Challenges and opportunities for trigger-level analyses

Antonio Boveia (a personal view)
6 October 2020

supported in part by US DOE grant DE-SC0011726

What is trigger-level analysis?

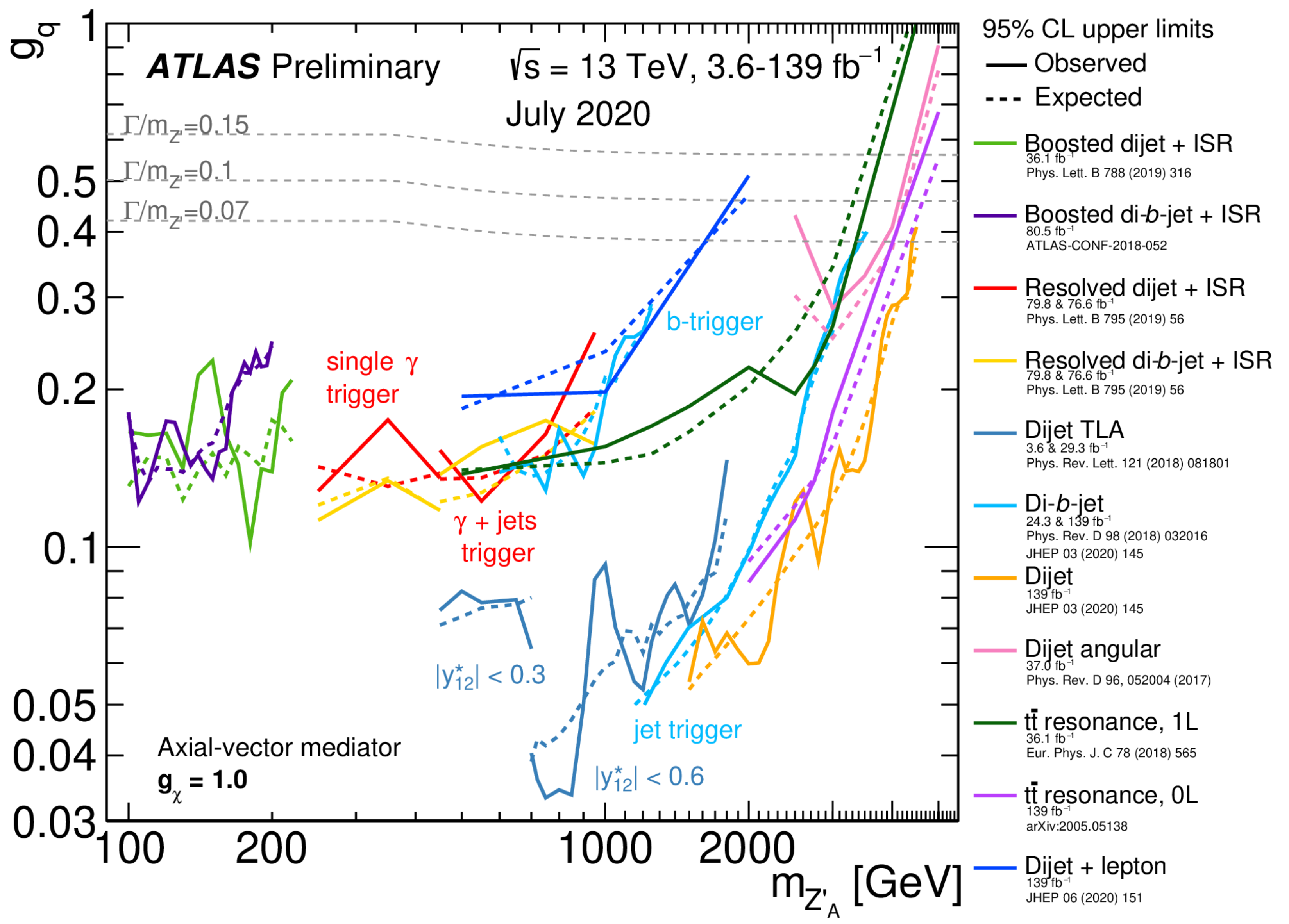
The trigger is a fact of life—we cannot record all of collision data, even if we wanted to do this.

In traditional "offline" analysis, we see only the events that survived the trigger, and live with the limitations. For example, **p_T thresholds** are sometimes higher than we would like; we may **miss a discovery** because we don't trigger on the events; etc. At ATLAS and CMS, these limitations prevent an exhaustive exploration of the electroweak scale!

However, the trigger system sees all the events. It performs real-time (or almost-real-time) analysis. In recent years (and especially for future upgrades), even the initial stages of these systems are becoming powerful enough to think about using them for serious analysis without reducing analysis sensitivity or introducing biases.

This allows real-time (or "trigger-level") analysis, where the trigger does not simply perform a decision but instead records an intermediate data reduction (trigger-level objects or partial event data) or even the final observables.

Initial implementations of this idea are already in use with ATLAS ("TLA"), CMS ("data scouting"), and LHCb ("turbo stream", "Tesla") and have been used for physics publications, primarily focused on the last stage of the trigger, reducing the data read out for each event to increase the event rate (at fixed bandwidth).



Opportunities for trigger-level analyses

These first publications are “proof of principle” and are now being followed up with more ambitious ideas.

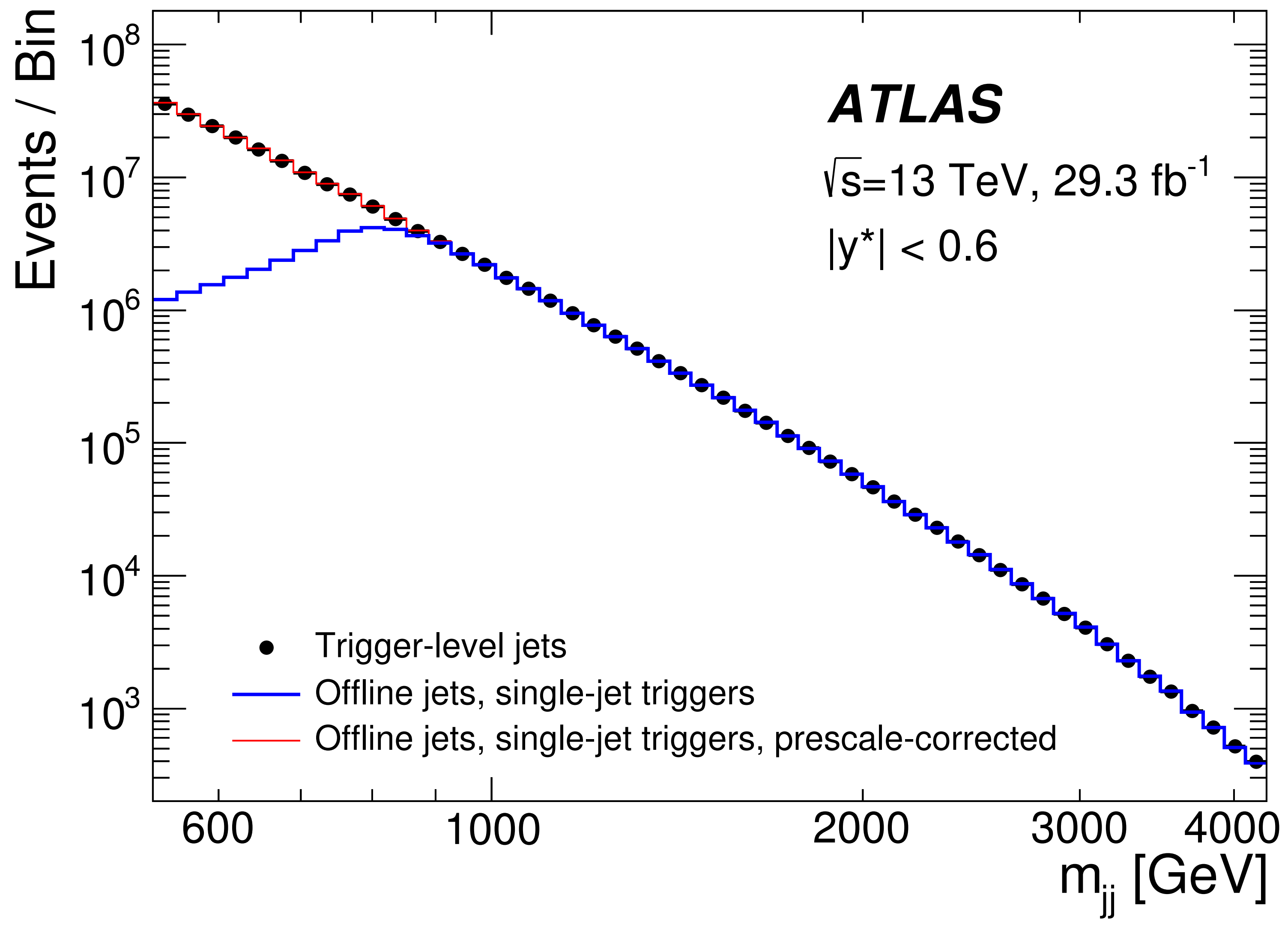
In the near-term (LHC Run 3), improvements to the trigger hardware and software make new things possible. For example, ATLAS will have a **larger online CPU farm** capable of running **tracking** for a much larger rate of events, and have **partial readout of full-detector information** around regions of interest. This will allow better **pile-up mitigation**, object **calibration** and resolution (e.g. jets), blended on-line and offline analysis (e.g. b- tagging), and other improvements. These will allow even **lower momentum thresholds** for the analyses done so far, and make **new channels** possible (e.g. low-mass scalar resonances decaying to jets, photons).

For the HL-LHC, more sophisticated trigger hardware (such as “global” trigger hardware, hardware track processors) make **bypassing even the hardware-level trigger** decision a possibility. For example, CMS is planning a 40 MHz scouting for the L1 system in Phase-2 and exploring dedicated hardware for ML-based anomaly detection.

These analyses are attractive projects especially for hardware and software developers, who are best positioned to take advantage of over-dimensioned systems.

Experimental challenges for trigger-level analysis

PRL 121 081801



Huge backgrounds and small signal **can require precision control** of all aspects of the analysis. This is difficult to achieve in the online environment, especially in a hardware system.

Real-time analyses requires a **separate data handling pipeline** (custom reconstruction for partial data, data quality, calibrations, analysis framework, etc.)

Advances are needed in reliable **machine learning** application in the trigger, especially for analysis. Offline reconstruction relies more and more on ML (e.g. b-tagging, tracking?). What aspects of this can migrate to the online environment, where calibrations may not be as sophisticated? How to ensure a strong correlation with offline algorithms?

Further work is needed on **compression**, flexible custom **data formats**, and toolkits for real-time detector **calibration**. It needs to be easier for **non-expert offline analysts** to design and deploy these techniques without the deep expert knowledge that they currently require.

Can we someday port a full offline analysis chain into the trigger and readout?

