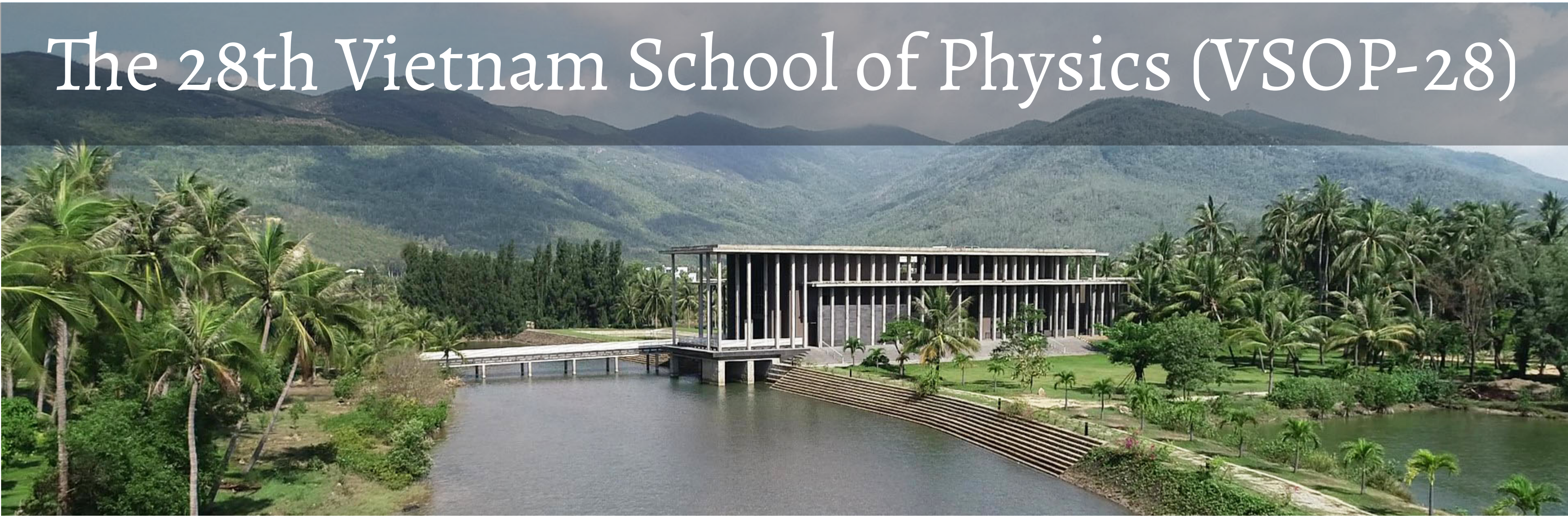


Experimental methods for physics at the LHC

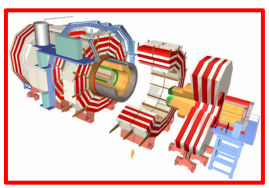
Sourabh Dube

IISER PUNE

July 24, 2022 to
August 5, 2022

Lecture 4:
Introduction to Analysis

A word on 'Monte Carlo' samples



Data

Generation

For a given physics process, obtain the 4-vectors of all outgoing particles

Simulation

Simulate what happens when these outgoing particles interact with the material of the detector

Digitization

Given the material interactions, what digital signals will be generated in the detector (like actual data)

Reconstruction

Use the digital signals to reconstruct what particles were seen in the detector.

Dedicated lectures on this later in the school....

	E	Px	Py	Pz	other information
N_e					
Ele1					
Ele2					
N_{mu}					
Mu1					
Mu2					
Mu3					
N_{jets}					
Jet1					
Jet2					
Jet3					
Jet4					
.					
.					
.					

1 event or 1 collision

We have many such collisions

E Px Py Pz other information

N_e

Ele1

Ele2

N_{mu}

Mu1

Mu2

Mu3

N_{jets}

Jet1

Jet2

Jet3

Jet4

.

.

.

Event counts

$$N = L \, \sigma \, B \, A \, \varepsilon$$

N is number of events (of a given process)

L is integrated luminosity (amount of data)

σ is the cross section of that process

B is branching fraction

A is acceptance

ε is efficiency

Event counts

$$N = L\ \sigma\ B\ A\ \varepsilon$$

N is number of events (Z to ll at the LHC)

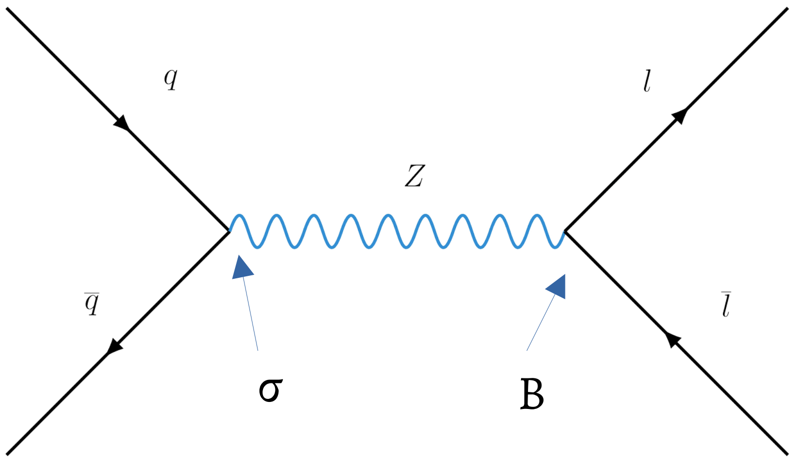
L is integrated luminosity (amount of data)

σ is the cross section of that process (pp→ Z)

B is branching fraction (Z→ ll)

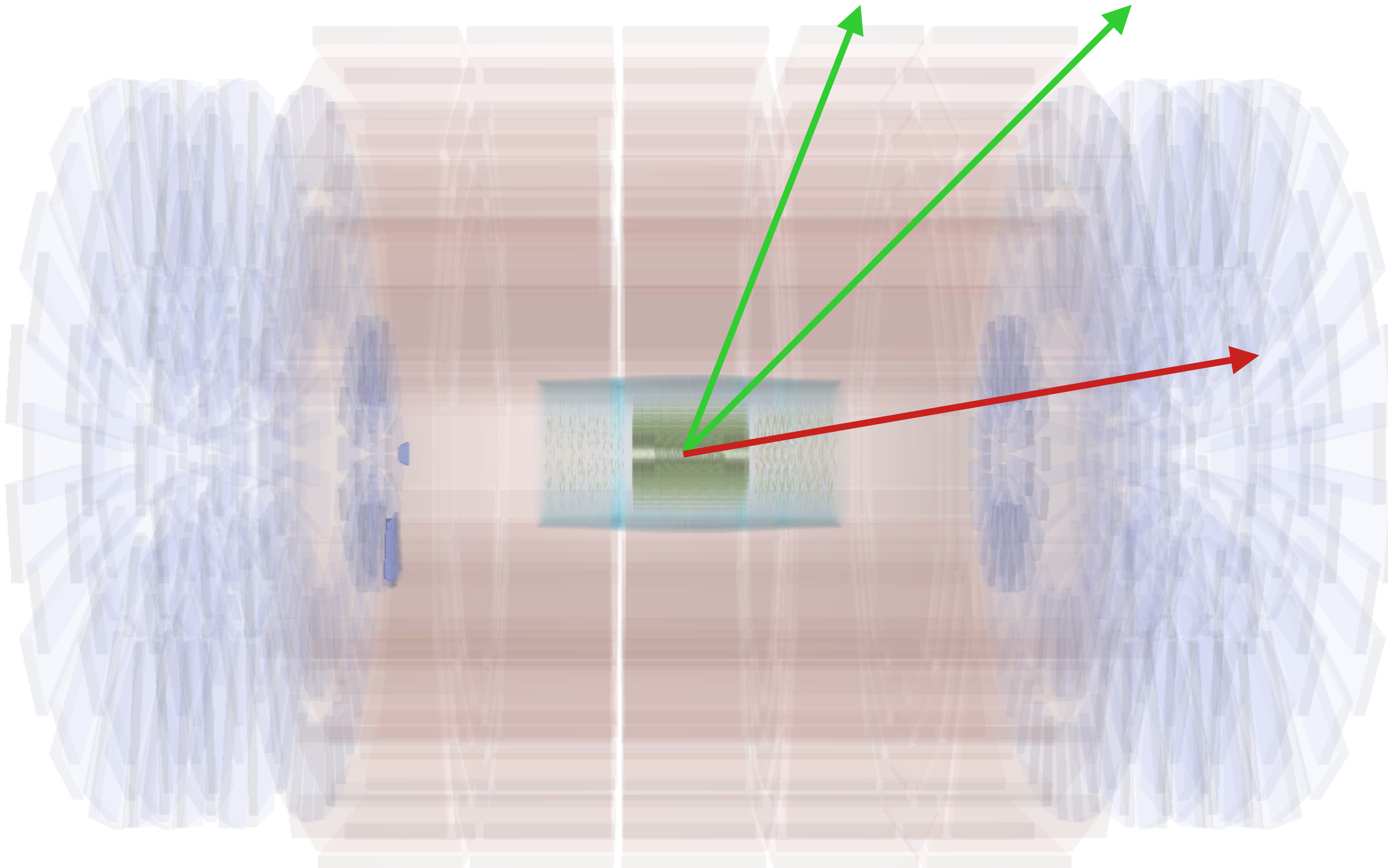
A is acceptance (Z→ll events within detector coverage)

ε is efficiency (of those how many are identified)



$$\text{Acceptance} = \frac{\text{Number that pass `selections'}}{\text{Total number}}$$

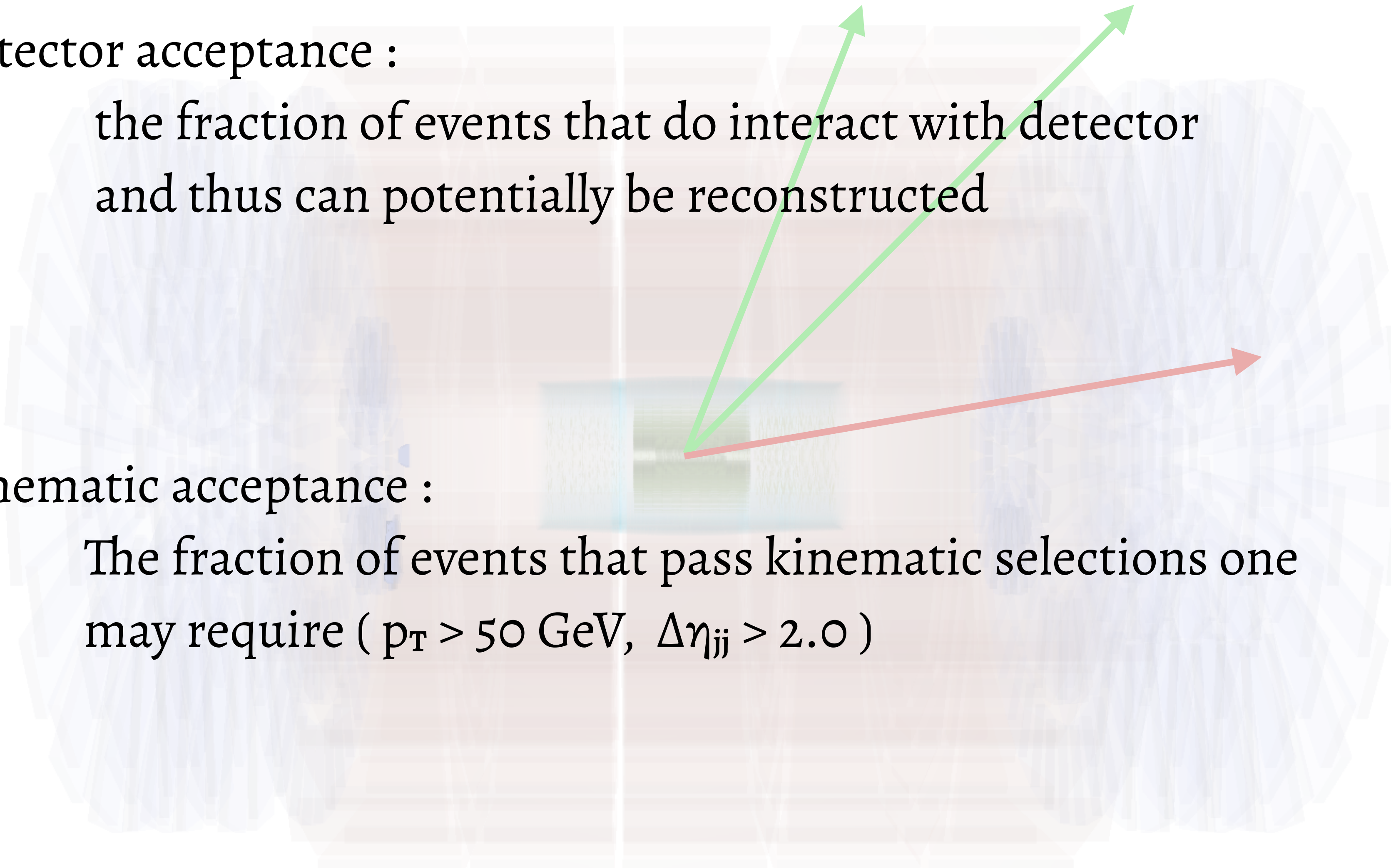
$$\text{Acceptance} = \frac{\text{Number that pass 'selections'}}{\text{Total number}}$$



$$\text{Acceptance} = \frac{\text{Number that pass 'selections'}}{\text{Total number}}$$

Detector acceptance :

the fraction of events that do interact with detector and thus can potentially be reconstructed



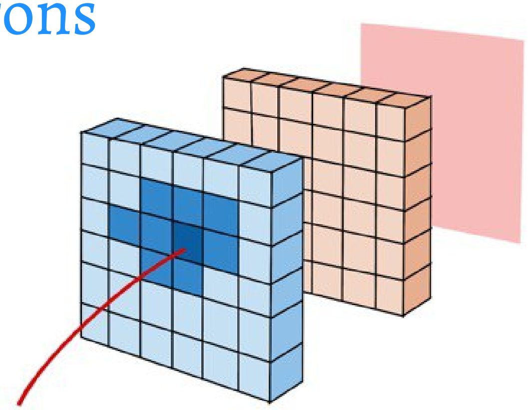
Kinematic acceptance :

The fraction of events that pass kinematic selections one may require ($p_T > 50$ GeV, $\Delta\eta_{jj} > 2.0$)

$$\text{Efficiency} = \frac{\text{Number that are correctly identified}}{\text{Number that can be identified}}$$

$$\text{Efficiency} = \frac{\text{Number that are correctly identified}}{\text{Number that can be identified}}$$

Electrons



A track that matches a calorimeter cluster — How good should the match be?

Energy/Momentum close to 1 — How close to 1?

Negligible HCAL deposit — How negligible?

Shape of shower as expected — How close to expected?

Isolated — How isolated?

Acceptance x Efficiency

Acceptance x Efficiency=

$$\frac{\text{Number that pass selections and can be detected}}{\text{Total number produced}} \times \frac{\text{Number that is detected}}{\text{Number that pass selections and can be detected}}$$

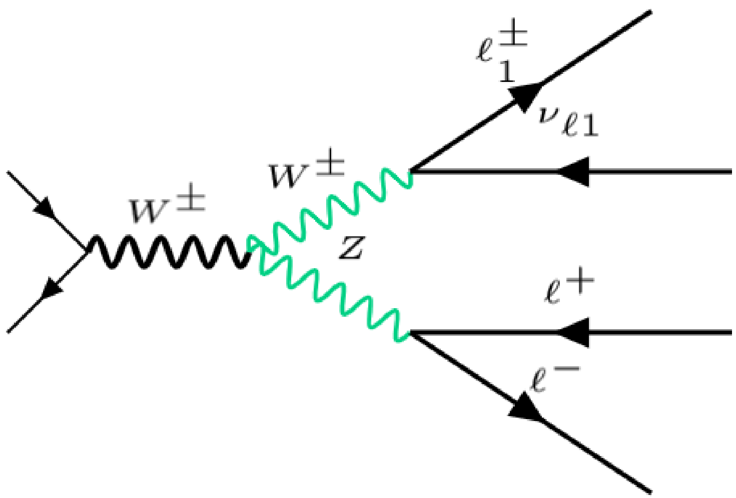
$$= \frac{\text{Number that is detected}}{\text{Total number produced}}$$

$$N = L\ \sigma\ B\ A\ \varepsilon$$

Number we will see = amount of data × rate of production × whether we reconstruct/identify the events.

