Data squeezing in CMS: techniques & physics results

Silvio Donato
(Univ. Pisa & INFN Pisa)
on behalf of
the CMS collaboration

Isola D'Elba
25 Maggio 2026

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INFN PISA

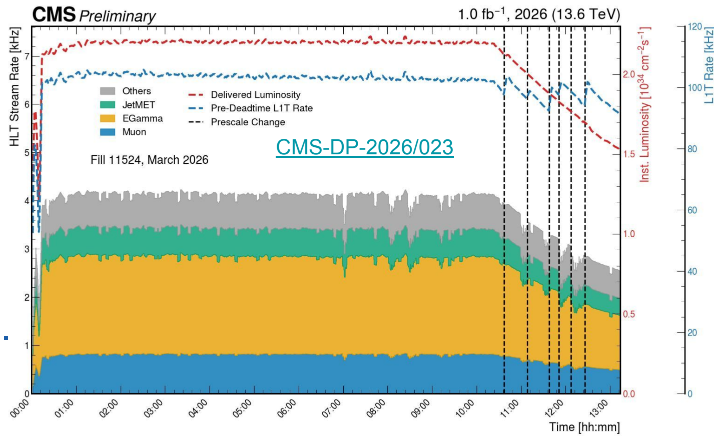
Outline

- The CMS trigger system
- Data scouting
  - Event size
  - Event content
  - Run-3 results
- Raw Prime/Second
  - The Heavy Ions challenge
  - The compression of strip detector raw data
  - The evolution: Raw Second
    - Entropy reduction
    - Results
- Outlook for HL-LHC
- Conclusions

The CMS trigger system

- Bunch crossing rate **40 MHz**
- Hardware trigger (L1): ~30 MHz → **100 kHz**
  - simplified readout (no tracker), small latency (<4μs).
- Software trigger (HLT): 100 kHz → **~3 kHz**.
  - full event readout available (**~1 MB/event**);
  - simplified reco: ~50k CPUs → **~500ms/event** on average.
- Storage and offline reconstruction

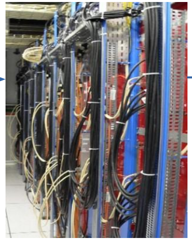
***Standard* HLT rate during a fill**



Simplified readout

40 MHz

L1 trigger



~100 kHz

Full readout

~100 kHz

~100 GB/s

HLT

~3 kHz

(~500 ms)

6 GB/s

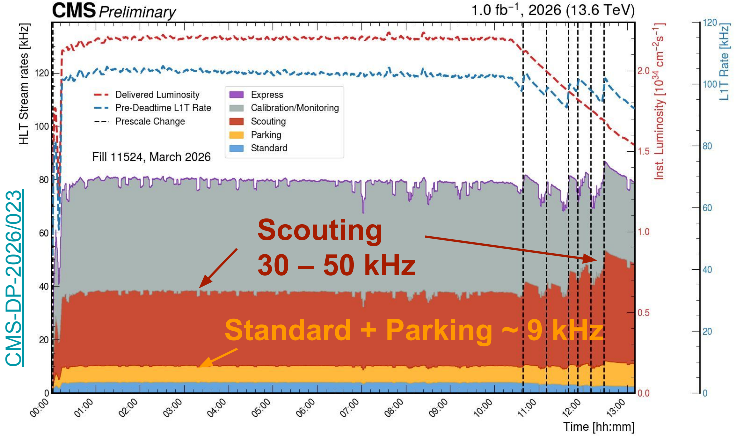
Data storage

Data Scouting

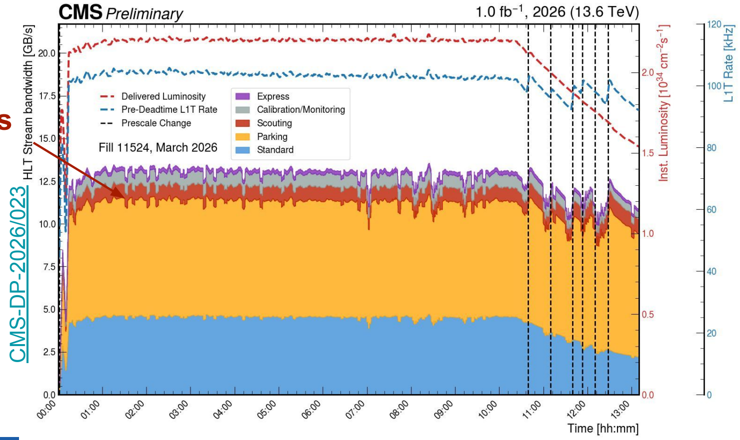
Data scouting and parking

- Since Run-1, CMS developed two techniques to go beyond these limits.
- **Data parking**: delay of the offline reconstruction to when resources are available.
  - More info in EXO-23-007.
- **Data scouting**: save the HLT reconstructed objects instead of full raw data
  - No offline reconstruction;
  - Smaller event size.
- **Scouting** has been revolutionized in Run-3:
  - larger rate, 30 – 50 (during lumi decay) kHz
    - ie. up to 50% of the L1-accepted events
  - save **all main physics objects** reconstructed at HLT.

HLT rate during a fill



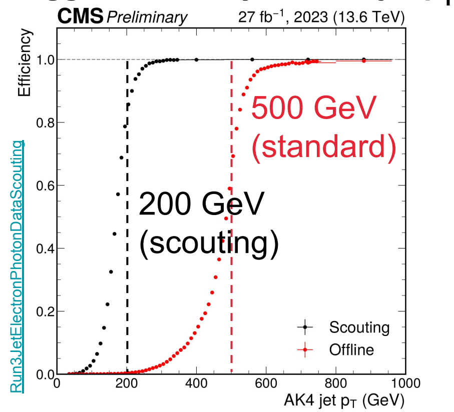
HLT bandwidth



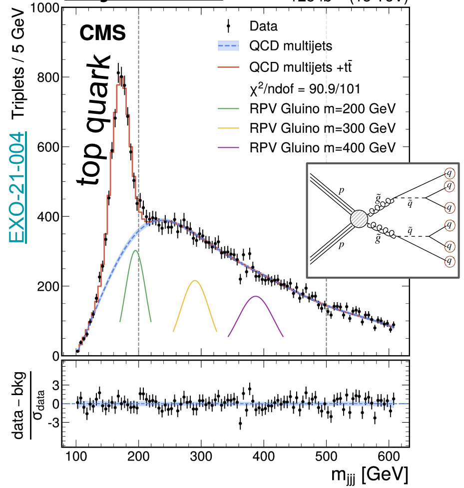
Data scouting event size

- High trigger rate → low trigger thresholds
  - eg. make possible to see **top** quark in **fully hadronic** channel
- Event size:
  - Full **raw** data (standard trigger) ~ **1000 kB/event**
  - **Scouting** data ~ **25 kB/event**, using fast compression (ZSTD lev. 3)
    - up to 50% of further lossless compression using slow compression (LZMA lev. 4)
  - Variables are stored with reduced numerical precision (**10 bits**) on the **mantissa**
- Run-3 data scouting is somehow similar to a small format (MINIAOD) used for data analysis

