Introduction to Physics Computing

Arnulf Quadt
CERN School of Computing 2019

Babeș-Bolyai University,
Politehnica University of Bucharest,
Cluj-Napoca, Romania

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

Powers of Ten

• Goal: understand fundamental structures and forces

\[ 10^{-15} \text{m} \quad 1 \text{m} \quad 10^{-6} \text{m} \quad 10^{3} \text{m} \quad 10^{0} \text{m} \]
### Fundamental structures & forces

- **From largest to smallest dimensions**
- **Reduction principle**
  - few fundamental building blocks
  - few fundamental forces

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

### Rutherford scattering

- **Need particles source & detectors**

### (sub) structure - atoms

<table>
<thead>
<tr>
<th>Periodensystem der Elemente</th>
<th>Periodensystem der Elemente</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 H</td>
<td>2 Li Be</td>
</tr>
<tr>
<td>11 Na Mg</td>
<td>12 Al Si P S Cl Ar</td>
</tr>
<tr>
<td>13 K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr</td>
<td></td>
</tr>
<tr>
<td>14 Mg Si S Cl Ar</td>
<td></td>
</tr>
<tr>
<td>15 P S Cl Ar</td>
<td></td>
</tr>
<tr>
<td>16 S Cl Ar</td>
<td></td>
</tr>
</tbody>
</table>

### Elementary building blocks

- ‘LEGO bricks’ of Spin-1/2 quarks & leptons + antiparticles

- 1995
- 2000
- 2012
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

What is dark matter?
...
Sun flowers

... look rotationally and mirror symmetric ....
A closer look

Examples for symmetry breaking

13 spirals cw,
6 spirals counter-cw!

34 spirals cw,
13 spirals counter-cw!

34 spirals cw,
21 spirals counter-cw!

21 spirals cw,
13 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 \( \Rightarrow \) massive Higgs-boson

History of the universe

Big bang in laboratory

Particle Accelerators
Particle Accelerators I

In your TV set, the electrons are accelerated to 20000 volts. In LEP, they are accelerated to 100 000 000 000 volts.

Particle Accelerators II

Heisenberg uncertainty relation:
\[ \Delta x \cdot \Delta p \geq \hbar \]

Einstein's equation:
\[ E = mc^2 \]

Linear accelerator:

Repeated acceleration:
- \(10^3\) eV = \(1\) keV
- \(10^6\) eV = \(1\) MeV
- \(10^9\) eV = \(1\) GeV
- \(10^{12}\) eV = \(1\) TeV

Tevatron' = TeV beam energy

Particle Accelerators III

Fermilab
- 1987 - 2011
- Tevatron PP
- 1.8 - 1.96 TeV
- CDF, DØ

CERN
- 1981 - 1990
- SPS FR
- 320 GeV
- H1, ZEUS

CERN
- 2008 - 2009
- LHC PP
- 14 TeV
- ATLAS, CMS, ALICE, LHC-B

Particle Accelerators IV

LEP/LHC

DESY
- HERA eP
- 320 GeV
- H1, ZEUS

CERN
- 1981 - 1990
- SPS FR
- 320 GeV
- H1, ZEUS

CERN
- 2008 - 2009
- LHC PP
- 14 TeV
- ATLAS, CMS, ALICE, LHC-B

DESY
- HERA eP
- 320 GeV
- H1, ZEUS

CERN
- 1981 - 1990
- SPS FR
- 320 GeV
- H1, ZEUS

CERN
- 2008 - 2009
- LHC PP
- 14 TeV
- ATLAS, CMS, ALICE, LHC-B
Large Hadron Collider - LHC

- CERN: europa.centre for particle physics
- Founded 1954
- LHC: PP collider
- High energies: \( \sqrt{s} = 7 \) \((14)\) TeV
- 40 Mio. collisions / sec
- 1st collisions in Nov. 2009
- Physics at 7 TeV since 31.3.2010
- Phys. at 13 TeV since 20.5.2015
- 4 Expts:
  - ATLAS, CMS, ALICE, LHC-B

Proton-Proton Collisions

- Large Hadron Collider @ CERN, Switzerland
  - pp collider at \( \sqrt{s} = 14 \) TeV
  - Luminosity \( L = 10^{34} \text{ cm}^{-2}\text{s}^{-1} \)

