Multivariate analysis of $H \rightarrow b\bar{b}$ in associated production of H with $t\bar{t}$-pair using full simulation of ATLAS detector

Sergey Kotov

MPI für Physik, München

DPG meeting, March 28, 2006

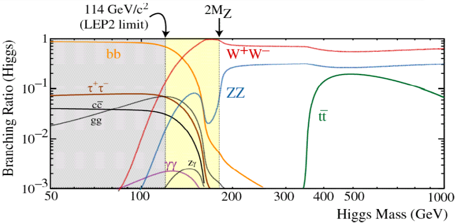
MAX-PLANCK-GESELLSCHAFT

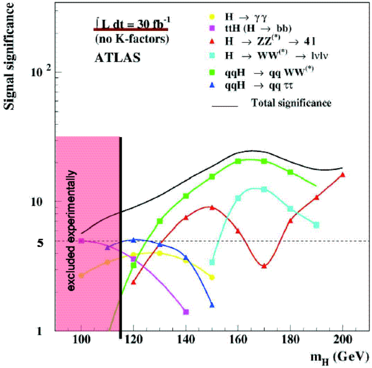
Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

Low mass SM Higgs boson overview

LEP2 experimental bounds on Higgs mass

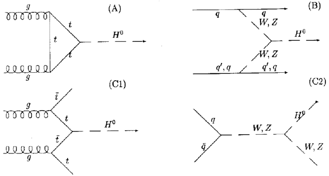
- precision measurements of EW observables: $m_H = 117^{+67}_{-45}$ GeV
- direct searches: $m_H > 114$ GeV





gg fusion

WW/ZZ fusion



associated $t\bar{t}H$

associated WH, ZH

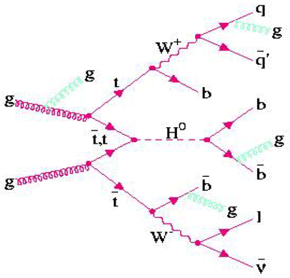
Signature channels for low mass SM Higgs

- $H \to \tau^+\tau^-$ in vector boson fusion
- $H \to \gamma\gamma$ in gluon fusion
- $H \to WW^* \to l\nu l\nu$ in vector boson fusion
- $H \to ZZ^* \to 4l$ in gluon fusion
- $H \to b\bar{b}$ in H associated production with $t\bar{t}$

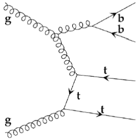
Channel description

Features of $t\bar{t}H$, $H \rightarrow b\bar{b}$ channel

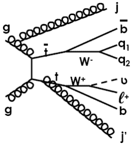
- Complex final state
  - 6 jets: 4 b-jets and 2 light jets
  - 1 high-p_t lepton (trigger)
  - missing energy from neutrino
  - additional jets from ISR/FSR
- Large backgrounds
  - combinatorial from mis-pairing of jets
  - irreducible from $t\bar{t}b\bar{b}$ events
  - reducible from $t\bar{t} + jets$ events
- Full reconstruction of event and very good b-tagging are needed



$t\bar{t}H$, $H \rightarrow b\bar{b}$ signal



$t\bar{t}b\bar{b}$ background



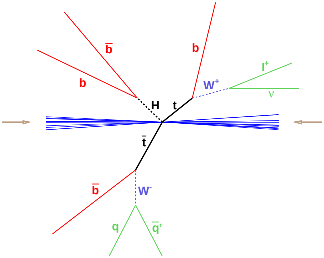
$t\bar{t}jj$ background

Expected number of events at LHC

Process	σ_{LO}, pb	BR	LHC events for $\mathcal{L}$ of 30 fb^{-1}	LHC events for $\mathcal{L}$ of 100 fb^{-1}	MC generator	FastSim sample	FullSim sample
$t\bar{t}H \rightarrow (bl\nu)(bjj)(b\bar{b})$	0.52	0.20	3.15k	10.5k	Pythia	1M	42k
$t\bar{t}b\bar{b} \rightarrow (bl\nu)(bjj)b\bar{b}$ (QCD)	8.1[a]	0.29	70.5k	235k	AcerMC	1.8M	92k
$t\bar{t}b\bar{b} \rightarrow (bl\nu)(bjj)b\bar{b}$ (EW)	0.9	0.29	7.8k	26k	AcerMC	200k	—
$t\bar{t} \rightarrow (bl\nu)(bjj) + jets$	500	0.29	4.3M	14.5M	Pythia	4M	327K

