VBF H->bb

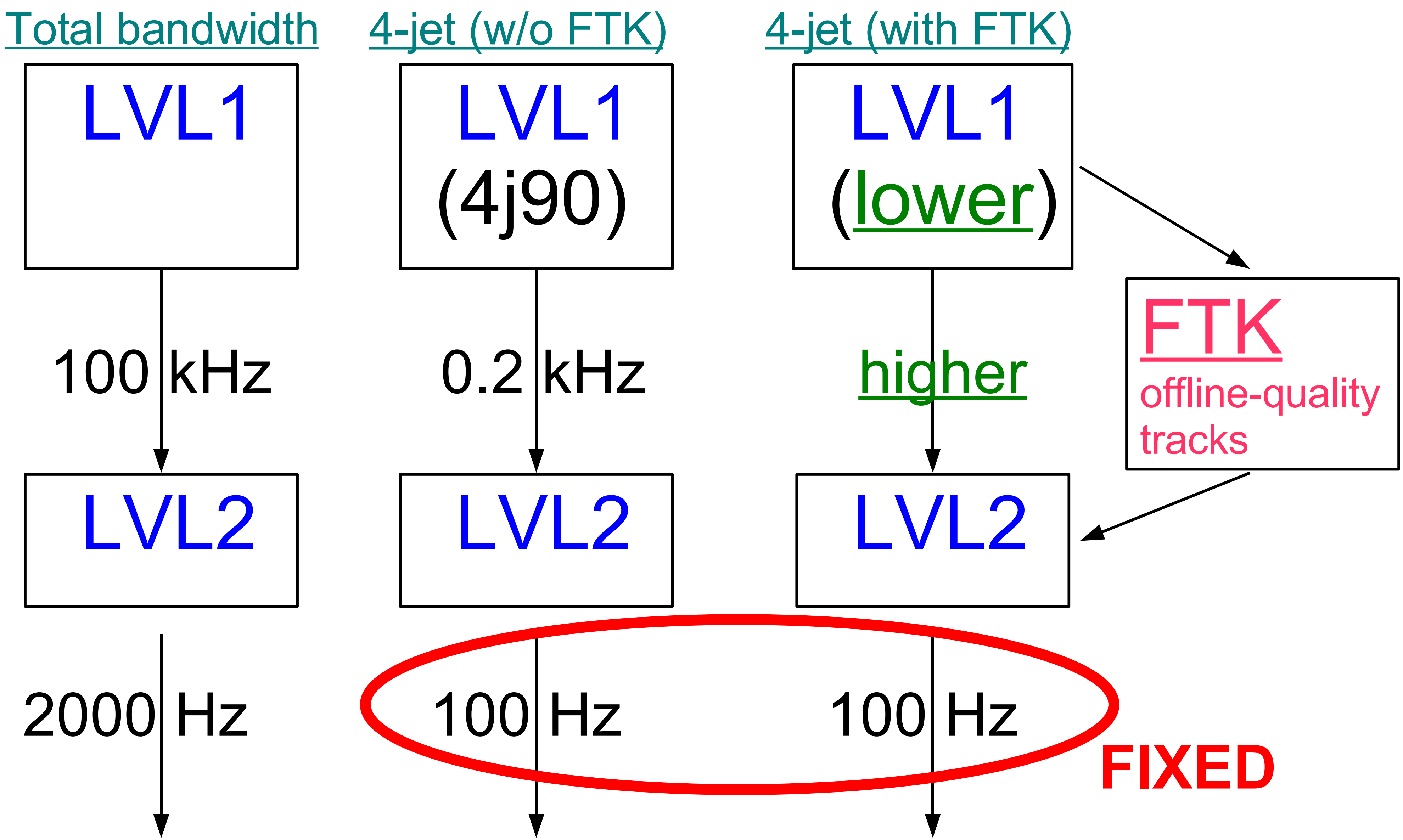
Anton Kapliy
(with help from Mel, Erik, Kohei)
University of Chicago

ATLAS meeting
Monday, November 21

Plan

- Motivation
- Physics process
- Analysis details
- Results
- Things to be done

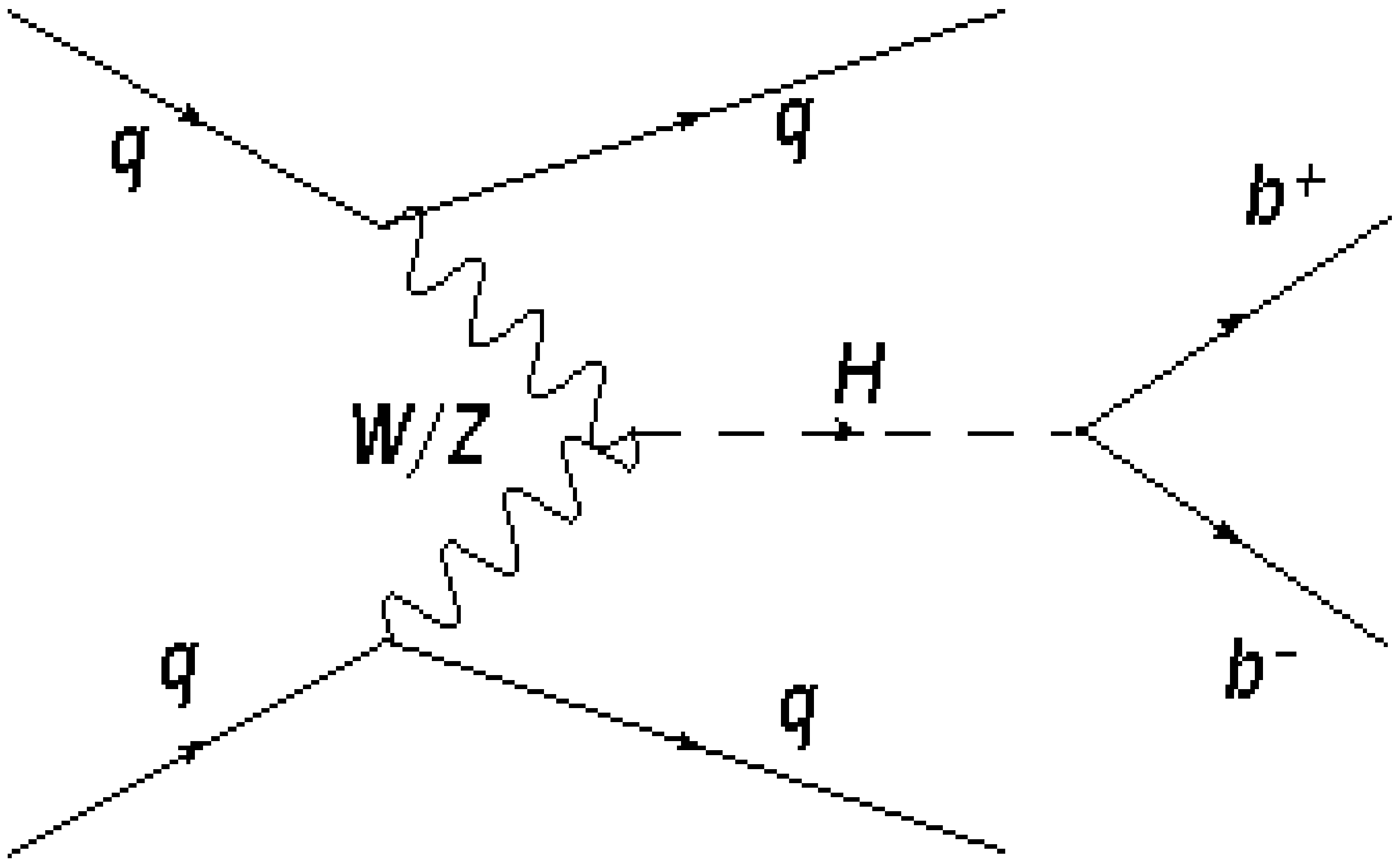
Motivation: FTK physics case



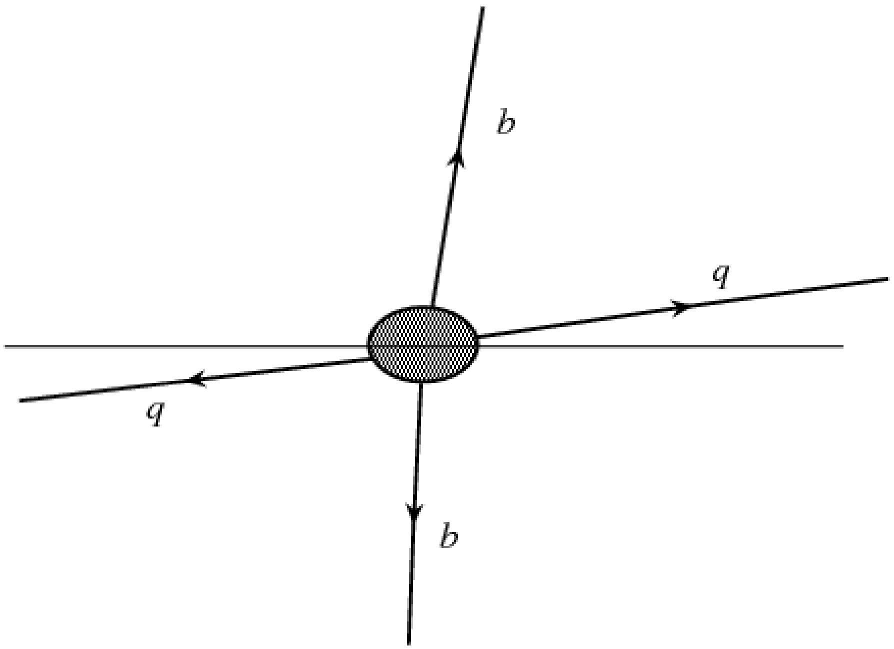
FTK { **improves b-tagging**
allows to <u>decrease</u> LVL1 thresholds

Physics process

- Standard Model Higgs decaying to b-bbar.
- Production mechanism: Vector Boson Fusion (VBF)



Process details



No color exchange between interacting partons ==>

- Two *b-jets* in the central Eta-region
- Two *forward jets*
- Little activity in-between

