PART 2

SMARTHEP

REAL-TIME ANALYSIS FOR
SCIENCE AND INDUSTRY

What will this lecture be about?

Introduction

- Definitions and basic concepts

Input to the physics

- The data: **trigger**, data preparation
- The theory: Monte carlo simulations
- Reconstruction, or how to translate detector signals to particles

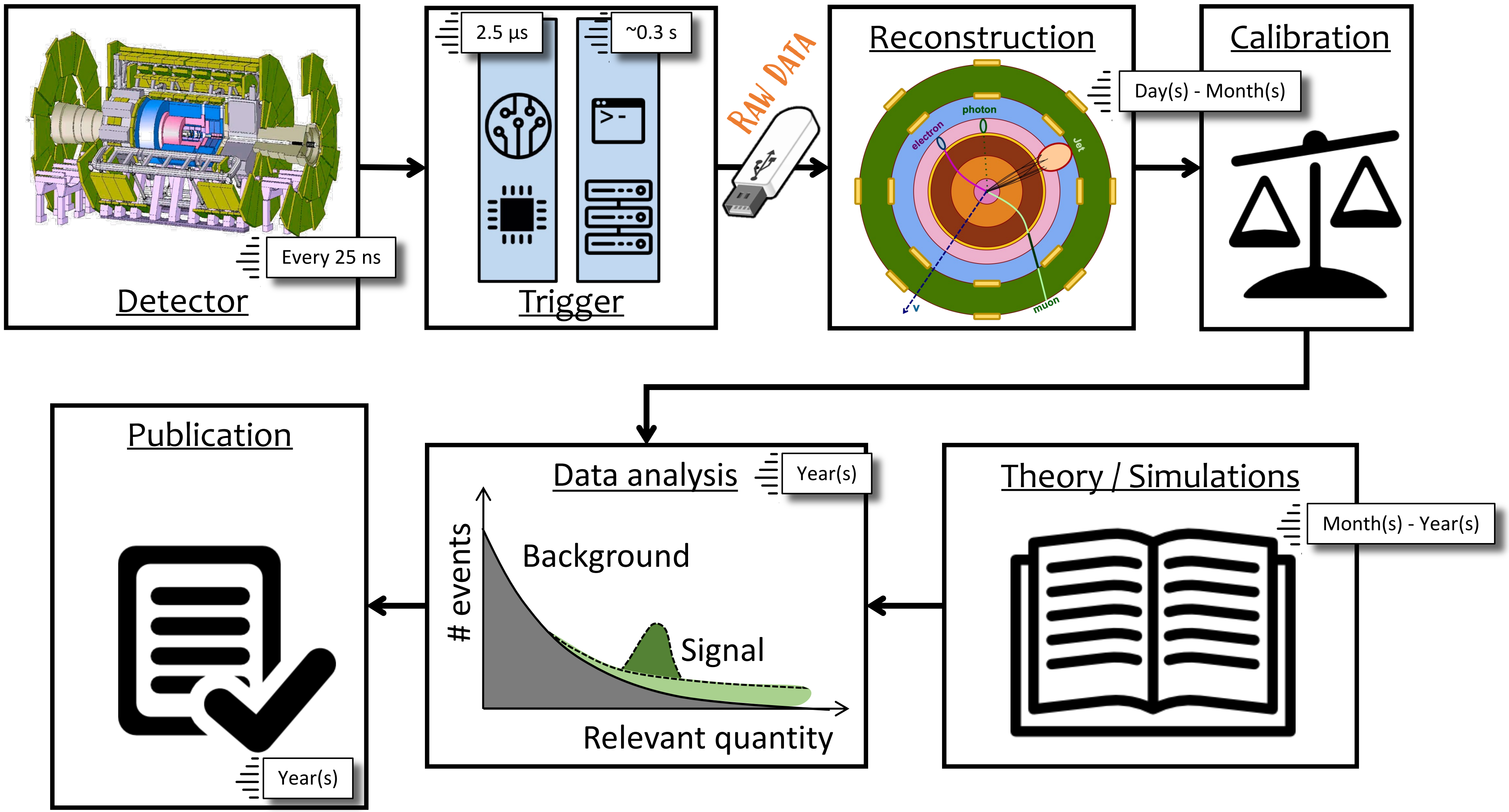
Physics analyses

- Through example, step-by-step
- Discussion of analysis methods

SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

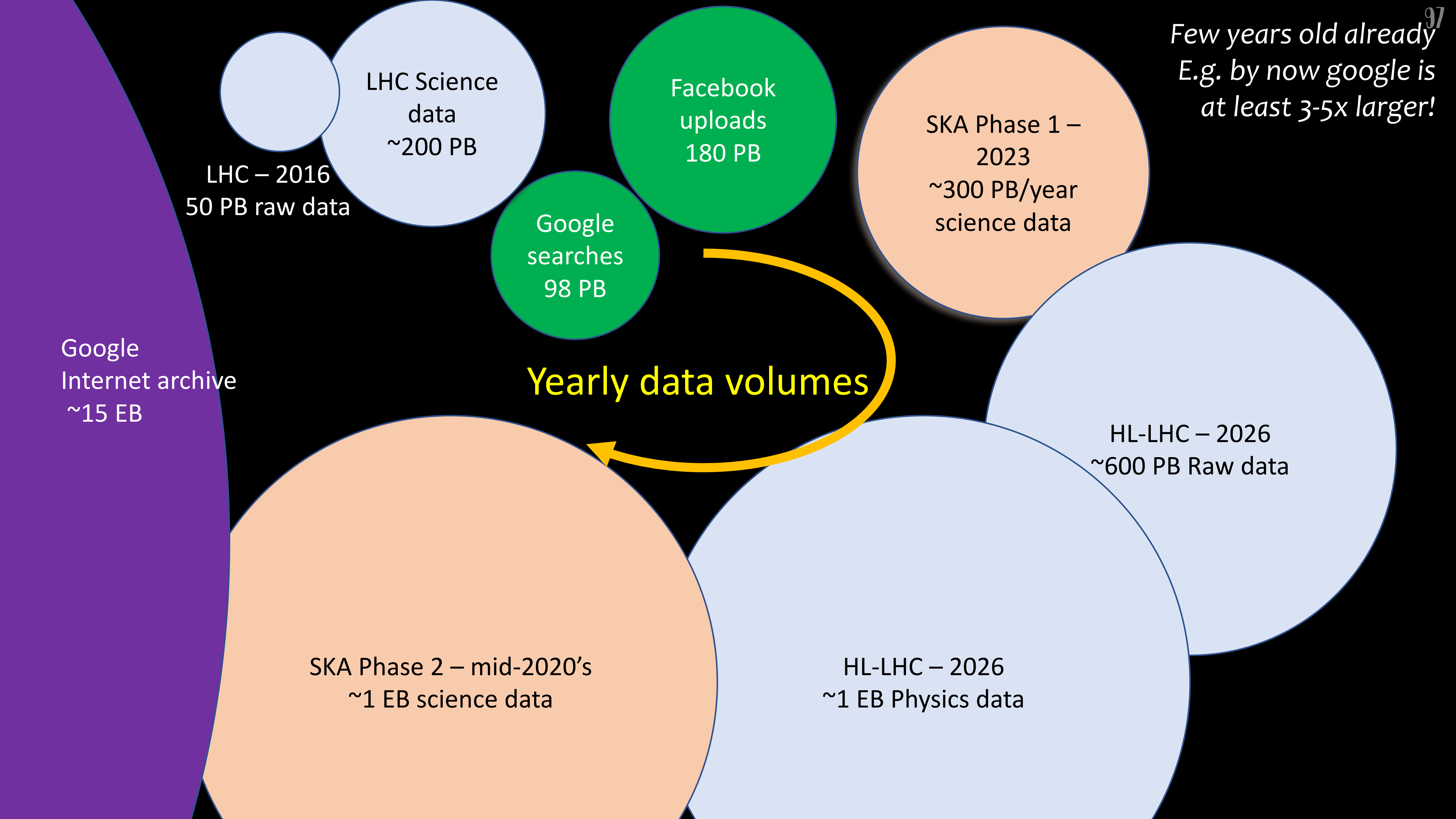
Is there a topic you would like to add to this material?
If so: please let me know at the end of this lecture and I will see if I can add it!

The lifetime of a collision event



Data Preparation

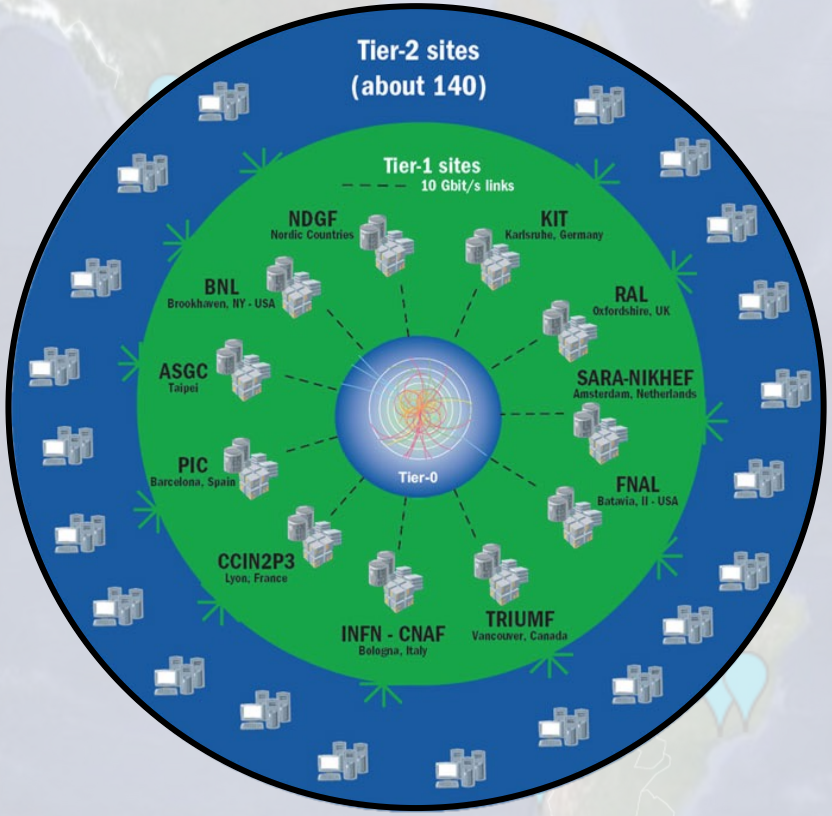

Let's first talk about
Computing!


Few years old already
E.g. by now google is
at least 3-5x larger!
LHC Science
data
~200 PB
LHC – 2016
50 PB raw data
Facebook
uploads
180 PB
Google
searches
98 PB
SKA Phase 1 –
2023
~300 PB/year
science data
Yearly data volumes
Google
Internet archive
~15 EB
HL-LHC – 2026
~600 PB Raw data
SKA Phase 2 – mid-2020's
~1 EB science data
HL-LHC – 2026
~1 EB Physics data

Worldwide LHC Computing Grid

an international collaboration to distribute and analyse LHC data

Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists.

