Measurement of observables sensitive to colour reconnection in $\boldsymbol{t\bar{t}}$ events with the ATLAS detector at $\sqrt{\boldsymbol{s}} = \mathbf{13\ TeV}$

Shayma Wahdan

BERGISCHE
UNIVERSITÄT
WUPPERTAL

Seminar at TU Dortmund

May 25, 2023

Content

Theoretical context

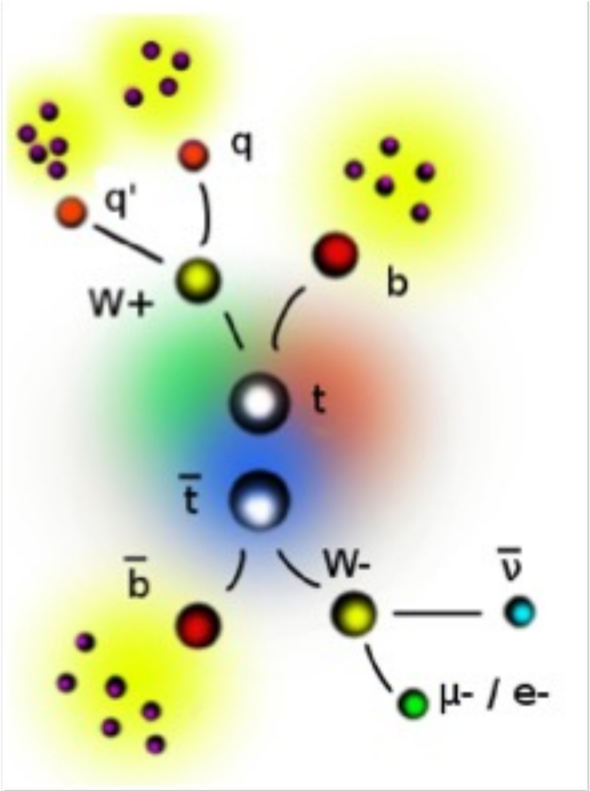
- The top-quark
- Colour reconnection modelling
- Sensitive observables

Analysis

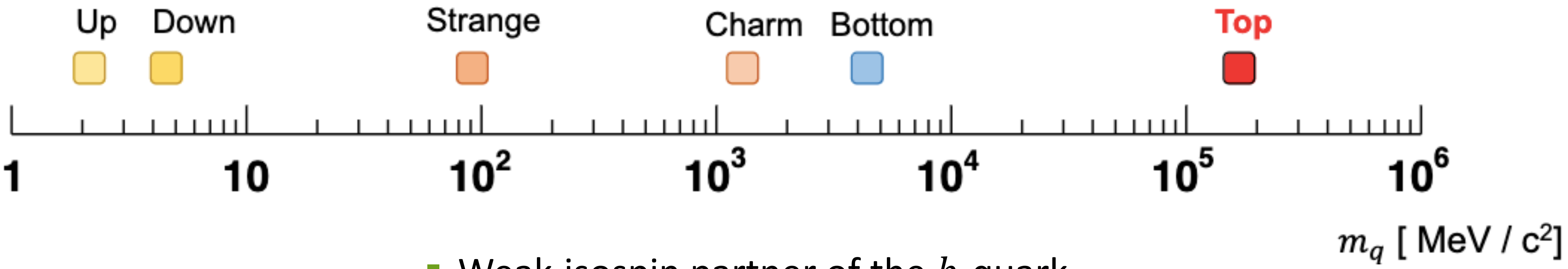
- Event selection and Background estimation
- Unfolding the sensitive observables
- Estimate systematic uncertainties

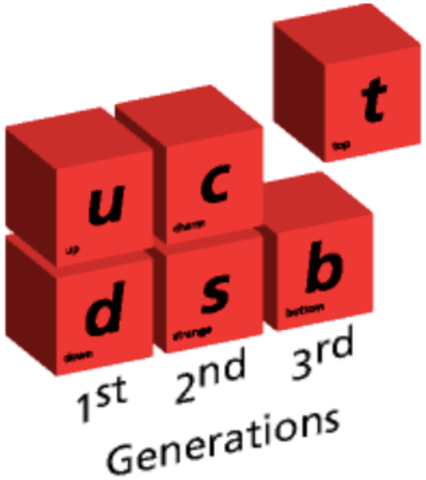
Results

- Compare unfolded data to MC predictions



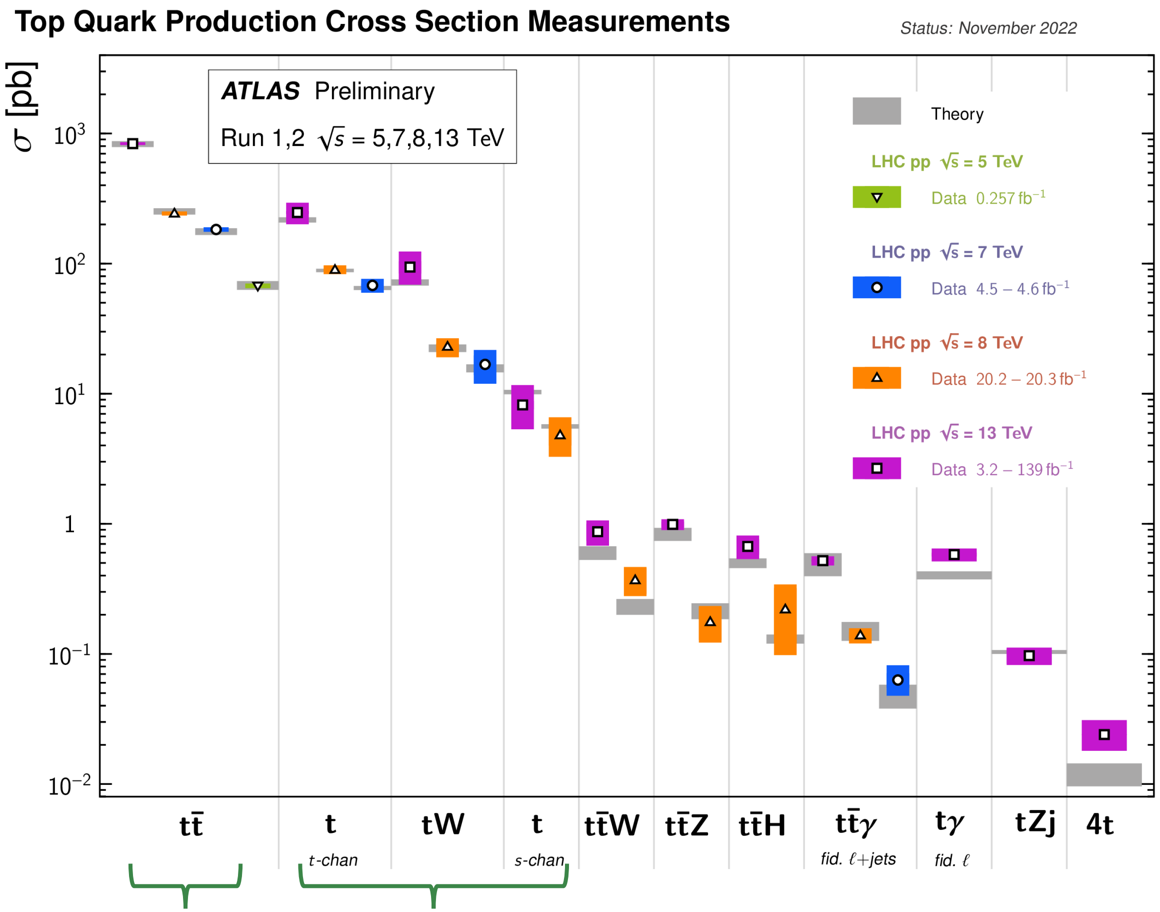
The top quark





- Weak-isospin partner of the b-quark
- Electric charge: $+2/3\,e$
- Spin: $1/2$
- The heaviest elementary particle $m_{\mathrm{t}} = 172.69 \pm 0.48\ \mathrm{GeV}$
- Very short lifetime $\tau \approx 0.5 \times 10^{-24}\ \mathrm{s}$
  $\Rightarrow$ decay before forming hadron
- Large Yukawa coupling to the Higgs boson $y_t \sim 1$
  $\Rightarrow$ connection to EW symmetry breaking

Top-quark processes



pairs

Single top-quark

- Dominant production: in pairs via the strong interaction

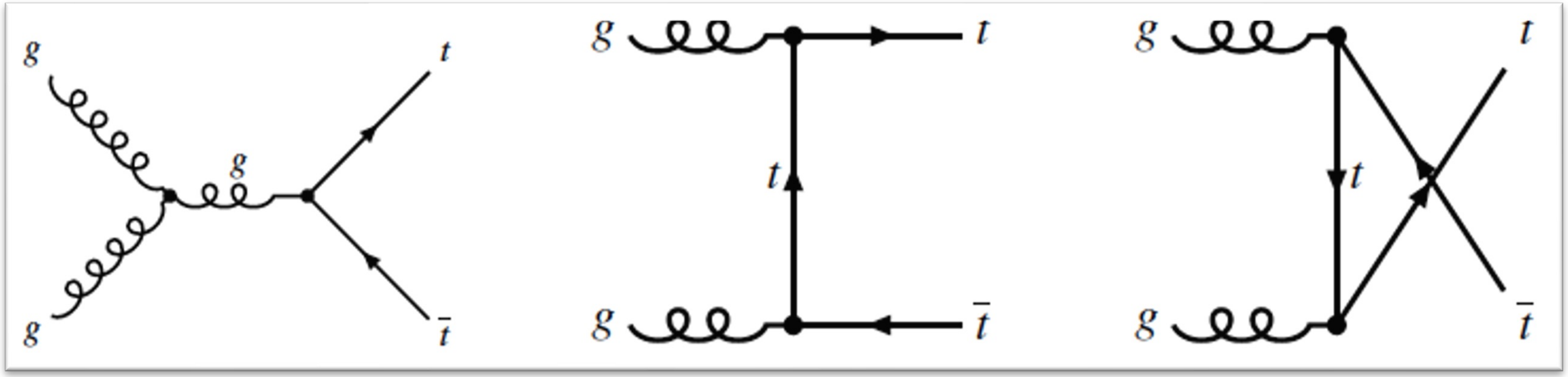
- Very high production rate at the LHC
  ⇒ Produced more than **100M** pairs during Run 2
  $(N = L\,\sigma = 139\mathrm{fb}^{-1} \cdot 832\mathrm{pb} \approx 116\ \mathrm{M})$

Top-quark pair production

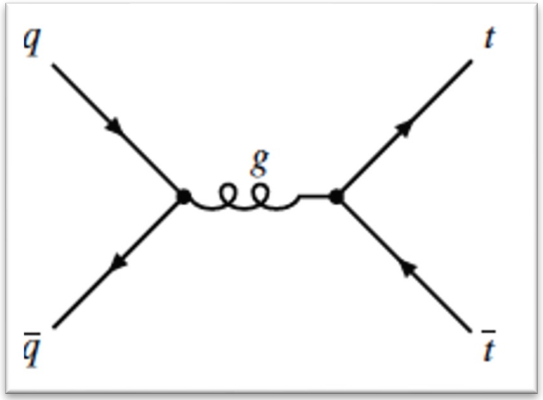
- At LO, produced either by gluon-gluon fusion or quark-antiquark annihilation

- **Mainly via gluon-gluon fusion at the LHC (90%)**

gluon-gluon fusion (90%)

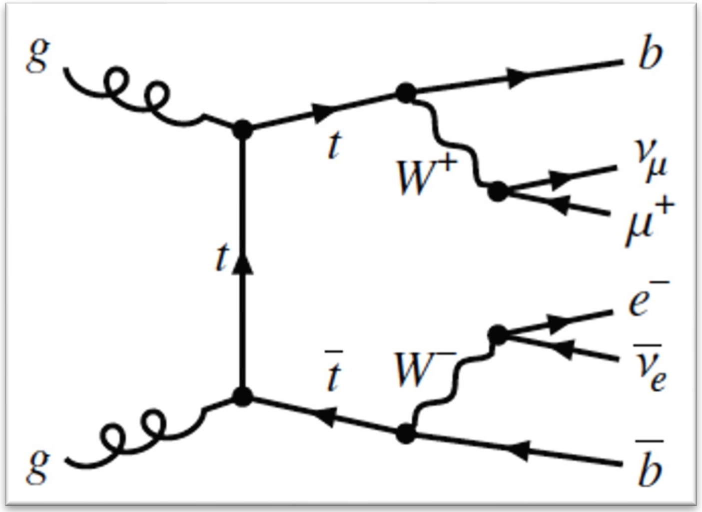


$q\bar{q}$ annihilation (10%)



Top-quark pair decay

- top-quark decay via the weak interaction
- Since $V_{tb} \approx 1, \mathcal{B}\,(t \rightarrow Wb) \approx 100\%$



- $t\bar{t}$ has three decay channels, defined by the decay mode of the W-boson from the top-quark

Channel	Branching ratio
Dilepton	10.5 %
lepton+jets	43.8%
All-hadronic	45.7%

- **Measurement in the dilepton channel**
  - highest signal-to-background ratio
  - small branching ratio is not a limitation, as we have a large dataset

one important uncertainty is colour-reconnection

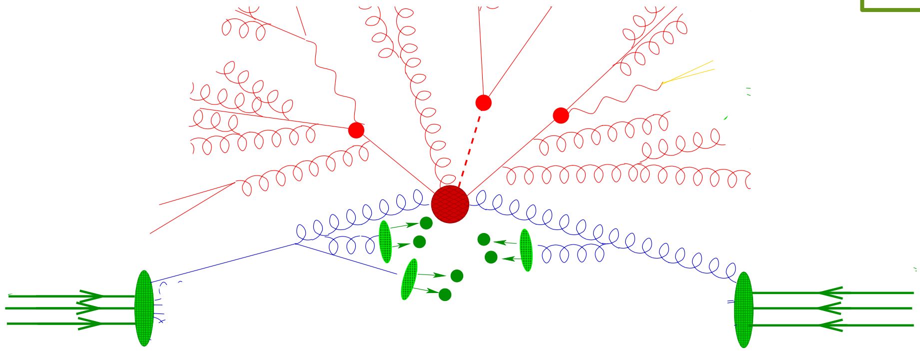
What is colour reconnection?

