Introduction to Physics Computing

Rudi Frühwirth

Institute of High Energy Physics

Austrian Academy of Sciences, Vienna

CSC 2009, Göttingen

Outline of the lectures

- Introduction
- Event Filtering
- Calibration and alignment
- Event Reconstruction
- Event Simulation
- Physics Analysis
- Data Flow and Computing Resources

What is Physics Computing?

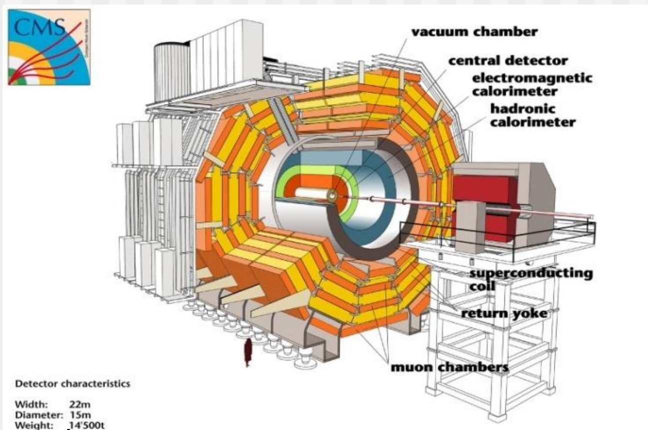
- Input: A few petabytes of data
- Output: A few hundred physics papers
- Data reduction factor of 10^7 to 10^8 !!
- How is it done?

<Digression>

</Digression>

CERN School of Computing

It’s simple … is it?




```
Paper paper15
Data higgsdata
...
paper15=make_paper(higgsdata)
...
```

INSTITUTE OF PHYSICS PUBLISHING — JOURNAL OF PHYSICS G: NUCLEAR AND PARTICLE PHYSICS

J. Phys. G: Nucl. Part. Phys. **31** (2005) 857–871 doi:10.1088/0954-3899/31/8/017

Electroweak phase transition in an extension of the standard model with a real Higgs singlet

S W Ham[1], Y S Jeong[2] and S K Oh[1,2]

[1] Center for High Energy Physics, Kyungpook National University Daegu 702-701, Korea
[2] Department of Physics, Konkuk University, Seoul 143-701, Korea

Received 1 December 2004
Published 24 June 2005
Online at stacks.iop.org/JPhysG/31/857

Abstract
The Higgs potential of the standard model with an additional real Higgs singlet is studied in order to examine if it may allow the strongly first-order electroweak phase transition. It is found that there are parameter values for which this model at the one-loop level with a finite-temperature effect may allow the desired phase transition. Those parameter values also predict that the masses of the neutral scalar Higgs bosons of the model are consistent with the present experimental bound, and that their production in e^+e^- collisions may be searched at the proposed ILC with $\sqrt{s} = 500$ GeV in the near future.

1. Introduction

The possibility of baryogenesis by means of electroweak phase transition has recently been widely examined, since the electroweak baryogenesis can, in principle, be tested in the future accelerator experiments [1]. If the electroweak phase transition is strongly first order, it can fulfil the departure from thermal equilibrium which is one of the three conditions required by Sakharov that are necessary for the dynamic generation of the baryon asymmetry during the evolution of the universe [2]. It has already been observed that in the standard model (SM) the electroweak phase transition cannot be strongly first order unless the mass of the scalar Higgs boson is smaller than its lower bound set experimentally by LEP [3]. The sufficient strength of the first-order electroweak phase transition is essential for preserving the generated baryon asymmetry at the electroweak scale. In the literature, a number of articles have been devoted to study the possibility of accommodating the strongly first-order electroweak phase transition in various models beyond the SM [4].

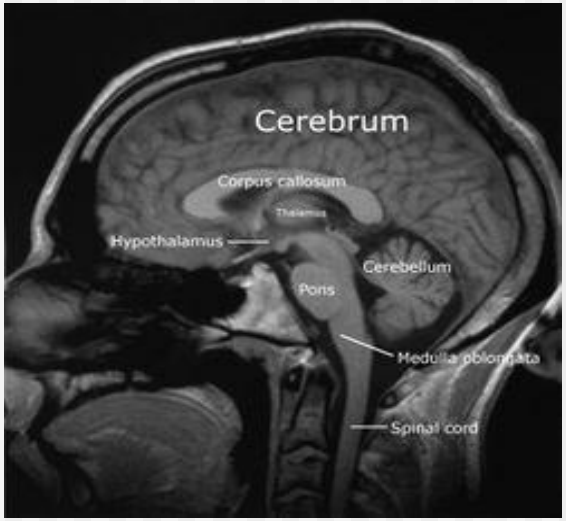
Among them, an interesting possibility has been investigated several years ago, where an extension of the SM with a real Higgs singlet field has been adopted within the context of the electroweak phase transition [5]. We consider that the model is inspiring, because adding a real Higgs singlet field is the simplest extension of the Higgs sector of the SM. In that model, the strength of the first-order electroweak phase transition has been stronger than that in the case of the SM. Due to the presence of cubic terms in the tree-level Higgs potential, a strongly

0954-3899/05/080857+15$30.00 © 2005 IOP Publishing Ltd Printed in the UK 857

CERN School of Computing

At LHC we need…

- Millions of lines of code (C++,Python, …)
- Hundreds of neural networks (BNNs, not ANNs)
- Large infrastructure
  - Customized hardware
  - PC farms
  - Database and storage systems
  - Distributed analysis facilities
  - The grid



What happens to the data?

- Event filtering, tagging and storage
- Calibration, alignment
- Event reconstruction
- Persistency
- Event simulation
- Physics analyses

Step by step

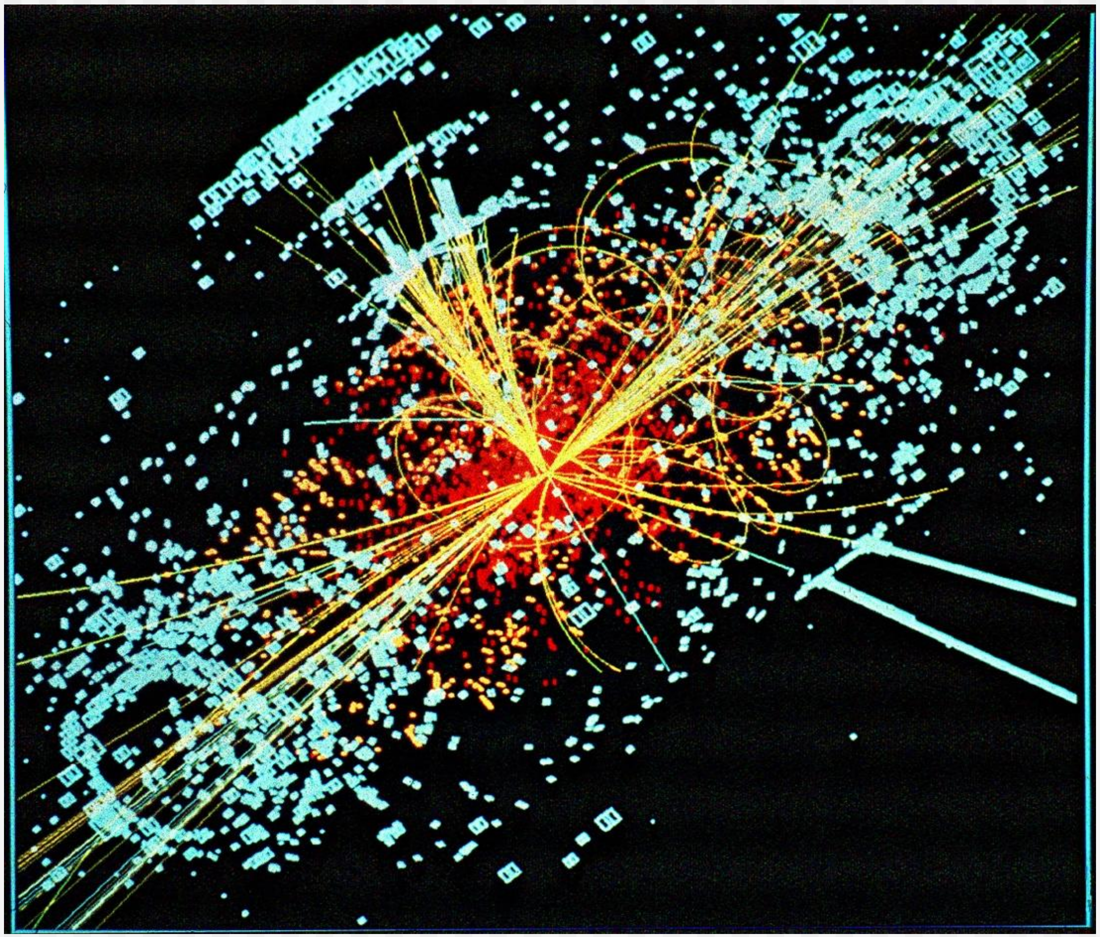
- Each step involves some data reduction
  - data are ignored or thrown away (online)
  - data are compressed (offline)

- In each step the data get closer to be interpretable in physical terms
- Some steps are repeated many times until the output is satisfactory

The technical challenge

- Very high event rate (40 Mhz)
- Large event size (>1MB)
- Large background of uninteresting events
- Large background in each event
  - many interactions in each beam crossing
  - many low-momentum particles

The technical challenge (ctd)



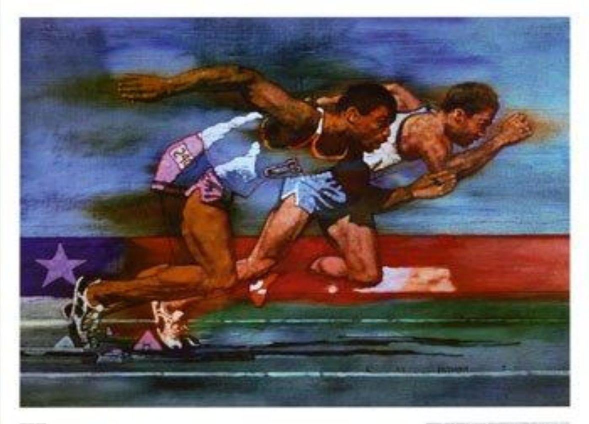
The social challenge

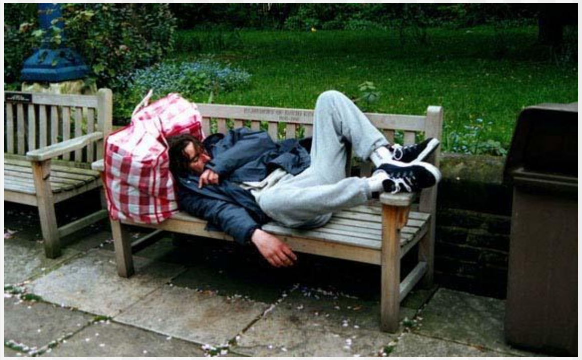
- Large number of physicists doing analysis
  - CMS: 183 institutes in 39 countries
- High pressure, competitive spirit
  - Important discoveries to be made
  - Fast turnaround required

CMS and ATLAS chasing the Higgs

Online vs Offline computing

- **Online**
  - In real time, fast!
  - Decisions are irreversible
  - Data cannot be recovered
- **Offline**
  - From almost real time to long delays
  - Decisions can be reconsidered
  - Data can be (and frequently are) reprocessed





Physicist reconsidering

Online processing

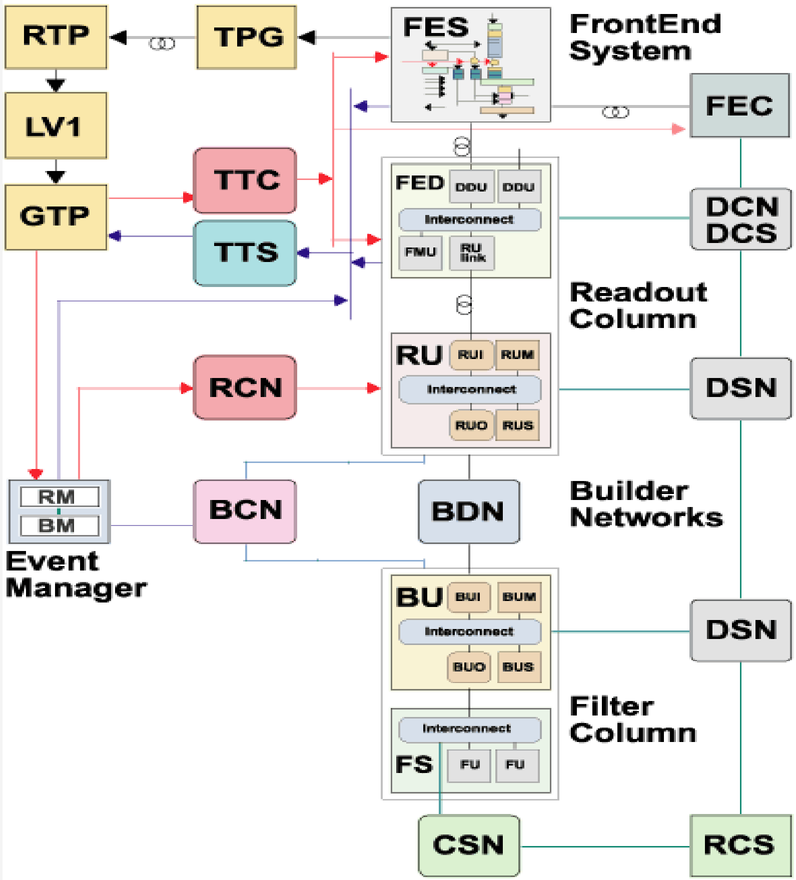
- Trigger: event selection
  - Needs only a small subset of the detector data
  - Fast, very little dead-time
  - Gives “green” or “red” light to the data acquisition

