Extreme Computing

GGI 2025
Philip Chang
University of Florida

Disclaimer:

Today, assuming audience with not much background

Therefore, to some this may be rudimentary

But, it never hurts to spell it out $\Rightarrow$ *helps us discuss interesting stuff!*

Questions for each experiment

UF
Chang
Florida

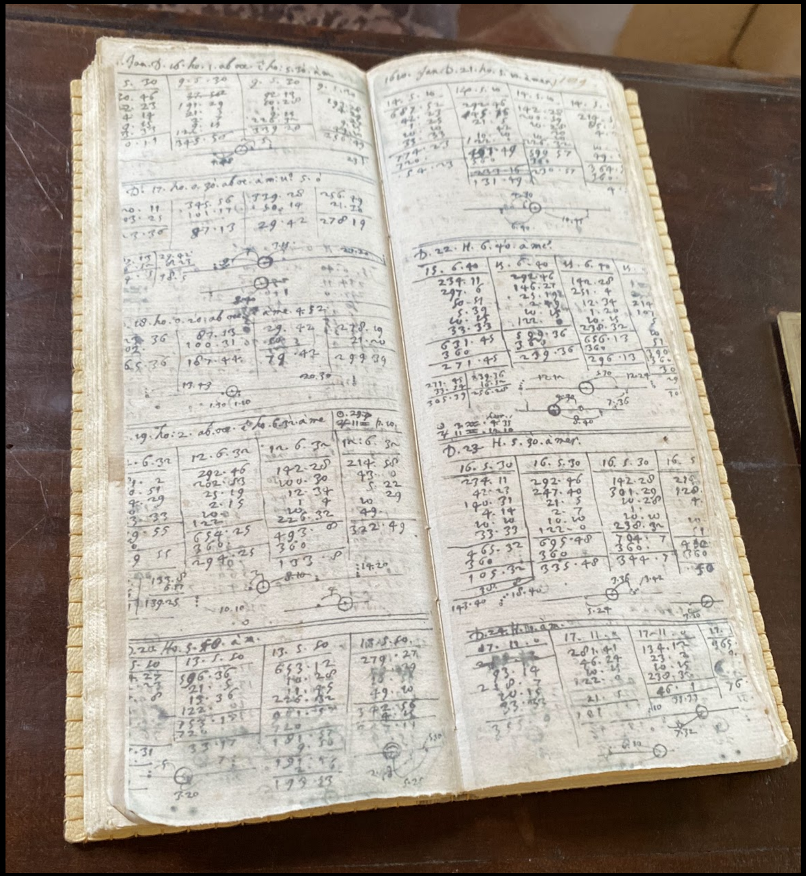
Questions for each experiment

UF
Chang
Florida

Galileo's jovilabe

CPU

Galileo's calculation notes



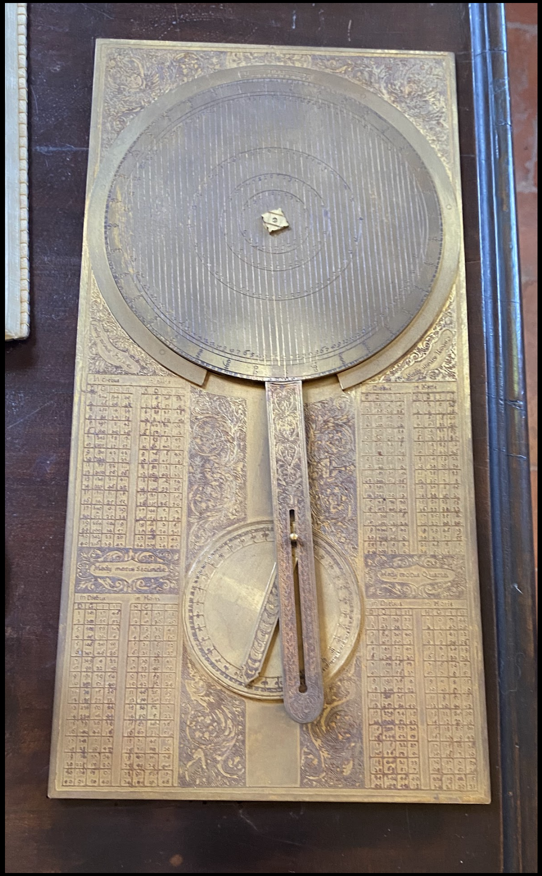
Storage

How many CPUs? How much storage?

Questions for each experiment

UF
Chang
Florida

Galileo's jovilabe



Galileo's calculation notes

CPU

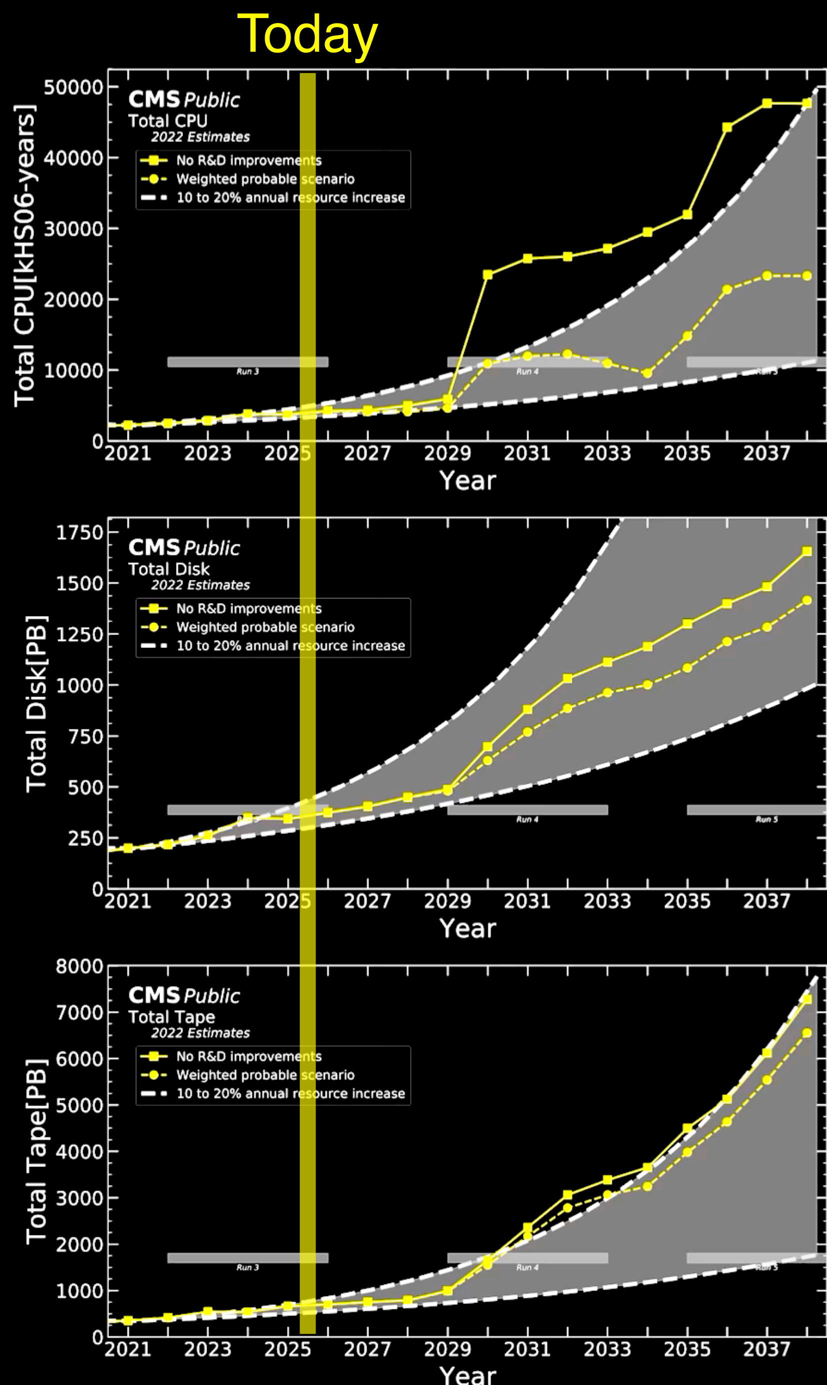
Storage

How many CPUs? How much storage?

Today even more complicated: GPUs? FPGAs? NVMes?

Example answers to the questions

UF
Chang
Florida



CPU

← Need **50M** “HS06-years”

← Plan to reduce to **22M**

(current ~3.5M)

HDD

Generally “OK”

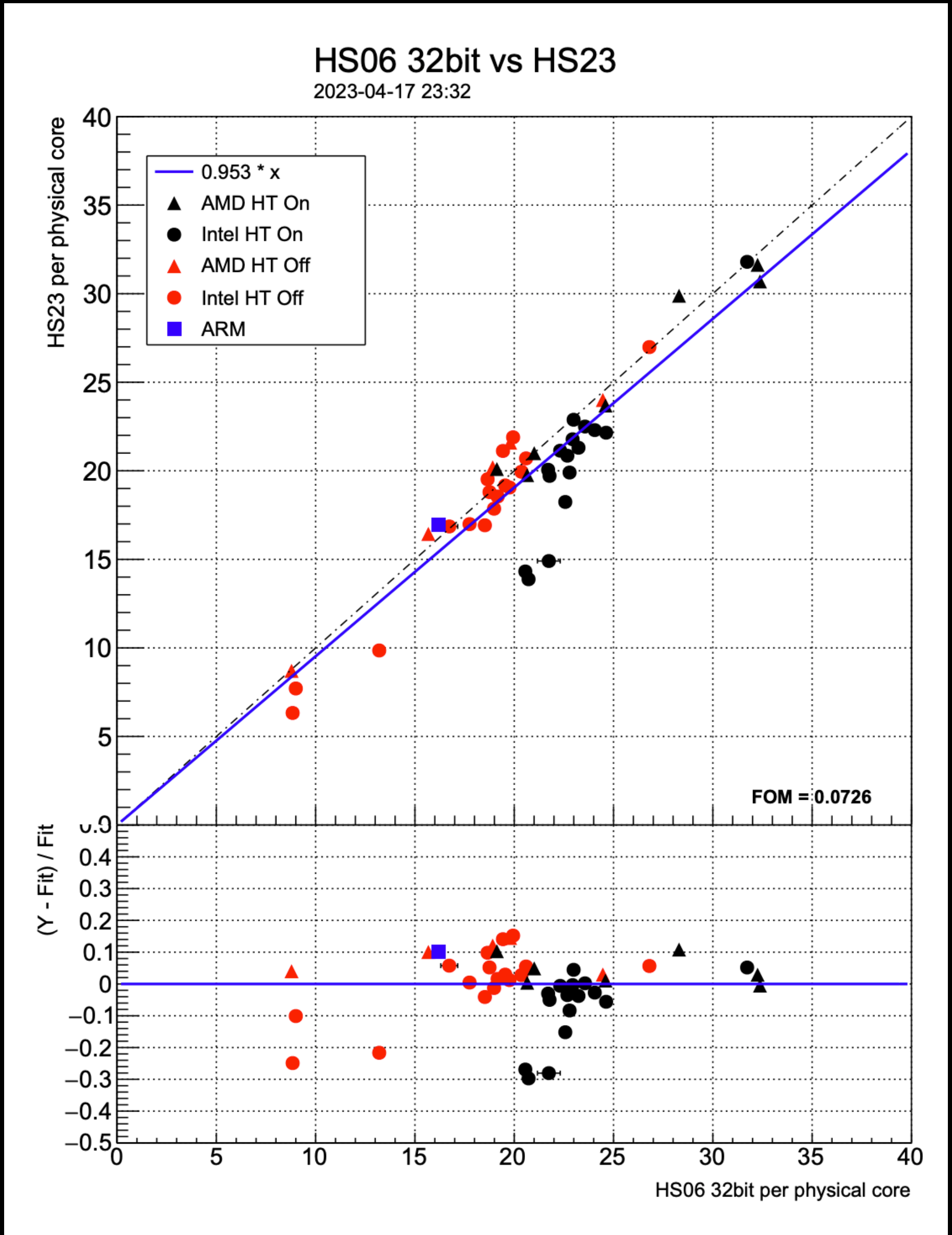
TAPE

Probably “OK”

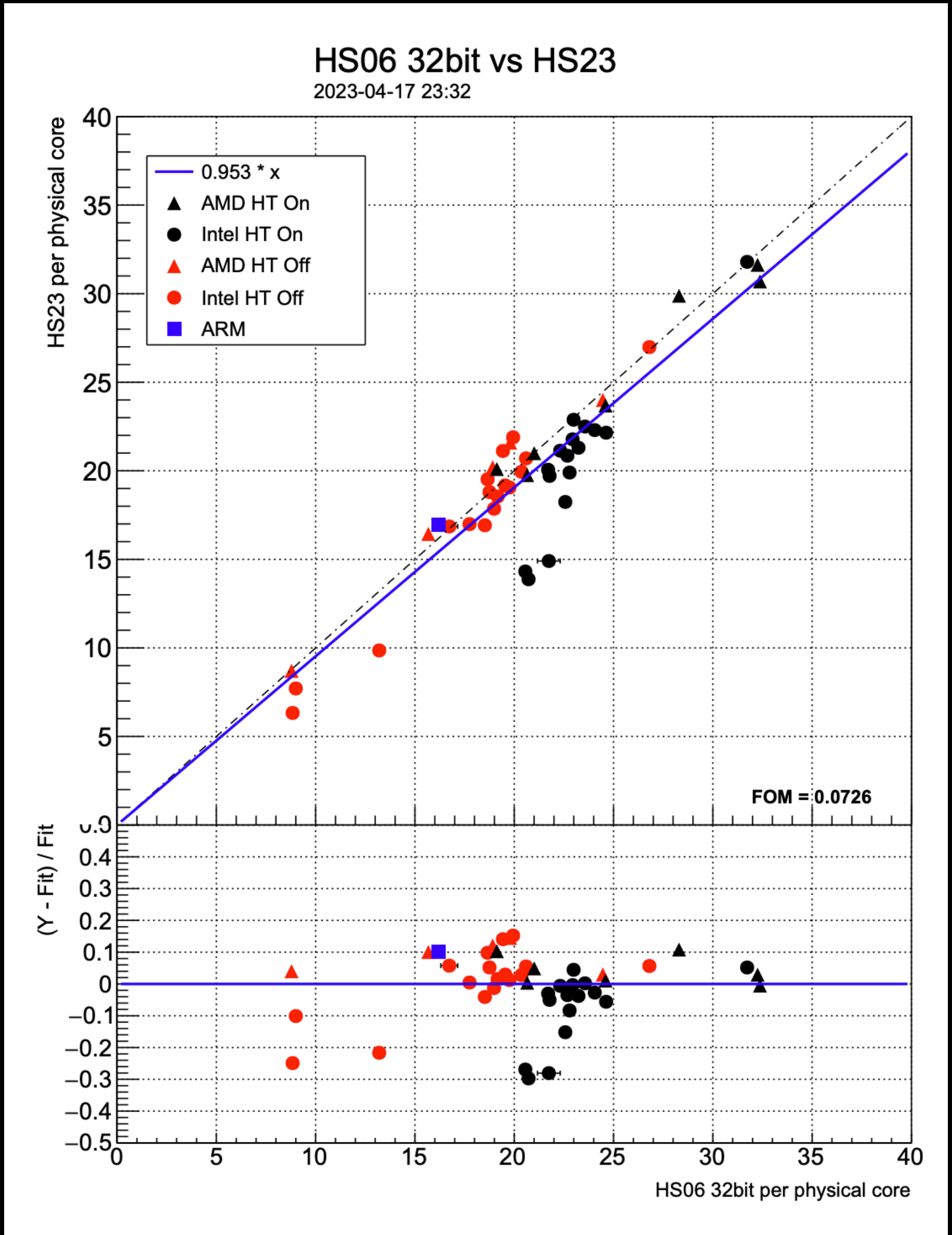
https://cds.cern.ch/record/2815292?ln=en

HS06 / HS23 / HEPScore

UF Chang Florida



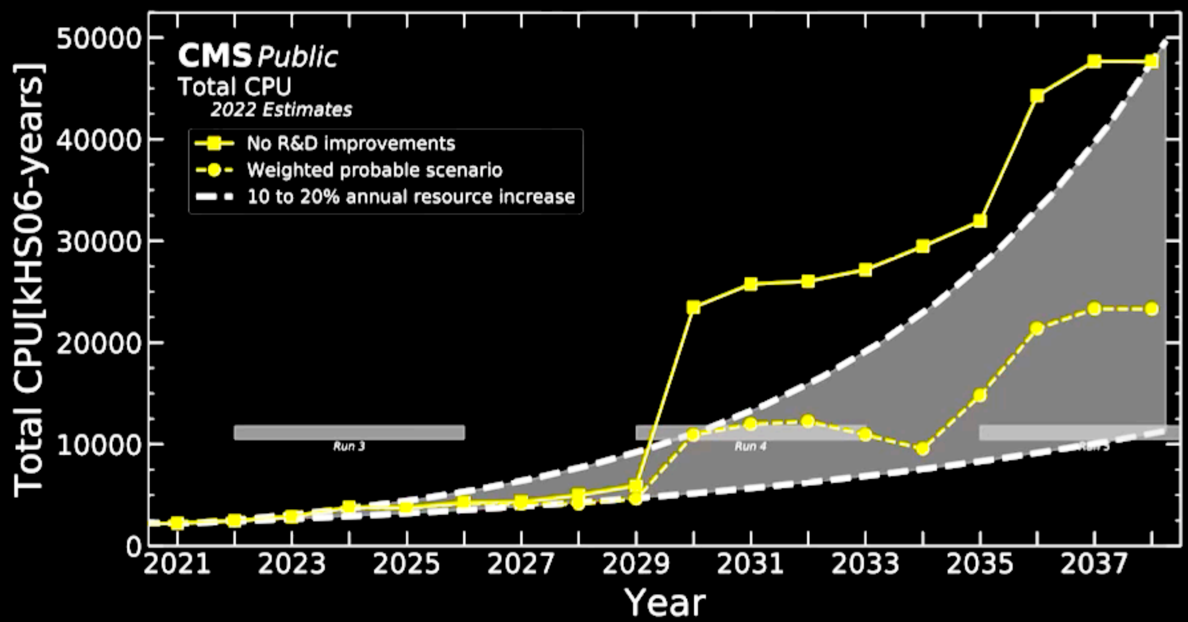
HS06 / HS23 / HEPScore



All you have to remember is that roughly 10 - 20 HS06-sec = 1 second

Conversion

UF
Chang
Florida



Therefore 50M HS06-years ~ 3M core-years

If there is only 1 CPU in the world, it will take 3 million years
if you have 3 million CPUs, it will take one year

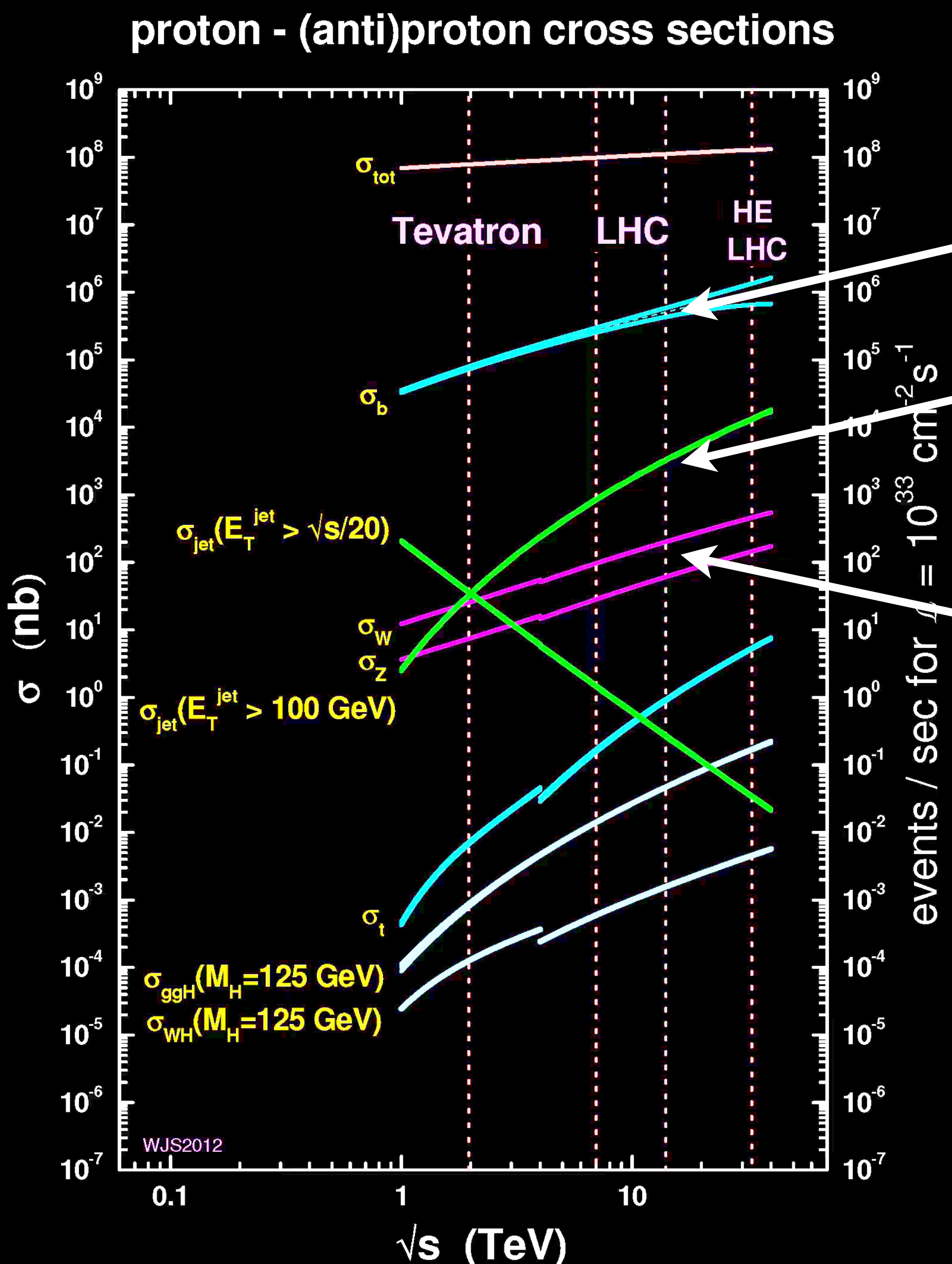
Actual model is quite complicated.

But in the following few slides I will motivate the numbers in "back-of-the-envelope" style.

More details can be found here: (for CMS example)
https://cds.cern.ch/record/2815292?ln=en

How many events produced?

