Analysis of Four-Lepton Event “Non-Prompt” Lepton Backgrounds

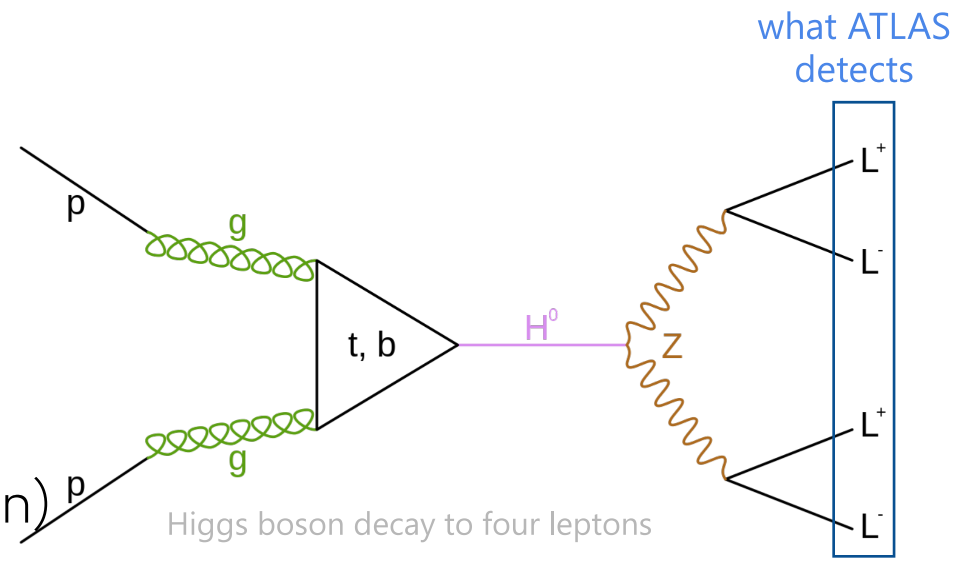
Iria Wang, Brandeis University

Introduction

Four-Lepton (4ℓ) event signature: two same-flavor, opposite-charge e or μ pairs

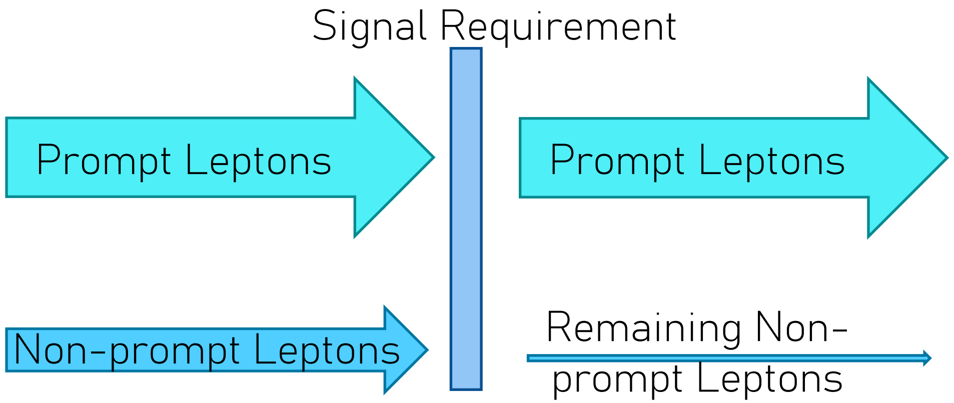
4ℓ ideal for Standard Model (SM) studies

- contribution from SM processes (Z boson, Higgs boson)
- sensitive to new physics beyond the SM



Prompt leptons – produced directly from hard scatter interaction

Non-prompt leptons – secondary decays of hadrons or mis-reconstruction



→ signal requirement imperfect

4ℓ dataset contaminated by 3 prompt + 1 non-prompt

Signal requirement – detect and reject non-prompts

Data from Run 2 of the Large Hadron Collider

Data Driven Fake Factor Analysis

Define: **Fake Efficiency (f)** = P(pass signal requirement | non-prompt)

Dilepton events – opposite-charge pairs, additional leptons are likely non-prompt

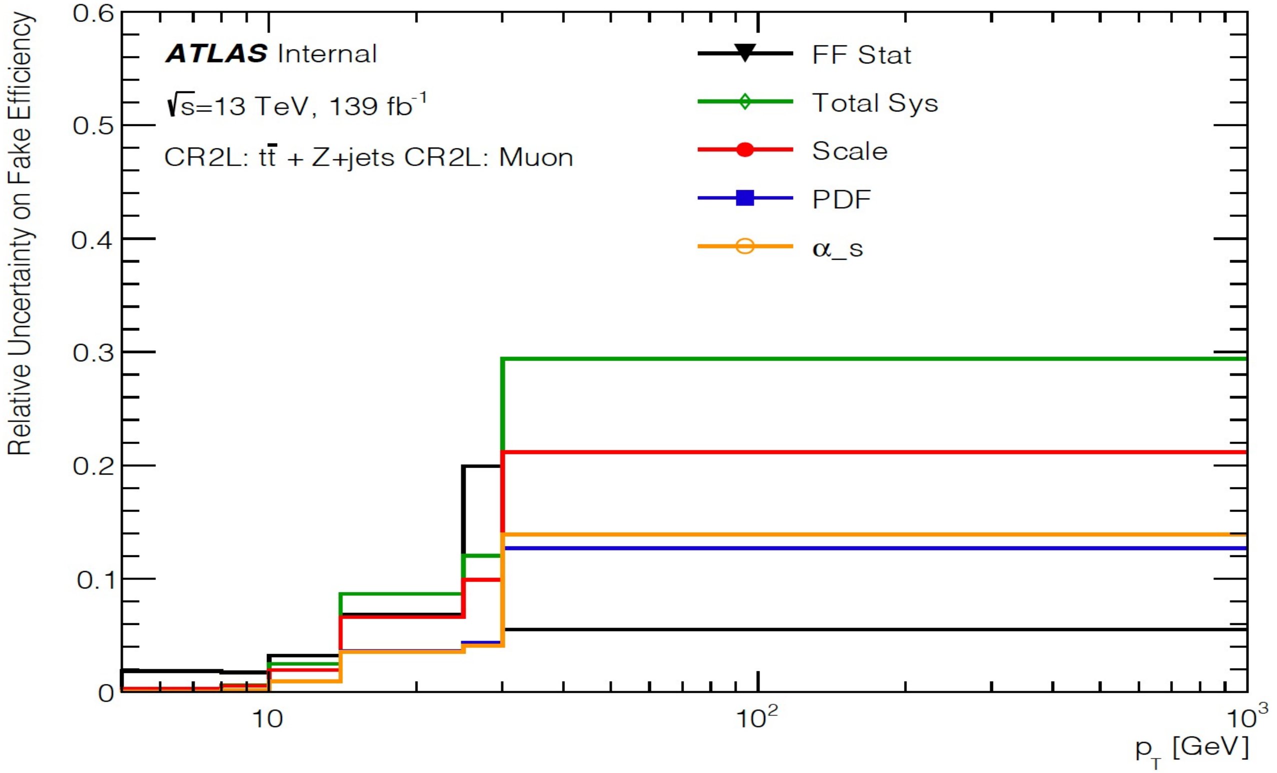
For dilepton events:

MC = Monte Carlo

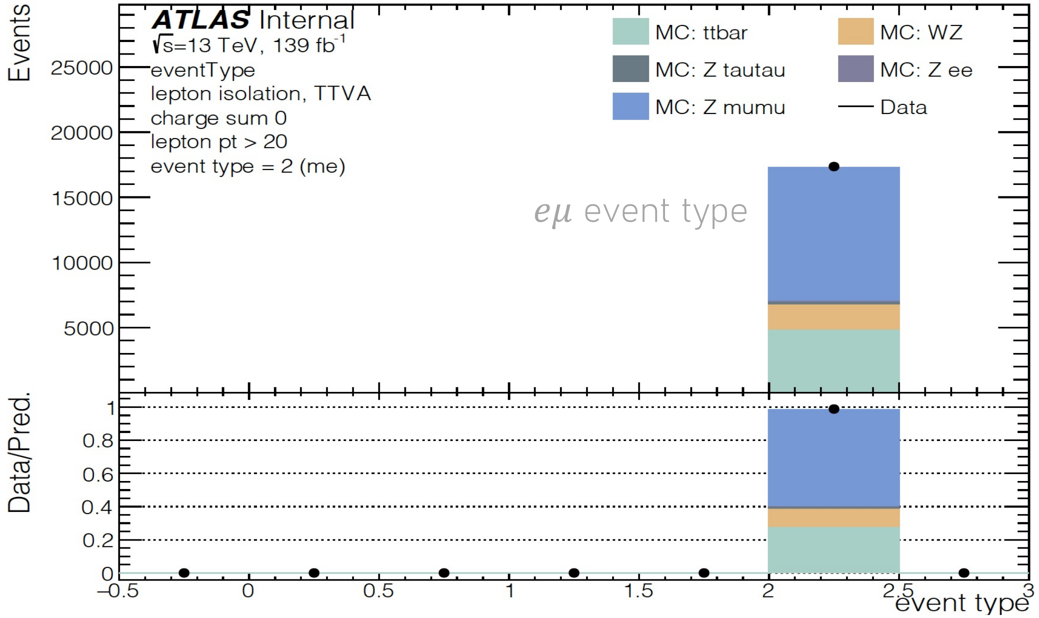
$$\textit{Fake Efficiency}_{Data} = \frac{\textit{Baseline Leptons Passing Signal Requirement} - \textit{Classified Prompt Leptons Passing Signal Requirement from MC}}{\textit{All Baseline Leptons} - \textit{Classified Prompt Leptons from MC}}$$

Baseline leptons = additional leptons from data
Classified leptons = additional leptons from MC

Project Goal: Study non-prompts to precisely identify f and reduce uncertainty (systematic errors from MC)