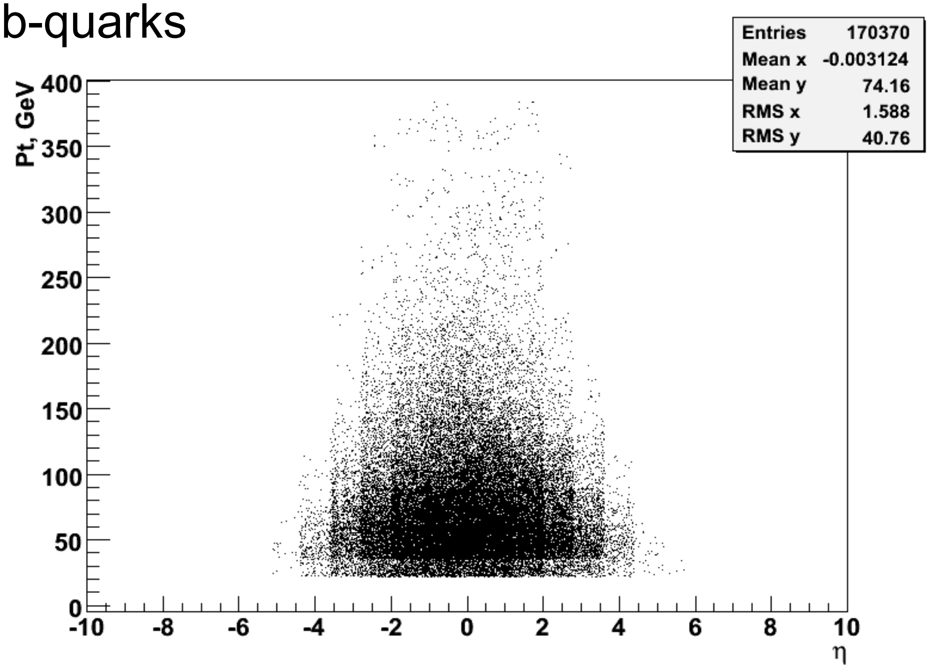
Mass range: 115 GeV to ~160 GeV (2x W mass)

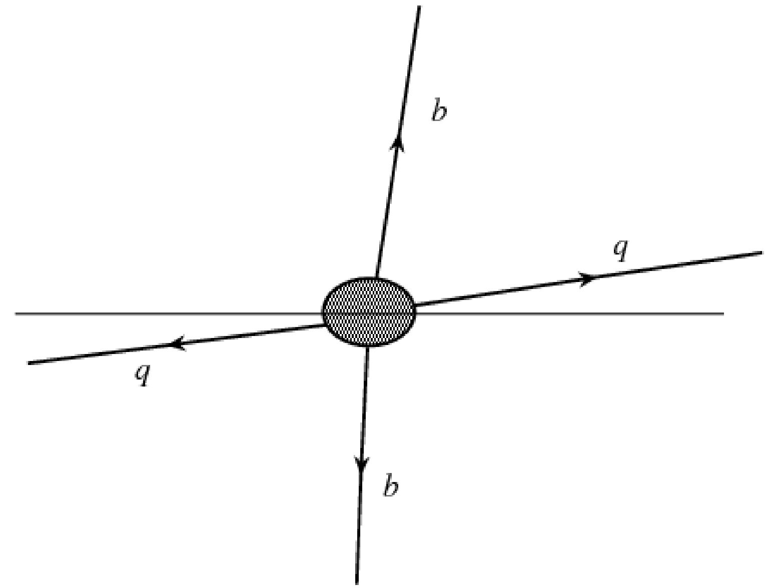
In this talk: **115 GeV SM Higgs**

Xsec = 3.15 pb (via Pythia 6.3) – <u>very</u> low

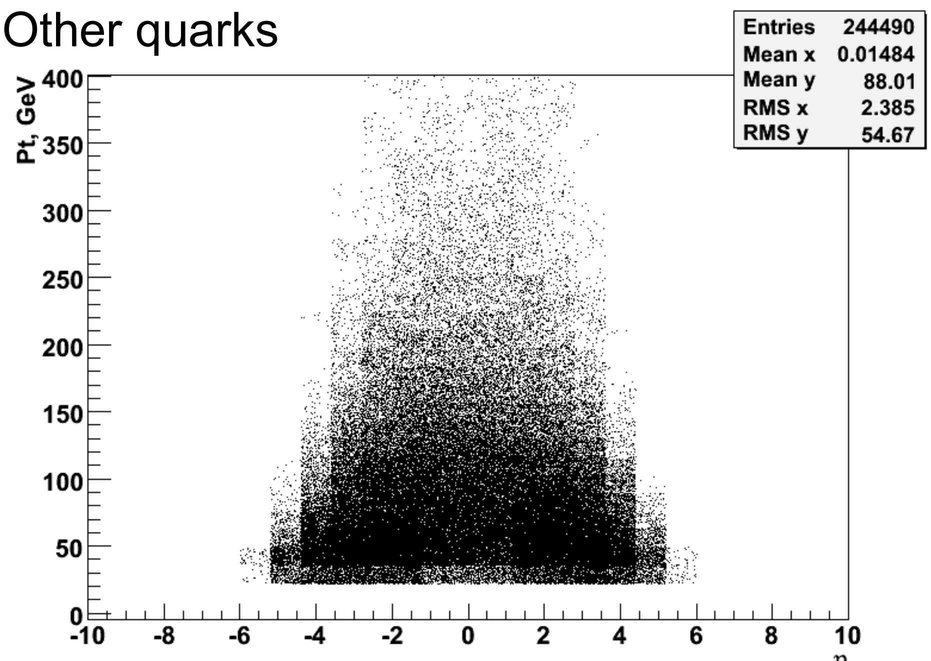
MC truth

b-quarks

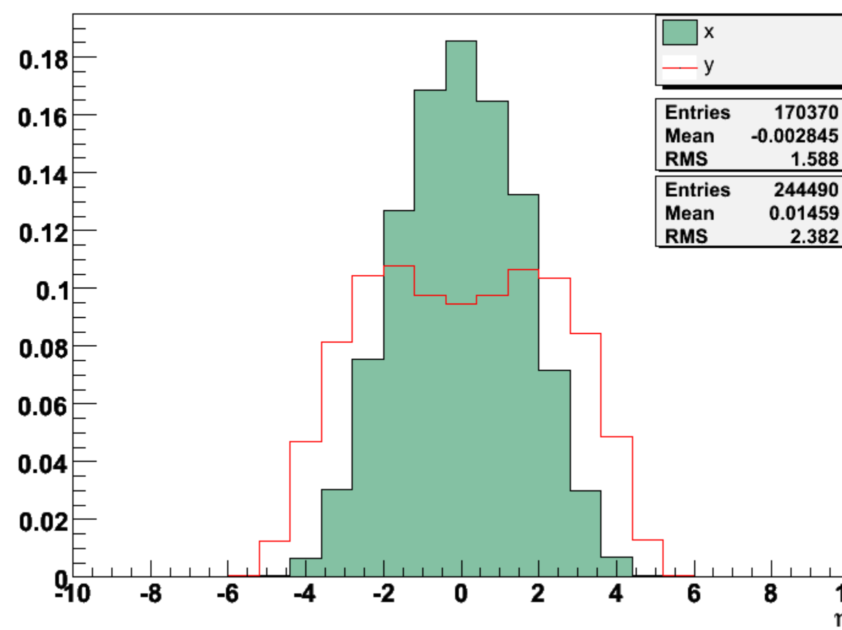




Other quarks



Eta of b-quarks (green) & other quarks (red)



Background

Generic QCD background: Erik's Sherpa samples.

Erik showed that 2->2 processes are negligible after trigger.

==> Using 2->3 only with yt = 25, 40, 50 GeV

yt = 25 GeV: Xsec = **3.2E7 pb**
yt = 40 GeV: Xsec = **3.8E6 pb**
yt = 50 GeV: Xsec = **1.3E6 pb**
for signal : Xsec = **3.15 pb**

Looks pretty bleak!

Generation & Simulation

Signal: 115 GeV VBF H->bbar
Generated in Athena 11.0.42
(Pythia 6.3)
Fast-simulated in Atlfast
(output to CBNT)

Background: generic QCD
Generated in Sherpa

Fast-simulated in Atlfast
(output to CBNT)

All analysis was done using AtlfastB kinematic variables:

- ✓ PTJETB/EEJETB 4-momenta (includes corrections)
- ✓ KFJET jet type (excludes b-tagging)

b-tagging

ATLFAST – primitive b-tagging capability (*KFJETB*):

KFJET:	Tag prob		#A	# < tag prob	tagged?	KFJETB:
u	0.01	Throws random #s →	0.41	0.41<0.01	NO	u
d	0.01		0.91	0.91<0.01	NO	d
c	0.1		0.05	0.05<0.1	**YES**	**b**
b	0.6		0.24	0.24<0.6	**YES**	**b**

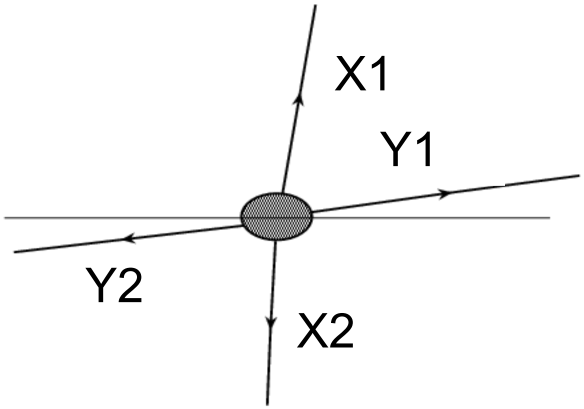
<u>Losing a lot of statistics!</u>

MANUAL B-TAGGING:

We can construct a <u>binomial decomposition</u> of all b-tagging possibilities and consider ALL of them in our analysis. We can gain a lot of statistics compared to ATLFAST b-tagging!

