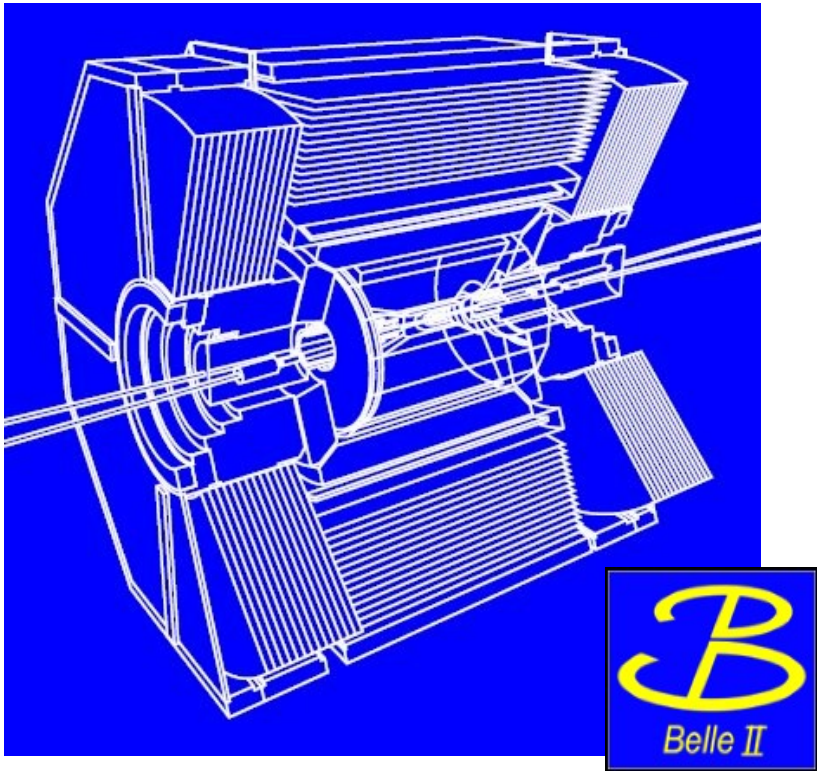
Belle 2 High Level Trigger

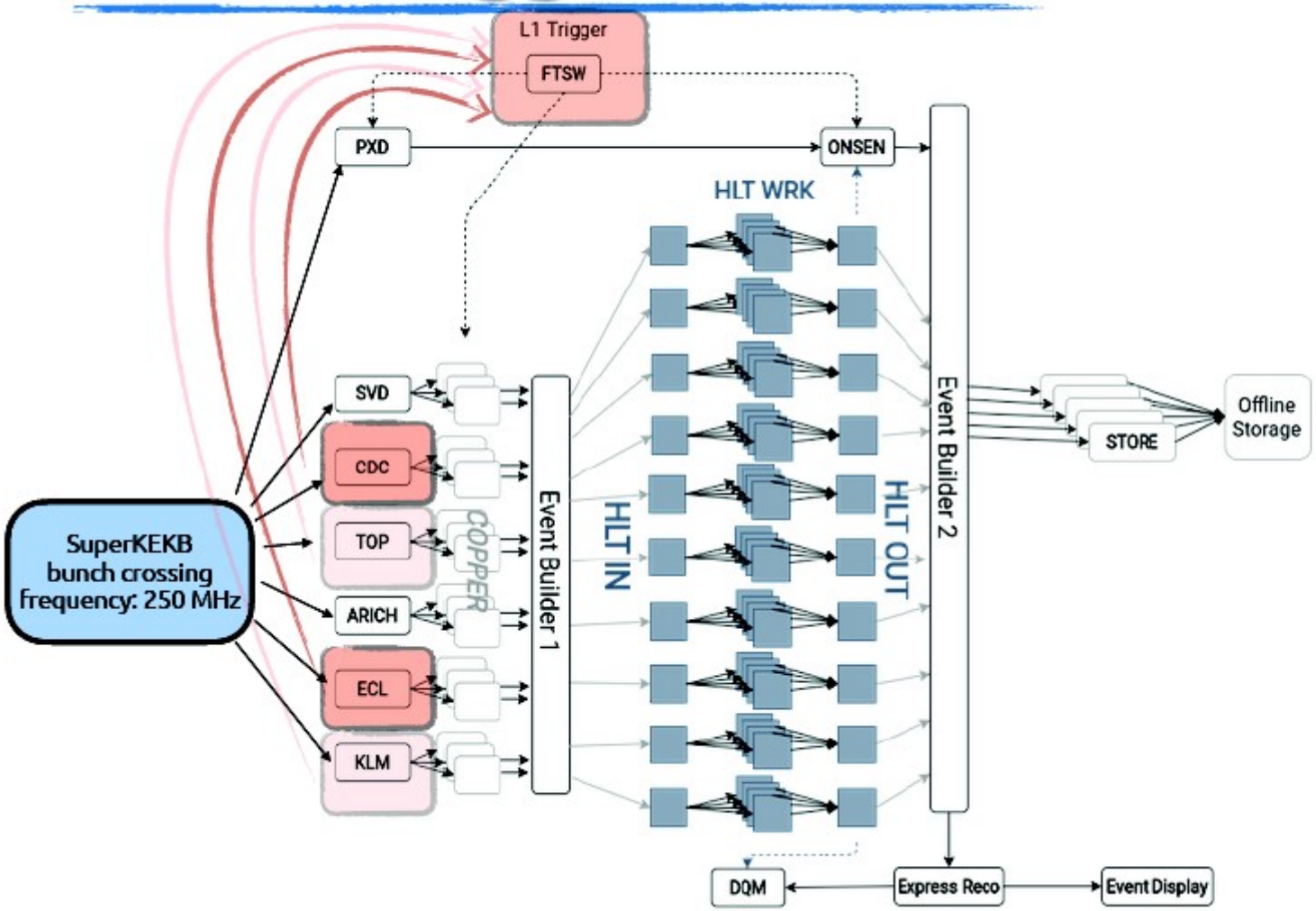


Karim Trabelsi

karim.trabelsi@in2p3.fr

2022/12/01

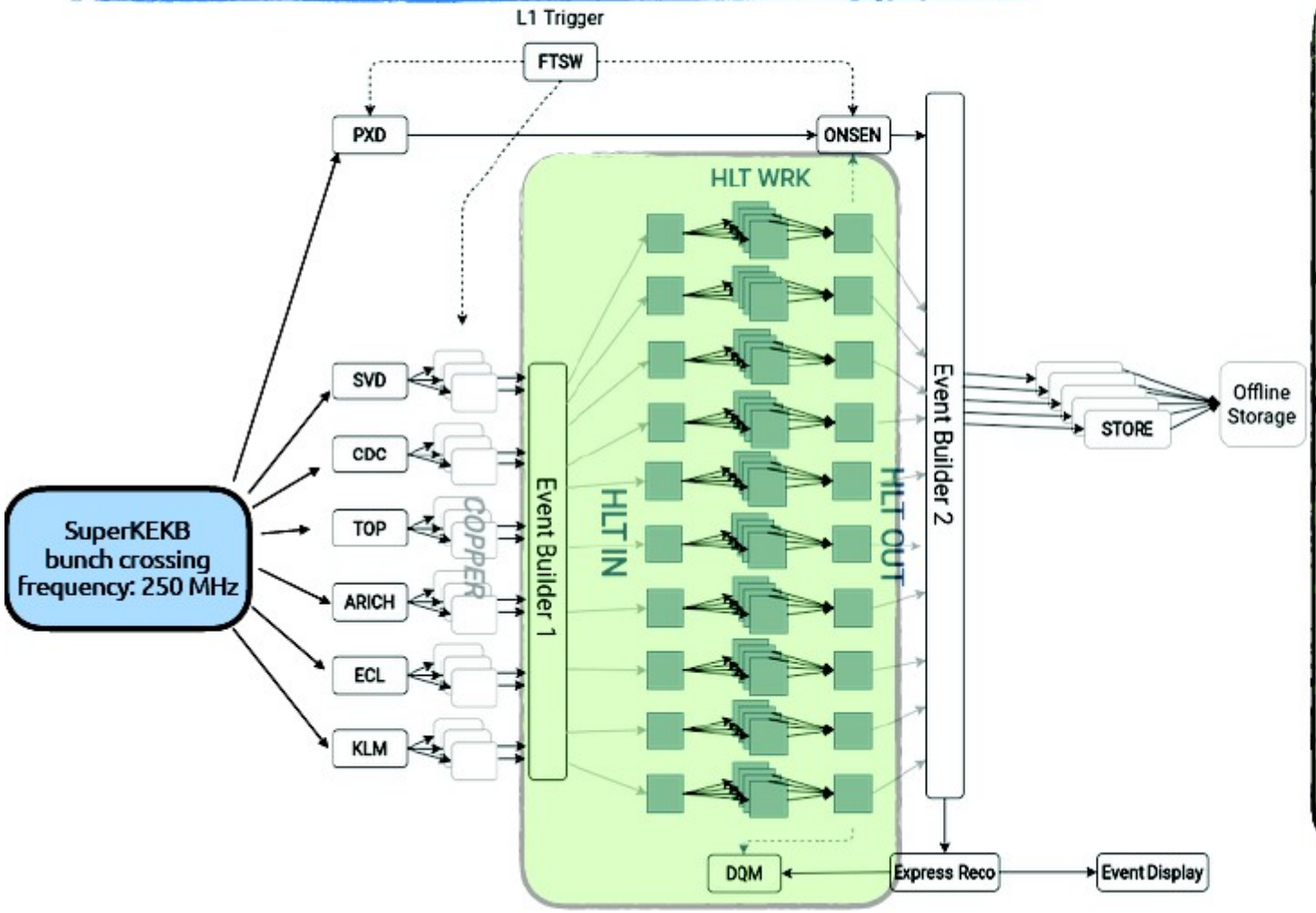
Belle II trigger dataflow: Level 1 trigger