Trigger efficiency vs sum jet p_T



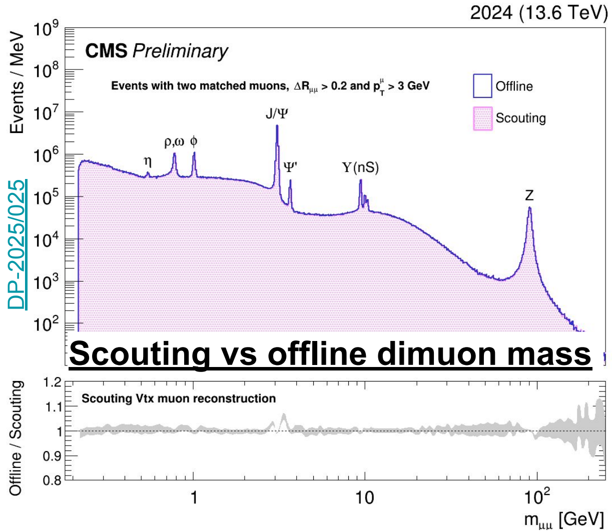
Trijet mass



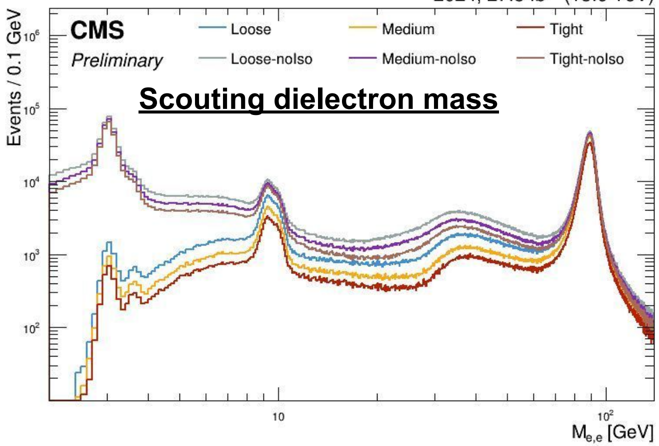
Data scouting event content

- Scouting contains:
  - **Particle Flow candidates**: p_T > 0.6 GeV and |η| < 3
  - **Jets**: p_T > 20 GeV and |η| < 5
  - **Muons**: p_T > 3 GeV
  - **Electrons and Photons**: p_T > 1 GeV,|η| < 2.65, HCAL/ECAL energy fraction <0.2
  - All **primary vertices**
- Recent changes:
  - Raw **PPS** data allowing to reconstruct and tag offline very forward intact **protons**
  - Energy and timing of **hits** from **HCAL** and **ECAL** to look for anomalous signatures
  - **Tracks** with p_T > 1 GeV and their covariant fit matrix, necessary for B physics analyses
    - 2022-23: All tracks reconstructed with **pixel** detector (r~16 cm)
    - 2024-: Tracks from the primary vertex reconstructed with **pixel+strip** (r~105 cm)
- Excellent performance, similar to offline, especially for muons

DP-2025/025

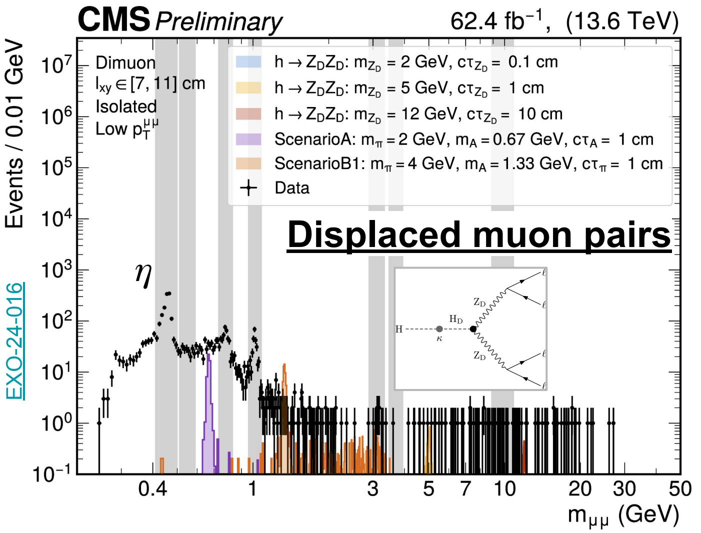


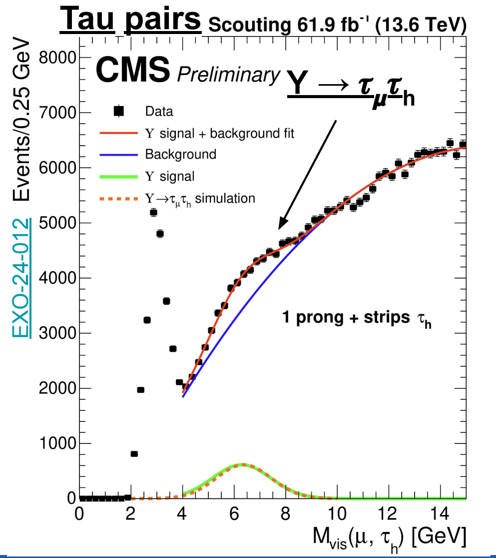
DP-2026/048

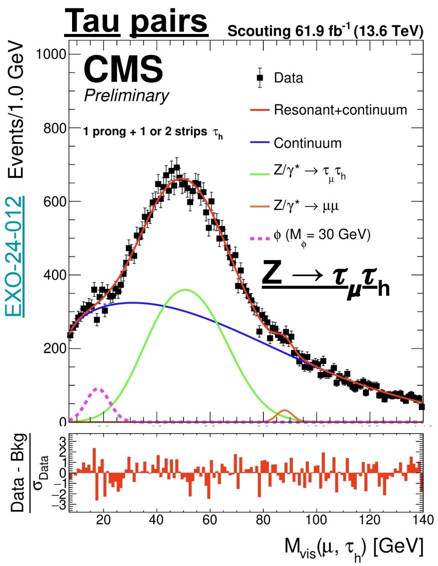


Data scouting - Run-3 data results

- Search for displaced muons (22-23 data, 62.4 fb^{-1}) → already competitive with Run-2
- Possibility to reconstruct objects offline (eg. τ) from scouting PF candidate enabled **the first search at the LHC for τ-pair resonance below 60 GeV**
  - Υ(9.4 GeV) → $\tau\tau$ clearly visible, perfect agreement with simulation





