The Role of Effective Event Reconstruction in the Higgs Boson Discovery at CMS

Slava Krutelyov (Texas A&M)
for CMS Collaboration

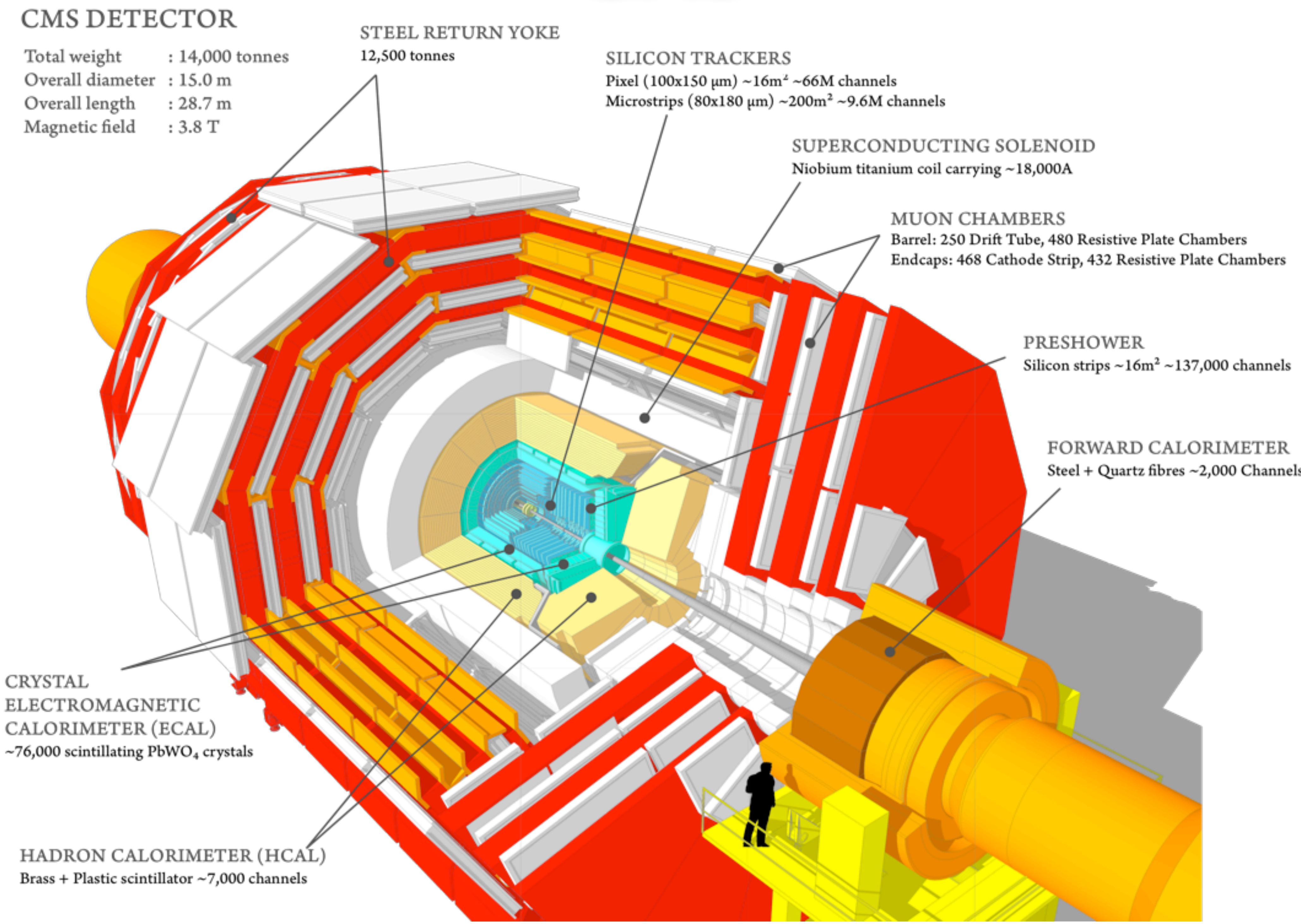
CHEP 2013
Amsterdam, Netherlands

October 14, 2013

Outline

- CMS physics and Higgs boson discovery
- Event reconstruction
- Solutions to effective event reconstruction
  - ✓ Flexible tracking
  - ✓ Software optimization
  - ✓ Avoid re-reconstruction
  - ✓ Better combination of information: multivariate techniques
- Summary and outlook

CMS



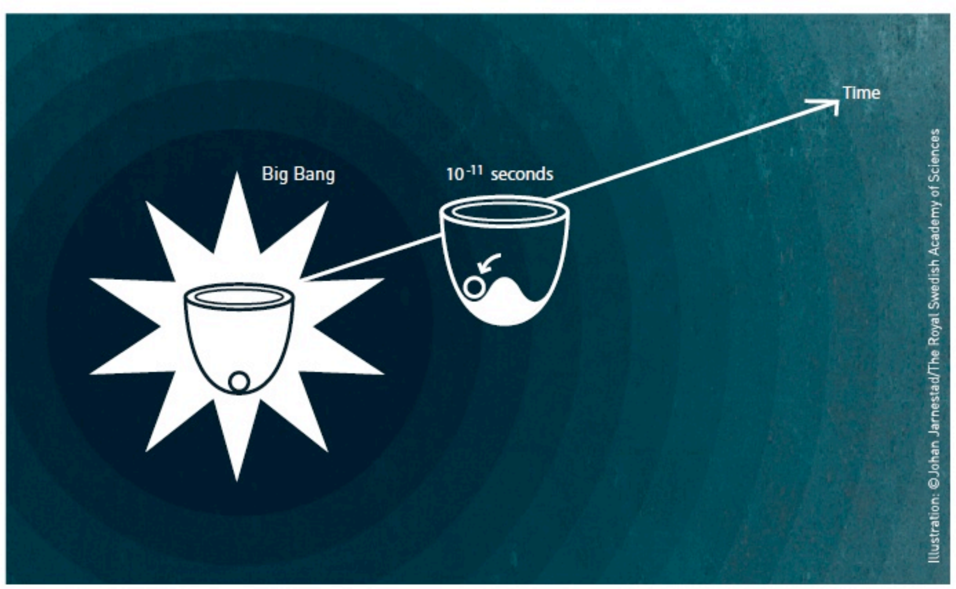
- CMS Collaboration
  - Around 4300 active people from close to 200 institutes in 42 countries
  - 2700 thousand physicists (+ 700 undergraduates); 900 engineers

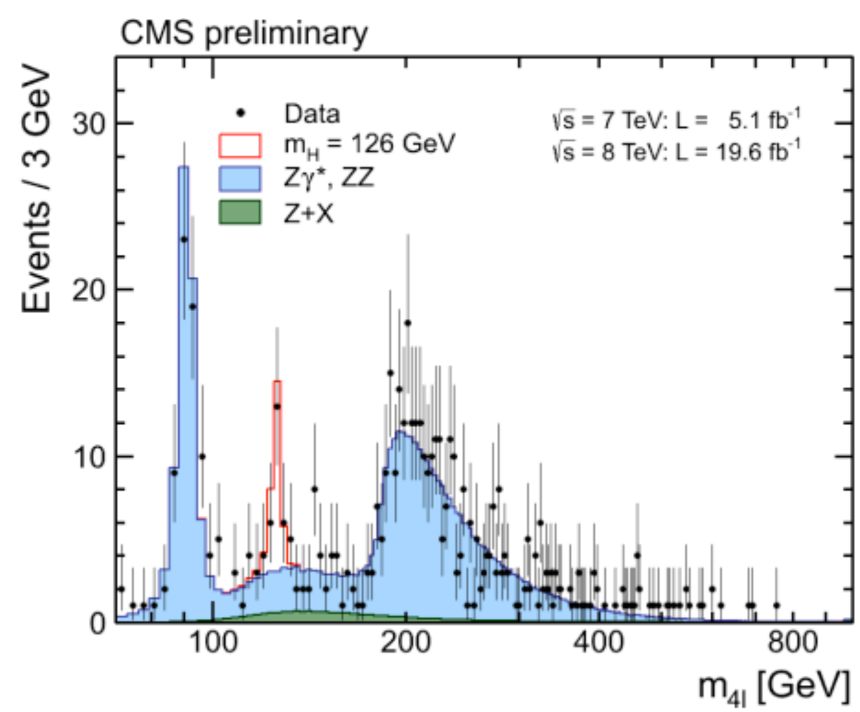
Scientific Success of CMS

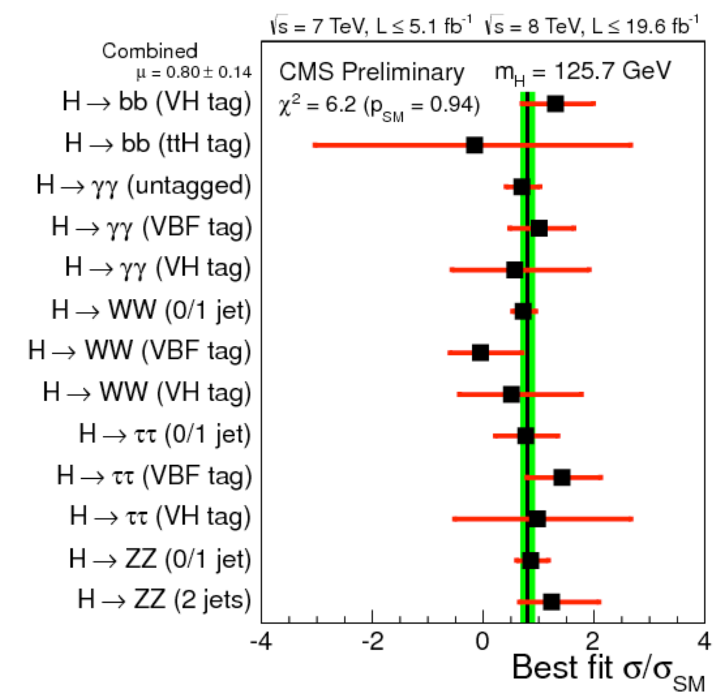
- Success is best measured with publications in journals: close to 300
  - **https://cds.cern.ch/collection/CMS%20Papers?ln=en**

- Most recent high mark is the contribution to the Nobel Prize in 2013
  - The Nobel Prize in Physics 2013 was awarded jointly to François Englert and Peter W. Higgs "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider"

