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

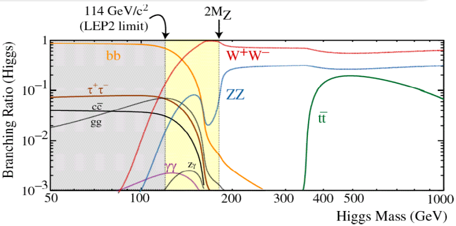
ATLAS-D Top meeting, May 19, 2006

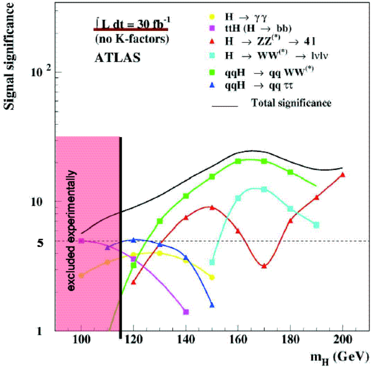
Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

Low mass SM Higgs boson overview

LEP2 experimental bounds on Higgs mass

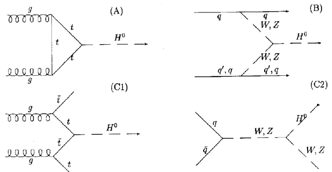
- precision measurements of EW observables: $m_H = 117^{+67}_{-45}$ GeV
- direct searches: $m_H > 114$ GeV





gg fusion

WW/ZZ fusion



associated $t\bar{t}H$

associated WH, ZH

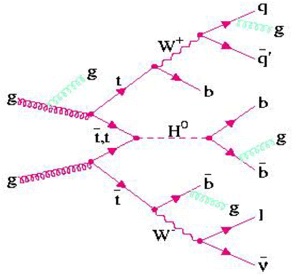
Signature channels for low mass SM Higgs

- $H \to \tau^+\tau^-$ 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 $t\bar{t}$

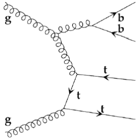
Channel description

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

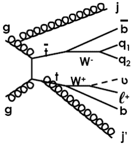
- Complex final state
  - 6 jets: 4 b-jets and 2 light jets
  - 1 high-p_t lepton (trigger)
  - missing energy from neutrino
  - additional jets from ISR/FSR
- Large backgrounds
  - combinatorial from mis-pairing of jets
  - irreducible from $t\bar{t}b\bar{b}$ events
  - reducible from $t\bar{t} + jets$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



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



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



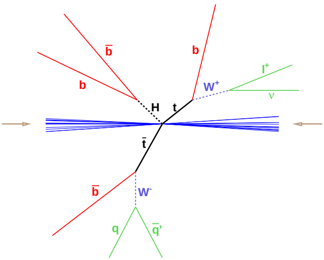
$t\bar{t}jj$ background

Expected number of events at LHC

Process	σ_{LO}, pb	BR	LHC events for $\mathcal{L}$ of		MC generator	FastSim sample	FullSim sample
			30 fb^{-1}	100 fb^{-1}			
$t\bar{t}H \to (bl\nu)(bjj)(b\bar{b})$	0.52	0.20	3.15k	10.5k	Pythia	1M	42k
$t\bar{t}b\bar{b} \to (bl\nu)(bjj)b\bar{b}$ (QCD)	8.1[a]	0.29	70.5k	235k	AcerMC	1.8M	72k
$t\bar{t}b\bar{b} \to (bl\nu)(bjj)b\bar{b}$ (EW)	0.9	0.29	7.8k	26k	AcerMC	200k	—
$t\bar{t} \to (bl\nu)(bjj) + jj$	420	0.29	3.6M	12M	Alpgen	1M	30K

