Section 1

MuTau Analysis: Data/QCD Event Count Issue

Simple Presentation for Advisors

Slide 1: The Problem

Observation

- **Data events: 1,708** (expected ~3,400)
- **QCD events: 234** (expected ~468)
- Both at approximately **50% of expected values**

Physics Check Failed

Data should be MC simulation (fundamental expectation)
Currently: **Data** $<$ **MC**
This proves something is wrong with the processing!

Slide 2: When Did This Start?

Processing Change

Changed from **artificial pileup protons** → **real proton data**

Old method (like DY/ttjets):

`Process events → Add artificial protons (100% get protons)`

New method (Data/QCD):

`Process events → Use real protons from detector (16.9% have protons)`

Expected impact: ~6x reduction (100% → 16.9%)

Observed impact: ~2x reduction

Something else is also wrong!

Slide 3: Diagnostic Results - Event Flow

One typical file (1M events):

```
Raw NanoAOD:                1,023,514 events
After lumi + HLT:           1,023,514
Has protons (both arms):      172,895 (16.9%)
+ Muon ID:                    138,903 (80% pass)
+ Tau ID cuts:                     65 (99.95% REJECTED!)
+ pT cuts:                          1
+ Eta cuts:                         1

Final: 1 Data event, 0 QCD events per file
```

99.95% rejection at Tau ID step!

Slide 4: Root Cause - Tau VSmu Distribution

Tau ID Requirements:

- **tau_id1 (VSjet):** $> 63 \rightarrow 83\%$ pass
- **tau_id2 (VSe):** $> 7 \rightarrow 85\%$ pass
- **tau_id3 (VSmu):** $> 1 \rightarrow$ **2.2% pass**

VSmu Acceptance:

VSmu Value	Events	Percentage
$= 0$ (fails)	991,842	96.91%
> 0 (has some score)	31,672	3.09%
> 1 (current cut)	22,479	**2.20%**

Only 3% of taus have ANY VSmu discrimination!

Slide 5: Why Is VSmu So Low?

Checked Raw NanoAOD:

```
Row 6,  Tau[1]: pT=30.6,  VSmu=15
Row 11, Tau[0]: pT=37.0,  VSmu=15
Row 14, Tau[0]: pT=36.6,  VSmu=0
Row 14, Tau[1]: pT=27.9,  VSmu=15
Row 14, Tau[2]: pT=26.2,  VSmu=15
Row 16, Tau[0]: pT=100.7, VSmu=1
Row 16, Tau[1]: pT=41.8,  VSmu=15
```

In raw NanoAOD: ~10-15% of taus have VSmu > 0

In our fase0 output: Only 3% have VSmu > 0

Why the difference?

Slide 6: The Bug - Wrong Tau Selection

Current Fase0 Logic:

```
.Define("tau_pt", "Tau_pt[0]")     # Always picks FIRST tau
.Define("tau_id3", "Tau_idDeepTau2017v2p1VSmu[0]")
```

