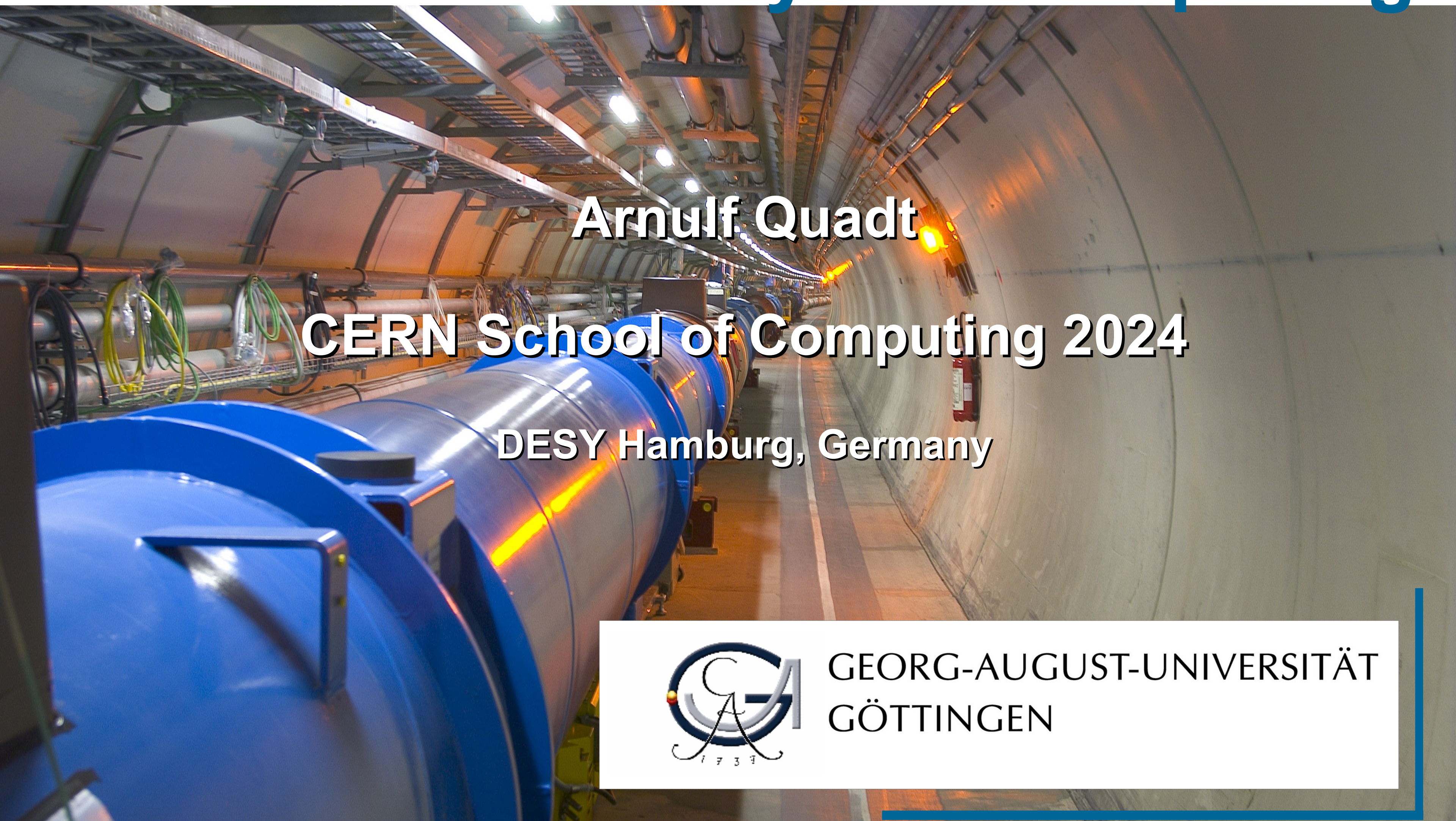
Introduction to Physics Computing



Arnulf Quadt

CERN School of Computing 2024

DESY Hamburg, Germany

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

CERN School of Computing

Outline of the lecture

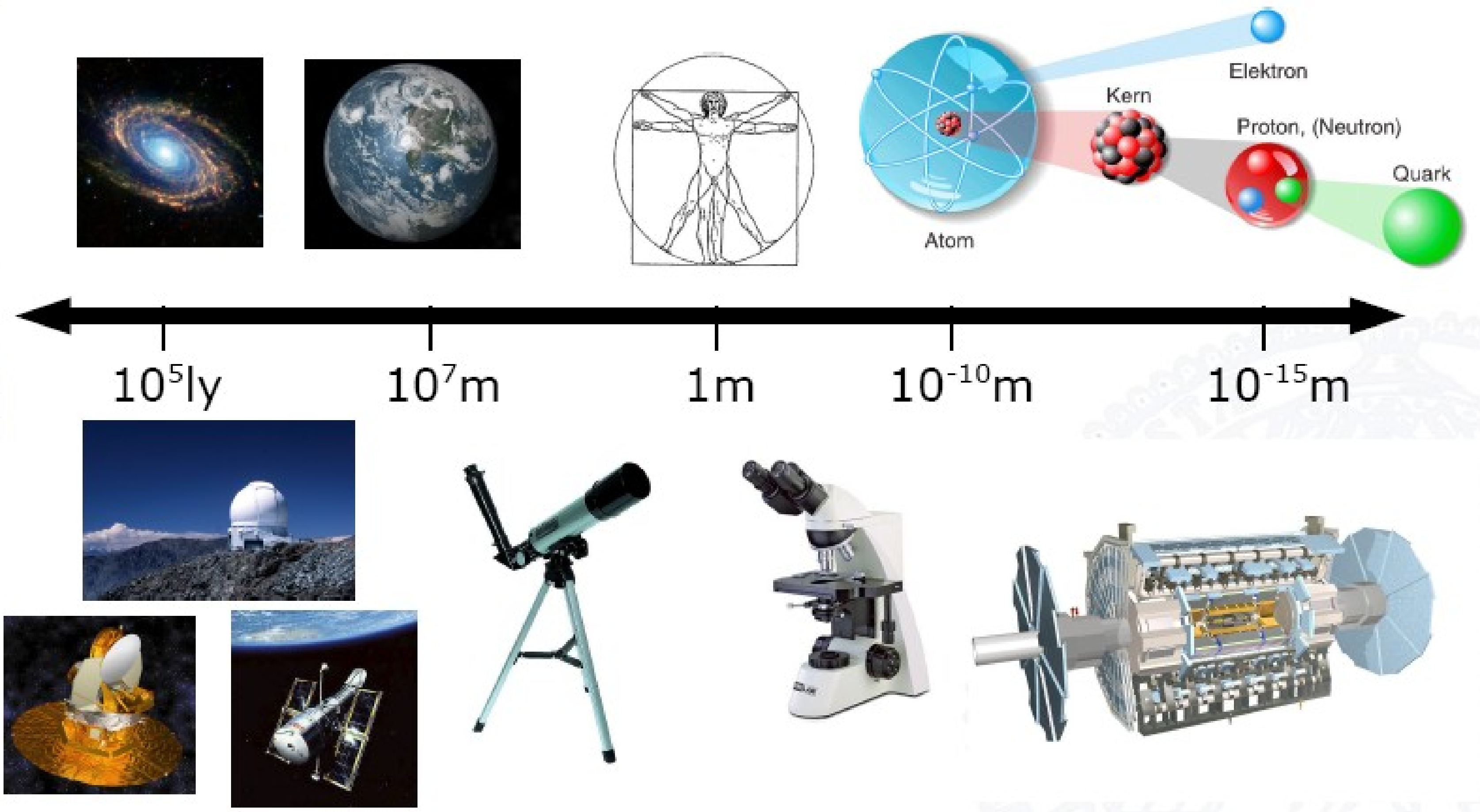
- **Introduction**
- **Various aspects of Physics Computing:**
  - Event Filtering
  - Calibration and alignment
  - Event Reconstruction
  - Event Simulation
  - Physics Analysis
  - Data Flow and Computing Resources

Introduction

CERN School of Computing

Powers of Ten

- **Goal: understand fundamental structures and forces**



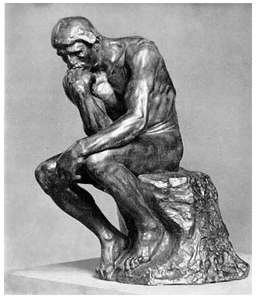
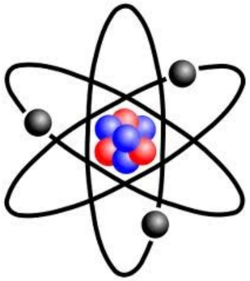
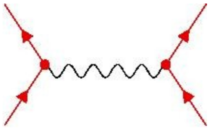
Fundamental structures & forces

- From largest to smallest dimensions
- Reduction principle

→ few fundamental building blocks

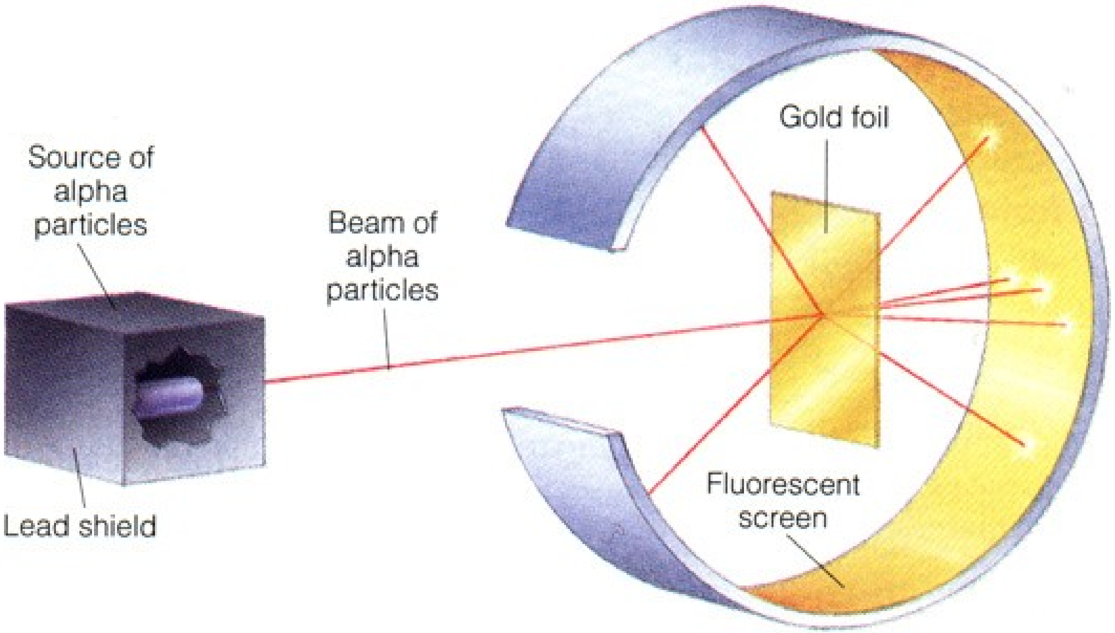
→ few fundamental forces


... erforschen was die Welt im Innersten zusammenhält ...

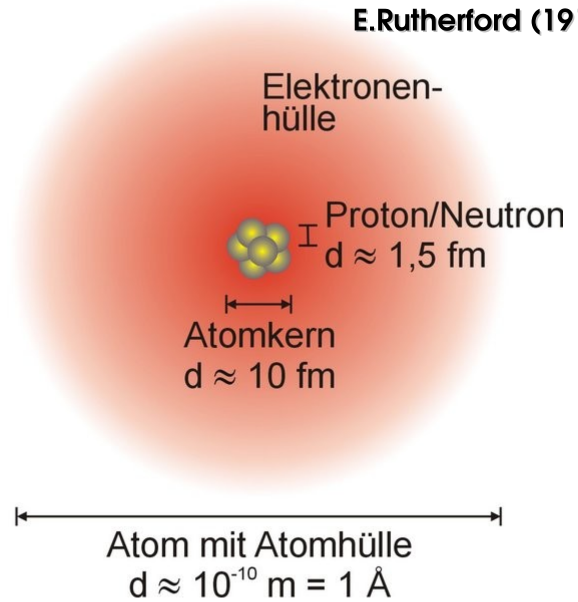




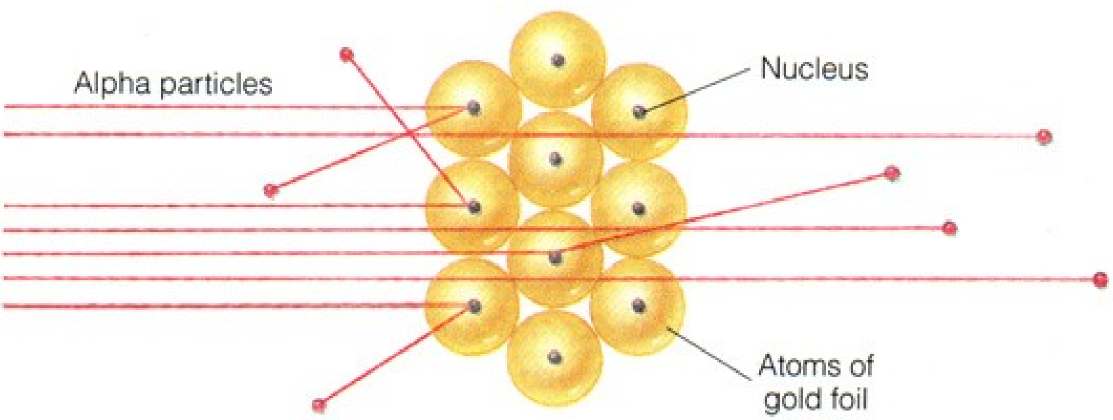
CERN School of Computing

Rutherford scattering



E.Rutherford (1912)

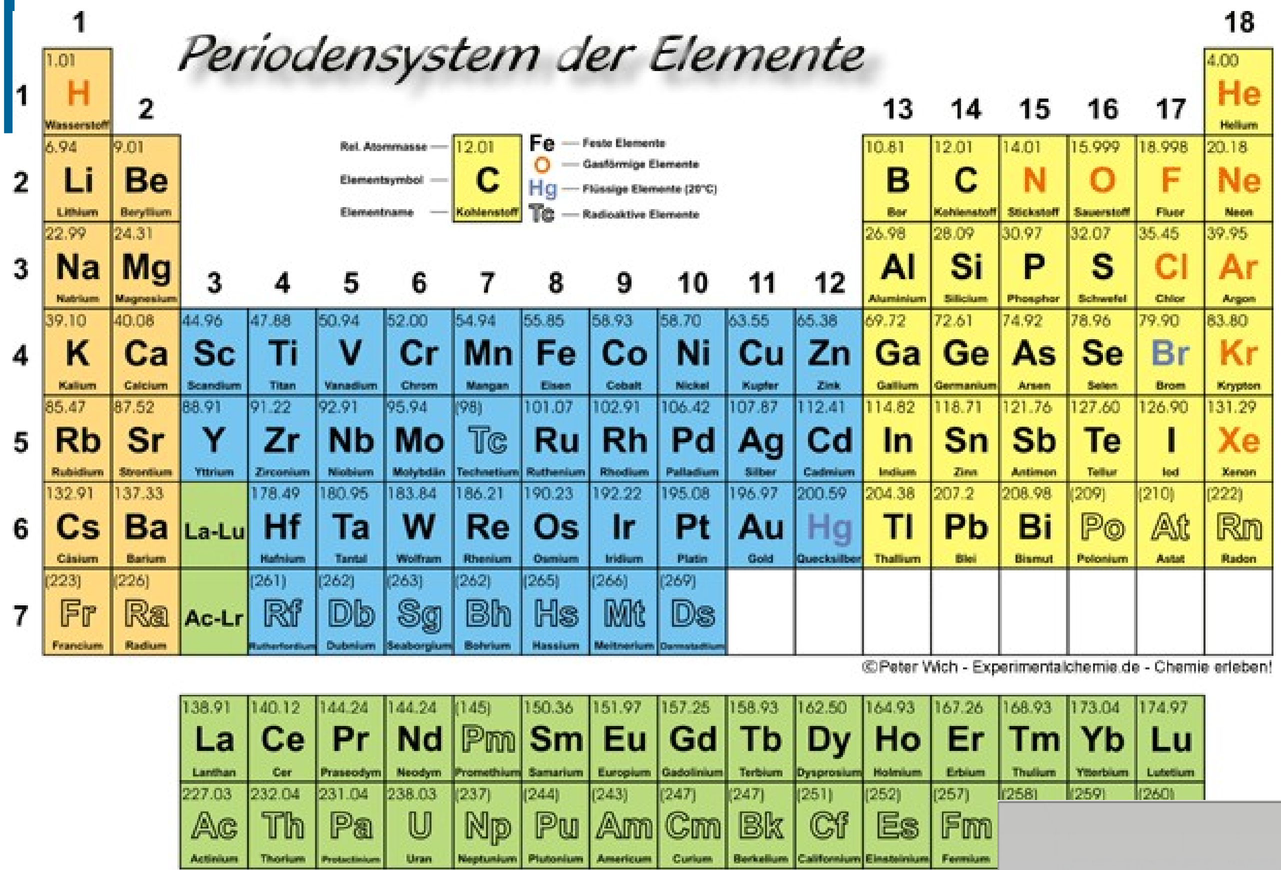


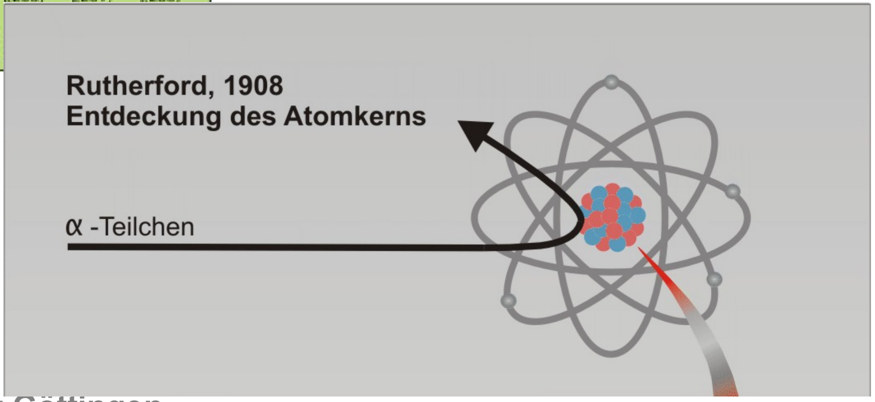


- **Need particles source & detectors**

CERN School of Computing

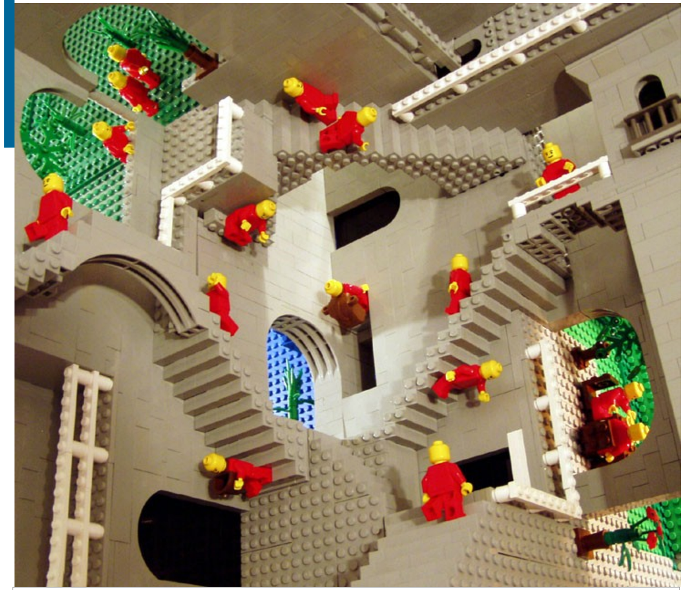
(sub) structure - atoms

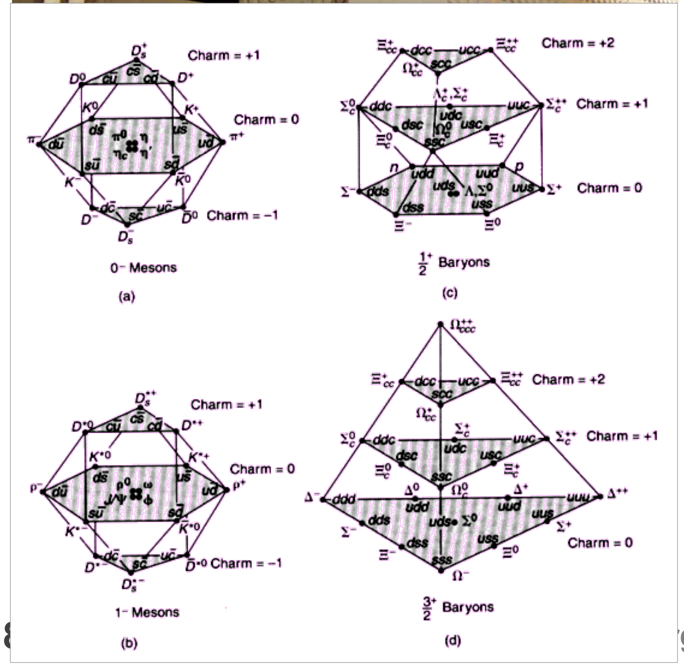




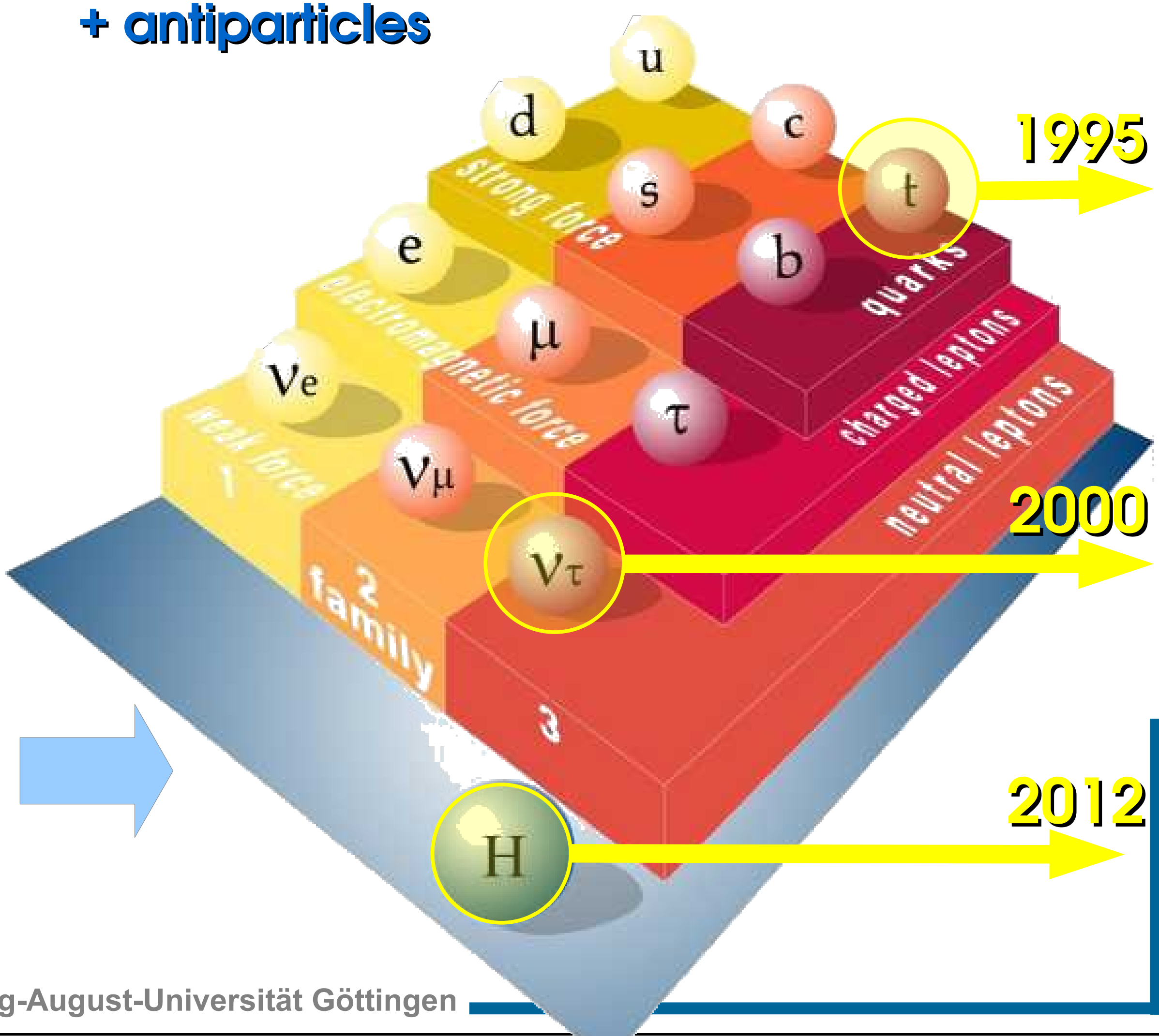
CERN School of Computing

Elementary building blocks



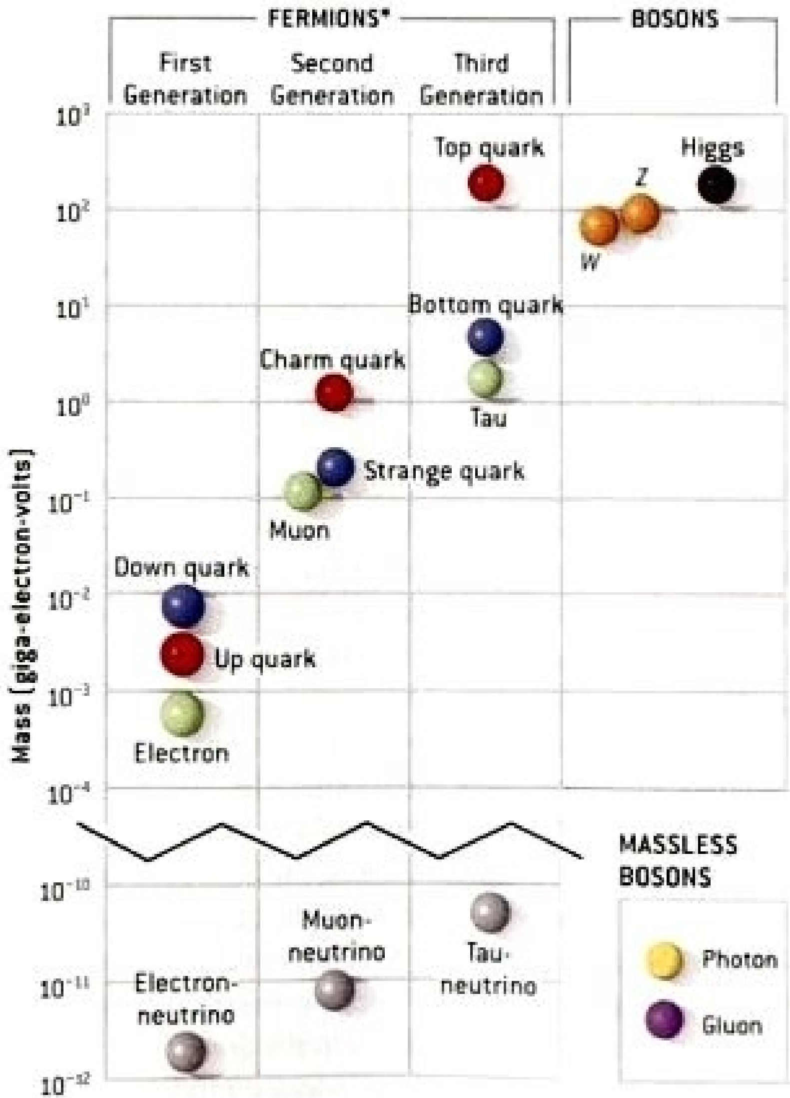


`LEGO bricks' of
Spin-1/2 quarks & leptons
+ antiparticles



CERN School of Computing

Elementary building blocks (ctd)



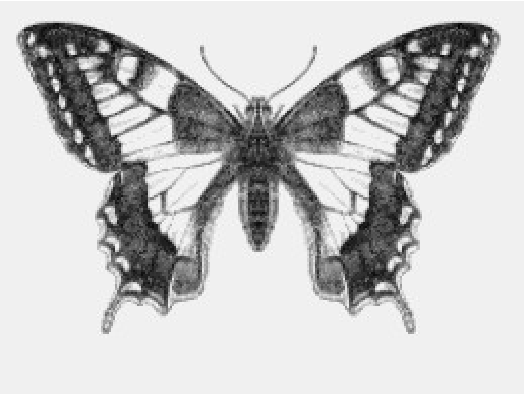
Open questions:

- Properties and role of fermions?
- Really only three generations?
- Relation between leptons & quarks?
- Mass and role of neutrinos?
- Origin of mass and hierarchy?
- Is there only ONE Higgs?
- Quark mixing and CP-violation

Physics Nobel Prize 2008

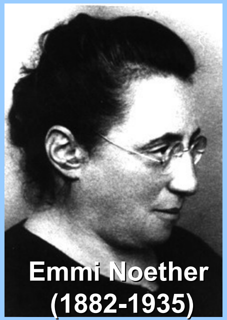
- What is dark matter?
- ...

Symmetries in Nature



Natur is full of symmetries ⇨ simple description

symmetry ⇔ conservation law



physics laws independent of

origin of time axis ⇨ conservation of energy

origin of space axis ⇨ conservation of momentum

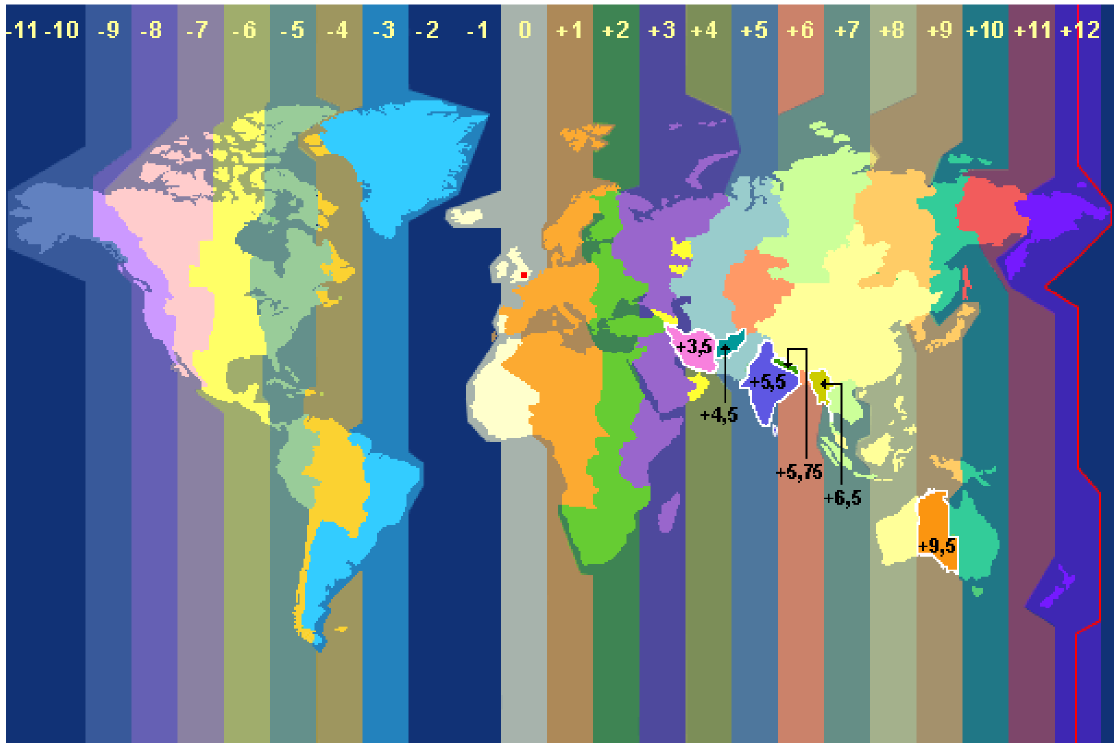
direction of space axis ⇨ conservation of angular momentum

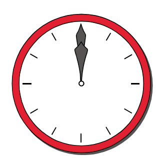
Symmetry breaking ⇨ new physical phenomena

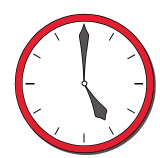
CERN School of Computing

Example: gauge symmetries

local (time) gauge symmetry: time zones





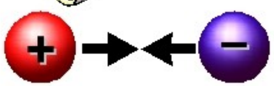


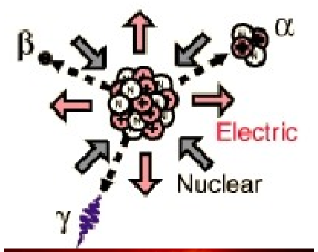
... physics laws (and every day live) do not depend on choice of time = zero ...