[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
  - AOD containers "MuonCollection", "ElectronCollection"
  - 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
  - AOD container "BJetCollection"
  - p_t>15 GeV and $|\eta|$<3
  - standard ATLAS b-tagging cut: $jetWeight >$ 3
- $\geq$ 2 light jets
  - AOD container "BJetCollection"
  - p_t>15 GeV and $|\eta|$<3
  - b-tagging cut (anti-b-tag): $jetWeight <$ 0.1

- several jet reconstruction algorithms are available (Cone4, Cone7, Kt)
- Cone4 and Cone7 algorithms will be compared throughout this talk

Event reconstruction: making combinations

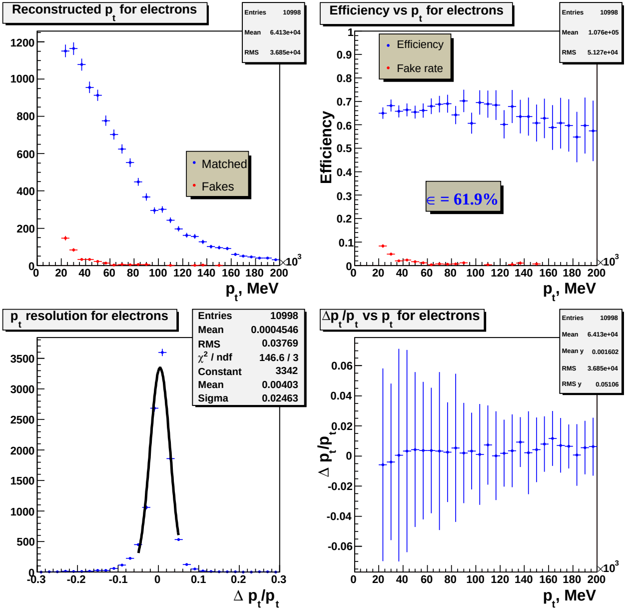
Making 4 b-jet + 2 light jets + 1 lepton combinations and selecting the best one

- use events which pass preselection criteria (1 isolated lepton, 4 b-jets, 2 light jets)
- determine p_ν from p_l and p_{miss} using m_W constraint (if fails, use approximation $p^z_\nu = p^z_l$)
- 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

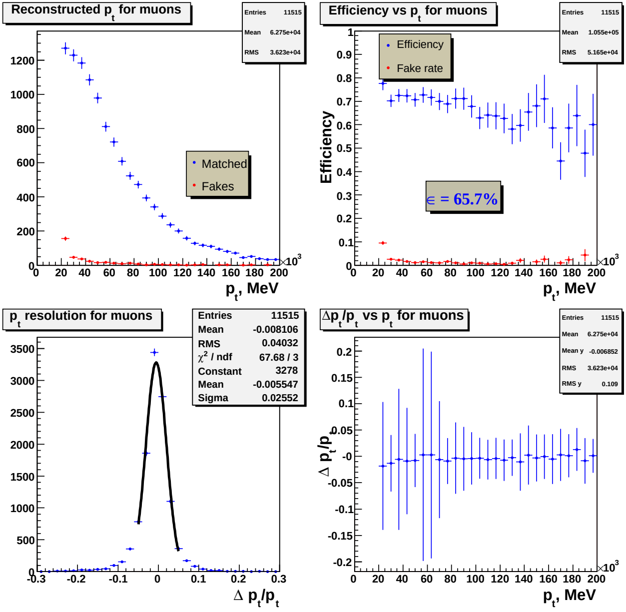
- ATLAS TDR: $\Delta m_{t\bar{t}} = \sqrt{(m_{bl\nu} - m_t)^2 + (m_{bjj} - m_t)^2}$
- $t\bar{t}$-pair likelihood in ATL-PHYS-2003-024 analysis
- this analysis uses neural network evaluation parameter

