Top quark physics in CMS

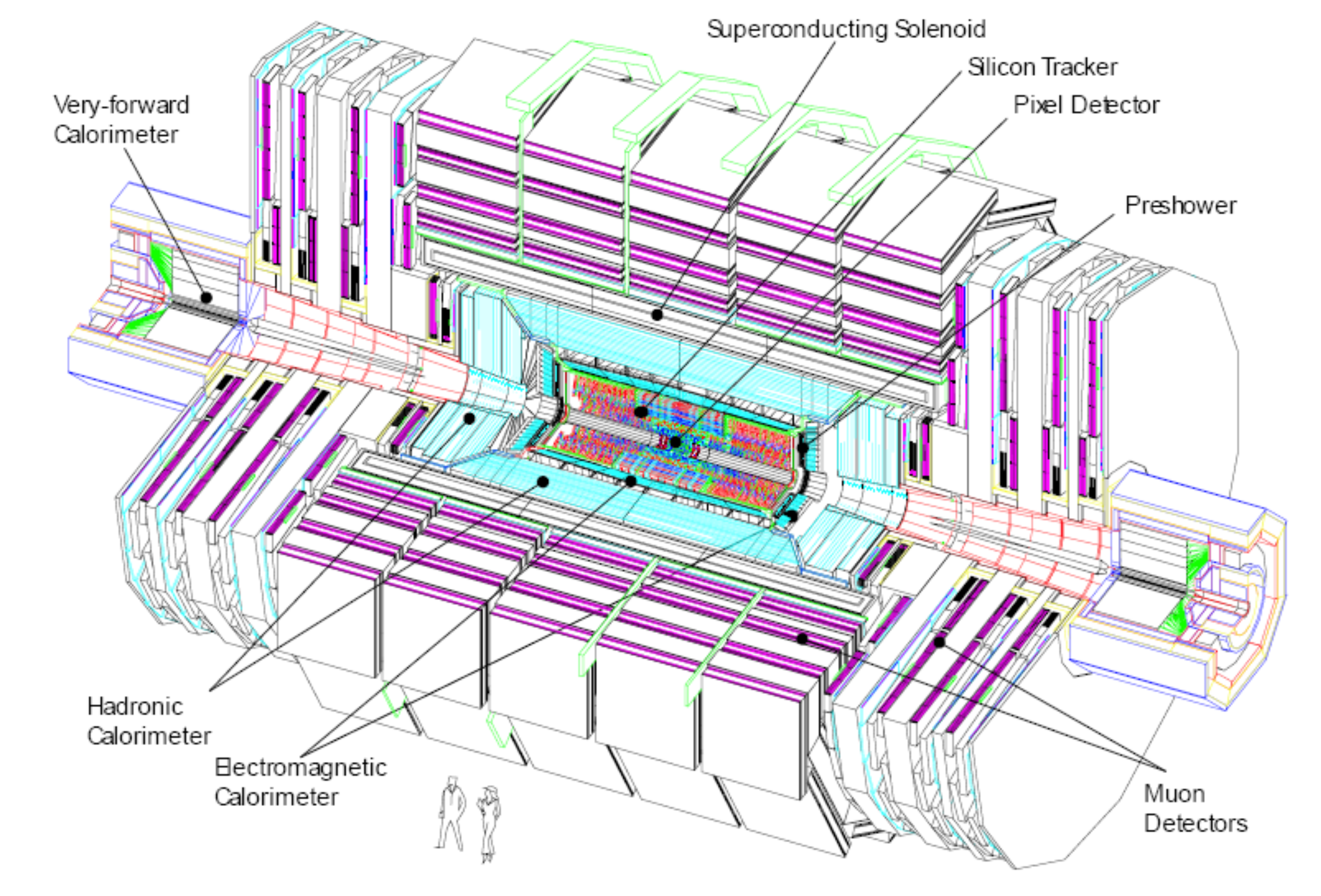
- How to prepare for top quark physics in CMS
  - *key aspects of the CMS detector relevant for top quark physics*
  - *strategy deployed by the CMS collaboration towards data*
- Obtaining a top quark sample
  - *event selection & background estimations*
- First physics analyses exploring the top quark domain
  - *using top quarks for calibration, differential distributions, …*
- Preparing for physics beyond top quark physics

Jorgen D’Hondt

Vrije Universiteit Brussel – IIHE

Top quark physics Workshop – March 13, 2009 – Lisbon

The CMS detector in a nutshell



CMS versus ATLAS

Rick Cavanaugh

	ATLAS	CMS
Ecal+Hcal pion resolution	$\frac{\sigma}{E} = \left(\frac{41.9\%}{\sqrt{E}} + 1.8\%\right) \oplus \frac{1.8}{E}$	$\frac{\sigma}{E} = \frac{90\%}{\sqrt{E}} \oplus 7\%$ *e/h calibrated*
MET resolution (TDR)	$\sigma(\not{E}_T)/\Sigma E_T \approx 53\%/\sqrt{\Sigma E_T}$ *e/h calibrated*	$\sigma(\not{E}_T)/\Sigma E_T \approx 123\%/\sqrt{\Sigma E_T} + 2\%$ *e/h uncalibrated*
Inner tracker resolution (TDR)	$\sigma(p_T)/p_T = 1.8\% + 60\%\ p_T$ (p_T in TeV)	$\sigma(p_T)/p_T = 0.5\% + 15\%\ p_T$ (p_T in TeV)
B field inner region	2 Tesla : p_T swept < 350 MeV	4 Tesla : p_T swept < 700 MeV

ATLAS has 2x better calorimetry, CMS has 4x better tracking!

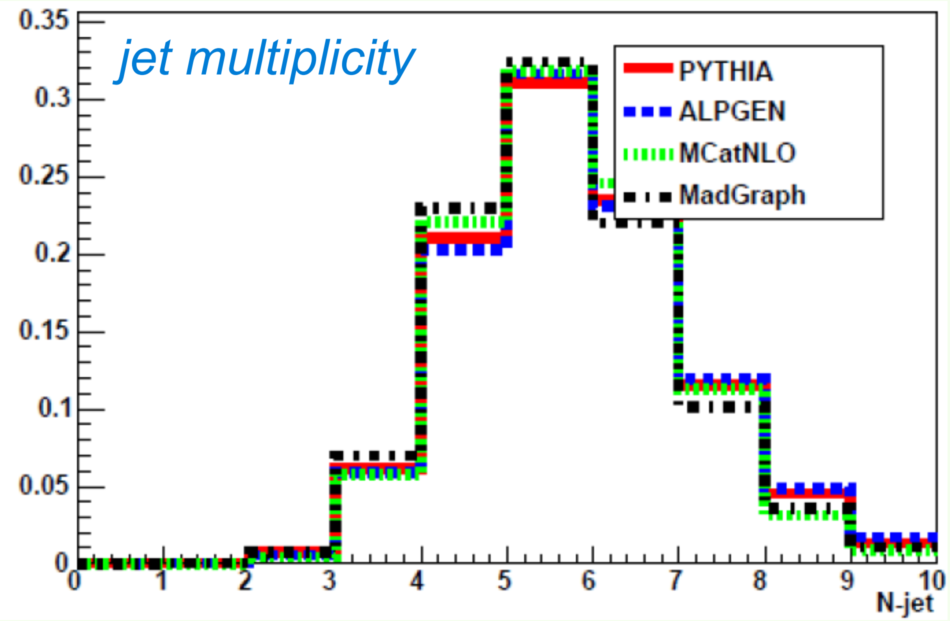
Motivation to implement Particle Flow tools combining the calorimeter with the tracking system. Today all main analyses are using only the calorimeter information to reconstruct jets.

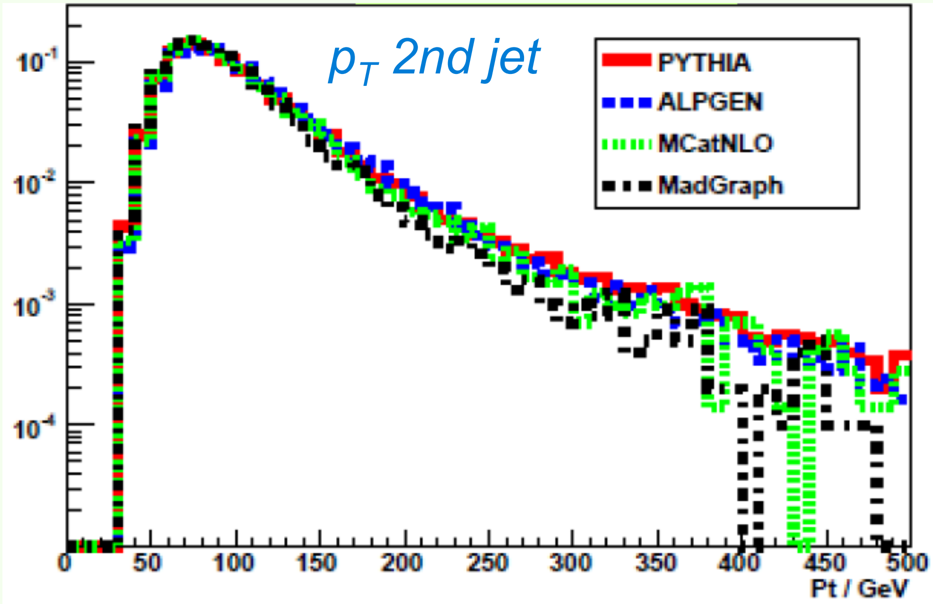
Simulation is being used today

Main aspects of the simulation being used in CMS today:

- **Matrix Element generators for ttbar/W/Z: MadGraph & AlpGen.**
- **Single-top: didn't look enough at single-top as a background for ttbar.**
- **Multi-jet production: mainly biased PYTHIA samples.**

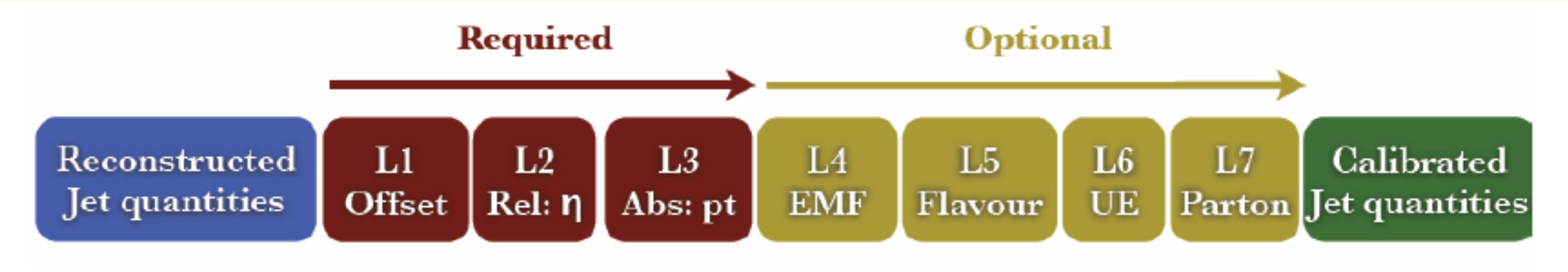
The simulation samples follow closely the changes in the software.
Used to be 14 TeV, now we move to 10 TeV simulations.





