18.-21.08.2022

Software in der Teilchenphysik

Wie man Erkenntnisse aus den Daten gewinnen kann

Lukas Kretschmann

Bergische Universität Wuppertal, Germany

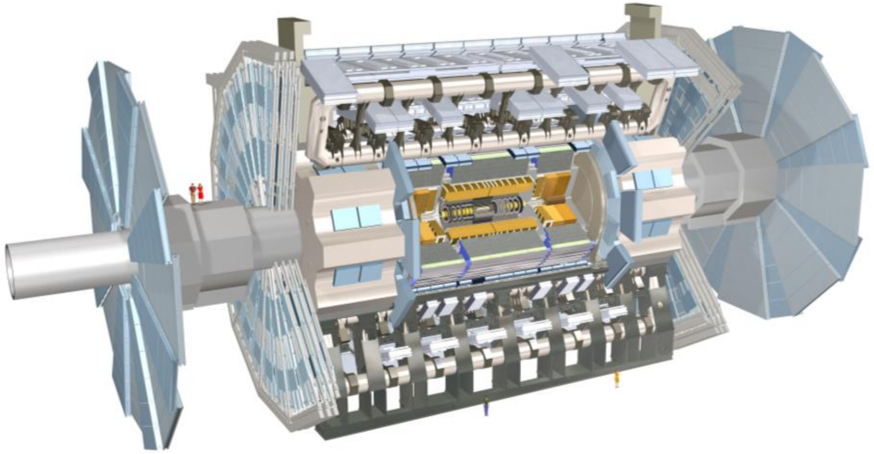
ATLAS
EXPERIMENT

NETZWERK
TEILCHENWELT

Overall Goal / Processing Chain

Overall Goal

Getting from this…



Status: March 2022

Argentina
Armenia
Australia
Austria
Azerbaijan
Belarus
Brazil
Canada
Chile
China
Colombia
Czech Republic
Denmark
France
Georgia
Germany
Greece
Israel
Italy
Japan
Mongolia
Morocco
Netherlands
Norway
Palestine
Philippines
Poland
Portugal
Romania
Russia
Serbia
Slovakia
Slovenia
South Africa
Spain
Sweden
Switzerland
Taiwan
Turkey
UAE
UK
USA
CERN
JINR

ATLAS Collaboration

181 institutions (245 institutes) from 42 countries

…to this

The European Physical Journal

volume 77 · number 8 · august · 2017

EPJ C

Recognized by European Physical Society

Particles and Fields

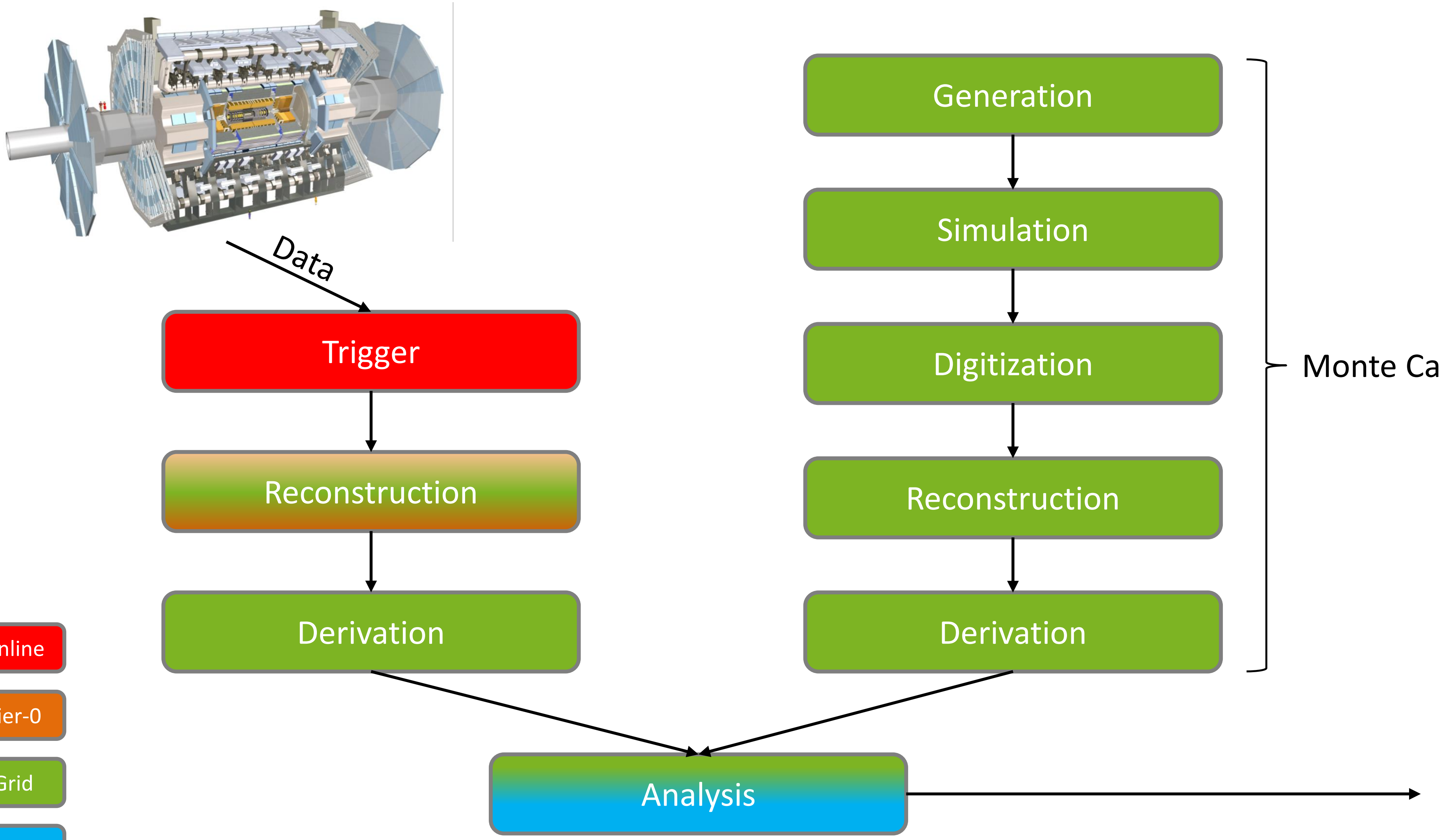
Predicted values of $R_t = \sigma_{\text{tot}}(tq)/\sigma_{\text{tot}}(\bar{t}q)$ calculated with NLO accuracy in QCD in the five flavor scheme using different NLOPDF sets compared to the measured value. The dashed white line indicates the central value of the measured R_t value. The combined statistical and systematic uncertainty of the measurement is shown in green, while the statistical uncertainty is represented by the yellow error band.

From the ATLAS Collaboration: Fiducial, total and differential cross-section measurements of *t*-channel single top-quark production in *pp* collisions at 8TeV using data collected by the ATLAS detector.

Società Italiana di Fisica

Springer

Overall Goal / Processing Chain



Trigger

Reconstruction

Derivation

Generation

Simulation

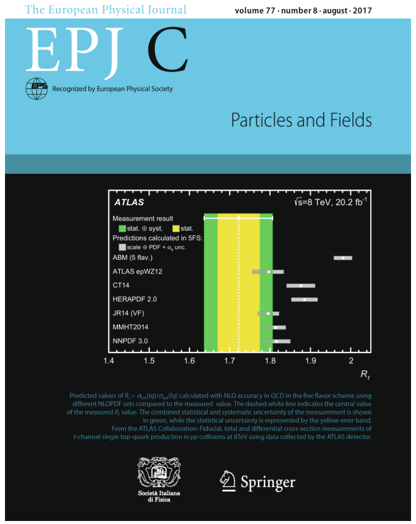
Digitization

Reconstruction

Derivation

Monte Carlo Simulation

Analysis


The European Physical Journal volume 77 · number 8 · august · 2017

EPJ C

Recognized by European Physical Society

Particles and Fields

Predicted values of $R_t = \sigma_{tot}(tq)/\sigma_{tot}(\bar{t}q)$ calculated with NLO accuracy in QCD in the five flavor scheme using different NLOPDF sets compared to the measured value. The dashed white line indicates the central value of the measured R_t value. The combined statistical and systematic uncertainty of the measurement is shown in green, while the statistical uncertainty is represented by the yellow error band. From the ATLAS Collaboration: Fiducial, total and differential cross-section measurements of *t*-channel single top-quark production in *pp* collisions at 8TeV using data collected by the ATLAS detector.

Società Italiana di Fisica

Springer

🟥 Online

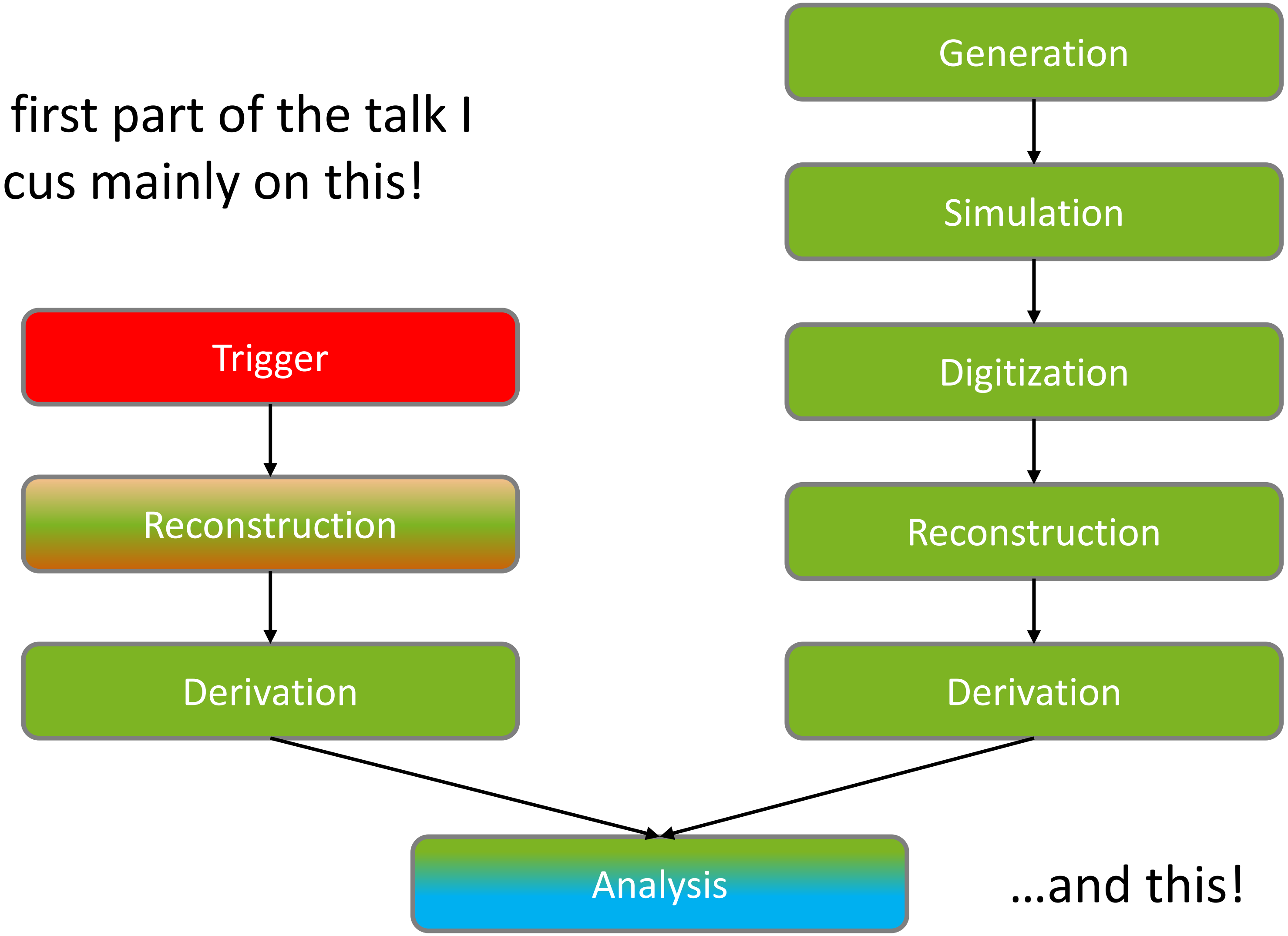
🟧 Tier-0

🟩 Grid

🟦 Local

Overall Goal / Processing Chain

In the first part of the talk I will focus mainly on this!



…this…

…and this!

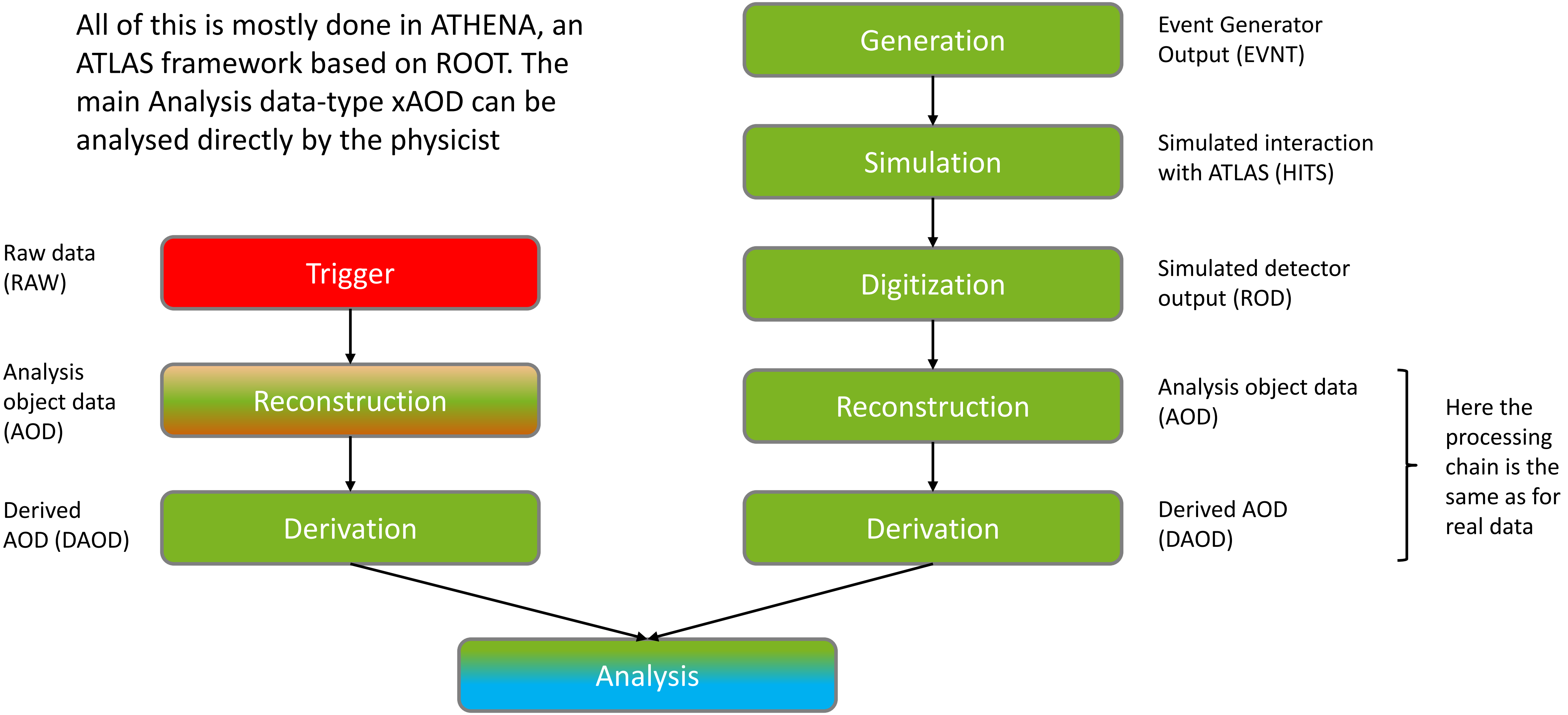
In the second part **we** will do something with this…

…depending on the time we have…

Data flow in ATLAS

Data Formats in ATLAS

All of this is mostly done in ATHENA, an ATLAS framework based on ROOT. The main Analysis data-type xAOD can be analysed directly by the physicist



Raw data (RAW) – Trigger

Analysis object data (AOD) – Reconstruction

Derived AOD (DAOD) – Derivation

Generation – Event Generator Output (EVNT)

Simulation – Simulated interaction with ATLAS (HITS)

Digitization – Simulated detector output (ROD)

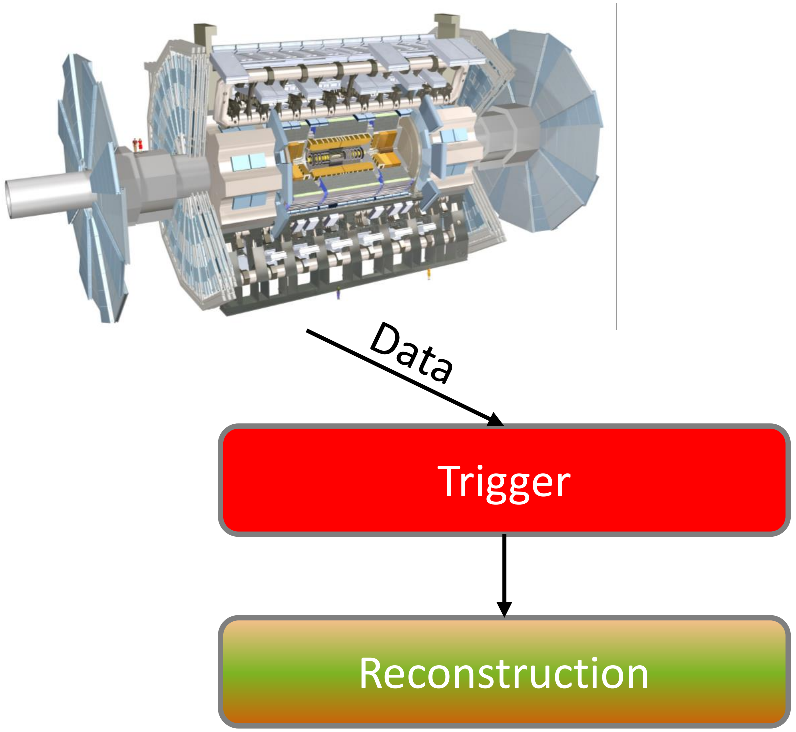
Reconstruction – Analysis object data (AOD)

Derivation – Derived AOD (DAOD)

Here the processing chain is the same as for real data

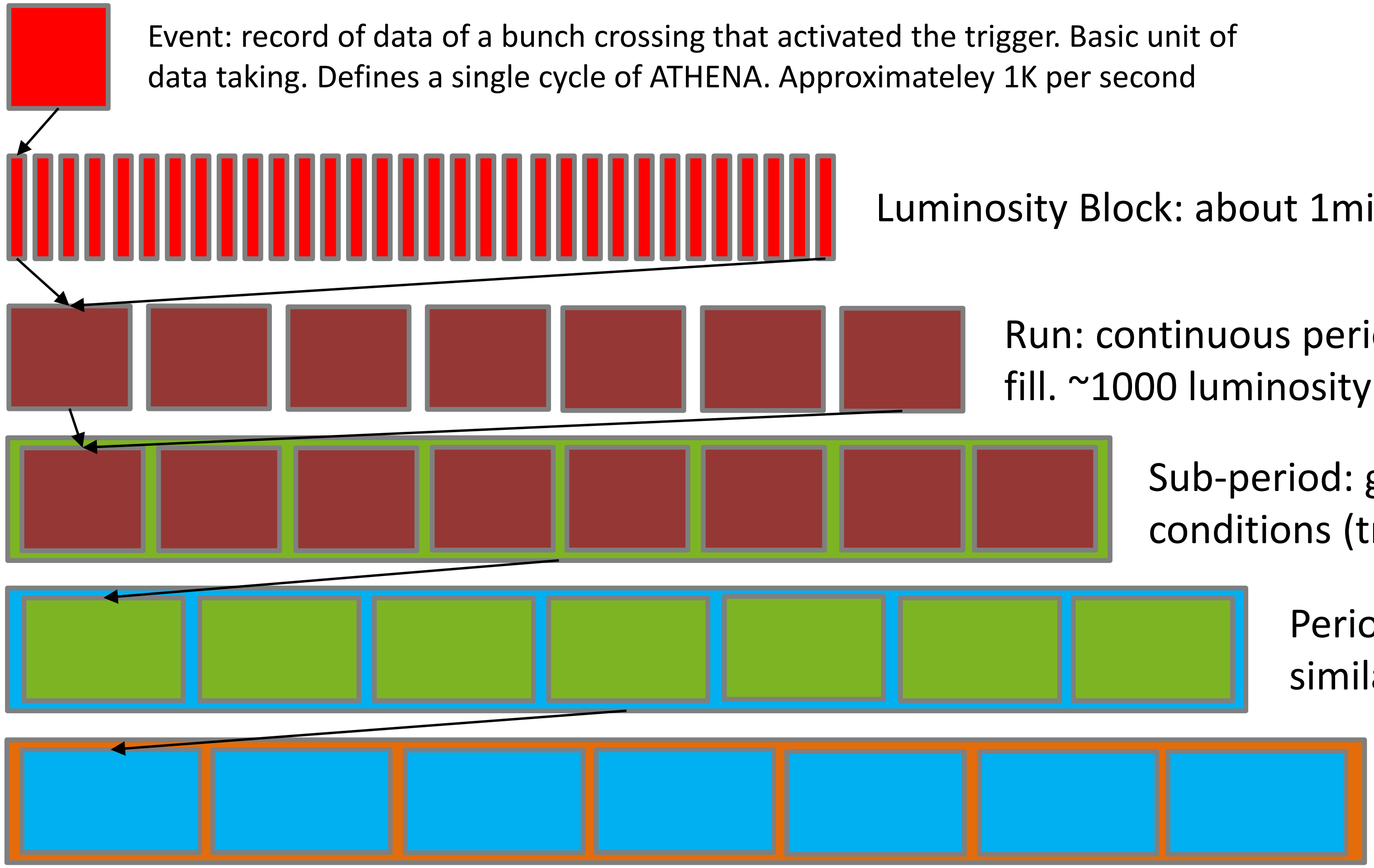
Analysis

Trigger and Data Preparation



First we need to understand how data is divided up in ATLAS…

ATLAS Data-Stream



Event: record of data of a bunch crossing that activated the trigger. Basic unit of data taking. Defines a single cycle of ATHENA. Approximateley 1K per second

Luminosity Block: about 1min of data taking. ~100k events

Run: continuous period of data taking. Usually a single LHC-fill. ~1000 luminosity blocks

Sub-period: group of runs taken with similar conditions (trigger, LHC parameters). ~5-10 runs

Period: group of sub-periods taken with similar conditions. ~5-10 sub-periods

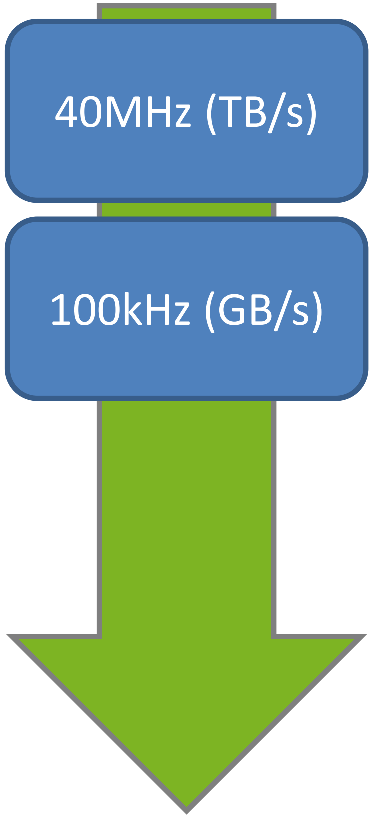
LHC Run: long period of data taking between shutdowns. ~20-30 periods per LHC run