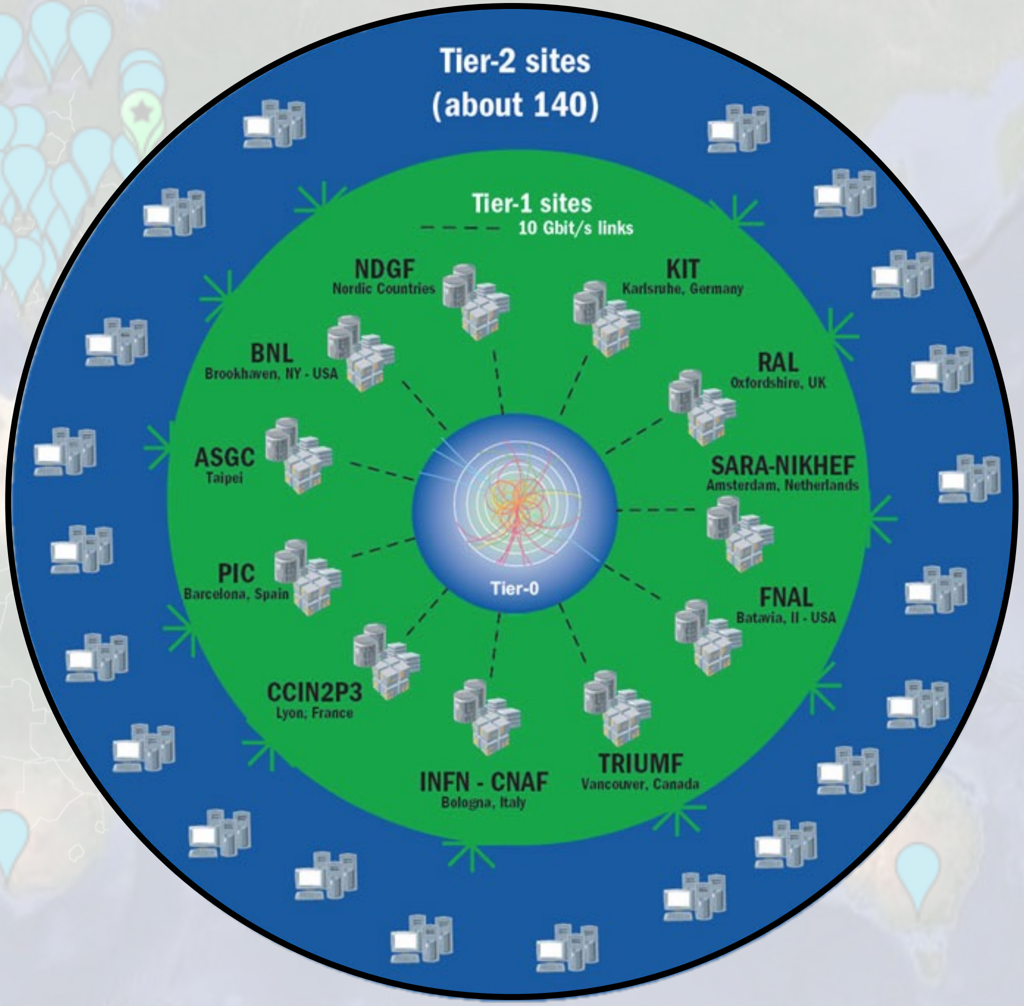


- 161 sites, 42 countries
- 1 M CPU cores
- 1 EB of storage
- > 2 M jobs/day
- > 100 PB moved/month
- accessed by 10k users
- 10-100 Gb links

Network proved better than anyone imagined: Any job can run anywhere

Worldwide LHC Computing Grid
The Tier System

- Tier-0 (CERN):
  - Data recording, reconstruction and distribution
- Tier-1:
  - Permanent storage, re-processing, analysis
- Tier-2:
  - Simulation, end-user analysis



ATLAS DATA MANAGEMENT

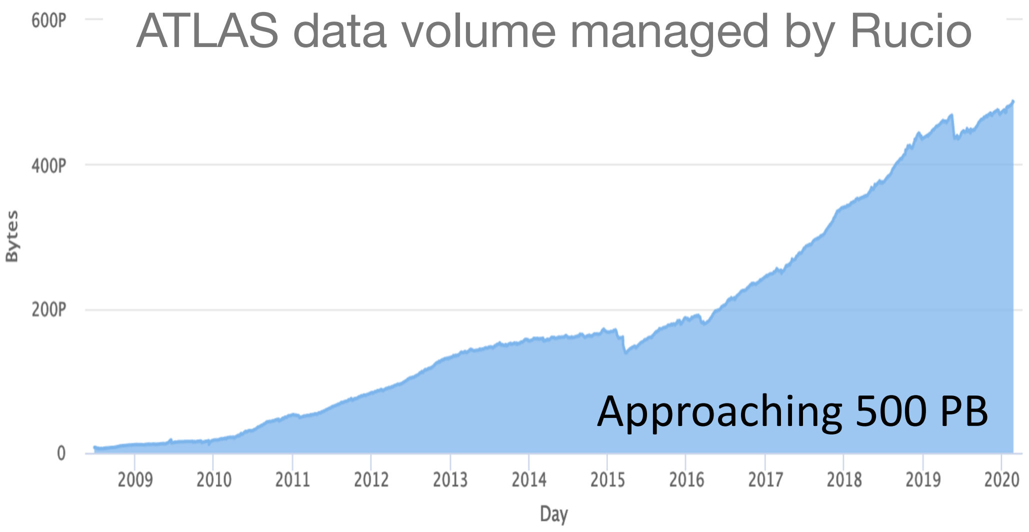
RUCIO

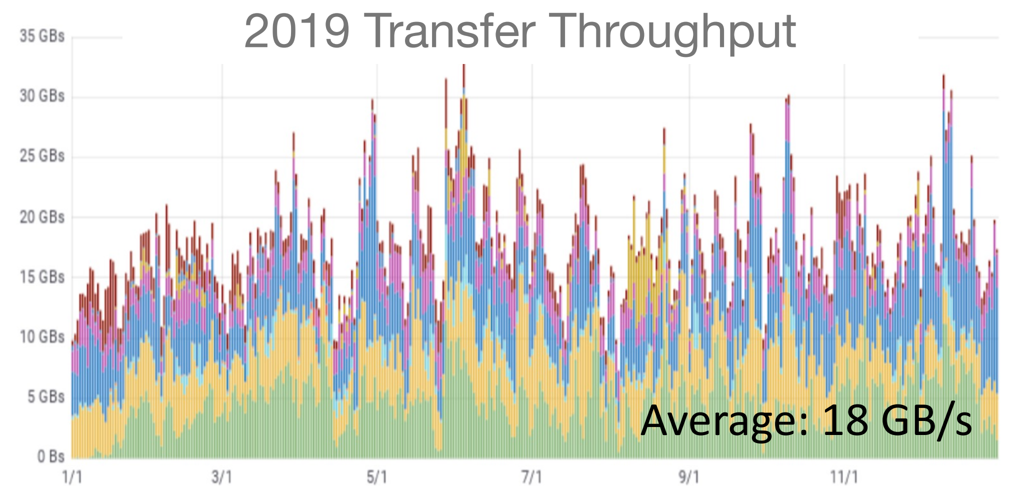
- Data storage
- Access
- Replication
- Deletion

- Scalable
- Policy-driven
- Monitorable
- Supporting "FAIR" data principles

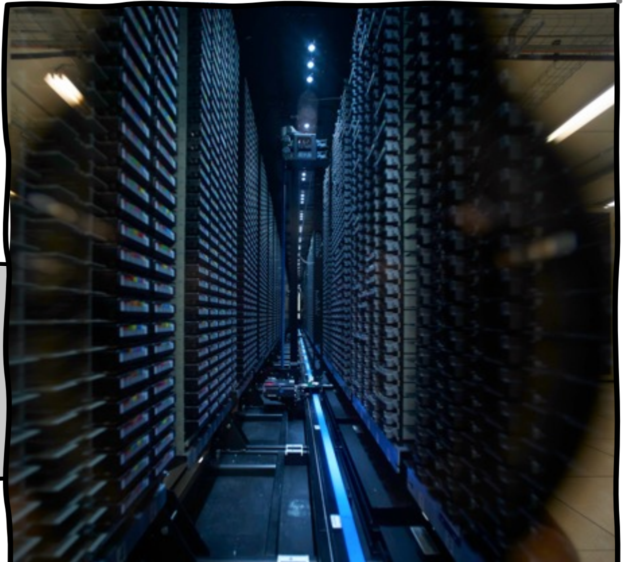
Findable
Accessible
Interoperable
Reusable

Now established in the HEP community and beyond





Hardware


Magnetic tapes, retrieved by robotic arms, are used for long-term storage

Storage

Tape (at CERN) about 270 PB
- Most reliable and cost-effective technology for large-scale archiving
- Data stored there infinitely

Disk about 200 PB
- Data for initial processing
- Copies for further processing / user analysis
- Data in disks gets staged from tape, on demand

Processing power

CPUs
- Mainly GRID
- About 400k cores

GPUs
- Mostly for RnD
- Few 10s

Also considering for the future: FPGA accelerators

Opportunistic resources
- Online farm, 100k cores
- High Performance Computers, primarily in the US
- Volunteer computing

Nvidia GeForce

LHC@home Volunteer computing for the LHC

SOFTWARE

athena ⊕
Project ID: 53790

70,356 Commits · 34 Branches · 1,374 Tags · 2.6 GB Files · 2.6 GB Storage · 124 Releases

The ATLAS Experiment's main offline software repository

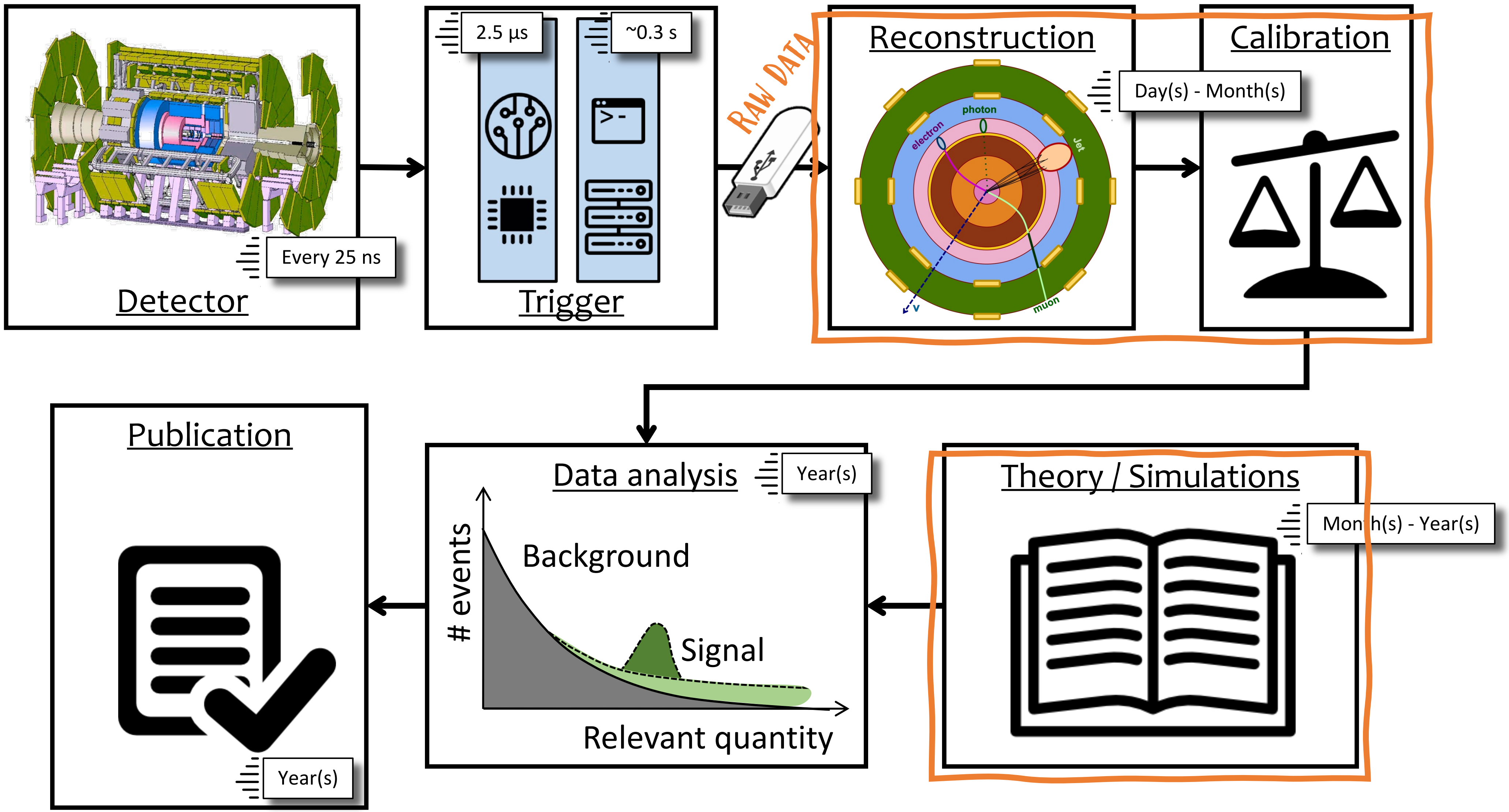
- **All software organized in packages in Git.** For example:
  https://gitlab.cern.ch/atlas/athena
- **All software open source, copyrighted and licenced (Apache 2)**
  - "Copyright (C) 2002-2020 CERN for the benefit of the ATLAS collaboration"
  - For open use – but also for crediting developers who move out of academia
- **Thorough tracking of software developments a key of success**
  - Via the Jira software, supported by CERN IT Jira Software
  - Multiple releases exist for merging of new code with existing one
  - Automated tools run nightly to verify code sanity & performance
  - Globally the software projects are coordinated with careful planning
- **Software Tools**
  - Databases
  - Analysis tools: ROOT is the workhorse!