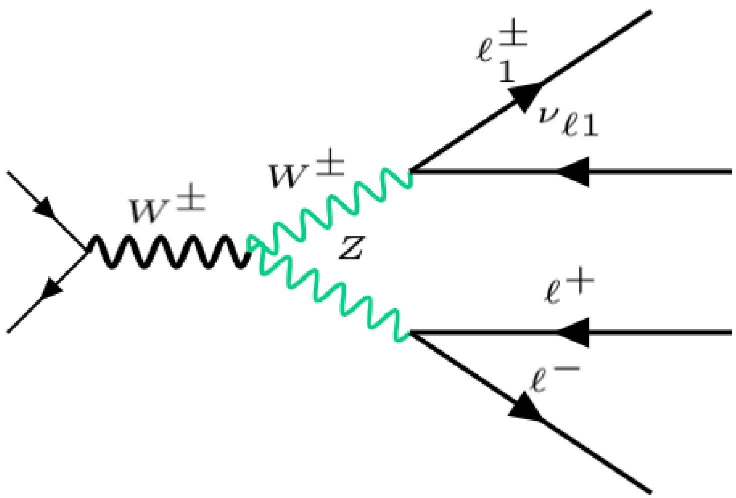
An example of acceptance

Consider looking for collisions which are from WZ production



An example of acceptance

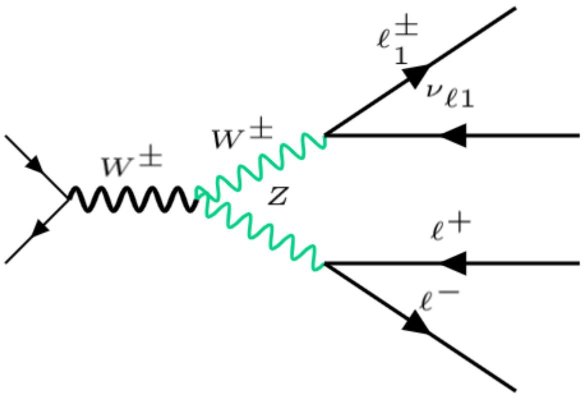
Consider looking for collisions which are from WZ production

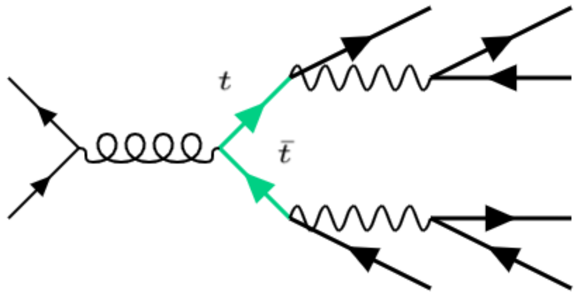


All events with 3 muons with pT > 20 GeV each

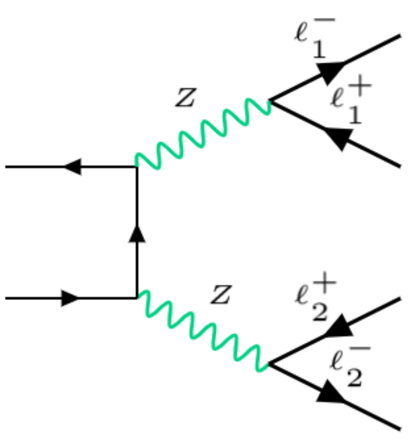
An example of acceptance

Consider looking for collisions which are from WZ production

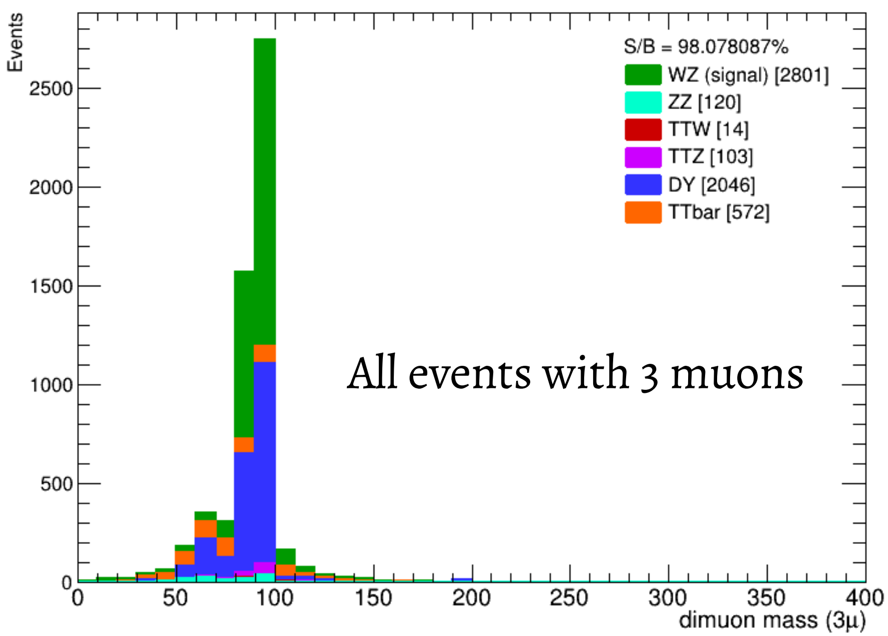




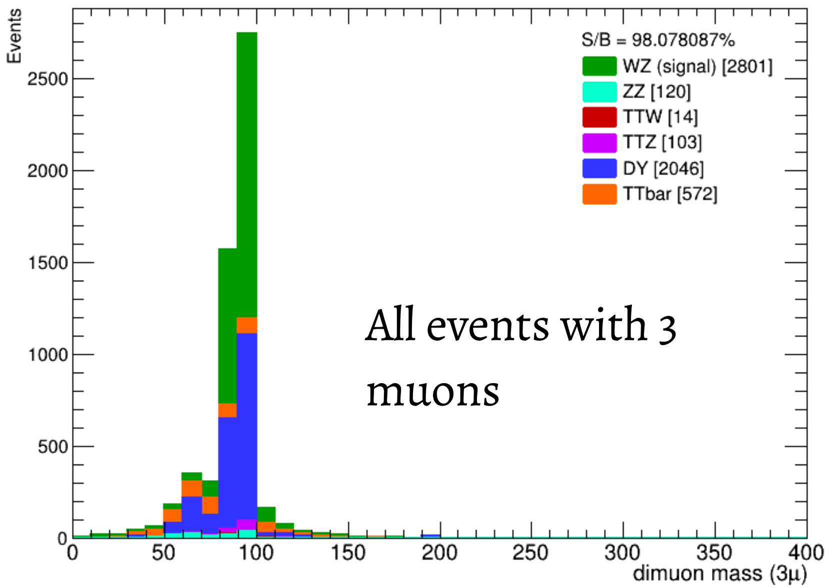
$t\bar{t}$

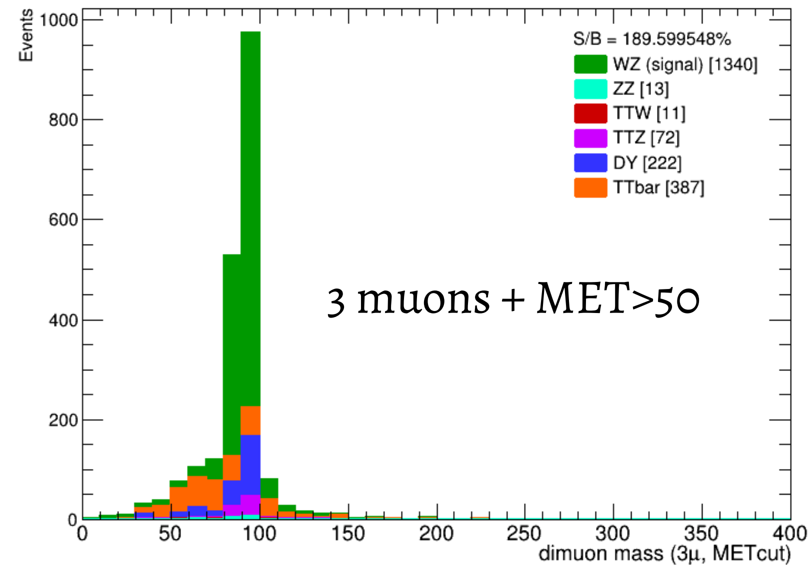


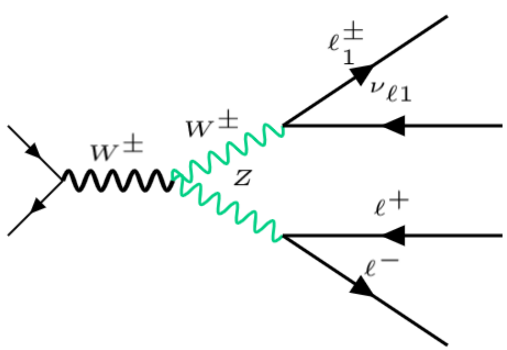
ZZ

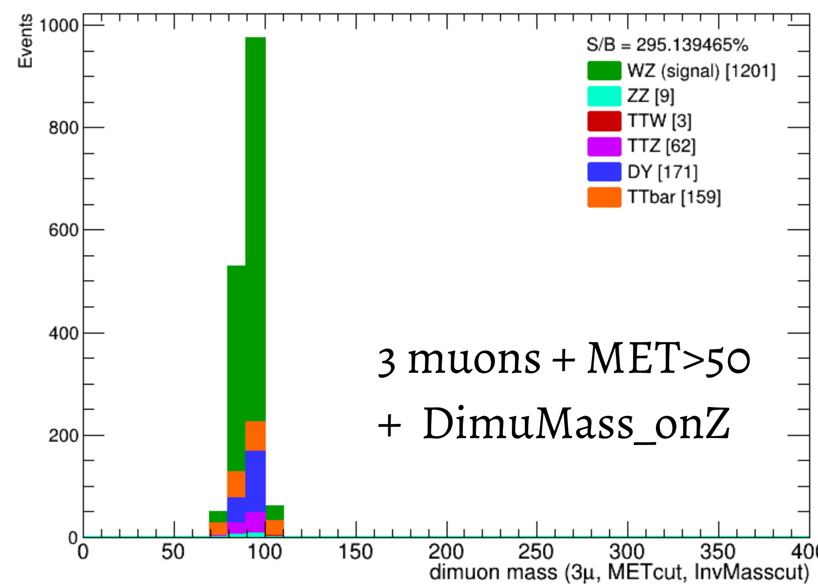


An example of acceptance





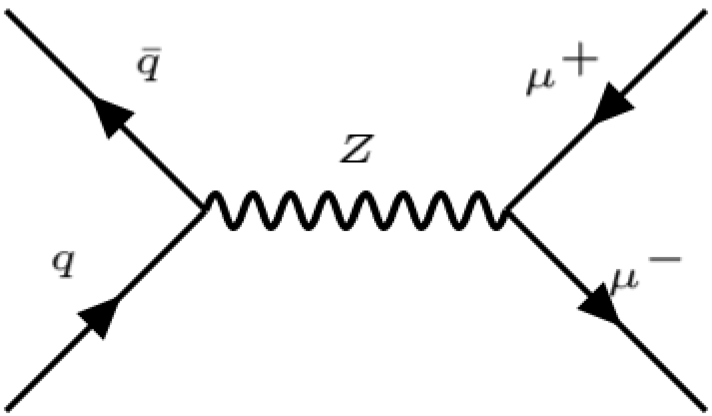




Measure cross section

We know Z exists. We know one possible decay is $Z \rightarrow \mu\mu$

Is the cross section calculated by human beings the correct value?



Measure cross section

We know Z exists. We know one possible decay is $Z \to \mu\mu$

Is the cross section calculated by human beings the correct value?

Let us nail down some particulars.

What quality requirements should we make to select muons?
Say we require some selections (q) for muons.
In addition we also require some kinematical conditions (k)

Measure Z cross section

$$N = L\ \sigma\ B\ A\ \varepsilon$$

$$\sigma B = N / (L\ A\ \varepsilon)$$

Fiducial cross section

$(\sigma B)_{fid} = N / L$

N is the observed events in data passing selections.

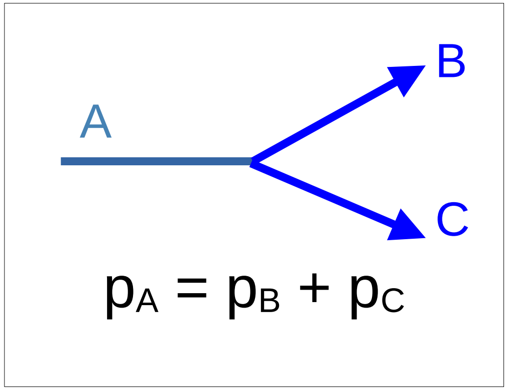
L is the integrated luminosity of our dataset.

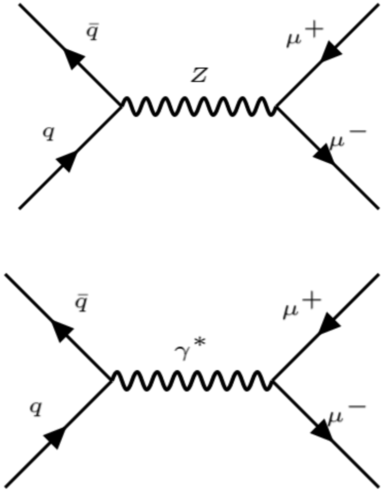
A is the acceptance of our selections, depends on k

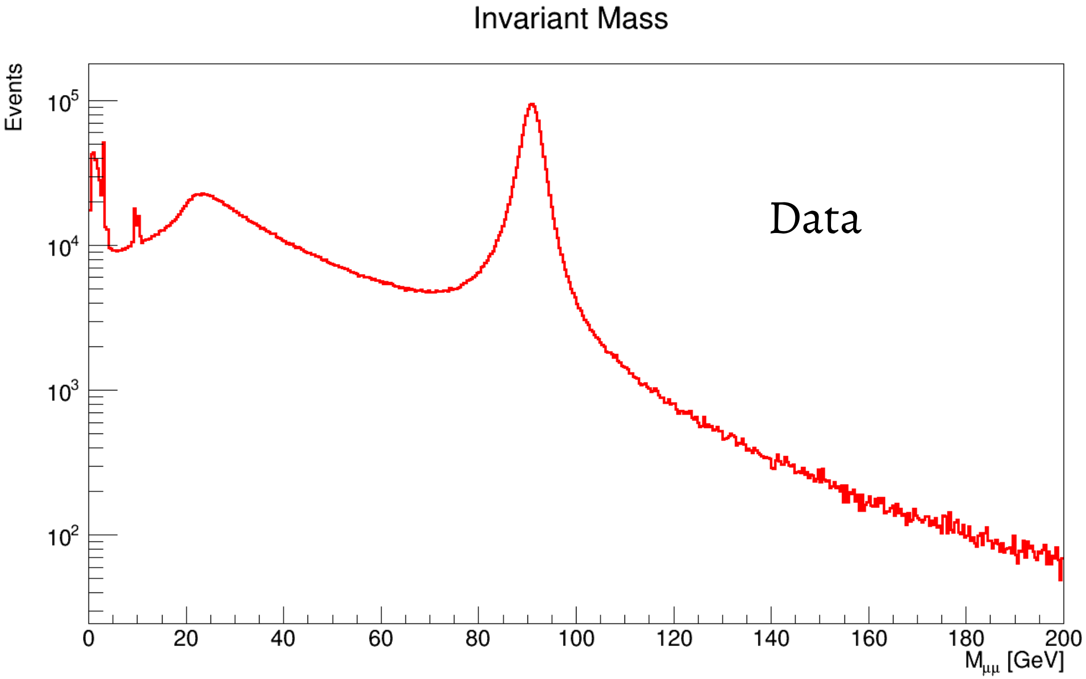
ε is the efficiency of our selections, depends on q

Measure Z cross section

$$\sigma B = N / (L\,A\;\varepsilon)$$

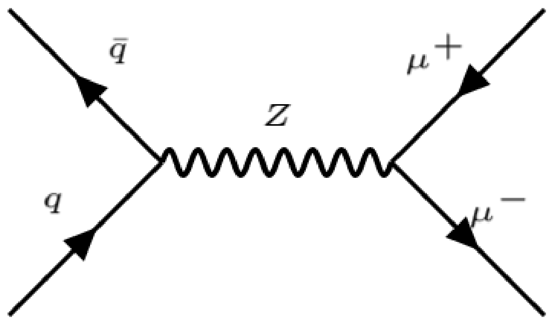


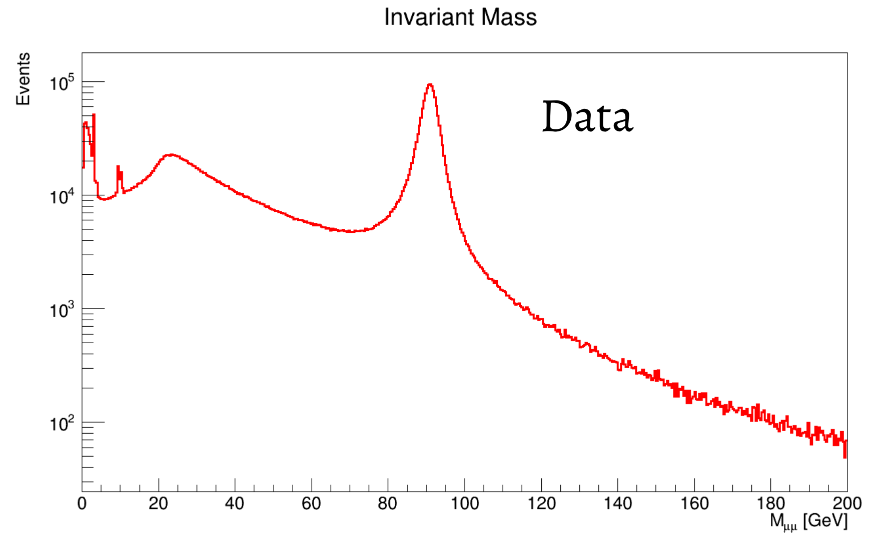




(Two muons, each with p_T>20 GeV, $|\eta|$<2.4 *etc.*)

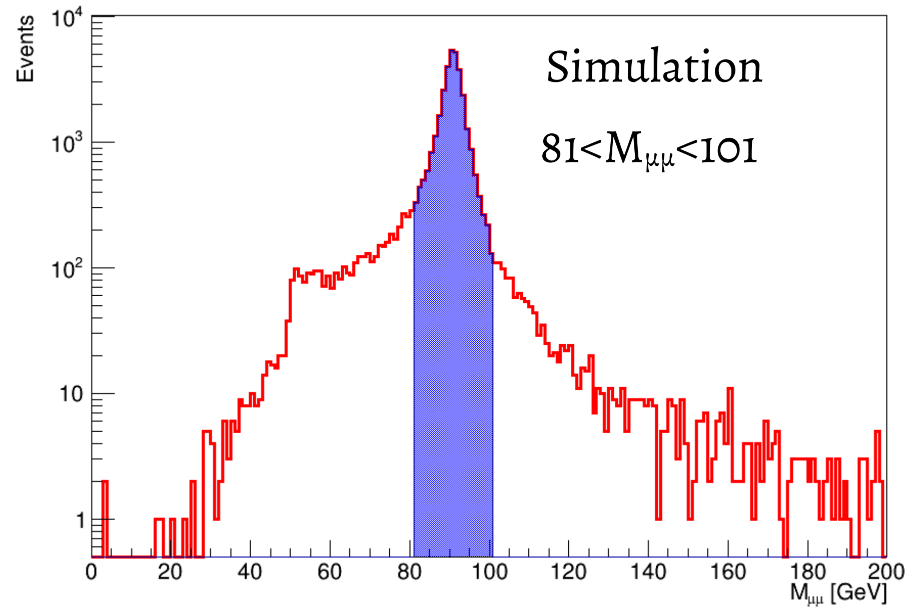
How do we determine A?





