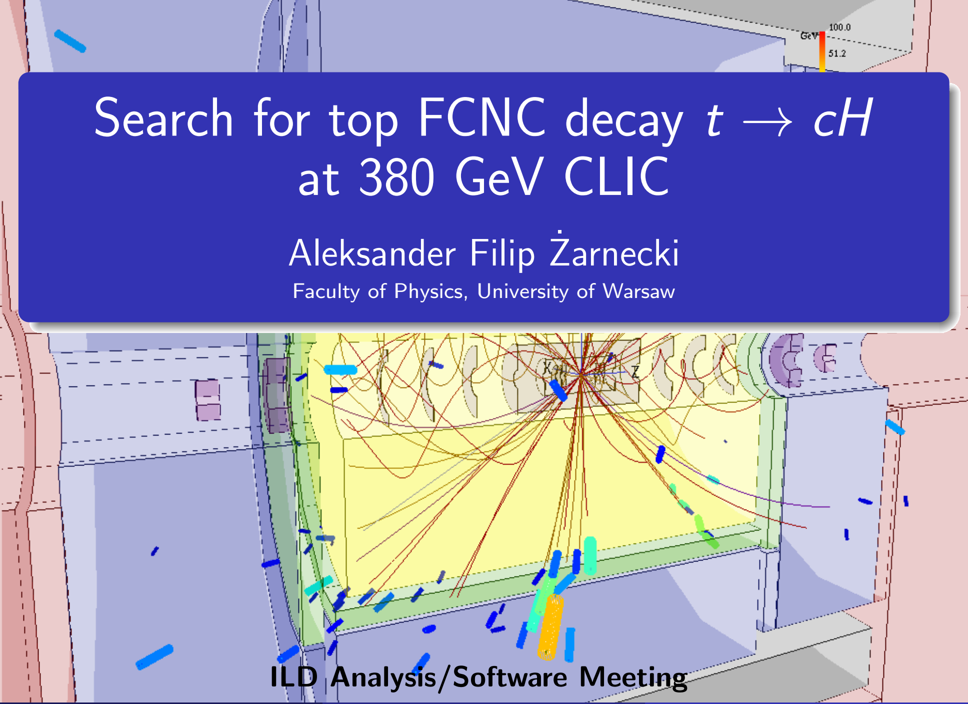


Search for top FCNC decay $t \to cH$ at 380 GeV CLIC

Aleksander Filip Żarnecki
Faculty of Physics, University of Warsaw

ILD Analysis/Software Meeting

Outline

1. Analysis overview
   - Motivation
   - Event simulation and reconstruction
   - Selection and classification of $t\bar{t}$ events
   - Kinematic fit
   - FCNC event selection

2. Some technical details
   - Clustering quality estimate
   - b-jet energy correction
   - BDT optimisation
   - LCFI+ problem

Motivation

In the Standard Model, FCNC top decays are strongly suppressed (CKM+GIM):

$$
\begin{aligned}
BR(t \rightarrow c\,\gamma) &\sim 5 \cdot 10^{-14} \\
BR(t \rightarrow c\,Z) &\sim 1 \cdot 10^{-14} \\
BR(t \rightarrow c\,g) &\sim 5 \cdot 10^{-12} \\
BR(t \rightarrow c\,H) &\sim 3 \cdot 10^{-15}
\end{aligned}
$$

Any signal is a direct signature of "new physics"...

Decay $t \rightarrow c\,H$ is most interesting:

- well constrained kinematics
- test of Higgs boson couplings
- seems to be most difficult for LHC

Estimated HL-LHC reach:
(Snowmass 2013/ATLAS 2016)

$$BR(t \rightarrow qH) \sim 2 \cdot 10^{-4}$$

Two Higgs Doublet Model (2HDM) as a test scenario:

- one of simplest extensions of the SM
- $BR(t \rightarrow c\,H)$ up to 10^{-2} (tree level) and 10^{-4} (loop level)

Full simulation for CLIC @ 380 GeV

Dedicated samples generated with WHIZARD 2.2.8
Signal: SARAH implementation of 2HDM(III), BR($t \rightarrow ch_1$) $= 10^{-3}$

Beam spectra for CLIC taken from file (350 GeV scaled to 380 GeV)
Beam polarization of -80%/0% (for e^-/e^+)

Hadronization done in PYTHIA 6.427
quark masses and PYTHIA settings adjusted to CLIC CDR
Standard event processing with CLIC_ILD_CDR500 configuration

Samples considered in the study

- dedicated FCNC signal sample $e^+e^- \longrightarrow cH\bar{t},\ t\bar{c}H$
  Higgs boson decay restricted to $H \rightarrow b\bar{b}$
- test sample of SM background $e^+e^- \longrightarrow t\bar{t}$ for simulation validation
- full 6-fermion sample as produced for CLIC $t\bar{t}$ studies
- 4-fermion and quark-pair samples (recently included in the analysis)

Event processing

DST files processed with MARLIN, ilcsoft v01-17-09 (ilcDIRAC)

- Using input PFO collection with loose timing cuts
  LooseSelectedPandoraPFANewPFOs - default for 380 GeV
- Isolated lepton identification IsolatedLeptonFinder
- LcfiPlus v00-07
  - primary and secondary vertex finder
  - jet finding with Valencia algorithm
  - vertex corrections and flavour taging

Analysis steps on root level:

- pre-selection and event classification
  selection of hadronic and semi-leptonic $t\bar{t}$ candidates
- kinematic fit for SM decay and FCNC decay hypotheses
- final signal-background discrimination

Event samples

Signal and background samples considered in the analysis.