ROOT
Data Analysis Framework

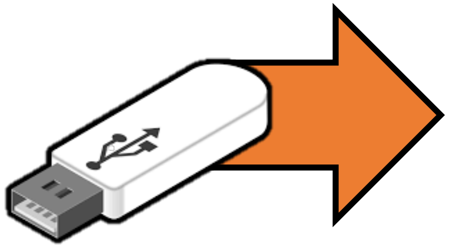
- **Analysis-specific software developed by teams available to whole collaboration!**

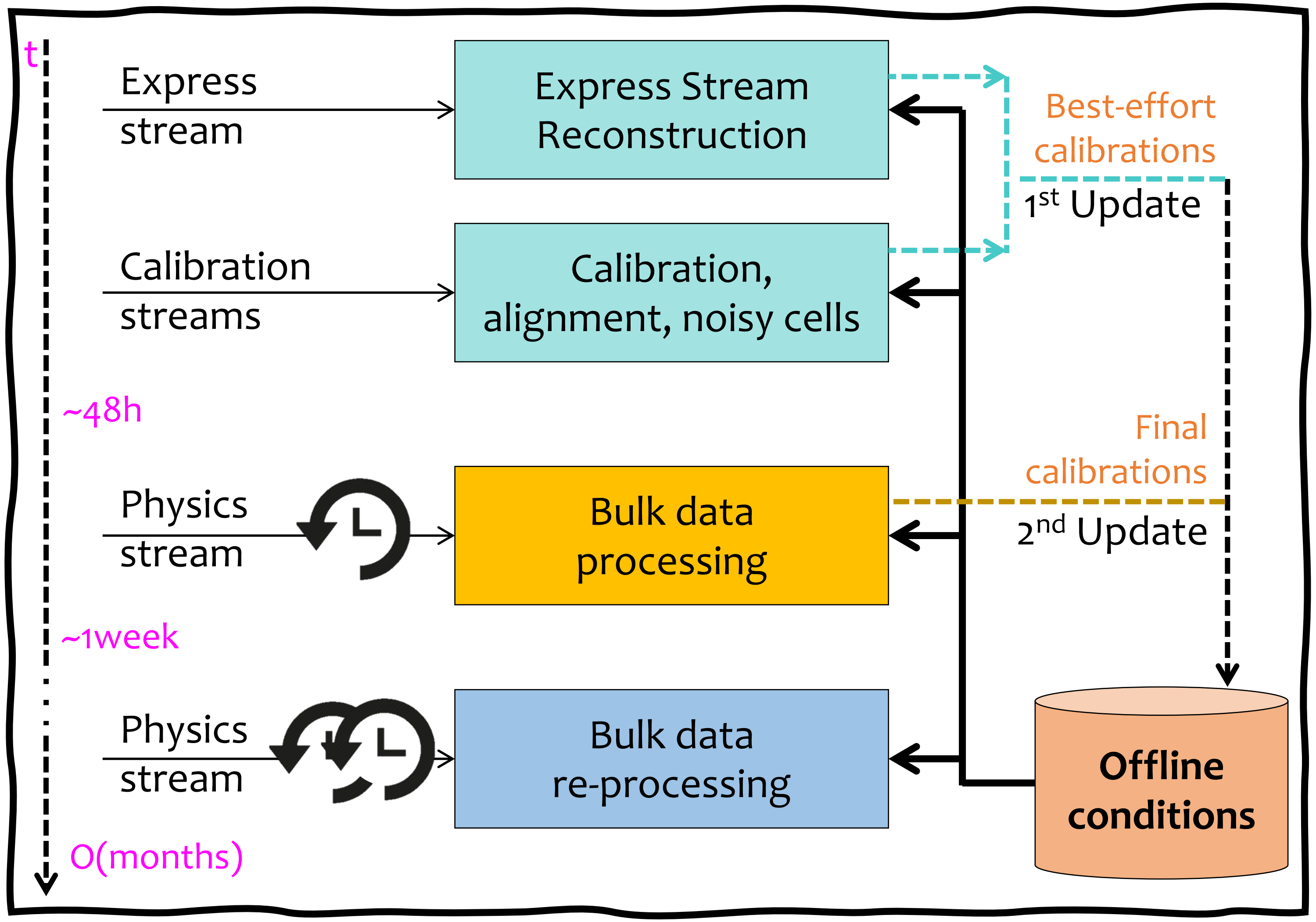
Data Preparation

The lifetime of a collision event

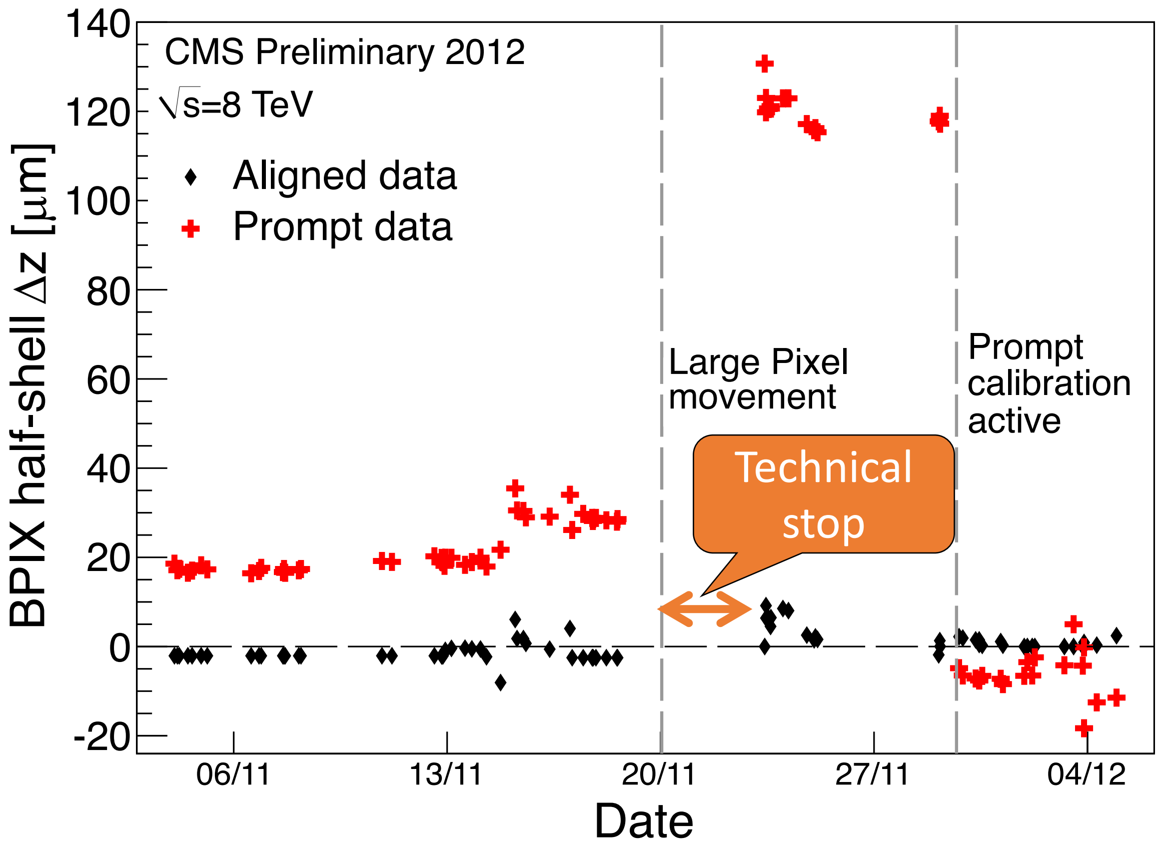


The event at Tier-0





E.G. ALIGNMENT

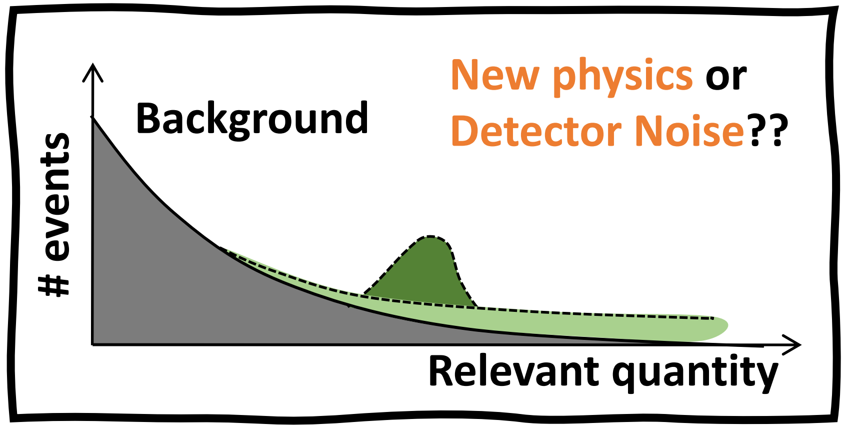


Day-by-day value of the relative longitudinal shift between the two half-shells of the BPIX as measured with the primary vertex residuals, for the last month of pp data taking in 2012.

DATA QUALITY

✓ The data we analyze have to follow norms of quality such that our results are trustable.

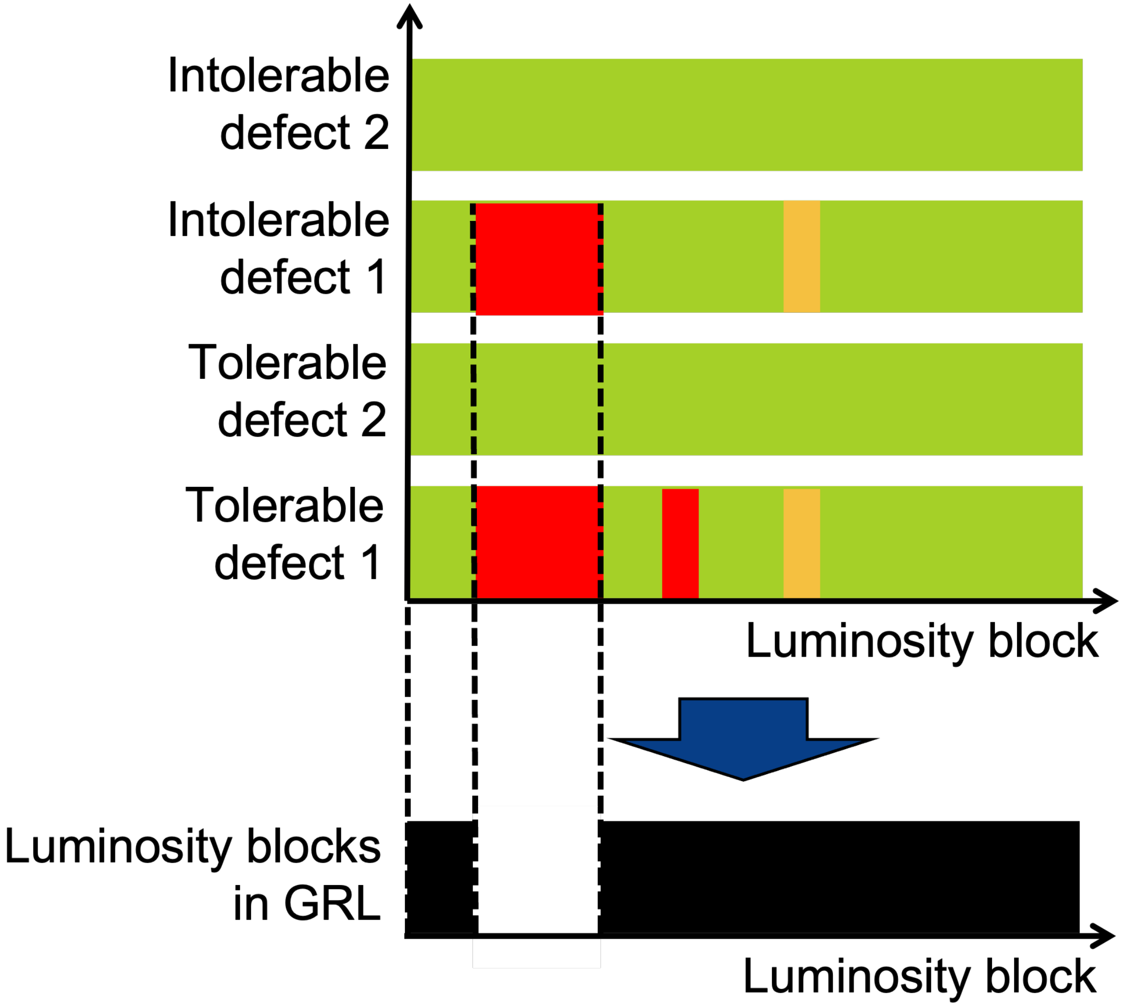
- **Online:** Fast monitoring of detector performance during data taking, using dedicated stream, "express stream".
- **Offline:** More thorough monitoring at two instances:
  - Express reconstruction; fast turn-around.
  - Prompt reconstruction: larger statistics.