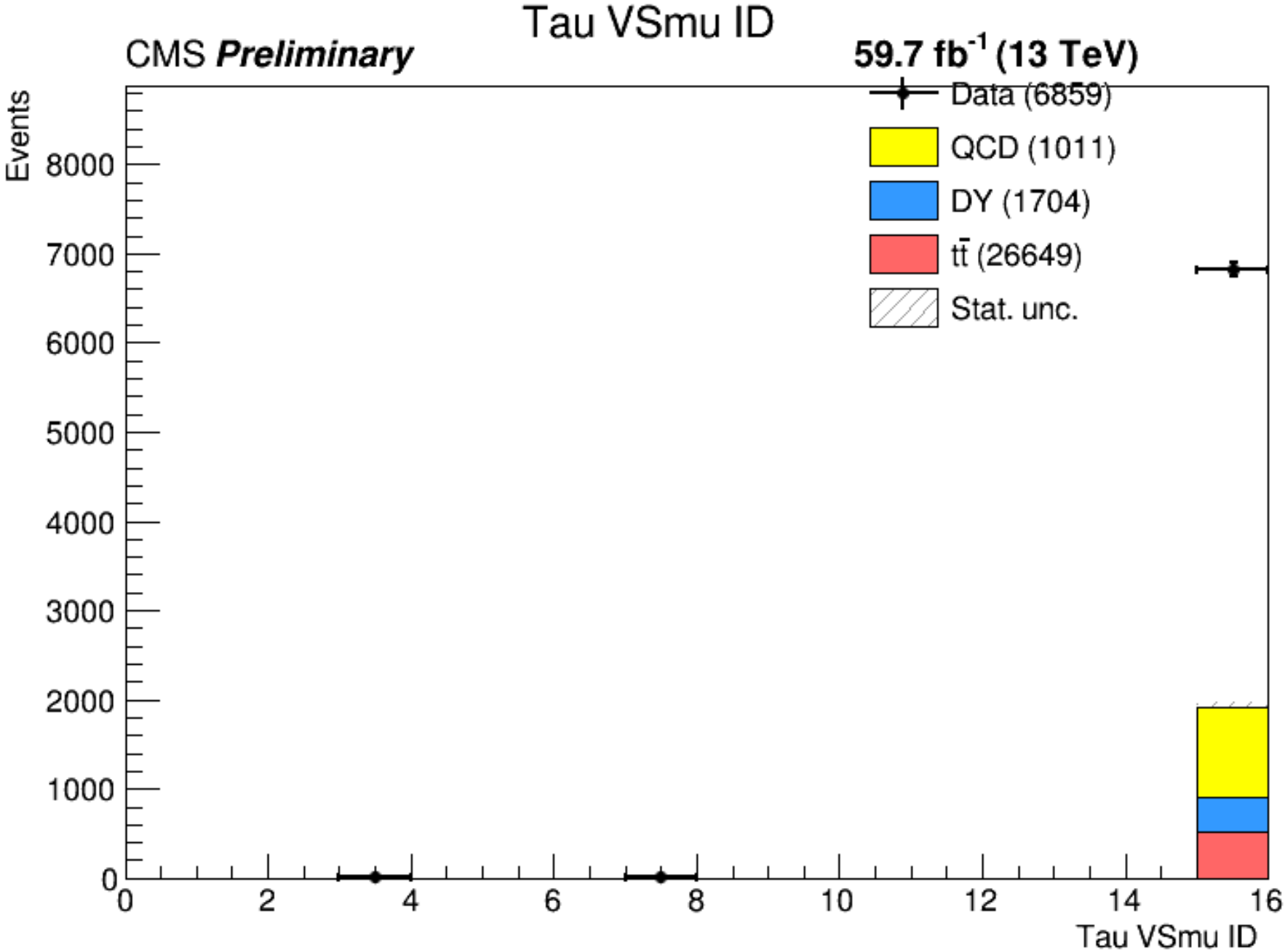
[0] = Highest pT tau (not best quality!)

Example (Row 14 from NanoAOD):

