# Ideas and Plans towards CMS Software and Computing for Phase 2

Maria Girone, David Lange
for the CMS Offline and Computing group

April 12 2015

# CMS Upgrade Strategy - Overview

**Upgrades 2013/14 ongoing:**

- Completes muon coverage (ME4)
- Improve muon trigger (ME1), DT electronics
- Replace HCAL photo-detectors in forward and outer (HPD → SiPM)

**Complete original detector**
**Address operational issues**
**Start upgrade for high PU**

LS1 LS2 LS3

**Phase 1 Upgrades 2017/18/19:**

- New Pixels, HCAL SiPMs and electronics, L1-Trigger
- Preparatory work during LS1:
  - new beam pipe
  - test slices of new systems (Pixel cooling, HCAL, L1-trigger)

**Phase 2 Upgrades: 2023-2025 (Technical Proposal in preparation)**

- Further Trigger/DAQ upgrade
- Barrel ECAL Electronics upgrade
- Tracker replacement/ Track Trigger
- End-Cap Calorimeter replacement

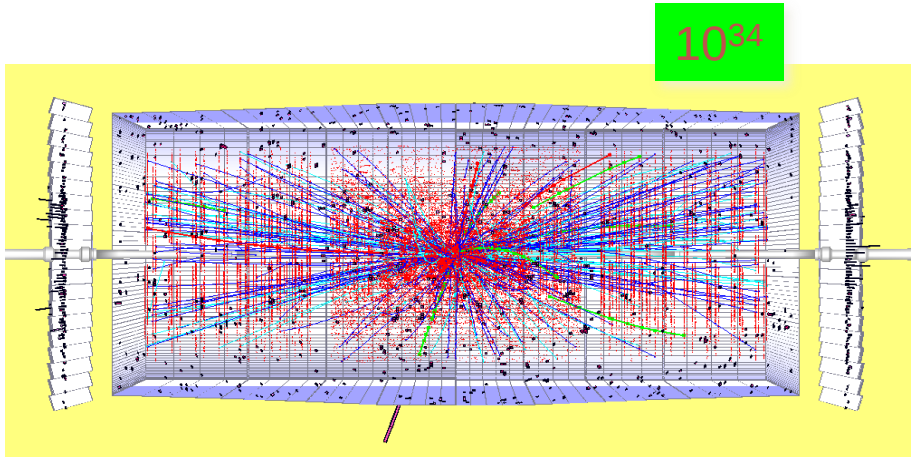
**Maintain/Improve performance at high PU**

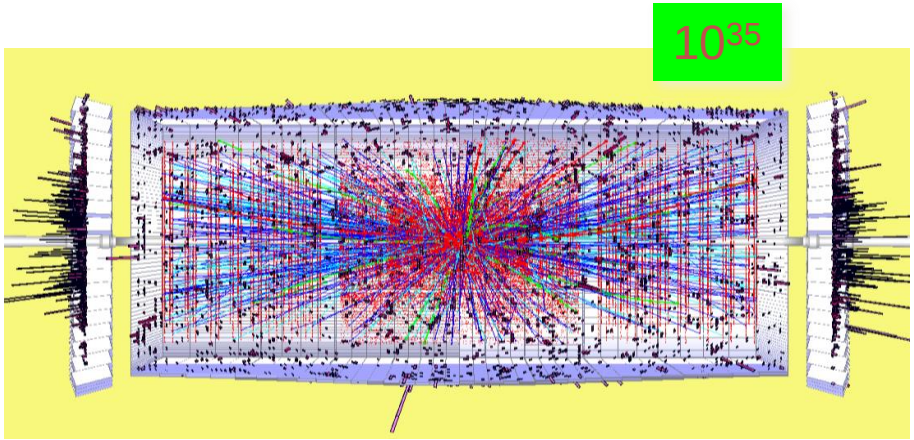
**Maintain/Improve performance at extreme PU. Sustain rates and radiation doses**

# For CMS, The big step up in computing scale is the HL-LHC program and Phase-2 upgrade

CMS at Phase 2 (HL-LHC):

- The CMS detector configuration is still to be determined
- Much higher output rate of trigger (potentially 10kHz)
- Even higher luminosity and pileup (140+ interactions/crossing)




**HL-LHC presents a big increase in the scale of the CMS computing program**

# Fitting Offline and Computing into this plan

- Offline and Computing take an incoming rate and complexity of events to be handled within a resource envelope
  - We are constrained on both ends!