Online processing (ctd)

- Data acquisition
  - Interfaces to detector hardware
  - Builds complete events from fragments
  - Sends them to the higher level event filter(s)
  - Writes accepted events to mass storage
  - Very complex system

CERN School of Computing

Complexity of Data acquisition

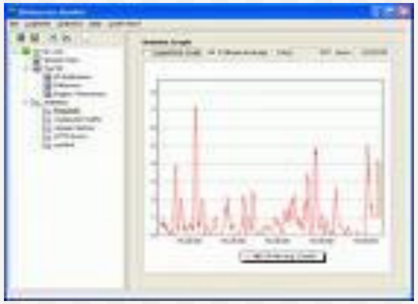


FLT	**FIRST LEVEL TRIGGER**
TPG	Trigger Primitive Generator
TDL	Trigger Data Link
RTP	Regional Trigger Processor
LV1	Calo & Trigger Processors
GTP	Global Trigger Processor
TTC	Timing, Trigger & Control
TTS	Trigger & Throttle System
EVB	**EVENT BUILDER**
BN	Builder Networks
BDN	Builder Data Network
RCN	Readout Control Network
BCN	Builder Control Network
EVM	Event Manager
RM	Readout Manager
BM	Builder Manager
FES	Front End System
DDL	Detector Data Link
RC	Readout Column
FED	FrontEnd Driver
FMU	Fast Monitoring Unit
DDU	Detector Dependent Unit
RDL	Readout Data Link
RU	Readout Unit
RUI	Readout Unit Input
RUM	Readout Unit Memory
RUO	Readout Unit Output
RUS	Readout Unit Supervisor
FC	Filter Column
BU	Builder Unit
BUI	Builder Unit Input
BUM	Builder Unit Memory
BUO	Builder Unit Output
BUS	Builder Unit Supervisor
FS	Filter Subfarm
FU	Filter Unit
HLT	High Level Trigger
RCS	**RUN CONTROL SYSTEM**
DCM	DAQ operation control and monitor
DDB	DAQ data bases
DCS	Detector Control System
DCN	Detector Control Network
FEC	FrontEnd Control
CDL	Control Data Link
CS	**COMPUTING SERVICES**
CSN	Computing and Services Network
DSN	DAQ Services Network
DA	Data archives

Computing and Communication main subsystems

Online processing (ctd)

- **Monitoring**
  - Detector status
  - Data acquisition performance
  - Trigger performance
  - Data quality check



- **Control**
  - Configure systems
  - Start/stop data taking
  - Initiate special runs (calibration, alignment)
  - Upload trigger tables, calibration constants, …

Event selection

- Primary collision rate: 40 Megahertz
- Recording rate: 100 Hertz
- How is this achieved?
  - Multilevel trigger – chain of yes/no decisions
  - Very fast first level: (Programmable) hardware
  - Slower higher level(s): Software on specialized or commodity processors

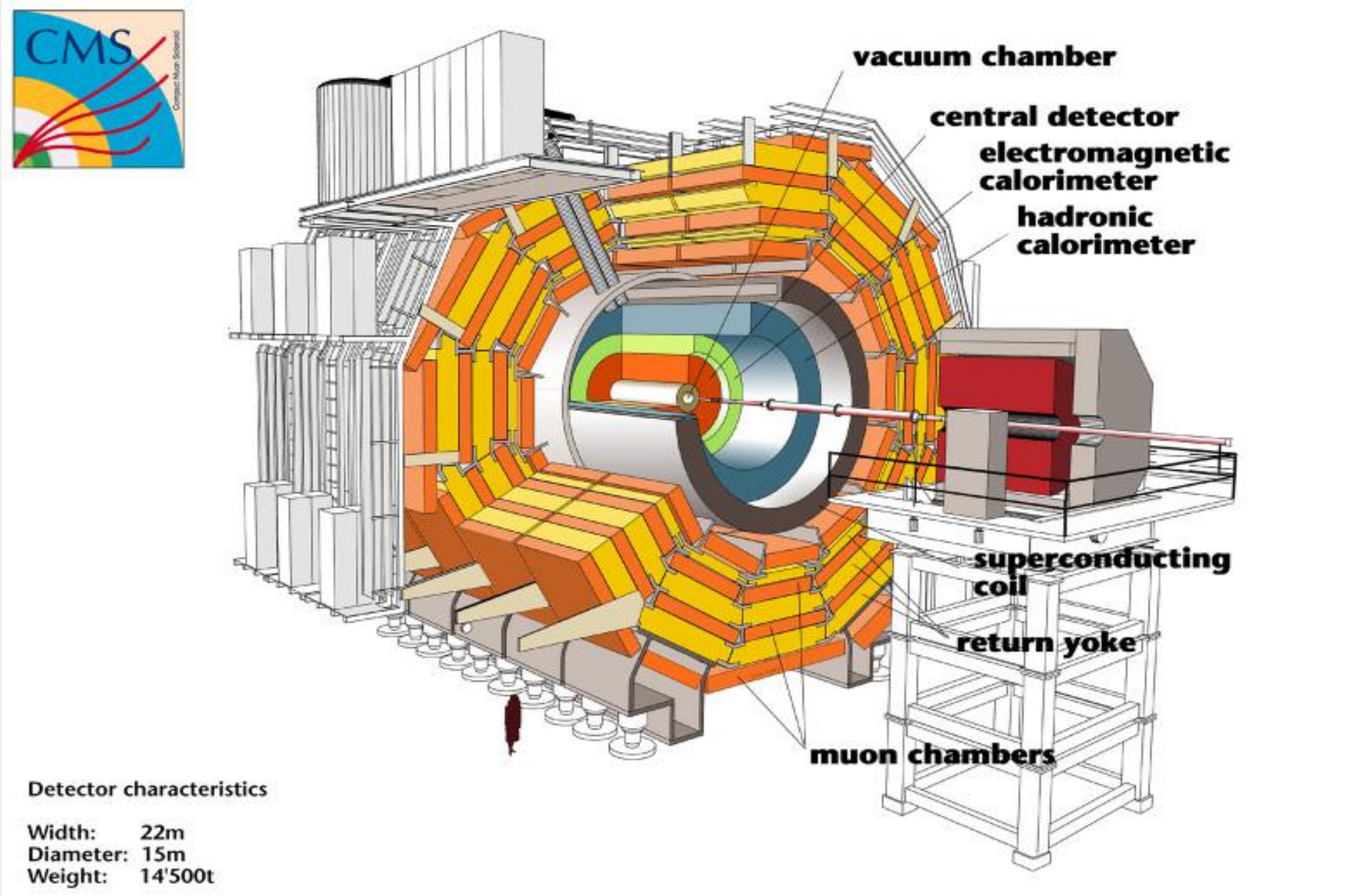
Event selection (ctd)

- Reliable
  - Rejected data are lost forever
  - Continuous monitoring
- Cautious
  - Do not lose new physics
- Versatile
  - Many different trigger channels run parallel
  - Trigger conditions can be changed quickly

Multilevel selection

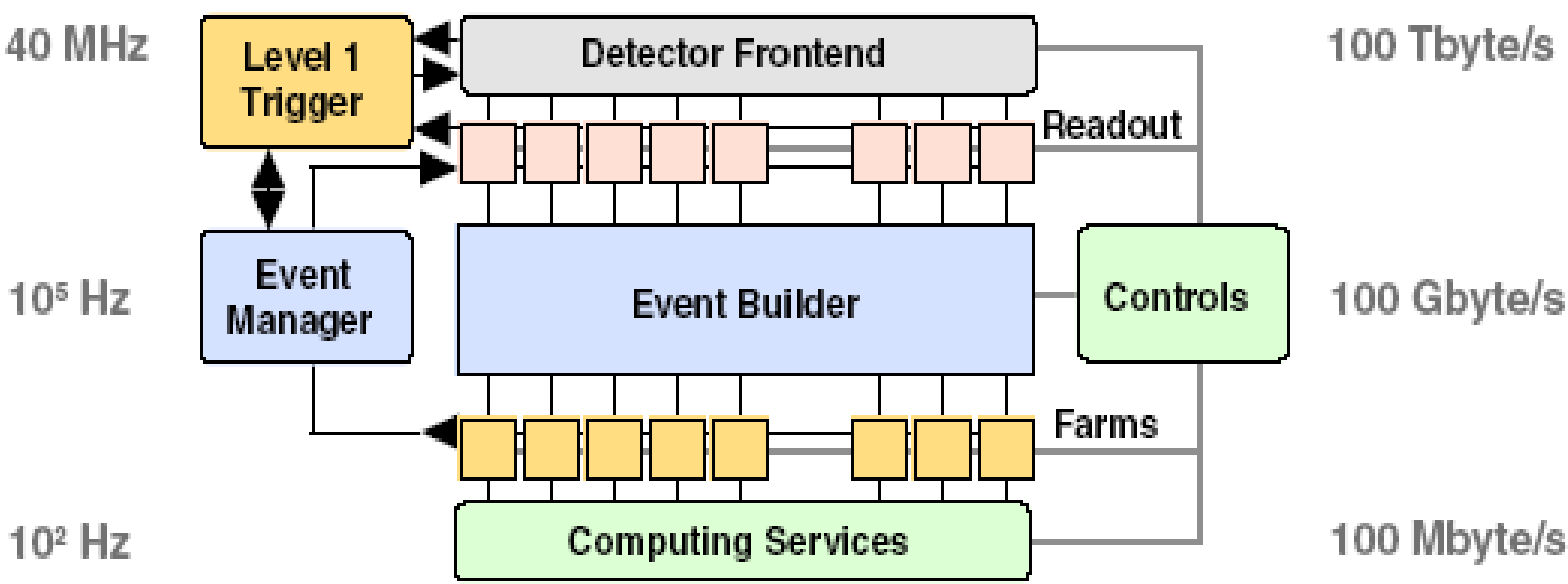
- Dead-time has to be minimized
- Many events can be discarded very quickly – Fast Level 1 Trigger
- Only the surviving ones are scrutinized more carefully – High Level Filter(s)
- Triggers are tailored to specific physics channels (Higgs, top, WW, ZZ, …)

Example: CMS


CMS
vacuum chamber
central detector
electromagnetic calorimeter
hadronic calorimeter
superconducting coil
return yoke
muon chambers
Detector characteristics
Width: 22m
Diameter: 15m
Weight: 14'500t

CERN School of Computing

Trigger/DAQ layout



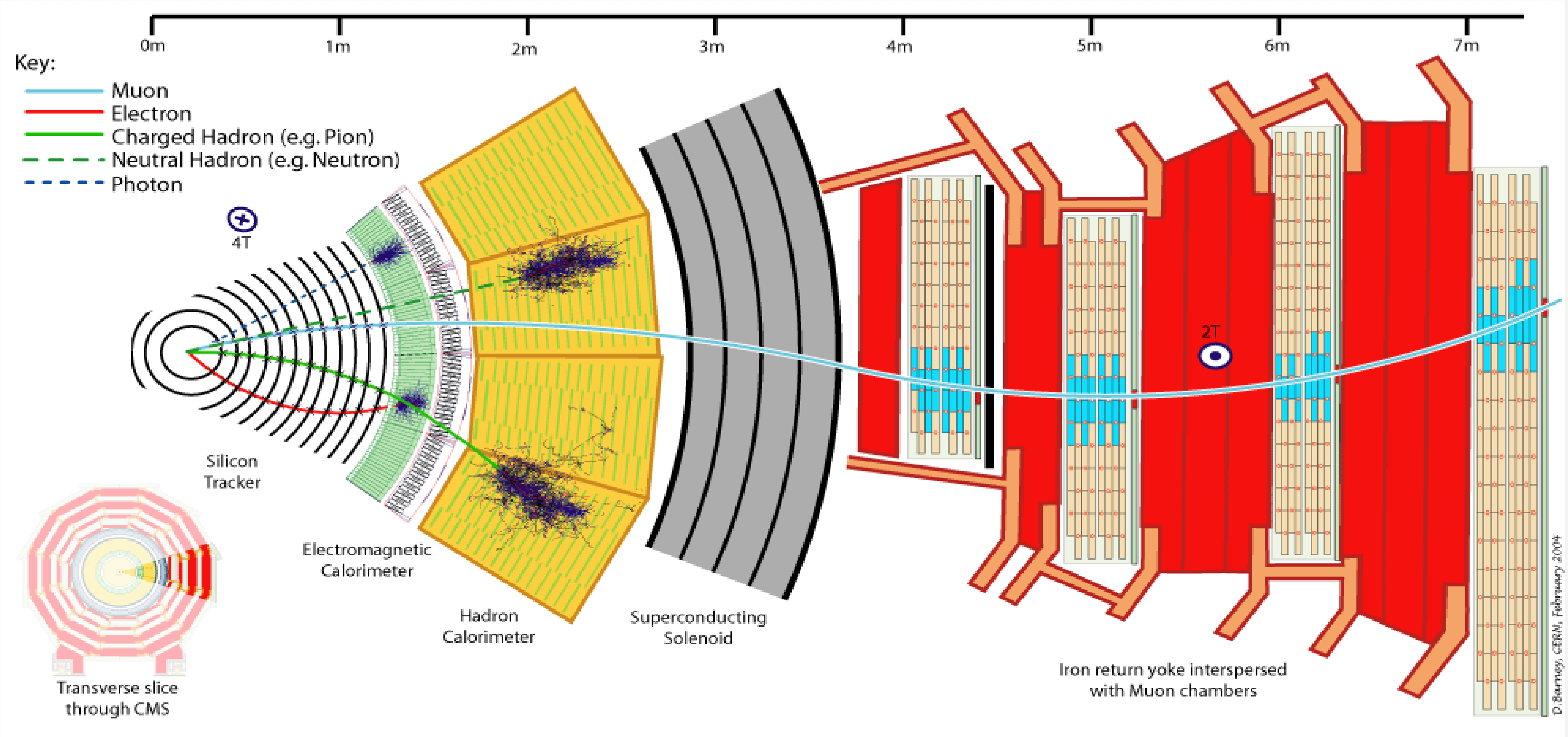
CERN School of Computing

What CMS subdetectors measure

- Inner tracker (pixels+strips)
  - Momentum and position of charged tracks
- Electromagnetic calorimeter
  - Energy of photons, electrons and positrons
- Hadron calorimeter
  - Energy of charged and neutral hadrons
- Muon system
  - Momentum and position of muons

