ANALYSIS STATUS: DATASETS AND TRIGGER SELECTION

Workshop Roma-Napoli, Napoli, 18/12/2024

Run 3 ATLAS dataset

- ATLAS data is collected in AOD (Analsys Object Data), particularly in skimmed AOD (Derived AOD → DAOD)
  - DAOD_PHYS is the unified data format, but it does not contain constituents information…
- **Our new requested DAOD_LLJ1 derivation:**
  - DAOD_PHYS format as baseline, with addition of constituents for jets
- **UFO collection for large-R jets, upgrade to previous TrackTopoClusters**
  - Possiblity to add new features for ML with respect to LHCOlympics datasets
  - Constituents taste, associated tracks information
- **Trigger and event skimming applied**
  - Due to significant size
- **Data**
  - 2022 and 2023 production;

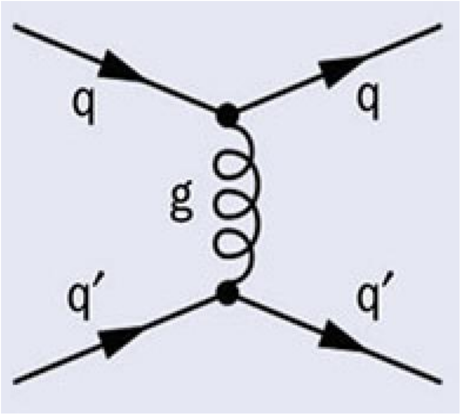
```
### 1 jet object skimming
sel_1jet_template = "((count (abs({0}eta) < 2.8 && {0}pt > 150*GeV && {0}m > 30*GeV)  >= 1))"
topology_selection_1jet = "({})".format(
    " || ".join([sel_1jet_template.format(j) for j in largeRJetsForSkimming])
)
```

```
### trigger skimming
TriggersList = [
    ### baseline run-2
    'HLT_j360_a10_lcw_sub_L1J100',
    'HLT_j420_a10_lcw_L1J100',
    'HLT_j460_a10t_lcw_jes_L1J100',
    ### new run-3
    'HLT_j460_a10sd_cssk_pf_jes_ftf_preselj225_L1J100',
    'HLT_j460_a10_lcw_subjes_L1J100',
    'HLT_j460_a10r_L1J100',
    ### new run-3 mass cut
    'HLT_j420_35smcINF_a10sd_cssk_pf_jes_ftf_preselj225_L1J100',
    'HLT_j420_35smcINF_a10t_lcw_jes_L1J100',
]
```

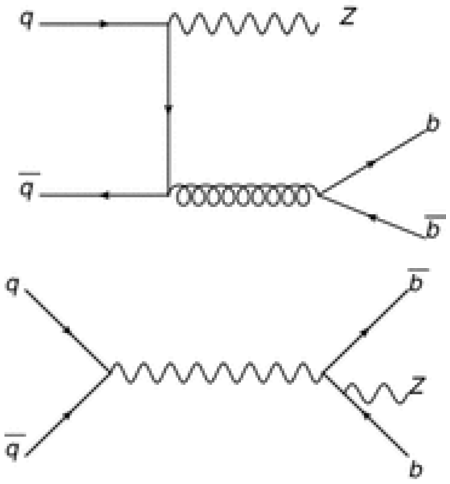
Run3 MC production: background

- mc23d only, due to size constraints
  - Most dominant: QCD di-jet, divided in pT slices (JZ 0-9incl)
  - Can be used for machine learning validation or background estimation

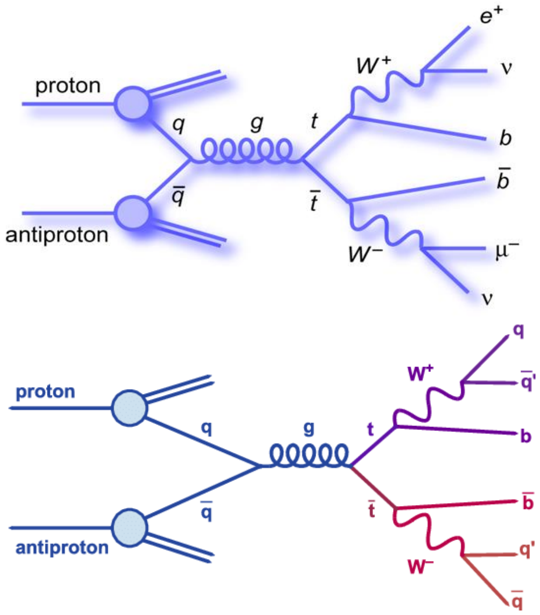
QCD di-jet (> 97%)



W/Z + jets (~2%)



$t\bar{t}$ couples (~1%)

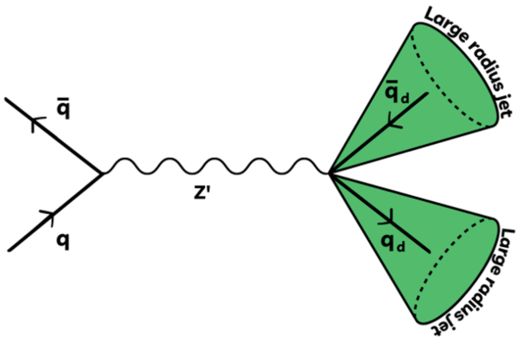


Run3 MC production: signal

- mc23d only, due to size constraints
  - HVT signals;
  - Darkjets signals;
  - 3 prong signals;

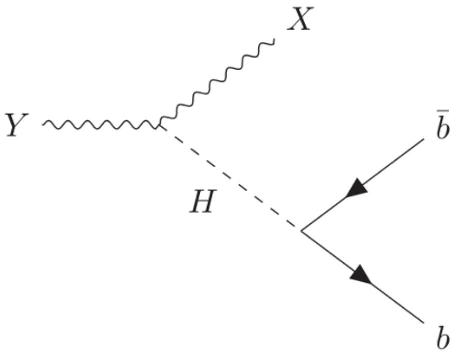
- Produced for machine learning validation and results interpretation

