Long Lived Machine Learning

G. Watts (UW/Seattle)

For the ATLAS Collaboration

ACAT 2017
21-25 August 2017
University of Washington, Seattle

Join us for talks, round tables, and discussions on

- **Machine Learning in Physics**
- **New algorithms for Particle Physics**
- **Advances in algorithms for theory computations**
- **Tools and infrastructure for computing in physics**

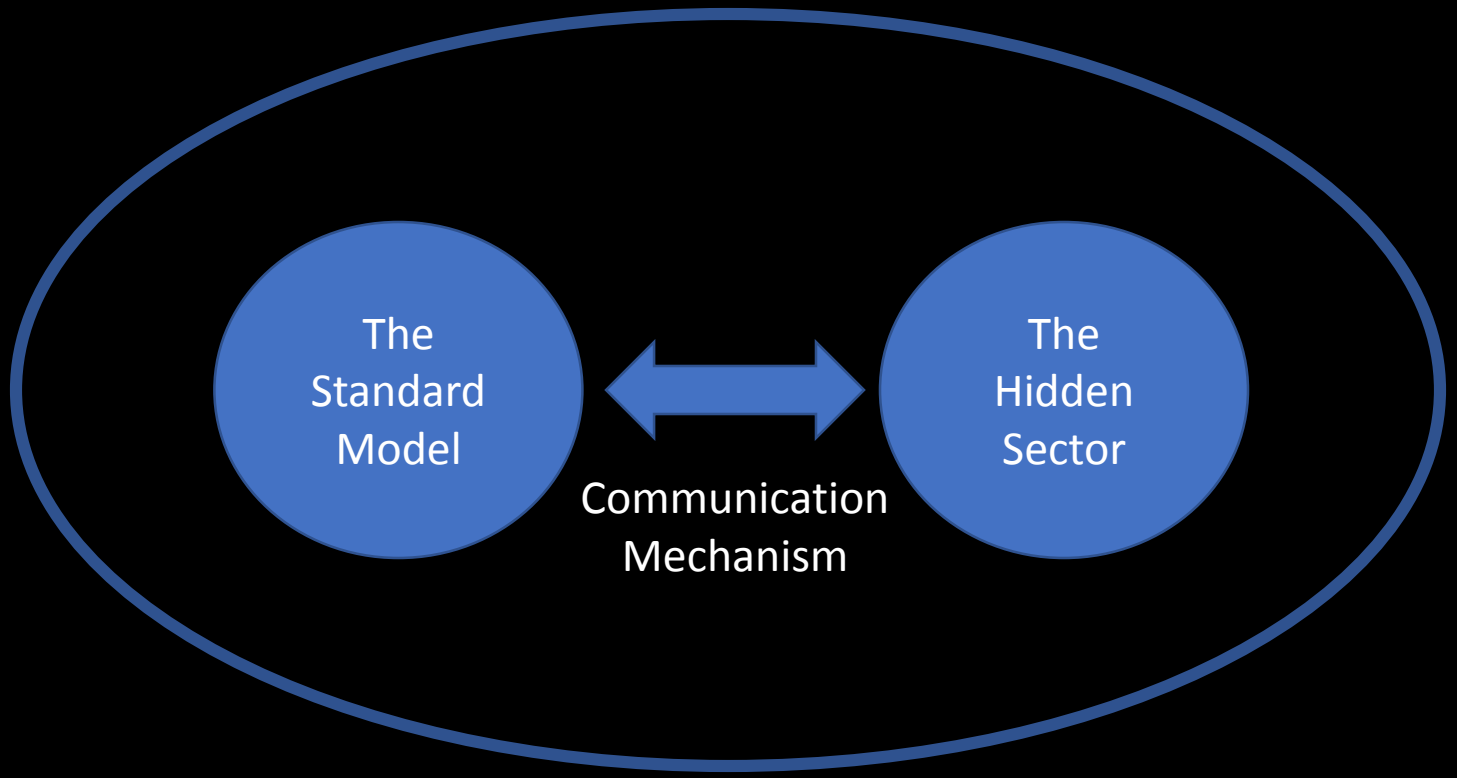
There is a fundamental shift occurring in how computing is used in research in general and data analysis in particular. The abundance of cheap, powerful, easy to use computing power in the form of CPUs, GPUs, FPGAs, etc., has changed the role of computing in physics research over the last decade. The rise of new techniques, like deep learning, means the changes promise to keep coming.

18th International Workshop on Advanced Computing and Analysis Techniques in Physics Research