The 2013 Nobel Prize in Physics - Press Release



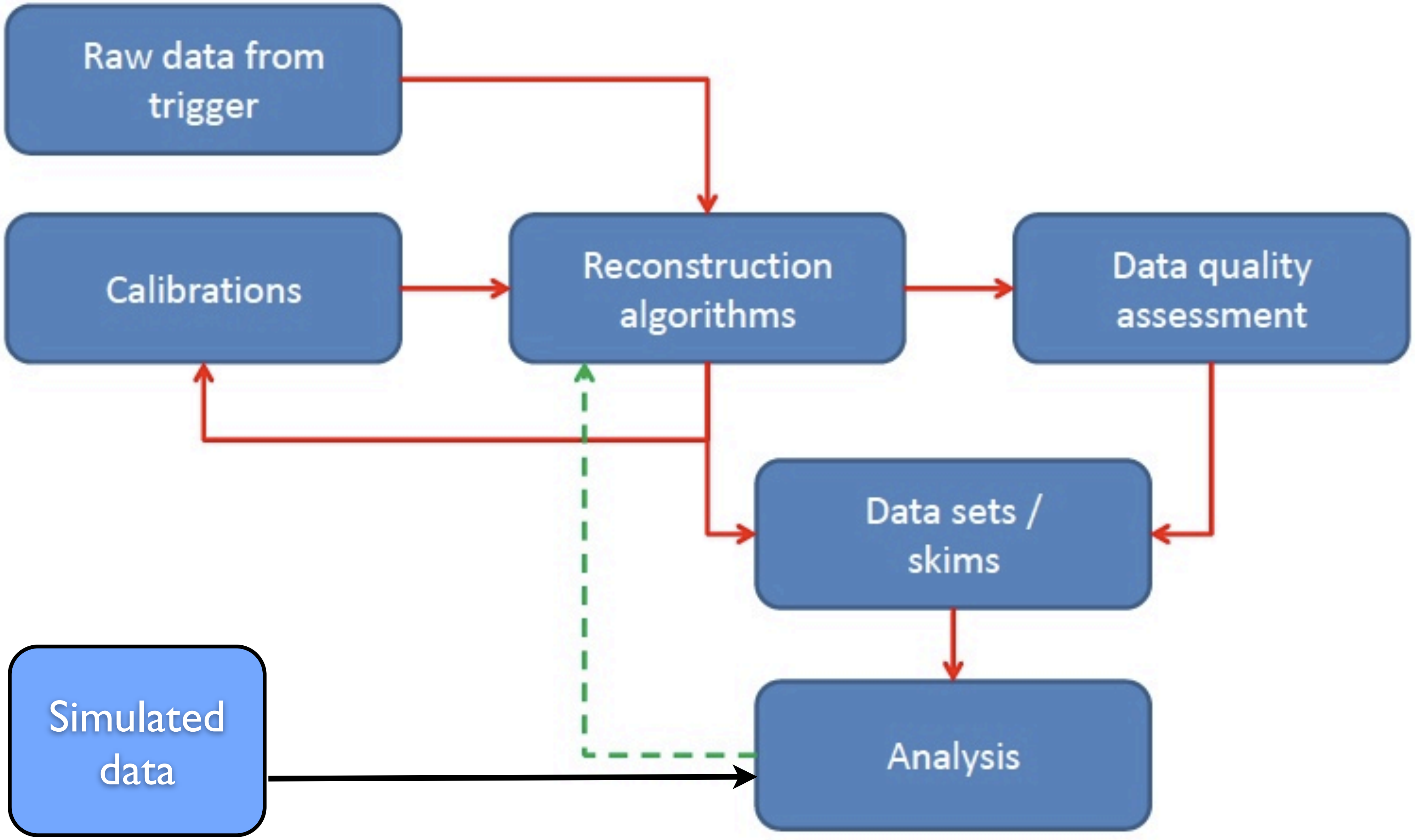




Contributions from many

- This success comes from effective work of all areas, to name some:
  - the detector groups
  - collision event selection in real time (trigger)
    - **See talk by D. Trocino on Thursday on CMS High Level Trigger**
  - offline software
    - **See talks/posters by C. Jones, D. Funke, A. Giammanco, P. Lujan, others**
  - computing infrastructure
    - **See, e.g., talk by O. Gutsche on CMS Computing in Run 1**
  - data quality assurance, detector data calibration
    - **See, e.g., talk by F. De Guio on data quality monitoring**
      - **talk by R. Castello on alignment and calibrations**
  - physics analyses
  - ...

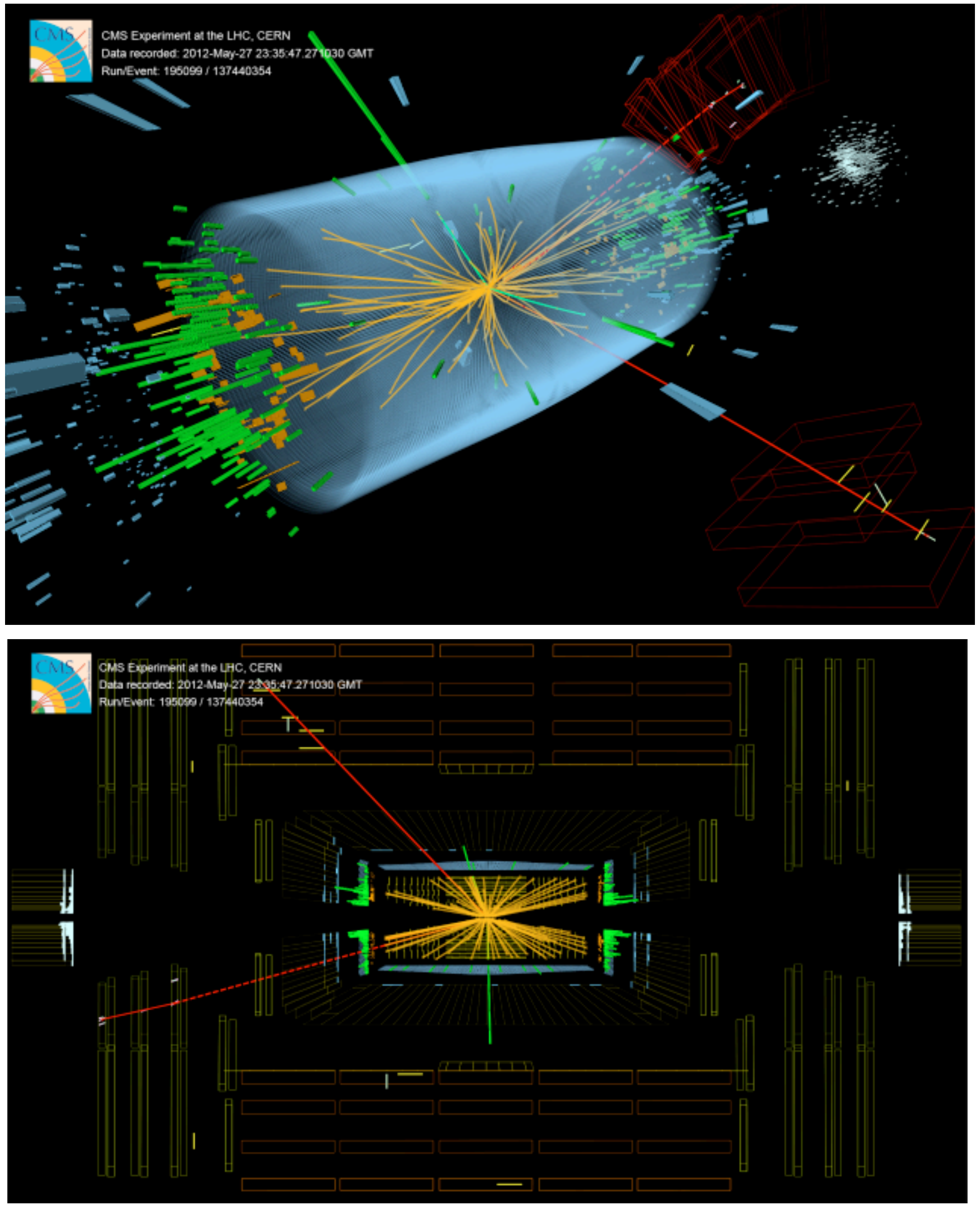
Interconnects

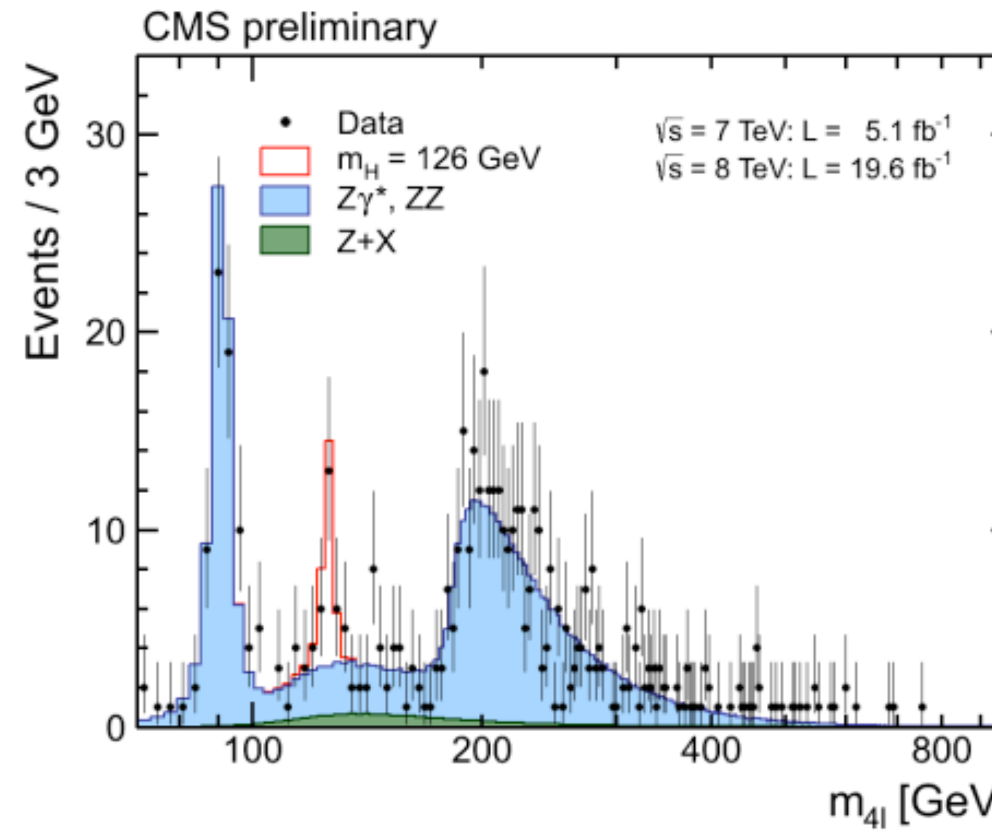


- It takes quite a few logical steps to get from the raw data or simulated inputs to the pretty plots
- Efficient event reconstruction is the cornerstone

Event Reconstruction

CMS-PHO-EVENTS-2012-007





- A picture of an actual 2e2μ event with mass ~126 GeV
- Not all events are this exciting, need to reconstruct all to see if important
  - ✓ Inputs to reconstruction: about 5 billion collision events in 2012