Jet calibration strategy

Factorized approach into natural pieces with additional optional corrections:



Allows a thorough understanding of each individual part of a systematic uncertainty on the jet energy scale (factorized uncertainties).

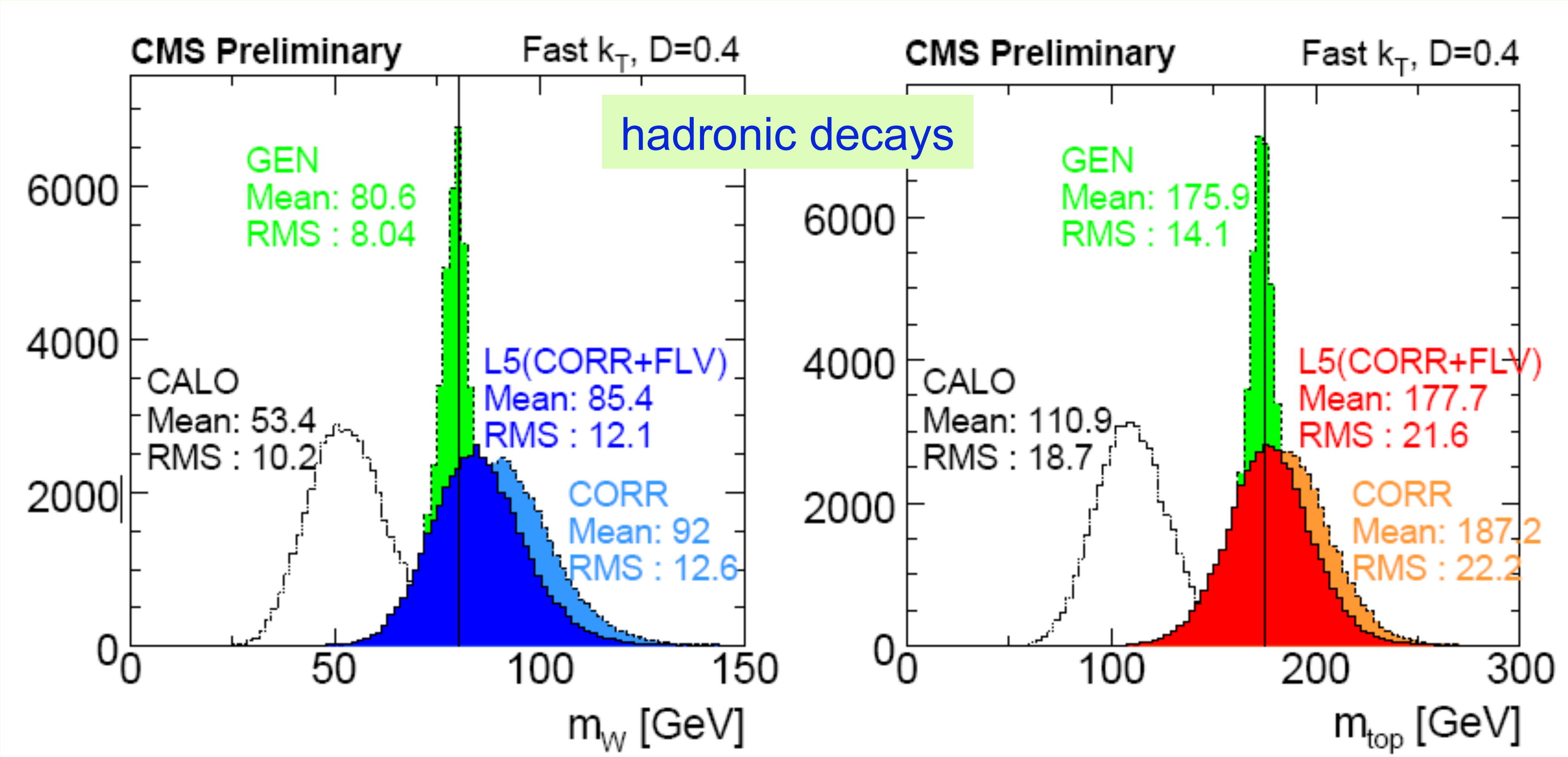
Most of the factors can be measured directly from collision data:

- **L1**: pile-up & threshold effects found in min-bias and zero-bias events.
- **L2**: jet response vs. η relative to barrel found using di-jet balance, etc.
- **L3**: jet response vs. p_T found in barrel using γ/Z + jets, top, etc.

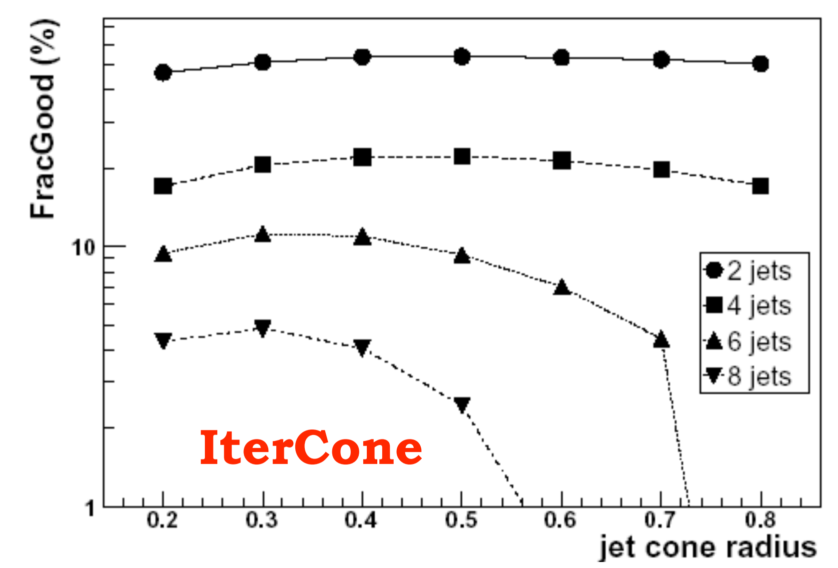
Lots of work in progress and being put in place for first data later this year.

Performance of jet algorithms

The jet reconstruction performance in ttbar events is studied by selecting events with one ("lepton+jets") or zero ("alljets") electron or muon in the final state from a ttbar ALPGEN sample with no additional jets ("ttbar +0 jets"). Only events are considered for which all three decay products of one or both t(tbar) decay(s) can be uniquely matched to reconstructed calorimeter jets.

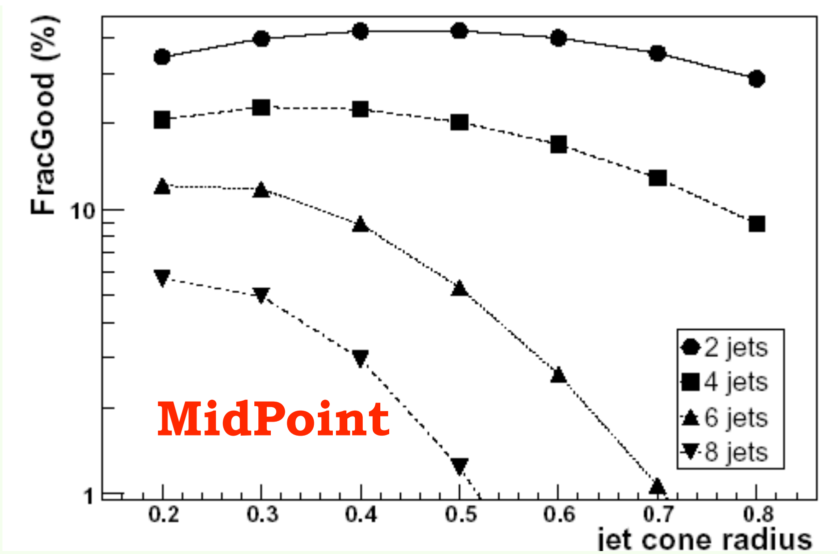


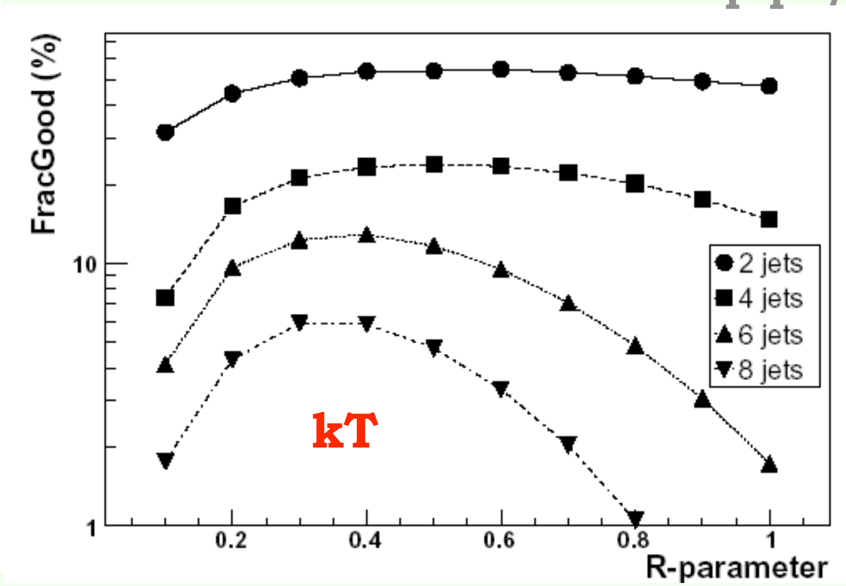
Optimization of parameters



Optimize the matching between the parton and jet kinematics for several benchmark processes (here top quark processes: single-top, top pairs and ttH). Need flexibility of the framework to allow optimization (eg. calibration for several parameters settings).