acat2017@uw.edu
https://indico.cern.ch/event/acat2017

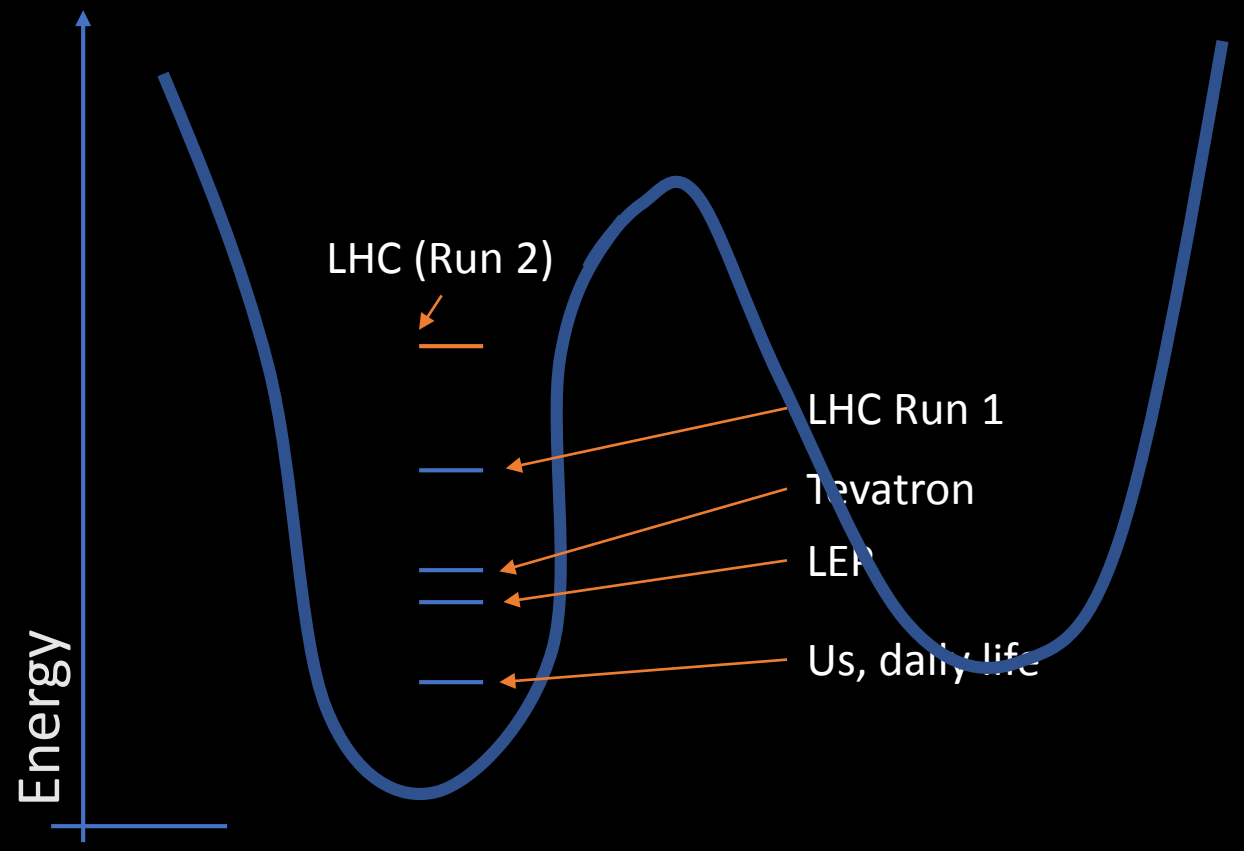
Longed Lived Particle Signatures

A Little Physics Context

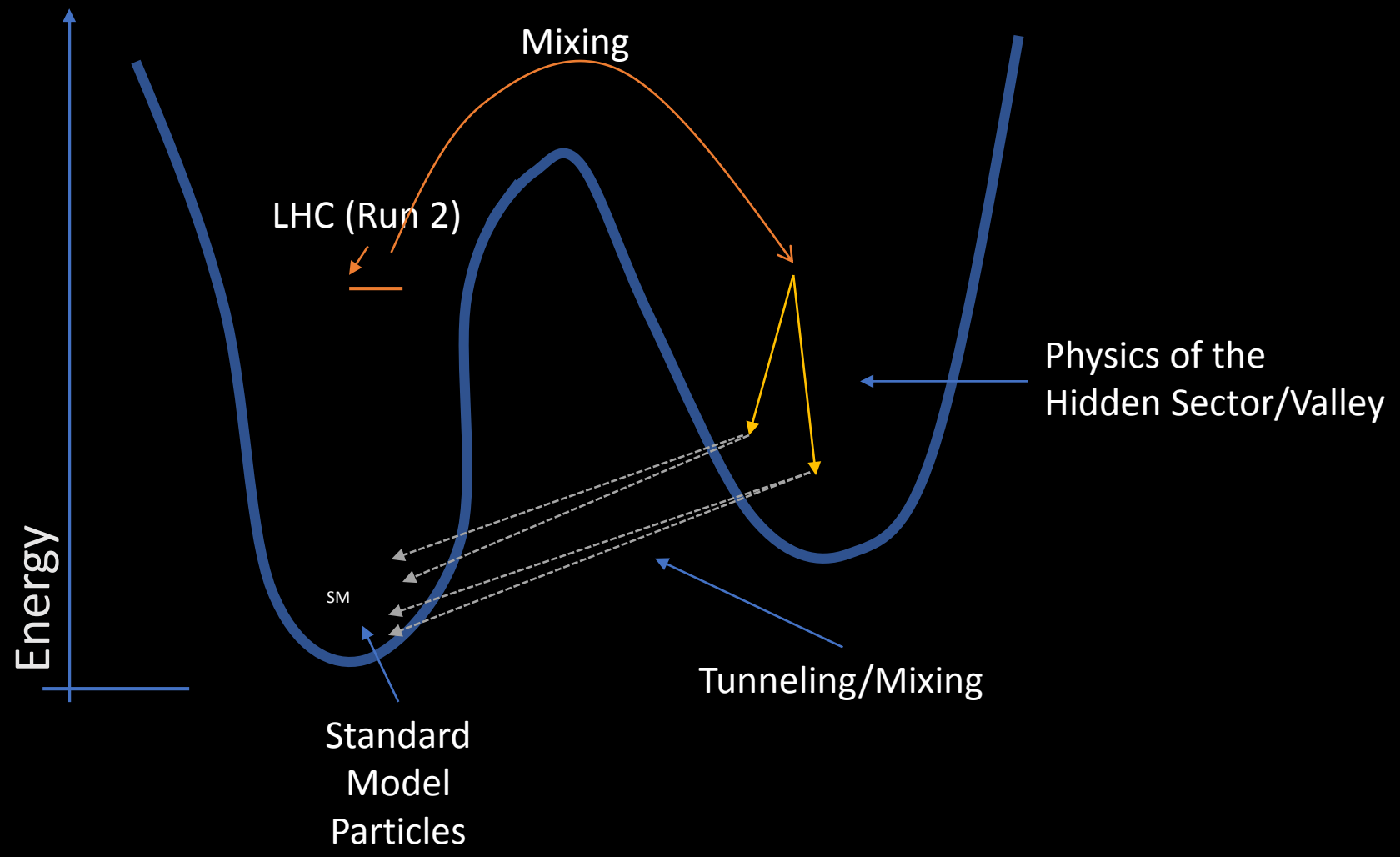


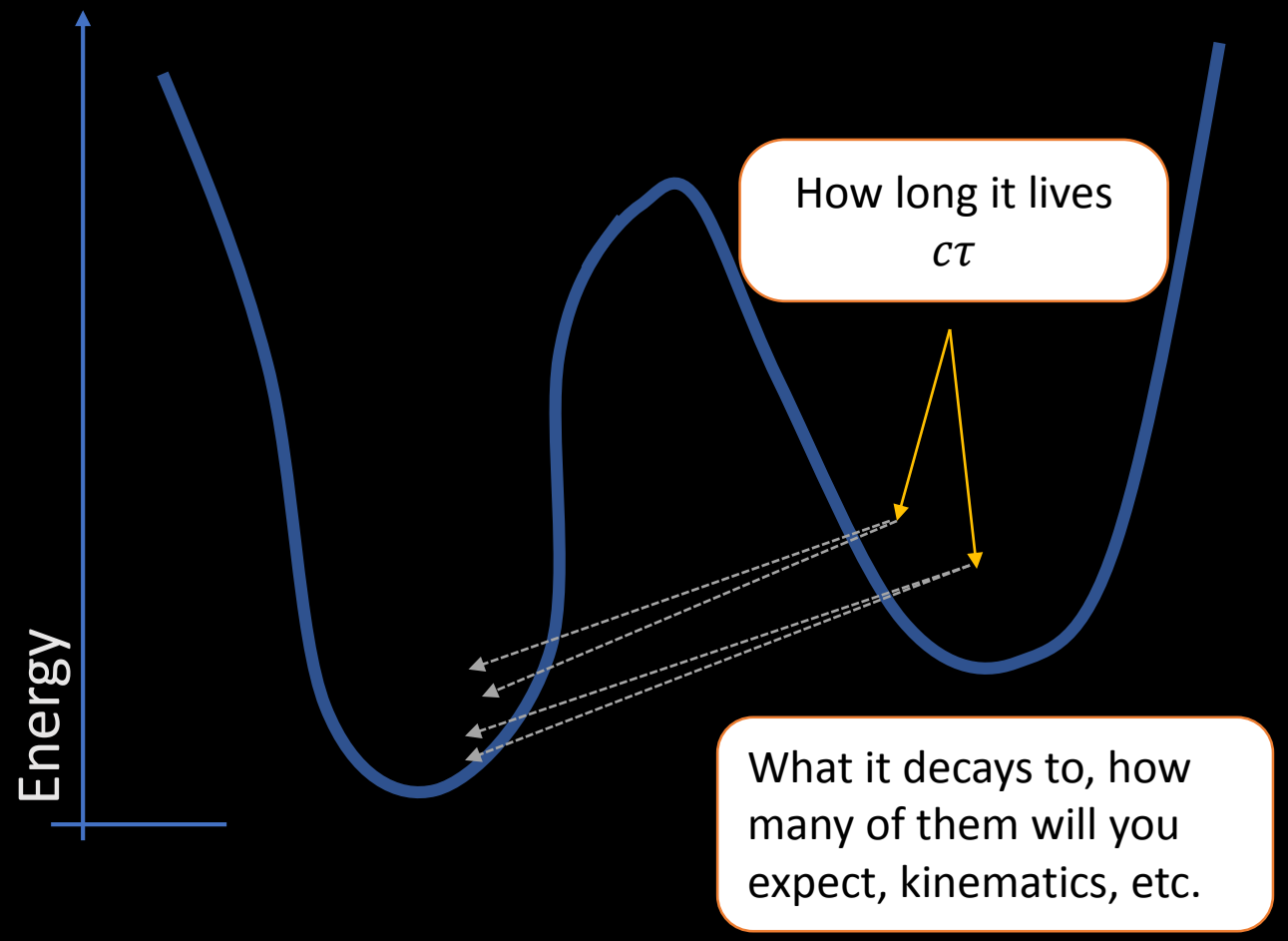
Hidden Sector does interact via Gravity

Why haven't we seen anything yet?



Why haven't we seen anything yet?




Energy
How long it lives
$c\tau$
What it decays to, how many of them will you expect, kinematics, etc.

