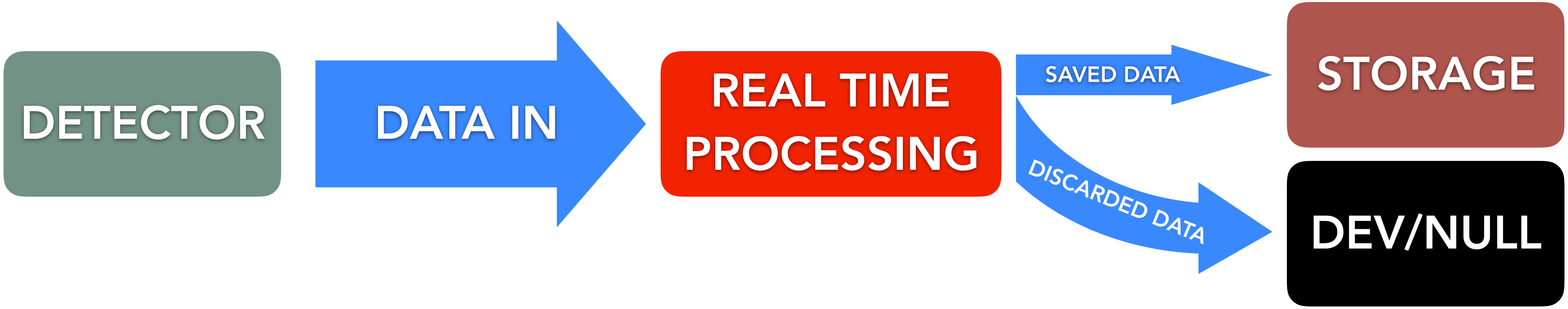
Conceptualization, implementation, and commissioning of real-time analysis in the High Level Trigger of the LHCb experiment

Vladimir V. Gligorov, CNRS/LPNHE

Habilitation a diriger les recherches, May 14th 2018

Eastbound and down : an introduction to real-time analysis

Q : What is real-time?



A : Any processing of data before it is permanently recorded

Why do we need to process data before recording it?

	LHCb	CMS/ATLAS
Data volume at detector	~30 Eb/year	~1 Zb/year

Why do we need to process data before recording it?

Data volume at detector

LHCb

~30 Eb/ year

Global internet dataflow 2015

~640 Eb/ year

CMS/ATLAS

~1 Zb/ year

Because HEP detectors produce too much data to store

Data volumes @ LHC after real-time processing

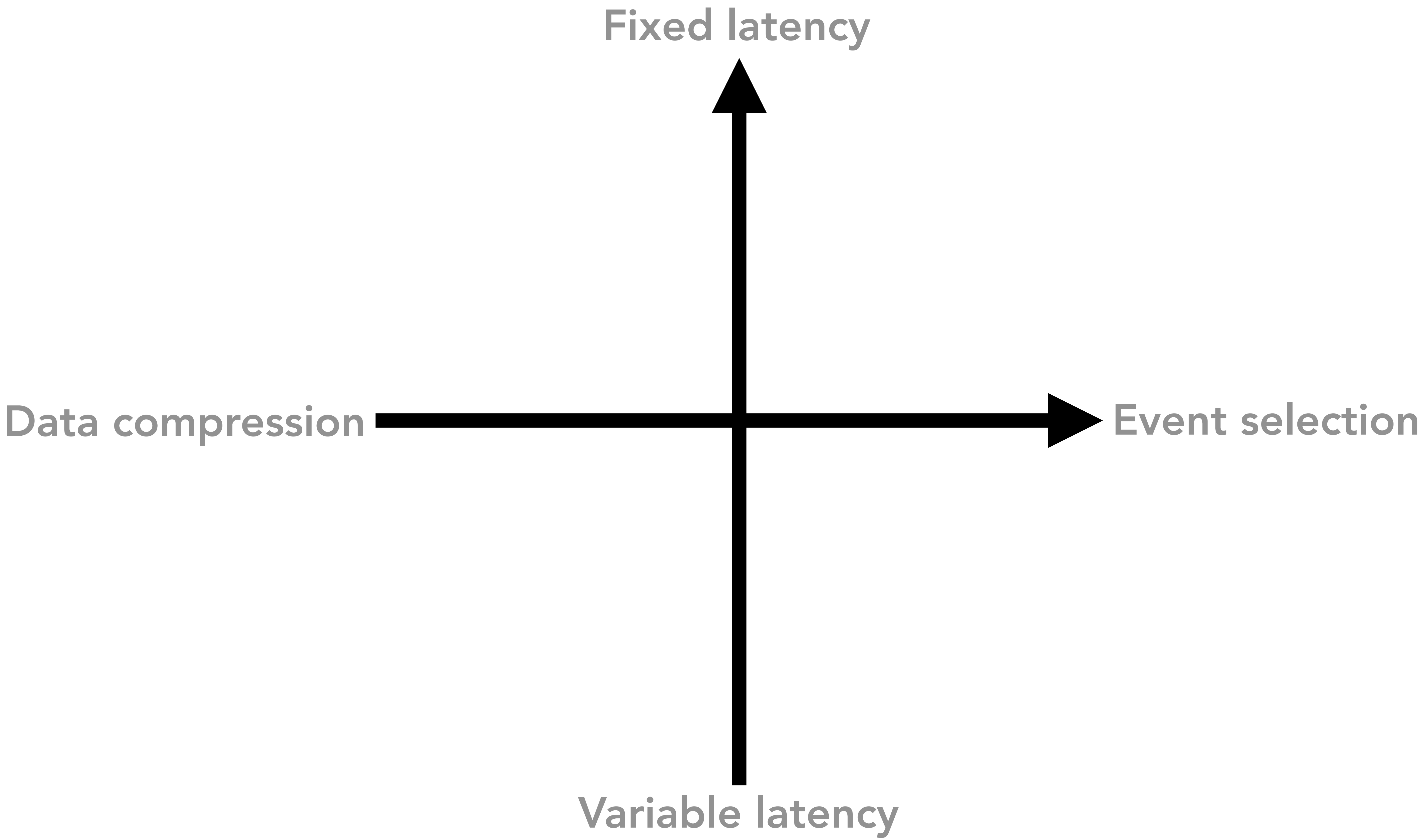
	LHCb	CMS/ATLAS
Data volume at detector	~30 Eb/ year	~1 Zb/ year
Data volume for analysts	~20 Pb/ year	~40 Pb/ year

Global internet dataflow 2015

~640 Eb/ year

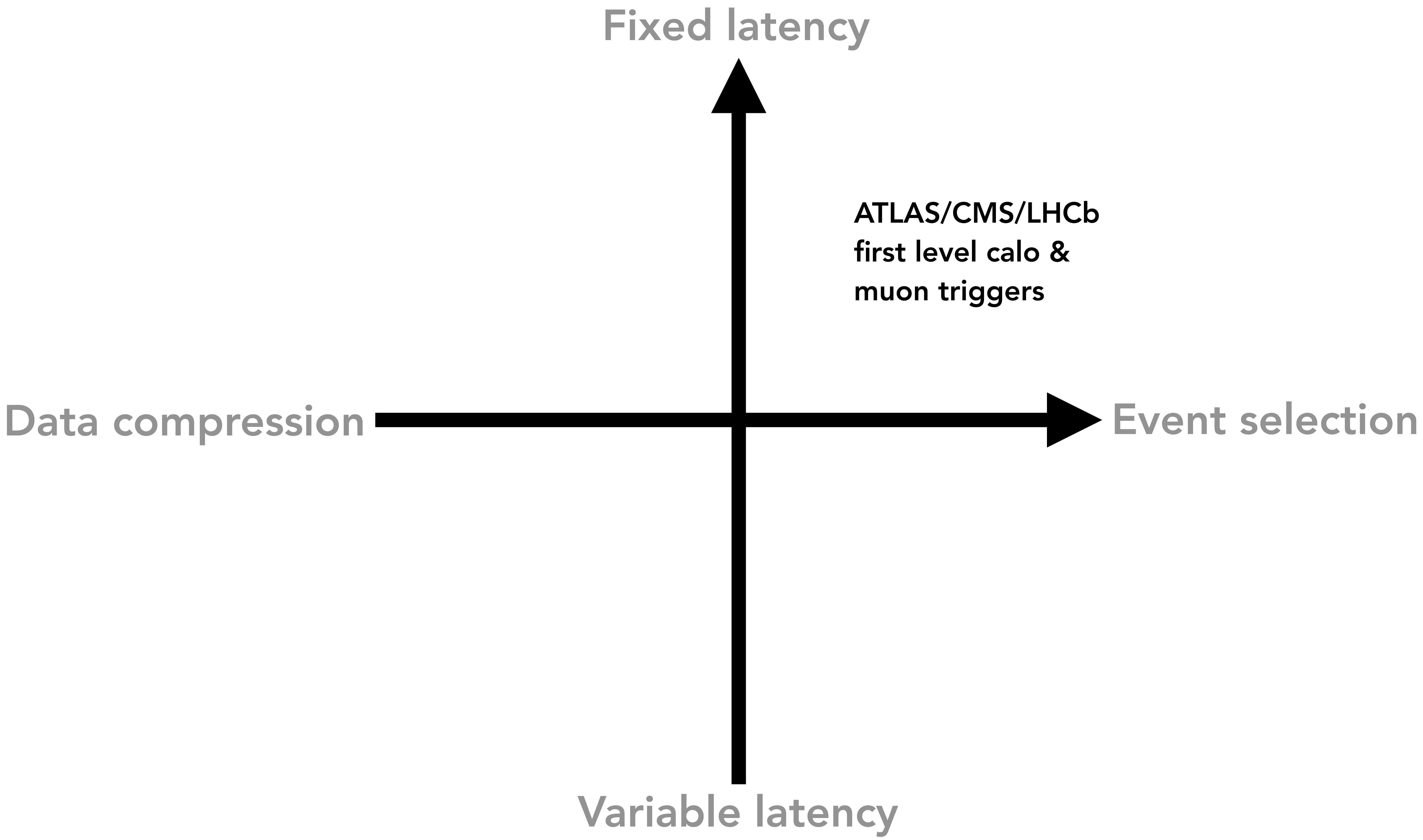
Real-time processing reduces data by 3-5 orders of magnitude

What kinds of real-time data processings exist?



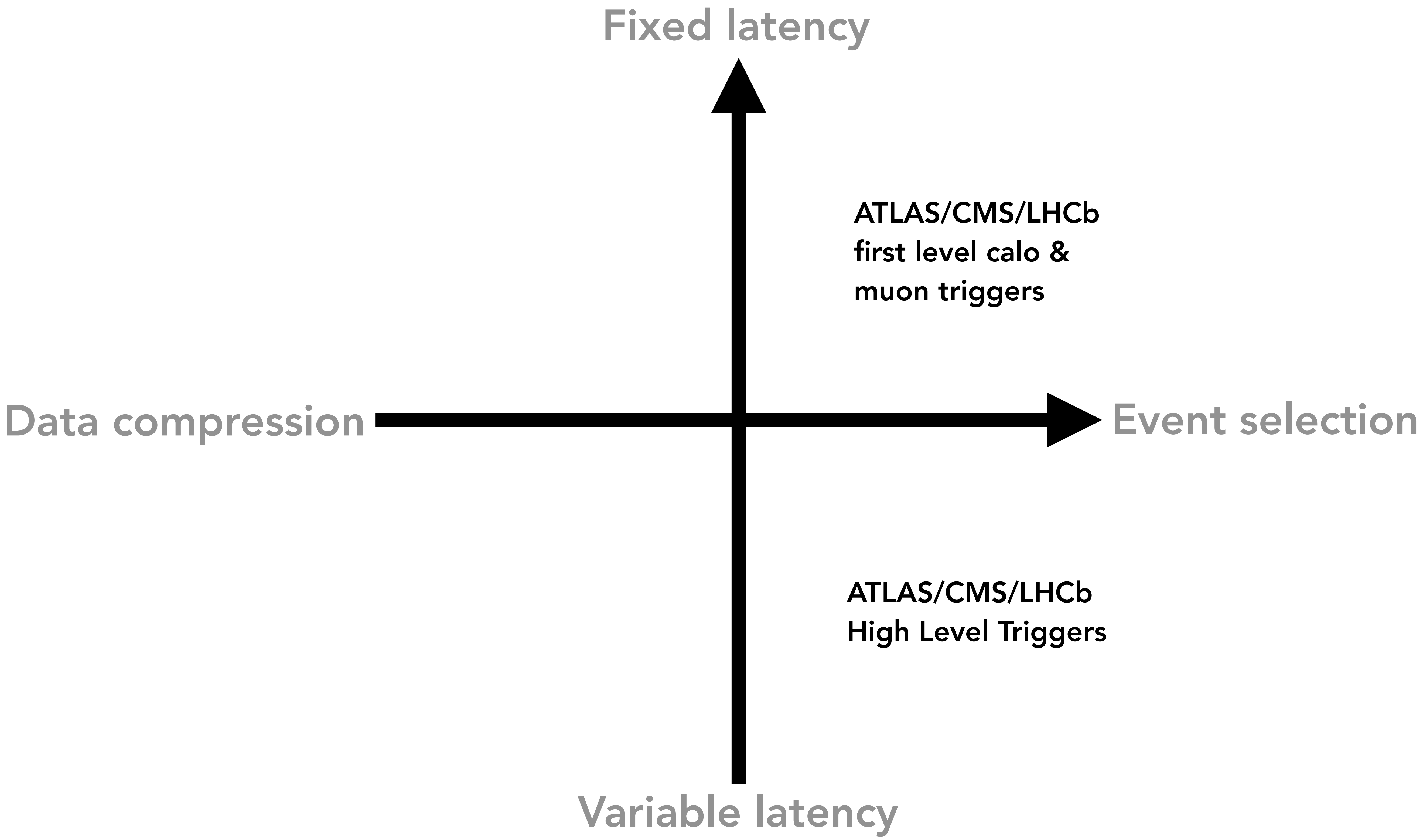
Distinguish fixed & variable latency, selection & compression

What kinds of real-time data processings exist?



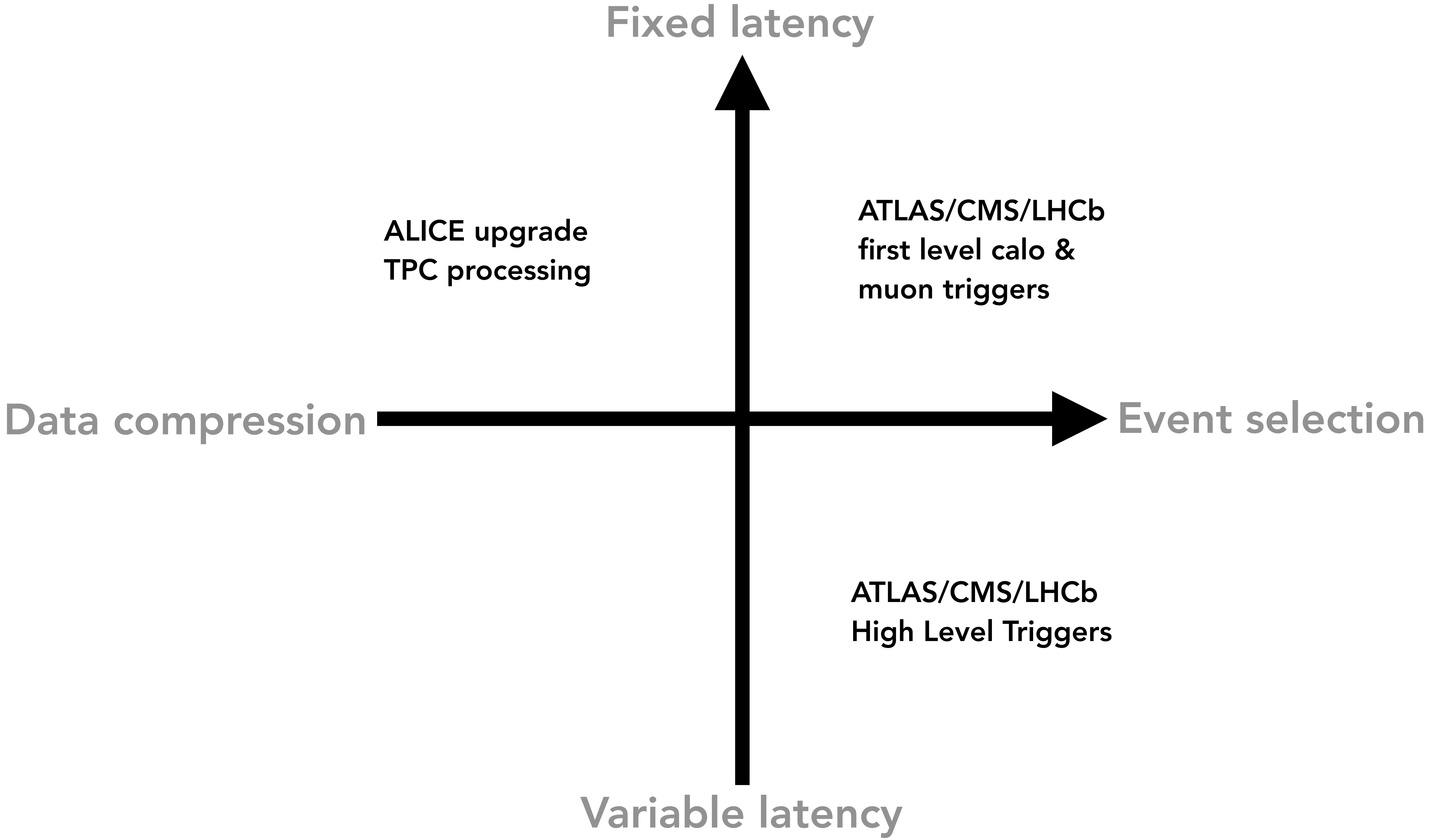
Distinguish fixed & variable latency, selection & compression

What kinds of real-time data processings exist?



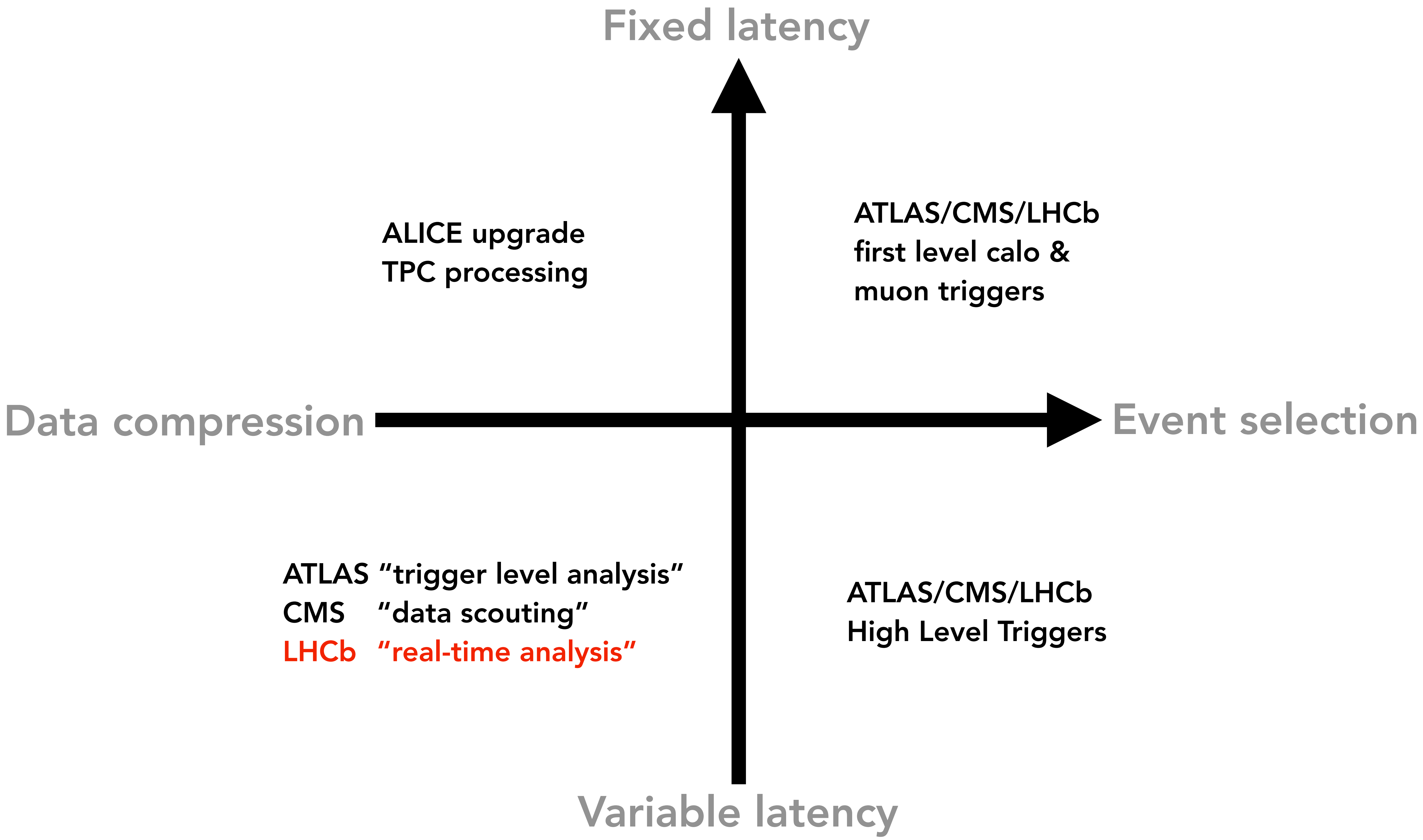
Distinguish fixed & variable latency, selection & compression

What kinds of real-time data processings exist?



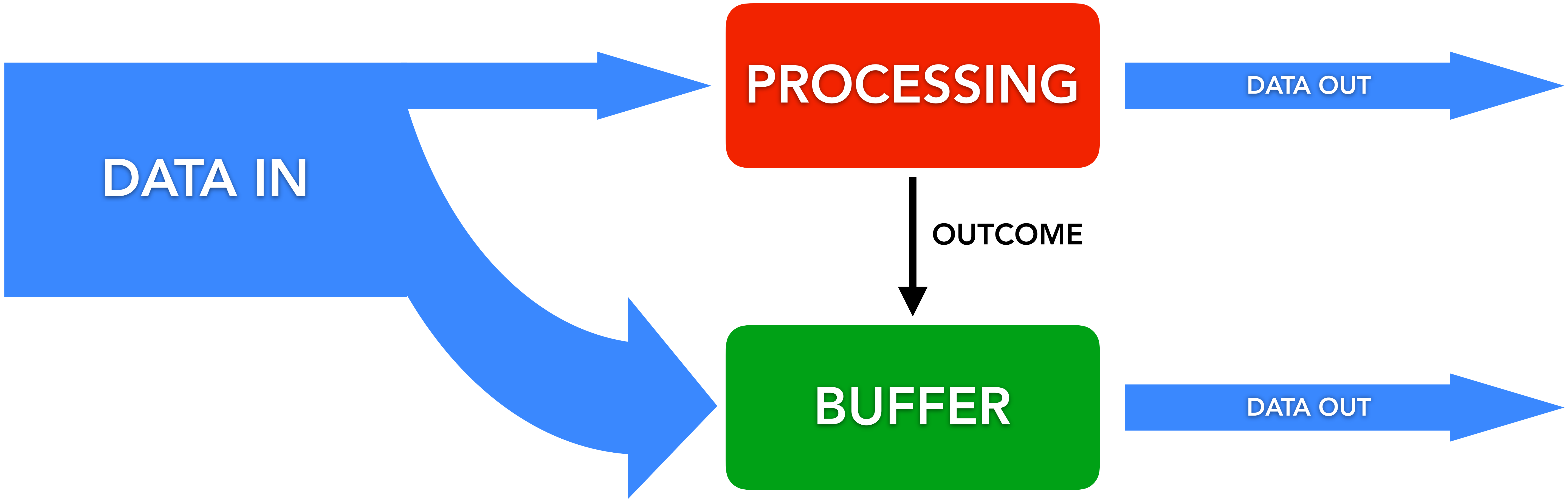
Distinguish fixed & variable latency, selection & compression

What kinds of real-time data processings exist?



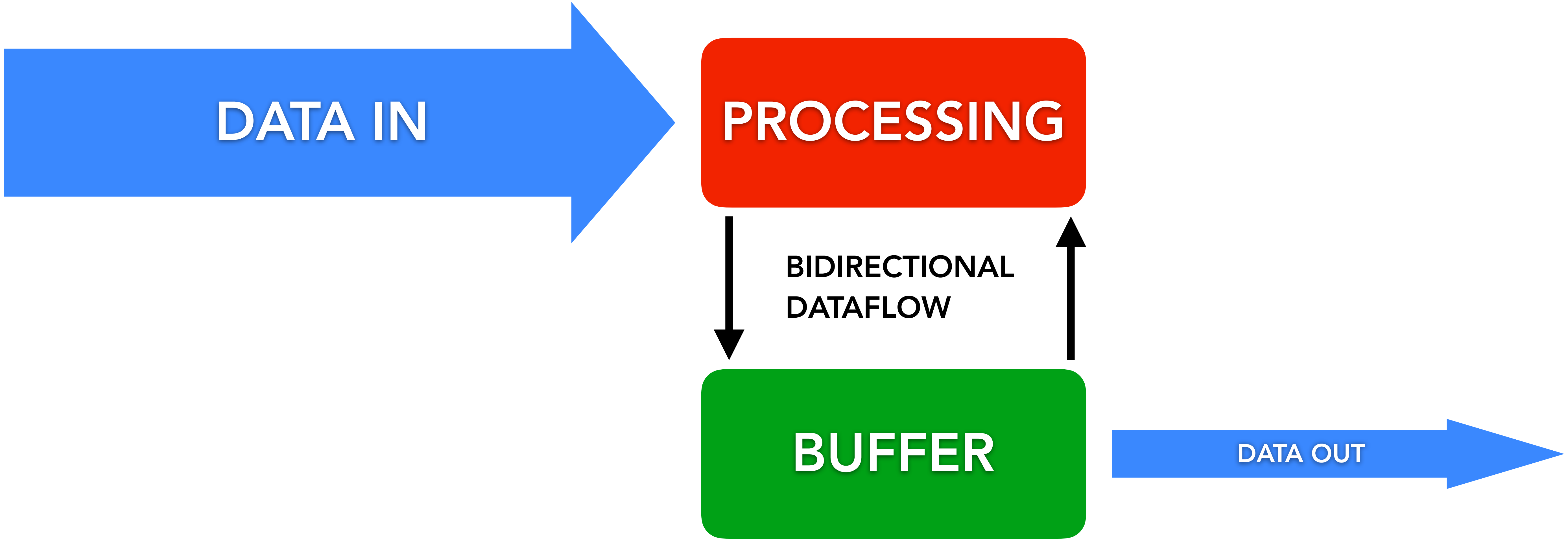
Distinguish fixed & variable latency, selection & compression

Fixed latency processing



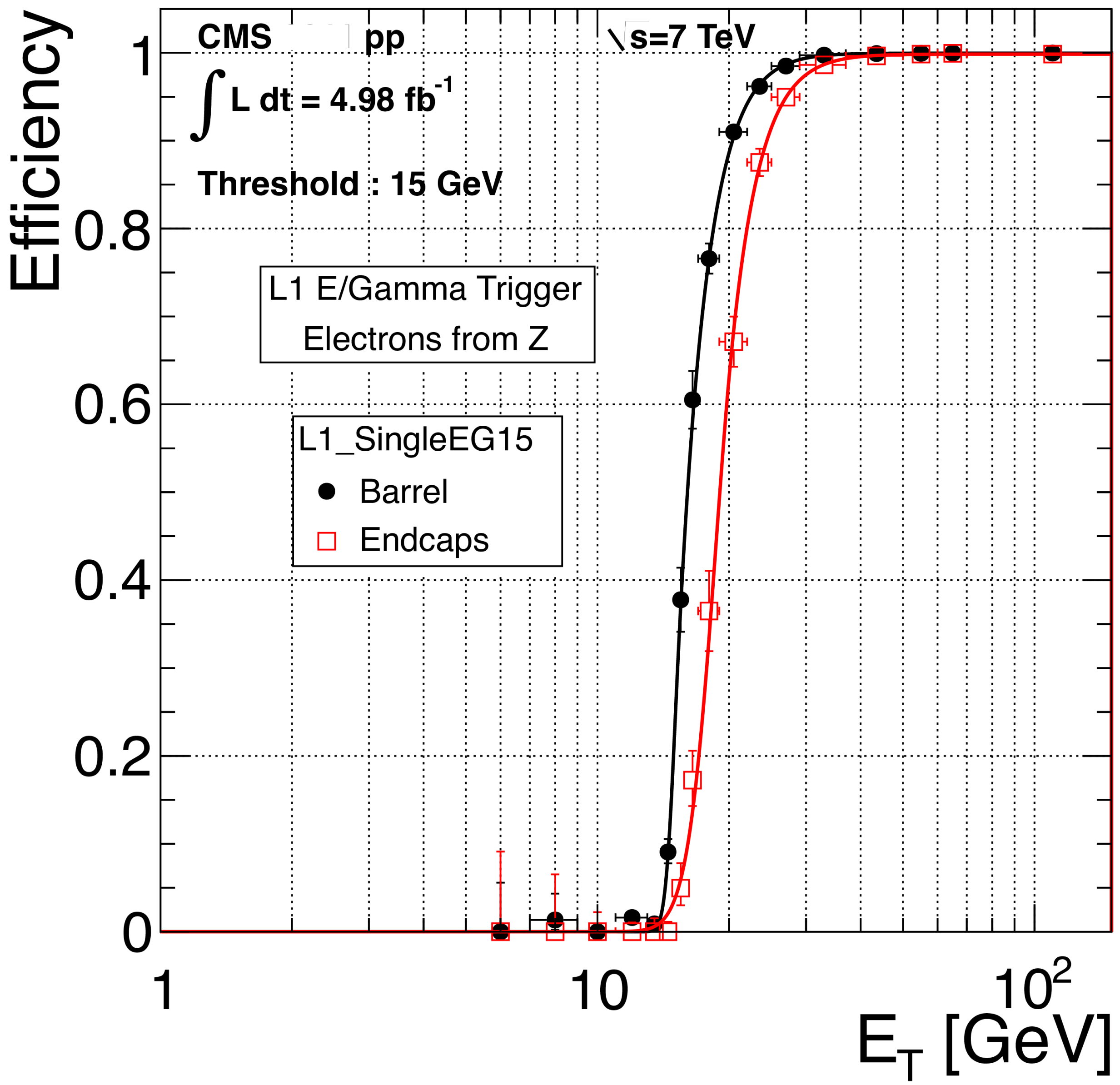
Typically used when processing controls detector readout