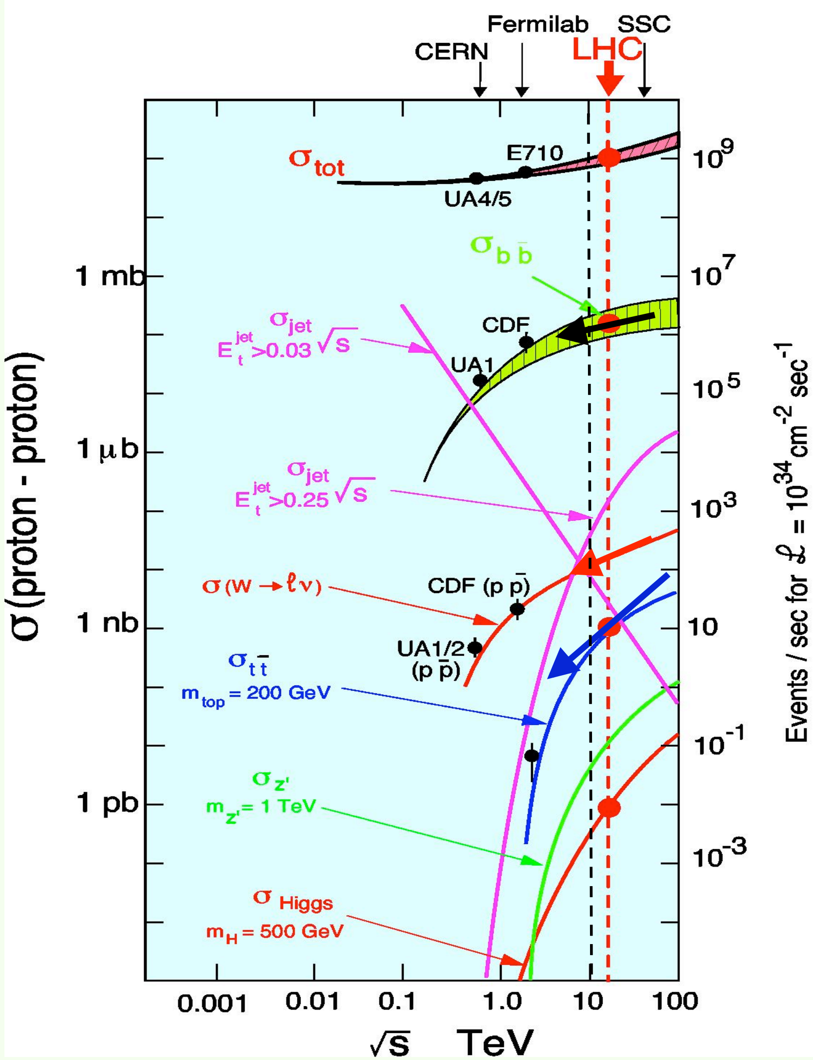
Les houches hep-ph/0604120





Top Quark physics at 10/14 TeV

- **Today all our analyses are performed with simulation of 14 TeV**

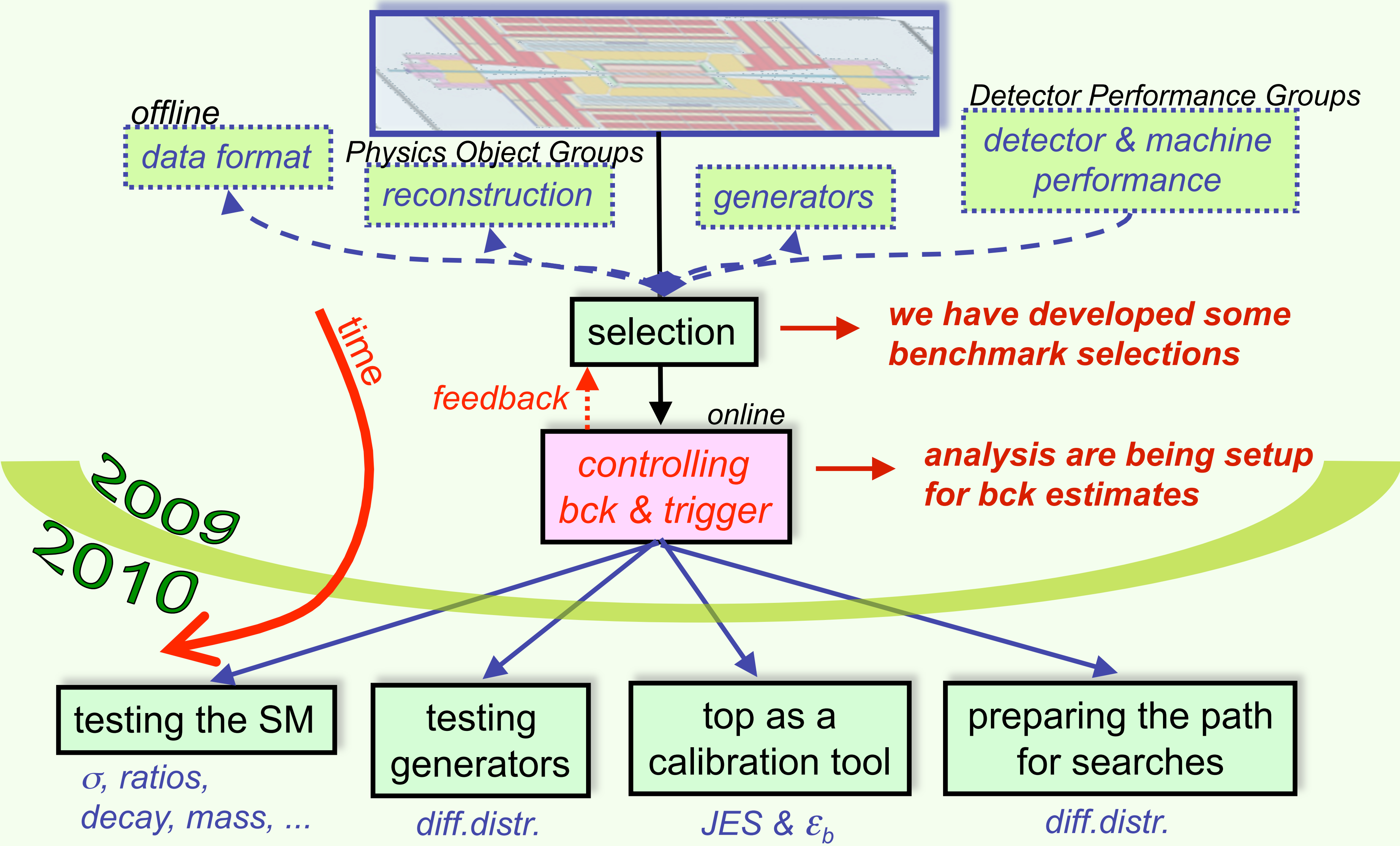


- **Cross section of the top signal is dropping faster from 10 to 14 TeV compared to the background processes**

	10 TeV σ_{LO}MadGraph	14 TeV σ_{LO}MadGraph
Top pairs	317 pb	750 pb
W+jets	40 nb	61 nb

- **Kinematics of the events is about similar (hence assumed equal)**
- **Efficiencies do not scale, S/B does!**
- **S/B scale = 1 → 0.66 & N_{signal} = 1 → 0.42**
- ***For this talk:* *take a 14 TeV analysis with 10/pb to be equivalent to a 10 TeV analysis with 25/pb...***

Getting ready to learn something

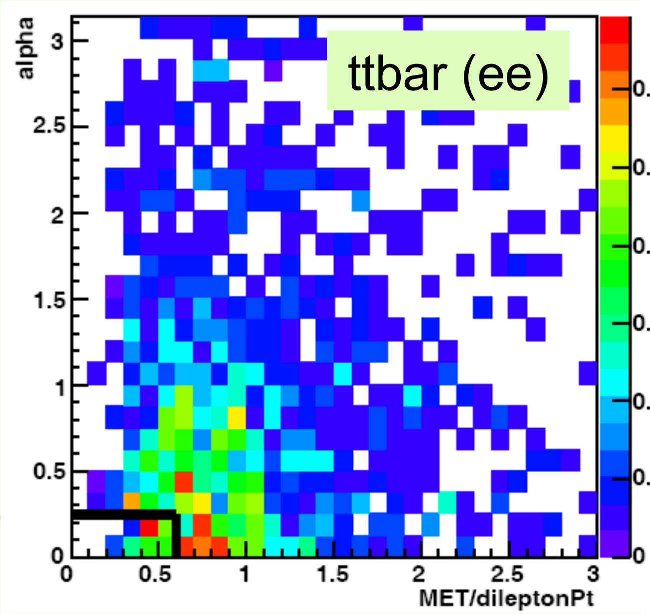


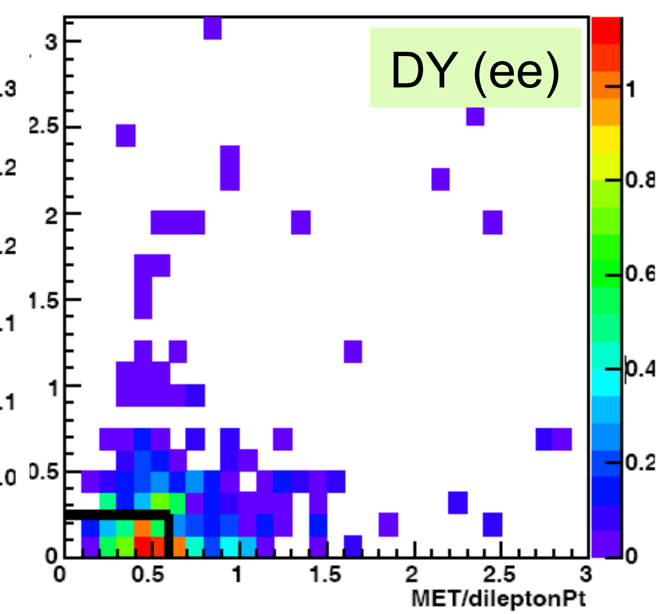
Selection: di-leptons

- **Small branching ratio but can obtain a pure event sample**
- **Trigger based on single lepton triggers as cuts in the analyses are higher than the HLT single-lepton thresholds**

ee mode	$\mu\mu$ mode	$e\mu$ mode
HLT1ElectronRelaxed 17	HLT1MuonNonIso 16	HLT1ElectronRelaxed 17 OR HLT1MuonNonIso 16

