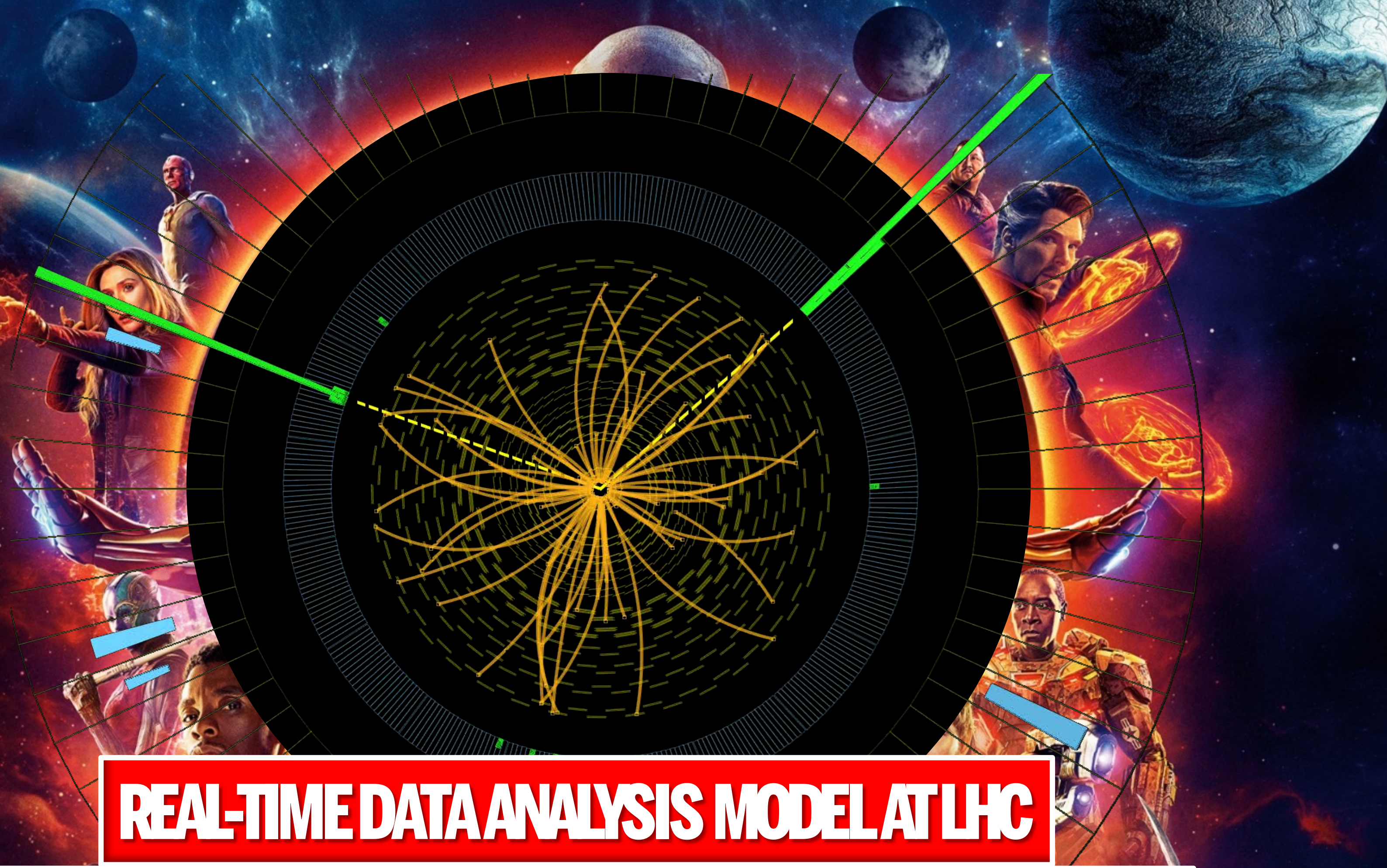
CHEP 2019

Adelaide



REAL-TIME DATA ANALYSIS MODEL AT LHC AND CONNECTIONS TO OTHER EXPERIMENTS & FIELDS

ARANTZA OYANGUREN IFIC-VALENCIA

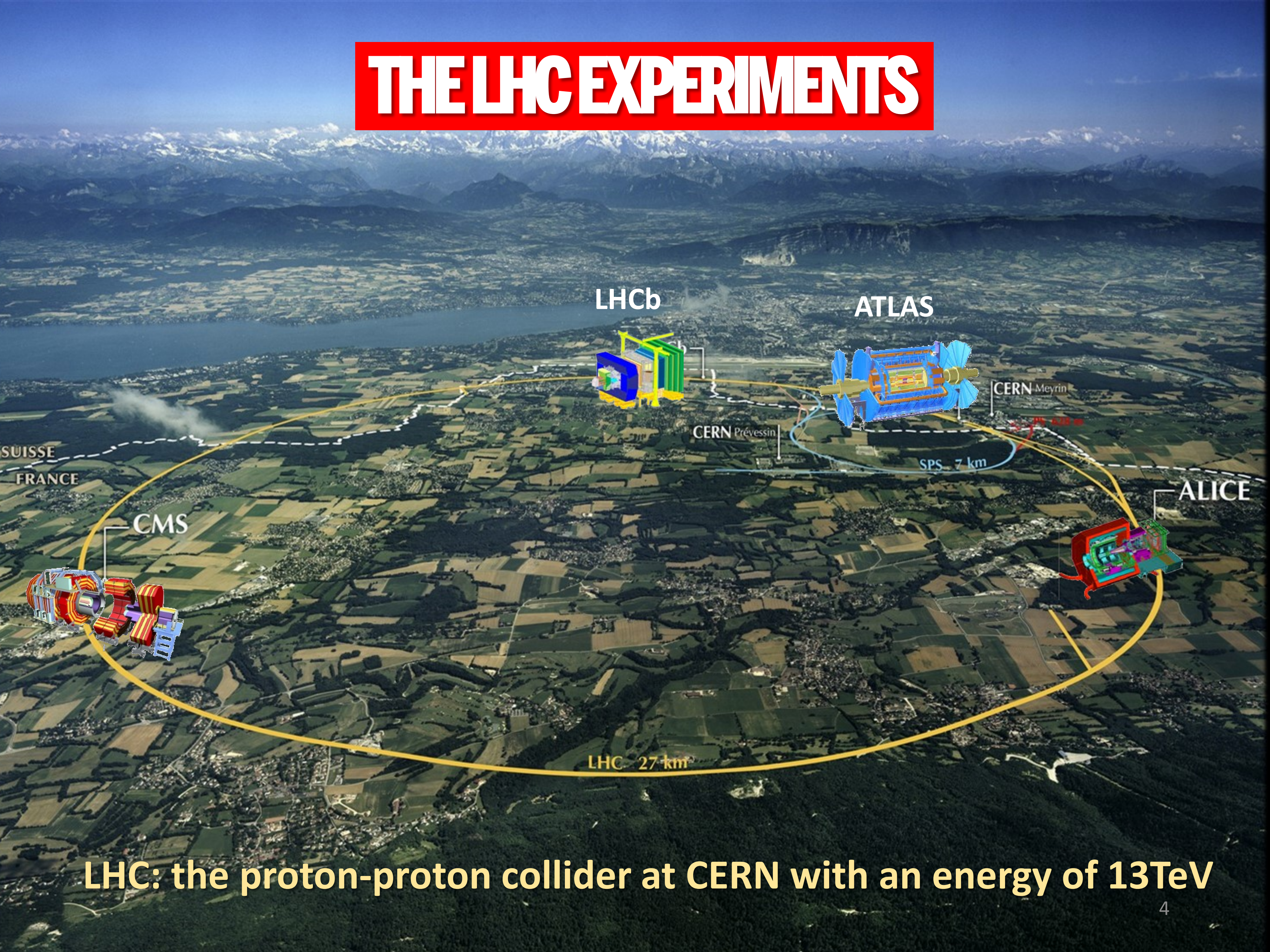
5th Nov. '19

OUTLINE

- The LHC experiments
- Real-Time Analysis
- Alignment and calibration
- Using accelerators
- Connections to other experiments and fields
- Conclusions

THE LHC EXPERIMENTS

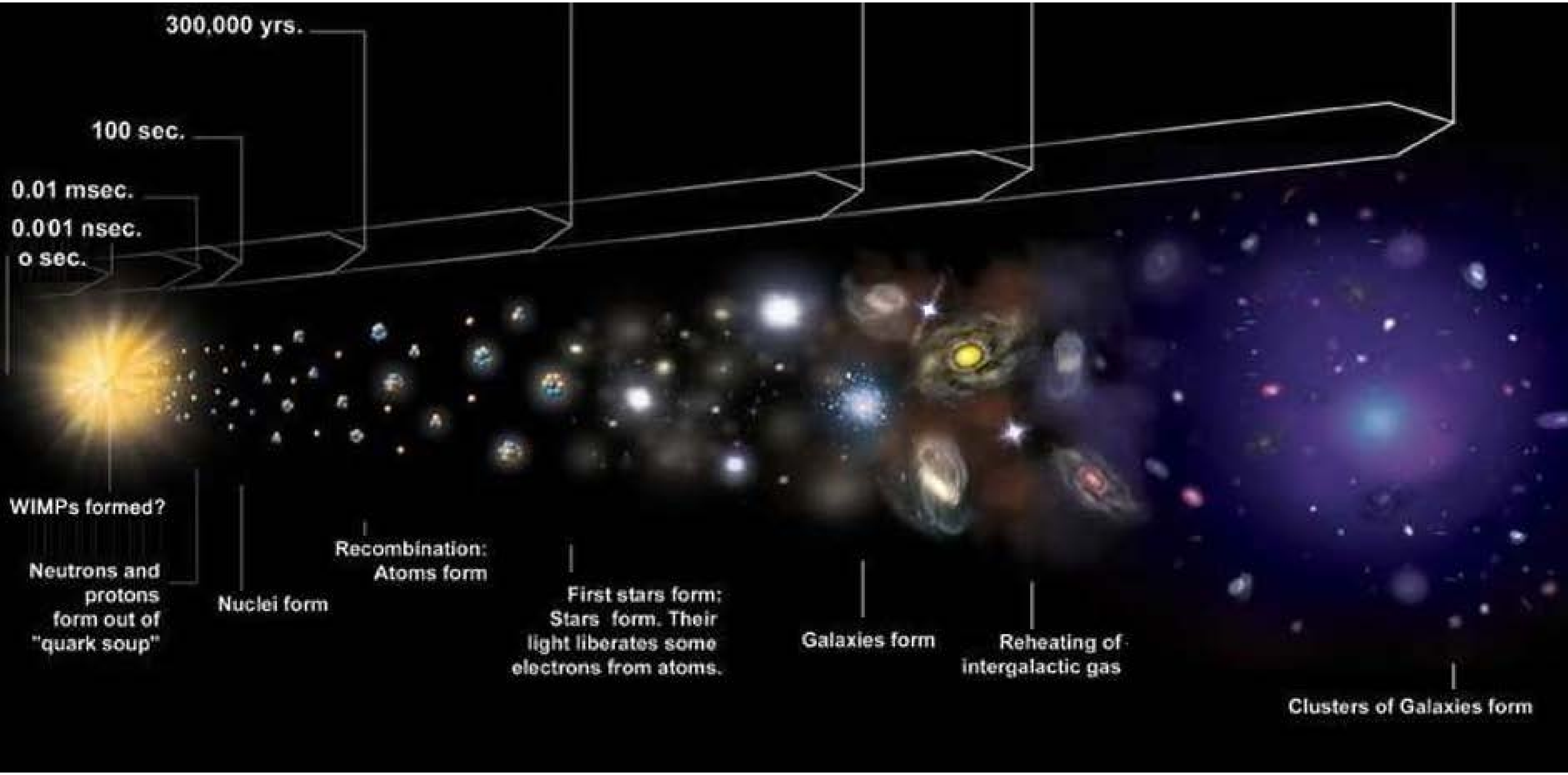
THE LHC EXPERIMENTS



LHC: the proton-proton collider at CERN with an energy of 13TeV

THE LHC EXPERIMENTS

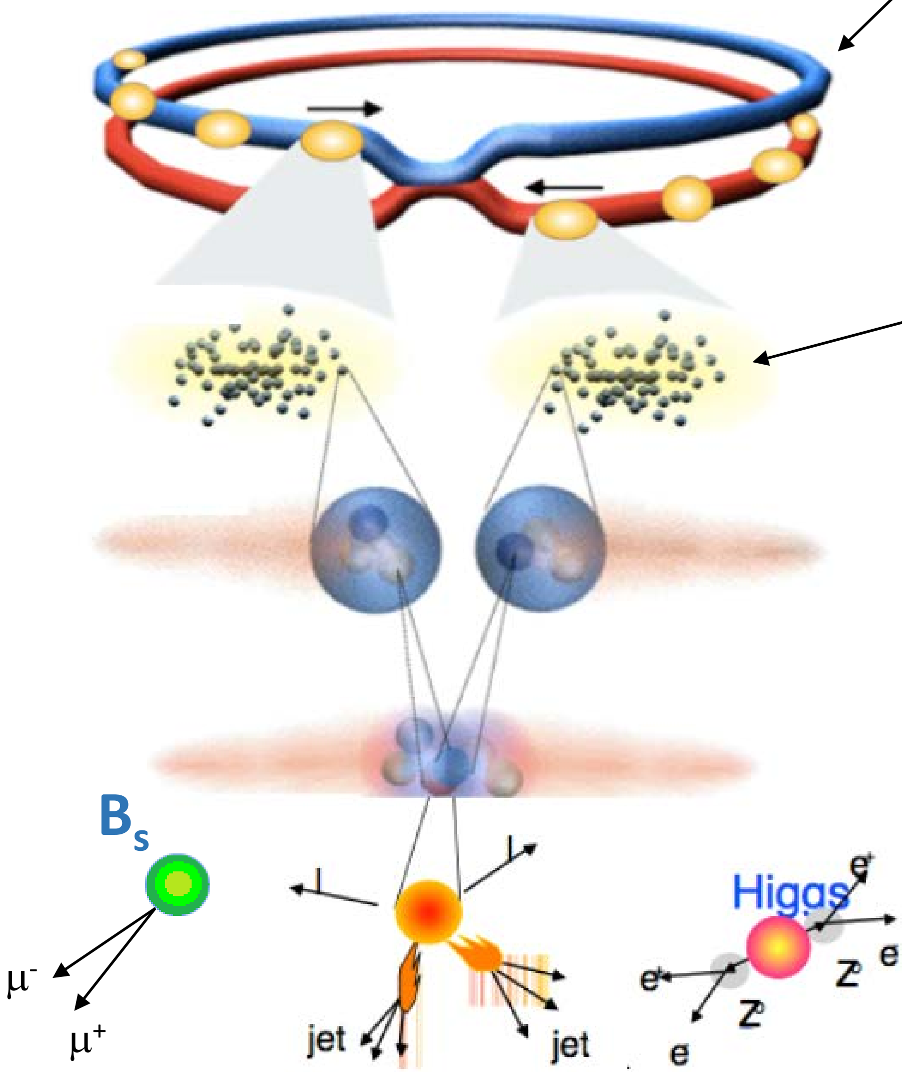
Aiming to discover new particles that explain the gaps of our well-established Standard Model of Particle Physics



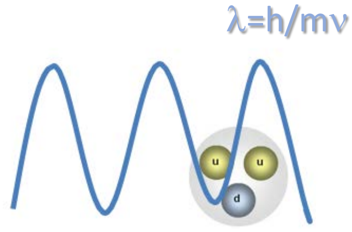
What is dark matter?   What is mass?   Where is the antimatter?

