Reconstructing Z→bb Signal and Measuring b-specific Energy Scale at CDF

- **Introduction: Z→bb and b-jet energy scale**
- **Measurement method**
- **Preliminary results on data (no final numbers … yet!)**

Top Workshop In2p3: from DØ to ATLAS

Grenoble 9-10.10.06

Julien Donini
University of Padova, INFN

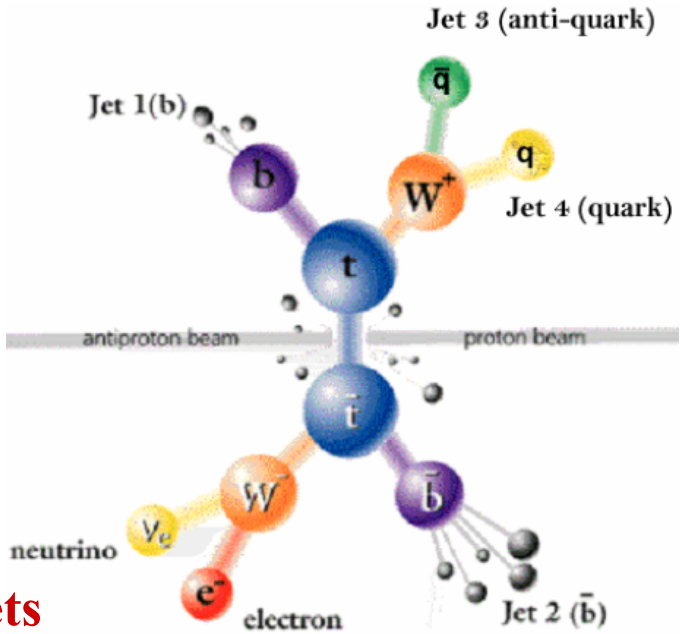
Introduction

A large part of **Tevatron** physics is done from jets. The best resolution achievable on jet energy is needed for QCD studies, **Higgs** boson searches

… and **Top quark** measurements:

All the top quark decay modes involve jets

- t→Wb (~100%)
- W→jj (~67%)



Jet Energy Scale is generally the **largest source of syst. uncertainty** in **top mass** measurement.

b-jets have different characteristics from generic jets

- Harder **fragmentation**
- B hadron **decays** (21% semi-leptonic)

Necessity to determine specific **b-JES**: data/MC scale factor.

Can be estimated using, specific **physic processes**: γ-b jet, **Z→bb resonance.**

Z→bb and b-JES at CDF

Extract a $Z^0 \rightarrow bb$ resonance in CDF Run II data using a dedicated trigger. Use this signal as a tool to:

- **study data/MC properties for b-jets:**
  - **determine b-specific jet energy scale**

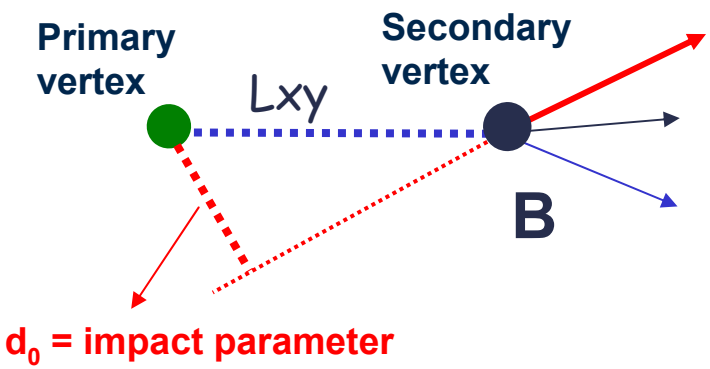
    Useful in particular for **top physics**: reduce JES systematic uncertainty on m_{top}.
  - **bbbar resonance: data/MC resolution studies.**
- **develop and test algorithm to improve b-jet energy resolution:**
  - **NN** (training with MC, apply to Z resonance in data),
  - other multivariate approache: **Hyperball** (exploit correlation between variables to correct b-jet energy).

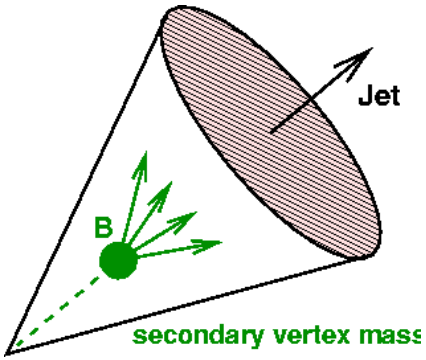
  It's crucial to improve bbbar pair dijet mass resolution for **Higgs** searches.

b-jet identification at CDF (1)

Heavy flavor jets (b, c quarks):

- **secondary** vertex
- charged tracks with **high impact parameter**



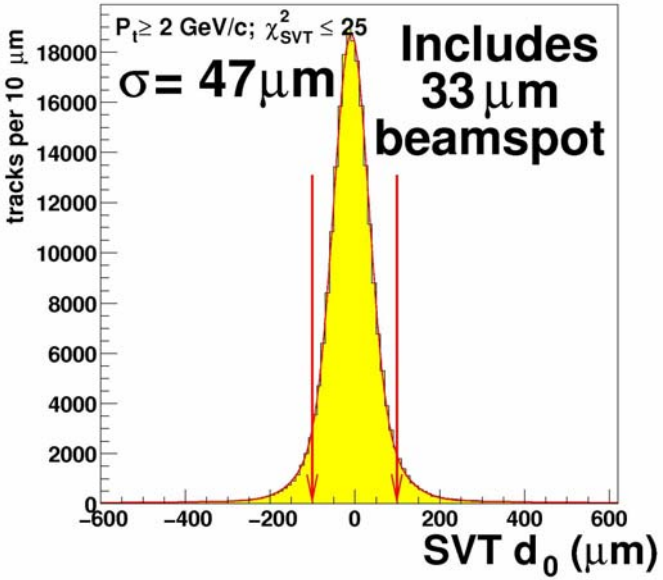


CDF II: specific L2 trigger (SVT):

Information given by the internal **trackers are used** to select events with **high impact parameter (d0) tracks.**

SVT has proven crucial for most of CDF II's B physics program.

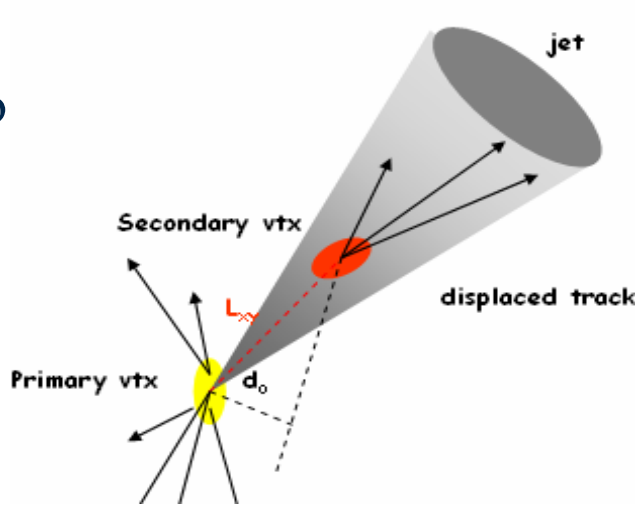
Silicon Vertex Tracker (SVT)