- DarkJets



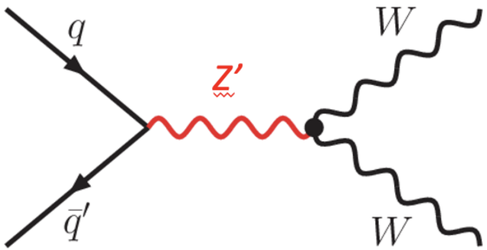
Z’ (3500 GeV) → $q_d\bar{q}_d$ (20 GeV) → large-R jets

- YXH



Y (3400 GeV) → X (200 GeV) H (125 GeV) → large-R jets

- VVJJ



Z’ (1900 GeV) → WW (80 GeV) → large-R jets

Easyjet

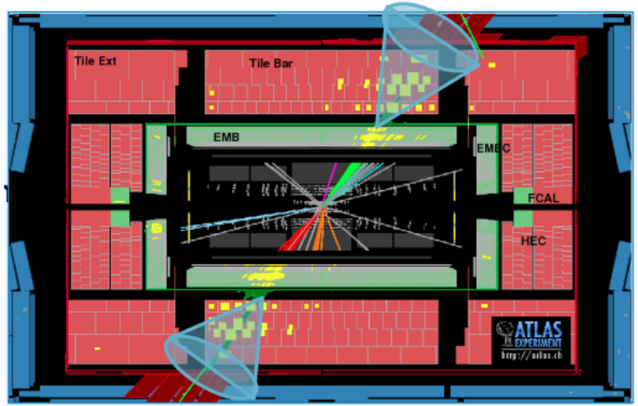
- Framework chosen to convert DAOD datasets (very hard to analysize) in ROOT ntuples (much easier to handle): https://gitlab.cern.ch/atlas-roma1-napoli/easyjet/-/tree/rebase?ref_type=heads

- Preselection applied on jets during conversion:
  - p_T > 200 GeV
  - $|\eta| < 2$

- Branches available in final ntuples:
  - Large-R jets and small-R jets 4 momenta
  - Large-R jets and muons trigger items
  - Trigger matching to large-R jets info for each trigger item included
  - Large-R jets constituents 4 momenta
  - Jets associated tracks 4 momenta,
  - Leptons (muons and electrons) 4 moments

- First production on CERN GRID available at:
  - /eos/atlas/atlascerngroupdisk/phys-hmbs/mbl/AD_JJ/

- Addition of new info after first production:
  - #jets, #constituents per jet, 4-momenta of leading dijet couples
  - PDGID of truth particles to be included in MC samples



FastFrames

- Postprocessing step on Easyjet ntuples, necessary to include event weight for MC samples: https://gitlab.cern.ch/atlas-roma1-napoli/adjj-fast-frames
  - **Event weight** takes into account the actual number of events expected under certain conditions of **luminosity, cross section and pile up**
  - No need to generate all events in this way

- FF is very handy to combine minintuples of samples into one single ntuple, add selections and new variables and decide branches to keep
  - Selections for objects definition: leading jet m > 50 GeV and subleading jet m > 50 GeV
  - nJets > 2

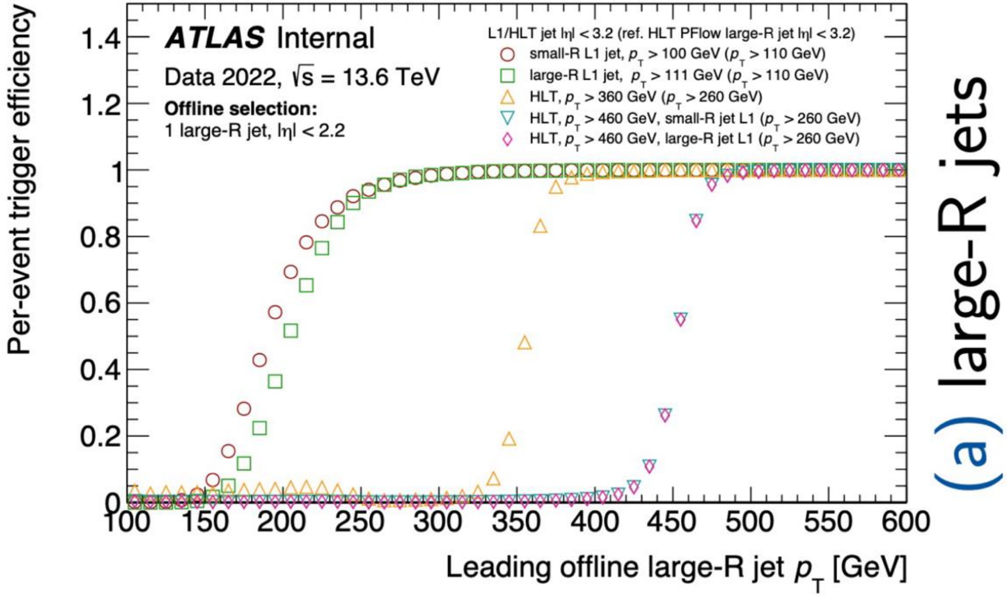
- Relatively fast, takes longer if constituents info is included
  - Size limit at 50 Gb for output, MC samples with consitituenst exceed this limit

- Produced ntuples are stored in mine and Graziella’s eos workspace

Trigger Efficiency Study

- Trigger study is important to build our signal region where we assure the full trigger efficiency
- Recommendations already exist, but we want to understand if it’s possible to reduce pT threshold to help our model agnostic search
  - In principle, new physics could be hidden in jets below that threshold, so the more we can recover the better

Plots from ATL-COM-DAQ-2022-133



An alternative trigger efficiency selection

- **Historic DBL trigger strategy**: trigger item passed ⊕ leading jet p_T > 560 GeV ⊕ m_{JJ} > 1.3 TeV
- **Question: could it be lower to recover low-p_T jets?**

Trigger matching of reconstructed objects implemented, can be used at our advantage!