THE LHC EXPERIMENTS

Proton-proton collision



2028 bunches of protons per beam

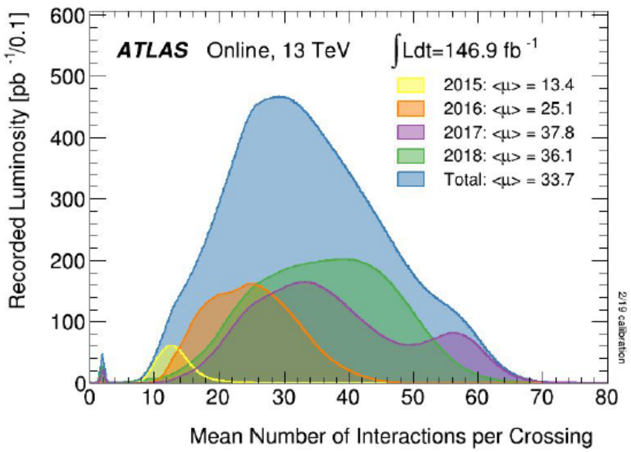


Beam energy of 7 TeV
(access to ~10^{-16} cm)

Luminosity 10^{34} cm^{-2} s^{-1}

10^{11} protons per bunch

Crossing rate 40 MHz, i.e. 40 M collisions/s



About 1 MB data per collision

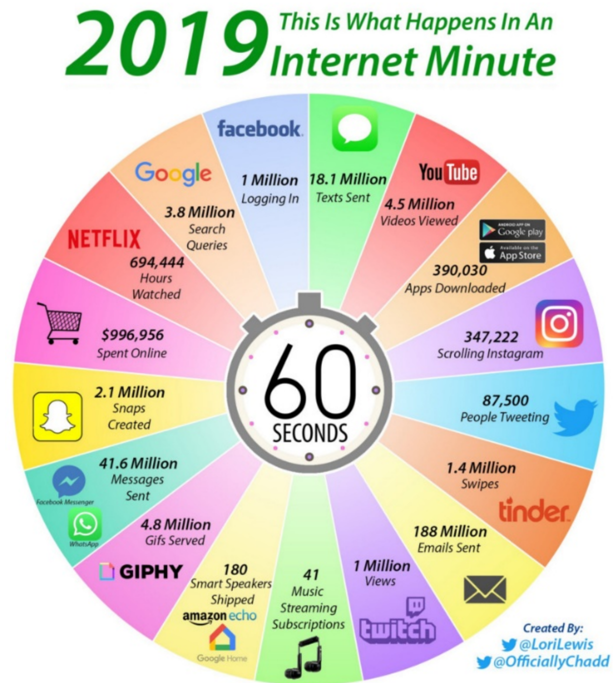
→ **40 TB/s**

THE LHC EXPERIMENTS

British Library, London

~ **12 TB**

(25 M books)

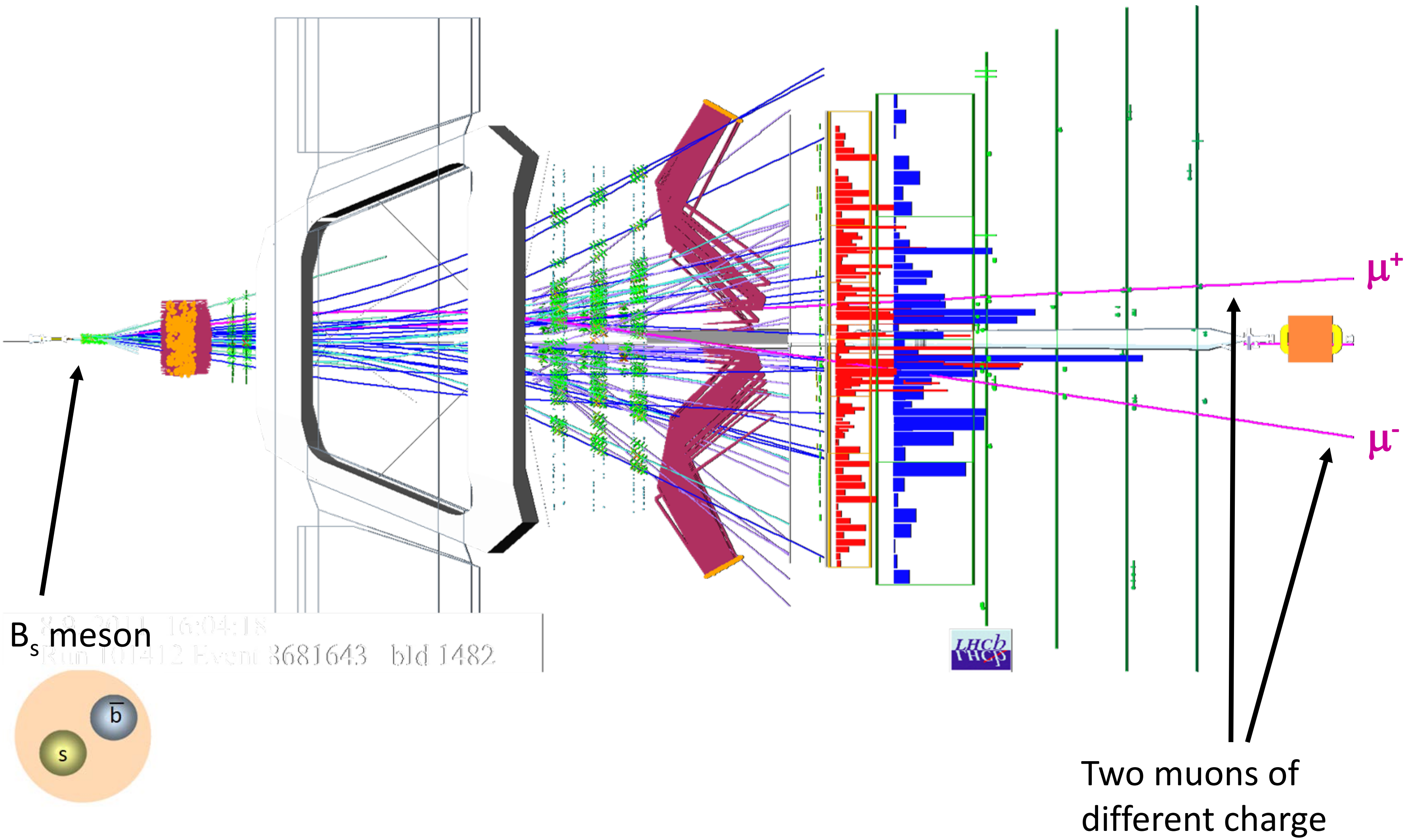


Google + + ~

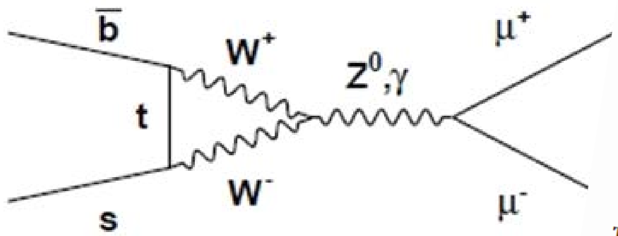
0.3 TB/s

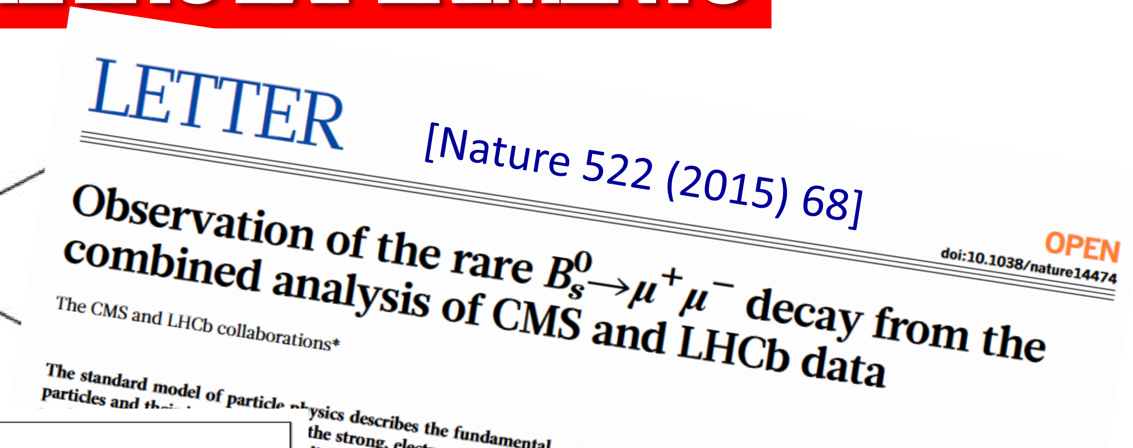
THE LHC EXPERIMENTS

Searching for very rare processes:



THE LHC EXPERIMENTS




LETTER

doi:10.1038/nature14474

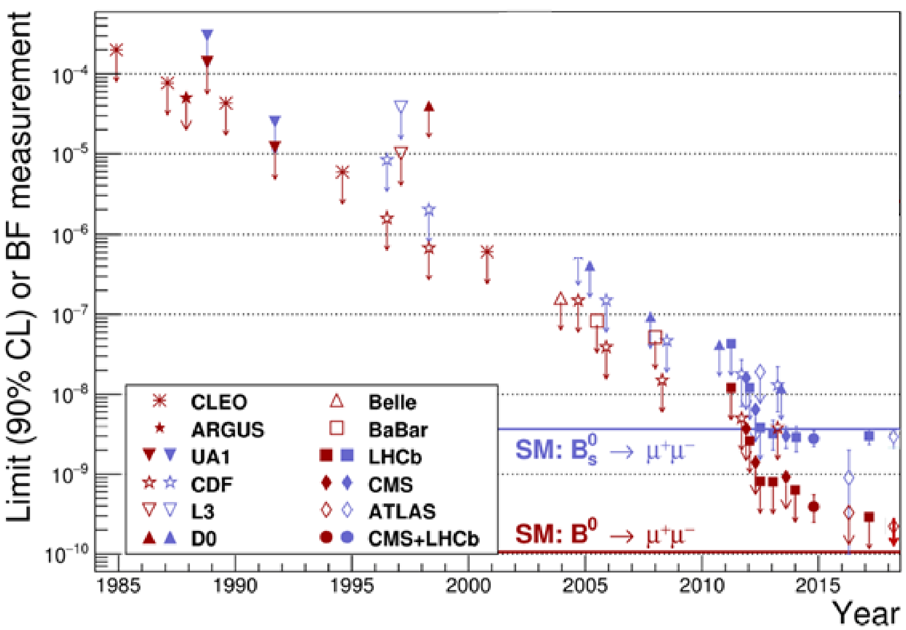
OPEN

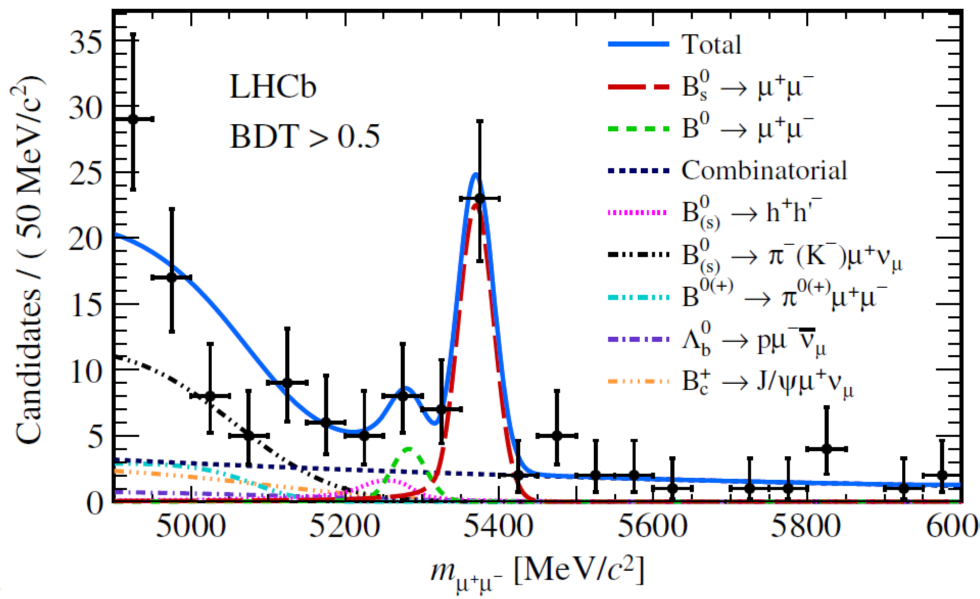
Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data

The CMS and LHCb collaborations*

The standard model of particle physics describes the fundamental particles and their ... the strong, electr...

[Nature 522 (2015) 68]



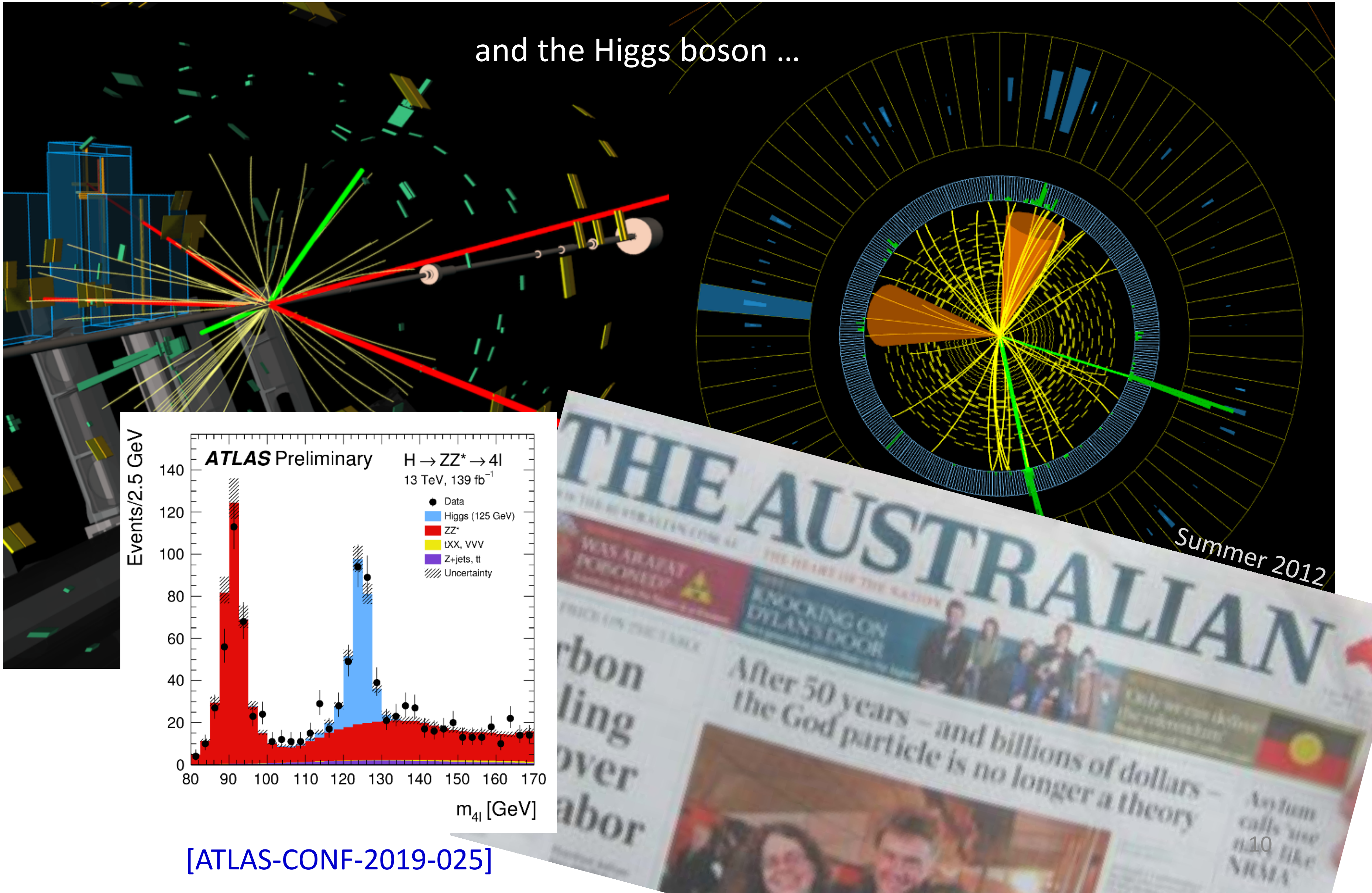


[PRL 118 (2017) 191801]

Searched for since the 80’s, observed now by LHCb, CMS and ATLAS experiments

It confirms the prediction of our Standard Model

THE LHC EXPERIMENTS



[ATLAS-CONF-2019-025]

THE LHC EXPERIMENTS

1.000.000.000 events/s rate in LHC collisions
0.00001 event/s where a new particle could appear

What events do we keep?