Reconstruction efficiency and resolution: electrons



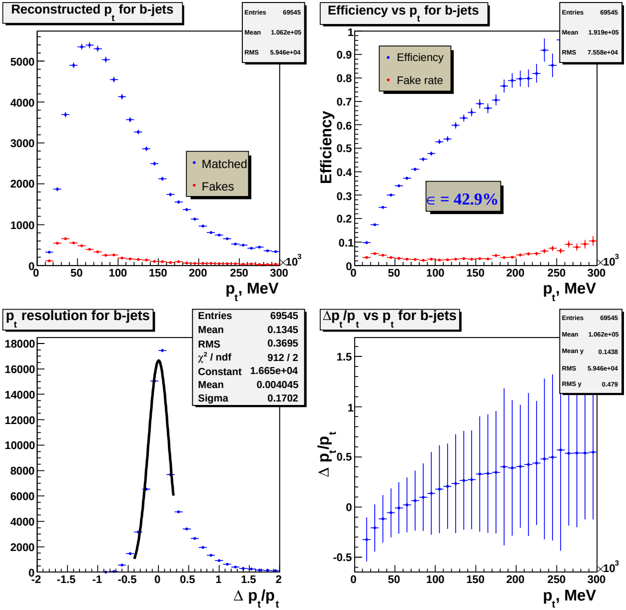
- p_t mean value $\sim$ 64 GeV
- average efficiency $\sim$ 62%
- mean p_t shift $\sim$ 0.4%
- p_t resolution $\sim$ 2.5%

Reconstruction efficiency and resolution: muons



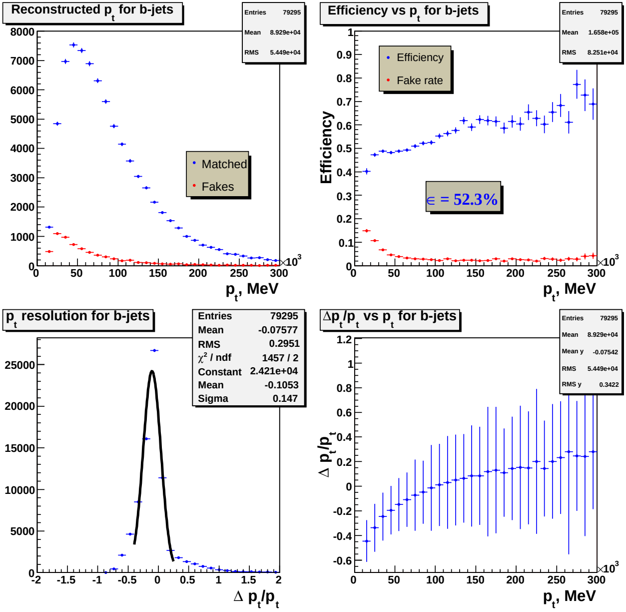
- p_t mean value $\sim$ 63 GeV
- average efficiency $\sim$ 66%
- mean p_t shift $\sim -0.5\%$
- p_t resolution $\sim$ 2.5%

Reconstruction efficiency and resolution: b-jets, JetCone=0.7



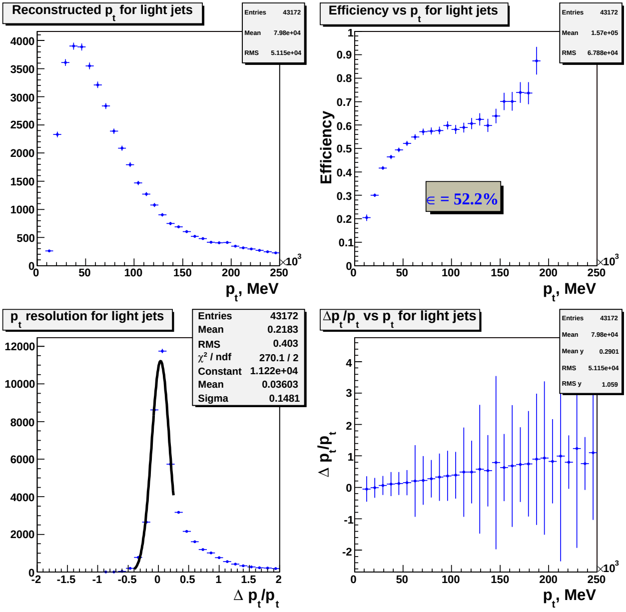
- p_t mean value $\sim$ 106 GeV
- average efficiency $\sim$ 43%
- mean p_t shift $\sim$ 0.4%
- p_t resolution $\sim$ 17%