Tag&Probe method: tag the leading jet if it matches the trigger item and use the sub-leading p_T to probe if it has passed the trigger

$$efficiency\,(trig\ item) = \frac{(leading\ jet\ matching == 1)\ \&\&\ (subleading\ jet\ matching == 1)}{leading\ jet\ matching == 1}$$

- Reason: Trigger skimming applied on our DxAOD, for this reason trigger efficiency study can't be made using the leading jet (i.e. bootstrap technique) without exploiting an auxiliary trigger item due to the bias
  - Efficiency curves are fitted with a sigmoid function σ(x)
  - Used to extrapolate transverse momentum value at plateau

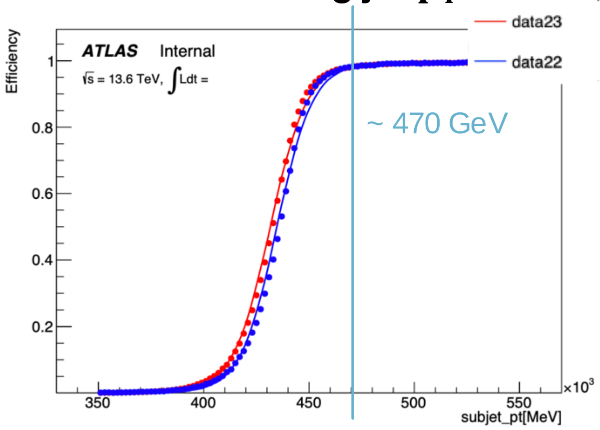
$$\sigma(x) = \frac{L}{1 + e^{-\frac{x-a}{b}}} \qquad \begin{cases} L = \text{plateau} \\ a = \text{centre of sigmoid} \\ b = \text{interval of rise} \end{cases}$$

It actually works!

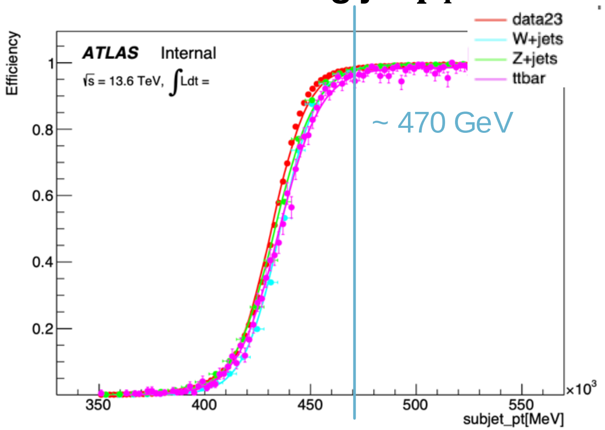
- Several trigger items studied
  - Best unprescaled trigger item: HLT_j460_a10_sd_cssk_pf_ftf_preselj225_L1J100

$$\sigma(x) = \frac{L}{1 + e^{-\frac{x-a}{b}}}$$

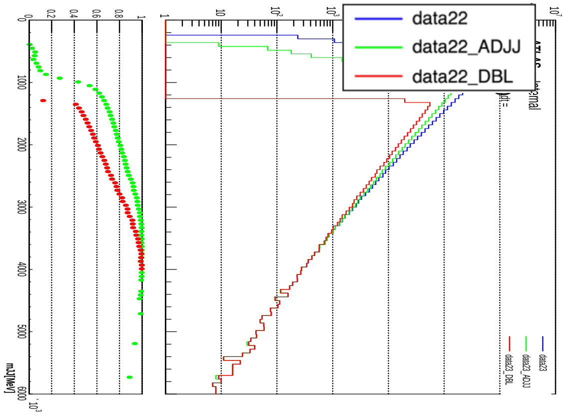
Sub-leading jet p_T



Sub-leading jet p_T



Invariant mass m_{JJ}

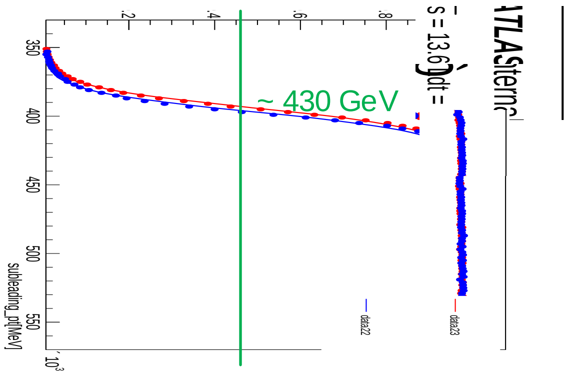


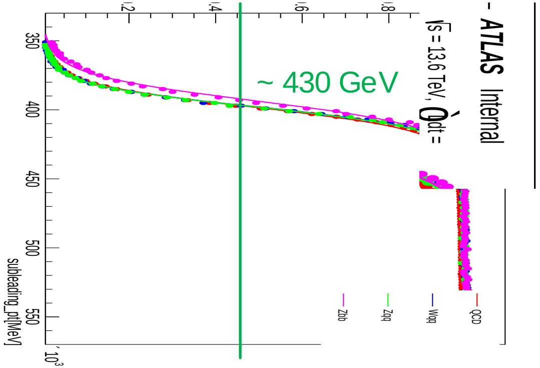
trigger item passed ⊕
leading jet p_T > 560 GeV ⊕
m_{JJ} > 1.3 TeV