→ the *trigger* systems

THE LHC EXPERIMENTS

Impossible to select all the data: need to select the events of interest

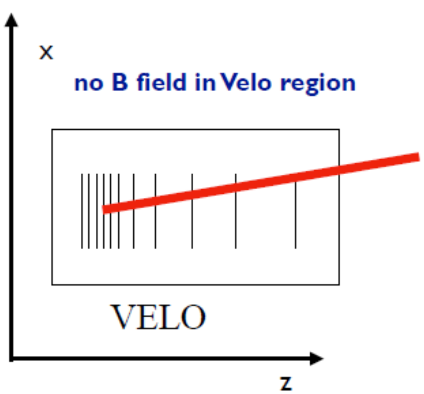
Traditional trigger systems:

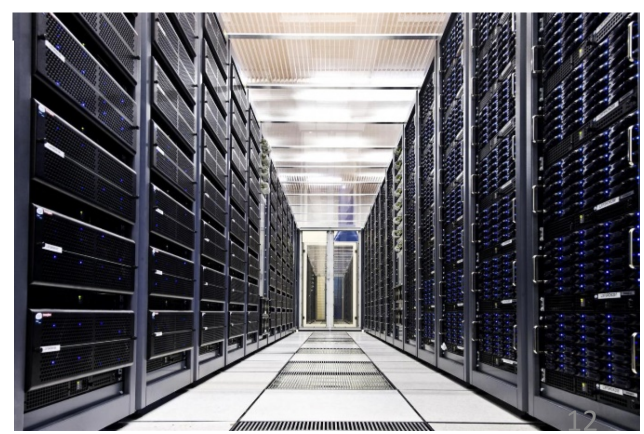
First level (L0/L1) → Y/N ? → First High Level (HLT1) → Y/N ? → Second High Level (HLT2) → Y/N ?

Custom electronics (FPGAs), information from calorimeters and muon stations

Processor farm, fast information from tracking

Processor farm, detailed information to reconstruct the event

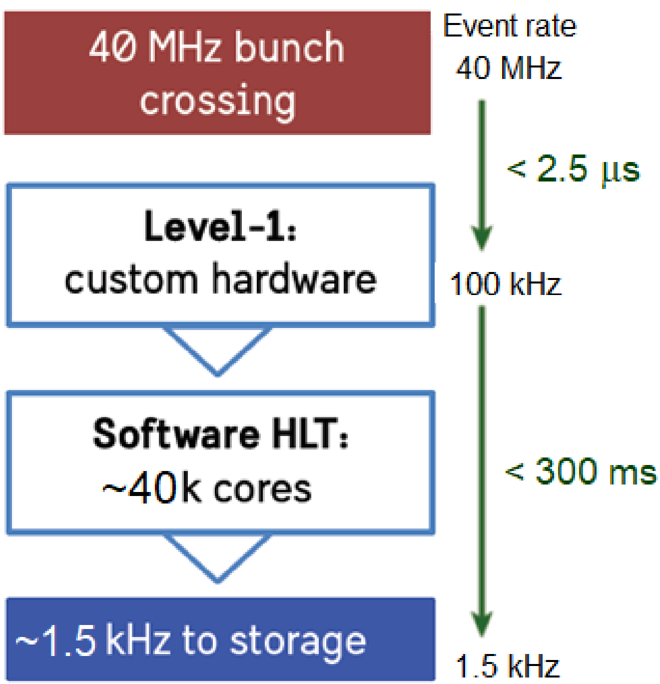




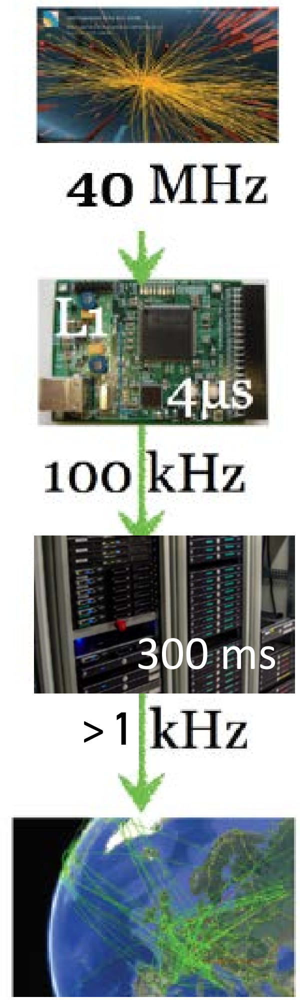
THE LHC EXPERIMENTS

The trigger systems
(last data taking period)

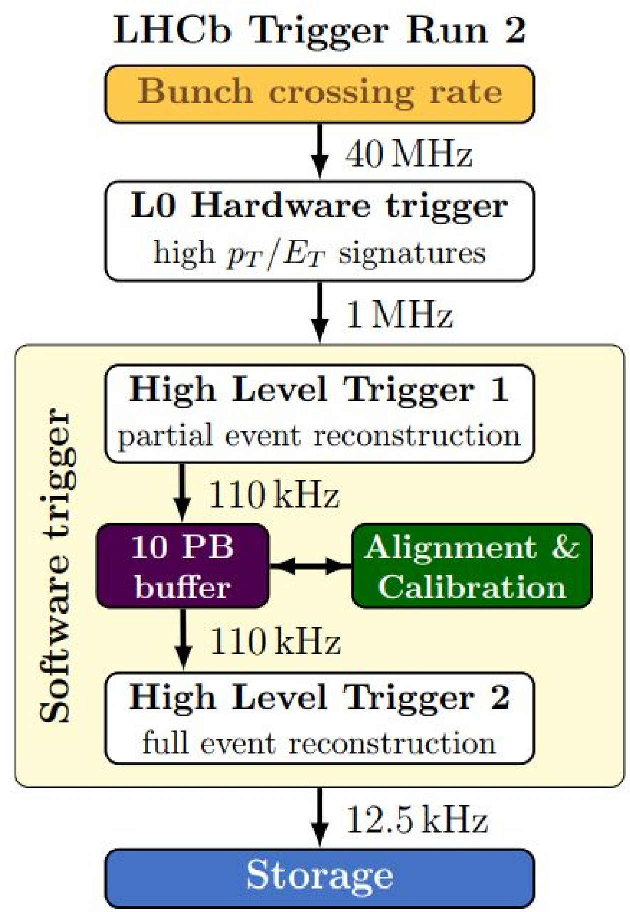
ATLAS



CMS



LHCb

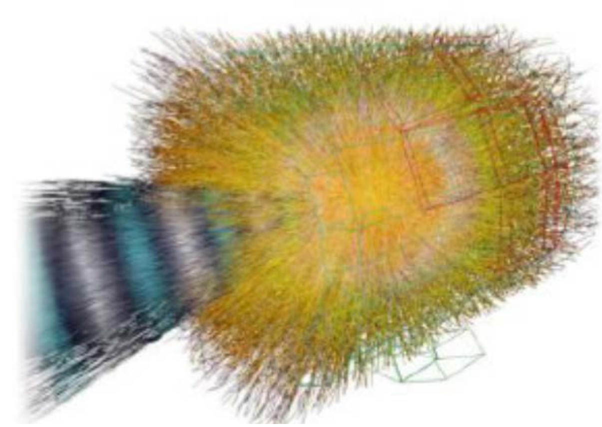


THE LHC EXPERIMENTS

The trigger systems
(last data taking period)

ALICE

Study of quark-gluon plasma properties through p-p, p-Pb and Pb-Pb collisions



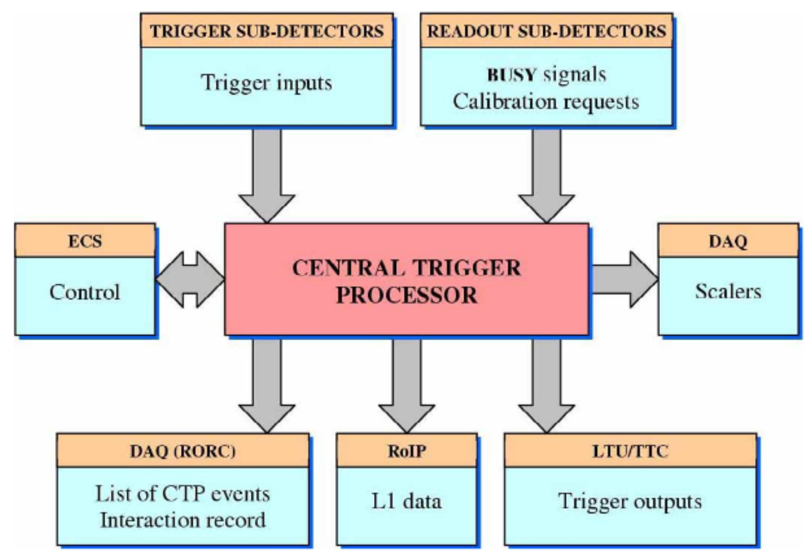
Ion rate of 10 KHz
(minimum bias events)

↓

Specific hardware triggers (L0, L1, L2)
with different latencies

↓ 500 Hz – 2 kHz

HLT: No rejection, data compression



(See next plenary talk by R. Shahoyan)

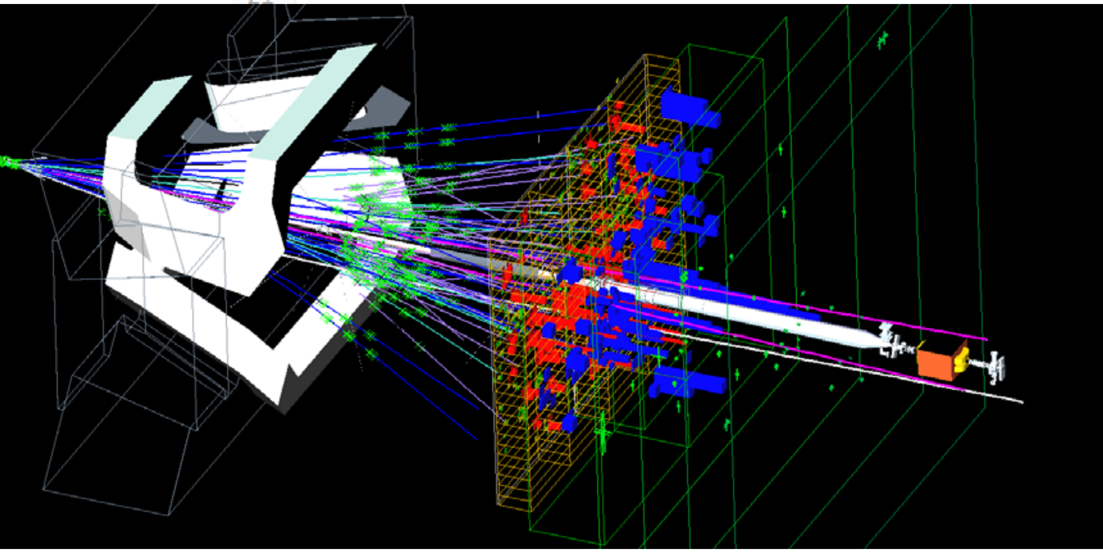
THE LHC EXPERIMENTS

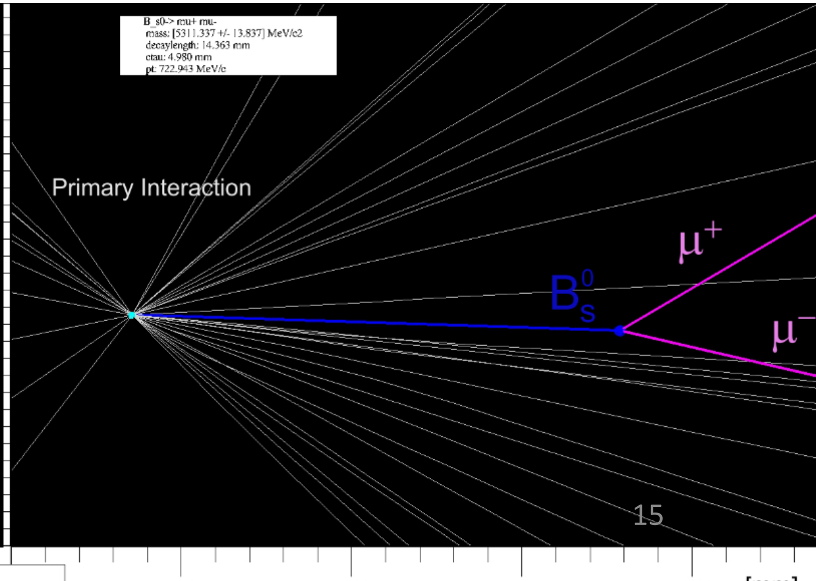
In practice, hundreds of ***trigger paths***, each selecting for a particular physics signature

Just an example: selecting $B_s \to \mu^+\mu^-$ at LHCb

```

from GaudiKernel.SystemOfUnits import GeV, MeV, mm, picosecond

from Hlt2Lines.Utilities.Hlt2LinesConfigurableUser import Hlt2LinesConfigurableUser
class DiMuonLines(Hlt2LinesConfigurableUser) :

    __slots__ = { 'Common' :        {'TrChi2'      :   10,
                                      'TrChi2Tight':    5},

                   'DiMuon' :        {'MinMass'     :    0 * MeV,
                                      'Pt'          :  600 * MeV,
                                      'MuPt'        :  300 * MeV,
                                      'VertexChi2' :    9},
```

(See talk by N. Nolte @ Track5)





THE LHC EXPERIMENTS

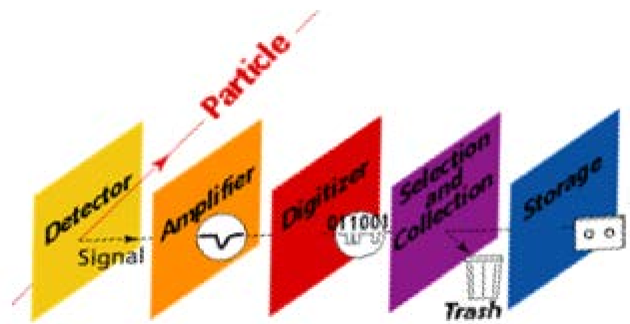
How many data can we record?