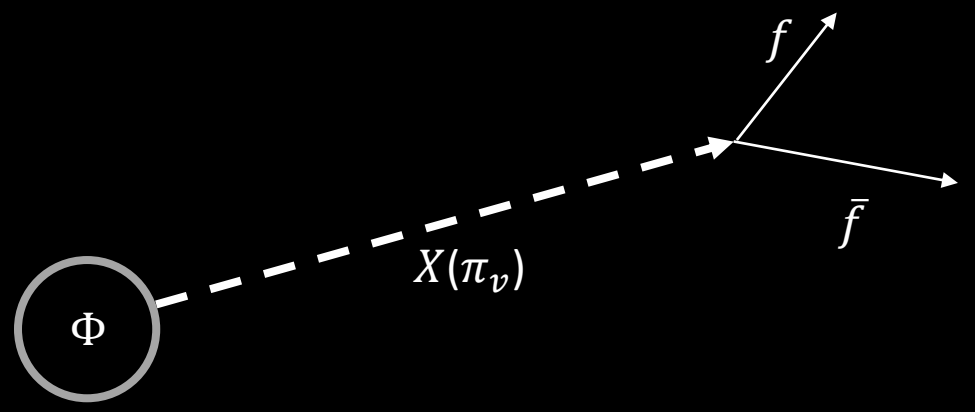
The mixing means the end result will be Standard Model Particles

- Couples to leptons (M_X small)

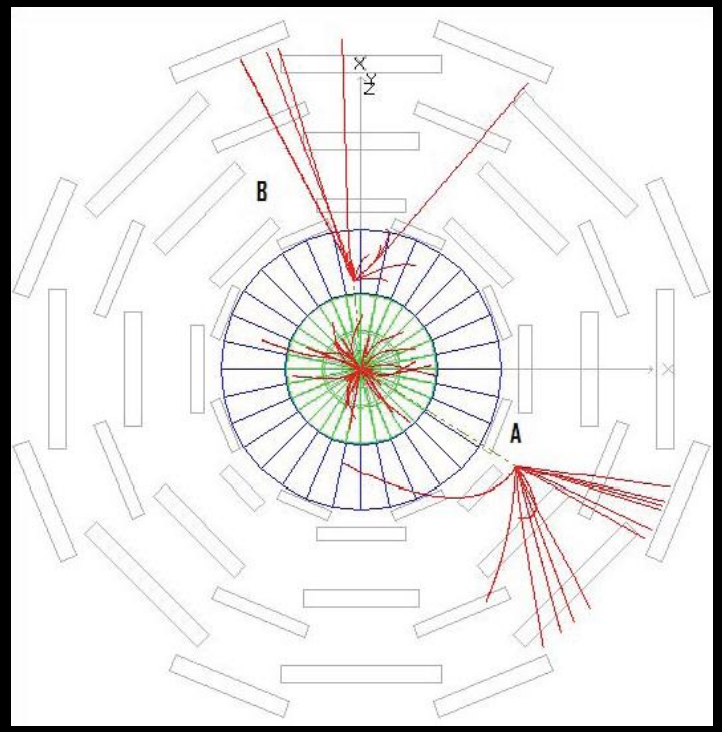
  "lepton-jets" – jets of leptons

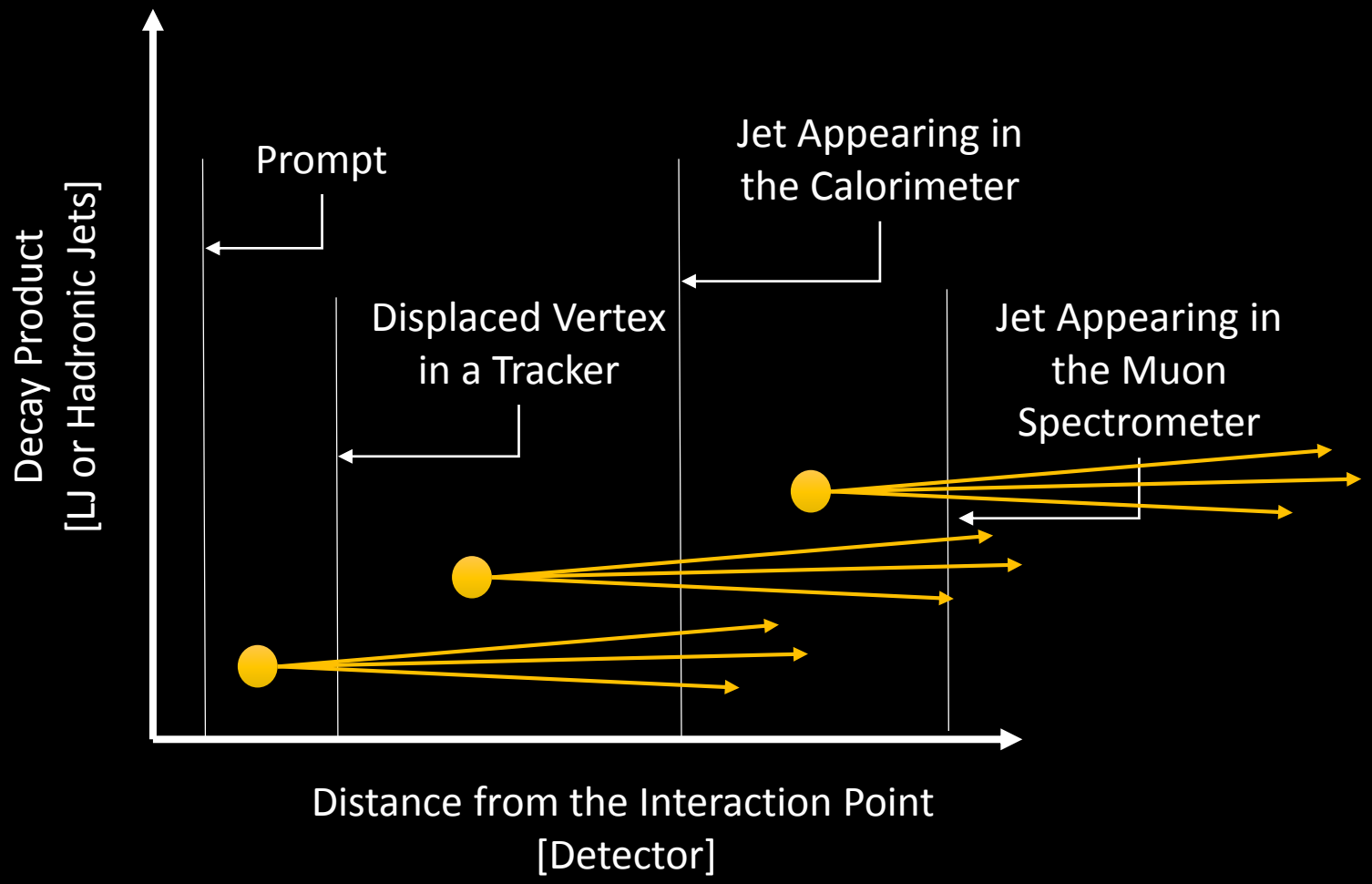
- Couples to heavy fermions (quarks)

