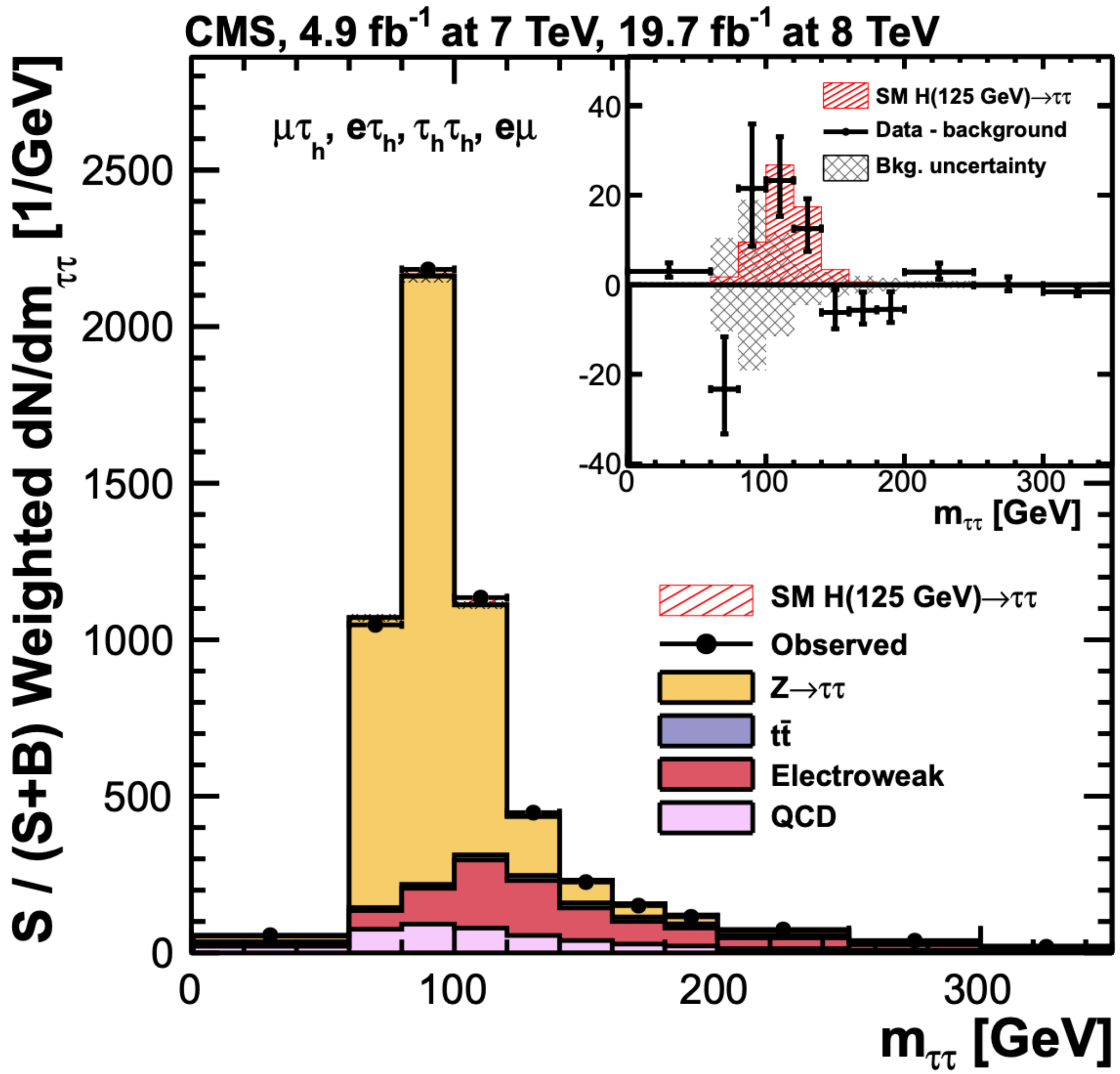
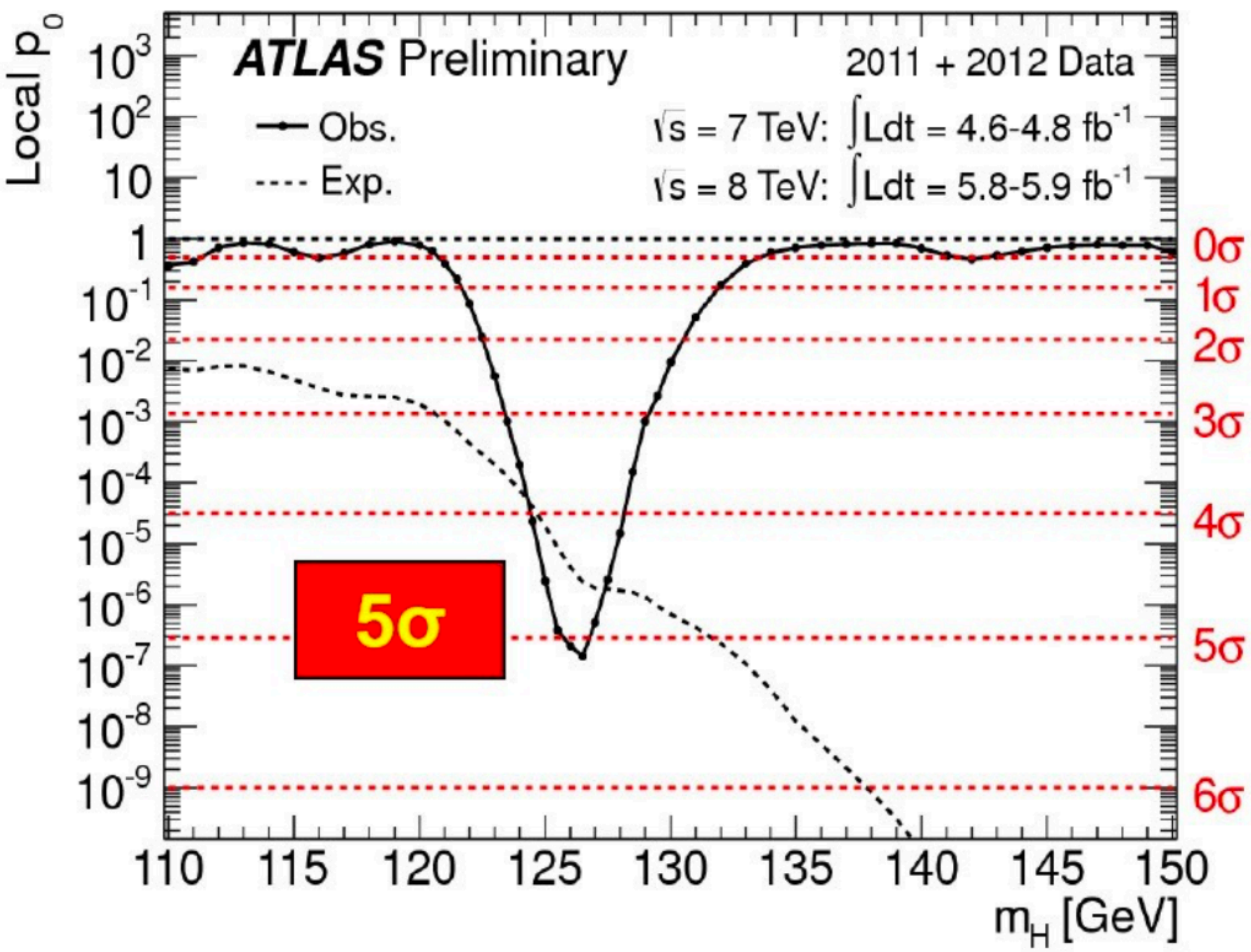


Lecture 13: Deep Learning Regressions

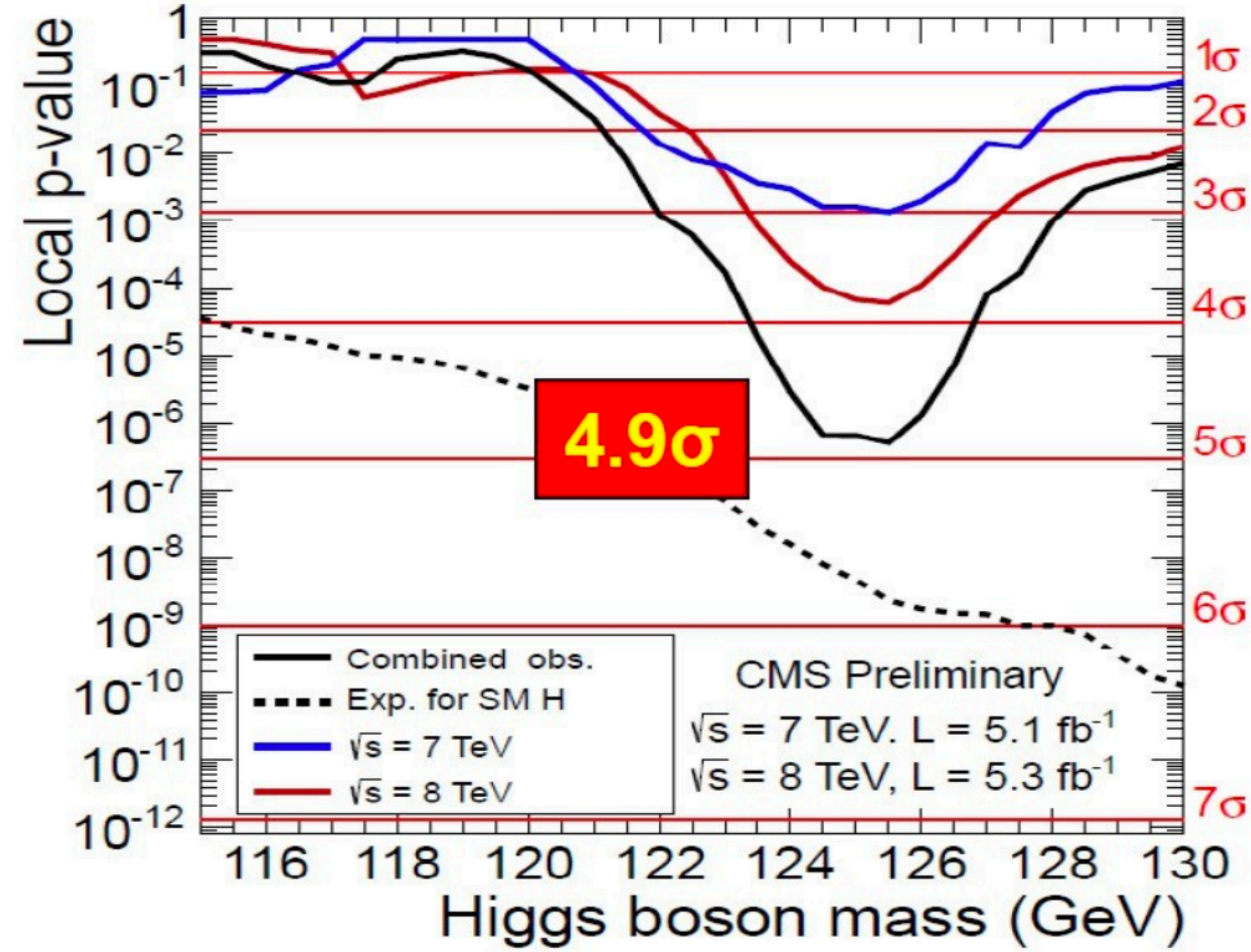
What you may not know?

At the Higgs discovery

ATLAS *γγ, 4l updated with ~6 fb^{-1} of 8 TeV data*



CMS *All channels updated with ~5 fb^{-1} of 8 TeV data*



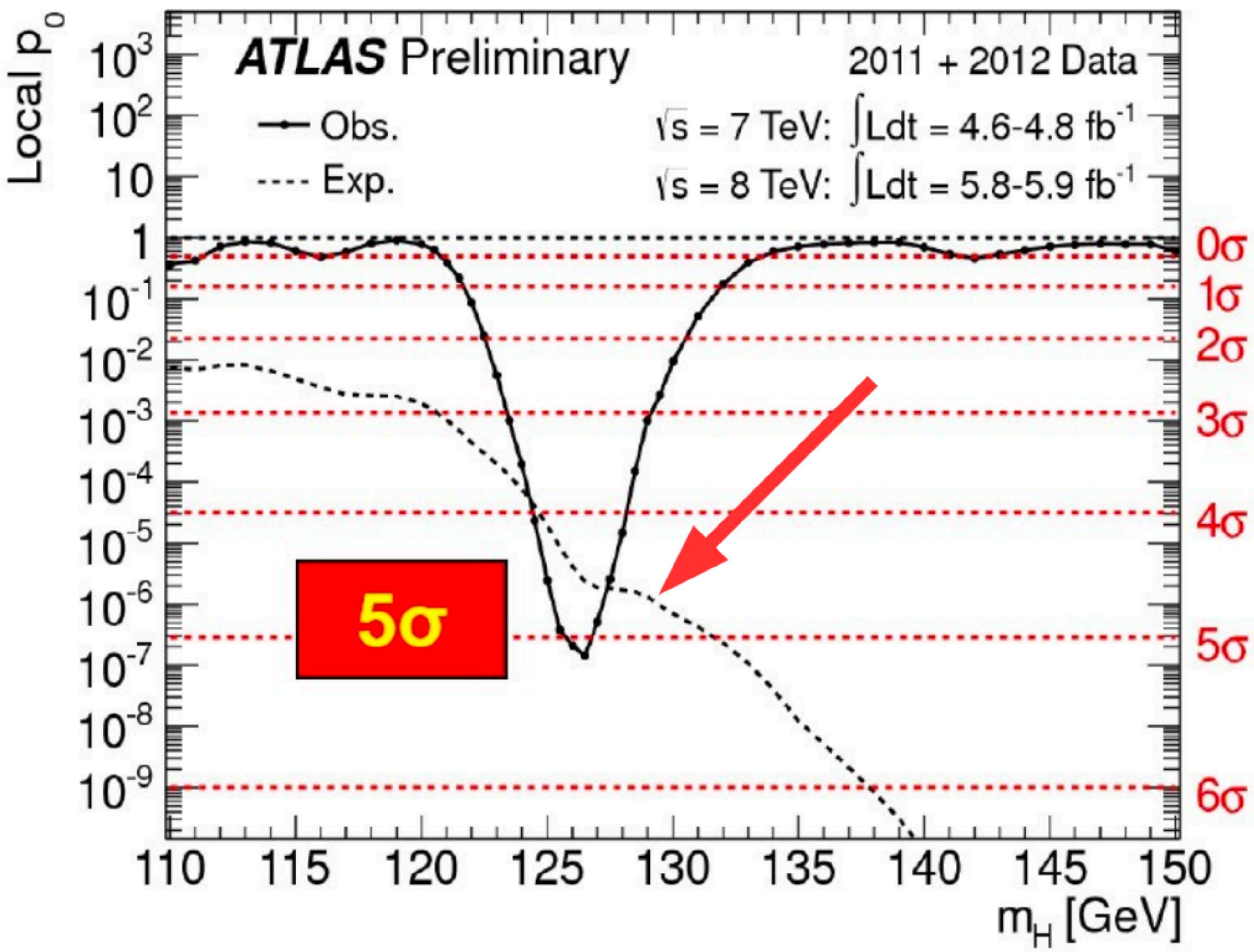
Largest local excess:
5σ at m_H= 126.5 GeV

Largest local excess:
4.9σ around m_H= 125 GeV
(using H→γγ and H→4l: 5.0σ)

A big difference was present

ATLAS

γγ, 4l updated with ~6 fb^{-1} of 8 TeV data

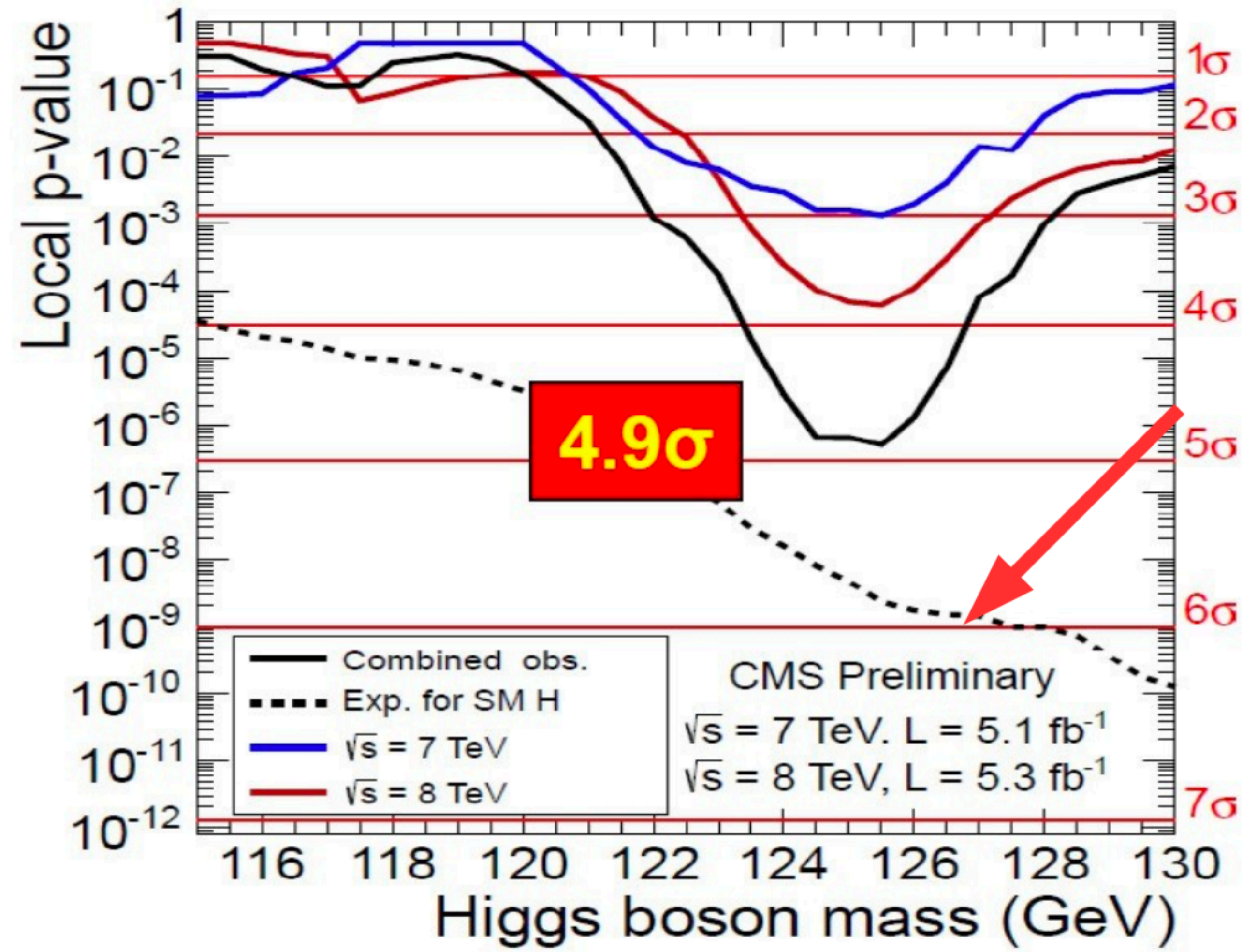


Largest local excess:

5σ at m_H= 126.5 GeV

CMS

All channels updated with ~5 fb^{-1} of 8 TeV data



Largest local excess:

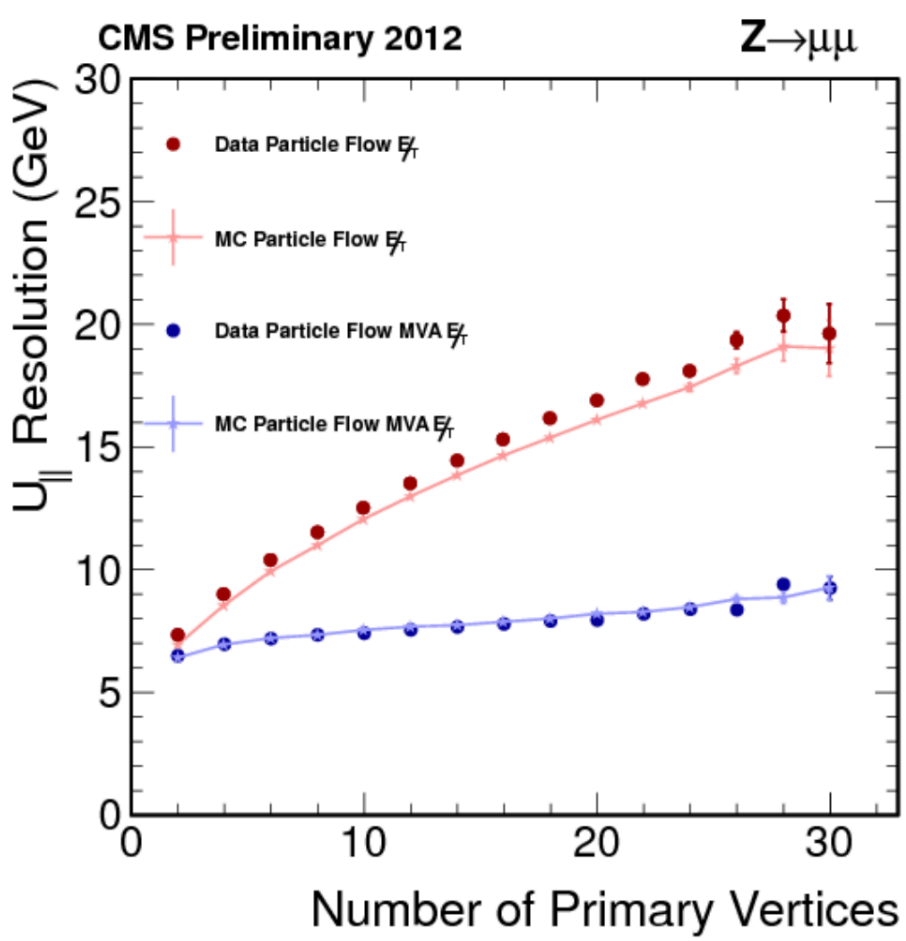
4.9σ around m_H= 125 GeV

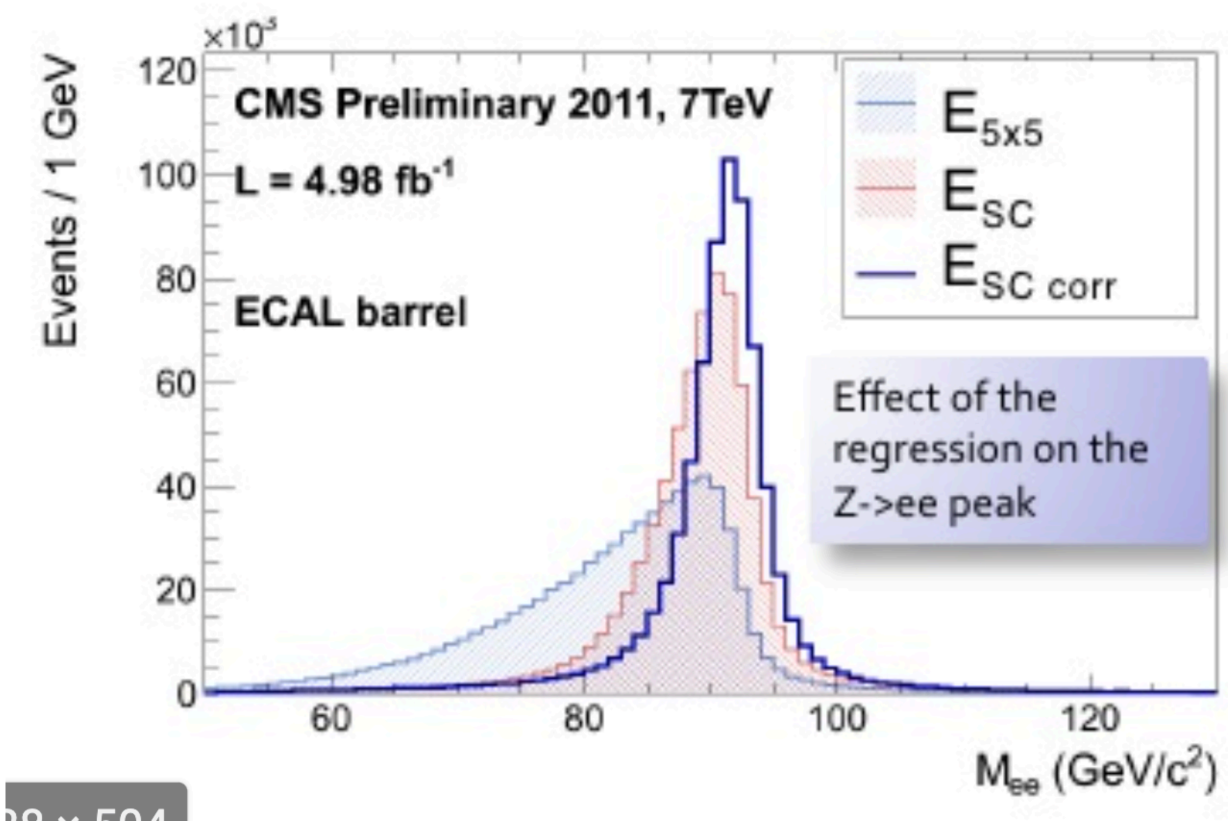
(using H→γγ and H→4l: 5.0σ)

CMS was nearly 30% more sensitive Despite an excess of same size

What caused the difference?

- A few things, but the big one was deep learning
- In particular, two novel deep learning approaches
    - These approaches involved deep learning regression





Overview

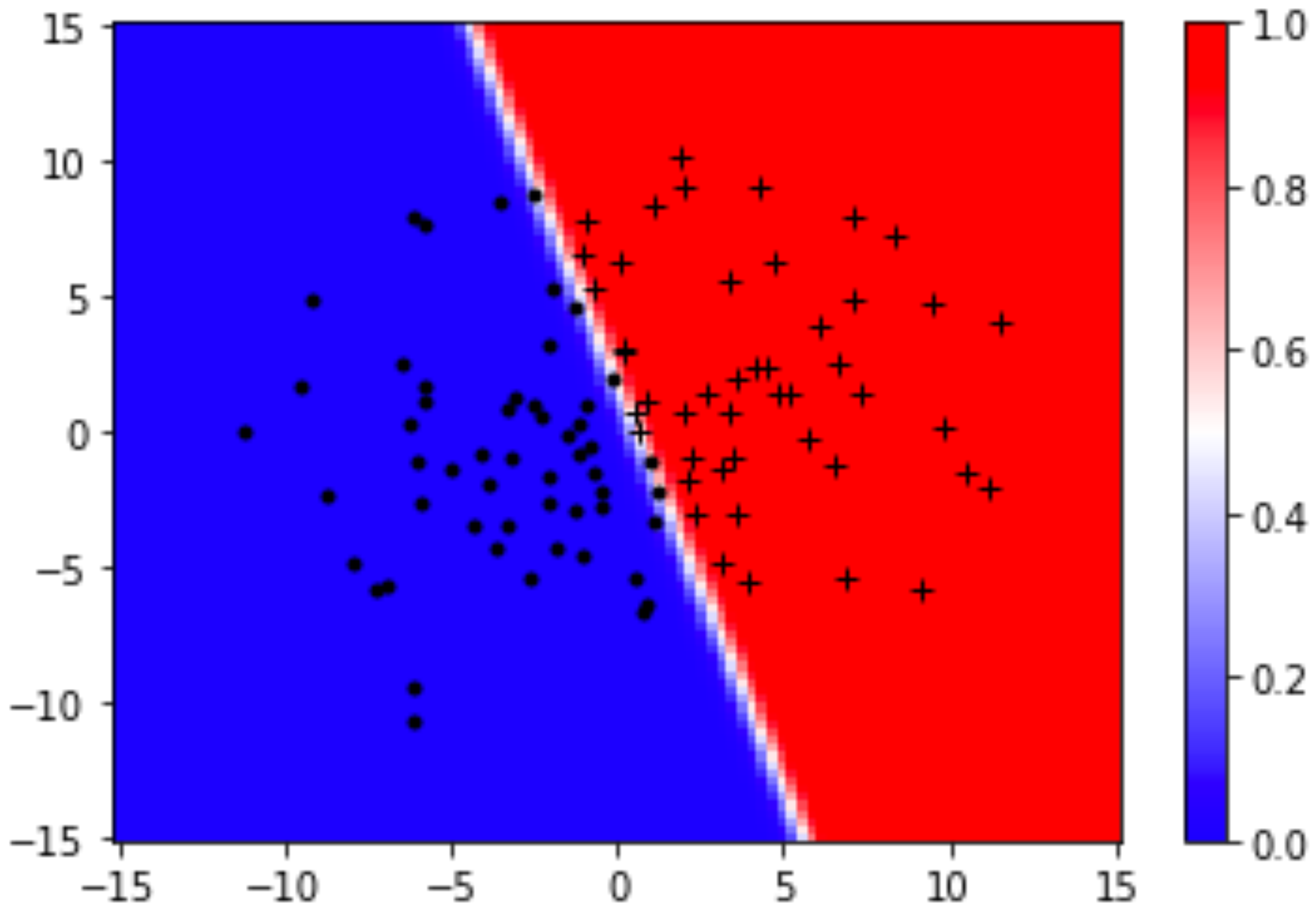
- In this lecture we are going to talk about
    - Deep Learning Regression
- Regression uses all the usual deep learning tools
    - Tries to solve a different problem than other DL lecture
    - Additionally it combines many of the concepts in fitting
- Lets review previous lectures to understand

Deep Learning

- In the past lectures we focused on :
  - Deep learning based classification

How do I separate to classes of points?

A



B

Deep Learning

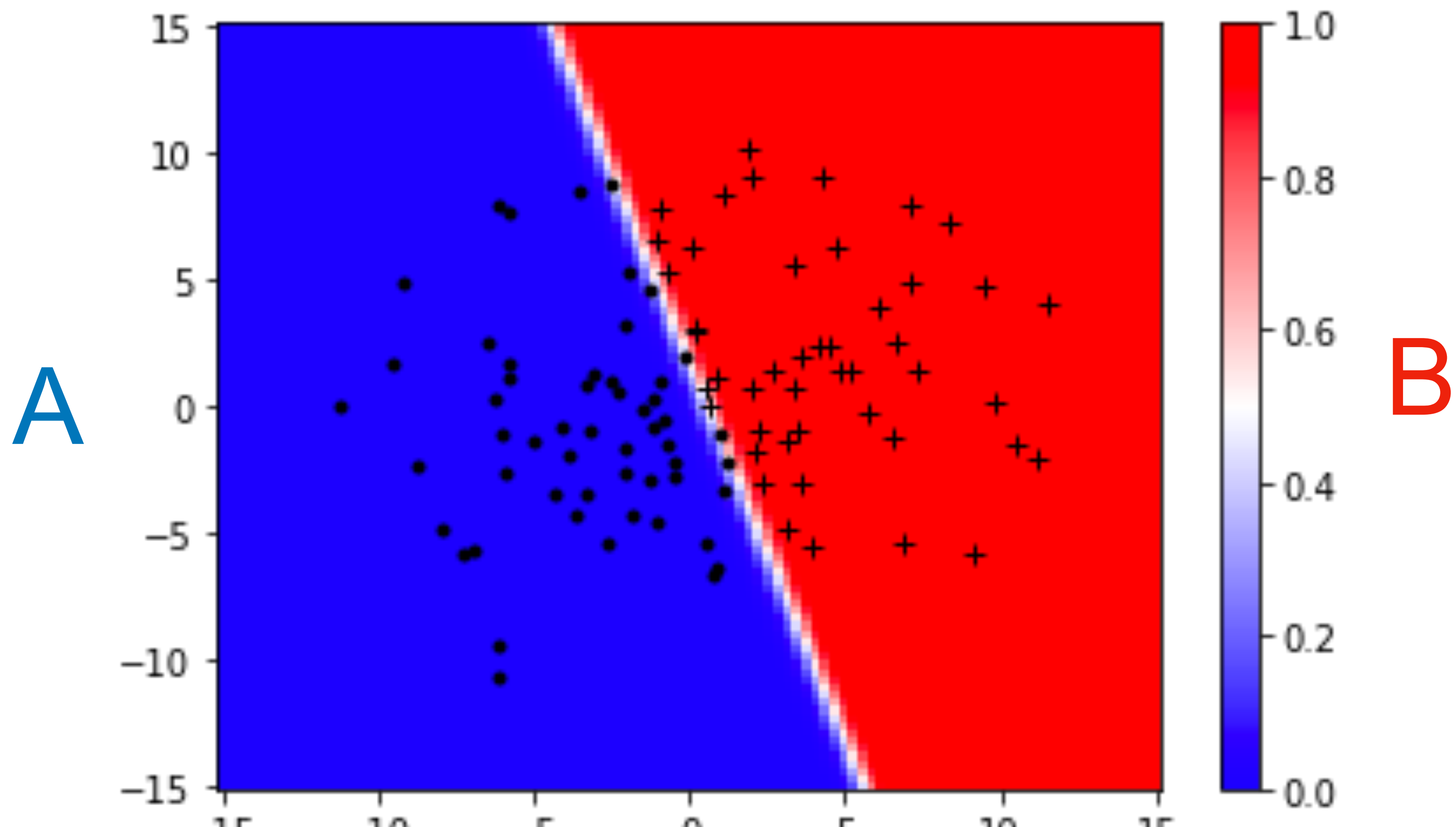
- In the past lectures we focused on :
    - Deep learning based classification

How do I separate to classes of points?

Minimize Loss:

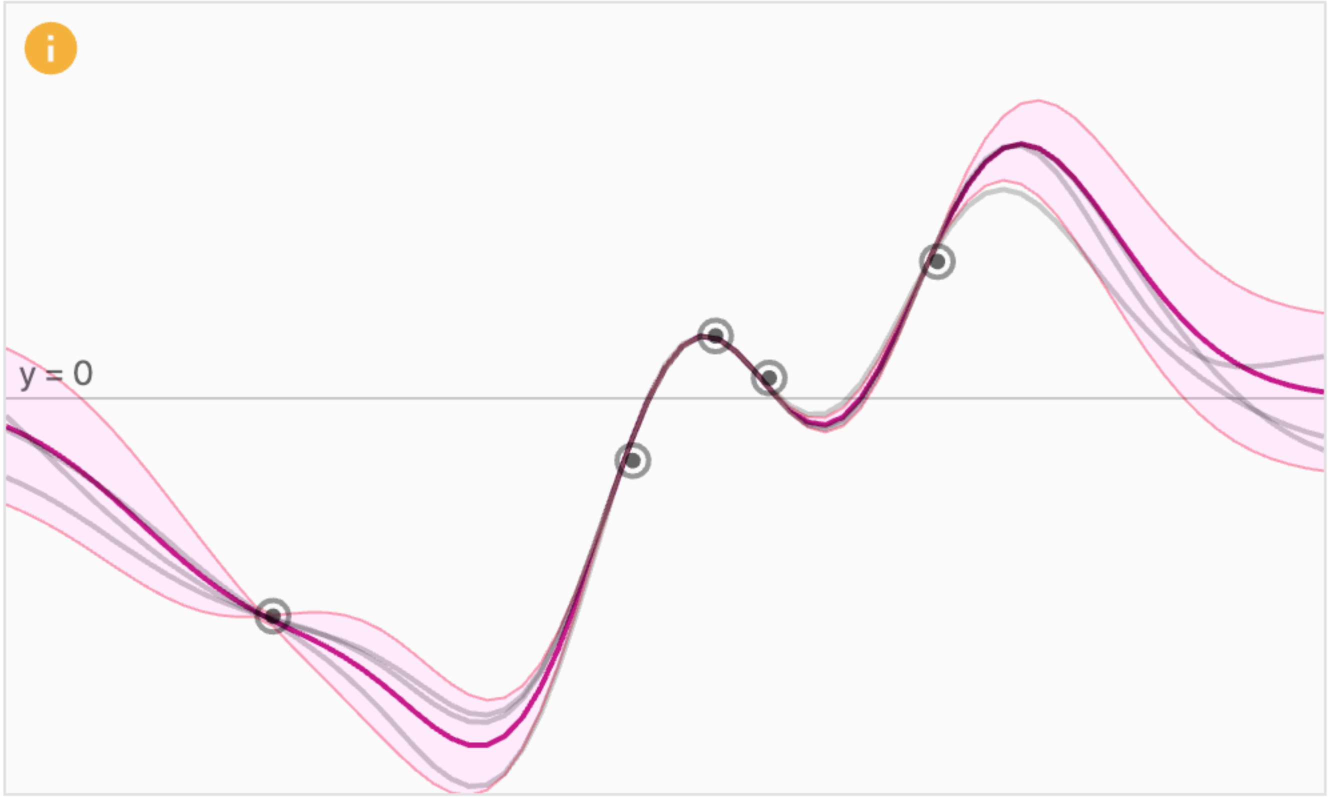
$$\mathscr{L} = B_{true}\log(p(B) + A_{true}\log(p(A))$$

$$\mathscr{L} = (1 - A_{true})\log(1 - p(A)) + A_{true}\log(p(A))$$



Interpolation

- How do I take a continuous set of points and connect them?
  - We have considered two separate approaches
    - Fitting a range of polynomials
    - Spline Interpolation and Gaussian Processes