The need of storage is given by the trigger bandwidth:

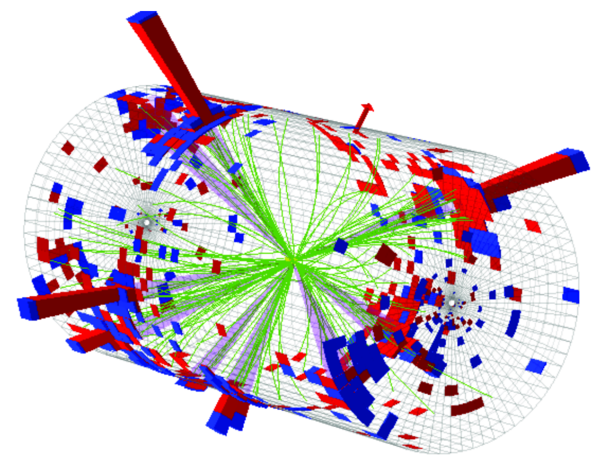
Bandwidth [GB/s] ~ Trigger output rate [kHz] x Average event size [MB]

> 1 GB/s

and increasing in the next periods of data taking



~1.5 kHz (ATLAS & CMS)
12.5 kHz (LHCb)

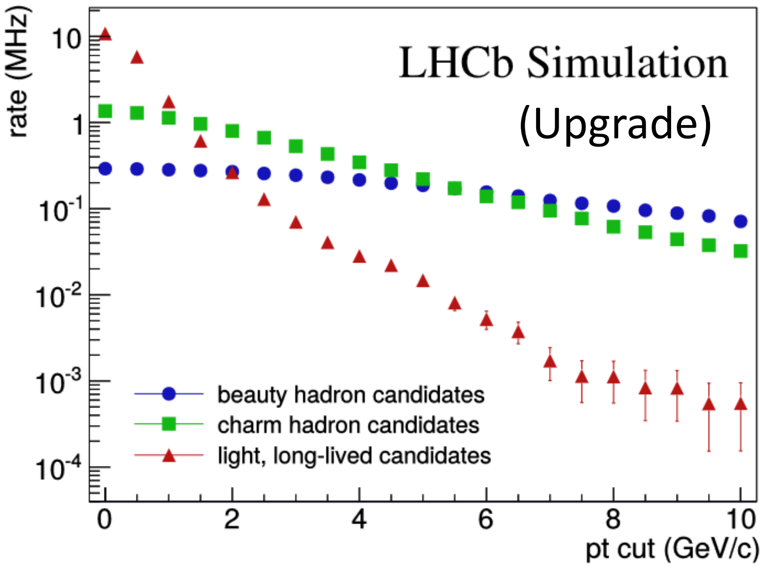


Raw event data size
~1 MB (ATLAS and CMS)
~0.1 MB (LHCb)

REAL TIME ANALYSIS

REAL TIME ANALYSIS

Bandwidth [GB/s] ~ Trigger output rate [kHz] x Average event size [MB]



[LHCb-PUB-2014-027]

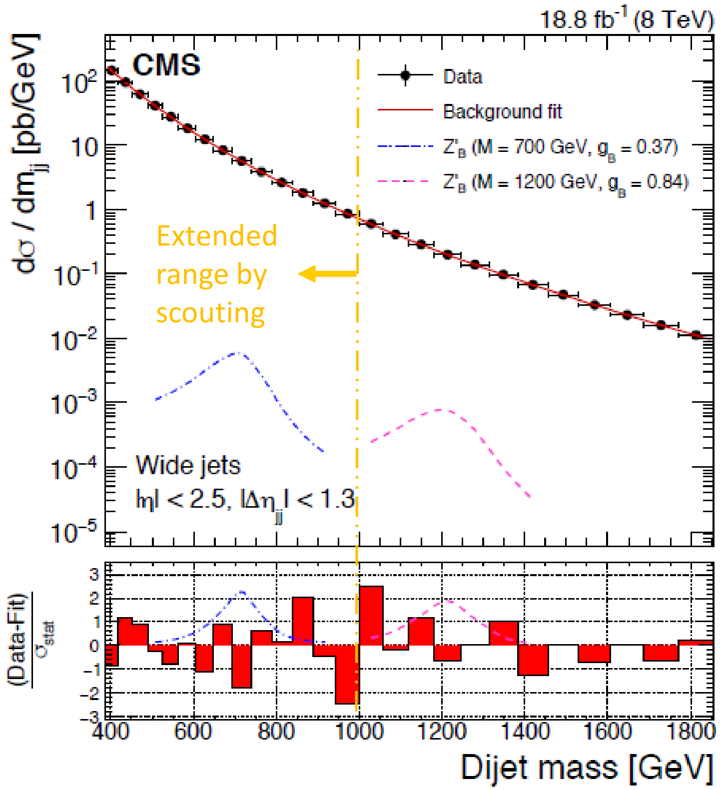
- The **trigger rate saturates**: we cannot reduce the trigger output, all our selected events are interesting signals!
- We need then to **reduce the event size**: Instead of taking the raw data, store only the relevant information

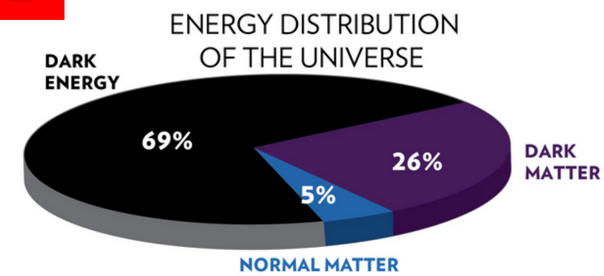
→ Need to reconstruct and analyse the events to select them in **real time**, and keep the important data

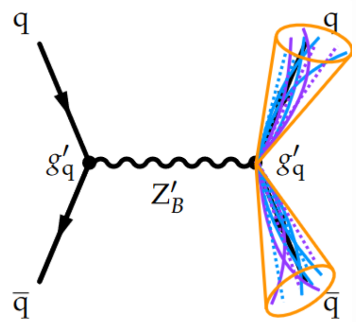
REAL TIME ANALYSIS

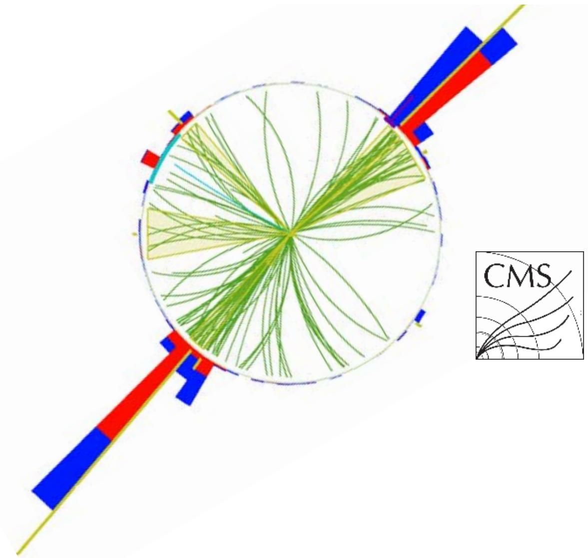
Scouting at CMS

★ **Searching for new resonances (related to dark matter)**









[CMS-EXO-11-094], Scouting applied in CMS since 2012
[CMS, PRL 117 (2016) 031802], [CMS, JHEP 08 (2018) 130]

Idea:

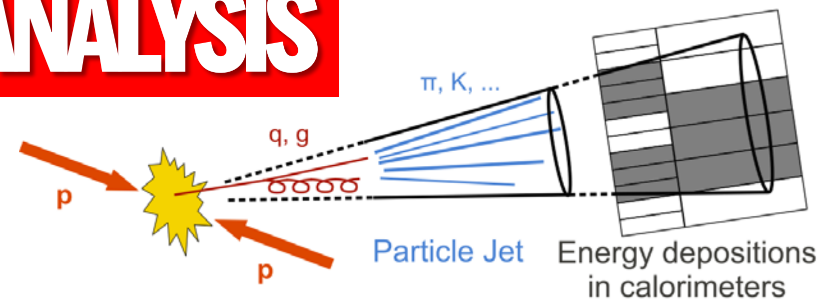
→ reconstruct and store **only** the information needed to perform some analyses

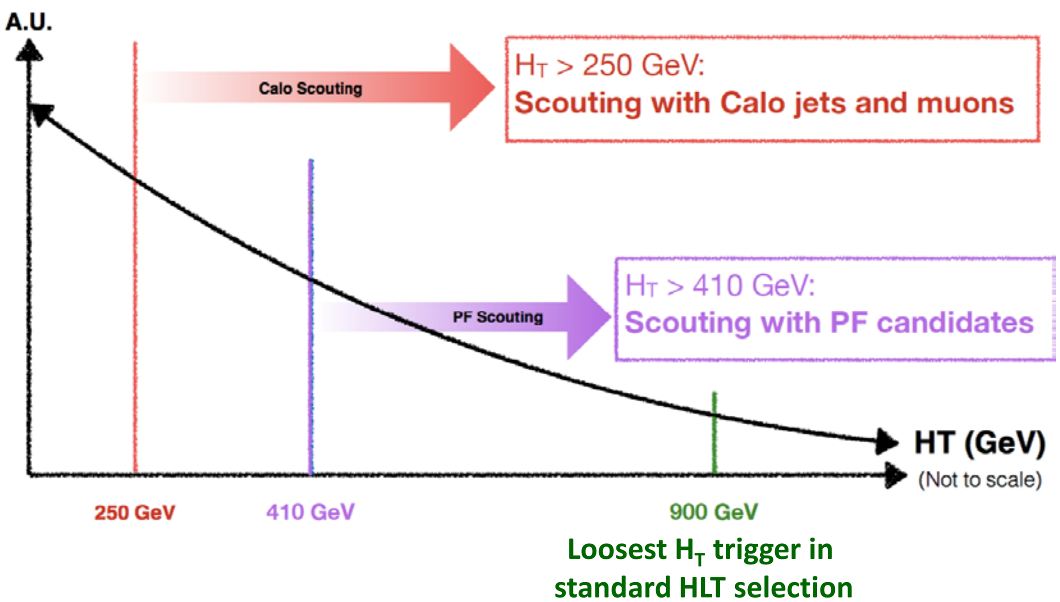
→ do physics analysis with these objects (from the trigger output): **Real Time Analysis**

REAL TIME ANALYSIS

Scouting at CMS

$H_T = \Sigma\,|\,\text{jet } p_T\,|$





CaloScouting: calorimeter jets, missing E_T (MET), vertices, muons

PF Scouting (Particle-Flow): PF jets and MET, PF candidates, vertices, muons

Stream	Rate (Hz)	Event Size	Bandwidth (MB/s)
PhysicsMuons	420	0.86 MB	360
PhysicsHadronsTaus	345	0.87 MB	300
ScoutingCaloMuon	4580	8.9 KB	40
ScoutingPF	1380	14.8 KB	20

[LHC FILL 7334, 10/18]

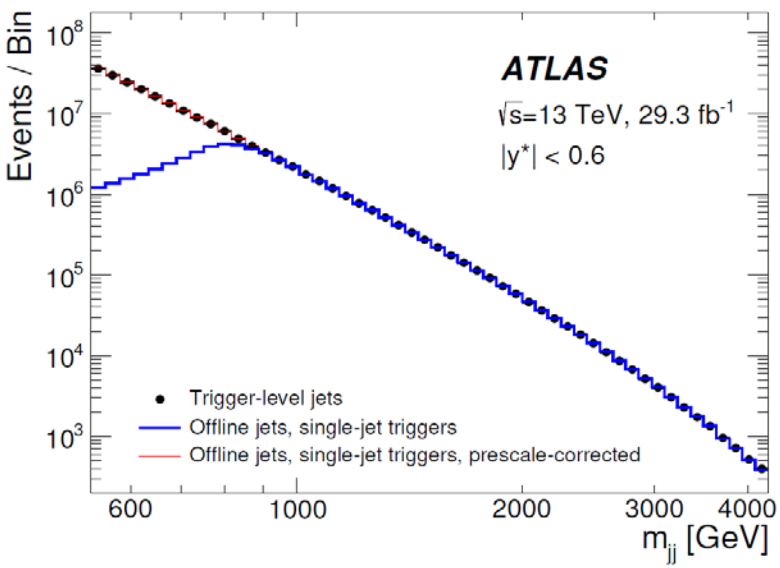
(See talk today @ Track1 by H. Sakulin)

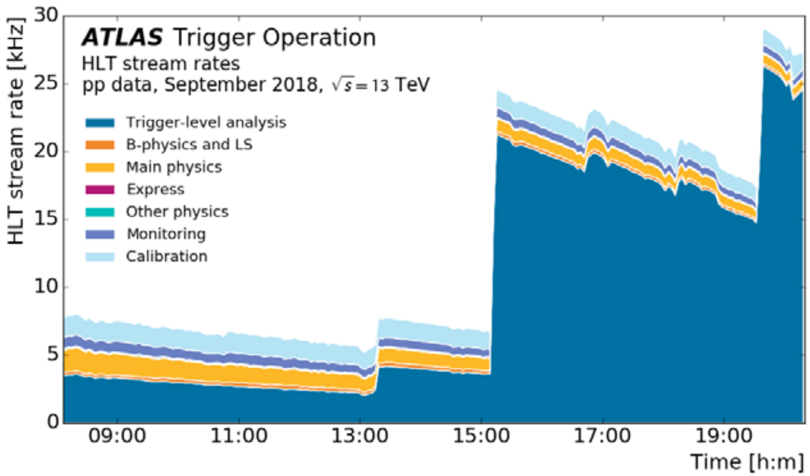
REAL TIME ANALYSIS

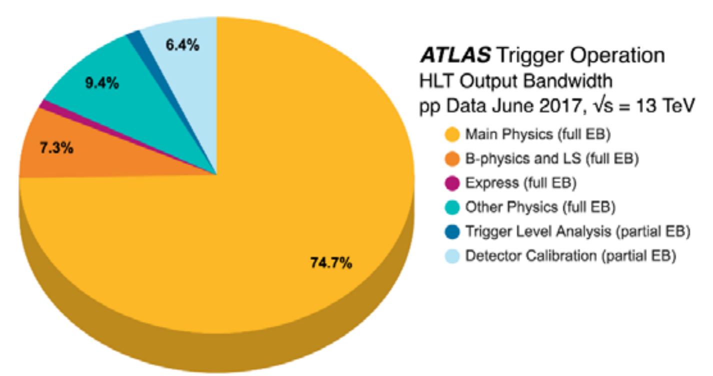
TLA (Trigger-object Level Analysis) at ATLAS

