Karlsruher Institut für Technologie

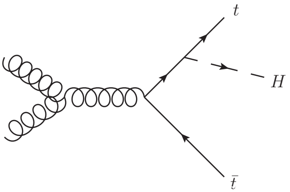
Search for $t\bar{t}H$ production in boosted topologies

GRK Plenary Workshop 2015, Freudenstadt

Karim El Morabit, Marco Harrendorf, Ulrich Husemann, Patricia Lobelle, Hannes Mildner, Tobias Pfotzer, Matthias Schröder, **Shawn Williamson** | September 30, 2015

INSTITUT FÜR EXPERIMENTELLE KERNPHYSIK (IEKP)

$t\bar{t}$H Production

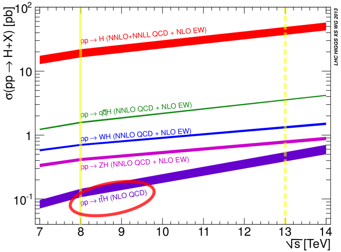


Special: Top-Higgs Yukawa Coupling

- Also large contributions in loop processes (e.g. $gg \rightarrow$ H, H $\rightarrow \gamma\gamma$)
  $\rightarrow$ Coupling disguised by other contributions
- Special for $t\bar{t}$H: Direct access to the top-Higgs Yukawa coupling

Challenging: Search

- Predicted by SM, but not yet sensitive for discovery
- Higgs production channel with very small cross section
- Dominating background processes



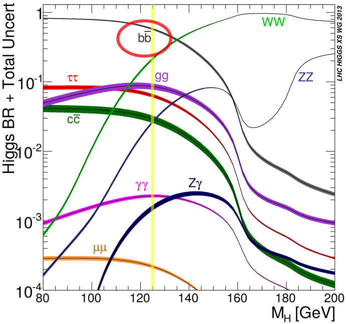
13 TeV: $\sim$4 times higher cross section compared to 8 TeV

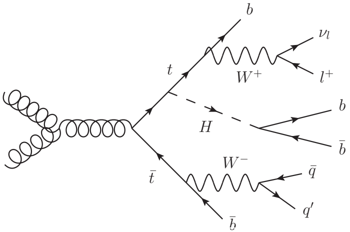
Final State Particle Decays

$t\bar{t}$ Decay: Semileptonic

- Hadronic top decay:
  Large branching fraction & background
- Leptonic top decay:
  Small branching fraction & background

Semileptonic $t\bar{t}$ decay: Compromise between clean signature and good statistics





Higgs Decay: $H \to b\bar{b}$

- Largest branching fraction at $m_{Higgs} = 125$ GeV (≈ 58 %)
- Overwhelming background in hadron collisions

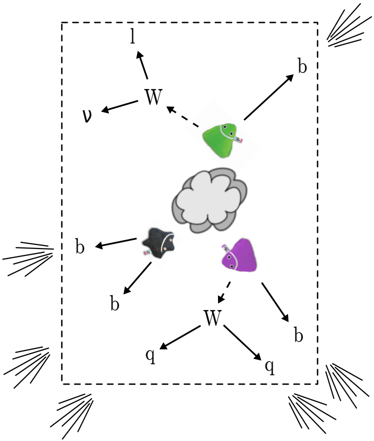
Semileptonic $t\bar{t}H(H \rightarrow b\bar{b})$ Event Topology

Decay Signature

- Four bottom quarks (Higgs boson and top quarks)
- Two light quarks (hadronic W boson)
- Charged lepton + MET (leptonic W boson)

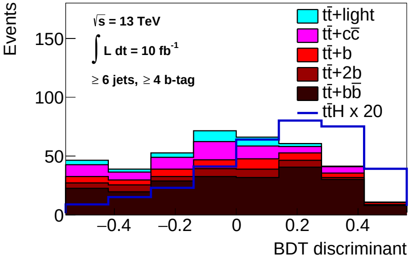
Resulting challenges:

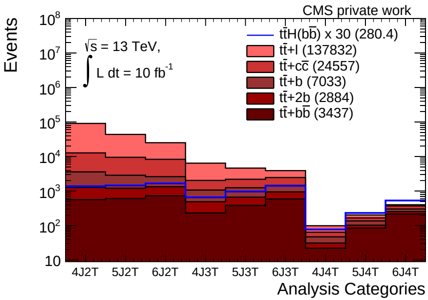
- Dominating irreducible background processes ($t\bar{t}(+b\bar{b})$ production)
- Many Final state particles $\rightarrow$ Combinatorial problem in event reconstruction



$t\bar{t}H(H \to b\bar{b})$ Analysis

1. Event selection
   - Single lepton
2. Event categorization based on
   - Jet multiplicity
   - b-tag multiplicity



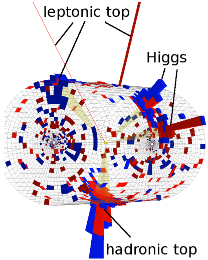


3. Event reconstruction
   - Assign jets to decay products
   - Minimize χ^2 based on particle masses
4. Final BDTs:
   - Discriminate $t\bar{t}H$ against backgrounds
   - Separately trained for each category

Boosted Semileptonic $t\bar{t}H(H \rightarrow b\bar{b})$ Event Topology

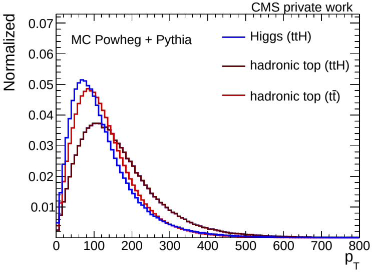
Plehn, Salam, Spannowsky (2009): Boosted kinematic regime

Analysis of events containing top quarks and Higgs bosons with large transverse momenta



But Why?

- Production of three heavy particles
  $\rightarrow$ Small phase space
- Heavy particles only moderately boosted



Still ...

- For $t\bar{t}H$ high p_T particles occur
  - more often compared to $t\bar{t}$
  - a lot more compared to QCD, Electro-Weak

  $\rightarrow$ Background reduction

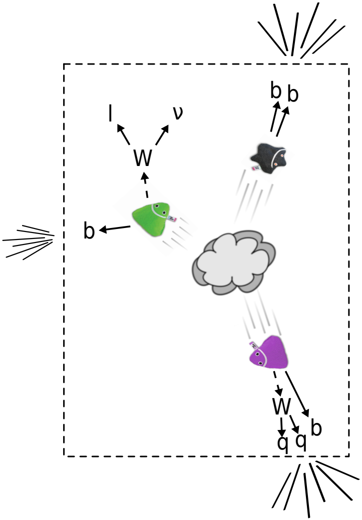
Boosted Semileptonic $t\bar{t}H(H \to b\bar{b})$ Event Topology

The real Reason: Reconstruction

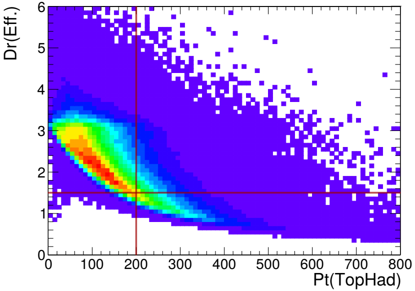
- Massive particles pass momentum to decay products
- Decay products are collimated $\to$ Simplified reconstruction

Approach:

- Cluster decay products with large distance parameters $\to$ fat jet
- Analyze jet substructure with dedicated algorithms
- Combine information for identification of top quarks and Higgs boson

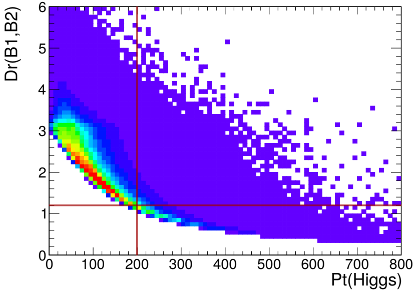


Decay Products Distance



- Distance of decay products
  - Eff. ΔR: Hadronic top quark (upper left)
  - ΔR: Higgs boson (bottom right)
- Differential in massive particle momentum

- Distance required to cluster decay products into one jet
- Based on MC particle level information
- Larger distances for hadronic top due to higher mass



