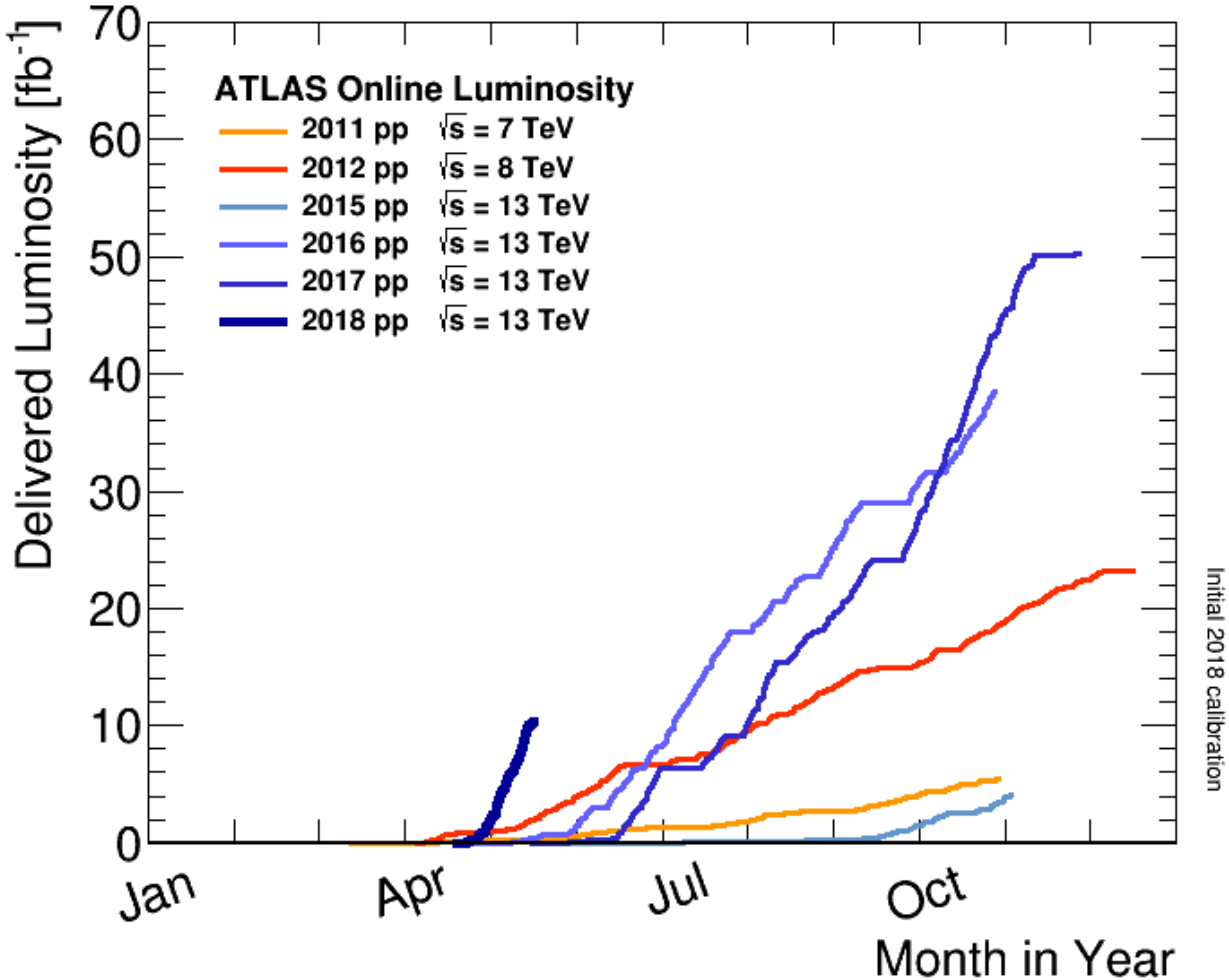
An Update on the LHC Monojet Excess

David Shih
NHETC, Rutgers University

Planck 2018, Bonn

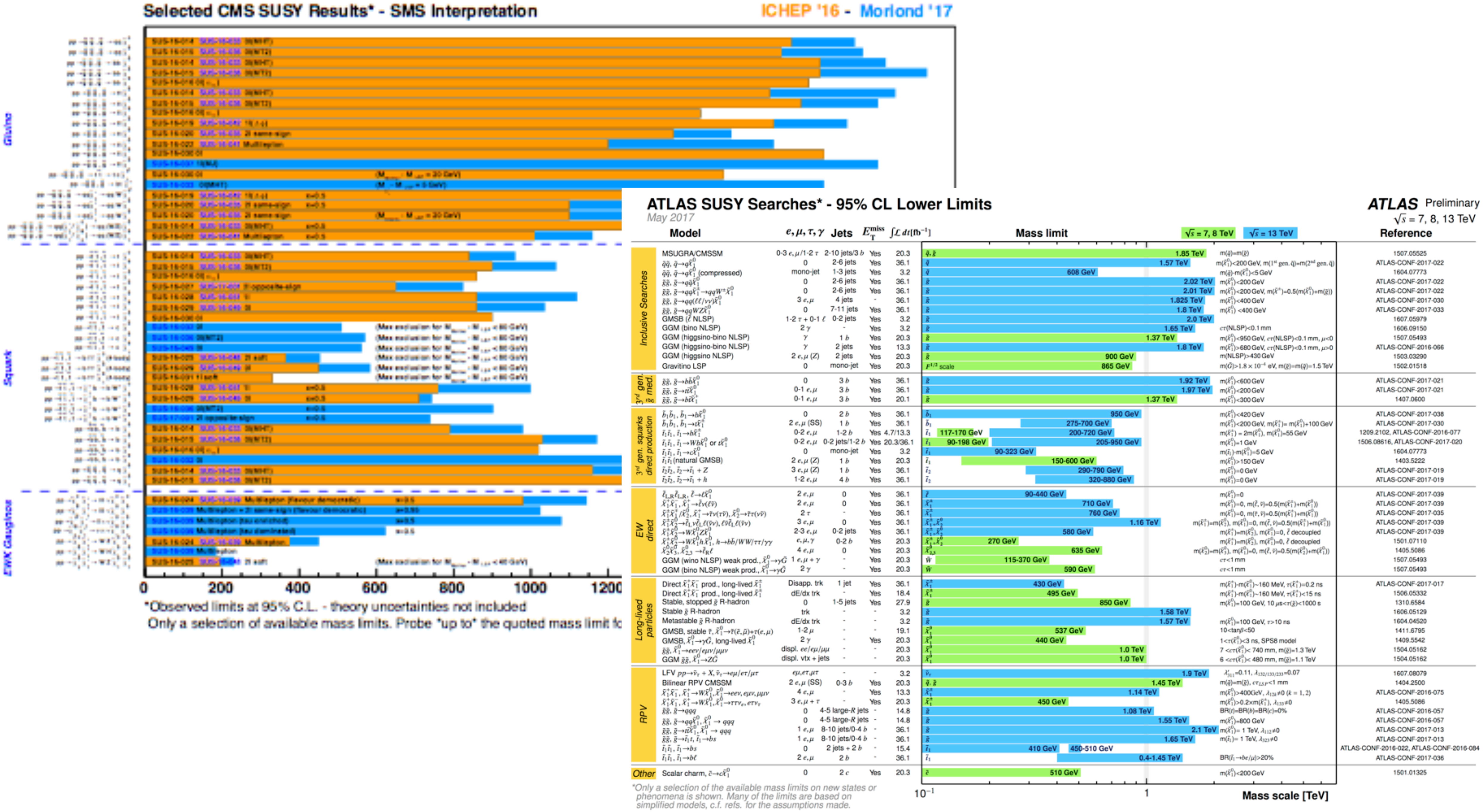
Pouya Asadi, Matt Buckley, Anthony DiFranzo, Angelo Monteux & DS
1707.05783, 1712.04939 and work in progress

The LHC has been performing spectacularly.



Last year, we reached an important milestone:

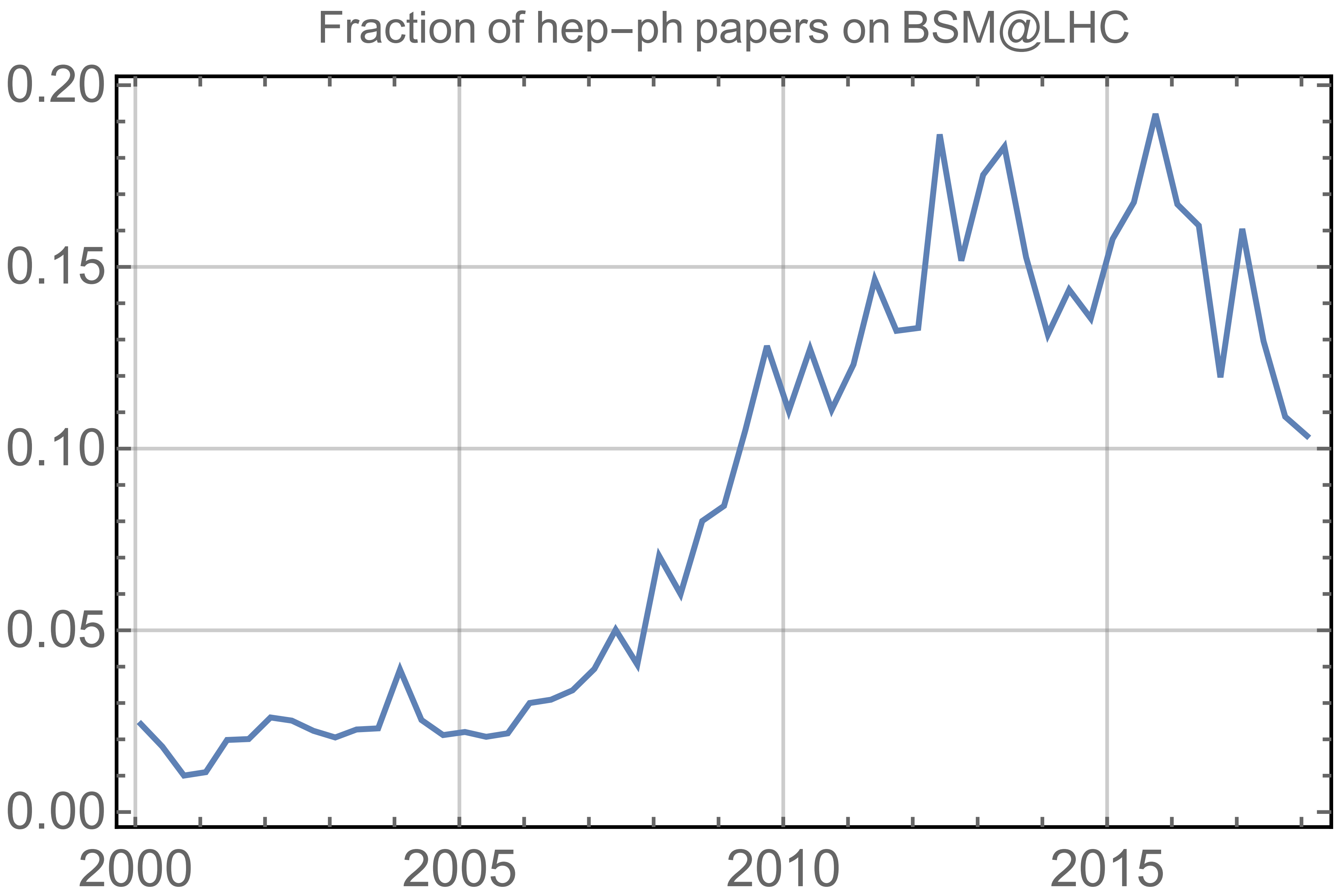
~40/fb at 13 TeV



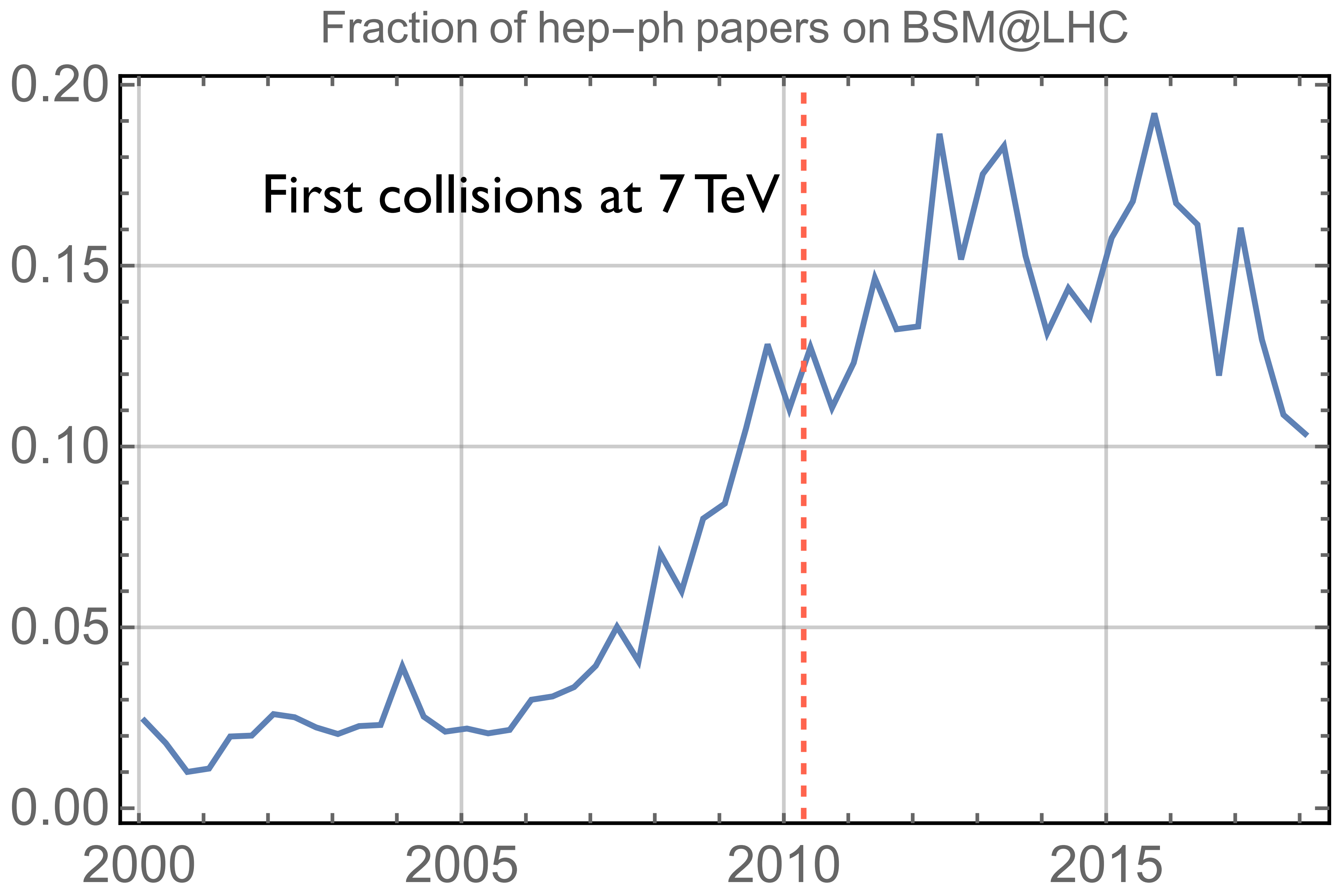
Many null results were presented by ATLAS and CMS.

Where are the discrepancies, excesses, anomalies?