- **What is monitored?**
  - Noise in the detector.
  - Reconstruction (tracks, clusters, combined objects, resolution and efficiency).
  - Input rate of physics.
  - All compared to reference histograms of data that has been validated as "good".

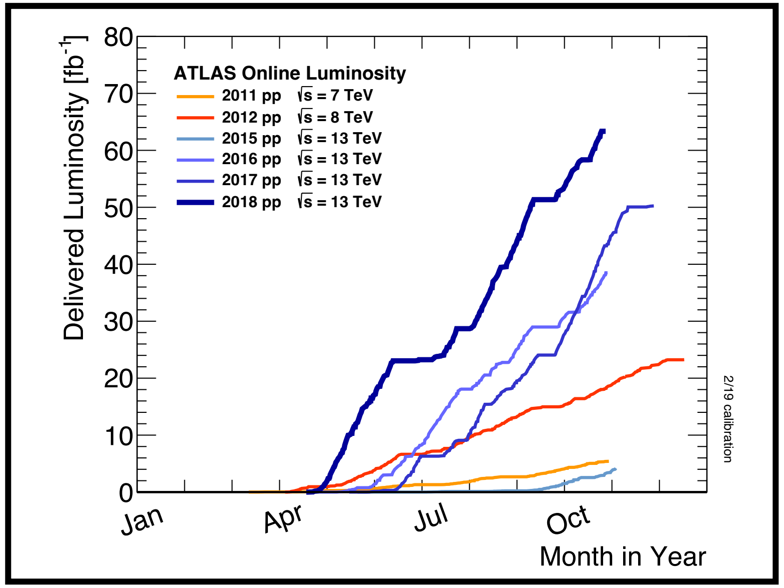
Data Quality and "GRL"

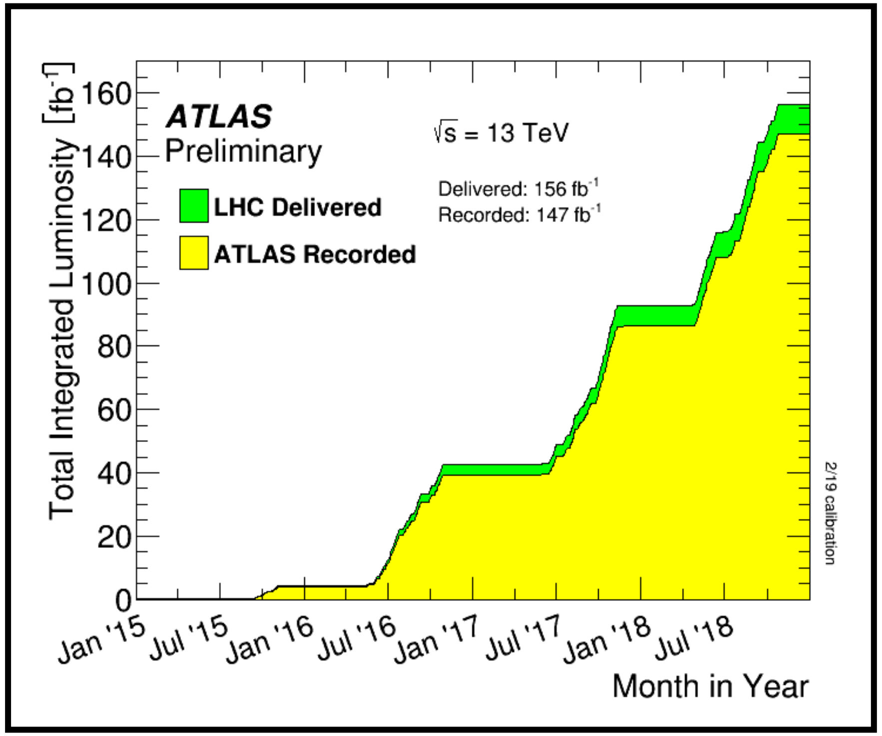
Good Run List

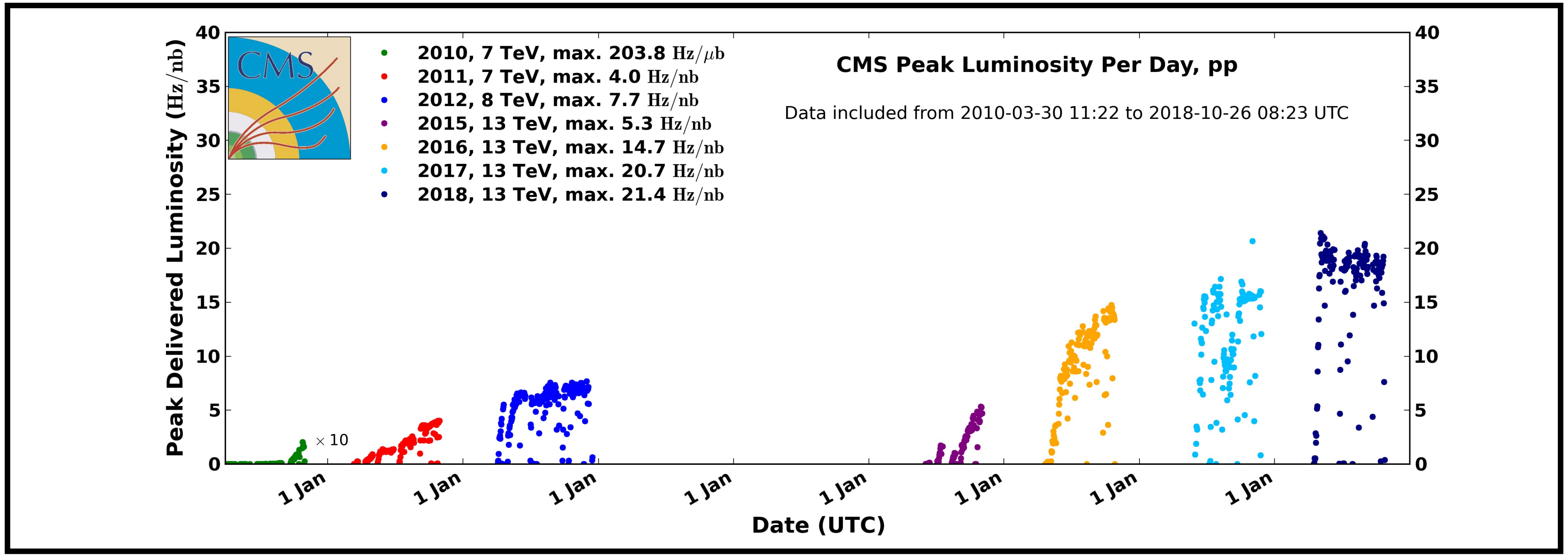


Short period during which data taking conditions are (expected to be) absolutely stable. Used for data-quality assessment and luminosity determination

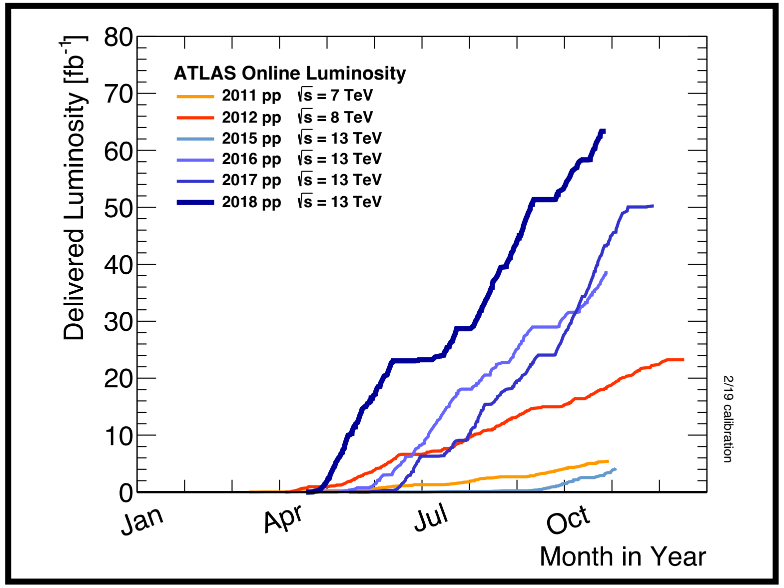
LUMINOSITY

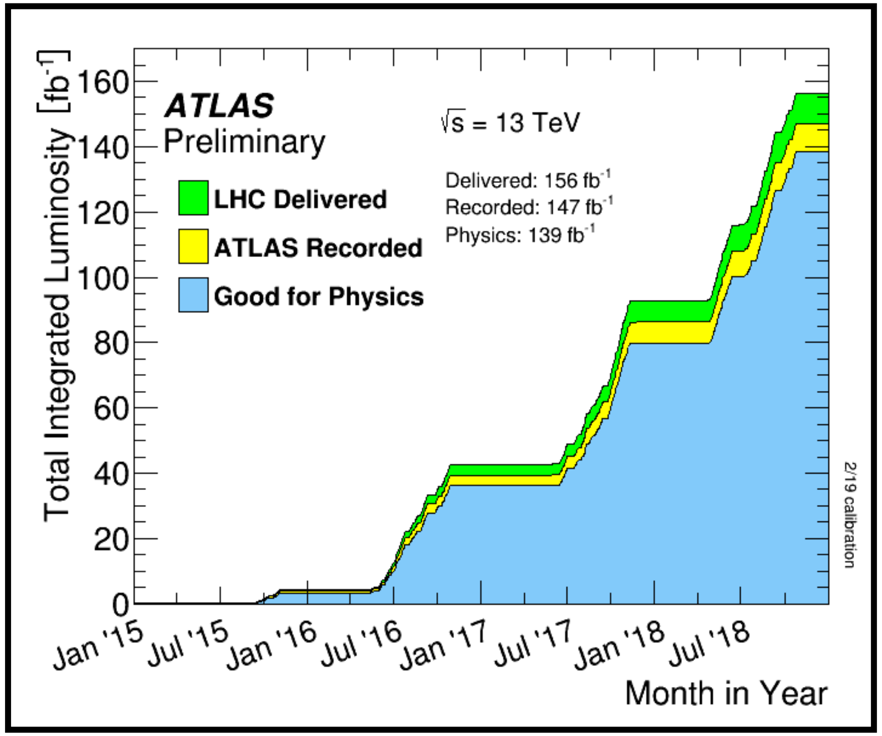

ATLAS Online Luminosity
2011 pp √s = 7 TeV
2012 pp √s = 8 TeV
2015 pp √s = 13 TeV
2016 pp √s = 13 TeV
2017 pp √s = 13 TeV
2018 pp √s = 13 TeV
Delivered Luminosity [fb-1]
Jan
Apr
Jul
Oct
Month in Year
2/19 calibration