Boosted Hadronic Top Reconstruction

Fat Jets

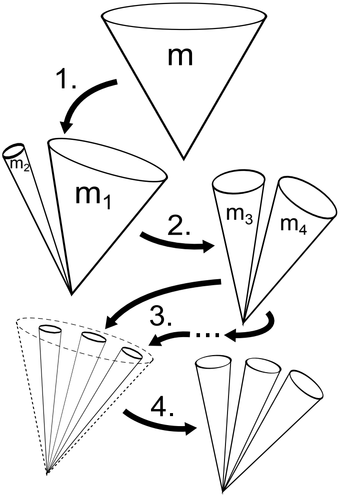
- Cambridge-Aachen algorithm $R = 1.5$

Subjet Finding

- HEP Top Tagger Algorithm
- Declustering of fat jets
- Subjet finding based on mass drop
- Subsequent jet grooming
- Returns three subjets assigned to hadronic top decay products

Additional Substructure Information

- N-Subjettiness
- Optimal R
- Groomed jet masses

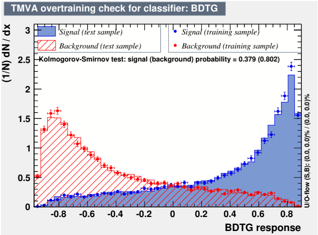


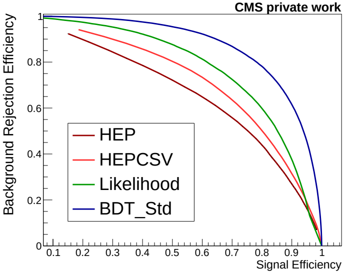
Boosted Hadronic Top Identification

Identification based on continuous discriminators for later optimization

Cut Based Discriminators

- Standard HEP Top Tagger (HEP): Cuts on invariant masses of subjet combinations
- b-tagging HEP Top Tagger (HEPCSV): HEP Top Tagger with b-tagging subjet assignment





Multivariate Discriminators

- Likelihood Top Tagger: Based on HEP-Top-Tagger variables
- BDT Top Tagger (BDT_Std): BDT trained on substructure variables

Boosted Higgs Reconstruction

Fat Jets

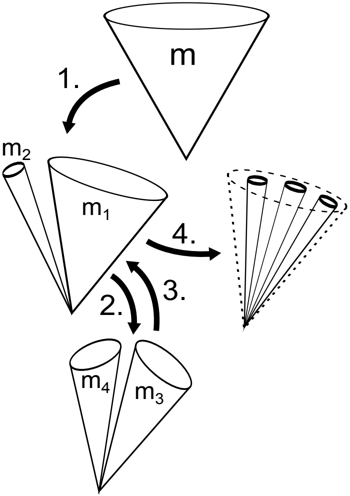
- Cambridge-Aachen algorithm $R = 1.2$

Subjet Finding

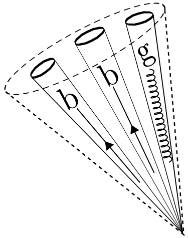
- Subjet Filterjet Algorithm
- Declustering of fat jets
- Subjet finding based on mass drop
- Subsequent jet grooming
- Returns an arbitrary number of subjets

Additional Substructure Information

- N-Subjettiness
- Groomed jet masses



Boosted Higgs Identification

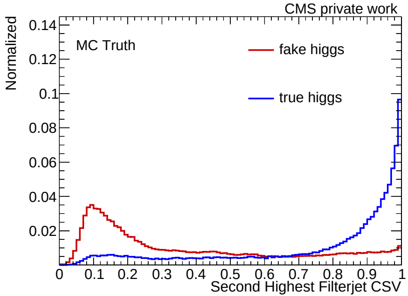


Higgs Discriminator

- Second highest b-tagging output value among filtered subjets
- Higgs mass unused $\rightarrow$ final BDT
- 8 TeV studies $\rightarrow$ improvement by MVA Higgs Tagger

Future Improvements

- Increase fat jet cone size
- CMS Higgs Tagger being developed
  - Double secondary vertex reconstruction
  - N-Subjettiness
  - Groomed masses



Event Selection Double Boosted Category

- Exactly one selected lepton
- One selected hadronic top candidate
- One selected Higgs candidate

Lepton Selection (e/μ)