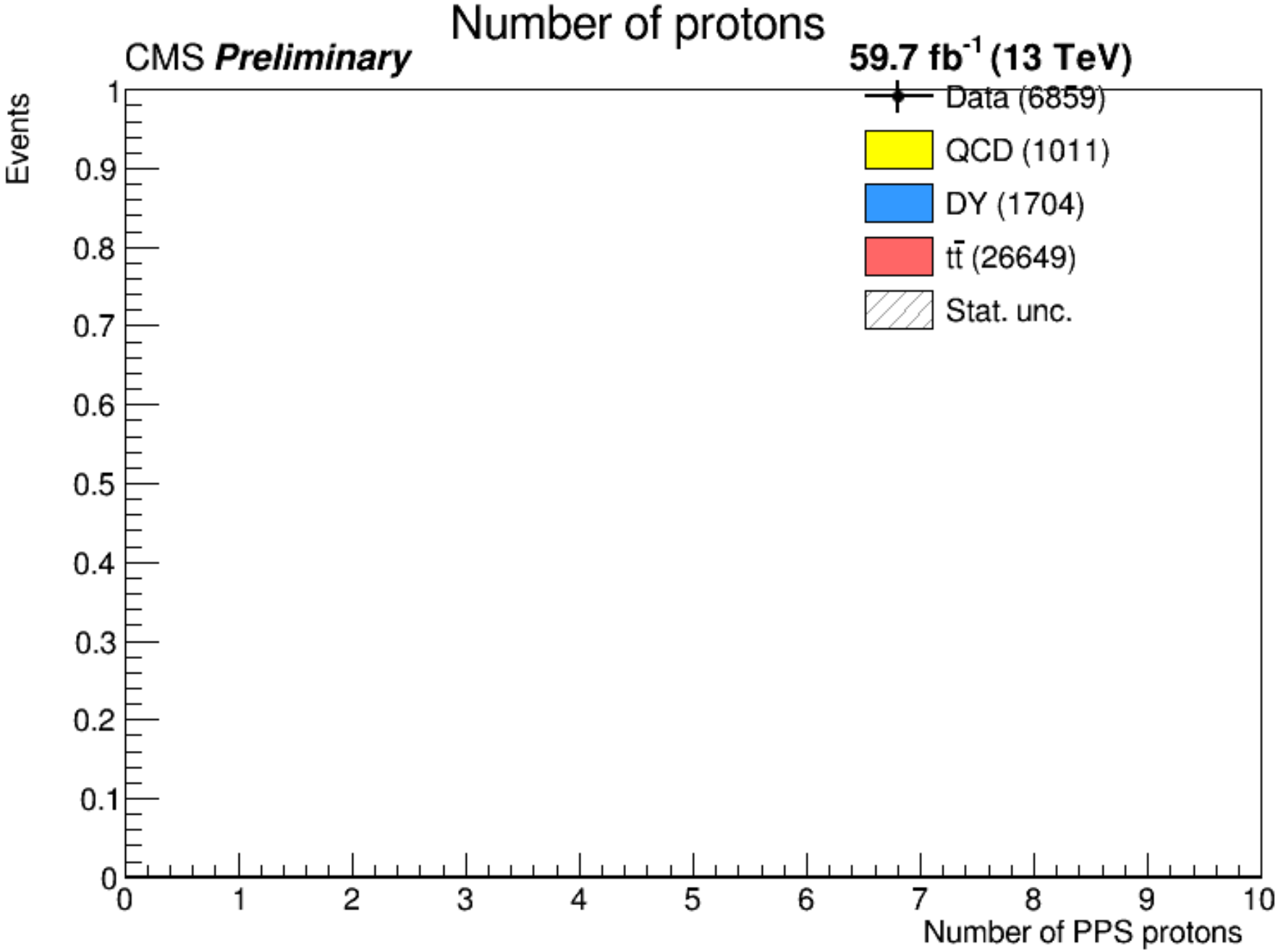
```
Tau[0]: pT=36.6 GeV, VSmu=0  ← PICKED (highest pT, bad quality)
Tau[1]: pT=27.9 GeV, VSmu=15 ← IGNORED (lower pT, good quality)
Tau[2]: pT=26.2 GeV, VSmu=15 ← IGNORED (lower pT, good quality)
```

In MuTau channel with real muon present: - Highest pT "tau" is often a misidentified muon or jet - Real taus are at lower pT - **We're systematically picking the wrong tau!**