Preliminary Analysis

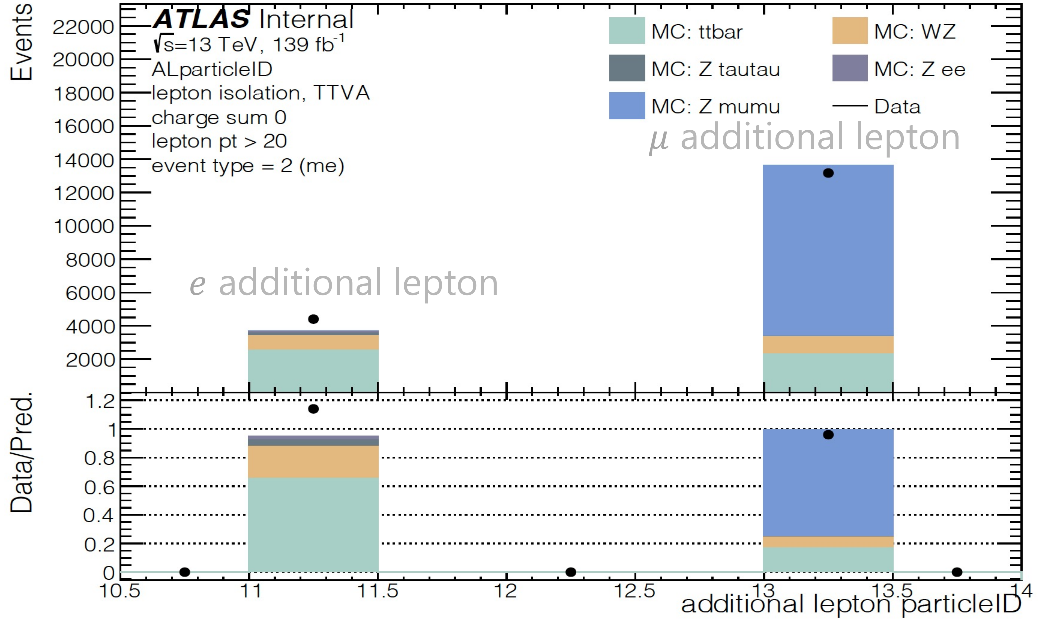


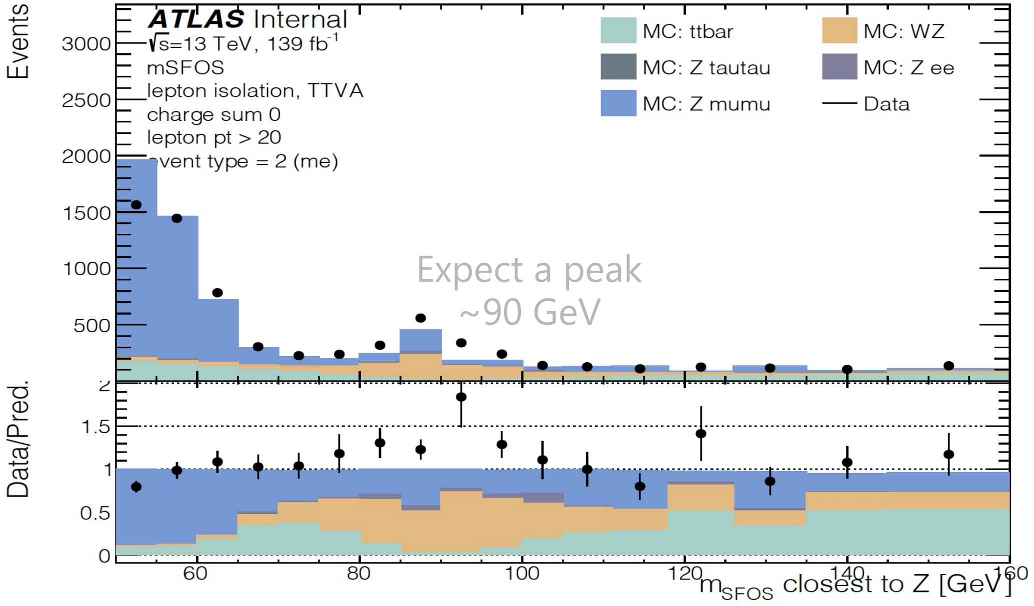
Select Z boson decay and top quark pair (ttbar) from dilepton via comparison to MC

- $e\mu$ for ttbar, ee or $\mu\mu$ for Z boson decay

Unexpected contamination of Z boson decay in $e\mu$ ttbar region

- Z → $\mu\mu$ + additional non-prompt lepton e: reconstructed with μ as additional





Suppression of Prompt Sources

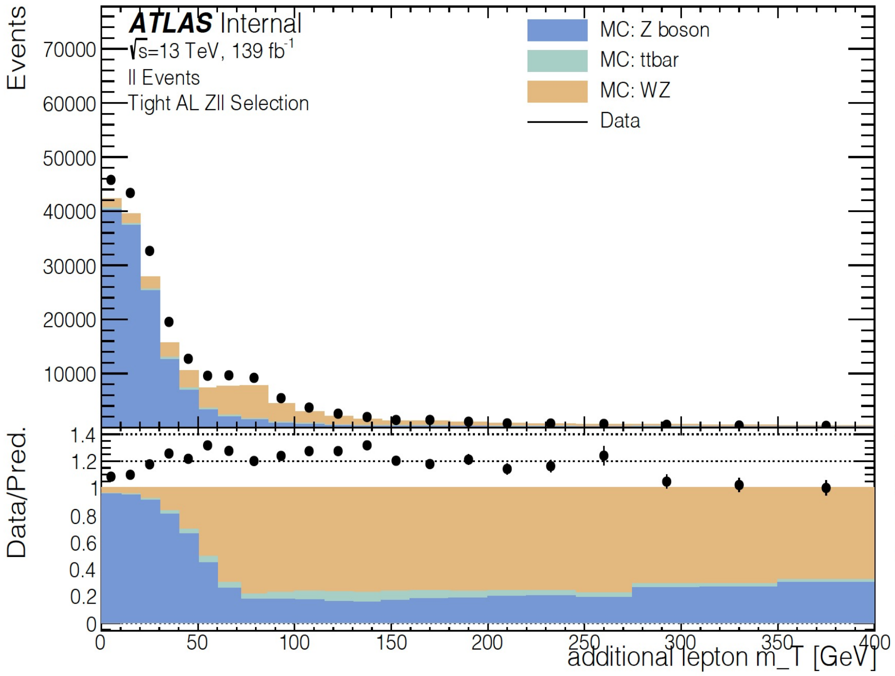
Prompt sources falsify measurement i.e., mis-ID Z boson pair, W/Z boson production