Notebook

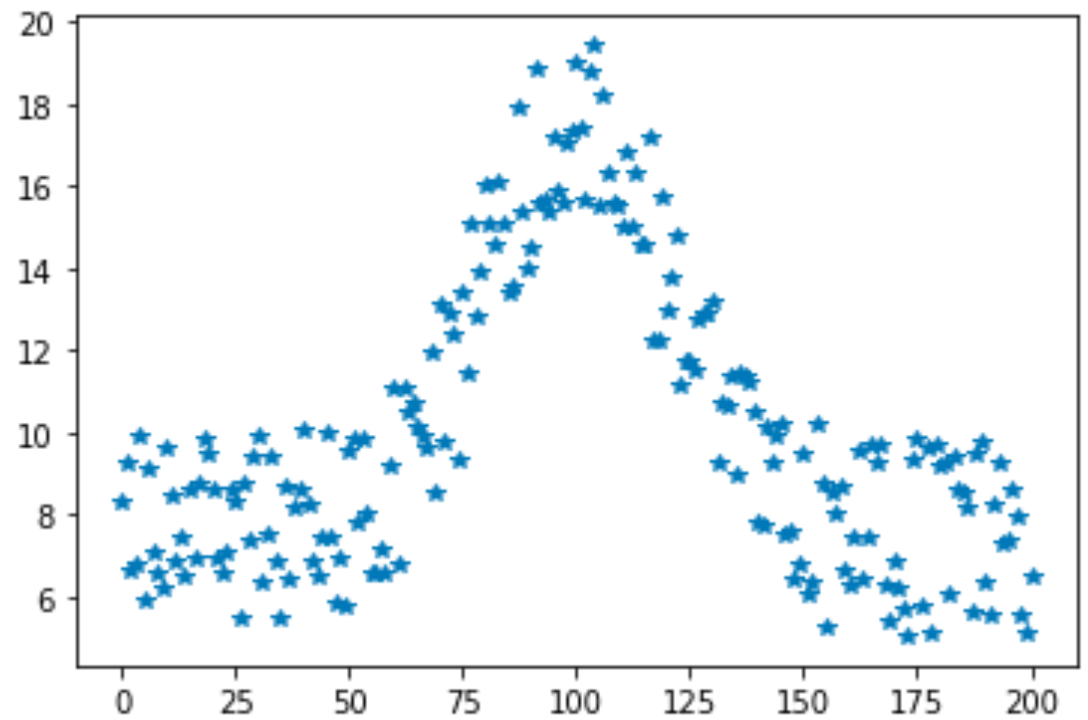
- https://colab.research.google.com/drive/1jmBNDxG2lLoYv2_WLawQbo2CGiJX91Oo?usp=sharing

https://github.com/MIT-8s50/course/tree/main/Lecture15

Fitting Any Distribution

- Between minimizing the likelihood and statistics we know what to do to get a fit that describes the data well. With interpolation and gaussian processes, we can connect the dots. However there are limitations what if we want to do something more complicated!

- **Challenge:** Fit the points below without guessing a function.



To the notebook

How do we do w/NN?

- With an NN all we are doing is a minimizing a loss
  - This loss can be any loss in the end
  - Really **Whatever we want**!
- A common loss is so-called Mean Squared Error
- $$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\boxed{y_i} - f(\boxed{\vec{x}})\right)^2$$

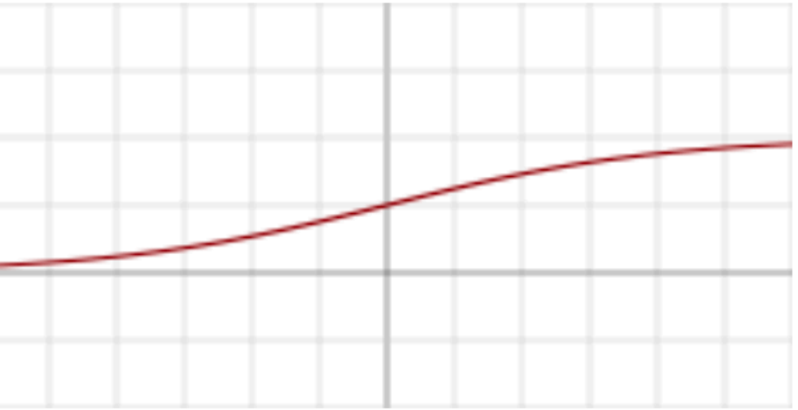
This is our input to our Neural Network it can be a vector of arbitrary size

This our target data in the training it can also be a vector of arbitrary size

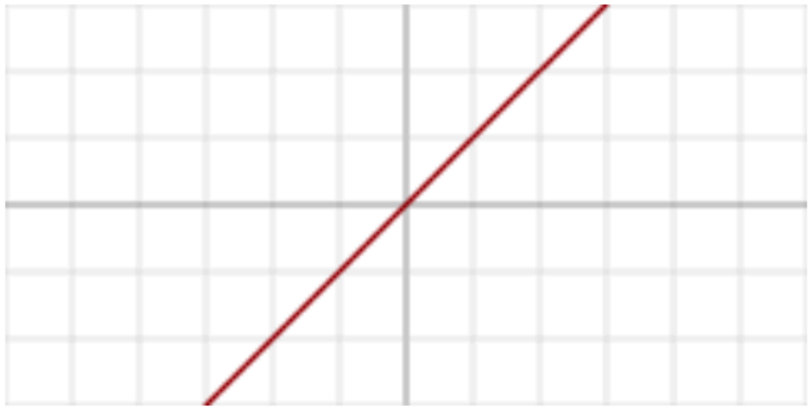
To the Notebook

Activation Functions

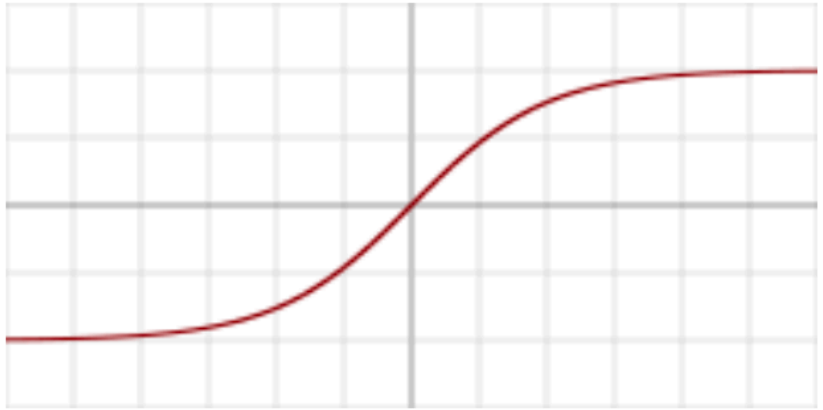
Sigmoid



Linear



Tanh

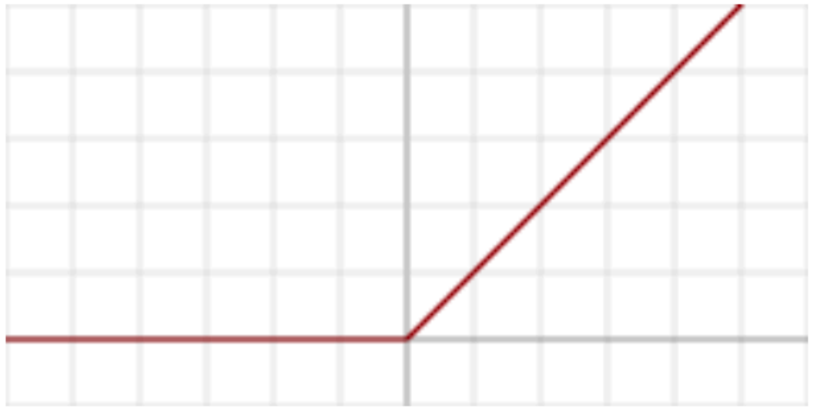


Softmax

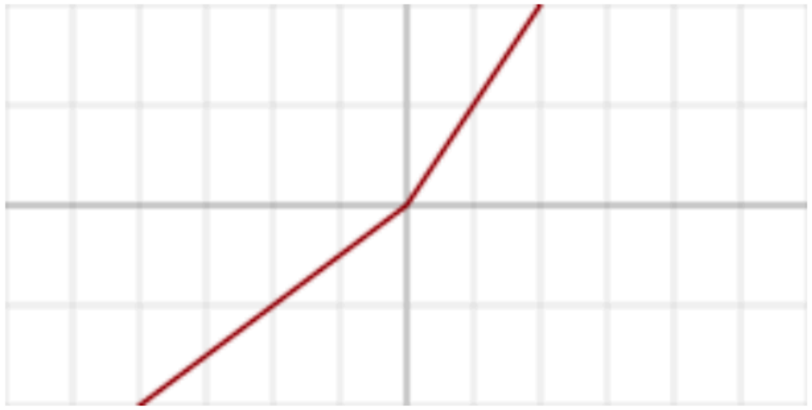
(multiclass)

$$\frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}}$$

ReLU

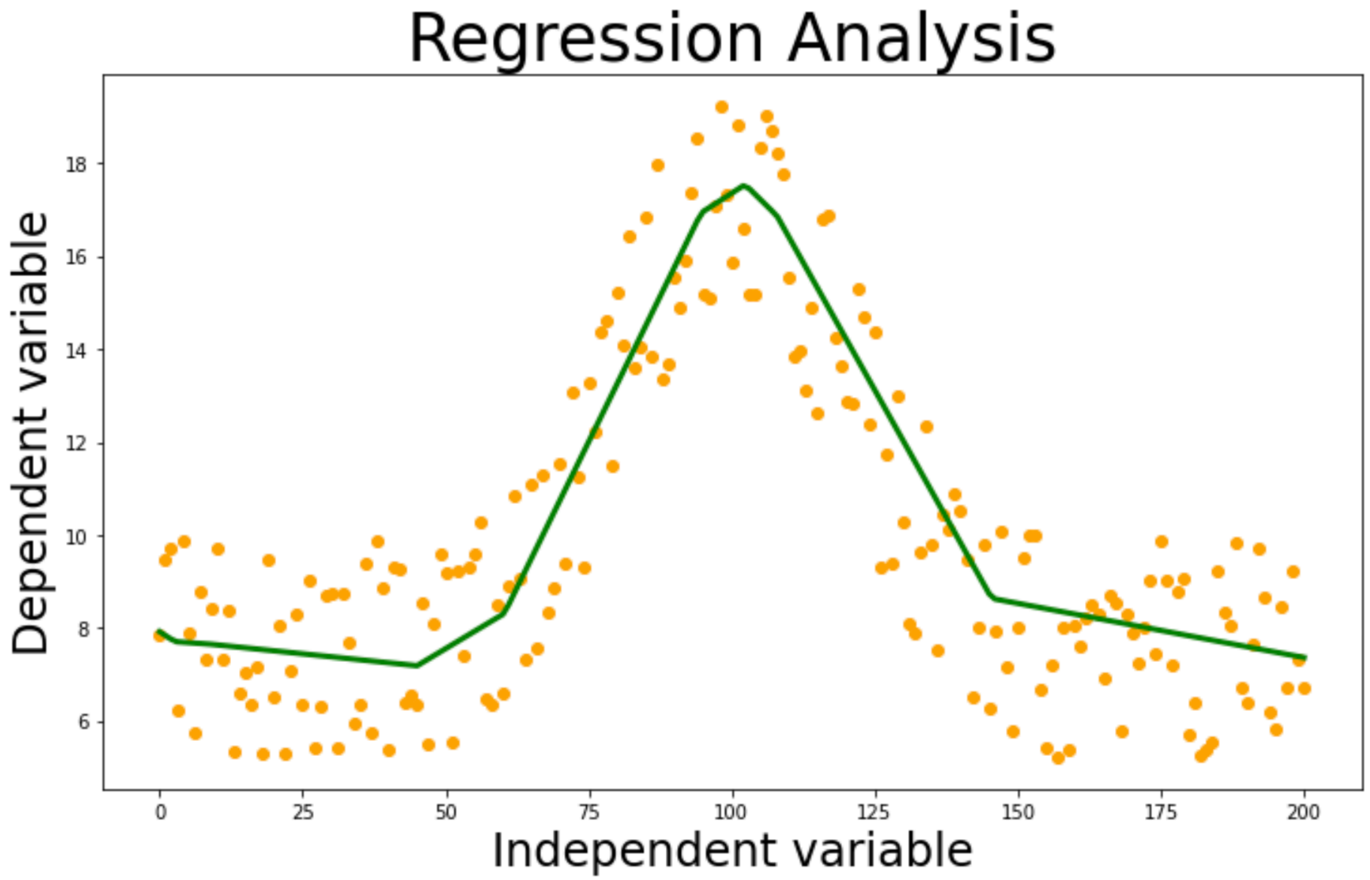


LeakyReLU

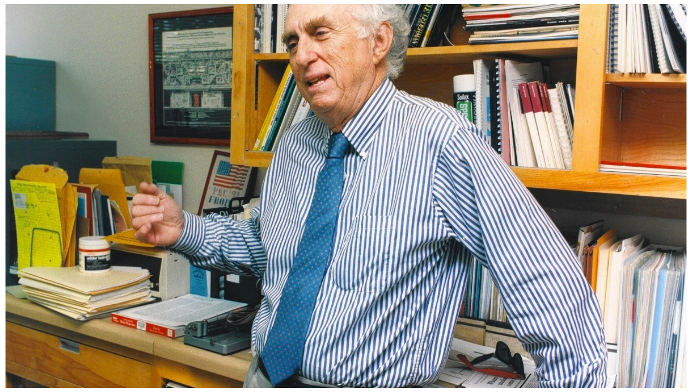


Parameter Extraction

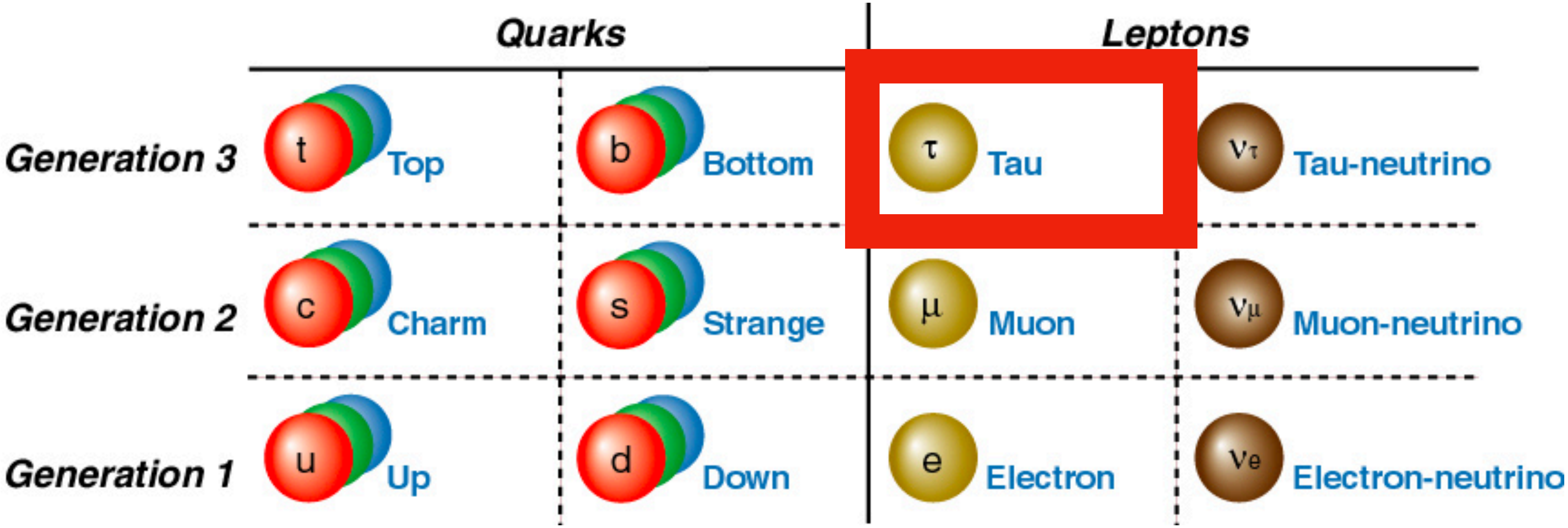
- Despite being able to fit such a distribution
  - There is a limit to how much we can do
  - The functional form for this distribution is complicated
    - To get a mean and a resolution, requires reverse engineering



Lets Solve A Real Problem



- Let's look at the tau lepton



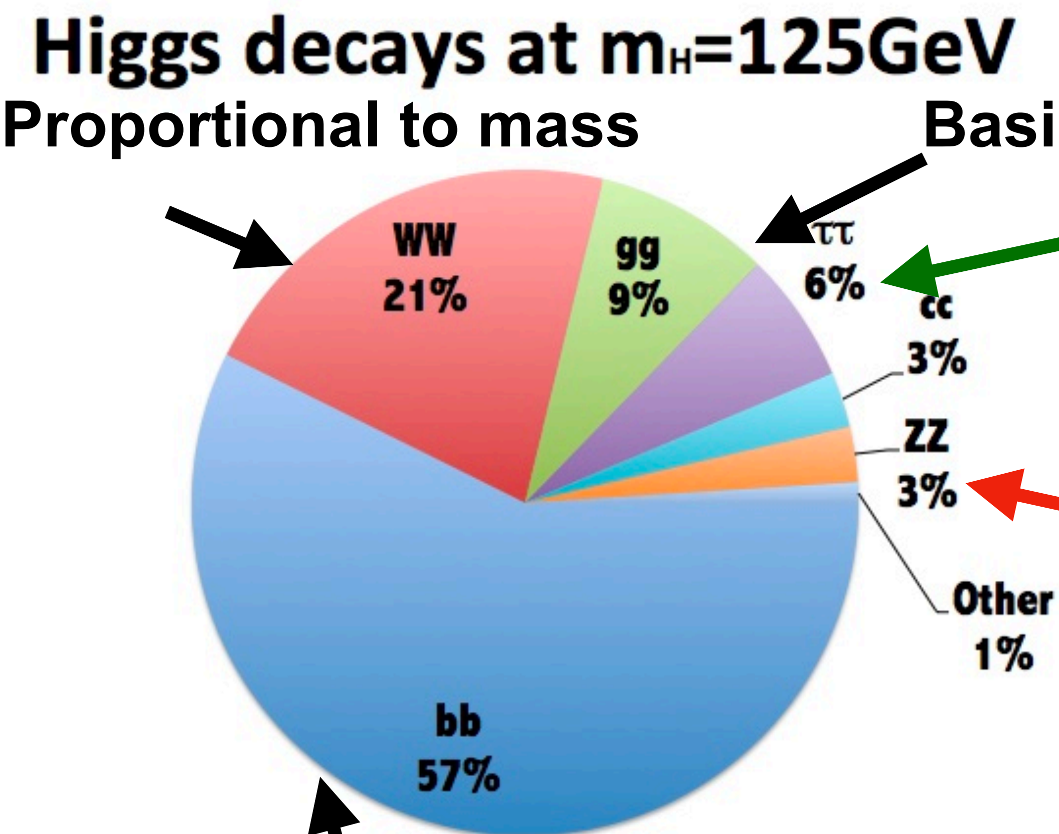
The Tau is the heaviest of the leptons (electron-like)
What makes it so special?