Ingredients for Effective Reconstruction

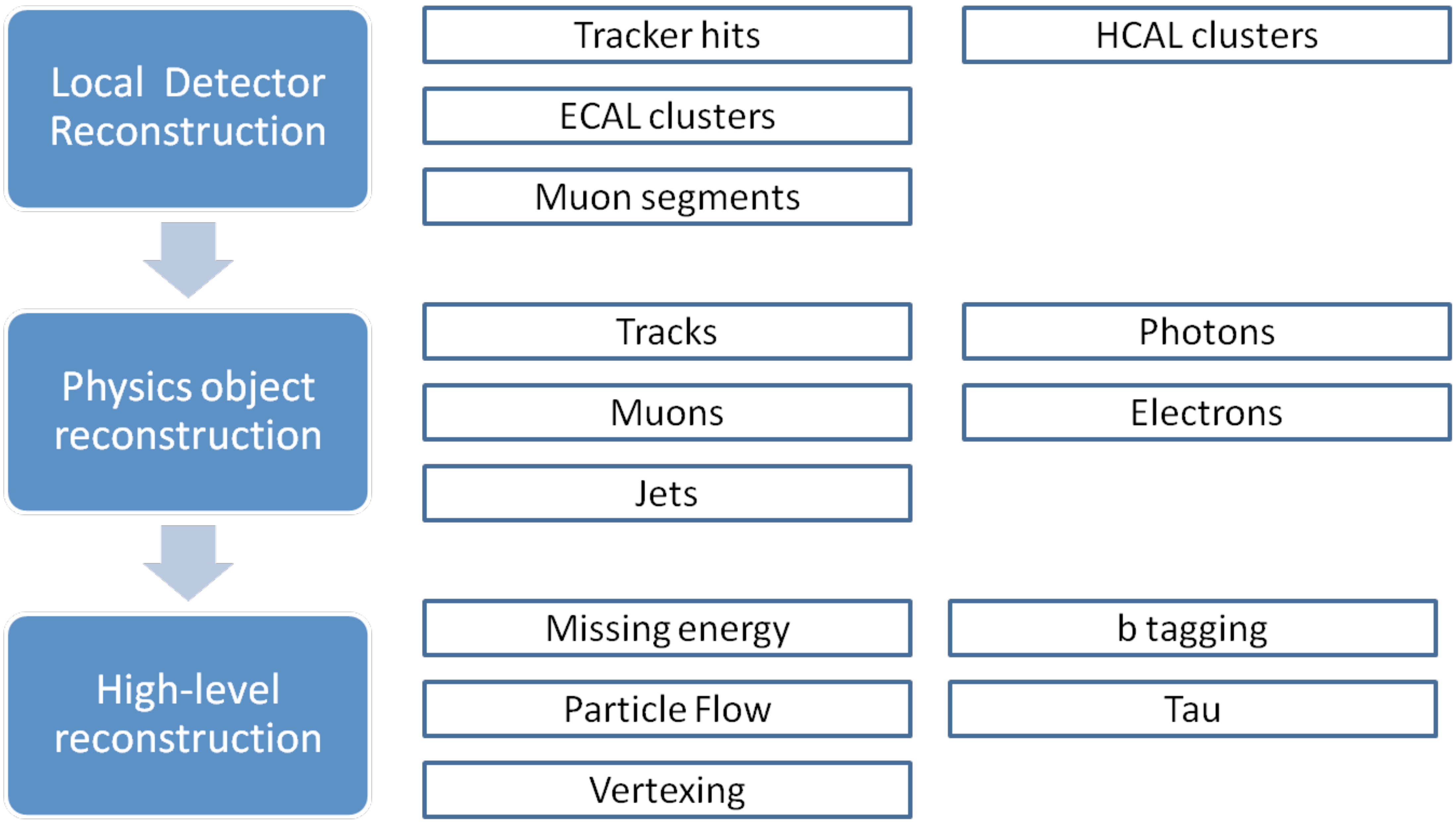
effective (ɪˈfɛktɪv)

— ***adj***

1.productive of or capable of producing a result

- We have to manage in constraints of limited resources for the given input data volume
- Goal is to provide the best quality and reconstruct all inputs
  - about 5 B collision events recorded; about 6 B events simulated ==> ~10^{10} events
- Constraints: time, CPU, and storage
  - ~25 thousand cores available (~160 kHEP-spec06) in Tier-0+Tier-1
  - ~16 PB disk and 37 PB tape at T0+T1 in 2012; 15 PB more disk at Tier-2
- Solutions
  - Flexible software
    - **Tune up at configuration level**
  - Optimized software
    - **Remove useless operations; benefit from compilers, use CPU vectorization**
  - Minimize the number of times events should be (re)reconstructed
    - **Prompt reconstruction**
  - Store enough information, take the most out of it
    - **Small enough Analysis Objects Data (AOD)**