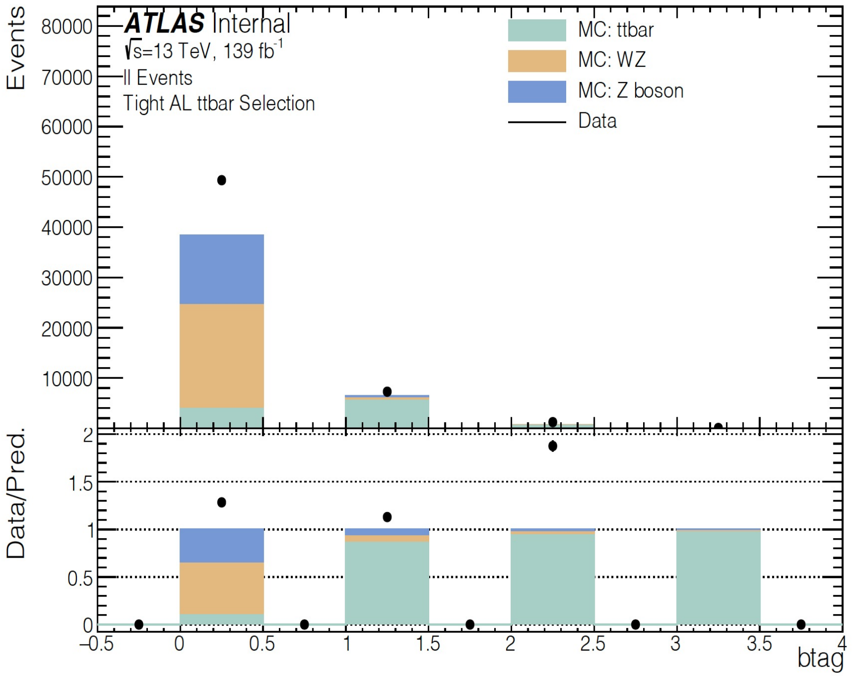
1. Create two subtrees, signal ID

   Z boson tree: attempt to pair ee or $\mu\mu$

   ttbar tree: attempt to pair $e\mu$

2. Selection cuts on kinematic variables to suppress prompt sources



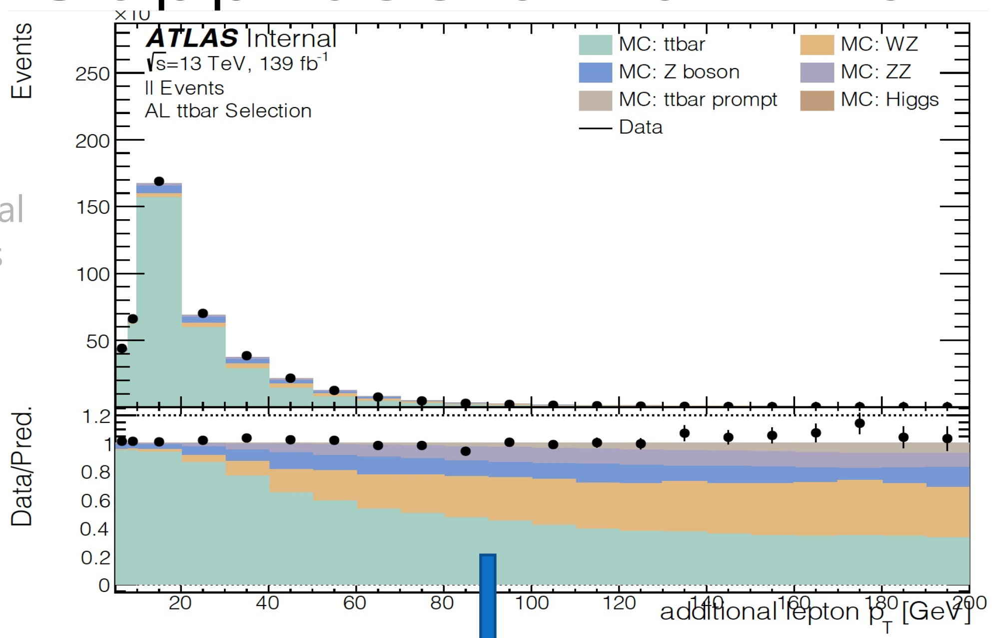


Example of selection cuts

- Z boson: W/Z peak after m_T 50 GeV
- ttbar: Prompts clustered at btag = 0

Suppression of Prompt Sources

ttbar additional leptons

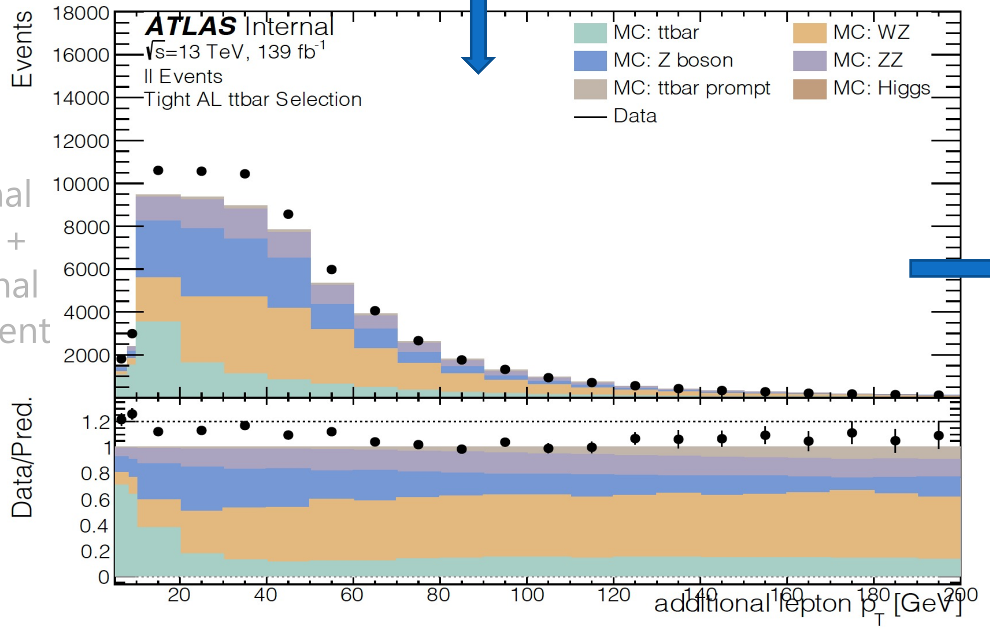


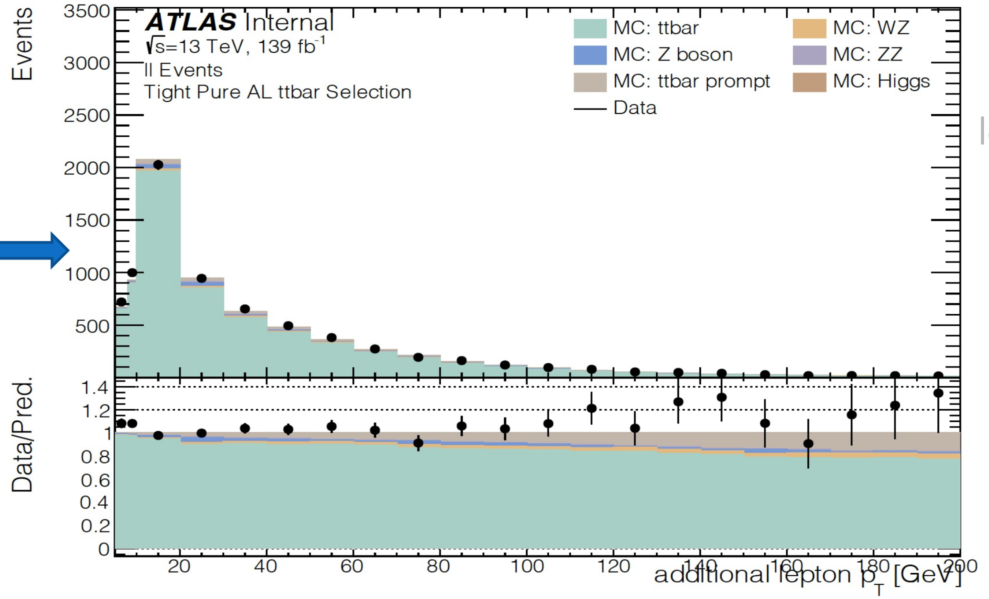
- Data-MC disagreement: Other potentially contributing processes
- Balance systematic vs statistical uncertainty

Selection cuts:

- btag > 0
- 80 GeV < m_{SFOS} < 100 GeV

ttbar additional leptons + pass signal requirement

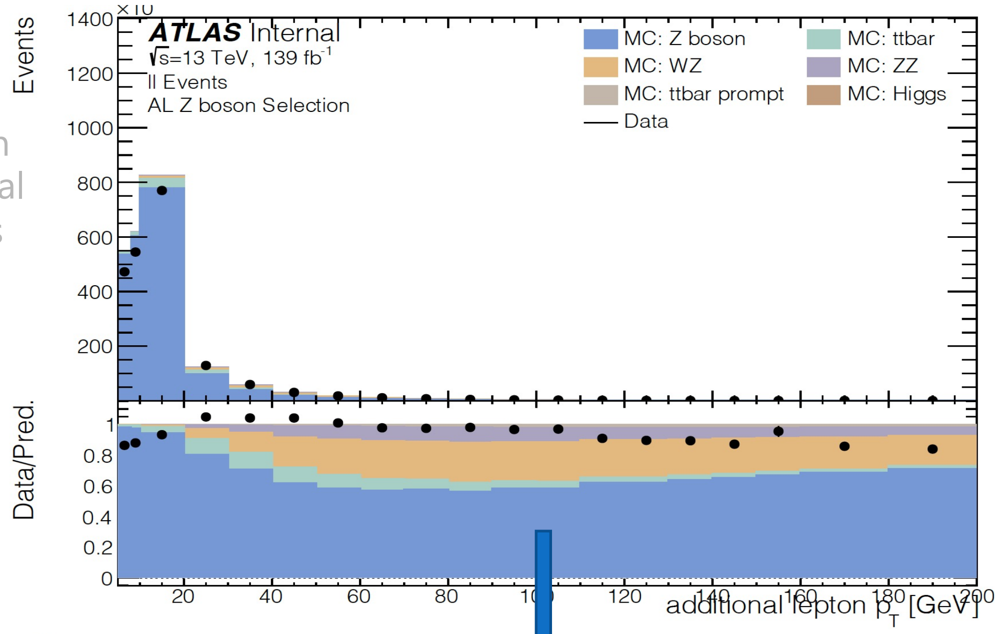




ttbar additional leptons + pass signal requirement + prompt lepton suppression

Suppression of Prompt Sources

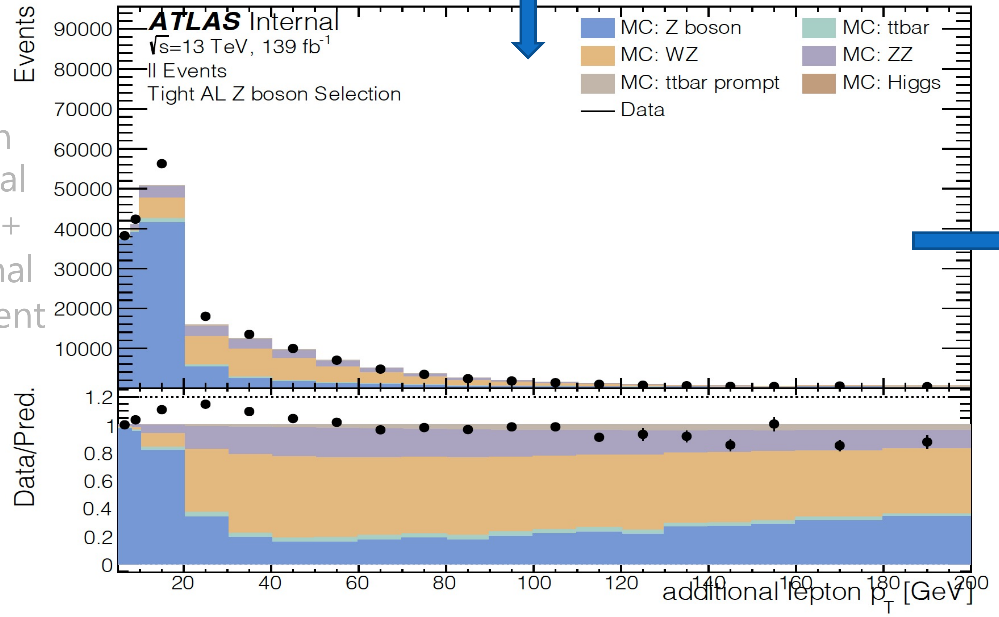
Z boson additional leptons

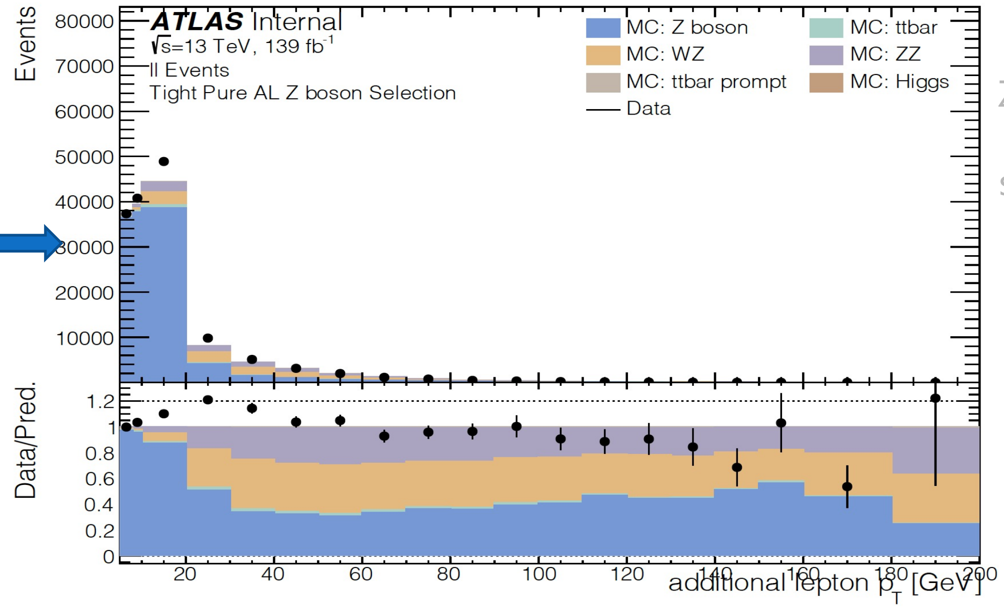


Selection cuts:

- $dR > 0.8$;
- pair lepton p_T < 200 GeV
- 2nd lepton p_T < 70 GeV
- Additional lepton m_T < 50 GeV