Higgs Decays

Not Proportional to mass

Basically impossible to probe

Our Best Bet for heavy objects

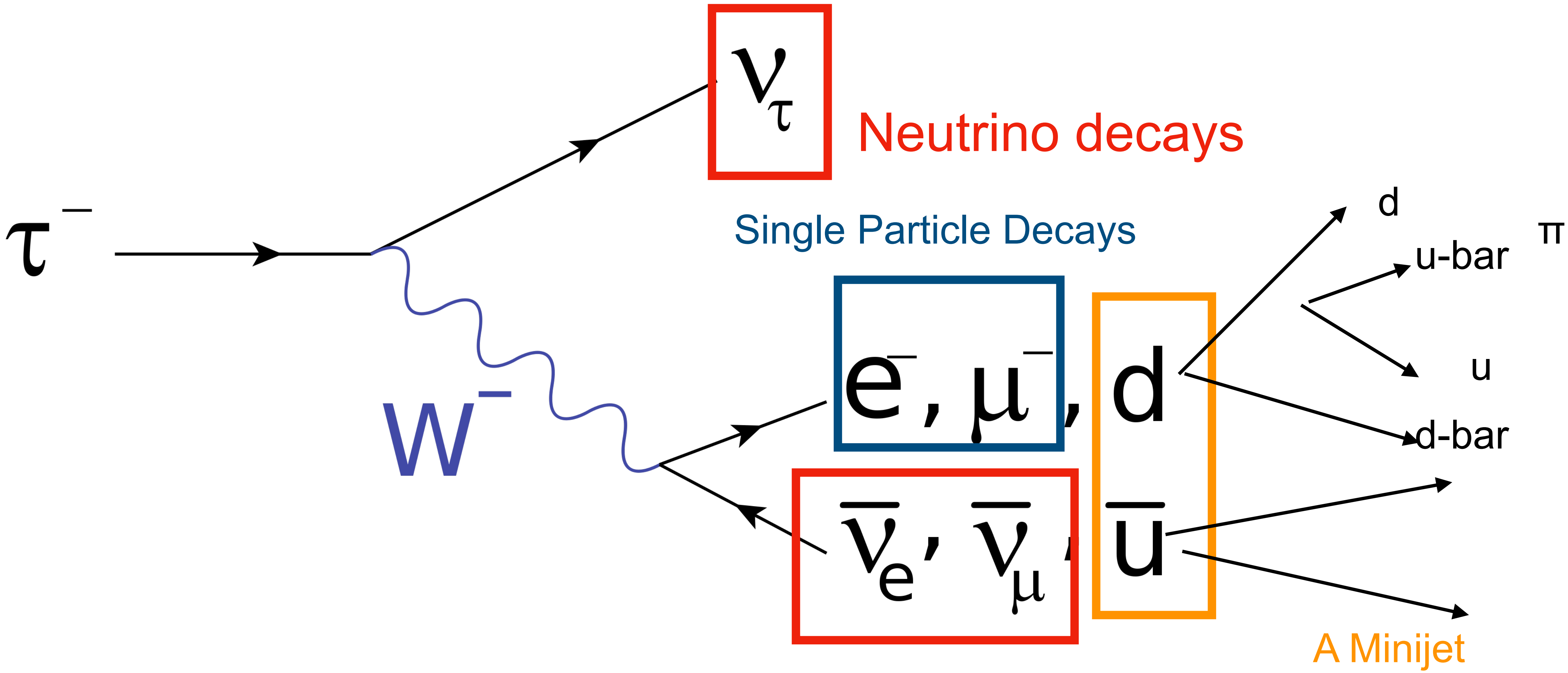


Main Discovery channels (<2% of Higgs)

Almost impossible to probe

- Higgs probability of decay to quarks and leptons is proportional the mass of the particle. Taus are very heavy particles. Higgs decays to them 6% of the time. That’s great. **It was the first channel we could actually probe the proportionality to mass.**

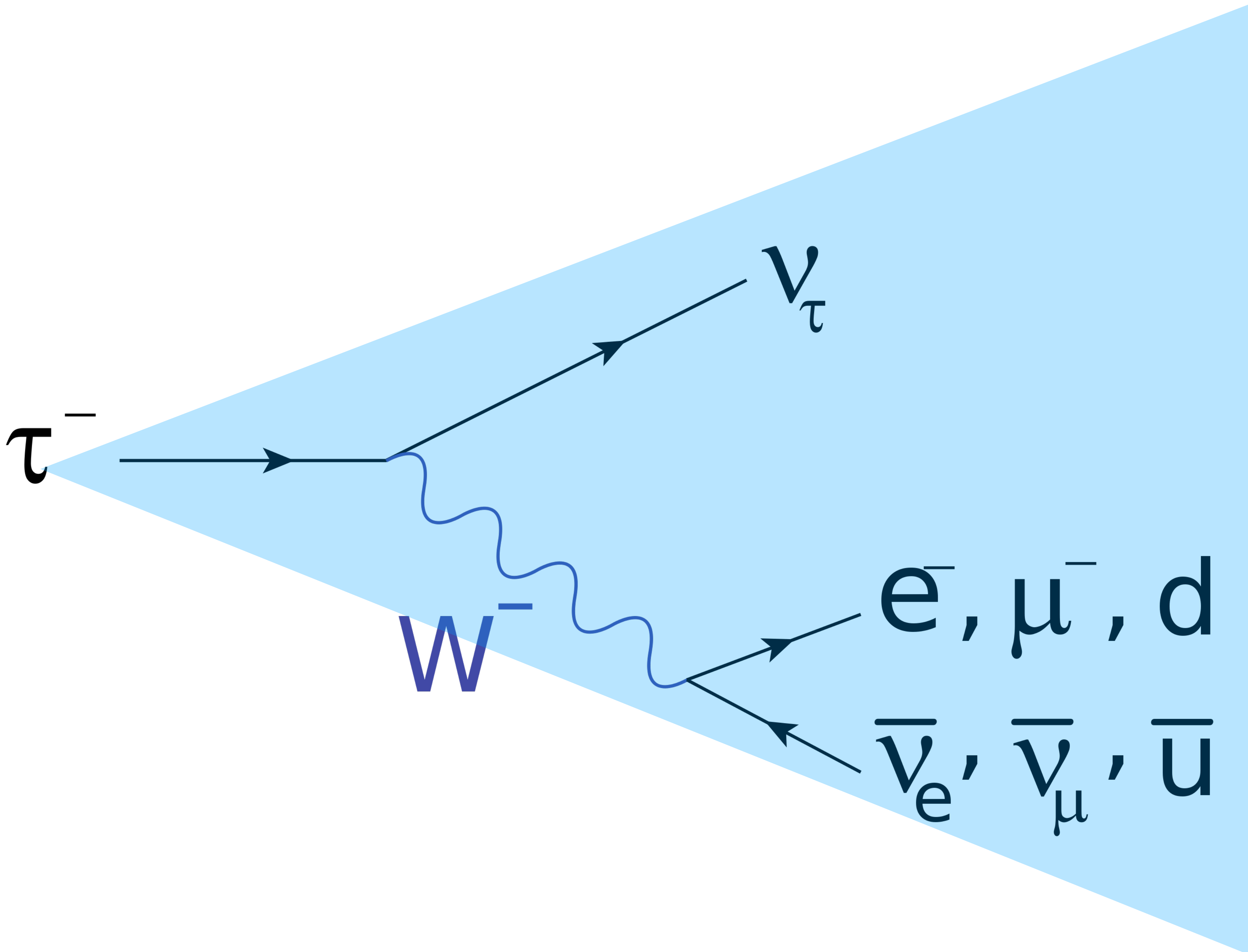
Tau Decays



Neutrino Decays: The probability of a neutrino interaction is too small to see at the LHC. These particles are invisible

Single Particle decays:These events just give us one particle e or μ
Minijet: Decays to quarks give us a shower of particles in small jet

Problem

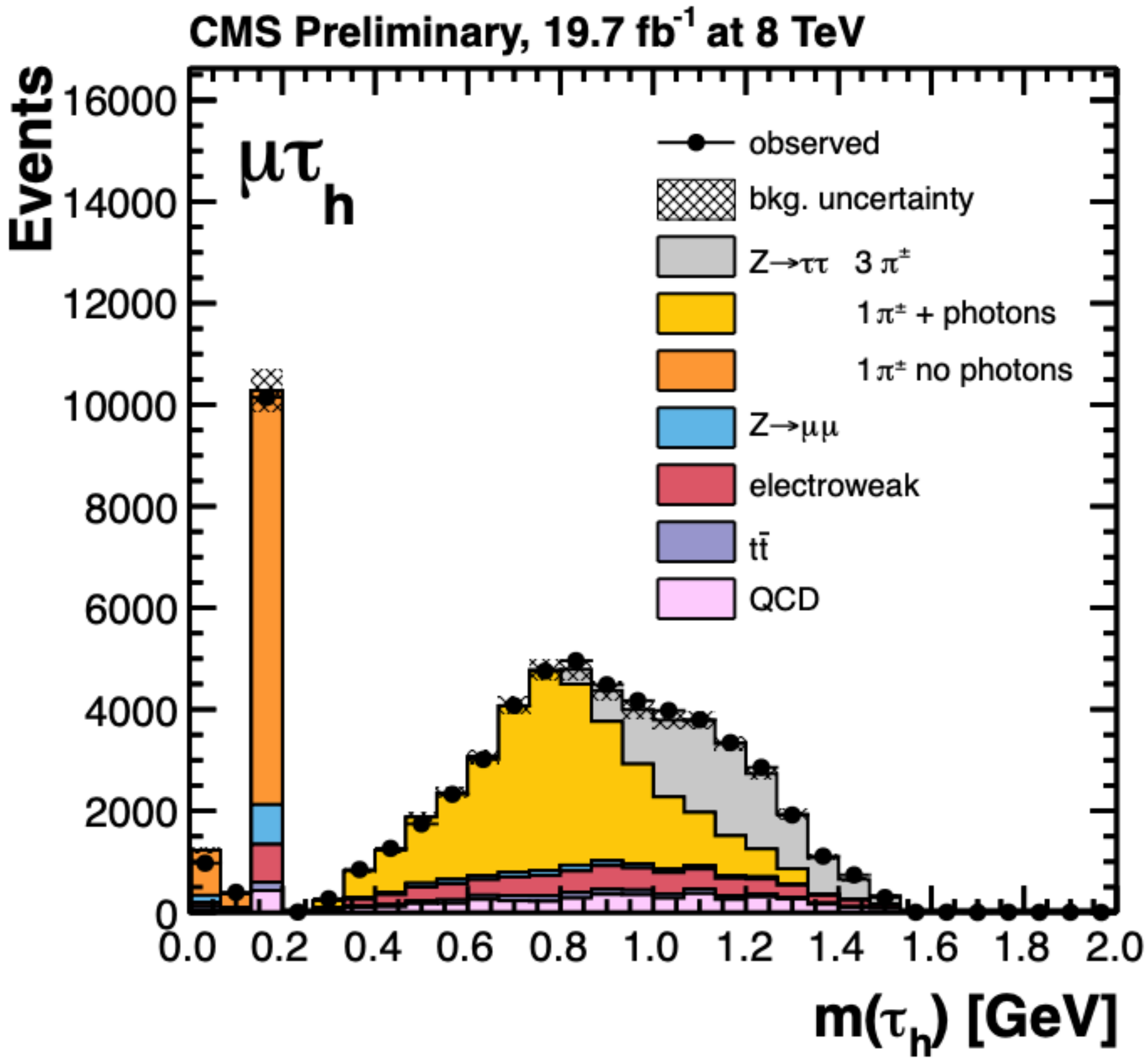


Take a jet
And Sum all the particles

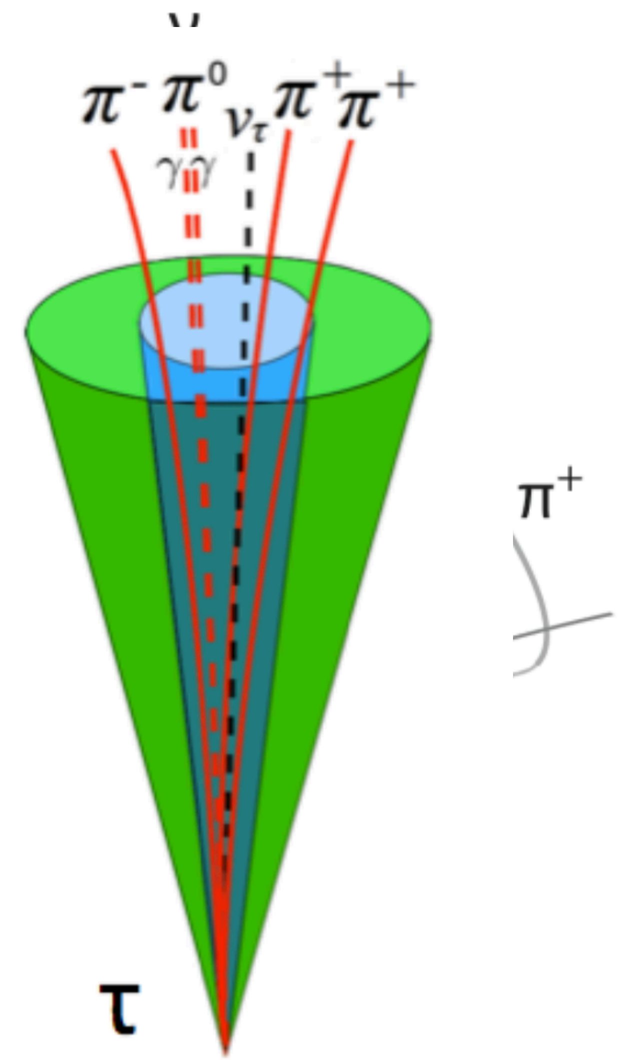
Can we go from
Jet p→Tau p

Can we guess direction of the neutrinos and reconstruct the original tau energy?