CERN School of Computing

What CMS subdetectors measure



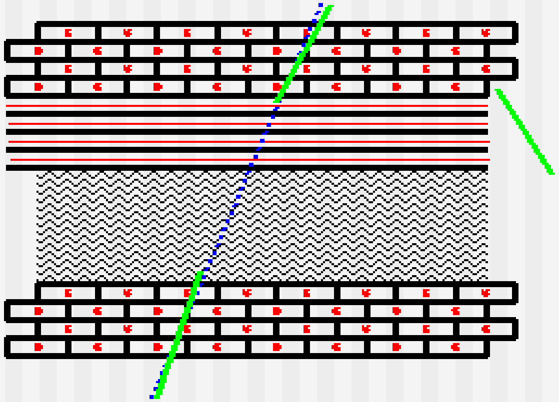
CMS L1 trigger

- Input rate: 40 MHz
- Output rate: 30 – 100 kHz
- Latency: 3.2 μs (128 BX)
- Pipelined, dead-time < 1%
- Available time for calculations: 1.25 μs
- 2 detector systems: muons/calorimeters
- 3 main steps: local/regional/global

CMS L1 calorimeter trigger

- Calorimeter trigger:
  - Two types of calorimeters: hadronic, electromagnetic
  - Local: Computes energy deposits
  - Regional: Finds candidates for electrons, photons, jets, isolated hadrons; computes transverse energy sums
  - Global: Sorts candidates in all categories, does total and missing transverse energy sums, computes jet multiplicities for different thresholds

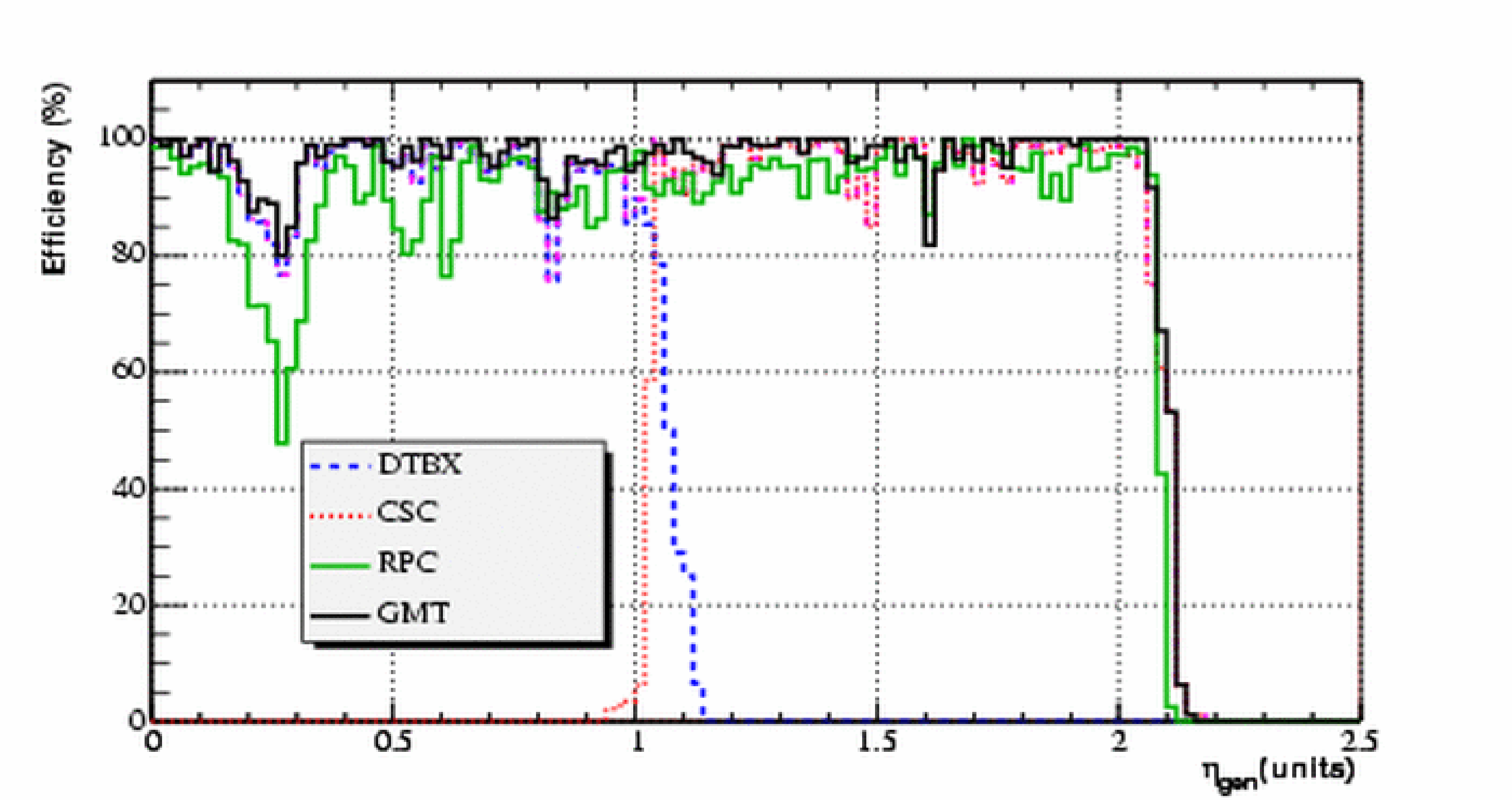
CMS L1 muon trigger



- Muon trigger:
  - Three types of muon detectors
  - Local: Finds track segments
  - Regional: Finds tracks
  - Global: Combines information from all regional triggers, selects best four muons, provides energy and direction

CERN School of Computing

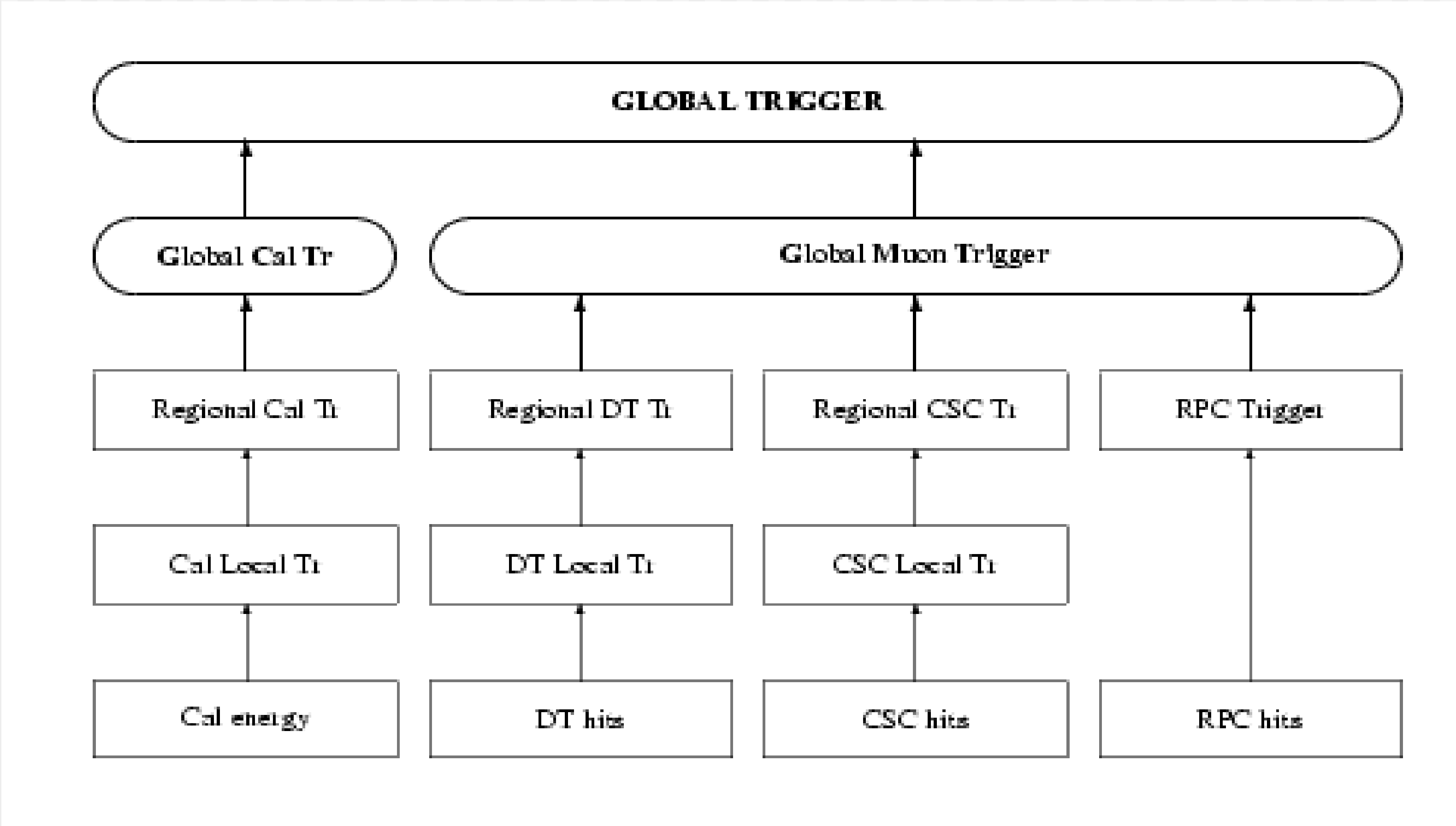
Efficiency of global muon trigger



CMS L1 global trigger

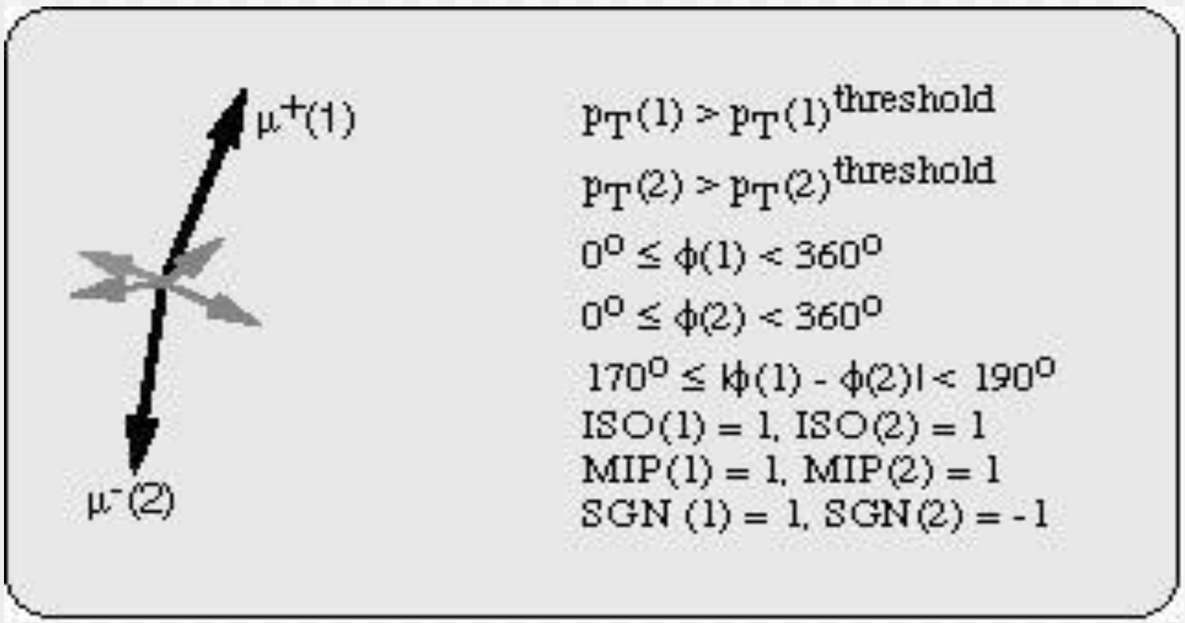
- Global trigger:
  - Final decision logic
  - 28 input channels (muons, jets, electrons, photons, total/missing E_T)
  - 128 trigger algorithms running in parallel
  - 128 decision bits
  - Apply conditions (thresholds, windows, deltas)
  - Check isolation bits
  - Apply topology criteria (close/opposite)