[ATLAS trigger]

Similar strategy as CMS scouting: store only jet 4-momenta and some summary info of the event







[ATLAS, PRL 121 (2018) 081801]

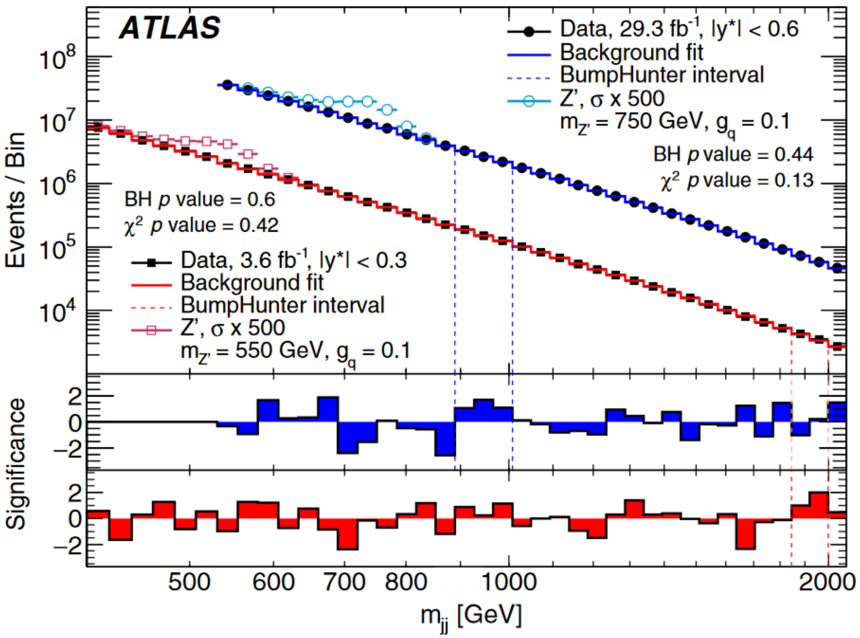
(See talk today @ Track1 by A. Boveia)

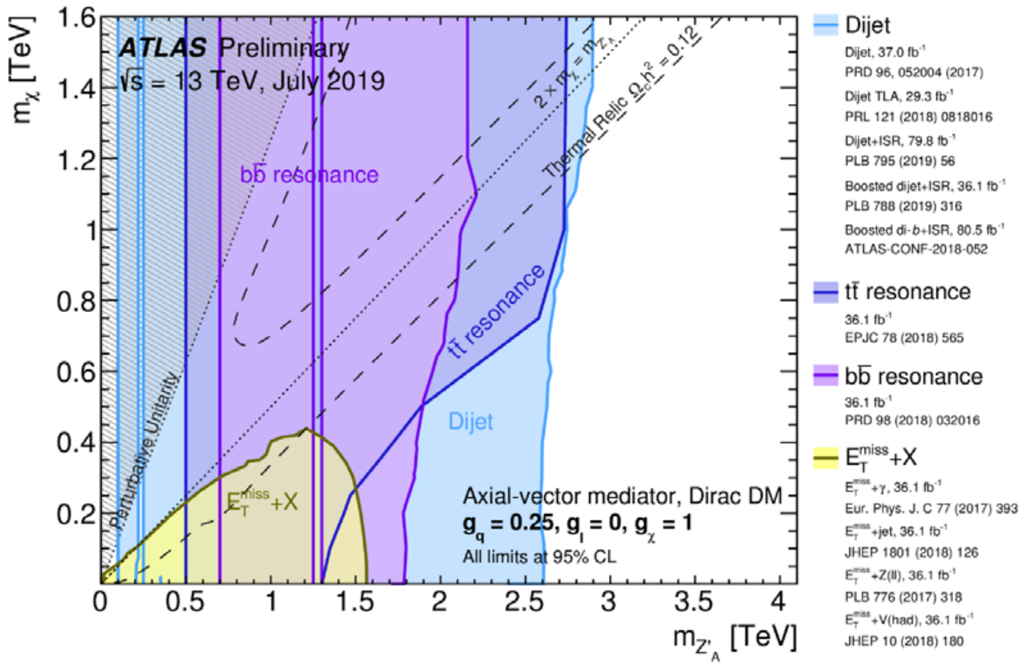
REAL TIME ANALYSIS

TLA (Trigger-object Level Analysis) at ATLAS

* **Searching for low-mass di-jet resonances**

  [ATLAS, PRL 121 (2018) 081801]

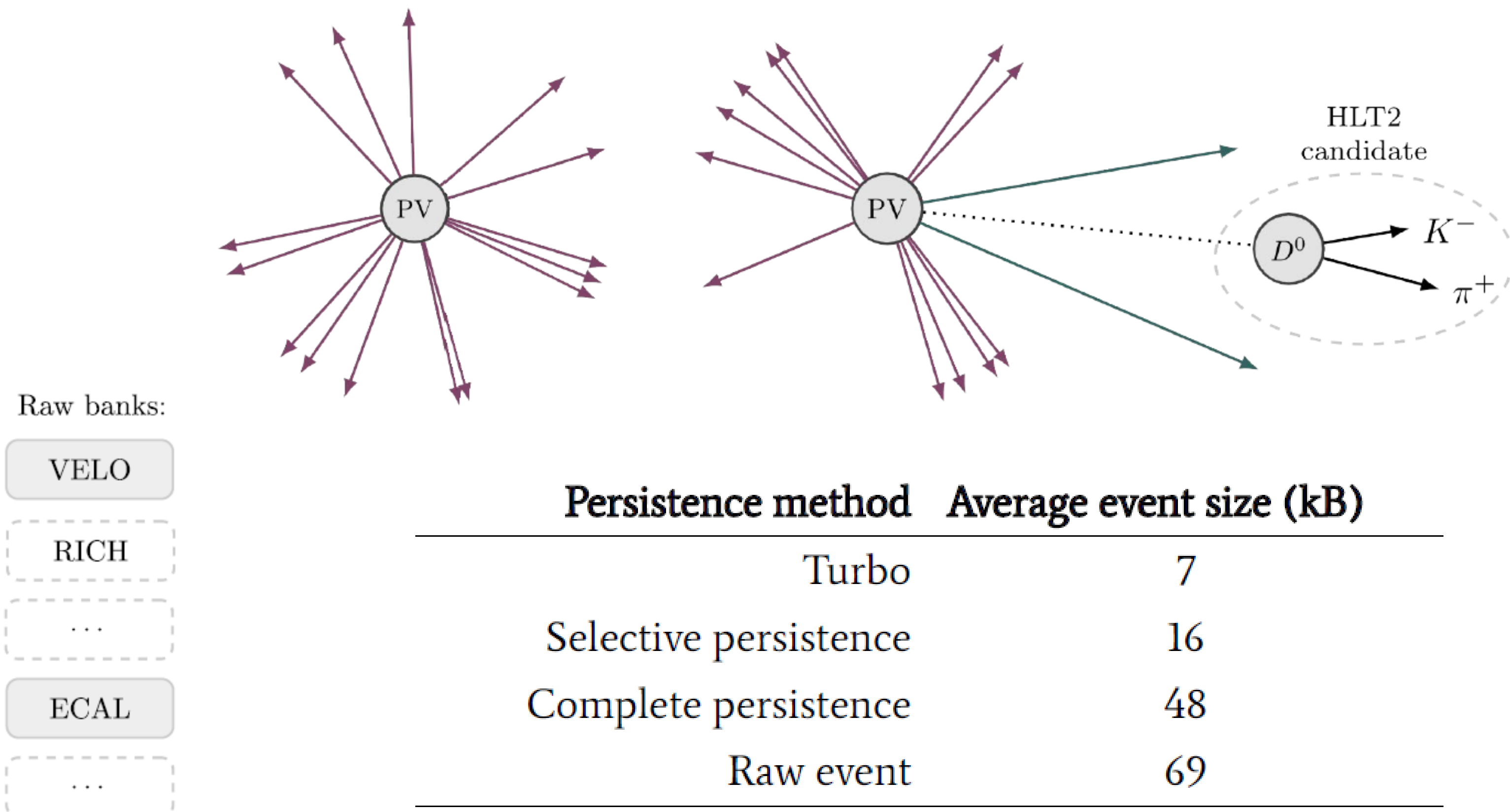




REAL TIME ANALYSIS

Turbo at LHCb

It exploits the event topology and saves only a subset of the objects which are relevant for a posterior analysis. One can use several persistence levels:



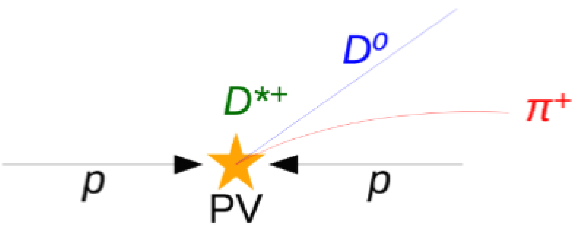
Persistence method	Average event size (kB)
Turbo	7
Selective persistence	16
Complete persistence	48
Raw event	69

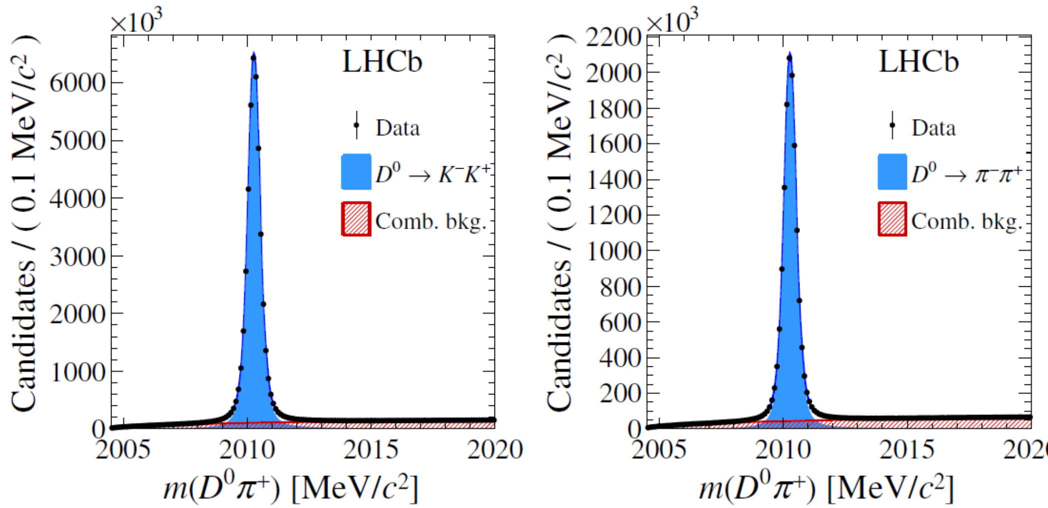
[Comput.Phys.Commun. 208 (2016) 35], applied in LHCb since 2015

REAL TIME ANALYSIS

★ **Observation of CP Violation in Charm Decays**

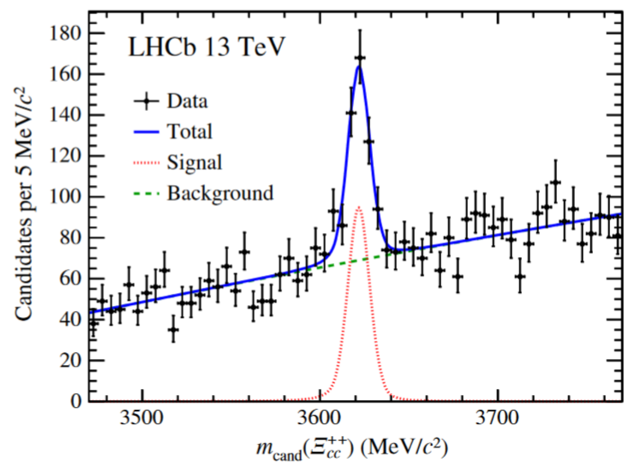
[LHCb, PRL 122 (2019) 211803]





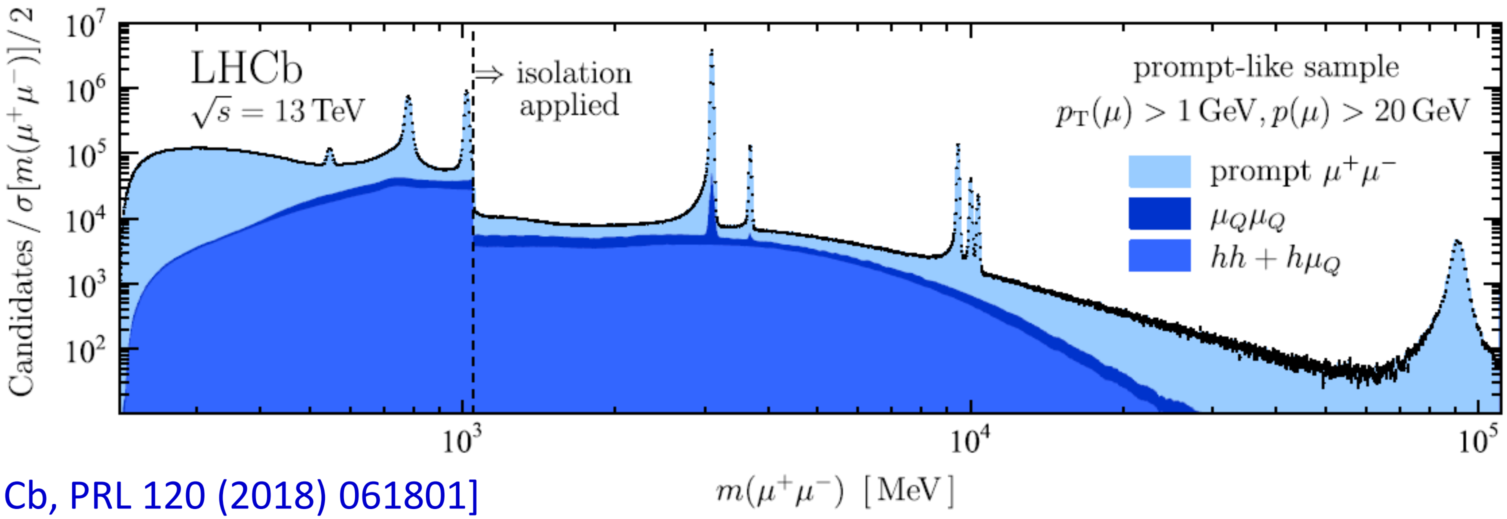
★ **Observation of the Doubly Charmed Baryon Ξ^{++}_{cc}**

[LHCb, PRL 119 (2017) 112001]