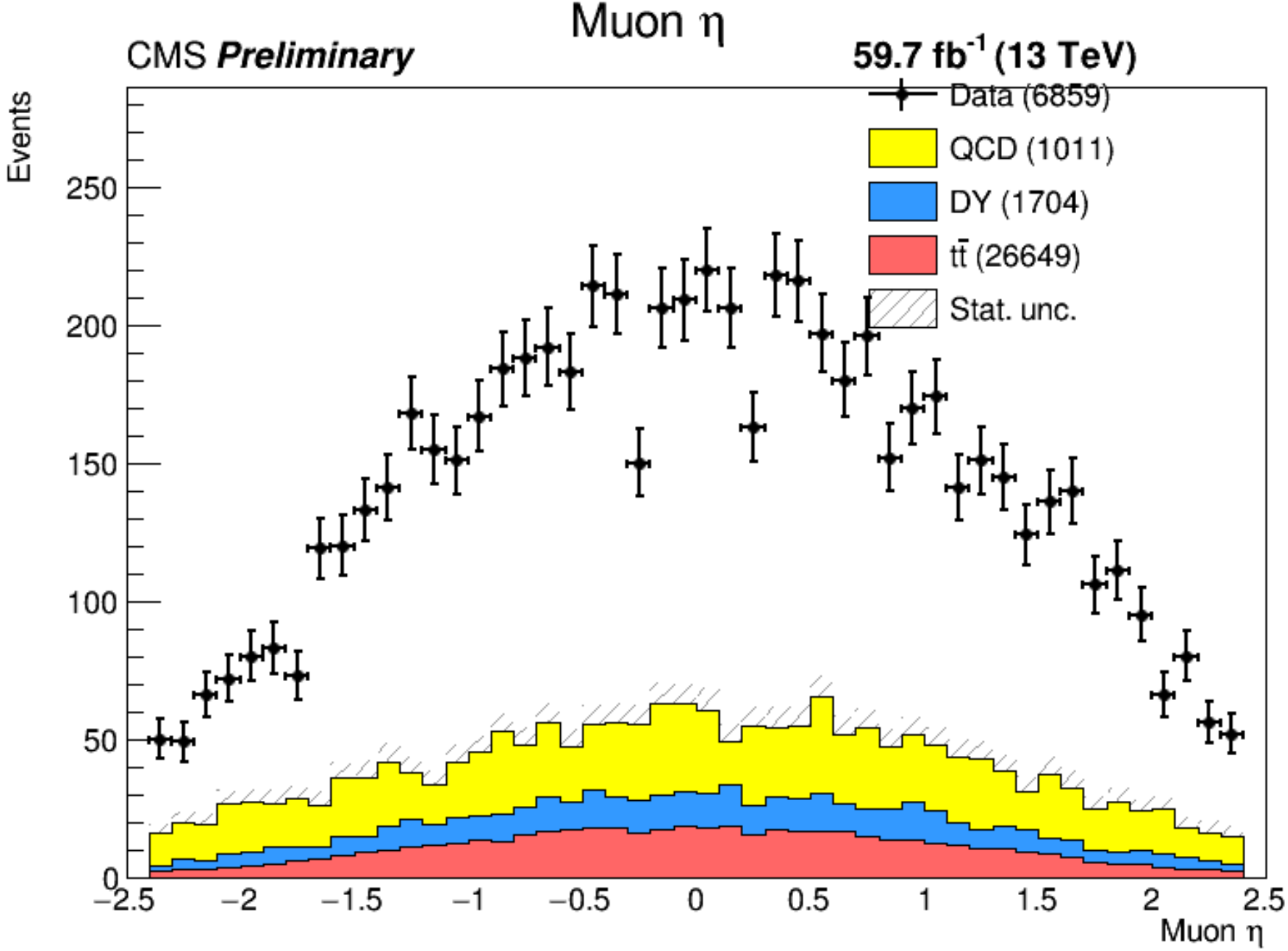
Slide 7: Diagnostic Plots - tau_id3 (VSmu)



