Monte Carlo Generators in ATLAS

Zach Marshall (LBL)

Berkeley Searches Workshop

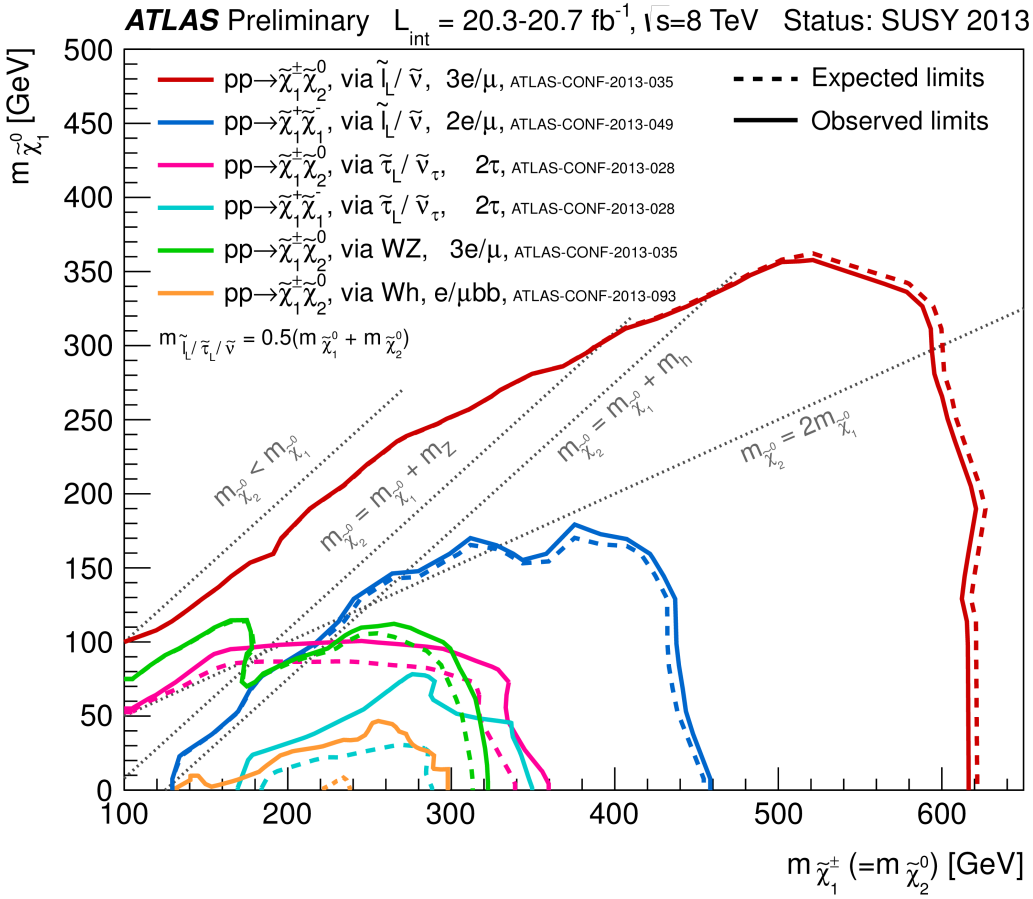
27 Jan 2014

This Talk's Disclaimer

- I'm not going to talk about soft physics
  - Even though we tune to minbias and assume that similar modeling for the UE in super high-pT events
  - Even though it's not clear that hadronization models are accurate for both LEP and the LHC
- I'm not really going to talk about the Higgs
  - Not my area – besides, it's a background now
- I'm not going to talk about heavy ion physics
- I'm assuming you're all good at reading (skipping some basics)
- I'm going to tell you an extra bit about how we do things
  - Some of this is beyond what is "normally" in a paper
  - My view is that you could get this information by asking the authors, and that given infinite time you would do that, so I'm not revealing any sensitive information by telling you how we operate