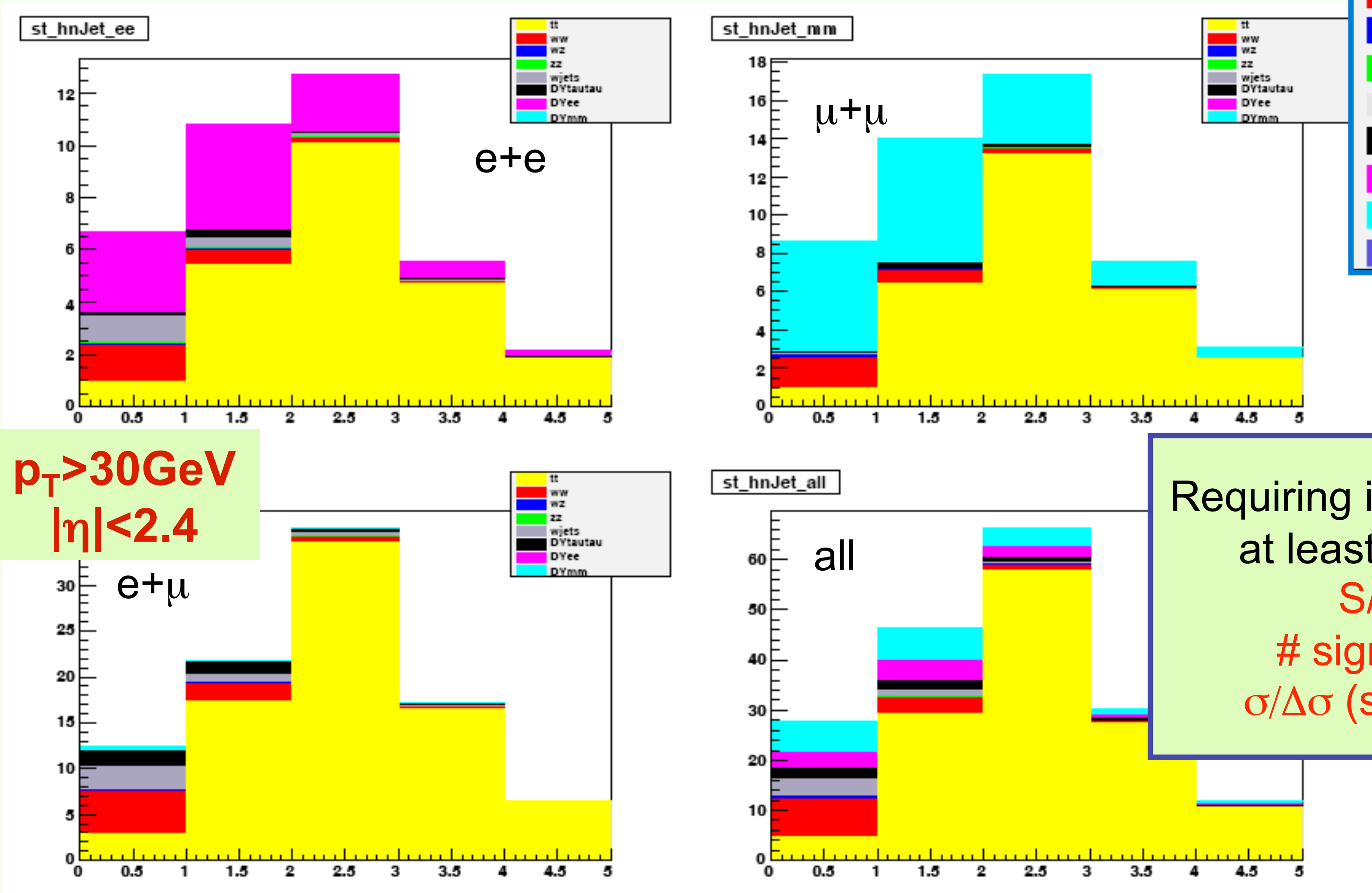
- **Muons (+isolation) :**
  - $|d_0^{X,Y}|<2.5$mm,
  - #hits≥7,
  - χ^2/ndf<5
- **Electrons (+isolation) :**
  - e/γ "tight" eID,
  - $|d_0^{X,Y}|<400\mu$m,
  - no μ in ΔR=0.1
- **ME_T>30 GeV & $\phi(ME_T,\ell\ell)$ or $ME_T>0.6\ p_T(\ell\ell)$**





Results: di-leptons

- **After Z veto [76,106] GeV, the resulting jet multiplicities**



p_T>30GeV
|η|<2.4

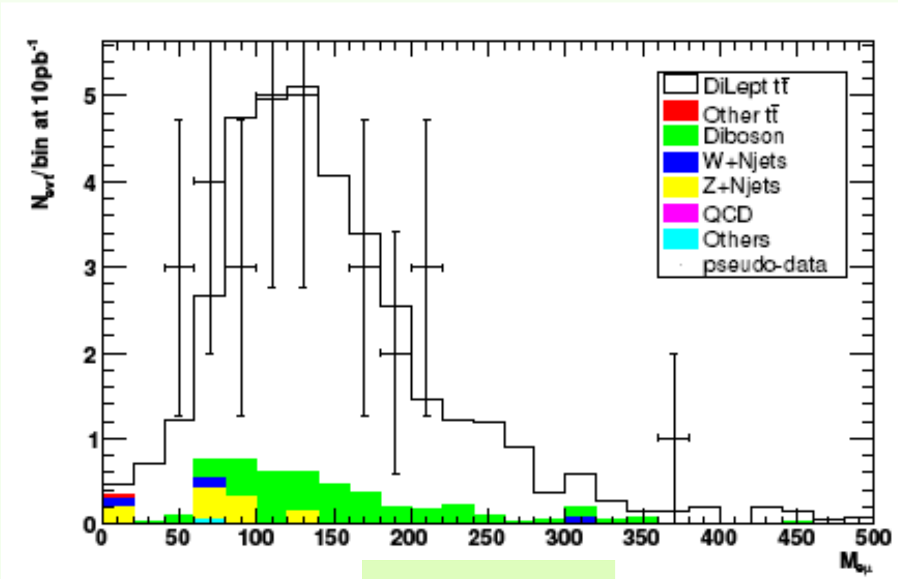
Requiring in the selection at least 2 jets give:
S/N ~ 8
signal ~ 100
σ/Δσ (stat) ~ 10%

back of the envelop

10/pb at 14 TeV

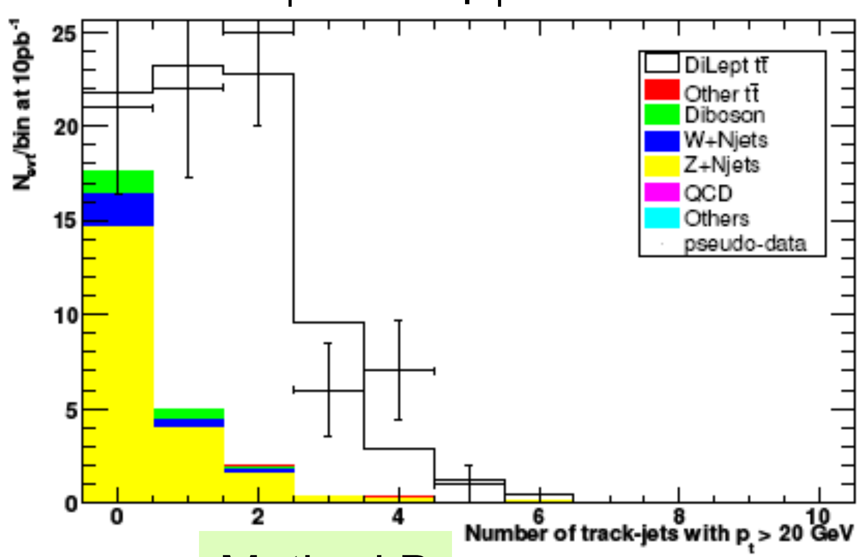
Special di-leptons: e+μ

- **Di-lepton e+μ channel also visible with only lepton cuts**
  - **A. Increased kinematic thresholds of p_T > 40 or 35 GeV on the lepton**
  - **B. p_T>25 or 20 GeV thresholds + cut on at least 2 track-jets p_T>20GeV**
- **Plots after the event selection:**



Method A

k_T D=0.7 p_T>20GeV



Method B

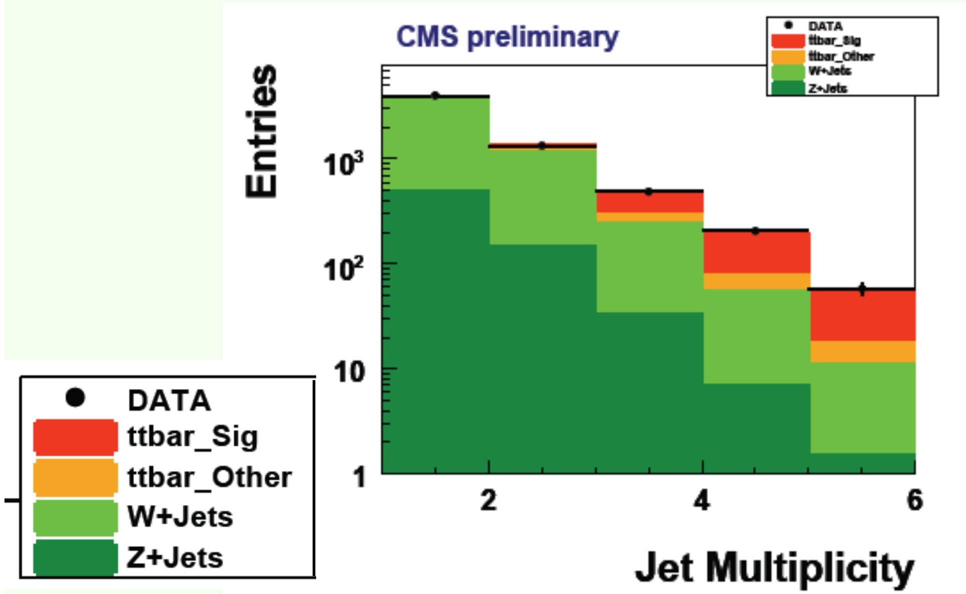
- **Small signal but also very few background → very early cross section!**

Selection: lepton+jets

10/pb at 14 TeV

- **Larger branching ratio but only one isolated lepton (here muon channel)**
- **Apply the `HLT1MuonNonIso` (p_T>30GeV → 91% efficiency plateau)**
  - **Exacly 1 muon with p_T>30GeV & $|\eta|$<2.1 + isolation**
  - **At least 4 jets with E_T>40GeV & $|\eta|$<2.4**

	$t\bar{t}$ (signal)	$t\bar{t}$ (other)	W+jets	Z+jets	QCD
Preselection	749	527	7474	1430	–
4 Jets $p_T > 40$ GeV	240	137	85	17	–
1 Muon $p_T > 30$ GeV	166	32	59	9	38



Requiring in the selection
at least 4 jets and
tight isolation give :
S/N ~ 1.5
S/N (m_t window) ~ 3
S/QCD ~11.6
signal ~ 128
$\sigma/\Delta\sigma$ (stat) ~ 15-20%