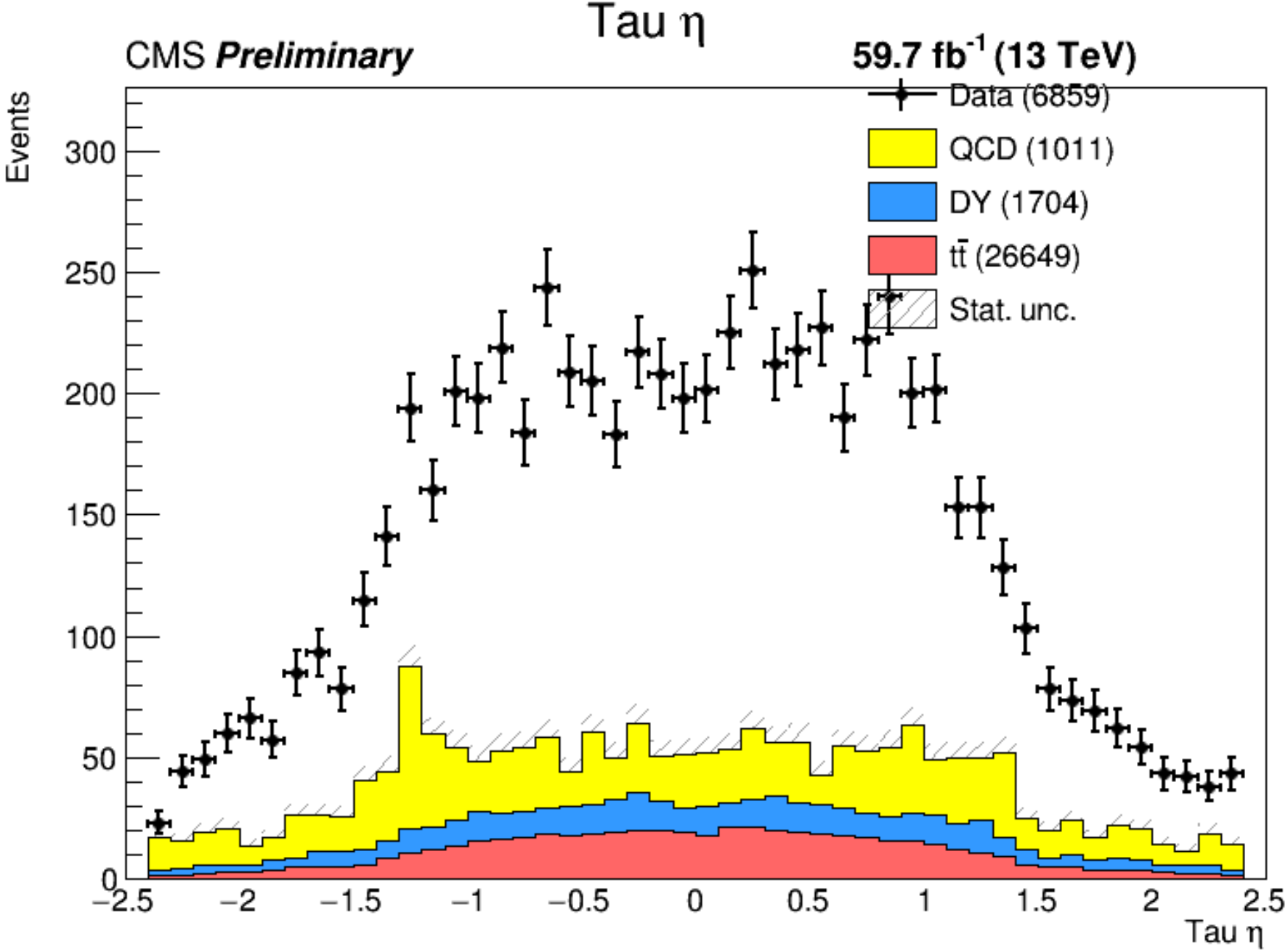
Slide 8: Diagnostic Plots - Number of Protons