- $p_T > 30\,\text{GeV}$
- $|\eta| < 2.1$
- Isolated

Had. Top Candidate Selection

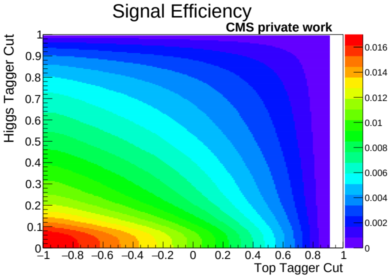
- $p_T > 200\,\text{GeV}$
- $|\eta| < 2.0$
- Select top jet with highest top identification value

Higgs Candidate Selection

- $p_T > 200\,\text{GeV}$
- $|\eta| < 2.0$
- Select Higgs jet with highest Higgs identification value
- Removing of overlap with hadronic top candidate:
  - Distance to hadronic top candidate $\Delta R > 1.5$
  - No sharing of subjets with hadronic top candidate

Event Selection Performance

Full Event selection except for the cuts on the discriminator values

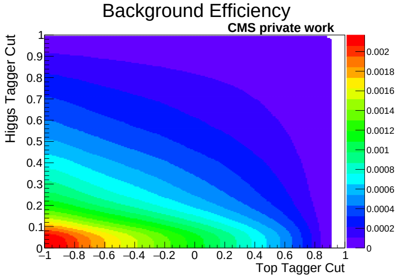


Performance Analysis

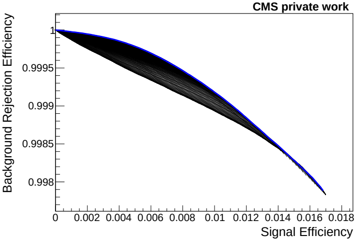
- Performance evaluated for all combinations of
  - Higgs discriminator cut
  - Hadronic top discriminator cut
- Analyzed for $t\bar{t}H$ (signal) and $t\bar{t}$ (main background)

Benchmark variables

- Signal selection efficiency
- Background selection efficiency
- Signal over background (S/B)
- Signal Significance ($S/\sqrt{B}$)
- Event reconstruction efficiency



Event Selection Optimization

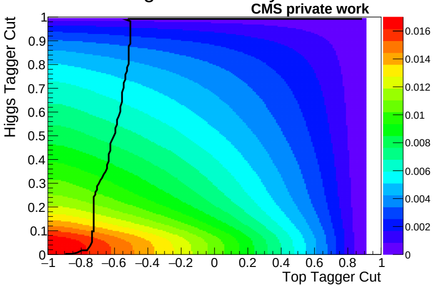


Receiver Operating Characteristics (ROCs)

- Selection performance analysis for each cut combination
- Various cut combinations with same signal efficiency → Broad ROC space

Selection Optimization

- Search for best performing cuts
- Choose cut combination with lowest background efficiency for given signal efficiency → Continuous line in two dimensional cut space



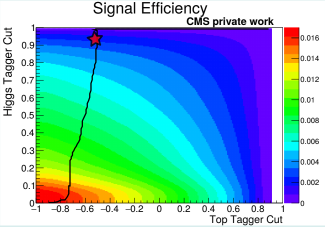
Boosted Event Selection Working Point

- Working points close to line of best cuts deliver best performance
- Choice of working point ambiguous → Optimization for best limits

Example: High Purity Working Point

- Top Identification Value ≥ -0.53
- Higgs Identification Value ≥ 0.96

Signal Eff.	Background Eff.	S/B	$S/\sqrt{B}$
0.00170	0.00004	0.014	0.26



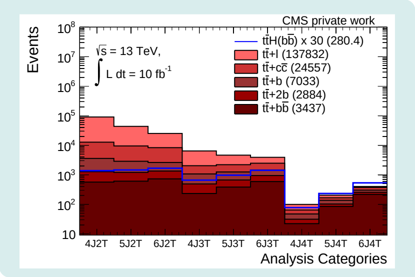
- Signal efficiency $\approx$ 40 times higher than background efficiency