1. Produce simulation of Z

2. In simulation, implement all the selections (k) that we aim to do in the data.



$$\frac{\text{Number that pass selections and can be detected}}{\text{Total number produced}}$$

How do we determine ε ?

$$\frac{\text{Number that is detected}}{\text{Number that pass selections and can be detected}}$$

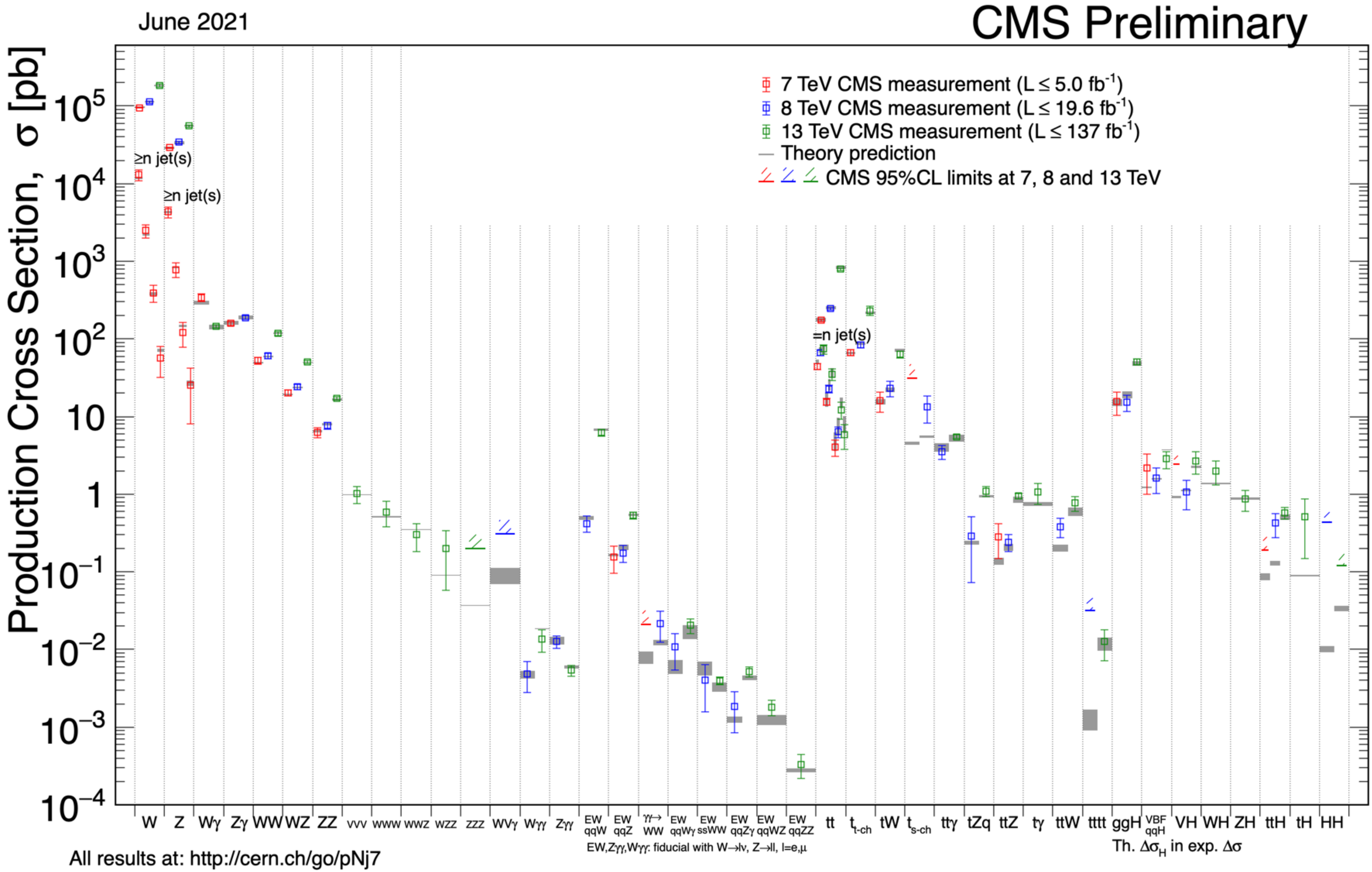
We calculate efficiency (effect of q) for one muon, ε_{μ}

Then the total efficiency of the event is $\varepsilon = \varepsilon_{\mu} \cdot \varepsilon_{\mu}$

Here q includes all the reconstruction and identification requirements.

Stick all numbers in…

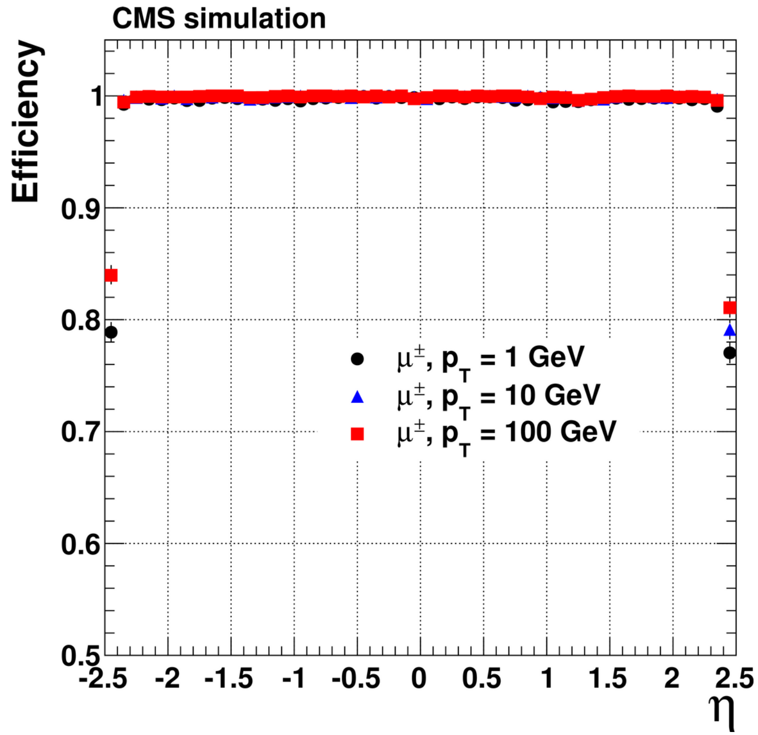
$$\sigma B = N / (L\, A\, \varepsilon)$$

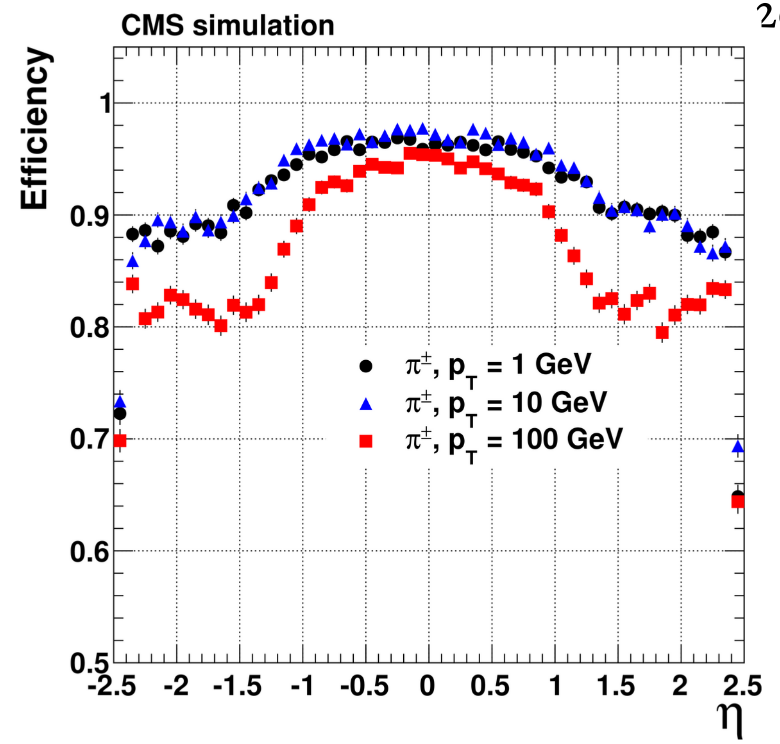

June 2021
CMS Preliminary
Production Cross Section, σ [pb]
7 TeV CMS measurement (L ≤ 5.0 fb⁻¹)
8 TeV CMS measurement (L ≤ 19.6 fb⁻¹)
13 TeV CMS measurement (L ≤ 137 fb⁻¹)
Theory prediction
CMS 95%CL limits at 7, 8 and 13 TeV
≥n jet(s)
≥n jet(s)
=n jet(s)
W Z Wγ Zγ WW WZ ZZ VVV WWW WWZ WZZ ZZZ WVγ Wγγ Zγγ EW qqW EW qqZ γγ→WW EW qqWγ EW ssWW EW qqZγ EW qqWZ EW qqZZ tt t t-ch tW t s-ch ttγ tZq ttZ tγ ttW tttt ggH VBF qqH VH WH ZH ttH tH HH
EW,Zγγ,Wγγ: fiducial with W→lν, Z→ll, l=e,μ
Th. Δσ_H in exp. Δσ
All results at: http://cern.ch/go/pNj7

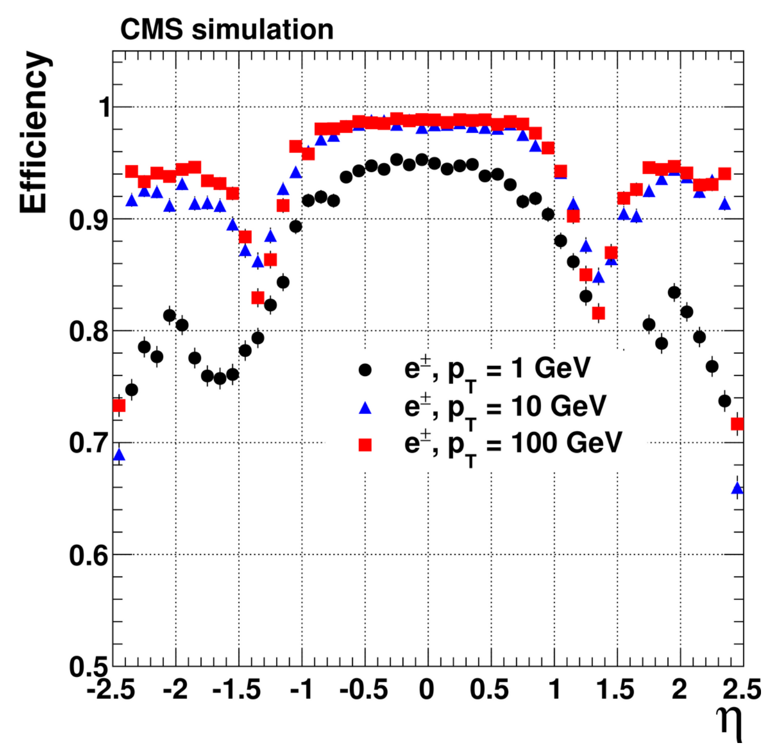
Let us now see some performance plots to get a sense of the numbers