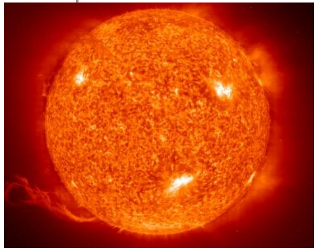
CERN School of Computing

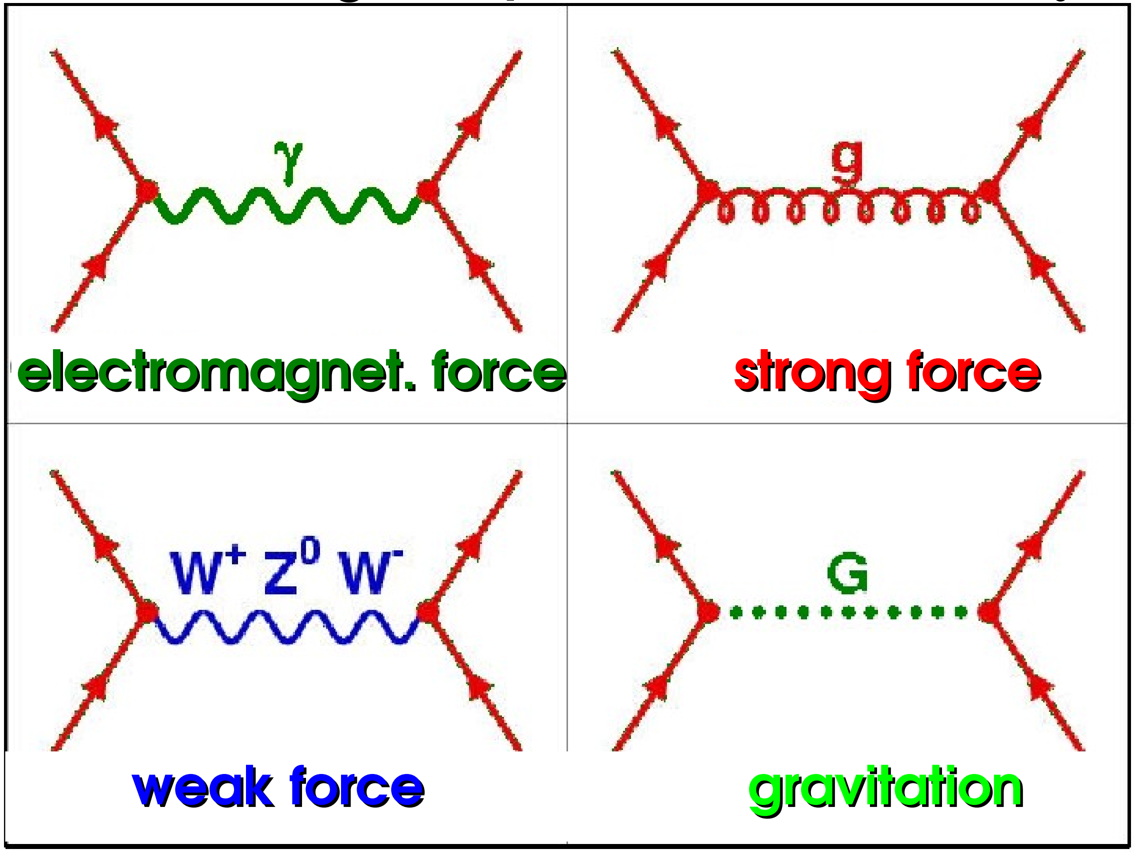
Fundamental interactions

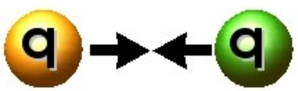
- Quantum field theory, local gauge symmetry
- Interaction between spin-1/2 fermions
  via exchange of spin-1 vector bosons (γ,W,Z,g)





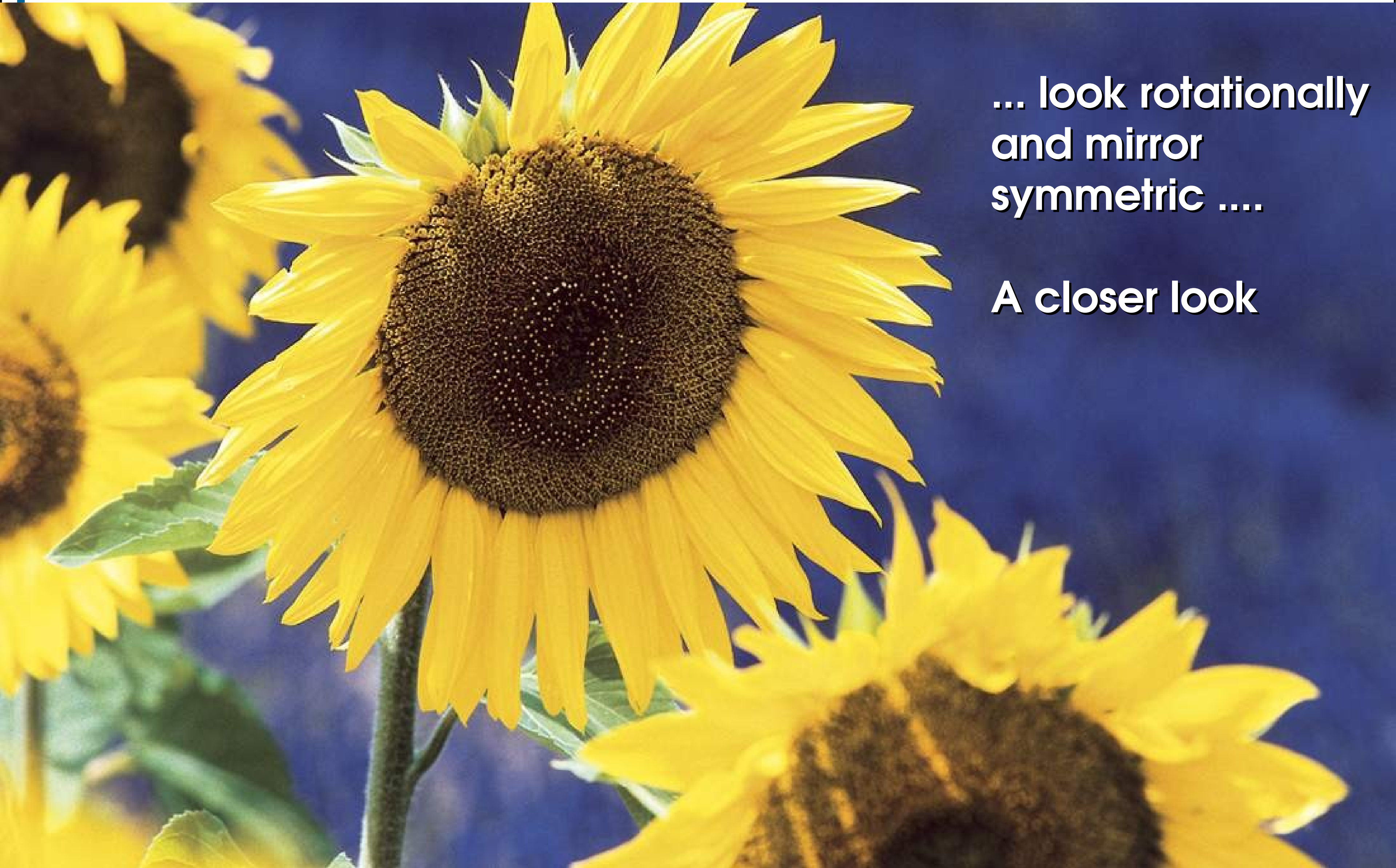






CERN School of Computing

Sun flowers

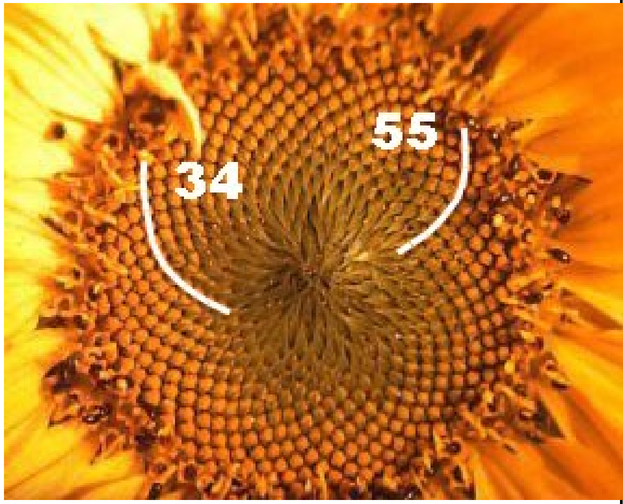


34 spirals clock-w, 21 spirals counter clock-wise !

Examples for symmetry breaking

**13 spirals cw,
8 spirals
counter-cw!**

**21 spirals cw,
13 spirals
counter cw!**

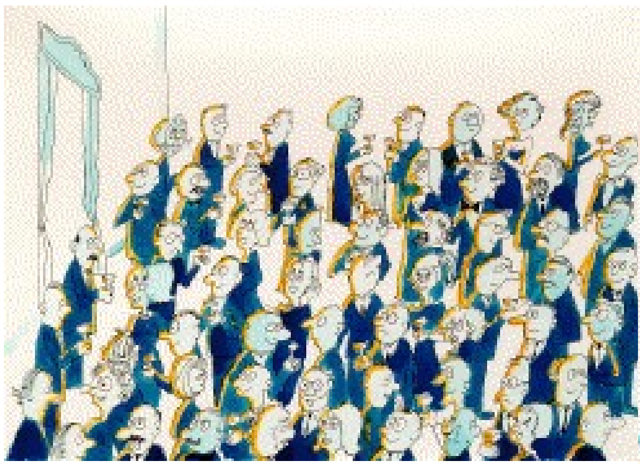


**34 spirals cw,
21 spirals
counter-cw!**

Particle masses – Higgs mechanics

Initially in Standard Model ⇨ massless particles
BUT we know that they have mass !?

One possible explanation: The Higgs Mechanism
(electroweak symmetry breaking)



Higgs fields fills space

... a particle in
Higgs field ...

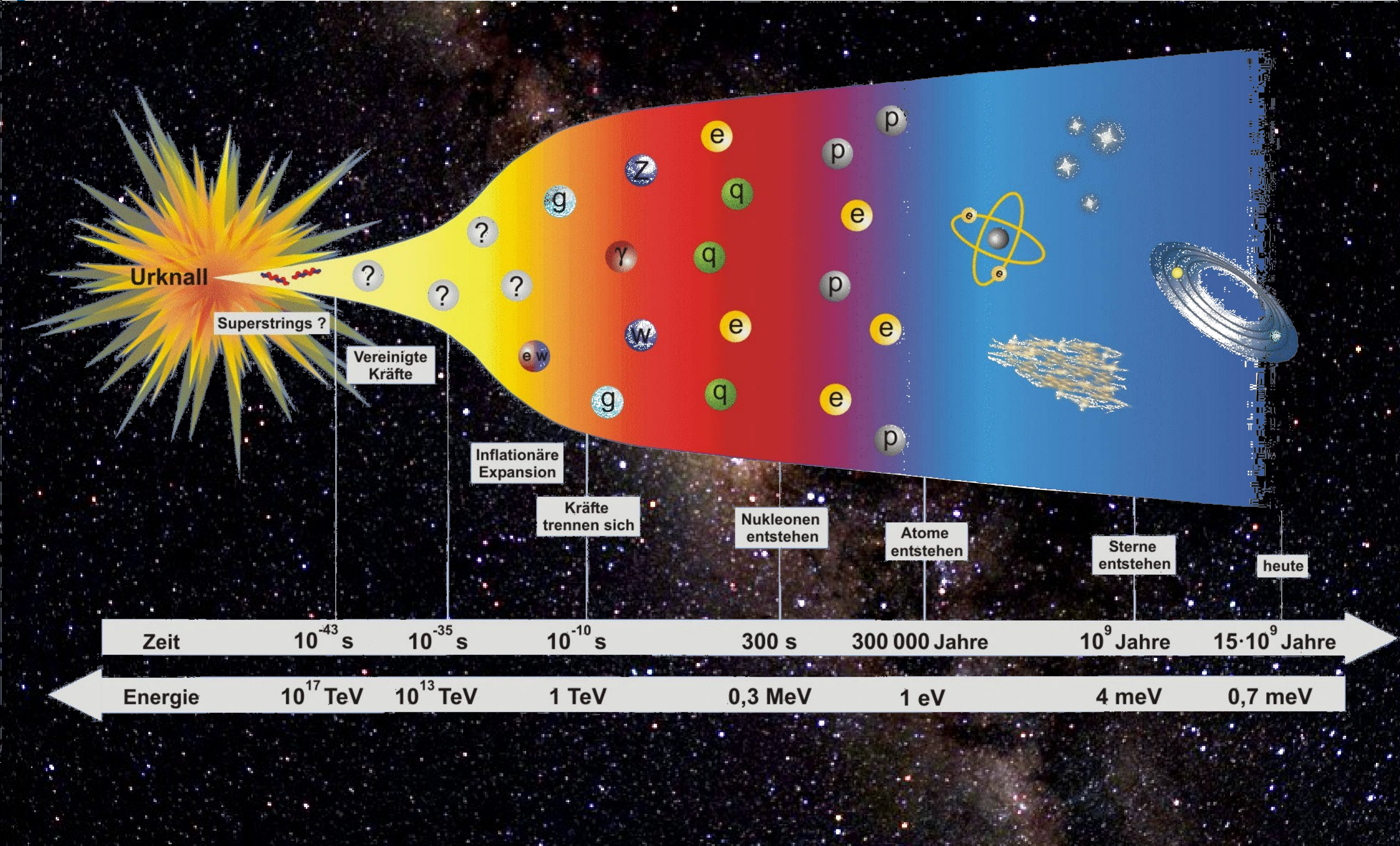
... couples to field ...
inertia = mass

Particle masses – Higgs mechanism

Excitation of
Higgs field

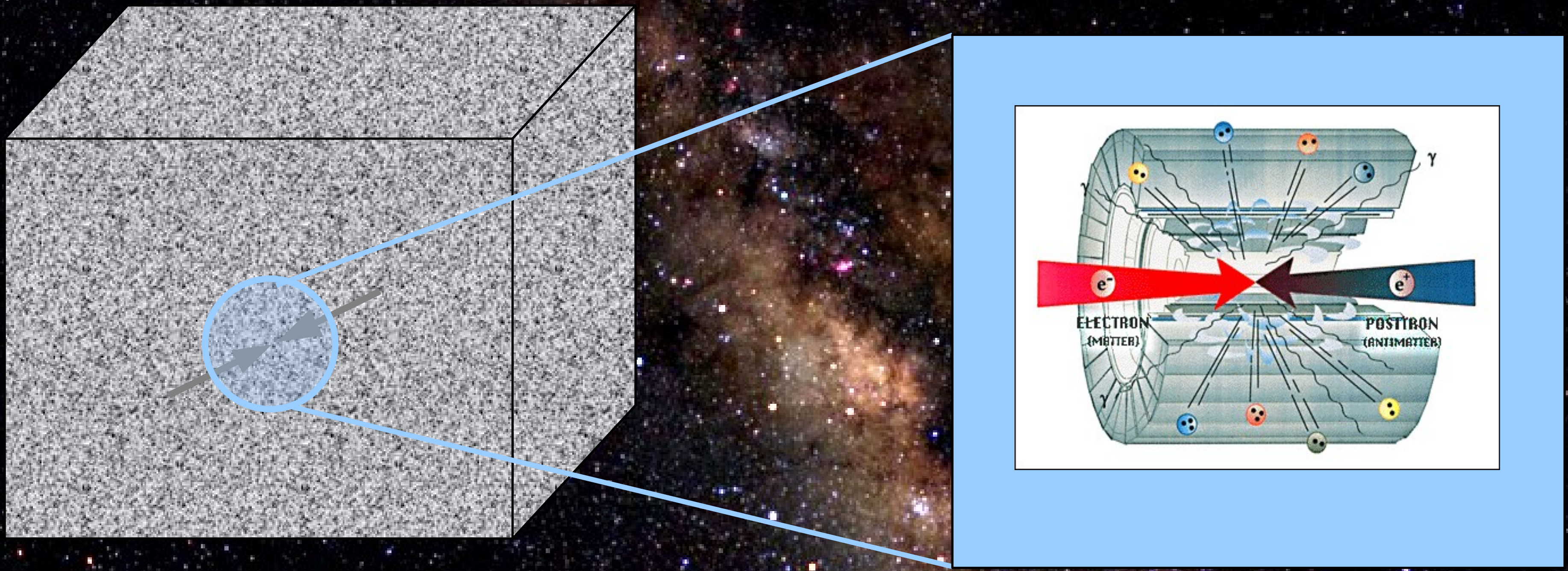
Excited Higgs field
≙ massive Higgs-boson

History of the universe



Big bang in laboratory

Matter-antimatter collisions
At high energy (=temperature)

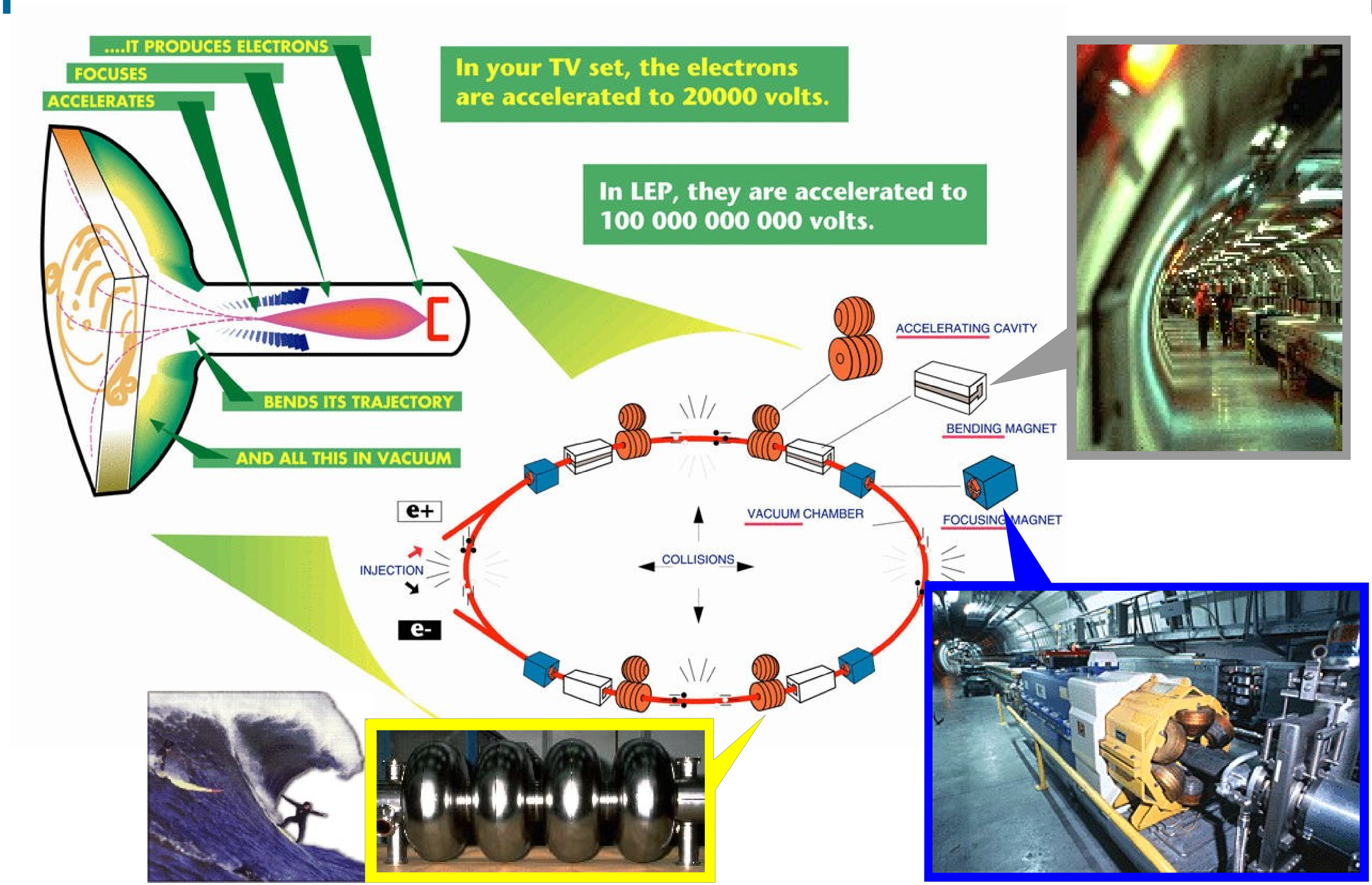


All particles have high energy (temperature) and collide uncontrolled

Individual collisions controlled, selected and recorded

Particle Accelerators

Particle Accelerators I



CERN School of Computing

Particle Accelerators II

studies of
small structures or
Heavy particles

requires

Particles accelerated
to high energies

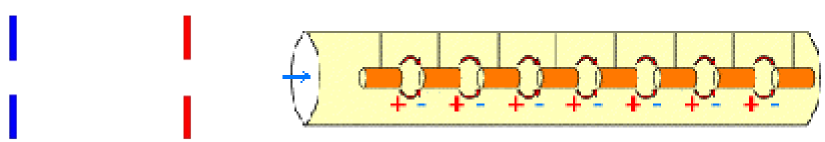
Heisenbergs
uncertainty relation

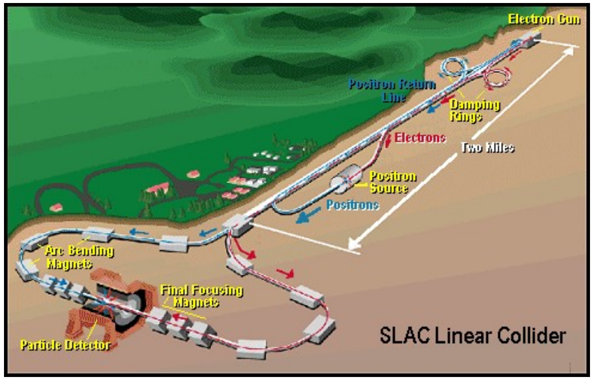
$\Delta x \cdot \Delta p \geq \hbar$

$E = m c^2$

Einstein

linear accelerator:



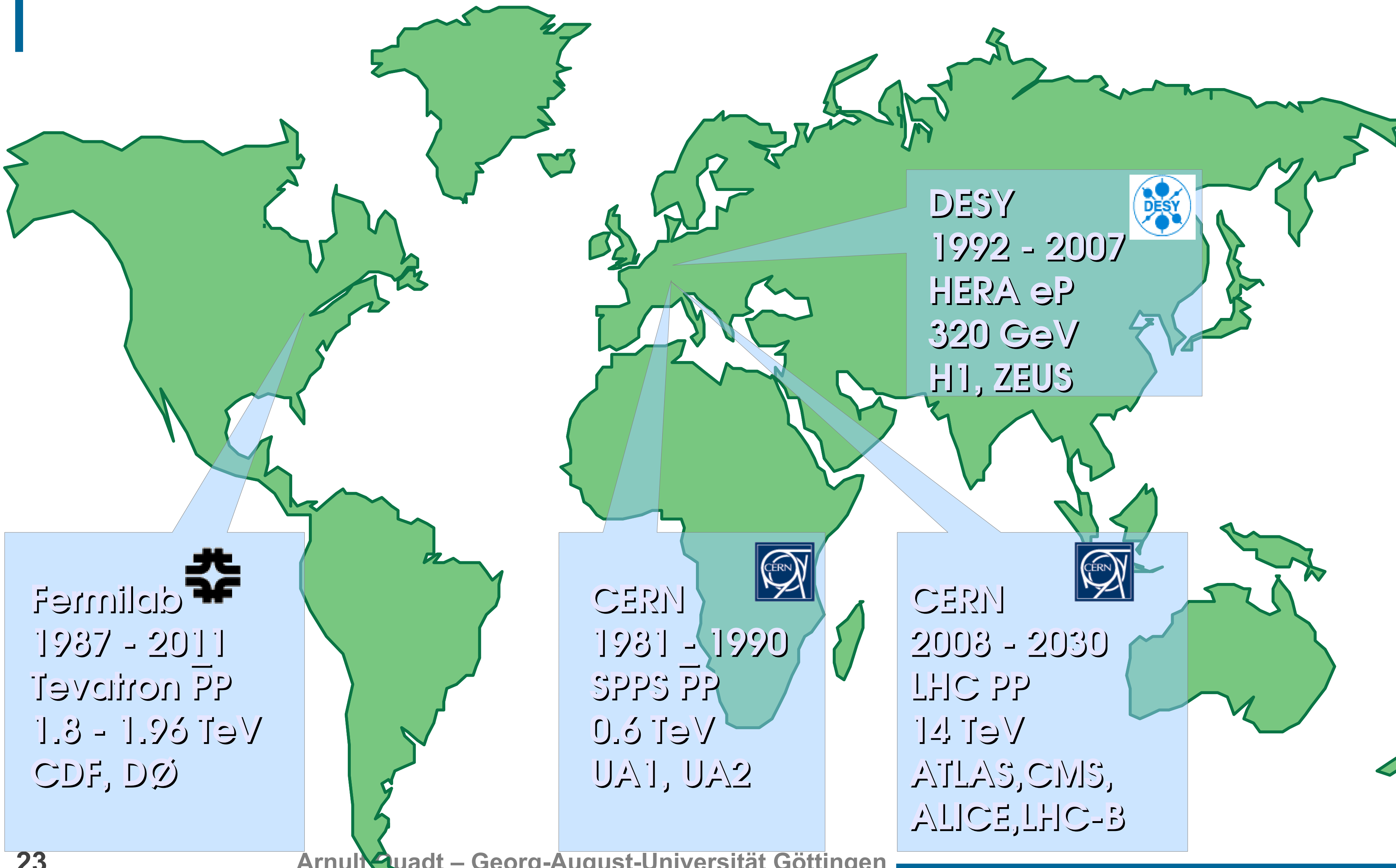


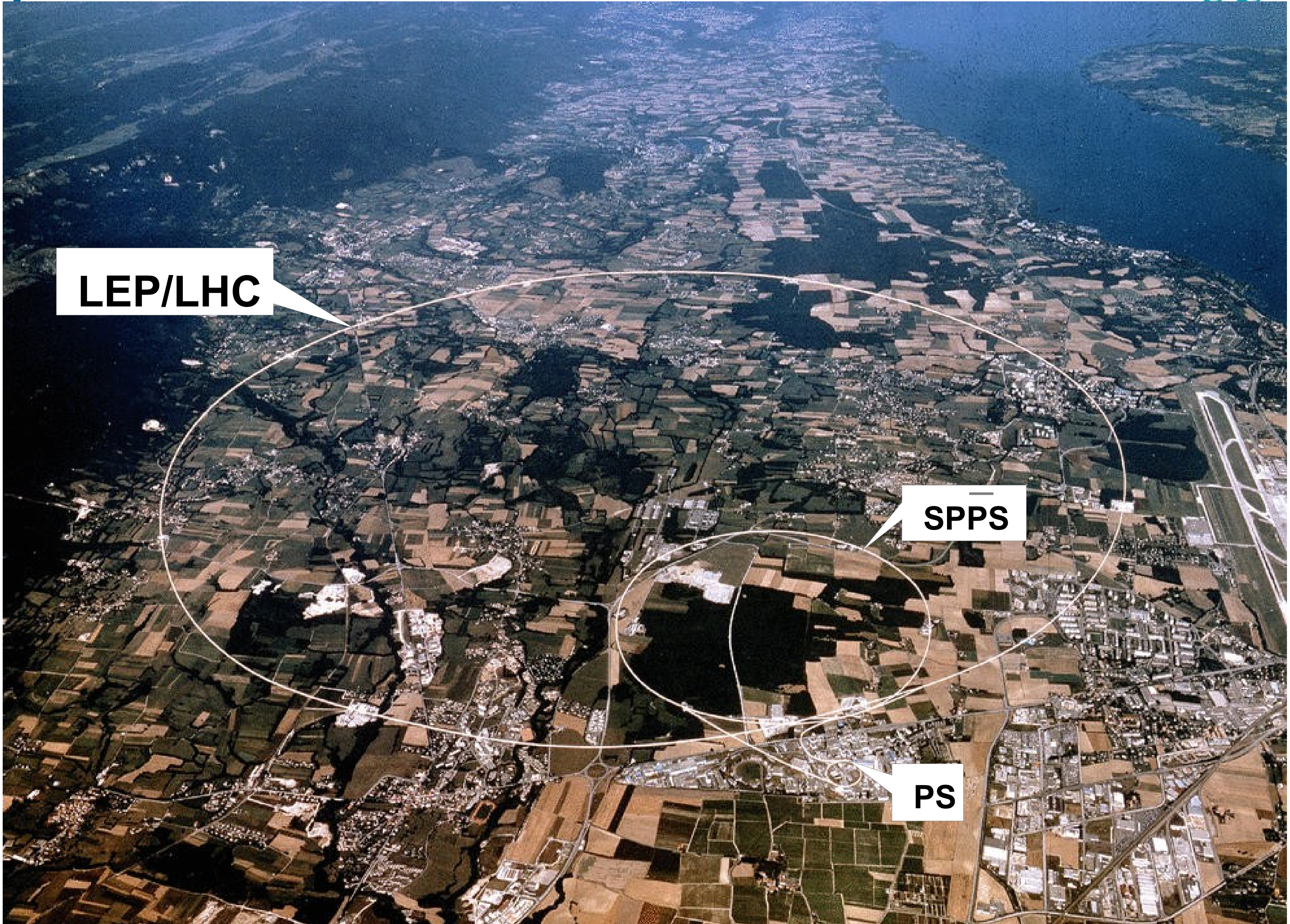
repeated acceleration : 1, 2, ... 1 000 eV =	10^3 eV =	1 keV
1 000 000 eV =	10^6 eV =	1 MeV
1 000 000 000 eV =	10^9 eV =	1 GeV
1 000 000 000 000 eV =	10^{12} eV =	1 TeV

