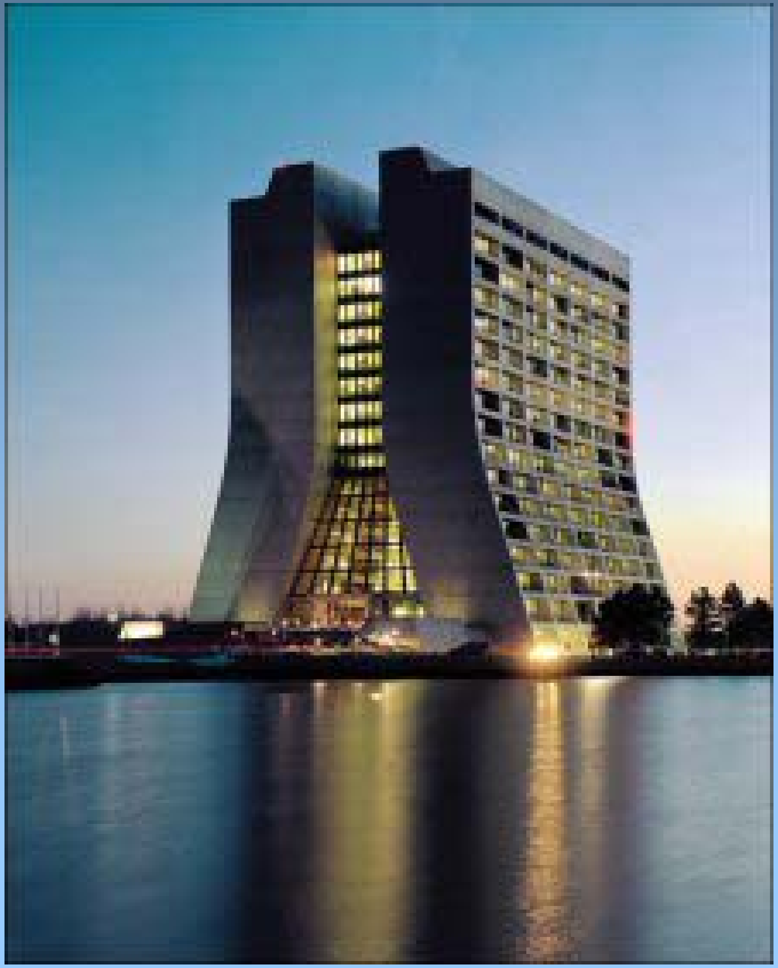
Extraction of a $Z \rightarrow b\bar{b}$ Signal in CDF Run II Data



Julien Donini

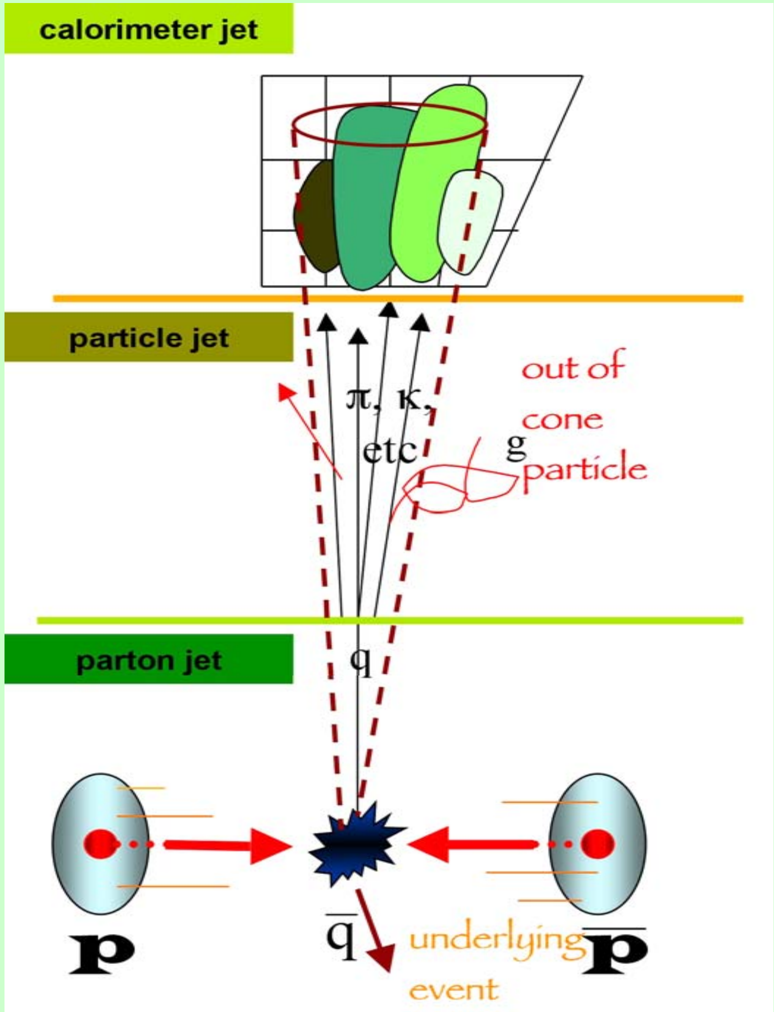
University of Padova and INFN

- ❖ Jet energy scale issues
- ❖ Understanding our dataset
- ❖ Z → bb signal extraction
- ❖ What's next

Tev4LHC at CERN 29.04.05

Jets at Tevatron

Jets are complex objects measured by a calorimeter 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 Scale in CDF II

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:

M_{top} = 173.2 ± 2.9 (stat.) ± 3.4 (syst.) GeV/c^2
Systematics on JES = 3.1 GeV ! *(CDF II, 318 pb^{-1} J.F Arguin)*

Jet corrections in CDF are performed at several levels:

- ☐ ***(f_{rel})* Relative Corrections**
  - Make response uniform in η
- ☐ ***(UEM)* Multiples Interactions**
  - Energy from different ppbar interaction increases jet energy
- ☐ ***(f_{abs})* Absolute Corrections (calorimeter → particle)**
  - Calorimeter is non-linear and non-compensating
- ☐ ***(UE)* Underlying Events**
  - Energy associated with the spectator partons in a hard collision
- ☐ ***(OOC)* Out-of-Cone (particle → parton)**
  - Particle level to parton level

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

Motivation for Z→ bb Studies

Determination of the b-jet energy scale

- ☐ The current jet energy corrections are **generic**. However b-jets have different properties (fragmentation, decay, mass) and need a specific treatment.
- ☐ The results on many **physical processes** depend on the resolution on b-jet energy:

Reconstructing and analysing the Z → bb resonance is still the best method to determine b specific jet corrections.