  → Which means two (displaced) jets



Detection strategy depends on where in the detector the decay happens

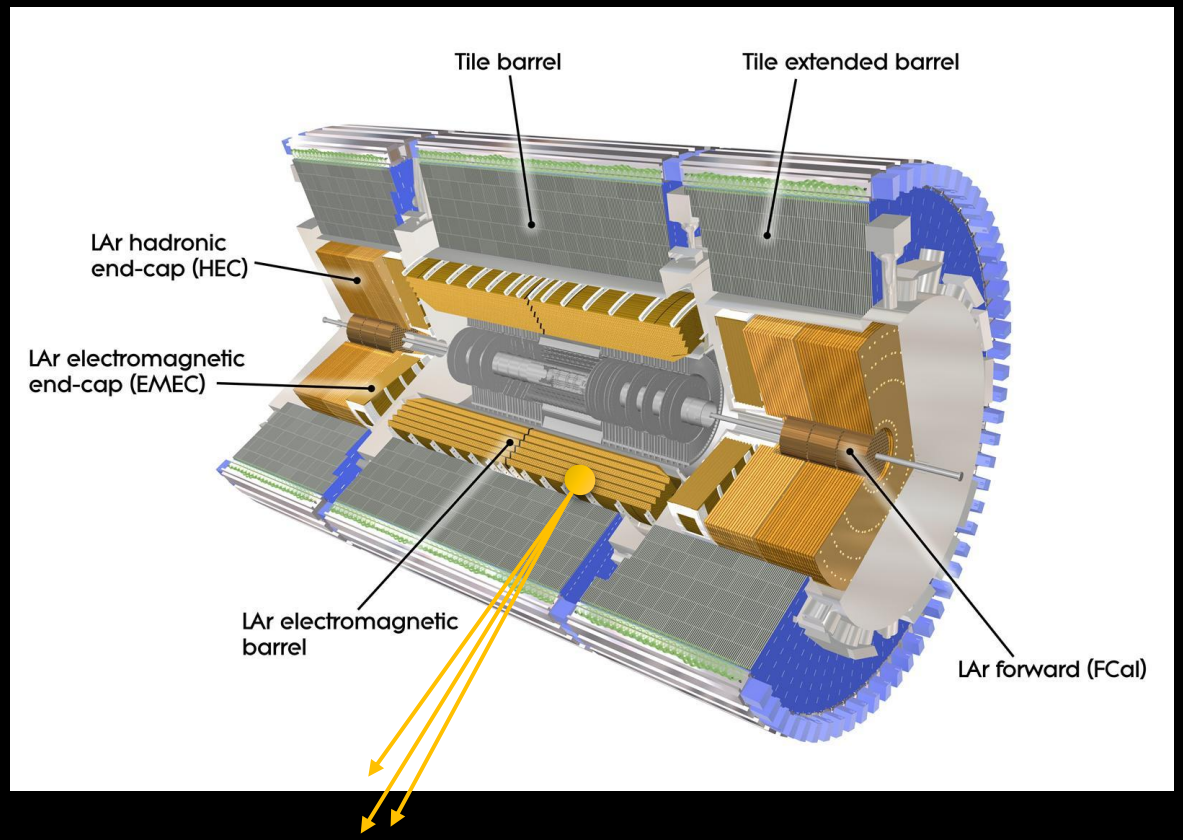




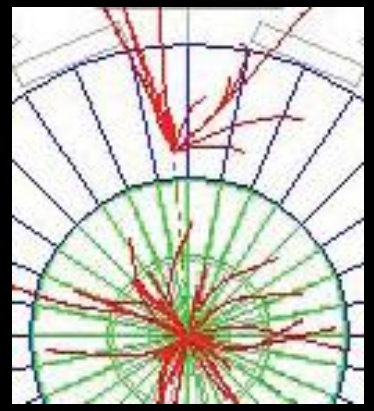
Lets concentrate on the calorimeter

Identifying Displaced Jets in the ATLAS Calorimeter

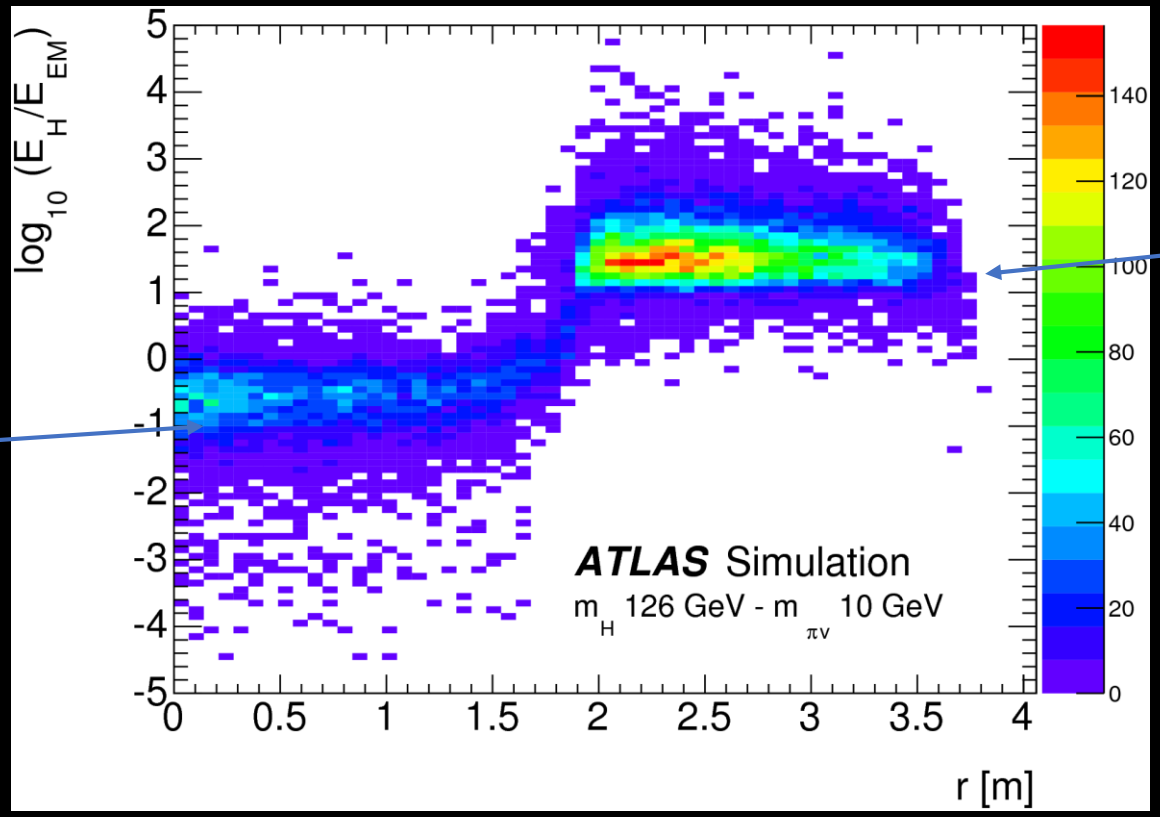
Decays in the Calorimeter



1. Look for the “appearance” of energy the Calorimeter
2. Little or no activity in the tracker

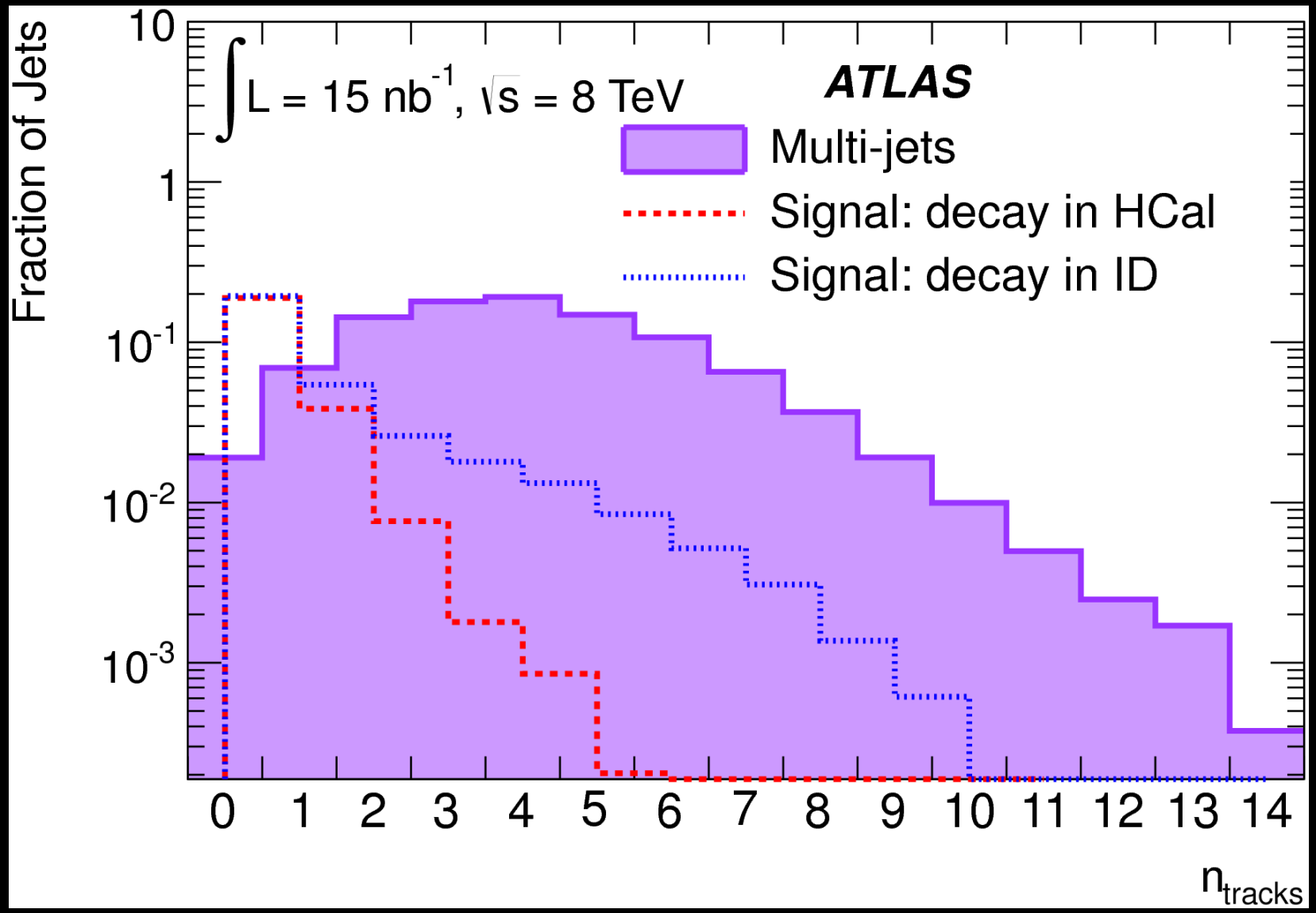


“CalRatio”