Further reading

HEP Software Foundation, Software and Computing R&D for the 2020s

J. Albrecht, et al., HEP Community White Paper on Software trigger and event reconstruction

A. Boveia, Trigger level analysis technique in ATLAS for Run 2 and beyond, CHEP 2019

H. Sakulin, 40 MHz Level-1 Trigger Socuting for CMS, CHEP 2019

ATLAS Collaboration, Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System

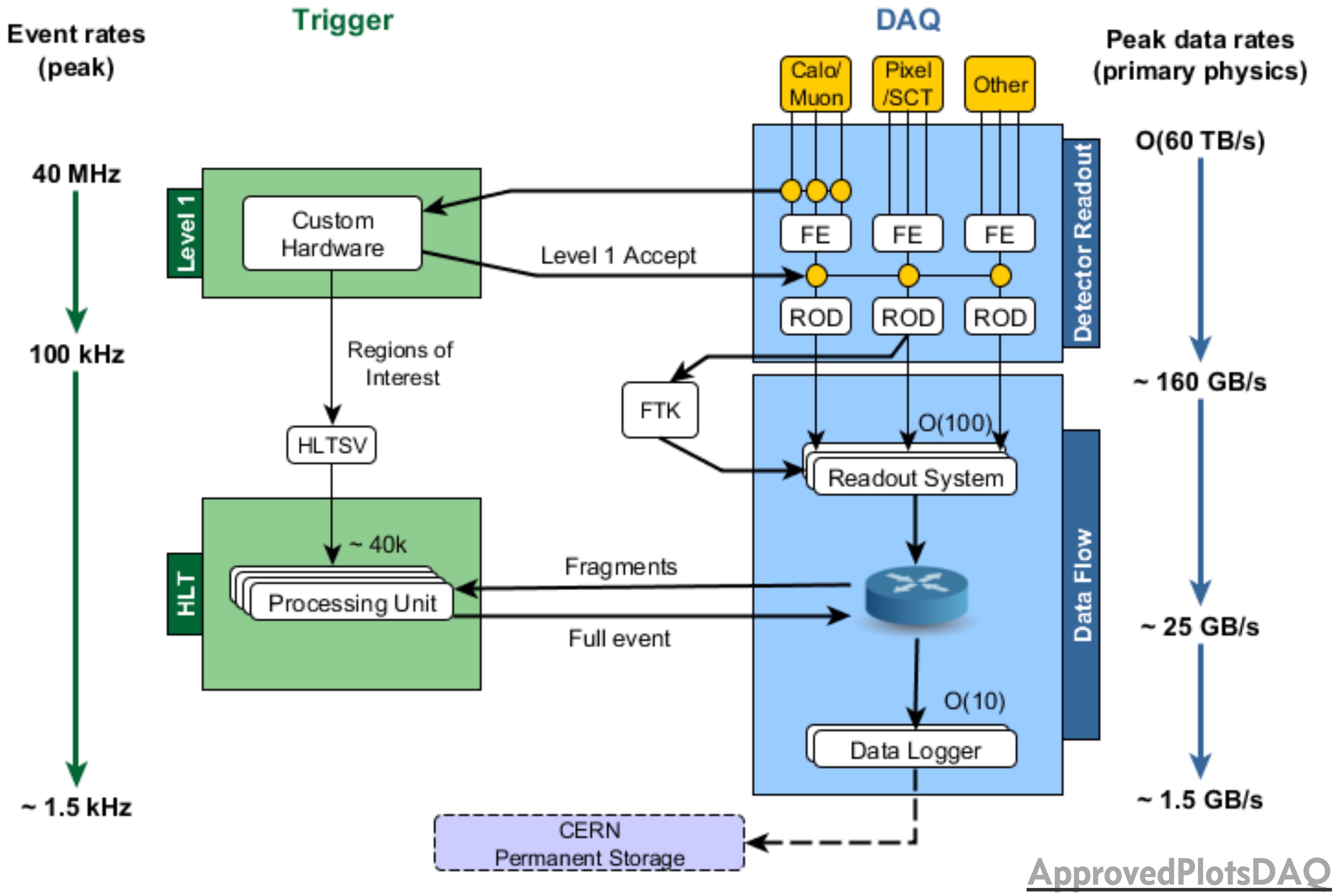
CMS Collaboration, The Phase-2 Upgrade of the CMS Level-1 Trigger

Additional slides

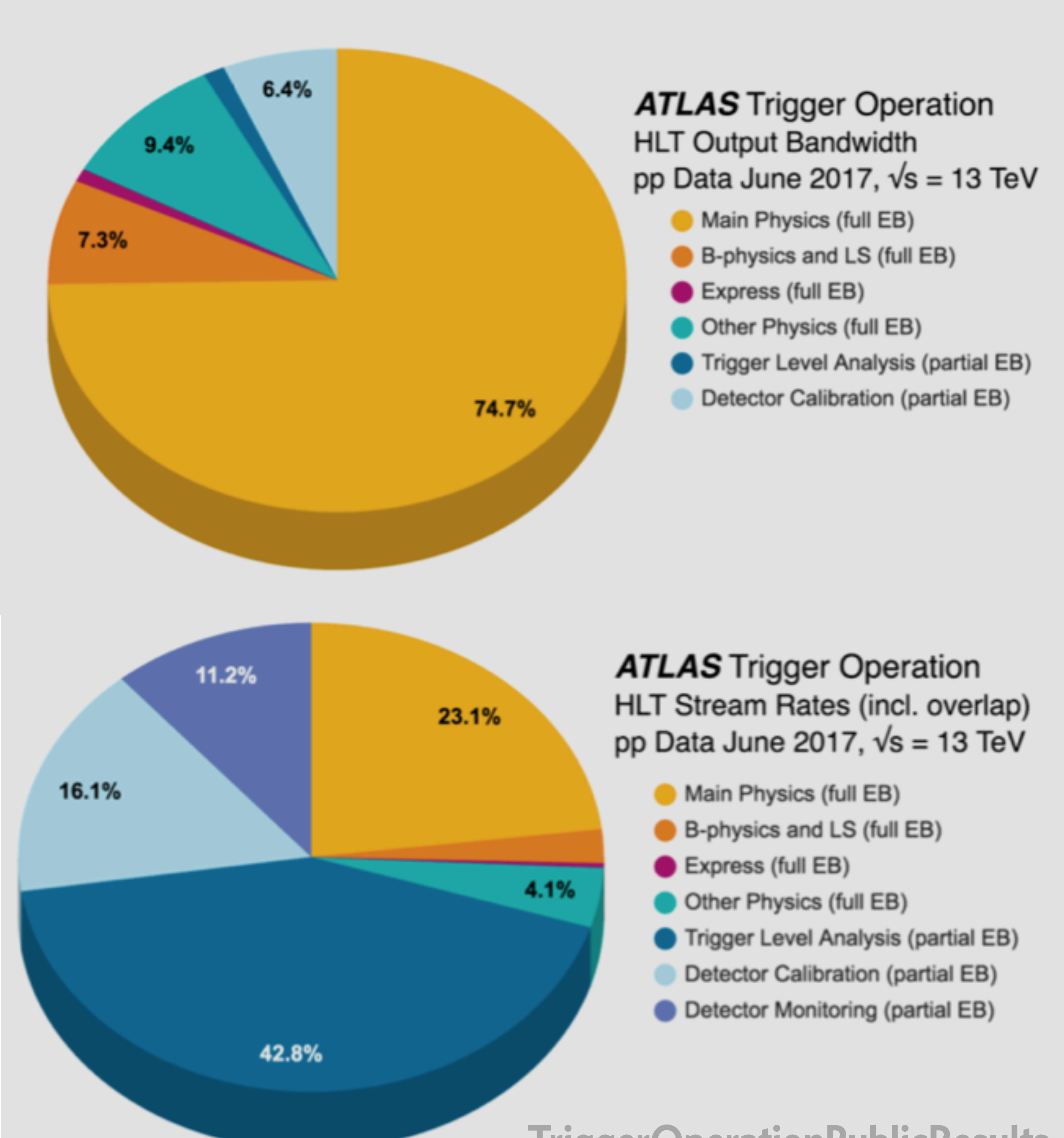
Overview of ATLAS trigger system during Run-2

The trigger system for ATLAS during Run 2 consisted of a **L1 hardware system** (accepting 100 kHz) and an **HLT software system** (accepting ~1 kHz of physics triggers).

Along with other **upgrades** to the L1 system, it also featured a **L1Topo processor** (allowing topological algorithms such as selection on angular distance between two L1 jets) and an **upgraded CTP** (providing e.g. more room for topo- and analysis-specific L1 items).



ApprovedPlotsDAQ



TriggerOperationPublicResults

ATLAS trigger menu largely driven by inclusive triggers generically useful to many analyses and recorded in a **"main" stream**. Average **1 kHz** and **1 MB/event**.

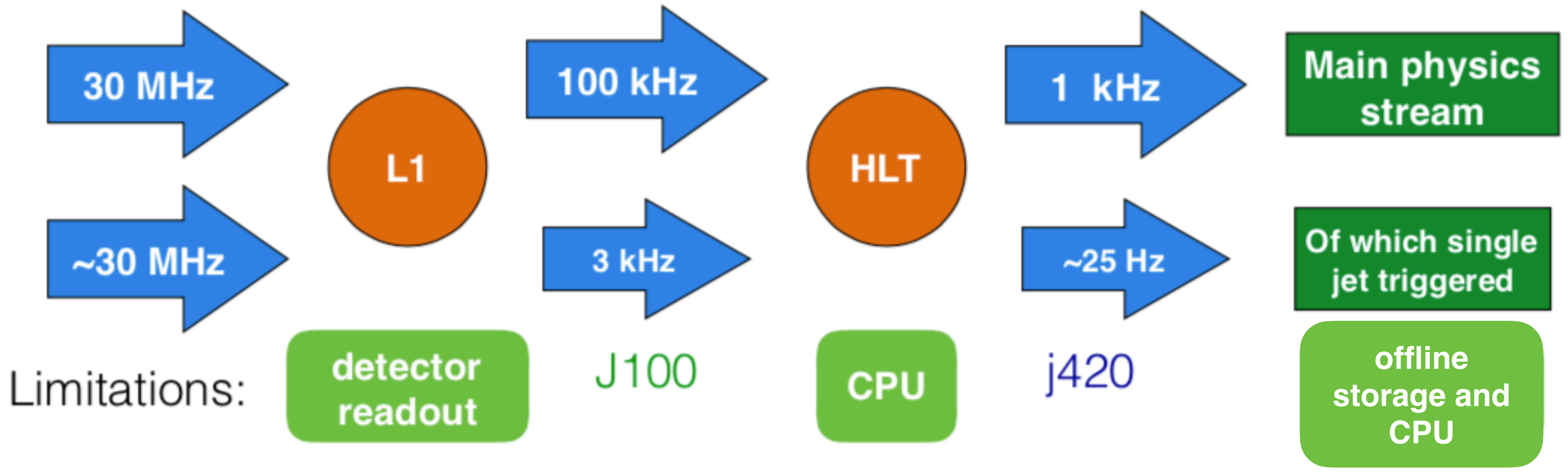
Additional **flavour physics streams**: dedicated triggers, can use delayed/custom reconstruction, or partial-event readout (e.g. only subdetectors in 1.5×1.5 area around a track satisfying pre-selection). Non-PE stream averages **200 Hz** and **1 MB/event**.

Trigger-Level Analysis stream: stores HLT reconstruction only. Discussed in this talk. Recorded up to **26 kHz** peak rate at an average **5 kB/event**.

In 2018, 32 streams total: about half with full event information, half with partial event building (PEB).

Trigger menu limitations during Run 2

Main menu limitations are **L1 rate** (multi-jet, taus, flavour physics), **HLT CPU** (b-tagging of low-p_T jets), and **HLT rate** (most triggers).