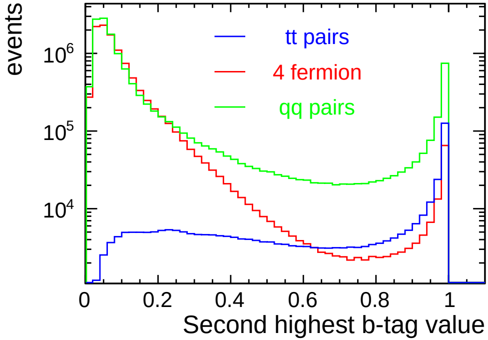
All samples processed with standard CLICdp simulation and analysis chain.
Assuming 500 fb^{-1} collected at 380 GeV, with polarization of -80%/0%.
FCNC signal for $BR(t \to cH) \times BR(H \to b\bar{b}) = 10^{-3}$

Sample	Cross section	Expected events	MC event sample
FCNC signal	1.64 fb	819	99 301
6 fermion	820 fb	410 000	1 014 966
4 fermion	21 pb	10 500 000	7 067 836
quark pair	26 pb	13 000 000	2 968 551

First analysis stage focused on reduction of huge non-$t\bar{t}$ backgrounds

Initial selection cut based on LCFI+ flavour tagging

To suppress non-$t\bar{t}$ background contribution, two jets are required to have b-tag of at least 0.2 (from 6-jet or from 4-jet final state reconstruction)



Removes 80% of $q\bar{q}$ events and 92% of 4-fermion sample.
FCNC signal efficiency of about 98% (90% for SM $t\bar{t}$ sample).

Two signal channels: fully hadronic and semi-leptonic $t\bar{t}$ events

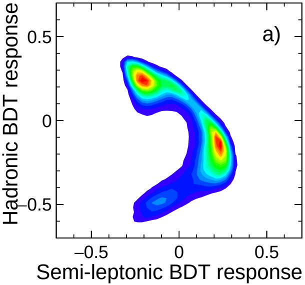
Classificantion:

used two BDTs for event selection: "hadronic" and "semi-leptonic"

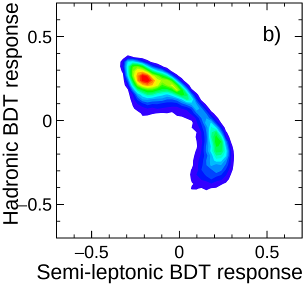
based on total energy-momentum, event shape and jet parameters (y_{min}, y_{max}), lepton ID

$\Rightarrow$ improved efficiency/purity, as compared to cut-based approach

SM $t\bar{t}$ sample



FCNC $t\bar{t}$ decays



Event classification

Two signal channels: fully hadronic and semi-leptonic $t\bar{t}$ events

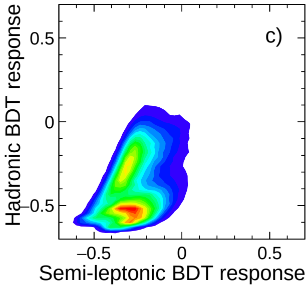
Classificantion:

used two BDTs for event selection: "hadronic" and "semi-leptonic"

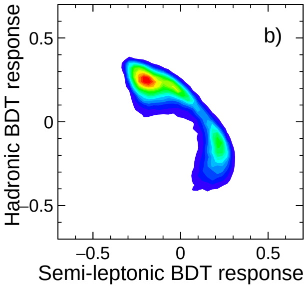
based on total energy-momentum, event shape and jet parameters (y_{min}, y_{max}), lepton ID

$\Rightarrow$ improved efficiency/purity, efficient rejection of non-$t\bar{t}$ background

SM 4-fermion sample



FCNC $t\bar{t}$ decays



Two signal channels: fully hadronic and semi-leptonic $t\bar{t}$ events

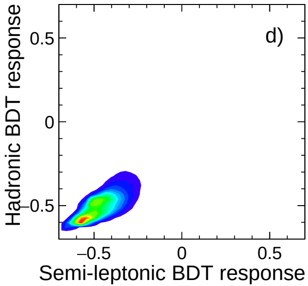
Classificantion:

used two BDTs for event selection: "hadronic" and "semi-leptonic"

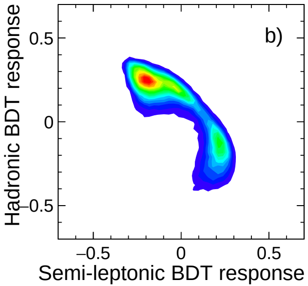
based on total energy-momentum, event shape and jet parameters (y_{min}, y_{max}), lepton ID

$\Rightarrow$ improved efficiency/purity, efficient rejection of non-$t\bar{t}$ background

SM $q\bar{q}$ sample



FCNC $t\bar{t}$ decays



Event classification

Two signal channels: fully hadronic and semi-leptonic $t\bar{t}$ events

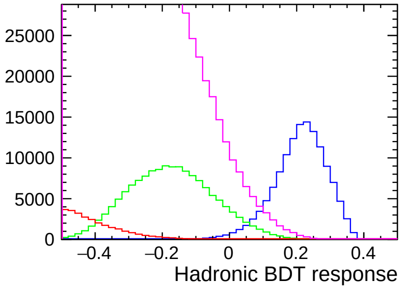
Classificantion:

used two BDTs for event selection: "hadronic" and "semi-leptonic"