CMS L1 trigger



CMS L1 trigger example

- back-to-back opposite sign isolated muons



CMS L1 trigger software

- Algorithms are developed in C++
- They are tested by extensive simulation studies (→ Event Simulation)
- Manual translation into VHDL (***V***ery high speed integrated circuit ***H***ardware ***D***escription ***L***anguage)
- Comparison with C++ implementation

High level filter

- Further data selection:
  - 30 – 100 kilohertz input rate
  - 100-150 Hertz output rate
- Event tagging:
  - Reconstruct physics objects
  - Mark events having interesting features
  - Facilitates quick access later

High level filter (ctd)

- More detailed analysis of event and underlying physics
- Runs on standard processors (commodity PCs)
- CMS: 1 stage
- ATLAS: 2 stages (LVL2,Event filter)

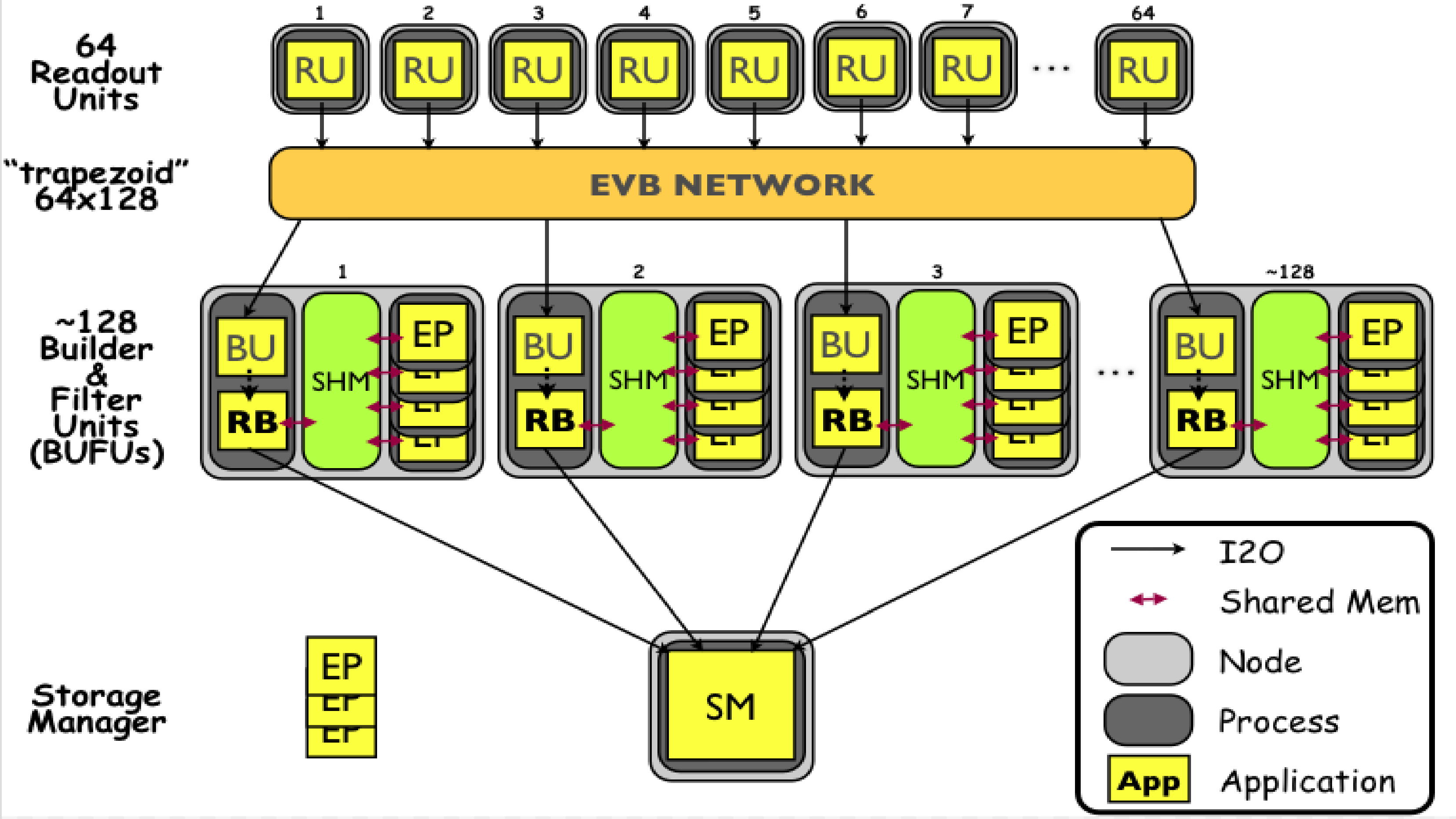
CMS High level trigger

- Has to keep pace with the L1 Output
- Solution: massive parallelism
- Filter farm
  - 720 PCs with dual quad-core Intel Harpertown @2.6 GHz, 16 GB RAM
  - Up to 200 events/s per PC
  - Decision time: ~ 40 ms

CMS High level trigger (ctd)

- Same software framework as in “offline” reconstruction
- Transparent exchange of algorithms with off-line code
- Regional reconstruction
  - Concentrates on region(s) found by Level 1
- Partial reconstruction
  - Stop as soon specific questions are answered

CMS HLT farm – schema …



… and reality

CERN School of Computing

Output of the high level trigger

- Raw data are sent to Tier-0 farm
  - Detector data (zero compressed)
  - Trigger information + some physics objects
  - O(50) primary datasets, depending on trigger history, O(10) online streams
- Physics: 1.5 MB @ 150 Hertz = 225 MB/sec
- Alignment/Calibration: 100 MB/sec

Output of the high level trigger (ctd)

- Total: 325 MB/s (~ 1/6th of maximum bandwidth)
- LHC runs for ~ 10^7 sec/year
- >3 PB per year!

CERN School of Computing

Tier-0 Processing

- Archive raw data on mass storage
- First event reconstruction without or with a small delay
- Archive reconstructed data on mass storage
  - 200 to 800 MByte/event, depending on physics
  - Reconstructed objects (hits/clusters, tracks, vertices, jets, electrons, muons)
- Send raw and processed data to Tier-1

Offline Processing

- Calibration
  - Convert raw data to physical quantities
- Alignment
  - Find out precise detector positions
- Event reconstruction
  - Reconstruct particle tracks and vertices (interaction points)
  - Identify particle types and decays
  - Impose physics constraints (energy and momentum conservation)

Offline Processing (ctd)

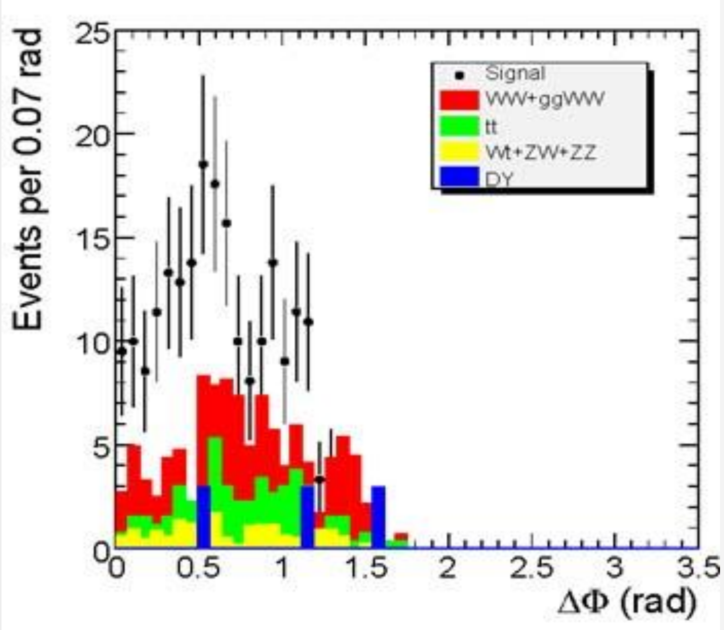
➢ Simulation

- Generate artificial events resembling real data as closely as possible
- Needed for background studies, corrections, error estimation, …

Monte Carlo Method

Offline Processing (ctd)

- Physics analysis
  - Extract physics signals from background
  - Compute masses, cross-sections, branching ratios, discovery limits, …



  - Requires sophisticated multivariate techniques
  - **Series of lectures and exercises on data analysis methods later in this track**

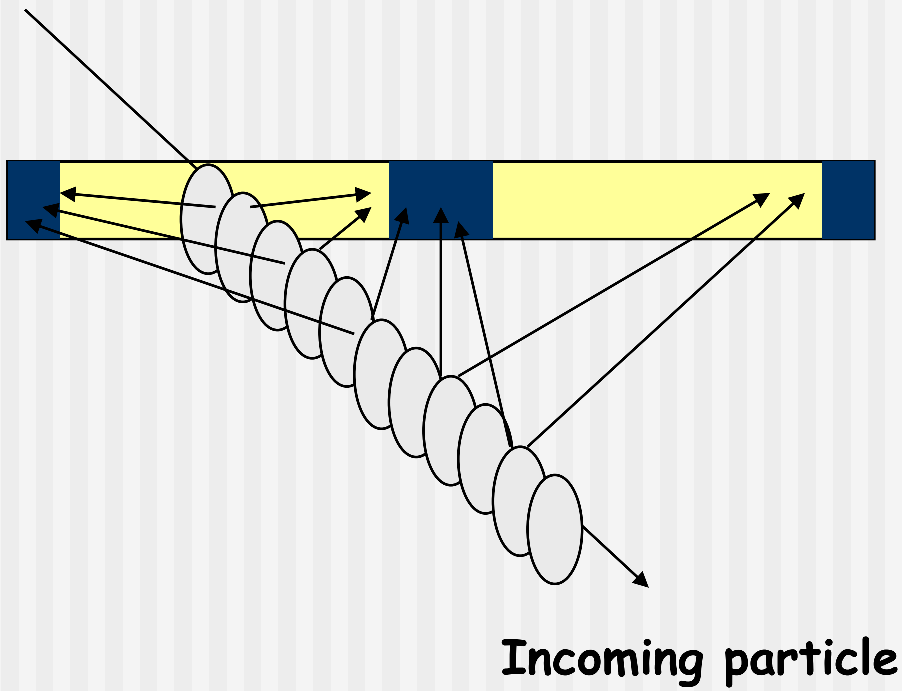
Calibration: From bits to GeV and cm

- Raw data are mostly ADC or TDC counts
- They have to be converted to physical quantities like energy or position
- Very detector dependent
- Every detector needs calibration
- Calibration constants need to be updated and stored

CERN School of Computing

Silicon Tracker calibration

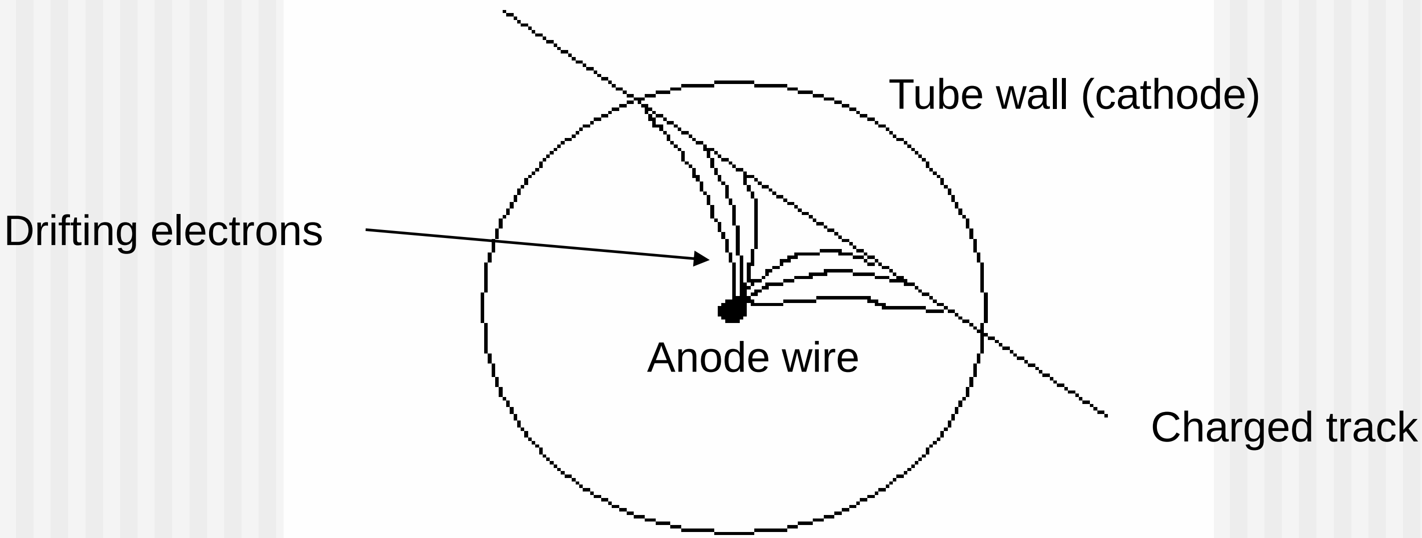
- Incoming particle creates electric charge in strips or pixels



Silicon Tracker calibration (ctd)

- Charge distribution depends on location of crossing point and crossing angle
- Solve inverse problem: reconstruct crossing point from charge distribution and crossing angle
- Test beam, real data

Drift tube calibration



Drift tube calibration (ctd)

- Incoming particle ionizes gas in tube
- Electrons/ions drift to anode/cathode
- Drift time is measured
- Must be converted to drift distance
- Time/distance relation must be determined (not always linear)
- Test beam, real data

Alignment: Where are the detectors?

