# Using Tags for Top Preselection

**Marcello Barisonzi**
**DESY**

# Overview

- What is the ATLAS FDR?
- The Analysis Model Forum
- Montecarlo vs. ”Real Data”
- The data TAG
- TAG Analysis for Top Physics
- Conclusions

# What is the ATLAS FDR?

- The ATLAS Full Dress Rehearsal is an exercise to simulate as close as possible the behaviour of data reconstruction and distribution to the GRID, with realistic commissioning physics samples
- Two phases: FDR-1 (Jan/Feb 08) and FDR-2 (Apr/May 08)
- Goal for FDR-1: production of two data samples of 0.8 $pb^{-1}$ each corresponding to a 10- and 1-hr run at $10^{31}$ & $10^{32}$ luminosity
- Goal for FDR-2: aim for $10^{33}$ luminosity
- https://twiki.cern.ch/twiki/bin/view/Atlas/FullDressRehearsal

# The Analysis Model Forum

- Last December, forum started to discuss analysis of data
- Nobody wants to use AOD files: too cumbersome, complicated, user data not supported
- Key is production of **D**erived **P**hysics **D**ata:
  - **Thinning**: dropping unused data (multiple jet, e, mu, tau)
  - **Slimming**: reduce object information (e.g. Pt but no Px,Py)
  - **Skimming**: skip uninteresting events
- Each Physics WG has to define its own objectives, can add proprietary data

# The Analysis Model Forum (*cont.)*

- No official way of producing DPD, several option will be tried out at the FDR:
- AOD--  : Light version of AOD, dropping unused Containers
- AthenaROOT : POOL files readable by ROOT. Sounds promising, but it's not there yet
- Event/Top/SUSYView : used by almost everyone, hated by CERN diehards.

# Montecarlo vs. ”Real Data”

- MC sample analysis is like being a DJ: cut up & mix.
- Cut up the physics into individual samples, one or more for each process; simulate and reconstruct them separately
- Mix the reconstructed samples after the analysis, by applying weights proportional to the sample luminosity
- With the FDR, we get ONE sample, and weights are unknown
- We cannot even analyse the whole data sample (~2.4 Tbyte), how do we select?
- In the Real World will be even worse!

# The Data TAG

- ATLAS Software has a tagging package (RecExCommon) that easies the task to select our events
- TAGs are small (~1kb/event) but quite detailed summaries of the contect of a data sample
- TAGs reside both in ROOT files and in DataBases at Tiers
- Crude analysis can be run on TAG file before full sample is analysed
- **To produce DPD files, we must start selecting TAGs!**

