What happens in PPD-land

Gianluca Cerminara (CERN)
Markus Klute (MIT)

Outlook

Three main topics

- Exploitation of 2016 data
- Planning of 2017 preparation
- Consolidation of operations and workflows

Outlook

Three main topics

- Exploitation of 2016 data
- Planning of 2017 preparation
- Consolidation of operations and workflows

Re-reco of 2016 data

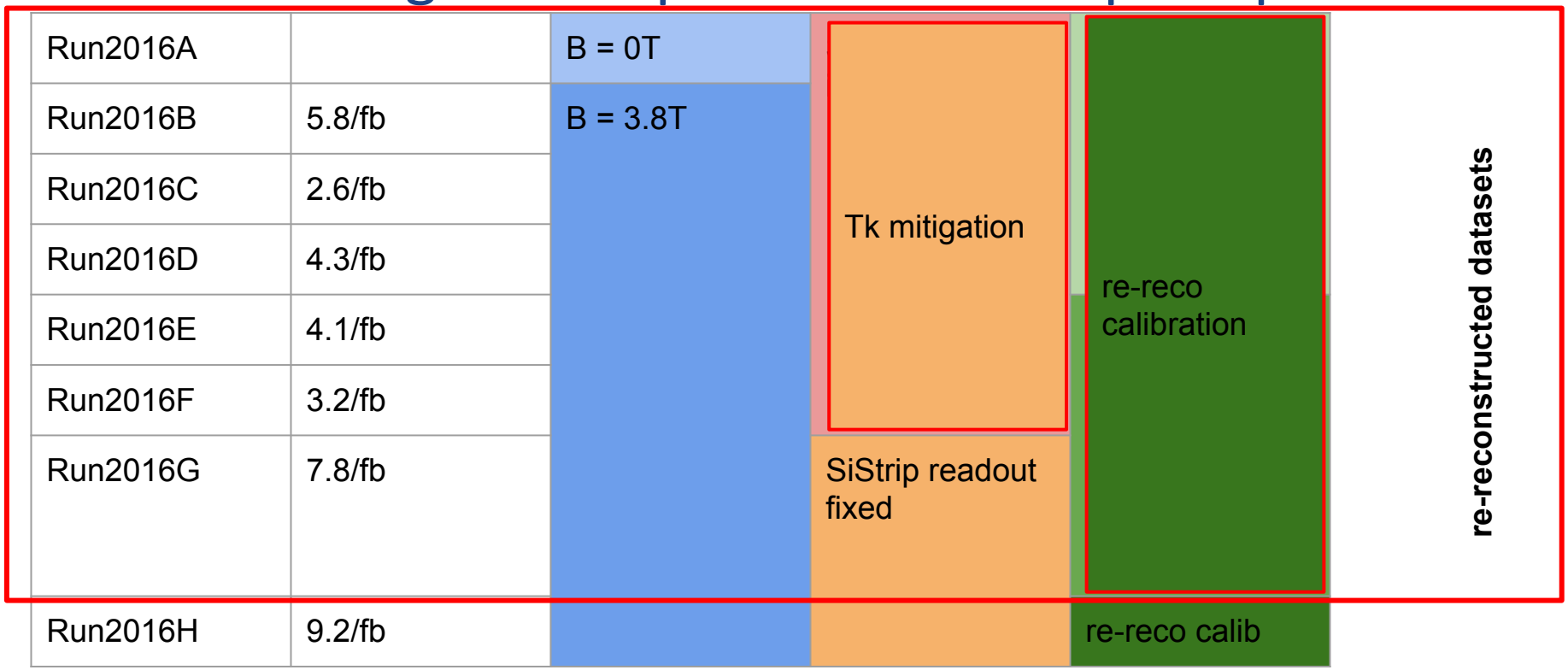
- non-homogeneous performance of prompt reco

Run2016A		B = 0T	SiStrip dynamic ineff.	Cosmics based calibrations
Run2016B	5.8/fb	B = 3.8T		
Run2016C	2.6/fb			
Run2016D	4.3/fb			
Run2016E	4.1/fb			New alignment & calibration deployed (derived on June dataset)
Run2016F	3.2/fb			
Run2016G	7.8/fb		SiStrip readout fixed	
Run2016H	9.2/fb			re-reco calib