Reconstruction efficiency and resolution: b-jets, JetCone=0.4



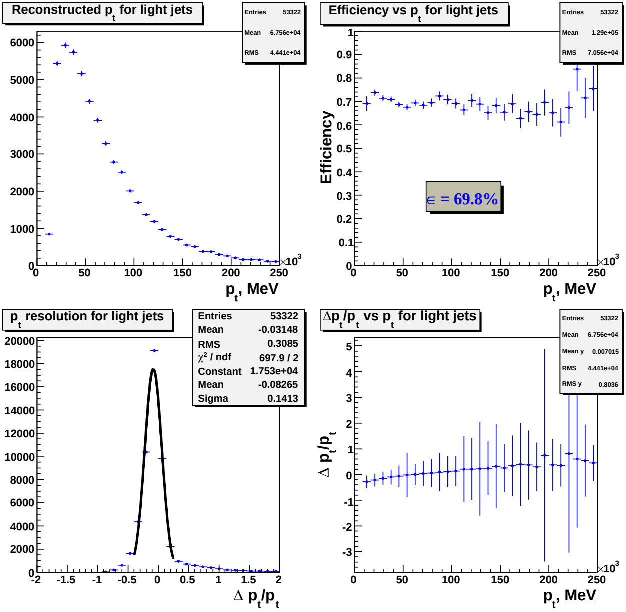
- p_t mean value $\sim$ 89 GeV
- average efficiency $\sim$ 52%
- mean p_t shift $\sim -10\%$
- p_t resolution $\sim$ 15%

Reconstruction efficiency and resolution: light jets, JetCone=0.7



- p_t mean value $\sim$ 80 GeV
- average efficiency $\sim$ 52%
- mean p_t shift $\sim$ 4%
- p_t resolution $\sim$ 15%

Reconstruction efficiency and resolution: light jets, JetCone=0.4



- p_t mean value $\sim$ 68 GeV
- average efficiency $\sim$ 70%
- mean p_t shift $\sim -8\%$
- p_t resolution $\sim$ 14%

Reconstruction efficiencies

Preselection efficiencies for JetCone=0.4 (JetCone=0.7)

Particle	$t\bar{t}H$ Kinematical acceptance, %	$t\bar{t}H$ Reconstruction efficiency, %	$t\bar{t}b\bar{b}$ Kinematical acceptance, %	$t\bar{t}b\bar{b}$ Reconstruction efficiency, %
e	88.1 (88.4)	61.9 (61.8)	88.3 (88.4)	65.9 (66.2)
μ	88.1 (88.0)	65.7 (65.8)	88.3 (88.3)	69.1 (69.1)
b-jet	93.4 (93.4)	52.3 (42.9)	75.7 (75.7)	47.5 (41.1)
light jet	48.7 (48.6)	69.8 (52.2)	59.6 (59.6)	76.0 (61.9)

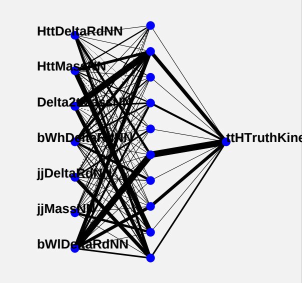
- 15 GeV p_t cut on light jets required by anti-b-tagging algorithm considerably decreases kinematical acceptance
- standard ATLAS b-tagging algorithm has 60% efficiency on average
- using Cone4 over Cone7 algorithm increases average jet multiplicity from 5.7 to 7.1 which means an overall improvement factor of 1.2-1.35 in jet reconstruction efficiencies
- on the other hand, jets reconstructed with Cone4 algorithm have their p_t underestimated by $\sim$ 10% which means negative shifts in invariant mass distributions