- Tracking detectors are very precise instruments
- Silicon strip detector: ~ 50 μm
- Pixel detector: ~ 10 μm
- Drift tube: ~ 100 μm
- Position needs to be known to a similar or better precision

Example: CMS tracker

CERN School of Computing

Alignment

- Mechanical alignment
- Measurements taken before assembly
- Switching on the magnetic field
- Laser alignment
- Alignment with charged tracks from collisions, beam halo and cosmic rays

Alignment (ctd)

- Difficult because of huge number of parameters to be estimated (≈ 100000)
- Continuous process
- Alignment constants need to be updated and stored

Environmental data

- Calibration data
- Alignment data
- Temperatures, gas pressures, …
- Machine parameters
- Need to be made persistent

Detector related software

- Configuration
  - Load trigger files, set thresholds, set HV, set amplifier gains, …
- Slow control
  - Measure and adjust temperature, gas pressure, dark currents, …
- Monitoring
  - Check trigger rates, detector efficiency, cluster sizes, wire maps, …

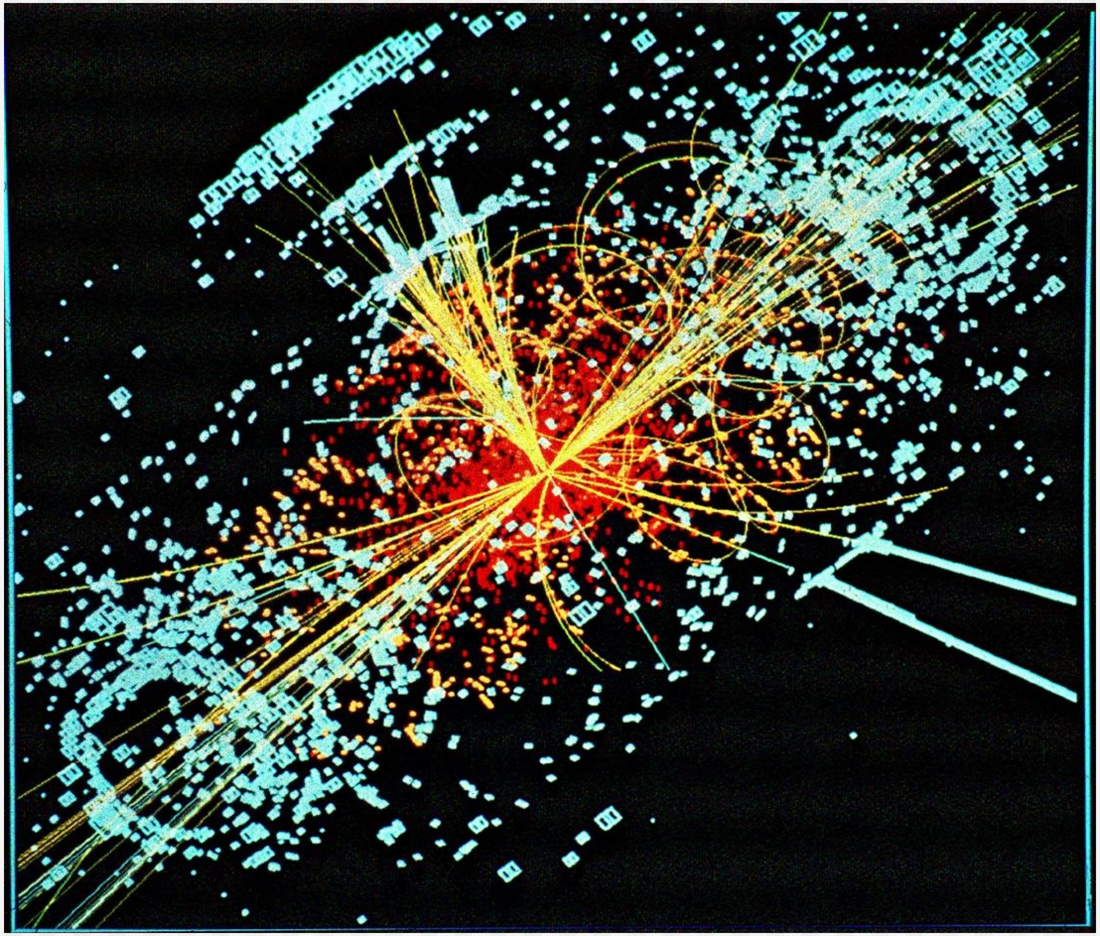
Event reconstruction

- Find out which particles have been created where and with which momentum
- Some of them are short-lived and have to be reconstructed from their decay products
- Some of them (neutrinos) escape without leaving any trace

Event reconstruction (ctd)

- Reconstruct charged particles
- Reconstruct neutral particles
- Identify type of particles
- Reconstruct vertices (interaction points)
- Reconstruct kinematics of the interaction
- Not trivial, very time-consuming …

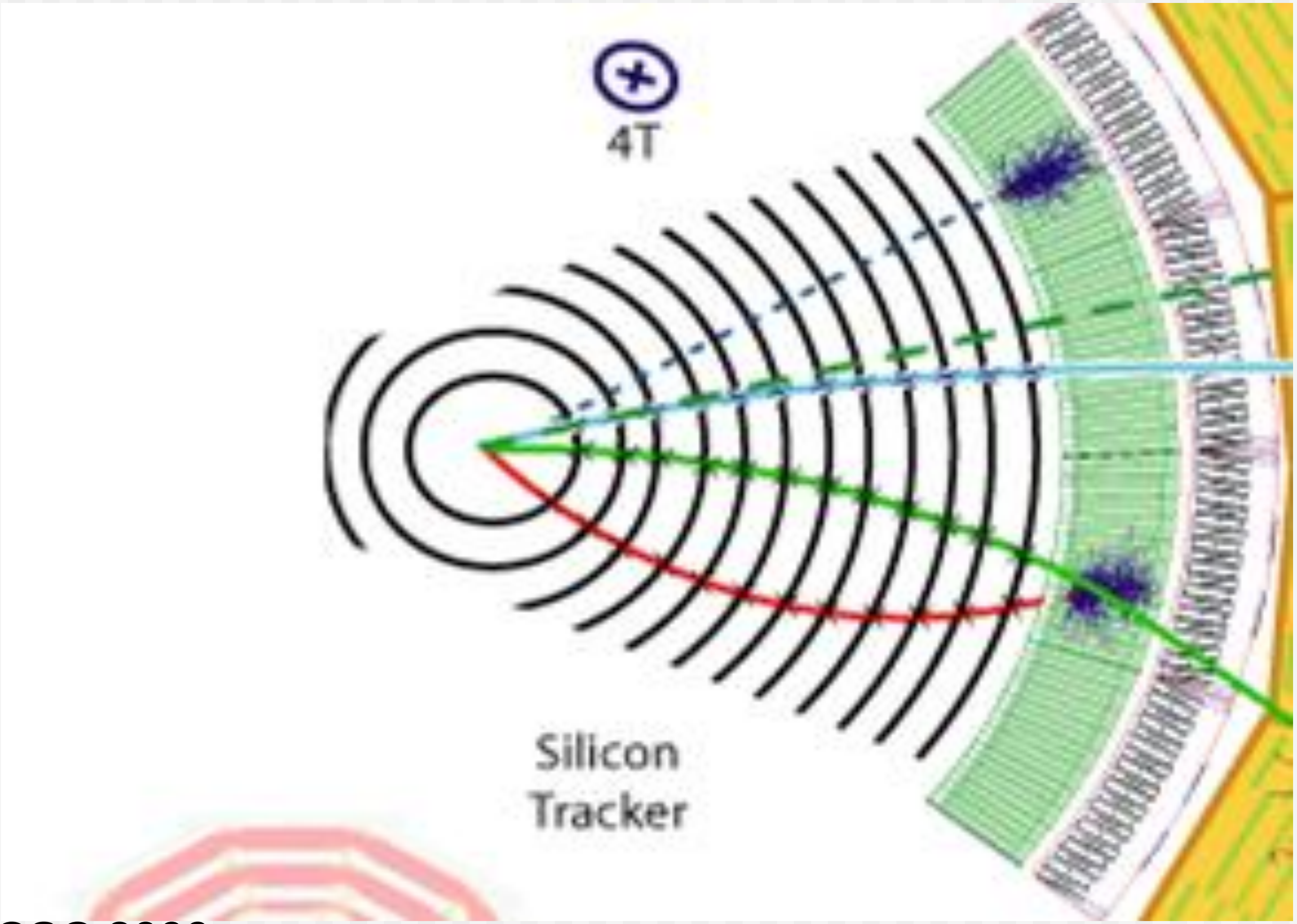
CMS: Higgs decay into two jets



Charged particles

- Charged particles are detected by tracker and calorimeters
- Muons also reach the muon system
- Very high number of low-momentum charged particles
- Select by threshold on transverse momentum

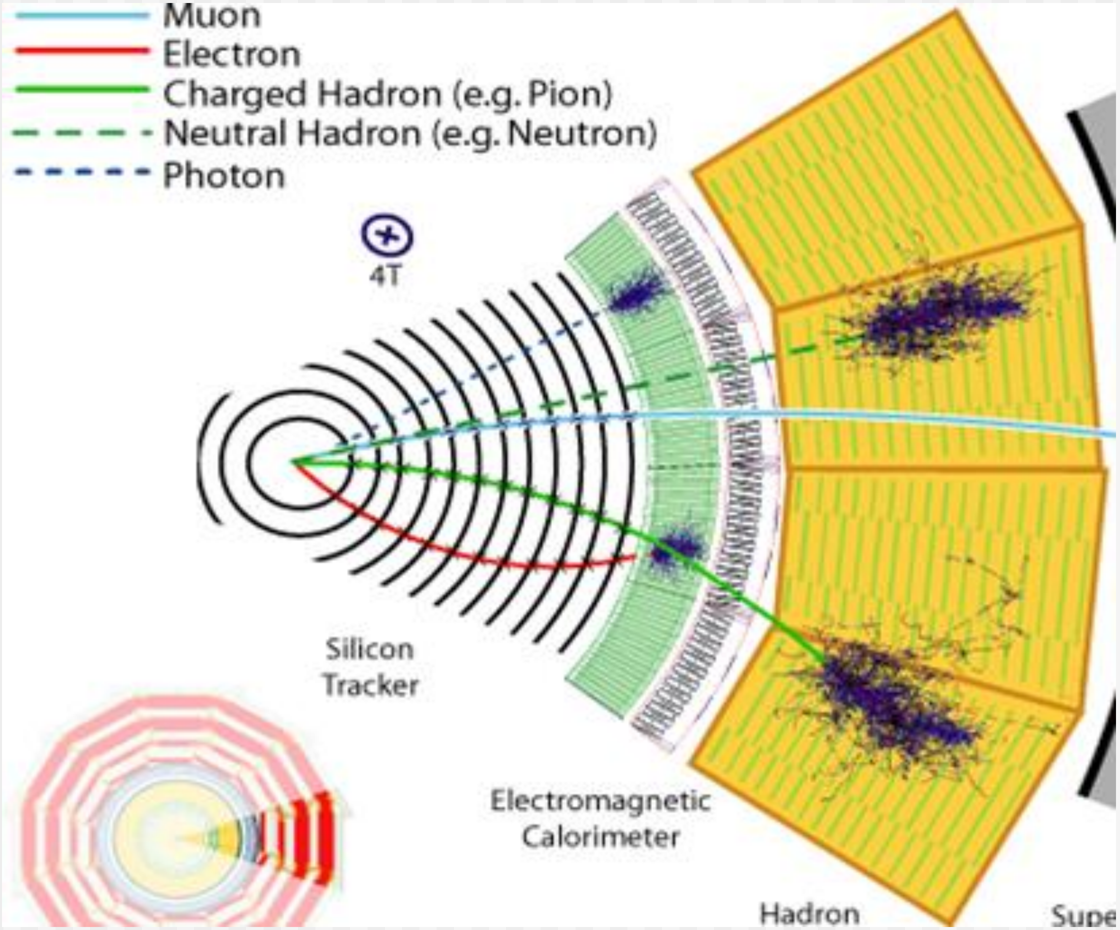
Charged particles (ctd)



Neutral particles

- Neutral particles are detected mainly by calorimeters (e.g. photons, neutrons)
- They should deposit their entire energy
- Some of them decay into two (or more) charged tracks which are detected by the tracker (e.g. K^0)
- Some of them escape without leaving a trace (neutrinos)

Neutral particles (ctd)



Reconstruction of charged particles

- Trajectory is curved because of the magnetic field
- Position is measured in a number of places –"hits"
- Determine track parameters (location, direction, momentum) plus errors from the position measurements
- Data compression