UF
Chang
Florida



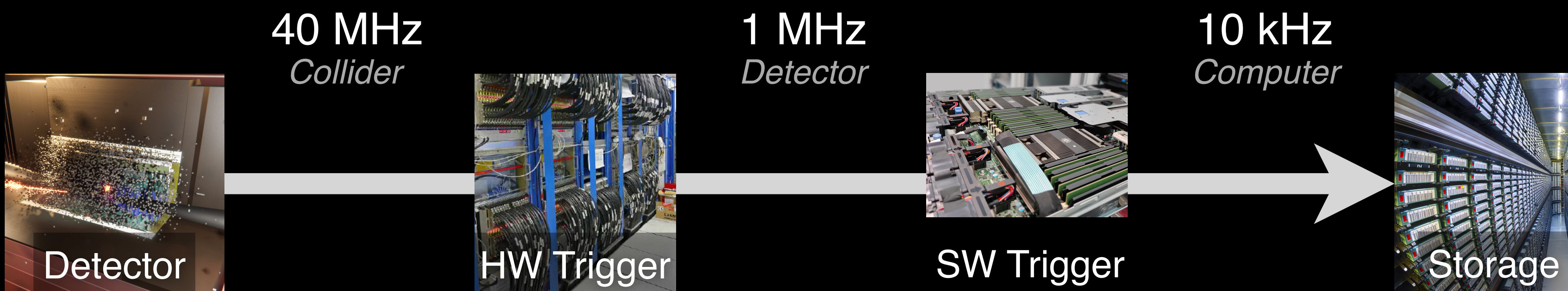
cross section tells us rates

b-jets	~O(MHz)	10T evts
$E_{T\text{-Jet}}$ > 100 GeV	~O(kHz)	10B evts
W / Z	~O(100 Hz)	1B evts

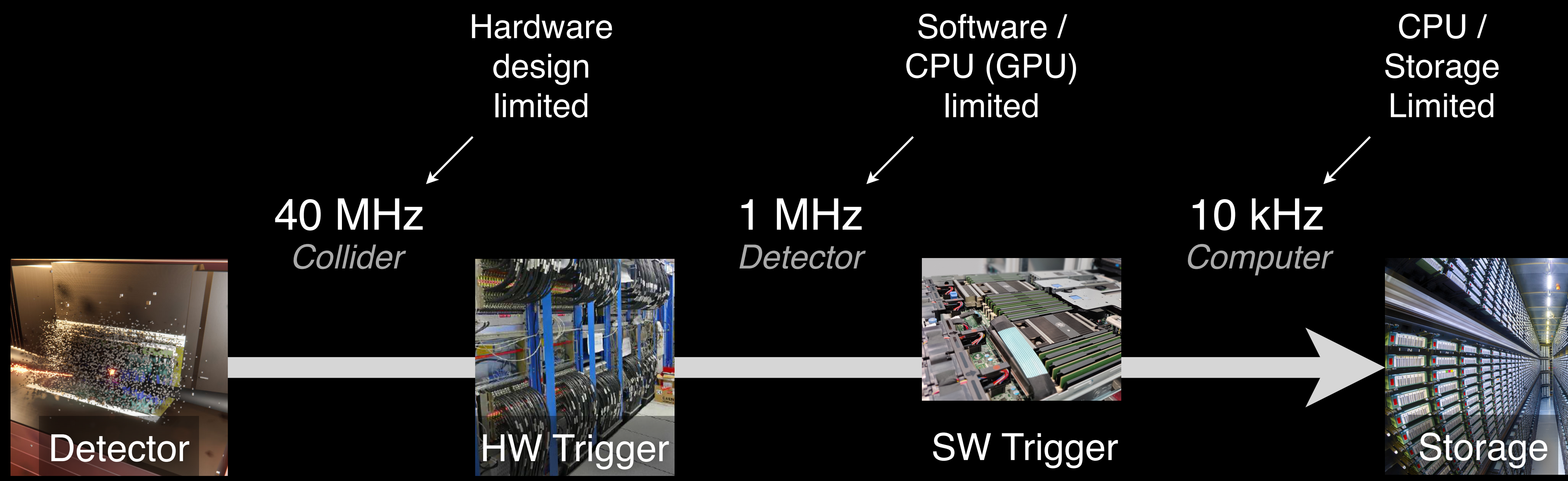
10^7 seconds / year
(~230 days 12 hours operations)

How many events can we save?

Chang
Florida

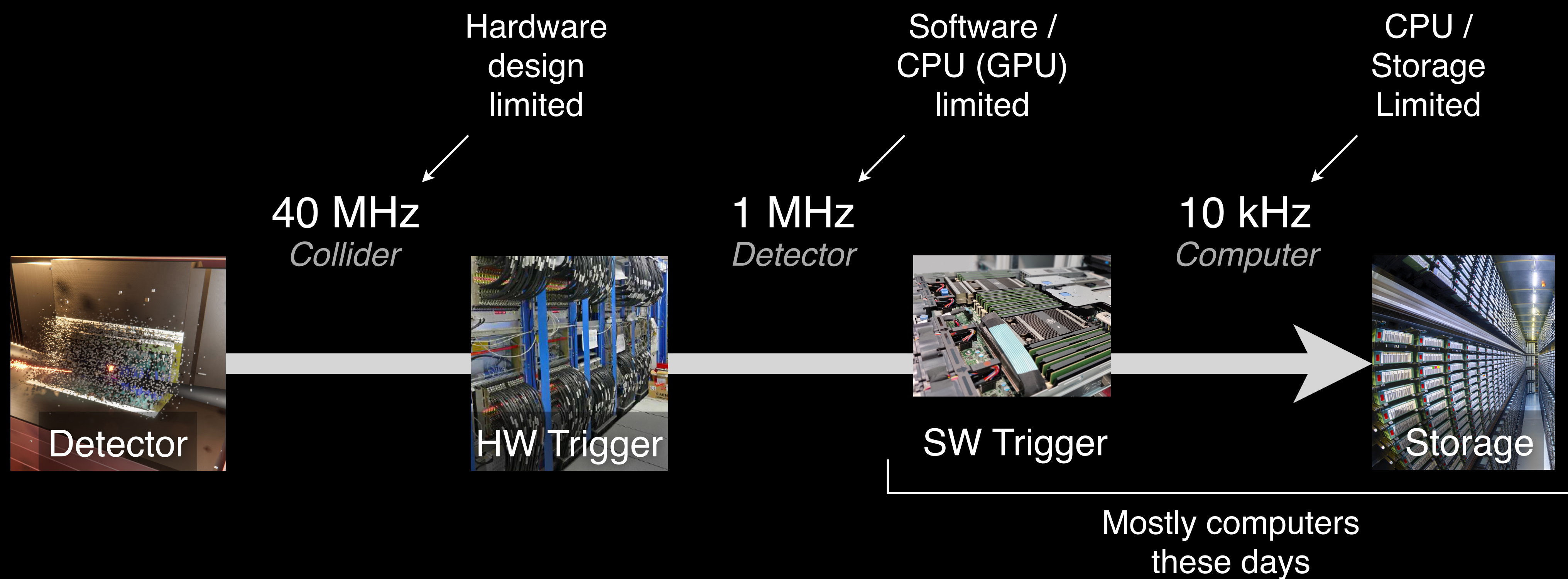


How many events can we save?



How many events can we save?

UF
Chang
Florida



How many events can we save?

UF
Chang
Florida

Hardware design limited

Software / CPU (GPU) limited

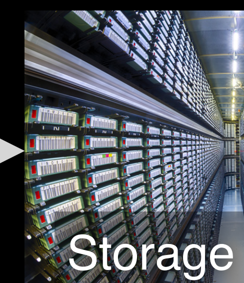
CPU / Storage Limited

40 MHz
Collider

1 MHz
Detector

10 kHz
Computer

SW Trigger



Mostly computers these days

There are efforts to make all of this computing based

LHCb Run 3 pure software trigger: J. Phys.: Conf. Ser. 878 012012

How many events can we save?

UF
Chang
Florida

Hardware design limited

Software / CPU (GPU) limited

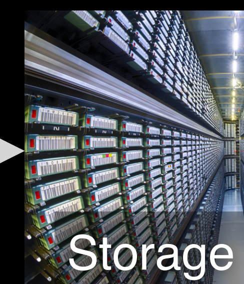
CPU / Storage Limited

40 MHz
Collider

1 MHz
Detector

10 kHz
Computer

SW Trigger



Mostly computers these days

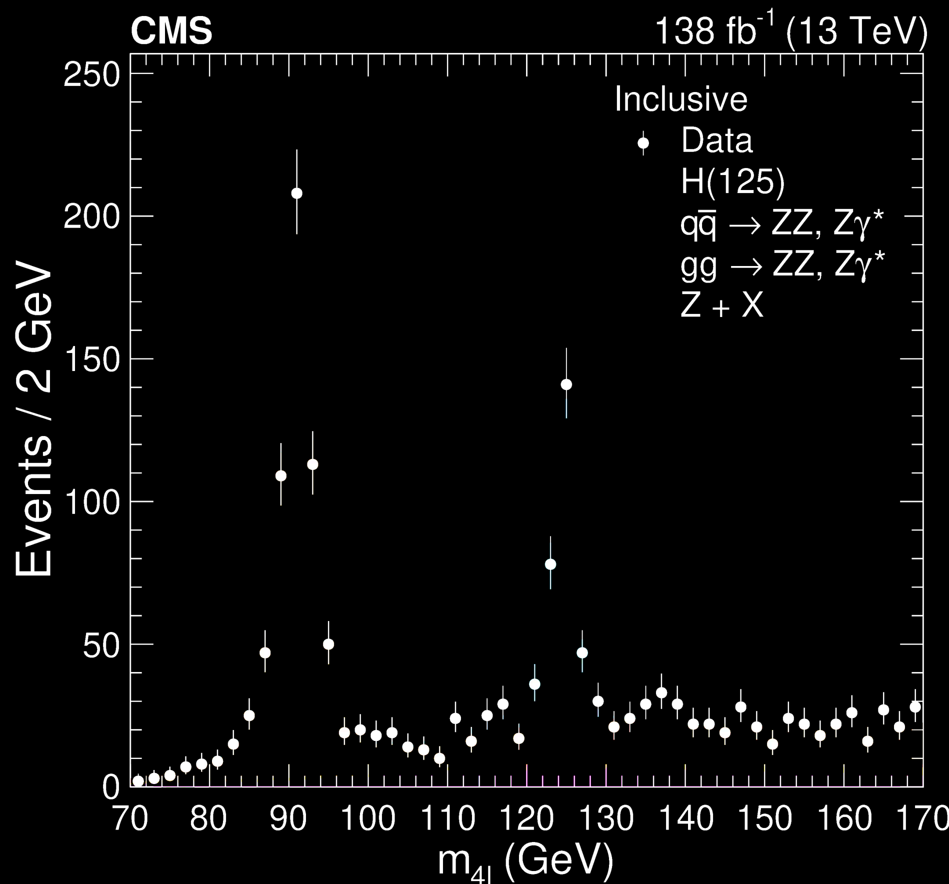
There are efforts to make all of this computing based

LHCb Run 3 pure software trigger: J. Phys.: Conf. Ser. 878 012012

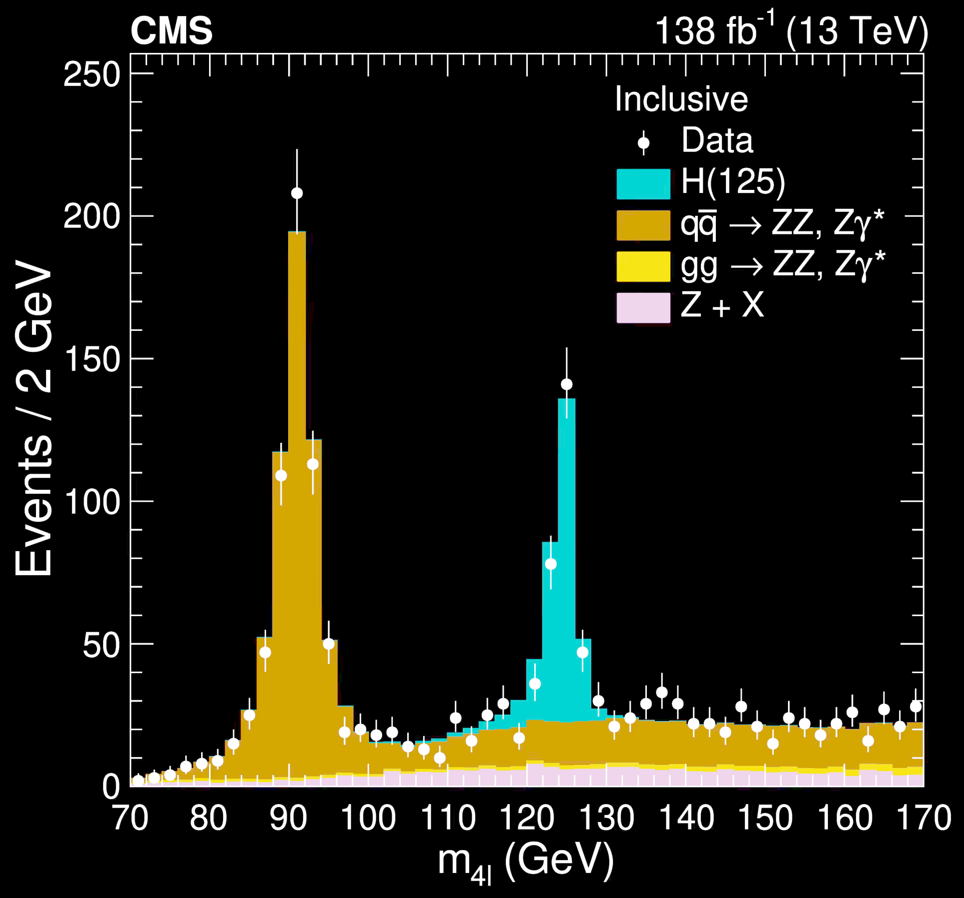
(10 kHz) × (10^7 seconds / year) = 100B events / year

How much simulated events?

UF
Chang
Florida

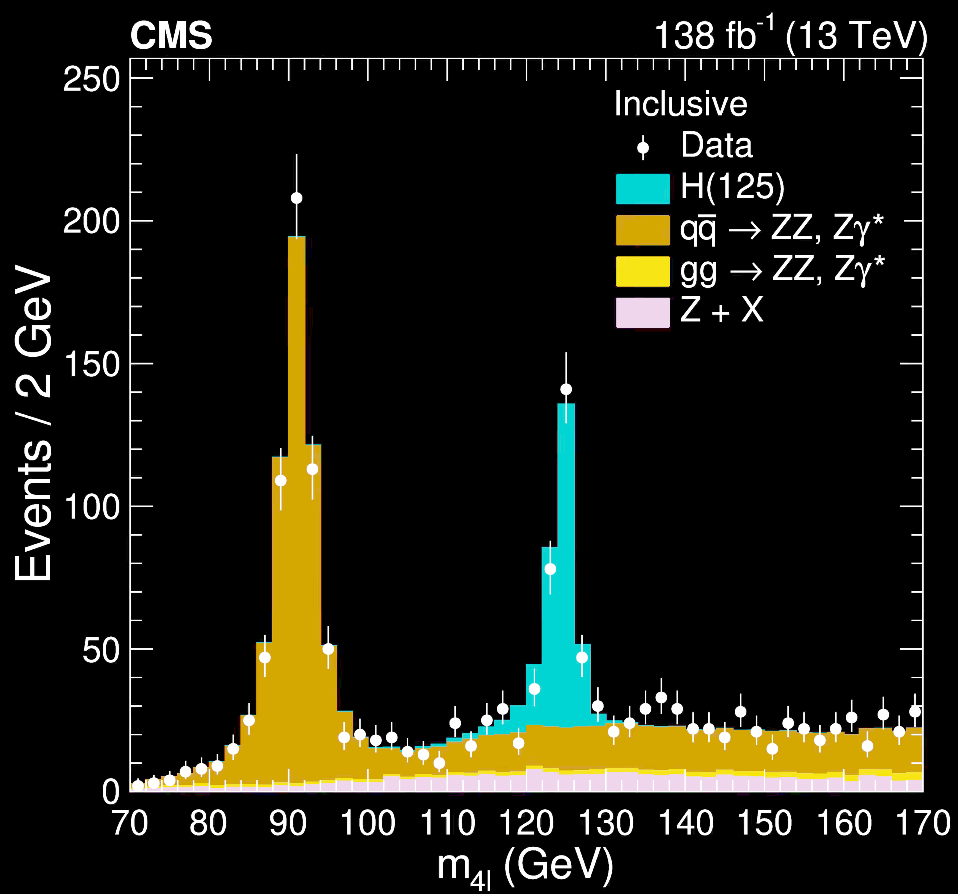


How much simulated events?



How much simulated events?

UF
Chang
Florida

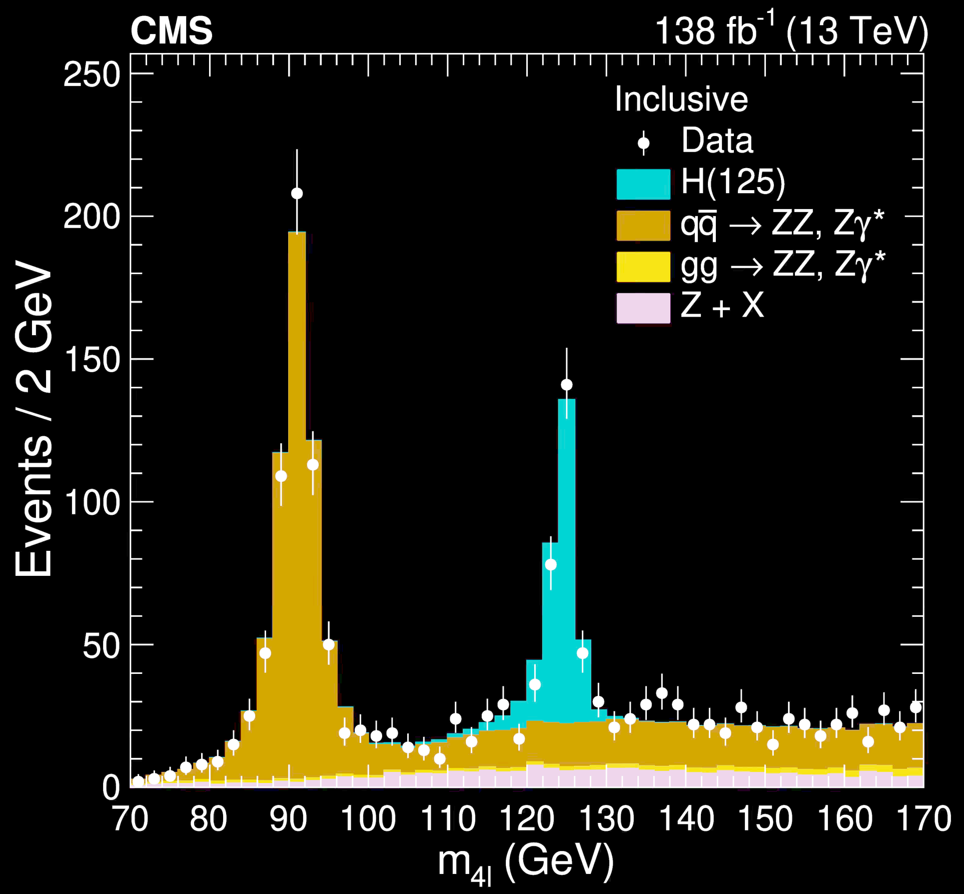


$f_{MC} \sim 1.5$ for LHC

per data event how many simulated events

(Experiment dependent. Physics goal dependent)

How much simulated events?



$f_{MC} \sim 1.5$ for LHC

per data event how many simulated events

(Experiment dependent. Physics goal dependent)

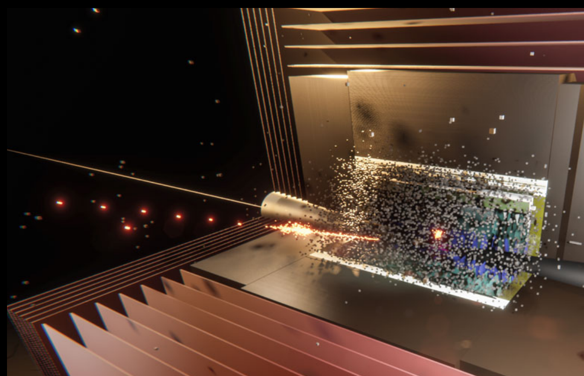
(100B events / year) × (1 + f_{MC}) = 250B events / year

How much storage?

UF
Chang
Florida

Raw data

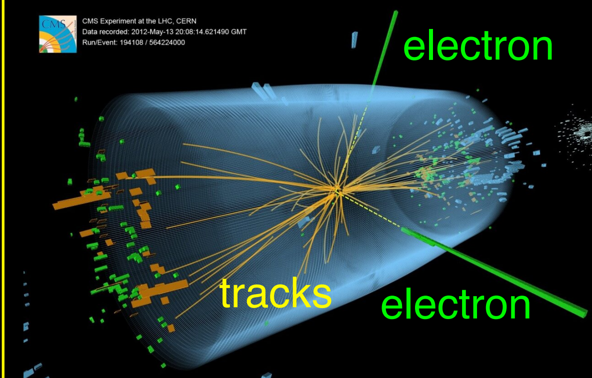
(channel readout)



6 MB

Useful Analysis data

(physics objects)

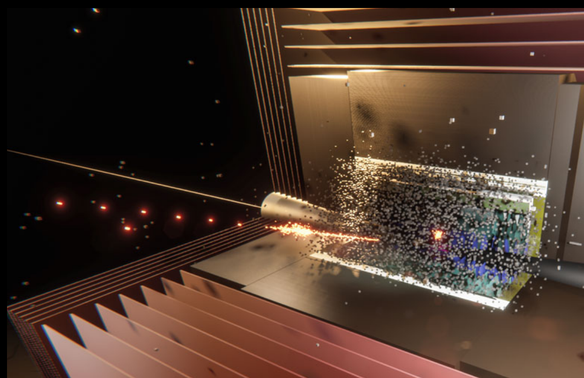


0.004 MB - 2 MB

How much storage?

Raw data

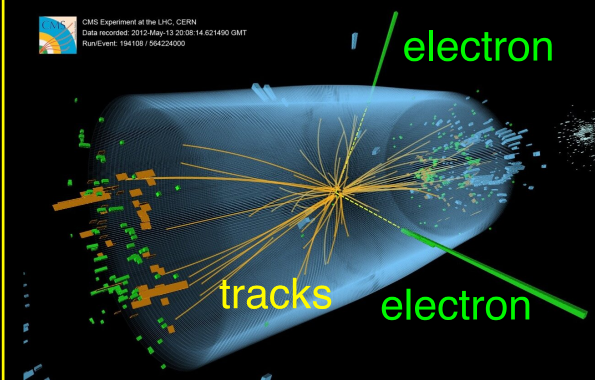
(channel readout)



6 MB

Useful Analysis data

(physics objects)



0.004 MB - 2 MB

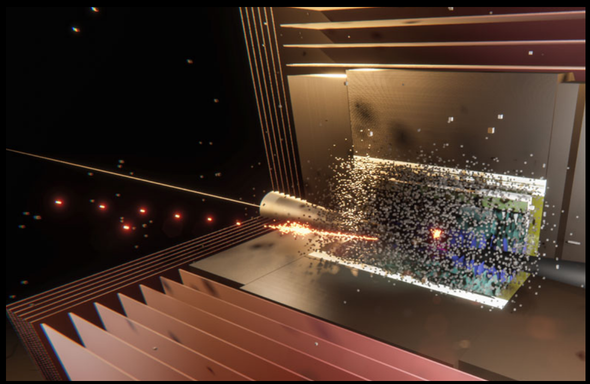
250B events × 6 MB = 1.5 Exabyte (~ $35M disk)

UF
Chang
Florida

How much storage?