L1 Trigger

- Purpose: **suppress the background** rate, retaining ~100% of $b\bar{b}$ events with high efficiency also for $c\bar{c}$ and $\tau^+\tau^-$
- Output rate:
  - Now: about **10 kHz**
  - Expected at target luminosity: 30 kHz
- latency: few μs
- Strategy:
  - processing on **FPGA,**
  - using OR of different, **orthogonal**, trigger lines (**CDC, ECL**)⇒ conservative approach

Belle II trigger dataflow: HLT



HLT

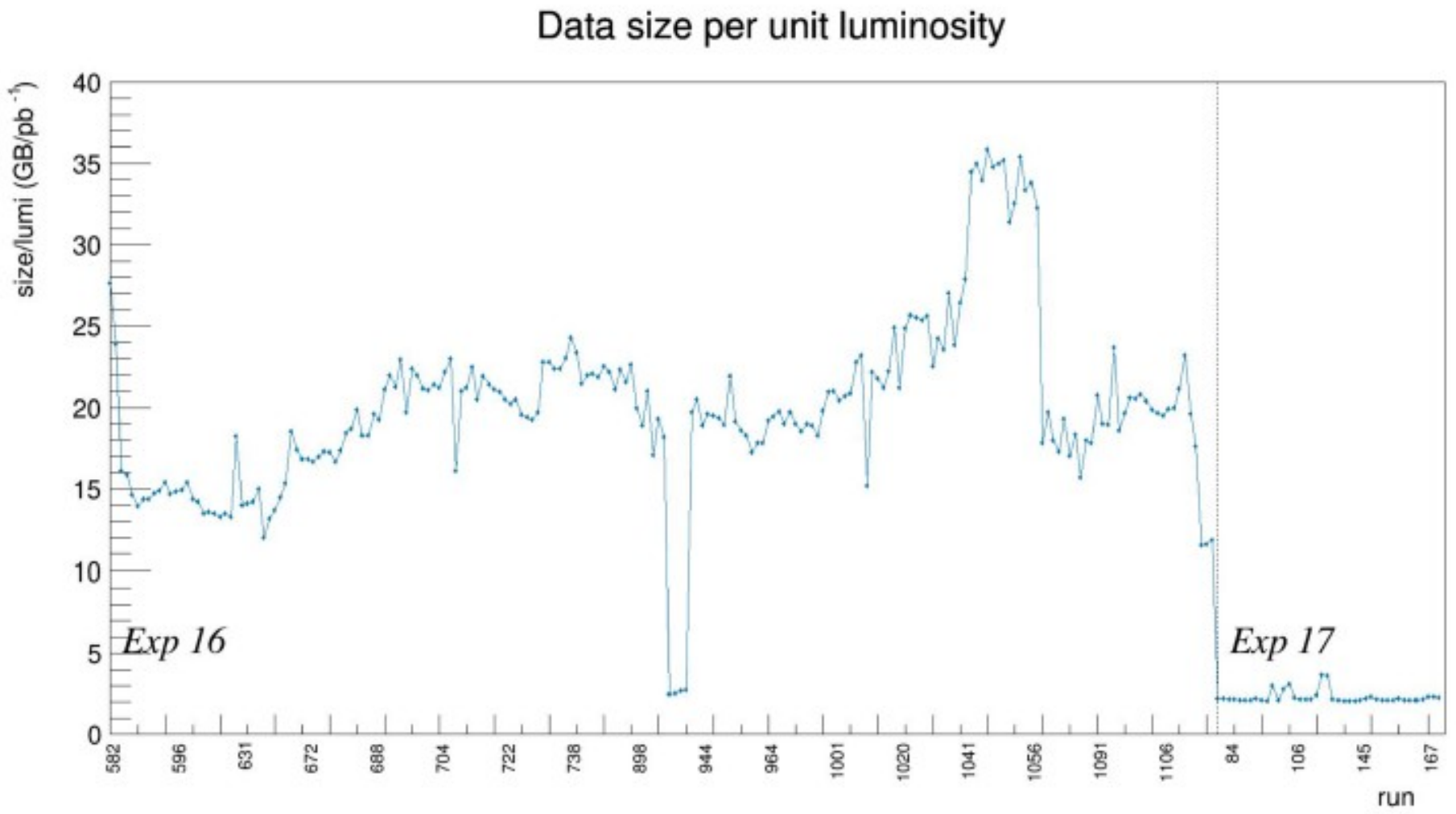
- Purpose:
  - reduce the trigger rate to a **storable rate**
  - run **DQM**
  - produce the **ROIs** for the PXD
  - assign the skim flag
- Output rate: ($\varepsilon \simeq 10 - 20\,\%$)
  - Now: about **2 kHz**
  - Expected at target luminosity: 6 kHz
- **Processing time: 300 ms**
- **Budget time (N_{proc}/L1 rate): 400 ms**
- Strategy: **fast reconstruction** on CPU
- hardware:
  - Now: 10 units, about 500 cores per unit–> **2 x 4800 processors**
  - After LS1: +3 units (to sustain 20 kHz input rate)

Belle 2 High Level Trigger is on

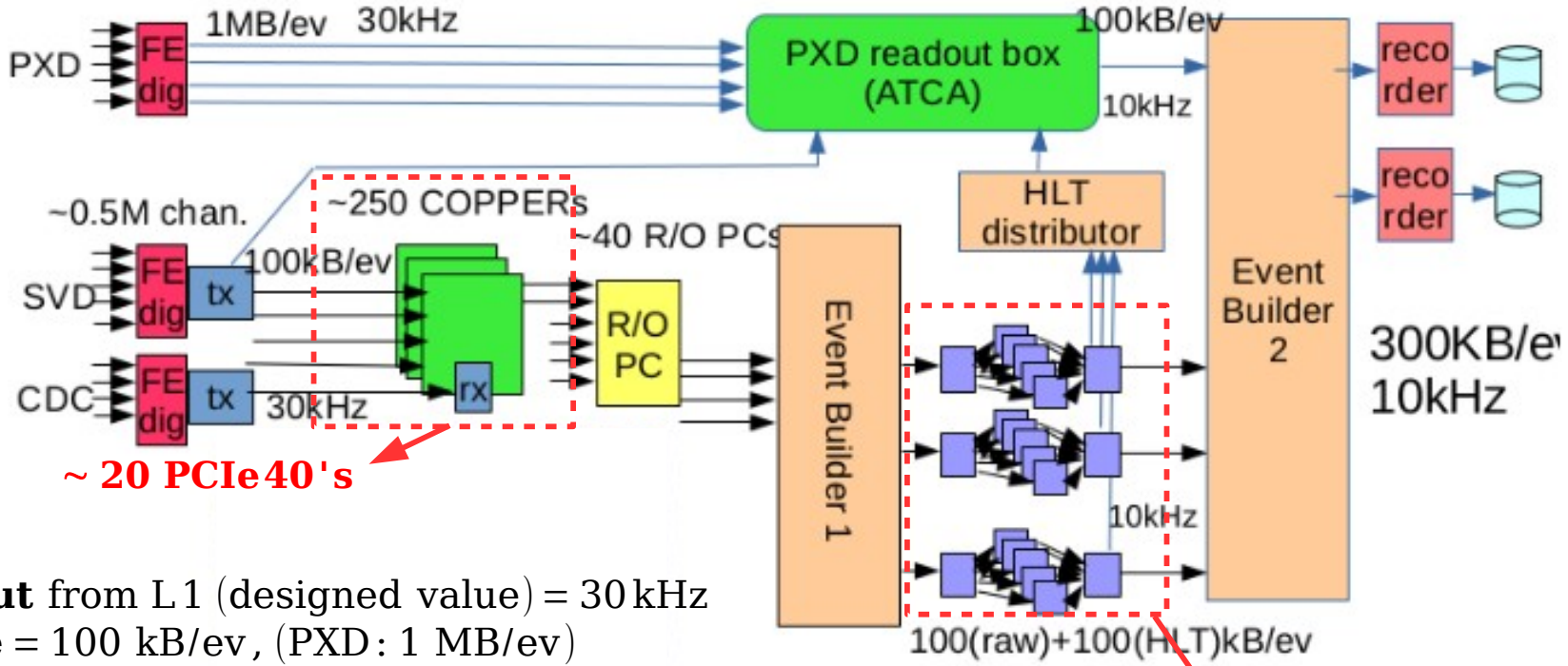
⇒ **main functions of HLT**

– **trigger rate: reduction by a factor 5(*)**

(*) **or more depending also on how loose is the L1 trigger**



Belle 2 High Level Trigger



- max. **input** from L1 (designed value) = 30 kHz
- event size = 100 kB/ev, (PXD: 1 MB/ev)
- max. **output** = 10 kHz
- event size = (100 + 100 (PXD)) kB/ev

