Identifying b-Jets in hadronic collisions

a tool for discoveries

Sandro De Cecco

sandro.dececco@lpnhe.in2p3.fr

Université de Paris VI & VII et LPNHE IN2P3/CNRS
Laboratoire de Physique Nucleaire et des Hautes Energies

LIP seminar

Lisboa, 20-01-2010

Caveat:

- This talk mixes information taken from my personal experience in:
  - Jet b-tagging at Tevatron and in particular in CDF
  - Some consideration on LHC Jet and b-Jet physics with ATLAS view
- I will not stress on physics results in themselves, but rather:
  - Focus on experimental strategies for b-tagging
  - Make main points connected with phenomenolgy
  - Have a transversal approach between experiments and physics channels.

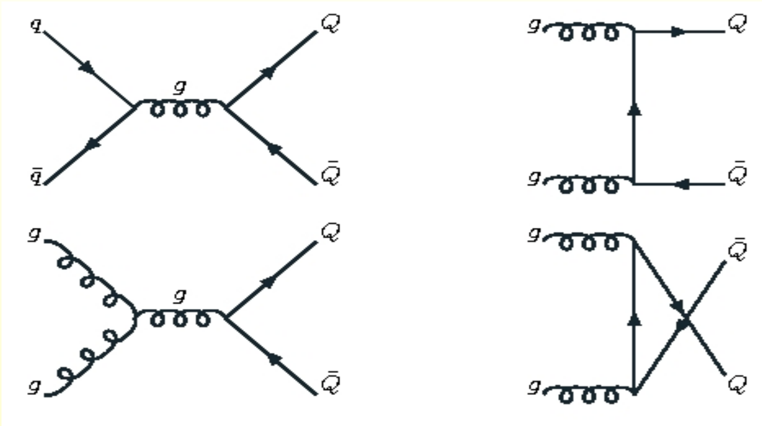
Introduction

- Lots of interesting physics involves high-p_T b-quarks
  - Top production - BR($t \to Wb$) ~ 100%
  - Higgs searches
    - For m_H < 135 GeV/c^2, $H \to b\underline{b}$ is the most common decay
    - SUSY $b\underline{b}g \to b\underline{b}A \to b\underline{b}b\underline{b}$ at high $\tan\beta$
  - Sparticle searches – sbottom, stop decays into b+X
- There are well established ways to b-tag and more exotic ones for future

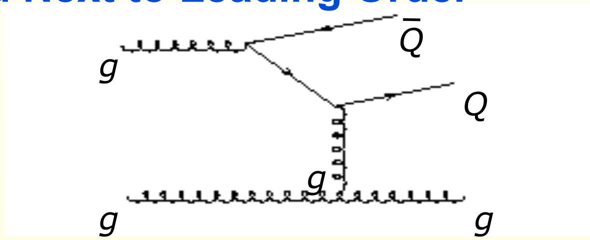
- Outline:
  - b quarks production at collider
  - Characteristics of b-jets
  - B-tagging algorithms
  - Efficiency and fake rate measurements, the CDF experience
  - Recent algorithmic advances
  - Some highlights on the LHC
  - Conclusions

bb pairs production in hadronic collisions

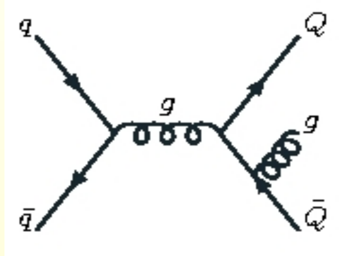
Leading Order and Next to Leading Order



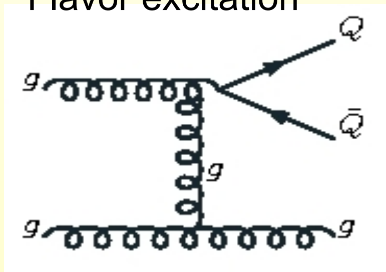
Flavor creation



Flavor excitation



and … radiative corrections



Gluon splitting

… but: b/c-jets or B/D hadrons are the observables rather than b/c-quark

<u>*Observed*</u>

Proton structure

$$\frac{d\sigma\left(p\bar{p} \to B/DX\right)}{d\,p_T\left(B/D\right)} = \frac{d\sigma\left(q\bar{q}/gg/qg \to b/cX\right)}{d\,p_T\left(b/c\right)} \otimes F^{p\bar{p}} \otimes D^{b/c \to B/D}$$

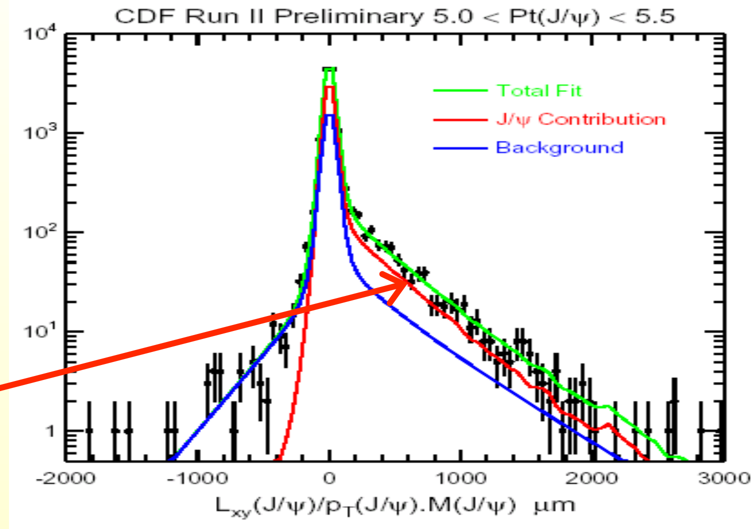
NLO QCD

Fragmentation

Factorizes into a calculable part and a non-calculable but universal piece

Inclusive b cross section - low P_T

- Run I: b cross-section ~ 3x old NLO QCD
- Theoretical approaches: Next-to-Leading-log resummations, non perturbative fragmentation function from LEP, new factorization schemes…
- **Experimentally: unbinned maximum likelihood fit to the J/ψ decay time in the R-φ plane to extract the b fraction**

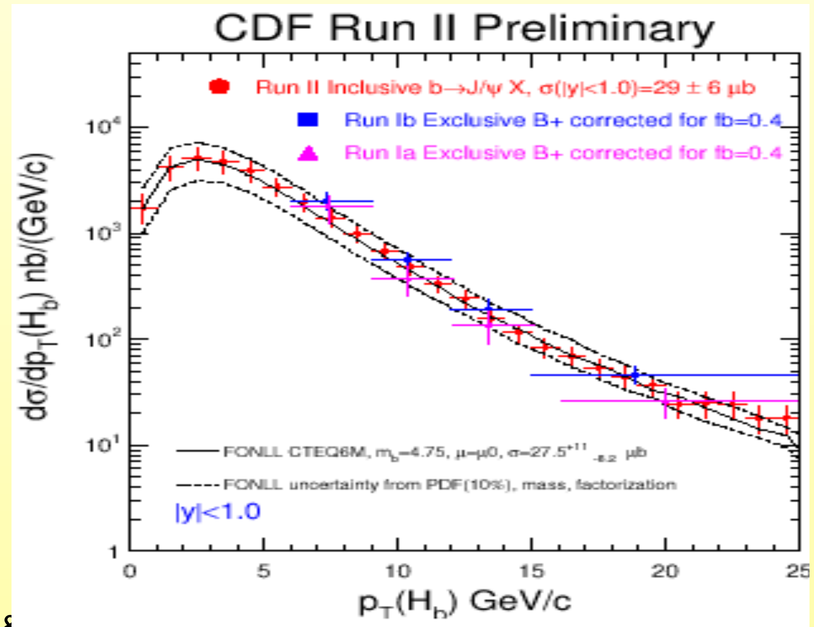


Run II: **Bottom Quark Production cross-section:**

$$\sigma(p\bar{p} \to bX)\big|_{|y|<1.0} = (29.4 \pm 0.6(stat) \pm 6.2(sys))\ \mu b$$

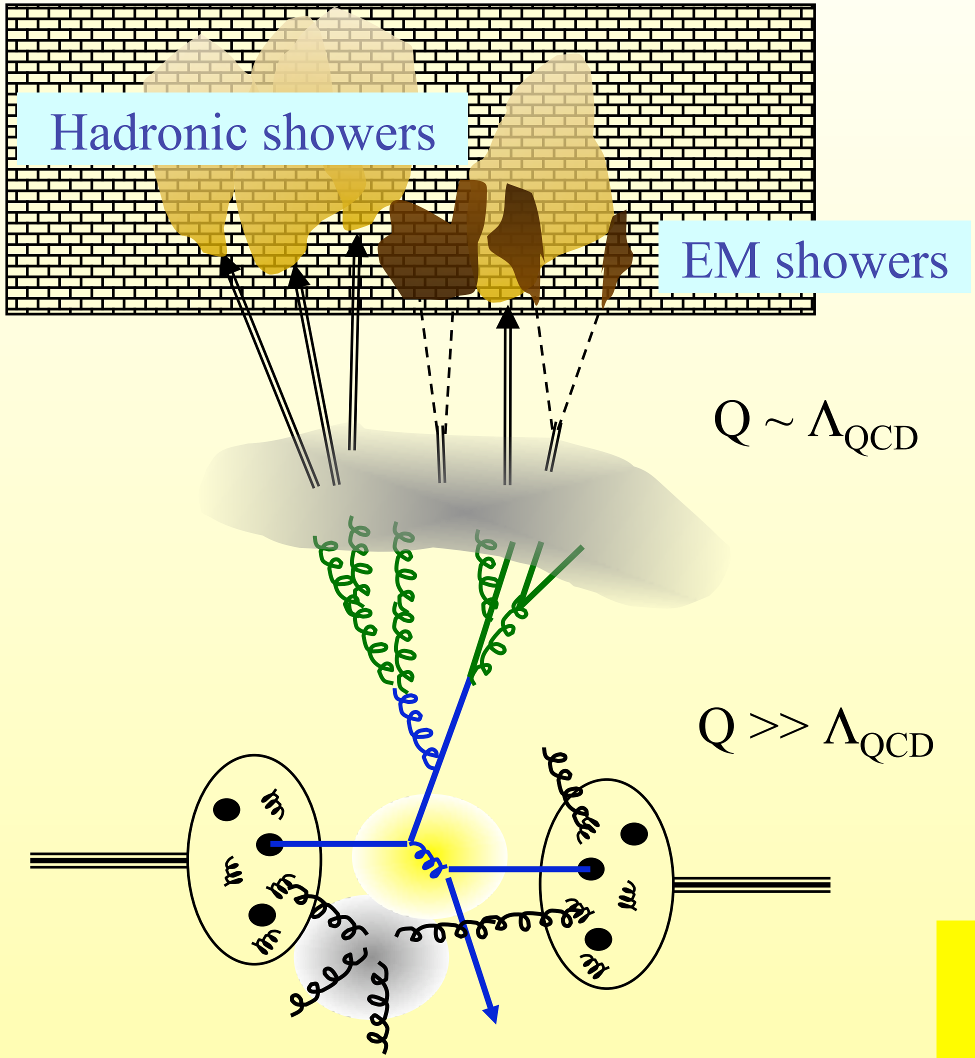
PRD 71, 032001 (2005)

Good agreement →



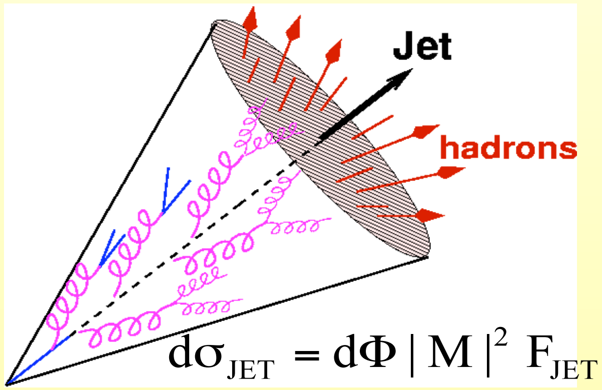
FONLL $\sigma(p\bar{p} \to bX)\big|_{|y|<1.0} = (27.5^{+11}_{-8.2})\ \mu b$

Jet definitions

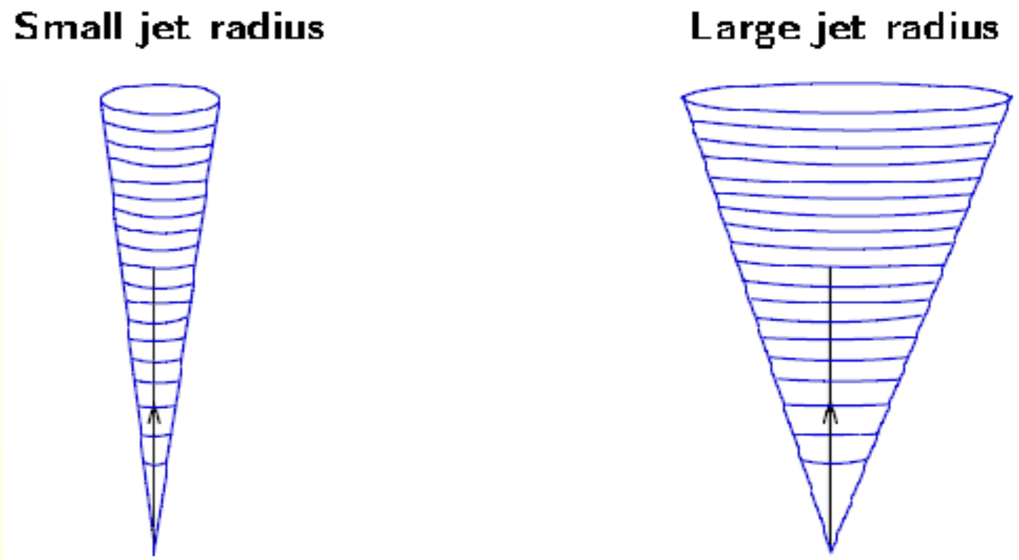


Unfold detector to Particle level
↳ Correct for efficiency, resolution

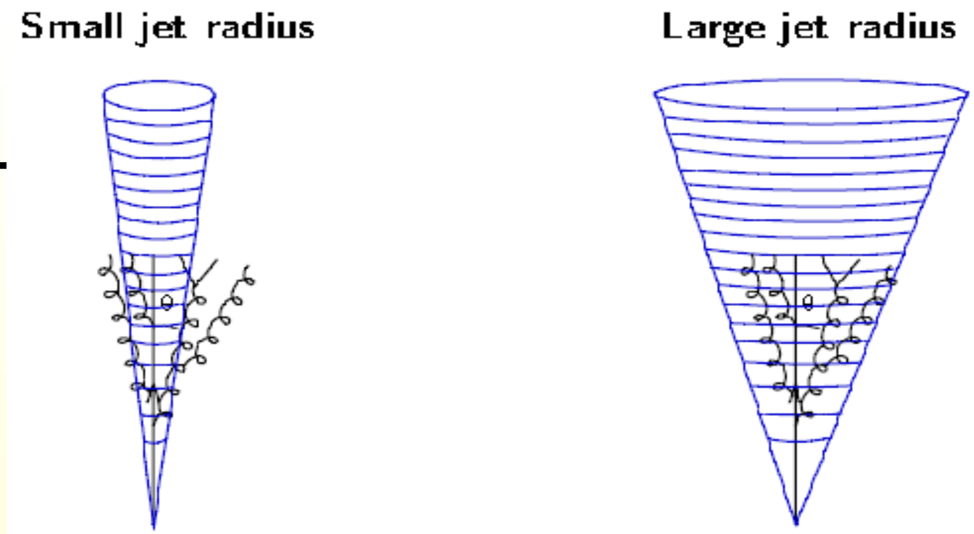
Correct theory (pQCD) for
non-perturbative effects
↳ Underlying Event, Fragmentation



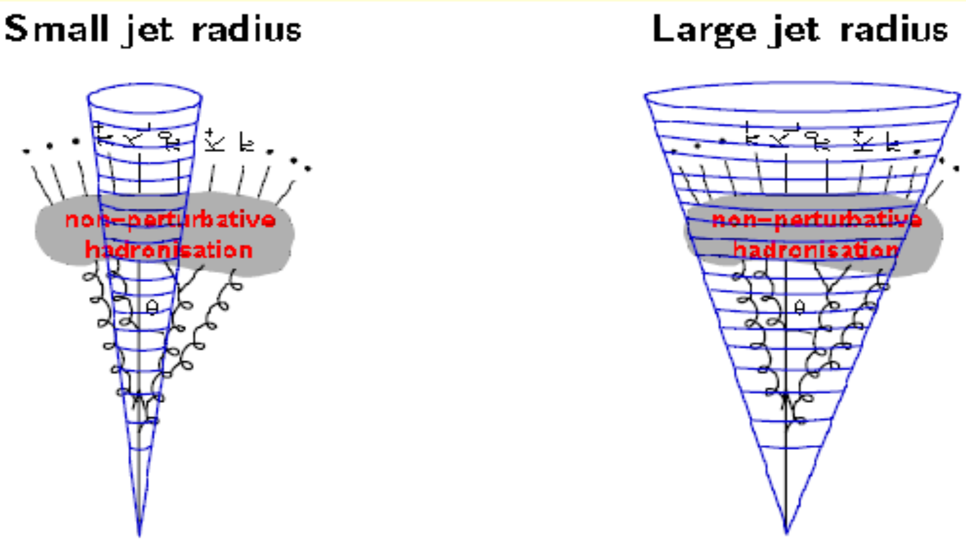
Well defined jet algorithm required
↳ At calorimeter, hadron and parton levels



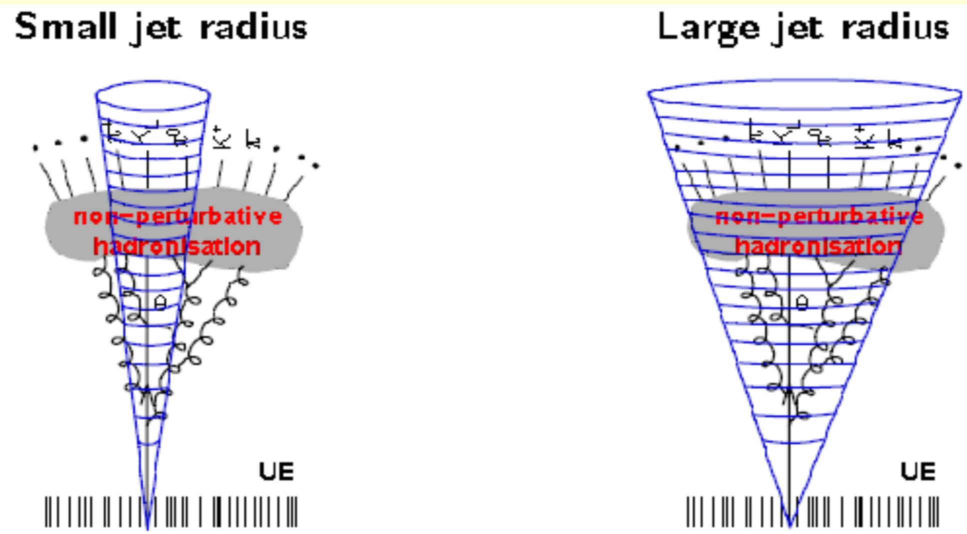
single parton @ LO: **jet radius irrelevant**



perturbative fragmentation: **large jet radius better**
(it captures more)

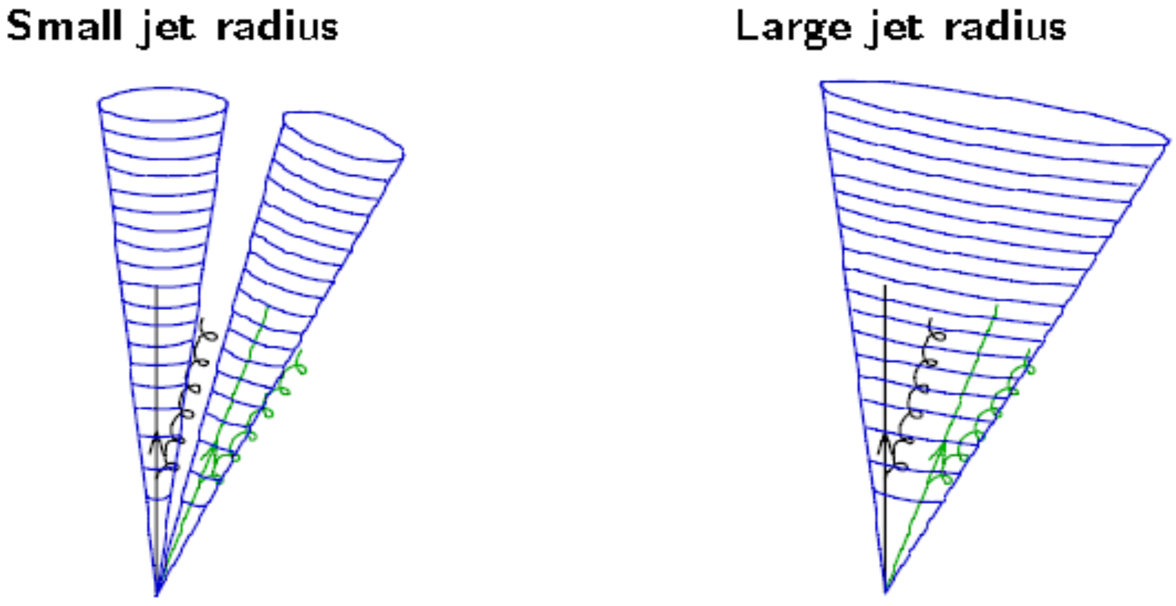


non-perturbative fragmentation: **large jet radius better**
(it captures more)



i. underlying ev. & pileup "noise": **small jet radius better**
(it captures less)

Jet issues



multi-hard-parton events: **small jet radius better**
(it resolves partons more effectively)

4-way tension in many measurements:

Prefer small R	prefer large R
resolve many jets (e.g. $t\bar{t}$)	minimize QCD radiation loss
limit UE & pileup	limit hadronisation

Jet definition choice

- A jet is not a parton: it's (sort of) what you choose it to be.
- It's easier to think in terms of partons (LO, NLO pQCD) with IR/Collinear safe jet algorithms. And gives sense to pQCD predictions
- ∃ many cones algs. Not equivalent. Many are IR/Coll unsafe.
  xC-SM → SISCone; xC-PR → anti-k_t
- "The best" jet definition *does not exist*
- To get the most out of jet-algs.,
  - Understand the interplay of physical scales — high p_t → larger R
  - Try out different combinations of algorithm & R
  - Check Variations of alg. & R don't change extracted physical quantities
- Special cases (e.g. boosted W/t/...) benefit from special techniques
  e.g. seq. recomb. "jet-decomposition" is a powerful tool

