Persistency Issues for LHCb Run III

- Basic numbers and requirements
- The new paradigm
- Implications of the computing model
- What is getting onto us

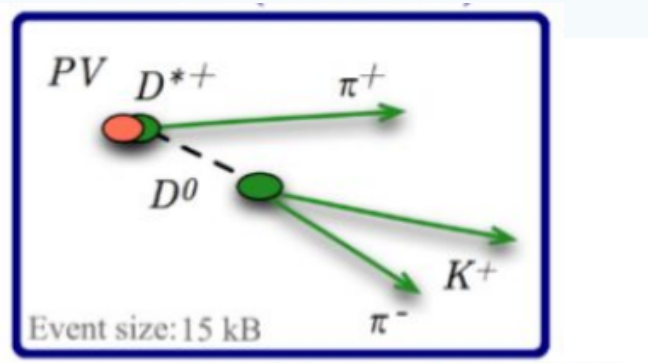
Markus Frank PH/LBC

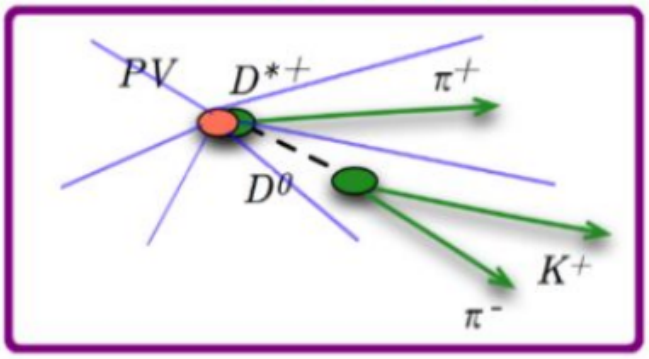
Scope

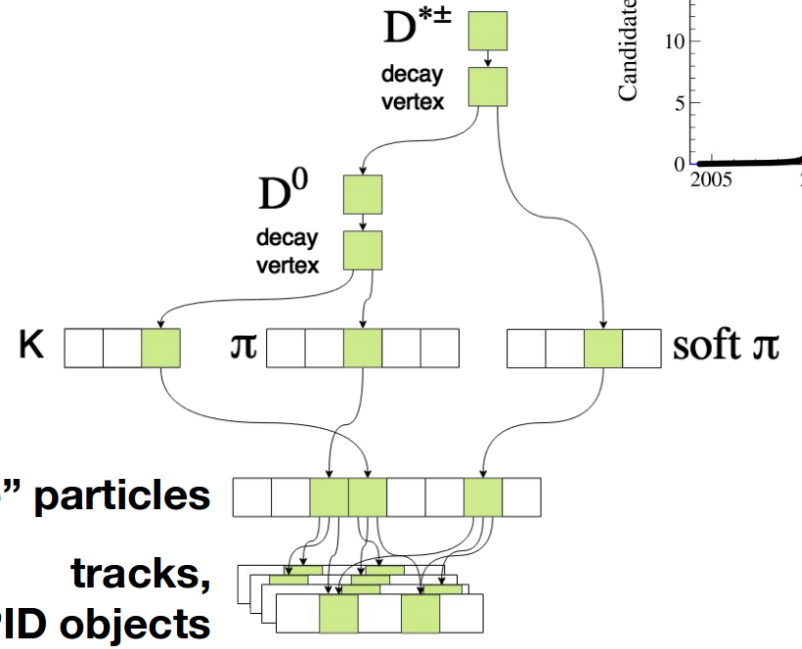
- Run I and Run II data are (nearly) history
  - Things worked out not too badly
    - Reconstruction and analysis I/O works based on ROOT trees
- On the agenda now:
  Upgrade, upgrade, upgrade: Run III
  - All numbers to be taken with a grain of salt
  - Dependencies on changes to event model etc. unknown
  - This is in three years from now
  - Situation similar to 2005: Exact facts notknown
    - Will still try to state the problems we shall likely face
    - LHCb faces a different situation
      Small events, many streams etc.
      problems probably even getting emphasized

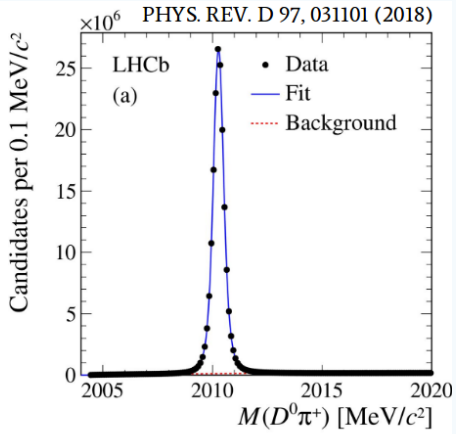
Data Format: Dramatic Changes Ahead

- **There will no longer be raw data: game changer**
  - **High level objects - no offline reconstruction possible**









Online Data I/O

- At 40 MHz we cannot afford to save raw data
- TESLA output format (3/4 of data volume [?])
  - Strategy: Preselect useful primary vertices and secondaries, throw away anything else
    - ~10 topological streams depending on physics content
    - MDF sequential files
    - Specialized data packing
  - Today: average 30 kB / events [spread: 15-80 kB]
    Expected: similar size and spread
  - Data rate: 5-10 GB/s signal events
  - 50 PB / year assuming 5 x 10^6 seconds collisions per year

Offline Data I/O

- Starting from 50 PB online data in 10 streams
  - Further preselections depending on physics
  - ~100 offline streams for physics analysis
    - 500 TB per stream with smallish overlaps
    - Up to 100 % data retention => refinement depending on physics
    - ROOT format
- Should the full data volume be availible at all time?
  - Idea: Keep 20 % ie. 100 TB per stream on disk
  - Only open access for 'final' analyses
- 100 PB total data volume per year
    - 50 PB from online + 100 x 500 TB

Event Model Dependencies

- Direction not clear
  - SoA or AoS
  - In split mode roughly the same at file level
  - CPU wise of course SoA is much simpler for ROOT
- Has clearly an effect on the analysis model
  - … but not too much known at the moment

Offline Streams and Analysis Model

- 500 TB / stream / year
- As 15 years ago. Only different scale:
  - Group productions for mini-, micro-DST, N-tuples
  - O(5kB), O(10TB) O(2 x 10^9 events)
  - Depends on analysis needs
  - Requires sparse reading of data O(< few %)
  - 1 … 2 refinement cycles per quarter

Offline Streams and Analysis Model

- Expect same problems as for stripping
  - 10 … 20 simultaneous output streams
  - Memory explosion for splitting
  - Any I/O buffer gets multiplied by 10...20
  - In the past this led to absolutely contrary optimizations
    - NO splitting: object I/O
    - Small buffers, relatively often flushed to disk
    - Could not take any advantage of work done by the other LHC experiments

Conclusions

- Showed the roadmap for the LHCb Run III data usage
- Facts and numbers are far from fixed
- ROOT event data I/O is an integral part of any analysis activity
  - Streams, mini, micro-DST, N-tuples
- Problems from Run I/II likely to not have vanished
  - Memory usage *is* an issue for LHCb ROOT I/O