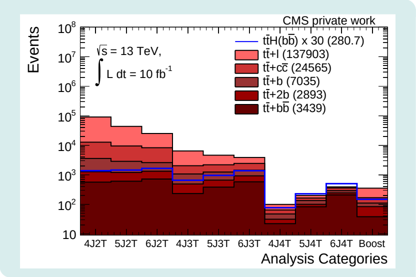
Top Rec. Eff.	Higgs Rec. Eff.	Event Rec. Eff.
69%	57%	45%

- Very high reconstruction efficiency compared to standard analysis ($\approx$ 20 %)

Event Categorization

- Boosted category covers phase space with only few events
- Reconstruction of events not resolved by standard jet analysis
- Combination of boosted analysis category with standard analysis categories





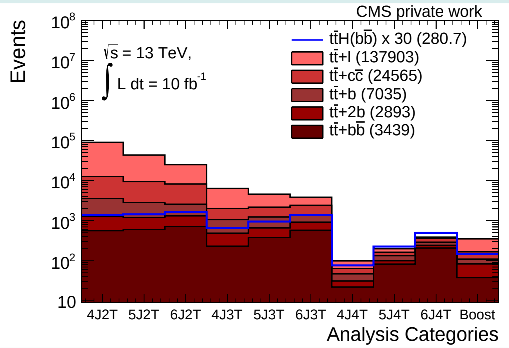
- Boosted event category overlaps with standard analysis categories
- Assign overlapping events to boosted category
  $\rightarrow$ Better reconstruction efficiency

Event Yields

Expected event yields for $\int \mathcal{L}\, dt = 10 fb^{-1}$:

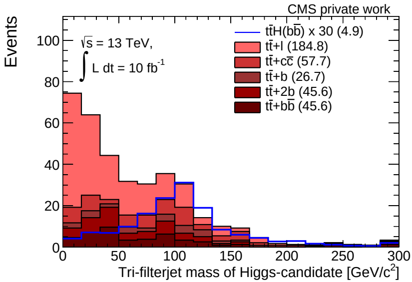
Boosted category features

- Signal event yields comparable to standard categories
- Large fraction of selected background from $t\bar{t}$ + light flavor
- Comparably strong $t\bar{t} + b\bar{b}$ background suppression



- Large $t\bar{t}$ + 2b background contribution
  - Radiated $b\bar{b}$ pair merged into one jet
  - Very likely to fake boosted Higgs boson

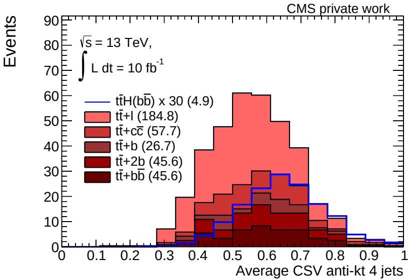
Background Discrimination



- Further background discrimination achieved by
  - Additional selection cuts
  - Final BDTs
- Variables benefit from high reconstruction efficiency

Variable Examples

- Event reconstruction variables: Mass of reconstructed Higgs candidate (upper plot)
- Std. jet b-tagging variables: Average b-tagging output of all standard jets (right plot)



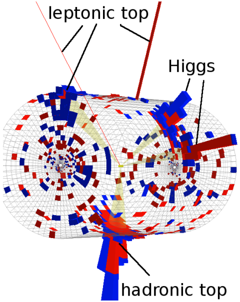
Conclusion

Subsequent Analysis Steps

Training of final BDT, fit, limit calculation and optimization in progress

The boosted $t\bar{t}H$ analysis features:

- Boosted jet reconstruction specialized on $t\bar{t}H$ search
- Optimized selection for signatures with a boosted Higgs boson and a boosted hadronic top
- High reconstruction efficiency brings further background discrimination



Outlook

- MVA Higgs tagger
- Larger cone size
- Additional boosted categories
- Full analysis

The End

Thank you for your attention!