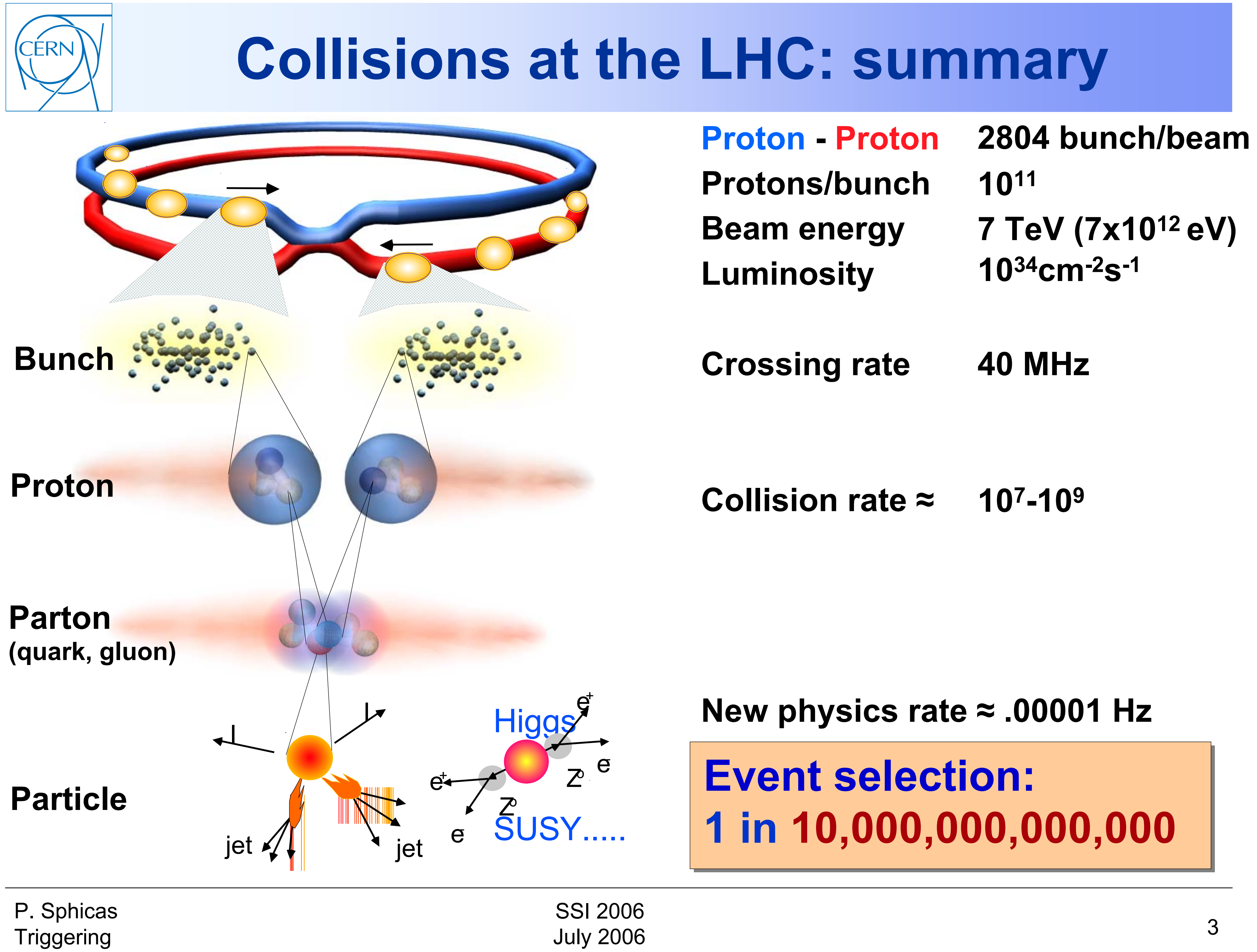
Variable latency processing



Typically used when data has already been read out

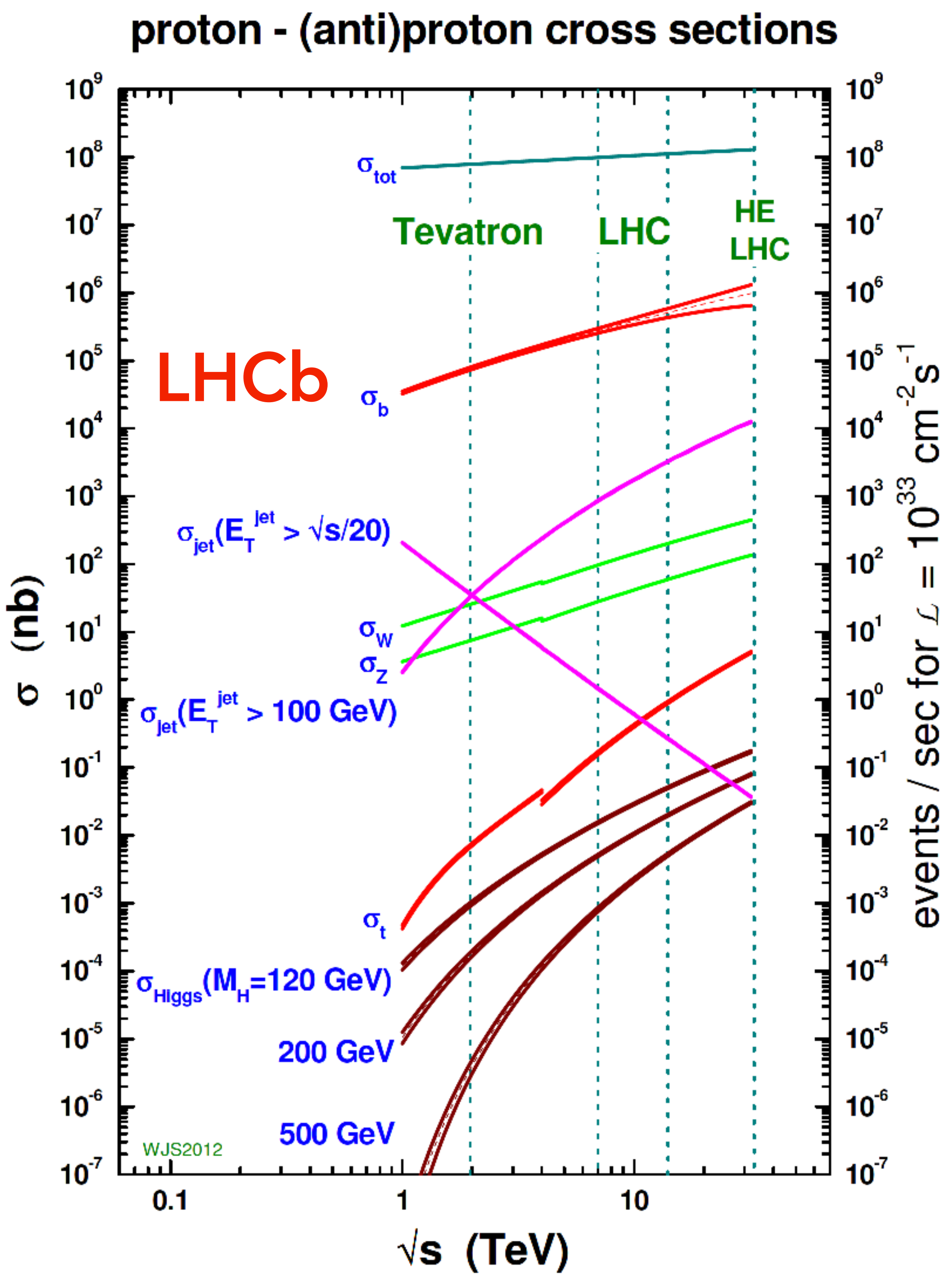
Traditional real-time processing, or "triggering"



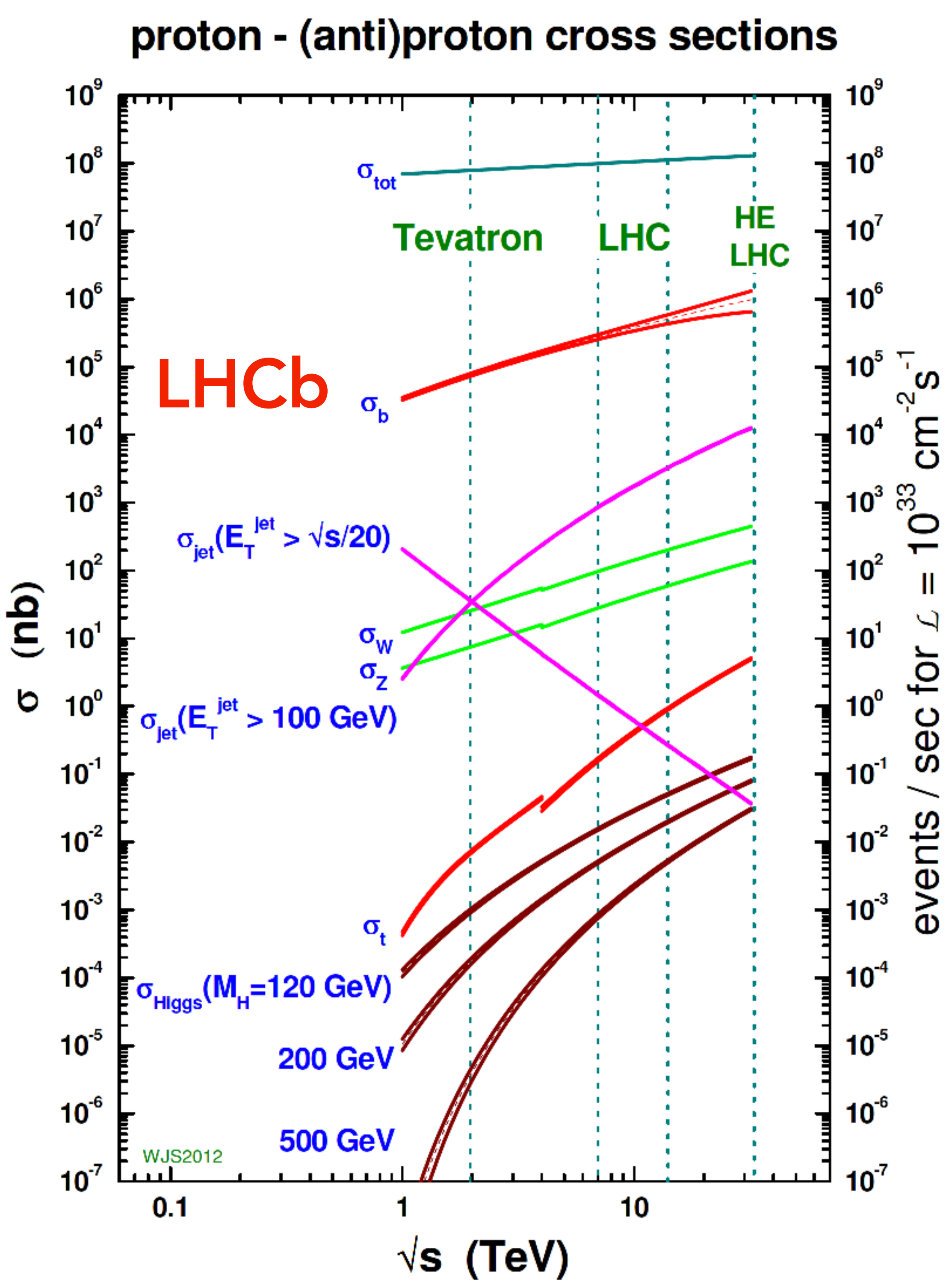


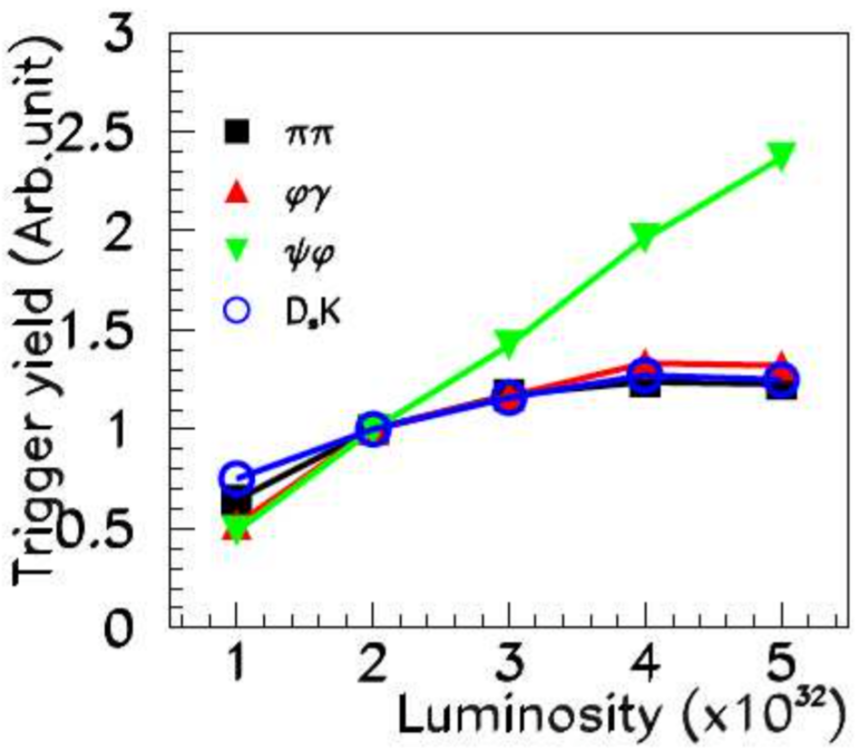
Driven by fixed-latency selection, analysis on efficiency plateau

Modern real-time processing, or "real-time analysis"



Modern real-time processing, or "real-time analysis"





LHC increases its luminosity by generating multiple pp interactions in a single bunch crossing

Fixed latency triggers select bunch crossings

Beyond some luminosity, all bunch crossings contain signal.
Select interactions, not bunch crossings => real-time analysis.

No possibility to work on efficiency plateau!

Largely compression not selection, variable latency by necessity

Before we proceed... credit where it is due

The work described in this habilitation is the result of an enormous team effort by many of my LHCb colleagues

I was lucky enough to coordinate a particularly brilliant High Level Trigger team, who came together ex-nihilo to make real-time analysis possible despite the lack of any funding agency support for our work.

Good ideas are cheap, teams which are able to bring those good ideas to life are very hard to find. I hope that this won't be the last challenge we tackle together.

Acknowledgements

During the years in which the real-time analysis described in this thesis was conceived, implemented, and commissioned I had the priviledge to work in the High Level Trigger team of LHCb, and I suspect many years will pass before I am again fortunate enough to collaborate with so many brilliant and generous people at once. None of the work described in this HDR would have been possible without them, or without the reconstruction, alignment, calibration, online, and offline computing teams of LHCb, who embraced the idea of real-time analysis and brought it to life. You know who you are, this work is ours, and I simply hope to have done it justice in the writeup.

That being said, I have used code written by my colleagues to produce many of the plots in this document, and I wish to acknowledge those cases more specifically. The results presented in the "haystack of needles" chapter, which form the historical physics case for real-time analysis in LHCb, were produced in collaboration with Conor Fitzpatrick, who wrote all the code. Similarly, Mike Sokoloff wrote the code used to produce most of the plots in the "making time less real" chapter. The analysis of charm cross-sections was done in collaboration with a large number of colleagues, and while I have highlighted some of my own technical contributions most of the scripts and code used to produce the results were written by others, in particular Christopher Burr, Dominik Mueller, Alex Pearce, and Patrick Spradlin. A special thank you is due Alex and Dominik for maintaining a reproducible version of the analysis framework such that I could rerun this code more than two years after the fact and remake most of the analysis plots myself.

I must also pay special respect to two people who really went above and beyond in making real-time analysis possible in LHCb. Silvia Borghi led the development and deployment of the real-time detector alignment and calibration, and has spent much of her evenings since 2015 perfecting every detail and training the next generation of LHCb reconstruction and calibration experts. Provela si ekipu kroz nevreme pravo, šta da ti kažem sem drugarice bravo![1] And Gerhard Raven... is simply the *nec plus ultra* of real-time data processing in High Energy Physics. Gerhard wrote much if not most of the code behind LHCb's High Level Trigger, and laid the foundations on which all the work described in this HDR stands. Ave maestro.

Before we proceed… credit where it is due

The work described in this habilitation is the result of an enormous team effort by many of my LHCb colleagues

I was lucky enough to coordinate a particularly brilliant High Level Trigger team, who came together ex-nihilo to make real-time analysis possible despite the lack of any funding agency support for our work.

Good ideas are cheap, teams which are able to bring those good ideas to life are very hard to find. I hope that this won't be the last challenge we tackle together.

Acknowledgements

During the years in which the real-time analysis described in this thesis was conceived, implemented, and commissioned I had the priviledge to work in the High Level Trigger team of LHCb, and I suspect many years will pass before I am again fortunate enough to collaborate with so many brilliant and generous people at once. None of the work described in this HDR would have been possible without them, or without the reconstruction, alignment, calibration, online, and offline computing teams of LHCb, who embraced the idea of real-time analysis and brought it to life. You know who you are, this work is ours, and I simply hope to have done it justice in the writeup.

That being said, I have used code written by my colleagues to produce many of the plots in this document, and I wish to acknowledge those cases more specifically. The results presented in the "haystack of needles" chapter, which form the historical physics case for real-time analysis in LHCb, were produced in collaboration with Conor Fitzpatrick, who wrote all the code. Similarly, Mike Sokoloff wrote the code used to produce most of the plots in the "making time less real" chapter. The analysis of charm cross-sections was done in collaboration with a large number of colleagues, and while I have highlighted some of my own technical contributions most of the scripts and code used to produce the results were written by others, in particular Christopher Burr, Dominik Mueller, Alex Pearce, and Patrick Spradlin. A special thank you is due Alex and Dominik for maintaining a reproducible version of the analysis framework such that I could rerun this code more than two years after the fact and remake most of the analysis plots myself.

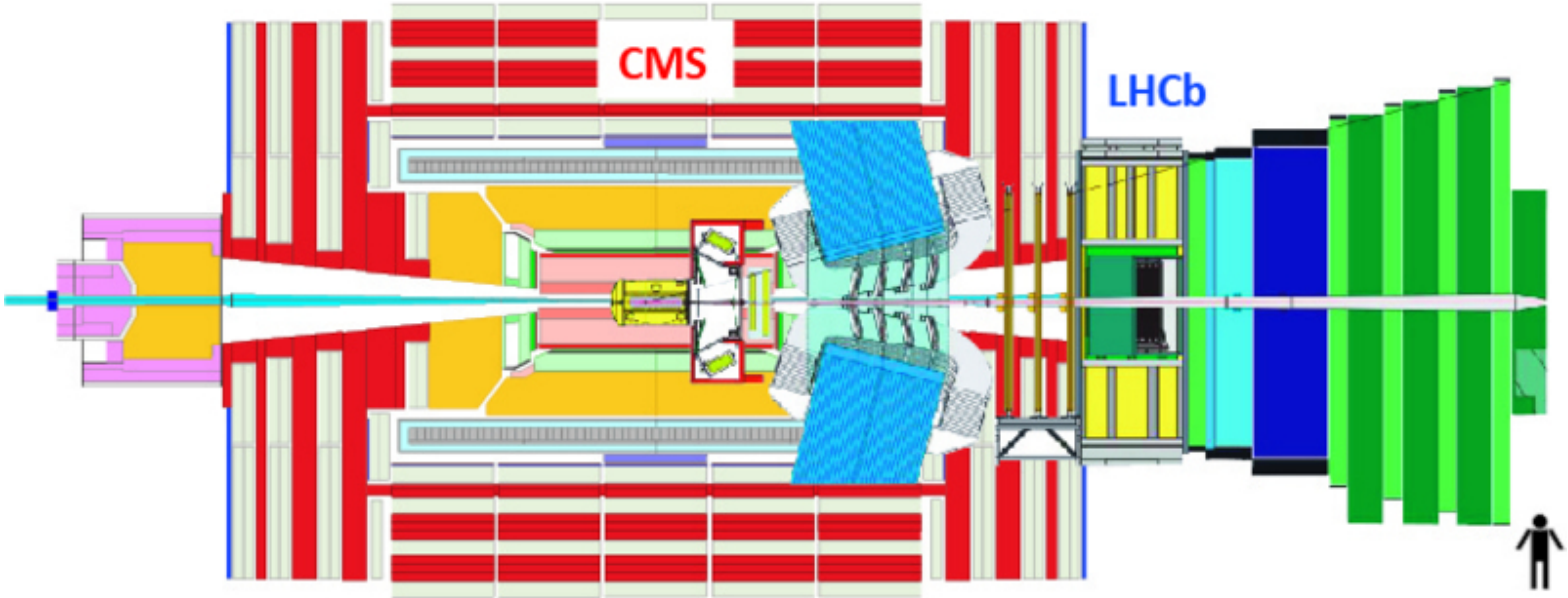
I must also pay special respect to two people who really went above and beyond in making real-time analysis possible in LHCb. Silvia Borghi led the development and deployment of the real-time detector alignment and calibration, and has spent much of her evenings since 2015 perfecting every detail and training the next generation of LHCb reconstruction and calibration experts. Provela si ekipu kroz nevreme pravo, šta da ti kažem sem drugarice bravo![1] And Gerhard Raven... is simply the *nec plus ultra* of real-time data processing in High Energy Physics. Gerhard wrote much if not most of the code behind LHCb's High Level Trigger, and laid the foundations on which all the work described in this HDR stands. Ave maestro.

You enjoyed this work? Great, now hire the rest of the team.

Mise-en-scène :

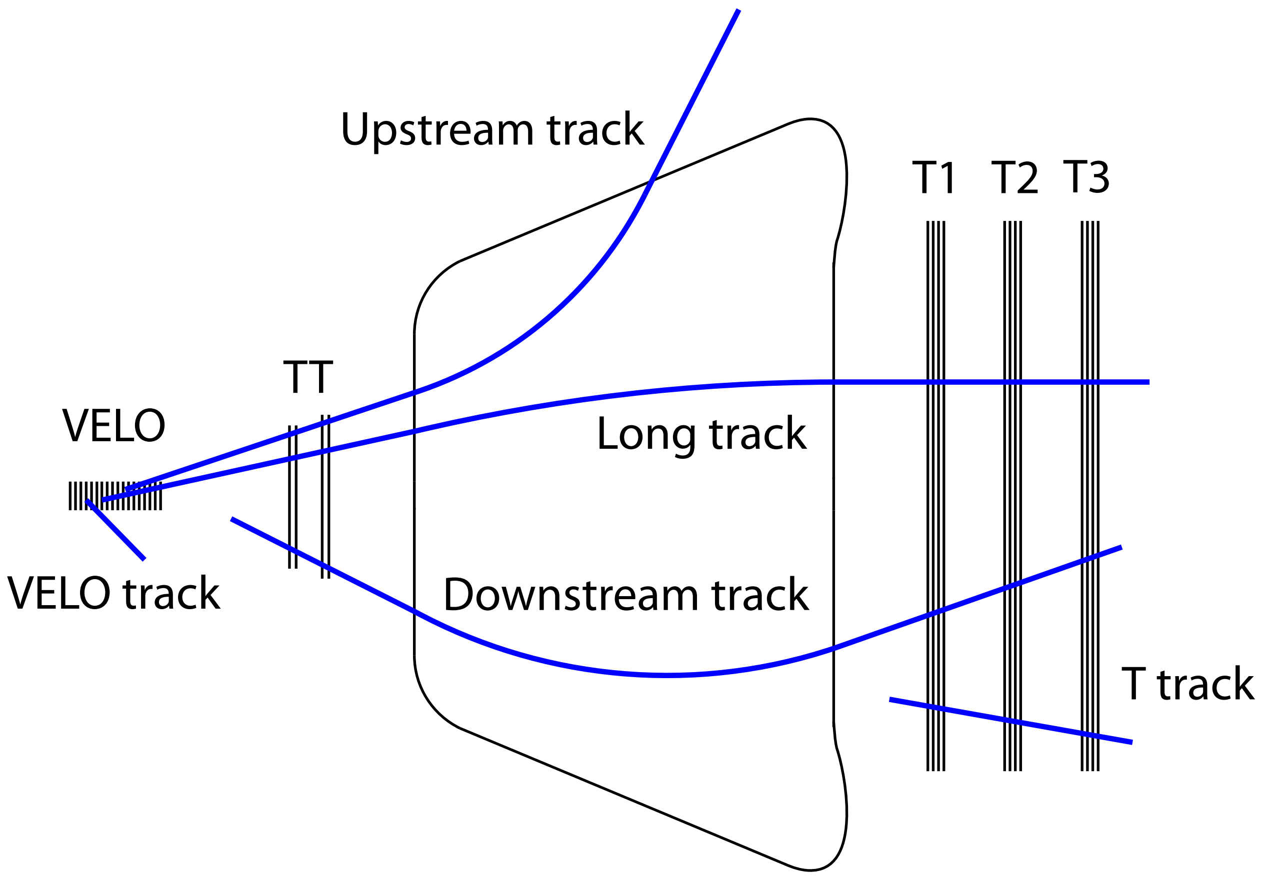
the LHCb detector and analysis methodology

The LHCb detector at the LHC



Forward spectrometer optimized for precision physics

Reconstruction philosophy and role of subdetectors



Tracker : chaged particle reconstruction

↓

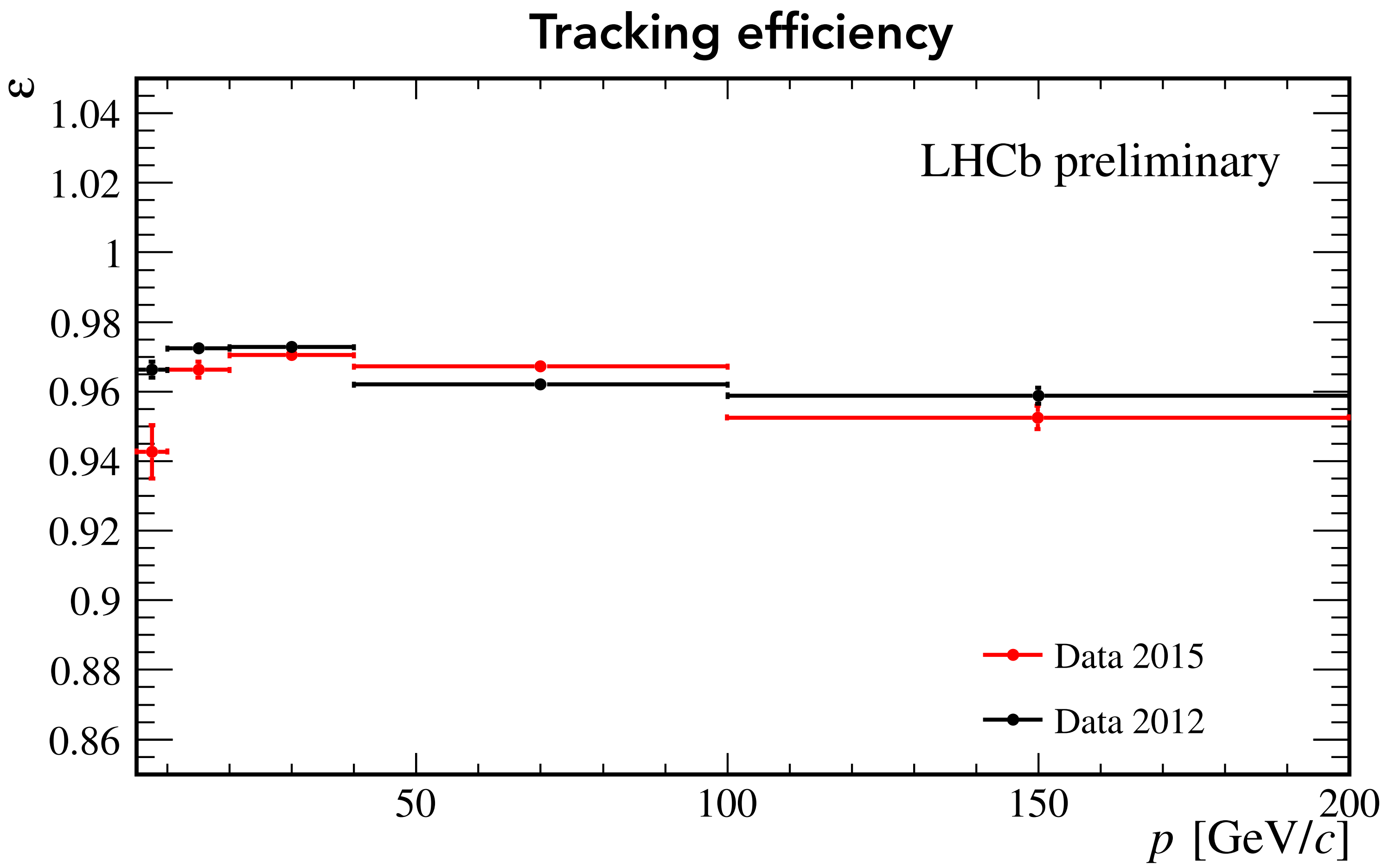
Particle identification : RICH, Muon,

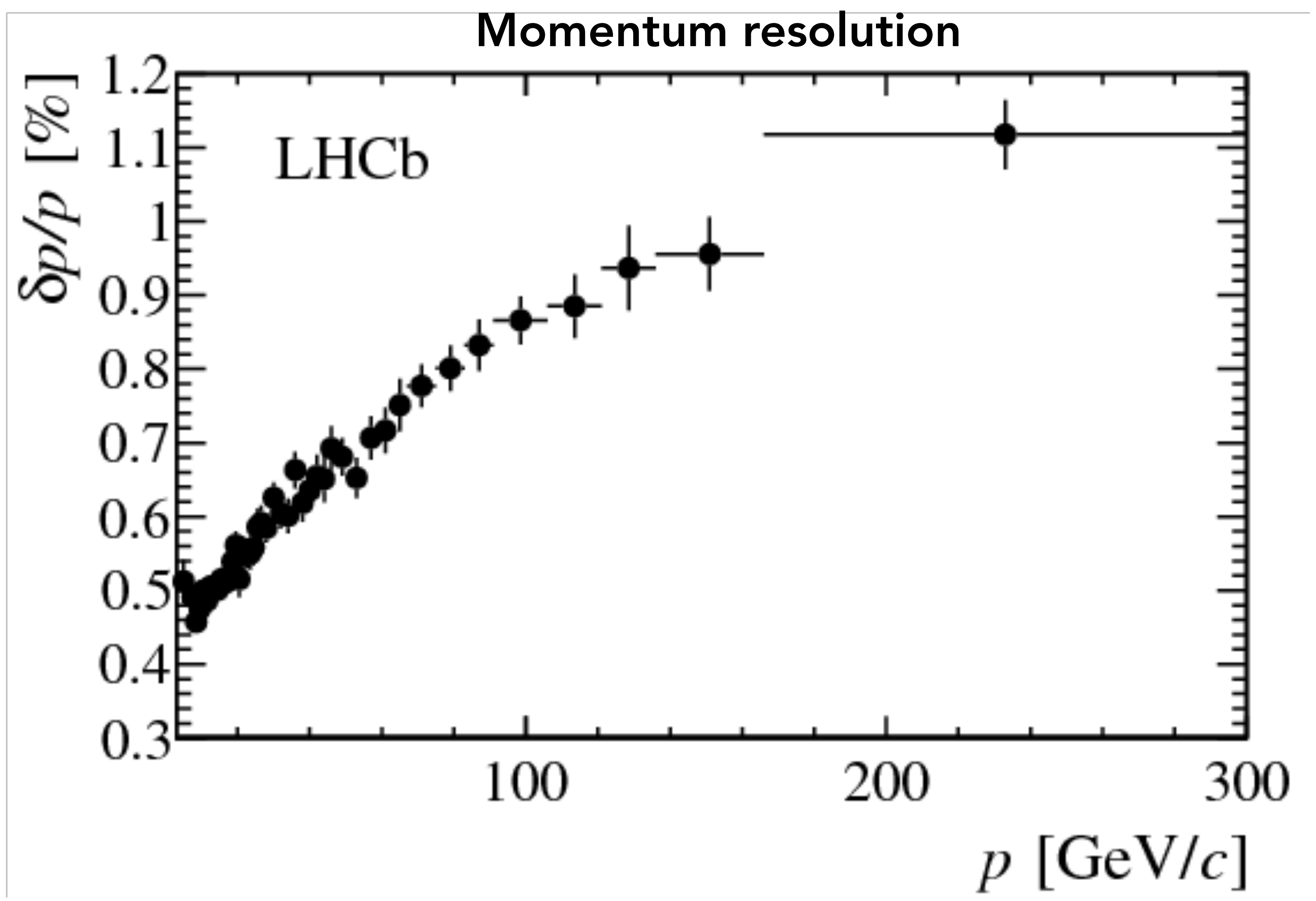
↓

Neutral reconstruction : ECAL

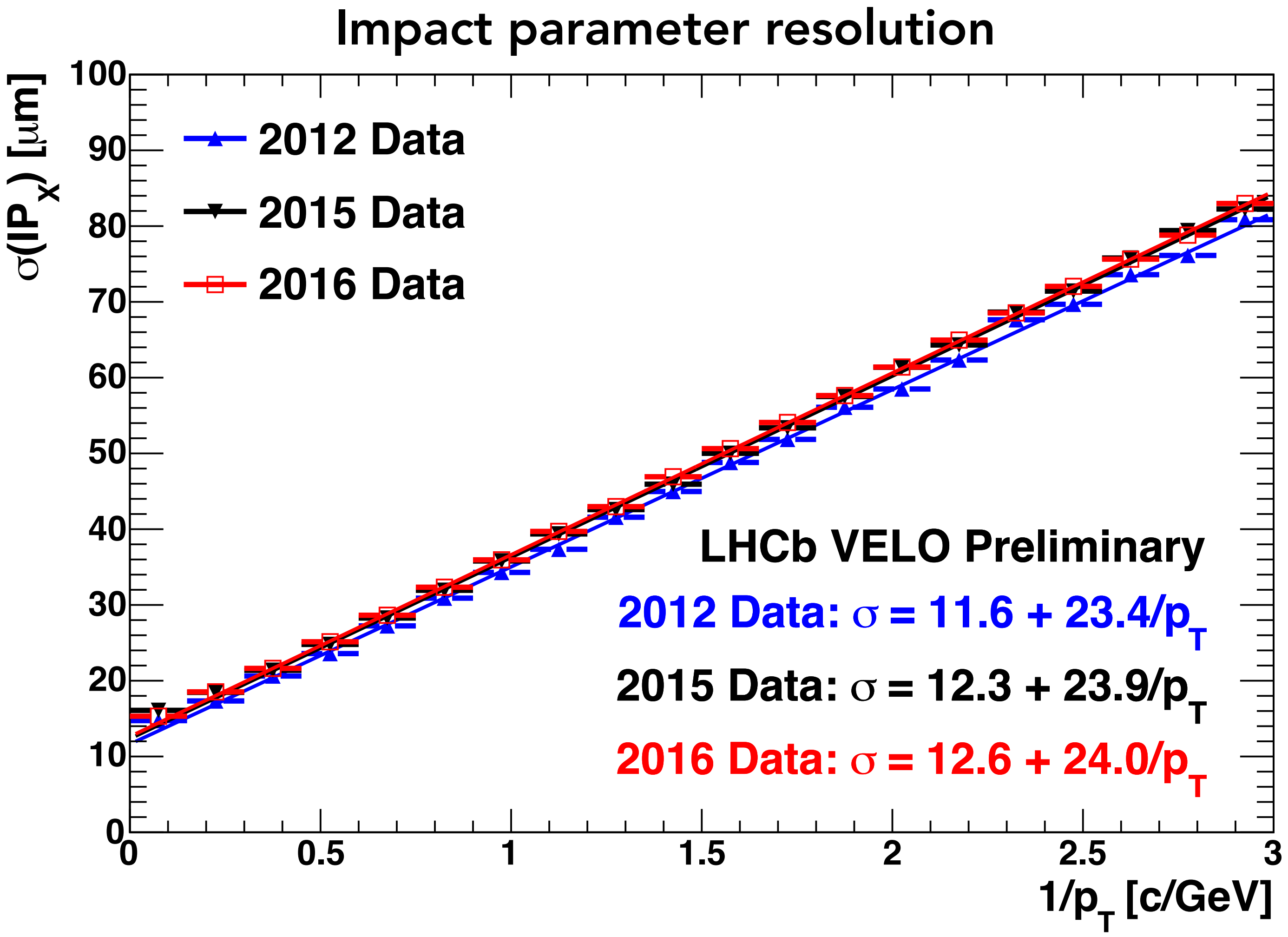
Optimized for charged particles w/some neutral capability