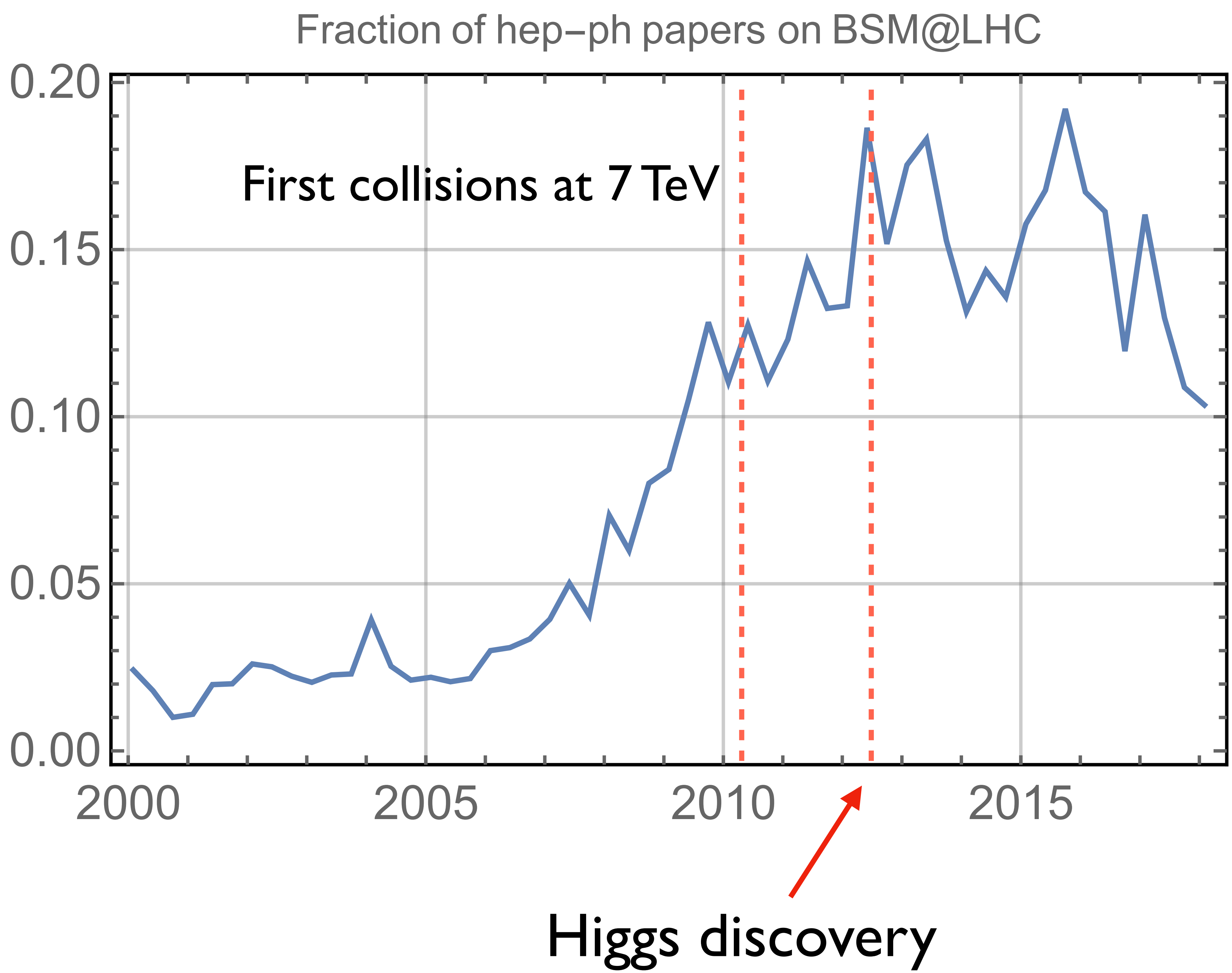
Theorists are losing interest…



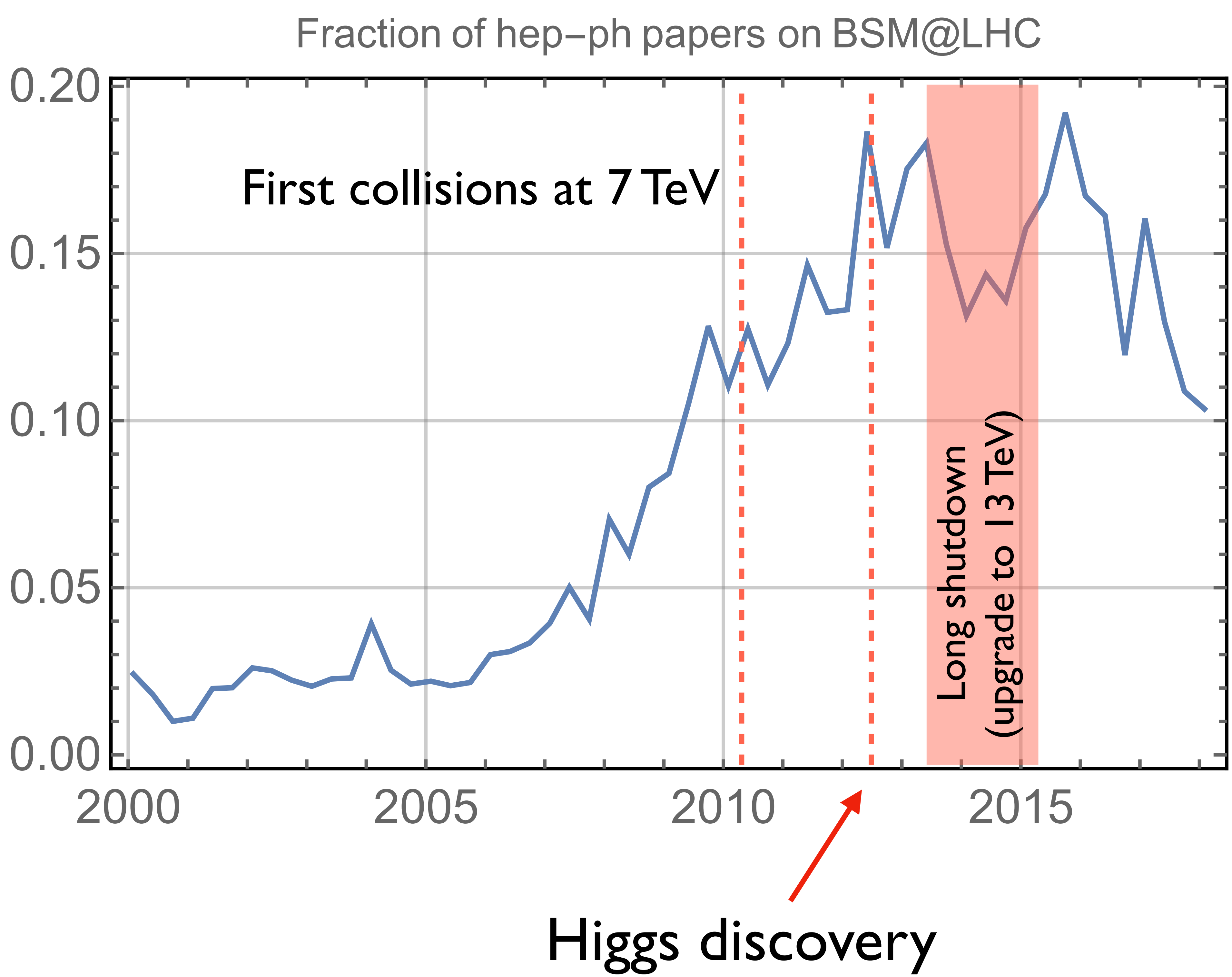
Theorists are losing interest…



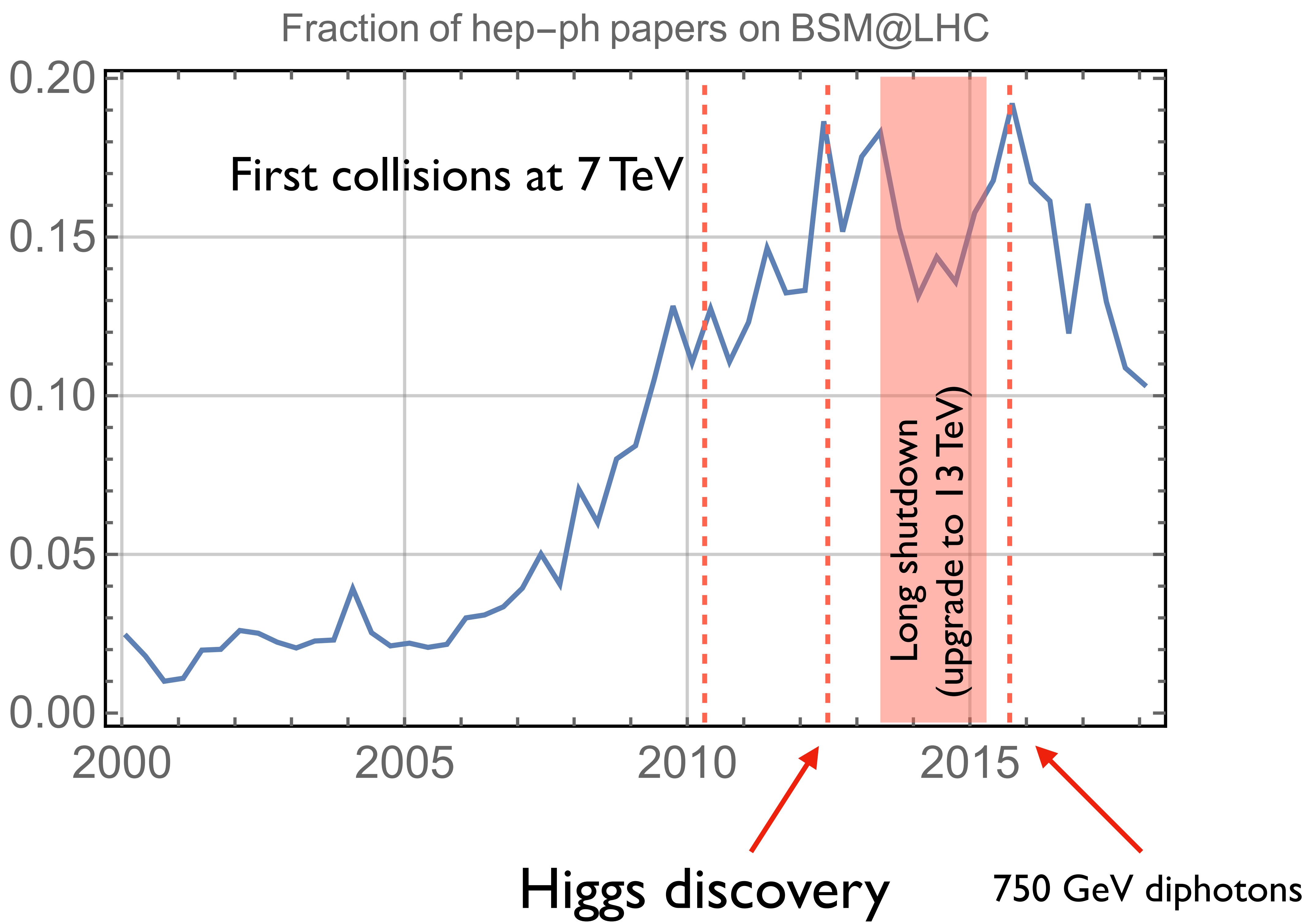
Theorists are losing interest…



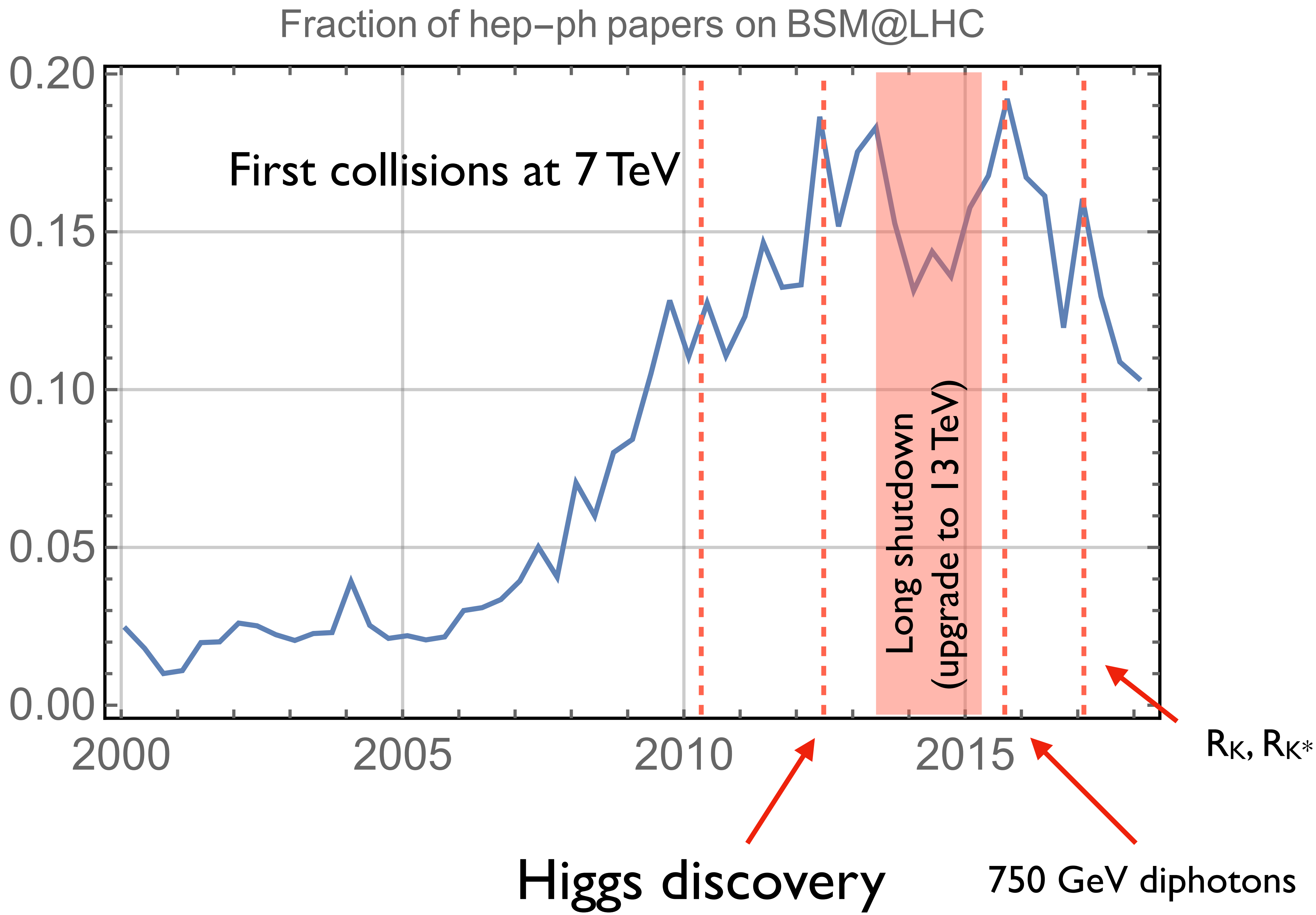
Theorists are losing interest…



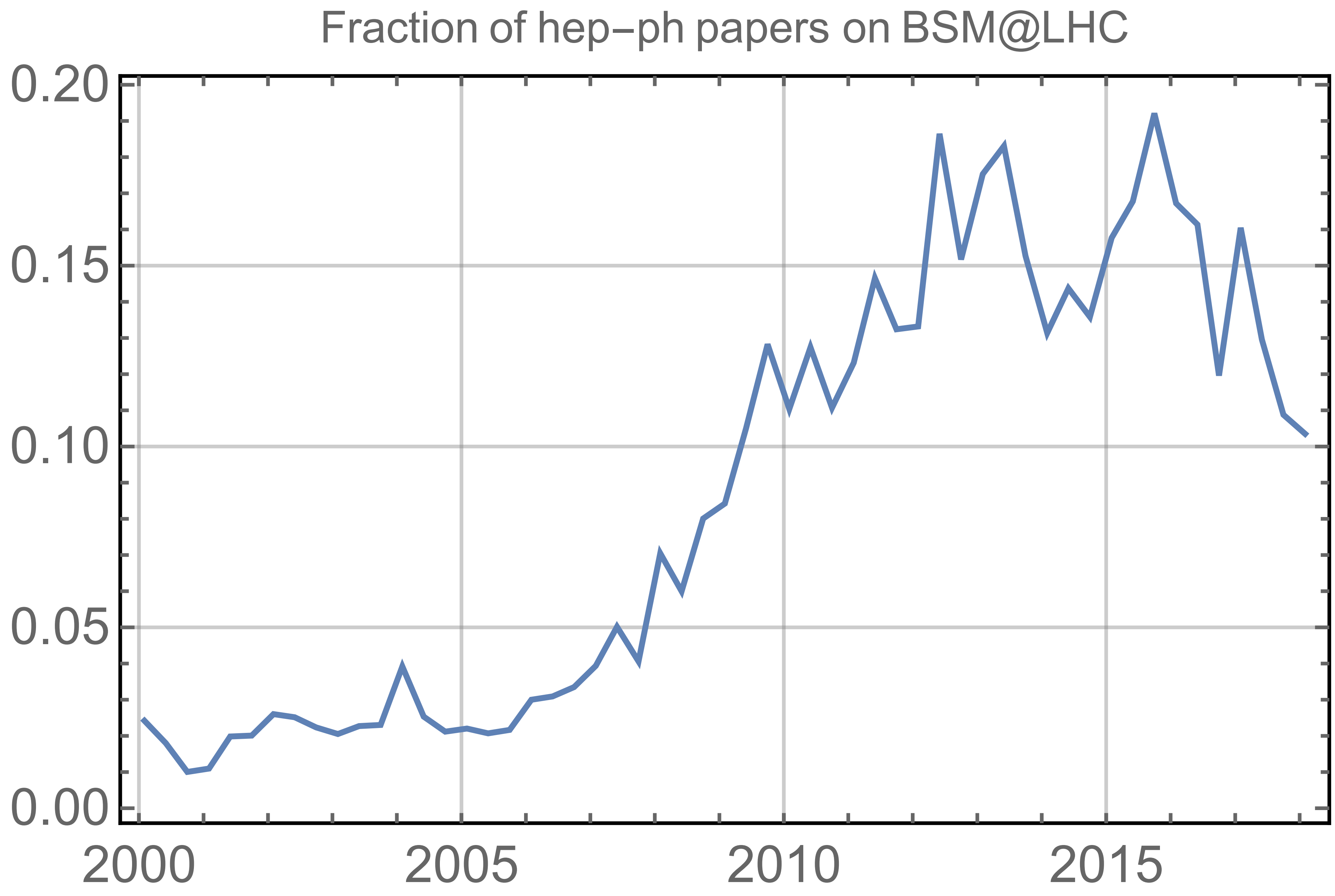
Theorists are losing interest…



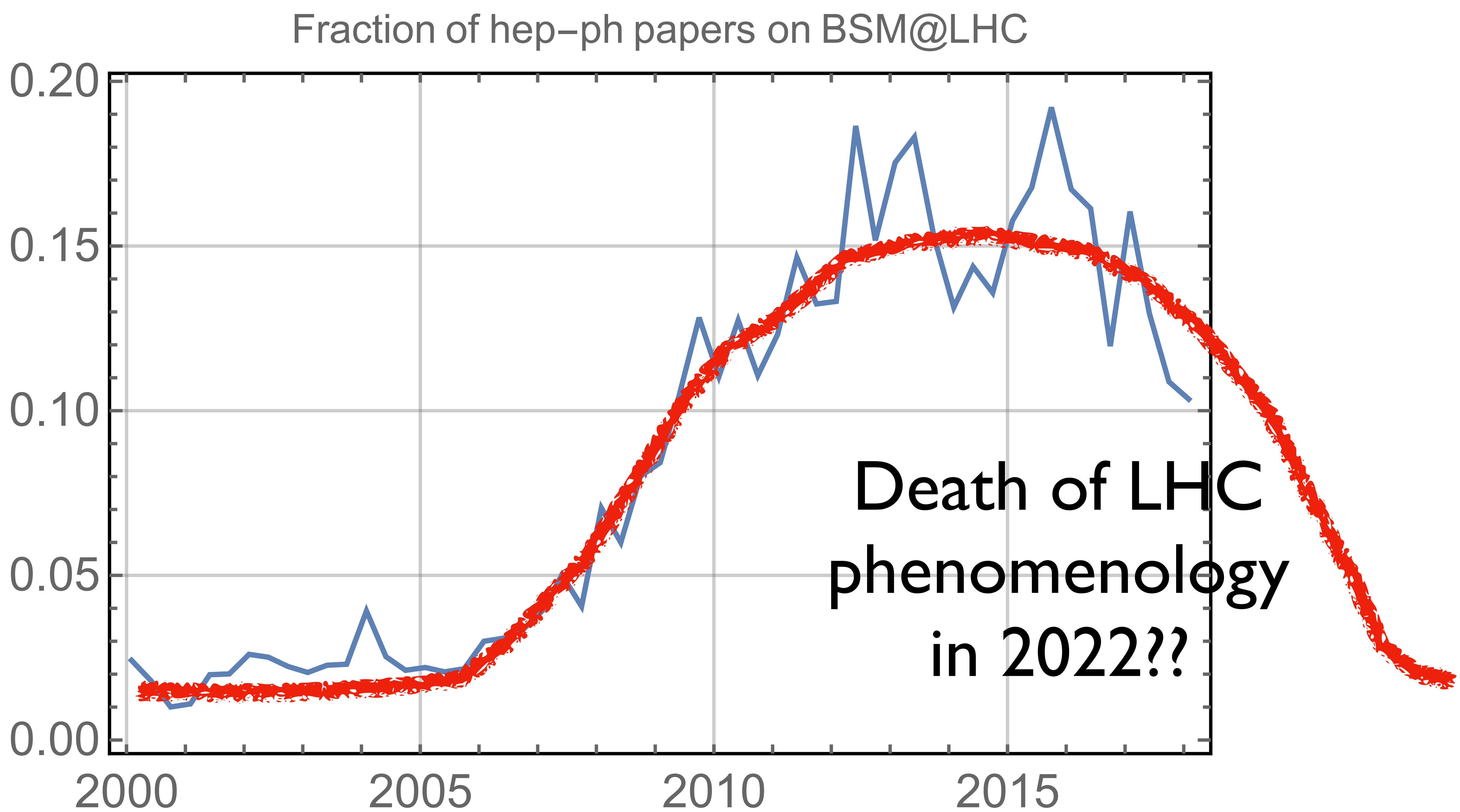
Theorists are losing interest…



Theorists are losing interest…

