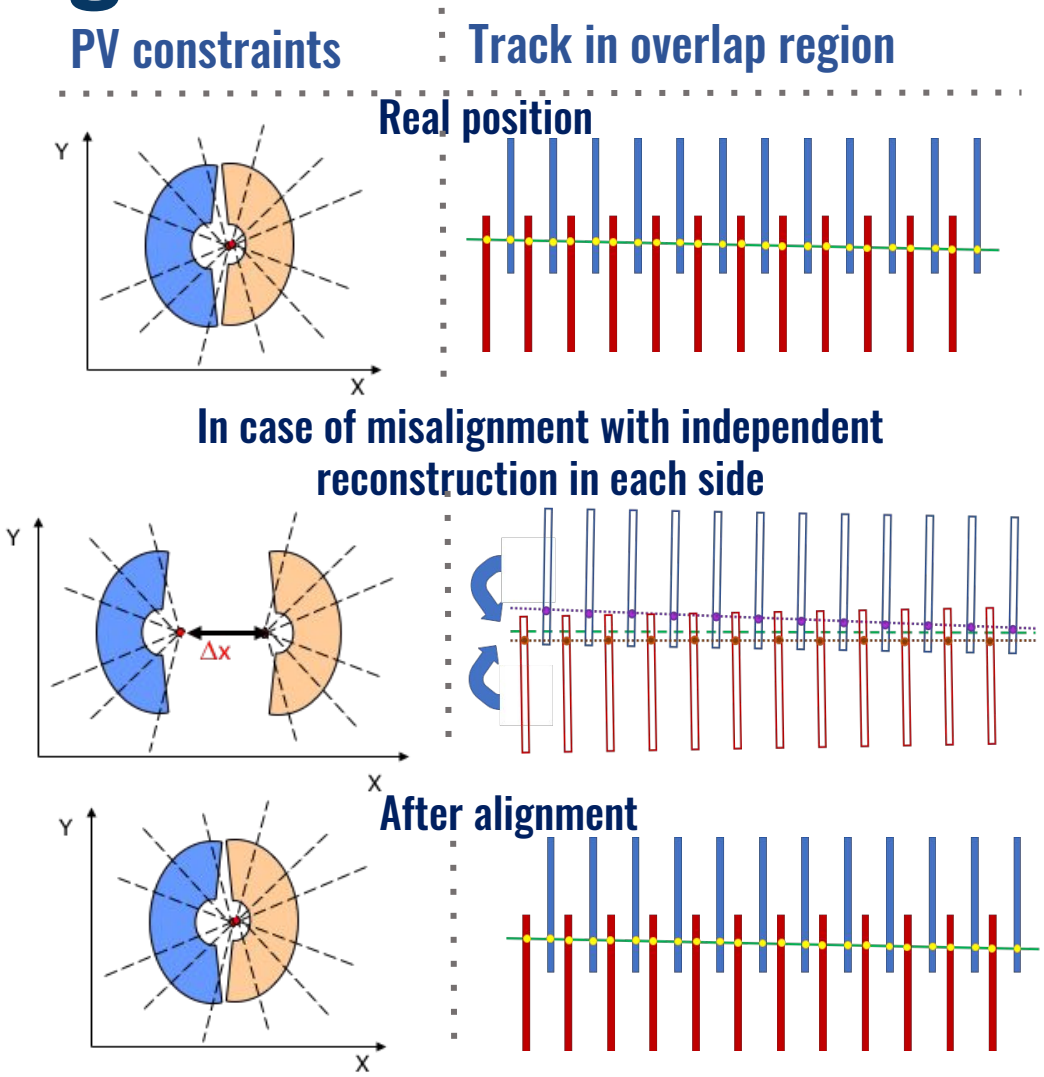
- Runs automatically for each fill
- Provides new constants as soon as possible
- Fully automated





