A Comprehensive Bandwidth Testing Framework for the LHCb Upgrade Trigger System

Luke Grazette, Ross Hunter†, Ella Noomen, Nicole Skidmore, Sascha Stahl, Mika Vesterinen, Shunan Zhang

†University of Warwick, U.K.

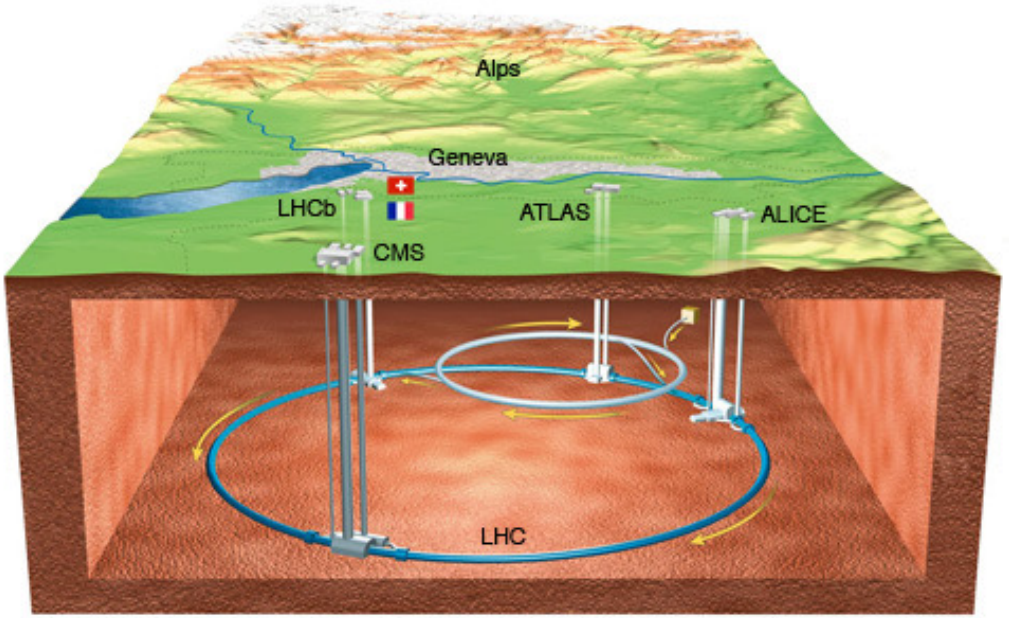
on behalf of the LHCb Collaboration

27th Computing in High Energy and Nuclear Physics Conference,

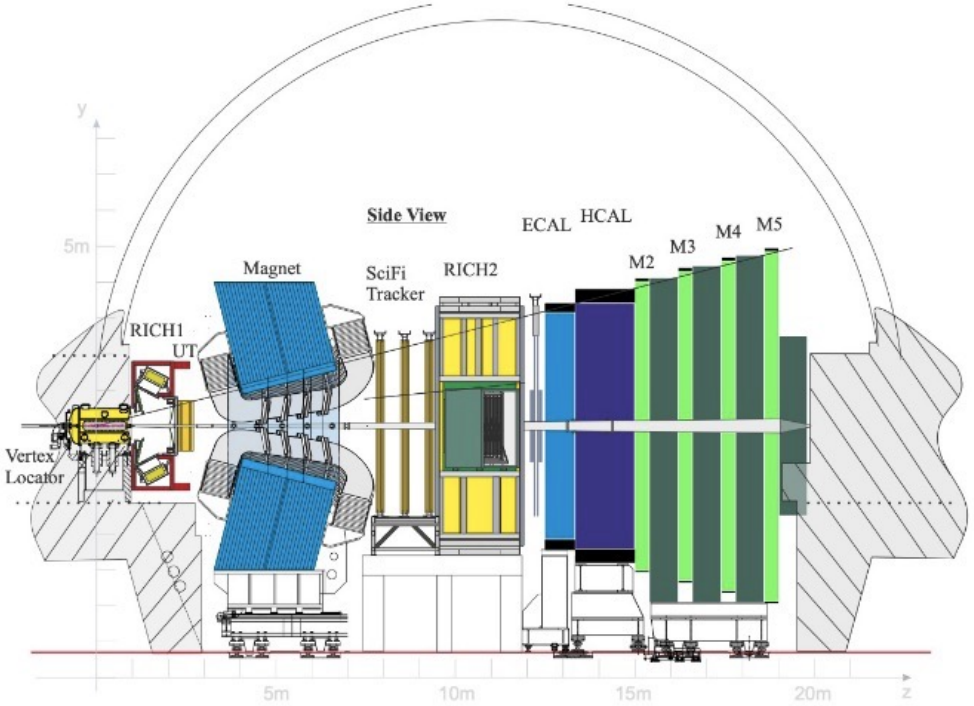
19th -25th October 2024, Kraków, Poland

European Research Council
Established by the European Commission

The upgraded LHCb experiment

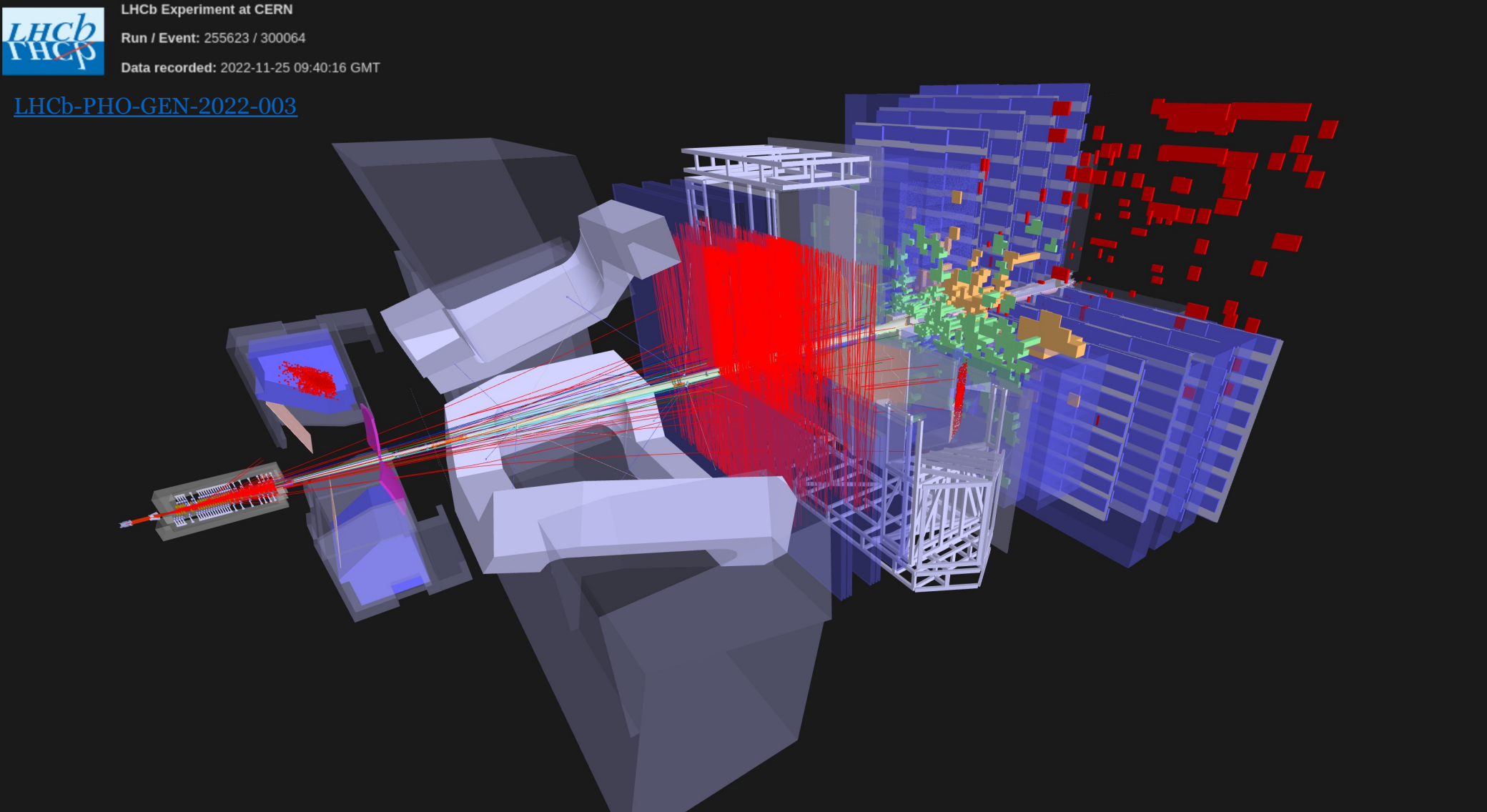


[LHCb Outreach]



[JINST 19 (2024) P05065]