Search for low mass Higgs boson

- ☐ The successful extraction of a bbbar resonance from the large QCD background provides precious knowledge and tools useful for **low mass Higgs searches** (H →bb).
- ☐ In particular the **understanding and modelisation of the QCD** background is of crucial importance.
- ☐ Once we achieve the extraction of the signal we can test **improved algorithms** that increase the **mass resolution** of a dijet decay.

Demonstrating that we are able to obtain a 10% resolution on bb resonance would have a big impact on Tevatron chances for light Higgs discovery !

Z→bb Specific Trigger

SVT based Trigger

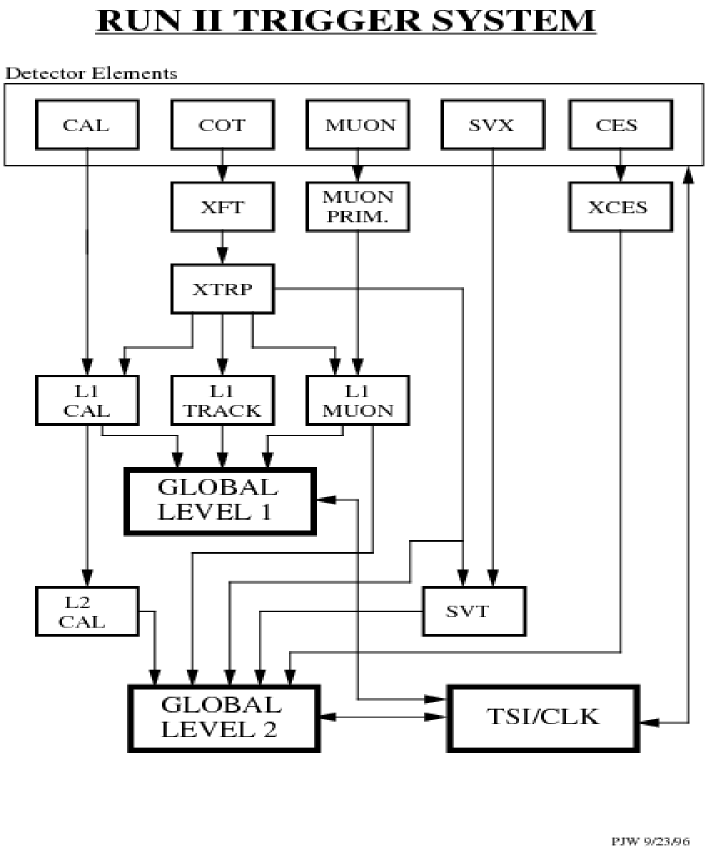
☐ Information given by the internal **trackers** (SVX and COT) are used at the **Level 2** of CDF trigger system to select events with **high impact parameter (d_0) tracks.**

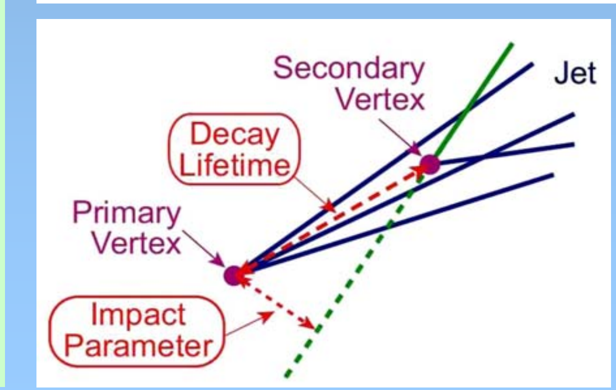
☐ This relies on the **SVT hardware** device which is able to measure P_t and impact parameter (to within 50 μm) of **charged tracks** in less than 20 μs. SVT has proven crucial for most of CDF II's B physics program

Z → bb trigger selects events with

- **Two displaced tracks:**
  - $d_0 > 160$ μm, $P_t > 2$ GeV
- **Two $E_t > 10$ GeV jets**

Efficiency on Z → bb is 4-5 %, but better than lepton trigger (<1%) which are biasing the jet E_t measurement.





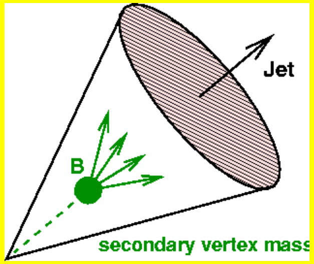
Tagging b-jets

Tagging b-jets: three methods are well-tested and used:

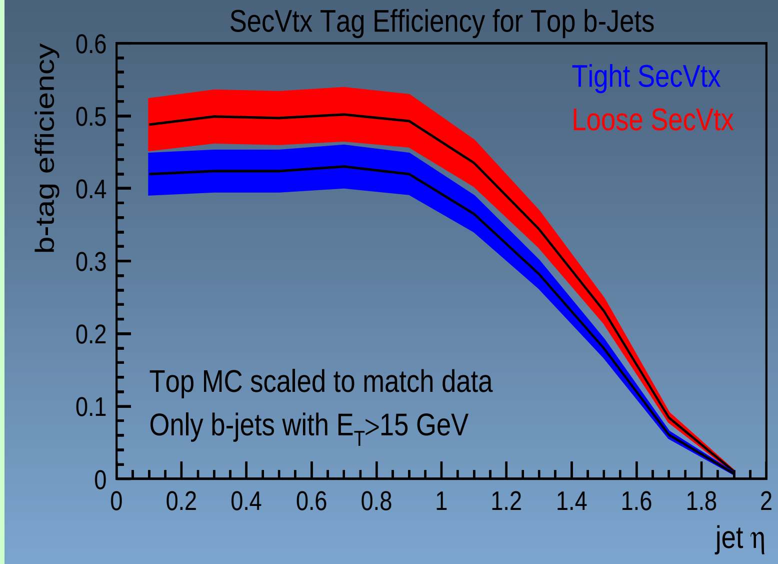
- ☐ Soft lepton tagging
- ☐ **Secondary vertex tagging**
- ☐ Jet Probability tagging