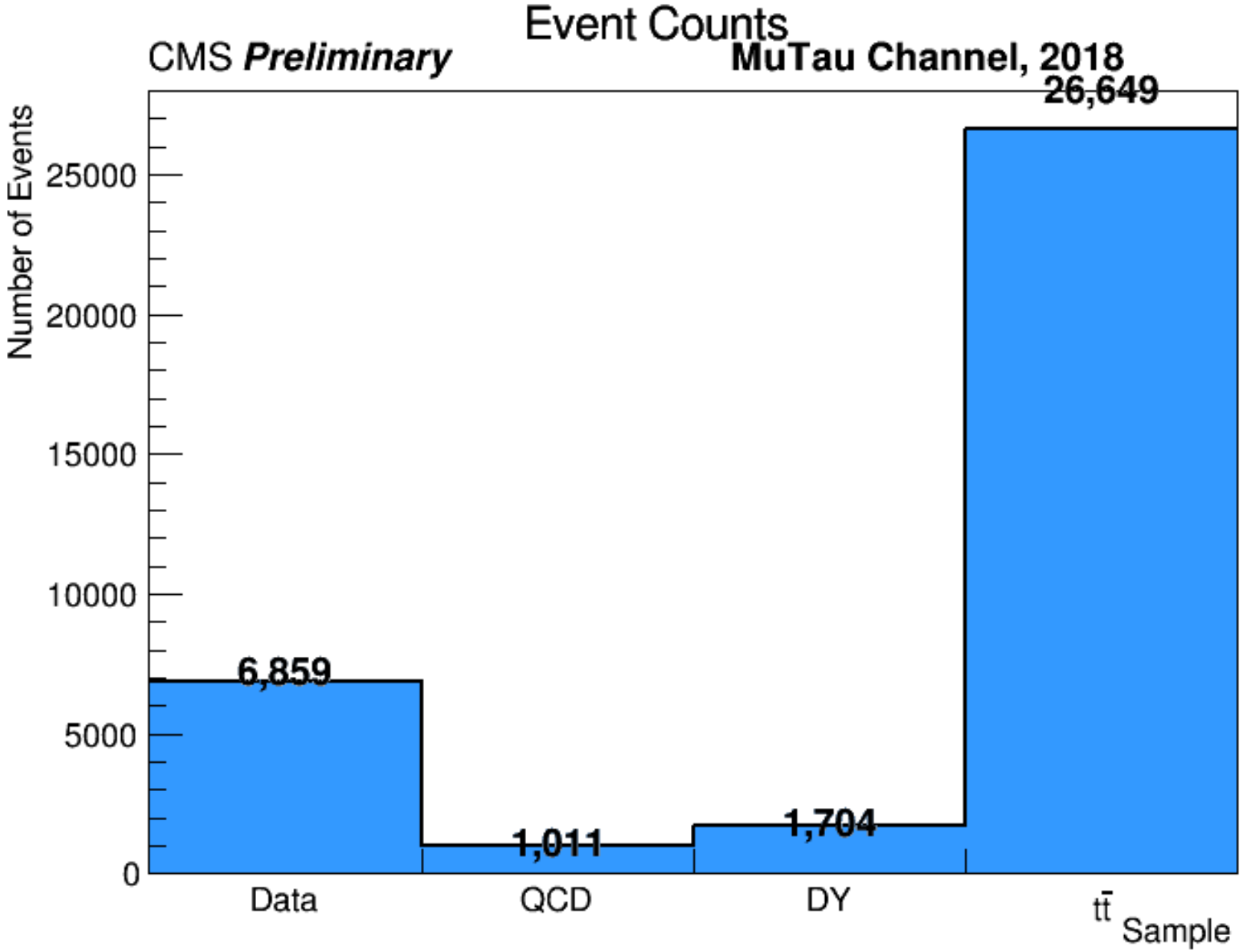
Slide 9: Diagnostic Plots - Muon Eta



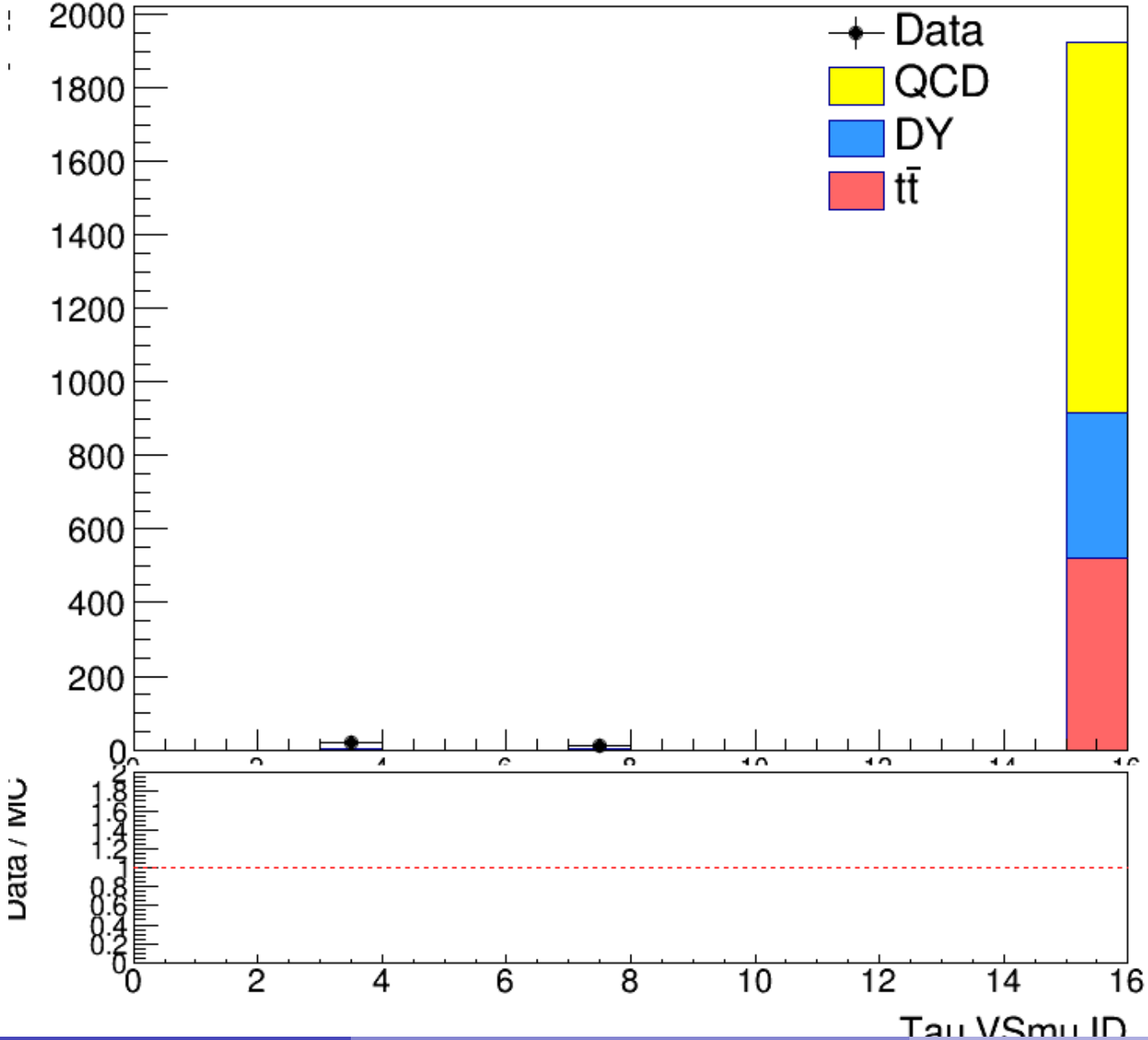
Slide 10: Diagnostic Plots - Tau Eta



Slide 11: Diagnostic Plots - Event Counts



Slide 12: Diagnostic Plots - Data/MC Ratio



Slide 13: The Solution

Modify Fase0 to Select Best Quality Tau

Instead of:

```
.Define("tau_pt", "Tau_pt[0]")  # Highest pT
```

Use:

```
.Define("best_tau_idx", """
    int best_idx = -1;
    int best_vsmu = -1;

    // Loop through all taus
    for (size_t i = 0; i < Tau_pt.size(); ++i) {
        // Require minimum quality
        if (Tau_pt[i] < 100) continue;
        if (Tau_idDeepTau2017v2p1VSjet[i] < 63) continue;
        if (Tau_idDeepTau2017v2p1VSe[i] < 7) continue;

        // Pick tau with best VSmu
        if (Tau_idDeepTau2017v2p1VSmu[i] > best_vsmu) {
            best_vsmu = Tau_idDeepTau2017v2p1VSmu[i];
            best_idx = i;
        }
```