The Anti-Kt Algorithm

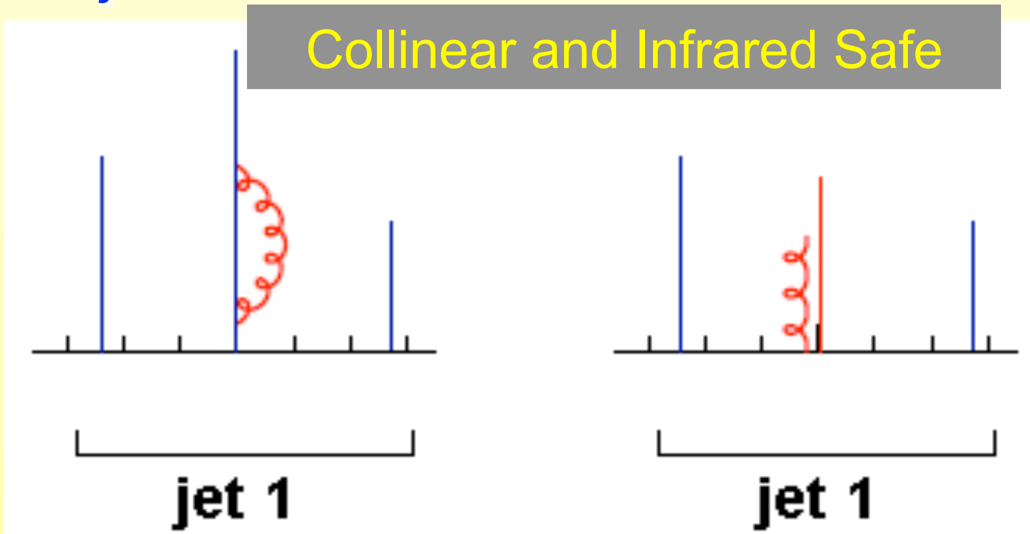
ATLAS default now (together with cone)

- Infra Red an collinear Safe.
- Good Speed (Better that the SISCone)
- Behaves like a idealized come algorithm, Conical Jets, active and passive areas are equal.
- Distance Definition:

$$d_{ij} = \min(k^{2p}_{ti},k^{2p}_{tj})\ \Delta^2_{ij}/R^2$$ *with p =-1.*

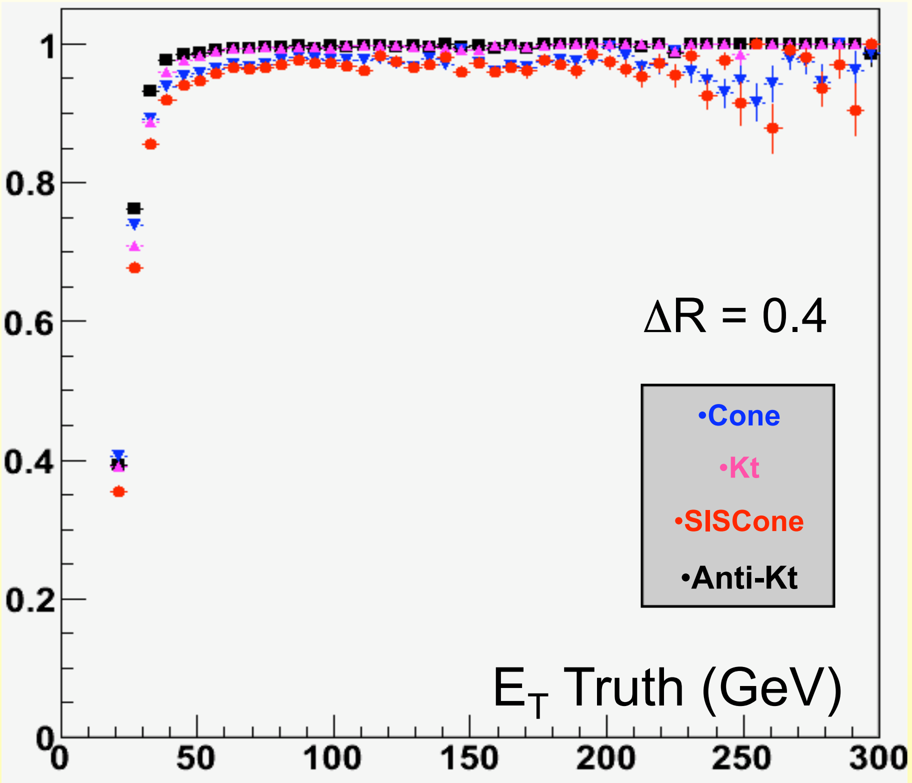
$$d_{ij} = \min\left(\frac{1}{k_{ti}^2}, \frac{1}{k_{tj}^2}\right)\frac{\Delta R_{ij}^2}{R^2}, \qquad d_{iB} = \frac{1}{k_{ti}^2}$$

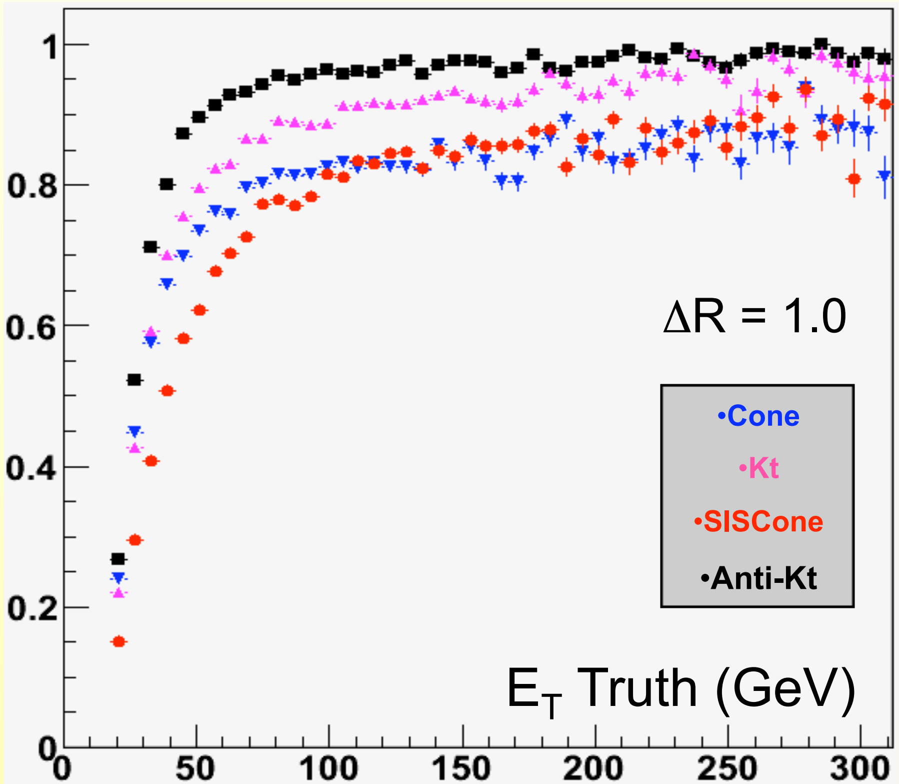
M. Cacciari,G. Salam, hep-ph 0704.0292v2



Top Jets reconstruction efficiency

ATLAS - Jet algorithm workshop january 09 and Lisbon hadronic workshop June 09
→ contribution to Atlas default Jet algo choice

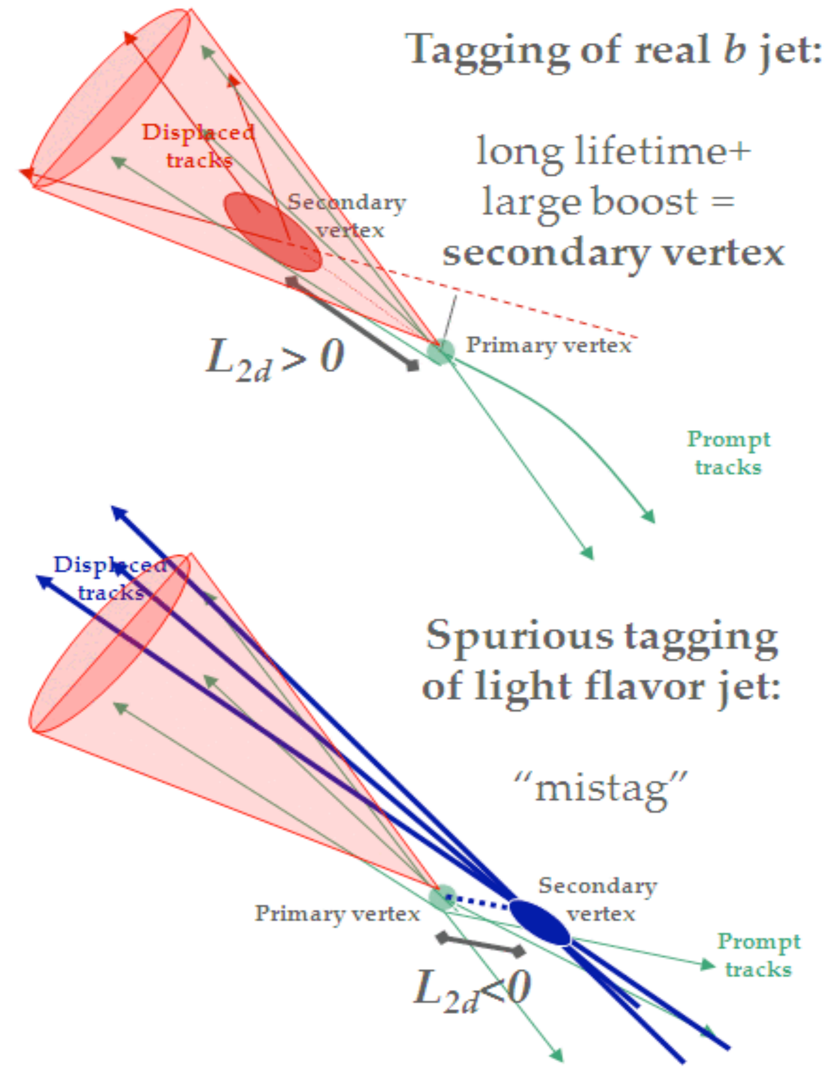




Work done at LPNHE with **HELEN** program fellow student from
Universidad de los Andes – Merida, Venezuela

- *To b or not to b ?*

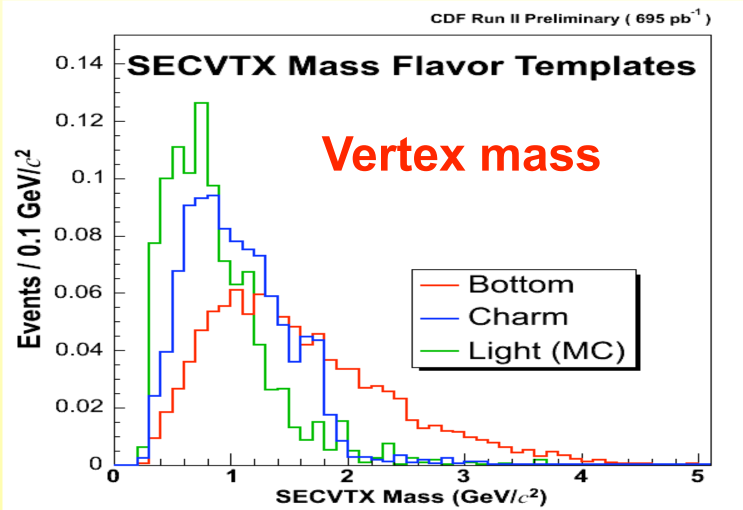
High P_T Jets and b tagging



Secondary Vertex b-tag Algorithm (CDF SecVtx)

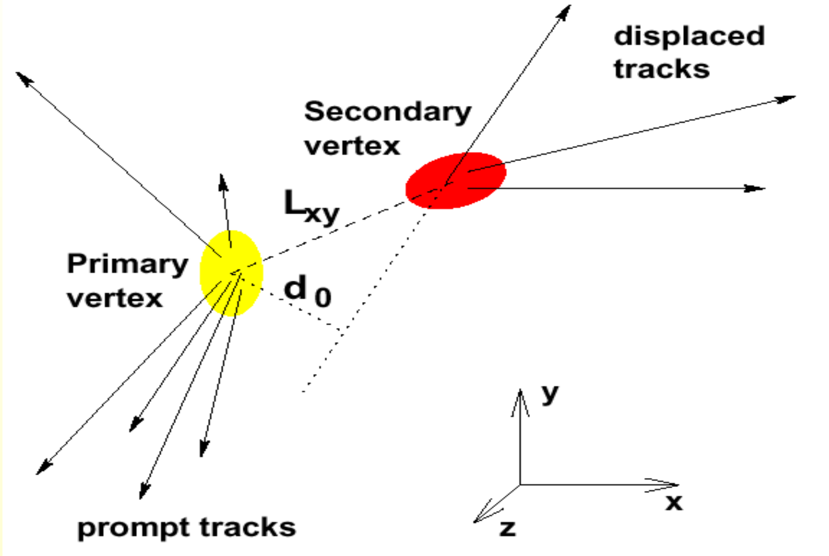
- Select tracks within the Jet Cone
- Reconstruct 2 or 3 tracks vertex
- Cut on vertex displacement significance to reduce light Jets rate

- **Secondary vertex mass** has some separating power between light, charm and b Flavors, as like as other variables

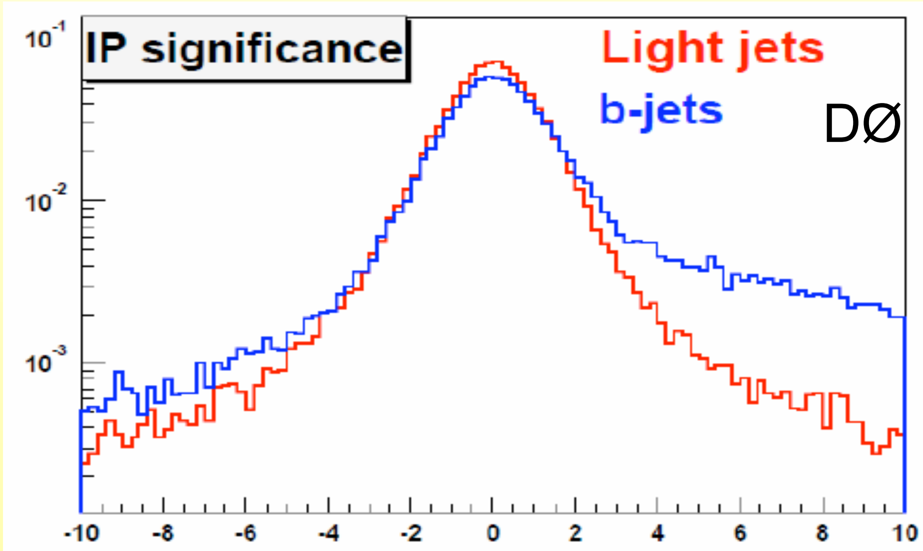


Signatures of B-Jets

- High b-hadron mass
  - 5.3 GeV/c^2
- Relatively long lifetime
  - 1.5 ps → $c\tau$ = 450 μm
- Hard fragmentation
  - b-hadron retains ~70% of b-quark momentum



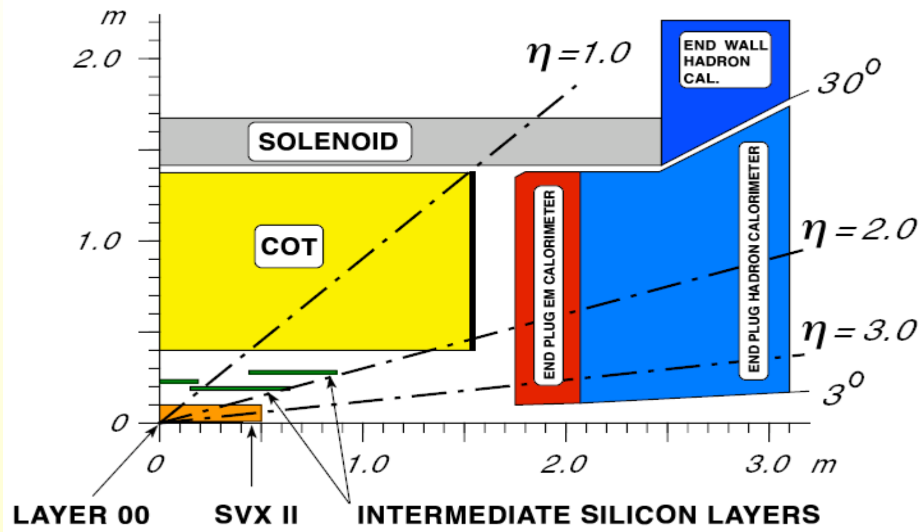
- Put 'em all together and you get
  - High-p_T tracks
  - Large impact parameters
  - Secondary vertices (few mm)

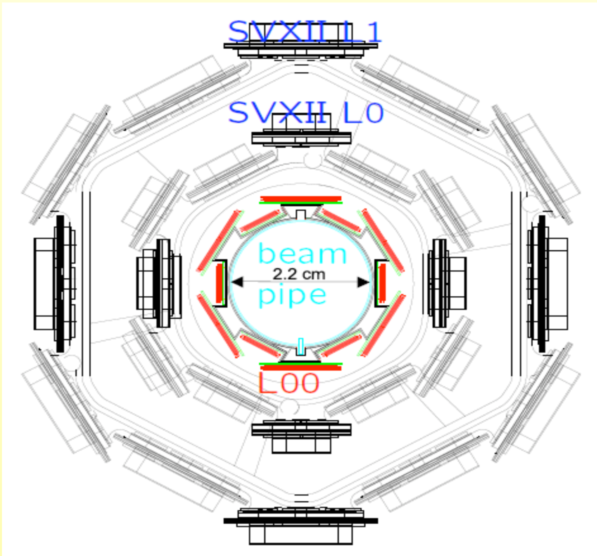


- Lepton production
  - ~10% b → $\ell\nu$
  - Also ~10% c/$\underline{c}$ → $\ell\nu$
  - High-p_T tracks
  - High-p_T relative to b-jet

The CDF Detector

- COT: 96-layer wire drift chamber
  - 4 axial, 4 stereo superlayers
  - 1.4 T magnetic field
- "Central" lepton ID out to |η|<1.1
- Silicon tracker
  - 95 cm long Layer 00
    - Radiation hard
  - 3 SVX-II barrels, each 29 cm long





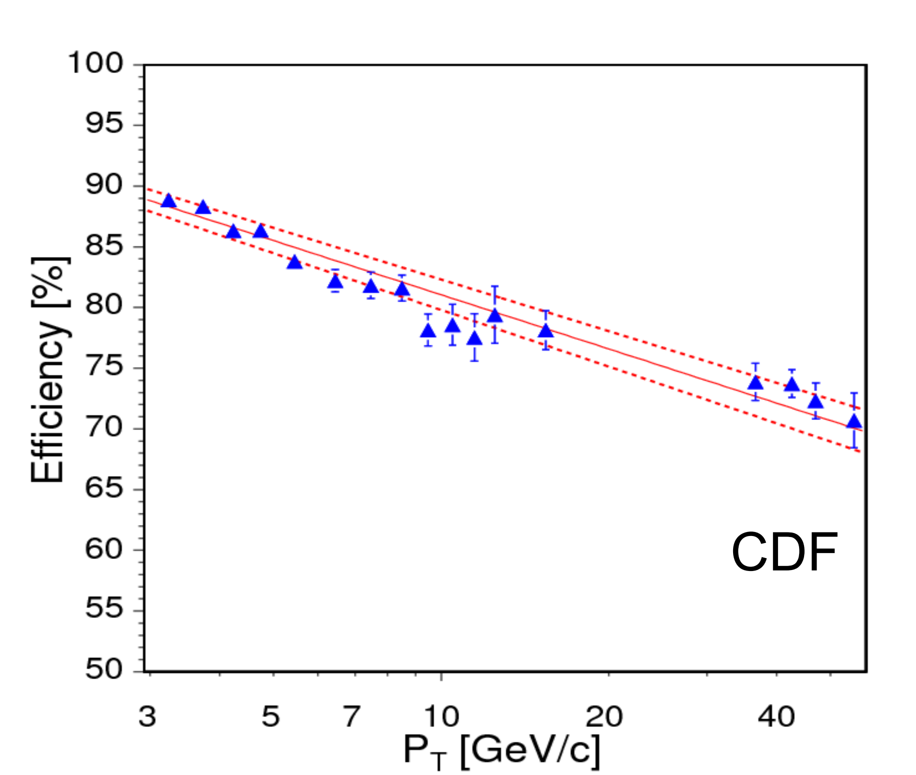
- L00 – singled-sided, r = 1.2 cm
- SVX-II – 5 double-sided layers
  - L0: rϕ – rz, r=2.5 cm
  - L1: rϕ – rz, r=4.5 cm
  - L2: rϕ – 1.2° stereo, r=6.5 cm
  - L3: rϕ – rz, r=8.5 cm
  - L4: rϕ – 1.2° stereo, r=10.5 cm
- Total sensor area 6 m^2
- 720k electronics channels

Detector Issues

- Reference for impact parameter calculations
  - Tevatron beam profile is ~30 μm in xy
  - Can improve by computing event-by-event primary vertex
  - Fit all tracks to a common point – discard outliers – iterate
  - Final resolution is 10-30 μm depending on event topology
- Poorly-reconstructed tracks are a killer – quality cuts are critical
  - Both experiments use silicon hit multiplicity and fit χ^2
  - Remove K_S/Λ decay products
  - Raising p_T cuts helps control fake tag rate
- Inactive detector regions and data/simulation mismatch
  - CDF: model inactive silicon ladders in the simulation run-by-run
  - DØ: everything relative to “taggable” jets (couple of silicon tracks) – get taggability rate from data (~80%)
  - Both experiments apply a scale factor to simulation efficiency to match what’s seen in the data (0.7 – 0.9, depending on tagger)

Soft Lepton Tagging

- Tag b-tags by identifying a lepton from b or c decay
- CDF & DØ both do it for muons
- ID requirements somewhat different than high-p_T case
  - Can't use calorimeter energy
  - Track-stub angular matching is more effective
- Typical ID efficiency 80-90%
  - Per-jet efficiency ~10%
- Mis-ID rates ~0.5% per track
- Electrons are more difficult
  - Had it in Run I
  - Under development at CDF