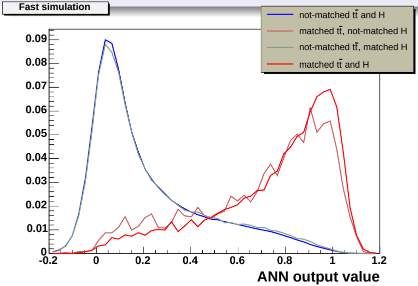
The neural network structure and performance

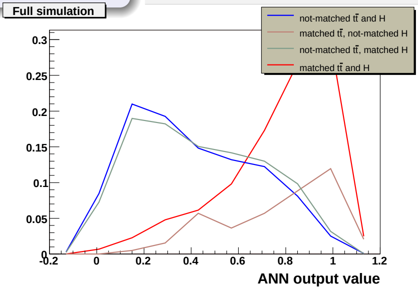
Due to limited size of the full simulation sample, fast simulation sample was used to train the ANN

ANN variables

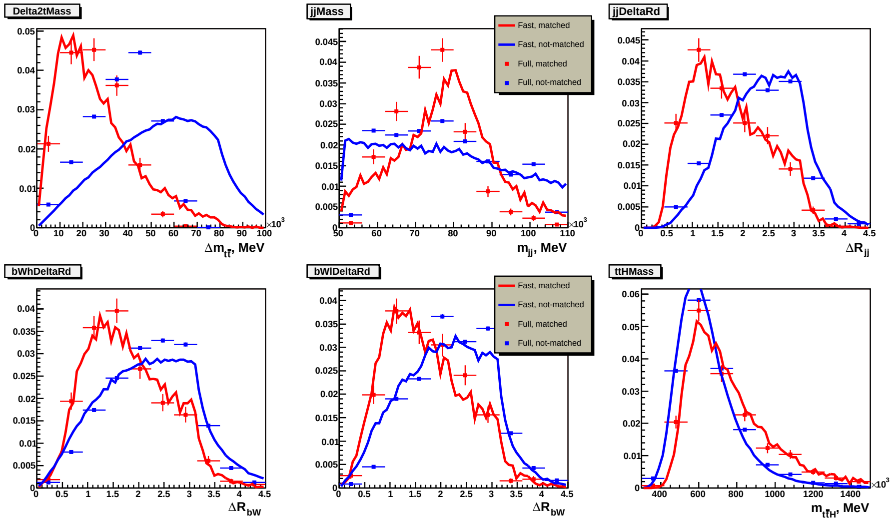
- TDR's evaluator, $\Delta m_{t\bar{t}} = \sqrt{(m_{bl\nu} - m_t)^2 + (m_{bjj} - m_t)^2}$
- invariant mass of two light jets from W_{had}
- invariant mass of $t\bar{t}$-H system
- ΔR between two light jets from W_{had}
- ΔR between b-jet and W_{had} from the same t-quark
- ΔR between b-jet and W_{lep}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