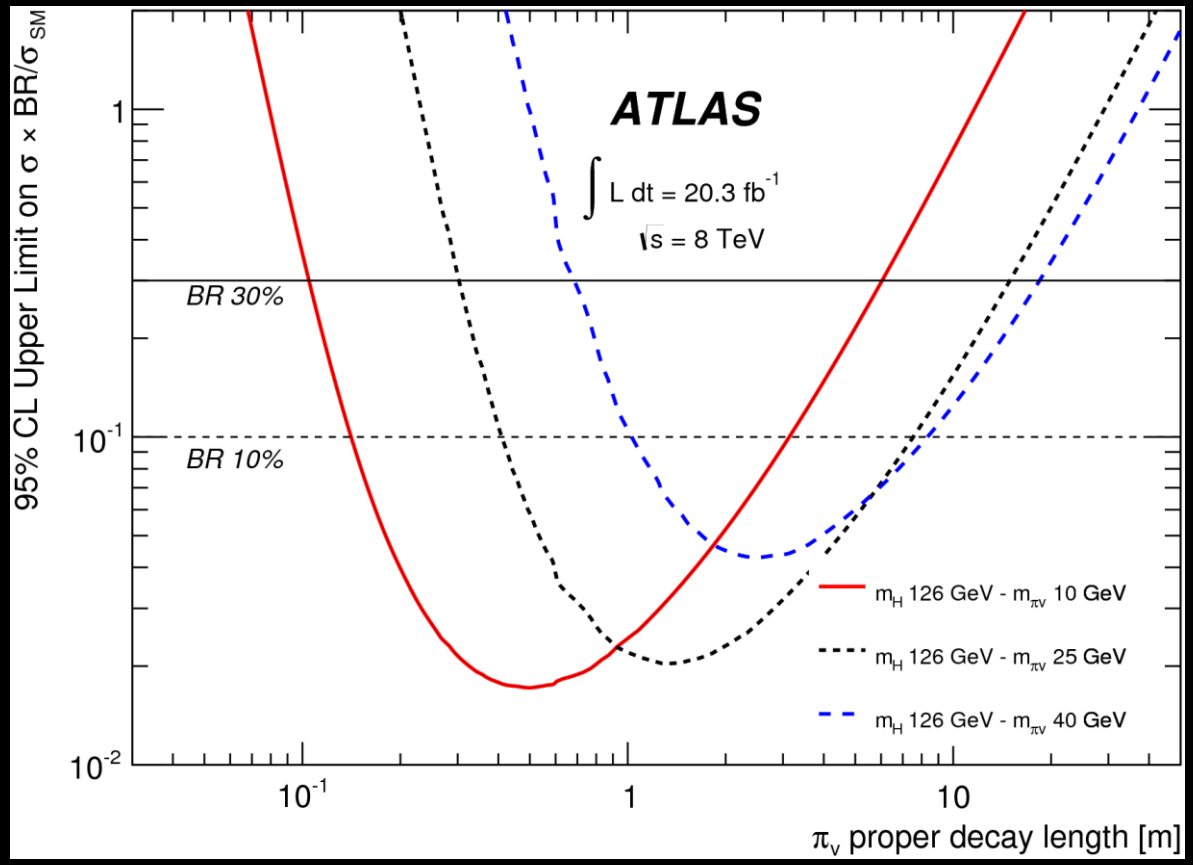


Signal Like

Background Like

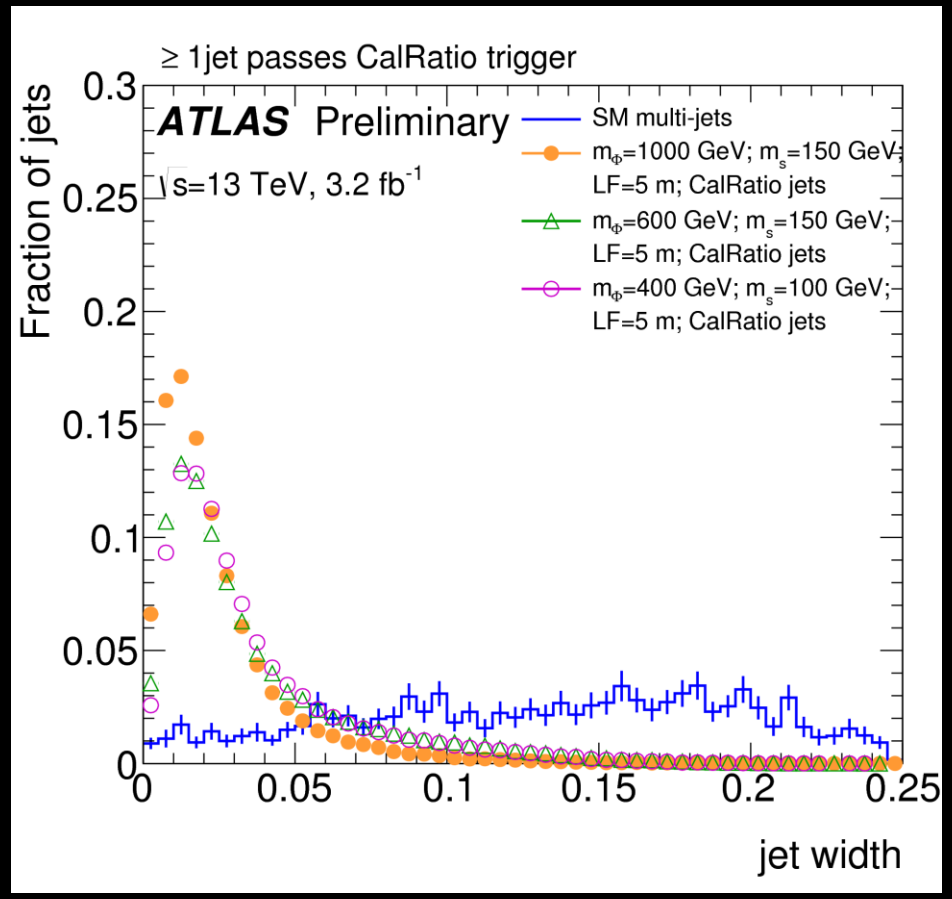

Fraction of Jets
10
1
10^{-1}
10^{-2}
10^{-3}
$\int L = 15\ nb^{-1}$, $\sqrt{s}$ = 8 TeV
ATLAS
Multi-jets
Signal: decay in HCal
Signal: decay in ID
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
n_{tracks}