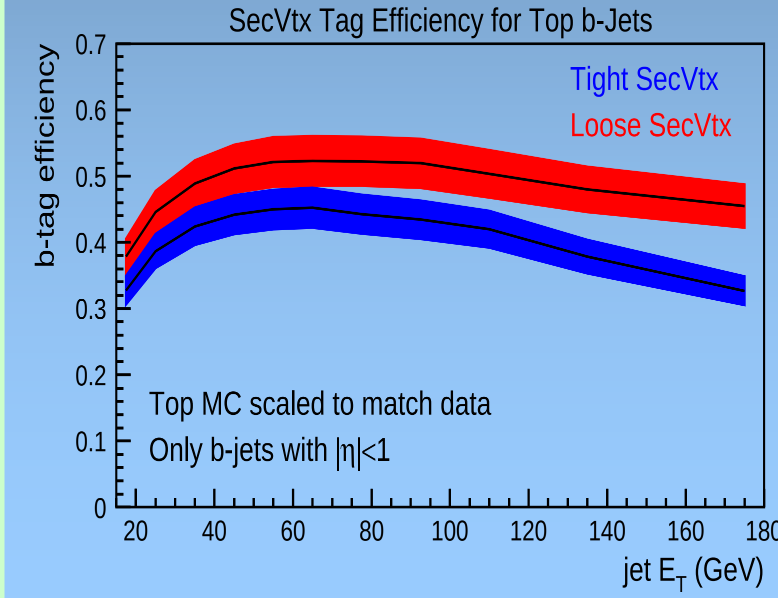
SecVtx tagging

- 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%**





Sample Composition Studies

☐ Once data is collected one has an **enhanced fraction of bbbar events**, but also an important part of **light quark and gluons** in the sample.

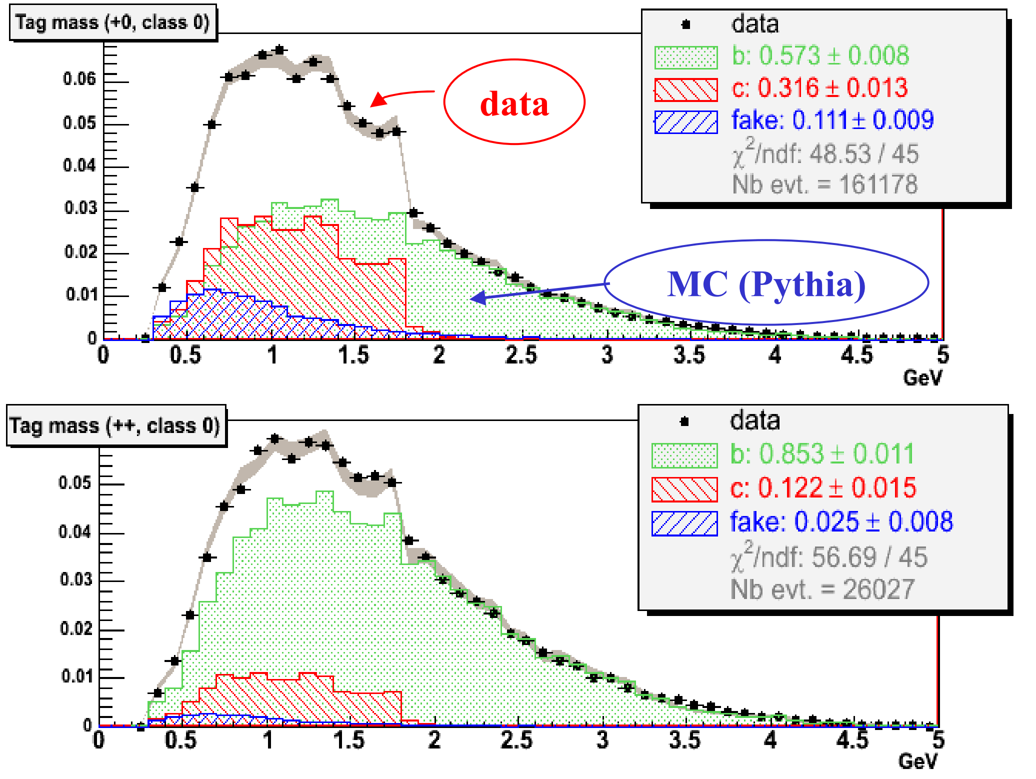
☐ **Sample composition**

- events effectively containing a **secondary vertex** (quarks b/c),
- events with **bad reconstructed tracks** (light quarks, gluons) which pass the SVT requirement.

Fits to the **invariant mass of the reconstructed secondary vertices** allow to estimate the b/c/g fraction in tagged jets.

Fraction of b jets in tagged **di-jet** central events (E_t > 10 GeV, $|\eta|<1$):