Detectors / Experiments

Caverns 100m underground
Built like a bottle ship

ATLAS and CMS experiments

**Built like a bottle ship**

**ATLAS and CMS experiments**

**Date rate and size**

<table>
<thead>
<tr>
<th>Rate</th>
<th>RAW</th>
<th>ESD</th>
<th>AOD</th>
<th>Monte Carlo</th>
<th>Monte Carlo % of real</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hz</td>
<td>MB</td>
<td>MB</td>
<td>kB</td>
<td>MB/evt</td>
<td><em>percent</em></td>
</tr>
<tr>
<td>ALICE HI</td>
<td>100</td>
<td>12.5</td>
<td>2.5</td>
<td>250</td>
<td>100</td>
</tr>
<tr>
<td>ALICE pp</td>
<td>100</td>
<td>1</td>
<td>0.04</td>
<td>4</td>
<td>0.4</td>
</tr>
<tr>
<td>ATLAS</td>
<td>200</td>
<td>1.6</td>
<td>0.5</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>CMS</td>
<td>150</td>
<td>1.5</td>
<td>0.25</td>
<td>50</td>
<td>2</td>
</tr>
<tr>
<td>LHCb</td>
<td>2000</td>
<td>0.025</td>
<td>0.025</td>
<td>22</td>
<td>22 DVDs per sec.</td>
</tr>
</tbody>
</table>

**Expected pictures: Higgs decay**
Search for new physics

Search for needle in the hay stack

Physics

Expectations and measurements
**Higgs production at the LHC**

- **H→γγ**:
  - Rare channel
  - Best for low Higgs masses

- **H→WW(*)**:
  - γννν: very important for intermediate masses
  - lνq: high rate, important at high mass

- **H→ZZ(*)**:
  - 4l: golden channel
  - llντ: 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**

<table>
<thead>
<tr>
<th>MH [GeV]</th>
<th>WW→lνlν</th>
<th>ZZ→4l</th>
<th>γγ</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>127</td>
<td>1.5</td>
<td>43</td>
</tr>
<tr>
<td>150</td>
<td>390</td>
<td>4.6</td>
<td>16</td>
</tr>
<tr>
<td>300</td>
<td>89</td>
<td>3.8</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Higgs → γγ**

Inclusive/Weighted Mass Spectra

- Excesses visible in the inclusive mass spectra
- CMS has a slightly wider fit range

**ATLAS**

- Excess of 4.5σ

**CMS**

- Excess of 4.1σ
Higgs production at the LHC

- 4e Candidate
  - $M_{Z_1} = 92 \text{ GeV}/c^2$
  - $M_{Z_2} = 27 \text{ GeV}/c^2$
  - $M_{4\ell} = 126 \text{ GeV}/c^2$

- 4µ Candidate
  - $M_{Z_1} = 90 \text{ GeV}/c^2$
  - $M_{Z_2} = 25 \text{ GeV}/c^2$
  - $M_{4\mu} = 119 \text{ GeV}/c^2$

Somewhat better S/B in CMS
(for instance lower reducible background)
(also 7 TeV ATLAS analysis does not have latest improvements in electron reconstruction)
**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

**Higgs combination (4th July 2012)**

ATLAS Preliminary

- Obs.
- Exp.
- ±1σ
- ±2σ

**As a Layman: We have it!**
The technical challenge at LHC

- 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

- 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

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

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

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

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
Complexity of Data acquisition

- 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, ...

Online processing (ctd)

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

ATLAS detector

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, calorimeter
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

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
**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)) software-based
  - 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

---

**Towards High-Lumi LHC**

- 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**

- Energy: 7–8 TeV
  - Peak lumin: 0.8e34 cm^-2 s^-1
  - L1 peak rate: 70 kHz
  - HLT av. rate: 400 Hz

- Energy: 13 TeV
  - Peak lumin: 2.2e34 cm^-2 s^-1
  - L1 peak rate: 100 kHz
  - HLT av. rate: 1 kHz

- Energy: 14 TeV
  - Peak lumin: 2.0e34 cm^-2 s^-1
  - L1 peak rate: 100 kHz
  - HLT av. rate: 1 kHz

- Energy: 14 TeV
  - Peak lumin: 7.5e34 cm^-2 s^-1
  - L1 peak rate: 1 MHz
  - HLT av. rate: 10 kHz

---

**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

- 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

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

Run 347, Event 2566
Higgs candidate

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

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
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 $\sim 10^7$ 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
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

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

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
- Positions of detector elements need to be known to a similar or better precision

Example: CMS tracker

Wow, I will have to realign this…

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 in a database

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 …

CMS: Higgs decay into two jets
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

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

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

Tracks with hits

Hits only

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

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

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

Particle identification (ctd)

- Combining information from several detectors
  - Shower in electromagnetic calorimeter + no matching track in tracker $\rightarrow$ photon
  - Shower in electromagnetic 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
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

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
Vertex fitting

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

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

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

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, …)

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, …

And finally …

- 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
- $O(10^9)$ 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 $\rightarrow$ Tier-1 $\rightarrow$ Tier-2
- Data access and storage
  - Long-term storage, direct access

Data flow in ATLAS

- 200 Hz
  - RAW: ~1.6 MB/evt
- Event Summary Data (ESD): ~1.5MB/evt
- Analysis Object Data (AOD): ~150 kB/evt
- Derived data (dE/dx, dAOD, NTUP, ...)
- Distributed over grid

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
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

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
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