How does a Tau decay

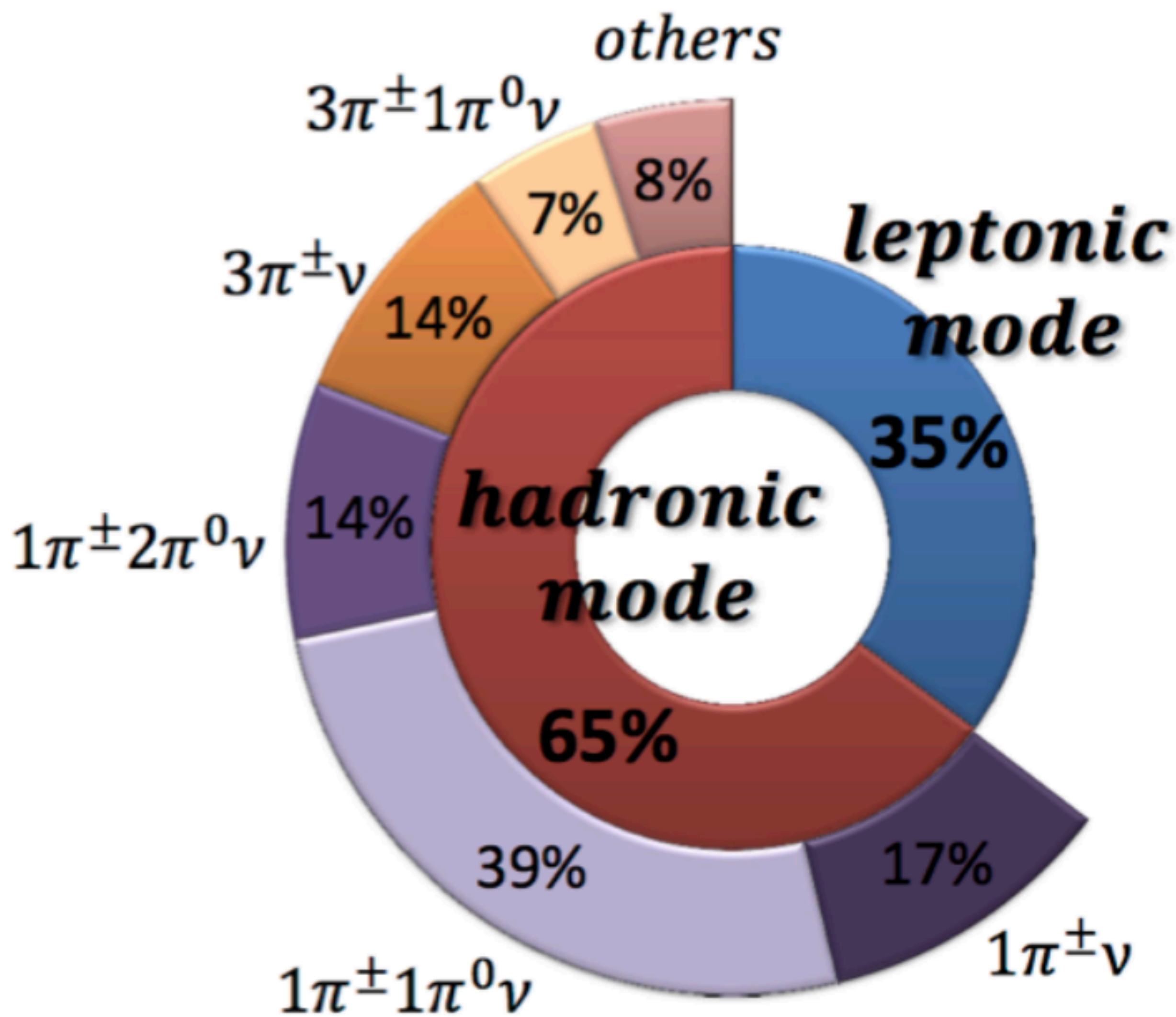


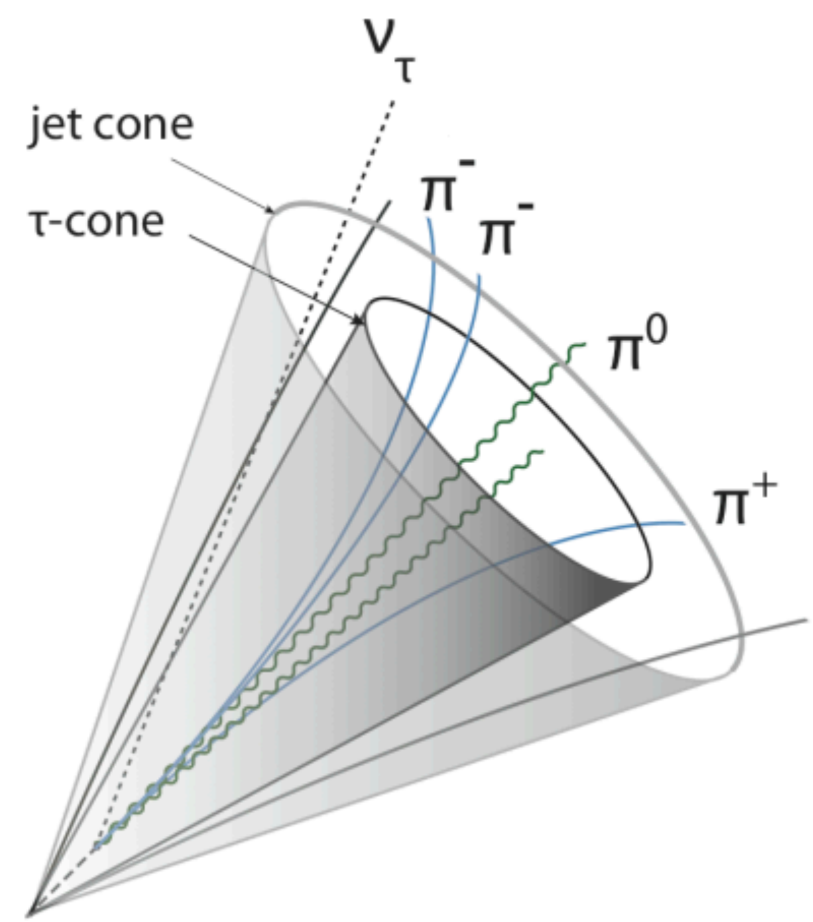
$$m_\tau = 1.76 \text{ MeV}$$



Taus have a small mass, which means they can be found within a small cone

How does a Tau decay

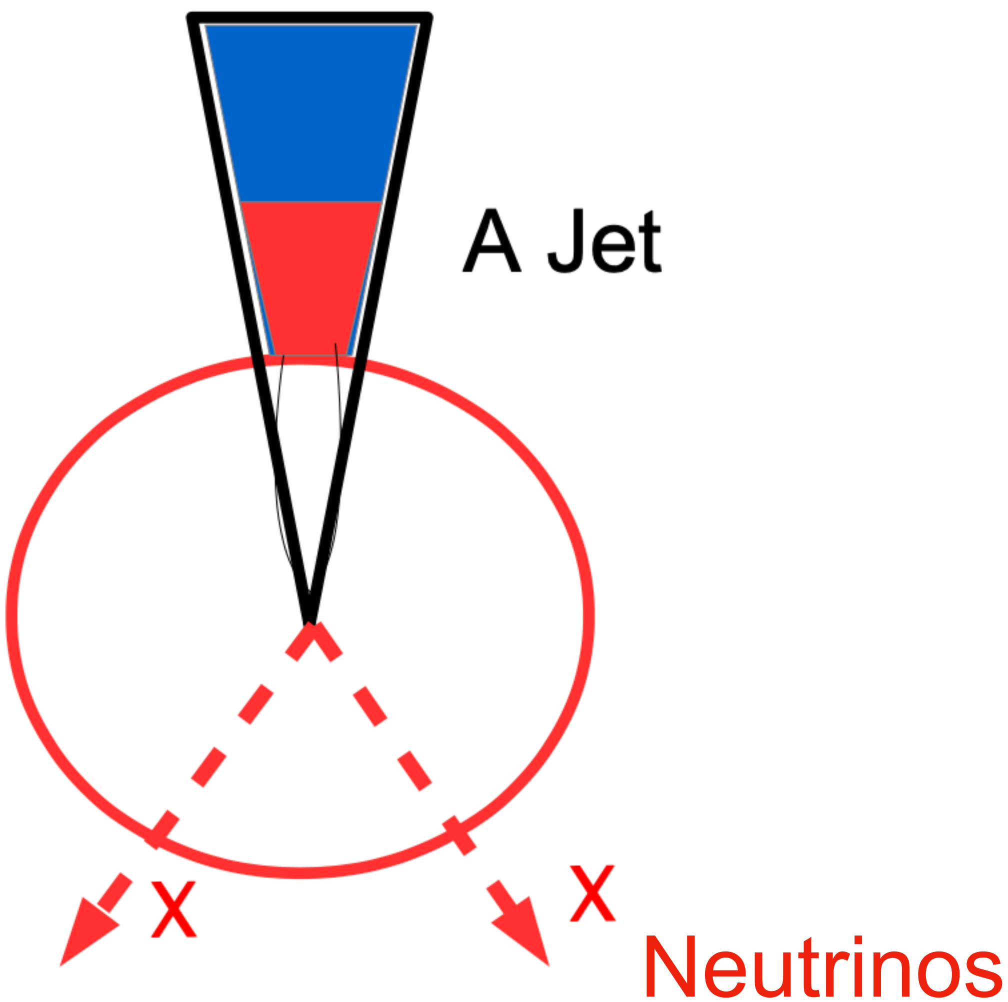


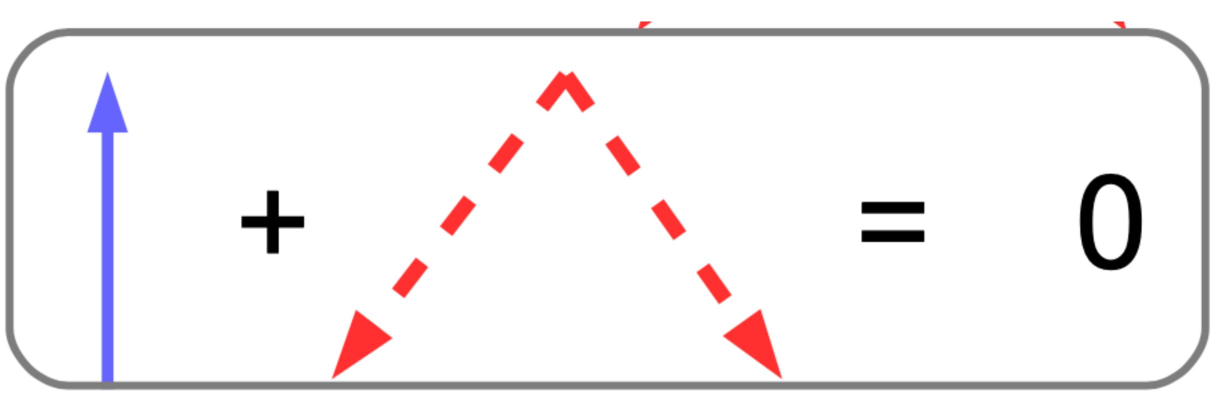


$$\pi^0 \to \gamma\gamma$$

We are looking for collection of 1-5 particles
Neutrino will fall in the same cone

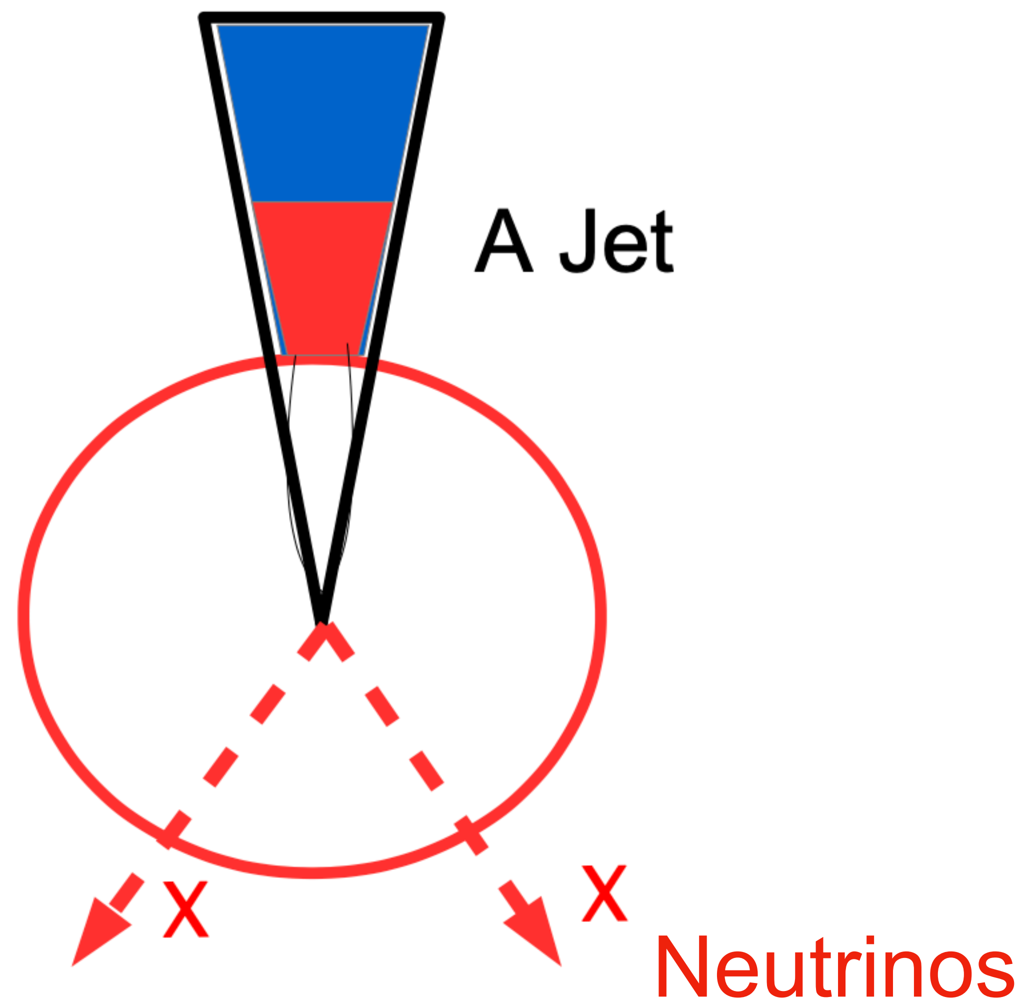
What we did for that result

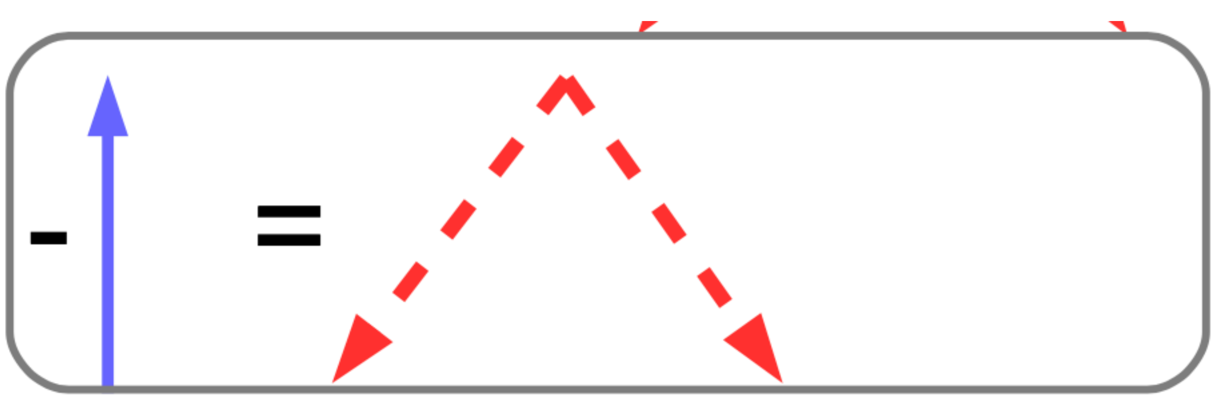




Conservation of transverse energy

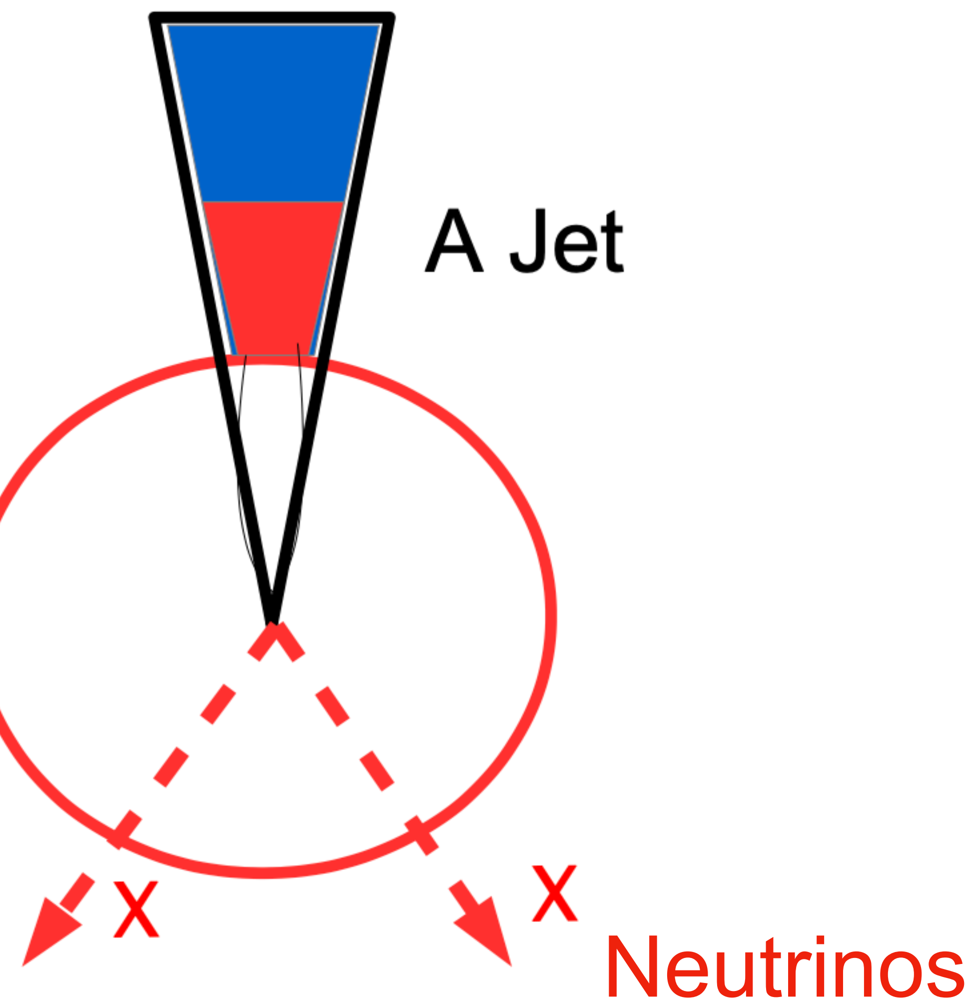
What we did for that result





Conservation of transverse energy

What we did for that result

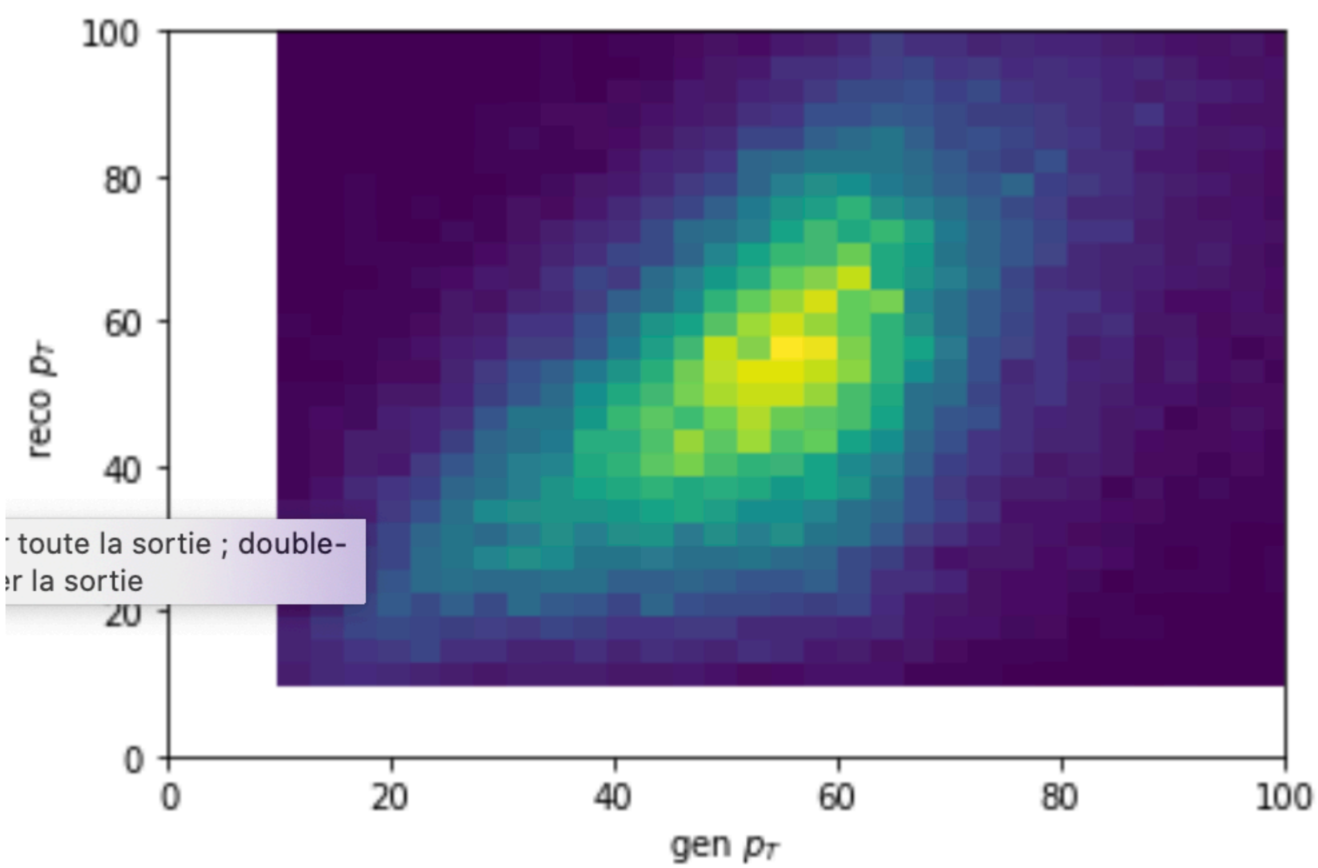


$$-\Sigma_{\text{All particles}} \vec{p}_T = \overrightarrow{MET}_{(E_T^{Miss})}$$