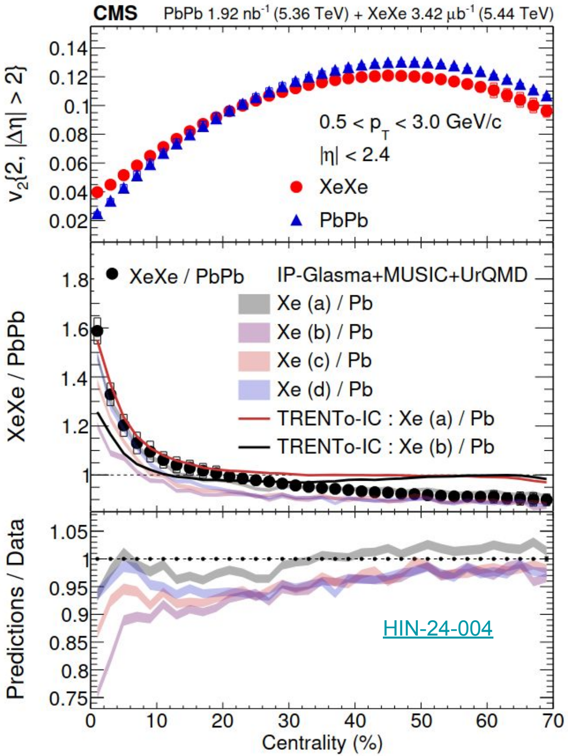


RawPrime and RawSecond

PbPb collisions and RawPrime

- For heavy ions runs, many data analysis requires unbiased data (eg. quark-gluon plasma studies).
  - The analysis sensitivity depends on the number of unbiased events collected.
  - CMS runs the DAQ at the limit to collect more events,
    - In 2024, the bandwidth has been increased to 32 GB/s by using a buffer of 250 TB of SSD disk.
- To squeeze more events at fixed bandwidth, CMS developed the **RawPrime compression** obtaining an event size of about 50%.

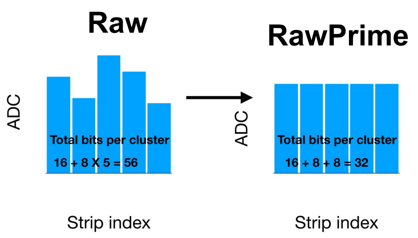
Elliptic flow in PbPb/XeXe



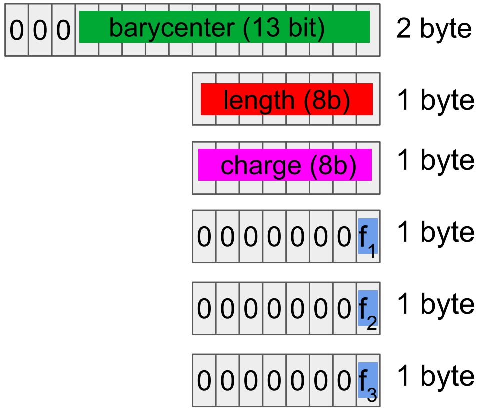
The Raw Prime idea: strip data compression

- Almost ⅔ of the event size of raw data of PbPb collisions is from the strip detector.
- The RawPrime idea is to replace the raw ADC amplitude numbers (8 bits x width) …
- …with the reconstructed strip cluster variables:
  - the **barycenter** in **0.1** of strips, max 7680 [strip module length/0.1] +
  - the **cluster length** in strips, max 255 [2^8-1] +
  - the **average charge** in ADC counts, max 255 [2^8-1] +
  - 3 quality flags, max 1 (boolean)
    - Note: boolean are saved as byte in C++/ROOT, but “0” are strongly compressed
- RawPrime was tested in 2022 and used since 2023

Strip detector

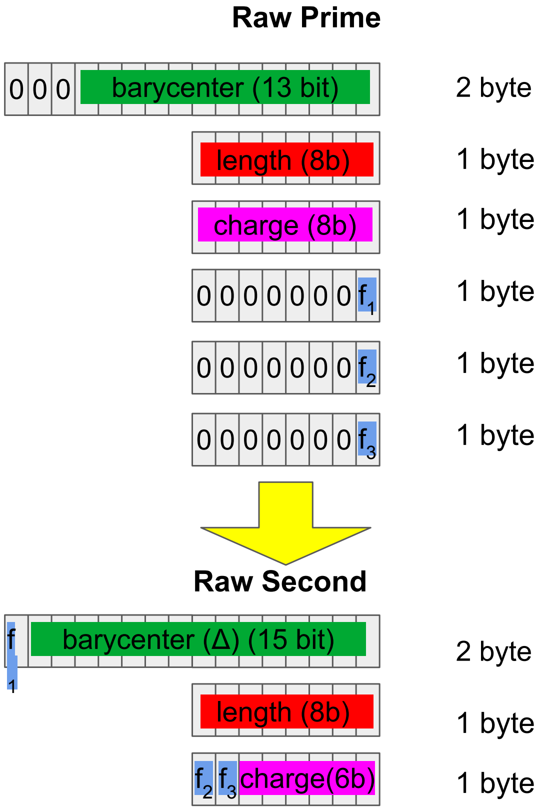


Bit count of Raw Prime



The evolution: Raw Second

- In its evolution (Raw Second), strip cluster data are formatted as follows:
  - the **barycenter (Δ)** (in **0.047** of strips, max 32767 [2^{15}-1]) +
  - the **cluster length** (in strips, max 255 [2^8-1]) +
  - the **average charge** (in **4** ADC counts, max 63 [2^6-1]) +
  - 3 **quality flags** (max 1)
- Changes:
  - Storing the difference between barycenter (Δ) positions instead of the barycenter (next slides)
  - Optimization of the number of bit used per variable:
    - more bits for barycenter (Δ) (13b → 15b), less for charge (8b → 6b)
  - Encoding quality flags in unused bits from other variables
- See detailed example in backup slides