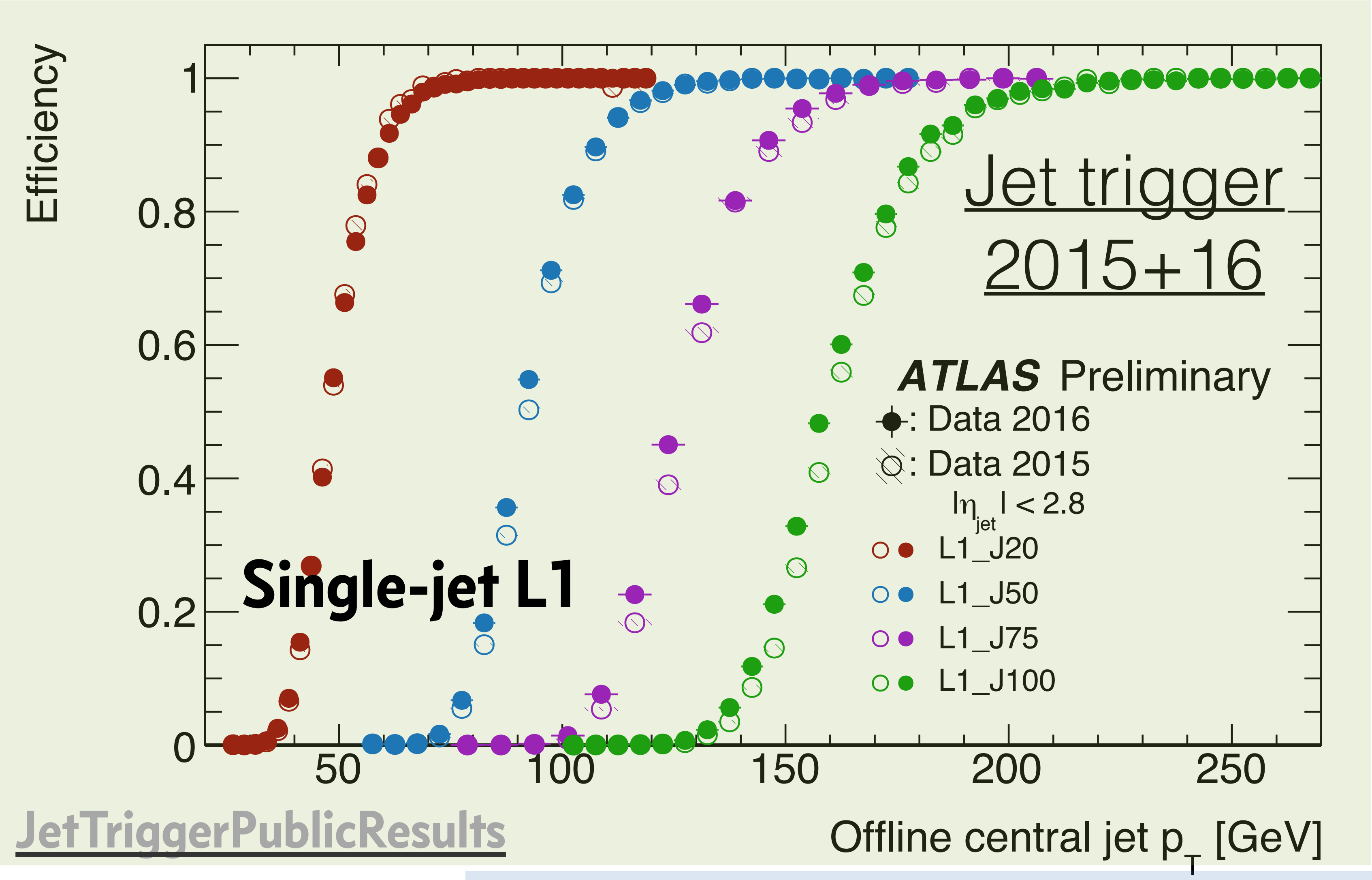


L1

Readout electronics set a hard limit of **100 kHz**.

Peak rate ~**95 kHz**.

Strong production (multi-)jet and flavour-physics triggers would quickly saturate this, without additional requirements (e.g. single-jet p_T thresholds)



HLT CPU

Processing power of HLT farm sets hard limit on what reconstruction can be run

Typically: pre-selection then offline-like (but speed-optimized) reconstruction

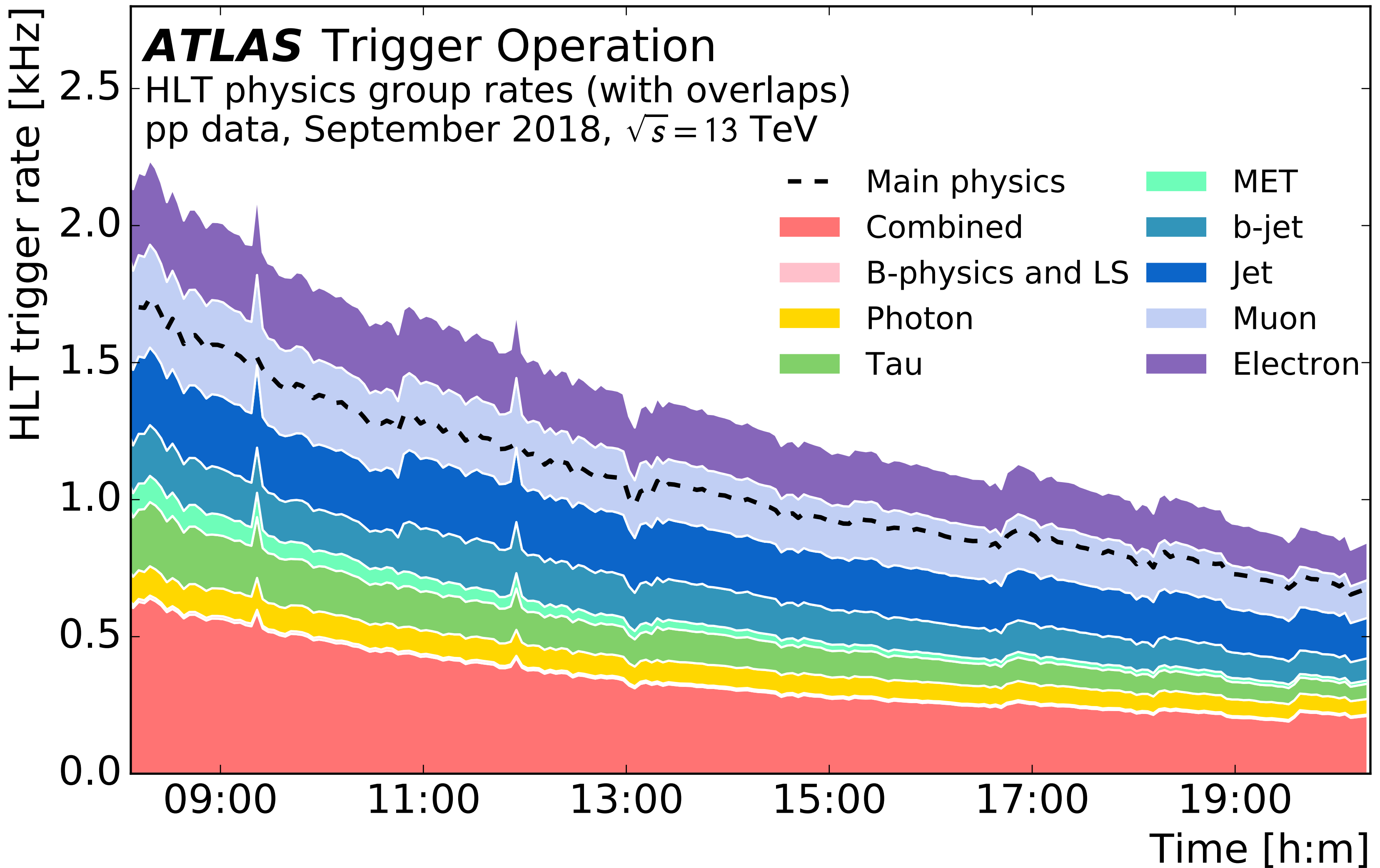
In particular, **tracking is not performed for jet triggers** (and for low-p_T b-jet candidates)

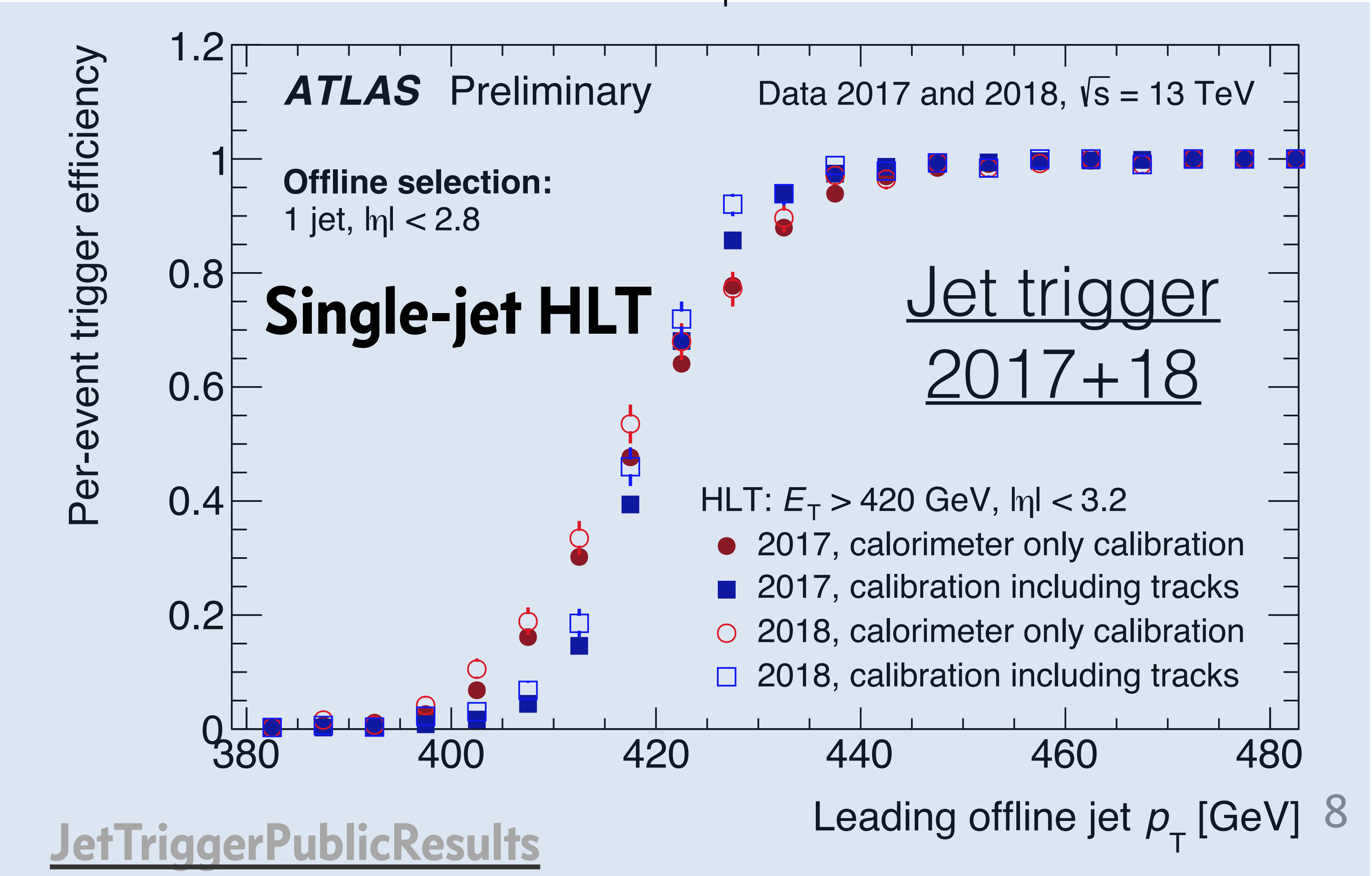
HLT rate

Soft limit of average **1 kHz** from data storage, processing, and maintenance needs

Jet triggers ~15% of total (~150–250 Hz)

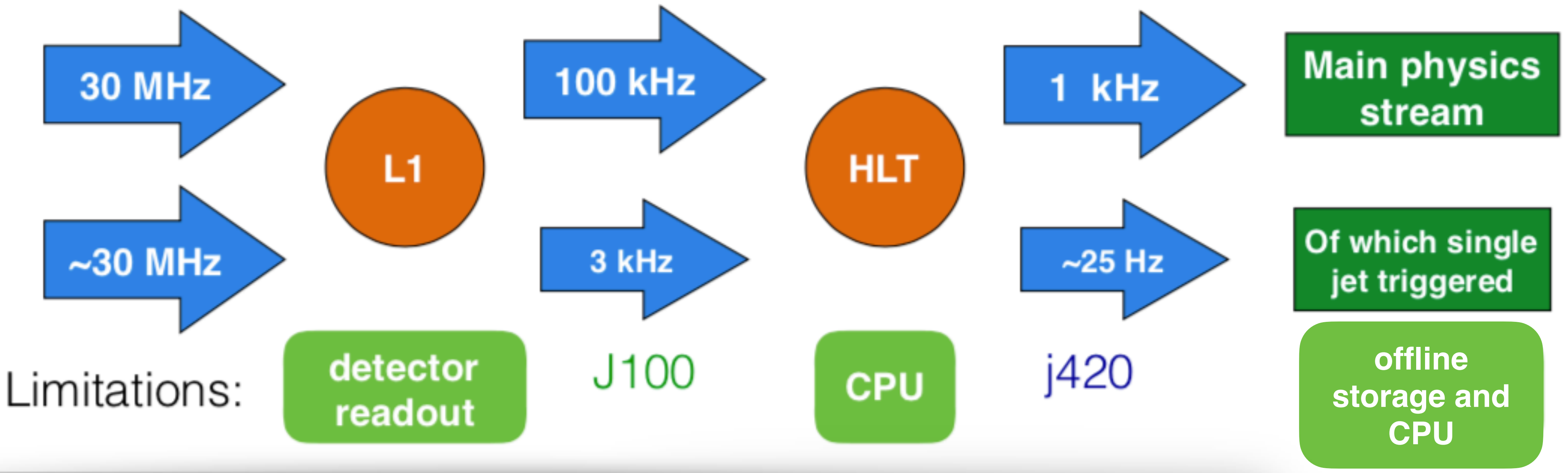
Single-jet triggers only unprescaled and fully efficient for offline p_T>~**440 GeV**





Trigger menu limitations during Run 2

Main menu limitations are **L1 rate** (multi-jet, taus, flavour physics), **HLT CPU** (b-tagging of low-p_T jets), and **HLT rate** (most triggers).



