Higgs and Jet Color-Connections

Jet *Super*structure and Event Kinematics

Jason Gallicchio

UC Davis

27 September 2011

- Higgs Status and $H \to b\bar{b}$ Motivation

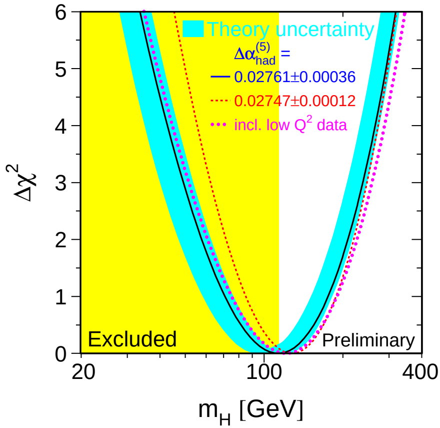
Outline

- Higgs Status and $H \to b\bar{b}$ Motivation
- Jets beyond (E, η, ϕ): Color Superstructure
  - Color Connection Primer
  - QCD radiation *between* and *within* jets
  - DØ Data on Higgs and Top and ATLAS work

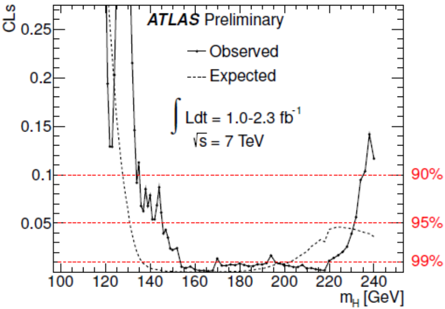
Outline

- Higgs Status and $H \to b\bar{b}$ Motivation
- Jets beyond (E, η, ϕ): Color Superstructure
  - Color Connection Primer
  - QCD radiation *between* and *within* jets
  - DØ Data on Higgs and Top and ATLAS work
- Multivariate Kinematics+Color
  - New kinematic observables
  - Ranking observables
  - Combining observables (multivariate Boosted Decision Trees)

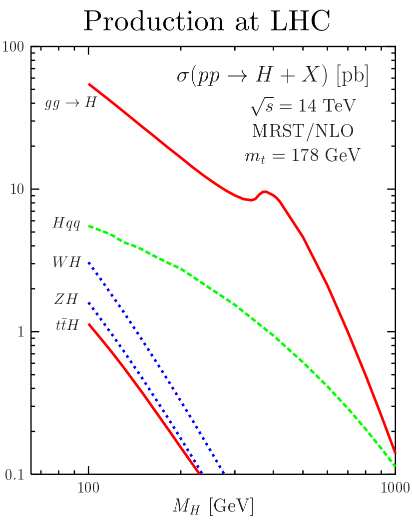
LEP Precision Electroweak



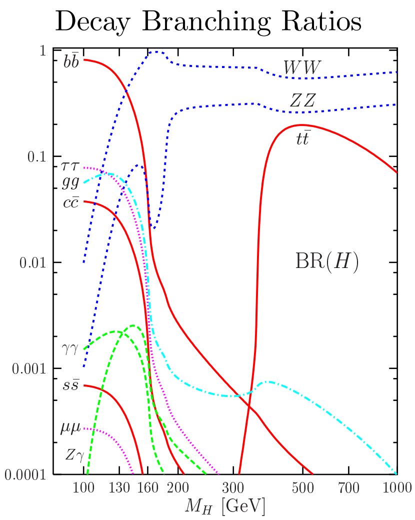
LHC Possibilities 2011



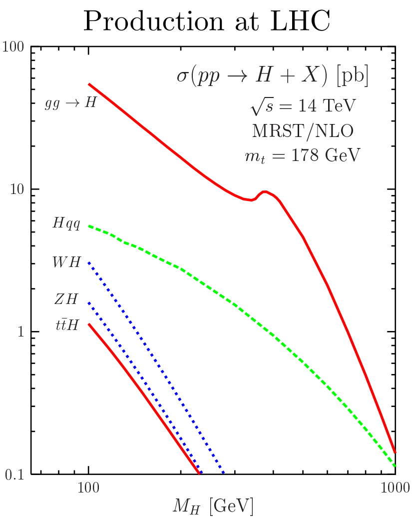
Higgs at LHC

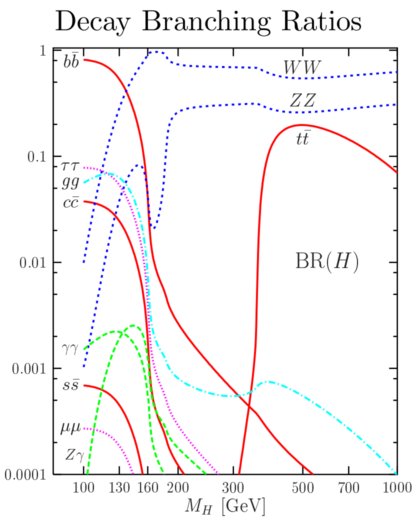


Production at LHC
$\sigma(pp \to H + X)$ [pb]
$\sqrt{s} = 14$ TeV
MRST/NLO
$m_t = 178$ GeV
$gg \to H$
Hqq
WH
ZH
$t\bar{t}H$
100
10
1
0.1
100
1000
M_H [GeV]



Decay Branching Ratios
$b\bar{b}$
WW
ZZ
$t\bar{t}$
$\tau\tau$
gg
$c\bar{c}$
BR(H)
$\gamma\gamma$
$s\bar{s}$
$\mu\mu$
$Z\gamma$
1
0.1
0.01
0.001
0.0001
100 130 160 200 300 500 700 1000
M_H [GeV]





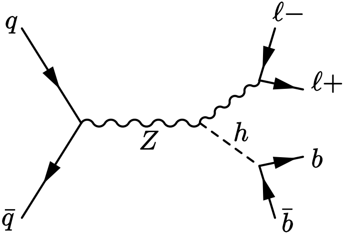
Improve search for $H \to b\bar{b}$ associated with a Z/W for $m_H \approx 120$ GeV.

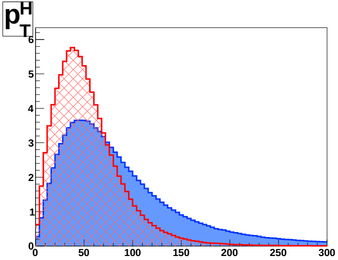
Higgs + Z

Kinematic variables to distinguish signal from background somewhat...

Good ones are p_T^H, $\Delta\eta_{b\bar{b}}$, $\Delta\phi_{b\bar{b}}$, ...

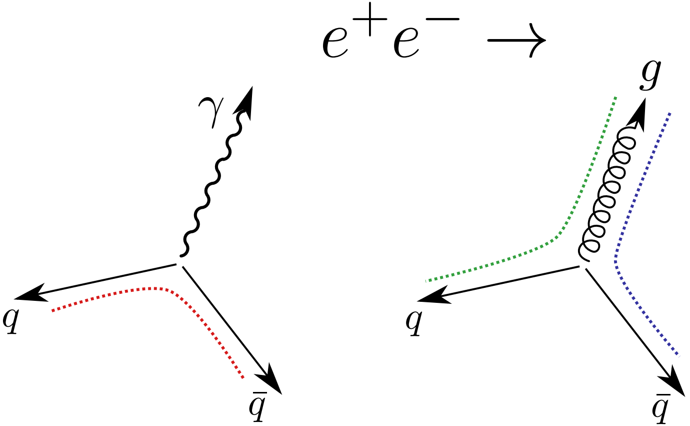
Higgs Signal





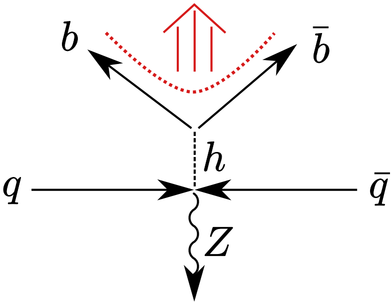
Full multivariate treatment after a long interlude...