- Talking concretely about Offline and Computing 10 years in advance is somewhat academic. CMS has been writing a technical proposal during LS1 and what follows is mostly based on the draft document
  - Disruptive technology changes can completely alter the landscape
    - We are counting on it
  - HEP doesn't drive the technology
    - We are a small piece of a much larger market. Industrial areas have surpassed us, and even other sciences are catching up

# Resource Growth

- Based on the technology section of the WLCG Computing Model Evolution document we currently we see 25% processing capacity and 20% storage increase per year for the same money
  - Means a doubling every 3(4) years for CPU (storage), so a factor of roughly 8 and 6 by the timescale of CMS Phase-II
    - This assumes flat funding (the best scenario we can hope for)
    - But past experience in technology evolution may not represent future performance

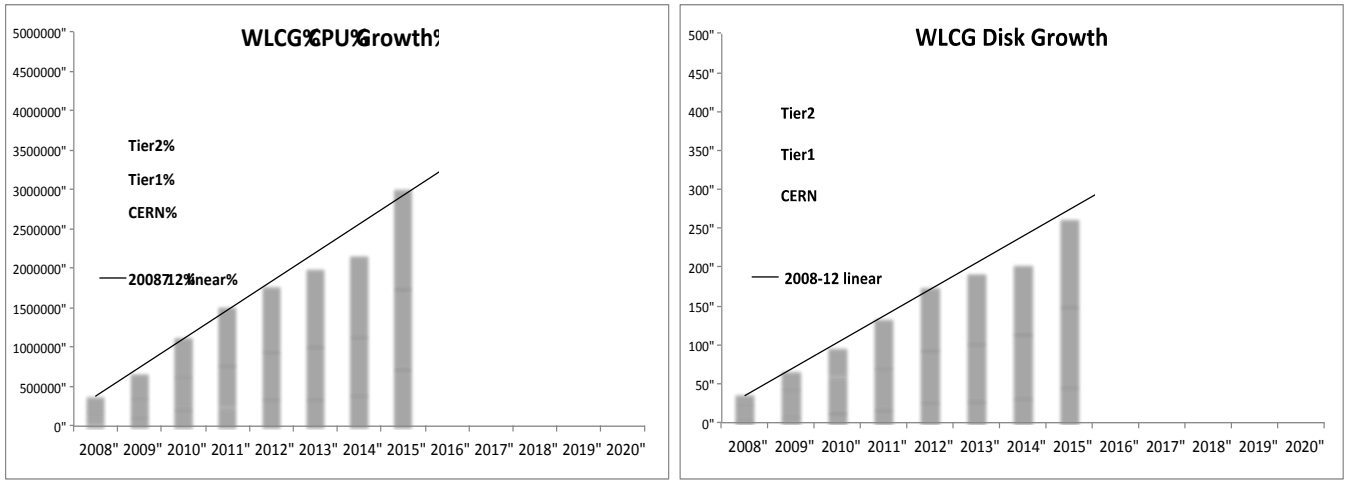



Figure 6.1: Shows the CPU and disk growth through the first 7 years of the program.

# Estimating Resource Needs

- CMS is planning for 5-7.5kHz of data in CMS Phase-II. In this scenario CMS would collect 25B-37B raw events per year
  - Estimating from the current software and using the upgrade simulation we see that each of these events is more complicated to reconstruct and larger than the events we will collect in 2015

| | Detector | Pile-up (Ave./crossing) | Reconstruction time (Ratio to Run 2) | AOD size (Ratio to Run 2) |
|---|---|---|---|---|
| Run3 | Phase 1 | 50 | 4 | 1.4 |
| | Phase-II | 140 | 20 | 3.7 |
| Run4 | Phase-II | 200 | 45 | 5.4 |

# Size of the processing problem

- Factoring in the trigger rate and taking a weighed average of the data and simulation tasks we see the computing challenge is 65-200 times worse than Run2

| Detector | HLT output rate (kHz) | Total |
|---|---|---|
| Phase 1 | 1 | 3 |
| Phase-II (140) | 5 | 65 |
| Phase-II (200) | 7.5 | 200 |

Scale of computing resource needs relative to Run 2 including the increase in projected HLT output rate

- Anticipating a factor of 8 in CPU improvements and a factor of 2 in code improvement, we have a deficit of a factor of 3-15
- Anticipating a factor 6 in storage improvements and having by Phase II events 4-5 times larger, we have still a deficit of 4-5 in storage

# Scale of solutions

- It is unlikely we will get a factor of 5 more money, nor will the experiment be willing to record a factor of 5 lewer events
  - Big improvements are needed

- CMS is investigating many areas.

