CMS Software 101

Yuri Gershtein

- Organization & People
- Simulation
- Reconstruction
- Data Handling & Analysis Tools

Disclaimer

- this is not a tutorial
- if by now you suspect that CMS must have a detailed detector simulation program you'll learn it's name and who develops it
- if you imagine that CMS reconstruction program finds electrons, you'll learn what kind of algorithms are deployed to do that

Accounts

- how to get CERN CMS accounts

http://cmsdoc.cern.ch/comp/comp_quick_guide.html#AccountCreation

- how to get UAF FNAL accounts

http://www.uscms.org/scpages/general/uaf/accounts.html

- possible (but hard) to do if you’re not in the corresponding country

Who Develops it?

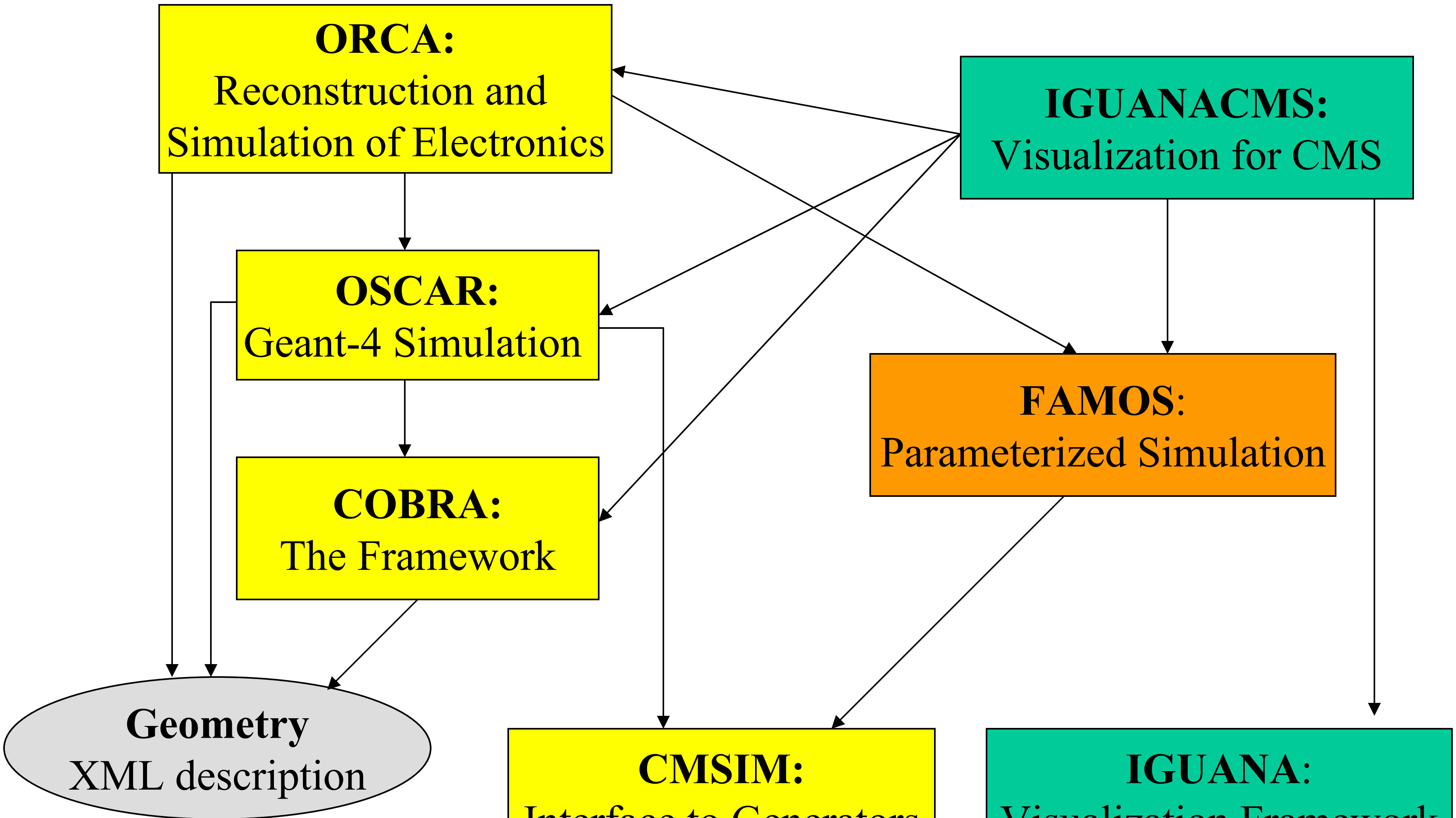
- Computer professionals
  - common between LHC experiments or even wider (Grid)
- PRS (Physics, Reconstruction and Selection) ID Groups:
  - ECAL/e/gamma: Chris Seez (London) and Yves Sirois (Ecole Polytechnique)

    general software responsibility: David Futyan and Emilio Meschi
  - HCAL/jet/met: Chris Tully (Princeton) and Jim Rohlf (Boston)

    Sunanda Banerjee (simulation), Olga Kodolova (calibration),
    Salavat Abdullin (reconstruction) and Sasha Nikitenko (physics objects)
  - Muon/mu: Ugo Gasparini (Padova) and Darin Acosta (Florida)
  - Tracker/b/tau: Marcello Mannelli (CERN) and Lucias Silvestris (Bari)

    Tommaso Boccali (simulation), Teddy Todorov (reconstruction), Fabrizio Palla (b algorithms), and Sasha Nikitenko (tau algorithms)
- LPC groups are beginning to contribute to work done in corresponding PRS groups

Software Projects



External Software

- LHC Computing Grid applications
  - SEAL – common tools for LHC experiments
  - POOL – persistency framework
  - PI – physicist interface project
- Externals
  - CERNLIB, Geant-4, ROOT, CLHEP, etc…
- Validation tool
  - OVAL
- Build tool
  - SCRAM

Documentation

- http://cmsdoc.cern.ch/cmsoo/cmsoo.html

Welcome to the CMS Object-Oriented page

Object Oriented Projects

Project Name	View	XRef	Project Description
ORCA	CVS	LXR	Object-oriented Reconstruction for Cms Analysis
CMSDAQSIM	CVS	N/A	CMS DAQ Simulation
CRISTAL	CVS	N/A	The CRISTAL workflow management system
TriDAS	CVS	N/A	Trigger Data Aquisition System
IGUANA	CVS	LXR	Interactive Graphical User ANAlysis
DDD	CVS	N/A	Detector Description Database
OSCAR	CVS	LXR	Object oriented Simulation for Cms Analysis and Reconstruction
FAMOS	CVS	LXR	CMS Fast Simulation
COBRA	CVS	LXR	Coherent Oject-oriented Base Reconstruction and Analysis
SCRAM	CVS	N/A	Software Configuration, Release And Management
Geometry	CVS	LXR	Geometry
IGUANACMS	CVS	N/A	Interactive Graphics and User ANAlysis for CMS

Download Projects

Software risk assessments

Bug Reporting System (Savannah)

Project dependency diagram

- CMS OO Software Architecture
- CMS TestBeams Offline Documentation
- CMS Documentation

Tutorials

- Occur on a regular basis
- Transparencies and video recordings are available at

`http://cmsdoc.cern.ch/cms/software/tutorials/`

CMS OO Software Tutorials

2nd series: starting 19. Mar. 2004

- Introduction to CMS software - The Basics: slides
- Simulation and Reconstruction - The Full Chain: slides
- Reconstruction and DST Analysis: slides
- Analysing Data with PAX: slides
- Muon and MuonReco software: slides
- Tracker and TrackerReco software: slides
- ECAL and ElectronPhoton software: slides
- HCAL and JetMET software:

1st series: 11. Nov. 2003 - 30. Jan. 2004

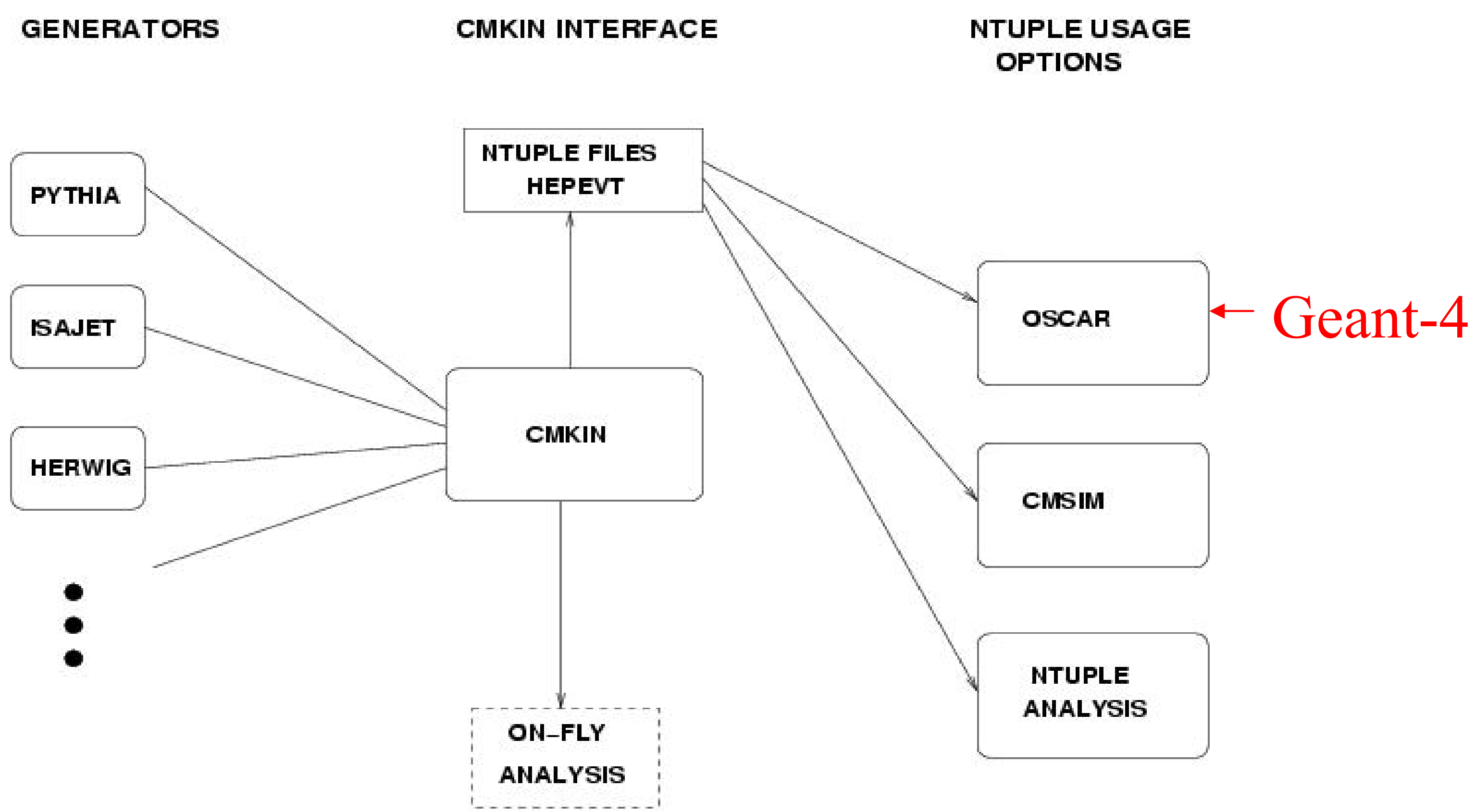
- New User Tutorial (basic introduction): slides, video
- The Full Simulation/Reconstruction Chain Tutorial: slides, video
- The Framework: slides, video
- Tracker-Track and Vertex Reconstruction: slides, video
- ECAL and Electron-Photon Reconstruction: slides, video
- HCAL and JetMET Reconstruction: slides, video
- Muon Reconstruction: slides, video

Event Generation: CMKIN

CMS Contact: ??? *US contact: ???*

Web documentation: ???

Written in FORTRAN – as are most of generators



Geant 4 Simulation: OSCAR

CMS Contact: Maya Stavrianakou ?

US Contact: Daniel Elvira ?

Web documentation: http://cmsdoc.cern.ch/oscar/

- Reads events from CMKIN Ntuple
- Uses Geant 4 to simulate how particles propagate through space, interact with the detector material, loose energy, etc. (SimTracks, SimVertices)
- Creates SimHits representing this information
- stores RawParticles, SimTracks, SimVertices and SimHits in POOL data files

Fast Simulation: FAMOS

CMS Contact: Patric Janot and Dave Bailey

US Contact: ?

Web documentation: http://cmsdoc.cern.ch/FAMOS/

- GeneratorToAnalysis
  - detector response parametrization
- ORCA Interface
  - enable to access to FAMOS objects in ORCA
- FAMOS Generic
  - its own framework
- Examples and Documentation
  - provided

ORCA

CMS Contact: Stephan Wynhoff and Norbert Neumeister

US Contact: Stephan Wynhoff and Norbert Neumeister?

Web documentation: http://cmsdoc.cern.ch/orca/

- Object Oriented Reconstruction for CMS Analysis
- ORCA is the program that you're most likely to use if you ever go beyond analyzing Ntuple's
- It is a very flexible program/environment which is now used for simulation, reconstruction and analysis

ORCA

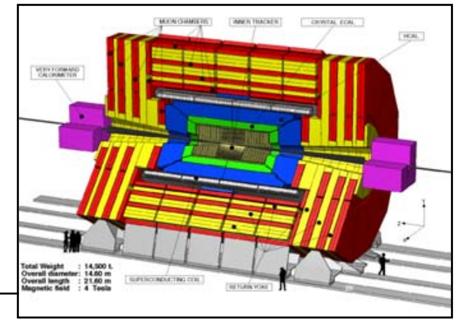
SimHits
Pileup Events

SimHits
Signal Event

Simulation of Electronics

Calorimetry
Tracker
Muon
Trigger

DIGI's
RecHits



CommonDet
CommonReco

Common Services

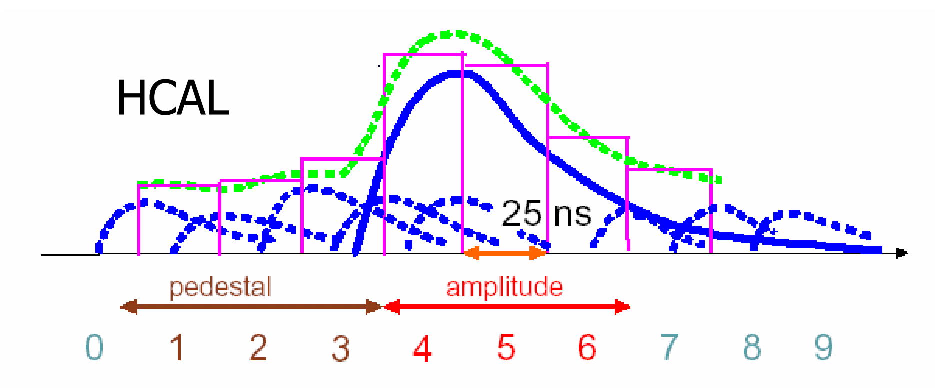
TrackReco, Vertex,
MuonReco, Jets

High Level Reconstruction

ElectronPhoton,
JetMetAnalysis,
MuonAnalysis,
bTauAnalysis

HLT and "Analysis"

- 20 interactions/ crossing
- 25 ns crossing time: faster than detector response – read out (simulate) crossings –5;+3
- 200 pileup events per 1 signal event!