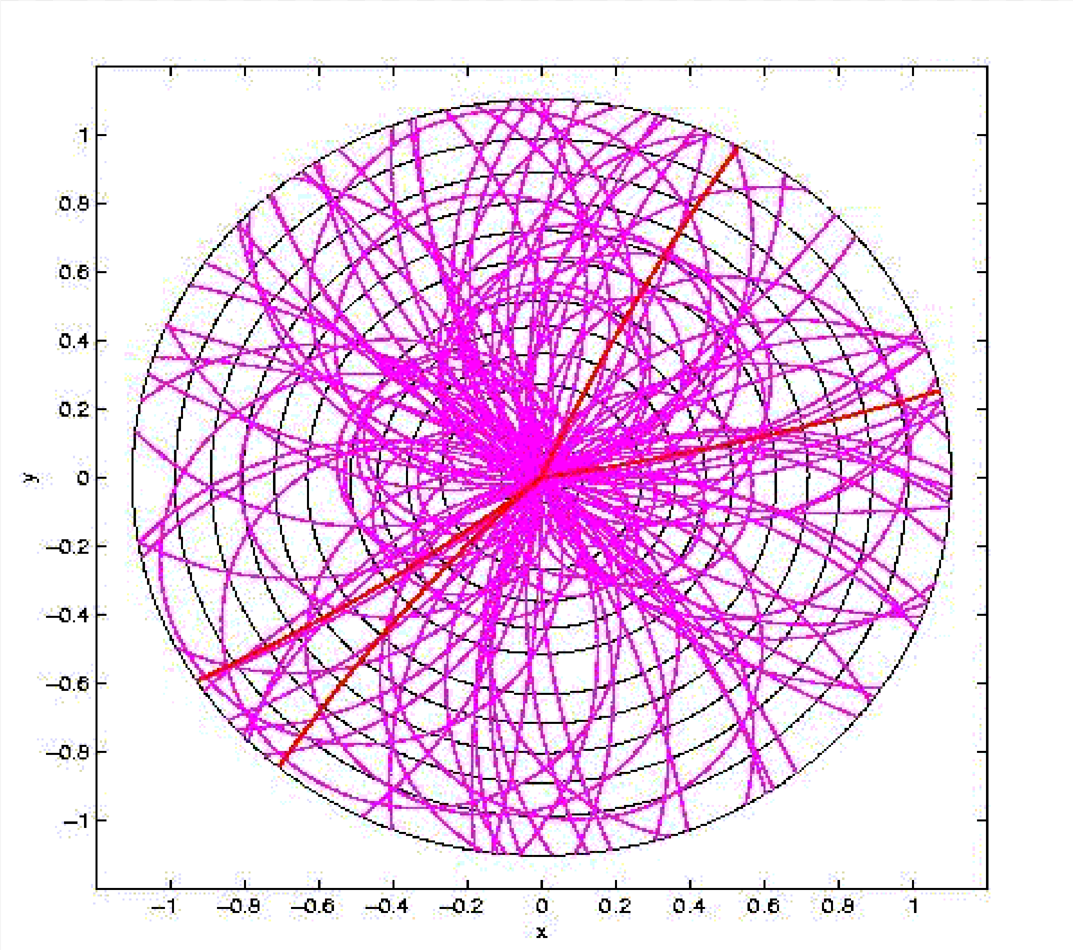
The difficulties

- Assignment of hits to particles is unknown
- Huge background from low-momentum tracks
- Additional background from other interactions in the same beam crossing and from adjacent beam crossings

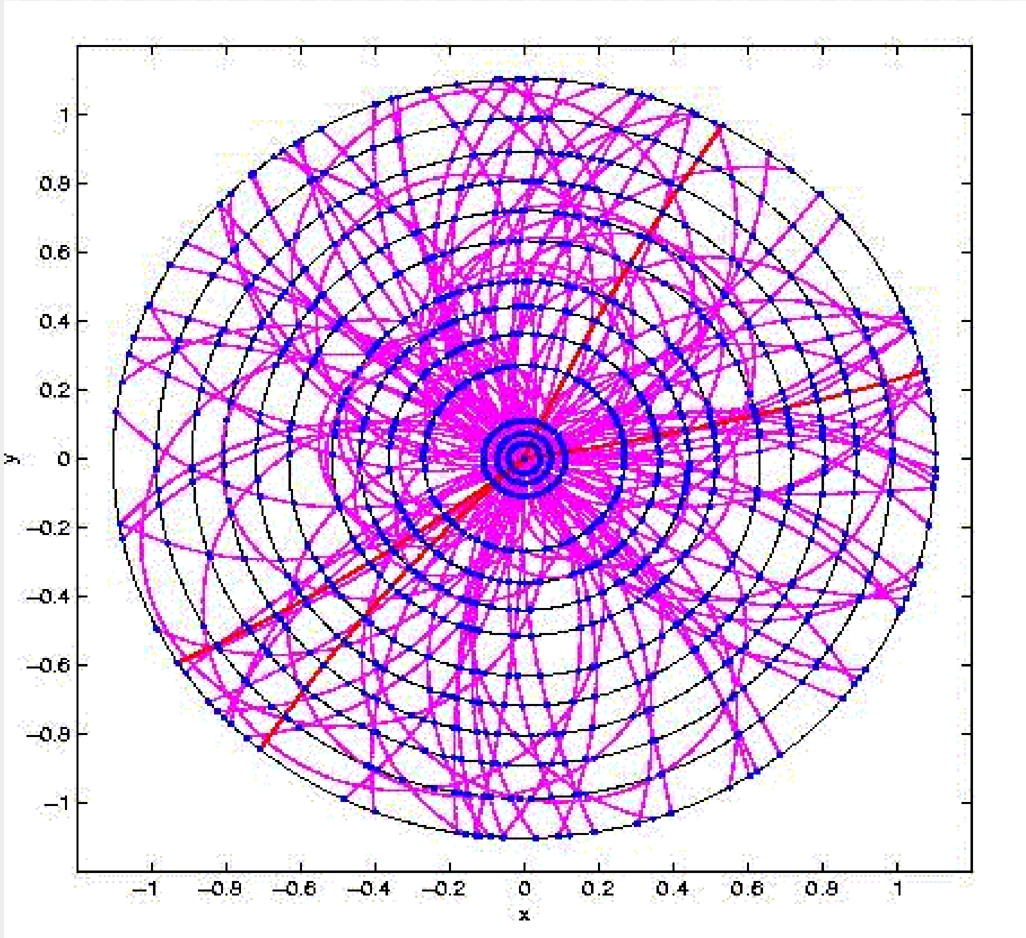
More difficulties

- Charged particles interact with all the material, not only the sensitive parts
- Multiple Coulomb scattering
  - Changes direction, but not momentum
- Energy loss by ionization
  - All charged particles
- Energy loss by bremsstrahlung
  - Mainly electrons and positrons

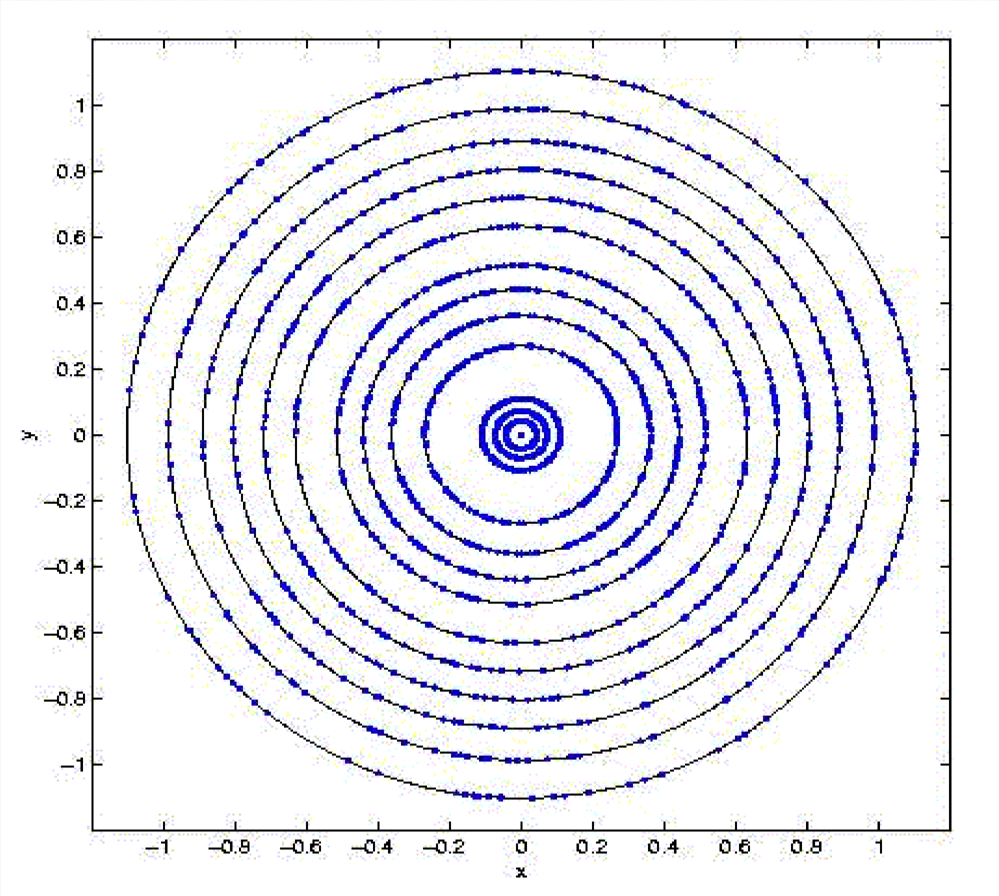
Tracks only



Tracks with hits



Hits only



?

Decomposition of the problem

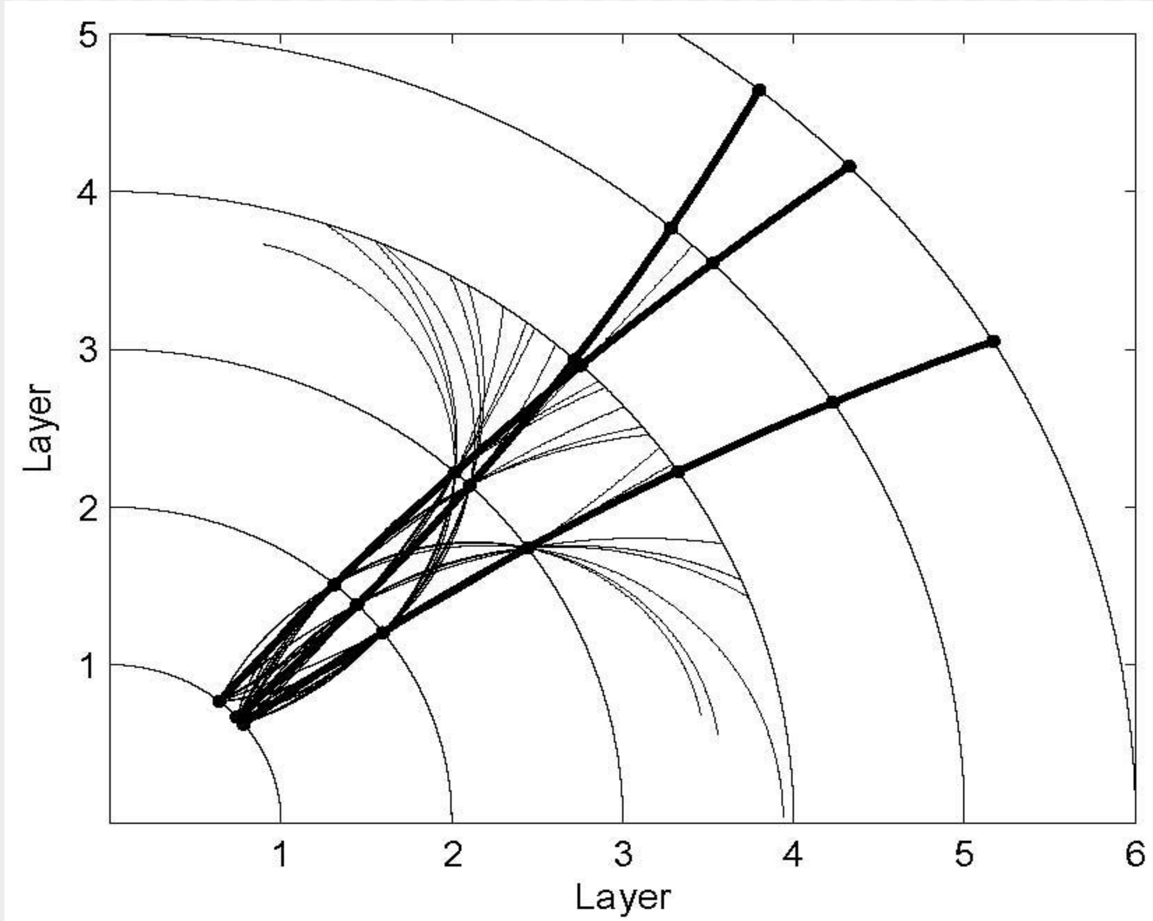
- Pattern Recognition or Track Finding
  - Assign hits to track candidates
- Parameter estimation or Track Fit
  - Determine track parameters + covariance matrix
- Test of the track hypothesis
  - Check chi-square, residuals, remove outliers

Track finding

- Depends a lot on the properties of the detector:
  - Geometry, configuration
  - Magnetic field
  - Precision
  - Occupancy
- Many solutions available
- No general recipe

A few track finding algorithms

- Track following
- Kalman filter
- Combinatorial Kalman filter
- Hough transform
- Artificial neural network



Track fit

- Determine track parameters
- Determine errors (covariance matrix)
- Test track hypothesis
- Reject outliers
  - Distorted hits
  - Extraneous hits
  - Electronic noise