Long History: 'String Effect' or 'Drag Effect' on planar events:



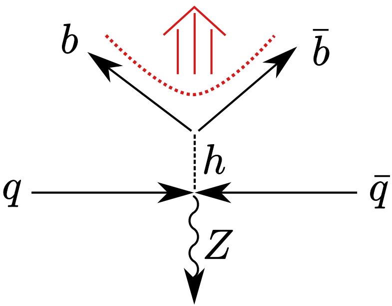
Gluons are treated as a colinear $q\bar{q}$ pair with *different* colors...
... equivalent up to $1/N_C^2$ corrections

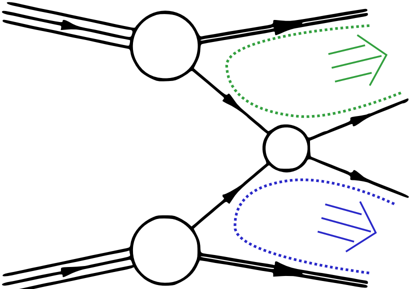
An Unexploited Handle for Higgs



Higgs Signal

bs form color singlet





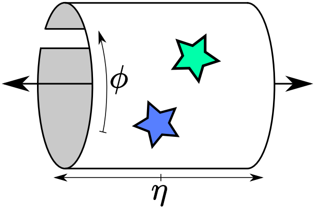
Higgs Signal

bs form color singlet

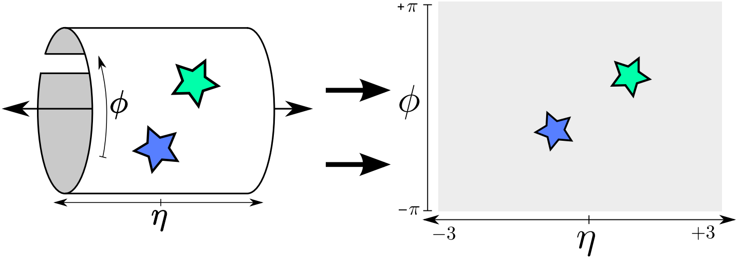
$Z + b\bar{b}$ QCD Background

bs color connected to beams

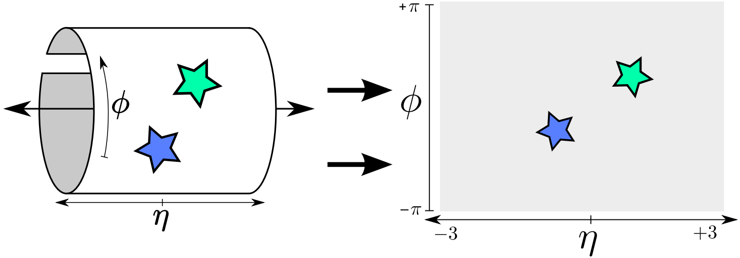
Showering Same Hard Event Millions of Times



Showering Same Hard Event Millions of Times



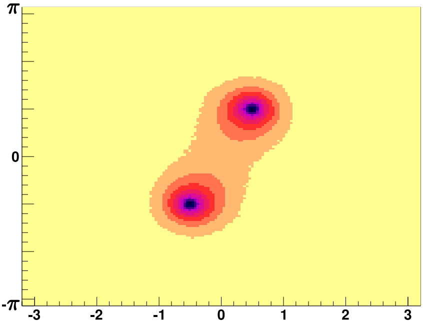
Showering Same Hard Event Millions of Times



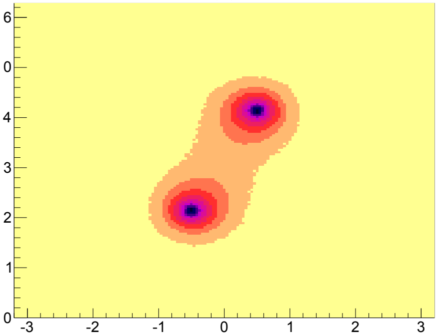
Higgs example:

$\Delta\eta_{b\bar{b}} = 1$
$\Delta\phi_{b\bar{b}} = 2$

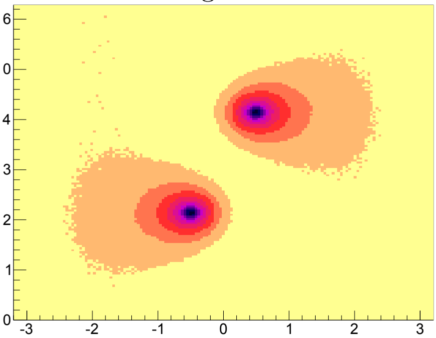
Add up p_T in each cell:



Higgs:

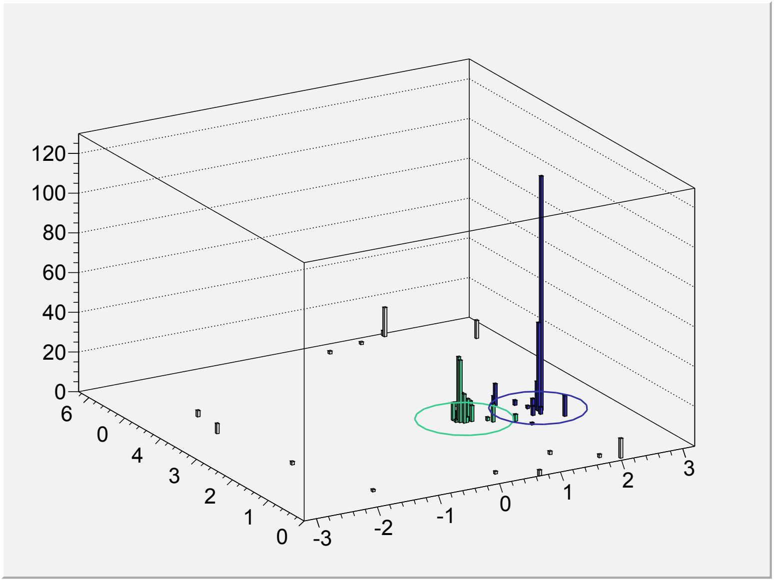


Background:

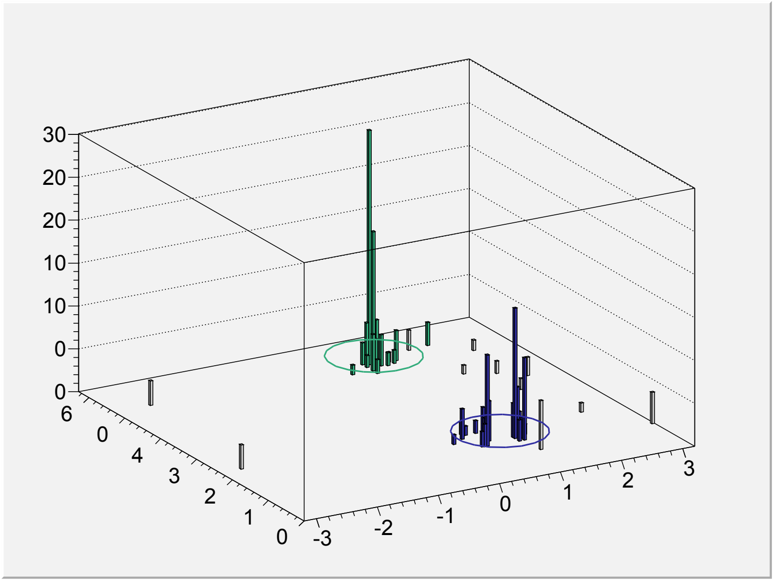


But event-by-event?

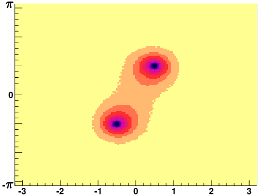
Higgs+Z Signal Event Example



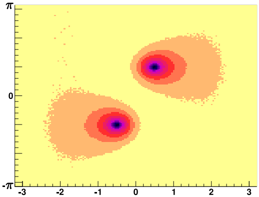
$Z + b\bar{b}$ Background Event Example



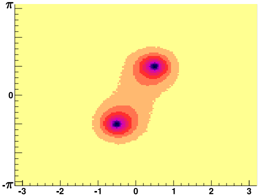
Higgs:



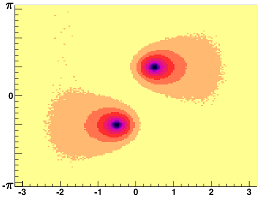
Background:



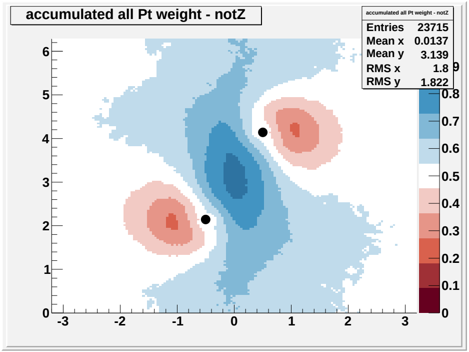
Higgs:



Background:



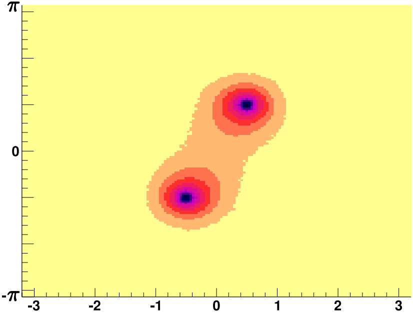
$\rightarrow$



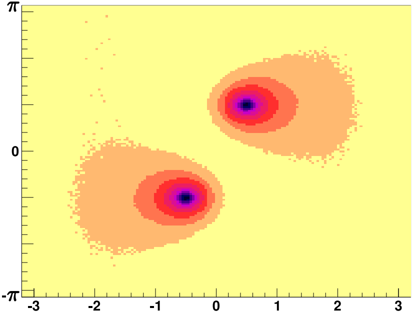
Important discrimination isn't at jet center — it's $\Delta R \approx 0.5 - 1.5$ away.