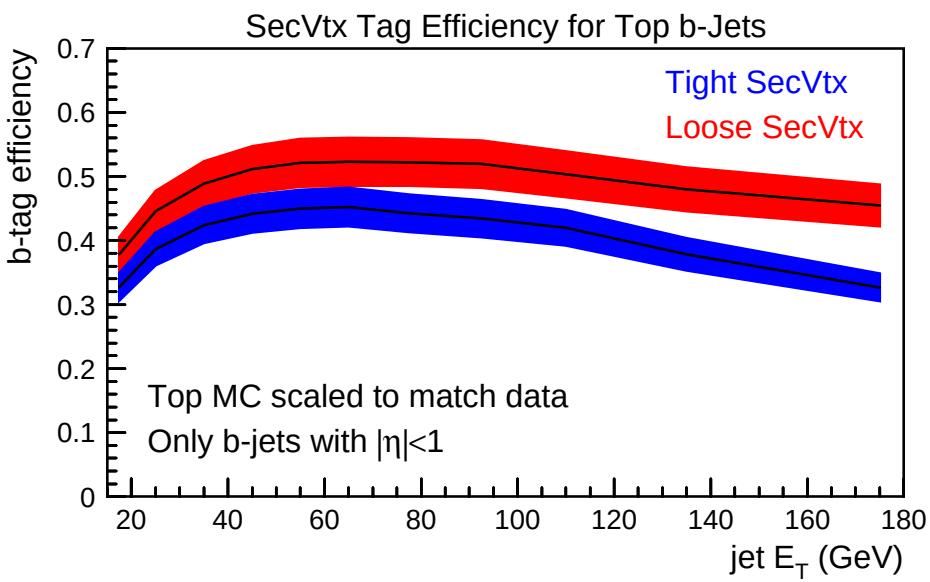


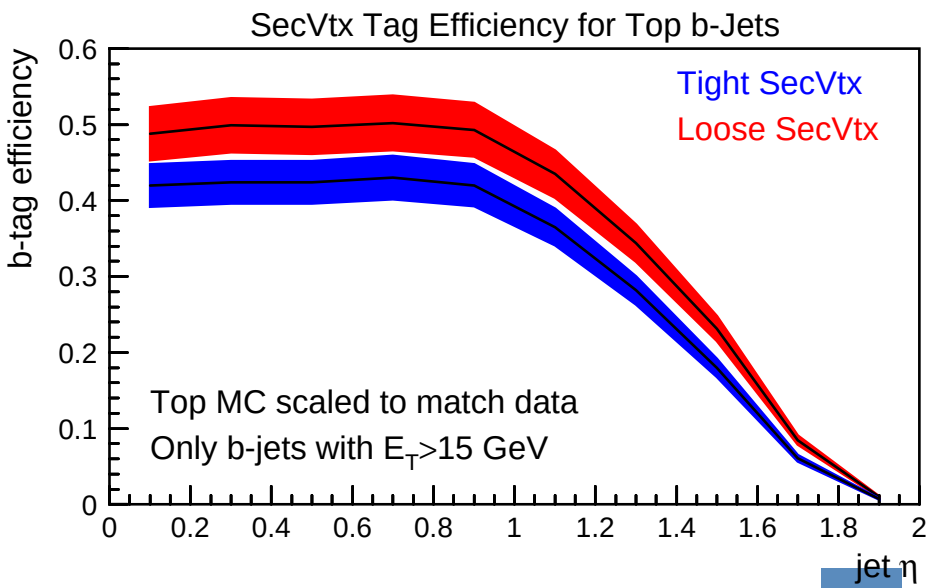
b-jet identification at CDF (2)

b-tagging: Secondary VerTex tagging (SecVtx)

- Tracks with **significant IP** are used in a iterative fit to **identify the secondary vertex inside the jet**
- Efficiency drops at low jet Et and high rapidity but is **45-50%** for central top b-jets
- Mistag rates are kept typically at **4-5%**







Event Selection

To observe the Z→bb resonance we need

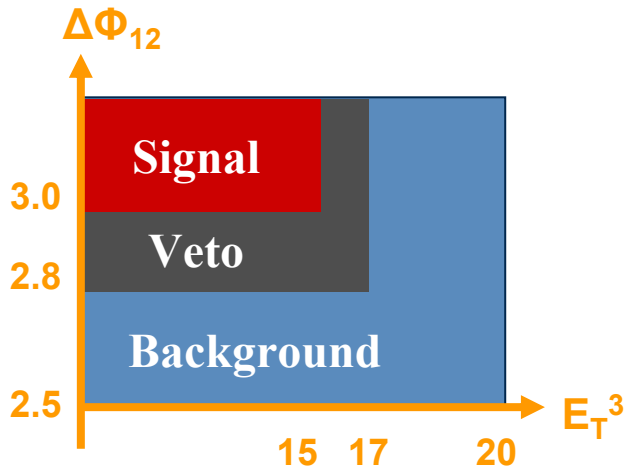
- **sample with high purity of bbbar events**
  - A **specific trigger**: basically it **requires 2 high Pt, displaced SVT tracks, two Et>10 GeV clusters**.
  - Event selection and tagging: high fraction of pure bbbar pairs, **~ 90% for events with 2 tags.**
- **low cut on jet E_t: dijet mass spectrum that peaks below the Z mass signal**
  - not too low or **trigger biases** appear at low dijet masses
- **kinematical variables to help separate signal from QCD background**
  - back-to-back events with low extra jet radiation helps reduce QCD background and ISR/FSR: **$\Delta\Phi$(jet 1, jet 2), E_t(3rd jet)**
- **Background model: data-driven.**

Data-driven BG Modeling (1)

[$\Delta\Phi_{12}$, E_t(3rd jet)] Space

Signal region: region with significant fraction of Z→bb signal (>2%),

Background region: excludes signal (and veto) region. Poor in signal (<<1%) .



Computation of the BG dijet invariant mass shape

Tag rate **as a function of dijet invariant mass** in **BG region**:

- **Tag Ratio (TR) = N(++)/N(00)** **[(+): tagged jet, (0): 'taggable' jet]**

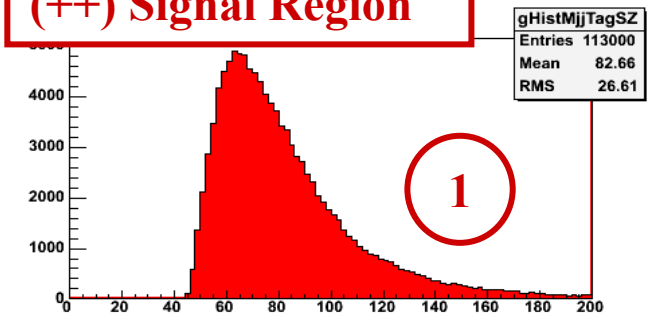
Number of BG events, **in signal region (SR)**:

- **BG shape: N_{BG}(++) = TR x N_{SR}(00)**

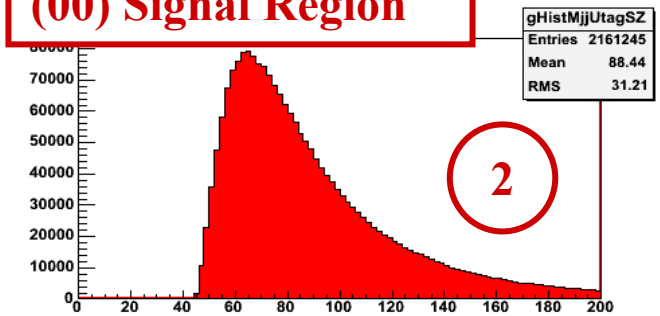
The ***BG shape*** thus obtained is then used to fit the *(++) data in the signal region*.

Data-driven BG Modeling (2)

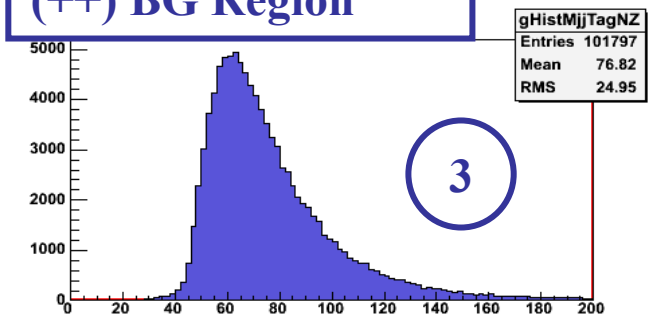
(++) Signal Region



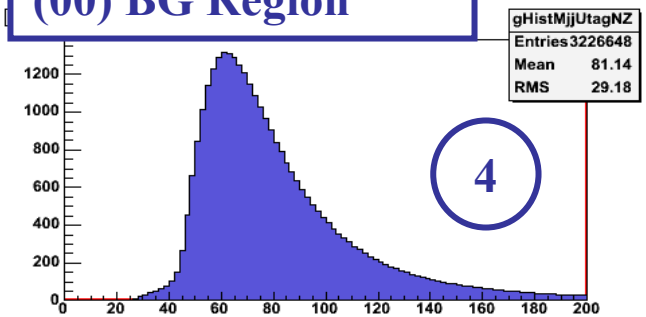
(00) Signal Region



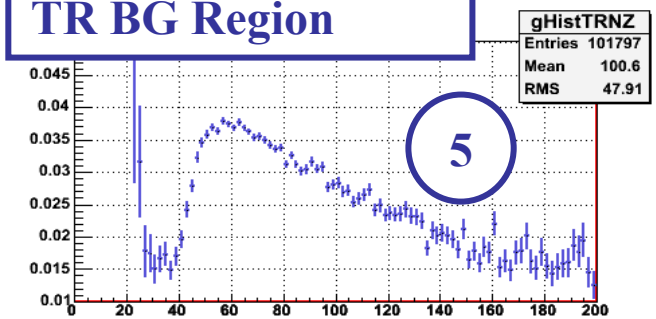
(++) BG Region



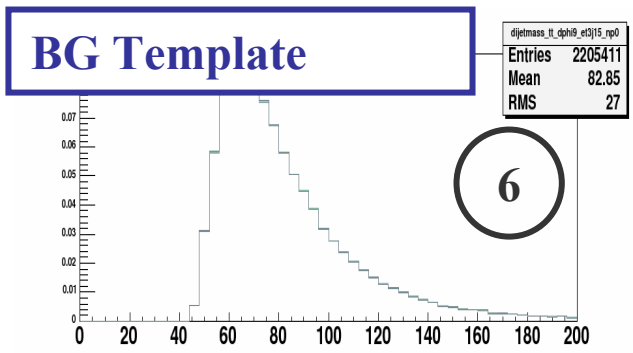
(00) BG Region



TR BG Region



BG Template



Tag rate vs mjj:

$3 / 4 \rightarrow 5$

BG template:

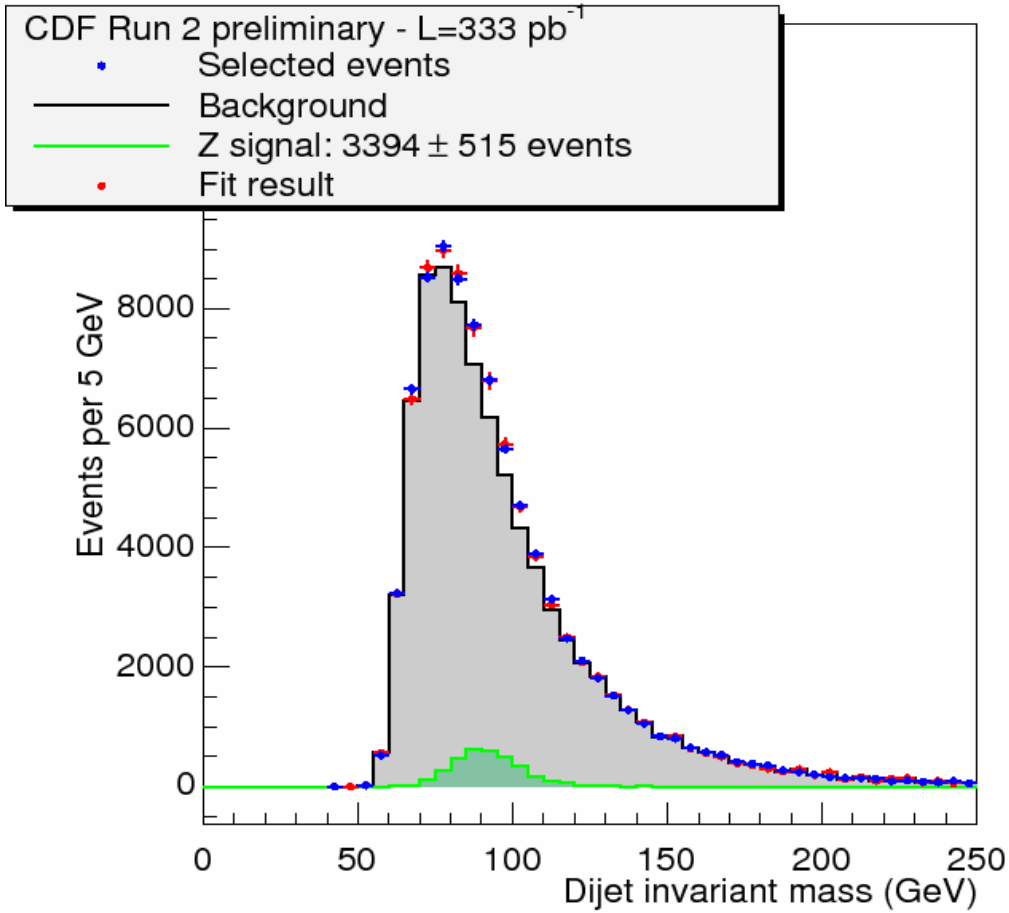
$5 \times 2 \rightarrow 6$

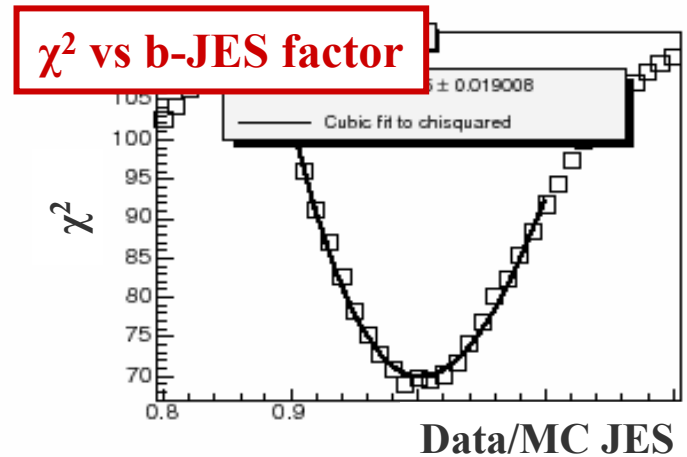
Fit to the data:

Use 6, + MC signal template.

Z→bb Signal Observation (02.05)

Data is fit to: BG shape (data-driven), and Z→bb MC templates (Pythia).





Varying data/MC scale of Z MC templates allows to **fit for b-JES**.

About **3400 events of signal** were reconstructed (out of 86000 dijets events).
Statistical uncertainty on **b-JES ~ 2%.**

Extraction of Z→bb Signal

Complex trigger, many modifications during the first two years of data taking.

- **difficult to combine data** and use an unique BG model;
- need to perform a **tight E_t jet selection** to avoid **trigger biases**;
- **systematic on b-JES** difficult to evaluate with signal that close to BG peak.

However first result seemed very promising !

… then a lot of effort was put on to:

- **Improve implementation of ZBB trigger**
  - lower selection on dijet events
  - separation signal from background turn-on
- **Improve fitting method**

Unbinned Likelihood

We use an unbinned likelihood procedure to measure the number of signal events and the b-JES scale factor in our data.

$$\mathcal{L}(SF) = \mathcal{L}_{shape}(SF) \times \mathcal{L}_{(n_s+n_b)}, \text{ with}$$

$$\mathcal{L}_{shape}(SF) = \prod_{i=1}^{N} \frac{n_s P_s(m_i; SF) + n_b P_b(m_i)}{n_s + n_b}, \text{ and}$$

$$\mathcal{L}_{(n_s+n_b)} = \frac{e^{-(n_s+n_b)}(n_s+n_b)^N}{N!}$$

n_s and n_b: numbers of signal and background events,

$P_s(m_i,SF)$ and $P_b(m_i)$: signal and background p.d.f's.

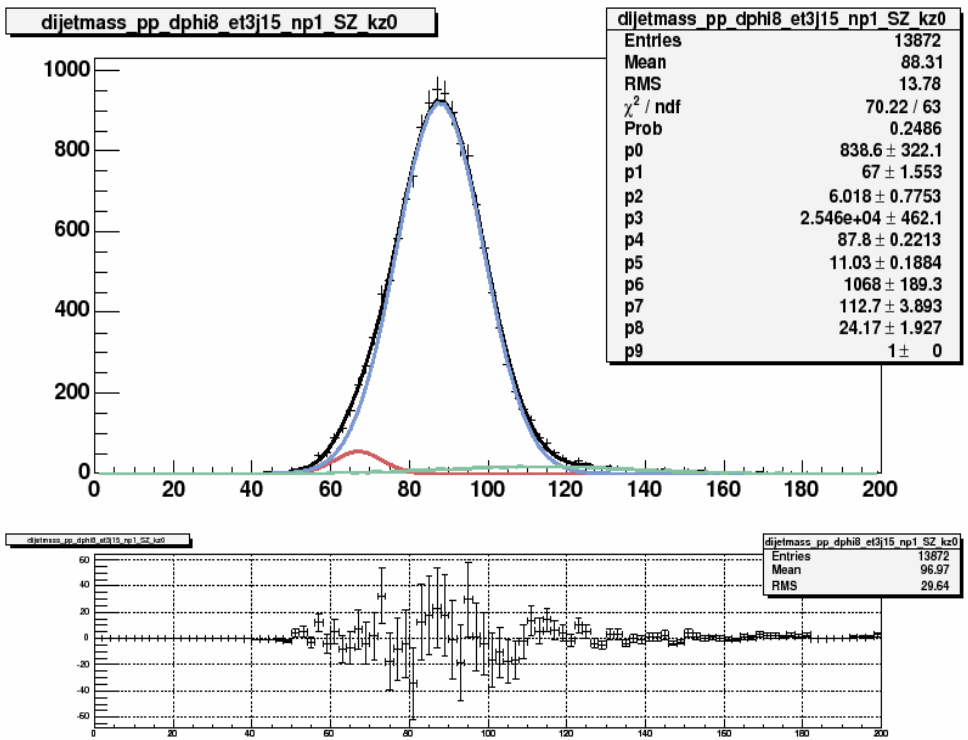
We minimise –ln(L) to find best SF hypothesis. Statistical error is given by the difference between this SF and the SF at $-\ln(L_{max}) + 0.5$.