- Re-reco of most part of the dataset injected in Sep
  - aim at homogenous (Tracking & Calibrations) for Moriond17

Re-reco of 2016 data

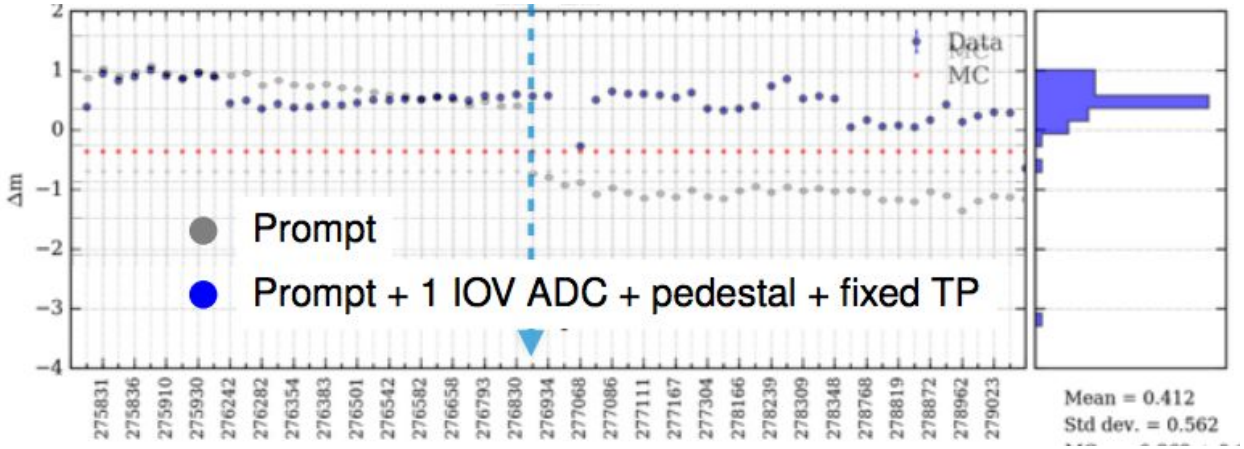
- non-homogeneous performance of prompt reco



- Re-reco of most part of the dataset injected in Sep
  - aim at homogenous (Tracking & Calibrations) for Moriond17

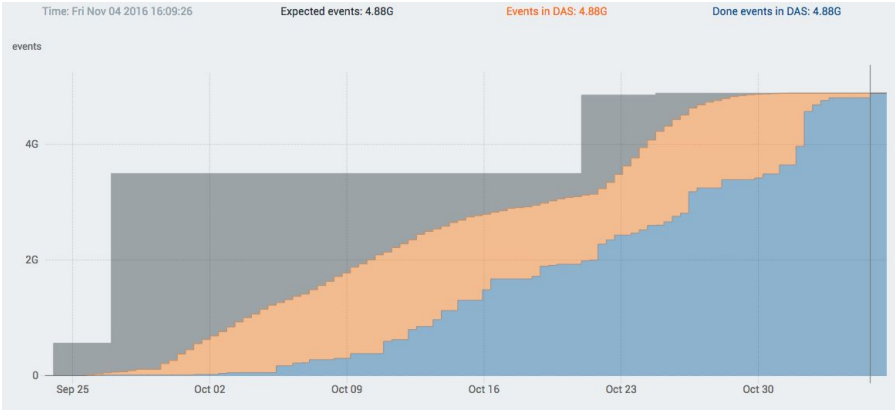
Improvements in re-reco

- Several improvements in Alignment & Calibrations
  - update in pixel conditions for rad. damage corrections
  - updated SiPixel alignment
  - fix stability of ECAL scale
  - HCAL corrections for radiation damage



- Mitigation of SiStrip dynamic inefficiencies in tracking algorithm
  - recover efficiency for b-tagging track selection