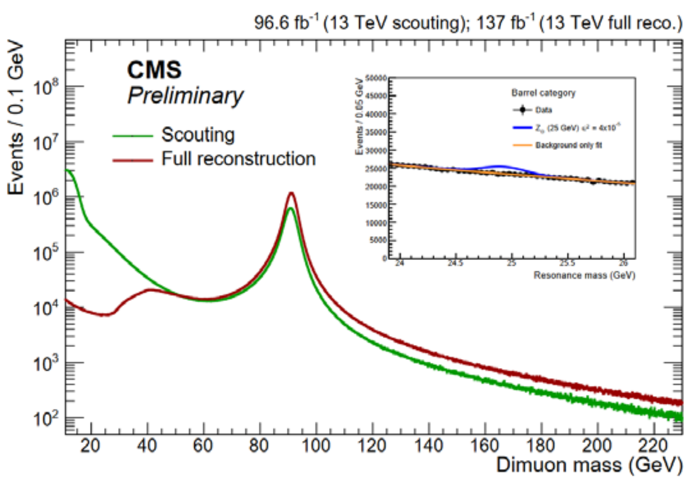


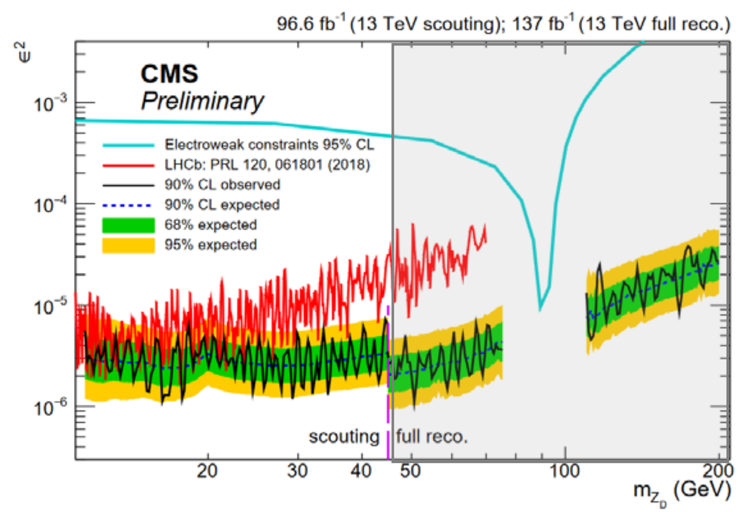
REAL TIME ANALYSIS

* **Search for Dark Photons decaying into two muons**



[LHCb, PRL 120 (2018) 061801]



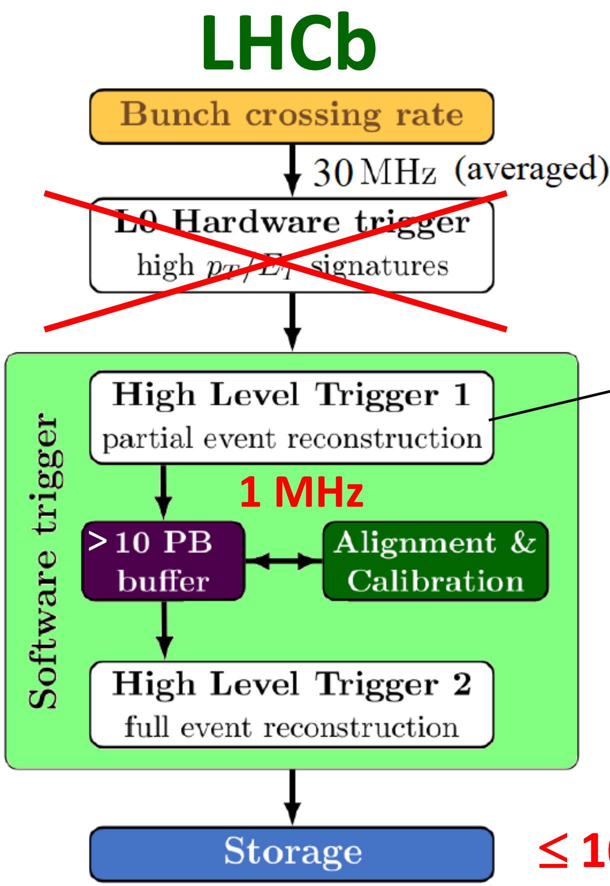


[CMS PAS EXO-19018]

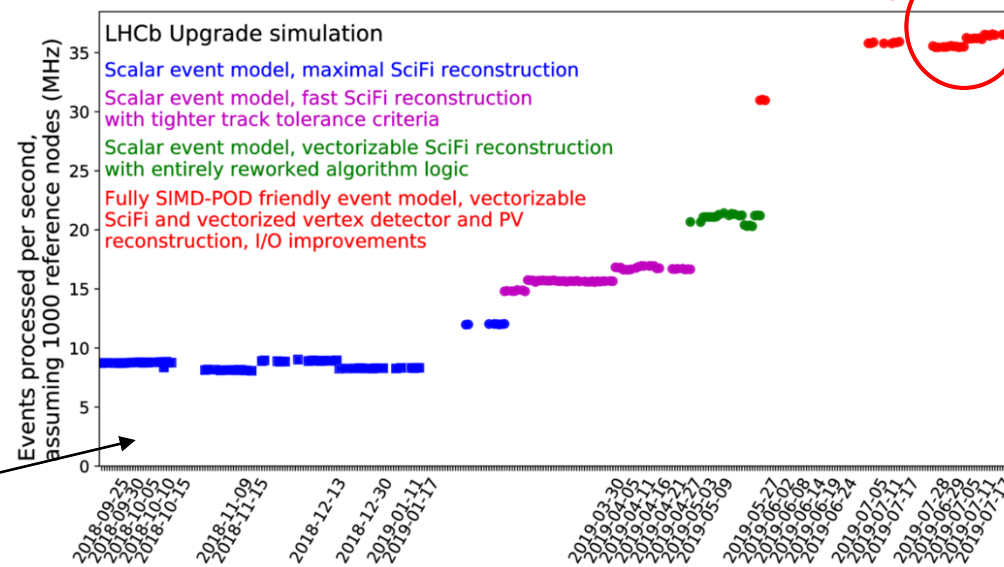
REAL TIME ANALYSIS

At LHCb, **new trigger strategy** for Run3 (2021 onwards):
Increasing the instantaneous luminosity x 5 and triggerless readout

Improved algorithms, data structures and vectorisation

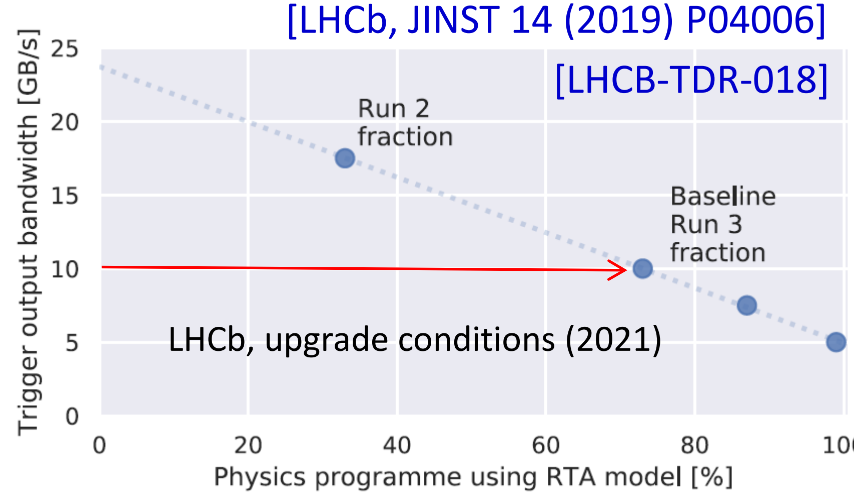


≤ 10 GB/s



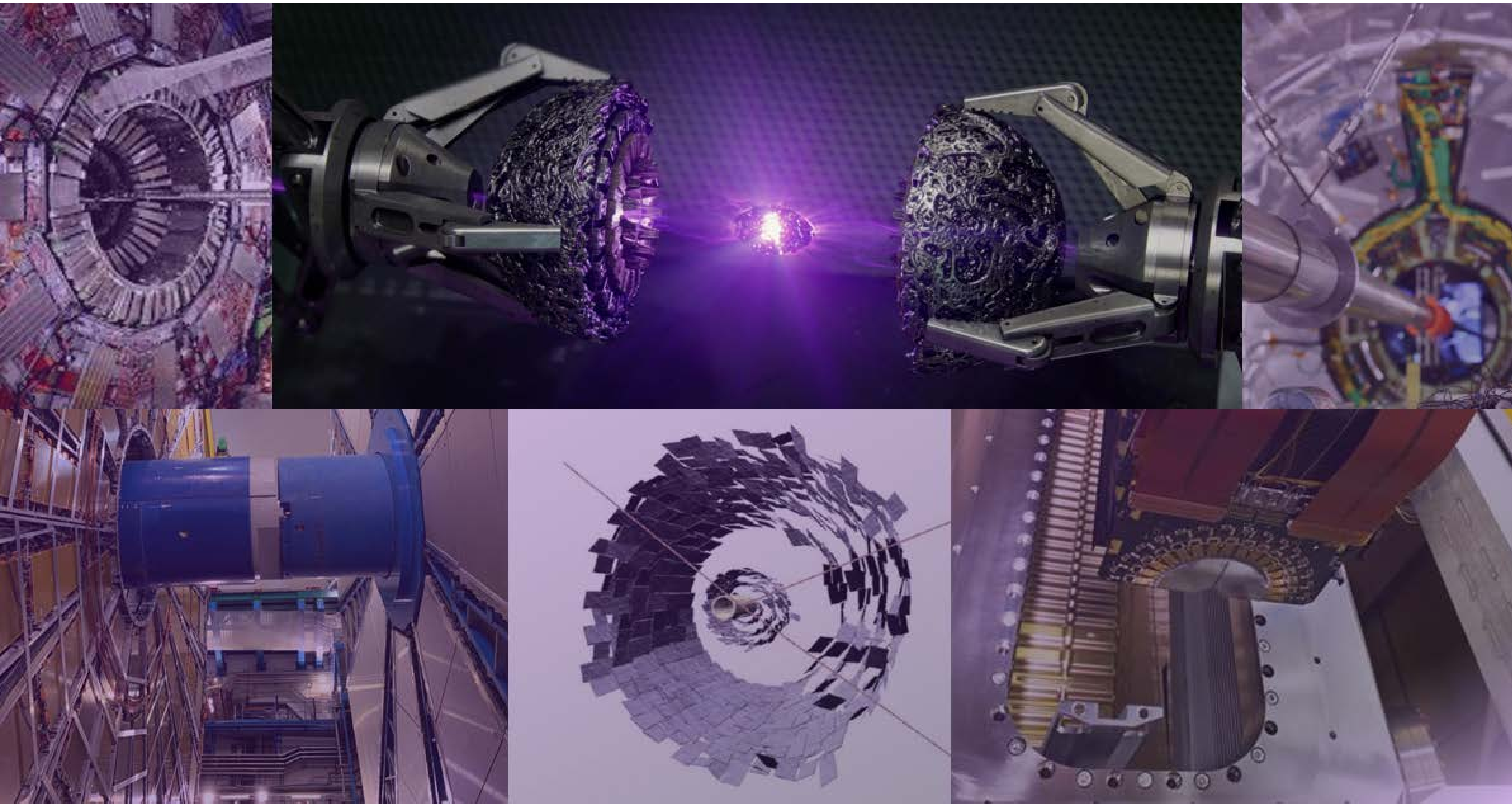
[LHCb, JINST 14 (2019) P04006]

[LHCB-TDR-018]



the Run3 raw event size is ~150 kB

ALIGNMENT AND CALIBRATION

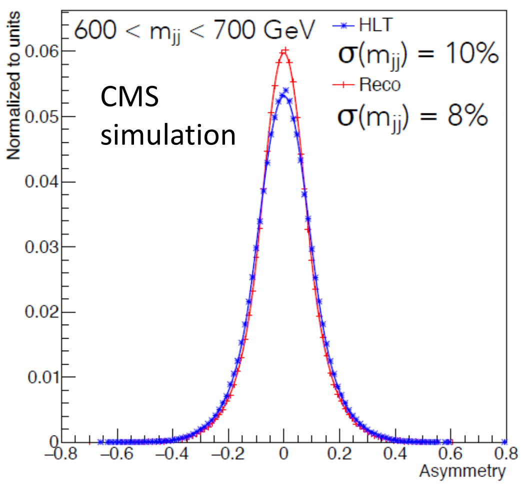


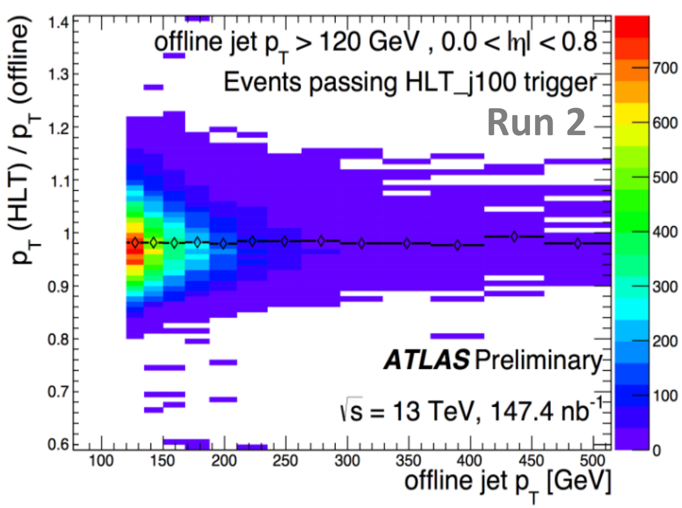
ALIGNMENT AND CALIBRATION

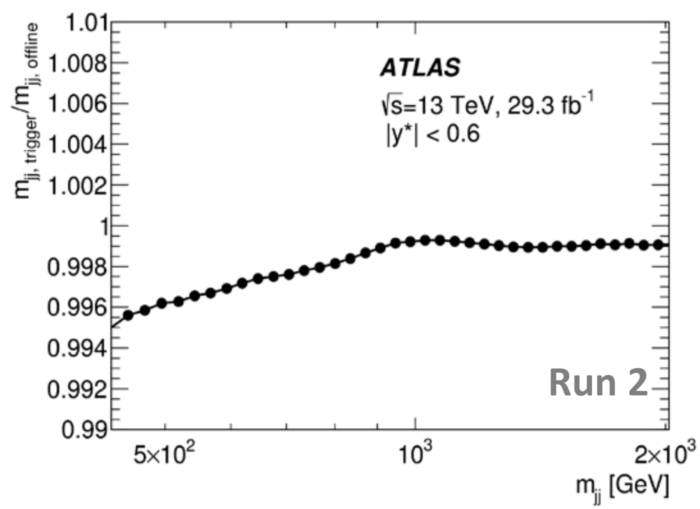
How well can we reconstruct the objects at trigger level ?