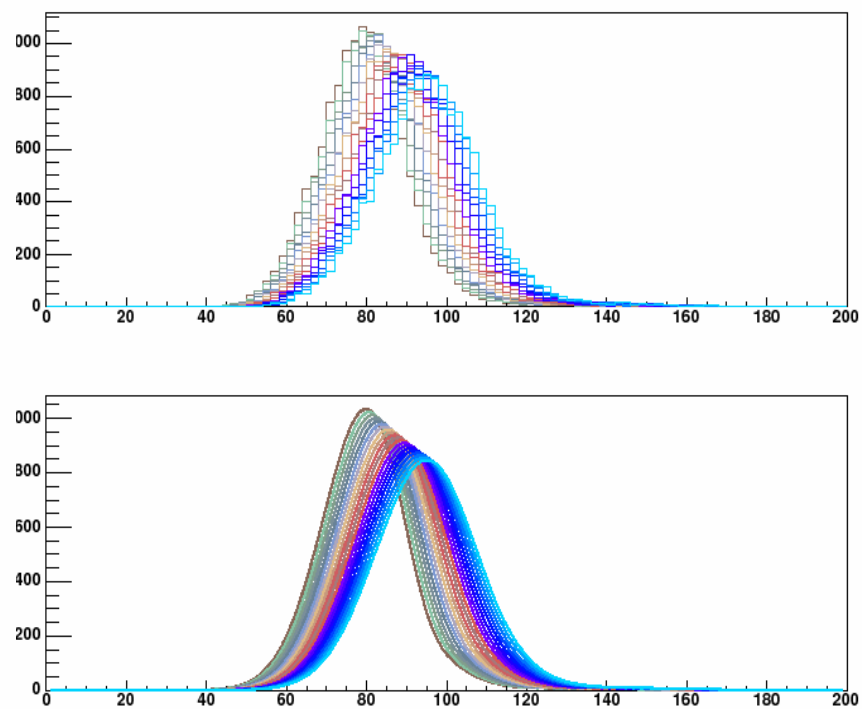
Construction of signal PDF

Dijet invariant mass shape is obtained from MC (pythia, Z→bb + MB events).

For events passing trigger simulation, event selection and tagging request (2 tagged leading jets) we apply different b jet Scale Factors: 0.9-1.1.

→ we construct a set of 20 MC Z→bb templates and fit each of them with a sum of three gaussian functions.





Signal PDF as a function of SF

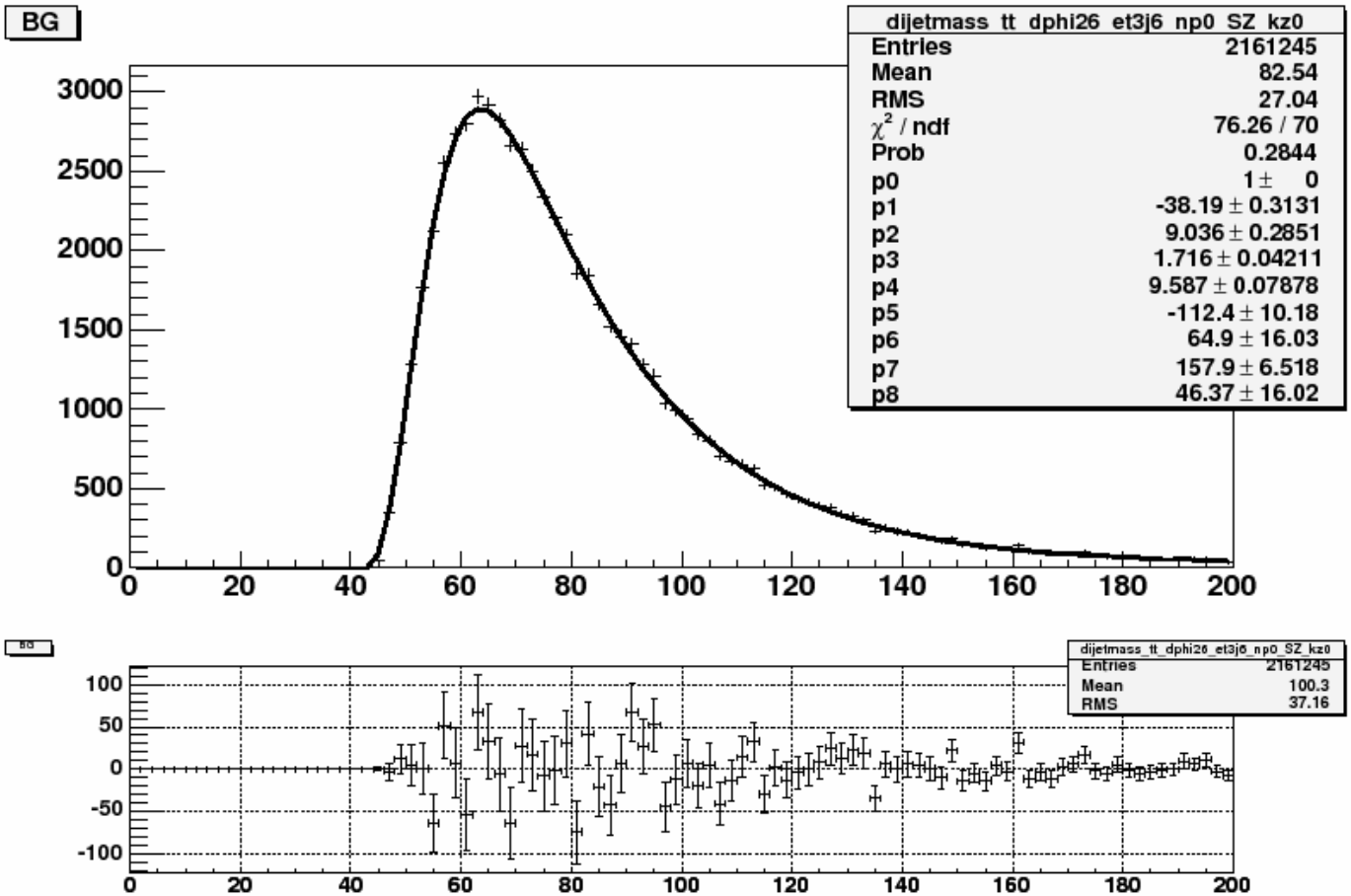
We construct on single signal PDF which has **b-jet energy scale factor (SF) as a parameter:**

$$P_s(m_{jj}, SF).$$

To do this we fit the 20 templates contemporarily and **let each parameter vary linearly with SF**.

Background PDF

Parameterization for data-driven BG model:



Pseudo-experiments

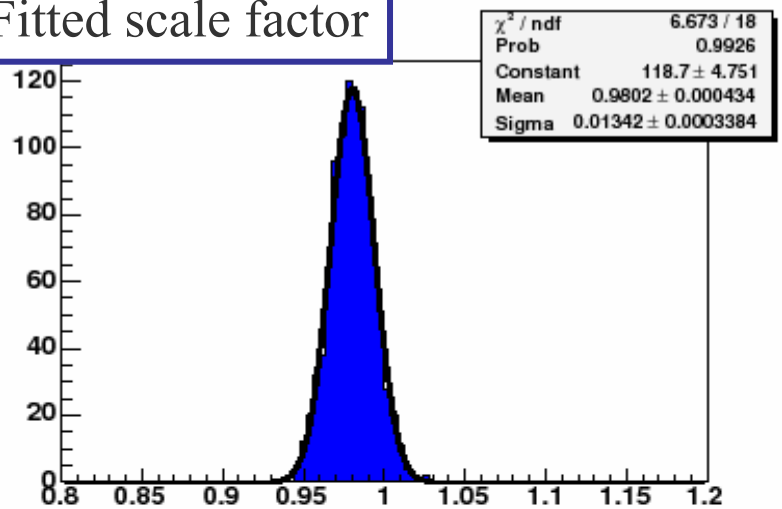
To check possible biases in the unbinned likelihood we perform pseudo-experiments.