# Inside the TAG

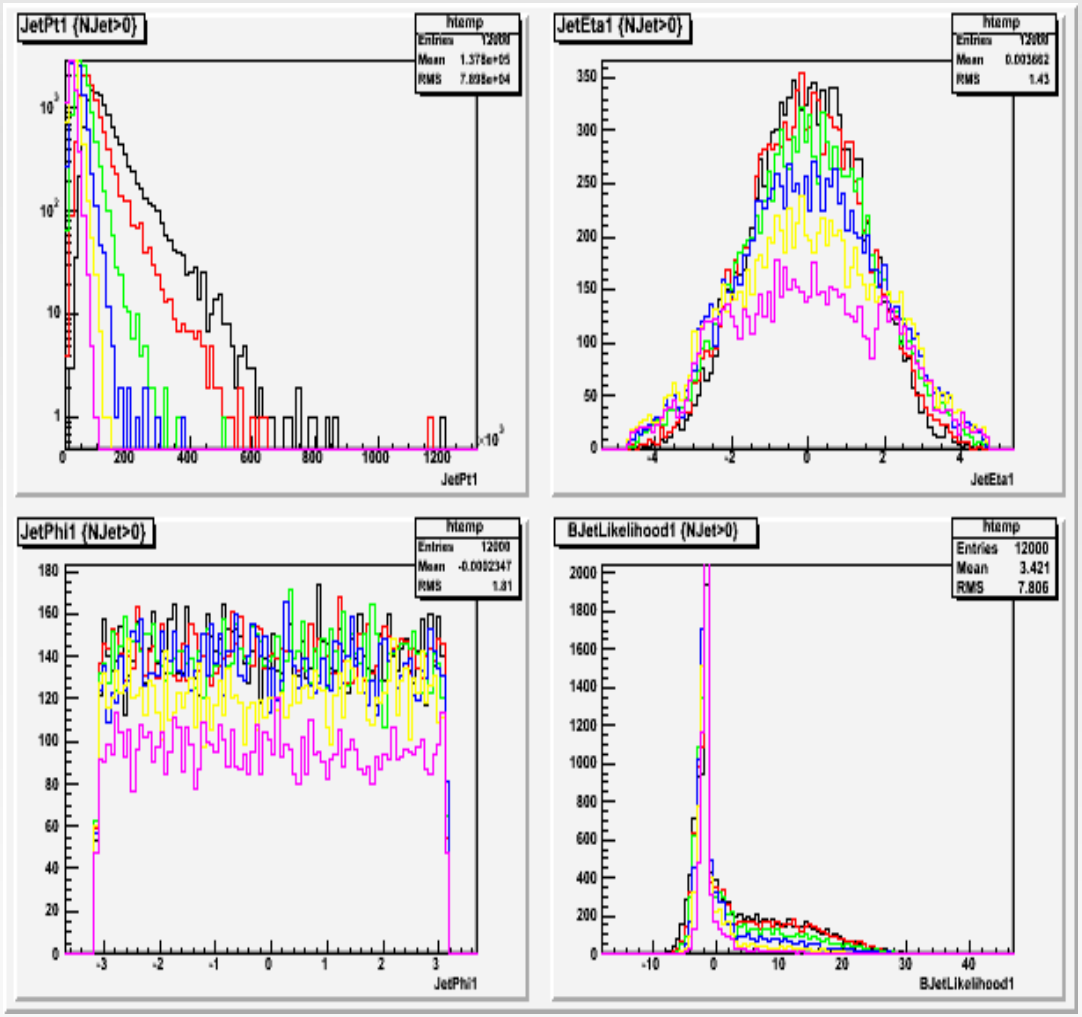
- The TAG contains many objects of interest:
  - First 6 jets in event, ordered by descending Pt
  - First 4 Electrons, 4 Muons, 2 Taus, 2 Photons, MissingEt
  - Trigger words
  - Detector Status & Data Quality
  - Luminosity block
  - **Physics WG-specific TagWord** (more on this later)
  - etcetera
- https://twiki.cern.ch/twiki/bin/view/Atlas/PhysicsAnalysisWorkBookTAG

# TAG for Top Physics

- Basic idea: Identify a set of cuts on tags, then produce DPD only on events that pass the cuts
- Cuts should be as loose as possible to maximize acceptance
- Final user can refine cuts, for this purpose, TAG must be copied in DPD file too
- Advanced idea: use the WG-specific 32-bit TAG word (*TopWord*) to discriminate between basic event topologies