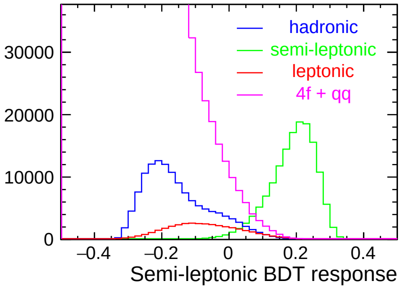
based on total energy-momentum, event shape and jet parameters (y_{min}, y_{max}), lepton ID

$\Rightarrow$ improved efficiency/purity, efficient rejection of non-$t\bar{t}$ background

Hadronic sample selection



Semi-leptonic sample selection



Kinematic fit

Signal hypothesis: three jets are required to have b-tag > 0.4
fourth jet required to have c-tag + b-tag > 0.4

χ^2 definition   for hadronic events

Mass ratios used to reduce influence of mass correlations

- signal hypothesis   top boost as additional constrain

$$\chi^2_{sig} = \left(\frac{M_{bqq} - m_t}{\sigma_t}\right)^2 + \left(\frac{M_{bbc} - m_t}{\sigma_t}\right)^2 + \left(\frac{\frac{E_{bqq}}{M_{bqq}} - \gamma_t}{\sigma_\gamma}\right)^2 + \left(\frac{\frac{E_{bbc}}{M_{bbc}} - \gamma_t}{\sigma_\gamma}\right)^2 + \left(\frac{\frac{M_{qq}}{M_{bqq}} - \frac{m_W}{m_t}}{\sigma_{R_W}}\right)^2 + \left(\frac{\frac{M_{bb}}{M_{bbc}} - \frac{m_h}{m_t}}{\sigma_{R_h}}\right)^2$$

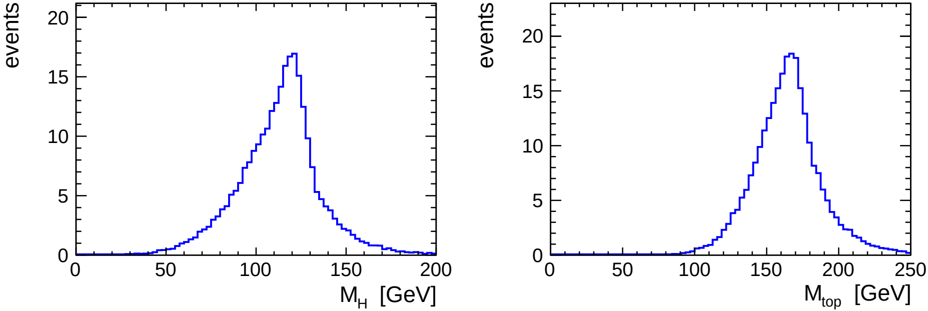
- similar for background hypothesis ($t\bar{t}$ hadronic decays)

$$\chi^2_{bg} = \quad \ldots \quad + \left(\frac{\frac{M_{qq}}{M_{bqq}} - \frac{m_W}{m_t}}{\sigma_{R_W}}\right)^2 + \left(\frac{\frac{M_{bq}}{M_{bqq}} - \frac{m_W}{m_t}}{\sigma_{R_W}}\right)^2$$

Results

Distributions of reconstructed invariant masses for FCNC event sample,
"signal" top decay reconstruction

Higgs boson mass | Top quark mass

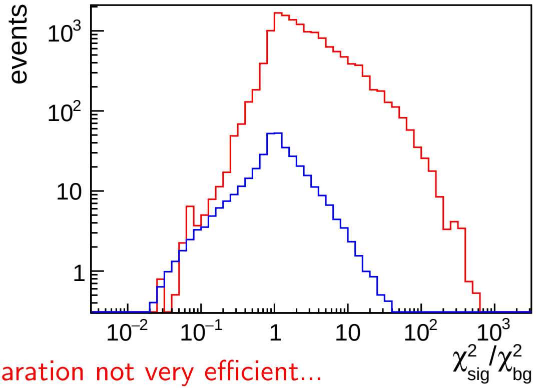


Invariant mass distributions significantly wider than expected !?...
Significant contribution of events with "poor" clustering,
mainly due to higher order QCD effects...

Signal/background discrimination

Kinematic fits for two hypotheses (FCNC signal and SM background) can be compared to discriminate between signal and background events.

χ^2 ratio for two hypotheses



Kinematic separation not very efficient...

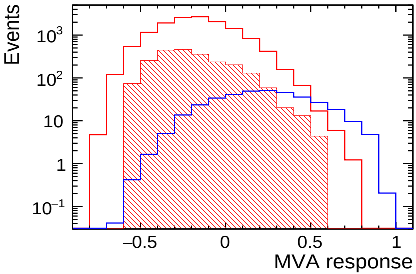
Multivariate analysis TMVA

Used for final signal vs background discrimination

Based on: event variables, flavour tagging and kinematic fit

New approach: one BDT trained on both samples,

hadronic and semi-leptonic decays



⇒ avoid complicated procedure for combining limits from both channels

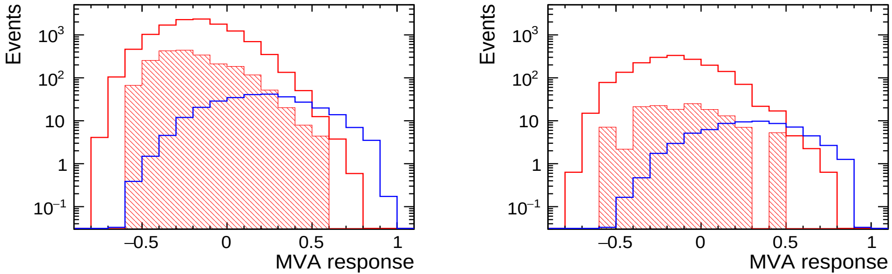
Multivariate analysis TMVA

Used for final signal vs background discrimination

Based on: event variables, flavour tagging and kinematic fit

New approach: one BDT trained on both samples, shown separately for

hadronic decays semi-leptonic decays



$\Rightarrow$ avoid complicated procedure for combining limits from both channels