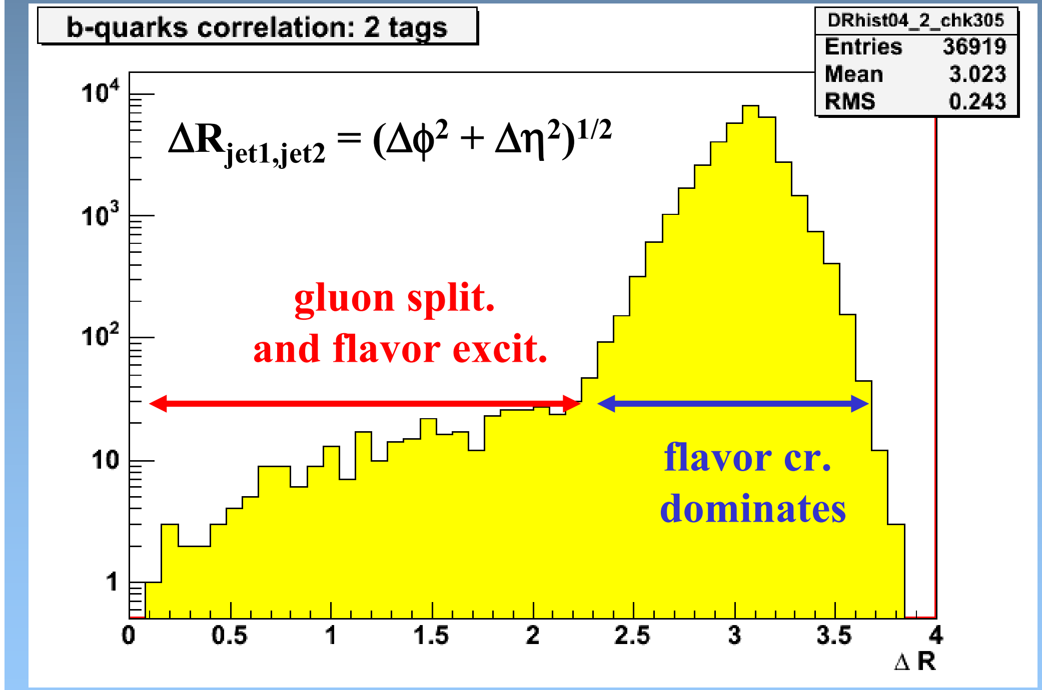
- **One tag (+0) : $F_b = 0.57 \pm 0.01$**
- **Two tags (++) : $F_b = 0.85 \pm 0.01$**

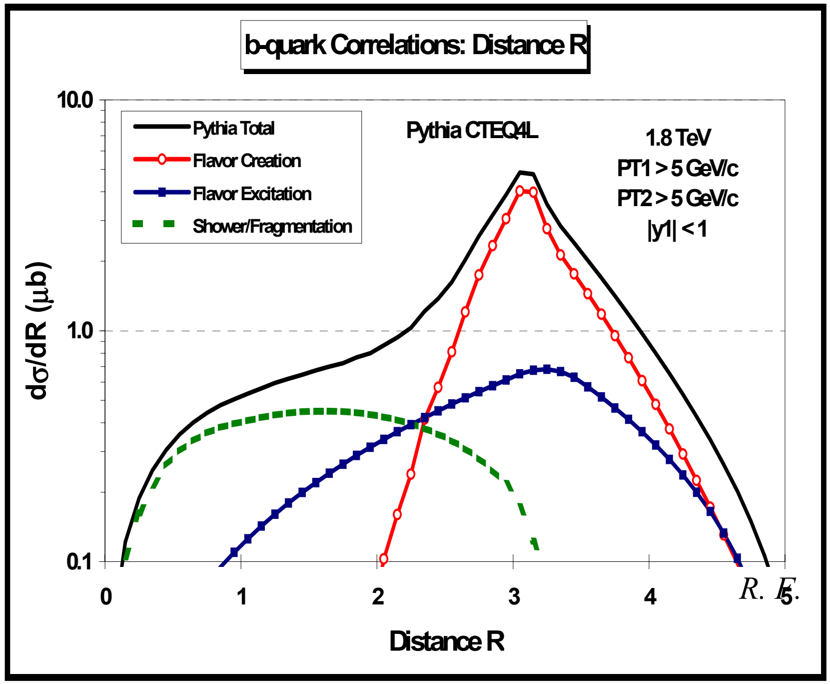


b-quark Correlations

We have a strong component of bb in our (++) di-jet events, but **what is the dominant process of b-quark production ?**

At Tevatron there are, at ‘leading-log’, three main sources of b-quarks: flavor creation, flavor excitation and gluon-splitting. The ΔR distribution of the two main jets can give us a hint on the relative weight of each processes in our data *(R. Field)*.





At first order dominant process appears to be flavor creation. The fraction of gluon splitting in our data is expected to be around 3-5 %.

Total Sample Composition

To estimate the total fraction of bb (f_{bb}), gluon-splitting/flavor excitation events (f_{bg}) and light flavor (f_{gg}) in our di-jet dataset we apply the following method:

1. Divide the data in different classes depending of the **number of muons** present in the two main jets (muon jets are often enriched in heavy flavors content):

   ⇒ **4 muon classes: (0μ, 0μ), (1 μ, 0 μ), (>1 μ, 0 μ), (≥1 μ, ≥1 μ)**
2. Count, for each class, the **number of events w/ 0, 1 or 2 positive or negative (i.e fake) tags:**

   ⇒ N_{00}, N_{+0}, N_{++}, N_{-0}, N_{+-}, N_{--}
3. Perform a **secondary vertex mass fit** for all muon classes with one or two tags:

   ⇒ $F_b(+0)$, $F_b(++)$

Simple equations relates the number of observed tags and the vertex mass fit results to the sample composition fraction and the tagging efficiency.

4. Thus, from a χ^2 fit on each muon classes we are able to extract the **fractions f_{bb}, f_{bg}, f_{gg}** and also to estimate the **b-tag efficiency.**

$$\chi^2 = \Sigma_{i=1}^{24} \frac{(N_i - N_i^{obs})^2}{\sigma^2_{N_i^{obs}}} + \Sigma_{i=1}^{4} \frac{(F_b^{(+0),i} - F_b^{(+0)meas,i})^2}{\sigma^2_{F_b^{(+0)meas,i}}} + \Sigma_{i=1}^{4} \frac{(F_b^{(++),i} - F_b^{(++)meas,i})^2}{\sigma^2_{F_b^{(++)meas,i}}}$$

Preliminary Fit Results

We tried different di-jet event selection on (most of) the available dataset

☐ **Two main jets with $E_t > 10$ GeV ($|\eta|<1$):**

N_{tot}: 2.6 10^6 events selected

⇒ fraction of ***bb*** events: $\mathbf{f_{bb} = 0.11 \pm 0.01}$

fraction of ***bg*** events: $\mathbf{f_{bg} = 0.05 \pm 0.01}$

☐ **Two $E_t > 20$ GeV jets with $\Delta\Phi > 2.5$, and no other jet with $E_t > 20$ GeV**

with this cleaner dijet-selection, the fraction of ***bg*** events is reduced:

N_{tot}: 6.2 10^5 events

⇒ $\mathbf{f_{bb} = 0.12 \pm 0.01}$

$\mathbf{f_{bg} = 0.02 \pm 0.01}$

b-tagging efficiency estimation for this dataset: $\mathbf{\varepsilon_b^+ = 0.45 \pm 0.02}$

These fits results are preliminary and must be taken with caution, but they seem to indicate that the component of **gluon-splitting/flavor excitation is small** in our data, in particular in clean back to back jets.