Z boson additional leptons + pass signal requirement

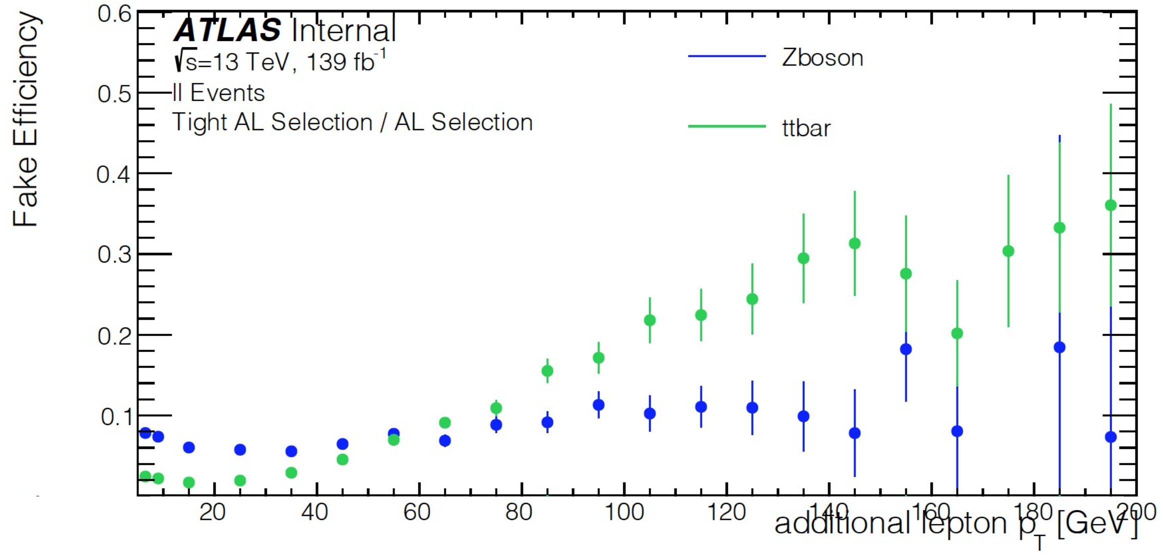


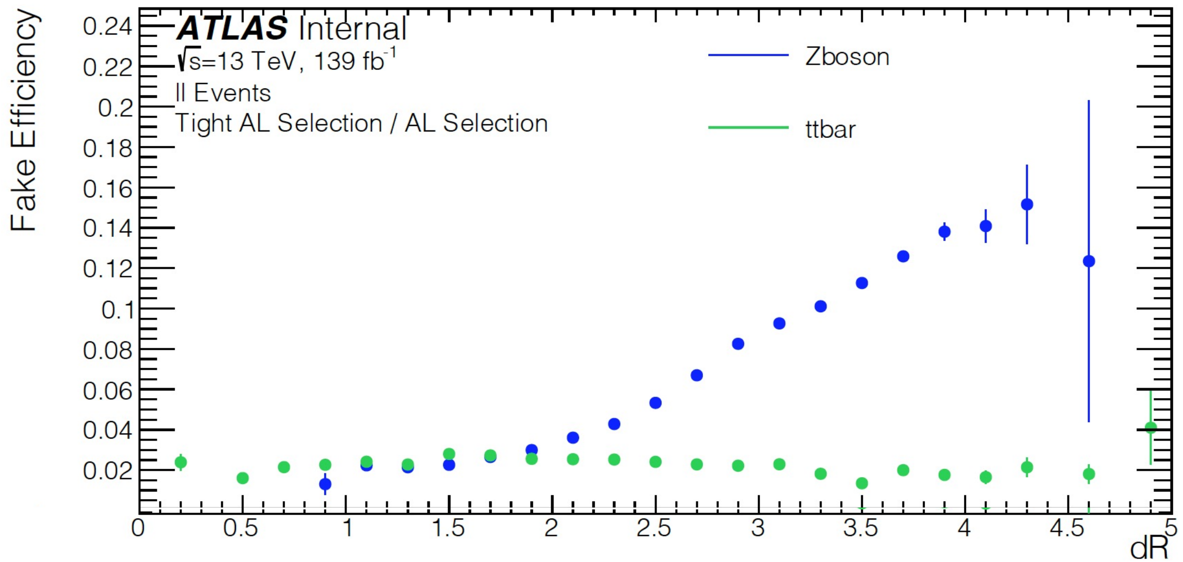


Z boson additional leptons + pass signal requirement + prompt lepton suppression

Rejection Power

Evaluate **rejection power** for analysis cuts → Plot fake efficiency for both regions





- Dependence of on additional some lepton kinematics, e.g., p_T
- Difference between regions is a measure of how strongly fake efficiency depends on composition
- Dependencies can also suggest contamination

Conclusions

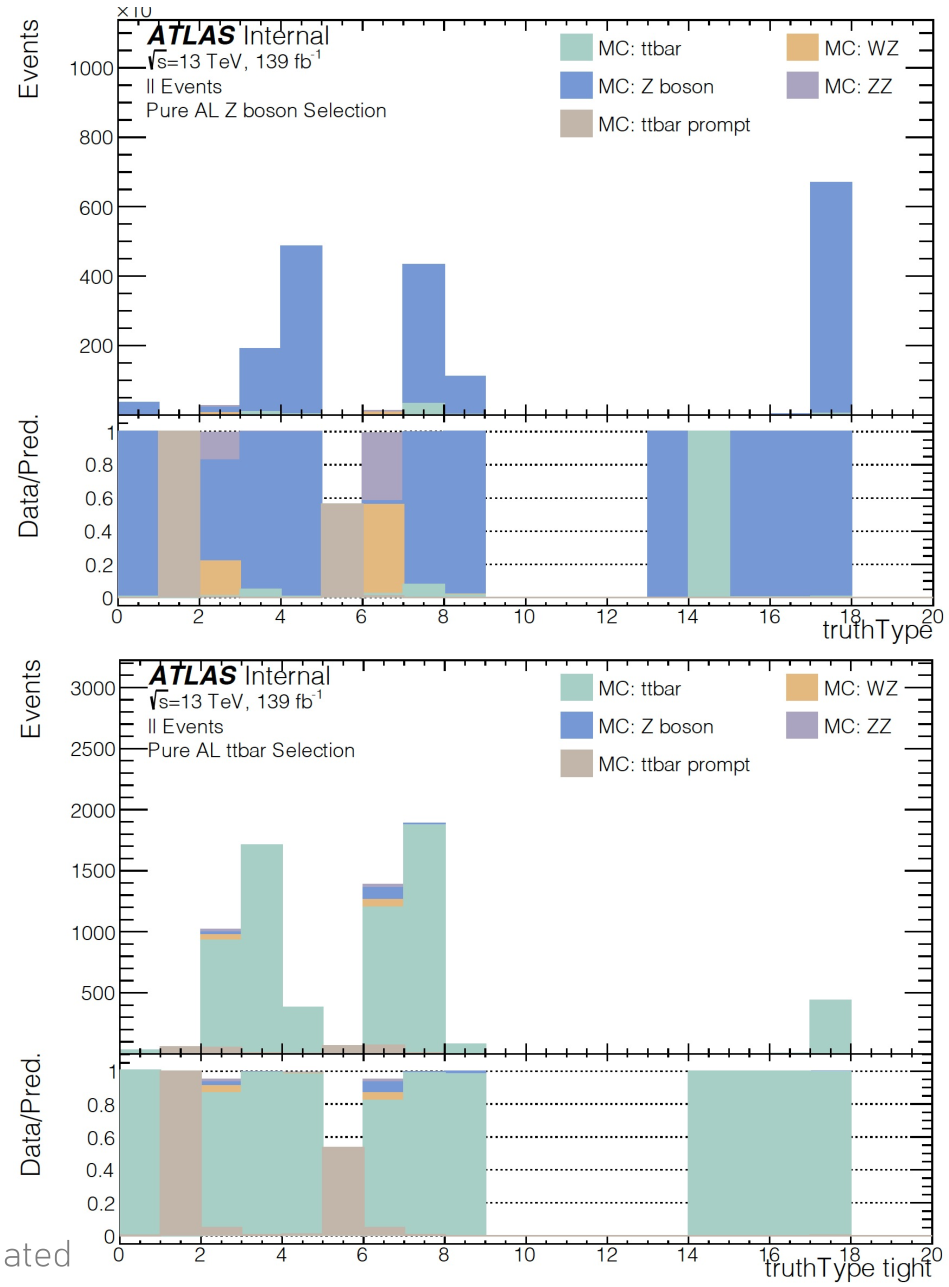
Observed and eliminated prompt sources from Z boson mis-IDs, WZ events, and other smaller contributions

- Z boson events in $e\mu$ region
- Data-MC disagreement
- Statistical vs systematic error
- Evaluation of rejection power
- Application to other samples

Next steps:

- Estimate "true rejection power" with generator information
- Compare to measured rejection power to validate approach

Generator information



Each integer is associated with an event type

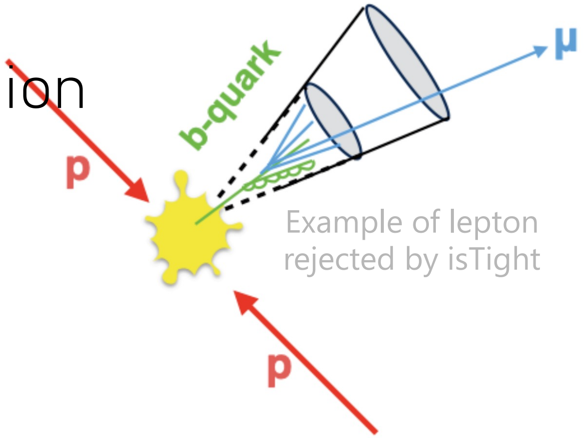
Thank you!

A huge thank you to Dr. Gabriella Sciolla for the research opportunities and her mentorship, Dr. Max Goblirsch for his guidance every step of the way, and Prajita Bhattarai for setting up the files and answering my questions. Thank you to the US ATLAS SUPER program for funding and hosting the symposium!

BACKUP SLIDES BELOW!