J. Instrum. 9 (2014) P10009

Tracking Efficiency

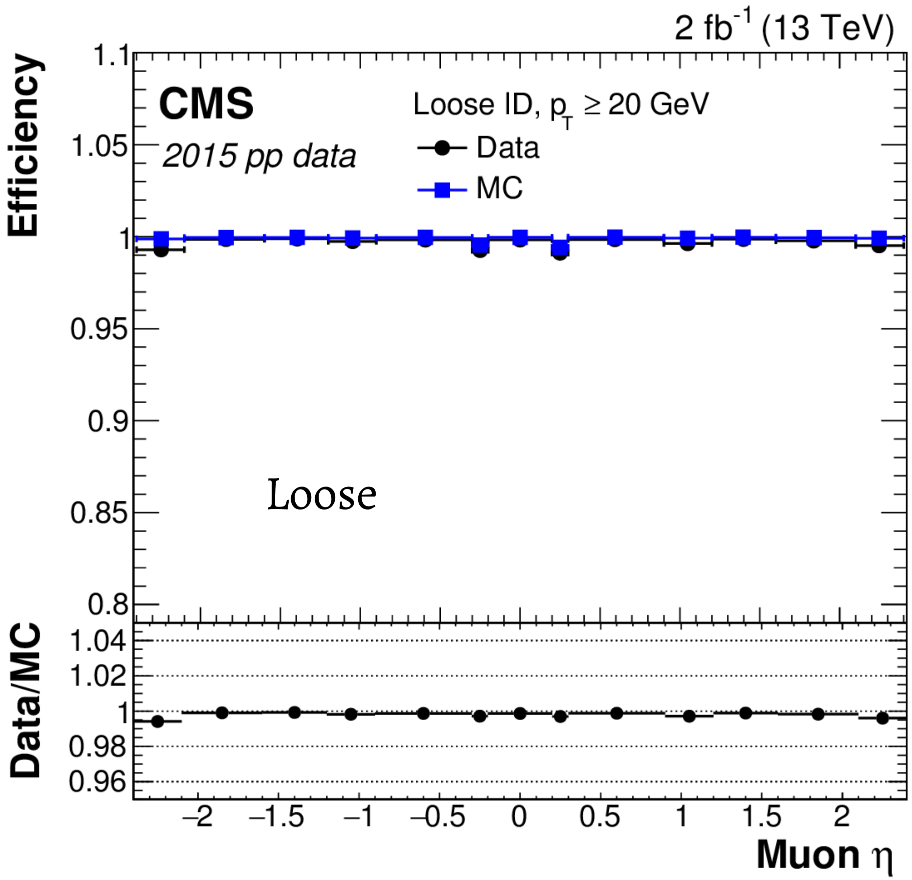


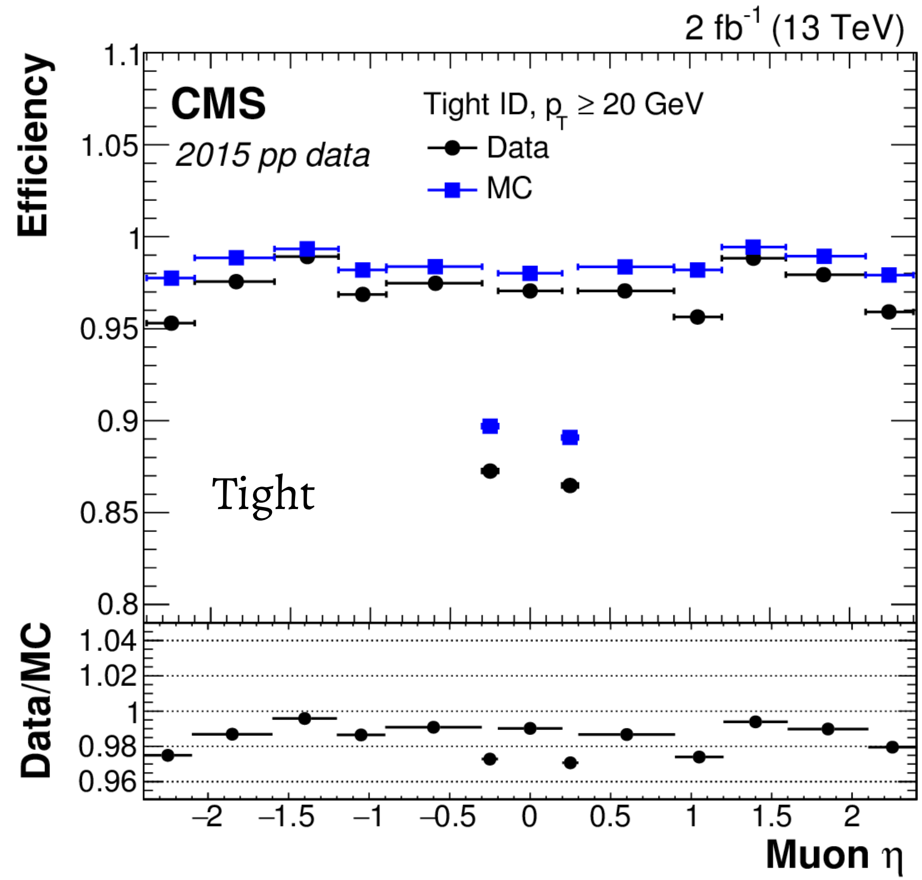




Muon efficiency

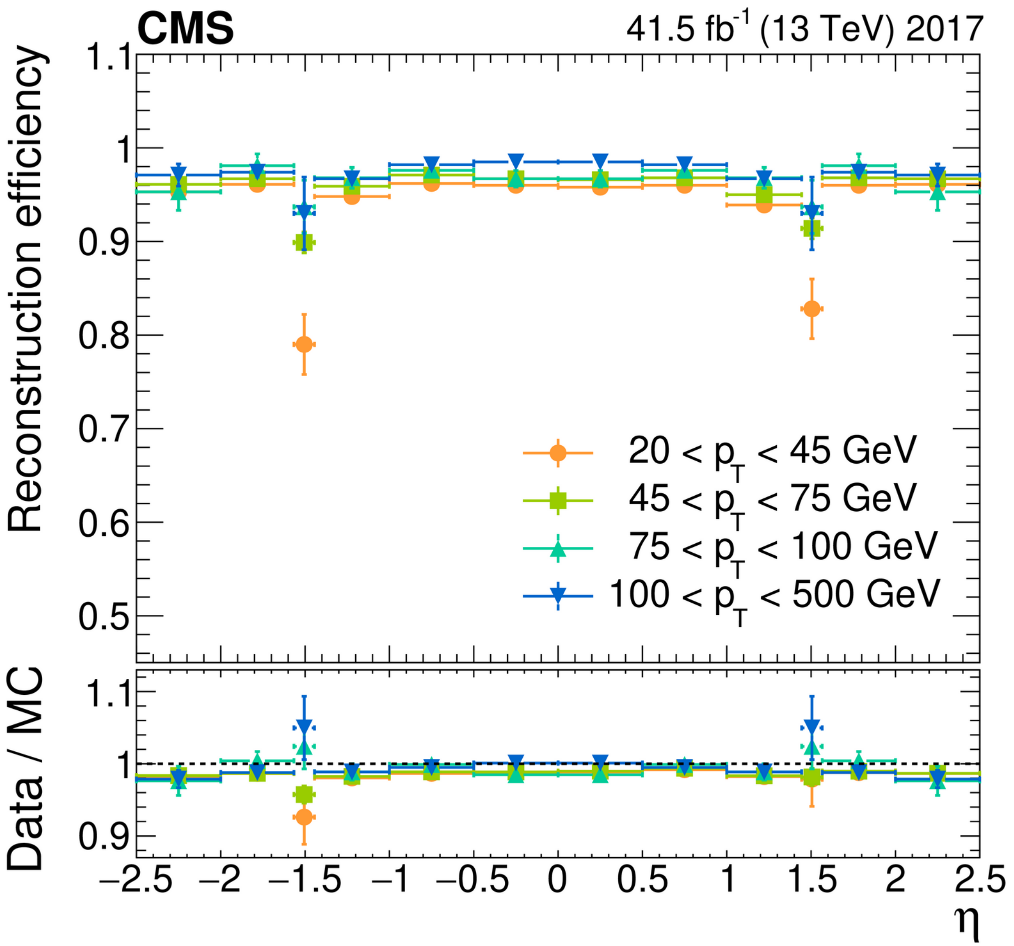
JINST 13 (2018) P06015

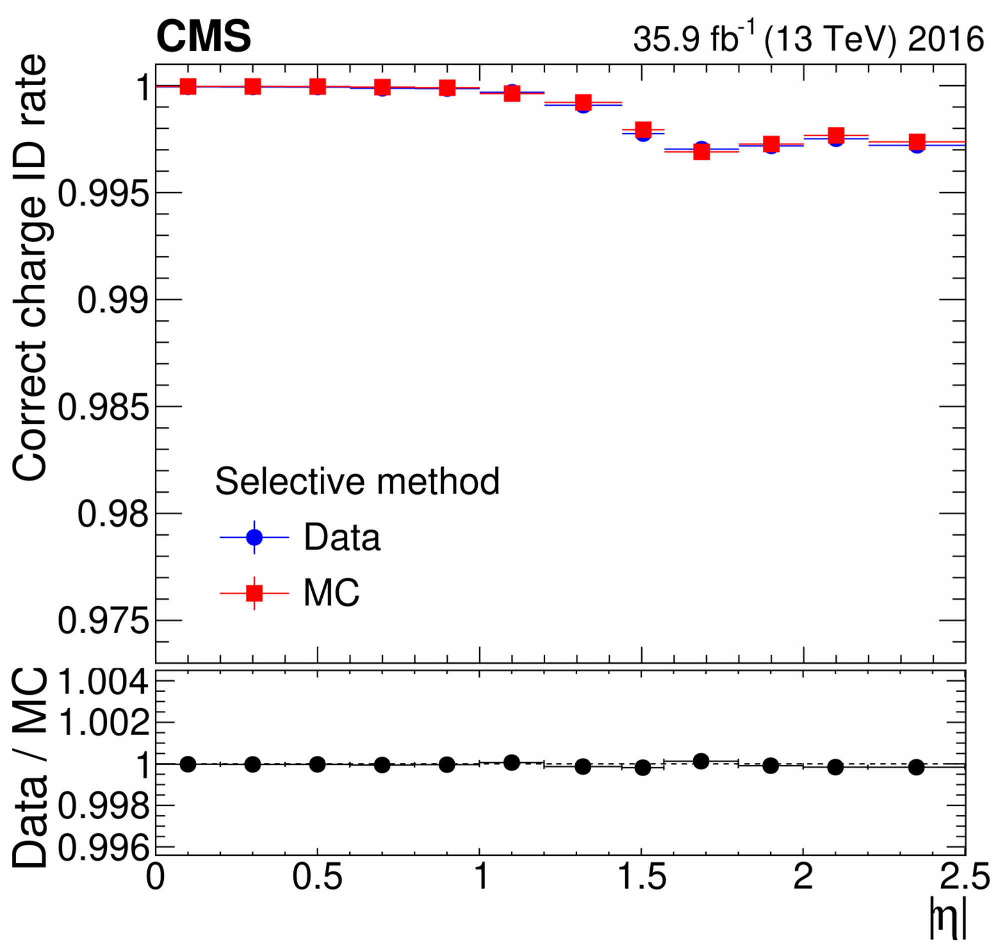




Electron efficiency

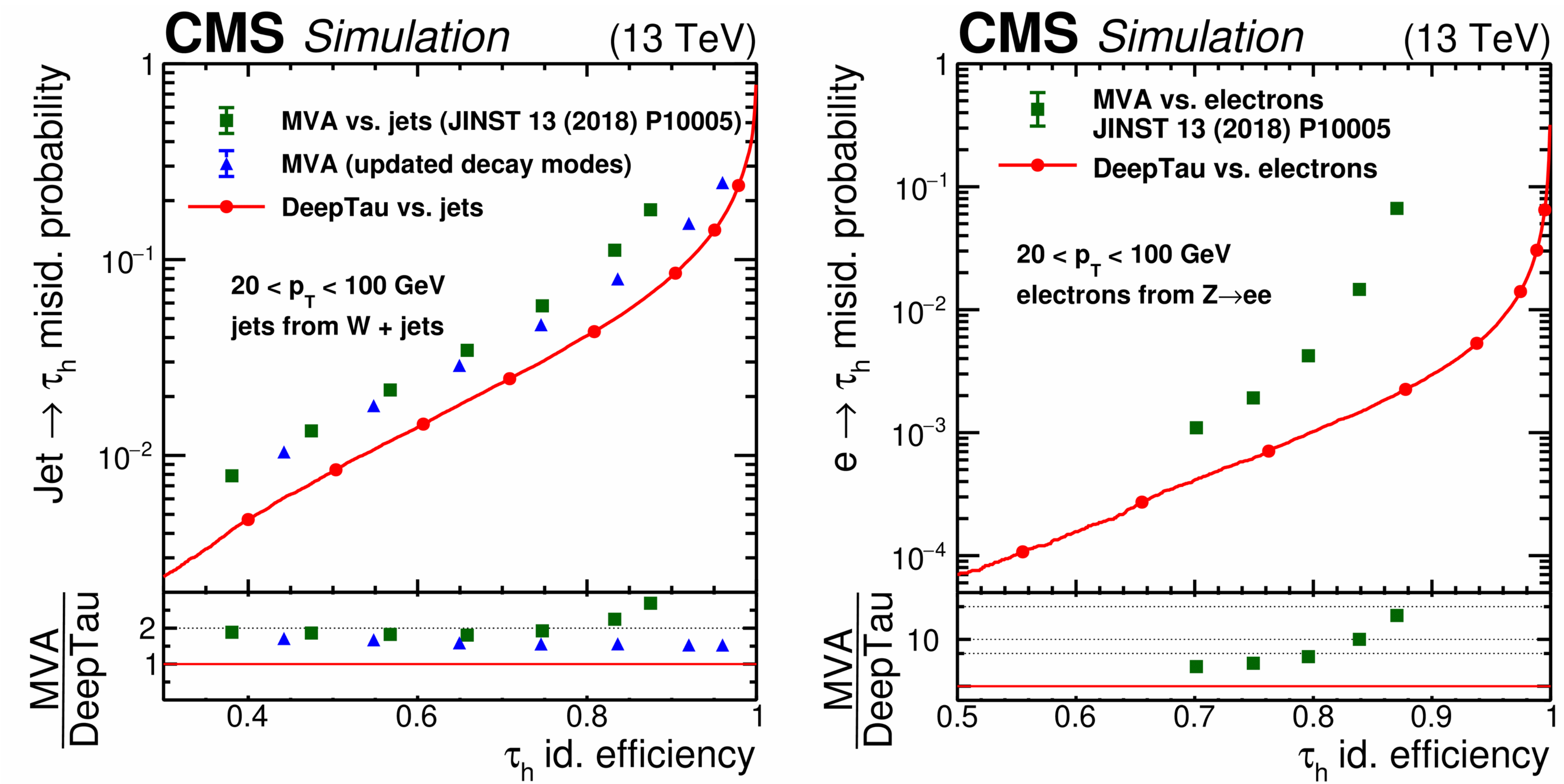
JINST 16 (2021) P05014





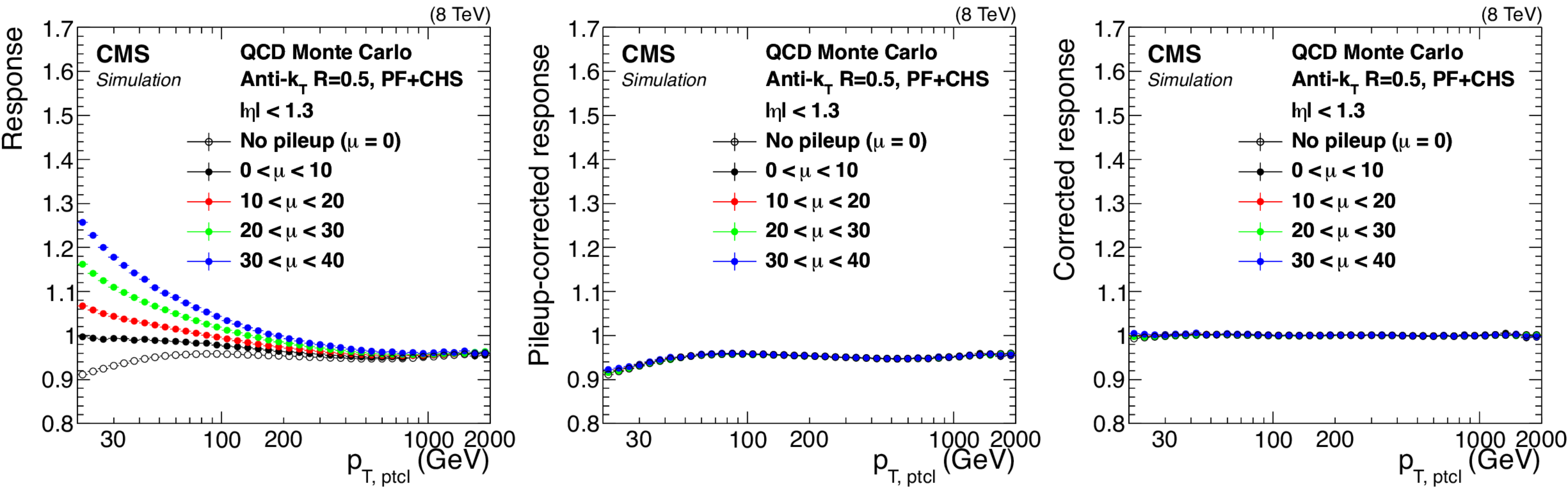
Tau efficiency

JINST 13 (2018) P10005

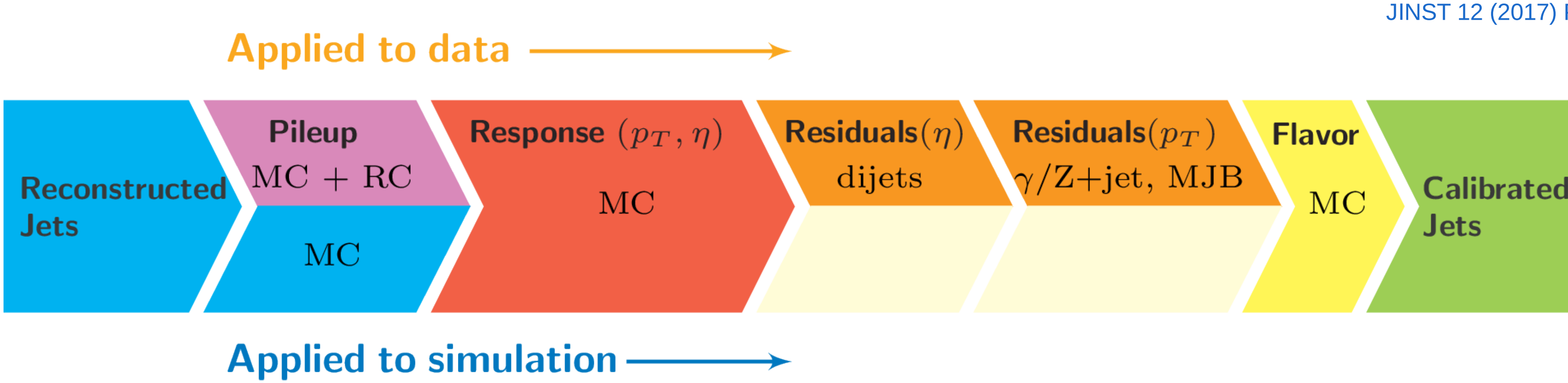


Jets

$$\text{Response} = \frac{\text{measured jet pT}}{\text{Truth jet pT}}$$



JINST 12 (2017) P02014



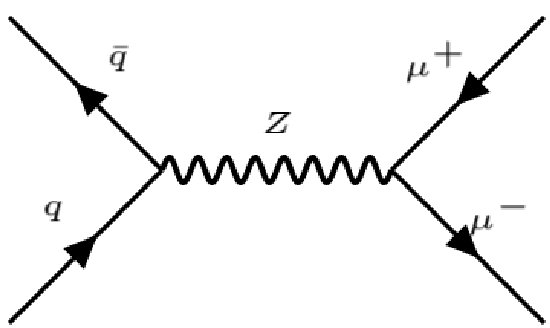
Z-tagging (leptons)

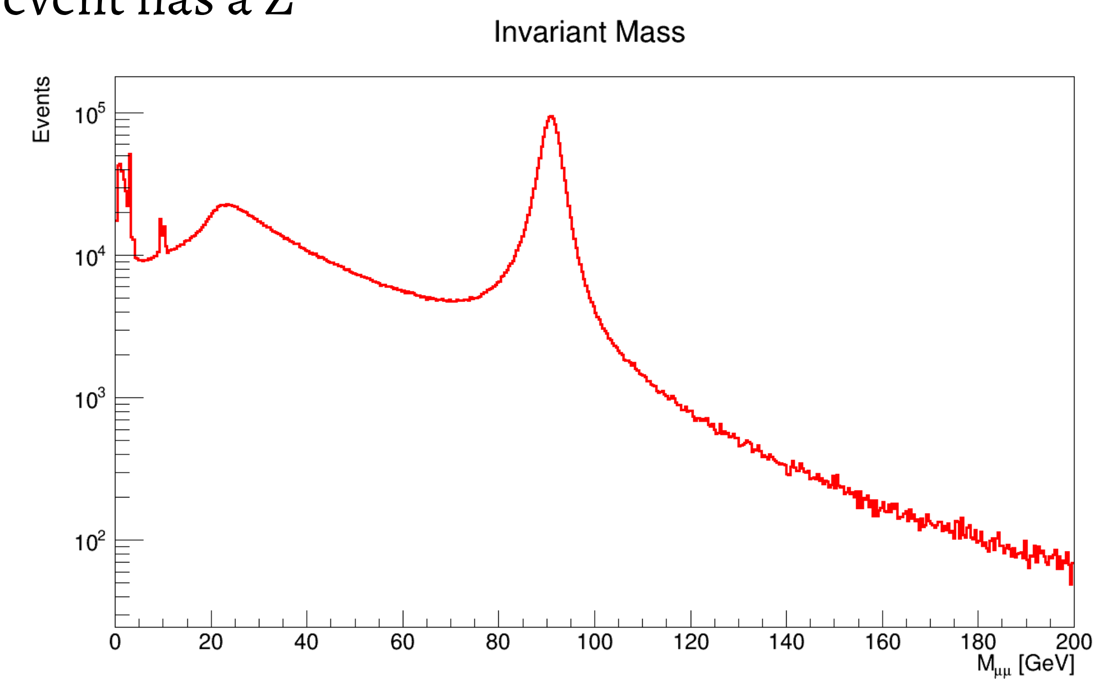
We saw how one can tag an event as having a Z boson