`Tevatron' = TeV beam energy

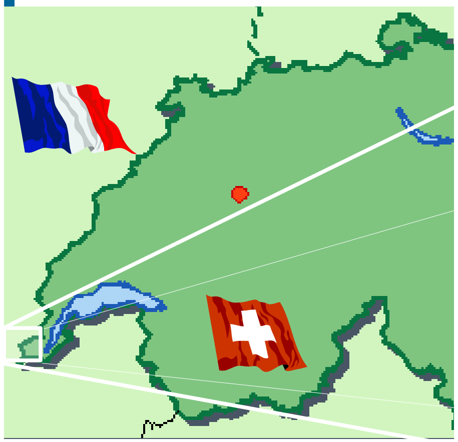
CERN School of Computing

Particle Accelerators III

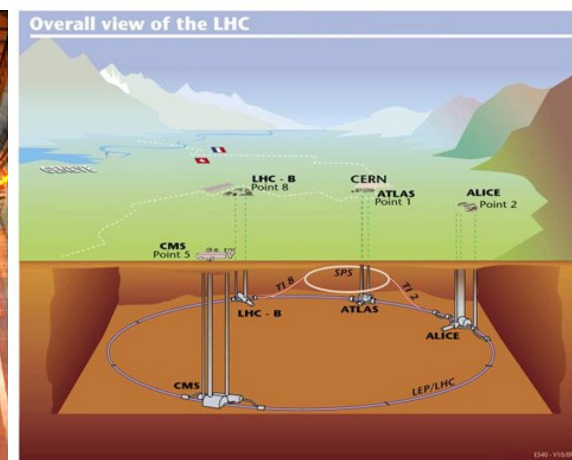



LEP/LHC
SPPS
PS

Large Hadron Collider - LHC



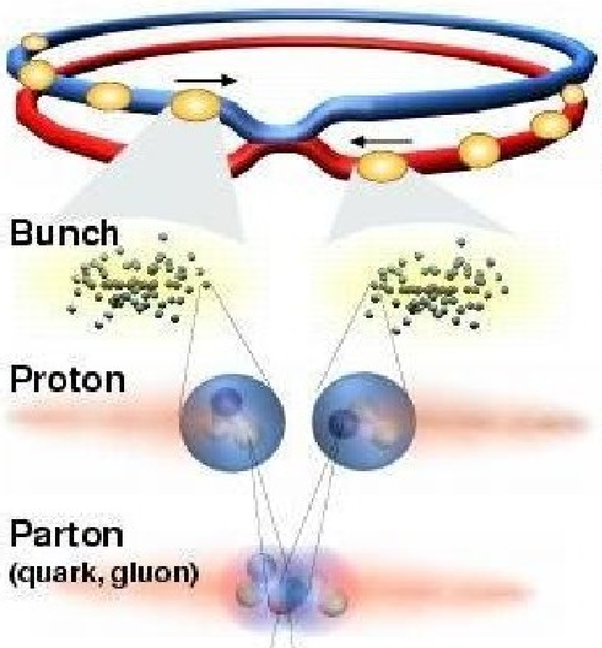
- CERN: europ.center for particle physics
- Founded 1954
- LHC: PP collider
- High energies: $\sqrt{s}$ = 7 (14) TeV
- 40 Mio. collisions / sec
- 1st beam: 10.Sept. 2008
- 1st collisions in Nov. 2009
- physics at 7 TeV since 31.3.2010
- Phys. at 13 TeV since 20.5.2015
- Phys. at 13.6 TeV since July 2022
- 4 Expts:
  ATLAS, CMS, ALICE, LHC-B



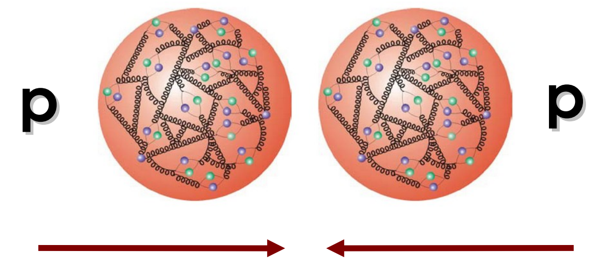
Detectors / Experiments

CERN School of Computing

Proton-Proton Collisions

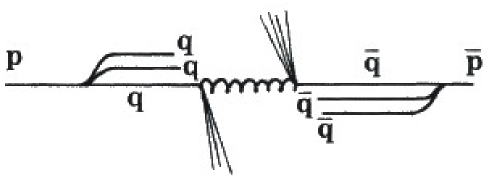


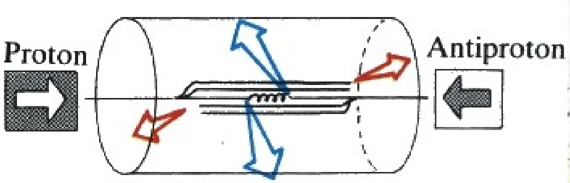
Large Hadron Collider @ CERN, Switzerland



