CMS Status

Journées LCG-France

Mathew Nguyen

20-07-2018

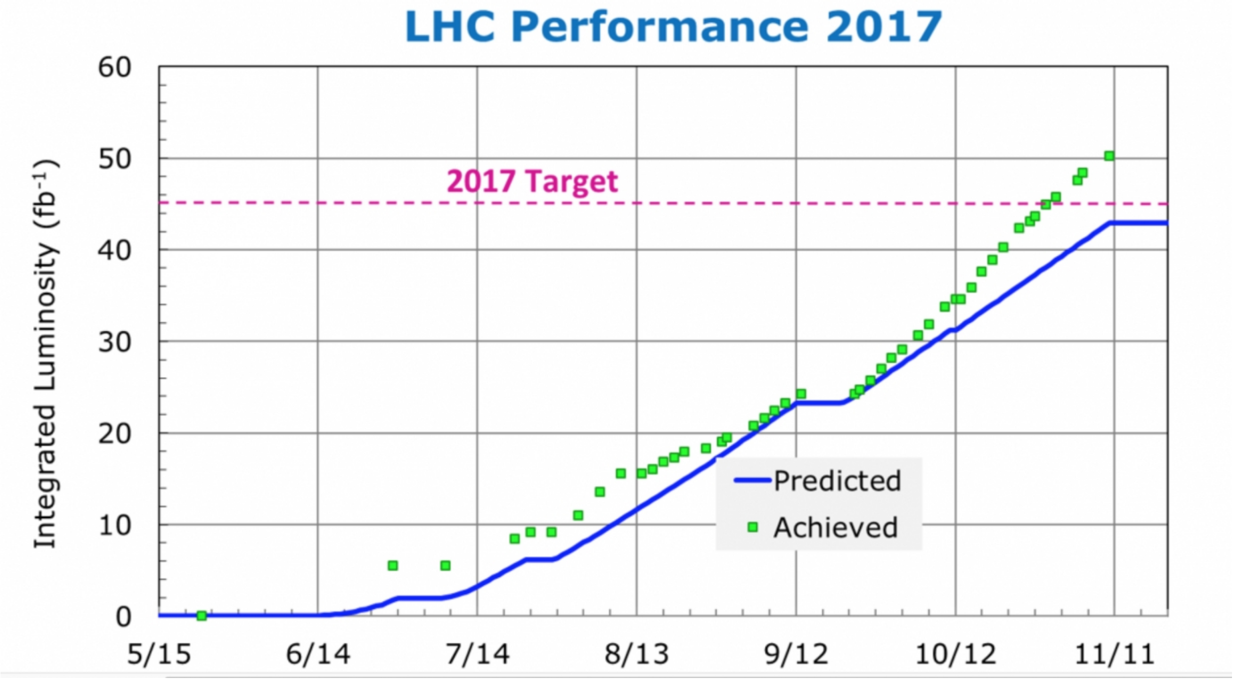
Outline

Talk is roughly time ordered:

- Challenges faced in 2017
- Developments for 2018
- Status of 2018 so far
- (Limited) information about the future

Primary sources:
Offline & computing week, 9 – 13 April
CMS week, 16 – 20 April
NB: Next week is CMS week, so information may soon be out of date

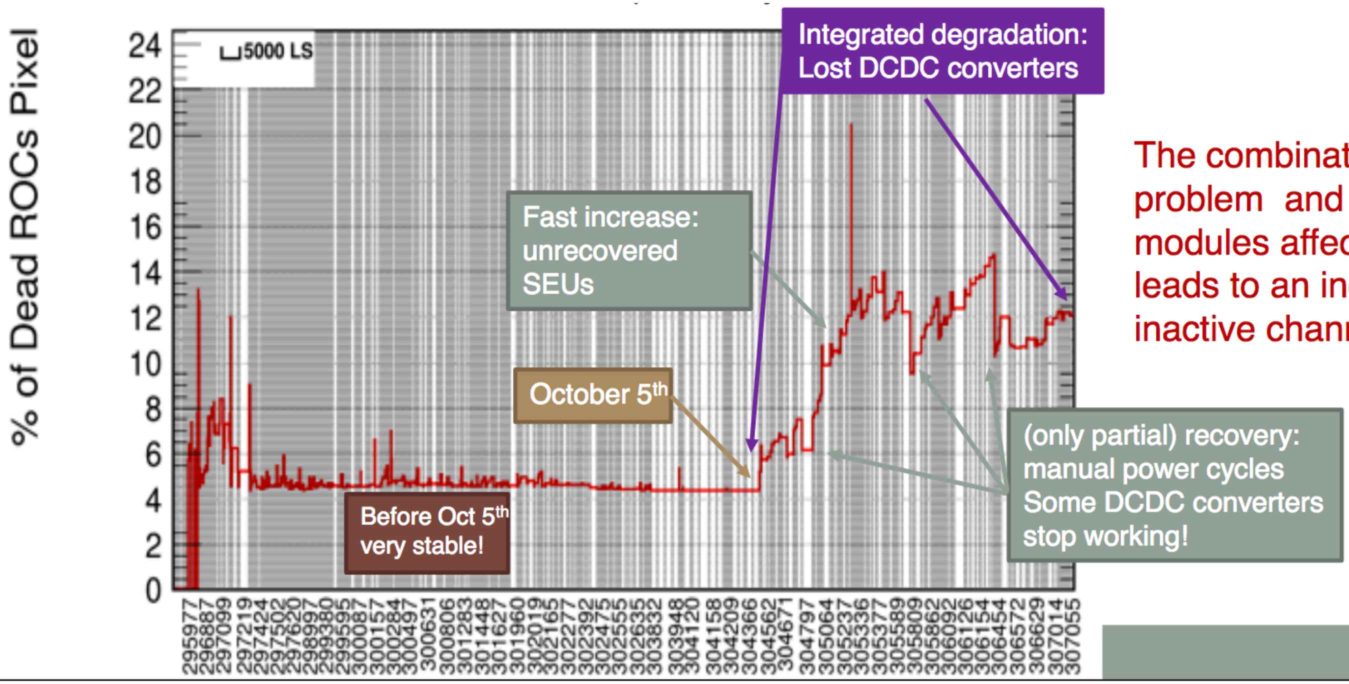
2017 data taking



- 2017 was a full LHC “production” year
- Instantaneous luminosity of 2 x 10^{34} cm^{-2} s^{-1}, 2x design
- CMS faced a couple of major challenges:
    1. Installation of new pixel detector and subsequent failures
    2. Computing resources stretched to maximum