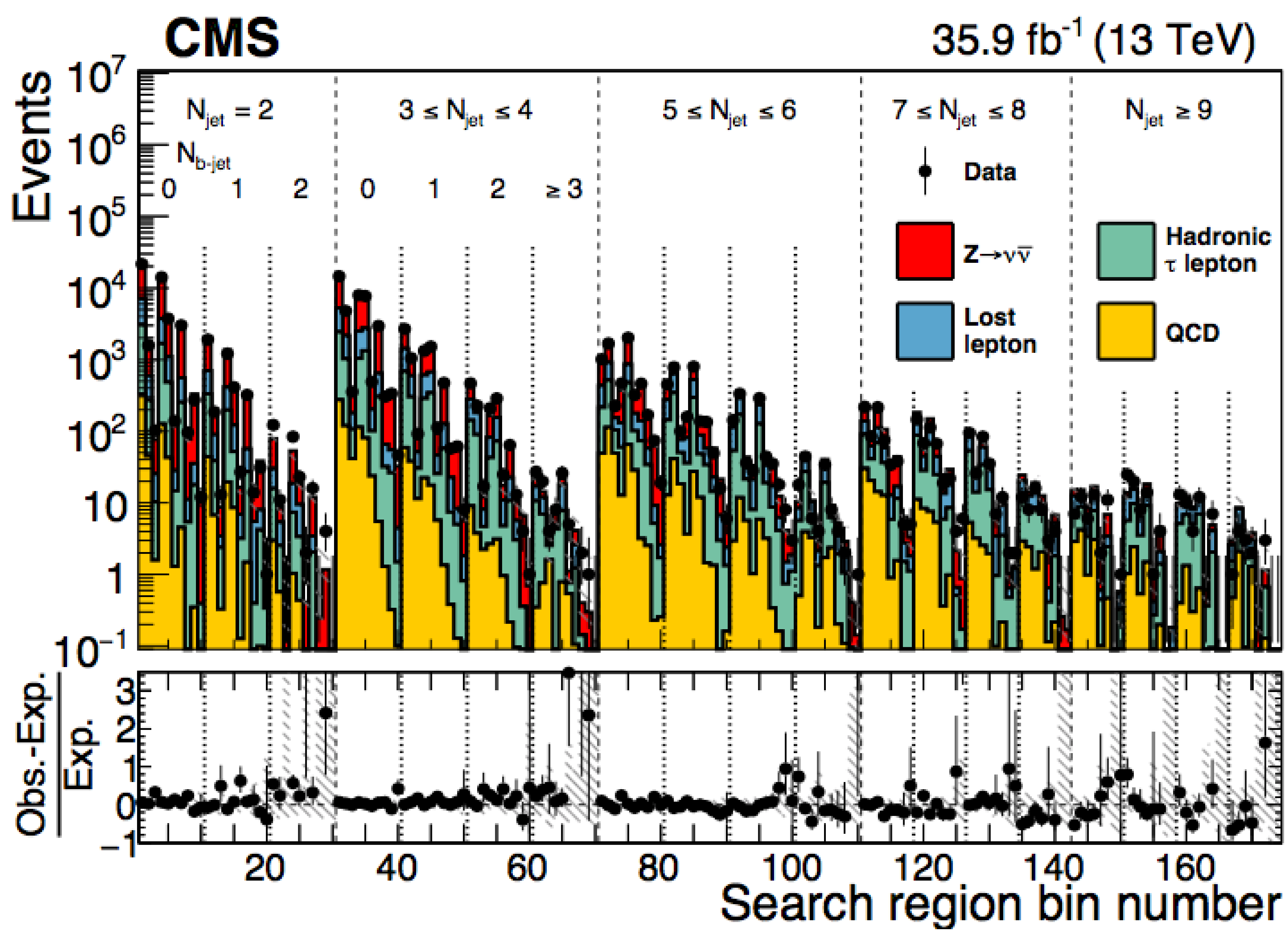


Theorists are losing interest…



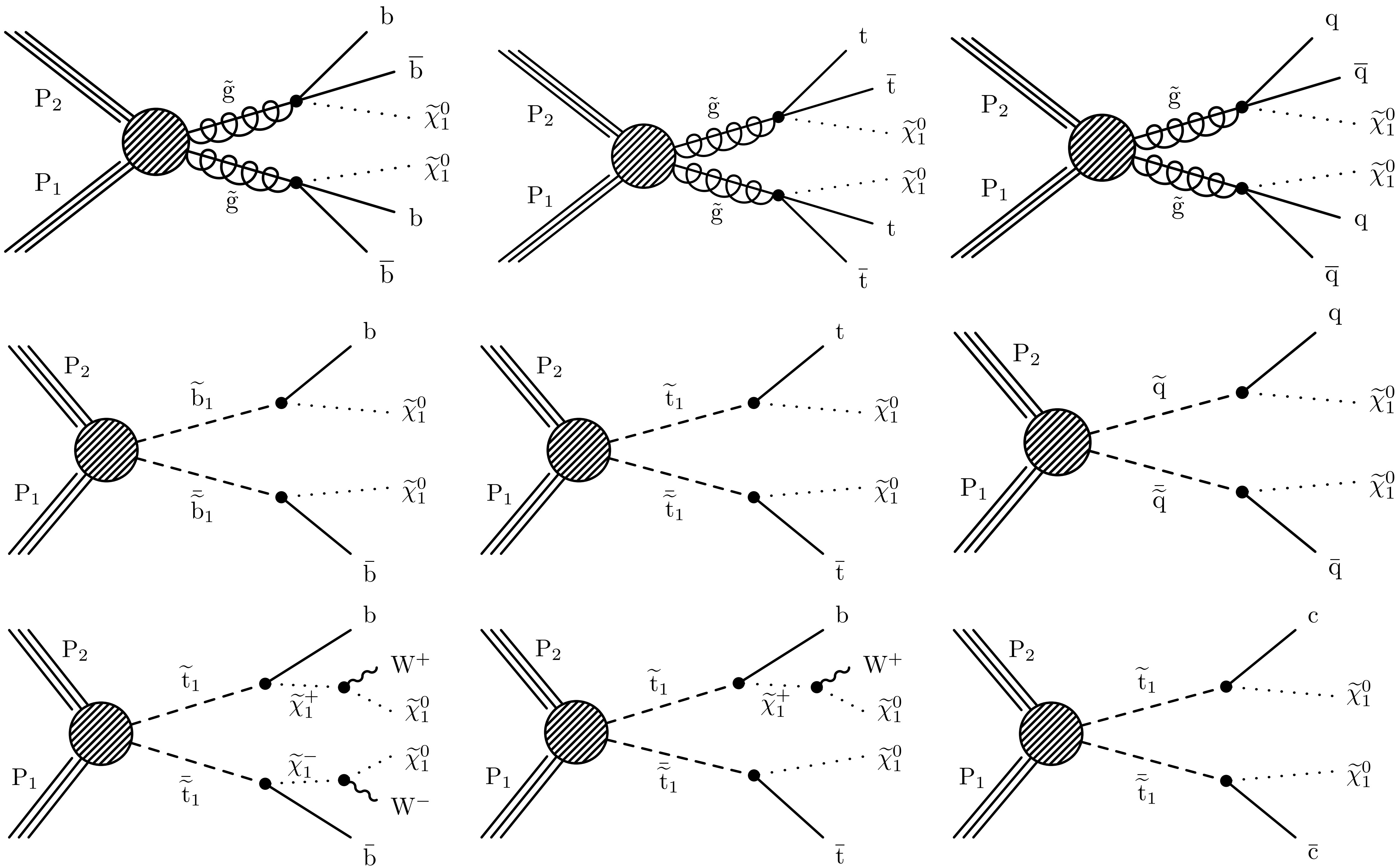
Is there really nothing interesting going on in the LHC data?

There can be hundreds of SRs in a typical LHC search…



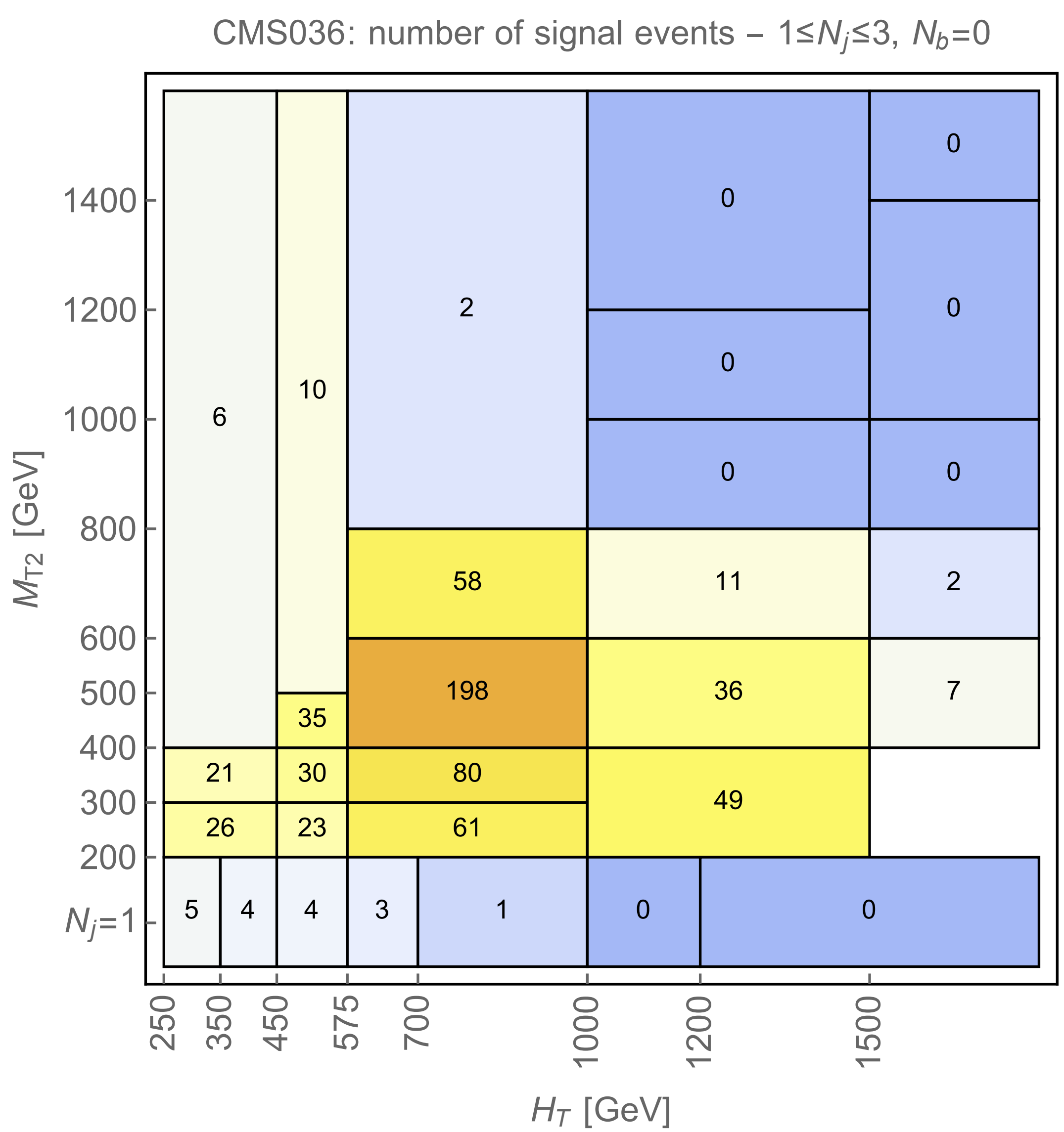
CMS-PAS-SUS-16-033 (jets+MHT)

…but all the limits are actually derived from a small set of simplified models…



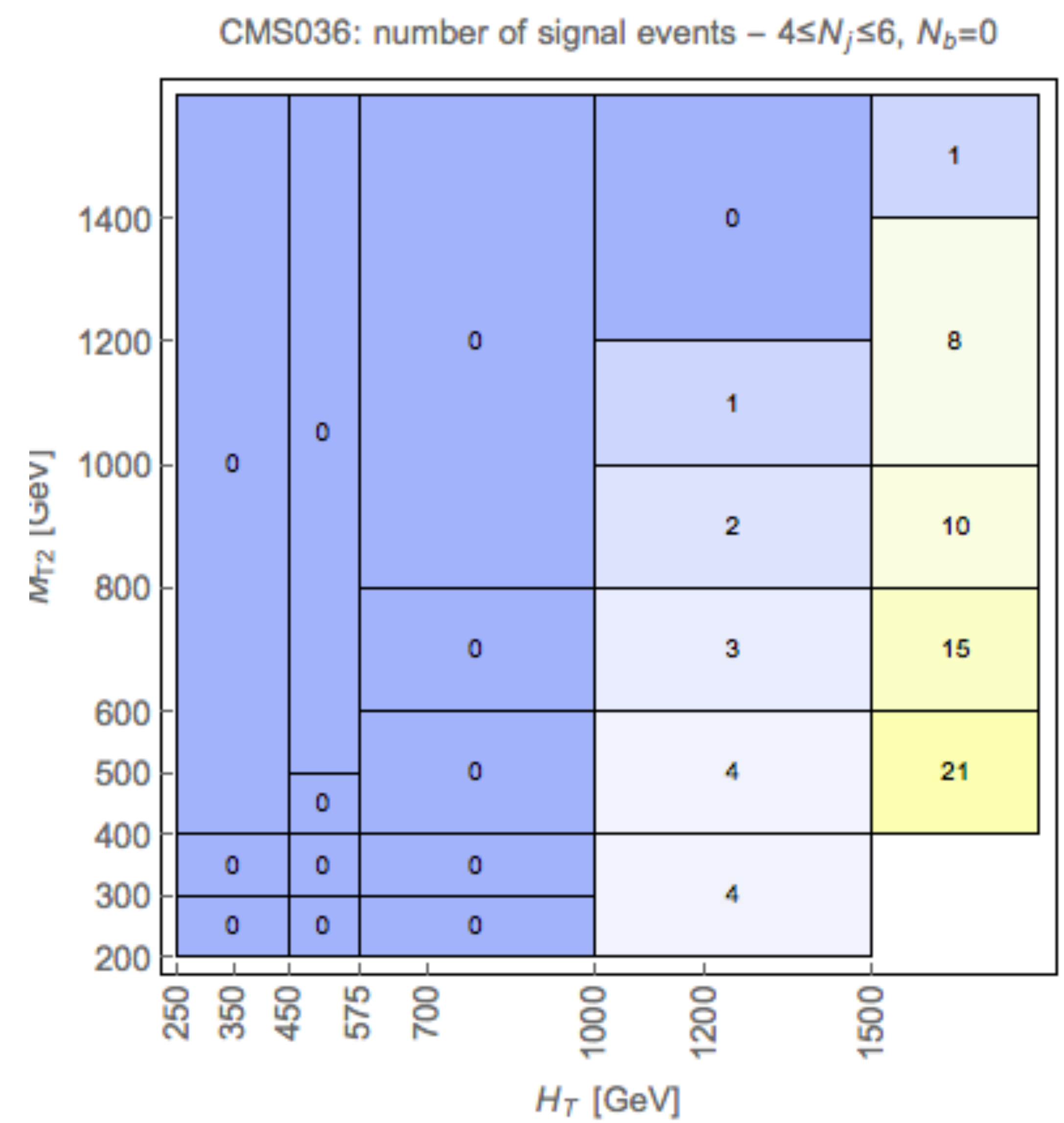
…and these simplified models only probe a small fraction of the SRs…