[a] Strongly depends upon factorization scale (up to a factor of 2). Here, $\mu_0 = (m_t + m_H)/2$

Event reconstruction: preselection



Preselection cuts

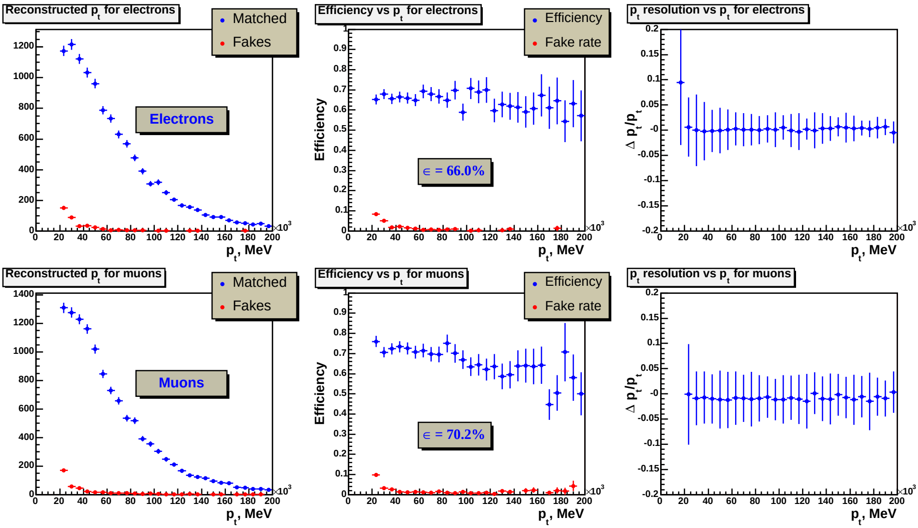
- $\geq$ 1 isolated leptons
  - p_t>20 GeV and $|\eta|$<2.7
  - $E_t < 10$ GeV within the isolation cone of $\Delta R = 0.4$
  - e-Id: EM cluster has a matched track in ID and the cluster shape is consistent with e-hypothesis
  - μ-Id: the combined fit of muon track has has good quality
- $\geq$ 4 b-jets
  - p_t>15 GeV and $|\eta|$<3
  - standard ATLAS b-tagging cut: $jetWeight > 3$
- $\geq$ 2 light jets
  - p_t>15 GeV and $|\eta|$<3
  - b-tagging cut (anti-b-tag): $jetWeight < 0.1$

Preselection efficiencies

Particle	Kinematical acceptance, %	Reconstruction efficiency, %
e	82.8	66.0
μ	82.3	70.2
b-jet	93.4	42.9
light jet	48.6	52.2

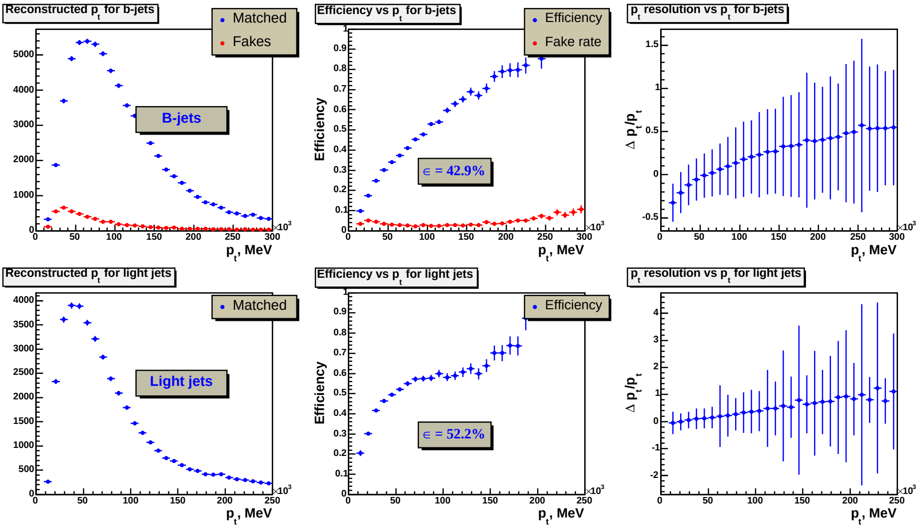
- 15 GeV p_t cut on light jets required by anti-b-tagging algorithm considerably decreases kinematical acceptance
- b-tagging algorithm has 60% efficiency, but there're fewer than expected reconstructed jets to tag