Ingredients

- Magnetic field
  - Constant or variable
- Track model
  - Solution of the equation of motion
  - Analytic (explicit) or numerical
- Error model
  - Observations errors
  - Process noise

Estimation of track parameters

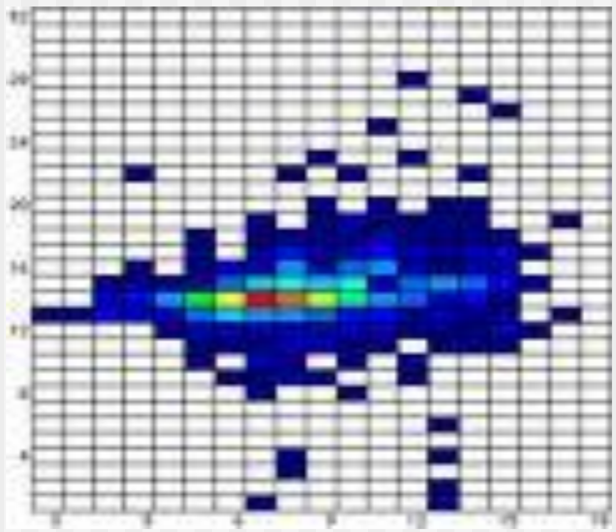
- Most estimators minimize a least-squares objective function
- Least-squares estimation
  - Linear regression
  - Kalman filter
- Robust estimation
  - Adaptive filter

Reconstruction of neutral particles

- Neutral particles are only seen by the calorimeters
- Photons are absorbed in the electromagnetic calorimeter
- Neutral hadrons are absorbed in the hadronic calorimeter
- Neutrinos are not detected directly

Shower finding

- An incident particle produces a shower in the calorimeter
- A shower is a cluster of cells with energy deposit above threshold

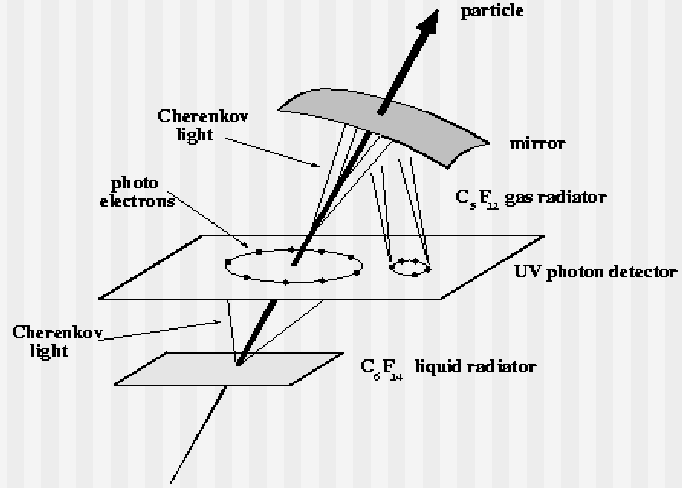


Shower finding (ctd)

- Overlapping clusters must be separated
- Various clustering techniques are used to find showers
- The algorithms depend on various characteristics of the calorimeter
  - Type (electromagnetic or hadronic)
  - Technology (homogeneous or sampling)
  - Cell geometry, Granularity

Particle identification

- Determining the type of a particle
- Dedicated detectors
  - Type (electromagnetic or hadronic)
  - Threshold Cherenkov
  - Ring imaging Cherenkov (RICH)
  - Transition radiation detector
  - Ionization measurements



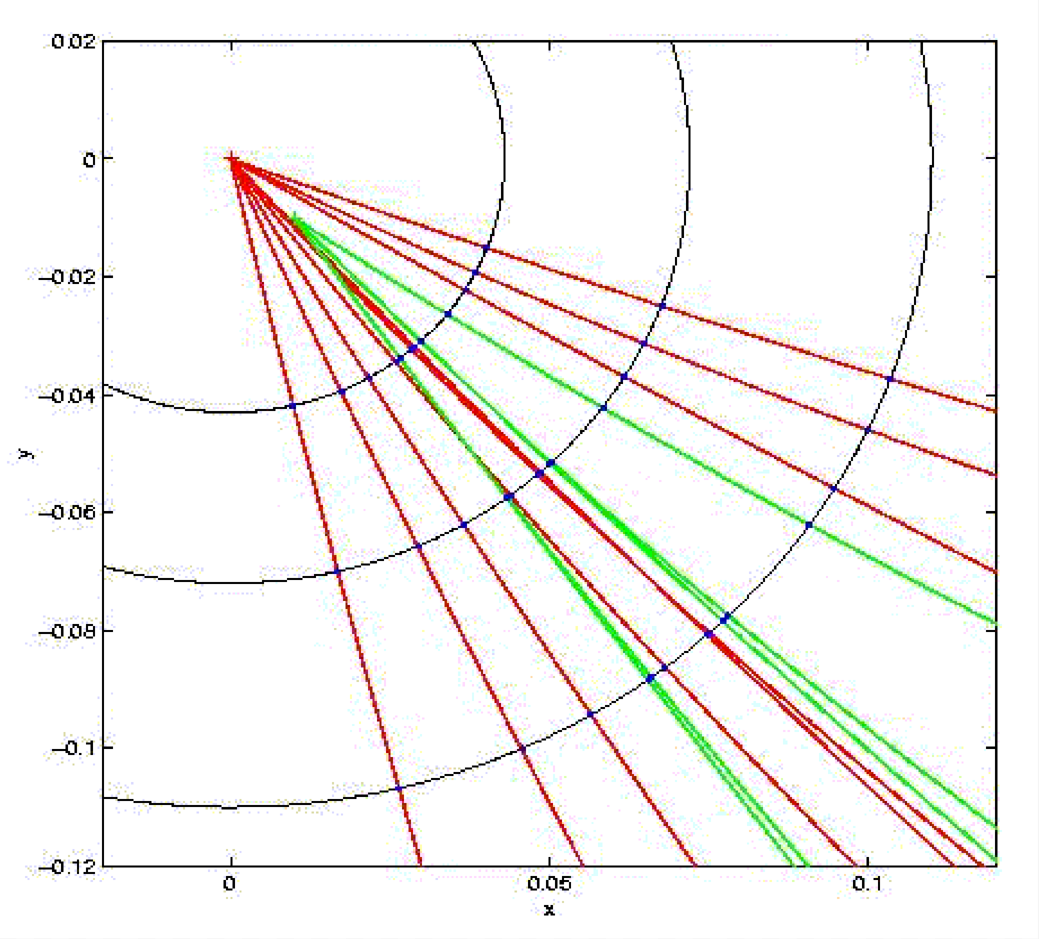
Particle identification (ctd)

- Combining information from several detectors
  - Shower in electro-magnetic calorimeter
    + no matching track in tracker $\rightarrow$ **photon**
  - Shower in electro-magnetic calorimeter
    + matching track in tracker $\rightarrow$ **electron/positron**
  - Shower in hadronic calorimeter
    + matching track in tracker $\rightarrow$ **charged hadron**
  - Track in muon system
    + matching track in tracker $\rightarrow$ **muon**

Vertex reconstruction

- Primary vertex: interaction of the two beam particles – *easy*
- Secondary vertices: decay vertices of unstable particles – *difficult*
- Emphasis on short-lived unstable particles which decay before reaching the tracker
- Data compression

Primary and secondary tracks



Primary tracks

Secondary tracks

The difficulties

- Association of tracks to vertices is unknown
- Secondary tracks may pass very close to the primary vertex
  - Especially if decay length is small
- Track reconstruction may be less than perfect
  - Outliers, distortions, incorrect errors

Decomposition of the problem

- Pattern Recognition or Vertex Finding
  - Assign tracks to vertex candidates
- Parameter estimation or Vertex Fit
  - Determine vertex location + covariance matrix, update track parameters
- Test of the vertex hypothesis
  - Check chi-square, residuals, remove outliers

Vertex finding

- Almost independent of the detector geometry
- Secondary vertex finding may depend on the physic channel under investigation
- Essentially a clustering problem
- Many solutions available

A few vertex finding algorithms

- Hierarchical clustering
  - Single linkage, complete linkage,…
- Non-hierarchical clustering
  - k-means, robust location (mode) estimation, iterated vertex fit
- Neural network/physics inspired
  - Competitive learning, deterministic annealing, superparamagnetic clustering, quantum cluste

Vertex fitting

- Most estimators minimize a least-squares objective function
- Least-squares estimation
  - Linear regression
  - Kalman filter
- Robust estimation
  - Adaptive filter

Persistency

- Event reconstruction produces physics objects
  - Tracks
  - Vertices
  - Identified particles
  - Jets
  - Tags
- Need to be made persistent

Persistency (ctd)

- Physics objects depend on
  - Alignment
  - Calibration
  - Version of the reconstruction program
  - Algorithm parameters
- Must be made persistent as well
- Tools: **ROOT**, POOL
  - Series of lectures and exercises later in this track

Simulation

- Why do we need simulation?
  - Optimization of detector in design phase
  - Testing, validation and optimization of trigger and reconstruction algorithms
  - Computation of trigger and reconstruction efficiency
  - Computation of acceptance corrections
  - Background studies
  - Systematic error studies

Simulation steps

- **Physics generation**
  - Generate particles according to specific physics processes
- **Event simulation**
  - Track particles through the detector, using detector geometry and magnetic field
  - Simulate interaction of particles with matter
  - Generate signals in sensitive volumes
  - Simulate digitization process (ADC or TDC)
  - Simulate trigger response

Simulation steps (ctd)

- Reconstruction
  - Treat simulated events exactly as real events
  - Keep (some) truth information: association of hits to tracks, association of tracks to vertices, true track parameters, true vertex parameters, …
  - Make everything persistent

Physics generation packages

- General purpose event generators
  - Hadron-hadron, hadron-lepton, lepton-lepton collisions
  - PYTHIA/JETSET, also known as “Lund Monte Carlo”
  - Herwig++, Hadron Emission Reactions With Interfering Gluons
  - PANDORA, event generator for linear collider studies, collisions of electrons, positrons and photons
- Specialized generators

CERN School of Computing

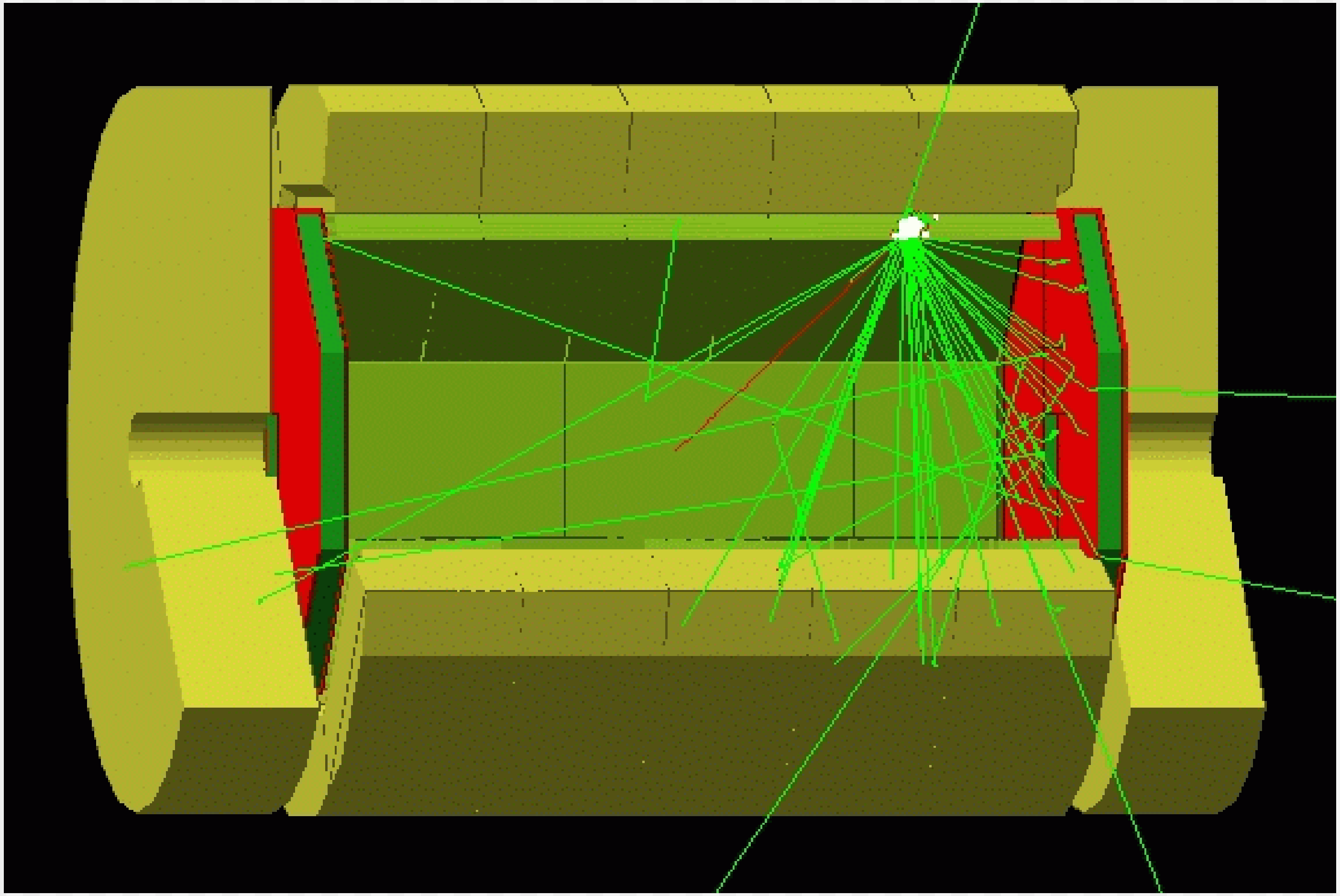
Event simulation

- Was frequently (and still sometimes is) experiment-specific
- Now there is a widely used standard: GEANT
  - GEANT3: procedural, FORTRAN
  - GEANT4 : object oriented, C++