- **N evts** generated/pseudoexp = 100,000
- vary **input signal fraction** from 0.5% to 3.0%
- vary **input b-JES scale factor** from 0.95 to 1.05
- perform **1000 pseudo-experiments** for each different selection.

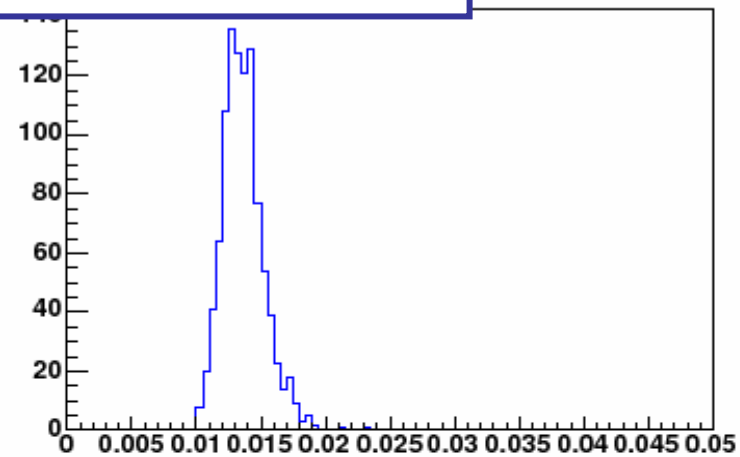
Pseudo-exp example: result on SF

Input signal fraction: 3%, input b-jet SF: 0.98

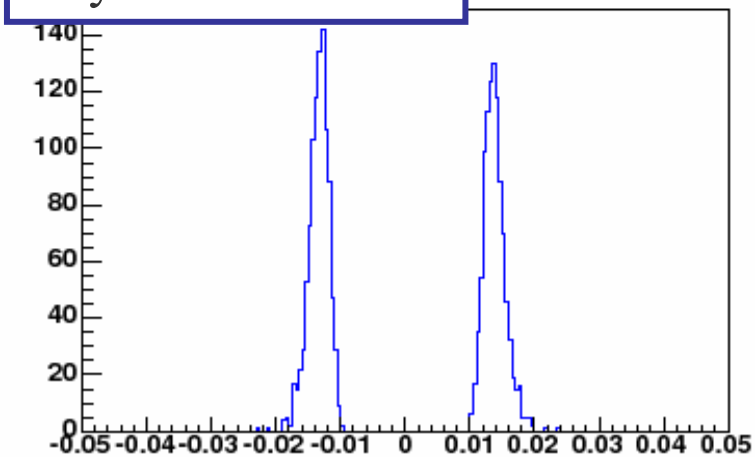
Fitted scale factor



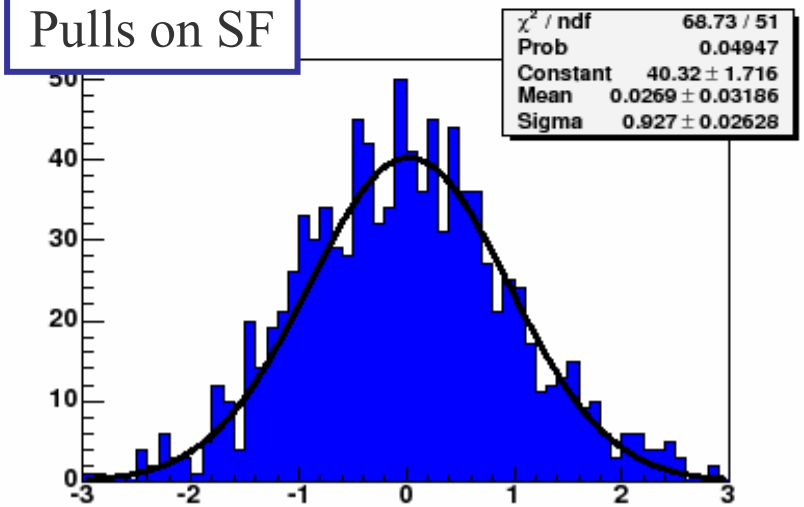
Parabolic error on SF



Asymmetric errors

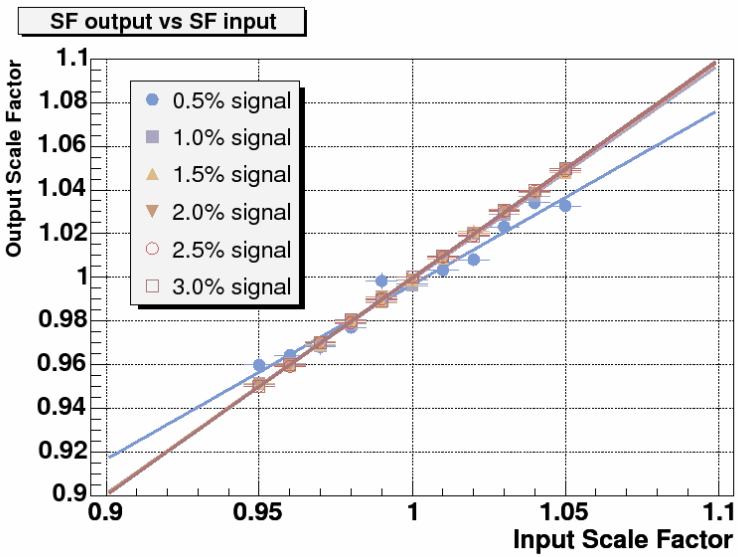


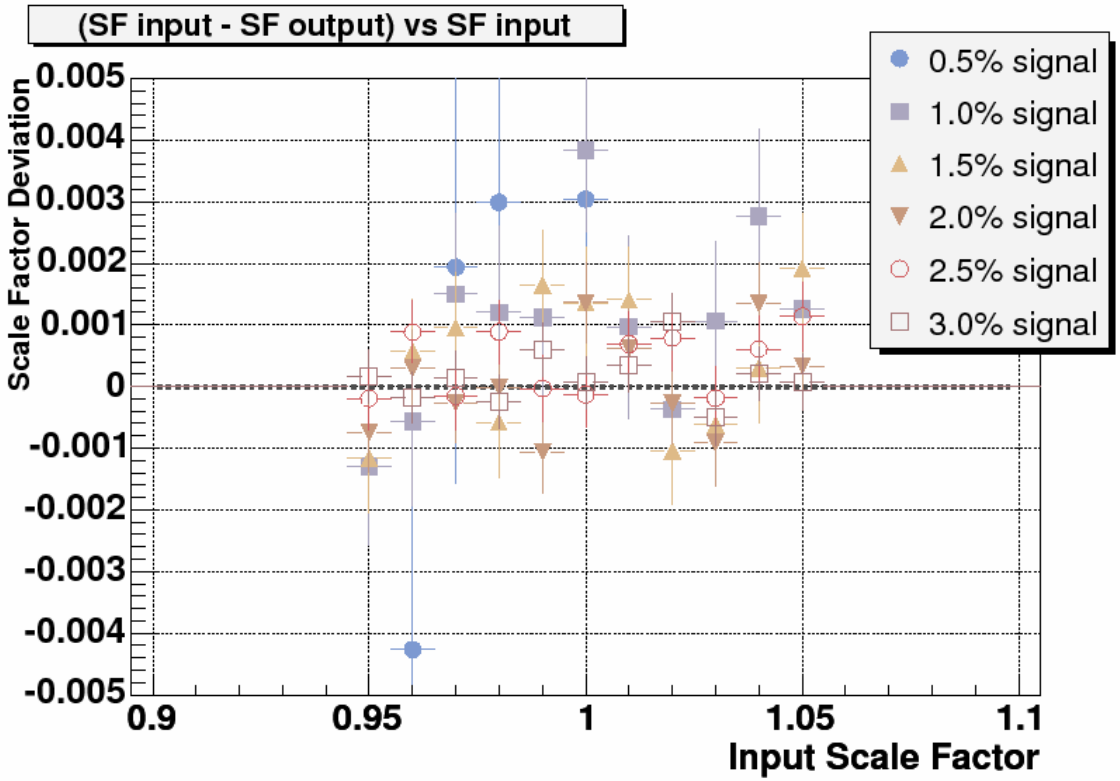
Pulls on SF



Linearity studies

SF output vs SF input





Pseudo-experiment **behave well** down to signal fraction 1.5% (1500 signal events). For lower S/B we begin to observe small biases in SF, and an overestimation of N of fitted signal events.