Typically select events with at least two leptons,

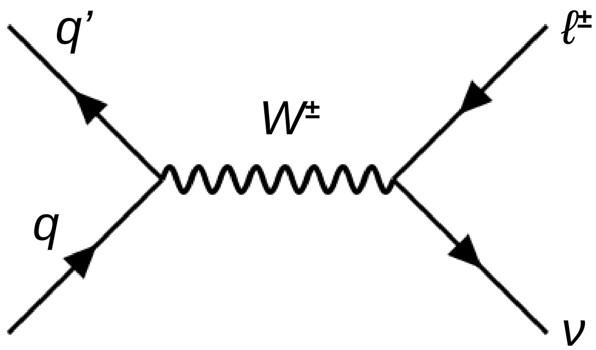
If their invariant mass is consistent with that of Z

(i.e. within some window), then the event has a Z

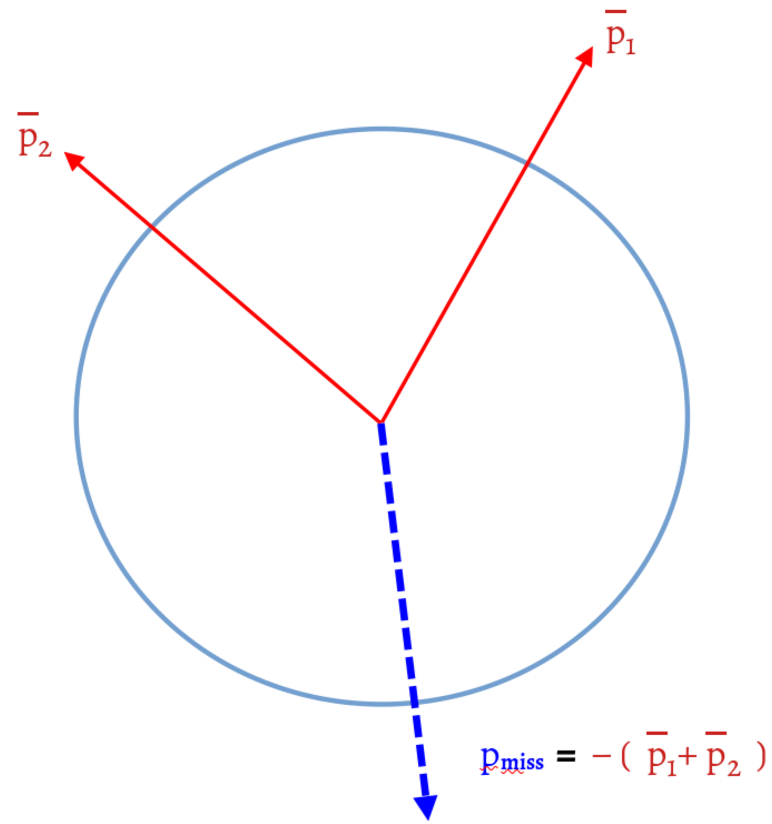




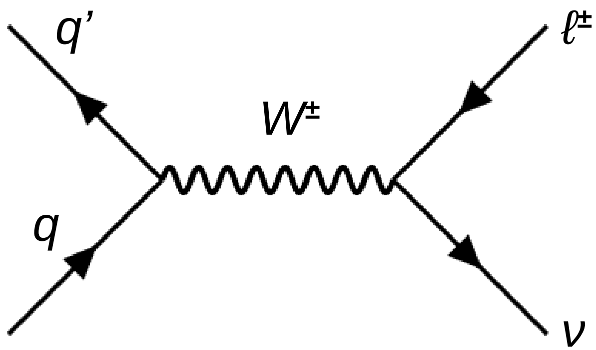
What about W ?



W needs us to reconstruct missing p_T

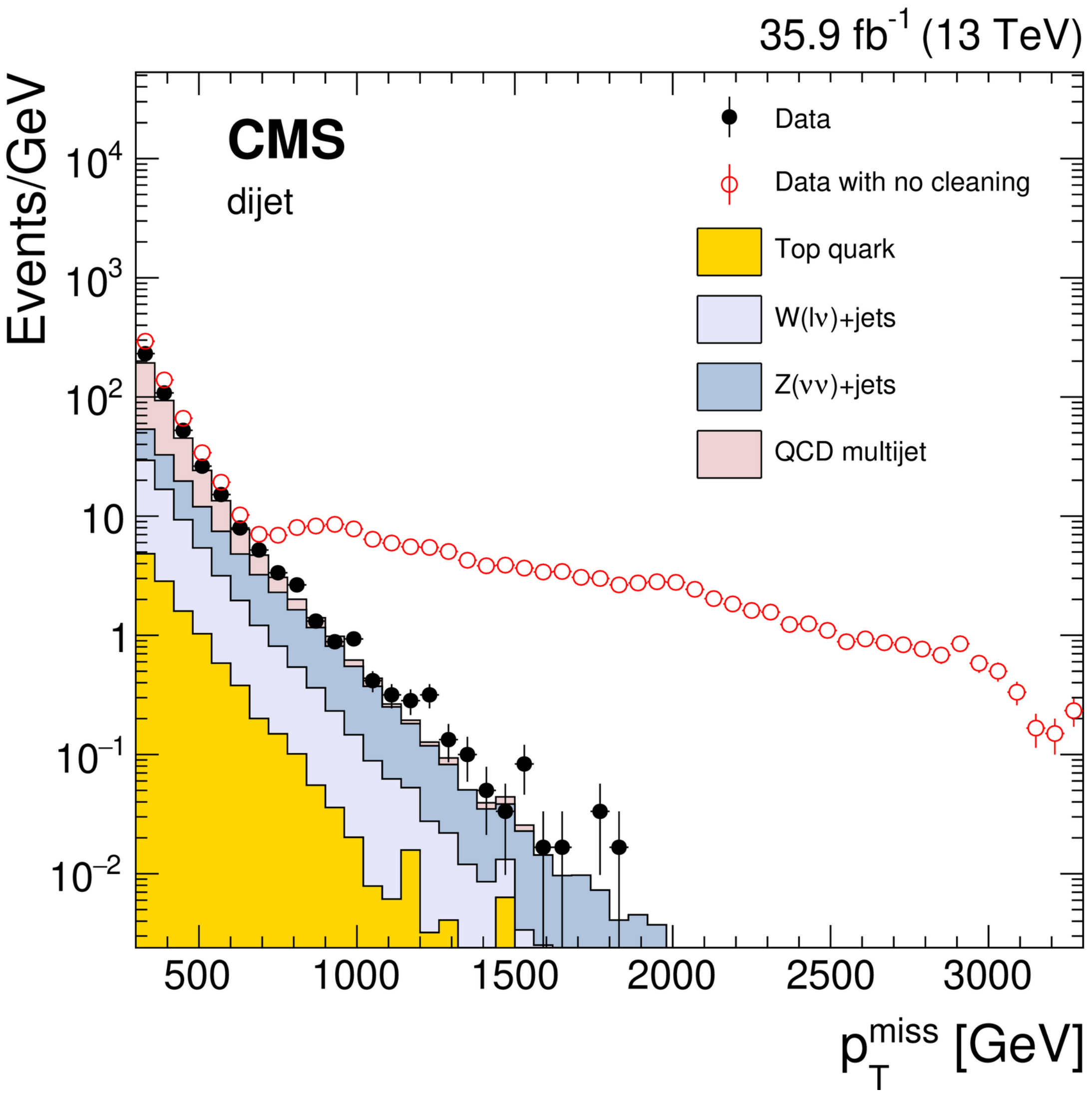


What about W ?



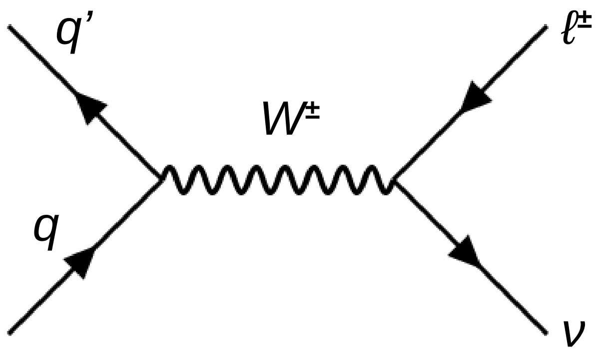
W needs us to reconstruct missing p_T

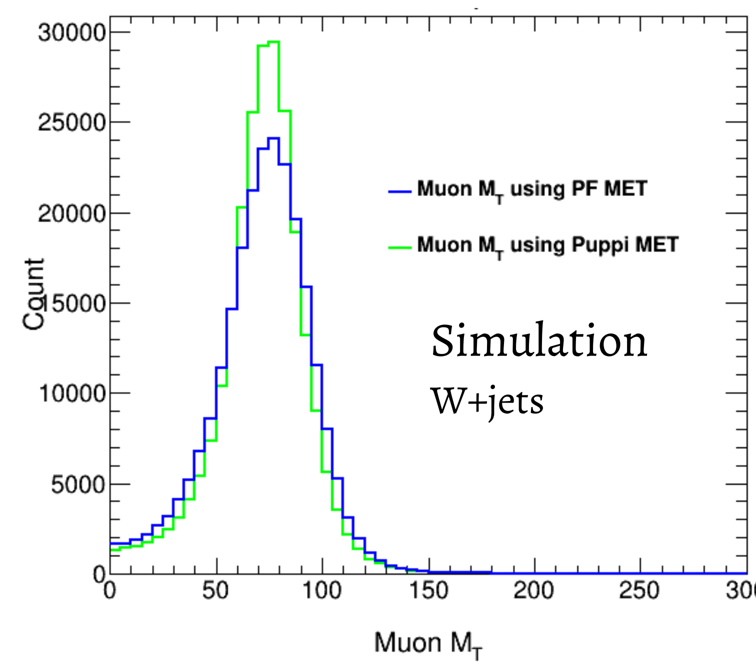
Before reconstructing pTmiss:
Cleaning: remove anomalous spikes
Filtering: remove noisy events (electronic noise, beam halo muons, poorly ID'ed muons)

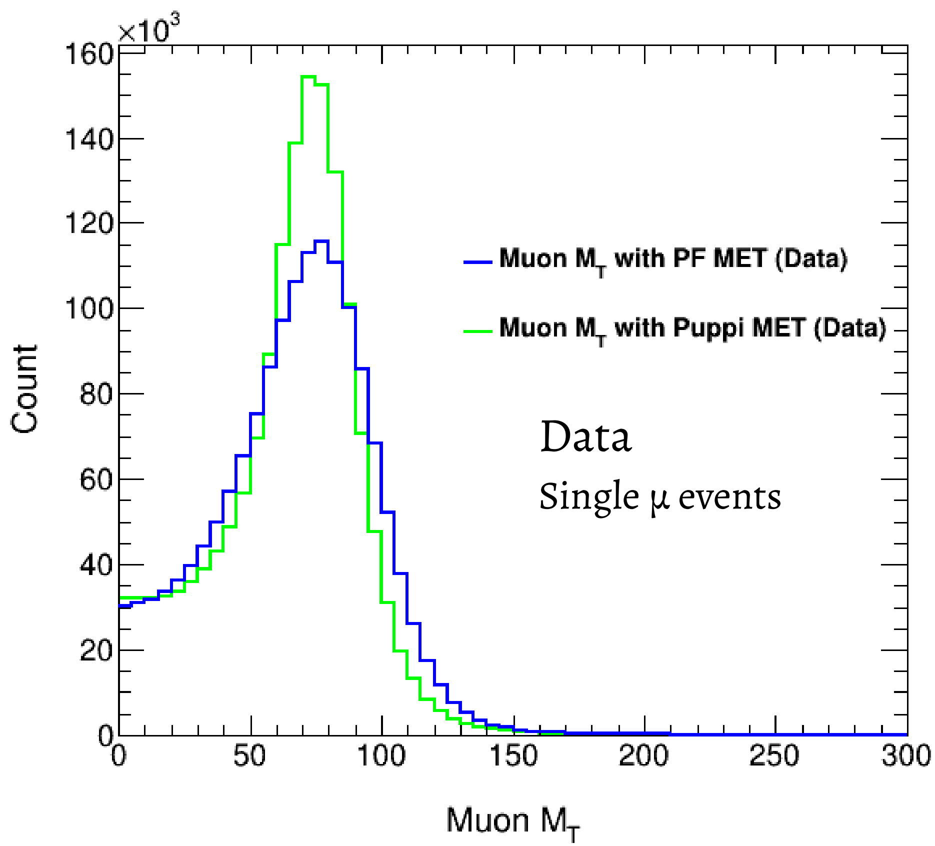


JINST 14 (2019) P07004

W-tagging (leptons)







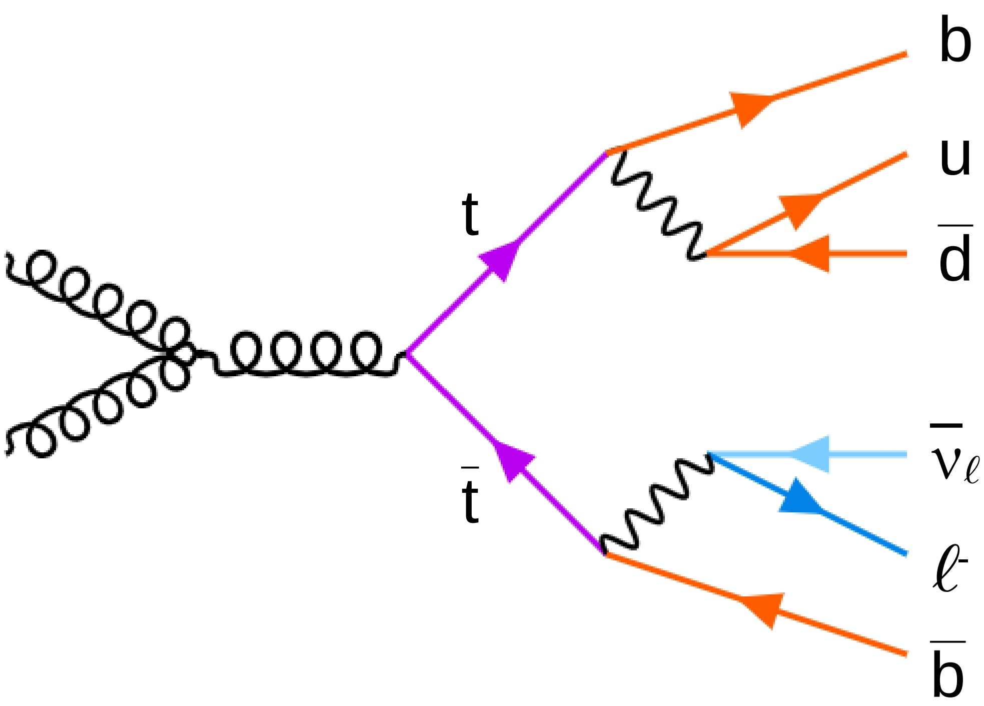
$$M_T = \left[2\, p_T^{miss}\, p_T^{\ell}\, (1-\cos(\Delta\phi))\right]^{1/2}$$

Back to analysis

Counting experiments

- Counting experiment
  - Looking for some process (We call this signal).
  - Make an event selection that we like.
  - Other SM processes will also pass selection (We call this background)
  - Predict number of background events
  - Check data, observation.
    - Does the observation match background prediction?
    - Or does it match background + some additional signal?