Conservation of transverse energy

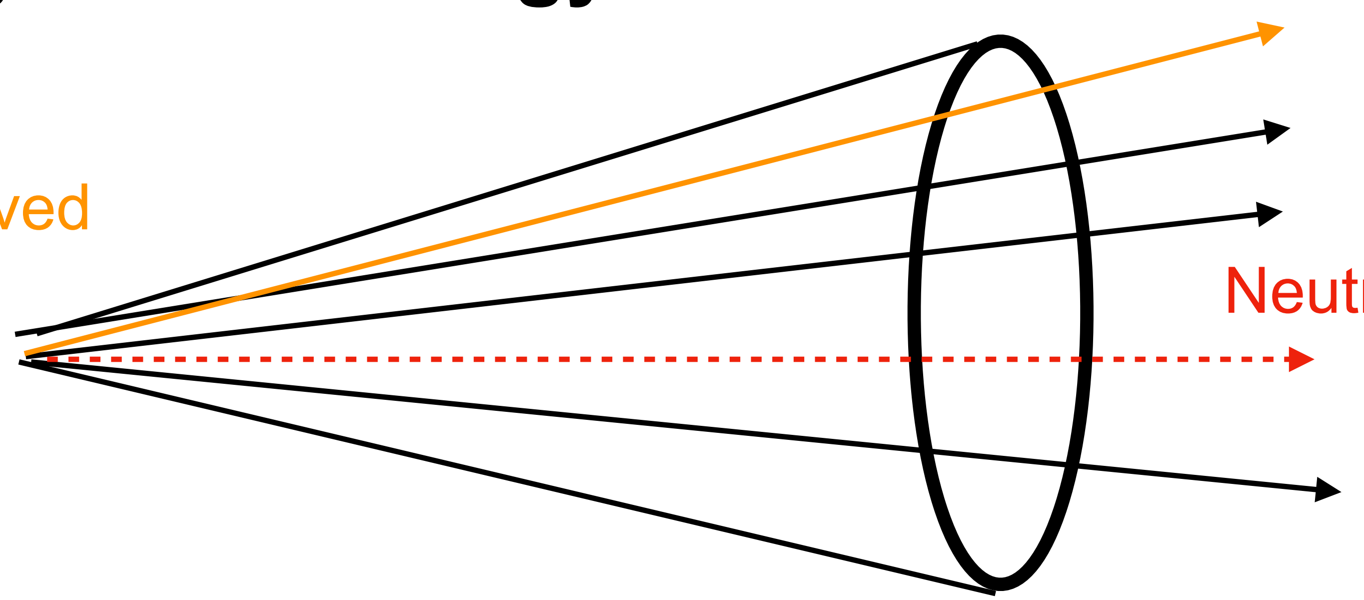
Some Correlation

- In this case, we want to try to use the tau momentum
  - Goal here is to rely on the fact that there is some correlation
  - The tau momentum can predict the total tau energy



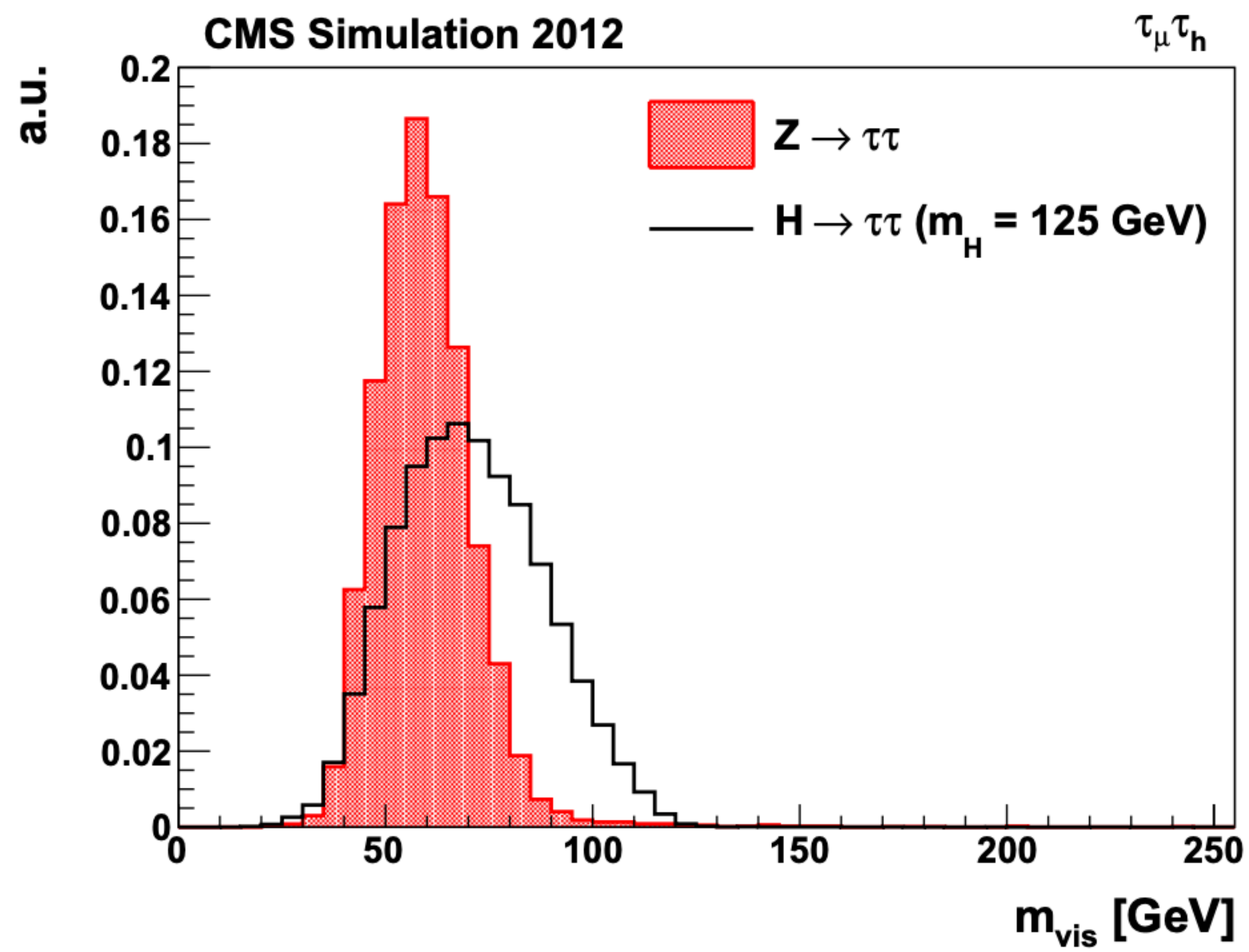
NN Problem

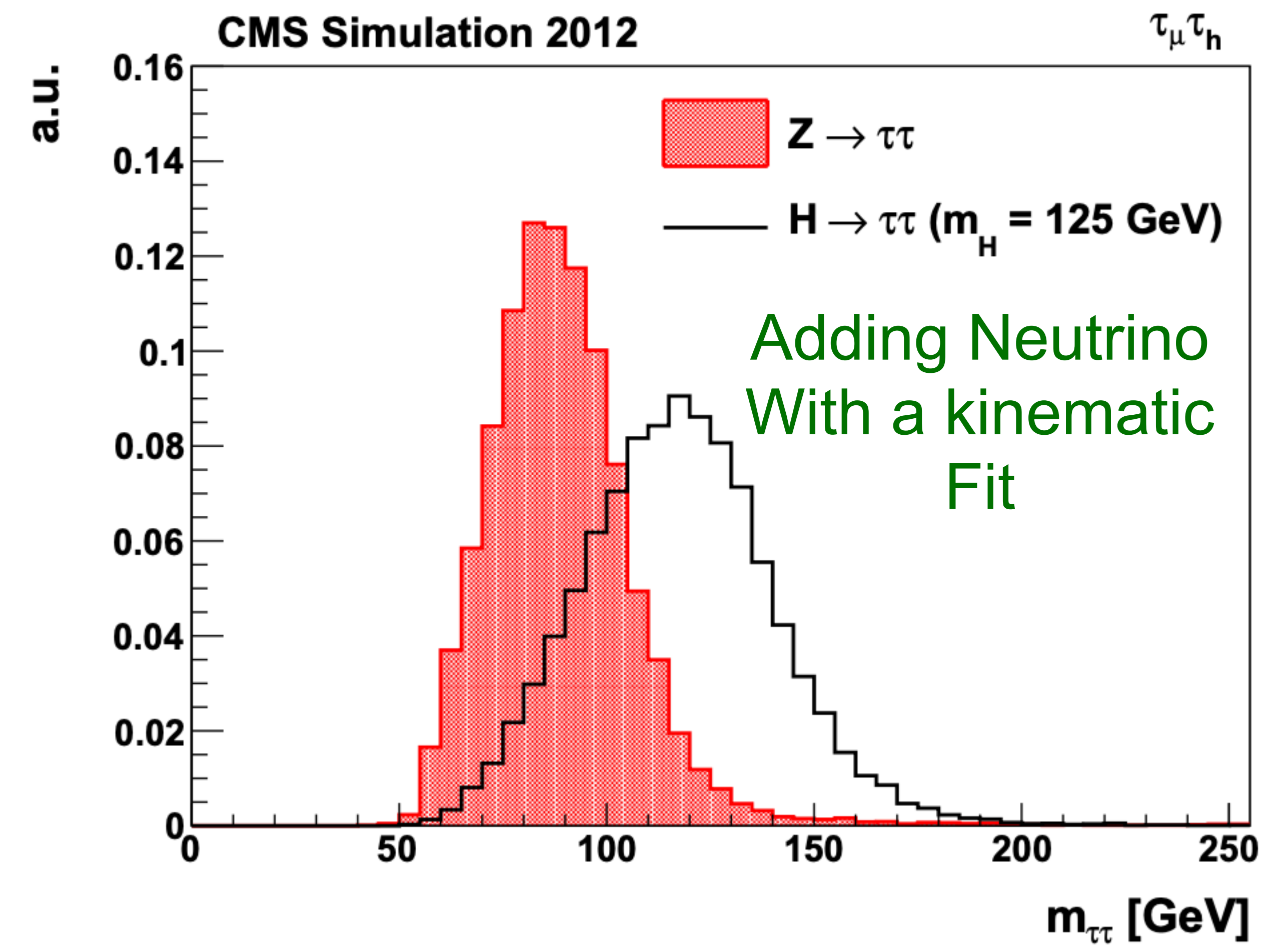
Can we guess direction of the neutrinos and reconstruct the original tau energy?



- simple: $p_T\ \tau = NN(p_T^{jet})$
- reduced scale: $\frac{p_T\ \tau}{p_T^{jet}} = NN(p_T^{jet})$
- Complex: $\frac{p_T\ \tau}{p_T^{jet}} = NN(\overrightarrow{p_1}, \overrightarrow{p_2}, \overrightarrow{p_3}, \overrightarrow{p_4}, \overrightarrow{p_5})$

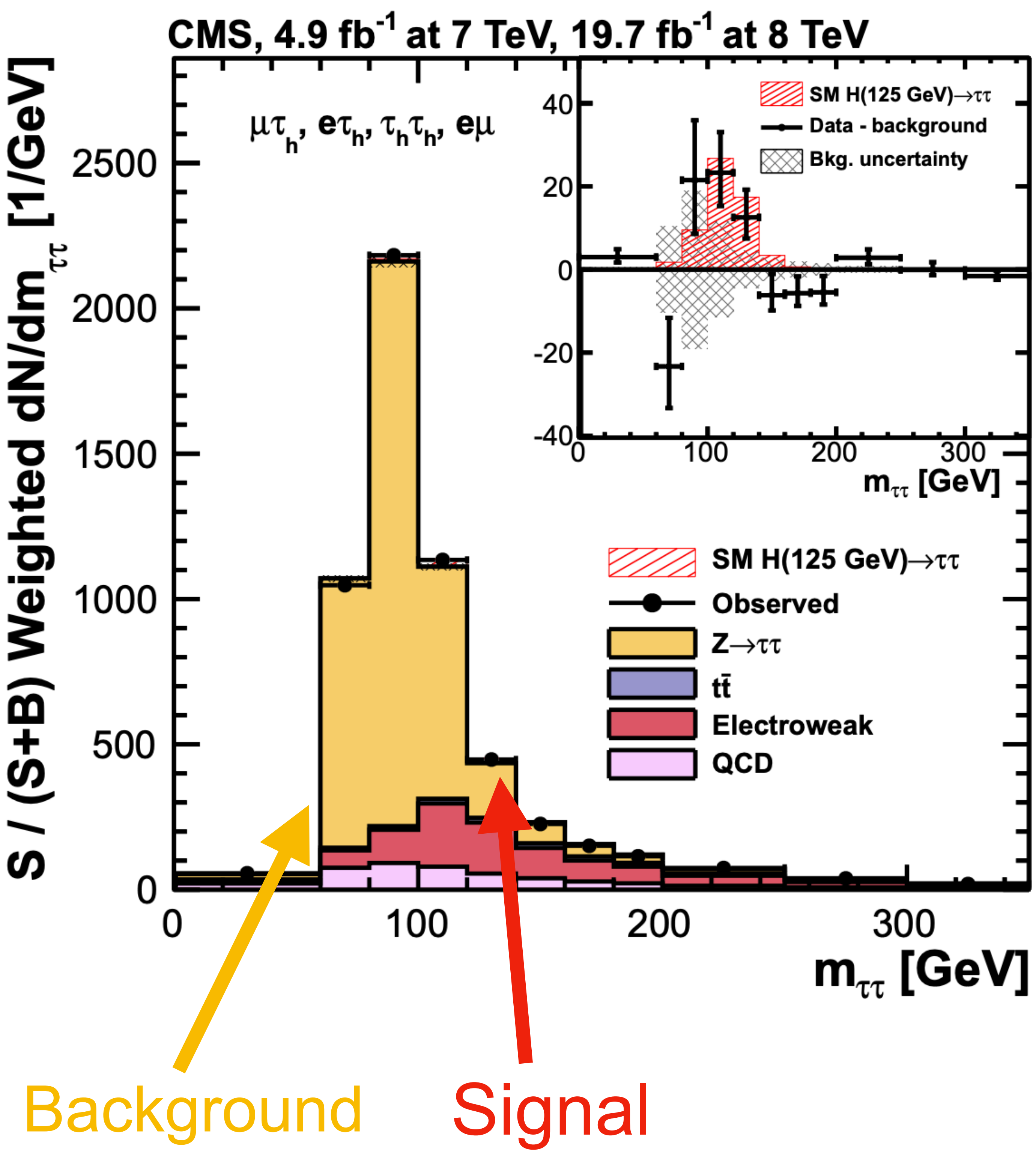
Why this?





- Finding the Higgs boson is hard we need to separate
  - Higgs boson mass peak from the Z boson mass
- When Higgs discovered didn't have the NN tech to add neutrinos