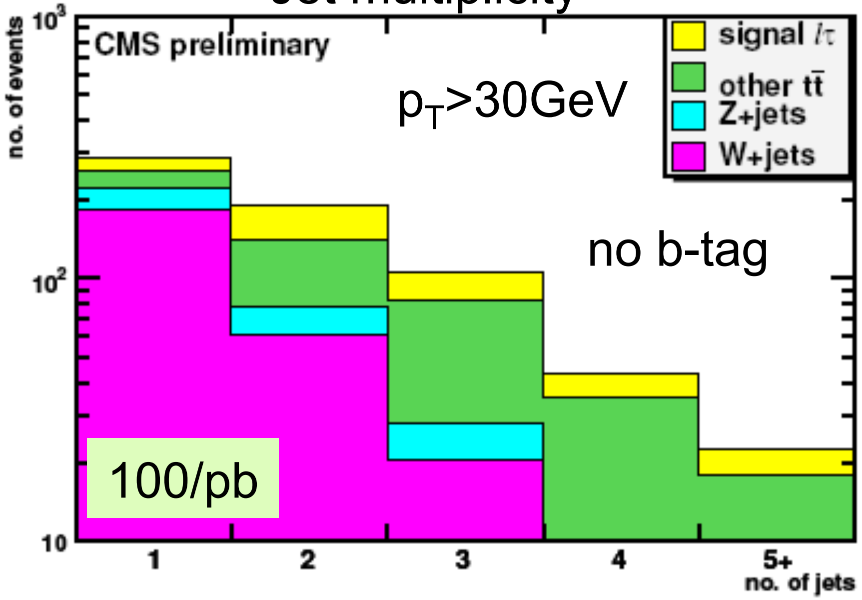
back of the envelop

LISBON

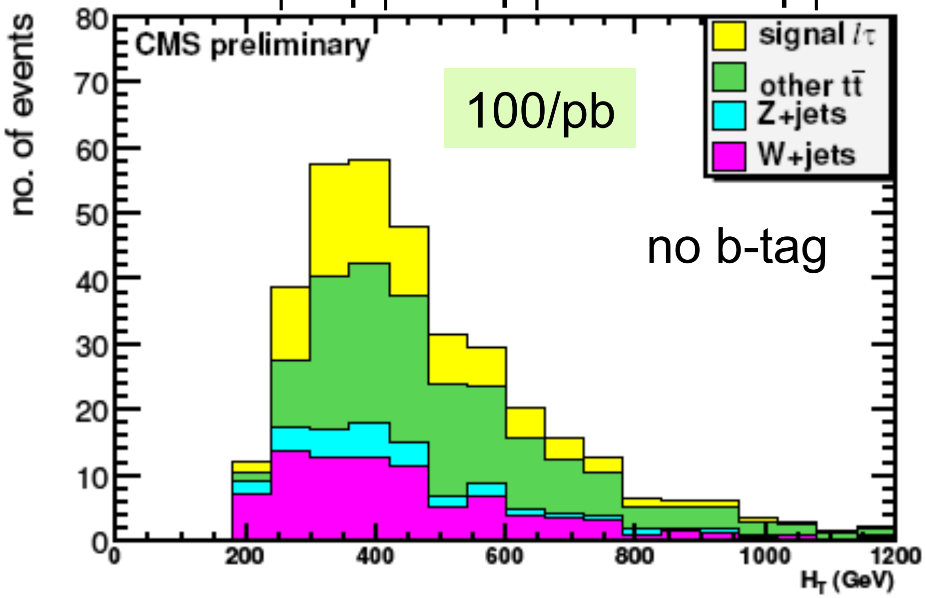
Tau's visible this year?

- **Dedicated event selection (isolated lepton + MET>60GeV + 2 b jets)**
- **One tau lepton (CaloTau) with general tau tagging algorithms**
- **An opposite charge is required from lepton and leading track in tau**

Jet multiplicity



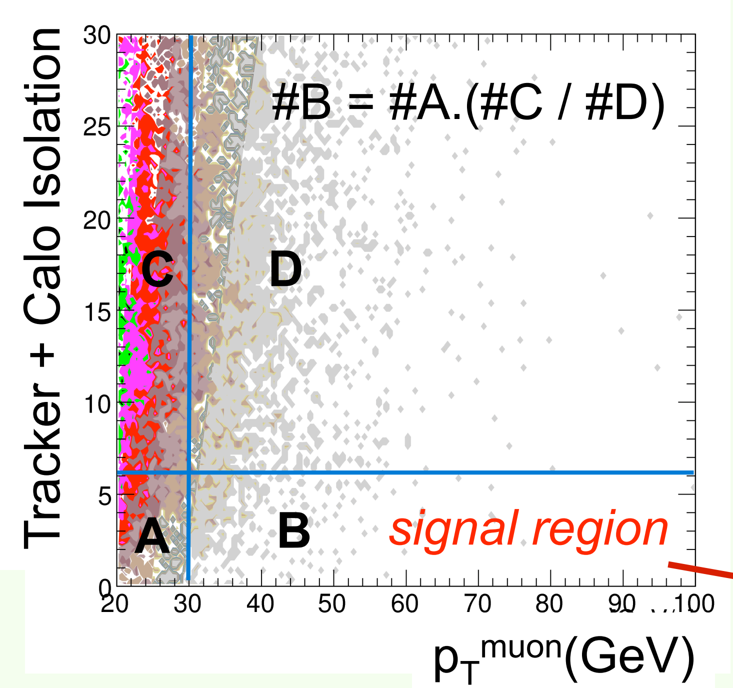
$H_T = p_T^{e,m} + p_T^{\tau} + MET + p_T^{jets}$



- **After this selection**
  - **1 prong : S/N ~ 0.40 (S~7.3 events for 10/pb at 14 TeV → S/sqrt(B)~2)**
  - **3 prong : S/N ~ 0.14 (S~1.3 events for 10/pb at 14 TeV → S/sqrt(B)<1)**

Estimating background from data

- QCD background in μ+jet channel via "ABCD" method
- Analysis using a biased PYTHIA multi-jet sample of 8.7/pb
- Selection: p_T^{μ}>20GeV (|η|<2.1) & E_T^{jet}>30GeV (|η|<2.4) & HLT1MuonNonIso



- Estimate the QCD level in the signal region taking ttbar and W+jet events into account

QCD jets W+jets ttbar	p_T^{μ}<30GeV	p_T^{μ}>30GeV
Iso>6GeV	16384+8+38	342+6+13
Iso<6GeV	1903+63+128	**44+51+101**

estimation

real number =40 events

effect of signal bck in A/C/D regions is 4 events

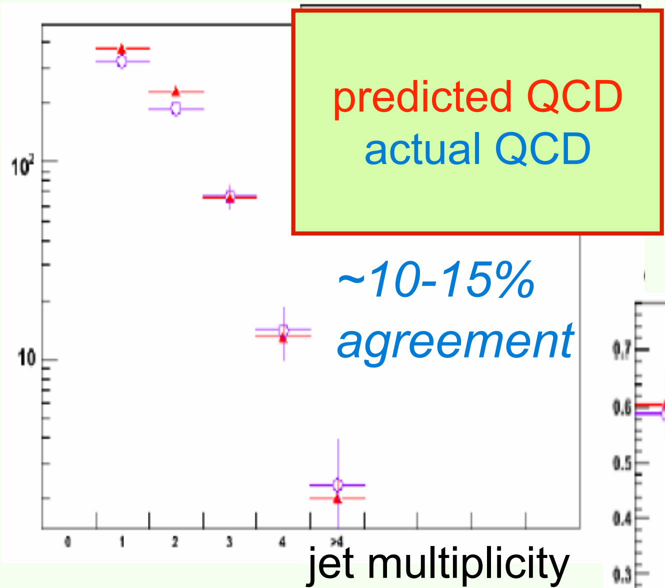
- QCD estimation stable versus the cuts on p_T^{μ} & Iso applied
- Work in progress to get the ABCD method with alternative observables

Estimating background from data

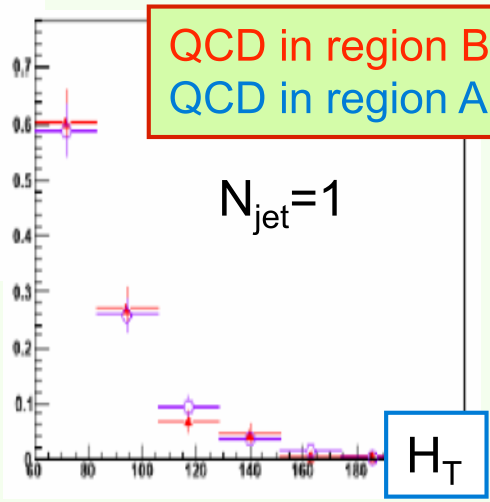
- **QCD background in μ+jet channel via "ABCD" method**
- **Selection: nominal selection used in this channel (cfr. previous slides)**
- **Different relative isolation observable:**

$$RelIso \equiv \frac{P_T^{\mathrm{lepton}}}{P_T^{\mathrm{lepton}} + \mathrm{sum}} > 0.92$$

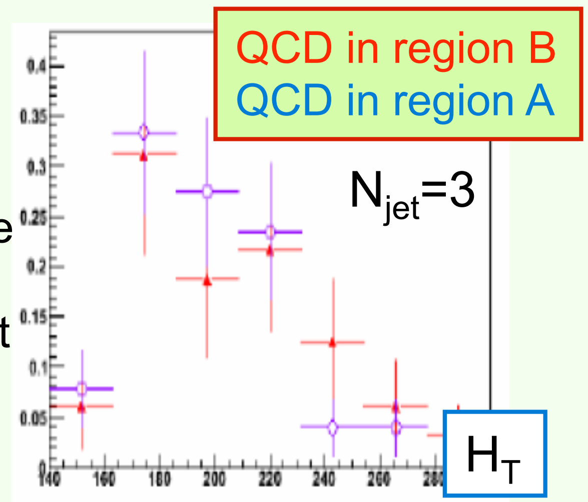
with sum = ΣP_T(tracks)+EM+HAD in a 0.3 cone around the muon



- **Using two different variables**
  - **sign. of d_0 impact parameter muon**
  - **isolation variable RelIso**
- **Predict data-driven distributions by assuming distributions independent of d_0 significance**