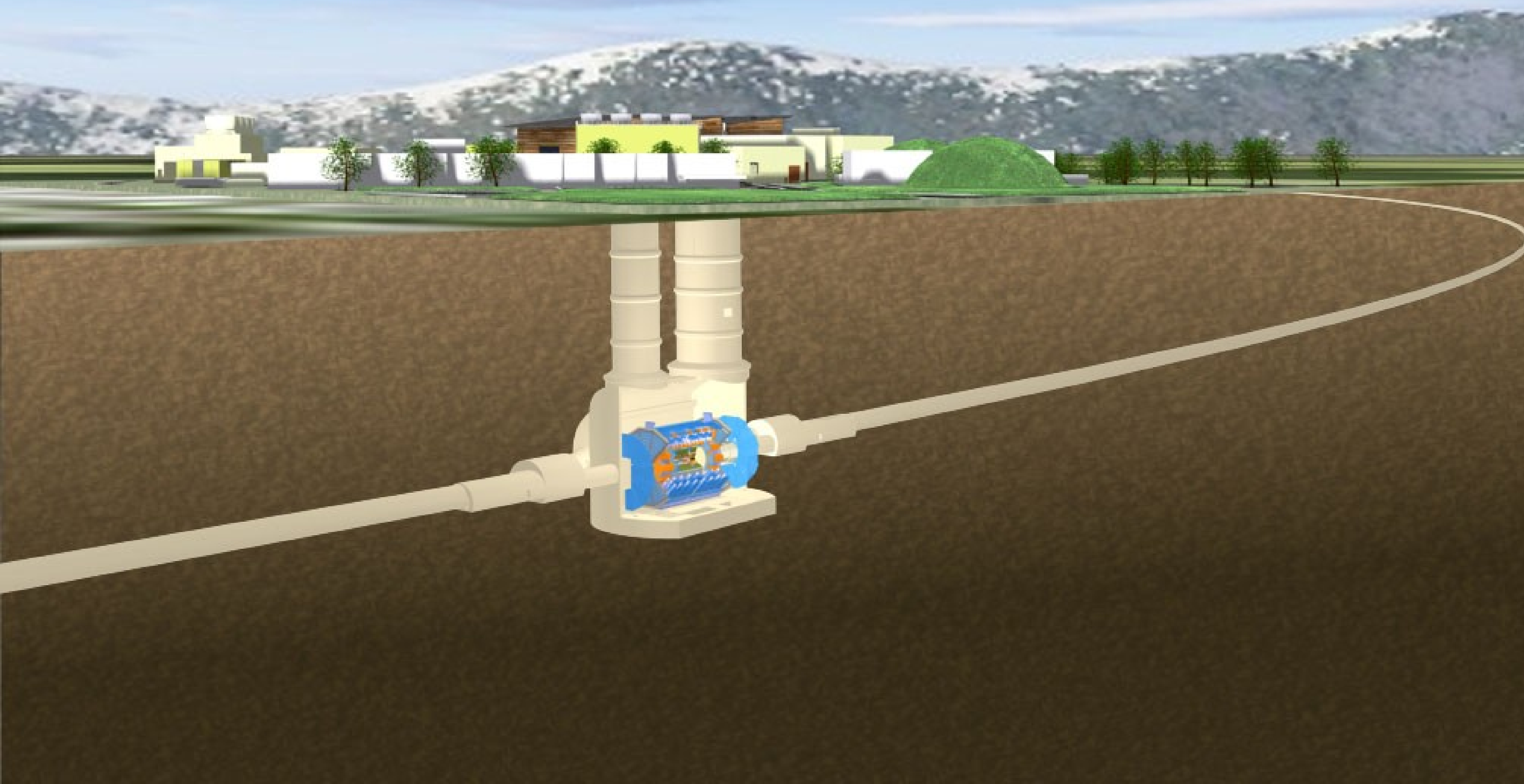
pp collider at $\sqrt{s}$ = 14 TeV

Luminosity L = 10^{34} cm^{-2}s^{-1}

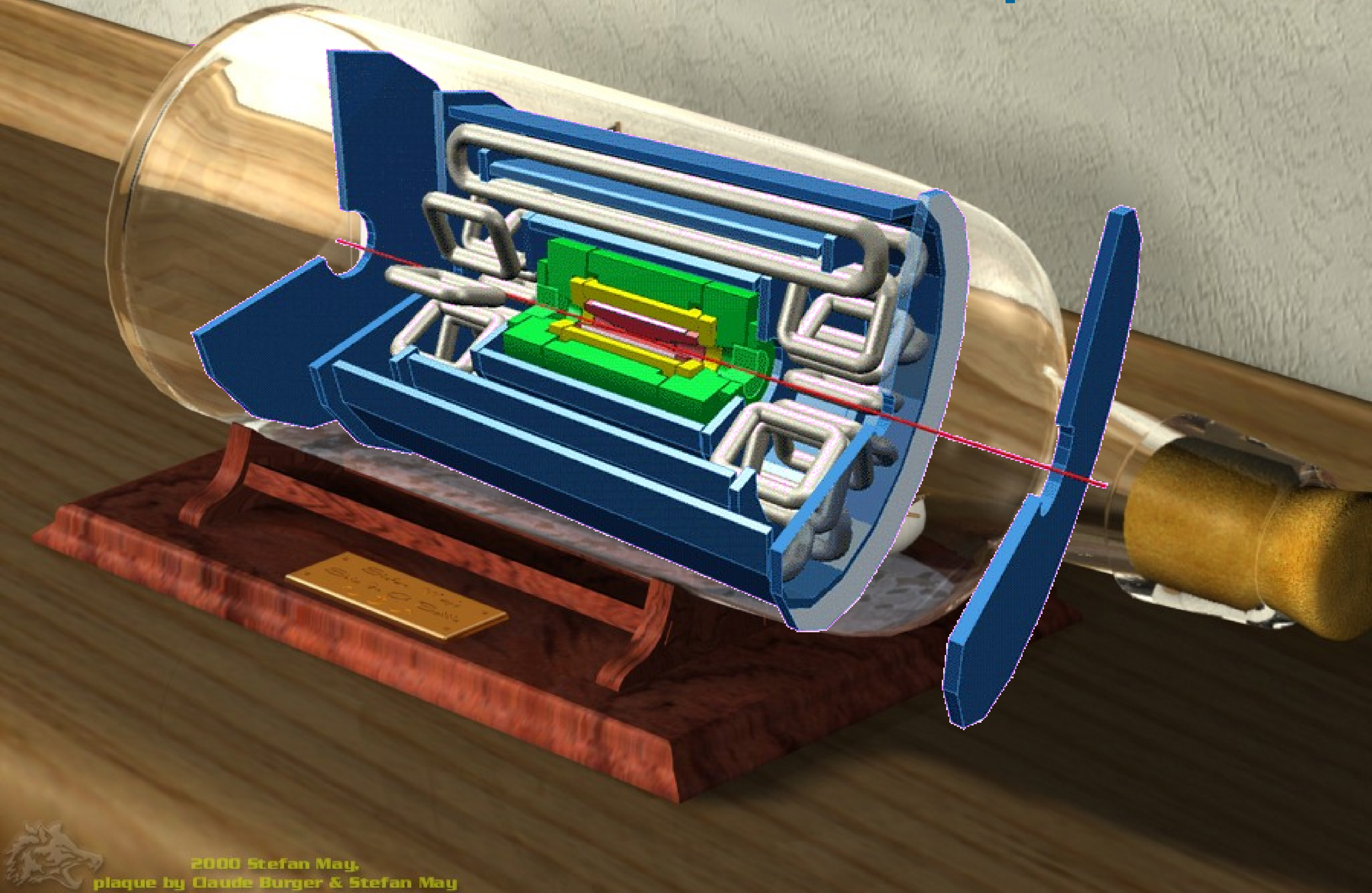




Caverns 100m underground

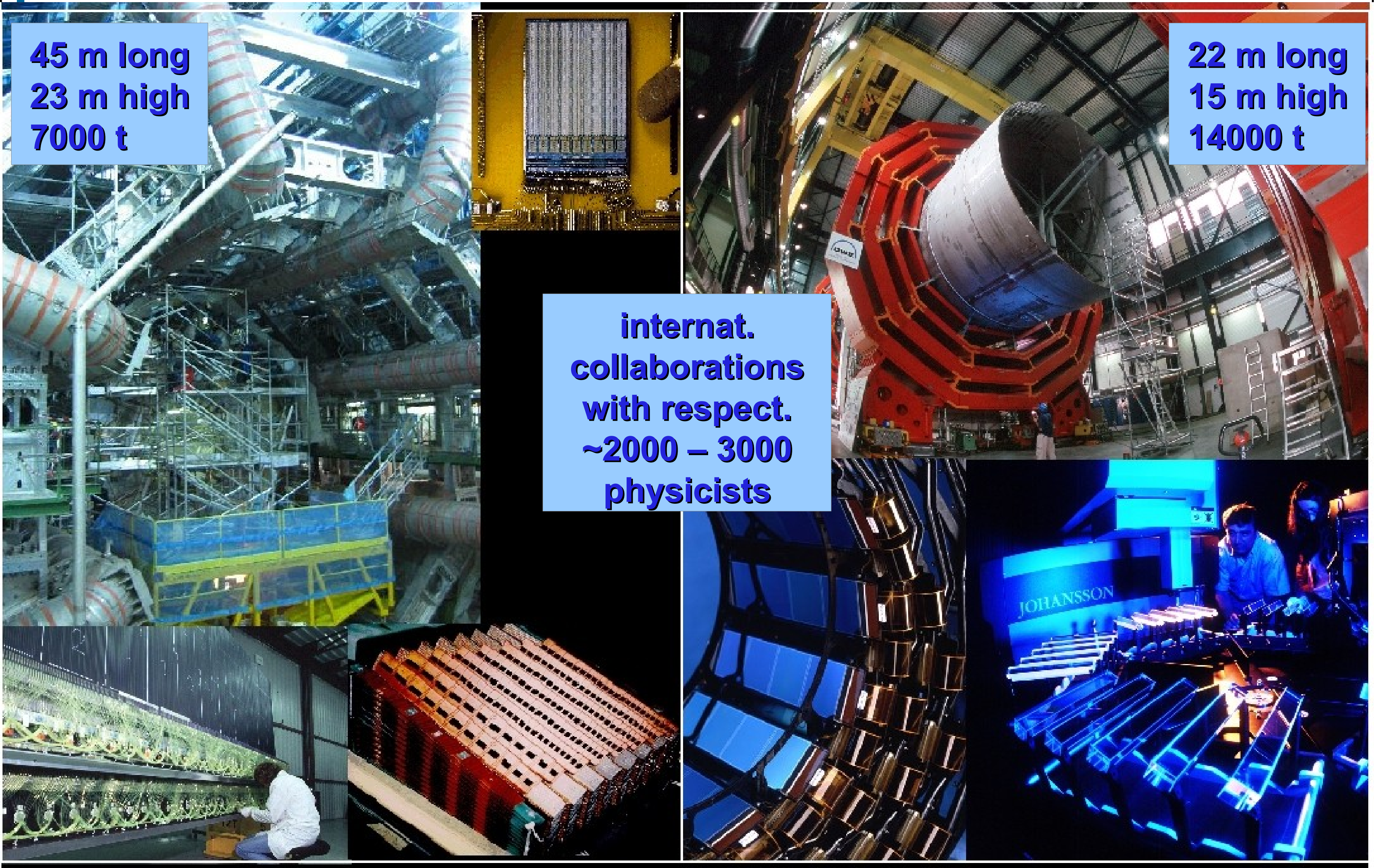


Built like a bottle ship



CERN School of Computing

ATLAS and CMS experiments

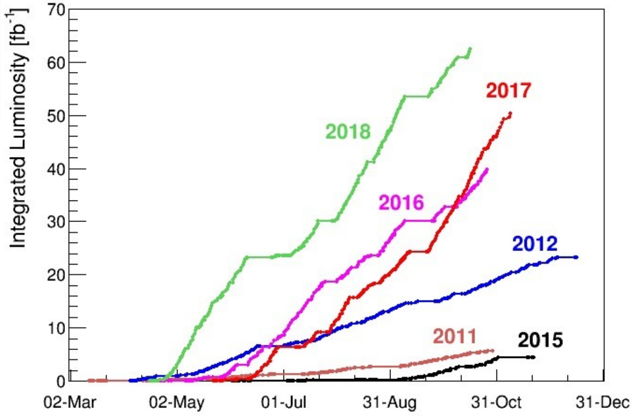


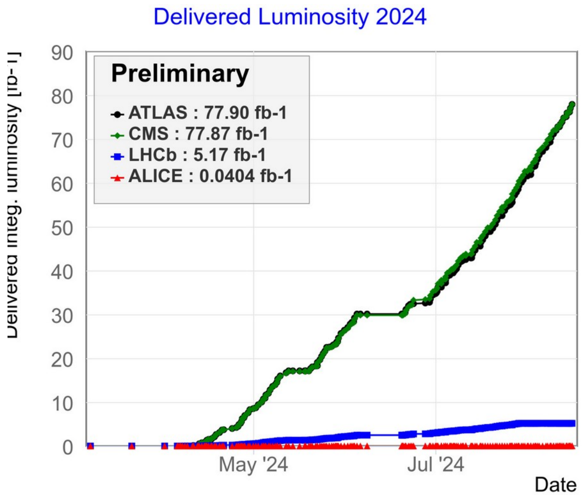
45 m long
23 m high
7000 t

internat.
collaborations
with respect.
~2000 – 3000
physicists

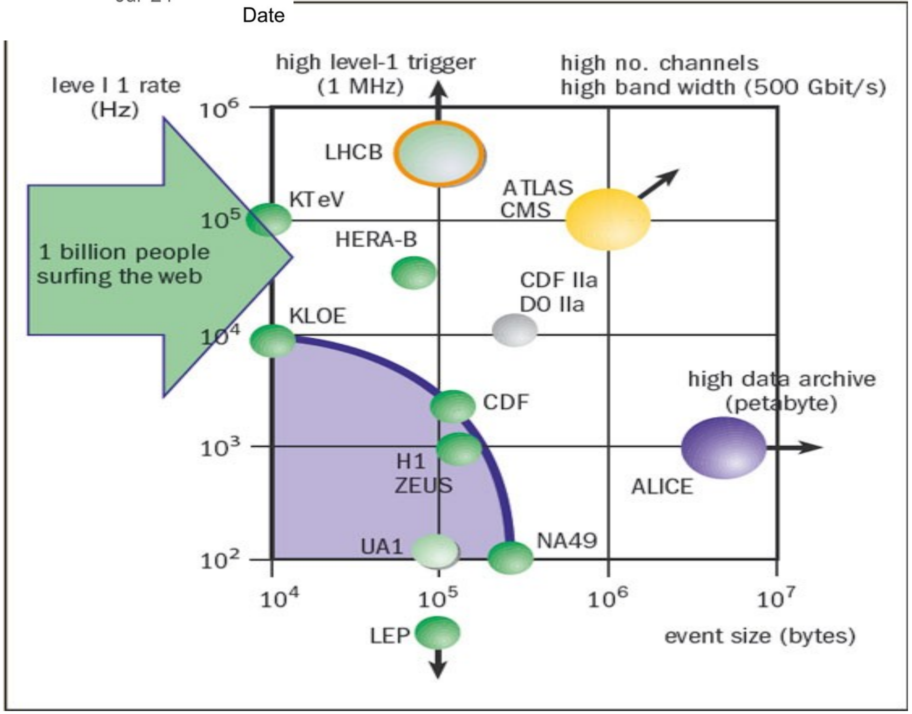
22 m long
15 m high
14000 t

Date rate and size



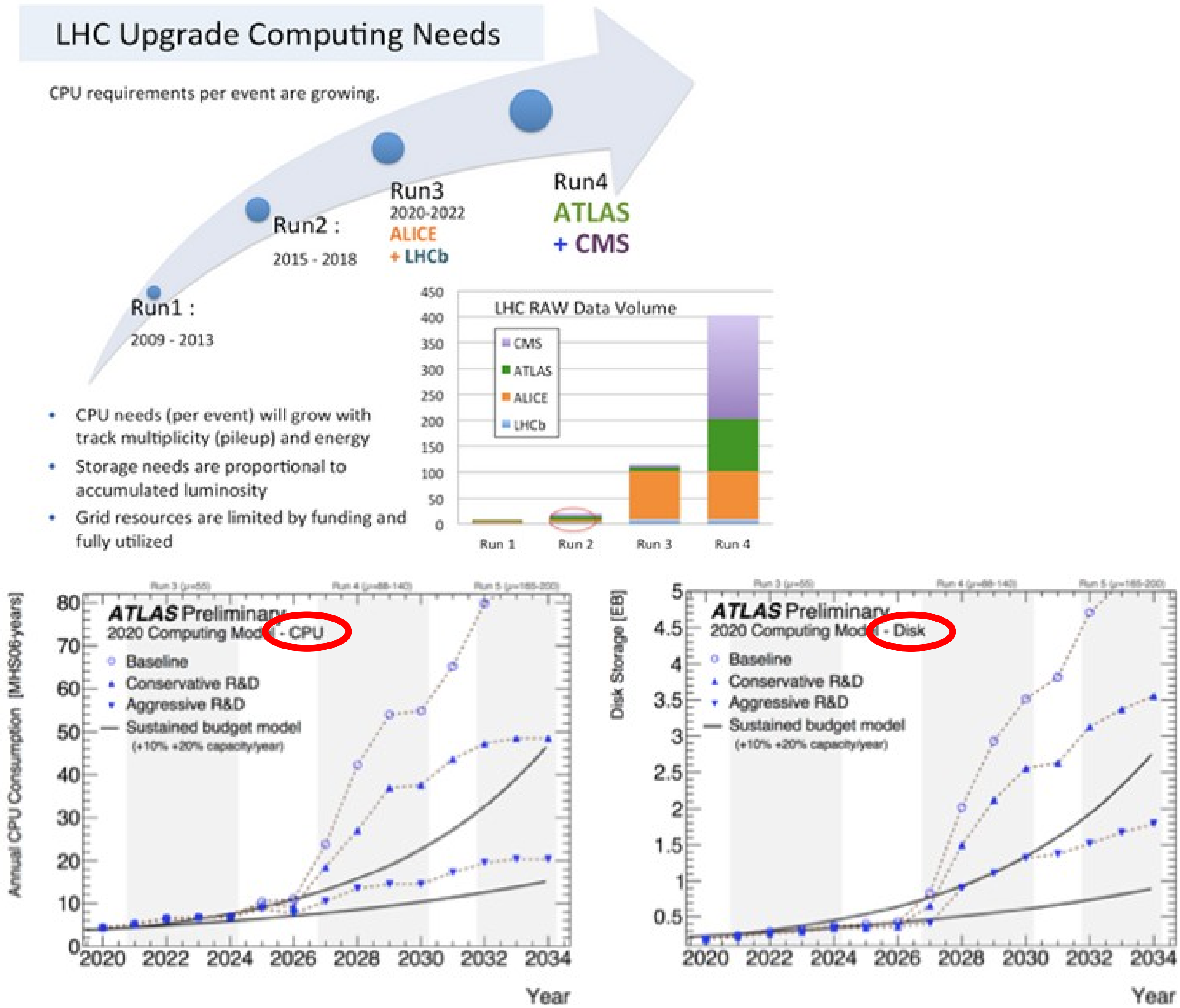


data at LHC:
15 PetaByte /year

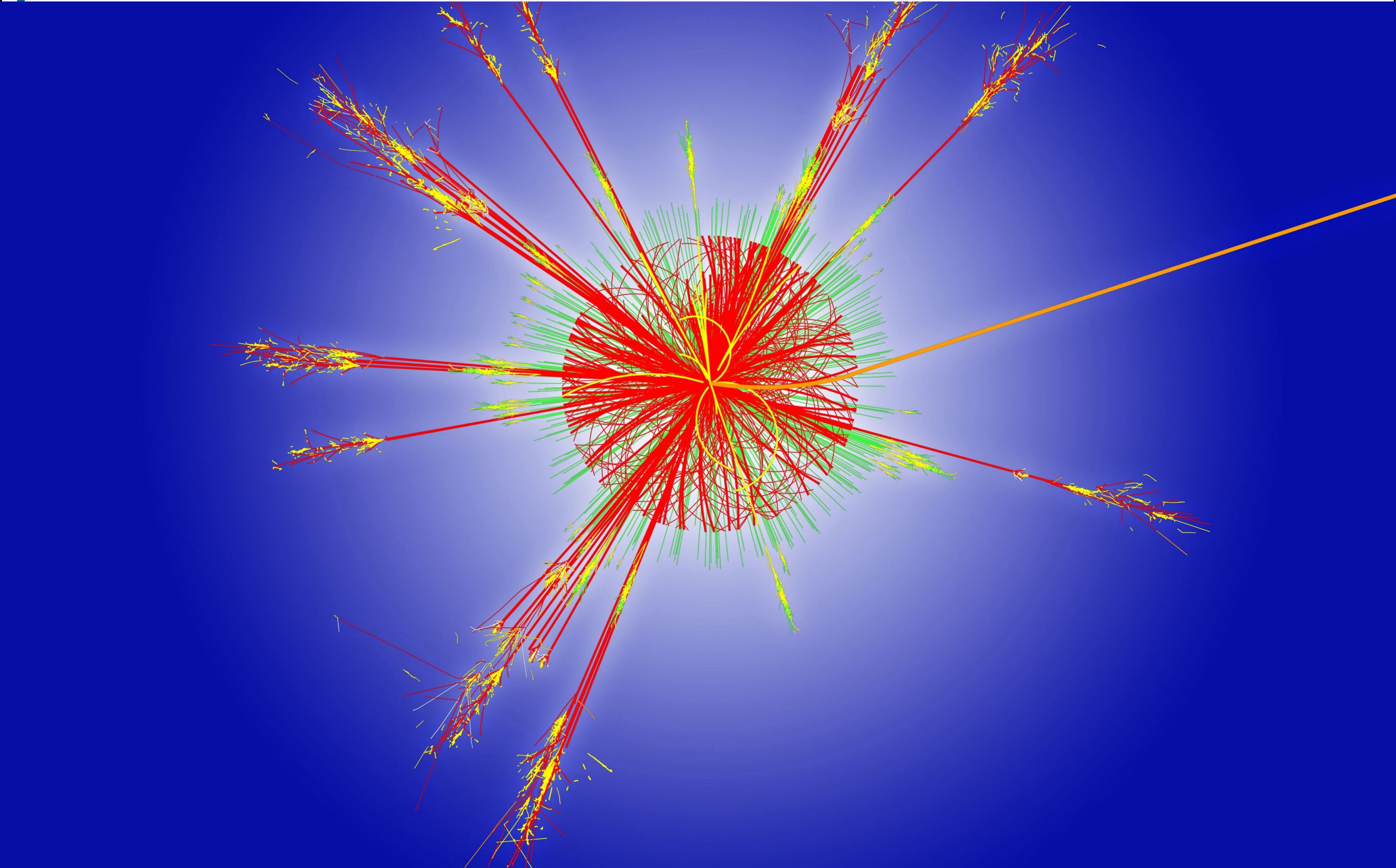


CERN School of Computing

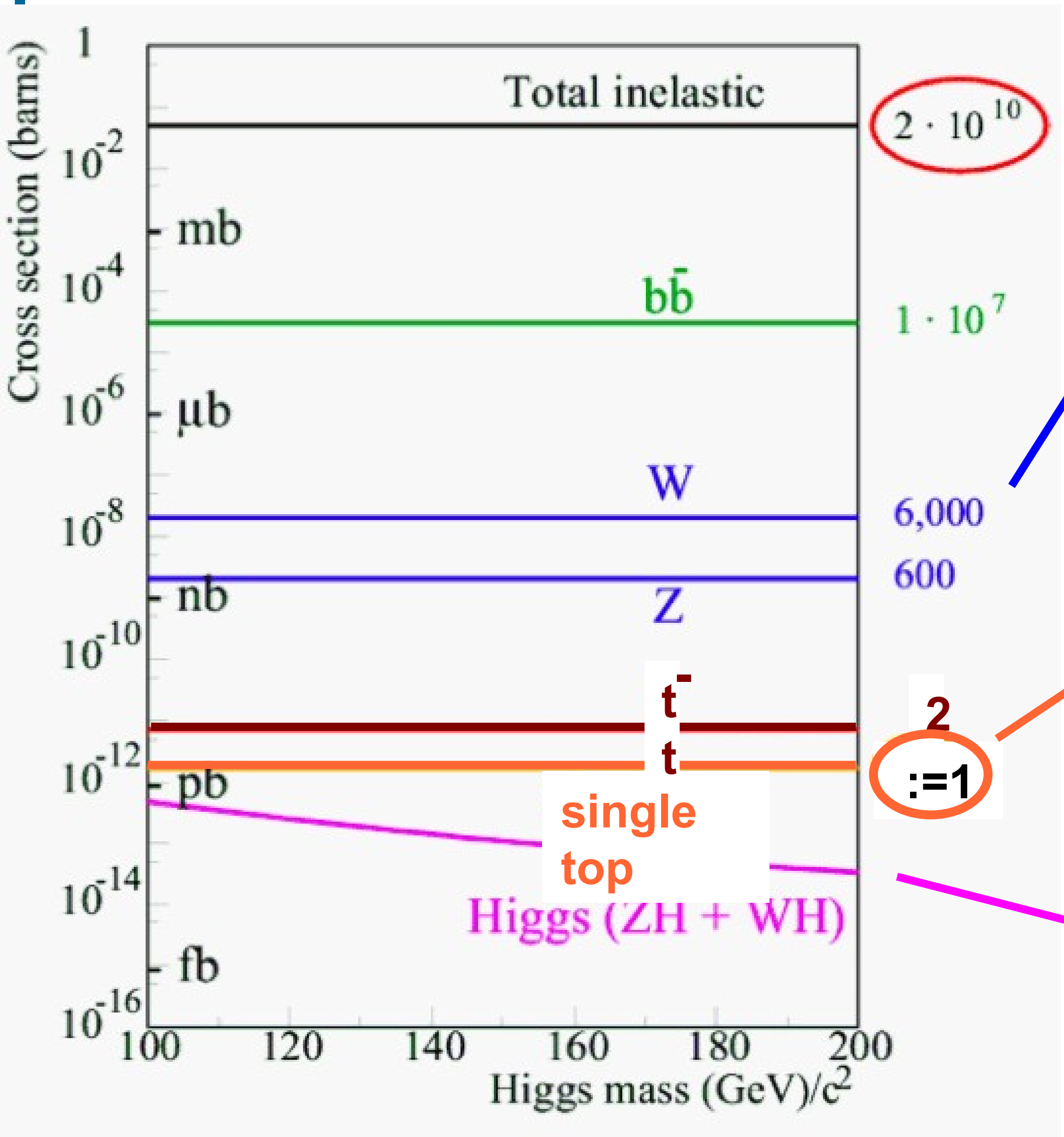
Date rate and size

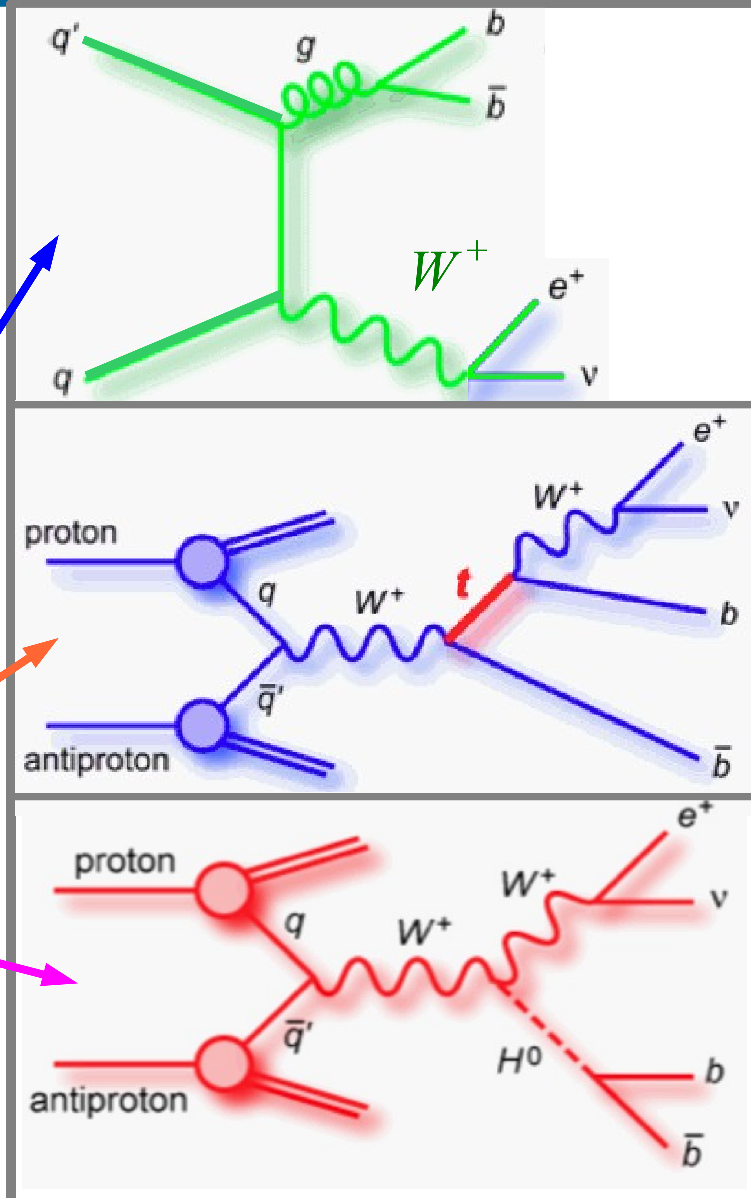


Expected pictures: Higgs decay



Search for new physics



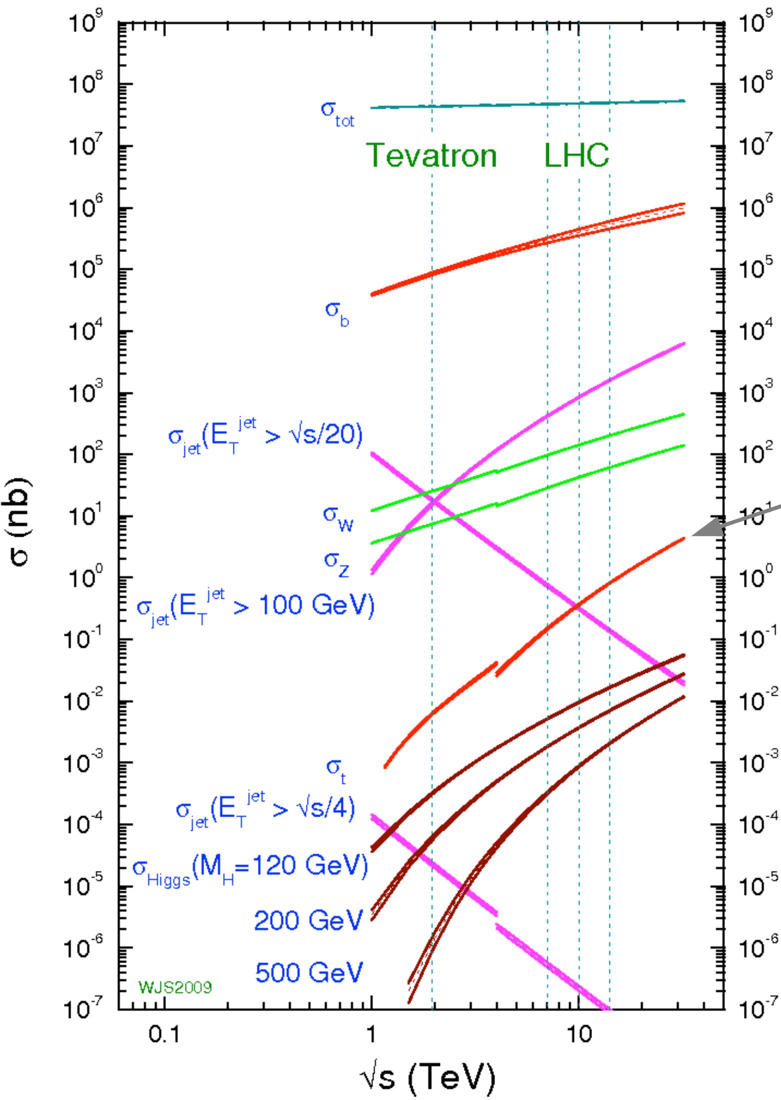


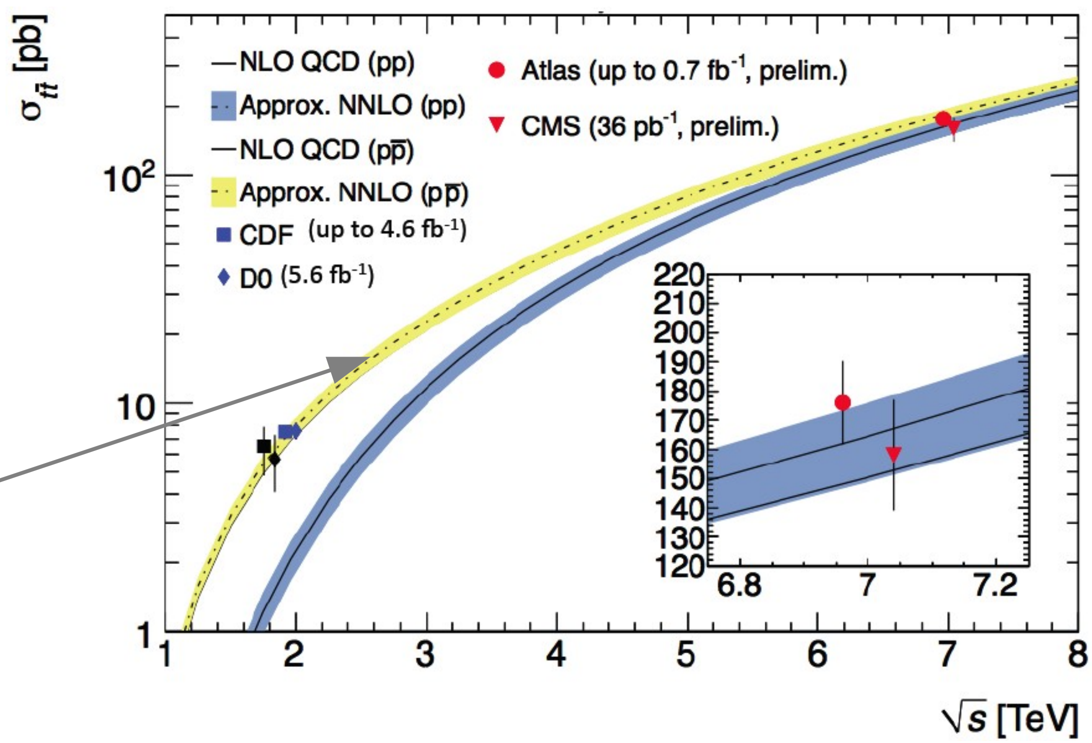
Search for needle in the hay stack

Physics

Expectations and measurements

proton - (anti)proton cross sections

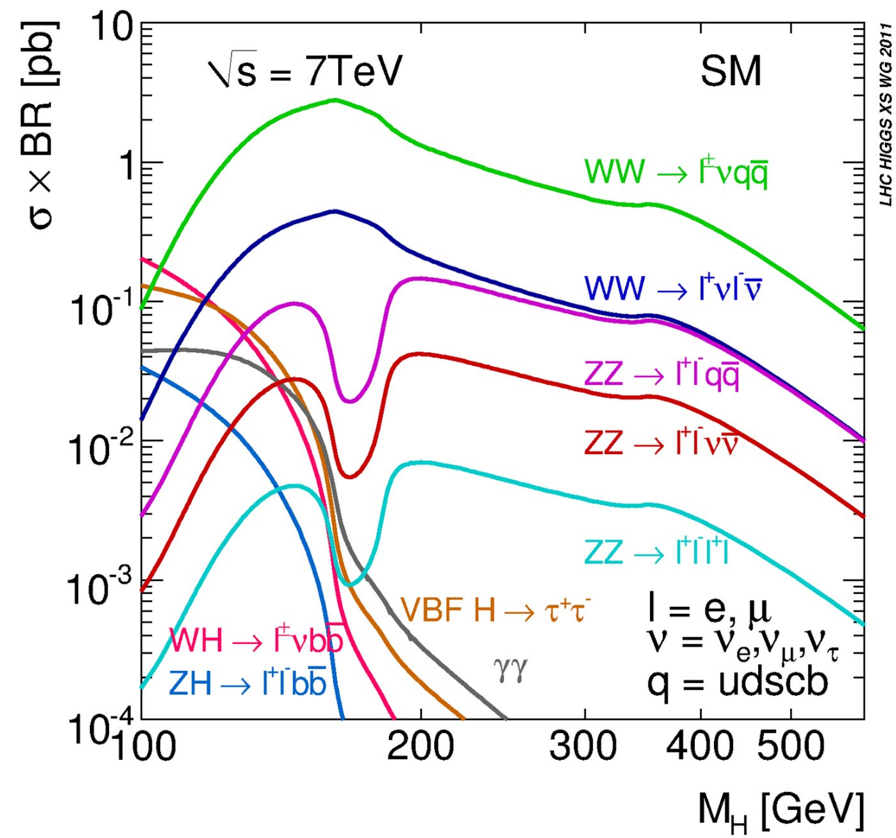




CERN School of Computing

Higgs production at the LHC

- **H→γγ:**
  - → rare channel
  - → best for low Higgs masses
- **H→WW(*):**
  - →lvlv: very important for intermed. masses
  - →lvqq: high rate, important at high mass
- **H→ZZ(*):**
  - → 4l: golden channel
  - → llvv: good for high masses
  - → llbb: also at high masses
- **H→ττ:**
  - → good signal-background ratio
  - → important at low masses, rare channel
  - → very important for Higgs properties

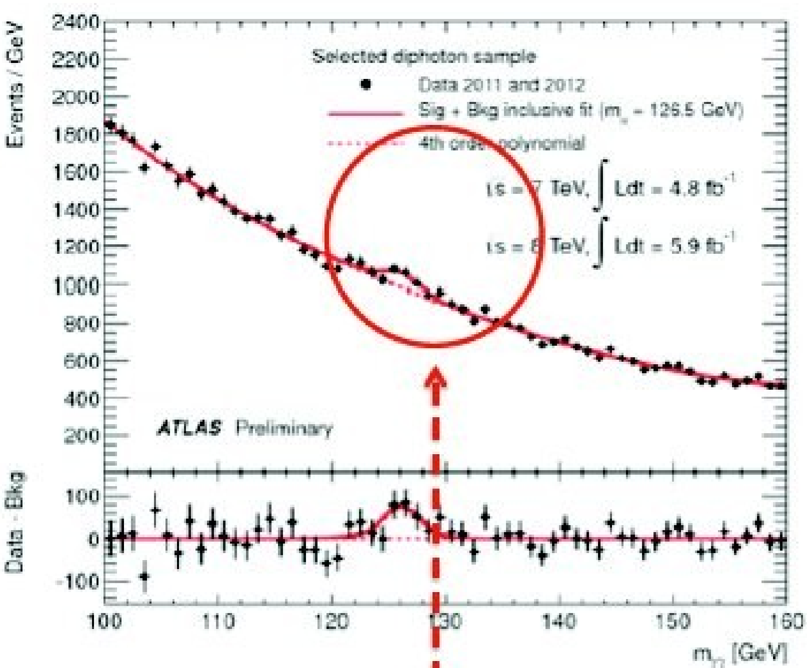


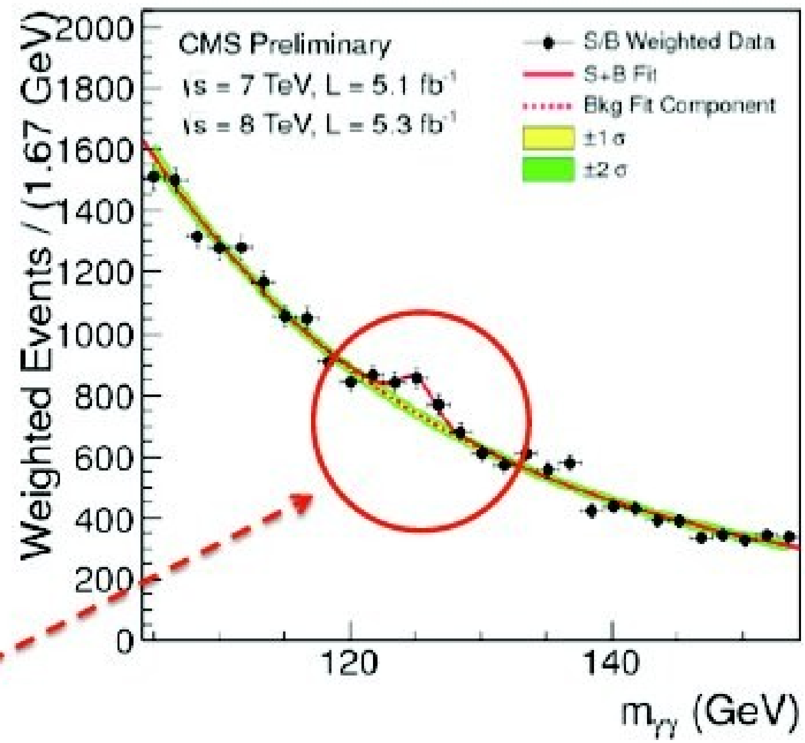
Expected nr. events

MH [GeV]	→WW→lvlv	→ZZ→4l	→γγ
120	127	1.5	43
150	390	4.6	16
300	89	3.8	0.04

Higgs → $\gamma\gamma$

Inclusive/Weighted Mass Spectra



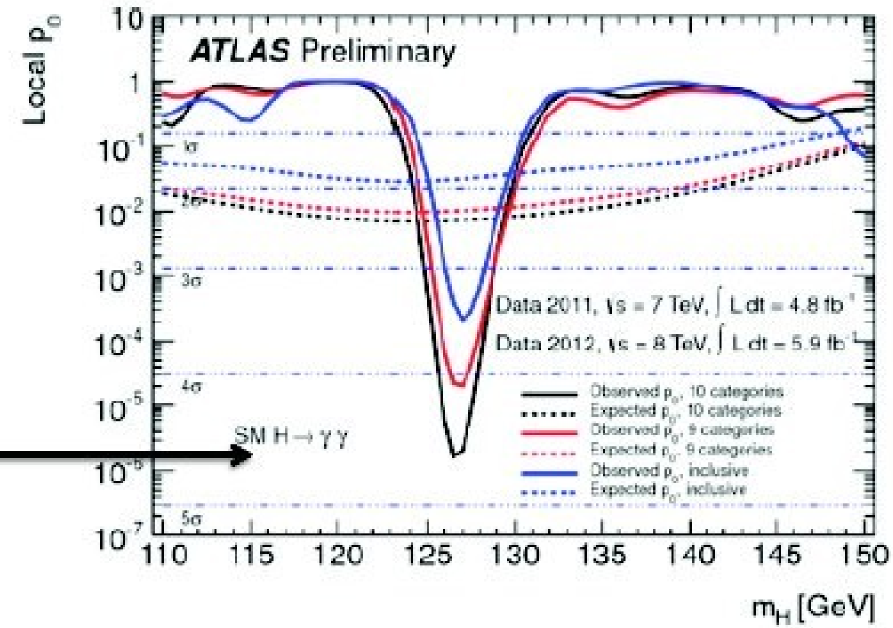


CMS has a slightly wider fit range

Excesses visible in the inclusive mass spectra

CERN School of Computing

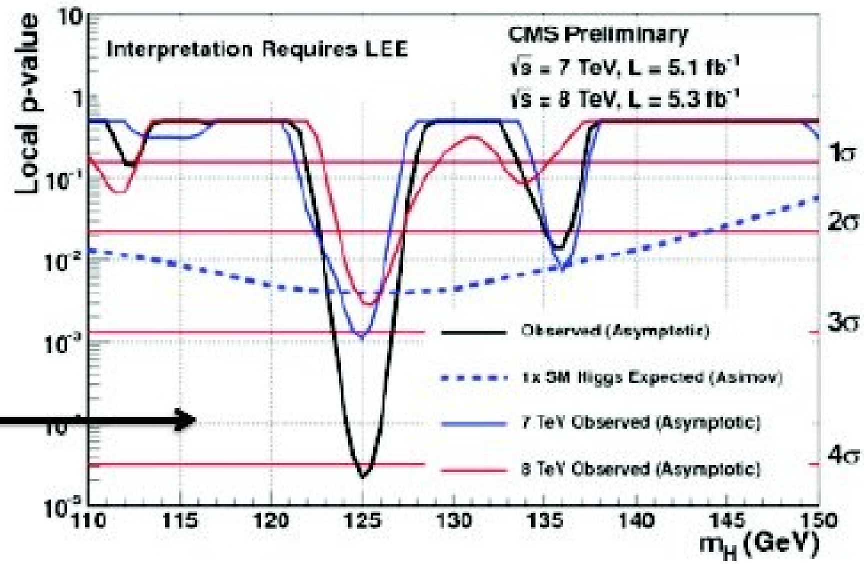
Higgs production at the LHC



ATLAS

Excess of

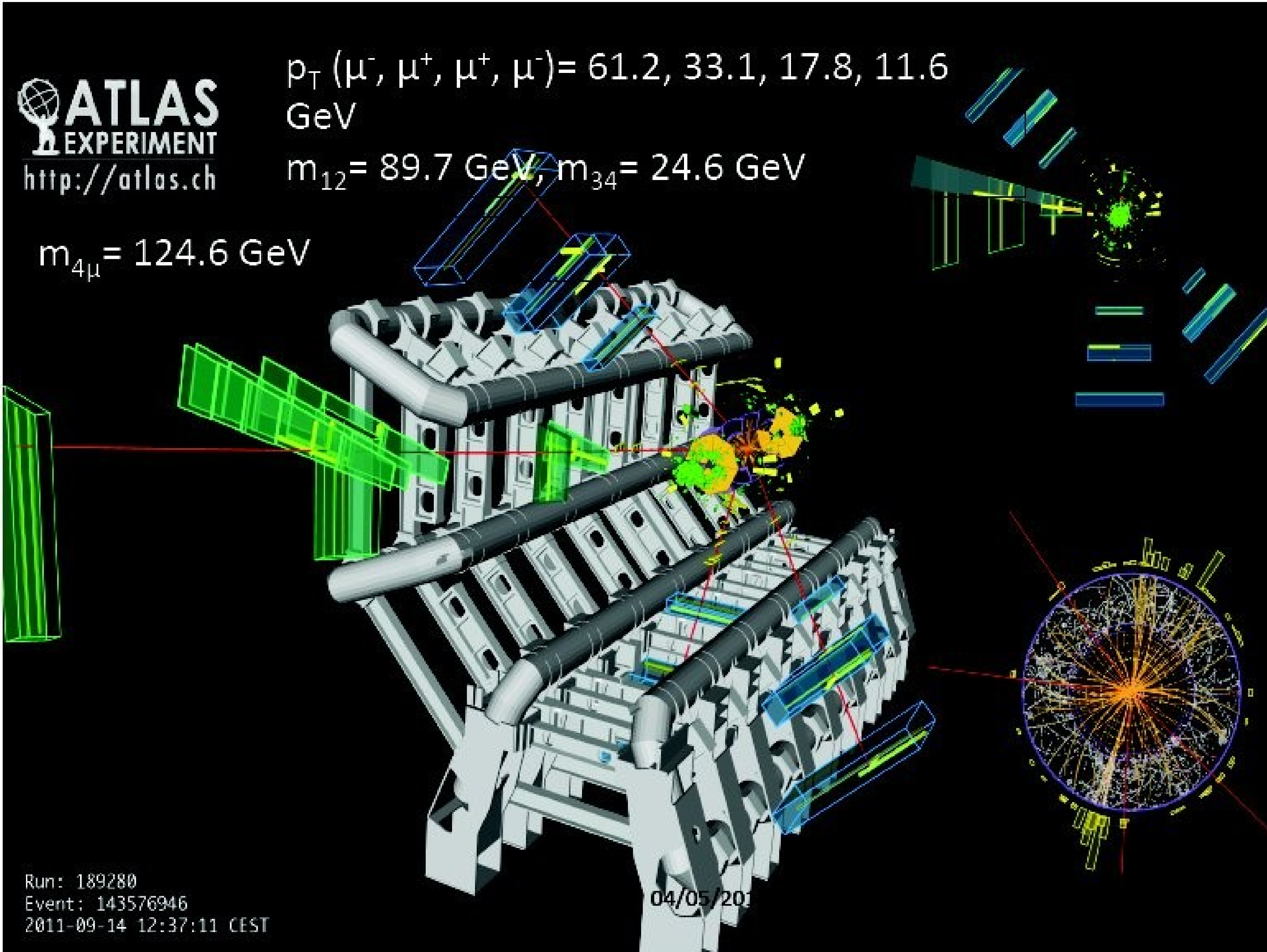
4.5σ

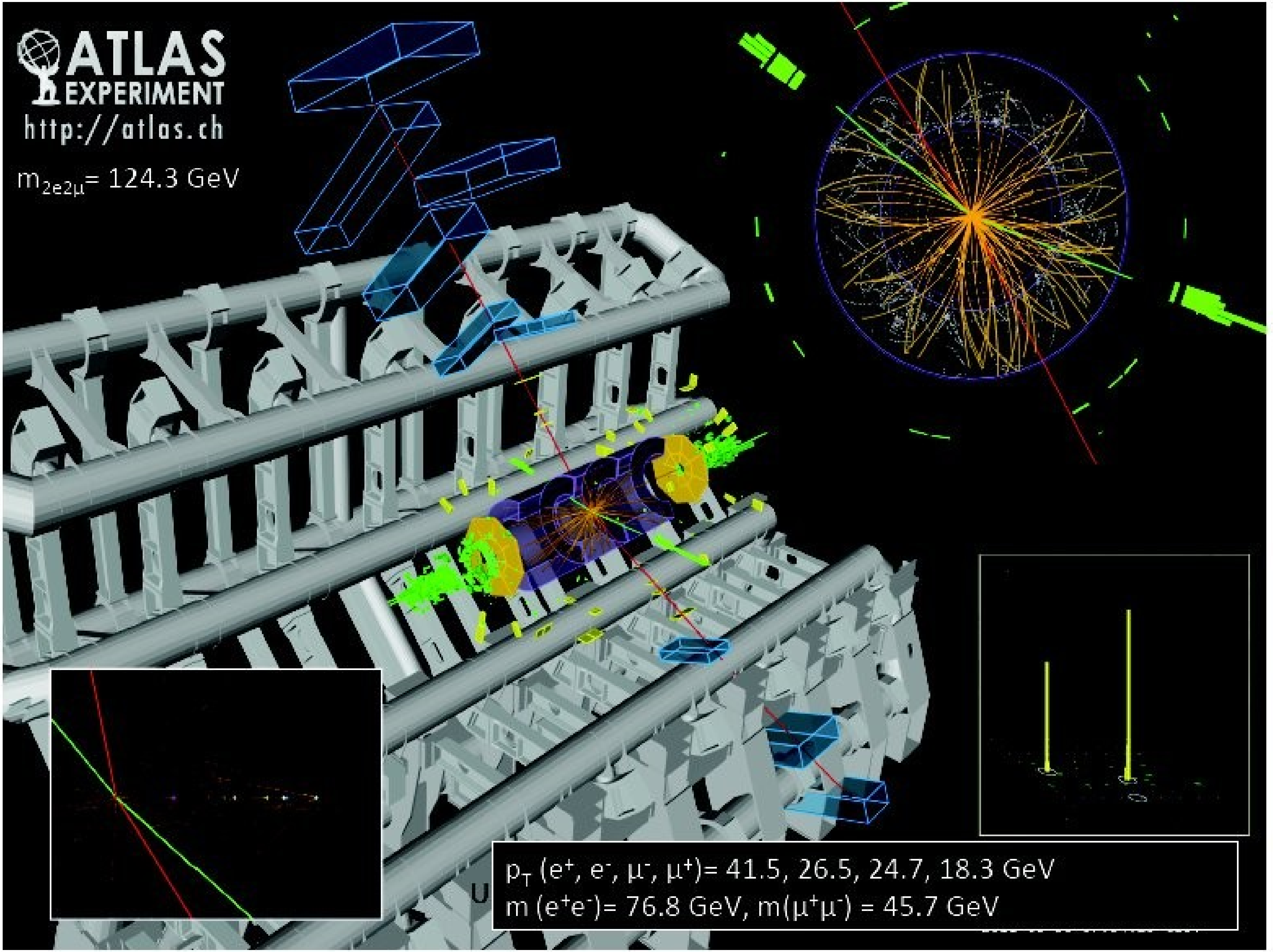


CMS

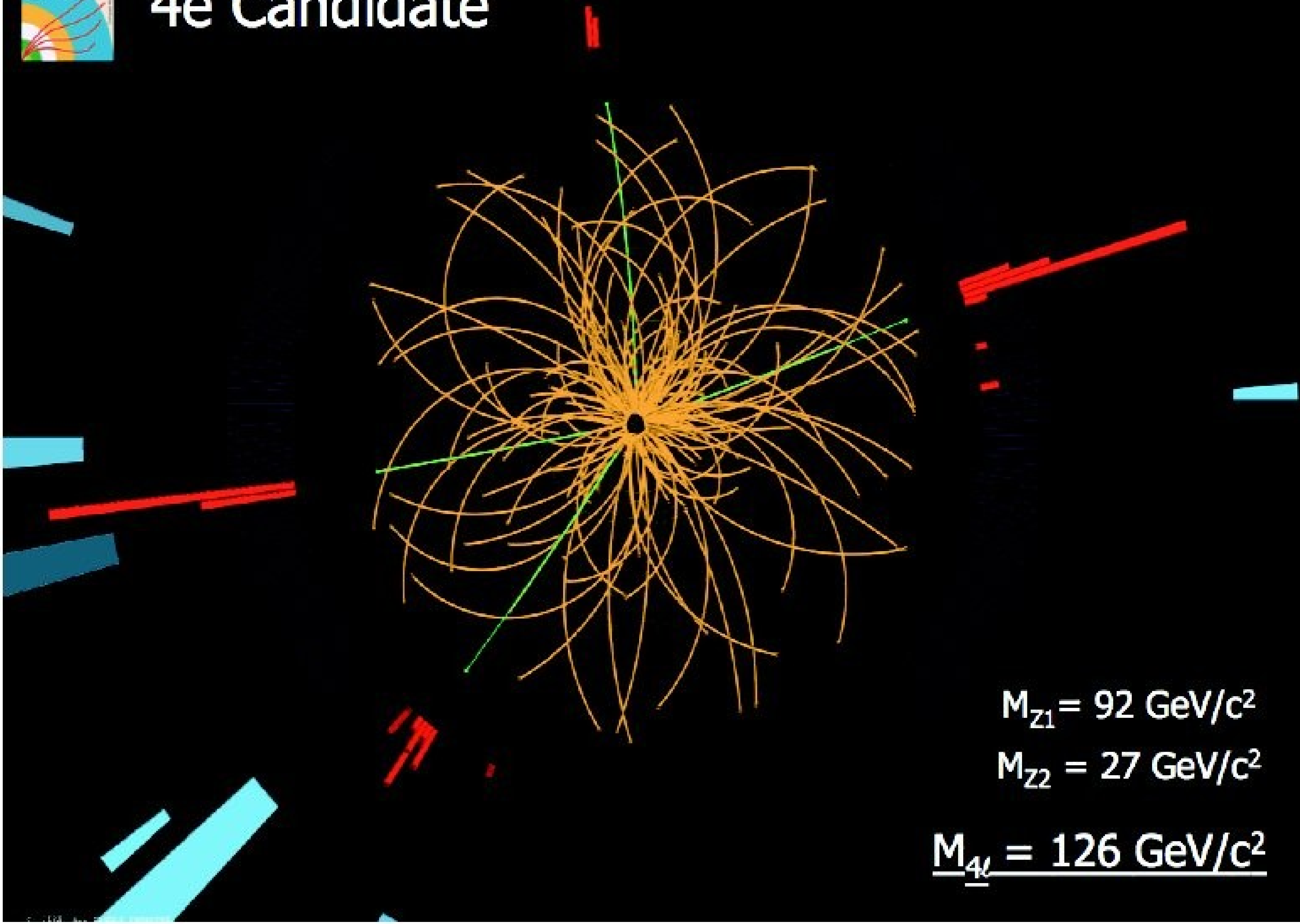
Excess of

4.1σ


ATLAS EXPERIMENT
http://atlas.ch
p_T (μ^-, μ^+, μ^+, μ^-)= 61.2, 33.1, 17.8, 11.6 GeV
m_{12}= 89.7 GeV, m_{34}= 24.6 GeV
$m_{4\mu}$= 124.6 GeV
Run: 189280
Event: 143576946
2011-09-14 12:37:11 CEST


