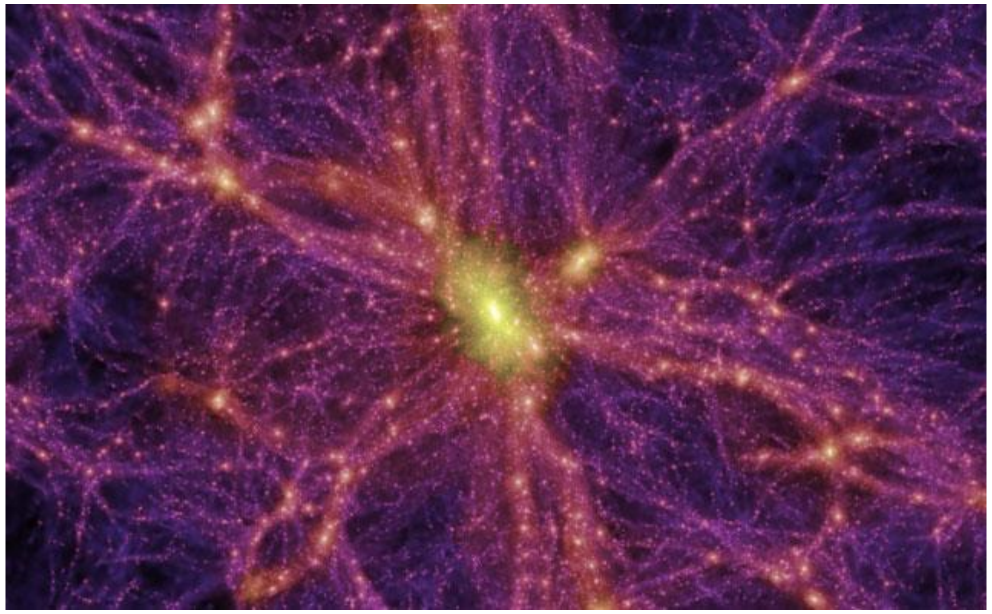
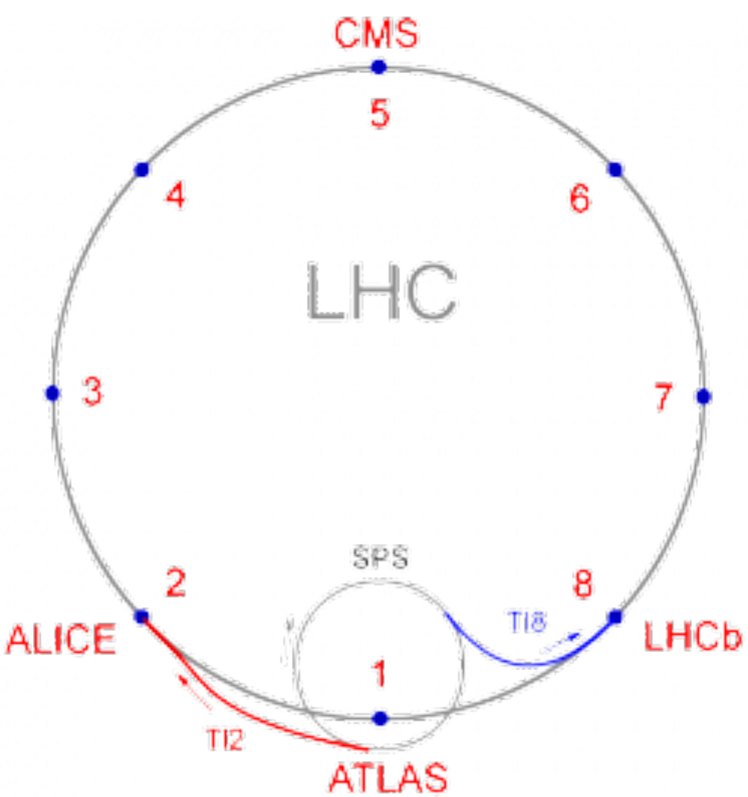


Thoughts on MVA for Dark Matter Searches at the LHC



Jamie Gainer
UH Manoa
April 5, 2017

Multivariate Analyses

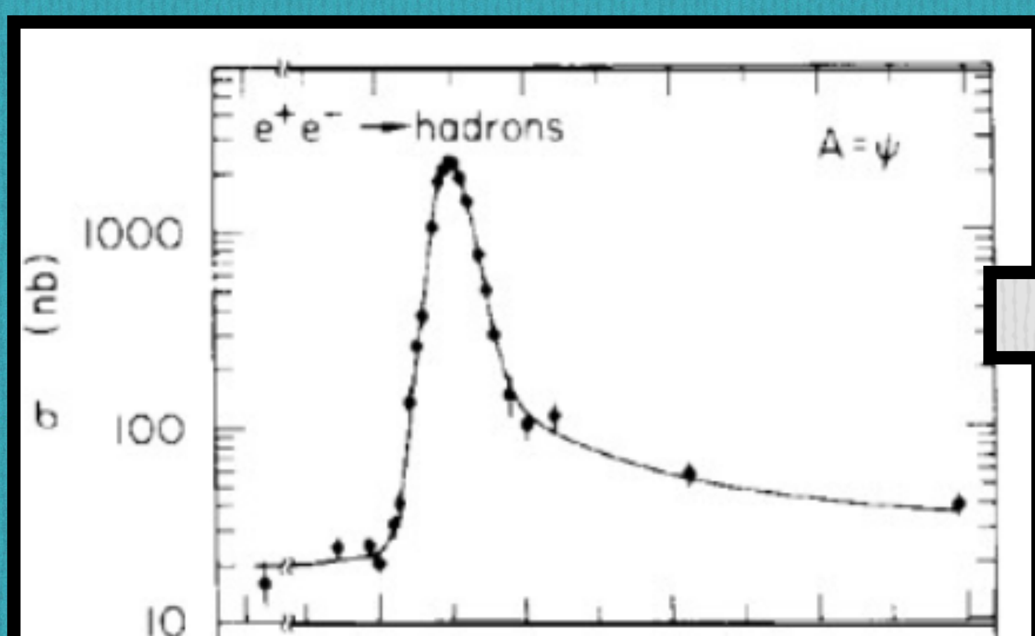
- An analysis using more than one variable