Selection efficiencies

Cut	FCNC signal	$t\bar{t}$/6 fermion	4 fermion	quark pairs
Preselection	98.6%	88%	8.5%	19.9%
Classification	98.9%	90%	5.1%	1.1%
Signal selection	45%	3.6%	2.8%	3.3%
BDT response	16.6%	0.17%	<0.1%	0.5%
Total	7.3%	$4.8 \cdot 10^{-5}$	$< 10^{-7}$	$3 \cdot 10^{-7}$

Selection efficiencies

Cut	FCNC signal	$t\bar{t}$/6 fermion	4 fermion	quark pairs
Preselection	98.6%	88%	8.5%	19.9%
Classification	98.9%	90%	5.1%	1.1%
Signal selection	45%	3.6%	2.8%	3.3%
BDT response	16.6%	0.17%	<0.1%	0.5%
Total	7.3%	$4.8 \cdot 10^{-5}$	$< 10^{-7}$	$3 \cdot 10^{-7}$

Expected limit 95% CL

With estimated background of 24 events and signal efficiency of 7.3%

$$BR(t \to cH) \times BR(H \to b\bar{b}) \quad < \quad 1.6 \cdot 10^{-4}$$

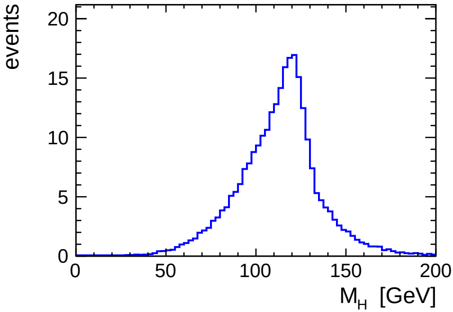
in agreement with results presented at LCWS'2017. Considered final

Going into more details...

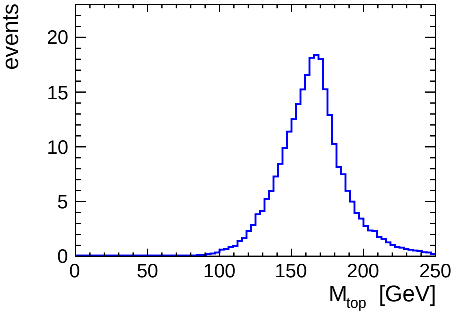
Results

Distributions of reconstructed invariant masses for FCNC event sample, "signal" top decay reconstruction

Higgs boson mass



Top quark mass

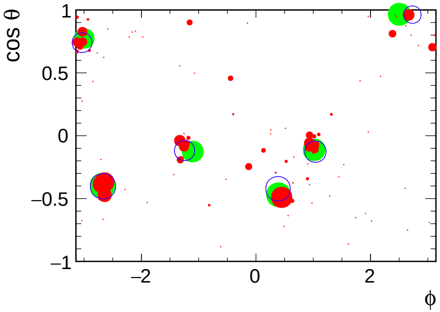


Invariant mass distributions significantly wider than expected !?...

Kinematic fit

Reconstructed PFOs and the clustering results compared to parton level

"good" event



- partons
- reconstructed particles (PFOs)
- Valencia jets (LCFI+)
- anti-k_T jets

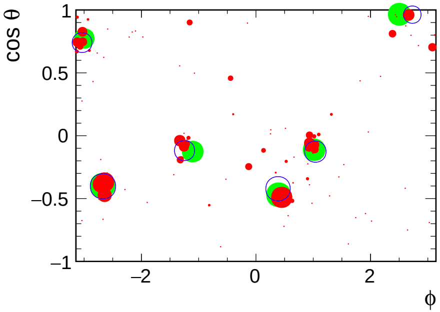
size reflects energy (log scale)

⇒ Kinematic fit works OK!

Kinematic fit

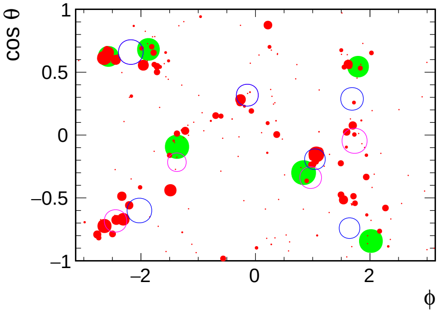
Reconstructed PFOs and the clustering results compared to parton level

"good" event



⇒ Kinematic fit works OK!

"bad" event

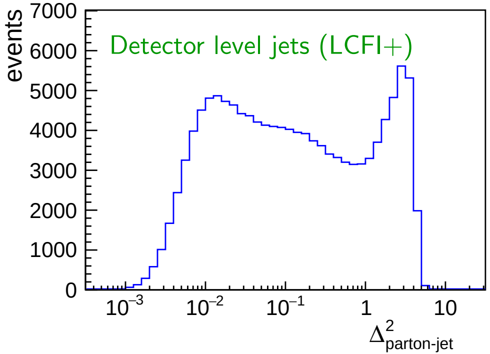


⇒ Can not discriminate between signal and background...

