Introduction to ATLAS

- part 1: ATLAS Detector (and LHC)
- part 2: Physics programme in ATLAS
- part 3: Event Reconstruction and Physics Performance
- part 4: Physicists' tools analyses in ATLAS



Wolfgang Liebig





Lecture Notes

• Wiki of IFT-UiB (Meetings & Seminars)

– https://wikihost.uib.no/ift/index.php/Particle_Physics_group



Part 4: Data and Physicists' Tools

- Collaboration structure
 - who are we? How are physics results discussed and agreed upon?
- Data formats
 - from RAW to Ntuples
- Offline software
- What is in the data?
 once more looking at tracks, jets...
- Meta-data: data about the detector
 as opposed to data from the detector

Essays





ATLAS Collaboration



Norway in ATLAS

Number of authors and institutions by country

- Two large groups:
 Bergen and Oslo
 - 2005: # authors can compete with larger countries
- Activities in Bergen
 - SCT construction and maintainance/op.
 - IBL upgrade project
 - $-\mu$ and τ performance
 - mainly B-physics and SUSY searches
 - top, higgs physics





Organising 4000 Members

- Internal rules and management structure defined
- Influence by members and institutes
 - through collaboration
 board: 174 seats,
 - 1 institute = 1 vote
 - other means, like candidate proposals
- Collaboration board
 - elects most management positions
 - votes on changes to rules
- Further 'democratic' structures
 - limitation of management
 Mandates to 2 (+2) years



Look at structure of physics groups

Internal and Public Information

- In the past everything that ATLAS did used to be world-readable – Plots, write-up, software, bugs, news
- With beginning of data-taking ATLAS has put information protection into place
- Internal plots and write-ups strictly separated from ATLAS-labeled official results
 - Example: to non-ATLAS members I am allowed to show only 'approved' plots, i.e. ATLAS, ATLAS_Preliminary label
 - exceptions made for students and funding applications
 - aim is to give coherent message to the planet while maintaining free discussion culture inside
 - indeed the LHC experiments have become very popular among colleagues, media, interested people
- Authentification mechanisms in place by CERN and GRID to identify ATLAS members or (prospective) authors

 internal distinction made between member and active physicist



So how is one to write a paper/conference talk/public plot in ATLAS?

Outline of Publication Procedure

Initiation (or late delivery of widely expected result)

- physicists present their work to their Activity WG
- open questions addressed
- agreement between competing analyses
- Activity WG (conveners) agree it should be published

Editorial Board

- WG conveners ask management for EB
- EB=Editorial Board of indep. experts for
 - reviewing and signing off paper/note

Circulation in ATLAS

- EB discusses with authors and agrees to circulate
- announced to all active physicists inviting to comment
- deadline of ~2 weeks

Second iteration possible

Comments and Seminar

- number of people commenting can be $\sim 5...50$
- comments need to be answered
- answers followed also by EB
- seminar given on plenary meeting to present result and address major comments in person

Publication

- final round and signed off by EB
- comments and sign-off by physics coord. and publication committee

- published on ATLAS page, arXiv

2011

part 4: Data and – another round of editing/review

Journal

8

Activity Working Groups in ATLAS







– etc

2011

Working with ATLAS Data



Athena Framework and Releases

• Why a framework?

- runs at the right time
- with the right input data
- using the correct conditions data
- correctly writes the results to disk
- and many more features
- All the offline processing work is done in athena
 - (combined) reconstruction as in 3rd lecture
 - ATLAS simulation
 - preparation of physics analysis data (or even full analysis)
 - user friendliness often was an issue, greatly improved
- Underlying programming language is C++
 – python for configuration scripts

Minor version, to distinguish special purposes – p–p data, trigger, heavy ion

Releases

- regular major releases
- defined by a list of packages (SVN) and their versions
- now open: 17.0.6, 17.2.0, 17.5.0