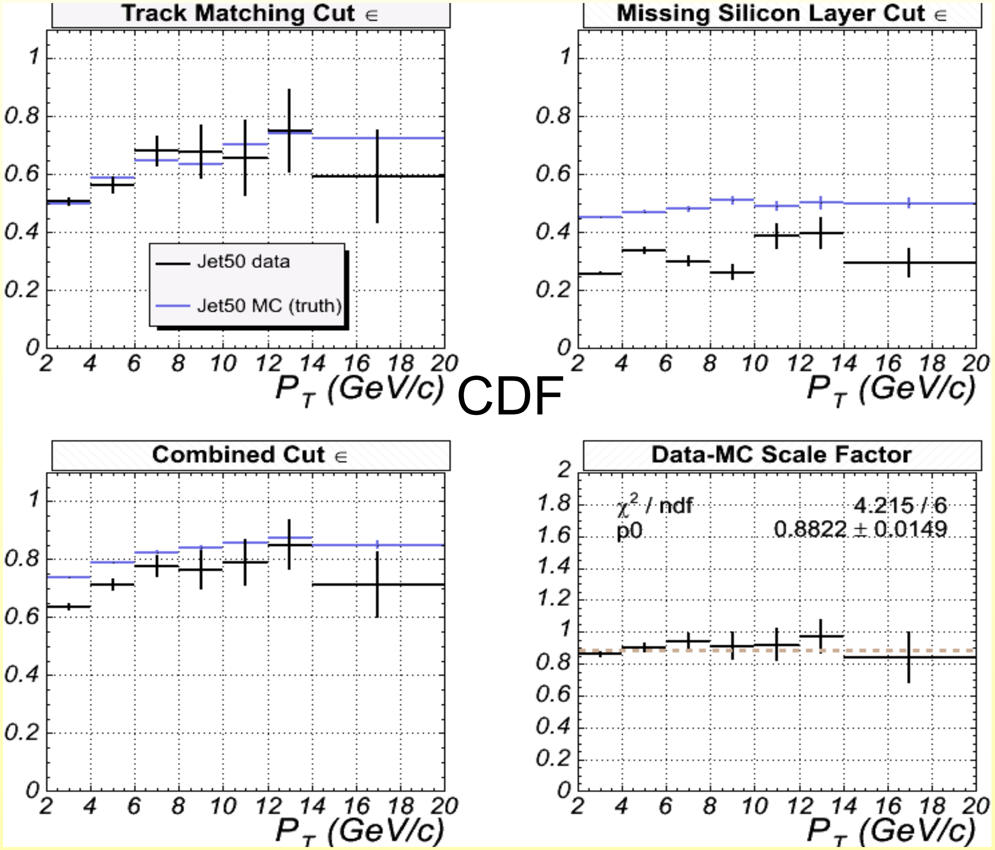


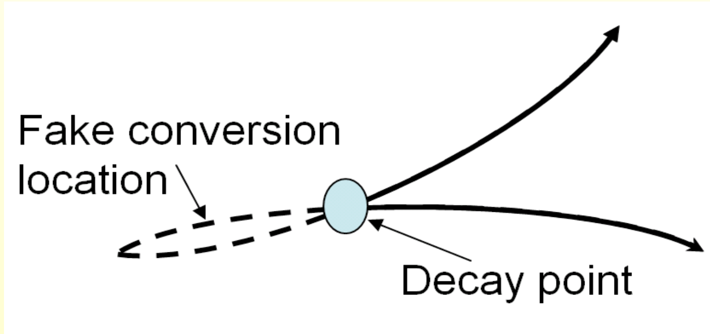
Soft lepton taggers use different information than lifetime taggers

Can use overlap rate for calibration

Soft Electron Tagger

- Additional information looking for "soft" (pT>2GeV) HF electrons within Jets
- Conversions electrons are main background
- Complementary soft electron tagged samples for tagger performance studies



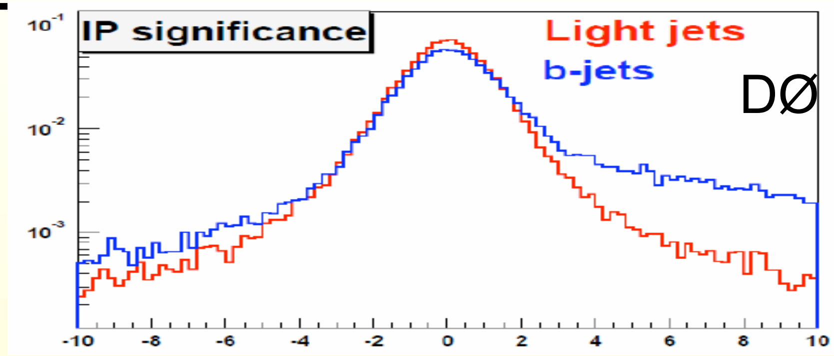


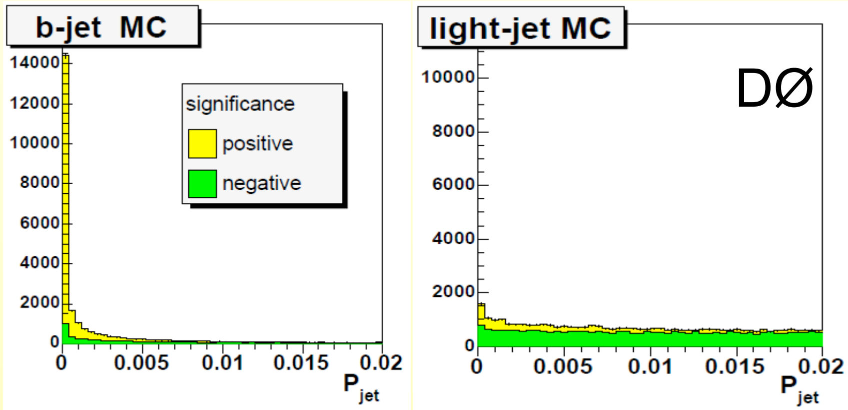
Need model for conversions to identify them: use partner legs

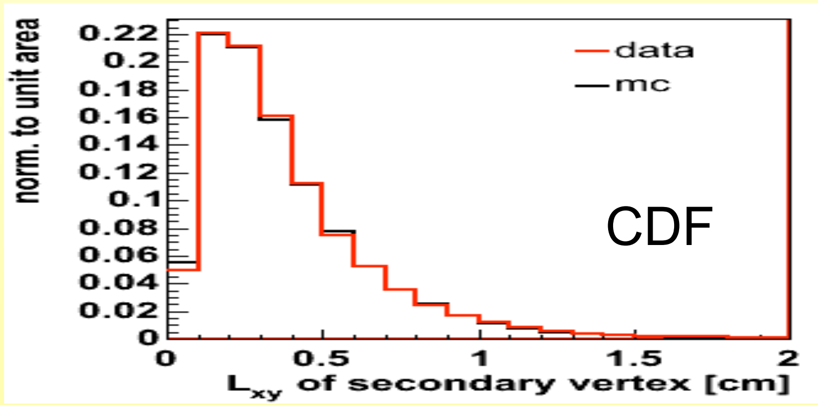
-In development also in ATLAS exploiting fine ecal segmentation and excellent tracking system.

Impact Parameter Taggers

- Count displaced tracks (DØ)
  - three 2σ or two 3σ
  - sign IP's against jet direction
- Jet probability (CDF & DØ)
  - Joint probability for all tracks to come from primary vertex
  - Track resolution derived from negative IP tracks
  - Use only positive IP tracks in probability calculation
- Displaced vertex (CDF & DØ)
  - Fit displaced tracks (above p_T cuts) to a common vertex
  - Prune tracks and cut on fit χ^2
  - Cut on L_{xy} significance
- All algorithms can be tuned by adjusting cuts – 2-6 operating points





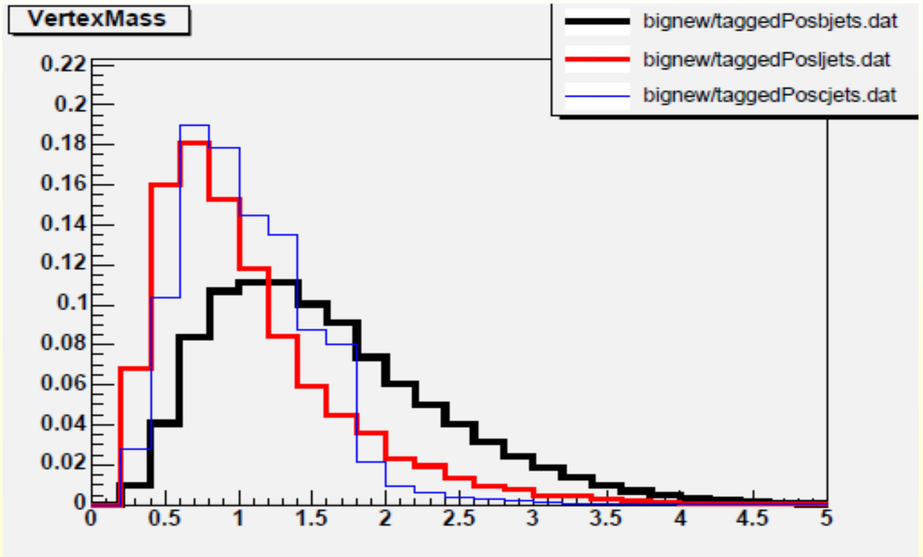


Multivariate Algorithms

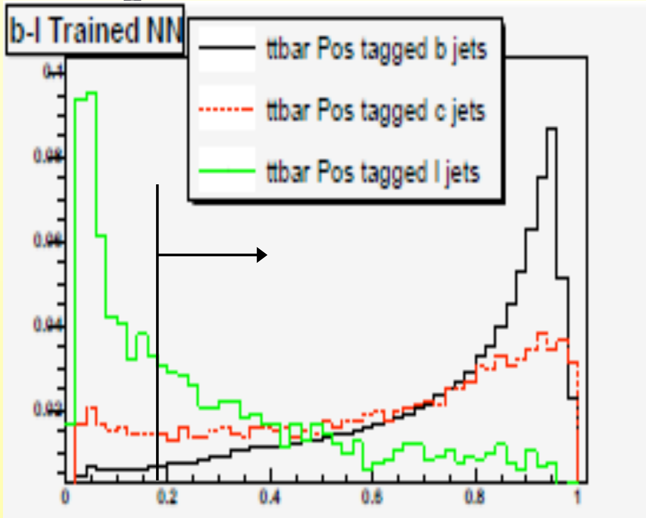
- Taggers are correlated, but not 100%
  - Can gain by combining them
  - CDF has a simple “combined tagger” – logical OR of displaced vertex and jet probability taggers
    - Gain 15-25% efficiency, at cost of factor two mistag rate
- Use more information than just the tagger outputs
  - Displaced vertex tagger gives you more than just yes/no
    - Many vertex properties discriminate signal/background
- Both experiments have new multivariate taggers
  - CDF: start with displaced vertex tag, try to reject charm/light while preserving b-tags
  - DØ: no displaced vertex prerequisite – can optimize for better purity or enchanced efficiency

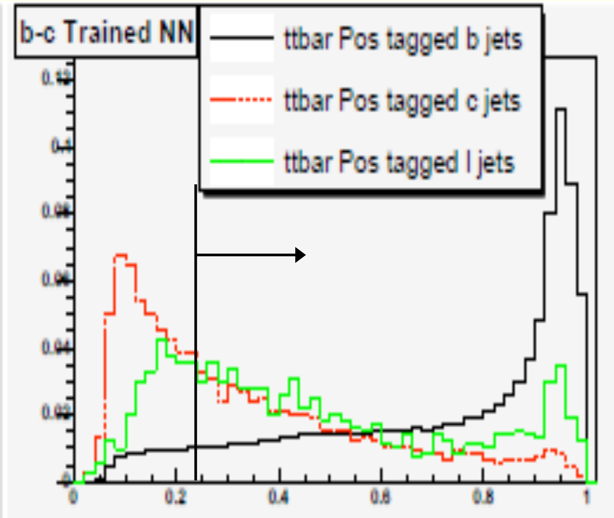
Secondary Vertex + Multivariate Tagger (CDF)

- Two 16-variable neural networks
  - Vertex mass, L_{xy}, χ^2, p_T
  - High-IP track p_T, multiplicity
  - Jet probability
  - and so much more!
- Train one to separate b-vs-light, the other b-vs-c
- Choose cuts to preserve 90% of b
  - Reject 45% of c, 65% of light



- Already in use
- Top cross section
- WH search
- Similar, dedicated algorithm for single-top search





- *Evaluating the performances of b-tagging*

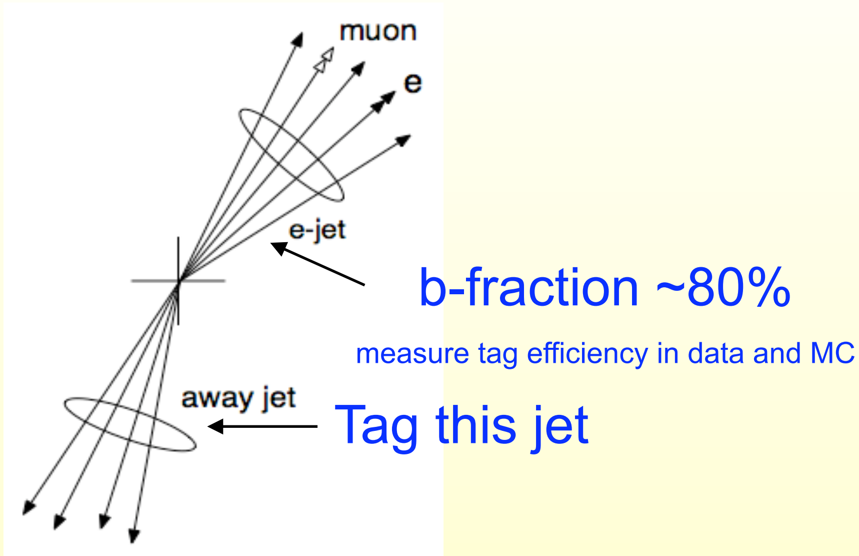
the Tevatron experience on impact parameter tags

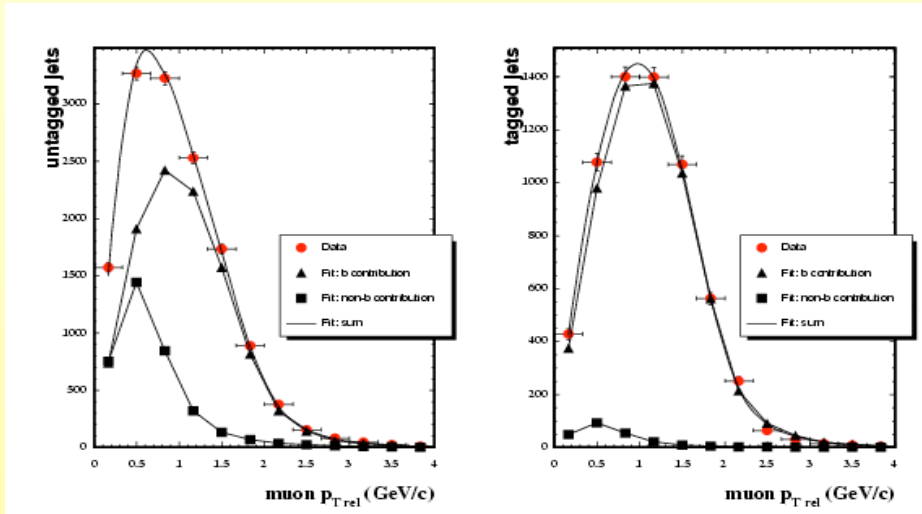
Efficiency Measurement (CDF)

- Based on 8 GeV/c electron and muon data samples
- Tag away jet to enhance b-fraction
- Generate matching MC samples

- Method A: muon p_{Trel} fits
  - Fit muon p_{Trel} against jet in tagged and untagged jet
  - Extract numbers of b-jets in each and compute efficiency
  - Systematics ~3%, mostly from modeling of p_{Trel} templates

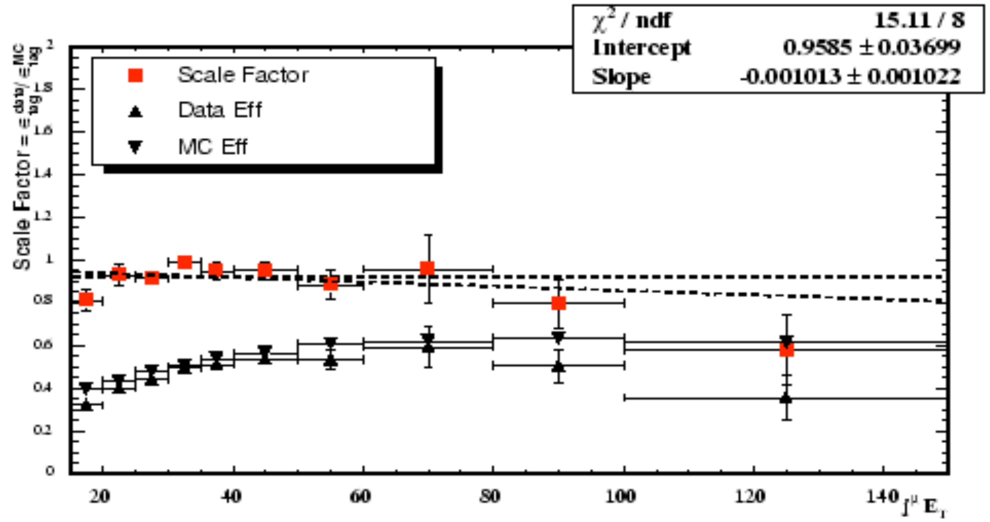
- Method B: electron double-tags
  - Infer non-b component from electron jet single-tag rate and conversion sample
  - Systematics ~5%, mostly from mistag subtraction



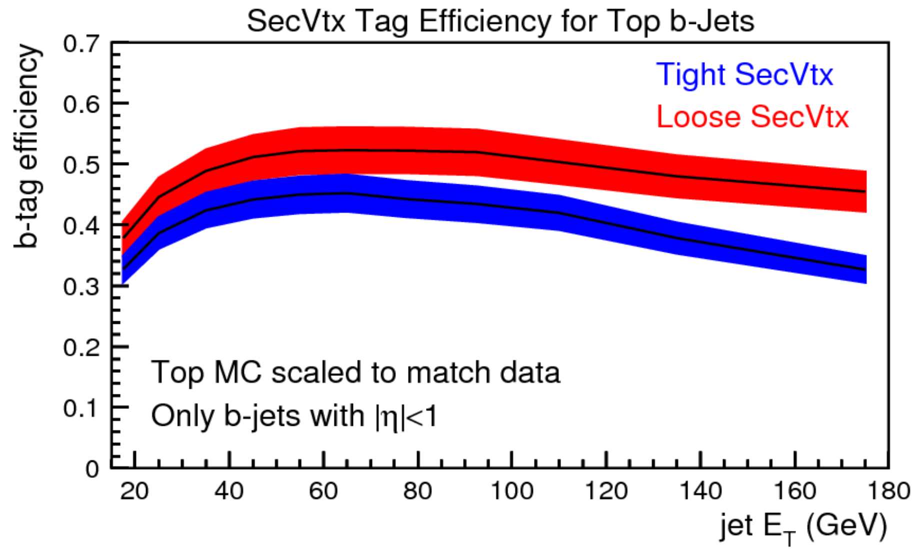


Efficiency Measurement (CDF)

- Systematics quoted earlier for each method were only the “internal” ones
- Additional systematics related to extrapolation to physics samples (in particular, b-jets from top)



- Jet E_T dependence
  - Convolute binned scale factor with E_T spectrum from top
  - 7% uncertainty



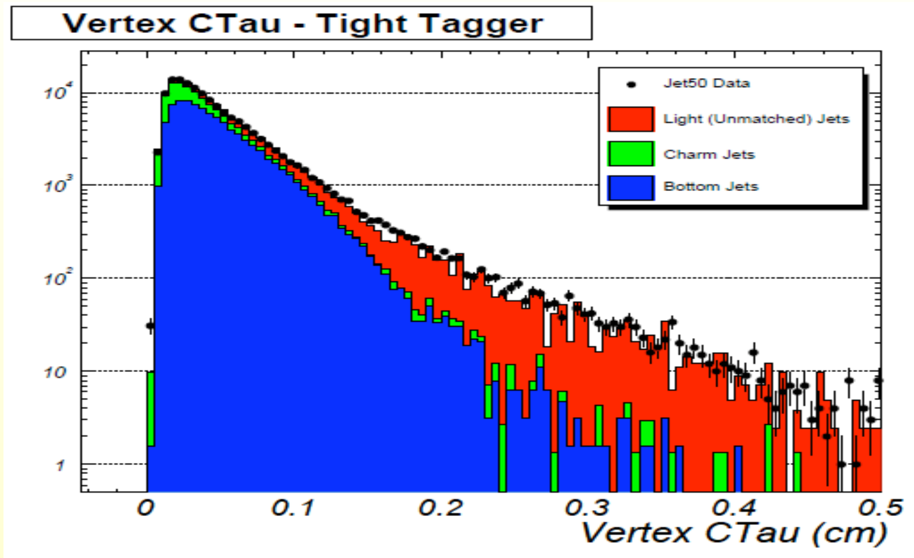
- Similar procedure for jet η
  - 1% uncertainty
- 2% uncertainty assigned for differences between semileptonic/ generic B-decays
- Total uncertainty on data/MC scaled factor : 7.3%

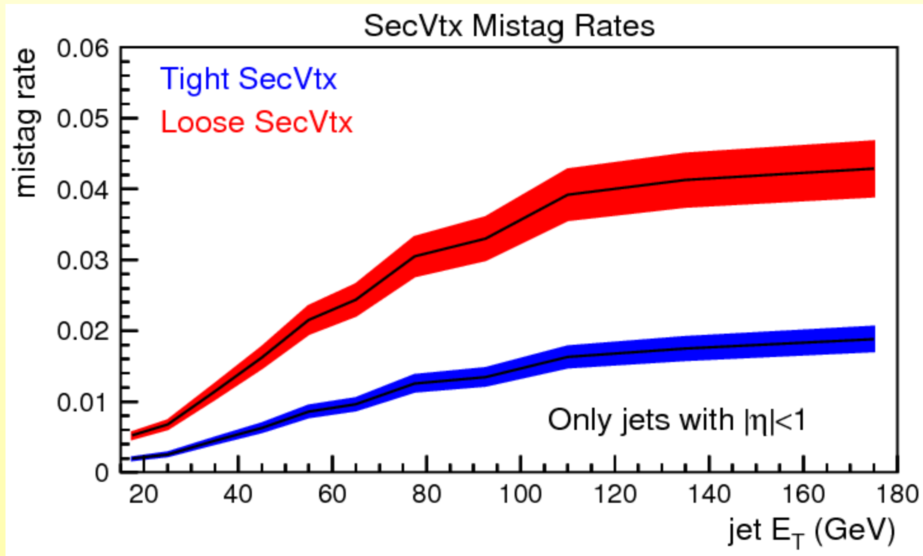
Fake Tag Rates

- Fake tags (mistags) mostly from misreconstructed tracks
  - +/- symmetry
- Estimate fake rates by
  - Using tracks with negative signed impact parameters
  - Using displaced vertices behind the primary w.r.t. the jet direction

- Both experiments use parametrizations of “negative” tag rates (in E_T, η, etc) to predict mistag rate in data samples