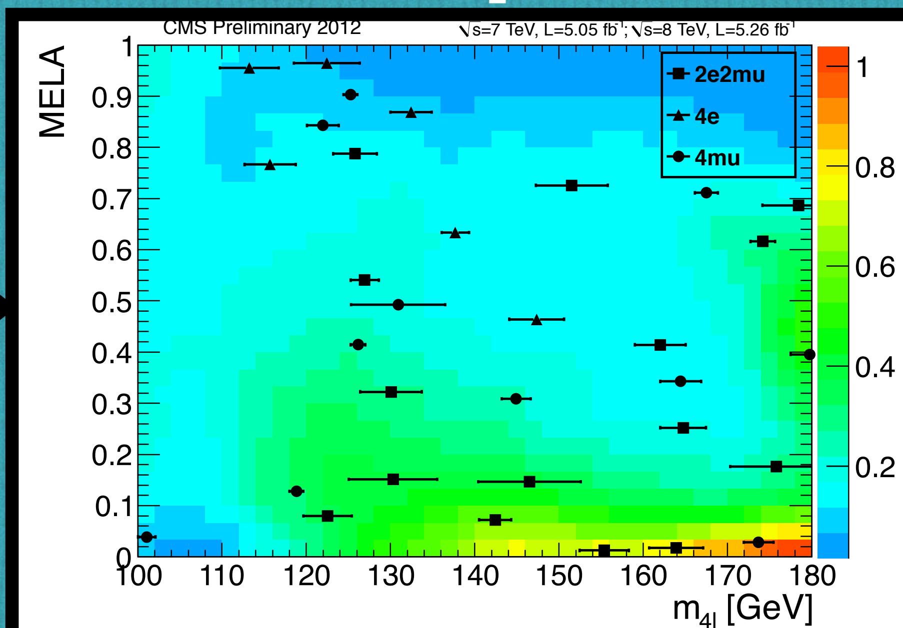
Captain Obvious

- Often refers to sophisticated/ powerful methods (neural nets, boosted decision trees, the matrix element method) which are contrasted with, e.g., binning data in a single variable

Revolution in Experiment!!!