Raw data

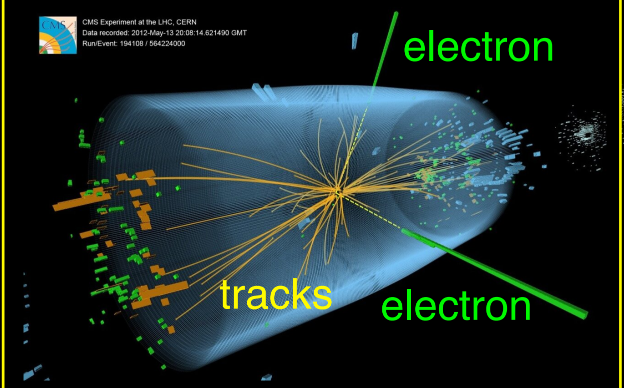
(channel readout)



6 MB

Useful Analysis data

(physics objects)



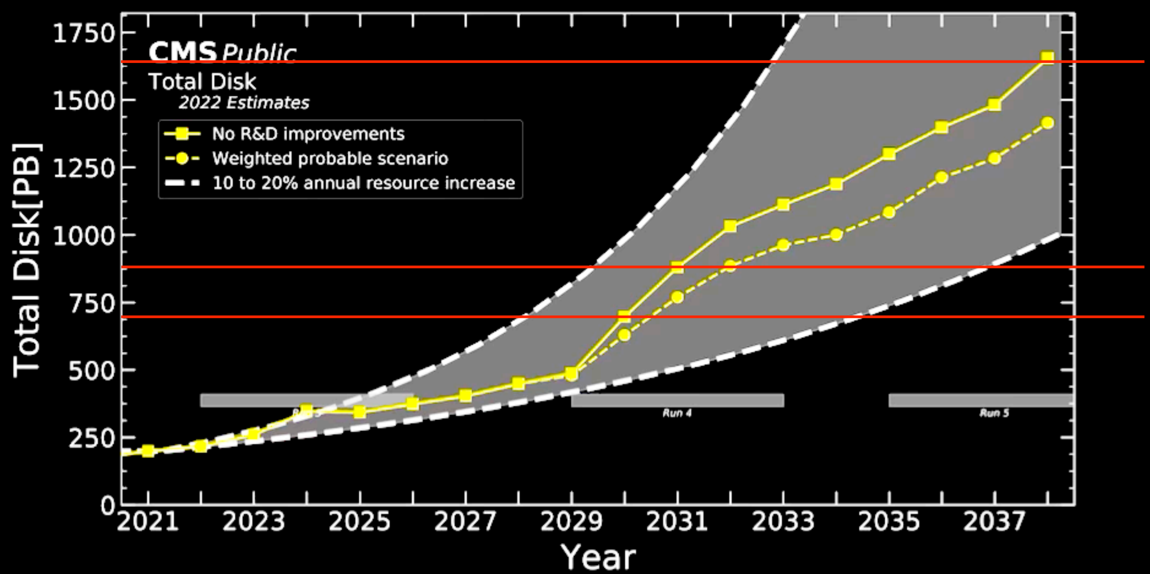
0.004 MB - 2 MB

250B events × 6 MB = 1.5 Exabyte (~ $35M disk)

Disk random access possible
(i.e. “get me so and so event”)

Tape is order of mag cheaper
Tape cannot do random access

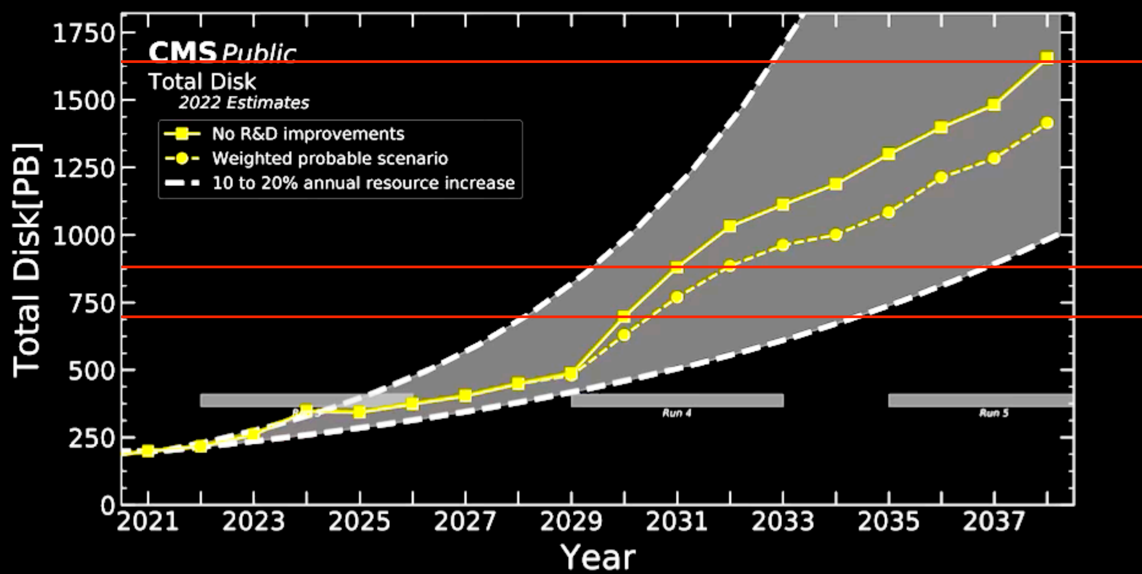
Caveat

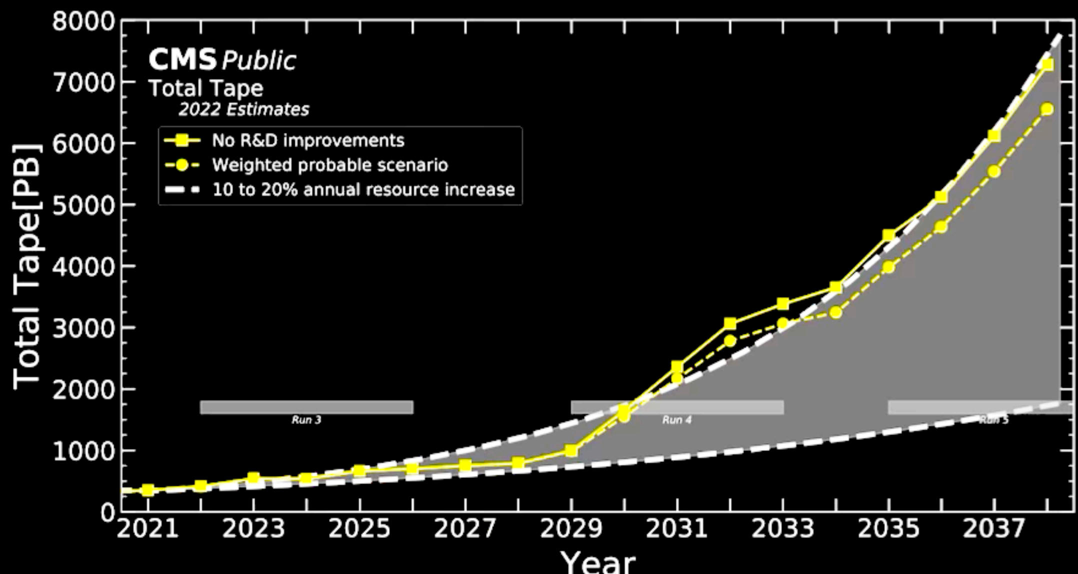


Capped ~1.5 EB

N.B. Disks only increases by a little

Caveat





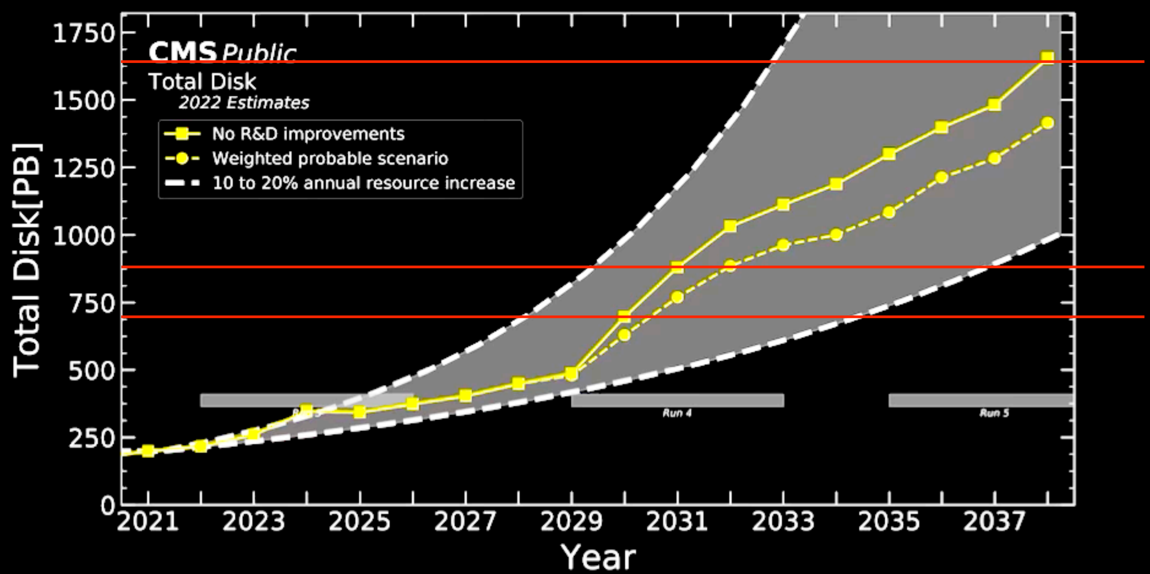
Capped ~1.5 EB

N.B. Disks only increases by a little

Tape does increase by EB

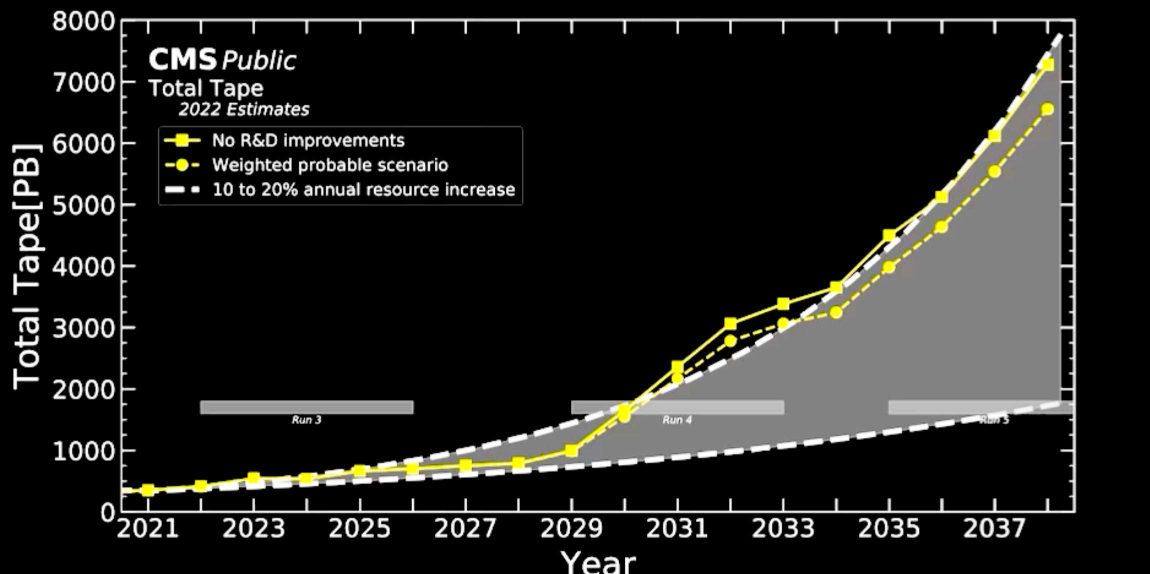
Caveat

UF
Chang
Florida



Capped ~1.5 EB

N.B. Disks only increases by a little

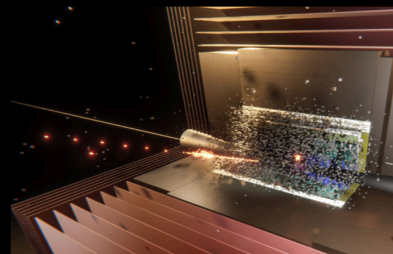


Tape does increase by EB

Raw data are moved to tape, and smaller size
Analysis format data are saved on disk

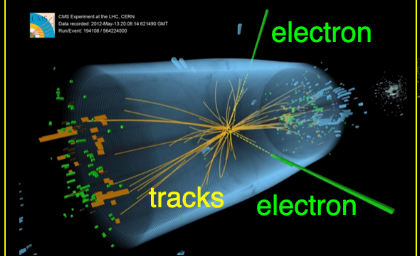
Each event takes how much CPU time?

Raw data
(channel readout)



6 MB

Useful Analysis data
(physics objects)



0.004 MB - 2 MB

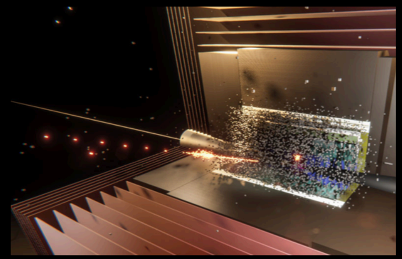
Monte Carlo Simulation

Reconstruction

Each event takes how much CPU time?

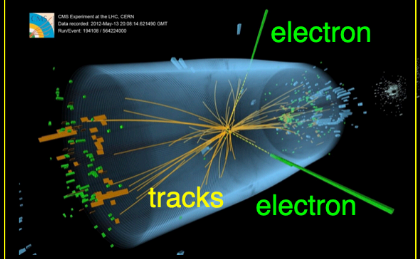
UF
Chang
Florida

Raw data
(channel readout)



6 MB

Useful Analysis data
(physics objects)



0.004 MB - 2 MB

Monte Carlo Simulation

Reconstruction

CMS	200 PU
“Simulation” (Gen + Sim)	111 sec
“Reconstruction” (Digi + PU mix + Reco)	300 sec

t = ~7 min

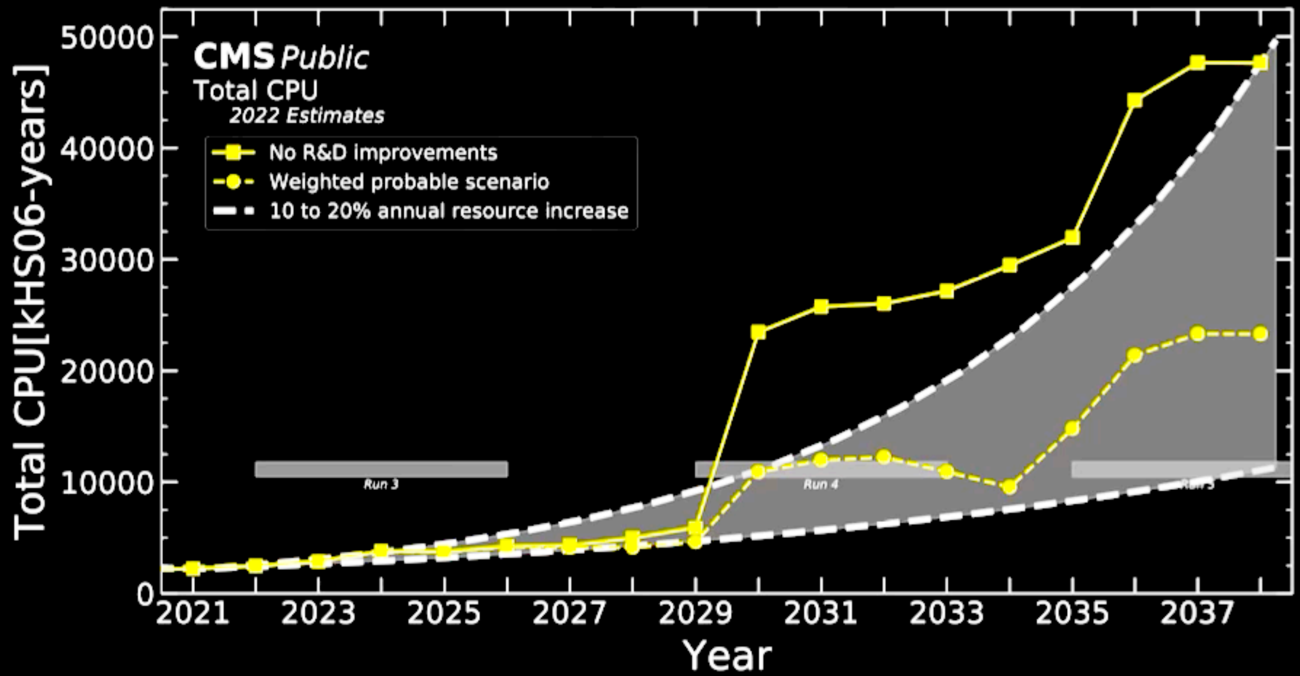
CMS Phase-2 Computing Model Document (converting 17 HS06-sec as 1 sec)

How many total CPUs?

(250B evts) × (7 core-min/evt) = 3M core-years

How many total CPUs?

(250B evts) × (7 core-min/evt) = 3M core-years



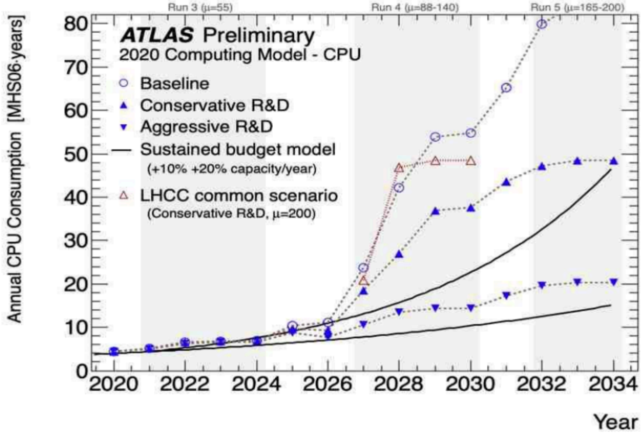
Per core ~$80

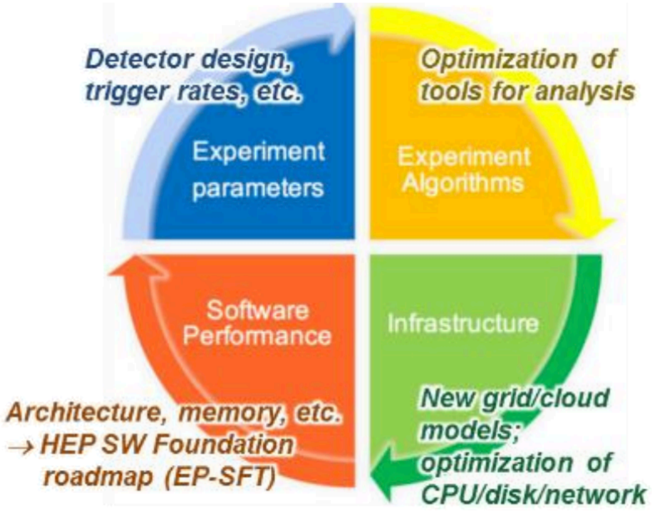
⇒ $250M

Computing in the HL-LHC Era

U.S. DEPARTMENT OF ENERGY | Office of Science

- **Simple extrapolation leads to an unsustainable place**
  - If the current software and computing approach is applied, costs can quickly exceed the entire U.S. HEP budget ("$1B problem")
- **Our goal is to match demonstrable experiment needs with a realistic funding profile — we want the science to succeed**
  - How do the software and computing models evolve?
    - much was developed beginning 15 years ago
    - they need to function 15 years from now
  - To what extent can we leverage HPC capabilities?
  - What is the optimum balance between CPU, disk, and networking?
  - R&D investments: what activities are being done or planned to address the HL-LHC software and computing challenges?
- **What is the optimum balance between people and hardware?**
  - Goal: assess computing resources and needs early enough to help inform experiments and funding agencies for successful operations during the HL-LHC era
- **For efforts towards a strategic plan, HEP Software Foundation prepared Community White Paper: https://arxiv.org/pdf/1712.06982.pdf** *(Dec. 2017)*
  - Additional documentation prepared by the LHC experiments during last few years





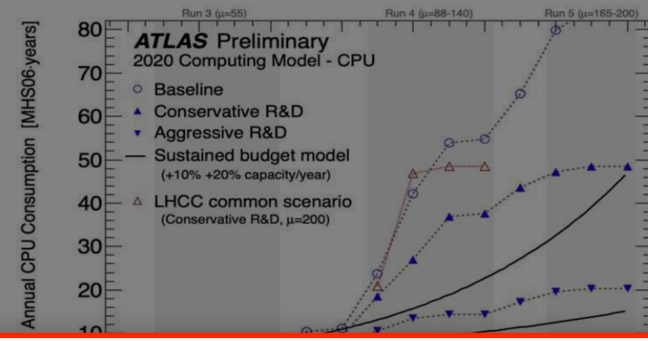
https://indico.fnal.gov/event/22303/contributions/246857/attachments/157751/206557/FY2022-DOE-PI-Meeting-Snowmass-Energy-Frontier-Program-PATWA.pdf

From DOE program manager

Computing in the HL-LHC Era

U.S. DEPARTMENT OF ENERGY | Office of Science

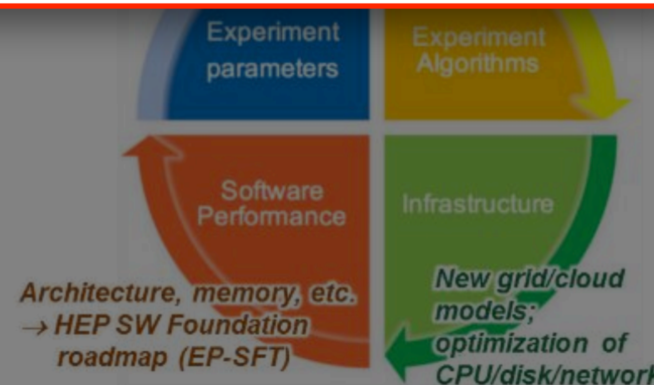
- **Simple extrapolation leads to an unsustainable place**
  - If the current software and computing approach is applied, costs can quickly exceed the entire U.S. HEP budget ("$1B problem")
- **Our goal is to match demonstrable experiment needs with a realistic funding profile — we want the science to succeed**
  - How do the software and computing models evolve?



- **Simple extrapolation leads to an unsustainable place**
  - If the current software and computing approach is applied, costs can quickly exceed the entire U.S. HEP budget ("$1B problem")

HL-LHC software and computing challenges?

- **What is the optimum balance between people and hardware?**
  - Goal: assess computing resources and needs early enough to help inform experiments and funding agencies for successful operations during the HL-LHC era
- **For efforts towards a strategic plan, HEP Software Foundation prepared Community White Paper:** *https://arxiv.org/pdf/1712.06982.pdf (Dec. 2017)*
  - Additional documentation prepared by the LHC experiments during last few years



https://indico.fnal.gov/event/22303/contributions/246857/attachments/157751/206557/FY2022-DOE-PI-Meeting-Snowmass-Energy-Frontier-Program-PATWA.pdf

Estimating per year

UF
Chang
Florida

$N_{evt} = (1 + f_{MC})\,(\text{"Rate"} \times 10^7\ \text{sec})$

250B 1.5 10kHz

Estimating per year

UF
Chang
Florida

$$N_{evt} = (1 + f_{MC})\ (\text{"Rate"} \times 10^7\ \text{sec})$$

250B — 1.5 — 10kHz

$$D_{size} = N_{evt} \times \text{"Raw data size"}$$

1.5 EB — 250B — 6MB / evt

Estimating per year

UF Chang Florida

$$N_{evt} = (1 + f_{MC})\,(\text{"Rate"} \times 10^7\ \text{sec})$$

250B — 1.5 — 10kHz

$$D_{size} = N_{evt} \times \text{"Raw data size"}$$

1.5 EB — 250B — 6MB / evt

$$C_{core} = N_{evt} \times \text{"Processing time"}$$

3.3M core-year — 250B — 7 min / evt

Estimating per year

$$N_{evt} = (1 + f_{MC}) \, (\text{"Rate"} \times 10^7 \text{ sec})$$

N_{evt}: 250B; f_{MC}: 1.5; "Rate": 10kHz

$$D_{size} = N_{evt} \times \text{"Raw data size"}$$

D_{size}: 1.5 EB; N_{evt}: 250B; "Raw data size": 6MB / evt

$$C_{core} = N_{evt} \times \text{"Processing time"}$$

C_{core}: 3.3M core-year; N_{evt}: 250B; "Processing time": 7 min / evt

In the future… GPU, FPGA…

CMS current resources

Currently we pledge to deliver 4.1M HS06 (~ 250k cores)

CMS current resources

UF
Chang
Florida

Currently we pledge to deliver 4.1M HS06 (~ 250k cores)

I estimate 250 FTEs supporting computing and R&D (for CMS)

(Not counting staff support activity from data center)

CMS current resources

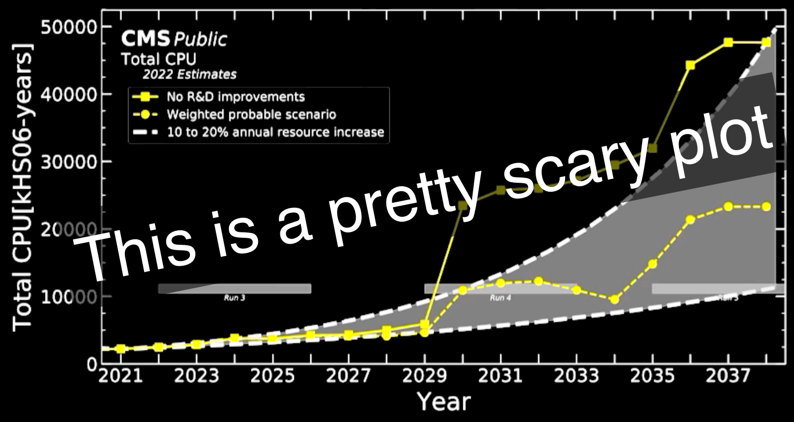
Currently we pledge to deliver 4.1M HS06 (~ 250k cores)

I estimate 250 FTEs supporting computing and R&D (for CMS)

(Not counting staff support activity from data center)

We will have to increase to 22M HS06 (or more)

⇒ What is the impact on FTE?



What about future colliders?

Spoiler: Generally OK…..

Future Colliders (per year)

UF
Chang
Florida

N.B. Not official numbers (take this with many grains of salt…)

	fMC	Rate	Time	Size	N_{evt}	D_{disk}	C_{CPU}
HL-LHC	1.5	10kHz	7 min	6 MB	250B	1.5 EB	3.3M
FCC-ee	4	200kHz	0.1 min	1 MB	10T	10 EB	2M
FCC-hh	2	10kHz	20 min	50 MB	300B	15 EB	11M
μC (10 km)	4	1kHz	20 min	50 MB	50B	5 EB	1.9M

Caveats: These are "back-of-the-envelope" numbers which is approximately correct with their CDR or supporting documents. For more detail please consult the documents.

Future Colliders (per year)

UF
Chang
Florida

N.B. Not official numbers (take this with many grains of salt…)

	fMC	Rate	Time	Size	N_{evt}	D_{disk}	C_{CPU}
HL-LHC	1.5	10kHz	7 min	6 MB	250B	1.5 EB	3.3M
FCC-ee	4	200kHz	0.1 min	1 MB	10T	10 EB	2M
FCC-hh	2	10kHz	20 min	50 MB	300B	15 EB	11M
μC (10 km)	4	1kHz	20 min	50 MB	50B	5 EB	1.9M

Caveats: These are "back-of-the-envelope" numbers which is approximately correct with their CDR or supporting documents. For more detail please consult the documents.