Trigger

- Trigger: hardware and software that decides whether a event is written to disk
  → LHC deliveres events at up to 40MHz (1 bunch crossing every 25ns)
  → We can afford to write up to 1kHz
  → This means we can keep one in 40.000 events

- Two-step ATLAS trigger mechanism:

  → Level 1: Implemented in hardware inside ATLAS
  → very little time to decide to decide whether to keep an event or not, we can only use parts of the detector that can be quickly read out (Muon,Calo) (NEW Small Wheel)


40MHz (TB/s)

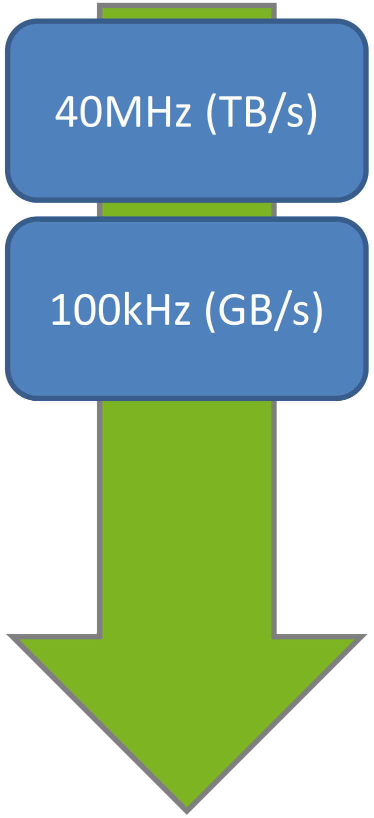
100kHz (GB/s)

Trigger

- Trigger: hardware and software that decides whether a event is written to disk
  → LHC deliveres events at up to 40MHz (1 bunch crossing every 25ns)
  → We can affor
  → This means w

- Two-step ATLAS

  → Level 1: Impl
  → very little tim can only use
  parts of the all Wheel)


40MHz (TB/s)
100kHz (GB/s)

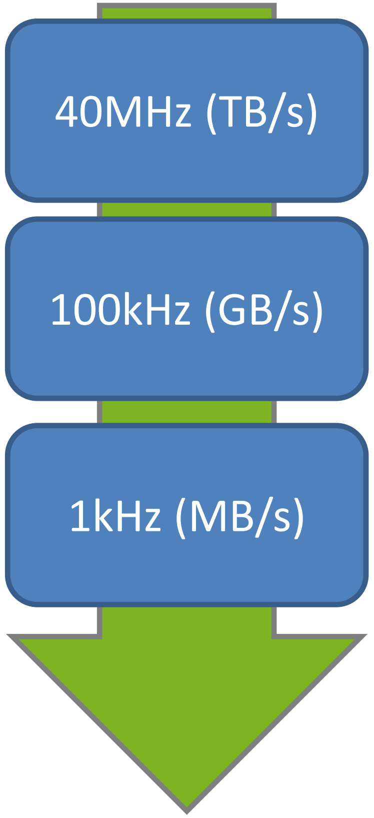
Trigger

- Trigger: hardware and software that decides whether a event is written to disk
  → LHC deliveres events at up to 40MHz (1 bunch crossing every 25ns)
  → We can afford to write up to 1kHz
  → This means we can keep one in 40.000 events

- Two-step ATLAS trigger mechanism:

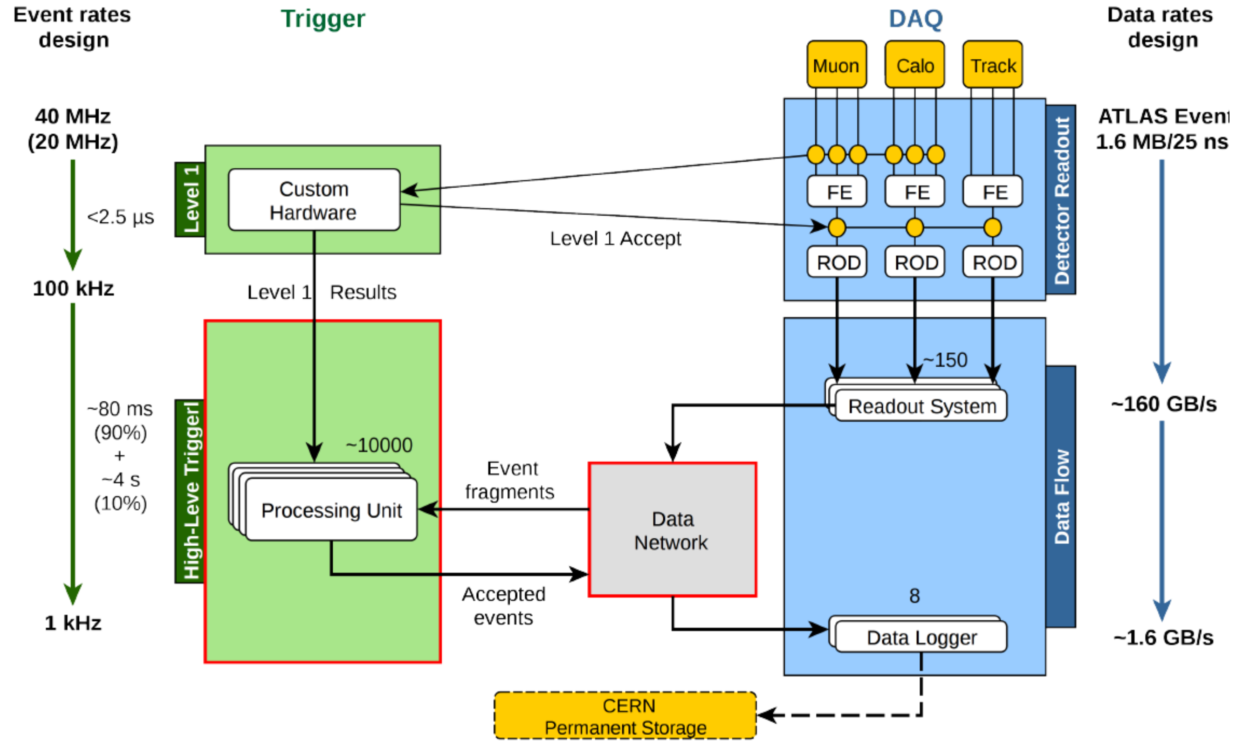
  → Level 1: Implemented in hardware inside ATLAS
  → very little time to decide to decide whether to keep an event or not, we can only use parts of the detector that can be quickly read out (Muon,Calo) (NEW Small Wheel)

  → High Level: implemented in software running on dedicated machines nearby ATLAS
  → has a bit more time, so can use the full detector information, including the ID (Pixel, SCT, TRT)


40MHz (TB/s)
100kHz (GB/s)
1kHz (MB/s)

Trigger

- The types of signatures that the trigger accepts are encoded in the trigger menu
- Trigger is not 100% efficient for any signatures
- Estimating trigger efficency is a crucial part of many physics analysis
  → Match reconstructed objects with the equivalent trigger
- Precise understanding if trigger is needed for all physics analyses



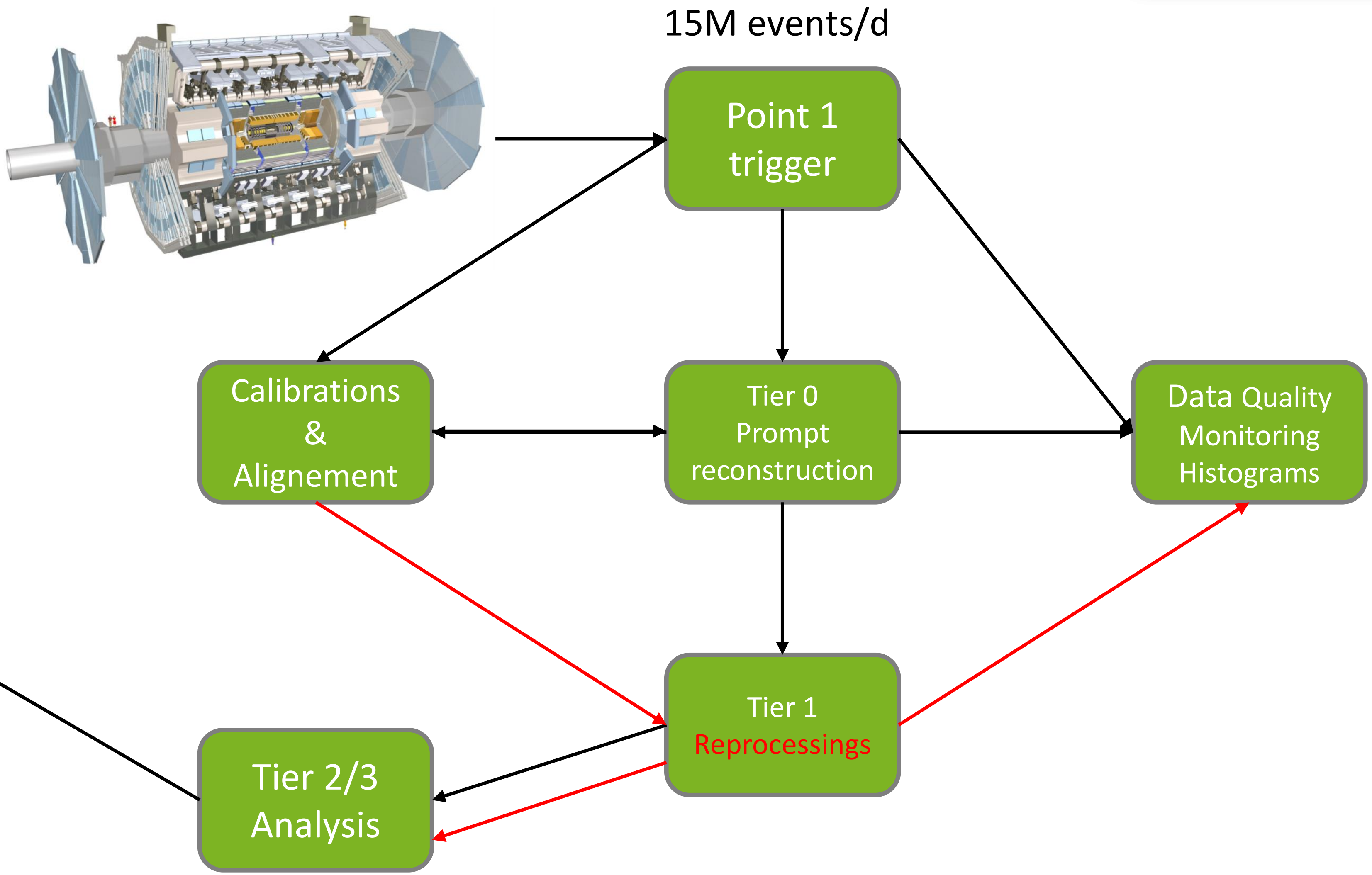
Recorded	Per event	Per year
Raw data	1.6 MB	3200 TB
Reconstructed data	1 MB	2000 TB
Physics data	0.1 MB	200 TB

Data Preparation

Data Preparation

Trigger

Reconstruction



15M events/d

Point 1 trigger

Calibrations & Alignement

Tier 0 Prompt reconstruction

Data Quality Monitoring Histograms

Tier 1 Reprocessings

Tier 2/3 Analysis

The European Physical Journal

volume 77 · number 8 · august · 2017

EPJ C

Recognized by European Physical Society

Particles and Fields

Predicted values of $R_t = \sigma_{tot}(tq)/\sigma_{tot}(\bar{t}q)$ calculated with NLO accuracy in QCD in the five flavor scheme using different NLOPDF sets compared to the measured value. The dashed white line indicates the central value of the measured R_t value. The combined statistical and systematic uncertainty of the measurement is shown in green, while the statistical uncertainty is represented by the yellow error band.
From the ATLAS Collaboration: Fiducial, total and differential cross-section measurements of *t*-channel single top-quark production in *pp* collisions at 8TeV using data collected by the ATLAS detector.

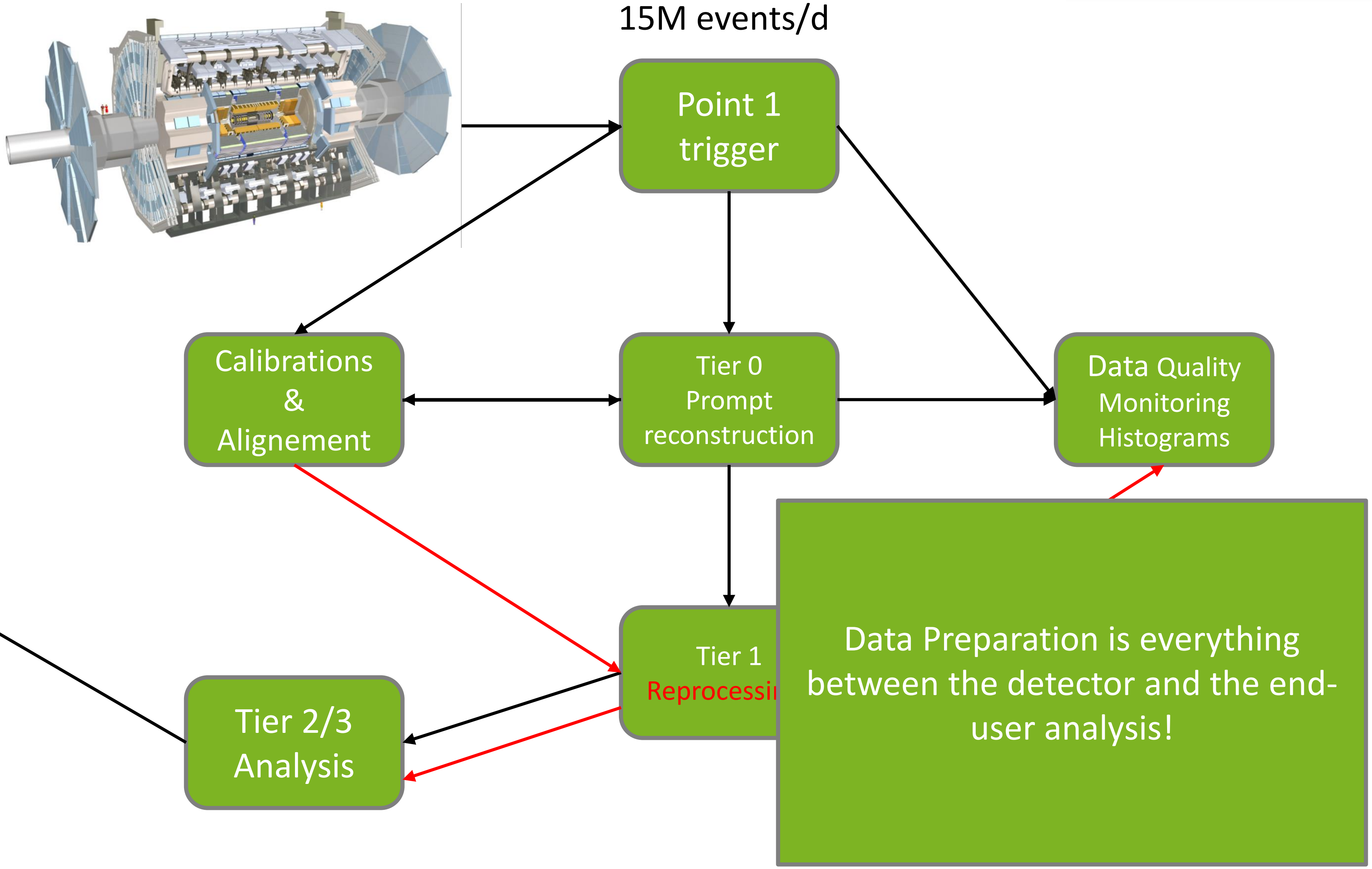
Società Italiana di Fisica

Springer

Data Preparation

Trigger

Reconstruction



15M events/d

Point 1 trigger

Calibrations & Alignement

Tier 0 Prompt reconstruction

Data Quality Monitoring Histograms

Tier 1 Reprocessi

Tier 2/3 Analysis

Data Preparation is everything between the detector and the end-user analysis!

The European Physical Journal volume 77 · number 8 · august · 2017

EPJ C

Recognized by European Physical Society

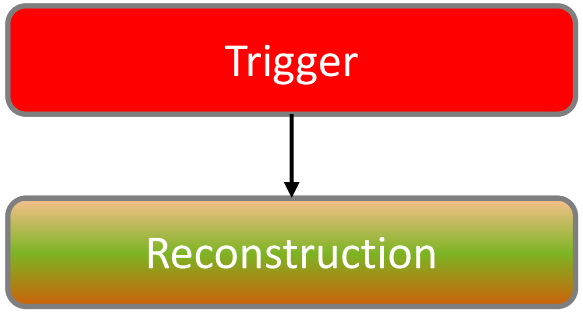
Particles and Fields

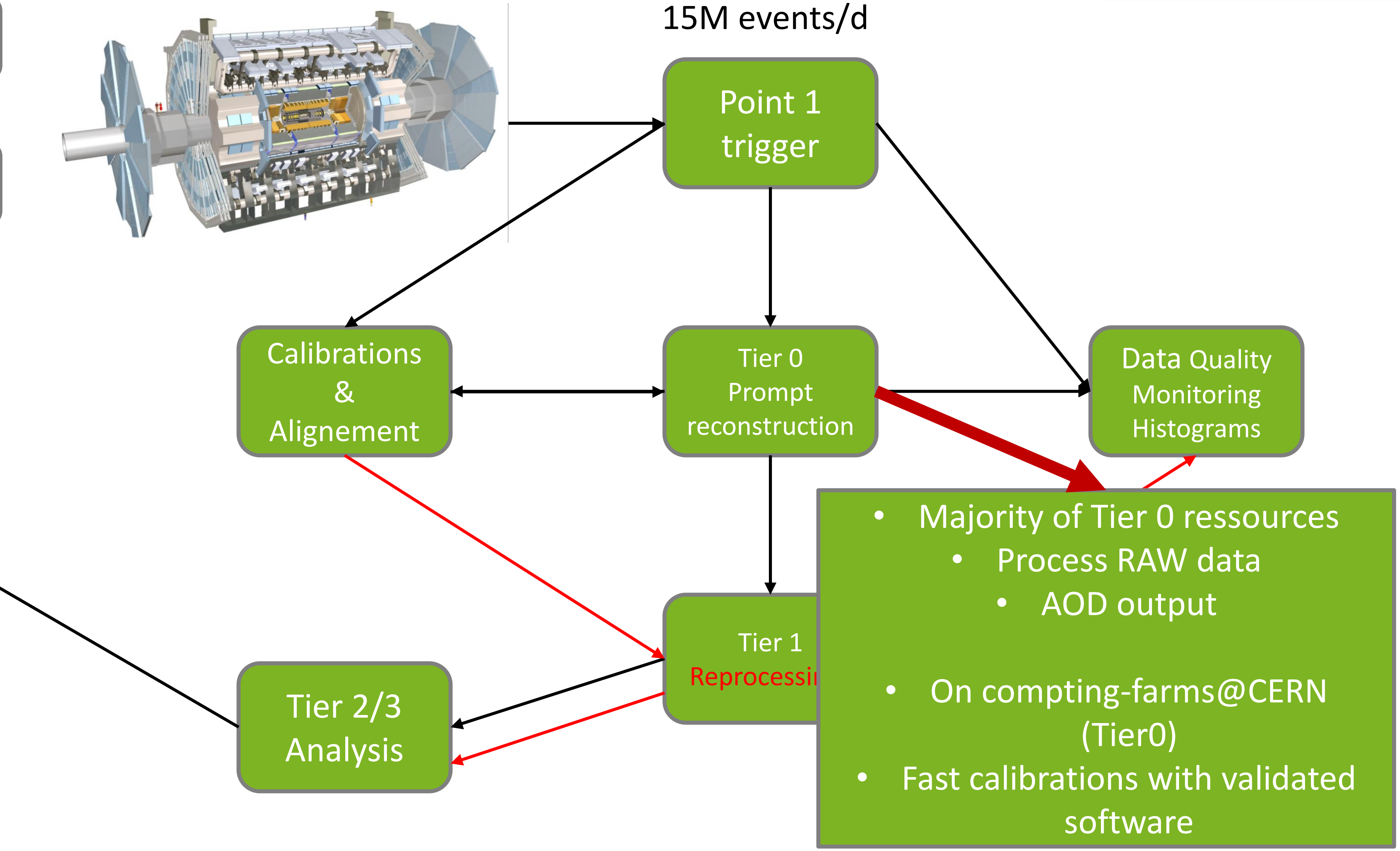
Predicted values of $R_t = \sigma_{\mathrm{tot}}(tq)/\sigma_{\mathrm{tot}}(\bar{t}q)$ calculated with NLO accuracy in QCD in the five flavor scheme using different NLOPDF sets compared to the measured value. The dashed white line indicates the central value of the measured R_t value. The combined statistical and systematic uncertainty of the measurement is shown in green, while the statistical uncertainty is represented by the yellow error band.
From the ATLAS Collaboration: Fiducial, total and differential cross-section measurements of *t*-channel single top-quark production in *pp* collisions at 8TeV using data collected by the ATLAS detector.

Società Italiana di Fisica

Springer

Data Preparation





- Majority of Tier 0 ressources
  - Process RAW data
    - AOD output
- On compting-farms@CERN (Tier0)
- Fast calibrations with validated software


The European Physical Journal

volume 77 · number 8 · august · 2017

EPJ C

Recognized by European Physical Society

Particles and Fields

Predicted values of $R_t = \sigma_{\mathrm{tot}}(tq)/\sigma_{\mathrm{tot}}(\bar{t}q)$ calculated with NLO accuracy in QCD in the five flavor scheme using different NLOPDF sets compared to the measured value. The dashed white line indicates the central value of the measured R_t value. The combined statistical and systematic uncertainty of the measurement is shown in green, while the statistical uncertainty is represented by the yellow error band.
From the ATLAS Collaboration: Fiducial, total and differential cross-section measurements of *t*-channel single top-quark production in *pp* collisions at 8TeV using data collected by the ATLAS detector.

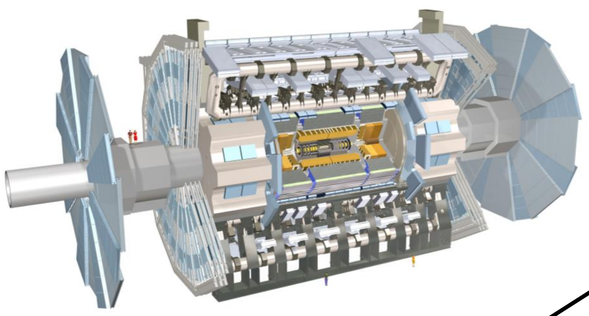
Società Italiana di Fisica

Springer

Data Preparation

Trigger

Reconstruction



15M events/d

- Over time knowledge is improved, bugfixes etc.

→ Reprocessing → better data-set

Calibrations
&
Alignement

Tier 1
Reprocessings

Tier 2/3
Analysis

The European Physical Journal

volume 77 · number 8 · august · 2017

EPJ C

Recognized by European Physical Society

Particles and Fields

Predicted values of $R_t = \sigma_{tot}(tq)/\sigma_{tot}(\bar{t}q)$ calculated with NLO accuracy in QCD in the five flavor scheme using different NLOPDF sets compared to the measured value. The dashed white line indicates the central value of the measured R_t value. The combined statistical and systematic uncertainty of the measurement is shown in green, while the statistical uncertainty is represented by the yellow error band.

From the ATLAS Collaboration: Fiducial, total and differential cross-section measurements of *t*-channel single top-quark production in *pp* collisions at 8TeV using data collected by the ATLAS detector.