L1

Readout electronics se

Peak rate ~**95 kHz**.

Strong production (m saturate this, without

HLT CPU

Processing power of H

Typically: pre-selection

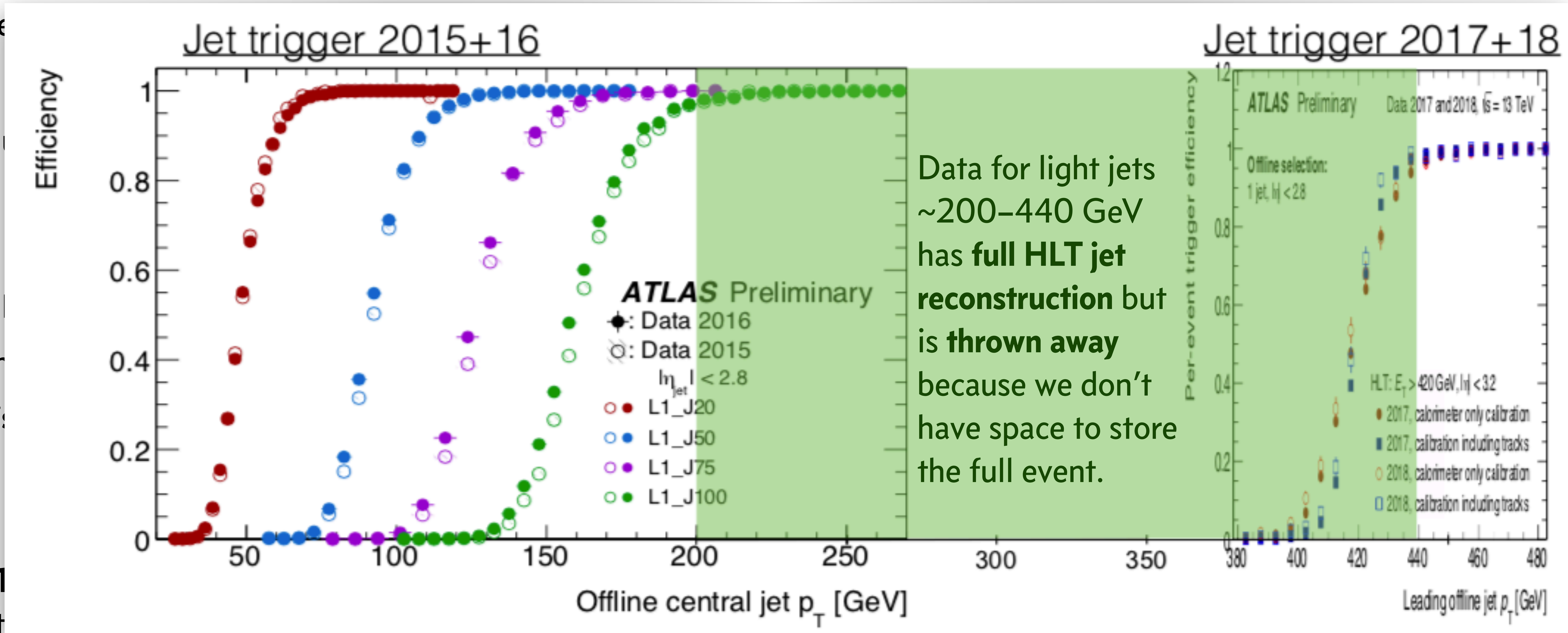
In particular, **tracking i** candidates)

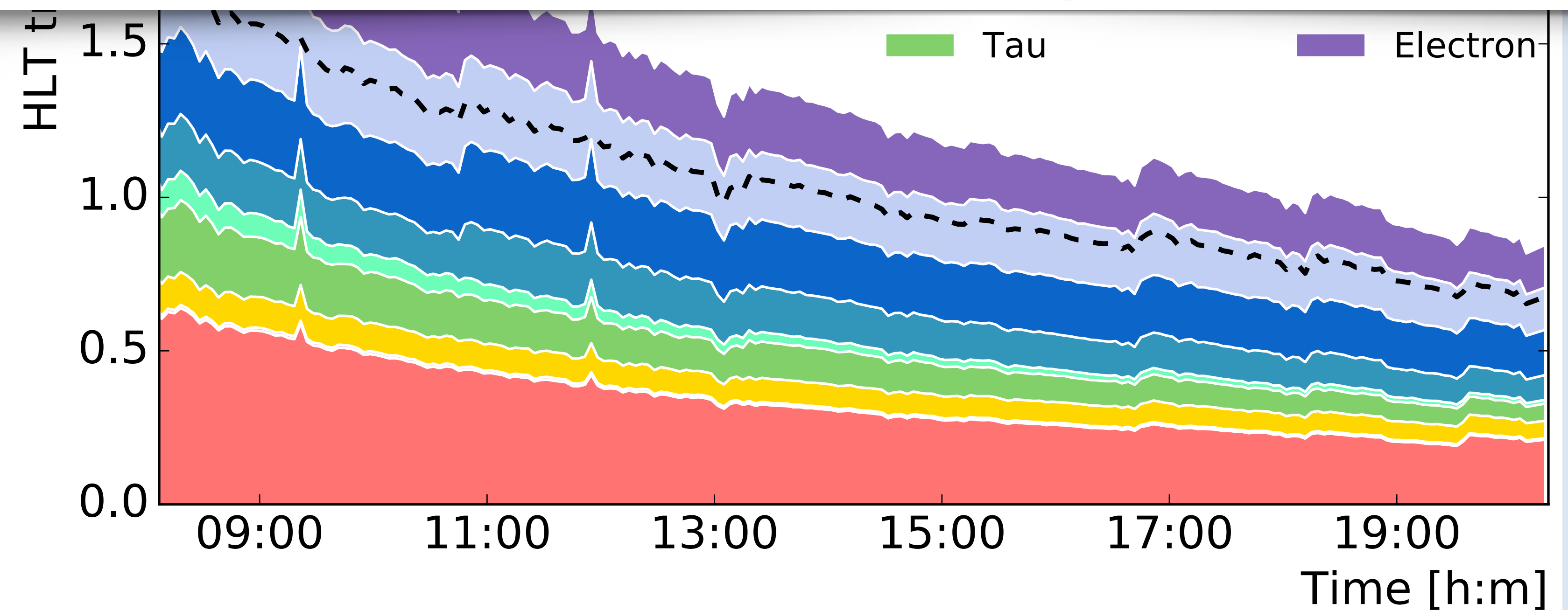
HLT rate

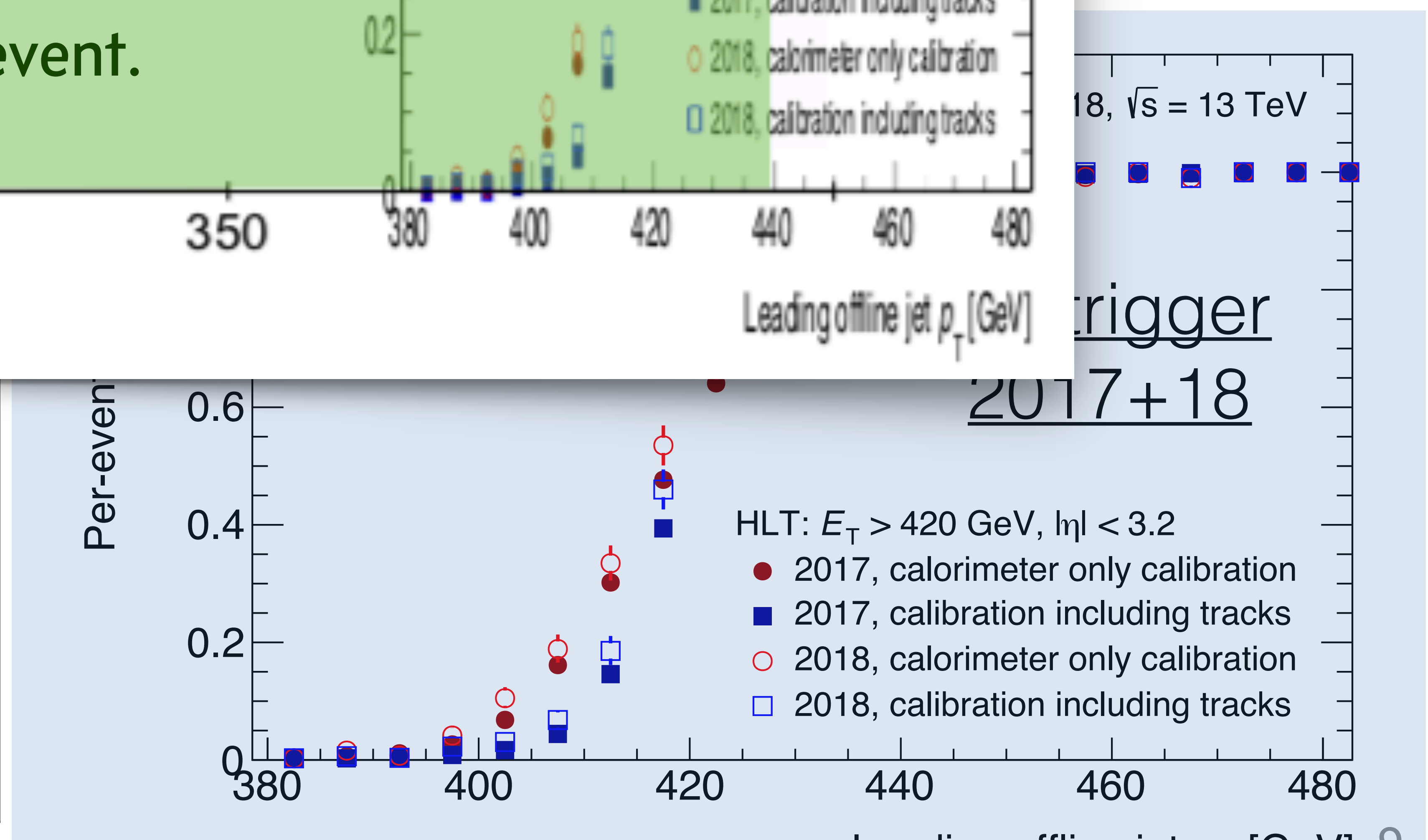
Soft limit of average 1 processing, and maint

Jet triggers ~15% of total (~150–250 Hz)

Single-jet triggers only unprescaled and fully efficient for offline p_T>~**440 GeV**