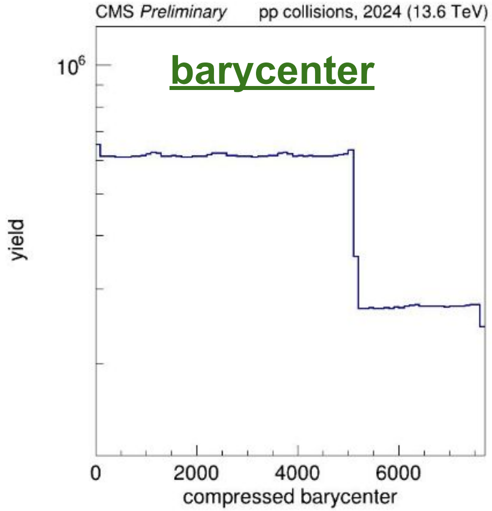


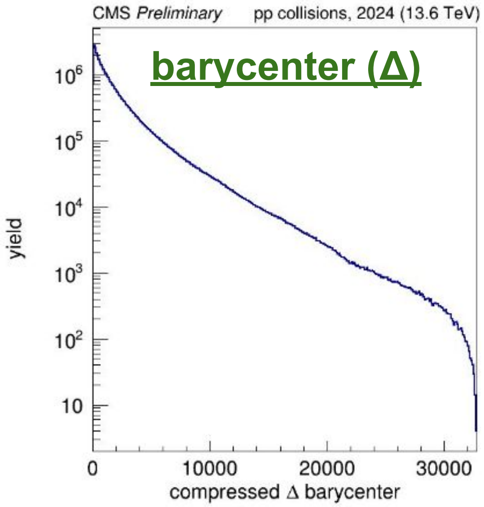
Barycenter: Raw Prime vs Raw Second:

- The probability distribution of barycenter (Δ) more peaked than barycenter
  → lower entropy zero→ stronger compression

Raw Prime

Flat distribution
=
High entropy
=
Low
compression





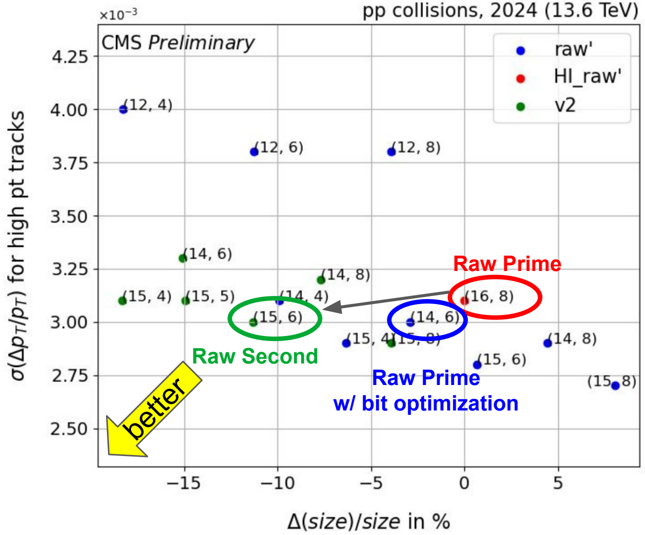
Raw Second

Peaked distribution
=
Low entropy
=
High compression

Bits optimization & Results

- Raw Second outperforms Raw Prime both in data quality and in data compression (-18%)
  - -7% comes from the encoding of the flags in other variables (not included in the plot)
- Further data reduction achievable by applying a cut on the cluster charge to remove the out-of-time pileup cluster and noise (-13%)
- **Raw Second** went **online** for testing in **2025**, showing a **reduction** of **11.4%** on the full event size wrt Raw Prime

Relative difference in p_T due to compression for tracks with p_T>1 GeV vs data compression, for different number of bits (barycenter, charge)



Outlook and Conclusions

Outlook

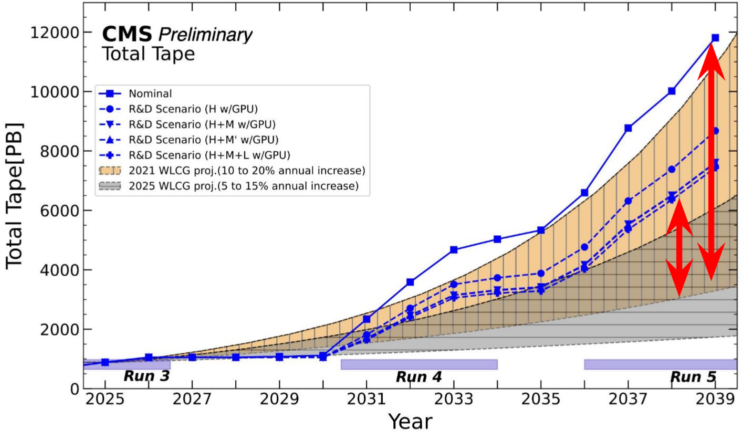
- In Phase-2 the event size will increase from **1.4 MB/ev** to **5.9 MB/ev**
  - because of higher pileup, new detectors with higher granularity and larger acceptance
- The WLCG projections have been updated
  - post-COVID **inflation**, **war**, **AI** revolution
- We are above the projection of ~8000 PB
  - ~3000 PB if including some R&D project (eg. RNtuple)
- Data compression challenge is becoming very **urgent**!

Table 1.2: CMS Phase-2 trigger and DAQ projected running parameters, compared to the design values of the current (Phase-1) system.

CMS-TDR-022

CMS detector	LHC Phase-1	HL-LHC Phase-2	
Peak ⟨PU⟩	60	140	200
L1 accept rate (maximum)	100 kHz	500 kHz	750 kHz
Event Size at HLT input	2.0 MB [a]	6.1 MB	8.4 MB
Event Network throughput	1.6 Tb/s	24 Tb/s	51 Tb/s
Event Network buffer (60 s)	12 TB	182 TB	379 TB
HLT accept rate	1 kHz	5 kHz	7.5 kHz
HLT computing power [b]	0.7 MHS06	17 MHS06	37 MHS06
Event Size at HLT output [c]	1.4 MB	4.3 MB	5.9 MB
Storage throughput [d]	2 GB/s	24 GB/s	51 GB/s
Storage throughput (Heavy-Ion)	12 GB/s	51 GB/s	51 GB/s
Storage capacity needed (1 day [e])	0.2 PB	1.6 PB	3.3 PB