Proton-proton collision at high energies

Incoming proton

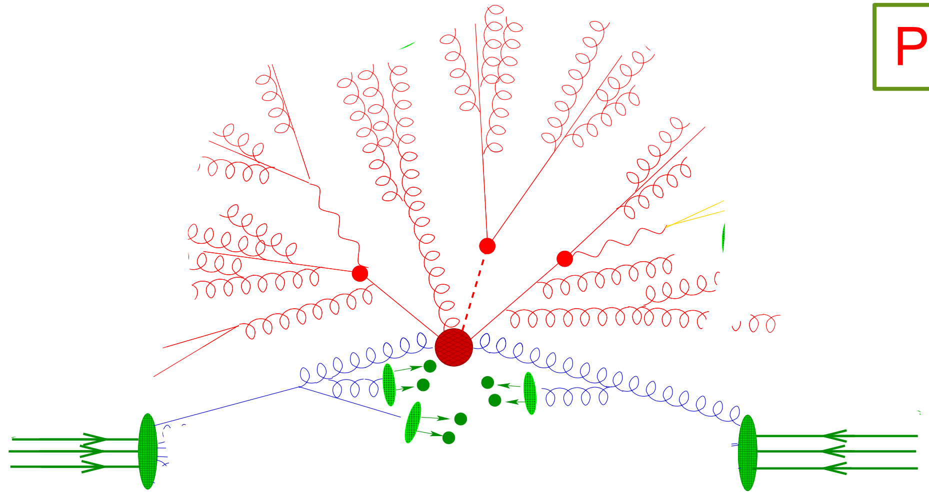
Incoming proton

Proton-proton collision at high energies



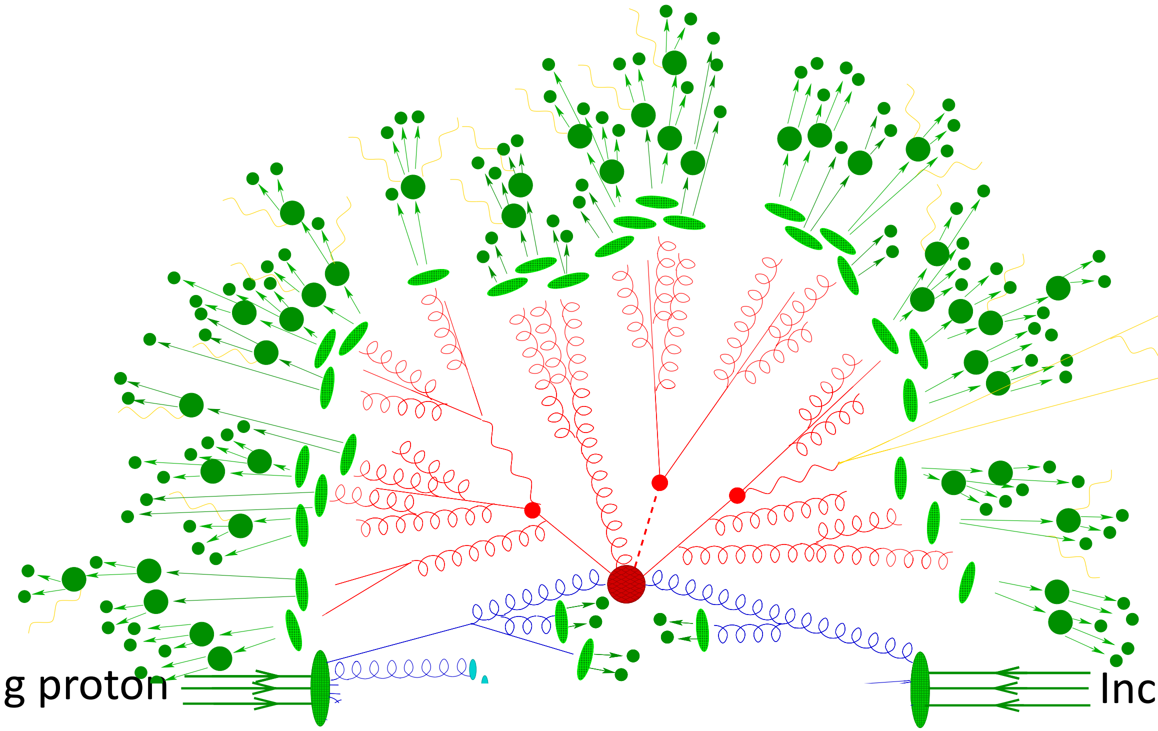
Hard partonic interaction: two partons with the highest momentum fraction interact.

Proton-proton collision at high energies



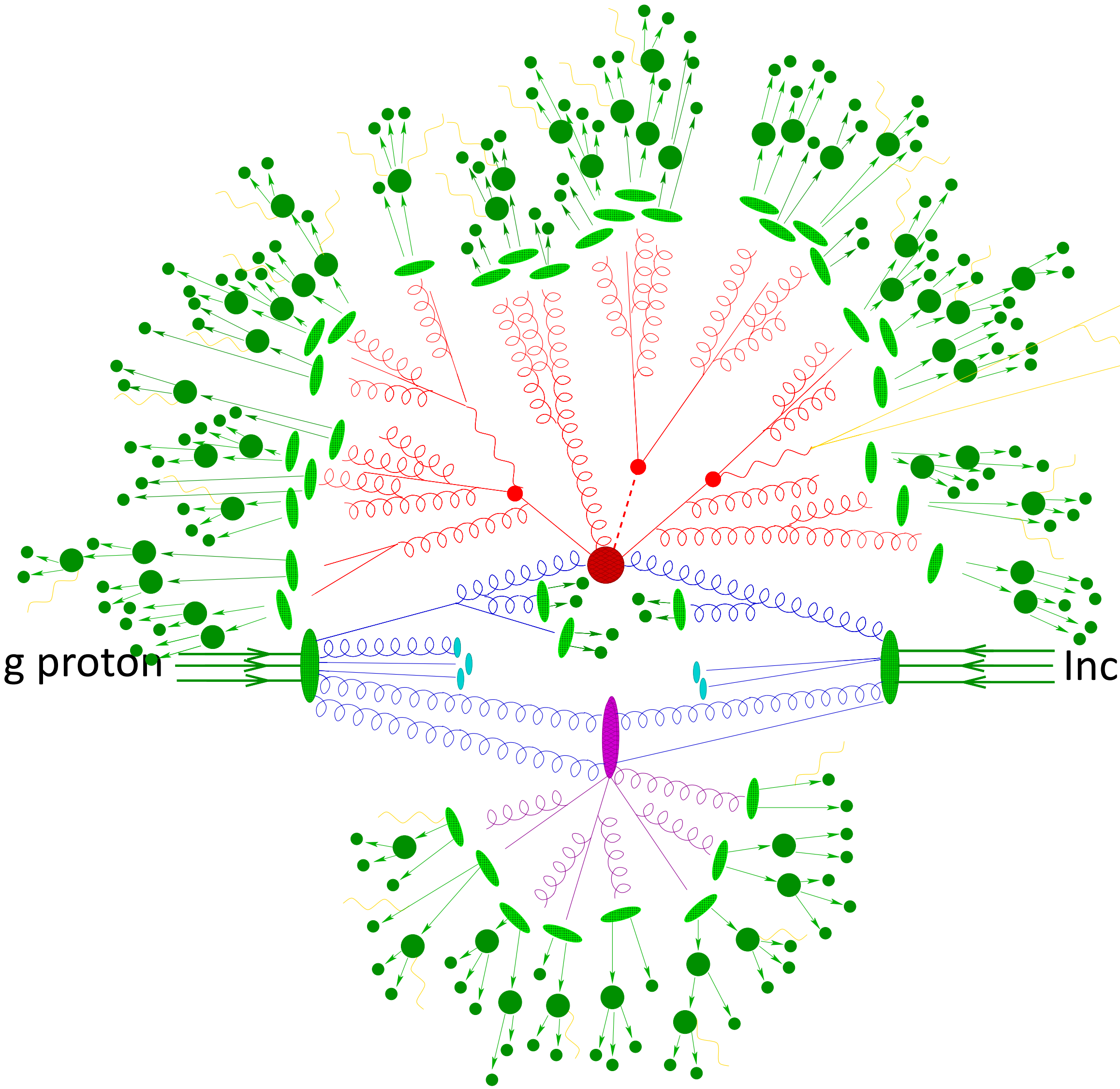
Parton shower: quark and gluons radiation

Proton-proton collision at high energies



Hadronisation: coloured partons combine to form colour-neutral hadrons

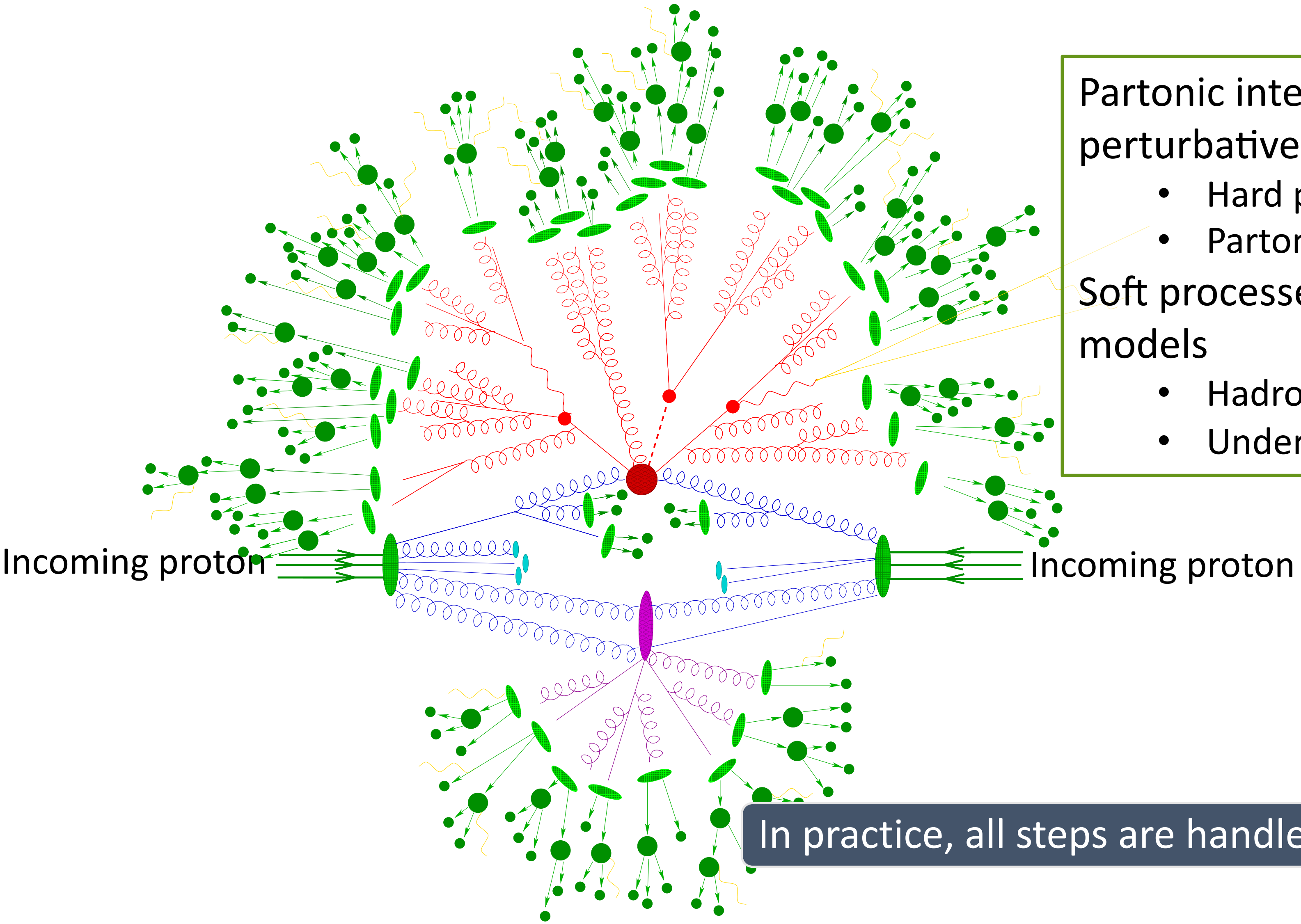
Proton-proton collision at high energies



Underlying events: Activity accompanying the hard scatter

- Multi-partonic interactions
- Beam remnants
- Radiation processes (ISR, FSR)

Proton-proton collision at high energies



Partonic interactions are calculable in perturbative QCD:

  - Hard process
  - Parton shower

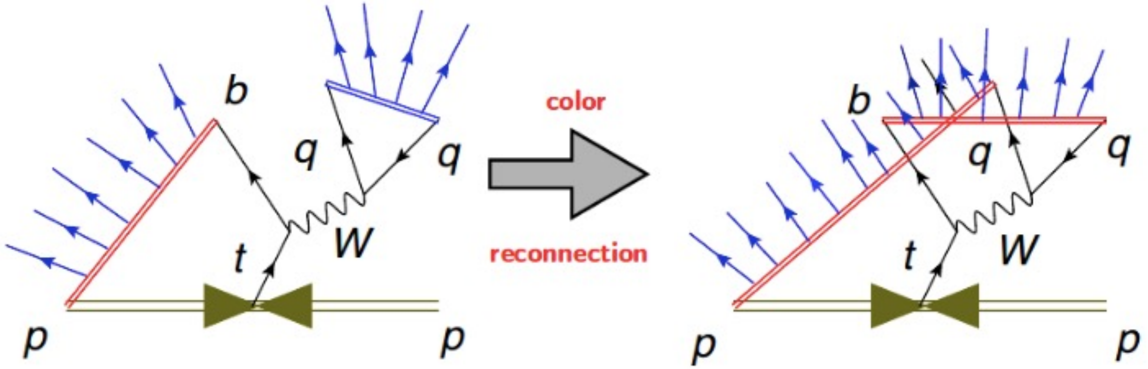
Soft processes rely on phenomenological models

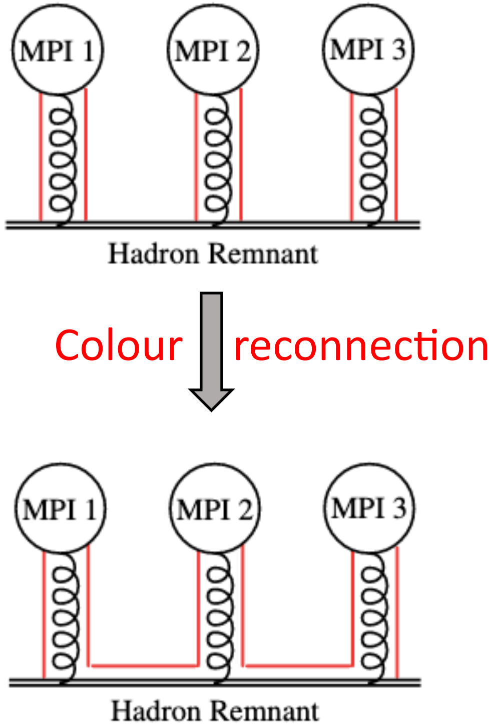
  - Hadronisation and decay
  - Underlying events

In practice, all steps are handled by Monte Carlo simulations

Colour reconnection (CR)

Mechanism of reassigning colour connections between partons during hadronisation



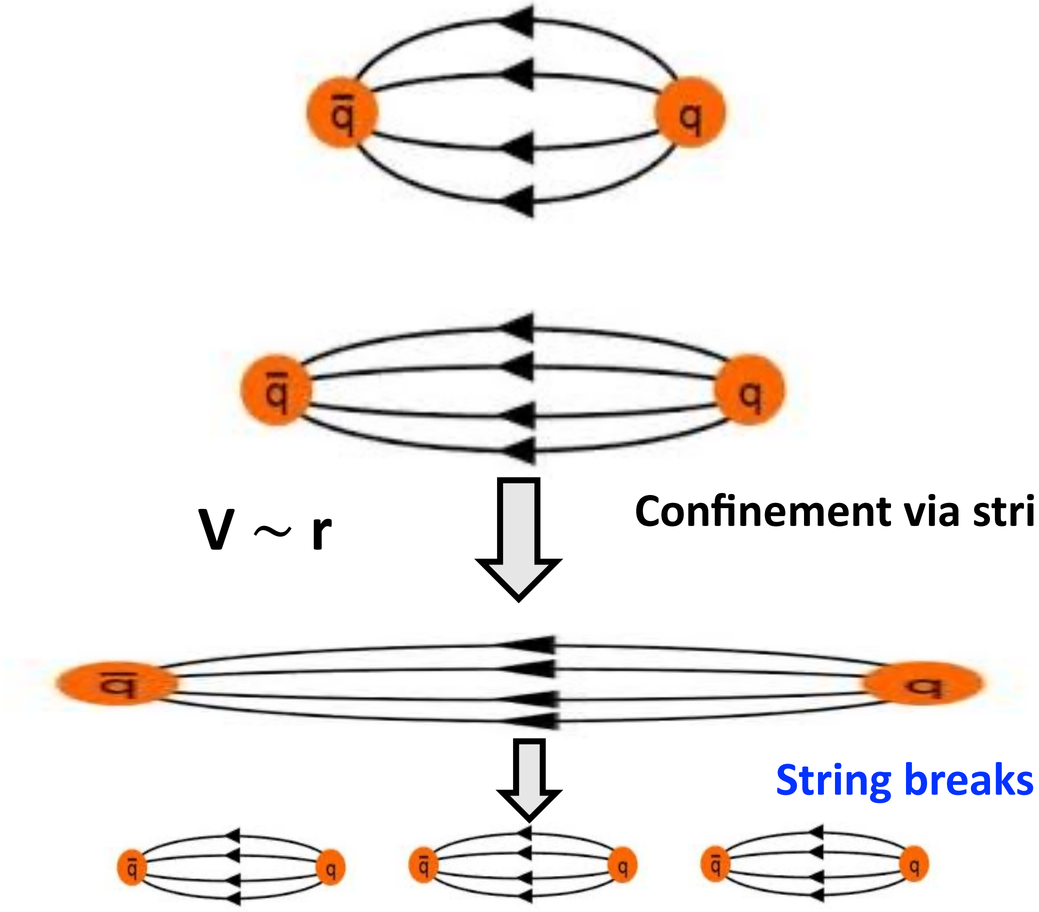


Colour reconnection

- In the leading colour (LC) approximation (before CR):
  - Each MPI is viewed as separate from all other systems in colour space
  - No strings stretched between different MPI systems
- CR allows different MPI systems to be colour-connected to each other (MPI hadronise collectively)
  - **Total colour charge reduced** w.r.t. LC approximation