Full replacement of tracking sub-detectors during Long Shutdown 2 (2018-2022).
Proton-proton data-taking restarted in 2022.



30 million visible events / second ; 5x collisions / event as 2018.

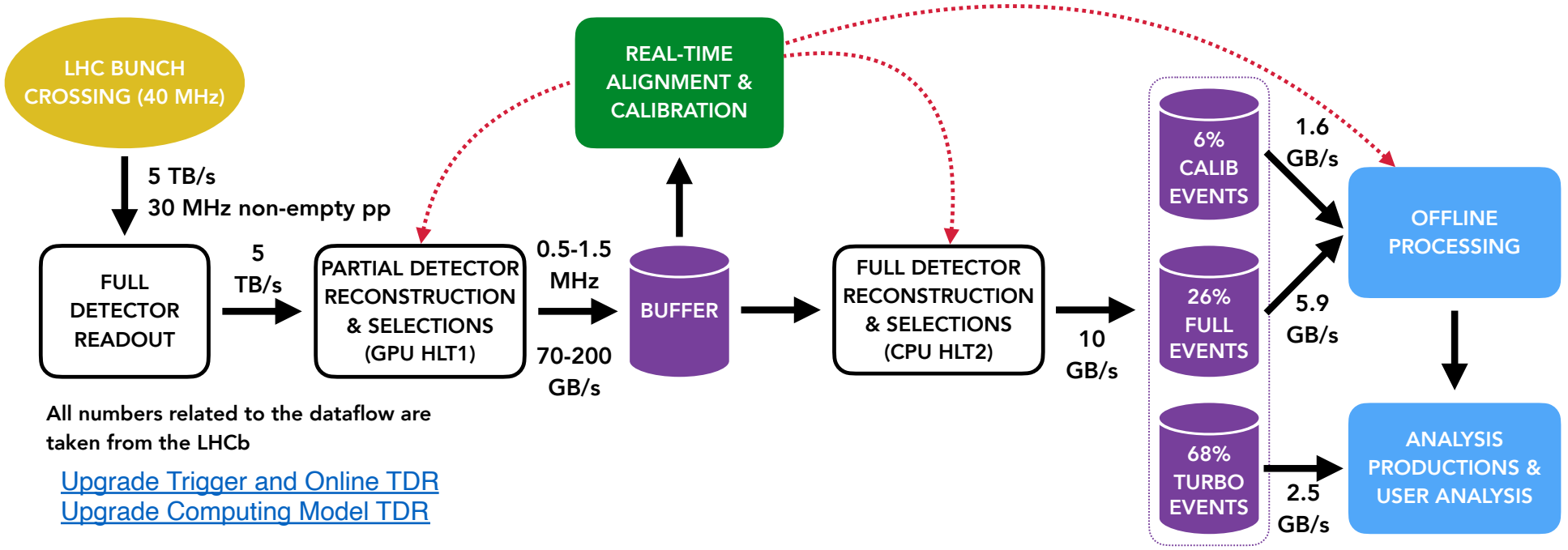
~5 TB / second of raw detector data!

Filter to an affordable data rate of interesting collisions with a trigger system.

The upgraded LHCb trigger

Full software system of reconstruction & selection algorithms:

LHCB-FIGURE-2020-016



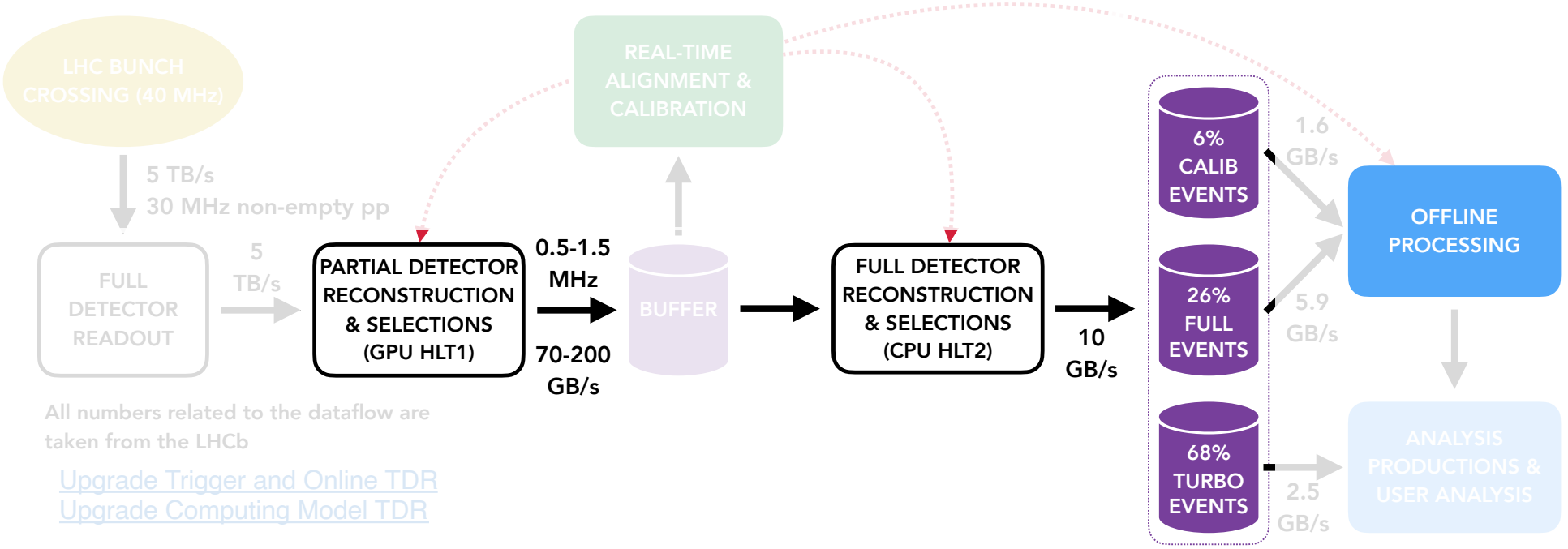
All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR
Upgrade Computing Model TDR