“Non-prompt” Backgrounds

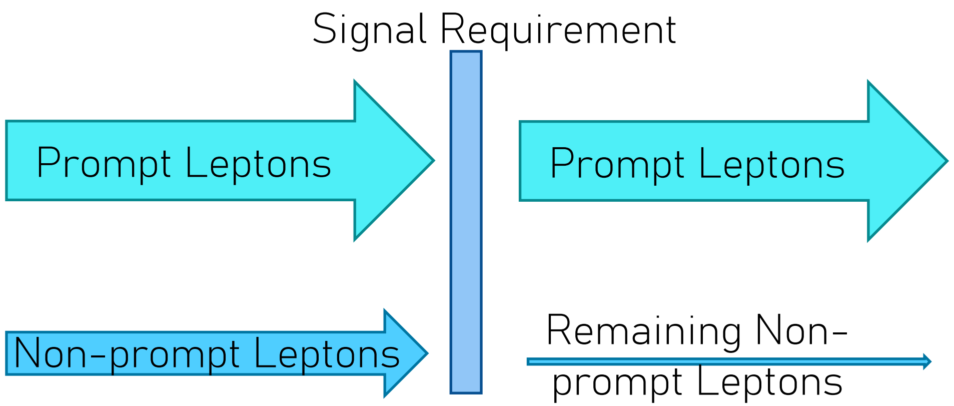
- **Prompt leptons** – produced directly in a hard scatter interaction
- **Non-prompt leptons** – secondary decays of hadrons or mis-reconstruction



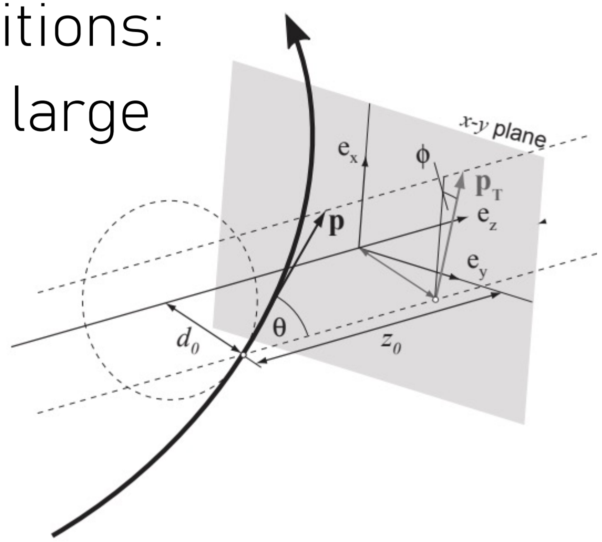
4ℓ dataset contaminated by 3 prompt + 1 non-prompt leptons

Signal requirement – detect and reject non-prompts, two conditions:

1. Isolation – reject if $isolation = sum(p_T, tracks\_in\_cone)$ is large
2. Impact parameters – reject if $|z_0|$ or $|d_0|$ non-negligible



→ signal requirement imperfect



Global track parameters e.g. wrt. perigee

Notes

Slide 1:

Require precise understanding of particle-detector interactions

- small likelihood of producing particle of interest + rapid particle decay -> ever-growing dataset, increasing statistical precision

Four-lepton events – interactions resulting in two same-flavor, opposite-charge lepton pairs

Ideal candidates for high-precision studies of the standard model

- (1) Contribution from interesting SM processes, including Z boson decay, Higgs boson production, on-shell ZZ production
- (2) Possible contribution from new physics beyond the SM, such as Supersymmetry
- (3) Allow clean separation between physics processes of interest and otherwise overwhelming backgrounds
- Therefore, these processes serve as probes of the SM and new physics

For this study, use data from Run 2 of the LHC

Ex. Figure shows Higgs boson production, bosons decay rapidly and resulting leptons are detected by ATLAS

Slide 2:

Such leptons are called "prompt", produced by some process of interest

- Leptons can also be produced by secondary decays of hadrons or as artifacts of mis-reconstructions — these are called "non-prompt"

While we have good separation from backgrounds, dataset can still be contaminated by events with three or fewer prompt leptons and one or more non-prompt

Idea: suppress non-prompt sources using a signal requirement

Signal requirement is based on surrounding activity and trajectory

- (1) Isolation: we expect leptons from other decays tend to come in jets
- (2) Impact parameters: closest approach of lepton track to beamline does not trace back to interaction point

Right figures show lepton that fails signal requirement

Bottom figure shows role of signal requirement: lets almost all prompt through while eliminating non-prompt

- Not 100% efficient

Slide 3:

Use data driven fake factor analysis to study remaining non-prompts surviving the signal requirement

Fake eff.: probability that a lepton passes the signal requirement given that it is non-prompt

- this is one of the main results we are trying to determine

Dilepton events: opposite charge pairs

- Idea: study non-prompts from the dilepton dataset where the third, additional lepton is most likely in the form of a non-prompt lepton

For these events define fake eff as non-prompt additional leptons passing signal req. over all non-prompt additional leptons

- In this study frequently use Monte Carlo simulation to predict data resulting from certain theoretical processes
- denominator: additional leptons from data minus prompt additional leptons
- numerator: same but passing signal req

This calculation has uncertainties (figure), most prominent one is systematic error from theory in monte carlo

Project goal: study non-prompt leptons using real collision data to improve data driven fake factor method

Slide 4:

Initial selection of two regions to study: Z boson decay and top quark pair decay (ttbar)

- Done via comparison to MC to estimate data composition
- Primarily use event type: ttbar has electron and muon, Z boson has electron electron or muon muon

Unexpected finding: substantial contamination of Z boson decay in e mu ttbar region

- To study this, break Z boson decay into individual processes
- event type = 2: e mu, large ratio of Z mumu
- additional lepton is a muon (particle ID = 13 is muon)
  - e and mu being paired as prompt leptons, muon flagged as additional
- mSFOS: invariant mass for same-flavor opposite-sign pairs
  - many electrons are actually photons mis-identified as an electron

Notes

Slide 5:

prompt sources such as these falsify the measurement

- eliminate mis-reconstructed Z boson events and W/Z boson production, which produces 3 genuine leptons

Resolve the issue from the previous slide: add signal ID restriction to ttbar region (checks for electron signatures)

Also separate into two trees, which helps with runtime and ttbar region

- Z boson tree: events with additional leptons, where prompt leptons are paired as ee or mumu

Experiment with selection cuts on kinematic variables to suppress prompt sources, examples:

- Z boson: W/Z events clustered after transverse mass of additional lepton > 50 GeV
- ttbar: b-tagging flags the decay of b quarks resulting from this interaction, so eliminating events with btag = 0 is very effective

Slide 6:

Encounter two challenges:

- MC-data disagreement — add some processes potentially contributing to this dataset, such as Higgs production
- Statistical vs systematic uncertainty: try to cut away prompt leptons to reduce systematic uncertainty, but cut away too much data and statistical uncertainty will grow
  - require a couple iterations of the study to balance this

Figures show suppression of prompt sources from ttbar region

- (1) initial set of additional leptons
- (2) application of signal requirement — non-prompt leptons from ttbar events suppressed, some remaining
- (3) add suppression of prompt sources, fairly effective

Slide 7:

same thing for Z boson decay region, initial selection -> pass signal req -> suppress prompt sources

use cuts on dR, transverse momentums, transverse mass

Slide 8:

need to evaluate rejection power for our analysis cuts, which is ability for signal criteria to suppress non-prompts — estimate fake efficiency over kinematic variables

- there is dependence on some additional lepton kinematics
  - eg. higher p_T leptons are more likely to pass signal requirements
- Z boson and ttbar regions prompt different compositions of prompt leptons
  - expect samples to demonstrate slightly different fake eff
  - difference is a measure of how strongly fake eff depends on composition
- Dependencies on unrelated variables, such as dR, can also suggest further contamination
  - require further studying using generator information from Monte Carlo

Slide 9:

Analyze non-prompt background sources by studying samples of non-prompt additional leptons

Eliminate prompt sources from our Z boson and ttbar selections

Challenges/obstacles observed

- substantial Z boson event contamination in e mu region
- Data-MC disagreement — suggests there is some processes in the data we are not simulating
- balancing stat vs sys errors
- evaluate and analyze rejection power

Next steps:

- Use generator information to estimate "true rejection power" from Monte Carlo
- compare this to measured rejection power to confirm validity of the approach

Notes

Role of simulation

- First thing: Simulation is not per se guaranteed to be representative of data
- The final measurement of f has to be done in data
- However, Simulation allows to do tests not possible in data: We can match reconstructed particles to the simulated particles, and thus know what is prompt and what is non-prompt
  - We can get the "true" value of f for the simulation by checking how many non-prompt leptons are "signal"
  - Then comparing this to our measurement allows to test our method
- Simulation is the main tool in ATLAS to obtain (standard model) predictions for our various measurements. We start by simulating the physics of the particle interactions, which gives us a list of particles with associated momenta / directions. These are then plugged into a detector simulation (our previous project!), which emulates how ATLAS will respond to the particles. The predicted detector response can be plugged into normal event reconstruction just like real collision data (same format). The output is then the predicted outcome of our measurements, with the bonus that we know what process really occurred and which particle caused which signal in the detector (which we can not know for sure in real data)

Transverse mass:

W decay: to muon/electron + neutrino

- For ATLAS, neutrinos are invisible
- We use the fact that in the x-y plane, the vector sum of all momenta should be 0
- This means: Vector sum of all invisible particles (in x-y plane) = - vector sum of all visible particles (in x-y plane)
  - = Missing transverse momentum
- If we assume a single neutrino is carrying away all the missing momentum, we can define a **transverse mass** for a given lepton in the event
  - MT = |(pt_Miss_x, pt_Miss_y,0,0) + (pt_lepton_x, pt_lepton_y,0,0)|^2 < full mass of the decaying particle

Particle momenta

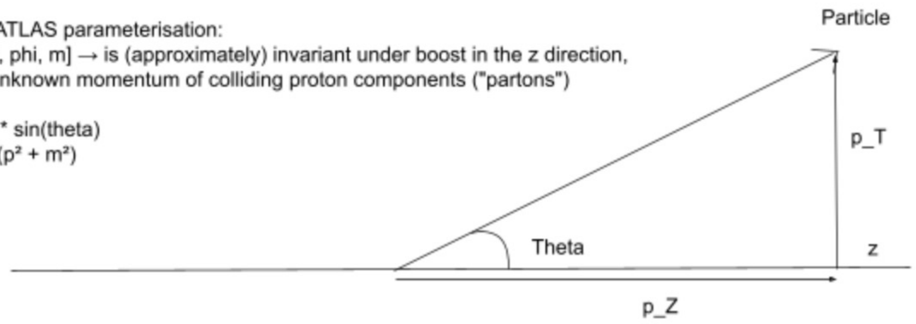
all formulas: natural units hbar = c = 1

Eta = - ln (tan (theta/2))

Eta of 0 corresponds to theta = 90 degreses ("outward" perpendicular to z)
Eta of +infinity corresponds to theta = 0 (along z)
Eta of -infinity corresponds to theta = 180 degrees (along negative z)