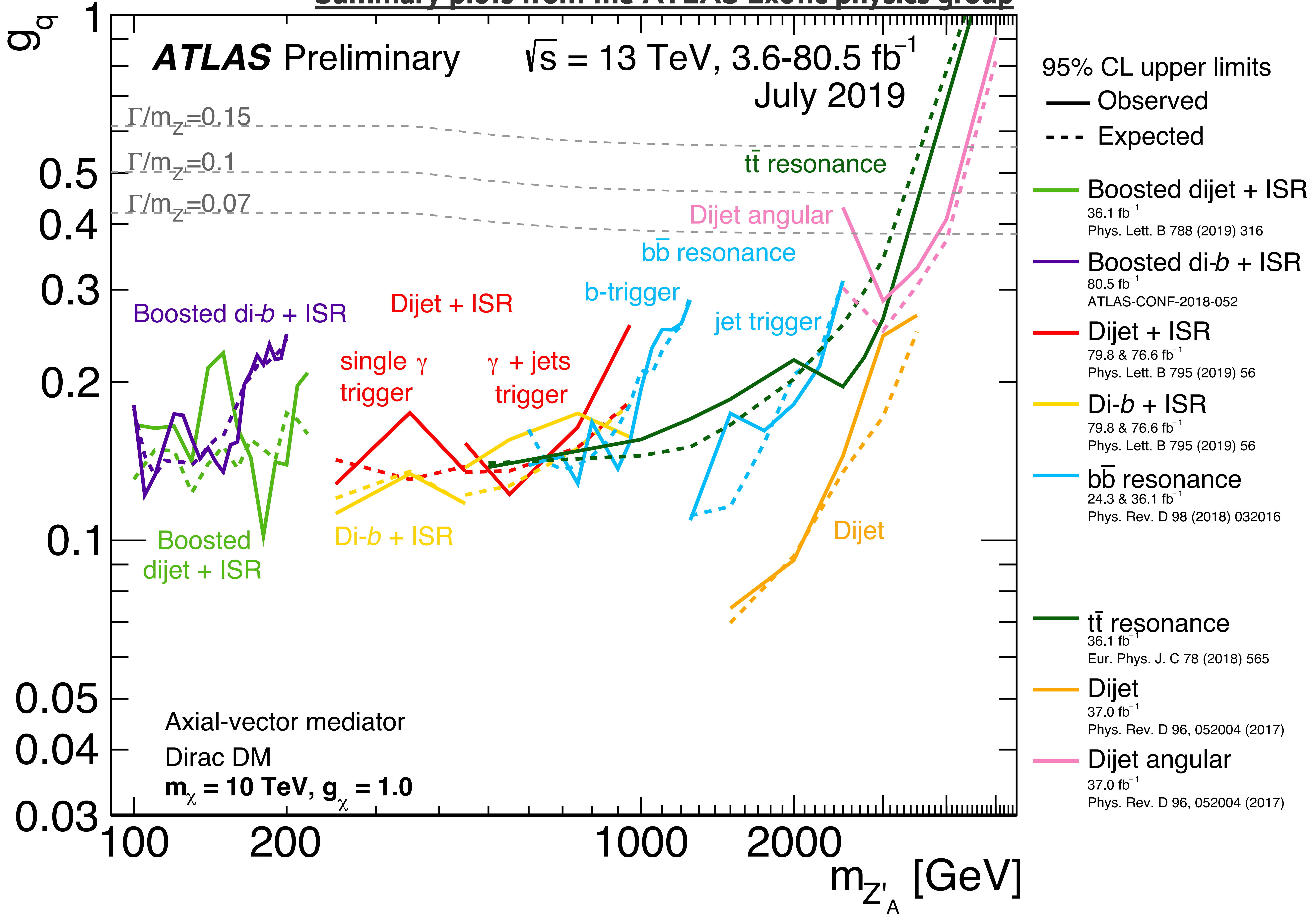




JetTriggerPublicResults

Why bother with "low p_T" jet data?

Summary plots from the ATLAS Exotic physics group



Two-body resonances are a historically fruitful search channel (J/psi, Z, Higgs) and a key component of the ATLAS search program. They are well-covered for most types of decays.

However, the HLT threshold for the single jet trigger (440 GeV) constrains dijet searches to the region **m_{jj}>~ 1.5 TeV** (~2x p_T).

The electroweak–TeV scale is special! The W, Z, Higgs, and top are all found there. We must study it as thoroughly as we can. Not even SM-like couplings (few * 0.01) are reached by the most sensitive search.

With a variety of alternate triggering strategies or more narrowly targeted searches, ATLAS can cover a wider range of dijet masses, but with **much less statistical power** than the full data would allow.

We have to do better!

Trigger-level analysis

To generically probe the entire range of EW–TeV dijet resonances with the full statistical power of the data, we need to work around all three trigger limitations (L1, HLT CPU, HLT rate).

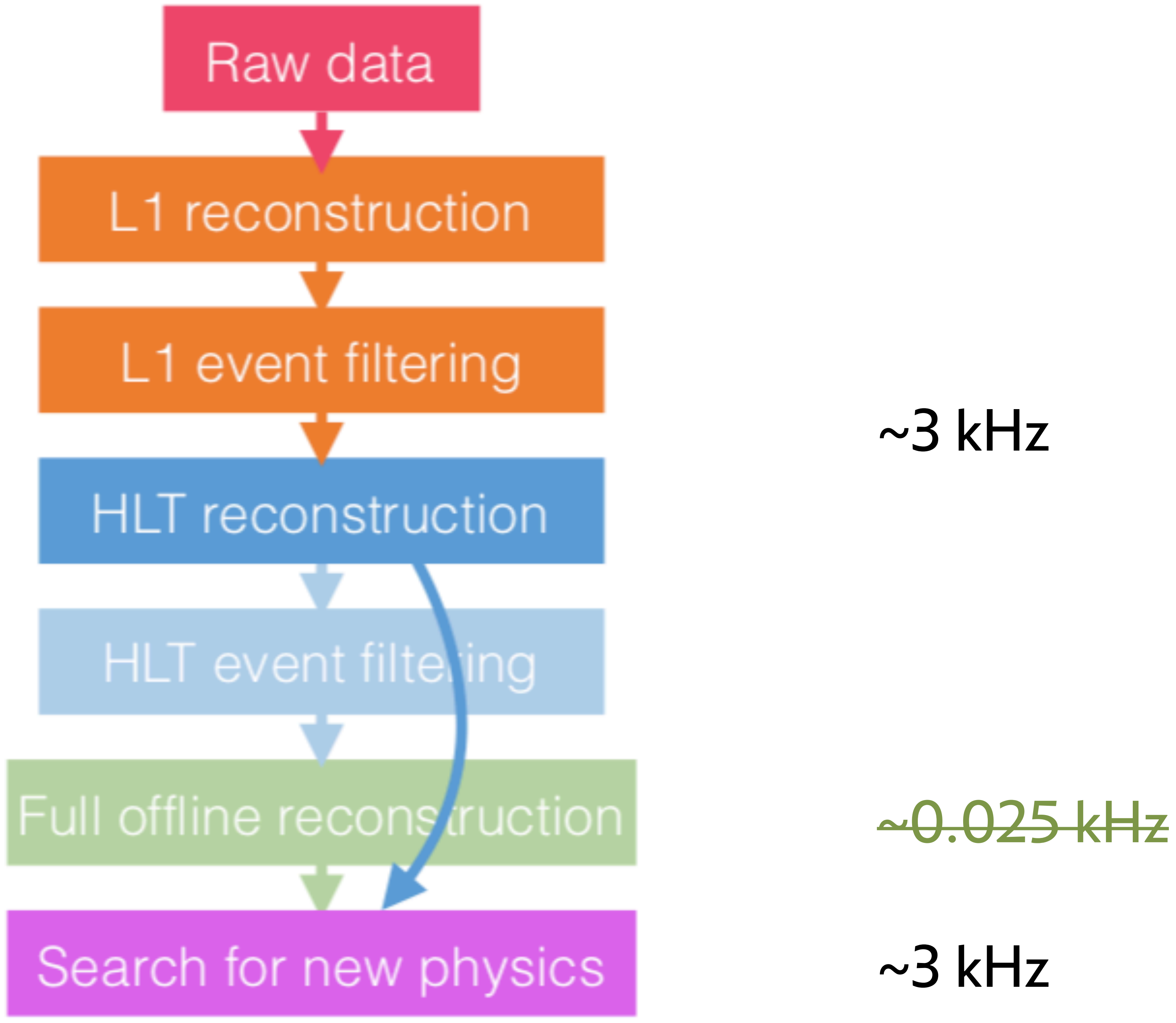
This can only be done within the trigger itself, i.e. trigger-level analysis (TLA).

Difference between L1 and HLT thresholds (200–440 GeV, shown earlier) suggests a first step for Run 2: improve (already good) and **analyze the HLT jet reconstruction at the L1A rate**; throw out the full data.

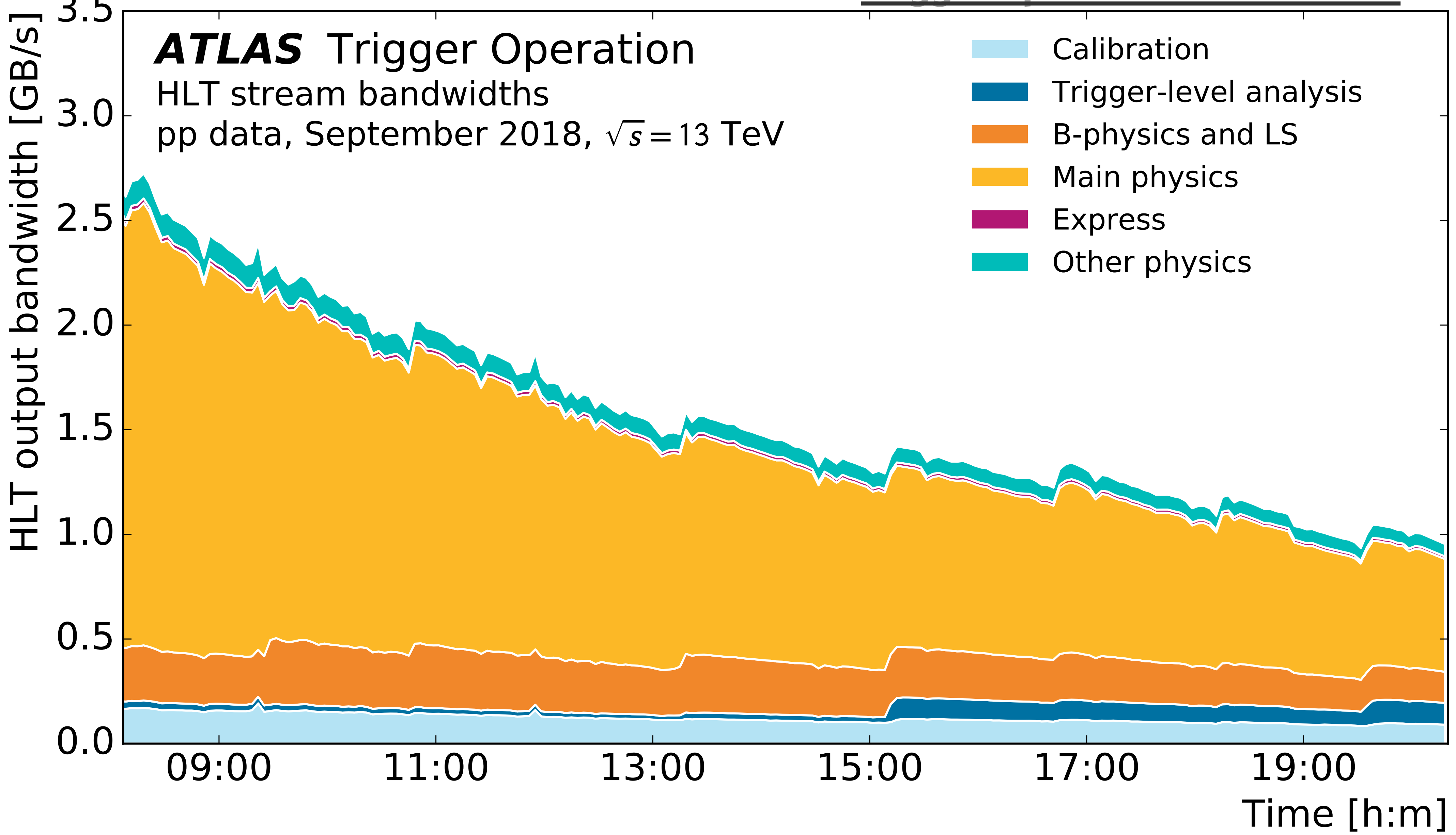
This technique also employed at LHCb (turbo stream) and CMS (data scouting).

TLA stream records **only HLT objects** (jet four-vectors, jet ID and calibration variables, etc.) for **specific L1A**.

