Top Reconstruction Algorithms in ATLAS

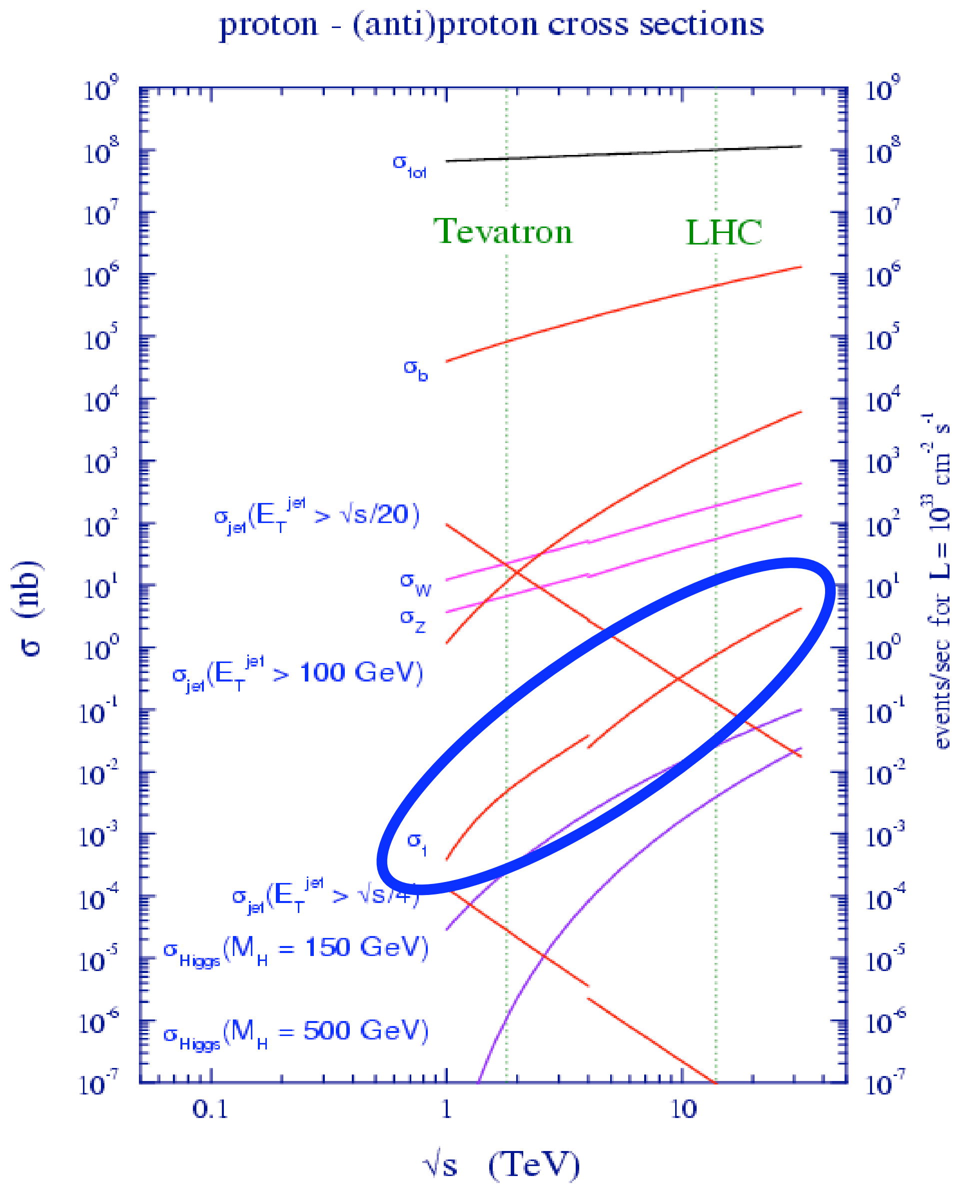
Gustaaf Brooijmans

Columbia University in the City of New York

3rd Top Workshop, Grenoble, Oct 23-25 2008

From Rare to Common

- Single + pair production ~1.2 nb (in SM)
  - Not much smaller than Z→ee
- Top quark physics, but also calibration sample, and significant background source
- Reconstruction algorithms with wide range of sophistication, tailored to specific purposes

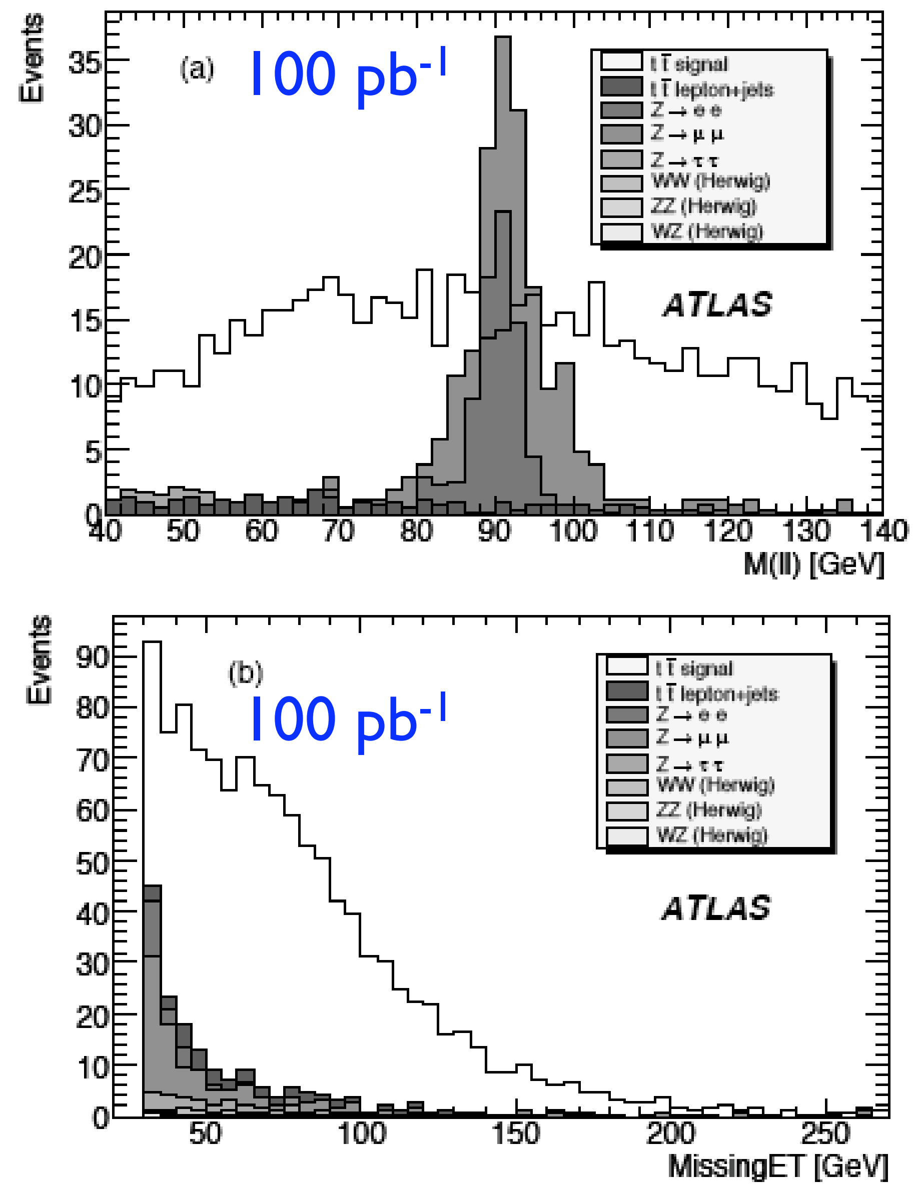


Types of Algorithms

- Selection-only
  - Require certain objects or sophisticated combinations of objects, no attempt to reconstruct top quark(s)
- "Individual" top quark reconstruction
  - "Classical": low and moderate p^T
  - High p^T
- Global Event Fitters
  - ➡ Top candidates are correlated
- Endpoints