KFJET:	KFJETB:	KFJETB:	KFJETB:	KFJETB:		
q	q	q	q	b		
q	q	q	b	q	...	2^NJET terms
q	q	b	b	q		
b	b	b	b	q		
Weight:	0.99*0.99* 0.9*0.6	0.99*0.99* 0.1*0.6	0.99*0.01* 0.1*0.6	0.01*0.99* 0.9*0.4		} Adds up to 1.0

Choosing jets



For each event/tagging possibility:

- Find highest-Pt b-jet; call it X1
- Find second-highest-Pt b-jet; call it X2
- For Eta>0, find the highest-Pt jet; call it Y1
- For Eta<0, find the highest-Pt jet; call it Y2

If two X's (b-jets) and two Y's (forward jets) found, proceed with analysis; otherwise **fail the event**.

Kinematic cuts - summary

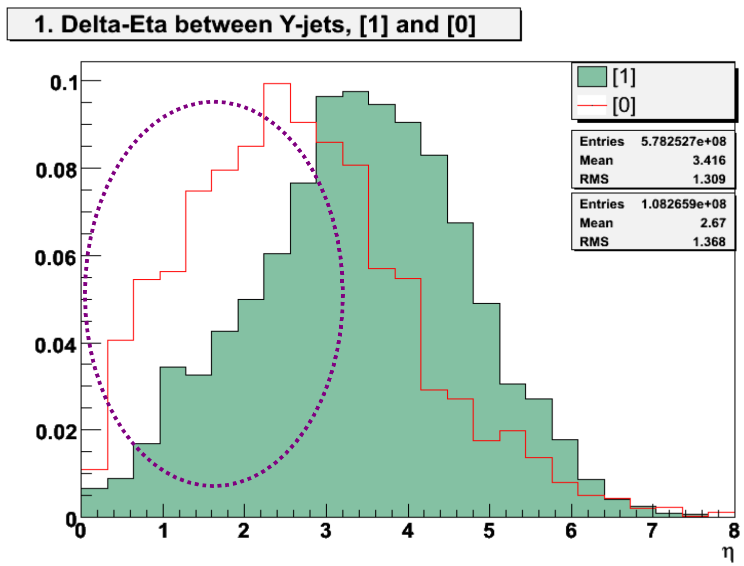
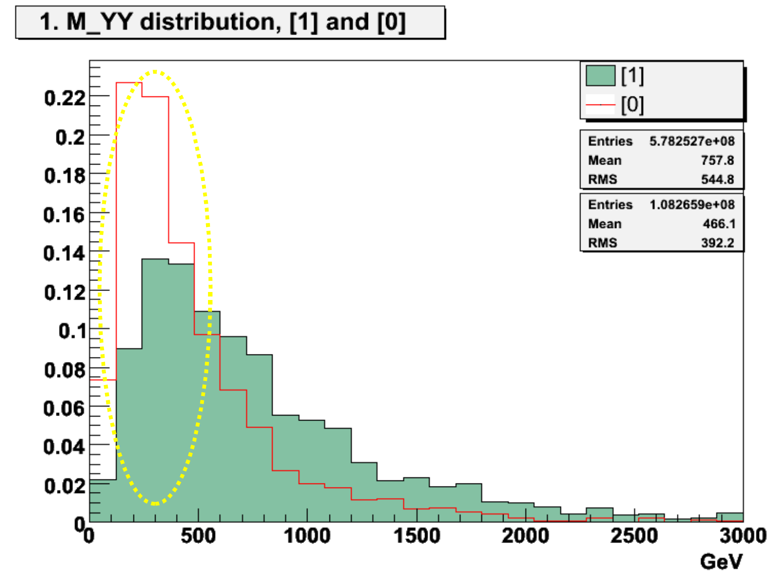
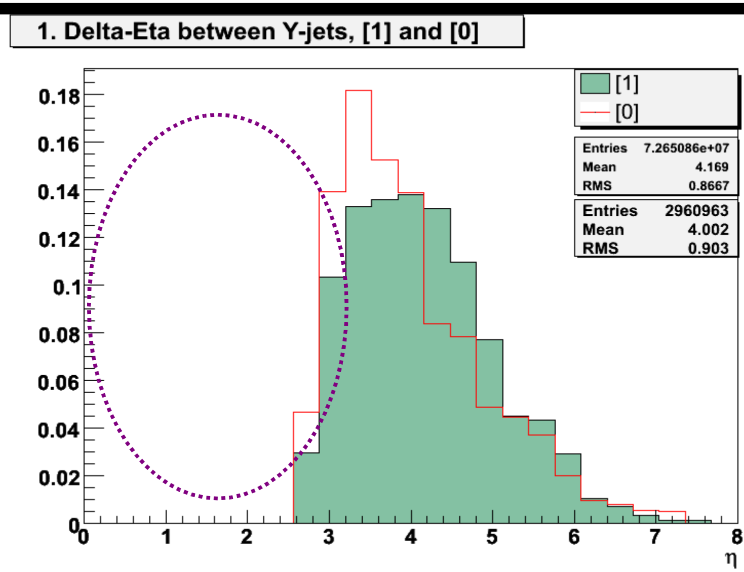
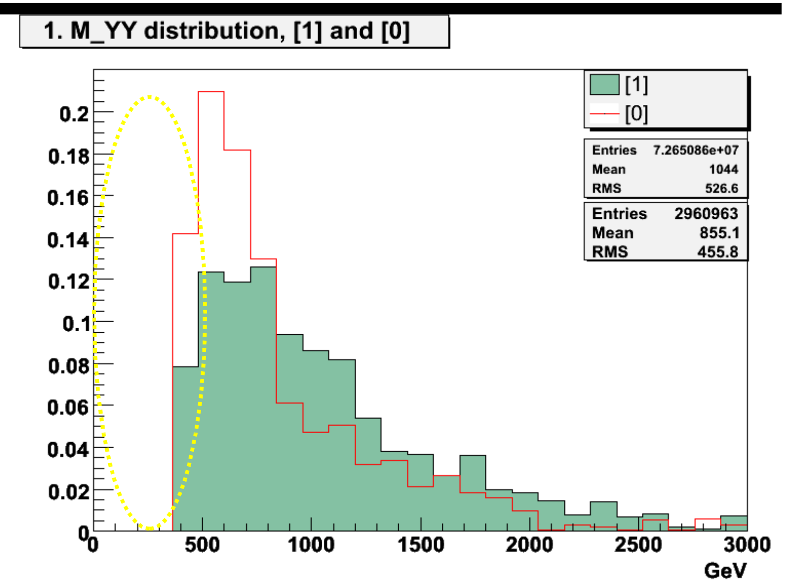
A variety of Eta, Pt, and Veto cuts were applied depending on SIG-vs-BG distributions.

For instance, for yt=25 Gev sample, 4j40, I used:

- *Pt of X-jets > 30 GeV*
- *Pt of Y-jets > 40 GeV*
- *|Eta[Y1]-Eta[Y2]| > 2.8*
- *Myy (invariant mass of Y-jets) > 400 GeV*

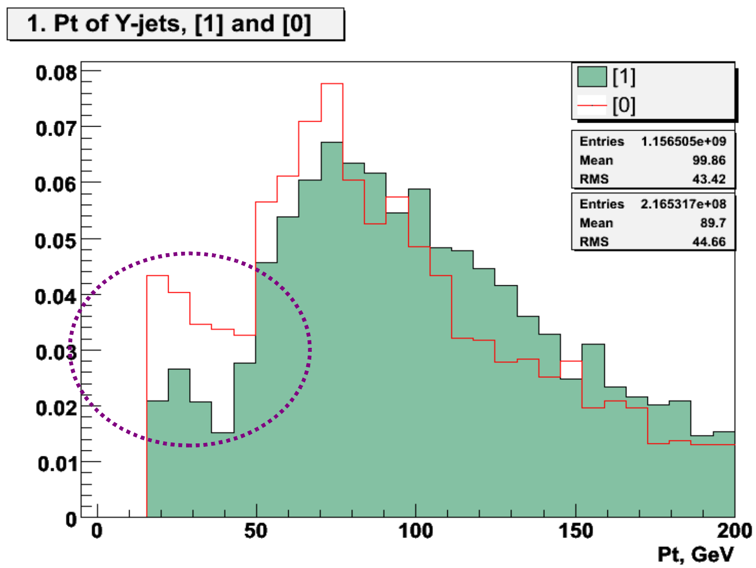
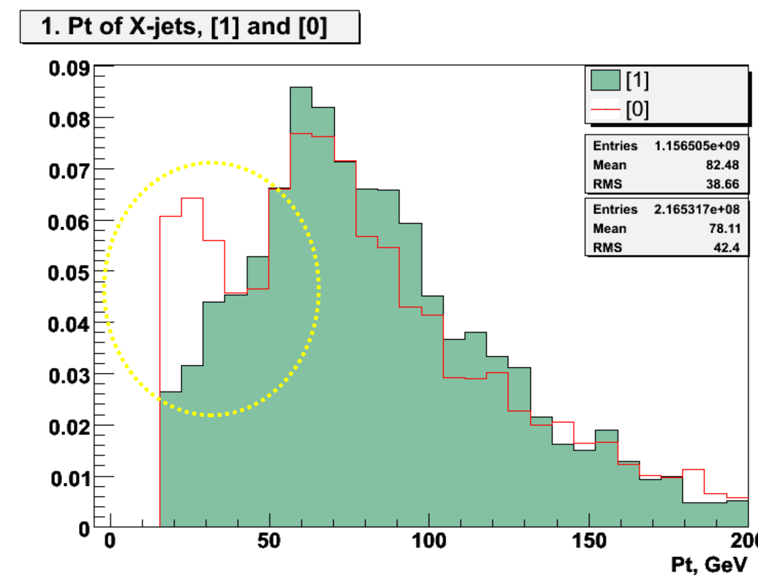
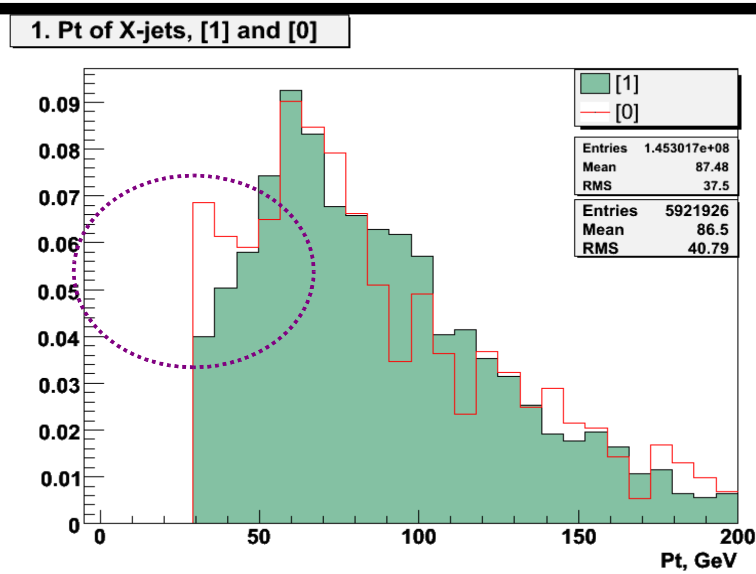
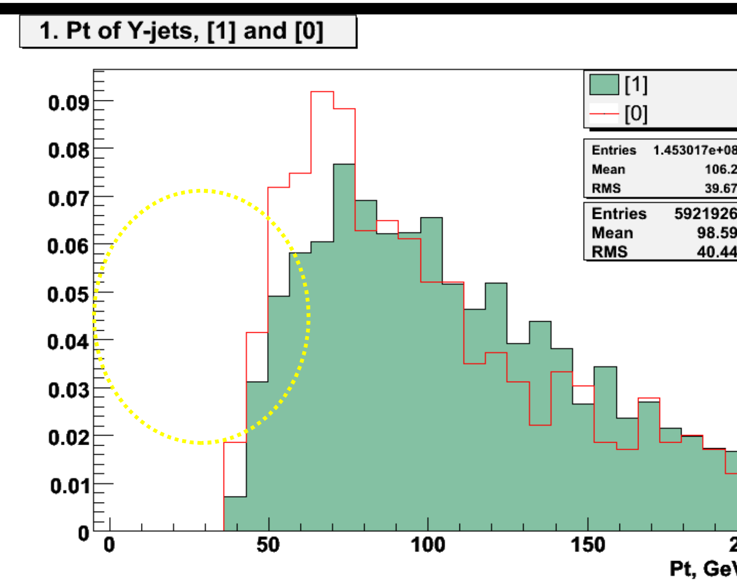
deltaEta>2.8; Myy>400GeV