The upgraded LHCb trigger

Full software system of reconstruction & selection algorithms:

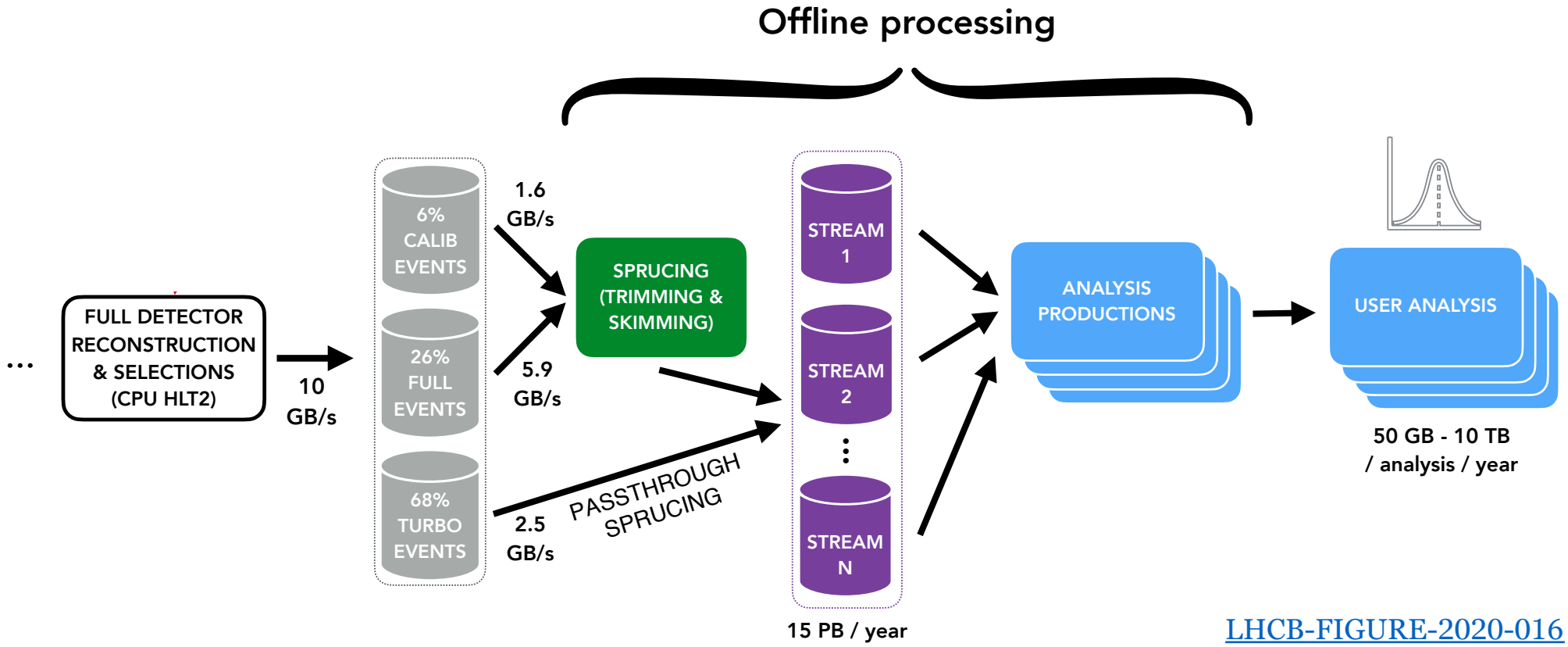
LHCB-FIGURE-2020-016



- GPU HLT1 at 30 MHz.
  - *The automated Bandwidth Division for the LHCb first-level trigger*
    Joshua Horswill, Tuesday Parallel (2)
- CPU HLT2 (more complex selections) writing to multiple streams (files).

The upgraded LHCb trigger

Full software system of reconstruction & selection algorithms:



LHCB-FIGURE-2020-016

- Sprucing (offline/deferred) writing to working-group (charm, $b \to c$ etc) streams (files).
  - Input for physics analysis.
  - *Sprucing and Analysis Productions: Offline data processing in LHCb without the pain*
    Nicole Skidmore, Monday Parallel (3)

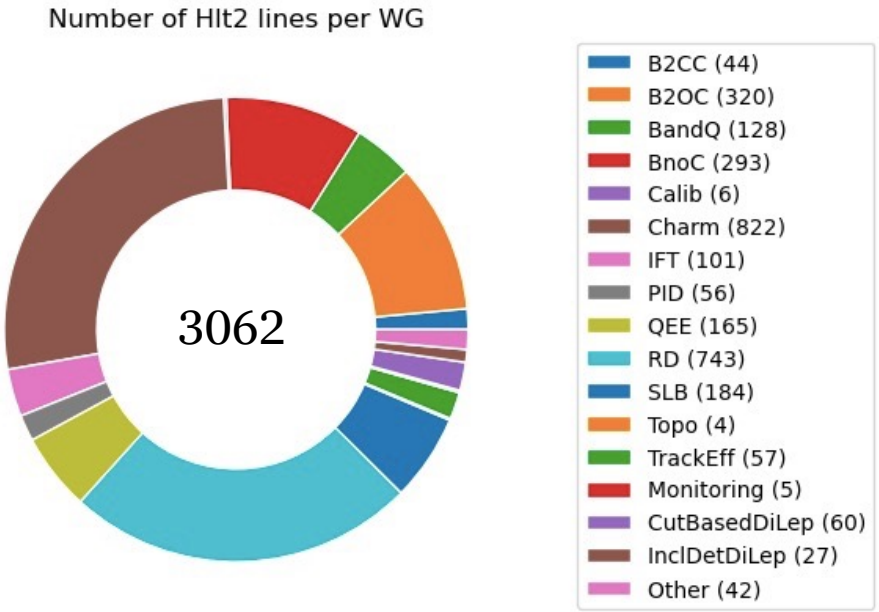
Trigger “lines” in LHCb

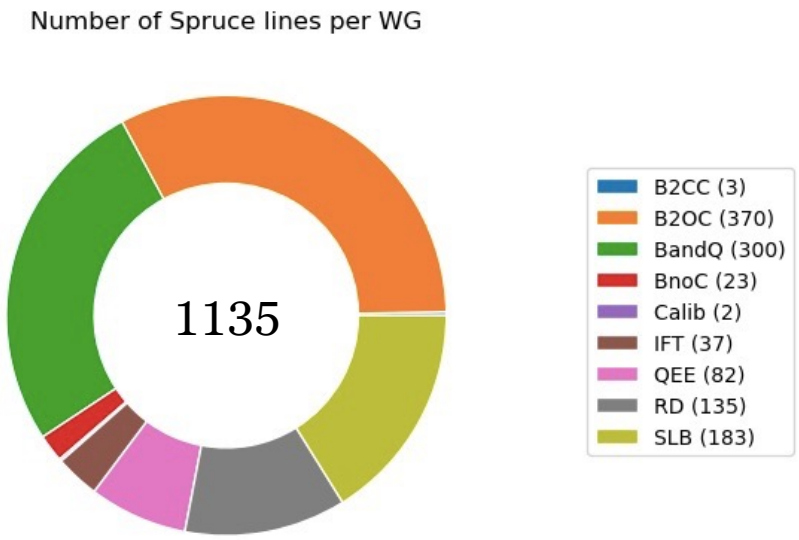
- Each stage = collection of “lines” (algorithms) targeting specific event types/signals.
  - e.g. two tracks from a displaced vertex, or topology matching $B_s \to J/\psi(\to \mu\mu)\phi(\to KK)$

- HLT1: O(50) lines.
  - See Joshua Horswill’s talk for more info.