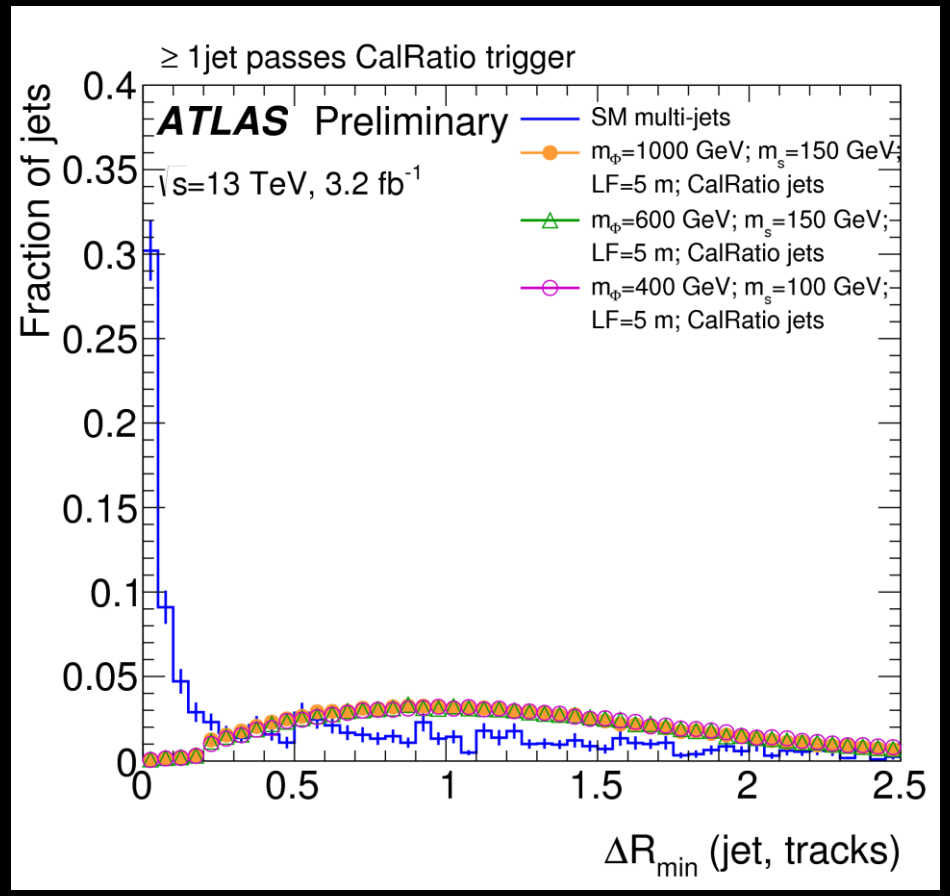
Can We Do Better?



There are lots more attributes connected with displaced jets

13 Jet Shape Variables





Boosted Decision Tree

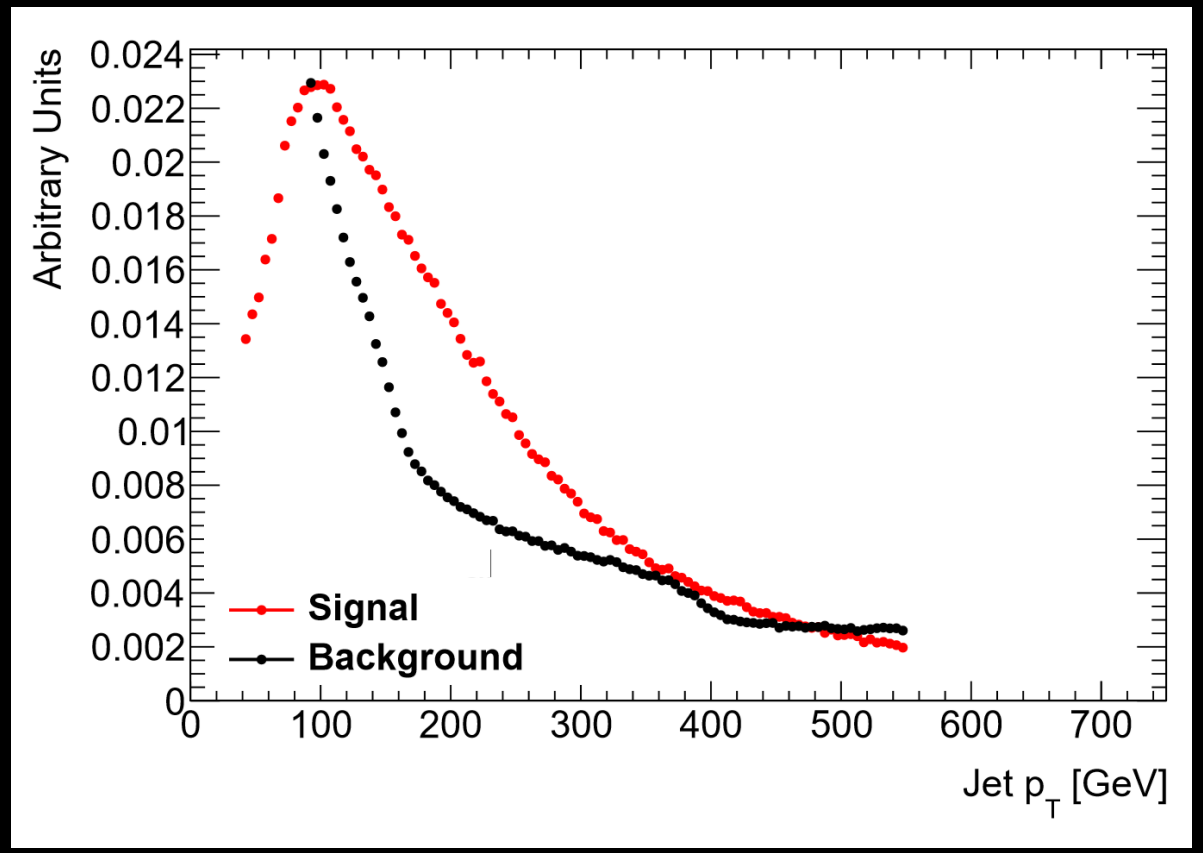
We used TMVA, and an old version at that

Training:

Jets with $p_T > 40$ GeV, $|\eta| < 2.4$
Signal if $\Delta R < 0.2$ to a π_v

We are just finding out
we are missing out here!

Reweight to flatten p_T spectra to avoid
model dependence



Training

A few parameter modifications from the defaults:

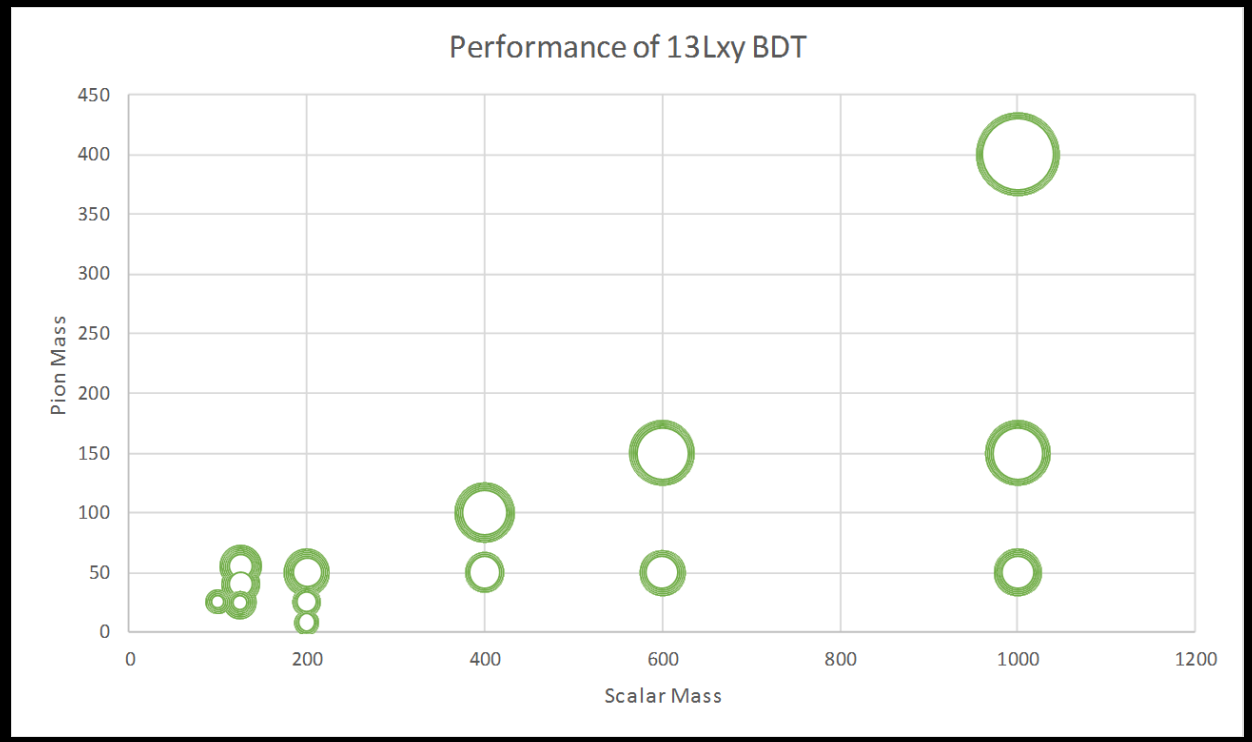
- Minimum Leaf Size: 1%
- Maximum Training Depth: 20
- Signal Training: 944,638
- Background Training: 1,499,996