New datasets already completed



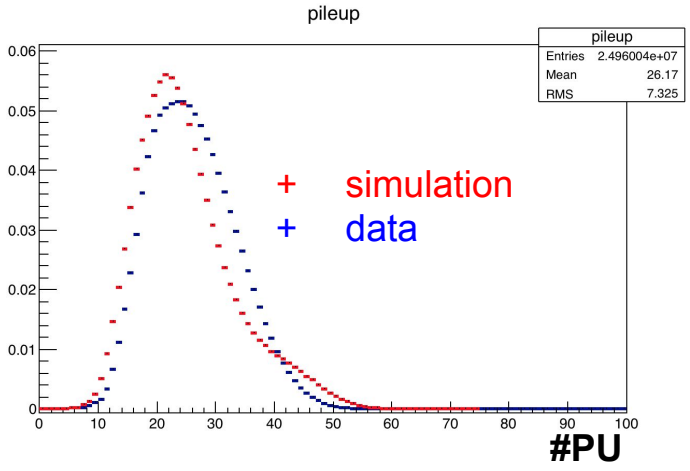
- 23Sep2016 re-reco completed in ~1 month processing
  - now focusing on scale factors and calibrations to exploit these datasets
- Working to mitigate some features of our dataset affecting high-energy E/gamma objects, muons and MET
  - ECAL slew rate effect (see talk at PPD general)
  - poorly reconstructed and duplicated muons, partly due to tracking mitigation for SiStrip dynamic ineff. (see talk)
- Will produce a new MiniAOD version as soon as an AOD level recipe is implemented

Planning Legacy 2016 re-reco

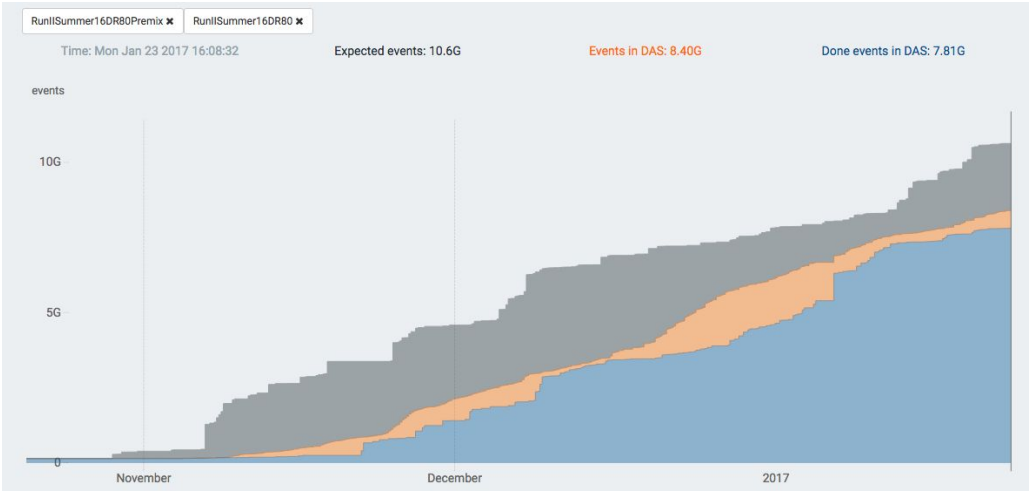
- meant to squeeze the ultimate performance
  - Tk alignment: remove weak modes in SiStrip endcaps
    - impact on muon pT scale at high rapidity
  - HCAL phi-dependent corrections & transition regions
  - HF radiation damage
  - ECAL inter-calibrations: improve resolution
- Some conditions already used to produce samples
  - aiming at final set by mid-March: ECAL driving schedule
- NOTE I: ideally the last re-reco of 2016 for a while
  - speak up if you have improvements in the pipeline
- NOTE II: important by product:
  - use this exercise to further develop calibration strategy for 2017

MC production for Moriond17

- New production for analysis on complete dataset
  - trigger simulation for all samples & with more recent menu (L1+HLT)
  - new PU profile: LHC reached ~1.5X10^{34}cm^{-2}s^{-1}
  - improved calibrations (e.g. pixel dynamic ineff.) → better data vs MC agreement