We organized the HLT switch on in Spring 2021

⇒ main functions of HLT

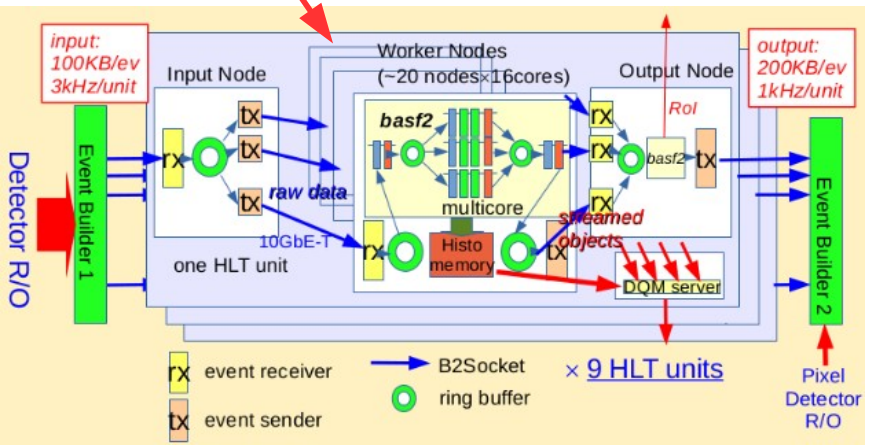
- trigger rate: reduction by a factor 5(*)

(*) or more depending also on how loose is the L1 trigger

- reconstruction without PXD → RoI feedback to Pixel Detector Readout
- tag events for calibration and physics skims
- monitoring (DQM on HLT/ExpressReco)

⇒ HLT activities

- performances + optimization
  → led by Vidya, KT, H. Grasland

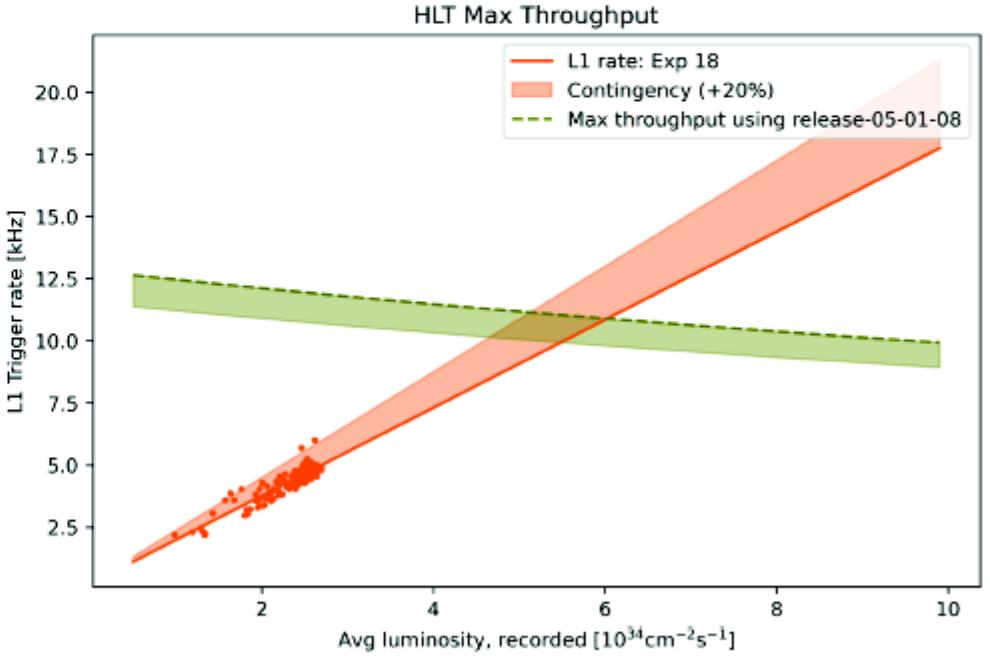


1 HLT unit, ~ 400 CPU cores/unit
each unit is completely independent
keep up with luminosity increase

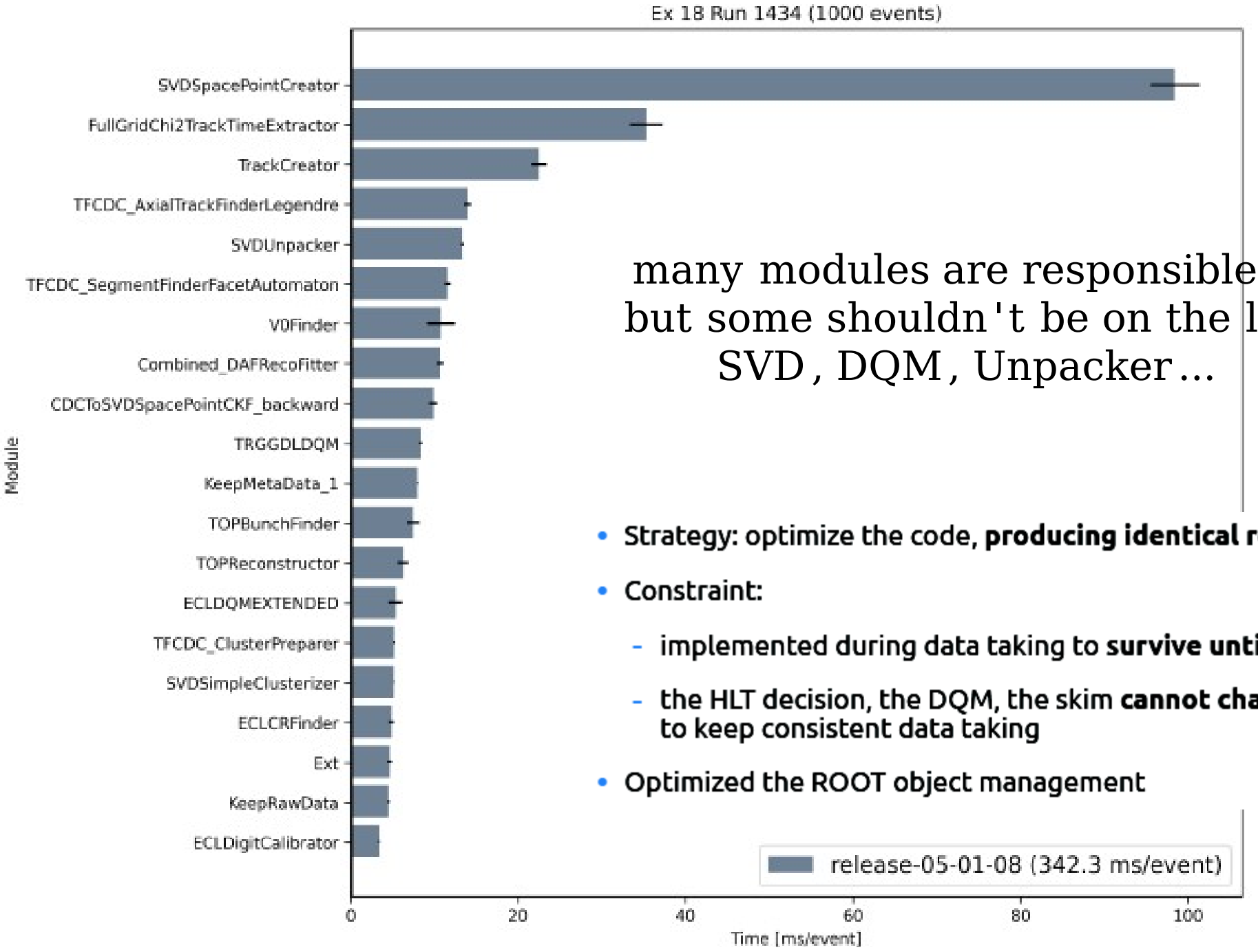
HLT limits (exp 18 ~ 2021 data taking)

Throughput = $N_{\text{processes}}$/process time

- L1 output (HLT input) **increase with luminosity** given the increased event rate
- Throughput decrease with luminosity given the **increasing complexity of the events** (higher background) which requires longer processing time
- In 2021 ($\mathscr{L} = 2 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$) Belle II realised that the conditions are not sustainable to reach the LS1
- Optimization of HLT is needed to increase the throughput (**decrease the processing time**)