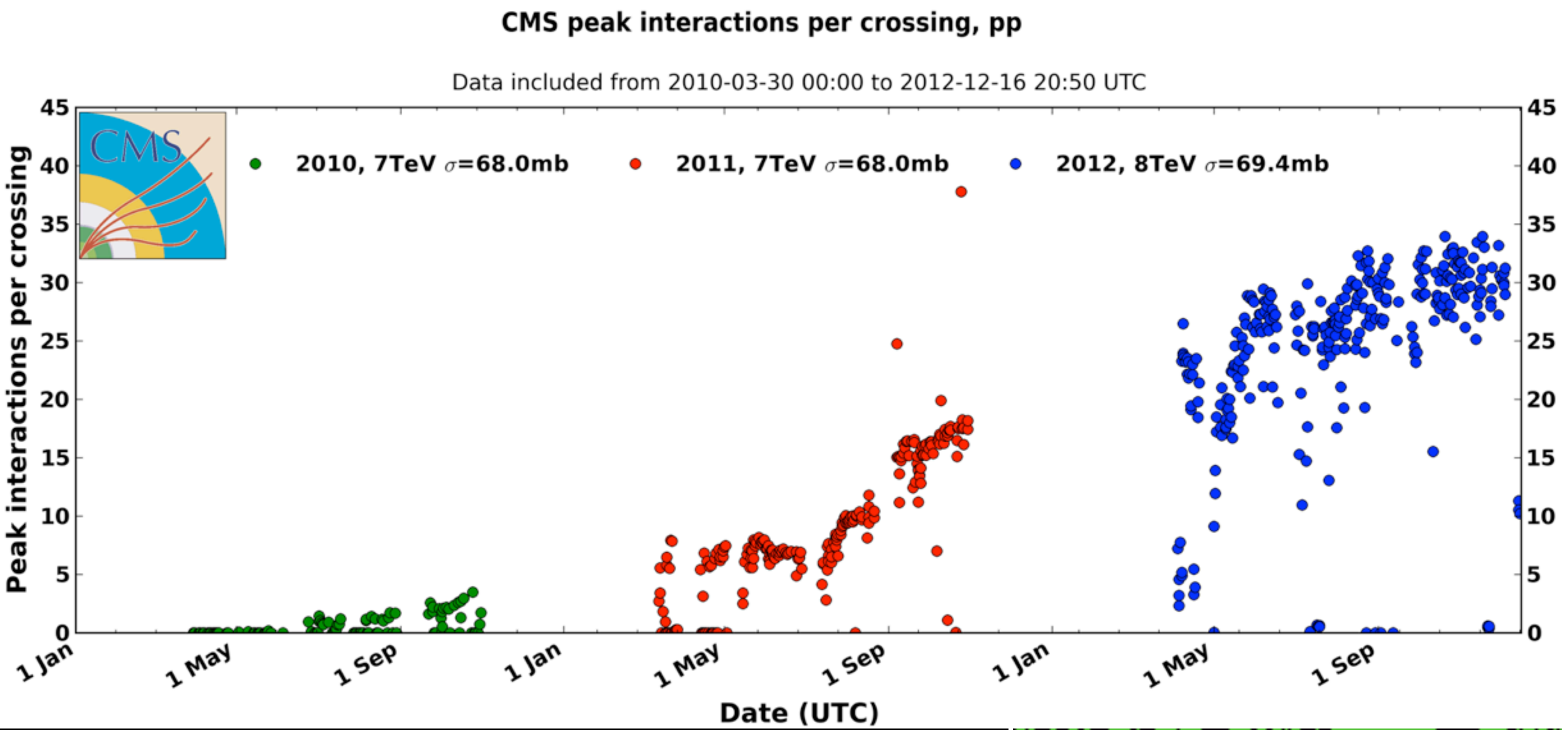
Offline Event Reconstruction

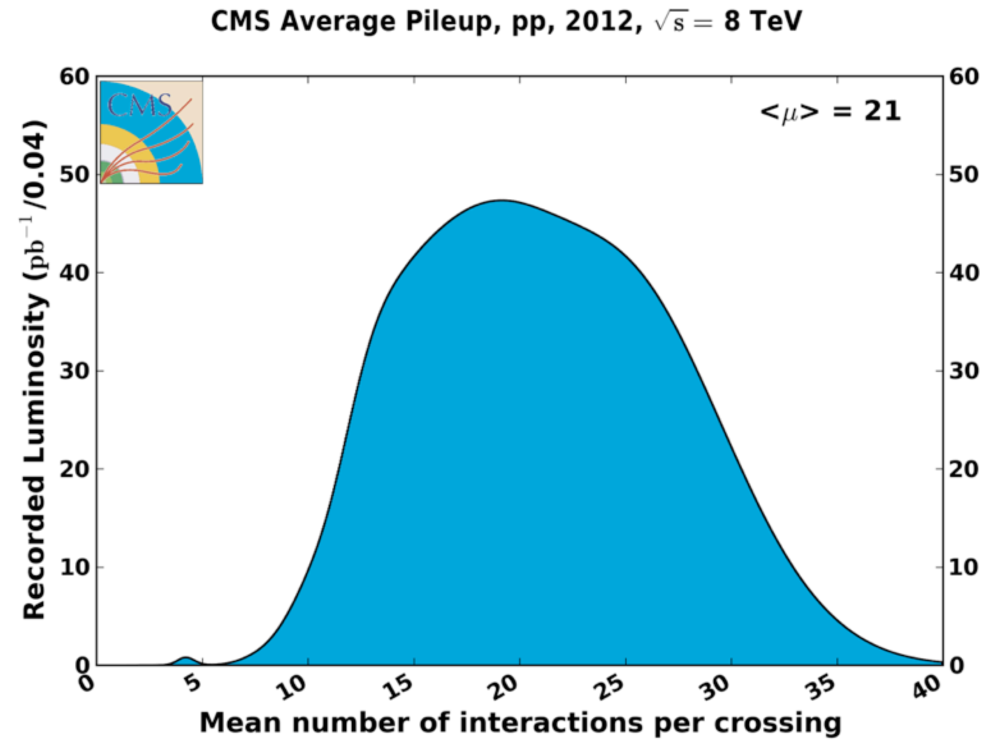


- Higher level objects require inputs from previous steps

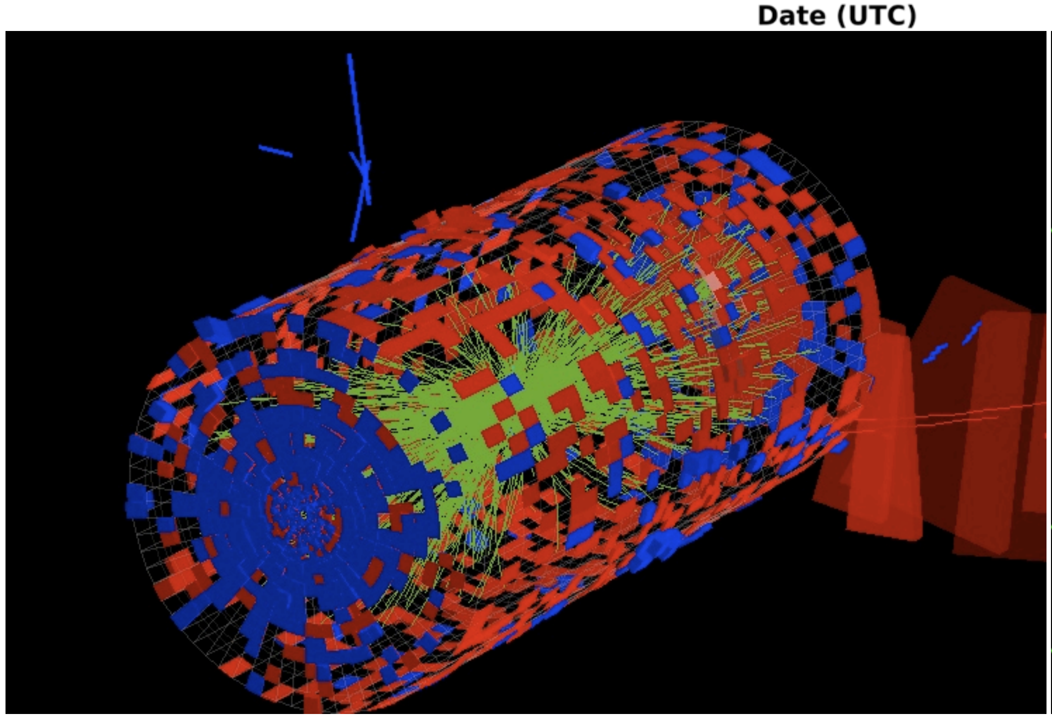
Pileup: major challenge in 2012

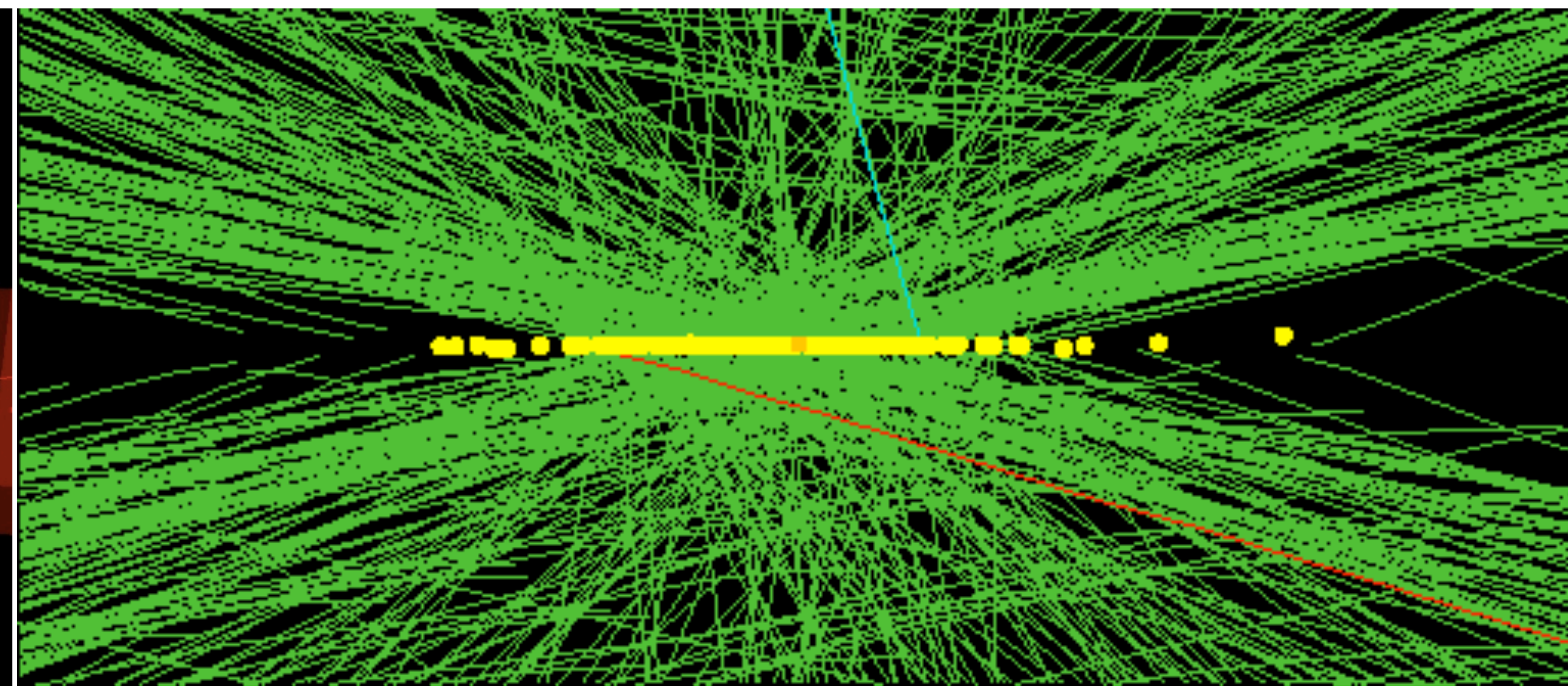
CMS Lumi (Public)





CMS-PHO-EVENTS-2012-006



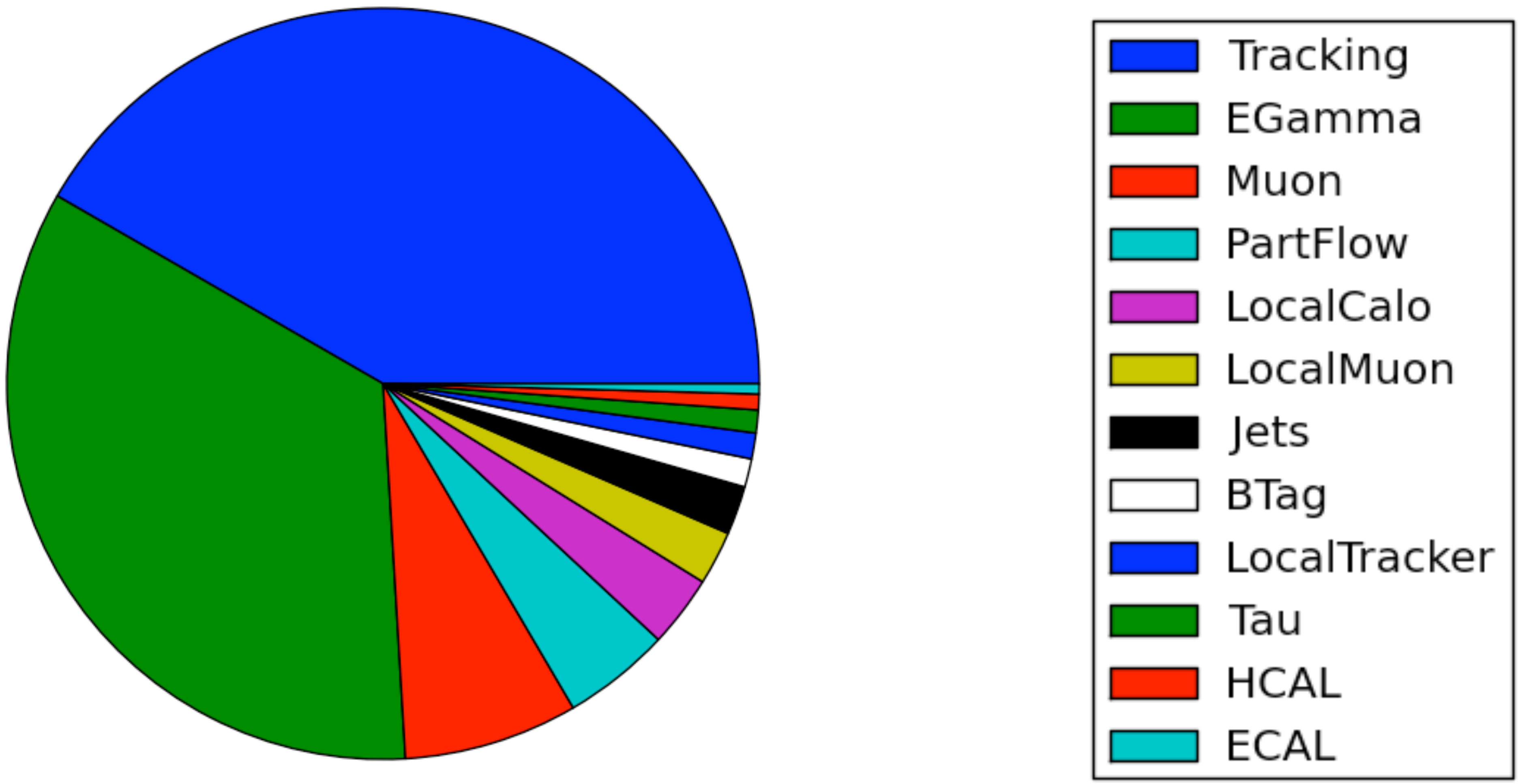


- The high-PU event is just an extreme example
- Not surprising, most of the effort to reconstruct this is spent in tracking

Reconstruction: CPU consumption

Top-quark pair production events

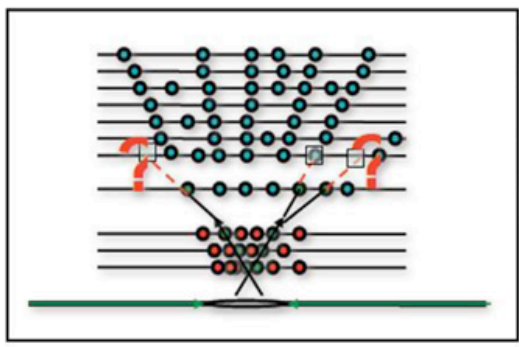
CMS CR-2011-002



- NB: timing breakdown before improvements made for 2012 data taking
- General tracking takes about 50% of CPU
- EGamma is dominated by special tracking for electrons and conversions ($\gamma \rightarrow ee$)

Iterative Tracking

- Track reconstruction goes in several iterations (6 iterations in 2012)
- Hits from tracks found in previous iterations are not used later
- Each iteration has seeding, pattern recognition, and refit steps
- We had this structure setup and working from day one of LHC collisions

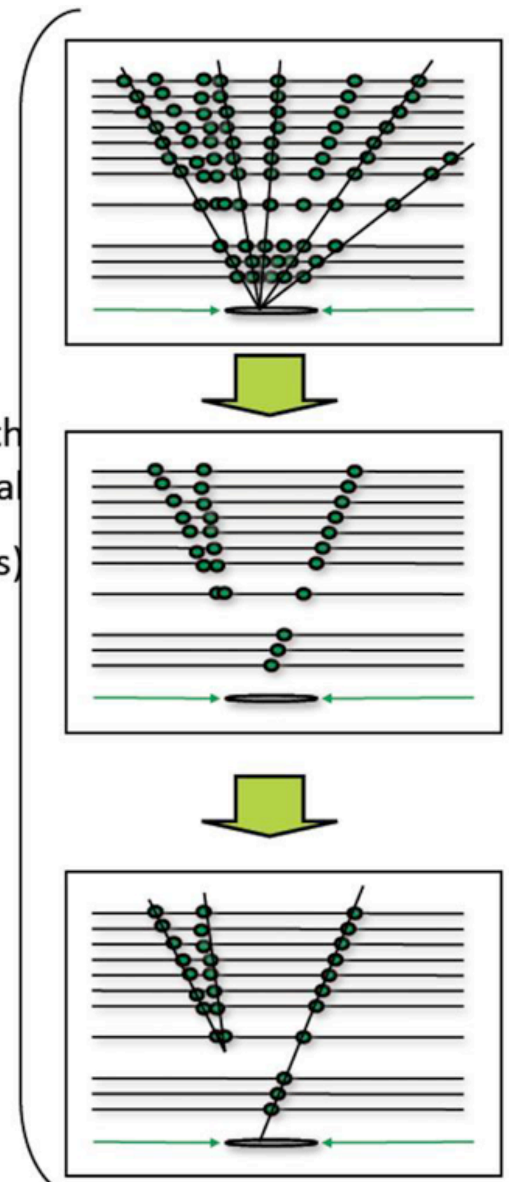


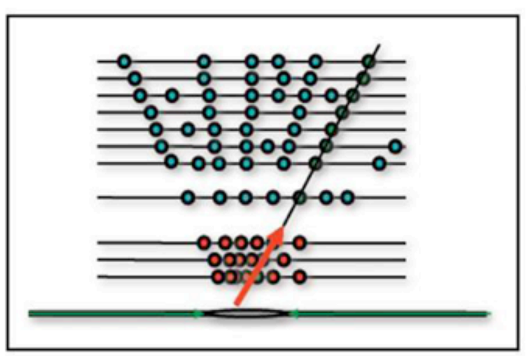
Seeding starts from innermost pixel layers.