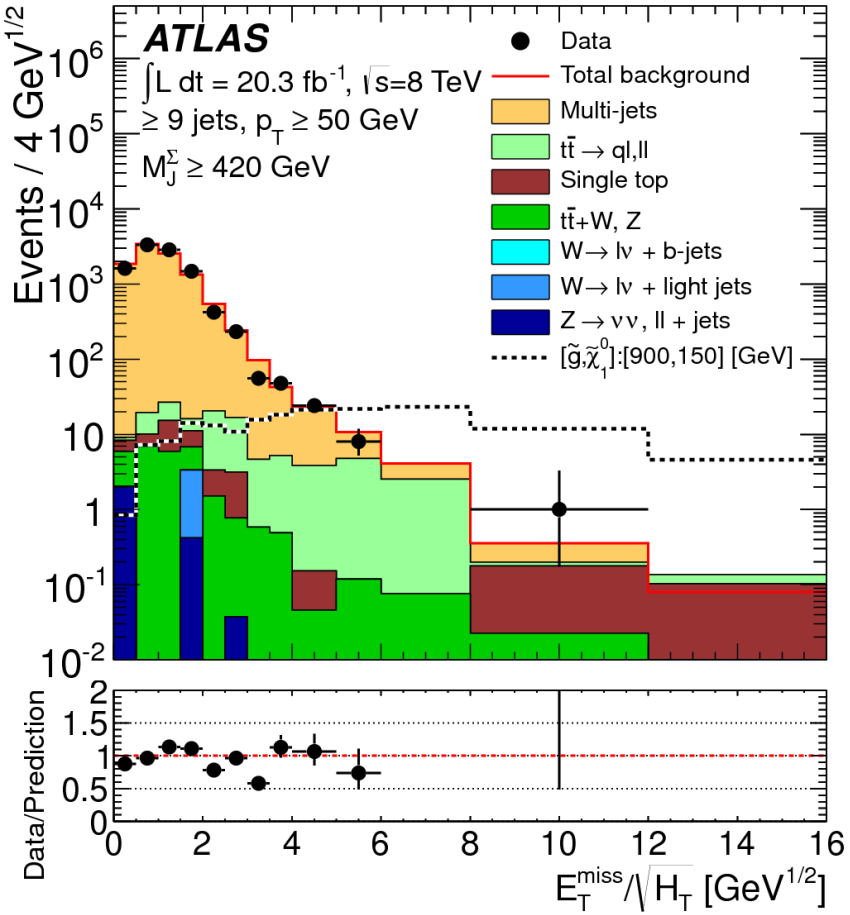
My Goal

Get you to stop worrying about the details of this



My Goal

…and start worrying about the details of this

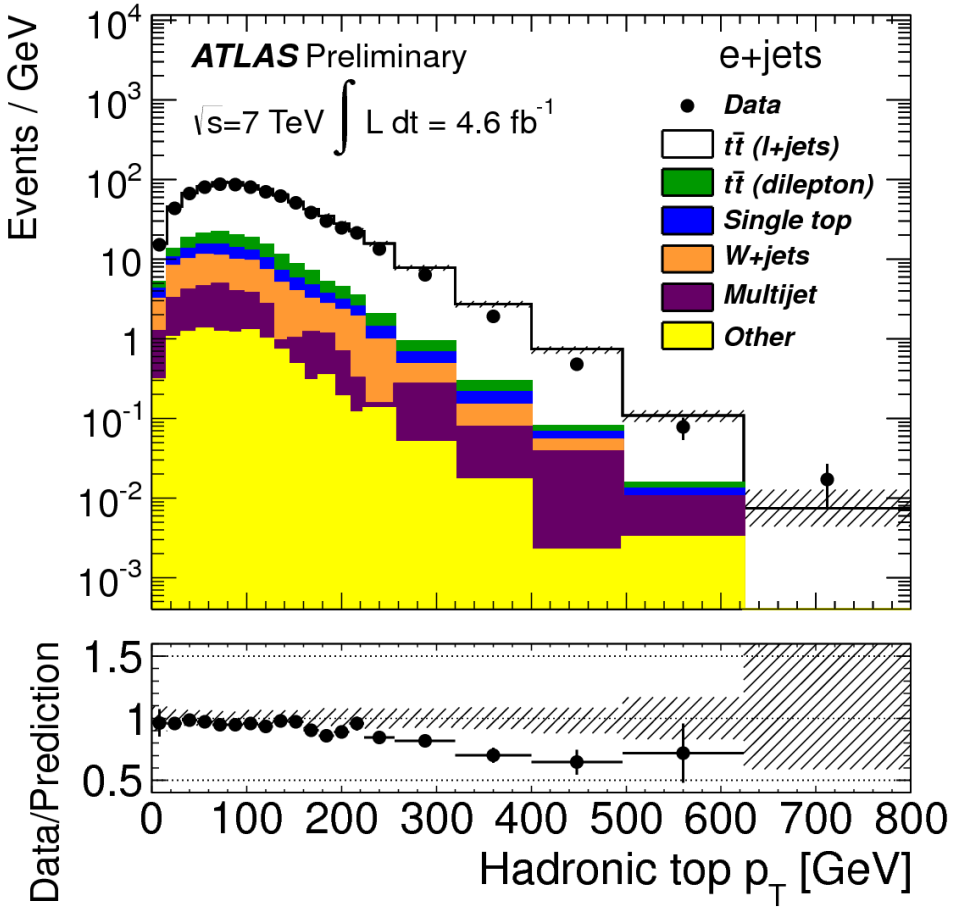


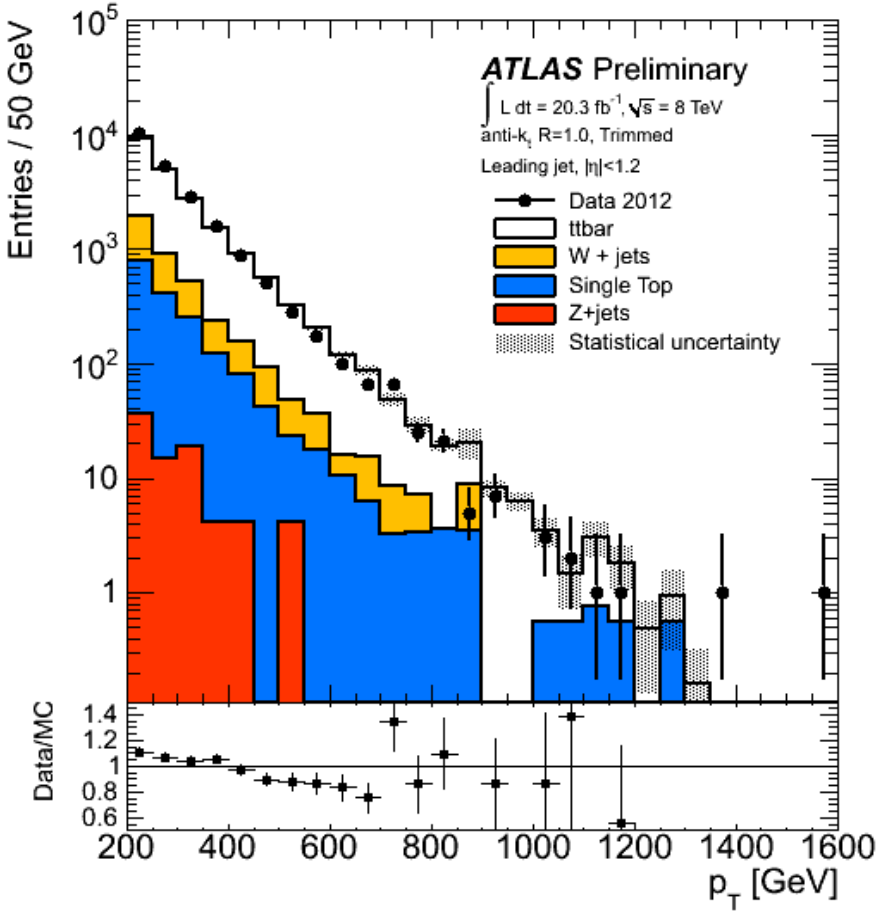
ATLAS Generators Cheat Sheet

- ~30 MC Generators (how many can you name?)
  - Actually, it's been over 50 in run 1, from my count!
- >50 Combinations (Powheg+Pythia vs Powheg+Jimmy)
- 34 000 samples in the last MC campaign
- MC for 2011: ~5B events
- MC for 2012: ~6.6B events
  - Not including re-runs
  - Fastest is a few minutes / 5k events. Slowest is 5 events per ***DAY***
- We are still getting better at event generation
  - Still trying to figure out the best methods for integration, for populating tails
  - That is one place where we really need help and better interactions with theorists and with each other

Basic Problem #1: Slopes

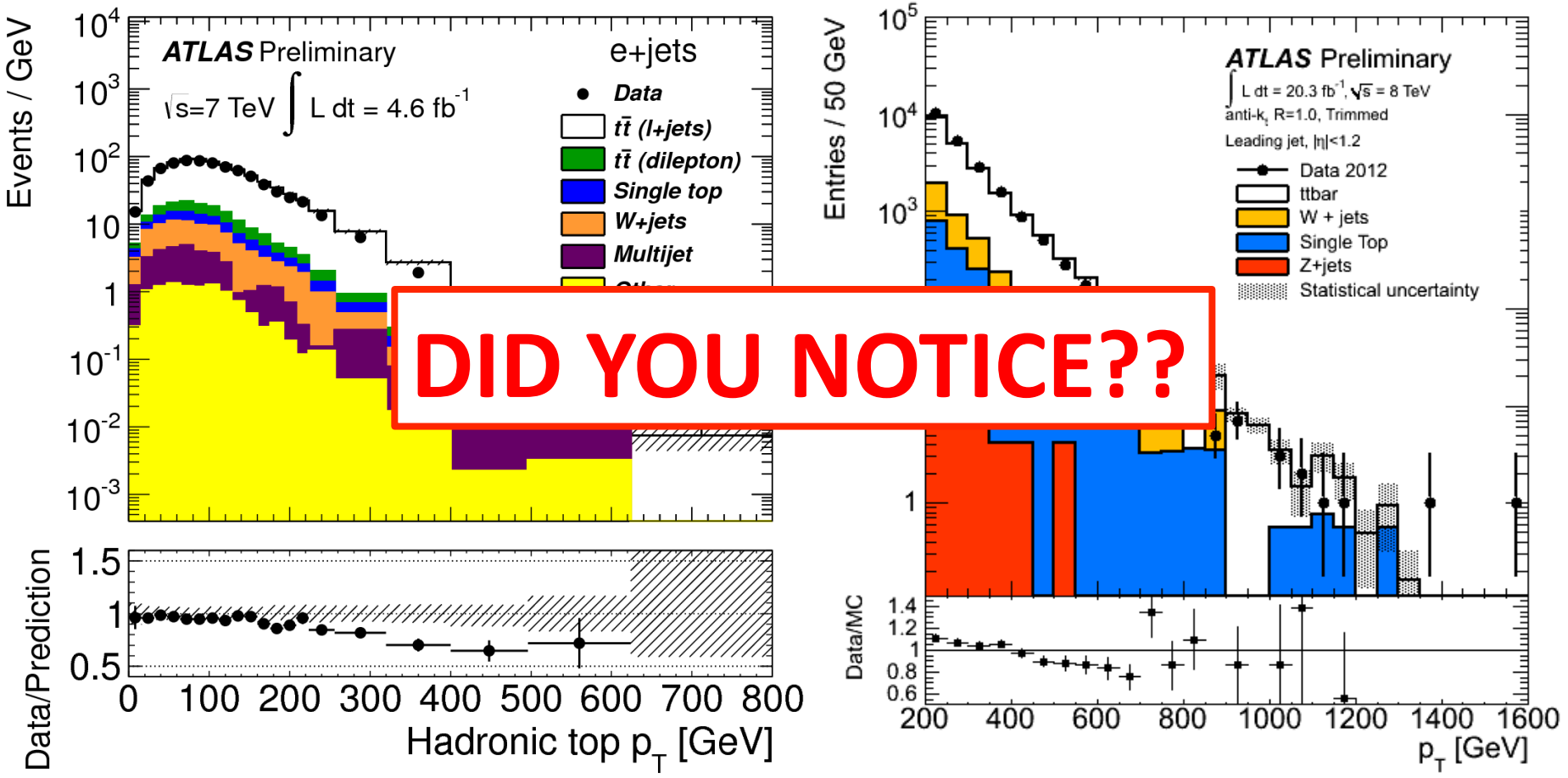
- Same picture for Alpgen+fHerwig and Powheg+Pythia 6, at 7 and 8 TeV
- We have no earthly clue of what this will look like at 14 TeV or whether this shape is linear, asymptotic, "just" a massive dip…