Multivariate Methods are Now Ubiquitous

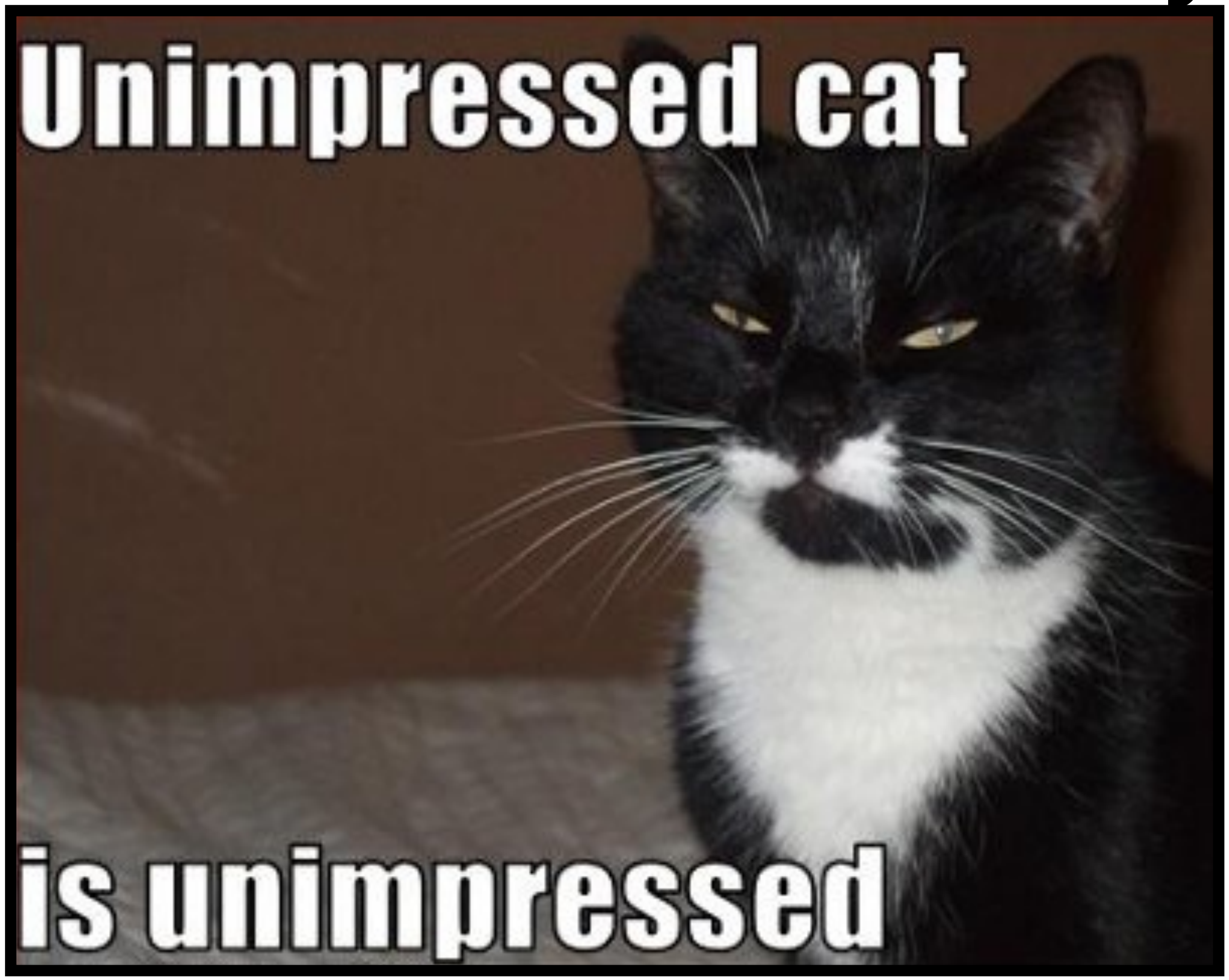




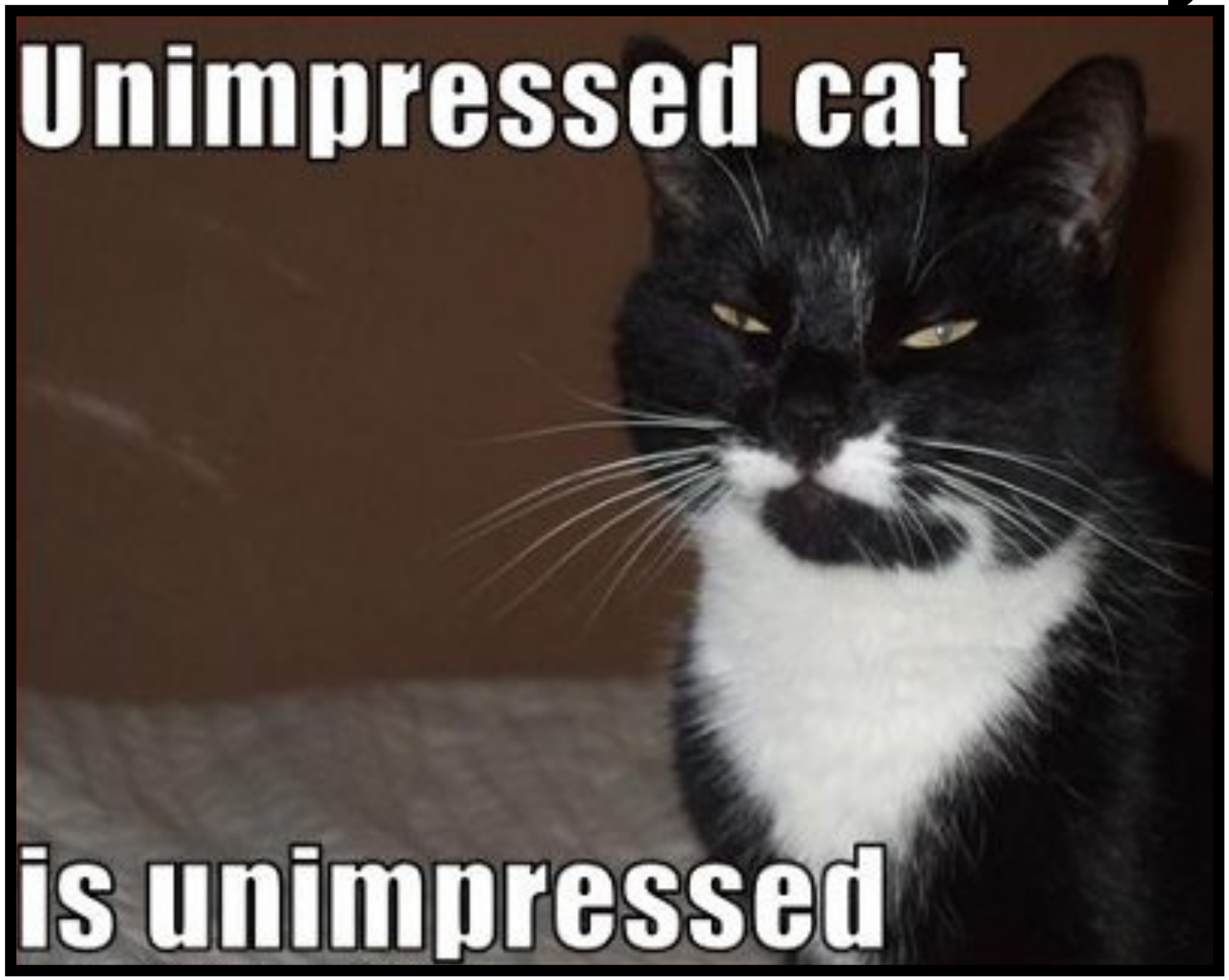
MVA information helpful even in discovering a narrow resonance in leptons!

Less So in Theory…

Less So in Theory…



Less So in Theory…



- I'm overstating the case. Plenty of work by theorists, especially on the Matrix Element Method.

- But not proportionate to the use of MVAs (especially neural nets and BDTs) in experiment

Understanding the Physics that Drives the Sensitivity

- The real problem is that it is hard to understand the physics reasons for increases in sensitivity that come from neural nets (black boxes?)
- Obviously the experimentalists don't want us to run neural nets for them, especially since some of the power comes from sensitivity to detailed aspects of the detector response
- But can we use theory tools to understand the additional sensitivity from MVAs?
- Yes, in the Matrix Element Method

Pauli

The Matrix Element Method

- Using the likelihood ~ probability
- Essentially the differential cross section normalized to the total cross section
- Once we take into account that what we measure in detectors isn’t the four momentum of a parton (transfer functions)
- And we integrate over particles we don’t see at all (like dark matter!)
- According to our friends Neyman and Pearson, the likelihood is an “optimal test statistic”

The Matrix Element Method

- And it works great in Higgs physics, especially Higgs to four-leptons

- But also in top physics, B physics, some SUSY…

- The big thing is that since your MVA output is essentially a differential cross section, the underlying physics can be understood

    - Heavy Higgs to four-leptons: sensitivity from different helicity amplitudes in signal and background, i.e., sensitivity driven by angular momentum considerations, spin of Higgs, etc.

    - SM Higgs to four-leptons, sensitivity driven by different propagators in signal (ZZ^*) and background ($Z\gamma^*$) leading to different distribution of M_{Z2}

The Matrix Element Method

- But it’s really hard to calculate likelihoods, especially for
  - many particle final states,
  - many jets— transfer functions matter!
  - many missing particles— hard integrals
  - reducible backgrounds-- need to understand the detector

MVA for Dark Matter Searches

- I'd love to be wrong, but the Matrix Element Method does not seem great for dark matter searches at the LHC
- Always have invisible particles— integrals. (Even worse if we also have neutrinos or more than two DM particles in the fianl state)
- How to deal with jets: for four-leptons can focus on leptons.
  QCD effects enter analysis only through pT and eta of the four-lepton system
- What's the analogue of this for, e.g., monojets?
- Vast advances , but still significant dependence on detector modeling (which goes into the MC expressions that are reweighted).
  For example, when do you miss the lepton in W + jets?
- Many dark matter searches— but I'm going to focus on maybe the most dramatic example: **monojets**

Monojets Then

SCIENCE

Supersymmetry theory decaying

Christine Sutton

DOES supersymmetry exist? Theorists in particle physics began to ask that question a few months ago. Initially, the answer seemed to be "maybe". Now it is changing to "maybe not", particularly in the light of recent calculations that show how certain phenomena can mimic some of the predicted effects of supersymmetry.