Z Signal: Data Selection

Total dataset: **21.5 M events**, corresponding to **333 pb^{-1}** of run goods for calorimeter and silicon tracker.

The signal extraction methods used are based on the assumption that **$Z \to bb$ process irradiates less energy**, outside from the main two leading jets, than b-quarks and generic QCD di-jet events.

Taking this into account we divide our dataset in two regions:

Signal zone (SZ)

back to back di-jet events with little extra-jet activity

- two main jets: $E_t > 20$ GeV, $\Delta\phi_{12} > 3.0$, $|\eta|<1.5$
- extra jets: $E_t < 10$ GeV

Normalization zone (NZ)

Region of background events that we use for normalization

- main two jets: $E_t > 20$ GeV, $\Delta\phi_{12} > 2.5$, $|\eta|<1.5$
- extra jets: $E_t < 20$ GeV
- signal zone excluded
- also exclude intermediate zone: $E_t < 12$ GeV, $\Delta\phi_{12} > 2.5$

Background Model

☐ The events that we use for **Z extraction** are those with **two tagged (++) jets** in the ***signal zone***.

☐ However we use the ***normalization zone*** to estimate the dijet invariant mass shape of the background in the ***signal region***. To do this we:

- extract the **ratio between tagged (++) and untagged (00)** events in the **NZ**, as a function of M_{jj}
- apply this tag rate calculation to **correct the dijet mass distribution of untagged events in the *signal region*.**

☐ This method is old and well tested. It basically assumes that the bias to the mass shape due to SecVtX tagging is uncorrelated with the bias to the mass shape due to the kinematic cuts.

Background Model

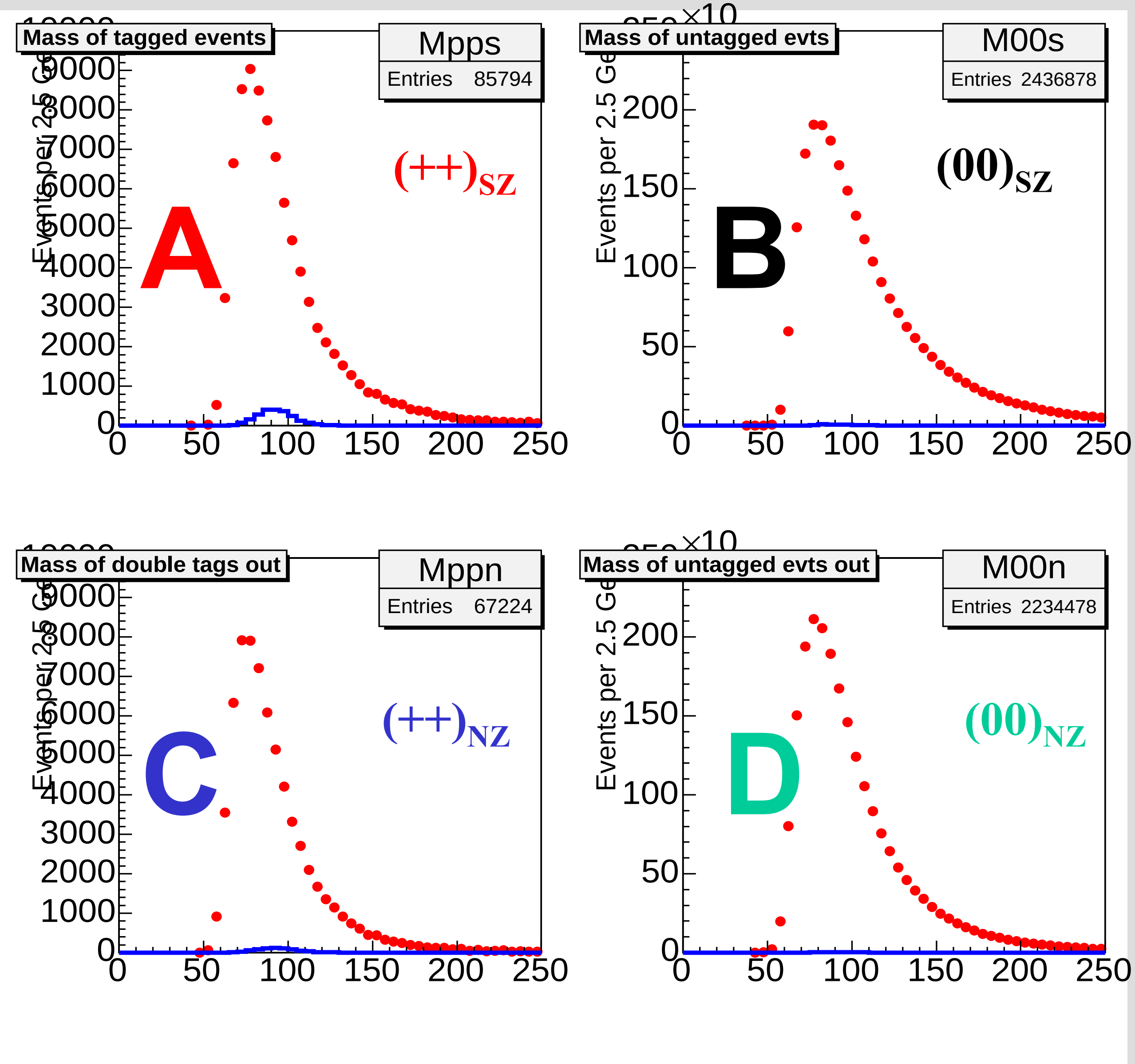
Top left: double tags In signal region

Top right: untagged events in signal region

Bottom left: double tags in background region

Bottom right:untagged events in background region.