Slide 14: Expected Improvement

Current:

```
VSmu acceptance: 3%
Data: 1,708 events
QCD: 234 events
Data < MC
```

After Fix:

```
VSmu acceptance: 10-15% (3-5x better)
Data: ~5,000-8,000 events
QCD: ~700-1,200 events
Data   MC
```

Remove debugging weight=2.0 (no longer needed)

Slide 15: Why This Happens in MuTau

Physics Context:

MuTau channel has a real muon in the event
Tau VSmu (Tau vs Muon discriminator): - Designed to reject taus that look like muons - In MuTau events, muon presence confuses reconstruction - Highest pT "tau" candidate is often: - The real muon misreconstructed as tau $\rightarrow$ VSmu $= 0$ - A jet near the muon $\rightarrow$ VSmu $= 0$ - A fake tau $\rightarrow$ VSmu $= 0$
Real taus: - Well separated from muon (delta_R > 0.4) - Often at lower pT than fakes - Have good VSmu scores (VSmu 1)
By always picking [0] (highest pT), we systematically select fakes!

Slide 16: Comparison with Old C++ Analysis

Old C++ (nanotry_data.cpp):

```
.Filter([](const vector<float>& aMu){
    return !aMu.empty() && aMu[0] >= 1.f;  // VSmu filter
  }, {"tau_id_antimu"});
```

Applied VSmu filter on [0], but: - Input was from "analyzer/ntp1" (pre-processed) - Likely already had tau selection/ordering different from raw NanoAOD - Or MC-based where tau[0] ordering was different

Our New Python Processing:

- Starts from **raw NanoAOD**
- Tau[0] = highest pT (physics-agnostic ordering)
- Need to **explicitly select best quality tau**

Slide 17: Alternative: Quick Test

Before reprocessing all 393 files, test with looser cut:

In `fase1_data.py` and `fase1_qcd.py`:

```
# Change from:
.Filter("tau_id3 > 1", "Tau VSmu (Loose)")

# To:
.Filter("tau_id3 > 0", "Tau VSmu (VLoose)")
```

Expected: Event count increases by ~40%

If this works → **Confirms VSmu is the bottleneck**

Then implement full solution (best tau selection)

Slide 18: Action Plan

Step 1: Run Diagnostic Plots

```
python3 diagnostic_plots.py
```

Review plots with advisor to confirm hypothesis

Step 2: Quick Test (Optional)

Loosen VSmu cut to verify it's the issue

Step 3: Implement Best Tau Selection

Modify `fase0_qcd_data.py` with best tau selection logic

Step 4: Test on One File

```
python3 fase0_qcd_data.py   # with line_number=0
```

Verify VSmu acceptance improves

Step 5: Reprocess All Files

```
python3 fase0_qcd_data.py   # with line_number=-1
```

Slide 19: Summary

Problem:

- Data/QCD at 50% of expected values
- 99.95% rejection at Tau ID (VSmu) step
- Data $<$ MC (physics check failed)

Root Cause:

- Fase0 picks highest pT tau (tau[0])
- In MuTau channel, tau[0] is often fake (VSmu=0)
- Real taus at lower pT are ignored
- Only 3% acceptance instead of 10-15%

Solution:

- Select best quality tau (not just highest pT)
- Prioritize VSmu score
- Expected improvement: 3-5x more events
- Data will be MC

Slide 20: Questions?

Key Takeaways:

1. **Diagnostic plots** will show if Data has different VSmu distribution than MC
2. **The fix** is simple: select best quality tau instead of highest pT tau
3. **Expected outcome:** Event counts increase 3-5x, Data MC

Files Created:

- `diagnostic_plots.py` - Generate all diagnostic plots
- `FINAL_SUMMARY_AND_SOLUTION.md` - Complete technical details
- Full solution code ready to implement

Next Step:

Review diagnostic plots together, then decide on implementation strategy

Section 2

End of Presentation

End of Presentation

Generated with Claude Code