Search for fourth generation t' quarks at CMS

Gerrit Van Onsem

IIHE – Vrije Universiteit Brussel

Student Seminar 10/03/2011

Gerrit.Van.Onsem@cern.ch

Overview and introduction

Overview

- Fourth generation
- The LHC and the CMS detector
- Setup of the analysis
- The hypothesis test
- Exclusion plots
- Outlook and conclusions

• Aim: **search for 4th generation heavy top-like quark (t')**
In the following analysis: assumed production in $t'\bar{t}'$ pairs, and with decay t'→Wb

• Application of a **hypothesis test** (using the "S2-method", already applied previously on the 2010 CMS data as goodness-of-fit test of the Standard Model (SM))
→ SM-only ↔ SM + $t'\bar{t}'$

• **S2-method**: combine information of several observables of an event to look for a global deviation of data or pseudodata w.r.t. SM prediction.

A fourth generation?

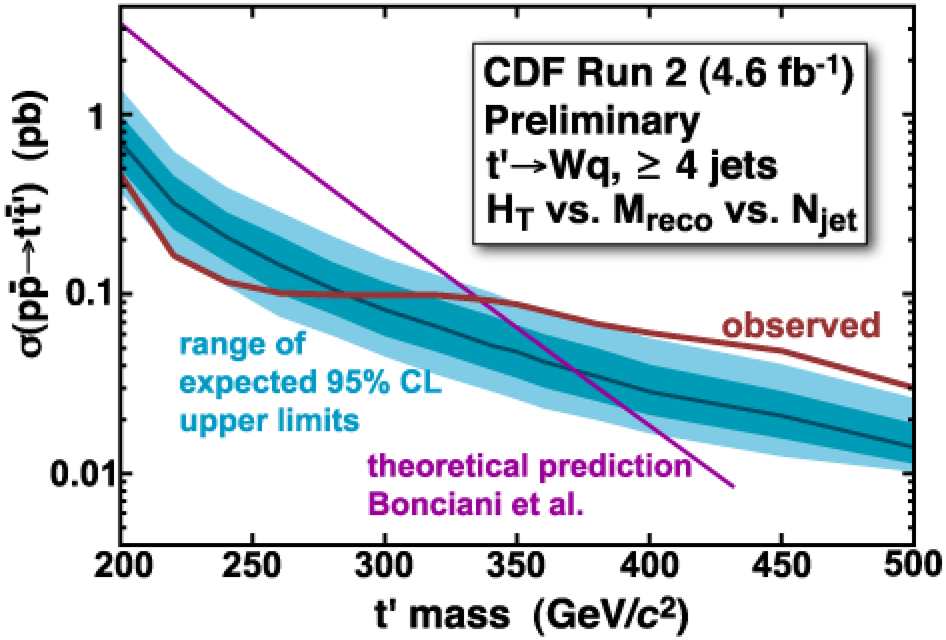
- No theoretical reasons exist that the number of fermion families should be three.

Moreover:

- ➢ Interesting interplay between 4th generation and the EW symmetry breaking
- ➢ CP violation can be enhanced to explain the baryon asymmetry in the universe
- ➢ A fourth family is **not** excluded experimentally

[see e.g. *"Four Statements about the Fourth Generation"*, arXiv:0904.4698v2 (2009)]

[CDF public conference note 10110]



- Searches at the **Tevatron** result in 95% C.L. upper limits on t' and b' masses:

  m(t') = 335 GeV/c^2

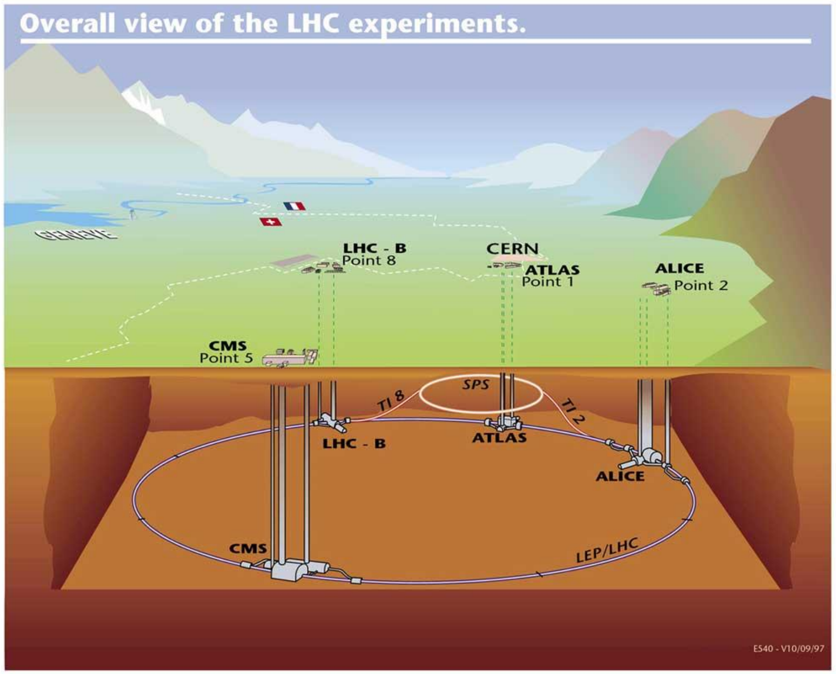
  m(b') = 338 GeV/c^2

- upper limit from **CMS** for b' quark: m(b') = 361 GeV/c^2, no limit for t' yet
  [arXiv:1102.4746v1]

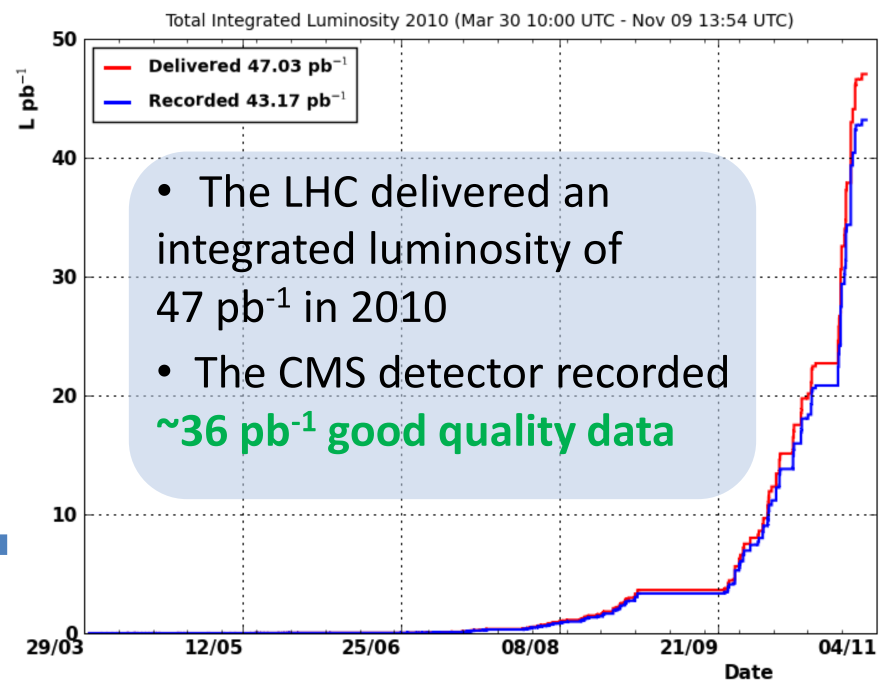
The Large Hadron Collider

Vrije Universiteit Brussel

- The **Large Hadron Collider** at CERN accelerates and collides protons at $\sqrt{s}$ = 7 TeV



- Physics motivation: search for Higgs boson, and physics beyond the SM (supersymmetry, new particles,…)