Slide from J. Catmore **Athena Components Scheme**



PAT

Athena Components I

- Algorithm: an application a piece of code that "does something"
 - All algorithms inherit from the Algorithm class, which contains three methods:
 - Initialize() run once at the start
 - Execute() run n times
 - Finalize() run once at the end
 - Algorithms are invoked centrally by the framework
 - Many algorithms can be run in a single job one after the other
- Data object: result of an algorithm, or the input to it
 - E.g. Track, Cluster, Muon, Electron, McEvent
- Service: globally available software entity which performs some common task
 - Message printing
 - Histogram drawing
- Event: a single pass of the execute() method, roughly corresponding to a physics event
- **JobOptions**: Python script which passes user instructions to Athena
 - Which algorithms to run, what order, configuration
 - Control of number of cycles, input/output files, runtime variables etc

Athena Components II

- **Tool**: piece of code that is shared between algorithms it can be executed as many times as you need in the execute() method of your algorithms
- Auditors: software which monitors the other components of the framework
- **Sequence**: execution order of the algorithms
- Filters: software which allows or forbids an event from passing to the next algorithm in the sequence or being written to disk
- **Transient Store (StoreGate)**: service which stores results of algorithms (data objects) and passes them to the next algorithm.
 - The data is held in the computer memory
- Persistent Store (POOL): format in which the data objects are written to disk
- Converter: software which enables the data objects used in the code to be written to and read from POOL without the details of the persistency being included in the objects themselves

ATLAS Data Formats

Monte Carlo

- Event generator output
- Digits/RAW
 - Simulation/detector output
- Event Summary Data (ESD)
 - Output of reconstruction
- Analysis Object Data (AOD)
 - Summary of reconstruction primary analysis data
- Tag
 - Thumbnail of each event used for identifying interesting events at the analysis stage
- dESD, dAOD
 - Data derived from ESD or AOD









ATLAS Data Processing Chain PAT



 Will use Tracking as example how reconstruction is integrated in processing chain

16

2011

Physics Analysis Concept

- Physics analysis can be done:
 - In directly with Athena to produce n-tuples readable by ROOT
 - such as with the D3PD-making tools (Thurs.)
 - directly with ROOT using Athena-ROOT-Access then your AOD, ESD etc is your n-tuple
- Analysis can be done in either framework
 - on every event
 - steered by the tag selection mechanism, thereby only selecting certain events

ılı

Processing Chain for Tracking



(•••••• : for commissioning phase or identified leptons)

- C++, but purposefully NOT object-oriented
 - data-objects and algorithms are strictly separated
 - several strong design reasons:
 cpu time, track conversion to disk, track constness
 - track user-friendlyness vs. flexibility to develop algorithms and use custom-configured algorithms



How does `the ATLAS Track class' look like?



ATLAS New Tracking



The TrackParticle

- The *TrackParticle* marks the boundary between tracking and analysis
- Remember: a TrackParticle is a smaller(*) version of a Track



I will only cover the tracking interface since the "AOD" interface is the same for all AOD objects ...

Athena

The latter provides a so called "kinematic interface" and gives you access to kinematic variables (p, E).

It also assures that one can iterate over AOD objects of different types.

(*) in terms of disk space

2011

Muons

- Muon on AOD contains merger of all identification algorithms
 - best identification defines kinematics
 - results from other id methods kept
 - overlap detected via shared hits

Priority for defining best muon

- 1. combined muon (ID+MS+CB track)
- segment-tagged muon (ST are more complex in reality)
- 3. stand-alone MS-only muon
- 4. calorimeter-tagged muon





Electrons/Photon Physics Object

- The 'ElectronAODCollection' and 'PhotonAODCollection' contain e/gamma objects with:
 - Author to determine which algorithm used
 - Four-momentum, with errors
 - CaloCluster
 - ★ calorimeter cluster information

Links to TrackParticles (if any)

- * tracking information for when interpreted as an electron
- ★ trackParticles are sorted so that those with silicon hits are before those without, and within the category, in deltaR/deltaφ.
- * the "best" trackParticle should be the first

Links to VxCandidates (if any)

- conversion vertex information for when interpreting as converted photons
- vertices are sorted so that those with two tracks are before those with one track, and within the categories, by conversion radius.