ATLAS
EXPERIMENT
http://atlas.ch
m2e2μ= 124.3 GeV
pT (e+, e-, μ-, μ+)= 41.5, 26.5, 24.7, 18.3 GeV
m (e+e-)= 76.8 GeV, m(μ+μ-) = 45.7 GeV

4e Candidate



$M_{Z1} = 92\ GeV/c^2$

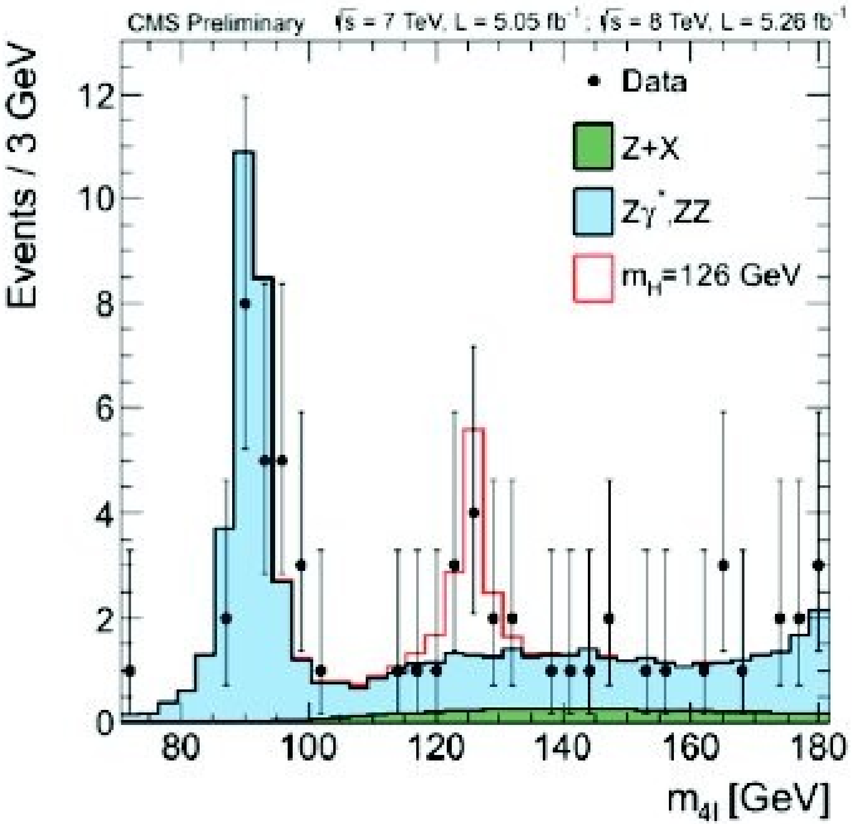
$M_{Z2} = 27\ GeV/c^2$

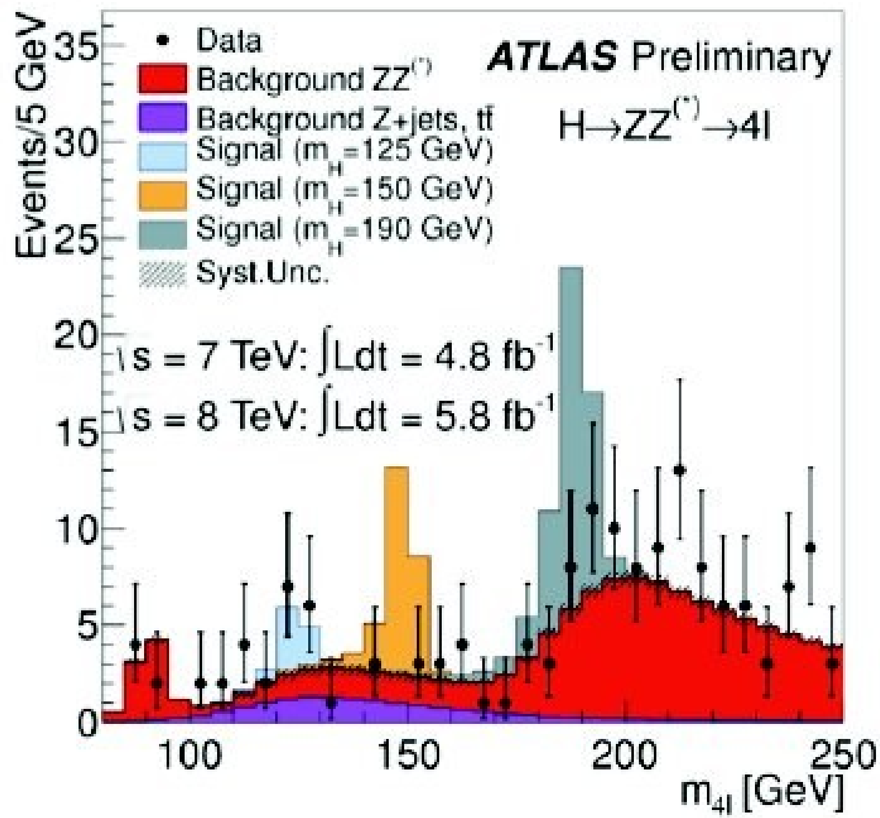
$\underline{M_{4\ell} = 126\ GeV/c^2}$

4μ Candidate

CERN School of Computing

Higgs production at the LHC





Somewhat better S/B in CMS

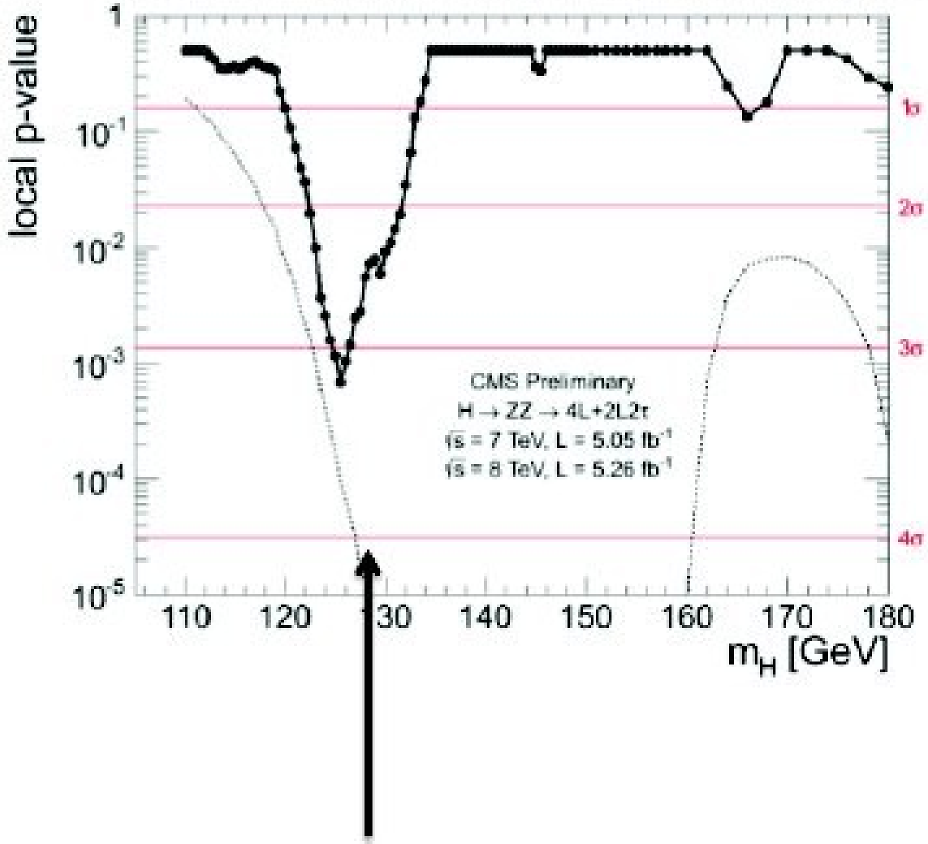
(for instance lower reducible background)

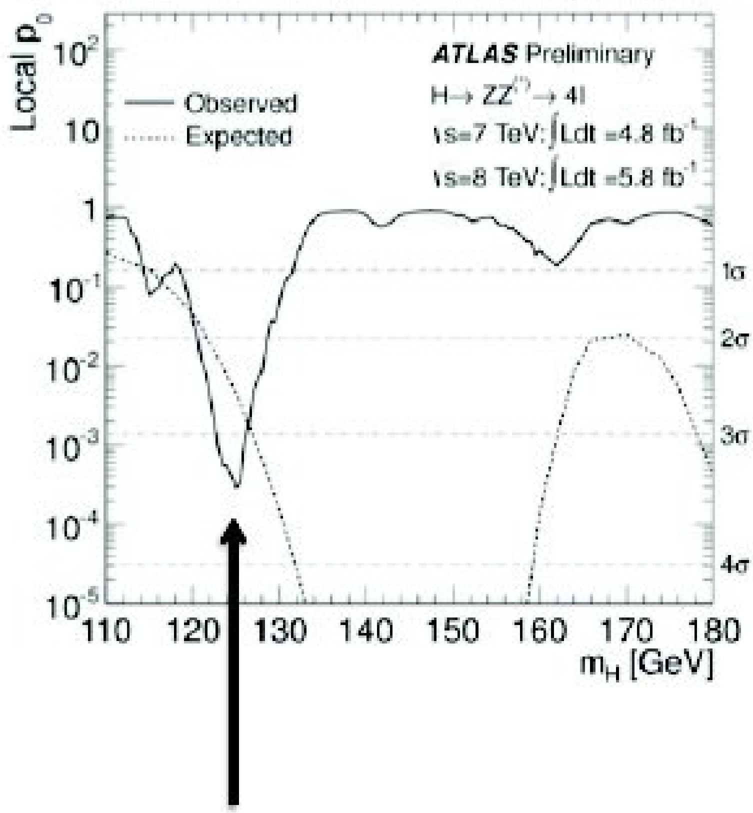
(also 7 TeV ATLAS analysis does not have latest improvements in electron reconstruction)

CERN School of Computing

Higgs production at the LHC

CMS better sensitivity from use of angular variables (20%) + better S/B at low mass (~20%)





CMS

Excess of

3.2σ

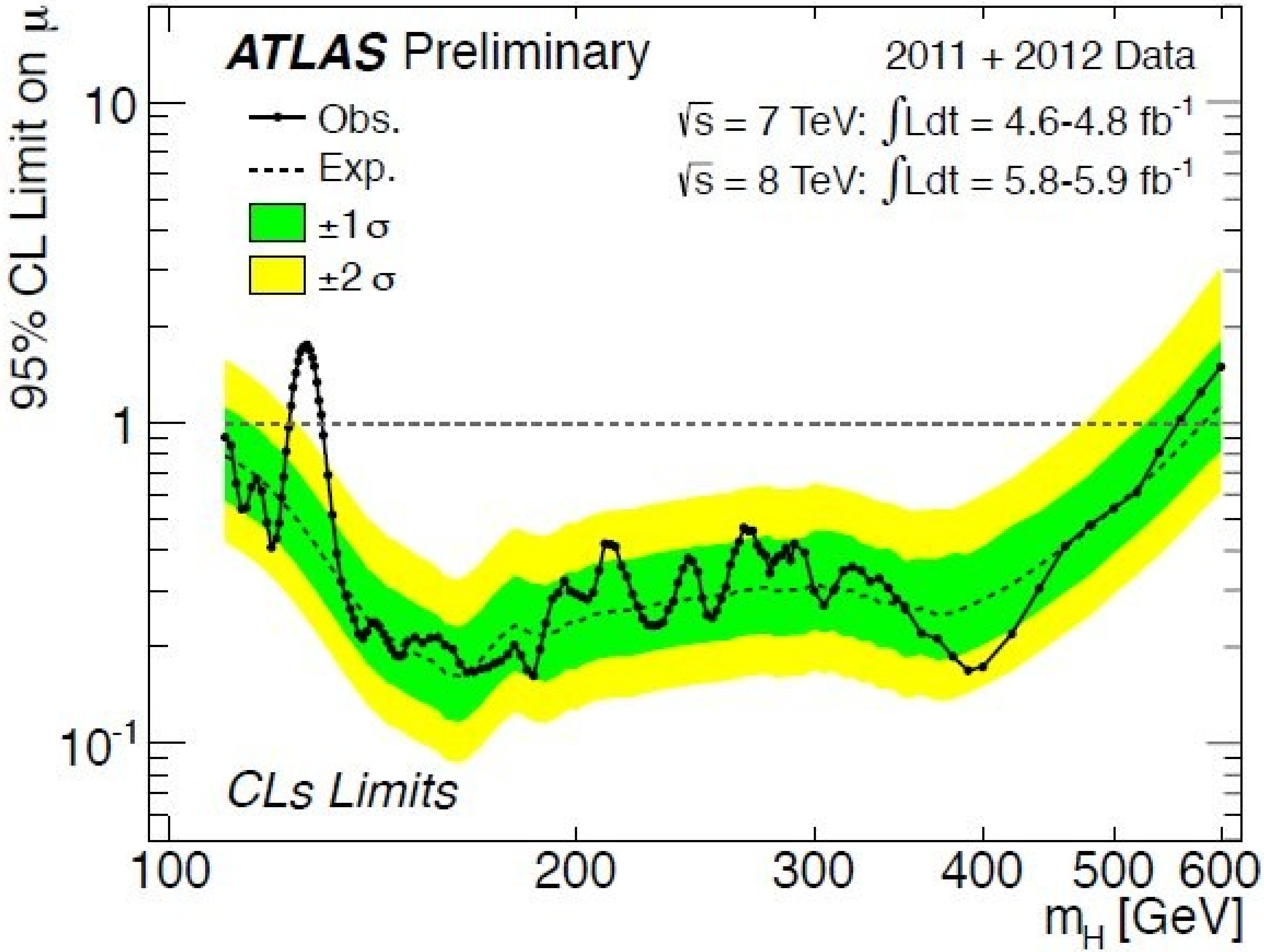
ATLAS

Excess of

3.4σ

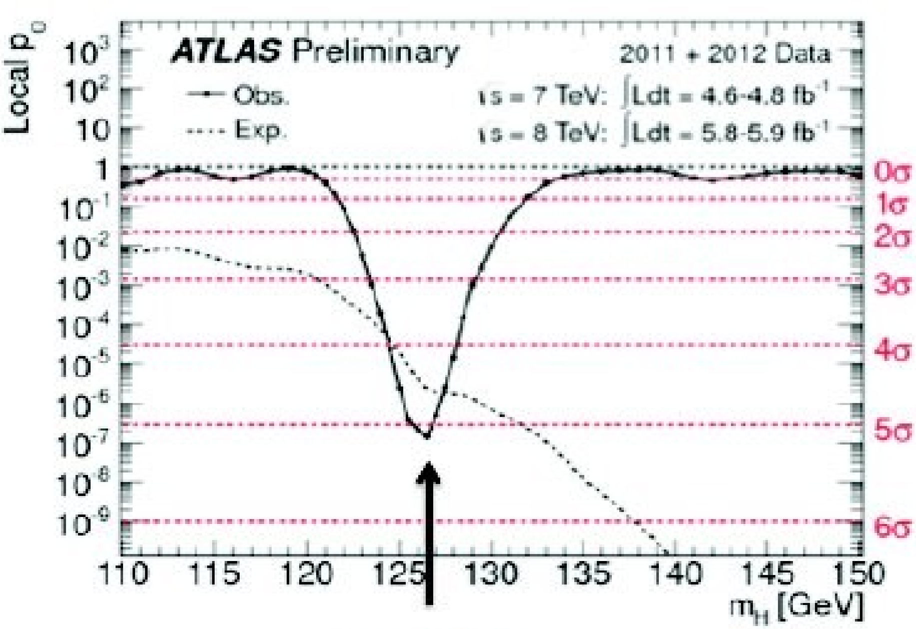
Higgs exclusion

Impressive Exclusion Range for both ATLAS and CMS



CERN School of Computing

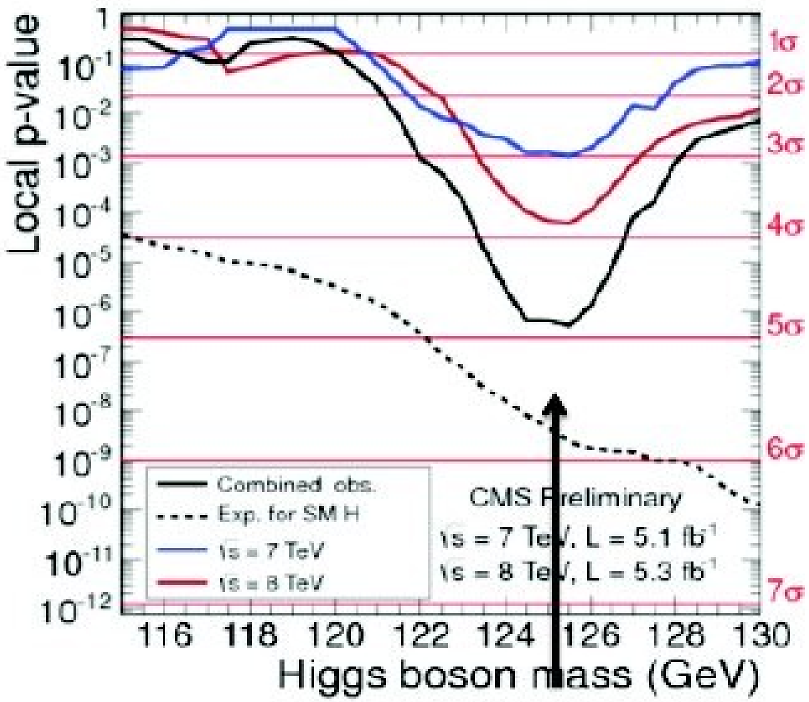
Higgs combination (4^{th} July 2012)



ATLAS

Excess of

5.0σ



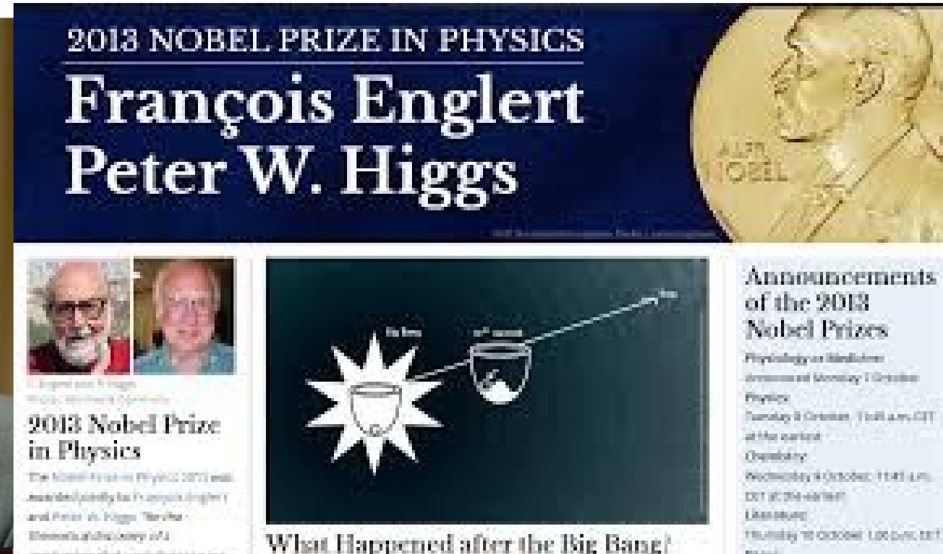
ATLAS

Excess of

4.9σ

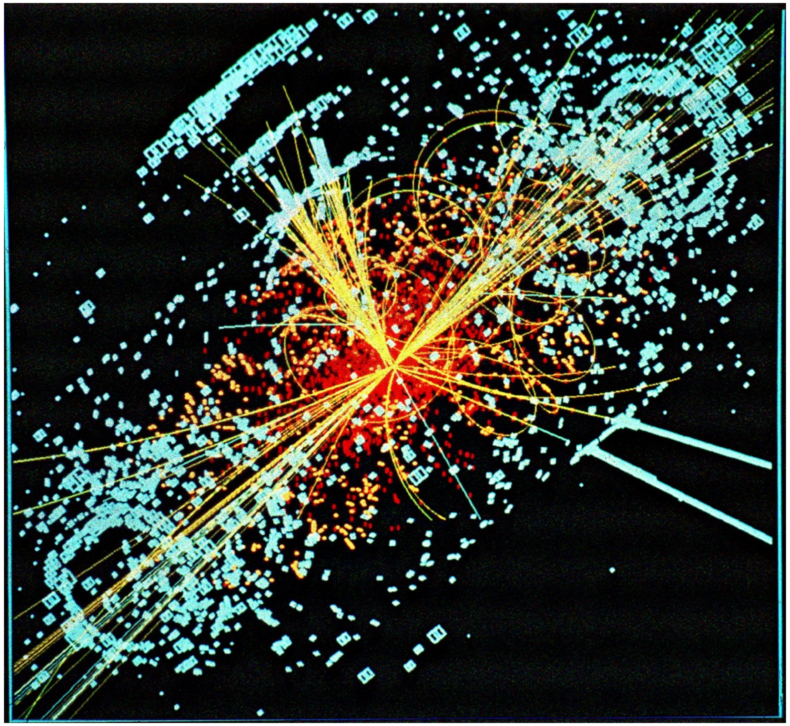
CERN School of Computing

As a Layman: We have it!



Computing

The technical challenge at LHC



Everything in LHC computing is connected to processing such data !!

The technical challenge at LHC (ctd)

- **Very high (design) event rate: 40 MHz**
- **Large event size: O(1) MB**
- **Large background of uninteresting events**
- **Large background in each event**
  - many interactions in each beam crossing
  - pile-up from adjacent beam crossing
  - many low-momentum particles

The technical challenge at LHC (ctd)

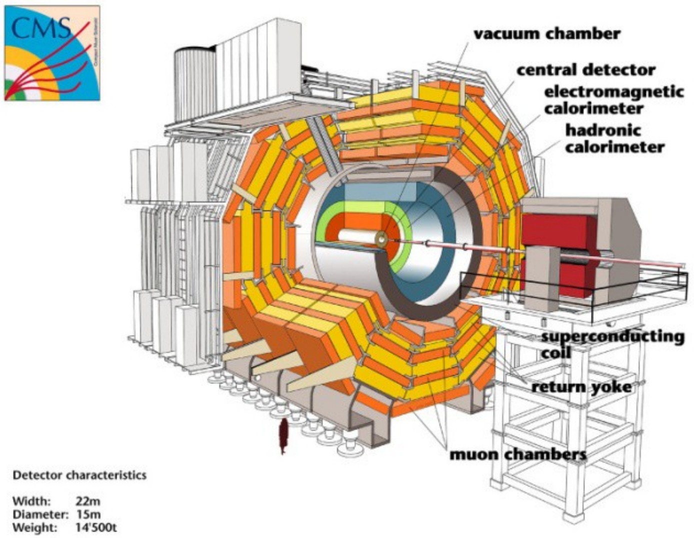
- **Large number of physicists doing analysis**
  - ATLAS and CMS experiments at the LHC: both consist of 170-180 institutes in about 40 countries
  - Distribution of data and programs
  - Bookkeeping is crucial
- **High pressure, competitive spirit**
  - Important discoveries to be (and have been) made
  - Computing has to be as fast as possible

CERN School of Computing

What is Physics Computing?

- Yearly input: A few petabytes of data
- Yearly output: A few hundred physics papers
- Data reduction factor of 10^7 to 10^8 !!
- How is it done?
- Will try to answer this question in this and tomorrow's lectures

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

Actually, 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

CERN School of Computing

What happens to the data?

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

Step by step

- **Each step involves some data reduction**
  - data are discarded (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 (offline reprocessing)**

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 reprocessed

Online processing

- **Trigger: event selection**
  - Needs only a (small) subset of the detector data
  - Fast, as little dead-time (time period when triggering system is insensitive to new data) as possible
  - Gives “green” or “red” light to the data acquisition

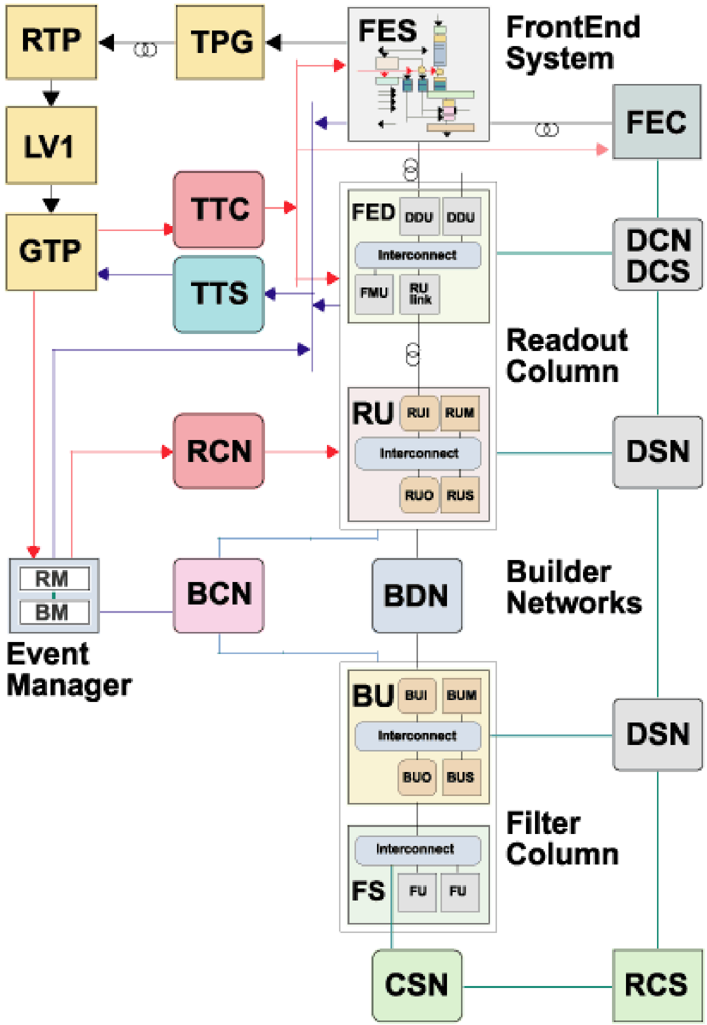
CERN School of Computing

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

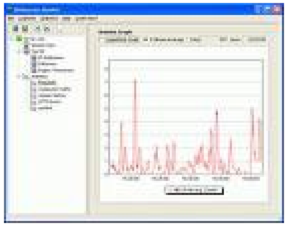


FLT	**FIRST LEVEL TRIGGER**
TPG	Trigger Primitive Generator
TDL	Trigger Data Link
RTP	Regional Trigger Processor
LV1	Calo & Trigger Processors
GTP	Global Trigger Processor
TTC	Timing, Trigger & Control
TTS	Trigger & Throttle System
EVB	**EVENT BUILDER**
BN	Builder Networks
BDN	Builder Data Network
RCN	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, …

CERN School of Computing

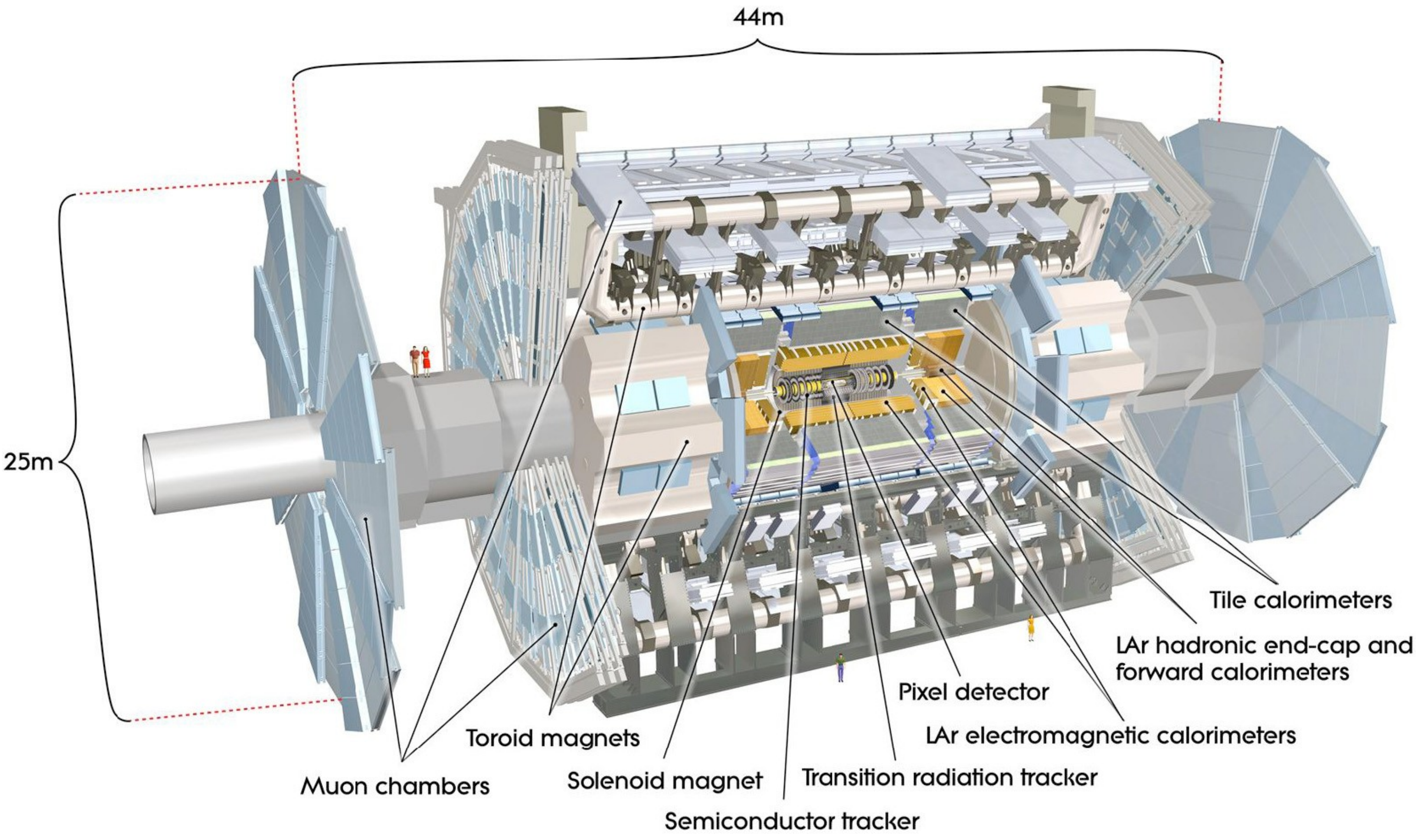
Event selection

- **Primary (design) collision rate: 40 MHz**
- **Recording rate: a few hundred Hz to kHz**
- **How is this achieved?**
  - Multi-level 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)

- Has to be reliable
- Rejected data are lost forever
- Continuous monitoring
- Do not lose new physics
- Must therefore be open to many different signatures of potentially new physics in the detector system