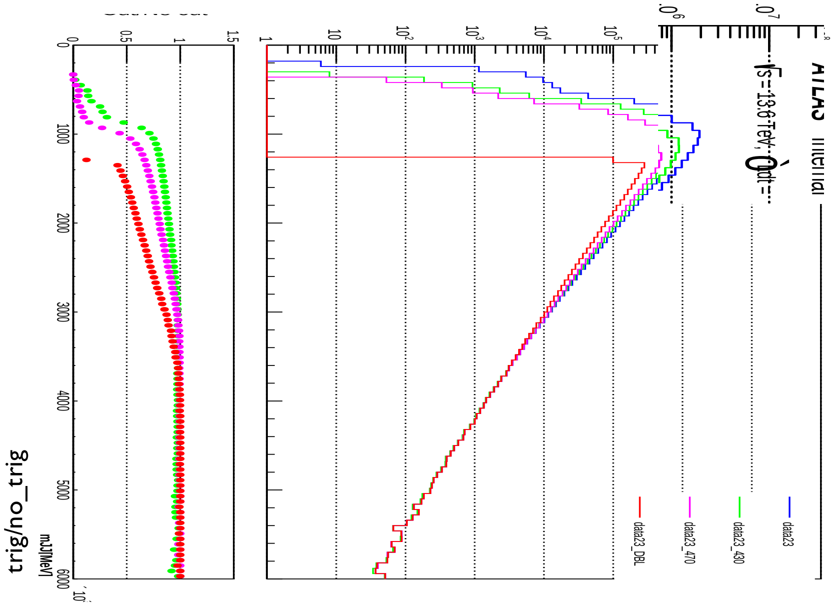
OPTIMIZING TRIGGER STRATEGY

- Current trigger selection: trigger item passed ⊕ trigger item matched with leading jet ⊕ leading jet p_T > 470 GeV
  - Trigger item: HLT_j460_a10_sd_cssk_pf_ftf_preselj225_L1J100
  - New checks on trigger items with mass cuts, 2 available

- Most promising one: HLT_j420_35smcINF_a10sd_cssk_pf_jes_ftf_preselj225_L1J100







data23_MBL:
trigger item passed ⊕
leading jet p_T > 560 GeV ⊕
m_{JJ} > 1.3 TeV

data23_470:
trigger item passed ⊕
leading jet matching ⊕
leading jet p_T > 470 GeV

data23_430 (new proposed):
trigger item passed ⊕
leading jet matching ⊕
leading jet p_T > 430 GeV

Preselections overview

Preselection = object definition + trigger selection

Object definition
\# jets > 2 ⊕
leading jet m > 50 GeV ⊕
subleading jet m > 50 GeV ⊕
jets pT > 200 GeV ⊕
-2 < jets η < 2 ⊕

Trigger selection 470
trigger item passed ⊕
trigger item matched with leading jet ⊕
leading jet p_T > 470 GeV

Trigger selection 430
trigger item passed ⊕
trigger item matched with leading jet ⊕
leading jet p_T > 470 GeV

- FastFrames ntuples produced with both trigger selections with constituents addition
  - 1M events max for QCD samples due to size

BACKUP

- Production of ntuples from our run 3 LLJ1 DxAOD based on EasyJet framework, large-R jets → Antikt10UFO jets.

- **News**:
  - Produced ntuple for data22, ~100k events.
  - Increased trigger list with new largeR-jet items, for both 2022 and 2023.
    - 2 items give problems with MC (#), can be turned off in their case.

trigger list 2022

```
'2022':
- 'HLT_j360_a10t_lcw_jes_L1J100'
- 'HLT_j420_a10sd_cssk_pf_jes_ftf_preselj225_L1J100'
- 'HLT_j460_a10sd_cssk_pf_jes_ftf_preselj225_L1J100'
- 'HLT_j460_a10t_lcw_jes_L1J100'
#- 'HLT_j460_a10r_L1J100'
#- 'HLT_j460_a10_lcw_subjes_L1J100'
- 'HLT_j420_35smcINF_a10sd_cssk_pf_jes_ftf_preselj225_L1J100'
- 'HLT_j420_35smcINF_a10t_lcw_jes_L1J100'
- 'HLT_mu24_ivarmedium_L1MU14FCH'
- 'HLT_mu50_L1MU14FCH'
- 'HLT_mu60_0eta105_msonly_L1MU14FCH'
- 'HLT_mu60_L1MU14FCH'
- 'HLT_mu80_msonly_3layersEC_L1MU14FCH'
```

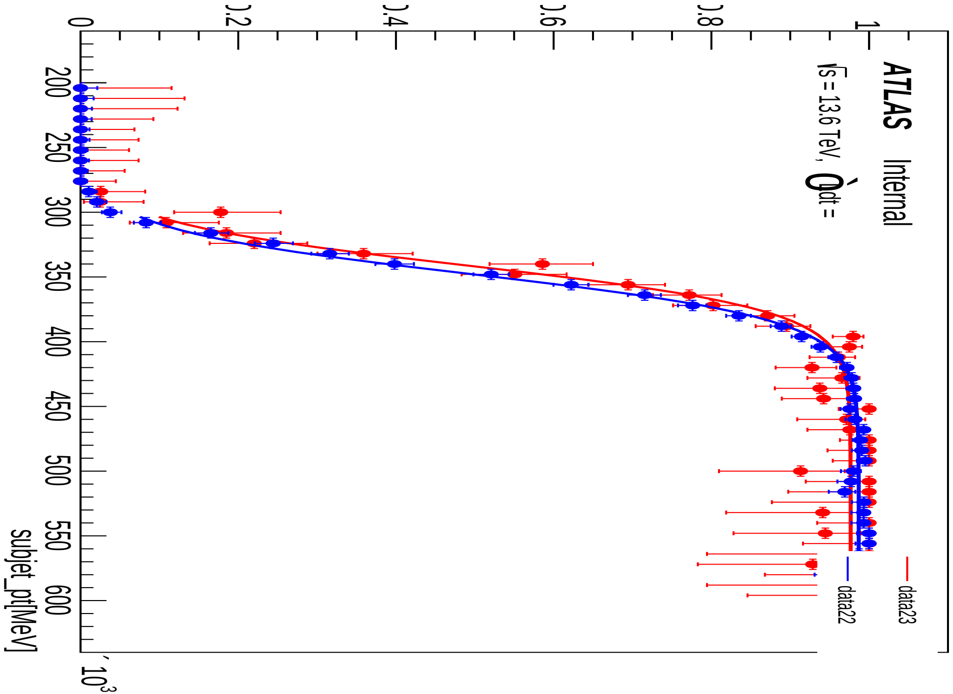
trigger list 2023

```
'2023':
- 'HLT_j360_a10t_lcw_jes_L1J100'
- 'HLT_j420_a10sd_cssk_pf_jes_ftf_preselj225_L1J100'
- 'HLT_j460_a10sd_cssk_pf_jes_ftf_preselj225_L1J100'
- 'HLT_j460_a10t_lcw_jes_L1J100'
#- 'HLT_j460_a10r_L1J100'
#- 'HLT_j460_a10_lcw_subjes_L1J100'
- 'HLT_j420_35smcINF_a10sd_cssk_pf_jes_ftf_preselj225_L1J100'
- 'HLT_j420_35smcINF_a10t_lcw_jes_L1J100'
- 'HLT_mu24_ivarmedium_L1MU14FCH'
- 'HLT_mu50_L1MU14FCH'
- 'HLT_mu60_0eta105_msonly_L1MU14FCH'
- 'HLT_mu60_L1MU14FCH'
- 'HLT_mu80_msonly_3layersEC_L1MU14FCH'
```