Future Colliders (per year)

Chang
Florida

N.B. Not official numbers (take this with many grains of salt…)

	fMC	Rate	Time	Size	N_{evt}	D_{disk}	C_{CPU}
HL-LHC	1.5	10kHz	7 min	6 MB	250B	1.5 EB	3.3M
FCC-ee	4	200kHz	0.1 min	1 MB	10T	10 EB	2M
FCC-hh	2	10kHz	20 min	50 MB	300B	15 EB	11M
μC (10 km)	4	1kHz	20 min	50 MB	50B	5 EB	1.9M

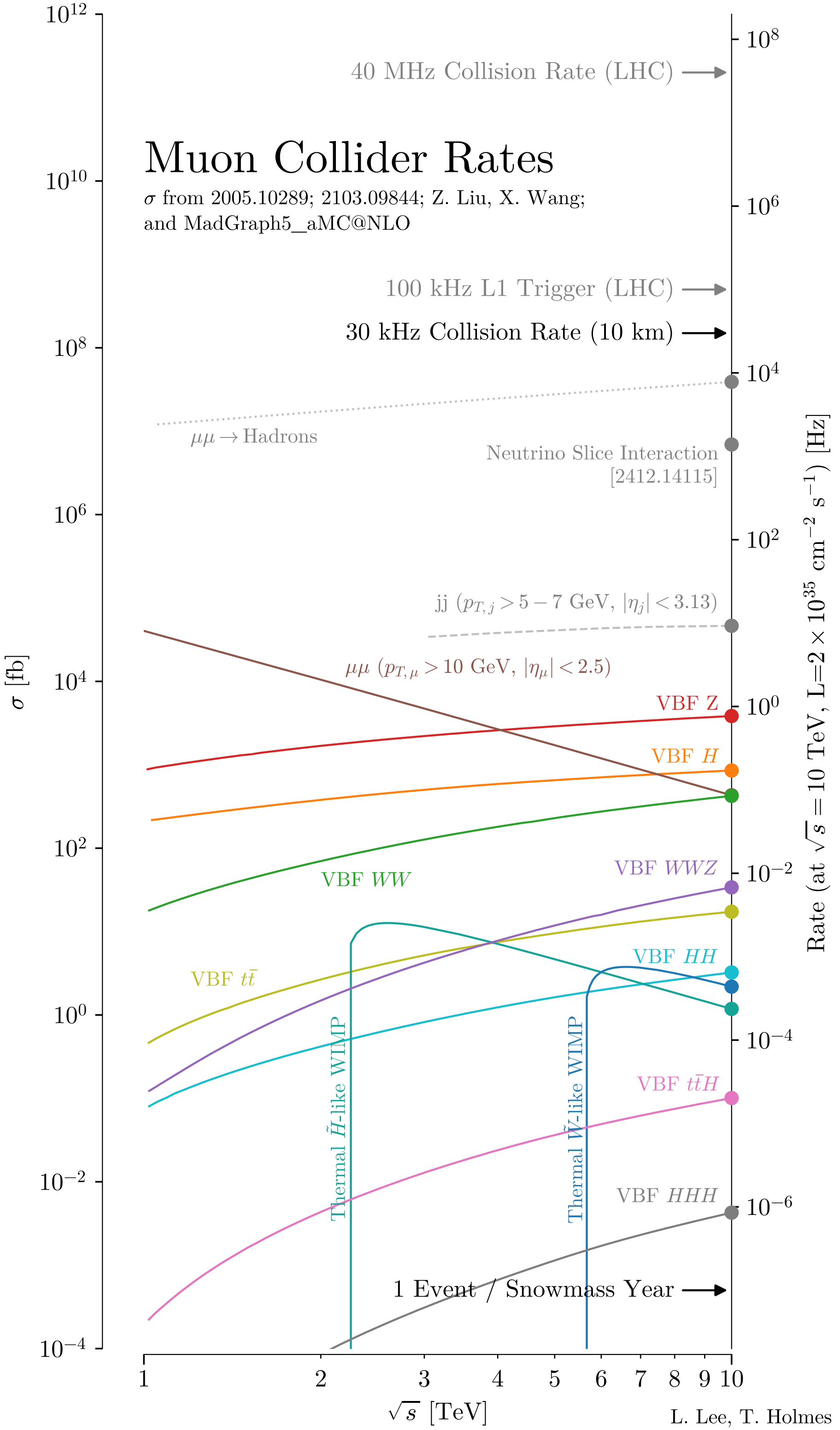
Caveats: These are "back-of-the-envelope" numbers which is approximately correct with their CDR or supporting documents. For more detail please consult the documents.

Future Colliders (per year)

N.B. Not official numbers (take this with many grains of salt…)

	fMC	Rate	Time	Size	N_{evt}	D_{disk}	C_{CPU}
HL-LHC	1.5	10kHz	7 min	6 MB	250B	1.5 EB	3.3M
FCC-ee	4	200kHz	0.1 min	1 MB	10T	10 EB	2M
FCC-hh	2	10kHz	20 min	50 MB	300B	15 EB	11M
μC (10 km)	4	1kHz	20 min	50 MB	50B	5 EB	1.9M

Caveats: These are "back-of-the-envelope" numbers which is approximately correct with their CDR or supporting documents. For more detail please consult the documents.



Muon Collider Rates

UF Chang Florida

30 kHz	10 TeV μC Collision rate
~ 10 Hz	jets (p_T > 5 - 7 GeV)
~ 1 Hz	W/Z
~ 0.2 Hz	H

Single jet P_T > 20 GeV trigger rate maybe 1 to 10 Hz

(Assuming BIB is dealt with)

Future Colliders (per year)

N.B. Not official numbers (take this with many grains of salt…)

	fMC	Rate	Time	Size	N_{evt}	D_{disk}	C_{CPU}
HL-LHC	1.5	10kHz	7 min	6 MB	250B	1.5 EB	3.3M
FCC-ee	4	200kHz	0.1 min	1 MB	10T	10 EB	2M
FCC-hh	2	10kHz	20 min	50 MB	300B	15 EB	11M
μC (10 km)	4	1kHz	20 min	50 MB	50B	5 EB	1.9M
μC (10 km)	4	10 Hz	60 min	50 MB	100M	1 EB	11k

Caveats: These are "back-of-the-envelope" numbers which is approximately correct with their CDR or supporting documents. For more detail please consult the documents.

Doing bare minimum ⇒ not extremely difficult

(However, FCC-hh is a bit hard but, I will likely never see it anyways.)

This is assuming HL-LHC works

⇒ all the HL-LHC work = future collider work

(e.g. Key4HEP, DD4Hep, ACTS, GPU, ML Reconstruction, …)

Doing bare minimum ⇒ not extremely difficult

(However, FCC-hh is a bit hard but, I will likely never see it anyways.)

This is assuming HL-LHC works
⇒ all the HL-LHC work = future collider work

(e.g. Key4HEP, DD4Hep, ACTS, GPU, ML Reconstruction, …)

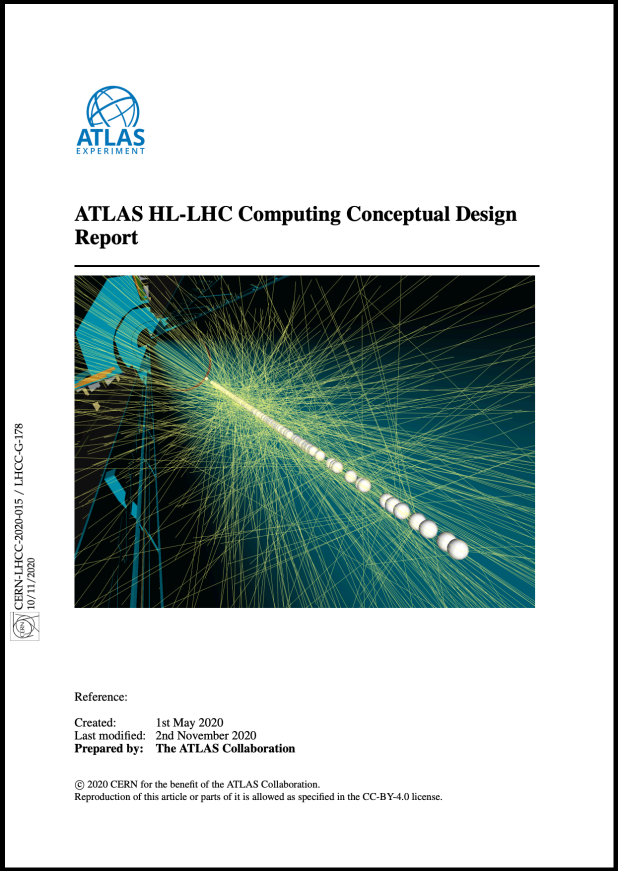
In computing for future colliders, we don't just prepare for what's coming, <u>we invent what's possible.</u>

There are many things that we can do with computers

(Software tools, ML reconstruction, tracking, event generation, GPU computing …)

But I want to focus on a couple of things

On-going HL-LHC R&D



ATLAS HL-LHC Computing Conceptual Design Report

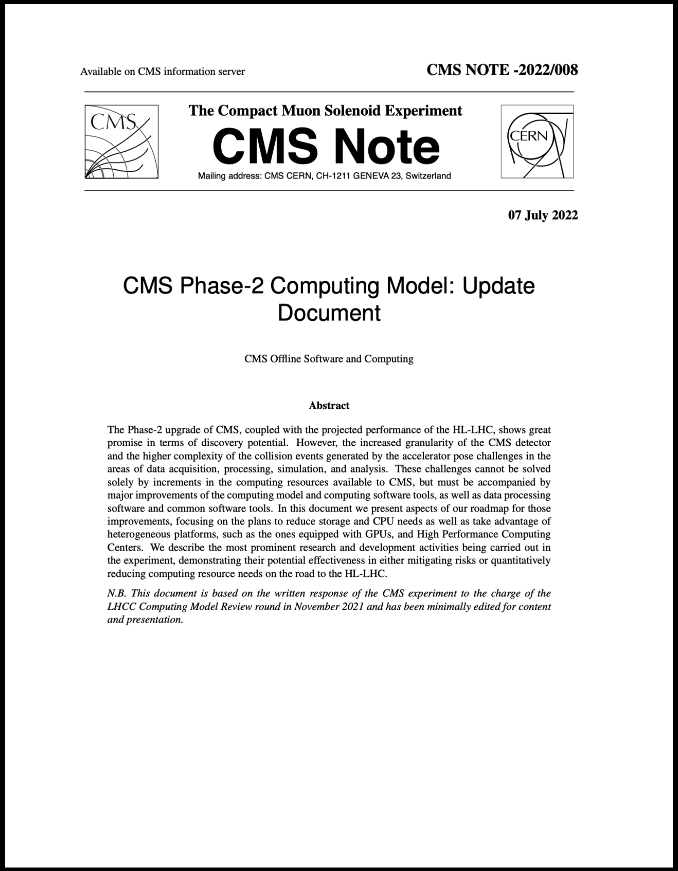
CERN-LHCC-2020-015 / LHCC-G-178
10/11/2020

Reference:

Created: 1st May 2020
Last modified: 2nd November 2020
Prepared by: The ATLAS Collaboration

© 2020 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

https://cds.cern.ch/record/2729668/files/LHCC-G-178.pdf



Available on CMS information server **CMS NOTE -2022/008**

The Compact Muon Solenoid Experiment
CMS Note
Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

07 July 2022

CMS Phase-2 Computing Model: Update Document

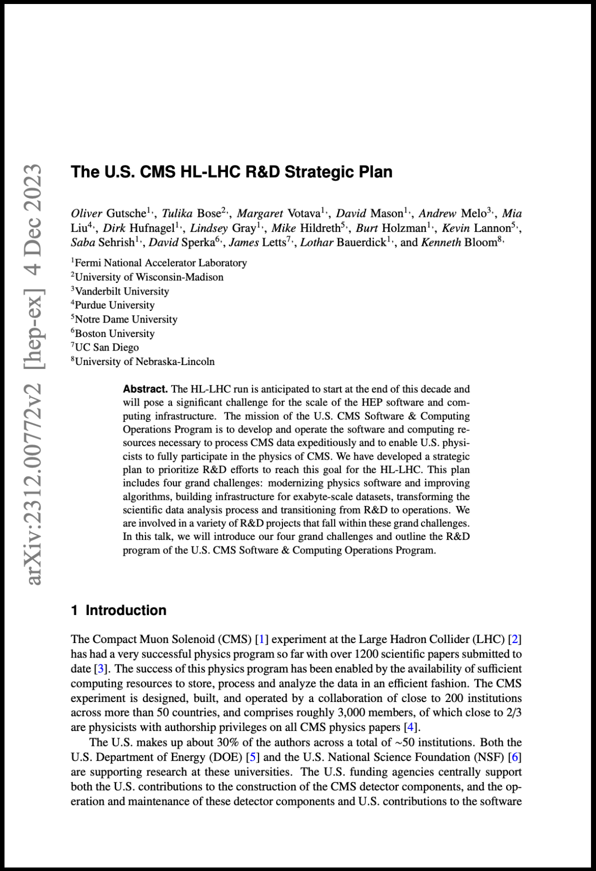
CMS Offline Software and Computing

Abstract

The Phase-2 upgrade of CMS, coupled with the projected performance of the HL-LHC, shows great promise in terms of discovery potential. However, the increased granularity of the CMS detector and the higher complexity of the collision events generated by the accelerator pose challenges in the areas of data acquisition, processing, simulation, and analysis. These challenges cannot be solved solely by increments in the computing resources available to CMS, but must be accompanied by major improvements of the computing model and computing software tools, as well as data processing software and common software tools. In this document we present aspects of our roadmap for those improvements, focusing on the plans to reduce storage and CPU needs as well as take advantage of heterogeneous platforms, such as the ones equipped with GPUs, and High Performance Computing Centers. We describe the most prominent research and development activities being carried out in the experiment, demonstrating their potential effectiveness in either mitigating risks or quantitatively reducing computing resource needs on the road to the HL-LHC.

N.B. This document is based on the written response of the CMS experiment to the charge of the LHCC Computing Model Review round in November 2021 and has been minimally edited for content and presentation.

https://cds.cern.ch/record/2815292?ln=en



arXiv:2312.00772v2 [hep-ex] 4 Dec 2023

The U.S. CMS HL-LHC R&D Strategic Plan

Oliver Gutsche[1], *Tulika* Bose[2], *Margaret* Votava[1], *David* Mason[1], *Andrew* Melo[3], *Mia* Liu[4], *Dirk* Hufnagel[1], *Lindsey* Gray[1], *Mike* Hildreth[5], *Burt* Holzman[1], *Kevin* Lannon[5], *Saba* Sehrish[1], *David* Sperka[6], *James* Letts[7], *Lothar* Bauerdick[1], and *Kenneth* Bloom[8]

[1]Fermi National Accelerator Laboratory
[2]University of Wisconsin-Madison
[3]Vanderbilt University
[4]Purdue University
[5]Notre Dame University
[6]Boston University
[7]UC San Diego
[8]University of Nebraska-Lincoln

Abstract. The HL-LHC run is anticipated to start at the end of this decade and will pose a significant challenge for the scale of the HEP software and computing infrastructure. The mission of the U.S. CMS Software & Computing Operations Program is to develop and operate the software and computing resources necessary to process CMS data expeditiously and to enable U.S. physicists to fully participate in the physics of CMS. We have developed a strategic plan to prioritize R&D efforts to reach this goal for the HL-LHC. This plan includes four grand challenges: modernizing physics software and improving algorithms, building infrastructure for exabyte-scale datasets, transforming the scientific data analysis process and transitioning from R&D to operations. We are involved in a variety of R&D projects that fall within these grand challenges. In this talk, we will introduce our four grand challenges and outline the R&D program of the U.S. CMS Software & Computing Operations Program.

1 Introduction

The Compact Muon Solenoid (CMS) [1] experiment at the Large Hadron Collider (LHC) [2] has had a very successful physics program so far with over 1200 scientific papers submitted to date [3]. The success of this physics program has been enabled by the availability of sufficient computing resources to store, process and analyze the data in an efficient fashion. The CMS experiment is designed, built, and operated by a collaboration of close to 200 institutions across more than 50 countries, and comprises roughly 3,000 members, of which close to 2/3 are physicists with authorship privileges on all CMS physics papers [4].

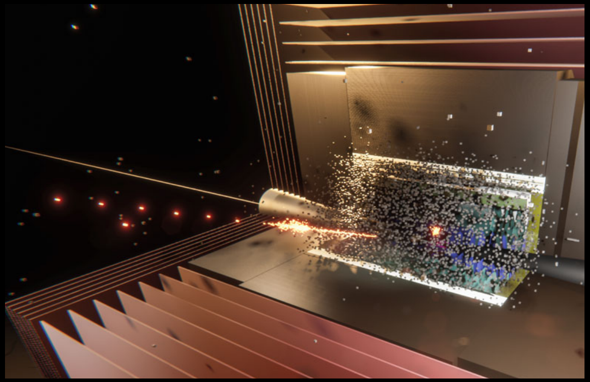
The U.S. makes up about 30% of the authors across a total of ~50 institutions. Both the U.S. Department of Energy (DOE) [5] and the U.S. National Science Foundation (NSF) [6] are supporting research at these universities. The U.S. funding agencies centrally support both the U.S. contributions to the construction of the CMS detector components, and the operation and maintenance of these detector components and U.S. contributions to the software

https://arxiv.org/pdf/2312.00772v2

RAW → Analysis

Raw data

(channel readout)

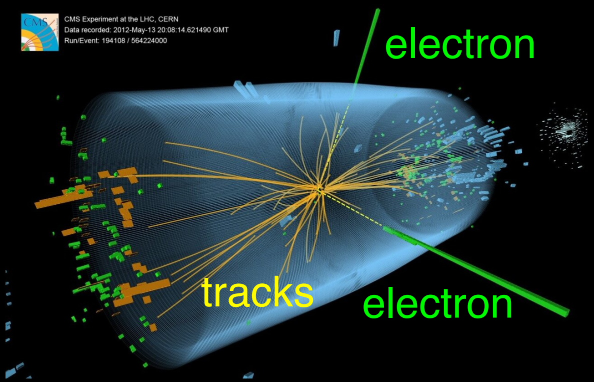


Fully
Machine
Learning

→

Useful Analysis data

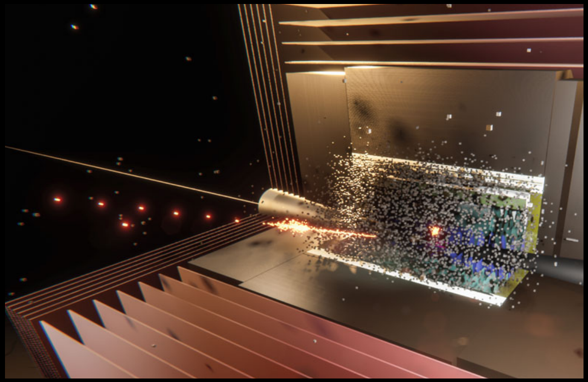
(physics objects)



RAW → Analysis

Raw data

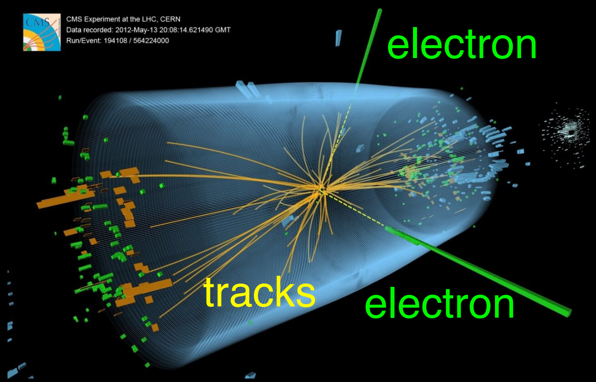
(channel readout)



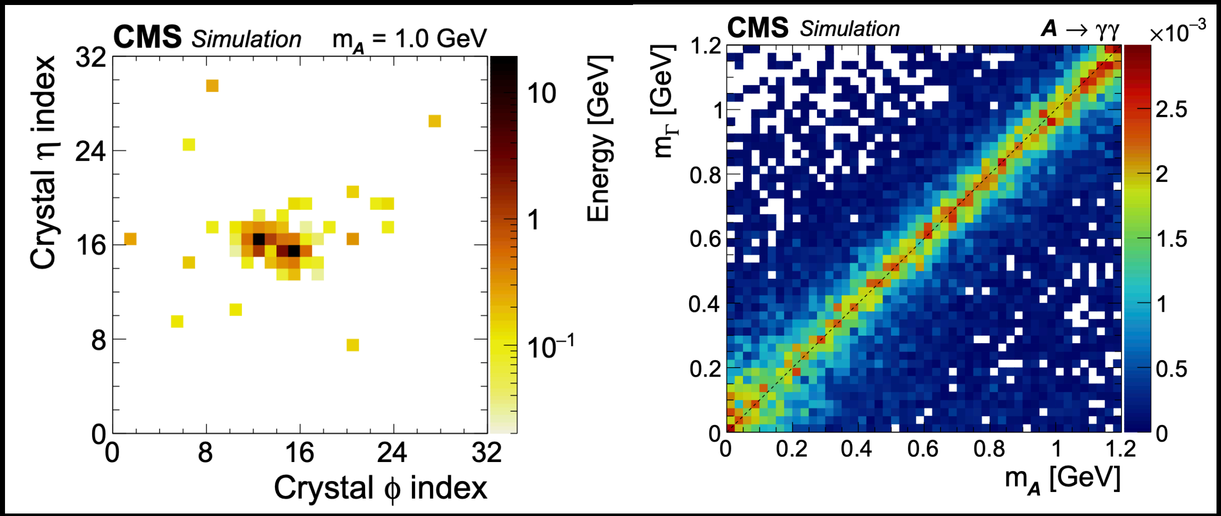
Fully Machine Learning

Useful Analysis data

(physics objects)

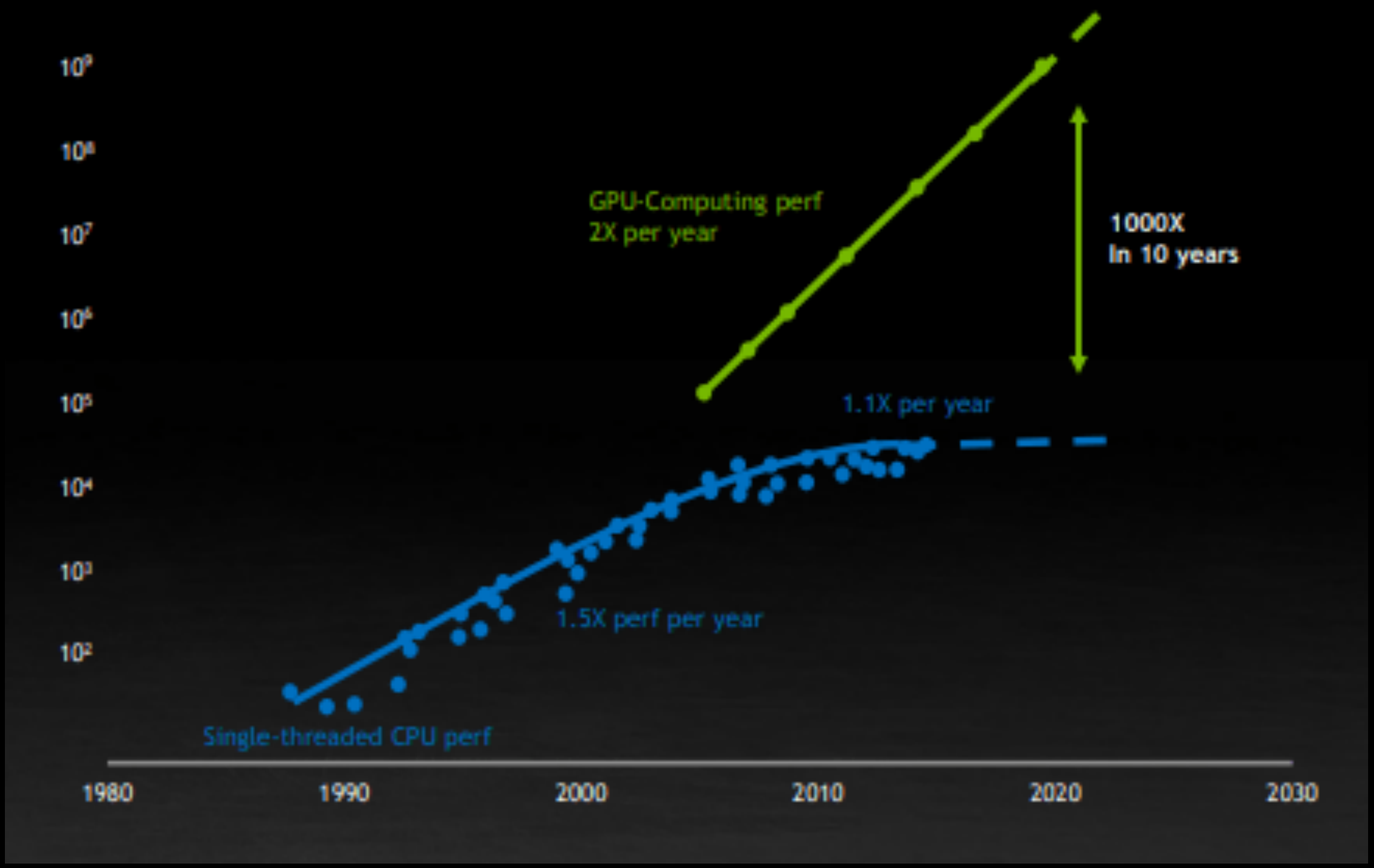


CMS-EGM-20-001



GPU (Huang's Law)