Systematic Errors:

- Many input variables depend on Jet Energy Scale
- Required re-running on full sample many many times

Measuring performance:

- What BDT cut gives you a straight-cut background rate?
- Calculate the signal efficiency



Ordered Variables

$\log_{10}(E_{\mathrm{H}}/E_{\mathrm{EM}})$	The base-10 logorithm of the ratio of the energy in the HCal to ECal. If no energy is present in the ECal, a large number is used as input to the BDT instead.
Jet Width	The p_{T}-weighted sum of the ΔR between each cluster the jet is built from and the jet axis.
Leading cluster Longitudinal length	How far the highest p_{T} cluster has spread in the longitudinal direction
Jet p_{T}	Training events are reweighed so that the jet p_{T} distribution is flat. This allows the BDT to look for correlations between other variables and the jet p_{T} without using jet p_{T} directly as a discriminating variable.
Leading cluster lateral width	How far the highest p_{T} cluster has spread in the latitudinal direction.
Leading Jet Cluster Shower Center	The distance from the inner edge of the ECal to the highest p_{T} cluster's center along the jet axis. A zero would mean this was on the edge of the ECal.
ΔR to Closest 2 GeVTrack	The ΔR between the jet axis and the closest 2 GeV track. Or 0.4 if there is no such track.
Radius of the leading cluster	The radius from the beam spot to the highest p_{T} cluster.

Cluster Energy Density	The $\frac{\sum E_i^2/V_i}{\sum E_i}$, where E_i is the energy of each cell, and V_i is the volume of the cell in the highest p_{T} cluster in the jet. Only positive energy cells are considered.
Number of tracks	The number of tracks with $p_{\mathrm{T}} > 2\ GeV$ within a $\Delta R < 0.2$ of the jet axis.
Sum p_{T} of all tracks	The sum p_{T} of all tracks within $\Delta R < 0.2$.
Layer 1 Energy Fraction	The fraction of the hadronic calorimeter's energy located in layer 1. The later in the calorimeter a LLP decays, the smaller value this variable will be.
Max track p_{T}	The maximum p_{T} of all tracks within $\Delta R < 0.2$.

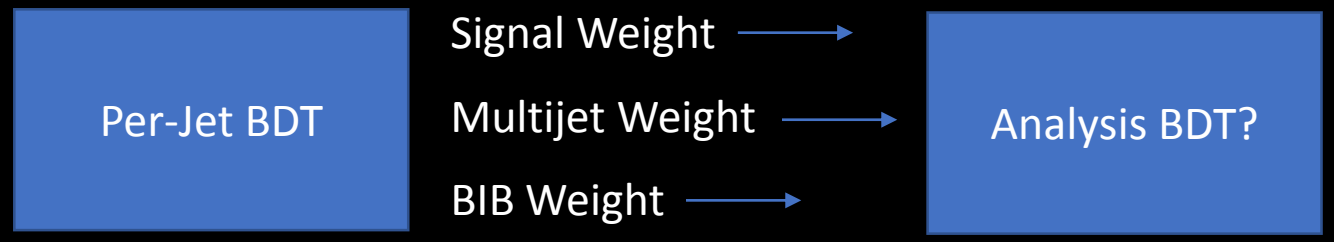
This result released as a conference note

- Same sensitivity with 3.2 fb^{-1} of data as the previous one with 20 fb^{-1}
  - Note the $\sqrt{s}$ change!
- Low mass sensitivity not as good in Run 2!

What is next?

We had to limit the sensitivity of the last BDT artificially.

- Beam induced background is ugly and looks exactly like our signal.
- Currently training a multi-class BDT to look for Multijet, signal, and Beam Induced Background



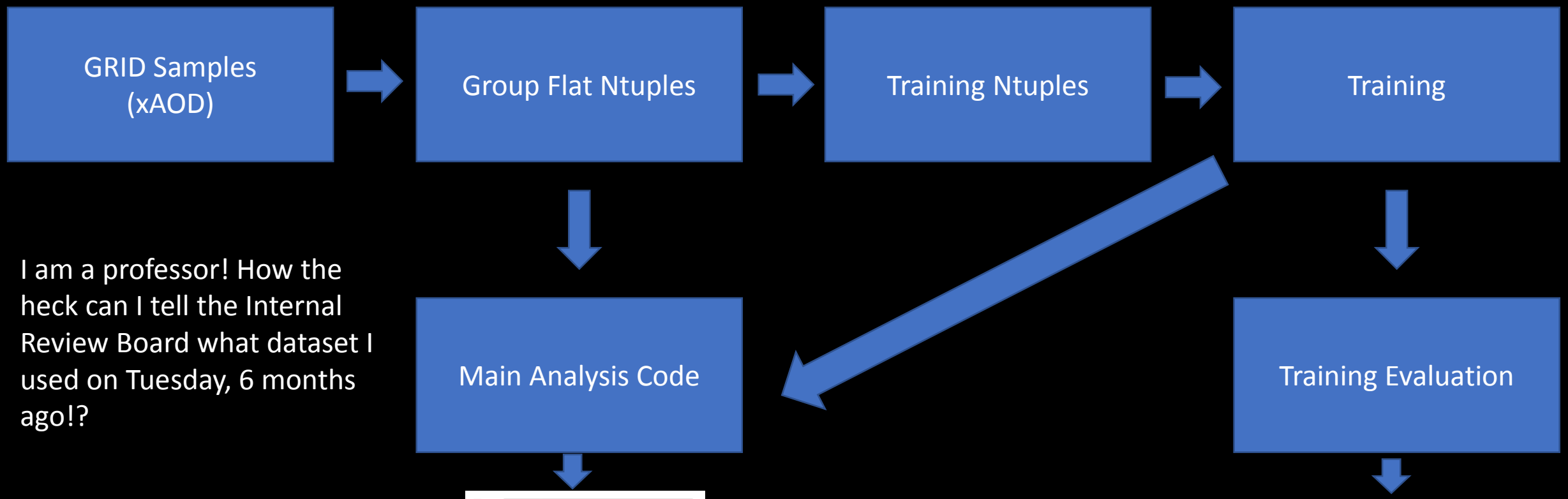
Training separately… because we were unaware of the tools available

Further in the future:

- The calorimeter is just a 3D CCD
- These jets have particular shapes
- Use Deep Learning to do image recognition.
- Some work along these lines in other contexts has already been done

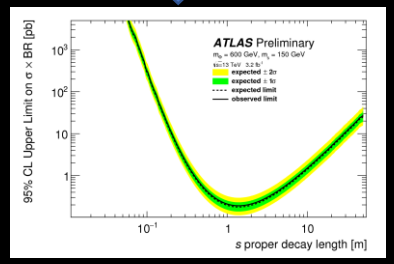
Peaches and Cream

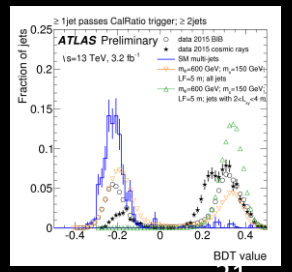
Pain…



I am a professor! How the heck can I tell the Internal Review Board what dataset I used on Tuesday, 6 months ago!?

