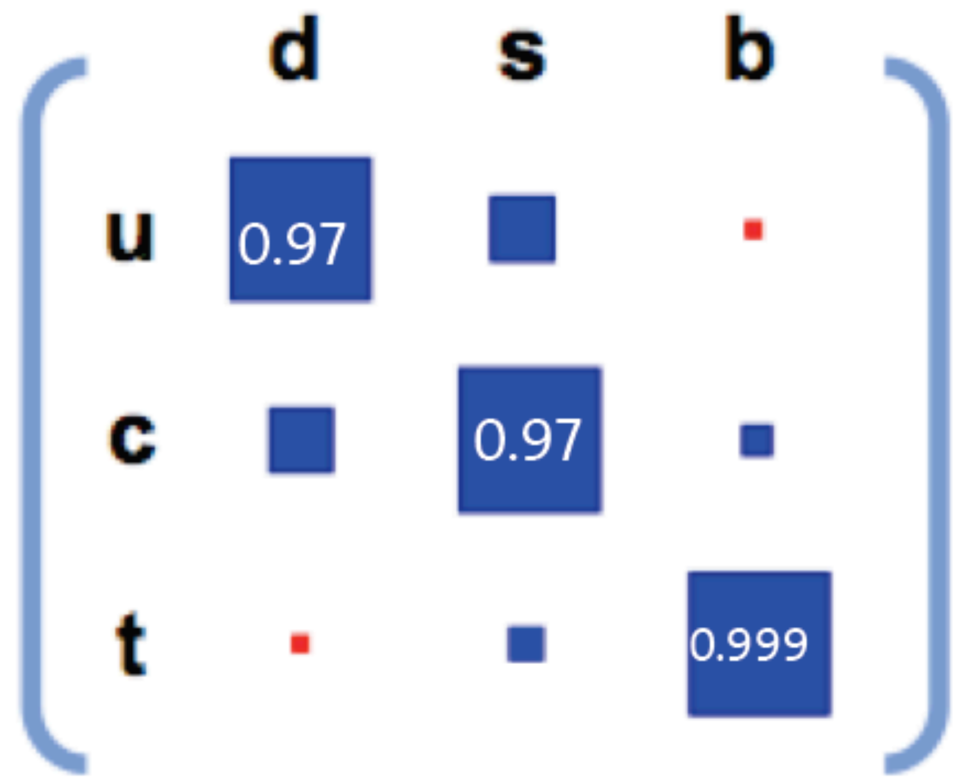
the top quark

Why is the Top quark special?

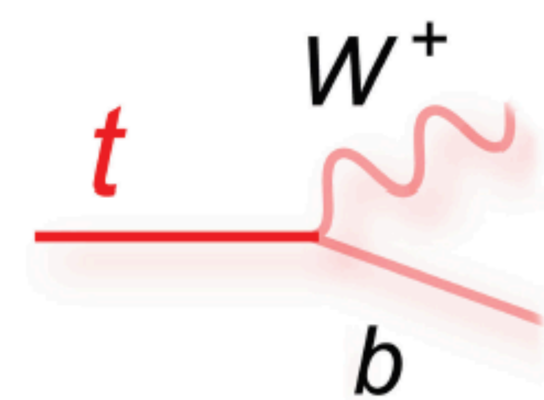
- It is the heaviest SM particle!
- Coupling to Higgs field (Yukawa couplings) are ~1
- Its live time is shorter than the characteristic hadronization time scale
  - Tops decay before fragmentation
  - Top quark decays carry information about spin correlations
- It decays exclusively in *W+b*

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$



Many searches for physics beyond the SM are connected to top physics:

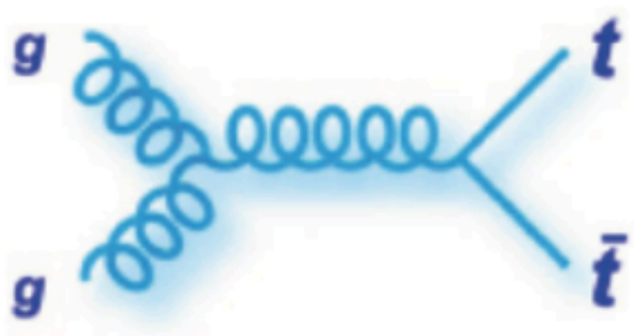
- Searches for fourth generation quarks
- SUSY searches (important background, but also final states with tops)
- *Z'* → ttbar ...