Società Italiana di Fisica

Springer

Data Preparation

How much data?
~Luminosity

How good is the data?

~Data Quality

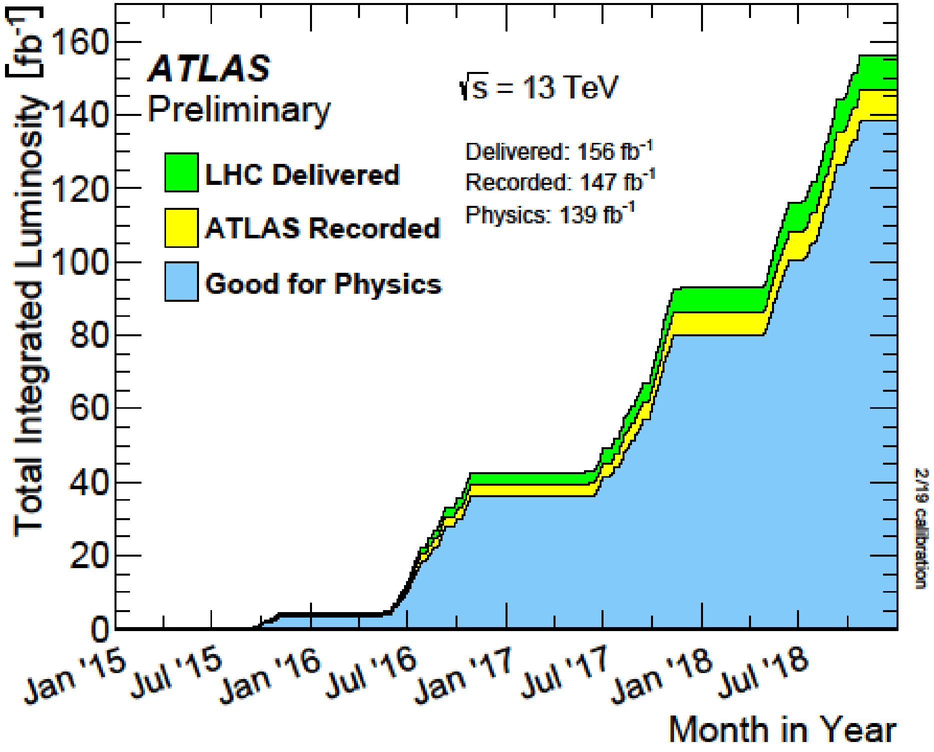
Not-data data???
~Non collision backgrounds/pileup

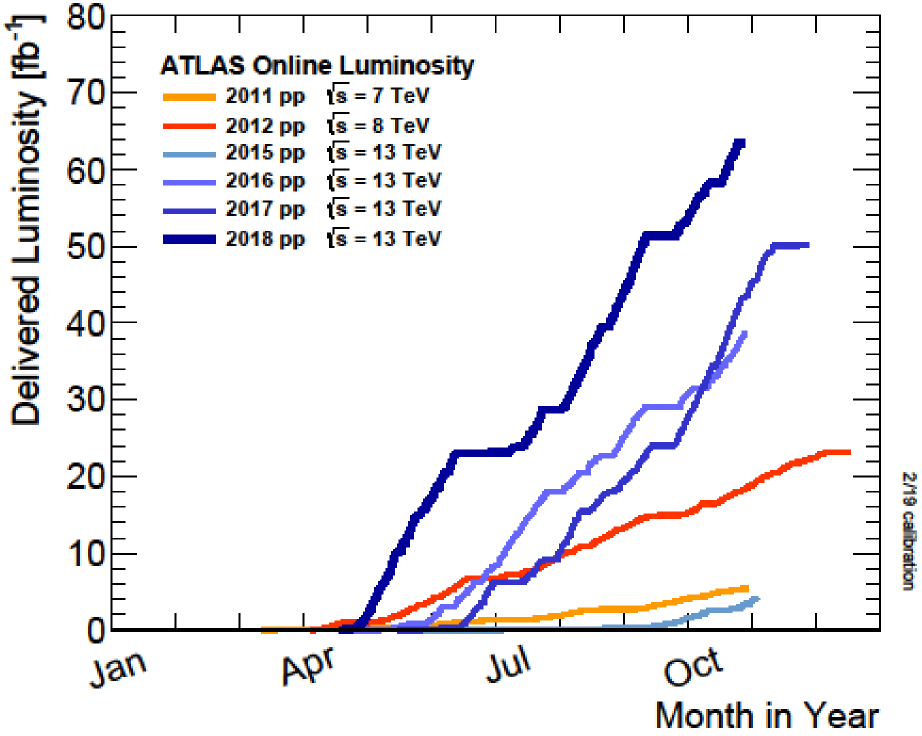
Data describing the data!
~Metadata

- Bulk reconstruction only starts when a run is released from the calibration loop (~48h after the run)
- GoodRunsLists mask out the bad luminosity blocks in analysis
- The Luminosity used in the analysis has to take this missing data into account

Data Quality and Luminosity

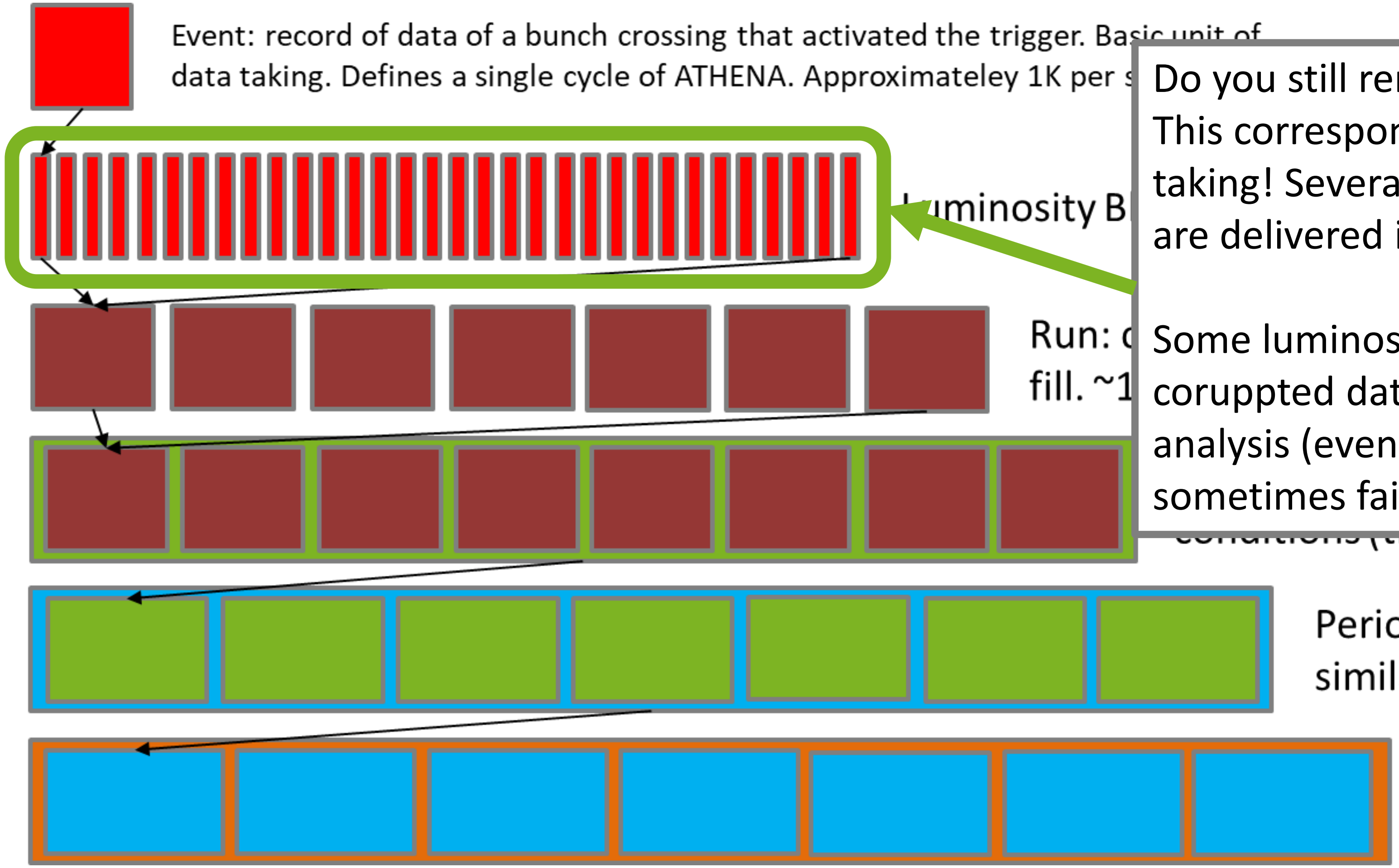
Data Quality and Luminosity





- Overall efficency of all systems: 95.5%
- Allowed to take huge amounts of data during Run1 and Run2
- In the coming Runs these datasets will be more than doubled at higher energies

Good Runs Lists



Event: record of data of a bunch crossing that activated the trigger. Basic unit of data taking. Defines a single cycle of ATHENA. Approximateley 1K per s

Luminosity Bl

Run: c e LHC-
fill. ~1

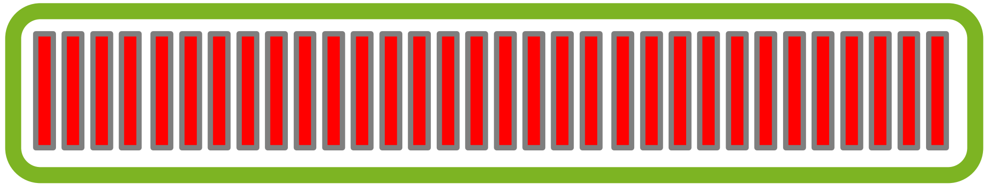
conditions (trigger, LHC parameters). ~5-10 runs

Period: group of sub-periods taken with similar conditions. ~5-10 sub-periods

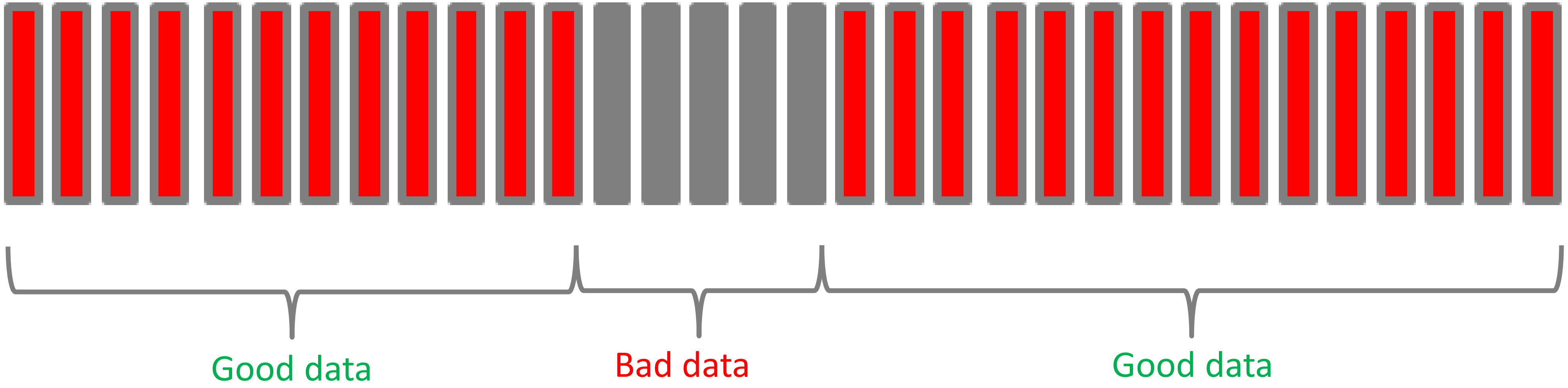
Do you still remeber these guys?
This corresponds to around 1min of data taking! Several hundred luminosity blocks are delivered in one run!

Some luminosity blocks can include bad or coruppted data, which should not be used in analysis (even ATLAS-Subsystems can sometimes fail or produce bad data…)

Good Runs Lists



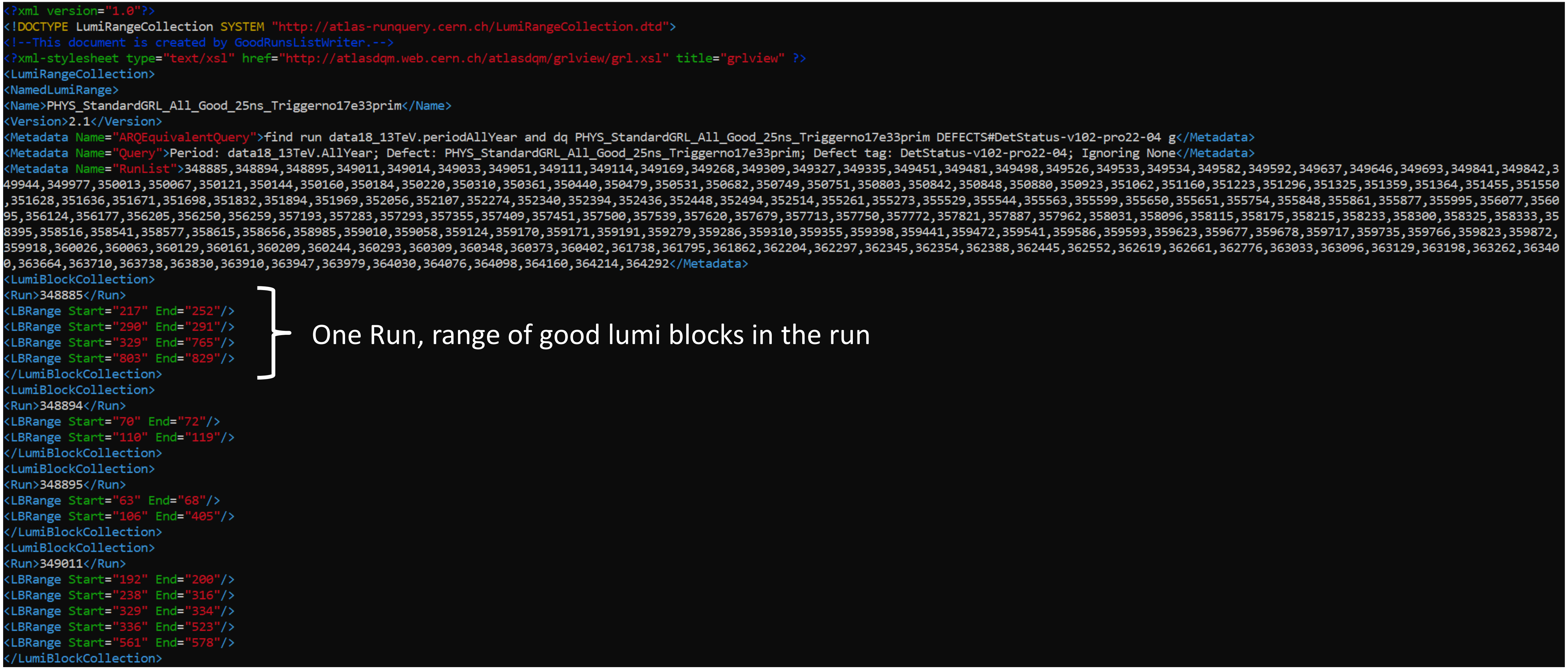
Luminosity Block: about 1min of data taking. ~100k events



- If even one system deliveres bad data, the luminosity block will be masked
- This information is encoded in a database
- When you do your analysis you need to filter out the bad luminosity blocks (GRL)

Good Runs Lists

BERGISCHE
UNIVERSITÄT
WUPPERTAL



```
<?xml version="1.0"?>
<!DOCTYPE LumiRangeCollection SYSTEM "http://atlas-runquery.cern.ch/LumiRangeCollection.dtd">
<!--This document is created by GoodRunsListWriter.-->
<?xml-stylesheet type="text/xsl" href="http://atlasdqm.web.cern.ch/atlasdqm/grlview/grl.xsl" title="grlview" ?>
<LumiRangeCollection>
<NamedLumiRange>
<Name>PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim</Name>
<Version>2.1</Version>
<Metadata Name="ARQEquivalentQuery">find run data18_13TeV.periodAllYear and dq PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim DEFECTS#DetStatus-v102-pro22-04 g</Metadata>
<Metadata Name="Query">Period: data18_13TeV.AllYear; Defect: PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim; Defect tag: DetStatus-v102-pro22-04; Ignoring None</Metadata>
<Metadata Name="RunList">348885,348894,348895,349011,349014,349033,349051,349111,349114,349169,349268,349309,349327,349335,349451,349481,349498,349526,349533,349534,349582,349592,349637,349646,349693,349841,349842,3
49944,349977,350013,350067,350121,350144,350160,350184,350220,350310,350361,350440,350479,350531,350682,350749,350751,350803,350842,350848,350880,350923,351062,351160,351223,351296,351325,351359,351364,351455,351550
,351628,351636,351671,351698,351832,351894,351969,352056,352107,352274,352340,352394,352436,352448,352494,352514,355261,355273,355529,355544,355563,355599,355650,355651,355754,355848,355861,355877,355995,356077,3560
95,356124,356177,356205,356250,356259,357193,357283,357293,357355,357409,357451,357500,357539,357620,357679,357713,357750,357772,357821,357887,357962,358031,358096,358115,358175,358215,358233,358300,358325,358333,35
8395,358516,358541,358577,358615,358656,358985,359010,359058,359124,359170,359171,359191,359279,359286,359310,359355,359398,359441,359472,359541,359586,359593,359623,359677,359678,359717,359735,359766,359823,359872,
359918,360026,360063,360129,360161,360209,360244,360293,360309,360348,360373,360402,361738,361795,361862,362204,362297,362345,362354,362388,362445,362552,362619,362661,362776,363033,363096,363129,363198,363262,36340
0,363664,363710,363738,363830,363910,363947,363979,364030,364076,364098,364160,364214,364292</Metadata>
<LumiBlockCollection>
<Run>348885</Run>
<LBRange Start="217" End="252"/>
<LBRange Start="290" End="291"/>
<LBRange Start="329" End="765"/>
<LBRange Start="803" End="829"/>
</LumiBlockCollection>
<LumiBlockCollection>
<Run>348894</Run>
<LBRange Start="70" End="72"/>
<LBRange Start="110" End="119"/>
</LumiBlockCollection>
<LumiBlockCollection>
<Run>348895</Run>
<LBRange Start="63" End="68"/>
<LBRange Start="106" End="405"/>
</LumiBlockCollection>
<LumiBlockCollection>
<Run>349011</Run>
<LBRange Start="192" End="200"/>
<LBRange Start="238" End="316"/>
<LBRange Start="329" End="334"/>
<LBRange Start="336" End="523"/>
<LBRange Start="561" End="578"/>
</LumiBlockCollection>
```

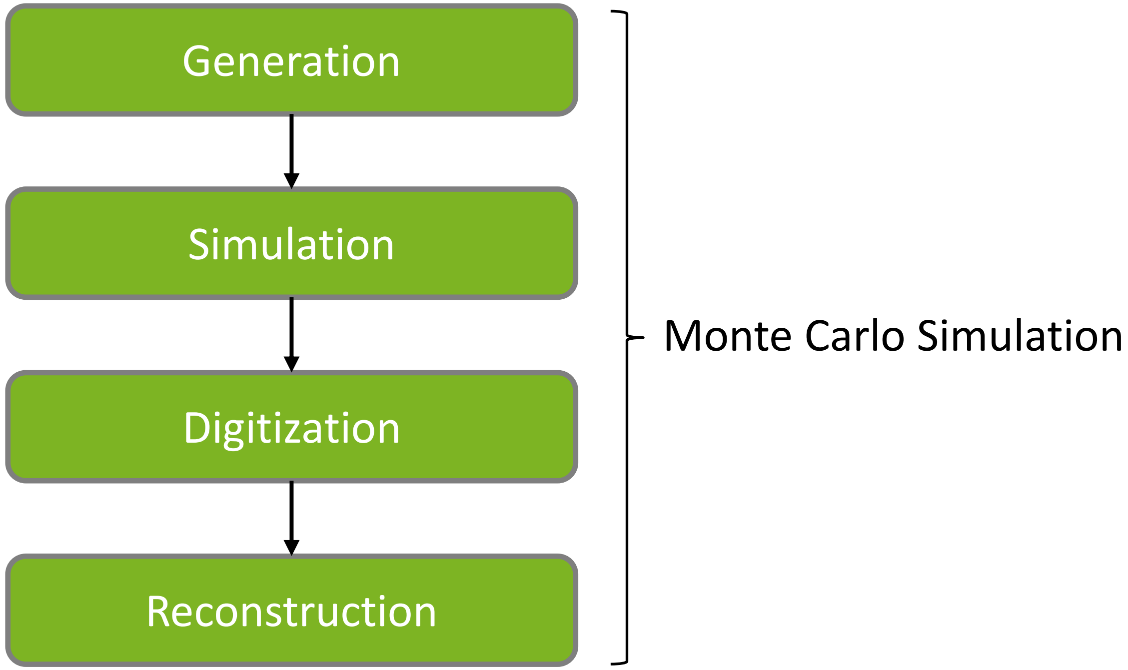
One Run, range of good lumi blocks in the run

Monte Carlo Production

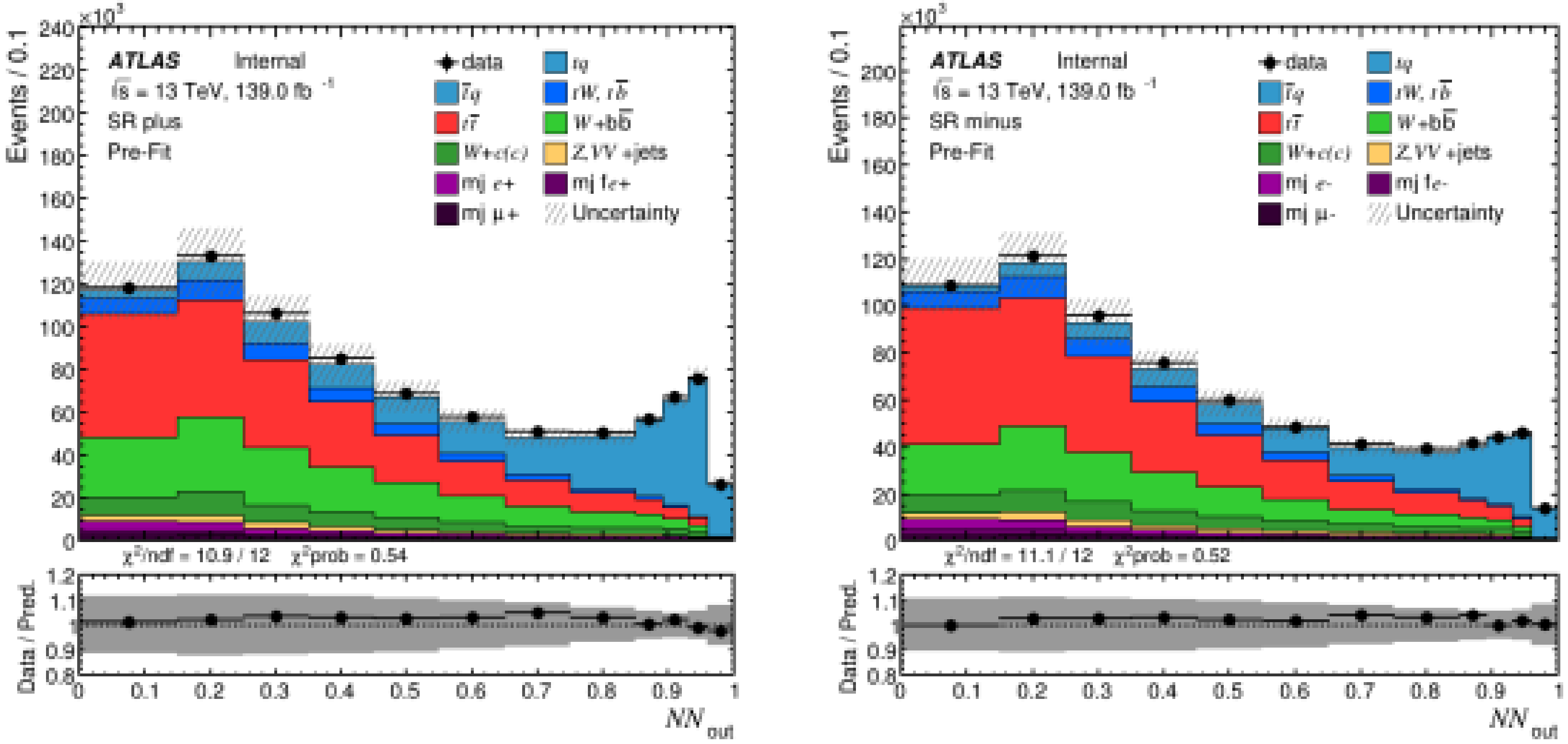
Monte Carlo Production

To…

- Understand performance of the detector
- Investigate reconstruction efficencies
- Modell expected backgrounds
- Modell signal processes
- Train multivariate classifiers (BDT,NN,DNN)
- Modell systematic uncertainties



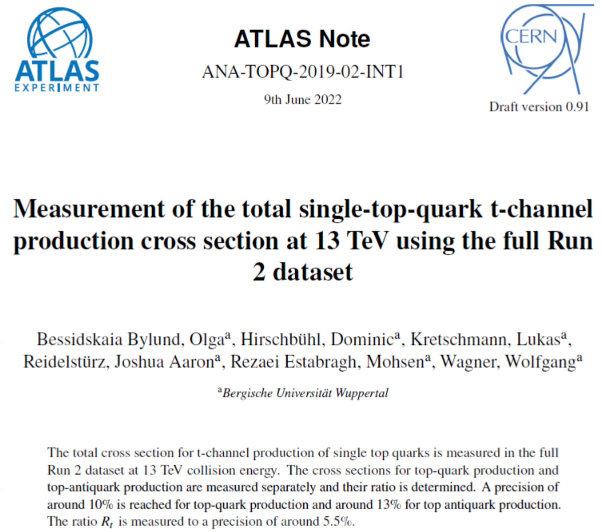
Monte Carlo Production



(a) ℓ^+ channel in the 2 jet 1 b-tag region

(b) ℓ^- channel in the 2 jet 1 b-tag region

Figure 6: Observed signal and simulated background neural network output distribution normalized to the number of expected events for the signal region in the plus (a) and the minus (b) channel. The uncertainty band represents the uncertainty due to statistical and systematic uncertainties in each bin. The ratio of observed to predicted (Pred.) number of events in each bin is shown in the lower histogram.



