How to do an ATLAS analysis

Stefan Richter (DESY)
on behalf of the Early Career Scientist Board
Some slides adapted from Ruth Pöttgen and Gabriel Facini

ATLAS Induction Day · October 21, 2019

Analysis?

- Scientific statement from experimentation
- Published numbers with uncertainties
- Broadly three types:
  - performance: this algorithm works this well
  - direct search: this new process exists or not
  - measurement: this (“known”) process looks like this
- Never forget what statement you intend to make. Analysis decisions should be based on that!

Goals of this session

- Before you start
- Some basic “ingredients” of an analysis
- Analysis roadmap
- Where to find what information
- Some best practices

How to decide what analysis to do?

How to decide what analysis to do?

- Existing theory/experiment tensions
- Thought interesting in the past
- “Because we can”
- (“My supervisor told me to do this”)

Ask *why* (before *how*)

- Your time and energy is limited!
- Make an *informed* decision
  - People: **Glance (people, mentors)**
    Finding the right people is half the fun & success!
  - Published results: **ATLAS**, **CMS**
  - Ongoing analyses: **Glance (analyses)**, asking conveners
  - Theory/phenomenology papers: **arXiv**

Life cycle of your result

- How long will it be useful for?
  - “Throw-away” search?
  - Legacy measurement?
  - Something in between?

- Reusability: publish enough information to actually use your result in secondary analysis
  E.g. publish correlations between bins given your uncertainties
  - **HepData**
  - **Rivet**
  - **CheckMATE**

2 — Analysis ingredients

Analysis ingredients

- Define process of interest
- Establish detector signature
  (using Monte Carlo simulation, experience, intuition)
- Select candidate events
- Estimate number of background events
- Histogram observables of interest
  Histogram = the “natural” data structure of HEP
- Analyse histograms statistically to extract parameters of interest ⇒ publish 🎉

- Do not take the above steps as set in stone! Your creativity is needed

ATLAS Data Flow

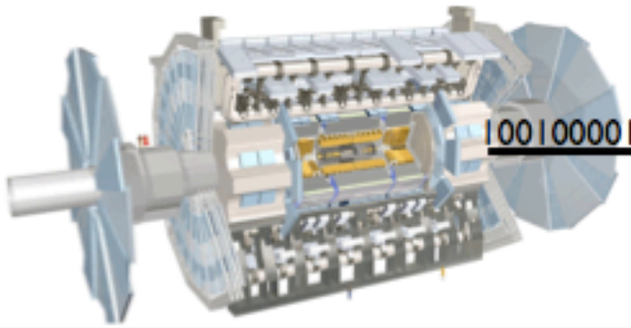
LHC/CERN

ATLAS

2-level trigger

40 MHz → 100 kHZ → 1.5 kHZ

raw data @ 1.5 GB/s



1001000011

Trigger & DAQ

10010101011

~10 PB/year

Worldwide LHC Computing Grid

Monte Carlo production

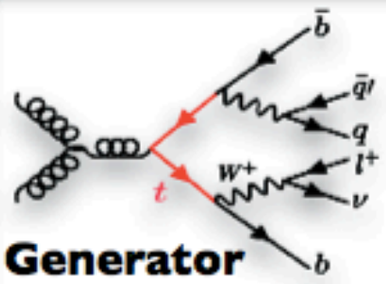
~100k CPUs over 100 PB

Local resources

Athena Framework

ROOT

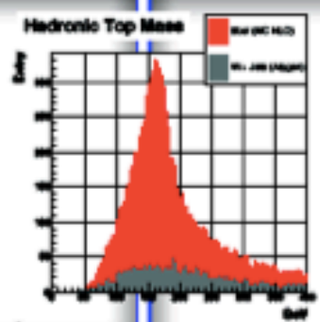
Detector Simulation



Generator

Reconstruction

Analysis



generated MC

simulated MC

reconstructed data/MC

ntuple

plots/ tables

QFT matrix element

primary kinematics

detector hits

tracks, clusters, jets

~GB-TB

Results!

Where to find…

- … information about ***all*** ATLAS physics activities:
  Physics Twiki
  Twikis are a pain to edit and the search is useless!

- … datasets (MC, data): **AMI**
  Can analyse them on the Grid or download them locally (if you have enough disk space)

- … mailing lists: **CERN e-groups**
  Good place to ask for help!