Typical ATLAS parameterisation:
[pT, eta, phi, m] → is (approximately) invariant under boost in the z direction, due to unknown momentum of colliding proton components ("partons")

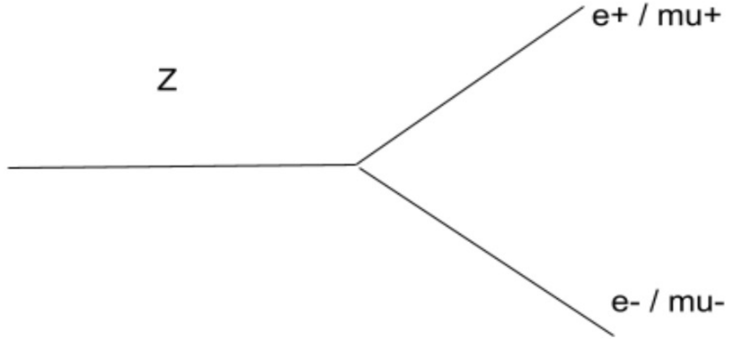
pT = |p| * sin(theta)
E = sqrt(p^2 + m^2)



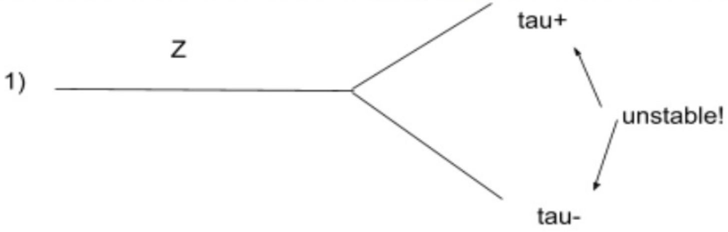
Notes

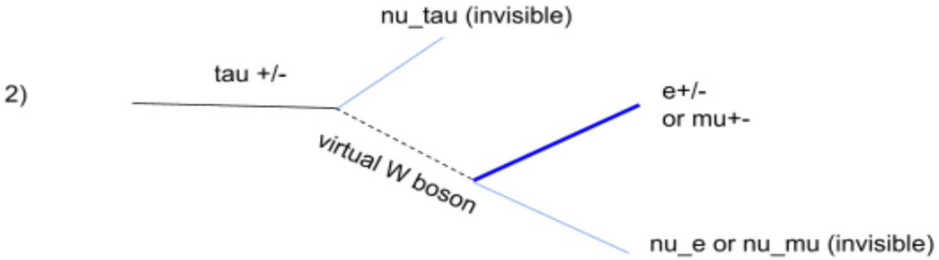
Z boson decays into leptons

- 2 "trivial" ones: mu+/mu- or e+/e-:

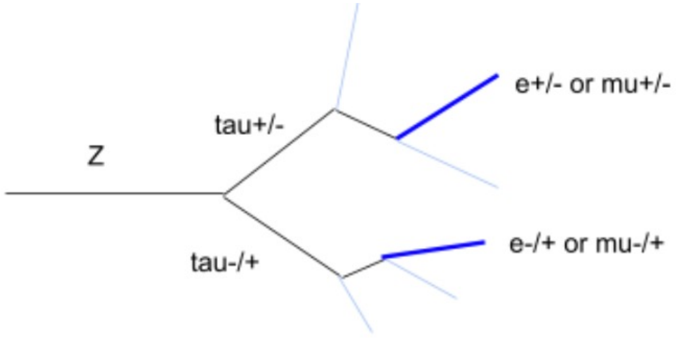


- Slightly more complicated: Z-> tau+ tau-





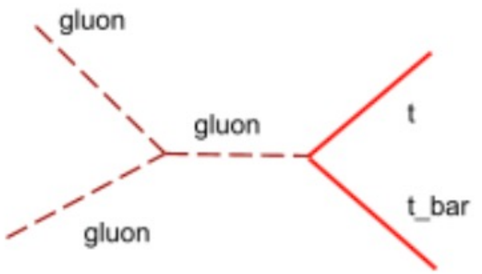
Together:



Notes

Top pair production

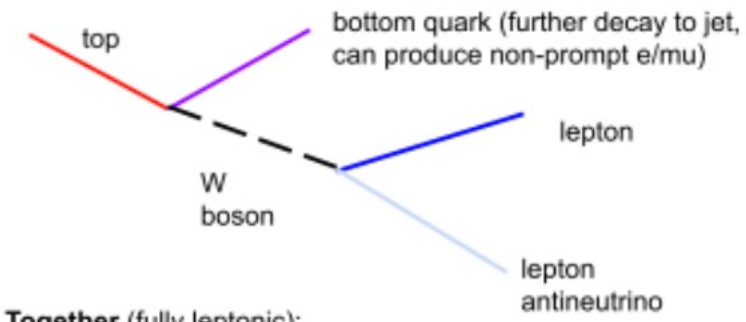
Top pair production:



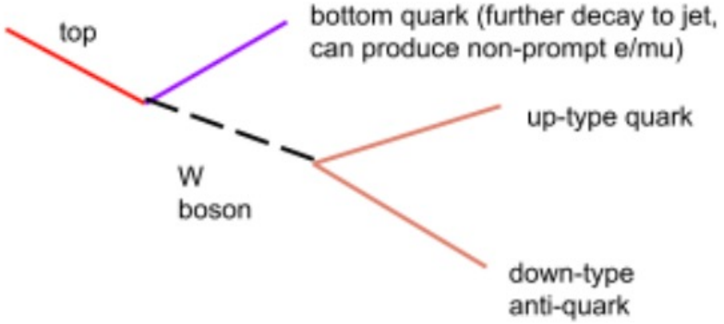
[also works with quark-antiquark on the left hand side]

Top quark decay: 2 modes:

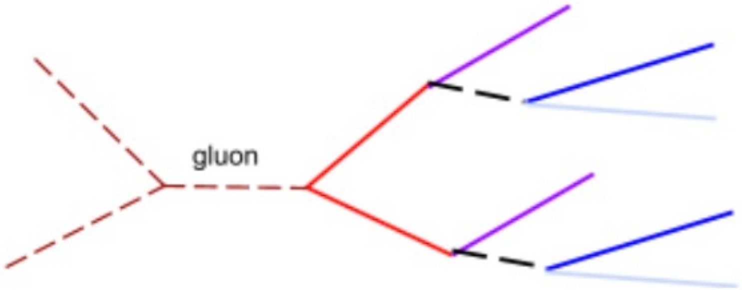
"<u>Semileptonic</u>" - interesting for us as prompt lepton



or "hadronic" - not interesting to us - no prompt leptons



Together (fully leptonic):



Final state:
2 charged leptons - opposite charges, can have any combination of flavours
2 b quarks (b and anti-b)
2 neutrino (neutrino + antineutrino)

Methods

Cuts – threshold requirements a particle must achieve to be included in a selection
Monte Carlo Simulation – predict data resulting from a particular SM process based on theory, used to infer data compositions

Goal: Study non-prompt leptons passing signal requirement

- ➢ Identify sample of decays to lepton pairs, where any additional leptons are likely non-prompt
- ➢ Analyze selection of additional non-prompt signal leptons from the sample and purify selection of prompt sources, thereby reducing the systematic errors from MC

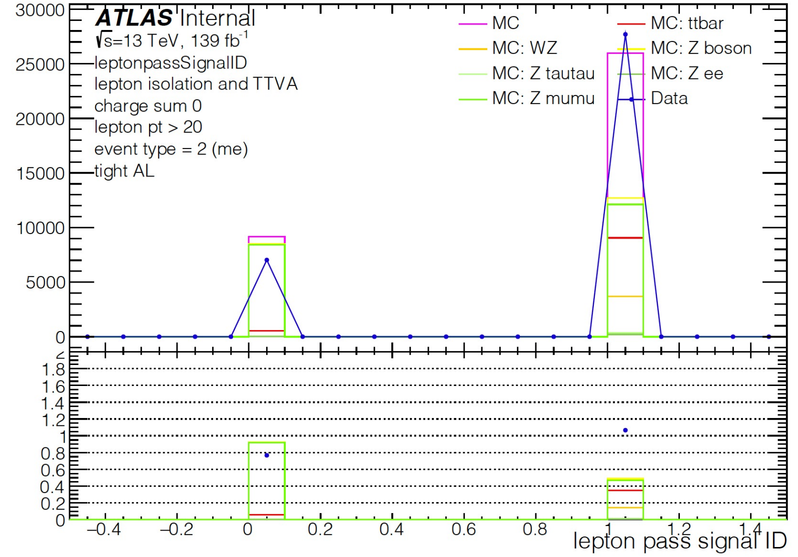
What to consider:

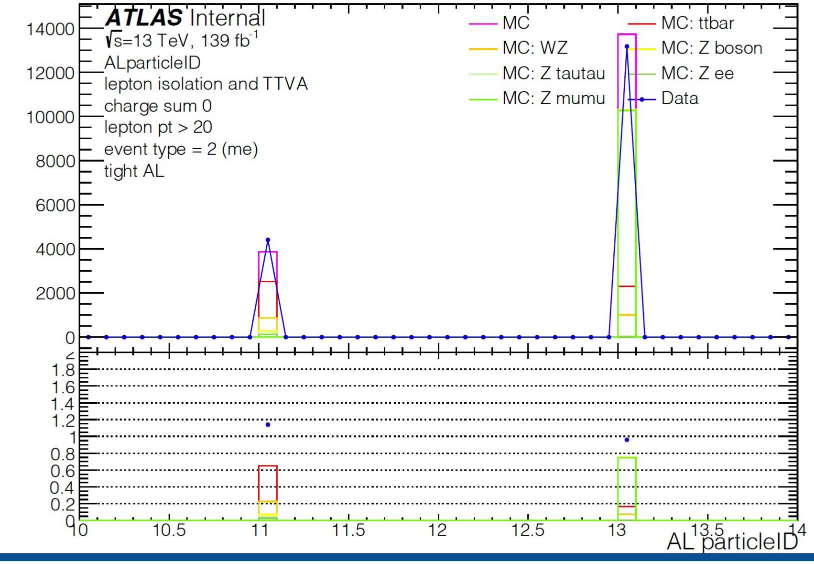
- Requires significant runtime and computing power
- Eliminating data can decrease systematic uncertainty but also increase statistical uncertainty
- Evaluate the efficacy of purification

Methods (notes)