- Status:



Outlook

Three main topics

- Exploitation of 2016 data
- Planning of 2017 preparation
- Consolidation of operations and workflows

Building blocks for 2017 readiness

- Several components need to be ready
  - new generators, new tune, and corresponding validation
  - simulation & digitization for new hardware
  - **new reconstruction algorithms to exploit new hardware**
  - **calibrations: both low level (detector level) and those depending on them (POG level)**
  - trigger simulation
- Requirements to be fulfilled along the way
  - need samples to prepare the calibrations
  - need samples to prepare the trigger menu
  - need a release to take data in GRun in February and for the 24/7 operations from March onwards
  - need to study possible PU scenarios for 2017 LHC
- Need to take care of the dependencies

Building blocks for 2017 readiness

- Need to converge on the HCAL choice
  - risk vs benefit to be assessed taking into account JetMET
- Need Tk alignment and calibrations to be exercised
  - SIM, local reco and tracking need to be fully functional
    - 81X considered a good starting point for this exercise → small production ongoing
- DPG level tuning & new tracking have knock-on effects on particle flow and POG level tuning
  - EM and Had PF-cluster corrections: can start once tracking and ECAL and HCAL reco & calibrations are stable
    - input: dedicated samples → need a dedicated release
    - output: need to go in DB and release → target the prod release
  - BTV will need to retrain all the algorithms