# Big uncertainty from projecting commodity computing

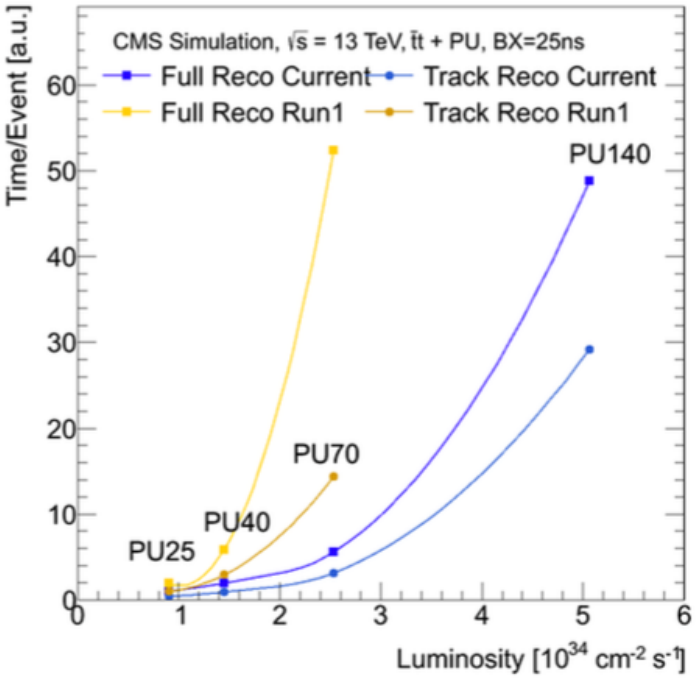
Without a big change in our model for cost for computing, we need to maximize performance/cost of commodity computing

1. Processor evolution: Move towards multi/many core driven by performance/Watt
2. Increasing memory bandwidth limitations
3. Storage and Disk evolution?
4. Networks evolution?

# Big uncertainty from the software even assuming today's computing technologies

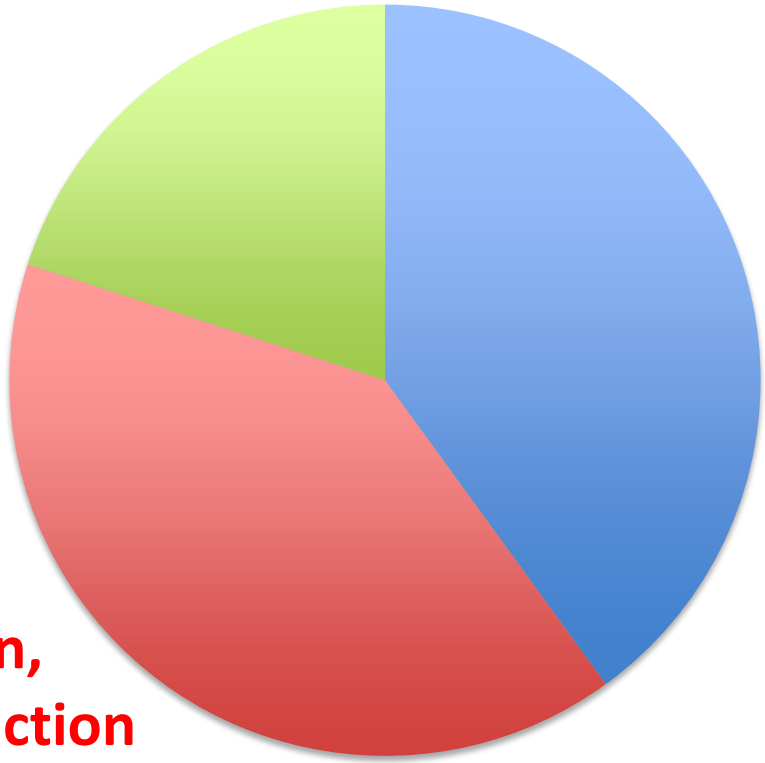
- We don't yet know the CMS Phase-II detector design:
  - Currently a 4x spread in CPU resources needed to reconstruct different possibilities
- We continue to achieve factors of performance improvement in our reconstruction software
  - Modernize and improve our track reconstruction
  - Pileup mitigation techniques continue to improve