23

- egDetail additional reconstructed information
- Particle IDentification (and Object Quality, from rel 16)

Jet Content on AOD

A Jet has 4 main features :

- A 4-momentum (IParticle) : provide all kinematic functions (e(), mass(), rapidity(), etc...)
- A list of constituents (before calib : $P_{\rm Jet} = \sum P_{\rm constituents}$), accessible from the jet
- Some attached "moments" and "associations" (to other particles)
 - B-tagging information
 - Energy per calorimeter sampling
 - jet width, ...
- "Calibration states". Jets also hold different momentums user can switch on/off
 - EMSCALE , CONSTITUENTSCALE, FINALSCALE(default, calibrated scale)



Jets in ESD and AOD

Jets are stored in **JetCollection** object in StoreGate.

Limited number of JetCollection are kept in ESD/AOD "AntiKt4TruthJets", "AntiKt4TowerJets", "AntiKt4TopoJets", Also in autumn 2010 reprocessing : "AntiKt4TopoEMJets" (might be a temporary collection) (+ same for AntiKt6, +pile-up truth jets)

Signification: (algotype)(main parameter)(input type)

- AntiKt = AntiKt algorithm
- 4 = size parameter is 0.4 in (η, Φ) plan
- Topo = Input to jet finding is topo clusters ("LCTopo" : inputs are Localy Calibrated clusters)

The "TopoEMJets" collections are identical to "TopoJets", except their default calibration is only a Scale factor (no hadronic weighting). 25

Tau-Jets on AOD

- Stored in TauJetContainer
 - Vector of the TauJet object
- TauJet object
 - Inherits from :
 - ParticleBase (info about origin and charge of particle)
 - P4ImplEEtaPhiM(from I4Momentum, 4-momentum of particle)
 - Contian tau-specific information:
 - Vector of links to TrackParticle's associated to tau candidate
 - Link to a jet (for Anti-kt jet tauRec seed)
 - CaloCluster with cells selected for discriminating variables/energy building
 - Pointer to TauPID object (results of identification algorithms)
 - Vector of links to TauDetails (extra information)

Information	Container	Base class	Availability
Basic	TauRecContainer	TauJet	ESD/AOD
Details	TauRecDetailsContainer	TauCommonDetails	ESD/AOD
ExtraDetails	TauRecExtraDetailsContainer	TauCommonExtraDetails	ESD only

Trigger

- access to trigger information facilitated by 2 tools
- TrigDecisionTool
 - for event level trigger information
 - holds LVL1 decision before/after dead-time veto
 - holds HLT decision
 - access to (potentially) lumi-block specific values: prescale values (for luminosity calculation) trigger chain definition
- TrigObjectMatching
 - for object level trigger information (e.g. tag-and-probe)
 - matches offline particle with trigger object



Data Quality, Data-Sets and Event Displays



Data Quality and Good Run List PAT

- DQ Flags are simple indicators of data quality, from many sources, set for each LumiBlocks
 - Detectors (divided by barrel, 2 endcaps)
 - Trigger (by slice)
 - Offline combined perf. (e/mu/tau/jet/MET/...)



- Combined performance groups define a recommended set of DQ flag criteria for physics objects
 - Physics analyses should not arbitrarily choose DQ flags
- Users in their working group decide which set of DQ flags is needed for each analysis, which can then be used to generate a GoodRunList (GRL)
 - Standard lists centrally produced by DQ group



ATLAS Data Quality



- to manage all this flood of information?
- Answer given by the **Run Query Tool**

MDTvsRPC: Green

GLOBAL: Green RPCBA:

Green

Green

RPC: Green

RPCBC:

MuonTracks: Green

 TGC: Green MuonPhysics: Red MuonSegments: Green

Ē

Run Query Tool

Query: find run 90270-90350 and events 100000+ / show ...