Basic Problem #1: Slopes

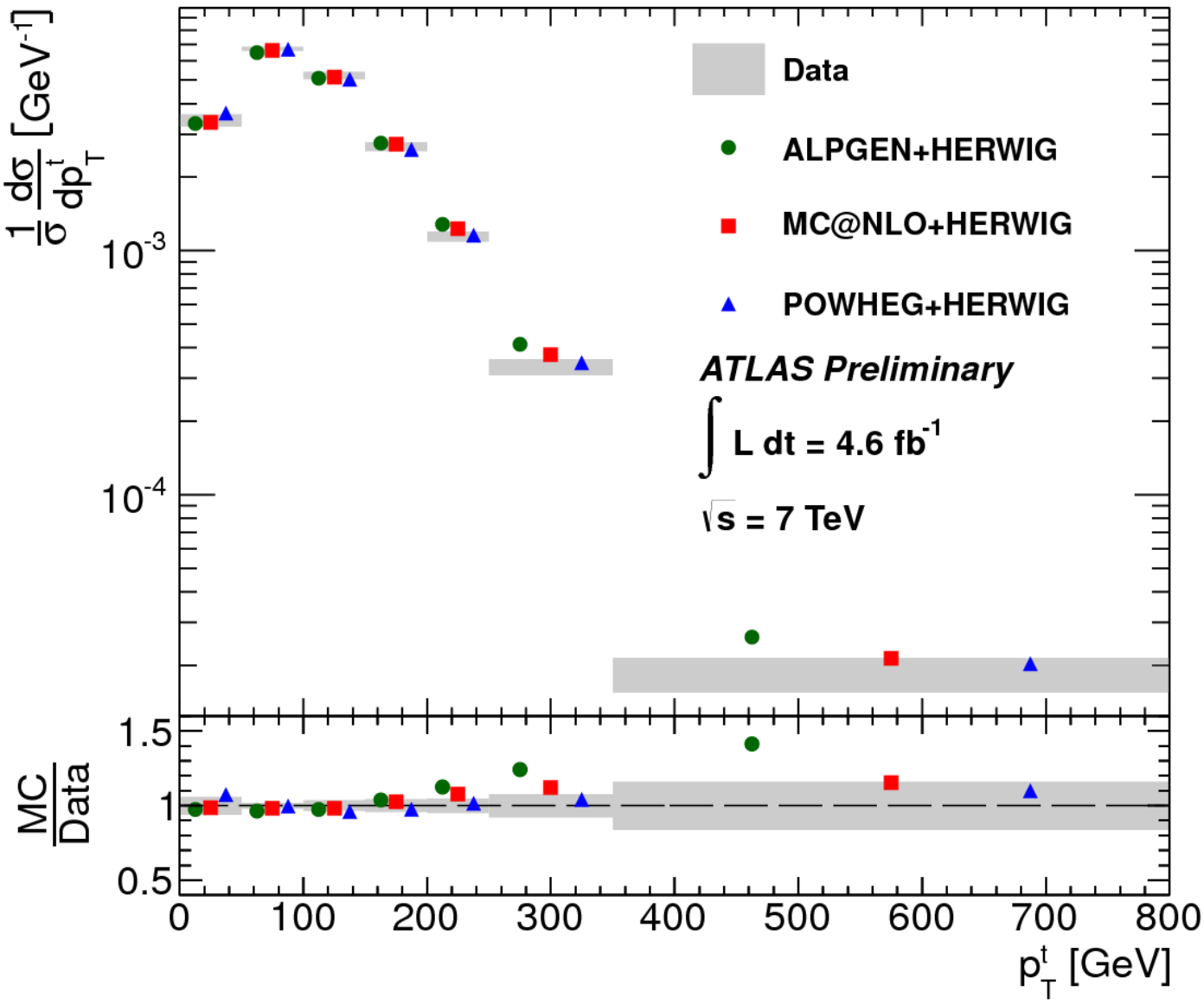
- Same picture for Alpgen+fHerwig and Powheg+Pythia 6, at 7 and 8 TeV
- We have no earthly clue of what this will look like at 14 TeV or whether this shape is linear, asymptotic, "just" a massive dip…



Consistently Everywhere

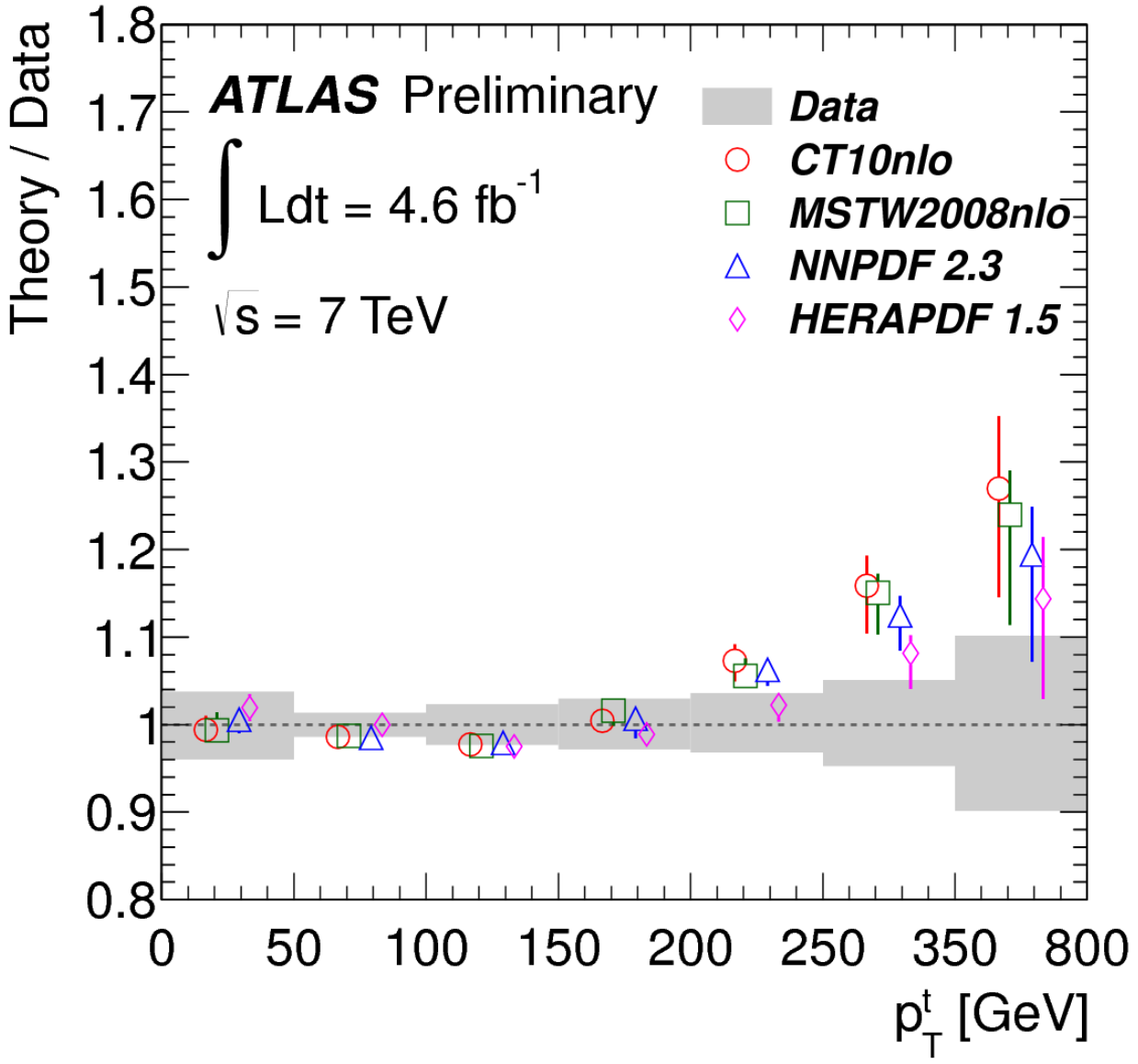
When the data are unfolded this is still true