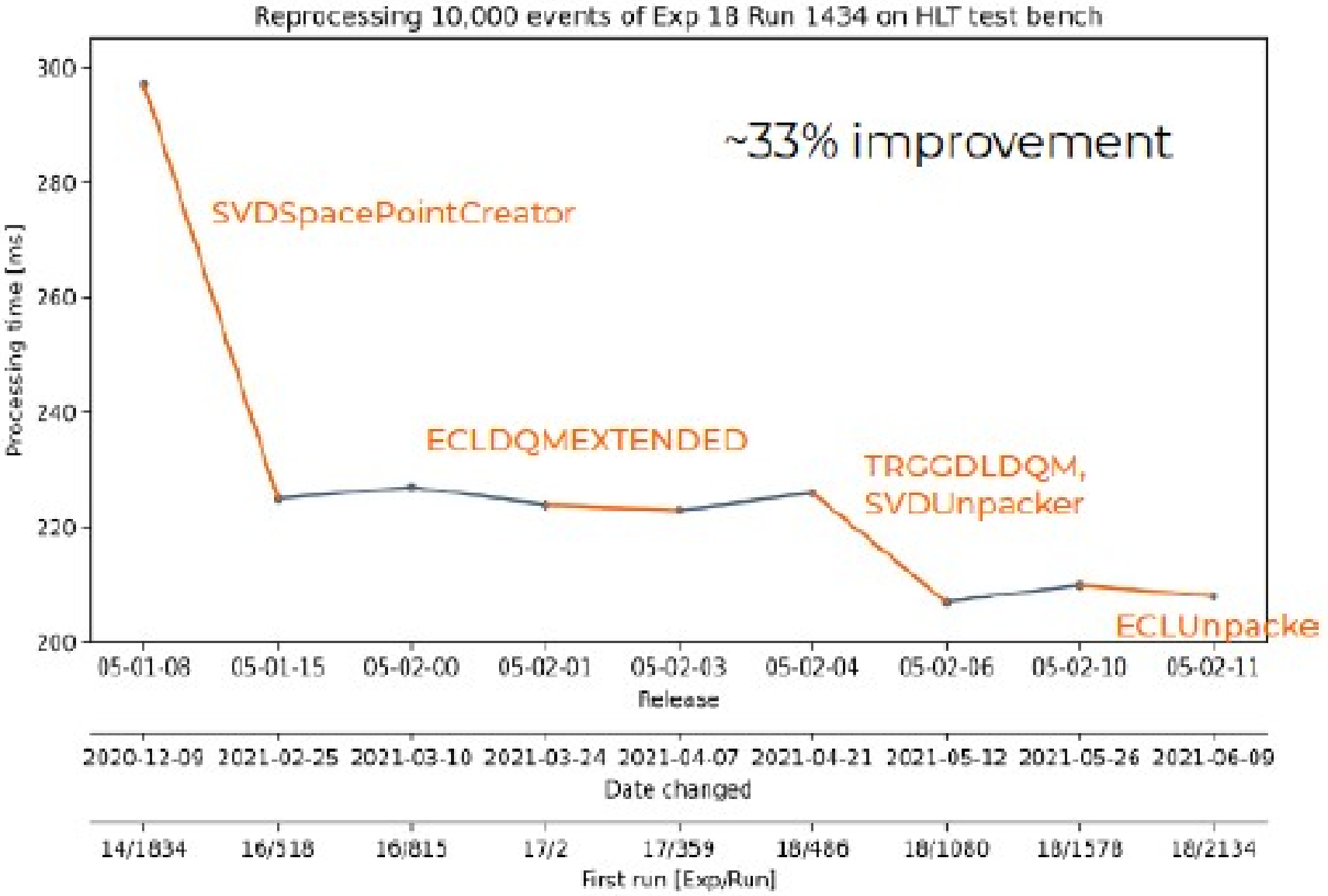
more than 300 ms/evt : from which modules ?



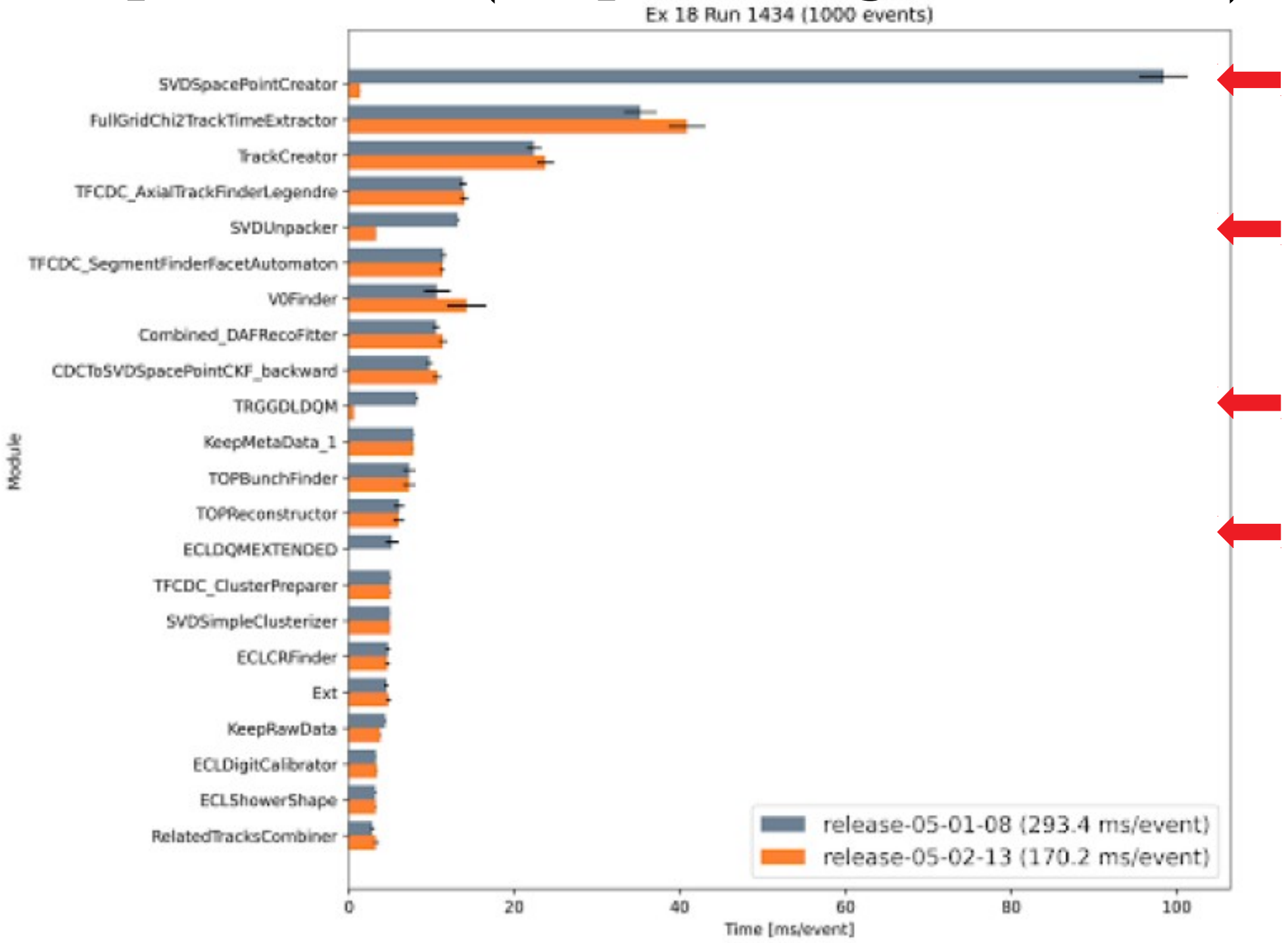
many modules are responsible…
but some shouldn't be on the list
SVD, DQM, Unpacker…

- Strategy: optimize the code, **producing identical results**
- Constraint:
  - implemented during data taking to **survive until LS1**
  - the HLT decision, the DQM, the skim **cannot change** to keep consistent data taking
- Optimized the ROOT object management

Optimization (Step 1, along release-05)