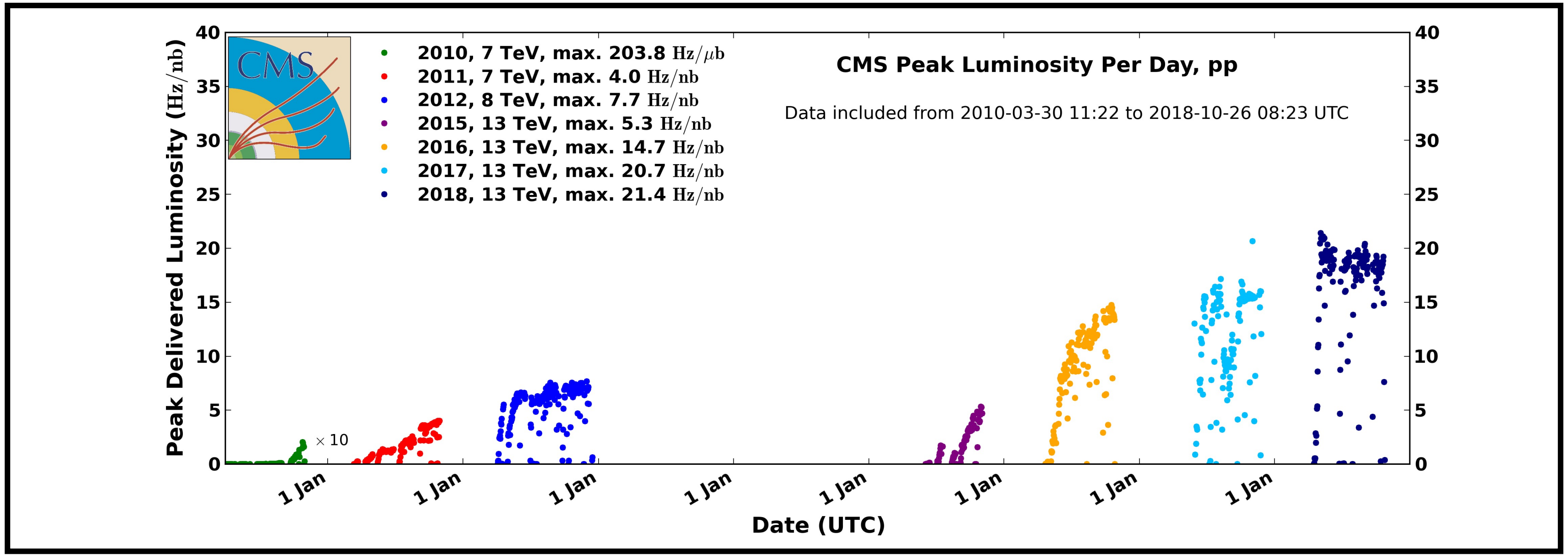

ATLAS
Preliminary
√s = 13 TeV
LHC Delivered
ATLAS Recorded
Delivered: 156 fb-1
Recorded: 147 fb-1
Total Integrated Luminosity [fb-1]
Jan '15
Jul '15
Jan '16
Jul '16
Jan '17
Jul '17
Jan '18
Jul '18
Month in Year
2/19 calibration


CMS Peak Luminosity Per Day, pp
Data included from 2010-03-30 11:22 to 2018-10-26 08:23 UTC
2010, 7 TeV, max. 203.8 Hz/μb
2011, 7 TeV, max. 4.0 Hz/nb
2012, 8 TeV, max. 7.7 Hz/nb
2015, 13 TeV, max. 5.3 Hz/nb
2016, 13 TeV, max. 14.7 Hz/nb
2017, 13 TeV, max. 20.7 Hz/nb
2018, 13 TeV, max. 21.4 Hz/nb
Peak Delivered Luminosity (Hz/nb)
× 10
1 Jan
Date (UTC)

LUMINOSITY


ATLAS Online Luminosity
2011 pp √s = 7 TeV
2012 pp √s = 8 TeV
2015 pp √s = 13 TeV
2016 pp √s = 13 TeV
2017 pp √s = 13 TeV
2018 pp √s = 13 TeV
Delivered Luminosity [fb-1]
0
10
20
30
40
50
60
70
80
Jan
Apr
Jul
Oct
Month in Year
2/19 calibration


ATLAS
Preliminary
√s = 13 TeV
LHC Delivered
ATLAS Recorded
Good for Physics
Delivered: 156 fb-1
Recorded: 147 fb-1
Physics: 139 fb-1
Total Integrated Luminosity [fb-1]
0
20
40
60
80
100
120
140
160
Jan '15
Jul '15
Jan '16
Jul '16
Jan '17
Jul '17
Jan '18
Jul '18
Month in Year
2/19 calibration


CMS
2010, 7 TeV, max. 203.8 Hz/μb
2011, 7 TeV, max. 4.0 Hz/nb
2012, 8 TeV, max. 7.7 Hz/nb
2015, 13 TeV, max. 5.3 Hz/nb
2016, 13 TeV, max. 14.7 Hz/nb
2017, 13 TeV, max. 20.7 Hz/nb
2018, 13 TeV, max. 21.4 Hz/nb
CMS Peak Luminosity Per Day, pp
Data included from 2010-03-30 11:22 to 2018-10-26 08:23 UTC
Peak Delivered Luminosity (Hz/nb)
0
5
10
15
20
25
30
35
40
× 10
1 Jan
1 Jan
1 Jan
1 Jan
1 Jan
1 Jan
1 Jan
1 Jan
Date (UTC)

Luminosity – the figure of merit

Intensity per bunch

Number of bunches

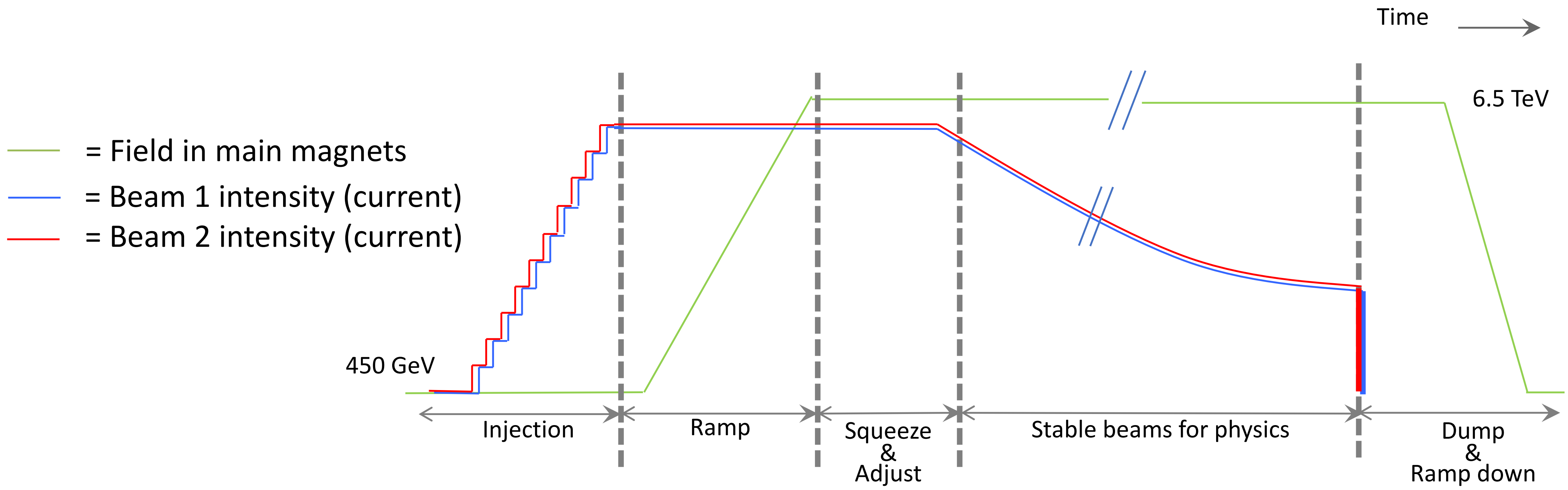
Revolution frequency

$$\mathcal{L} = \frac{k N_1 N_2 f}{4\pi \sigma_x \sigma_y}$$

Beam dimensions

More of less fixed parameters: Revolution frequency and Number of bunches

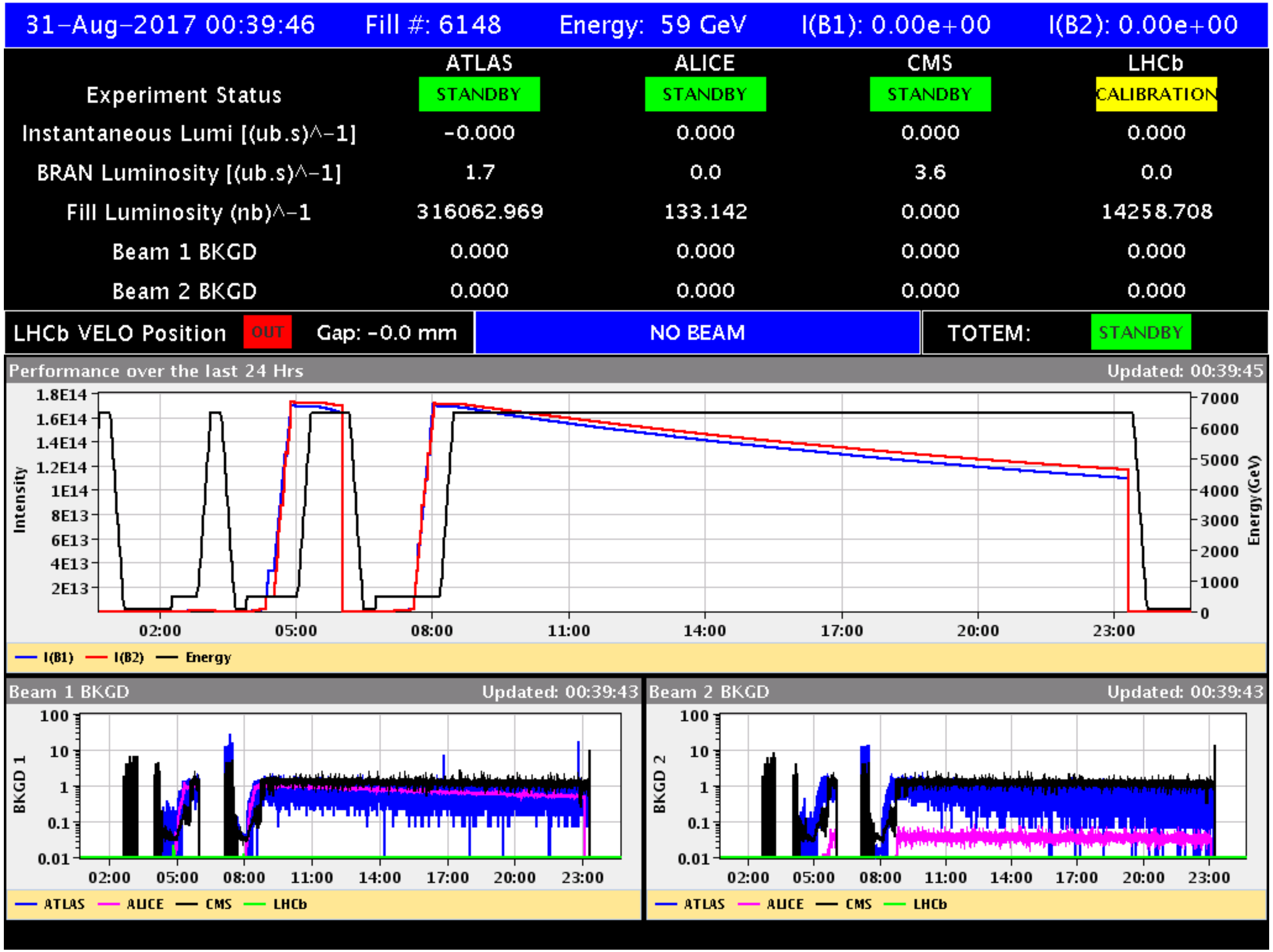
Luminosity – the figure of merit



- The LHC is built to collide protons at 7 TeV per beam, which is **14 TeV centre of Mass**
- In 2012 it ran at 4 TeV per beam, 8 TeV c.o.m.
- Since 2015 it runs at 6.5 TeV per beam, 13 TeV c.o.m
- In Run 3, starting this year, it will run at 6.8 TeV per beam, 13.6 TeV c.o.m

Why not 14 TeV?