Example: VELO alignment

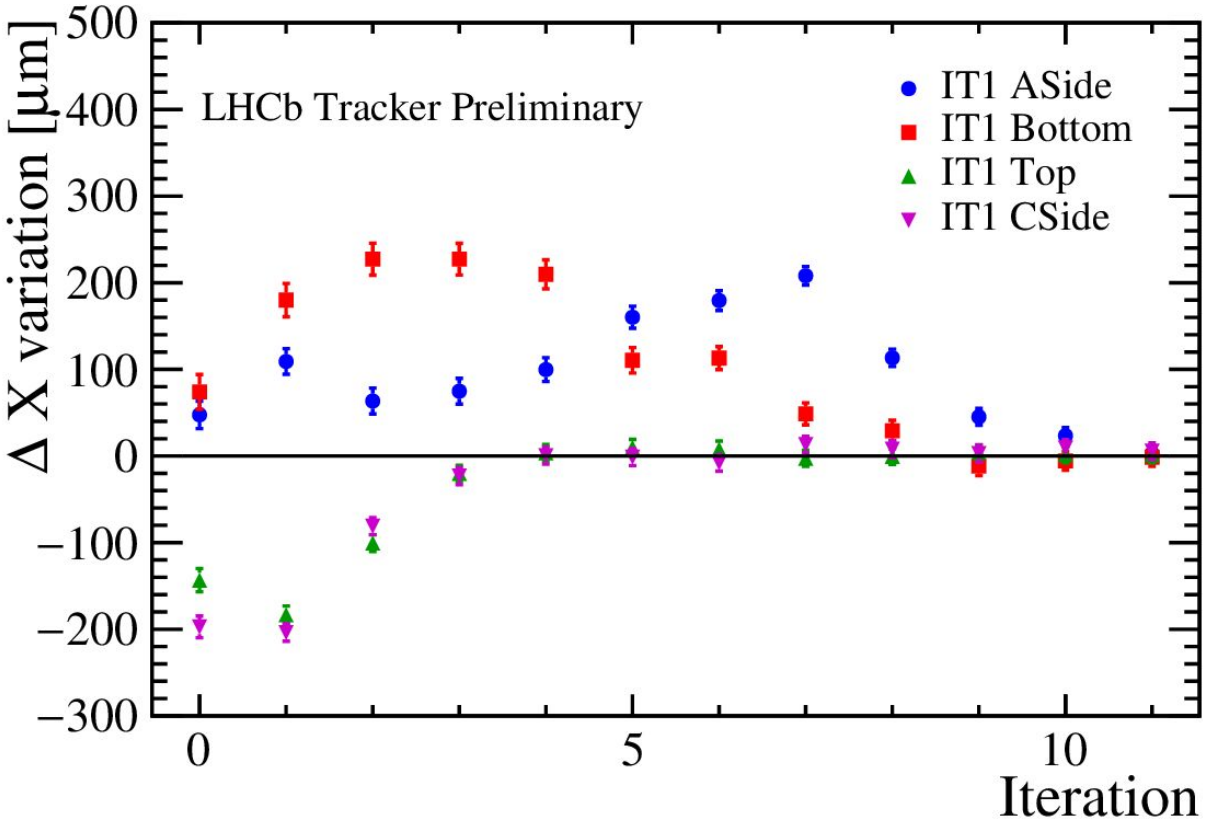
- VELO centred around the beam for each fill
  - Resolver X, Y position accuracy of 10 μm
- Kalman filter based method, minimizing the track hit residuals with PV constraints
- Automatic alignment of VELO halves in less than 10 minutes



Online data quality monitoring

- LHCb DQM software (Monet) presents per-run detector performance histograms to DQ shifter
- Previously a C++/ROOT application
- Now web-based
  - much easier remote use
  - fast deployments, easier support
- Reusable and extensible for other monitoring tasks such as verification of simulation software or verification of data selections changes

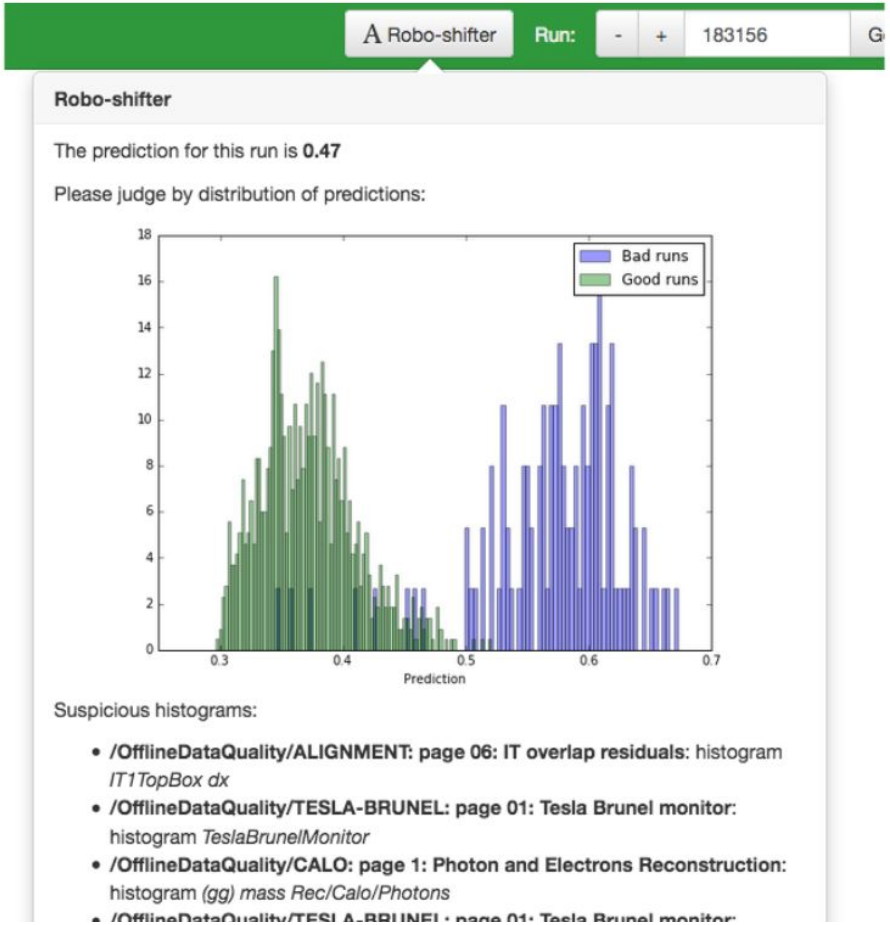
Online data quality monitoring



- Automatic actions affecting recorded data need monitoring
- Example: first tracker alignment convergence in Run II

Online DQ analysis example

- Robo-shifter is a ML based expert system designed to assists the DQ shifter
- Given run data it can predict probability of run being good or bad
- Provides potential problem sources extracted from decision trees
- Integrated in DQM



Summary and outlook

- Tremendously more data for LHCb in Run III
  - Drop events and event information as early as possible
- No turning back: analysis must be “real-time”
  - Align and calibrate the full detector
  - Produce complex objects and filter on them online
  - Monitor data quality on high-level physics quantities
- Push online analysis even further
  - Optimize Turbo event size (e.g. better format, compression)
  - Expand monitoring to full-blown analyses
  - Explore aggregating events in the trigger?