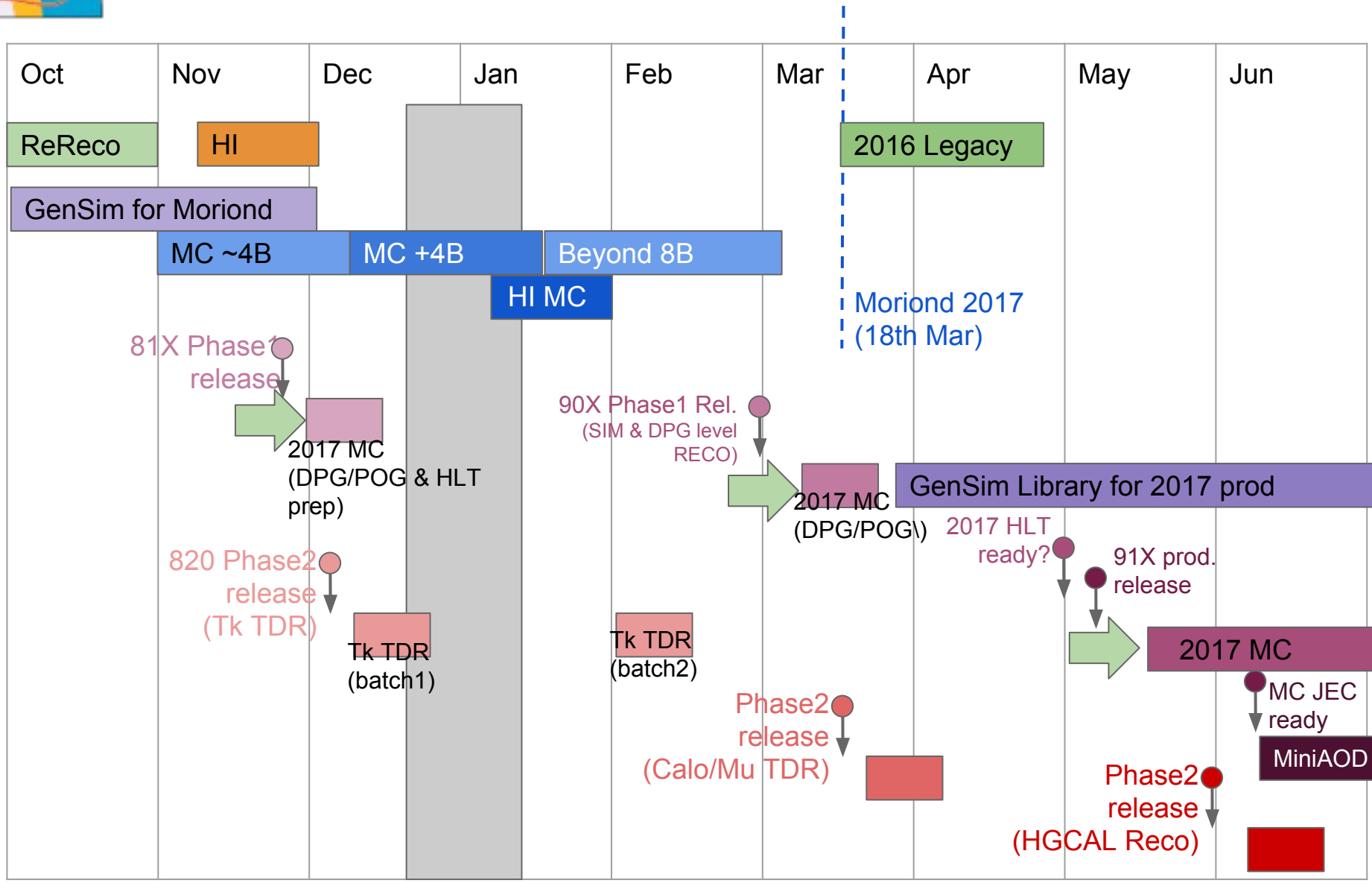
Samples for physics

- 2017 puts us in a new situation:
  - no jumps in energy/luminosity expected (we already have ~40/fb on disk from 2016)
    - **difficult to foresee many new results on 2017 data in time for summer conferences**
    - most of the pressure will be on performance plots → demonstrate functionality of Phase1 upgrade
  - no need for a huge MC library available @ day0
    - needs mostly POG driven at the beginning
    - could profit of “close-to-reality” HL/trigger