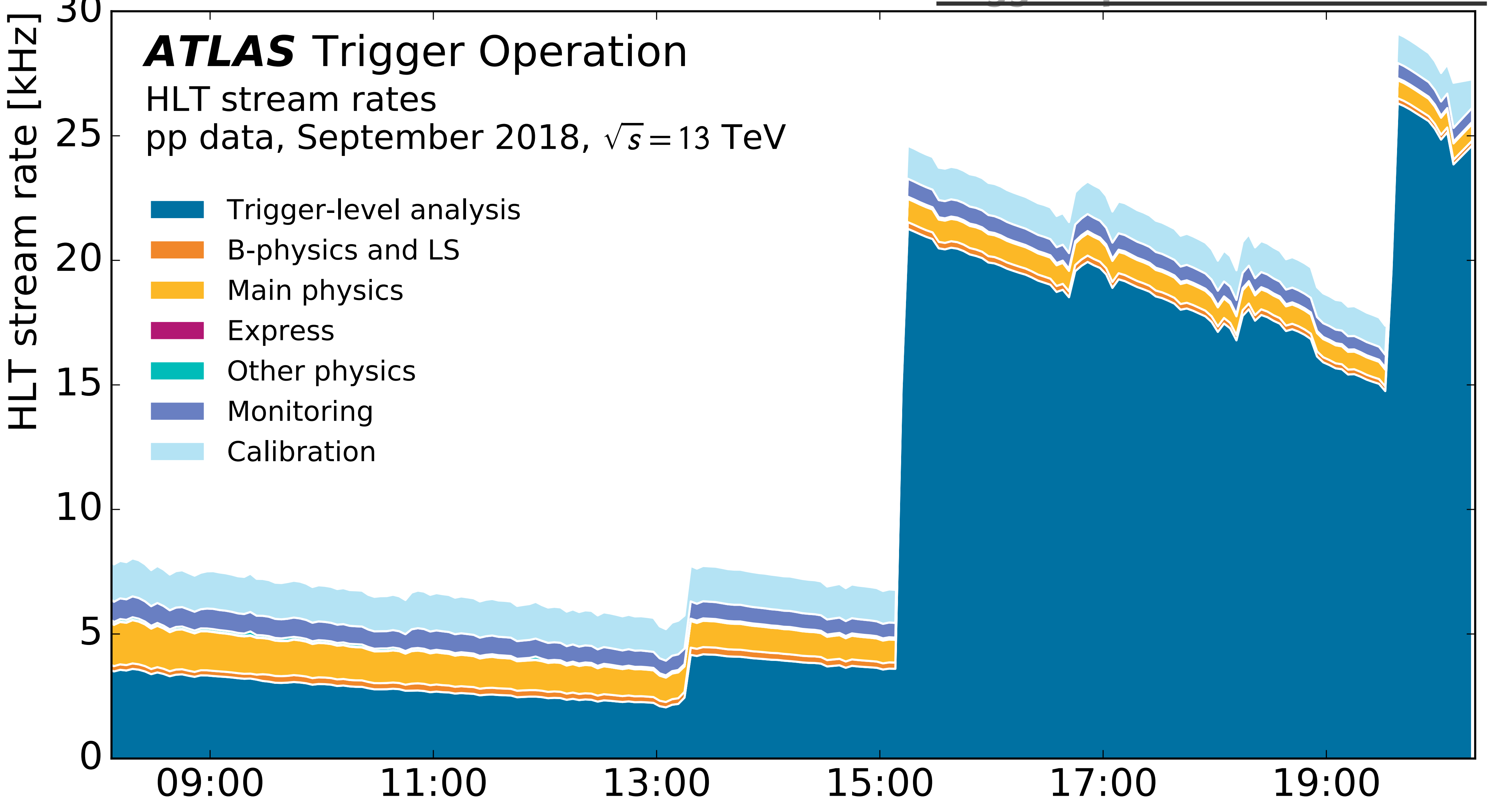
Throw out other information (e.g. **no tracking information kept** in Run 2, but **0.5% of full event** size.)



TriggerOperationPublicResults

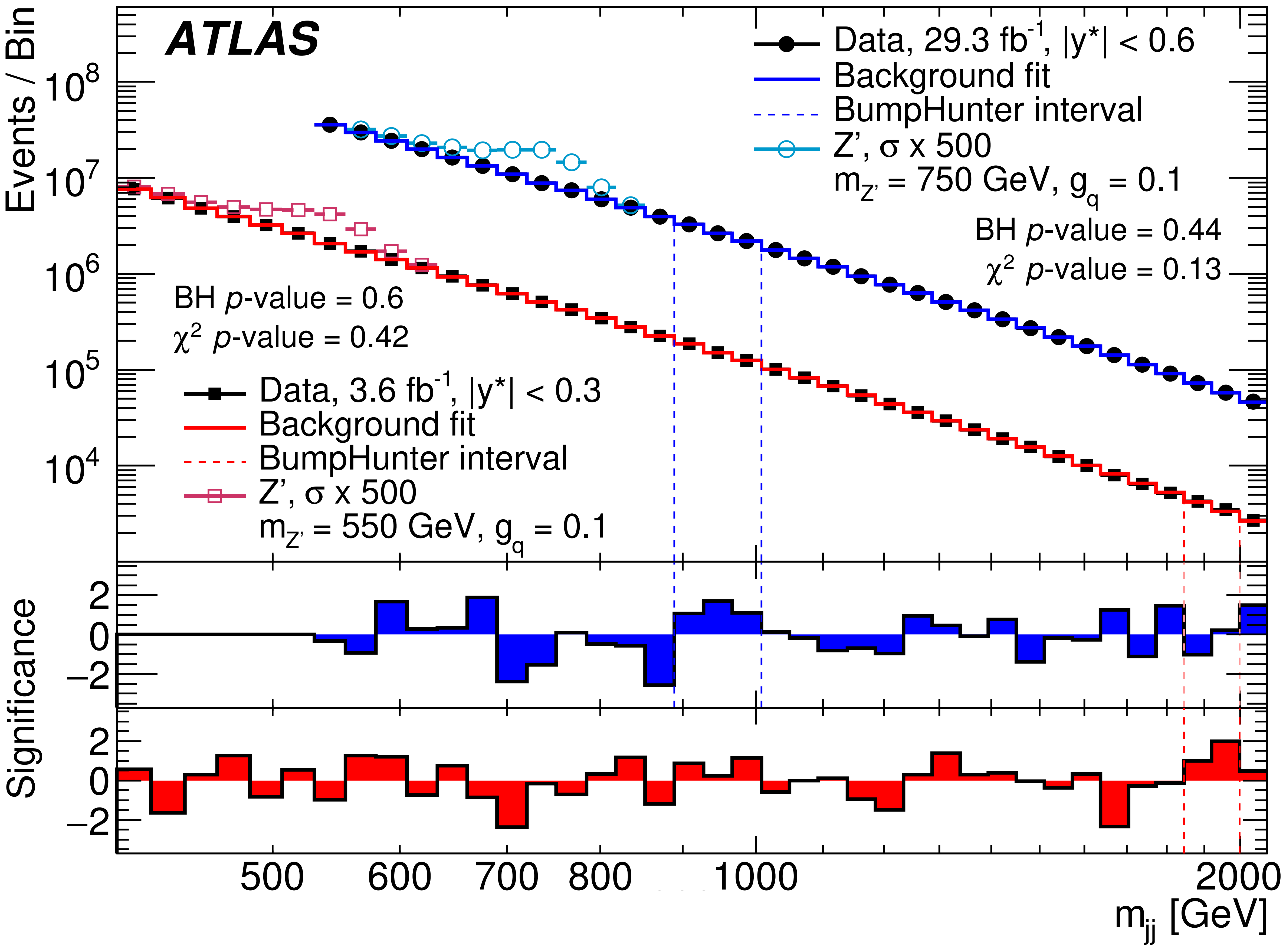


TriggerOperationPublicResults

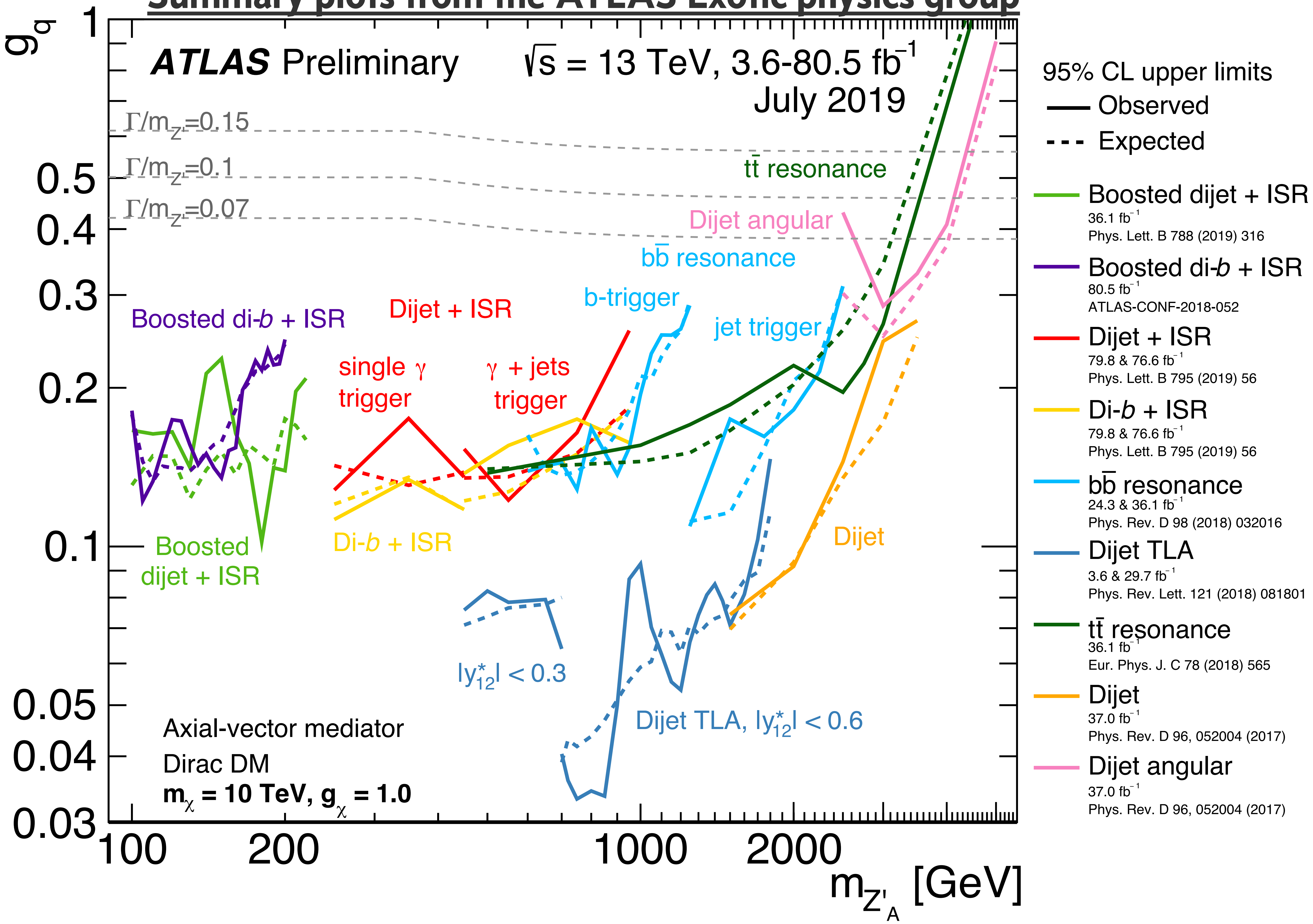


TLA results from first 1/4 of Run 2

PRL 121 081801



Summary plots from the ATLAS Exotic physics group



Analysis of **two mass ranges** with different L1 triggers (75 & 100 GeV) and different angular (y*) cuts.

Factor of **2–5x improvement in coupling** limits (roughly 1–2 orders of magnitude in cross section).

Does not yet use strategies for other trigger limitations.