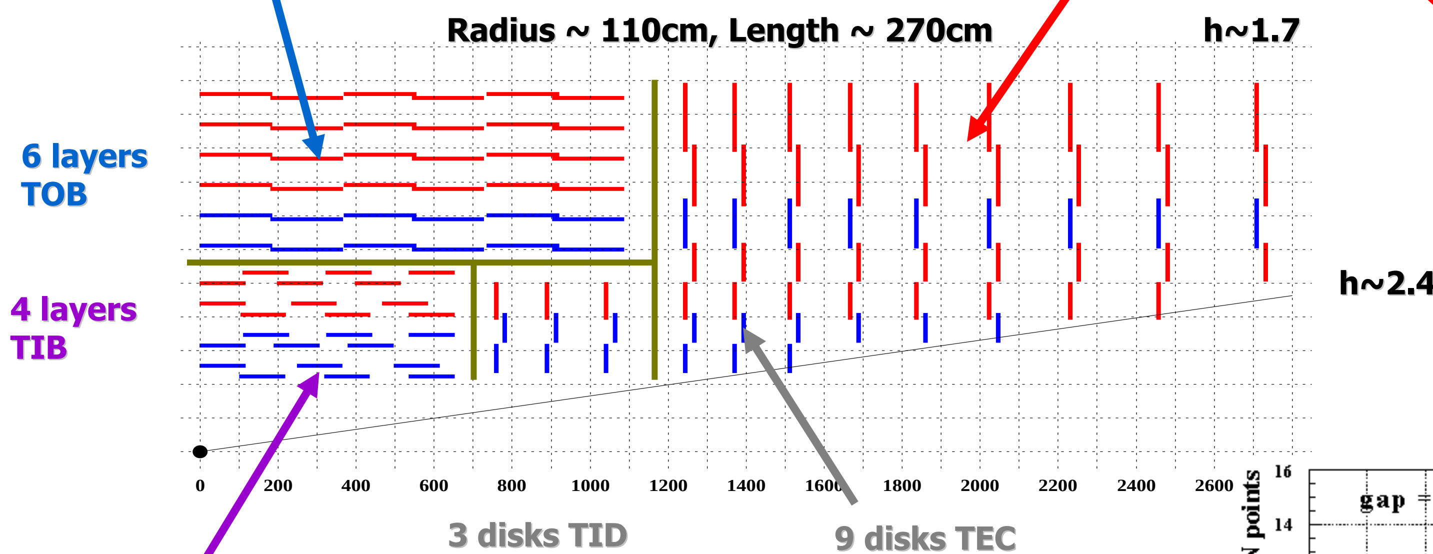
Framework

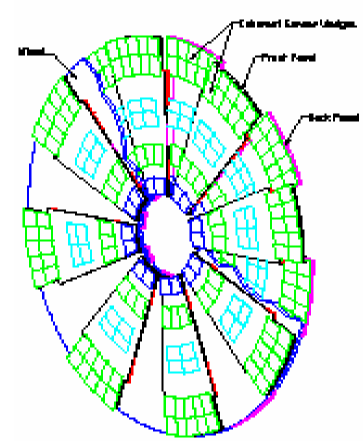
- System is built with enormous flexibility
- CARF – CMS Analysis and Reconstruction Framework
  - part of COBRA (Coherent OO Base for simulation, Reconstruction, and Analysis)
  - base event classes (objects, reconstruction, storage)
  - "action on demand" and "implicit invocation"
- All reconstruction modules register with the framework and invoked only when required

CMS Tracker

Outer Barrel (TOB): 6 layers
- Thick sensors (500 mm)
- Long strips

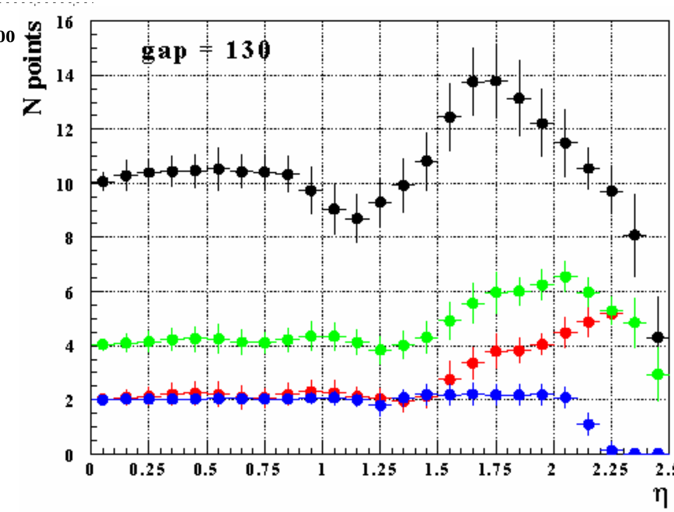
Endcap (TEC): 9 Disk pairs
- r < 60 cm thin sensors
- r > 60 cm thick sensors





Inner Barrel (TIB): 4 layers
- Thin sensors (320 mm)
- Short strips

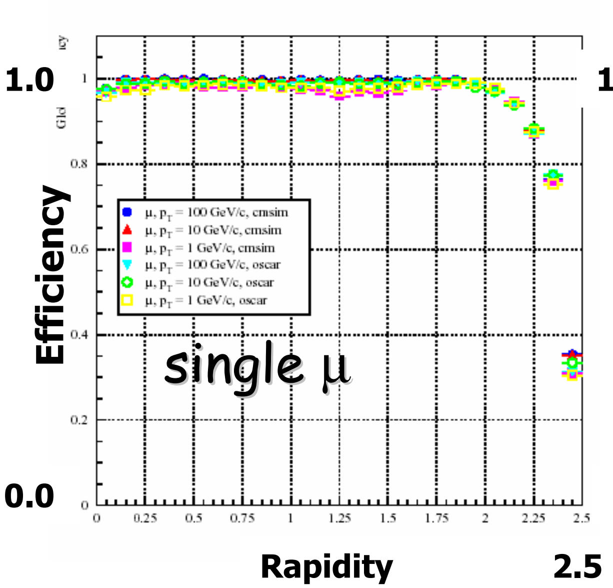
Inner Disks (TID): 3 Disk pairs
- Thin sensors

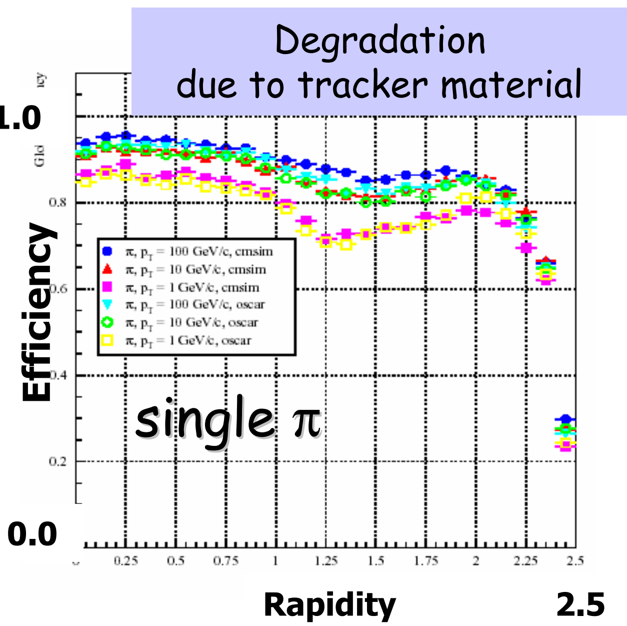


Black: total number of hits
Green: double-sided hits
Red: ds hits - thin detectors
Blue: ds hits - thick detectors

ORCA: Tracking

- Very few measurements per track
  - very precise
  - very low occupancy: 10^{-4} in pixels, <5% in strips
- A lot of material in the tracker
  - 0.5-1.5 X_0: electrons brem, photons convert
  - 0.1-0.4 λ: pions interact
- Track inside out: start in pixels and extrapolate to strips
  - Can use external seeds (i.e. a pixel hit and an EM cluster)

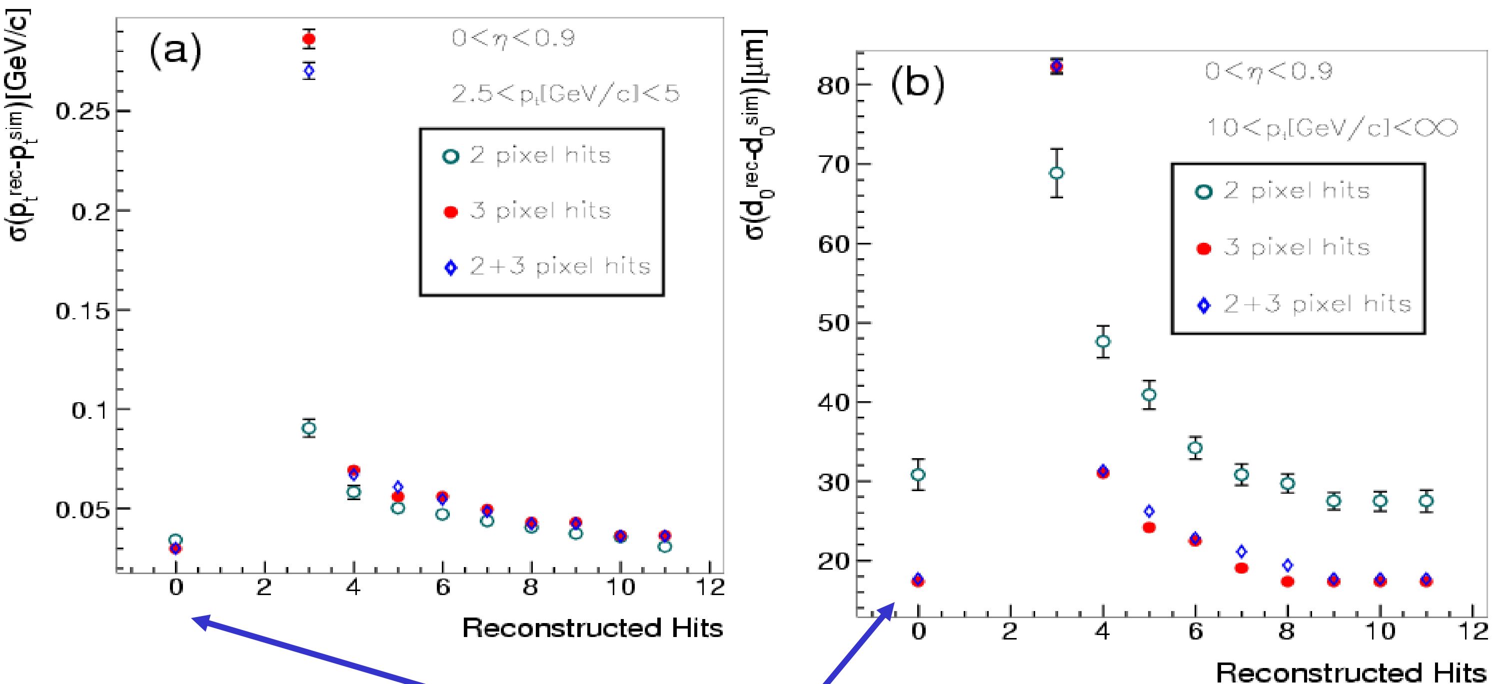




pT resolution plot

HLT Tracking

- Same algorithm and code
- HLT specific:
  - track only specific regions in the tracker, i.e. around an muon trigger candidate
  - stop extrapolation once the track has 4-5 hits to save time