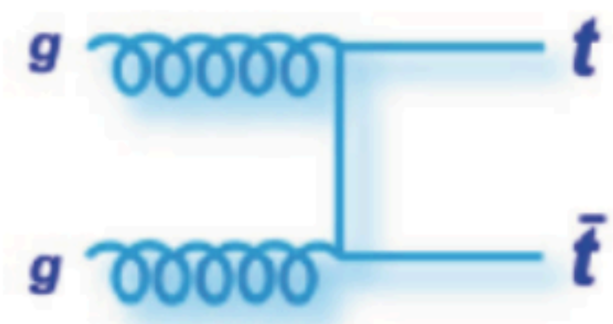


top quark production

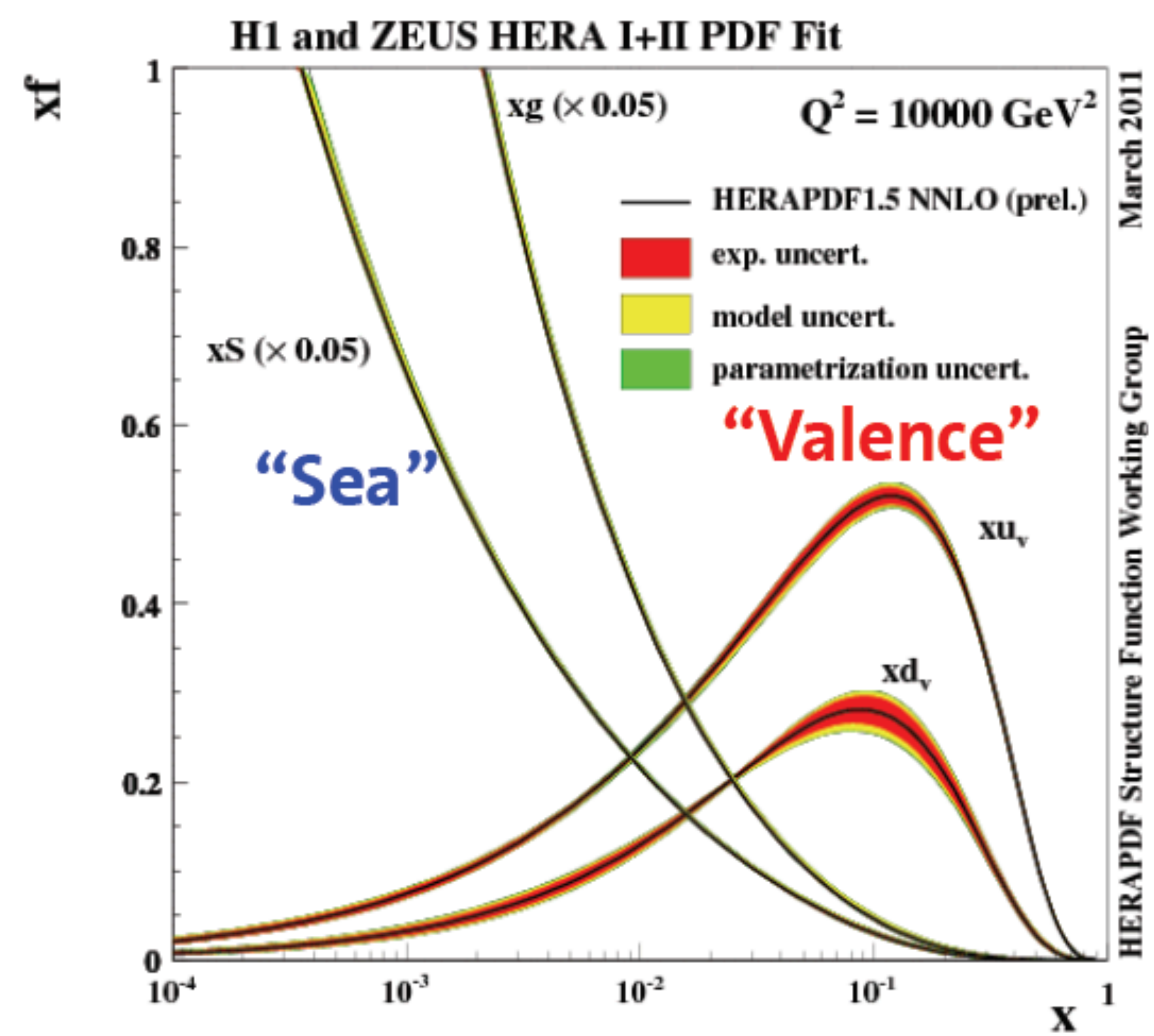
- LHC is a Top factory
- gg and qqbar are complementary at Tevatron and LHC