For both NLO and multileg generators



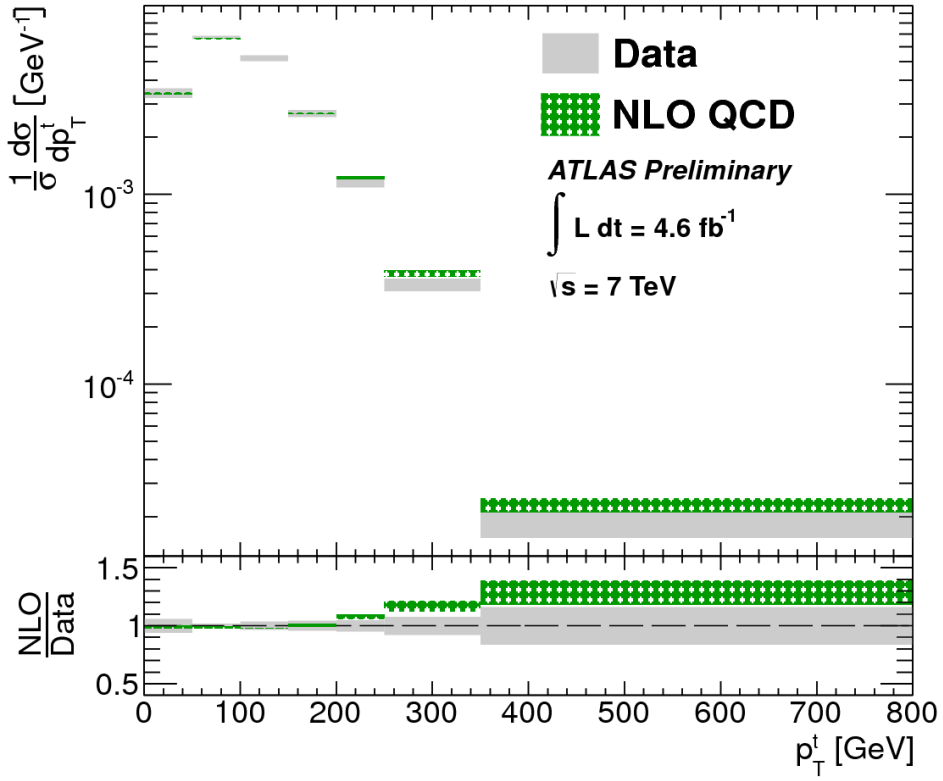
Consistently Everywhere

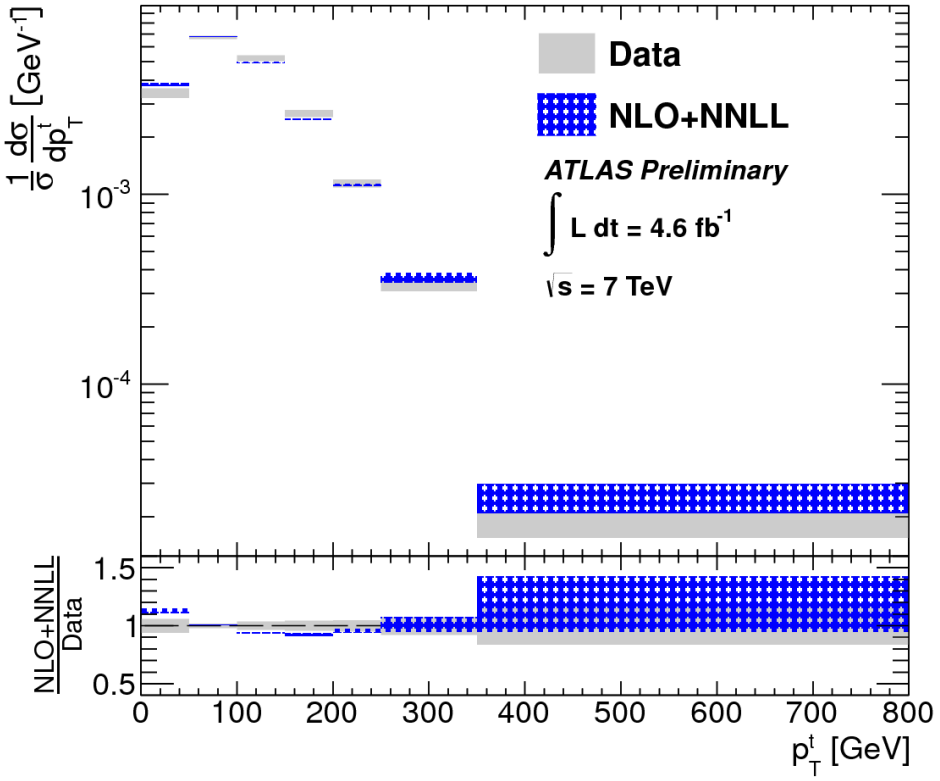
It looks like PDFs won't save you



Higher Order Corrections

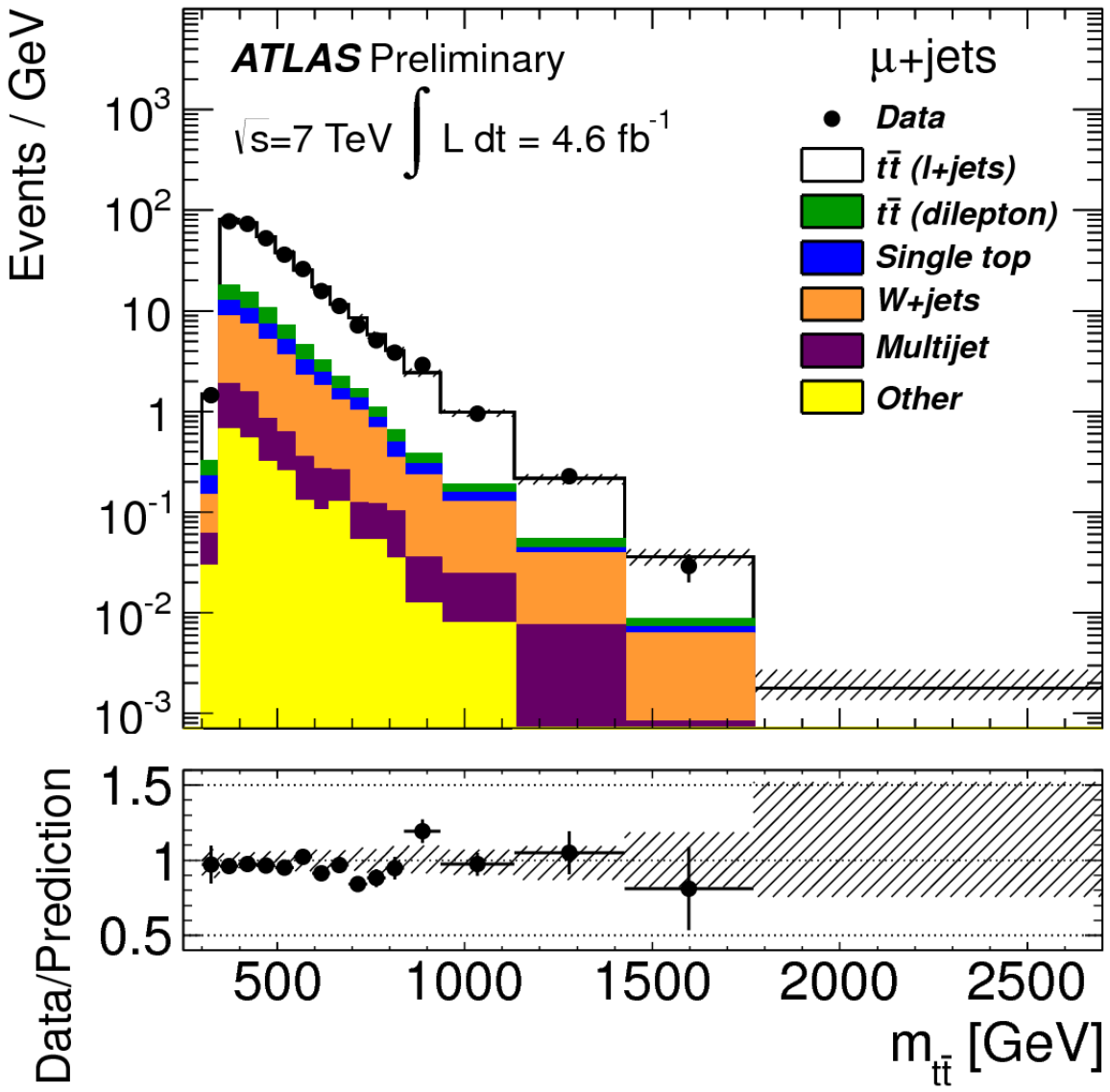
- We'll have to see what NNLO does
- NNLL mostly blows up the high-p_T error bars