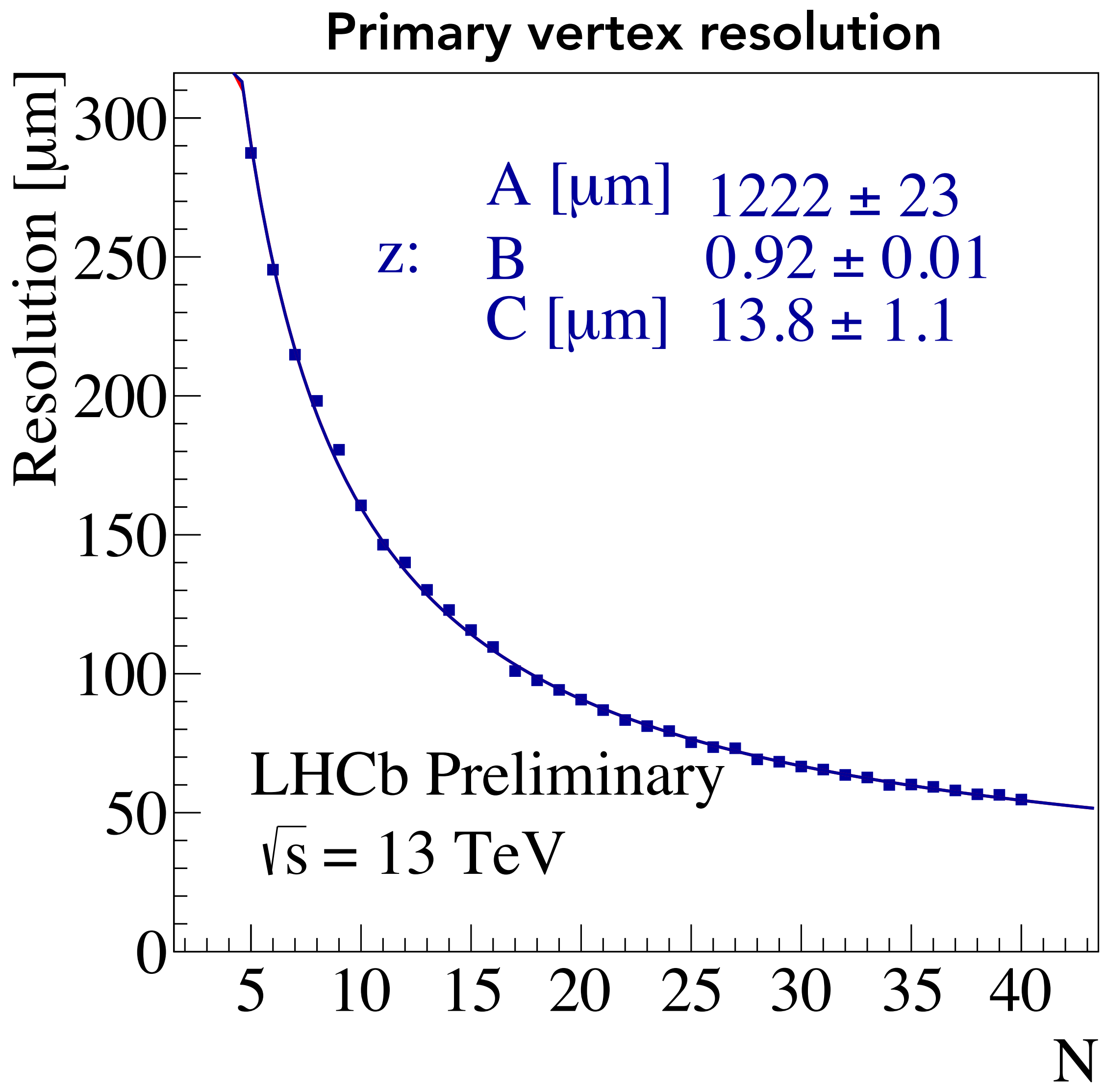
Charged particle reconstruction



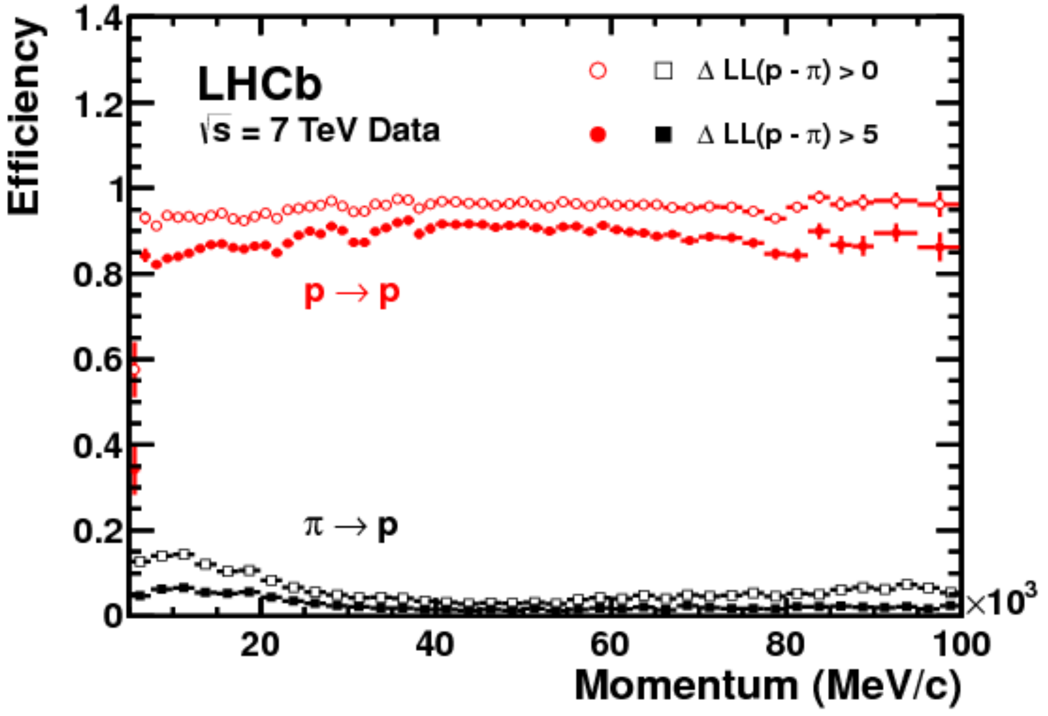


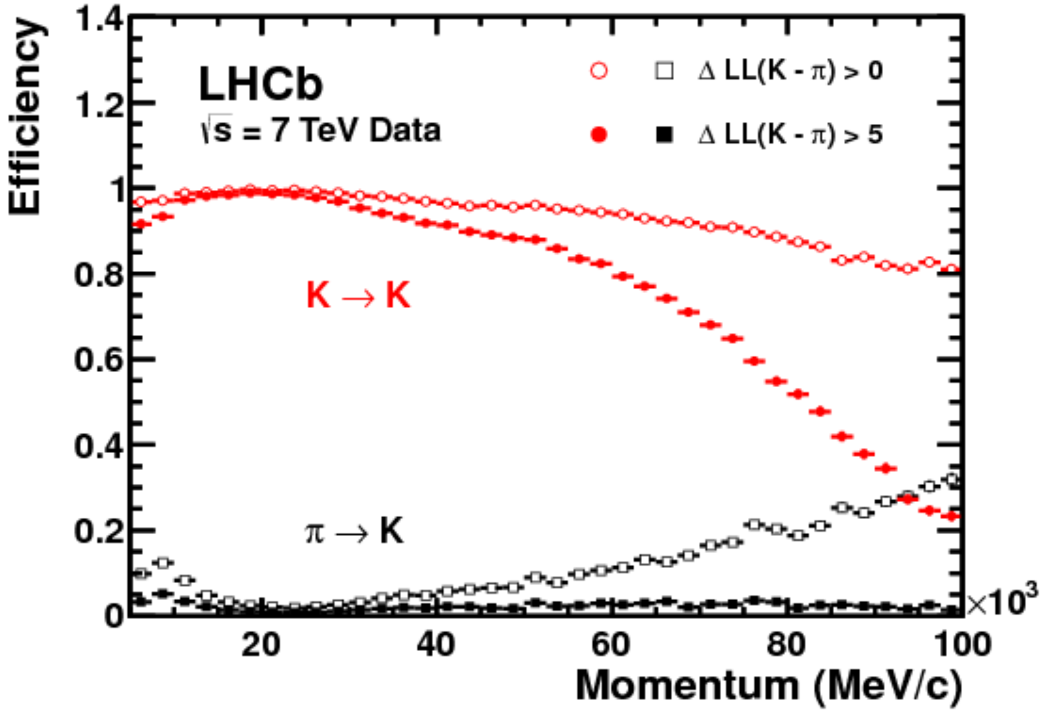
Vertexing performance

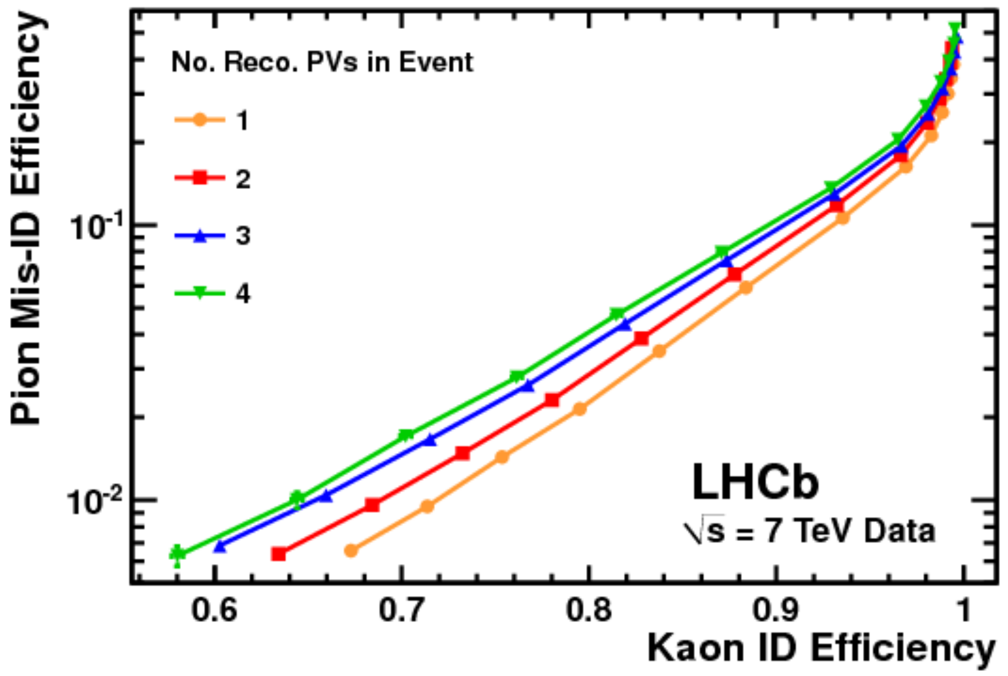




Particle identification @ LHCb





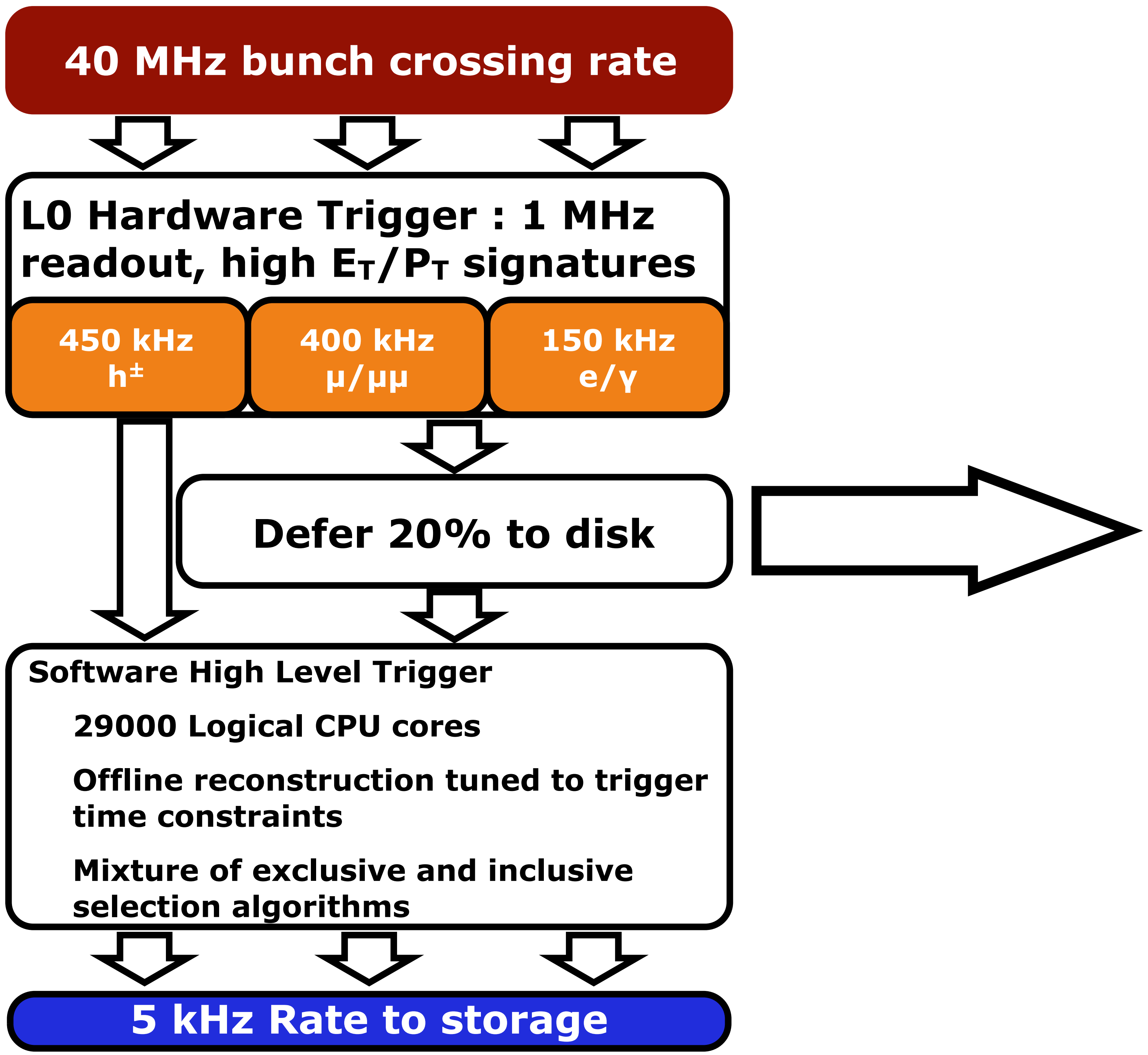


In addition to "usual" muon system and ECAL + preshower based electron identification, LHCb can separate charged hadrons using two Ring Imaging Cherenkov (RICH) detectors.

RICH detectors also contribute to electron/muon identification : in practice, all subdetector information is combined using neural networks to achieve the best possible particle identification.

Made possible by forward layout, shields photodetectors

Real-time data processing strategy during Run I



Ability to buffer events within High Level Trigger developed by online team during Run I, enabled real-time analysis in Run II

Largely fixed+variable latency *selection*

LHCb physics programme

Subject	Analyses	Historical?
b-hadrons	Searches for rare decays Time-integrated CP violation Time-dependent CP violation Dalitz measurements Angular measurements Radiative decays	YES
	Searches for forbidden decays	NO
c-hadrons	Searches for rare decays Searches for forbidden decays Time-integrated CP violation Time-dependent CP violation Dalitz measurements	NO
s-hadrons	Searches for rare decays Searches for forbidden decays	NO
Spectroscopy	Hadron masses Hadron quantum numbers Penta and tetraquark searches Hadron differential cross-sections Exclusive production of hadrons	NO
	Hadron widths or lifetimes	YES
Electroweak and top	EW boson differential cross-sections EW boson forward-backward asymmetries Single and double-top differential cross-sections	NO
Exotica	Direct searches for new particles	NO
Ion and fixed target physics	Hadron differential cross-sections EW boson differential cross-sections	NO

Greatly expanded since 2010 thanks to trigger flexibility

LHCb analysis methodology and role of calibration samples

Trigger Efficiency

Tag-and-probe calibration method exists & widely used

Tracking efficiency

Tag-and-probe

Existing	Developing
μ	e,π,K,p

Particle identification

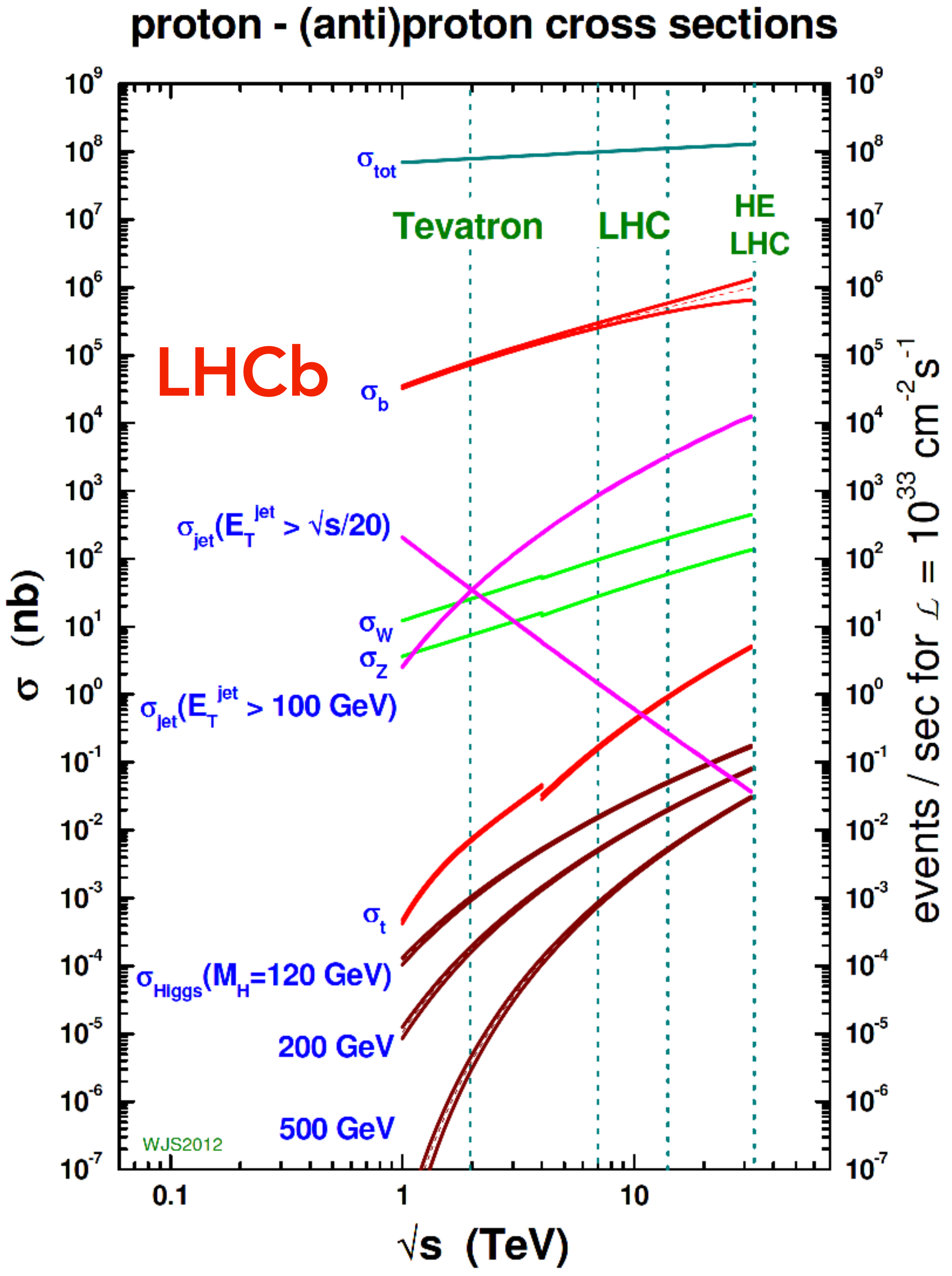
Tag-and-probe

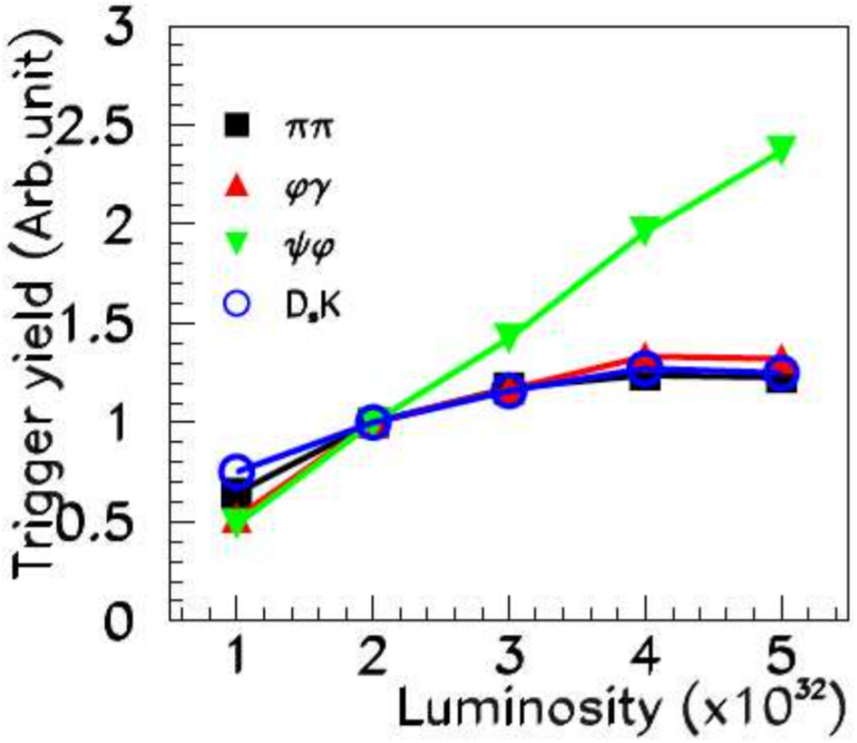
Tag-and-probe calibrations exist for all charged particle species and for π^0/γ, with new sources added over time to improve coverage

Data driven efficiency calibration key to precision physics

A haystack of needles : The necessity of real-time analysis in LHCb

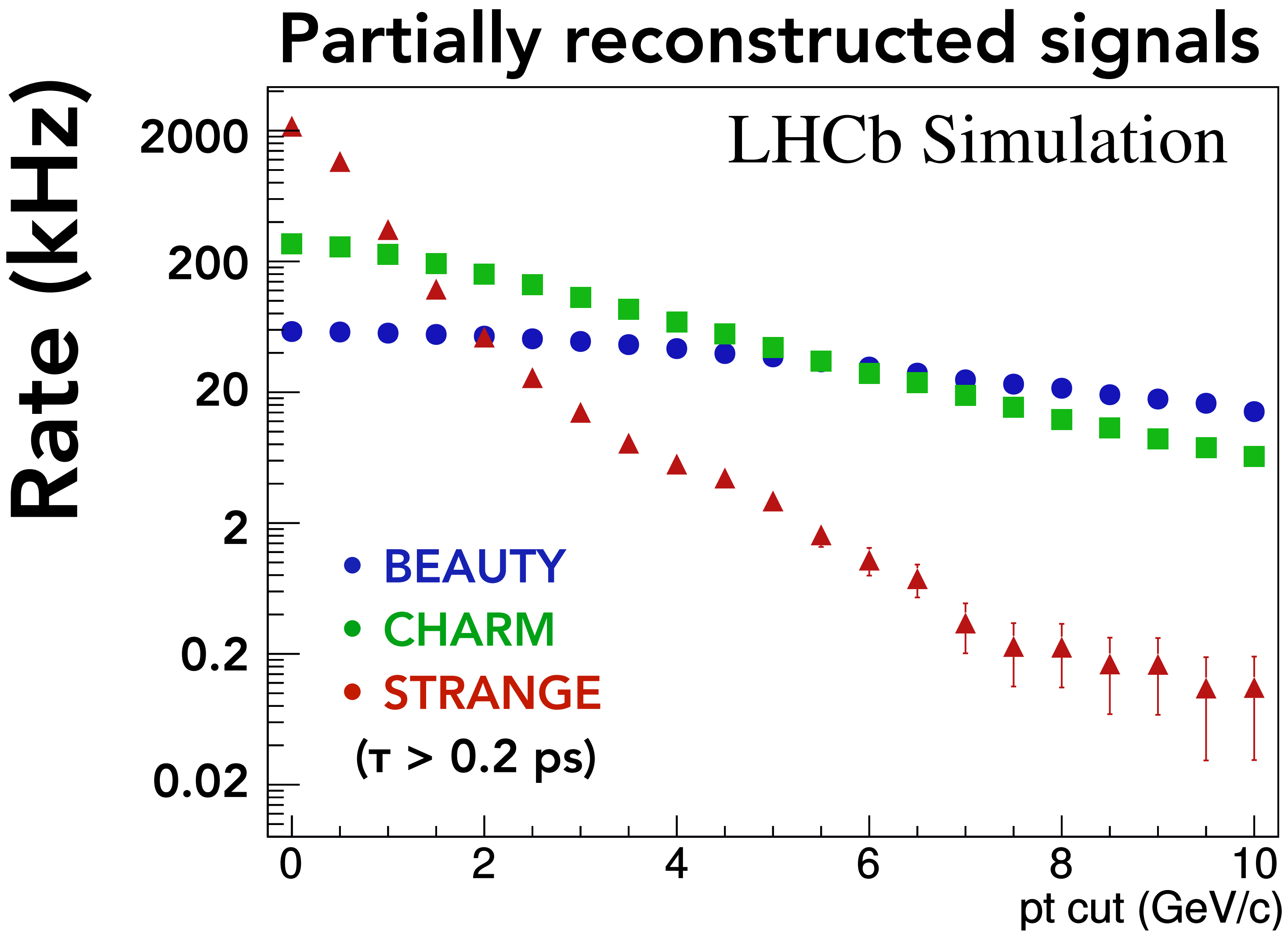
Why does LHCb not run at ATLAS/CMS luminosities today?





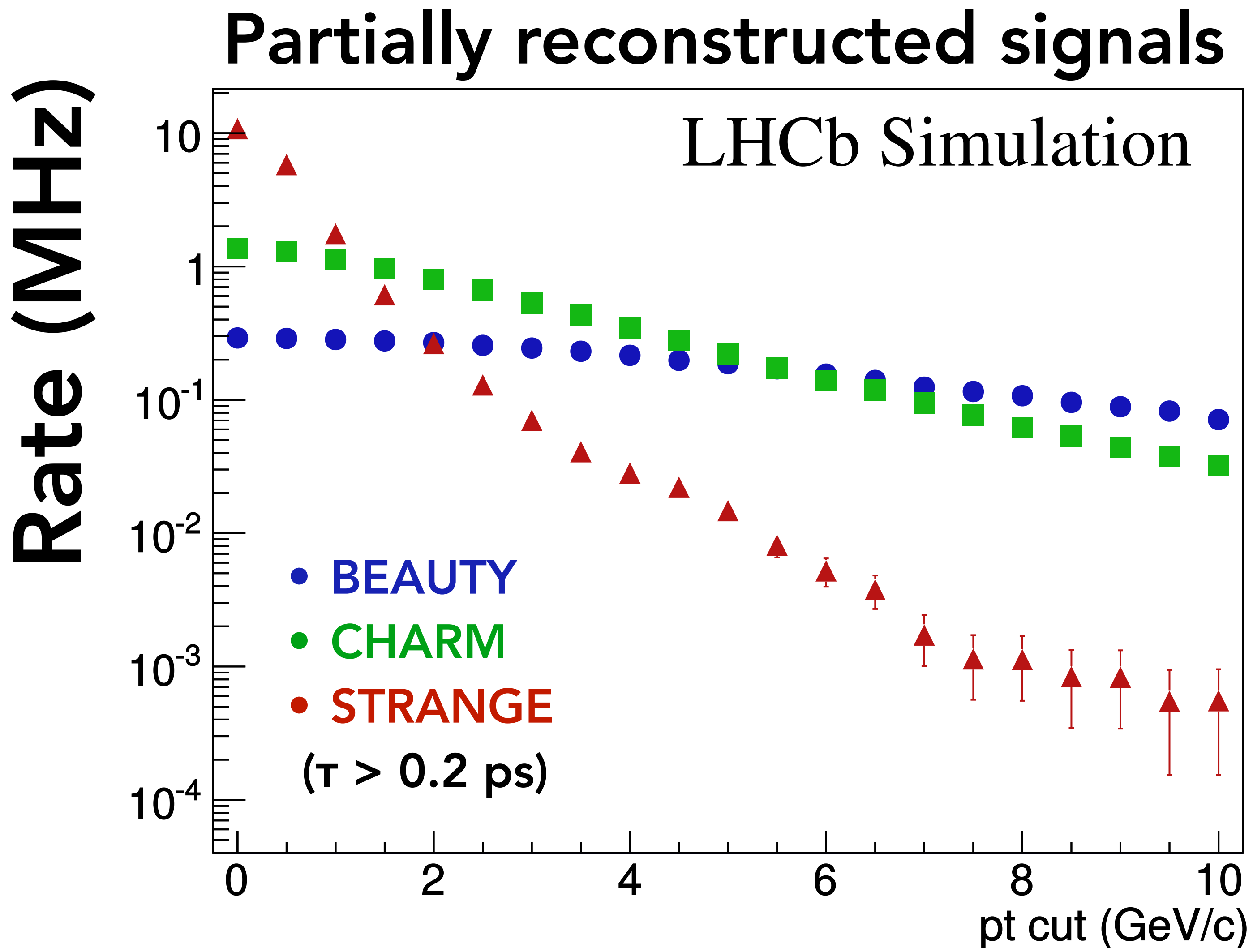
Fixed-latency trigger only effective up to around $4 \cdot 10^{32}$

Signal and data rates at LHCb in Run 2



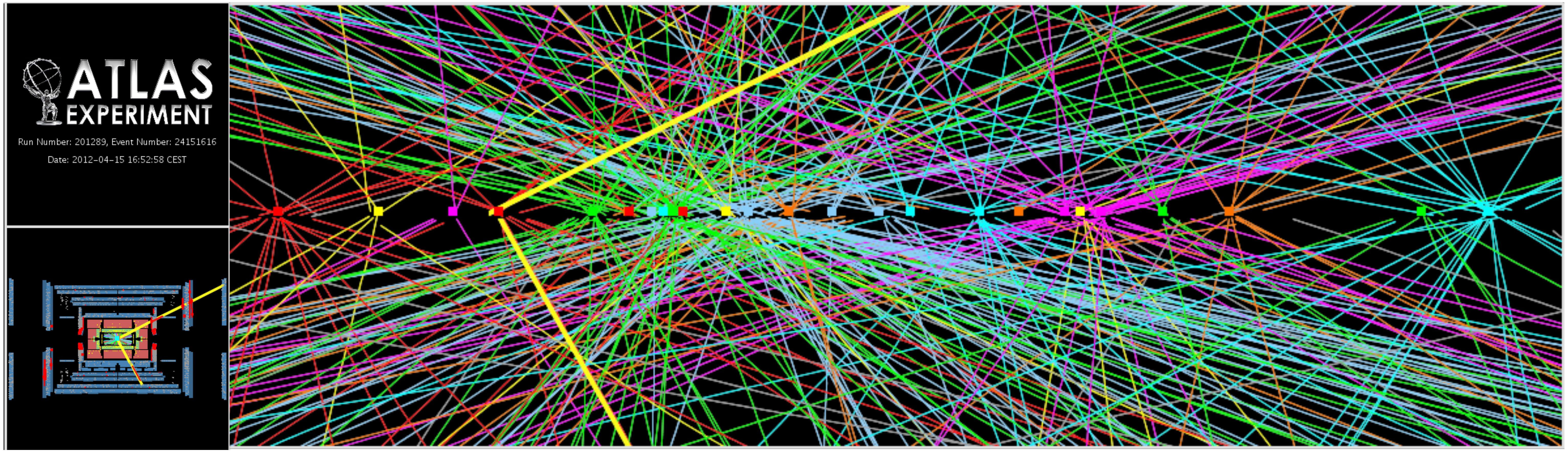
Already greatly exceeds allowed O(10kHz) bandwidth

Signal and data rates at LHCb in the upgrade



Plus data volume increases quadratically because of pileup

From selection to compression : real-time analysis



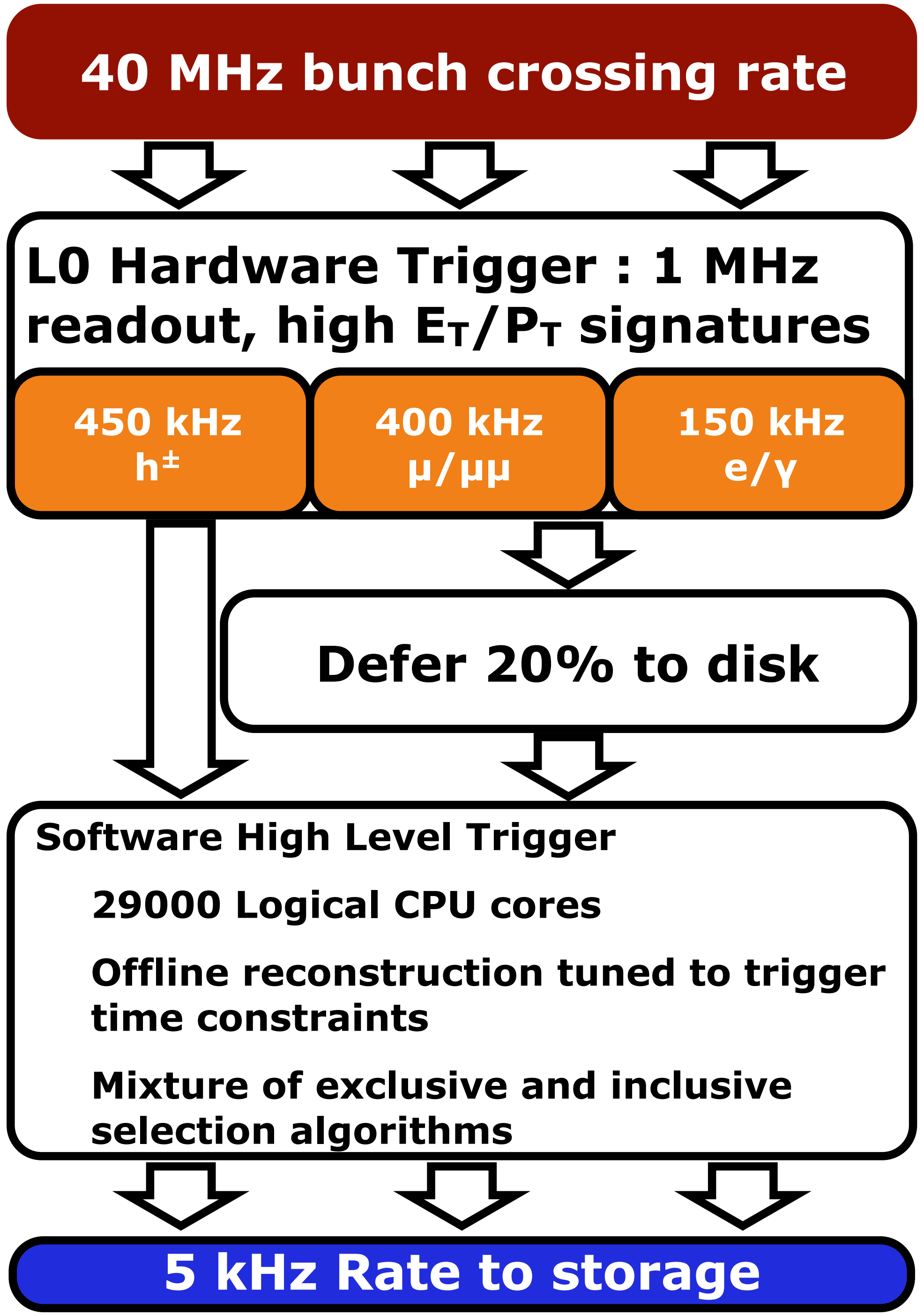
Most physics measurements require only a signal candidate and information about the specific pp collision which produced it → the rest is pileup