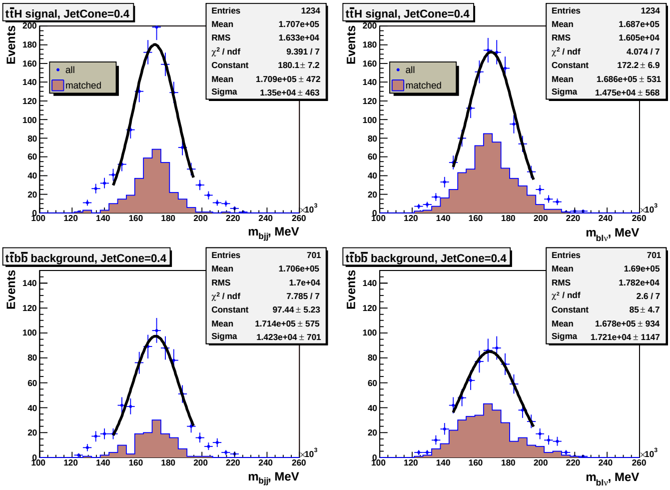


ANN variables: Full simulation vs Fast simulation signal samples



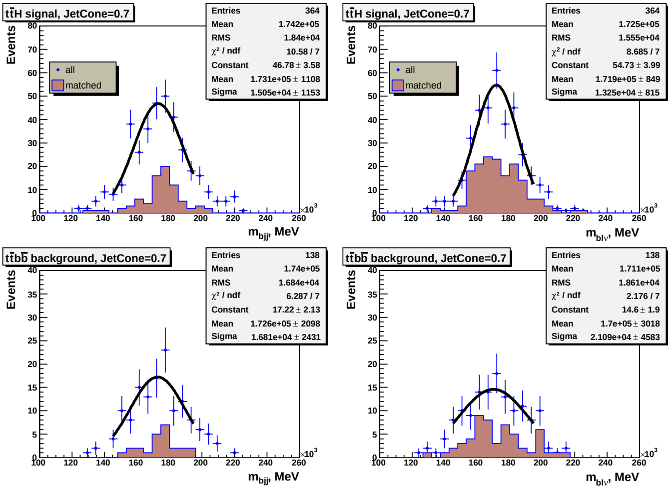
- most powerful discriminating variables are $\Delta m_{t\bar{t}}$ and ΔR_{jj}
- in full simulation $\Delta m_{t\bar{t}}$ variable has less power than in fast simulation

Reconstructed $t\bar{t}$-pair invariant mass distributions: JetCone=0.4



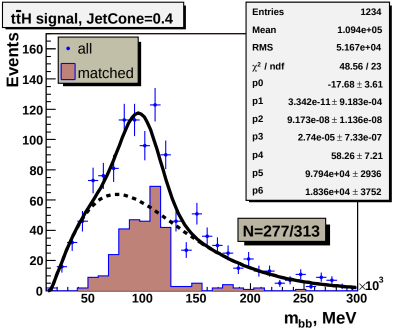
- shift in the reconstructed $m_t \sim -5$ GeV
- reconstructed m_t width $\sim$ 14 GeV

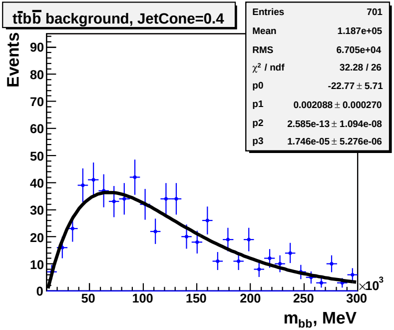
Reconstructed $t\bar{t}$-pair invariant mass distributions: JetCone=0.7



- shift in the reconstructed $m_t \sim -3$ GeV
- reconstructed m_t width $\sim$ 15 GeV