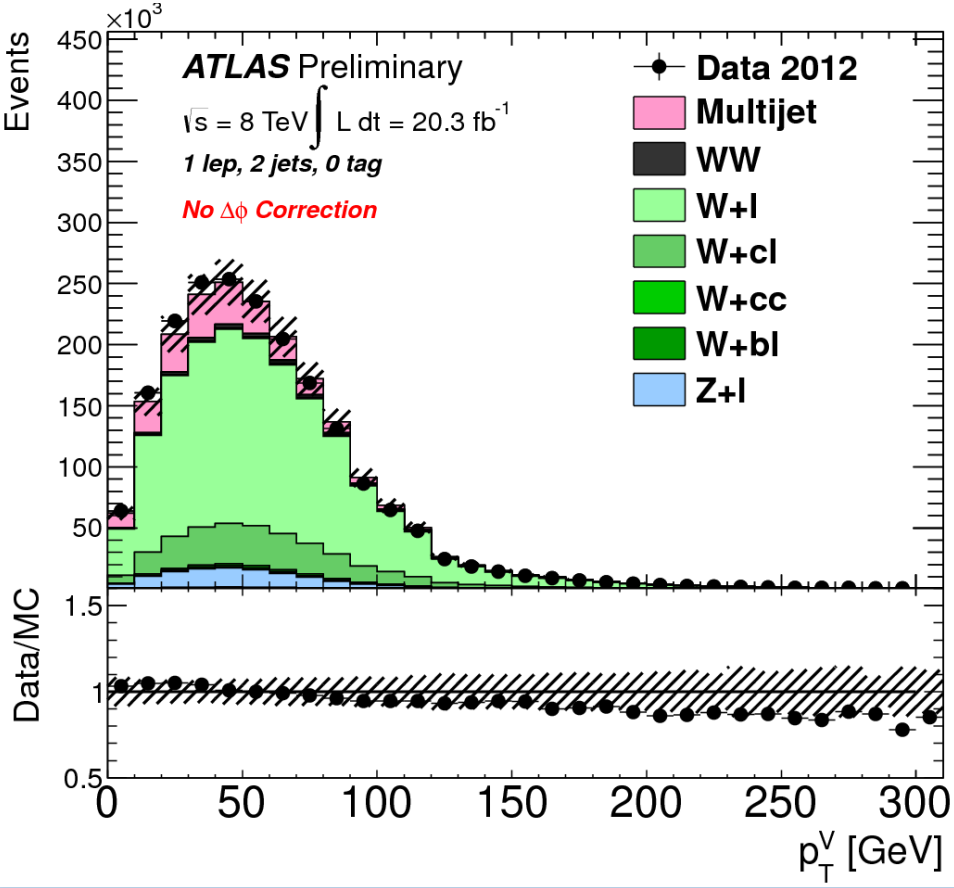
Different Observables

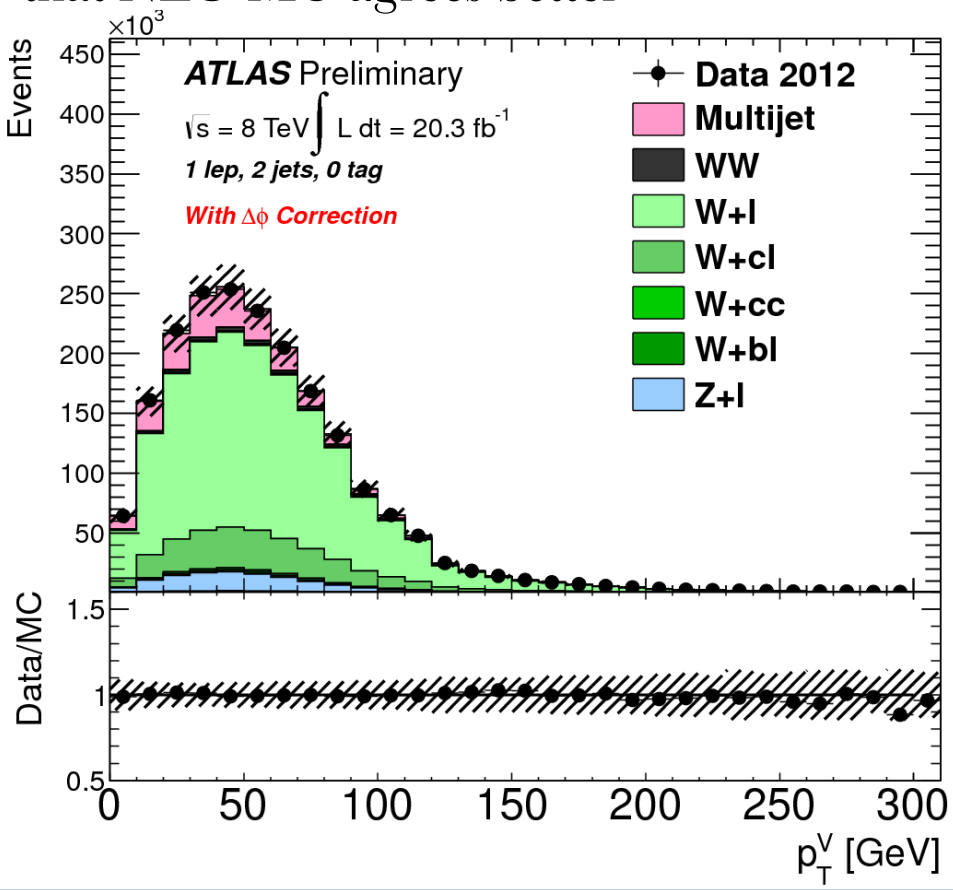
- Not true in all observables
- We don't want to pick our search observables based on this
- But maybe we should?
- The same applies to the discussion of uncertainties later
- This is something that theorists / generator folk could do…