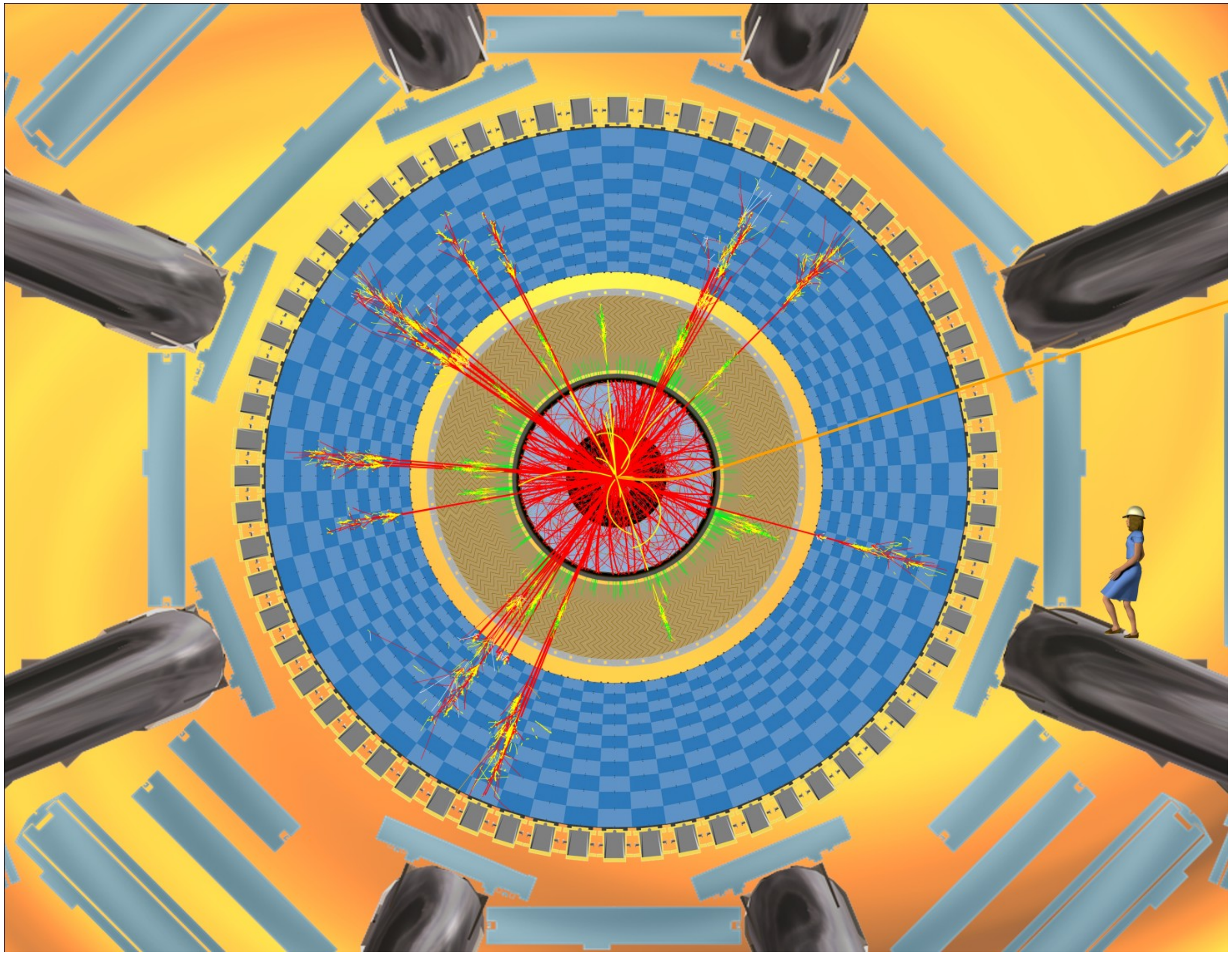
Example: ATLAS



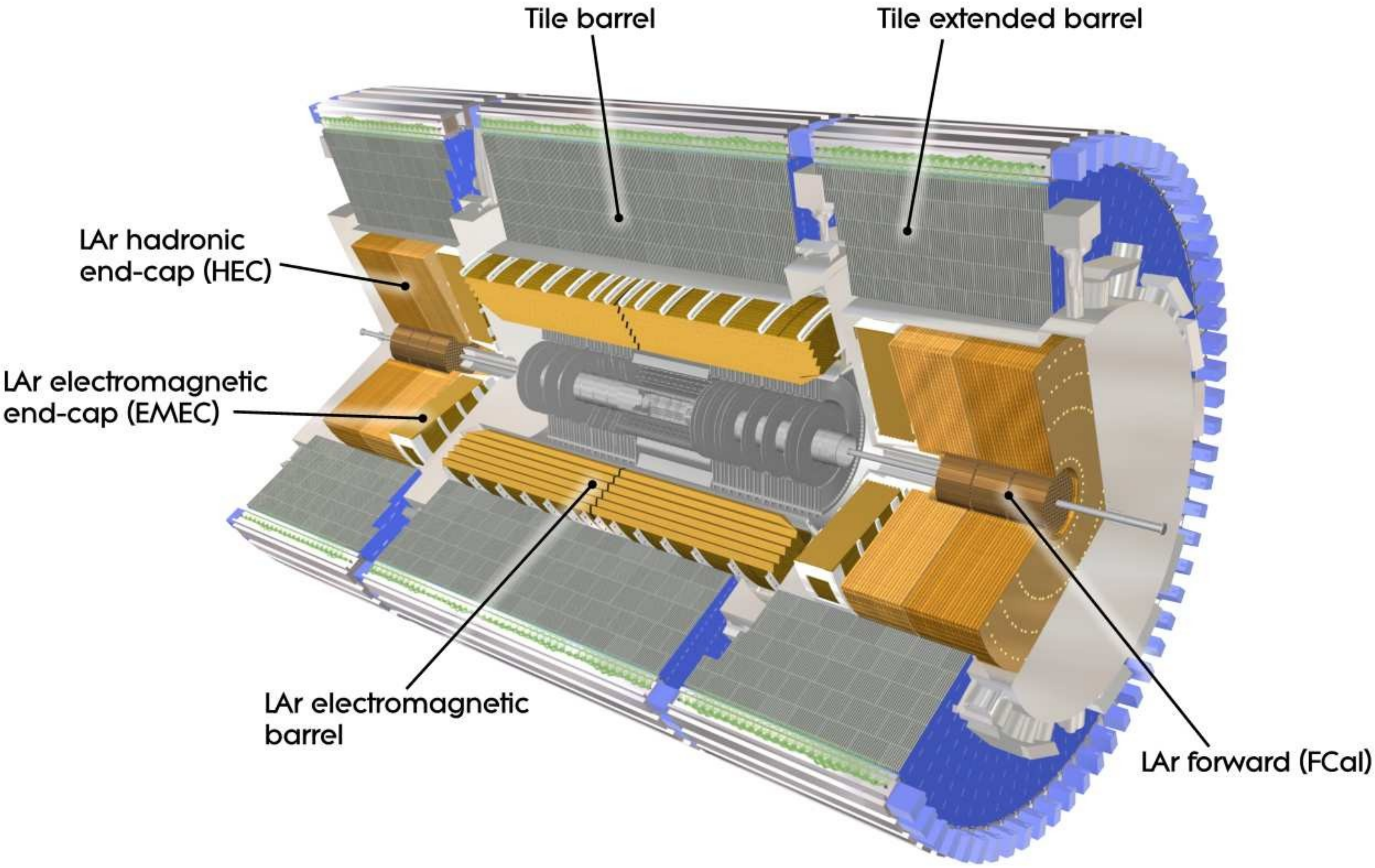
What ATLAS subdetectors measure

- **Inner detector**
  - Momentum and position of charged particles
- **Electromagnetic calorimeter**
  - Energy of photons, electrons and positrons
- **Hadron calorimeter**
  - Energy of charged and neutral hadrons
- **Muon system**
  - Momentum and position of muons

ATLAS detector

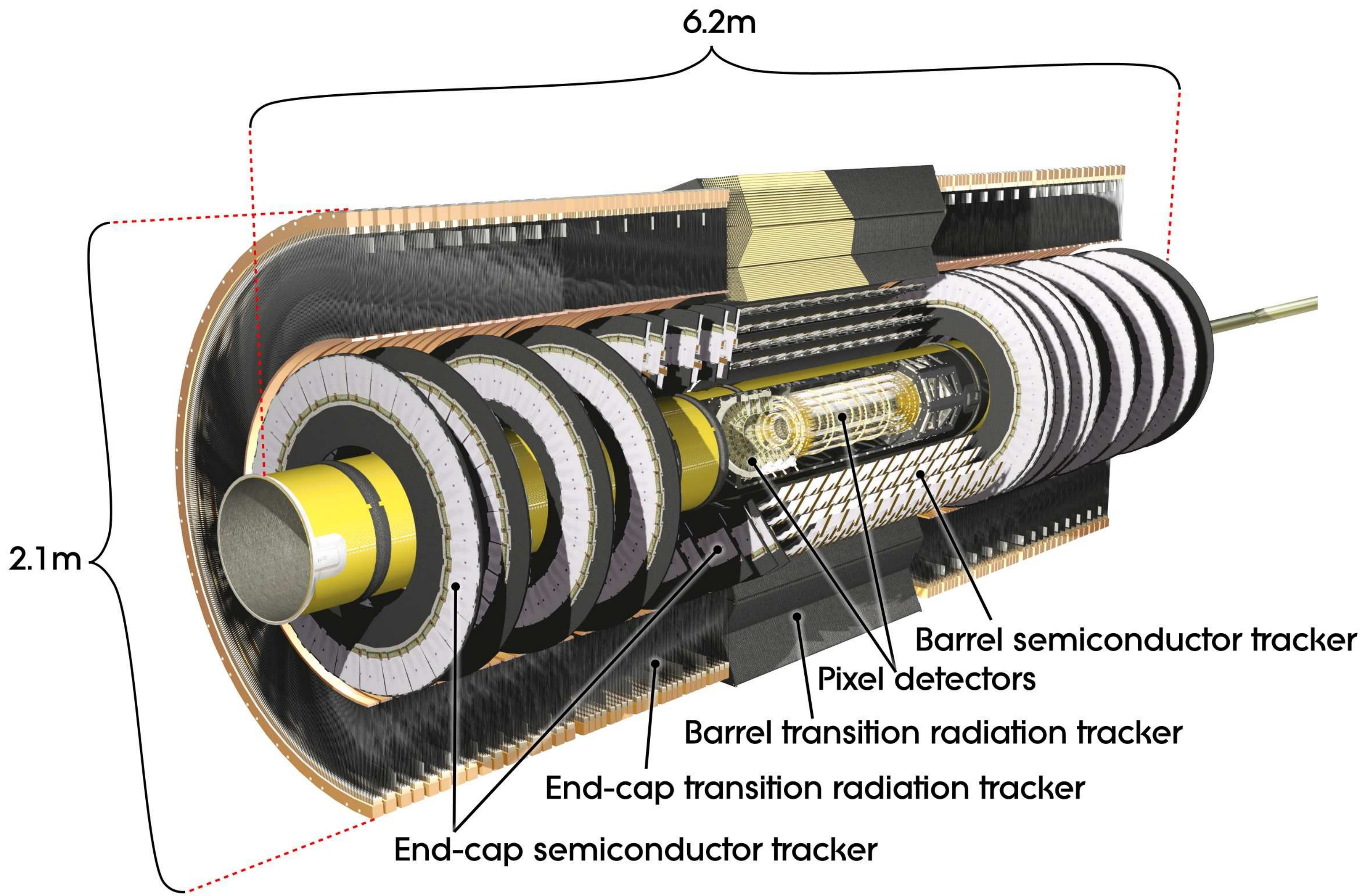


ATLAS detector, calorimeter



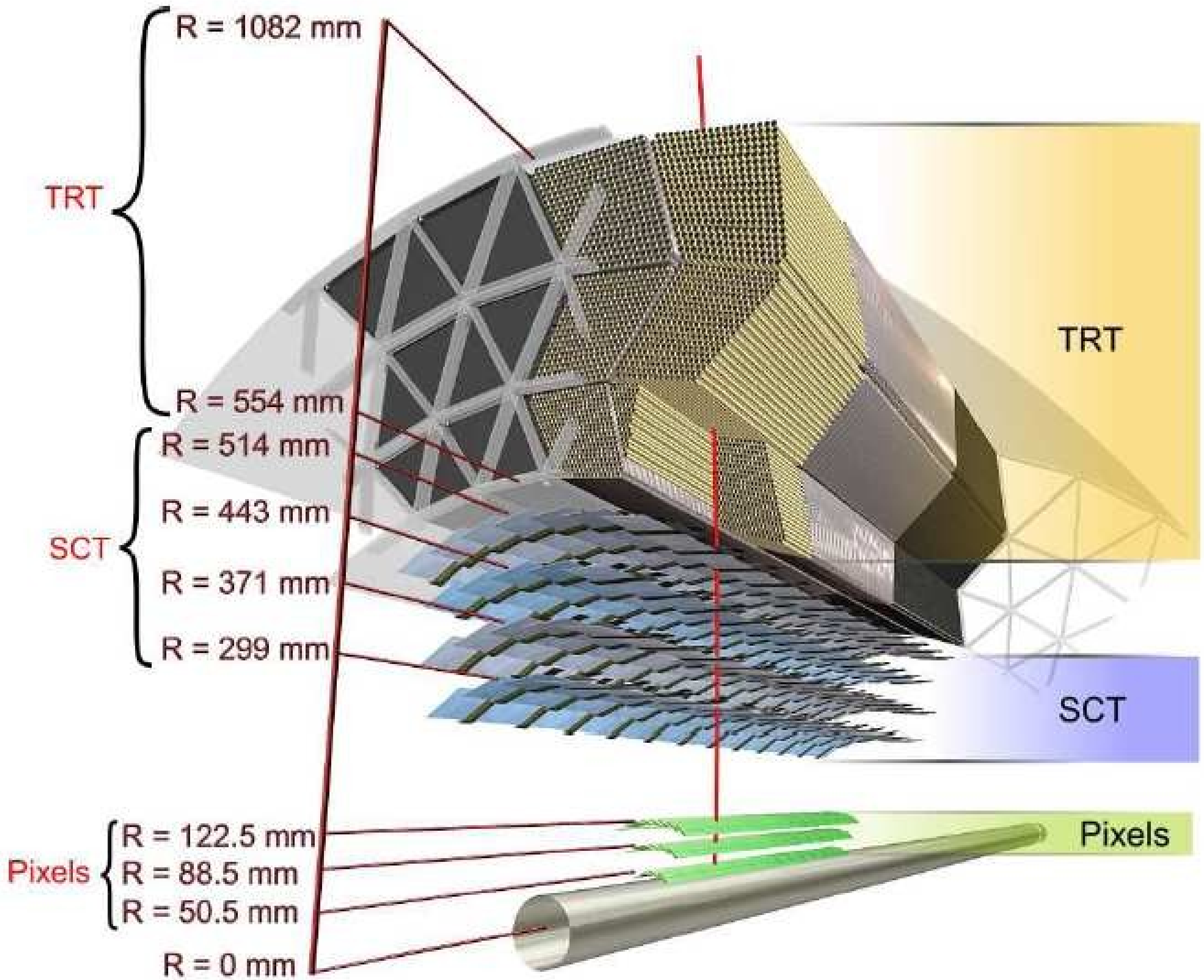
CERN School of Computing

ATLAS detector, inner tracker



CERN School of Computing

ATLAS detector, inner tracker



Event selection (ctd)

- **Overall guideline in designing trigger system: what are the essential features of interesting physics in the detectors?**
  - Typically high-energy particles moving transversely to the
  - beam direction
  - Results in large energy deposits in the calorimetric
  - systems, high-energy muons in the muon system, etc.
- **Multi-level trigger explores such features in various degrees of detail**

CERN School of Computing

Multi level selection

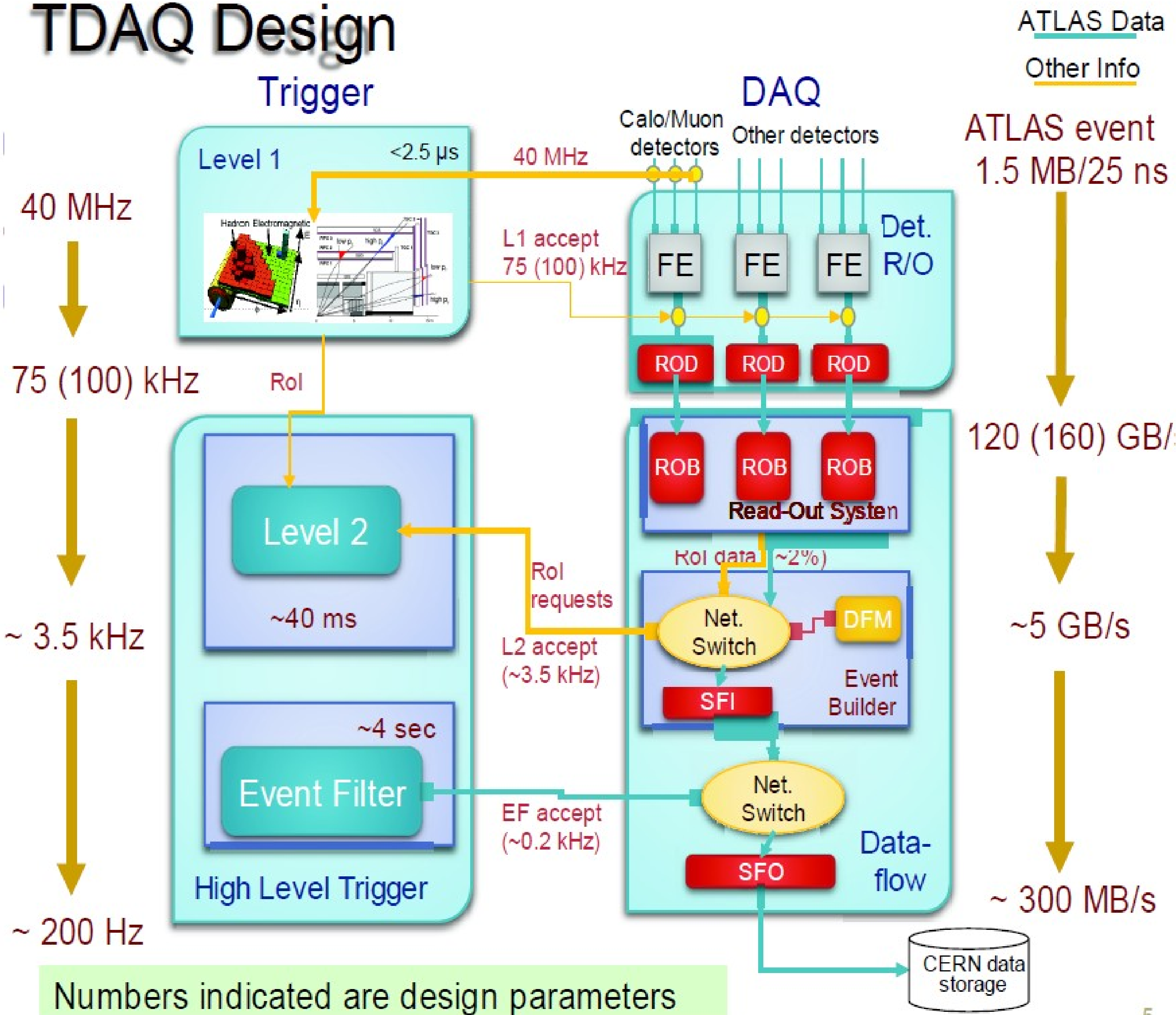
- **Many events can be discarded very quickly – fast level-1 trigger**
- **Only the surviving ones are scrutinized more carefully – high-level filter(s)**
- **Triggers are tailored to specific physics channels (Higgs, top, WW, ZZ, …)**
  - Many such hypotheses are investigated in parallel

CERN School of Computing

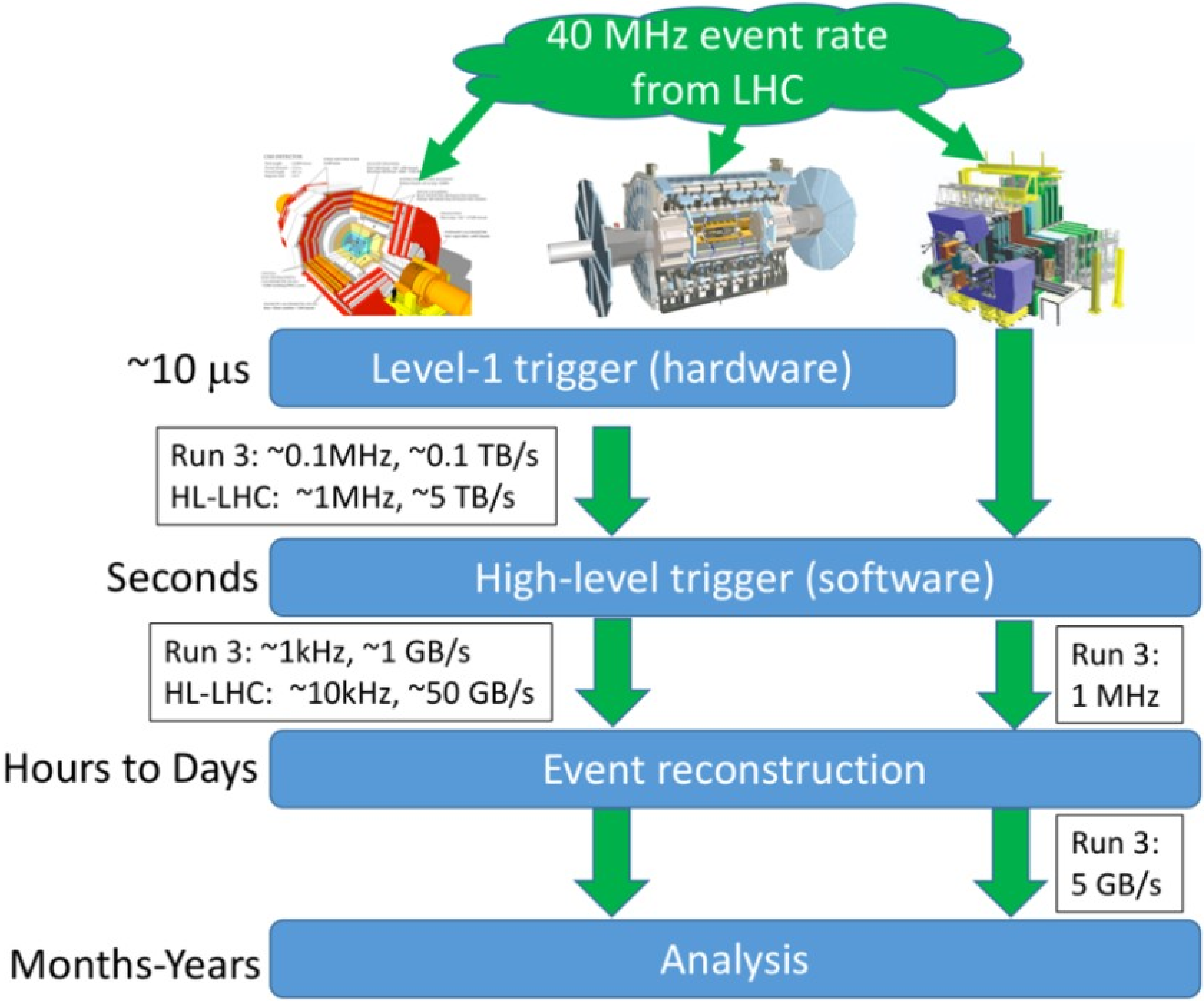
ATLAS triggering system

- **ATLAS has three-level trigger system**
  - Level 1 purely hardware-based (ASICs and FPGAs)
  - High-level trigger (level 2 and Event Filter (EF)) softwarebased
- **Level 1 uses information mainly from calorimeters and muon system**
- **Level 2 also includes information from Inner Detector, uses data from Regions of Interest (RoI) identified by level 1**
- **EF has access to complete set of data and uses same algorithms as offline event reconstruction**

TDAQ Design



Numbers indicated are design parameters


40 MHz event rate from LHC
~10 μs
Level-1 trigger (hardware)
Run 3: ~0.1MHz, ~0.1 TB/s
HL-LHC: ~1MHz, ~5 TB/s
Seconds
High-level trigger (software)
Run 3: ~1kHz, ~1 GB/s
HL-LHC: ~10kHz, ~50 GB/s
Run 3: 1 MHz
Hours to Days
Event reconstruction
Run 3: 5 GB/s
Months-Years
Analysis

CERN School of Computing

Towards High-Lumi LHC

LHC Injectors Upgrade

HiLumi HL-LHC PROJECT

nominal luminosity: 1e34 $cm^{-2}s^{-1}$

Run 1	Run 2	Run 3	Run 4 - 6
Energy: 7-8 TeV Peak lumi: 0.8e34 $cm^{-2}s^{-1}$ Peak <μ>: 40	Energy: 13 TeV Peak lumi: 2.2e34 $cm^{-2}s^{-1}$ Peak <μ>: 65	Energy: 14 TeV Peak lumi: 2.0e34 $cm^{-2}s^{-1}$ Peak <μ>: 60	Energy: 14 TeV Peak lumi: 5-7.5e34 $cm^{-2}s^{-1}$ Peak <μ>: 140-200
L1 peak rate: 70 kHz HLT av. rate: 400 Hz	L1 peak rate: 100 kHz HLT av. rate: 1 kHz	L1 peak rate: 100 kHz HLT av. rate: 1 kHz	L1 peak rate: 1 MHz HLT av. rate: 10 kHz

Two upgrade steps towards HL-LHC

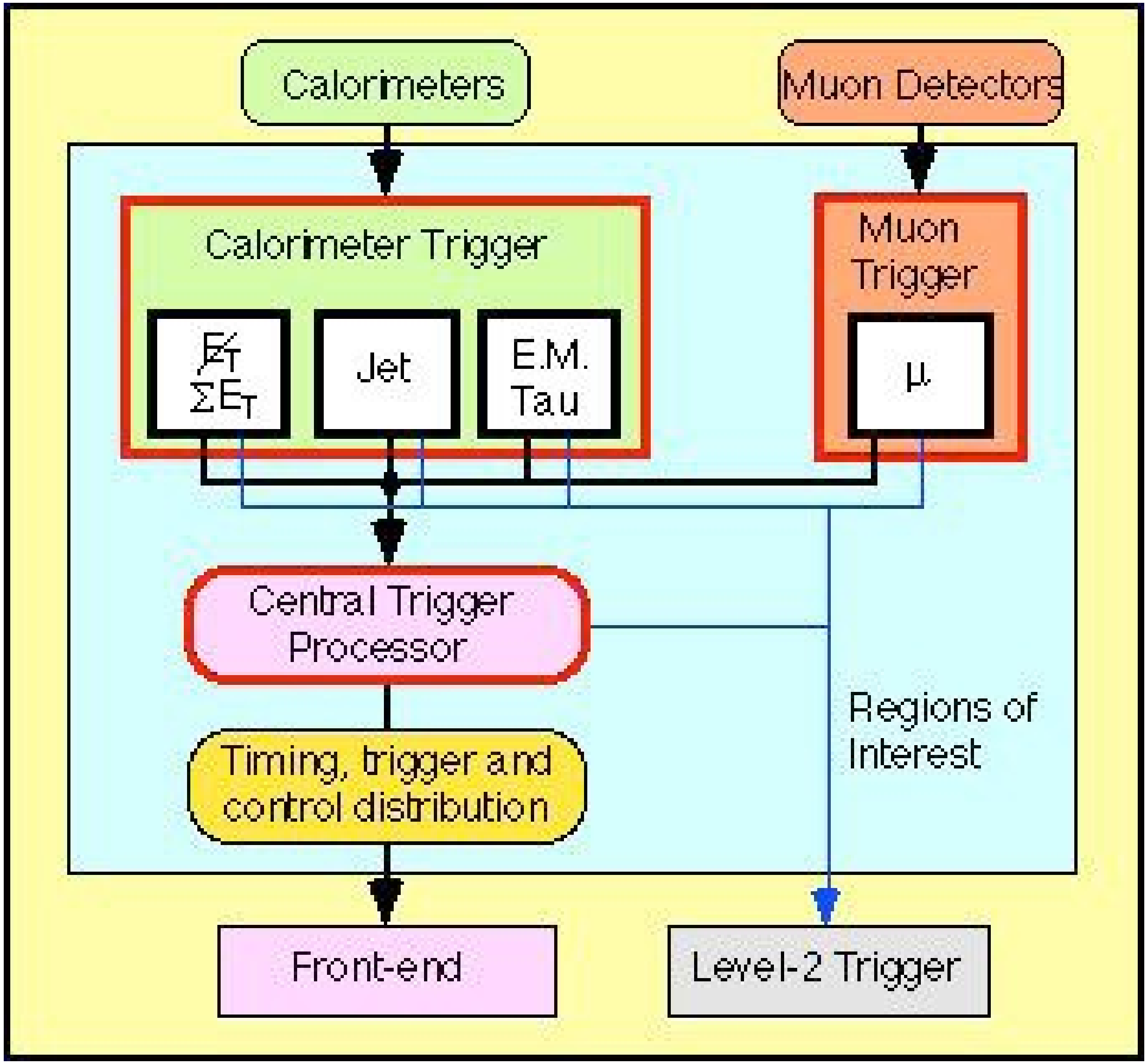
- LHC injector upgrade (LIU) during LS2
- Upgrade to HL-LHC upgrade during LS3
- Both upgrades are needed to reach the ultimate luminosity of 5–7.5 e34 $cm^{-2}s^{-1}$

ATLAS upgrades, including TDAQ and Trigger, match the LHC upgrades

ATLAS L1 trigger

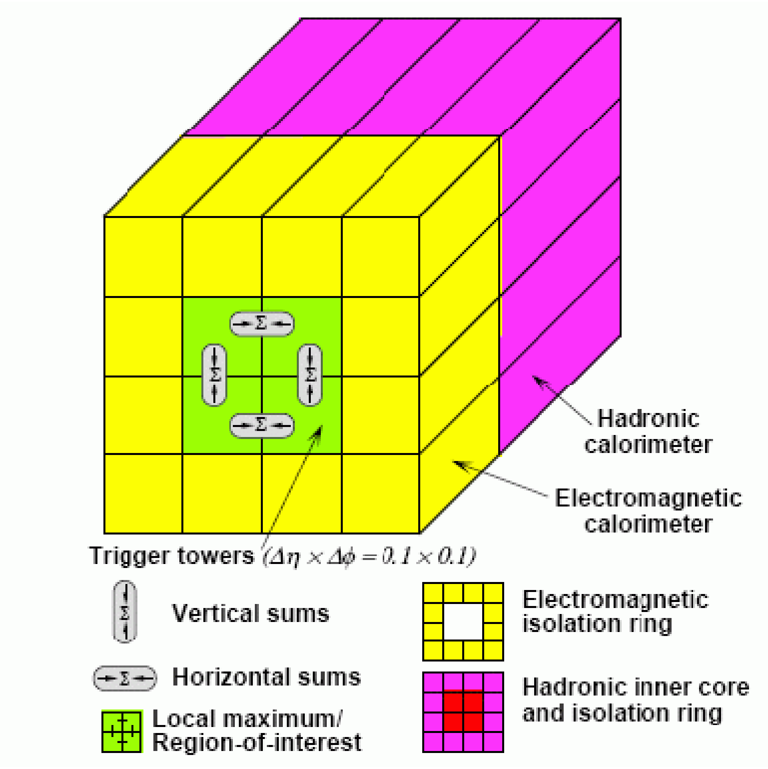
- Input (design) rate: 40 MHz
- Output rate: up to 100 kHz
- Latency (time to reach trigger decision): O(1 μs)
- Data pipelined until trigger decision can be made
- Mainly 2 detector systems: muons/calorimeters