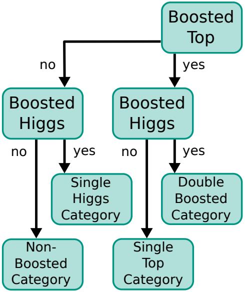
Sources:

- V. Khachatryan *et al.* [CMS Collaboration], “Search for the associated production of the Higgs boson with a top-quark pair”, JHEP **1409** (2014) 087
- T. Plehn, G. P. Salam and M. Spannowsky, “Fat Jets for a Light Higgs”, Phys. Rev. Lett. **104** (2010) 111801
- CMS Collaboration, “Boosted Top Jet Tagging at CMS,” CMS-PAS-JME-13-007
- D. E. Kaplan, K. Rehermann, M. D. Schwartz and B. Tweedie, “Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks”, Phys. Rev. Lett. **101** (2008) 142001

Backup

Analysis Strategy

1. Clustering of fat jet collections:
   - Separate jet collections for top and Higgs identification
     $\rightarrow$ Algorithms tailored to top/Higgs decay kinematics
   - Independently clustered from all particles
2. Event categorization based on boosted object selections:
   - Double boosted category
     $\rightarrow$ boosted Higgs & top candidate
   - Single boosted top categories
     $\rightarrow$ boosted top candidate & anti-k_T jets
   - Single boosted Higgs categories
     $\rightarrow$ boosted Higgs candidate & anti-k_T jets
3. Training of BDT against background processes
4. Fit to data and limit calculation



Event Yields

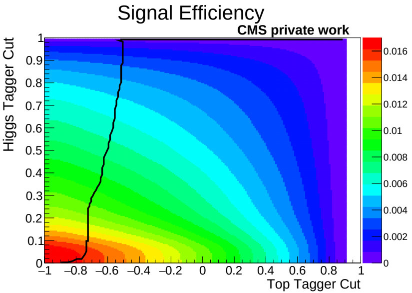
Expected event yields for $\mathcal{L} = 10fb^{-1}$:

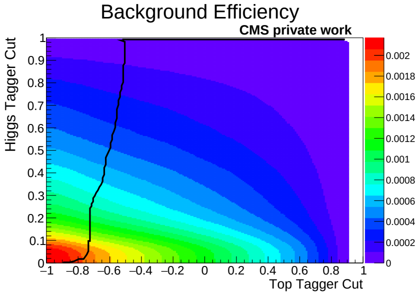
	4j3t	4j4t	5j3t	5j≥4t	≥6j2t	≥6j3t	≥6j≥4t	Boost
$t\bar{t}H$	21.8	2.6	31.9	7.6	55.0	45.9	16.7	4.9
$t\bar{t} + lf$	4404.8	35.1	2428.5	36.0	16842.9	1435.3	31.8	184.8
$t\bar{t} + b$	578.5	15.9	579.7	33.9	1265.6	512.0	47.7	26.7
$t\bar{t} + b\bar{b}$	232.0	22.2	382.9	82.8	720.6	579.7	208.6	38.0
$t\bar{t} + 2b$	257.9	9.2	279.2	17.6	603.6	343.6	35.5	45.6
$t\bar{t} + c\bar{c}$	964.3	17.1	944.7	29.7	5682.5	999.9	67.3	57.7
Total bkg	6437.5	99.5	4615.0	199.8	25115.2	3870.5	390.9	352.8

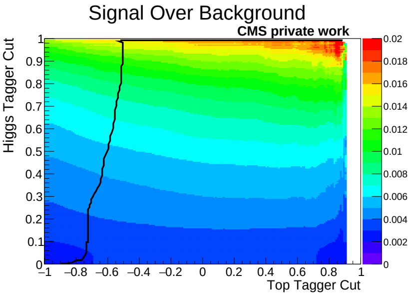
Boosted Category yields comparable with 4 jet 4 b-tag category:

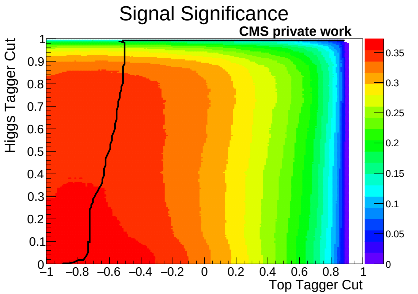
- Much more $t\bar{t}$ + light flavor background
  $\rightarrow$ Can be separated from signal by also taking anti-kt b-tagging information into account
- Less $t\bar{t}$ + $b\bar{b}$ background
- More $t\bar{t}$ + 2b backgound $\rightarrow$ Merged b jets misidentified as Higgs
- Small signal overlap with best performing categories

Selection Performance 1









Selection Performance 2