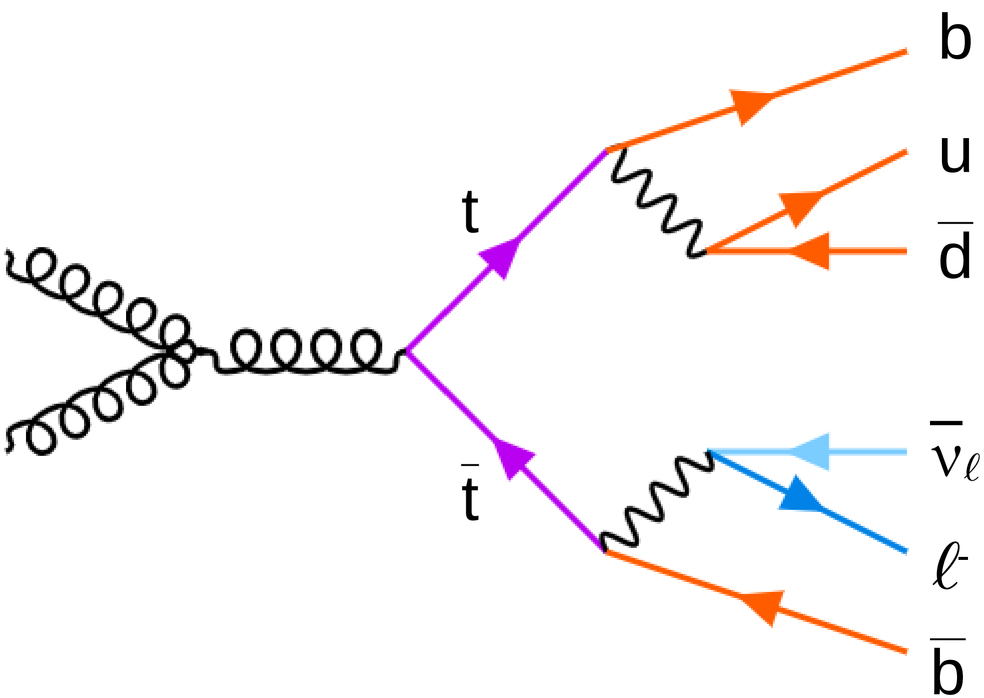
Top discovery



t →Wb is 100%

W^+ DECAY MODES		Fraction (Γ_i/Γ)
$\ell^+\nu$	[b]	(10.86 ± 0.09) %
$e^+\nu$		(10.71 ± 0.16) %
$\mu^+\nu$		(10.63 ± 0.15) %
$\tau^+\nu$		(11.38 ± 0.21) %
hadrons		(67.41 ± 0.27) %

Top discovery



Select a certain class of events from all collision events (lepton+jets, with b-tagged jets)

Expect some known processes (at that point) to contribute [WW,Wc].
Is the data completely explained by the known processes or not?

Top discovery

Select a certain class of events from all collision events
(lepton+jets, with b-tagged jets)

Expect some known processes (at that point) to contribute [WW,Wc].
Is the data completely explained by the known processes or not?

$N_{sig} = N_{tt} = \mathcal{L}(\sigma B)_{tt}(A\varepsilon)_{tt}$

$N_{bkg1} = \mathcal{L}(\sigma B)_{b1}(A\varepsilon)_{b1}$

$N_{bkg2} = \mathcal{L}(\sigma B)_{b2}(A\varepsilon)_{b2}$

$N_{bkg3} = \mathcal{L}(\sigma B)_{b3}(A\varepsilon)_{b3}$

$N_{bkg} = \Sigma_i N_{bkgi}$

Is the observation, N_{obs} consistent with N_{bkg}, or with $N_{bkg}+N_{sig}$

$N_{obs} \stackrel{?}{=} N_{bkg}$
$N_{obs} \stackrel{?}{=} N_{bkg} + N_{sig}$

Top discovery

Select a certain class of events from all collision events
(lepton+jets, with b-tagged jets)

Expect some known processes (at that point) to contribute [WW,Wc].
Is the data completely explained by the known processes or not?

$N_{sig} = N_{tt} = \mathcal{L}(\sigma B)_{tt}(A\varepsilon)_{tt}$

$N_{bkg1} = \mathcal{L}(\sigma B)_{b1}(A\varepsilon)_{b1}$

$N_{bkg2} = \mathcal{L}(\sigma B)_{b2}(A\varepsilon)_{b2}$

$N_{bkg3} = \mathcal{L}(\sigma B)_{b3}(A\varepsilon)_{b3}$

$N_{bkg} = \Sigma_i N_{bkg\,i}$

Is the observation, N_{obs} consistent with N_{bkg}, or with $N_{bkg}+N_{sig}$

$N_{obs} \stackrel{?}{=} N_{bkg}$

$N_{obs} \stackrel{?}{=} N_{bkg} + N_{sig}$

As we shall see, backgrounds can be estimated from data too….

Top discovery

N_{jet}	observed events	observed SVX tags	background tags expected
1	6578	40	50 ± 12
2	1026	34	21.2 ± 6.5
3	164	17	5.2 ± 1.7
≥ 4	39	10	1.5 ± 0.4

Table 1: Number of lepton+jet events in the 67 pb^{-1} data sample along with the numbers of SVX tags observed and the estimated background. Based on the excess number of tags in events with ≥ 3 jets, we expect an additional 0.5 and 5 tags from $t\bar{t}$ decay in the 1 and 2 jet bins respectively.

SVX tag = b-tag

Phys. Rev. Lett. 74, 2626 (1995)

Top discovery

Phys. Rev. Lett. 74, 2626 (1995)

Channel:	SVX	SLT	Dilepton
observed	27 tags	23 tags	6 events
expected background	6.7 ± 2.1	15.4 ± 2.0	1.3 ± 0.3
background probability	2×10^{-5}	6×10^{-2}	3×10^{-3}

The probability that background fluctuated to give this observation is 10^{-6}, which is 4.8 σ

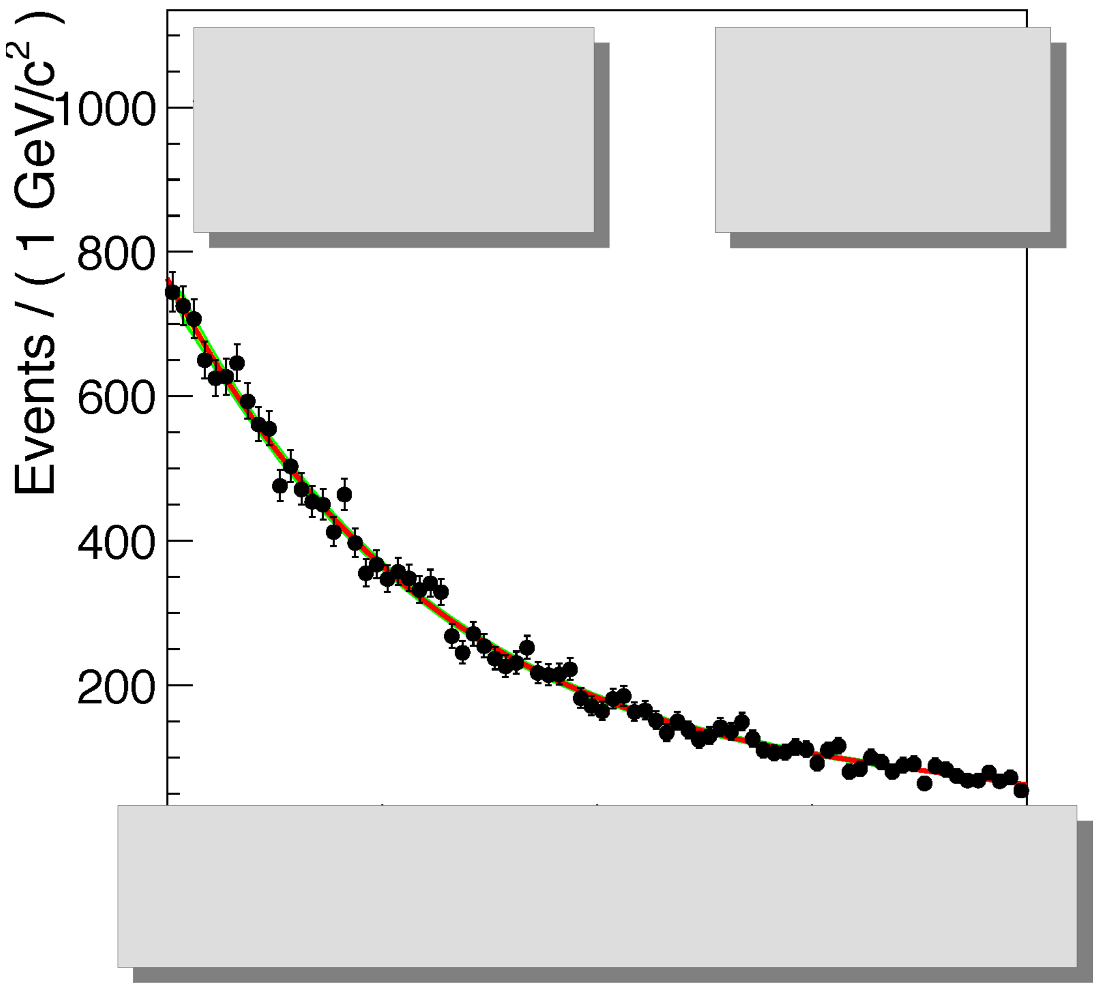
Since observation is not explained by background, there is additional signal present.. in this case the top quark!

A bump hunt

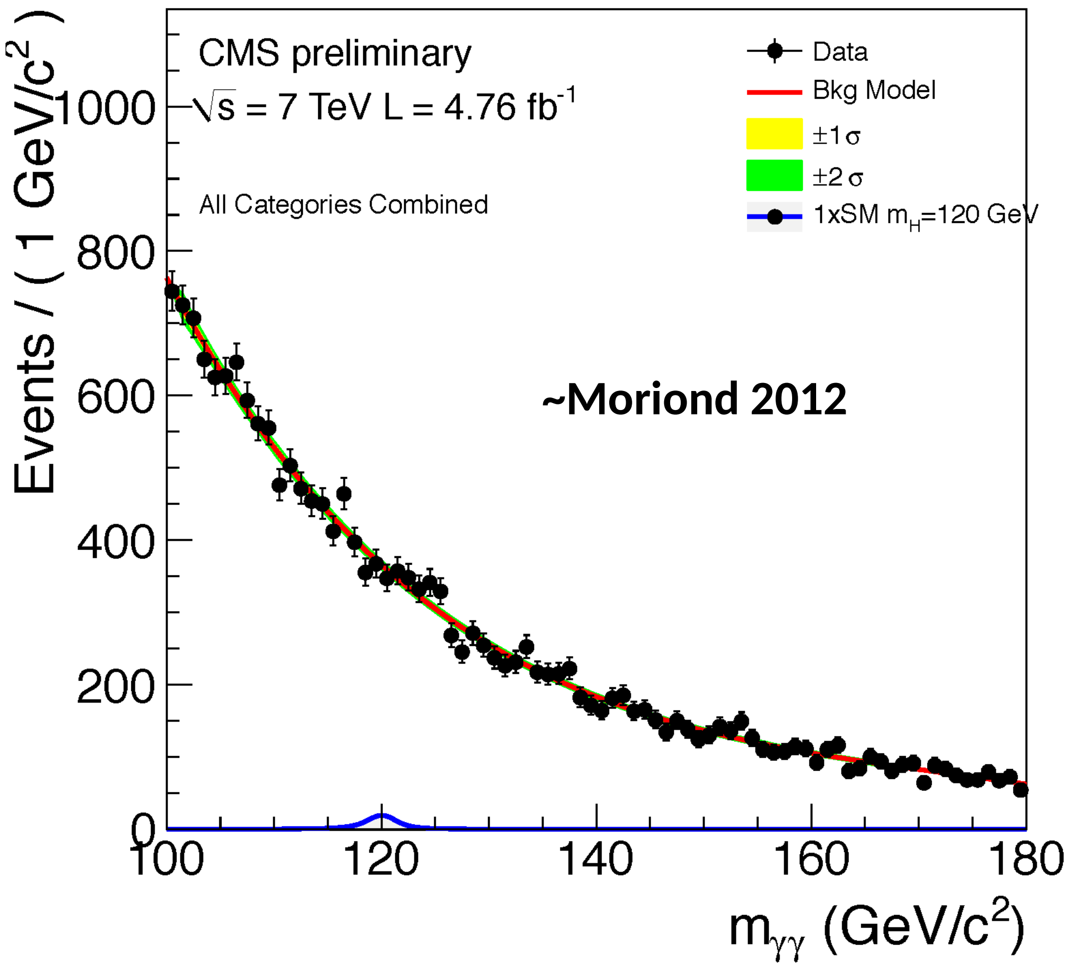
Bump hunt

- Looking for some process (We call this signal).
- Make an event selection that we like.
- Other SM processes will also pass selection (We call this background)
- Predict ~~number~~ distribution of background events
- Check data, observation.
    - Does the observed distribution match background prediction?
    - Or does it match background + some additional signal?