Figure from R. Steerenberg

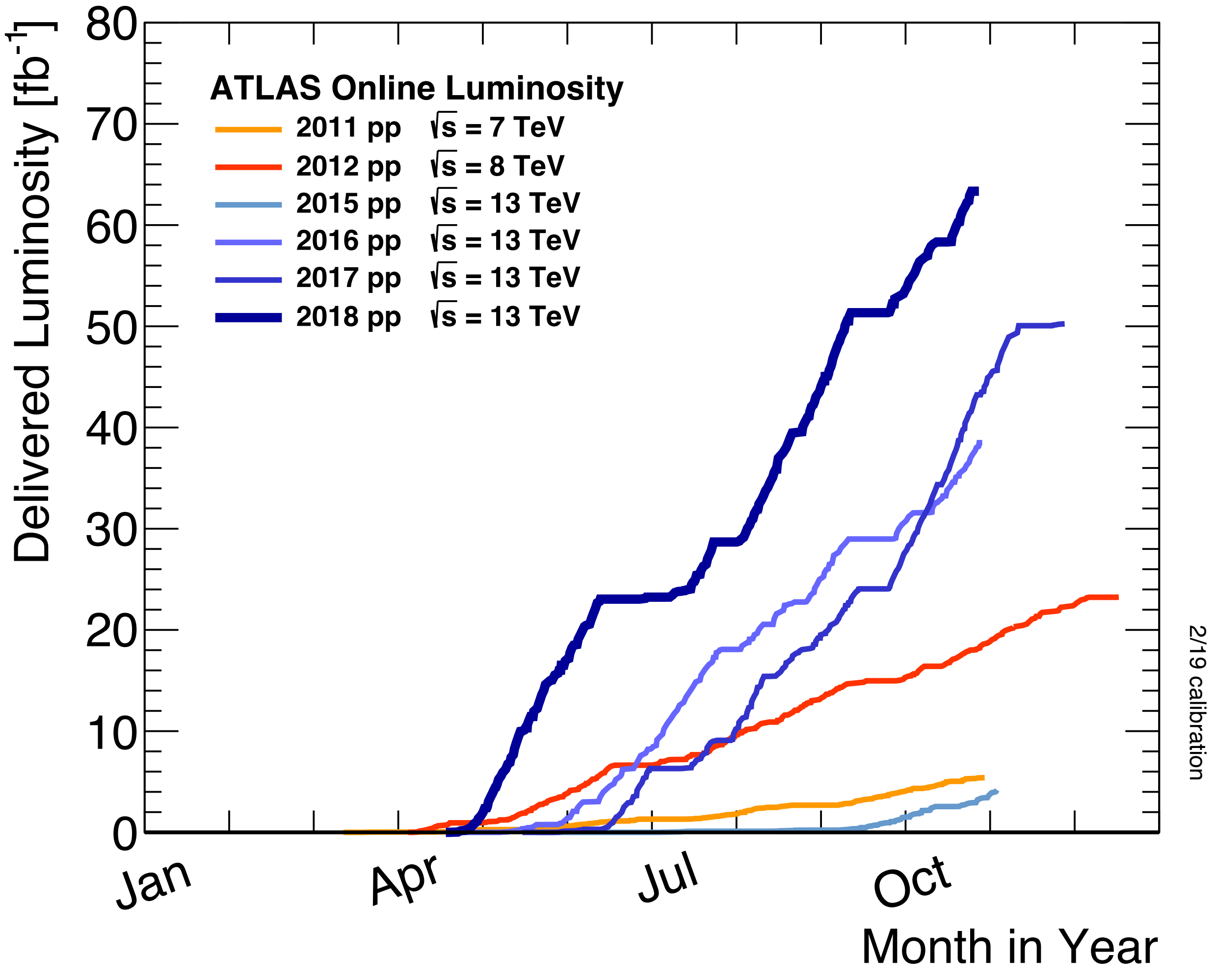


URL: https://op-webtools.web.cern.ch/vistar/vistars.php?usr=LHC1

Luminosity – the figure of merit

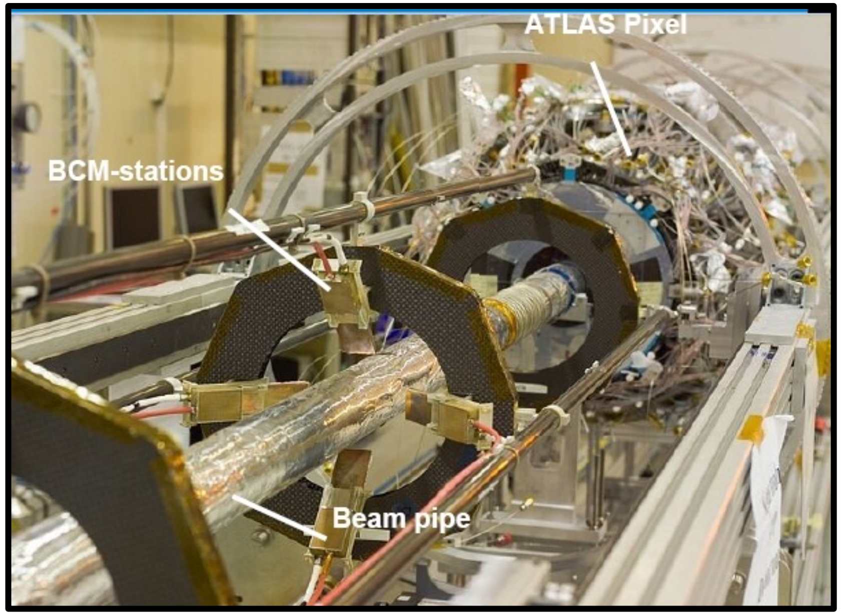
$$L = \int \mathcal{L} dt$$

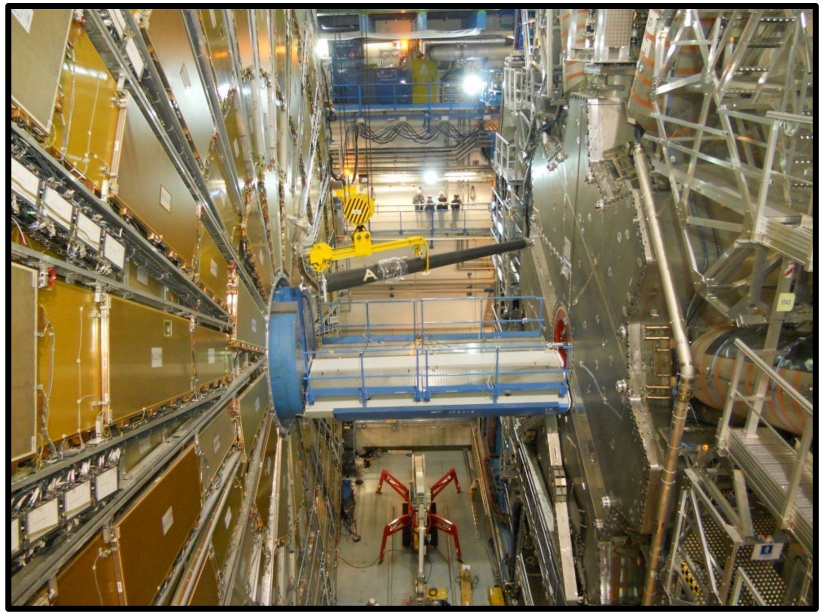
$$\sigma = \frac{\text{N events}}{L}$$



Luminosity Determination "Figure of merit"

- A measurement of the number of collisions per cm^2 and second.
- Multiple methods used for determining luminosity: reducing uncertainties.
- Principle detectors for luminosity determination on ATLAS:
  - Beam Conditions Monitor (BCM)
    - Designed for beam abort system
    - Diamond Sensors, $|\eta| \sim 4.2$
  - LUCID
    - Dedicated Luminosity Monitor
    - Cherenkov Tubes, $5.6 < |\eta| < 6.0$