Inside-out trajectory building

Iterative tracking with hits-removal

(6 iterations like this)



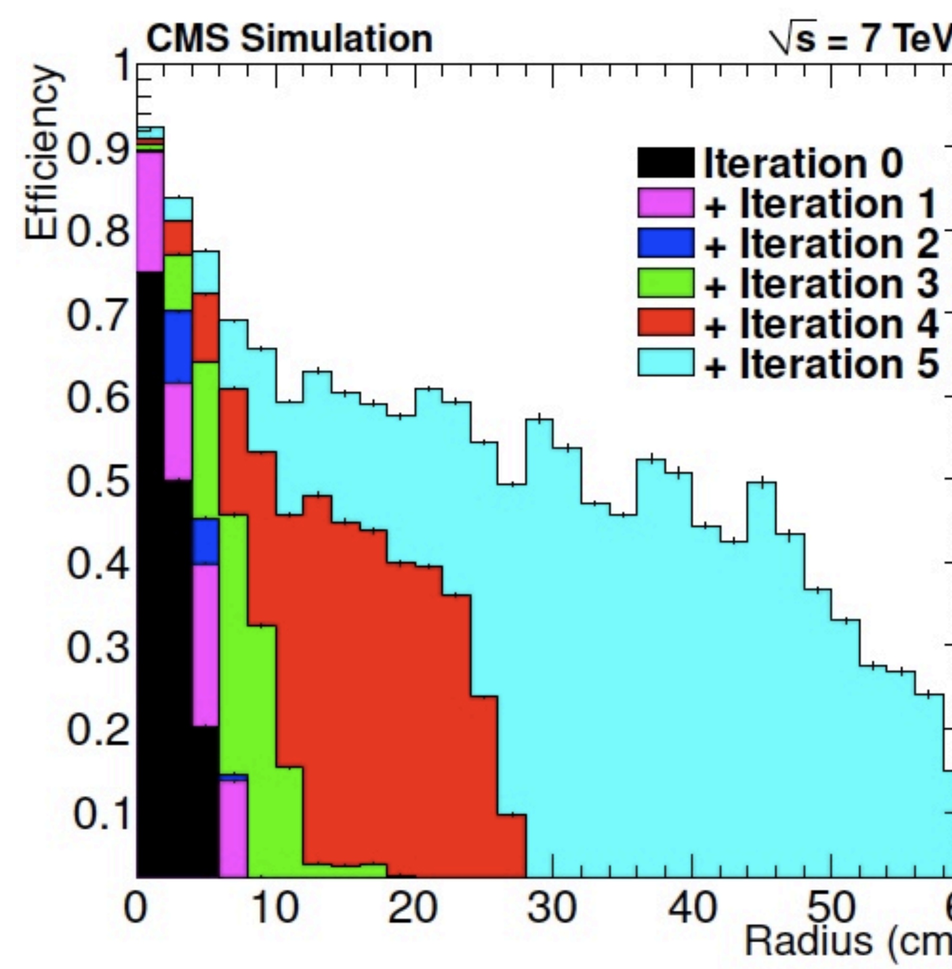


Final fit using **Kalman Filter/ Smoother**.

Parameters propagated through magnetic field inhomogeneities using **Runge-Kutta propagator**

Track Parameters (q/p,eta,phi,dz,d0)

CMS TRK-11-001



Iterative Tracking: Tunable Setup

- We have successfully managed to tune the iterative tracking steps to fit the tracking to the available CPU resources
  - ✓ It worked when we prepared to high-pileup environment in 2012
  - ✓ Similar strategy will be applied for the high-luminosity LHC with 140 pileup
- Below are some examples in settings

2011 Data taking settings

#step	seed type	seed subdetectors	P_T^{min} [GeV/c]	d_0 cut	z_0 cut
0	triplet	pixel	0.8	0.2 cm	3.0σ
1	pair	pixel/TEC	0.6	0.05 cm	0.6 cm
2	triplet	pixel	0.075	0.2 cm	3.3σ
3	triplet	pixel/TIB/TID/TEC	0.25-0.35	2.0 cm	10.0 cm
4	pair	TIB/TID/TEC	0.5	2.0 cm	12.0 cm
5	pair	TOB/TEC	0.6	6.0 cm	30.0 cm

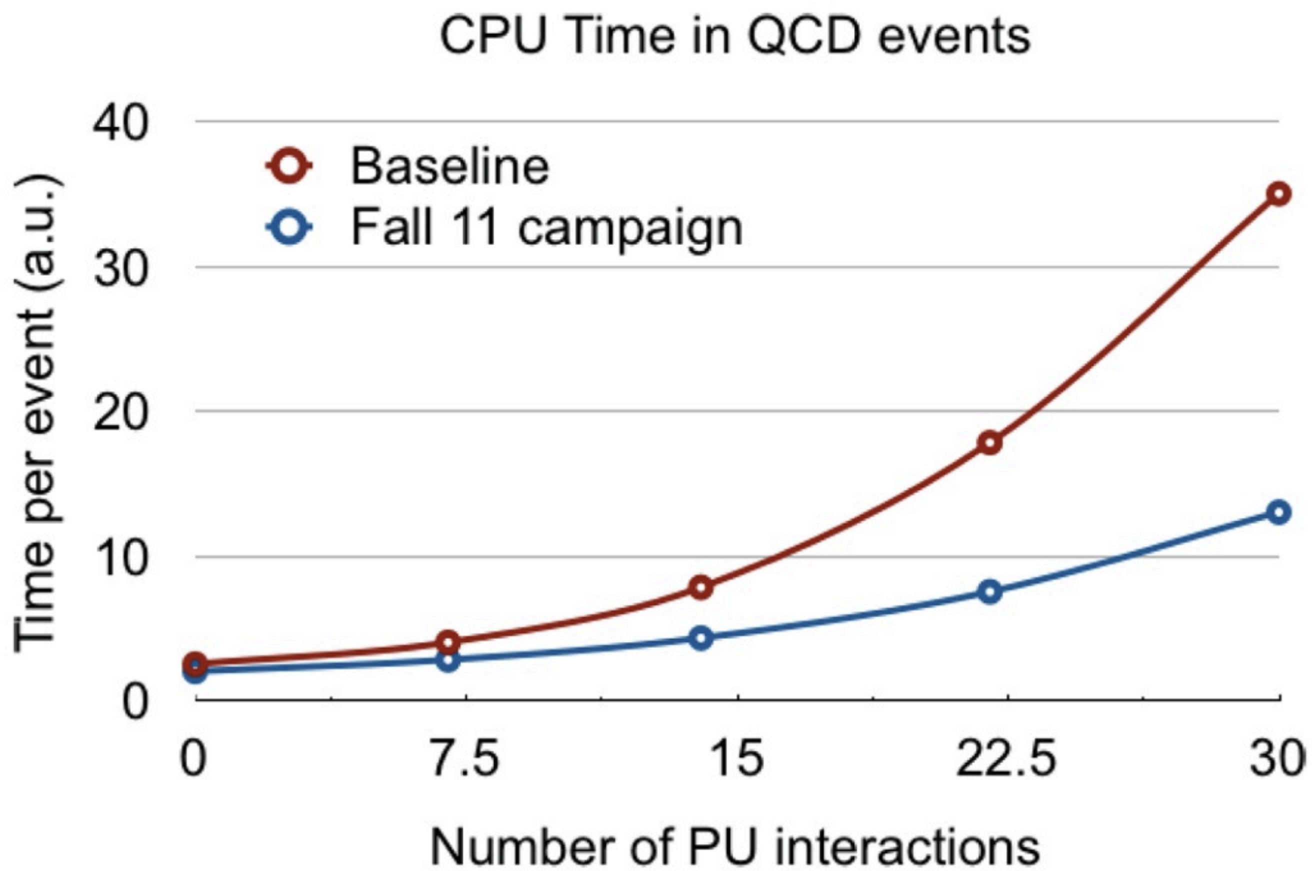
2012 Data taking settings

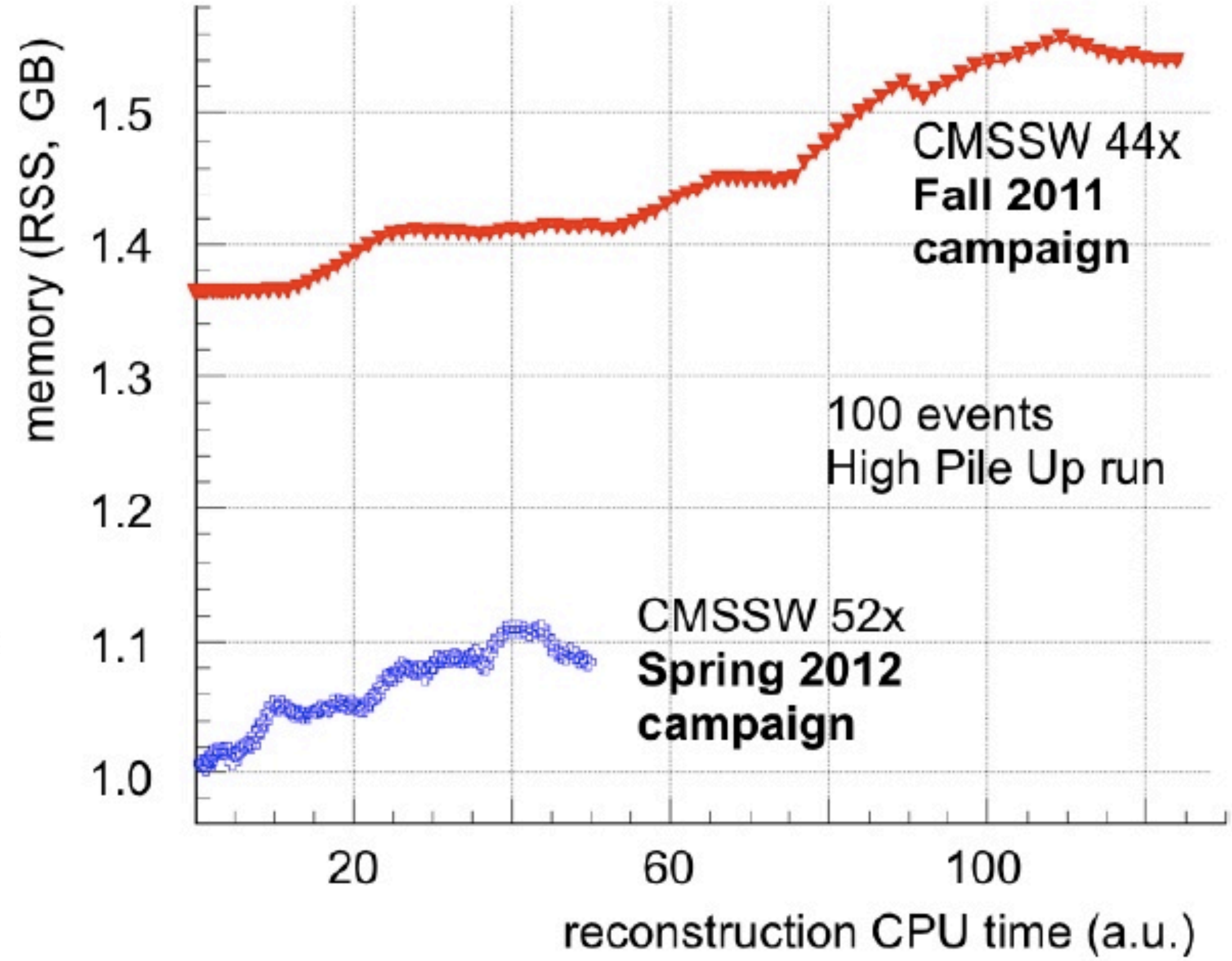
#step	seed type	seed subdetectors	P_T^{min} [GeV/c]	d_0 cut	z_0 cut
0	triplet	pixel	0.6	**0.02** cm	4.0σ
1	triplet	pixel	0.2	**0.02** cm	4.0σ
2	pair	pixel	0.6	**0.015** cm	0.09 cm
3	triplet	pixel	**0.3**	1.5 cm	**2.5σ**
4	triplet	pixel/TIB/TID/TEC	**0.5-0.6**	1.5 cm	10.0 cm
5	pair	TIB/TID/TEC	0.6	2.0 cm	10.0 cm
6	pair	TOB/TEC	0.6	2.0 cm	30.0 cm