W+jets MC

- We use Sherpa 1.4.1 (+CT10) to model W+jets at LO
- Clear discrepancy in $p_T(V)$ is fixed via reweighting the $\Delta\phi(j,j)$ distribution so that the MC agrees with the data
  - Some checks were made to show that NLO MC agrees better





So Many Questions

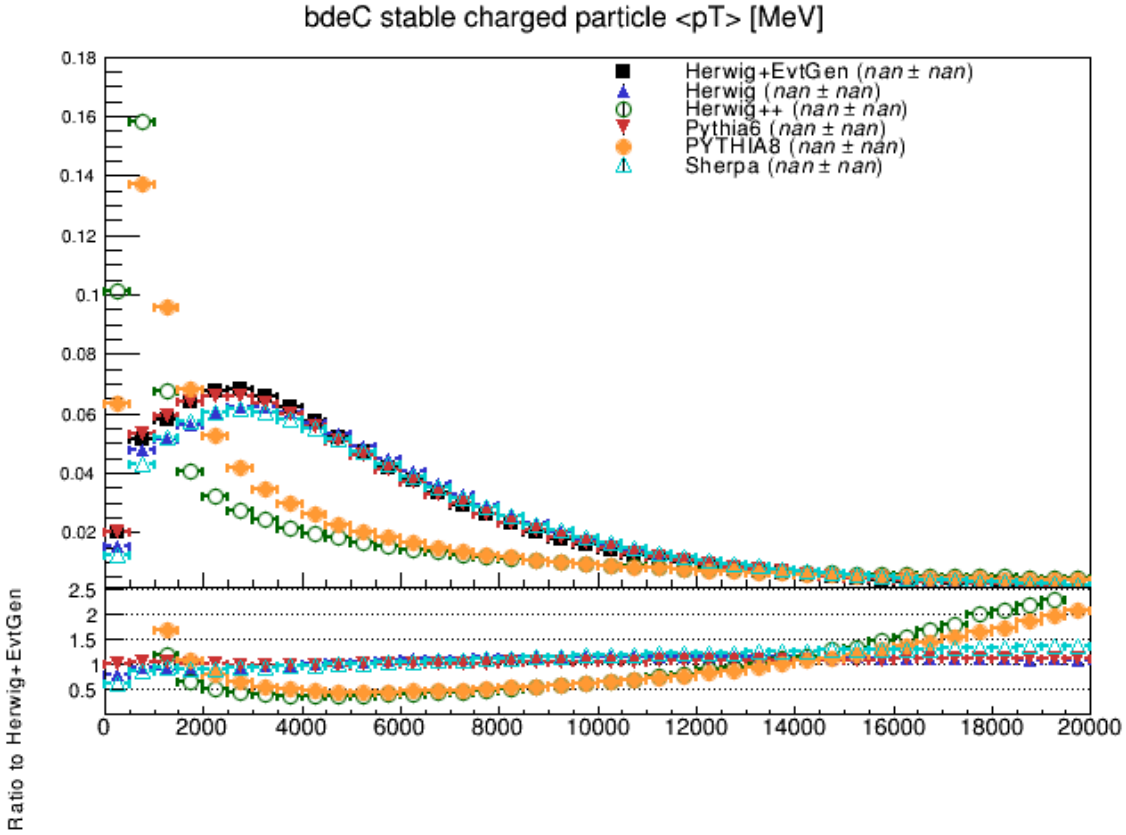
- Is there some variable that we should trust more than others for reweighting?
  - Notice here we took a $\Delta\phi$ rather than a p_T
- Do we believe our systematic uncertainties any more?
  - If we do, then surely they should cover any effect that we see
  - 1-sigma is still only 1-sigma, even if it's correlated and appears to look really bad or obvious
  - If we don't, then how do we assign new ones after reweighting?
- Do we have any sense of the origin of the discrepancies?
  - If we did, surely we'd use whatever generator did a better job…
  - If we are going to reweight, why bother with the fancy generator in the first place?
- If we believe this is a physics effect, why does *only Sherpa* have a problem? If other generators do, what do we do with the distributions? Reweight in the same way?
  - What about different processes?
  - Of course, the Higgs signal is generated in a different way…
  - Same question for SUSY, of course!!

Basic Problem #2: Heavy Flavor

- The number of heavy flavor problems we've had is *incredible*.
- Treating quarks as massive or massless we're getting better at
  - But massless in the matrix element and massive in the shower (MadGraph+Pythia/new Sherpa) causes momentum imbalance!
- But allowing heavy flavor to come from a matrix element or a parton shower causes *endless* problems for us
  - Think: where does your g→bb come from?
- We deal with this via something we call heavy flavor overlap removal. Automatic in Sherpa. By hand in Alpgen. Not treated at all in MadGraph.
  - I've met (many) fewer than a dozen people on ATLAS who claim to completely understand this issue – and no people outside of ATLAS who claim to understand what ATLAS are doing.
  - Surely generator authors have to deal with this too?
- We NEED to get better at this in time for run 2!!

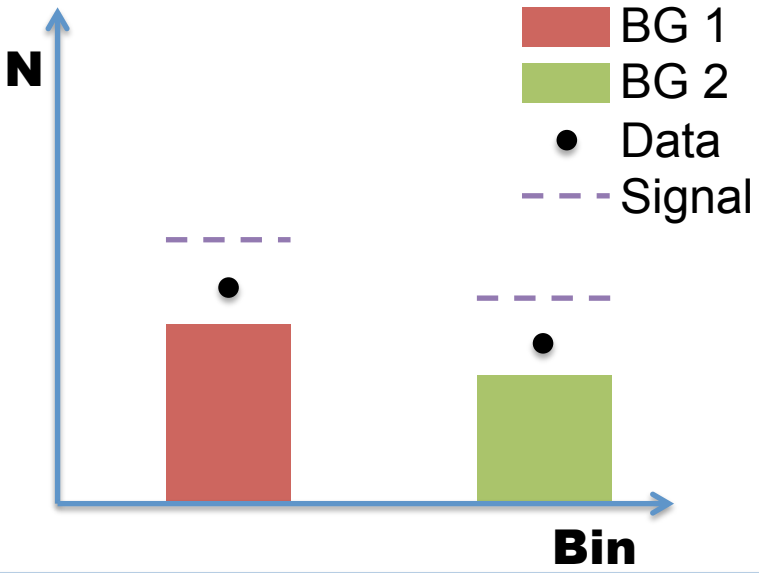
B-fragmentation

- Simple things are still *very* poorly understood
- These manifest themselves as apparent detector changes
- They can also have *huge* impacts on our detector systematics
- I'm still amazed at how variant some of these things are!!



Basic Problem #3: Correlations

- Surprisingly common myth: to be conservative, one should de-correlate samples or uncertainties
- This is *totally bogus*. Even in a simple case it breaks.
- We *must* treat correlations properly. And now the big problem: WHAT IS CORRELATED?
  - Simple example: are W+jets and Z+jets expectations correlated?
  - Are results from different generators ever correlated?



If I treat two backgrounds as *uncorrelated* when they are correlated – or vice-versa – we can get "too" aggressive or to conservative a limit depending on what we observe!

Basic Problem #4: Small Backgrounds

- Until this summer, ATLAS was including ttWW background and was *not* including tZ backgrounds
- Did anybody notice?
  - The cross-section is ~1/3 that of ttZ; ~700 events expected in 2012
- Rarely are there k-factors for these processes, but they can be quite large
  - MadGraph can do these backgrounds, as can Sherpa, but specialty generators have to be set up for the process in advance
  - Genuine question: are there any caveats to Sherpa / aMC@NLO's ability to generate complex final states? Should we just be using those things for all our rare backgrounds to get at least NLO accuracy?
  - Of course, that'd imply a lot of time spent on generating tiny backgrounds…
- What happens if we see an excess? Is there someone looking into backgrounds like this?