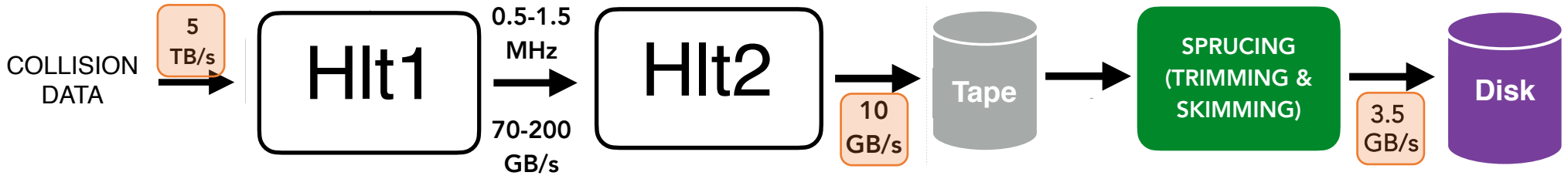
Trigger “lines” in LHCb

- Each stage = collection of “lines” (algorithms) targeting specific event types/signals.
  - e.g. two tracks from a displaced vertex, or topology matching $B_s \to J/\psi(\to \mu\mu)\phi(\to KK)$
- HLT1: O(50) lines.
  - See Joshua Horswill’s talk for more info.
- HLT2 and Sprucing: O(4000) lines!
  - 100s of authors, 10000s of parameters.





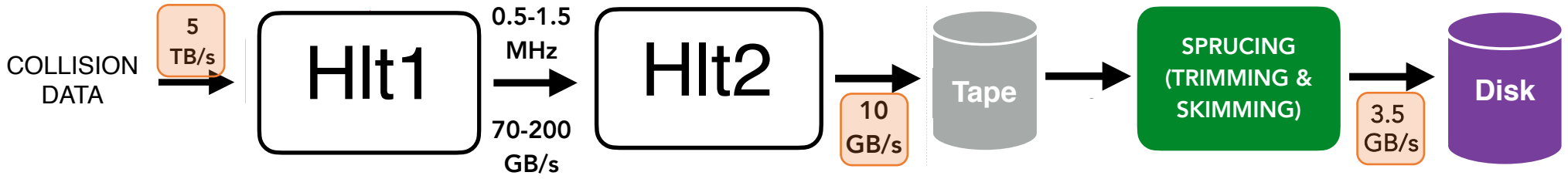
Bandwidth division in LHCb



- Target bandwidth (GB/s) ~ storage budget (CHF).

Upgrade Computing Model TDR

Bandwidth division in LHCb



Upgrade Computing Model TDR

- Target bandwidth (GB/s) ~ storage budget (CHF).
- Easy enough to measure coarsely…

```
[ross@lhcb-storage ~]$ ls -lrt --si my_trigger_output.file
-rw-r-----  1 ross ross  HUGE Oct  3 13:52 my_trigger_output.file
```

…but difficult to share fairly between all the trigger lines.

Bandwidth division in LHCb

Upgrade Computing Model TDR

- Target bandwidth (GB/s) ~ storage budget (CHF).
- Easy enough to measure coarsely…

```
[ross@lhcb-storage ~]$ ls -lrt --si my_trigger_output.file
-rw-r----- 1 ross ross HUGE Oct 3 13:52 my_trigger_output.file
```

…but difficult to share fairly between all the trigger lines.

- HLT1: O(50) lines; automated, unbiased division.
  - See Joshua Horswill's talk for more info.
- HLT2 & Sprucing: >4000 lines; automated division is intractable; divide "by-hand".
  - Requirements less stringent: 10x data reduction in HLT2, ~50x in HLT1; storage is ~adaptable.

Bandwidth division in LHCb

Upgrade Computing Model TDR

- Target bandwidth (GB/s) ~ storage budget (CHF).
- Easy enough to measure coarsely…

```
[ross@lhcb-storage ~]$ ls -lrt --si my_trigger_output.file
-rw-r-----  1 ross ross  HUGE Oct  3 13:52 my_trigger_output.file
```

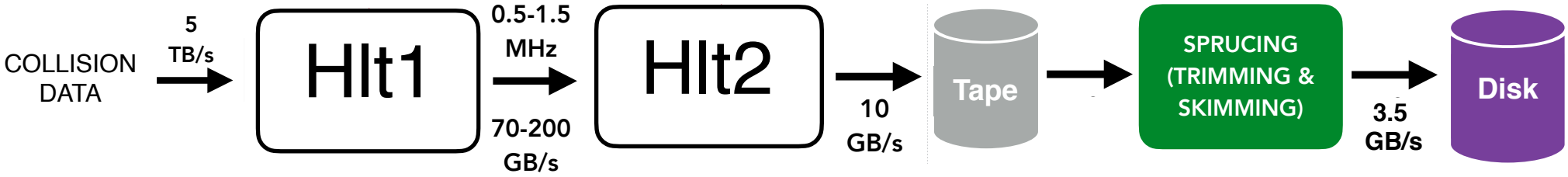
…but difficult to share fairly between all the trigger lines.

- HLT1: O(50) lines; automated, unbiased division.
  - See Joshua Horswill's talk for more info.
- HLT2 & Sprucing: >4000 lines; automated, fair division is intractable; divide "by-hand".
  - Requirements less stringent: 10x data reduction in HLT2, ~50x in HLT1; storage is ~adaptable.

Bandwidth decisions need monitoring tools that work at micro (author of 1 line) and macro (trigger management) levels.