Focus on Jets Themselves — Pull

Signal Accumulated $\mathbf{p_t}$

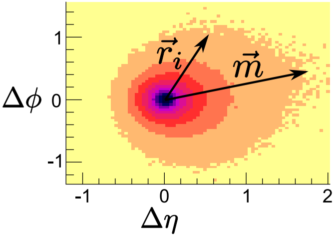


Background Accumulated $\mathbf{p_t}$



Pull

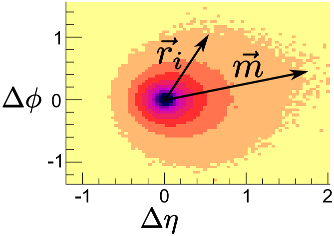
Add up particles or calorimeter energy deposits *within* a jet:



Pull Vector $\vec{m} = \sum_{i \in \text{jet}} \frac{p_T^i \, |r_i|}{p_T^{jet}} \, \vec{r}_i$ where $\vec{r}_i \equiv (y_i - y_{\text{jet}}, \phi_i - \phi_{\text{jet}})$

Pull

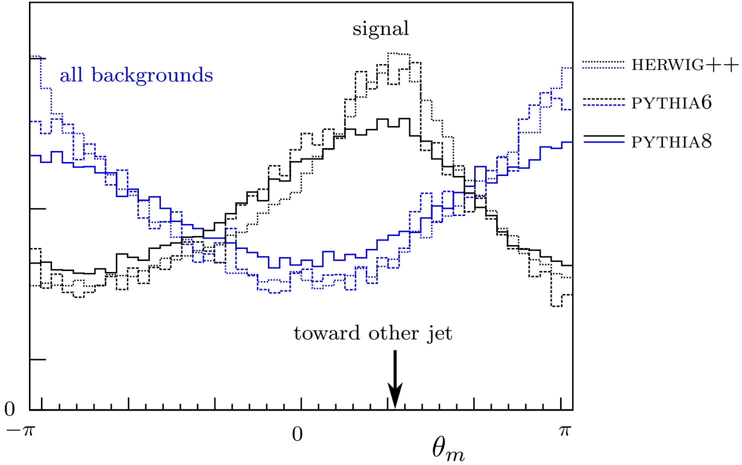
Add up particles or calorimeter energy deposits *within* a jet:



$$\text{Pull Vector } \vec{m} = \sum_{i \in \text{jet}} \frac{p_T^i \, |r_i|}{p_T^{jet}} \, \vec{r}_i \qquad \text{where} \qquad \vec{r}_i \equiv (y_i - y_{\text{jet}}, \phi_i - \phi_{\text{jet}})$$

- Angle of moment $\vec{m}$ gives "pointing" direction of teardrop
- Length of moment $|\vec{m}|$ doesn't help much

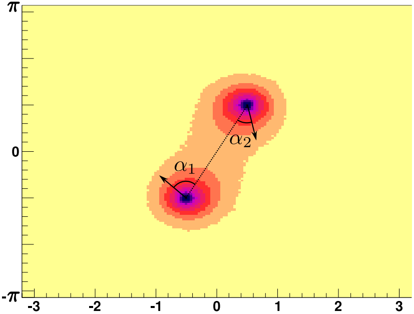
Different Shower Programs, Same Event



Distribution of the pull angle (one b-jet) with $\Delta y_{b\bar{b}} = 1$ and $\Delta\phi_{b\bar{b}} = 2$

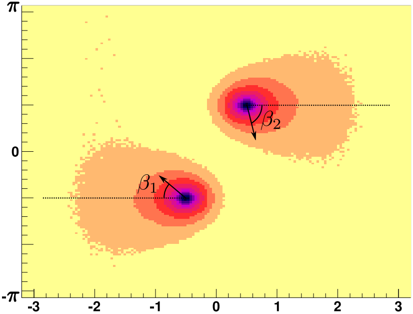
Pull Angle w.r.t. where it *Should* Point

Signal Accumulated $\mathbf{p_t}$



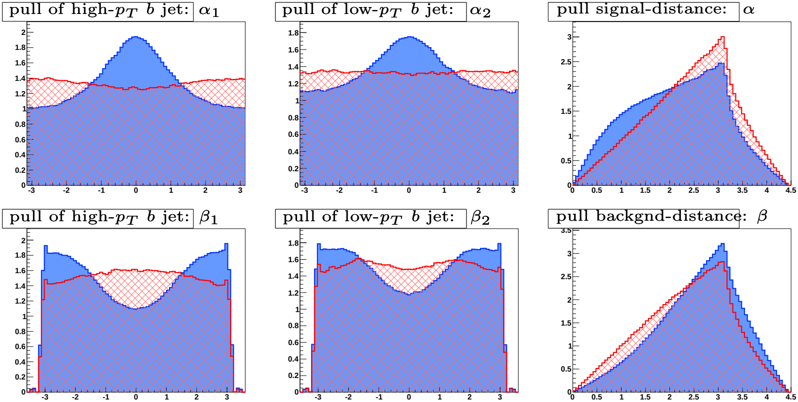
Signal pull: $\alpha = \sqrt{\alpha_1^2 + \alpha_2^2}$

Background Accumulated $\mathbf{p_t}$

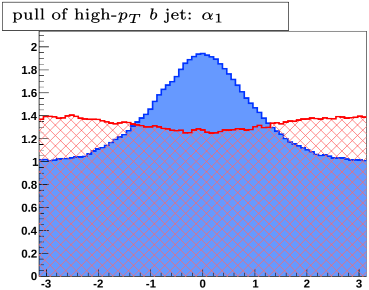


Background pull: $\beta = \sqrt{\beta_1^2 + \beta_2^2}$

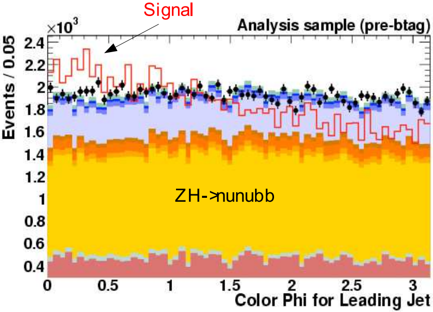
Pull Distributions for Full Z+Higgs Search



α_1

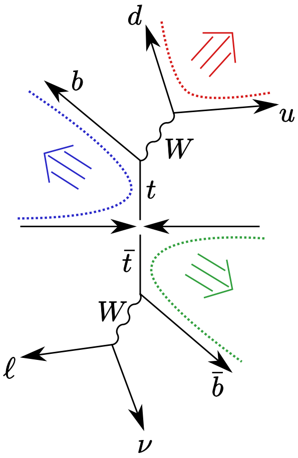


$|\alpha_1|$

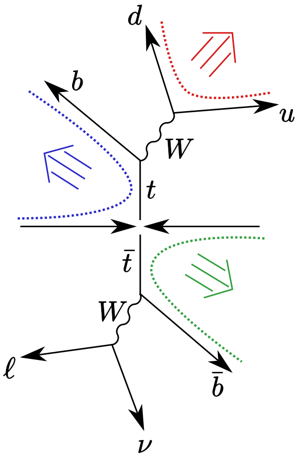


DØ Note 6087-CONF Aug 2010, Andy Haas: $ZH \to b\bar{b}\nu\bar{\nu}$
(data consistent with flat background)

Andy claims 5% improvement in multivariate search



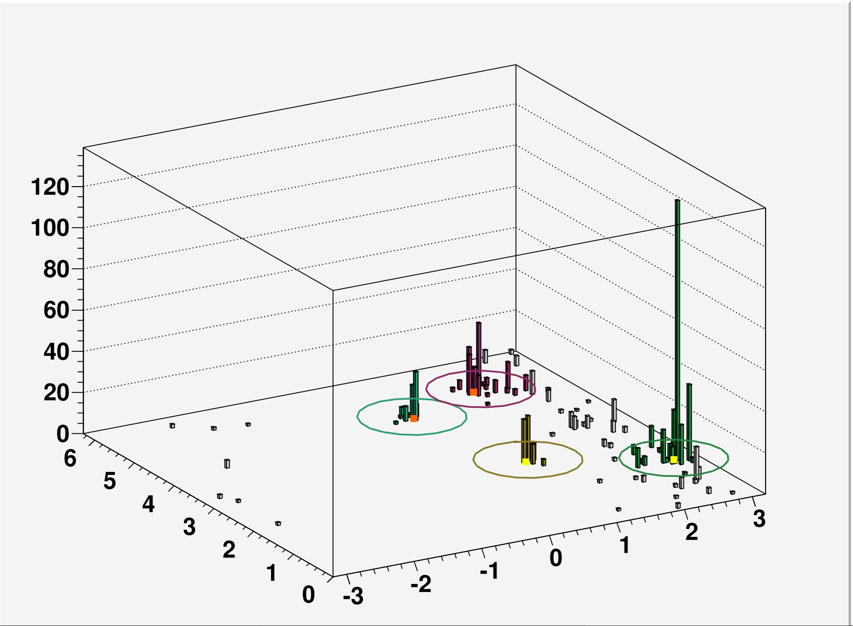
- The two b jets are color-connected to the beam (like $Z + b\bar{b}$ background earlier)
- The two light quark jets from W are color-connected to each other