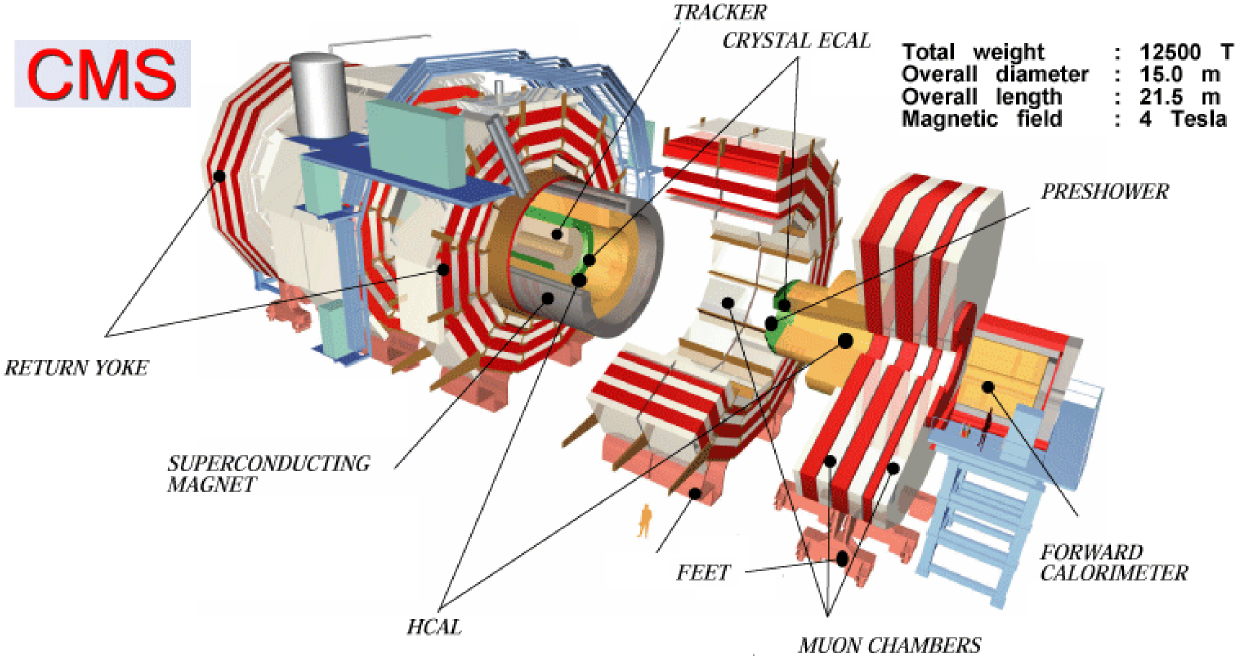


- The LHC delivered an integrated luminosity of 47 pb^{-1} in 2010
- The CMS detector recorded **~36 pb^{-1} good quality data**

Prospects for 2011: several fb^{-1} of data at 7 TeV!

The Compact Muon Solenoid

- CMS: general-purpose detector, to study any physics that comes up at TeV scale.



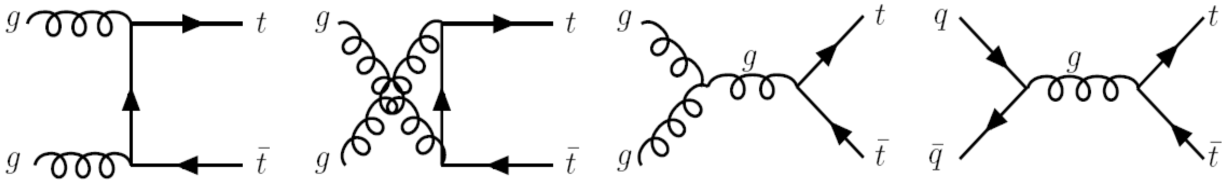
- Trigger and data acquisition systems needed to cope with large data flow

- pseudorapidity

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right)$$

- Consists of subdetectors, build around a large 3.8 T superconducting magnet:
  - **Silicon tracker**: records hits for later use in particle track reconstruction
  - **Electromagnetic calorimeter** (ECAL): direction and energy measurement of photons and electrons
  - **Hadronic calorimeter** (HCAL): measures energy of hadrons
  - **Muon System**: dedicated to muon detection

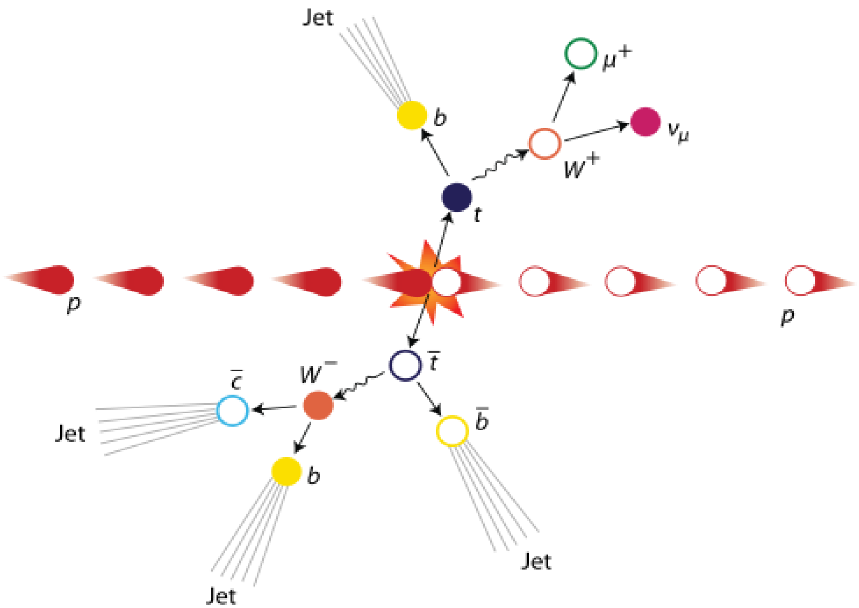
Top quark topologies



$m^{top} \approx 172\,\mathrm{GeV/c}^2$

- Top quark pair ($t\bar{t}$) **production** at LHC

  $\sigma_{NLO}(7\,\mathrm{TeV}) \approx 157\,\mathrm{pb}$



- top **decays** almost exclusively to *W* boson and *b* quark, and *W* decays hadronically ($W \rightarrow q\bar{q}$) or leptonically ($W \rightarrow \ell \bar{\nu}_\ell$)

- **Semi-muonic decay channel**:

  $t\bar{t} \rightarrow bW\bar{b}W \rightarrow bq\bar{q}\bar{b}\mu\nu_\mu$

- **4th generation quarks t'**
  - when pair produced, and **assuming** t'→Wb, the final state of these events looks similar to top quark pair events (= dominant background)

Vrije Universiteit Brussel

Samples

- **SM Monte Carlo** samples: 38X Fall 10, PAT-ified in CMSSW_3_8_5_patch3:

Process	σ_{eff} (NLO) (pb)	Sample
TTJets	157.5	/TTJets_TuneD6T_7TeV-madgraph-tauola/Fall10-START38_V12-v2/AODSIM
WJets	31314	/WJetsToLNu_TuneD6T_7TeV-madgraph-tauola/Fall10-START38_V12-v1/AODSIM
ZJets	3048	/DYJetsToLL_TuneD6T_M-50_7TeV-madgraph-tauola/Fall10-START38_V12-v2/AODSIM
Single top (t)	20.93	/TToBLNu_TuneZ2_t-channel_7TeV-madgraph/Fall10-START38_V12-v2/AODSIM
Single top (tW)	10.6	/TToBLNu_TuneZ2_tW-channel_7TeV-madgraph/Fall10-START38_V12-v2/AODSIM
QCD (mu+jets)	84679.3 (LO)	/QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM

- **Data**: Full 2010 Mu dataset: L_{int} **= 36.1 pb^{-1}**
  - Nov4 ReReco of Run2010A and Run2010B, PAT-ified in CMSSW_3_8_6 patch1

- **t-prime MC** samples (t' pair): 36X user samples PAT-ified in CMSSW_3_8_6_patch1 :

t' pair production	σ (NLO) (pb)	Sample
$m_{t'}$ = 300 GeV	7.3	/tptpjj_m300_7tev_madgraphlhe_iv_fnal/vorobiev-tptpjj_m300_7tev_RECO
$m_{t'}$ = 350 GeV	2.9	/tptpjj_m350_7tev_madgraphlhe_iv/vorobiev-tptpjj_m350_7tev_RECO
$m_{t'}$ = 400 GeV	1.3	/tptpjj_m400_7tev_madgraphlhe_igor/vorobiev-tptpjj_m400_7tev_RECO
$m_{t'}$ = 450 GeV	0.6	/tptpjj_m450_7tev_madgraphlhe_igor/vorobiev-tptpjj_m450_7tev_RECO