Blue=SIG
Red=BG

	deltaEta of Y-jets	Myy of Y-jets
Before:		
After:		

Pt[X]>30GeV; Pt[Y]>40GeV

Blue=SIG
Red=BG

	Pt of X-jets (aka b-jets)	Pt of Y-jets (aka forward jets)
Before:		
After:		

LVL1 Trigger

Erik's parametrizations:
Atlfast Jet Pt versus Full-Simulated 8x8 CaloCluster Et.

Kohei's studies of trigger rates:
With FTK, we can lower the 4JXX trigger by 20-30 GeV.

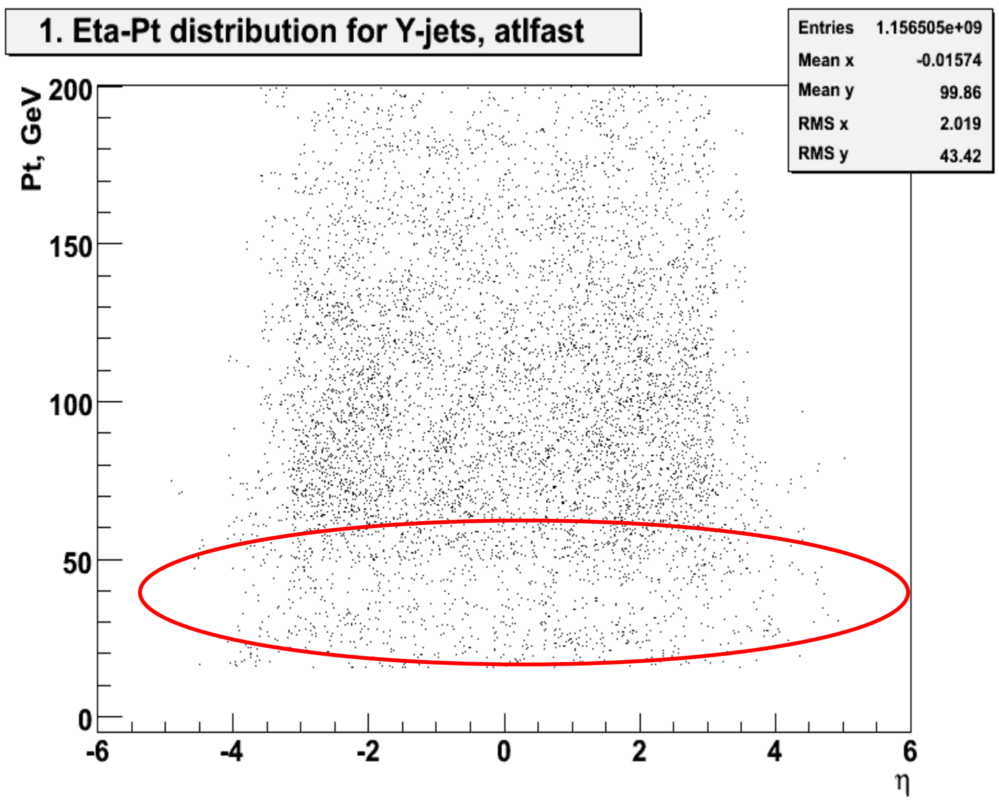
In this talk, the following triggers are considered:

4j40, 4j50, 4j60 ← **8x8 CaloCluster Et**

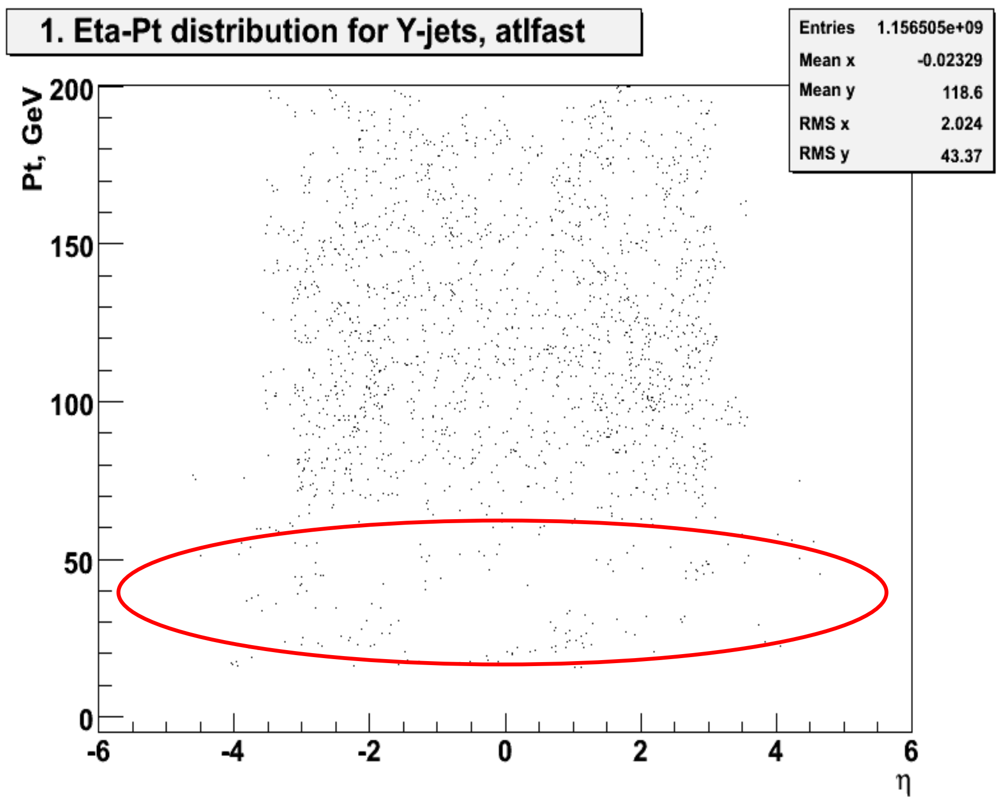
Each event is accepted with weight given by probability to pass a trigger.