Selection Only

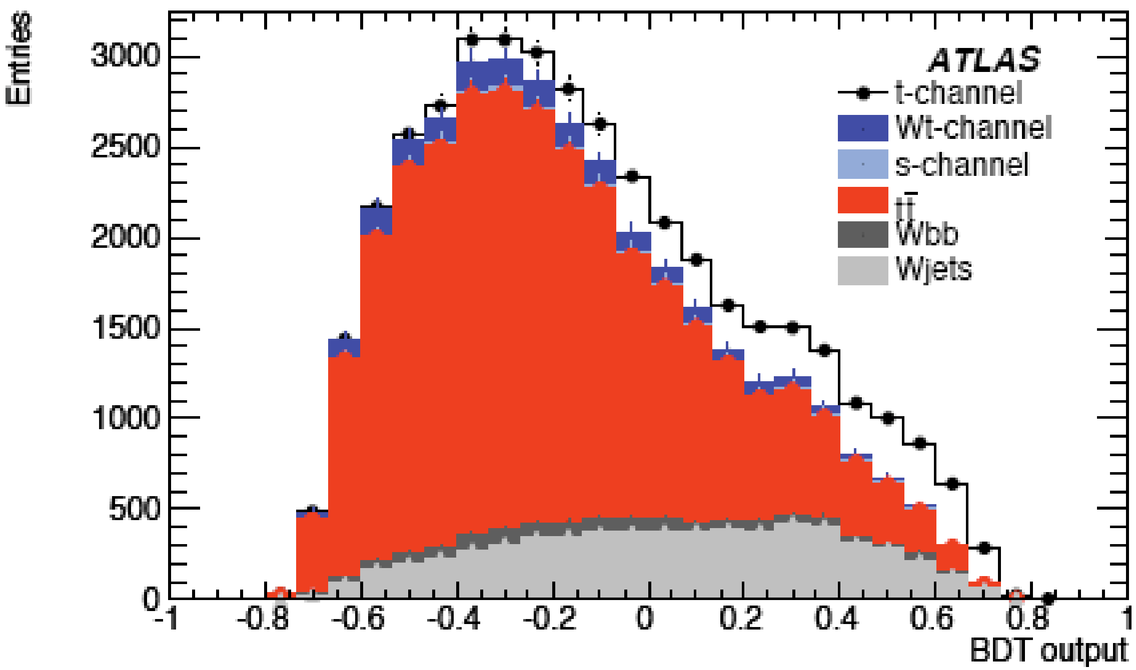
- In a sense, simple, since make no attempt to reconstruct top quark(s)
- E.g. dileptons: require
  - 2 good leptons $p^T > 20$ GeV
  - Z veto
  - 2 jets $p^T > 20$ GeV
  - MET > 25-35 GeV

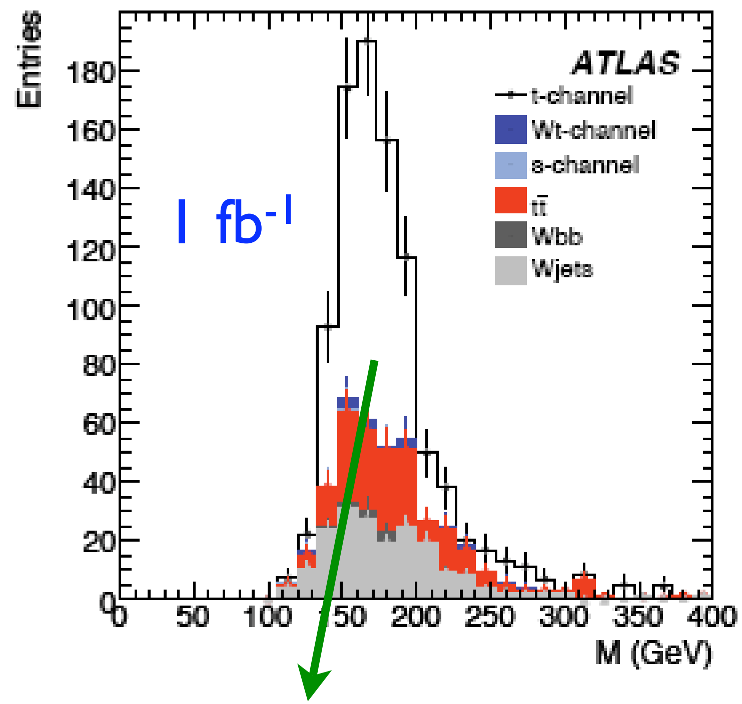


ee, eμ and μμ channels together → S/B ~4

Also works in lepton + (4) jets

- But even selection-only can be complex
  - E.g. in single top, use BDT to "select" events





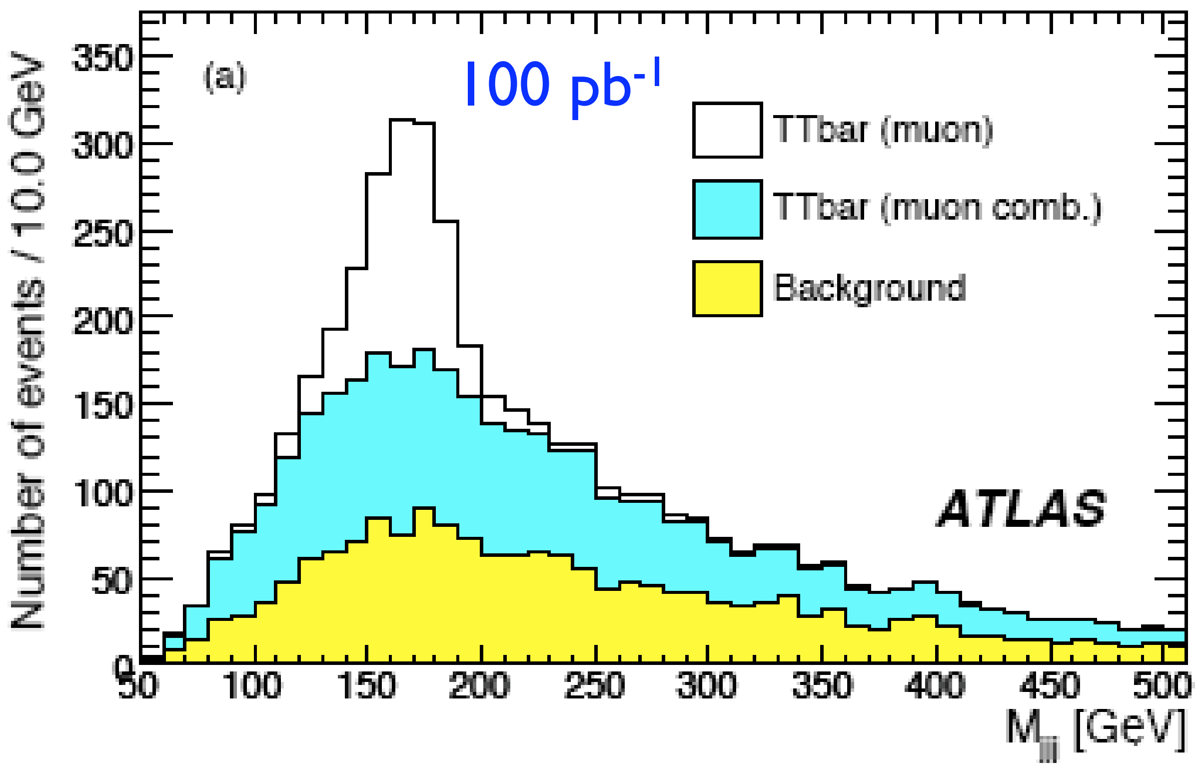
Leptonic top candidate mass after BDT output > 0.6

"Individual Top Reconstruction"

- Reconstruct individual top quarks "independently" of the rest of the event
- Mostly relevant in the context of $t\bar{t}$ events
  - Main problem is combinatorics
    - Alleviated somewhat through b-tagging
- At high p^T, no combinatoric problem
  - But need special techniques to reject light quark & gluon jets
- Don't introduce correlations $\rightarrow$ less sensitivity to detailed understanding of efficiencies and resolutions