ATLAS Physics Activity

Physics coordinators: Pierre Savard, Klaus Mönig

Physics Twiki

High priority: please help on combined performance and simulation developments and analysis

- Open tasks are described at the following TWikis:
  Tracking | **EGamma** | **Muon** | **Tau** | **JetETmiss** | **Flavour tagging** | **Simulation** | PMG

Key links

Run 2 Analysis Planning
Results for Conferences and Deadlines for papers and CONF notes
ATLAS physics and other workshops
Results including 2018 data
Planning for Run 3 physics

Run-2 Information

LHC schedule 2017 LHC schedule 2018
Run-2 planning and operation
Data and Monte Carlo Datasets for Analysis
Luminosity
Trigger for Run-2:
Menu, L1Topo, Scouting
TADA – Fast physics monitoring
Software:
Release21 software release planning
Run-2 data:
Recent DataSummary
Data 2016 summary, Data 2017 summary, Data 2018 summary
Good Run Lists, Tier-0 processing, Data-set lister, Run-2 reprocessing
Run-2 streaming setup
Report on Run-2 B-physics stream
Validation ART

Combined Performance (CP) Groups

Inner Tracking	M. Danninger, N.E. Pettersson
E/gamma	J.-B. de Vivie, M. Boonekamp
Muon	F. Sforza, S. Zambito
Tau	C. Grefe, S. Tsuno
Jet/EtMiss	C. Young, S. Schramm
Flavour Tagging	M. Kagan, C. Pollard

Physics Analysis (PA) Groups

Heavy Ions	M. Spousta, A.M. Sickles
B Physics & Light States	U. de Sanctis, J. Walder
Standard Model	B. Malaescu, A. Pilkington
Top	E. Shabalina, W. Wagner
Higgs	K.Tackmann, G. Piacquadio
Supersymmetry	Z. Marshall, F. Meloni
Exotics	M.-H. Genest, C. Gwilliam
Higgs & Diboson Searches	B. Murray, V. Cavaliere
Physics Modelling	F. Siegert, S. Amoroso

• • •

Know your triggers

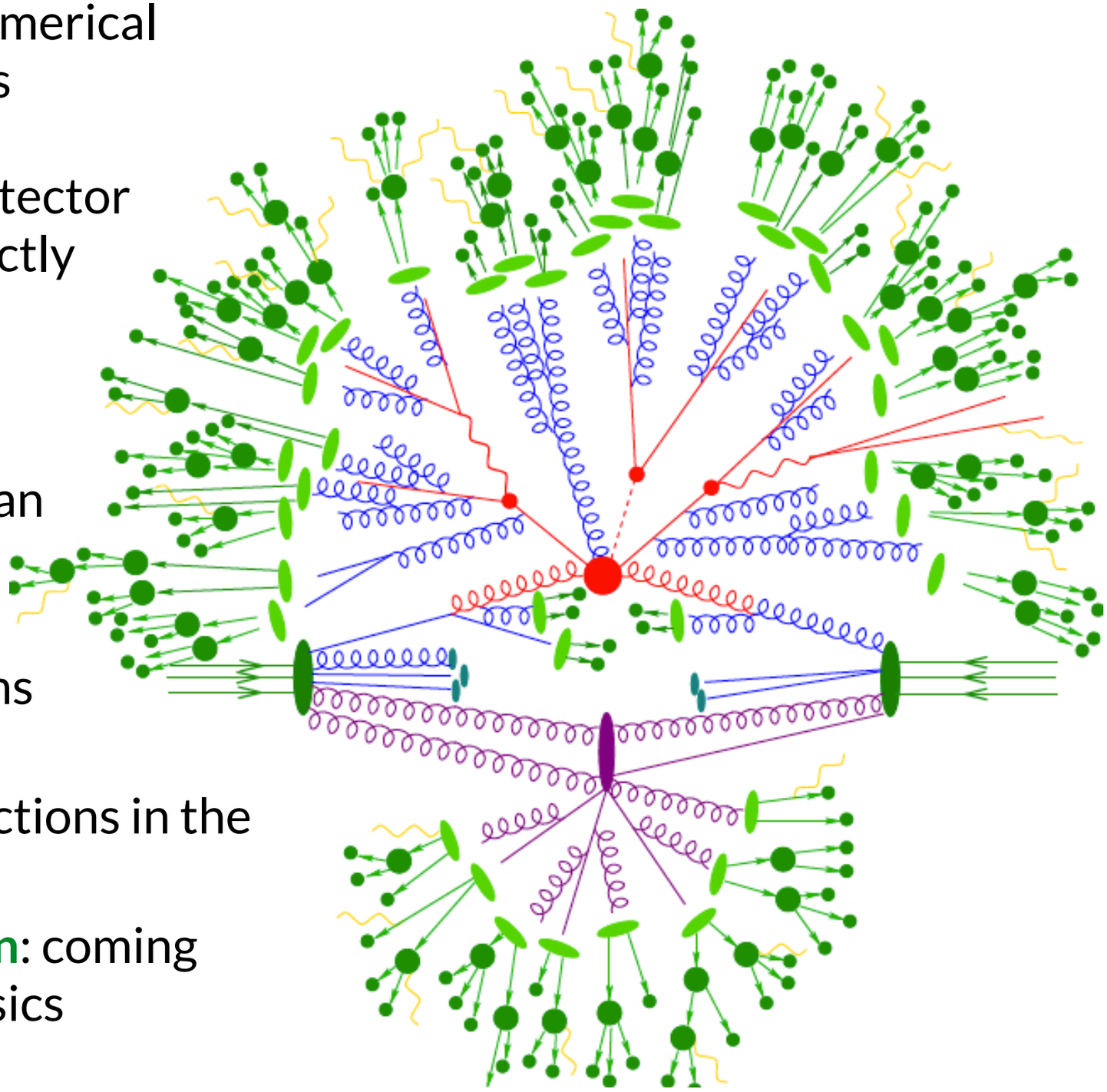
- What's not triggered is gone forever
- Limited total bandwidth
- Sometimes hard limit on what analyses are possible (cf. Higgs → $b\bar{b}$, needs associated V → $\ell\nu$)
- It'll only get worse with higher instantaneous luminosity (Run 3, then HL-LHC)