UF
Chang
Florida



NVIDIA

Online Computation

UF
Chang
Florida

Muon Collider has relatively low rate

Online Computation

UF
Chang
Florida

Muon Collider has relatively low rate

But large number of channels and hits
mean that event size are large

Online Computation

Muon Collider has relatively low rate

But large number of channels and hits
mean that event size are large

This limits how many events
we can read out

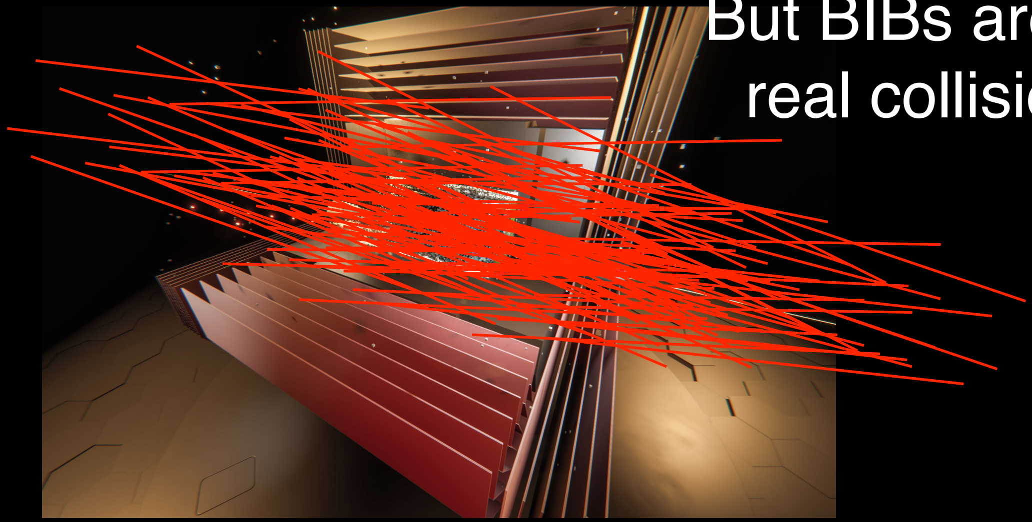
Online Computation

UF
Chang
Florida

Muon Collider has relatively low rate

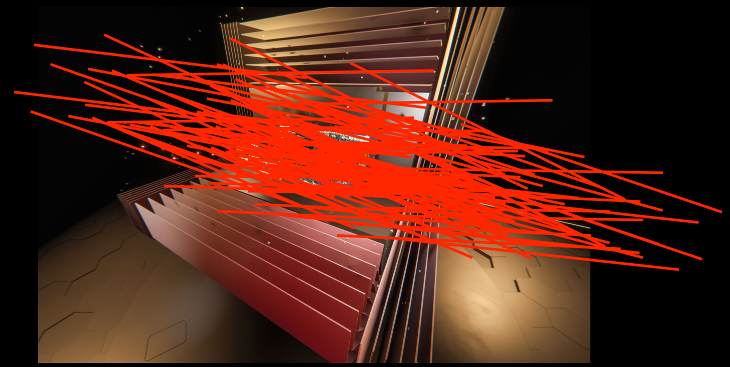
But large number of channels and hits mean that event size are large

This limits how many events we can read out



But BIBs are not real collisions

Data Compression / Filtering



If one can suppress the readout
(e.g. putting processor close to
detector readout level)

"Triggerless" or "streaming"

⇒ One could be saving
entire experimental events

Future Colliders (per year)

UF
Chang
Florida

N.B. Not official numbers (take this with many grains of salt…)

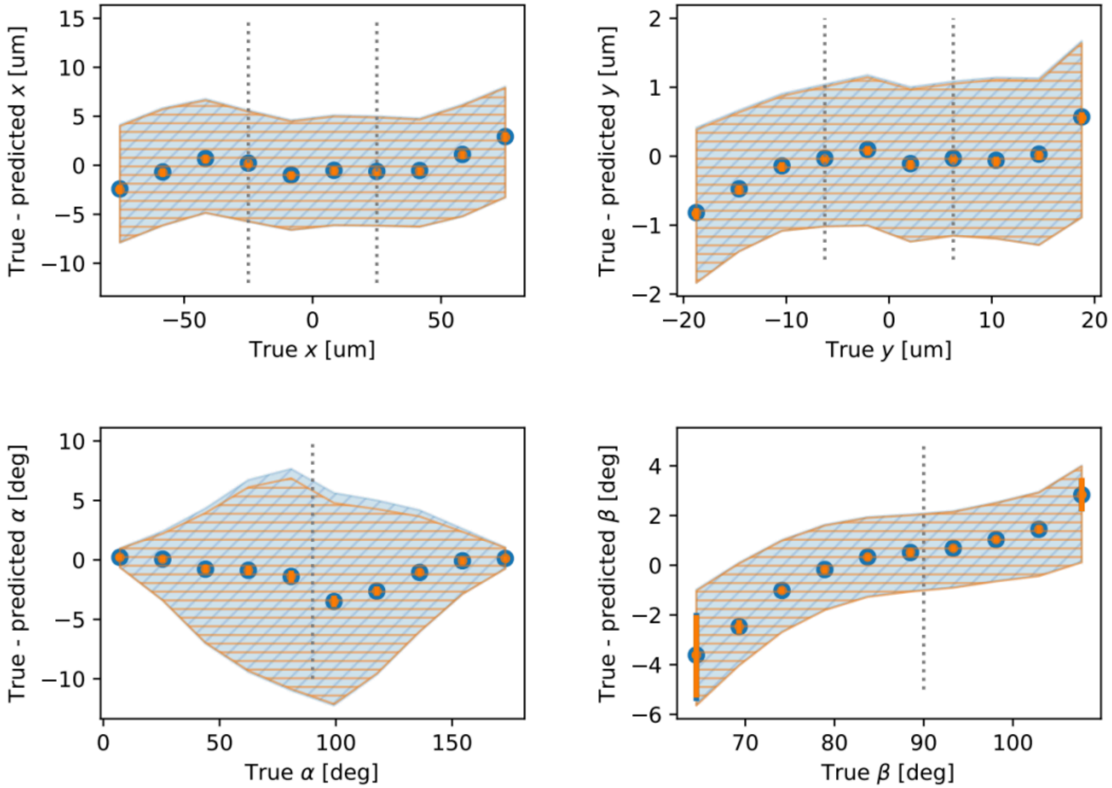
	fMC	Rate	Time	Size	N_{evt}	D_{disk}	C_{CPU}
HL-LHC	1.5	10kHz	7 min	6 MB	250B	1.5 EB	3.3M
FCC-ee	4	200kHz	0.1 min	1 MB	10T	10 EB	2M
FCC-hh	2	10kHz	20 min	50 MB	300B	15 EB	11M
μC (10 km)	4	1kHz	20 min	50 MB	50B	5 EB	1.9M
μC (10 km)	4	10 Hz	60 min	50 MB	100M	1 EB	11k
μC streaming	9	30kHz	0.1 min	1 MB	3T	3 EB	0.6M

Caveats: These are "back-of-the-envelope" numbers which is approximately correct with their CDR or supporting documents. For more detail please consult the documents.

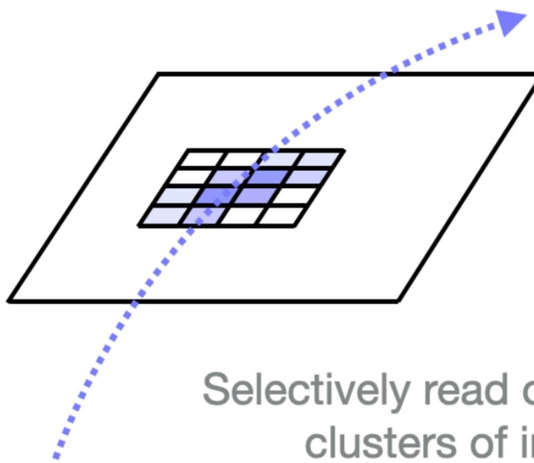
Example

fastml/smart-pixels

Simple regression algorithm
in QKeras and hls4ml

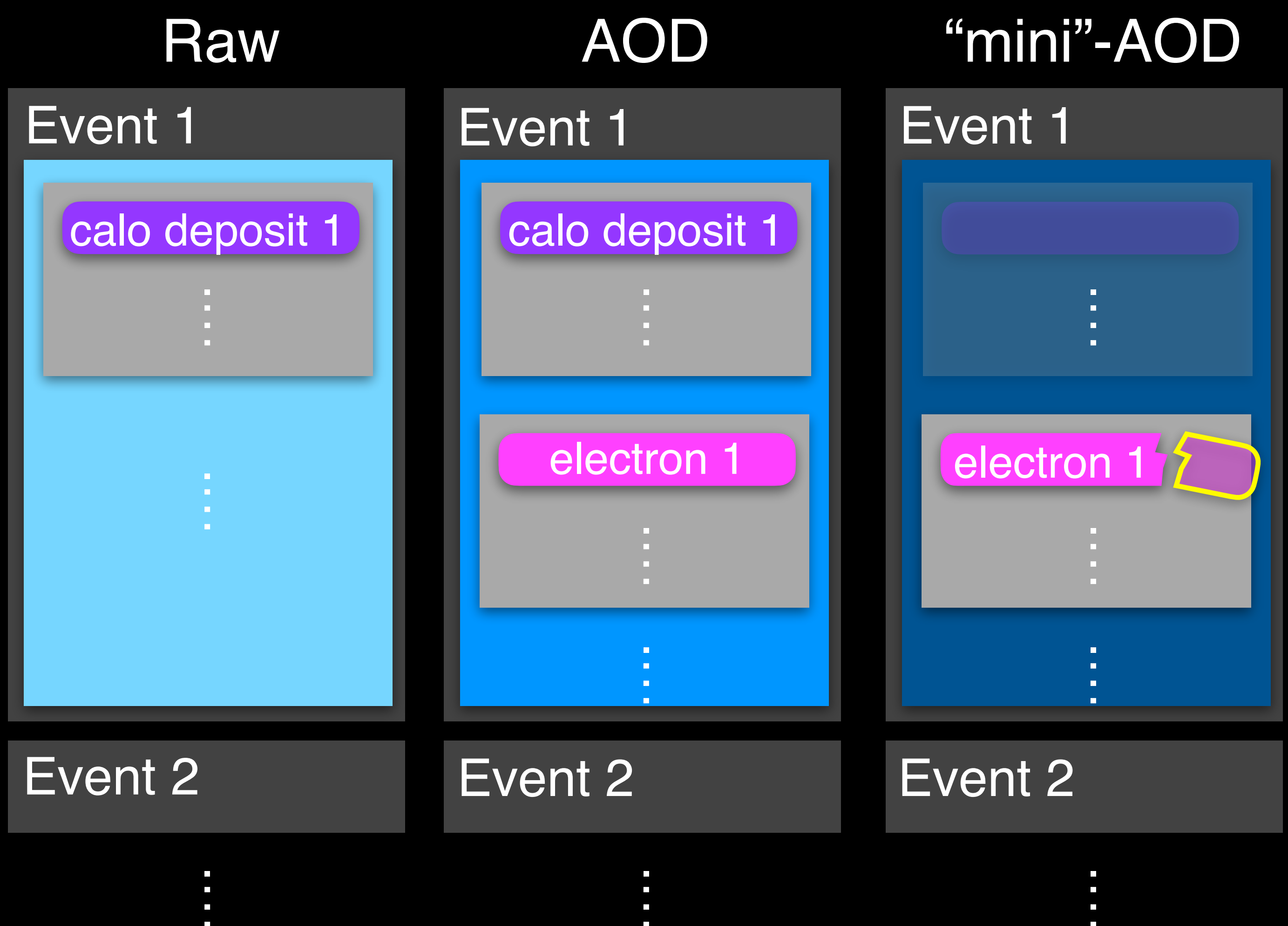


Infer properties of incident particle from cluster of pixel hits



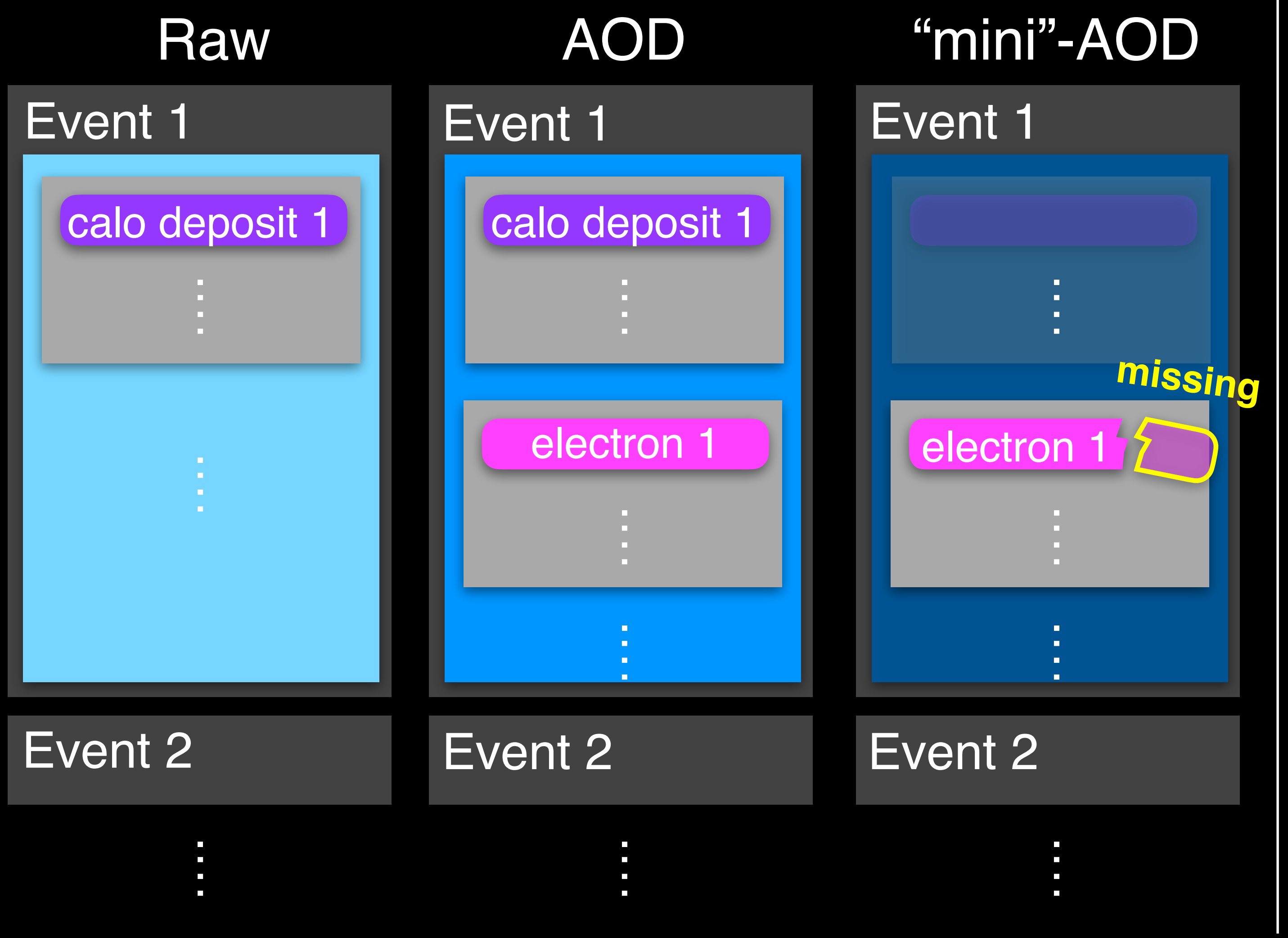
Selectively read out/featurize clusters of interest