Overview of CR models in PYTHIA 8

Based on reconstructing a colour potential that minimise the total string length



V ~ r

Confinement via string

String breaks forming hadrons

Longer string → more hadrons

String length measure:

$$\lambda \approx \sum_{ij} \ln\left(1 + \frac{m_{ij}^2}{m_0^2}\right)$$

Parton pairs

typical hadronic mass scale

Overview of CR models in PYTHIA 8

MPI-based model (CR0)

1. Starting from lowest p_T interaction calculate reconnection probability

$$P_{\mathrm{rec}}\,(p_T) = \frac{(R_{\mathrm{rec}}\,p_{\mathrm{T0}})^2}{(\underbrace{R_{\mathrm{rec}}}_{\text{CR range}}\,\underbrace{p_{\mathrm{T0}}}_{\text{Soft dampening scale}})^2 + p_{\mathrm{T}}^2}$$

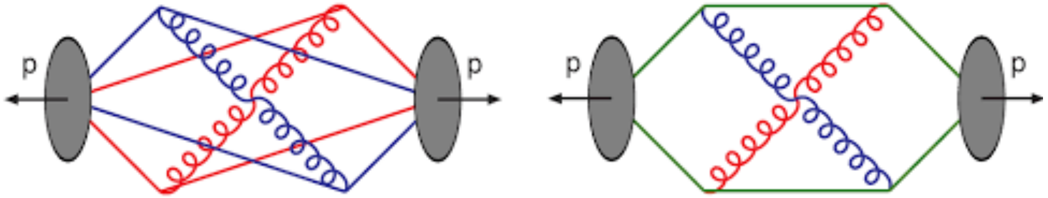
$p_{\mathrm{T}} \downarrow \Longrightarrow P_{\mathrm{rec}} \uparrow$

softer systems easier to reconnect

2. Iterate (1) for all interactions; if $\mathrm{P_{rec}} > \alpha \in [0,1]$ do reconnection

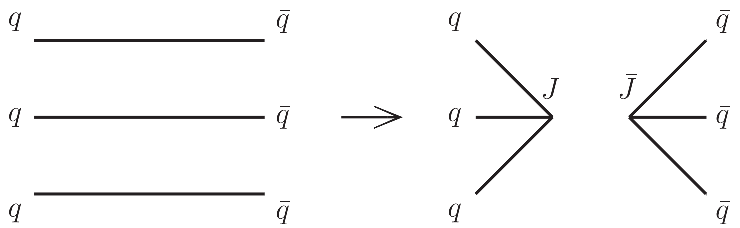
α → stochastic

3. Move gluons from softer interactions to high p_T dipole that minimizes the increase in λ

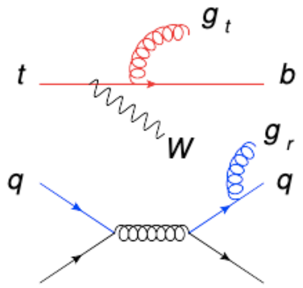


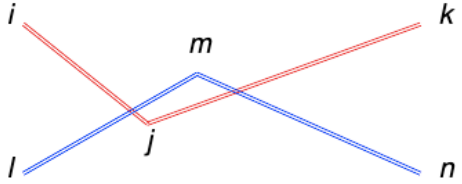
Overview of CR models in PYTHIA 8

- QCD-based (CR1):
  - is a more complete treatment of the QCD multiplet structure
  - includes reconnections of dipoles, which can produce structures of three (anti-)colour indices (junctions) → Improves description of baryon production
  - based on string minimization

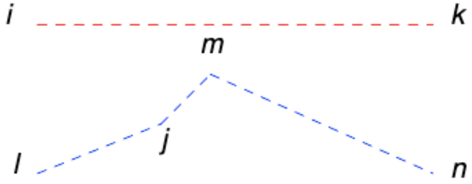


- Gluon-move (CR2):
  - only gluons are considered for reconnection, no quarks reconnection
  - each gluon reconnect to all MPI systems (not only the ones for softer MPIs)
  - As QCD-based model, also based on the minimization of string length



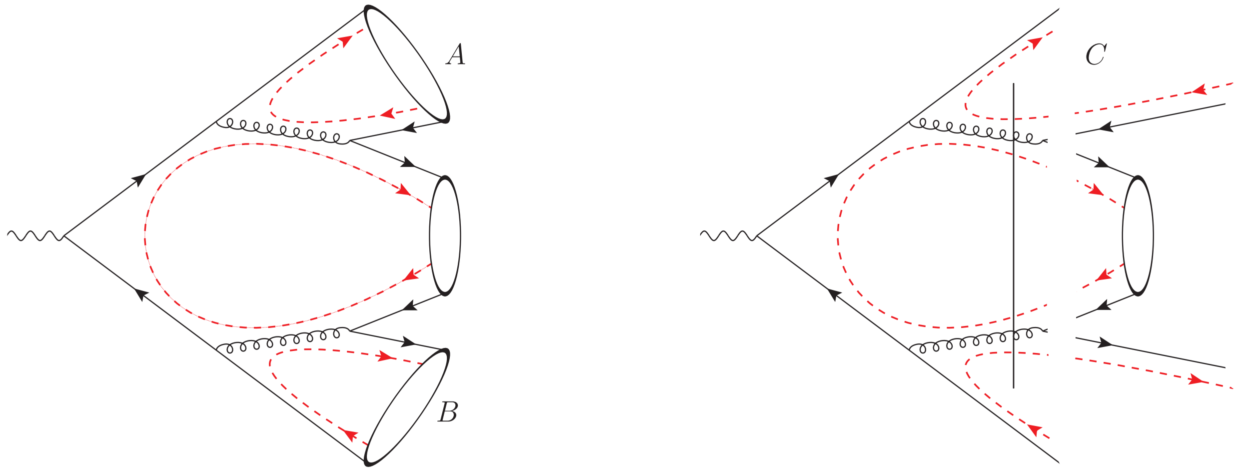


original configuration



resulting configuration after moving the gluon

The default CR model in Herwig 7

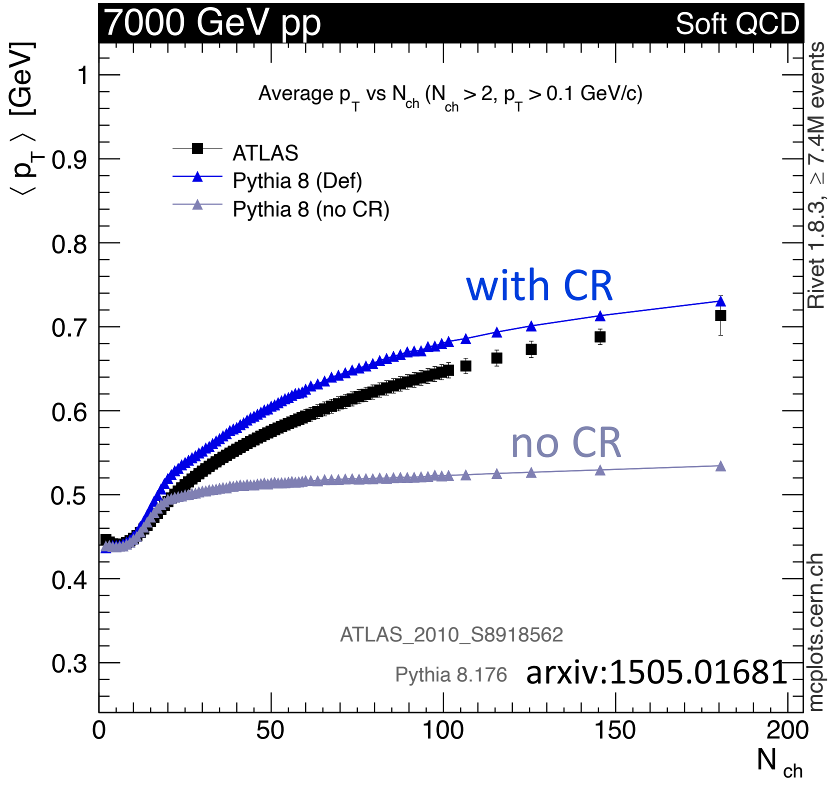


- Iterating over quarks in all clusters, try reconnection
- Select reconnection which minimises $m_\mathrm{C} + m_\mathrm{D}$ if $m_\mathrm{C} + m_\mathrm{D} < m_\mathrm{A} + m_\mathrm{B}$
- Accept reconnection with probability P_reco

Why do we need colour reconnection?

- Explains the rising trend of $\langle \mathrm{p_T} \rangle$ vs. n_{ch}
  ↪ CR is needed to describe the data

- It can shed light in the quest for precise SM measurements, such as the top-quark mass:
  - top-quark decays take place right in the middle of the showering/hadronization region
  - so quarks(and gluons) produced in the decay are subject to CR
  - $t \rightarrow bu\bar{d}$, b for sure is colour-connected somewhere else, giving mass ambiguities

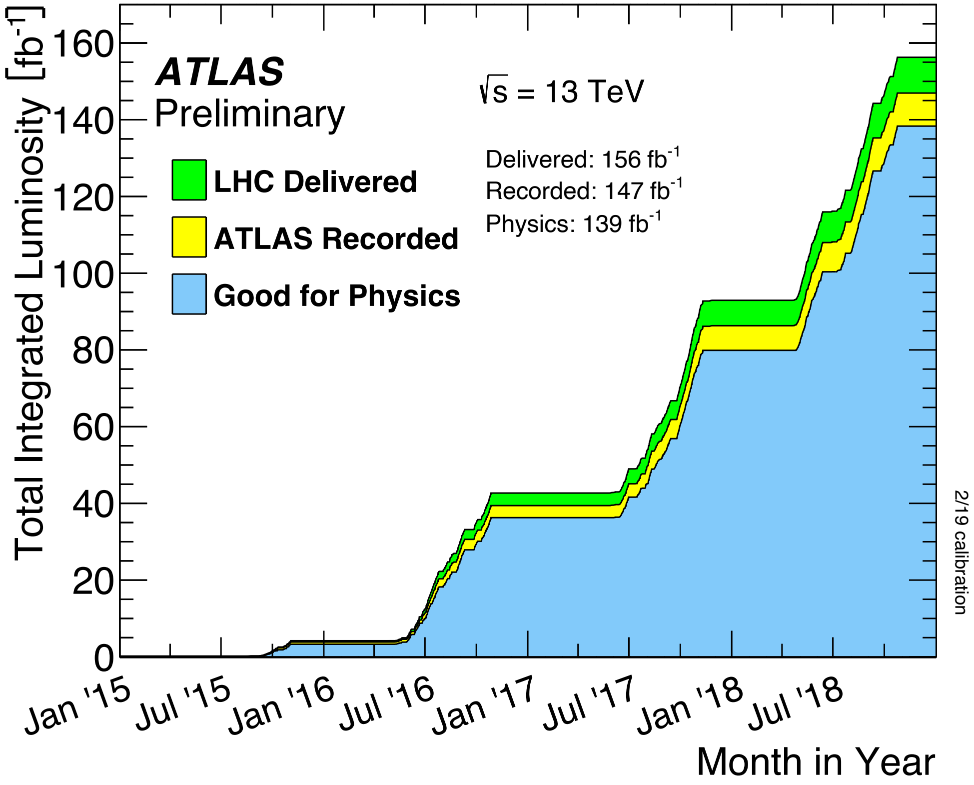


arxiv:1505.01681

➢ CR is one of the dominant systematics in m_{top} measurements (up to $400\ \mathrm{MeV}$)

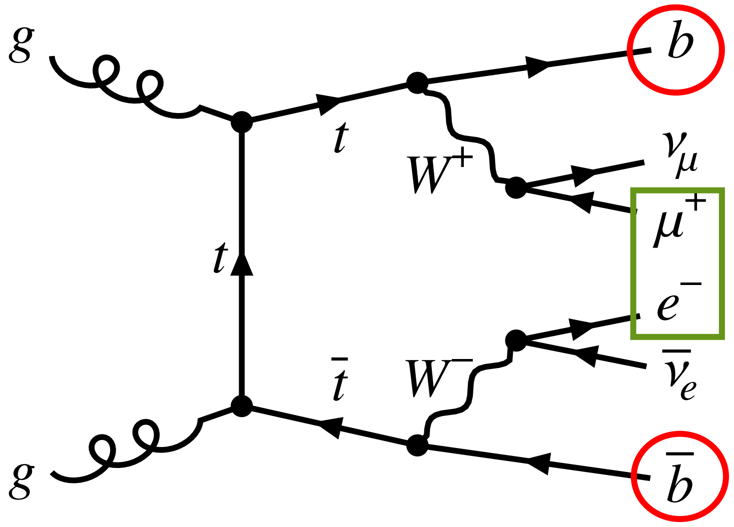
prescription to estimate this uncertainty is not well defined
⇒ In Pythia 8, there are currently more than 15 CR models

Analysis



ATLAS
Preliminary
$\sqrt{s}$ = 13 TeV
🟩 LHC Delivered
🟨 ATLAS Recorded
🟦 Good for Physics
Delivered: 156 fb^{-1}
Recorded: 147 fb^{-1}
Physics: 139 fb^{-1}
Total Integrated Luminosity [fb^{-1}]
0
20
40
60
80
100
120
140
160
Jan '15
Jul '15
Jan '16
Jul '16
Jan '17
Jul '17
Jan '18
Jul '18
Month in Year
2/19 calibration

Event selection

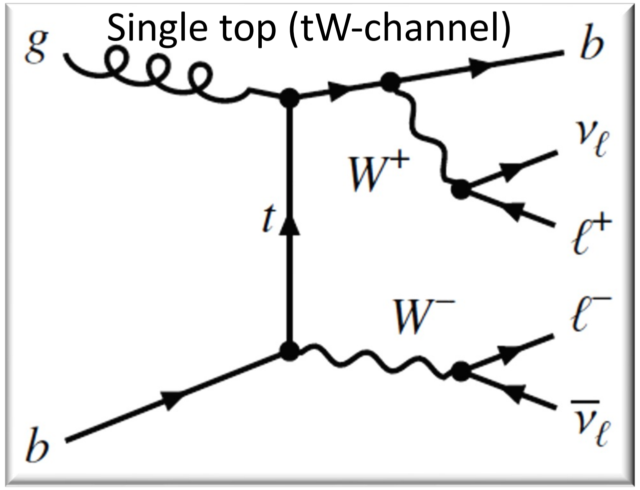


- Full Run 2 data, $\int L\ dt = 139\ fb^{-1}$

- Select $t\bar{t}$ 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
  - $m_{\ell\ell} > 15$ GeV