	LHC (7 TeV)	Tevatron (1.96 TeV)
gg	~80%	~15%
qqbar	~20%	~85%



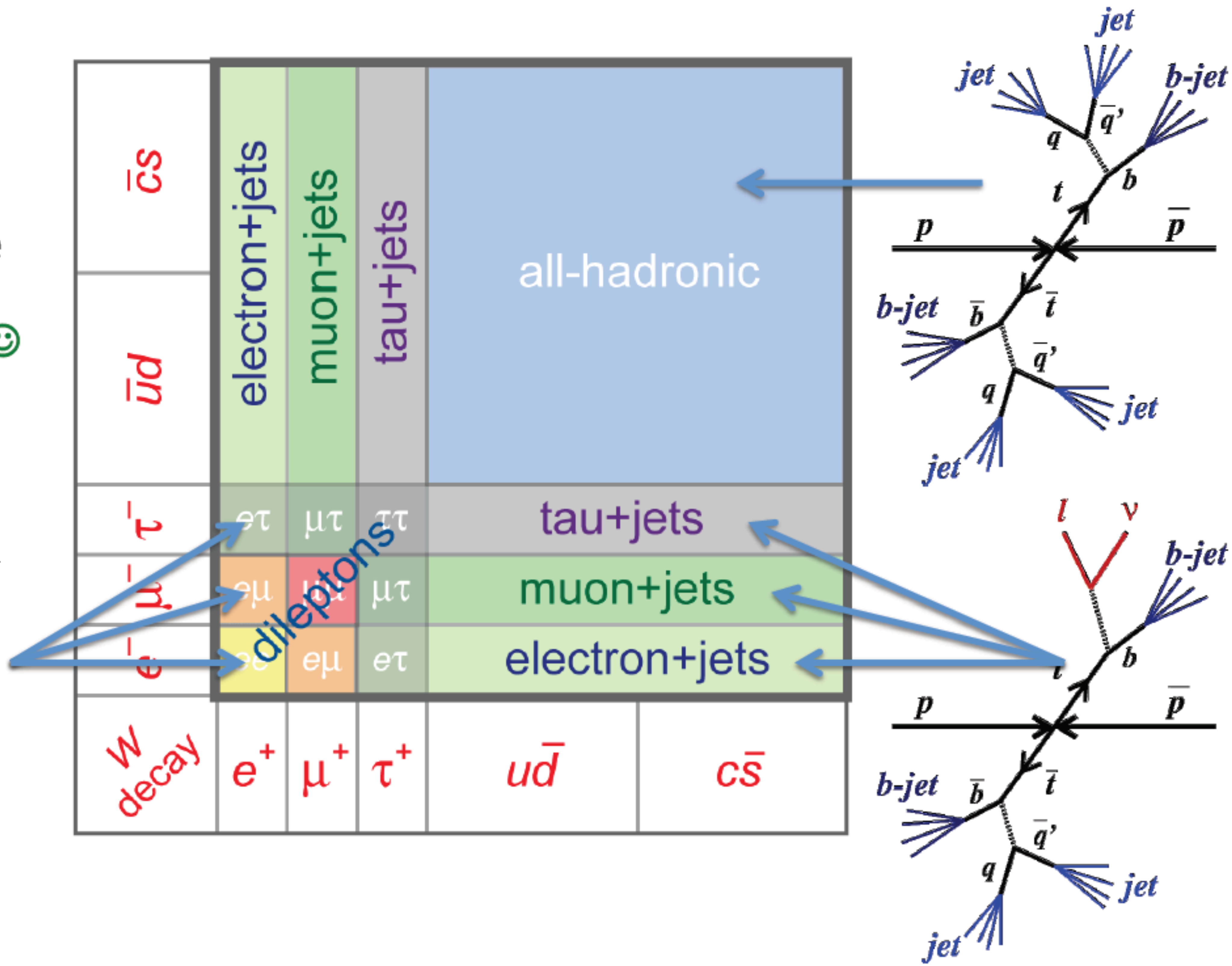


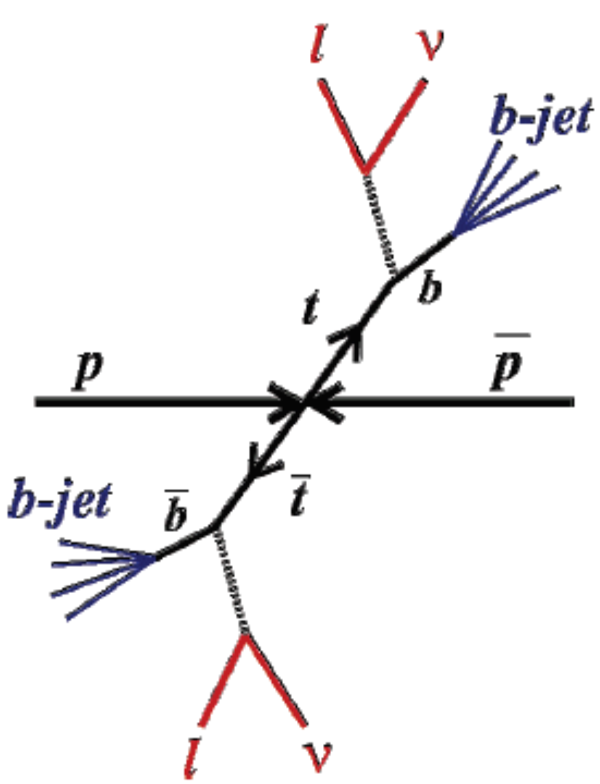
σ_{tt}[pb]	Tevatron	LHC (7 TeV)	LHC (8 TeV)
NLO	$6.68^{+0.36+0.23}_{-0.75-0.22}$	$158.1^{+19.5+6.8}_{-21.2-6.2}$	$226.2^{+27.8+9.2}_{-29.7-8.3}$
NNLO	$7.00^{+0.21+0.29}_{-0.31-0.25}$	$160.9^{+11.1+7.2}_{-11.5-6.7}$	$229.8^{+16.5+9.7}_{-16.7-9.0}$
NNLL	$7.15^{+0.21+0.30}_{-0.20-0.25}$	$162.4^{+6.7+7.3}_{-6.9-6.8}$	$231.8^{+9.6+9.8}_{-9.9-9.1}$



Decay channels with leptons:

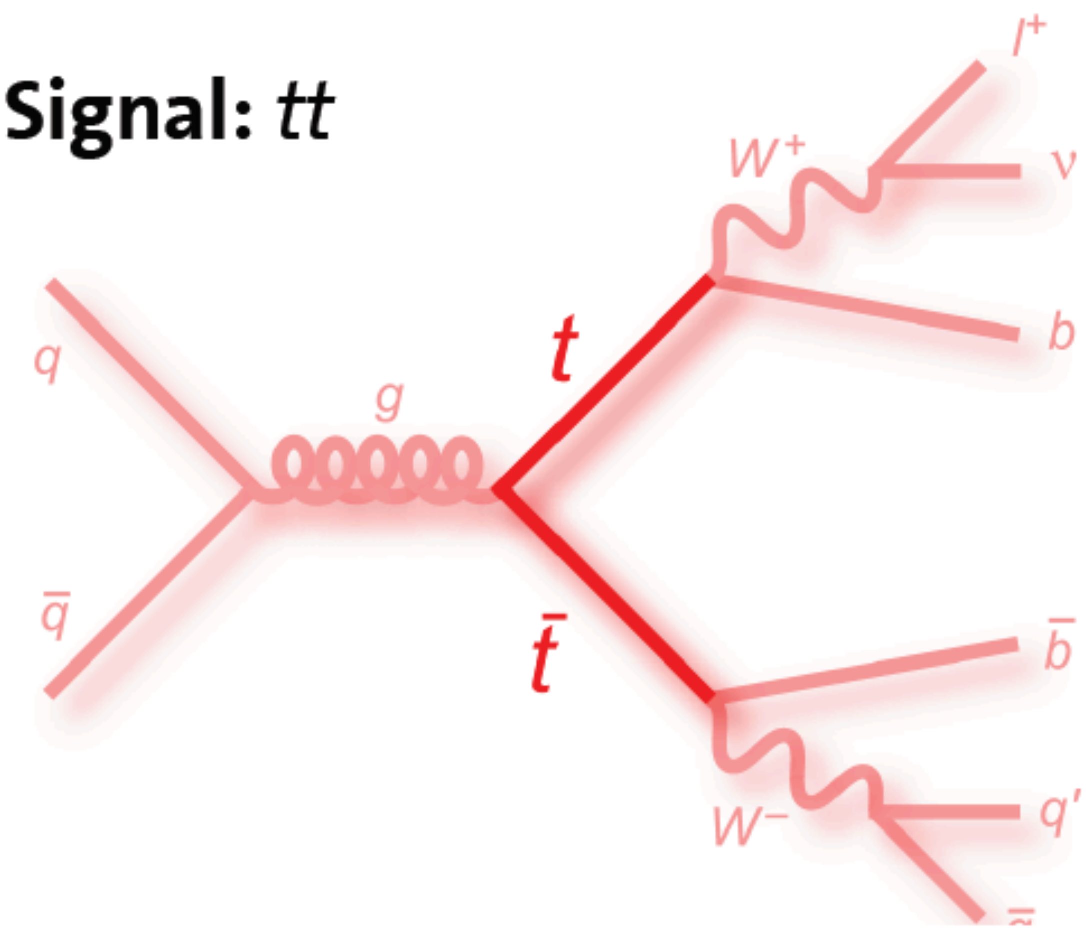
- Low branching ratio ☹
- Clean signature
- Smaller combinatorics ☺
- Smaller backgrounds ☺