LS1 performance improvements



# Targets for improvement in CMS processing capacity

Simulation, user analysis, calibration, monitoring



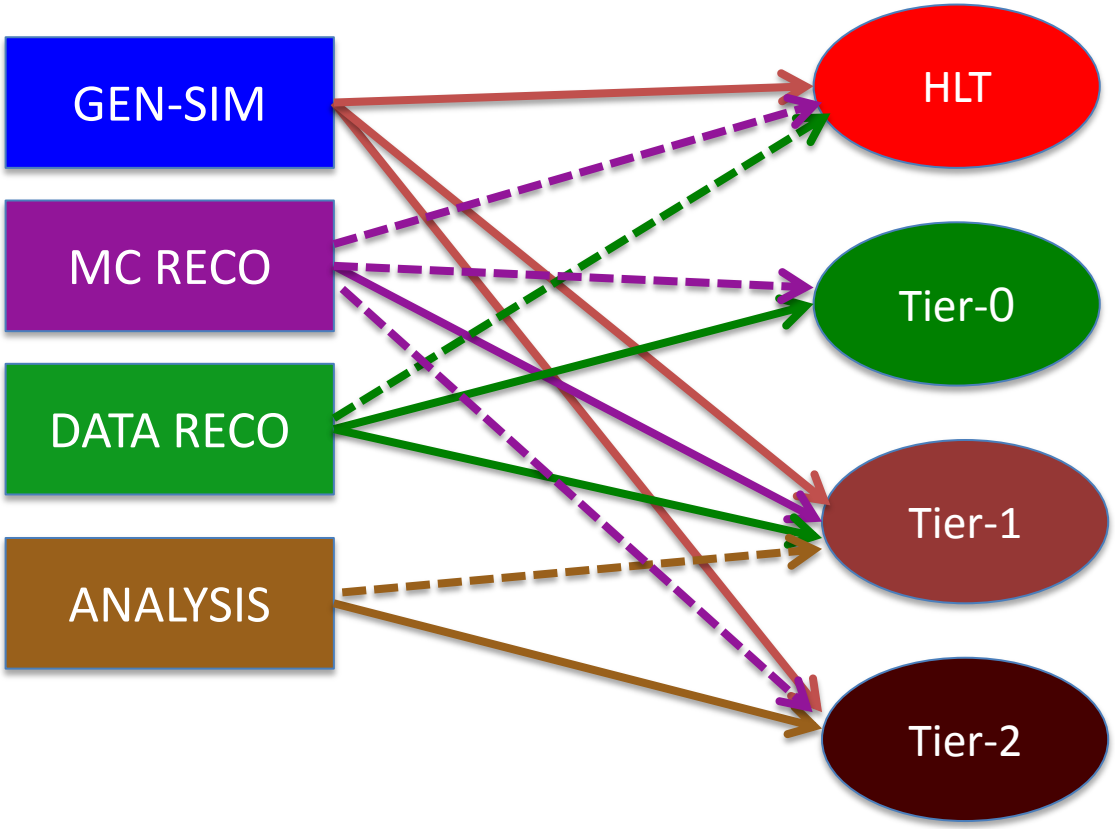
Reconstruction: Improving the number of events that can be reconstructed per computing unit per Swiss Franc is the single biggest savings