Backgrounds to $t\bar{t}$ events

- Single top, $t\bar{t}V$, Z +jets and diboson
  - Expected number of events calculated using their theoretical cross-sections
- Fake lepton background
  - At least one of the leptons is wrongly reconstructed as prompt
    - An electron from photon conversion ($e \rightarrow \gamma \rightarrow e$)
    - e or μ from the decay of a bottom or charm hadron
  - estimated with partial data driven approach using $e^{\pm}\mu^{\pm}$ same (charge) sign control region

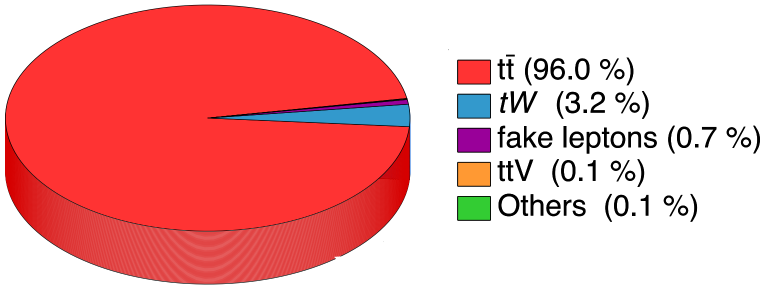


$$N^{\text{fake}} = R \cdot \left(N^{\text{data},SS} - N^{\text{prompt},SS}\right)$$

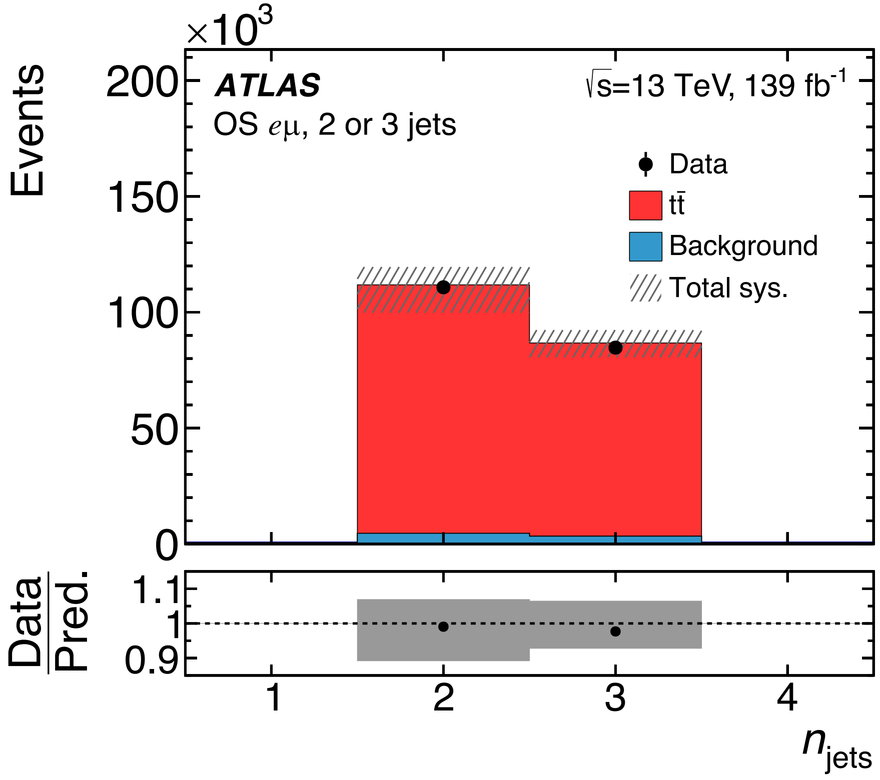
Ratio of OS-to-SS events with fake leptons

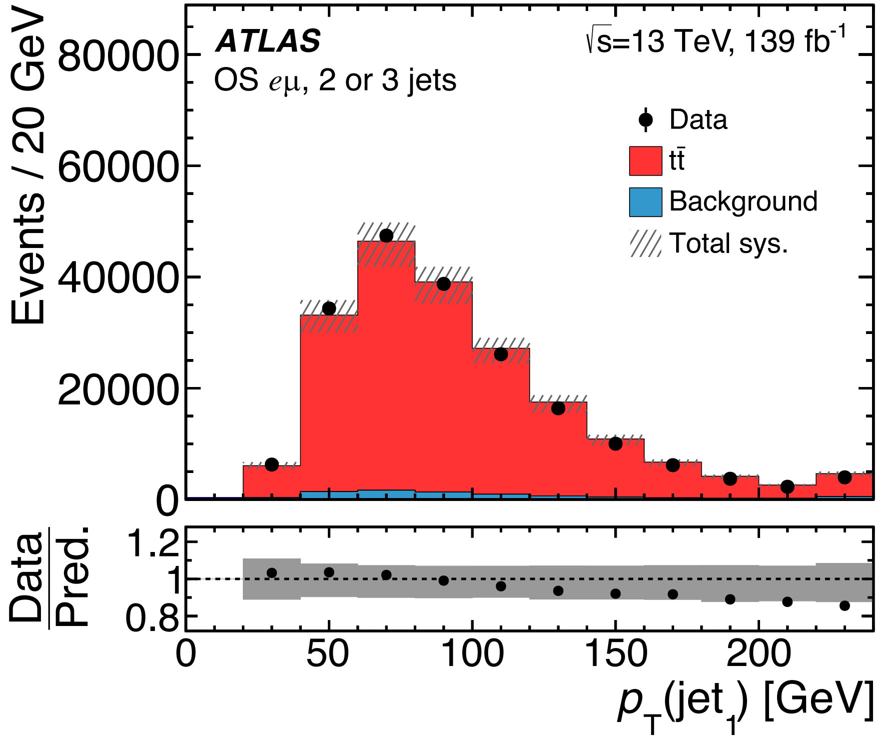
Observed same-sign events

Predicted same-sign events with prompt leptons



Data-MC comparison plots





- Uncertainty band includes MC statistical, theoretical and systematics uncertainties
- MC describes the data well and deviations are covered by uncertainties

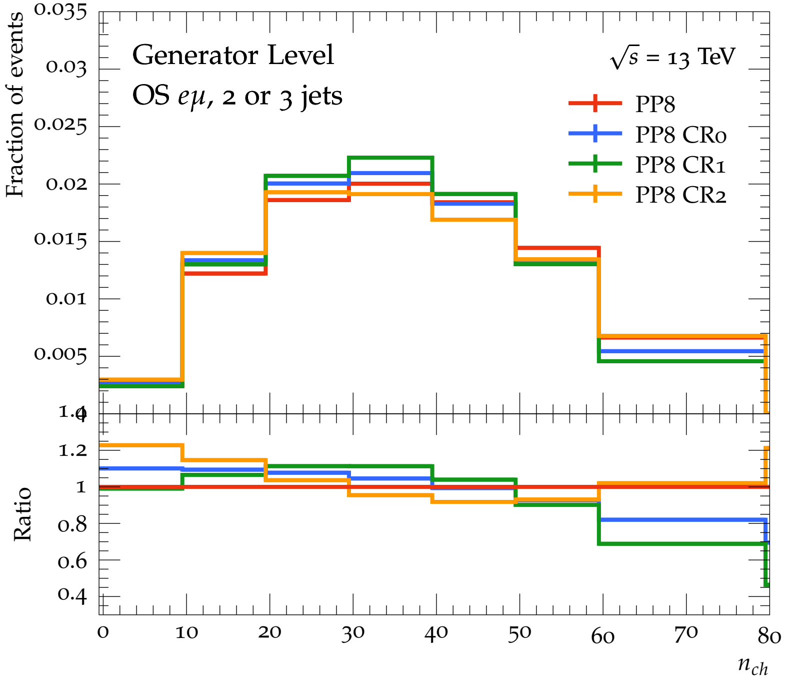
➢ Backgrounds to $t\bar{t}$ events COMPLETED

Sensitive observables

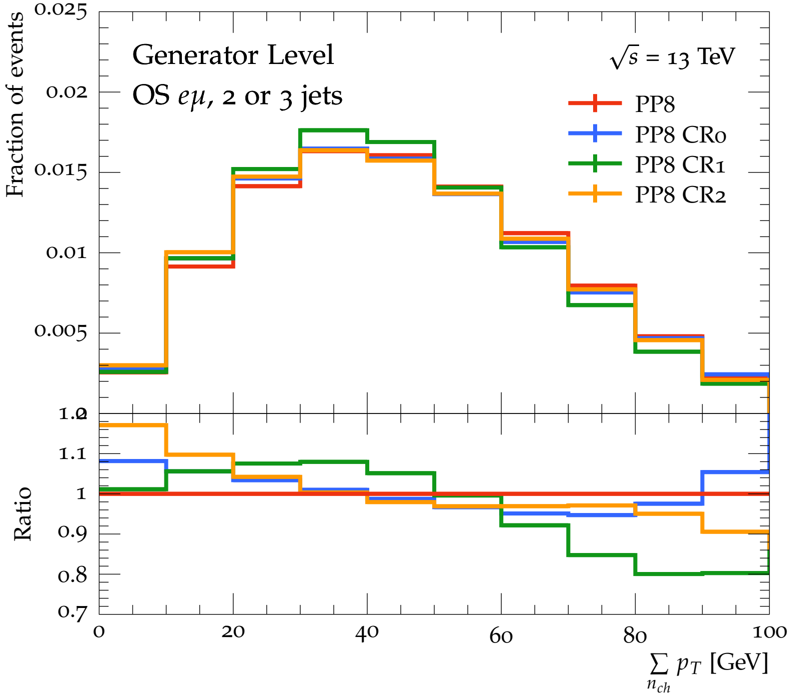
CR0-CR2: tuned to ATLAS data

ATL-PHYS-PUB-2017-008

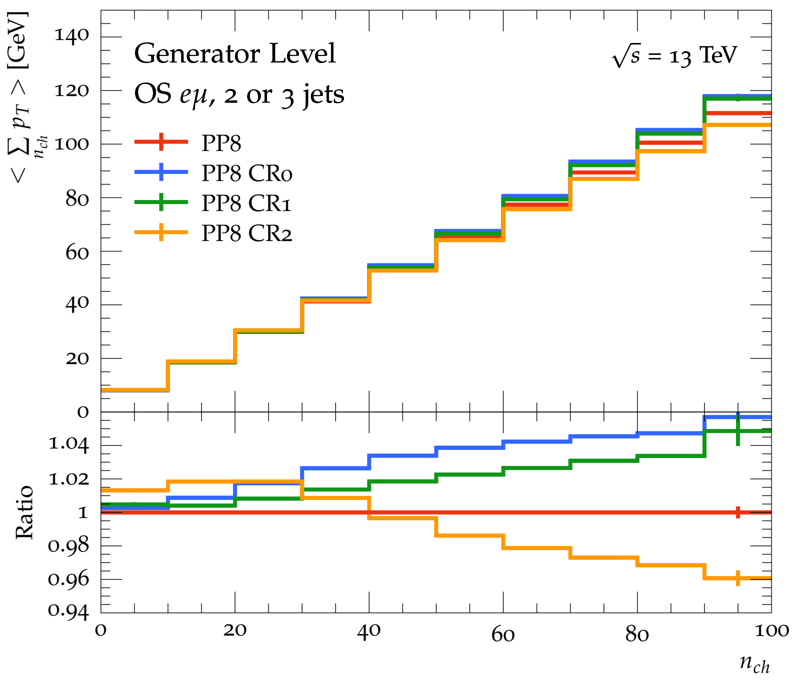
Charged particles multiplicity



Scalar sum of transverse momentum

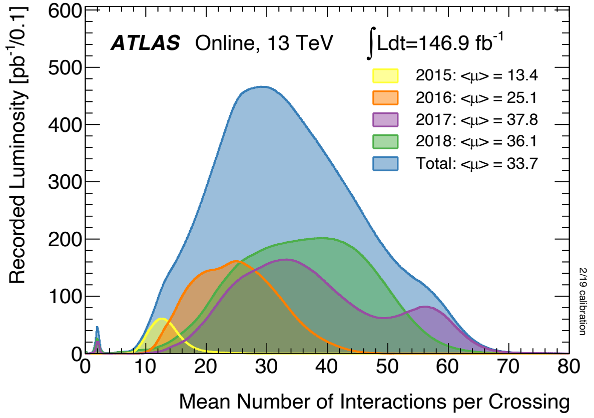


$\sum_{n_{ch}} p_T$ as a function of n_{ch}



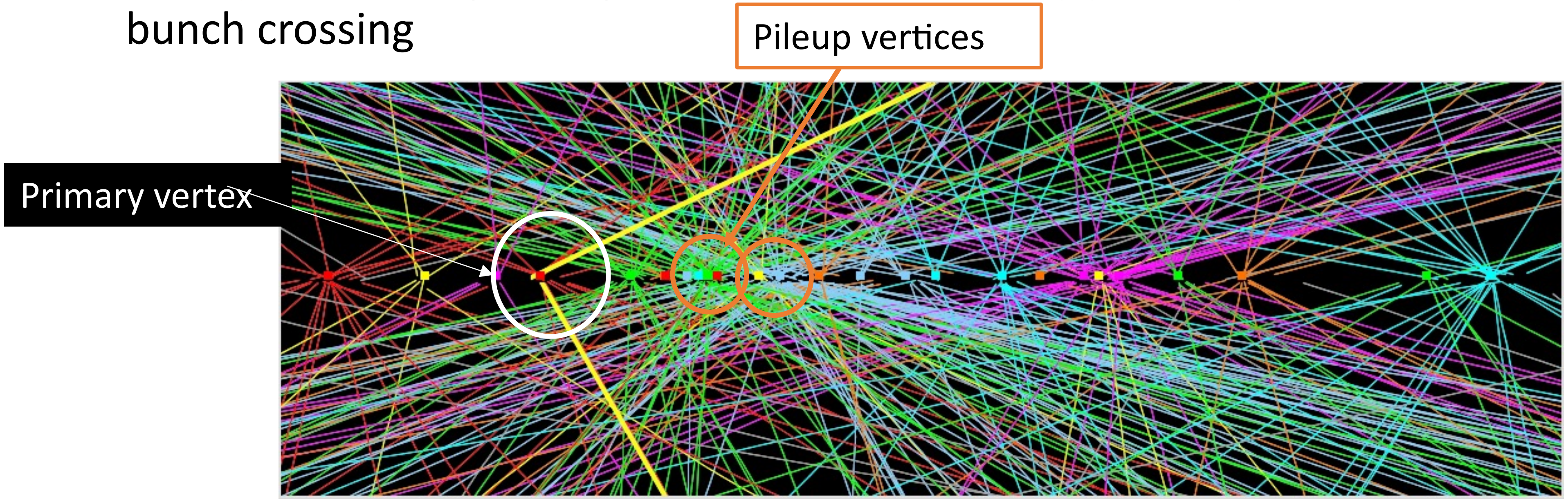
The observables use tracks outside jets because tracks inside jets does not contribute significantly to the discrimination power between CR models.