Watch for improved results with **the full Run 2 dataset!**

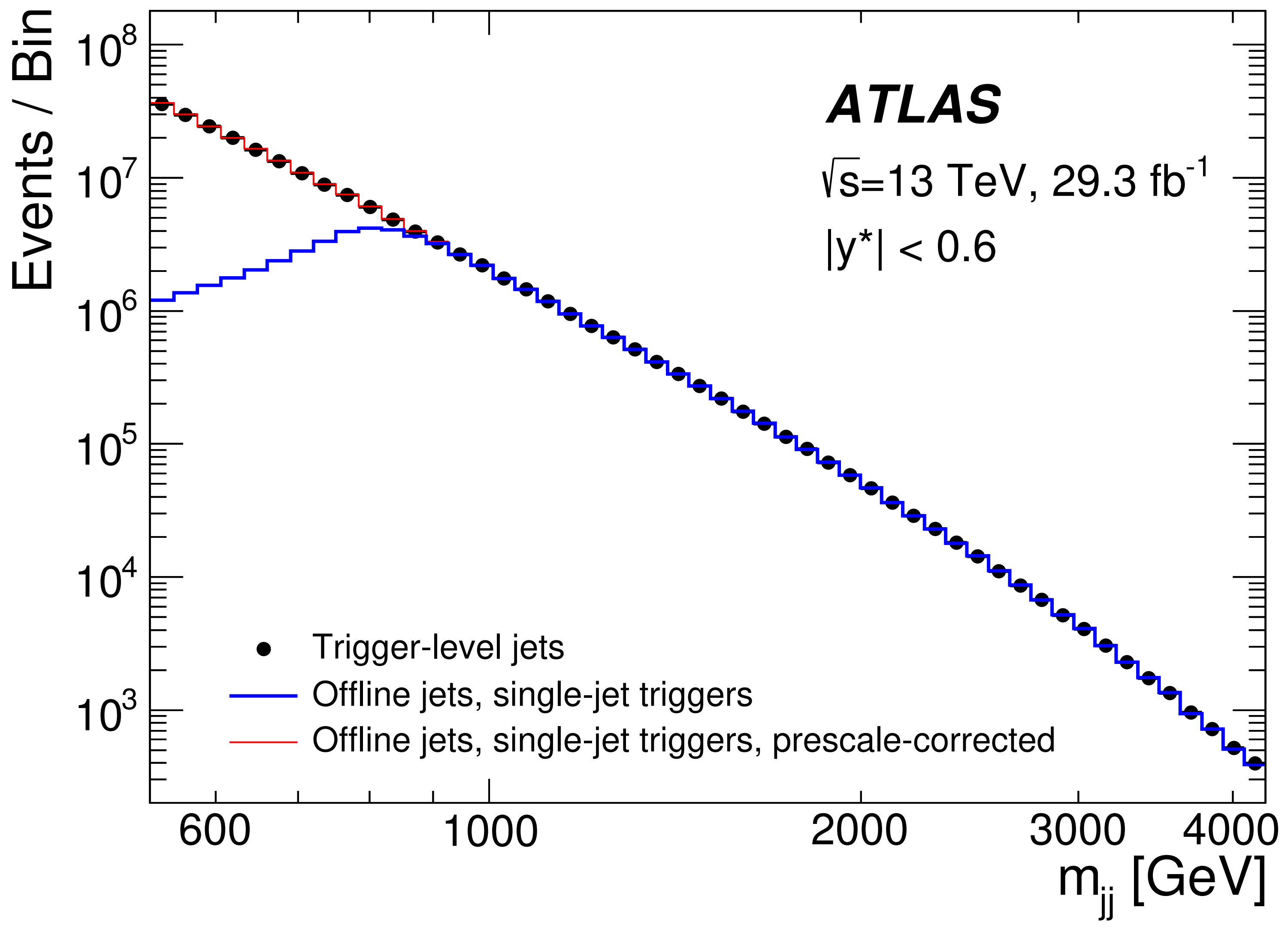
Challenges of TLA in Run 2

Huge background and small signal **requires very precise control** of all aspects of the analysis.

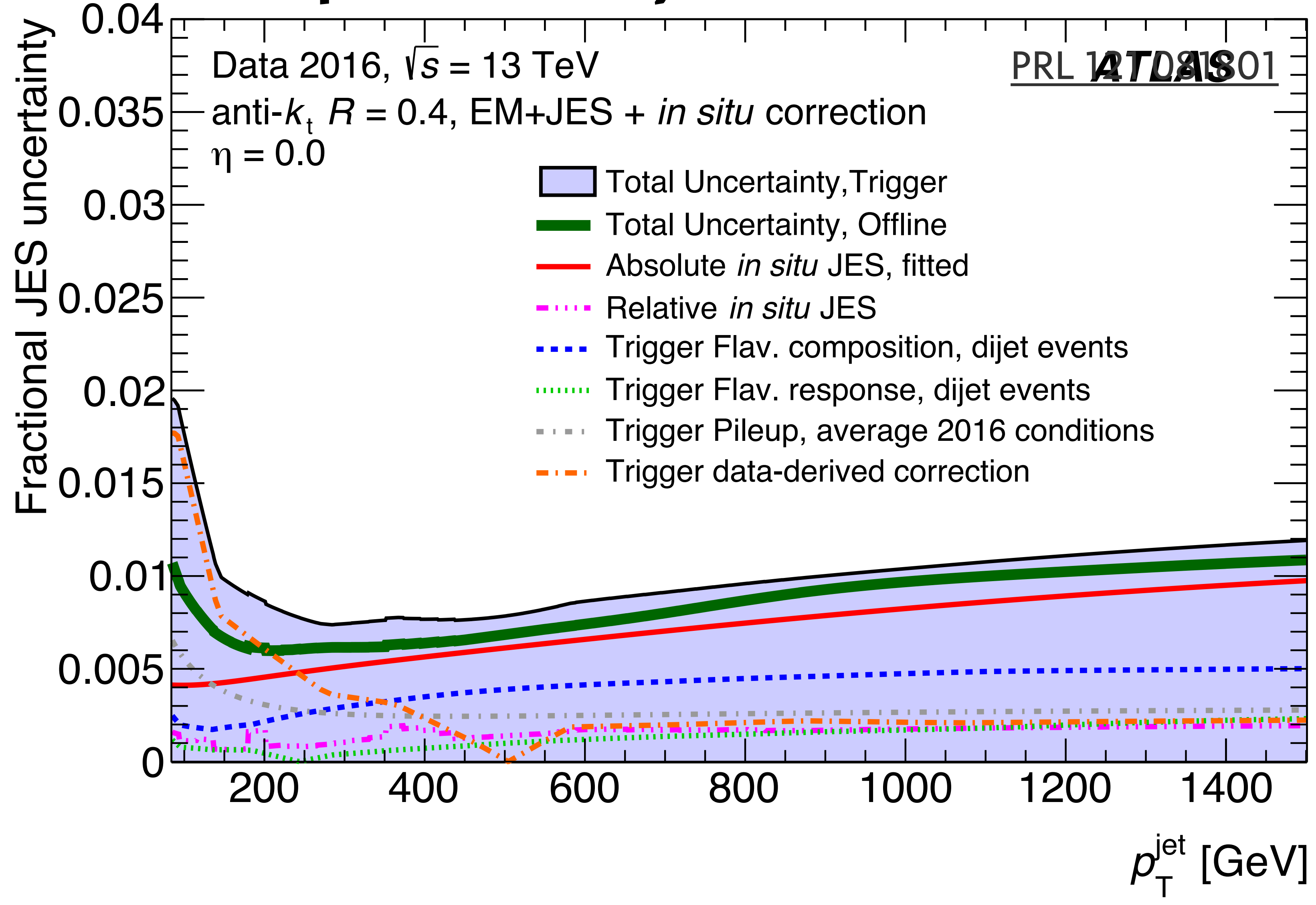
Partial-event data requires a **separate data handling pipeline** (non-standard reconstruction, data cleaning, HLT object calibrations, etc.)

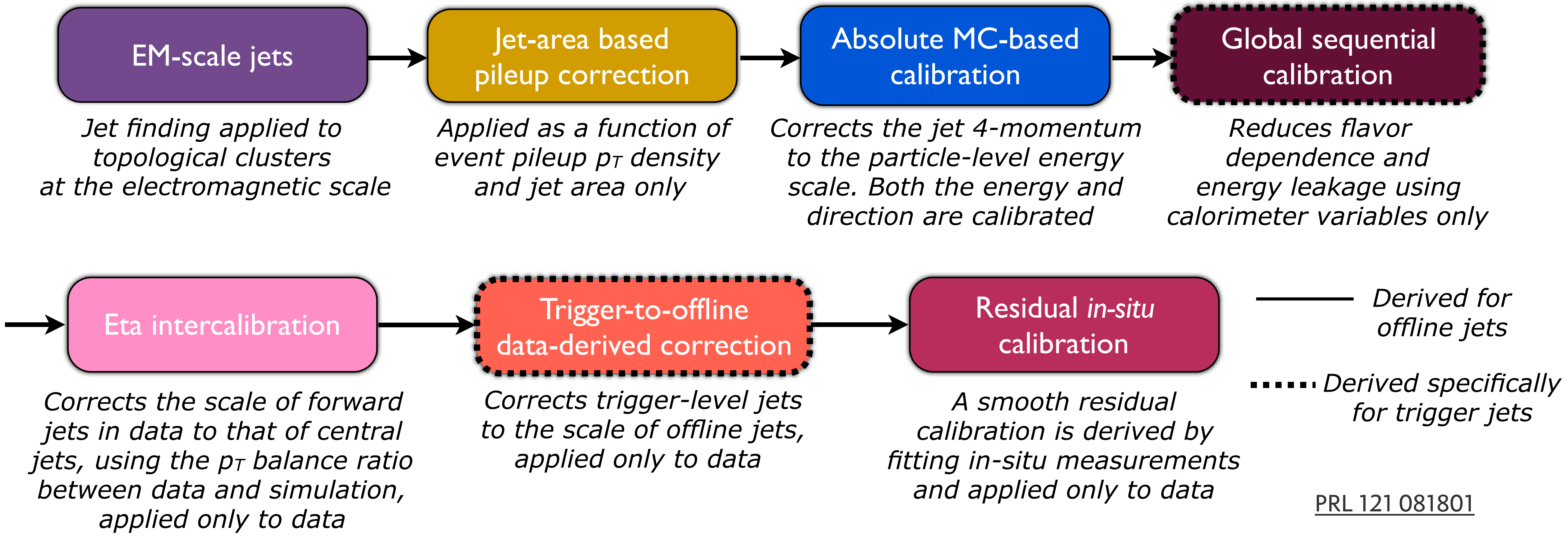
Without tracking, pile-up suppression is difficult for low-p_T jets.