Full tracker performance

Vertexing

- Interaction region is 5 cm, ~ 20 vertices/event
  - 2.5 mm between vertices
  - Track Z resolution is
    - ~ 0.1(0.7) mm for 1 GeV and η=0 (2)
    - ~ 0.03 (0.1) mm for 10 GeV and η=0 (2)
- Especially important for analyses like $h \rightarrow \gamma\gamma$
  - correct E_T determination
  - track isolation w.r.t correct vertex
- heavy flavour tagging
  - bTauAnalysis

Electrons & Photons

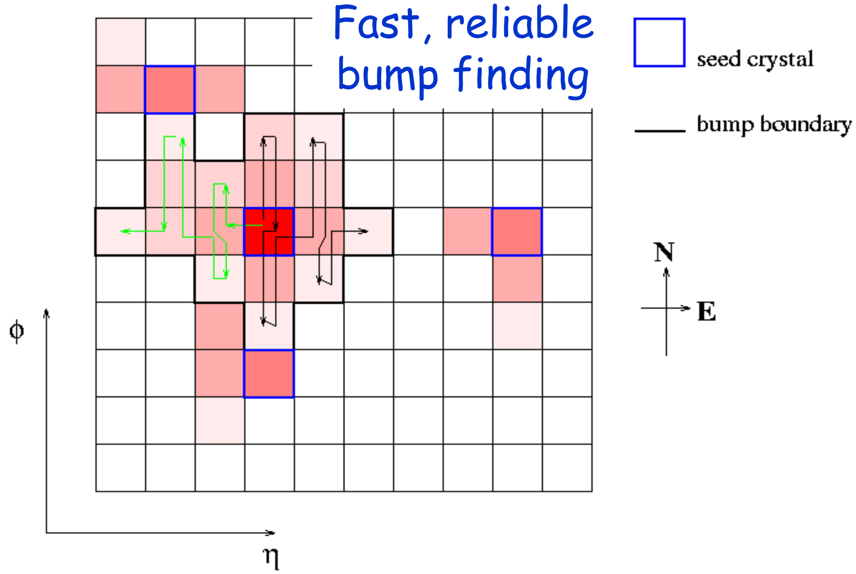
Documented in CMS note 2001/034

Consists of the following steps

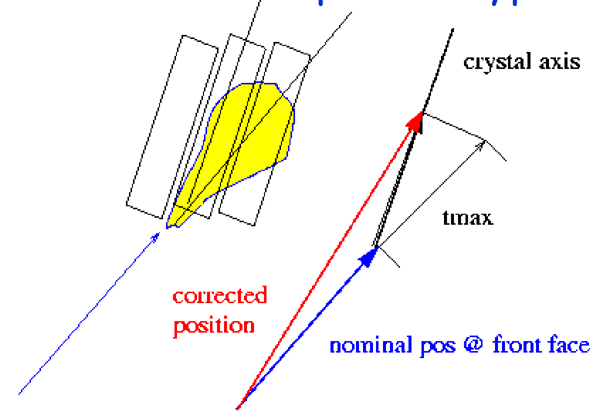
- ✓ Reconstruct cell energies from time frames
- ✓ Define a basic calorimetric cluster as a collection of cells with energy deposition
- ✓ Get the best position resolution
- ✓ Get the best brem recovery
- ✓ Get the best energy resolution

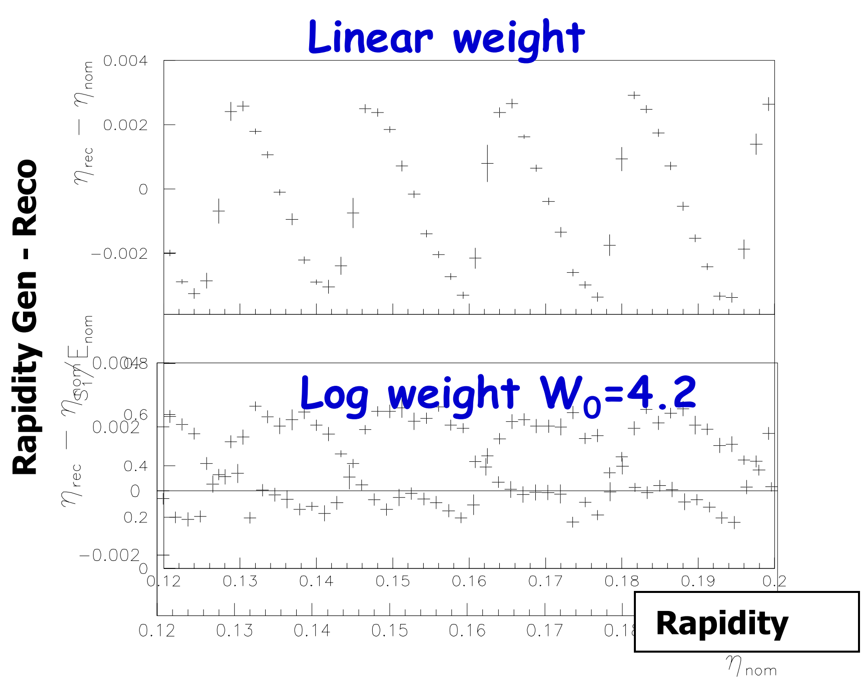
Island Clustering and Position Corrections

Fast, reliable bump finding



Shower depth correction
Log dependence on energy
(and electron or photon hypothesis!)





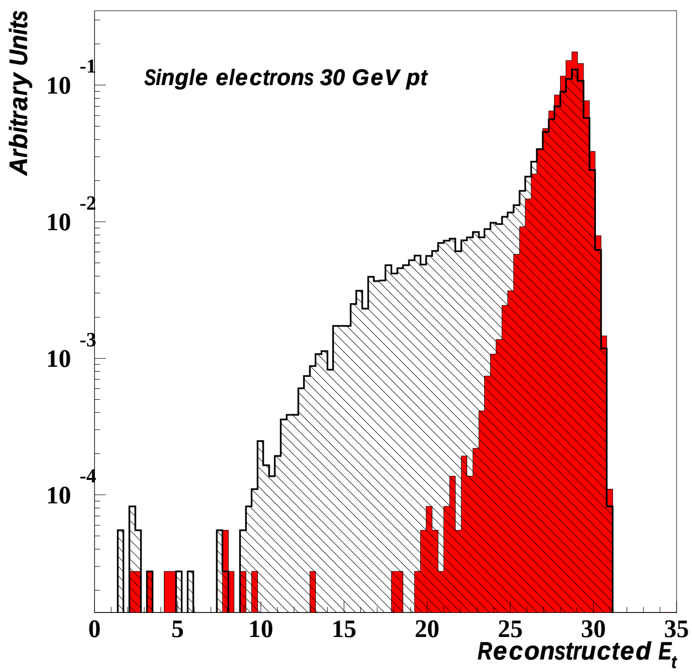
Shower position correction

$$W_i = W_0 + \log\left(\frac{E_i}{\sum_i E_i}\right)$$

Brem recovery

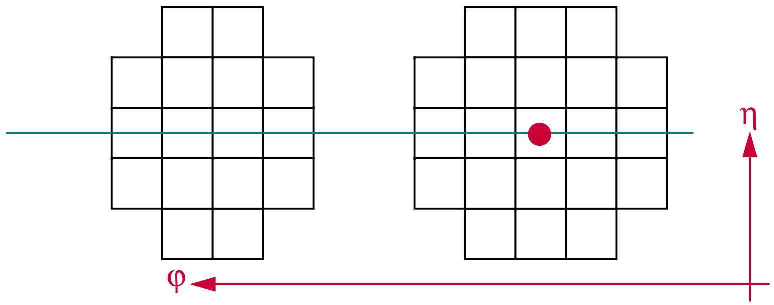
SuperCluster of island clusters

- Energy deposition from brem well aligned in η
- Use narrow η window
- Collect clusters along ϕ