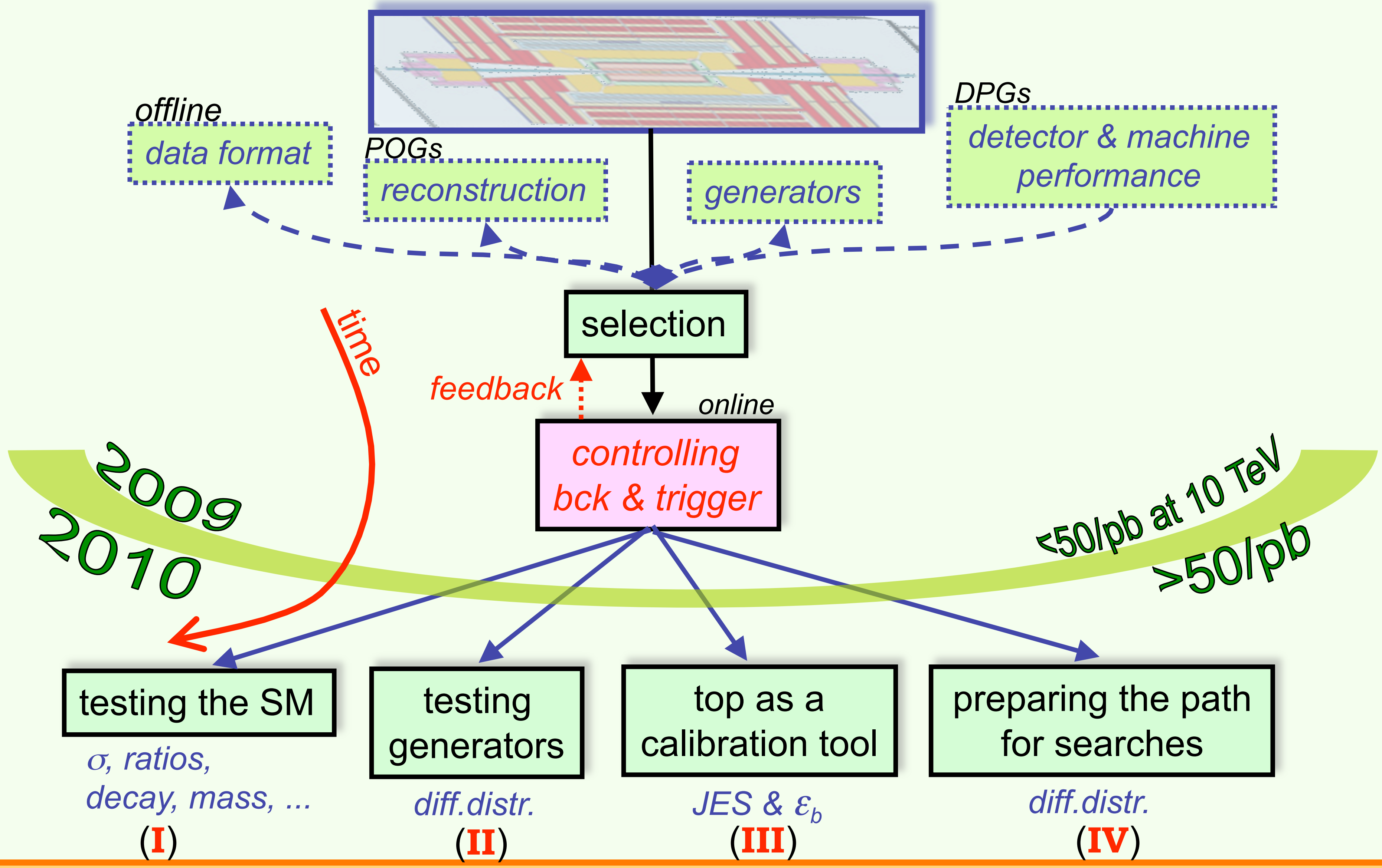


reasonable
good
agreement



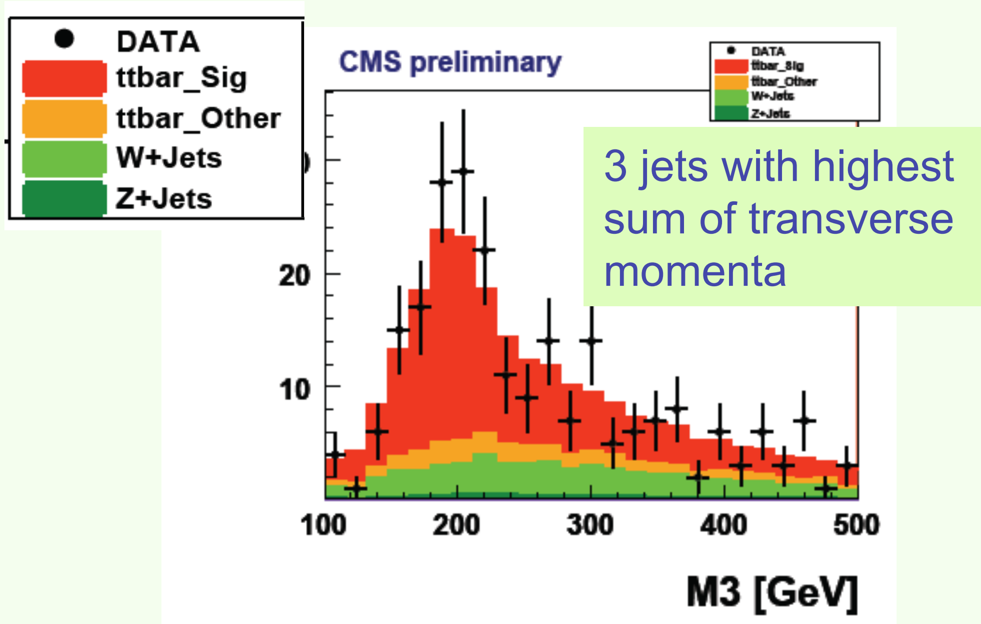
Not enough simulation to test in the signal region $N_{jet} \geq 4$

Getting ready to learn something



(I) Testing the Standard Model

- **After the event selection we have to convience ourselves that we see the Standard Model top quark, hence measure its properties**

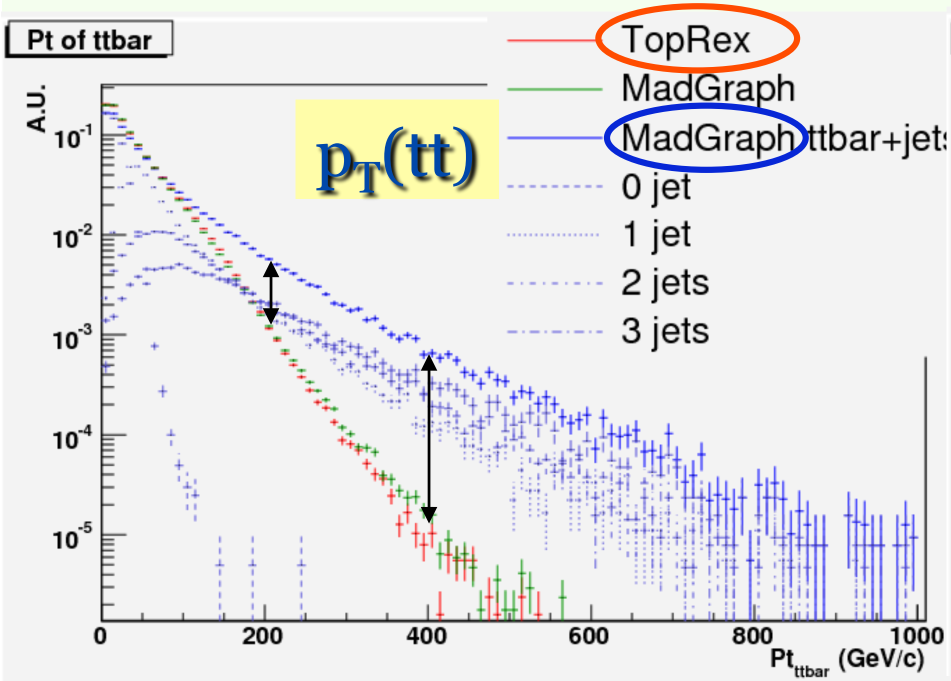


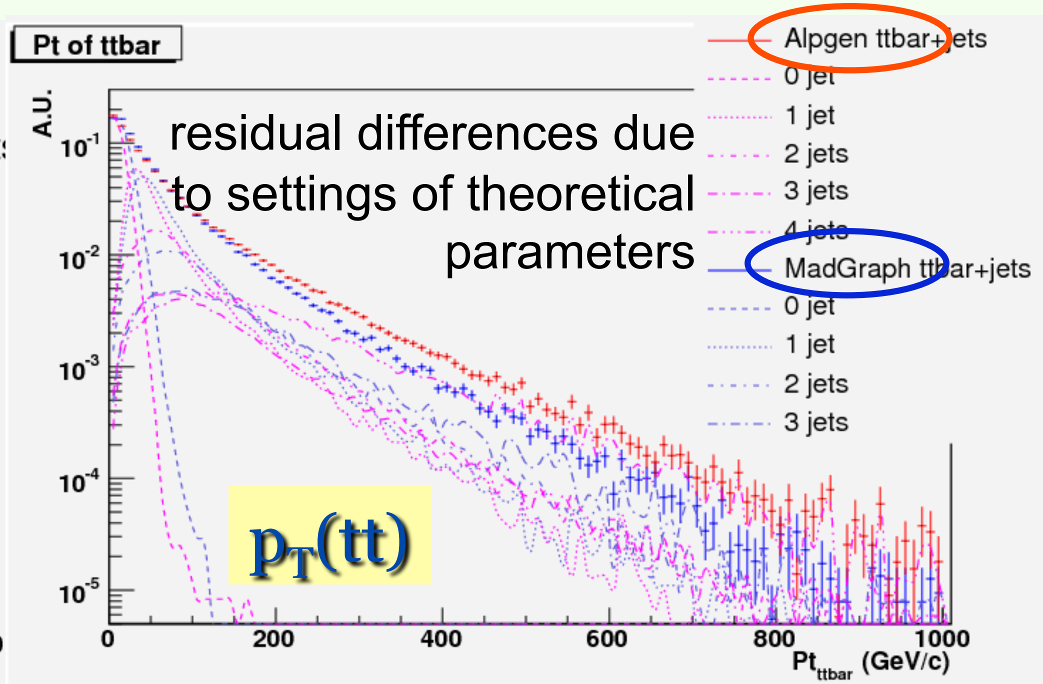
- **Usually this requires to combine the jets into a t→bW→bjj tree**
- **Several methods explored from simple choices to multi-variable Likelihood Ratios**
- **We reach jet combination efficiencies of ~30% from simple to ~70% of advanced methods and looking in a window around m_{top}**

- **Ongoing activity to estimate the W+jet background from data itself**
- **Trivial but important remark**
  - **If you select the 4 highest E_T jets in the event, it happens only in ~20-30% of the events that you find that these jets match the 4 primary quarks**