Reconstruction efficiencies and resolutions: leptons



- due to high jet activity the efficiencies are somewhat lower than expected
- the p_t resolutions are $\sim$ 4% for electrons and $\sim$ 5% for muons

Reconstruction efficiencies and resolutions: jets



- in high jet multiplicity events overlapping of jets considerably deteriorates jet energy calibration and resolution

Event reconstruction: making combinations

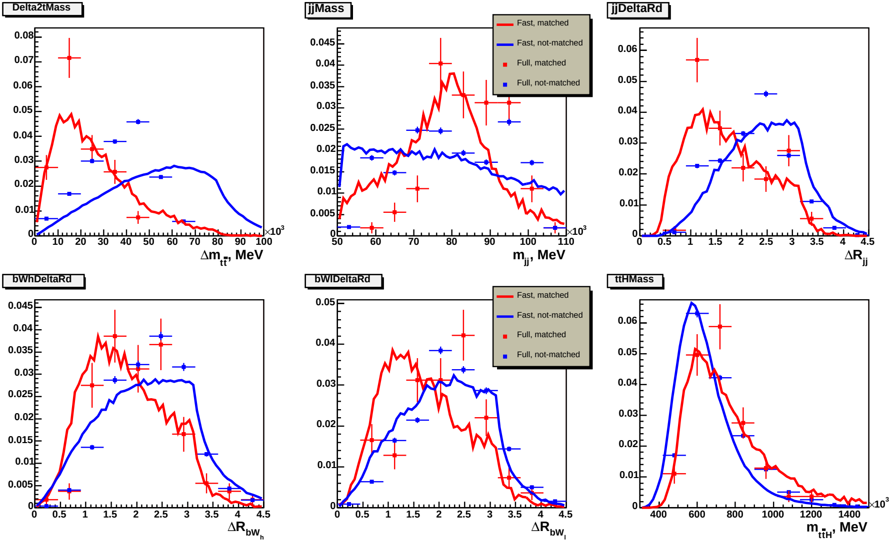
Making 4 b-jet + 2 light jets + 1 lepton combinations and selecting the best one

- use events which pass preselection criteria (1 lepton, 4 b-jets, 2 light jets)
- determine p_ν from p_l and p_{miss} using m_W constraint (if fails, use approximation $p_\nu^z = p_l^z$)
- reconstruct “leptonic” $W \to l\nu$ from lepton and neutrinos
- reconstruct “hadronic” $W \to jj$ from jj combinations with $|m_{jj} - m_W| <$ 35 GeV (the jets 4-momenta rescaled to get the nominal W mass)
- permute over all combinations of reconstructed W_{lep}, W_{had}, and 4 b-jets
- calculate the evaluation parameter for each combination
- from each event select the combination with the highest value of this parameter
- plot invariant mass distributions from these best combinations and look for a Higgs peak

Various evaluation parameters of $t\bar{t}$-pair reconstruction

- ATLAS TDR: $\Delta m_{t\bar{t}} = \sqrt{(m_{bl\nu} - m_t)^2 + (m_{bjj} - m_t)^2}$
- $t\bar{t}$-pair likelihood in ATL-PHYS-2003-024 analysis
- this analysis uses neural network evaluation parameter

ANN variables: Full simulation vs Fast simulation signal samples



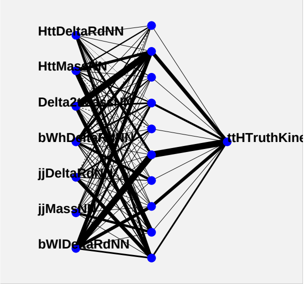
- most powerful discriminating variables are $\Delta m_{t\bar{t}}$ and ΔR_{jj}

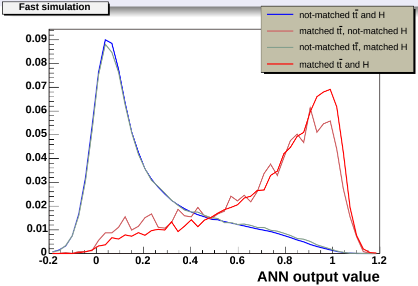
The neural network structure and performance

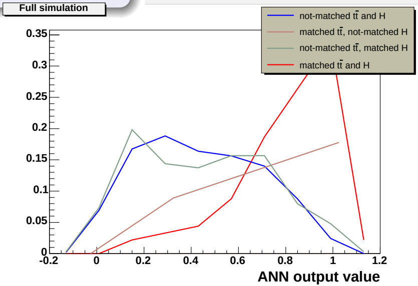
Due to limited size of the full simulation sample, fast simulation sample was used to train the ANN