Backgrounds to primary hard-scatter tracks



Backgrounds to primary hard-scatter tracks

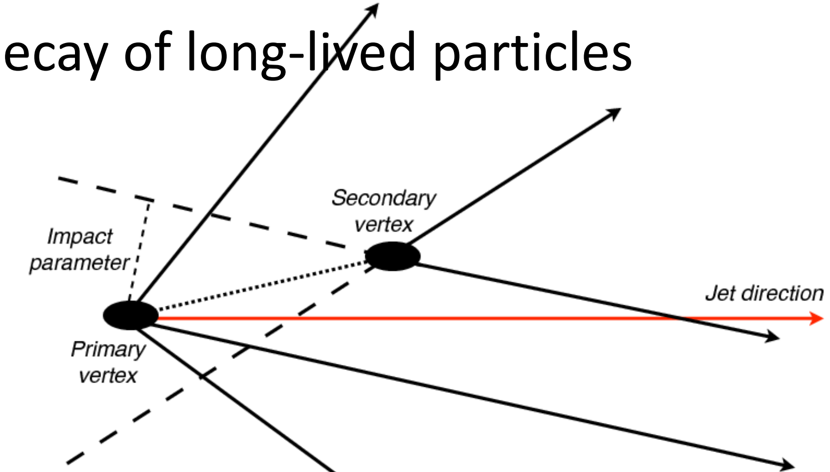
1. Pile-up tracks, originating from additional nearby proton-proton collisions at the same bunch crossing



Pileup vertices

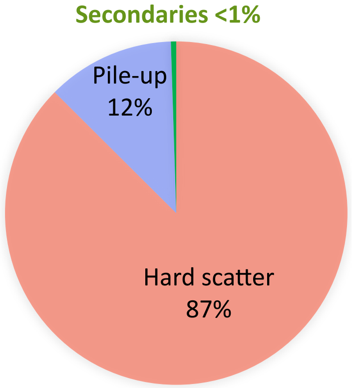
Primary vertex

2. Secondary tracks, decay of long-lived particles



Pile-up background estimation

Even after these requirements tracks are still diluted with pileup and secondary tracks

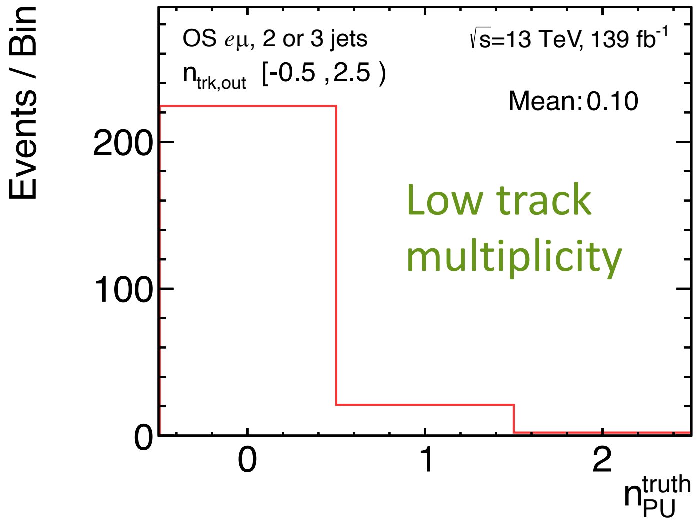


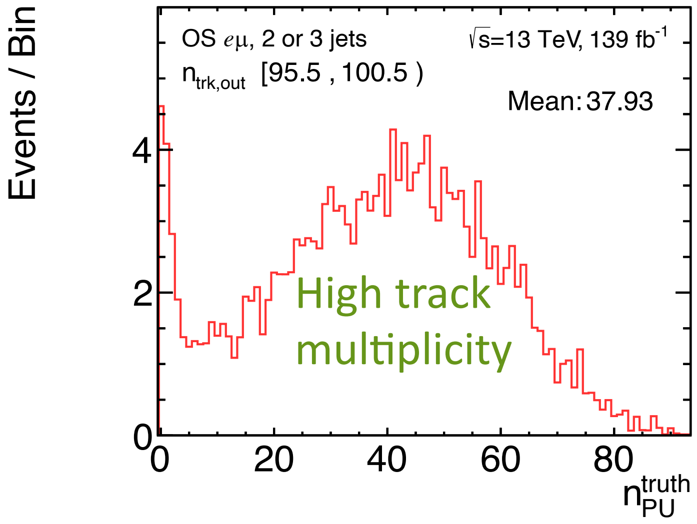
- we can not subtract the MC prediction directly from data, as MC does not perfectly model collision data
  - therefore, we have devised a method which gives a closure in MC
  - then, we subtract this contribution from data in a stochastic way after correcting for Data-MC mismodelling

$$\underbrace{n_{\mathrm{trk,prim}}}_{\text{\# of primary tracks}} = \underbrace{n_{\mathrm{trk,out}}}_{\text{\# of selected tracks}} - \overbrace{c_{\mathrm{PU}}}^{\text{Scale factor (pile-up)}}(\mu, n_{\mathrm{trk,out}}) \cdot \underbrace{n_{\mathrm{PU}}}_{\text{\# of pile-up tracks (stochastically)}} - \overbrace{c_{\mathrm{sec}}}^{\text{Scale factor (secondaries)}} \cdot \underbrace{n_{\mathrm{sec}}}_{\text{\# of secondary tracks (stochastically)}} ,$$

Estimate n_{PU}

- From simulation, create templates of $n_{\mathrm{PU}}^{\mathrm{truth}}$ in bins of track multiplicity

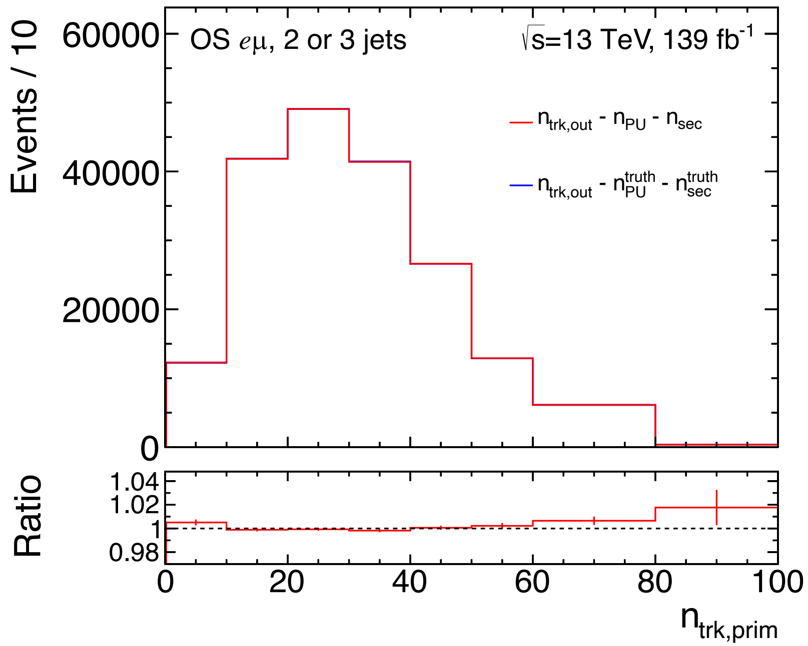




- For data event, with a known track multiplicity
  ⇒ draw a random number from the template of $n_{\mathrm{PU}}^{\mathrm{truth}}$ correspond to the given multiplicity

Closure test

- Compare the estimated hard-scatter track multiplicity with the true one

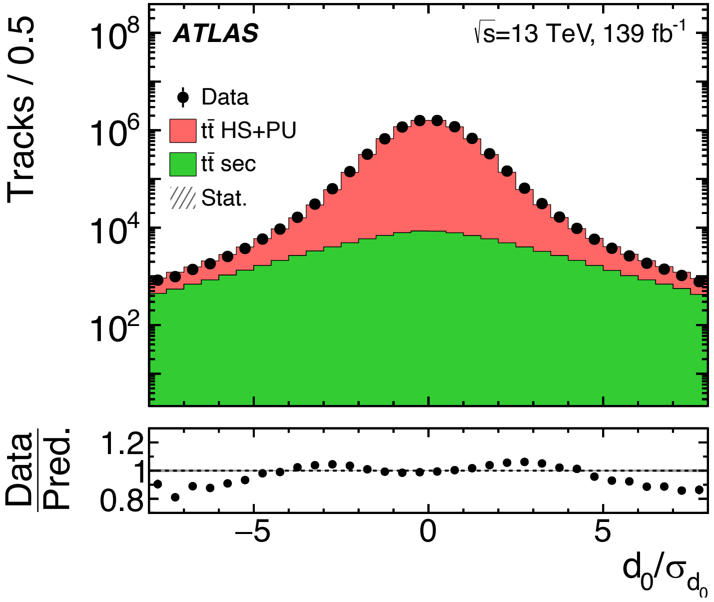


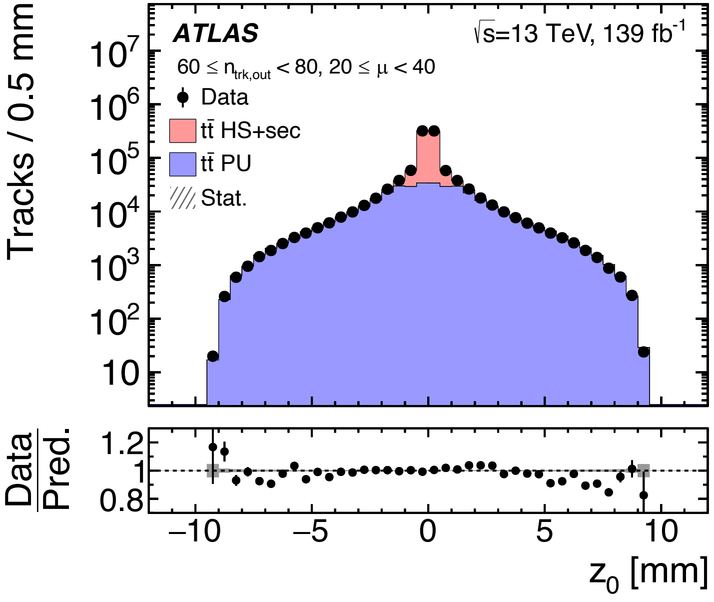
- Non-closure is < 2%, which is taken as an uncertainty

Scale factors estimation

μ: number of interactions per bunch crossing

- Perform a binned maximum-likelihood fit to data
  - C_{sec} : Secondary tracks scale factor → d_0/σ_{d_0}
  - $C_{\text{PU}}\,(n_{\text{trk,out}}, \mu)$: Pile-up scale factor → z_0





C_{sec} = 2.34 ± 0.02

$c_{\text{PU}}\ (n_{\text{trk,out}}, \mu)$

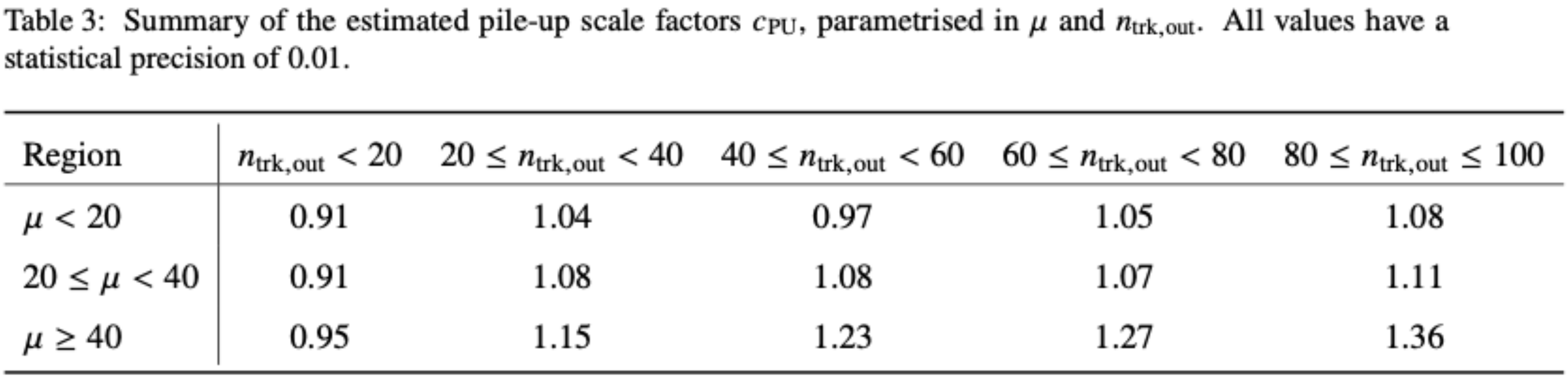


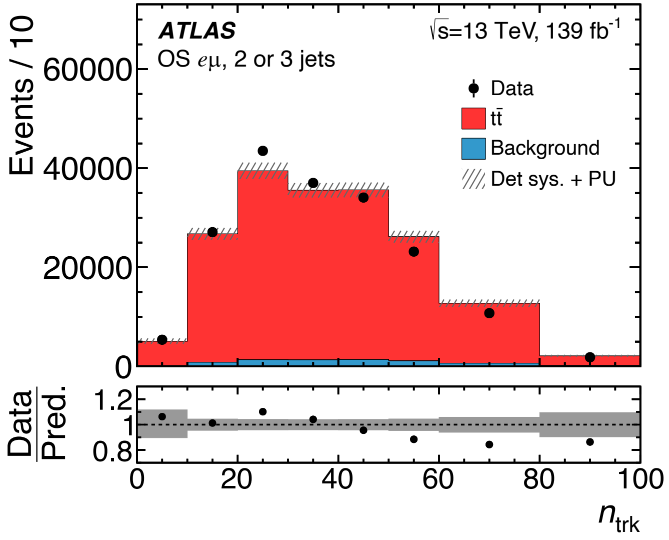
Table 3: Summary of the estimated pile-up scale factors c_{PU}, parametrised in μ and $n_{\text{trk,out}}$. All values have a statistical precision of 0.01.

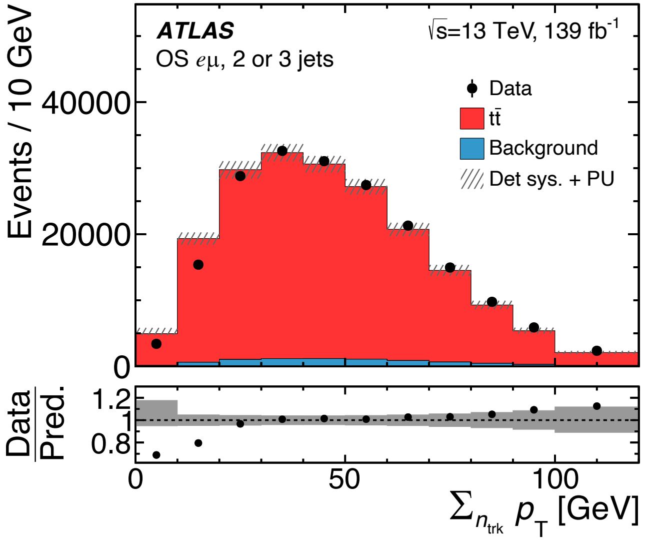
Region	$n_{\text{trk,out}} < 20$	$20 \leq n_{\text{trk,out}} < 40$	$40 \leq n_{\text{trk,out}} < 60$	$60 \leq n_{\text{trk,out}} < 80$	$80 \leq n_{\text{trk,out}} \leq 100$
$\mu < 20$	0.91	1.04	0.97	1.05	1.08
$20 \leq \mu < 40$	0.91	1.08	1.08	1.07	1.11
$\mu \geq 40$	0.95	1.15	1.23	1.27	1.36

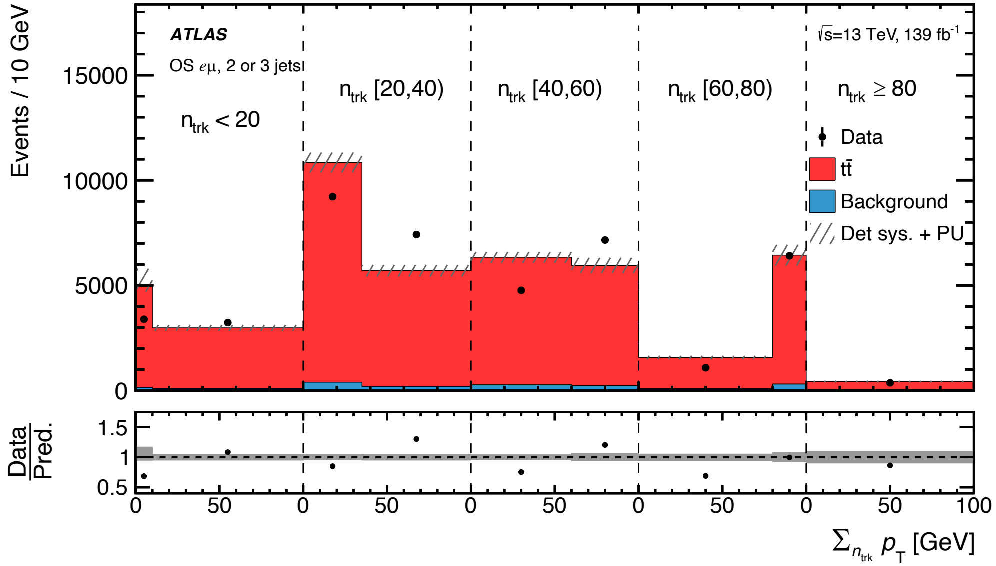
- Backgrounds to primary-hard scatter tracks also COMPLETED

Reconstruction-level observables

- All background contributions are estimated







But, Data distributions are distorted by detector effects and they differ from their true value

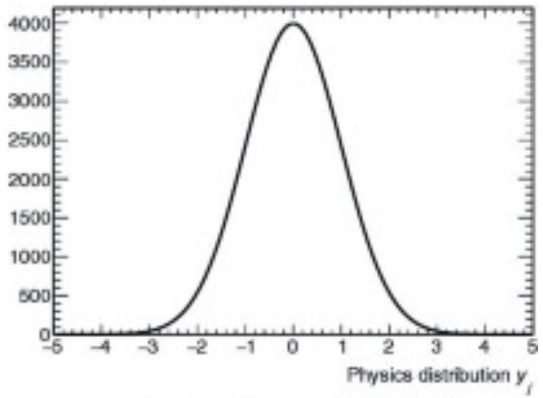
ACTION NEEDED

How to correct for detector effects?