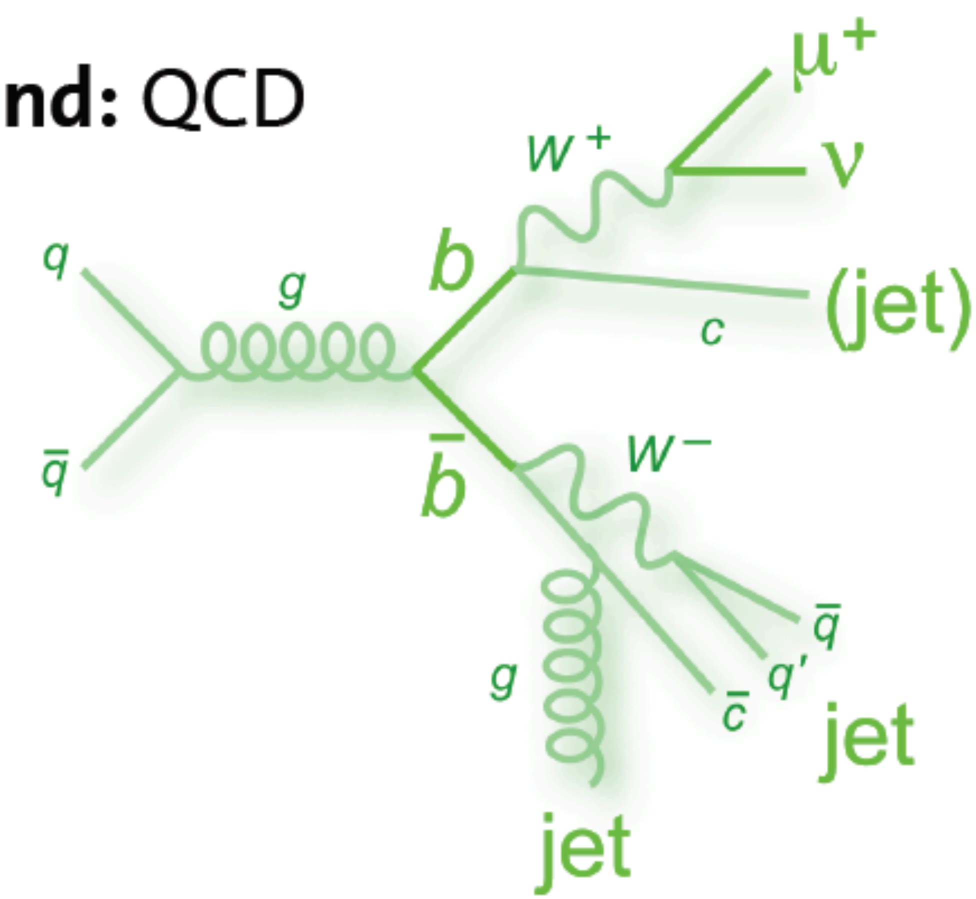


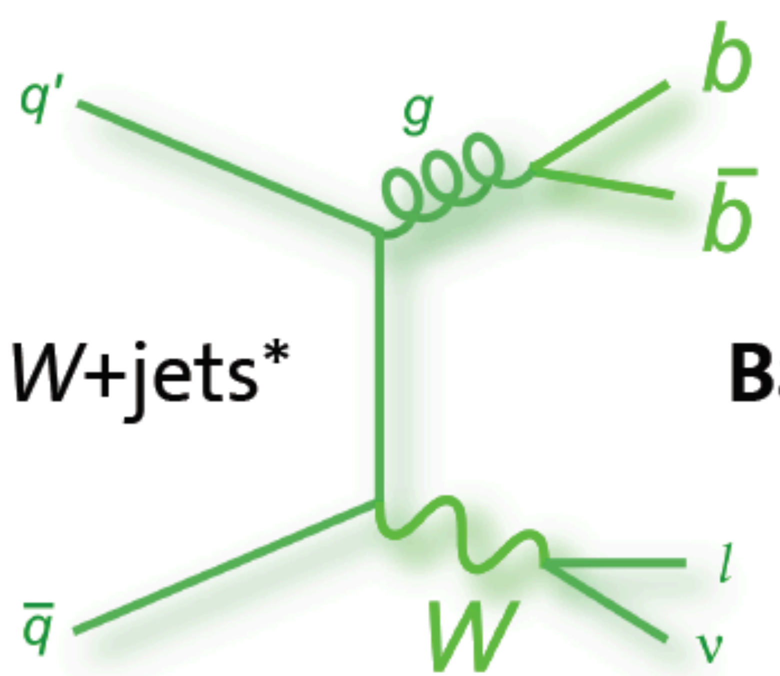
signal and background

Signal: *tt*

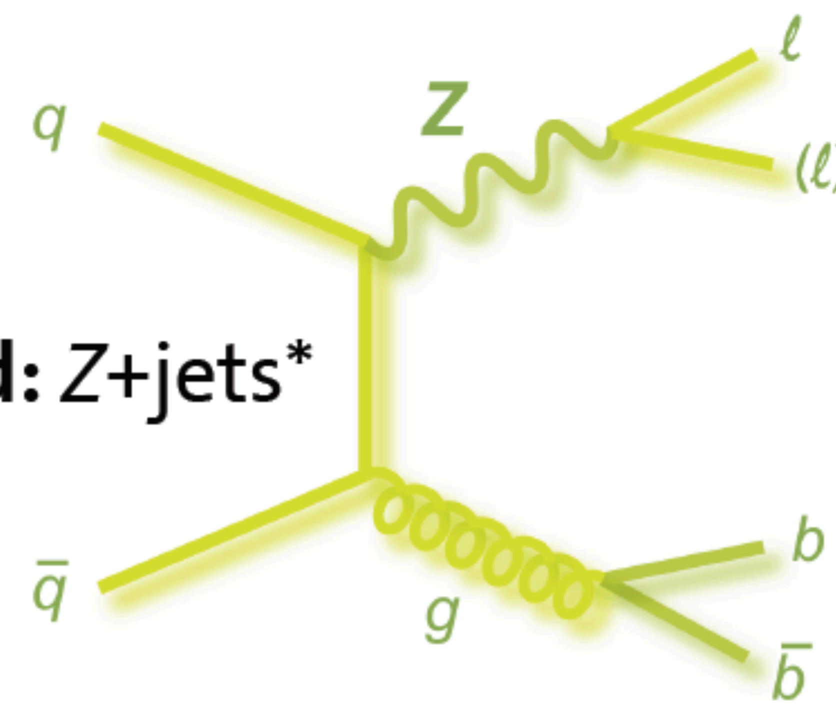


Background: QCD





Background: *W*+jets*



Background: *Z*+jets*

* not necessarily b-jets

b-jets

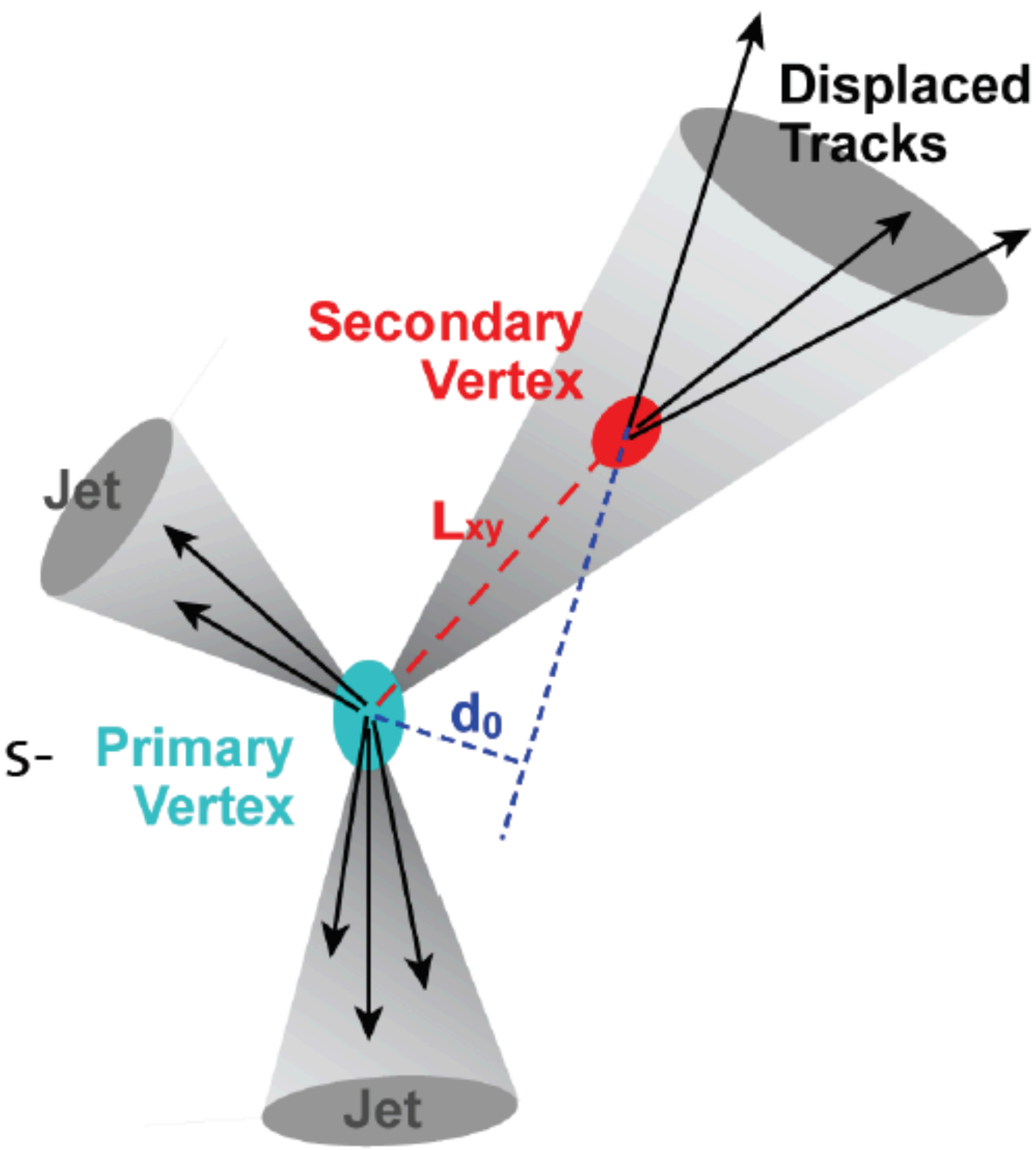
At hadron colliders (LHC) B-hadrons are produced inside of jets:

- **Their lifetime (1.5ps) and**
- **The Lorentz boost**

→ Displaced decay vertices

For example:

- Look for displaced tracks and vertices within jets **(b-jet tagging)**
- **This tutorial:** Track counting of high impact parameter tracks is used (efficiency ~50%, mis-Tag rate ~1%)