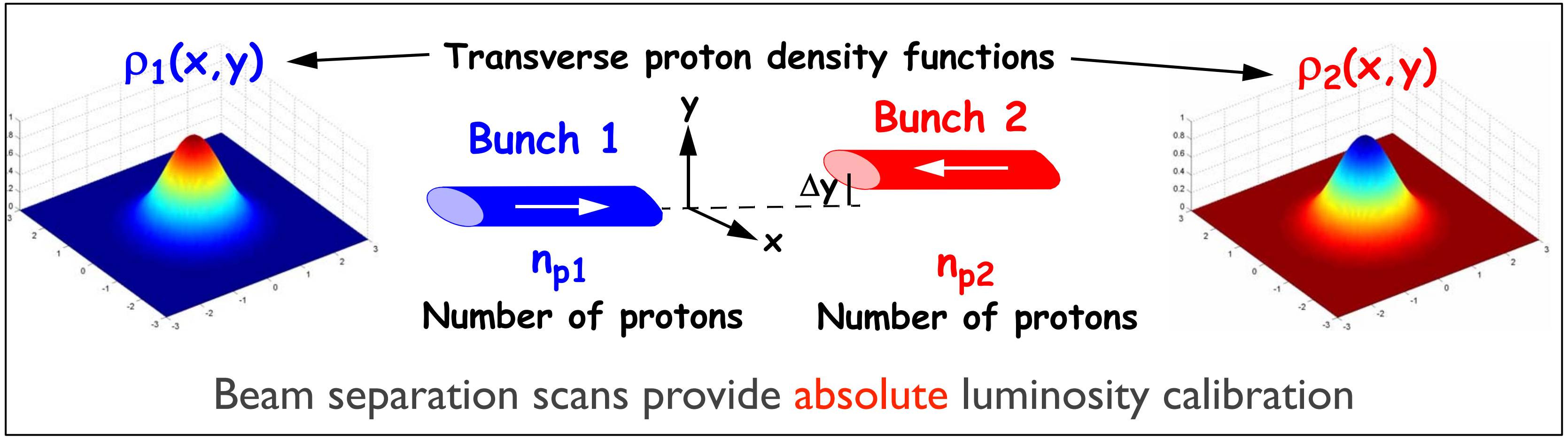




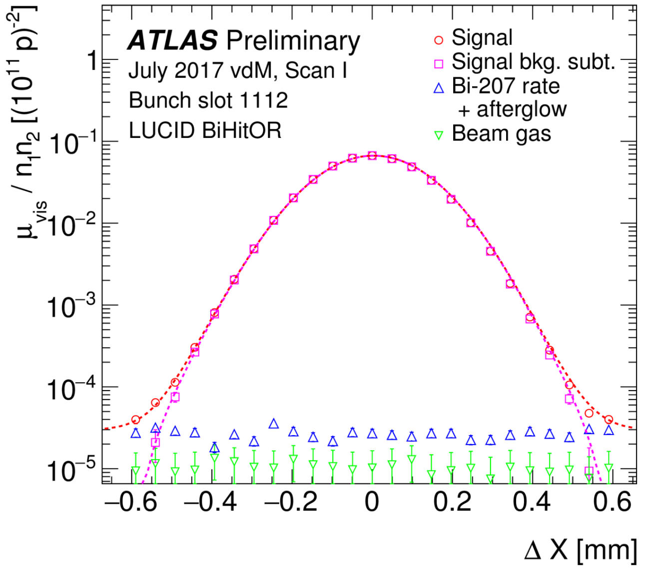
LUCID 2 installation in 2014

Luminosity Determination – VdM scans

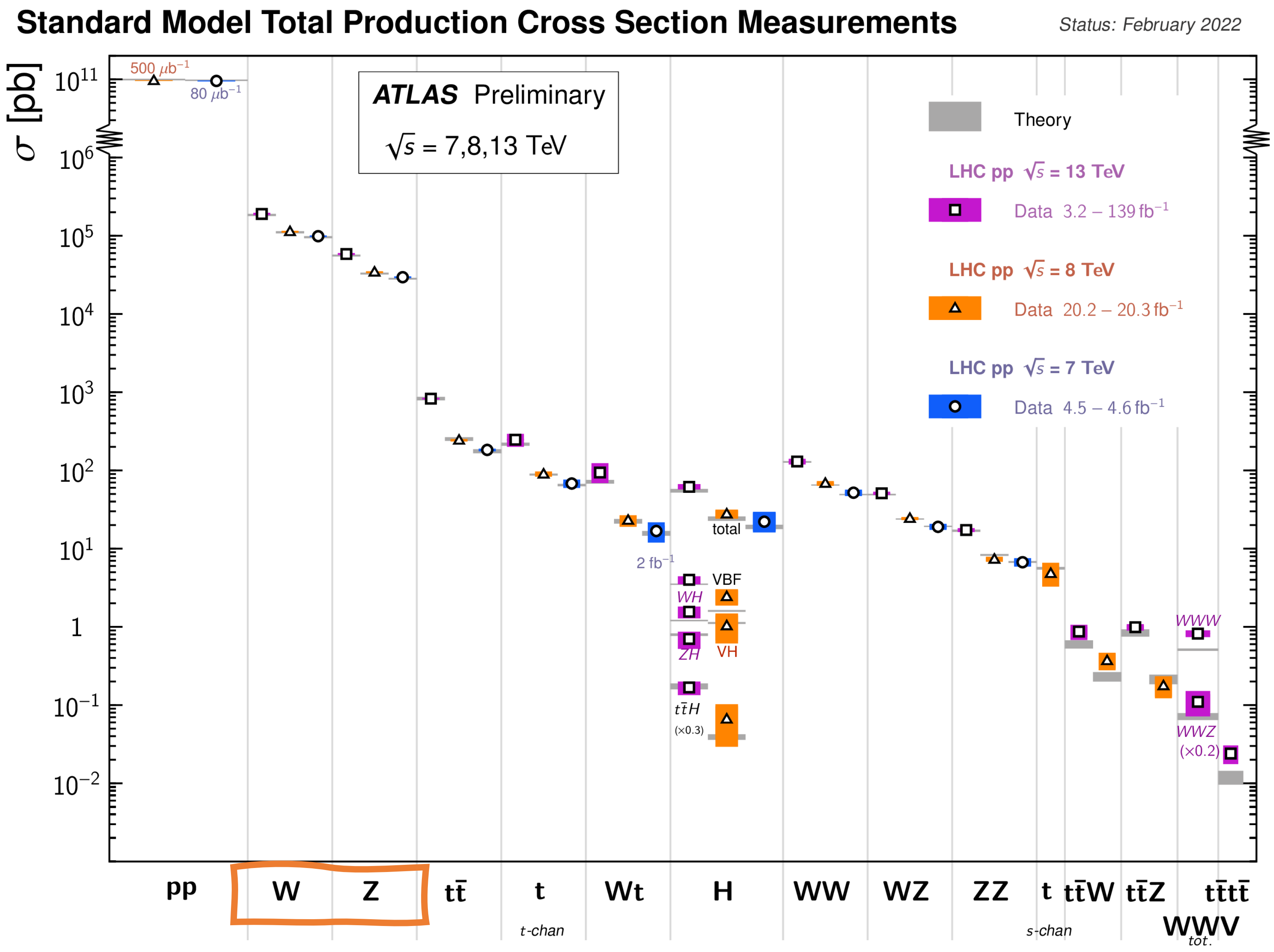
- Normalization is done with beam-separation scan (Van-der-Meer scan). Requires careful control of beam parameters: beams moved vertically and horizontally, varying how they overlap
- Study μ = f(ΔX) – calculating the combined size of both beams in the horizontal and the vertical directions



From http://cds.cern.ch/record/1490292/files/ATL-DAPR-SLIDE-2012-627.pdf



- Determine the total number of protons in each colliding bunch from the measurement of the beam currents
- Result: luminosity measurement with very small uncertainties (order of few %) with very fast turn-around time.



Standard Model Total Production Cross Section Measurements
Status: February 2022
ATLAS Preliminary
√s = 7,8,13 TeV
σ [pb]
Theory
LHC pp √s = 13 TeV
Data 3.2 – 139 fb⁻¹
LHC pp √s = 8 TeV
Data 20.2 – 20.3 fb⁻¹
LHC pp √s = 7 TeV
Data 4.5 – 4.6 fb⁻¹
500 μb⁻¹
80 μb⁻¹
2 fb⁻¹
total
VBF
WH
ZH
VH
tt̄H
(×0.3)
WWW
WWZ
(×0.2)
pp
W
Z
tt̄
t
t-chan
Wt
H
WW
WZ
ZZ
t
s-chan
tt̄W
tt̄Z
tt̄tt̄
WWV
tot.

A TINY BIT OF **MONTE CARLO**

Why do we need Monte Carlo Simulation?

- **We only build one detector: how does this influence the physics we are doing?**
  - How do we compromise physics due to detector design?
  - How would a different detector design affect measurements?
  - How does the detector behave to radiation?
- **In the detectors we only measure voltages, currents, times: how do we go from these to particles?**
  - It's an *interpretation* to say that such-and-such particle caused such-and-such signature in the detector.
  - Simulating the detector behavior we correct for inefficiencies, inaccuracies, unknowns.
- **We need a theory to tell us what we expect and to compare our data against.**
- **A good simulation is the way to demonstrate to the world that we understand the detectors and the physics we are studying.**

Monte Carlo Production Chain

How much processing time needed for each step?

Event Generation
simulate the physics process.

From < 1s to a few hours / event.

Detector Simulation
simulate the interaction of the particles with the detector material.

From 1 to 10min / event

Digitization
Translate interactions with detector into realistic signals.

Reconstruction
Go from signals back to particles, as for real data.

From 5 to 60s / event

Monte Carlo Generators

Various models of the physics of interest

Even more outside ATLAS and CMS!

Courtesy: Z. Marshall

Our LHC Simulation: The Dream

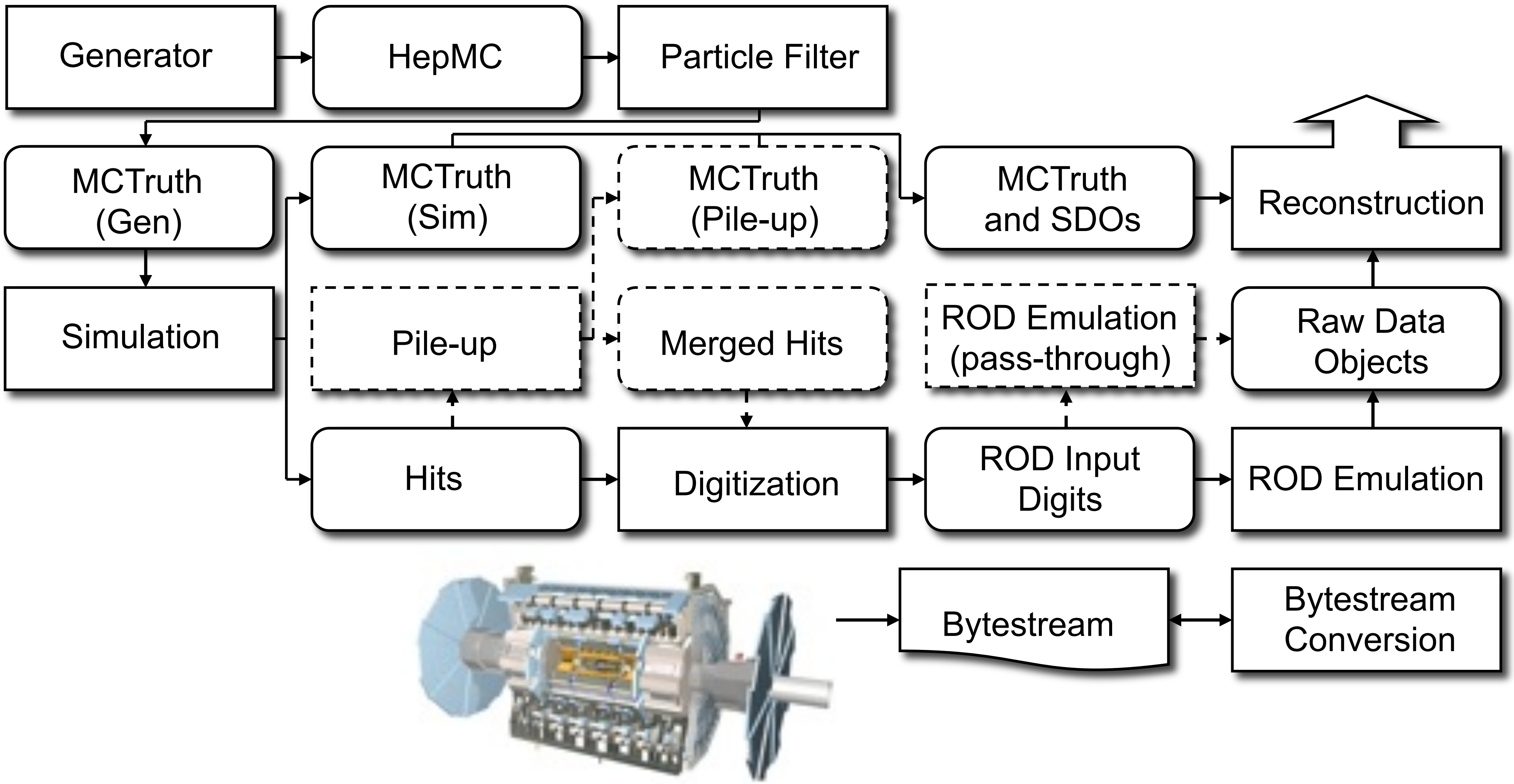


Illustration: Z. Marchall

Our LHC Simulation: The Reality?