ATLAS Note
ANA-TOPQ-2019-02-INT1
9th June 2022

Draft version 0.91

Measurement of the total single-top-quark t-channel production cross section at 13 TeV using the full Run 2 dataset

Bessidskaia Bylund, Olga[a], Hirschbühl, Dominic[a], Kretschmann, Lukas[a], Reidelstürz, Joshua Aaron[a], Rezaei Estabragh, Mohsen[a], Wagner, Wolfgang[a]

[a] *Bergische Universität Wuppertal*

The total cross section for t-channel production of single top quarks is measured in the full Run 2 dataset at 13 TeV collision energy. The cross sections for top-quark production and top-antiquark production are measured separately and their ratio is determined. A precision of around 10% is reached for top-quark production and around 13% for top antiquark production. The ratio R_t is measured to a precision of around 5.5%.

Monte Carlo Production

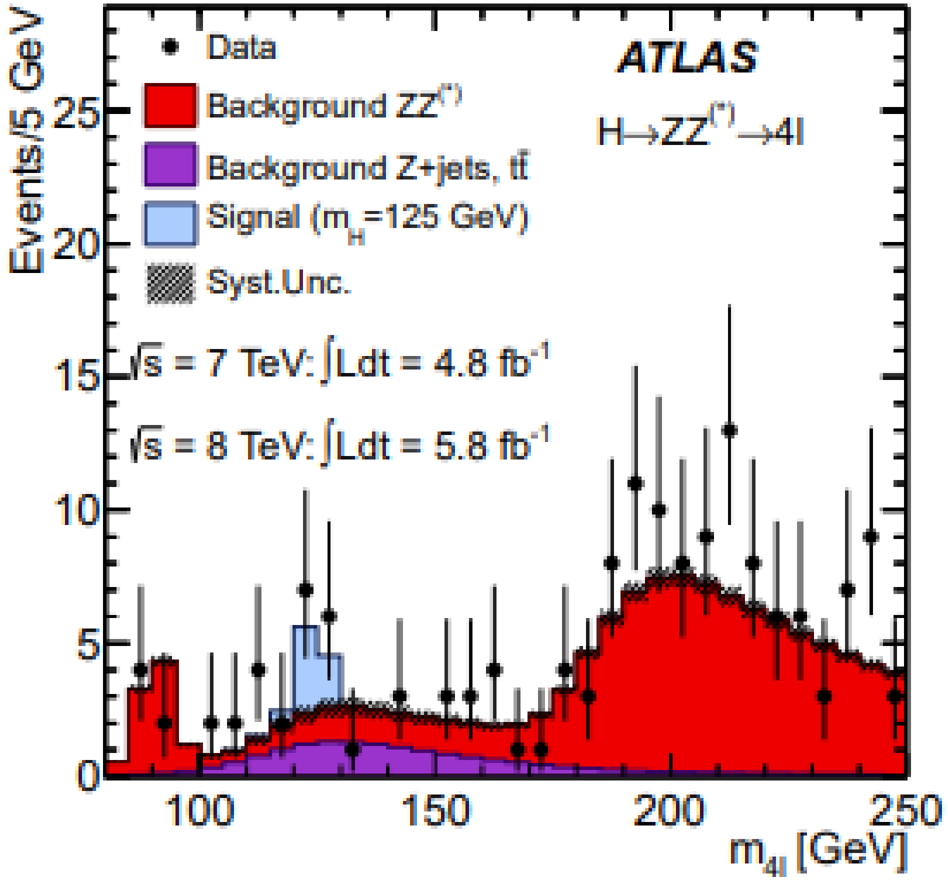


Figure 2: The distribution of the four-lepton invariant mass, $m_{4\ell}$, for the selected candidates, compared to the background expectation in the 80–250 GeV mass range, for the combination of the $\sqrt{s}$ = 7 TeV and $\sqrt{s}$ = 8 TeV data. The signal expectation for a SM Higgs with m_H = 125 GeV is also shown.

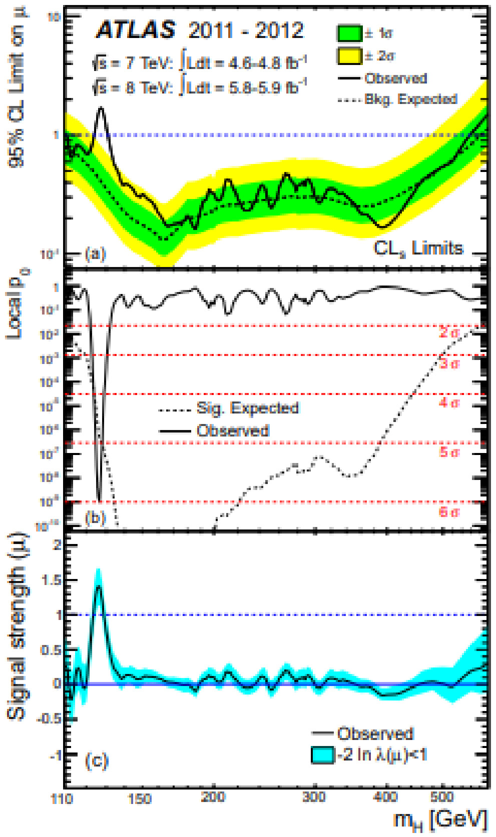
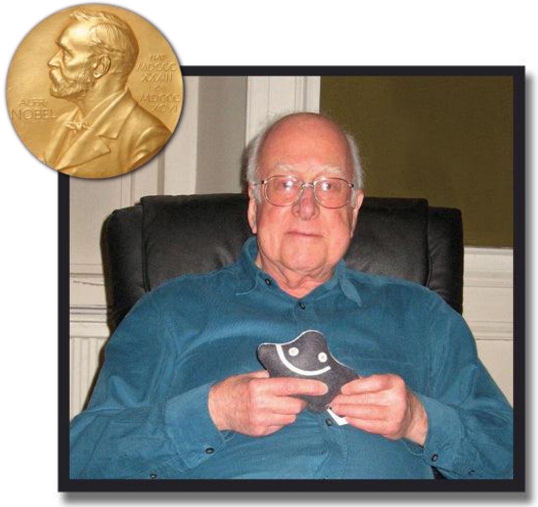


Figure 7: Combined search results: (a) The observed (solid) 95% CL limits on the signal strength as a function of m_H and the expectation (dashed) under the background-only hypothesis. The dark and light shaded bands show the $\pm1\sigma$ and $\pm2\sigma$ uncertainties on the background-only expectation. (b) The observed (solid) local p_0 as a function of m_H and the expectation (dashed) for a SM Higgs boson signal hypothesis (μ = 1) at the given mass. (c) The best-fit signal strength $\hat{\mu}$ as a function of m_H. The band indicates the approximate 68% CL interval around the fitted value.

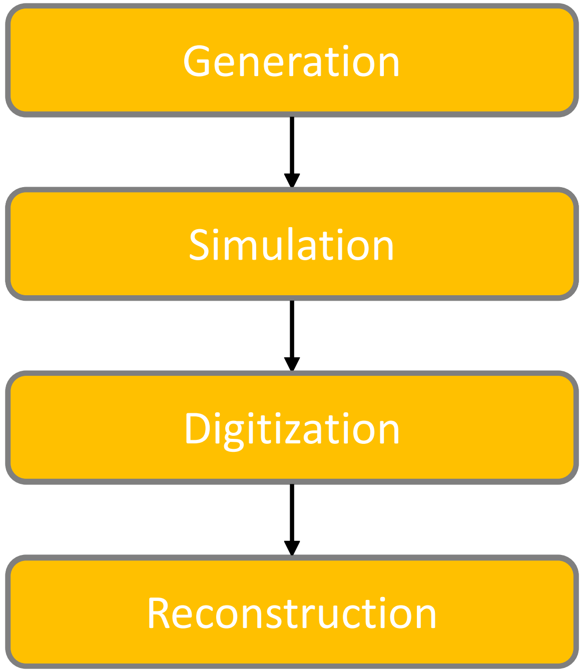


"Professor Higgs loves his boson, he will use it as a paperweight."

Monte Carlo Production on the GRID

- Eventgeneration (EvGen):
  Simulation of the interaction between the quarks and gluons in the colliding protons, the subsequent parton showering and hadronization and decays into stable particles (Pythia/Herwig)
- Detector Simulation:
  Calculation of how the particles from the generator interact with the detector, how they shower into secondaries, how much energy they deposit (Delphes/Geant4)
- Digitisation:
  Tuning the simulated energy deposits into a detector response that looks like raw data from the real detector

→ Then: same processing as for real data



Generation

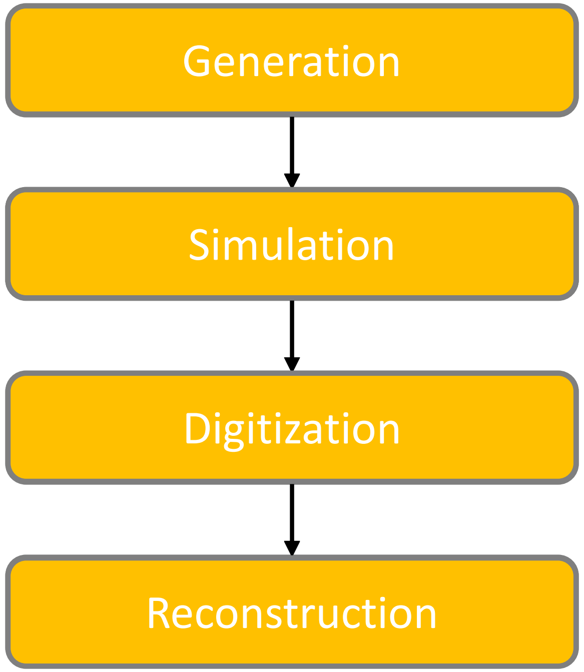
Simulation

Digitization

Reconstruction

Monte Carlo Production on the GRID

- Analysis data MC data look the same but in MC the original generated events („truth“) are available as well as reconstructed objects („reco“)
- Low momentum events must be injected into the chain to simulate the presence of multiple proton-proton collisions („pile-up“) in a given bunch crossing
  - Average number of collisions per bunch crossing is a function of the LHC parameters, unknown in Monte Carlo production
  - MC events are reweighted to account for discrepancies between real and simulated pileup

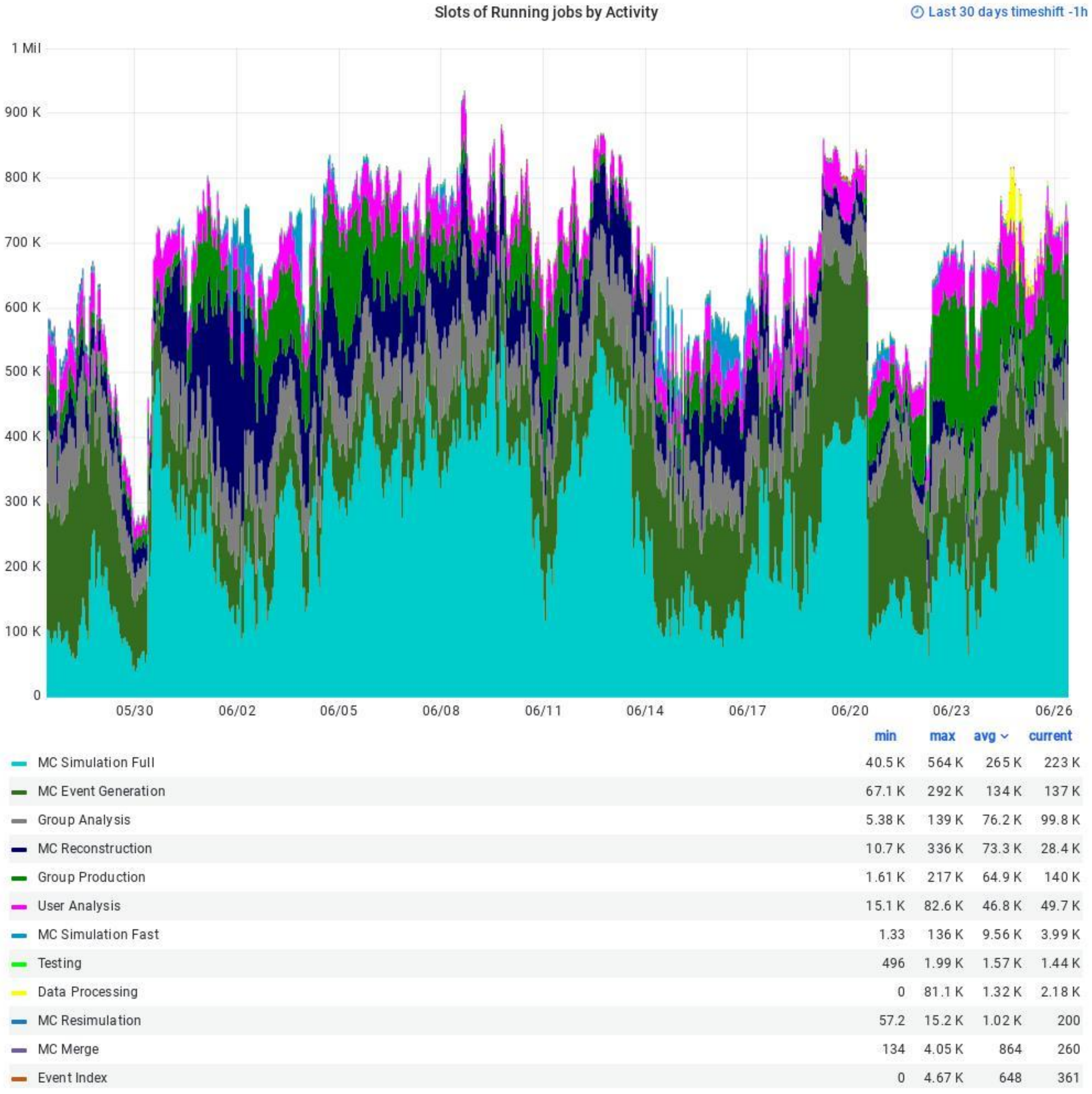

Generation

Simulation

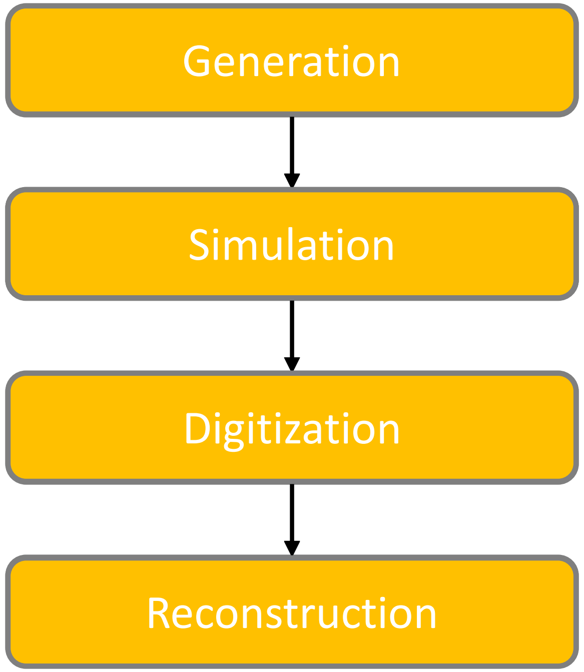
Digitization

Reconstruction

Monte Carlo Production on the GRID



	min	max	avg ⌄	current
MC Simulation Full	40.5 K	564 K	265 K	223 K
MC Event Generation	67.1 K	292 K	134 K	137 K
Group Analysis	5.38 K	139 K	76.2 K	99.8 K
MC Reconstruction	10.7 K	336 K	73.3 K	28.4 K
Group Production	1.61 K	217 K	64.9 K	140 K
User Analysis	15.1 K	82.6 K	46.8 K	49.7 K
MC Simulation Fast	1.33	136 K	9.56 K	3.99 K
Testing	496	1.99 K	1.57 K	1.44 K
Data Processing	0	81.1 K	1.32 K	2.18 K
MC Resimulation	57.2	15.2 K	1.02 K	200
MC Merge	134	4.05 K	864	260
Event Index	0	4.67 K	648	361



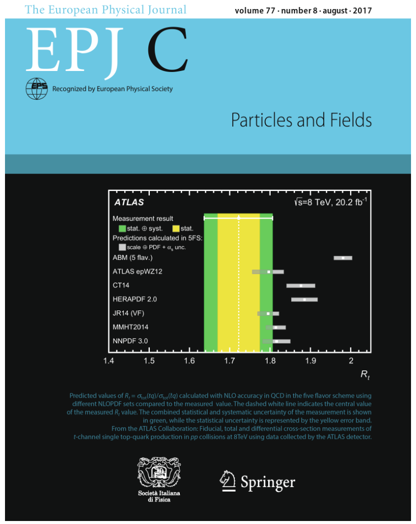
Generation

Simulation

Digitization

Reconstruction

Diving into analysis



The European Physical Journal

volume 77 · number 8 · august · 2017

EPJ C

Recognized by European Physical Society

Particles and Fields

Predicted values of $R_t = \sigma_{tot}(tq)/\sigma_{tot}(\bar{t}q)$ calculated with NLO accuracy in QCD in the five flavor scheme using different NLOPDF sets compared to the measured value. The dashed white line indicates the central value of the measured R_t value. The combined statistical and systematic uncertainty of the measurement is shown in green, while the statistical uncertainty is represented by the yellow error band.

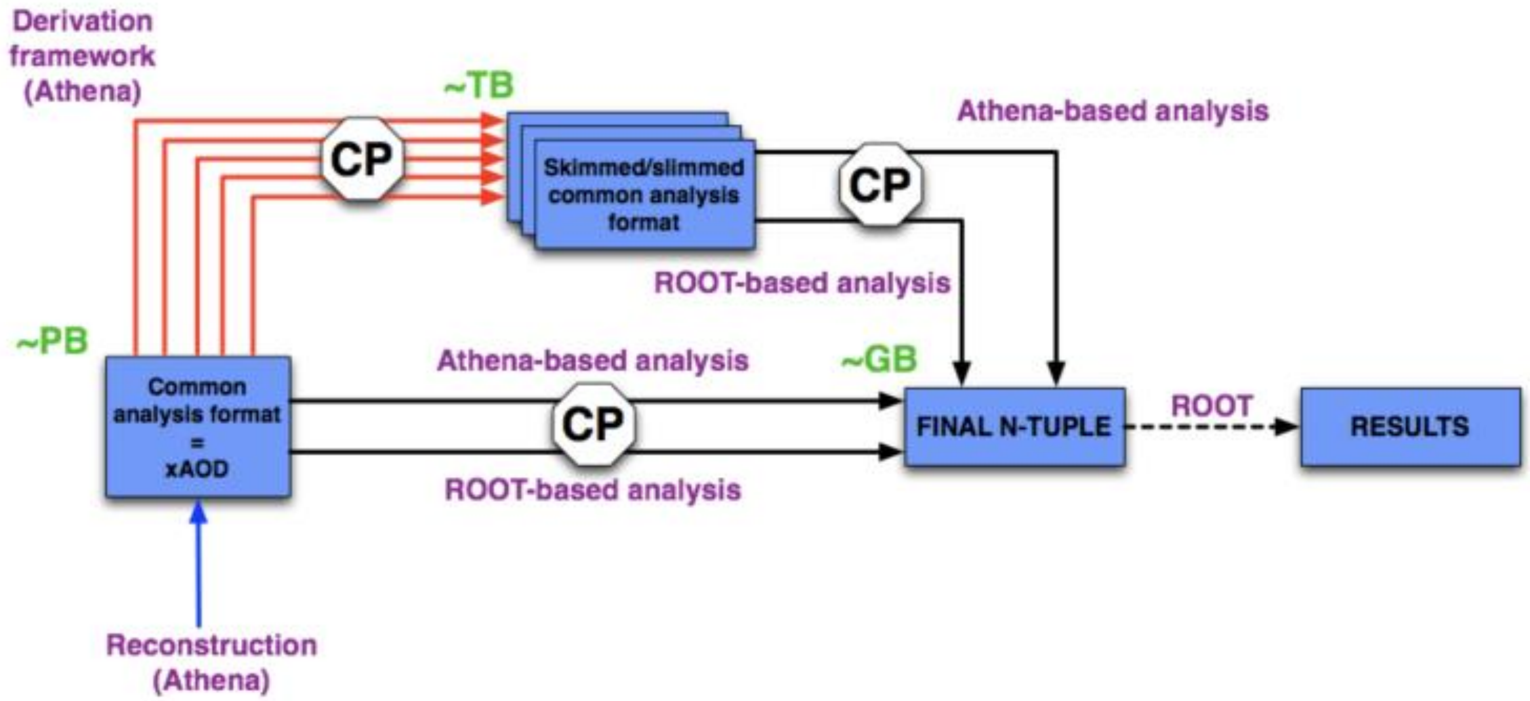
From the ATLAS Collaboration: Fiducial, total and differential cross-section measurements of *t*-channel single top-quark production in *pp* collisions at 8TeV using data collected by the ATLAS detector.