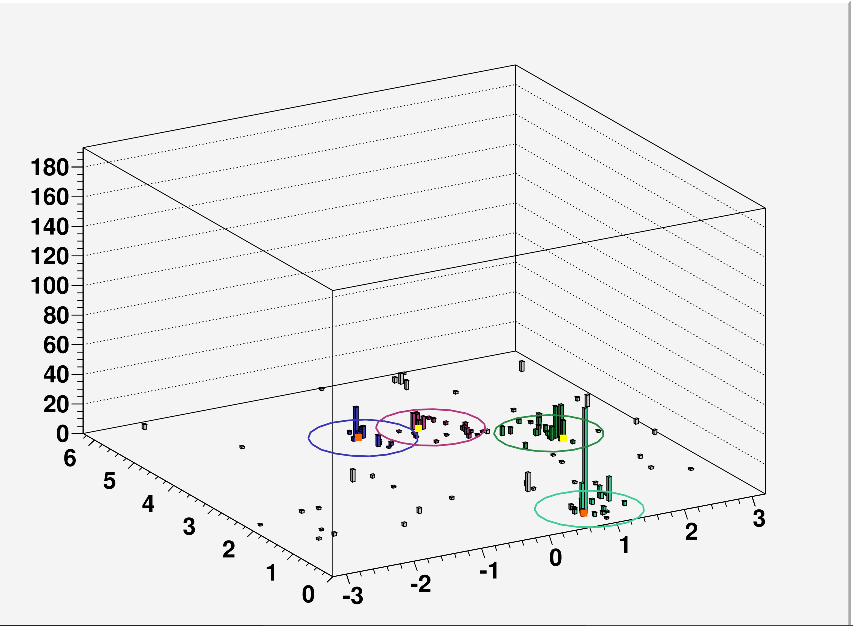
- The two b jets are color-connected to the beam (like $Z + b\bar{b}$ background earlier)
- The two light quark jets from W are color-connected to each other

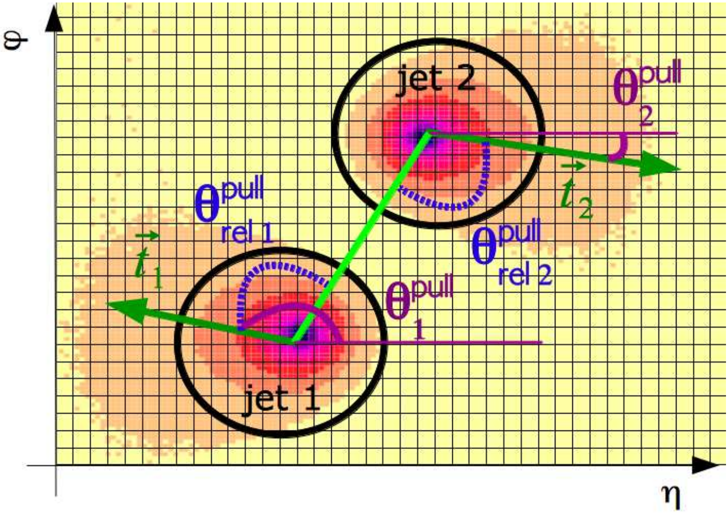
Test QCD and the Monte Carlos:
Given b tags and *clean* top sample, what do the pulls look like?

$t\bar{t}$ Event Example 1



$t\bar{t}$ Event Example 2

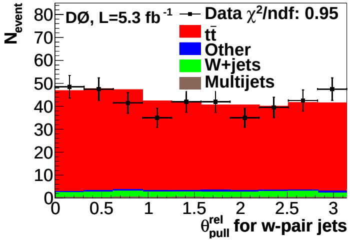


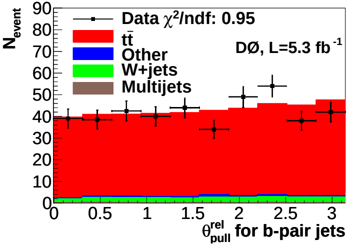


DØ Andy Haas, Yvonne Peters, **arXiv:1101.0648v1** [hep-ex] Jan 2011

$t\bar{t}$ Pull in DØ Data!

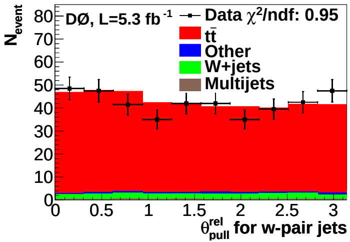
$t\bar{t} Jets$:

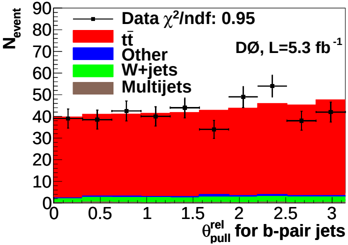




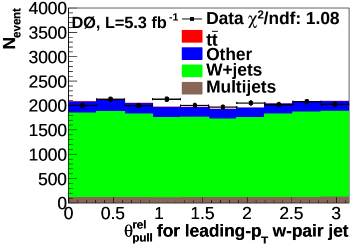
$t\bar{t}$ Pull in DØ Data!

$t\bar{t}Jets$:

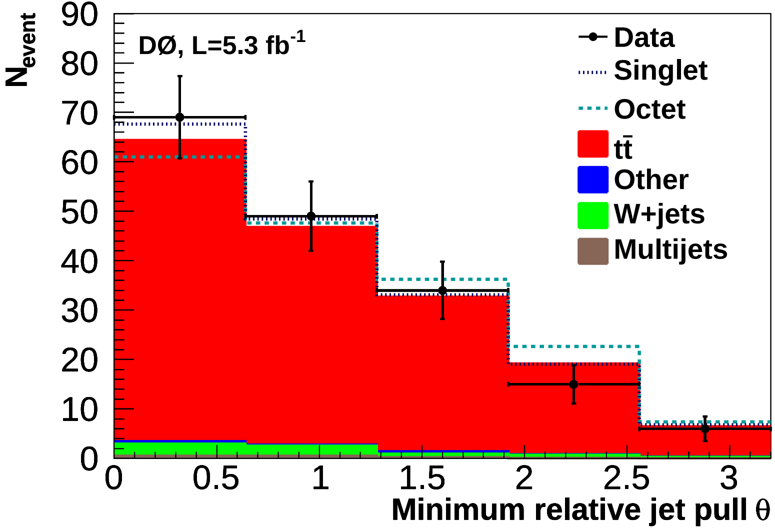




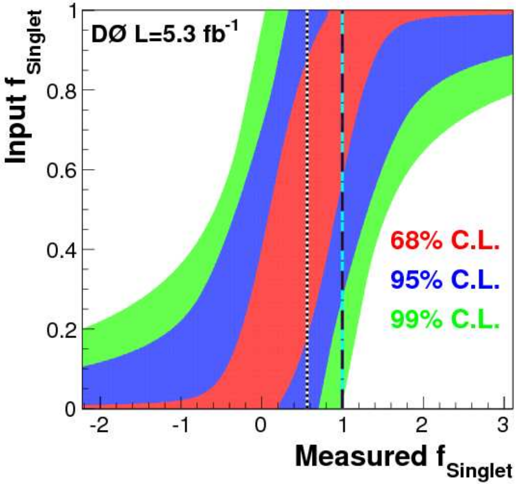
$W + 2jets$ (no b tags, so not from top):



W looks like a color singlet, not octet



W looks like a color singlet, not octet



$f_{\text{Singlet}} = 0.56 \pm 0.36(\text{stat}) \pm 0.22(\text{syst})$

... or exclude $f_{\text{Singlet}} = 0$ to three standard deviations

Improvements and ATLAS Plans

- Measure W's f_{Singlet} to 10% at ATLAS (DØ got 40%)

Improvements and ATLAS Plans

- Measure W's f_{Singlet} to 10% at ATLAS (DØ got 40%)
- $t\bar{t}$ cross section $\times$ acceptance at ATLAS is 25x to 80x Tevatron

Improvements and ATLAS Plans

- Measure W's f_{Singlet} to 10% at ATLAS (DØ got 40%)
- $t\bar{t}$ cross section $\times$ acceptance at ATLAS is 25x to 80x Tevatron
- Better calorimeter, tracking, and calo/track matching

Improvements and ATLAS Plans

- Measure W's f_{Singlet} to 10% at ATLAS (DØ got 40%)
- $t\bar{t}$ cross section $\times$ acceptance at ATLAS is 25x to 80x Tevatron
- Better calorimeter, tracking, and calo/track matching
- Use true boost-invariant jet rapidity (not pseudo-rapidity)

Improvements and ATLAS Plans

- Measure W's f_{Singlet} to 10% at ATLAS (DØ got 40%)
- $t\bar{t}$ cross section $\times$ acceptance at ATLAS is 25x to 80x Tevatron
- Better calorimeter, tracking, and calo/track matching
- Use true boost-invariant jet rapidity (not pseudo-rapidity)
- Take into account primary vertex location (affects p_T and rapidity)

Improvements and ATLAS Plans

- Measure W's f_{Singlet} to 10% at ATLAS (DØ got 40%)
- $t\bar{t}$ cross section $\times$ acceptance at ATLAS is 25x to 80x Tevatron
- Better calorimeter, tracking, and calo/track matching
- Use true boost-invariant jet rapidity (not pseudo-rapidity)
- Take into account primary vertex location (affects p_T and rapidity)
- Use pull in ZH and in $t\bar{t}H$ searches

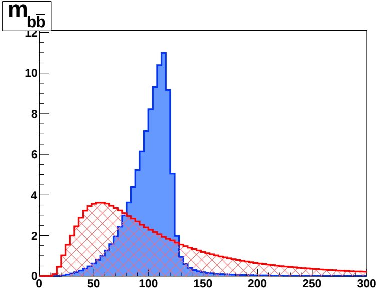
(This is a way to measure color of new particles!)