[NLO cross sections at 7 TeV: Berger and Cao, arXiv:0909.3555v2]

Vrije Universiteit Brussel

MC corrections and event selection

- **Corrections and scale factors** applied to the MC to account for data-MC differences
  - Jet Energy Resolution
  - Leptonic branching ratio correction
  - Trigger- and lepton efficiencies

Data-driven factors to scale the SM Monte Carlo (to be multiplied with cross section):

- $\beta_{ttbar} = 1.06$, $\beta_{WJets} = 1.34$, $\beta_{ZJets} = 1.32$, $\beta_{singletop} = 1.0$, $\beta_{QCD} = 2.09$

Analysis selection:

- Reference selection **SelV4** of the CMS Top Lepton+Jets group (in sync)

"exactly 1 isolated muon with $p_T > 20$ GeV/c, and at least 4 jets with $p_T > 30$ GeV/c"

- In following: Particle Flow jets (MC: L2L3 corrections START38_V14, data: L2L3 and residual corrections GR_R_38X_V15)

TopLeptonPlusJets group twiki (mu): https://twiki.cern.ch/twiki/bin/view/CMS/TopLeptonPlusJetsRefSel_mu

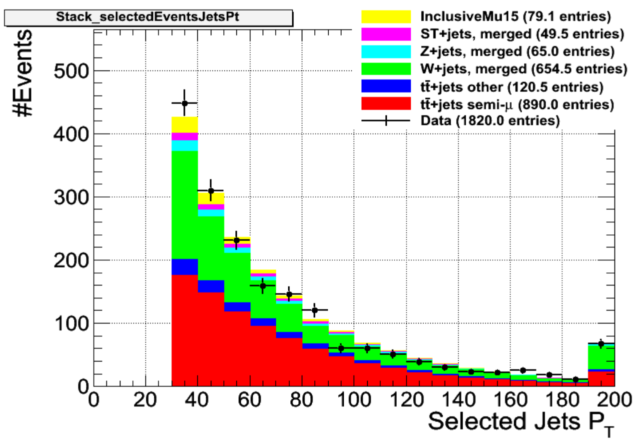
- The t' pair selection efficiencies of the used samples are 8 to 9% (for ttbar this is ~4%)

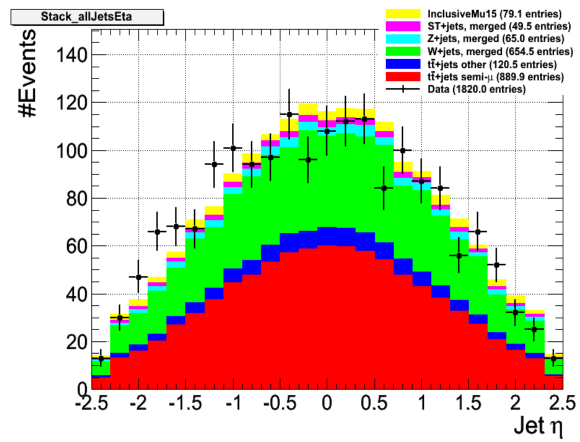
Vrije Universiteit Brussel

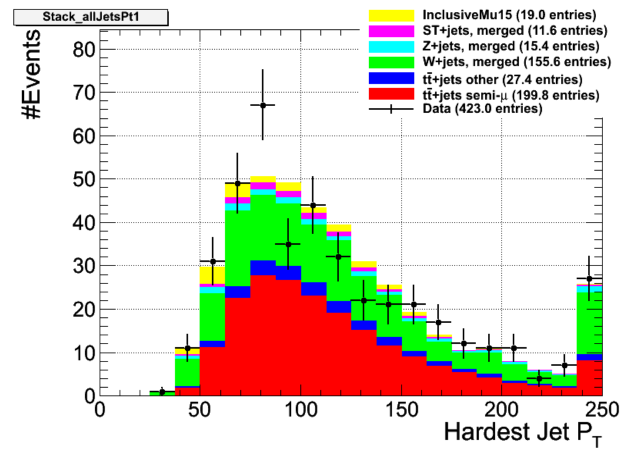
MC vs data

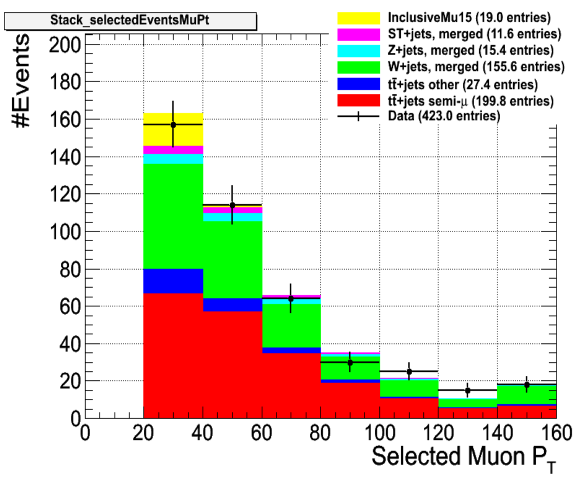
selection	TTJets semi-mu	TTJets other	WJets	ZJets	Single top	QCD	Sum SM MC	Data
RefSelV4	199.8	27.4	155.6	15.4	11.6	19.0	**428.8**	**423**

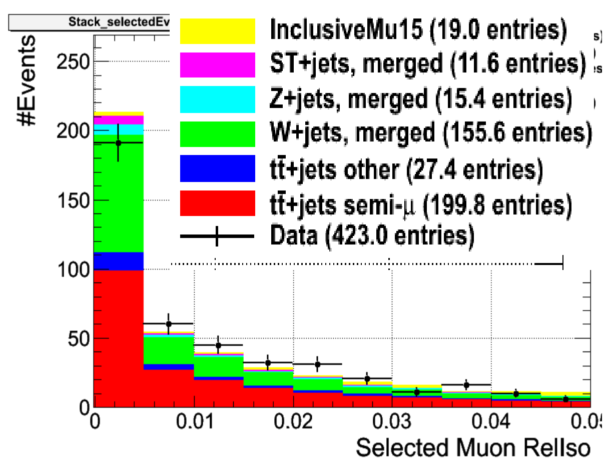
- Some data-MC comparisons of variables after event selection: → **agreement**

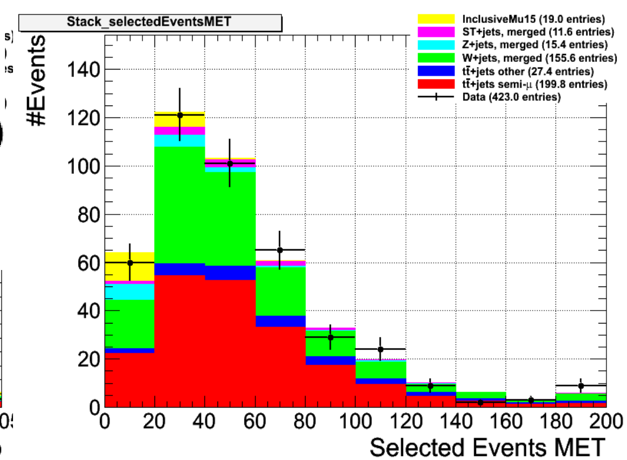






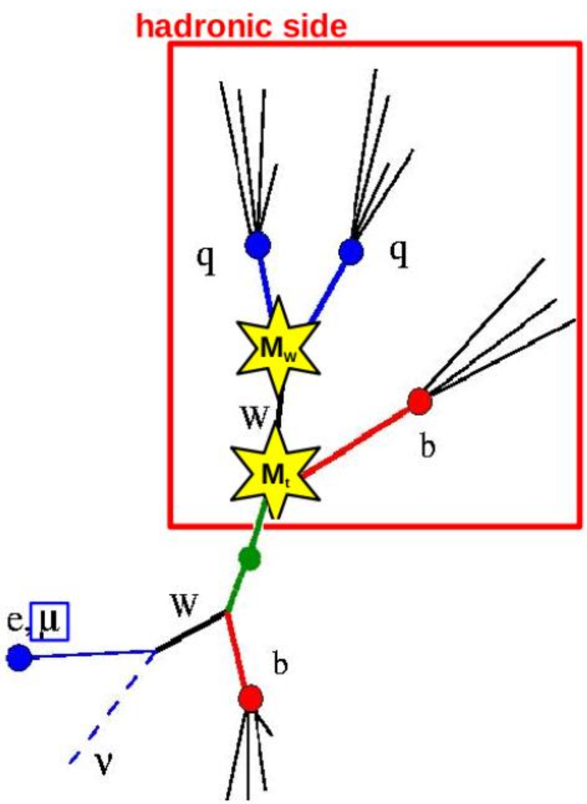






Jet Combinations via TMVA

- multivariate analysis (MVA) technique to **identify which are the light- and b-jets on the hadronic side**
- The 4 highest p_T jets after selection are used
- For each event, there exist 12 possible ways of combining the jets with the quarks
- The good jet combination is obtained by this procedure:
  - The 6 'most' discriminating variables are combined with a Likelihood Ratio
  - The jet combination with the maximum MVA output is chosen