CMS-PAS-SUS-16-036 (jets+MT2)



$$pp \to \tilde{q}\tilde{q}, \quad \tilde{q} \to q\tilde{\chi}_1^0$$

$$m_{\tilde{q}} = 800 \text{ GeV}, \quad m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$$



$$pp \to \tilde{g}\tilde{g}, \quad \tilde{g} \to qq\tilde{\chi}_1^0$$

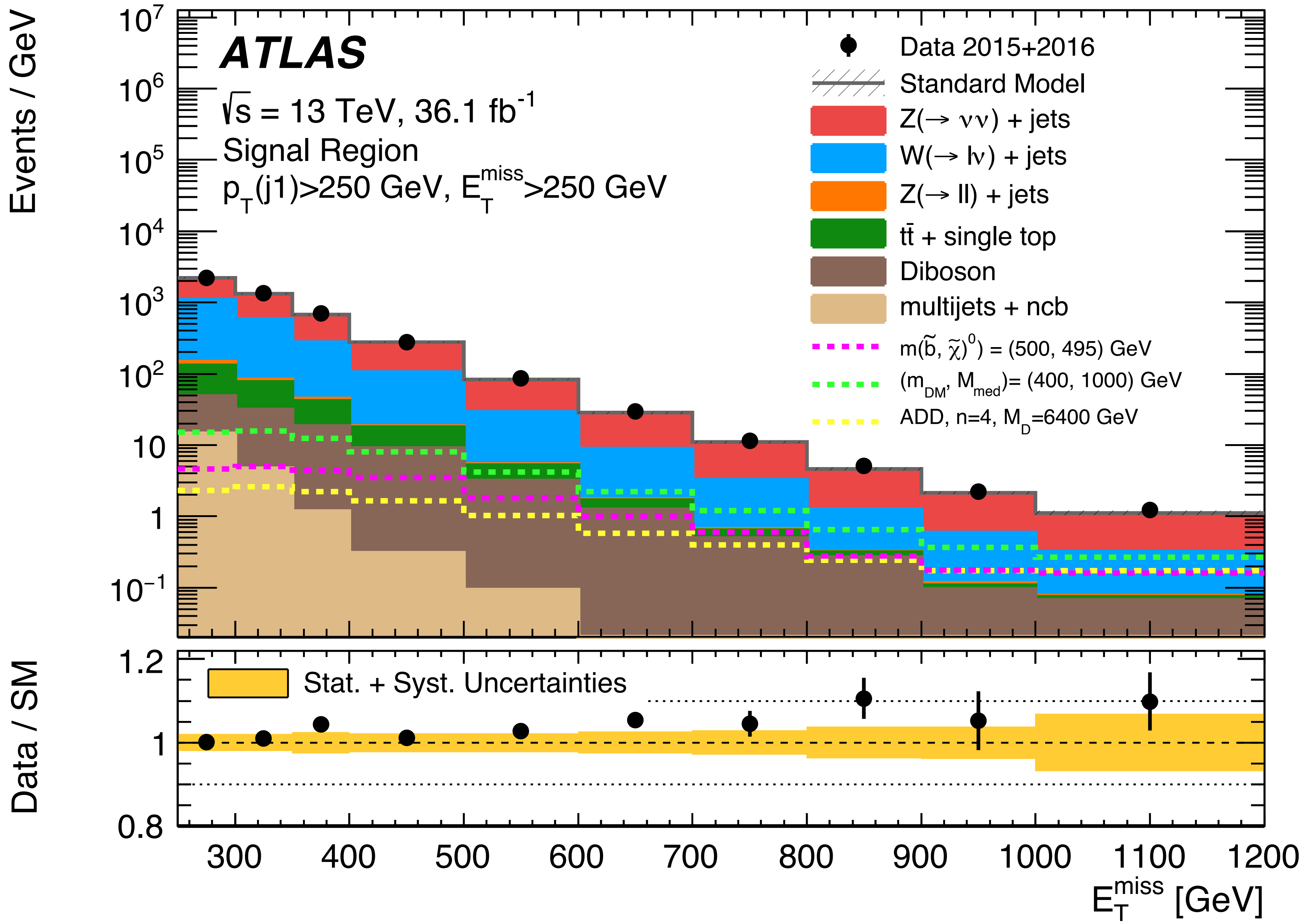
$$m_{\tilde{g}} = 1600 \text{ GeV}, \quad m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$$

What is going on in the rest of the data not probed by any "official" simplified model?

Nobody knows!

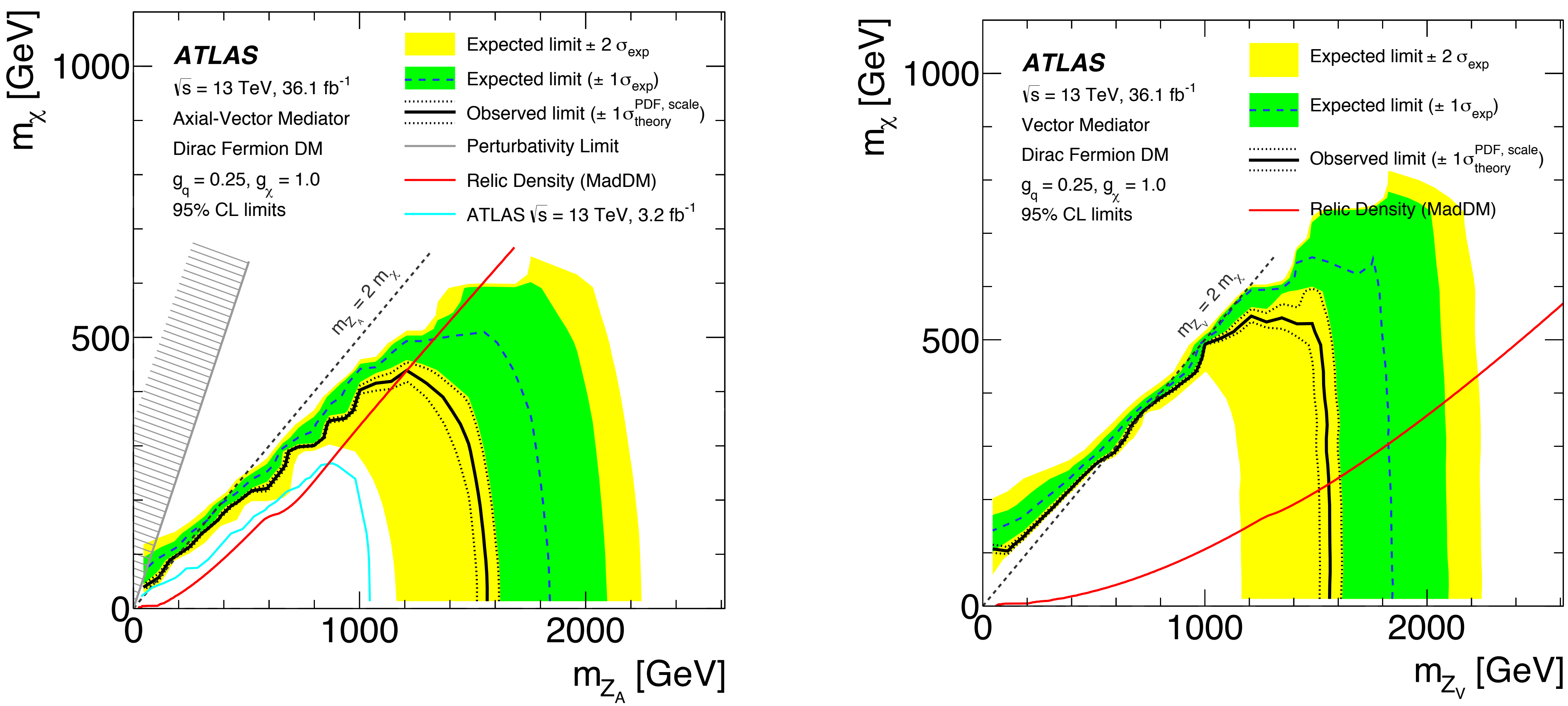
There could be potentially interesting discrepancies there!

Case in point: ATLAS monojets



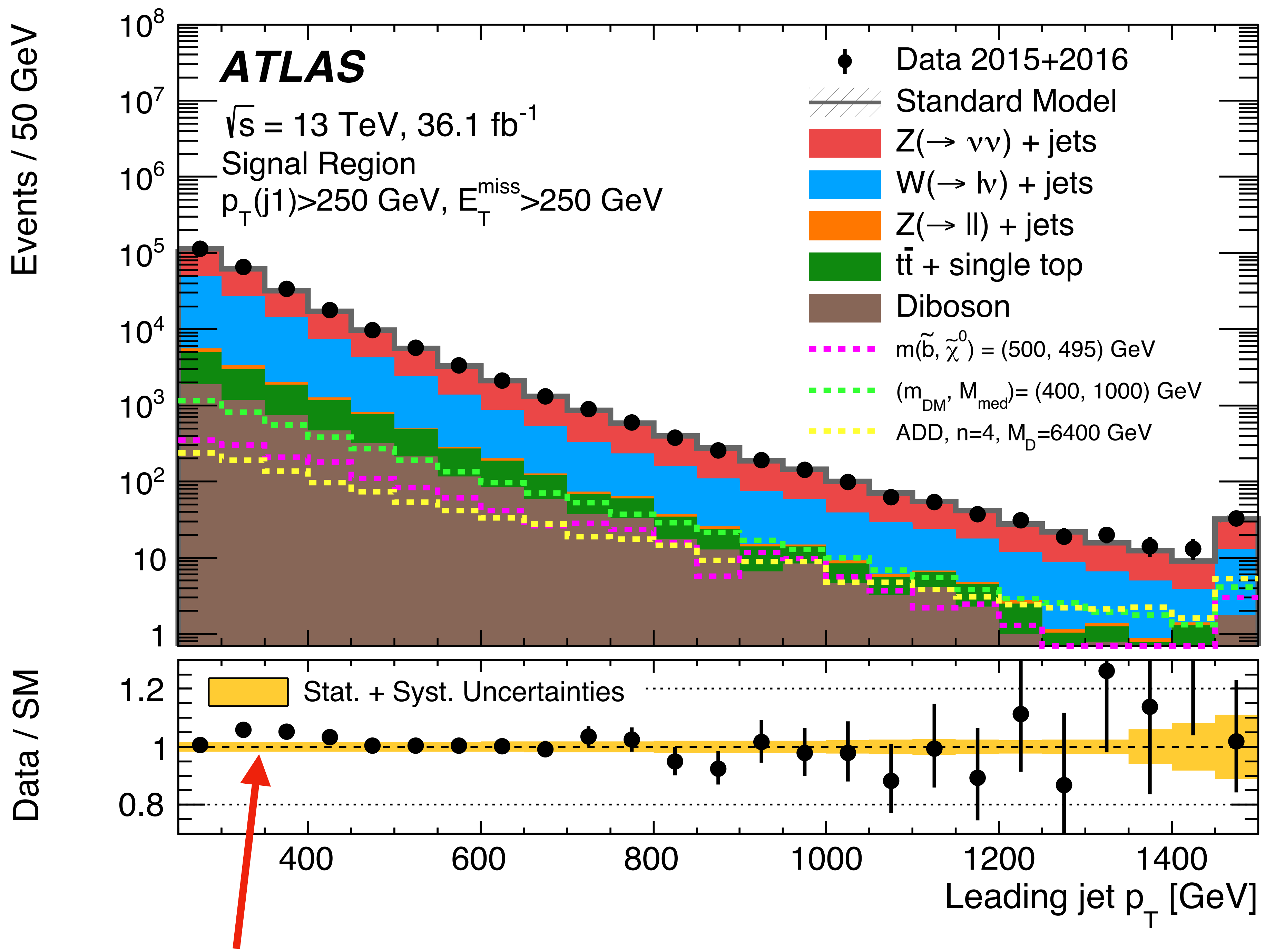
MET distribution looks mostly fine, maybe a bit high…

Nothing but limits and "nothing-to-see-here" boilerplate officially…



Results of a search for new phenomena in final states with an energetic jet and large missing transverse momentum are reported. The search uses proton–proton collision data corresponding to an integrated luminosity of 36.1 fb^{-1} at a centre-of-mass energy of 13 TeV collected in 2015 and 2016 with the ATLAS detector at the Large Hadron Collider. Events are required to have at least one jet with a transverse momentum above 250 GeV and no leptons (e or μ). Several signal regions are considered with increasing requirements on the missing transverse momentum above 250 GeV. Good agreement is observed between the number of events in data and Standard Model predictions. The results are translated into exclusion limits in models with pair-produced weakly interacting dark-matter candidates, large extra spatial dimensions, and supersymmetric particles in several compressed scenarios.

But taking a closer look reveals pretty big discrepancies in the leading jet pT distribution…



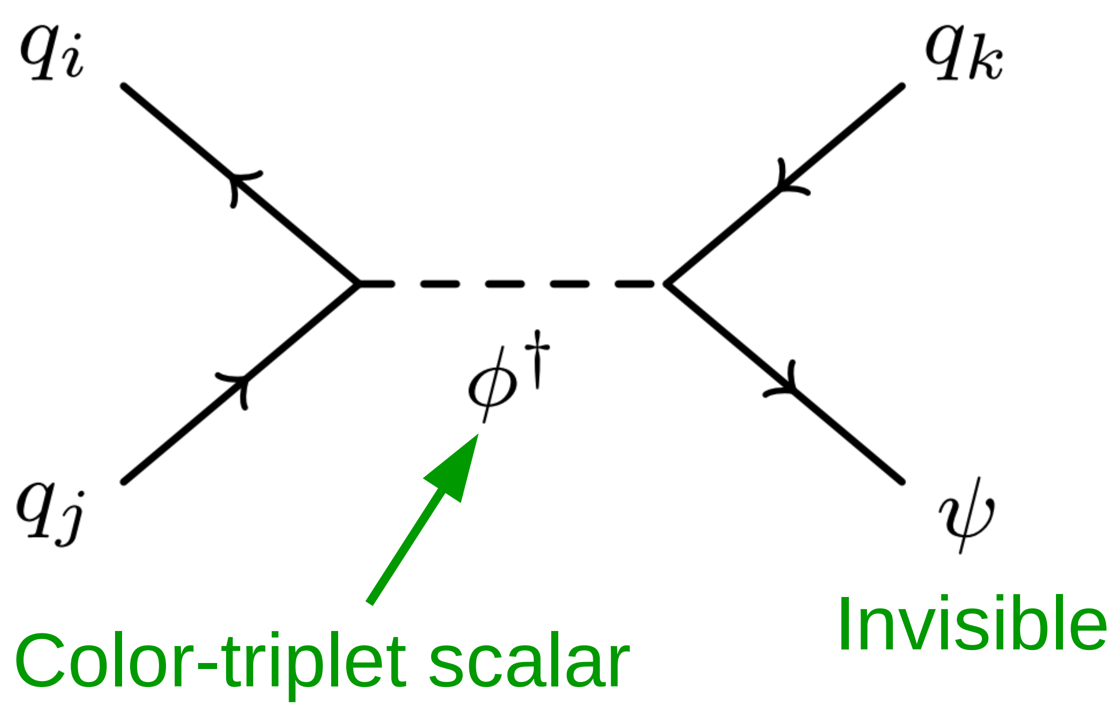
What's that feature? How significant is it?

Monojet excess counts

pT bin	Observed	Predicted	Pull
250-300	113837	113069±1889.5	
300-350	65430	61768.4±1040.4	3.4σ
350-400	33571	31905.8±553.6	2.9σ
400-450	17720	17166.4±294.5	1.7σ
450-500	9726	9679.8±163.9	

(from HEPDATA database)

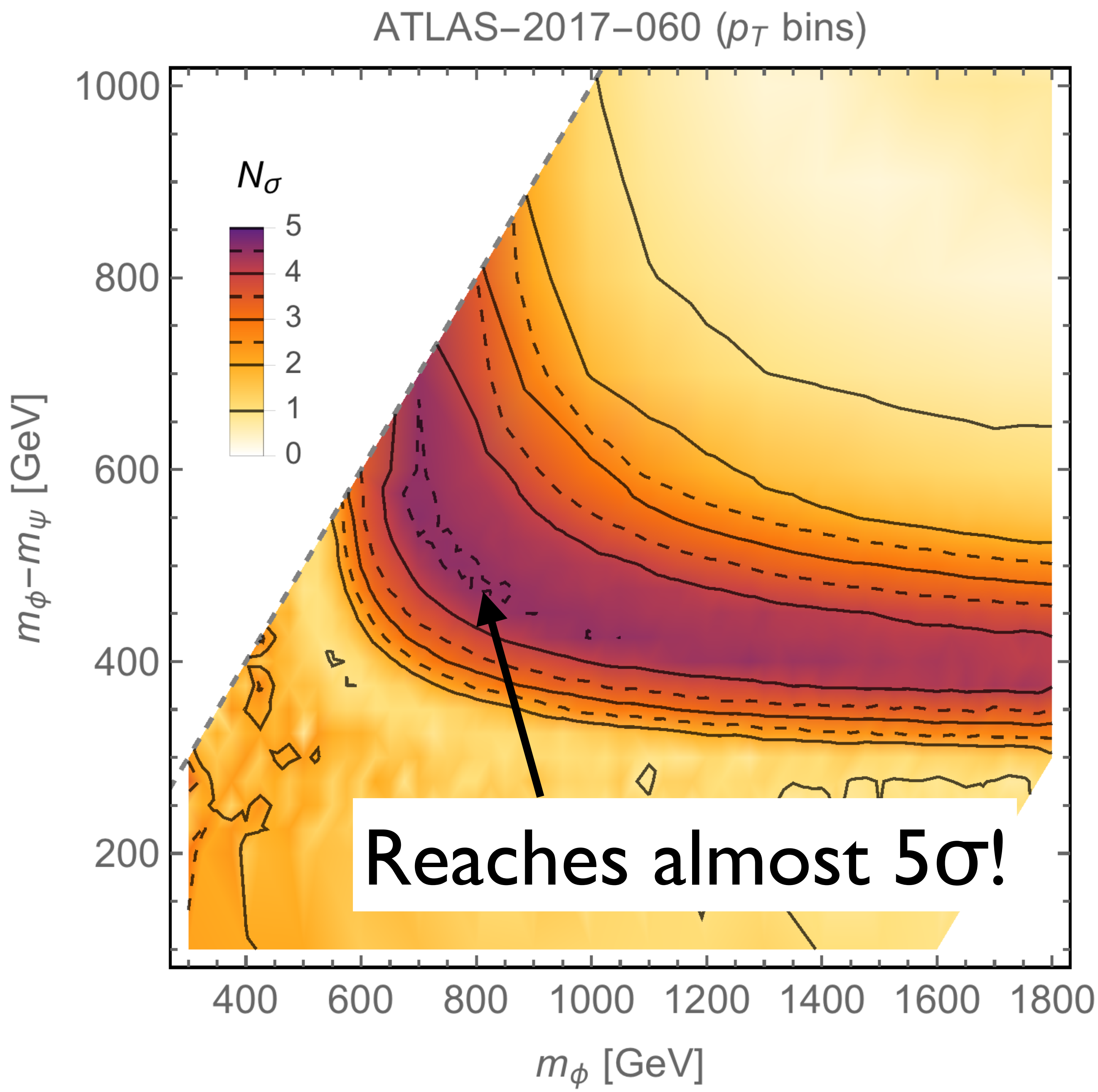
Using the numbers that ATLAS provided on HEPDATA, we can fit the excess to a model:



the “mono-phi model”

Turns out this is the UV completion for the “hylogenesis” model of asymmetric dark matter