Improving the $H \to b\bar{b}$ search

arXiv:1010.3698 with theorists Matt Schwartz and Brock Tweedie and ATLAS people: Michael Kagan, John Huth, and Kevin Black.

Higgs Window

Higgs Invariant Mass $m_{b\bar{b}}$ using anti-k_T $R=0.5$ jets



ZH signal (solid blue) and $Zb\bar{b}$ background.
Normalized to same area (or you couldn't even see signal.)

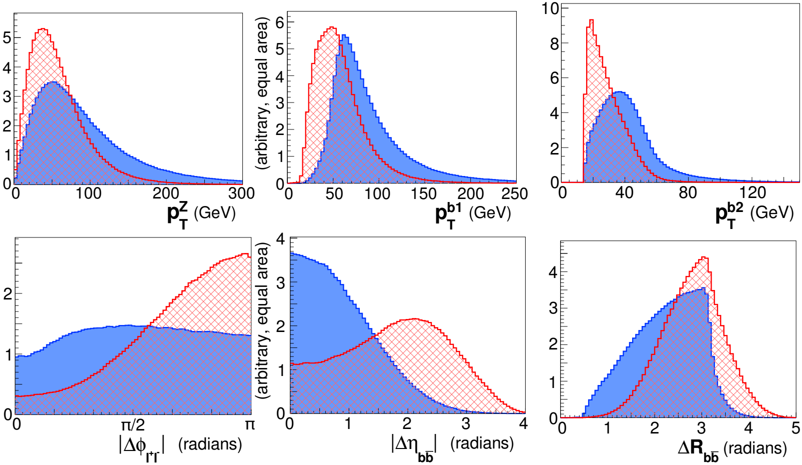
Pick initial window: $90\,\mathrm{GeV} < m_{b\bar{b}} < 124\,\mathrm{GeV}$(justified later)

LHC and TVT cross sections for ZH and QCD

	LHC (14 TeV)		Tevatron (1.96 TeV)	
Integrated Luminosity, $\int \mathscr{L}$	$30\ fb^{-1}$		$10\ fb^{-1}$	
	$pp \to ZH$	$pp \to Zb\bar{b}$	$p\bar{p} \to ZH$	$p\bar{p} \to Zb\bar{b}$
Xsec times Branching Ratio	33.4 fb	57,200 fb	3.63 fb	1250 fb
After Generator-Level Cuts	31.5 fb	26,000 fb	3.40 fb	570 fb
Two b Tags % (of Gen-Level)	57%	25%	81%	25%
Higgs Window % (of Gen-Level)	40%	4%	52%	3%
Initiated by gg	0%	90%	0%	27%
Xsec (in Higgs Window)	12.3 fb	1100 fb	1.8 fb	14.9 fb
Events (Xsec $\times \int \mathscr{L}$)	370	33,700	18	149
Starting B/S	91.1		8.2	
Starting $S/\sqrt{B}$	2.02		1.47	

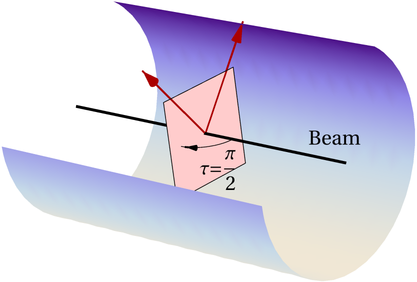
hard-parton level cuts	detector level cuts
$p_T^b > 7$ GeV	$p_T^b > 15$ GeV
$p_T^\mu > 3$ GeV	$p_T^\mu > 6$ GeV
$p_T^e > 3$ GeV	$p_T^e > 20$ GeV (LHC), 10 GeV (Tevatron)
$\|\eta_b\| < 5$ and $\|\eta_\ell\| < 5$	$\|\eta_b\| < 2.5$ and $\|\eta_\ell\| < 2.5$

Standard Kinematic Variables



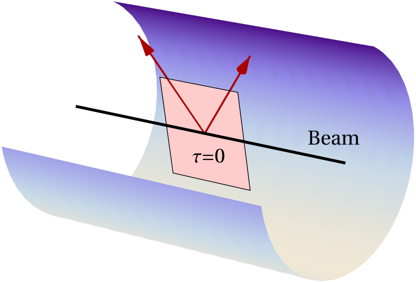
All In Higgs Mass-Window: $90\,\mathrm{GeV} < m_{b\bar{b}} < 124\,\mathrm{GeV}$ (and equal area)

Twist $\tau = \pi/2$



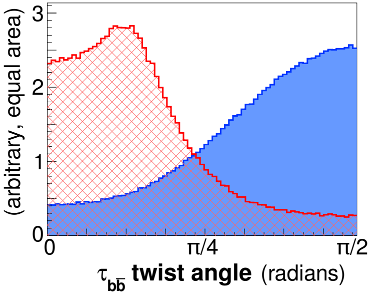
Higgs-Like

Twist $\tau = 0$

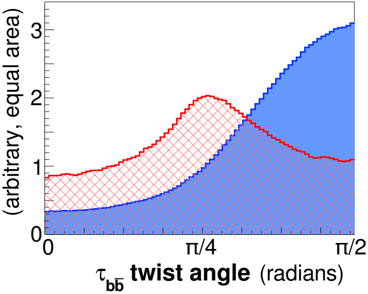


QCD-Like

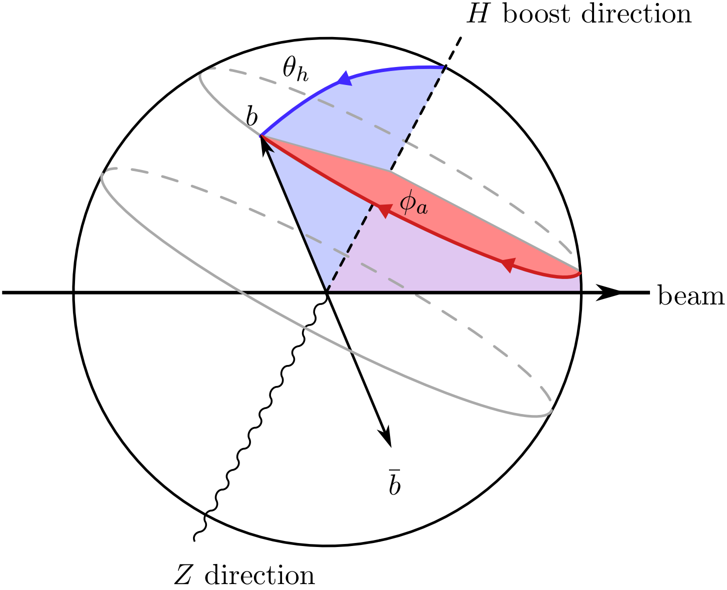
Parton level with no cuts



Jet level with detector cuts

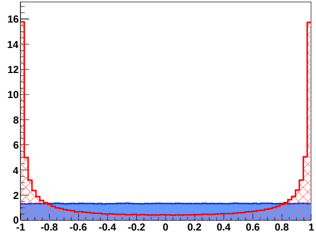


Higgs Rest Frame: Helicity and Azilicity

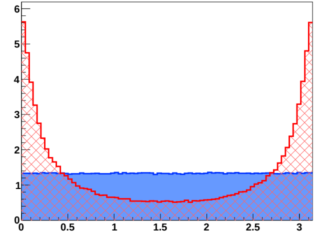


Madgraph hard partons with no cuts:

H Frame $cos(\theta_h)$ helicity angle

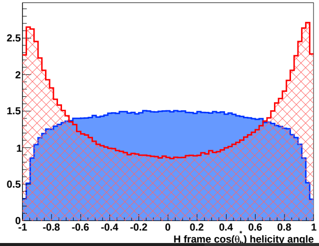


H Frame ϕ_a azilicity angle

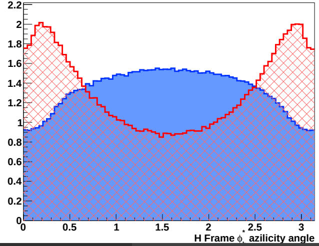


Showered and reconstructed, with detector cuts:

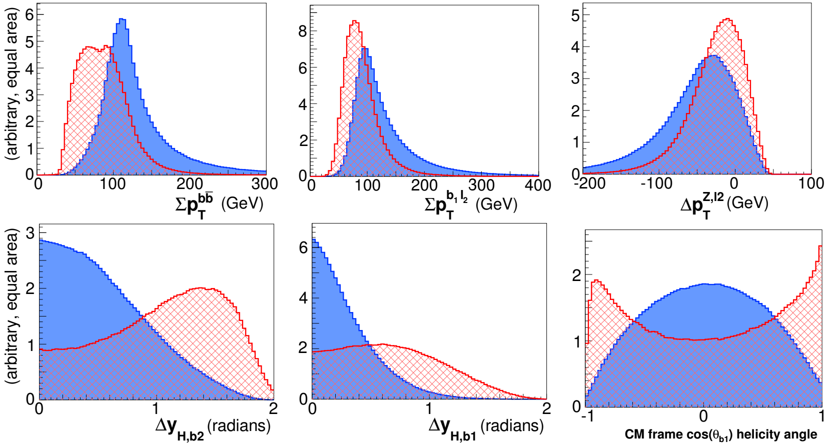
H Frame $cos(\theta_h)$ helicity angle



H Frame ϕ_a azilicity angle

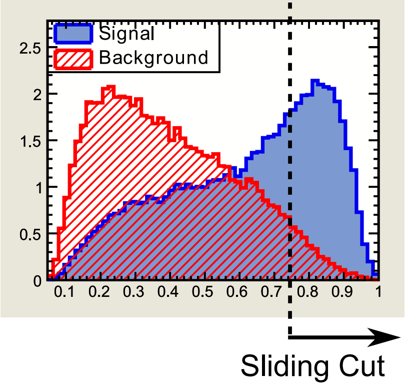


Menu-Method Variables: ‘One from Column A...’

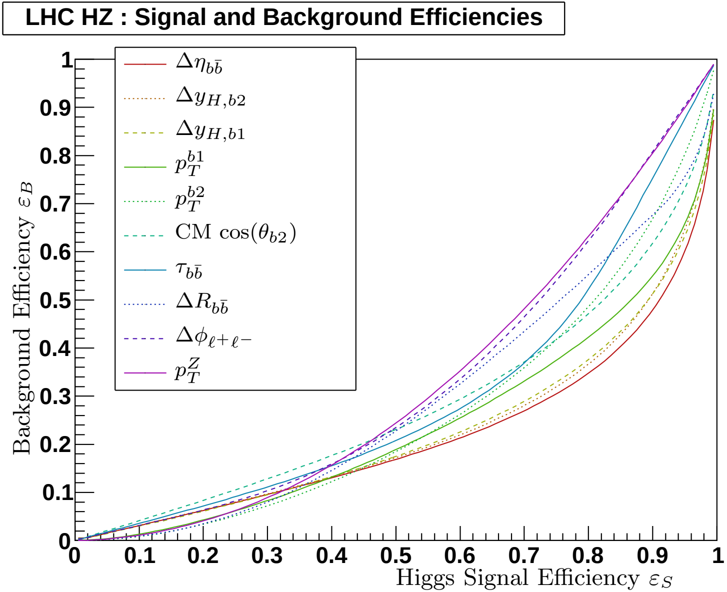


How do we pick the most useful variables?

We want to find clean variables that can be used for separation.

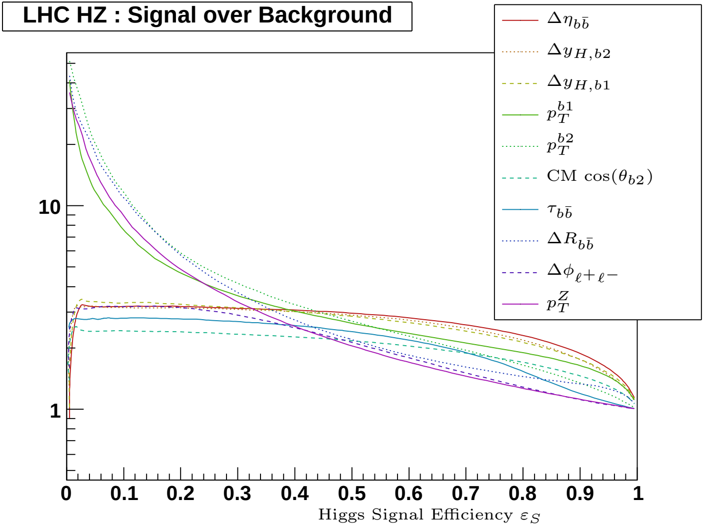


Background vs Signal Efficiency "ROC Curve"

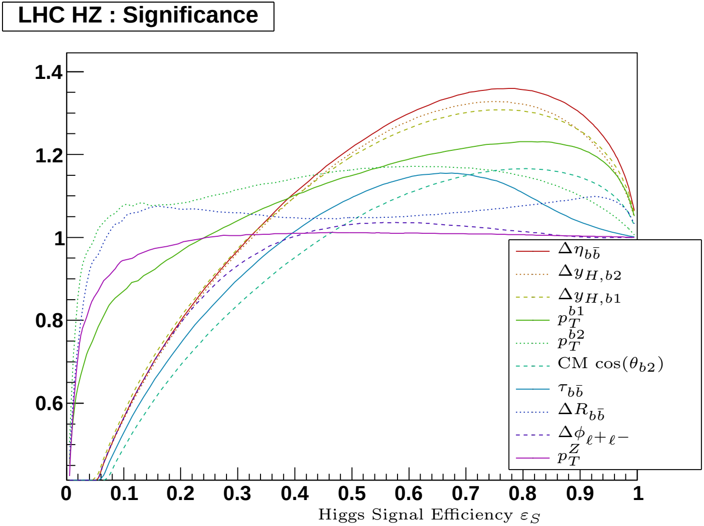


Signal Over Background Improvement

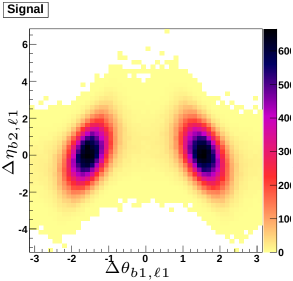
$$\frac{S}{B} \quad \xrightarrow{\text{cut}} \quad \frac{\varepsilon_S S}{\varepsilon_S B} = \left(\frac{\varepsilon_S}{\varepsilon_B}\right)\frac{S}{B}$$

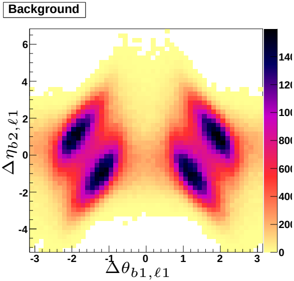


$$\sigma \equiv \frac{S}{\sqrt{B}} \quad \xrightarrow{\text{cut}} \quad \frac{\varepsilon_S S}{\sqrt{\varepsilon_B B}} = \left(\frac{\varepsilon_S}{\sqrt{\varepsilon_B}}\right)\sigma$$