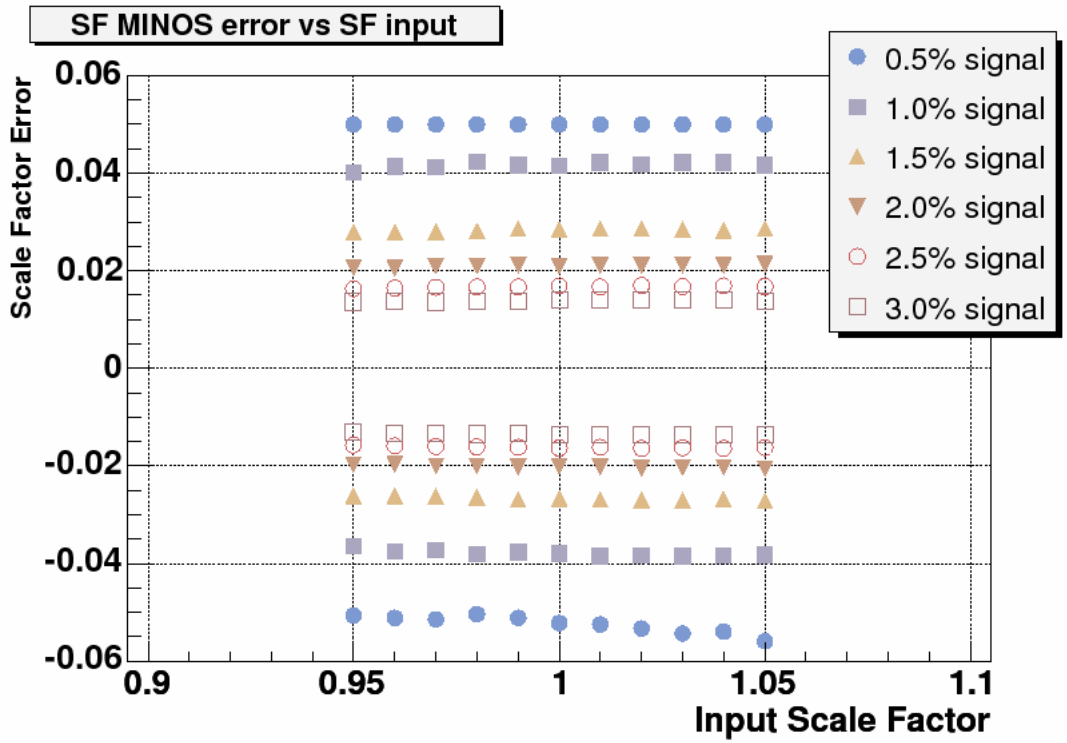
Not much of a concern in our selected sample: N signal expected to be > 2%.

Pseudo-experiments

Expected uncertainty on fitted b-jet energy scale and number of signal events:

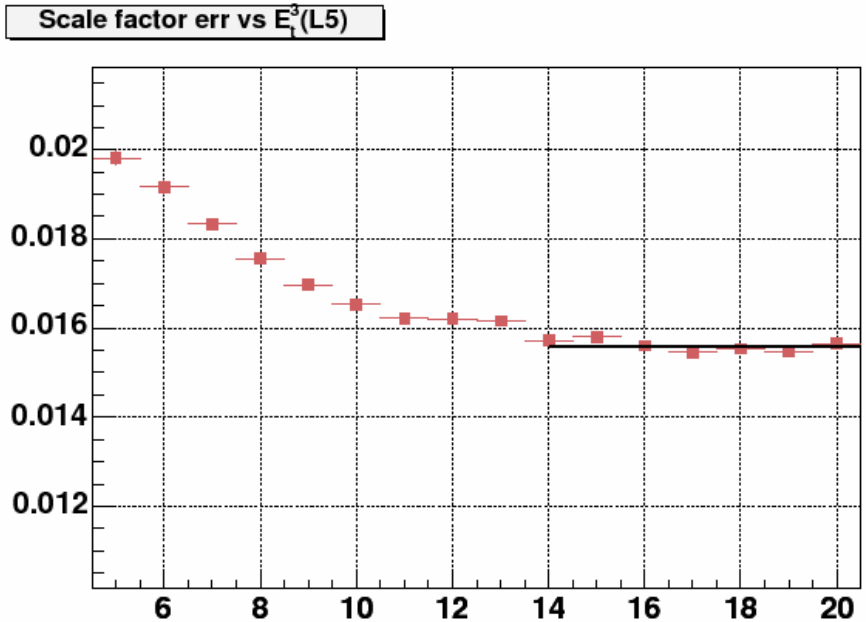
Input signal fraction	b-JES error	N signal error
1.5 %	0.025	340 events
3%	0.014	320 events

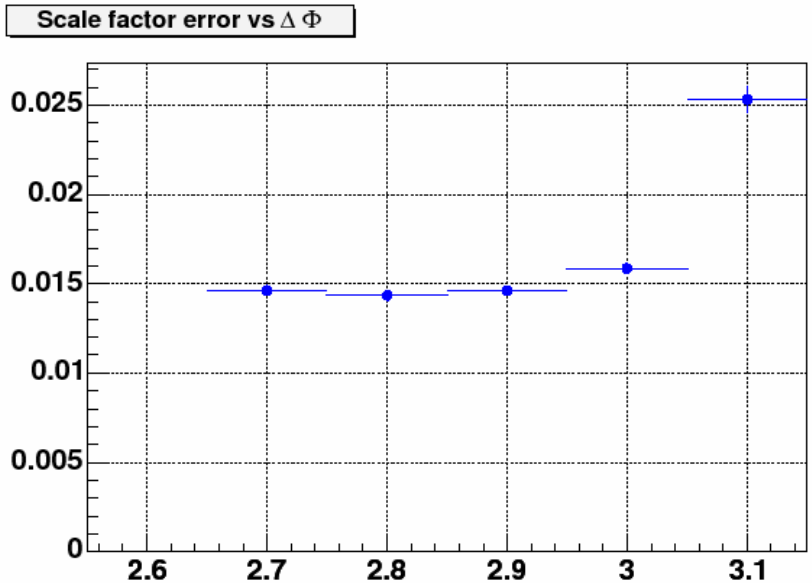
Asymmetric uncertainties on b-JES as a function of input SF



Signal Region Selection

Expected uncertainty on fitted SF as a function of kinematical cuts.





Maximizing significance reduces error on SF.

High cut on $E_t^{3rd\ jet}$ is favored

Rather flat for **E_t>14 GeV**

Not too high cut to **keep statistic in background region**

Select **back to back dijet events**

Flat for ΔΦ<3

Not to low cut to **keep stat in background region.**

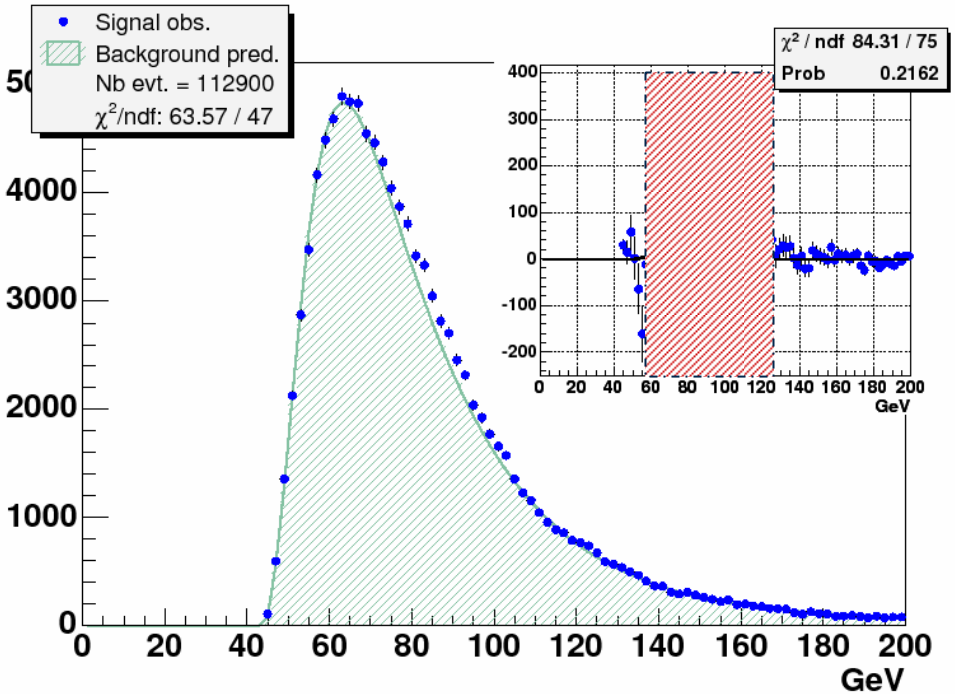
Signal region cuts: **[$\Delta\Phi_{12}$>3, E_t(3rd jet)<15 GeV]** sounds a good compromise.