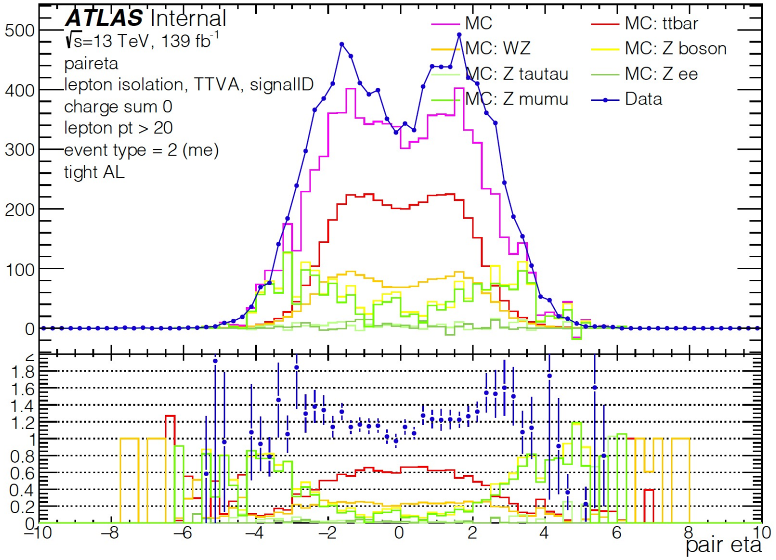
1. Preliminary analysis and event selections
   Isolation of Z+jets and comparison to SM MC simulations
   Preliminary study of misidentified leptons and their kinematic properties (including invariant masses, transverse momenta, angular correlation variables…)
   Requires dealing with large runtimes and computations power
2. Purification of Non-prompt lepton sample from prompt sources
   Reduces systematic uncertainty from MC
   Must minimize systematic uncertainty while maintaining enough data to keep statistical uncertainty low
3. Identify rejection power of frequently used analysis cuts
   Evaluate efficacy of purification cuts
4. Use generator info to estimate "true" rejection power

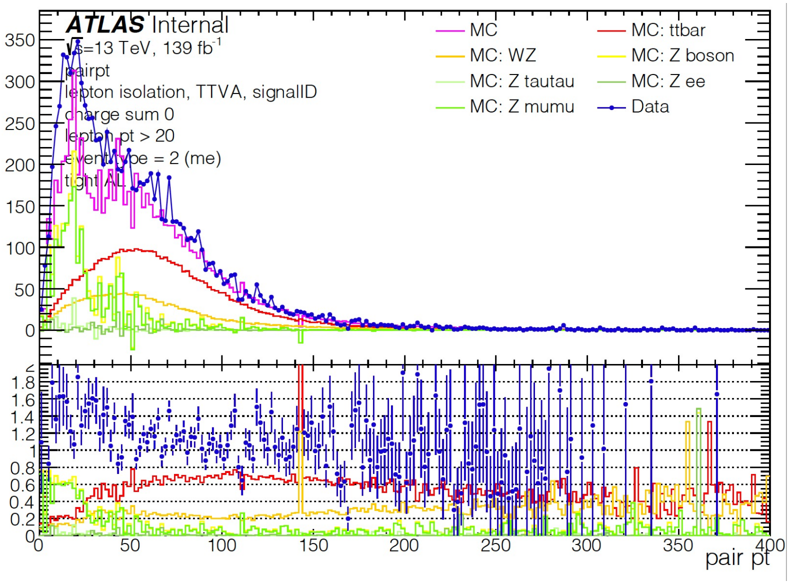
Purifying ttbar selection (old ver.)

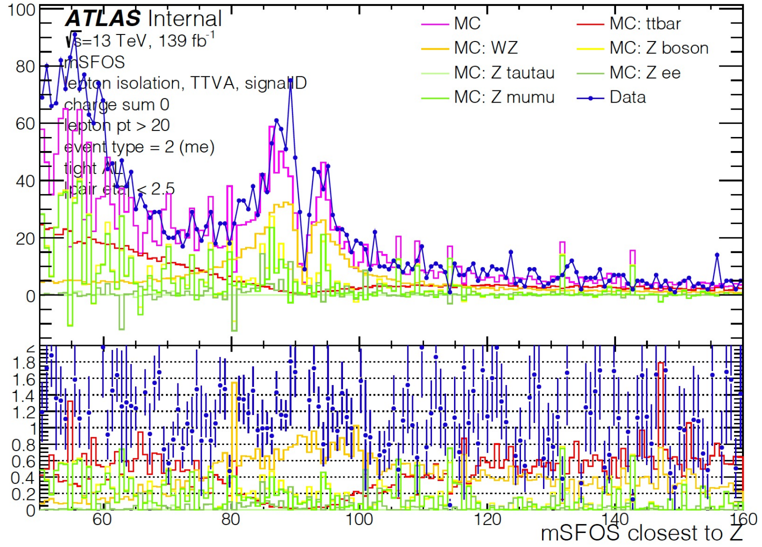




Preliminary selection plots



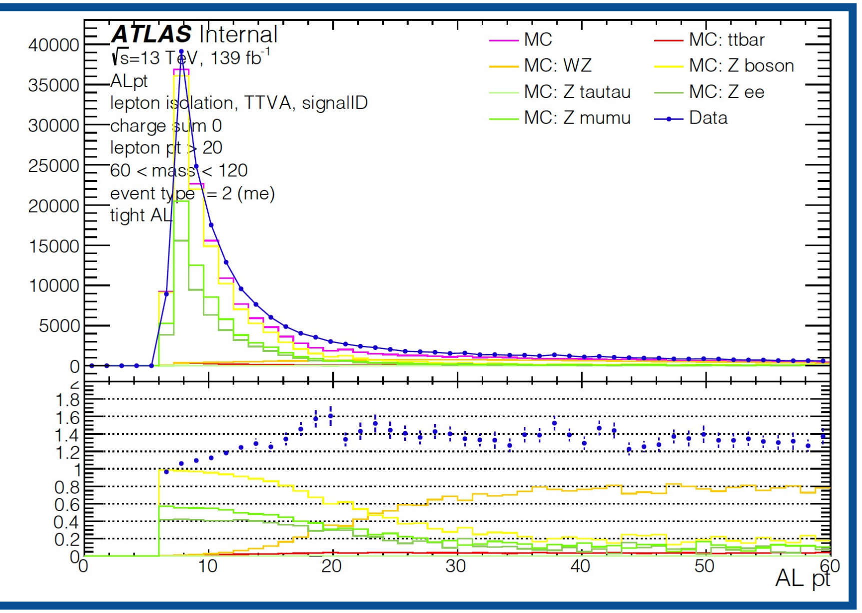


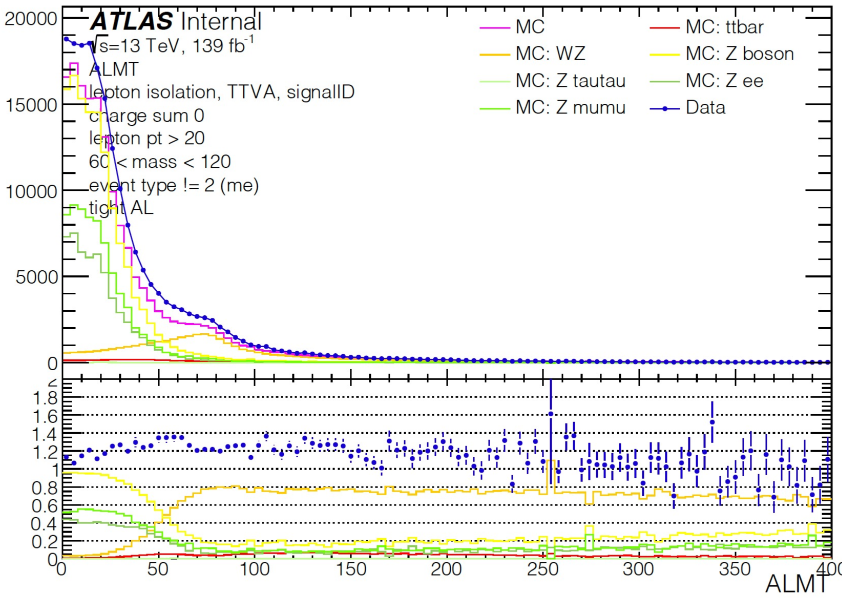


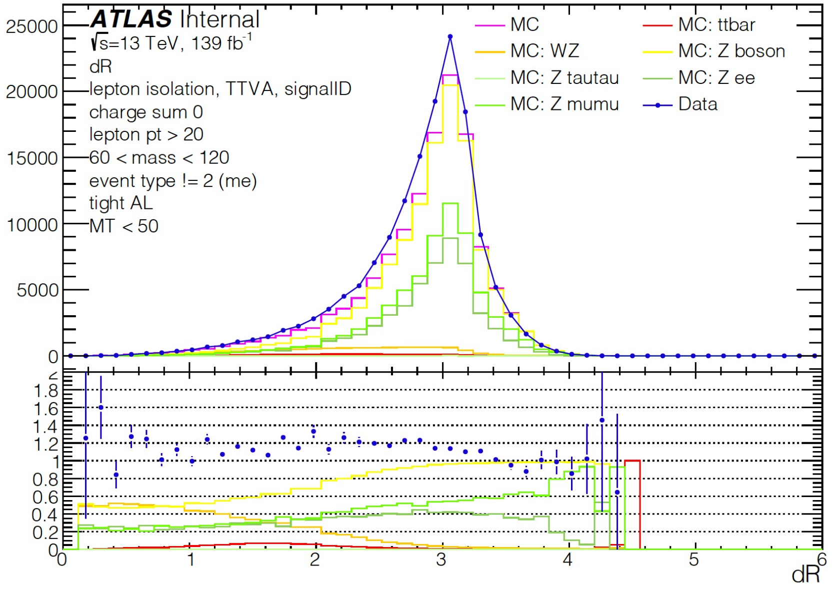
Added cuts:

- Eliminate Z boson events
  - SignalID: veto if 1+ lepton does not pass
  - |pair eta| < 2.5
  - pair pt > 30
- Eliminate WZ events
  - mSFOS < 75 or > 105
  - pair m > 25
  - dR < 4