Unprescaled run 3 triggers twiki here

Trigger study on data

- Trigger item: HLT_j360_a10t_lcw_jes_L1J100
  - 2023 and 2022 data
- Tigger Item is <u>prescaled</u> for 2023 data



$$\text{fit function (sigmoid)} = \frac{p2}{1 + e^{-\frac{x - p1}{p0}}}$$

- **fit and extrapolation values**

data23

```
p0                         =      17241.7    +/-    1047.67
p1                         =       341034    +/-    1508.44
p2                         =     0.976619    +/-    0.00605004
Found cut off at 0.97 efficiency: 427025 MeV
```

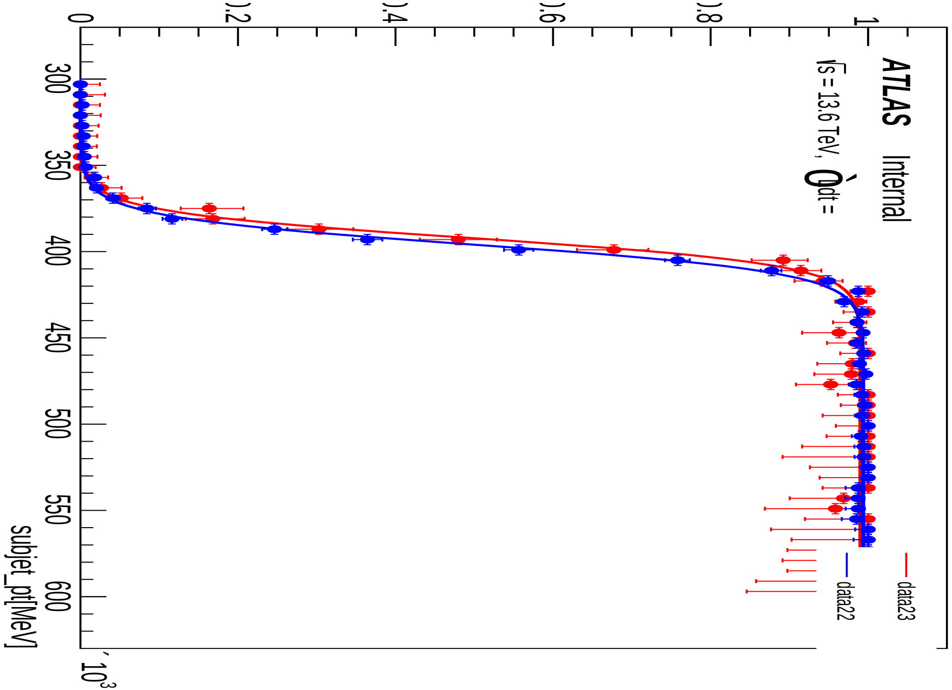
data22

```
p0                         =      17694.9    +/-    382.35
p1                         =       347636    +/-    580.226
p2                         =     0.986821    +/-    0.00191717
Found cut off at 0.98 efficiency: 435537 MeV
```

Trigger study on data

- Trigger item: HLT_j420_a10sd_cssk_pf_jes_ftf_preselj225_L1J100
  - 2023 and 2022 data
- This one is also <u>prescaled</u> for 2023, not for 2022

$$\text{fit function (sigmoid)} = \frac{p2}{1 + e^{-\frac{x - p1}{p0}}}$$



- **fit and extrapolation values**

data23

```
p0                                  =     7705.95    +/-    378.696
p1                                  =      392137    +/-    637.458
p2                                  =    0.990094    +/-    0.00280787
Found cut off at 0.99 efficiency: 463507 MeV
```

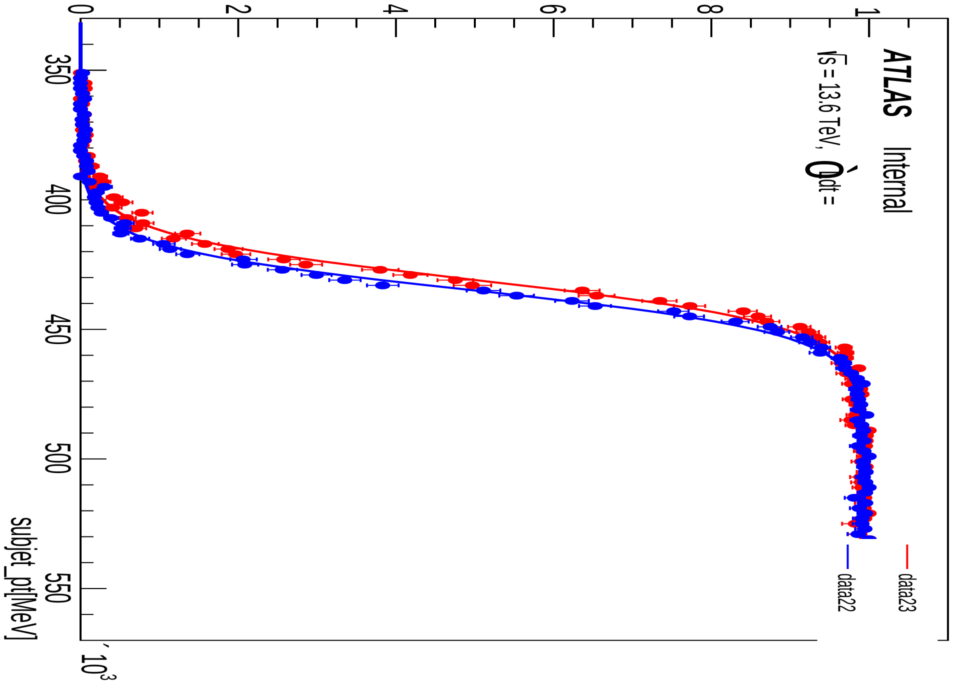
data22

```
p0                                  =     8120.13    +/-    159.395
p1                                  =      395867    +/-    265.546
p2                                  =    0.993706    +/-    0.000985083
Found cut off at 0.99 efficiency: 441240 MeV
```