Bandwidth testing framework

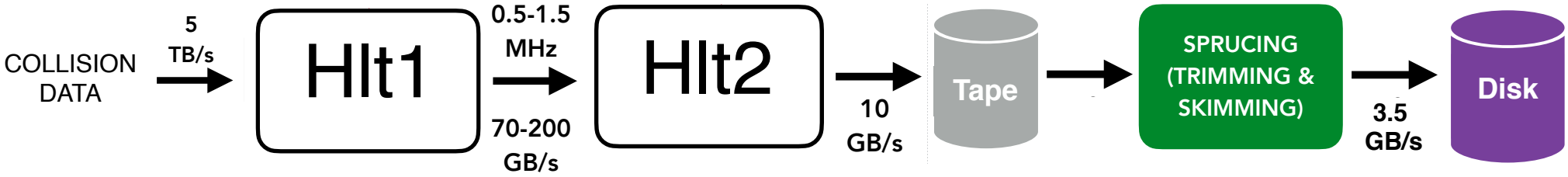
- Real trigger & offline processing (simplified):



Upgrade Computing Model TDR

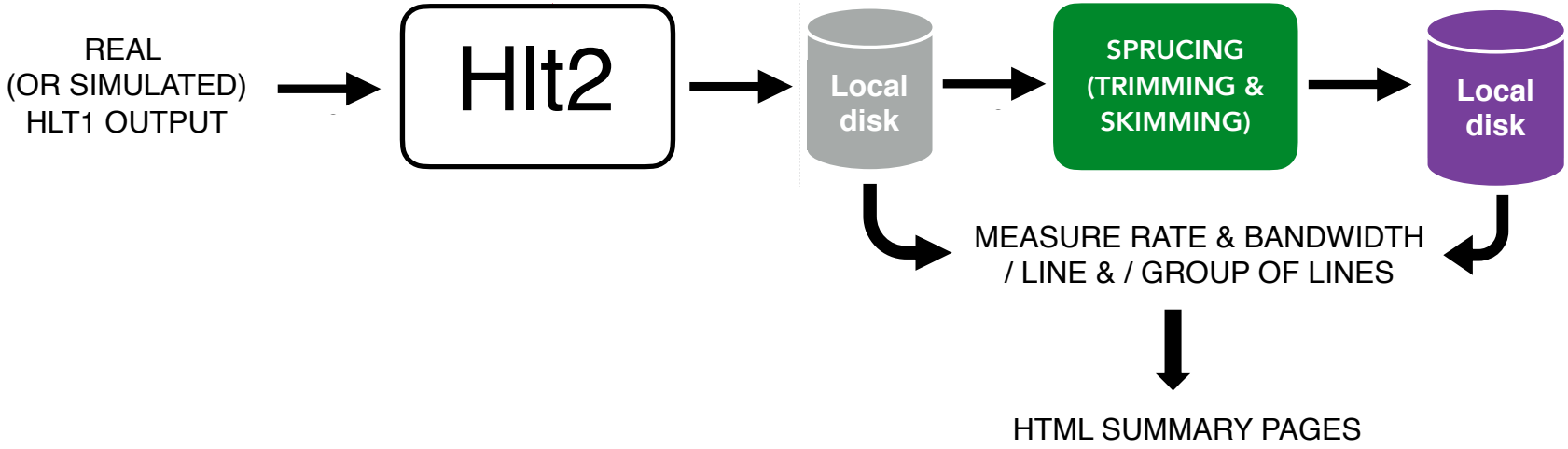
Bandwidth testing framework

- Real trigger & offline processing (simplified):



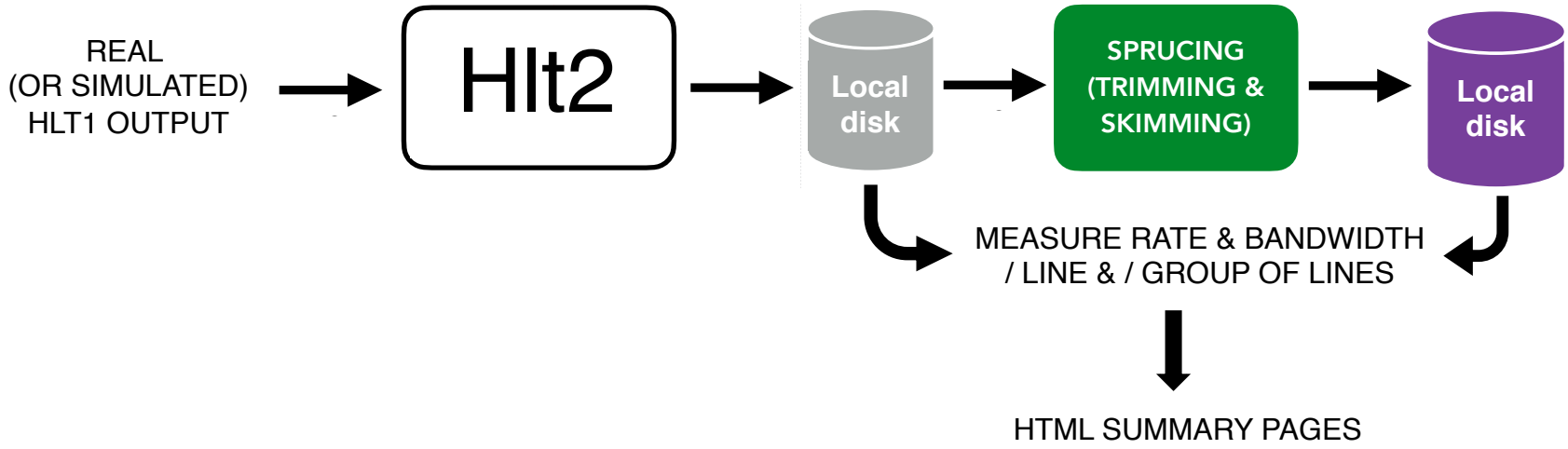
Upgrade Computing Model TDR

- UpgradeRateTest on a test machine:



Bandwidth testing framework

- UpgradeRateTest:

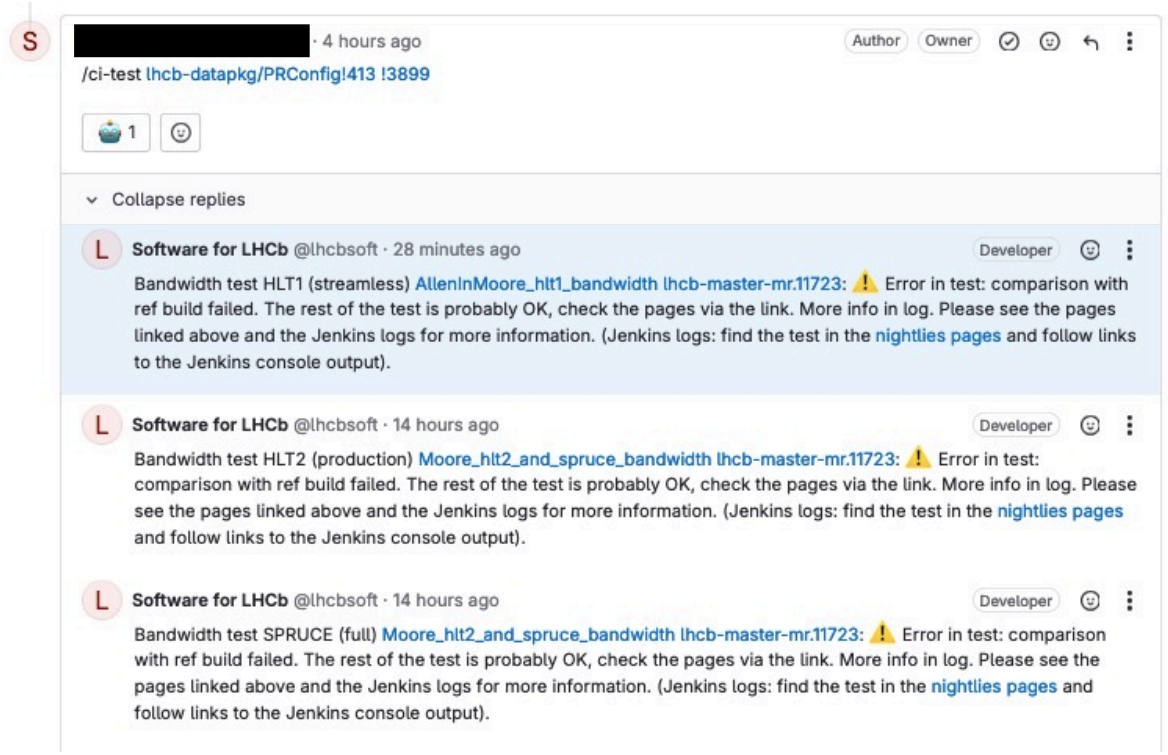


- Run on test machine from build of LHCb trigger software stack,
- Tests run nightly and at request of software testers on Gitlab merge requests,