The Full Challenge



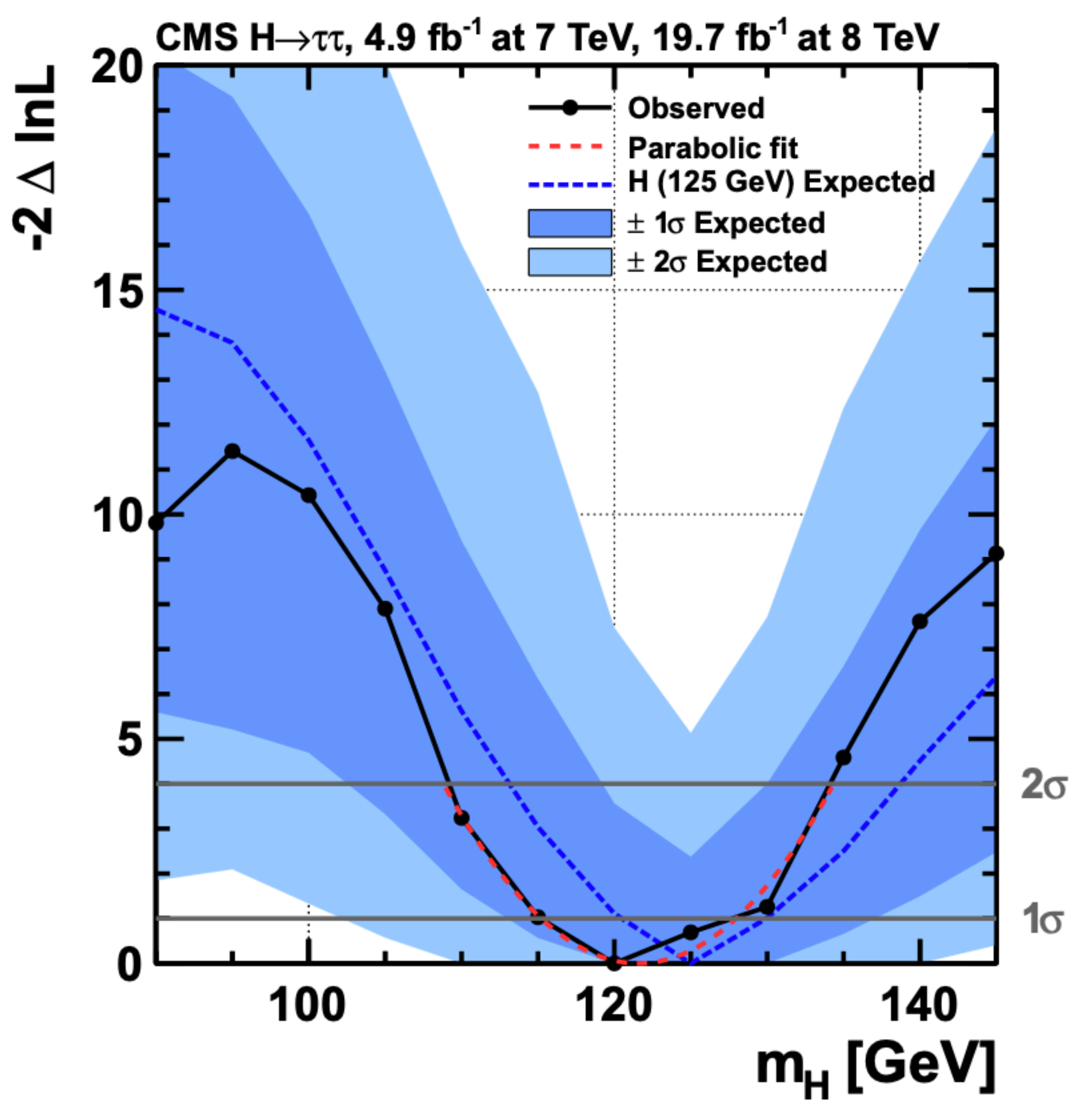
Plot is a composite of 70 separate fits

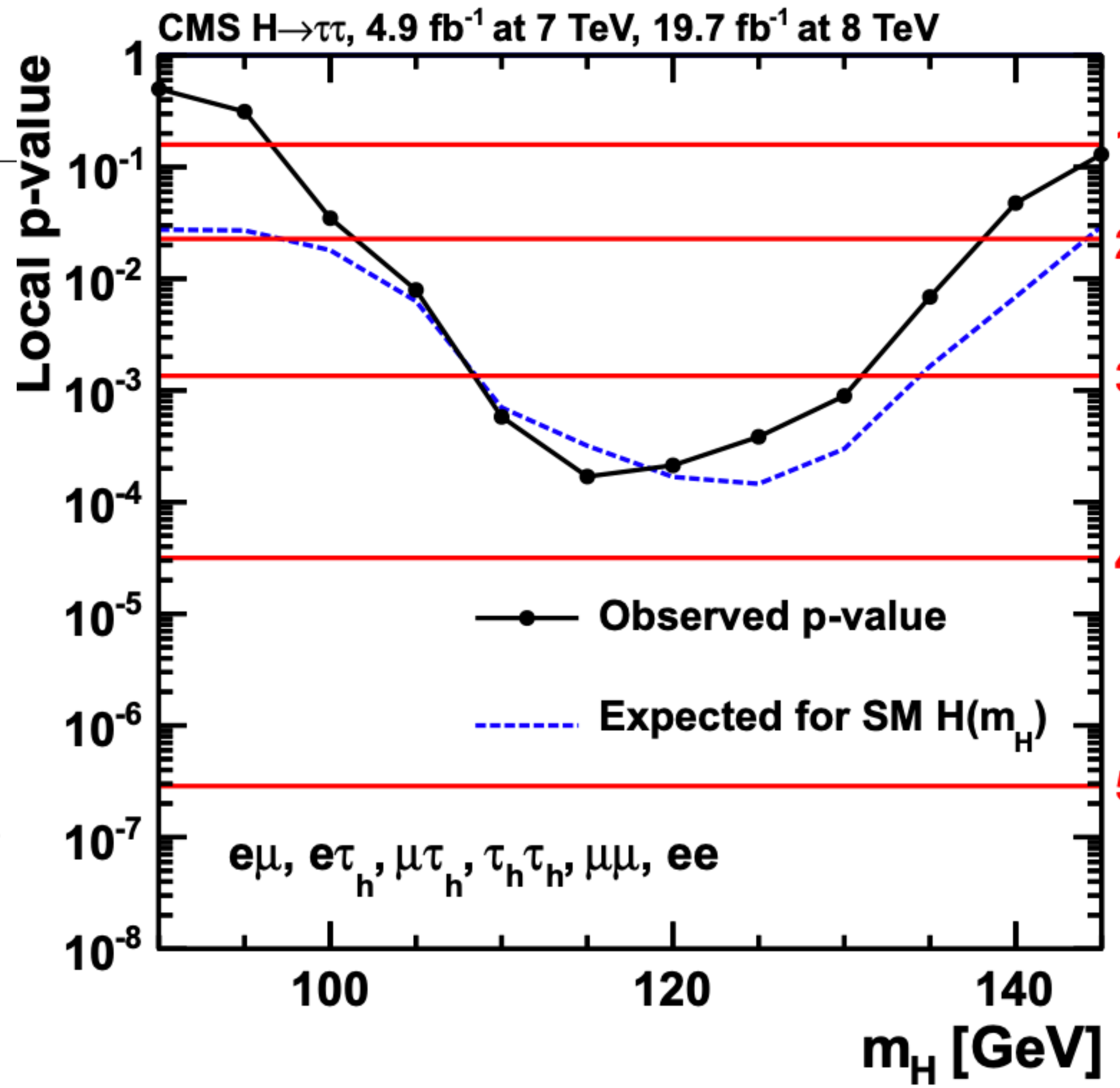
There were > 2000 Floated parameters

Fit took 24h to run

Higgs to Tau Tau Bound

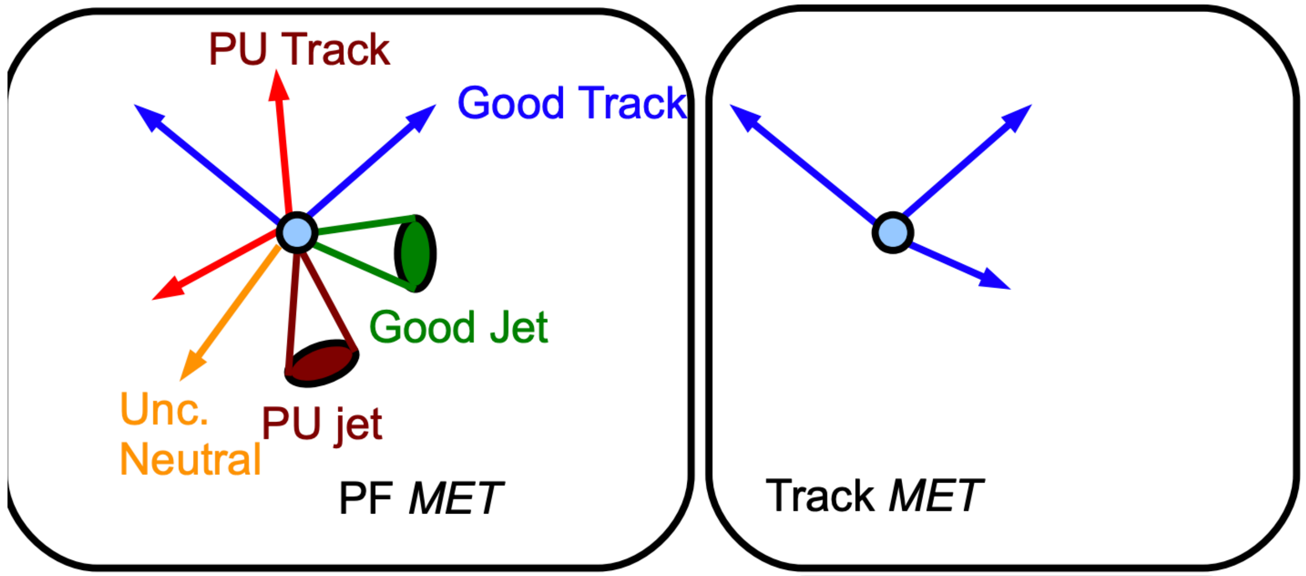
- Best fit





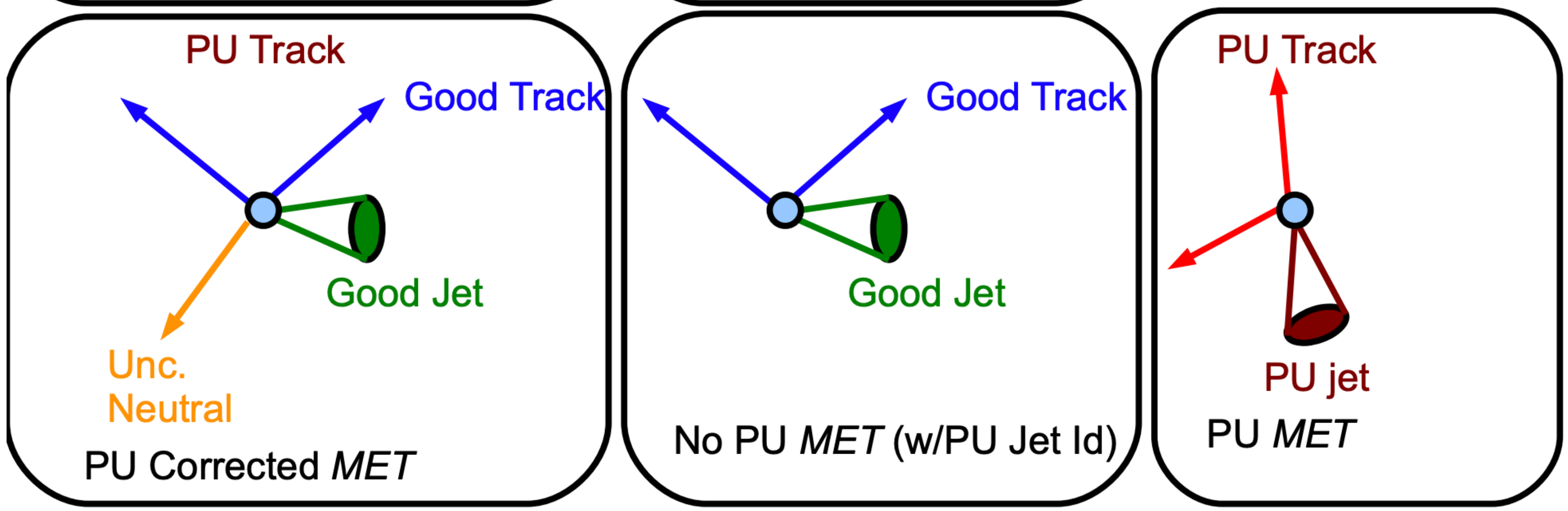
What we did for that result

MET Variables (Inputs to a new Algo)



5 Different *METs*
Each targeted on a different aspect

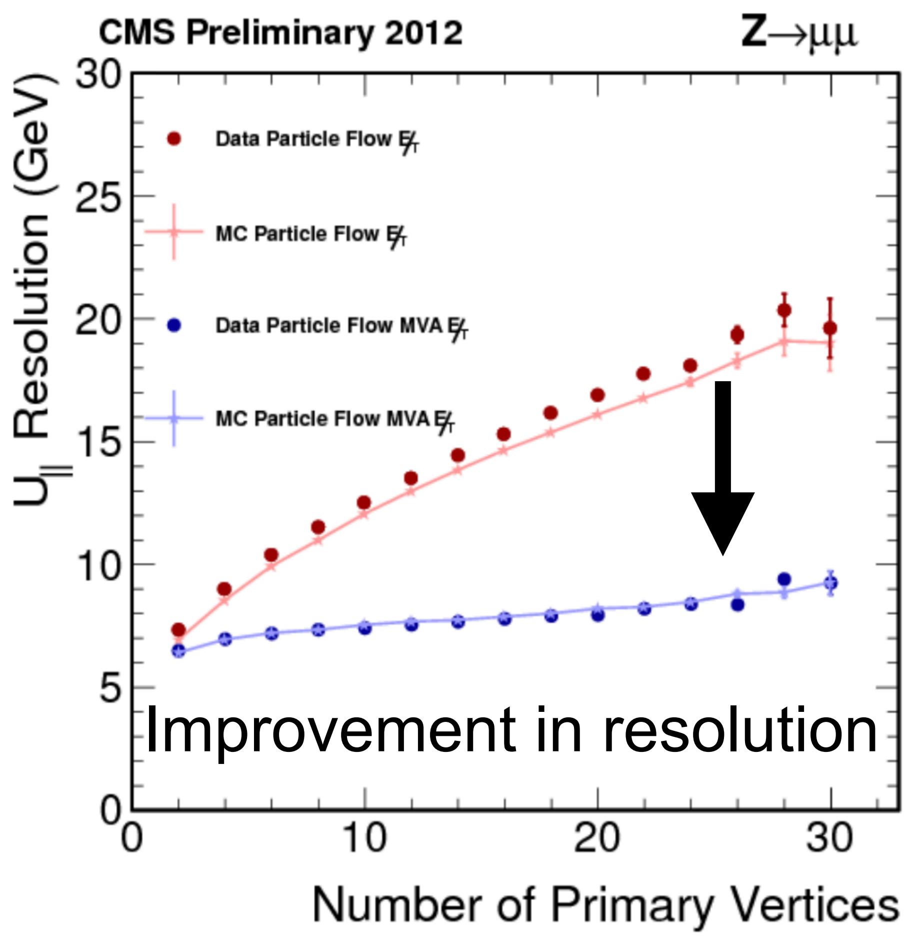
Take recoil
(except PU *MET*)



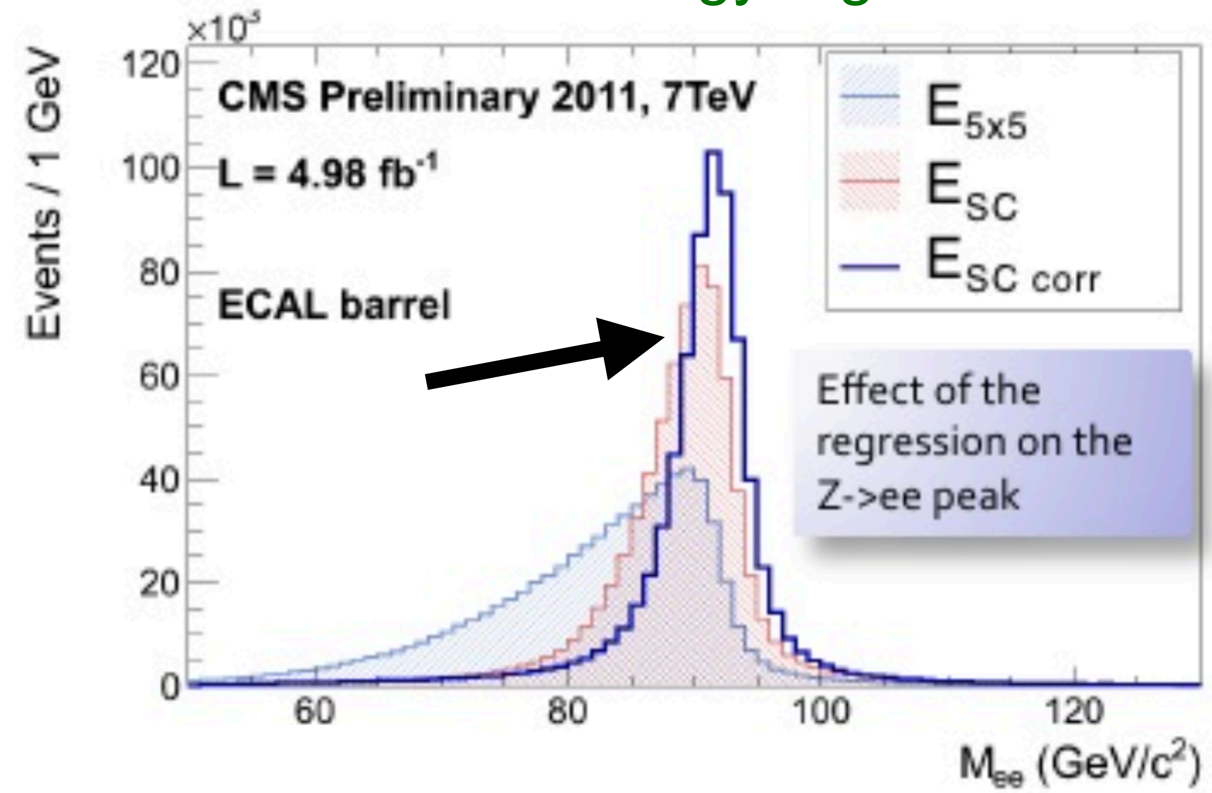
All of these separate *MET* calculations were put into 1 single regression

- We did end up a using an NN regression for that plot

Impact of Regression



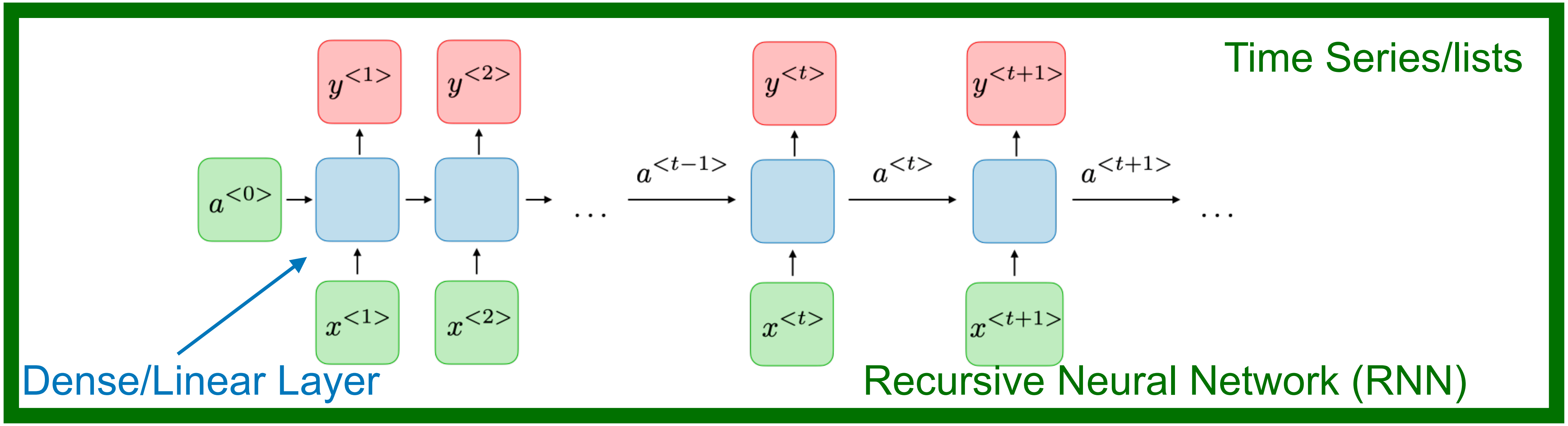
Photon energy regression

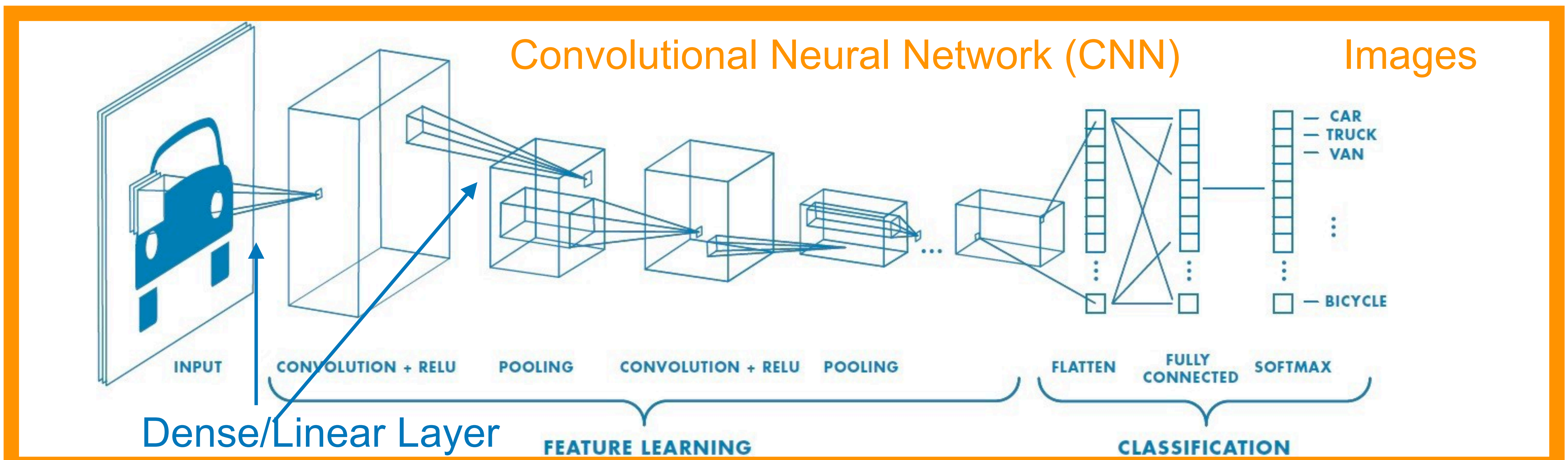


- Regression ended improving the Higgs sensitivity by 30%
  - Both in the diphoton channel and Higgs to tau leptons
  - This is teh difference between 2σ and 3σ

When are different NN geometries useful?