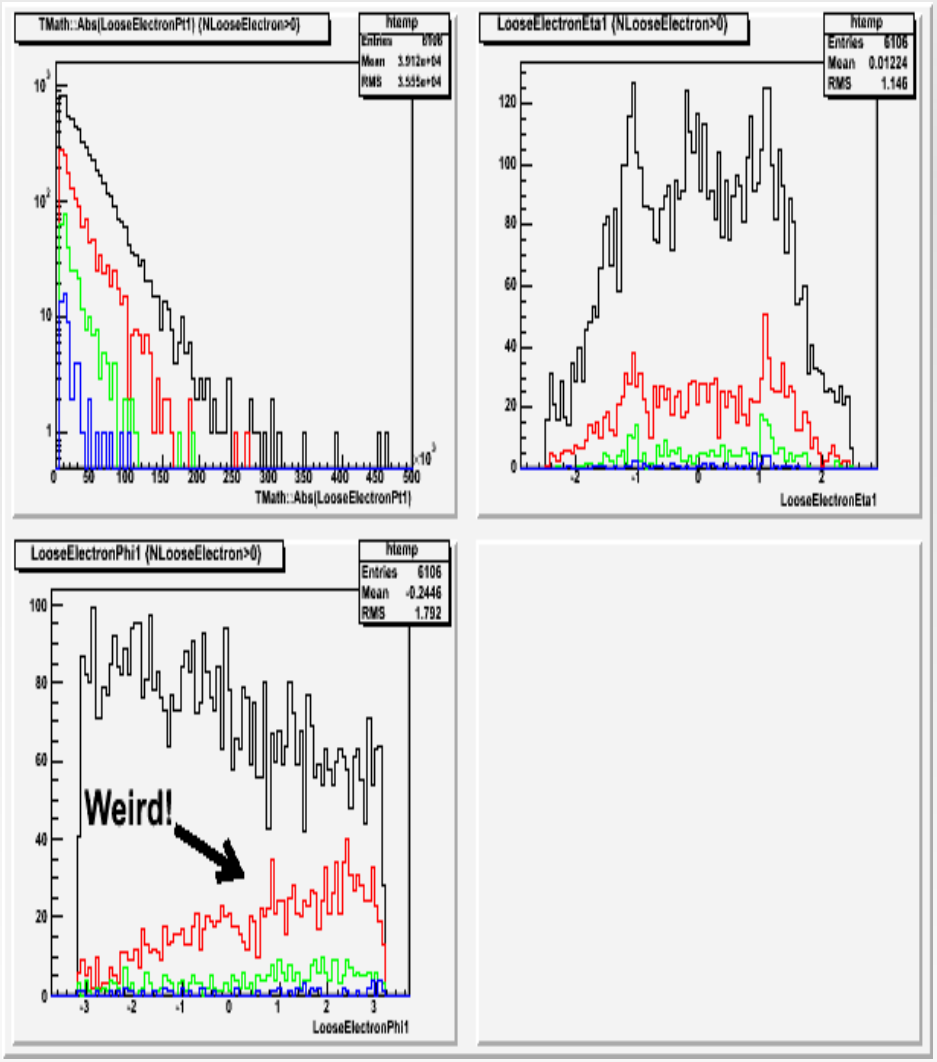
# TAG for ttbar (5200) sample

- Ordered Jets seem OK
- BTagging information is present, though it's better not to use it (60% efficiency)