Tracking and calorimeter object reconstruction is less sophisticated as compared to the “offline” [Processing timing ~ 0.1 s vs 10 s at CMS]

→ Less iterations in tracking reconstruction and using smaller regions
→ Trigger jets are corrected to the scale of “offline” reconstruction jets
(using simulation and calibrated objects from data (ex: $Z \to \ell^+\ell^-$, photons, multijets)



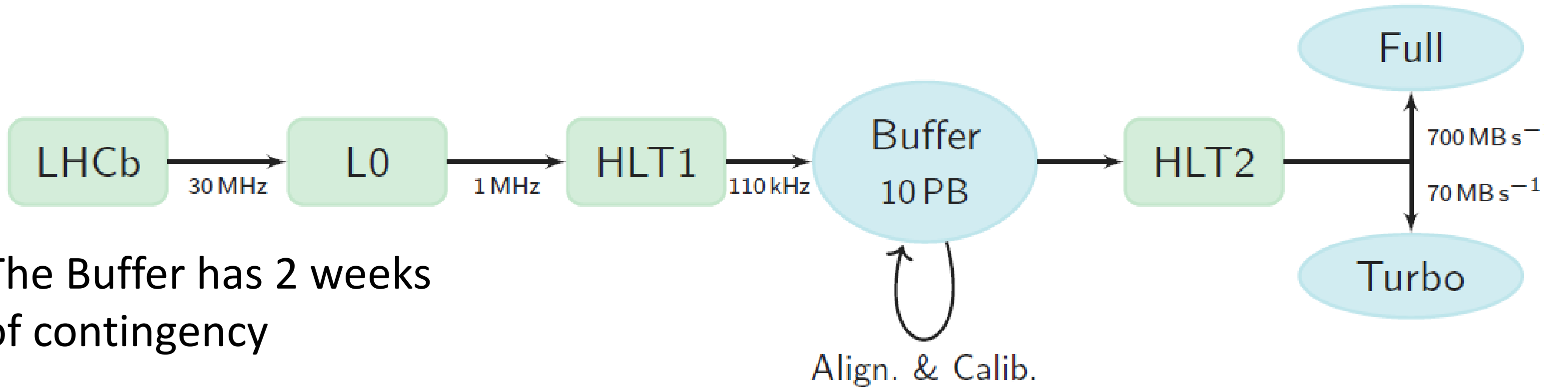




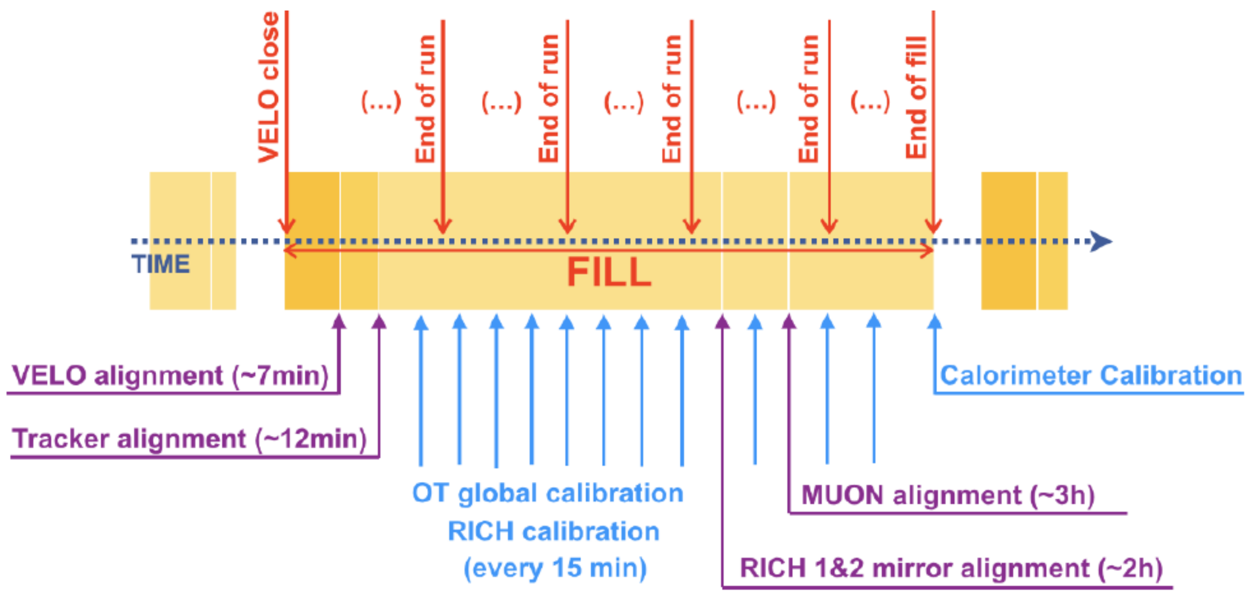
Good agreement between trigger and offline jet reconstruction and ATLAS and CMS

ALIGNMENT AND CALIBRATION

At LHCb all detectors are **aligned & calibrated online** using a selected set of events from the HLT1 output, stored in a 10 PB Buffer, before the data go to HLT2



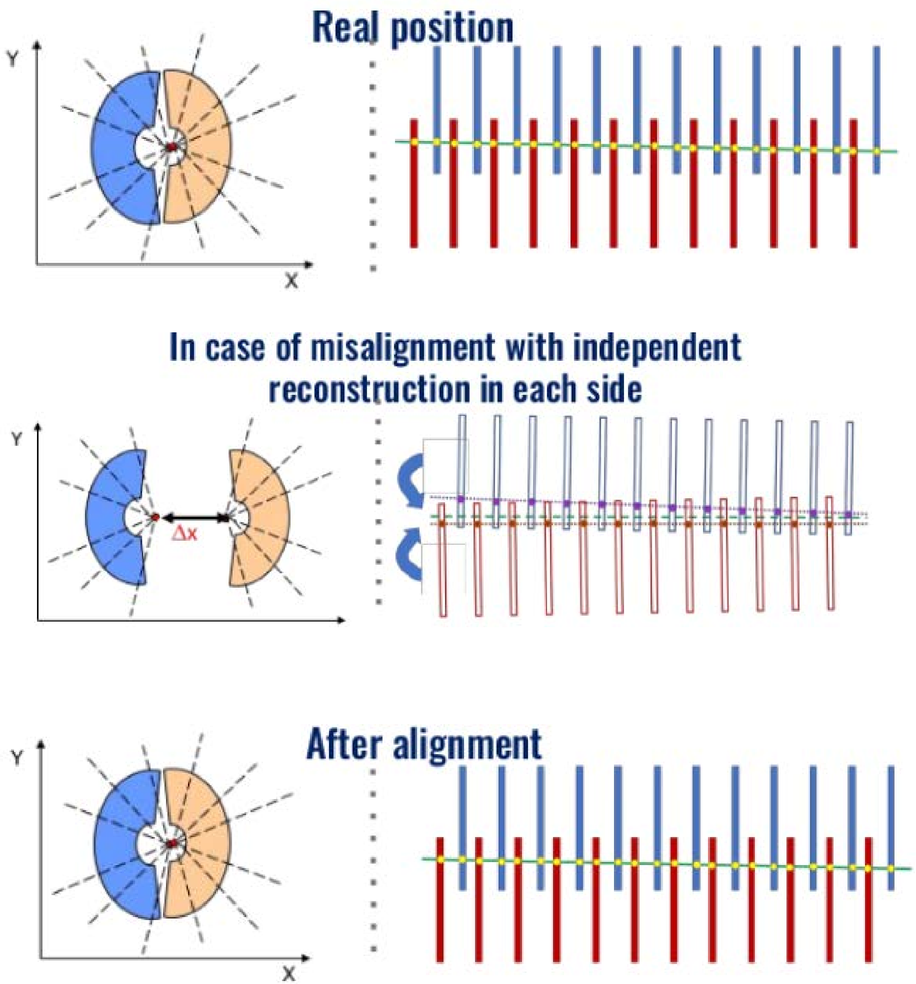
The Buffer has 2 weeks of contingency

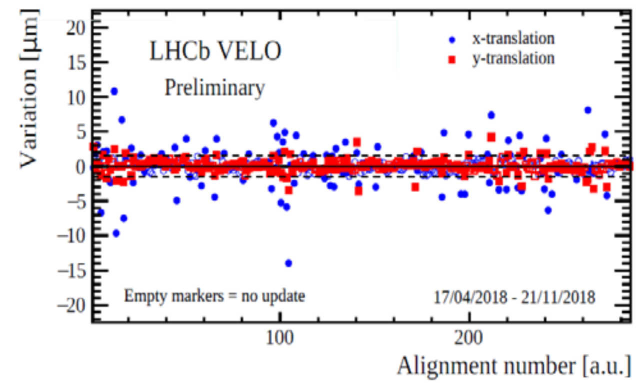


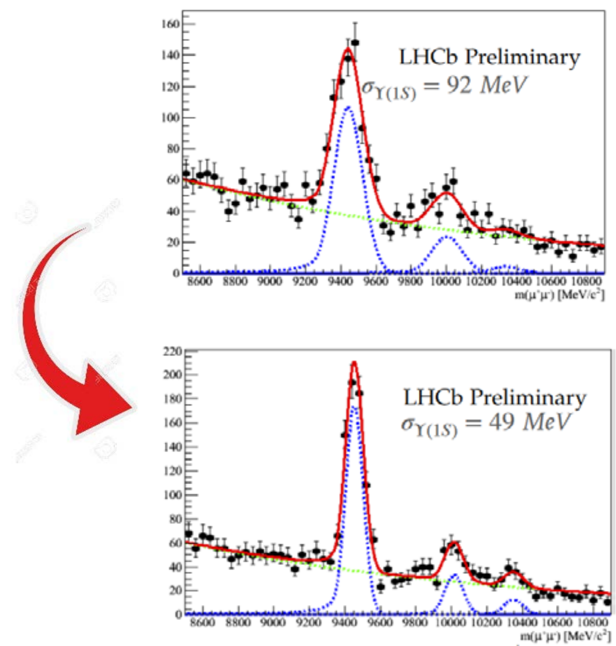
(...) - time needed for both a data accumulation and running the task

ALIGNMENT AND CALIBRATION

Ex: at LHCb the VELO centers itself around beam at start of each fill, aligned with a Kalman filter using track hit residuals with PV constraints



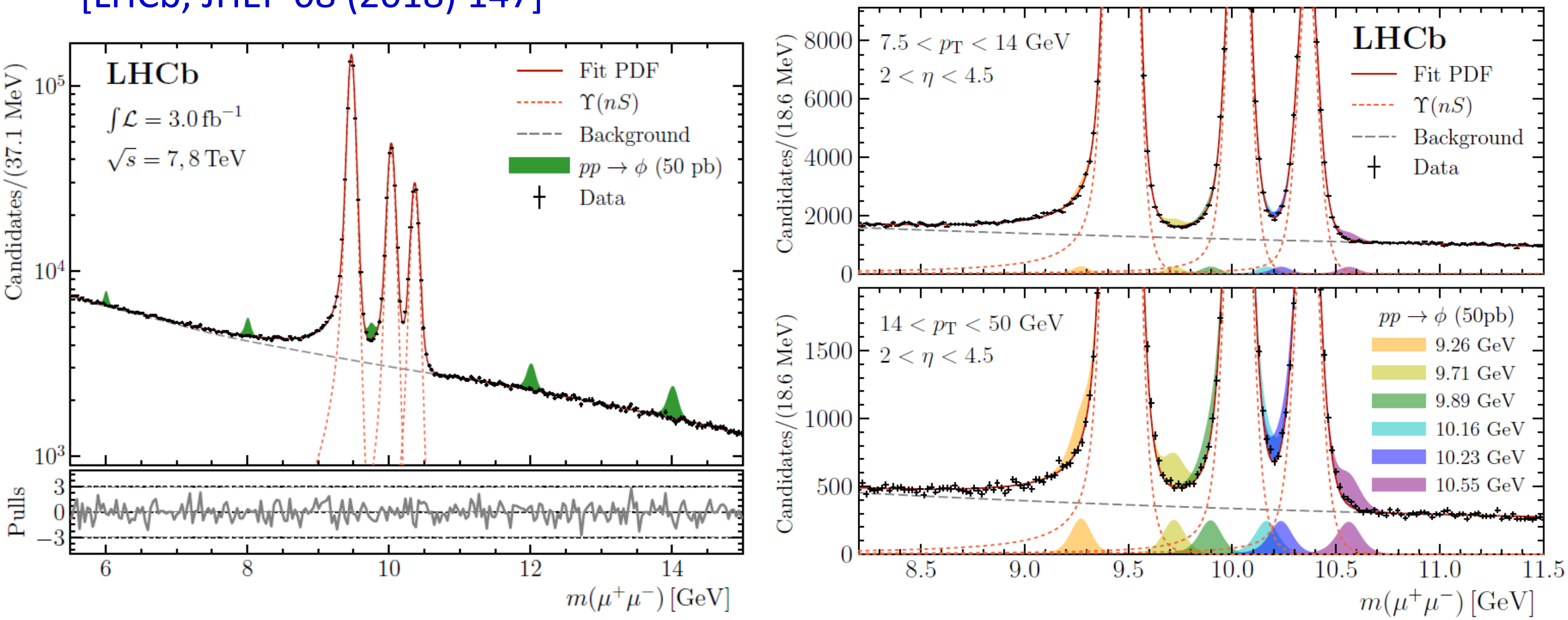




ALIGNMENT AND CALIBRATION

★ **Search for dimuon resonance (new spin-0 bosons)**

[LHCb, JHEP 08 (2018) 147]



- ▶ Fully aligned and calibrated physics objects in real time → allow to perform analysis at the same level as the offline
- ▶ Data reprocessing not needed → fast and fresh analyses, results delivered in few days
- ▶ Reduced systematics in HLT2 selections

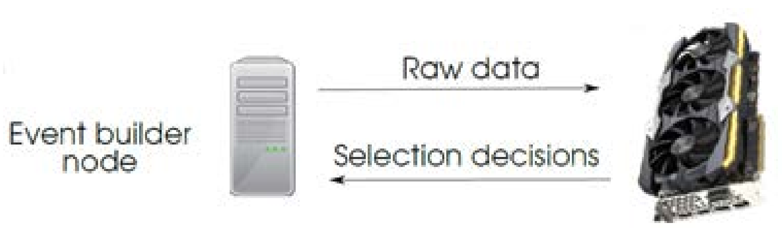
USING ACCELERATORS

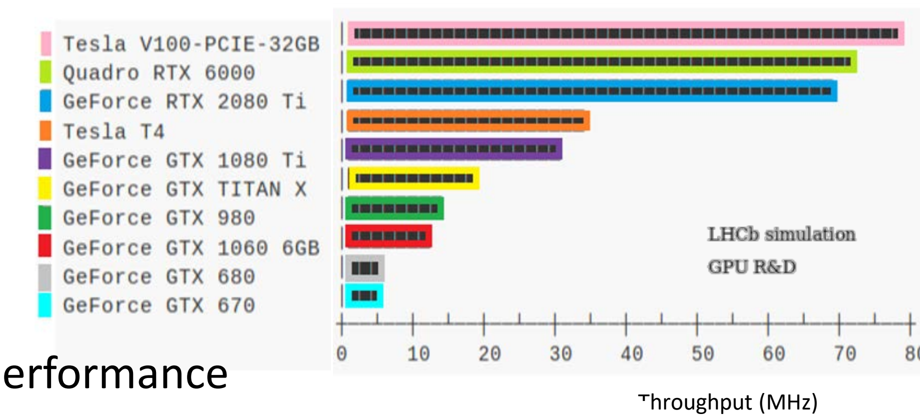
USING ACCELERATORS

Timing is an issue… how can we be faster?