Unfolding

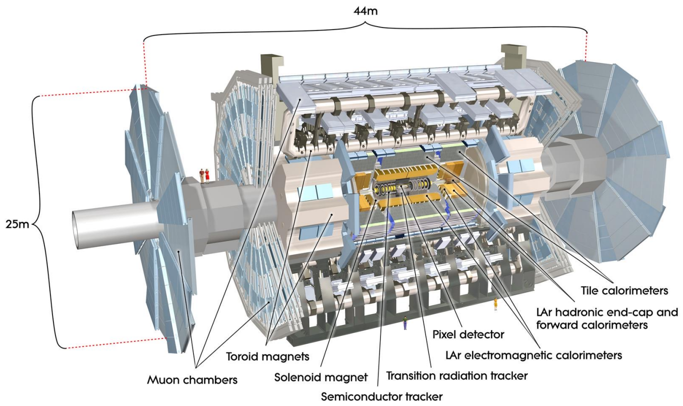
- Procedure to correct for detector effects (finite resolution, and limited efficiency and acceptance)

Truth (particle-level) ⊗ Detector effect = Observed data



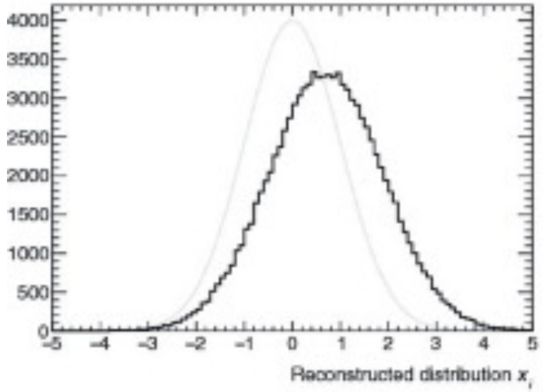
Physics distribution y_j
(particle-level)

⊗



Detector response

=



Reconstructed distribution x_i
(detector-level)

Unfolding

Extraction of the differential cross-section

The master-formula:

Data

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^{i}} = \frac{1}{\mathcal{L} \cdot \Delta X^{i} \cdot \epsilon_{\mathrm{eff}}^{i}} \cdot \sum_{j} R_{ij}^{-1} \cdot f_{\mathrm{acc}}^{j} \cdot \left(N_{\mathrm{obs}}^{j} - N_{\mathrm{bkg}}^{j} \right)$$

Truth

Extraction of the differential cross-section

DATA

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\mathcal{L} \cdot \Delta X^i \cdot \epsilon_{\mathrm{eff}}^i} \cdot \sum_j R_{ij}^{-1} \cdot f_{\mathrm{acc}}^j \cdot \left(\underline{N_{\mathrm{obs}}^j - N_{\mathrm{bkg}}^j}\right)$$

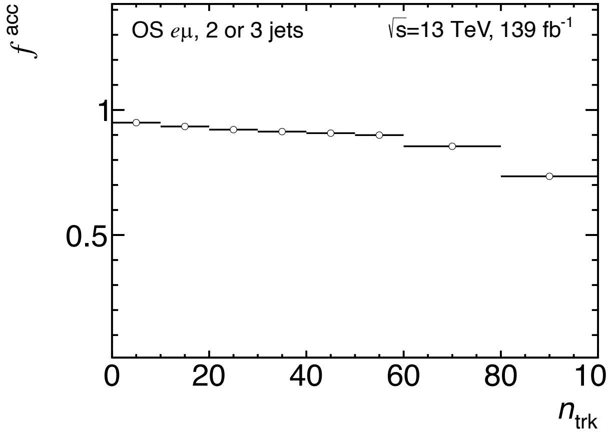
TRUTH

1. Subtract background events from data

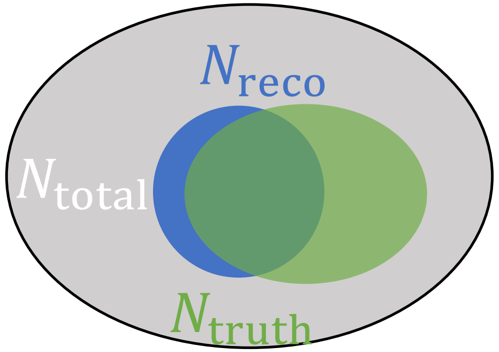
Extraction of the differential cross-section

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

Obtained from signal MC



1. Subtract background events
2. f^j_{acc}: correct for $t\bar{t}$ events that fall outside the fiducial acceptance

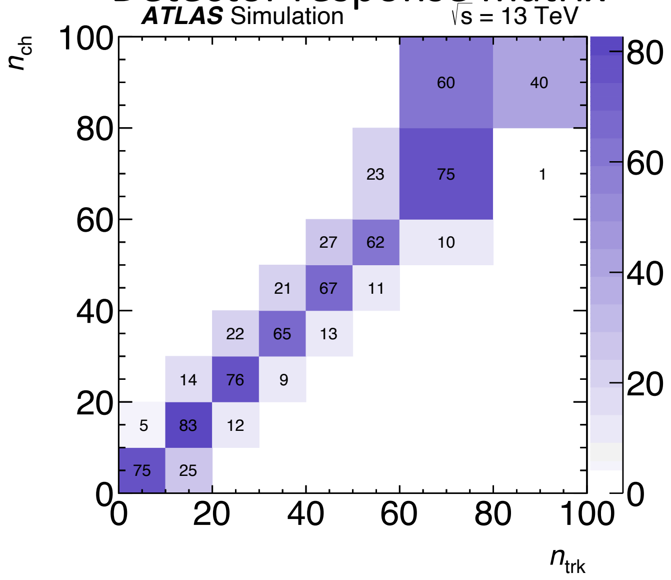


Extraction of the differential cross-section

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

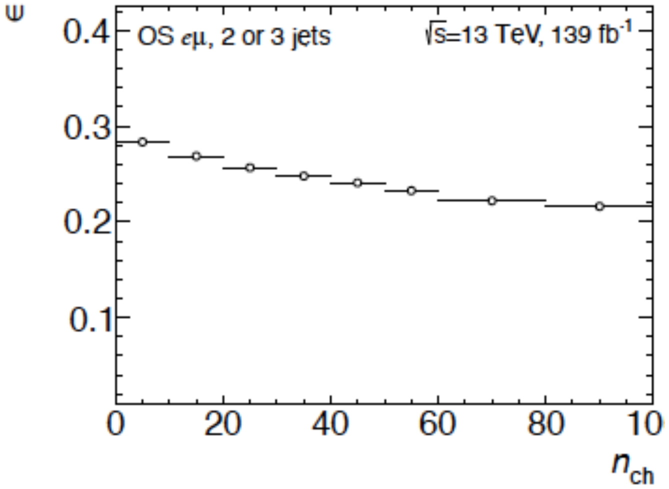
1. Subtract background events
2. Correct for events that are in the reco-level but aren't in truth
3. Remove the detector effects
   ⇒ using the Iterative Bayesian Unfolding method

Detector response matrix

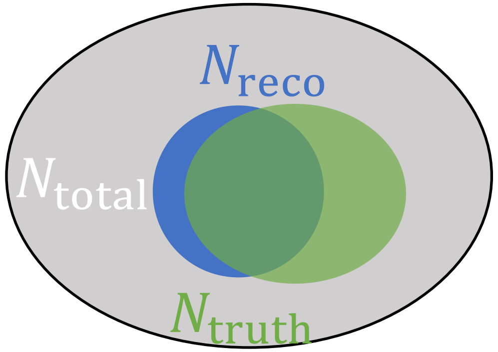


Extraction of the differential cross-section

$$\frac{\mathrm{d}\sigma_{t\bar{t}}}{\mathrm{d}X^i} = \frac{1}{\mathcal{L}\cdot\Delta X^i\cdot\underline{\epsilon_{\mathrm{eff}}^i}}\cdot\sum_j R_{ij}^{-1}\cdot f_{\mathrm{acc}}^j\cdot\left(N_{\mathrm{obs}}^j - N_{\mathrm{bkg}}^j\right)$$



1. Subtract background events
2. Correct for events that are in the reco-level but aren't in truth
3. Remove the detector effects
4. Extrapolate to the truth phase-space



Extraction of the differential cross-section

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

Integrated luminosity

Bin width

1. Subtract background events
2. Correct for events that are in the reco-level but aren’t in truth
3. Remove the detector effects
4. Extrapolate to the truth phase-space
5. Convert event count to cross-section

Unfolding validation test – Stress test

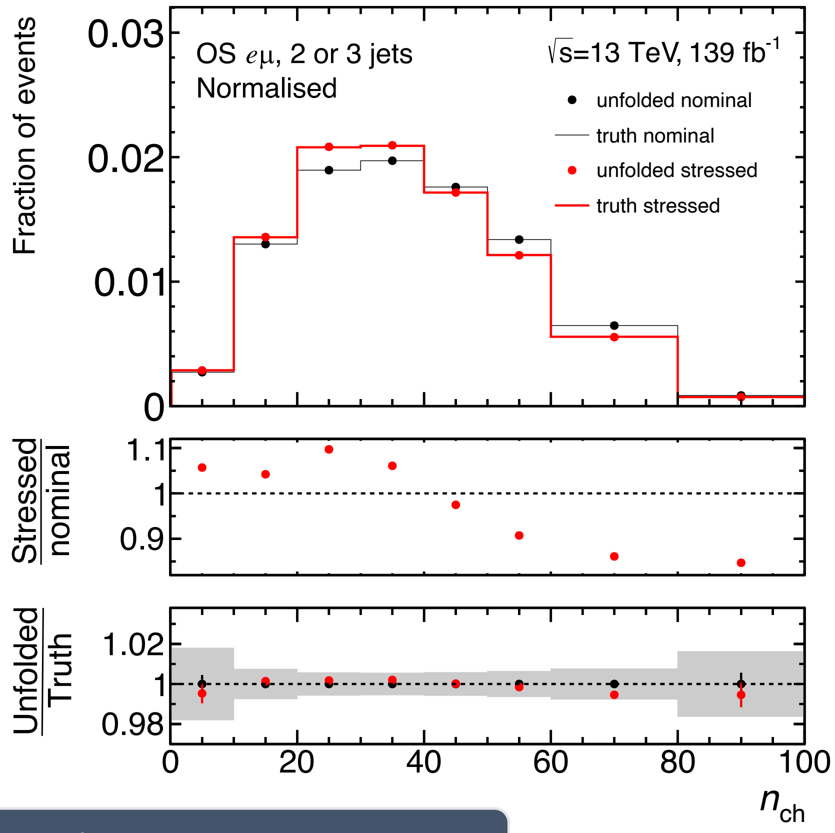
- Aims to verify that the IBU able to recover a truth distribution different from the predicted
  ↪ ability to maintain unexpected features in the data

Procedure:

- re-weight the truth distribution
- unfold the corresponding reweighted reco. using nominal migration matrix

Reweighting function:

- Data-driven reweighting $\left(\frac{\text{Data} - \text{bkg}}{\text{signal}}\right)$



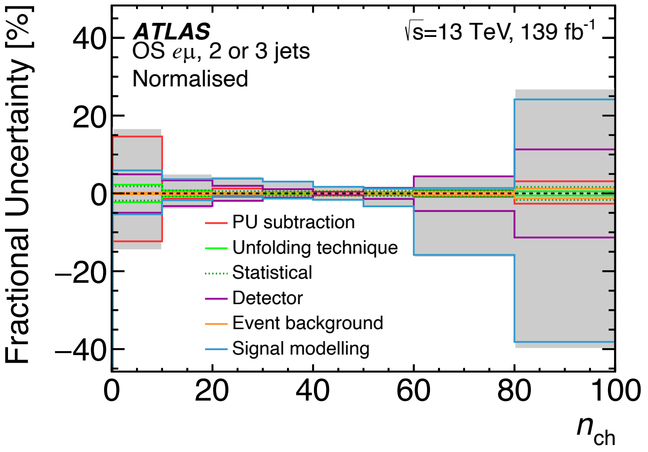
Any (non-closure)deviation between the unfolded and the truth is taken as uncertainty