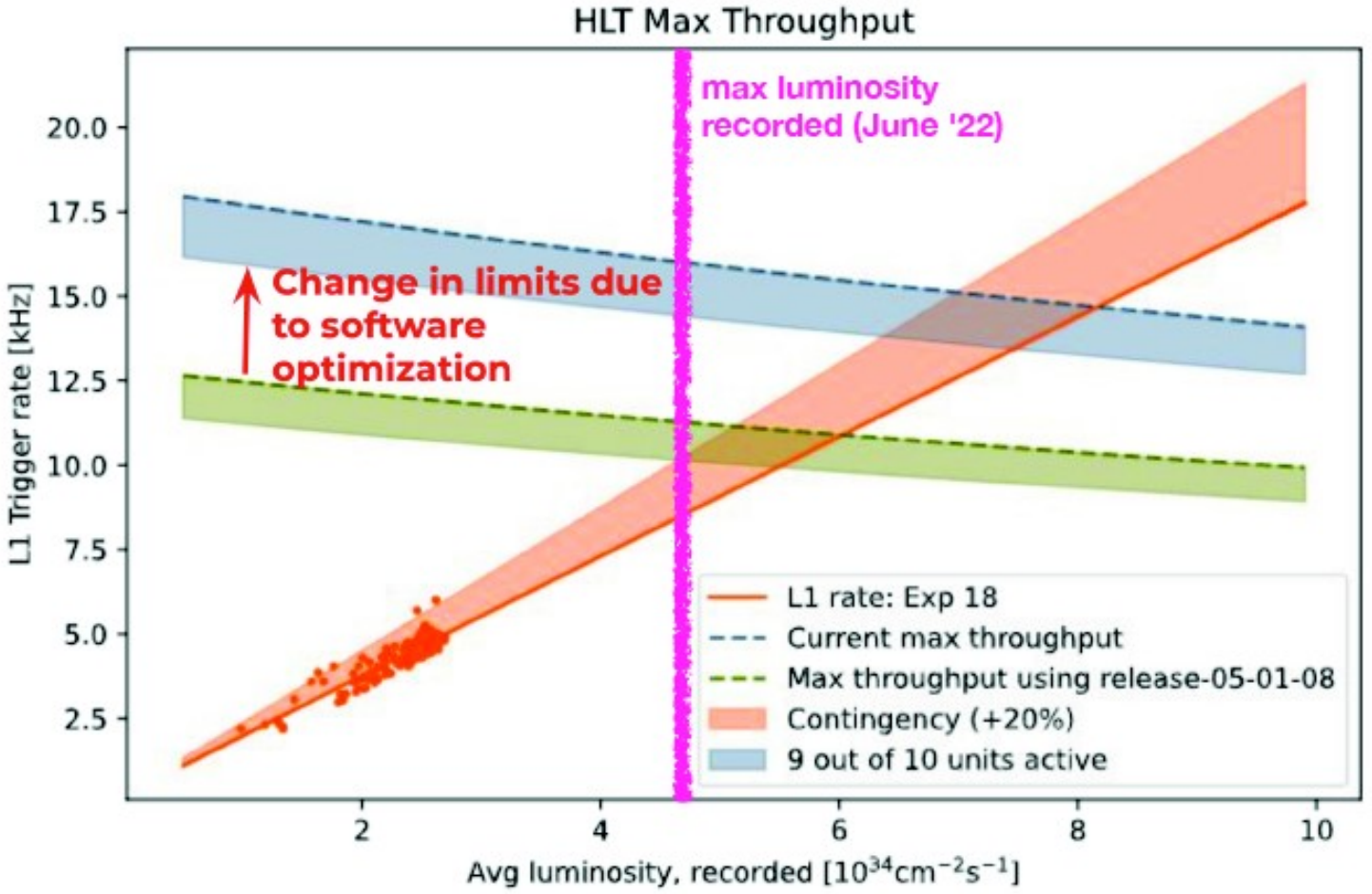


Optimization (Step 1, along release-05)