ATLAS L1 trigger



ATLAS L1 calorimeter trigger

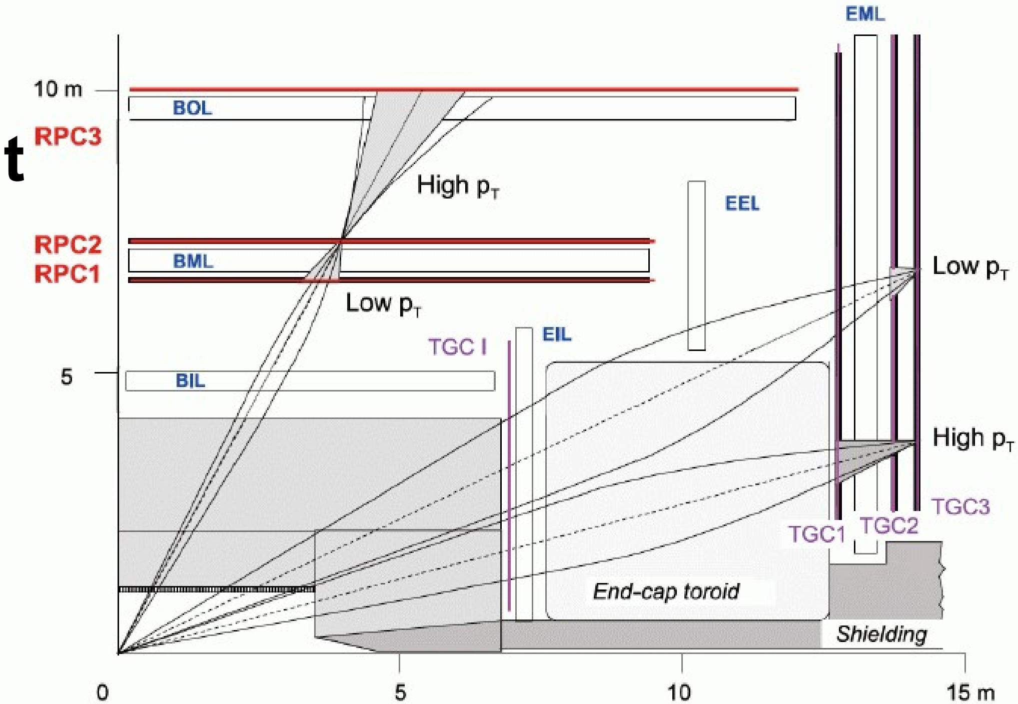
- **High-energy objects in an event:**
  - Electrons/photons
  - Hadronic decays of tau lepton
  - Jet candidates
- **Global event properties:**
  - Total transverse energy (ET)
  - Missing ET
  - Jet sum ET
- **Sends to Central trigger:**
  - Multiplicity of electrons/photons and jets passing thresholds
  - Thresholds passed by total and missing ET



CERN School of Computing

ATLAS L1 muon trigger

- **Dedicated muon trigger chambers with good time resolution:**
  - RPCs (barrel region)
  - TGCs (endcap regions)
  - Search for patterns of
- **measurements consistent with high momentum muons coming from collision point**



ATLAS L1 CTP

- **Central Trigger Processor**
- **L1 inputs are combined to form L1 items**
  - e.g. an input EM10 (electromagnetic cluster above 10 GeV) can be used in the generation of several L1 items:
    - L1_EM10: At least one EM cluster above 10 GeV
    - L1_2EM10: At least two EM clusters, each above 10 GeV
    - L1_EM10_MU6: An EM cluster above 10 and a muon above 6 GeV.
- **A L1 Accept is generated and sent to the detector readout electronics only if at least one L1 item survives.**

High-Level Filter

- **Further data selection:**
  - Up to 100 kHz input rate
  - A few hundred Hz 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 (in contrast to ATLAS two-stage solution)

CMS high-level trigger

- **Has to keep pace with the L1 Output (up to 100 kHz)**
- **Solution: massive parallelism**
- **Filter farm**
  - O(10000) cores
  - Decision time: O(100) ms

CERN School of Computing

CMS high-level trigger (ctd)

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

CERN School of Computing

Output of CMS high-level trigger

- **Raw data are sent to Tier-0 farm (at CERN)**
  - Detector data (zero compressed)
  - Trigger information + some physics objects
  - O(50) physics datasets, depending on trigger history, O(10) online streams (calibration/monitoring/alignment)
- **Physics: O(1) MB @ a few hundred Hertz = a few hundred MB/sec**
- **Alignment/Calibration: O(50) MB/sec**

Output of CMS high-level trigger (ctd)

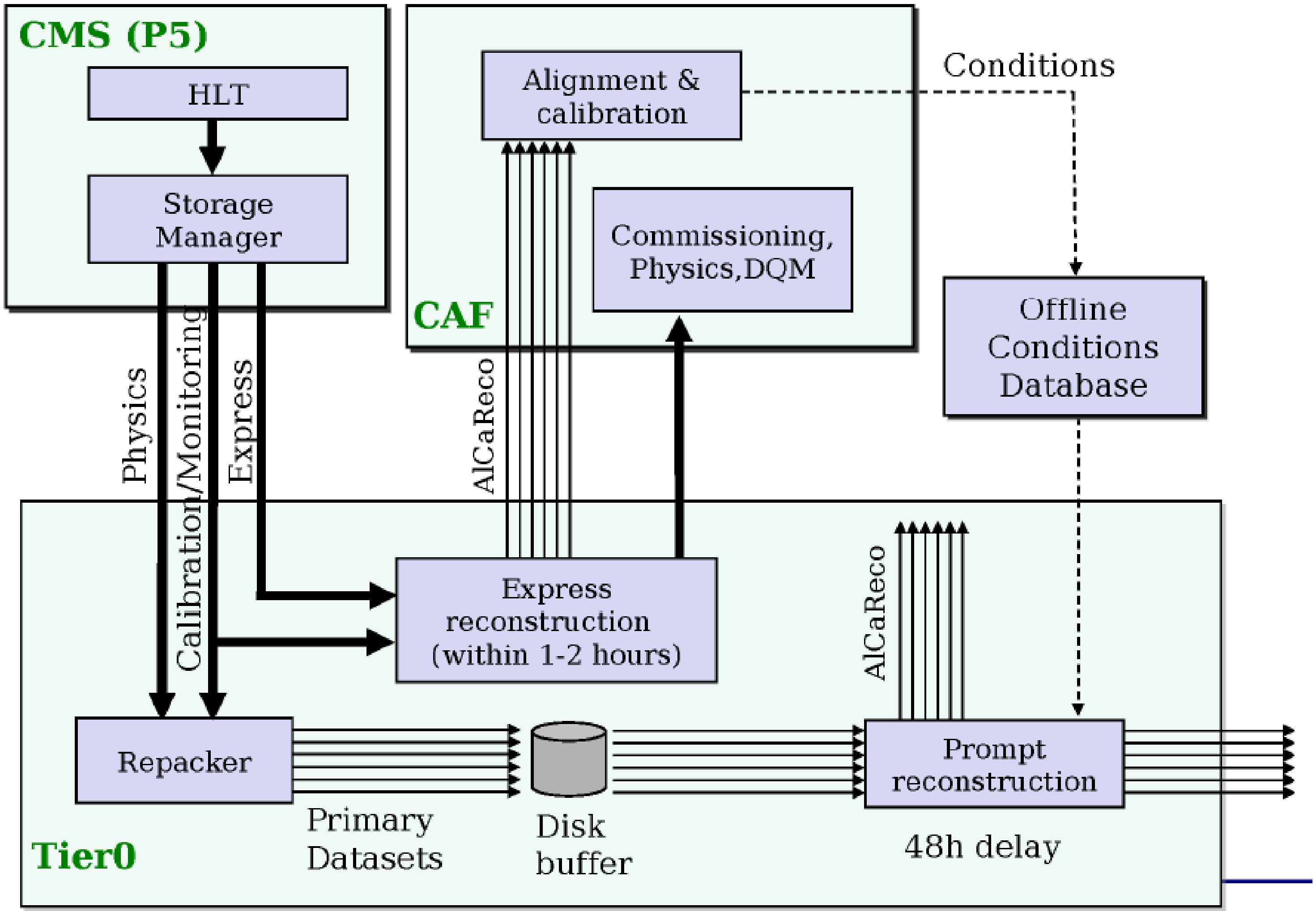
- LHC runs for ~ 10^7 sec/year
- A few PB per year at design luminosity

Tier-0 processing

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

CERN School of Computing

Tier-0 processing (ctd)



CERN School of Computing

Summary, event selection

- **Selecting a small subset of all collision events for offline analyses**
  - Reducing from 40 MHz collision rate to recording rate of a few hundred Herz
- **Multi-level triggering system**
  - Looking for signatures of potentially interesting physics in detectors
  - First level purely hardware-based with pipelined data
  - Higher level(s) software-based, massively parallelized on filter farms

CERN School of Computing

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)

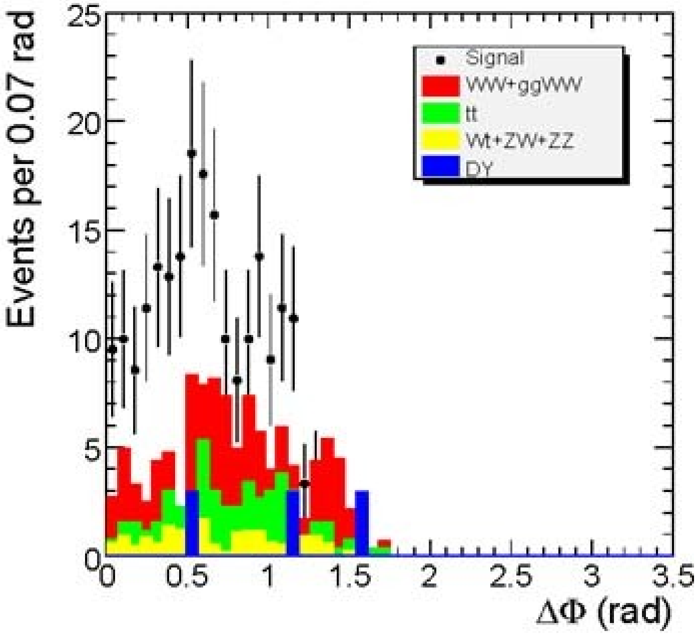
Offline Processing (ctd)

- **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, …



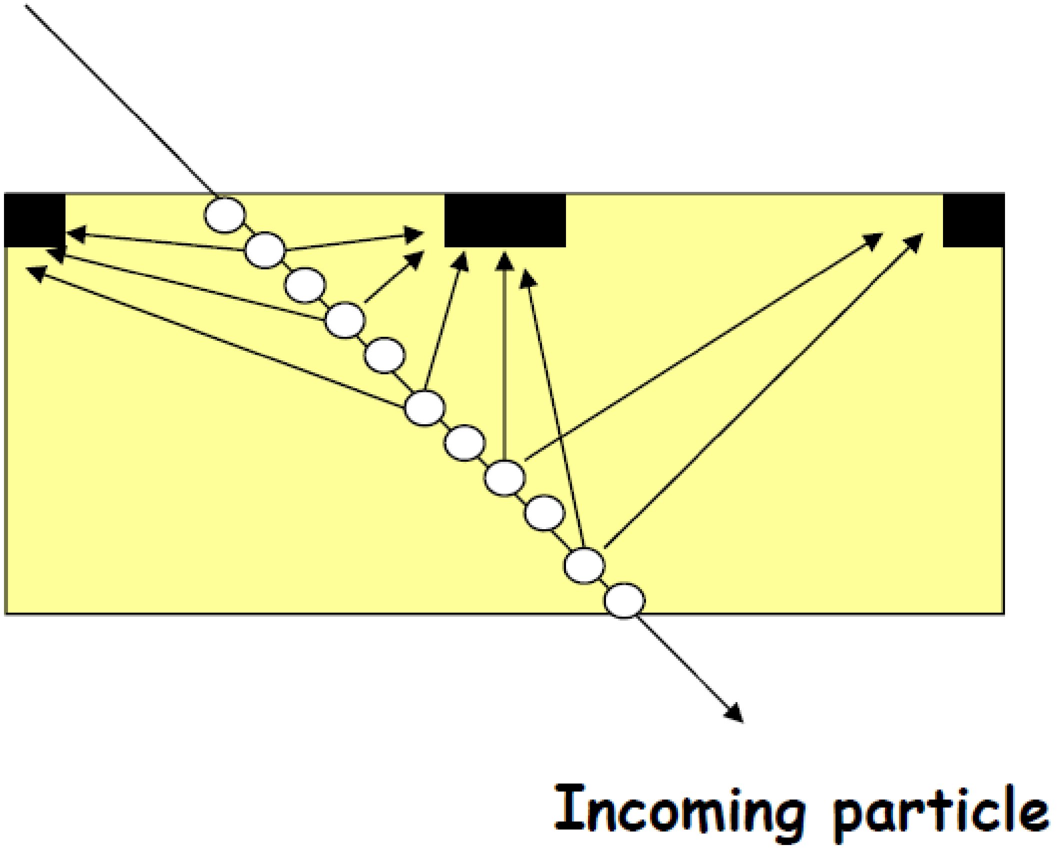
- Requires sophisticated multivariate techniques
- Series of lectures and exercises on data analysis methods later in this theme

Calibration: From bits to GeV and cm

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

Silicon Tracker calibration

- **Incoming particle creates electric charge in strips or pixels**



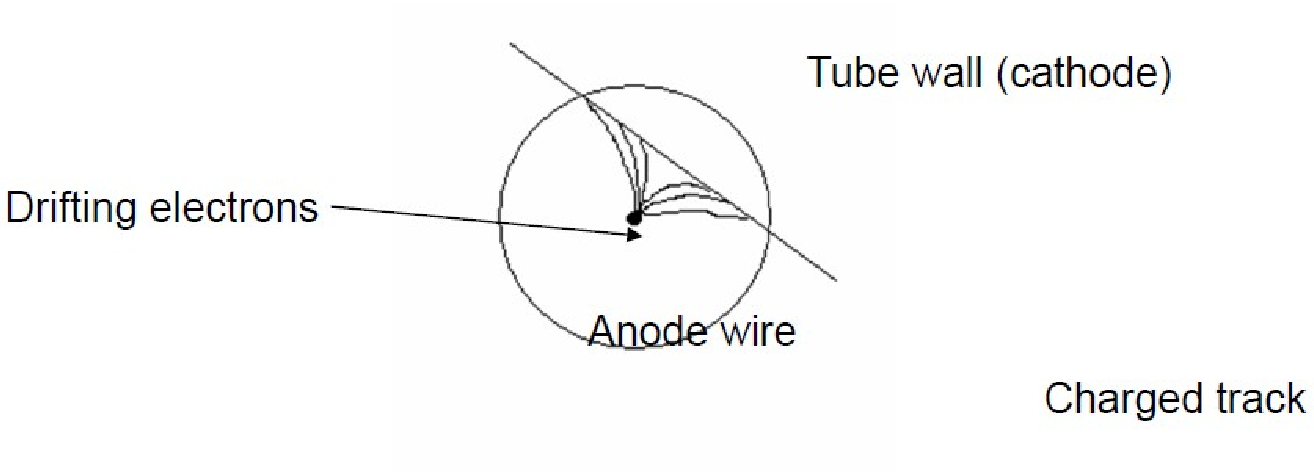
CERN School of Computing

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

CERN School of Computing

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
- Positions of detector elements need 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

CERN School of Computing

Alignment (ctd)

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

CERN School of Computing

Event reconstruction

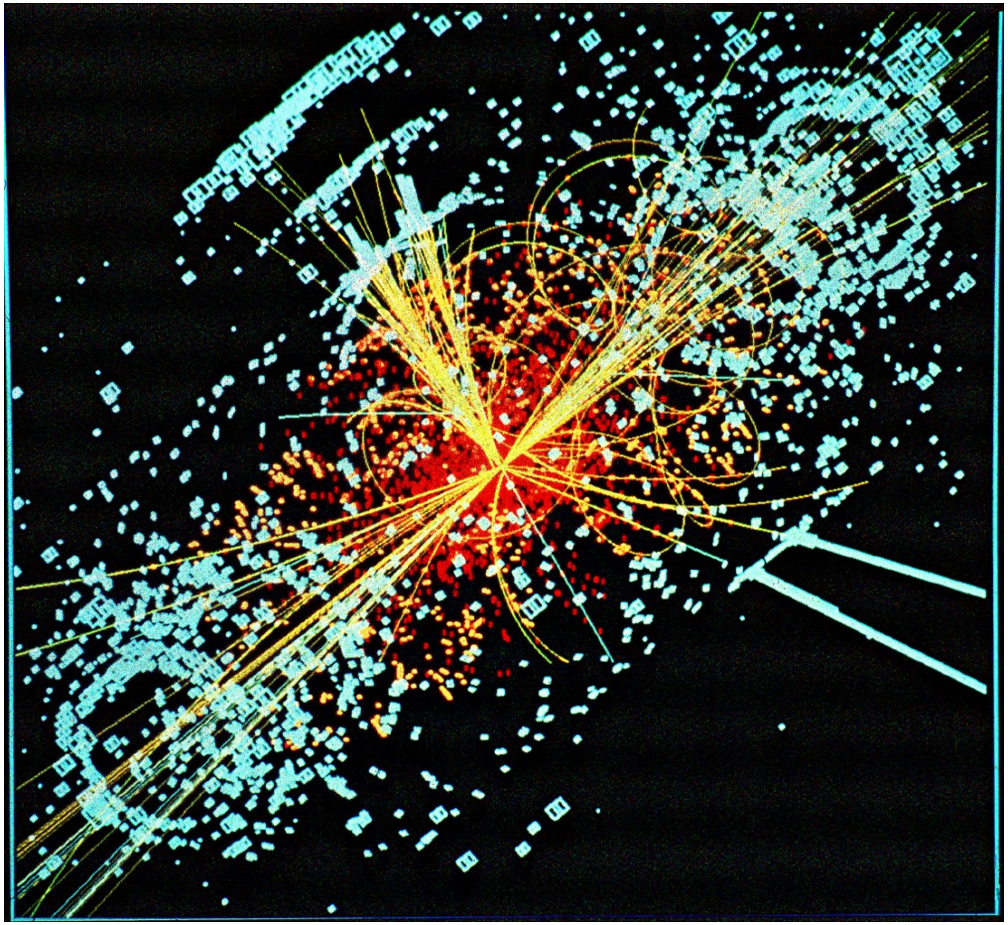
- Find out which particles have been created where and with which momentum
- Many can be observed directly
- Some are short-lived and have to be reconstructed from their decay products
- Some (neutrinos) escape without leaving any trace

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 …

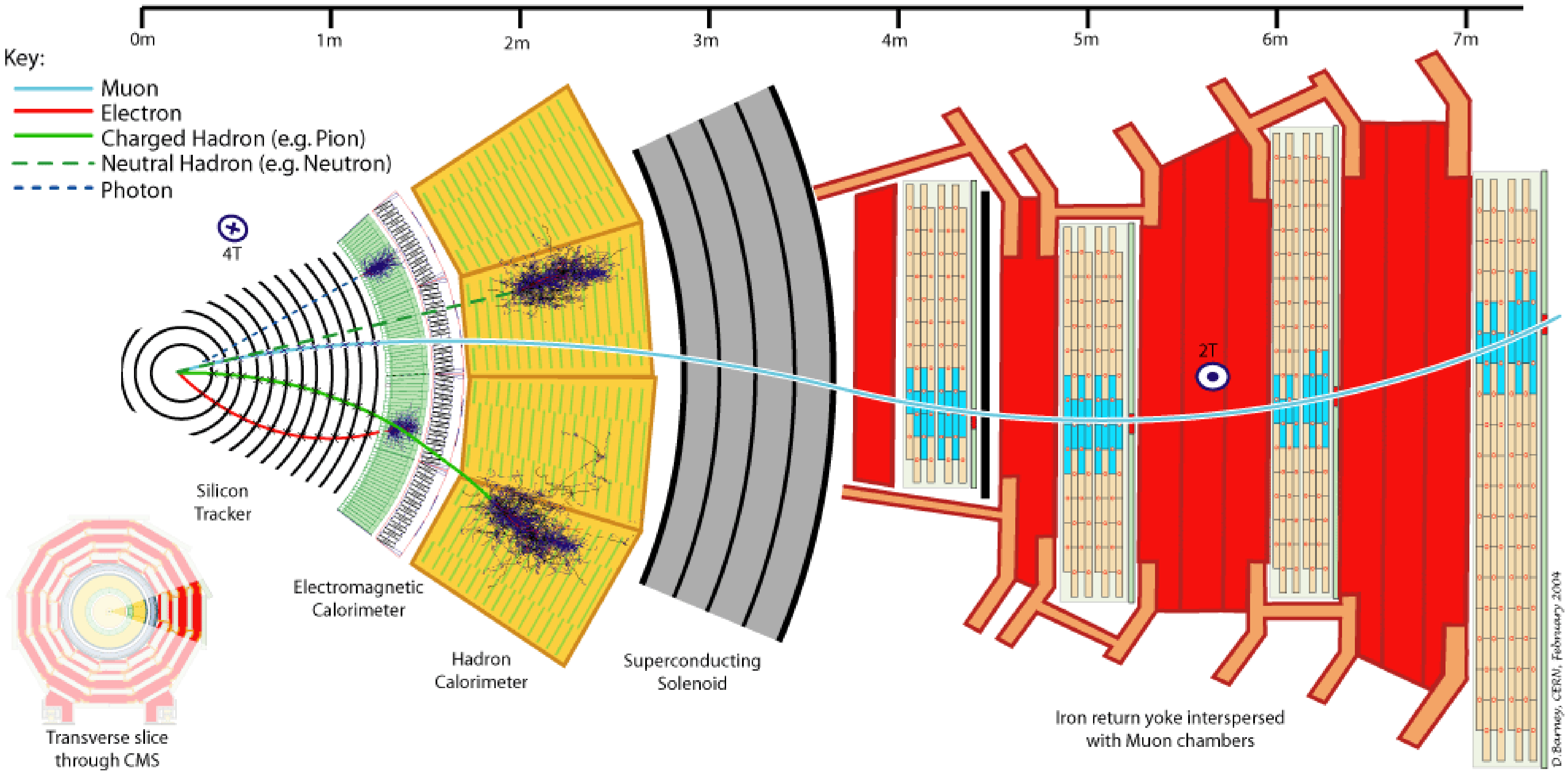
Event reconstruction (ctd)

CMS: Higgs decay into two jets



CERN School of Computing

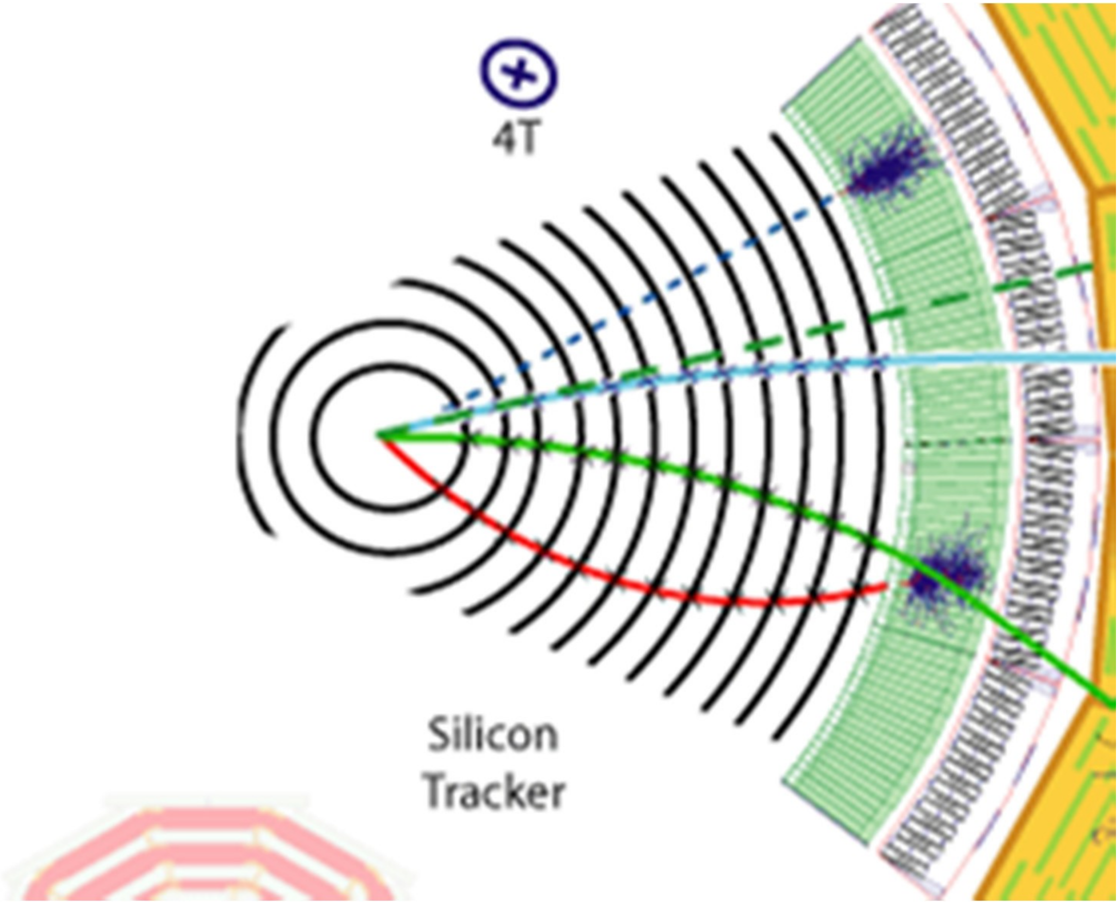
What CMS subdetectors measure



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**

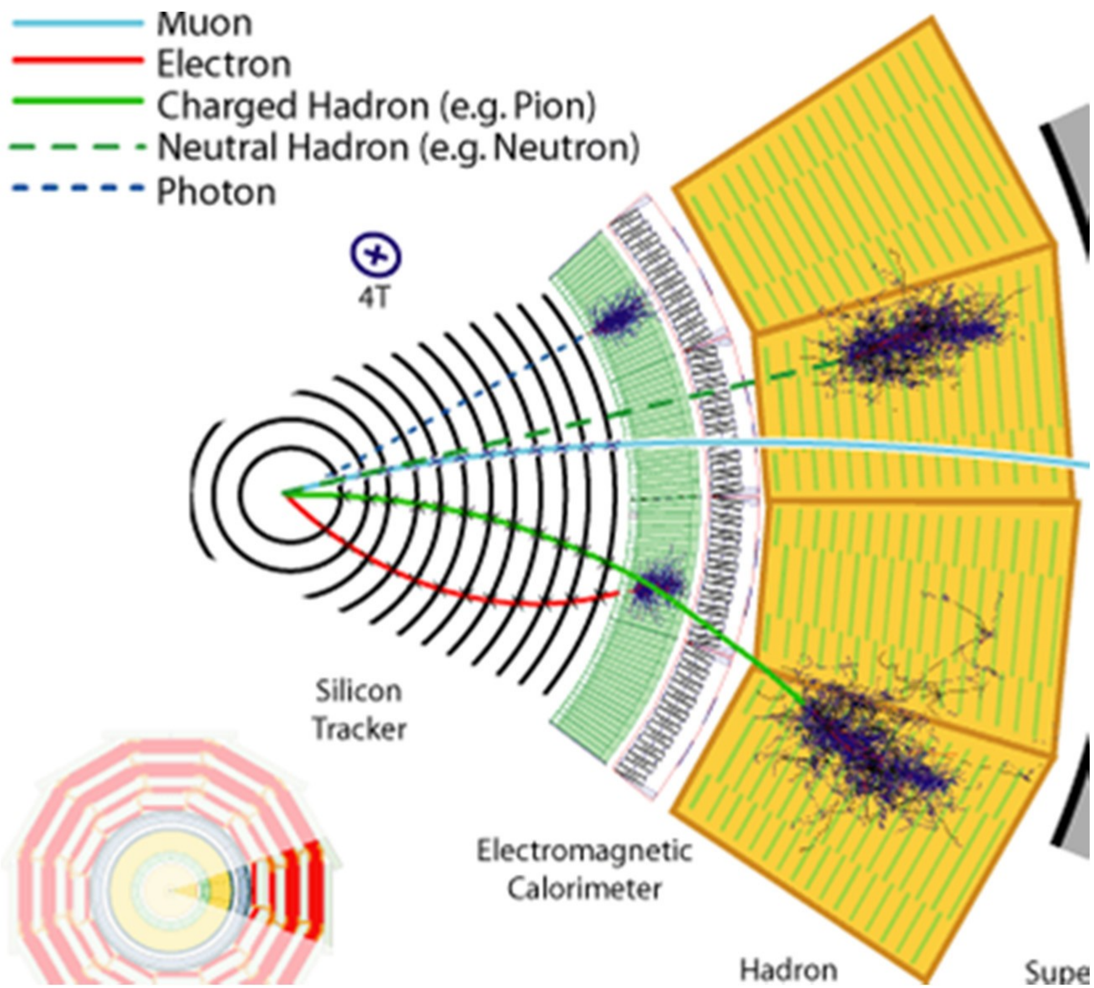
Charged particles (ctd)



Neutral particles

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

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 their estimated uncertainties from the position measurements**
- **Data compression**

CERN School of Computing

The difficulties

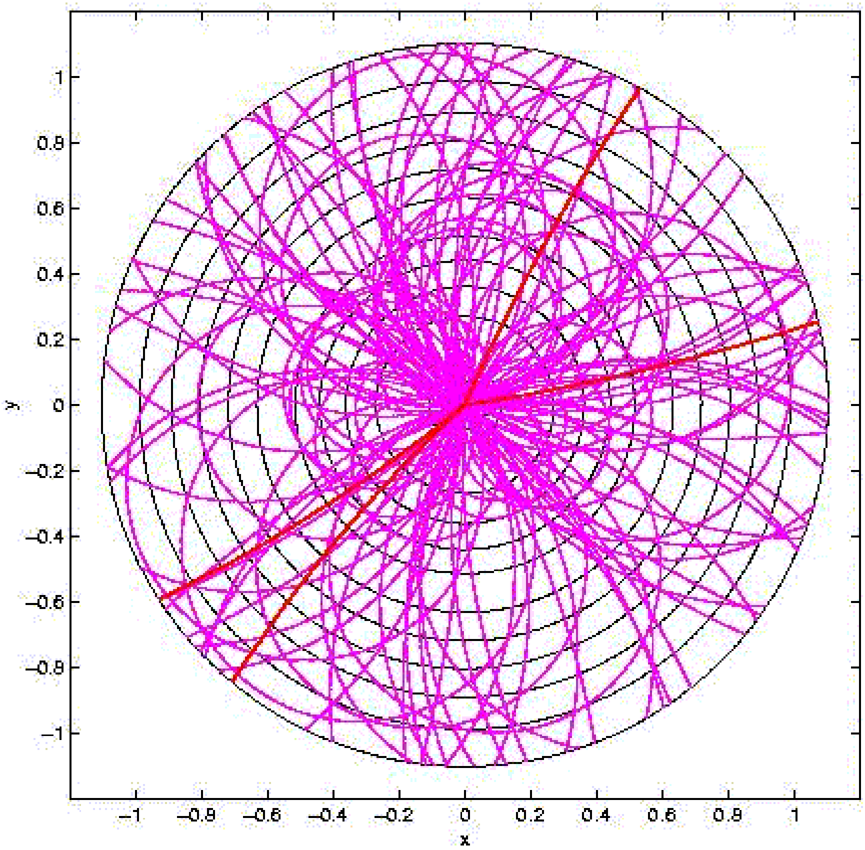
- Assignment of hits to particles is unknown
- Huge background from low-momentum tracks
- Additional background from other interactions in the same beam crossing, from adjacent beam crossings and from noise in the electronics