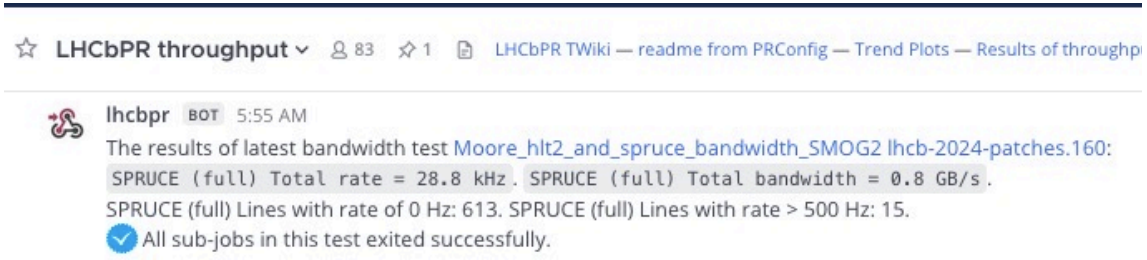
- Scheduled with Jenkins with LHCbPR (*EPJ Web Conf.* 214 (2019) 05042),
- 100k input events takes ~1 hour.

Constructive test feedback

- On Gitlab merge request:



- Mattermost (instant messaging):



...following the links...

HTML “dashboard”

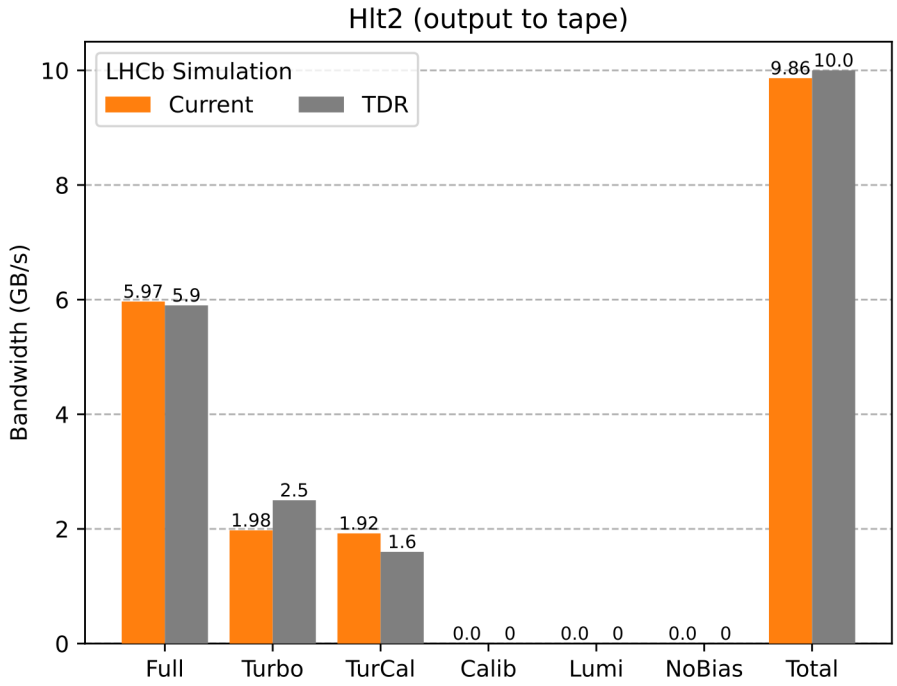
- Big-picture:

All sub-jobs in this test exited successfully.

This page contains the results of the hlt2 bandwidth test with the production streaming configuration. Scroll down to see:

- Summary of main results,
- Details of the streaming configuration,
- Links to other html pages produced by this test,
- Bar charts of rate and bandwidth for each WG within each stream (HLT2 only),
- A pie chart of all lines split by WGs (HLT2 and sprucing only),
- Information about the input sample,
- Stacked histograms of all lines, split by WG, of rate/bandwidth metrics.

Main results:

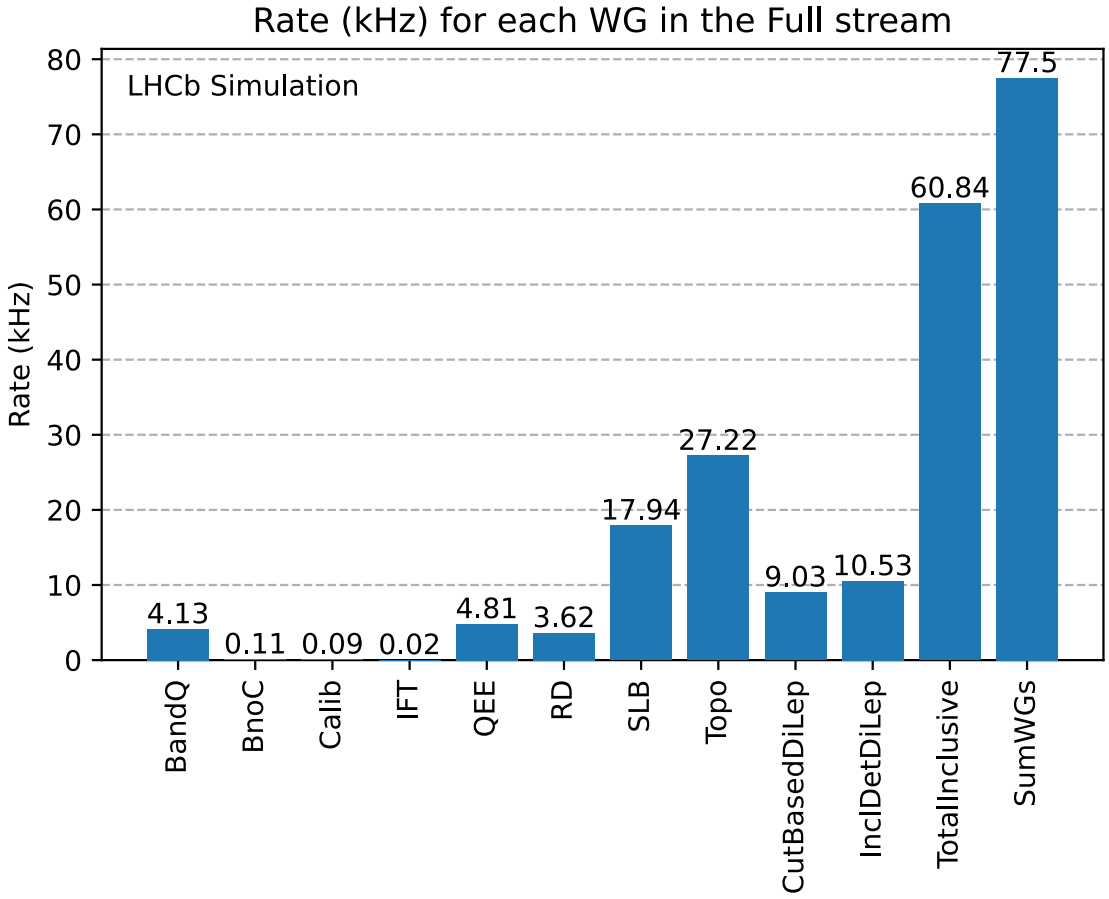


[LHCb-FIGURE-2024-034](#)