The higher the luminosity, the larger the fraction of event data caused by pileup

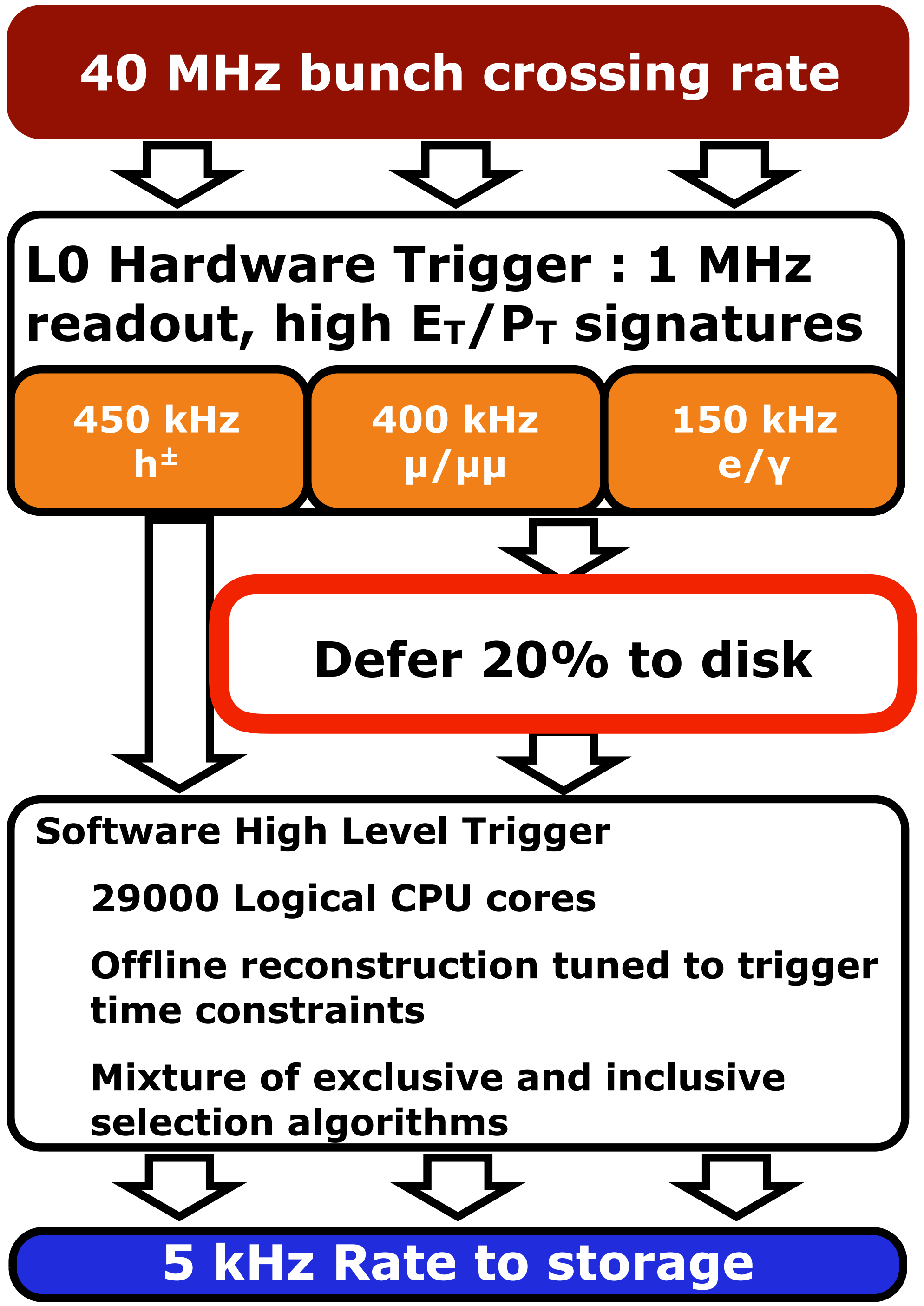
Hence create more room for signal by compressing & removing pileup in real-time!

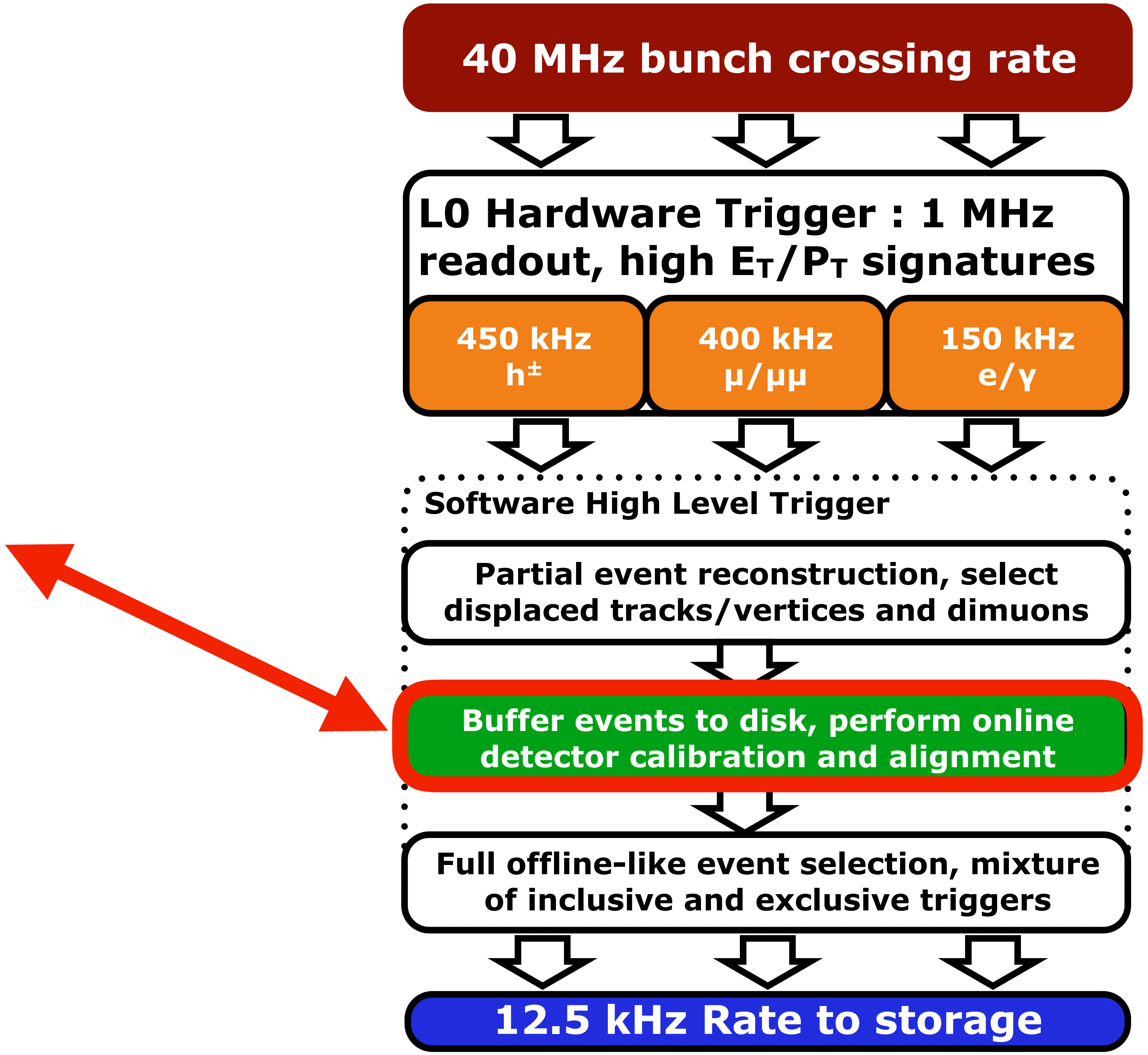
Requires the ability to carry out precise pileup suppression

Run 2 as a proving ground for the detector upgrade



Run 2 as a proving ground for the detector upgrade





Switch to real-time analysis in Run 2 to learn for the upgrade

A cunning plan:

the requirements for

real-time analysis

The necessary ingredients of a precision physics measurement

Alignment

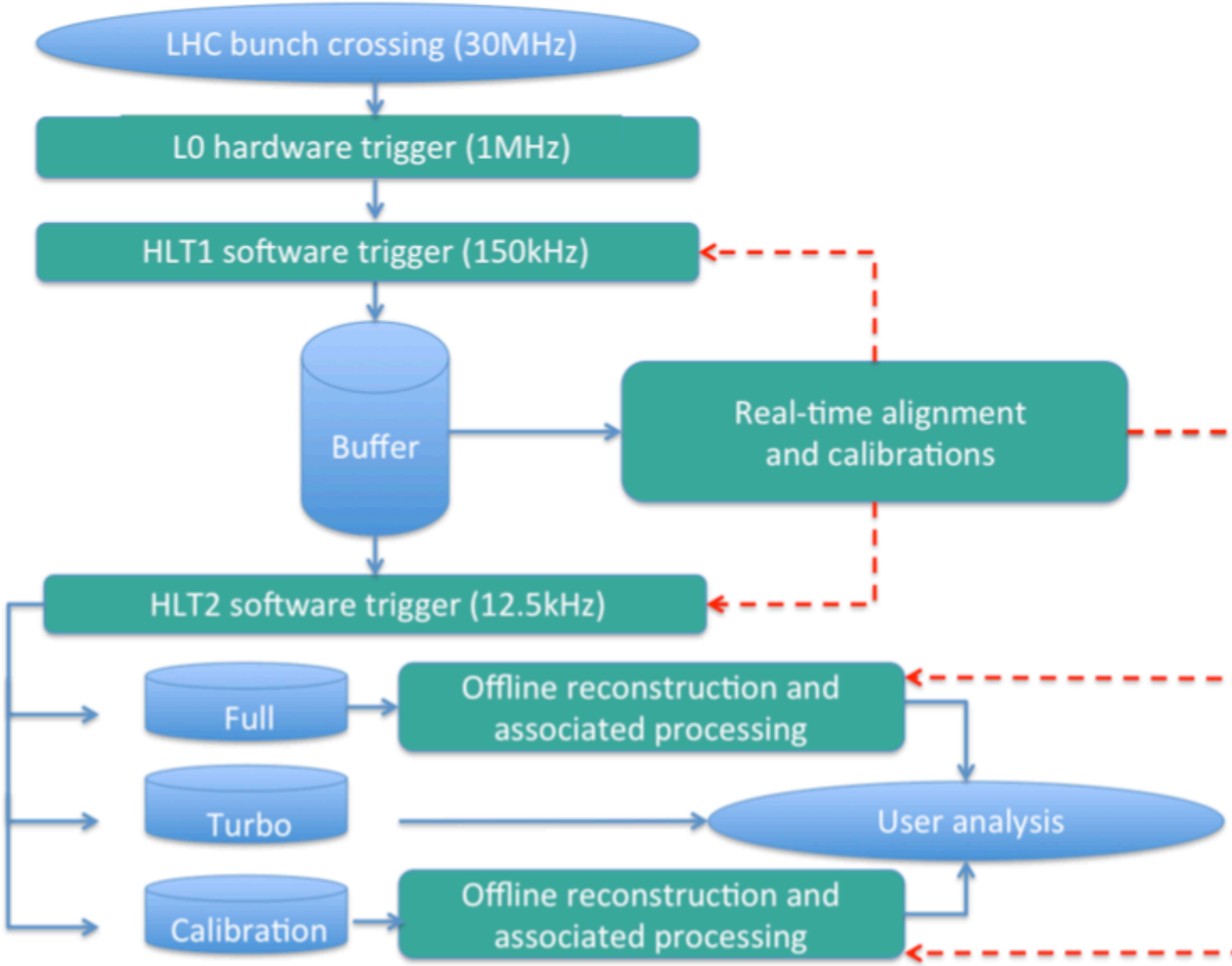
Full detector reconstruction

Calibration

Signal and control samples

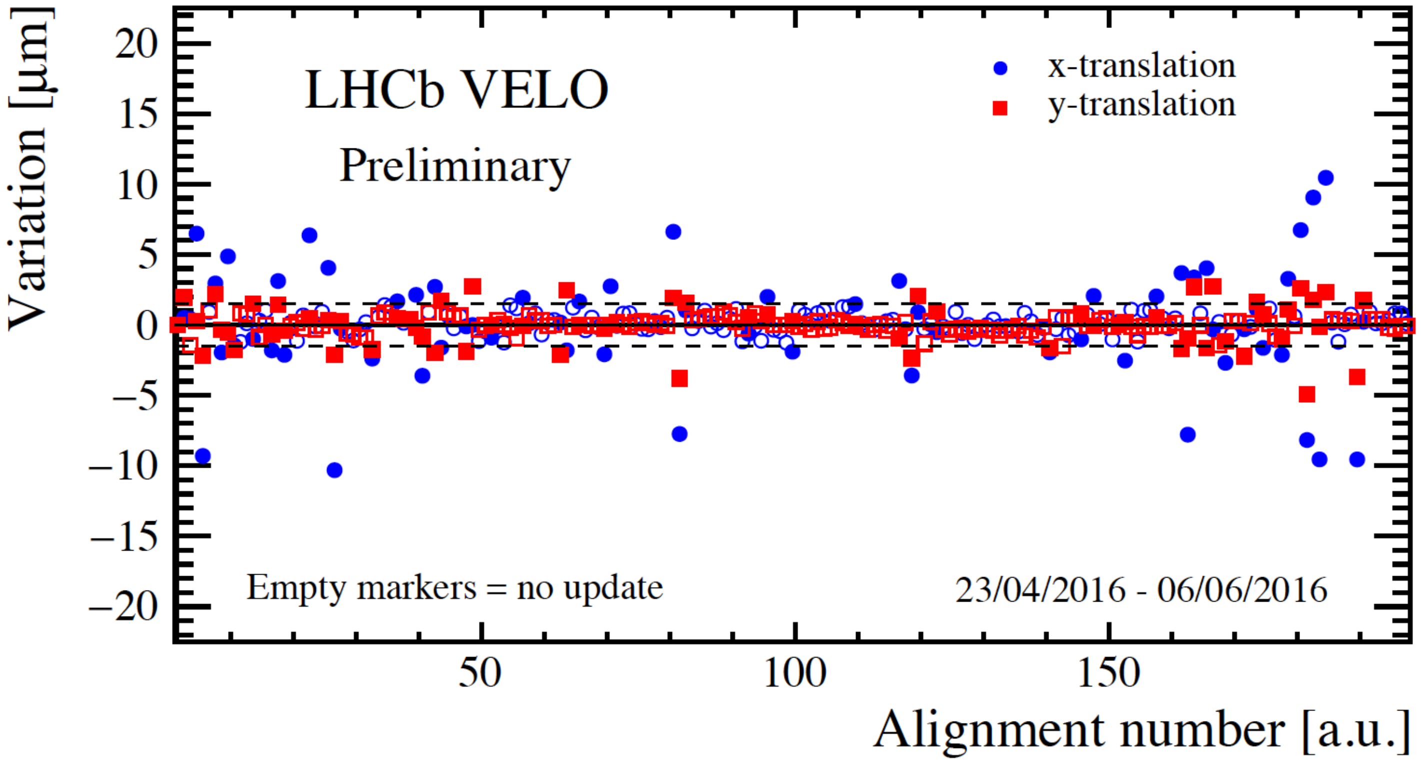
Monitoring & Data Quality

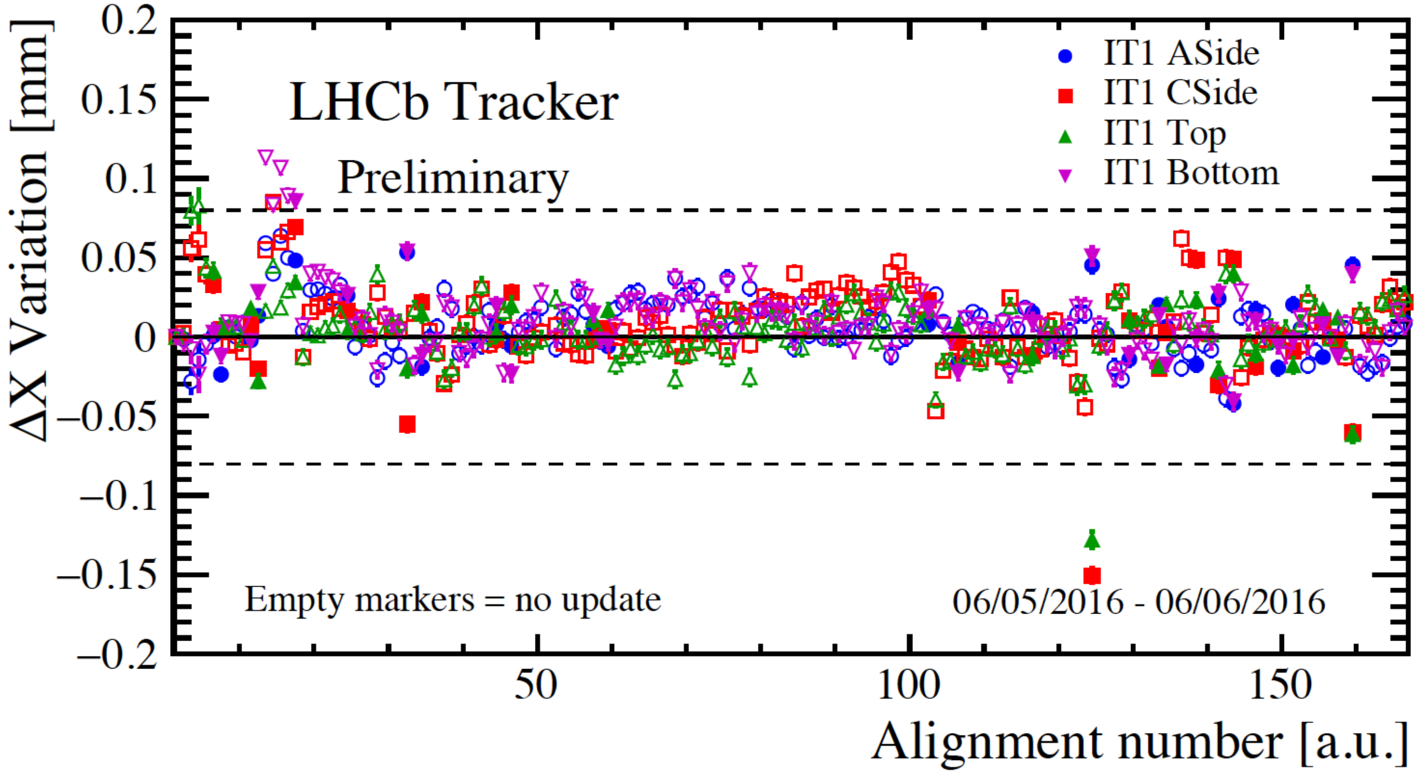
Addressing the requirements by splitting the HLT



Splitting the HLT enables the parallelization of alignment&calib

Aligning the detector in real time





Automated tracker alignment performed once per fill (in 8 min)

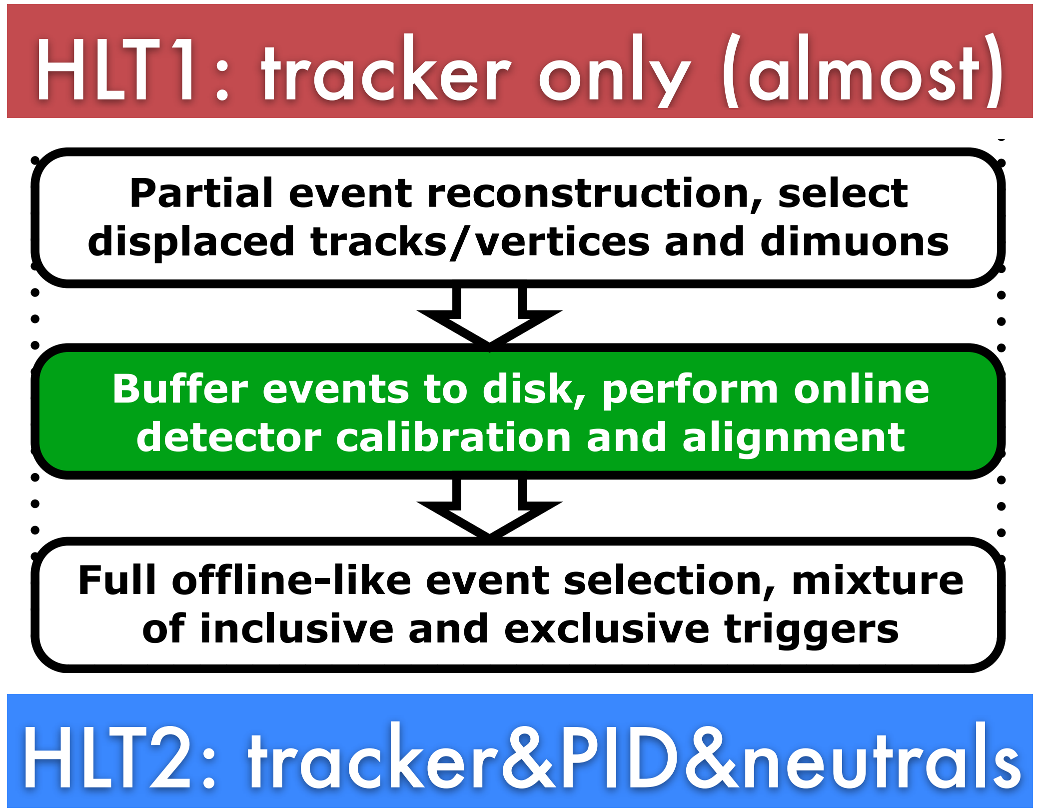
Fully reconstructing the detector in real time

No possibility of performing a full reconstruction at the 1 MHz rate coming out of the L0 trigger, therefore

1) Optimize the disk buffer between HLT1 and HLT2 to create time for the full offline reconstruction in HLT2

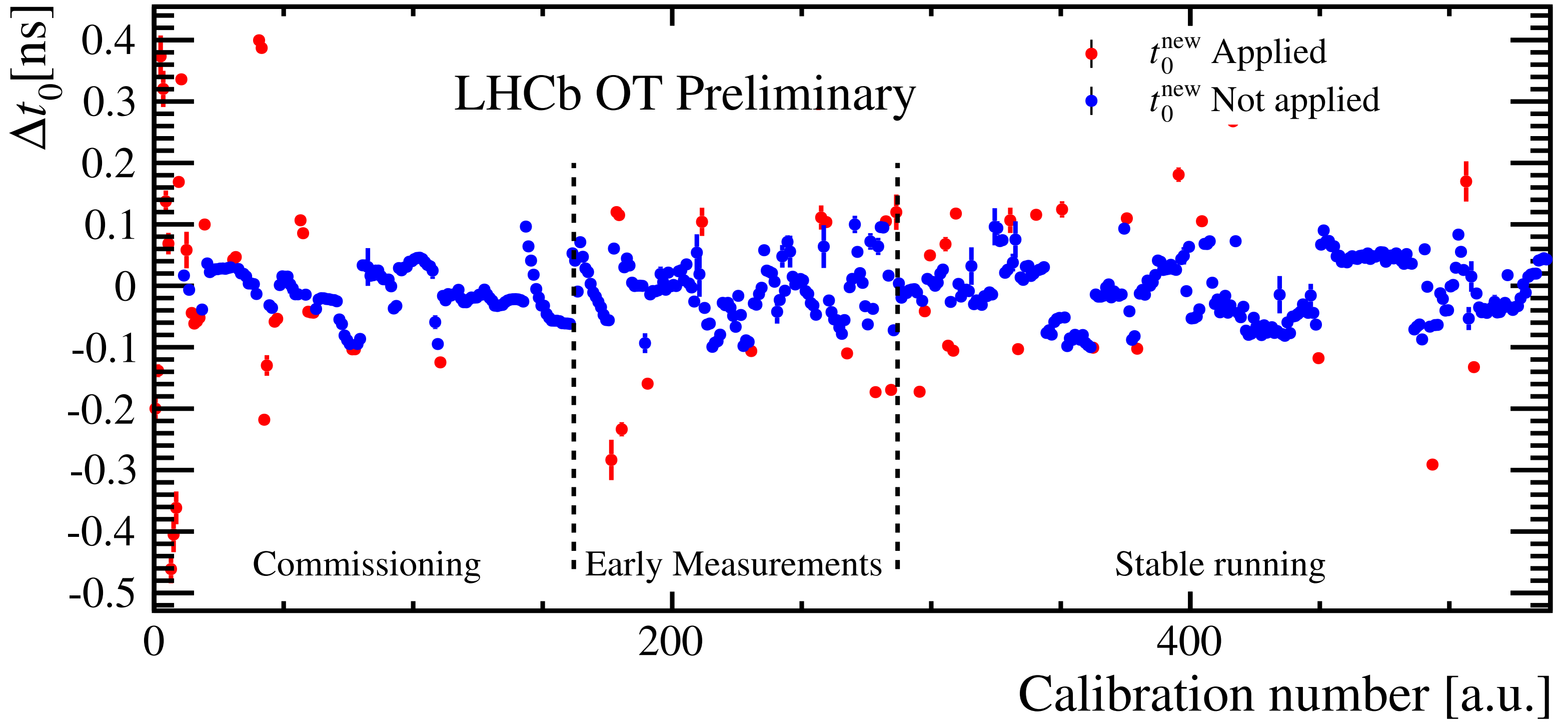
2) Optimize the vertex detector reconstruction so it could run with offline quality in HLT1

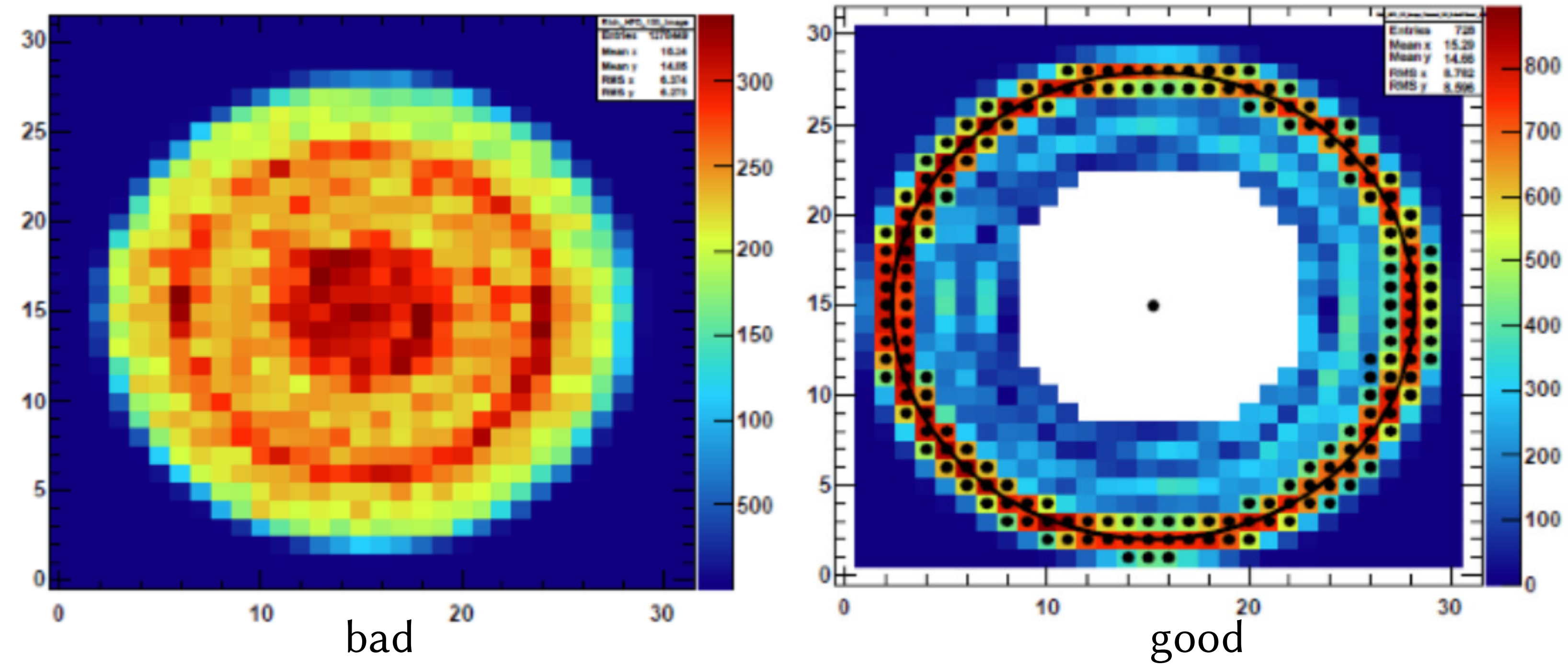
3) Show that the HLT2 tracker reconstruction can be factorized in particle kinematics, making the HLT1 reconstruction the high momentum subset of HLT2

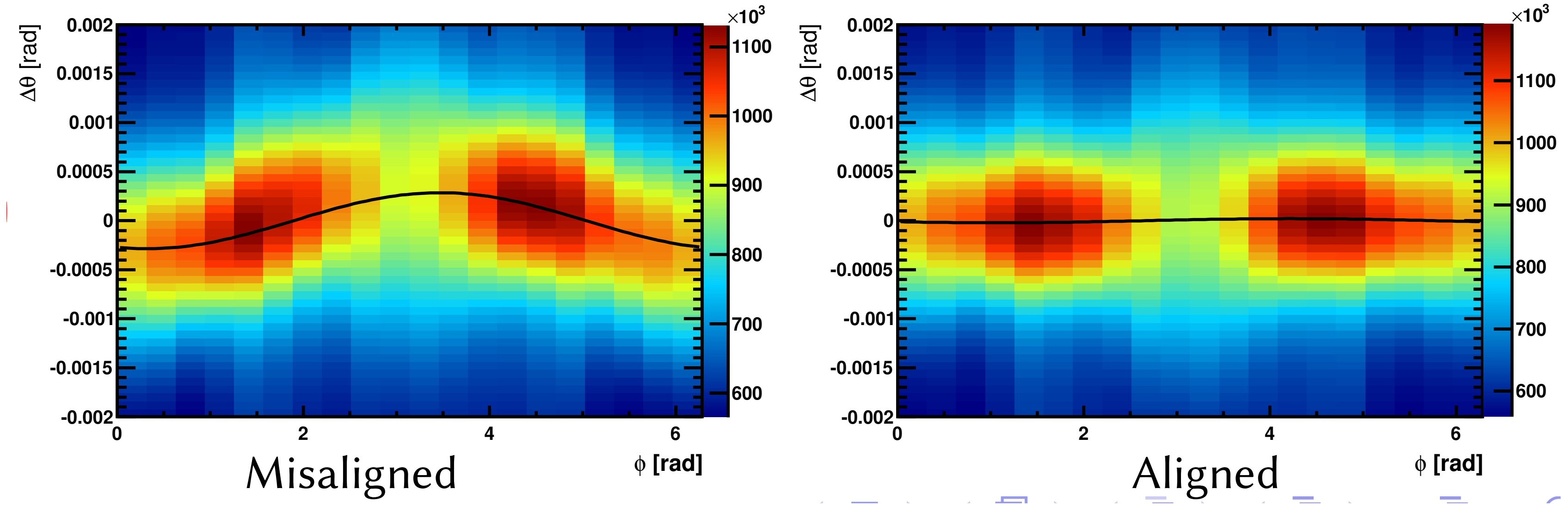


Enabled by putting the disk buffer between HLT1 and HLT2

Calibrating the detector in real time

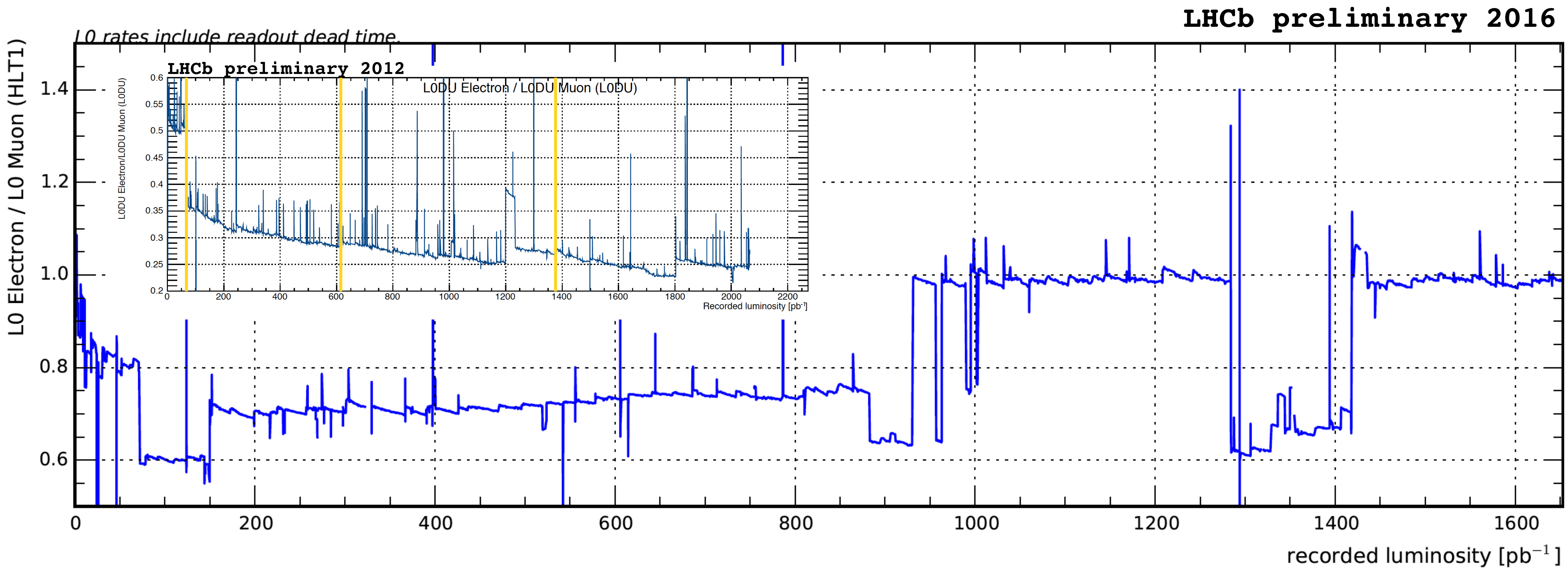






Automated RICH and straw-tube tracker alignment

Calibrating the calorimeter in real time



Automated occupancy-based calorimeter ageing correction

Selecting calibration samples in real time

Alignment/Calibration task	Sample	Size of sample
VELO	random triggers	$\mathcal{O}(100\text{k})$ events
Tracker	$D^0 \to K^-\pi^+$, high momentum tracks	$\mathcal{O}(200\text{k})$ events
Tracker vertical alignment	magnet off tracks	5-10M events
Muon system	$J/\psi \to \mu^+\mu^-$	$\mathcal{O}(250\text{k})$ events
RICH mirrors	equal occupancy triggers	$\mathcal{O}(3\text{M})$ events
RICH image	random triggers	
RICH refractive index	random triggers	
CALO coarse	random triggers	$\mathcal{O}(100\text{k})$ events
CALO fine	$\pi^0 \to \gamma\gamma$	