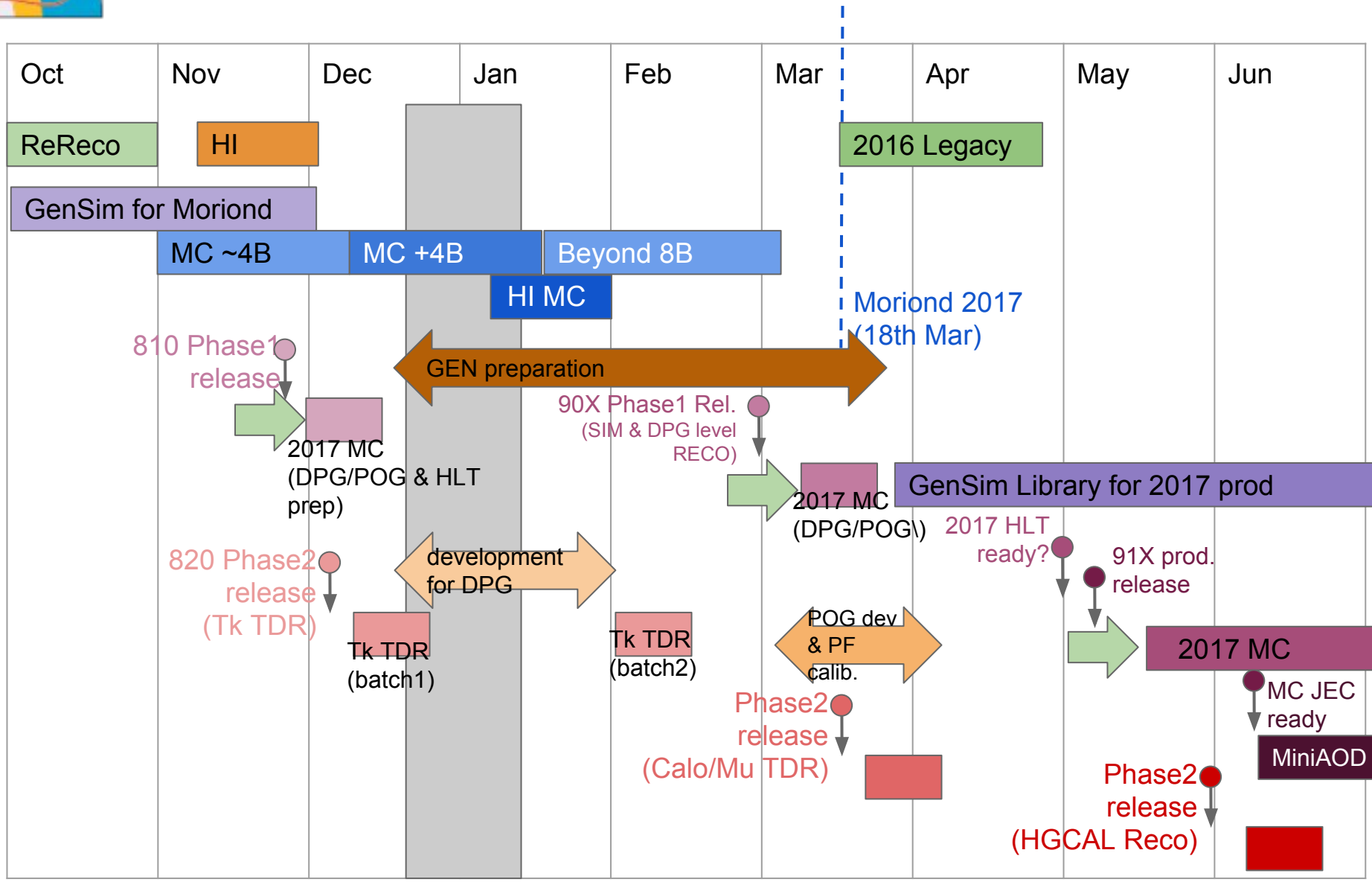
Building a schedule for 2017 production

- t_0=0 ⟶ start of MC production for analysis/performance plot
  - could profit of a library of GEN-SIM produced in advance
- t_0 - ~2 months: need to finalize “low level” reconstruction & calib. (Tracking & DPG local reco)
  - allows POG to derive higher level calibrations & object ID
    - e.g. PF cluster corrections
  - could be used to produce first GEN-SIM library
- 81X: need SIM and DIGI good enough to start dev. of local-reco algos and POG level tuning
  - samples for alignment and calibration development
  - samples for POG development
    - already several requests by POGs for samples with pre-releases

Production schedule



Production schedule



Outlook

Three main topics

- Exploitation of 2016 data
- Planning of 2017 preparation
- Consolidation of operations and workflows

PPD workshop @ CERN, 6-7 Feb

- Focus on consolidation of workflows for 2017 and beyond
  - MC request management & preparation of 2017 campaigns
  - preparation of DQM for Phase-I detector & infrastructure consolidation
  - effectiveness of validation workflows, including Phase-II releases
  - review of secondary datasets and skims
- Would like to use CMS week to prepare the discussion in depth

AlCa/Db workshop [27 Feb → 1 Mar 2017 (indico)]

- Preparation of 2017 activities and the start of data taking
- **Goal: plans for 2017 performance, MC production & data taking operations. Topics**:
  - In-situ calibration & alignment in 2017
    - 1.1 Alignment and calibration for the 2017 data taking startup
    - 1.2 High level calibration for startup: PF, egamma regressions, BTV training
    - 1.3 AlCa updates after the data taking start, w/ emphasis on the phase1 detectors
    - 1.4 Regular Alca updates and their validation: review of the 2016 operations, prospects for 2017
  - AlCa scenarios for 90X Monte Carlo production
    - 2.1 startup MC scenario to be used for TSG studies and
    - 2.2 asymptotic MC scenario: which int-lumi-sensitive conditions need to change between early and late 2017?
  - Automation of calibration measurements (PCL) and AlCa-Reco/RAW usage & content
    - 3.0: Prompt Calibration Loop: review of 2016 operation and prospects for 2017, multi-run harvesting exploitation
    - 3.1: Prompt Calibration Loop: which new workflows can replace manual regular measurements
    - 3.3: review of ALCARECO usage in 2016, and needs for 2017 in terms of event content and rates
    - 3.4: integration and validation of new AlCa streams / AlCaReco workflows, triggers to be used.
    - 3.5: transition away from AFS
    - 3.6: inputs from T0, monitoring
  - O2O and infrastructure: review of 2016 & improvements for 2017
    - 4.1 Experience from 2016, splitting of critical and non-critical workflows
    - 4.2 Offline tests of o2o software as part of the release integration
    - 4.3 Review and updates of AlCa/Db tools and infrastructure for alca/db contacts