Supersymmetry is symmetry so neat that many physicists want it to be true—the more so because supersymmetry might offer a pathway to a quantum theory of gravity and hence to a completely unified theory of all the fundamental forces (*New Scientist*, 15 March 1984, p 28). The main feature of supersymmetry is that it links the particles of matter—"fermions"—with the particles that carry the fundamental forces—"gauge bosons". The two categories of particle simply become different facets of the same underlying object.

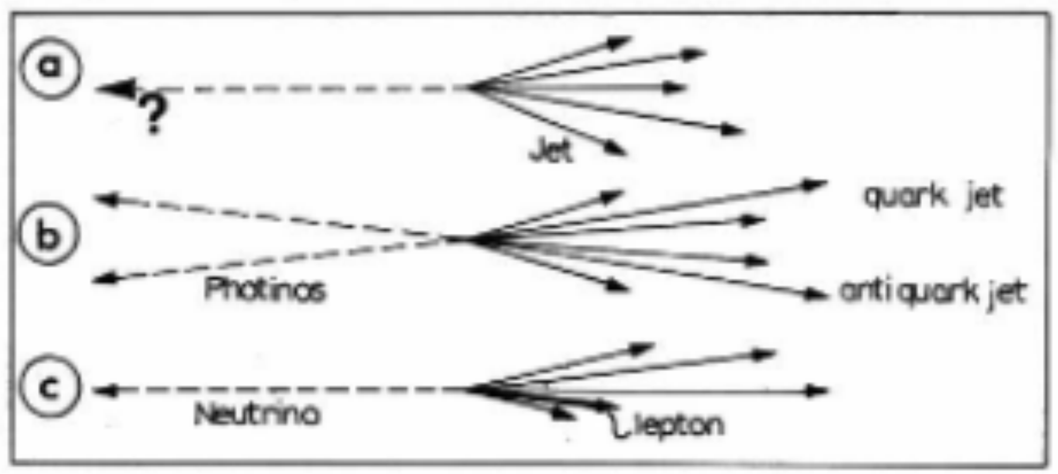
The possible evidence that

theoretical speculation about what they represent; it seemed at first that even as few as five examples were too many to account for by standard theories of particle physics. The proponents of supersymmetry were among those who snatched the monojets to

each incorporating some hitherto unobserved phenomenon, such as a new kind of neutrino; or quarks built from still smaller entities; or a new type of strong nuclear force, apparent only at the high energies explored in the proton-antiproton collisions at CERN.

Now, some detailed analysis by three physicists at CERN has put an experimental cat among the theoretical pigeons. James Stirling, Ronald Kleiss and Steve Ellis (normally at the University of Washington in Seattle) have looked very carefully into the question of whether conventional physics can explain the monojets. They calculate a higher number of monojets from normal sources than previous studies have suggested. Indeed, they claim that monojets "can arise at a significant rate" from the decays of W and Z particles which are not identified as such (CERN preprint TH.4144/85).

The charged W and neutral Z particles are the carriers of



A "monojet" (a) is a jet of energetic charged particles, with energy "missing" in the opposite direction, produced when a proton and an antiproton collide. This could, for example, be due to the decay of a squark and antisquark, into two undetected photinos and a quark and antiquark, each of which produce coalescing jets (b). In the alternative scenario, a W particle produced at the same time as a jet (c) will decay into an unseen neutrino and a charged lepton, which may be lost in the jet

- New Scientist article from June 13, 1985 describing the excitement over an "excess" of monojets in UA1 as well as Ellis, Kleiss, and Stirling's determination that SM backgrounds could explain the excess

And Now

Spring 2017 (ongoing)

Arrive at a joint estimation of theory uncertainties for *precision DM searches* at colliders (e.g. mono-jet)

ATLAS EXPERIMENT erc European Commission Horizon 2020 European Union funding for Research & Innovation LUND UNIVERSITY 3

Motivation and goal for this focused effort:

- Ambitious program of LHC mono-X searches up to HL-LHC needs:
  - Understand impact of electroweak corrections
  - Reaching 1% uncertainties on theoretical modelling
  - Understand correlations between uncertainty sources