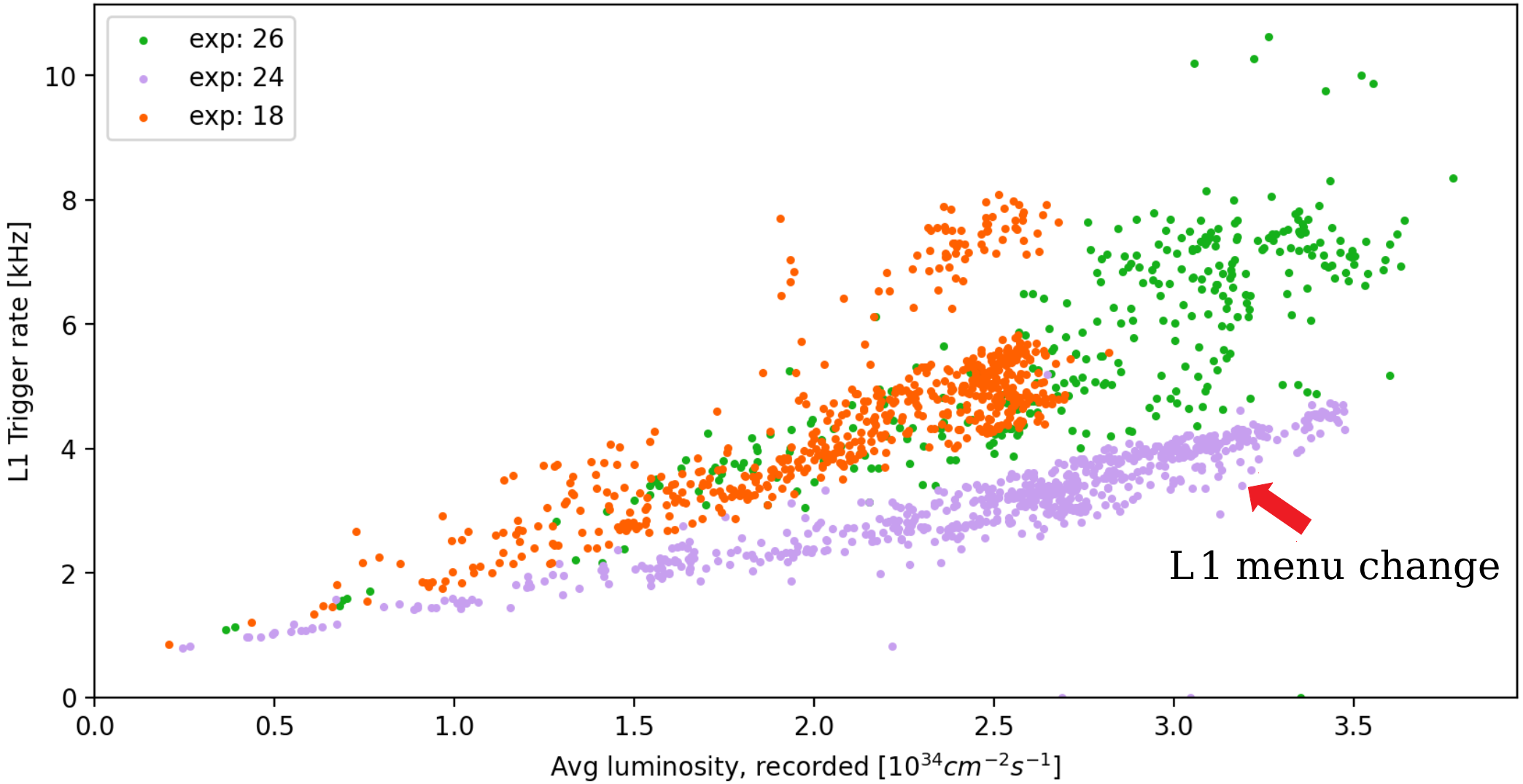


First optimization impact

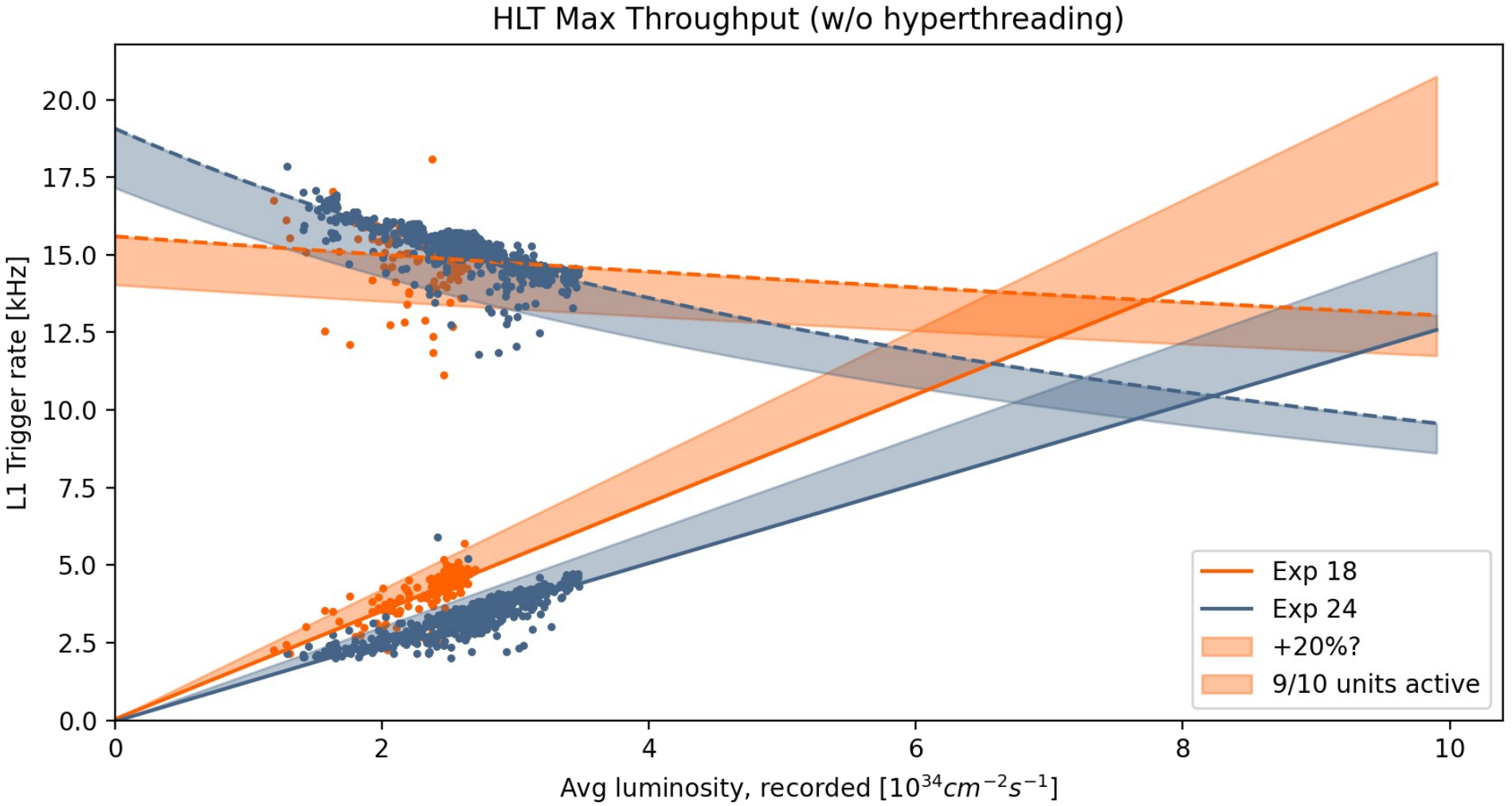
- Thanks to this optimization work we will survived until LS1!



L1 trigger versus Luminosity

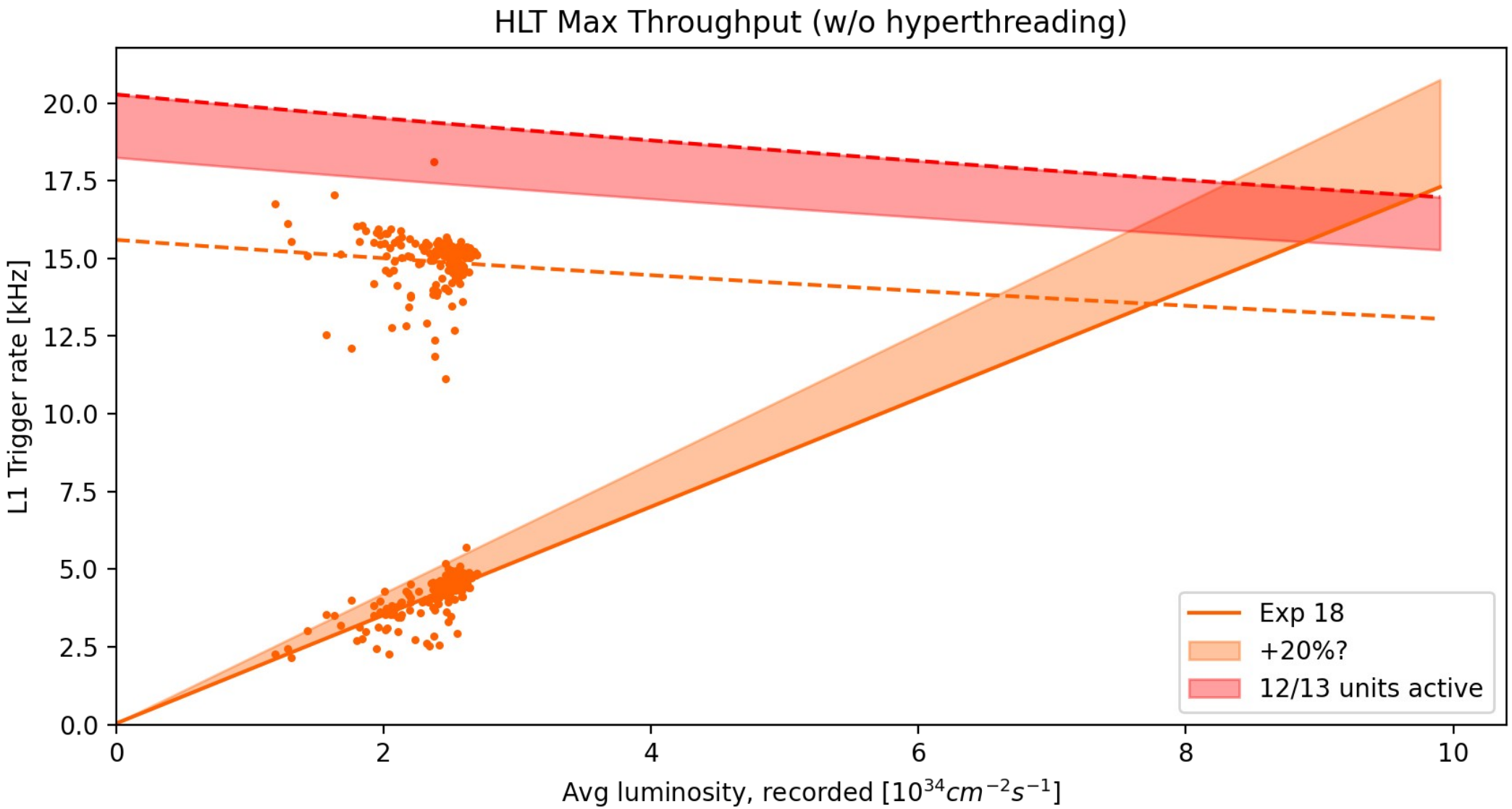


exp 18 versus exp 24



following conditions of exp 18 → can operate until 13 kHz
(corresponding to $6 \times 10^{34}/\mathrm{cm}^2/\mathrm{s}$)