Detector challenge: DCDC converters

- Pixel detector replaced during 2016 – 2017 EYETS
- Towards end of the year started to lose channels during power cycle
- Failures come from “DCDC converters” used to power detector
- Had to replace 1st layer of pixels along w/ DCDC converters in 2018
- Luckily, investigations show only a modest impact on performance
- Problem is now finally understood and mitigations steps are in place



Integrated degradation: Lost DCDC converters

Fast increase: unrecovered SEUs

October 5th

Before Oct 5th very stable!

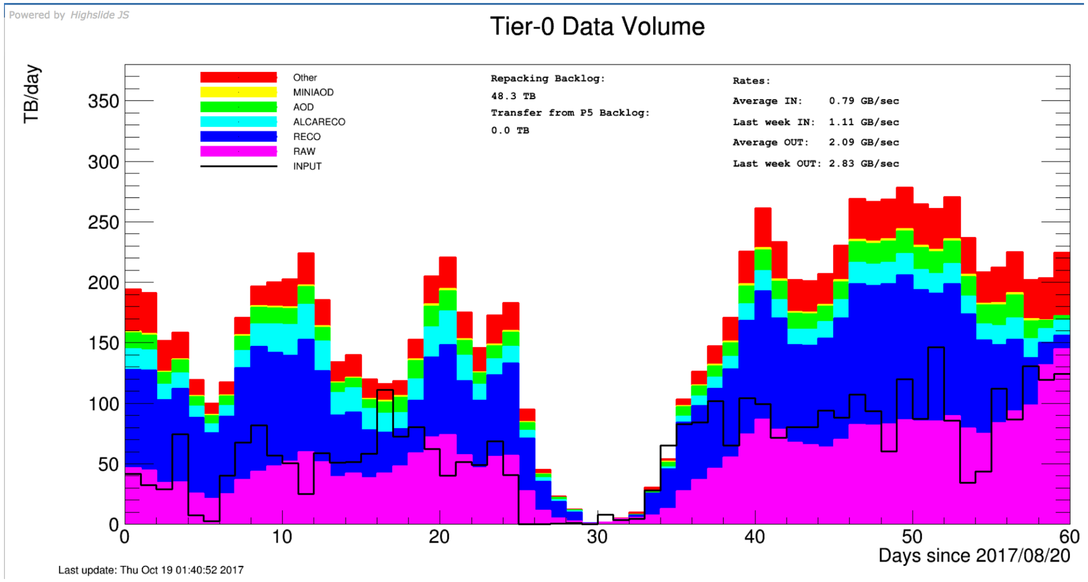
(only partial) recovery: manual power cycles Some DCDC converters stop working!

The combination of the DCDC converter problem and the difficulty to recover the modules affected by SEU without a power cycle, leads to an increase of the fraction of the inactive channels