Muons are very useful objects:

- Detected in the muon chambers which are "shielded" by thick absorbers (e.g. steel return yoke)

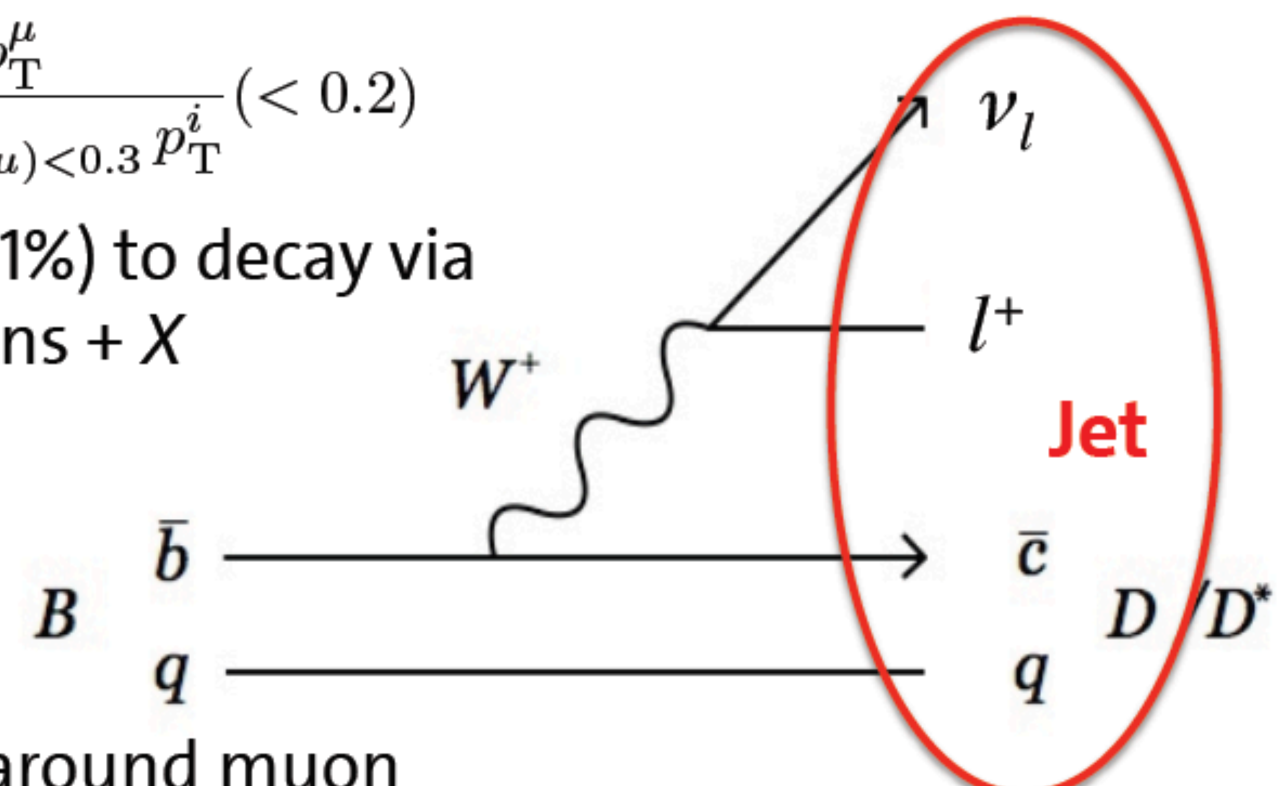
→ Other particles have negligible probability to reach this detector

→ Very clean object ID

- Processes with **isolated** muons are rare compared to QCD jet events
- Muon chambers can be used to trigger events (low p_T threshold compared to other objects)

$$\mathrm{Iso}_{\mu}^{\mathrm{rel}} = \frac{p_{\mathrm{T}}^{\mu}}{\sum_{\Delta R(i,\mu)<0.3} p_{\mathrm{T}}^{i}} (< 0.2)$$

- *B*-mesons have significant probability (~11%) to decay via the electroweak interaction in light leptons + *X*



→ muons "inside" jets

→ suppress by isolation

→ limited amount of additional activity around muon

what and how much

- **Data:** 50 pb^{-1} (~1%) of CMS datasets
- **Monte Carlo:** Set of background processes are generated with full detector simulations

How is the information stored?

- Flat ROOT trees of only most fundamental object/event properties:

→ No CMSSW dependence, no reconstruction details accessible

- **4-vectors (*px, py, pz, E*)** of leading objects
  - Jets: p_T > 30 GeV; else: p_T > 10 GeV
- +Isolation (ΔR<0.3), +charge, +b-Tag, +Jet quality depending on object
- **MC: +event weight, +IsoMuPt24 trigger bit** (for *tt* only)**, +MC truth** (parton level of semi-leptonic *tt* events)

samples with events

filename	type	#events	x-section	int. lumi.	trig. only
`data.root`	data	469384		50 pb^{-1}	yes
`ttbar.root`	sim. $t\bar{t}$ signal	36941	165 pb	50 pb^{-1}	no
`wjets.root`	sim. W plus jets background	109737	31300 pb	50 pb^{-1}	yes
`dy.root`	sim. Drell-Yan background	77729	15800 pb	50 pb^{-1}	yes
`ww.root`	sim. WW background	4580	43 pb	50 pb^{-1}	yes
`wz.root`	sim. WZ background	3367	18 pb	50 pb^{-1}	yes
`zz.root`	sim. ZZ background	2421	6 pb	50 pb^{-1}	yes
`qcd.root`	sim. QCD multijet backgr.	142	10^8 pb	50 pb^{-1}	yes