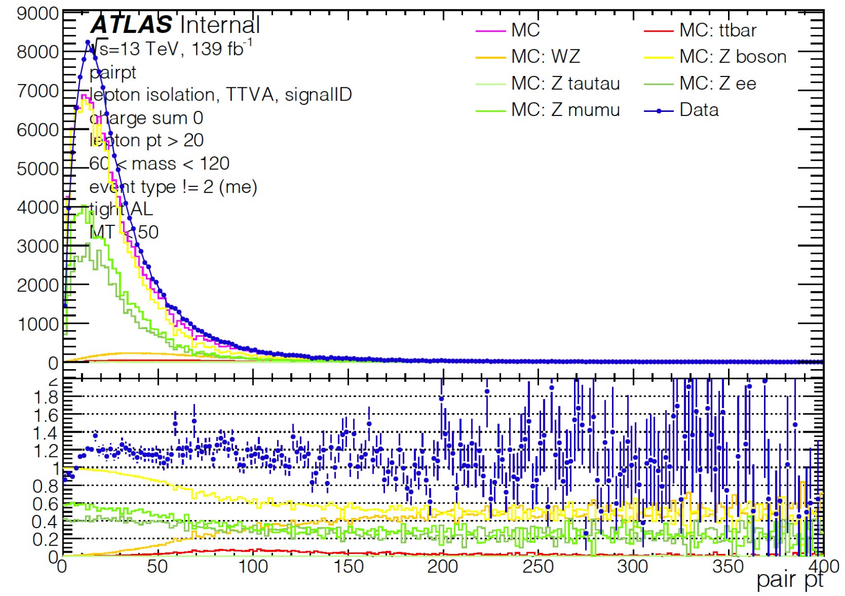
Purifying Z boson selection

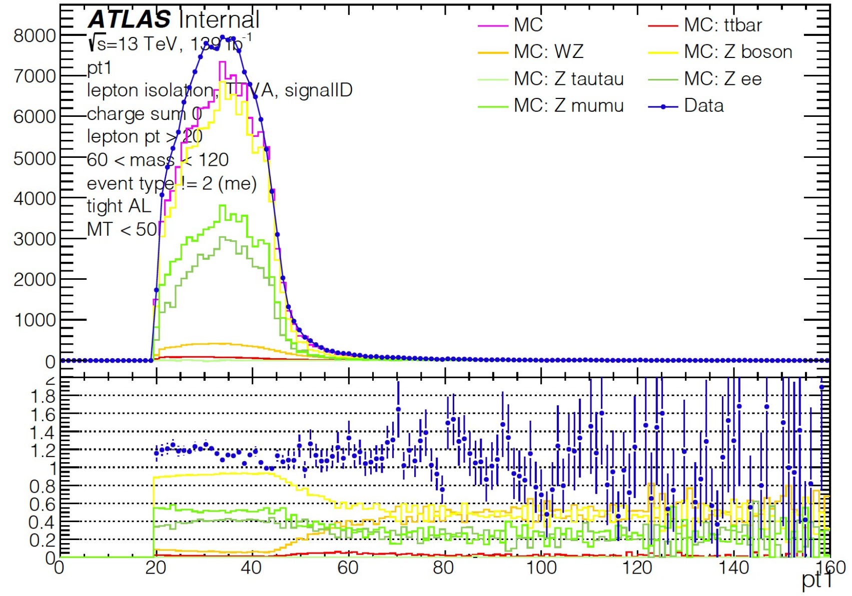






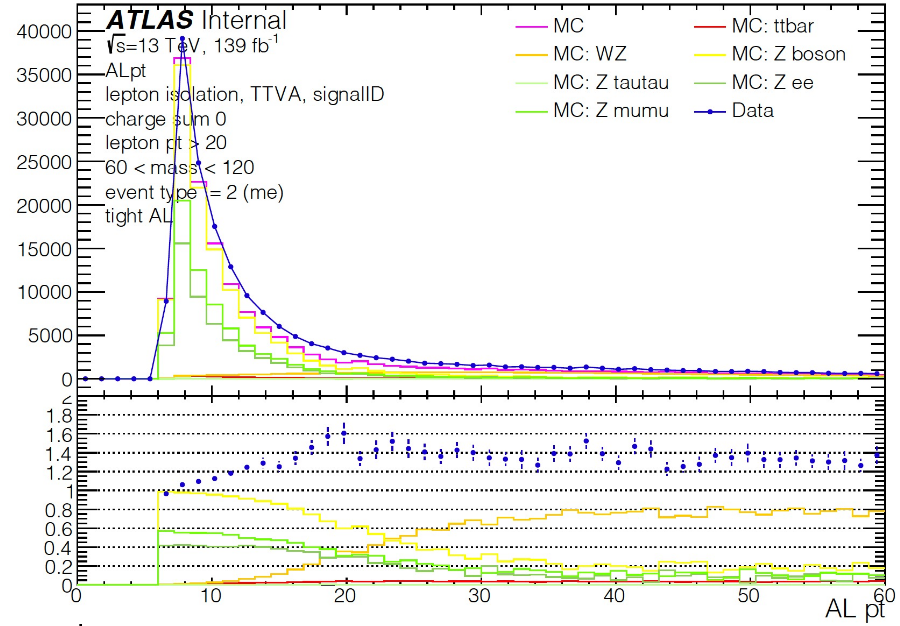
Preliminary selection plots



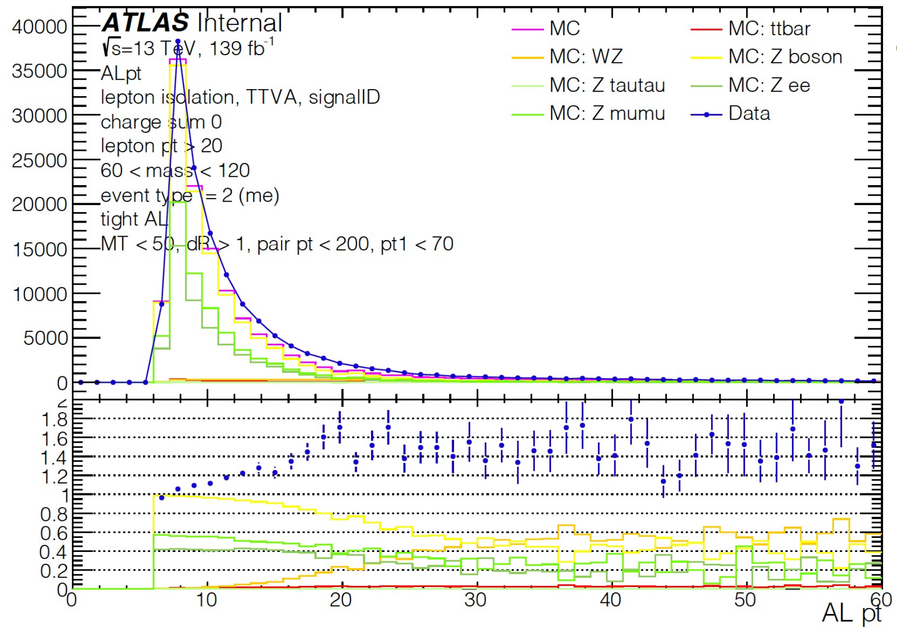


Added cuts:

- Eliminate WZ events
  - mT < 50
  - dR > 1
  - pair pt < 200
  - pt1 <70



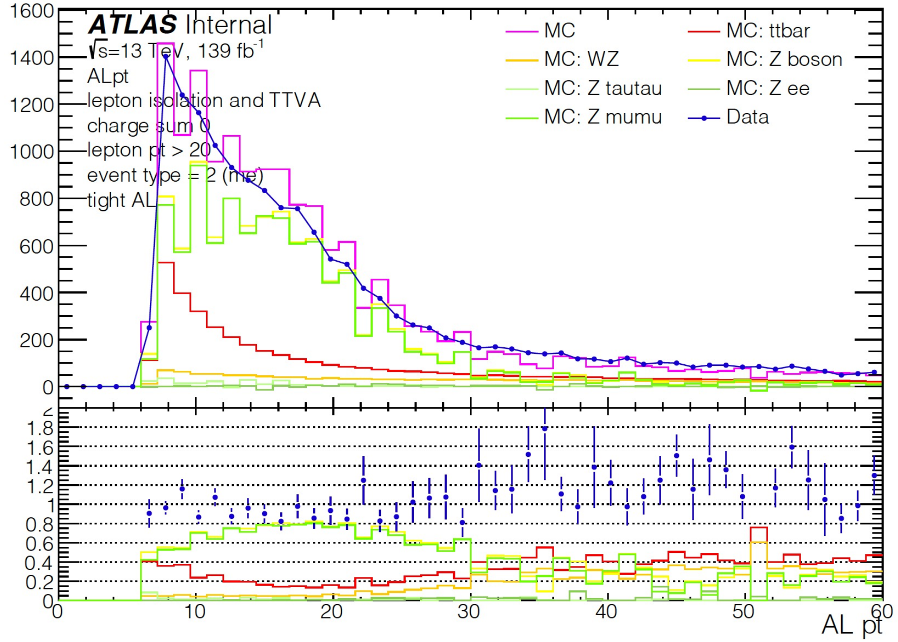
before



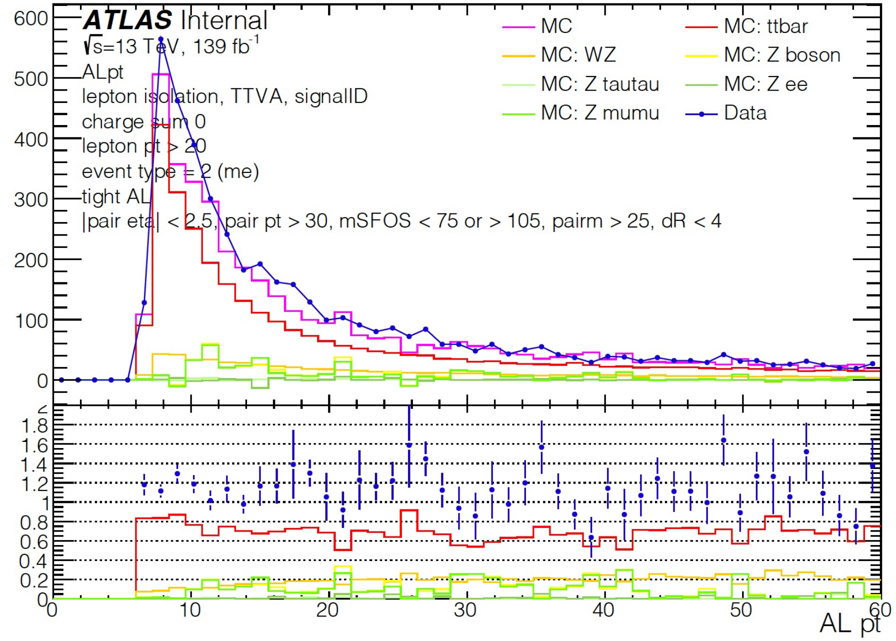
after

Z boson selections

ttbar selections



before



after