Getting rid of data tiers



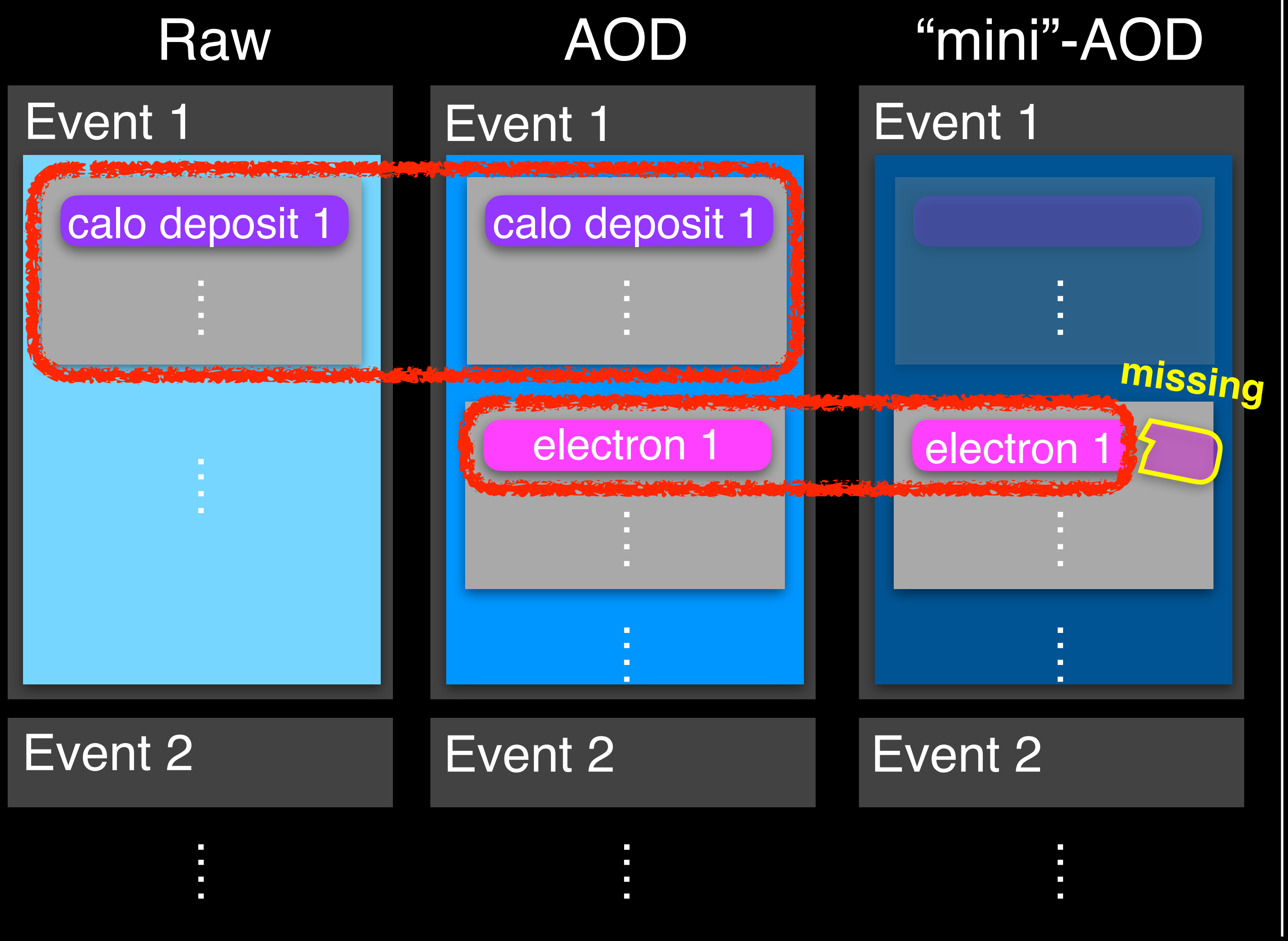
Getting rid of data tiers

UF
Chang
Florida

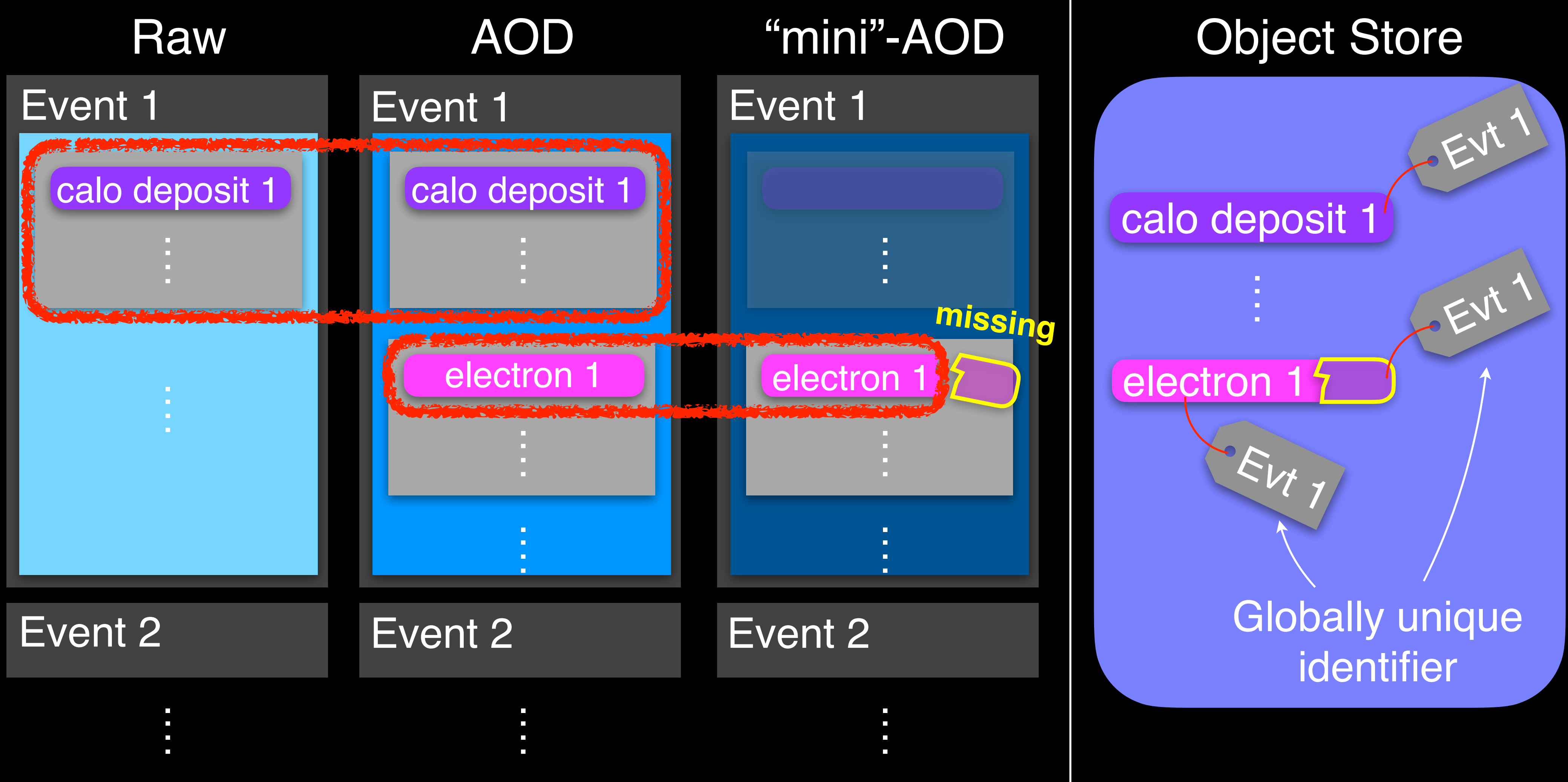


Getting rid of data tiers

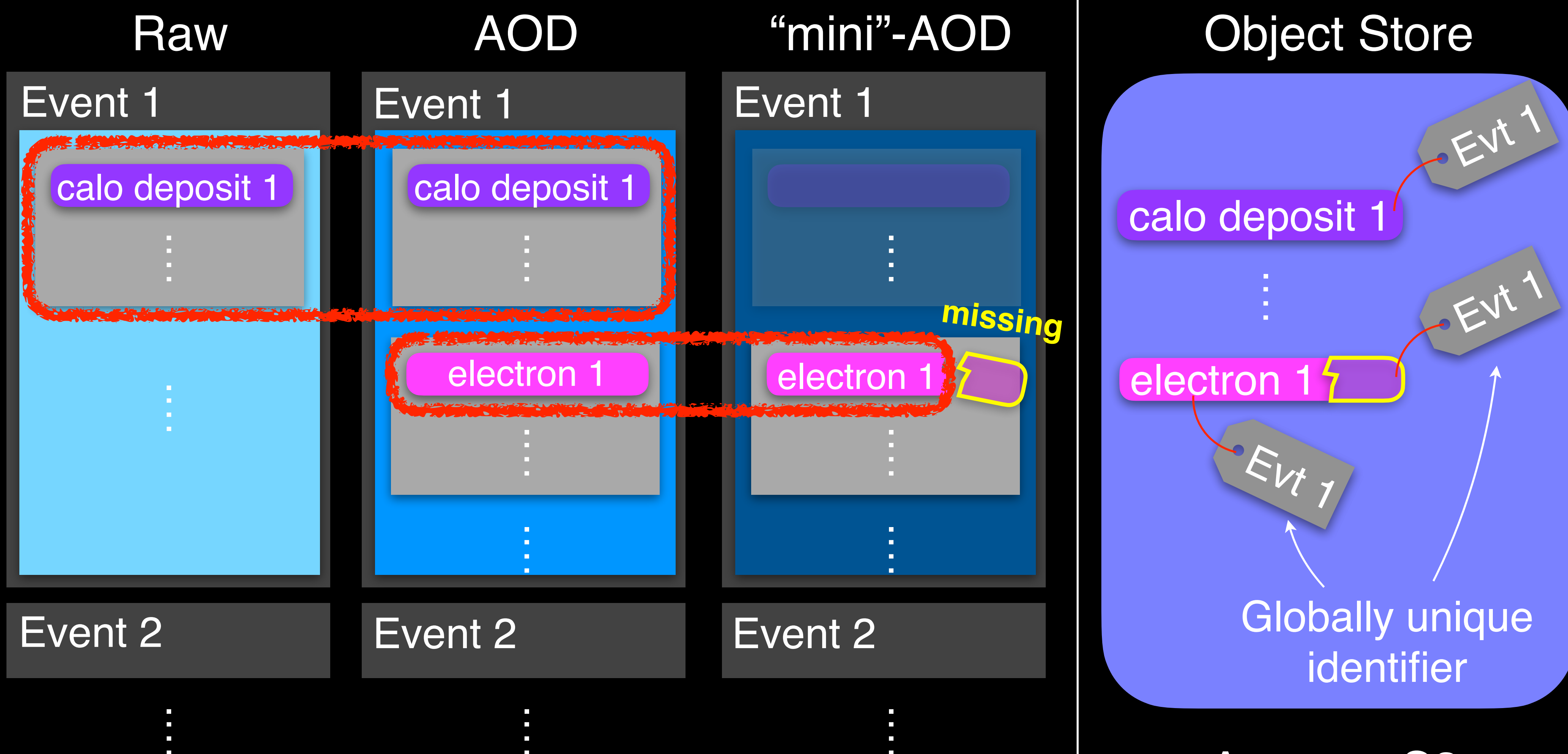
UF
Chang
Florida



Getting rid of data tiers

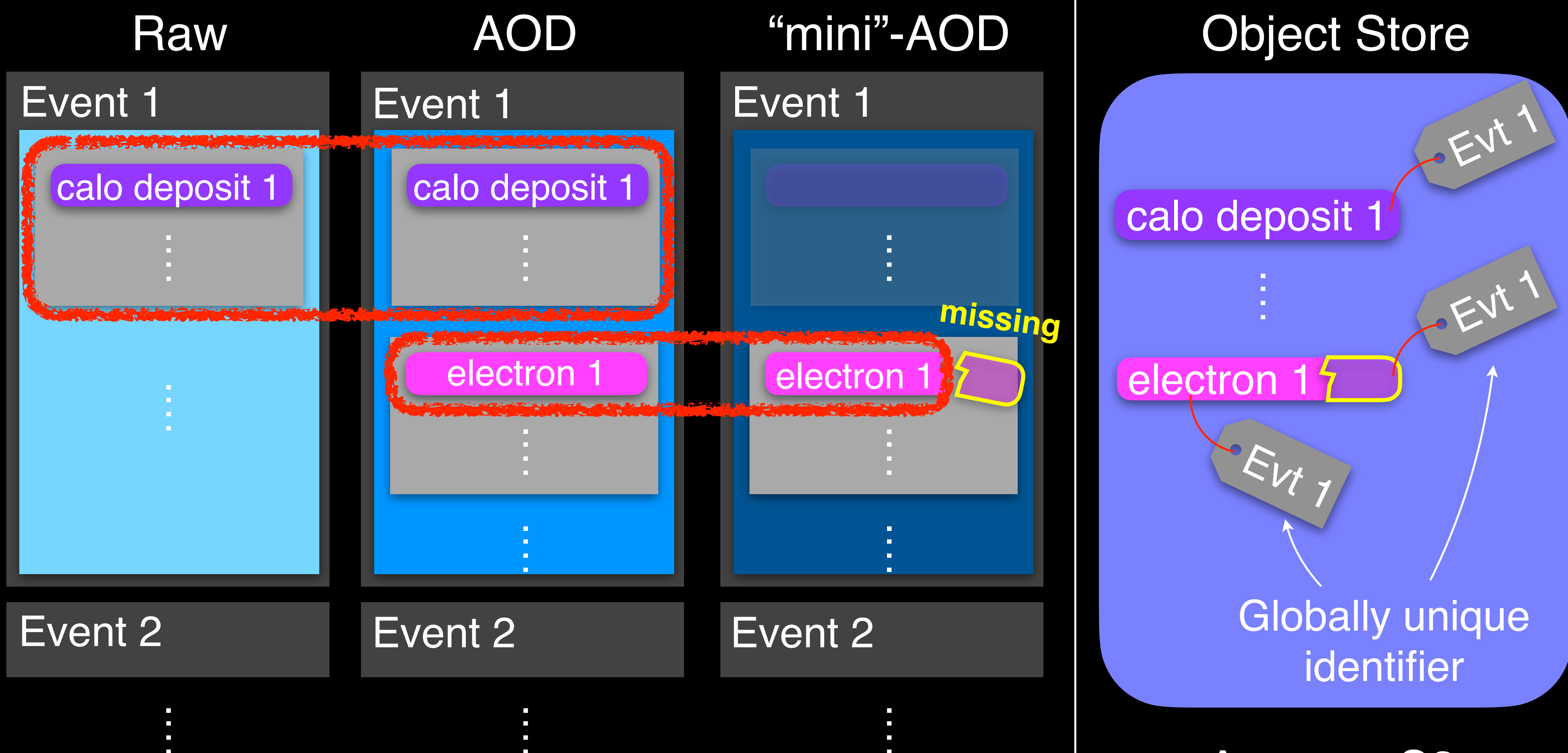


Getting rid of data tiers



Amazon S3
Azure Blob storage
Google Cloud Storage

Getting rid of data tiers

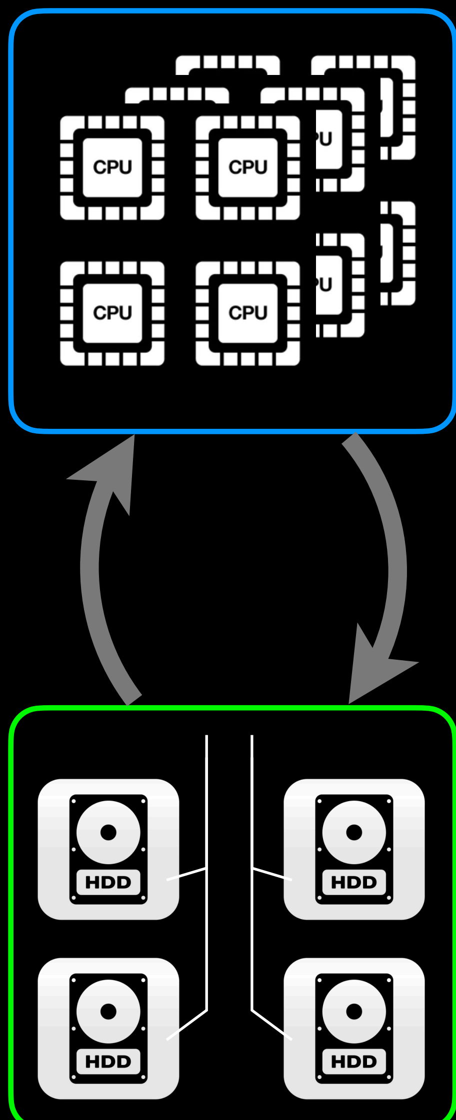


Amazon S3
Azure Blob storage
Google Cloud Storage

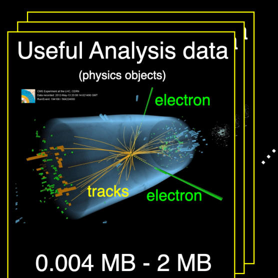
Once we forgo data-tiers, we can access "lower-level information" on demand

Analysis Facilities

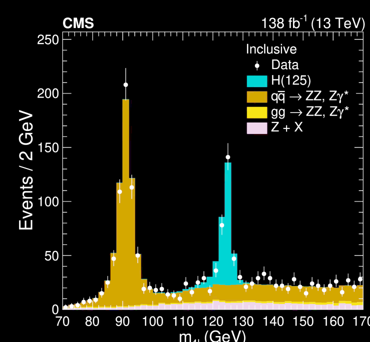
CPU farm



Disk Storage

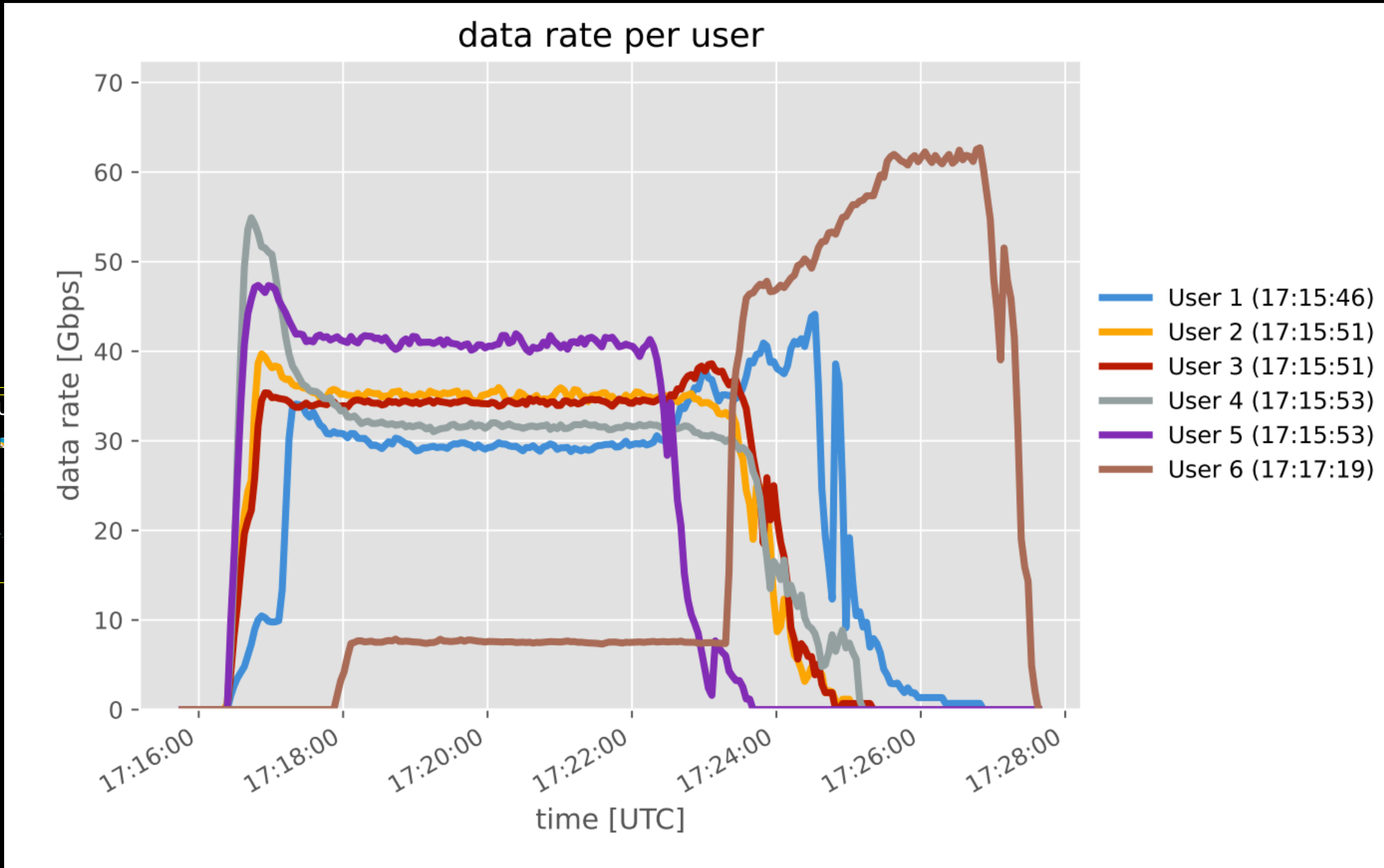


Often O(10MB) per CPU due to network



O(100) Gbps

Analysis Facilities

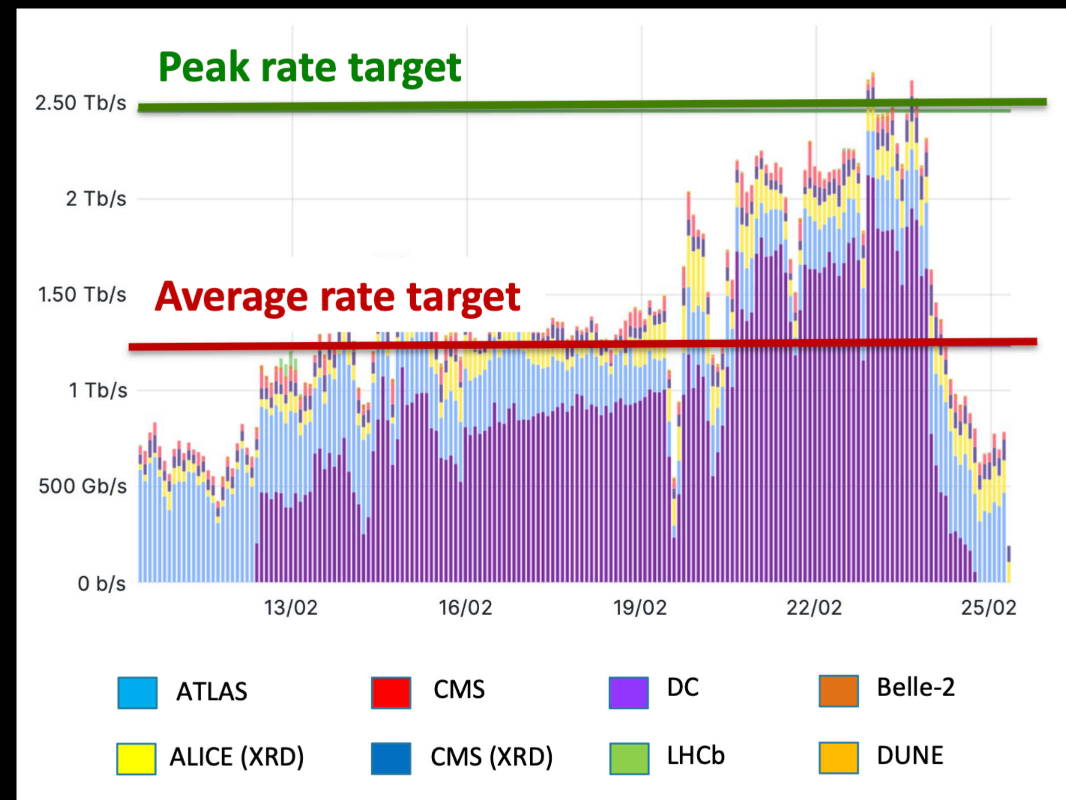


Disk Storage

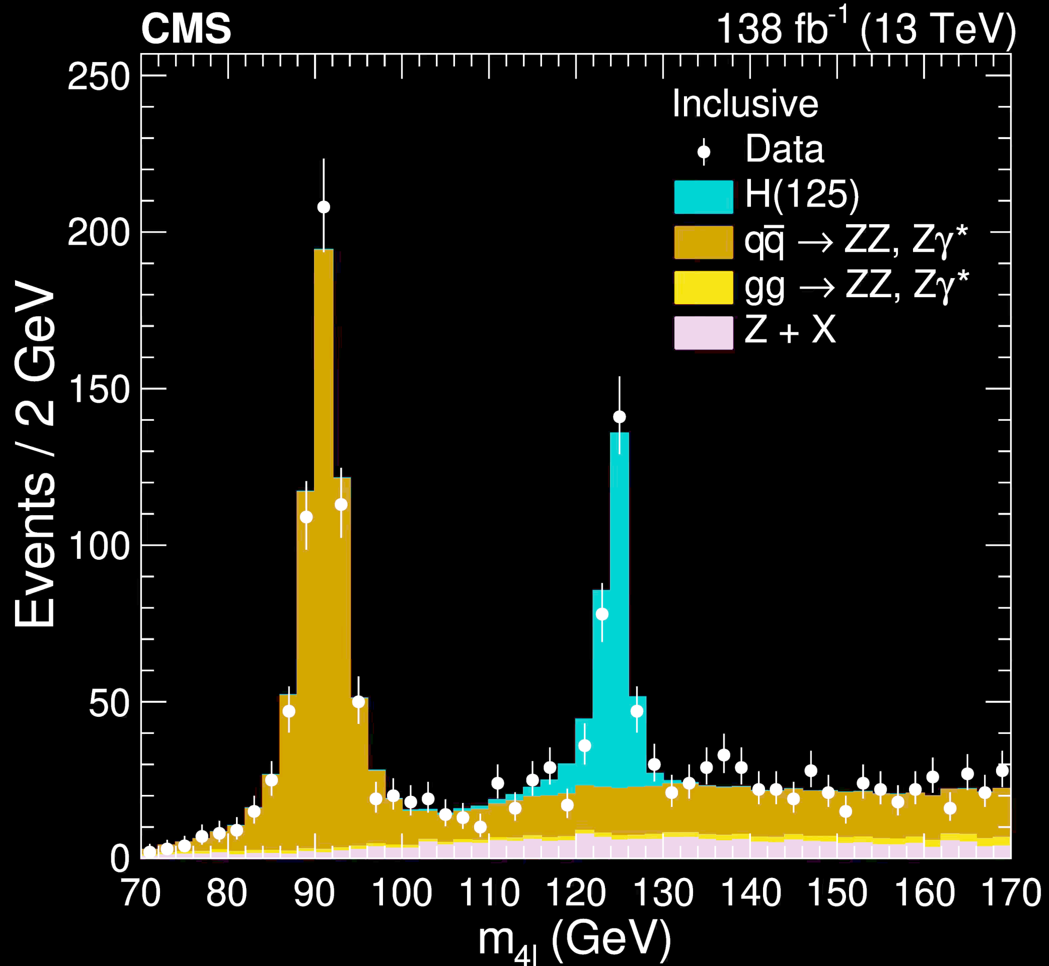
Networking Challenges

UF
Chang
Florida

WLCG
Worldwide LHC Computing Grid

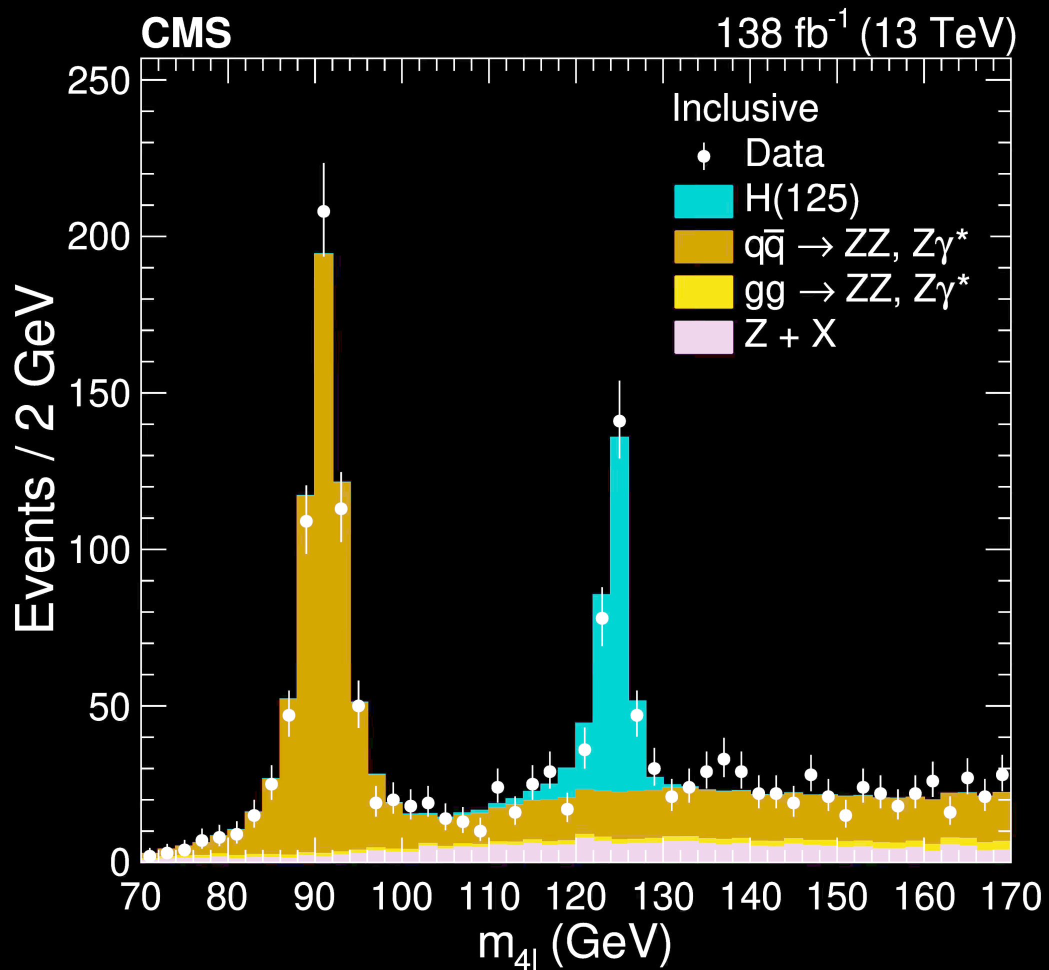


Importance of data skimming



How many analyses make plots like this?

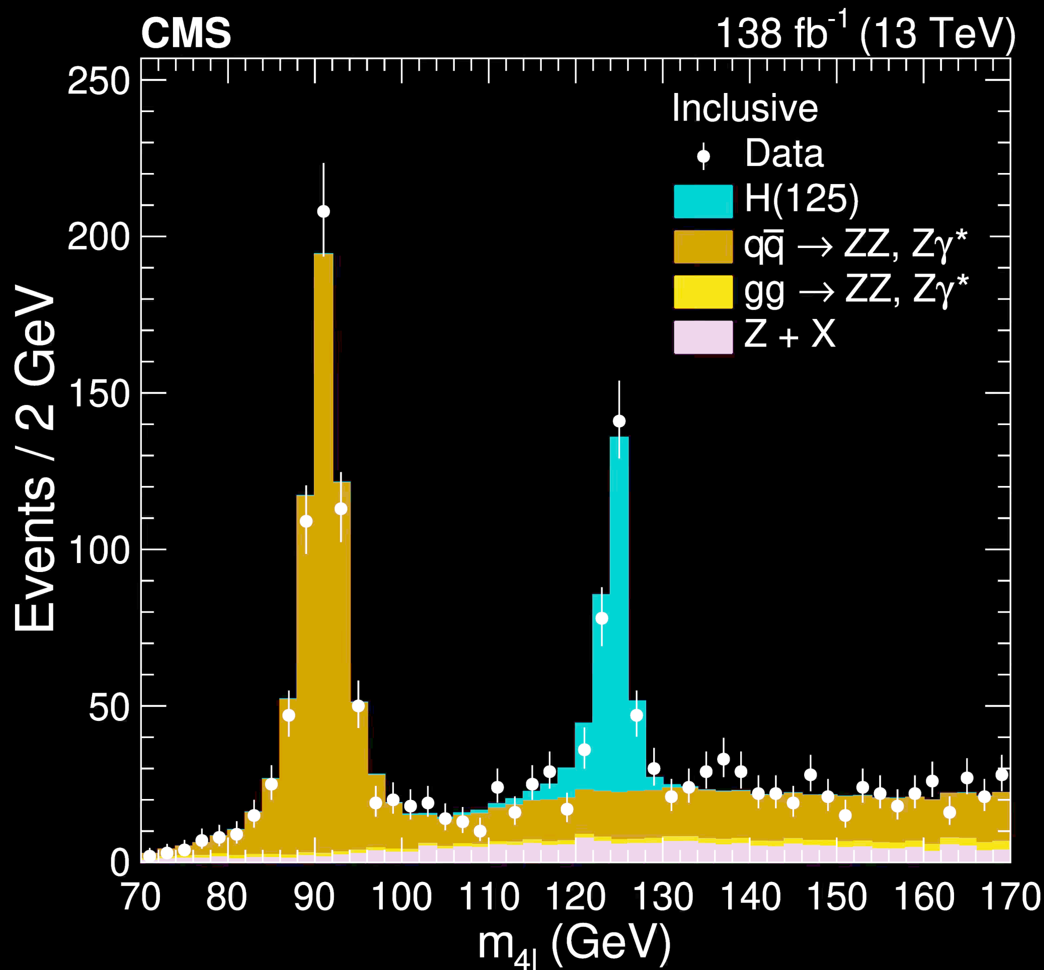
Importance of data skimming



How many analyses make plots like this?