Reconstructed Higgs invariant mass distributions: JetCone=0.4



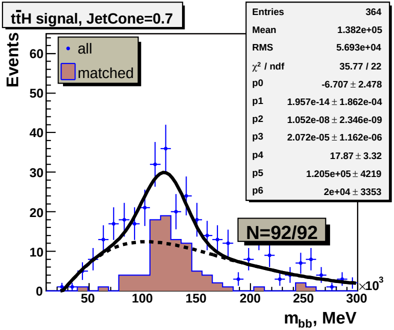


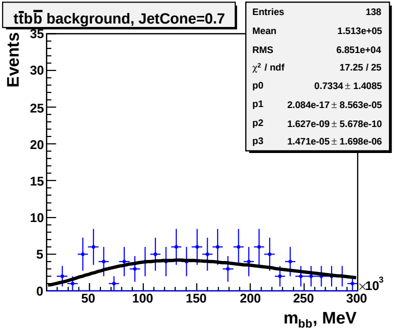
Efficiencies

	$t\bar{t}H$ sample		$t\bar{t}b\bar{b}$ sample	
	ϵ, %	Events	ϵ, %	Events
	100	40132	100	71993
W_{lep}	56.8	22776	60.1	43275
W_{had}	44.4	17816	47.3	34050
4 b-jets	3.17	1271	1.03	741
fit area	0.7	277	—	—

- the shape of distributions reasonably well described by function $(a_0 + a_1 x + a_2 x^2)e^{-bx} + Gaussian$
- the reconstructed Higgs mass is shifted from the nominal by ∼22 GeV
- Higgs mass width is ∼18 GeV

Reconstructed Higgs invariant mass distributions: JetCone=0.7



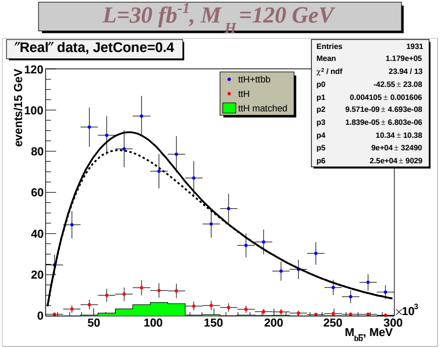


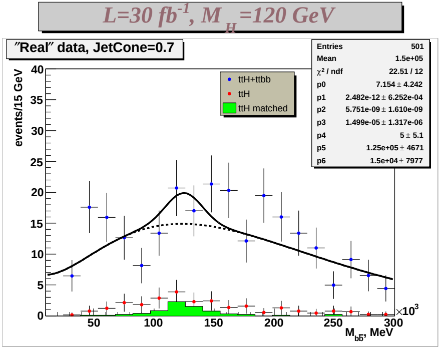
Efficiencies

	$t\bar{t}H$ sample		$t\bar{t}b\bar{b}$ sample	
	ϵ, %	Events	ϵ, %	Events
	100	41882	100	48349
W_{lep}	56.8	23810	60.2	29122
W_{had}	33.0	13834	38.8	18737
4 b-jets	0.92	388	0.32	153
fit area	0.22	92	—	—

- $(a_0 + a_1 x + a_2 x^2)e^{-bx} + Gaussian$ function still describes distributions quite well
- the reconstructed Higgs mass is close to the nominal
- Higgs mass width is ∼20 GeV

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



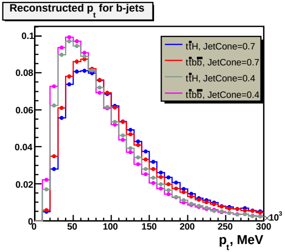


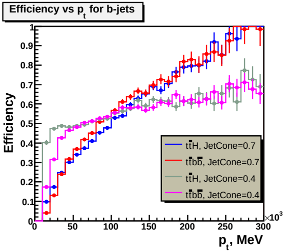
- it's hard to extract the signal, unless the background shape is well known from MC, so that one can constrain background contribution to the fit

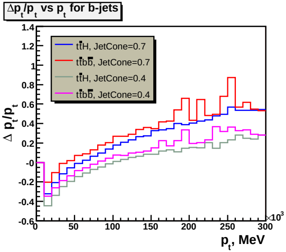
Conclusions and plans

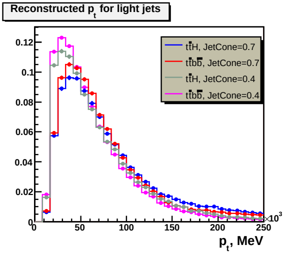
- final event reconstruction efficiency strongly depends upon jet reconstruction algorithm (a factor of 3 between JetCone4 and JetCone7)
- standard ATLAS b-tagging algorithm performance is satisfactory
- it would be difficult to extract the $H \to b\bar{b}$ signal from data without good understanding of the background shape
- things to do
  - determine the shape of reducable $t\bar{t}jj$ background (still not enough full simulation samples)
  - rerun the analysis with Kt and TopoCluster jet reconstruction
  - with more statistics, retrain the neural network on full simulation data