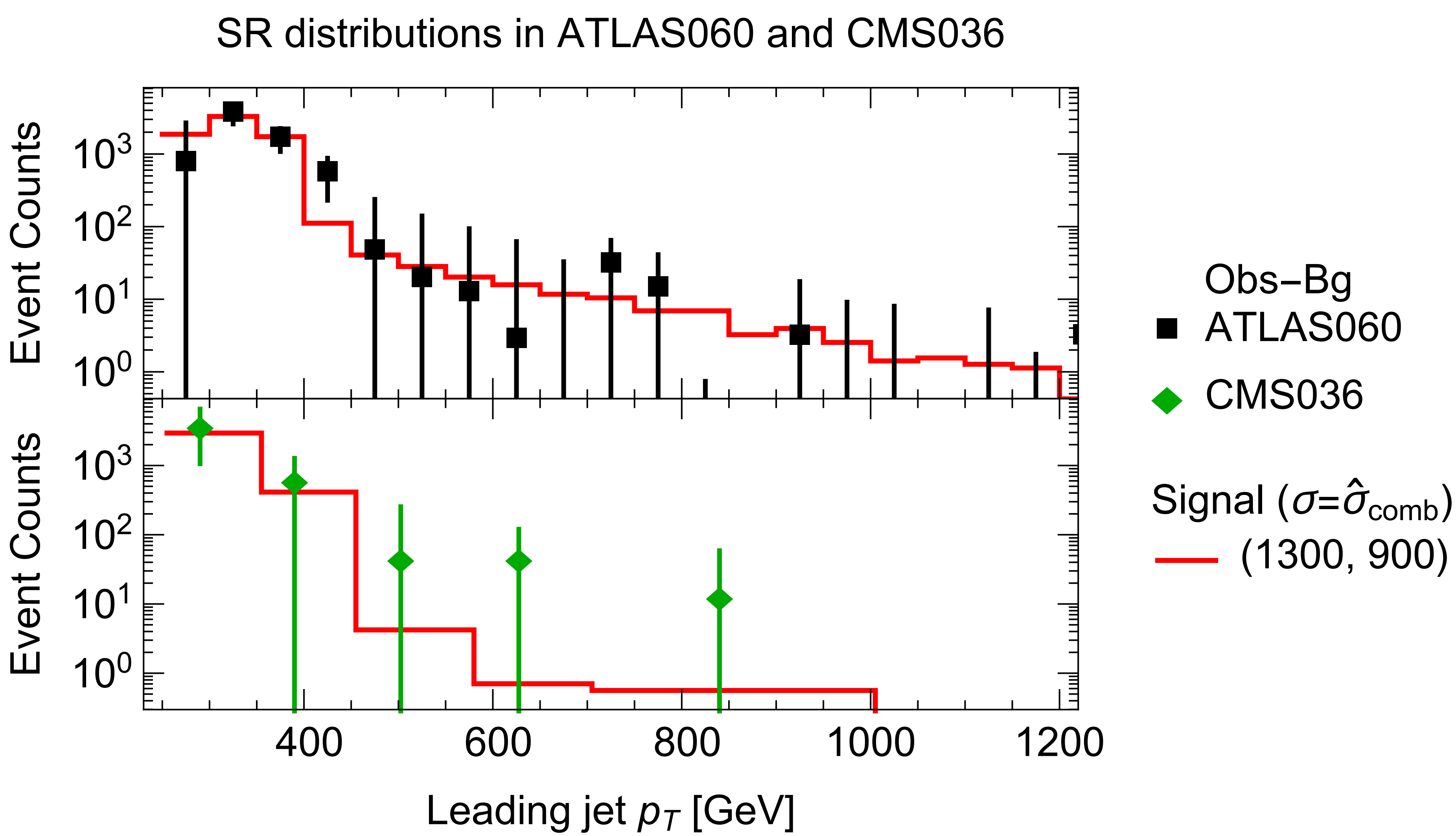
Davoudiasl, Morrissey, Sigurdson & Tulin 1008.2399, 1106.4320



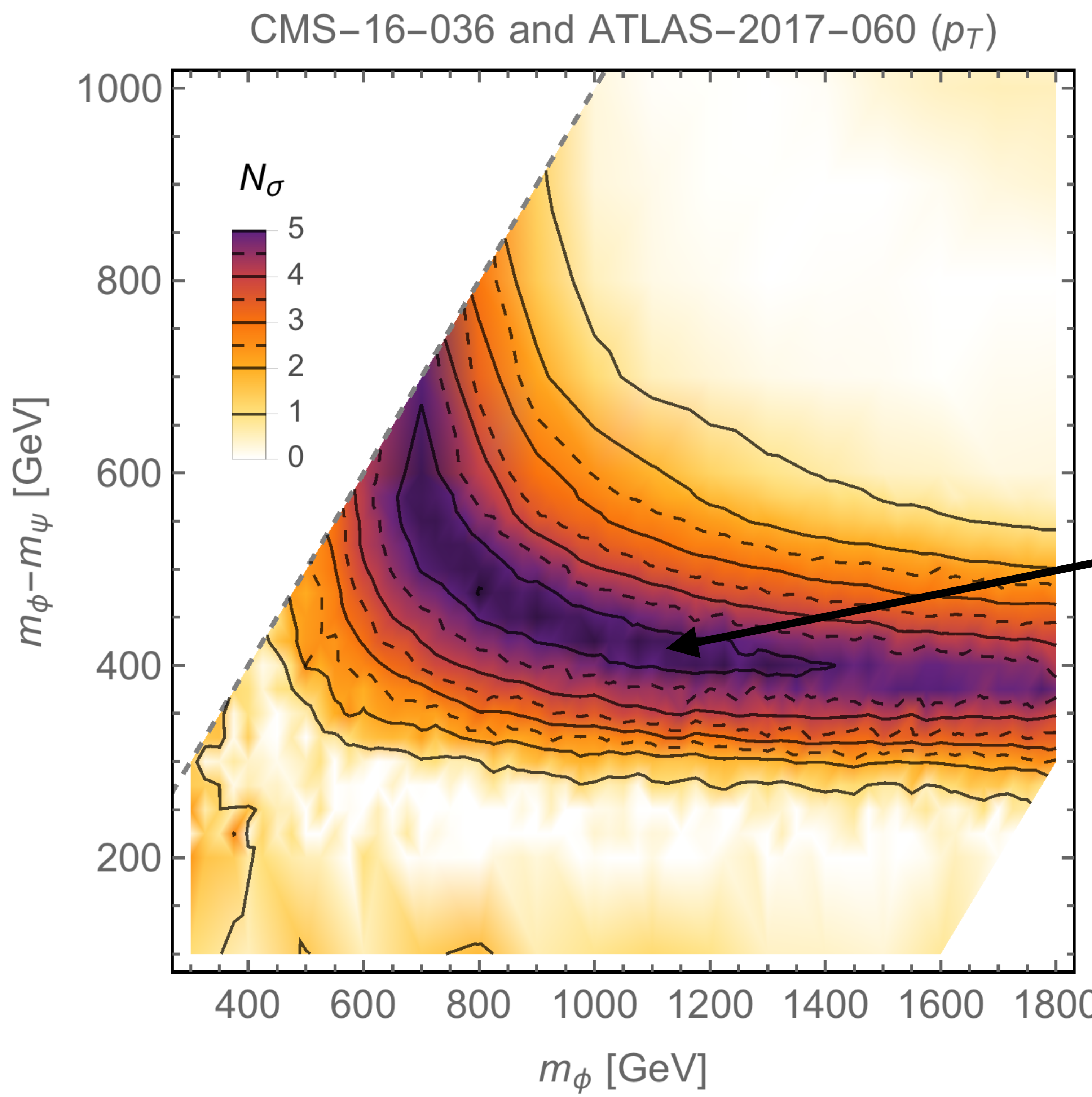
(Checked: best fit signal strength is compatible with MET distribution. Ideally would do a fit to joint MET-jet pT distribution)

Compatibility with other searches

Similar feature seen in the monojet bins of CMS jets+MT2 search (CMS036)!

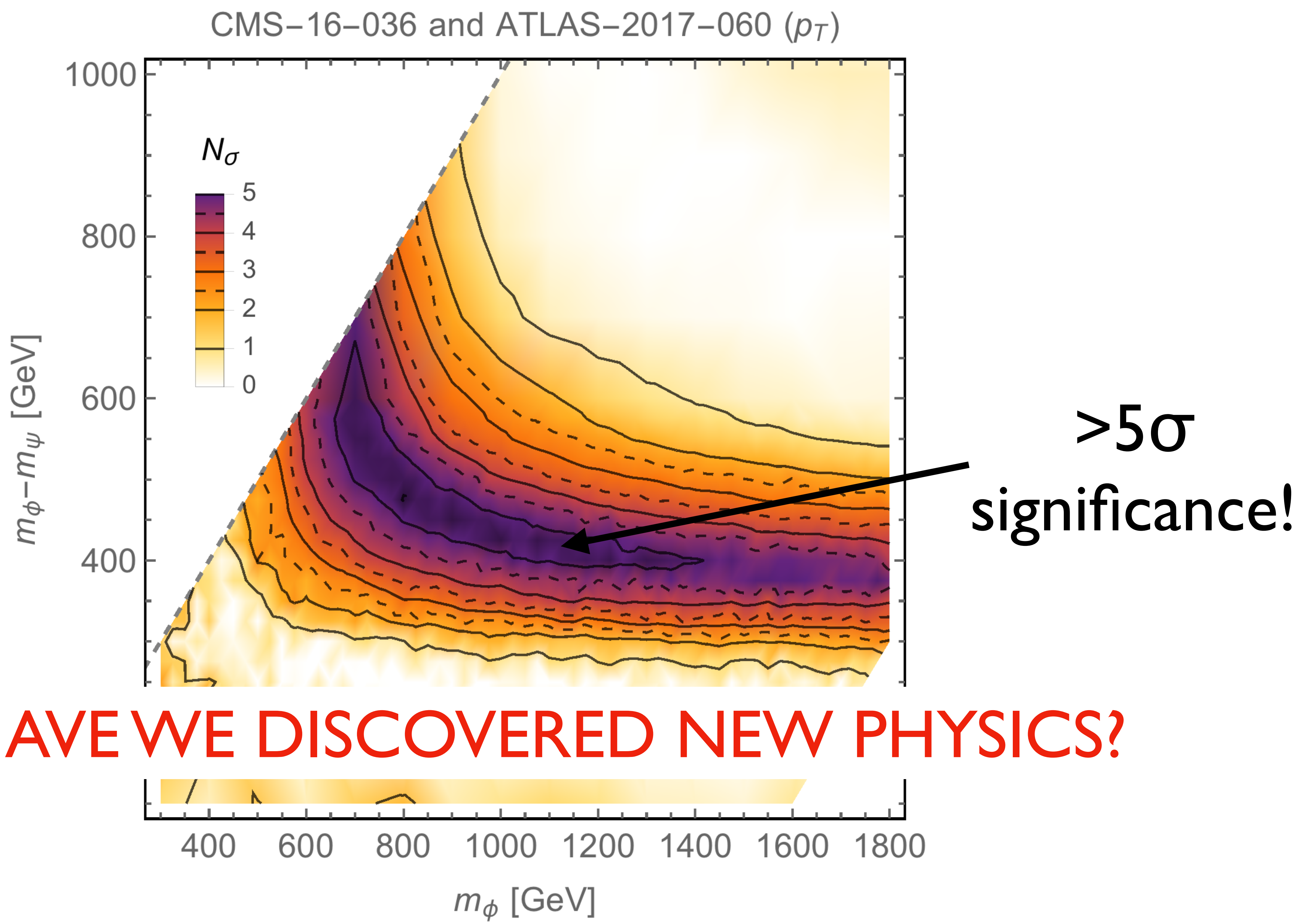


Combined fit to ATLAS monojets and CMS jets+MT2 results in over 5σ local significance!!



>5σ significance!

Combined fit to ATLAS monojets and CMS jets+MT2 results in over 5σ local significance!!



HAVE WE DISCOVERED NEW PHYSICS?

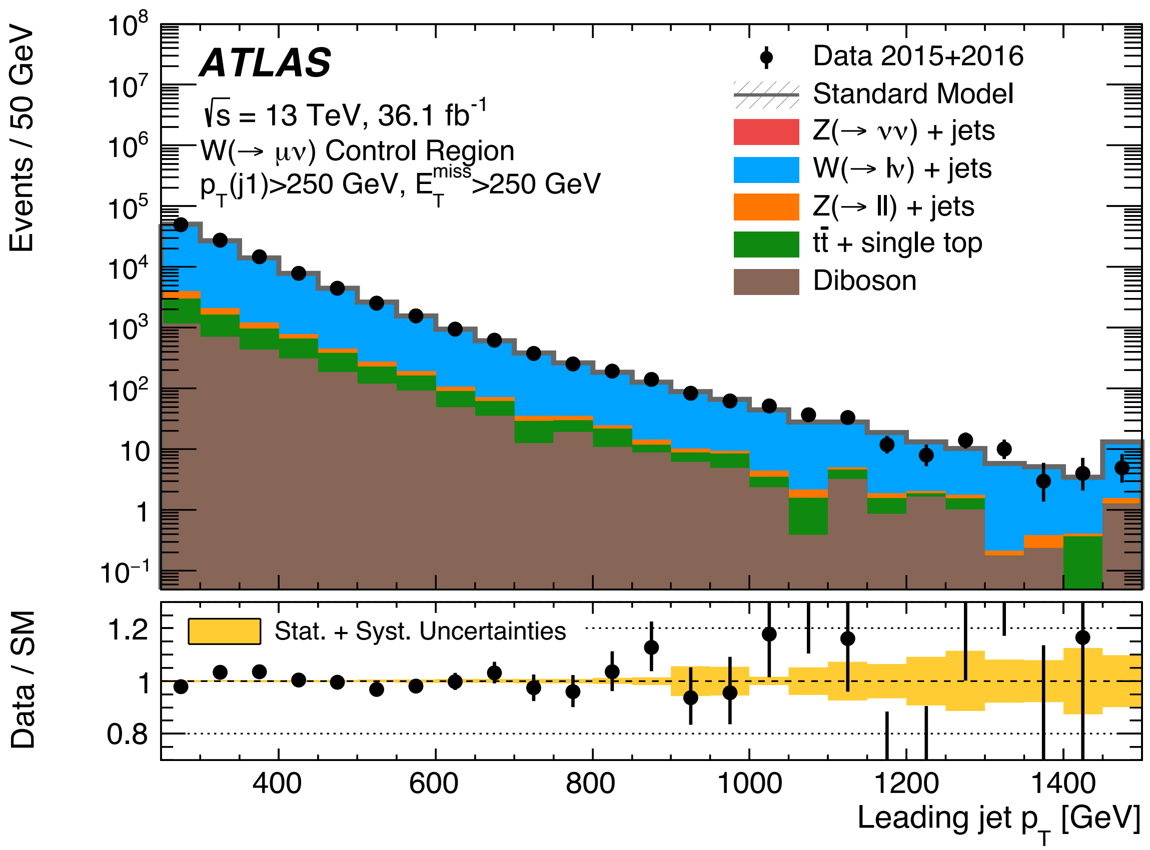
It certainly doesn't seem like a statistical fluctuation…

…but it could have a Standard Model explanation:

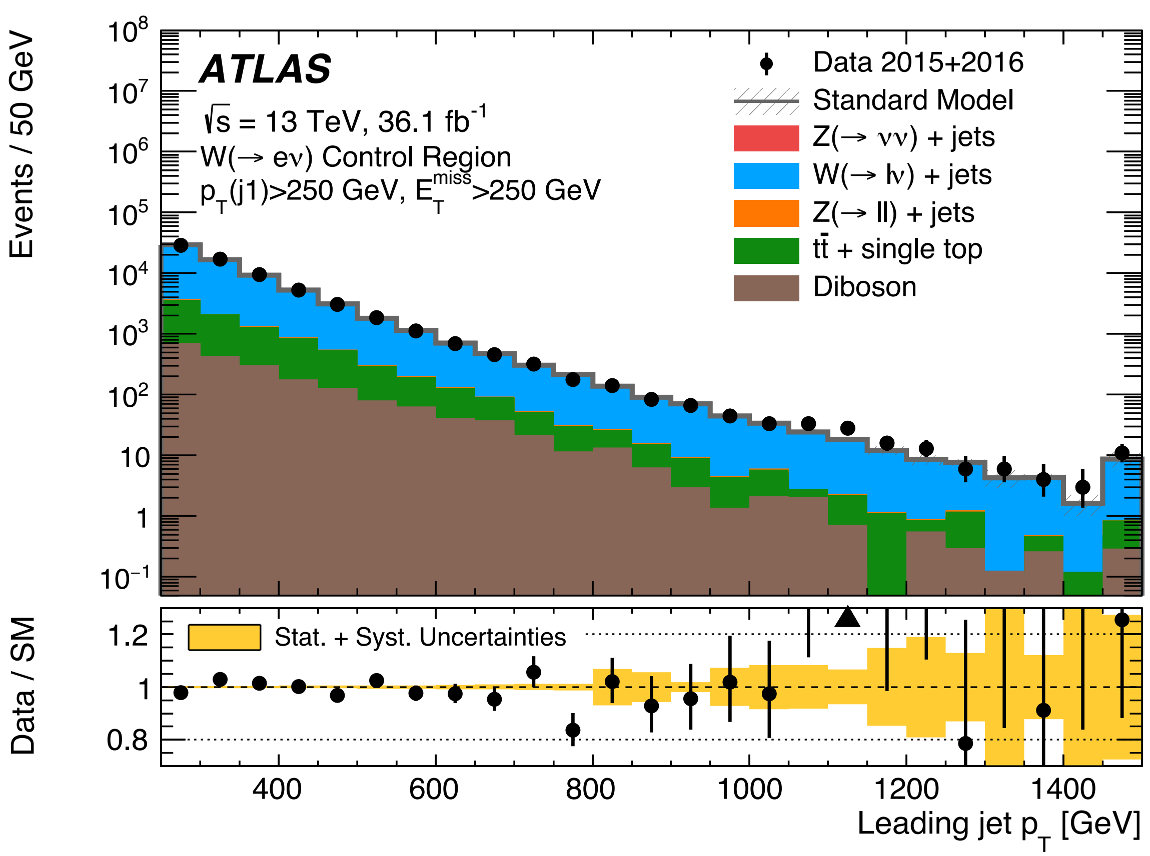
The excess is systematics dominated.