Clustering quality

"Distance" Δ^2 defined to quantify the agreement between generator level partons and particle or detector level jets

$$\Delta^2_{\text{parton-jet}} = \min_{\text{all combinations}} \sum [\sphericalangle(\vec{p}_{jet}, \vec{p}_{parton})]^2$$

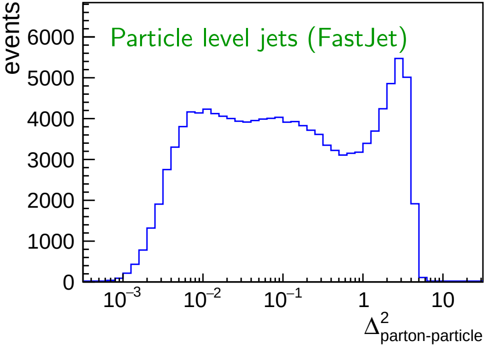


Significant contribution of poorly reconstructed events ($\Delta^2 > 1$)

Clustering quality

"Distance" Δ^2 defined to quantify the agreement between generator level partons and particle or detector level jets

$$\Delta^2_{\text{parton-jet}} = \min_{\text{all combinations}} \sum [\sphericalangle(\vec{p}_{jet}, \vec{p}_{parton})]^2$$

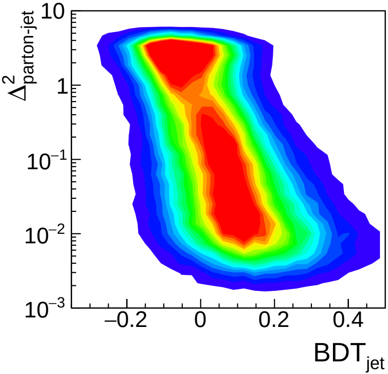


Significant contribution of poorly reconstructed events ($\Delta^2 > 1$)

Clustering quality estimate

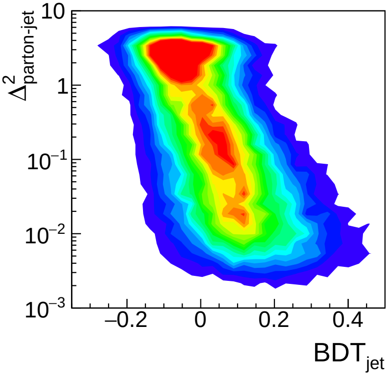
Dedicated BDT implemented to recognize events with "bad" clustering
based on jet variables and comparison of different jet algorithms (!)

SM $t\bar{t}$ background



FCNC signal

hadronic channel

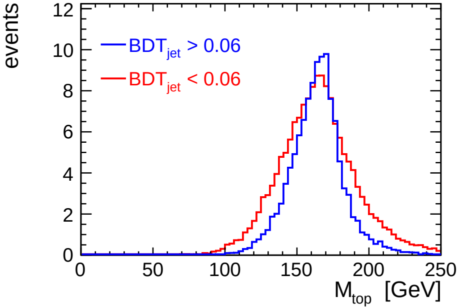


Clustering quality estimate

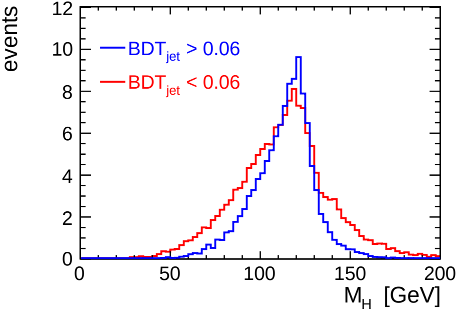
Dedicated BDT implemented to recognize events with "bad" clustering based on jet variables and comparison of different jet algorithms

Kinematic fit result for FCNC sample (signal top decays)

Top quark mass



Higgs boson mass



Kinematic fit

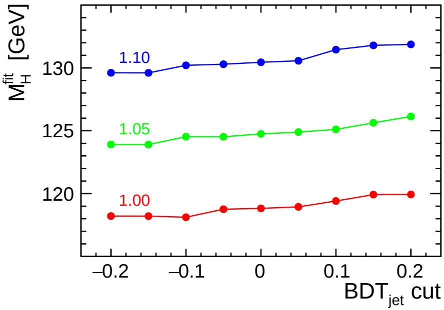
b-jet energy correction

No visible shift in $W^{\pm}$ boson invariant mass (two light quark jets).
Significant shift in reconstructed Higgs boson and top quark masses.

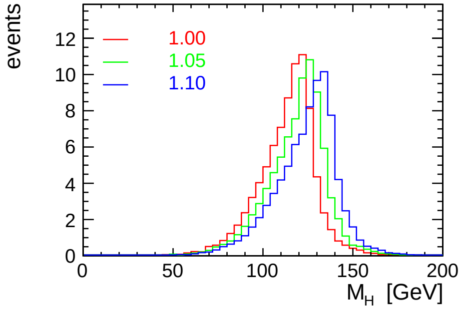
⇒ additional 5% energy correction for *b*-jets

Higgs boson reconstruction

Maximum position vs quality cut



Reconstructed mass distribution

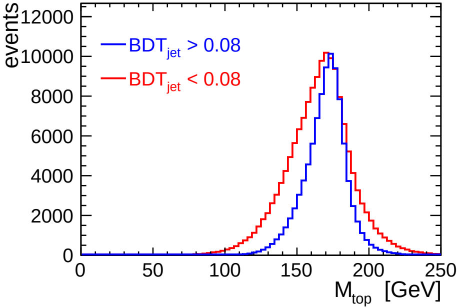


Clustering quality estimate

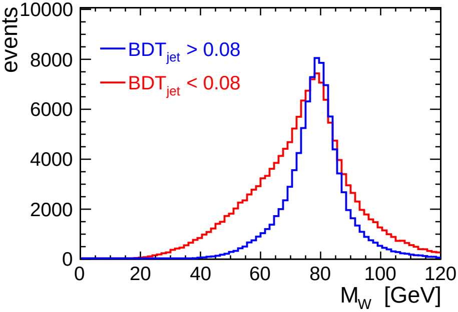
Dedicated BDT implemented to recognize events with “bad” clustering
based on jet variables and comparison of different jet algorithms