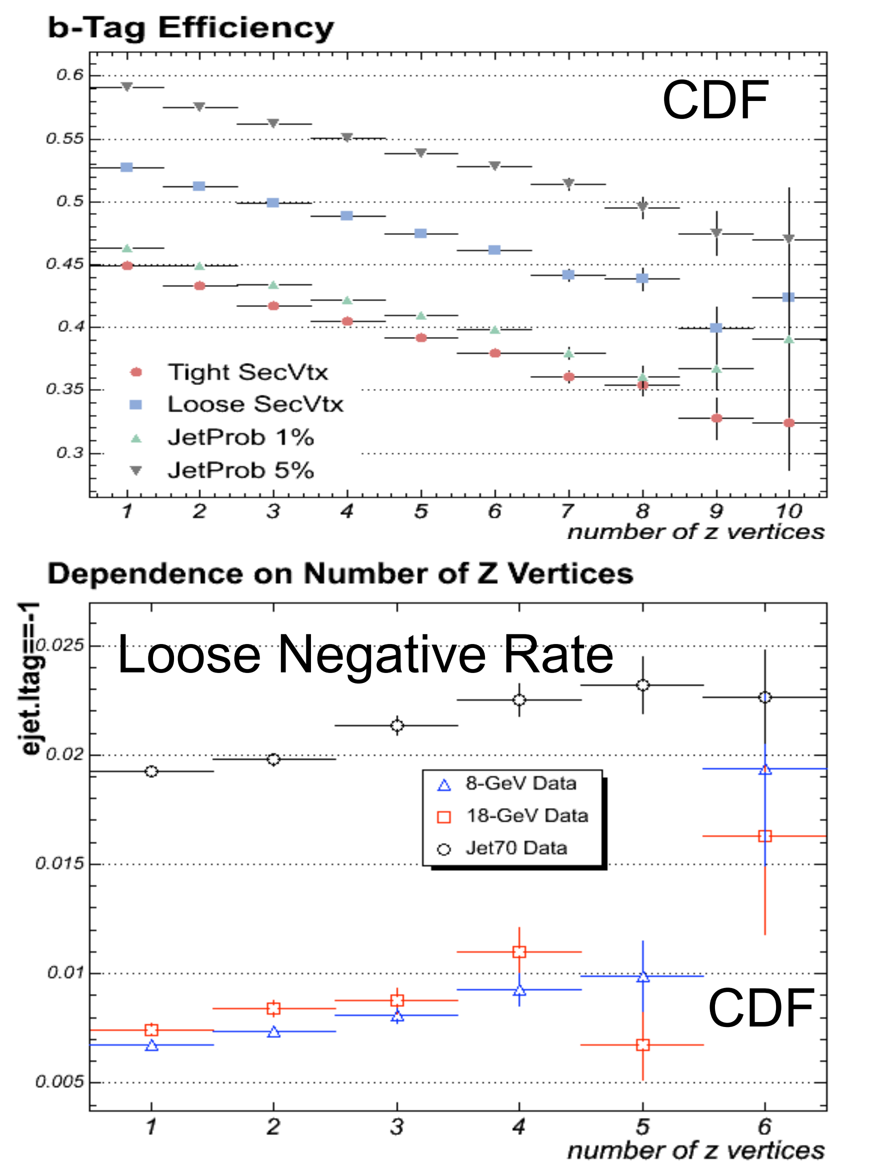
- Complications
  - Negative tags from heavy flavor (~10-15%)
  - Mistag rate not exactly symmetric – stray K_S/Λ and interactions with detector material
  - Fit POS-NEG excess in pseudo-cτ to estimate (~40%)
  - Net effect ~30% enhancement over negative tage rate





B-Tagging at High Luminosity

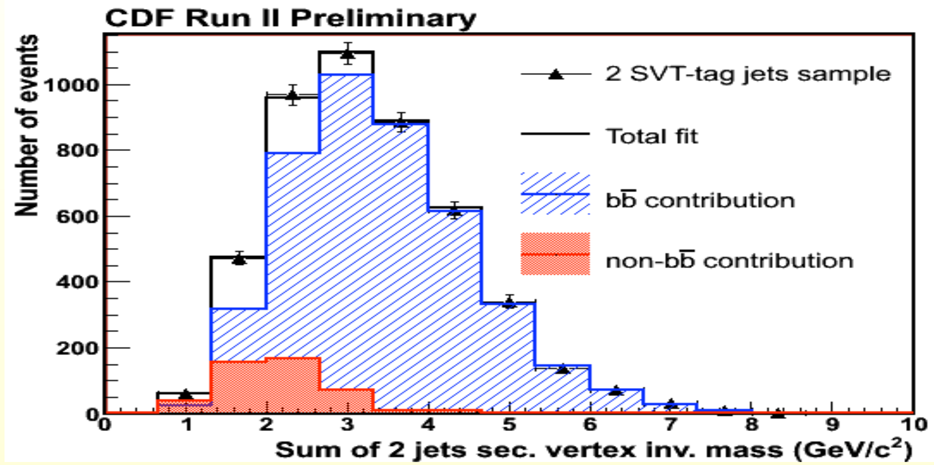
- Starting to study effects of high luminosity on b-tagging (results are extremely preliminary)
- N_{vtx} α Luminosity
  - At 3E32 $cm^{-2}s^{-1}$, $<N_{vtx}> = 4$
- MC study of top events with extra minimum-bias overlaid indicates that efficiency decreases
  - High tracker occupancy
  - Hits merge together
  - Pattern recognition harder
- Negative tag rates in data rise
  - Silicon hit merging?
  - Another interaction nearby?



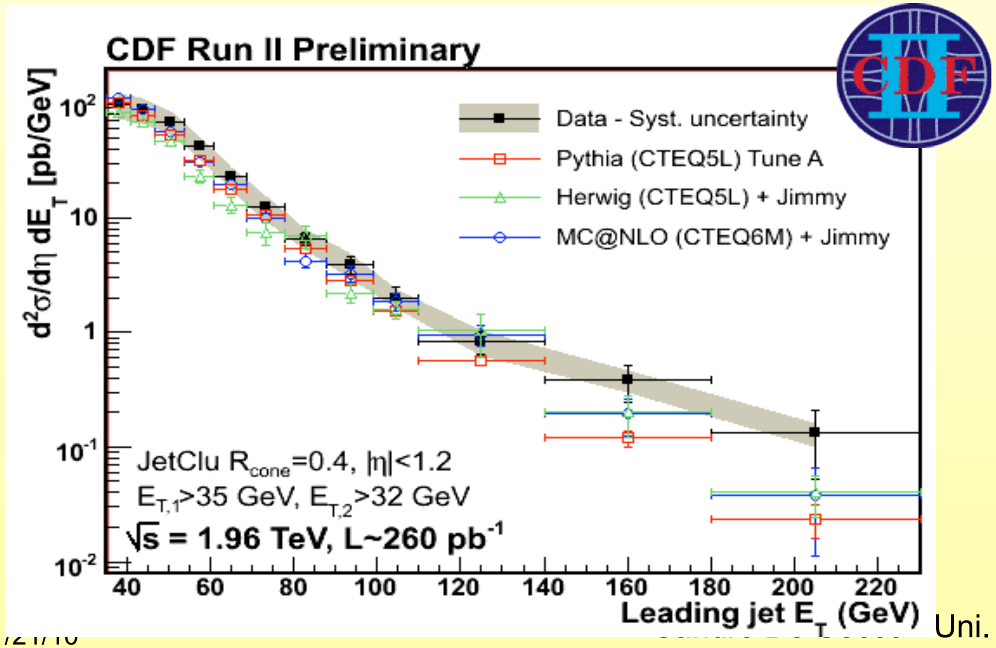
- *… or directly triggering on b-Jets*

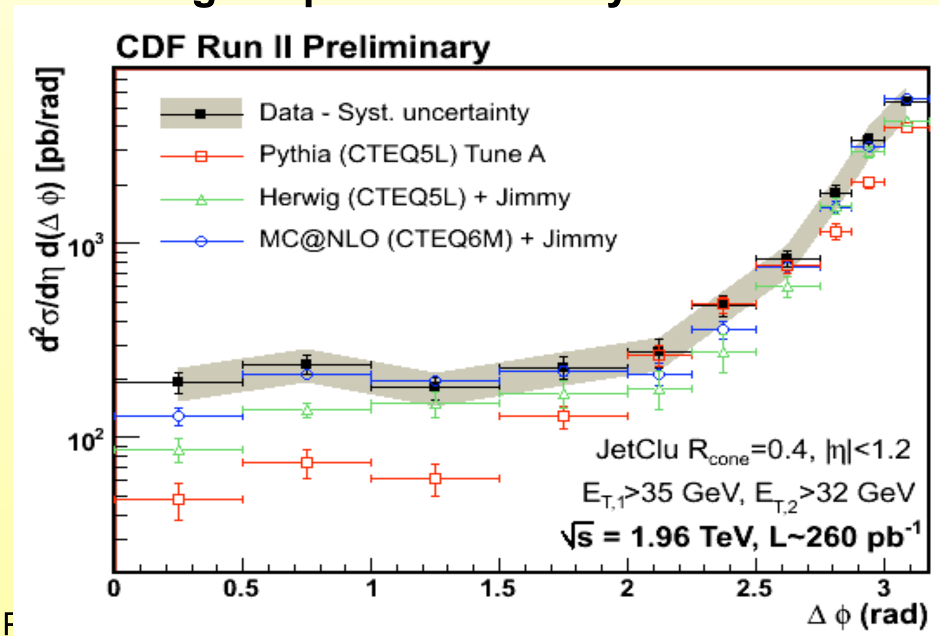
Inclusive bb di-jets production

- Specific Trigger based on L2 SVT (Secondary Vertex Trigger) used.
- Sensitive to the different production mechanisms
  - Flavor creation at high $\Delta\phi$
  - Flavor excitation or gluon splitting at low $\Delta\phi$



Purity ~ 85 % : extracted from data using shape of secondary vertex mass



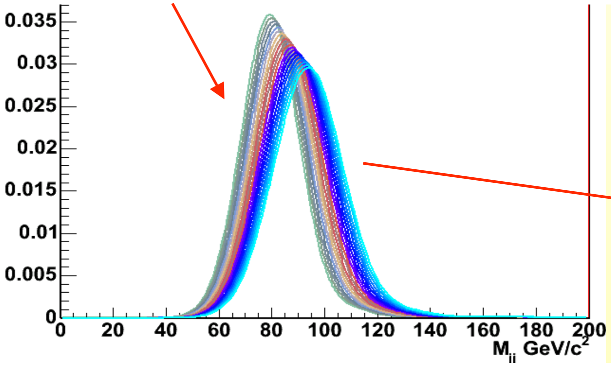


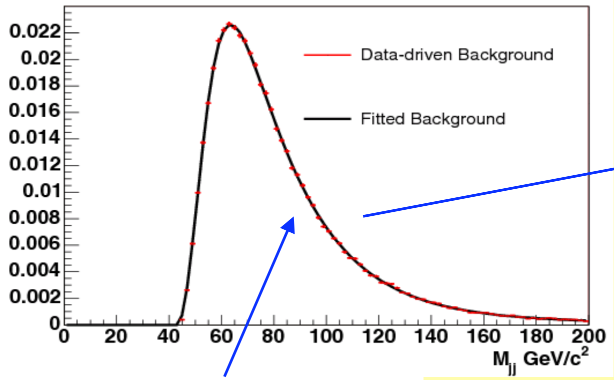
b-Jets Energy Scale from Z→bb

- Generic Jet Energy Calibration of the Detector needs specific correction for b-Jets
- Reduction of uncertainty in b-JES important for Top mass measurement
- Test algorithms to improve b-jet energy resolution is crucial for low mass Higgs search.
- Use of L2 SVT based b-tag di-jet trigger → Extract a signal, measure data/MC b-JES

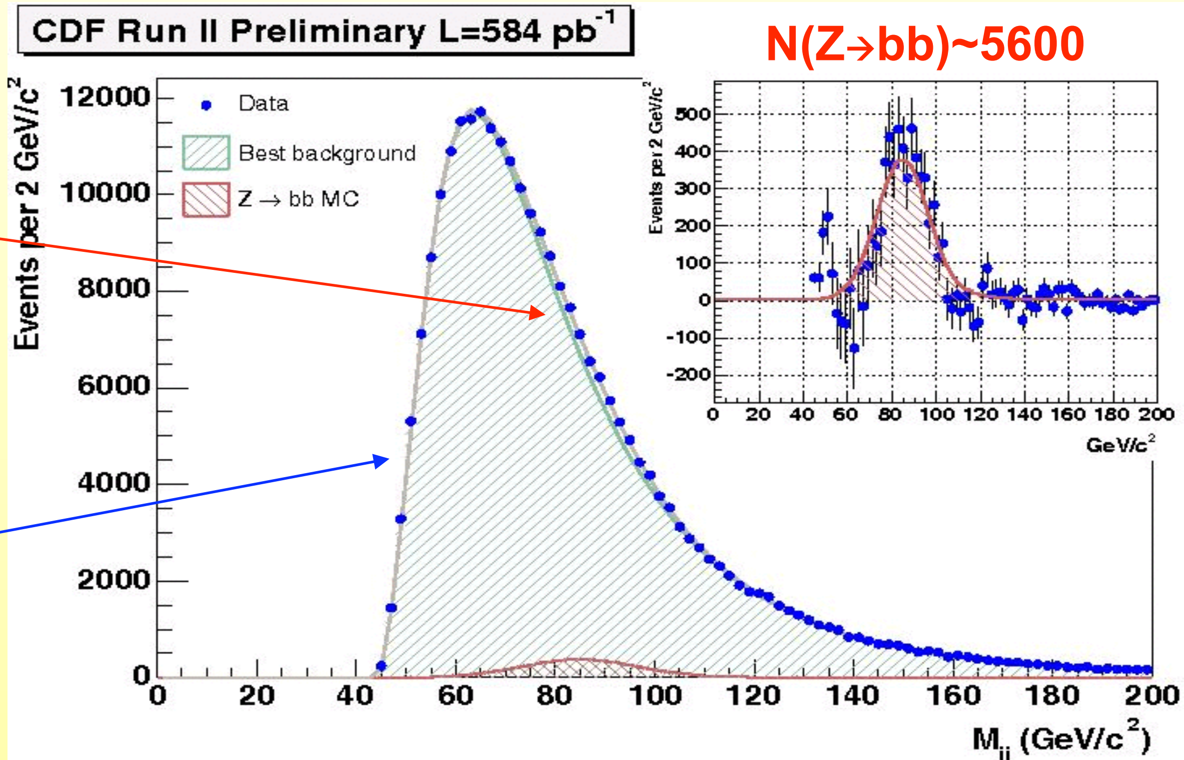
Result on 600 pb^{-1} → b-jet Data/MC Energy Scale Factor = 0.974 ± 0.020

Z→bb signal templates for 0.9<SF<1.09





Background template

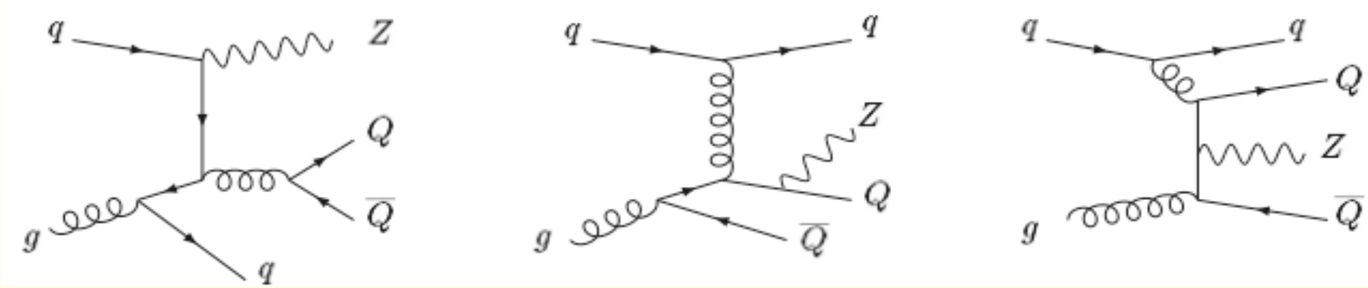


- *from inclusive b-Jets production…*

- *… to gauge boson + HF associated production*
- *Main background to Top, Higgs and NP searches*

- *→ need for precise knowledge of b-tag performances*

W / Z /γ bosons + HF Jets production



The study of W / Z / γ + Jets production is very relevant:

- to QCD:
  - High q^2 (~$M_{\mathbf{boson}}$) interactions, perturbative theory
  - Less leading diagrams (wrt pure Jets production)
  - Test Monte Carlo generators (ALPGEN…)
  - Probe for protons PDF (in particular W/Z + H.F.)

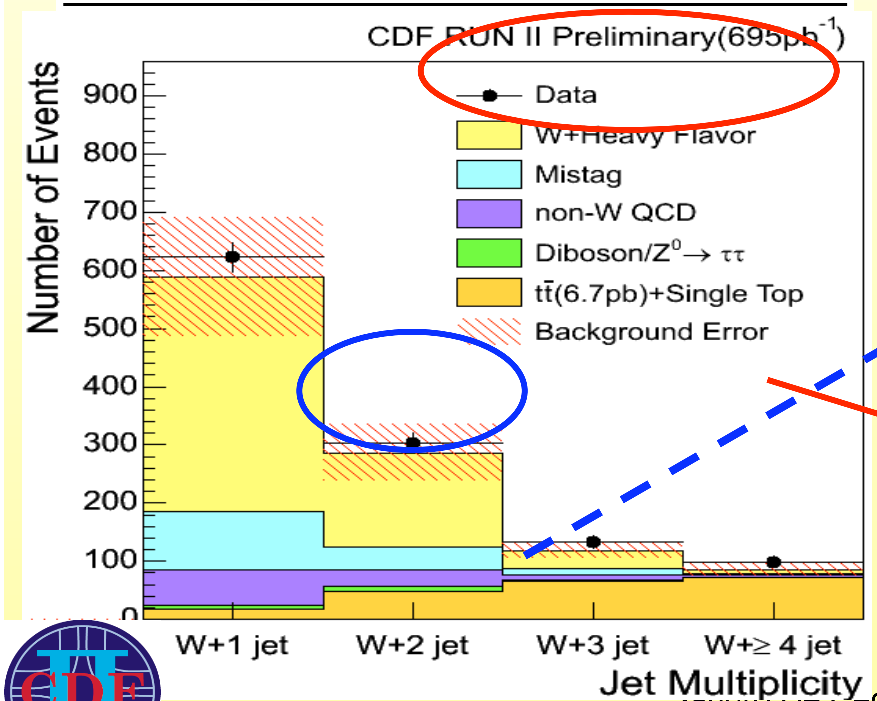
- for many high p_T analysis where W/Z+HF is a main background:
  - ✓ Top quark cross section and mass
  - ✓ Single Top cross section
  - ✓ Search for low mass Higgs boson
  - ✓ Several SUSY searches

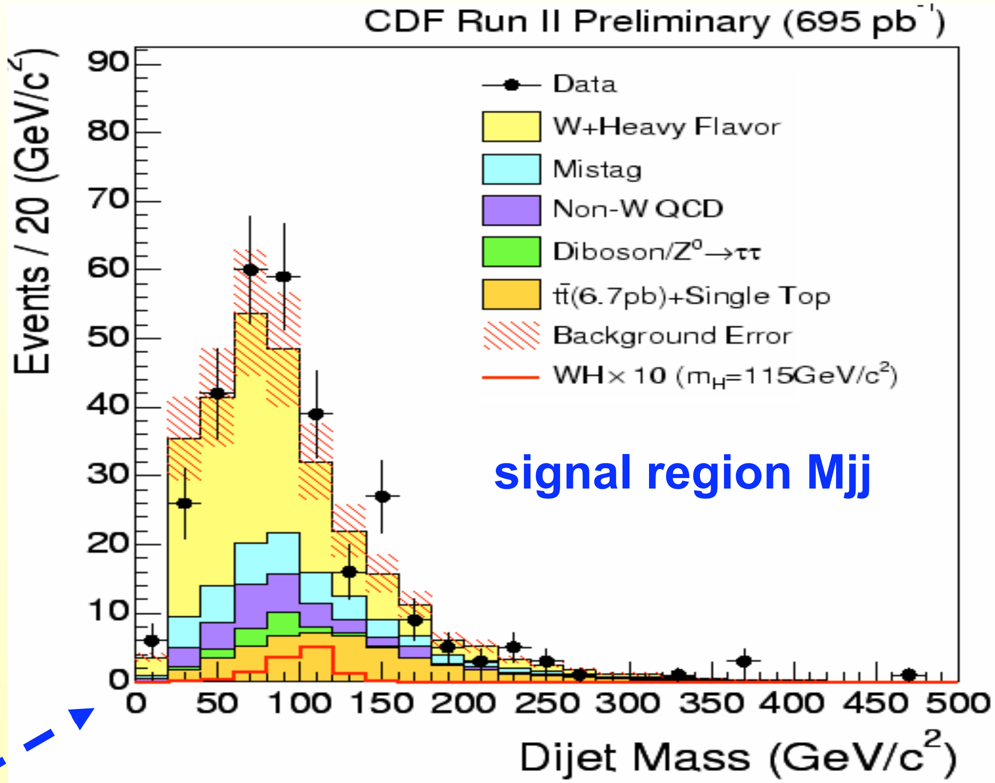
… and plays crucial role also at LHC !