f r 89465-93000 and duration 2m+	runs that had a duration of at least two minu	utes	
f r 90270+ and ev 100k-	runs with run number >= 90270 and < 100k	events	
f r 90100-90150 and det any sct	runs with at least one SCT in and show all participating detectors		
f r 90270-90350 and st *IDCos* 10k+ and	d st *RPC*		
	runs where sum of *IDCos* streams has mo events and one *RPC* stream is enabled	ore than 10000	
f r 90270-90350 and mag s and not mag	it i		
	runs with solenoid on and toroid off	through comm	and-line
f r 90270-91350 and dq lar g	runs where all LAr-EM are "green"	or web interface	
f r 90270-90350 and dq em y+ and dq pi	ixb y+		
	runs where LAr-EM and Tile DQ flags are a Pixel-B is "G"	t least "Y" and	
f r 90270-90350 and dq any pix n.a.			
	runs where at least one Pix qualitity flag is unset		
f t 10.9.2008-13.9.2008 and ptag data08	_1beam*		
	runs with single-beam project tag, between	10th and 13th Sep 08	
f r 91890-92070 and tr EF_e5*	runs with triggers matching pattern EF_e5*	[not case sensitive]	
2011	part 4: Data and Physics Tools		31

Luminosity

- Raw input recorded online in time units called *luminosity blocks* ~ 2 minutes inverval
- Actual luminosity derived offline and stored in COOL data-base
- Key ingredients to define luminosity in physics analysis
 - time range / machine conditions, such as:
 - 7 TeV collisions from stable beams
 - data approved for physics analyses in the *good run list*
 - detector, trigger and reconstruction data quality
 - trigger configuration matches to the one needed by physics analysis
 For example:
 - EF_e25_loose active
 - both e25 and mu20 active and unprescaled
- Luminosity precision 3.4%



Coherent Data-Sets

- Three sources of data can be combined as a coherent data-set
 - real data from calibrated offline reconstruction at CERN farm (the 'Tier0')
 - real data from earlier periods reprocessed with the same software version
 - MC simulation re-processed or generated with the same software version

• For example:

- Data 2011 Mar-Aug (release 16, reprocessed with 17.0.x)
- Data 2011 Sep-Oct (release 17.0.x at the 'Tier0')
- MC simulation from the 'MC11b' production
- Data 2010 was not reprocessed! 2010 makes up only 0.8% of 2011's data sample

• Production catalogue data base: 'AMI' – Atlas Metadata Interface

- Information about samples: size, commands for evt generation + reconstruction, formats, availabilities
- web and command-line interface
- GRID certificate in ATLAS virtual organisation needed to access it



Distributed Computing – GRID

- During early planning for LHC computing the need for de-centralizing was recognised
 - a single centre would not supply enough resources for the LHC experiments
 - space@cern, power and financial resources (no-one would fund such a centre)
 - efficiencies
- Computing GRIDs were fashionable at that time, like today's cloud
- Aim is seamless connection of storage and computing resources distributed across the planet
 - job should go to where the data is without needing to know where the job actually ran
 - data moving only for registered / requested activities
- A lot of framework applications created to make heterogeneous and distributed environment work 'seamlessly'
 - data-management
 - book-keeping of jobs and cpu resources
 - production system
 - installation of experiments' software and data-bases
 - downtime management

CGCC Enabling Grids for E-sciencE

ATLAS Grids

• In fact ATLAS uses three grids:

- EGEE in Europe, Asia and Canada
- OSG in USA
- NorduGrid in Nordic countries
- This is a fact which we mostly try and hide from you
 - But there is a lot of complexity here...



. Anali

LCG

G. Stewart Phys. Analysis Tutorial

GRID Hierarchies



- I Tier-0: CERN
- I0 Tier-I: National Computing Centres (BNL, RAL, IN2P3, ...)
- 40 Tier-2: Regional Computing Centres (ScotGrid, Frascati, Toronto, ...)
- Composed of multiple individual sites
 - ~80 T2 Analysis queues in PanDA



Who Does What?