Kinematic fit result for SM $t\bar{t}$ background sample   5% b-jet correction

Top quark mass



W boson mass

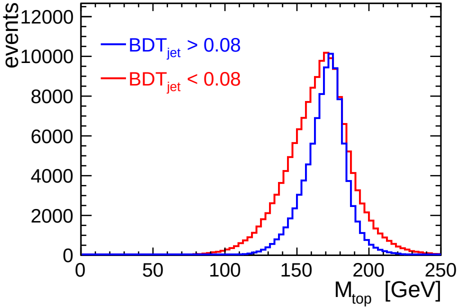


Clustering quality estimate

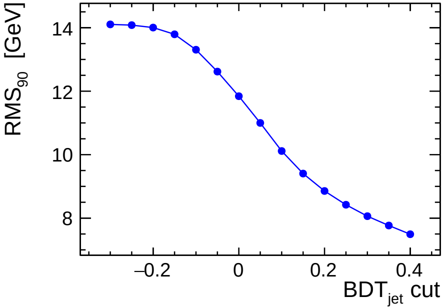
Dedicated BDT implemented to recognize events with "bad" clustering based on jet variables and comparison of different jet algorithms

Kinematic fit result for SM $t\bar{t}$ background sample 5% b-jet correction

Top quark mass



Top mass resolution



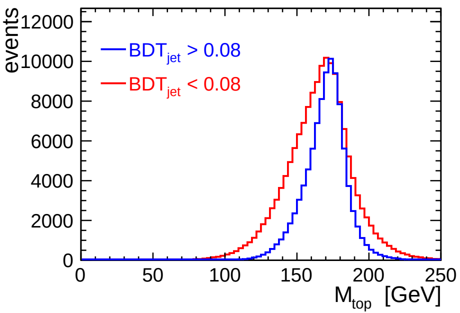
Kinematic fit

Clustering quality estimate

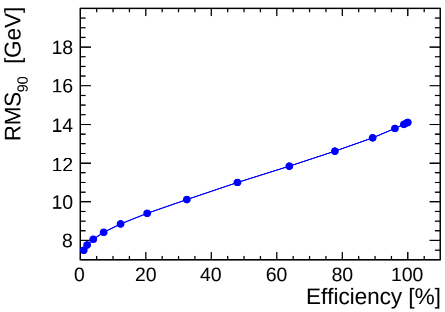
Dedicated BDT implemented to recognize events with "bad" clustering
based on jet variables and comparison of different jet algorithms

Kinematic fit result for SM $t\bar{t}$ background sample   5% b-jet correction

Top quark mass



Top mass resolution

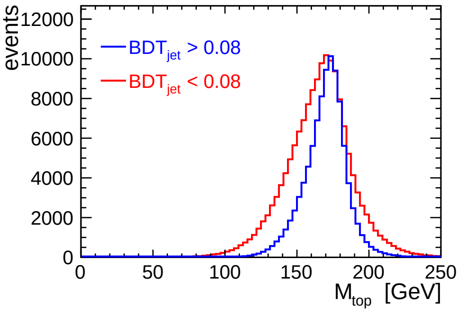


Clustering quality estimate

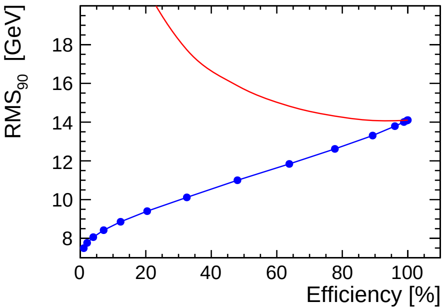
Dedicated BDT implemented to recognize events with "bad" clustering
based on jet variables and comparison of different jet algorithms

Kinematic fit result for SM $t\bar{t}$ background sample 5% b-jet correction

Top quark mass



Top mass resolution

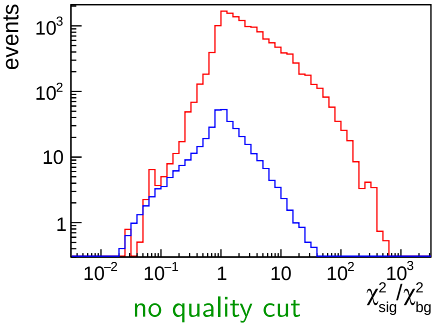


Quality cut does not improve statistical precision...

Signal/background discrimination

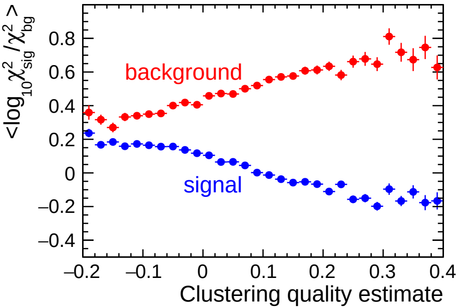
Kinematic fits for two hypotheses (FCNC signal and SM background) can be compared to discriminate between signal and background events.