Background in **A** is computed as **B*C/D**

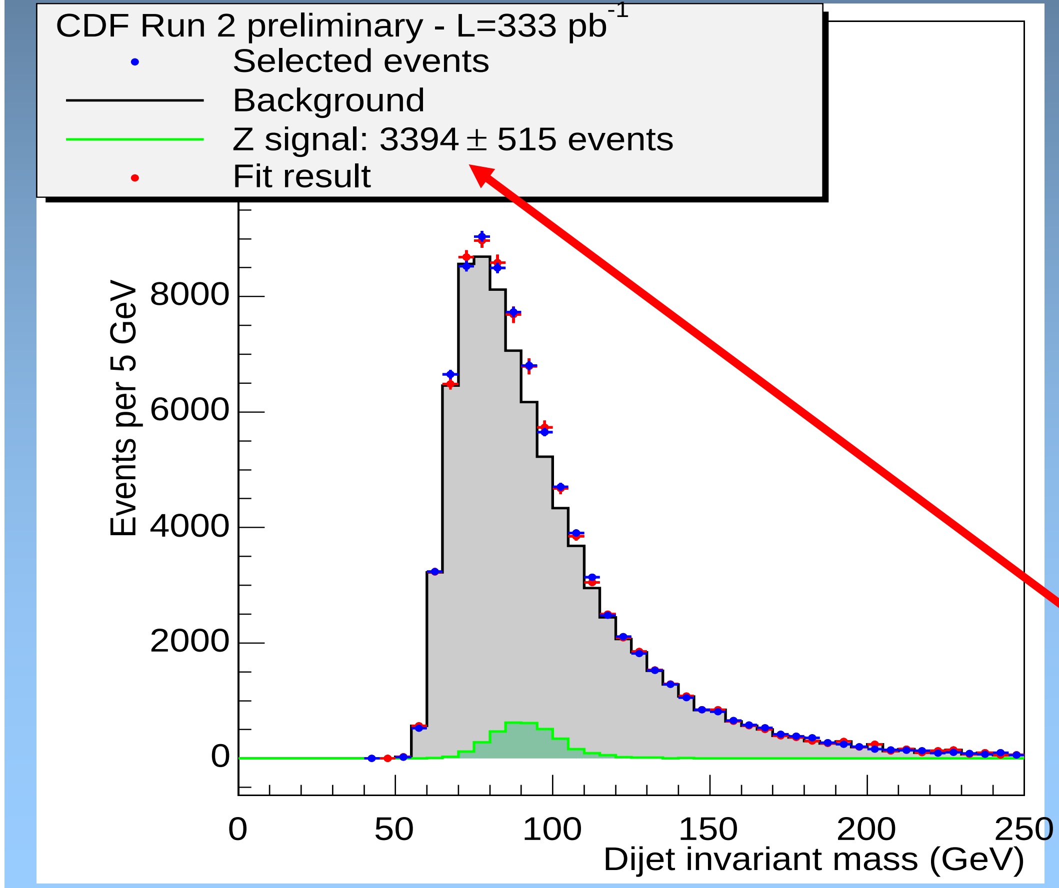


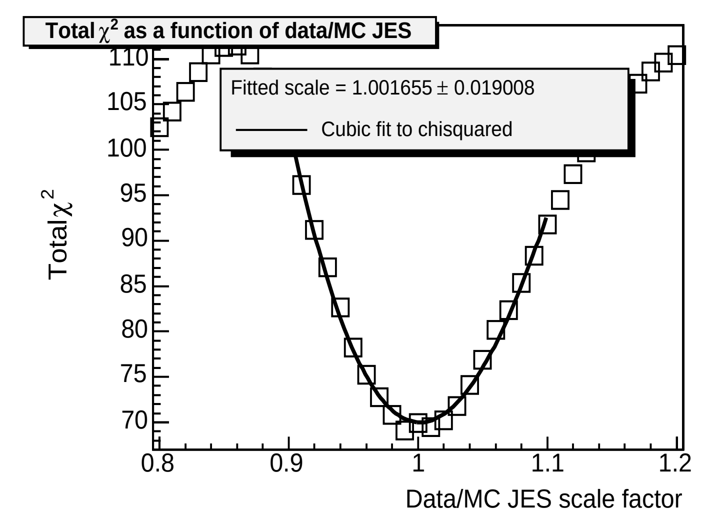
Fit and Data/MC Jet Energy Scale Factor

☐ We create **$Z \to bb$ signal templates with varying data/MC JES factors**

☐ We can then **fit the tagged data** to the sum of **background and signal templates**, for varying JES.

☐ The fit converges nicely and gives the JES and the **number of reconstructed Z's**





Among 85,720 events selected ($L=333pb^{-1}$) CDF finds

3400 ± 500 év. $Z \to bb$ decays

Test of the Model

- ☐ To verify the correctness of our fitting procedure, we perform **pseudoexperiments.**
  - ⇒ Background template and signal templates are fluctuated within Poisson statistics, then the sum is fit and the best value of the data/MC jet energy scale factor is found.

This allows us to verify the **absence of biases** and the **fit stability vs k**.

- ☐ Same goes for the **checks of the background model** and for the **signal efficiency studies**. Work here is still preliminary but show good results.

- ☐ **Solid results and validation of the analysis are expected for this summer**

Conclusion

We showed that a significant Z signal could be extracted from CDF Run II data … and next ?