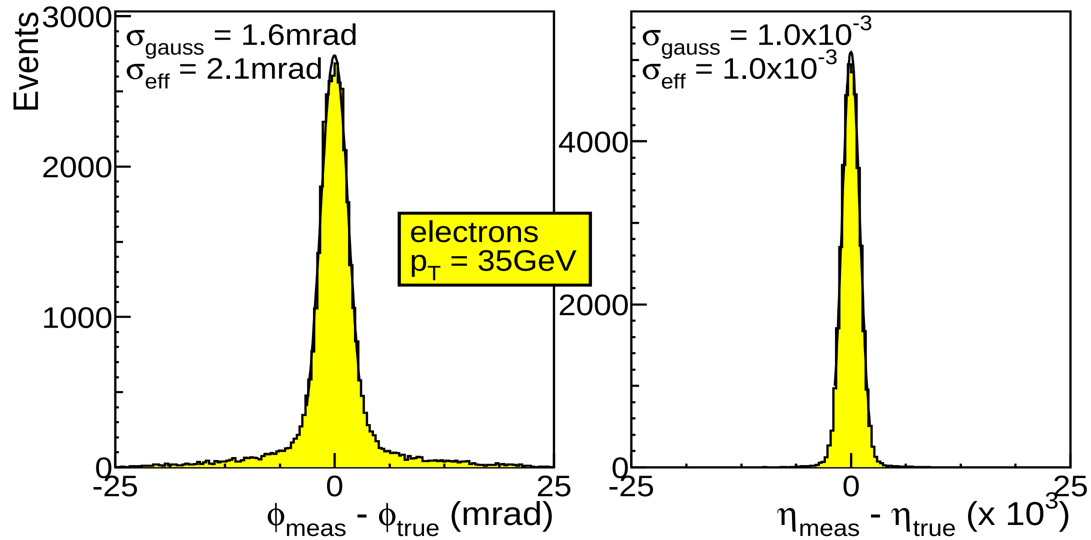


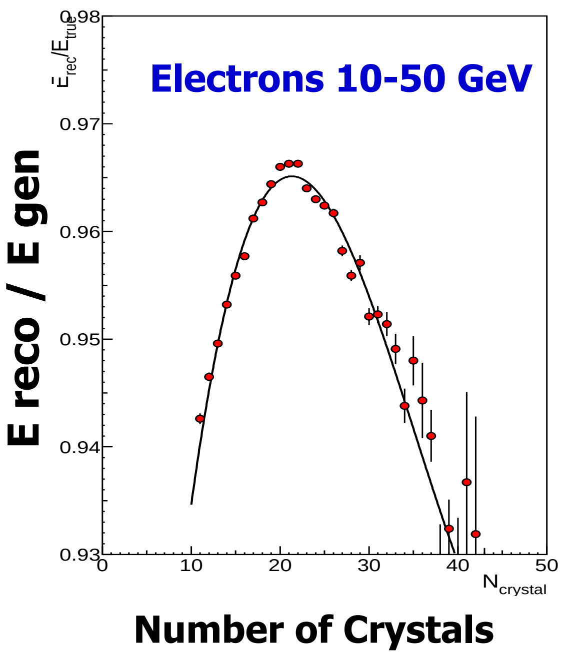
Hybrid Algorithm

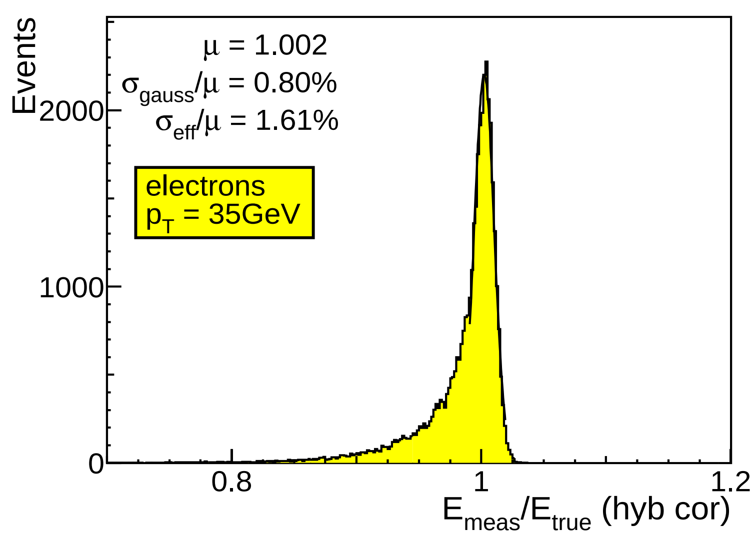
- Use η–ϕ geometry of barrel crystals
  - Start from a seed crystal (as for island)
  - Take a fixed domino of 3 or 5 crsytals in η
  - Search dynamically in ϕ



Resolutions and Energy Scale

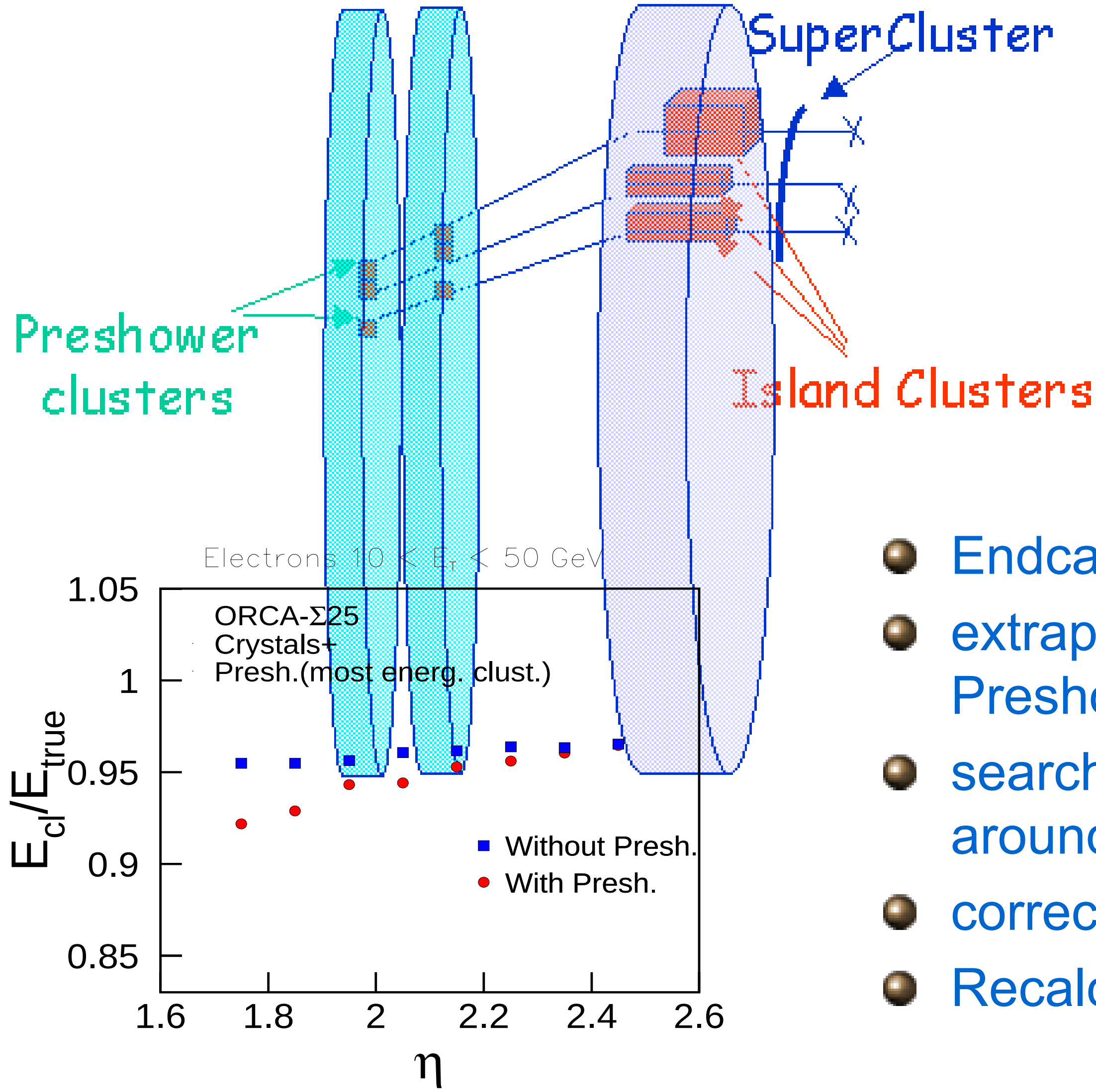






Energy resolution of unconverted photons $\sigma_{eff}/E \sim 0.9$ %

Preshower matching



- Endcap SuperCluster
- extrapolate components to Preshower
- search PS cluster in narrow road around extrapolated point
- correct component energy
- Recalculate SuperCluster energy