W+HF backgrounds in HIGGS WH→ lν bb

= 1 jet tagged jet sample

Jet Multiplicity	w/o NN tag	w/ NN tag
Before b-tagging	10647	10647
Mistag	119.4 ± 19.4	41.8 ± 9.0
$Wb\bar{b}$	130.4 ± 44.6	120.2 ± 41.1
$Wc\bar{c}$	48.9 ± 16.7	33.7 ± 11.5
Wc	47.4 ± 12.4	25.0 ± 6.5
$t\bar{t}$(6.7pb)	43.1 ± 7.3	37.8 ± 6.4
Single Top	22.7 ± 2.4	20.1 ± 2.1
Diboson/$Z^0 \rightarrow \tau\tau$	17.0 ± 2.5	10.6 ± 1.7
non-W QCD	44.4 ± 7.7	29.5 ± 5.1
Total Background	473.4 ± 66.9	318.8 ± 54.7
Observed($\geq$1tag)	514	332



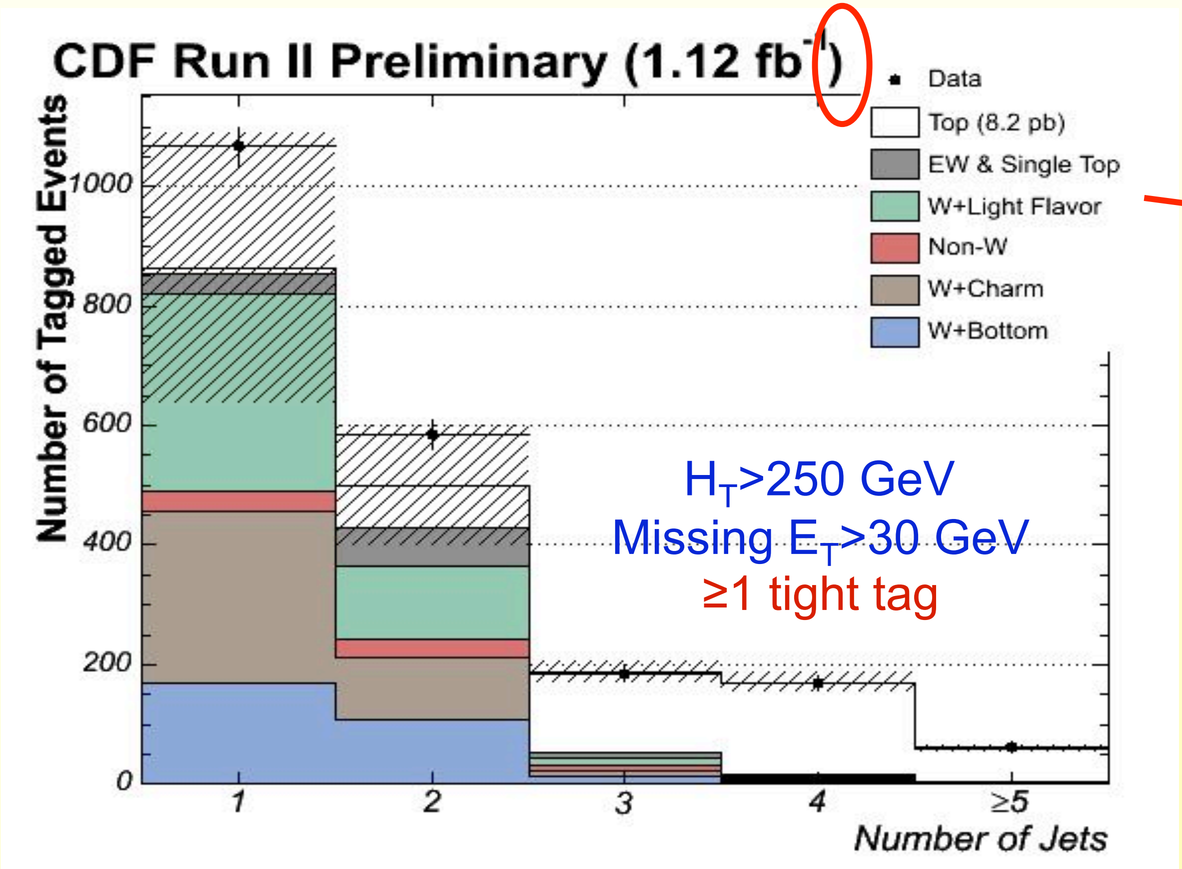


Expectations based on:

- **MC predictions for EW and top backgrounds**
- **Fake tags (mistag matrix) applied to light jets**
- **Heavy Flavor background:**

admixture of MC and data driven procedure

… and … in Top analysis



Expectations based on:

- MC predictions for EW and top backgrounds
- Fake tags (mistag matrix) applied to light jets
- CDF Method 2 admixture of MC and data driven procedure

σ_{tt} = 8.2 ± 0.5 ± 1.0 pb

… how is this expectation obtained? →

→ HF estimates: CDF method 2

- Method 2 = Heavy flavor background estimate in the b-tagged W + jets sample
  - Signal = *tt*, single top, WH

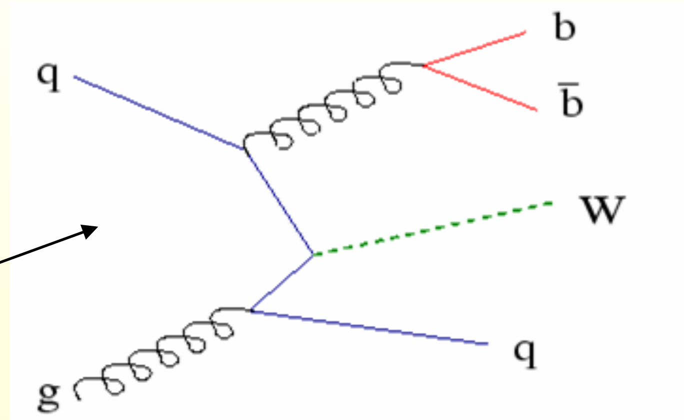
- **Main backgrounds**
  - W + HF (W+*bb*, W+*cc*, W+*c*)

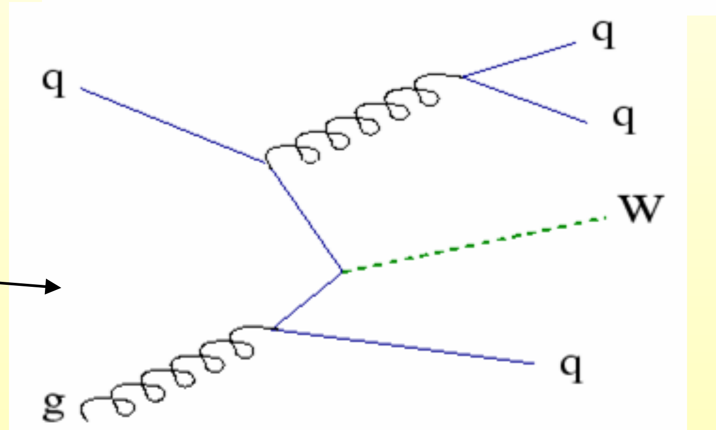
MC (HF fractions) and Data

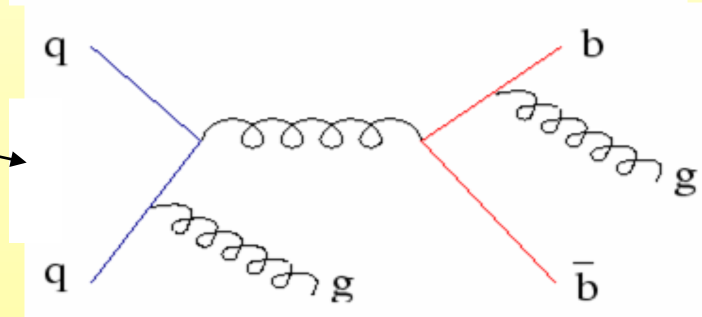
Data based

Data based nds (diboson, Z+jets, etc.)

MC based



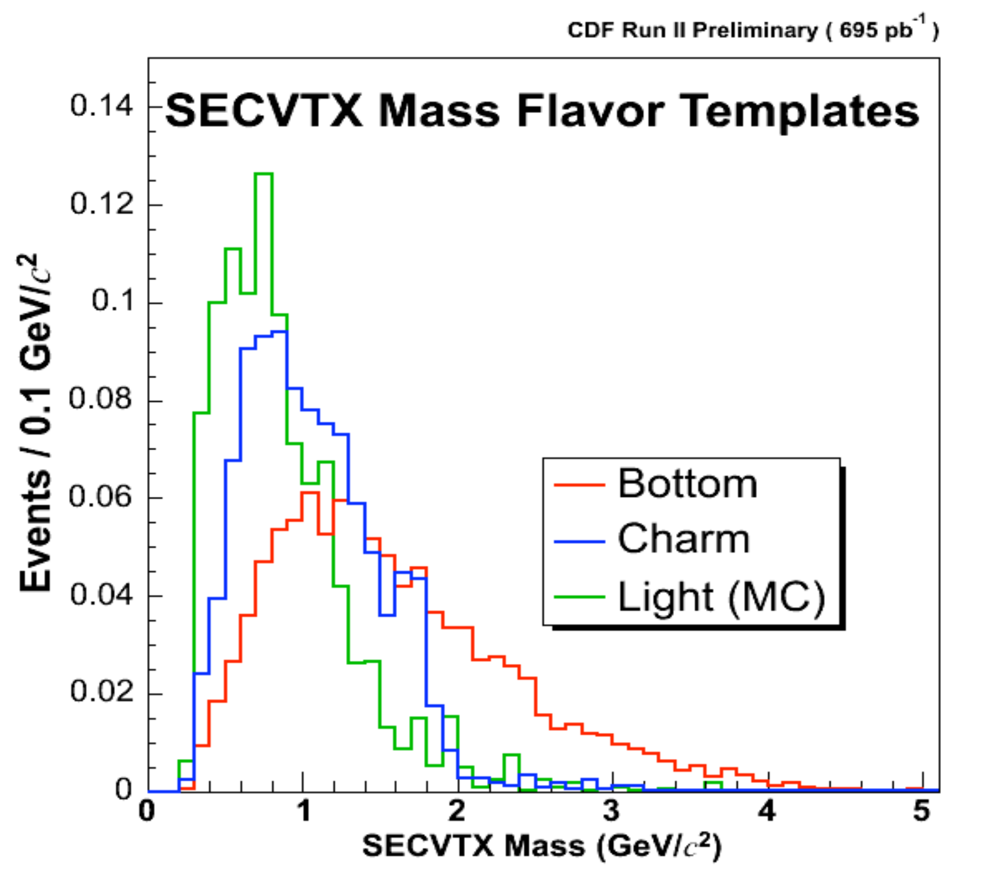


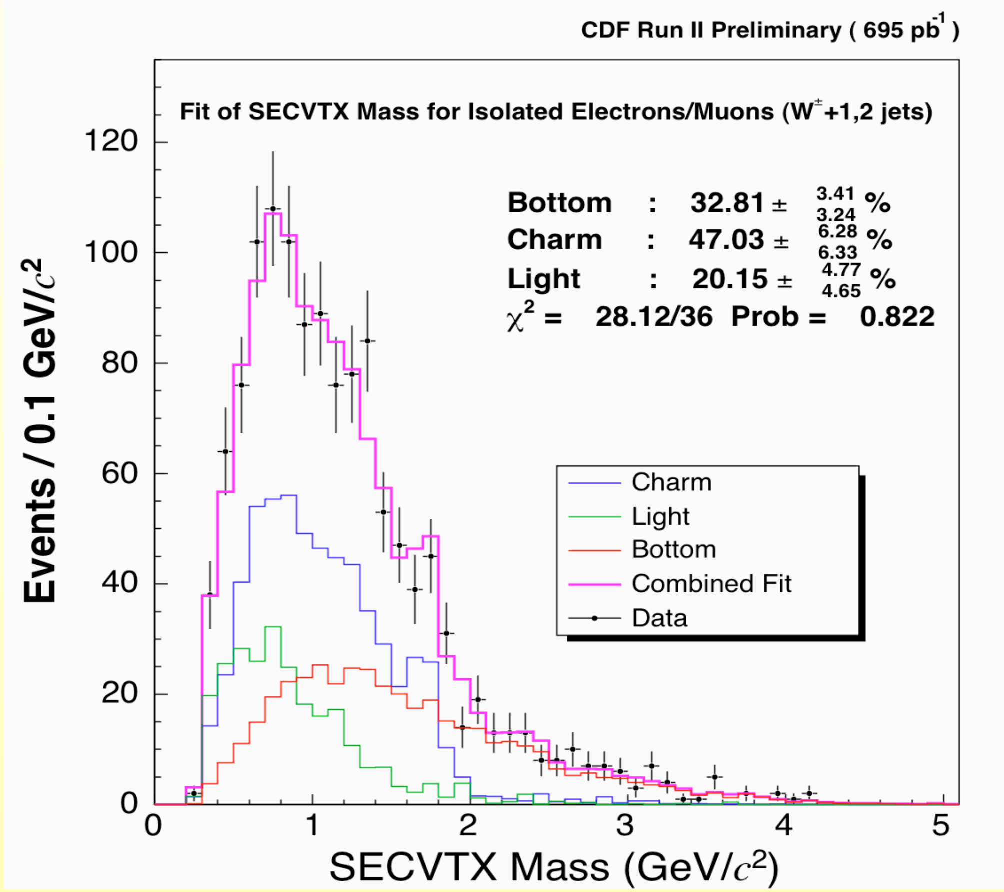


- *At what level do we know W/Z + HF production ?*
- *Present measurements at Tevatron*
- *On going studies and perpectives*

W + bb production results

In secondary-vertex-tagged sample, fit for light, c, b contributions

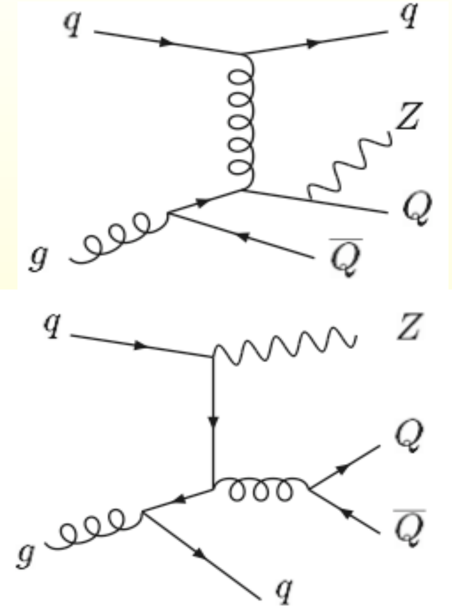


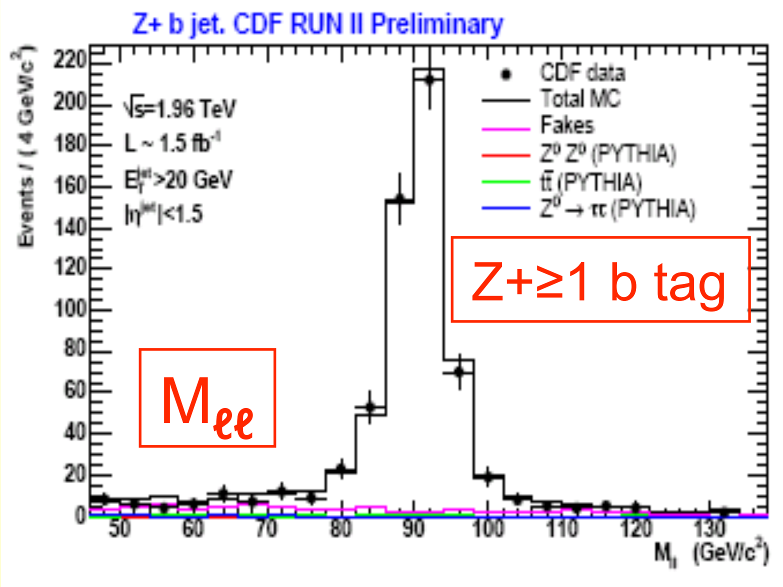


measure: 0.90±0.32 pb for Wbb (E_T(j)>20 GeV, $|\eta|$<2.0

LO calculation (ALPGEN): 0.74±0.18 pb

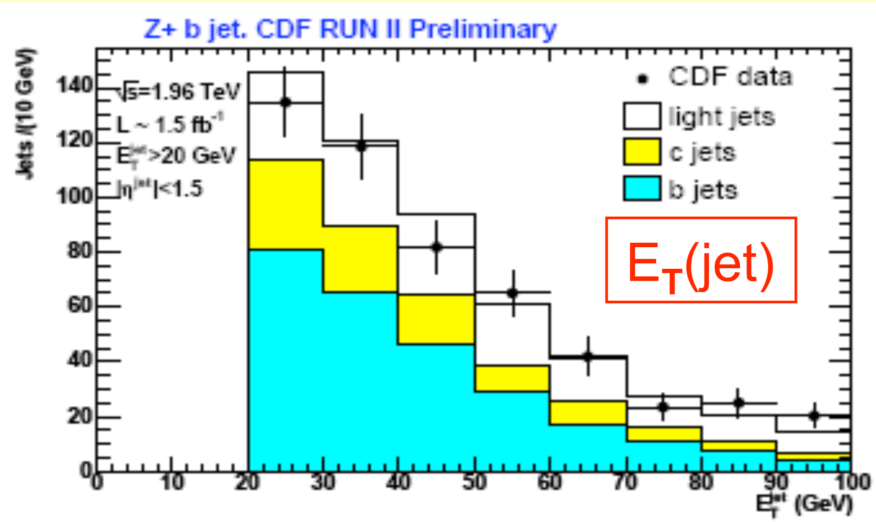
Z+b jets analysis

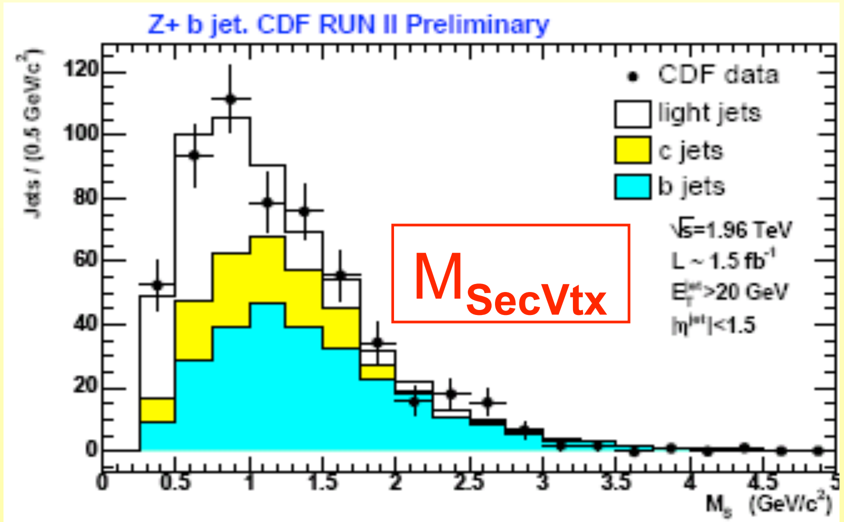




Ask for the Jet to have a Secondary Vertex Tag and

…exploit light,charm to b separation from sec.vtx. Mass





Z + HF x-section from Tevatron to LHC

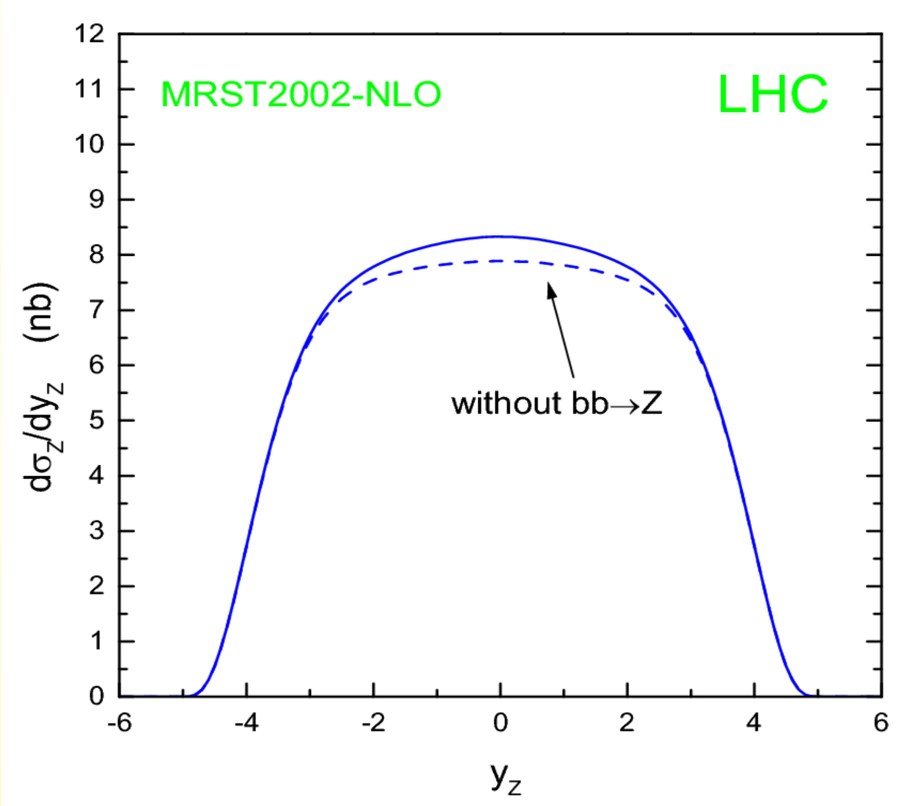
from hep-ph/0312024 - Tevatron : $p_T \geq 15$ *GeV,* $|\eta| < 2$, $\Delta R_{jj} > 0.7$

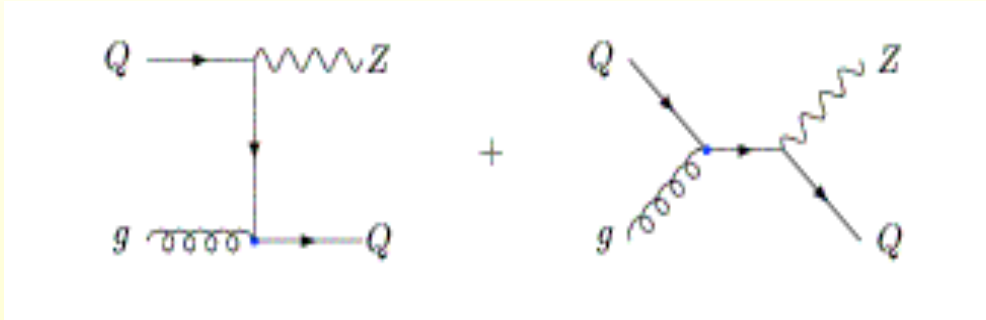
Tevatron

Production (pb)	ZQ	$Z(Q\bar{Q})$	ZQj	$ZQ\bar{Q}$	Zj	Zjj
$gb \to Zb$	10.4	0.169	2.19	0.631	-	-
$q\bar{q} \to Zb\bar{b}$	3.32	1.92	-	1.59	-	-
$gc \to Zc$	16.5	0.130	3.22	0.49	-	-
$q\bar{q} \to Zc\bar{c}$	5.66	6.45	-	1.70	-	-
$q\bar{q} \to Zg, gq \to Zq$	-	-	-	-	870	137

LHC
~1000
~50
~1400
~90
~16000

LHC example: b quark PDF from Z+b





b$\bar{b}$->Z @ LHC is ~5% of entire Z production

Spread of existing pdf's gives a 10% spread in the prediction of the SM Higgs cross section

Even bigger impact on MSSM Higgs production (sensitive to bb coupling)

- *New generation of Jet Flavor tagger*

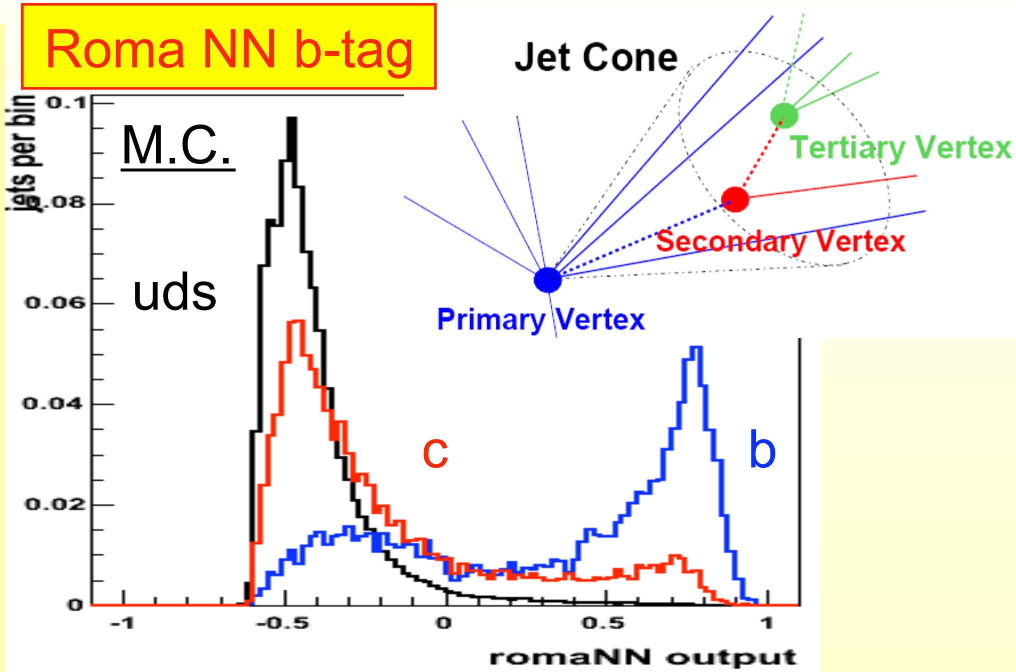
- *Highlights on first attempt to measure directly:*
  Z+b , Z+c , Z+uds