Detector description

- Geometry
  - Shape
  - Placement relative to mother volume
  - Symmetries
- Material
  - Composition
  - Density
  - Radiation length, interaction length, …

An example detector model



Physics analysis

- Event selection
  - Multidimensional criteria
  - Statistics, neural networks, genetic algorithms, …
- Signal extraction
  - Study background
  - Determine significance of signal

- Corrections
  - Detector acceptance, reconstruction efficiency, …
  - From simulated data

Physics analysis (ctd)

- Computation of physical quantities
  - Cross sections, branching rations, masses, lifetimes, …
- … and of their errors
  - Statistical errors: uncertainty because of limited number of observations
  - Systematic errors: uncertainty because of limited knowledge of key assumptions (beam energy, calibration, alignment, magnetic field, theoretical values, background channels, …)

Analysis tools

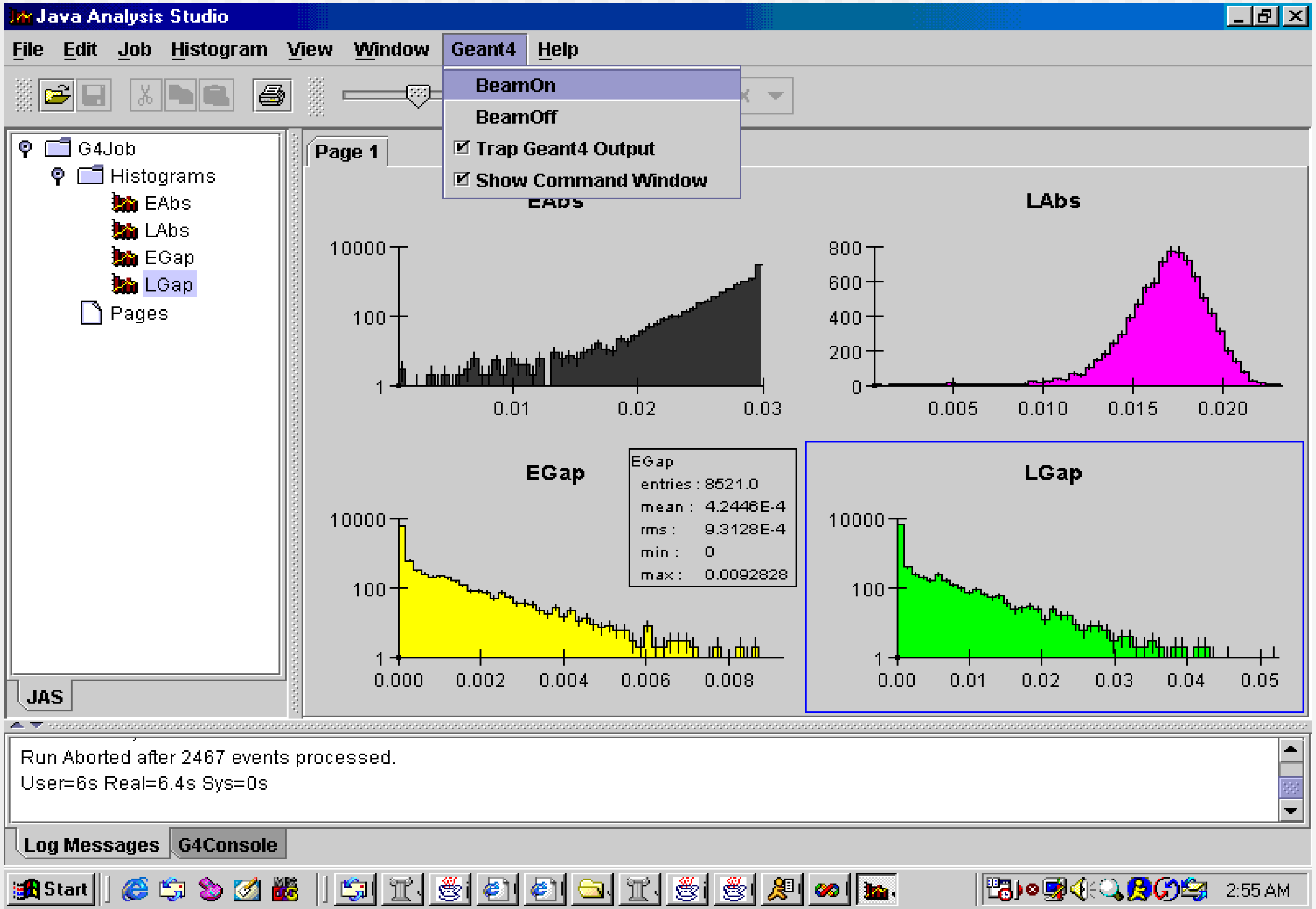
- Need versatile tools for
  - Multidimensional selection
  - Event display and interactive reprocessing
  - Histogramming
  - Plotting
  - Fitting of curves and models
  - Point estimation, confidence intervals, limits
  - …

CERN School of Computing

Analysis tools (ctd)

- **ROOT**
  - Data analysis and persistency, but also detector description, simulation, data acquisition, …
  - Series of lectures and exercises later in this track
- JAS
  - Java Analysis Studio (SLAC)
- WIRED
  - Platform independent event display (Java, SLAC)

JAS screenshot



And finally …

FERMILAB-PUB-07/094-E

Measurement of the Λ_b lifetime in the exclusive decay $\Lambda_b \to J/\psi\ \Lambda$

We have measured the Λ_b lifetime using the exclusive decay $\Lambda_b \to J/\psi\ \Lambda$, based on 1.2 fb^{-1} of data collected with the D0 detector during 2002–2006. From 171 reconstructed Λ_b decays, where the J/ψ and Λ are identified via the decays $J/\psi \to \mu^+\mu^-$ and $\Lambda \to p\pi$, we measured the Λ_b lifetime to be $\tau(\Lambda_b) = 1.218^{+0.130}_{-0.115}(\text{stat}) \pm 0.042(\text{syst})$ ps. We also measured the B^0 lifetime in the decay $B^0 \to J/\psi(\mu^+\mu^-)K^0_S(\pi^+\pi^-)$ to be $\tau(B^0) = 1.501^{+0.078}_{-0.074}(\text{stat}) \pm 0.050(\text{syst})$ ps, yielding a lifetime

CERN School of Computing

Well, actually …


Measurement of the Λ_b lifetime in the exclusive decay $\Lambda_b \to J/\psi\,\Lambda$

Distributed analysis

- Physics analysis will take place in many labs all over the world
- Physicists need fast access to event data and corresponding calibration, alignment and bookkeeping data … and to simulated data
- We need the grid!

The LHC Computing Grid

- Global collaboration of more than 140 computing centers in 34 countries
- Four-tiered model
- Data storage and analysis infrastructure
- $O(10^5)$ CPUs
- >25 PByte disk storage (tiers 0 and 1)

Data management

- Dataset bookkeeping
  - Which data exist?
- Dataset locations service
  - Where are the data?
- Data placement and transfer system
  - Tier-0 → Tier-1 → Tier-2
- Data access and storage
  - Long-term storage, direct access

CERN School of Computing

Datasets in CMS

- RAW: Raw data (1-1.5 MB)
  - Detector data, L1 trigger results, HLT results, reconstructed HLT objects
- RECO: Reconstructed data (200-800 kB)
  - Reconstructed objects (hits, clusters, tracks, vertices, muons, electrons, jets)
- AOD: Analysis object data (50-100 kB)
  - High-level reconstructed objects (tracks, vertices, muons, electrons, jets)

Datasets in CMS (ctd)

- TAG: Tagging data (10 kB)
  - Run/event number, some high-level physics objects
- Non-event data
  - Construction data (information on sub-detectors)
  - Equipment management data (detector geometry, electronics)
  - Configuration data (front-end electronics)
  - Conditions data (run conditions, calibration, alignment)

Data flow in CMS

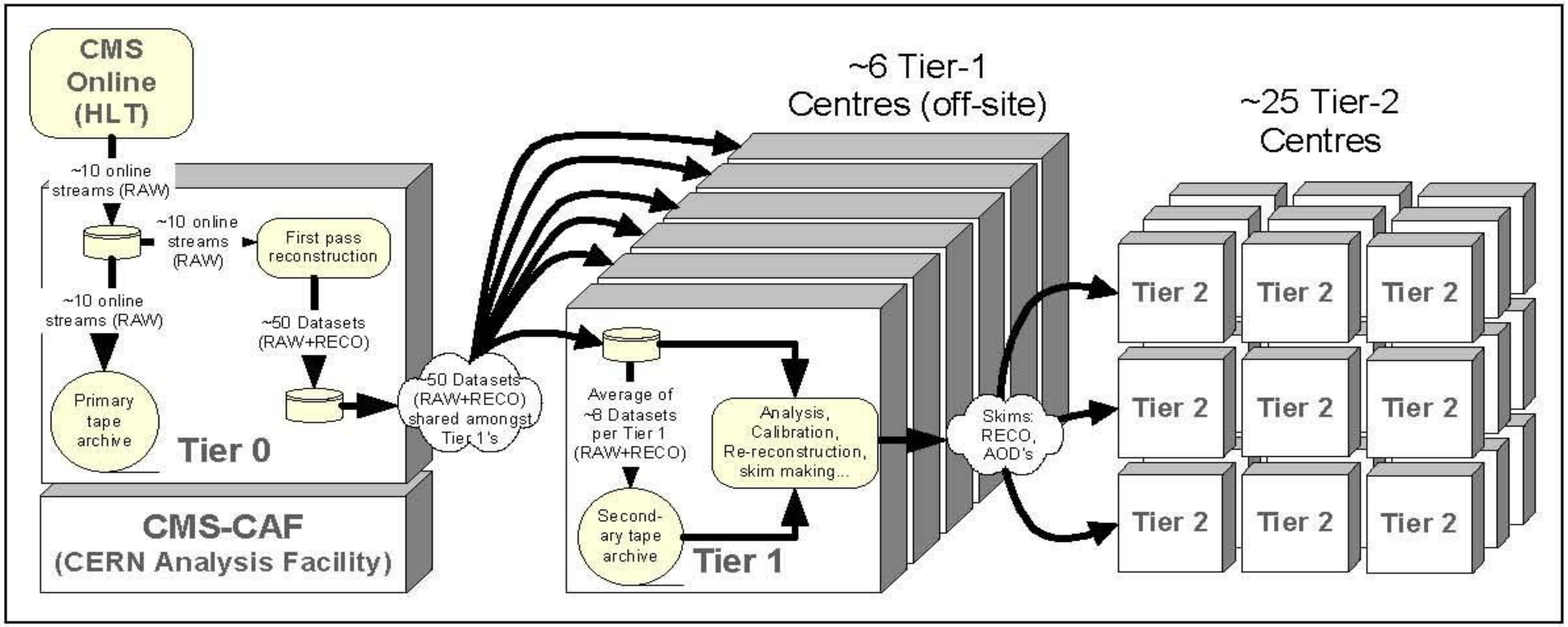


Figure 2.1: Schematic flow of bulk (real) event data in the CMS Computing Model. Not all connections are shown - for example flow of MC data from Tier-2's to Tier-1's or peer-to-peer connections between Tier-1's.

Tiered resources

- Tier-0 (CERN)
  - First pass of reconstruction
  - Primary archive on mass storage
- Tier-1 (~10 centers)
  - Global and local services
  - (Only) copies of certain samples
  - Reconstruction
  - Monte Carlo production

Tiered resources (ctd)

- Tier-2 (~25 centers)
  - Associated to a particular Tier-1 centre
  - Local use
  - CMS controlled use
  - Opportunistic use
- Tier-3 (~100 centers)
  - Coordination with a specific Tier-2 centre
  - Local use
  - No guaranteed support, no guaranteed availability

Additional resources

- CMS-CAF (CERN Analysis Facility)
  - Ready access to RAW and RECO data
  - Short turnaround
  - Operation critical tasks: detector diagnostics, calibrations for HLT, trigger optimization, testing of new trigger algorithms
  - Main repository for software and documentation

Summary

- Physics computing involves:
  - Event filtering with multilevel trigger
  - Persistency of raw data
  - Calibration and alignment
  - Persistency of calibration, alignment and environmental data
  - Event reconstruction
  - Persistency of reconstruction objects and metadata

Summary (ctd)

- Physics computing involves:
  - Simulation of many million events
  - Persistency of simulated raw data and truth information
  - Reconstruction of simulated events
  - Persistency of reconstruction object and truth information
  - Distributed physics analysis and event viewing
  - Persistency of high-level physics objects

Outlook on the track

- ROOT
  - 3 hours of lectures (A. Naumann, B. Bellenot)
  - 3 hours of exercises (A. Naumann, B. Bellenot)
- Data analysis
  - 4 hours of lectures (A. Heikkinen, I. Puljak)
  - 4 hours of exercises (A. Heikkinen, I. Puljak)


The End