CMS DP -2026/004



Outlook

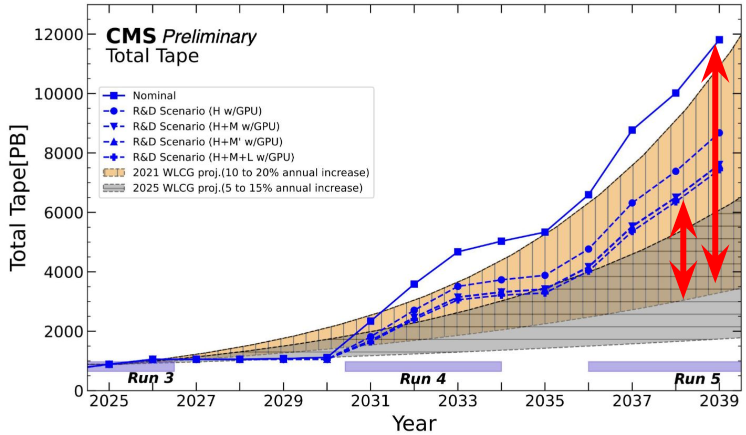
- The idea of saving **low-level reconstructed objects** instead of raw data (Raw Prime) should be investigated and extended to other detectors,
  - esp. the high-granularity calorimeter and inner tracker.
- More info in this report at the 1st NGT workshop.
- NGT aims to perform an offline-like reconstruction at **750 kHz**
  - L1 scouting will run run even at **40 MHz**
- How can we store all these data?
  - Here a strong lossy compression is mandatory
    - 51 GB/s / 750 kHz = 68 kB / event
    - 51 GB/s / 30 MHz = 1.7 kB / event

Table 1.2: CMS Phase-2 trigger and DAQ projected running parameters, compared to the design values of the current (Phase-1) system.

	LHC	HL-LHC	
CMS detector	Phase-1	Phase-2	
Peak ⟨PU⟩	60	140	200
L1 accept rate (maximum)	100 kHz	500 kHz	750 kHz
Event Size at HLT input	2.0 MB [a]	6.1 MB	8.4 MB
Event Network throughput	1.6 Tb/s	24 Tb/s	51 Tb/s
Event Network buffer (60 s)	12 TB	182 TB	379 TB
HLT accept rate	1 kHz	5 kHz	7.5 kHz
HLT computing power [b]	0.7 MHS06	17 MHS06	37 MHS06
Event Size at HLT output [c]	1.4 MB	4.3 MB	5.9 MB
Storage throughput [d]	2 GB/s	24 GB/s	51 GB/s
Storage throughput (Heavy-Ion)	12 GB/s	51 GB/s	51 GB/s
Storage capacity needed (1 day [e])	0.2 PB	1.6 PB	3.3 PB

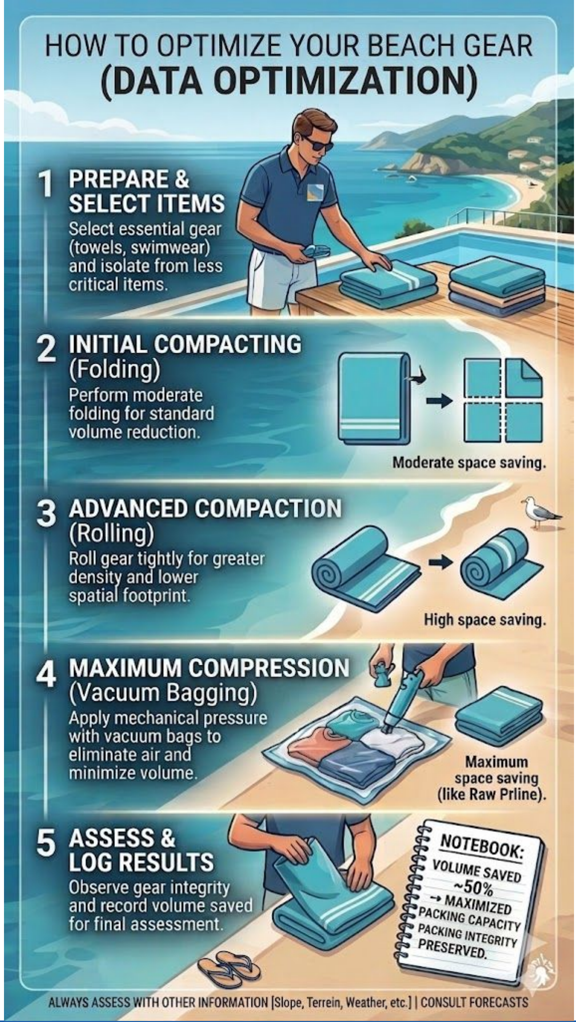
CMS-TDR-022

CMS DP -2026/004



Conclusions

- Data compression is key to increase the trigger rate in HEP.
- CMS has addressed this challenge in two ways.
- **Data scouting** has been revolutionized since Run-1 getting a better and complete reconstruction and running at higher rate
  - Scouting compression is very strong (~ x100)
    but it contains only high-level objects.
- Since 2023, CMS uses **Raw Prime** in HI runs.
  - About -50% in event size → x2 in trigger rate.
    by replacing the strip raw data with low-level reconstructed objects (clusters)
  - An improved version (**Raw Second**) has been deployed online for testing
    - achieved a further -11% in event compression.
- As HL-LHC is approaching, the data compression challenge is becoming more and more urgent !



Thank you!

This study was funded by the European Union - Next Generation EU, Mission 4 Component 1 CUP I53D23001070006

Backup

Barycenter and module ID: Raw Prime

- In addition to the clusters, in Raw Prime, the strip module IDs are also saved
  - together with the <u>index</u> of the first cluster belonging to each strip module

Barycenter-0: 3
Barycenter-1: 5
Barycenter-2: 18
Barycenter-3: 19

Barycenter-4: 8
Barycenter-5: 9

Barycenter-6: 12

Strip module ID: 321321

Index 1st cluster: 0

Strip module ID: 321323

Index 1st cluster: 4

Strip module ID: 321325

Index 1st cluster: 6

Barycenter(Δ) and module ID(Δ): Raw Second

- In Raw Second, the differences are saved instead of values
  - The collection of the index of the 1st cluster is removed (unnecessary)

(*) 22 is the length of the strip module and can be obtained from the strip module ID

Barycenter-0(Δ): 3 (=3-0)
Barycenter-1(Δ): 2 (=5-3)
Barycenter-2(Δ): 13 (=18-5)
Barycenter-3(Δ): 1 (=19-18)

Strip module ID(Δ): 321321 (=321321 - 0)

~~Index 1st cluster: 0~~

Barycenter-4(Δ): 11 (=8+22(*)-19)
Barycenter-5(Δ): 1 (=9-8)

Strip module ID(Δ): 2 (=321323 - 321321)