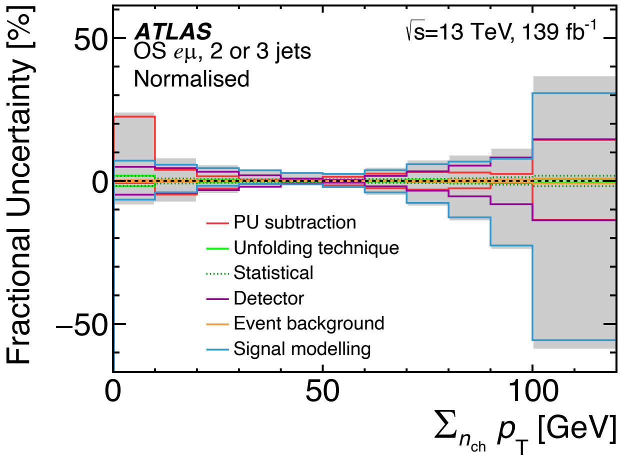
Systematic uncertainties

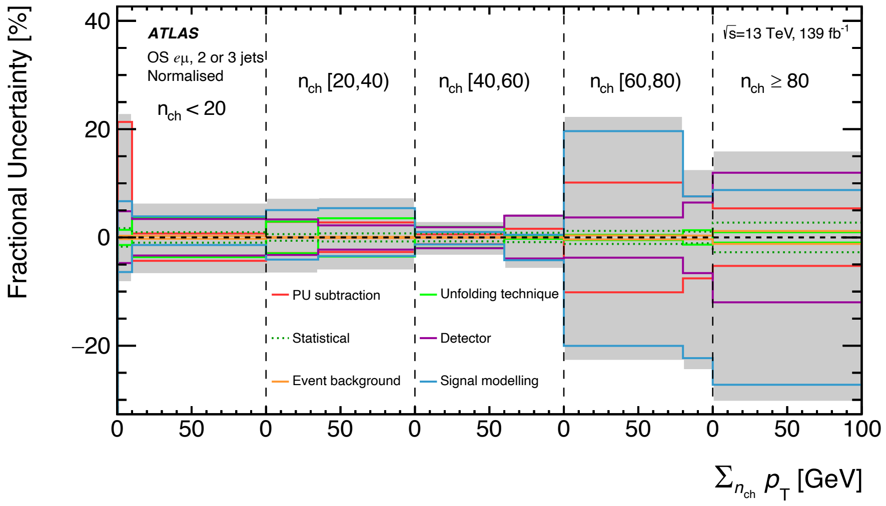
- Experimental uncertainties, related to objects reconstruction (tracks, jets, ..)
  - ➢ The varied prediction is unfolded
  - ➢ Uncertainty is the relative difference of the unfolded systematic variation w.r.t nominal unfolded distribution
- Signal modelling uncertainties:
  - Parton shower (Powheg+Herwig713 vs. Powheg+Pythia 8)
  - Colour reconnection
  - ➢ Uncertainty is the difference between the unfolded and the particle-level distribution of the systematic variation
- Background modelling uncertainties
  - Event-based backgrounds
  - Track-based backgrounds
- Unfolding techniques

Systematic uncertainties

- Dominant uncertainties:
  - pile-up tracks background estimation
  - Signal Modelling uncertainties





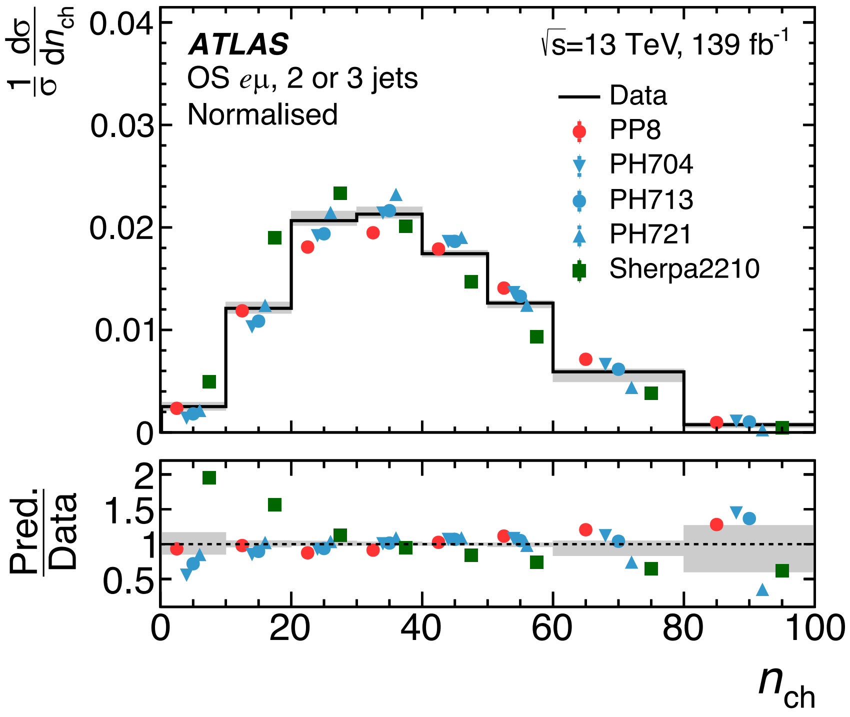


Results and Summary

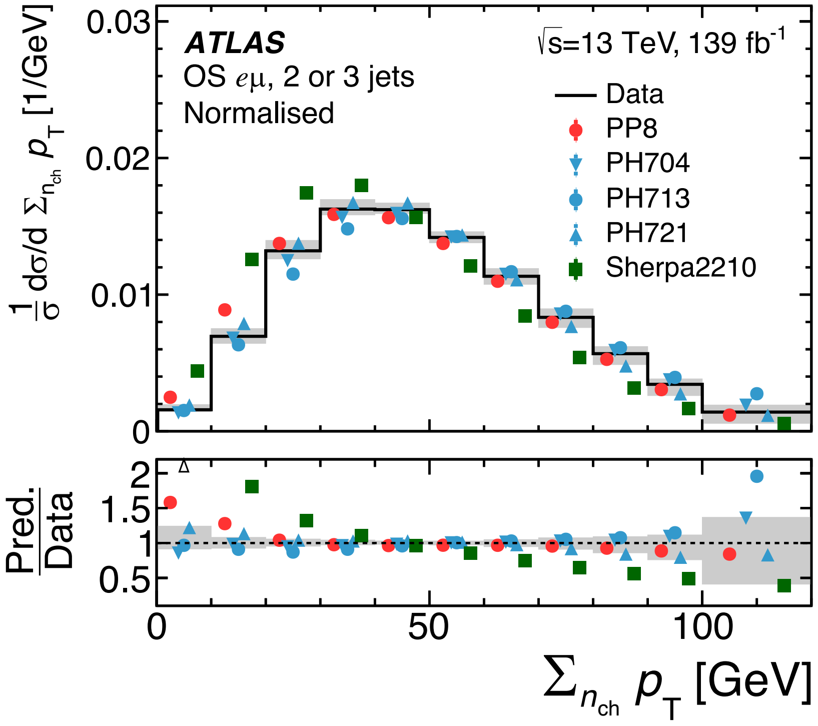
Results (1)

- Measured distributions are compared to the prediction from different generators

Charged particles multiplicity



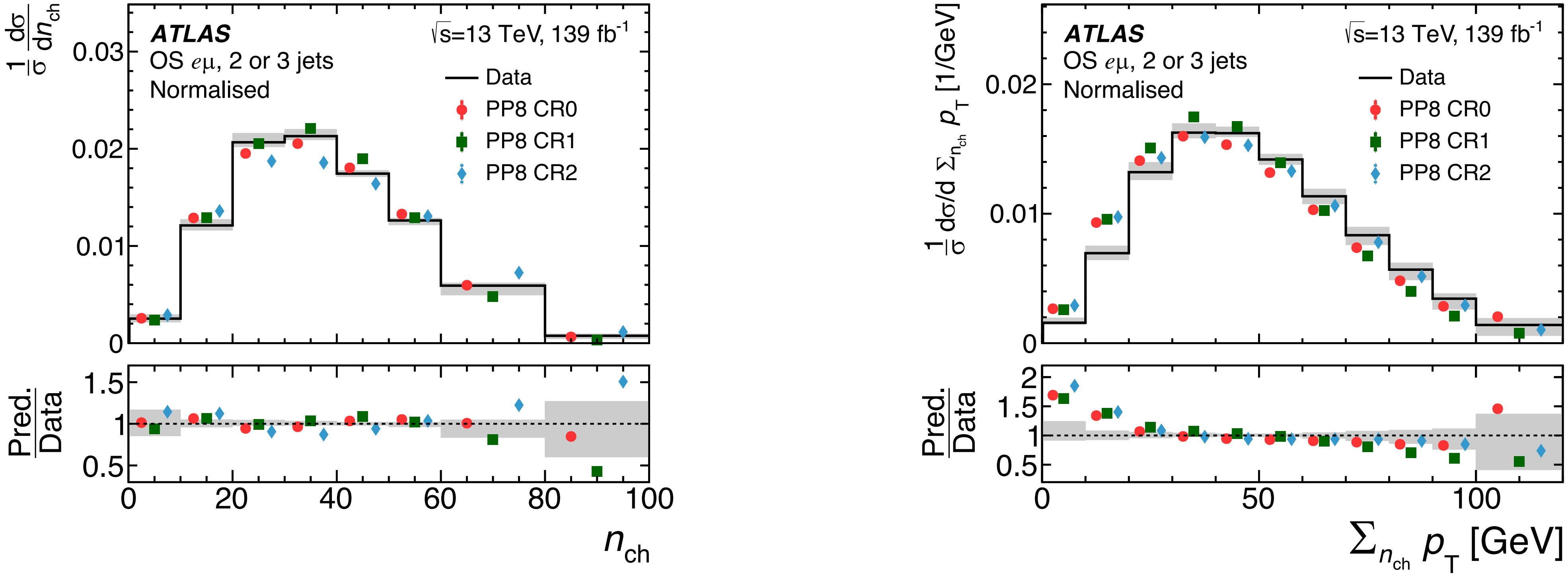
Charged particles $\sum p_T$



- Measured data disagree with the predictions from Sherpa 2, which does not include CR effects
- The n_{ch} is approximately equally well described by Pythia 8 and Herwig 7
- The $\sum_{n_{\mathrm{ch}}} p_T$ has a better agreement with Herwig 7, especially in $\sum_{n_{\mathrm{ch}}} p_T < 20$ GeV

Results (2)

- Measured distributions are compared to the prediction from different CR models in PYTHIA 8

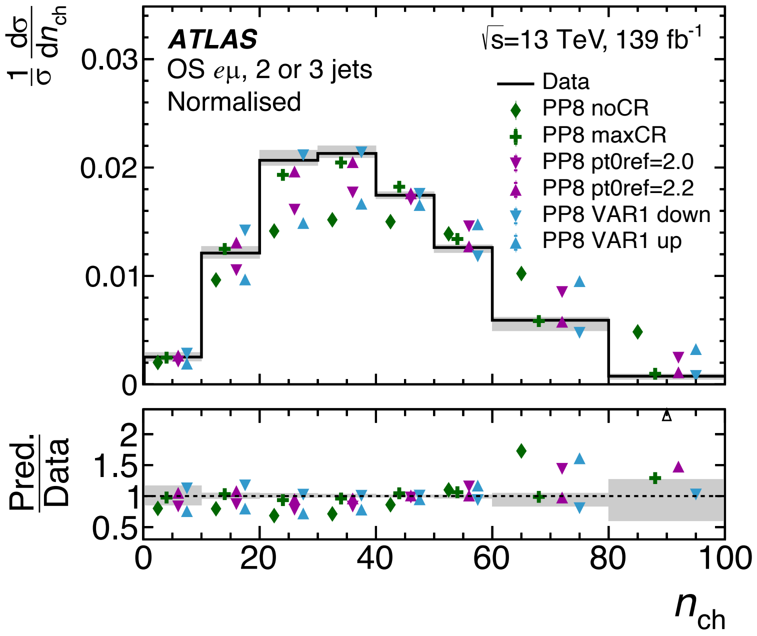


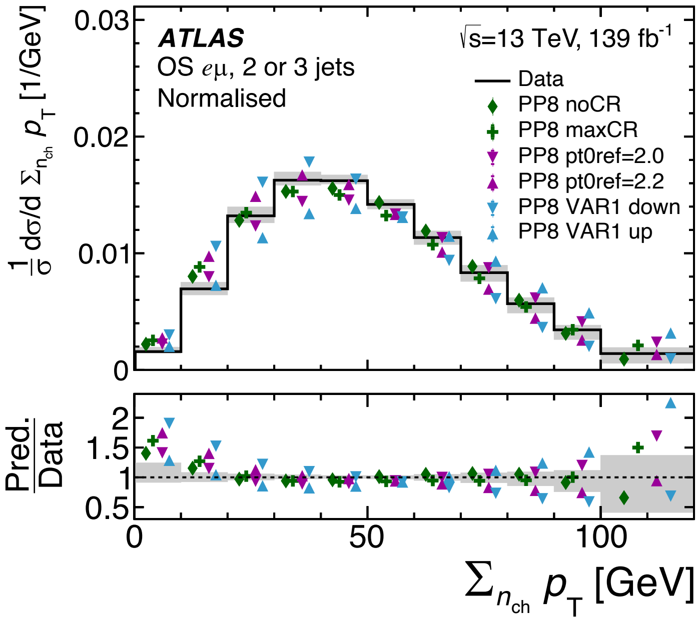
- The n_{ch} is best described by the CR0 (MPI-based) model
- Similar to the nominal PP8 A14 tune, none of the models can describe $\sum_{n_{\text{ch}}} p_T < 20\ \text{GeV}$ well

Results (3)

Comparison with CR and UE parameters variation in Pythia 8

variation	Varied parameter
noCR	R_{range}= 0 (default 1.71)
maxCR	R_{range}=10
Var1 down	MPI α_s= 0.121 and R_{range} = 1.69 (default MPI α_s = 0.126)
Var1 up	MPI α_s= 0.131 and R_{range} = 1.73





- ➢ noCR does not describe the data
- ➢ Maximal probability still compatible with the measurement
- ➢ Clear sensitivity to pt0ref and Var1 → future tuning both parameter should be included

Summary

Three observables sensitive to colour reconnection are measured in $t\bar{t}$ events

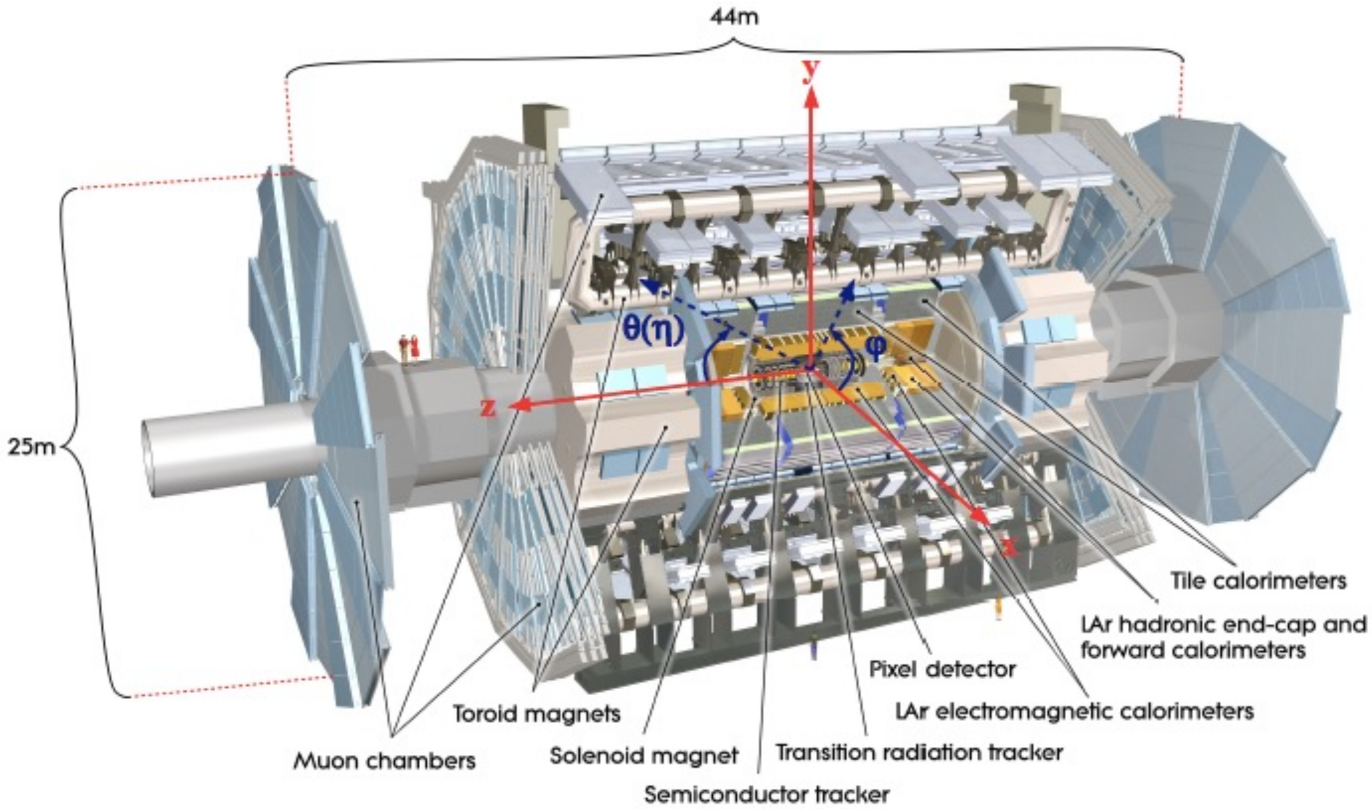
- charged particle multiplicity (n_{ch})
- scalar sum of charged-particles transverse momenta ($\sum_{n_{ch}} p_T$)
- $\sum_{n_{ch}} p_T$ in bins of n_{ch}