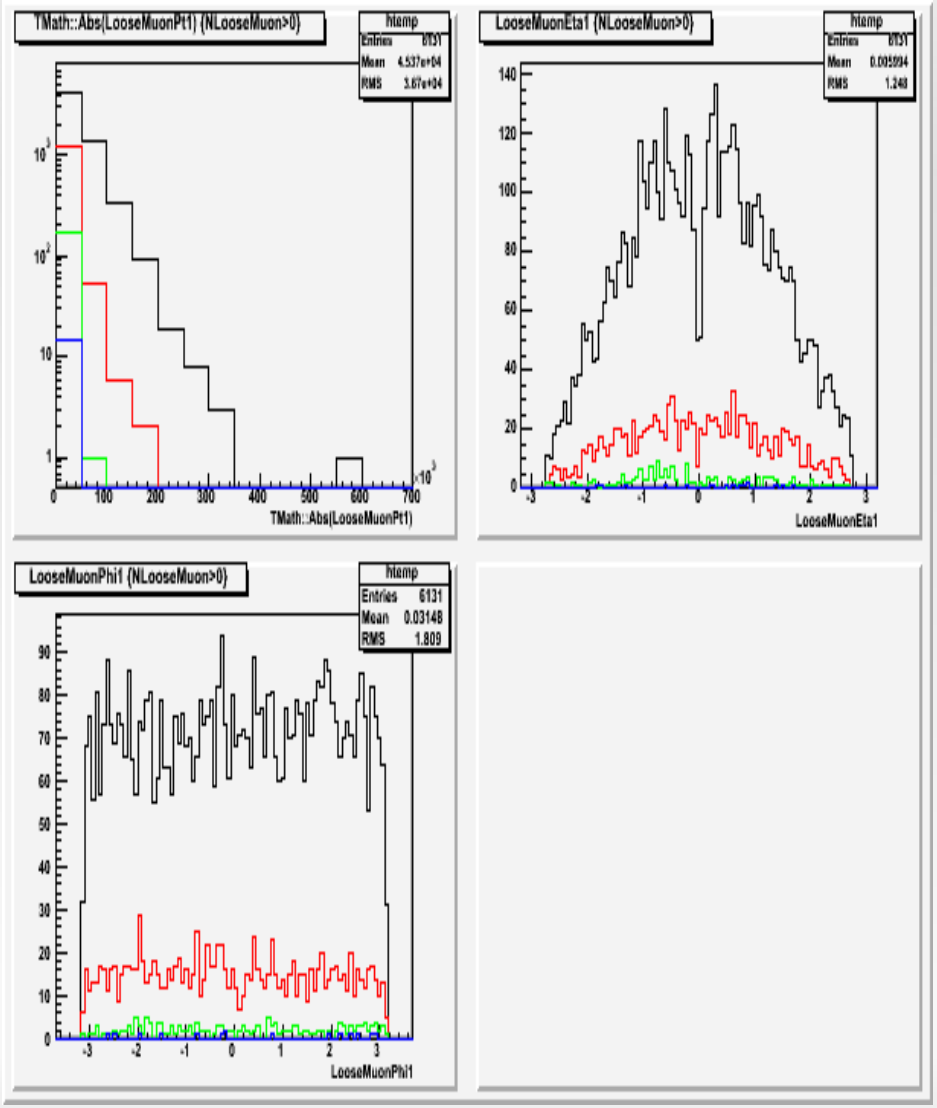
# TAG for ttbar (5200) sample

- WARNING!
  Electron Pt is signed, to encode charge
- Something funny is happening to phi spectrum, have to investigate



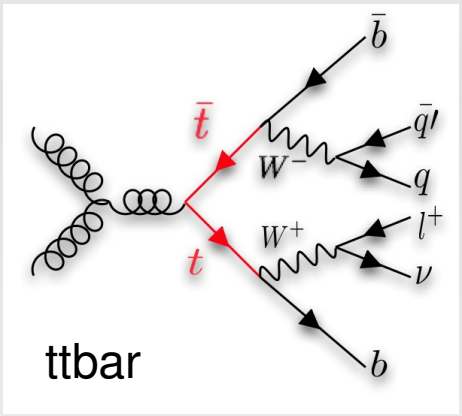
# TAG for ttbar (5200) sample

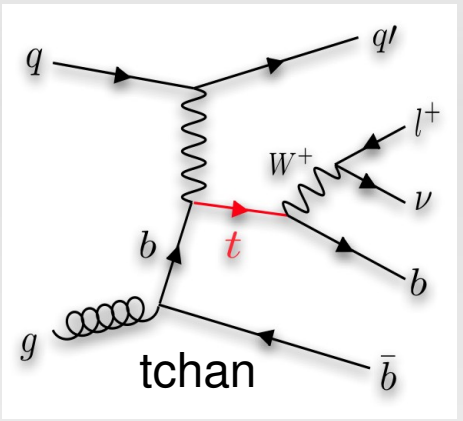