- *go beyond method 2 like analysis*

A new CDF H.F. Neural Network Tagger

- The underlying idea is improving the ability to boost performance exploiting *as much information as possible*:
  - number of vertices (not just ***A SINGLE*** secondary vertex a' la SecVtx) and their masses;
  - displaced but un-vertexed tracks (tracks not belonging neither to secondary vertex nor to the primary);
  - Soft leptons, JetProb, global jet variables
- Best to combine all info together in a single discriminant: use (a series of) *Neural Networks* (NN)
- NN training performed on generic sample to be used in different analysis (top, single top, low mass Higgs, …)

Roma NN b-tag

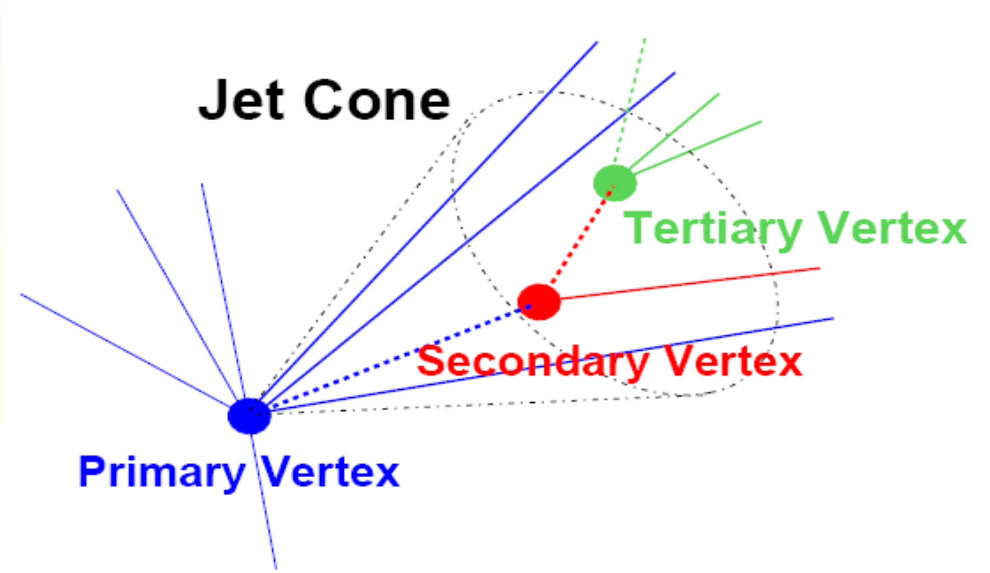


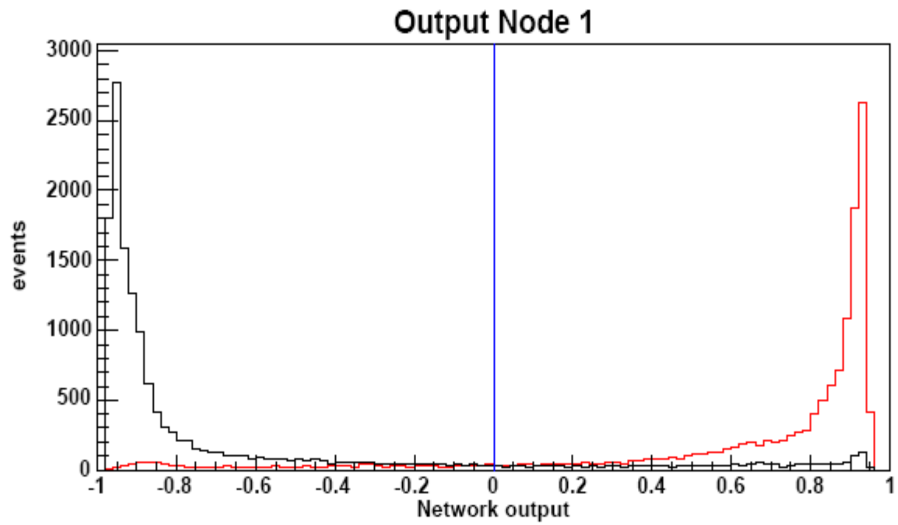
- **A *Continuous* tagging** variable allow weighting events to get the most of it in terms of search sensitivity
- **GOOD *b-c-uds statistical separation*** allows the fraction of c jet to be studied as well, HIGH Jet taggability: ~ 75%
- **Increased *per-jet efficiency*** at same background rate than SecVtx:

1st step: new vertexing algorithm

Uses all tracks in the jet to try to find multiple vertices: primary+all secondaries.No cut on decay lenght as usual b-tags.

- b-hadrons often produces more than one distinct vertex
- Jet containing two b-hadrons have even more

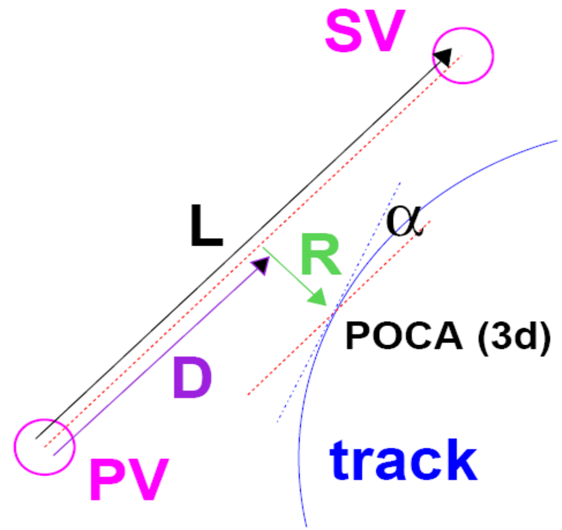


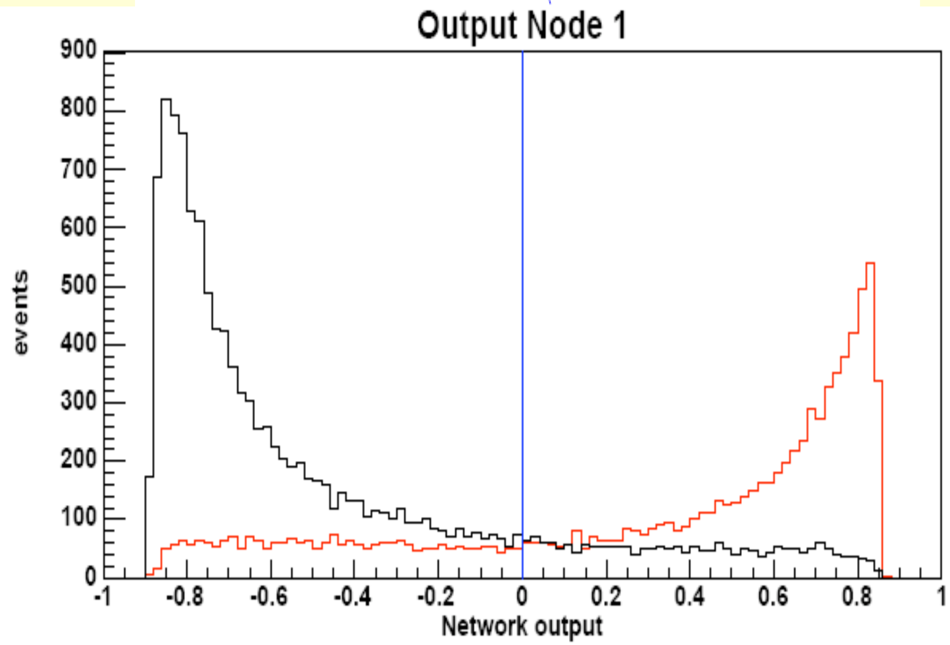


- ➢ Tracks are unambiguously assigned to "closest" vertex based on their χ^2
- ➢ Iterative procedure ends when no more track can be associated to any vertex
- ➢ List of vertices found and associated tracks passed to next stage of NN (un-vertexed tracks and HF discrimination)

2^{nd} step: track NN

- Selects un-vertexed tracks from HF decays
- Select 5 most significant variables and train network to distinguish:
- good track-vertex combinations (true HF decay track and true HF vertex decay)
- from other combinations (fake track-fake vertex,fake-signal etc.)

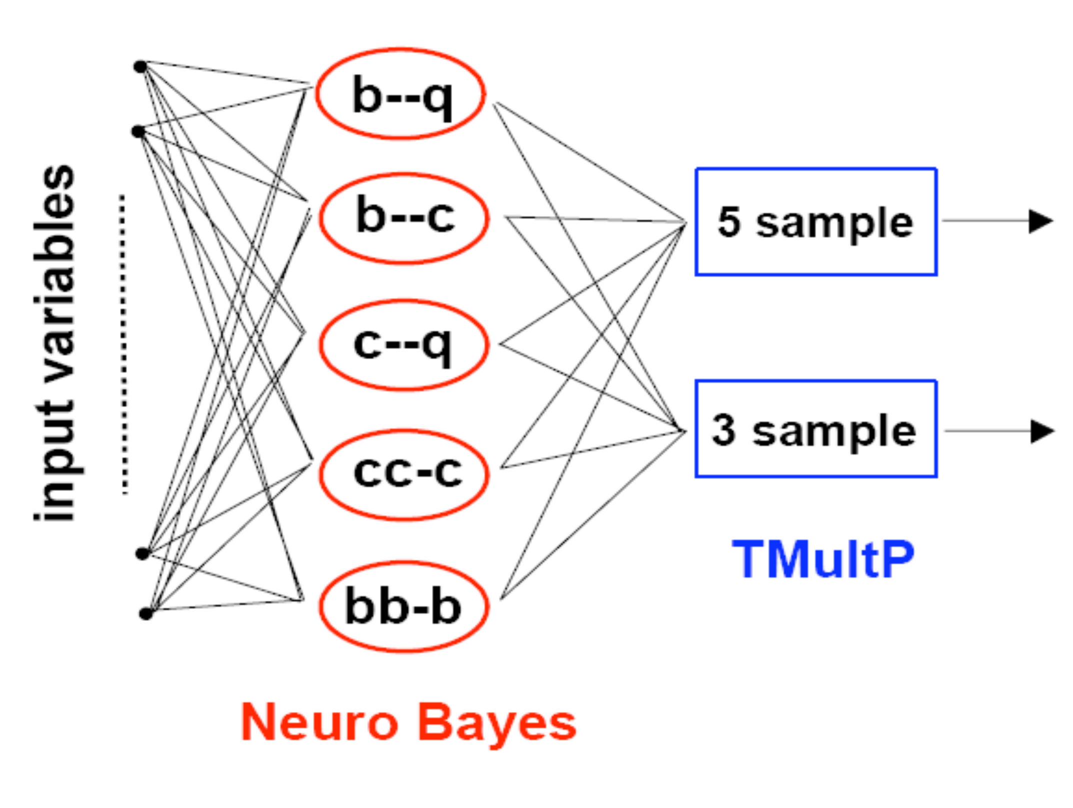




3^{rd} step: Jet NN

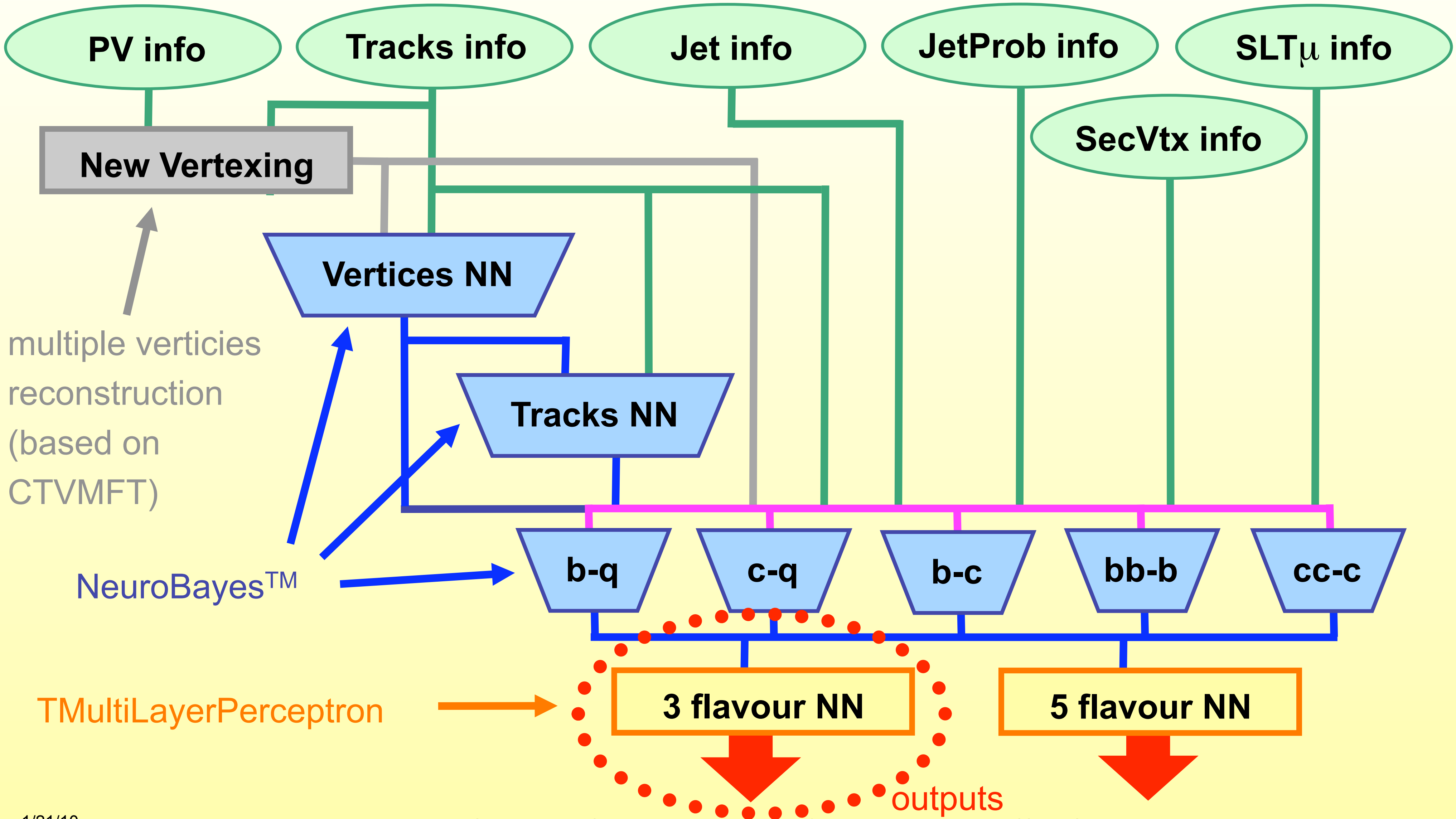
- Combine tracks and vertices NN outputs, muons, SecVtx, JetProb, etc.

- 5 different NN's specialized to separate two by two light, c, b, ccbar, bbbar jets



- Combine the 5 outputs to give a 5 flavour separation (light, c, ccbar, b, bbbar) or 3 flavour (light, c, b) in a single tagging variable
- Present studies just on the simpler **3 flavors NN final output**

NN tagger structure

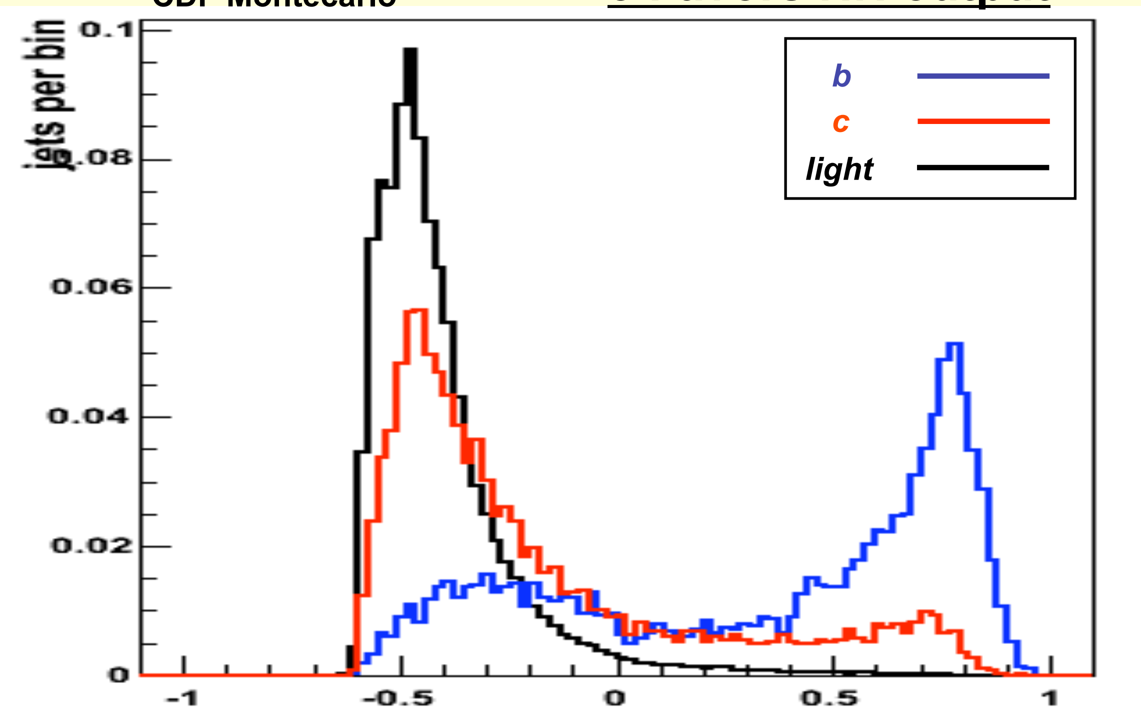


Roma NN final output

- **A** ***Continuous*** **tagging** variable allow weighting events to get the most of it in terms of search sensitivity
- **GOOD** ***b-c-uds statistical separation*** allows the fraction of c jet to be more reliably studied.
- **HIGH** ***Jet taggability***: ~ 75%

CDF Montecarlo

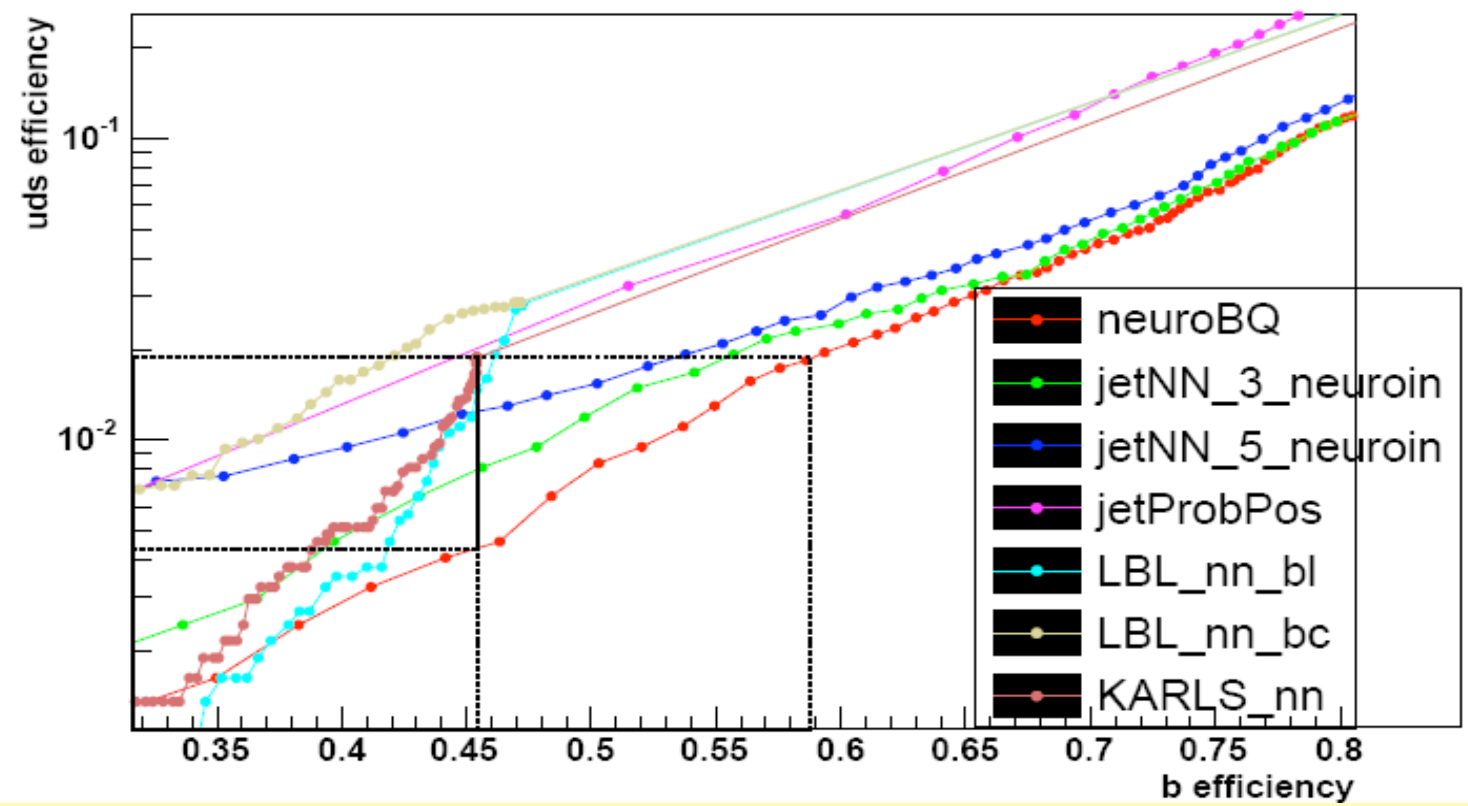
3 flavors NN output



Advantages

- Work in pre-tag sample (Roma NN Jet taggability ~75%)
- Classify jets according to purity
- More efficiency at same background rate than SecVtx

- Similar or better performances than Karlshrue and Berkeley Neural Networks (proved on MC)