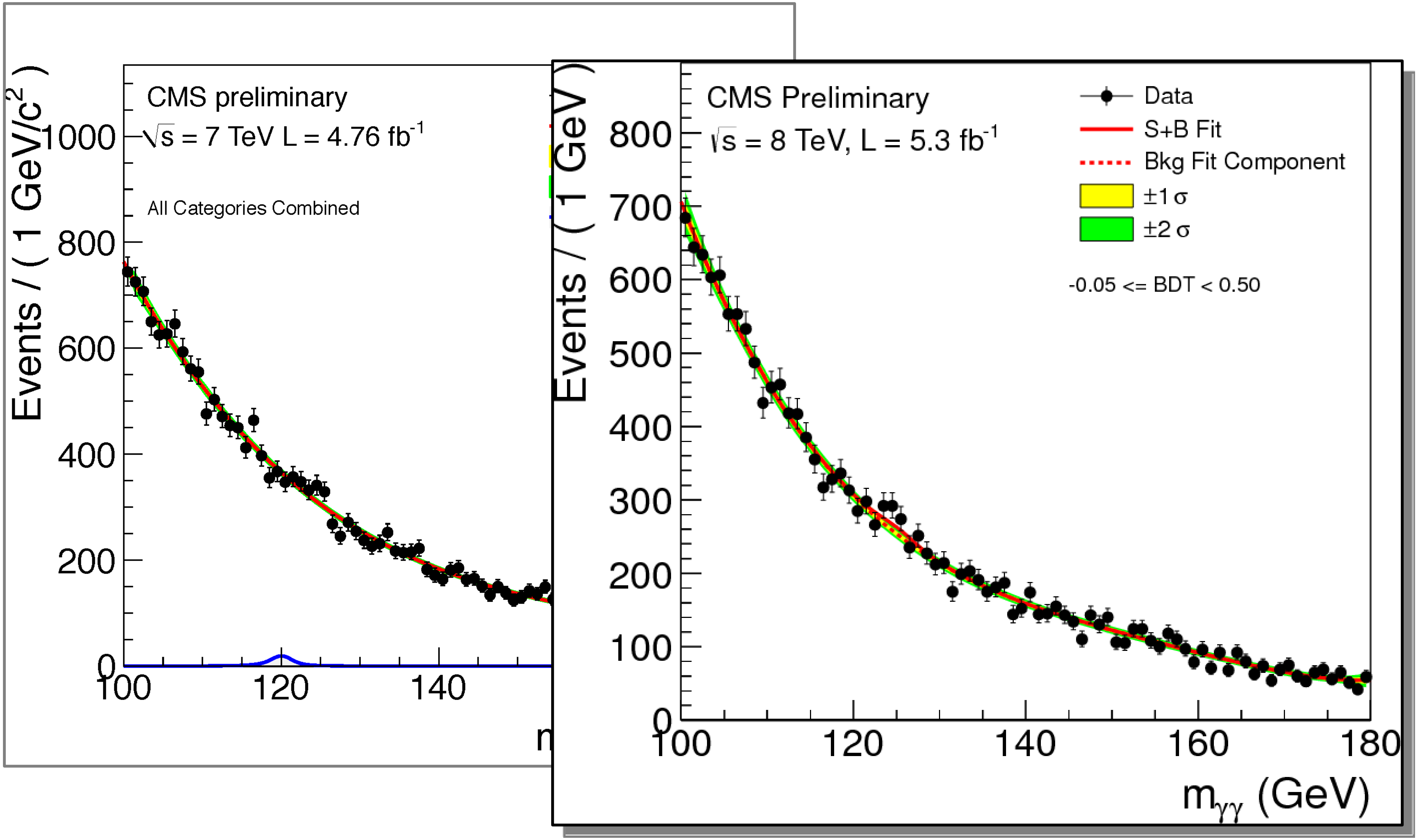
Some bump hunt



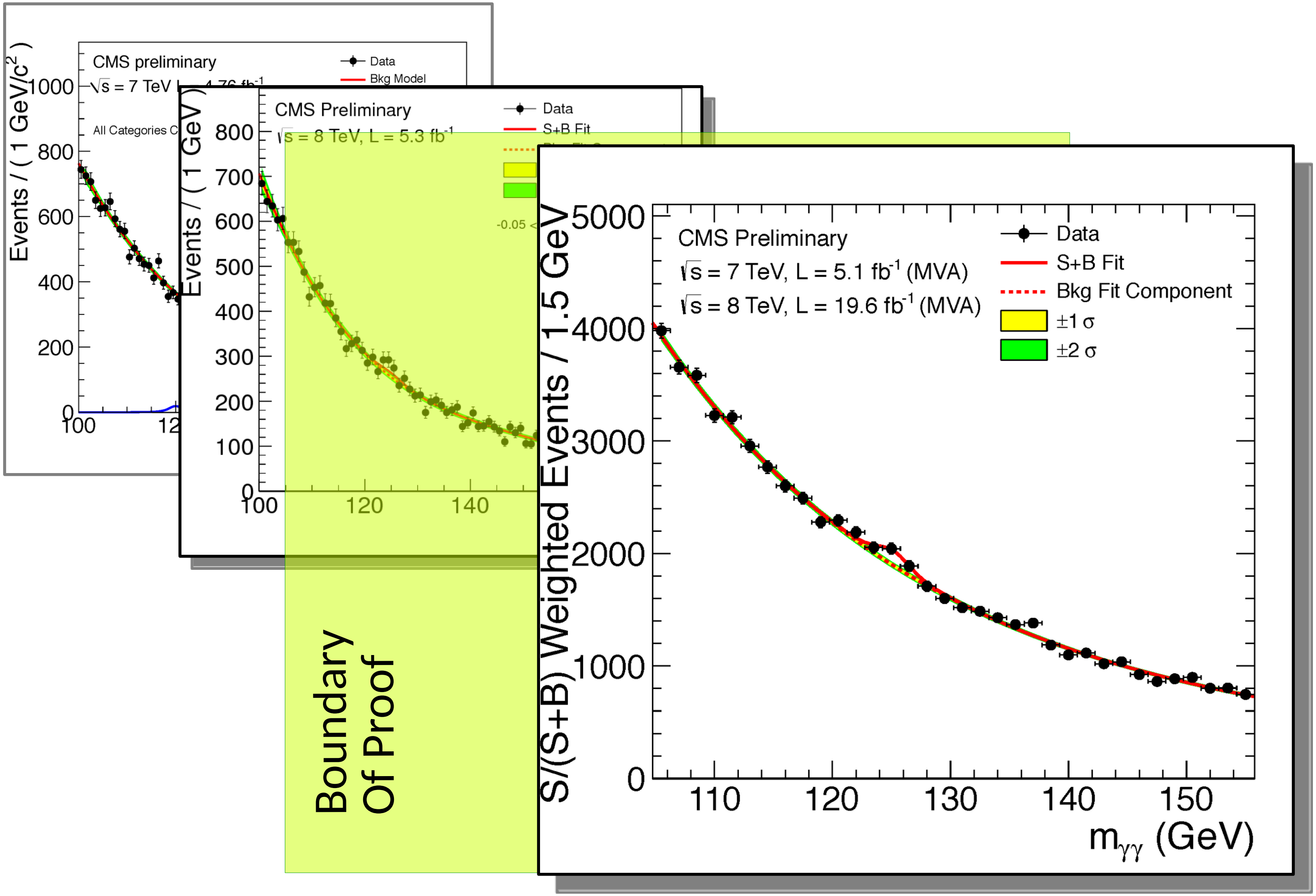
Higgs search



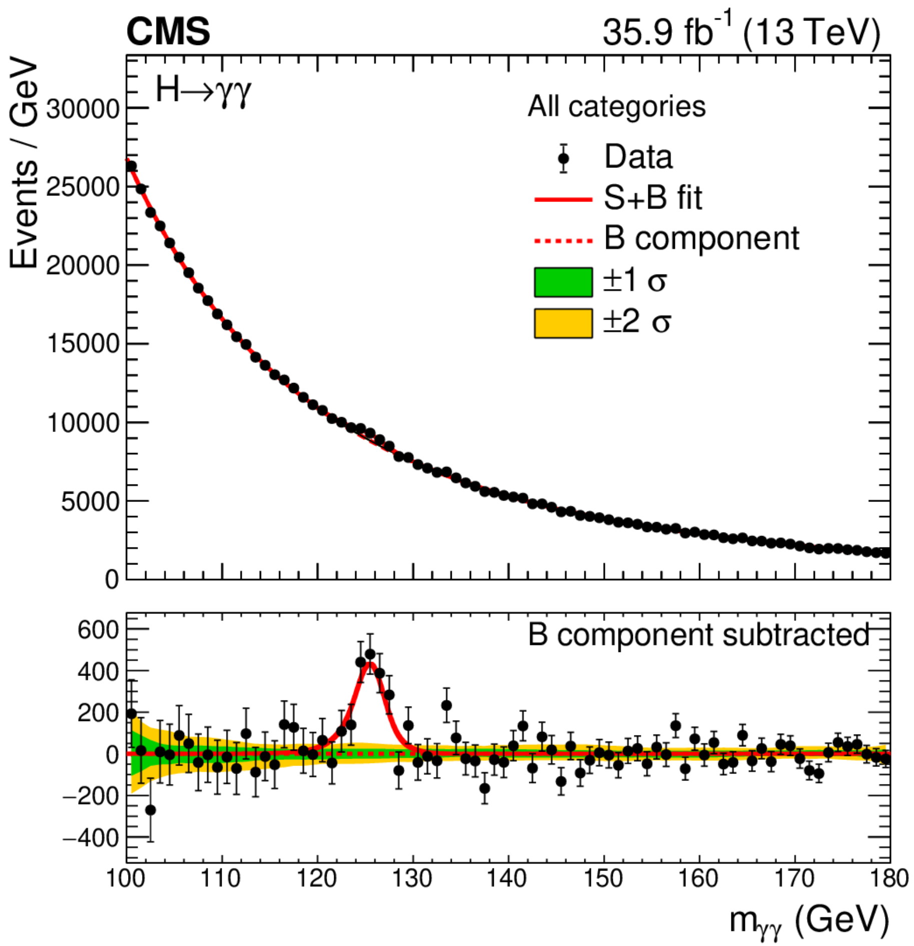
Higgs search

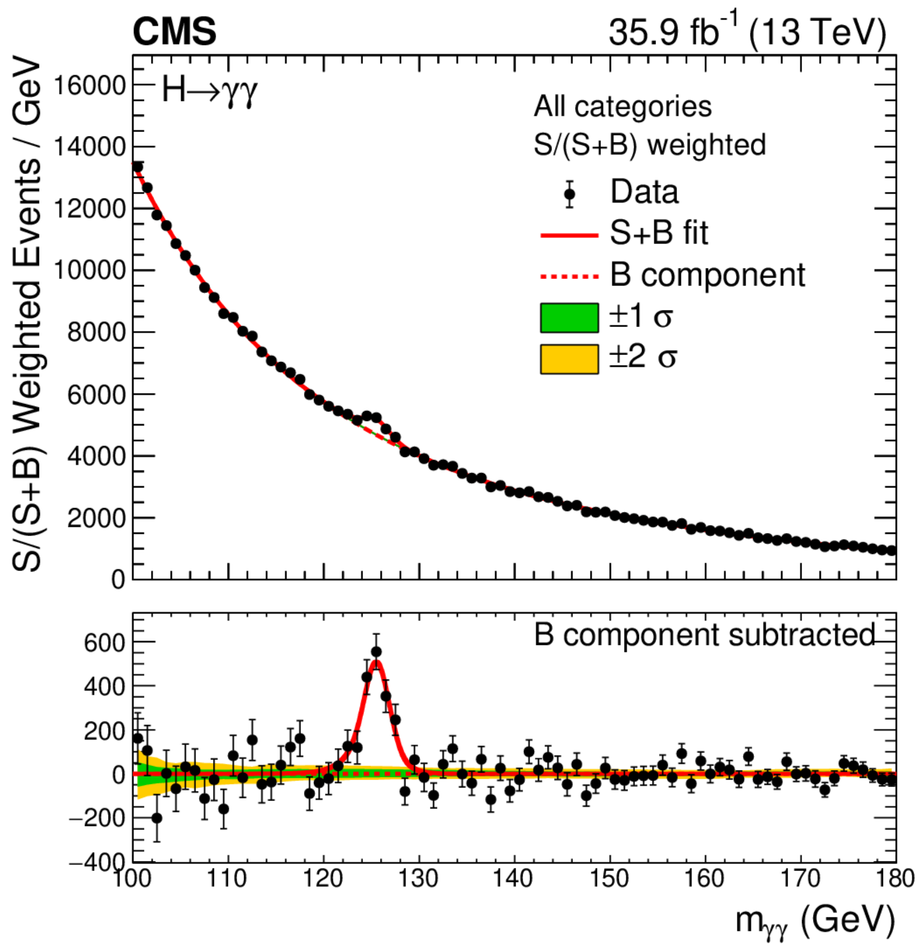


Higgs search



With 13 TeV





Phys. Lett. B 805 (2020) 135425

Limits…

What happens when the observation matches the background?

Thus no signal is found (obviously)… but can we say something about the signal?

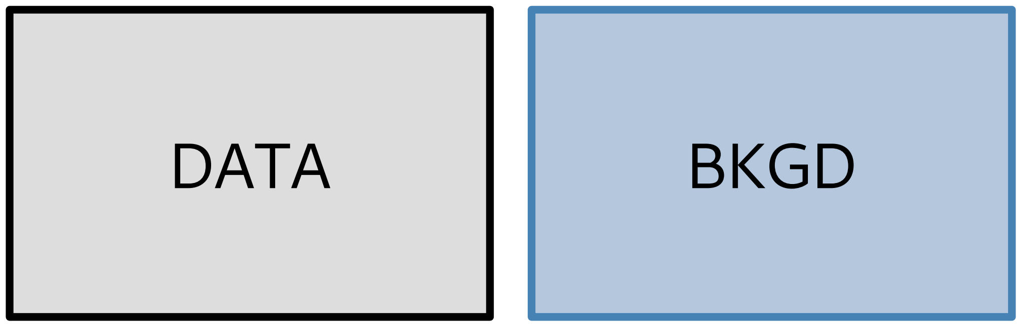
Limits…

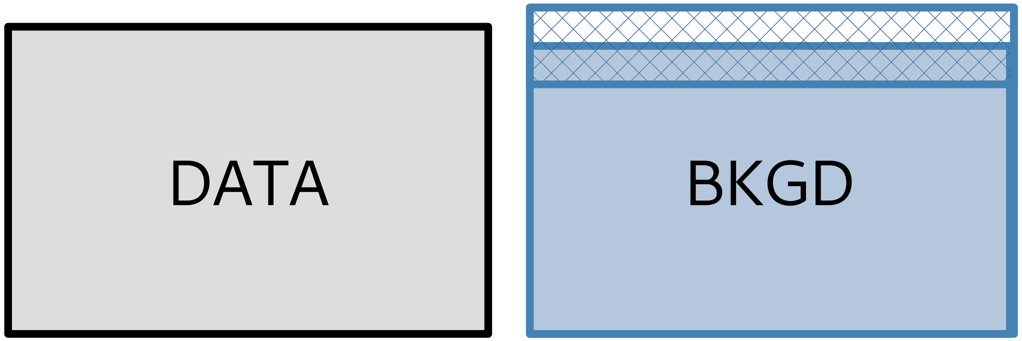
What happens when the observation matches the background?

Thus no signal is found (obviously)… but can we say something about the signal?

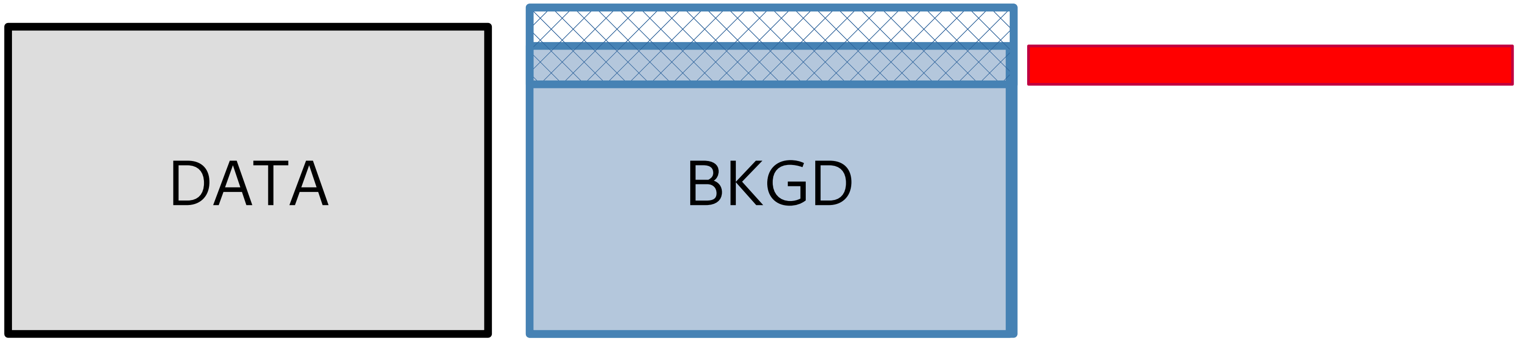
Two options
(a) Model for signal does not predict reality OR
(b) Collider not producing enough events model says.

Leads to constraints on model, or ruling out some parameter space for models.


DATA
BKGD

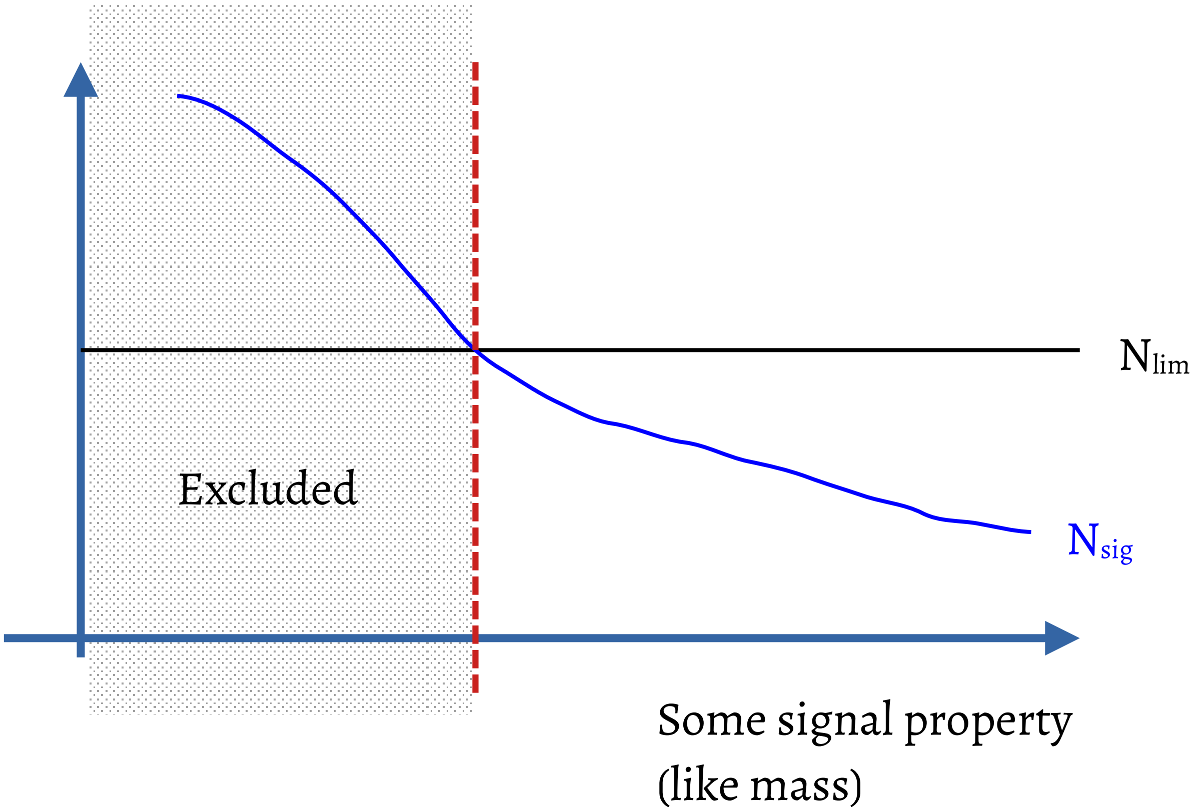


But there are uncertainties involved, certainly in the background prediction… but also inherent in the data.



But there are uncertainties involved, certainly in the background prediction... but also inherent in the data.

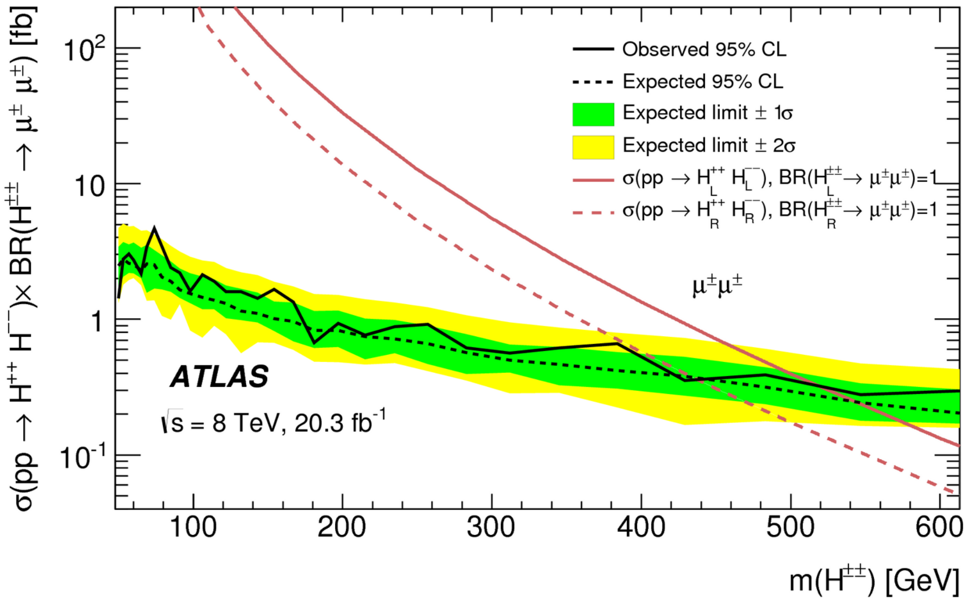
What is the maximum amount of signal that could be hidden in the data without us noticing?