Società Italiana di Fisica

Springer

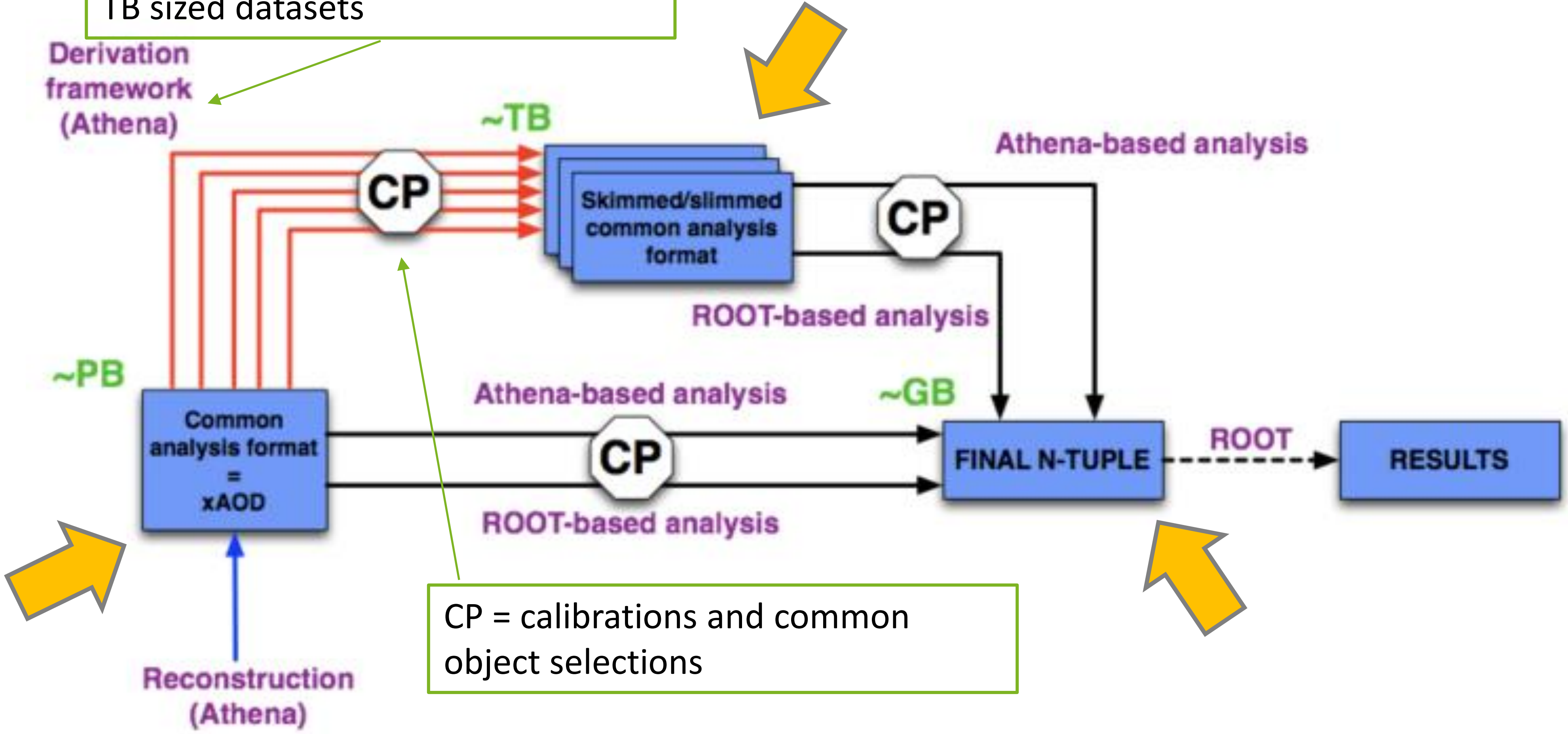
We are coming closer…

ATLAS analysis model



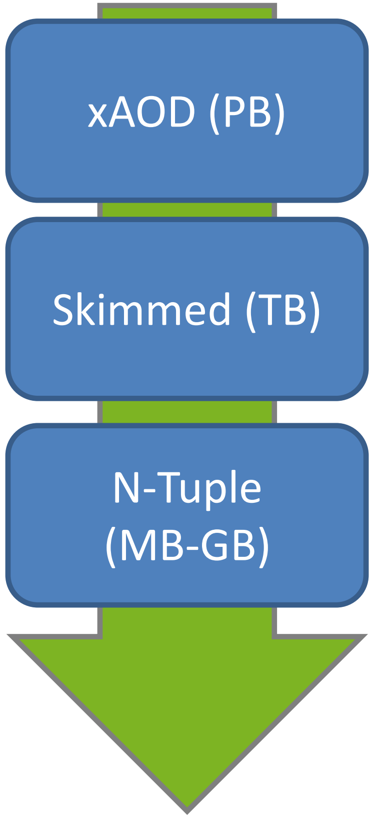
- xAOD: ROOT readable data format produced by reconstruction algorithms
- Derivation framework: AODs too big to analyse, centrally produced according to needs of physics groups, DAODs can then be analysed
- Analysis framework: used in the end-user analysis, reads in xAOD objects and applies all tools from CP (combined performance) groups [Athena/AthAnalysisBase]
- University groups often have their own analysis framework (Wuppertal: Kolophon/Arachne)

ATLAS analysis model



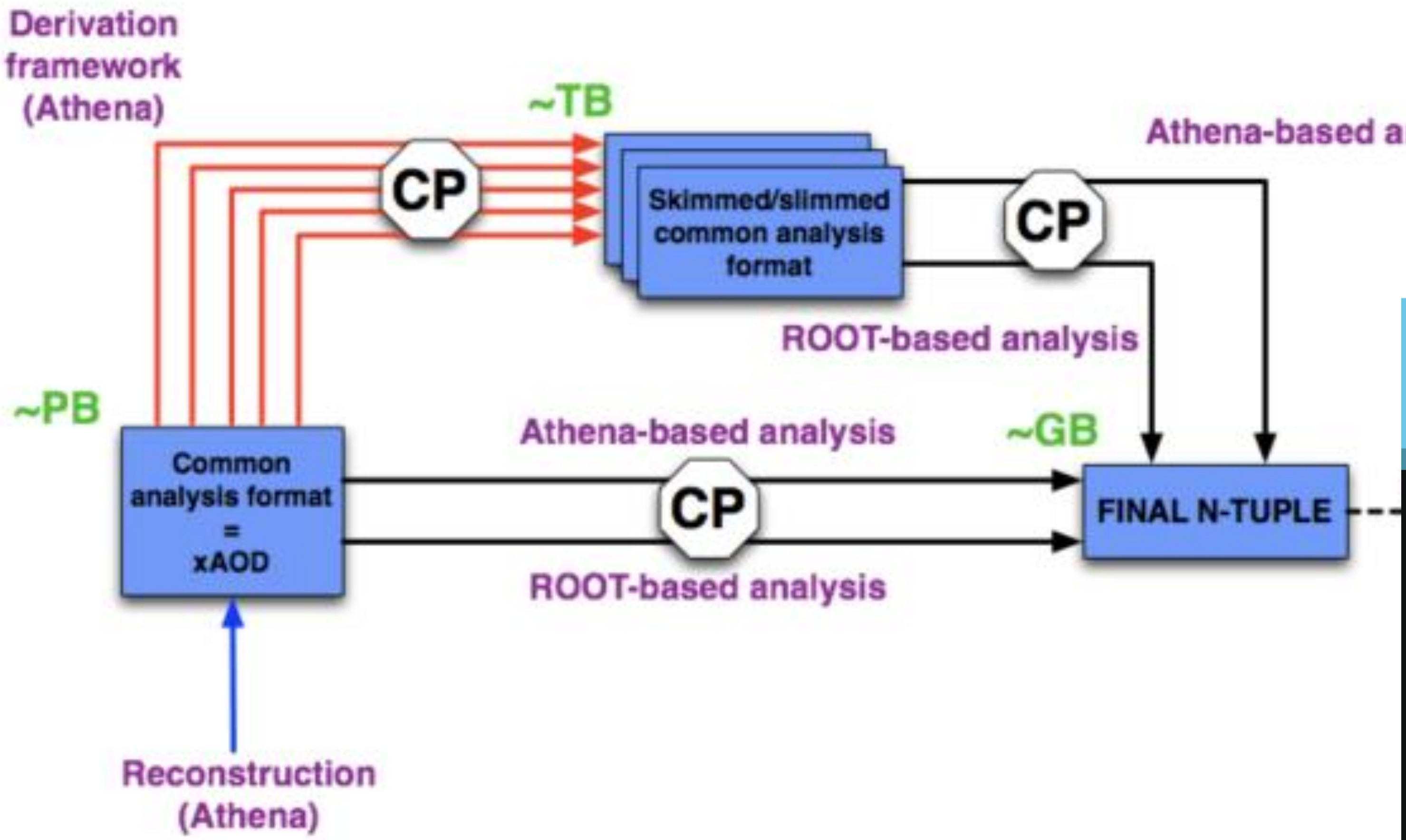
Derivation framework reduces PB to TB sized datasets

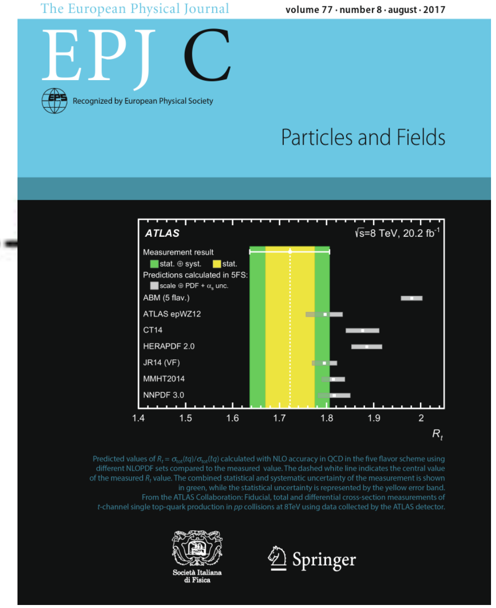
CP = calibrations and common object selections



- xAOD: information about the event, information about reconstructed objects (tracks, muons, electrons, jets….

ATLAS analysis model





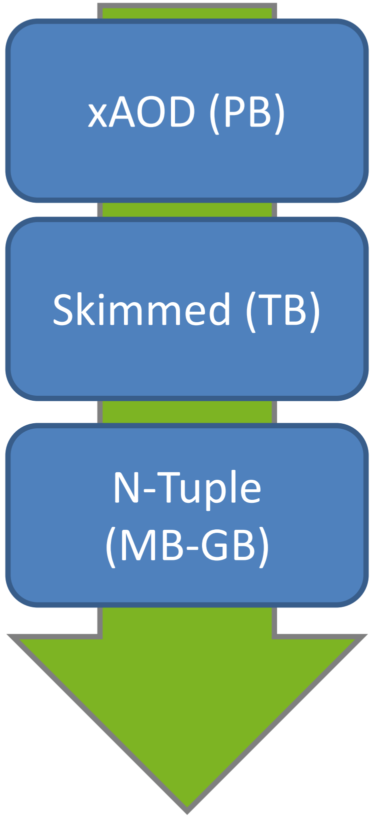
The European Physical Journal volume 77 · number 8 · august · 2017

EPJ C

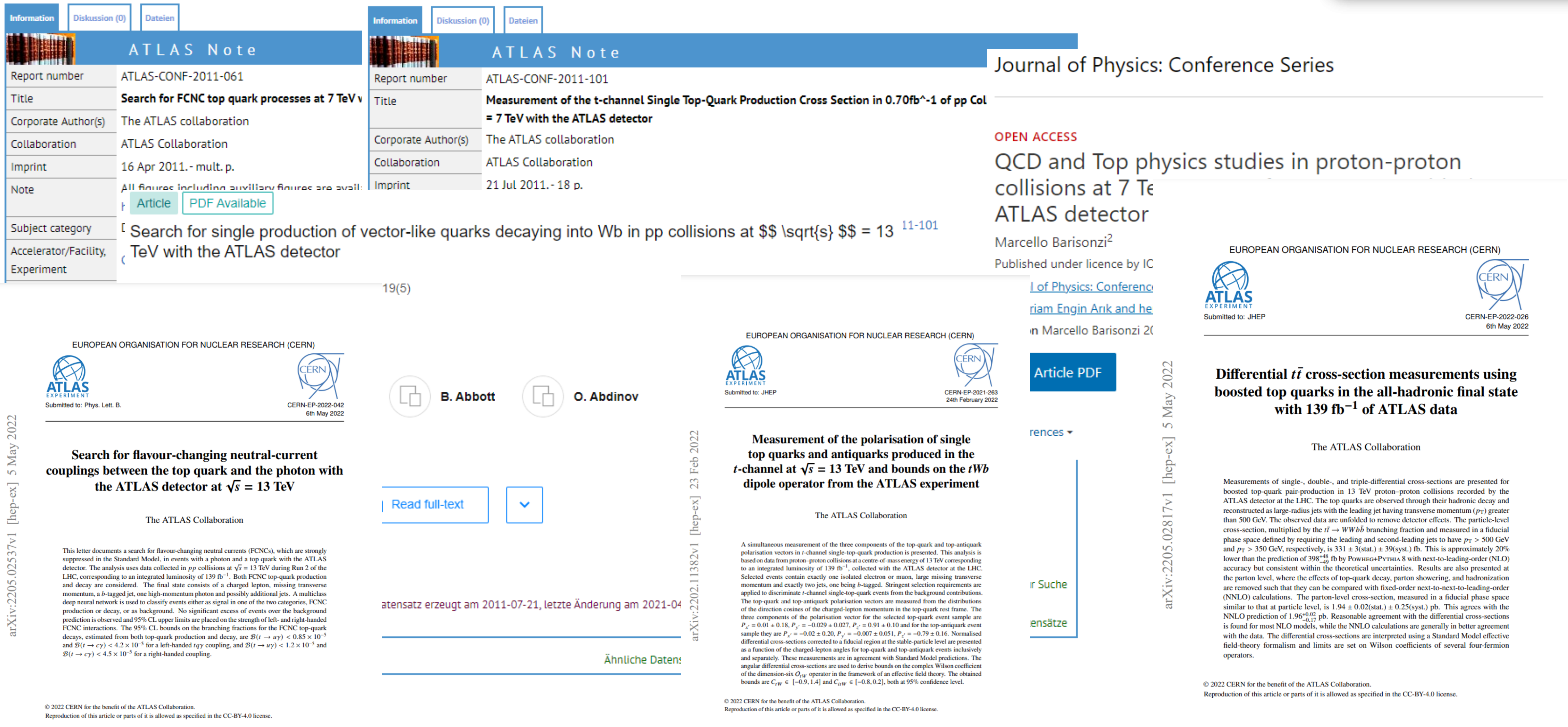
Recognized by European Physical Society

Particles and Fields

Springer



Finally… results!



ATLAS Note

Report number	ATLAS-CONF-2011-061
Title	Search for FCNC top quark processes at 7 TeV v
Corporate Author(s)	The ATLAS collaboration
Collaboration	ATLAS Collaboration
Imprint	16 Apr 2011. - mult. p.
Note	All figures including auxiliary figures are avail
Subject category	
Accelerator/Facility, Experiment	

ATLAS Note

Report number	ATLAS-CONF-2011-101
Title	Measurement of the t-channel Single Top-Quark Production Cross Section in 0.70fb^-1 of pp Col = 7 TeV with the ATLAS detector
Corporate Author(s)	The ATLAS collaboration
Collaboration	ATLAS Collaboration
Imprint	21 Jul 2011. - 18 p.

Journal of Physics: Conference Series

OPEN ACCESS

QCD and Top physics studies in proton-proton collisions at 7 Te
ATLAS detector

Marcello Barisonzi[2]

Published under licence by IC

l of Physics: Conferenc
riam Engin Arık and he
n Marcello Barisonzi 2(

Search for single production of vector-like quarks decaying into Wb in pp collisions at $$ \sqrt{s} $$ = 13 TeV with the ATLAS detector

11-101

19(5)

B. Abbott
O. Abdinov

atensatz erzeugt am 2011-07-21, letzte Änderung am 2021-04

Ähnliche Datens

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: Phys. Lett. B.

CERN-EP-2022-042
6th May 2022

arXiv:2205.02537v1 [hep-ex] 5 May 2022

Search for flavour-changing neutral-current couplings between the top quark and the photon with the ATLAS detector at $\sqrt{s}$ = 13 TeV

The ATLAS Collaboration

This letter documents a search for flavour-changing neutral currents (FCNCs), which are strongly suppressed in the Standard Model, in events with a photon and a top quark with the ATLAS detector. The analysis uses data collected in pp collisions at $\sqrt{s} = 13$ TeV during Run 2 of the LHC, corresponding to an integrated luminosity of 139 fb^{-1}. Both FCNC top-quark production and decay are considered. The final state consists of a charged lepton, missing transverse momentum, a b-tagged jet, one high-momentum photon and possibly additional jets. A multiclass deep neural network is used to classify events either as signal in one of the two categories, FCNC production or decay, or as background. No significant excess of events over the background prediction is observed and 95% CL upper limits are placed on the strength of left- and right-handed FCNC interactions. The 95% CL bounds on the branching fractions for the FCNC top-quark decays, estimated from both top-quark production and decay, are $\mathcal{B}(t \to u\gamma) < 0.85 \times 10^{-5}$ and $\mathcal{B}(t \to c\gamma) < 4.2 \times 10^{-5}$ for a left-handed $tq\gamma$ coupling, and $\mathcal{B}(t \to u\gamma) < 1.2 \times 10^{-5}$ and $\mathcal{B}(t \to c\gamma) < 4.5 \times 10^{-5}$ for a right-handed coupling.

© 2022 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: JHEP

CERN-EP-2021-263
24th February 2022

arXiv:2202.11382v1 [hep-ex] 23 Feb 2022

Measurement of the polarisation of single top quarks and antiquarks produced in the t-channel at $\sqrt{s}$ = 13 TeV and bounds on the tWb dipole operator from the ATLAS experiment

The ATLAS Collaboration

A simultaneous measurement of the three components of the top-quark and top-antiquark polarisation vectors in t-channel single-top-quark production is presented. This analysis is based on data from proton–proton collisions at a centre-of-mass energy of 13 TeV corresponding to an integrated luminosity of 139 fb^{-1}, collected with the ATLAS detector at the LHC. Selected events contain exactly one isolated electron or muon, large missing transverse momentum and exactly two jets, one being b-tagged. Stringent selection requirements are applied to discriminate t-channel single-top-quark events from the background contributions. The top-quark and top-antiquark polarisation vectors are measured from the distributions of the direction cosines of the charged-lepton momentum in the top-quark rest frame. The three components of the polarisation vector for the selected top-quark event sample are $P_{x'} = 0.01 \pm 0.18$, $P_{y'} = -0.029 \pm 0.027$, $P_{z'} = 0.91 \pm 0.10$ and for the top-antiquark event sample they are $P_{x'} = -0.02 \pm 0.20$, $P_{y'} = -0.007 \pm 0.051$, $P_{z'} = -0.79 \pm 0.16$. Normalised differential cross-sections corrected to a fiducial region at the stable-particle level are presented as a function of the charged-lepton angles for top-quark and top-antiquark events inclusively and separately. These measurements are in agreement with Standard Model predictions. The angular differential cross-sections are used to derive bounds on the complex Wilson coefficient of the dimension-six O_{tW} operator in the framework of an effective field theory. The obtained bounds are $C_{tW} \in [-0.9, 1.4]$ and $C_{itW} \in [-0.8, 0.2]$, both at 95% confidence level.

© 2022 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: JHEP

CERN-EP-2022-026
6th May 2022

arXiv:2205.02817v1 [hep-ex] 5 May 2022

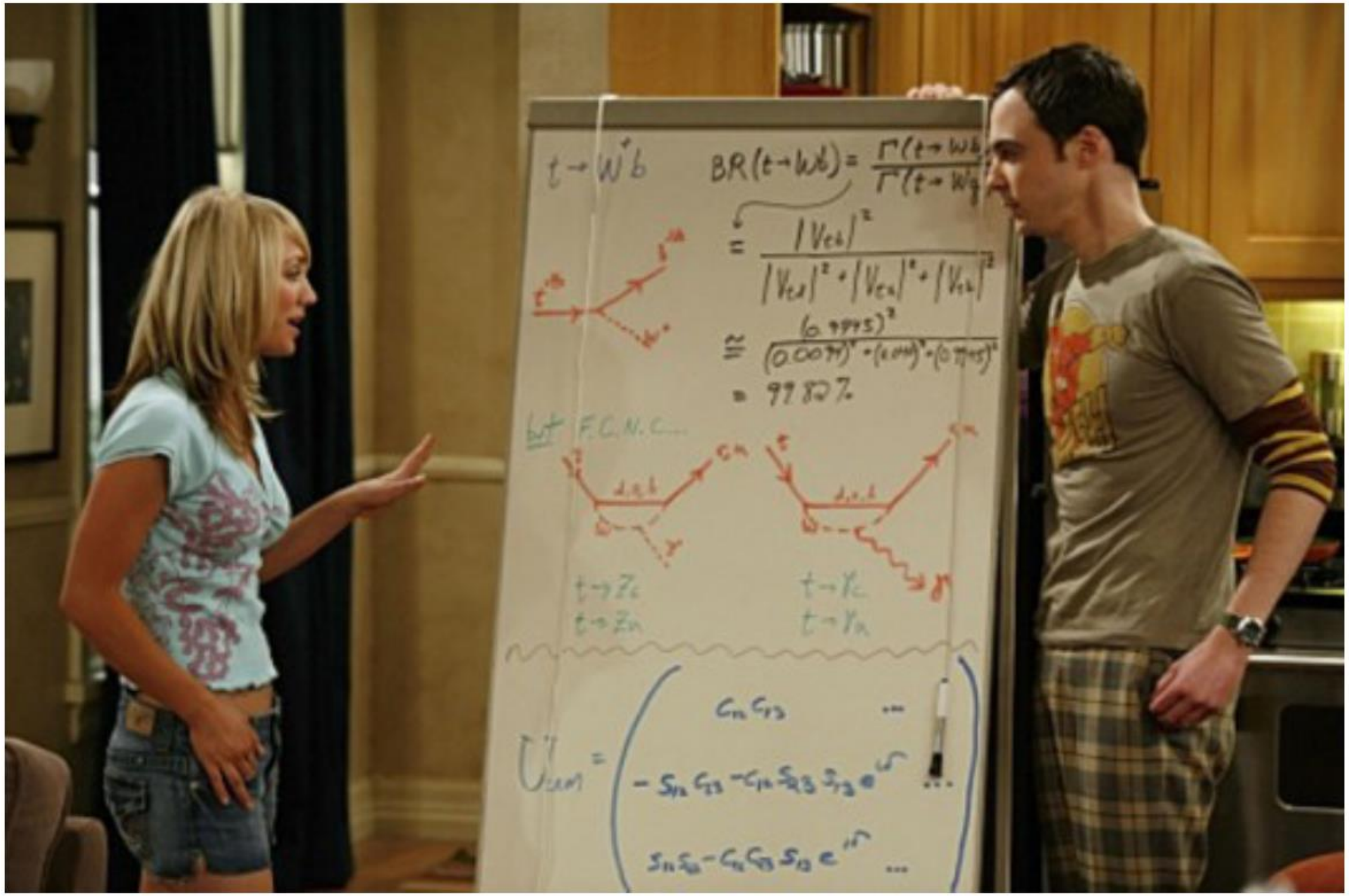
Differential $t\bar{t}$ cross-section measurements using boosted top quarks in the all-hadronic final state with 139 fb^{-1} of ATLAS data

The ATLAS Collaboration

Measurements of single-, double-, and triple-differential cross-sections are presented for boosted top-quark pair-production in 13 TeV proton–proton collisions recorded by the ATLAS detector at the LHC. The top quarks are observed through their hadronic decay and reconstructed as large-radius jets with the leading jet having transverse momentum (p_{T}) greater than 500 GeV. The observed data are unfolded to remove detector effects. The particle-level cross-section, multiplied by the $t\bar{t} \to WWb\bar{b}$ branching fraction and measured in a fiducial phase space defined by requiring the leading and second-leading jets to have $p_{\mathrm{T}} > 500$ GeV and $p_{\mathrm{T}} > 350$ GeV, respectively, is 331 ± 3(stat.) ± 39(syst.) fb. This is approximately 20% lower than the prediction of 398^{+48}_{-49} fb by Powheg+Pythia 8 with next-to-leading-order (NLO) accuracy but consistent within the theoretical uncertainties. Results are also presented at the parton level, where the effects of top-quark decay, parton showering, and hadronization are removed such that they can be compared with fixed-order next-to-next-to-leading-order (NNLO) calculations. The parton-level cross-section, measured in a fiducial phase space similar to that at particle level, is 1.94 ± 0.02(stat.) ± 0.25(syst.) pb. This agrees with the NNLO prediction of $1.96^{+0.02}_{-0.17}$ pb. Reasonable agreement with the differential cross-sections is found for most NLO models, while the NNLO calculations are generally in better agreement with the data. The differential cross-sections are interpreted using a Standard Model effective field-theory formalism and limits are set on Wilson coefficients of several four-fermion operators.