Source: Narain, LHCC

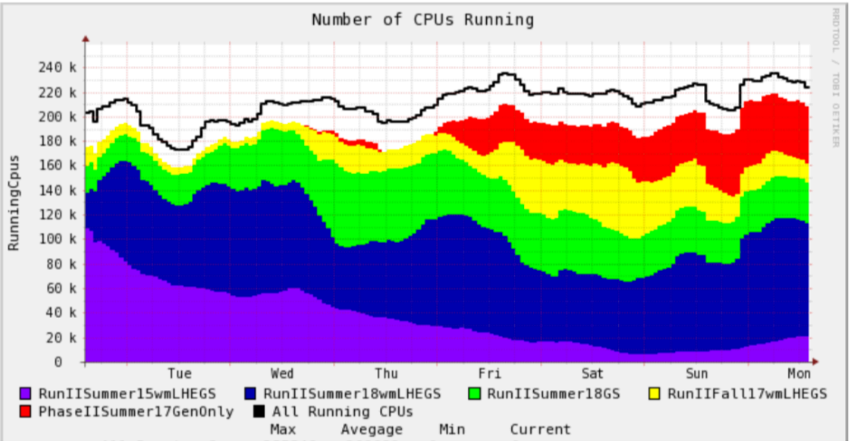
Computing challenge: T0→T1 x-fers

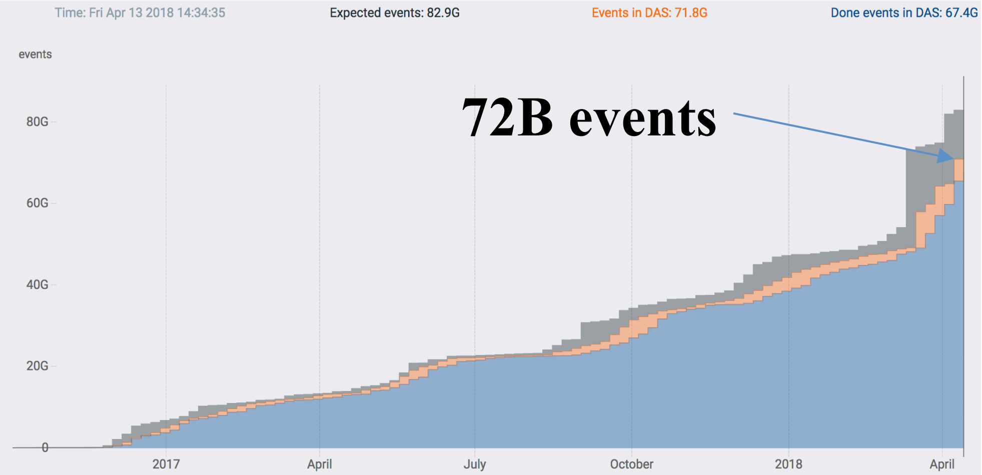
- Phase change after TS2 (mid-Sept), w/ higher than expected pile-up
- Enormous pressure placed on transfer of data from T0 to T1s
  - 3 GB/s T0 output
  - Up to 10 PB / week (70 PB total)
- Spike in data volume + chronic under-pledging of CMS T1s
  → disk constantly at quota, requiring lots of manual intervention



Full resource utilization

- Starting from xmas 2017, HLT (50k cores) enabled as a production resource (important test for 2019+ operations)
- During EYETS routinely reached a record 200k cores
  - 90% of 2017 data processed by January 1st
  - > 10B 2017 MC events + 1 B early 2018 MC + phase II MC
- More than 70 B events processed in 18 months





72B events

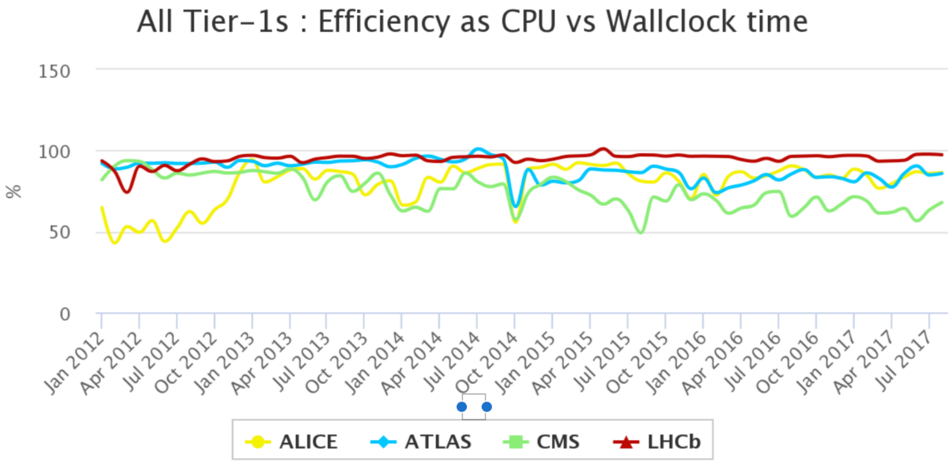
Special runs

No heavy-ion run in 2017, but:

- XeXe pilot run in Oct
- pp “reference run”: HLT rate increased from 1 kHz to 35 kHz, recording 4PB of data. Large backlog of transfers from disk to tape
- Low energy, high beta* run

The mystery of CMS CPU inefficiency

A task force was formed to investigate growing CPU inefficiency vs. time w.r.t. ATLAS

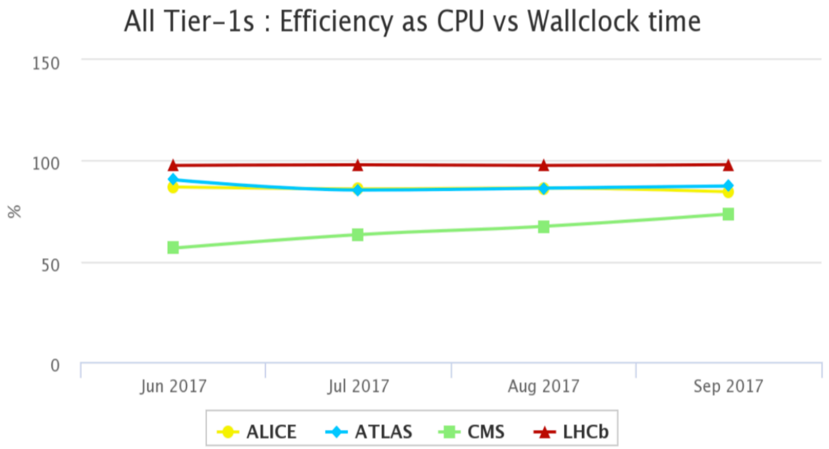


CMS software is fully multi-threaded
→ expect reduced CPU efficiency in exchange for reduced memory.
→ Not sufficient to explain inefficiency

Inefficiency considered into two pieces

1) Submission infrastructure
2) Job-level, i.e., "payload" inefficiency

Performance improvements related to 1)