J. Phys.: Conf. Ser. 396 022044

Software Optimization

J. Phys.: Conf. Ser. 396 022044



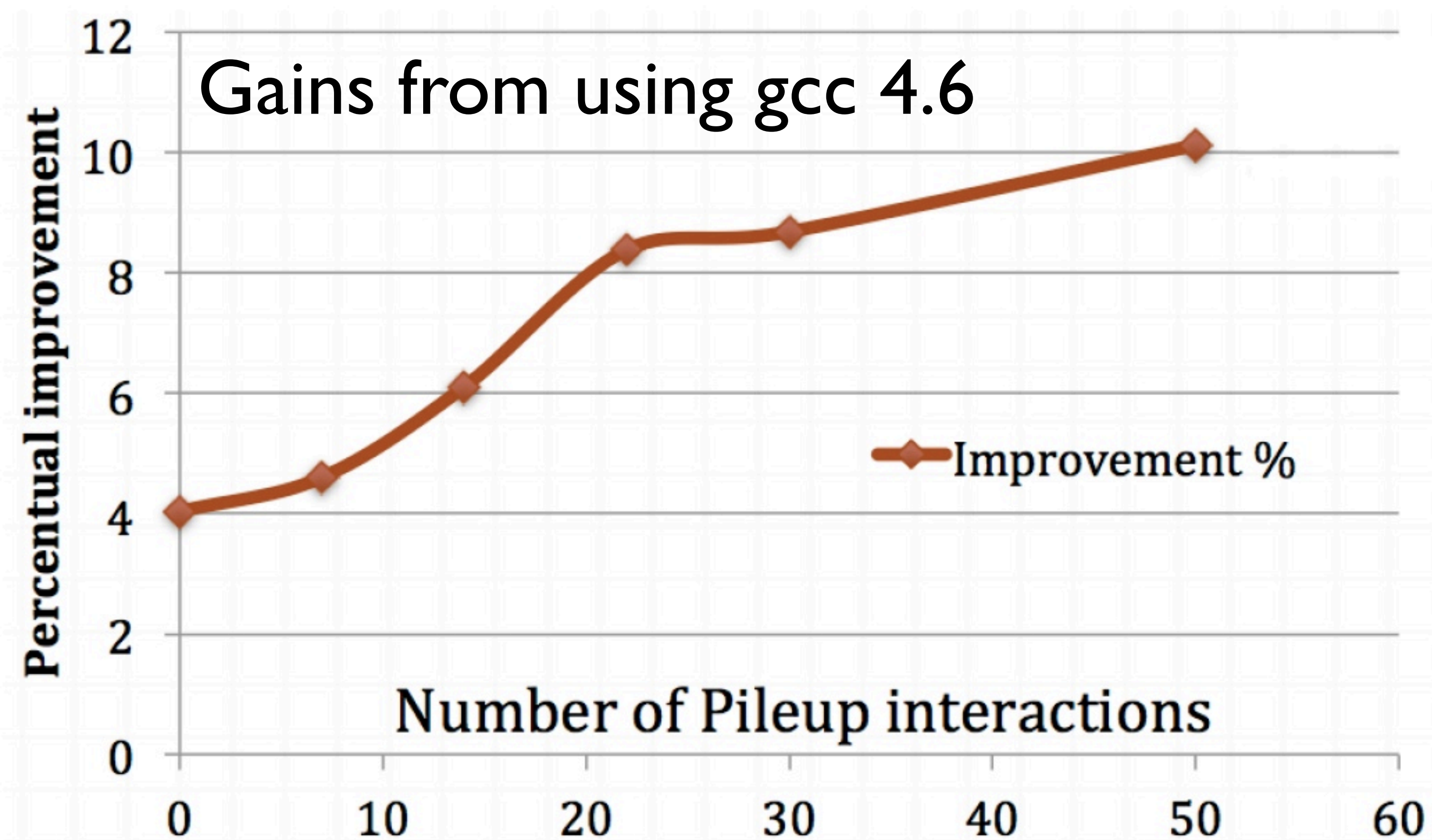


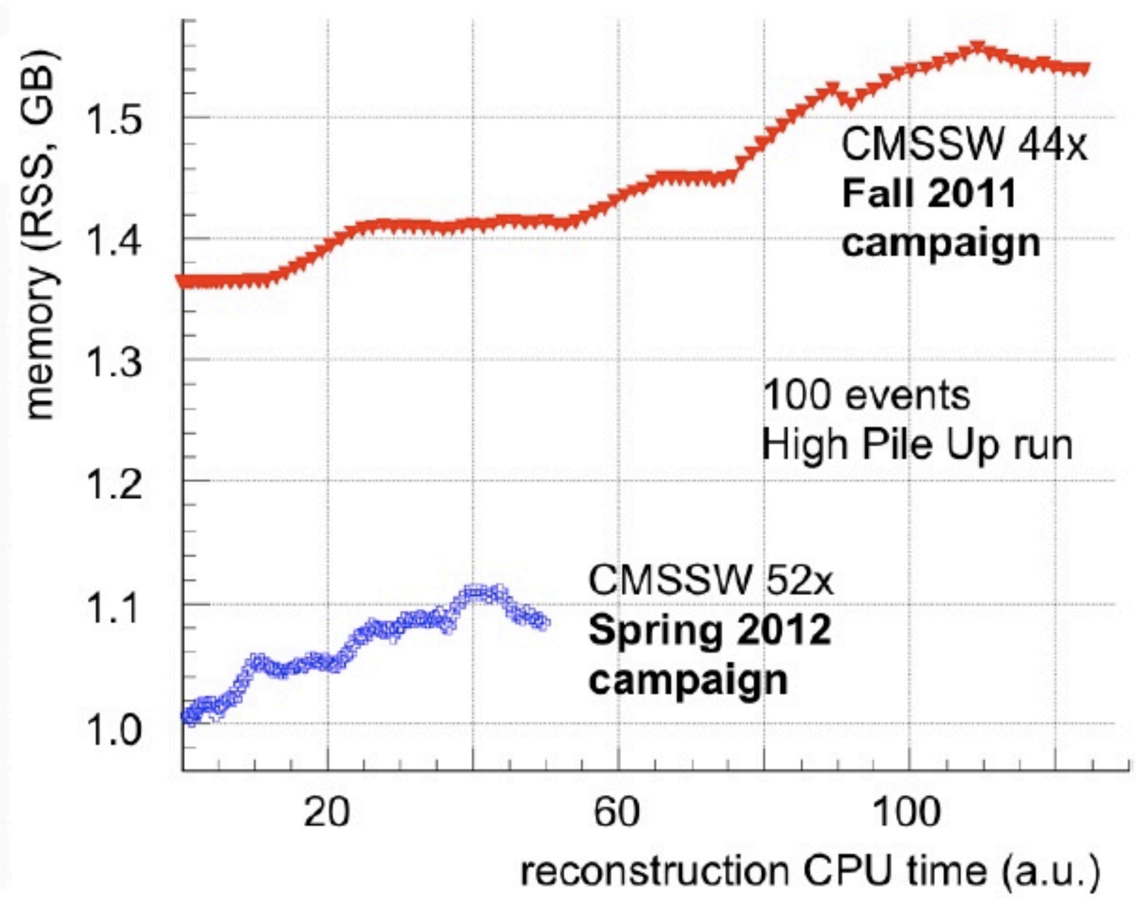
- Two major steps made in 2011 ("Fall11") and early 2012 ("Spring12") to cope with expected high pileup of 2012
  - ✓ NB: late 2011 data taking was already pushing the limits of available resources with the software version used throughout 2011
- Achieved major reduction in memory and CPU utilization

"Technical" Improvements

- We managed to get a lot from improving code quality, using new compiler, new root version, and a new memory allocation (jemalloc)