© 2022 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

Finally… results!



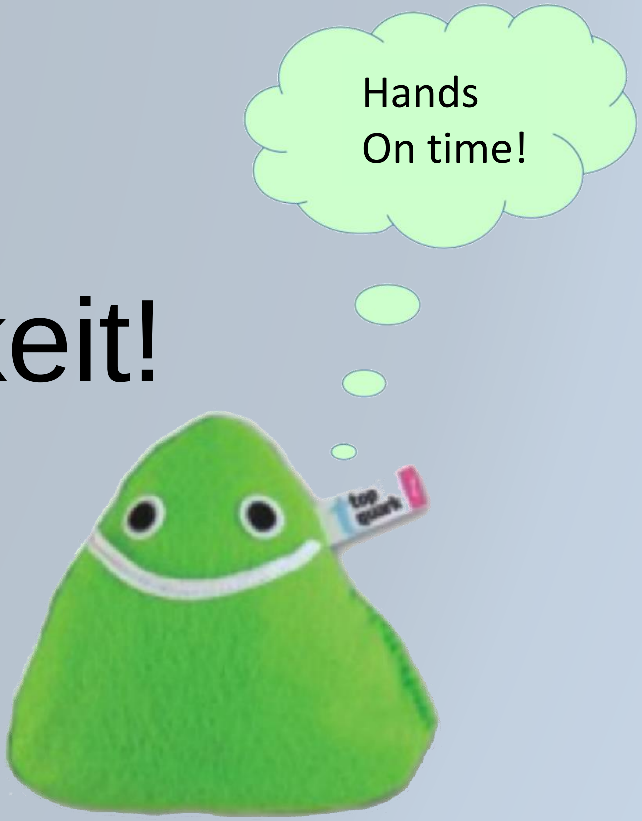
And sometimes this research even goes Hollywood…

https://arxiv.org/abs/1203.0529

Fellow-Treffen 2022 – Würzburg

18.-21.08.2022

Vielen Dank für Ihre Aufmerksamkeit!



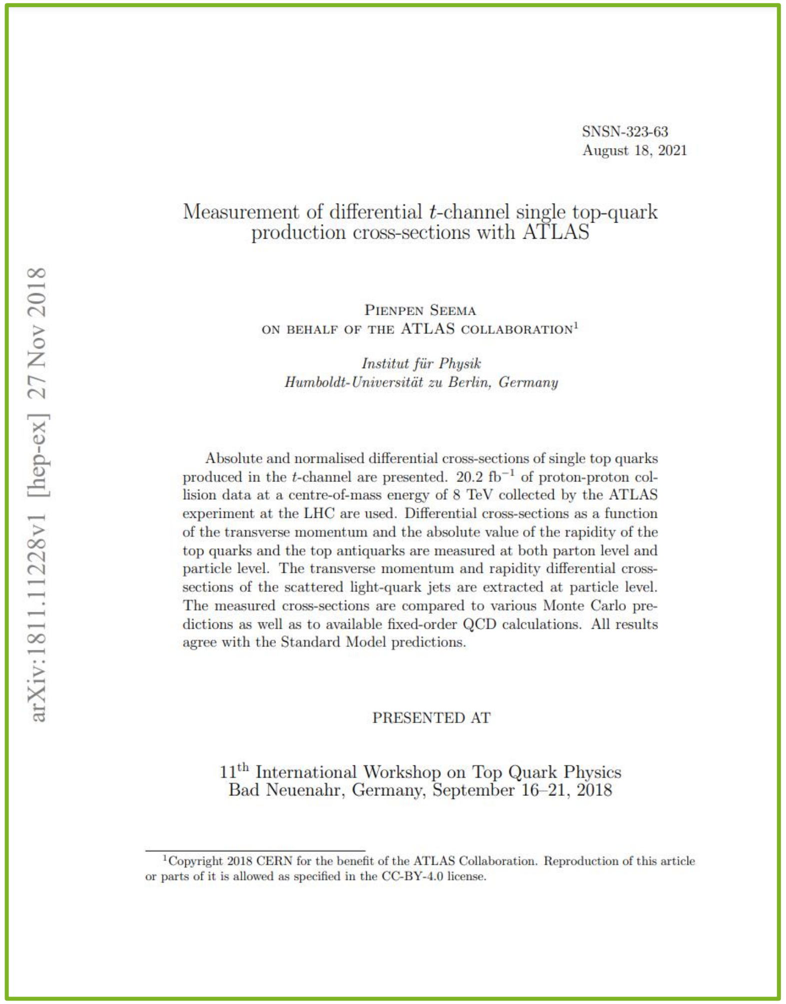
Back-

My current research – Unfolding Studies

Introduction

- Unfolding is to complicated to explain this procedure in under one hour
- Therefore I will focus on a more practical point-of-view
- I will explain the terminology that is used and explain some other core features of unfolding

➢ There are no stupid questions, so please ask them whenever you want!



SNSN-323-63
August 18, 2021

Measurement of differential t-channel single top-quark production cross-sections with ATLAS

PIENPEN SEEMA
ON BEHALF OF THE ATLAS COLLABORATION[1]

Institut für Physik
Humboldt-Universität zu Berlin, Germany

arXiv:1811.11228v1 [hep-ex] 27 Nov 2018

Absolute and normalised differential cross-sections of single top quarks produced in the t-channel are presented. 20.2 fb^{-1} of proton-proton collision data at a centre-of-mass energy of 8 TeV collected by the ATLAS experiment at the LHC are used. Differential cross-sections as a function of the transverse momentum and the absolute value of the rapidity of the top quarks and the top antiquarks are measured at both parton level and particle level. The transverse momentum and rapidity differential cross-sections of the scattered light-quark jets are extracted at particle level. The measured cross-sections are compared to various Monte Carlo predictions as well as to available fixed-order QCD calculations. All results agree with the Standard Model predictions.

PRESENTED AT

11th International Workshop on Top Quark Physics
Bad Neuenahr, Germany, September 16–21, 2018

---

[1] Copyright 2018 CERN for the benefit of the ATLAS Collaboration. Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

More in depth overview of unfolding:
https://indico.cern.ch/event/794004/contributions/3326125/attachments/1802168/2939911/UnfoldingReview.pdf

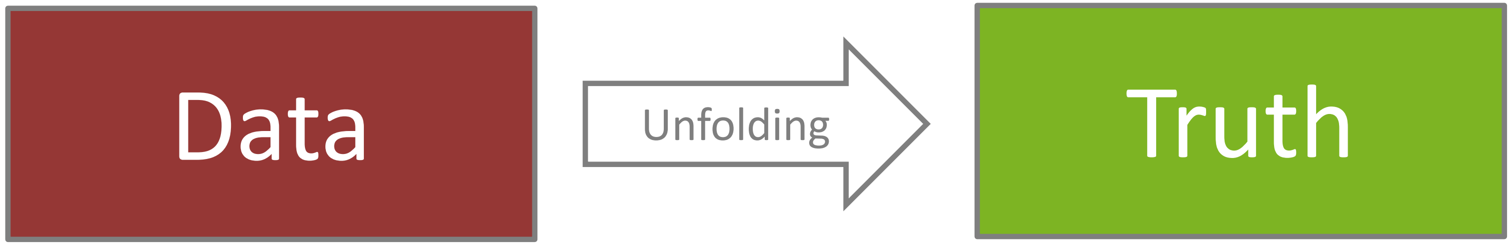
https://indico.cern.ch/event/795477/contributions/3378783/attachments/1848586/3033879/JayHowarth_ATLAS_TopWorkshop_UnfoldingTutorial.pdf

What is unfolding?

- You can also look at unfolding as a „deconvolution“ technique where we have some signal f convoluted with some noise N to give some data d

$$f * N = d$$

- In particle physics f and d are discrete data in the form of binned histograms and N is a combination of detector smearing and acceptance effects
- Unfolding is basically the process of coming from d to f
- f and d are histograms (vectors of data), N has the form of a matrix and the process of going from d to f is a matrix inversion problem



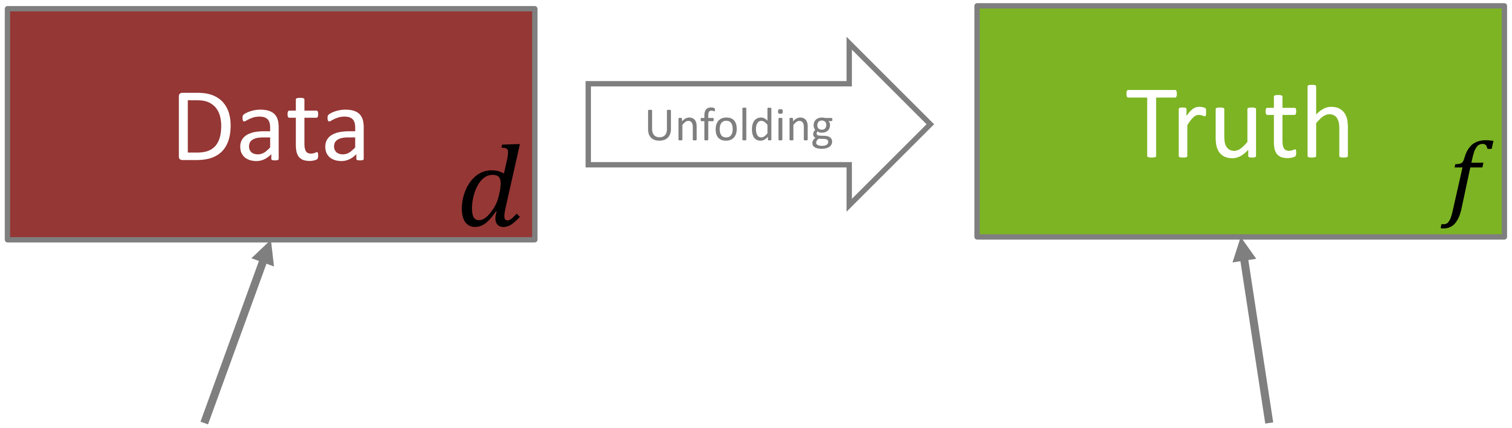
Data

Unfolding

Truth

What is the problem with matrix inversion?

- So what happens if we invert the matrix and apply it to our data vector?



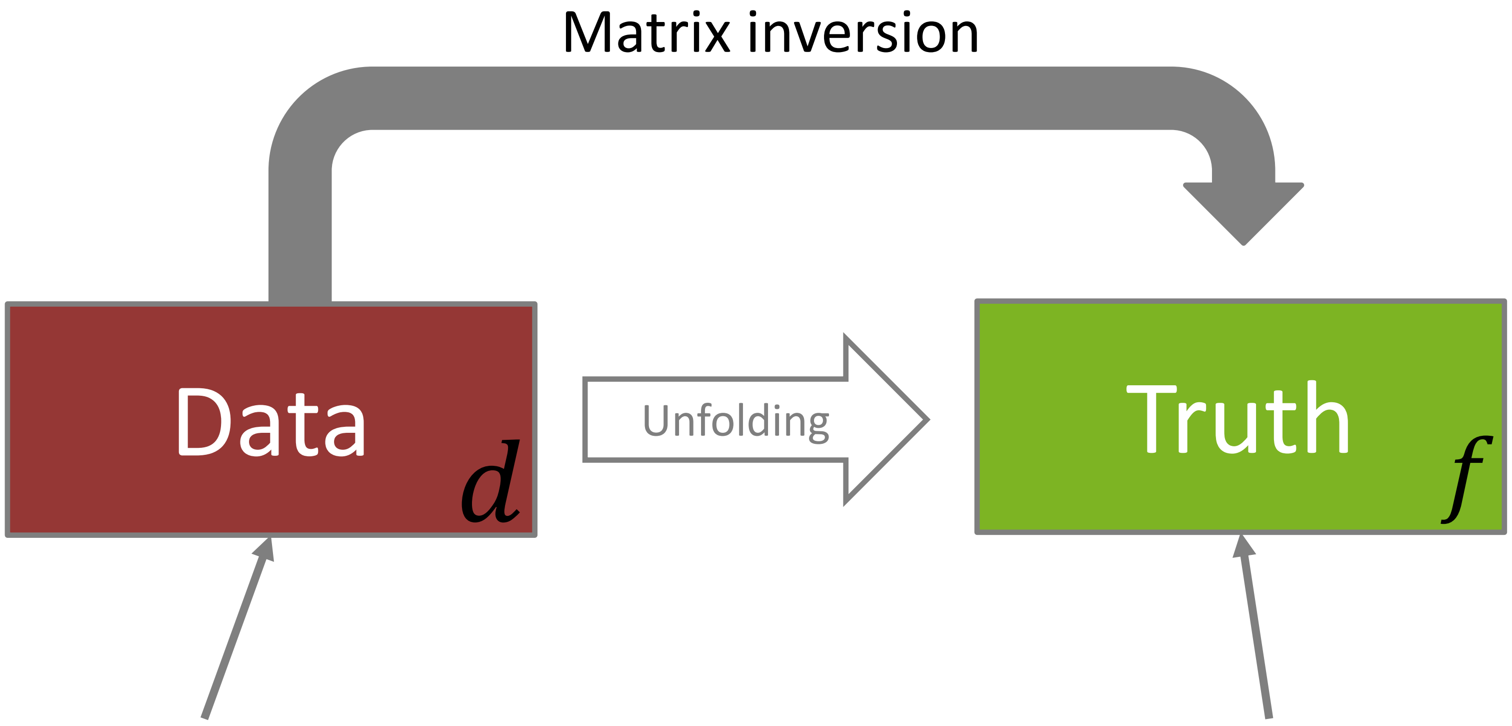
$$f * N = d$$

$$f = N^{-1} * d$$

$$V_f = N^{-1} V_d (N^{-1})^T$$

What is the problem with matrix inversion?

- So what happens if we invert the matrix and apply it to our data vector?



Matrix inversion

Data d

Unfolding

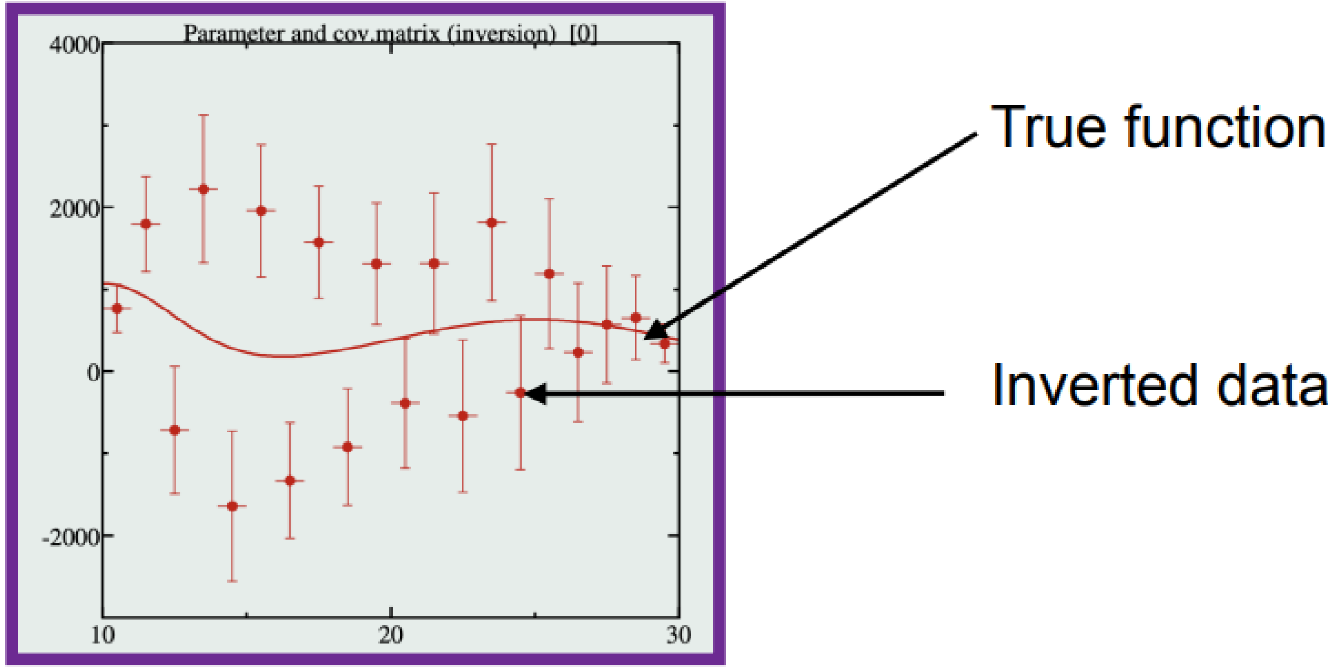
Truth f

$$f * N = d$$

$$f = N^{-1} * d$$

What is the problem with matrix inversion?

- This matrix is usally built from Monte Carlo (MC), we also call it the response matrix
- The big problem is, if we invert this matrix, statistical fluctuations get amplified and you get stupid results. We need to dampen (regularise) these fluctuations somehow, and this is what almost all unfolding methods do.



True function

Inverted data

Covariance provides insight into how two variables are related to one another. More precisely, **covariance refers to the measure of how two random variables in a data set will change together**. A positive covariance means that the two variables at hand are positively related, and they move in the same direction.

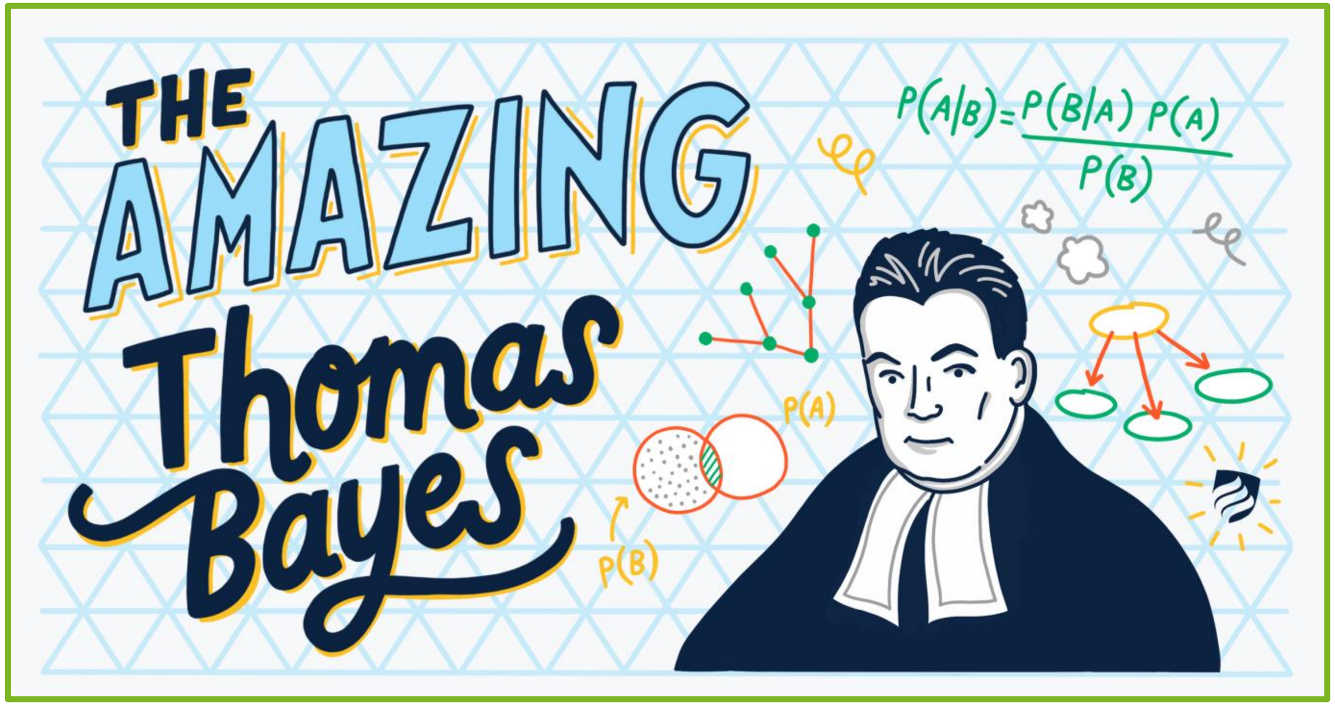
- Large anti-correlations in the covariance matrix allow this result
  (covariance-matrix gives covariance between each pair of elements of a given random vector)

What is the problem with matrix inversion?

- This matrix is usally built from Monte Carlo (MC), we also call it the response matrix
- The big problem is, if we invert this matrix, statistical fluctuations get amplified and you get stupid results. We need to dampen (regularise) these fluctuations somehow, and this is what almost all unfolding methods do.

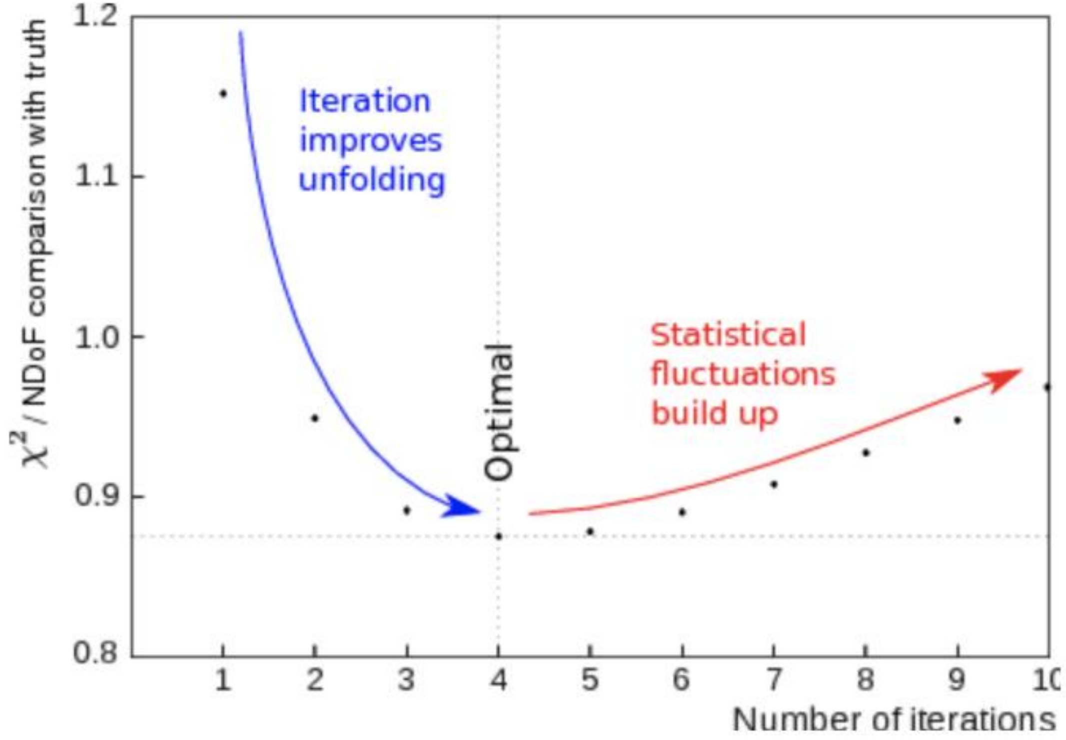
- There are multiple methods of unfolding to tackle this problem:
  - Iterative Bayes‘ theorem (IBU)
  - Singular Value Decomposition (SVD)
  - Bin-by-bin corrections
  - Matrix inversion
  - ….