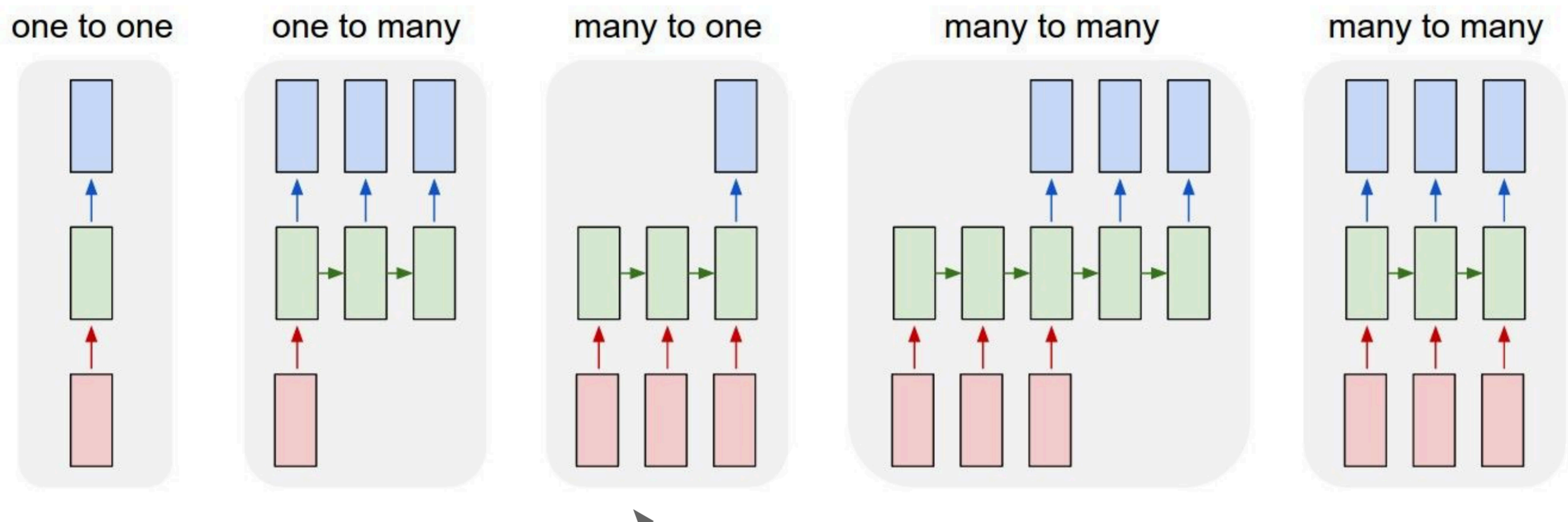
- Recall form Earlier





Using an RNN

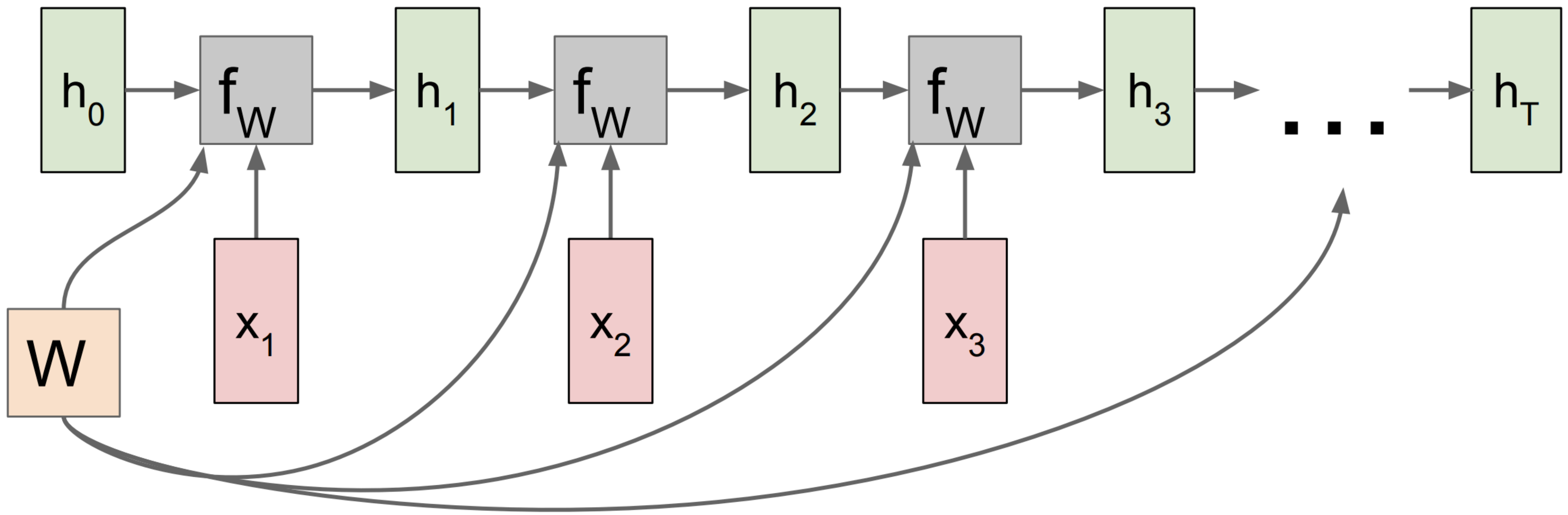
- Recusive neural network takes input one by tone



e.g. **Sentiment Classification**
sequence of words -> sentiment

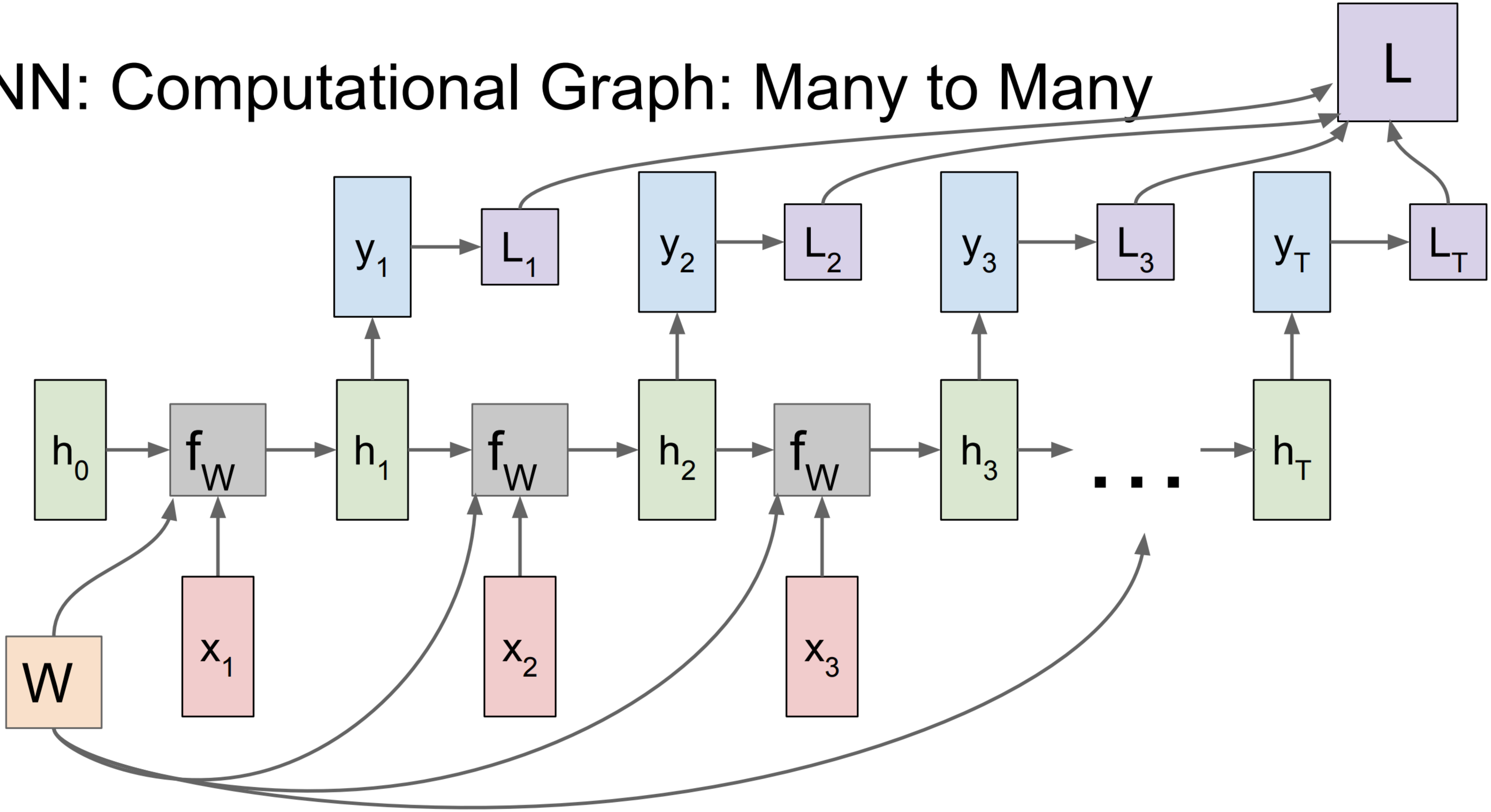
Using an RNN

Re-use the same weight matrix at every time-step

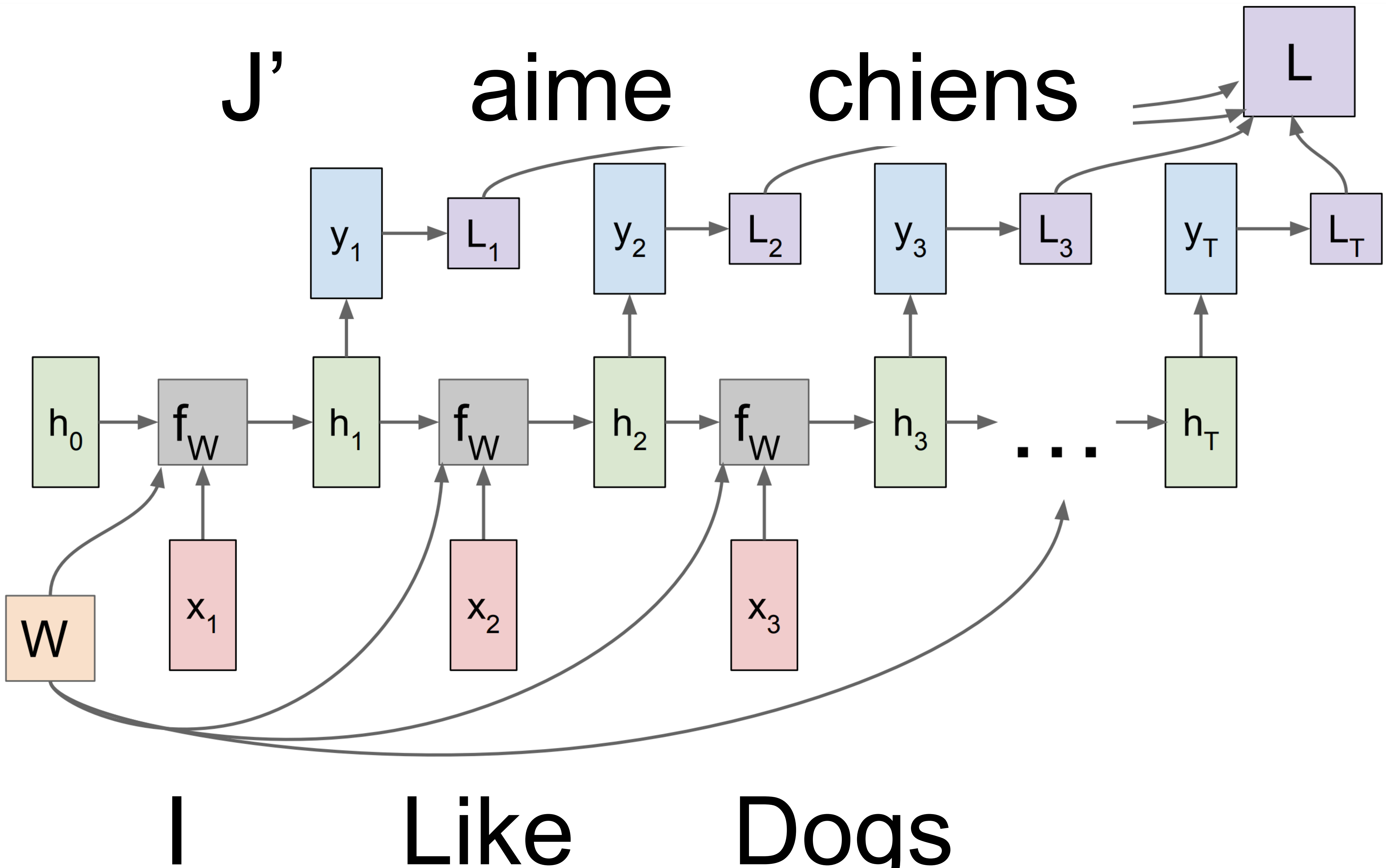


Using an RNN

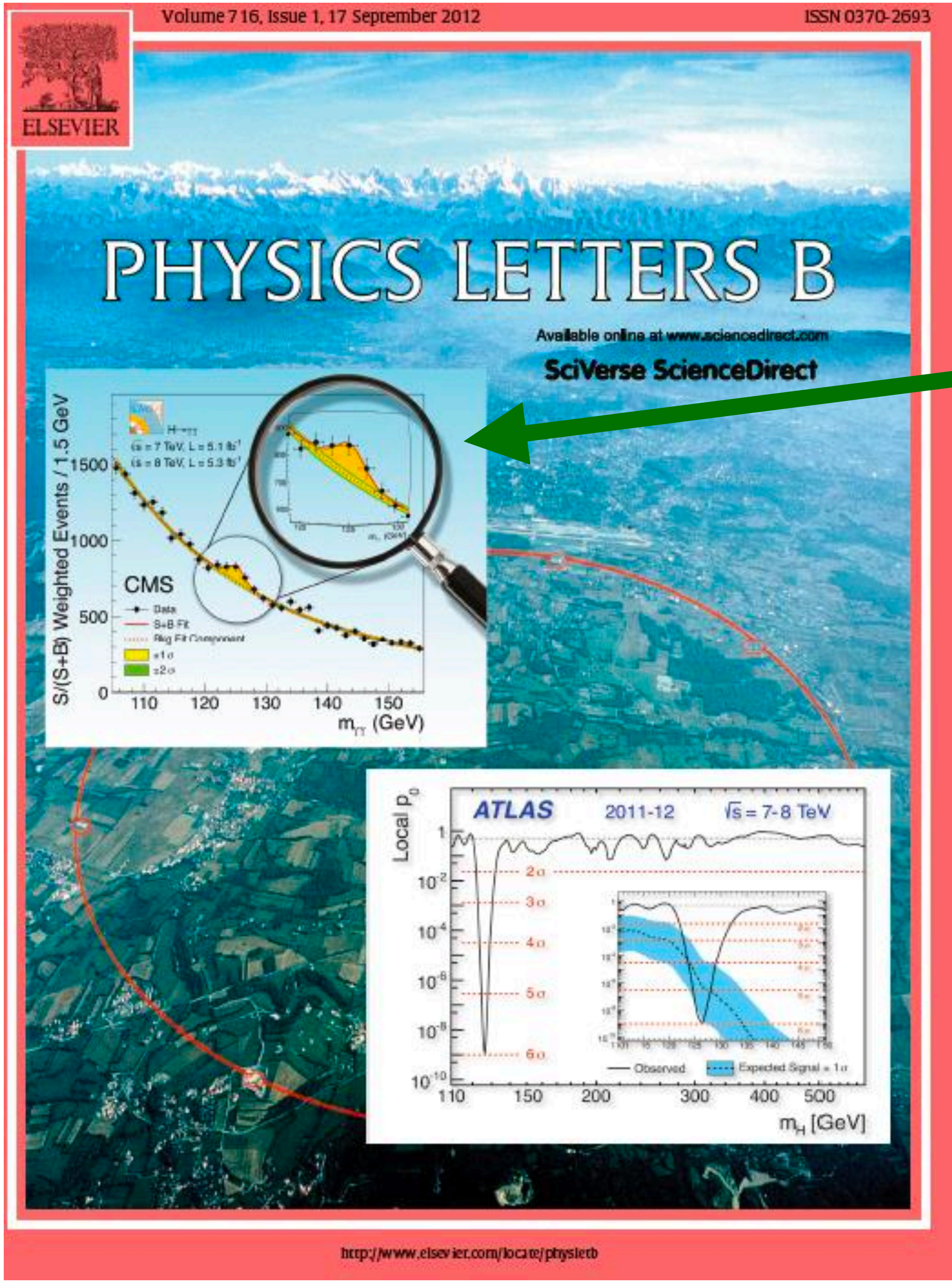
RNN: Computational Graph: Many to Many



Using an RNN


J'
aime
chiens
L
y_1
L_1
y_2
L_2
y_3
L_3
y_T
L_T
h_0
f_W
h_1
f_W
h_2
f_W
h_3
...
h_T
W
x_1
x_2
x_3
I
Like
Dogs

A Point



That Plot has a photon energy NN regression

Summary

- This class we showed the flexibility of the NN
- The real insight here is that we modified the loss
- We tried to solve a problem different than classification
- **You can solve many more**

Bonus

Are you Hungry?

- Lets do something fun:
  - Online there is a recipe list of about 100k recipes
- Challenge:
  - Lets try to generate our own recipes
- Any ideas of how you can do this?