Trigger study on data

- Trigger item: HLT_j460_a10sd_cssk_pf_jes_ftf_preselj225_L1J100
  - 2023 and 2022 data
- Good candidate for data 2023

$$\text{fit function (sigmoid)} = \frac{p2}{1 + e^{-\frac{x - p1}{p0}}}$$



- **fit and extrapolation values**

data23

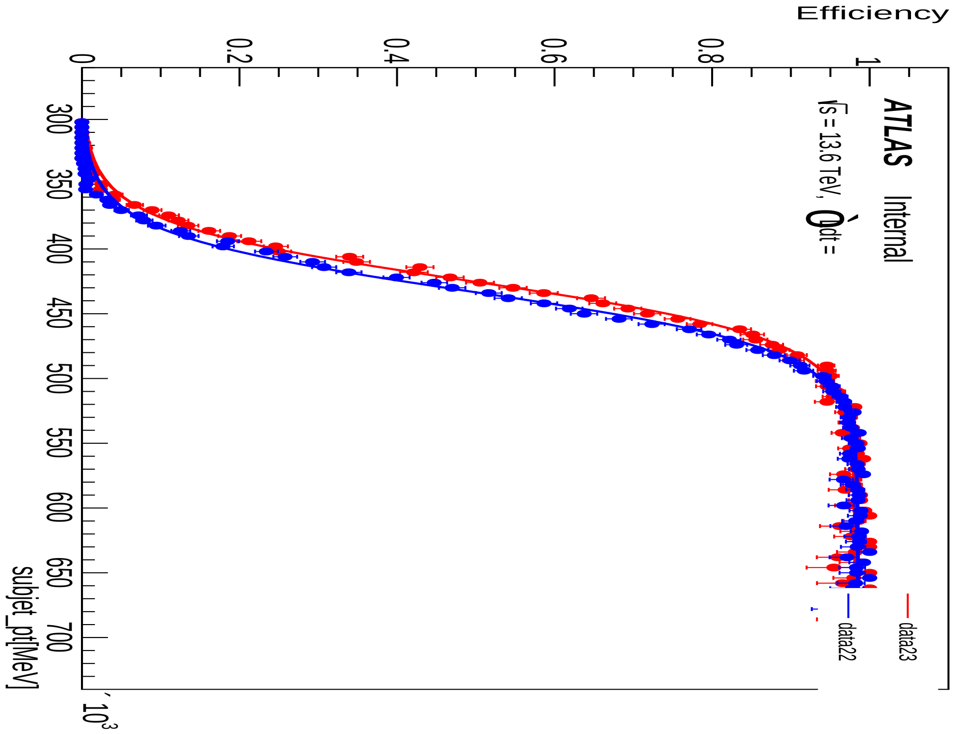
```
p0                        =     8737.24   +/-   119.947
p1                        =      430779   +/-   196.361
p2                        =    0.993001   +/-   0.000923239
Found cut off at 0.99 efficiency: 481444 MeV
```

data22

```
p0                        =     8290.88   +/-   105.462
p1                        =      434886   +/-   176.388
p2                        =    0.993693   +/-   0.000797063
Found cut off at 0.99 efficiency: 481243 MeV
```

Trigger study on data

- Trigger item: HLT_j460_a10r_L1J100
  - 2023 and 2022 data



$$\text{fit function (sigmoid)} = \frac{p2}{1 + e^{-\frac{x - p1}{p0}}}$$

- **fit and extrapolation values**

data23

```
p0                                  =      22423.4    +/-    267.65
p1                                  =       424553    +/-    391.524
p2                                  =     0.981706    +/-    0.00166781
Found cut off at 0.97 efficiency: 523602 MeV
```