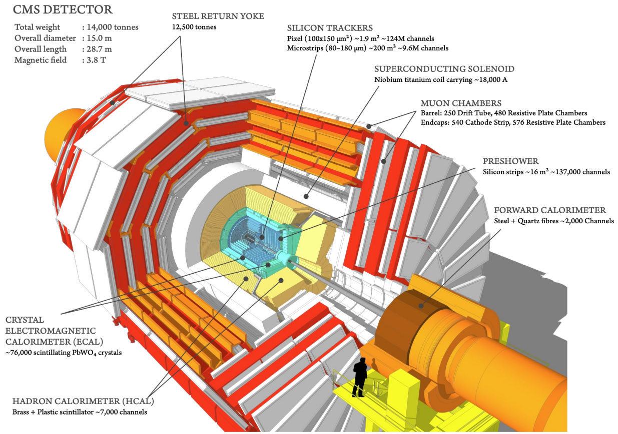
~~Index 1st cluster: 4~~

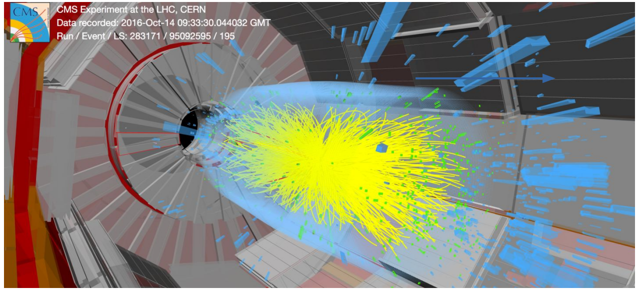
Barycenter-6(Δ): 25 (=12+22(*)-9)

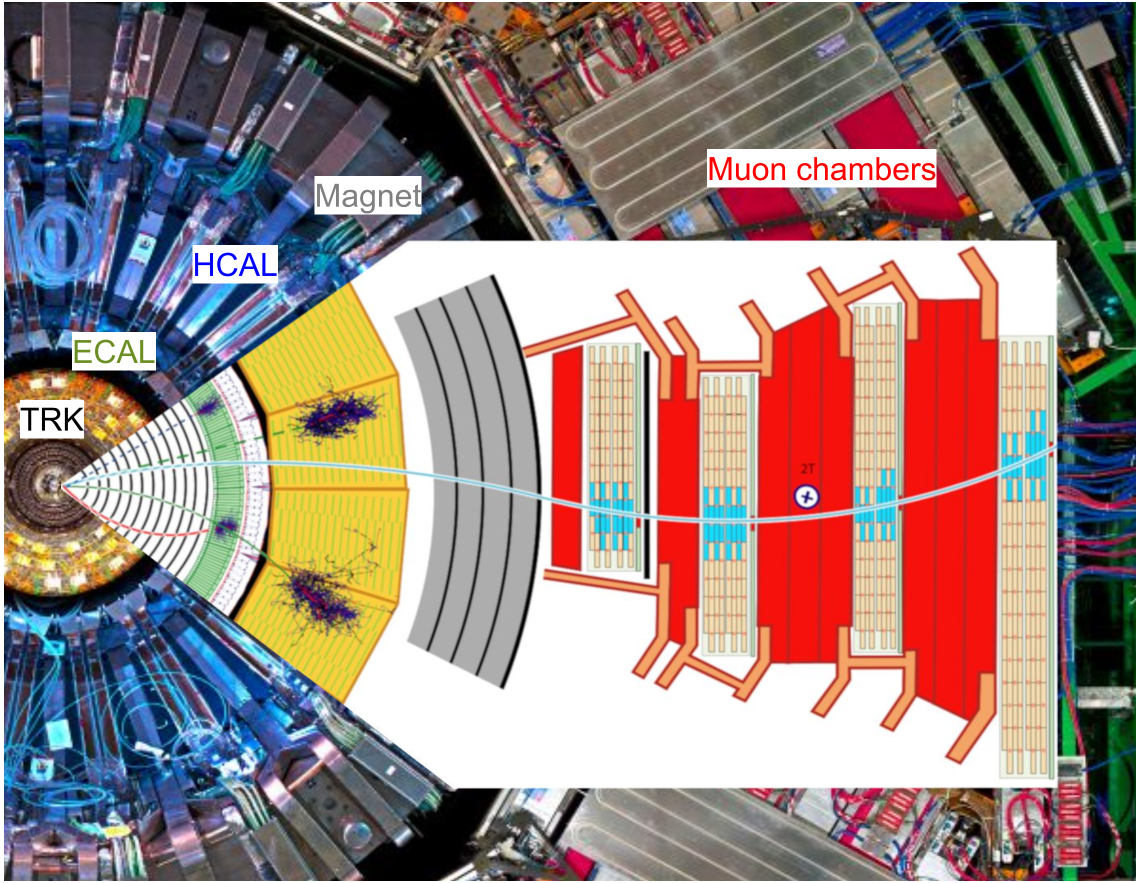
Strip module ID(Δ): 2 (=321325 - 321323)

~~Index 1st cluster: 6~~

The CMS detector







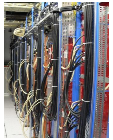
doi:10.1088/1748-0221/19/11/P11021

CMS trigger menu

Simplified readout

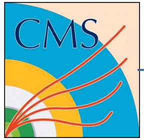
40 MHz →

L1 trigger



100 kHz (<4μs) →

Full readout



100 kHz
100 GB/s →

HLT

~2 kHz
(~300 ms)
~2 GB/s →

Data storage & reconstruction

2018 CMS trigger menu

HLT path	L1 thresholds [GeV]	HLT thresholds [GeV]	Rate [Hz]
Single muon	22	50	49
Single muon (isolated)	22	24	230
Double muon	22	37, 27	16
Double muon (isolated)	15, 7	17, 8	32
Single electron (isolated)	30	32	180
Double electron	25, 12	25, 25	16
Double electron (isolated)	22, 12	23, 12	32
Single photon	30	200	16
Single photon (isolated), barrel only ($\lvert\eta\rvert < 1.48$)	30	110	16
Double photon	25, 12	30, 18	32
Single tau	120	180	16
Double tau	32	35, 35	49
Single jet	180	500	16
Single jet with substructure	180	400	32
Multijets with b tagging	$H_T > 320$ jets > 70, 55, 40, 40	$H_T > 330$ jets > 75, 60, 45, 40	16
Total transverse momentum	360	1050	16
Missing transverse momentum	100	120	49