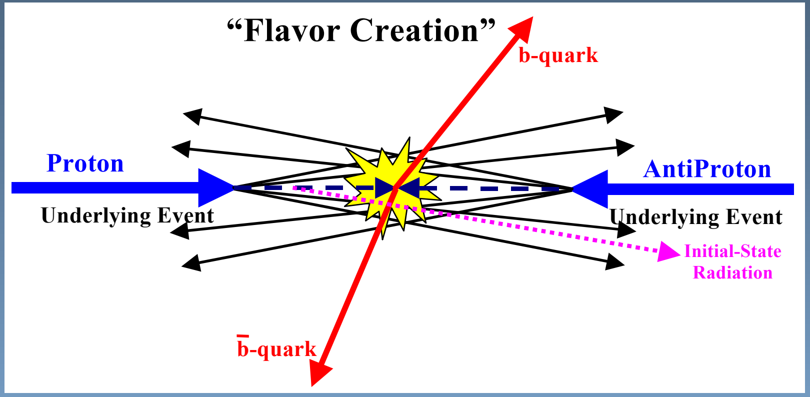
- **Improve the background shape model** to extract the best possible Z → bb signal.
- With enough statistics we will then be able to **constraint the b-jet energy scale**. With a 10 000 Z signal we should be able to determine the b-jet scale to within 1%.
- Such a signal will also allow us to **perform detailed studies of resolution optimization algorithms**
  - for instance jet resolution algorithms, used for the h → bb, that **combine tracks and calorimeter towers** *(H1 algorithm)* will be studied on the Z bbar dataset.

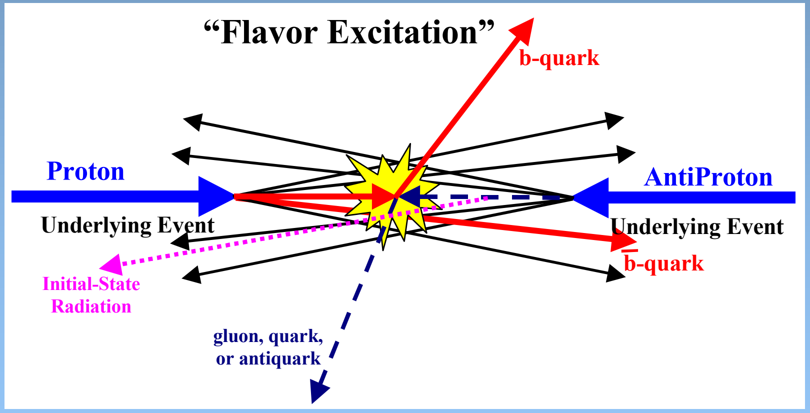
... More on this at the next TeV4LHC !

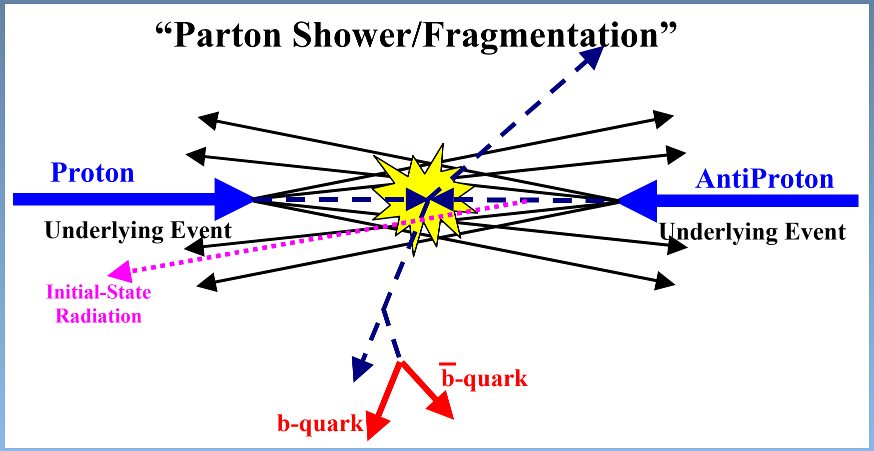
Bckup Slides

B-quark production at Tevatron

Rick Field

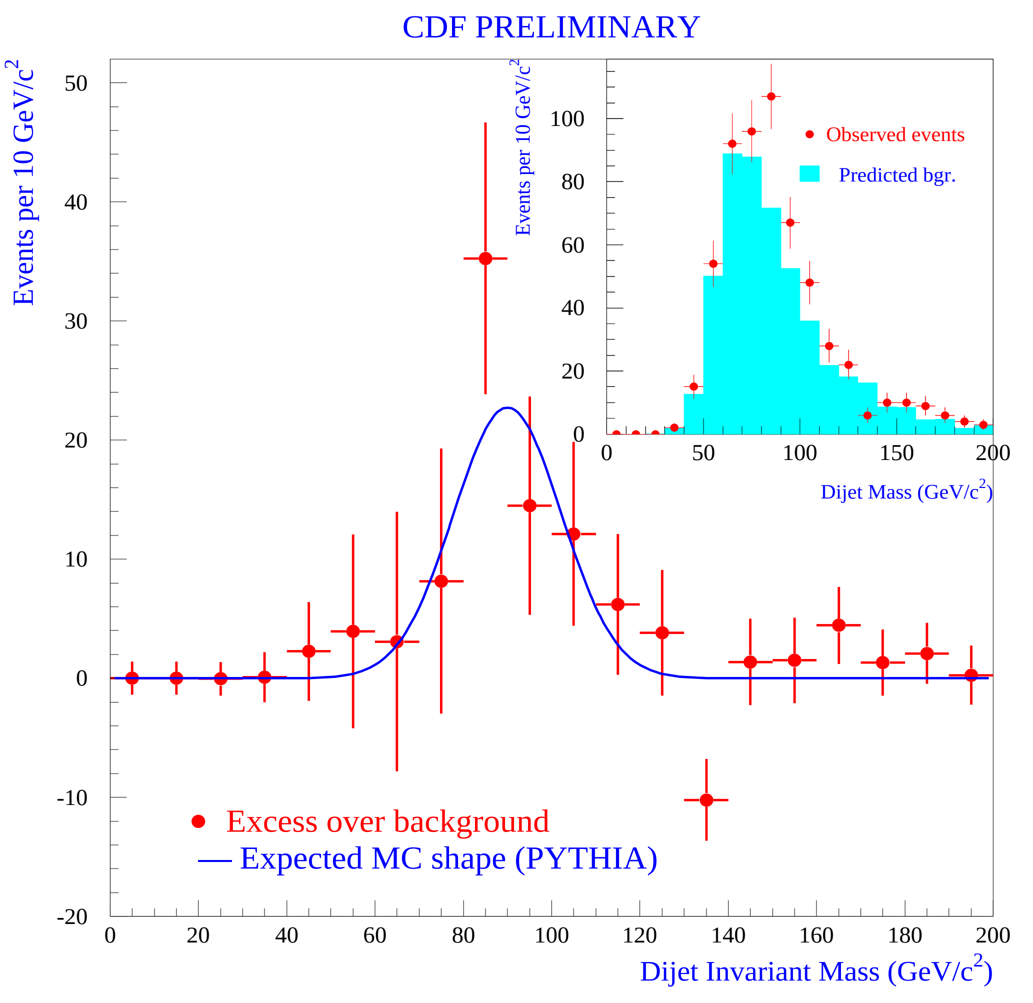






All three sources are important at the Tevatron!