Particle physics analyses are "embarassingly parallelizable"

Importance of data skimming

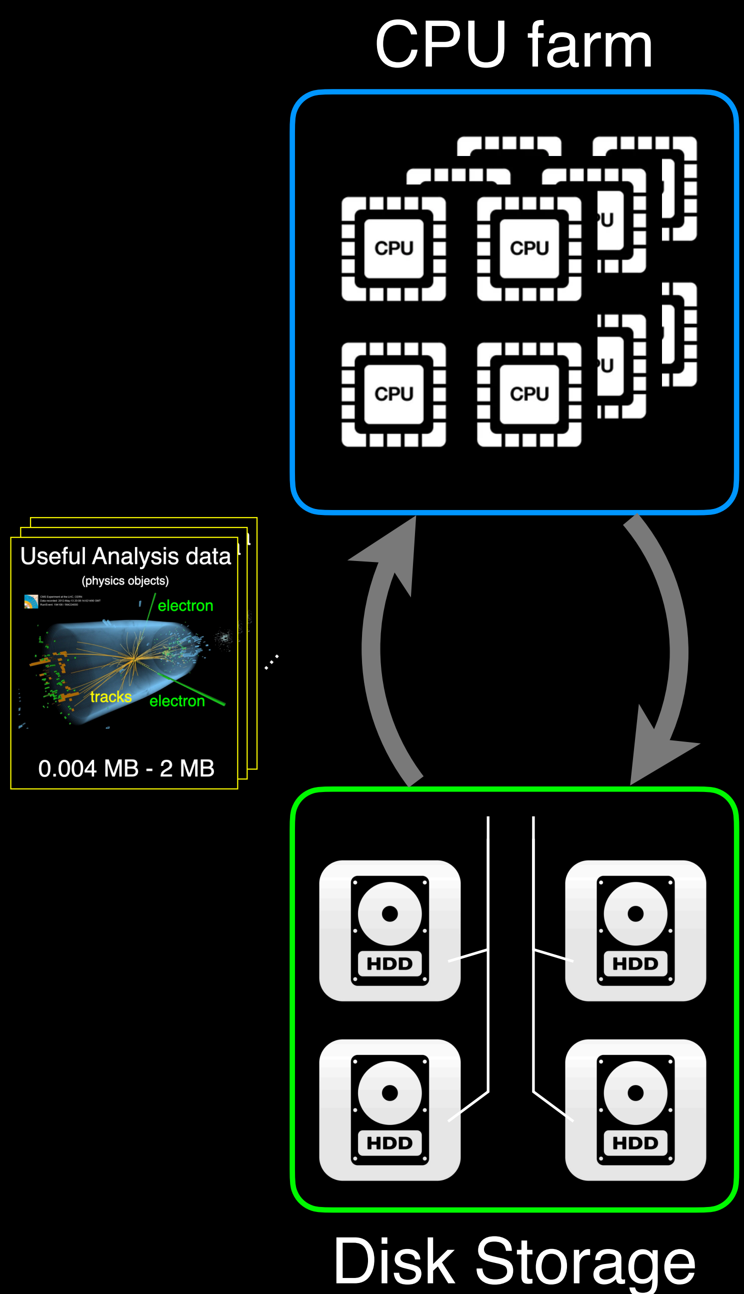


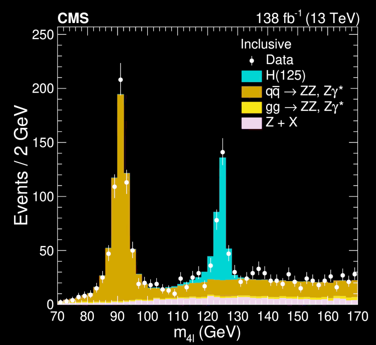
How many analyses make plots like this?

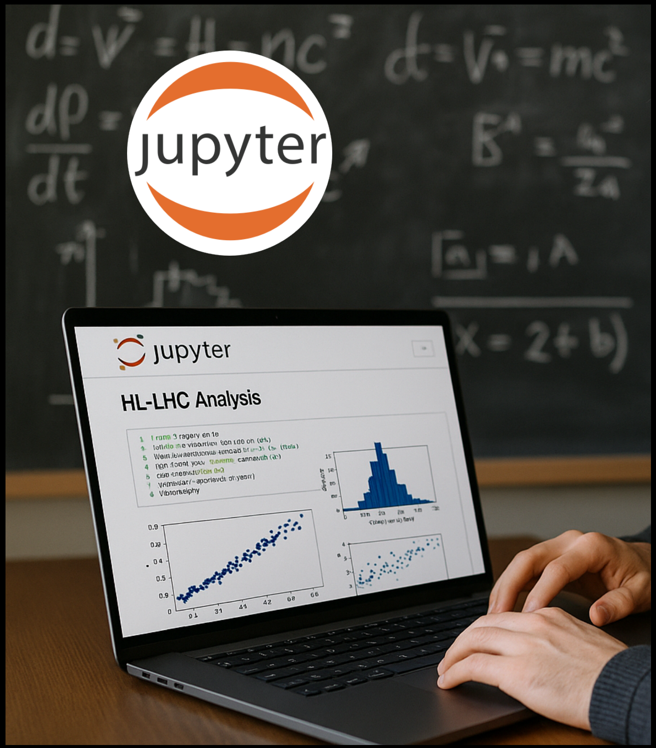
Particle physics analyses are "embarassingly parallelizable"

e.g. "select events with 4 leptons"

Analysis Facilities



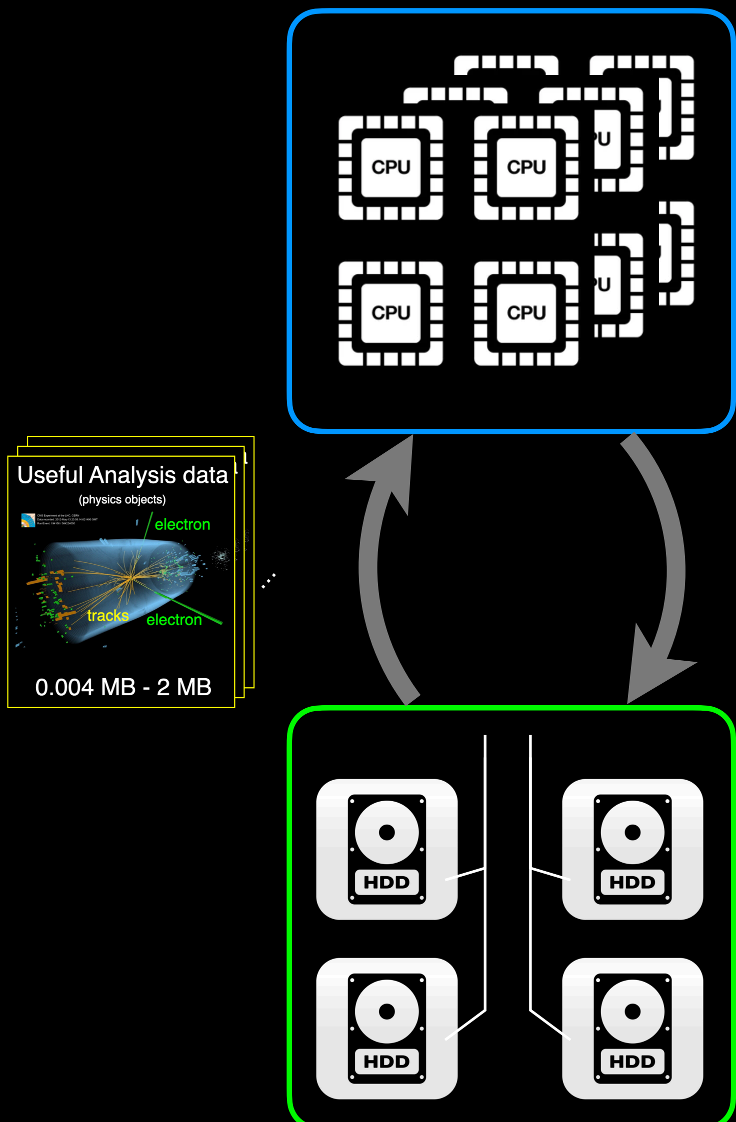




Analysis Facilities

UF
Chang
Florida

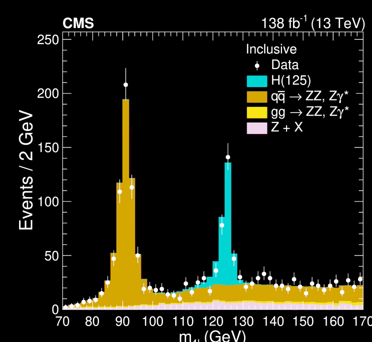
CPU farm



Disk Storage

Often O(10MB/s) per CPU due to network

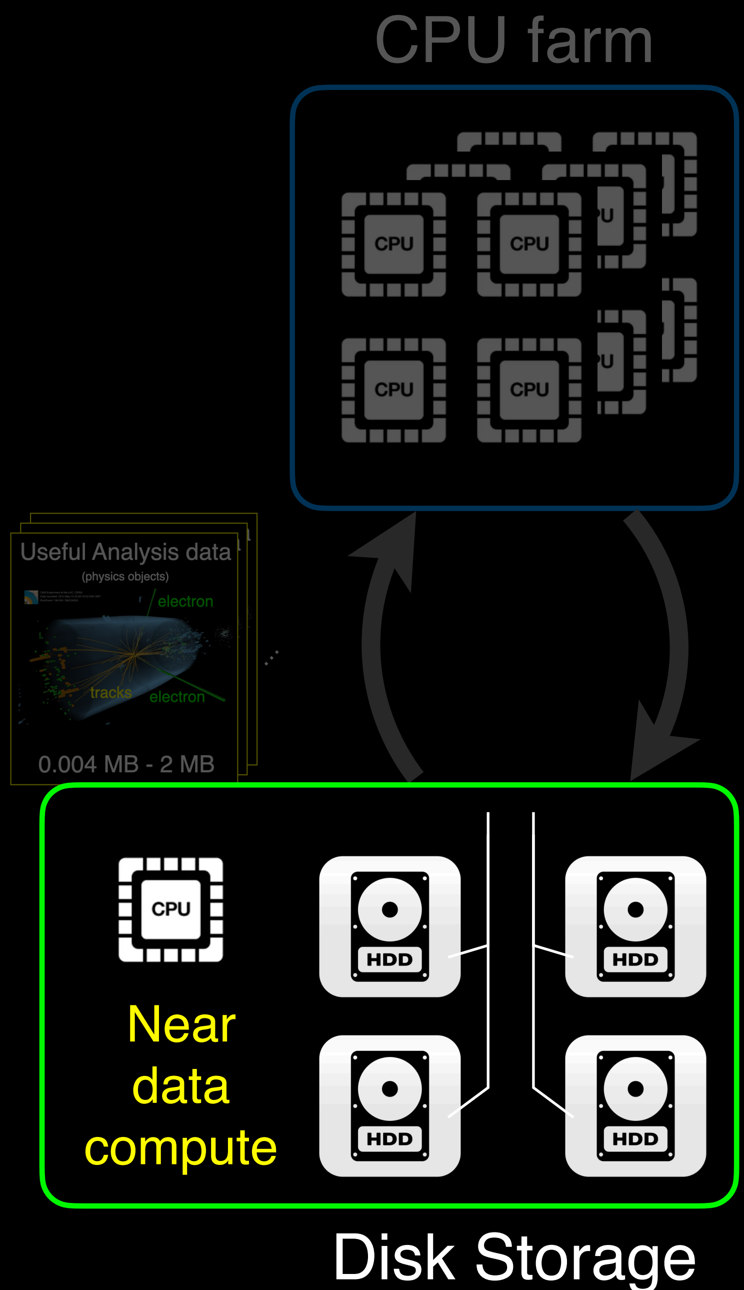
But they can take O(10) GB/s!

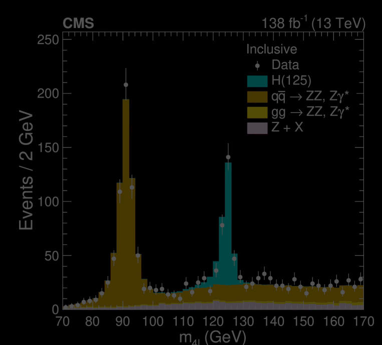


O(100) Gbps

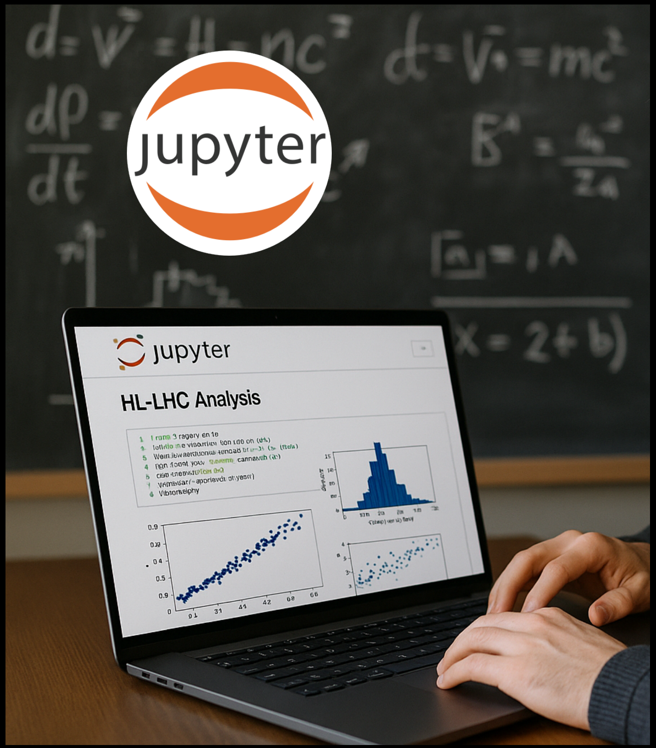
Instead…

UF
Chang
Florida

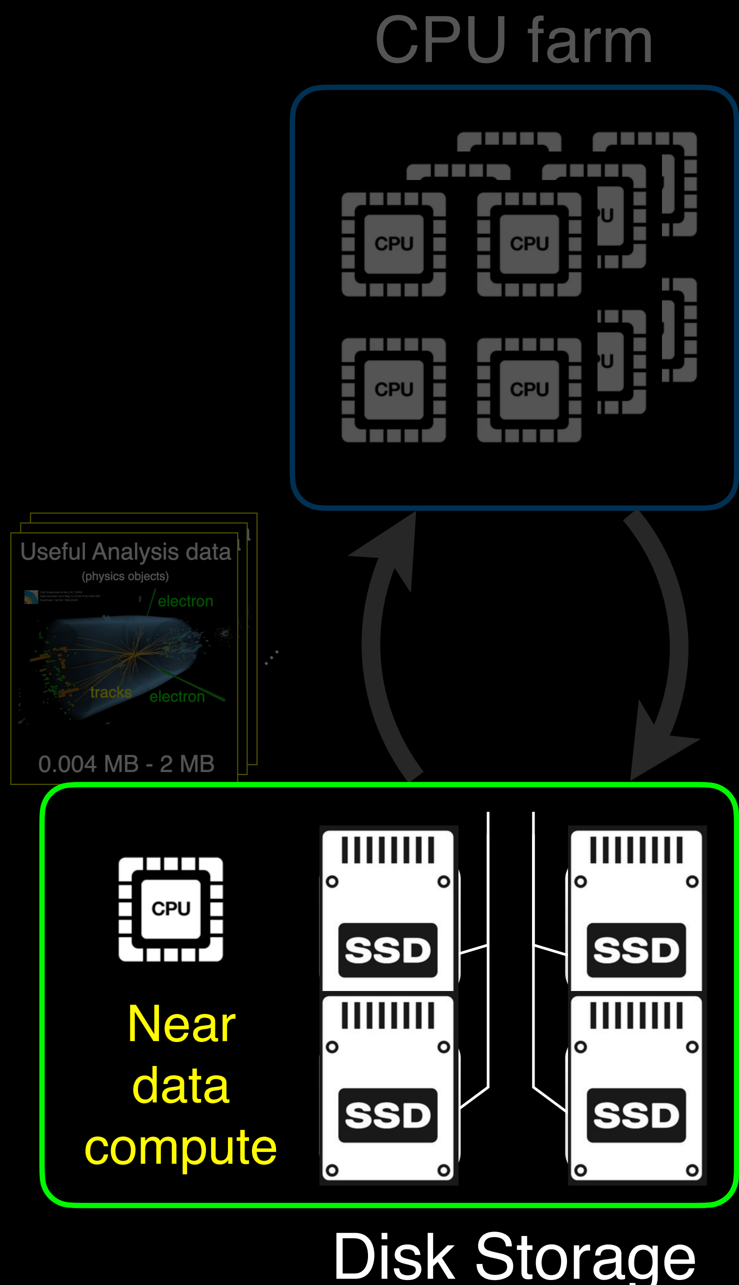


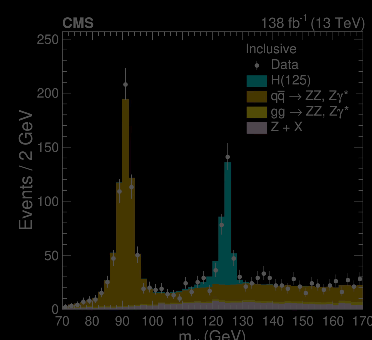


Utilize full bandwidth and skim the data O(100) Gbps



Instead…



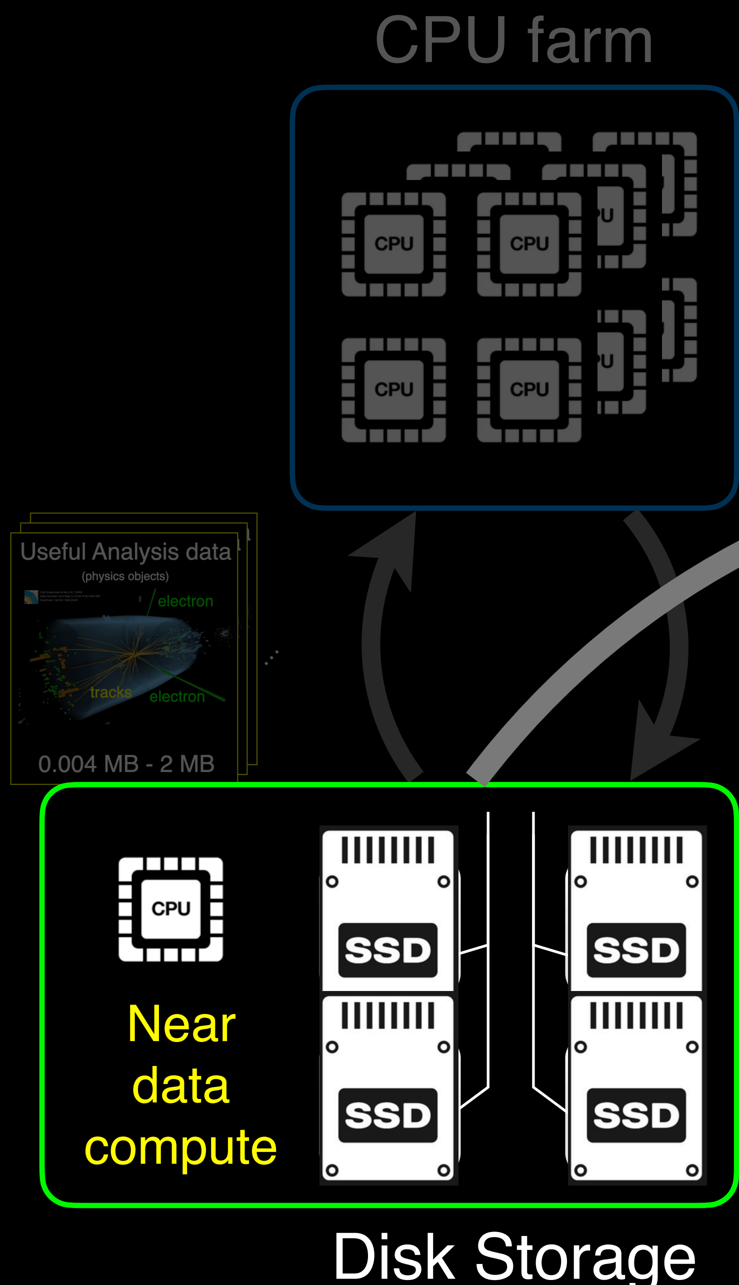


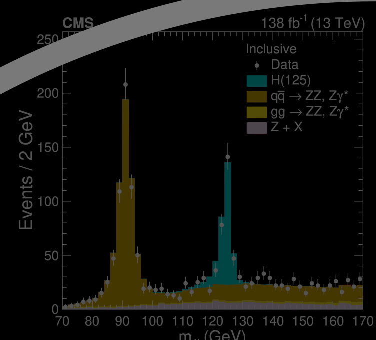
Utilize full bandwidth and skim the data O(100) Gbps

Likely require NVMe

Instead…

UF
Chang
Florida





Utilize full bandwidth and skim the data O(100) Gbps

Transfer O(10 GB) data and run your analysis on laptop

Likely require NVMe

Summary

UF
Chang
Florida

- I presented "back-of-the-envelope" style of computing needs
- Various future colliders have its own challenges
- HL-LHC challenges that we are already working to solve are directly applicable to future collider computing challenges

Summary

- I presented "back-of-the-envelope" style of computing needs
- Various future colliders have its own challenges
- HL-LHC challenges that we are already working to solve are directly applicable to future collider computing challenges

In computing for future colliders, we don't just prepare for what's coming, <u>we invent what's possible.</u>

Summary

UF
Chang
Florida

- I presented "back-of-the-envelope" style of computing needs
- Various future colliders have its own challenges
- HL-LHC challenges that we are already working to solve are directly applicable to future collider computing challenges

In computing for future colliders, we don't just prepare for what's coming, <u>we invent what's possible.</u>

- End-to-end event reconstruction using machine learning
- Getting rid of data-tier structure and more flexibility
- Data compression on detector readout to allow "triggerless" approach
- Overcoming networking challenges via near-data compute

Backup

Title

Table 3: CMS preliminary resource request for 2026 in the default scenario where 2026 is a shutdown year and the alternate scenario where 2026 is a data taking year. The percentage changes with respect to the approved 2025 request are shown, as well as the different between the alternate and default scenarios.