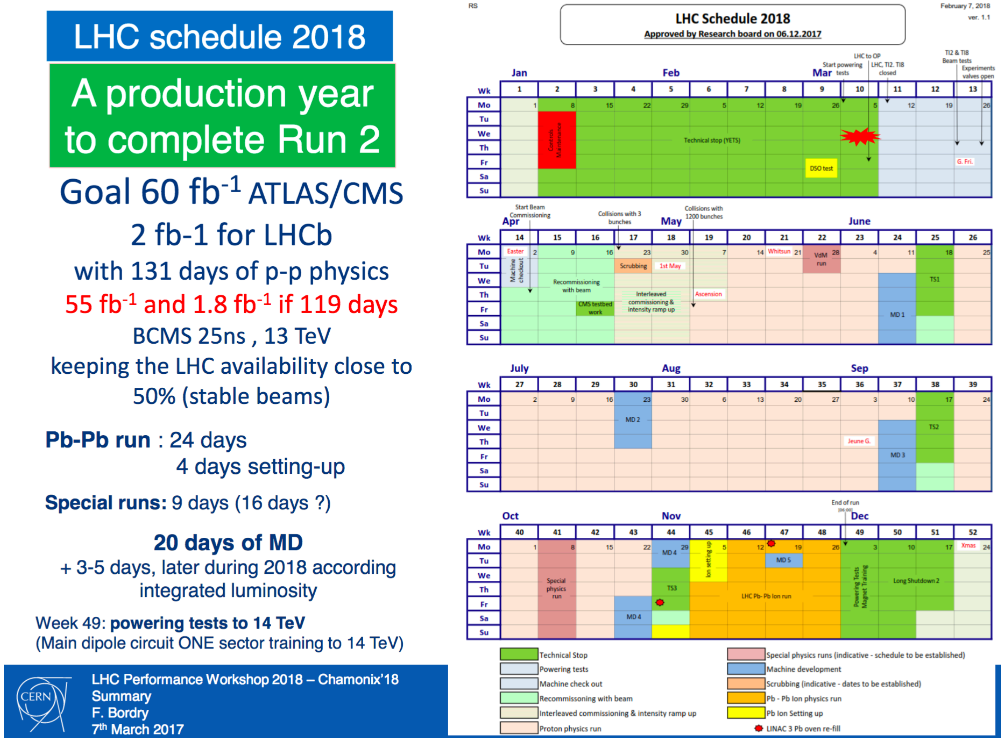


1) A number of improvements were made to grid submission to better handle fluctuating job pressure
2) Residual inefficiency mostly on payload side. Major source is multithread jobs executing external (typically Fortran) code.

2018: Another production year

Integrated luminosity goal from LHC

https://indico.cern.ch/event/705545/attachments/1613081/2562222/Chamonix_Summary_Fk_Bordry_7_March_2018.pdf



LHC schedule 2018

A production year to complete Run 2

Goal 60 fb^{-1} ATLAS/CMS

2 fb-1 for LHCb

with 131 days of p-p physics

55 fb^{-1} and 1.8 fb^{-1} if 119 days

BCMS 25ns , 13 TeV

keeping the LHC availability close to 50% (stable beams)

Pb-Pb run : 24 days
4 days setting-up

Special runs: 9 days (16 days ?)

20 days of MD

+ 3-5 days, later during 2018 according integrated luminosity

Week 49: **powering tests to 14 TeV**
(Main dipole circuit ONE sector training to 14 TeV)

RS

LHC Schedule 2018
Approved by Research board on 06.12.2017

February 7, 2018
ver. 1.1

Jan · Feb · Mar

Start powering tests · LHC to OP · LHC, TI2. TI8 closed · TI2 & TI8 Beam tests · Experiments valves open

Wk	1	2	3	4	5	6	7	8	9	10	11	12	13
Mo	1	8	15	22	29	5	12	19	26	5	12	19	26
Tu		Controls Maintenance											
We						Technical stop (YETS)							
Th													
Fr									DSO test				G. Fri.
Sa													
Su													

Apr · May · June

Start Beam Commissioning · Collisions with 3 bunches · Collisions with 1200 bunches

Wk	14	15	16	17	18	19	20	21	22	23	24	25	26
Mo	Easter 2	9	16	23	30	7	14	Whitsun 21	VdM run 28	4	11	18	25
Tu	Machine checkout			Scrubbing	1st May								
We		Recommissioning with beam										TS1	
Th				Interleaved commissioning & intensity ramp up		Ascension							
Fr			CMS testbed work								MD 1		
Sa													
Su													

July · Aug · Sep

Wk	27	28	29	30	31	32	33	34	35	36	37	38	39
Mo	2	9	16	23	30	6	13	20	27	3	10	17	24
Tu				MD 2									
We												TS2	
Th										Jeune G.			
Fr											MD 3		
Sa													
Su													

Oct · Nov · Dec

End of run [06:00]

Wk	40	41	42	43	44	45	46	47	48	49	50	51	52
Mo	1	8	15	22	MD 4 29	Ion setting up 5	12	● 19	26	3	10	17	Xmas 24
Tu								MD 5					
We		Special physics run								Powering Tests Magnet Training	Long Shutdown 2		
Th					TS3		LHC Pb- Pb Ion run						
Fr					●								
Sa				MD 4									
Su													

🟩 Technical Stop	🟥 Special physics runs (indicative - schedule to be established)
▢ Powering tests	🟦 Machine development
▢ Machine check out	🟧 Scrubbing (indicative - dates to be established)
🟩 Recommissoning with beam	🟧 Pb - Pb Ion physics run
▢ Interleaved commissioning & intensity ramp up	🟨 Pb Ion Setting up
▢ Proton physics run	● LINAC 3 Pb oven re-fill