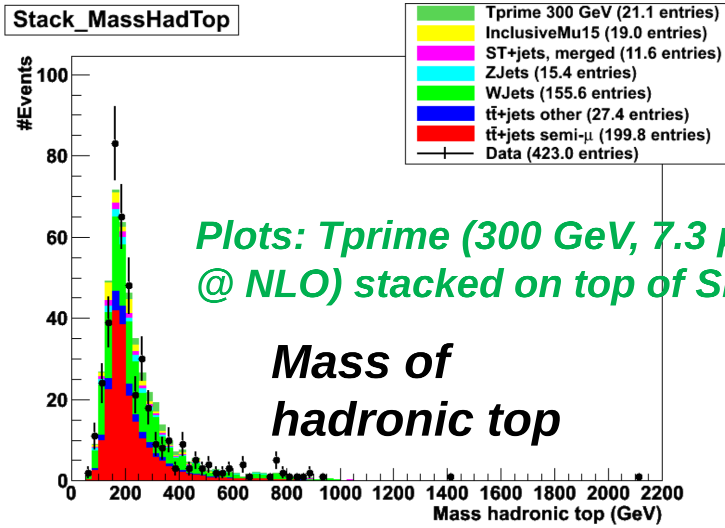


- The 6 variables used are $\Delta\Omega(t^h,\mu),\ \Delta\Omega(t^h,b^l),\ \Delta\Omega(b^l,\mu),\ \frac{p_T^{t^h}}{\sum p_T^{3\text{ jets}}},\ m_{W^h},\ bTag$

In this way we make the good jet combination in 54% of the cases where 4 leading jets can be matched to the 4 ttbar quarks

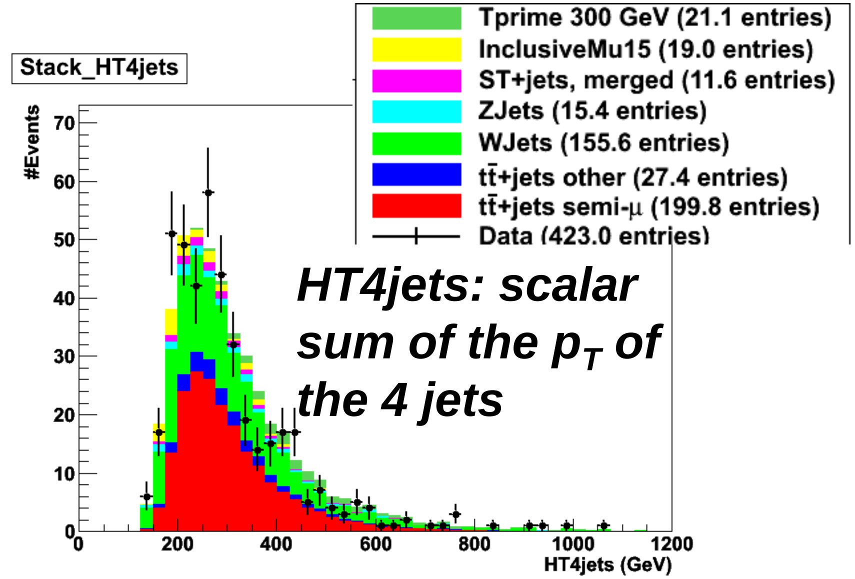
Kinematic variables

- We can construct the variables:

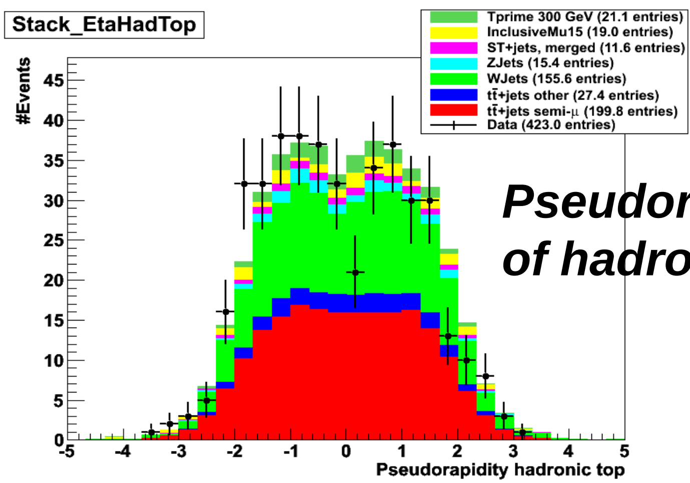


Plots: Tprime (300 GeV, 7.3 pb @ NLO) stacked on top of SM

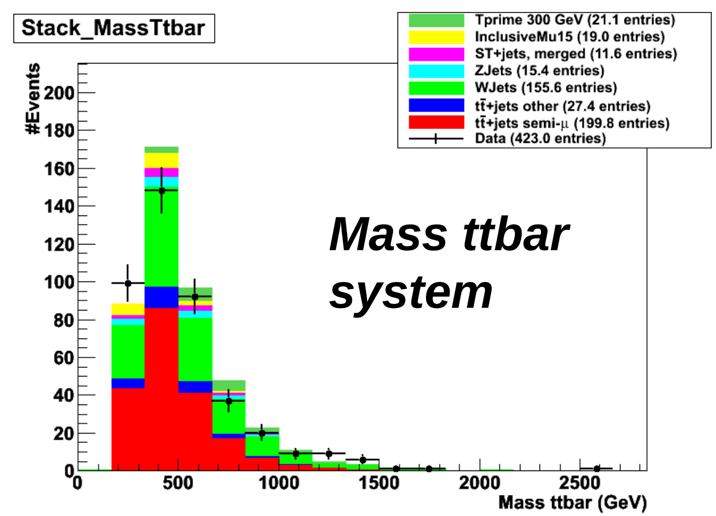
Mass of hadronic top



HT4jets: scalar sum of the p_T of the 4 jets



Pseudorapidity of hadronic top



Mass ttbar system

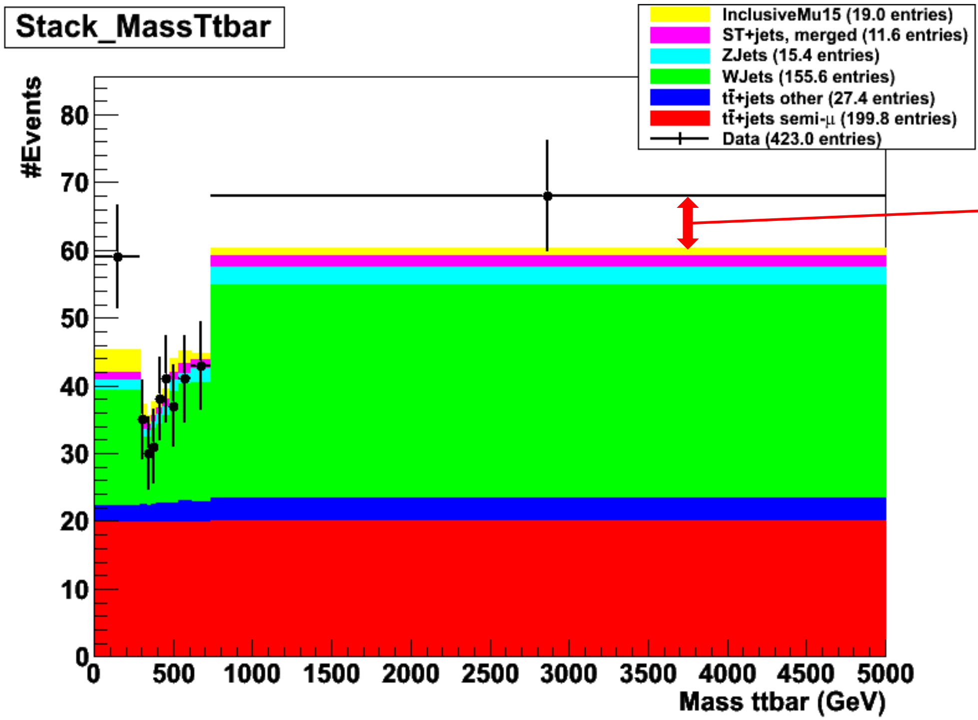
- Binning in following analysis: uniform TTJets semi-mu MC distribution after event selection, 10 bins (we only want enough events per bin). Plots in back-up.

Vrije Universiteit Brussel

The statistical procedure (1)

• Consider *N* **kinematic variables**, each variable *k* has corresponding 1-dimensional binned distributions

- of **data** or **pseudodata** (bin contents d_j, statistical variance $Var[d]_j$)
- of **SM prediction** (bin contents y_j, variance $Var[y]_j$)



• For each bin *j* and observable *k*: a "**squared bin significance**" is computed

$$s_{k,j}^2 = \frac{(d_{k,j} - y_{k,j})^2}{Var[d]_{k,j} + Var[y]_{k,j}}$$

"Measure for deviation between data and prediction in the bin"

• Systematics can be included in squared bin significance definition:

$$Var[y] = Var[y]_{stat} + Var[y]_{syst}$$

or can be treated via nuisance parameters during pseudoexperiments

The statistical procedure (2)

- A certain event has variable (k = {1,…4}) values in particular bins (j = {1,…8}) . Combined (w.r.t. observables) **S-weight** of an event *i:*

$$S_i = \sum_{k=1}^{N} s^2_{k,j_k}$$

→ 1 value for each event. In the presence of New Physics, events with a high S-weight are 'New Physics-like'.

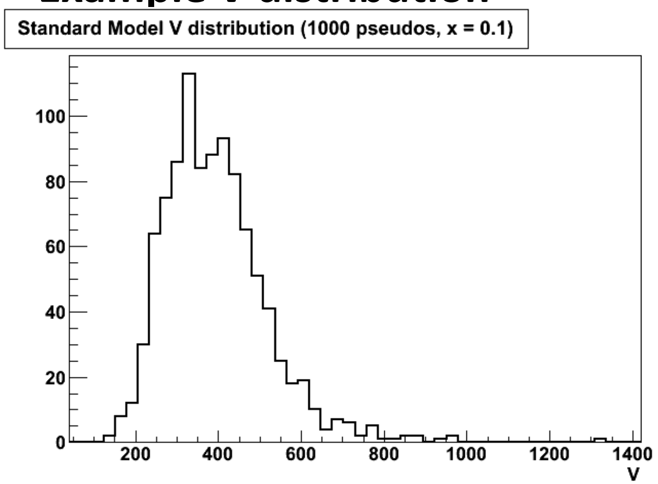
- Take *x* % of the events to obtain 'highest S-weight subsample'. Combine the S-weights:

$$V = \sum_{x\%\,\text{highest}} S_i$$

→ 1 value for whole dataset; *V* will be the **test statistic** in the hypothesis test / goodness-of-fit. *V*-distributions can be made via **pseudoexperiments** with resampling techniques.

↓ pick random events from MC samples (number of events equals random value from Poisson distribution with mean $n = L_{\text{int}}\sigma$)

Example V distribution



- *Note:* choosing a small *x* means you look at the events with the largest deviation w.r.t. the prediction

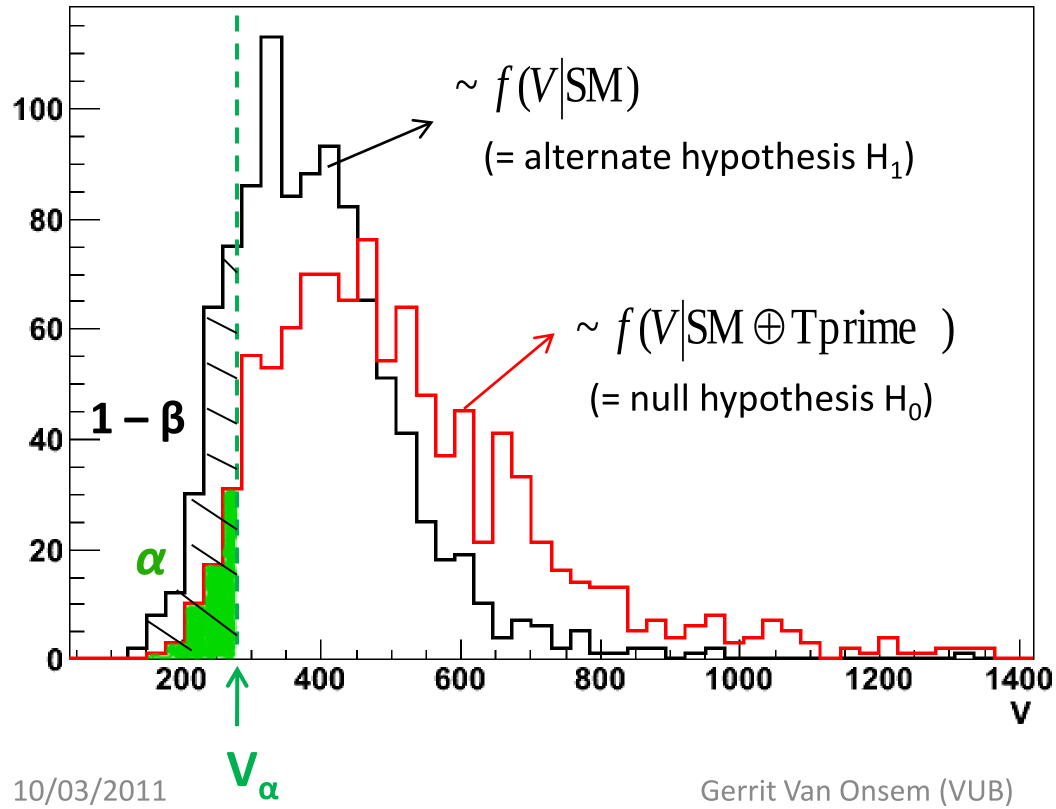
Hypothesis test (1)

- Application on the 4 chosen kinematic variables: obtain two V-distributions
  - 1000 SM pseudoexperiments → normalized to 1 → $f(V|\mathrm{SM}) = H_1$
  - 1000 SM+Tprime pseudoexperiments → $f(V|\mathrm{SM} \oplus \mathrm{Tprime}) = H_0$

Choose significance level $\alpha = 0.05$ and calculate V_α according to:

$$\alpha = \int_0^{V_\alpha} f(V|\mathrm{SM} \oplus \mathrm{Tprime})\,dV$$

V distributions (1000 pseudos, x = 0.1)



The **power of the test** $1-\beta$ is given by:

$$1-\beta = \int_0^{V_\alpha} f(V|\mathrm{SM})\,dV$$

(Hashed region in plot)