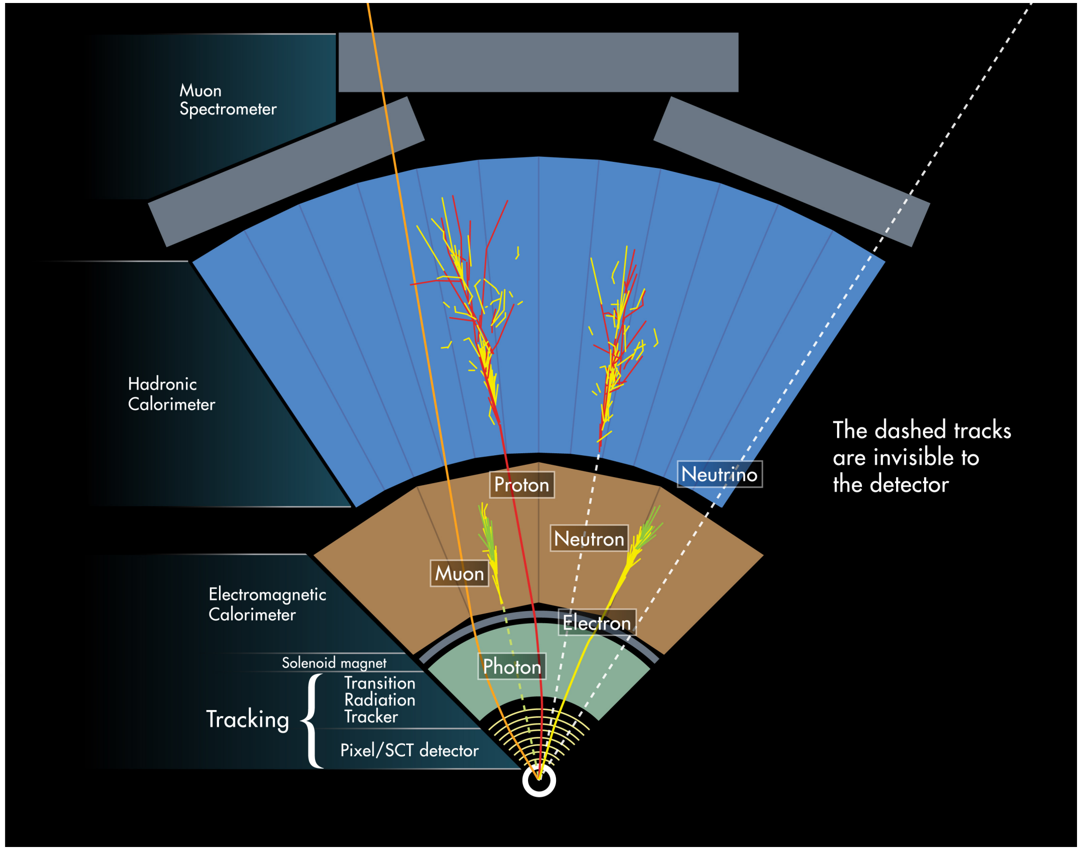
- Dominant uncertainties are signal modelling:
  - Track-background subtraction (low n_{ch})
  - Parton shower (mainly tail)

➢ Paper is accepted by EPJC journal and available as preprint on arXiv arXiv:2209.07874

➢ The result can be used as input for future tuning of MC (CR and MPI parameters)

Backup

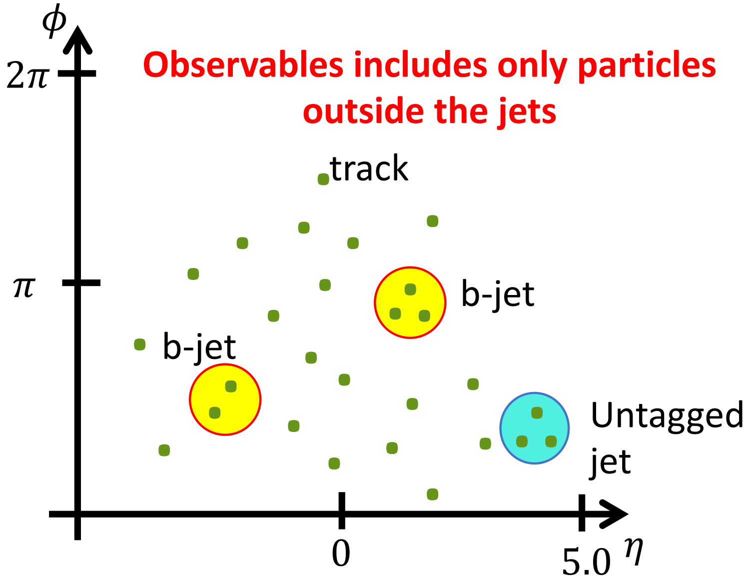




The dashed tracks are invisible to the detector

Backup: tracks selection

- $p_T > 500$ MeV
- $|\eta| < 2.5$
- 9 (11) silicon hits for $|\eta| < 1.65$ ($|\eta| > 1.65$)
- 1 IBL or B-layer hit
  - IBL: is the innermost layer of ATLAS pixel detectors

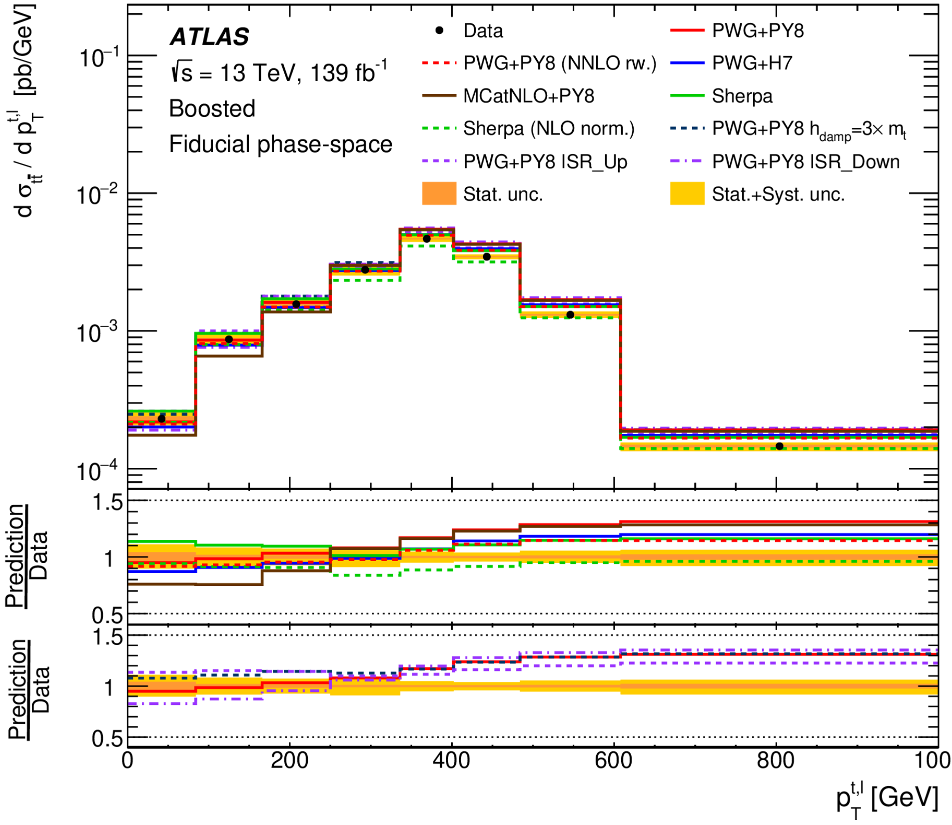


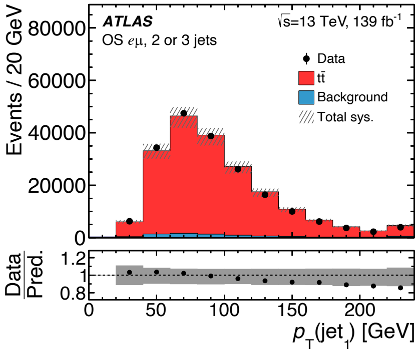
Observables includes only particles outside the jets

- Additional cuts:
  - $\Delta R(\text{track}, \text{jet}) > 0.4$
  - $\Delta R(\text{track}, \text{lep}) > 0.01$
  - $|d_0| < 0.5$ mm
  - $|z_0 \cdot \sin(\theta)| < 0.5$ mm w.r.t. PV

Backup: p_T (jet$_1$) miss-modelling

- In agreement with previous cross-section measurements by ATLAS and CMS
- Consistently observed a softer top-quark p_T spectrum (and related distributions)
- This discrepancy is at least partially due to missing NNLO corrections





Backup - Overview of CR models in Pythia8

Parameters summary

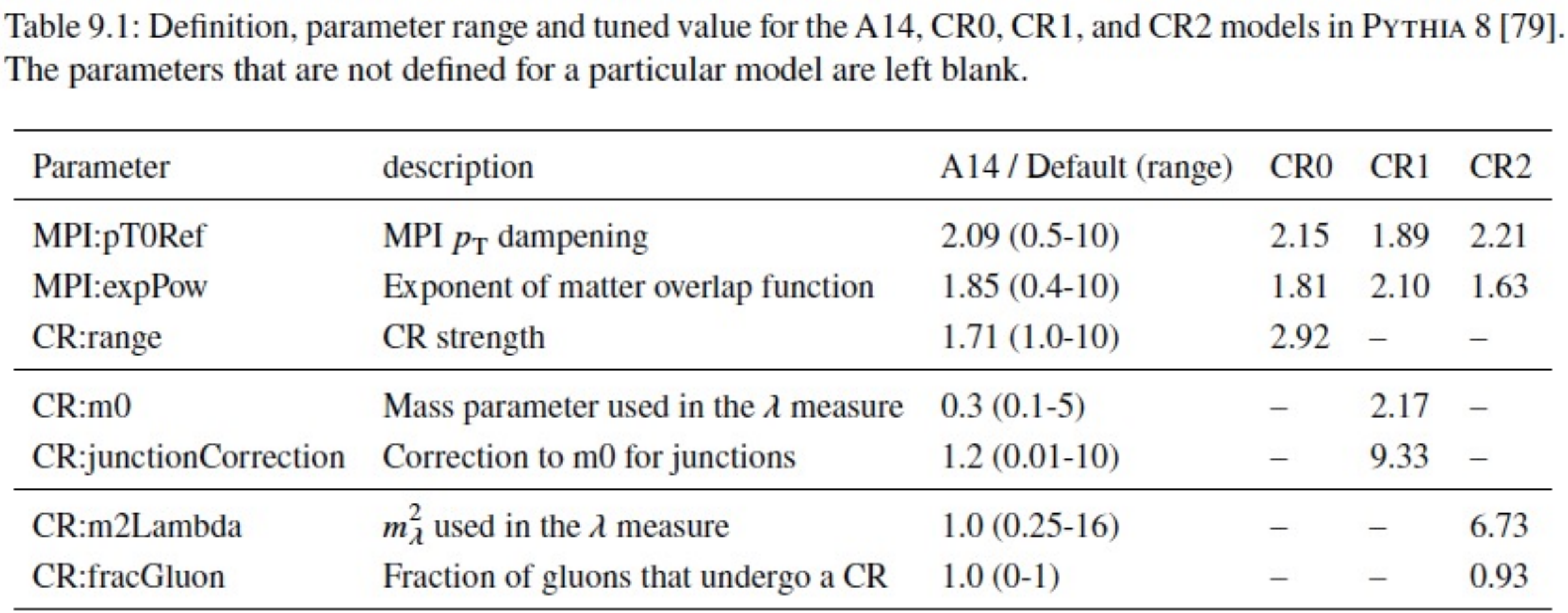


Table 9.1: Definition, parameter range and tuned value for the A14, CR0, CR1, and CR2 models in PYTHIA 8 [79]. The parameters that are not defined for a particular model are left blank.

Parameter	description	A14 / Default (range)	CR0	CR1	CR2
MPI:pT0Ref	MPI p_T dampening	2.09 (0.5-10)	2.15	1.89	2.21
MPI:expPow	Exponent of matter overlap function	1.85 (0.4-10)	1.81	2.10	1.63
CR:range	CR strength	1.71 (1.0-10)	2.92	–	–
CR:m0	Mass parameter used in the λ measure	0.3 (0.1-5)	–	2.17	–
CR:junctionCorrection	Correction to m0 for junctions	1.2 (0.01-10)	–	9.33	–
CR:m2Lambda	m^2_λ used in the λ measure	1.0 (0.25-16)	–	–	6.73
CR:fracGluon	Fraction of gluons that undergo a CR	1.0 (0-1)	–	–	0.93

Backup - Overview of CR models in Pythia8

MPI-based model (CR0)

reconnection probability

$$P_{\mathrm{rec}}\,(p_T) = \frac{(R_{\mathrm{rec}}\,p_{\mathrm{T0}})^2}{(R_{\mathrm{rec}}\,p_{\mathrm{T0}})^2 + p_{\mathrm{T}}^2}$$

CR range

Soft dampening scale

MPIs are calculated by using the pQCD cross section
- this diverges for small transverse momenta
- Solution: Introduction of a "smooth dampening" scale

$$\frac{d\sigma}{dp_T^2} \sim \frac{\alpha_S^2(p_T^2)}{p_T^4} \rightarrow \frac{\alpha_S^2(p_{T,0}^2+p_T^2)}{(p_{T,0}^2+p_T^2)^2}$$

- p_{T0} regularises the partonic cross-section to avoid divergence at low p_{T}

$$p_{\mathrm{T0}} = p_{\mathrm{T0}}^{\mathrm{ref}} \left(\frac{E_{\mathrm{CM}}}{E_{\mathrm{CM}}^{\mathrm{ref}}}\right)^{E_{\mathrm{CM}}^{\mathrm{pow}}}$$

  - $p_{\mathrm{T0}}^{\mathrm{ref}}$ is the value of p_{T0} at a reference energy $E_{\mathrm{CM}}^{\mathrm{ref}}$
  - $E_{\mathrm{CM}}^{\mathrm{pow}}$ us a tunable parameter

IBU

- use Bayes theorem recursively and use our truth distribution as a prior

$$\underbrace{P\,(\text{Parameter}|\text{Data})}_{\text{posterior truth pdf}} = \frac{\overbrace{P\,(\text{Data}|\text{Parameter})}^{\text{response matrix}}\;\overbrace{P(\text{Parameter})}^{\text{prior truth pdf, taken from nominal MC to speed convergence}}}{\underbrace{P(\text{Data})}_{\text{observed pdf}}}$$

Iterations optimisation

- With higher iterations, the negative correlations increases
  - correlations are expected to go from positive to negative it is expected to have a minimum

$$\rho_i = \sqrt{1 - \left(\mathrm{Cov}_{ii} \cdot \mathrm{Cov}^{-1}_{ii}\right)^{-1}}\,, \qquad \rho_{\mathrm{avg}} = \frac{1}{N_b} \sum_{i=1}^{N_b} \rho_i$$