clean environment
low thresholds

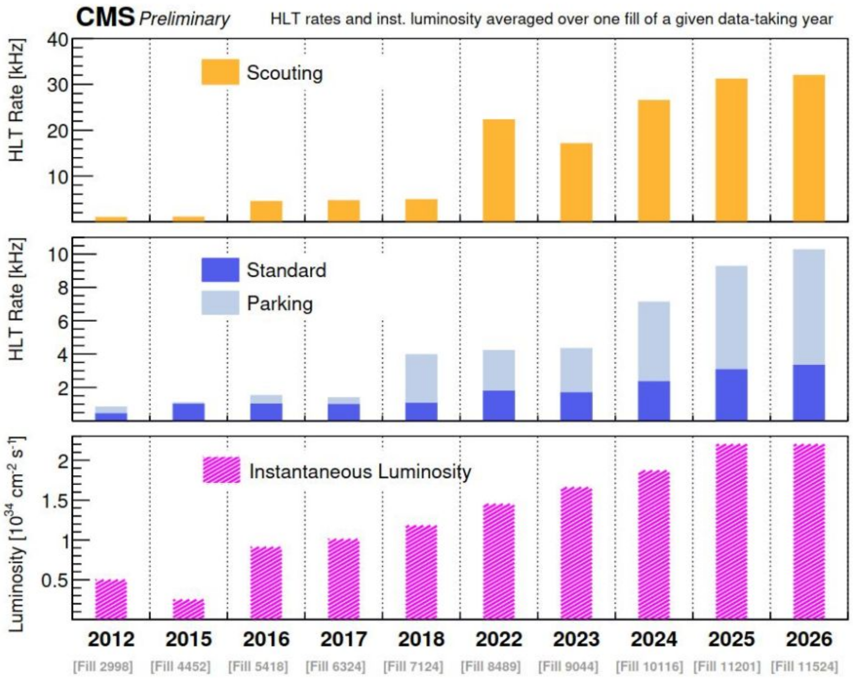
large background
high thresholds

doi:10.48550/arXiv.2403.16134

Rate evolution, scouting and parking

- The "standard" HLT rate cross-section increased with lumi since 2012,
  - **trigger cross section roughly constant**:
    - ~ 1 kHz/10^{34}cm^{-2}s^{-1}
- Large increase of **parking** rate in 2018.
  - **2018** idea: collect additional data to be processed during LHC shutdown 2 (2019-22)
  - **Run-3**: data processed promptly, if possible
- Revolution in **scouting** since 2022:
  - save **all** main physics objects reconstructed at HLT (tracks, vertices, muons, electrons, jets …) in **20%** of events processed (~20 kHz);
  - event size x100 smaller than full RAW.

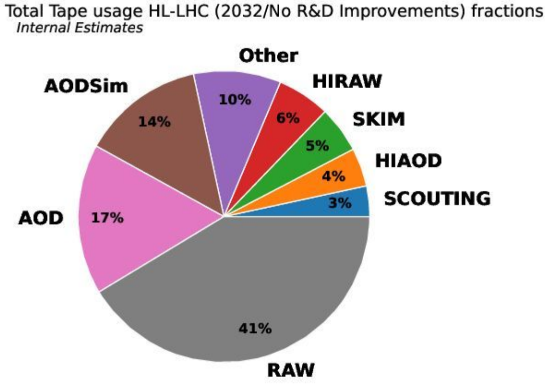
Trigger rate evolution

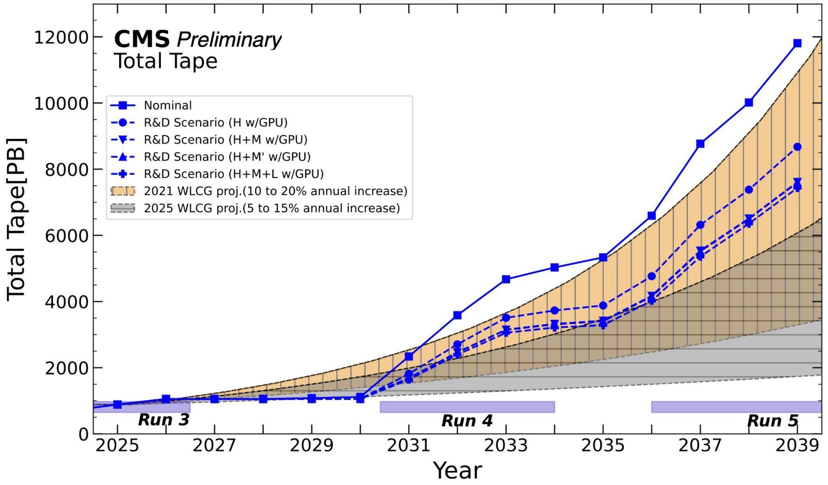


CERN-LHCC-2026-003

Tape

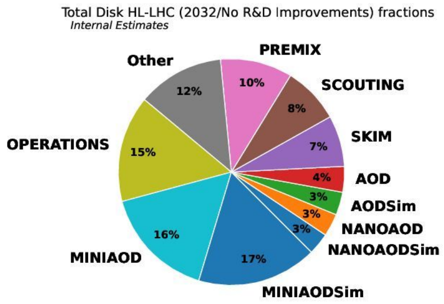
	Description	Impact (Fraction of total RECO CPU time)	Probability of Success and/or Adoption
Reco.1	Continuous Code Optimization	5% CPU speedup (per year for avg. evt.)	Success: High
Reco.2	Vertex algorithms: code vectorization GPU porting	Takes 6% of total time/evt. 80% speedup on CPU (>90% of CPU time offloaded to GPUs)	Adoption: High
Reco.3	Pixel-track seeding: GPU-targeted algoritms	Takes 12% of total time/evt. 60% speedup on CPU (>90% of CPU time offloaded to GPUs)	Adoption: High
Reco.4a	Tracking: mkFit algorithm + ML	Takes 10% of total time/evt. 65% speedup on CPU	Success: High Adoption: Medium
Reco.4b	Tracking: mkFit/LST combination + ML	50% mkFit CPU time done by LST (>90% of LST offloaded to GPUs)	Success: High Adoption: High
Reco.5a	HGCAL: TICL framework on GPUs	Takes 2% of total time/evt. (>90% time/evt. offloaded to GPUs)	Success: High (focusing on physics gains)
Reco.5b	Calorimeters: TICL for the barrels on GPUs	Takes 1.7% of total time/evt. (>90% time of calorimeter barrels offloaded to GPUs)	Success: High
Reco.6	Electrons: GPU-targeted seeding algorithms	Takes 3.5% time/evt. (>90% time offloaded to GPUs)	Success: Medium Adoption: Medium
Reco.7	particle flow: GPU-targeted reconstruction with ML	Takes 10% total time/evt. (>90% time offloaded to GPUs)	Success: Medium
Reco.8	HGCAL: ML reconstruction	Takes 3% of total time/evt. (>90% time offloaded to GPUs)	Success: Medium
Reco.9a	Tracking: Only LST and other GPU-based algorithms	Takes 15% of total time/evt plus an additional 5% for other tracking steps	Success: Low
Reco.9b	MTD: GPU-based + ML	Takes 3% of total time/evt	Success: Low
Reco.9c	Muon: GPU-based	Takes 5% of total time/evt	Success: Low
Reco.9d	Other: GPU-based algorithms	Up to 2% of total time/evt combing several small (in terms of CPU time) offloaded to GPU	Success: Low
DataF.1	RAW event size	Reduction with respect to Table 1 by 30%*	Success: Medium
DataF.2	Derived data event size	Reduction with respect to Table 1 by 15%*	Success: High
DataF.3	AOD dataset	reduction by 50%	Success: High