Species	Soft	Hard
$e^\pm$	—	$J/\psi \to e^+e^-$
$\mu^\pm$	$D_s^+ \to \mu^+\mu^-\pi^+$	$J/\psi \to \mu^+\mu^-$
$\pi^\pm$	$K_S^0 \to \pi^+\pi^-$	$D^{*+} \to D^0\pi^+$, $D^0 \to K^-\pi+$
$K^\pm$	$D_s^+ \to K^+K^-\pi+$	$D^{*+} \to D^0\pi^+$, $D^0 \to K^-\pi+$
$p^\pm$	$\Lambda \to p\pi^-$	$\Lambda \to p\pi^-$, $\Lambda_c^+ \to pK^-\pi^+$

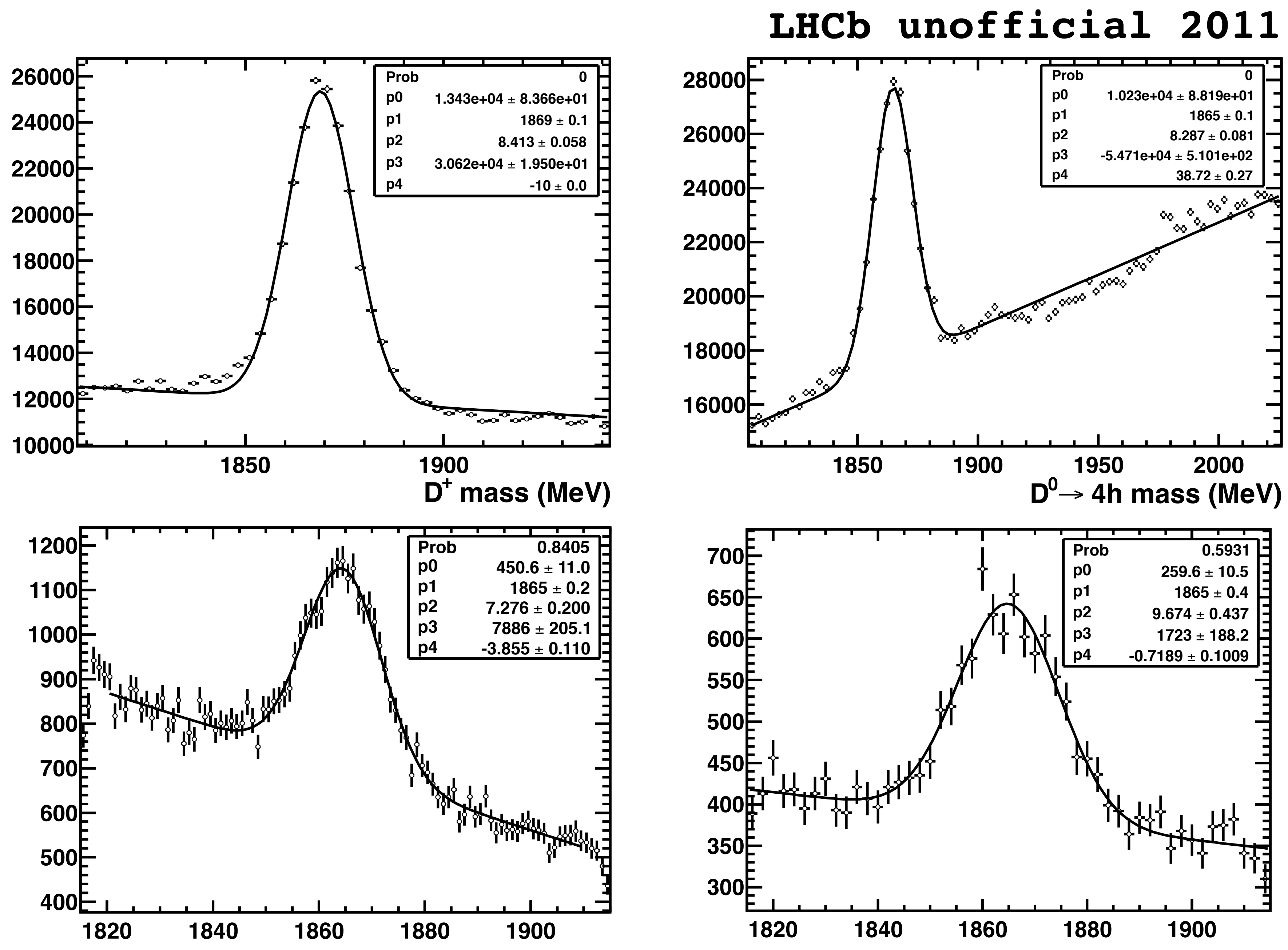
Must select all tag-and-probe calibration samples in real time!

Selecting signal samples in real time

During Run I, we already had clean fully reconstructed charm samples with good resolution coming out of the trigger

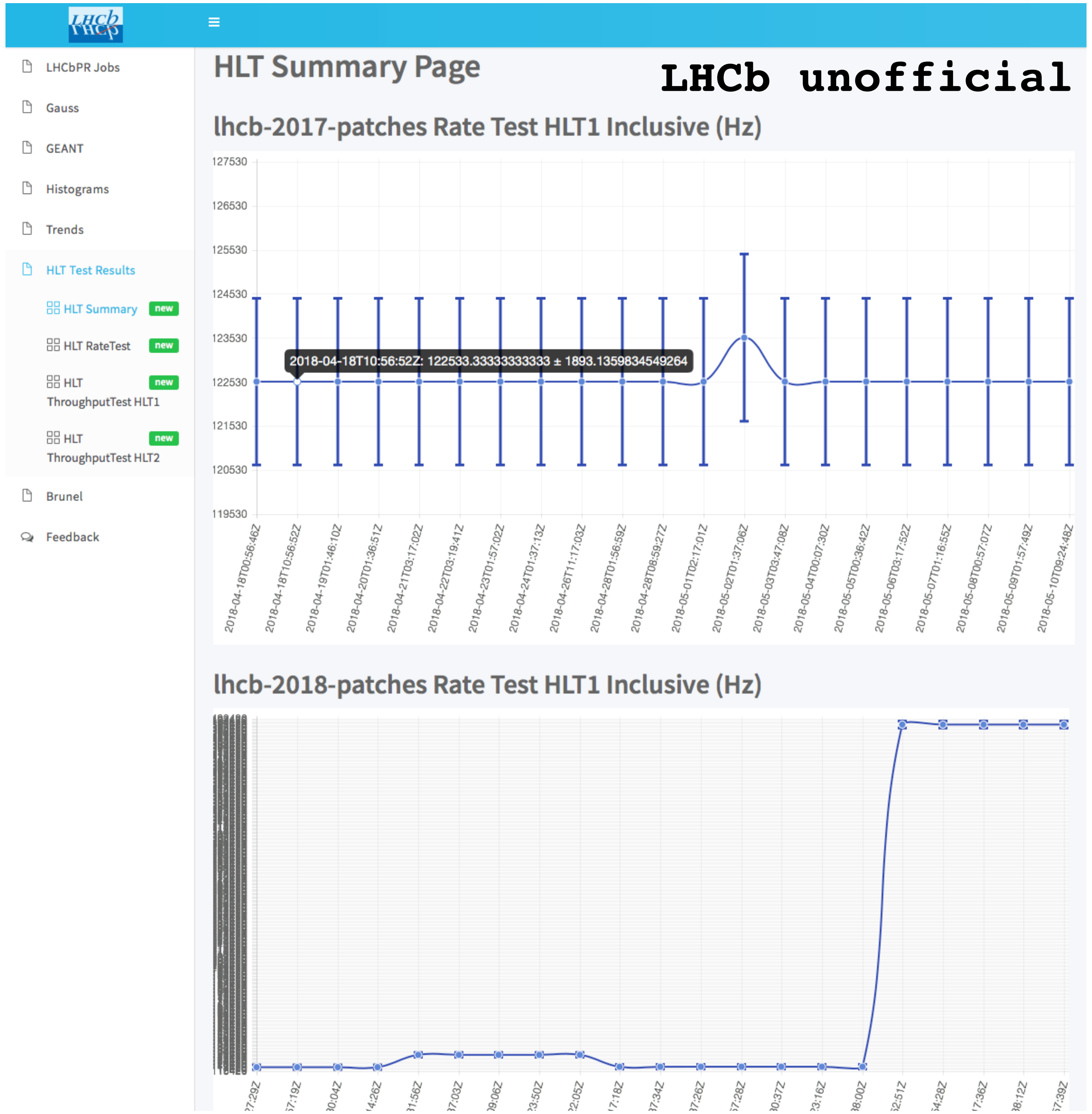
Moving to full real-time analysis, need to select many more signal channels, add background and control channel selections → 100s of selections!

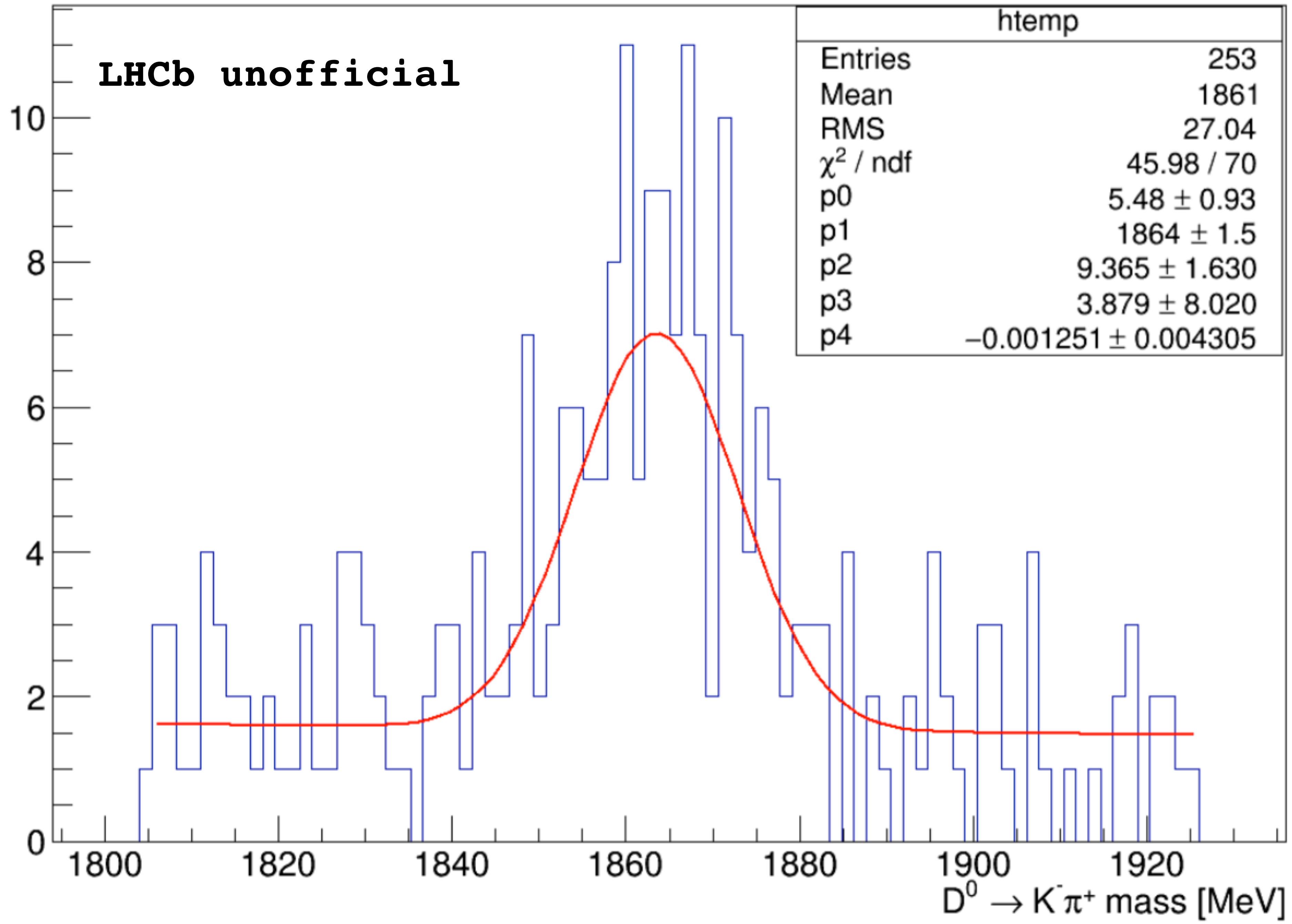
Controlling the timing of making charged particle combinations becomes crucial, enabled by having offline-quality particle identification information at start of HLT2



Basic ability already demonstrated in Run I, PID key to timing

Monitoring and software validation





For 2015 ad-hoc, git and software validation crucial later

À la recherche du temps reel: optimizing the cascade buffers

What is a cascade buffer?

Bigger data volume

Reconstruct high P_T leptons

Reconstruct pp vertices & select displaced leptons

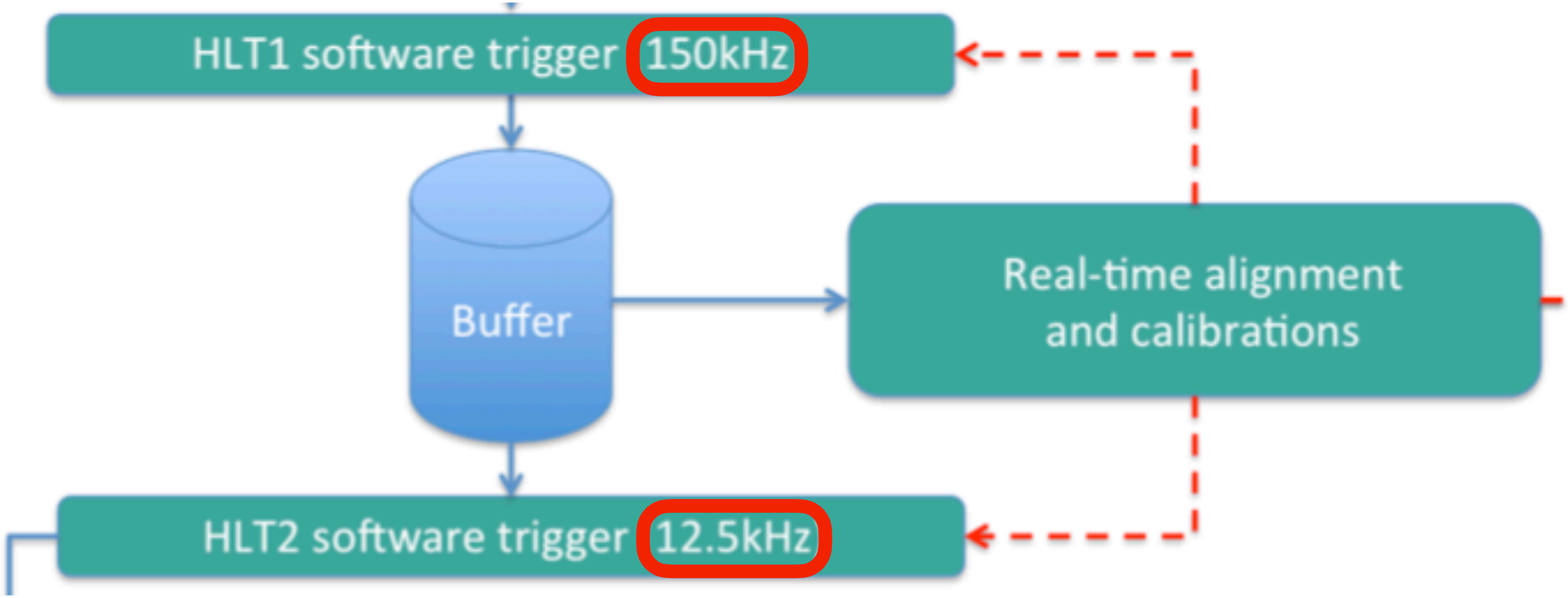
Reconstruct other charged particles & build B candidate

Build particle identification information & purify selection

More complex processing

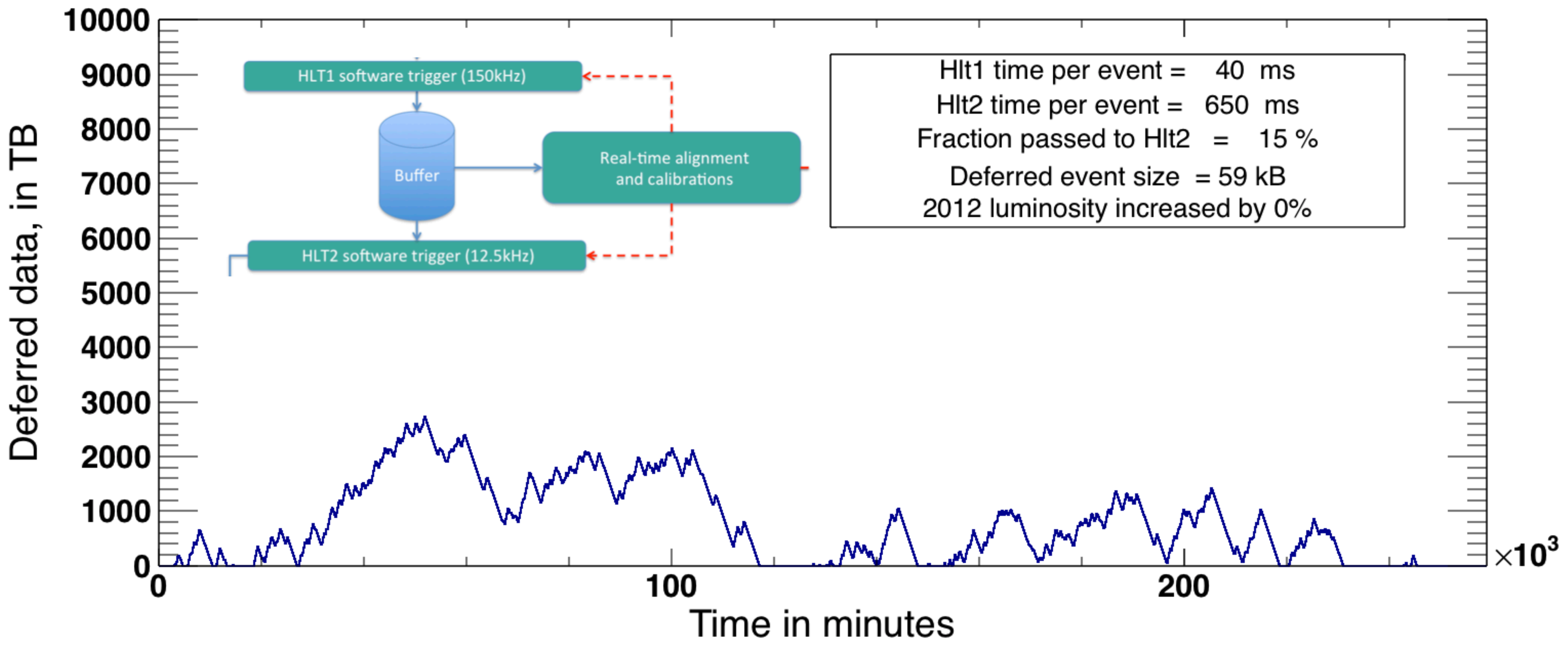
A staged data reduction using increasingly complex algorithms

What cascade did we optimize?



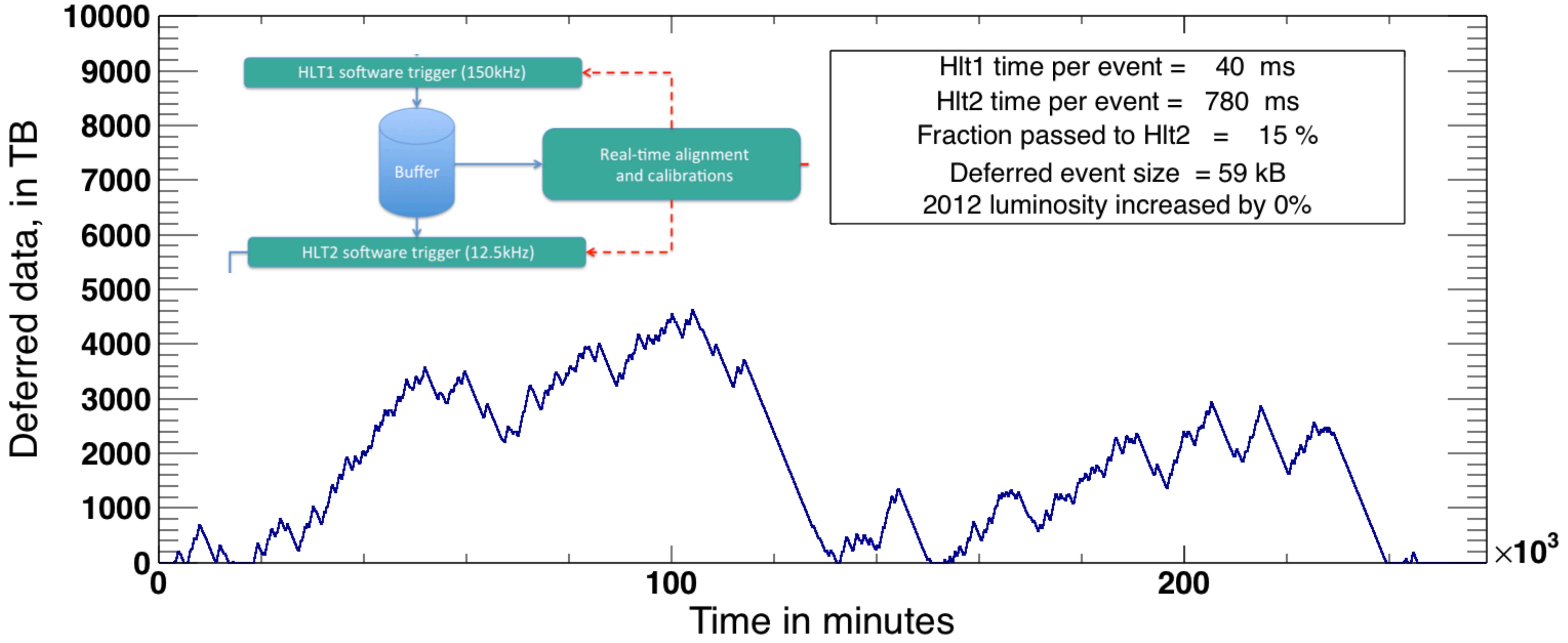
Balance retention of HLT1 against processing time of HLT2

Optimization of the Run 2 LHCb cascade buffer



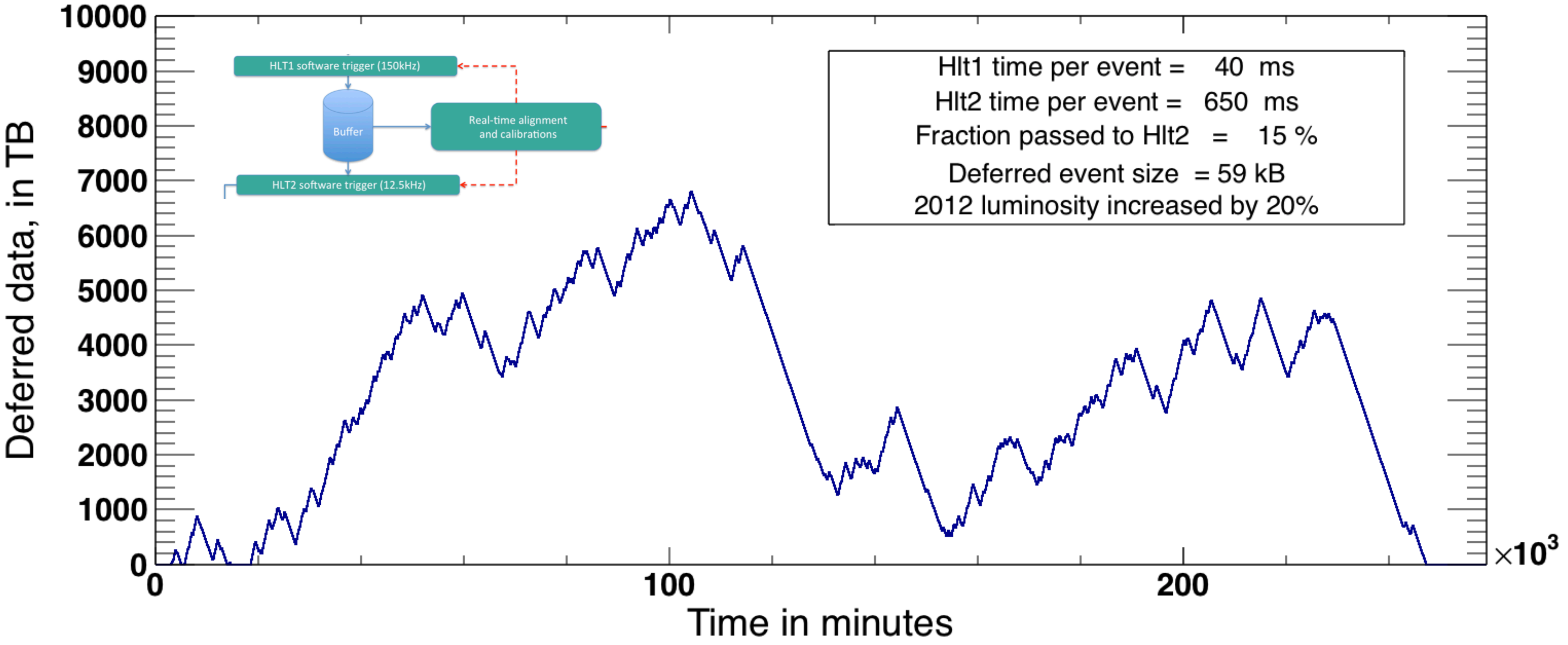
Use Run I LHC fill structure to simulate disk buffer usage

Optimization of the Run 2 LHCb cascade buffer



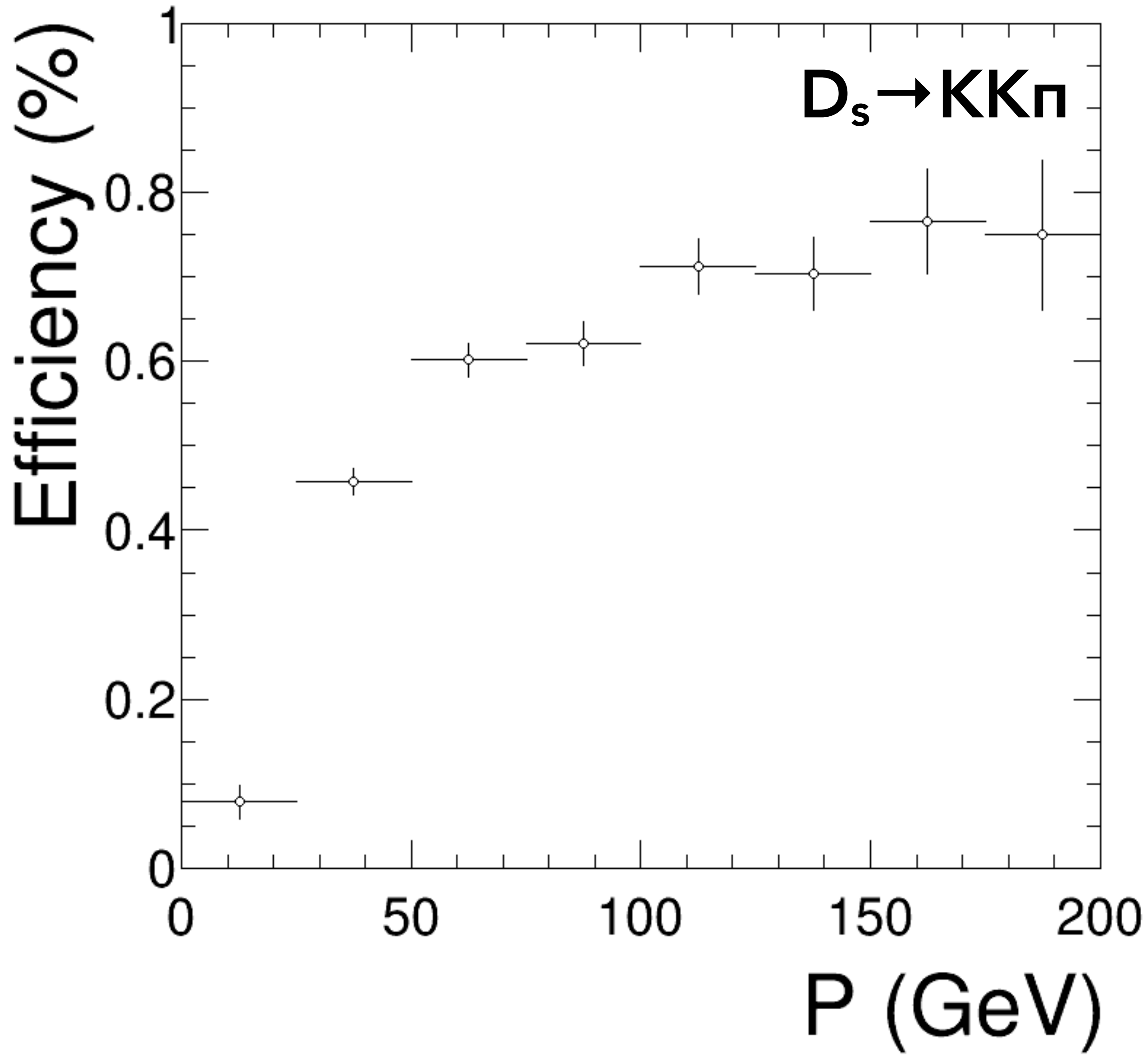
Use simulation to ensure robustness if timing estimates wrong

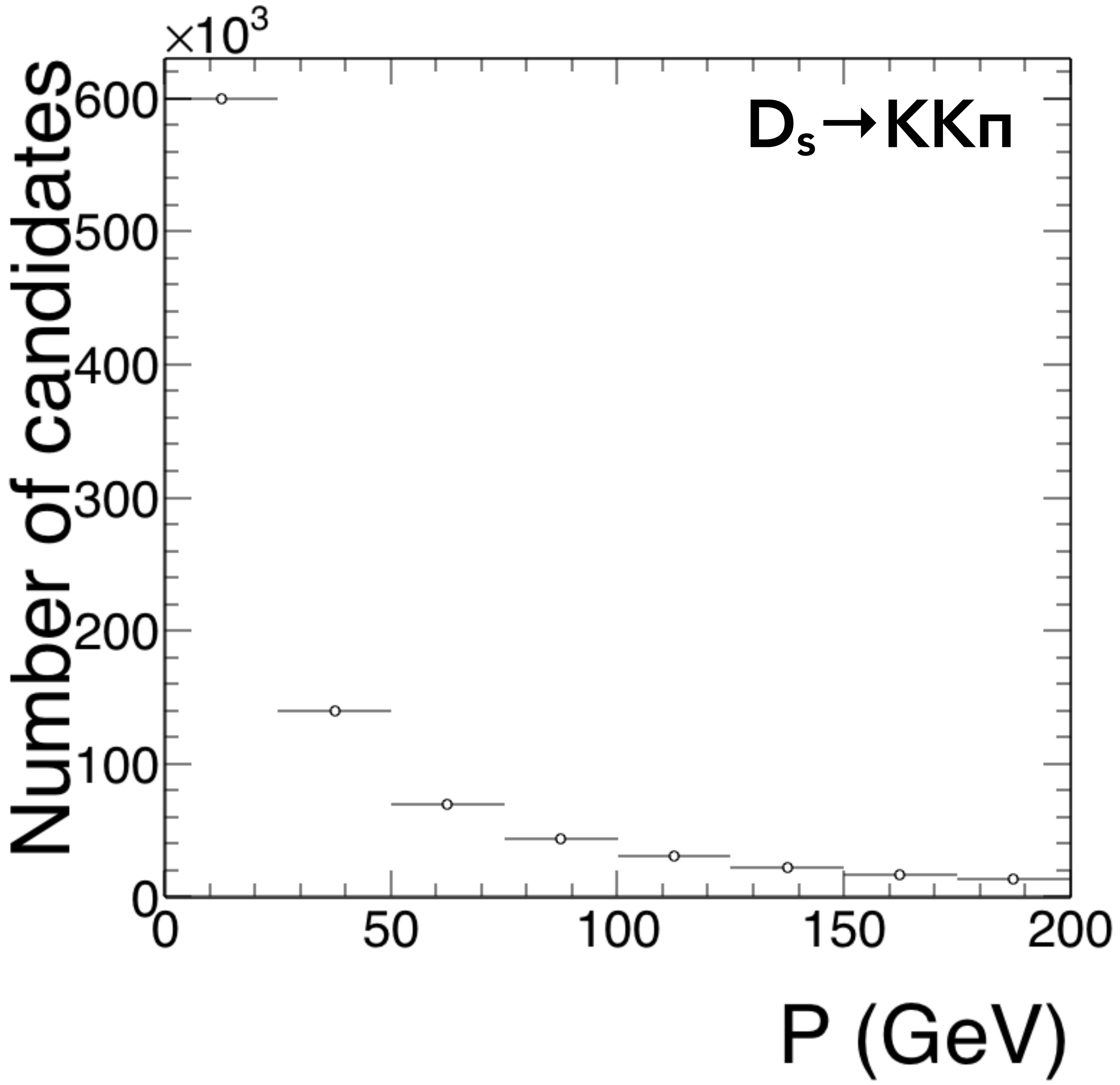
Optimization of the Run 2 LHCb cascade buffer