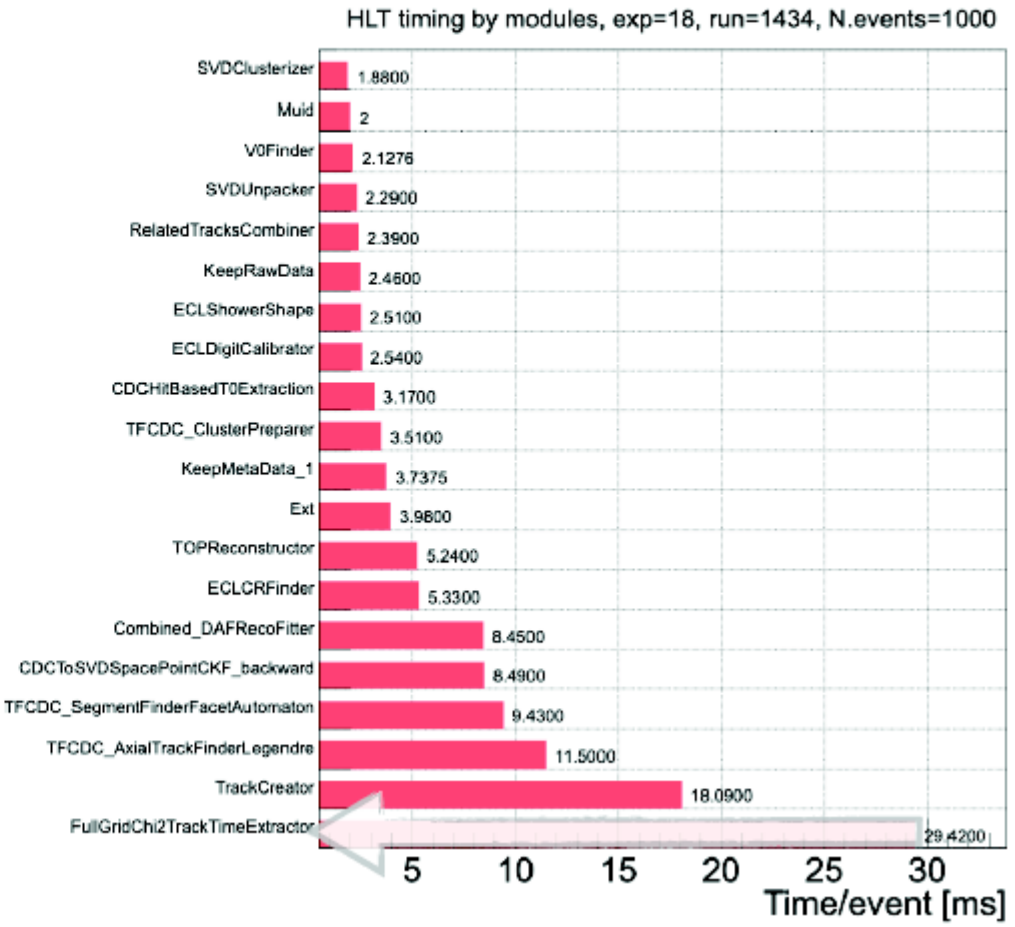
with 13 units



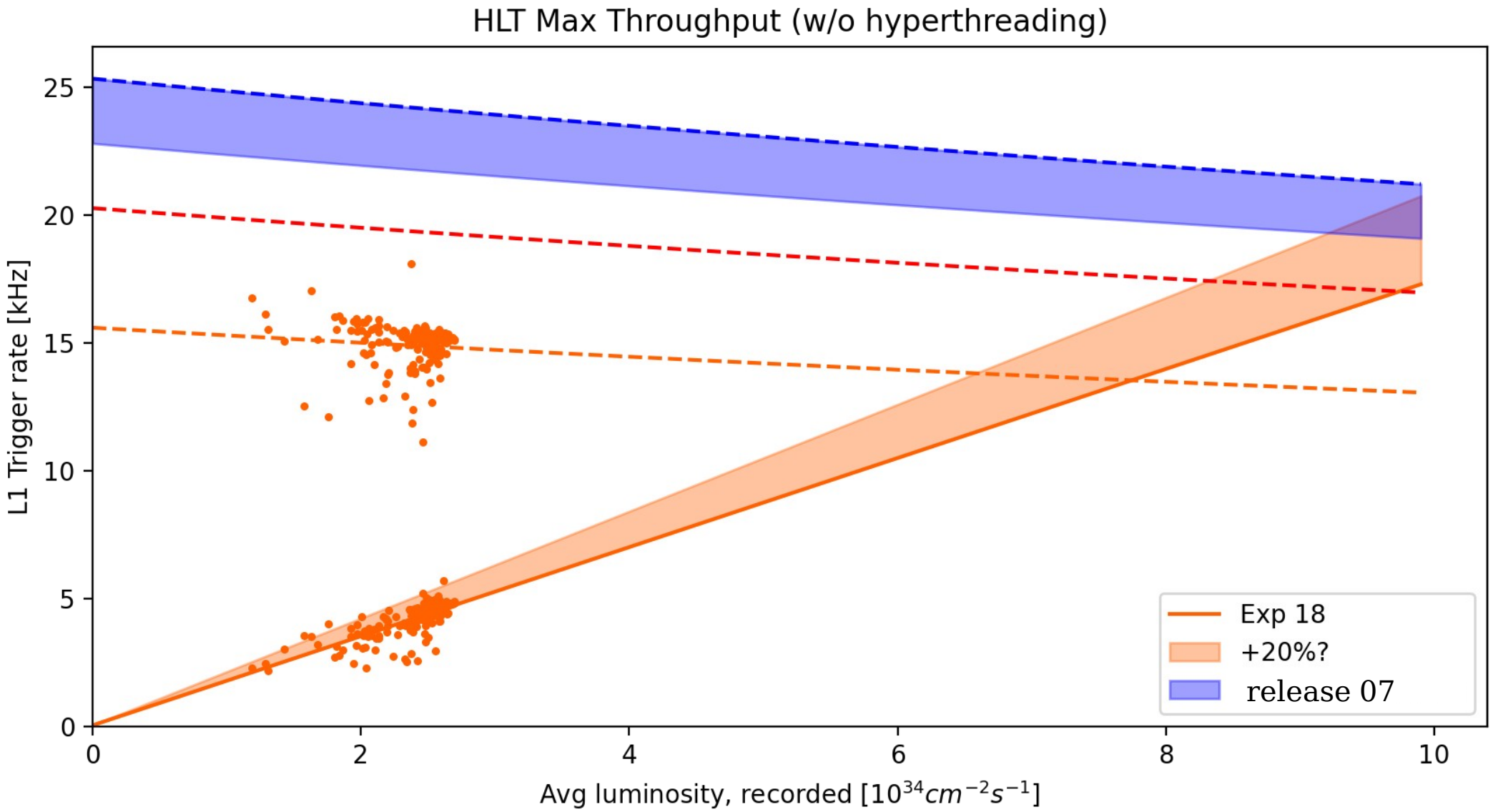
following conditions of exp 18 → can operate until 16 kHz
(corresponding to $8 \times 10^{34}/\mathrm{cm}^2/\mathrm{s}$)

Further optimization is needed

- Strategy: **modify the reconstruction** strategies, allowing also *small degradation*, to save processing time
- First achieved result: CDC Event Time estimation has been **replaced with SVD Event Time** estimation ⇒ 2000 times faster *[see backup]*
- Next step: reducing tracking processing time **(track fitting)**



with 13 units + release 07



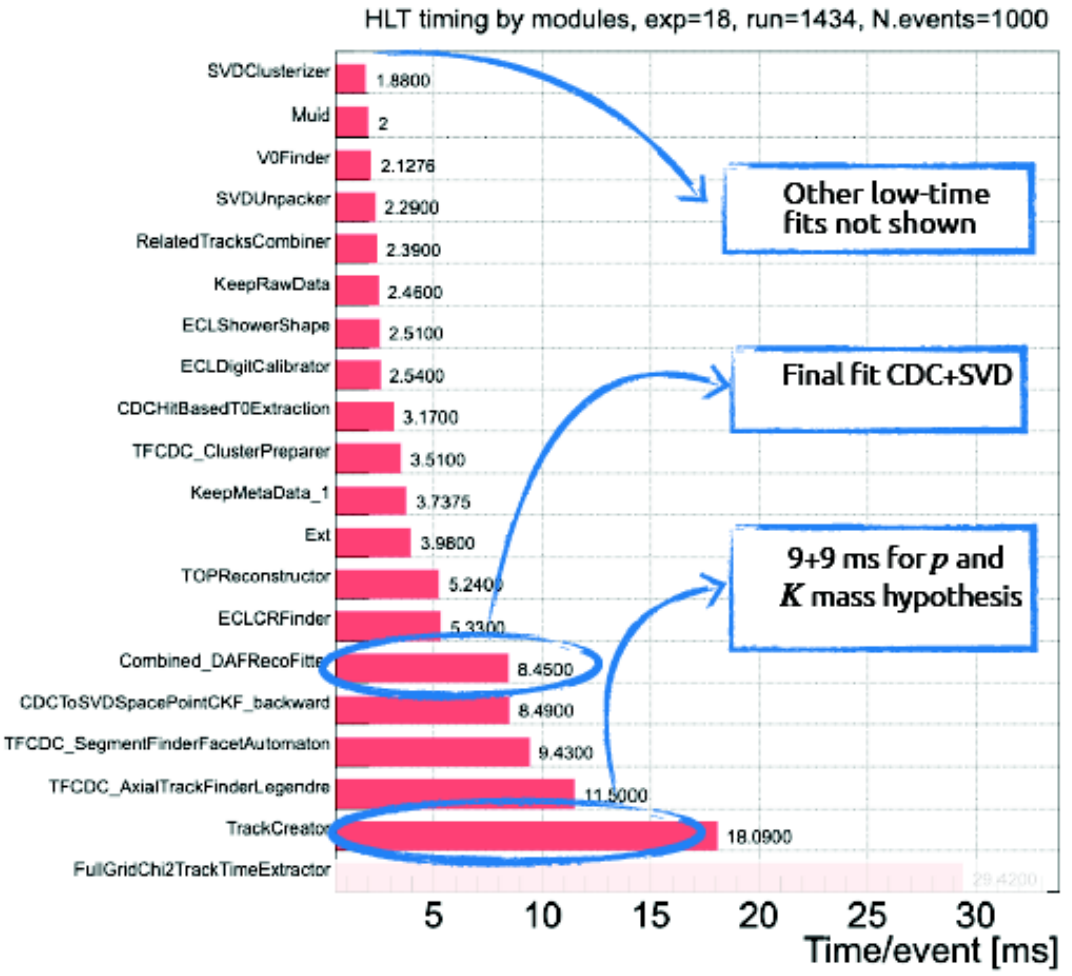
following conditions of exp 18 → can operate until 20 kHz
(corresponding to $9 \times 10^{34}/\mathrm{cm}^2/\mathrm{s}$)

V. Bertacchi (CPPM/IJCLab)

Track Fitter calls

- The fitter is **called ~5 times per track**, using a Deterministic Annealing Filter (**DAF**)
- With the current configuration the DAF takes **15 ms/track** for each call

The DAF its optimization has a **radical impact** on reconstruction CPU time (and tracking performance)