CERN School of Computing

More difficulties

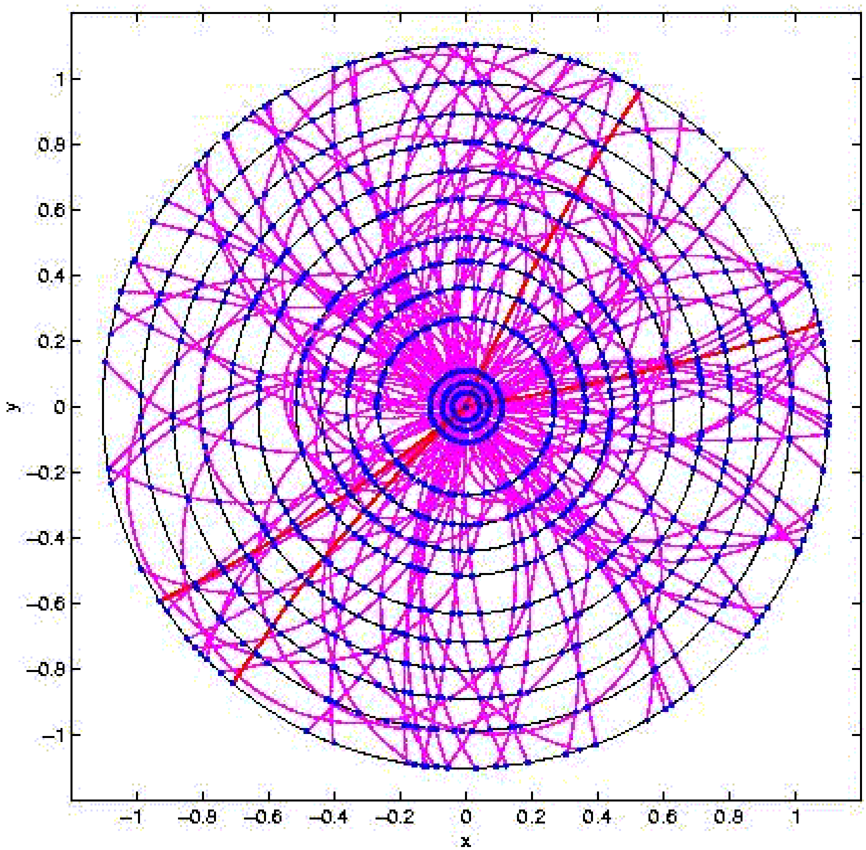
- **Charged particles interact with all the material, not only the sensitive parts**
  - **Multiple Coulomb scattering**
    - Changes direction, but not momentum
  - **Energy loss by ionization**
    - All charged particles, changes momentum
  - **Energy loss by bremsstrahlung**
    - Electrons and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

Tracks only

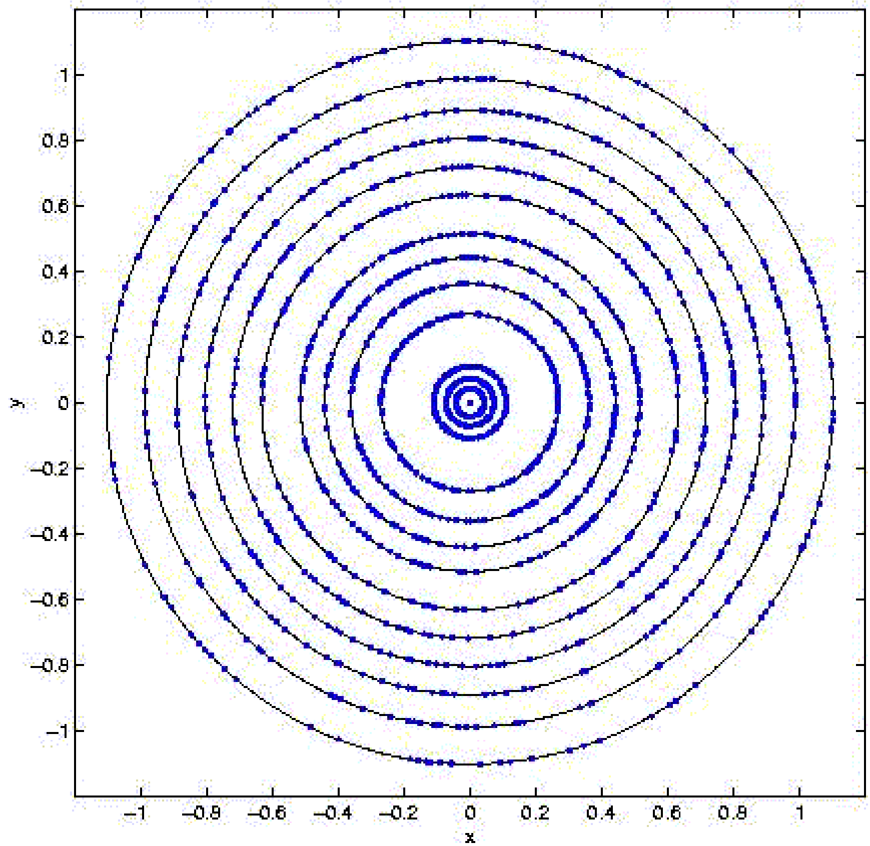


CERN School of Computing

Tracks with hits



Hits only



?

CERN School of Computing

Decomposition of the problem

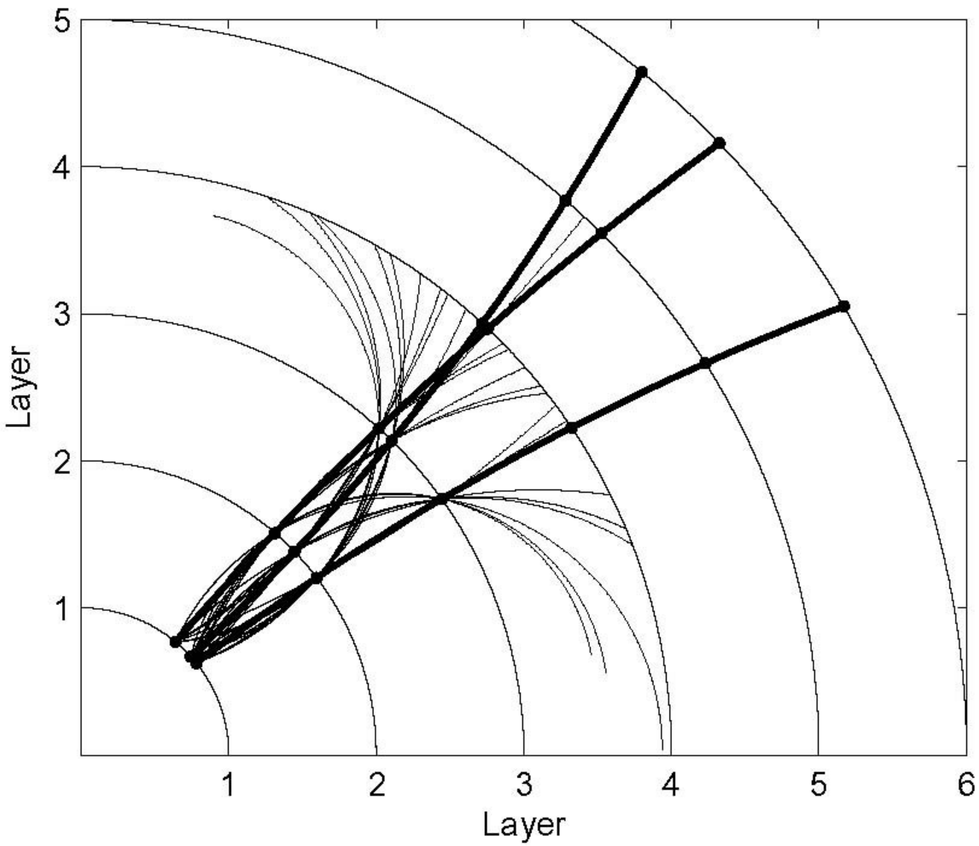
- **Pattern Recognition or Track Finding**
  - Assign detector hits to track candidates (collection of hits all believed to be created by the same particle)
- **Parameter estimation or Track Fit**
  - Determine track parameters + their estimated uncertainties (covariance matrix)
- **Test of the track hypothesis**
  - Is the track candidate the trace of a real particle?

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



CERN School of Computing

Track Fit

- **Determine (estimate) track parameters**
- **Determine uncertainties of estimated track parameters (covariance matrix)**
- **Test track hypothesis**
- **Reject outliers**
  - Distorted hits
  - Extraneous hits
  - Electronic noise hits

Ingredients

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

Estimation of track parameters

- **Most estimators minimize a least-squares objective function**
  - Linear regression
  - Kalman filter
- **Robust estimation**
  - Adaptive filter
  - Automatic suppression of outlying hits

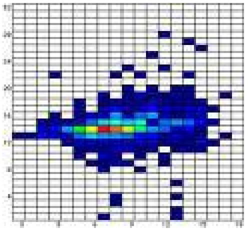
CERN School of Computing

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**



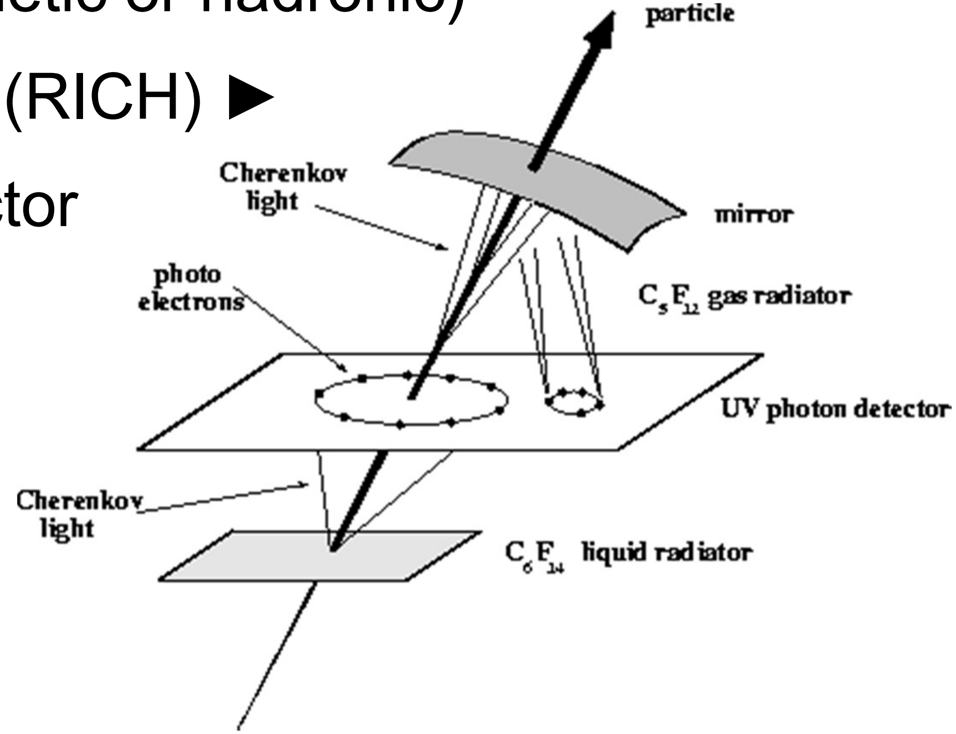
CERN School of Computing

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**
  - Calorimeter (electromagnetic or hadronic)
  - Ring imaging Cherenkov (RICH) ►
  - Transition radiation detector
  - Ionization measurements



CERN School of Computing

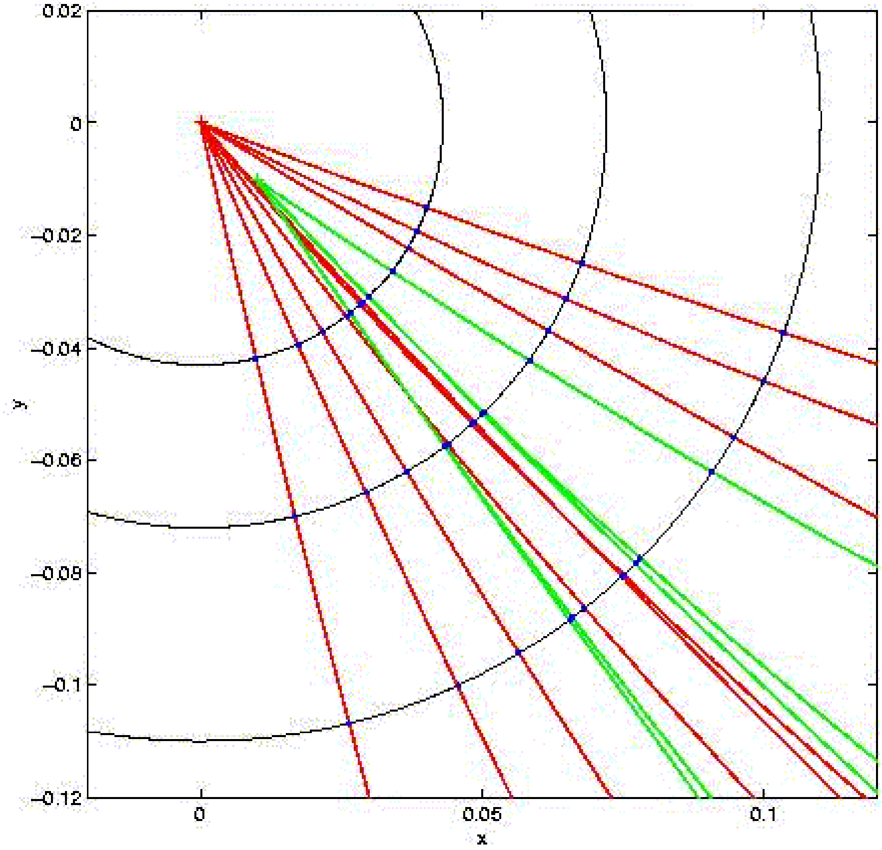
Particle identification (ctd)

- **Combining information from several detectors**
  - Shower in electromagnetic calorimeter + no matching track in tracker → photon
  - Shower in electromagnetic calorimeter + matching track in tracker → electron/positron
  - Shower in hadronic calorimeter + matching track in tracker → charged hadron
  - Track in muon system + matching track in tracker → 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 (and vice versa)**
  - Especially if decay length is small
- **Track reconstruction may be less than perfect**
  - Outliers, distortions, incorrect errors

CERN School of Computing

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**
  - Is the vertex candidate a real vertex?

Vertex finding

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

A few vertex finding algorithms

- **Hierarchical clustering**
  - Single linkage, complete linkage, …
- **Machine learning**
  - k-means, competitive learning, deterministic annealing, …
- **Estimation based**
  - robust location estimation, iterated vertex fit

CERN School of Computing

Vertex fitting

- **Most estimators minimize a least-squares objective function**
  - Linear regression
  - Kalman filter
- **Robust estimation**
  - Adaptive filter
  - Automatic suppression of outlying tracks

CERN School of Computing

Kinematical fitting

- **Impose physical constraints**
  - Momentum conservation
  - Energy conservation
- **Test mass hypotheses**
  - See whether kinematics are compatible with the decay of a certain particle
- **Reconstruct invisible particles**

Storage

- **Event reconstruction produces physics objects**
  - Tracks
  - Vertices
  - Identified particles
  - Jets
  - Tags
- **Need to be stored**

Storage (ctd)

- **Preferred tool for event data: ROOT**
- **Physics objects depend on**
  - Alignment
  - Calibration
  - Version of the reconstruction program
  - Algorithm parameters
- **Must be stored as well (database)**

Summary, event reconstruction

- **Track reconstruction**
  - Charged: determine track parameters from hits
  - Neutral: find showers in calorimeters
- **Particle identification**
- **Vertex reconstruction**
  - Determine number of production points and their positions from the set of reconstructed tracks
- **Kinematic fitting**
  - Refine estimates by e. g. imposing physical constrain

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 geometrical acceptance corrections
  - Background studies
  - Systematic error studies

CERN School of Computing

Simulation steps

- **Physics generation**
  - Generate particles according to physics of the collision
  - General-purpose and specialized generators
- **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, …
  - Store everything

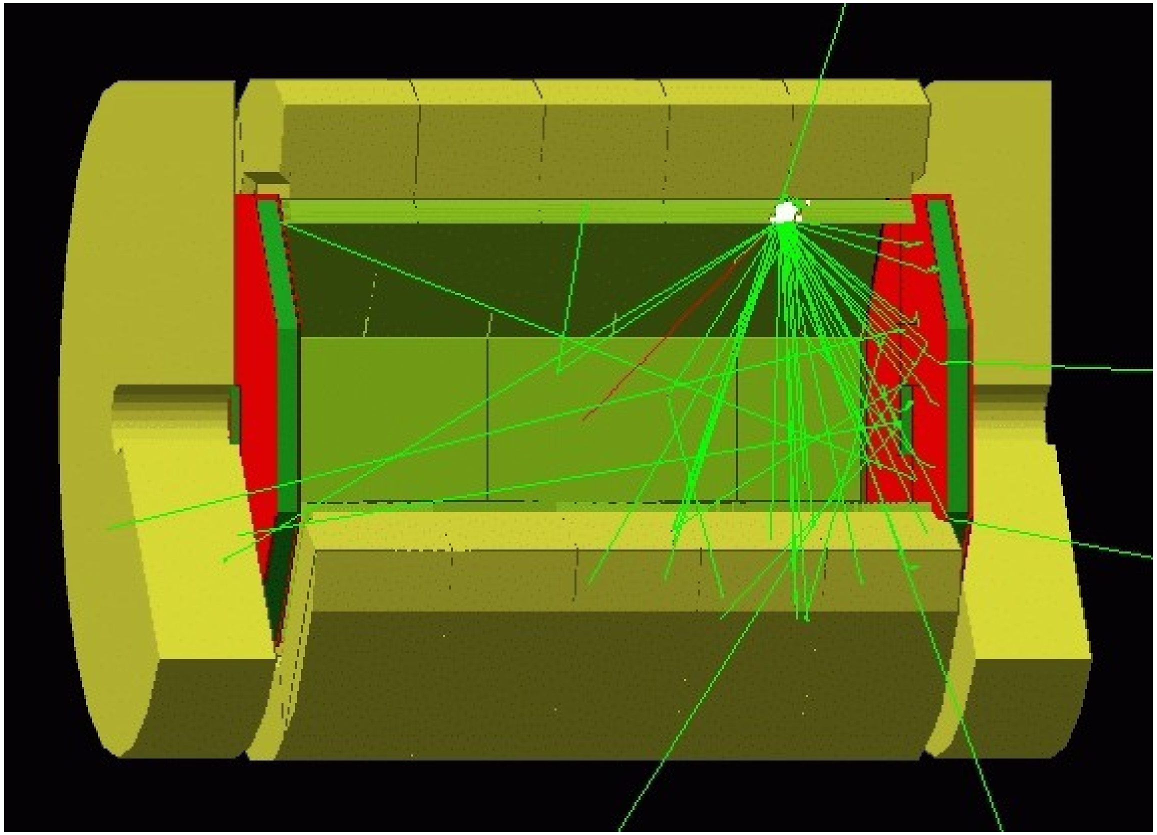
Event simulation

- **Was frequently (and still sometimes is) experiment-specific**
- **Now there is a widely used standard:**
- **GEANT4**
  - Object oriented, C++
  - Extremely general and versatile
- **Needs detailed description of the apparatus (sensitive and insensitive parts)**

Detector description

- **Geometry**
  - Partition the detector into a hierarchy of volumes
  - Describe their shape and their position relative to a mother volume
  - Use possible symmetries
- **Material**
  - Chemical composition, density
  - Physical properties: radiation length, interaction length, …

An example detector model



CERN School of Computing

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 and from real data

Physics analysis (ctd)

- **Computation of physical quantities …**
  - Cross sections, branching ratios, 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, …)

CERN School of Computing

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
- **Main tool currently used: ROOT**
  - Data analysis and storage, but also detector description, simulation, data acquisition, …

CERN School of Computing

And finally …

Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 0.9$ and 2.36 TeV

CMS Collaboration

ABSTRACT: Measurements of inclusive charged-hadron transverse-momentum and pseudorapidity distributions are presented for proton-proton collisions at $\sqrt{s} = 0.9$ and 2.36 TeV. The data were collected with the CMS detector during the LHC commissioning in December 2009. For non-single-diffractive interactions, the average charged-hadron transverse momentum is measured to be 0.46 ± 0.01 (stat.) $\pm$ 0.01 (syst.) GeV/c at 0.9 TeV and 0.50 ± 0.01 (stat.) $\pm$ 0.01 (syst.) GeV/c at 2.36 TeV, for pseudorapidities between -2.4 and $+2.4$. At these energies, the measured pseudorapidity densities in the central region, $dN_{\mathrm{ch}}/d\eta|_{|\eta|<0.5}$, are 3.48 ± 0.02 (stat.) $\pm$ 0.13 (syst.) and 4.47 ± 0.04 (stat.) $\pm$ 0.16 (syst.), respectively. The results at 0.9 TeV are in agreement with previous measurements and confirm the expectation of near equal hadron production in $\mathrm{p\bar{p}}$ and pp collisions. The results at 2.36 TeV represent the highest-energy measurements at a particle collider to date.

KEYWORDS: Hadron-Hadron Scattering

ARXIV EPRINT: 1002.0621

JHEP02(2010)041

Distributed analysis

- **Physics analysis takes 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

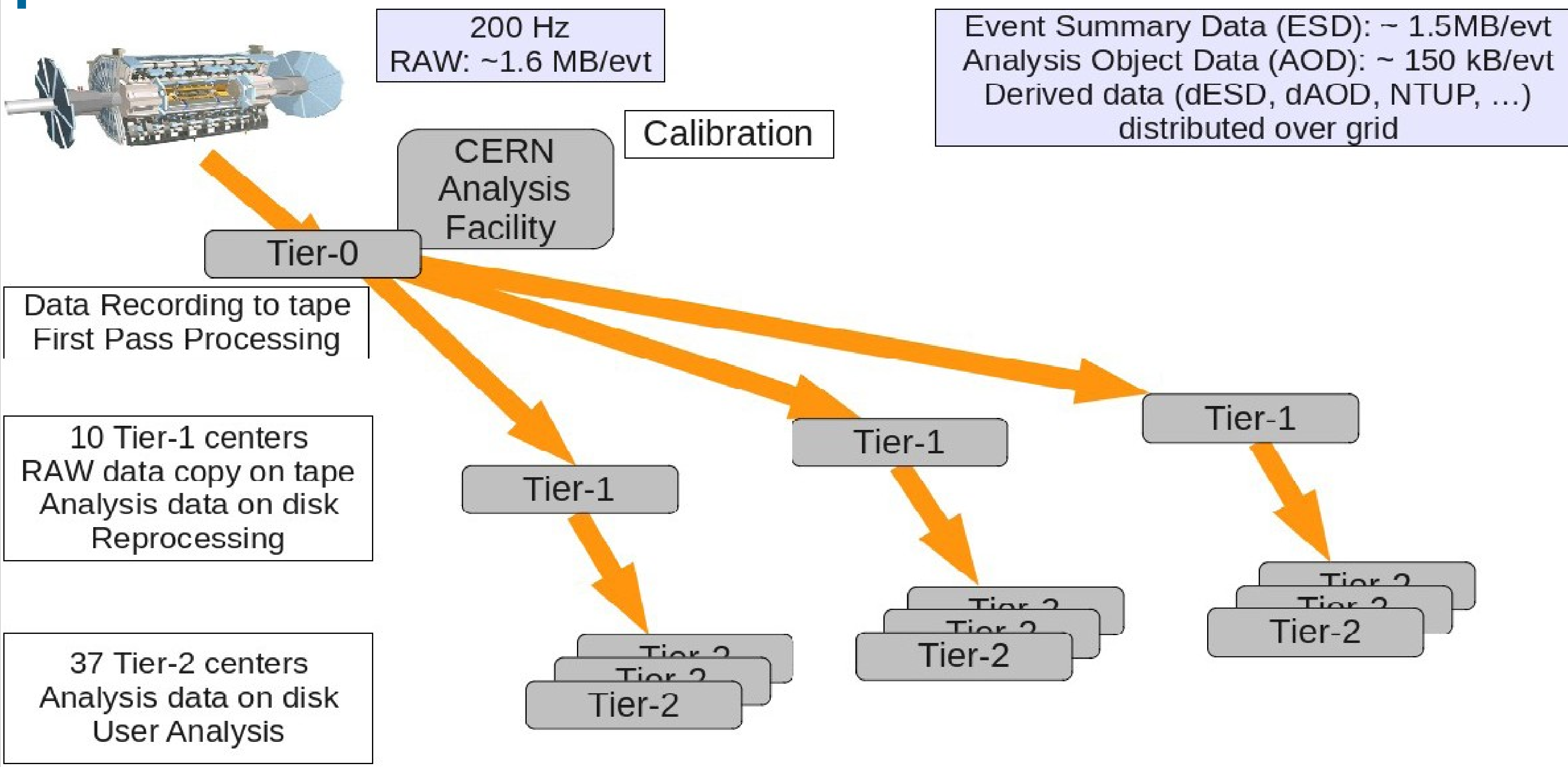
- Global collaboration of more than 170 computing centers in 36 countries
- Four-tiered model
- Data storage and analysis infrastructure
- $O(10^5)$ CPUs
- O(100) PByte disk storage (tiers 0 and 1)

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

Data flow in ATLAS

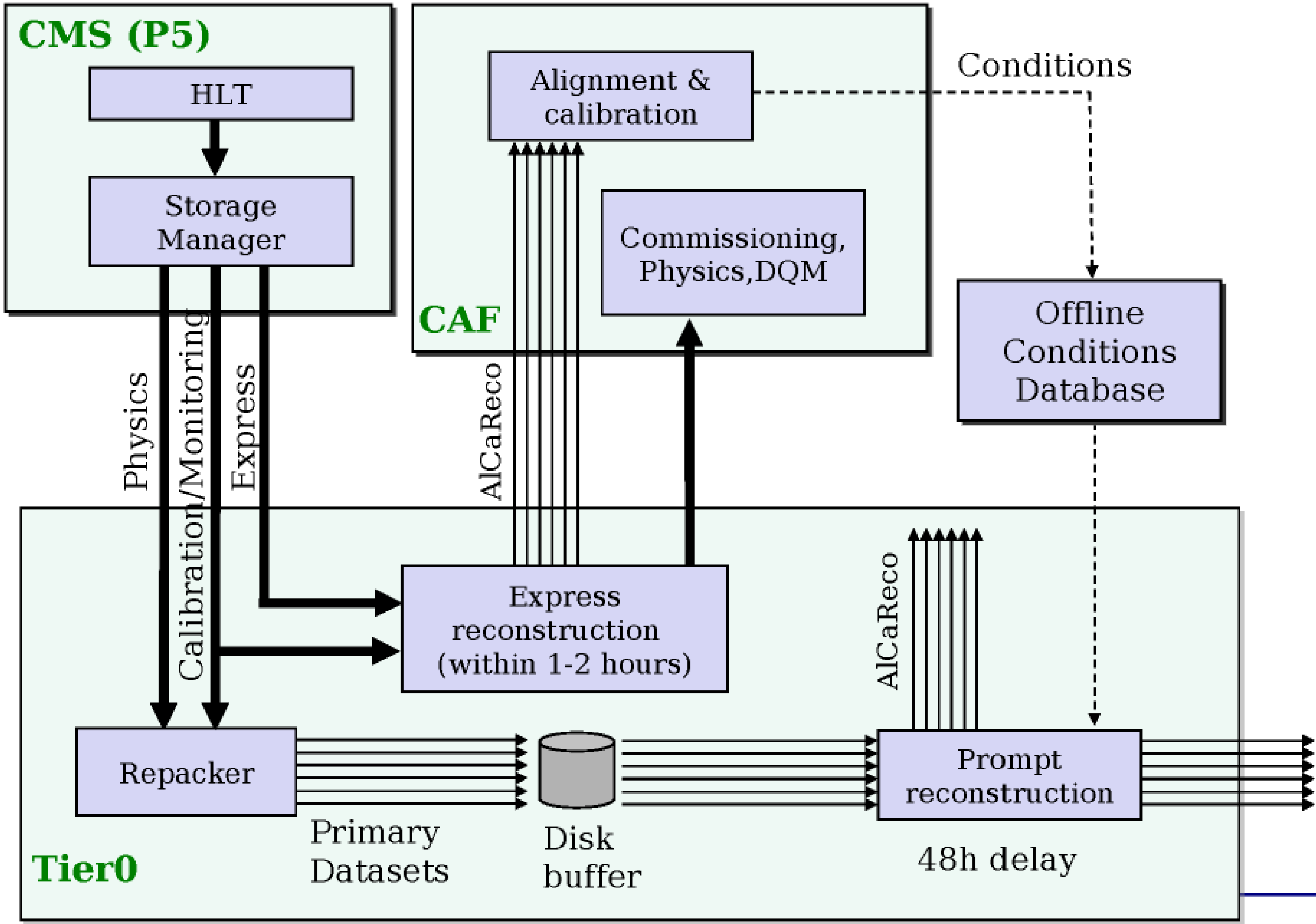


CERN School of Computing

Additional resources

- **CAF (CERN Analysis Facility)**
  - O(100) worker nodes, O(1000) cores (CMS)
  - Ready access to calibration and express streams
  - Fast turnaround
  - Operation critical tasks
    - trigger and detector diagnostics
    - alignment and calibration
    - monitoring and performance analysis
  - Physics data quality monitoring

Data flow in CMS-CAF



Summary

- **Physics computing involves:**
  - Event filtering with multi-level trigger
  - Storage of raw data
  - Calibration and alignment
  - Storage of calibration and alignment data
  - Event reconstruction
  - Storage of reconstruction objects and metadata

CERN School of Computing

Summary (ctd)

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

CERN School of Computing

Summary (ctd)

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

CERN School of Computing

Acknowledgments

- All illustrations from the ATLAS and CMS experiments are subject to © 2019 CERN (CERN Copyright)
- Thanks in particular to the ATLAS TDAQ community for providing facts and figures describing the ATLAS triggering system
- Thanks to Are Strandlie for his lecture at CSC 2013