Basic Problem #5: Systematics

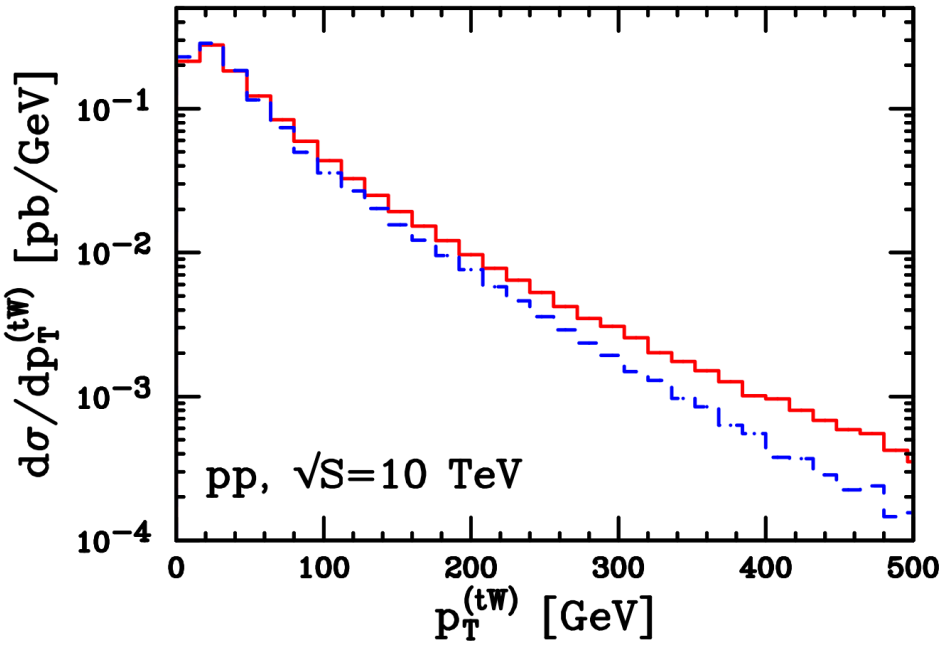
- We have a lot of comparisons to set systematics on generators
  - I like the work that's going on now to improve the handling of scale and PDF uncertainties in generator output – keep it up!!!
- We really have a poor understanding of what covers what, what over-covers, what under-covers, and what double counts
- We also do some things that deeply bother me, which I will now take the opportunity to point out
- On the whole, keep in mind that we are VERY concerned about over-trusting our MC, and (on the whole) when we here about some new effect, we will ADD a systematic uncertainty to account for it
  - We are still not great at removing statistical uncertainties from systematics, which means this will almost always inevitably INCREASE the total systematic uncertainty…
  - I am kinda surprised no group of theorists have ever come up with recommendations on this topic…

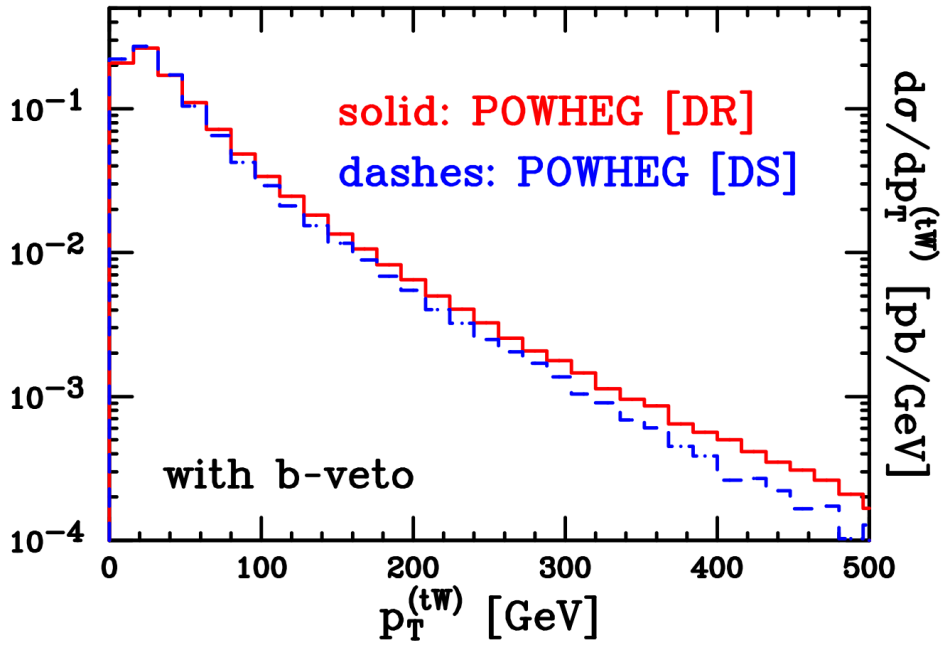
Example: Single Top

- If you read some recent SUSY papers, you should find these theory systematics for single top
- Cross section uncertainty
  - Including PDF and scales
- PDF uncertainty for shape
  - 40-50 intra-PDF variations + α_S + inter-PDF
- AcerMC comparison to account for ISR/FSR
  - Isn't this largely an α_S issue?
- Pythia6 vs fHerwig+Jimmy
  - Ahhh, the usual "we have two generators, compare them" systematic
- MC@NLO+fHerwig+Jimmy vs Powheg+fHerwig+Jimmy
  - "In principle" these should be identical – building in fluctuations?
  - Notice no multileg option… expect that soon enough
- Interference effects

Single Top Interference

- Two schemes for treating single top interference exist in Powheg: diagram removal and diagram subtraction
- We don't know which is reasonable
- So we take one as an asymmetric systematic uncertainty
- This makes a BIG difference at high-p_T!!
  - It blows my mind a little bit that this isn't a solved problem…