[Upgrade Computing Model TDR](#)

HTML “dashboard”

- Big-picture:



LHCb-FIGURE-2024-034

HTML “dashboard”

- In-depth:

FULL

BandQ

	Line	Total Retention (%)	Rate (kHz)	Exclusive Retention(%)	Exclusive Rate (kHz)	Avg Total Event Size (kB)	Total Bandwidth (GB/s)	Avg DstData Size (kB)	DstData Bandwidth (GB/s)
0	Hlt2_JpsiToMuMuDetachedFullDecision	0.065	0.651	0.032	0.32	92.7	0.0603	92.3	0.0601
1	Hlt2BandQ_BsForSpectroscopyFullDecision	0.017	0.17	0.007	0.0701	236	0.0402	235	0.0401
2	Hlt2BandQ_BdForSpectroscopyFullDecision	0.024	0.24	0.01	0.1	163	0.0393	163	0.0392
3	Hlt2BandQ_BuForSpectroscopyFullDecision	0.022	0.22	0.007	0.0701	161	0.0355	161	0.0354
4	Hlt2BandQ_LbForSpectroscopyFullDecision	0.013	0.13	0.002	0.02	212	0.0276	211	0.0275
5	Hlt2BandQ_OmegabForSpectroscopyFullDecision	0.009	0.0901	0.005	0.05	257	0.0231	256	0.0231
6	Hlt2_DiMuonJPsiTightFullDecision	0.027	0.27	0.014	0.14	76.3	0.0206	75.8	0.0205
7	Hlt2BandQ_DiMuonSameSignIncFullDecision	0.019	0.19	0.015	0.15	100	0.019	99.6	0.0189
8	Hlt2BandQ_Xib0ForSpectroscopyFullDecision	0.007	0.0701	0	0	267	0.0187	267	0.0187
9	Hlt2_Psi2SToMuMuDetachedFullDecision	0.018	0.18	0.011	0.11	102	0.0183	101	0.0183

Comparison between lhcb-2024-patches.159 and lhcb-2024-patches-mr.1628 under different streaming configurations

All comparison tables were made successfully

Changes in rates per line:

	Line	Rate (kHz)		
		Ref	**New**	**Change (%)**
0	SpruceQEE_IncSVTagDijetsDecision	0.36	0.35	-2.78

Tracking over time

LHCBPR

Results of Upgrade Rate Tests

Rate tests

BandwidthTest_lhcb-2024-patches.160_Moore_hlt2_and_spruce_bandwidth_SMOG2_x86_64_v3-el9-gcc13+detdesc-opt+g_2024-10-10_05:40:52_+0200
BandwidthTest_lhcb-2024-patches.160_AllenInMoore_hlt1_bandwidth_x86_64_v3-el9-gcc13+detdesc-opt+g_2024-10-10_03:54:36_+0200
BandwidthTest_lhcb-2024-patches.160_Moore_spruce_latest_bandwidth_x86_64_v3-el9-gcc13-opt+g_2024-10-10_03:18:05_+0200
BandwidthTest_lhcb-2024-patches.160_Moore_hlt2_and_spruce_bandwidth_x86_64_v3-el9-gcc13-opt+g_2024-10-10_02:13:56_+0200
BandwidthTest_lhcb-2024-patches.160_Moore_hlt2_bandwidth_x86_64_v3-el9-gcc13-opt+g_2024-10-10_01:21:40_+0200
BandwidthTest_lhcb-2024-patches-mr.1628_AllenInMoore_hlt1_bandwidth_x86_64_v3-el9-gcc13+detdesc-opt+g_2024-10-09_18:49:24_+0200
BandwidthTest_lhcb-master-mr.11723_Moore_hlt2_and_spruce_bandwidth_x86_64_v3-el9-gcc13-opt+g_2024-10-09_17:51:24_+0200
BandwidthTest_lhcb-master-mr.11723_AllenInMoore_hlt1_bandwidth_x86_64_v3-el9-gcc13+detdesc-opt+g_2024-10-09_16:00:53_+0200

Facilitating exploratory data analysis

- What if your quantity of interest isn't in the dashboard?
- All intermediary files are available for download: allows for fast insights.

Facilitating exploratory data analysis

- What if your quantity of interest isn't in the dashboard?
- All intermediary files are available for download: allows for fast insights.
- E.g. 20 lines of data-frame hacking found the worst discrepancies between real data and simulated data…

Comparison of rates and bandwidths for hlt2, production stream (by Rate):

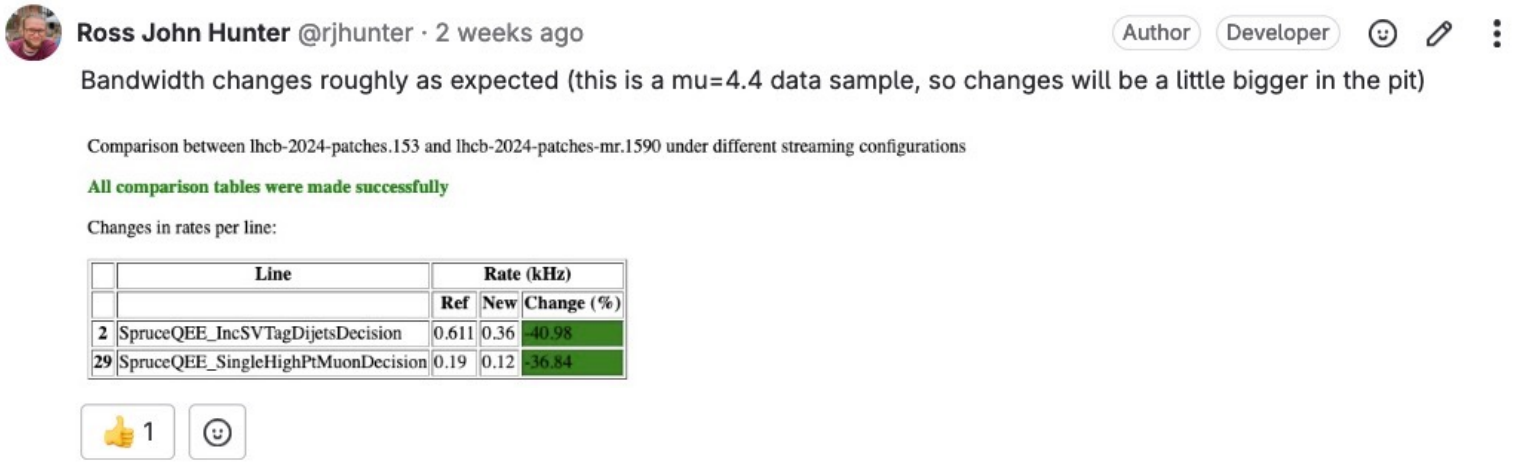
	Line	Rate (kHz)_mc	Rate (kHz)_data	Rate Data/MC
350	Hlt2QEE_MDS_BDT_nHitsDecision	0.11	9.07	82.454545
1507	Hlt2B2OC_LbToD0PPi_D0ToKsTTHHDecision	0.01	0.29	29.000000
1351	Hlt2B2OC_LbToD0PK_D0ToKsTTHHDecision	0.01	0.28	28.000000
1206	Hlt2BandQ_OmegabForSpectroscopyFullDecision	0.01	0.20	20.000000
1548	Hlt2BandQ_BuToDspPPbarDecision	0.01	0.18	18.000000
1402	Hlt2BandQ_EMDM_B2D3Pi_D2KS3Pi_TTDecision	0.02	0.28	14.000000
1050	Hlt2BandQ_LbToLcDsmPiPiDecision	0.01	0.14	14.000000
788	Hlt2B2CC_Bs0ToJpsiPhi_JpsiToPPDecision	0.03	0.39	13.000000
1073	Hlt2BandQ_XibToJpsiPKPi_JpsiToMuMuDecision	0.01	0.13	13.000000
1209	Hlt2B2OC_BcToD0Dsp_D0ToHHHH_DspToHHHDecision	0.01	0.12	12.000000