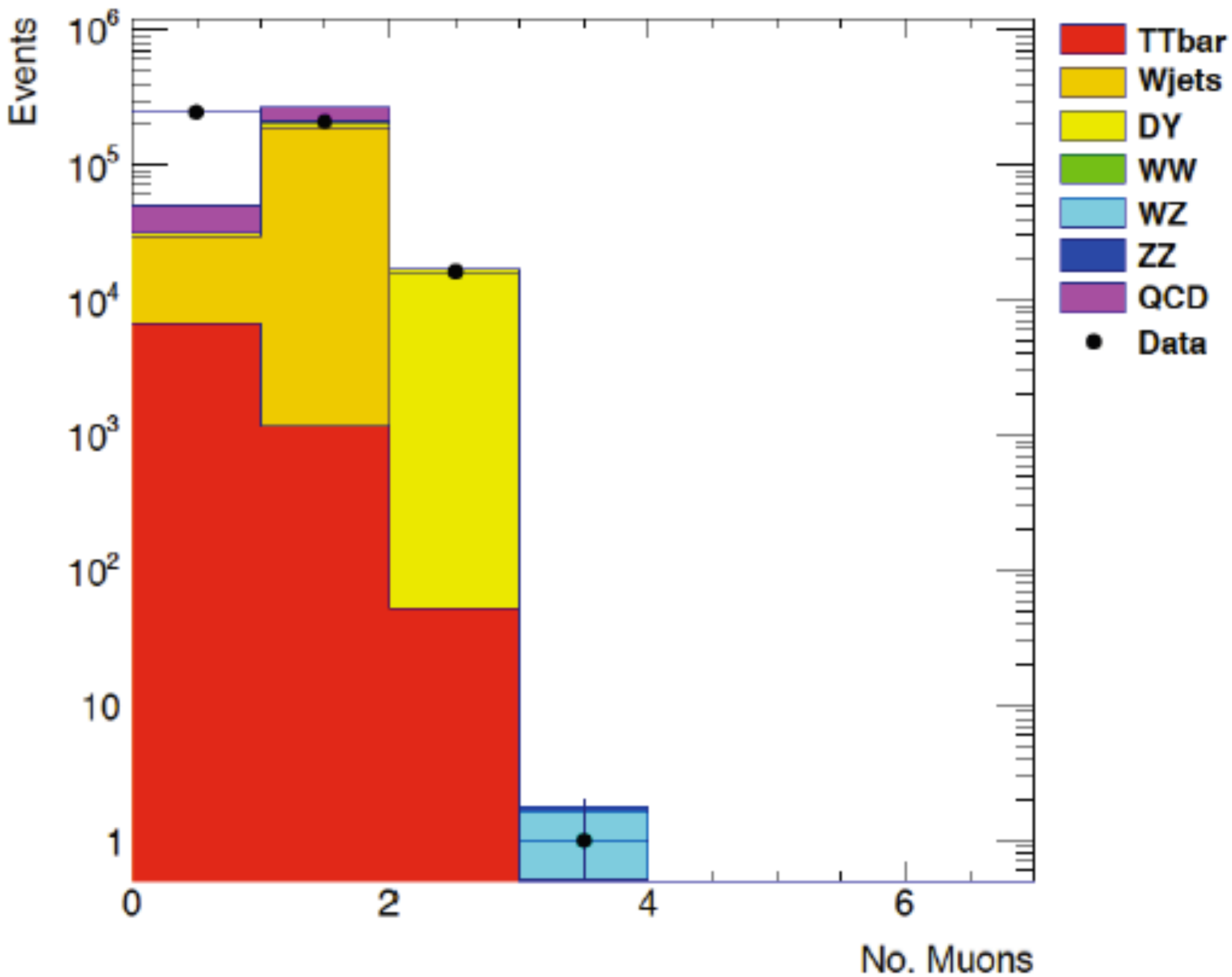
Table 1: Data and simulated Monte Carlo samples.