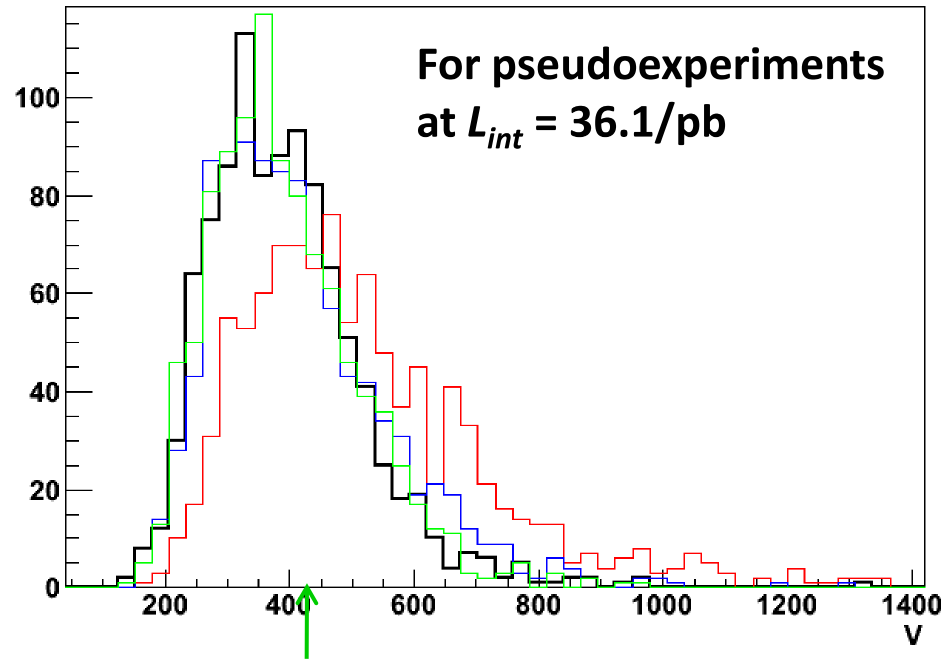
- If $1-\beta > 0.5$, we can say that the null hypothesis H_0 can be rejected on average at 95% C.L.

Hypothesis test (2)

- Consider different null hypotheses H_0 corresponding to different t' masses:
  - $m_{t'}$ = 300 GeV
  - $m_{t'}$ = 350 GeV
  - $m_{t'}$ = 400 GeV
  - $m_{t'}$ = 450 GeV (not plotted)

Always same alternate hypothesis H_1 = **SM-only** (black in plot)

V distributions (1000 pseudos, x = 0.1)



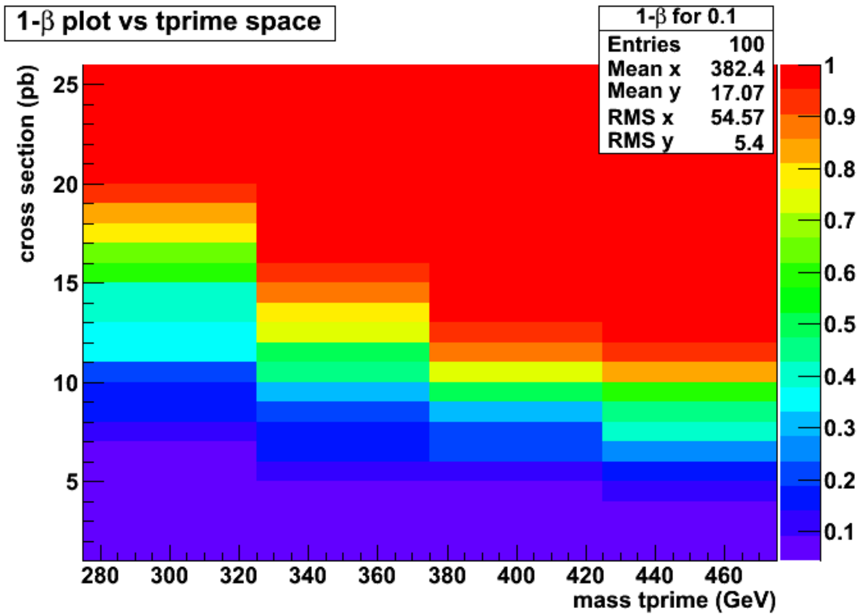
V_{data} = 435 (V value of data can be calculated as well)

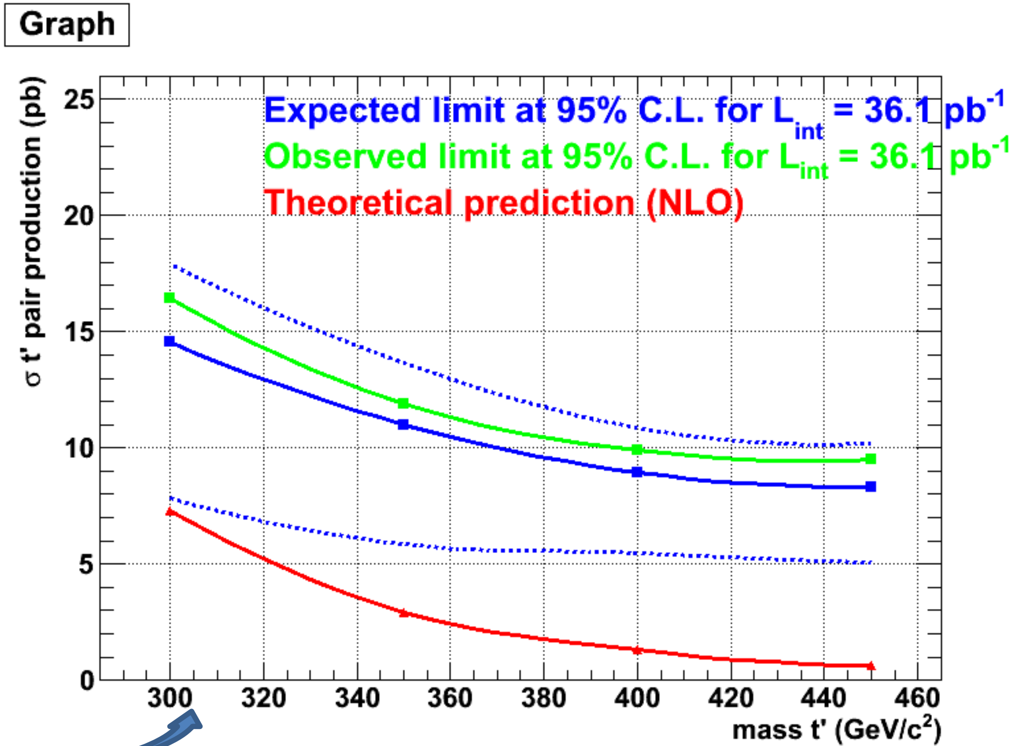
The power of the test $1 - \beta$ is < 0.5 in all these cases

The expectation is that these models cannot be excluded at 95% C.L. with 36.1/pb
With theoretical t' pair production cross sections (NLO)

Exclusion plot for L_{int} = 36.1/pb

- Scan plane of production cross section ($\sigma_{t't'bar}$) vs mass tprime ($m_{t'}$) by performing the hypothesis test and calculating 1 – β for 'each point' in the plane:





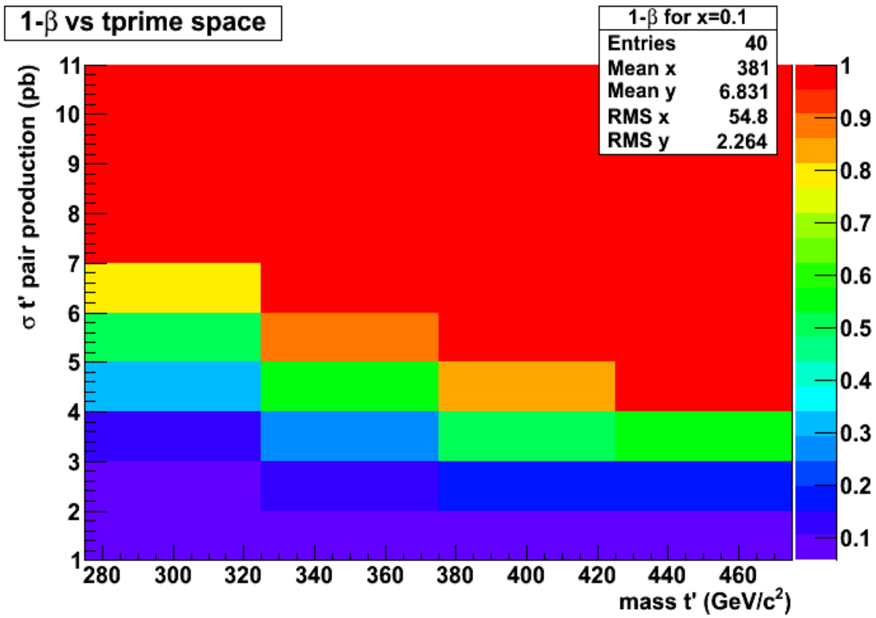
- Interpolation result in expected **exclusion band**
  - ➢ expected limit (solid line): 1 – β = 0.5
  - ➢ dashed lines: 1 – β = 0.16 and 1 – β = 0.84 (1-sigma band: in 68% of the cases a SM outcome lies in the dashed band)