And a similar feature is seen in $W \to \ell\nu$ control regions.

(But a naive rescaling doesn't completely account for the excess in the signal region.)

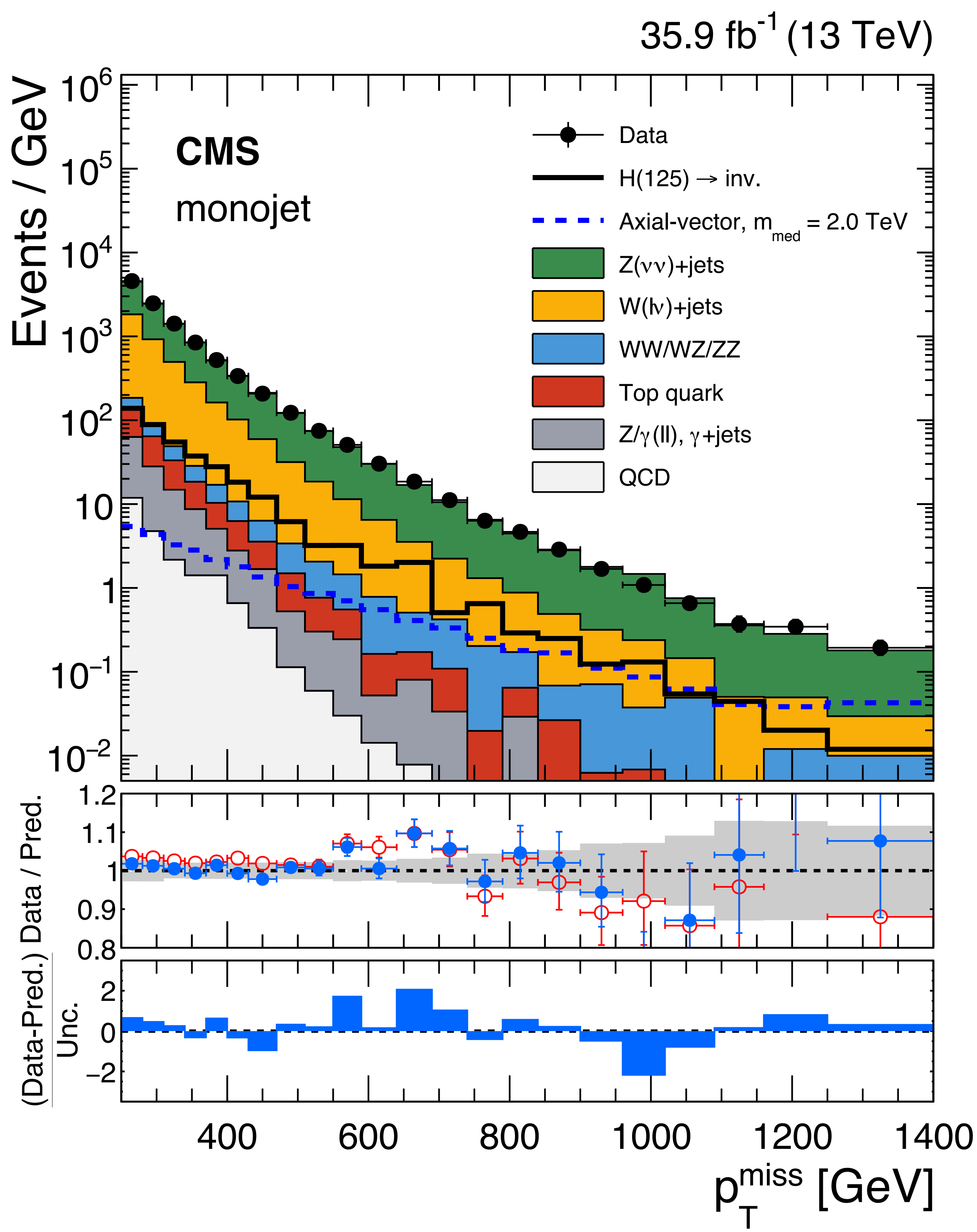


(b)



(d)

Also, nothing seen in CMS monojets.



Major differences in how CMS and ATLAS calculate their backgrounds.

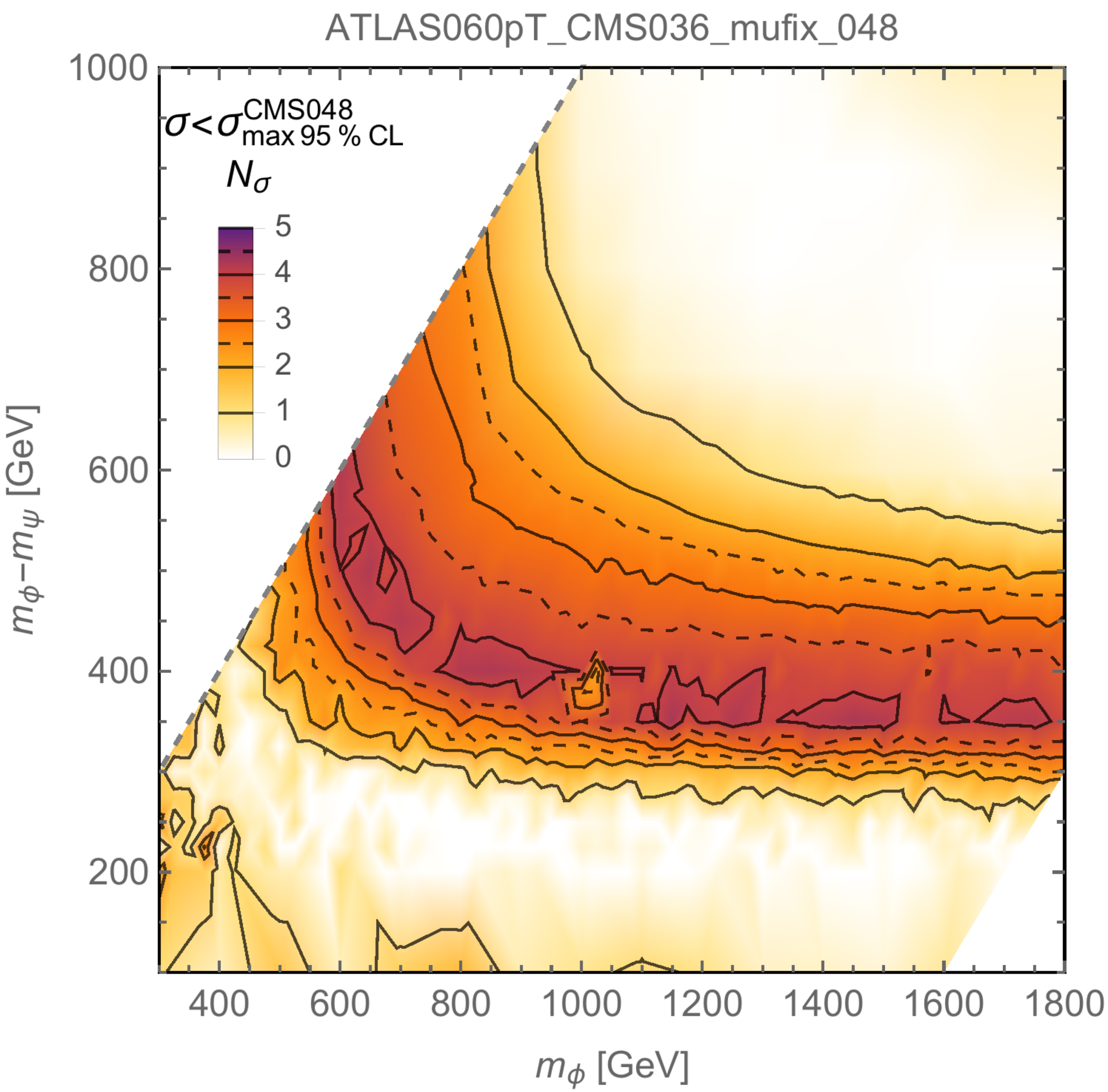
- CMS fits for the shape (beyond just nuisance parameters) as well as the normalization in the CRs
- ATLAS fits for only the normalization (plus nuisance parameters) in the CRs.

Who is doing it more correctly?

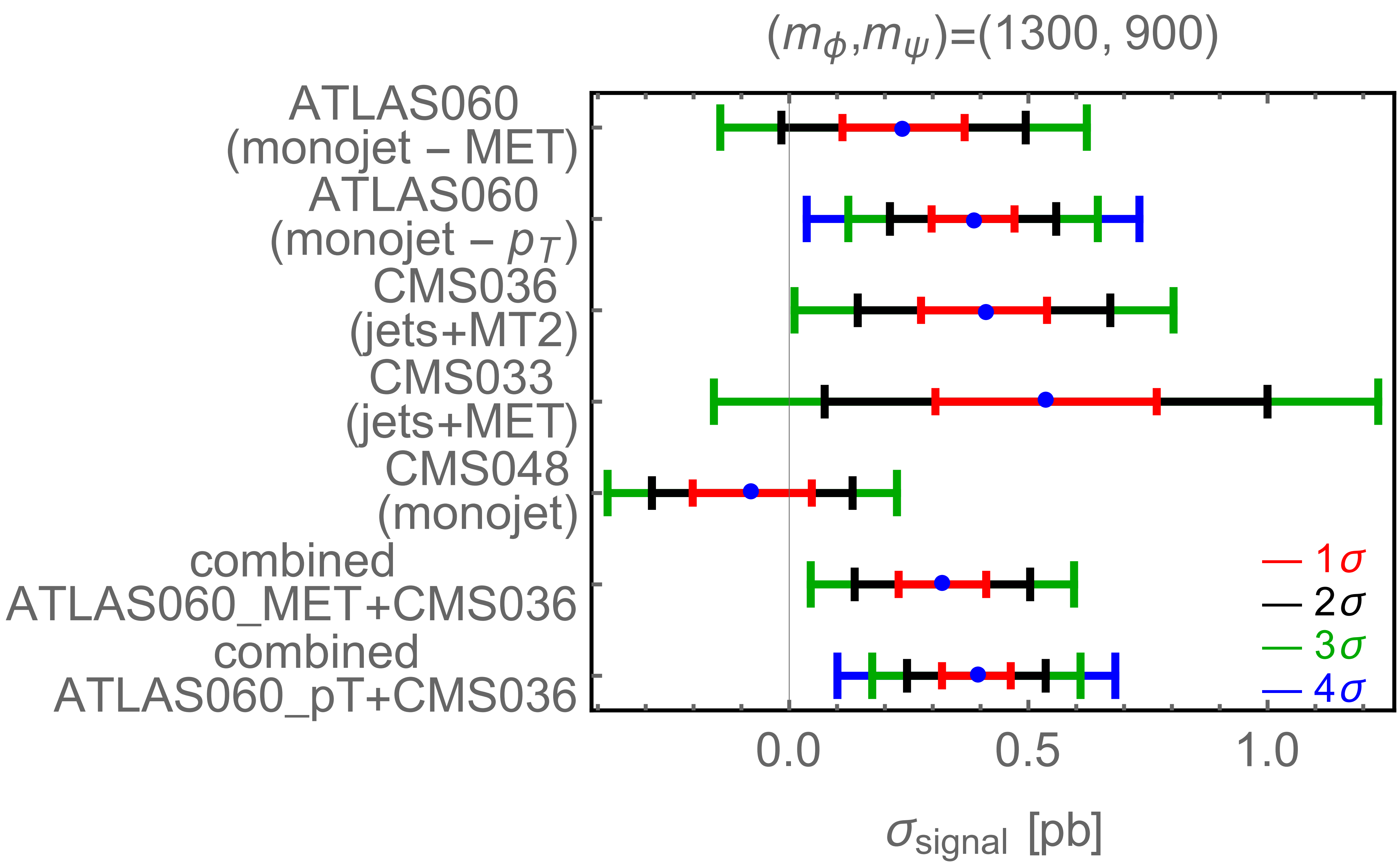
Could a BSM model populate both signal and control regions?

Best-fit consistent with CMS monojets 95% UL.

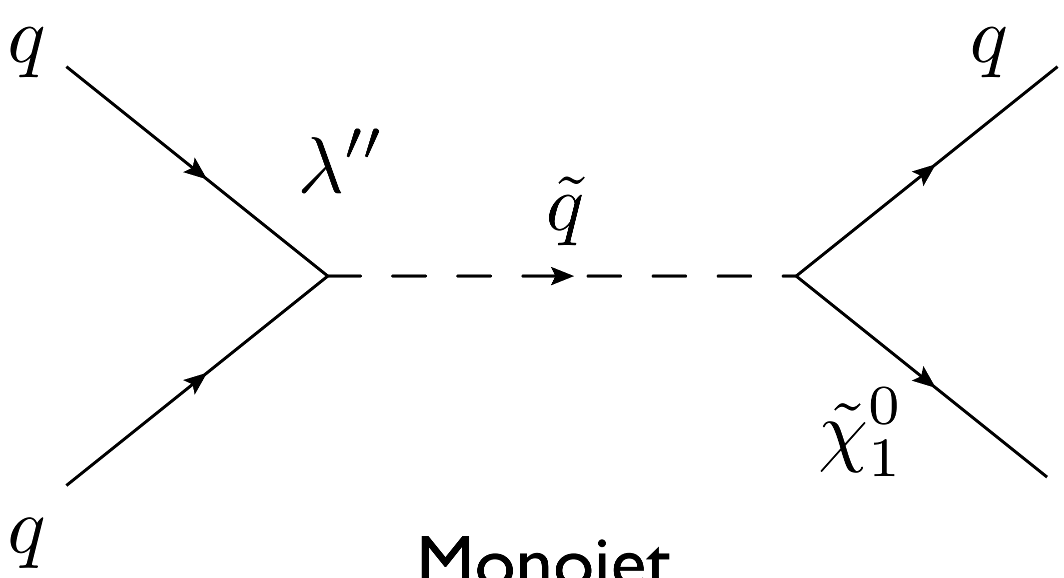
Still reaches 4.3σ local significance.



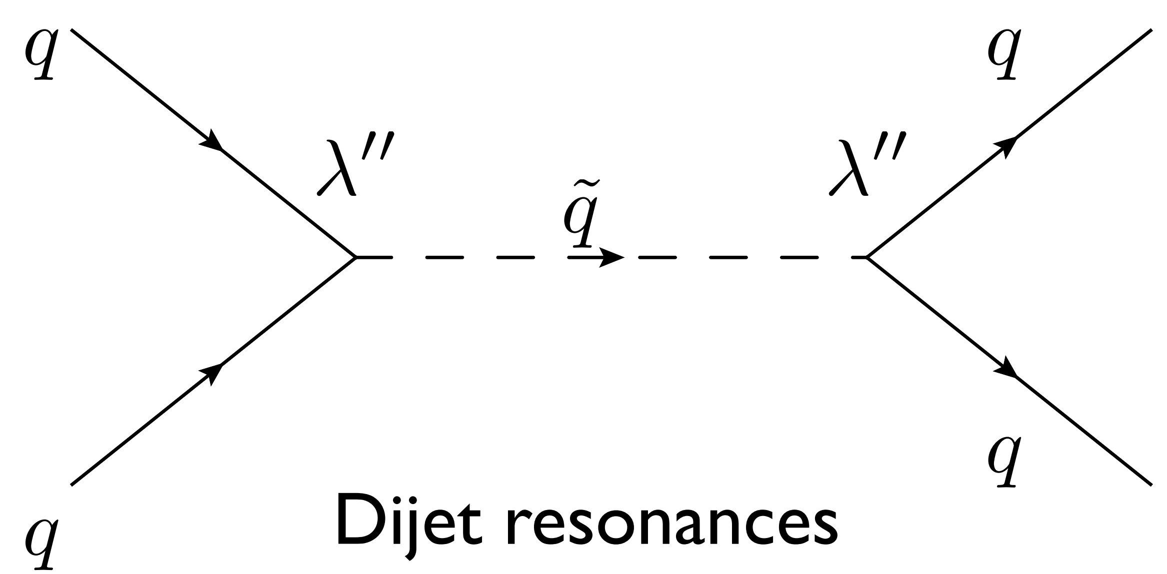
Summary of best-fit cross sections



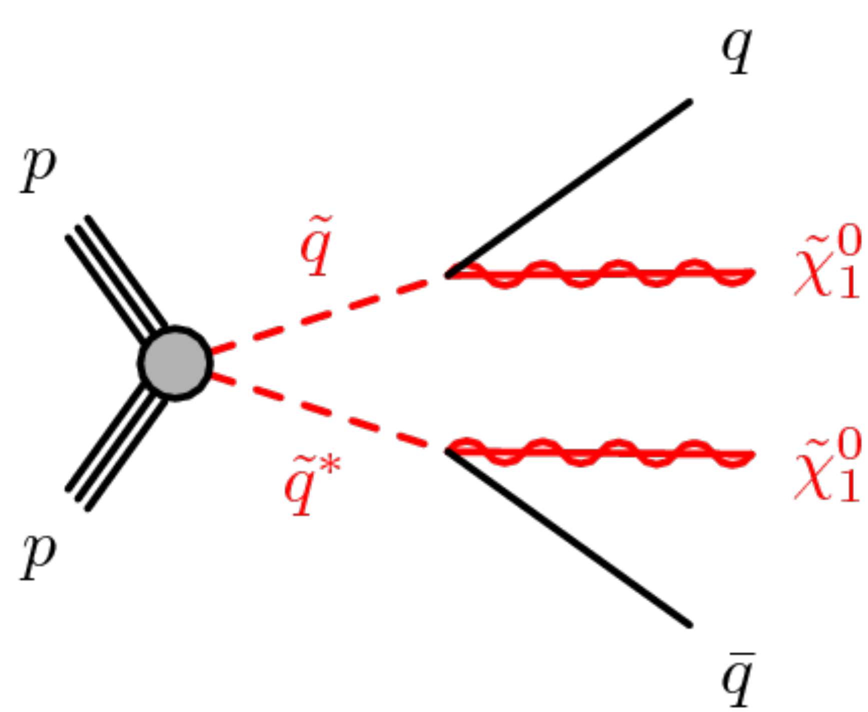
Correlated signatures



Monojet



Dijet resonances
model dependent



Pair production (2,3j+MET)
Irreducible, current limits 800-1000 GeV

Method of Rectangular Aggregations

Pouya Asadi, Matt Buckley, Anthony DiFranzo,
Angelo Monteux & DS (1707.05783)