CERN

LHC Performance Workshop 2018 – Chamonix'18
Summary
F. Bordry
7th March 2017

But even more demanding than 2017

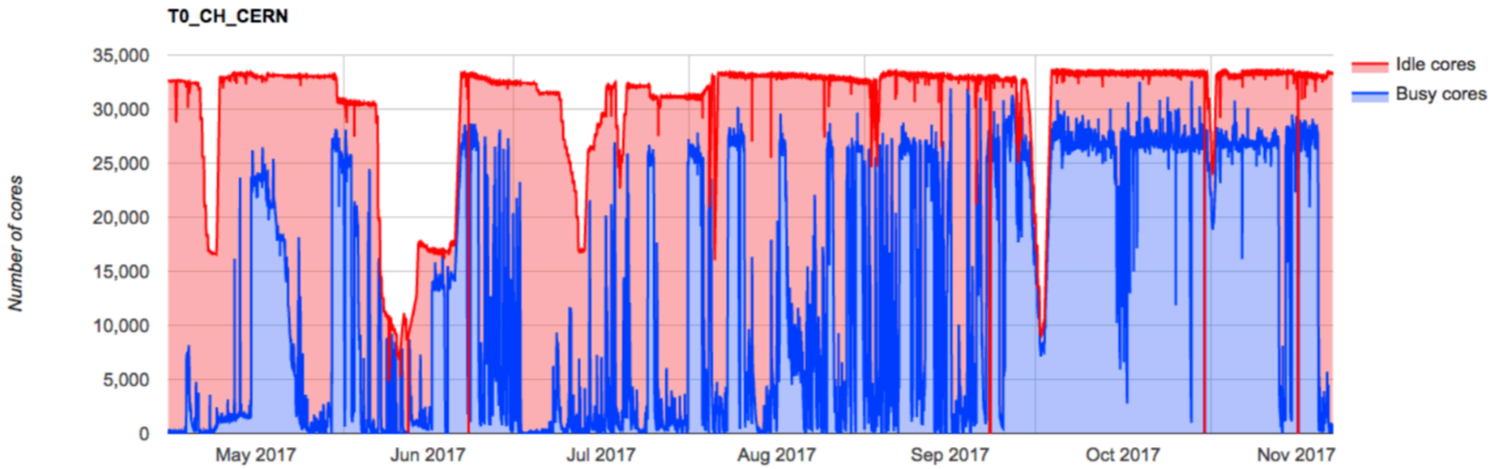
- Larger data volume
- Large trigger rate proposals for special runs (heavy-ions, low PU)
- "Parking" of data for b-physics, i.e., not reconstructed until LS2

A better CMS in 2018

- Tracker detector reinstalled
  - Failed DCDC converters replaced
  - 75% of barrel pixel modules replaced
  - Reconditioned and in stable operation with 97% active channels
- Phase I HCAL upgrade completed
  - HPDs replaced by SiPMs for higher photo-detection efficiency
  - Removed source of coherent noise
  - Increased longitudinal segmentation (2-3 → 6-7 readout depths)

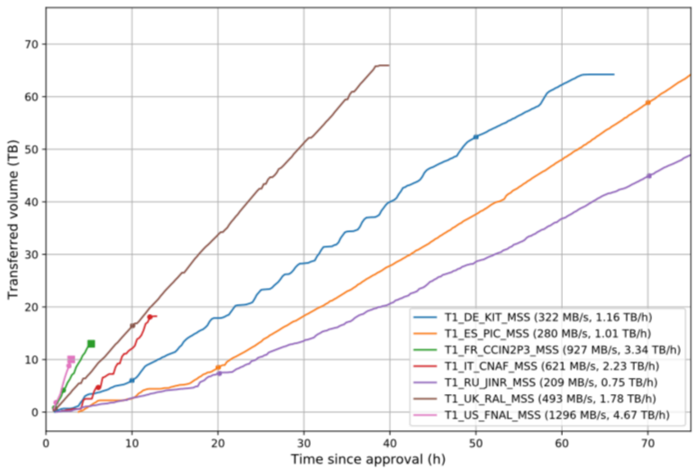
Also, a better T0 for CMS

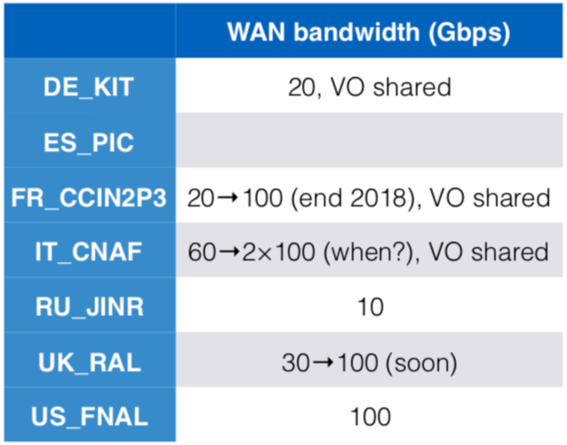
- T0 migration from a dedicated infrastructure to a shared one
  - With the CERN-T2, and with ATLAS!
  - Enabled by switch from OpenStack to HTCondor
- Much easier on CMS side to use CERN resources
  - T0 & T2 jobs are running on the same machines, driven by priority
  - No need to partition EOS space into distinct areas: T0 buffer can be enlarged at the expenses of Analysis Space
  - All together, a system which is easier to steer, also at the last second



Addressing T0→T1 bottleneck

- System tests conducted in March-April
- Results will be used to map datasets to T1s
- New disk-cleaning policy put into effect, removing datasets untouched for 100 days → enough space for first few months of data
- For CCIN2P3, 100 Gb connection will help (but only after this year)





	WAN bandwidth (Gbps)
DE_KIT	20, VO shared
ES_PIC	
FR_CCIN2P3	20→100 (end 2018), VO shared
IT_CNAF	60→2×100 (when?), VO shared
RU_JINR	10
UK_RAL	30→100 (soon)
US_FNAL	100