This is most people's view of the chain

and this is how we will treat it too, in lack of time…

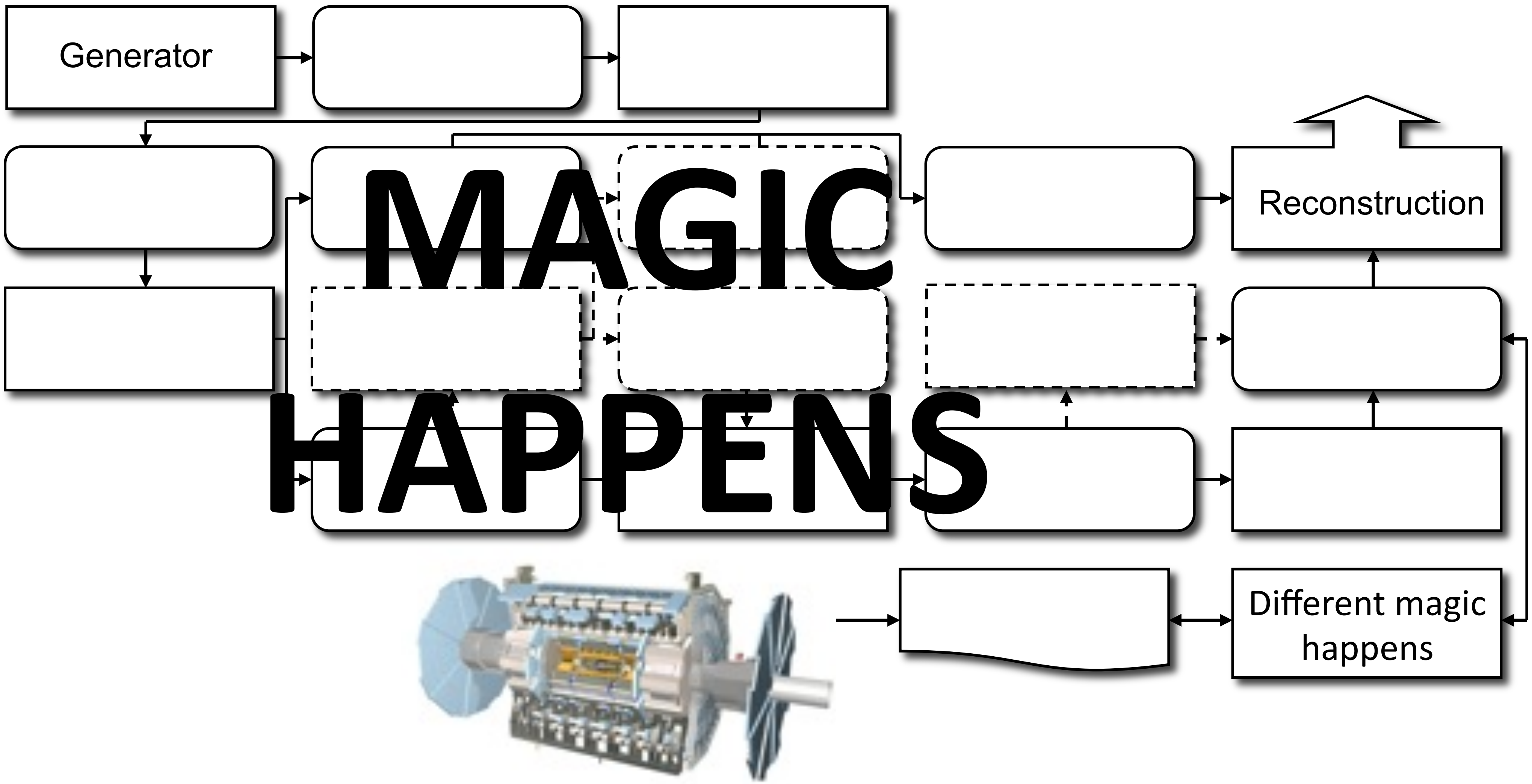
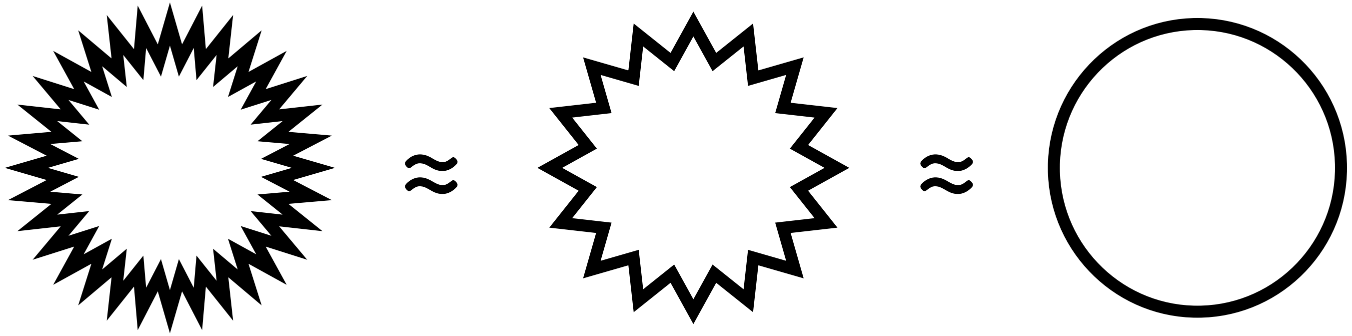
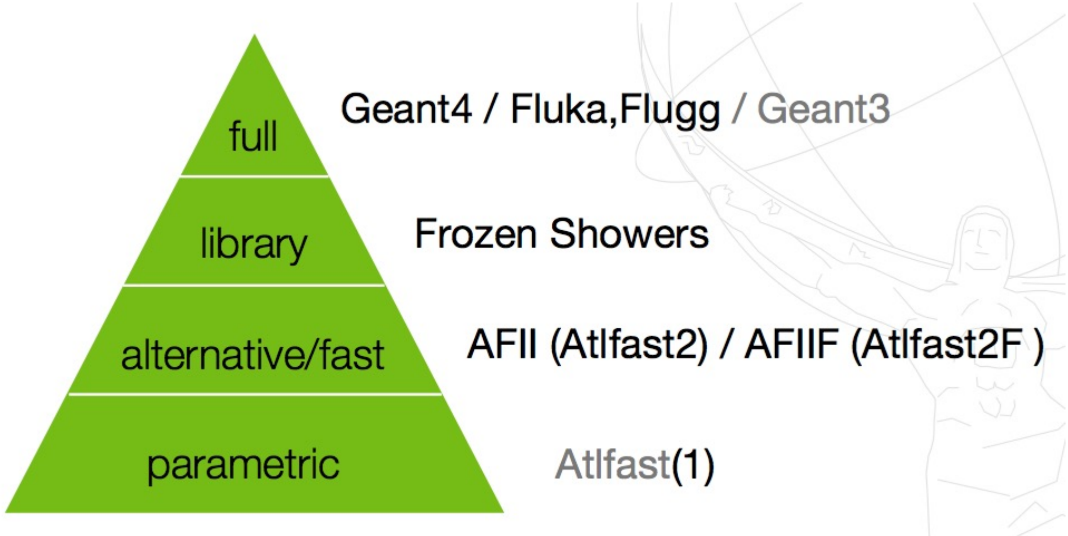


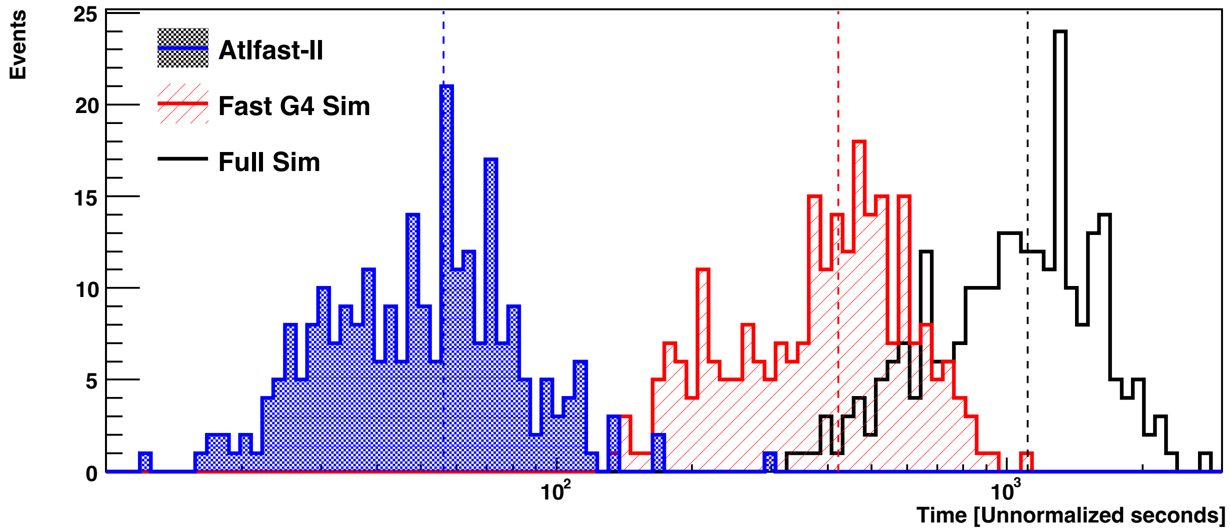
Illustration: Z. Marchall

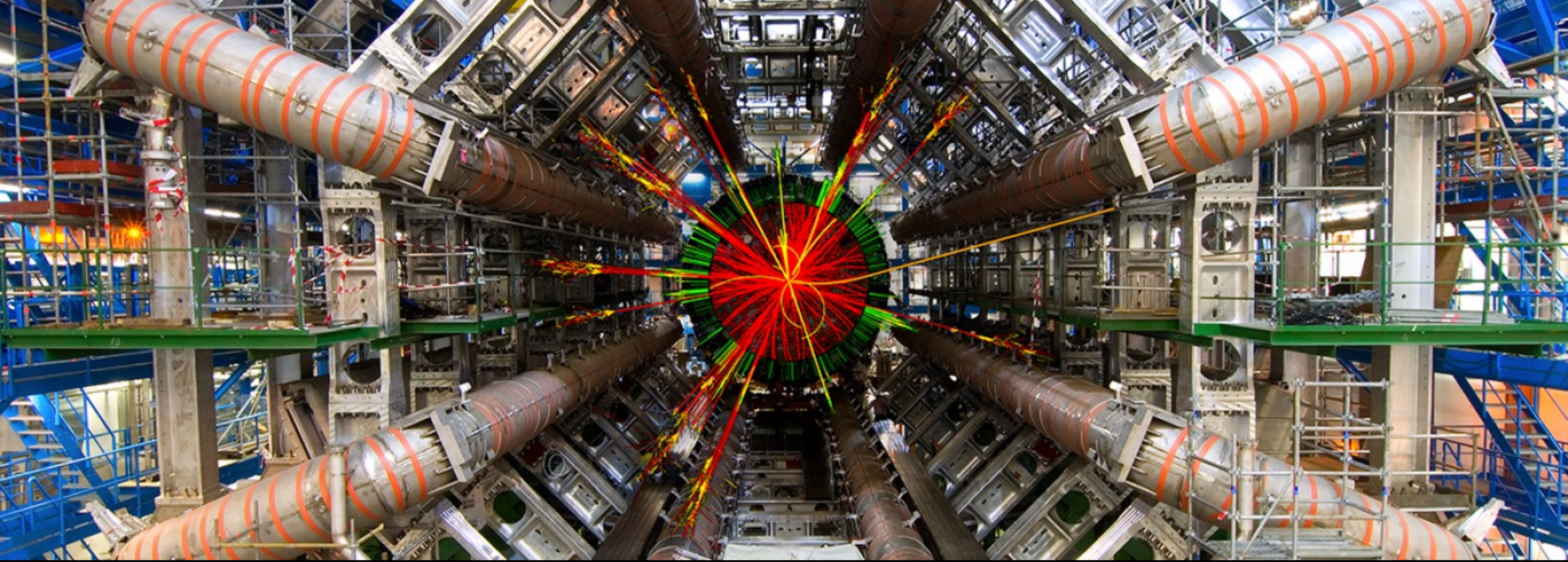
Simulation – Full and Fast



Simulation – Full and Fast







The ATLAS Open Data

Why? Guarantee openness and preservation of experimental data

New open data policy in support of open science from CERN & the LHC experiments

Peer-reviewed publications

- Open Access
- Followed by detailed data related to the results, available at hepdata.net

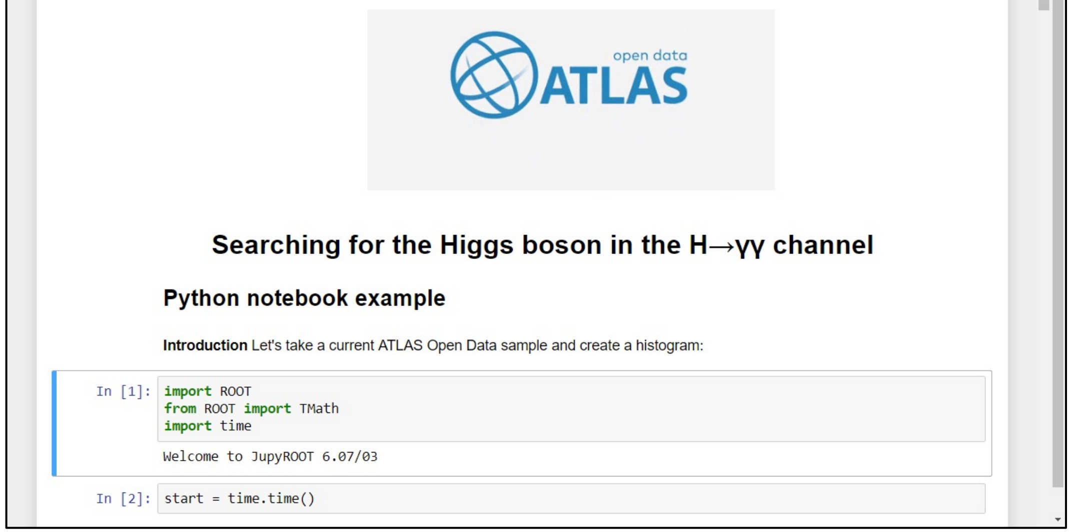
Purpose: Communicate results and maximize their scientific value

Data for outreach and education

- Selected and formatted (“light”) datasets
- Examples available in Jupyter notebooks
- Used in university classes, in growing numbers

Purpose: Maximize educational impact

More info: https://atlas.cern/resources/opendata



Reconstructed & calibrated data

- Followed by related metadata
- Accompanied by appropriate simulated data samples

Purpose: Algorithmic, performance and physics studies