Use simulation to ensure robustness if LHC overperformed

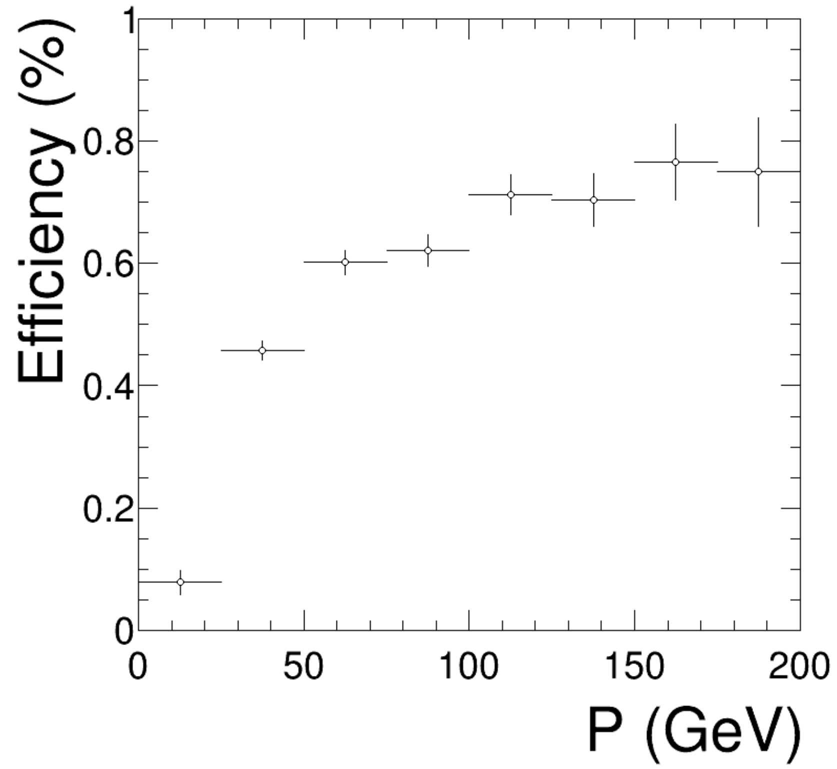
Factorizing the LHCb reconstruction

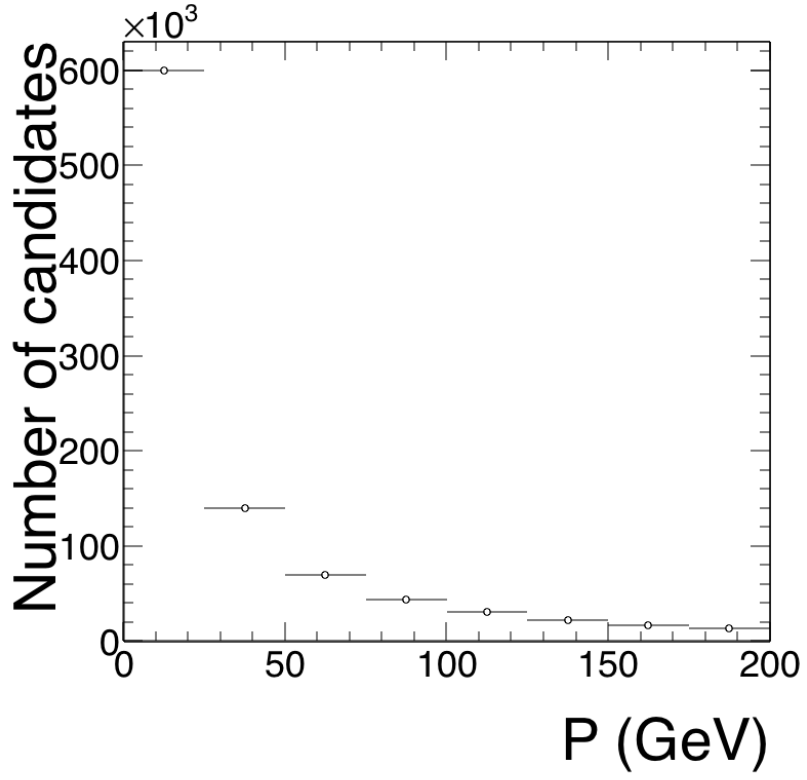




Remember, LHCb does not work on an efficiency plateau

Factorizing the LHCb reconstruction





Key objective : make it possible for HLT2 to run the full offline reconstruction

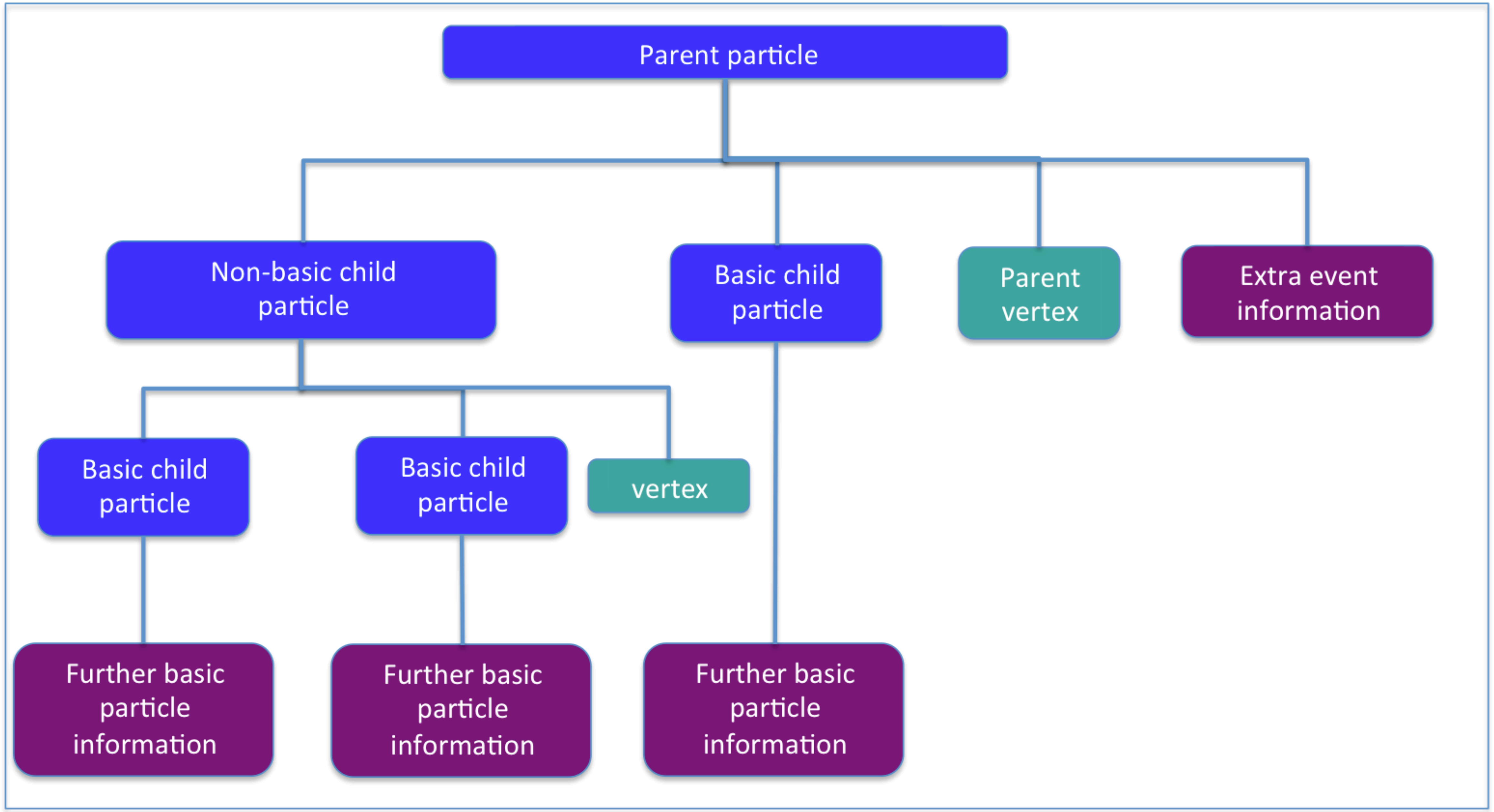
However for precision physics we do not work on an efficiency plateau : must understand in detail efficiency of HLT1 with respect to HLT2

Reoptimize tracking sequence so that HLT1 almost perfectly selects a high-momentum subset of tracks found by HLT2. This factorization of the tracking minimizes systematic uncertainties without losing absolute performance.

Crucial to ensure HLT1 "fast" reconstruction a subset of HLT2

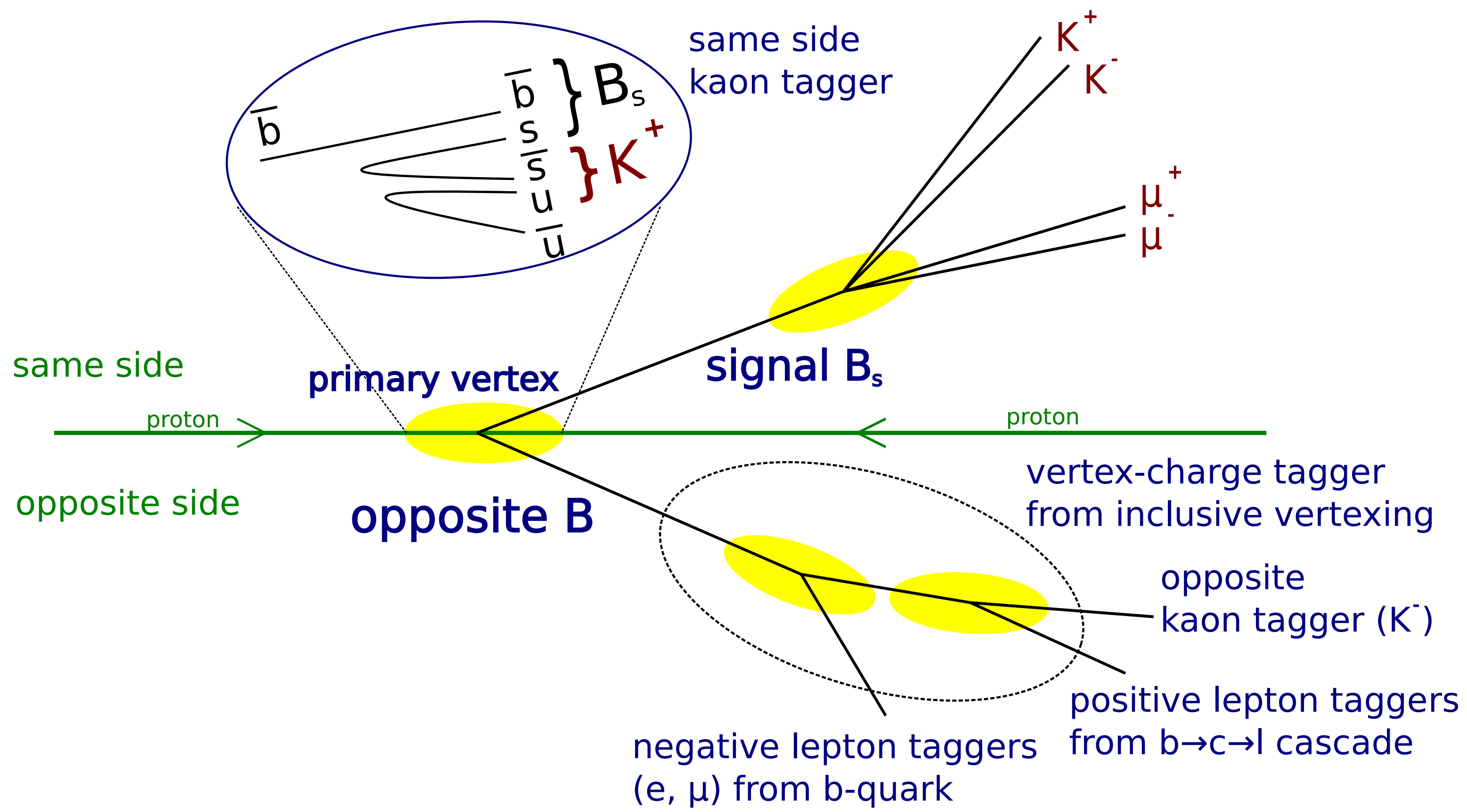
Il nous faut
une procédure:
persisting and
validating the data

Persisting analysis-quality data in real time



Enabled trigger to persist data in analysis format (huge job)

What exactly is a signal? The role of "event" information



Information about particles which are not directly part of the signal but allow us to infer some information about it, for example

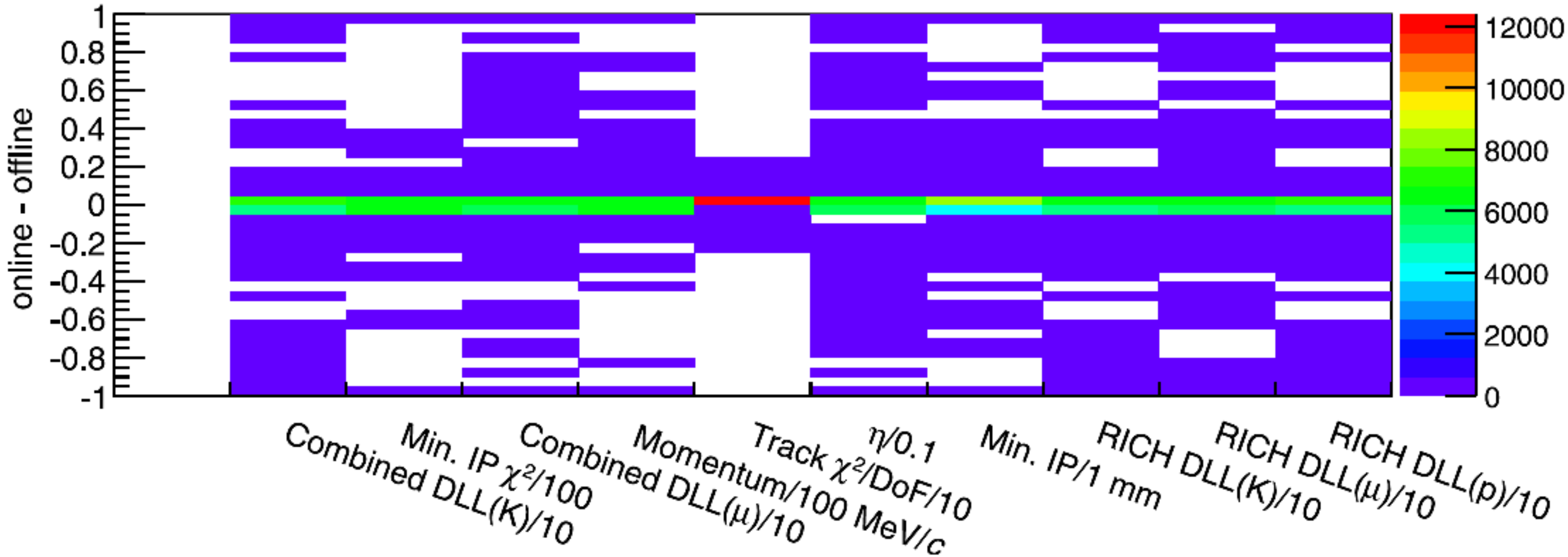
Isolation : how likely is this real signal and not combinatorial background?

Flavour tagging : infer production flavour of signal based on particles produced in same quark-fragmentation chain

Spectroscopy : search for excited states by combining Cabibbo-favoured beauty/charm decays with tracks from same pp interaction

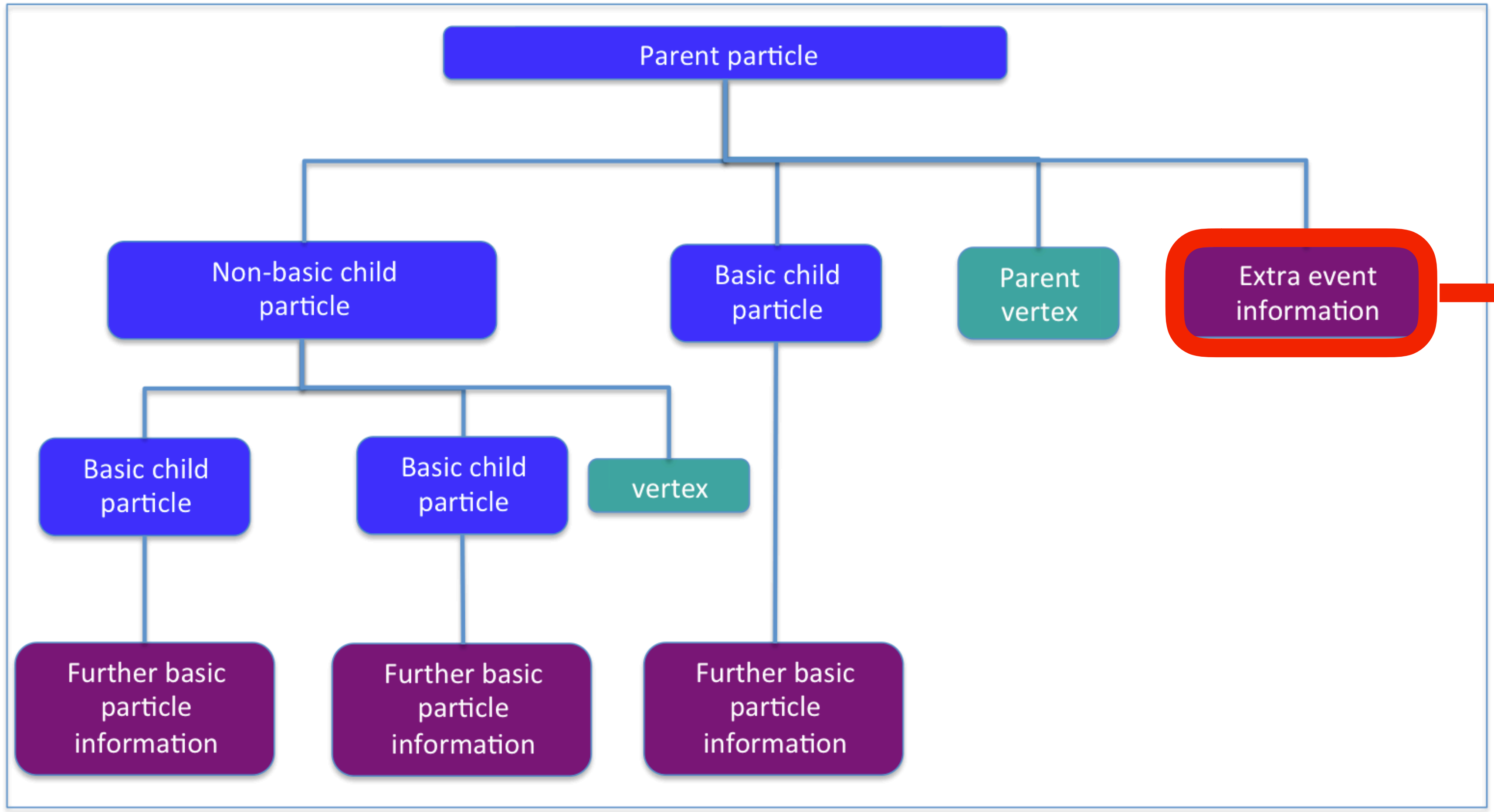
Vital for spectroscopy, searches (isolation), flavour tagging

Commissioning and validating the Run 2 analysis data



A lot of work comparing trigger-level and offline variables…

Evolution of data persistence in Run 2 and the LHCb upgrade



Today : mainly additional particles in event, ability to associate isolation variables to the signal

Tomorrow : any subset of other reconstructed objects or raw detector data required for a given analysis

Necessary for the upgrade where most analyses should be done in real time, takes pileup suppression aspect to its logical conclusion.

Goal : enable each analysis to save its own custom "event"

It is not all relative: measuring σ_{cc} in real time

Why measure charm cross-sections?

Pragmatic :

Proper validation of new analysis model requires full review process and publication

New collider energy, cross-sections immediately publishable

In addition, absolute cross-sections extremely sensitive to control of detector effects, validate all aspects of calibration

Physics :

Validate MC generator tunings and QCD hadronization models

Understand production of dominant background to rare Higgs/EW boson decays

Constrain production of high-energy atmospheric neutrinos from cosmic-ray induced charm hadron production

Marriage of pragmatic and physics motivation

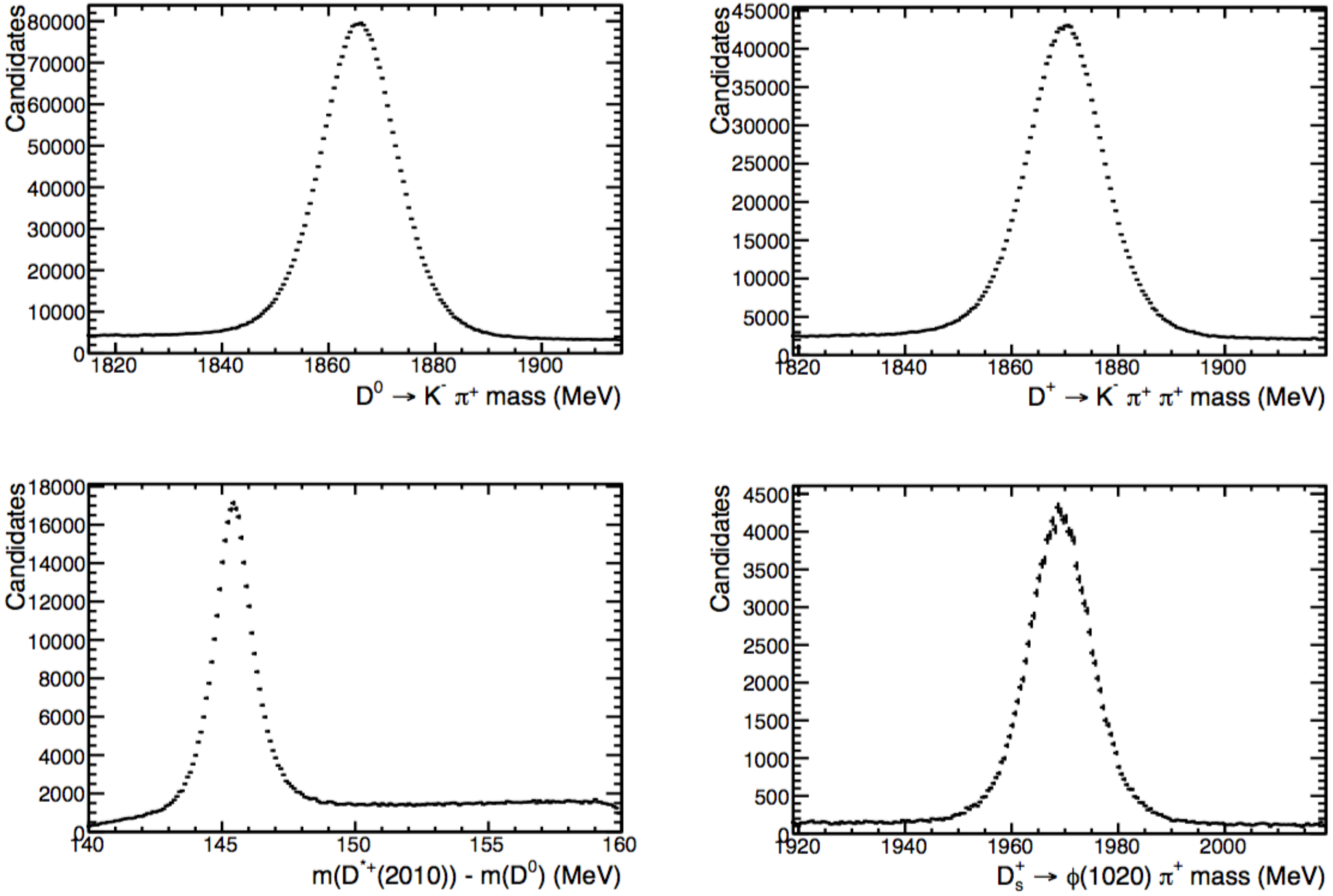
Signal selection

Random selection at hardware trigger level to avoid having to understand calorimeter

Single displaced particle at HLT1

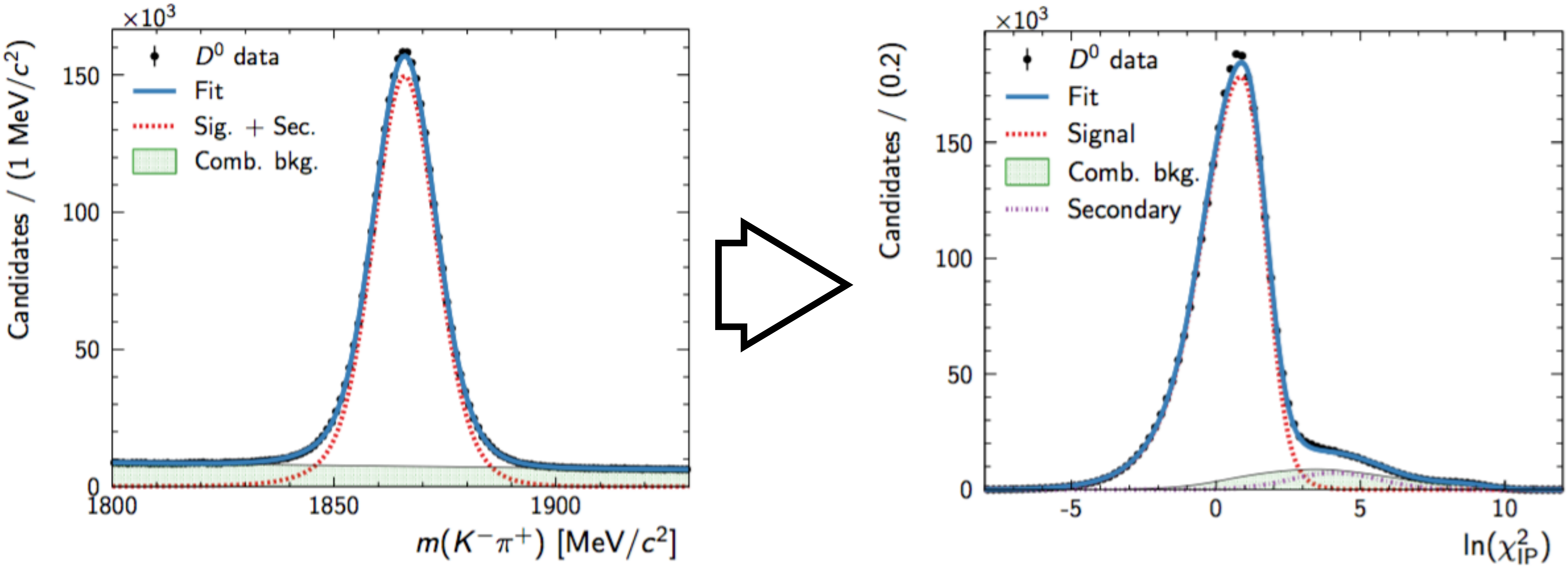
Full signal reconstruction at HLT2

Only measured D^0, D^+, D_s, and D^{*+} for the first paper, left baryons and other excited charm hadrons for future papers



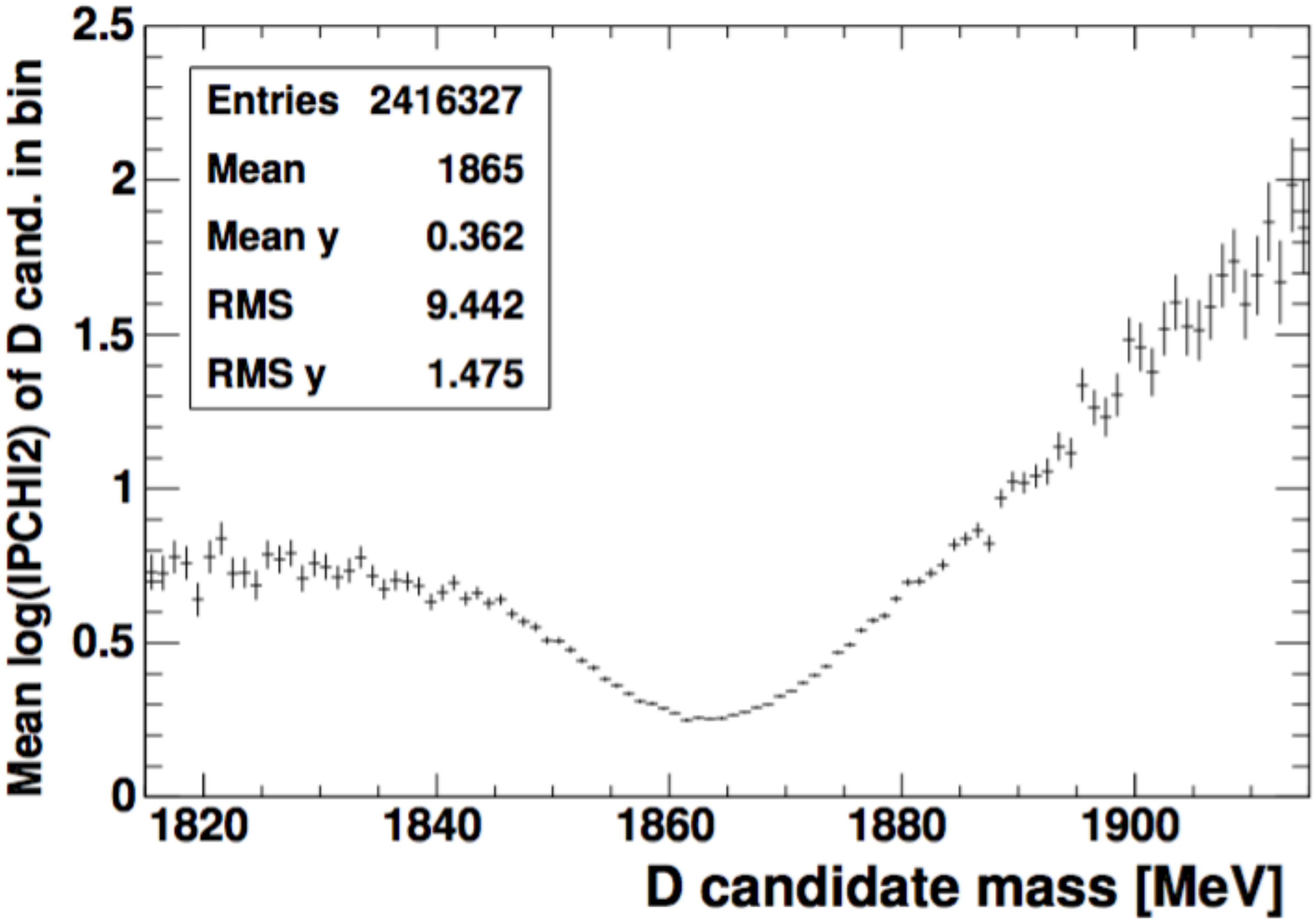
Systematics limited so no BDTs, just kinematics&displacement