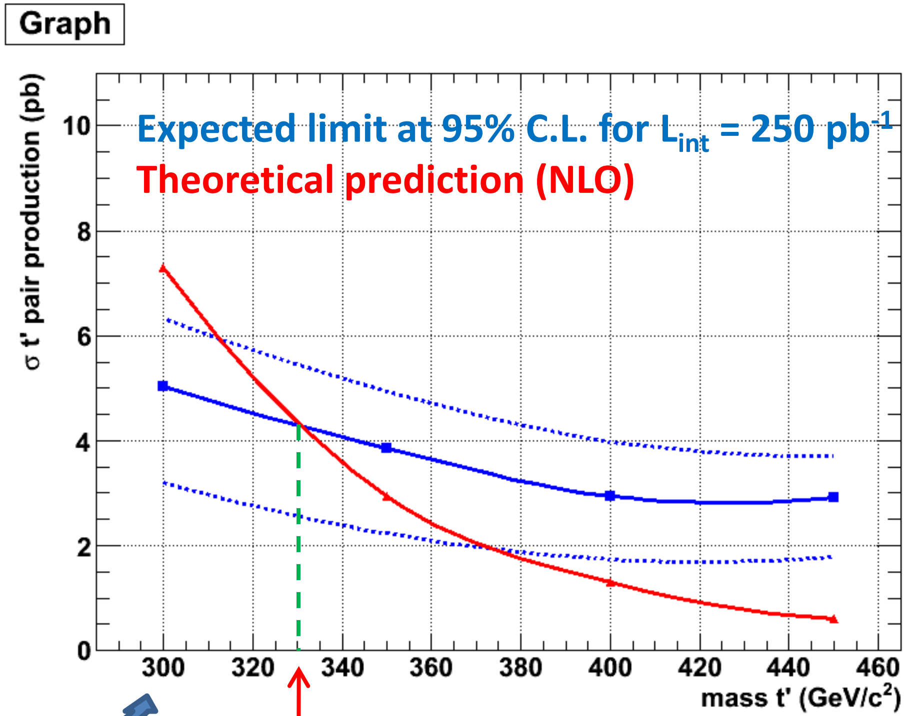
- **Observed exclusion limit at 95% C.L.**

Observed exclusion line lies above theoretical prediction: no t' masses can be excluded

Exclusion plot for L_{int} = 250/pb

- Do the same for **250/pb** (omitting QCD in pseudoexperiments, sample too small)





Interpolation

With 250/pb: expected limit on mass t' = 330 GeV/c²

Note: only first test that tools work, to be optimized later!

Outlook and conclusions

Vrije Universiteit Brussel

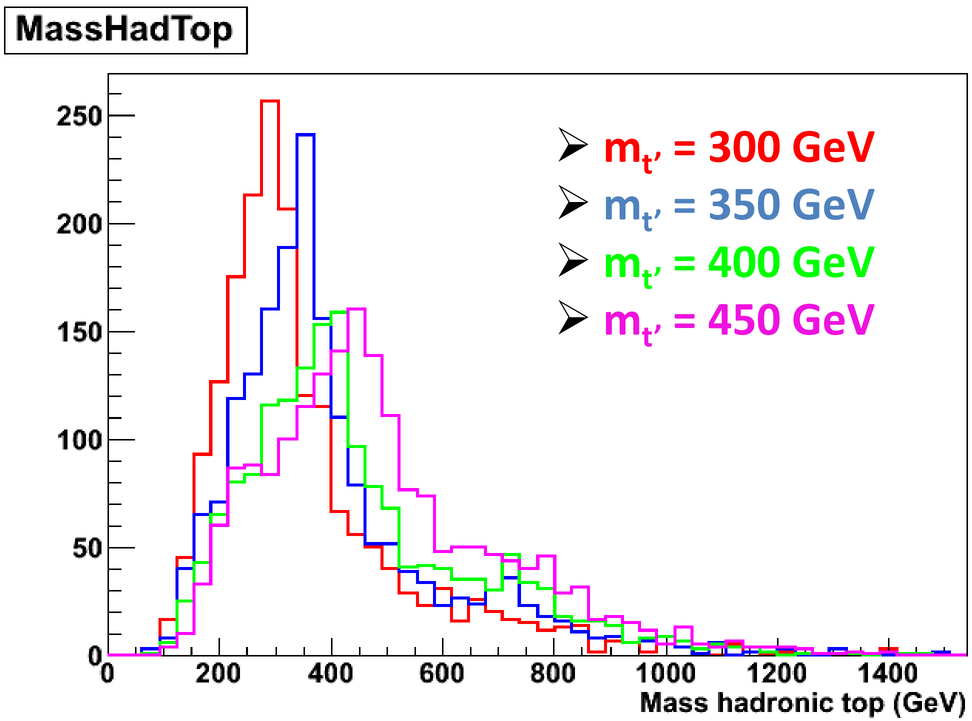
- A **new hypothesis test**, using kinematic variables in ttbar topologies, is applied to distinguish the Standard Model with t' models, to explore the possibilities of the method.
- **Note**: other (t'-)models can be used in the procedure as well

- An exclusion plot is obtained for 36.1/pb, and applied to the 2010 data at 7 TeV, and for 250/pb. **The results shown are a first demonstration: now the method can fine-tuned.**
- **Method can be optimized**:
  - extra cuts (e.g. harder cut on jet p_T)
  - b-tagging (e.g. Wjets background can be reduced a lot)
  - the effect of alternative or more variables can be explored (e.g. η(hadtop) not very sensitive)
  - taking pile-up into account
- **Systematical uncertainties** will be included (MC cross sections, Jet Energy Scale, …)