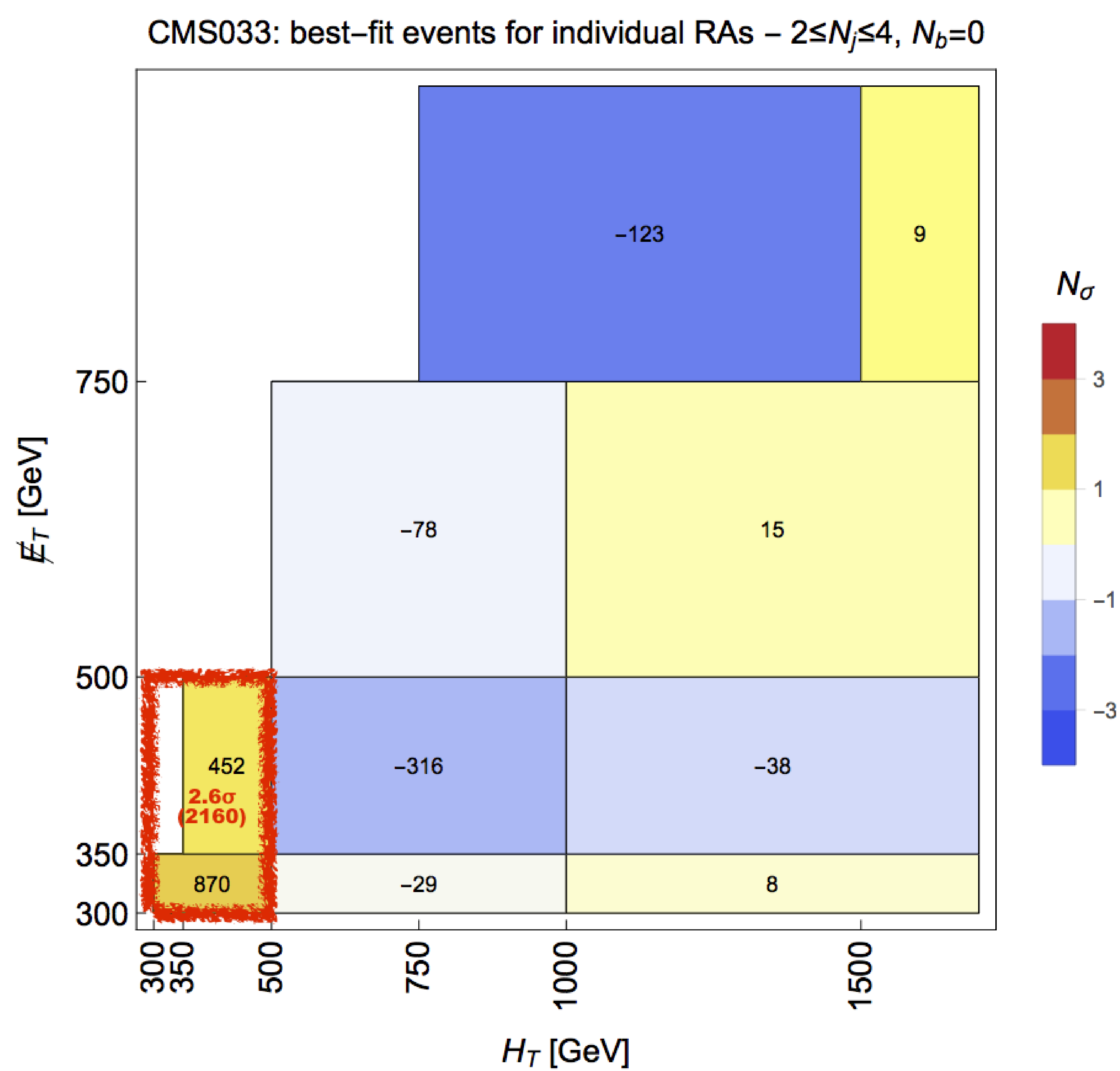
How did we find this ATLAS monojet excess?

It wasn't by chance. We developed a model-independent, data-driven way to probe the data for anomalies and excesses.

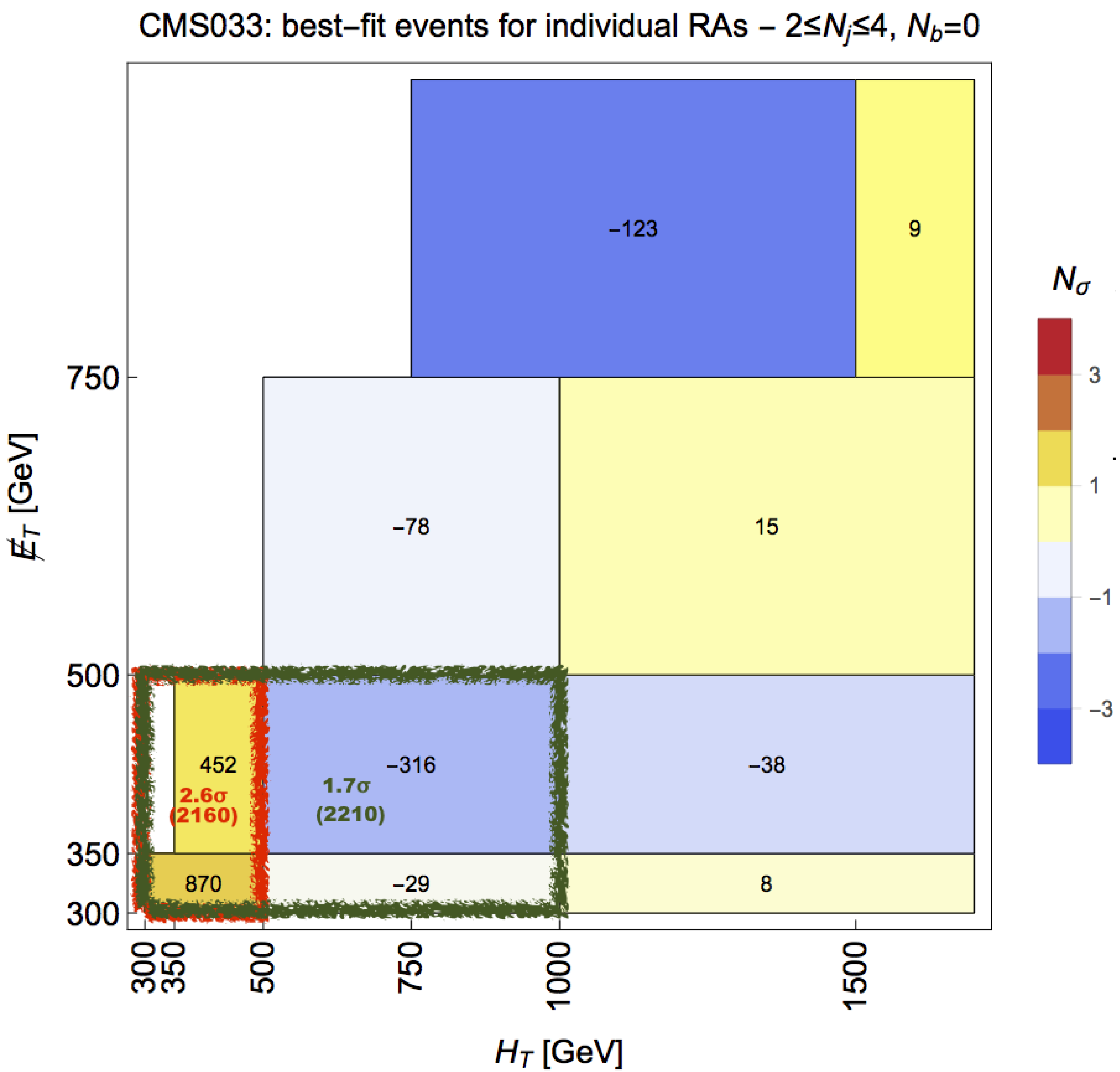
Idea: a true signal will usually populate multiple "neighboring" signal regions, while background fluctuations are more often confined to individual bins.

Consider all possible rectangles that aggregate together adjacent bins. Compute the likelihood of observing a deviation as large (or small) as observed in the data, assuming New Physics only contributes to that rectangular aggregation.

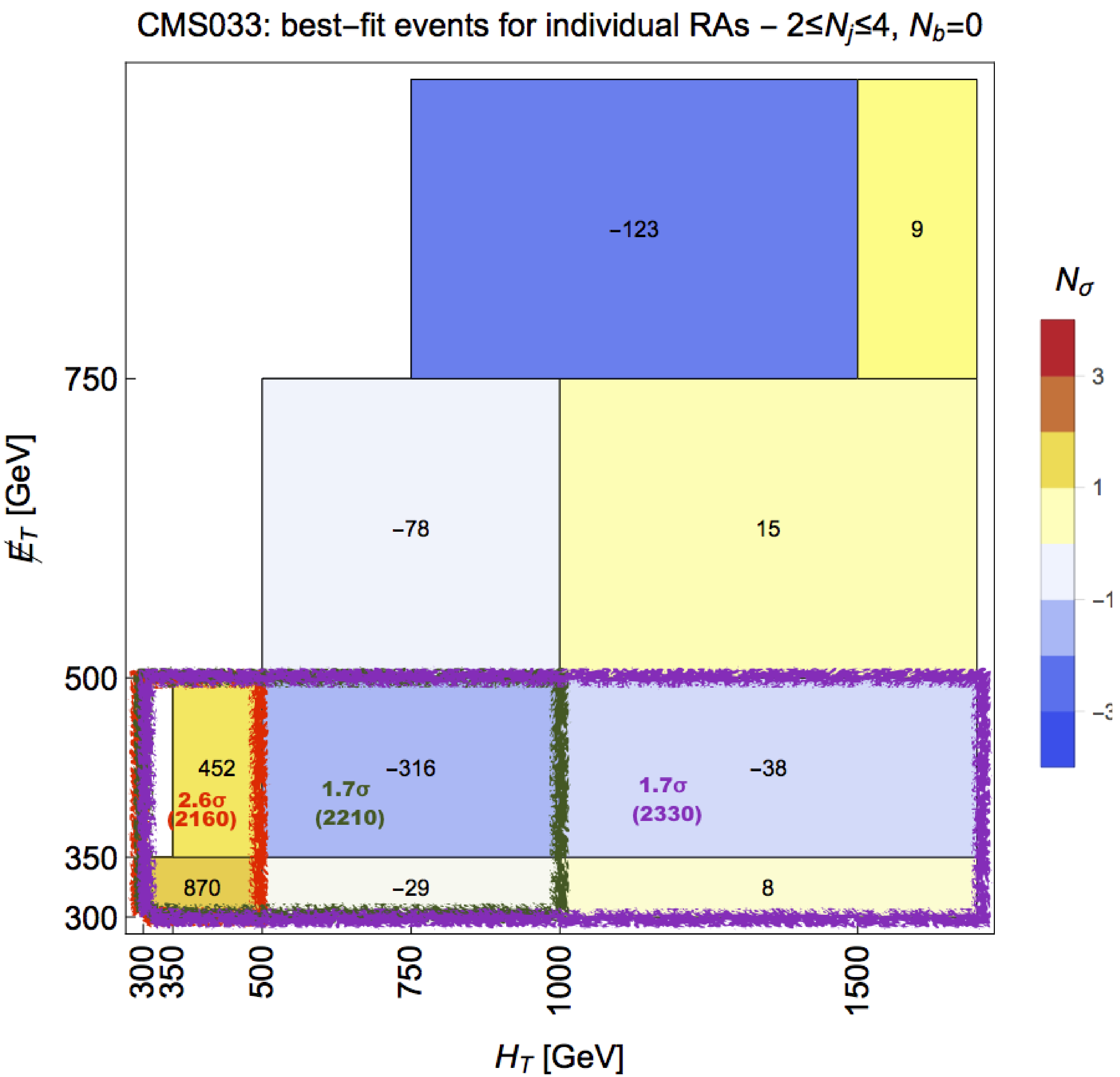
Example: rectangular aggregations in CMS033



Example: rectangular aggregations in CMS033



Example: rectangular aggregations in CMS033



We applied this technique to two "big" CMS SUSY searches:

- CMS-PAS-SUS-16-033 (CMS033): jets+MET :
  174 signal regions binned in four variables:
  $N_j \geq 2$, $N_b \geq 0$, MET$\geq$300GeV, $H_T \geq$300GeV

→ 7,000 aggregations

- CMS-PAS-SUS-16-036 (CMS036): jets+MT2
  213 signal regions binned in four variables:
  $N_j \geq 1$, $N_b \geq 0$, $M_{T2} \geq$ 200 GeV, $H_T \geq$ 250 GeV

→ 33,000 aggregations

2 viable excesses with $\gtrsim 3\sigma$ local significance in each search

Interestingly, one viable excess is shared between the two searches:

CMS036:

ROI		N_j	N_b	H_T (GeV)	M_{T2} (GeV)	N_σ	compatible?
2	b	$1-3$	0	$250-450$	$200-300$	2.95	✓
	d	$1-3$	0	$250-450^*$	$200-300$	2.74	✓

CMS033:

ROI		N_j	N_b	H_T (GeV)	$\not{E}_T$ (GeV)	N_σ	compatible?
2	a	$2-6$	0	300^*-500	$300-500$	2.96	✓
	c	$2-4$	0	300^*-500	$300-500$	2.64	✓
	d	$3-6$	0	300^*-500	$300-500$	2.57	✓

(Same data: we cannot statistically combine the significances.)

This a low N_j , low H_T , low MET region of parameter space.

the monojet excess!

Comments on LEE

Of course, with so many rectangular aggregations, there will be a big look elsewhere effect.

We view the main utility of the method as identifying "hot spots" in the data where potentially interesting anomalies could be hiding. There is a lot more data coming, and these hot spots are worth keeping an eye on. (Make sure the SRs are frozen if at all possible!) Having identified the hot spots, we no longer need to worry about the LEE in future data.

Also, keep in mind that the discovery threshold was always meant to be 5σ *local significance*.

Conclusions

There is something interesting going on in the low-pT region of the monojet search. Given the 4-5σ significance of the mono-phi model fit, it is unlikely to be a statistical fluctuation. It must be either SM mismodeling or new physics. We need more scrutiny from experts to help figure out what's going on!

We uncovered this excess using a model-independent, data-driven method for mining the signal regions of existing LHC searches for signal-like anomalies.

The method can have many more applications!

- There are several other excesses to explore in CMS jets+MET, and many more in the other searches for new physics
- The strategy can be extended in different directions (e.g. non- rectangular aggregations, signal templates).