Monte Carlo simulation

- Generate collisions according to the rules of the SM or some new theory via numerical integration using random numbers

- These events are put through a detector simulation and reconstructed exactly like data

- Some jargon:
  - **hard scatter(ing)**: think Feynman diagram and where the most energy transfer is happening
  - **parton shower**: softer emissions of coloured particles
  - **underlying event**: other interactions in the same proton-proton collision
  - **hadronisation & fragmentation**: coming down to non-perturbative physics



Know your Monte Carlo

- Physics Modelling Group (PMG)
- It's not just “some simulation”, it’s a formal theory prediction!
  Even though it's numerical — that's because we cannot analytically integrate complicated squared matrix elements over high-dimensional phase spaces
- ~~Powerful~~ *indispensable* tool
- Unphysical misuse possible — reading **this** helps you avoid it
- Get MC-literate by reading **this review**

Really know your MC?

- *"Compared to the Standard Model"*
  —or—
  "compared to the NLO QCD prediction in the narrow-width approximation calculated by Herwig 7, interfaced with the Pythia 8 parton shower using the Powheg method, using the Lund string hadronisation model and the NNPDF 3.0 NLO PDFs with the AZ tune of the underlying event, …"

- Think about what matters when

Performance & uncertainties

- CP groups provide recipes for physics objects
- PMG provides recipes for theory uncertainties
- Your analysis might have concerns beyond those! Study, measure in control regions, estimate…
- Arguably the hardest and most important part of your analysis

Statistics tools

- Some ATLAS standards and some common tools in ROOT: RooFit, RooStats, HistFactory, …
  ROOTless solutions being developed <3

- StatCom has **links to documentation** and this new **documentation under construction**

- Like everything, lots of jargon: Asimov dataset, profiling, pulls, rankings… **a crash course**

3 — Publishing

ATLAS approval

- Team *(“do”)* < subgroup < group < ATLAS *(“sign”)*

- The game: ask all of ATLAS if they think what you did is correct. If so, make it public.
  - Don’t take critique as a personal attack
  - In science, you are guilty until proven innocent 😉