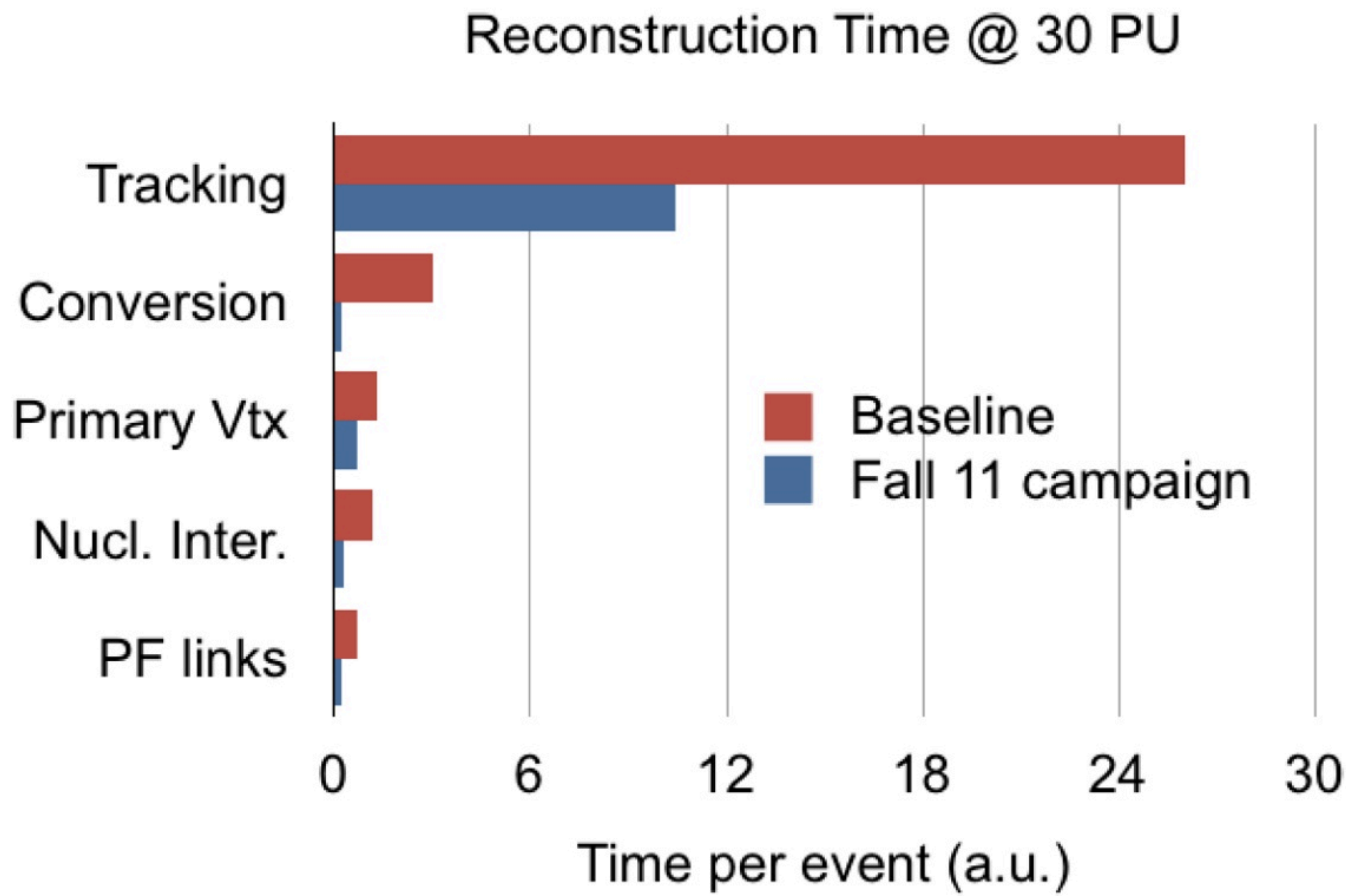
J. Phys.: Conf. Ser. 396 022044





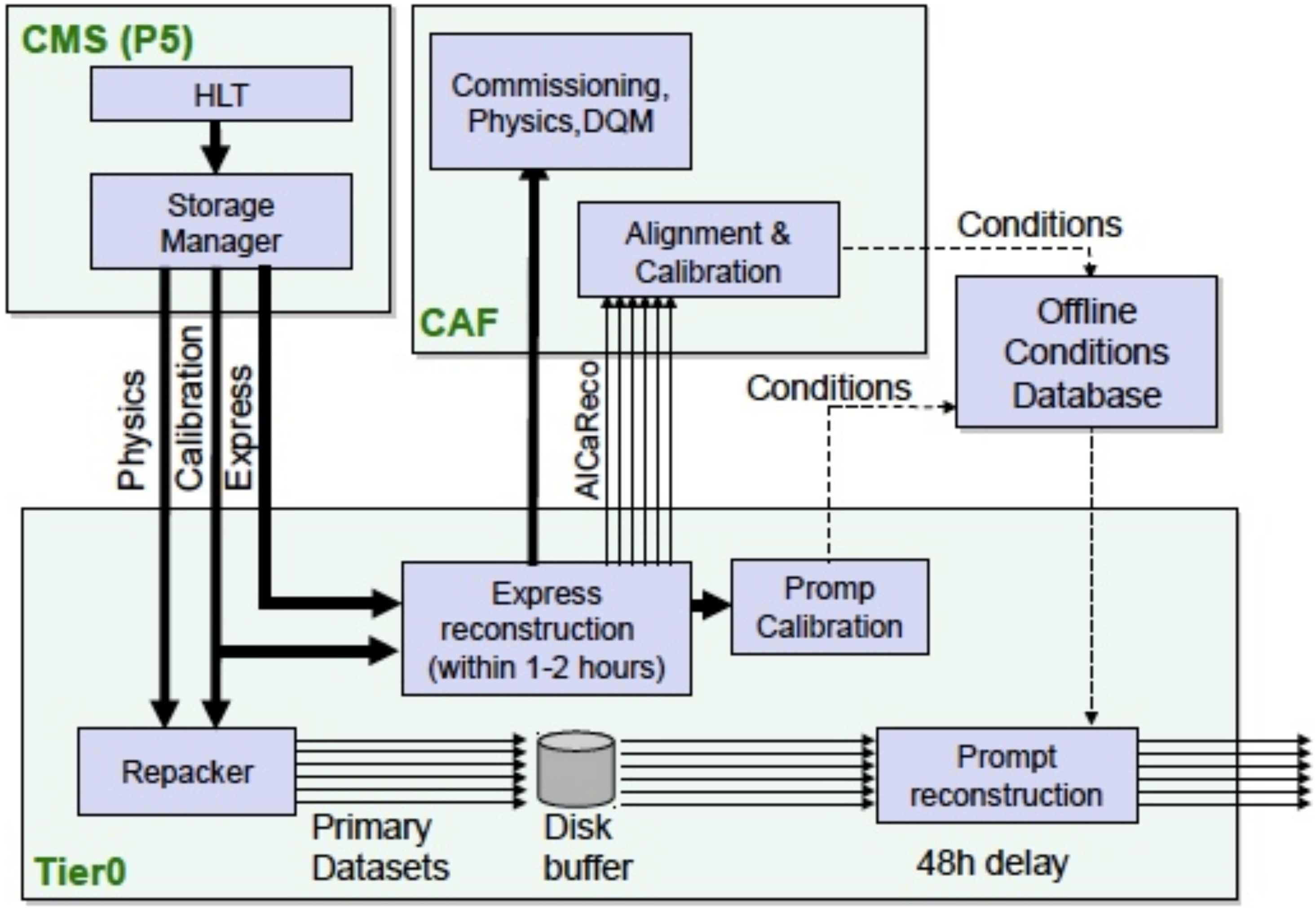
Algorithmic Improvements

J. Phys.: Conf. Ser. 396 022044



- ✓ (In addition to tuning the iterative tracking steps)
- ✓ Major reduction (x10) in conversion finding by improved seeding and tighter selections
- ✓ Nuclear interaction (displaced vertices) finding improved with tighter selections
- ✓ Other areas gained mostly from technical changes: e.g., using k_D-trees in PF links

Prompt Data Processing

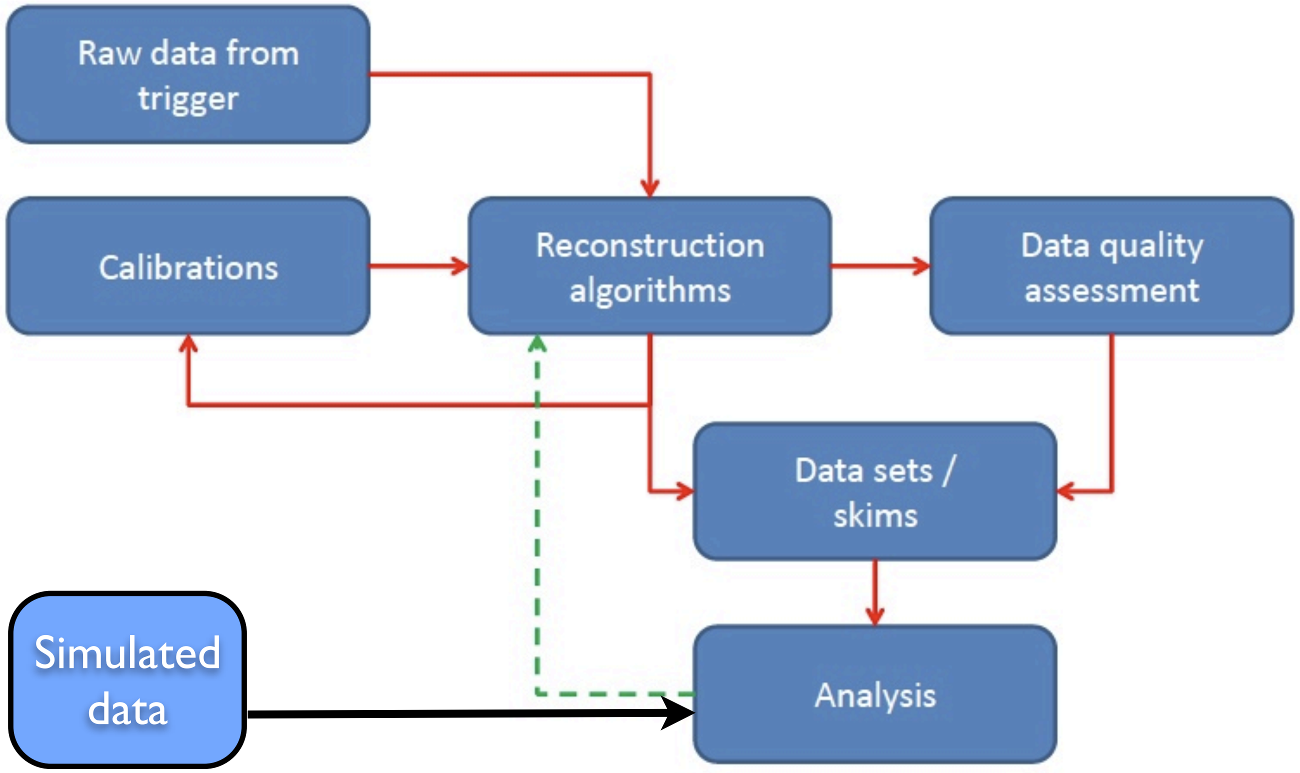


CMS CR-2012-145

- Resources are limited: crucial to have essential inputs right the first time
- Most preliminary and many published results on 2012 data used prompt reconstruction
  - ✓ Higgs boson discovery was made using data sets from prompt reconstruction
  - ✓ We had only one full reprocessing campaign (completed in May 2013)
- Recall: rereco of ~5 B events is several months CPU + analysis overhead

Improving Physics Performance: MVAs

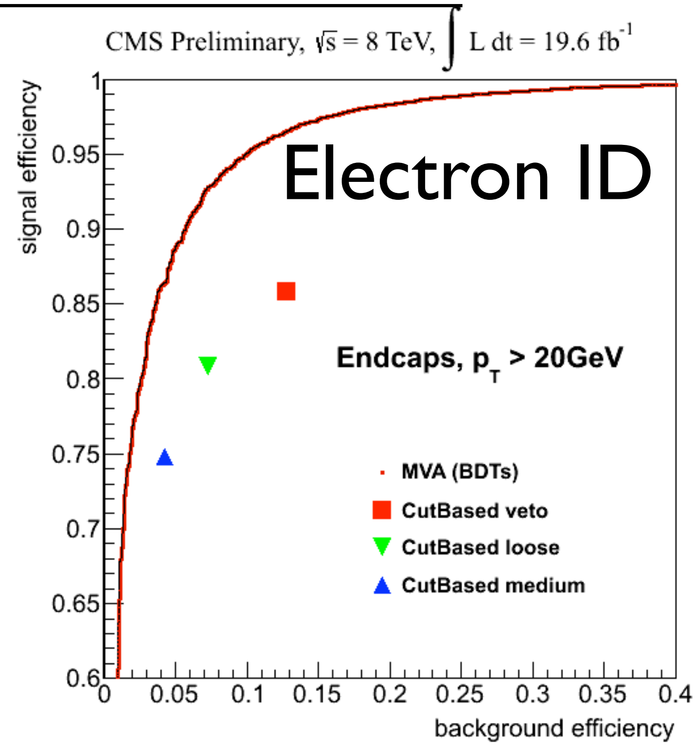
- Multivariate techniques are playing increasing role in high level object reconstruction
- Our analysis object data (AOD) format contains enough information to improve object selections or even kinematics measurements using multivariate techniques
- In 2012 we started closing the feed-back loop for many cases and compute MVA quantities now in the event reconstruction step
- Many Higgs boson analyses have pioneered these techniques



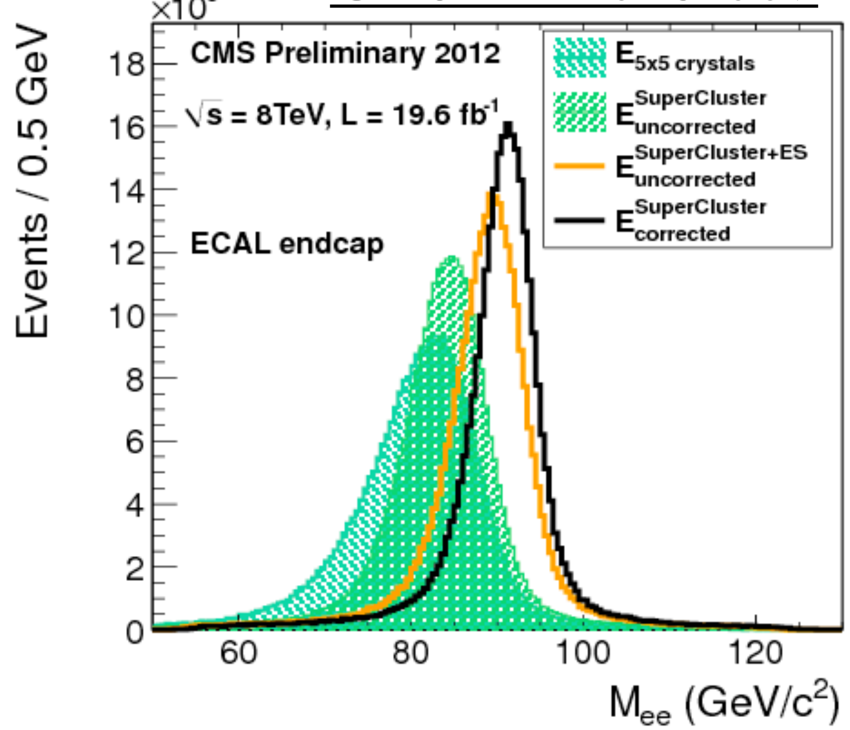
MVAs for electrons and photons

- Major improvements in ID and in resolution are possible
- Regression techniques are used to correct the energy, they improve the resolution
  - ✓ **Updated MVA training and calibration in data reprocessing of full 2012 dataset now gives the best results available**