- We use Iterative Bayesian Unfolding



Iterative Bayesian Unfolding

- Iterative Bayesian Unfolding is the most commonly used method
- In this method we use Bayes theorem recursiveley
- IBU has the unique feature that the more iterations you do, the weaker the regularisation gets (basically the unfolding gets worse)
- It is crucial to find the optimal stopping point



The master formula

DATA
(What we have)

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\mathcal{L} \cdot \Delta X^i \cdot \epsilon^i_{\mathrm{eff}}} \cdot \sum_j R^{-1}_{ij} \cdot f^j_{\mathrm{acc}} \cdot (N^j_{\mathrm{obs}} - N^j_{\mathrm{bkg}})$$

TRUTH
(What we want)

The master formula

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\mathcal{L} \cdot \Delta X^i \cdot \epsilon^i_{\mathrm{eff}}} \cdot \sum_j R^{-1}_{ij} \cdot f^j_{\mathrm{acc}} \cdot \underline{(N^j_{\mathrm{obs}} - N^j_{\mathrm{bkg}})}$$

1. Subtract the background events

The master formula

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\mathcal{L} \cdot \Delta X^i \cdot \epsilon^i_{\mathrm{eff}}} \cdot \sum_j R^{-1}_{ij} \cdot \underline{f^j_{\mathrm{acc}}} \cdot (N^j_{\mathrm{obs}} - N^j_{\mathrm{bkg}})$$

1. Subtract the background events
2. Correct for „non-fiducial“ events that are in the reco-level but not in truth (in-smearing)

$$f^j_{acc} = \frac{reco_j \wedge truth_j}{reco_j}$$

The master formula

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\mathcal{L} \cdot \Delta X^i \cdot \epsilon^i_{\mathrm{eff}}} \cdot \underline{\sum_j R^{-1}_{ij}} \cdot f^j_{\mathrm{acc}} \cdot (N^j_{\mathrm{obs}} - N^j_{\mathrm{bkg}})$$

1. Subtract the background events
2. Correct for „non-fiducial“ events that are in the reco-level but not in truth (in-smearing)
3. Remove the detector smearing

INSERT MIGRATION MATRIX HERE

The master formula

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\mathcal{L} \cdot \Delta X^i \cdot \underline{\epsilon^i_{\text{eff}}}} \cdot \sum_j R^{-1}_{ij} \cdot f^j_{\text{acc}} \cdot (N^j_{\text{obs}} - N^j_{\text{bkg}})$$

1. Subtract the background events
2. Correct for „non-fiducial“ events that are in the reco-level but not in truth (in-smearing)
3. Remove the detector smearing
4. Extrapolate to the truth phase-space (correct for events that are at truth level but not at reco)

$$\epsilon^i_{eff} = \frac{reco_j \wedge truth_j}{truth_j}$$

The master formula

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\underline{\mathcal{L} \cdot \Delta X^i} \cdot \epsilon^i_{\mathrm{eff}}} \cdot \sum_j R^{-1}_{ij} \cdot f^j_{\mathrm{acc}} \cdot (N^j_{\mathrm{obs}} - N^j_{\mathrm{bkg}})$$

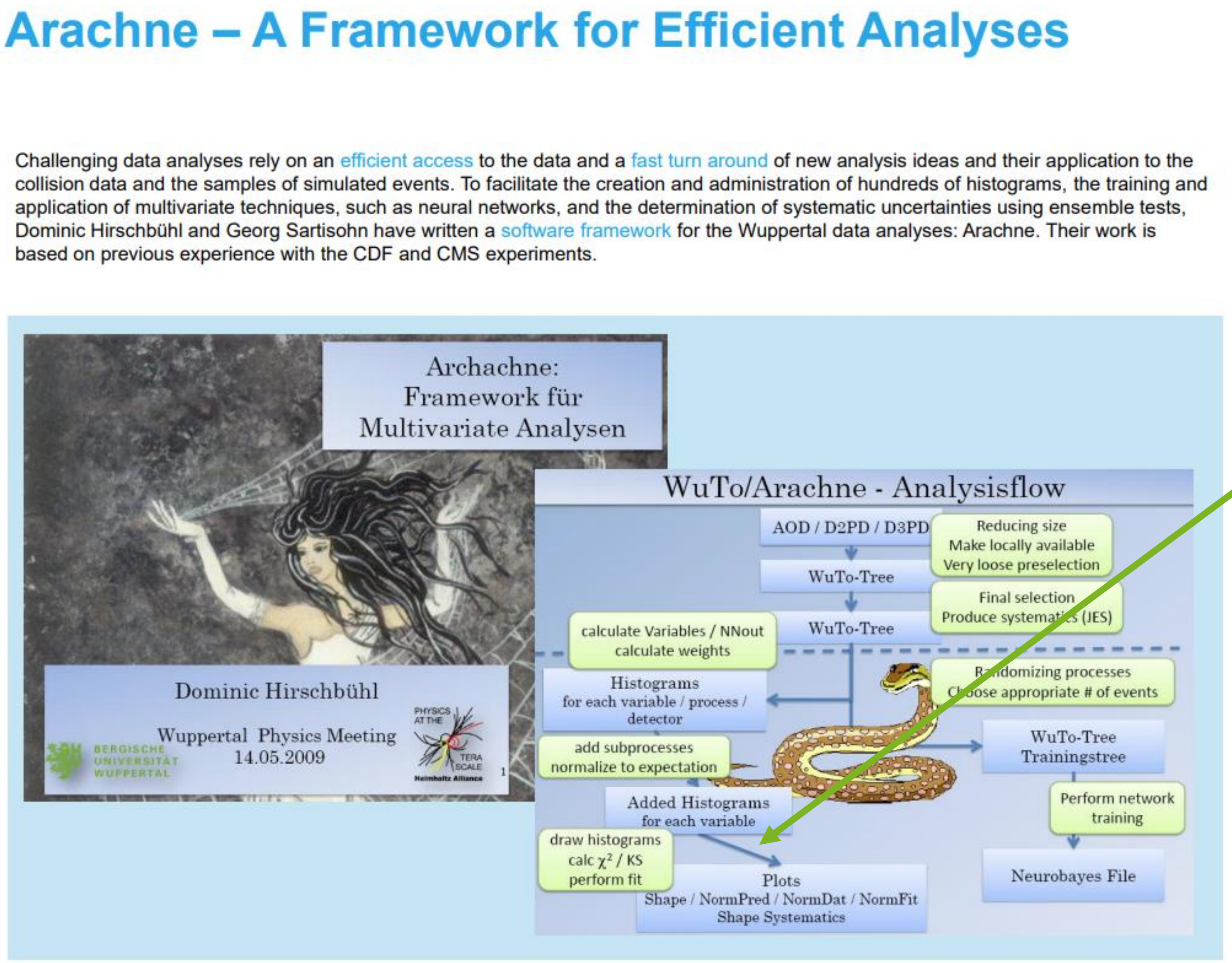
1. Subtract the background events
2. Correct for „non-fiducial“ events that are in the reco-level but not in truth (in-smearing)
3. Remove the detector smearing
4. Extrapolate to the truth phase-space (correct for events that are at truth level but not at reco)
5. Normalise to cross-section
6. Be happy!

The software

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\mathcal{L} \cdot \Delta X^i \cdot \epsilon^i_{\mathrm{eff}}} \cdot \sum_j R^{-1}_{ij} \cdot f^j_{\mathrm{acc}} \cdot (N^j_{\mathrm{obs}} - N^j_{\mathrm{bkg}})$$

- The core software for doing the unfolding is „RooUnfold“
- Me and two colleagues implemented RooUnfold in our groups analysis framework „Kolophon“ and „Arachne“
- Some serious developement is currently ongoing there to make the code future-proof, easy to use, implement new advanced functions (stress tests etc.) and plotting routines…
- More and more analysis are using it in our group, so we need to write documentations and make sure that it is easy to use for the individual analyser

The software



Arachne – A Framework for Efficient Analyses

Challenging data analyses rely on an efficient access to the data and a fast turn around of new analysis ideas and their application to the collision data and the samples of simulated events. To facilitate the creation and administration of hundreds of histograms, the training and application of multivariate techniques, such as neural networks, and the determination of systematic uncertainties using ensemble tests, Dominic Hirschbühl and Georg Sartisohn have written a software framework for the Wuppertal data analyses: Arachne. Their work is based on previous experience with the CDF and CMS experiments.

Archachne:
Framework für
Multivariate Analysen

Dominic Hirschbühl

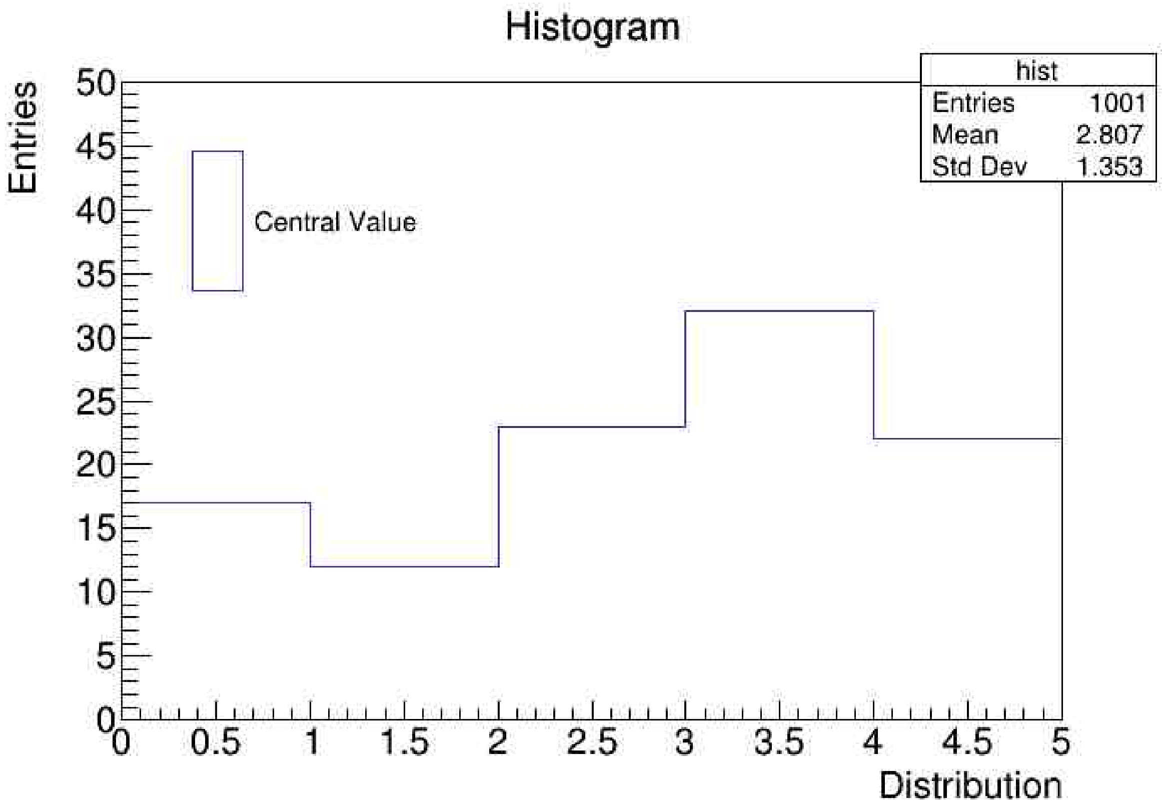
Wuppertal Physics Meeting
14.05.2009

Here we do the unfolding!

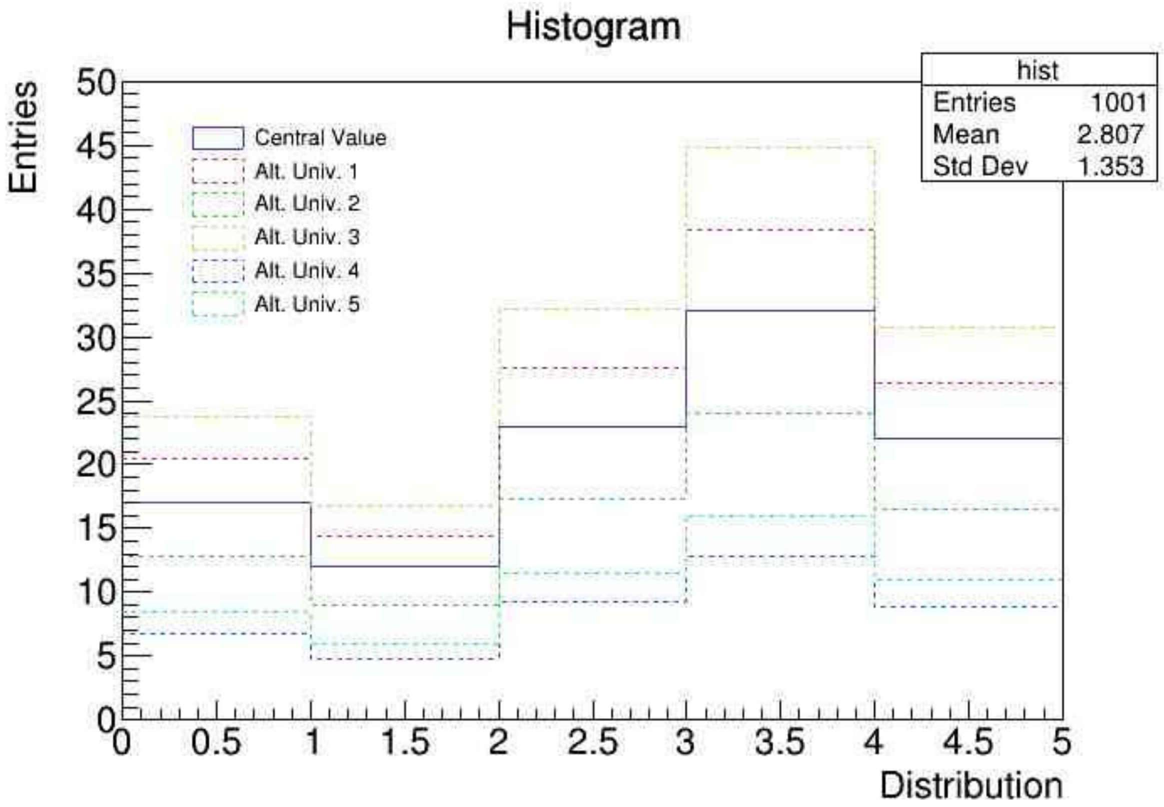
What do we study?

- There are two main methods of propagating systematic uncertainties

…and nobody has a real opinion on what is the right way to do it… so we are trying to find this out!



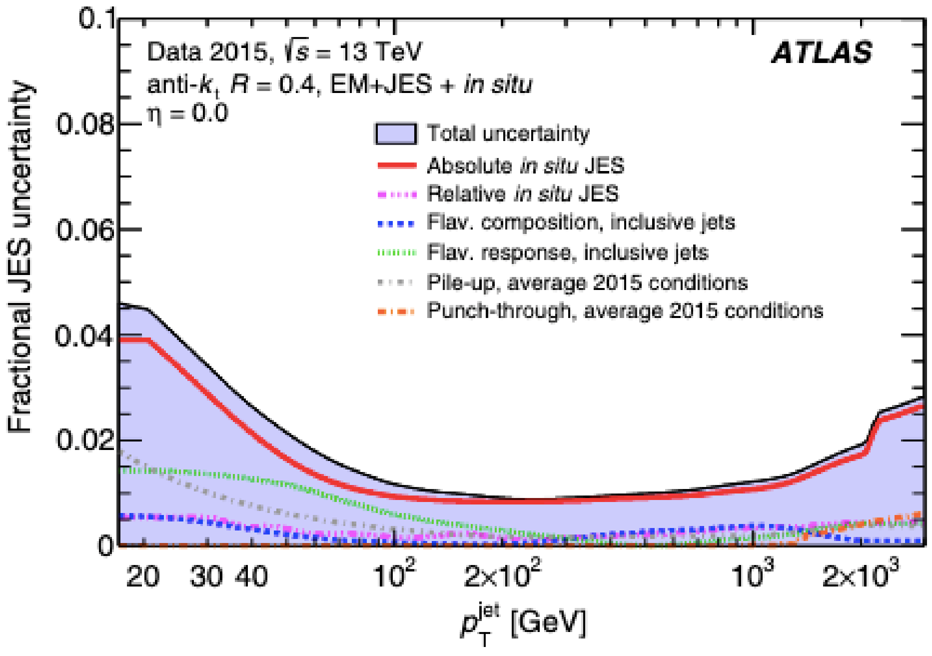
Nominal data



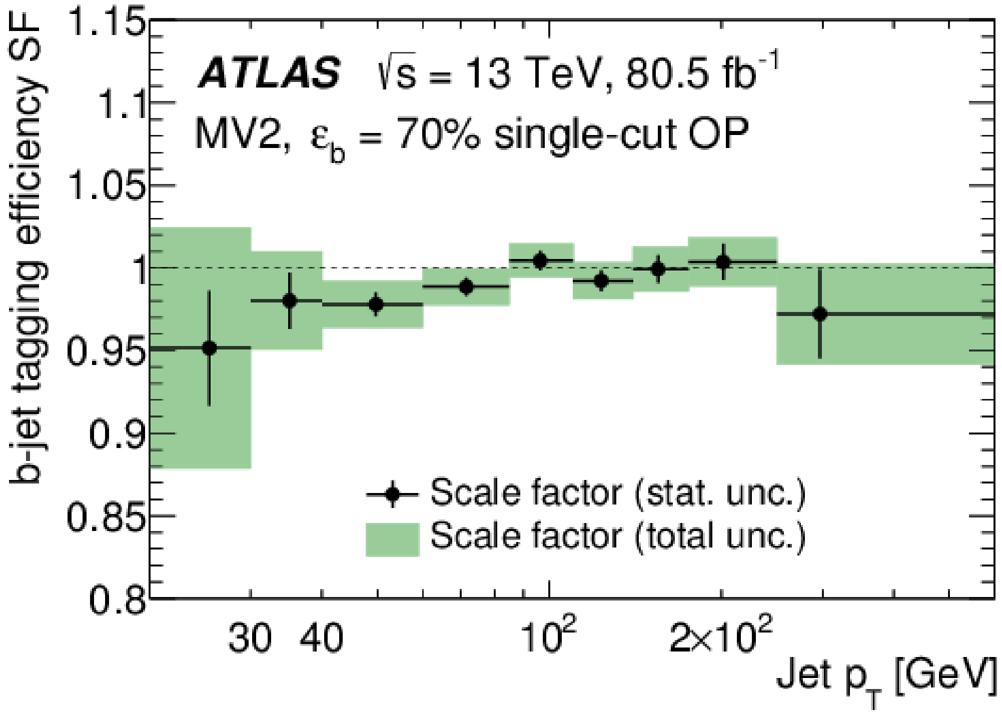
Data fluctuated by some systematic

What do we study?

- There are two main methods of propagating systematic uncertainties



(a)



What do we study?

- There are two main methods of propagating systematic uncertainties:

➢ Changing the input reco-file

➢ Changing the migration-matrix

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\mathcal{L}\cdot\Delta X^i\cdot\epsilon^i_{\mathrm{eff}}}\cdot\sum_j R^{-1}_{ij}\cdot f^j_{\mathrm{acc}}\cdot(N^j_{\mathrm{obs}} - N^j_{\mathrm{bkg}})$$

…or we could change the migration matrix (build it from this systematic data instead of the nominal one) and then put the nominal data through this (change migration-matrix)

If we want to unfold a systematic, the by the systematic affected distribution differs from the nominal one (see last slides), so we could change the number of events we observe (input reco-file)

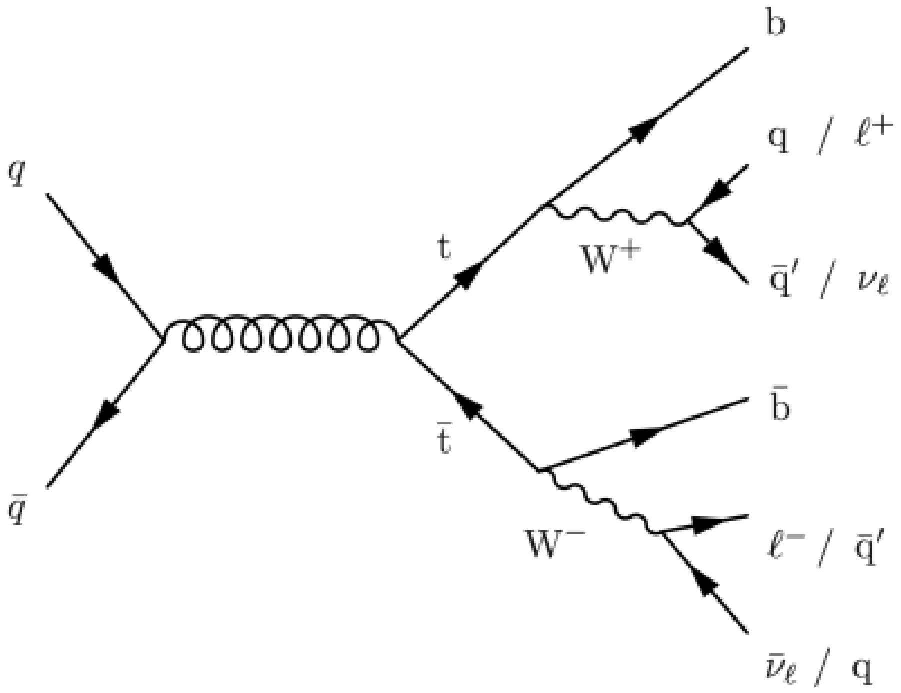
What do we study?

- There are two main methods of propagating systematic uncertainties:
- ➢ Changing the input reco-file
- ➢ Changing the migration-matrix

- Quick summary:
- ➢ The migration-matrix is normally built from the nominal data you get
- ➢ Then you can basically route you systematics data through the unfolding and see what the impact is INSERT PLOTS FOR SYSTEMATICS HERE
- ➢ Or you could change the matrix for every systematic and then put the nominal data through it

What do we study?