- Muons seem OK, though there are single high Pt muons (Reco fault?)
- Taus are sorely missed

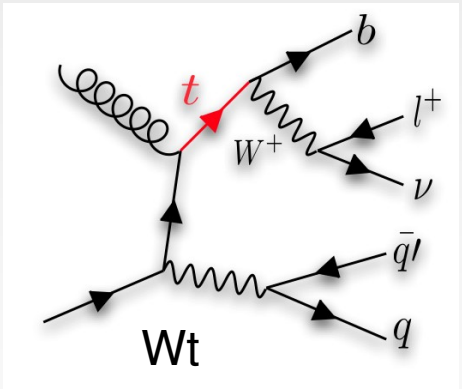


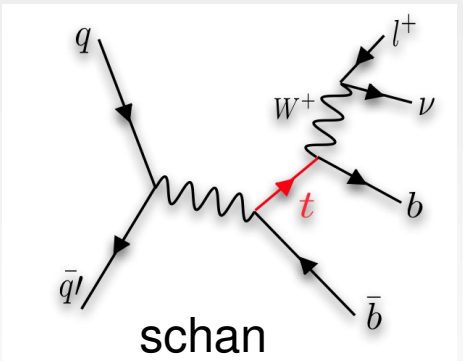
# Selection Cut

- I identify 4 topologies:
  - fully-leptonic ttbar
  - fully-hadronic ttbar
  - semi-leptonic ttbar
  - common single top
- And I choose 4 sets of cuts