Background Models Selection

Different BG models can be constructed **varying the [$\Delta\Phi$,E_t(3rd jet)] cuts** on the BG region.

A priory not obvious to favor one BG model respect to another: pseudo-exp. do not really give any indications, study on sample composition does not help much either.

Simple choice: select BG models that fit well **sidebands**, difference between the models will end up in **systematic error.**



Background fit to the dijet mass sidebands.

BG Shape Systematics

Systematic related to limited statistic for BG shape modeling:

Smear BG shape within its statistics and perform 500 **pseudo-experiments**, constructing as usual the 'data' template from signal PDF and unsmeared BG but fitting it with the smeared BG shape.

This procedure is repeated 100 times (BG will be smeared 100 times in total) and for each pseudo-experiment we plot the distribution of **mean number of fitted signal and mean fitted scale factor.**

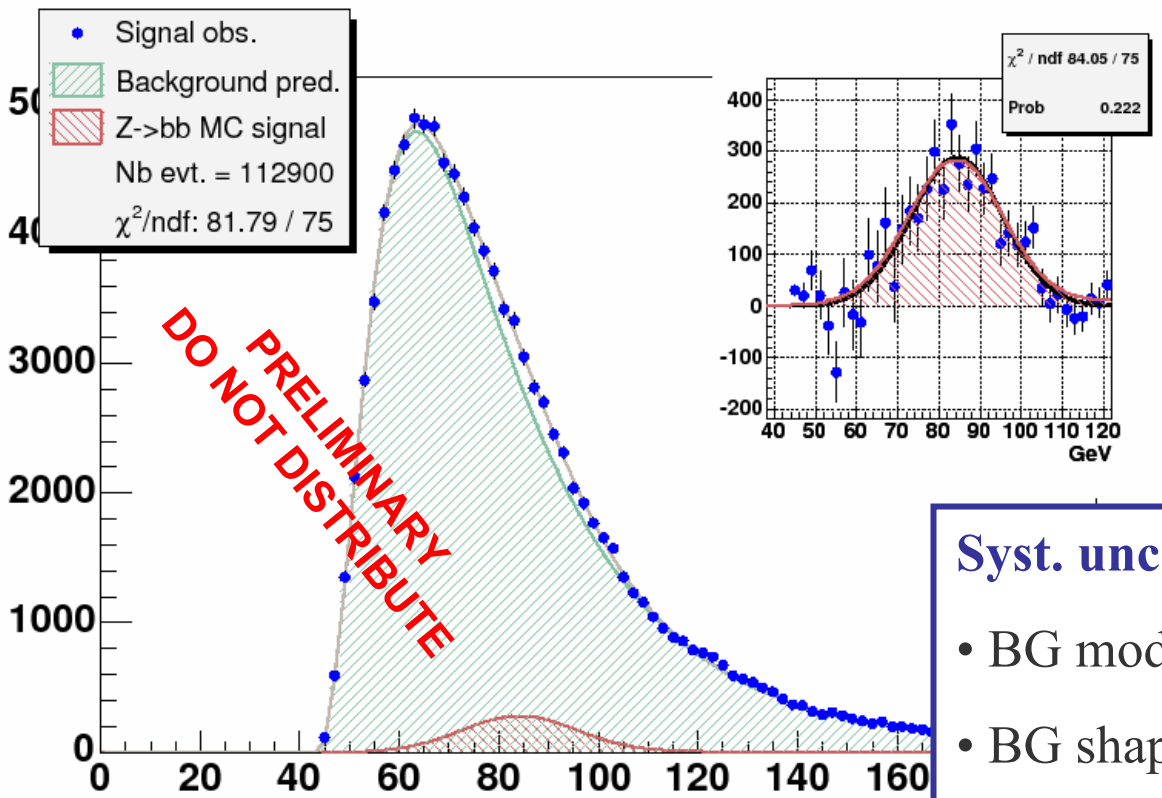
The RMS of these distributions give the BG systematic uncertainty on N signal and SF.

Estimated systematic uncertainty related to BG modeling:

b-JES	N signal
~ 0.010	~ 310 events

Fit Results on Data

Results of unbinned likelihood fit on data:



Statistical uncertainties

- N signal ~ 360 events
- Scale Factor ~ 0.011

In agreement with pseudo-exp predictions.

Syst. uncertainties on b-JES

- BG model selection ~ 0.012
- BG shape statistics ~ 0.010
- other systematics: JES, MC (PDF, ISR/FSR, generator): currently evaluating them, expected to be <<0.010

Expect a total error (stat. + syst.) on b-JES ~ 2%

Conclusions

We have described a method to measure both the number of Z→bb events and the b-specific jet energy scale in data.

All tools are ready to make final measurement

Analysis actually under process of approval by collaboration, public results in about 1 month !

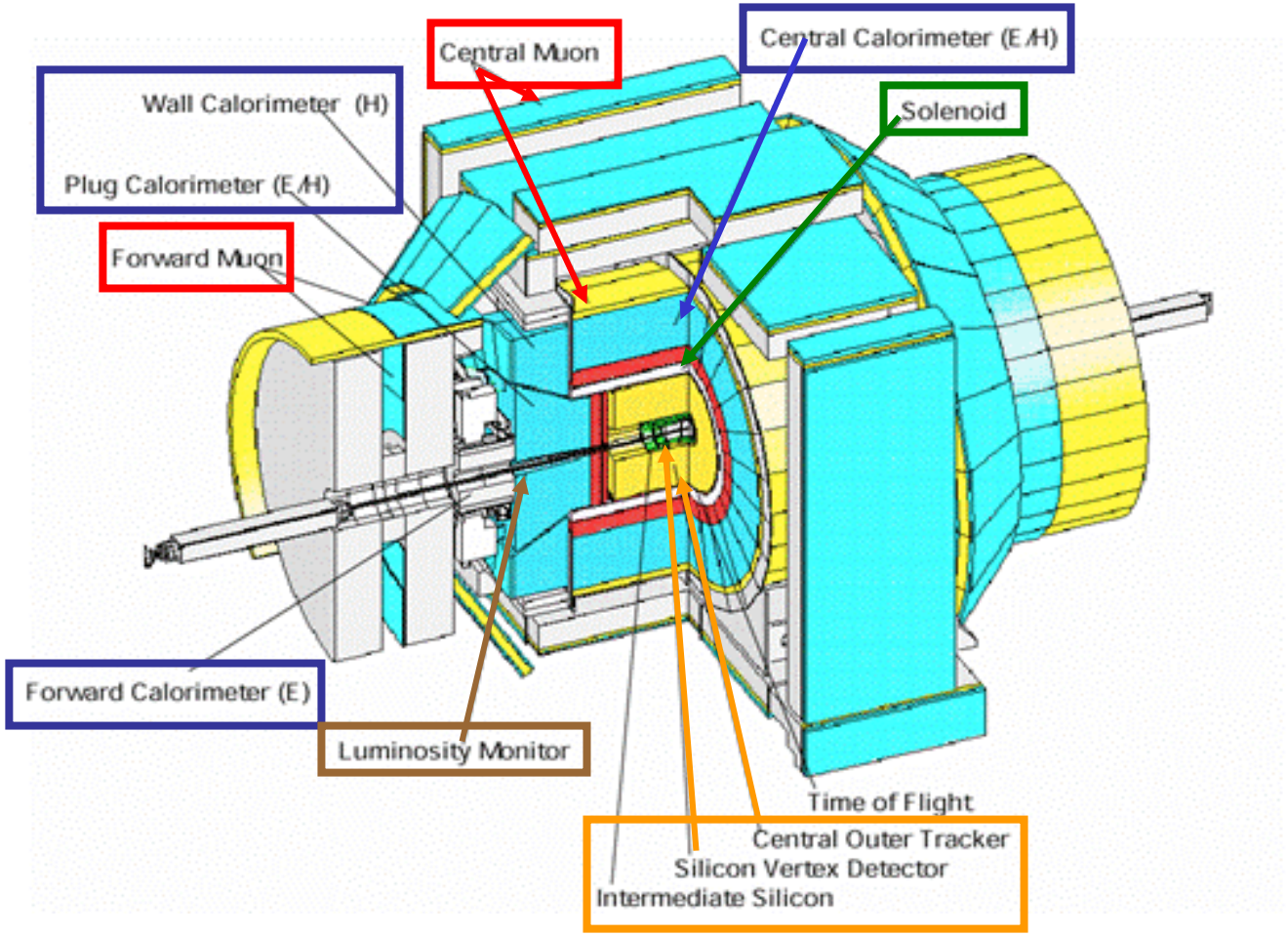
Short term future applications:

b-JES: top mass in dilepton channel

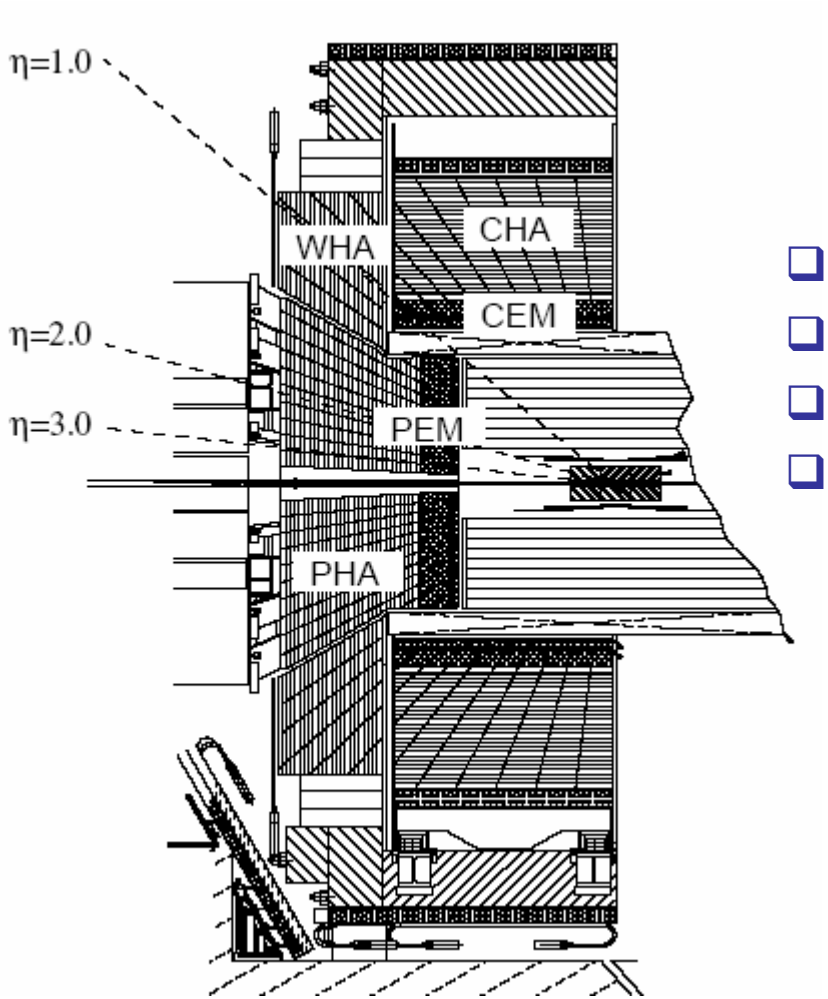
Z signal: test *on data* of algorithm to improve b-jet energy resolution.

BCKUP

CDF Detector



CDF Calorimeters



Calorimeters |η|<3.6

- EM: Lead scintillator sampling
- HA: Steel/iron scintillator sampling
- photomultipliers (one or two per tower)
- readout electronics upgraded for Run II

❑ Non-linear response to hadrons (non-compensating)

❑ Different response to electrons and hadrons

❑ Coarse granularity ($\Delta\eta \times \Delta\Phi \sim 0.1 \times 0.26$ in central)

❑ Low noise

Electrons :

$$\frac{\sigma_{E_T}}{E_T} = \frac{13.5\%}{\sqrt{E_T}} \oplus 1.5\% \quad \text{(CEM)}$$

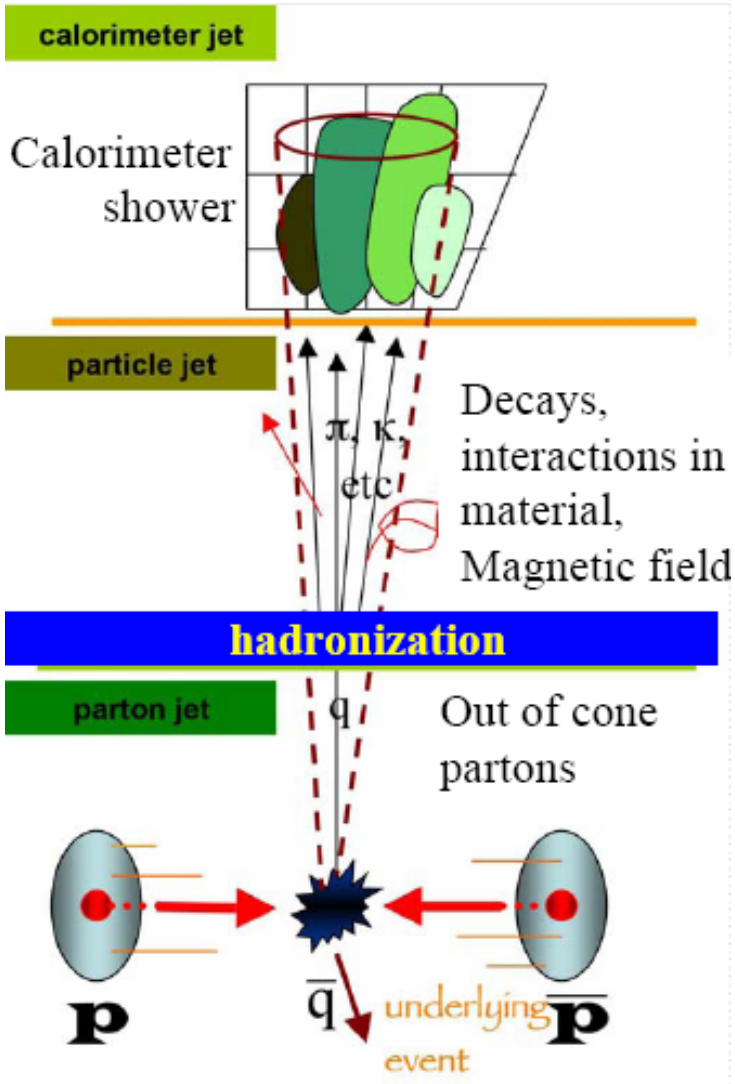
$$\frac{\sigma_E}{E} = \frac{16\%}{\sqrt{E}} \oplus 1\% \quad \text{(PEM)}$$

Pions :

$$\frac{\sigma_E}{E} \approx \frac{80\%}{\sqrt{E}}$$

Jets at Hadron Collider

Jets are complex objects measured by calorimeters and defined by algorithms



Detector properties:

- non-linearity energy response
- un-instrumented regions

Complex underlying event physics:

- spectator events
- gluon radiation (ISR and FSR)
- multiples ppbar interactions
- different jets types (light and heavy flavour, gluons, taus)

Reconstruction algorithms:

- out of cone energy

Need to correct for detector, algorithm and physics effects to obtain the true energy of the jets: **Jet energy scale (JES)**

Jet Energy Corrections

$$P_T(R) = \left[P_T^{raw}(R) \times f_{rel} - UEM(R)\right] \times f_{abs}(R) - UE(R) + OOC(R)$$

Jet corrections in CDF are performed at several levels:

- ***(f_{rel})* Relative Corrections: di-jet balance in data and MC**
  - Makes calorimeter response uniform in η
- ***(UEM)* Multiples Interactions: minimum bias data**
  - Energy from different ppbar interaction increases jet energy
- ***(f_{abs})* Absolute Corrections (calorimeter → particle): di-jet MC**
  - Calorimeter is non-linear and non-compensating
- ***(UE)* Underlying Events: min-bias data**
  - Subtract energy from spectator particles (ISR, beam-beam remnant)
- ***(OOC)* Out-of-Cone (particle → parton): di-jet MC**
  - Correct for particle losses outside the jet cone (FSR, hadronization)

Systematic uncertainties: estimated at each step by comparing data and MC

Validation: photon+jet and Z+jet samples are used for validation of Jet Corrections