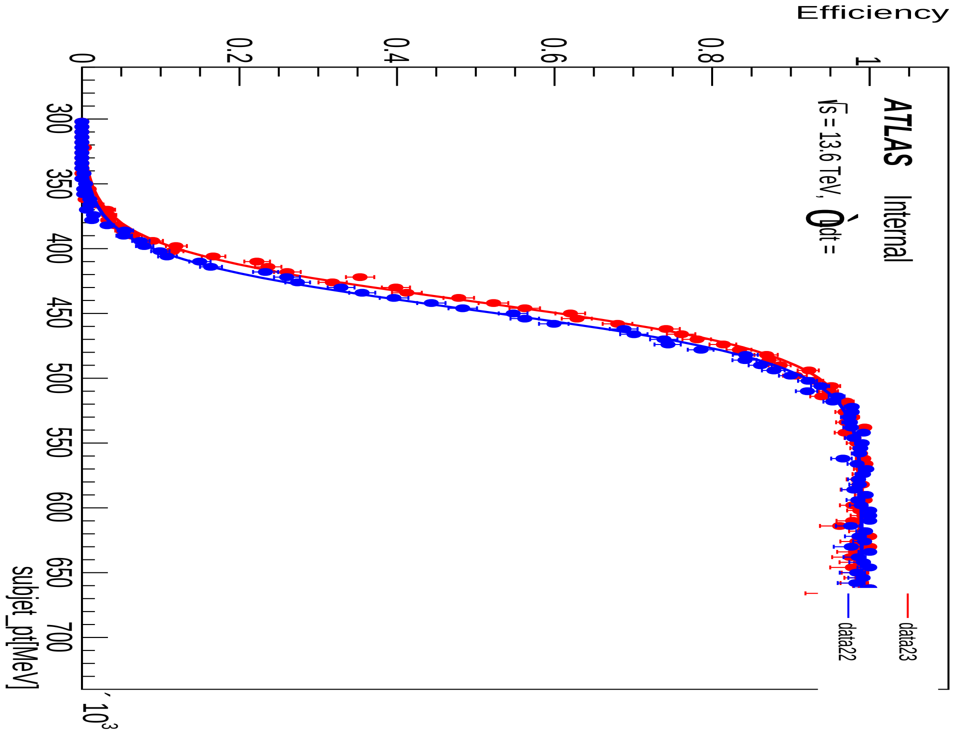
data22

```
p0                                  =      22383.8    +/-    245.315
p1                                  =       432760    +/-    358.861
p2                                  =     0.985058    +/-    0.00142138
Found cut off at 0.98 efficiency: 550647 MeV
```

Trigger study on data

- Trigger item: HLT_j460_a10t_lcw_jes_L1J100
  - 2023 and 2022 data

$$\text{fit function (sigmoid)} = \frac{p2}{1 + e^{-\frac{x - p1}{p0}}}$$



- **fit and extrapolation values**

data23

```
p0                              =       20584.8    +/-    259.742
p1                              =        440484    +/-    390.226
p2                              =      0.987447    +/-    0.00150964
Found cut off at 0.98 efficiency: 540932 MeV
```

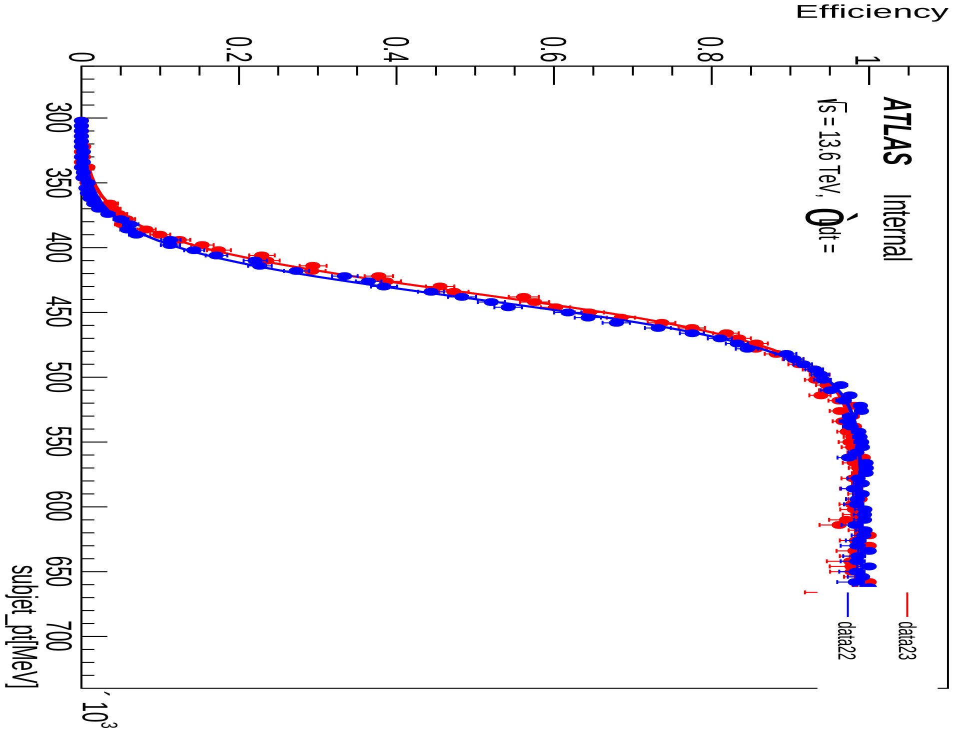
data22

```
p0                              =       20580.3    +/-    239.521
p1                              =        447298    +/-    363.356
p2                              =       0.99001    +/-    0.00127226
Found cut off at 0.98 efficiency: 541637 MeV
```

Trigger study on data

- Trigger item: HLT_j460_a10_lcw_subjes_L1J100
  - 2023 and 2022 data

$$\text{fit function (sigmoid)} = \frac{p2}{1 + e^{-\frac{x - p1}{p0}}}$$



- **fit and extrapolation values**

data23

```
p0                             =      20702.3    +/-    264.489
p1                             =       434846    +/-    390.602
p2                             =     0.983288    +/-    0.00165146
Found cut off at 0.98 efficiency: 552791 MeV
```