ANN variables

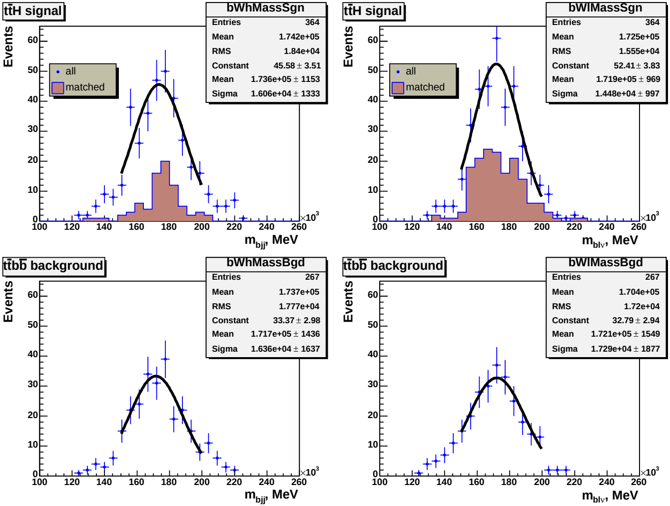
- TDR's evaluator, $\Delta m_{t\bar{t}} = \sqrt{(m_{bl\nu} - m_t)^2 + (m_{bjj} - m_t)^2}$
- invariant mass of two light jets from W_{had}
- invariant mass of $t\bar{t}$-H system
- ΔR between two light jets from W_{had}
- ΔR between b-jet and W_{had} from the same t-quark
- ΔR between b-jet and W_{lep} from the same t-quark
- ΔR between $t\bar{t}$ system and Higgs





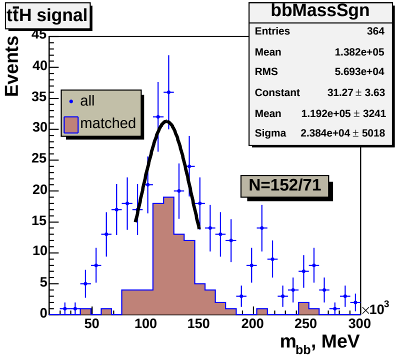


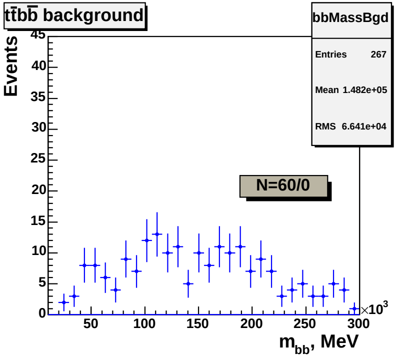
Reconstructed $t\bar{t}$-pair invariant mass distributions



- there's a small shift of ∼3 GeV in the reconstructed m_t
- the width of the reconstructed m_t is ∼16 GeV

Reconstructed Higgs invariant mass distributions



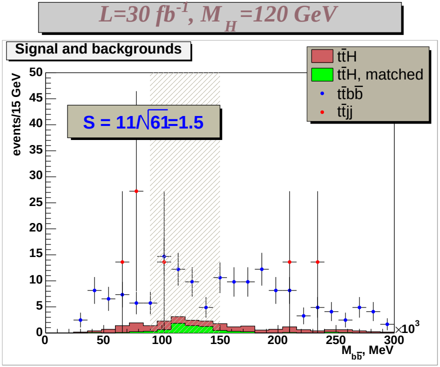


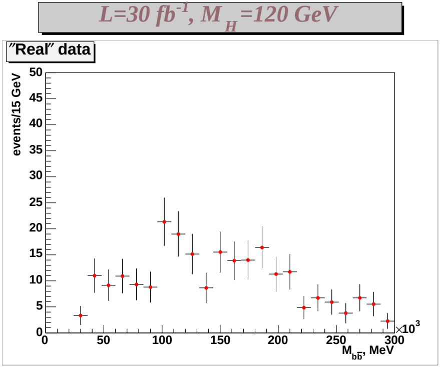
Efficiencies

	$t\bar{t}H$ sample		$t\bar{t}b\bar{b}$ sample	
	ϵ, %	Events	ϵ, %	Events
	100	42882	100	96053
W_{lep}	55.5	23810	58.3	56038
W_{had}	32.3	13834	37.9	36455
4 b-jets	0.84	362	0.28	267
m_H window	0.35	152	0.06	60