4j40 vs 4j60

Forward jets; 4j40



Forward jets; 4j60

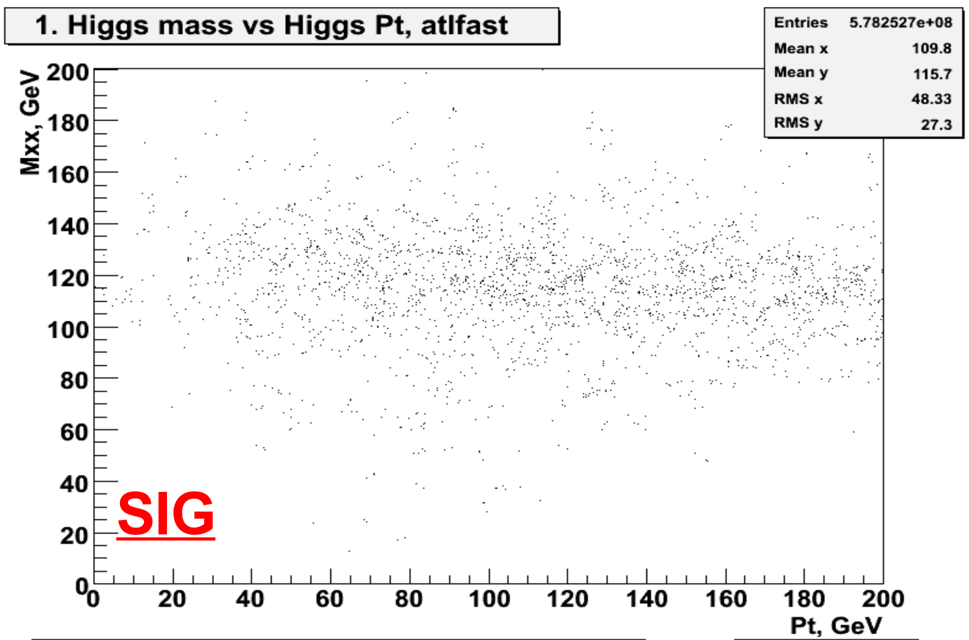


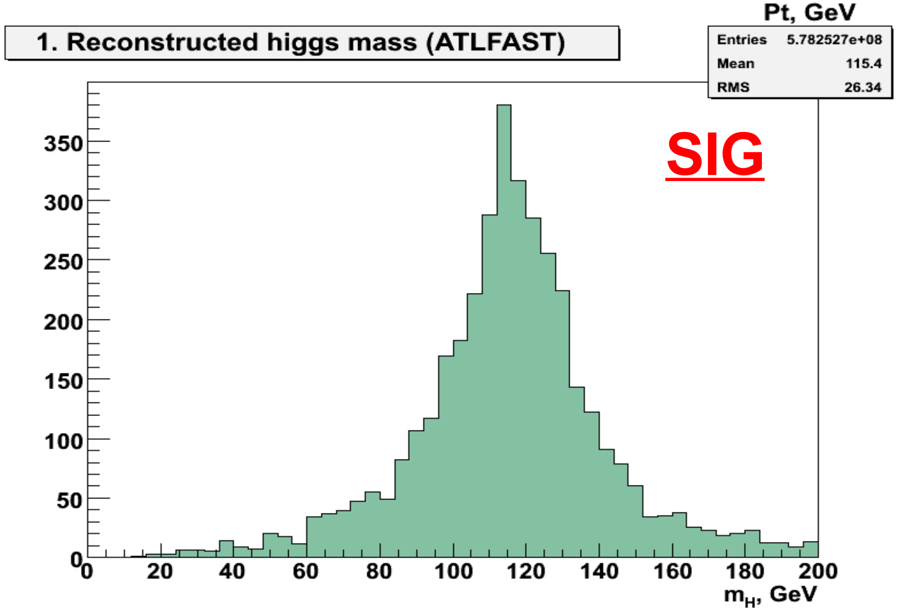
The region below Pt=60 GeV got almost empty once trigger was increased to 60 GeV.

Why are there some jets left below 60 GeV? That's because the trigger threshold = 8x8 CaloCluster Et, while the vertical scale is Atlfast Pt. The two are related by Erik's parametrization.

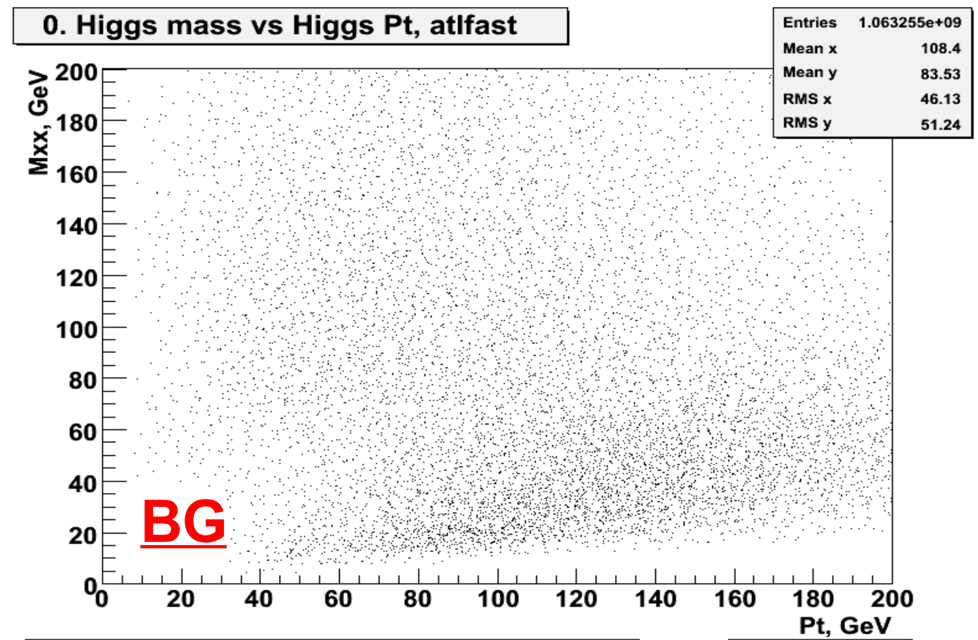
Mbb (115 GeV); yt=50; 4j40

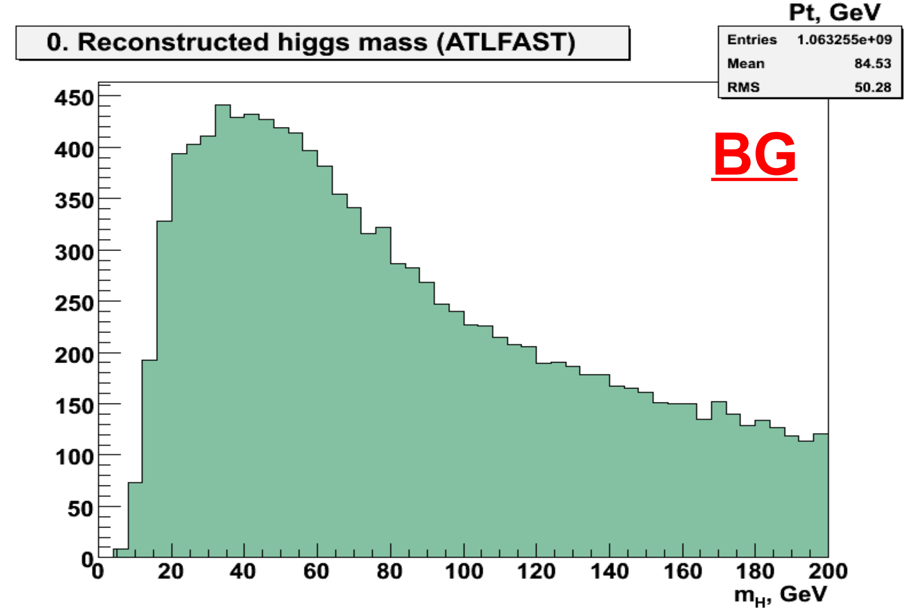
Reconstructed Higgs Mass-vs-Pt (SIG) (before kinematic cuts)





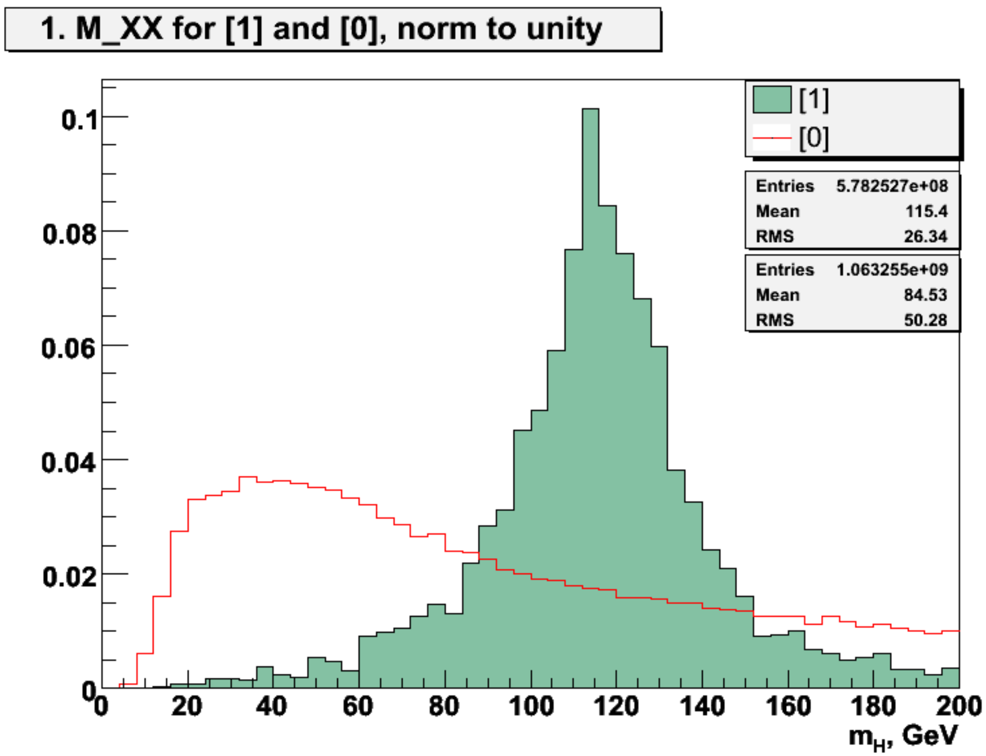
Reconstructed Higgs Mass-vs-Pt (BG) (before kinematic cuts)





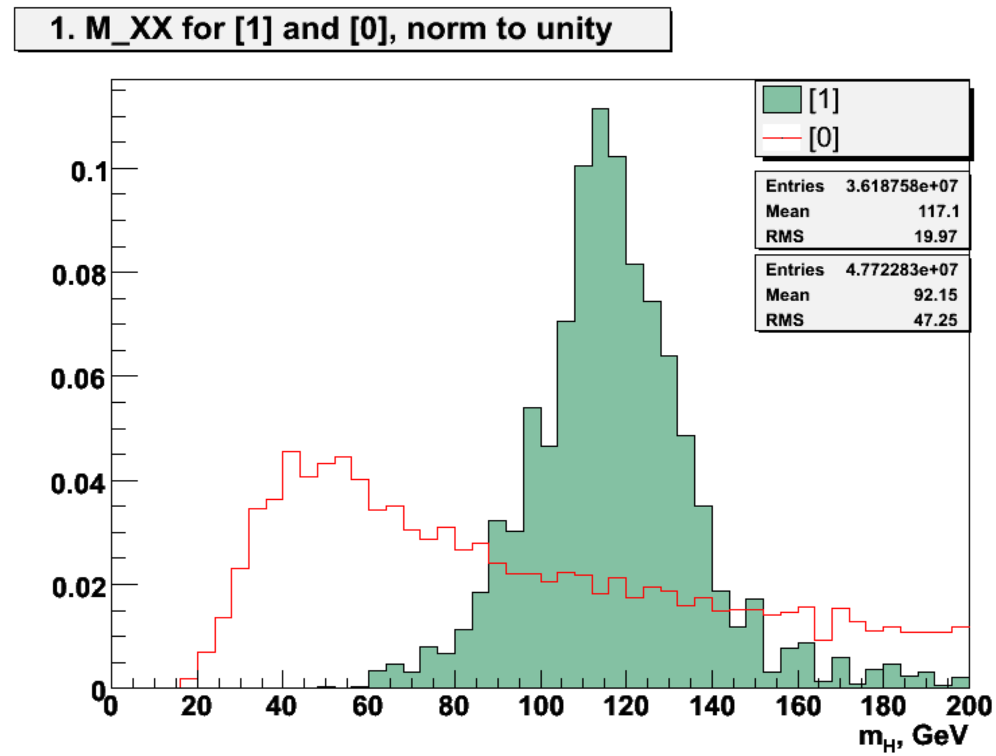
Mbb; yt=50; 4j40

Before kinematic cuts:



~4k SIG events
~16k BG events

After kinematic cuts:



~1.5k SIG events
~2k BG events

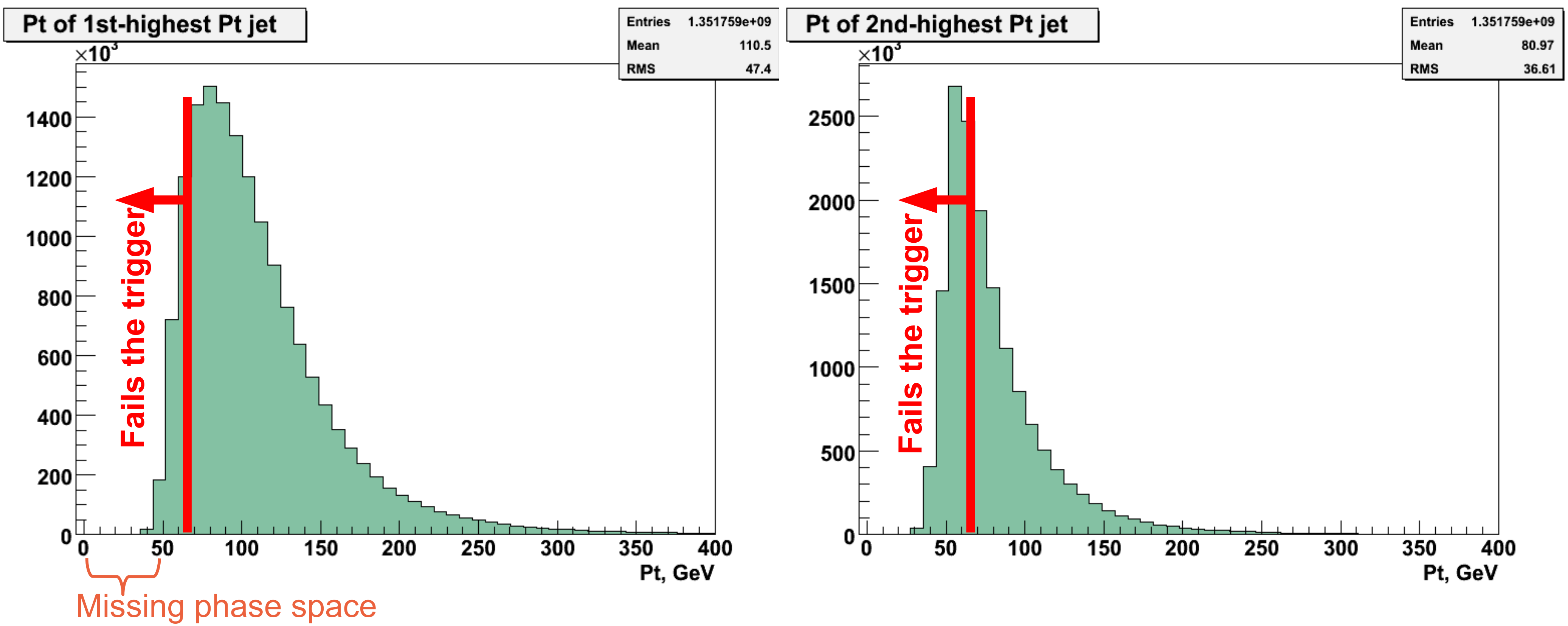
- Signal peaks close to 115 GeV
- Enough statistics to see the background shape (~falling exponent)

Mbb; yt=50; 4j40 – jet Pt's

yt=50 GeV sets the Pt scale for the jets: Pt > 50 GeV
But we require the energy deposition of 40 GeV in 8x8 Calo cells
Do we cover all of the phase space? - the answer appears to be Yes

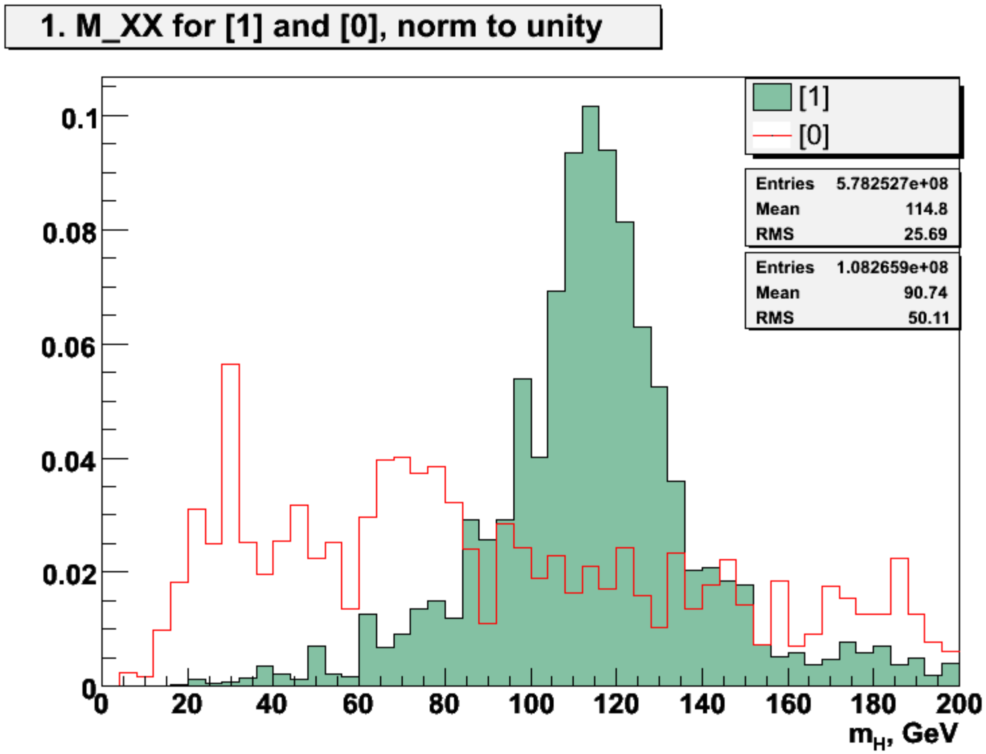
40 GeV 8x8 Et <--> **60-70 GeV** Atlfast Pt
(via Erik's and Kohei's studies)

Transverse momentum of the 1st and 2nd highest-Pt jets in the sherpa yt=50 sample



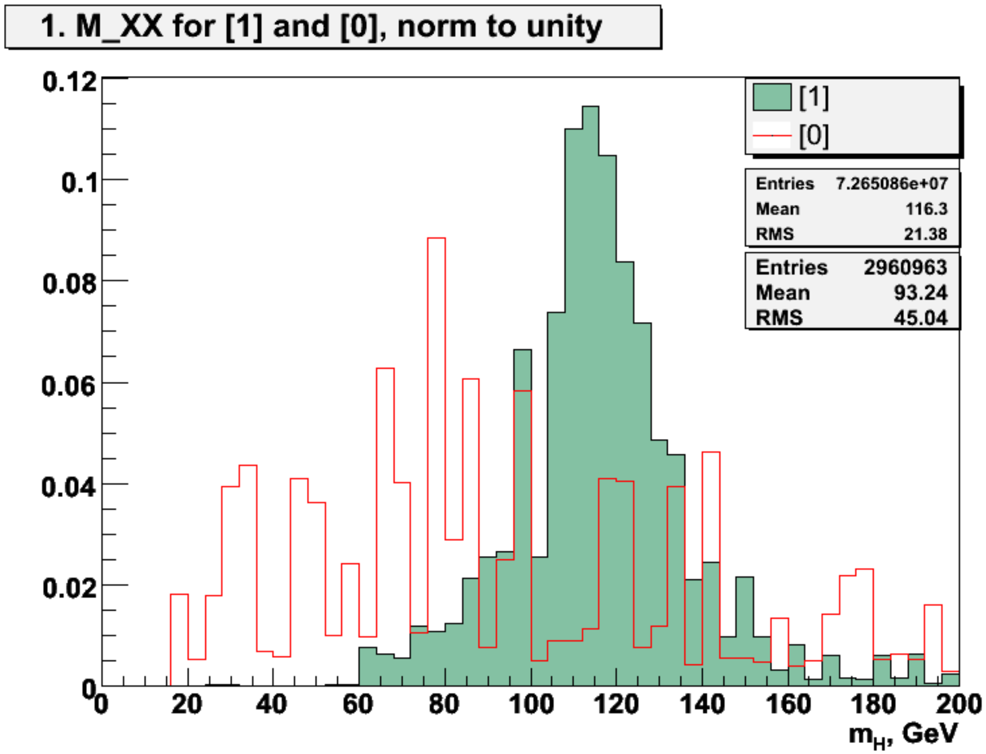
Mxx (b-bbar); yt=25; 4j60

Before kinematic cuts:



~1482 SIG events
~73 BG events

After kinematic cuts:



~821 SIG events
~16 BG events

- Signal still peaks close to 115 GeV
- **Not enough statistics** for BG (however, I've shown the worst case)

Significance

Assuming 300 fb^-1 of LHC data (~3 years at full luminosity)