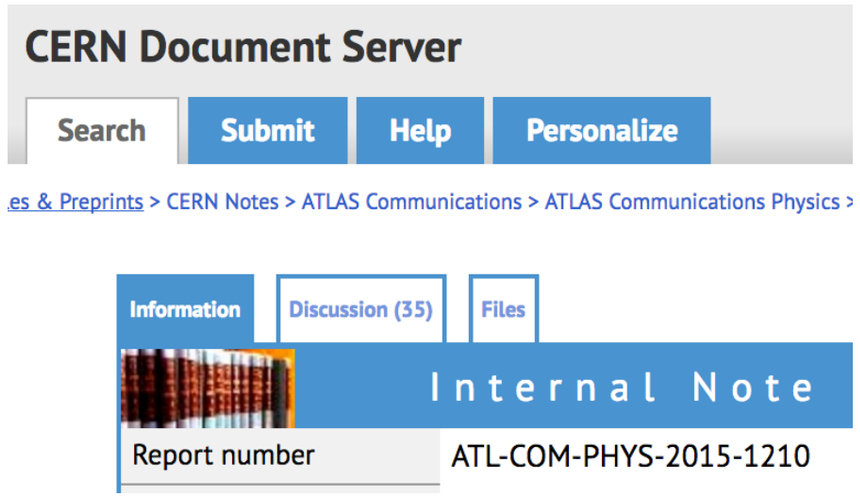
- ATLAS Publication Committee (PubCom) ensures overall quality of ATLAS publications

- Approval stages:
  subgroup → group → ATLAS → Physics coordination → PubCom → spokesperson (full explanation **here**)

- Accompanied by an Editorial Board (EB) that scrutinises your analysis. Formed when the analysis is mostly complete

Cern Document Server

- CDS
- Hosts public and internal documents
- Exchange between reviewers and authors of a document
- Reviewers post comments in the discussion tab, authors use the same functionality to reply to the comments, answer questions, clarify…
- All ATLAS documents get a report number and are stored in CDS



4 — Final thoughts

What is the statement I want to make?
Regularly, take a step back from the details and look at the big picture

Talk to people
Don’t get tunnel vision.
Being able to explain what you are doing is important

What is the plan?
Where do you want to be in a day, week, month… 2 years?

Communities

Many communities exist in ATLAS that are in place to gather expertise and offer advice
Visit and/or join these communities!

Physics Analysis & MC Groups

- B Physics & Light States
- Top
- Higgs
- Standard Model
- SUSY
- Exotics
- Heavy Ions
- Upgrade Physics
- Physics Modelling (PMG)

- Simulation Group
- MC Production Group

Performance Groups & Fora

- Inner Tracking
- Egamma
- Muon
- Tau
- Jet/Etmiss
- Flavour Tagging

- Isolation and Fake Forum (IFF)
- PDF Forum
- Astroparticle Forum
- Statistics Forum
- Machine Learning Forum