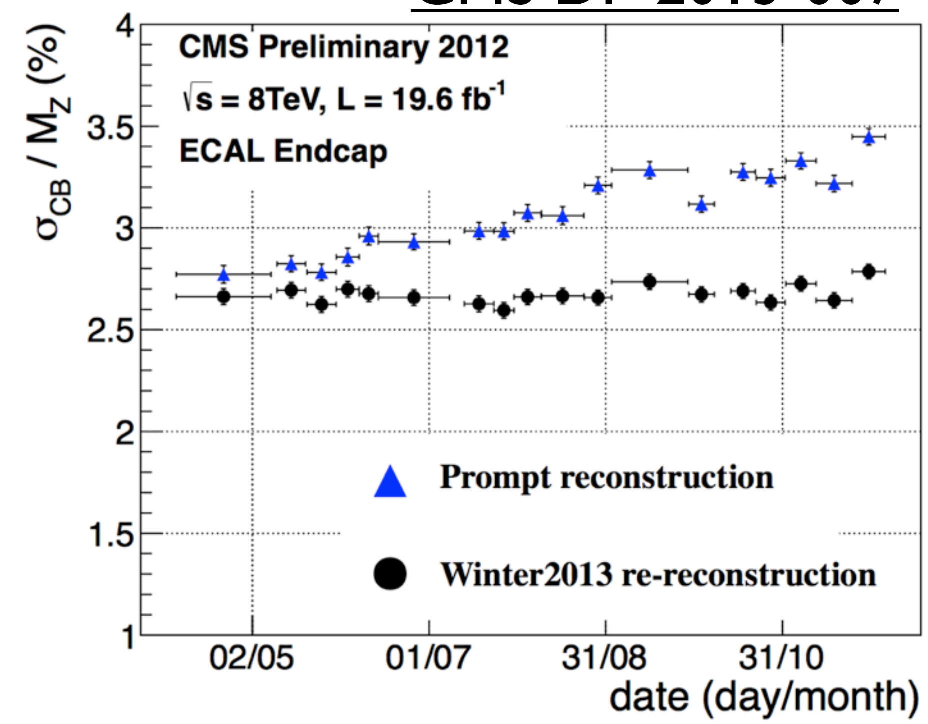
CMS DP-2013-003



CMS DP-2013-007



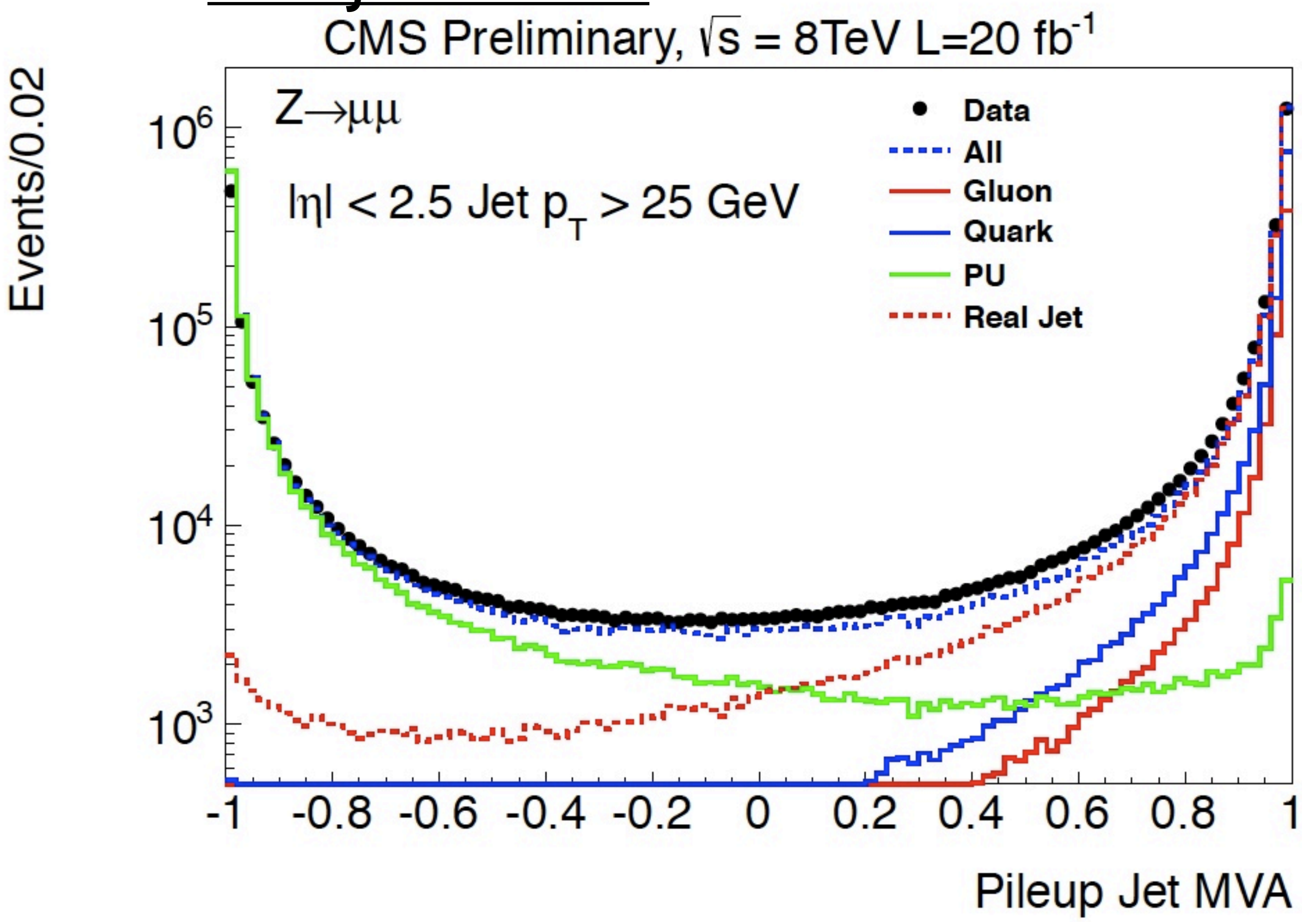
CMS DP-2013-007



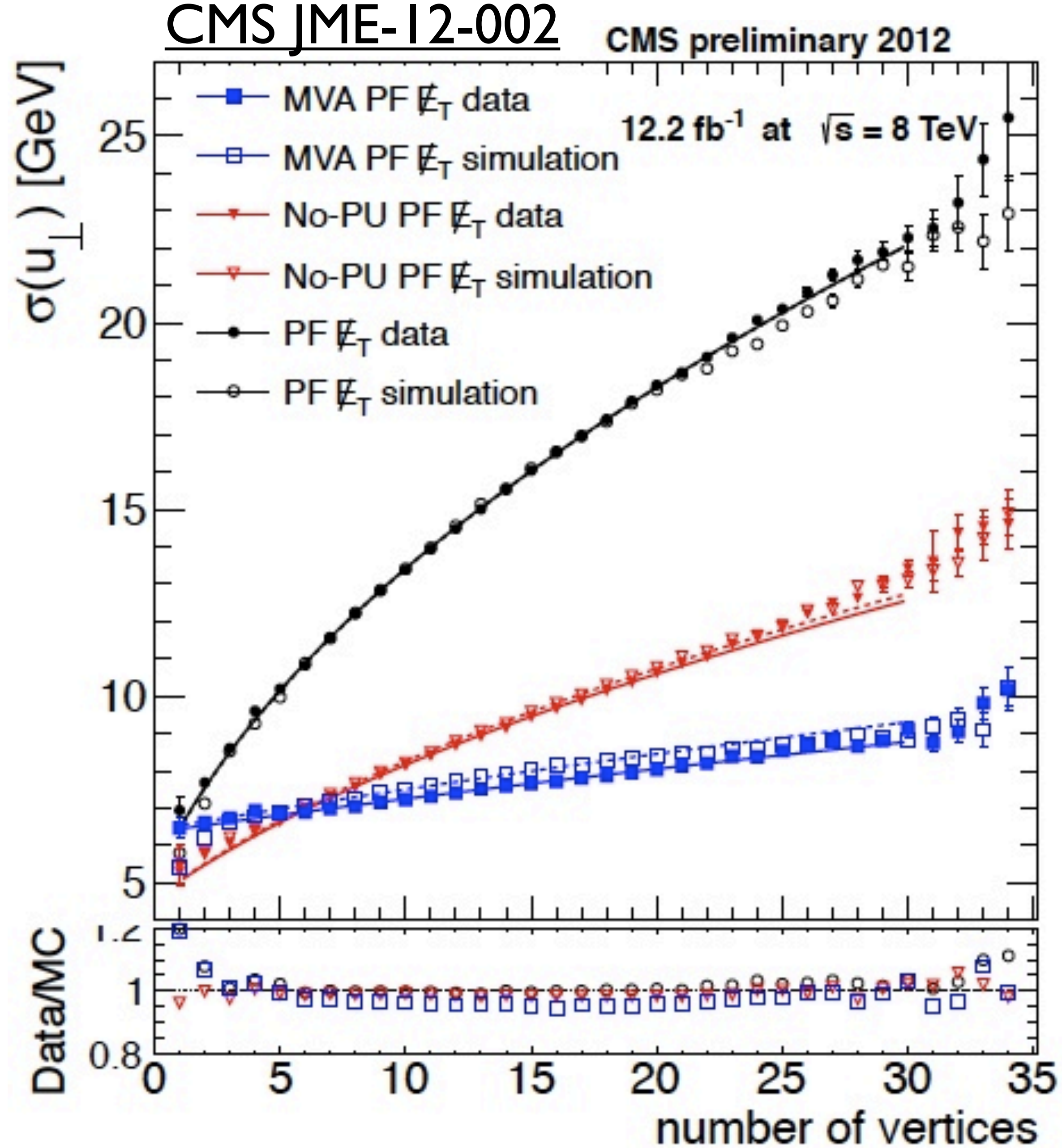
MVAs for jets/MET to cope with pileup

- Random clusters of particles coming from pileup can be rejected using MVA. This can be applied to either individual jets or in computation of missing transverse momentum (MET).

CMS JME-13-005



CMS JME-12-002



Summary and outlook

- Event reconstruction is made for all events used in physics analyses
  - ✓ The same setup is used for all events, both in real data and in simulation
- Higgs boson discovery wouldn't be timely possible without effective event reconstruction
- Major improvements in technical and algorithm level done for 2012 data taking lead to the successful physics program at CMS

- We are now preparing for upcoming Run 2 of the LHC and already making estimates of performance beyond that, through HL-LHC period in mid/late 2020s
  - ✓ Gains in performance will come both at "technical" and algorithmic level
    - execution in multiple threads (see today by C. Jones)
    - new tracking algorithms (see tomorrow by D. Funke)

Backup Slides