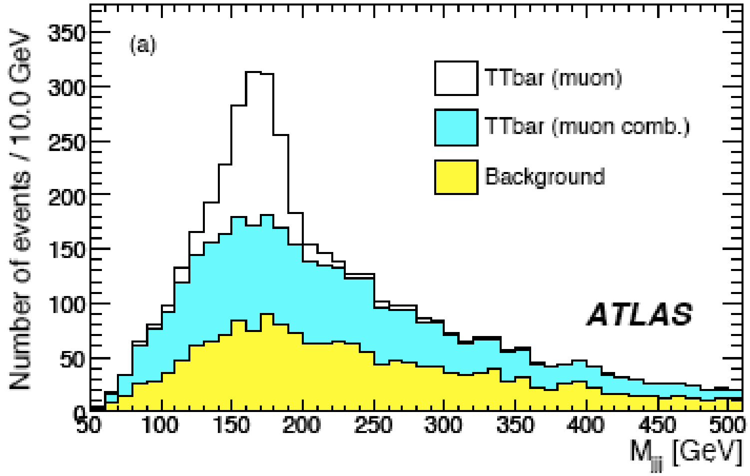
Commissioning Analysis

- "Keep it simple"
- Simplest:
  - e, μ $p^T > 20$ GeV
  - MET > 20 GeV
  - 3 jets $p^T > 40$ GeV
  - 4th jet $p^T > 20$ GeV

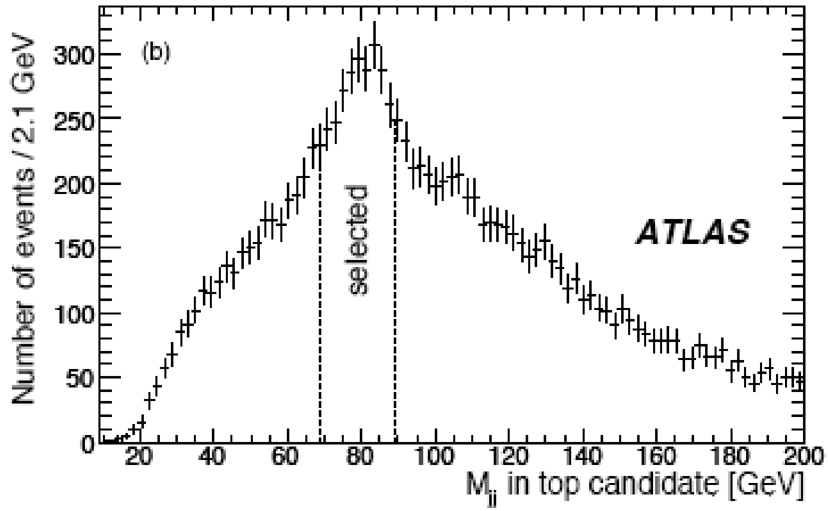


- Hadronic top is 3-jet combination with highest p^T sum
  - Largest background is signal combinatoric!

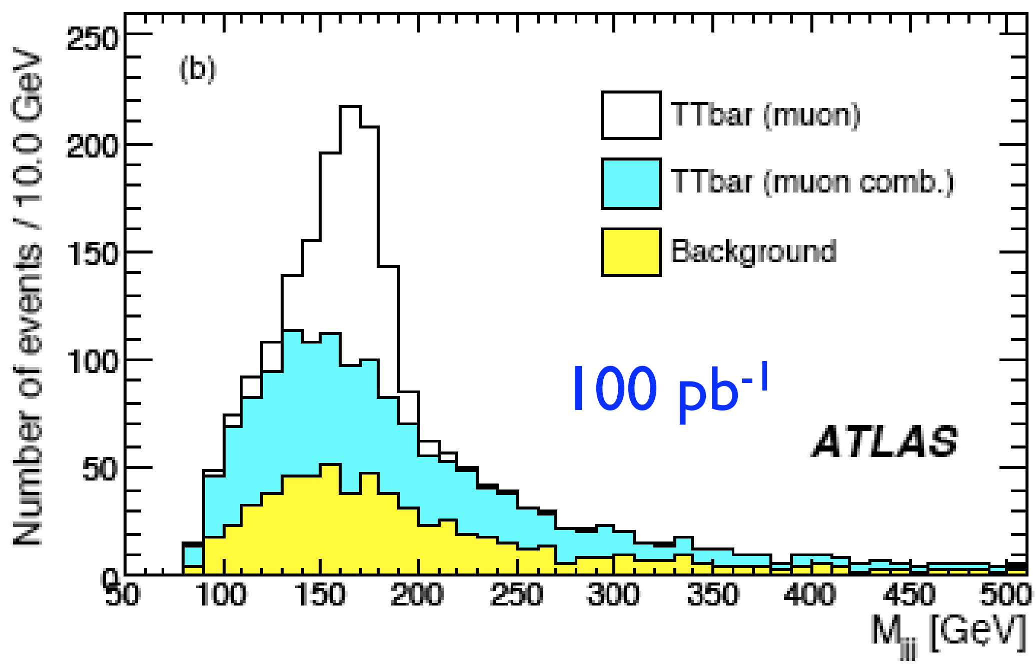
- Try to "purify" by requiring that among the three jets, one pair has mass = W