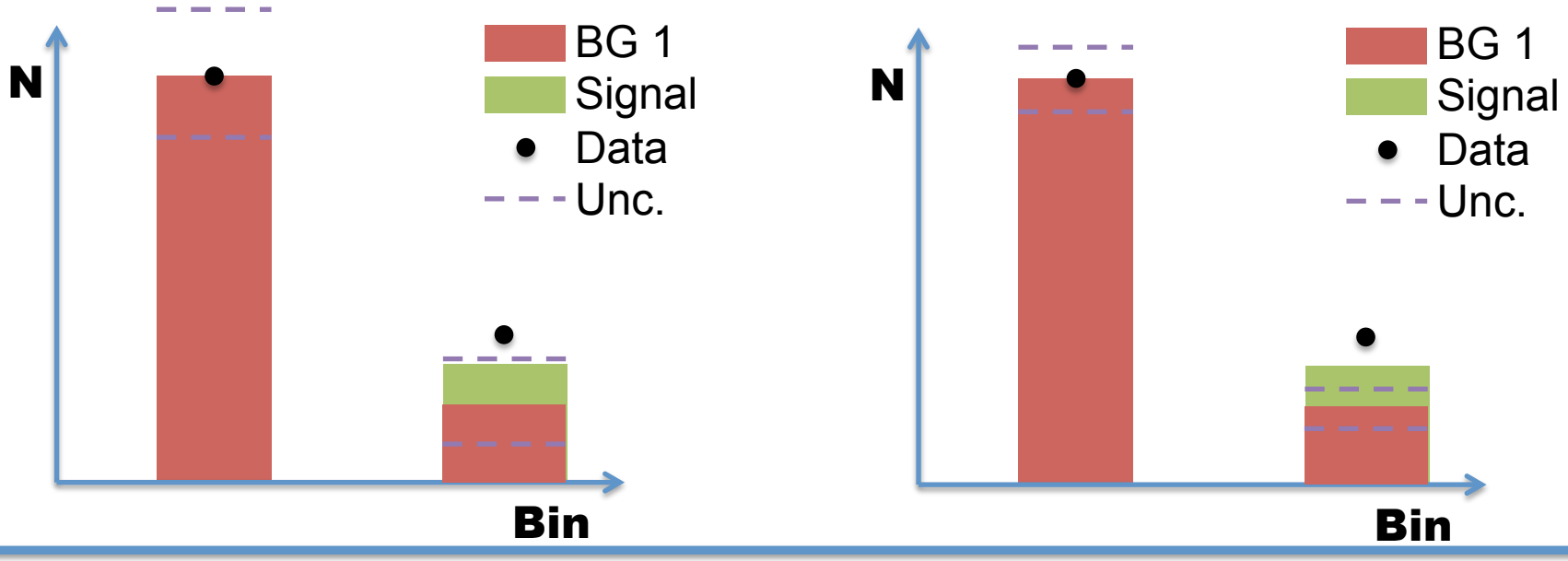


Example: Diboson Backgrounds

- For quite a while, Powheg was used for diboson backgrounds for many analyses
  - Our Sherpa samples were buggy, MC@NLO wasn't really available, Herwig is only leading order and doesn't treat VVbb well…
- Now aMC@NLO is coming strongly into favor
- This change is largely because aMC@NLO has a more clear breakdown of systematic uncertainties available to us
- Until this point we were using aMC@NLO in various configurations to generate systematic uncertainties to apply to the Powheg nominal expectation
  - No one is super comfortable with that!
  - And this isn't the only place; we've also used Alpgen scale variations as systematics on Sherpa Z+jets…

Basic Problem #6: Nuisances

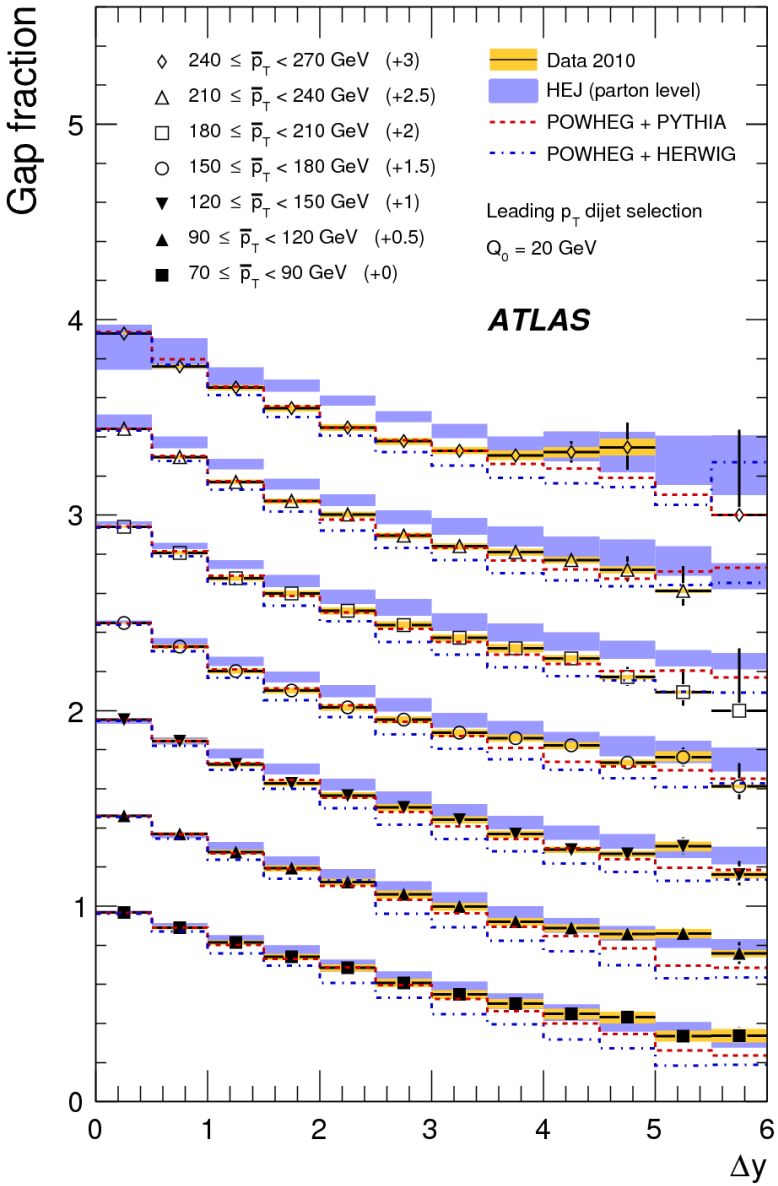
- For many searches (and heavily in the Higgs group) we use *fits* to the data
- Neither we nor CMS really ever show the results publicly
- The idea is to use our knowledge to constrain the prediction in the signal region beyond just the "expected counts"
- This is totally fair if you believe you really have a full understanding of the uncertainties!



Signal Samples

- For SUSY, we generate signals with Herwig++ and MadGraph
- For some signals it's clear that we care about the production of additional jets / the spectrum of the system
  - Mono-X searches, ISR-heavy searches, small-mass-splitting searches
- For others, it's really not clear, but consistency wins
  - The signal shouldn't change generators mid-grid
- Of course, systematics become a problem
  - Changing scales and settings works in MadGraph, but how does one get a good sense of comparable systematics in Herwig++?
- Little help from the theory community on this one
  - More often we get the question "how did you generate this sample" so that someone can copy our settings, NOT because they think our settings are incorrect and wish to help fix them…
  - Really we should all be working together on these

Problem #7: Crazy Generators



- Fraction of events without a jet (with a gap) between the two highest-p_T jets
- Studies of wide-angle QCD and soft-scale gluon emissions, help with modeling for gluon fusion-like variables
- **BFKL-motivated generators generally don't do as nice a job of describing the data**
  - **They were *designed* for this!**
- HEJ, parton-level only, presumably missing some parton shower effects that become important at high p_Ts
- Some differences are on the scale of the difference between Pythia and Herwig showering, others slightly larger
  - Alpgen doesn't do well here, actually!
- Generally well modeled by Powheg+Pythia

Reminder #1: We Love New Stuff

- Sherpa's now the workhorse for multi-leg Z+jets samples
- We've produced an *incredible* number of buggy Sherpa samples
  - Problems with b/c mass in the matrix element
  - Problem with showering the MPI
  - Scale variations not representing full uncertainty
  - Unstable particles that are declare stable
  - Bug in QED radiation
  - …
- All that to prove *how excited we are* to use fancy new generator technology when it is available!
- We are *heavily* testing Sherpa 2.0 now…

Reminder #2: We Try Lots of Stuff

- Snipped from a paper's multijet background description:

> H_{T}-based trigger. The observed number of events in this region are then projected into the signal region using transfer factors from MC simulation. Alternatively, in order to model the mis-measurement of jets in the calorimeter, jets in events collected with these single-jet triggers are smeared according to response functions estimated using data [10]. Both of these methods result in an estimate consistent with that derived in the main fit.

- This involved 7 independent background estimates including four different MC configurations
- That comes with two messages:
  - Theorists: don't be shy about asking us to try things out, or even asking us what we've tried
  - Experimentalists: maybe we should be less shy about asking theorists for help with these things

Conclusions

- ATLAS has an incredible diversity of generators in play
- There are some "known" (to us only?) issues that pervade our generated samples
  - Slopes in some distributions
  - Correlations between samples
  - Small backgrounds
  - Systematics that fully cover known effects
  - Treating nuisance parameters
- We would very much welcome help from the community in understanding (and resolving) these issues
  - We all agree that measurements are one good way to help – we are working on those, but guidance is welcome
- And we ourselves need to think hard about how to address them coherently in time for run 2
  - Some of these are only going to get worse!