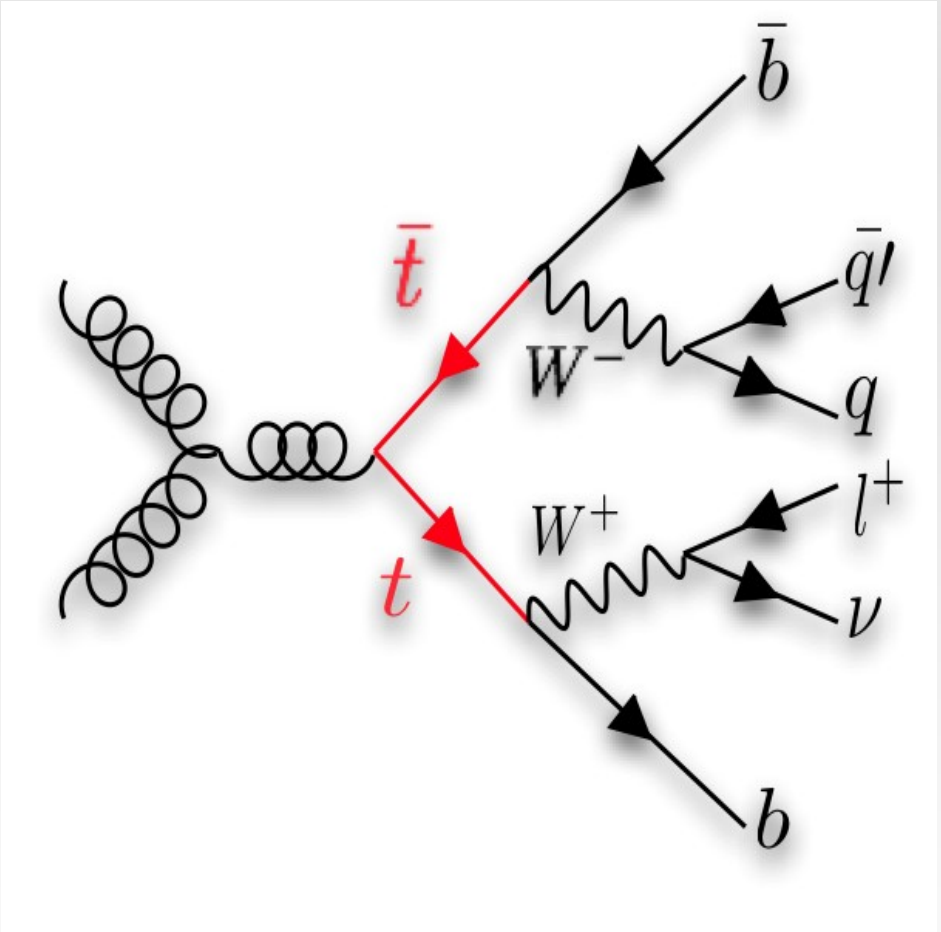



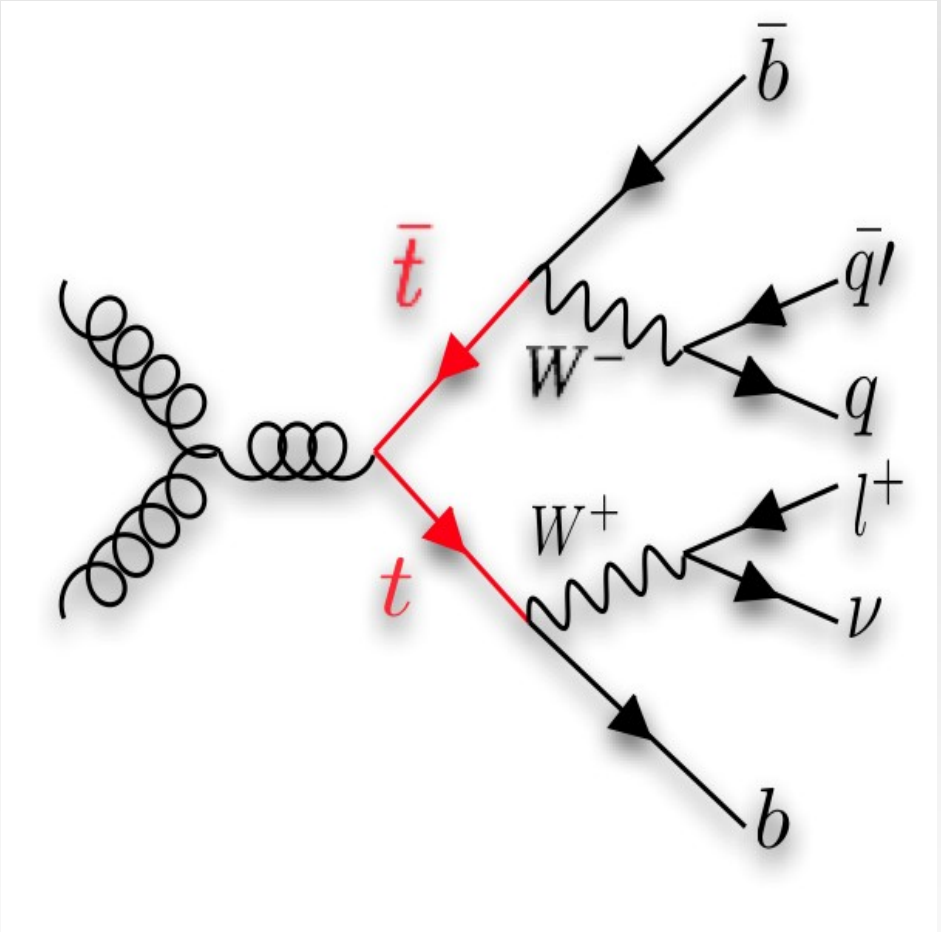
# Selection Cut

- Fully-leptonic ttbar:
  - >=2 Jets > 30 Gev
  - >=2 Leptons > 5 GeV
  - Missing Et > 20 GeV