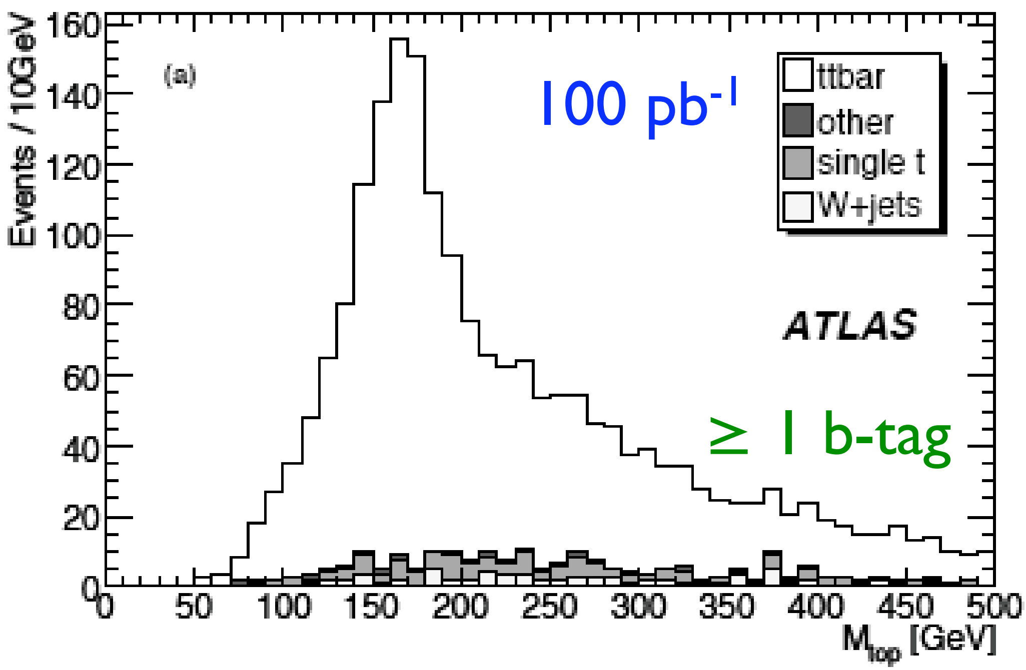
+

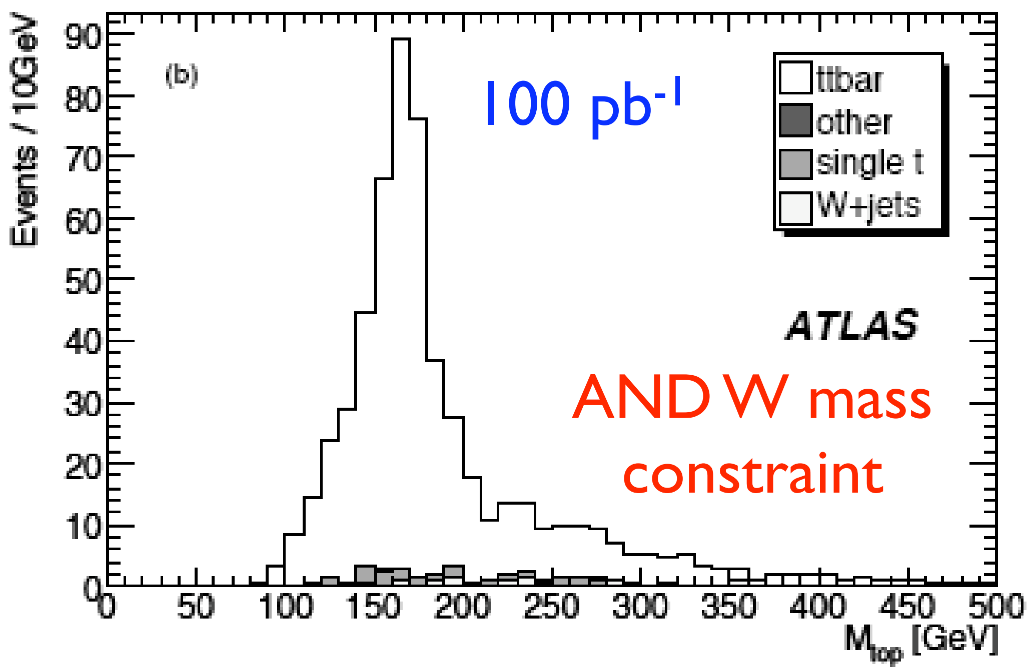


=



- b-tagging reduces the combinatorics & W+jets background substantially...
  - And with W mass constraint & b-tagging, achieve a purity of 95%
    - total efficiency ~1% (incl BR), vs ~7% for "simplest" analysis

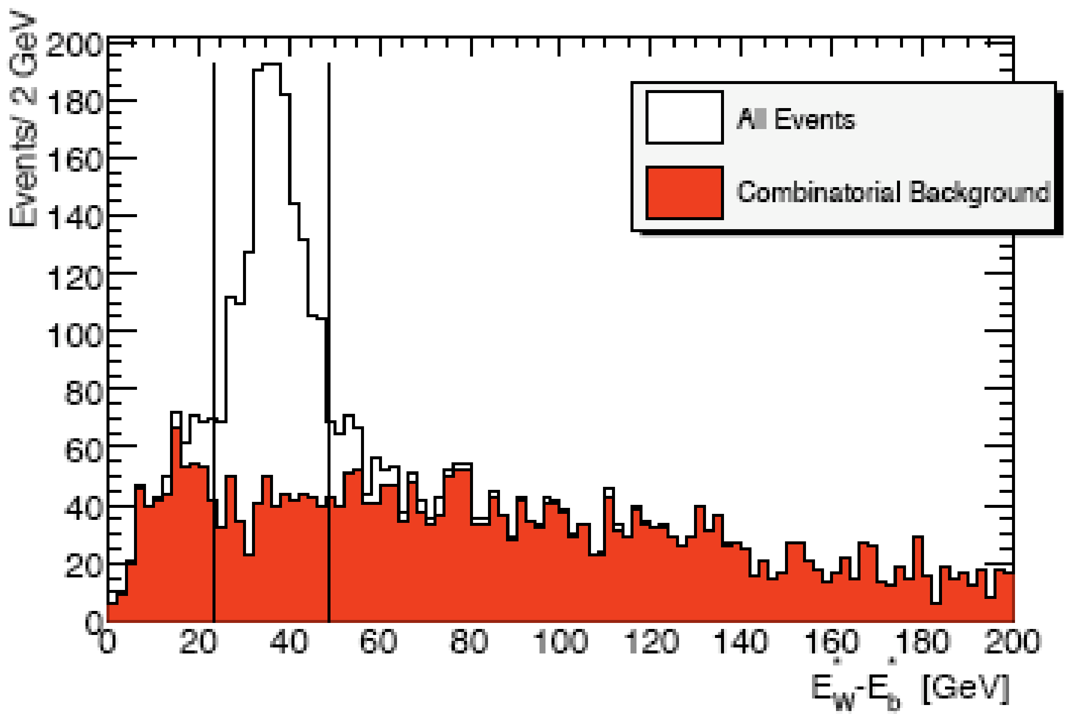




- There are other "purification" variables:

$$X_1 = E^*_{\mathrm{W}} - E^*_{\mathrm{b}} = E^*_{\mathrm{j1}} + E^*_{\mathrm{j2}} - E^*_{\mathrm{b}} = \frac{M^2_{\mathrm{W}} - M^2_{\mathrm{b}}}{M_{\mathrm{top}}}$$

(E^* = energy in top quark rest frame)



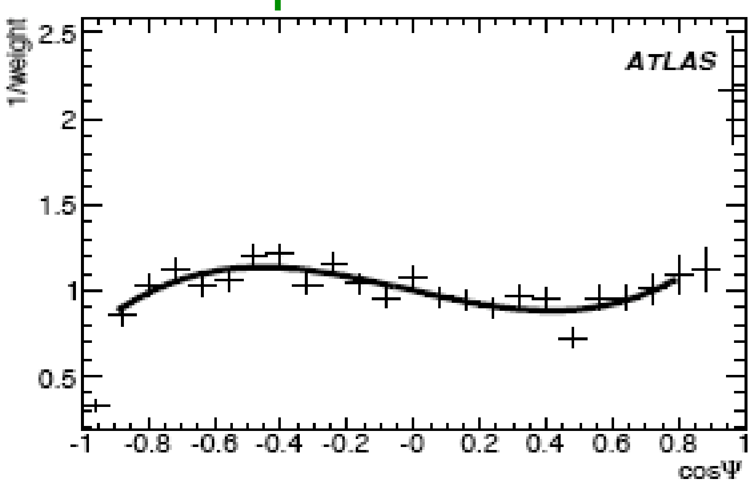
- Angular distributions
  - W polarization in top decay, top spin correlations

$$\frac{1}{N}\frac{dN}{d\cos\Psi} = \frac{3}{2}\left[F_0\left(\frac{\sin\Psi}{\sqrt{2}}\right)^2 + F_L\left(\frac{1-\cos\Psi}{2}\right)^2 + F_R\left(\frac{1+\cos\Psi}{2}\right)^2\right]$$

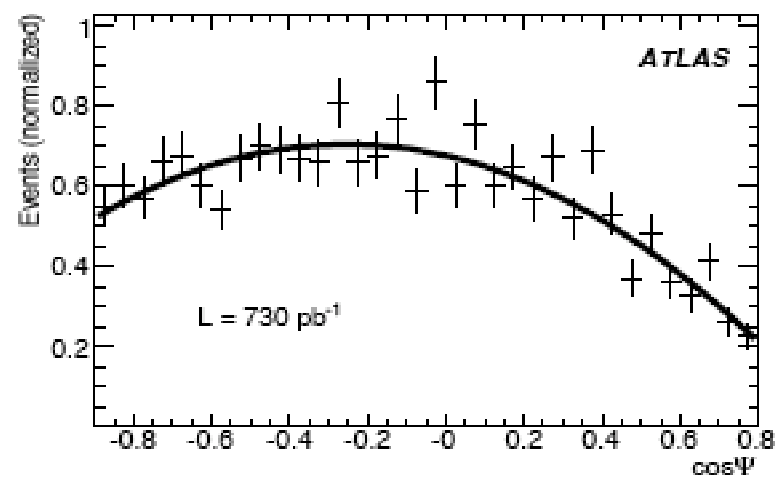
Higgs!