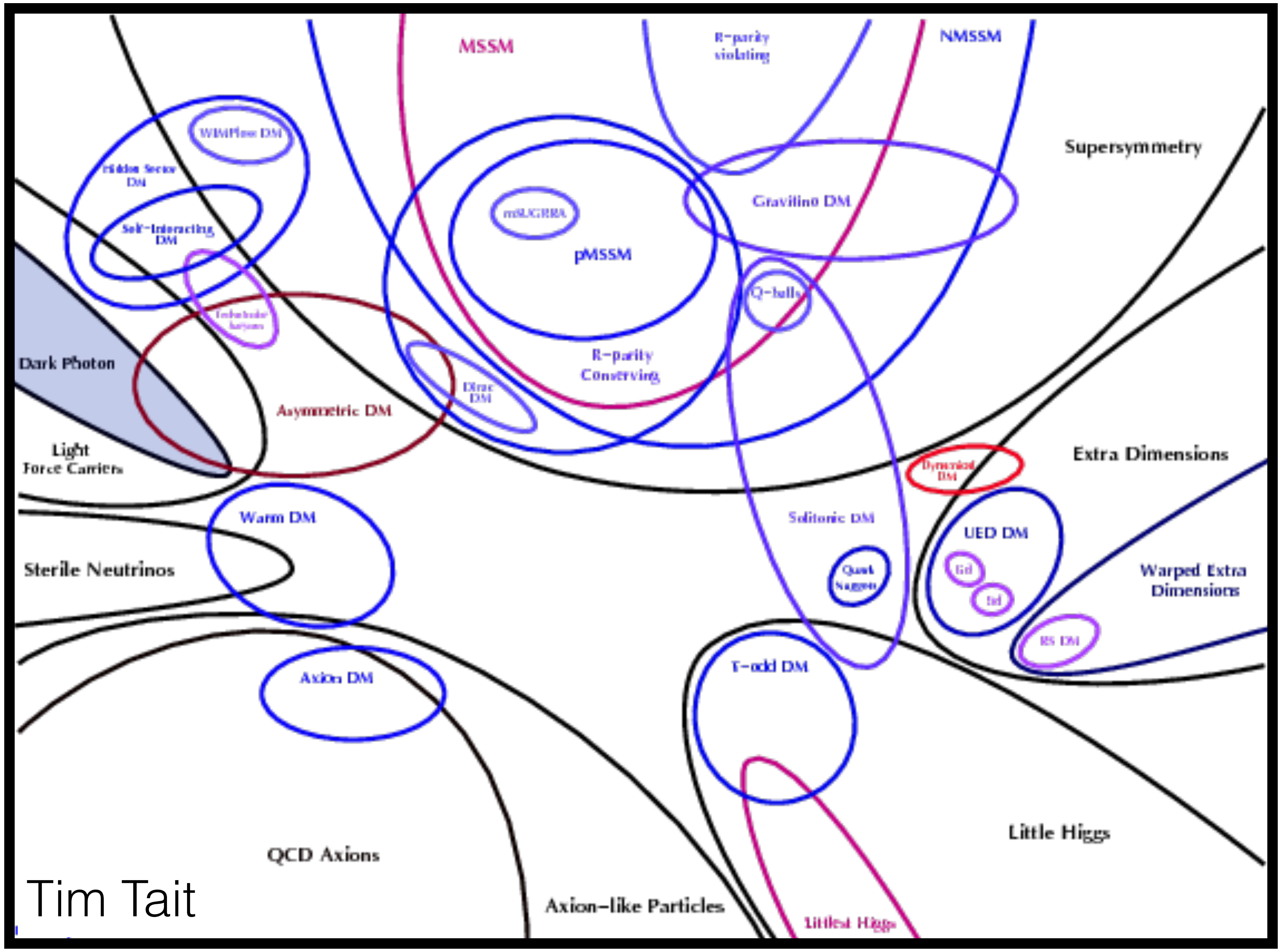
from Caterina Doglioni's talk on Monday

- See also Stefano Pozzorini's talk went into the details of the theoretical calculations needed to make this a reality

And Now

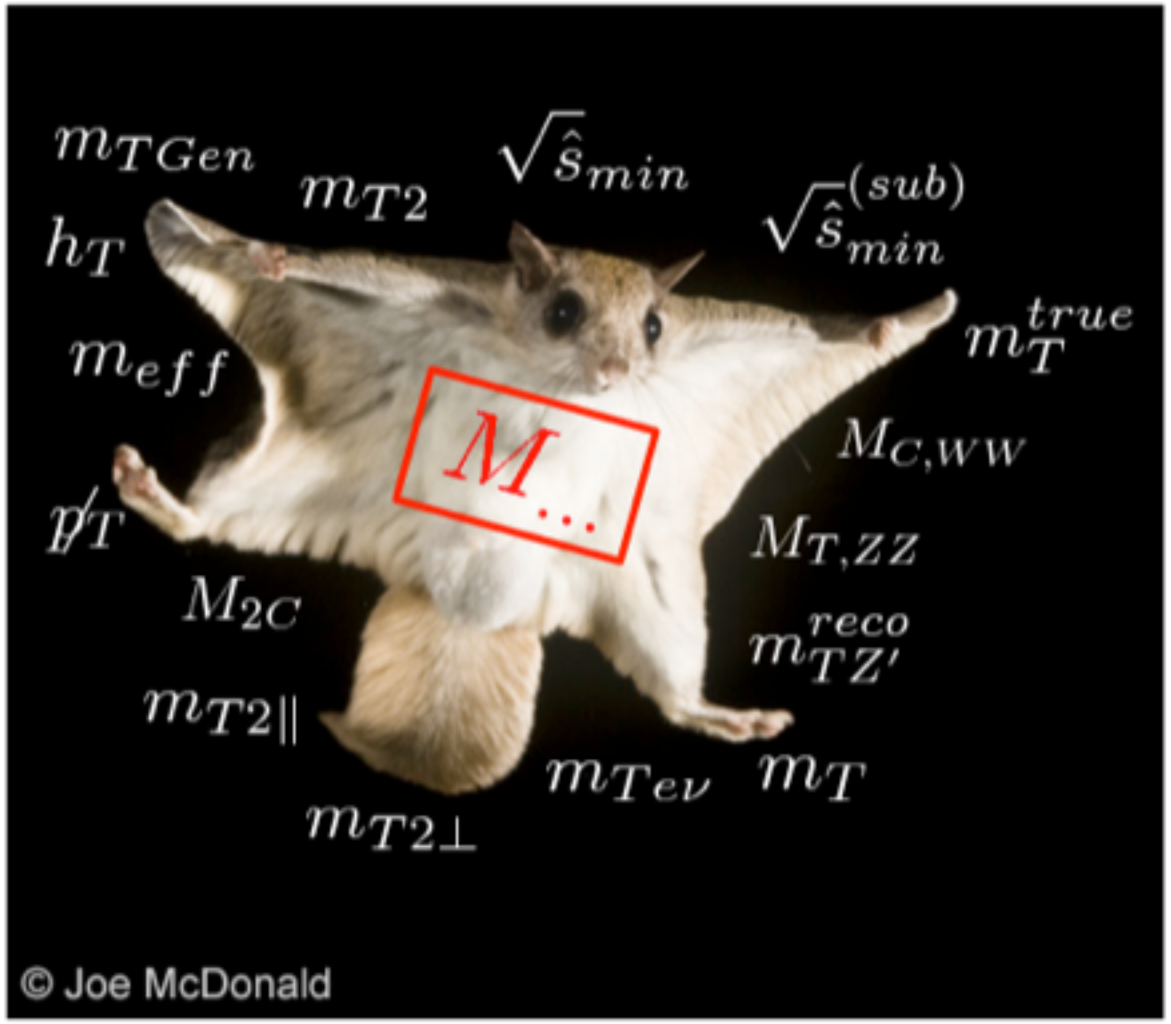
- I think a major theme of the monojet talks here is that we are entering the precision regime
- As confidence in experimental and theoretical modeling of backgrounds and signals grows, MVAs will play an increasing role
- How can we model-building/ testing theorists take this development into account in a useful way?
- As Doraval pointed out yesterday, reducing systematic errors in signal models can be challenging
- The challenge is heightened by the fact that there are…