Analysis selection, skimming, production of reduced user formats

# Blurring the Site Boundaries

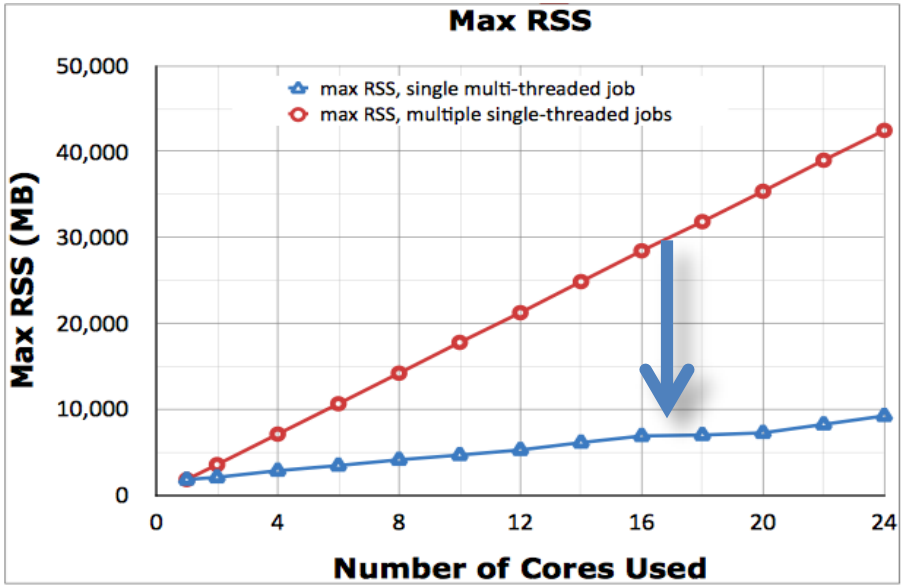
- In Run2 CMS computing resources are intended to work more like a coherent system than a collection of sites with specialized functions
  - Improved networks have been key to this

- Data Federation will make CMS datasets available transparently across the Grid
- One central queue for all resources and all workflows
- The HLT farm is now an integrated resource when we are not in data-taking mode



# Multi-threaded CMSSW Framework now in production commissioning

- We have developed next-generation framework for CMSSW based on a multi-threading approach
  - This gives CMS a natural platform to investigate highly parallelizable algorithms
- Short term focus: This Framework allows us to process higher Run 2 trigger rates efficiently and to adapt to computing technology trends
- Current results:
  - Good scaling in CPU performance up to at least 8 cores in a job
  - Substantial application memory savings in CMS reconstruction
- A development plan is in place to modify the FWK to scale up the threading performance to much higher levels over the next years.




Detailed reports during CHEP

# Big changes in CMS analysis data model: MiniAOD

- CMS has developed a new user analysis data format that is 5-10x smaller than our Run 1 format
  - Design and content based on Run 1 experience across analysis groups
  - Re-evaluation of physics requirements on stored objects and on variable precision.
  - Targets most analysis needs (80-90%)
- Potential for big analysis improvements in Run 2:
  - Increased analysis agility in limited computing resources: Recreating miniAOD is much faster than rerunning the reconstruction for a vast range of performance improvements
  - Substantial robustness improvements in analysis results by reducing user data processing
- Provides a basis for R+D towards more I/O performant analysis data models

Detailed reports during CHEP

# HL-LHC focused R+D Activities and ideas (Reports during CHEP on many areas)

- Towards using heterogeneous many-core resources
  - We Prototype tracking demonstrator for Intel Phi in development: Initial results from generic algorithm encouraging
- Trigger on many-core or specialized resources
- Simulation (G4 or beyond) on heterogeneous resources
- Resource scheduling, I/O, event processing frameworks for heterogeneous resources, parallel processing models
  - Developing a computing model simulation to study optimizing data access
- Improving tracking and other algorithms at high pile-up
  - Continuous improvement activities
- Evaluating improved cost/performance of using specialized centers for dedicated workflows

# HL-LHC focused R+D Activities (Reports during CHEP on many areas)

- Evaluating "big data" data reduction and selection techniques for I/O intensive analysis procedures
  - Efficient physics object skimming/tagging would also promote re-use of selection criteria
- Developing tools and infrastructure (profilers, dev. tools, etc.)
  - Prototyping container technology (e.g., DOCKER) for CMSSW
- Evaluating metrics for performance per power use, in addition to simply raw performance
- Investigating data analytics techniques
  - Ideas for next generation data popularity

# Outlook

- CMS is facing a a much more complex computing program for Phase-II due to detector changes, increased trigger rates and event complexity
  - Technology evolution alone will not close the gap between naïve resource extrapolations based on constant funding and resource needs for the HL-LHC era
  - In addition we need to evolve to be ready to effectively use future commodity computing technologies to be cost effective
- CMS is investigating a number of R&D areas, and is seeking to extend our collaboration between experiments, sites, and groups
- We have organized our R&D around weekly meeting (these are now open outside CMS)
  - Active collaboration between CMS and TechLab for R+D on (CMSSW) benchmark performance measurements on new architectures
  - CMS members closely following HEP software foundation (including via startup team)