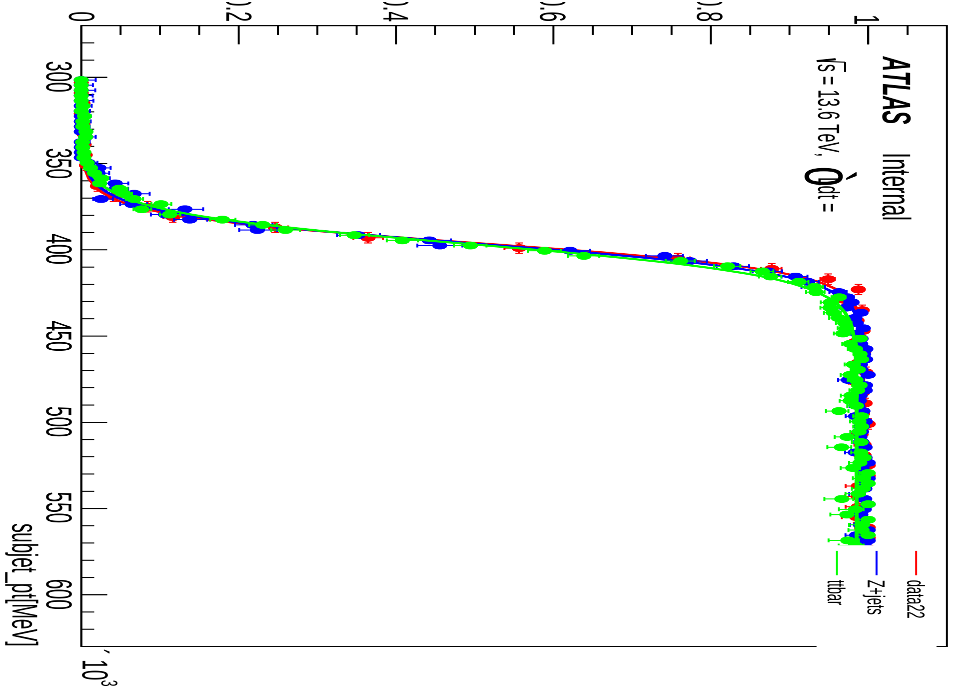
data22

```
p0                             =      20236.8    +/-    226.398
p1                             =       439348    +/-    343.221
p2                             =     0.990331    +/-    0.00115468
Found cut off at 0.99 efficiency: 601296
```

Trigger study on MC bkg

- Trigger item: HLT_j420_a10sd_cssk_pf_jes_ftf_preselj225_L1J100
  - Z + jet and ttbar fully hadronic
- Nice closure between curves and with data 2022

$$\text{fit function (sigmoid)} = \frac{p2}{1 + e^{-\frac{x - p1}{p0}}}$$



- **fit and extrapolation values**

Data 2022

p0	=	8120.13	+/−	159.395
p1	=	395867	+/−	265.546
p2	=	0.993706	+/−	0.000985083

Found cut off at 0.99 efficiency: 441240 MeV

Z + jet

p0	=	8756.98	+/−	170.396
p1	=	396203	+/−	281.215
p2	=	0.993021	+/−	0.000715131

Found cut off at 0.99 efficiency: 446924 MeV

ttbar

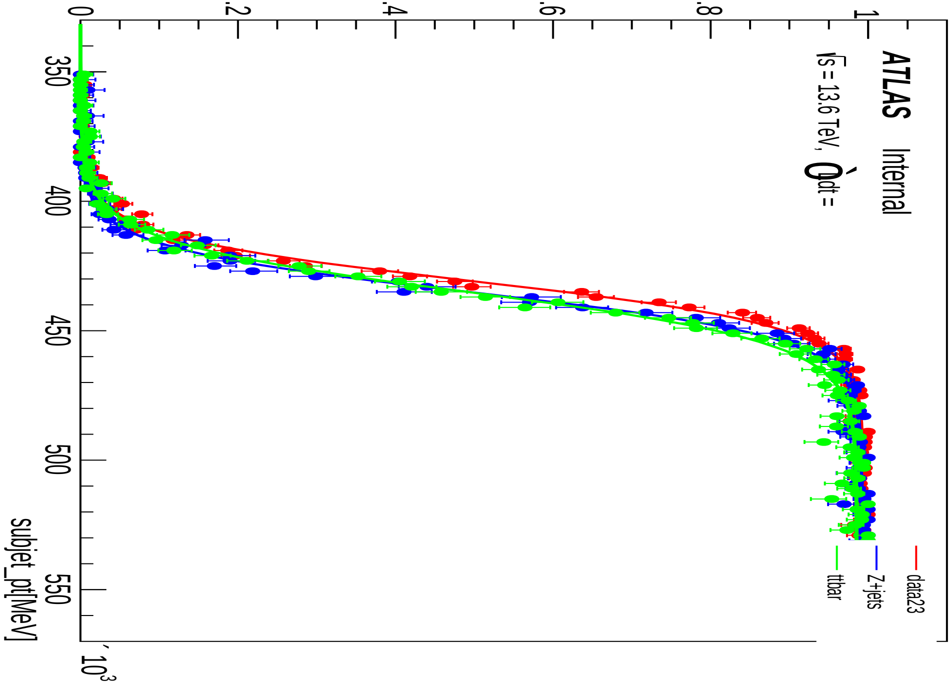
p0	=	9527.8	+/−	137.616
p1	=	396632	+/−	226.778
p2	=	0.985301	+/−	0.0011313

Found cut off at 0.98 efficiency: 446364 MeV

Trigger study on MC bkg

- Trigger item: HLT_j460_a10sd_cssk_pf_jes_ftf_preselj225_L1J100
  - Z + jet and ttbar fully hadronic
- Nice closure between curves and with data 2023

$$\text{fit function (sigmoid)} = \frac{p2}{1 + e^{-\frac{x-p1}{p0}}}$$



- **fit and extrapolation values**

Data 2023

```
p0                        =      8737.24    +/-    119.947
p1                        =       430779    +/-    196.361
p2                        =     0.993001    +/-    0.000923239
Found cut off at 0.99 efficiency: 481444 MeV
```

Z + jet

```
p0                        =      8704.49    +/-    184.957
p1                        =       435131    +/-    294.569
p2                        =     0.991279    +/-    0.00122803
Found cut off at 0.99 efficiency: 493027 MeV
```

ttbar

```
p0                        =      9872.54    +/-    175.105
p1                        =       435088    +/-    290.259
p2                        =     0.985346    +/-    0.00170914
Found cut off at 0.98 efficiency: 486535 MeV
```

ADJJ vs DBL

- **DBL**: trigger item passed ⊕ leading jet p_T > 560 GeV ⊕ m_{JJ} > 1.3 TeV
- **ADJJ**: trigger item passed ⊕ trigger item matched with leading jet ⊕ leading jet p_T > 470 GeV