  - Requires reconstructing top & W restframes, then measure angles

Acceptance & Efficiencies shape distribution

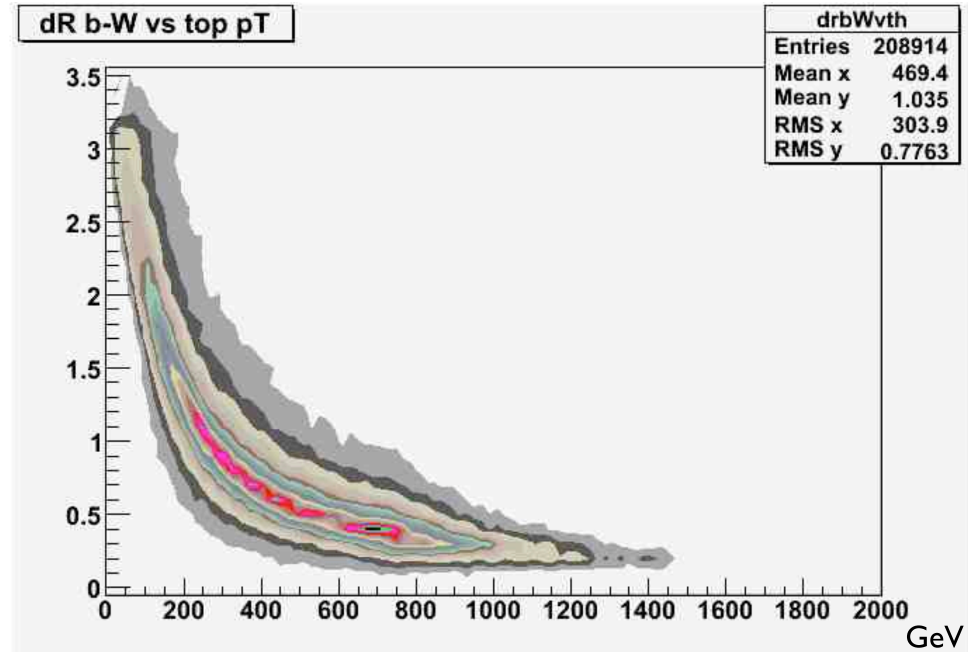


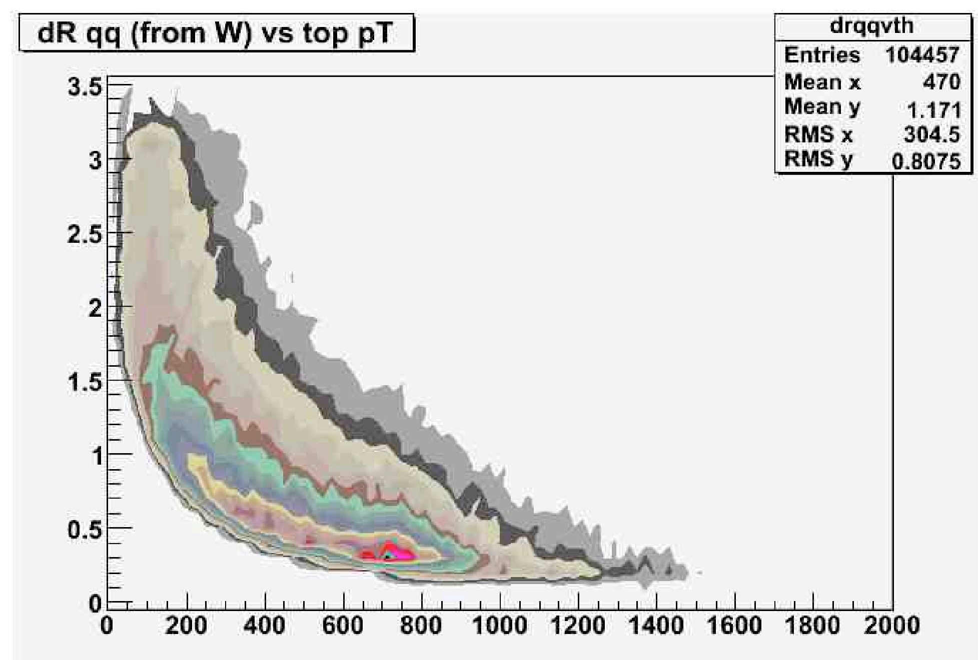
Corrected result



High p^T

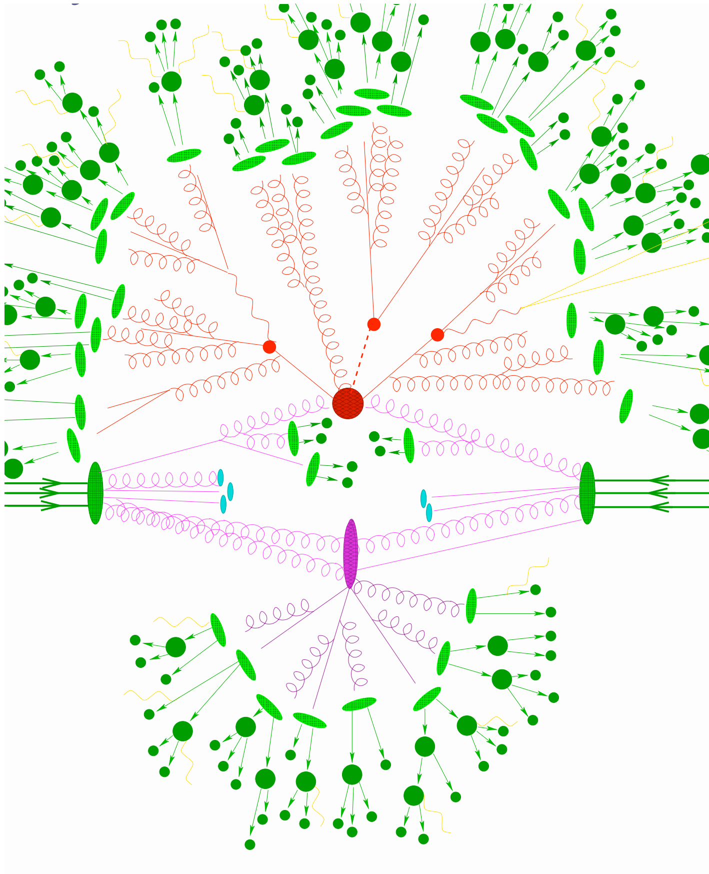
- Collimated decay products
  - $dR = \sqrt{(\Delta\eta^2 + \Delta\phi^2)}$
  - Typical jet radius ~0.5
  - But calorimeter segmentation much finer (especially in EM)
- For top pT > ~300 GeV
  - dR (qq from W) < 2 R_{jet}
  - dR (bW) < 2 R_{jet}
    - (No isolated lepton!)





Jet Structure

- Decay hadrons reconstructed as a single jet
  - But even if it looks like a single jet, it originates from a massive particle decaying to 2/3 hard partons, not one
- If I measured each of the partons in the jet perfectly, I would be able to:
  - Reconstruct the “originator’s” invariant mass
  - Reconstruct the direct daughter partons
- But
  - quarks hadronize -> cross-talk
  - my detector can’t resolve all individual hadrons



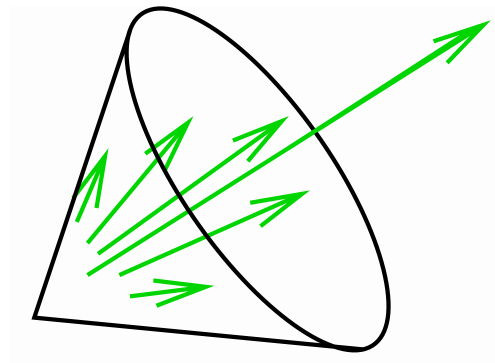
k_T Distance

- k_T jet algorithm is much better suited to understand jet substructure than cone:
  - Cone maximizes energy in an η x ϕ cone
  - k_T is a "nearest neighbor" clusterer