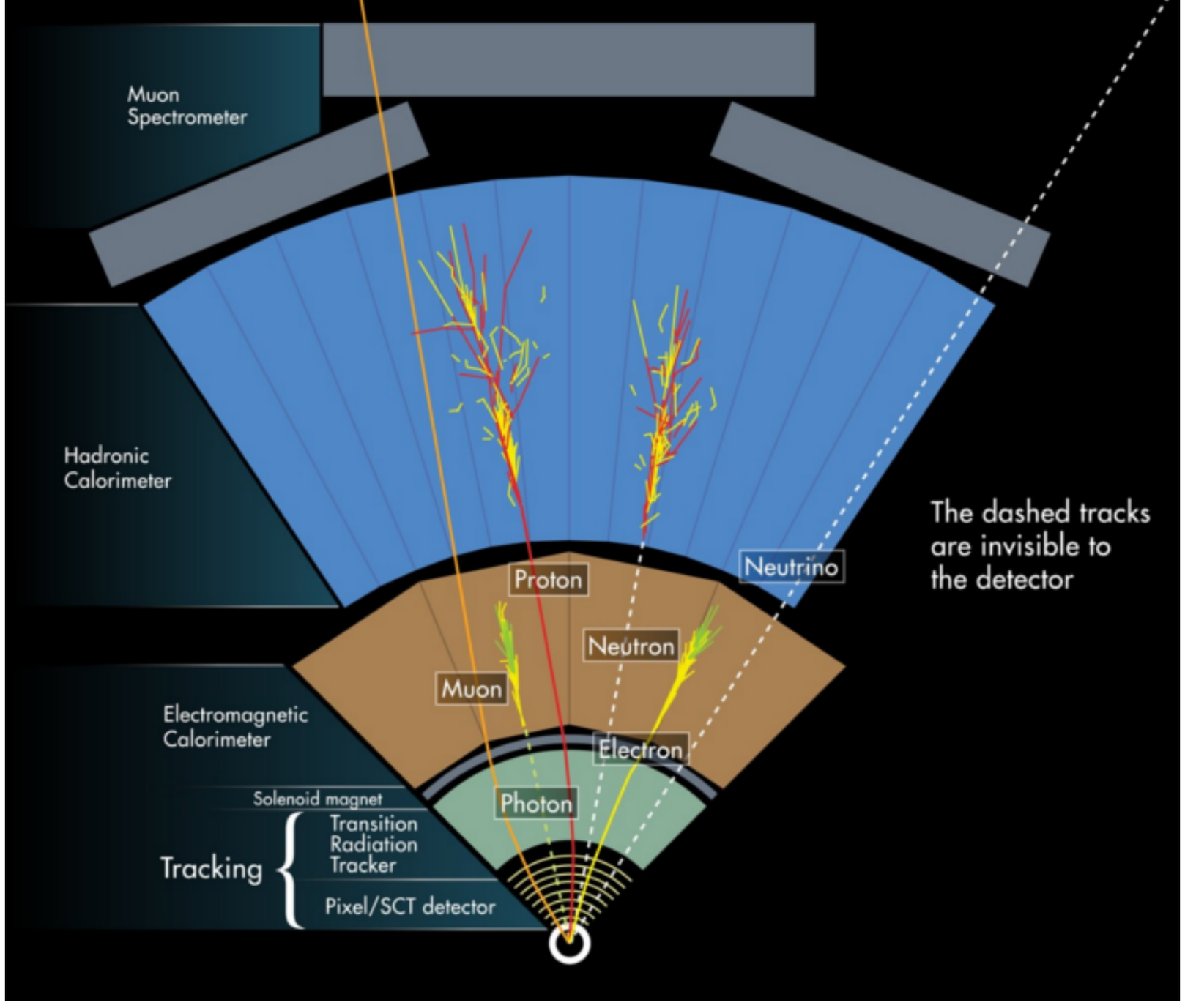
- Go to the Twiki
- Find the names
- Send a mail
- Have a chat

- Keep learning. If you hear a new buzzword, find out what it’s about. But not just buzzwords!
  particle flow, 750 GeV diphoton, deep learning, containerisation, pseudo-continuous b-tagging, …
- Question everything (incl. yourself), but stay humble
- You can join any meeting and ask any question

Conclusion

- Analysis is the job
- Requires a working experiment
- ATLAS is a big place & experts exist in all areas
  Discuss with them and try to be one yourself!
- Know something about everything and “everything” about something.
- Don’t make the mistakes others have made. Make new ones
- Have fun!

Backup



Reconstruction contains custom algorithms that do:

energy deposits → energy measurement (**Detector Groups**)

energy measurements → physics objects (**Combined Performance Groups**)

Types of public results

- All results are approved by the group the analysis is hosted in
  Depending on type of publication, additional approvals are needed

- Public plots — plots showing (usually) performance of detector requiring one week of circulation to ATLAS

- PUB note — simulation performance (no data) requiring one week of circulation to ATLAS and sign-off by two Readers

- CONF note — preliminary result requiring an EB, ATLAS circulation, and additional sign-offs

- Paper (peer-reviewed) — requires same as CONF with an additional ATLAS circulation and additional sign-offs

- Conference talks, proceedings: not covered here ;)

- Glance guides you through the process!

A Broad View

- **ATLAS is so big** it can be easy to lose perspective

- *Nevertheless, you should create a view from data collection to the final result and identify the critical points along the way, learn the details of all areas and build a wide base of experience. (**also consider time**!)*

- Things that should be avoided:

  - only focusing on the final part of the analysis chain i.e. optimization of cuts without learning about the objects themselves

  - only work on performance i.e. better electron efficiency but never pay attention to how that is used in an analysis

  - not work on the detector or upgrade

Get Dirty

- **Experimentalists get better with experience**. The more you do, the “better” you are at your job. So get after it!
  - realize your “elders” might have seen the same thing before and can teach you something - *but you have to ask them!*

- Reading ATLAS papers in review, paying attention in group & sub-group meetings is part of this too!