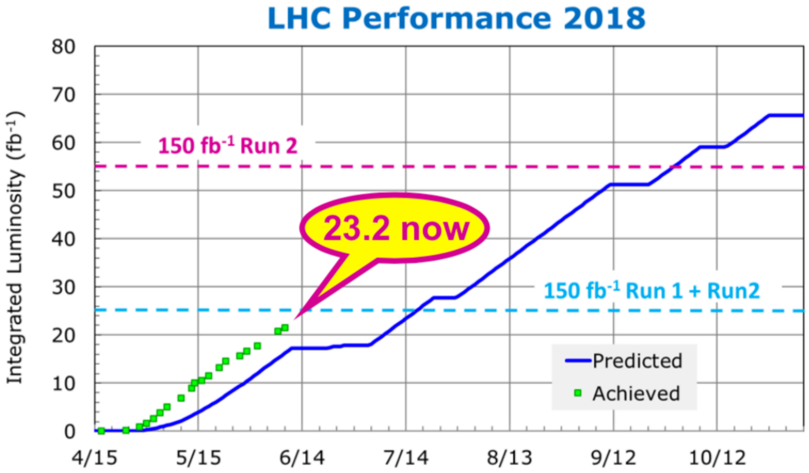
CRAB developments

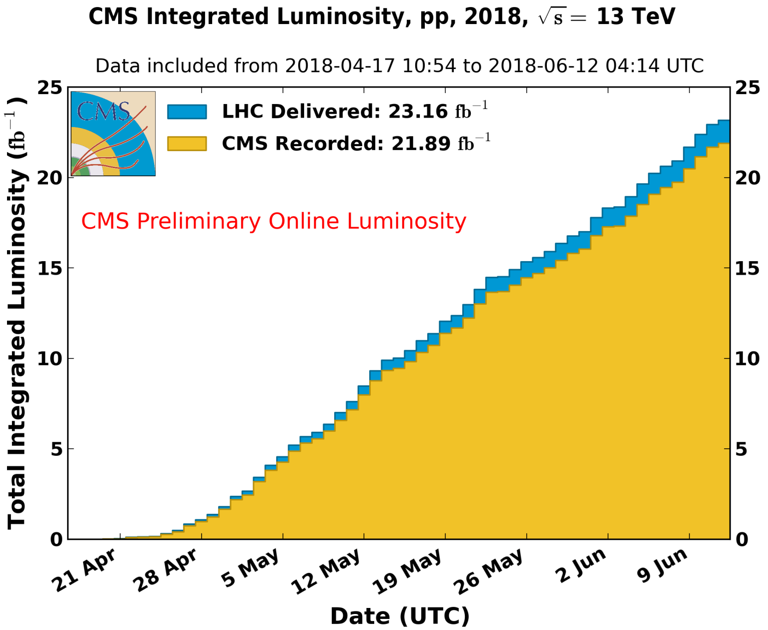
- CRAB is the CMS user interface to the grid ("CMS Remote Analysis Builder")
- For tape-resident datasets, now automatically requests staging (w/ some protections)
- Now splits jobs automatically to reduce running many small jobs, which are dominated by initialization time

2018 data so far

CMS taking data efficiency ~ 95%

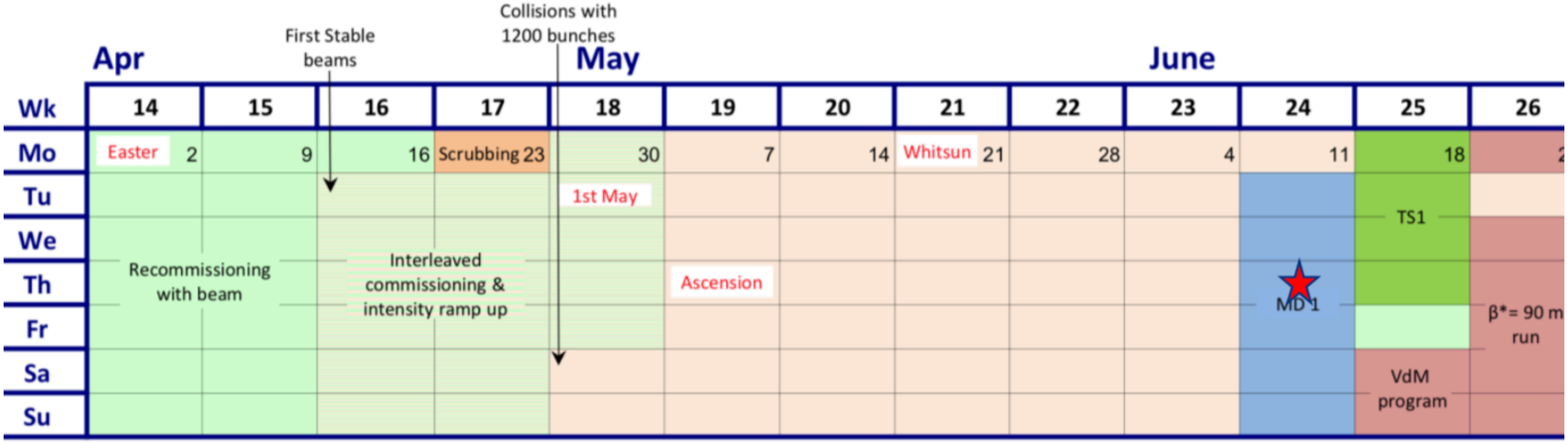
Almost two weeks ahead of the prediction





Current status

- Now in the first Machine Development week
- Then a short Technical Stop
- Followed by the VdM scan and β^* 90m run