CAF & CERN resources

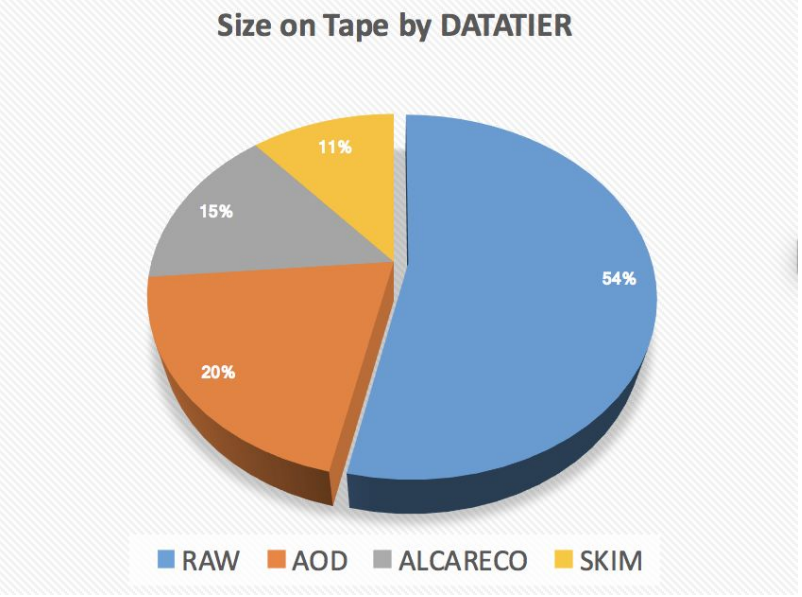
- CPU wise:
  - dedicated machines and batch queues ⟶ no clear feedback or stats on how they are used
  - future of LSF?
- space wise:
  - central EOS space for phedex managed transfers ⟶ DPG coord.
  - EOS space for DPGs (commissioning and AlCa) in /store/group/
    - access requests managed by DPG conveners (see twiki)
  - AFS space: still heavily used by several groups (e.g. Tk alignment)
    - should be phased out before next LS
    - need to understand if use-case can be addressed using EOS
- Need to understand the need for each of these items and plan the evolution of CAF