- the shape of the irreducible background is reasonably flat
- the reconstructed Higgs mass is close to the nominal with the width of $\sim$24 GeV

Expected signal after 30fb^{-1} of luminosity





Signal significance estimate

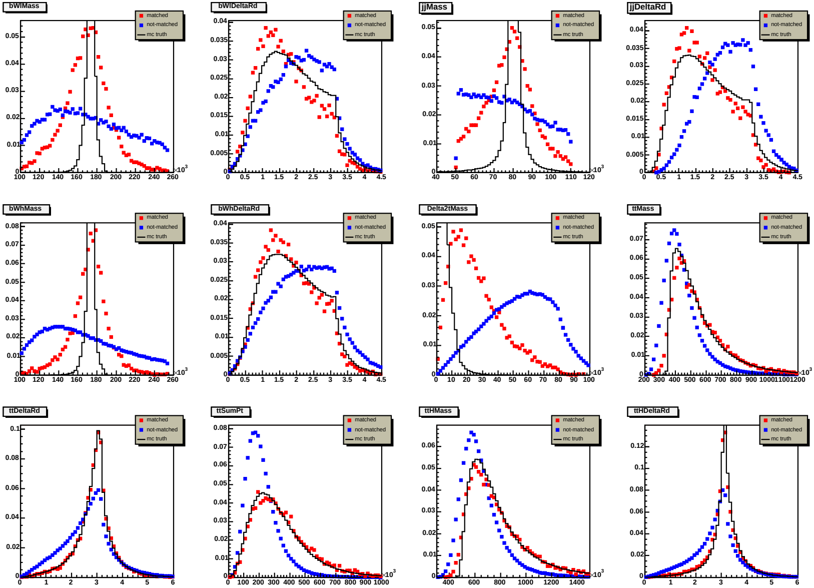
	$t\bar{t}H$	$t\bar{t}b\bar{b}$	$t\bar{t}jj$
Events in $m_H \pm$30 GeV window	152	60	1
Final efficiency, %	0.35	0.06	0.0003
Events normalized to 30 (100) fb^{-1}	11 (38)	48 (160)	13 (45)
Signal significance	1.5 (2.7)		

- it's hard to extract the signal, unless the background shape is well known from MC

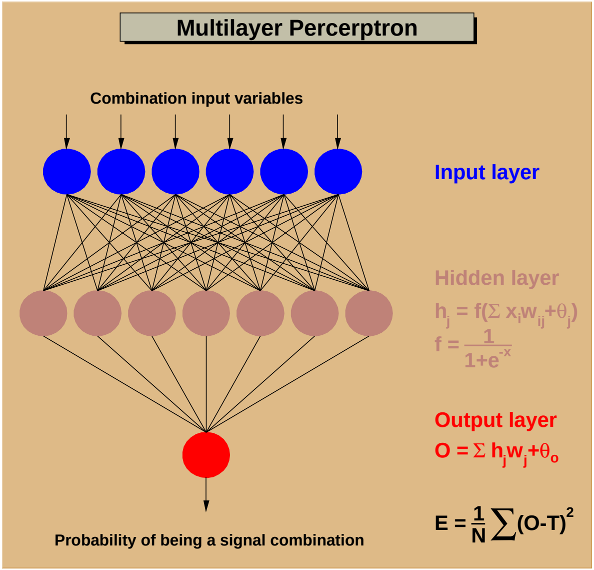
Conclusions and plans

- signal significance for $t\bar{t}H, H \to b\bar{b}$ channel in this study comes out quite low: $S = 1.5$ for 30 fb^{-1} of integrated luminosity (ATLAS TDR had $S = 3$)
- it would be difficult to extract the $H \to b\bar{b}$ signal from data without good understanding of the background shape
- ANN gives a small improvement in signal significance over standard TDR evaluator (~4%)
- still a lot of things can be done to improve the signal significance
  - smarter jet reconstruction algorithms are needed to deal with overlapping of jets in high jet multiplicity events (smaller jet cone size, TopoCluster jets)
  - b-jet reconstruction efficiency can be increased by loosening the b-tagging cut, with the downside of reduced suppression against $t\bar{t}jj$ background $\to$ room for optimisation
  - with more statistics, the neural network can be retrained on full simulation data

Distributions of the ANN variables in signal sample: Fast simulation



Neural network basics



- TMultiLayerPerceptron ROOT built-in class is used as neural network (1 hidden layer with 10 nodes)
- 11500 of matched and 12500 of non-matched combinations were used to train the neural network