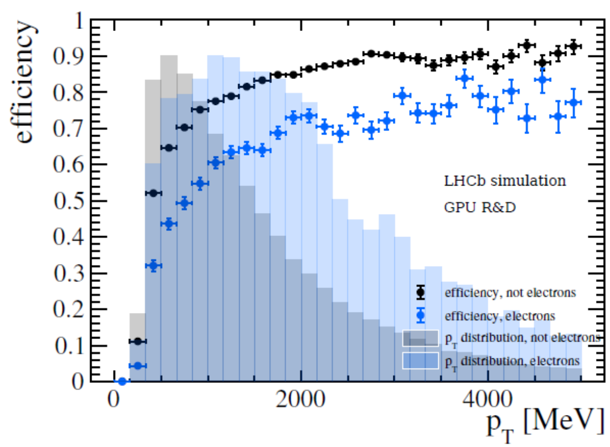
See talk by D. von Bruch (plenary tomorrow) and D. Cámpora (yesterday @ Track5)

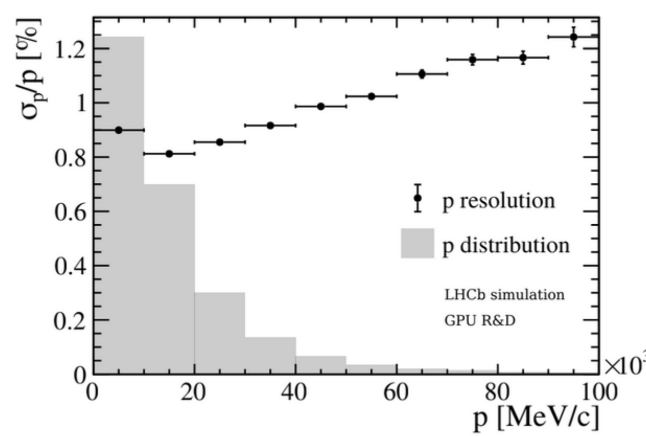
Ex: **Allen:** project at LHCb: running the entire LHCb HLT1 software on GPUs





→ High throughput with good physics performance





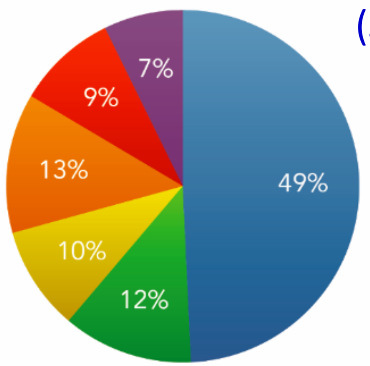
GPUs also at ALICE (next plenary and D. Rohr today @ TrackX), CMS and ATLAS (yesterday by Z. Chen and A. Bocci @ Track5)

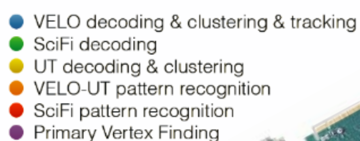
USING ACCELERATORS

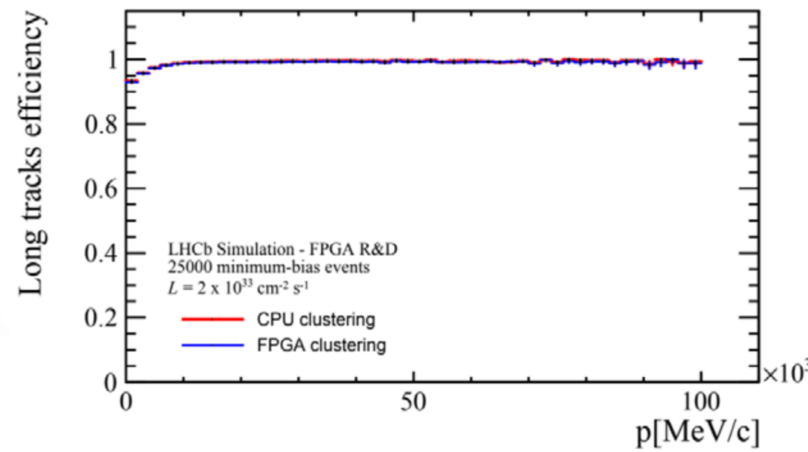
FPGAs can also be used to accelerate repetitive simple tasks before the HLT decision

Ex: at LHCb VELO clustering is being implemented on the PCIe40

(See talk by G. Tuci @ TrackX today)

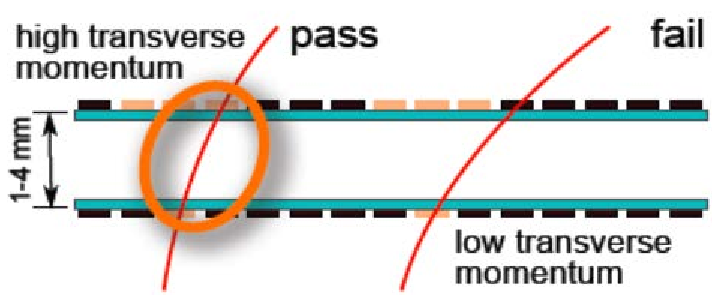


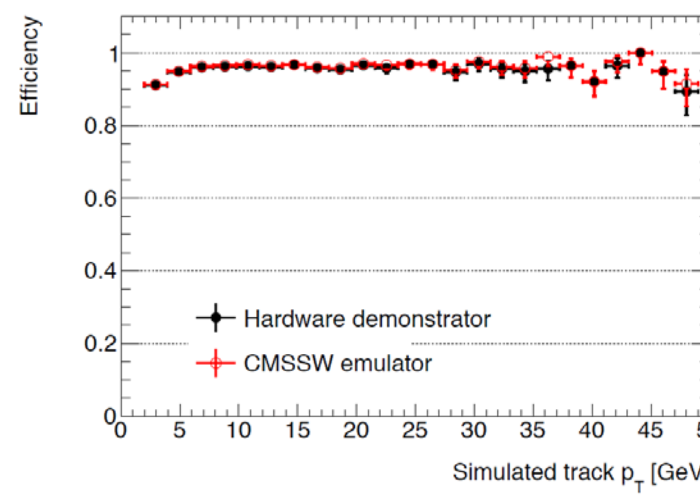




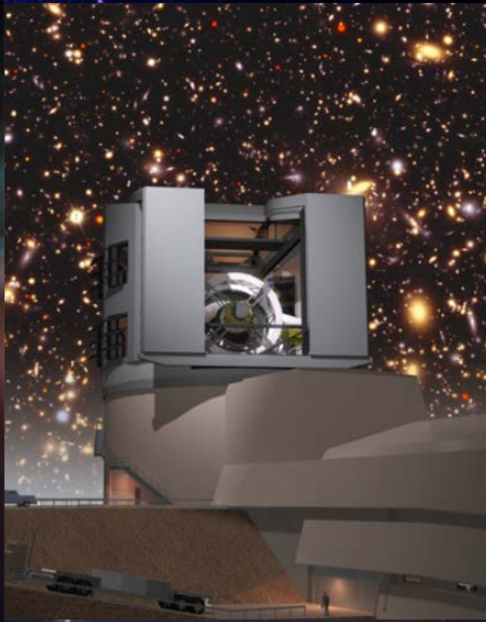
Ex: track finder for the L1 at CMS

[R. Aggleton et al JINST (2017) 12 P12019]





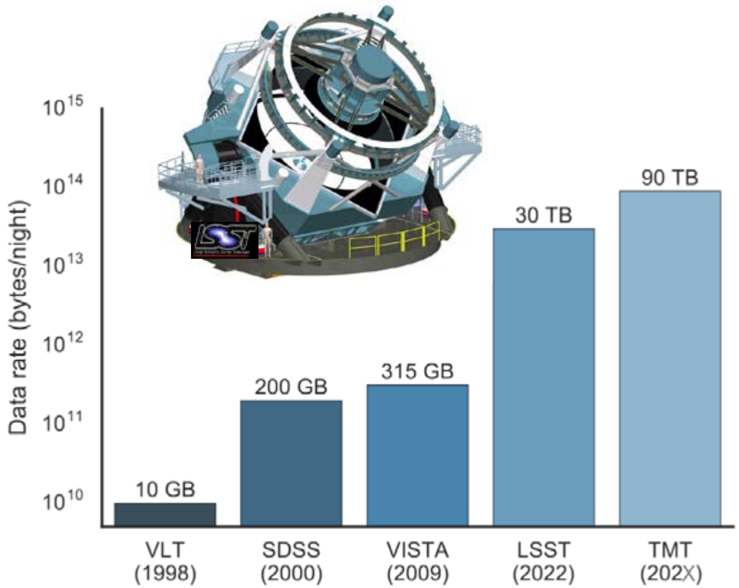
CONNECTIONS TO OTHER EXPERIMENTS AND FIELDS



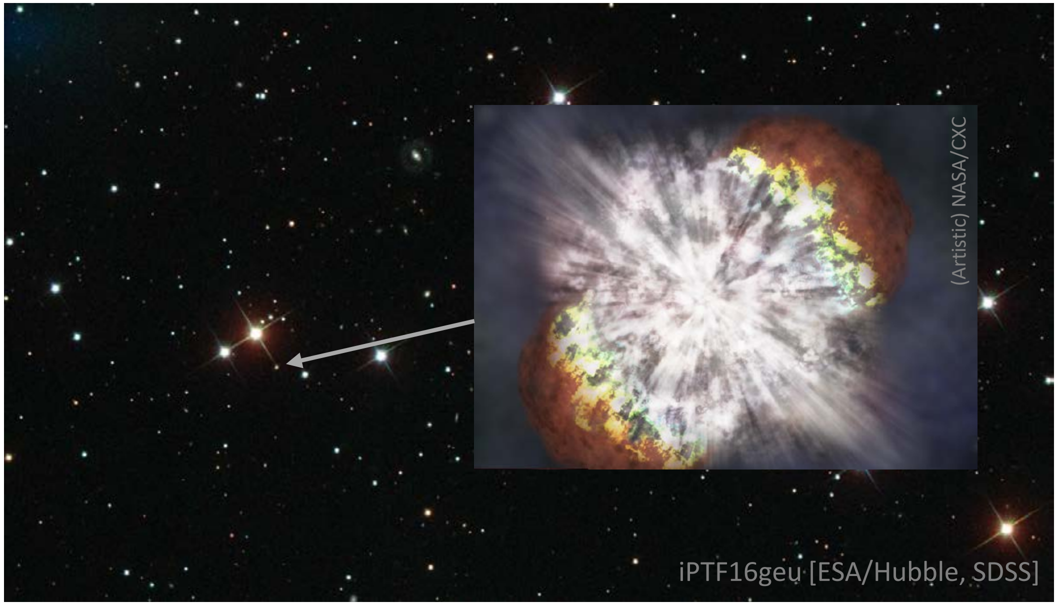
ASTRONOMY & COSMOLOGY

Large Sky Surveys aim to discover *transient* objects that change brightness over time-scales of seconds to months

→ Cosmic explosions:
supernovae, gamma-ray bursts…
→ Relativistic phenomena:
black hole formations, jets…
→ Potentially hazardous asteroids
etc…



[arXiv:1704.04650]



iPTF16geu [ESA/Hubble, SDSS]

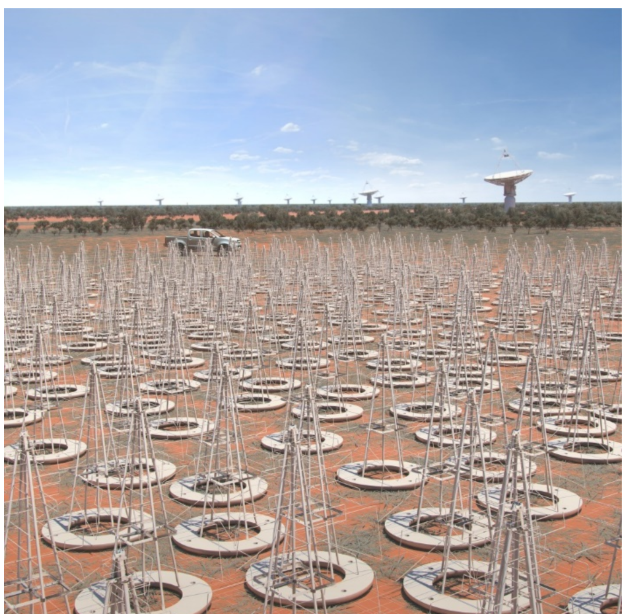
The Large Synoptic Survey Telescope (LSST) will take 800 images (30TB data) and will see millions of transient objects per night
→ need to alert in real time to other telescopes and instruments (latency ~60s)

Classification before taking decisions is crucial

ASTRONOMY & COSMOLOGY

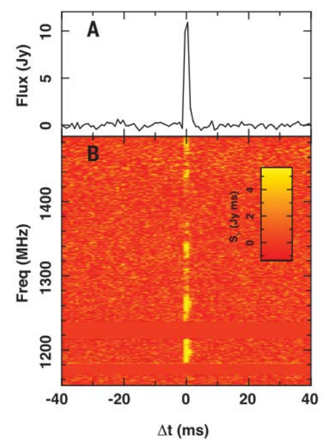
Not only optical and IR photometry, also in **radio astronomy**:
Ex: fast radio bursts, coming from distant astronomical sources last less than a millisecond

Square Kilometer Array (SKA) will receive 1 EB/s (raw data) from 2026 (EB = 10^6 TB)



130K low frequency antennas in Australia

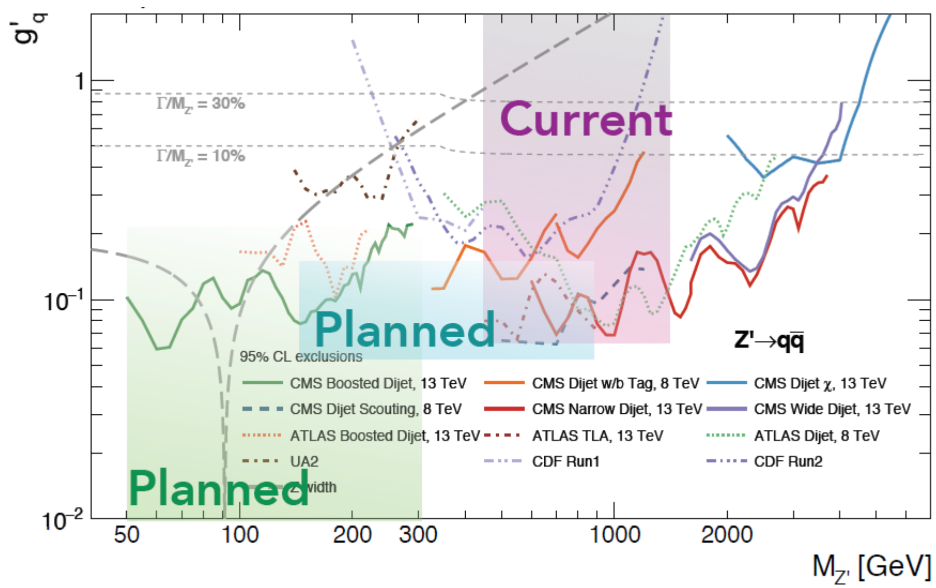
FRB 18092 seen by ASKAP
[Science 365 (6453), 565-570]



- Optical fibre-based synchronisation distribution system
- Calibration and data reduction (100 GB/s rate per site)
- Extensive use of ML techniques for clasification

CONCLUSIONS

- Data volumes and complexity are exponentially increasing
- Impossible to record all raw data within the available resources
- **Real Time Analysis** is crucial to analyse and take a decision to keep the key data
- It allows for extending the phase space regions for discoveries



- But we have to be also very careful with what we dismiss…