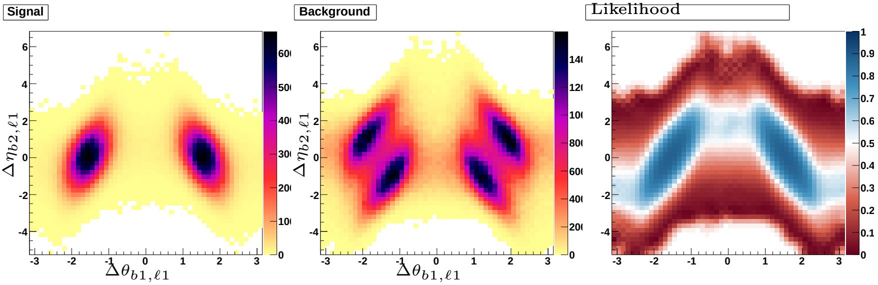


2D Likelihood

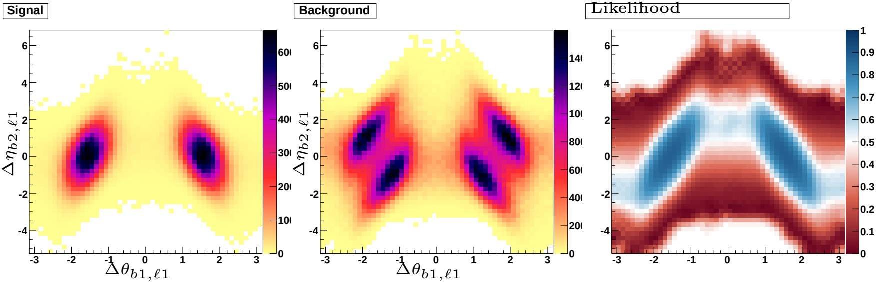


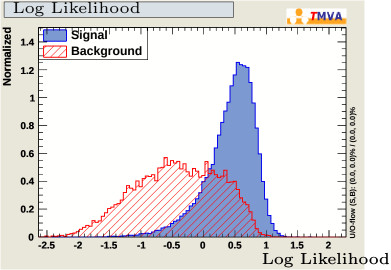


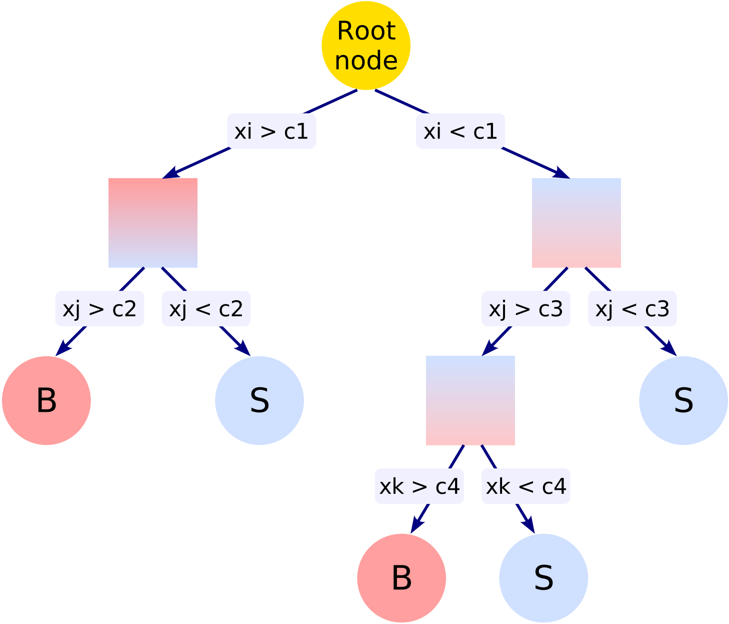
2D Likelihood



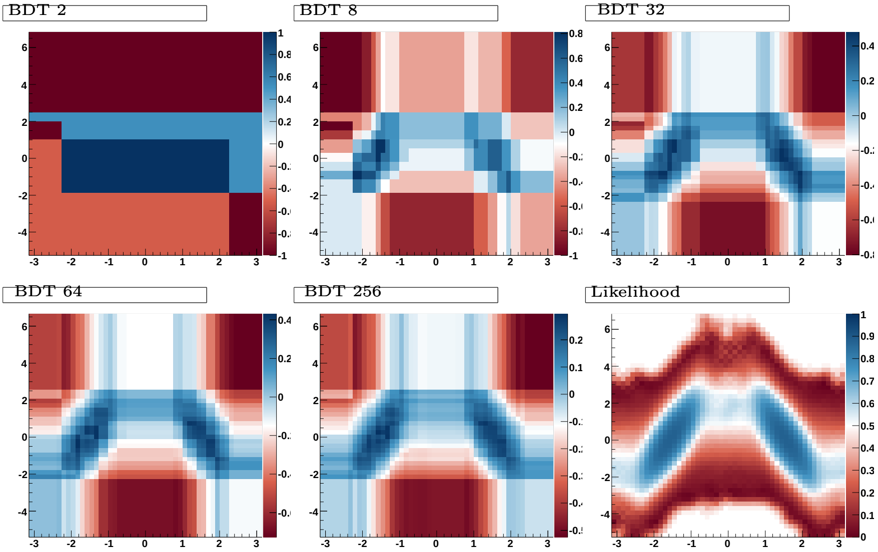
2D Likelihood



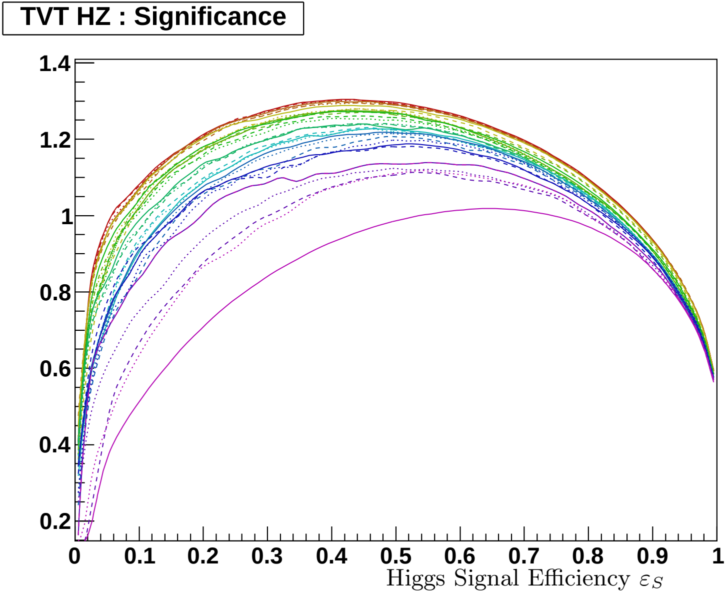


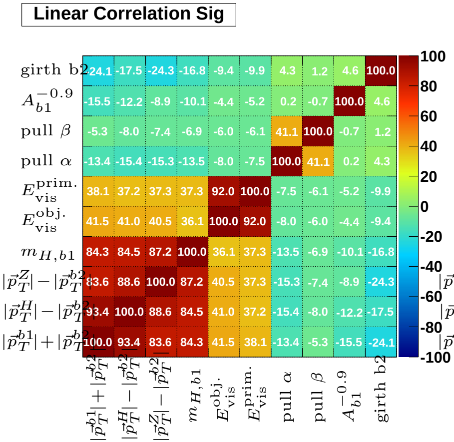

Root node
xi > c1
xi < c1
xj > c2
xj < c2
xj > c3
xj < c3
B
S
xk > c4
xk < c4
S
B
S

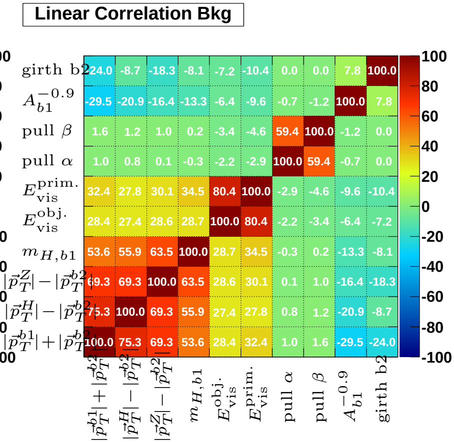
Boosted Decision Trees



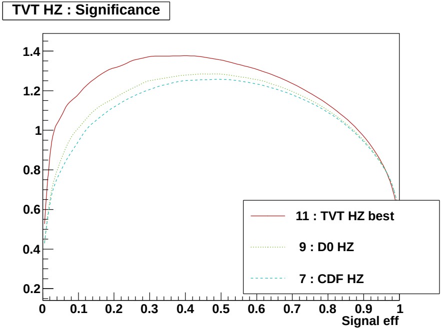
Tevatron ZH improvements, up to 10 variables

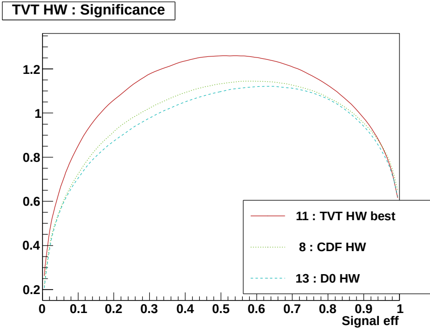



Linear Correlation Sig
girth b2 | 24.1 -17.5 -24.3 -16.8 -9.4 -9.9 4.3 1.2 4.6 100.0
$A_{b1}^{-0.9}$ | -15.5 -12.2 -8.9 -10.1 -4.4 -5.2 0.2 -0.7 100.0 4.6
pull β | -5.3 -8.0 -7.4 -6.9 -6.0 -6.1 41.1 100.0 -0.7 1.2
pull α | -13.4 -15.4 -15.3 -13.5 -8.0 -7.5 100.0 41.1 0.2 4.3
$E_{\mathrm{vis}}^{\mathrm{prim.}}$ | 38.1 37.2 37.3 37.3 92.0 100.0 -7.5 -6.1 -5.2 -9.9
$E_{\mathrm{vis}}^{\mathrm{obj.}}$ | 41.5 41.0 40.5 36.1 100.0 92.0 -8.0 -6.0 -4.4 -9.4
$m_{H,b1}$ | 84.3 84.5 87.2 100.0 36.1 37.3 -13.5 -6.9 -10.1 -16.8
$|\vec{p}_T^{Z}|-|\vec{p}_T^{b2}|$ | 83.6 88.6 100.0 87.2 40.5 37.3 -15.3 -7.4 -8.9 -24.3
$|\vec{p}_T^{H}|-|\vec{p}_T^{b2}|$ | 93.4 100.0 88.6 84.5 41.0 37.2 -15.4 -8.0 -12.2 -17.5
$|\vec{p}_T^{b1}|+|\vec{p}_T^{b2}|$ | 100.0 93.4 83.6 84.3 41.5 38.1 -13.4 -5.3 -15.5 -24.1
$|\vec{p}_T^{b1}|+|\vec{p}_T^{b2}|$, $|\vec{p}_T^{H}|-|\vec{p}_T^{b2}|$, $|\vec{p}_T^{Z}|-|\vec{p}_T^{b2}|$, $m_{H,b1}$, $E_{\mathrm{vis}}^{\mathrm{obj.}}$, $E_{\mathrm{vis}}^{\mathrm{prim.}}$, pull α, pull β, $A_{b1}^{-0.9}$, girth b2
100 80 60 40 20 0 -20 -40 -60 -80 -100