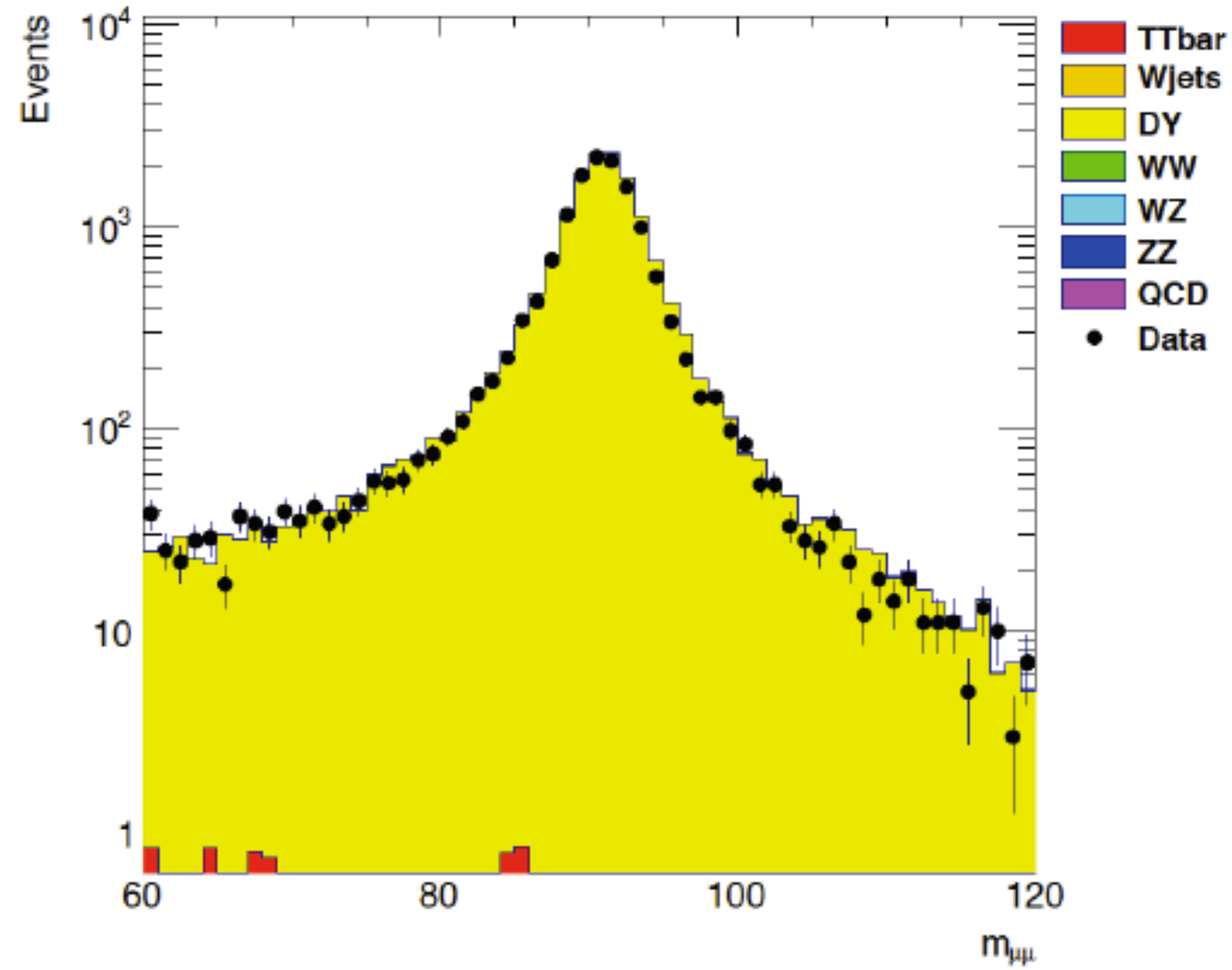
1 Warmup

The trigger for this tutorial selects events which contain one or more muons as discussed in the documentation and explanation.

- Find out how often there is more than one isolated, reconstructed muon in data (histogram of the muon multiplicity)! Where could these additional muons come from?
- Calculate the invariant mass of two muons of opposite charge (manually and/or using the ROOT functionality of adding two fourvectors)! Only use isolated muons.
- Display the invariant mass distribution of two muons in a histogram (hint: try different axis ranges)!
- Compare your results to MC simulation (display simulation and data in the same histogram). Make sure you select triggered events only for the simulated samples!

exercise 1


Events
10^6
10^5
10^4
10^3
10^2
10
1
0
2
4
6
No. Muons
TTbar
Wjets
DY
WW
WZ
ZZ
QCD
Data


Events
10^4
10^3
10^2
10
1
60
80
100
120
$m_{\mu\mu}$
TTbar
Wjets
DY
WW
WZ
ZZ
QCD
Data

how to measure a cross-section

1) Selection of signal events (here: *tt*) to ensure
   - High **trigger efficiency**: require one isolated muon! Details later.
   - High **signal acceptance = #(selected signal events)/#(all signal events)** → low statistical uncertainties
   - High **purity = #(selected signal events)/#(all selected events)** → small uncertainty from unknown backgrounds
2) From simulation: acceptance, purity and, trigger efficiency
3) Count selected data events N_{data}
4) Subtract expected background $N_{background} = N_{data}\,(1 - \text{purity})$
5) Correct for acceptance and trigger efficiency:

$$N_{signal,corr} = (N_{data} - N_{background}) \,/\, (\text{acceptance} \cdot \text{trigger efficiency})$$

6) Cross section $\sigma_{signal} = N_{signal,corr} \,/\, \text{Luminosity}_{integrated}$

trigger efficiency

- Event rate of a process with cross section $\sigma \rightarrow f = \sigma \cdot L_{\text{instantaneous}}$
- With $L_{\text{instantaneous}} = 10^{34}\ \text{s}^{-1}$ and $\sigma_{\text{total}} = 10^{8}$ pb $\rightarrow$ $f = 10^{6}\ \text{s}^{-1}$
- Typical recorded rate ~100 $\text{s}^{-1} \rightarrow$ online preselection (trigger)
- **For the presented example: trigger on isolated muon with p_T > 24 GeV**
- Online-offline differences $\rightarrow$ Determine trigger efficiency (e.g. from MC)

Trigger efficiency = #(triggered and selected)/#(selected)

  - Selection not necessarily signal selection (as long as independent) $\rightarrow$ e.g. isolated muon in p_T interval $\rightarrow$ **"turn on" curve**
- To evaluate statistical uncertainties correctly:
  - Don't use error propagation for ratio of two quantities; use binomial errors. **Better:** confidence intervals (otherwise error for efficiency of 1 is zero by definition)

```
TGraphAsymmErrors* g_eff = new TGraphAsymmErrors (h_SelTrig, h_Sel);
g_eff->GetXaxis()->SetTitle("Muon p_{T} [GeV]");
g_eff->GetYaxis()->SetTitle("Trigger efficiency");
g_eff->UseCurrentStyle();
g_eff->Draw("AP");
c2->Print("Trigger_eff.pdf");
```

2 Properties of top quark events

In this exercise we take the first steps towards a real measurement using top quark events. We need to understand how we can efficiently select top quark events and reject events without top quarks (background rejection) at the same time.

- Starting from the requirement of at least one isolated muon, compare several other distributions of event variables for simulated signal ($t\bar{t}$ events) and background.
- Try to find variables which are especially sensitive to separate signal from background (jet multiplicity, transverse momenta of jets and leptons, lepton isolation, b-tagging, missing transverse energy, angular distributions). Fill all these distributions into histograms and compare between signal, background and data.
- Apply cuts on these variables to enrich the signal over background. Try to optimize the signal over background ratio and estimate the purity that can be achieved (based on simulation only).
- Apply your selection cuts also on data. Compare the selection efficiency between data and simulation.

schedule

- we continue with exercise part 2 today
- Valentina and Andrey will be around to help you
- I have another meeting between 15:00 and 17:00 today
- Andrey and Valentina have a meeting between 16:00 and 17:00 (so you will be alone for one hour)
- I will be back at 17:00 and present a solution and explanation for the exercise part 2 (will be recorded in case this is too late for you)
- we will continue on Tuesday with exercise part 3
- you can ask questions also on Tuesday of course