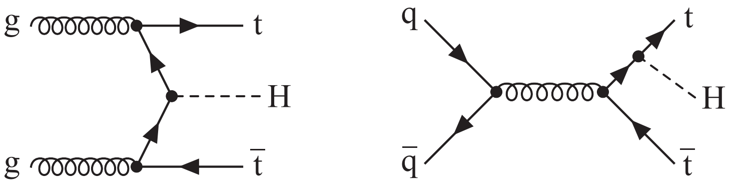
First Stable beams

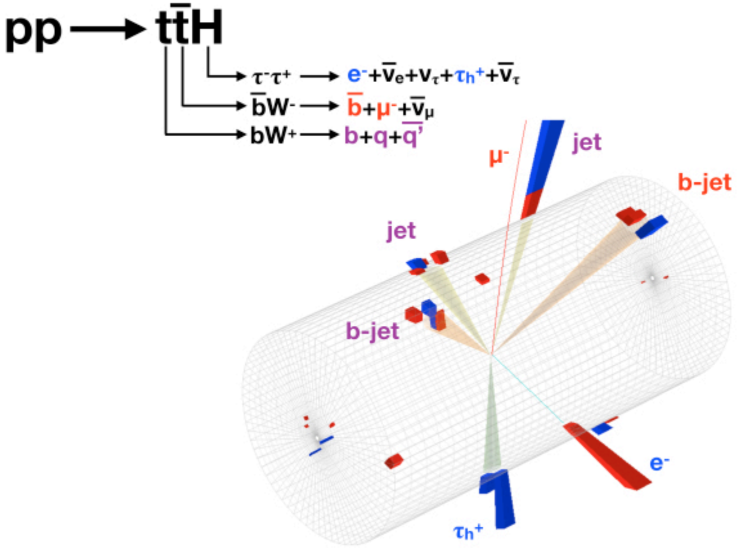
Collisions with 1200 bunches

	Apr				May					June			
Wk	14	15	16	17	18	19	20	21	22	23	24	25	26
Mo	Easter 2	9	16	Scrubbing 23	30	7	14	Whitsun 21	28	4	11	18	2
Tu					1st May							TS1	
We													
Th	Recommissioning with beam		Interleaved commissioning & intensity ramp up			Ascension					MD 1		β*= 90 m run
Fr													
Sa												VdM program	
Su													

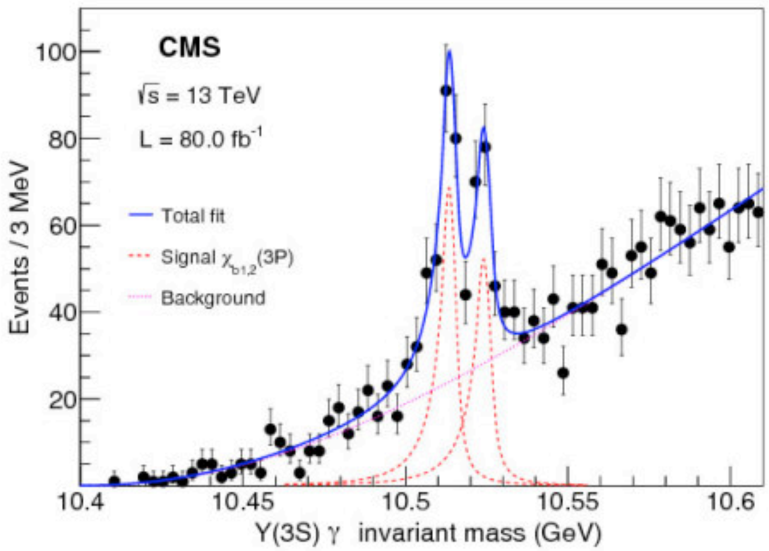
Recent physics news (LHCP)

Observation of ttH production




$pp \longrightarrow t\bar{t}H$
$\tau^-\tau^+ \longrightarrow e^- + \bar{\nu}_e + \nu_\tau + \tau_h^+ + \bar{\nu}_\tau$
$\bar{b}W^- \longrightarrow \bar{b} + \mu^- + \bar{\nu}_\mu$
$bW^+ \longrightarrow b + q + \bar{q}'$
μ-
jet
b-jet
jet
b-jet
e-
τh+

Observation of $\chi_{b1}(3P)$ and $\chi_{b2}(3P)$ states



PRL 120, 231801 (2018)

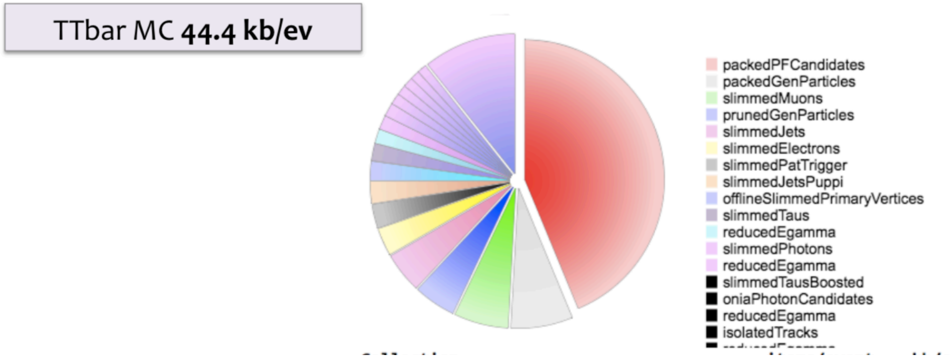
Data collected <= 2016

arXiv:1805.11192

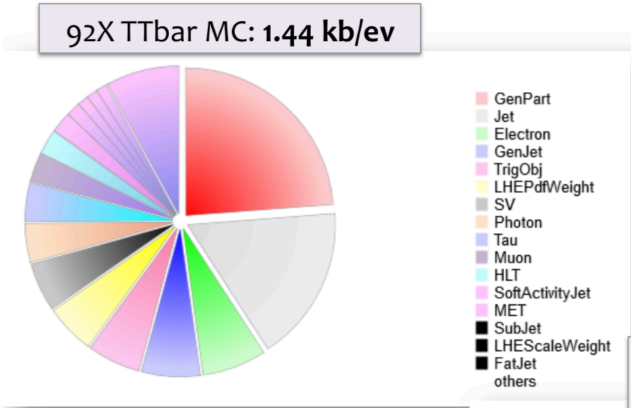
Includes data from 2017

Data format evolution

- Mini-AOD
  - Size: ~ 40 kB per event
  - Stable definition, produced ~ 2x per year
  - Widely used for analysis, increasingly used for upgrades