DAF

- For each call of the fitter the DAF (*Deterministic Annealing Filter*) is called
- The purpose of the DAF is to remove from the fit the **outlier hits** to improve the fit accuracy
- Method:
  - The DAF is **assigning weights** (in the range [0,1]) to each hit, accordingly to the residuals between the measurement and the Kalman Filter prediction.
  - The fit is r**epeated multiples times lowering an annealing temperature**
  - A **convergence criterion** is defined, based on the variation of the weights and the p-value of the fit (see next slide)
- Status: the **DAF has been never optimized**, and in the current configuration the convergence is not tuned ⇒ **extremely time consuming**

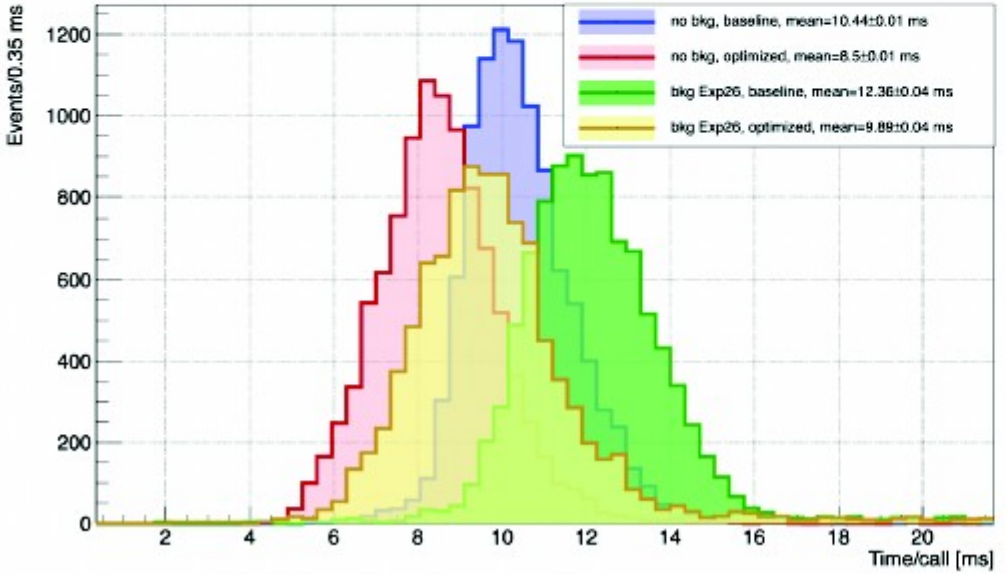
DAF demonstrative optimization

Test condition:
- CDC only tracking
- Muon Gun
- pT=1 GeV
- $\theta = 70°$

Changed some hyperparameters of the DAF *[see backup]*:

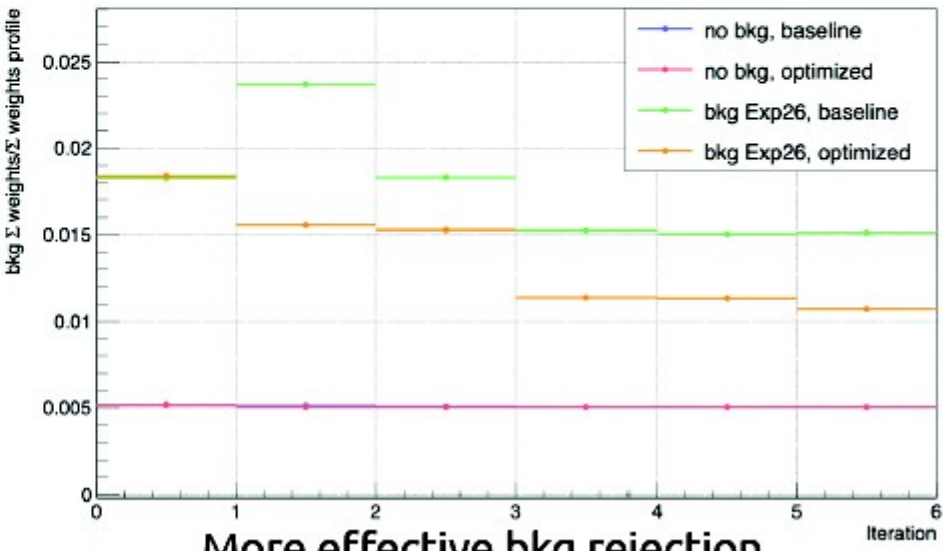
- to obtain reasonable **convergence behaviour** (use the iteration range, use mainly primary convergence criterion, exploit more wisely the p-value)
- having the CPU **time figure of merit**

DAFRecoFitter module CPU time



Better timing performance

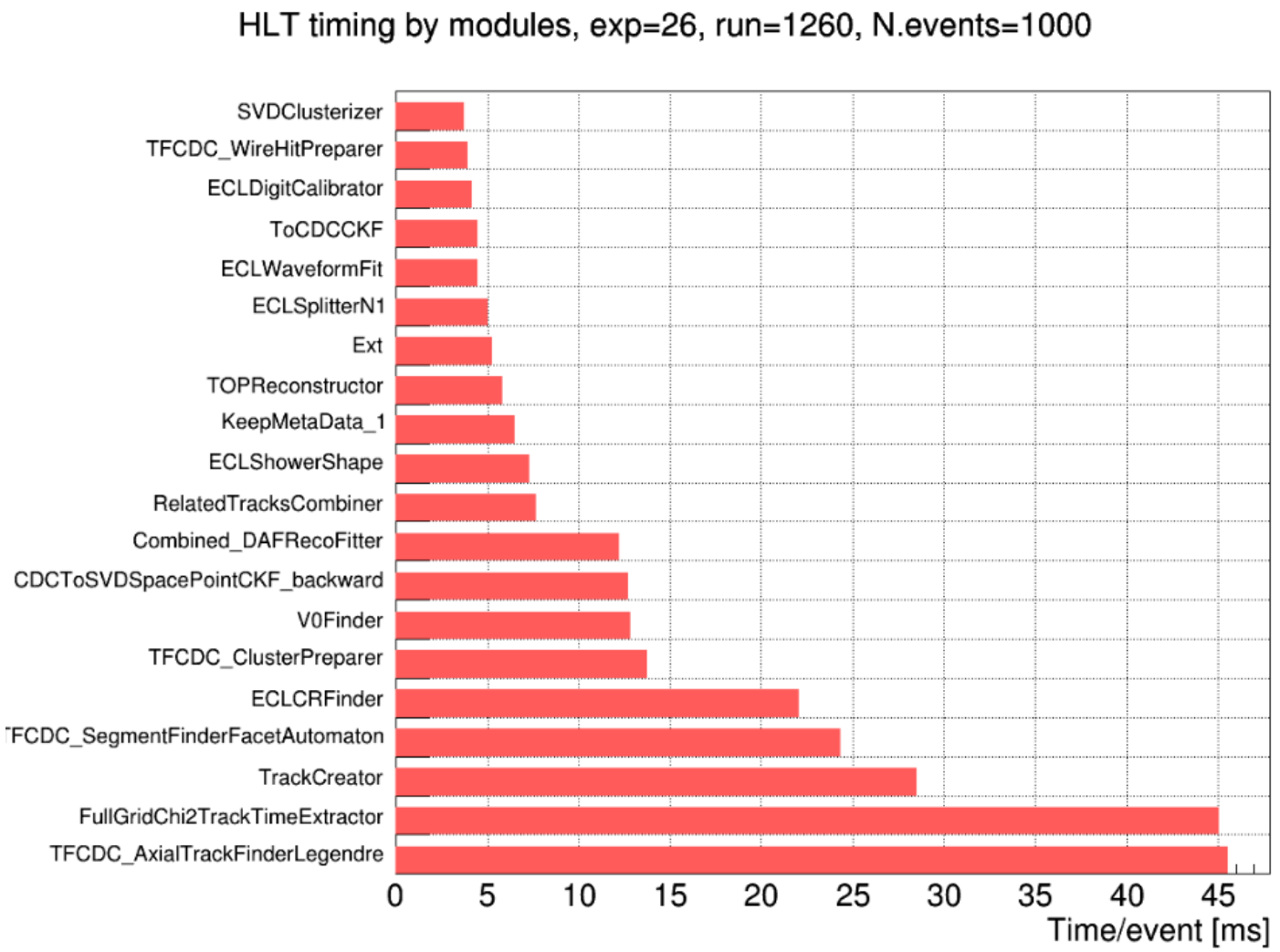
Un-purity of the tracks



More effective bkg rejection
(improving in high-bkg scenario)

promising results, possible CPU/event gain in order of 15% or so…
→ release-08

However, noticed pattern recognition quickly degraded (exp 26)



Summary

- Optimization during data taking (release-05) allowed us to survive until LS1 (13 kHz)
- with 3 HLT units more + release-08 (event time+ track fitting improvements)

⇒ should be able to reach 20 kHz (Luminosity ~ $10^{35}/cm^2/s$)

- need to carefully monitor tracking performance at higher luminosity
- need to make sure software development keeps CPU budget under control
- no clear path beyond 20 kHz… unless significant improvements in pattern recognition