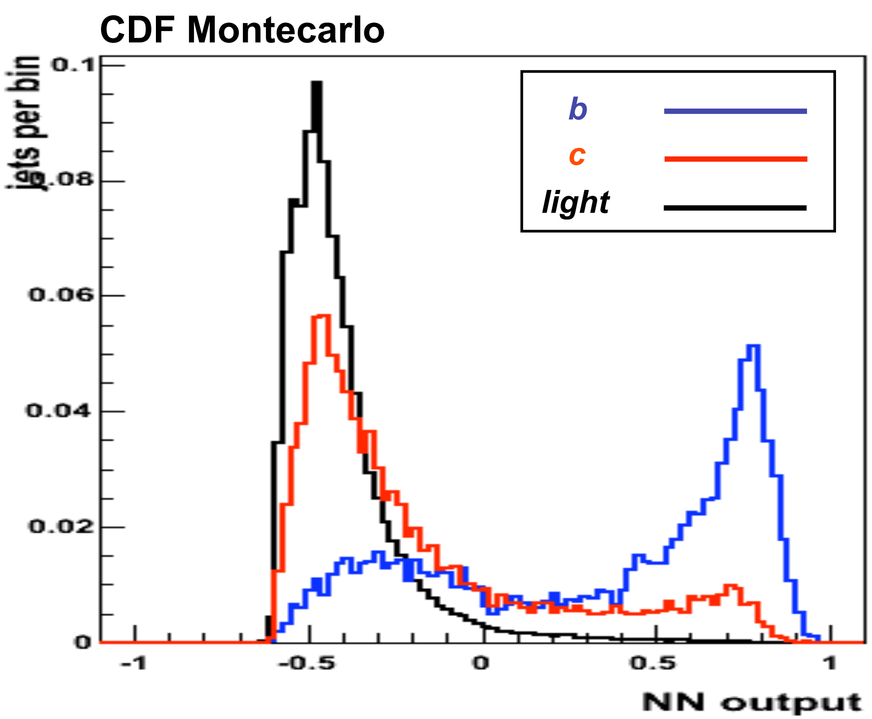


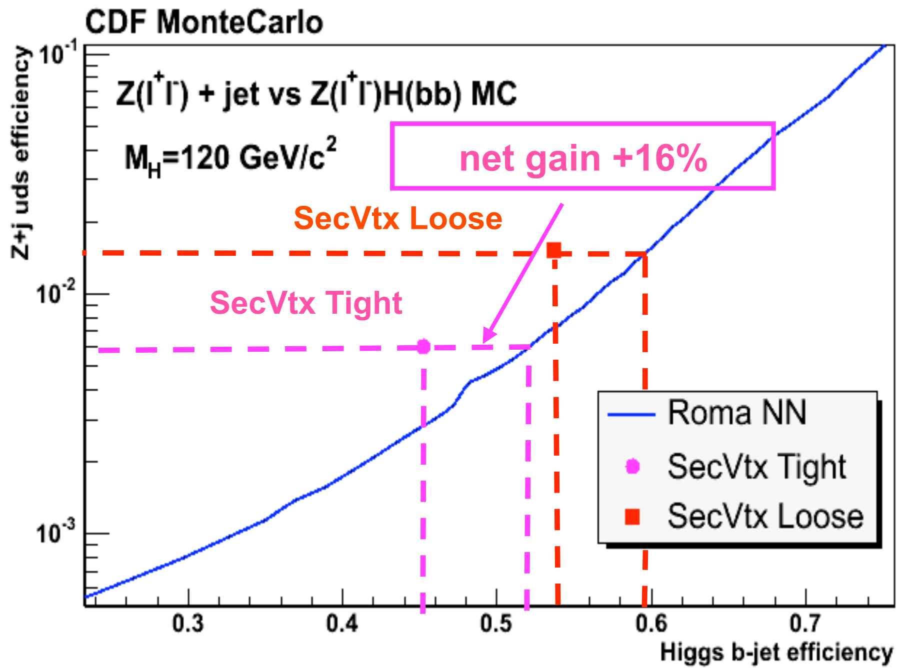
Advantages

- **Increased** ***per-jet efficiency*** at same background rate than SecVtx:

  +30% (relative) on a p-pbar → q-qbar MC di-jet (Jet 20 - btopqb)

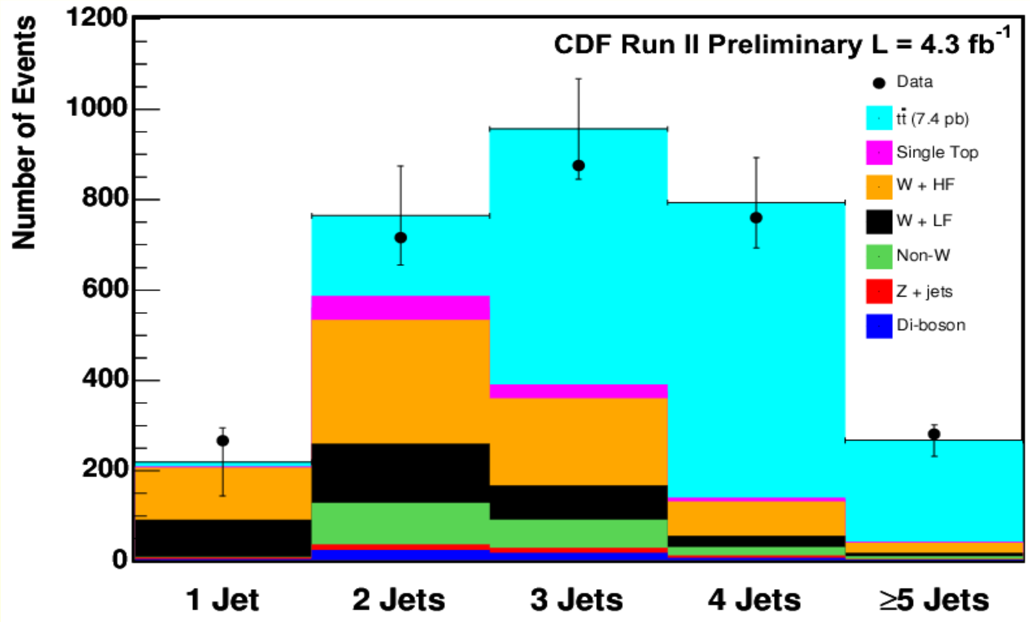
  +16% (relative) for Z+j compared to ZH(120) MC

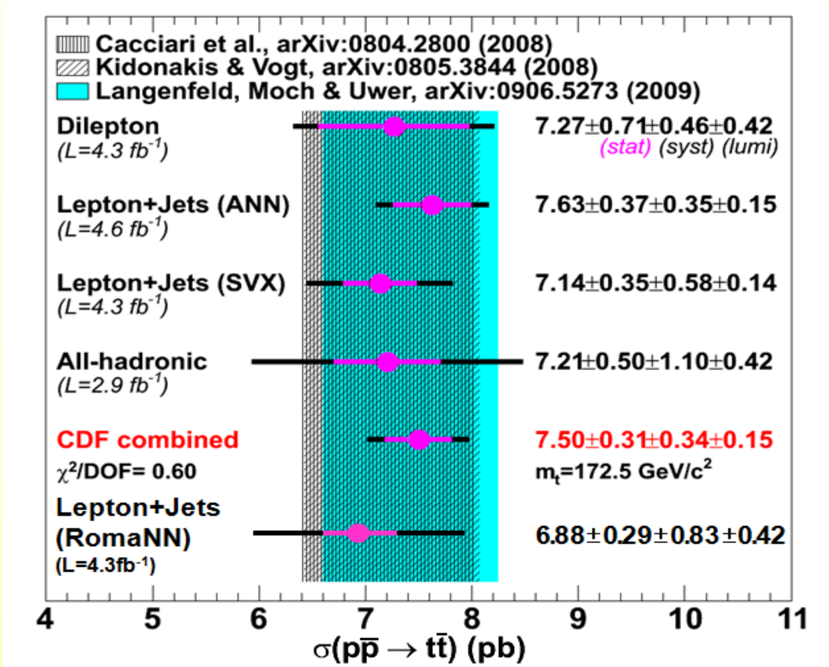




Top x-section in Lepton+Jets with cut based Roma NN b-tag

New !





- The cross section of pair produced top quarks has been measured to be: $6.88 \pm 0.29_{stat} \pm 0.83_{sys} \pm 0.42_{lumi}$
- using 4.3 fb^{-1} of collected data from the high Pt lepton triggers. The measurement is performed with TightRomaNN tagged lepton plus jets events with >= 3 tight jets, Ht >= 230 GeV, and MET >= 20 GeV. This measurement has taken the top mass to be 172.5GeV/c^2.

Higgs search in WH → ln bb with RomaNN in the pipeline …

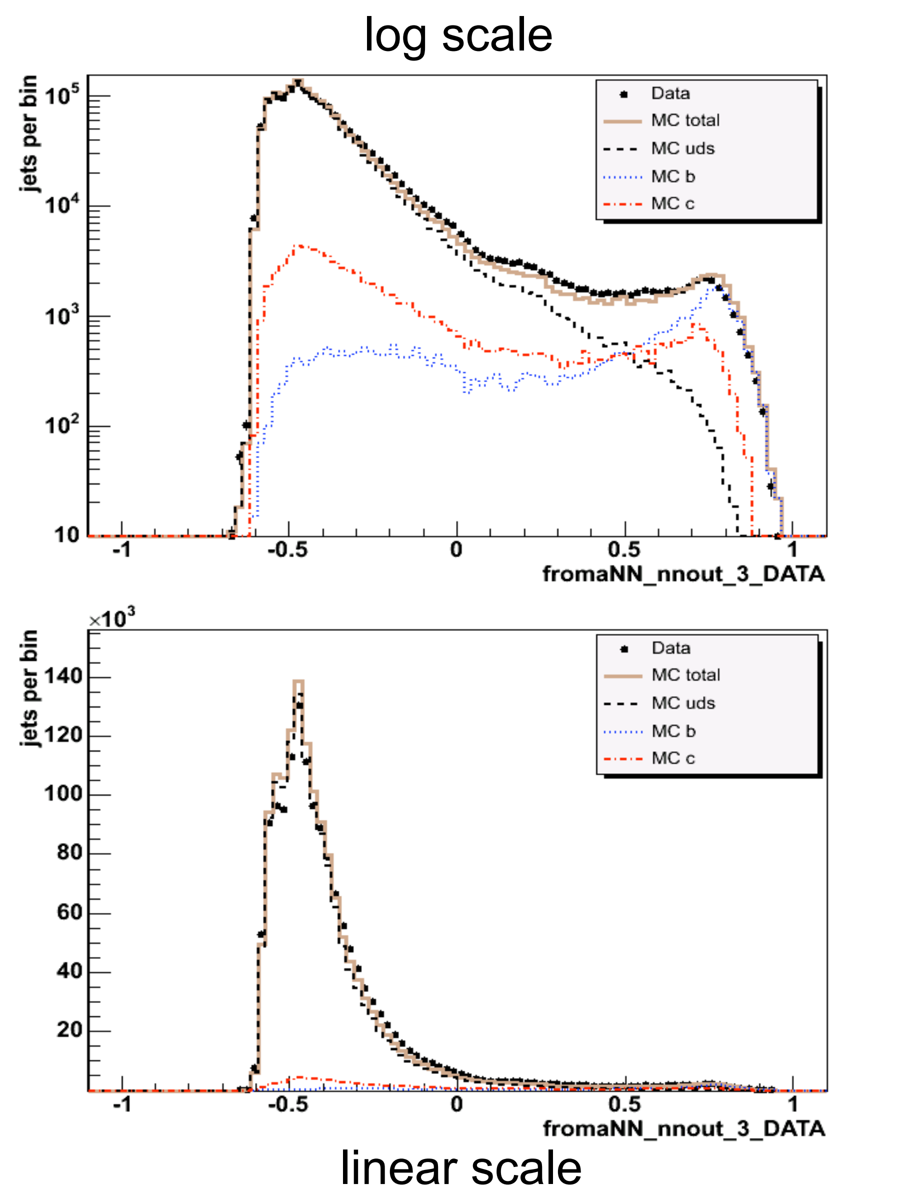
Tuning NN output to data (aka calibration)

- Iiterative procedure to extract ***correction functions*** (cf) = weights for b, c and light MC tagger *templates*

- *Data driven* extraction of the cf's exploiting the high statistic JET20, JET50 data samples and the corresponding di-jet MC samples

- Defined 3 independent and unbiased jet samples with different HF content but otherwise identical jet properties
  - each flavor has the same shape in all the 3 samples

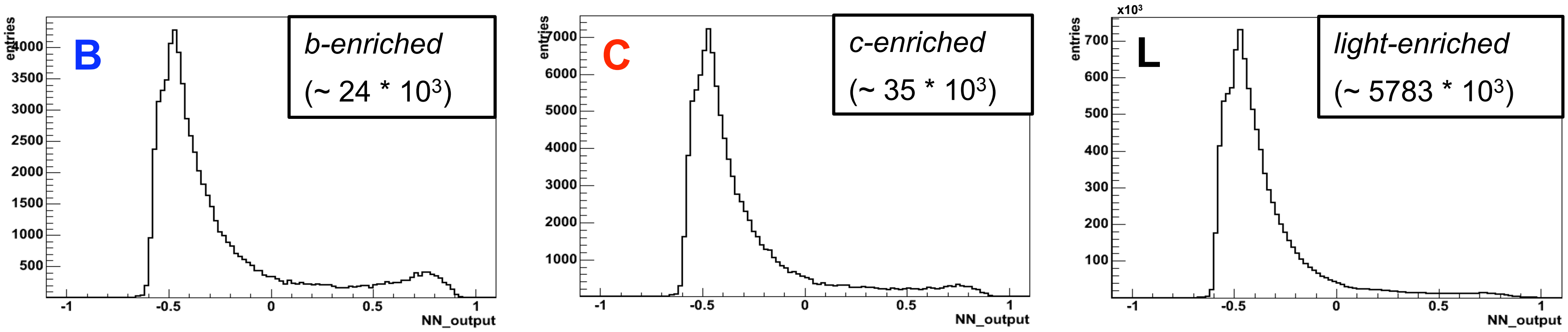
- Correct for Data/MC disagreement & measure HF content at the same time



MC(Pythia) predicted b/c/lightfractions

Calibration data-set

- Define 3 independent flavor enriched samples selecting di-jet events and asking for an *away jet*:
  - SecVtx tag and b/c LBL NN > 0.6 *b-enriched* (**B**)
  - SecVtx tag and b/c LBL NN < 0.6 *c-enriched* (**C**)
  - SecVtx untagged *light-enriched* (**L**)

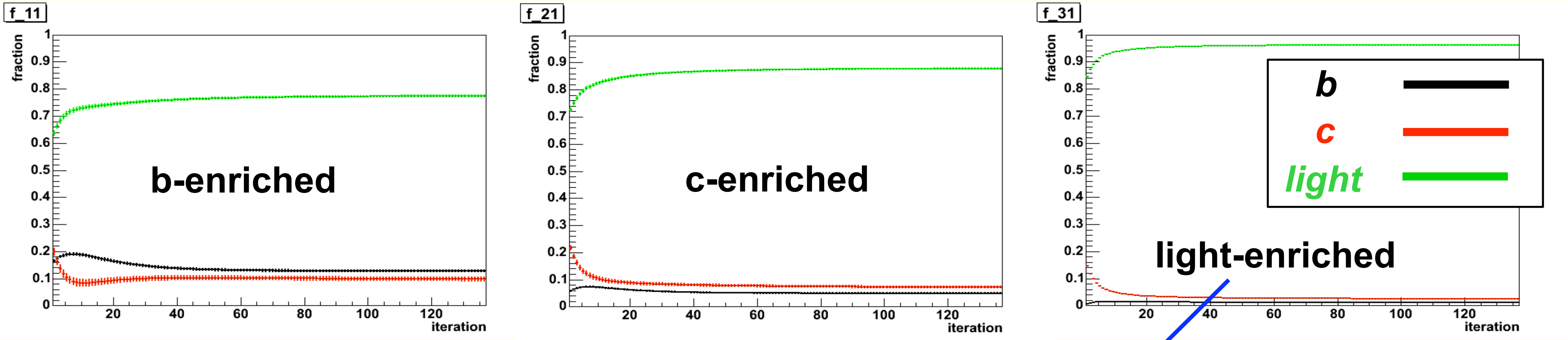


MC prevision

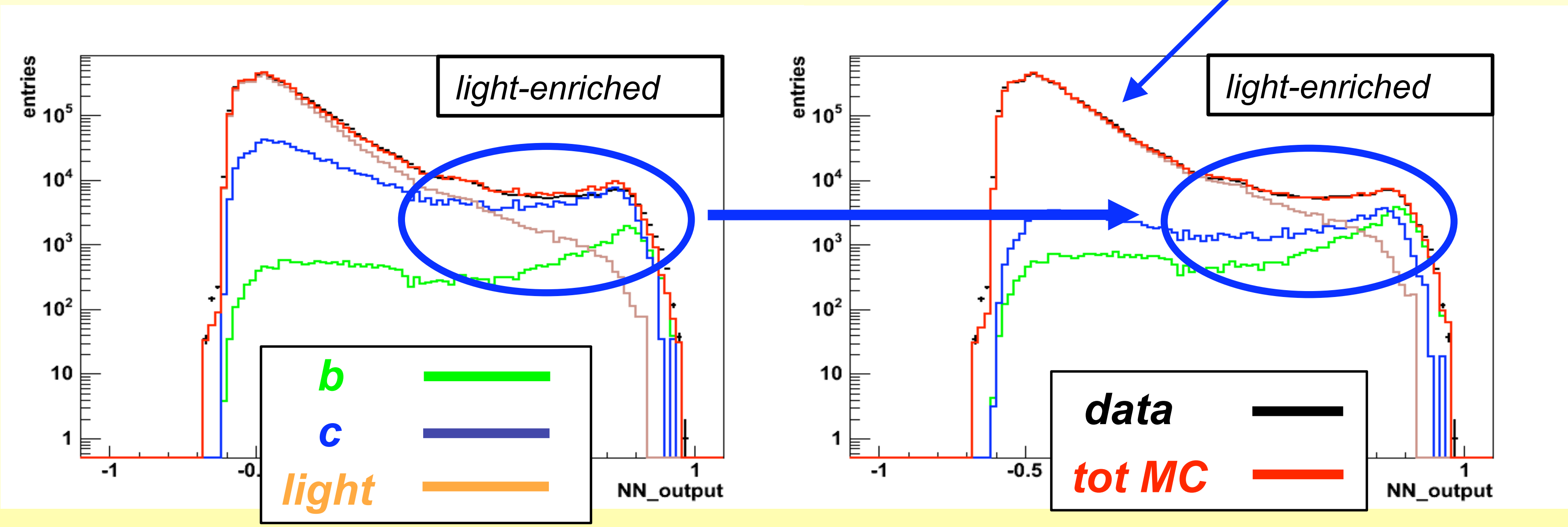
Sample	b-jets	c-jets	light-jets
B	0.234 ± 0.009	0.058 ± 0.004	0.708
C	0.096 ± 0.005	0.092 ± 0.004	0.812
L	0.022	0.052	0.926

Templates calibration with Jet DATA

Fraction evolution as a function of the iteration.



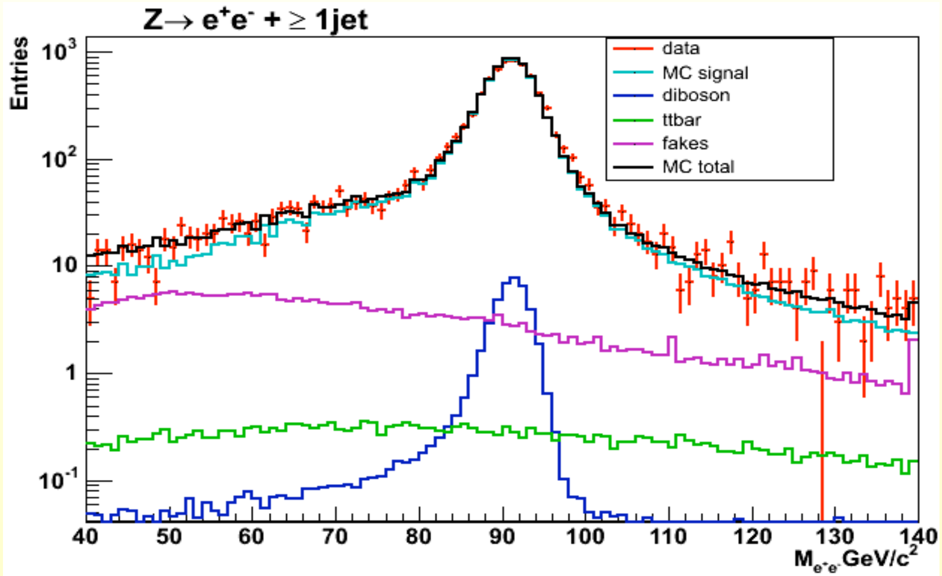
***before** correction*



***after** correction*

Roma NN full application: Z+HF

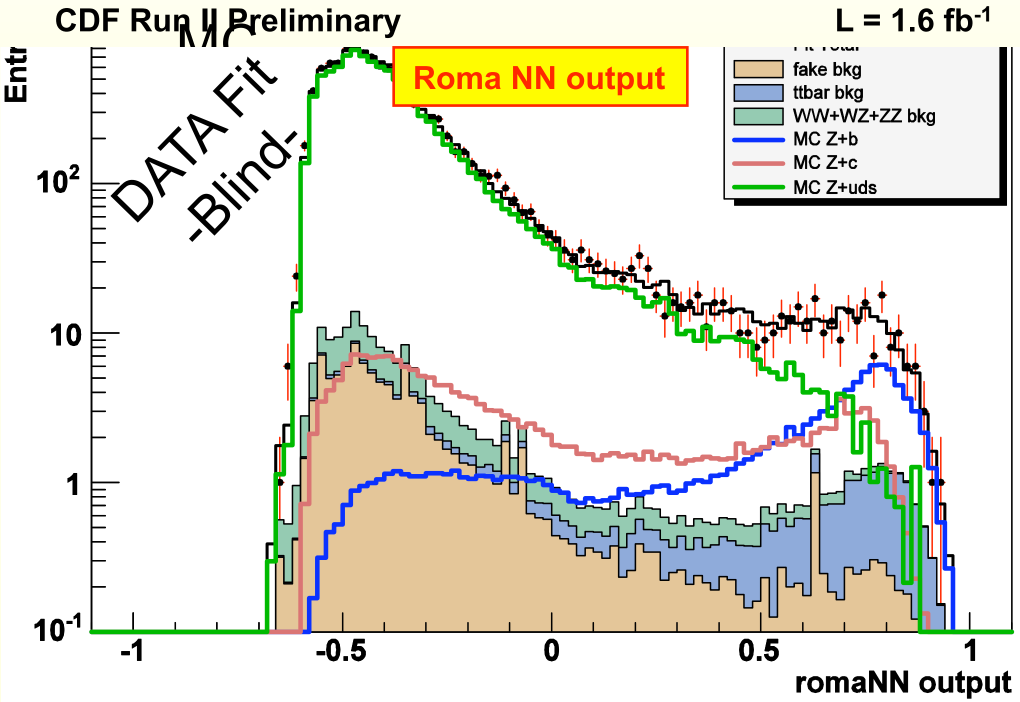
- Ideal play-ground to test the new Roma NN tagger
  - Easy signature
  - Low jet multiplicity
  - Analysis based on 1.1 fb^{-1}



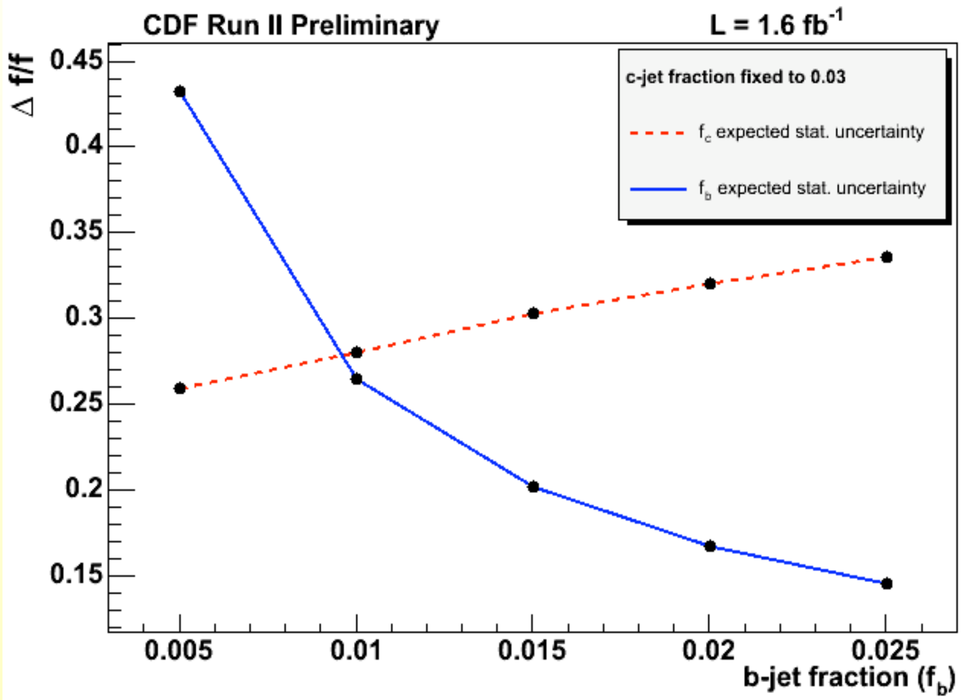
- Interesting measurement by itself: QCD generators, PDF's…
- Major background to ZH channel Higgs search
- Understand how to extrapolate corrected templates from generic di-jet samples to Z+jets signal sample:
  - STRATEGY: re-weight di-jet data to match E_T spectra in the target sample (eventually re-weight for other event related observables)
  - Test case for other extrapolations: to top, Higgs,…

Z+HF fractions with Roma NN b-tagging

Roma NN b-tag output fit in Z+jets events, L= 1.6 fb^{-1}



Toy MC estimate on b, c fractions uncertainty



- Could allow for the first time a simultaneous measurement of **Z+c** and **Z+b** cross sections.
- Application of the Roma NN tagger to other samples: one could fit top/higgs cross section together with the W/Z+HF background as a function of jet multiplicity.