- Nano-AOD
  - Size: ~ 1 kB per event !
  - ➢ Entire Run 2 = 50 TB !
  - Commissioning in 2018, targeting 50% usage
  - Potential game changer for long-term computing needs



LS2 plans

- Detector work
  - Pixel layer 1 and DCDC replacement
  - Muon CSC and GEM upgrades
  - …

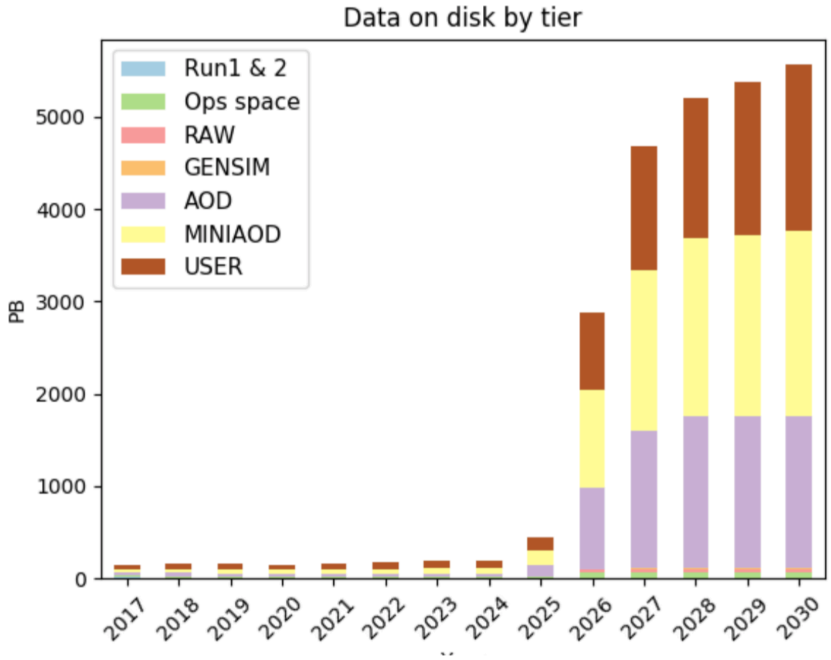
- Analysis of 150 fb^{-1} legacy dataset
- HL-LHC will increasingly compete for resources
  - Several TDRs still to done
  - Possible there will be a computing TDR

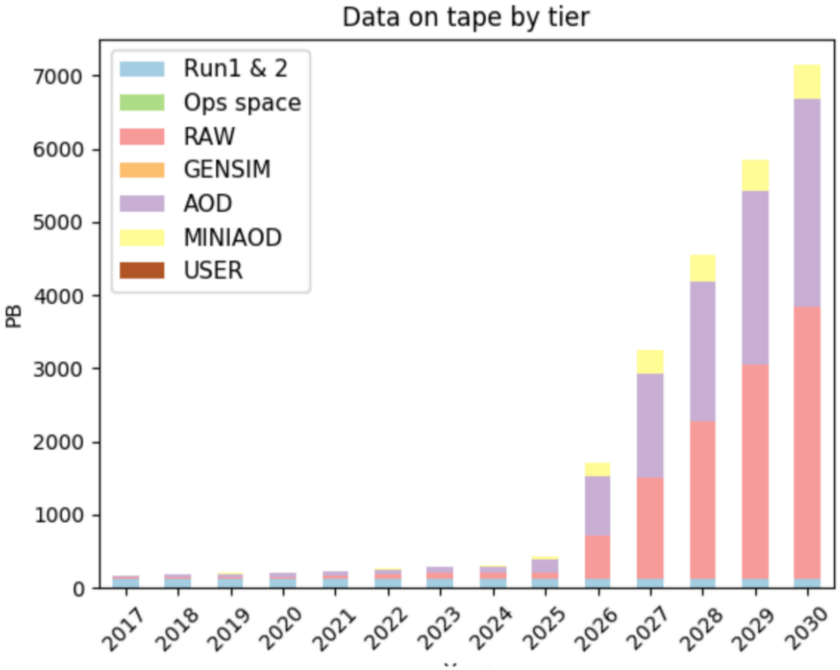
Computing during LS2

Preliminary!

- 2019 resources
  - Primary task: full reprocessing of Run 2 data and MC (60 B events)
  - HLT heavily utilized in model
  - T0 assumed to be 100% available
  - No major changes wrt 2018, except parked data reconstruction
  - Compensated by aggressive disk and tape cleaning
- Initial thoughts on 2020
  - Tail of Run 2 legacy production
  - Peak of Run 2 analysis
  - Run 3 MC production, including detector commissioning
  - Continued HL-HLC MC for TDRs

Long-term projections


Data on disk by tier


Data on tape by tier

U.S. CMS asked to estimate long-term computing needs, including HL-LHC era

Currently using a naïve extrapolation, mostly illustrates scale of the problem

Does not incorporate technological improvements, or proposed data-format evolution

Of course the main driver of resources is the trigger rate

Pledge flexibility

- Computing needs are variable, hitting peaks, e.g., during a reprocessing campaign
- Certain resources intrinsically variable
  - Opportunistic resources, e.g, Open Science Grid
  - Commercial clouds
  - Non-HEP supercomputing sites
- On-going reflection towards adapting the pledge system to dynamic resources
- Expect WLCG to receive a proposal from CMS along these lines

Summary

- 2017 was a record year in terms of data taking, but several challenges had to be overcome
  - Issues w/ pixel detector
  - Throughput from T0 to T1
- Lessons were learned for 2018. So far off to a good start, but we continue to find ways to push the envelope
  - Reconfiguration of T0
  - Improvements to CPU efficiency / job submission infrastructure
  - Introduction of ultra-compact data-tier
- Planning is becoming concrete for LS2. Will be intense despite the lack of new data.