χ^2 ratio for two hypotheses



no quality cut

average χ^2 ratio vs quality estimate

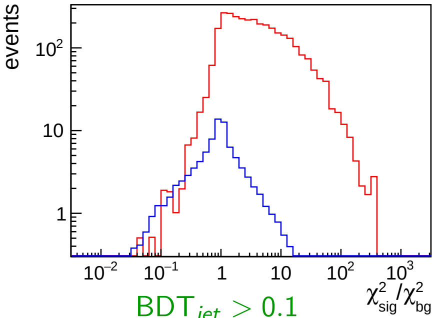


Kinematic fit

Signal/background discrimination

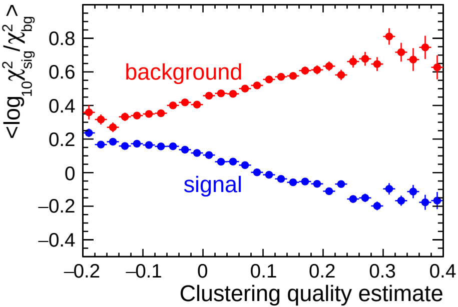
Kinematic fits for two hypotheses (FCNC signal and SM background) can be compared to discriminate between signal and background events.

χ^2 ratio for two hypotheses



$\mathrm{BDT}_{jet} > 0.1$

average χ^2 ratio vs quality estimate



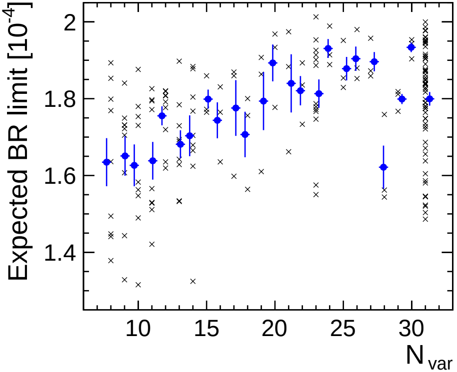
Quality cut does not improve (sufficiently) kinematic separation...

BDT classification TMVA

Used for final signal vs background discrimination

Earlier results: 31 variables used, including quality estimate

New results: reduction in number of variables improves BDT selection (!)



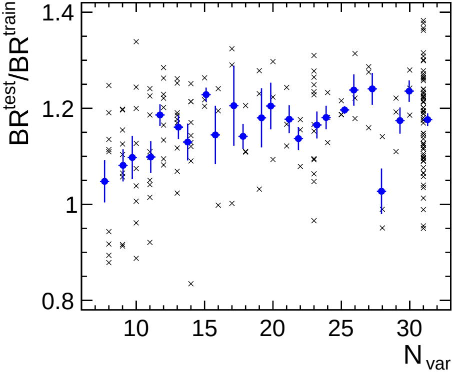
⇒ BDT with 11 input variables used for presented results (no quality!)

BDT classification TMVA

Used for final signal vs background discrimination

Earlier results: 31 variables used, including quality estimate

New results: reduction in number of variables improves BDT selection (!)



$\Rightarrow$ higher number of variables increases the chances of BDT overtraining!

LCFI+ problem

Marlin analysis flow

Aiming for reconstruction of 6j, 4j+l and 2j+2l final states in one job

Initial approach: (ver 1)

- take PFO collection with loose timing cuts: all PFOs
- vertex finding using all PFOs   LcfiplusProcessor
- identification of isolated leptons   IsolatedLeptonFinderProcessor

  ⇒ create new collection with isolated leptons removed: noiso PFOs
- jet clustering and flavour tagging for 6-jet hypothesis using all PFOs
- jet clustering and flavour tagging for 4 jets using noiso PFOs
- jet clustering and flavour tagging for 2 jets using noiso PFOs

  (but primary vertex reconstructed with all PFOs)

⇒ 4 jet and 2 jet clustering turned out to be extremely time consuming !...

Marlin processing times

example

[MESSAGE "Marlin"] Time used by processors (in processEvent()) :
[MESSAGE "Marlin"]
[MESSAGE "Marlin"] TagJets2 4.040871e+04 s in 499 events
==> 8.097938e+01 [s/evt.]

[MESSAGE "Marlin"] TagJets4 7.628240e+03 s in 499 events
==> 1.528705e+01 [s/evt.]

[MESSAGE "Marlin"] allVertexFinder 4.515380e+03 s in 499 events
==> 9.048858e+00 [s/evt.]

[MESSAGE "Marlin"] TagJets6 1.797500e+02 s in 499 events
==> 3.602204e-01 [s/evt.]

Processing time for clustering+tagging

$$6j : 4j : 2j \quad \sim \quad 1 : 40 : 220$$

LCFI+ problem

Marlin analysis flow

Aiming for reconstruction of 6j, 4j+l and 2j+2l final states in one job

Modified approach: (ver 2)

- take PFO collection with loose timing cuts: all PFOs
- vertex finding using all PFOs
- identification of isolated leptons IsolatedLeptonFinderProcessor

  ⇒ create new collection with isolated leptons removed: noiso PFOs
- vertex finding using noiso PFOs
- jet clustering and flavour tagging for 6-jet hypothesis using all PFOs
  (with primary vertex reconstructed with all PFOs)
- jet clustering and flavour tagging for 4 jets using noiso PFOs
- jet clustering and flavour tagging for 2 jets using noiso PFOs
  (with primary vertex reconstructed with noiso PFOs)

⇒ comparable processing times for all configurations !...