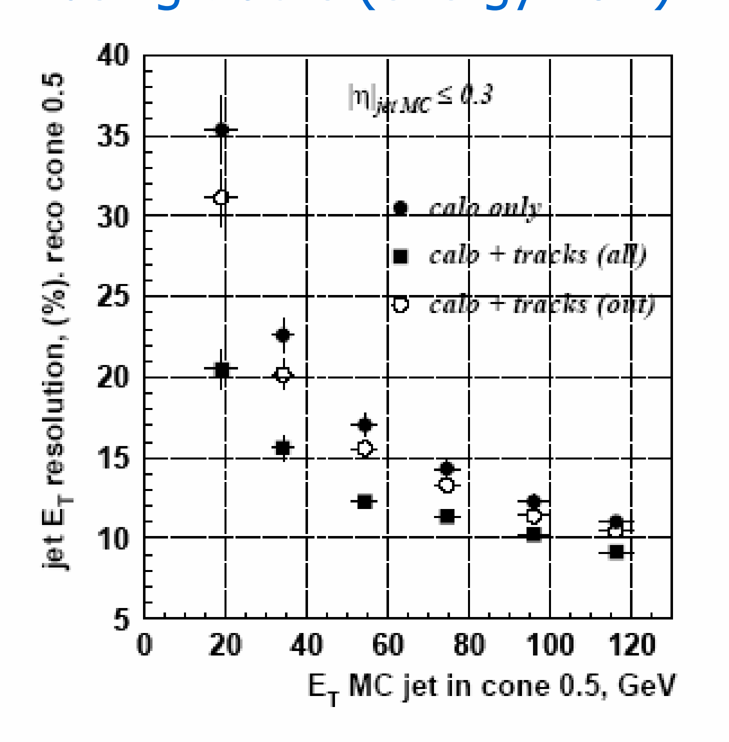
Jets & Missing E_T

- Mostly uses ECAL and HCAL information
- High Level Trigger: EcalPlusHcalTowers
  - Correspond to HCAL $\eta-\varphi$ towers
  - One HCAL tower matches 5x5 ECAL crystals (approximately in EC)
- Offline
  - Use both longitudinal and transverse segmentation (RecHits)
  - Refine jets and MET with tracks and muons
- Jet Algorithms available:
  - Cluster based:
    - Inclusive kT
    - Exclusive kT(d_{cut})
    - Exclusive kT(N_{jets})???
  - Cone based:
    - Simple Cone
    - Iterative Cone
    - MidPoint Cone
  - Advanced: JetPlusTracks: add out of cone tracks and substitute hadronic energy measurement with track pT

Tracking in Jets and MET

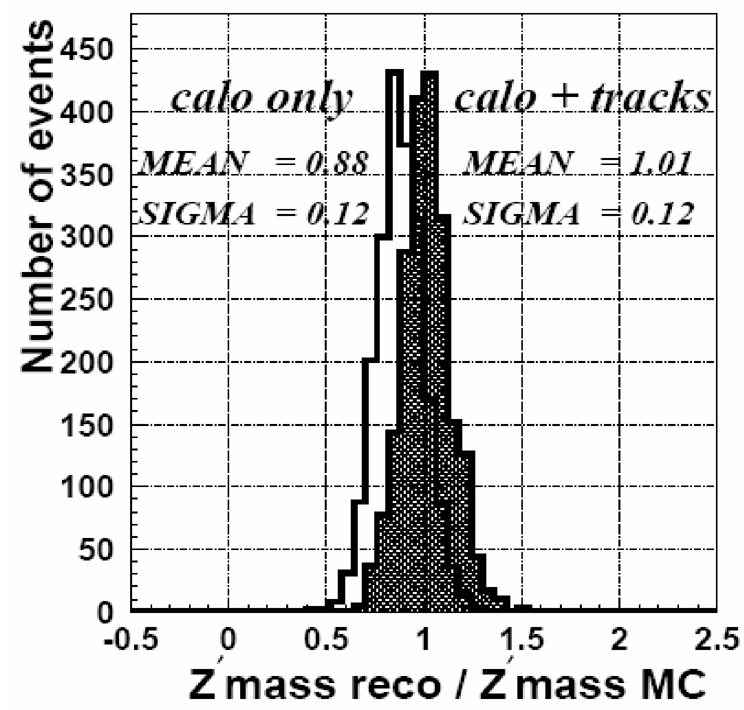
Jet resolution:

- correct for magnet-induced out-of cone leakage
- correct isolated hadronic clusters using tracks (energy flow)



Reconstruction of Z′→jj:

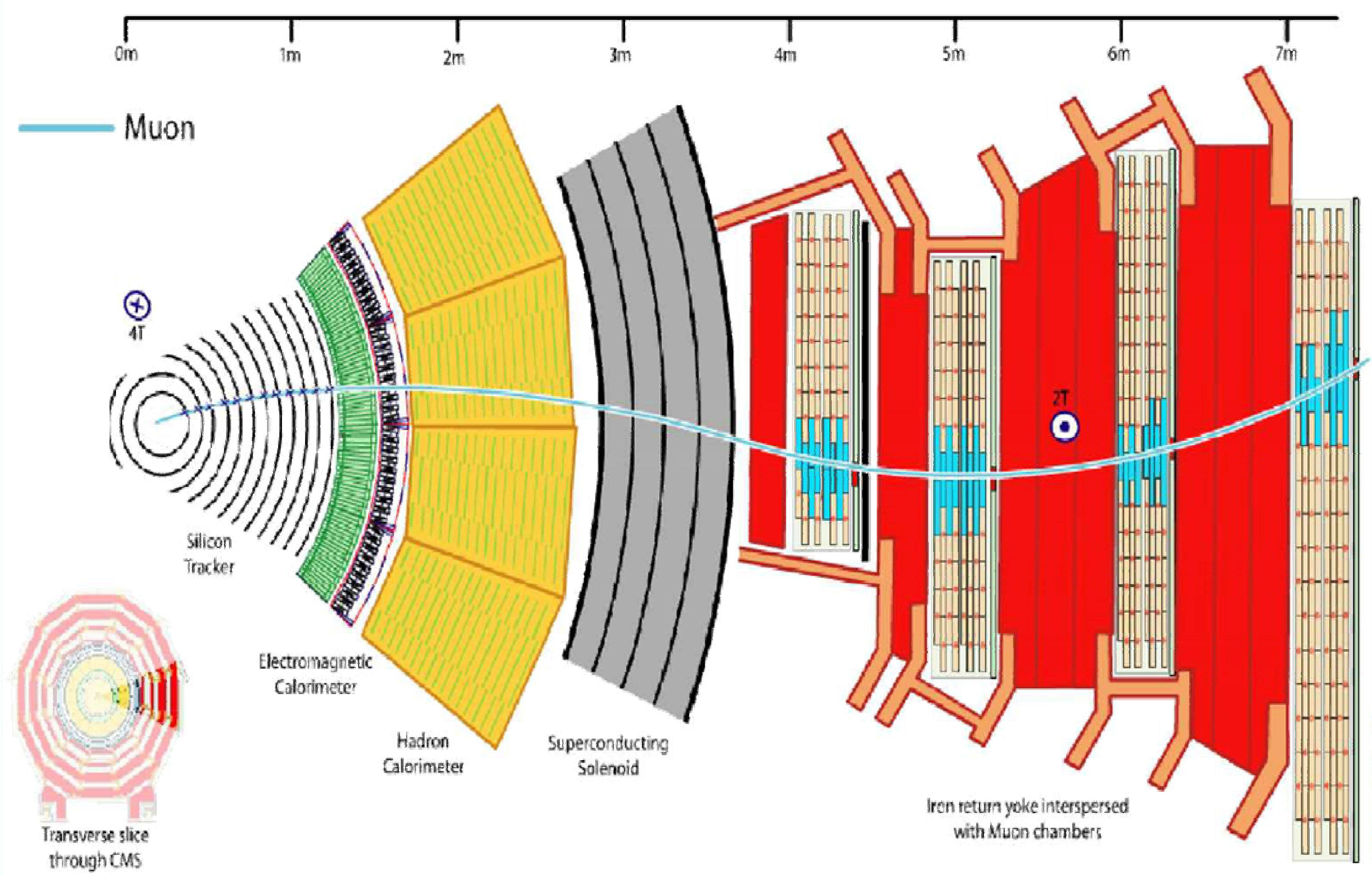
~15% improvement!