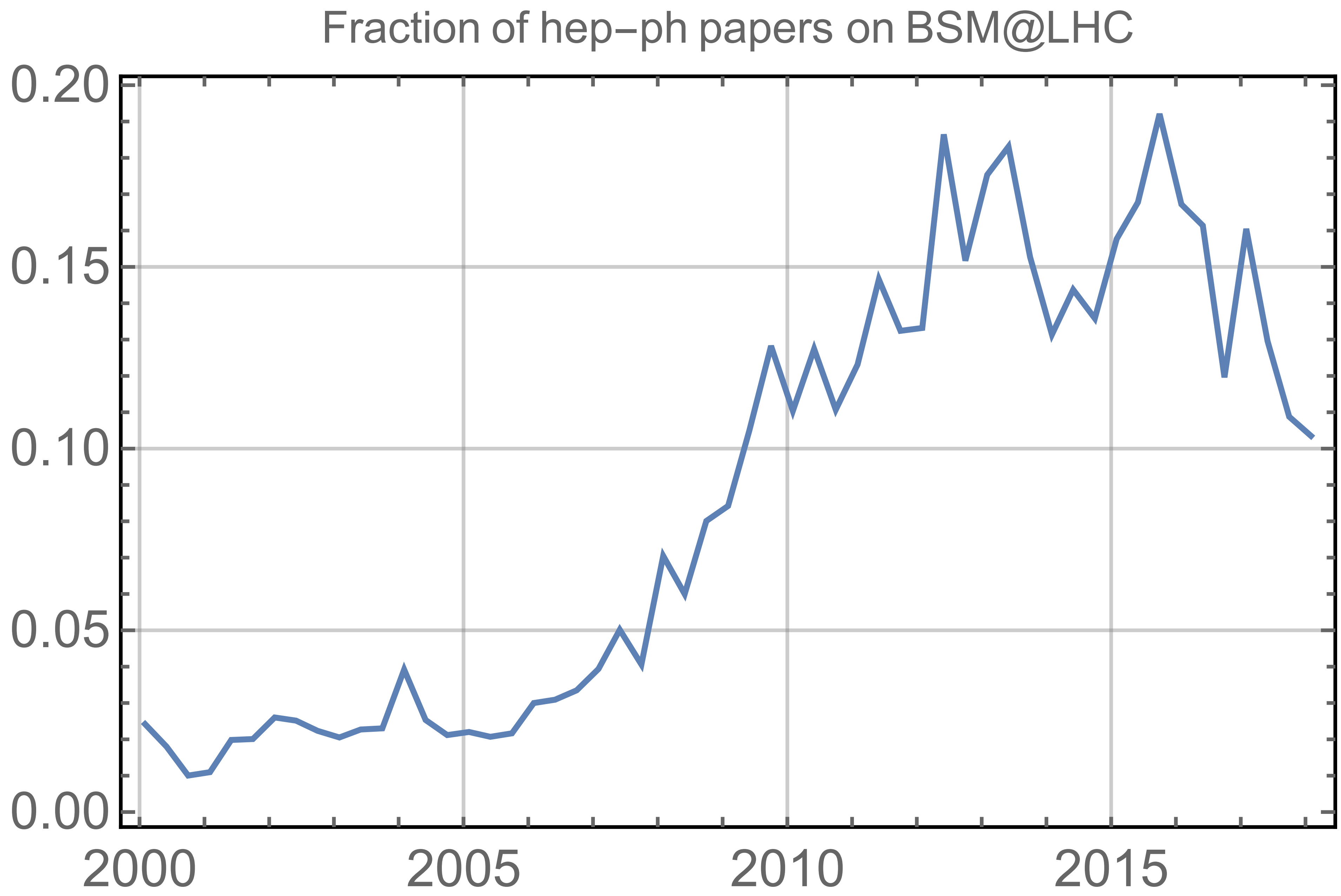
Conclusions

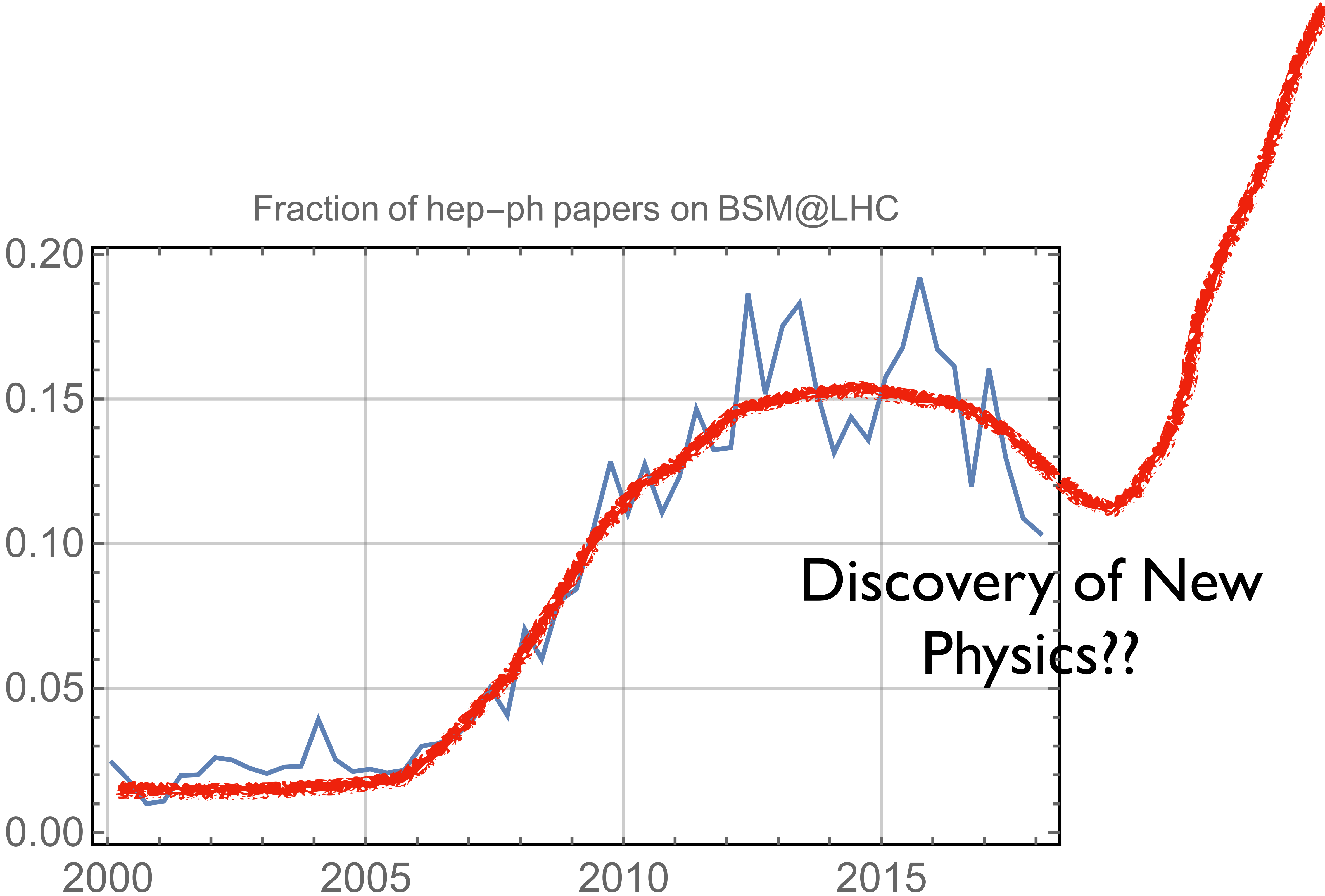
In the past, a "wait and see" approach made a lot of sense, but as the data comes in more slowly, it becomes increasingly motivated to sift the fluctuations for anomalies and attempt to fit them models. This could have many benefits, e.g.

- Reveal patterns of correlated fluctuations
- Provide a new target for search re-optimization
- Suggest new final states to search in
- Maybe one of the excesses will turn out to be real!

The official propaganda plots based on a handful of simplified models just don't do justice to the interesting things going on in the data.

Don't believe the official LHC propaganda!
There could still be new physics hiding in the data!


Fraction of hep-ph papers on BSM@LHC
0.20
0.15
0.10
0.05
0.00
2000
2005
2010
2015



Fraction of hep–ph papers on BSM@LHC
0.20
0.15
0.10
0.05
0.00
2000
2005
2010
2015
Discovery of New Physics??

Thanks for your attention!

Backup material

Statistics

We use the standard LHC profile likelihood approach:

$$\mathcal{L}(\mu, \theta) = \prod_i \frac{(\mu s_i + b_i + \theta_i)^{n_i} e^{-(\mu s_i + b_i + \theta_i)}}{n_i!} \exp\left(-\frac{1}{2}\theta^T V^{-1}\theta\right)$$

[CMS-NOTE-2017-001]

- n_i is the number of observed events in each bin.
- s_i is the number of BSM signal events, for a reference xsec
- μ is a cross section multiplier.
- b_i is the expected background count in the bin (extrapolated from control regions).
- θ_i is a nuisance parameter for the background b_i, and its variation is modulated by the covariance matrix V.

Maximizing the likelihood we get:

- local maximum for given μ: $\mathcal{L}(\mu, \hat{\theta}_\mu)$
- global maximum: $\mathcal{L}(\hat{\mu}, \hat{\hat{\theta}})$

[Cowan, Cranmer, Gross, Vitells, 1007.1727]

- significance:

$$q_0 \equiv \begin{cases} -2\ln\frac{\mathcal{L}(0,\hat{\theta}_0)}{\mathcal{L}(\hat{\mu},\hat{\hat{\theta}})} & \hat{\mu} \geqslant 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

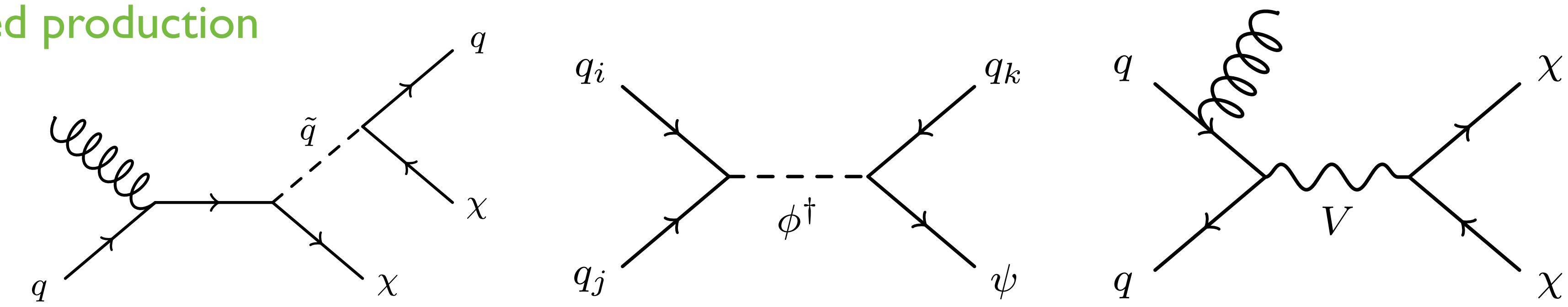
χ^2-distributed with 1 dof in the large N limit, $N_\sigma = \sqrt{q0}$.

Monojet excess: possible explanations

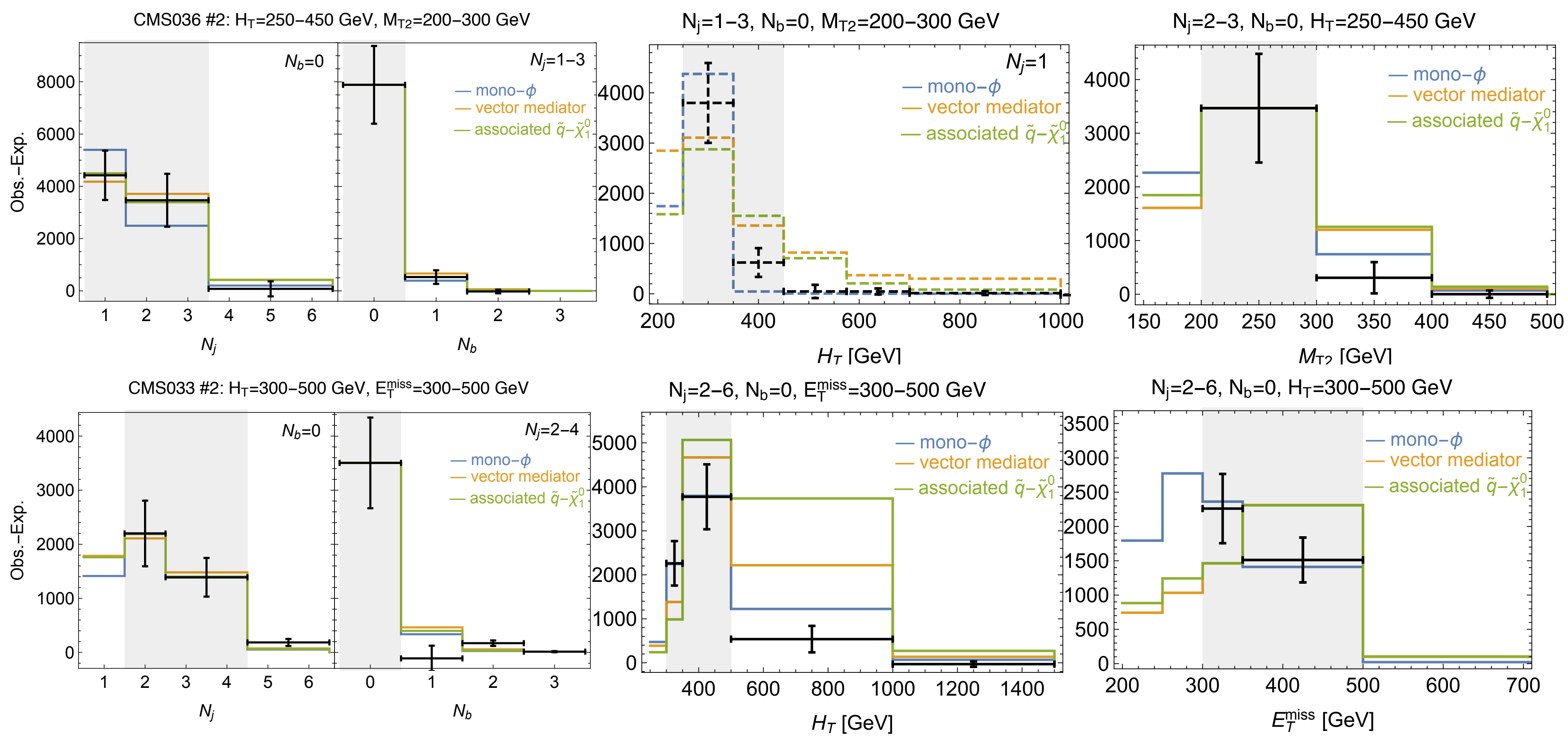
squark-neutralino associated production

“mono-squark”

vector-mediator DM



(m_{heavy}=1.2 TeV, $m_{invisible}$=800 GeV)

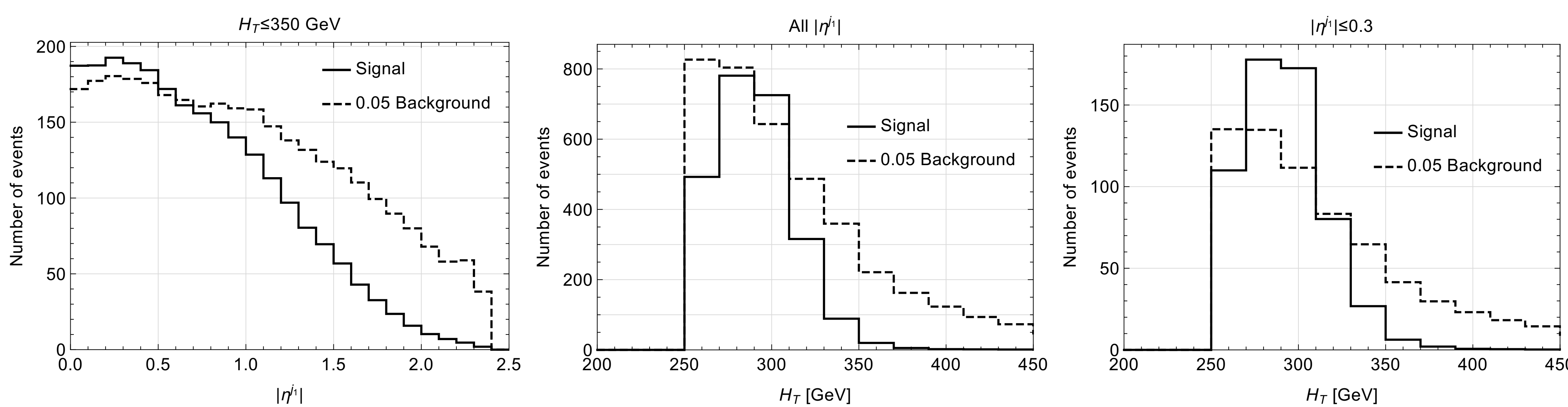


Only the mono-squark model is peaked enough in MT2 and HT to fit the excess!

Improving S/B

Search	N_j	N_b	H_T	$M_{T2}, \not{E}_T$	Obs.	Bg. (pre-fit)	Bg. (post-fit)	Best-fit Signal
CMS036	$1-3$	0	$250-450$	$200-300$	145144	137256 ± 8159	140391 ± 1524	4753
CMS033	$2-4$	0	$300-500$	$300-500$	58138	54550 ± 2246	55976 ± 780	2162
ATLAS060	≥ 1	-	≥ 250	$350-700$	74686	72645 ± 1140		2041

The significance of the excess is dominated by systematic errors. So adding more data might not improve the situation.



Can try to purify the signal over background with tighter cuts on eta and HT. Could gain a factor of ~ 1.5 in S/B this way.

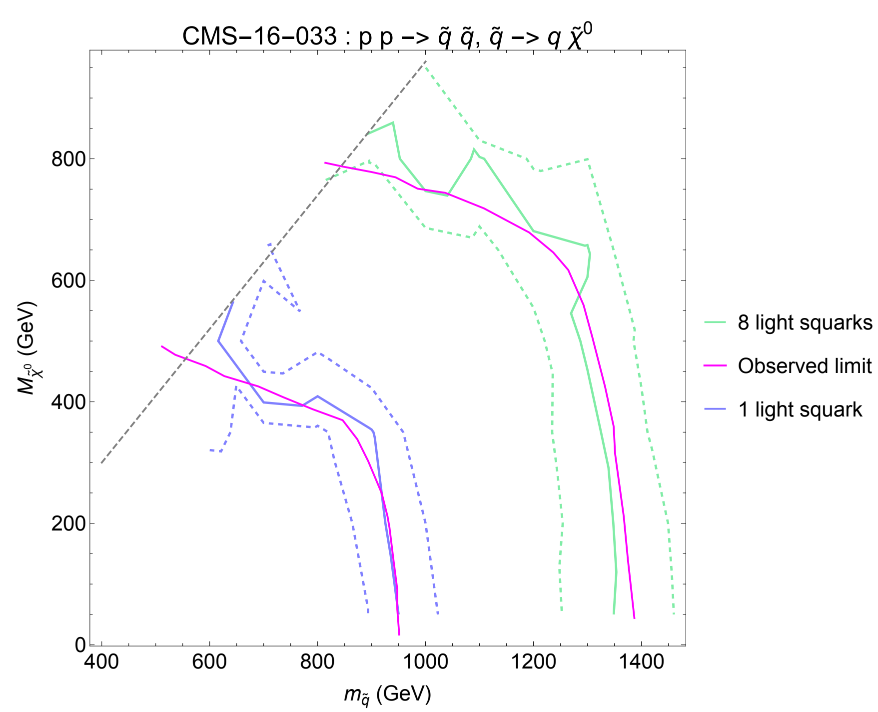
Our recasting pipeline

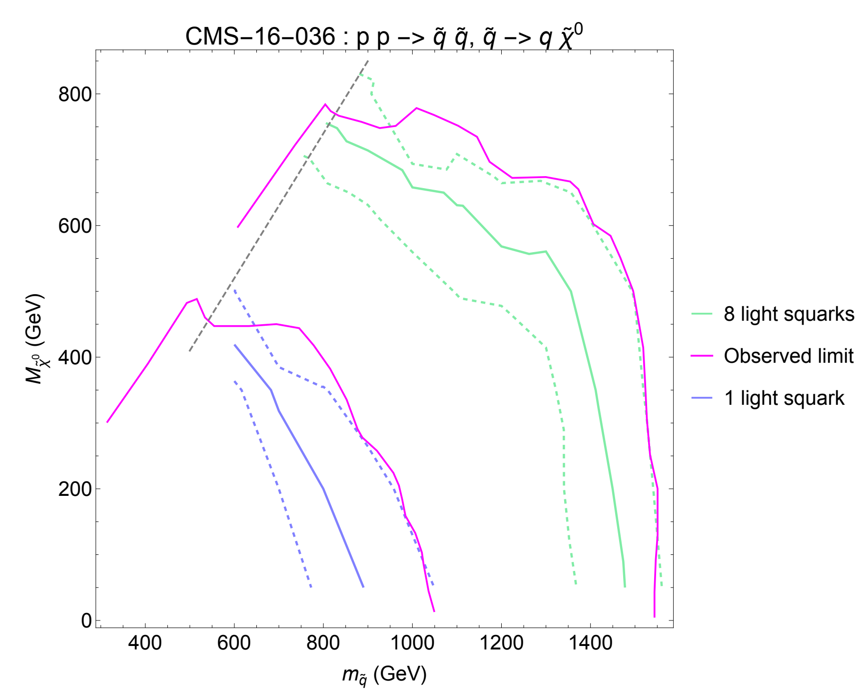
Buckley, Feld, Macaluso, Monteux & DS 1610.08059

Madgraph5; Prospino 2.1 — **Hard process (signal only)** → Pythia8.2 — **Showering, hadronization** → Delphes3 — **Detector simulation**

↓

Limit plots ← **Signal efficiencies** ← **Recasted ATLAS/ CMS analysis**





MET residuals