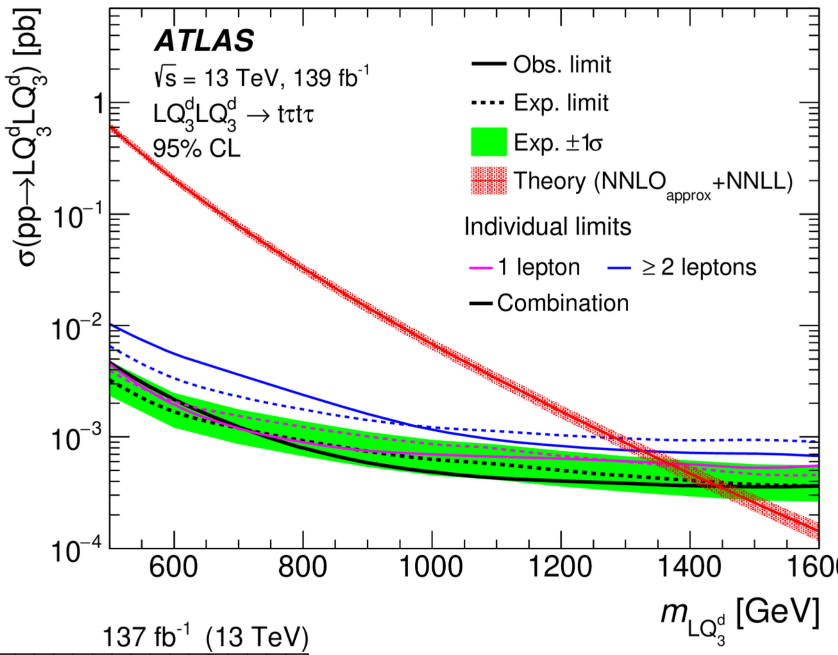


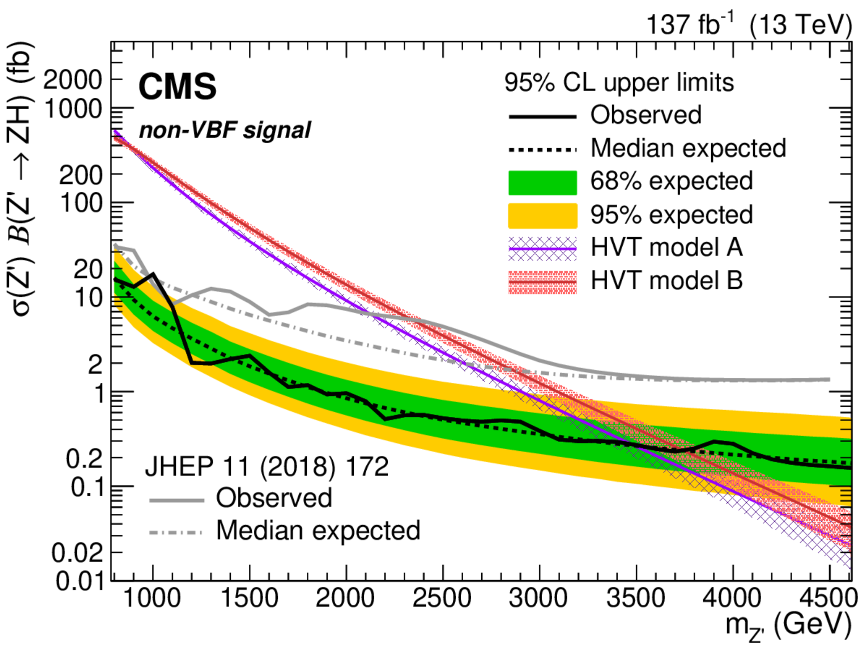
The number theory is predicting.

Examples

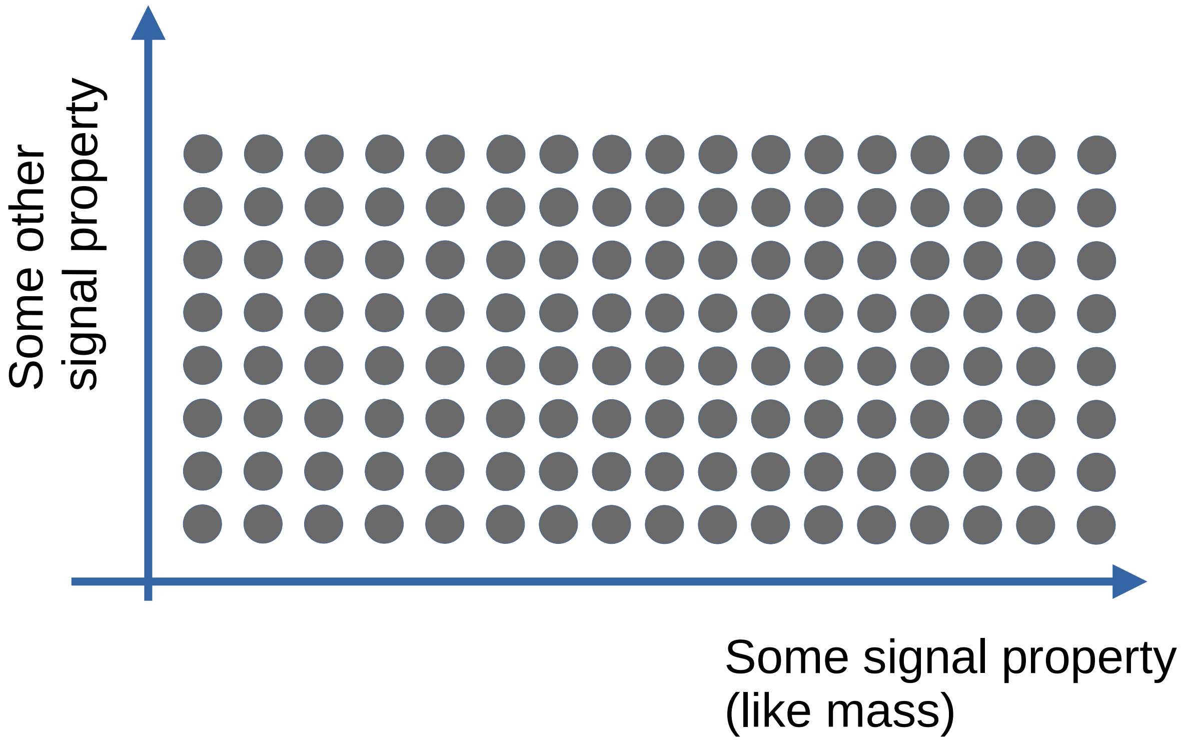
$$(\sigma B)_{lim} = N_{lim} / (L\,A\,\,\varepsilon)$$





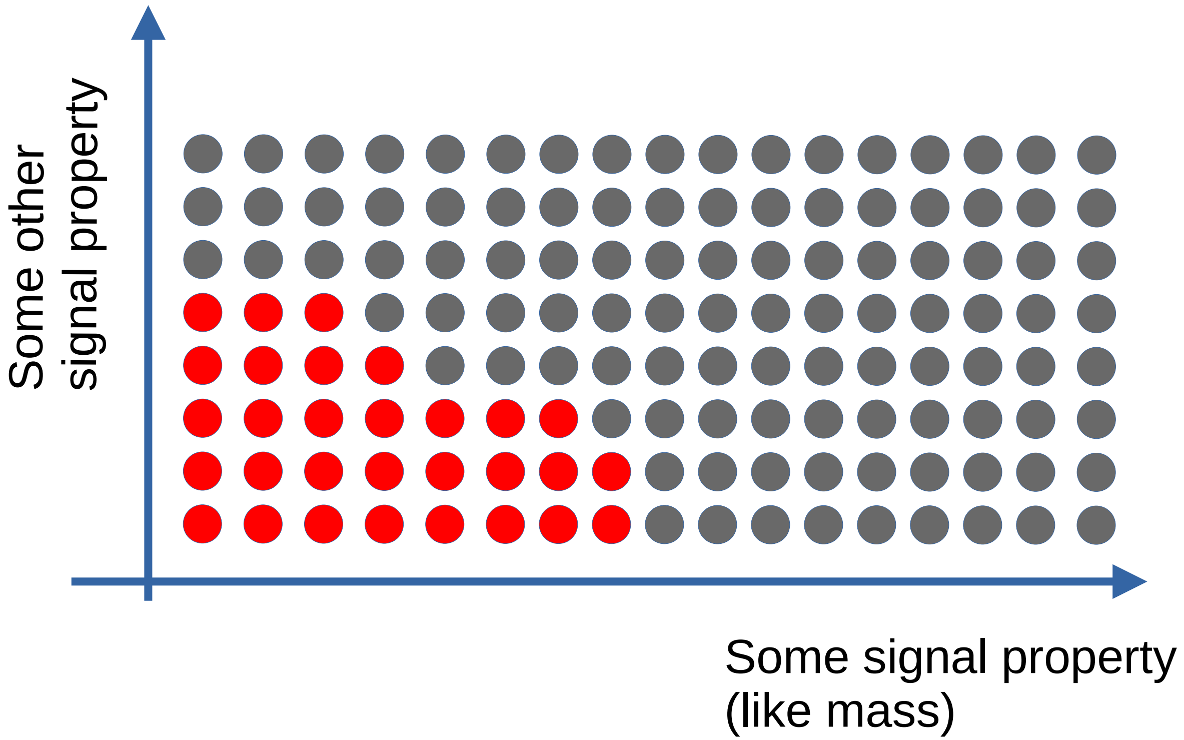


2D limits



At each point, compare N_{lim} and N_{sig}.

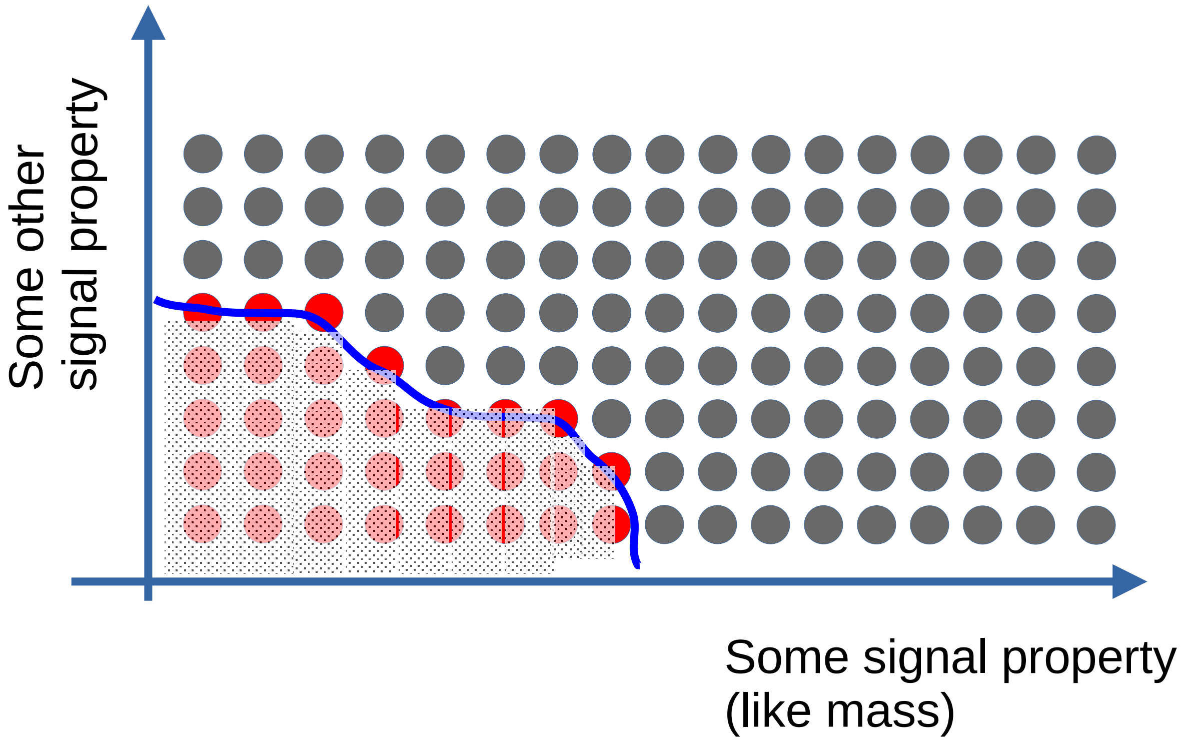
2D Limits



At each point, compare N_{lim} and N_{sig}.

Wherever $N_{sig} > N_{lim}$, we rule out the signal. ●

2D Limits

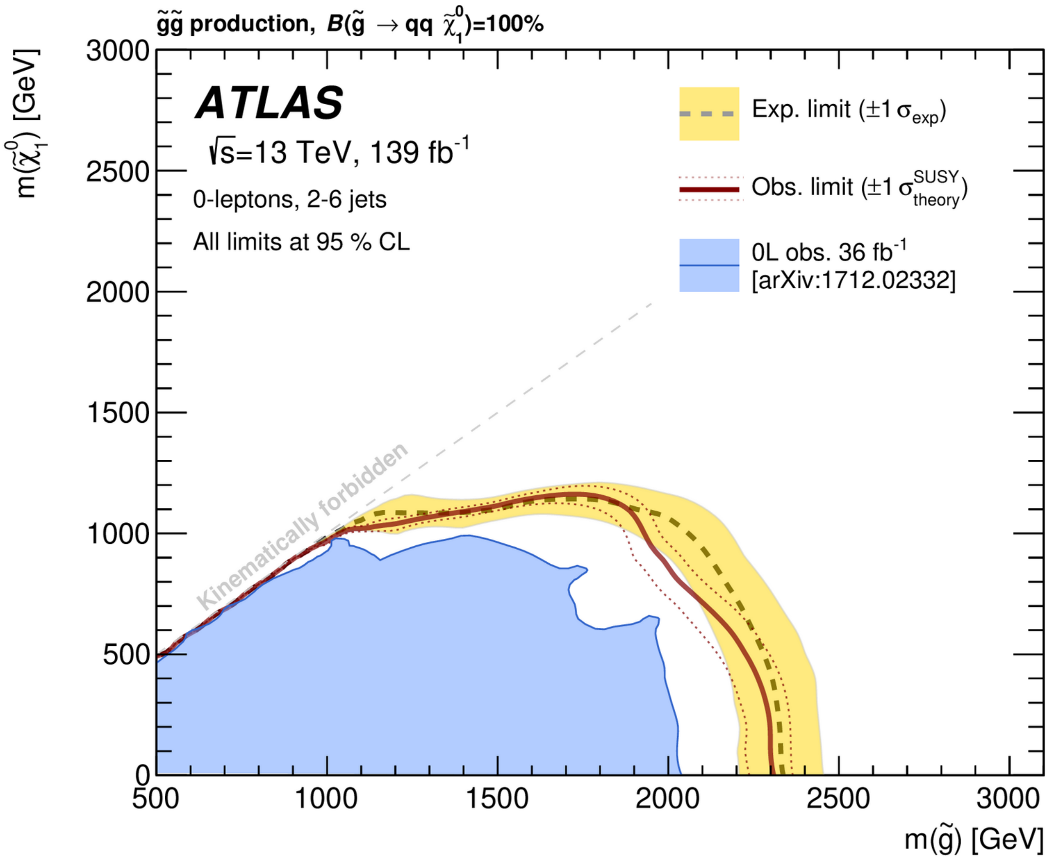


At each point, compare N_{lim} and N_{sig}.

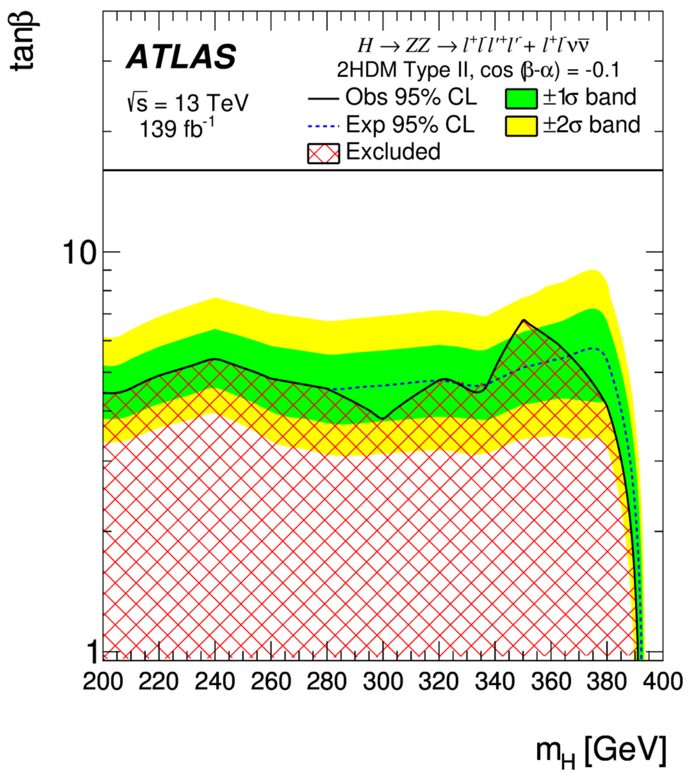
Wherever $N_{sig} > N_{lim}$, we rule out the signal.

Draw a contour around what is excluded.

Examples



g̃g̃ production, B(g̃ → qq χ̃₁⁰)=100%
ATLAS
√s=13 TeV, 139 fb⁻¹
0-leptons, 2-6 jets
All limits at 95 % CL
Exp. limit (±1 σexp)
Obs. limit (±1 σ SUSY theory)
0L obs. 36 fb⁻¹ [arXiv:1712.02332]
Kinematically forbidden
m(χ̃₁⁰) [GeV]
m(g̃) [GeV]



ATLAS
√s = 13 TeV
139 fb⁻¹
H → ZZ → l⁺l⁻l'⁺l'⁻ + l⁺l⁻νν̄
2HDM Type II, cos (β-α) = -0.1
Obs 95% CL
Exp 95% CL
Excluded
±1σ band
±2σ band
tanβ
m_H [GeV]