Potential benefits for MET:

pile-up subtraction

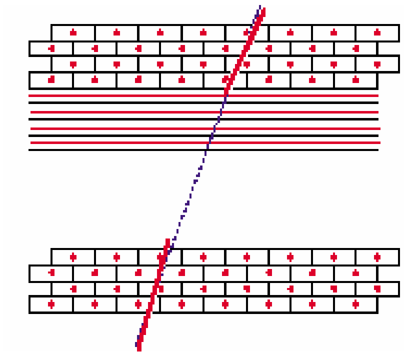
Muon Reconstruction



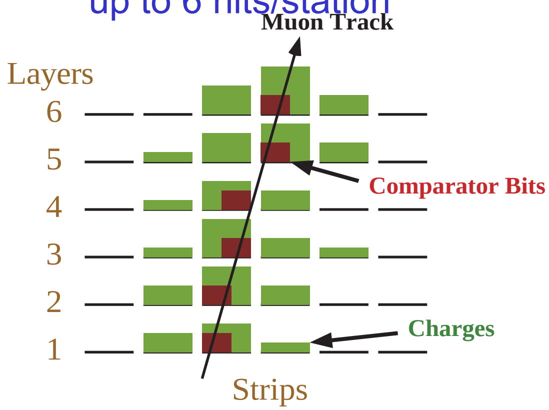
Standalone Muon Reconstruction

- All muon detectors (DT, CSC and RPC) are used
- Start by finding track segments in stations:
  - 2d hits in barrel, 3d hits in endcaps
- Fit:
  - Kalman filter technique applied to DT/CSC/RPC track segments
  - Use segments in barrel and 3D hits in endcaps
  - Trajectory *building* works from inside out
  - Apply χ^2 cut to reject bad hits
  - Track *fitting* works from outside in
  - Fit track with beam constraint
- Propagation:
  - Non constant magnetic field
  - Iron between stations, propagation through iron (more difficult than in tracker!)
  - GEANE used for propagation through iron

Barrel: Drift Tubes
up to 12 hits/station

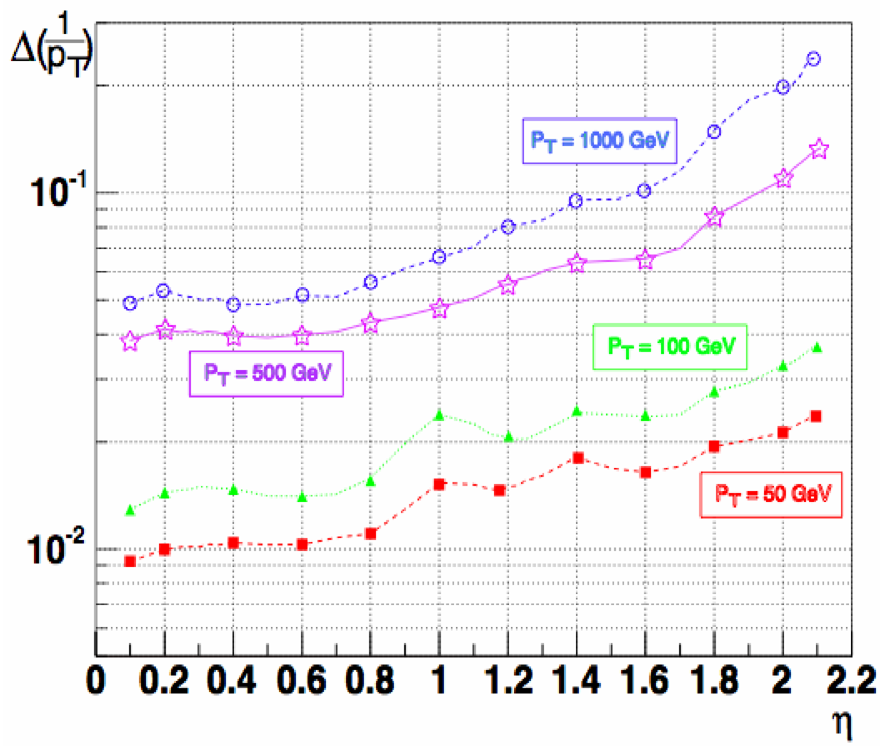


Endcap:
Cathode Strip Chambers
up to 6 hits/station



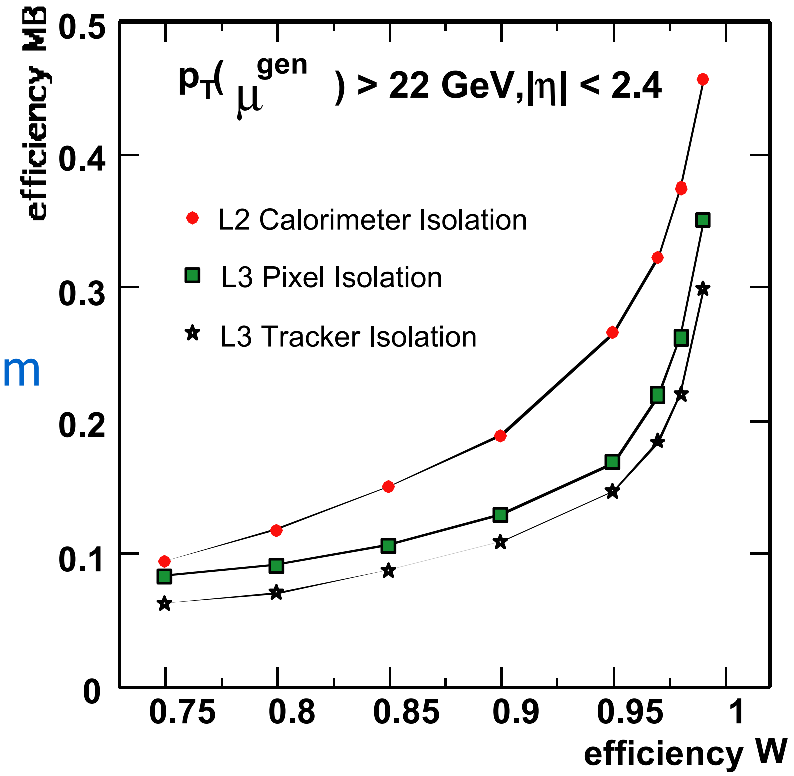
Global Muon Reconstruction

- Start with a local muon (~10% resolution @ 100 GeV)
- Extrapolate to the interaction point and find track seeds
  - can have many track seeds per muon
  - build a track propagating out, including hits in muon system
  - resolve ambiguities and do a final fit



Muon Isolation

- Very useful for jet rejection
- Calorimeter Isolation
  - ΣE_T from calorimeter towers in a cone around muon (sensitive to pile-up)
- Pixel Isolation
  - ΣP_T of 3-hit tracks in the pixel detector in cone around muon
  - Requires that contributing tracks come from the same primary vertex as the Level-3 muon (to reduce pile-up contamination)
  - Studies done for full pixel detector (no staging)
- Tracker Isolation
  - ΣP_T of tracks in the tracker (regional reconstruction around Level-3 muon)

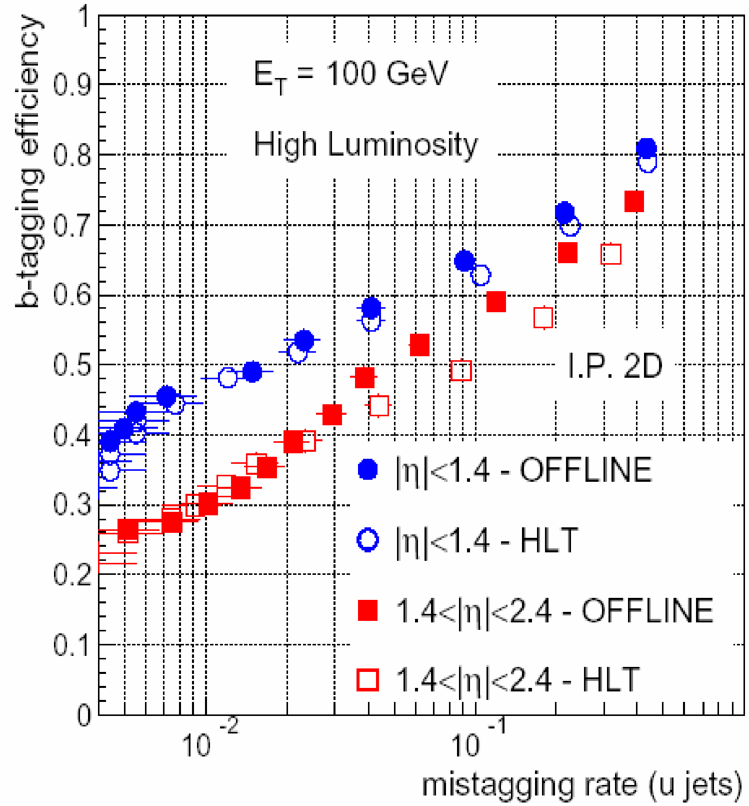


bTauAnalysis

Inclusive b tag at HLT possible, provided alignment under control …

Regional Tracking: Look only in
Jet-track matching cone
Loose Primary Vertex association

Conditional Tracking:
Stop track as soon as
Pixel seed found (PXL) / 6 hits found
If Pt<1 GeV with high C.L.



bTauAnalysis

Regional Tracking: Look only in
Jet-track matching cone
Loose Primary Vertex association

Conditional Tracking: Stop track as soon as
Pixel seed found (PXL) / 6 hits found (Trk)
If Pt<1 GeV with high C.L.