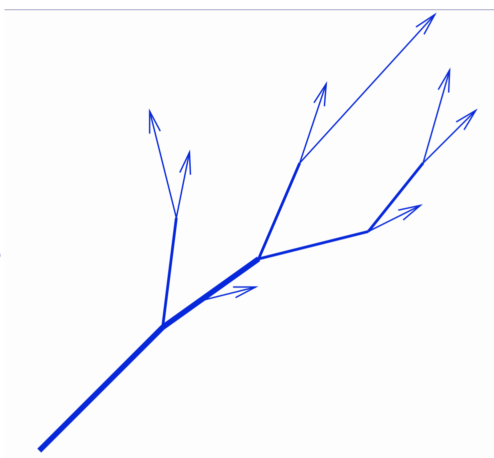
$$y_2 = \min\left(E_a^2, E_b^2\right) \bullet \theta_{ab}^2 / p_{T(jet)}^2$$

$$Y \text{ scale } = \sqrt{p_{T(jet)}^2 \bullet y_2}$$

- Can use the k_T algorithm on jet constituents and get the k_T distance (y-scale) at which one switches from 1 -> 2 (-> 3 etc.) jets
  - scale is related to mass of the decaying particle

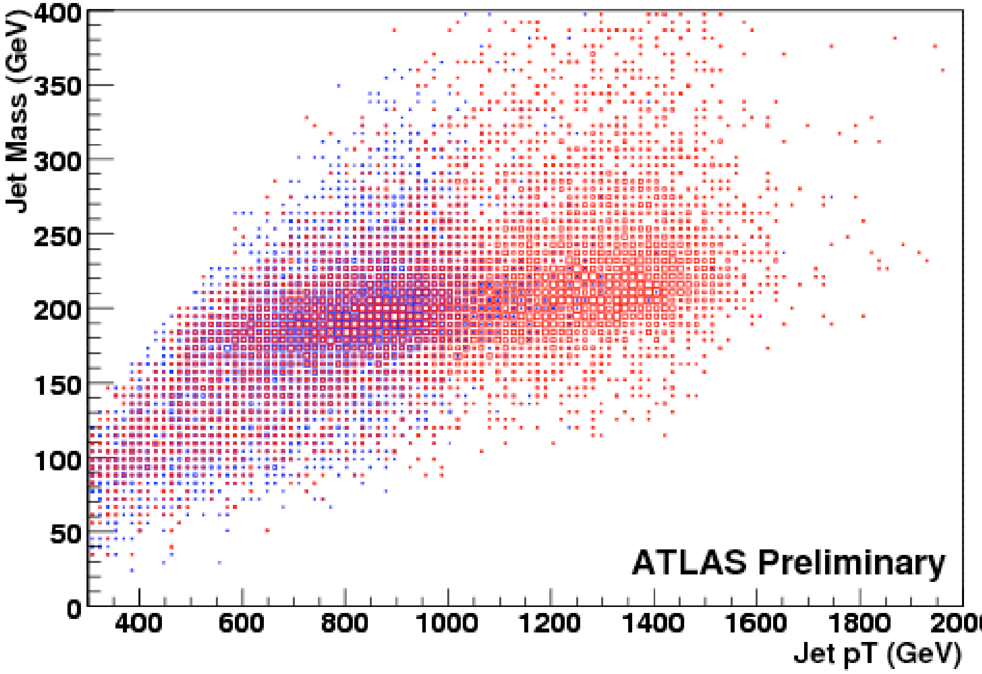


Cone

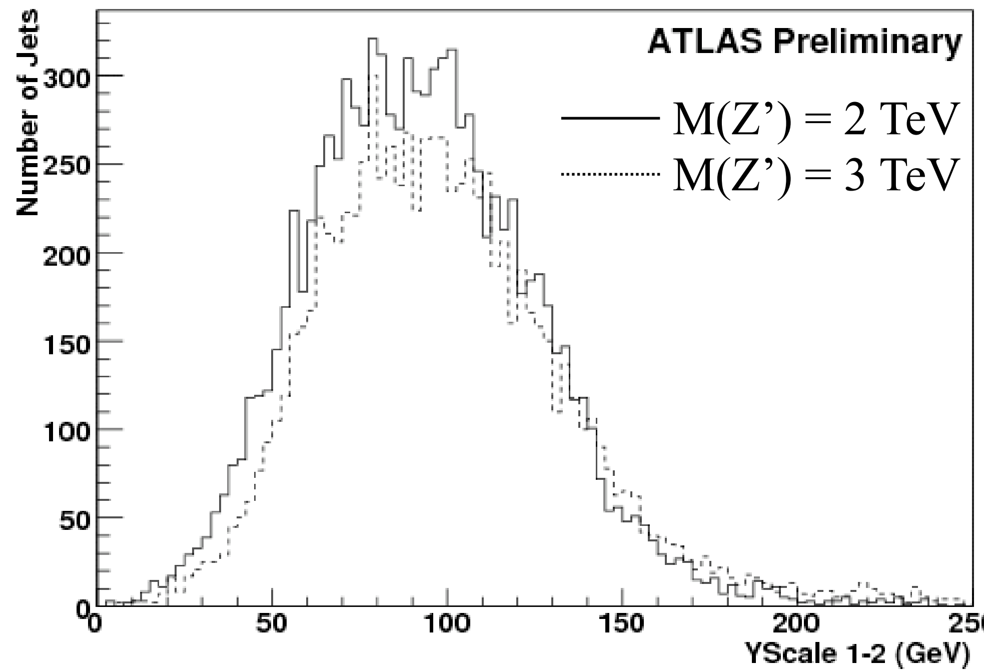


kT

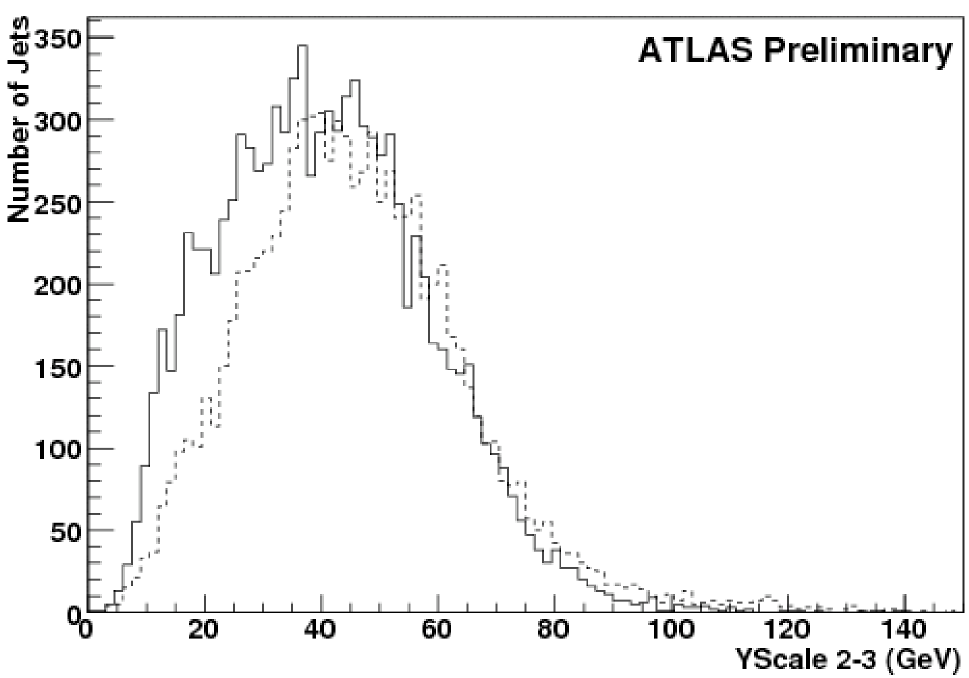
Z’ → tt



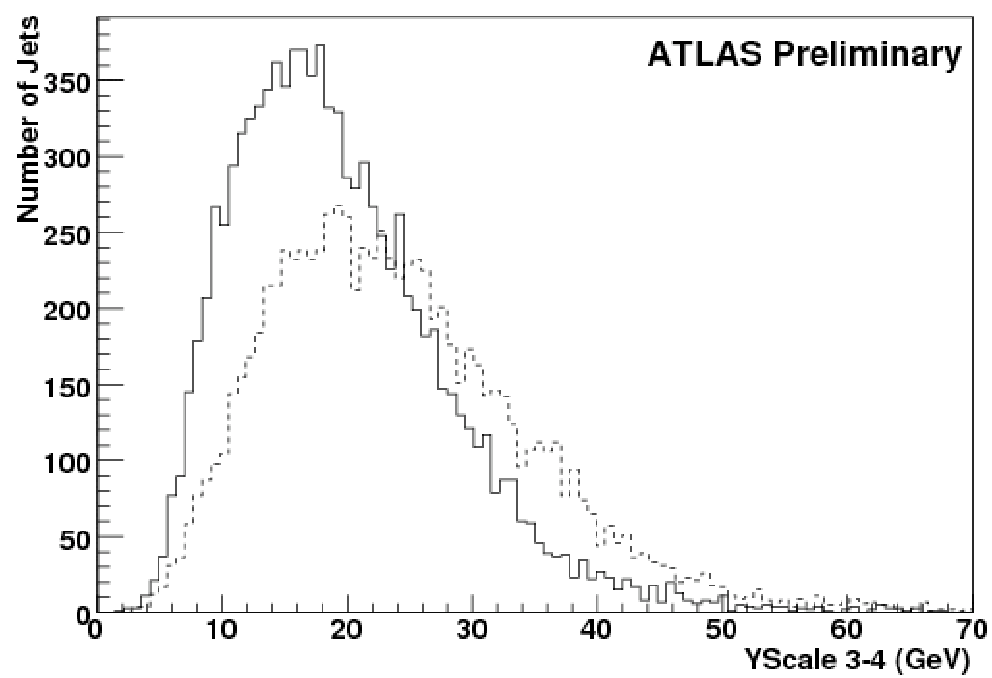
Jet Mass (GeV)
Jet pT (GeV)
ATLAS Preliminary



ATLAS Preliminary
M(Z’) = 2 TeV
M(Z’) = 3 TeV
Number of Jets
YScale 1-2 (GeV)

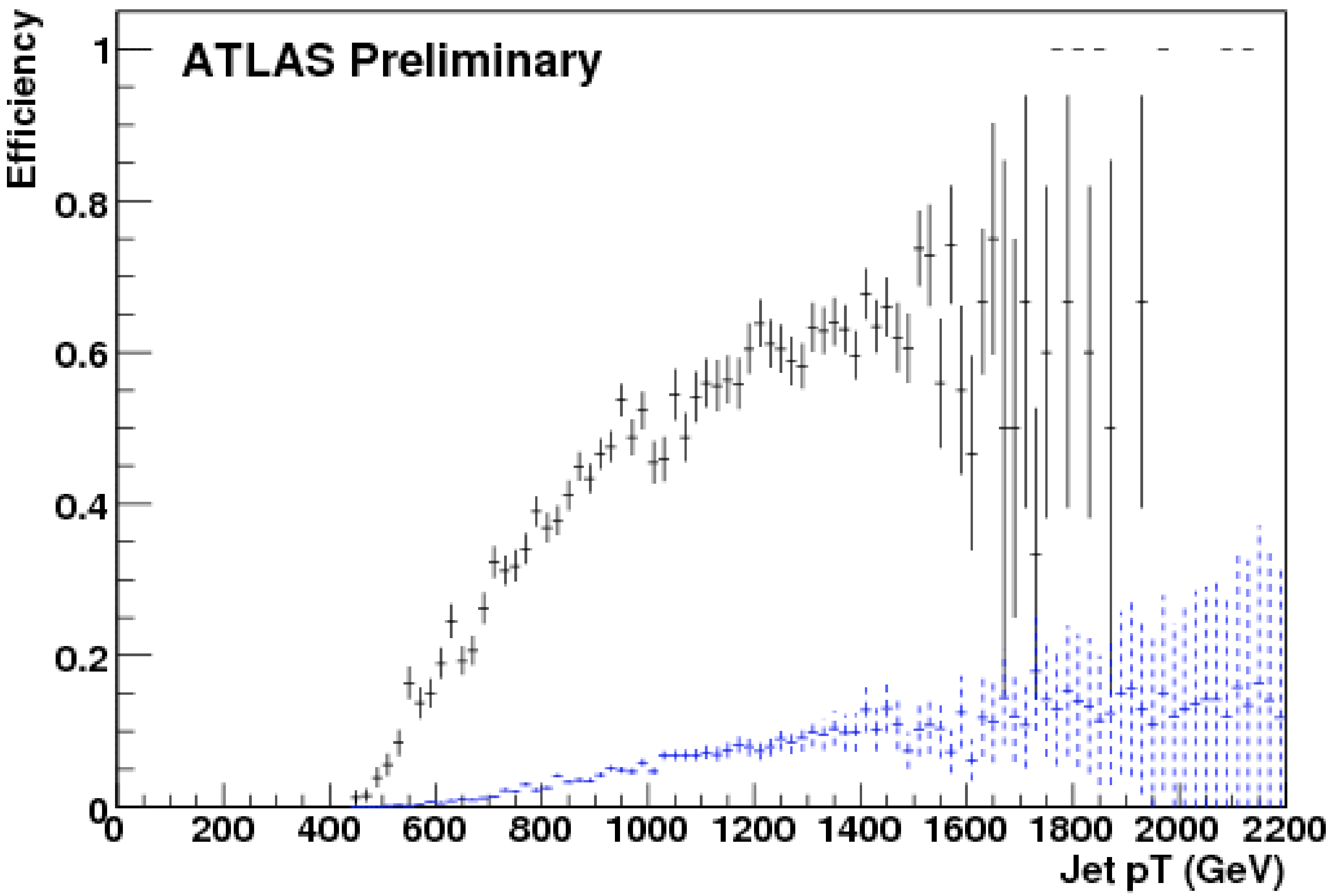


ATLAS Preliminary
Number of Jets
YScale 2-3 (GeV)



ATLAS Preliminary
Number of Jets
YScale 3-4 (GeV)

Result



Efficiencies

Jet p_T (GeV)	500	600	700	800	900	1000	1100	1200	1300	1400	1500
Top Samples (%)	5.6	19	32	37	47	45	56	64	63	68	74
Background Samples (%)	0.1	0.5	1.3	2.5	4.2	4.7	7.1	7.4	9.8	12.8	10.2

Global Fitters

- Test the overall event morphology against a signal hypothesis, yielding a figure-of-merit for each possible combination
  - Typically used when want to measure a property of the system, e.g. m_{tt}
  - Analysis can use best solution, or multiple solutions with weight related to figure of merit
- Crucial aspect is that top quark candidates <u>cannot</u> be considered individually
  - Any adjustment (like jet energy rescaling) changes figure of merit and other top candidate properties