LCFI+ problem

Marlin processing times same set of events

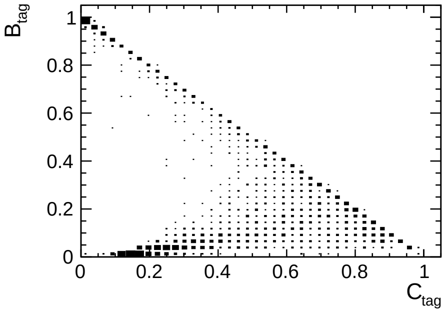
```
[ MESSAGE "Marlin"] Time used by processors ( in processEvent() ) :
[ MESSAGE "Marlin"]
[ MESSAGE "Marlin"] allVertexFinder      2.841660e+03 s in 499 events
                                         ==> 5.694709e+00 [ s/evt.]

[ MESSAGE "Marlin"] noisoVertexFinder    2.826240e+03 s in 499 events
                                         ==> 5.663808e+00 [ s/evt.]

[ MESSAGE "Marlin"] TagJets2             4.801900e+02 s in 499 events
                                         ==> 9.623046e-01 [ s/evt.]

[ MESSAGE "Marlin"] TagJets4             1.224400e+02 s in 499 events
                                         ==> 2.453707e-01 [ s/evt.]

[ MESSAGE "Marlin"] TagJets6             1.142300e+02 s in 499 events
                                         ==> 2.289178e-01 [ s/evt.]
```

Tagging performance

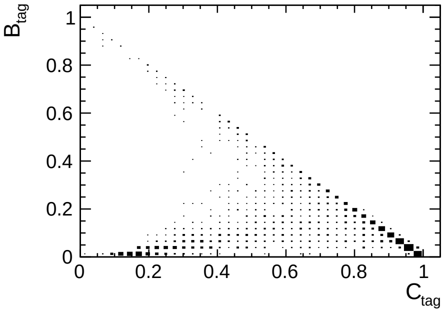
Largest difference observed for c-jet tagging!

Tagging results for 4 jet reconstruction, jet following *c* quark direction (signal sample, *c* quark from FCNC decay)

all PFOs vertex (ver 1)



noiso PFOs vertex (ver 2)

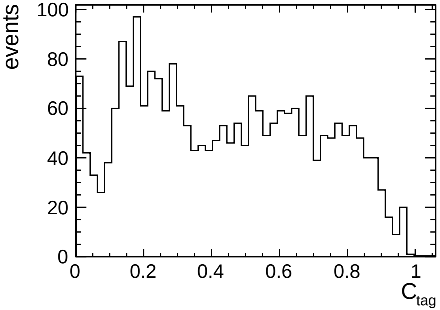


Tagging performance

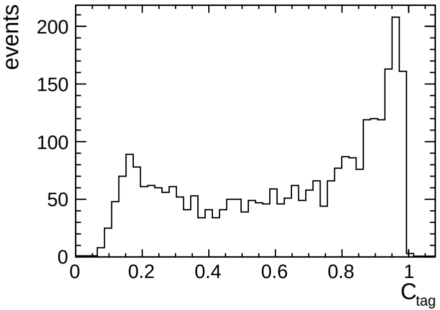
Largest difference observed for c-jet tagging!

Tagging results for 4 jet reconstruction, jet following *c* quark direction (signal sample, *c* quark from FCNC decay)

all PFOs vertex (ver 1)



noiso PFOs vertex (ver 2)



Thank you!

Parton level study presented at TopLC'2015 [arXiv:1604.08122]

Feasibility study with very simple detector modelling. Estimated limit:

$$BR(t \to cH) \times BR(H \to b\bar{b}) < 5 \cdot 10^{-5} \text{ (500 fb}^{-1} \text{ @ 380 GeV)}$$

LCWS'2016 results CLICdp-Conf-2017-005 [arXiv:1703.05007]

Cut based analysis using full simulation samples. Only hadronic final state, only 6-fermion background samples considered. Expected 95% C.L. limit:

$$BR(t \to cH) \times BR(H \to b\bar{b}) \quad < \quad 2.6 \cdot 10^{-4}$$

LCWS'2017 results CLICdp-Conf-2018-001 [arXiv:1801.04585]

Analysis based on BDT algorithms. Both hadronic and semi-leptonic final states considered. Only 6-fermion background samples included:

$$BR(t \to cH) \times BR(H \to b\bar{b}) \quad < \quad 1.6 \cdot 10^{-4}$$

This presentation

Including 6-fermion, 4-fermion and $q\bar{q}$ background samples.

Improved (and simplified) analysis: limit setting with single BDT

Event classification BDT

Variables used to classify $t\bar{t}$ events (and suppress non-$t\bar{t}$ backgrounds):

- total energy of the event, E_{tot},
- total transverse momentum, p_T,
- total longitudinal momentum, p_z,
- missing mass, M_{miss},
- sphericity and acoplanarity of the event, S and A,
- number of isolated leptons, N_{iso}
- energy of isolated lepton with highest p_T, E_{lep},
- minimum jet energy for the 6 jet final state, E^{jet}_{min},
- minimum (y_{min}) and maximum (y_{max}) distance cuts for 6, 4 and 2 jet reconstruction with Valencia algorithm (for 4 and 2 jet clustering, isolated leptons are not included in the clustering).

Signal selection BDT

Variables used for final signal-background discrimination:

- from kinematic fit of signal hypothesis
  - χ^2 value from the fit,
  - reconstructed Higgs boson mass,
  - reconstructed W mass from the spectator top decay,
  - smaller of the two b-tag values for the jets from Higgs boson decay,
  - c-tag and b-tag values for c quark from FCNC decay,
  - b-tag value for b-jet from the spectator top decay,
- from kinematic fit of background hypothesis
  - χ^2 value from the fit,
  - smaller of the two b-tag values for jets from top decays,
- responses of hadronic and semi-leptonic classifiers
  (as used to classify $t\bar{t}$ events at the first analysis stage)