(II) Testing the generators

- Physics generators use several parameters which receive a value motivated by theoretical arguments or via tuning on data
- Several distributions in top events are sensitive to those parameters

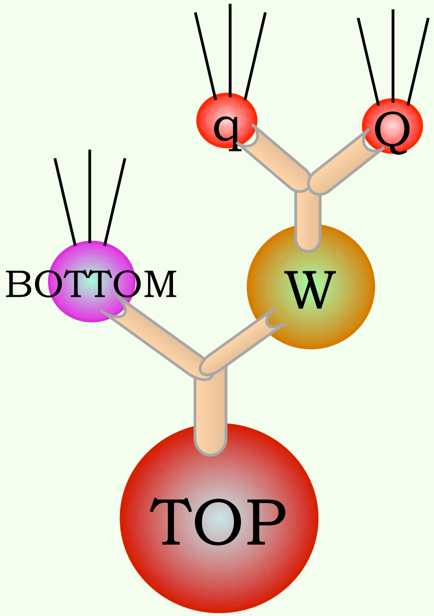




- beyond the MC validation of the generators
- effort has started to estimate these effects (joint with EWK/Top group)
- reconstruct and understand the top quark relevant differential distributions

(III) Top as a calibration tool

- In the top decay we have two mass constraints and one flavour constraint if we assume the Standard Model
  - m_W has been measured with a precision of 0.03 %
  - m_t has been measured with a precision of 0.8 %
  - flavour constraint $BR(t \rightarrow bW) = 1$



- Methods are put in place to use 100/pb of data at 10/14 TeV to estimate the b-tagging efficiency and the Jet Energy Corrections

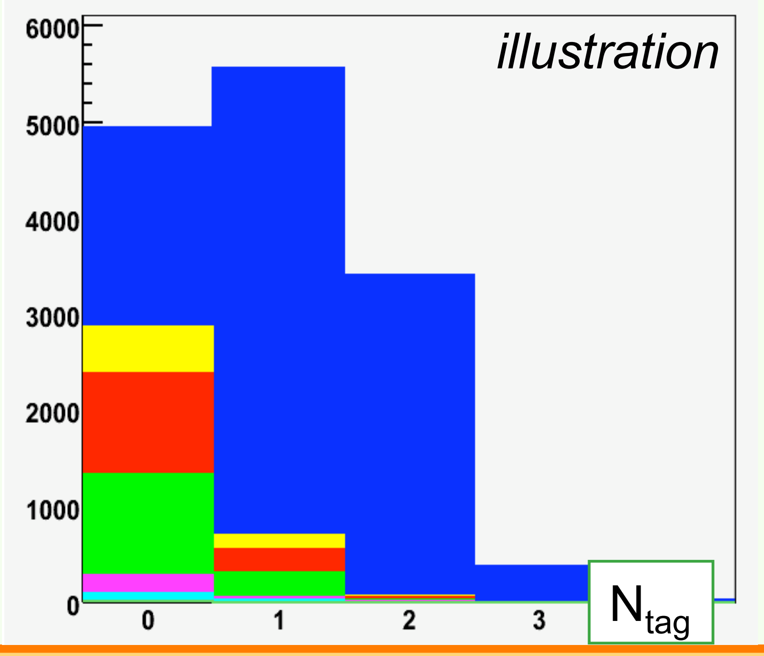
Top as a calibration tool: b-tag

Method 1 : 'tag consistency' method

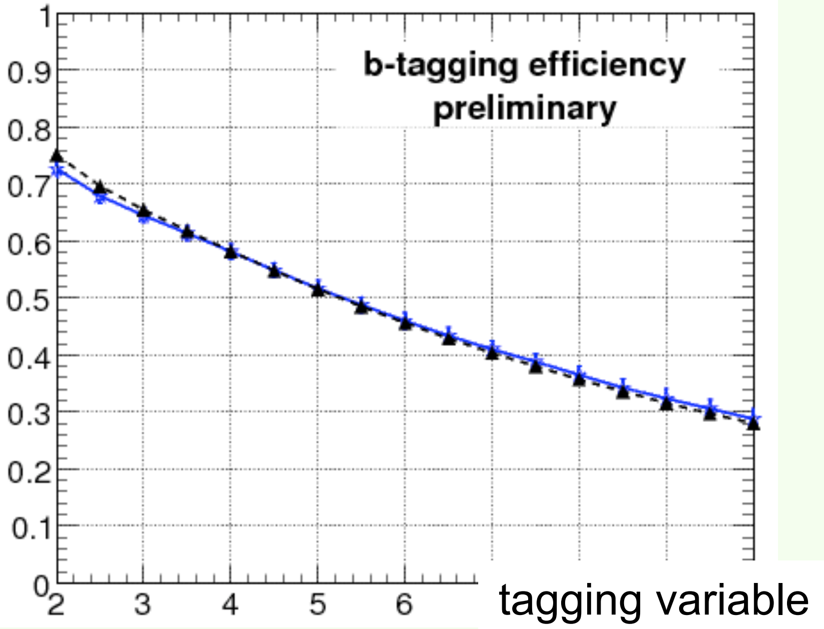
- reconstruct the N_{tag} distribution from semi-leptonic ttbar decays
- consistency between observed and expected number of tagged events
- maximize log likelihood and find ε_b

$$L = -\log\left(Poisson(N_1, \langle N_1 \rangle) \times Poisson(N_2, \langle N_2 \rangle) \times Poisson(N_3, \langle N_3 \rangle)\right)$$

- get N_{tag}=1, N_{tag}=2, N_{tag}=3 from data where the expected $\langle N_{tag} \rangle$ depends on the selection efficiency, the cross section and the b-tag efficiency



plot show agreement of MC ε_b and ε_b obtained with the method

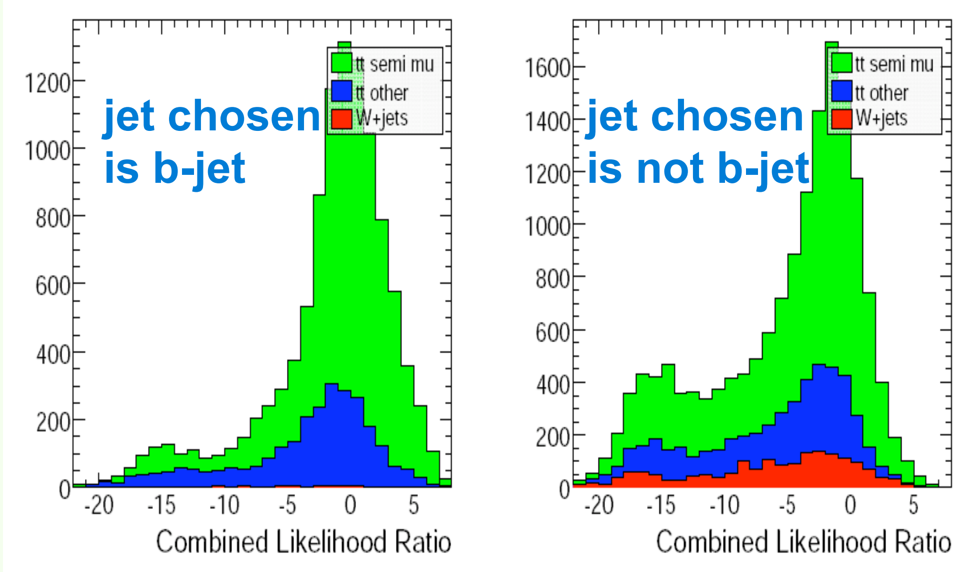


Top as a calibration tool: b-tag

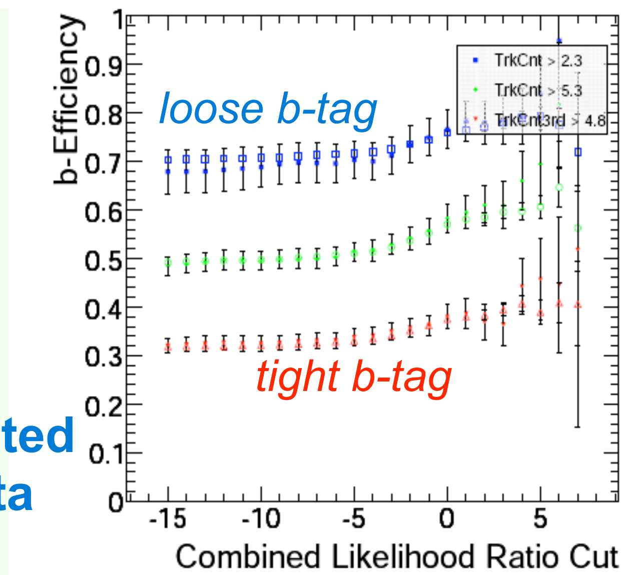
Method 2: Likelihood ratio method

- in both semi-leptonic and fully leptonic ttbar decay channel
- use likelihood ratio of several observables to select a pure b-jet sample
- b efficiency from