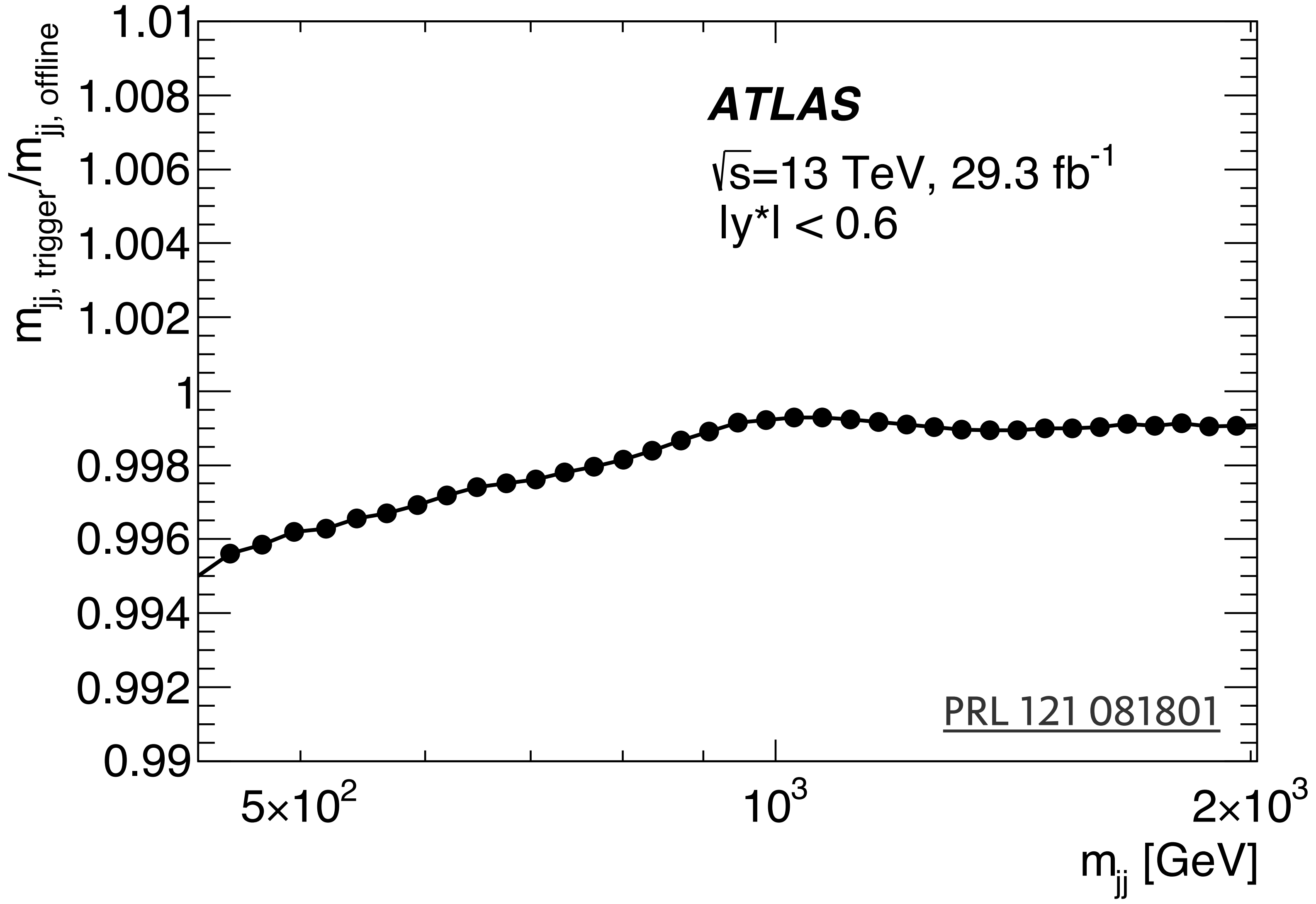
PRL 121 081801



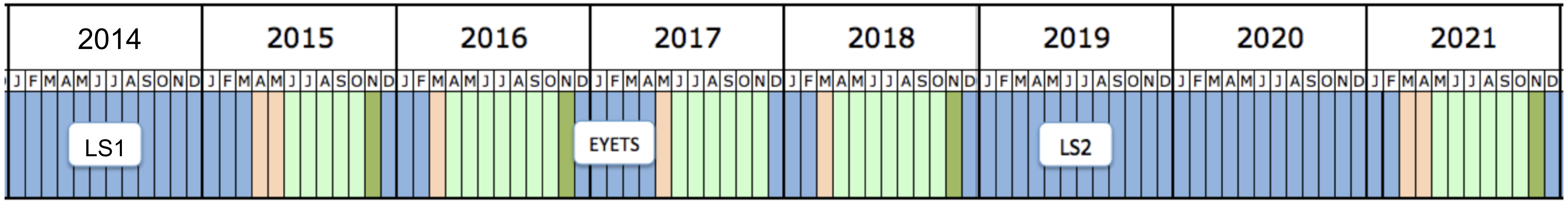
Example: Custom jet calibration







Trigger-level analysis in Run 2



TLA implemented in 2013–2014 see Eur. Phys. J. C (2017) 77:317

First TLA result with L1_J75 (3/fb)

Substantial work on HLT jet calibration

First TLA publication with L1_J75 and L1_J100 (30/fb) see PRL 121 081801

Final Run 2 jet calibrations and analyses underway + Run 3 preparations

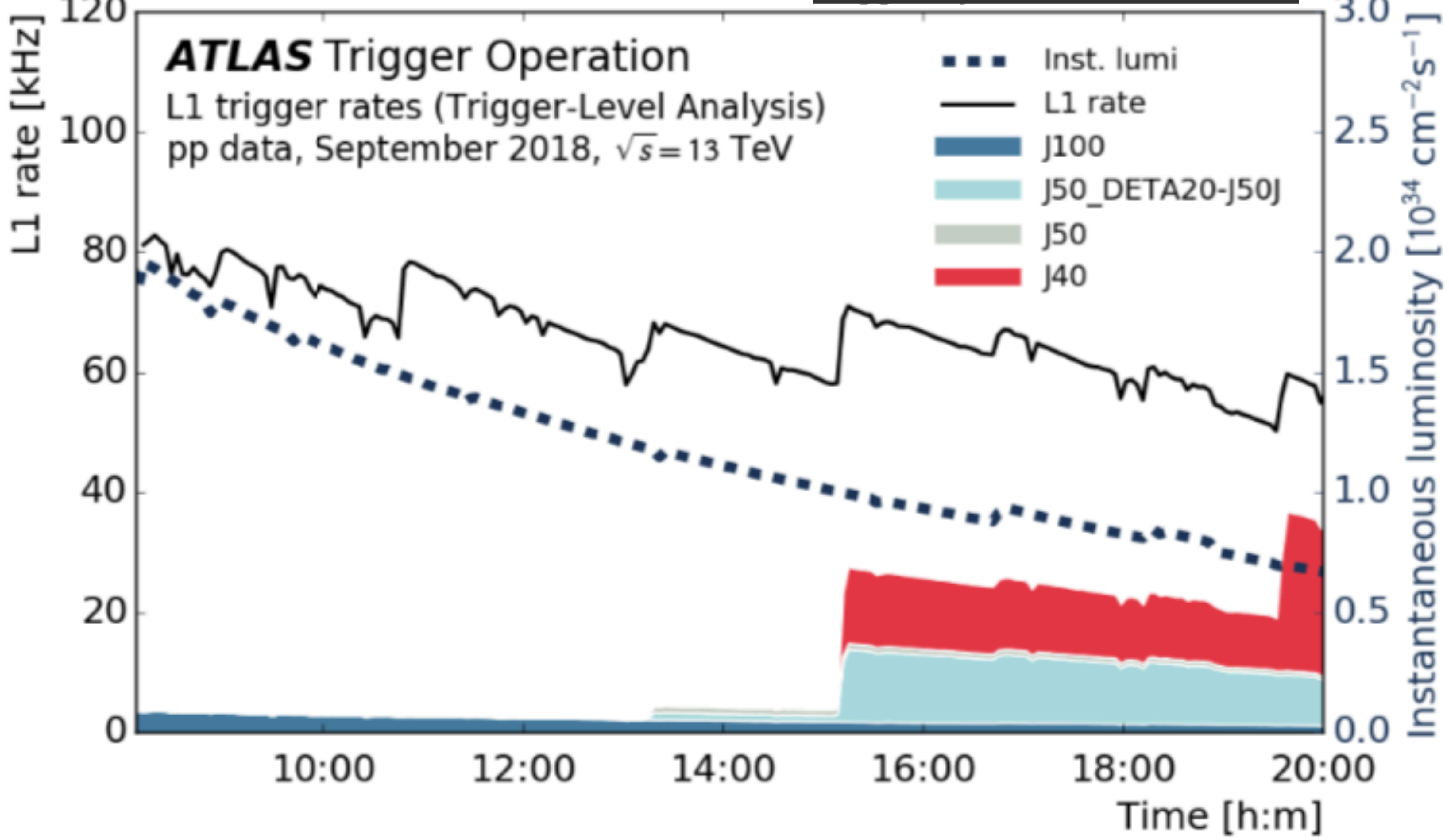
L1 rate remains most significant limitation

Two initial strategies explored for L1 limits: topological trigger and end-of-fill.

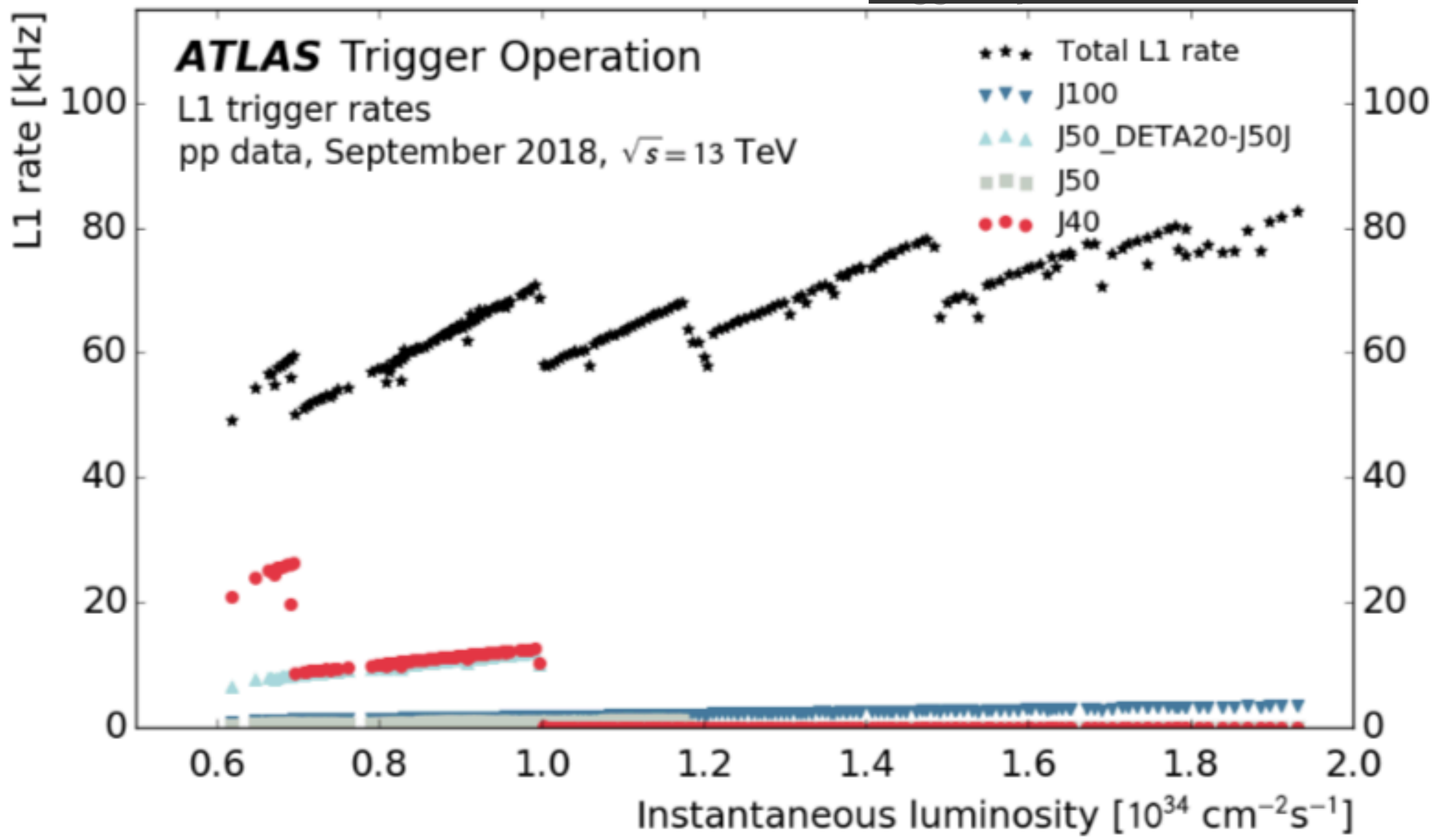
L1Topo processor allows angular cuts (pseudorapidity difference) to suppress main search background (t-channel dijet production).

Opportunistic **end-of-fill triggers** with lower L1 thresholds, in special situations using the majority of the L1 bandwidth.

TriggerOperationPublicResults



TriggerOperationPublicResults



Outlook for Run-3 and HL-LHC

Run 3 will bring several relevant improvements to the trigger hardware and software

Better HLT object calibration and resolution

- Possibility of particle-flow jets for Run-3 (better jet resolution at low p_T)

Pile-up mitigation for better control of lower-momentum (~<100 GeV) jets

- Possible software tracking for rejection of pile-up for lower-momentum jets; additional objects

Improved software flexibility for **partial-event readout**

Improved performance with **new L1 hardware**

HL-LHC

Powerful first-stage trigger capabilities with L0 Global Trigger upgrade and HLT Hardware Track Trigger

Storage and computing pressures increase

- but TLA also offers a solution, at least for some types of standard physics analyses

For details on these and further ideas, see also ATL-DAQ-PUB-2017-003 and related HSF-CWP-2017-01.