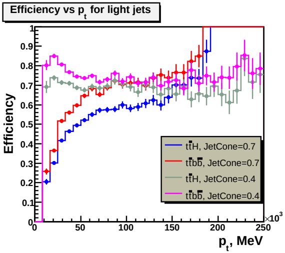
Efficiencies and resolutions: JetCone=0.4 vs JetCone=0.7

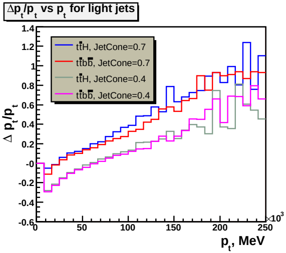




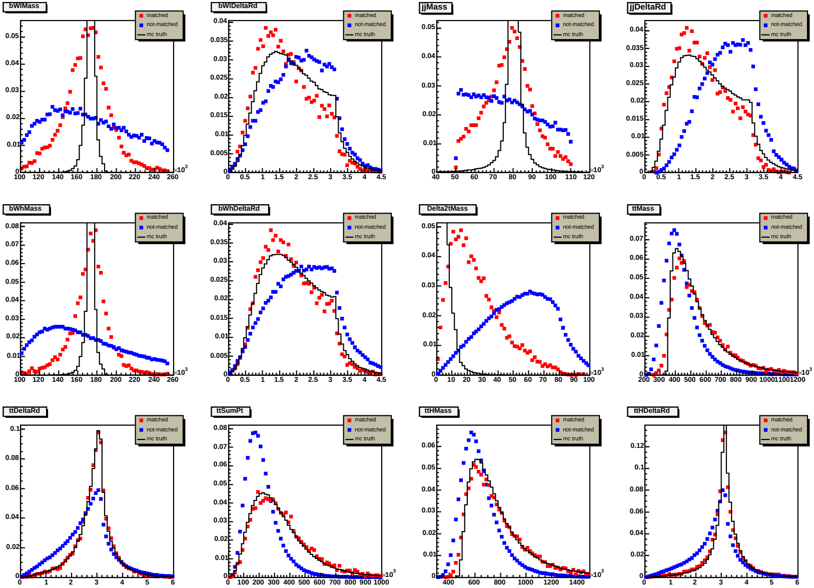




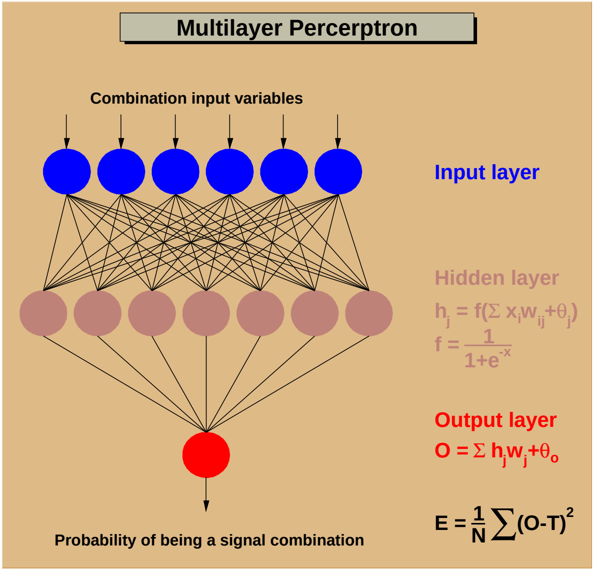




Distributions of the ANN variables in signal sample: Fast simulation



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