Yield measurement



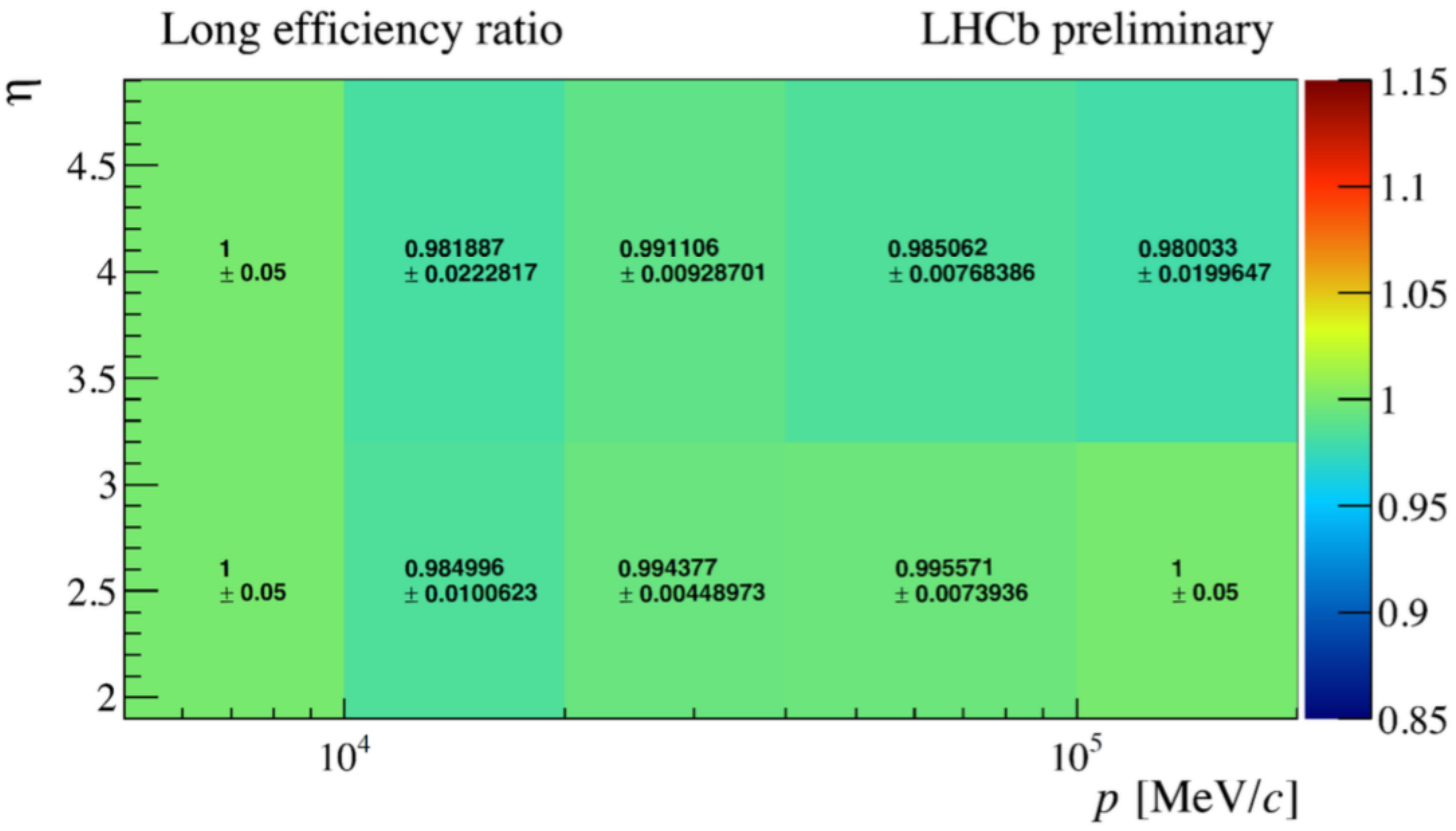
Two stage fit to D mass and impact parameter χ^2

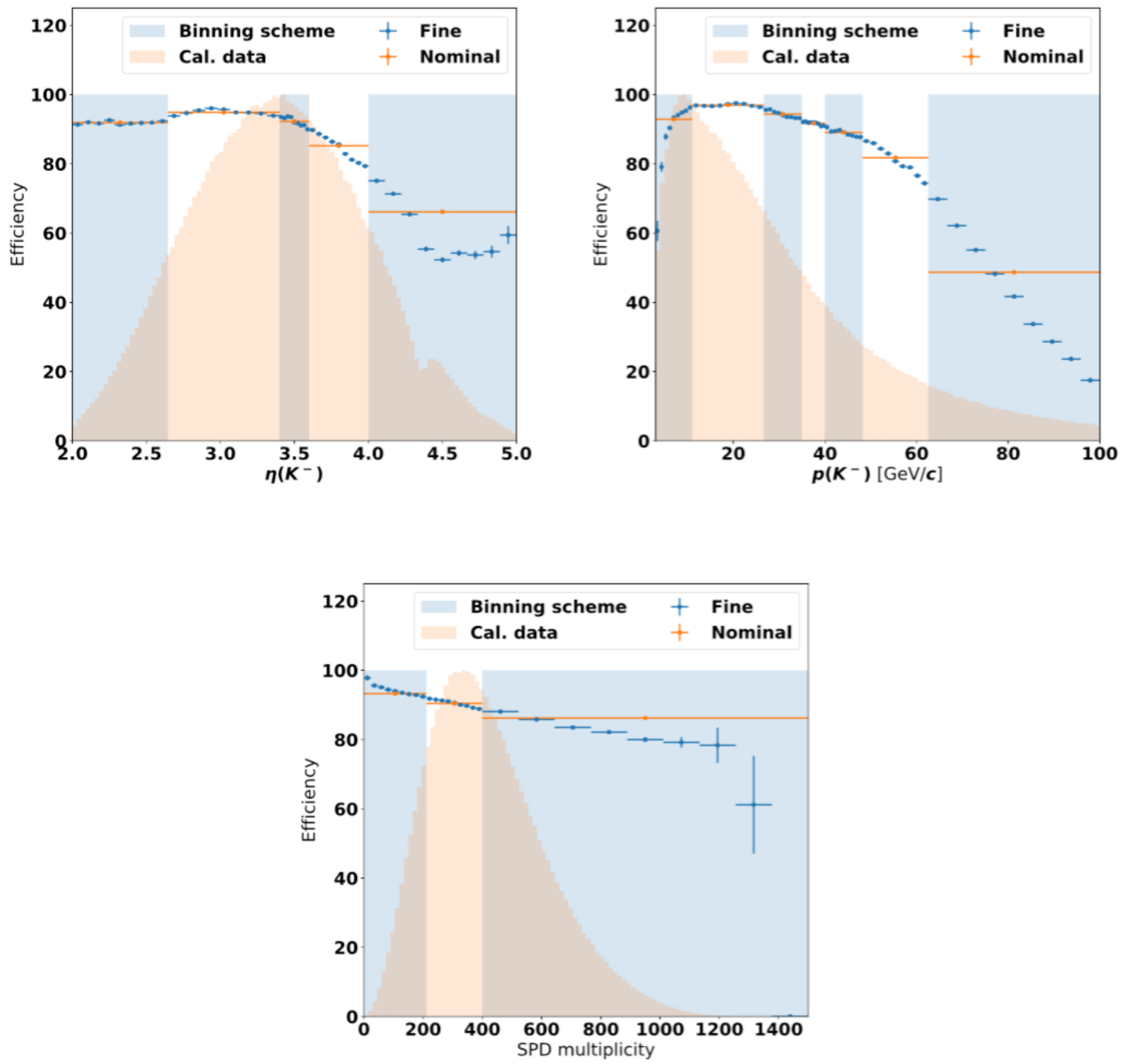
Why a two stage fit?



Because mass & impact parameter χ^2 are correlated for signal

Efficiency correction





Realistic trigger-level data-driven calibrations taken in parallel with the signals.

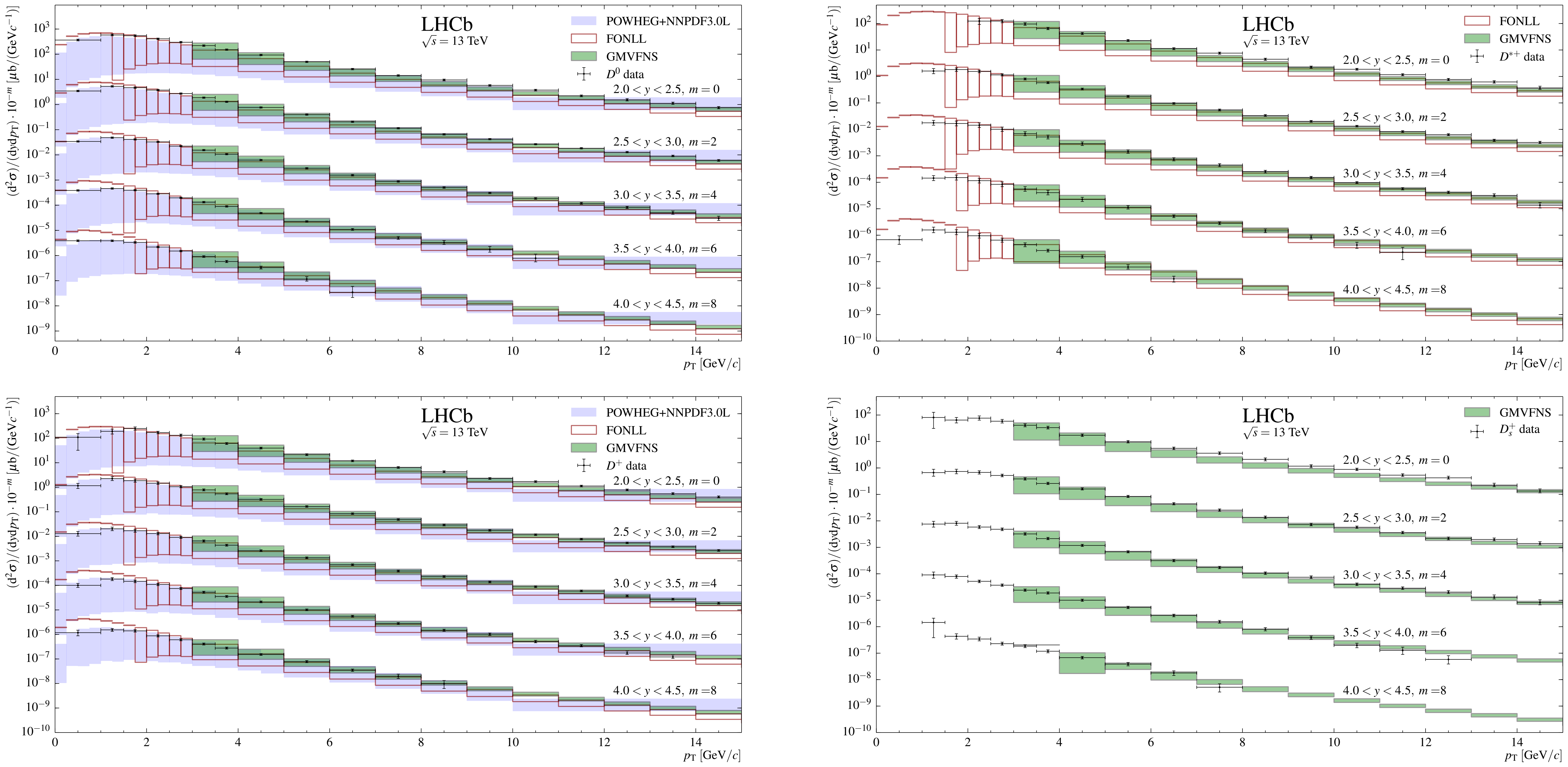
Kinematics from MC, PID and reconstruction corrected w/calib

Systematics

	Uncertainties (%)				Correlations	
	D^0	D^+	D_s^+	D^{*+}	bins	modes
MC stat.	1-26	1-39	1-55	1-23	0	0
MC modelling	1	1	0.2	0.9	0	0
Fit model	1-6	1-5	1-2	1-2	0	0
Tracking	3-10	3-14	4-14	5-11	90-100	90-100
PID cal.	0-2	0-1	0-2	0-1	0-100	0-100
PID binning	0-44	0-10	0-20	0-15	100	100
$\mathcal{BR}$	1.2	2.1	5.8	1.5	100	0-95
Luminosity	3.9				100	100

Dominated by luminosity and tracking efficiency systematic

Results and discussion



Main result, double differential cross-sections

LHCb repent! : errors in real-time analyses and their implications

From real-time analysis to delayed errata

Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV

[to restricted-access page]

INFORMATION

LHCB-PAPER-2015-041

PH-EP-2015-272

ARXIV:1510.01707 [PDF]

(SUBMITTED ON 06 OCT 2015)

JHEP03(2016)159 JHEP09(2016)013 JHEP05(2017)074

INSPIRE 1396331

ERRATA
JHEP 09 (2016) 013
JHEP 05 (2017) 074

TOOLS

CITED 111 TIMES

GET BIBTEX

HEPDATA

CONTACT

EDITORIAL BOARD EMAIL

Abstract

Production cross-sections of prompt charm mesons are measured with the first data from pp collisions at the LHC at a centre-of-mass energy of $13\,\mathrm{TeV}$. The data sample corresponds to an integrated luminosity of $4.98 \pm 0.19\,\mathrm{pb}^{-1}$ collected by the LHCb experiment. The production cross-sections of D^0, D^+, D_s^+, and D^{*+} mesons are measured in bins of charm meson transverse momentum, p_T, and rapidity, y, and cover the range $0 < p_\mathrm{T} < 15\,\mathrm{GeV}/c$ and $2.0 < y < 4.5$. The inclusive cross-sections for the four mesons, including charge conjugation, within the range of $1 < p_\mathrm{T} < 8\,\mathrm{GeV}/c$ are found to be

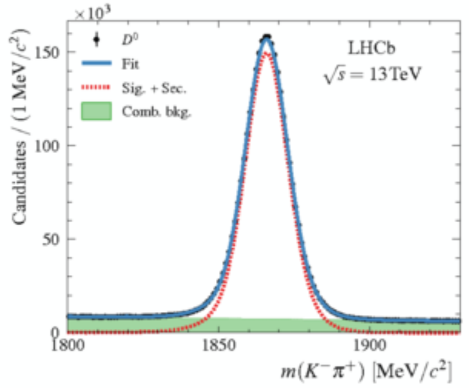
$$\sigma(pp \to D^0 X) = 2072 \pm 2 \pm 124\,\mu\mathrm{b}$$
$$\sigma(pp \to D^+ X) = 834 \pm 2 \pm 78\,\mu\mathrm{b}$$
$$\sigma(pp \to D_s^+ X) = 353 \pm 9 \pm 76\,\mu\mathrm{b}$$
$$\sigma(pp \to D^{*+} X) = 784 \pm 4 \pm 87\,\mu\mathrm{b}$$

where the uncertainties are due to statistical and systematic uncertainties, respectively.

Figures and captions

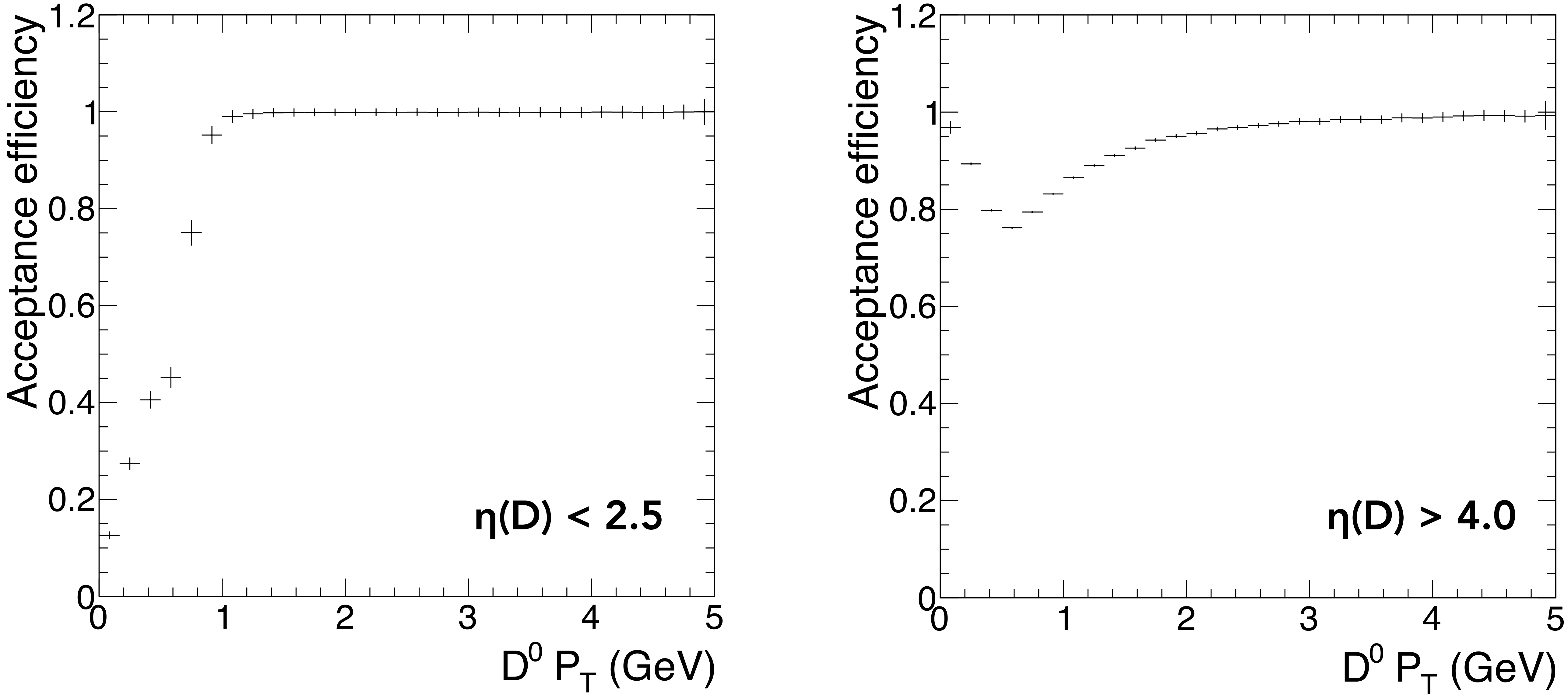
Distributions for selected $D^0 \to K^-\pi^+$ candidates: (left) $K^-\pi^+$ invariant mass and (right) $\ln \chi^2_{\mathrm{IP}}$ for a mass window of $\pm$\SI20\mevcc around the nominal D^0 mass. The sum of the simultaneous likelihood fits in each (p_T, y) bin is shown, with components as indicated in the legends.

Fig1a.pdf [60 KiB]
HiDef png [298 KiB]
Thumbnail [70 KiB]



Ironically nothing to do with the real-time part...

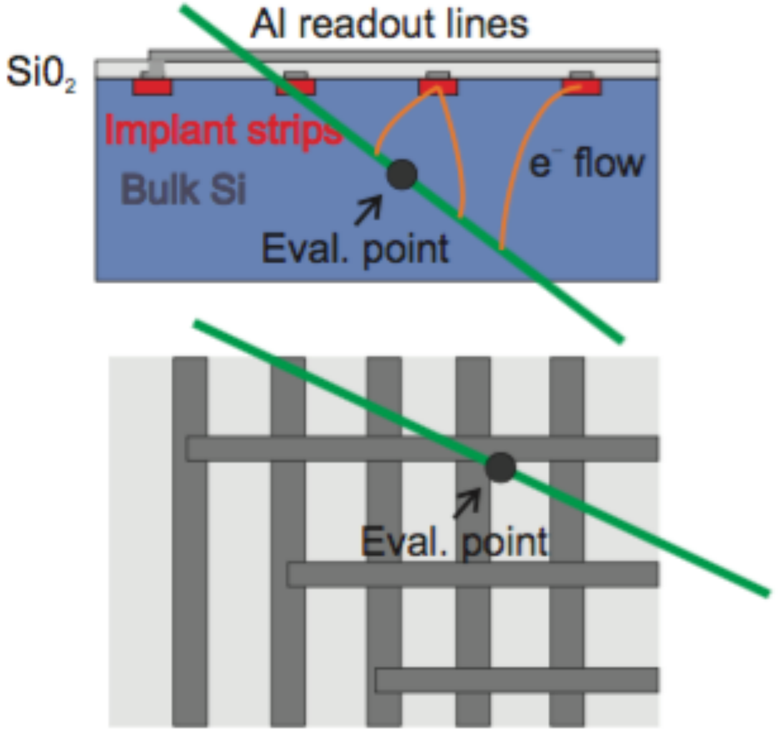
The first, specific, erratum

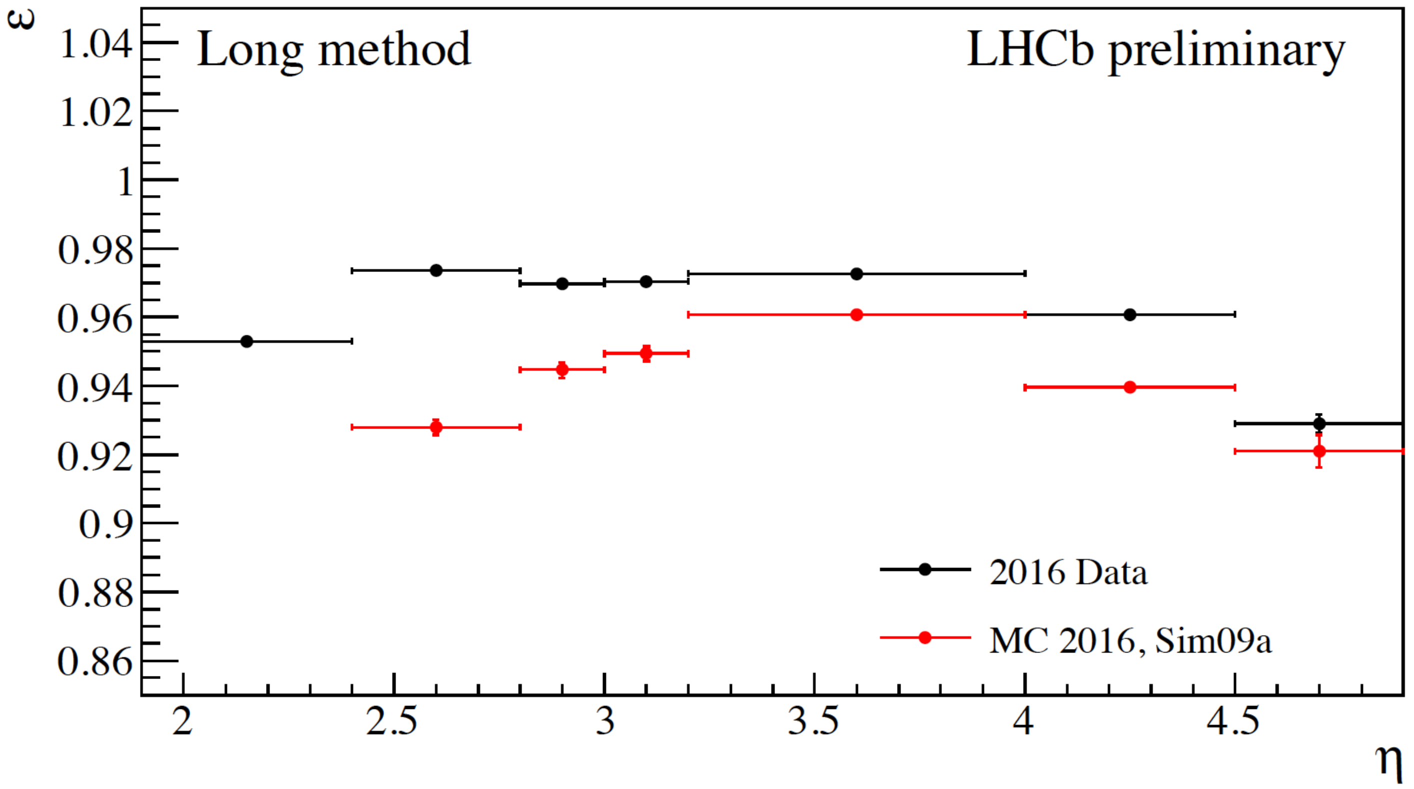


Required soft pion from $D^{*+} \to D^0\pi^+$ decay chain to be in acceptance for D^0 cross-section

Straightforward bug in acceptance calculation

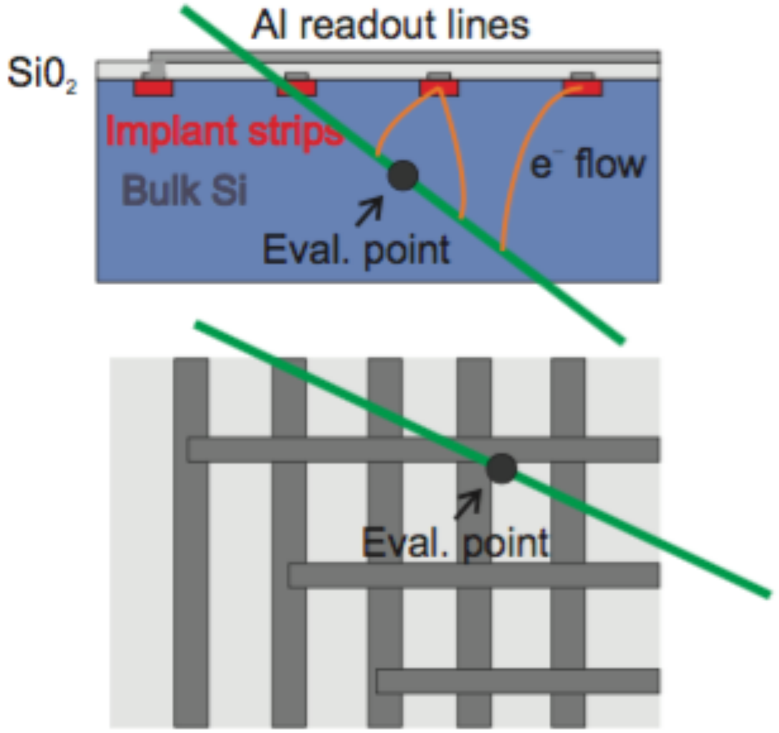
The second, general, erratum

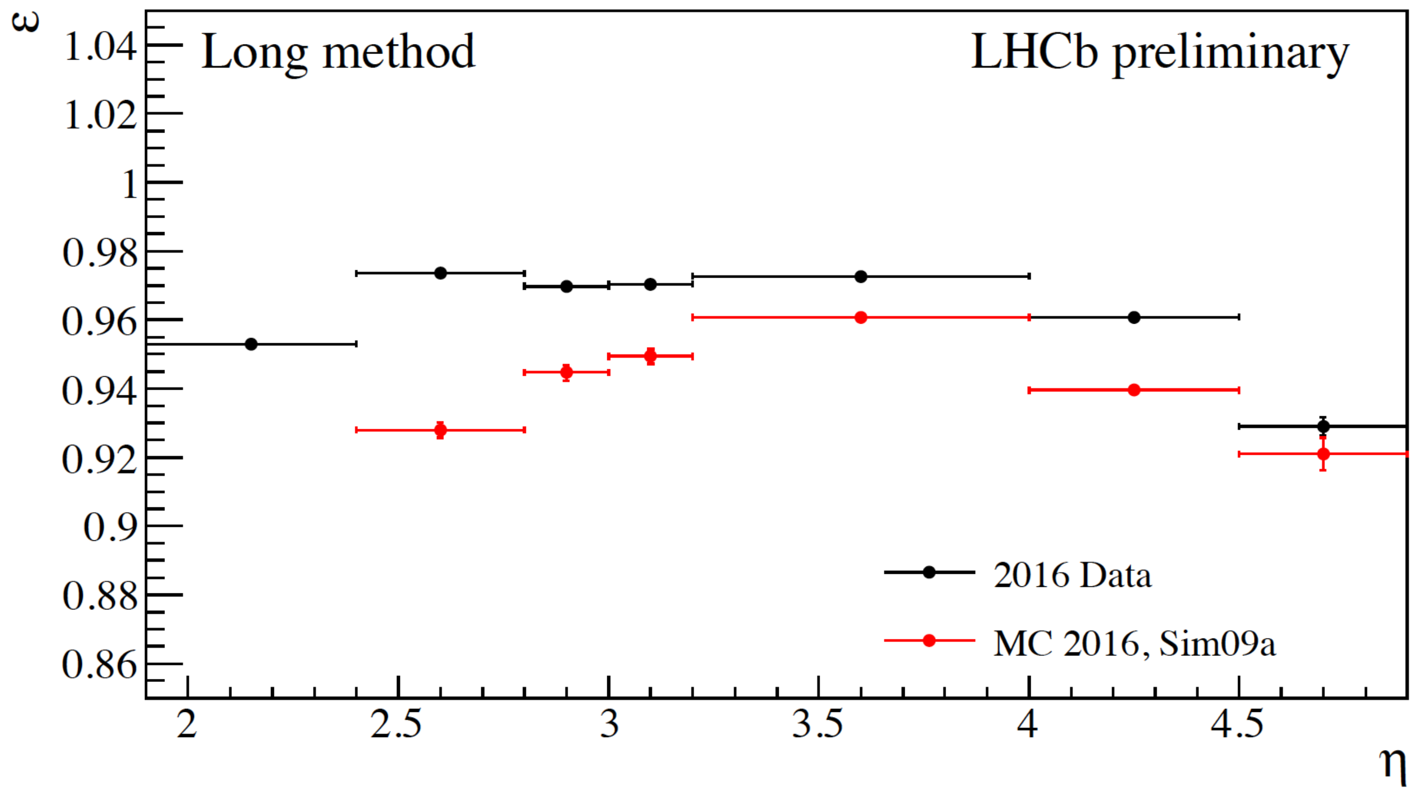




Mismodelled radiation damage impact on VELO hit efficiency

The second, general, erratum

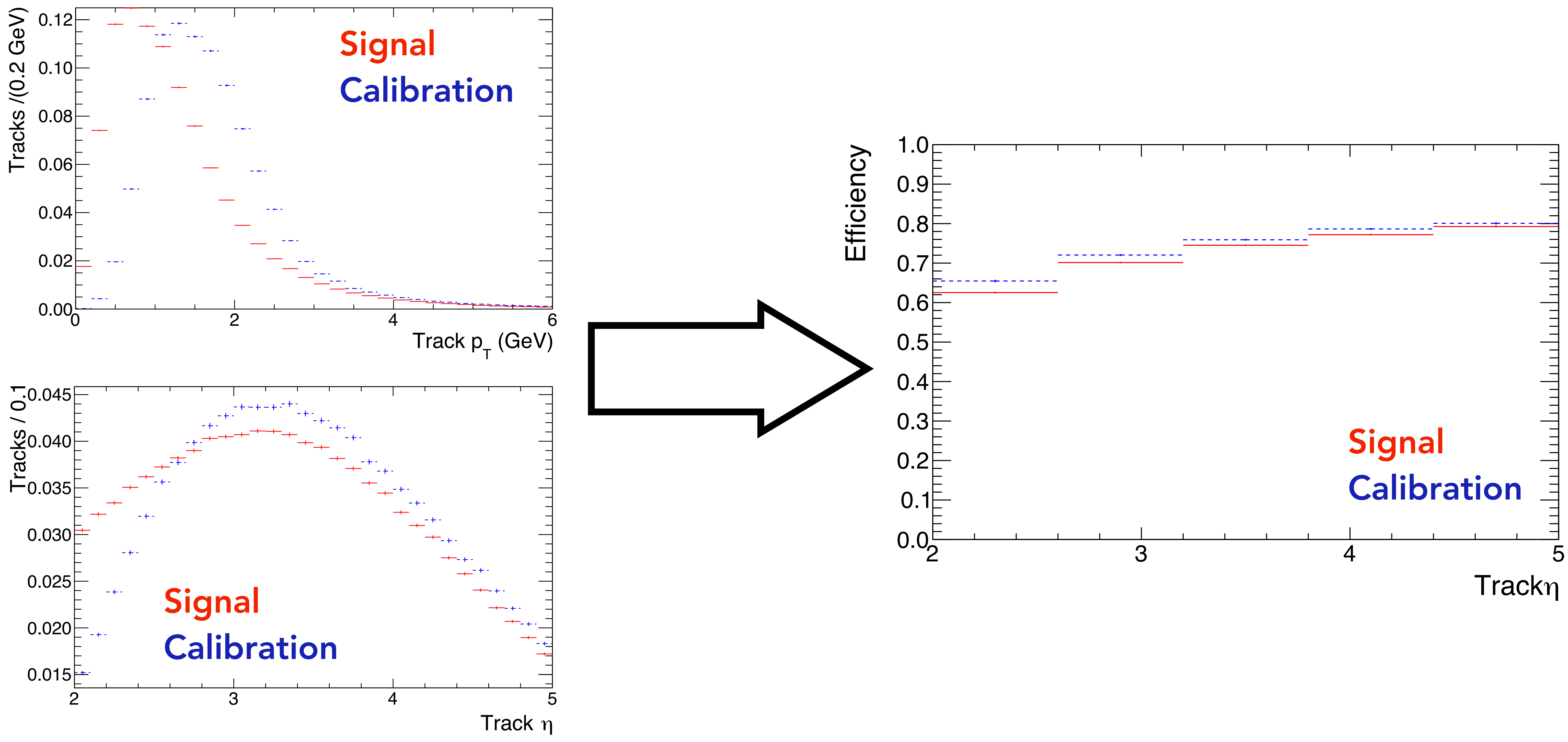




We should have just regenerated the MC but we thought nah, we can calibrate this away, save the computing resources… as Jim Libby once taught me (paraphrasing somewhat), don’t try to be clever if you can brute-force a problem.

Turns out that the calibration samples did not correct for it

A more detailed look at in-bin variations



Difference in signal/calibration kinematic variation within bins

Lessons for the future

Not as many as you might think

Was this an embarassing episode? Definitely.

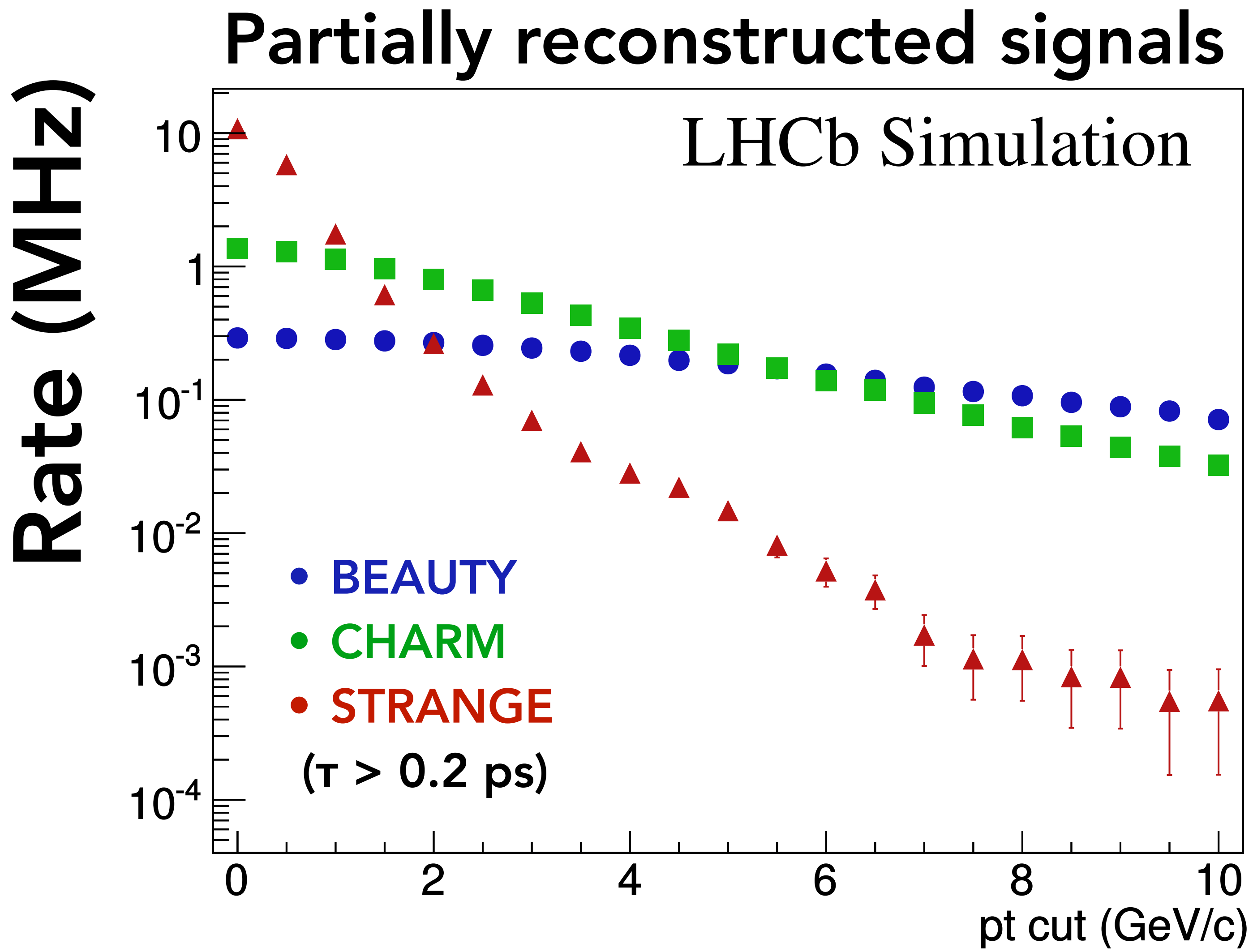
Do I wish I had insisted to redo the analyses with the corrected simulation instead of relying on calibration tables? Sure, although easy to say that now.