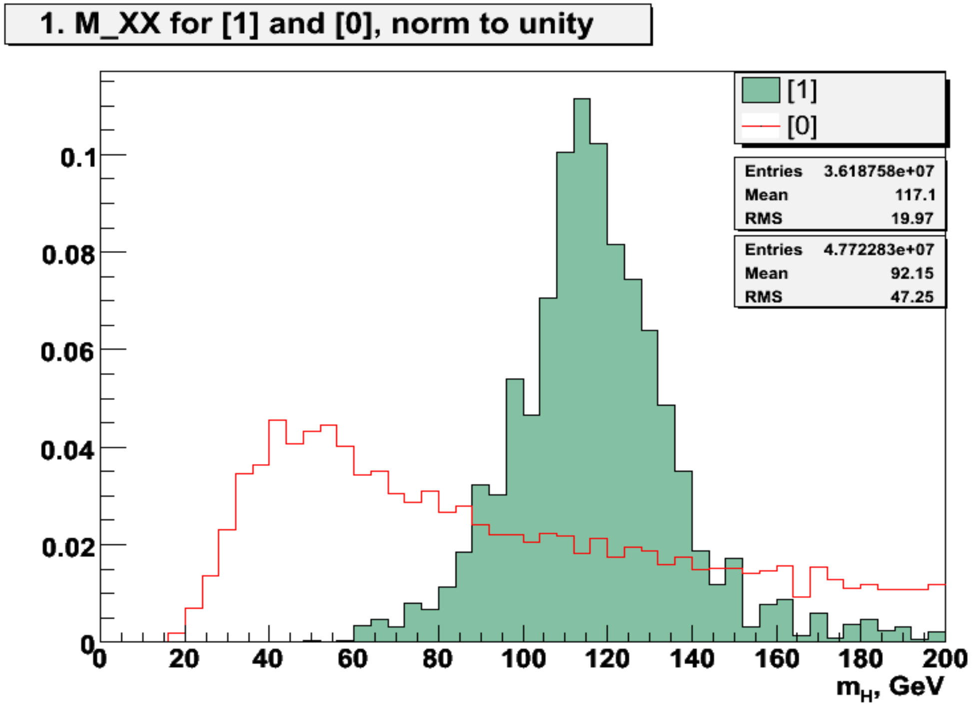
1st try:
#SIG/sqrt(#BG) = 2 to 2.5 (depending on trigger)
=> Looks hopeless as a discovery channel!

As a precision measurement channel:
We can exploit our approximate knowledge of Higgs Mass.

2nd try:
#SIG[80..140]/
sqrt(#BG[80..140])

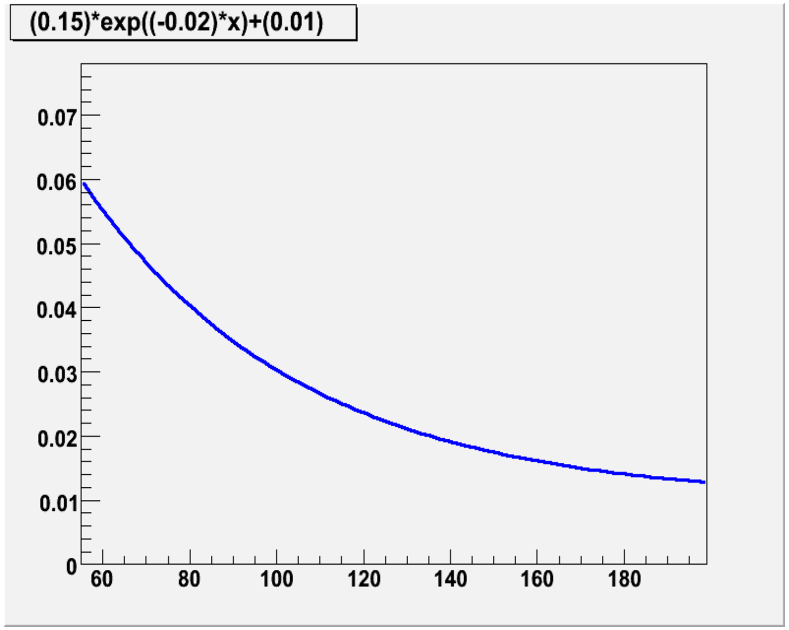
- nominator stays nearly the same;
denominator reduces significantly
=> significance improves

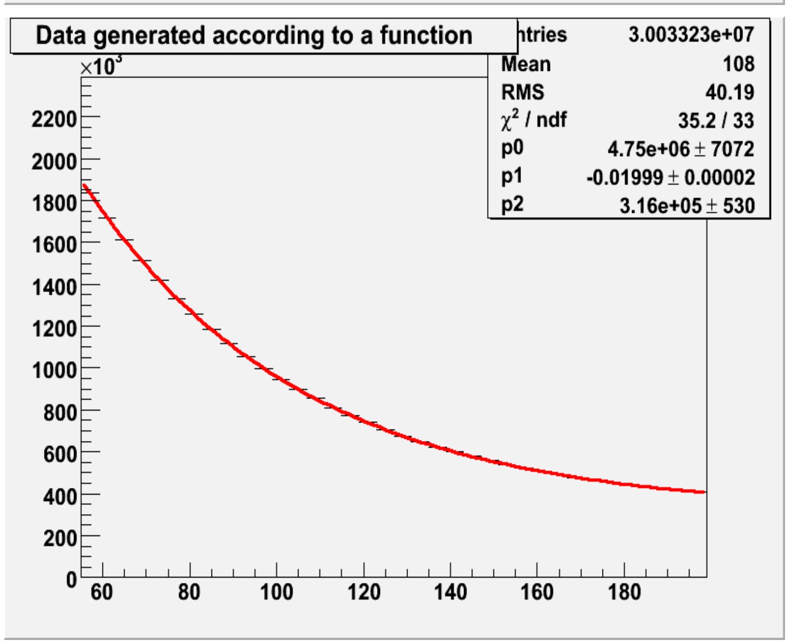
Doesn't take into account quality of BG fit!



Significance (improved)

3rd try:





1. Get enough BG data to deduce its shape. Suppose it follows a falling exponential between 55..200 GeV.

2. Generate random events whose distribution function is given by the above function. #events generated = #BG events expected after 300 fb^-1 of LHC data (calculated using pass prob and Xsection).

3. Fit the histogram to a given functional shape and obtain errors on fit parameters [red curve in the plot].

4. **Propagate these errors** to the error on the integral under the fitted curve over the region where we expect higgs signal (e.g 80..135). This value will be a denominator in #SIG[80..135]/delta{BG[80..135]}

In the plots presented: yt=50 GeV; trig=4j40
After 300fb^-1:
#BG[0..200]=33M BG; #BG[55..200)=30M
sqrt(#BG[55..200]} = 6330; delta{#BG[55..200]}=6329
sqrt(#BG[80..135]} = 2830; delta{#BG[80..135]}=3200

Red line shows that over the whole range, this algorithm just gives delta{#BG}~sqrt{#BG}

Significance - results

Significance as a function of LVL1 (8x8) threshold (4jXX):

		4j60	4j50	4j40
yt=25	#SIG/sqrt(#BG)	2.0+/-0.3	2.2+/-0.3	2.4+/-0.3
	#SIG/delta(#BG)	**2.7+/-0.5**	**3.2+/-0.5**	**3.6+/-0.5**
yt=40	#SIG/sqrt(#BG)	2.1+/-0.3	2.3+/-0.3	2.5+/-0.3
	#SIG/delta(#BG)	**2.9+/-0.5**	**3.2+/-0.4**	**3.8+/-0.5**
yt=50	#SIG/sqrt(#BG)	1.9+/-0.2	2.1+/-0.2	2.3+/-0.3
	#SIG/delta(#BG)	**2.7+/-0.5**	**2.9+/-0.3**	**3.5+/-0.4**

Summary:

- Results independent of yt cut => GOOD!
- 30% improvement in significance
- Significance is still less than 5 :-(

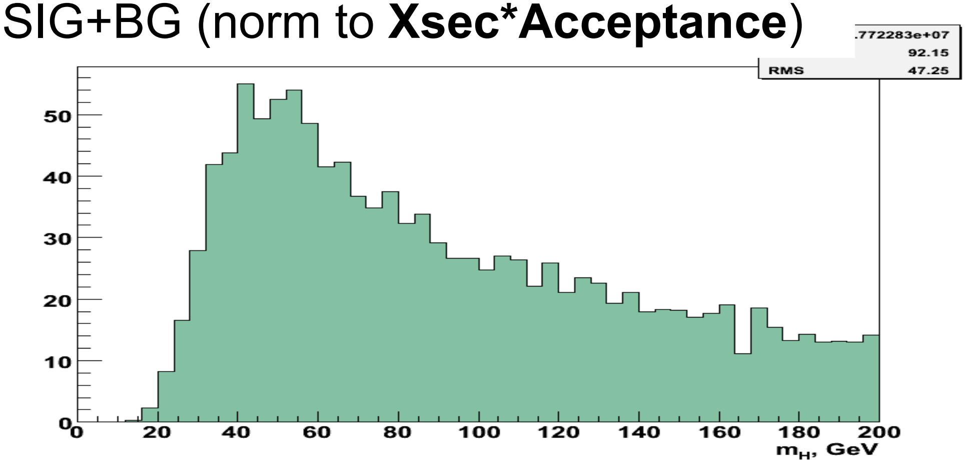
BUT

Not a discovery channel -> *significance not important*

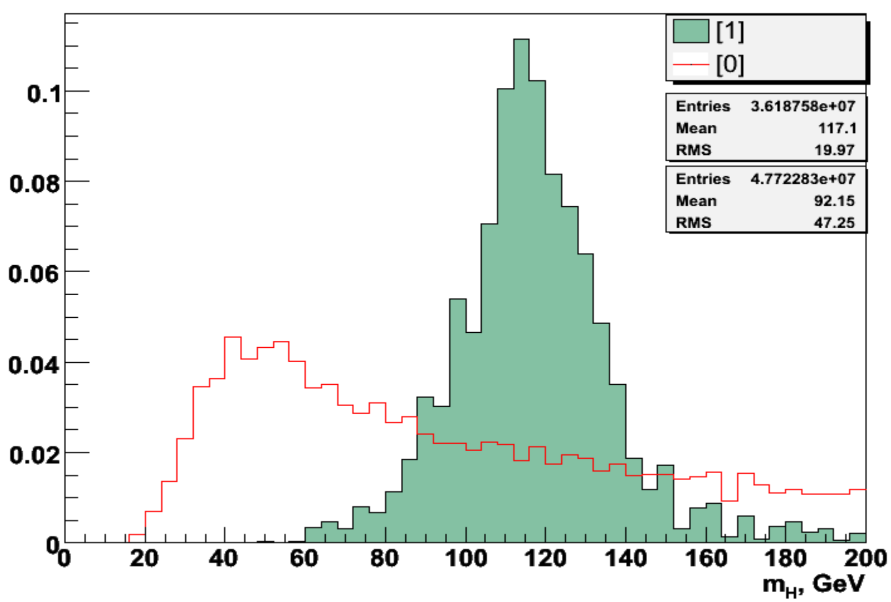
Interested in: **coupling of Higgs to b-quarks**

Coupling strenth - method

SIG+BG (norm to **Xsec*Acceptance**)



SIG,BG (norm to **unity**)



In real life:

1. From another channel (e.g., H->gamgam), we will get higgs mass and width.

2. Fit the data to a function (e.g., falling expo) between 50 GeV and 200 GeV. Exclude the expected higgs-region from the fit.