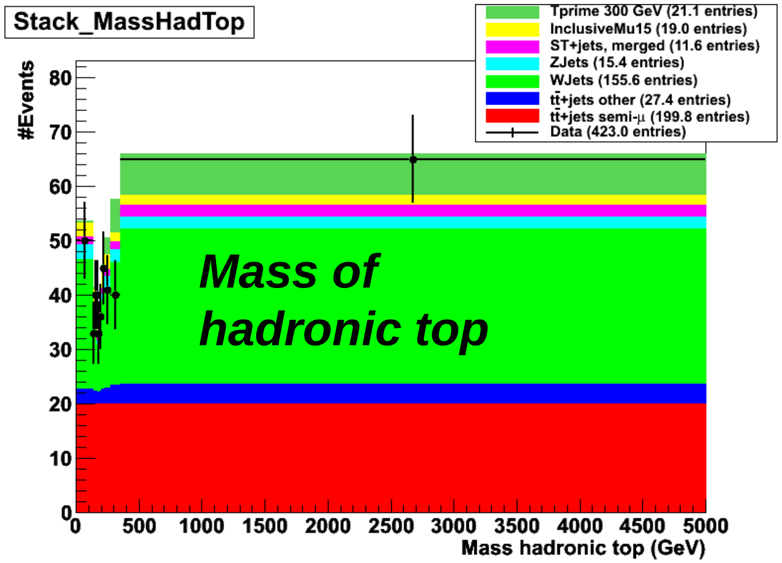
Back-up

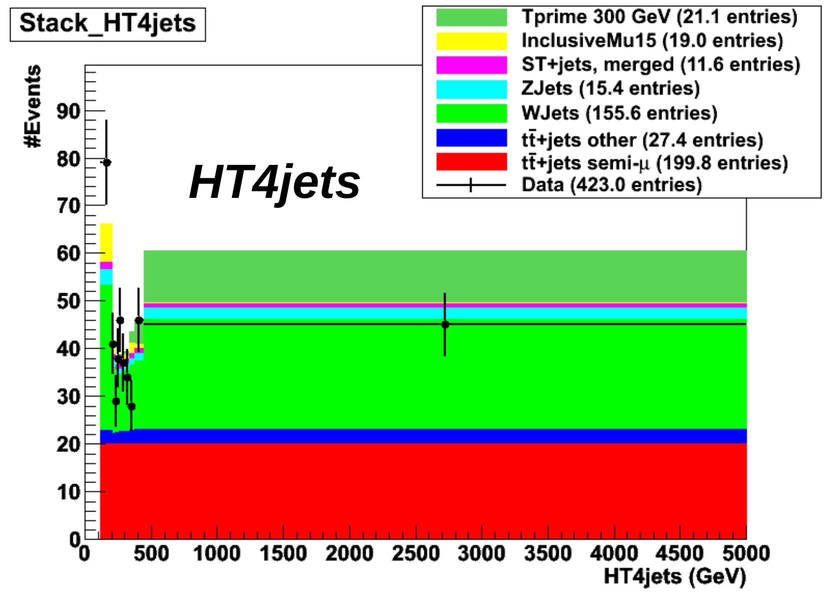
Mass of hadronic top

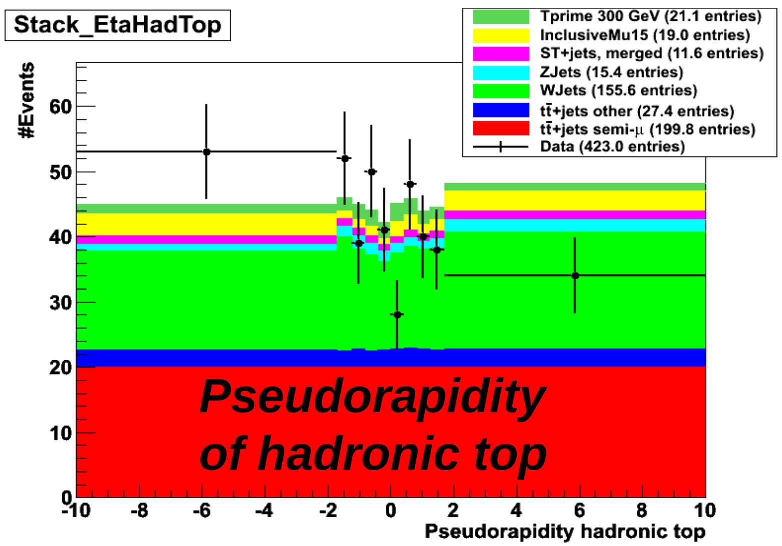
- Mass of the hadronic top for the different used t' samples
(not normalized w.r.t. each other)

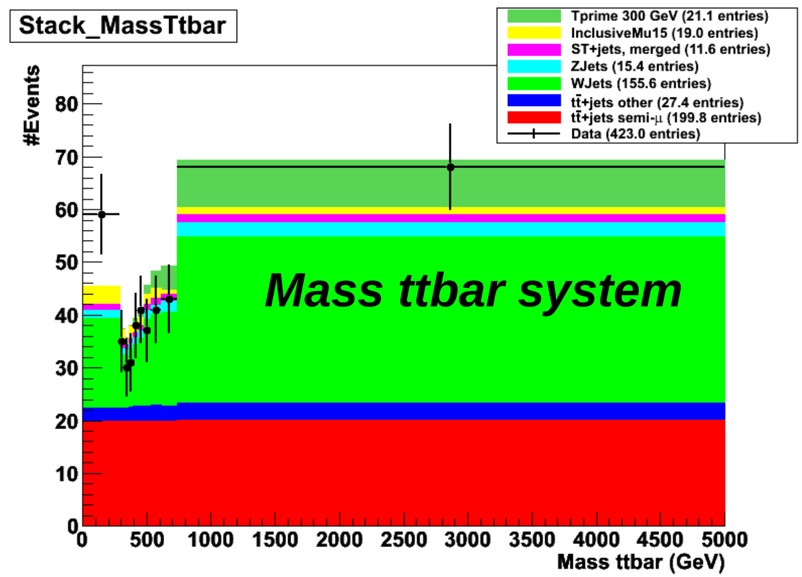


Kinematic variables with binning for the S2-method



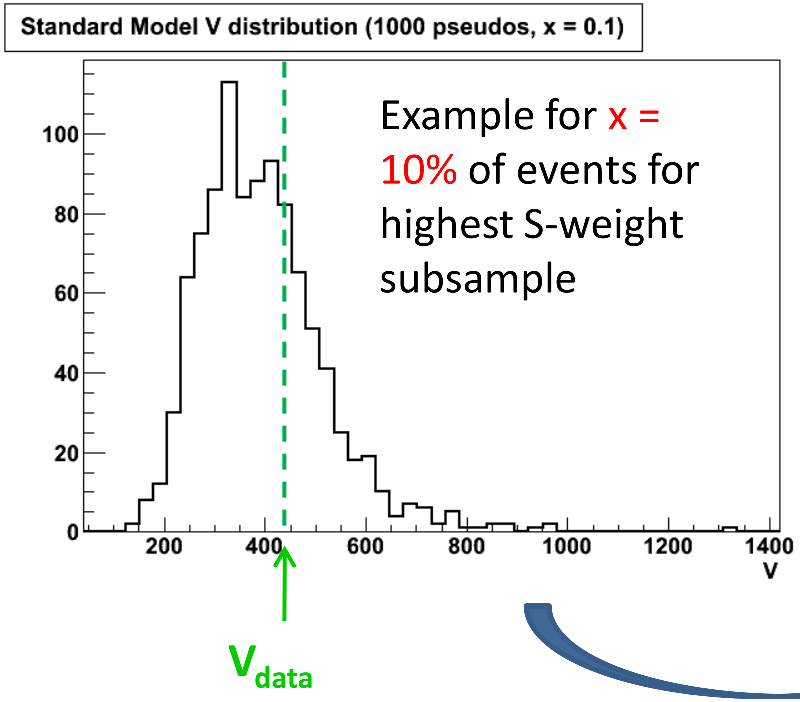


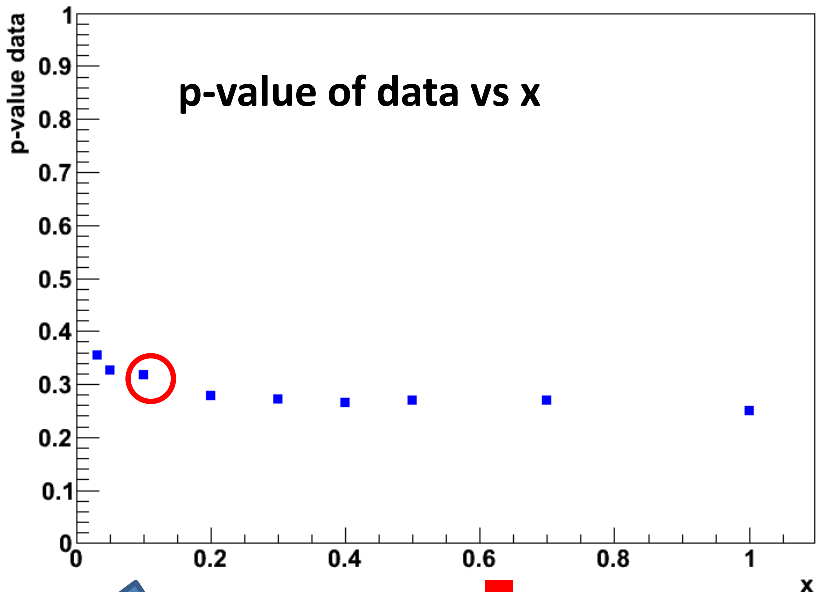




Goodness-of-fit of Standard Model to data

- Application of S2-method as goodness-of-fit test of the Standard Model with the 2010 data, using the 4 variables used in the tprime analysis





$$p = \left(\int_{V_{data}}^{+\infty} f(V)dV \right) / \int_{0}^{+\infty} f(V)dV$$

flat behaviour, p ~ 0.3 without systematics: **data agrees reasonably with SM**

Jet Combinations via TMVA (2)