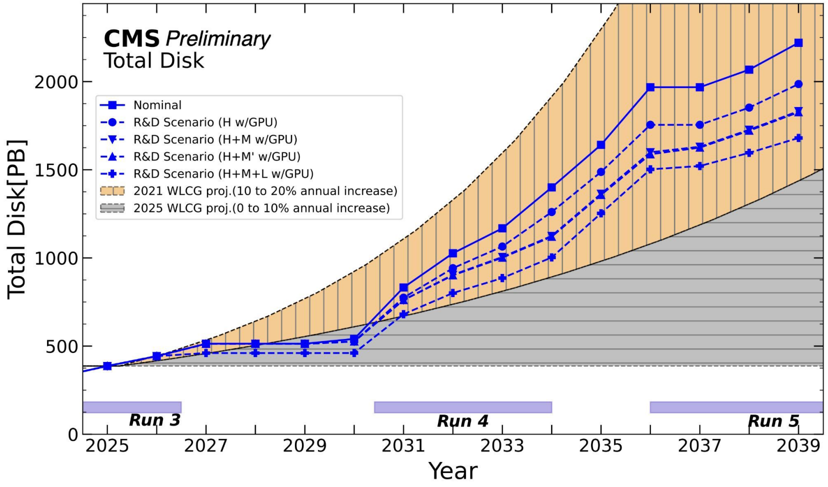




Disk

	Description	Impact (Fraction of total RECO CPU time)	Probability of Success and/or Adoption
Reco.1	Continuous Code Optimization	5% CPU speedup (per year for avg. evt.)	Success: High
Reco.2	Vertex algorithms: code vectorization GPU porting	Takes 6% of total time/evt. 80% speedup on CPU (>90% of CPU time offloaded to GPUs)	Adoption: High
Reco.3	Pixel-track seeding: GPU-targeted algoritms	Takes 12% of total time/evt. 60% speedup on CPU (>90% of CPU time offloaded to GPUs)	Adoption: High
Reco.4a	Tracking: mkFit algorithm + ML	Takes 10% of total time/evt. 65% speedup on CPU	Success: High Adoption: Medium
Reco.4b	Tracking: mkFit/LST combination + ML	50% mkFit CPU time done by LST (>90% of LST offloaded to GPUs)	Success: High Adoption: High
Reco.5a	HGCAL: TICL framework on GPUs	Takes 2% of total time/evt. (>90% time/evt. offloaded to GPUs)	Success: High (focusing on physics gains)
Reco.5b	Calorimeters: TICL for the barrels on GPUs	Takes 1.7% of total time/evt. (>90% time of calorimeter barrels offloaded to GPUs)	Success: High
Reco.6	Electrons: GPU-targeted seeding algorithms	Takes 3.5% time/evt. (>90% time offloaded to GPUs)	Success: Medium Adoption: Medium
Reco.7	particle flow: GPU-targeted recon-struction with ML	Takes 10% total time/evt. (>90% time offloaded to GPUs)	Success: Medium
Reco.8	HGCAL: ML reconstruction	Takes 3% of total time/evt. (>90% time offloaded to GPUs)	Success: Medium
Reco.9a	Tracking: Only LST and other GPU-based algorithms	Takes 15% of total time/evt plus an additional 5% for other tracking steps	Success: Low
Reco.9b	MTD: GPU-based + ML	Takes 3% of total time/evt	Success: Low
Reco.9c	Muon: GPU-based	Takes 5% of total time/evt	Success: Low
Reco.9d	Other: GPU-based algorithms	Up to 2% of total time/evt combining several small (in terms of CPU time) offloaded to GPU	Success: Low
DataF.1	RAW event size	Reduction with respect to Table 1 by 30%*	Success: Medium
DataF.2	Derived data event size	Reduction with respect to Table 1 by 15%*	Success: High
DataF.3	AOD dataset	reduction by 50%	Success: High





CPU

	Description	Impact (Fraction of total RECO CPU time)	Probability of Success and/or Adoption
Reco.1	Continuous Code Optimization	5% CPU speedup (per year for avg. evt.)	Success: High
Reco.2	Vertex algorithms: code vectorization GPU porting	Takes 6% of total time/evt. 80% speedup on CPU (>90% of CPU time offloaded to GPUs)	Adoption: High
Reco.3	Pixel-track seeding: GPU-targeted algoritms	Takes 12% of total time/evt. 60% speedup on CPU (>90% of CPU time offloaded to GPUs)	Adoption: High
Reco.4a	Tracking: mkFit algorithm + ML	Takes 10% of total time/evt. 65% speedup on CPU	Success: High Adoption: Medium
Reco.4b	Tracking: mkFit/LST combination + ML	50% mkFit CPU time done by LST (>90% of LST offloaded to GPUs)	Success: High Adoption: High
Reco.5a	HGCAL: TICL framework on GPUs	Takes 2% of total time/evt. (>90% time/evt. offloaded to GPUs)	Success: High (focusing on physics gains)
Reco.5b	Calorimeters: TICL for the barrels on GPUs	Takes 1.7% of total time/evt. (>90% time of calorimeter barrels offloaded to GPUs)	Success: High
Reco.6	Electrons: GPU-targeted seeding algorithms	Takes 3.5% time/evt. (>90% time offloaded to GPUs)	Success: Medium Adoption: Medium
Reco.7	particle flow: GPU-targeted reconstruction with ML	Takes 10% total time/evt. (>90% time offloaded to GPUs)	Success: Medium
Reco.8	HGCAL: ML reconstruction	Takes 3% of total time/evt. (>90% time offloaded to GPUs)	Success: Medium
Reco.9a	Tracking: Only LST and other GPU-based algorithms	Takes 15% of total time/evt plus an additional 5% for other tracking steps	Success: Low
Reco.9b	MTD: GPU-based + ML	Takes 3% of total time/evt	Success: Low
Reco.9c	Muon: GPU-based	Takes 5% of total time/evt	Success: Low
Reco.9d	Other: GPU-based algorithms	Up to 2% of total time/evt combing several small (in terms of CPU time) offloaded to GPU	Success: Low
DataF.1	RAW event size	Reduction with respect to Table 1 by 30%*	Success: Medium
DataF.2	Derived data event size	Reduction with respect to Table 1 by 15%*	Success: High
DataF.3	AOD dataset	reduction by 50%	Success: High