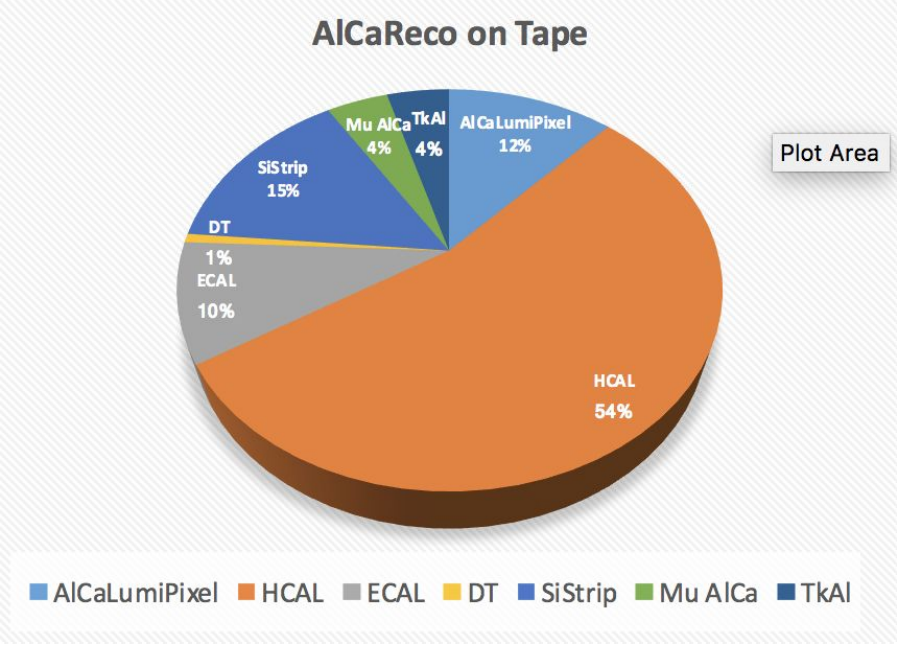
RECO, Skims and ALCARECO

- High LHC availability and trigger rate require careful evaluation of disk&tape resources
  - MiniAOD and AOD are the data-tier serving almost all analysis
    - go to disk & tape
  - RAW: is removed from disk after some time (tape only)
  - RECO is kept on disk only for a limited time and ONLY for prompt
- SKIMS & ALCARECO ⟶ only way to access content not in AOD
  - need careful review of their definition to make them affordable!!!

RECO, Skims and ALCARECO

- Optimization of Skims and AlCaReco might be important in the resource plan for 2017





	Size (PB)
RAW	9.9
AOD	3.7
ALCARECO	2.8
SKIM	2.1

Data Certification: a mature Workflow

- several changes decided during DQM review in LS1 and implemented in 2016
  - new express-based daily certification
  - suppression of offline shifters
- We want DPG feedback on the new approach
  - load on the DPG teams
  - integration with “prompt-feedback group” activity
  - schedule of the updates
  - status of the expertise
  - wishes for central tools → e.g. multi-run harvesting
- Failure scenarios studies:
  - also aim @ assessment of coverage for aspects which might fall in the cracks between DPG and POG expertise

DQM & Machine Learning

- Several groups independently expressed interest in ML applications for DQM and DC workflows
  - several already very advanced among them:
    - Yandex on Data certification
    - IBM on online DQM and slow control monitoring
    - projects proposed in the context of CERN OpenLab
    - students working on test automation
- huge potential for automation and improvements of tests
  - need to bring detector expertise in the game NOW
    - provide actual use cases and data → WE NEED YOU
- Trying to setup a DQM-ML team to steer this activity
  - if interested please contact us….or we will come to you with fancy proposals ;)

Run Registry Workshop

- RunRegistry: crucial tool for Data Certification workflow
- Currently part of WBM effort (although technologically not uniform)
  - WBM is going to be revamped for RunIII ⟶ also RR will need to be updated
  - Vilnius Group (in the process of becoming CMS member) taking responsibility of this project
- Need feedback on the current system to define the future application
  - workshop planned in Spring on these topic
  - will need feedback by all the users to define the requirements

DPG & PPD Organization

- PPD & DPGs: an important link
  - planning: contribution of DPG conveners & DPG coordinator in PPD coordination meetings
  - operation of AlCa, DQM and validation infrastructures
- expertise on the validation & DQM aspects needs to be promoted in the DPGs

- Some criticalities
  - seeking 3rd DQM convener for restart in 2017
  - team to maintain online beamspot workflow
    - essential to provide measurements to the HLT reconstruction
    - needs some DQM expertise

PPD Shifters

- Offline Run Manager:
  - close interaction with Run Coordination in 2016
    - crucial in commissioning periods and setup of special runs
  - small pool of shifters → we struggle to cover the full year
    - very few persons consider it for their central shift contribution
- DQM shifter:
  - supports central shift crew but also sub-system experts
  - pool of shifters is too limited
    - involvement of the sub-system experts would be welcome
- DB expert:
  - monitor status of DBs, not only for AlCa related issues
  - might need to better document its role in the P5 operations

In all cases: thin layer of expertise → CMS needs to find the way to promote participation

Conclusion

- DPG have crucial role in enabling full exploitation of CMS data by physics analysis
  - work on 2016 dataset still not concluded
  - plan for 2017 readiness is under development
- Next few weeks will be crucial for planning the recalibration of the Phase-I detector & consolidation of all data preparation workflows
  - need your feedback and participation in the planning phase