CMS		2025	2026 Preliminary		Increase with respect to 2025		
		Approved	Default	Alternate	Default	Alternate	Difference
CPU [kHS23]	Tier-0	1,180	1,180	1,180	0 (0%)	0 (0%)	0
	Tier-1	1,100	1,100	1,200	0 (0%)	100 (8%)	100
	Tier-2	1,900	1,900	2,000	0 (0%)	100 (5%)	100
	Total	**4,180**	**4,180**	**4,380**	**0 (0%)**	**200 (5%)**	**200**
Disk [PB]	Tier-0	70	70	73	0 (0%)	3 (4%)	3
	Tier-1	142	150	160	8 (5%)	18 (13%)	10
	Tier-2	175	185	195	10 (6%)	20 (11%)	10
	Total	**387**	**405**	**428**	**18 (5%)**	**41 (11%)**	**23**
Tape [PB]	Tier-0	442	442	462	0 (0%)	20 (5%)	20
	Tier-1	445	452	470	7 (2%)	25 (6%)	18
	Total	**887**	**894**	**932**	**7 (1%)**	**45 (5%)**	**38**

UF
Chang
Florida

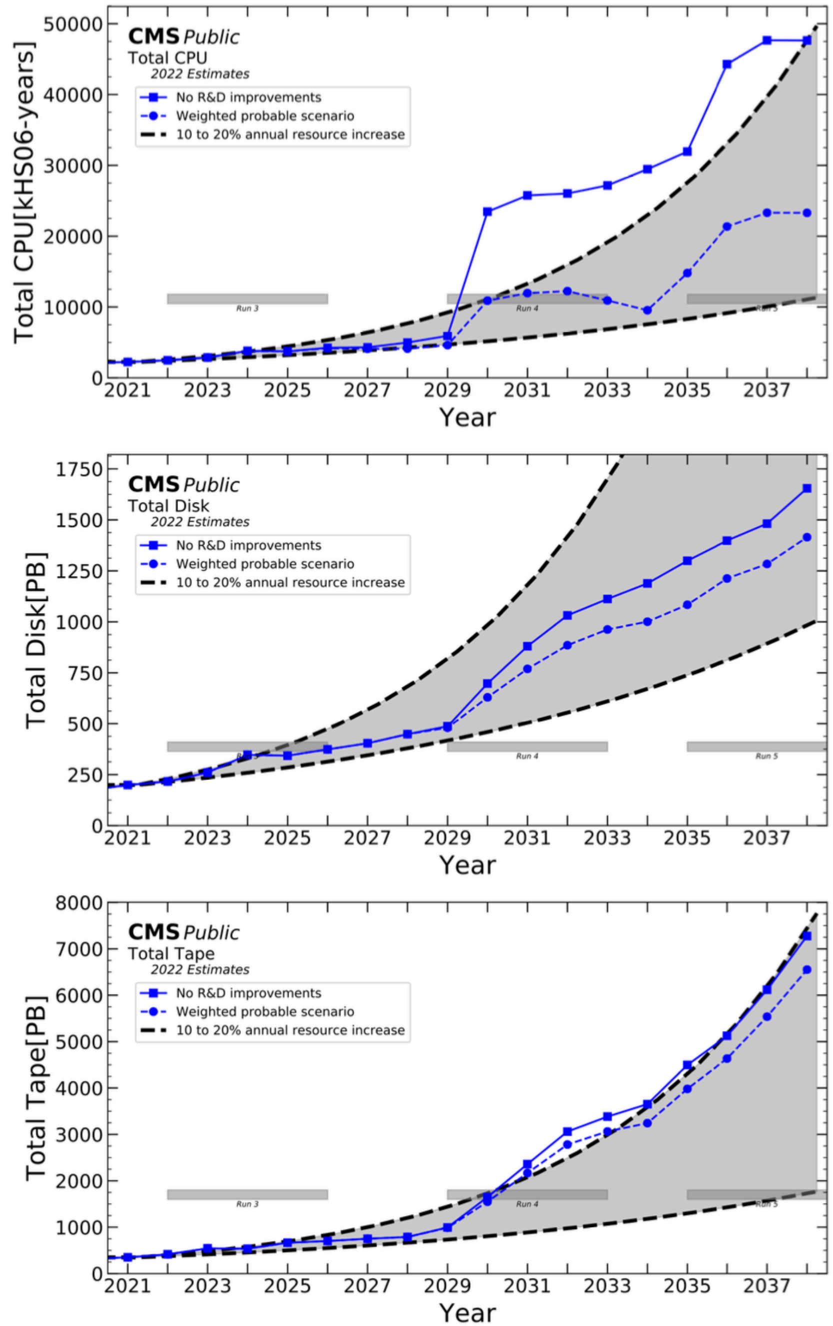
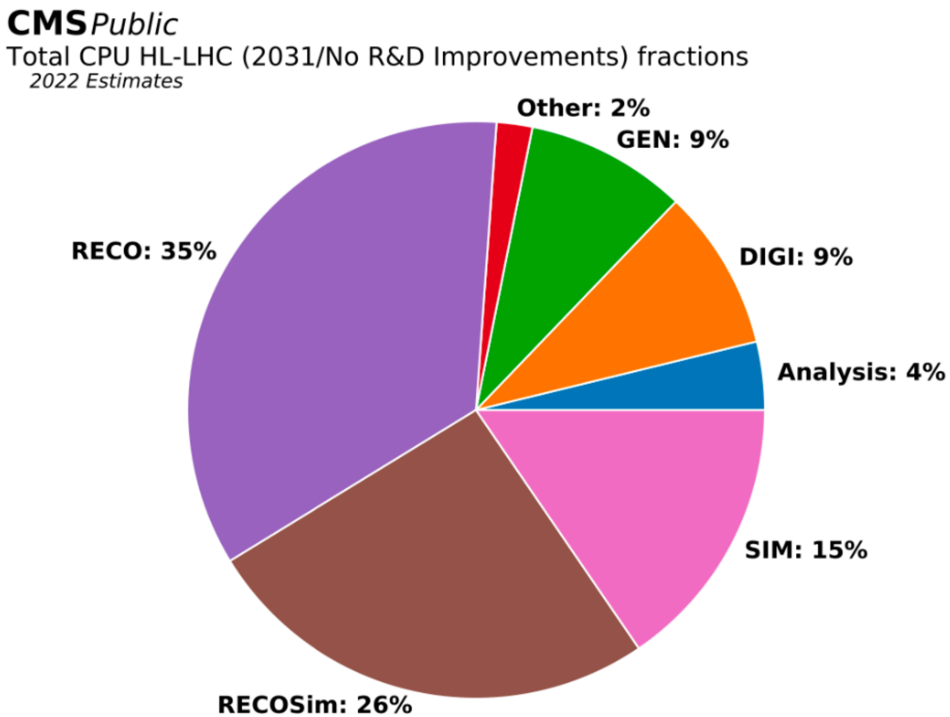
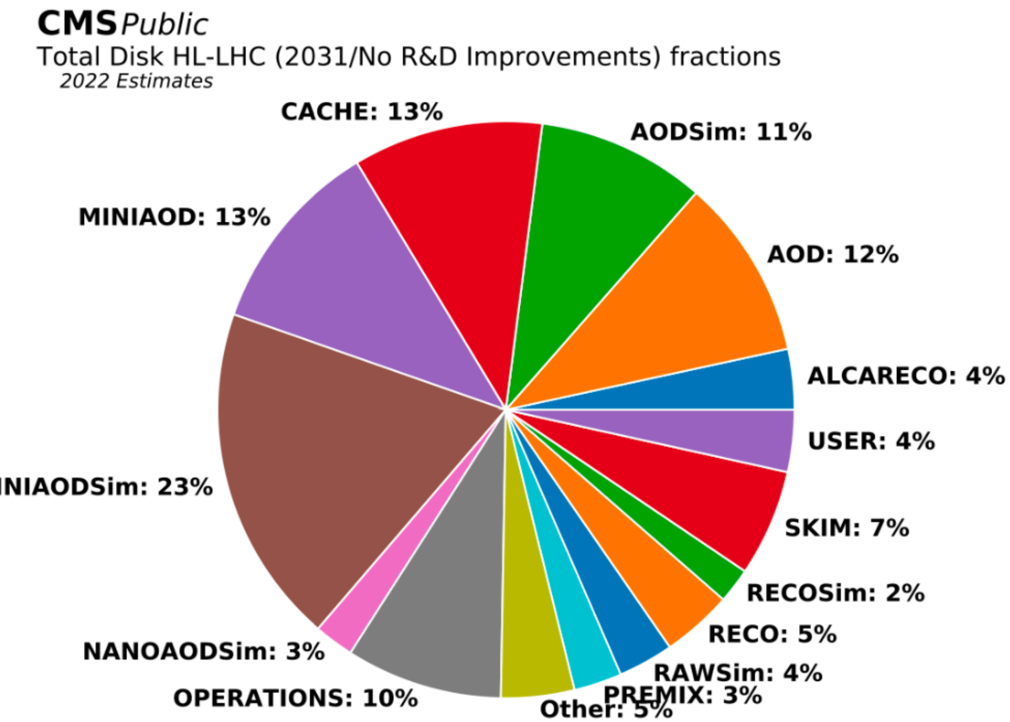
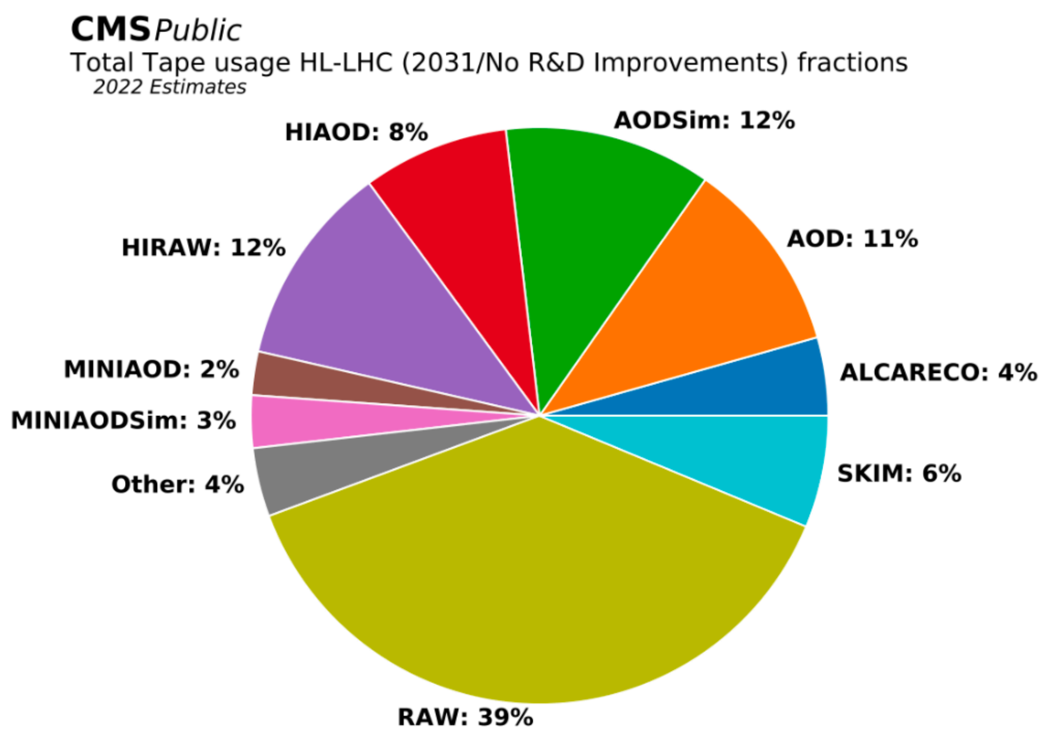


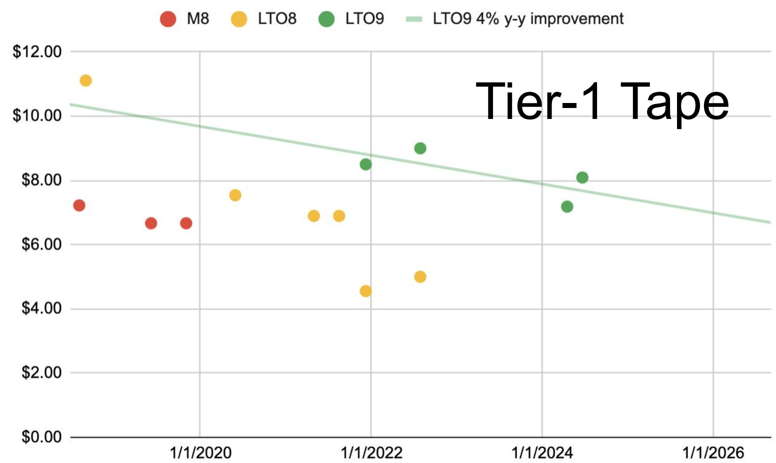
Figure 9: Updated projections of needed CPU, disk and tape needs into HL-LHC. On each plot a gray band represents the projected capacity of the resource within flat budget. Two lines are drawn, each corresponding to one of the two scenarios considered, *Baseline* and *Weighted Probable* (dashed line). The latter incorporates the improvements summarized in Table 16. The effect of GPUs is not represented in these plots. The tape projected needs increases almost linearly driven by the RAW data stored. In the legends, the *Baseline* scenario is described as "No R&D improvement" and the *Weighted Probable* scenario as "R&D most probable outcome".

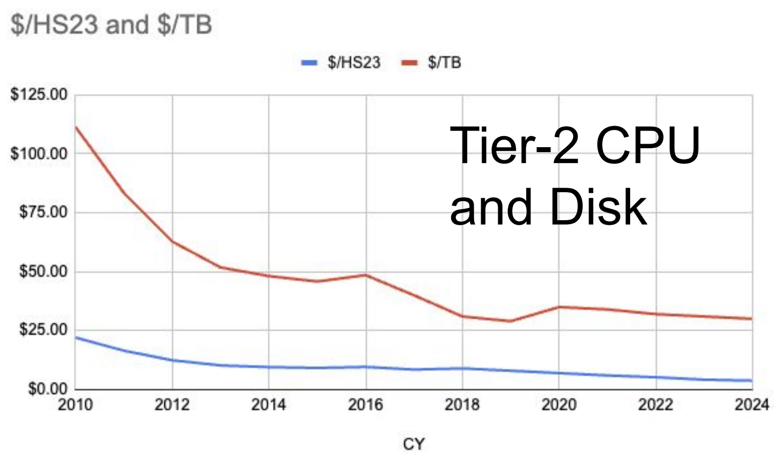

CMS Public
Total CPU HL-LHC (2031/No R&D Improvements) fractions
2022 Estimates
Other: 2%
GEN: 9%
DIGI: 9%
Analysis: 4%
SIM: 15%
RECOSim: 26%
RECO: 35%


CMS Public
Total Disk HL-LHC (2031/No R&D Improvements) fractions
2022 Estimates
CACHE: 13%
AODSim: 11%
AOD: 12%
ALCARECO: 4%
USER: 4%
SKIM: 7%
RECOSim: 2%
RECO: 5%
RAWSim: 4%
PREMIX: 3%
Other: 5%
OPERATIONS: 10%
NANOAODSim: 3%
MINIAODSim: 23%
MINIAOD: 13%

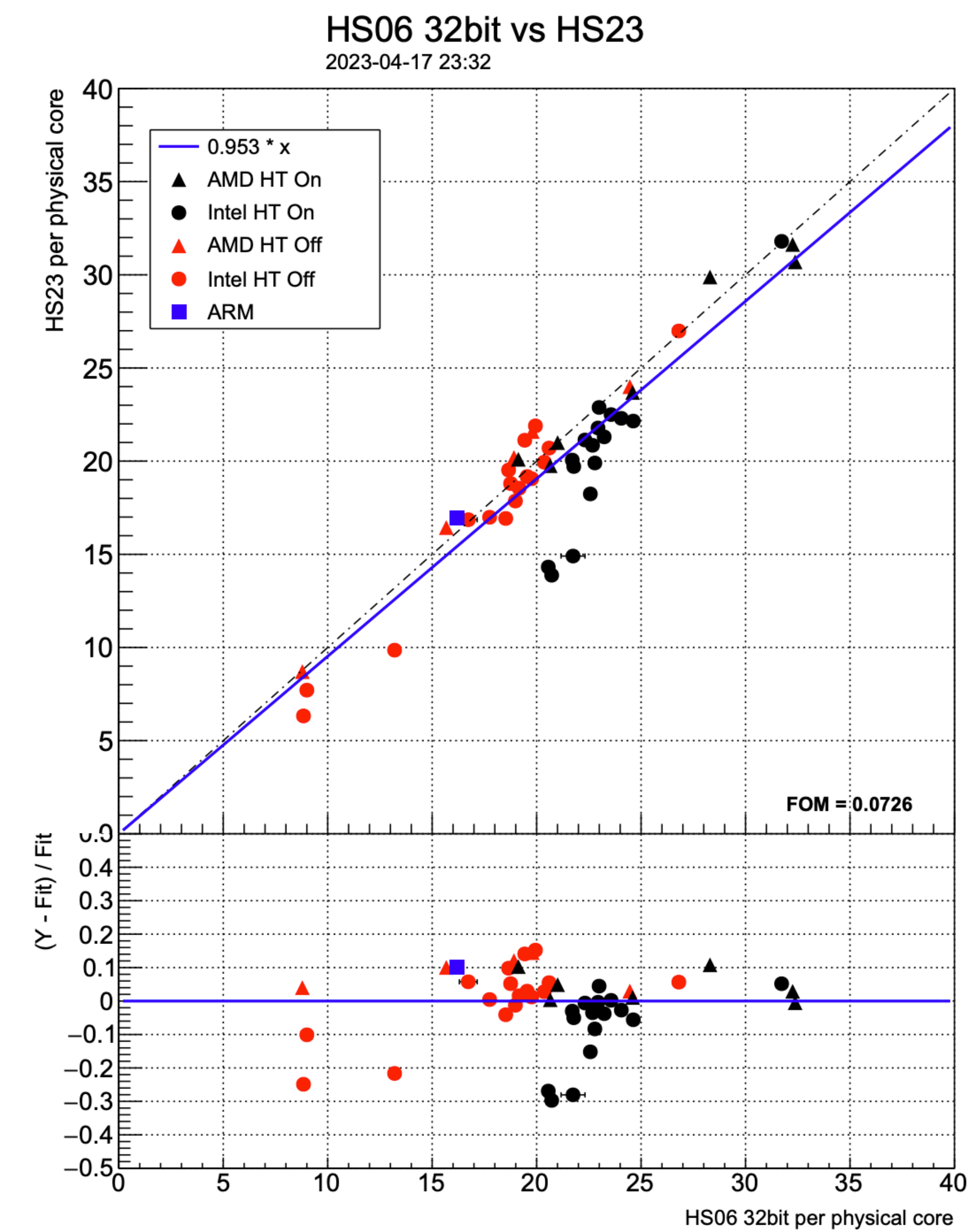

CMS Public
Total Tape usage HL-LHC (2031/No R&D Improvements) fractions
2022 Estimates
HIAOD: 8%
AODSim: 12%
AOD: 11%
ALCARECO: 4%
SKIM: 6%
RAW: 39%
Other: 4%
MINIAODSim: 3%
MINIAOD: 2%
HIRAW: 12%

Hardware cost evolution tracking





- Tier-1 tape
  - Tape media $/TB through 2024
  - Different media types color coded
  - M8 media was a temporary reformatting of LTO7 due to LTO8 unavailability in 2019-20.
    - Unlike LTO8, M8 media is unreadable by LTO9 drives, motivating early migration
  - Cost improvement of LTO9 media is slower than historical rates
    - This spring LTO9 media cost INCREASED by about 15%
    - (LTO8 increased much more)
- Tier-2 CPU and disk
  - Same trends as for Tier-1 CPU and disk



UF
Chang
Florida

Snowmass Recommendations

1. **Efficiently exploit specialized compute architectures and systems**. To achieve this will require the allocation of dedicated facilities to specific processing steps in the HEP workflows, in particular for "analysis facilities" (Sections II and V); designing effective benchmarks to exploit AI hardware (Section III); improved network visibility and interaction (Section VII); and enhancements to I/O libraries such as lossy compression and custom delivery of data (Section IV).

2. **Invest in portable and reproducible software and computing solutions to allow exploitation of diverse facilities.** The need for portable software libraries, abstractions and programming models is recognized across all the topics discussed here, and is especially called out in Processing (Section II), AI Hardware (Section III) and Storage (Section IV). Software frameworks to enable reproducible HEP workflows are also greatly needed (Sections V and VI).

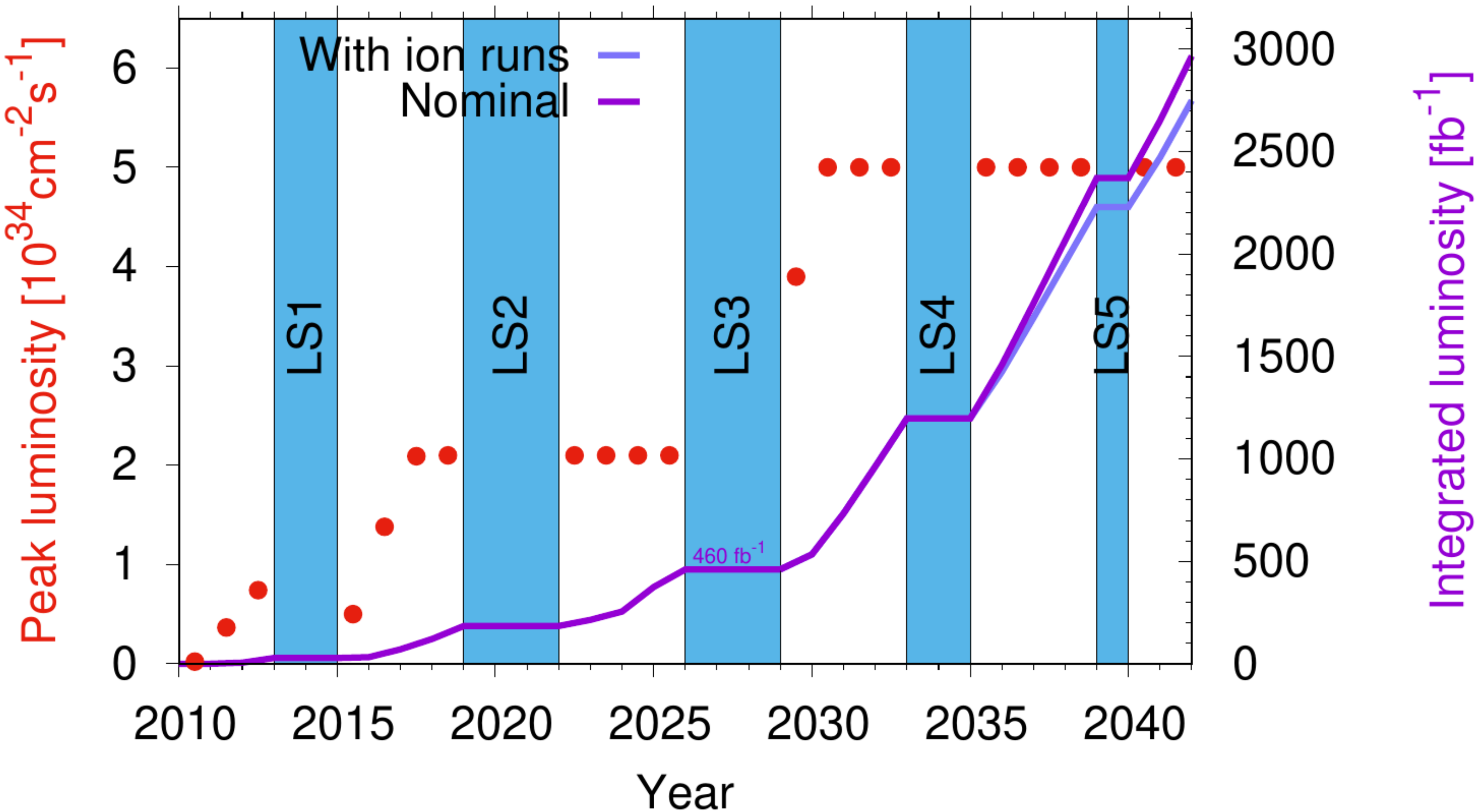
3. **Embrace disaggregation of systems and facilities**. The HEP community will need to embrace heterogeneous resources on different nodes, systems and facilities and effectively balance these accelerated resources to match workflows. To do so will require software abstraction to integrate accelerators, such as those for AI (Section III); orchestration of network resources (VII); exploiting computational storage (Section IV); as well as exploiting system rack-level dis-aggregation technology if adopted at computing centers.

4. **Extend common interfaces to diverse facilities**. In order to scalably exploit resources wherever they are available, HEP must continue to encourage edge-service platforms on dedicated facilities as well as Cloud and HPC (Section VI), develop portable edge-services that are re-usable by other HEP projects, and exploit commonality within

HEP and other sciences (Section VI). These interfaces will also need to extend into all aspects of HEP workflows, including data management and optimizing data movement (Sections VII, II and IV), as well as the deployment of compute resources for analysis facilities (Section V).

Title

UF
Chang
Florida



Title

U.S. CMS Operations Program

Quick program budget overview

- Software and Computing is the single largest area in the budget.
  - ~Half of S&C is equipment and operation of Tier-1 and Tier-2 facilities.
- Common Cost is set by our ~30% PhD headcount in CMS.
- Role of Risk Contingency and Management Reserve to be discussed in later presentations.
- Personnel support is for engineers, technical staff, computing professionals, not scientists.
  - We do provide travel/COLA support to scientists who provide Operations Program deliverables.