Many Models…

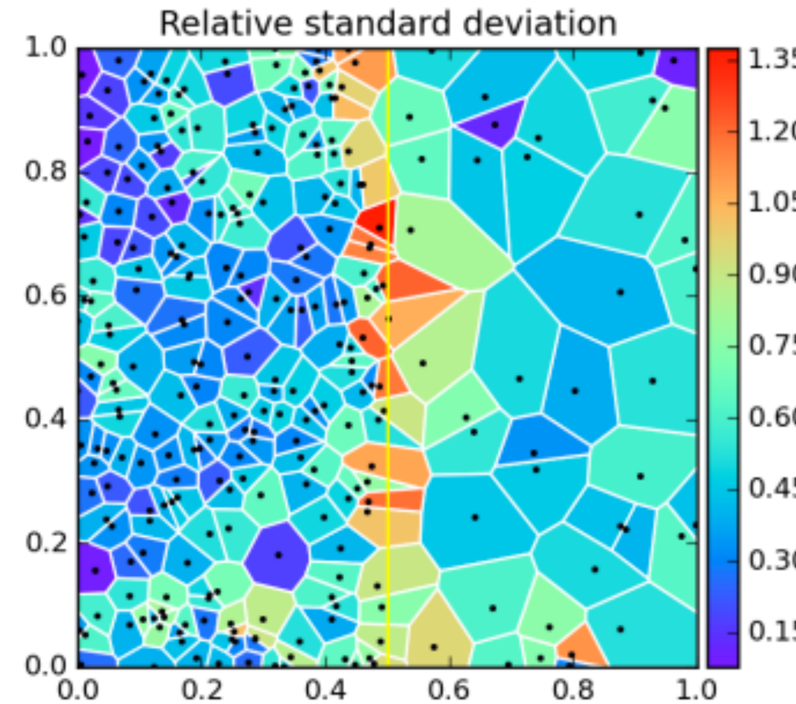


- Benchmarks essential
- But many possibilities, signatures
- We want to explore as many possibilities as possible, as well as possible

and Many Variables…



(Barr, Khoo, Konar, Kong, Lester, Matchev, Park, 2011)



(Debnath, JG, Kilic, Kim, Matchev, and Yang, 2016)

Also old standbys like jet pT, missing ET, HT, sphericity, delta phi, etc.

- Many choices of variables (not all independent)
- Need to determine best variables to add: may depend on signal model
- Useful to be able to understand/ project experimental choices, and/or suggest good variables for new/ challenging models

My Proposal: Quantile Binning

- My proposal is to consider the signal cross section in quantile bins for the background in the variable(s) of interest
- Quantile bins: construct bins so that each bin has the same number of events, or same cross section
- So we pick a variable, make bins in that variable so that each bin has the same background cross section, and then see what the signal cross section in each bin is
- (sort of an unboosted decision tree: we don’t vary our bin limits to optimize sensitivity)
- **Pros**:
    - easily investigate sensitivity of physics variables
    - some systematics will (somewhat!) cancel, e.g. effects (detector or theory) common to signal and background
- **Con**: not as sensitive as using full distributions.
    - But this is okay: point is to let theorists determine good variables for studying new model points and project future sensitivities in a not-totally-crazy way.
      Experiments can then do better with neural nets, etc.

Example: CMS Monojet Search

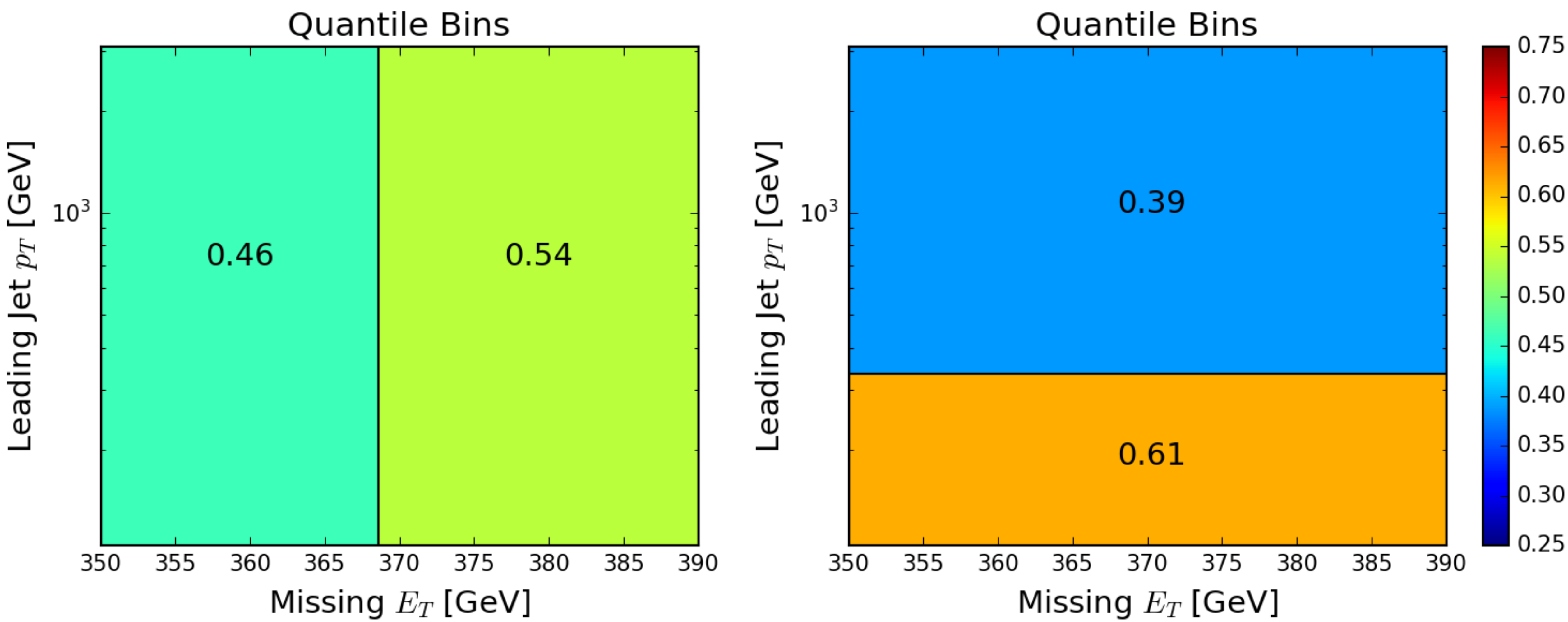
- See Osamu Jinnouchi's talk from Monday
- ATLAS most recent monojet search
  Phys. Rev. D94 (2016) 032005 (1604.07773)
  3.2 fb^{-1}
- CMS most recent monojet search
  PAS EXO-16-037 (1703.01651).
  12.9 fb^{-1}
- Using the CMS search as the basis for my example
- My signal model will be a 600 GeV stop decaying to a 570 GeV neutralino + charm jet (no charm tagging, though)
- That analysis uses jet mass/ substructure to also do a mono-V search, but I am restricting my attention to the mono-jet search