Z → bb in Run I


CDF PRELIMINARY
Events per 10 GeV/c^2
50
40
30
20
10
0
-10
-20
0
20
40
60
80
100
120
140
160
180
200
Dijet Invariant Mass (GeV/c^2)
● Excess over background
— Expected MC shape (PYTHIA)
Events per 10 GeV/c^2
100
80
60
40
20
0
0
50
100
150
200
Dijet Mass (GeV/c^2)
● Observed events
Predicted bgr.

The Z→ bb trigger at CDF II

- The CDF trigger system has 3 levels. The Z→bb trigger exploits most of its functionalities.

- **At L1, dijet events with charged tracks are collected by requiring 1 5-GeV calorimeter tower, plus two 2 GeV charged tracks (thanks to the XFT, an eXtremely Fast Tracker).**
- **At L2, the SVT is used to ask for two tracks with IP>160 um and two energy clusters with Et>5 GeV.**
- **At L3, a full speed-optimized reconstruction is done. Events with two Et>10 GeV jets containing hints of lifetime are selected.**

- The cross section (70 nb @L2) is largish for a calibration trigger. We are constantly fighting with rate increase with L…

Sample Composition Fit

- $N_{00} =$
  $N_{bb}(1-\epsilon_b^+-\epsilon_b^-)^2 + N_{bg}(1-\epsilon_b^+-\epsilon_b^-)(1-\epsilon_g^+-\epsilon_g^-) + N_{cc}(1-\epsilon_c^+-\epsilon_c^-)^2 + N_{cg}(1-\epsilon_c^+-\epsilon_c^-)(1-\epsilon_g^+-\epsilon_g^-) + N_{gg}(1-\epsilon_g^+-\epsilon_g^-)^2;$
- $N_{+0} =$
  $2N_{bb}(1-\epsilon_b^+-\epsilon_b^-)\epsilon_b^+ + N_{bg}(\epsilon_b^+(1-\epsilon_q^+-\epsilon_q^-)+\epsilon_q^+(1-\epsilon_b^+-\epsilon_b^-)) + 2N_{cc}(1-\epsilon_c^+-\epsilon_c^-)\epsilon_c^+ + N_{cg}(\epsilon_c^+(1-\epsilon_q^+-\epsilon_q^-)+\epsilon_q^+(1-\epsilon_c^+-\epsilon_c^-)) + 2N_{gg}(1-\epsilon_g^+-\epsilon_g^-)\epsilon_g^+;$
- $N_{++} =$
  $N_{bb}(\epsilon_b^+)^2 + N_{bg}\epsilon_b^+\epsilon_g^+ + N_{cc}(\epsilon_c^+)^2 + N_{cg}\epsilon_c^+\epsilon_g^+ + N_{gg}(\epsilon_g^+)^2;$
- $N_{-0} =$
  $2N_{bb}(1-\epsilon_b^+-\epsilon_b^-)\epsilon_b^- + N_{bg}(\epsilon_b^-(1-\epsilon_q^+-\epsilon_q^-)+\epsilon_q^-(1-\epsilon_b^+-\epsilon_b^-)) + 2N_{cc}(1-\epsilon_c^+-\epsilon_c^-)\epsilon_c^- + N_{cg}(\epsilon_c^-(1-\epsilon_q^+-\epsilon_q^-)+\epsilon_q^-(1-\epsilon_c^+-\epsilon_c^-)) + 2N_{gg}(1-\epsilon_g^+-\epsilon_g^-)\epsilon_g^-;$
- $N_{--} =$
  $N_{bb}(\epsilon_b^-)^2 + N_{bg}\epsilon_b^-\epsilon_g^- + N_{cc}(\epsilon_c^-)^2 + N_{cg}\epsilon_c^-\epsilon_g^- + N_{gg}(\epsilon_g^-)^2;$
- $N_{+-} =$
  $2N_{bb}\epsilon_b^+\epsilon_b^- + N_{bg}(\epsilon_b^+\epsilon_g^- + \epsilon_b^-\epsilon_g^+) + 2N_{cc}\epsilon_c^+\epsilon_c^- + N_{cg}(\epsilon_c^+\epsilon_g^- + \epsilon_c^-\epsilon_g^+) + 2N_{gg}\epsilon_g^+\epsilon_g^-;$

$$F_b^{(+0),i} = \frac{N_i^{obs}}{N_i^{(+0)obs}}(2F_{bb}^i\epsilon_b^+(1-\epsilon_b^+-\epsilon_b^-) + F_{bg}^i\epsilon_b^+(1-\epsilon_g^+-\epsilon_g^-)),$$

$$F_b^{(++),i} = \frac{N_i^{obs}}{2N_i^{(++)obs}}(2F_{bb}^i(\epsilon_b^+)^2 + F_{bg}^i\epsilon_b^+\epsilon_g^+).$$

$$\chi^2 = \Sigma_{i=1}^{24}\frac{(N_i-N_i^{obs})^2}{\sigma^2_{N_i^{obs}}} + \Sigma_{i=1}^{4}\frac{(F_b^{(+0),i}-F_b^{(+0)meas,i})^2}{\sigma^2_{F_b^{(+0)meas,i}}} + \Sigma_{i=1}^{4}\frac{(F_b^{(++),i}-F_b^{(++)meas,i})^2}{\sigma^2_{F_b^{(++)meas,i}}}$$