Linear Correlation Bkg
girth b2 | 24.0 -8.7 -18.3 -8.1 -7.2 -10.4 0.0 0.0 7.8 100.0
$A_{b1}^{-0.9}$ | -29.5 -20.9 -16.4 -13.3 -6.4 -9.6 -0.7 -1.2 100.0 7.8
pull β | 1.6 1.2 1.0 0.2 -3.4 -4.6 59.4 100.0 -1.2 0.0
pull α | 1.0 0.8 0.1 -0.3 -2.2 -2.9 100.0 59.4 -0.7 0.0
$E_{\mathrm{vis}}^{\mathrm{prim.}}$ | 32.4 27.8 30.1 34.5 80.4 100.0 -2.9 -4.6 -9.6 -10.4
$E_{\mathrm{vis}}^{\mathrm{obj.}}$ | 28.4 27.4 28.6 28.7 100.0 80.4 -2.2 -3.4 -6.4 -7.2
$m_{H,b1}$ | 53.6 55.9 63.5 100.0 28.7 34.5 -0.3 0.2 -13.3 -8.1
$|\vec{p}_T^{Z}|-|\vec{p}_T^{b2}|$ | 69.3 69.3 100.0 63.5 28.6 30.1 0.1 1.0 -16.4 -18.3
$|\vec{p}_T^{H}|-|\vec{p}_T^{b2}|$ | 75.3 100.0 69.3 55.9 27.4 27.8 0.8 1.2 -20.9 -8.7
$|\vec{p}_T^{b1}|+|\vec{p}_T^{b2}|$ | 100.0 75.3 69.3 53.6 28.4 32.4 1.0 1.6 -29.5 -24.0
$|\vec{p}_T^{b1}|+|\vec{p}_T^{b2}|$, $|\vec{p}_T^{H}|-|\vec{p}_T^{b2}|$, $|\vec{p}_T^{Z}|-|\vec{p}_T^{b2}|$, $m_{H,b1}$, $E_{\mathrm{vis}}^{\mathrm{obj.}}$, $E_{\mathrm{vis}}^{\mathrm{prim.}}$, pull α, pull β, $A_{b1}^{-0.9}$, girth b2
100 80 60 40 20 0 -20 -40 -60 -80 -100

Tevatron Improvement





Conclusions

Results:

- Tevatron Searches can be Improved 10% to 20%
- LHC's current "boosted Higgs" can be improved as much as 200%

Conclusions

Results:

- Tevatron Searches can be Improved 10% to 20%
- LHC's current "boosted Higgs" can be improved as much as 200%

Take Aways:

- Color-Connections and Jet Pull

Conclusions

Results:

- Tevatron Searches can be Improved 10% to 20%
- LHC's current "boosted Higgs" can be improved as much as 200%

Take Aways:

- Color-Connections and Jet Pull
- Ranking Variables (significance improvement)

Conclusions

Results:

- Tevatron Searches can be Improved 10% to 20%
- LHC's current "boosted Higgs" can be improved as much as 200%

Take Aways:

- Color-Connections and Jet Pull
- Ranking Variables (significance improvement)
- Boosted Decision Trees are not mysterious

Conclusions

Results:

- Tevatron Searches can be Improved 10% to 20%
- LHC's current "boosted Higgs" can be improved as much as 200%

Take Aways:

- Color-Connections and Jet Pull
- Ranking Variables (significance improvement)
- Boosted Decision Trees are not mysterious
- Correlated variables still provide improvement

Conclusions

Results:

- Tevatron Searches can be Improved 10% to 20%
- LHC's current "boosted Higgs" can be improved as much as 200%

Take Aways:

- Color-Connections and Jet Pull
- Ranking Variables (significance improvement)
- Boosted Decision Trees are not mysterious
- Correlated variables still provide improvement

Thank You

...

CDF ZH	DØ ZH	CDF WH	DØ WH
$m_{b\bar{b}}$ p_T^{b1} p_T^{b2} $\hat{s}$ p_T^Z H_T $\text{Sphericity}^{\text{obj.}}$	$m_{b\bar{b}}$ p_T^{b1} p_T^{b2} $\Delta R_{b\bar{b}}$ $\Delta R_{\ell+\ell-}$ ΔR_{ZH} $\text{CM} \cos\theta_H$ P_T^H H_T	$m_{b\bar{b}}$ $p_T^{\text{imbalance}}$ $m_{W,b1}$ $m_{W,b2}$ η_ℓ $\Sigma p_T^{b\bar{b}}$ p_T^W H_T	$m_{b\bar{b}}$ p_T^{b1} p_T^{b2} E_{b2} $\Delta R_{b\bar{b}}$ $\Delta\phi_{b\bar{b}}$ $\Delta\phi_{b1,\ell}$ p_T^H p_T^W $\hat{s}$ $\Delta R_{W,H}$ H_z $\text{CM} \cos\theta_H$

Our Top Variables: (GROUP LENGTH IN - GROUP)

LHC ZH		LHC WH		TVT ZH		TVT WH	
1-10	$\lvert\vec{p}_T^{b1}\rvert + \lvert\vec{p}_T^{b2}\rvert$	1-10	Δy_{WH}	1-10	$\lvert\vec{p}_T^{b1}\rvert + \lvert\vec{p}_T^{b2}\rvert$	1-10	Δy_{WH}
1-3	$\Delta\eta_{b\bar{b}}$	2-10	$\lvert\vec{p}_T^{b1}\rvert + \lvert\vec{p}_T^{b2}\rvert$	1-10	$\Delta\eta_{b\bar{b}}$	1-2	CM $\cos\theta_H$
1-2	$\Delta y_{H,b2}$	1-2	CM $\cos\theta_H$	1-7	$\Delta y_{Z,H}$	2-10	$\Delta y_{\ell+\ell-}$
2-4	$\lvert\vec{p}_T^{H}\rvert - \lvert\vec{p}_T^{b2}\rvert$	2-4	$\Delta\phi_{\ell+\ell-}$	1-10	twist $\tau_{\ell+\ell-}$	3-10	$H_T^{\text{prim.}}$
3-5	$\Delta y_{H,\ell 1}$	2-7	Twist $\tau_{b\bar{b}}$	2-10	Centrality	4-10	$\Delta\eta_{b\bar{b}}$
3-4	Δy_{ZH}	3-10	Twist $\tau_{\ell+\ell-}$	2-3	$\lvert\vec{p}_T^{b1}\rvert$	4-10	$\Delta p_T^{\ell^+\ell^-}$
4-10	$\lvert\vec{p}_T^{Z}\rvert + \lvert\vec{p}_T^{b2}\rvert$	3-10	$E_{\text{vis}}^{\text{prim.}}$	3-10	pull β	4-10	pull α
4-10	$m_{H,\ell 1}$	4-10	pull β	3-10	pull α	4-10	pull β
4-9	*Sphericity*	4-6	$\Delta\phi_{W,l2}$	3-10	$m_{H,\ell 2}$	7-10	avg. subj. p ($R_{\text{sub}}^{k_T}=0.$
5-9	$\Sigma y_{Z,b1}$	5-10	$m_{W,b1}$	3-5	$\cos\theta_{\ell 2}$ (Z Frame)	7-10	m_{b2}/p_T^{b2}
5-9	$\Sigma y_{Z,b1}$	9-10	$m_{W,b2}$	5-8	girth g_{b2}	8-9	$m_T^{b\bar{b}}$
5-9	$\Sigma y_{Z,b1}$	6-10	$\Delta R_{H,b2}$	6-10	angul. $A_{b2}^{-0.01}$	8-10	$m_{W,b1}$
5-10	$E_{\text{vis}}^{\text{obj.}}$	5-10	angul. $A_{b2}^{-0.1}$	7-9	m_{b2}/P_T^{b2}		
5-10	pull α	5-10	$\lvert\vec{p}_T^{W}\rvert - \lvert\vec{p}_T^{\ell 1}\rvert$	9-10	angul. $A_{b1}^{+0.01}$		
6-10	girth g_{b2}	8-9	pull α	9-10	$\Delta\phi_{b\bar{b}}$		
6-10	pull β_2						
8-10	angul. $A_{b1}^{-0.90}$						
9-10	pull β						
9-10	$\lvert\vec{p}_T^{H}\rvert + \lvert\vec{p}_T^{l2}\rvert$						