Example: CMS Monojet Search

E_T^{miss} [GeV]	$Z(\nu\bar{\nu})$+jets	$W(\ell\nu)$+jets	Top quark	Dibosons	Other	Total bkg.	Observed
200–230	71300 ± 2200	54600 ± 2300	2140 ± 320	1320 ± 220	2470 ± 310	132100 ± 4000	140642
230–260	39500 ± 1300	27500 ± 1200	1060 ± 160	790 ± 130	1090 ± 130	69900 ± 2200	73114
260–290	21900 ± 670	13600 ± 550	440 ± 65	364 ± 61	498 ± 65	36800 ± 1100	38321
290–320	12900 ± 400	7300 ± 290	210 ± 31	235 ± 40	216 ± 30	20780 ± 630	21417
320–350	8000 ± 280	4000 ± 170	107 ± 16	145 ± 24	124 ± 18	12340 ± 400	12525
350–390	6100 ± 220	2800 ± 130	74 ± 11	111 ± 19	87 ± 13	9160 ± 320	9515
390–430	3500 ± 160	1434 ± 66	30.1 ± 4.5	58.4 ± 9.9	33.4 ± 5.3	5100 ± 200	5174
430–470	2100 ± 98	816 ± 37	16.6 ± 2.5	42.4 ± 7.1	16.3 ± 2.7	3000 ± 120	2947
470–510	1300 ± 66	450 ± 20	7.4 ± 1.1	24.6 ± 4.1	9.6 ± 1.6	1763 ± 79	1777
510–550	735 ± 39	266 ± 13	5.2 ± 0.8	18.5 ± 3.1	7.0 ± 1.3	1032 ± 48	1021
550–590	513 ± 31	152 ± 8	2.4 ± 0.4	13.5 ± 2.3	1.1 ± 0.3	683 ± 37	694
590–640	419 ± 23	120 ± 6	1.5 ± 0.2	10.6 ± 1.8	2.1 ± 0.4	554 ± 28	554
640–690	246 ± 16	62.8 ± 3.8	1.3 ± 0.2	11.4 ± 1.9	1.0 ± 0.2	322 ± 19	339
690–740	139 ± 11	34.2 ± 2.4	0.6 ± 0.1	4.2 ± 0.7	0.20 ± 0.07	178 ± 13	196
740–790	97.2 ± 7.2	22.7 ± 1.7	0.22 ± 0.03	1.4 ± 0.2	0.63 ± 0.12	122 ± 8	123
790–840	59.8 ± 5.8	12.9 ± 1.2	0.13 ± 0.02	1.5 ± 0.3	0.05 ± 0.02	74.5 ± 6.6	80
840–900	64.3 ± 6.4	12.3 ± 1.1	0.24 ± 0.04	0.92 ± 0.1	0.03 ± 0.01	77.8 ± 7.2	68
900–960	31.5 ± 4.3	6.0 ± 0.7	0.21 ± 0.03	0.74 ± 0.1	0.01 ± 0.01	38.4 ± 4.8	37
960–1020	20.8 ± 3.0	3.4 ± 0.5	—	0.94 ± 0.2	0.01 ± 0.01	25.1 ± 3.4	23
1020–1090	16.3 ± 2.6	3.1 ± 0.5	0.04 ± 0.01	1.6 ± 0.3	0.01 ± 0.01	21.1 ± 3.0	12
1090–1160	8.1 ± 1.8	1.3 ± 0.3	—	—	—	9.4 ± 1.9	7
>1160	18.6 ± 2.7	2.7 ± 0.4	—	1.3 ± 0.2	—	22.6 ± 3.0	26