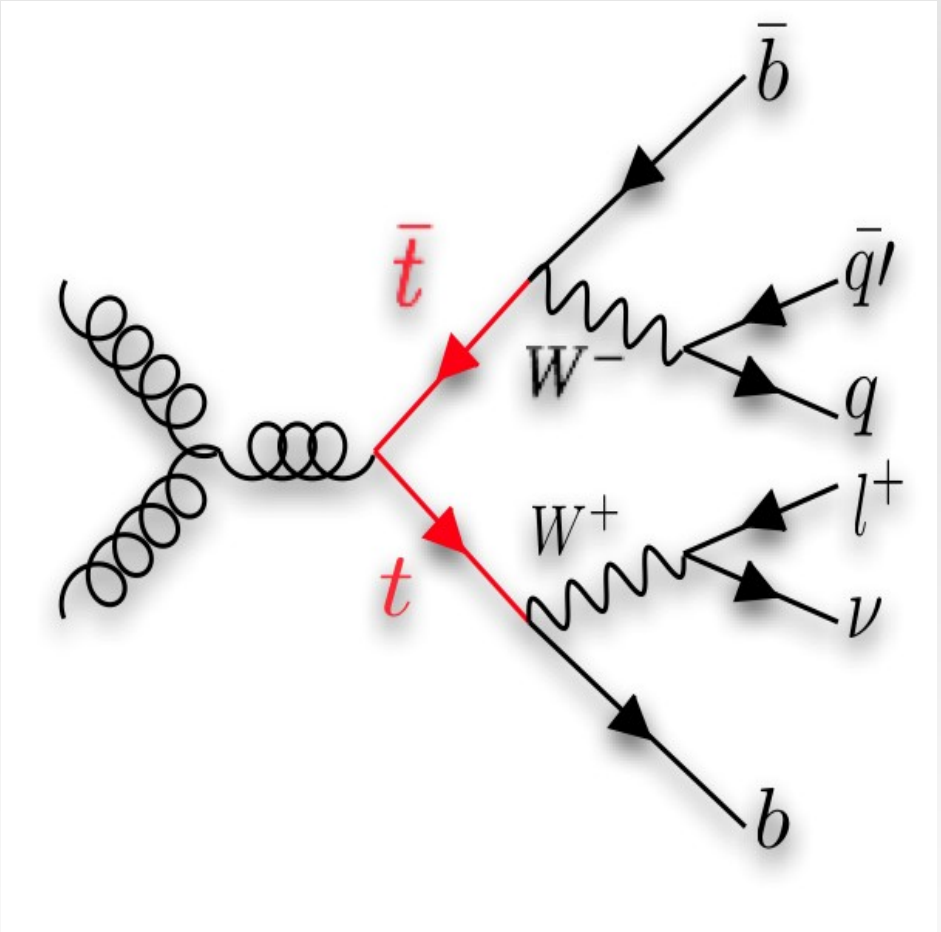
# Selection Cut

- Fully-hadronic ttbar:
  - >=4 Jets > 40 GeV
  - >=1 Jet > 30 GeV



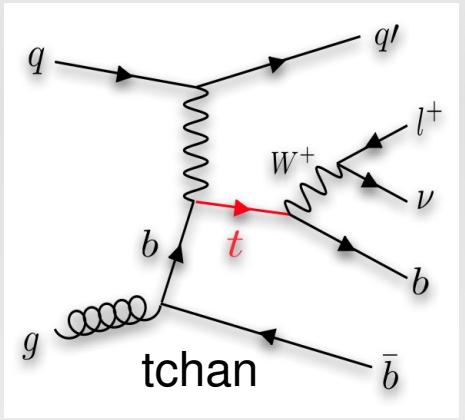
# Selection Cut

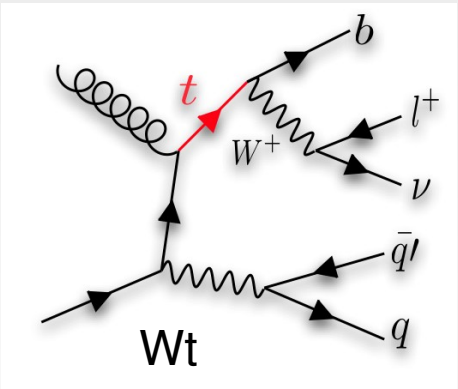
- Semi-leptonic ttbar:
  - >=2 Jets > 40 GeV
  - >=1 Jet > 30 GeV
  - >=1 Lepton > 5 GeV

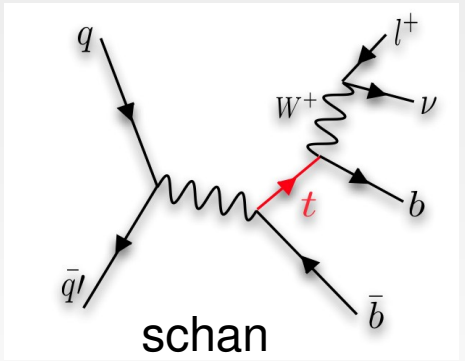


# Selection Cut

- Common single top
  - >=2 Jets > 30 Gev
  - >=1 Leptons > 5 GeV
  - Missing Et > 20 GeV



# Results

- Applying these cuts on TAGs obtained by the usual CSC data sample (13.0.30), I can predict the composition of the Top DPD data at FDR for 0.8 $pb^{-1}$
- Semi-leptonic topology: (ε=65%)
  240 correct events, ~30000 background (QCD, other topologies)
- Fully-hadronic topology: (ε=69%)
  205 correct events, ~120000 background events
- Common Single Top: (ε=53-66%)
  1-39 correct events, ~300000 background events
- **Topologies are not mutually exclusive!**

# Conclusion

- By applying the logical OR of all three topologies, one can achieve ~60-70% efficiency for all channels
- Discriminating power at this level sufficient to perform some rough topology recognition
- The resulting ntuple size is less than 500k, which is manageable with the current computing facilities
- Trigger efficiencies and more detailed background studies must be included
- To be investigated: Top Mass Reconstruction at TAG Level?