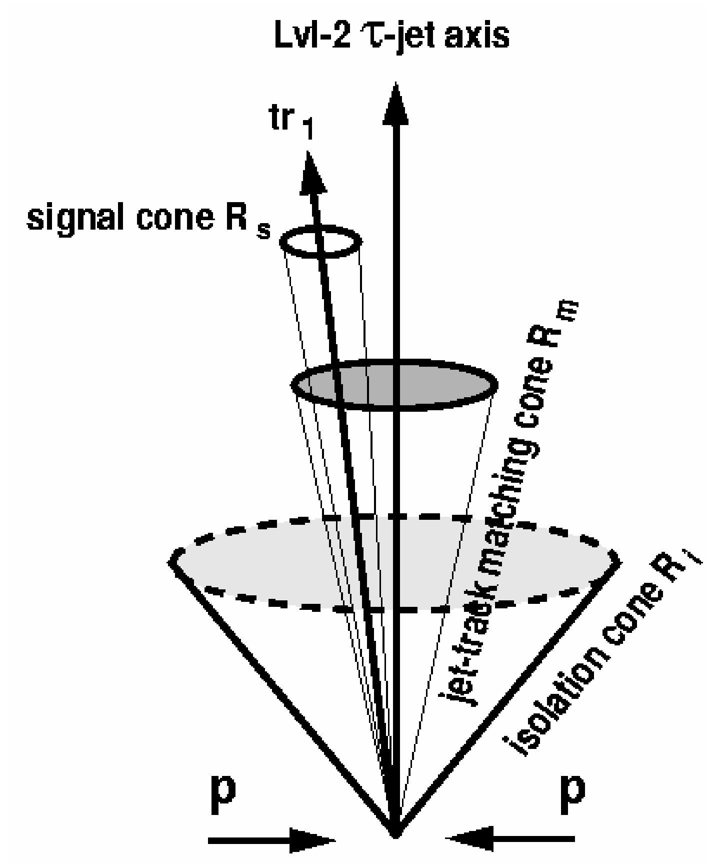
Reject event if no "leading track" found

Regional Tracking: Look only inside
Isolation cone
Loose Primary Vertex association

Conditional Tracking: Stop track as soon as
Pixel seed found (PXL) / 6 hits found (Trk)
If Pt<1 GeV with high C.L.

Reject event as soon as additional track found

$A^0/H^0 \to 2\tau \to 2\tau$ jets



Visualization

- IGUANA – Interactive Graphics for User Analysis

  CMS Contact: ???

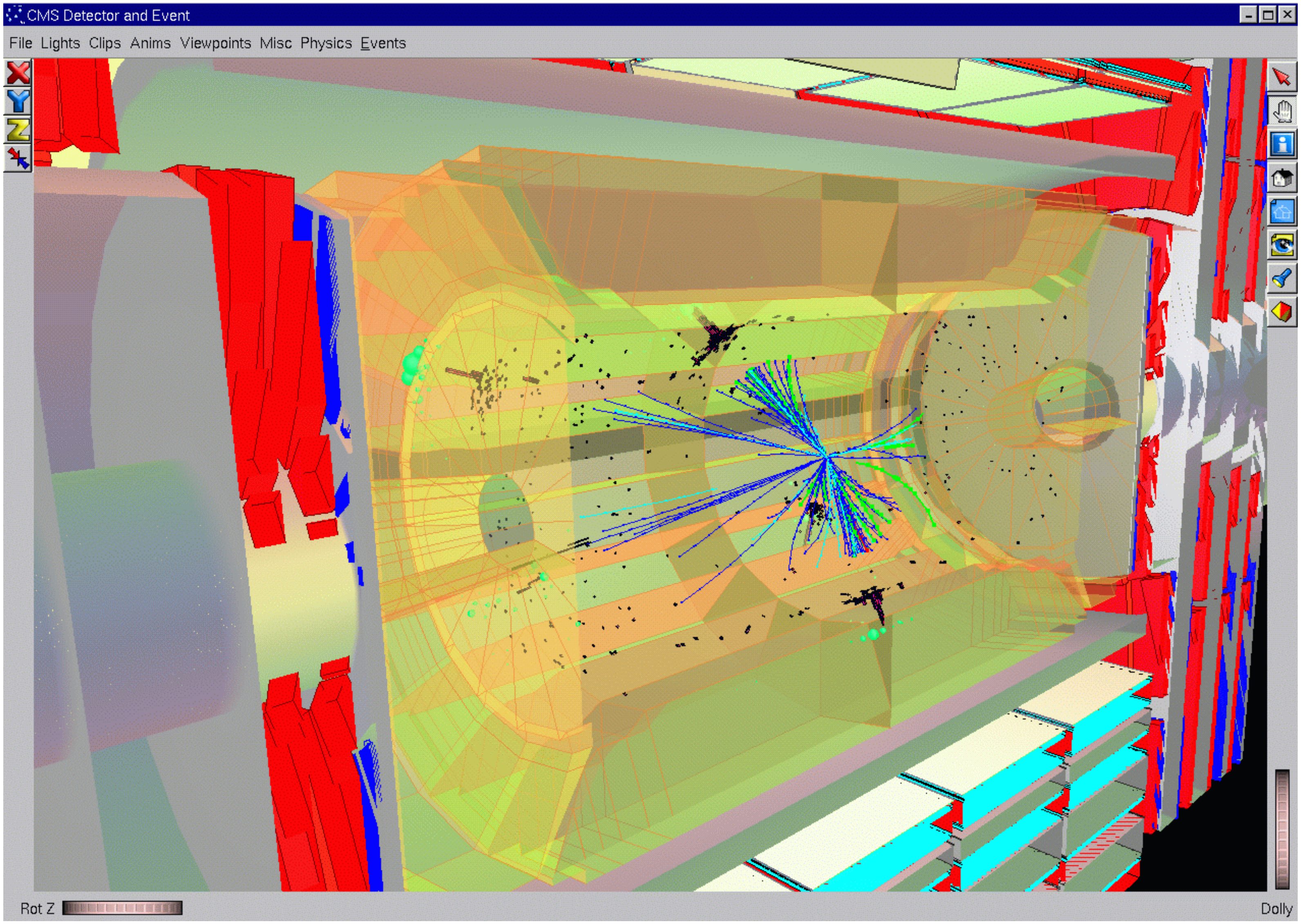
  US Contact: George Alverson ???

  Web documentation: http://cmsdoc.cern.ch/iguana/
  http://cmsdoc.cern.ch/iguanacms/

- More than just an event display!
- Can browse the event, print out objects, select objects, trigger "reconstruction on demand", etc…

IGUANA CMS:

- Black Hole production! M_{PL}=1 TeV, n=2



Data Model

Analysis Tools

- Right now there is no single CMS format for doing analysis
- The only complete data format is POOL
  - doing analysis means learning the framework and coding in C++
  - the work is going on on making POOL files readable in ROOT, though speed might prove to be problematic
- There are a number of system-specific ROOT formats
  - JetMET root-tuple
  - EGamma root-tuple
  - Track root-tuple

} enough information for some analyses, probably not enough for commissioning

Analysis Tools

- EGamma Ntuple is documented at
  http://cmsdoc.cern.ch/Physics/egamma/www/ntple.html
  - Branches: basic clusters, supercluster seeds, superclusters, preshower, L1 trigger info, pixel vertex, electron and photon HLT, Geant info, generator info, general event info
- JetMET Ntuple is documented at
  http://home.fnal.gov/~jdamgov/rootmaker/
  - Branches: some configuration parameters, generator info, pile-up info, generator jets for signal interaction and including pile-up, generator MET, unclustered energy, EGamma basic clusters, reconstructed jets, MET, L1 simulation for jets and taus, optional blocks with topological variables and L1 trigger primitives
- Track root tuple is documented at
  http://home.fnal.gov/xxxxxxxxxx

Summary

- The reconstruction is based on concept of "reconstruction on demand"
  - makes possible to run the same code off-line and in the HLT
- Major pieces of the software are in place
  - temptation to develop and tune "smarter" algorithms using the MC simulation which will not reproduce the data
  - need to start the experiment with most simple and robust algorithms
  - need a transition plan from simple to fancy (with implications to data format and management)