- …and helped us validate that all lines working on simulation worked in real data:

```
Lines with Rate = 0 in data but > 100 Hz in MC for spruce, wg stream:
Empty DataFrame
Columns: [Line, Rate (kHz)_mc, Rate (kHz)_data]
Index: []
Total number of lines: 0
```

Impact

- Micro level: MRs →
  Bandwidth informs merging.



Ross John Hunter @rjhunter · 2 weeks ago

Author Developer

Bandwidth changes roughly as expected (this is a mu=4.4 data sample, so changes will be a little bigger in the pit)

Comparison between lhcb-2024-patches.153 and lhcb-2024-patches-mr.1590 under different streaming configurations

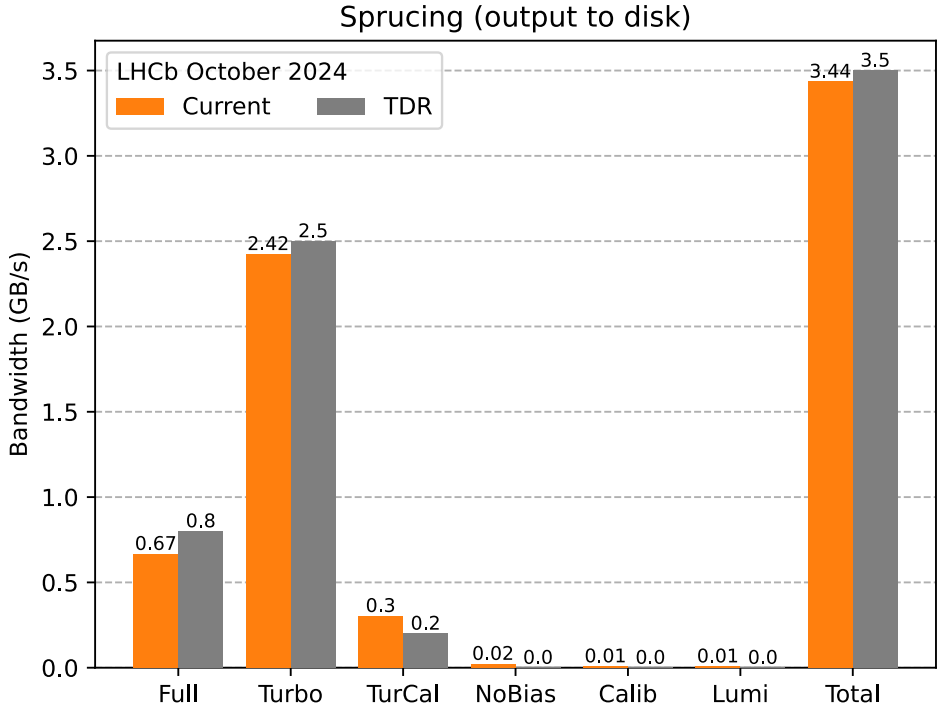
All comparison tables were made successfully

Changes in rates per line:

	Line	Rate (kHz)		
		Ref	New	Change (%)
2	SpruceQEE_IncSVTagDijetsDecision	0.611	0.36	-40.98
29	SpruceQEE_SingleHighPtMuonDecision	0.19	0.12	-36.84

1

- Macro level: current trigger status →
  (even allowing HLT1 to keep 20% more rate than design - see Joshua's Horswill talk.)



Upgrade Computing Model TDR
LHCb-FIGURE-2024-034

Summary

- The upgraded LHCb experiment is operating at 5x previous collision rate, with a full-software heterogeneous-architecture trigger system.
  - More info: *Performance of the LHCb heterogeneous software trigger* - Alessandro Scarabotto, Monday Parallel (2)

Summary

- The upgraded LHCb experiment is operating at 5x previous collision rate, with a full-software heterogeneous-architecture trigger system.
  - More info: *Performance of the LHCb heterogeneous software trigger* - Alessandro Scarabotto, Monday Parallel (2)

- Stringent operational targets necessitate a comprehensive bandwidth monitoring & testing framework for 100s of contributors and 1000s of selection algorithms.

Summary

- The upgraded LHCb experiment is operating at 5x previous collision rate, with a full-software heterogeneous-architecture trigger system.
  - More info: *Performance of the LHCb heterogeneous software trigger* - Alessandro Scarabotto, Monday Parallel (2)

- Stringent operational targets necessitate a comprehensive bandwidth monitoring & testing framework for 100s of contributors and 1000s of selection algorithms.

- Automated “UpgradeRateTest” framework emulates our trigger system, measures bandwidth nightly and on-demand.

Summary

- The upgraded LHCb experiment is operating at 5x previous collision rate, with a full-software heterogeneous-architecture trigger system.
  - More info: *Performance of the LHCb heterogeneous software trigger* - Alessandro Scarabotto, Monday Parallel (2)
- Stringent operational targets necessitate a comprehensive bandwidth monitoring & testing framework for 100s of contributors and 1000s of selection algorithms.
- Automated “UpgradeRateTest” framework emulates our trigger system, measures bandwidth nightly and on-demand.
- Feedback comes as instant-messaging notifications and on Gitlab MRs, leading to a HTML dashboard designed for management and individual line authors.

Summary

- The upgraded LHCb experiment is operating at 5x previous collision rate, with a full-software heterogeneous-architecture trigger system.
  - More info: *Performance of the LHCb heterogeneous software trigger* - Alessandro Scarabotto, Monday Parallel (2)

- Stringent operational targets necessitate a comprehensive bandwidth monitoring & testing framework for 100s of contributors and 1000s of selection algorithms.

- Automated “UpgradeRateTest” framework emulates our trigger system, measures bandwidth nightly and on-demand.

- Feedback comes as instant-messaging notifications and on Gitlab MRs, leading to a HTML dashboard designed for management and individual line authors.

- Framework enables bandwidth-driven decision-making, and has helped us build a trigger that fits within operational constraints, even at 20% higher input than originally planned.

Thank you for your attention.
Any questions?