χ^2

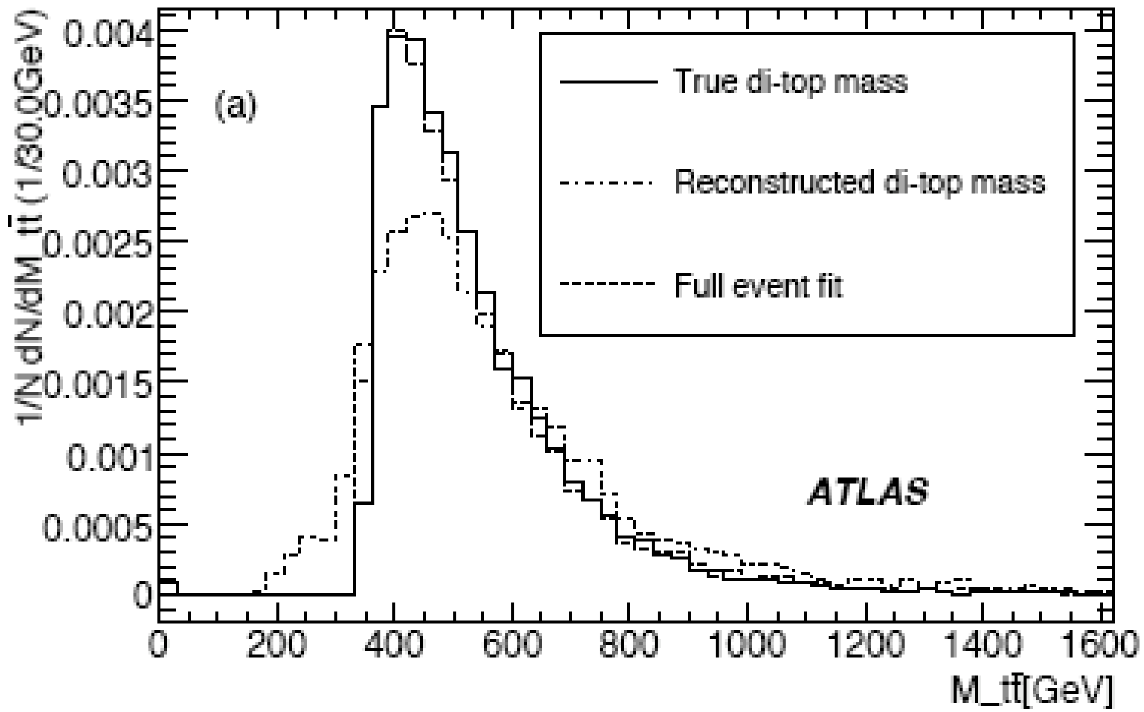
- "Classical" global fitter

$$\chi^2 = \sum_{\text{jets}}\left(\left(\frac{\eta_i^m - \eta_i^f}{\sigma_\eta^i}\right)^2 + \left(\frac{\phi_i^m - \phi_i^f}{\sigma_\phi^i}\right)^2\right) + \sum_{\text{jets,lepton}}\left(\frac{E_i^m - E_i^f}{\sigma_E^i}\right)^2 + \sum_{x,y,z}\left(\frac{p_{i\nu}^m - p_{i\nu}^f}{\sigma_{i\nu}}\right)^2$$
$$+\left(\frac{m_{jj} - M_W^{PDG}}{\sigma_W}\right)^2 + \left(\frac{m_{l\nu} - M_W^{PDG}}{\sigma_W}\right)^2 + \left(\frac{m_{jjb_h} - m_{top}^{fit}}{\sigma_t}\right)^2 + \left(\frac{m_{l\nu b_l} - m_{top}^{fit}}{\sigma_t}\right)^2.$$

  - But multiple approaches, depending if allow to rescale some measurements
    - And what the purpose is!

$$\chi^2 = \sum_{i=1}^{4}\left(\frac{\alpha_i E_i - E_i}{\sigma_{\text{jets}}(\alpha_i E_i)}\right)^2 + \left(\frac{\lambda \not{E}_T - \not{E}_T}{\sigma_{\not{E}_T}(\lambda \not{E}_T)}\right)^2 + \sum_{\text{type=lep,had}}\left(\frac{M_W^{\text{type}} - M_W^0}{\Gamma_W}\right)^2 + \sum_{\text{type=lep,had}}\left(\frac{M_{\text{top}}^{\text{type}} - M_{\text{top}}^0}{\Gamma_{\text{top}}}\right)^2$$

- Example: m_{tt}
  - “Reconstructed” = “simple” reconstruction
  - Full event fit is χ^2-based



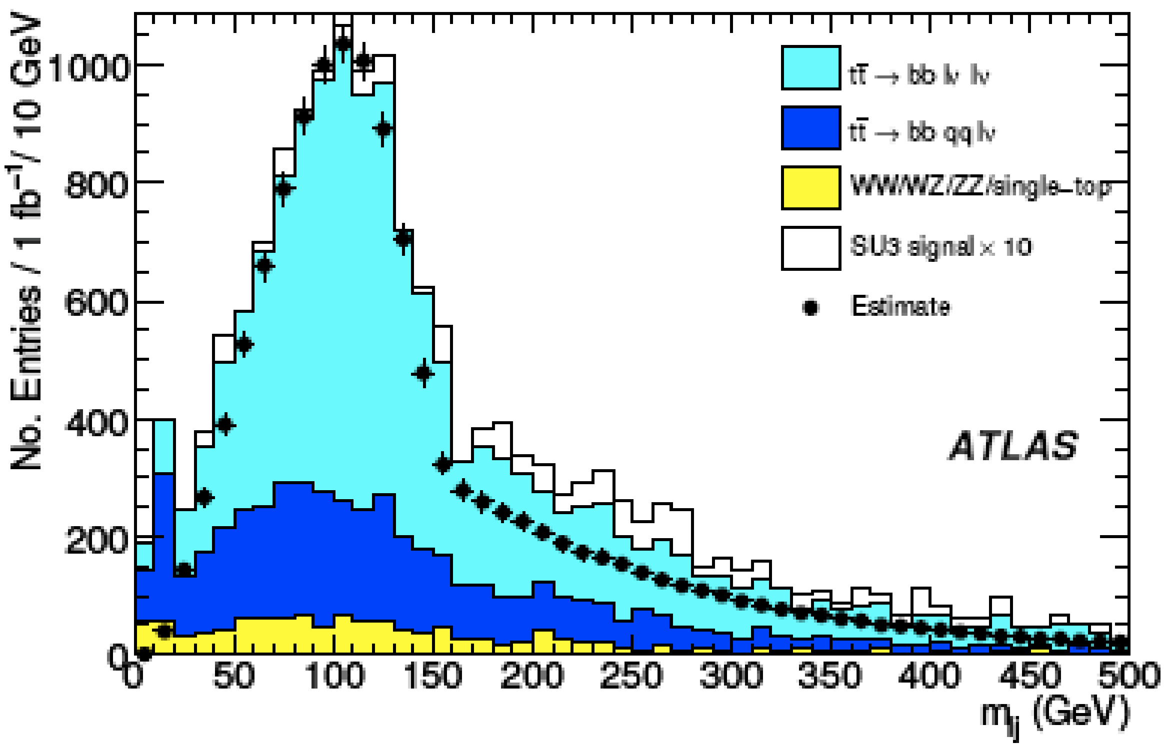
- Other examples: polarization measurements, rare decays, ...

- χ^2 fits assume gaussian probability density functions
  - Often have tails in resolution functions
    - Can be reduced by tight acceptance criteria, at the expense of efficiency
  - With good detector, get improvement by constraining W and top to Breit-Wigner rather than gaussian
- Alternative solution is to use likelihood approach
  - Signals & resolutions can be represented by appropriate *pdf's*
  - Under development...
- Or, use matrix element approach...

Endpoints

- Studies of SUSY signals (two escaping particles) have yielded many distributions with characteristic endpoints
  - No attempt to reconstruct individual decaying particles
  - In top dilepton events, have $m_{lj}^{\mathrm{max}} \simeq \sqrt{m^2(\mathrm{t}) - m^2(\mathrm{W})}$

~(155 GeV)



Points: events from "data" for which top quark kinematics were derived using m_W, m_{top} and MET constraints, then top quarks were "redecayed" with MC

⇒ eliminate many MC uncertainties (UE, extra jets, etc.)

- Many other similar variables
  - E.g. "contransverse mass" (arXiv:0802.2879)
  - For dilepton tt:

$$M_{CT}^{max} = \frac{m^2(\mathrm{t}) - m^2(\mathrm{W}) + m^2(\mathrm{b})}{m(\mathrm{t})}$$

Conclusions

- Many different top reconstruction techniques
  - From the simplest event selection to very high tech fitting methods
  - Tailored to specific needs:
    - Calibration
    - Background
    - Top physics
  - Provide necessary redundancy for complex final state
    - Sensitive to many "features" of detector and reconstruction
- Eager to test on data!