But in the end the fundamental problem was that nobody realized the in-bin variation of the calibration samples was large compared to the difference in efficiencies between data and simulation until someone finally checked the results between the "corrected" and "calibrated" simulation and found a discrepancy. We all learned something there.

Keep making calibrations even more fine grained…

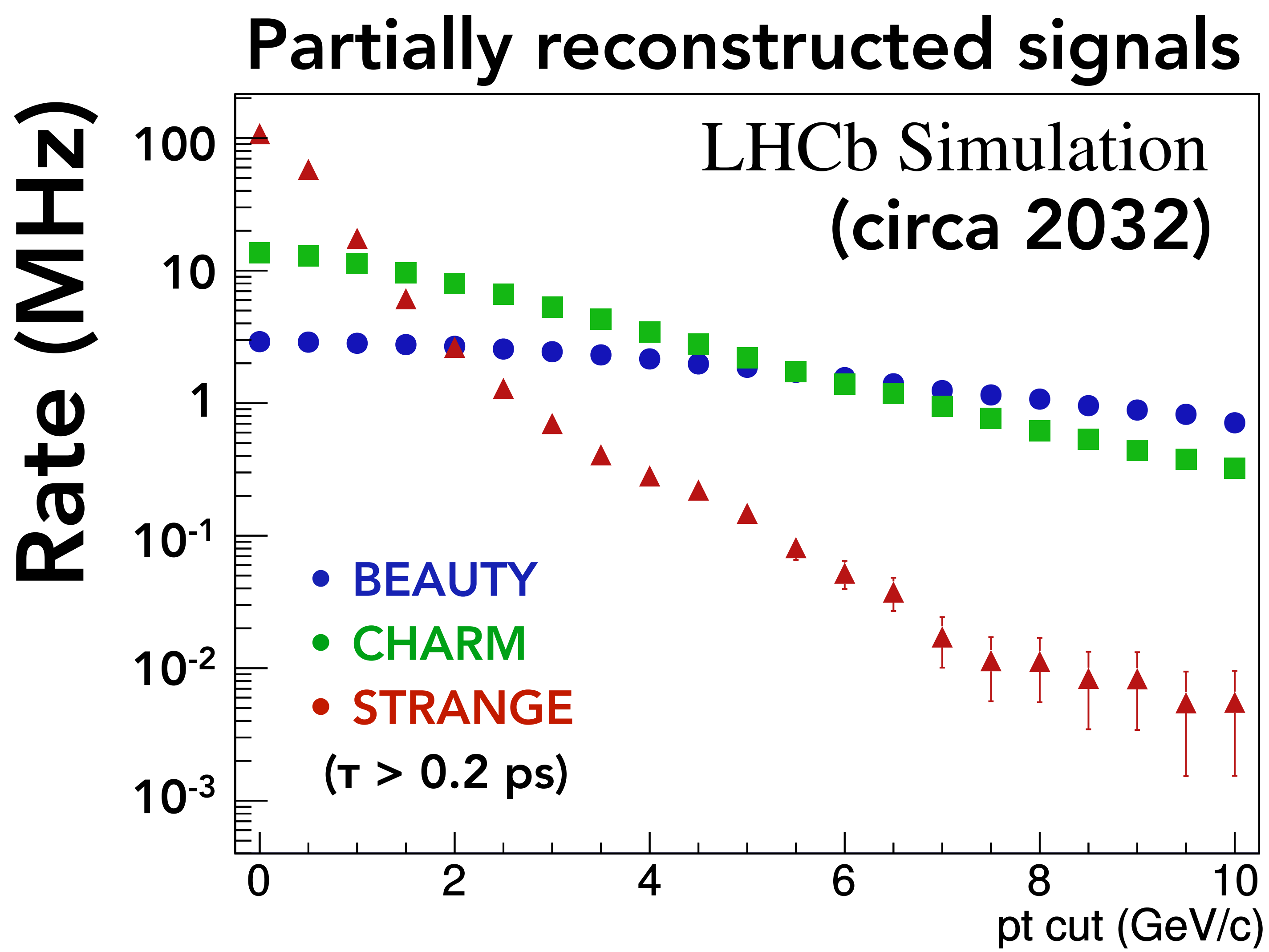
Staring at the sun: the future of real-time analysis

Evolution of real-time analysis towards the LHCb upgrade...



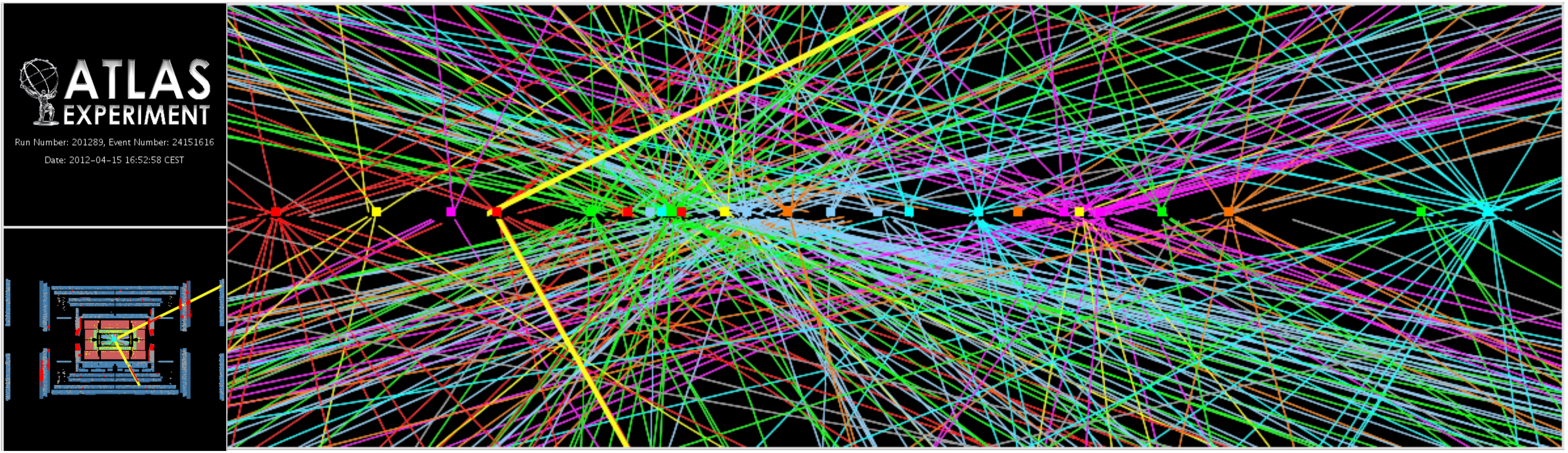
Still just about room for a first level selective trigger

...and a potential second upgrade



But at $2 \cdot 10^{34}$, even that will no longer be possible

Why is real-time analysis here to stay?



Almost all bunch crossings will contain interesting signal, most proton-proton collisions will not

➡ Our triggers should select collisions, not bunch crossings

Requires ~offline-quality real-time reconstruction, detector alignment&calibration

Requires access to “rest of event” information (tagging, isolation…) in real-time

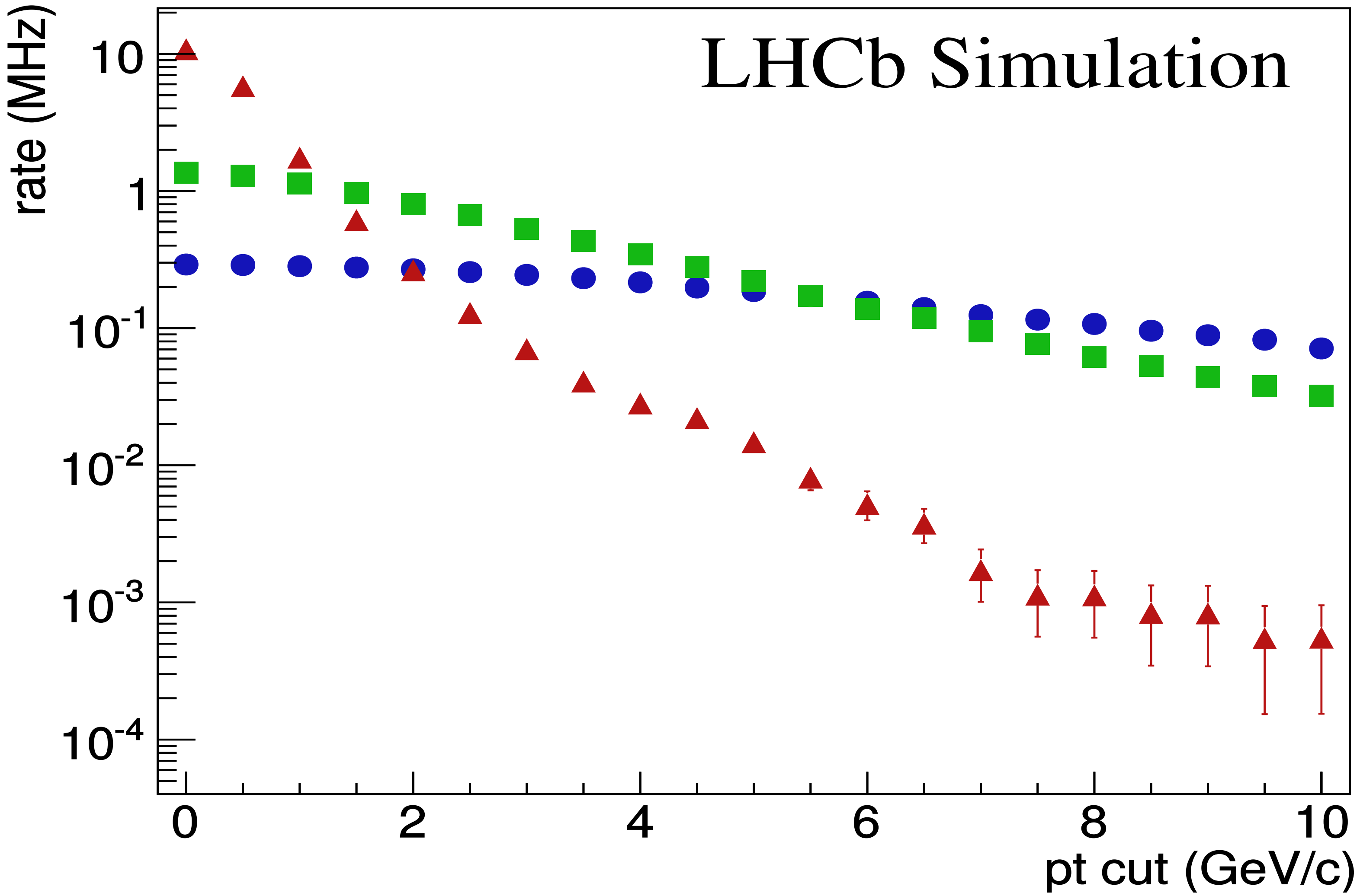
Fundamentally because it is driven by physics, not technology.

Resource constraints facing real-time analysis at LHCb

CMS detector	LHC Run-2	HL-LHC Phase-2		LHCb Upgrade	LHCb Upgrade II
Peak ⟨PU⟩	60	140	200		
L1 accept rate (maximum)	100 kHz	500 kHz	750 kHz	30 MHz	30 MHz
Event Size	2.0 MB [a]	5.7 MB [b]	7.4 MB	130 kB	1.5 MB ?
Event Network throughput	1.6 Tb/s	23 Tb/s	44 Tb/s	~40 Tb/s	~500 Tb/s
Event Network buffer (60 seconds)	12 TB	171 TB	333 TB	??	??
HLT accept rate	1 kHz	5 kHz	7.5 kHz	50-100 kHz (?)	??
HLT computing power [c]	0.5 MHS06	4.5 MHS06	9.2 MHS06	??	??
Storage throughput	2.5 GB/s	31 GB/s	61 GB/s	5-10 GB/s (?)	50 GB/s ?
Storage capacity needed (1 day)	0.2 PB	2.7 PB	5.3 PB	??	??

LHCb upgrade already has to process ATLAS/CMS HL-LHC data volumes in the software trigger, on 1/10th of the budget and 5 years earlier. Feel free to ask me about Upgrade II...

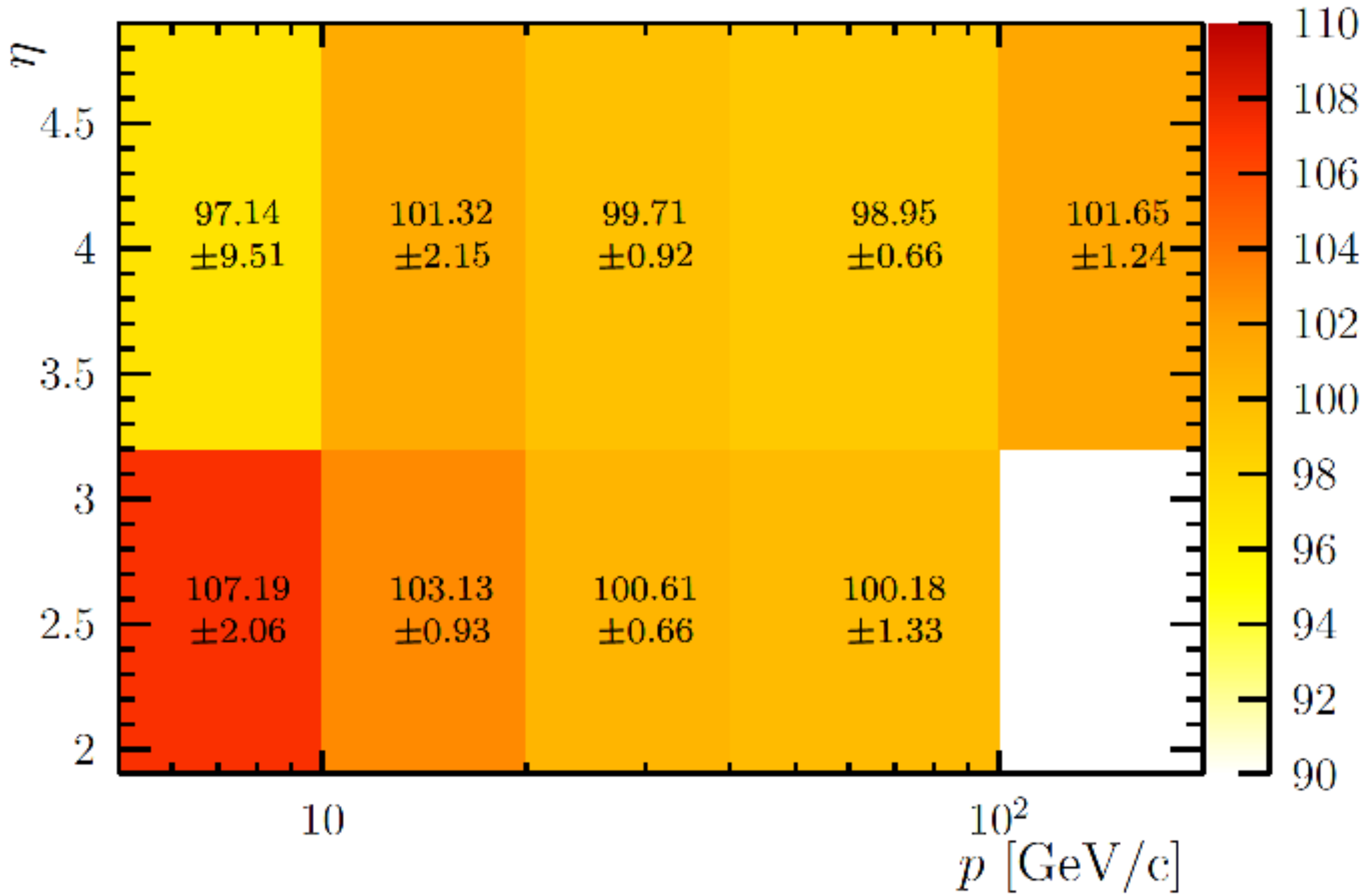
Même si le défi peut sembler insurmontable, rappelez-vous…



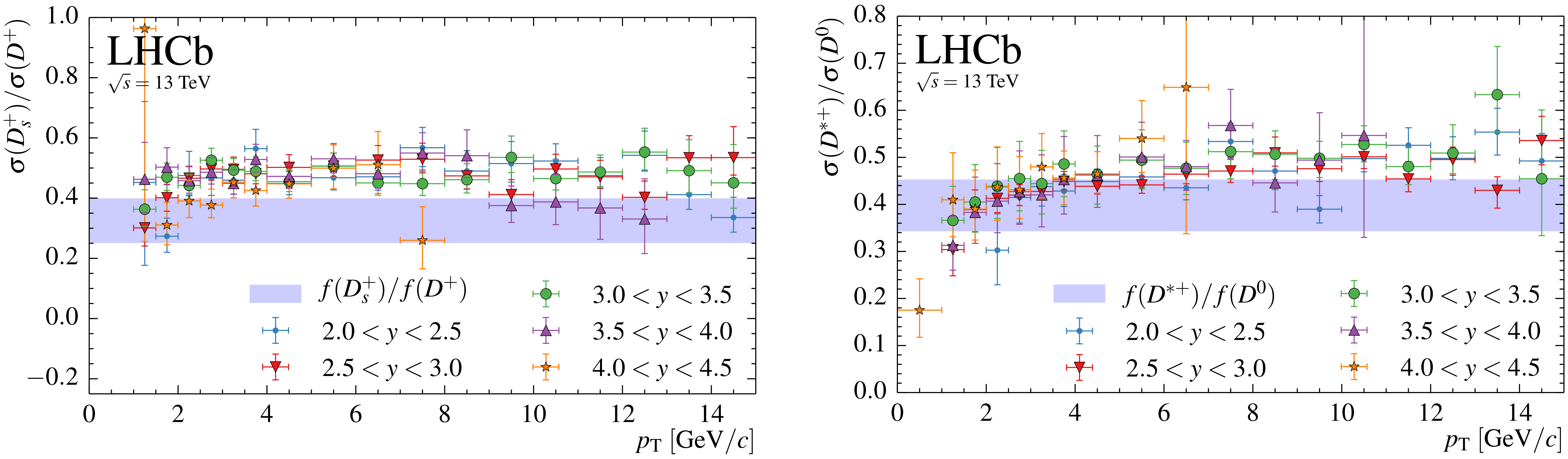
…et sepultus resurrexit : certum est, quia impossibile – Tertullian

More data

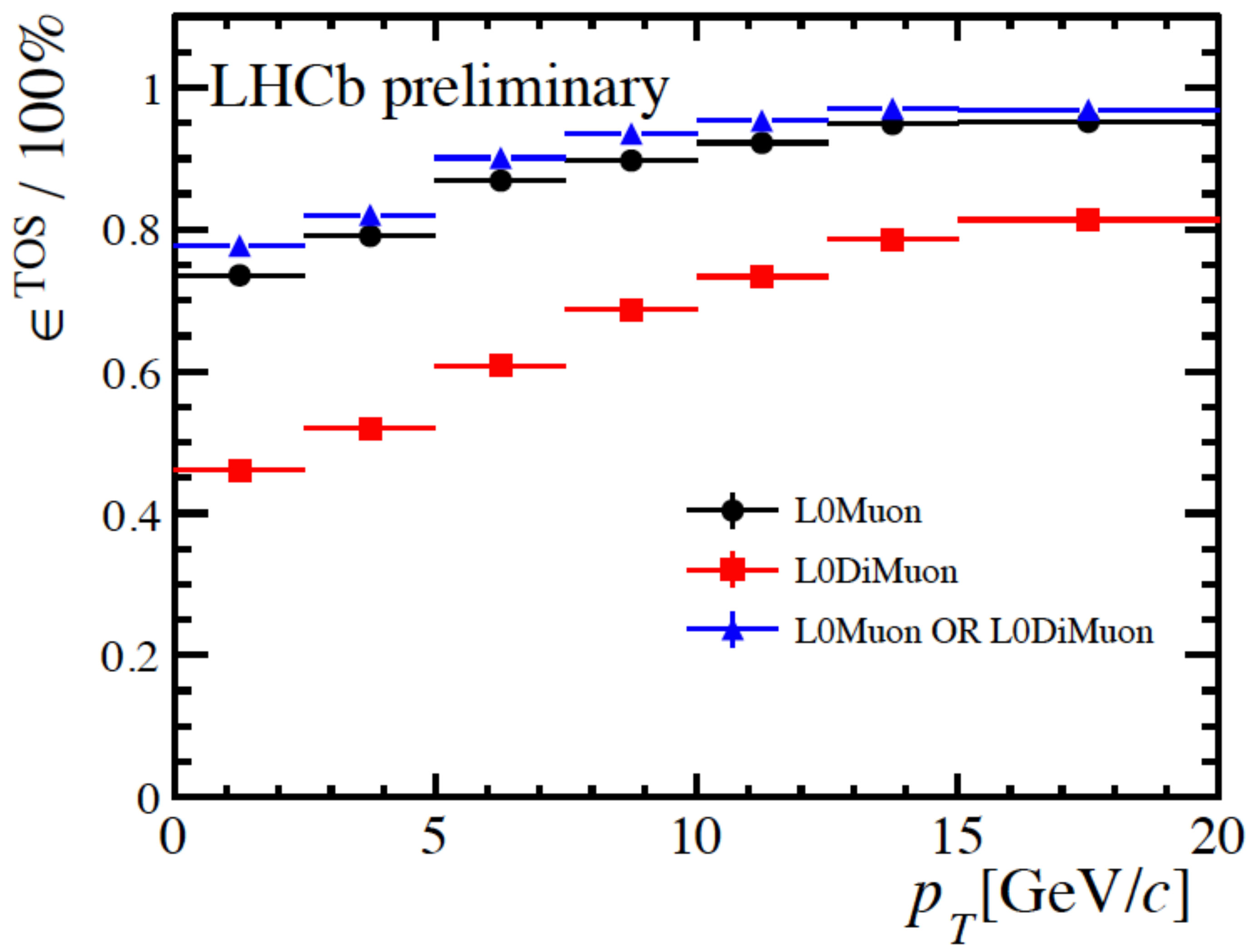
2015 bugged tracking correction table

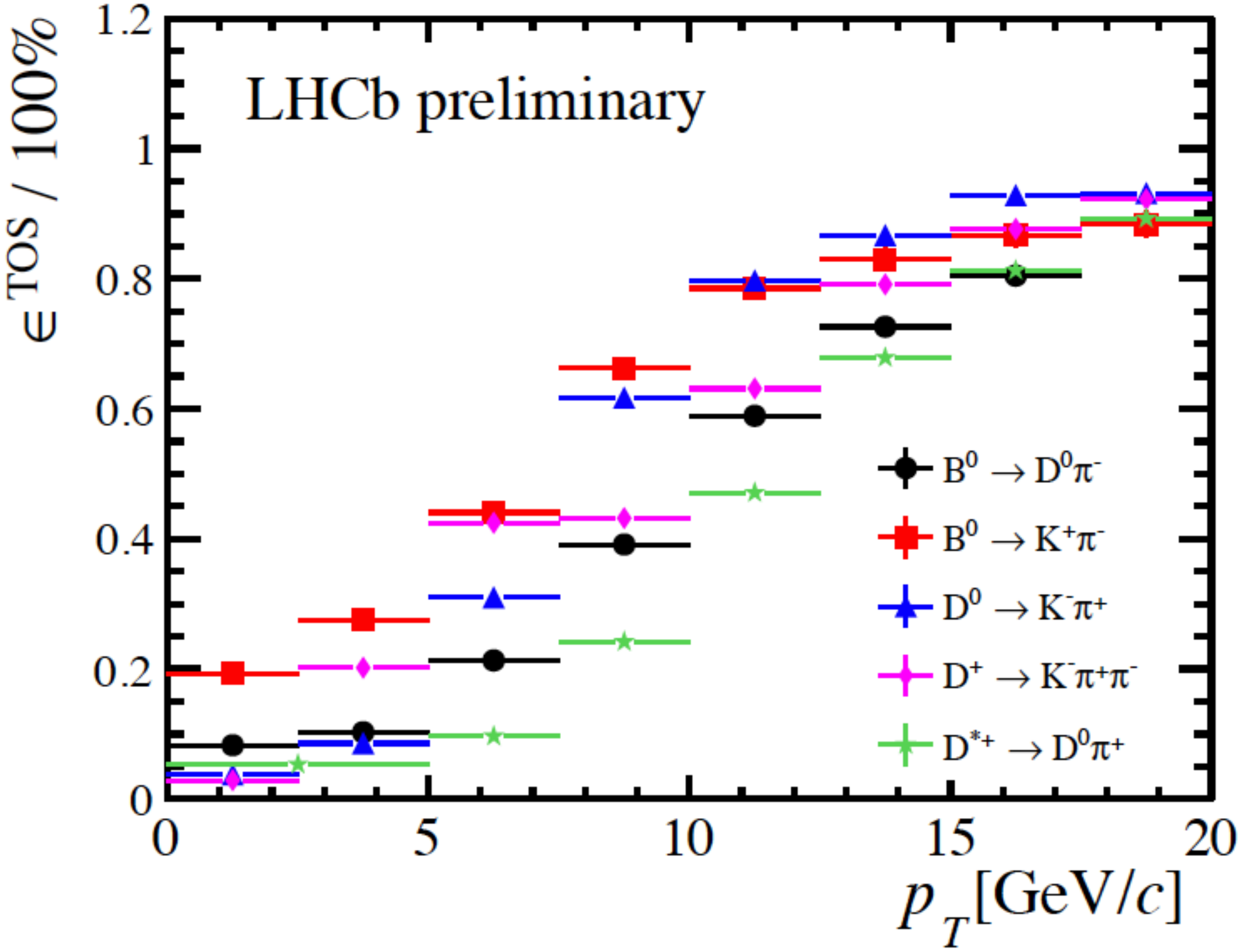


Ratios of charm cross-sections

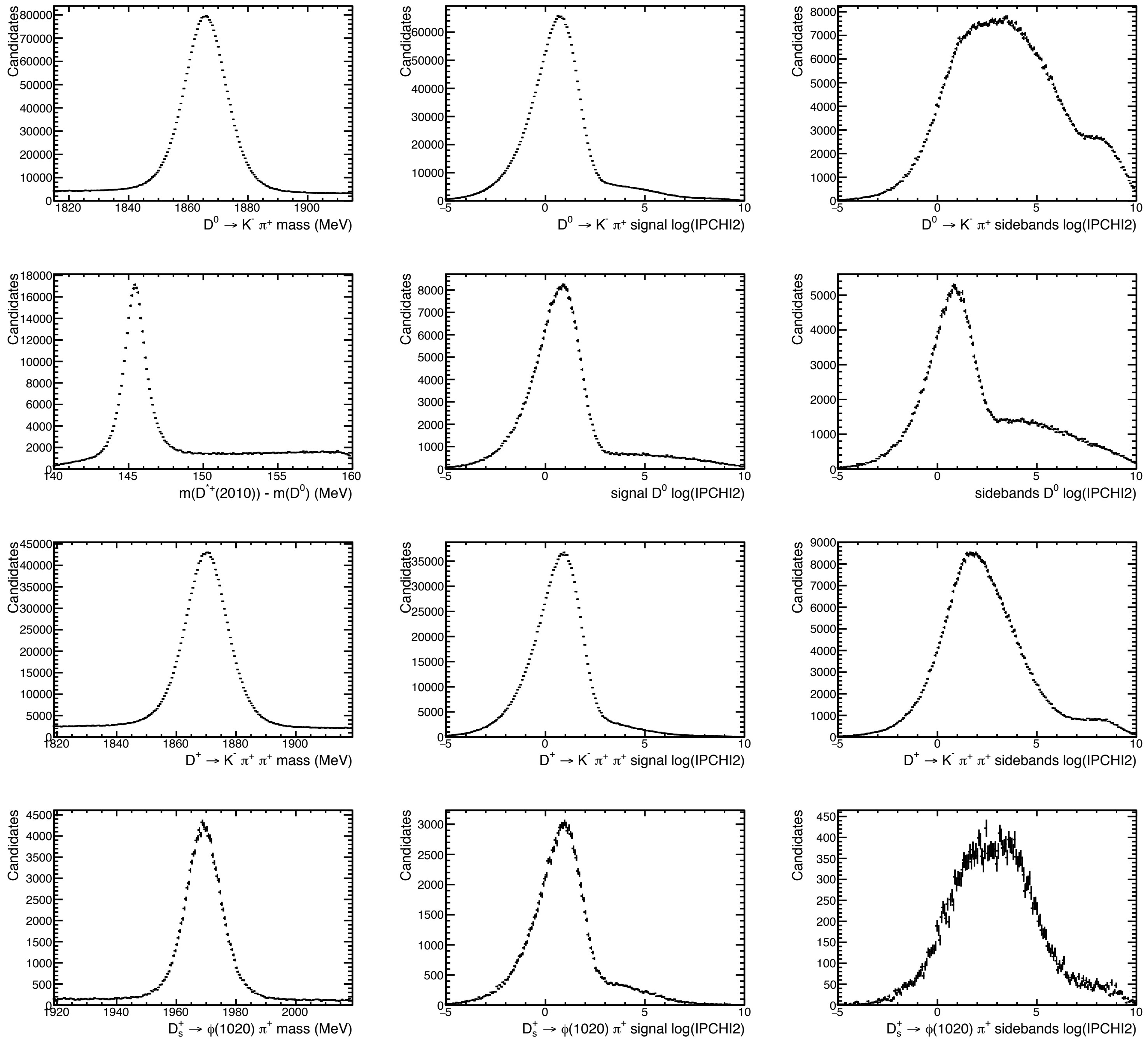


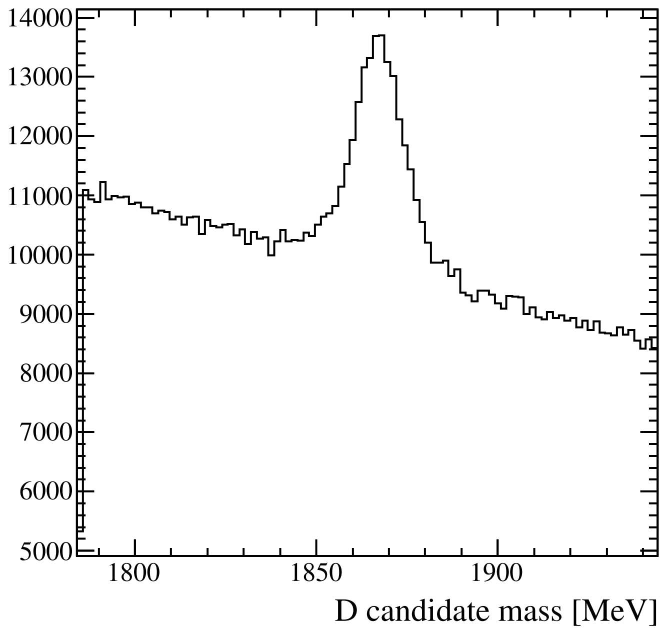
LHCb hardware trigger efficiency

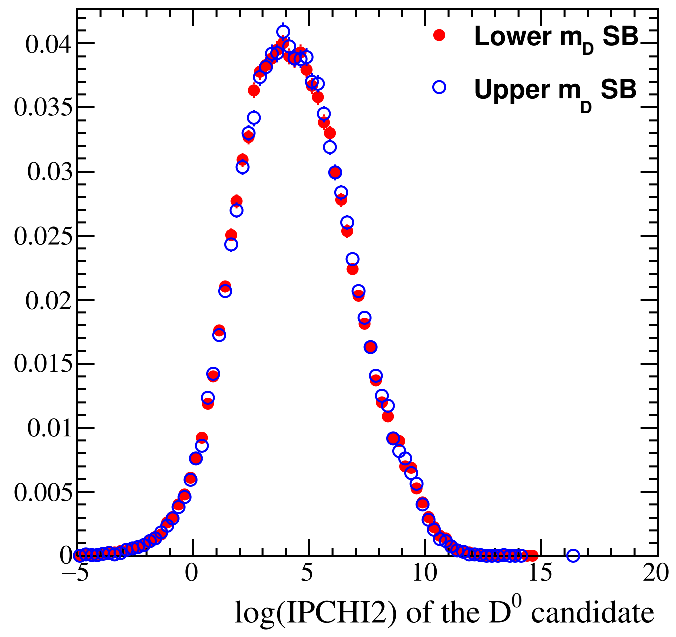




Charm background subtraction







Huge IP backgrounds in 2015