3. Interpolate the fit to the expected-higgs region and predict #BG.

4. Subtract #BG from #SIG+#BG to get #SIG

In this study:

1. The signal is around 80..135 GeV. Get #SIG and #BG from Atlfast plots (on the right).

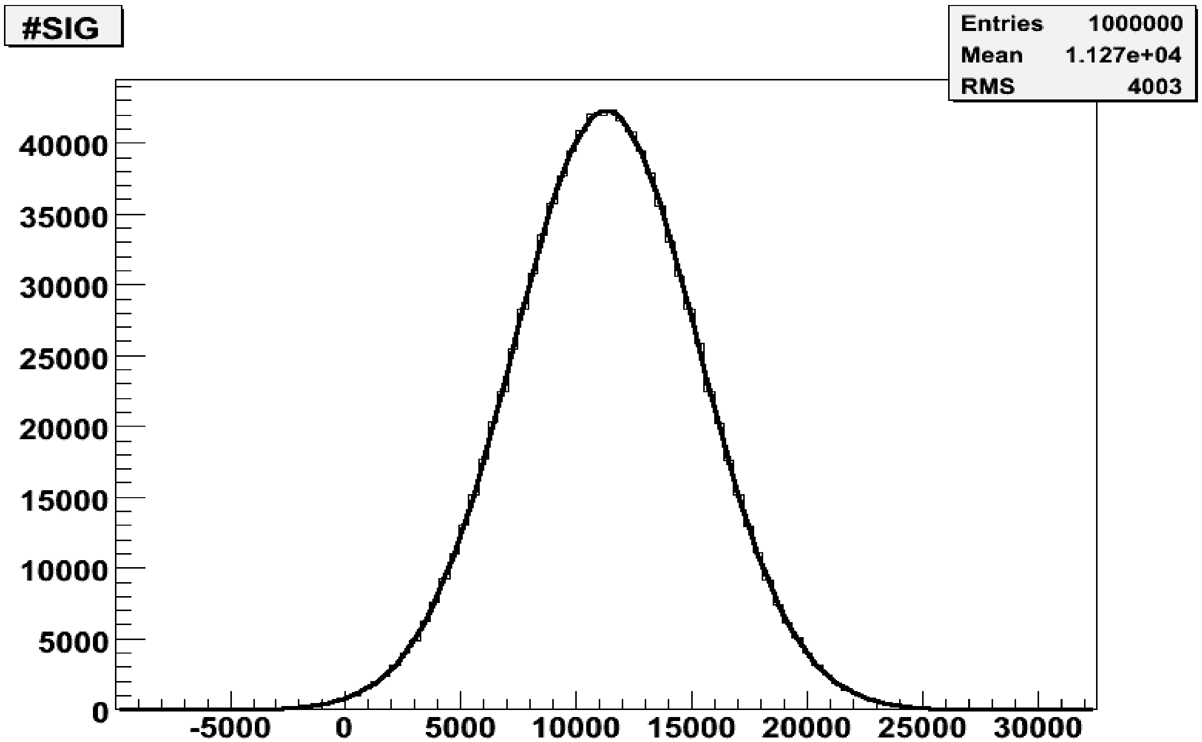
2. Let #BG and #TOT=(#SIG+#BG) be gaussian-distributed vars with sigma=sqrt.

3. Run a toy-MC that gets #BG and #TOT from the gaussian and computes #SIG.

4. Fit the result to a gaussian; get sigma of #SIG.

Coupling strenth - results

#SIG=(#TOT-#BG)
<u>1M MC trials</u>
(4j40)



In 80..135 GeV

Trigger	Via Atlfast reconstructio		#SIG+#BG	Monte-Carlo results		d(#SIG)/#SIG
	#SIG	#BG	#TOT	MC #SIG	MC d(#SIG)	% error
4j60	4412	2019241	2023653	4410	2011	**46%**
4j50	7175	4247721	4254896	7172	2916	**41%**
4j40	11271	8014561	8025832	11267	4004	**36%**

- Improved our knowledge of coupling by 10%
- However, we are still talking about ~40% uncertainty!

Problems and TODO

- Ignoring pile-up; using 300 fb^-1 and assuming high luminosity – don't really know acceptable LVL1 thresholds.
- Multi-threshold LVL1 – possibility of improvement (Kohei's work)
- Large Eta problem – FIXME.
- In significance estimation, do function fit to a combined SIG+BG
- Explicit reconstruction of SIG out of SIG+BG combined histo
- Understand how we'd translate #SIG events expected to the coupling strength (i.e. How to compare with other higgs couplings)

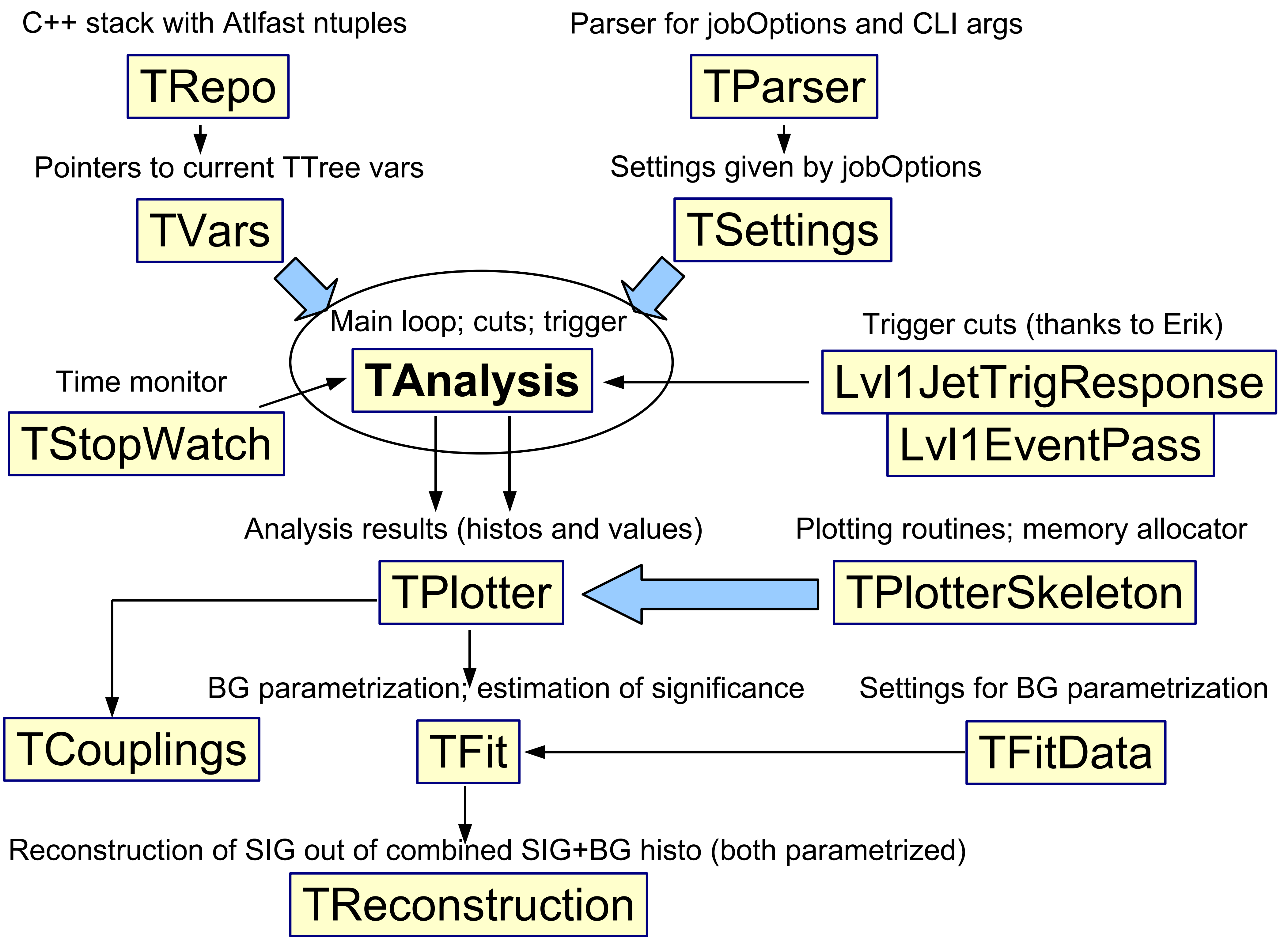
Reinventing the wheel

Dual purpose:

study physics <-> learn C++/ROOT

Designed a <u>stand-alone analysis framework</u>:

- Customizable collection of cuts (14 types)
- Extensible to other physics channels
- Athena-style job-options (>50 settings)
- Multiple stages (cuts; BG param; reco)
- Ability to resume from a previous stage
- Fully scriptable => easy to run via Condor


C++ stack with Atlfast ntuples
TRepo
Pointers to current TTree vars
TVars
Parser for jobOptions and CLI args
TParser
Settings given by jobOptions
TSettings
Main loop; cuts; trigger
TAnalysis
Time monitor
TStopWatch
Trigger cuts (thanks to Erik)
Lvl1JetTrigResponse
Lvl1EventPass
Analysis results (histos and values)
TPlotter
Plotting routines; memory allocator
TPlotterSkeleton
TCouplings
BG parametrization; estimation of significance
TFit
Settings for BG parametrization
TFitData
Reconstruction of SIG out of combined SIG+BG histo (both parametrized)
TReconstruction