Partly ATLAS' Fault
Partly the tools
Partly the analysis team

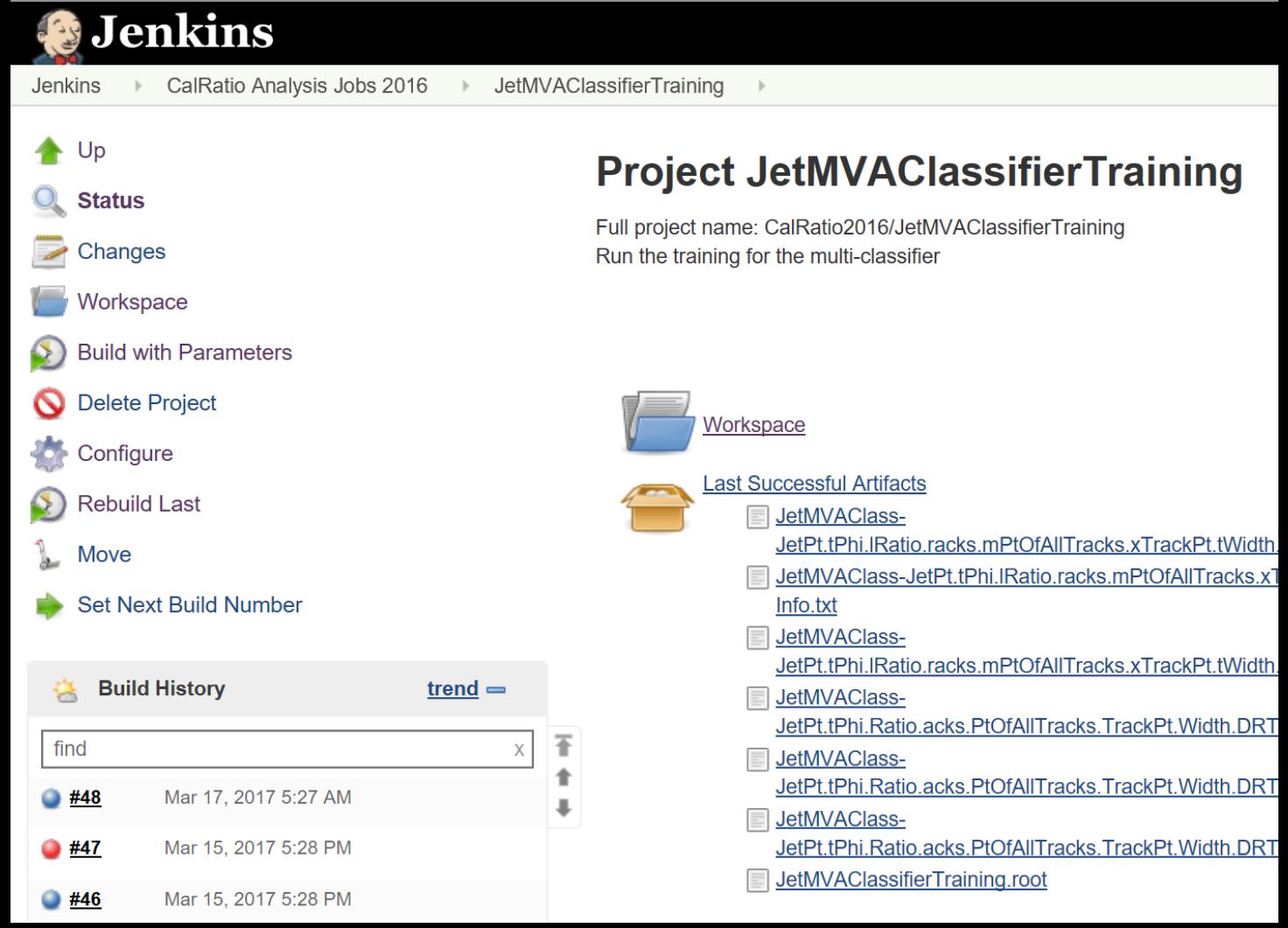




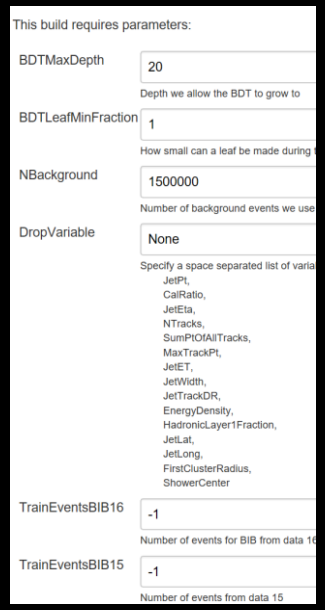
Sample	TrainingEvents	AddVariable	BackgroundSample	FlattenBy	Performance	BDTLeafMinFraction	BDTTrainingTreeDepth	JobId
1000pi50lt5m	2000000		UseBackgroundAll	JetPt	0.318192988	0.5	30	632
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.317231812	0.5	40	636
1000pi50lt5m	1000000		UseBackgroundAll	JetPt	0.316992295	1	30	614
1000pi50lt5m	2000000		UseBackgroundAll	JetPt	0.316908309	1	50	610
1000pi50lt5m	2000000		UseBackgroundAll	JetPt	0.316908309	1	40	639
1000pi50lt5m	1000000		UseBackgroundAll	JetPt	0.316283077			592
1000pi50lt5m	1000000		UseBackgroundAll	JetPt	0.316283077			617
1000pi50lt5m	2000000		UseBackgroundAll	JetPt	0.316071556			608
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.31521303			613
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.31521303			634
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.31521303			635
1000pi50lt5m	1000000		UseBackgroundAll	JetPt	0.314706001			618
1000pi50lt5m	2000000		UseBackgroundAll	JetPt	0.314656232	1	30	604
1000pi50lt5m	2000000		UseBackgroundAll	JetPt	0.314093212	0.5	50	611
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.313707497	0.5	50	641
1000pi50lt5m	1000000		UseBackgroundAll	JetPt	0.312852081	0.5	50	620
1000pi50lt5m	2000000		UseBackgroundAll	JetPt	0.312811644	0.25	30	606
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.312581459	0.5	30	623
1000pi50lt5m	2000000		UseBackgroundAll	JetPt	0.312506804	0.25	40	609
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.311819361	1	30	622
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.311819361	1	40	625
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.311819361	1	50	628
1000pi50lt5m	1000000		UseBackgroundAll	JetPt	0.311810029	0.5	30	615
1000pi50lt5m	2000000		UseBackgroundAll	JetPt	0.310478691	0.25	50	612
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.308836292	0.25	30	624
1000pi50lt5m	1000000		UseBackgroundAll	JetPt	0.308833181	0.25	50	621
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.308400808	0.25	40	637
1000pi50lt5m	1000000		UseBackgroundAll	JetPt	0.304630756	0.25	40	619
1000pi50lt5m	1500000		UseBackgroundAll	JetPt	0.30166635	0.25	50	603
1000pi50lt5m	1000000		UseBackgroundAll	JetPt	0.301237087	0.25	30	616

Testing the training required a huge amount of repeated runs.
Also a disaster to track.

Automated Everything



- Tracks from GRID to TMVA training XML
- Scripts to download and scan log files automatically to generate performance tables



Command line configuration parameters

TMVA Training Simpler?

```
var training = flatSignalTrainingData
                    .AsSignal(isTrainingEvent: e => !(e.EventNumber % 3 == 1))
                    .Background(flatBackgroundTrainingData, isTrainingEvent: e => !(e.EventNumber % 3 == 1))
                    .UseVariables(varList);

                // Build options (like what we might like to transform.
                var m1 = training.AddMethod(ROOTNET.Interface.NTMVA.NTypes.EMVA.kBDT, "BDT")
                    .Option("MaxDepth", options.BDTMaxDepth.ToString())
                    .Option("MinNodeSize", options.BDTLeafMinFraction.ToString())
                    .Option("nCuts", "200")
                    ;

                // Do the training
                var trainingResult = training.Train("JetMVATraining");
```

Does the required flexibility really require the existing TMVA complexity?

Obviously I should be exploring some of the tools mentioned here

Conclusions

- How to give theorists object efficiency plots for reinterpretation of the results
- The Community White Paper processes
  - The groups are not islands
  - Analysis complexity is rising too fast
  - For us: ML is too small a part of the overall experiment
- Automate all tools
- Make sure their use can easily be moved to the “cloud”
  - Config files, network access required, keep to a minimum
- I’ve been involved in ML for a long time
  - Since DZERO single top discovery days
- This is the first time I did it
  - Rather than a student or post-doc of mine. ☺
- I’m bullish in that given my limited analysis time I was able to contribute to an analysis
  - Obvious measure of TMVA and ML’s success and accessibility
- But, I am a “pet” analyzer