$$\epsilon_b = \frac{1}{x_b}[x_{tag} - \epsilon_o(1 - x_b)]$$



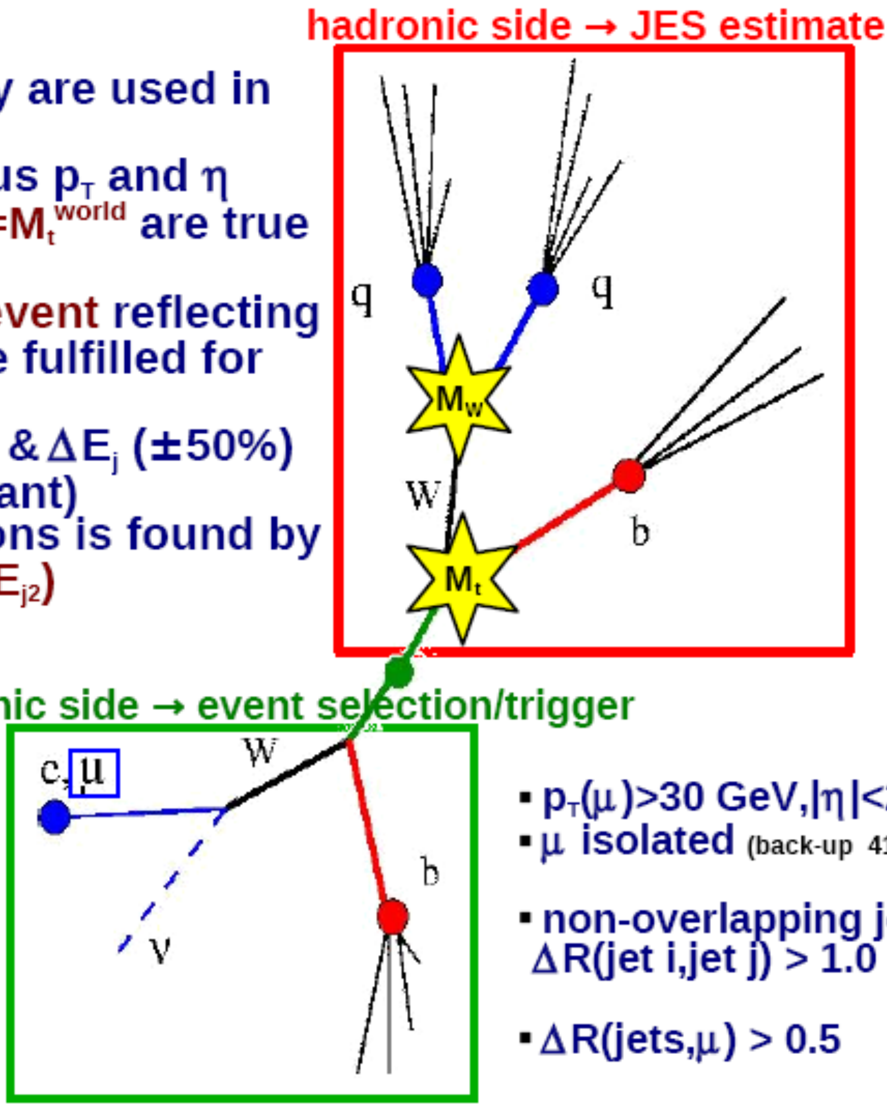
- the fraction of tagged jets x_{tag} is obtained on the purified sample
- the mistag rate and the purity of the sample is currently obtained from MC



- Plot shows the b-tag efficiency as function of the combined likelihood ratio cut compared to MC expected
- Now looking for a way to estimate the purity from data

Top as a calibration tool: JEC

- The 3 jets from the hadronic top decay are used in an event-by-event kinematic fit
- Jet resolutions are parametrized versus p_T and η
- The constraints $m_W^{rec}=M_W^{world}$ and $m_t^{rec}=M_t^{world}$ are true at parton level
- Kinematic fit returns a $P(\chi^2)$ for each event reflecting the probability that the constraints are fulfilled for this event
- A whole range of JES corrections ΔE_b & ΔE_j (±50%) is scanned for each event (E/|p| constant)
- The best estimate of the JES corrections is found by minimizing the function $\chi^2(\Delta E_b, \Delta E_{j1}=\Delta E_{j2})$

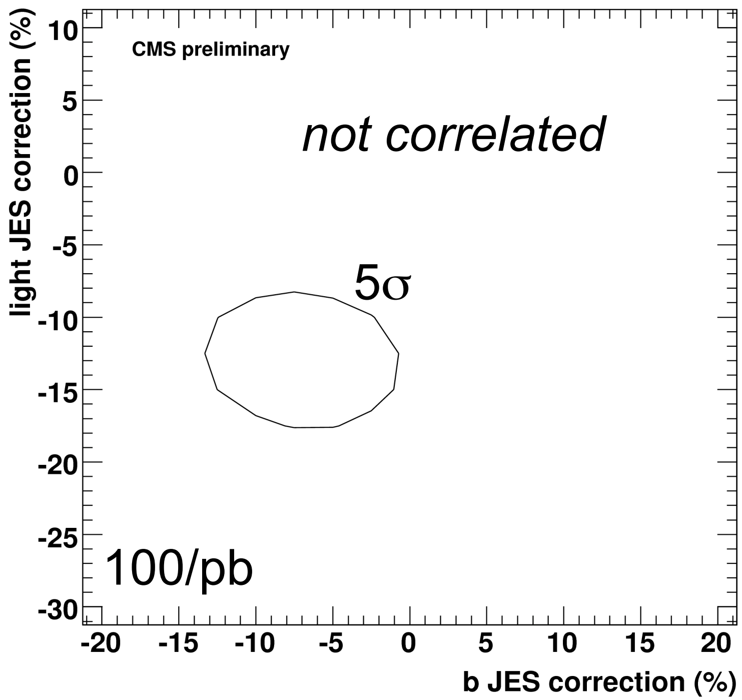


- To reduce the process background a tight event selection is applied
- A likelihood ratio is constructed to identify the correct jet combination
- A cut on this likelihood ratio is made to reduce combinatorial background
- To reduce contributions from mis-reconstructed events cuts are made on the probability of the kinematic fit

- $p_T(\mu)>30$ GeV, $|\eta|<2.1$
- μ isolated (back-up 41)
- non-overlapping jets: ΔR(jet i,jet j) > 1.0
- ΔR(jets,μ) > 0.5

Top as a calibration tool: JEC

- Hence for each event a $\chi^2 = \chi^2(\Delta E_b, \Delta E_{q1}, \Delta E_{q2})$
- When estimating an inclusive correction we can put $\Delta E_{q1} = \Delta E_{q2}$
- Hence we obtain a confidence interval in 2 dimensions: ΔE_b & ΔE_{q1},

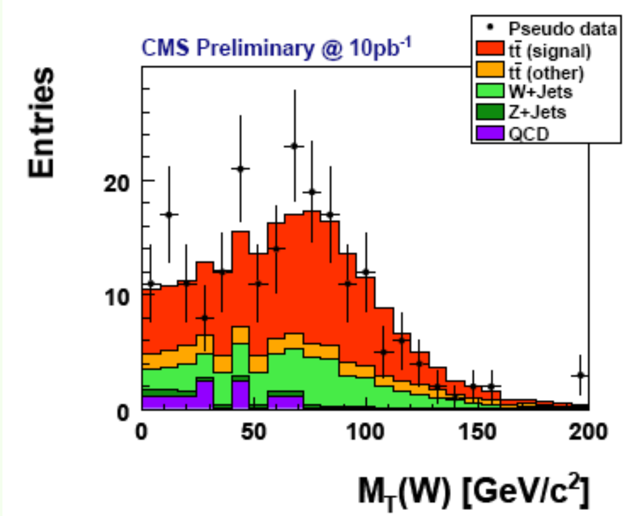


- The residual jet energy correction is
  - $\Delta E_b = -7.0 \pm 0.9$ %
  - $\Delta E_q = -12.9 \pm 0.9$ %
- These uncertainties are corrected to have a unity width of the pull distribution
- These data-driven numbers agree well with the MC expectation and the method can therefore serve as a measure and a closure test for JEC's.

- with 100/pb we could have a precision of about 1% on the JEC
- effort to project vs (p_T,η)-jet to profit optimal from this data, and going towards a combine JEC/m_{top} measurement

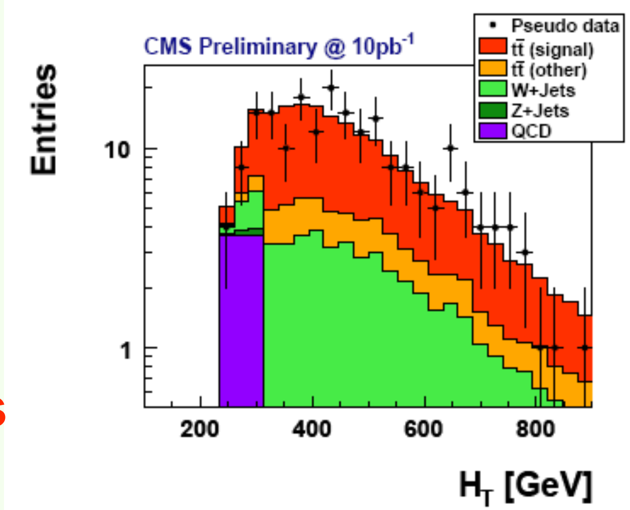
(IV) Preparing the path for searches

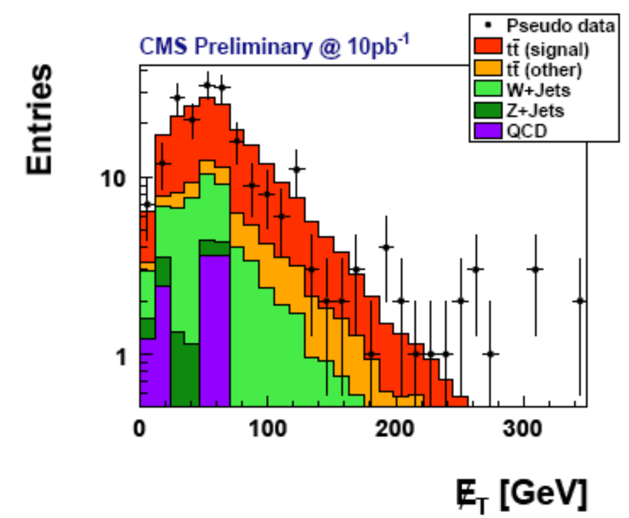
- **Several differential distributions can go beyond testing the Standard Model and are sensitive to new physics**
- **We need to understand the SM part of the distribution before we start looking in the part sensitive to new physics**
- **Including the systematic effects...**



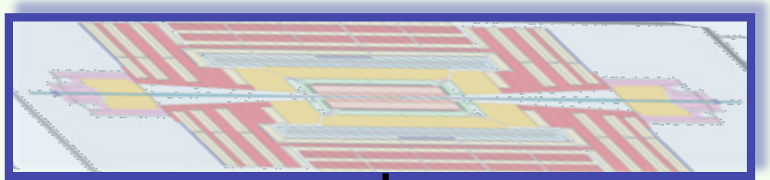
- **Examples:**
  H_T, MET, p_T^{top}, p_T^{ttbar}, p_T^{lept}, m_{ll}, m_T(l+MET), topo. variables, ...

- **Need to increase the activity and ideas in this direction**





En route to data…



offline

data format

we have settled on unified formats for Physics Analyses

Physics Object Groups

reconstruction

generators

Detector Performance Groups

detector & machine performance

selection

finalizing, now effort to equalize them

online

controlling bck & trigger

- **main effort and good progress**
- **ideas for bck estimates are being implemented and tested**
- **settling on control triggers**

2009

2010

testing the SM

good progress

testing generators

need more effort for data driven validation

top as a calibration tool

good progress

preparing the path for searches

need some kick-off

Summary/Outlook

- The ATLAS and CMS detectors are different, hence do not expect the same performance of reconstruction tools
- The physics however is the same, hence we challenge the same limits (choice of correct jets, estimation of bck, …)
- The focus in CMS is on the early physics (<50/pb) and how to use known physics to commission our experiment
- I'm sure we can learn a lot from each others work

Some important publicity…

TOP2010 Conference

30th of May – 5th of June 2010
Brugge, Belgium

CP3 - IIHE