- There is a long discussion about which method is the best and if there are diffrences between these two methods
- Goal is to study the impact of both methods:
- ➢ First propagate the systematic uncertainty by changing the input reco-file
- ➢ Propagate the systematic uncertainty by changing the migration-matrix
- ➢ Compare these two results obtained for some systematic errors



- $t\bar{t}$ dilepton channel:
  - $e^{\pm}$ and $\mu^{\mp}$, $p_T > 25$ GeV
  - 2 or 3 jets, $p_T > 25$ GeV
  - 2 b-tags @ 70% efficiency working point

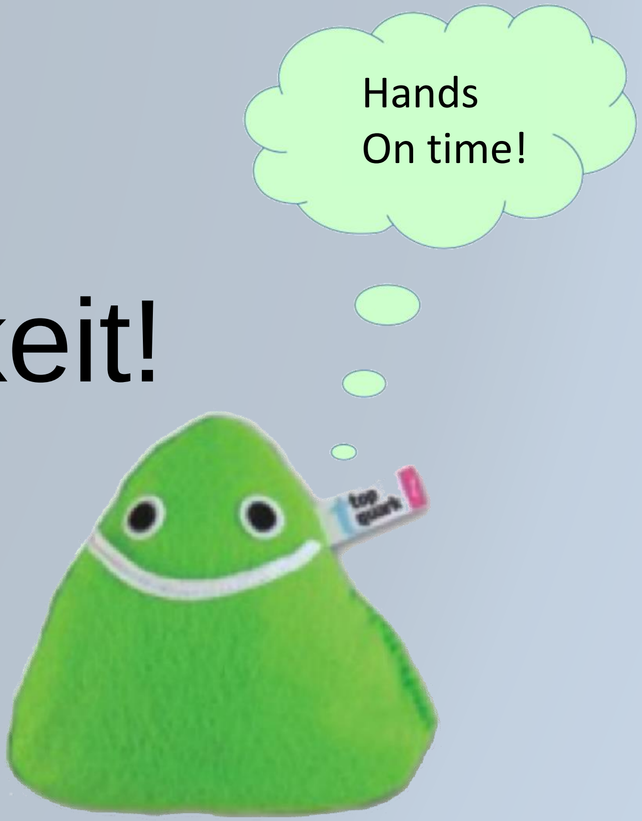
What do we study?

- There is a long discussion about which method is the best and if there are diffrences between these two methods
- Goal is to study the impact of both methods:
- ➢ First propagate the systematic uncertainty by changing the input reco-file
- ➢ Propagate the systematic uncertainty by changing the migration-matrix
- ➢ Compare these two results obtained for some systematic errors

- Based on: „Measurement of observables sensitive to colour reconnection”
- For now, we just study the 1D-variables
- Particle-level observables: n_{ch} and $\sum_{n_{ch}} p_t$
- Detector-level observables: $n_{trk,out}$ and $\sum_{n_{trk,out}} p_t$

BERGISCHE UNIVERSITÄT WUPPERTAL

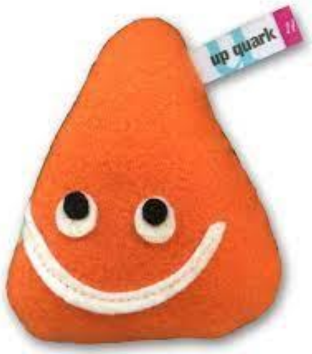
Vielen Dank für Ihre Aufmerksamkeit!



ATLAS EXPERIMENT

NETZWERK TEILCHENWELT

Back-



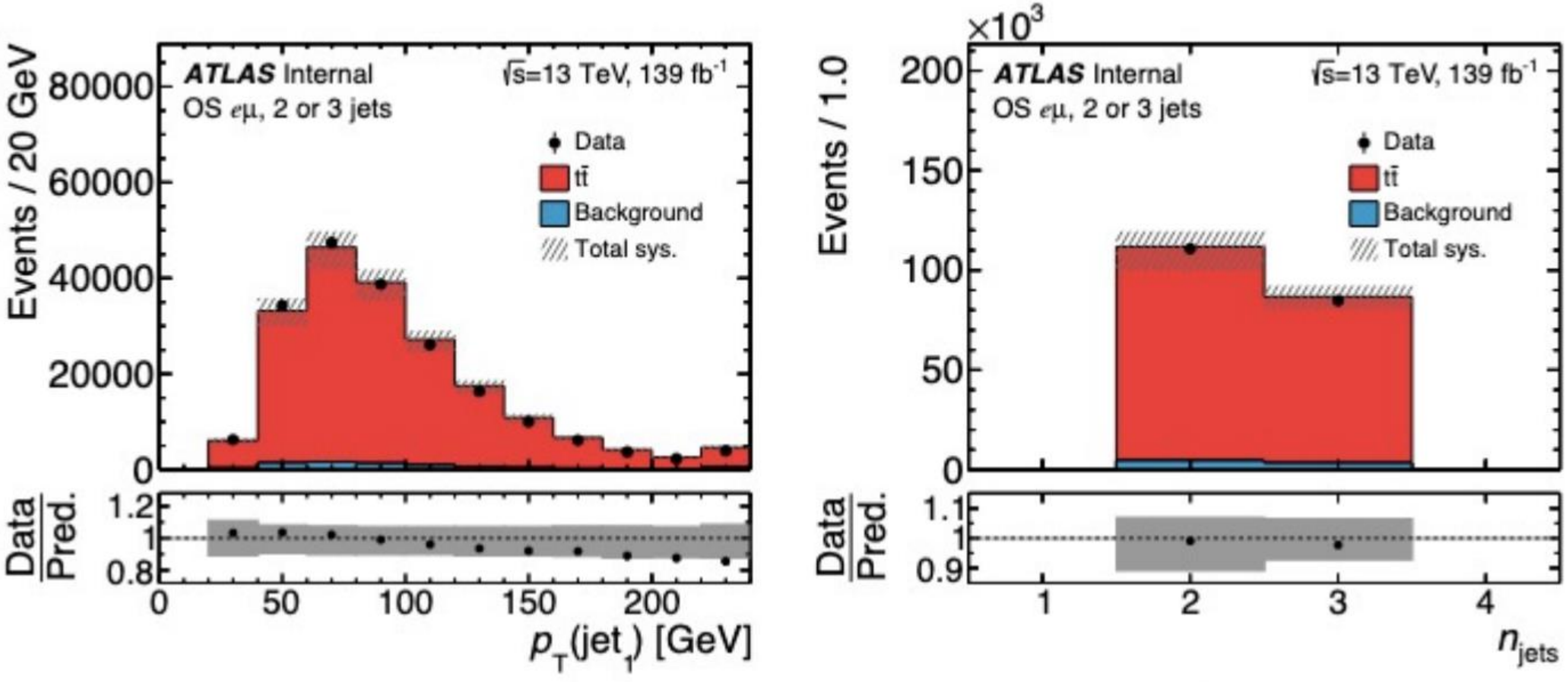
Measurement of observables sensitive to colour reconnection

PCM: https://indico.cern.ch/event/1188349/

Paper Draft: https://cds.cern.ch/record/2778339/files/ATL-COM-PHYS-2021-593.pdf

Glance: https://atlas-glance.cern.ch/atlas/analysis/papers/details?id=13766

Measurement of observables sensitive to colour reconnection



(a) (b)

Figure 2: Distributions of (a) the leading jet p_T and (b) the jet multiplicity in events with an opposite-sign $e\mu$ pair and two or three jets, of which two are b-tagged jets. The observed data is compared with the expectation from simulated $t\bar{t}$ and background events, where the background includes contributions from tW, $t\bar{t}+V$, Z+jets and diboson processes and the estimated fake-lepton background. The lower panel shows the ratio of data to the prediction in each bin. The hatched (grey) uncertainty band includes the MC statistical uncertainty, background normalisation uncertainties, detector systematic uncertainties and $t\bar{t}$ modelling systematic uncertainties. Events beyond the x-axis range are included in the last bin.

Measurement of observables sensitive to colour reconnection

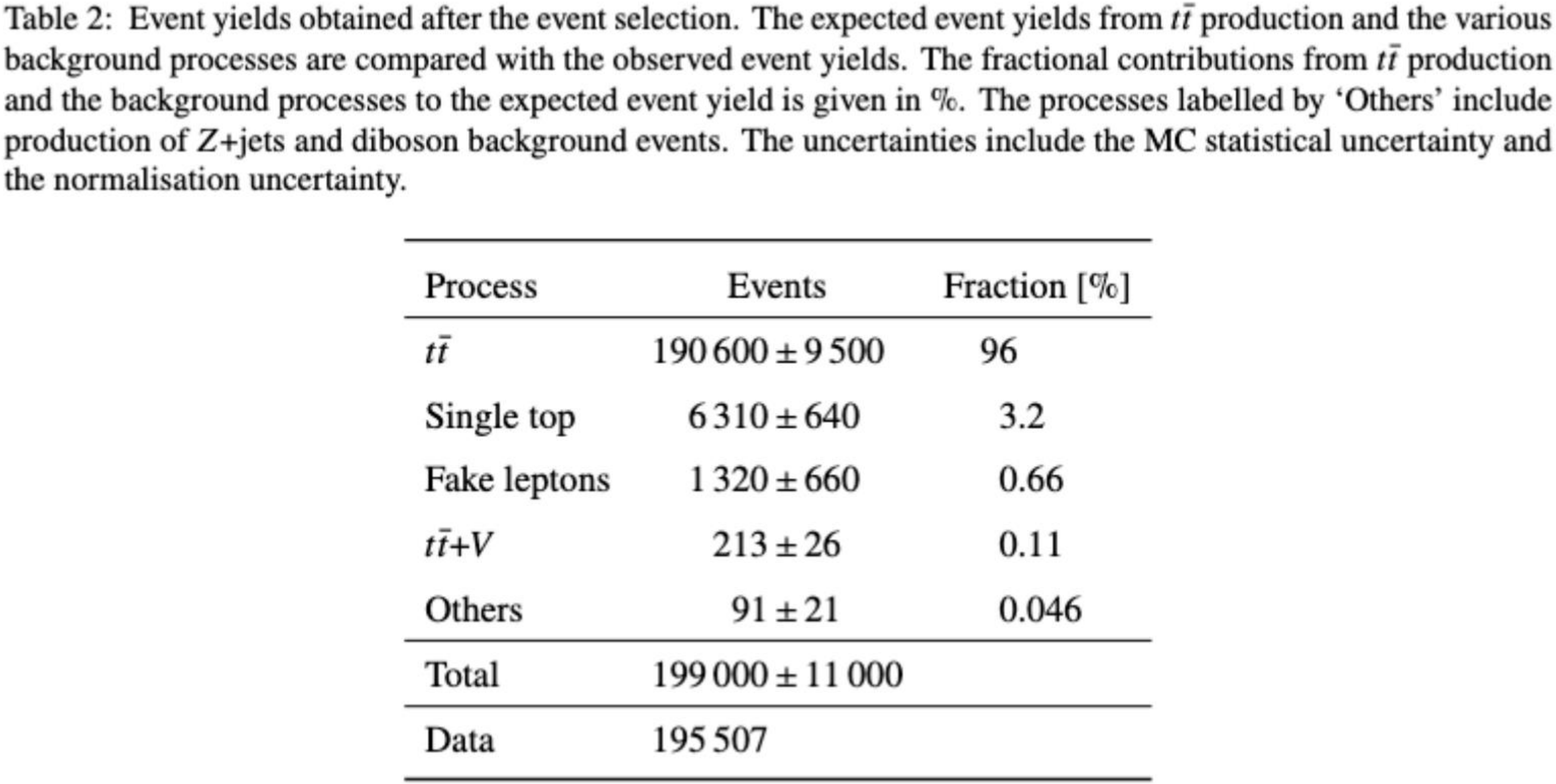
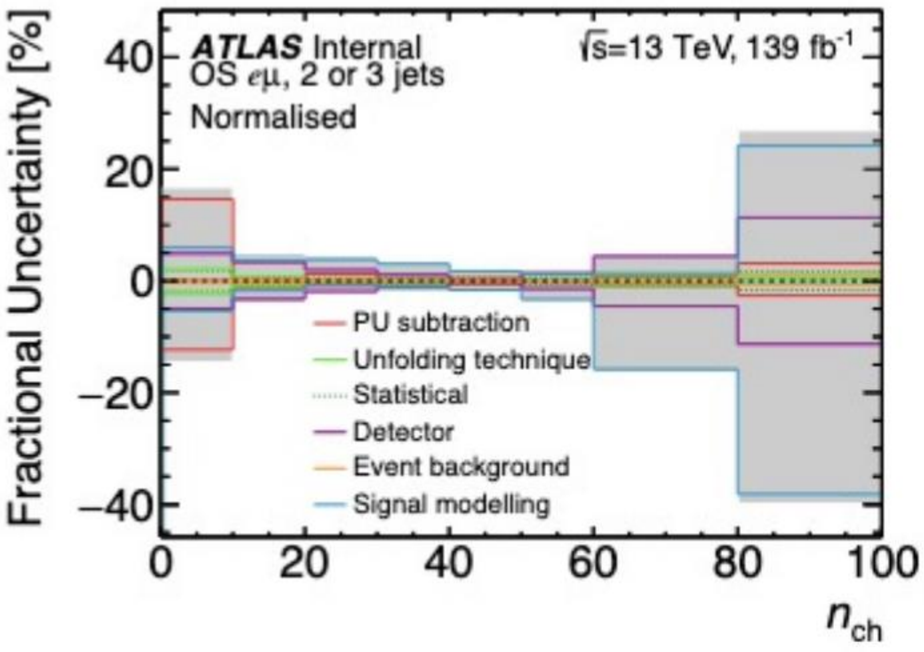


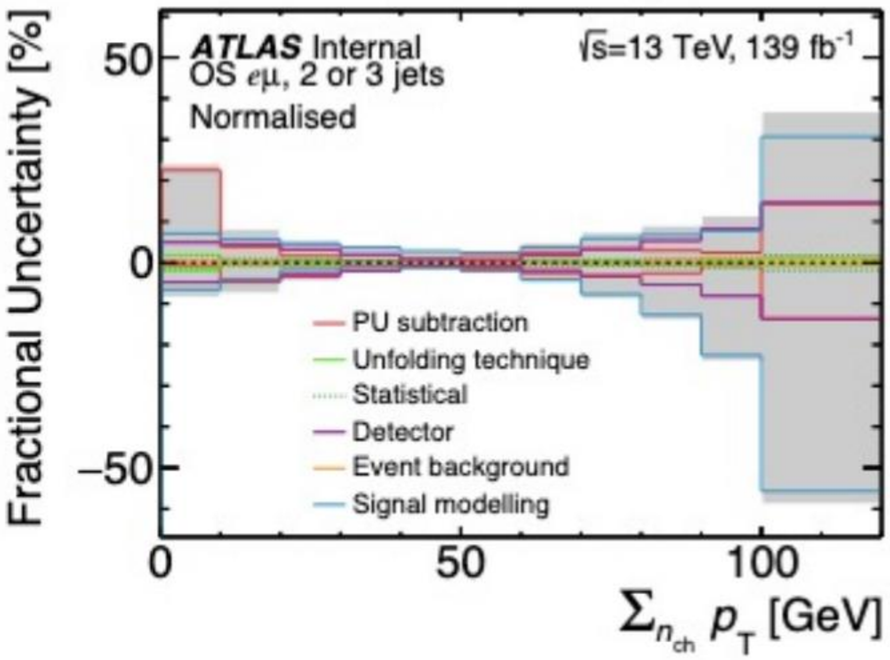
Table 2: Event yields obtained after the event selection. The expected event yields from $t\bar{t}$ production and the various background processes are compared with the observed event yields. The fractional contributions from $t\bar{t}$ production and the background processes to the expected event yield is given in %. The processes labelled by ‘Others’ include production of Z+jets and diboson background events. The uncertainties include the MC statistical uncertainty and the normalisation uncertainty.

Process	Events	Fraction [%]
$t\bar{t}$	$190\,600 \pm 9\,500$	96
Single top	$6\,310 \pm 640$	3.2
Fake leptons	$1\,320 \pm 660$	0.66
$t\bar{t}+V$	213 ± 26	0.11
Others	91 ± 21	0.046
Total	$199\,000 \pm 11\,000$	
Data	$195\,507$	

Measurement of observables sensitive to colour reconnection

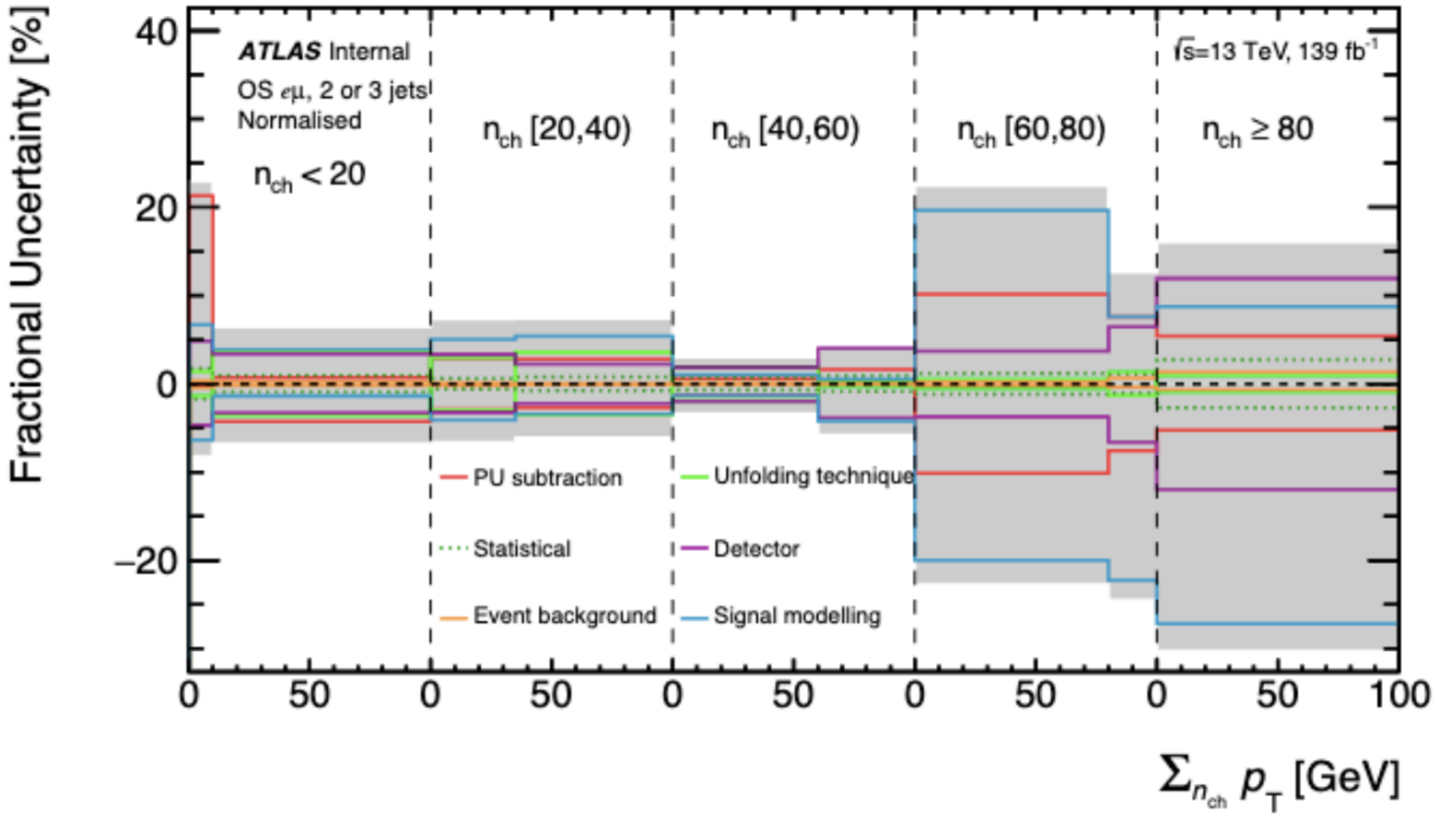


(a)



(b)

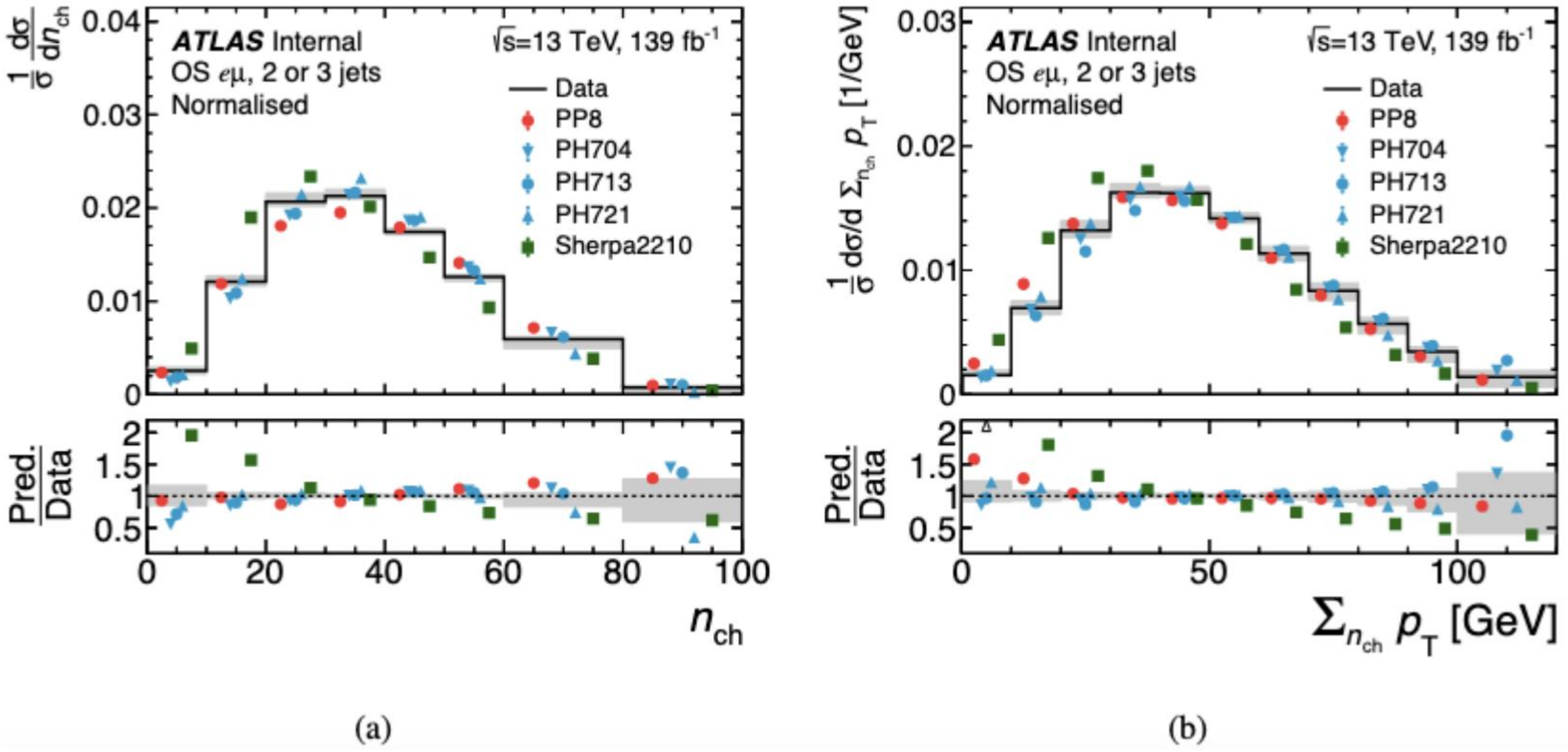
Measurement of observables sensitive to colour reconnection



(c)

Figure 8: Fractional uncertainties in the measured normalised differential cross-section as a function of (a) n_{ch}, (b) $\sum_{n_{\mathrm{ch}}} p_{\mathrm{T}}$, (c) $\sum_{n_{\mathrm{ch}}} p_{\mathrm{T}}$ in bins of n_{ch} for all systematic and statistical uncertainties. The x-axis in (c) is split into five bins of n_{ch} by the dashed vertical lines and $\sum_{n_{\mathrm{ch}}} p_{\mathrm{T}}$ is presented in each bin. The grey bands represent the sum in quadrature of the presented components. Events beyond the x-axis range are included in the last bin in (a) and (b), and in the corresponding last bin of each slice in (c).

Measurement of observables sensitive to colour reconnection



(a)

(b)