Tier-0:

EOS

- Copy RAW data to CERN <u>Caster</u> Mass Storage System tape for archival
- Copy RAW data to Tier-1s for storage and subsequent reprocessing
- Run first-pass calibration/alignment (within 24 hrs)
- Run first-pass reconstruction (within 48 hrs)
- Distribute reconstruction output (ESDs, AODs & TAGS) to Tier-1s
- Tier-1s (x10):
 - Store and take care of a fraction of RAW data (forever)
 - Run "slow" calibration/alignment procedures
 - Rerun reconstruction with better calib/align and/or algorithms
 - Distribute reconstruction output to Tier-2s
 - Keep current versions of ESDs and AODs on disk for analysis
 - Run large-scale event selection and analysis jobs for physics and detector groups
- Tier-2s (~x40):
 - Run analysis jobs
 - Run simulation (and calibration/alignment when/where appropriate)
 - Keep current versions of AODs and samples of other data types on disk for analysis
- Tier-3s:
 - Provide access to Grid resources and local storage for end-user data
 - Contribute CPU cycles for simulation and analysis if/when possible

Event Displays

- projective event display: Atlantis
 - event projected onto planes of physics interest: transverse plane, calo layer ($\eta\Phi$)
 - two technical parts:
 - JiveXML, part of athena to extract event info
 - Atlantis, reads XML files + provides Java-based display, independent of athena or architecture
- 3-dim geometry and event view: Persint
 - OpenGL based
 - powerful way of displaying muon spectrometer details
- 3-dim display inside athena: VP1
 - perspective display of geometry and event
 - based on Qt4 and Inventor/OpenGL
 - active during event processing
 - powerful tool for debugging & developing geometry, event data model and reconstruction









Atlantis





Persint



Virtual Point 1





Summary

- Started with adressing how a big collaboration like ATLAS works
- Today's focus was the main practical aspects and tools needed for physics analysis
- It is not needed to master them all right from the start!
 - working e.g. from centrally produced n-tuples is much easier
 - still it is good to know how the data are/were treated or understand technical discussions
 - when needed, detailed instructions always available from colleagues, tutorials and documentation
- Essays (~2 pages)
 - pile-up
 - top physics
 - Higgs searches



Pile-Up

- Pile-up as function of machine conditions, in/out of time
- Physics nature of pile-up collisions
- Reconstruction of pile-up and recognition of signal vertex
- Pile-up effects in physics analyses

ATLAS Soft QCD results https://twiki.cern.ch/twiki/bin/view/AtlasPublic/StandardModelPublicResults

ATLAS Pile-up measurements https://twiki.cern.ch/twiki/bin/view/AtlasPublic/InDetTrackingPerformanceApprovedPlots#Pile_up



B-Physics

- Spectrum of hadrons with b-content, decay channels
- Kinematics and identification of b-decays
- First B-physics results in ATLAS
 - can be overview or a more detailed description of one result
- Physics goals

Particle Data Group on B hadrons http://pdg.lbl.gov/2011/reviews/rpp2011-rev-b-meson-prod-decay.pdf

ATLAS papers on J/Psi and Y cross-sections http://arxiv.org/abs/1104.3038 http://arxiv.org/abs/1106.5325



Top Physics

- Standard model theory
- Production and decay of top quarks at LHC
- Identification of events with top quarks
- Measurements with top quarks
 - again can be overview or a more detailed description of one result

Top physics at Hadron Colliders, A Quadt, http://pdg.lbl.gov/2009/reviews/rpp2009-rev-top-quark.pdf

ATLAS papers https://twiki.cern.ch/twiki/bin/view/AtlasPublic

Hadron Collider Physics Conference, Paris http://indico.in2p3.fr/conferenceOtherViews.py?view=cdsagenda&confId=6004



Higgs Physics

- Standard model and electroweak symmetry breaking
- Higgs production and decay at LHC (std model Higgs boson only)
- Identification of Higgs decays, rejection of background

 again can be overview or a more detailed description of one analysis
- Current limit/sensitivity on Higgs boson as function of its mass

PDG Higgs review (2008 – outdated!) http://pdg.lbl.gov/2011/reviews/rpp2011-rev-higgs-boson.pdf

ATLAS papers https://twiki.cern.ch/twiki/bin/view/AtlasPublic

Hadron Collider Physics Conference, Paris http://indico.in2p3.fr/conferenceOtherViews.py?view=cdsagenda&confId=6004