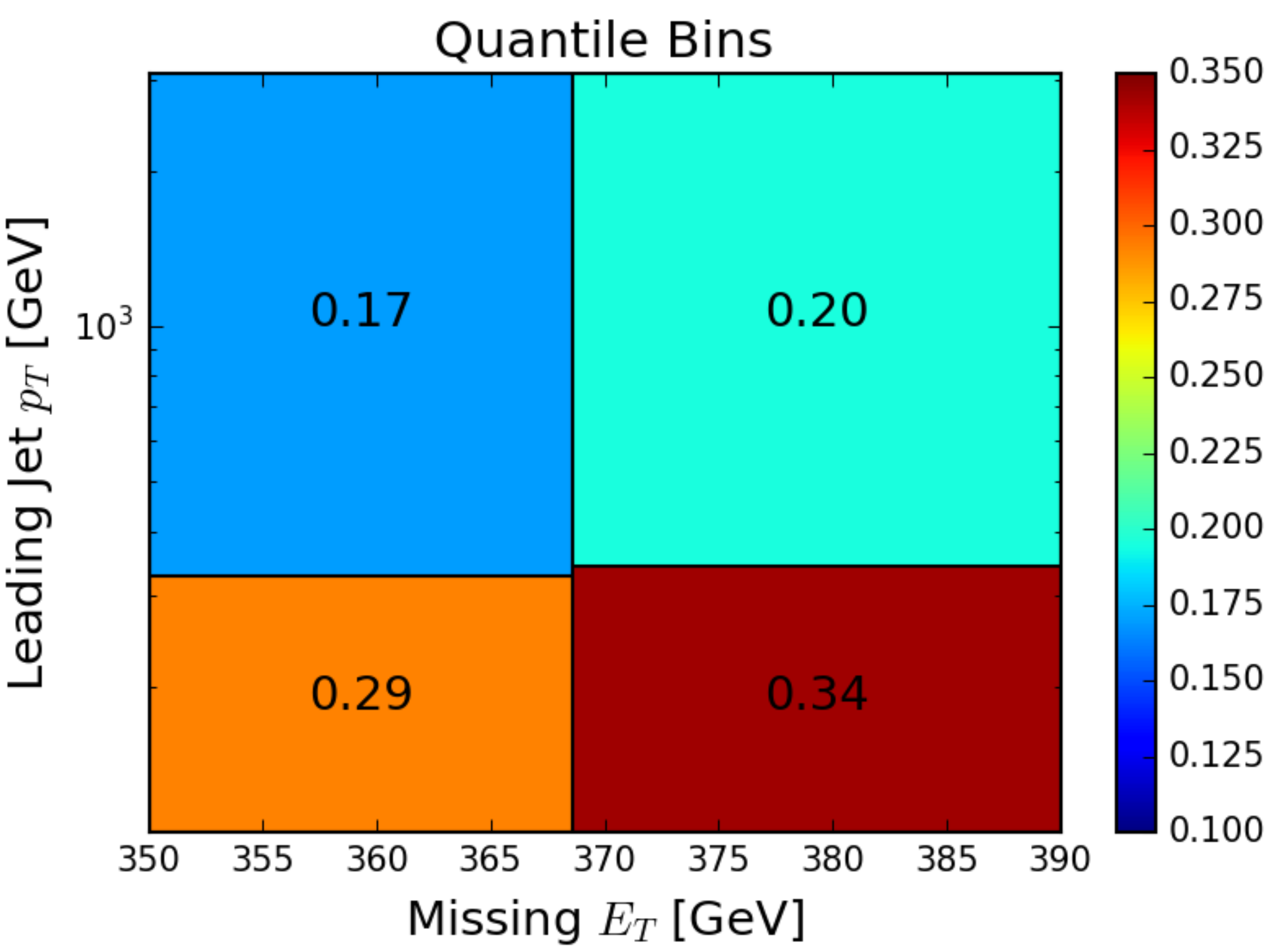
- We're looking at a single missing E_T bin

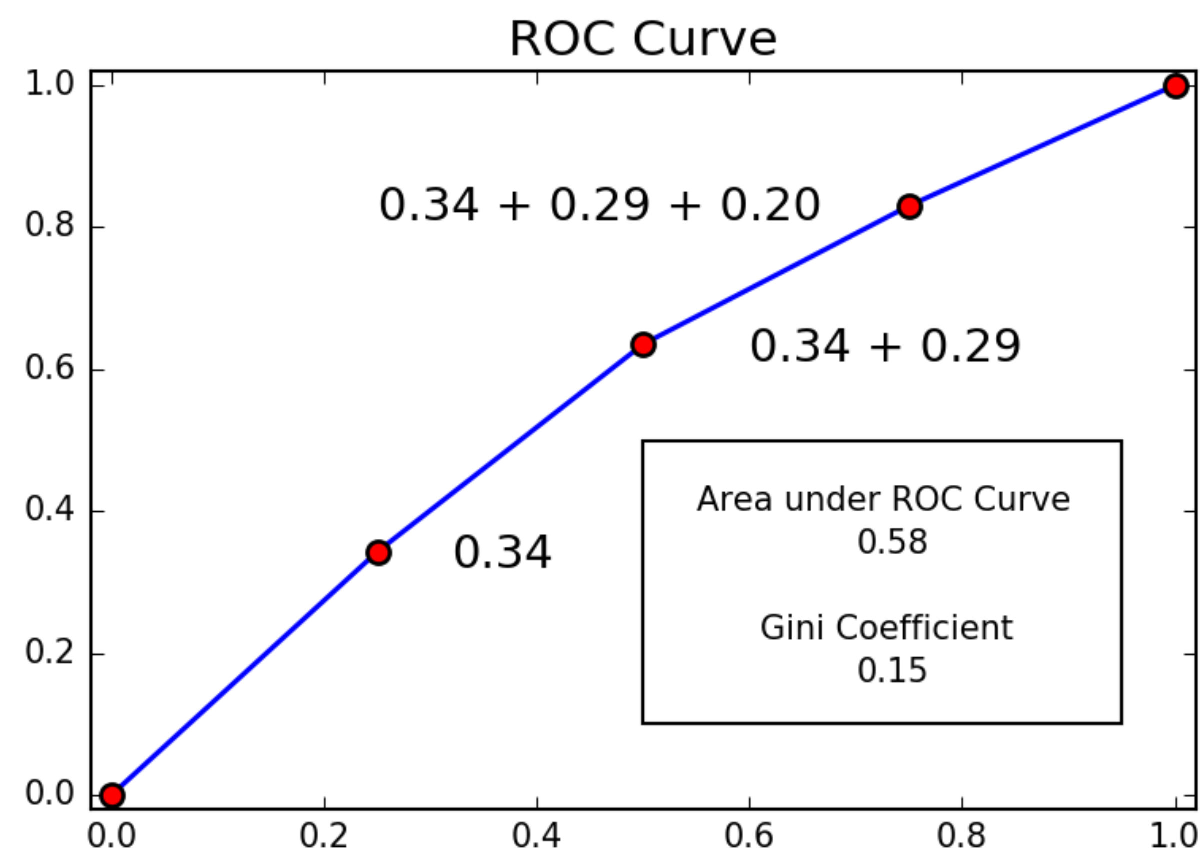
Example: CMS Monojet Search



- Within our bin, dividing into quantile bins in leading jet pT is more useful than a further subdivision in Missing E_T

Example: CMS Monojet Search





- Here we consider quantile bins in p_T and MET, slight increase in sensitivity over the two p_T quantile bin analysis
- Simple translation between signal fractions in quantile bins and ROC curves: common measure of the sensitivity of analyses

Conclusions

- MVA are a huge part of experimental searches
  - We theorists need to respond to this development in a useful way
- Simple MVAs like quantile binning provide a simple and robust framework for evaluating variables and communicating with experiment
- May be especially useful in the context of dark matter searches