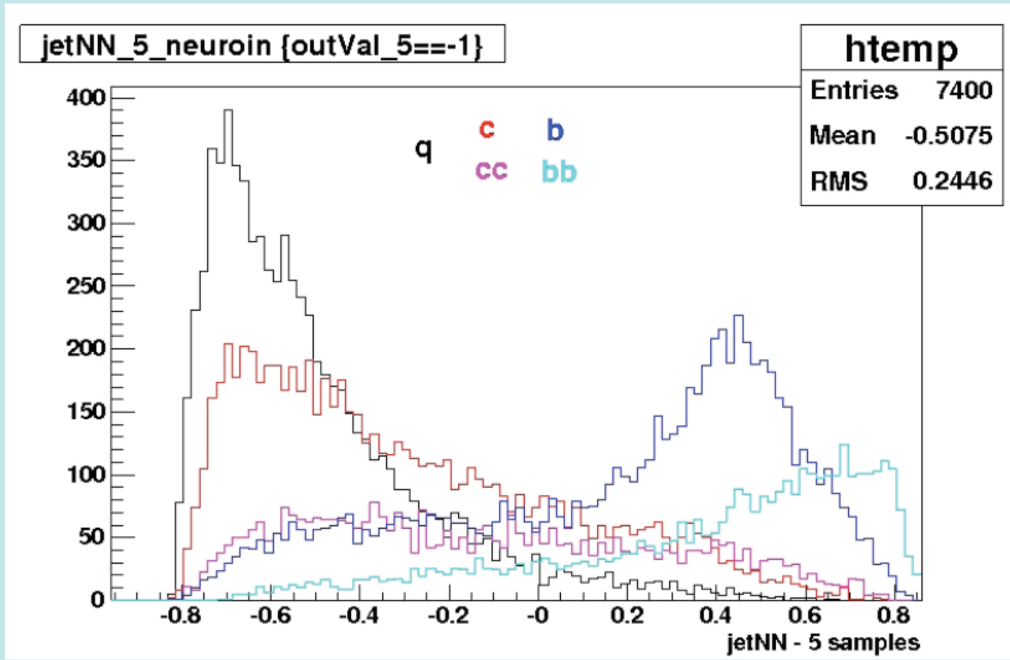
New generation Jet Flavor Tagger, Outlook

- The whole concept of Method II could be revisited also with the use of a powerfull, high efficiency and well simulated b-tagging tool → change paradigma:

- No need of working out the tagged sample from the pretag sample and questionable HF fraction from MC

- In principle one could fit the top/ higgs cross section and the W+b/c component in each jet multiplicity bin taking in to account fake, and other EWK bkg! *… to be followed up … and discussed…*

***HF sample composition** for Higgs, Top and searches analysis is becoming more and more crucial*

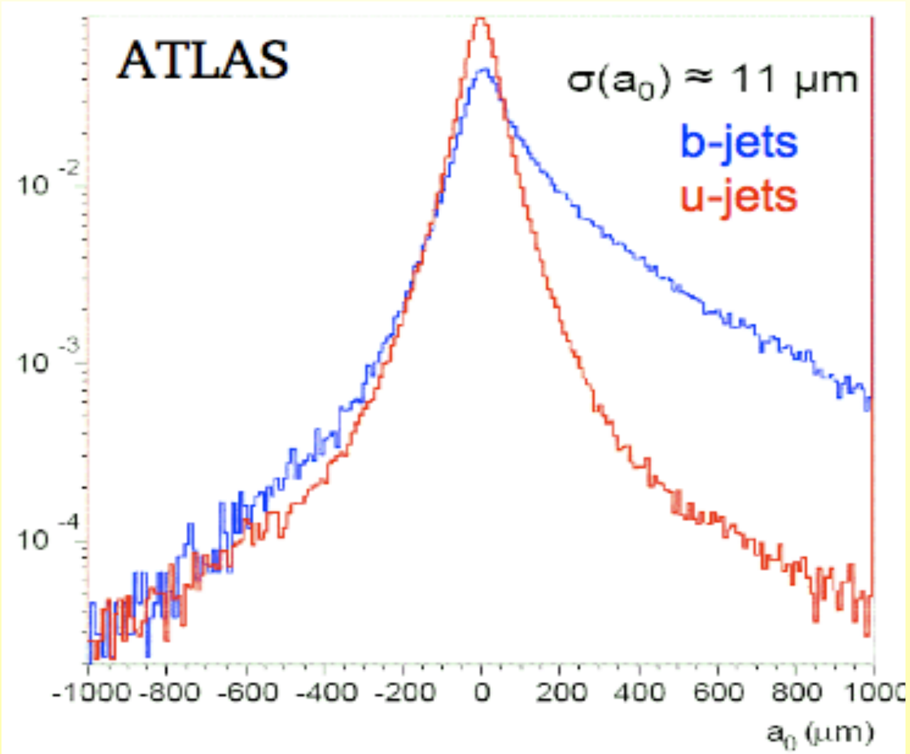


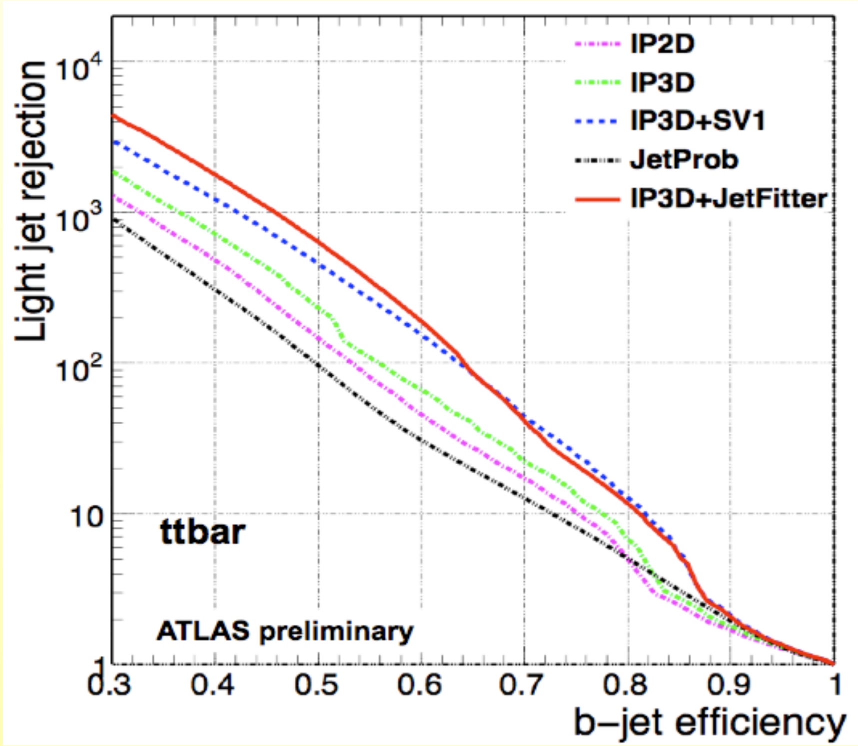
- *Further possibility: building block are there the to tackle also the b/bbar (c/ ccbar) jet issues*

- *a look to LHC now*

B-tagging in ATLAS

- Many studies has been made for soft-leptons, IP tags and multivariate
- Also b-Jet trigger at HLT level based on likelyhood
- For now concentrate on IP tag validation of performances, **on MonteCarlo:**

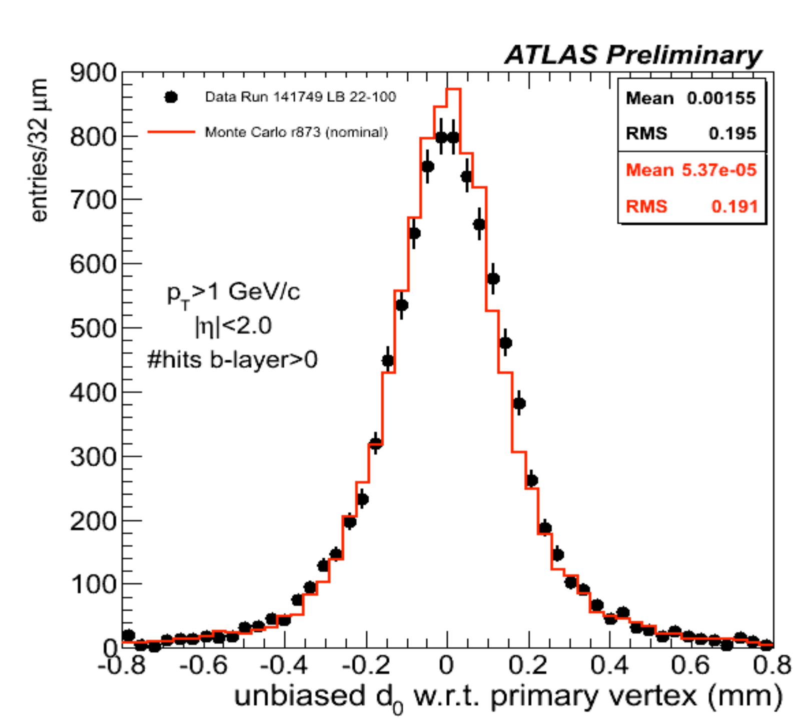




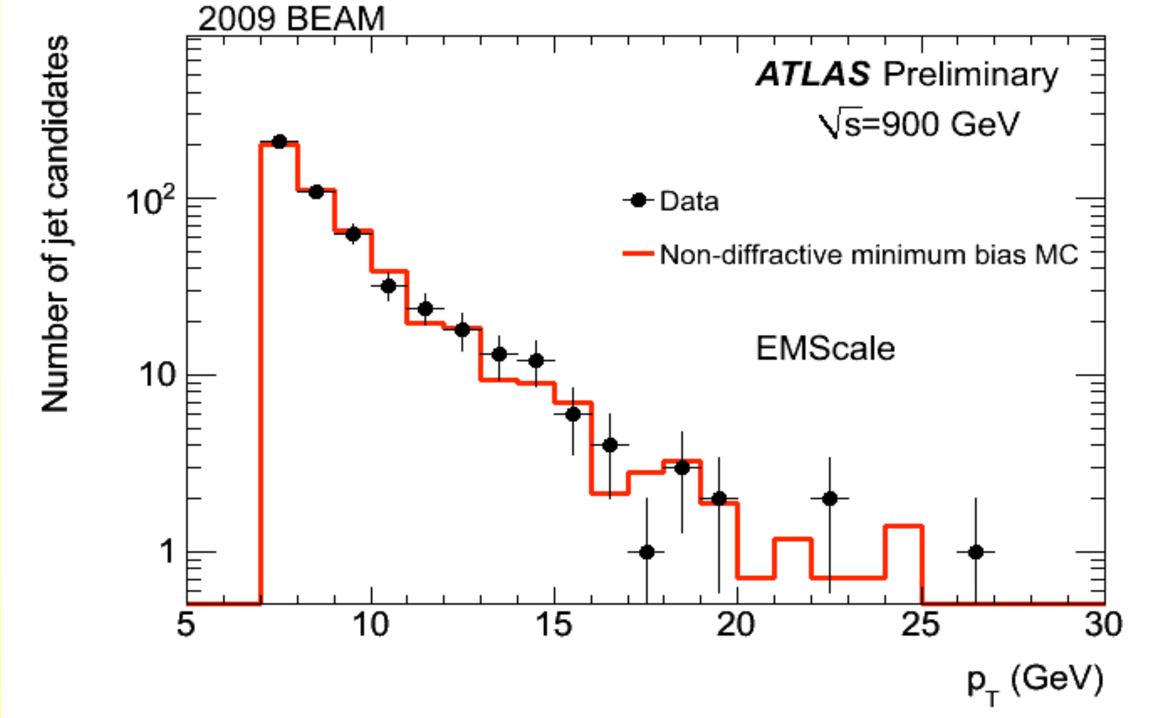
B-tagging in ATLAS

- For the first data concentrate on IP tag validation **… and on DATA:**

Tracks Impact parameter



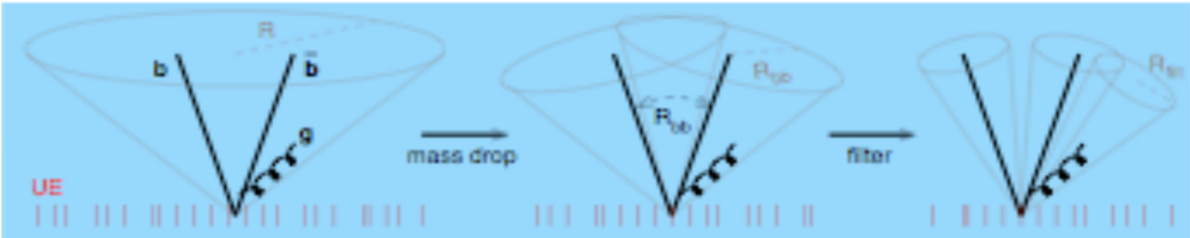
Anti-KT cone 0.4 Jets



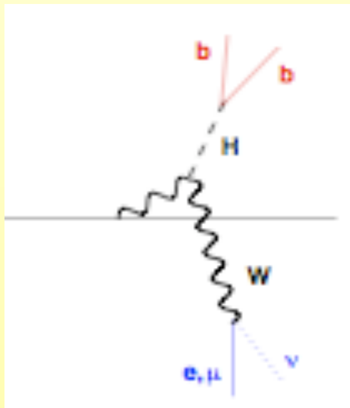
Future: boosted objects decaying in bb ?

At LHC one can have highly boosted sizeable fraction of heavy particles decaying in bb and merging in single reconstructed JET:

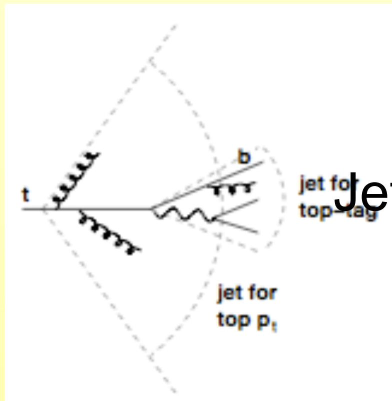
• try to explore Jet substructure, guided by b-tagging interest regions



Possible applications to:



WH with H→bb   or



Jet “ top – tagging”

Summary:

- Reviewed main b-tagging strategies and their use
- Highlight the relevance of b-tag control for background processes studies
- W/Z + HF processes not yet fully studied, now starting to characterize at the Tevatron.
- New generation NN Jet Flavor Tagger being developed at CDF and being tested in Z+b, Z+c measurement and used in Top and Higgs analysis
- *… lot of knowledge progresses on HF production and b-tag are an important Tevatron legacy to the LHC … a rich research ground for next years…*
- B-tagging is a fun and challenging topic
  - Design/calibration of taggers has a lot of physics in it
  - Not a solved problem – still room for good ideas

Thank you

Backup

Method 2 on One Slide

- Start from the W+jets data before tagging (pretag)
- Estimate MC-based backgrounds and Non-W

$$N_{MC} = \left(\sum_i \sigma_i \varepsilon_i \right) \int \Lambda dt \qquad N_{NonW} = f_{nonW} N_{Data}^{Pretag}$$

$$N_W = N_{Data}^{Pretag} - N_{nonW} - N_{MC} - N_{signal}$$

- Subtract to get pretag W yield

$$N_{W_{HF}}^{btag} = N_W F_{HF} \varepsilon_{btag}$$

- Multiply by heavy flavor fraction and tagging efficiency to get W+bb, W+cc, W+c

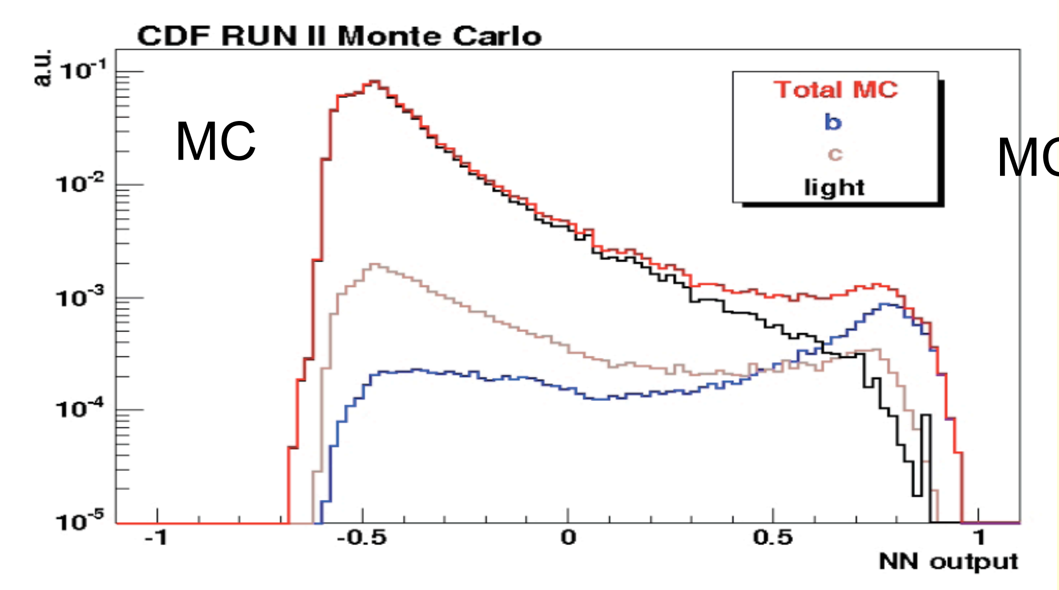
$$N_{W_{LF}}^{btag} = N_{Data}^{NegativeTag} \left(\frac{N_W - N_{W_{HF}}}{N_{Data}^{Pretag}} \right)$$

- Subtract W+HF from pretag and get W+light flavor from mistag matrix

$$N_{MC}^{btag} = \left(\sum_i \sigma_i \varepsilon_i \varepsilon_{btag} \right) \int \Lambda dt \qquad N_{NonW}^{btag} = f_{nonW} N_{Data}^{btag}$$

- Estimate tagged backgrounds from MC-based and non-W backgrounds

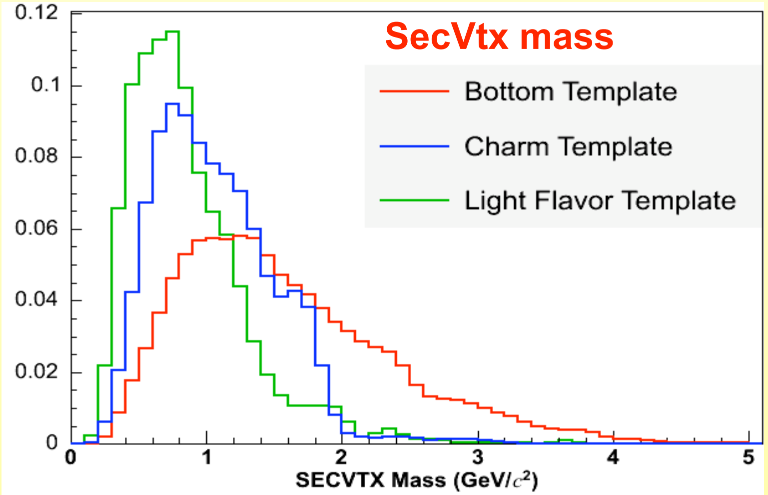
Z+HF fractions expected uncertainties



MC

Fitted fractions with statistical error (MC)	
b fraction	0.0200 ± 0.0036
c fraction	0.040 ± 0.014
light fraction	0.940 ± 0.011

- Used MC templates to evaluate expected uncertainty
- Generate 1000 experiments with 10000 events each with input fractions of b, c, light 2,4,94 [%]:
  - Allow a measurement of **Z+c** cross section with ~30% relative precision
  - Statistical separation for b -jets better than **SecVtx mass** by roughly 15% for equal luminosity



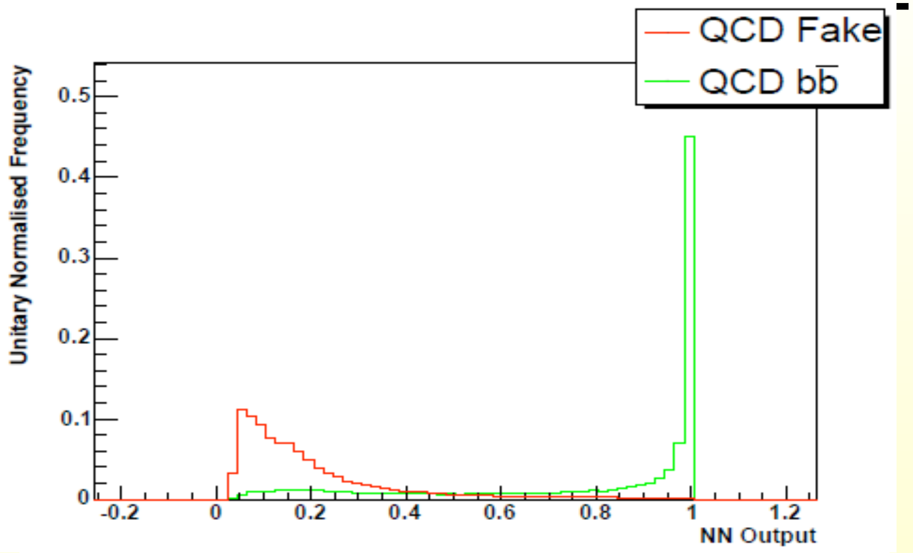
Multivariate Tagger (DØ)

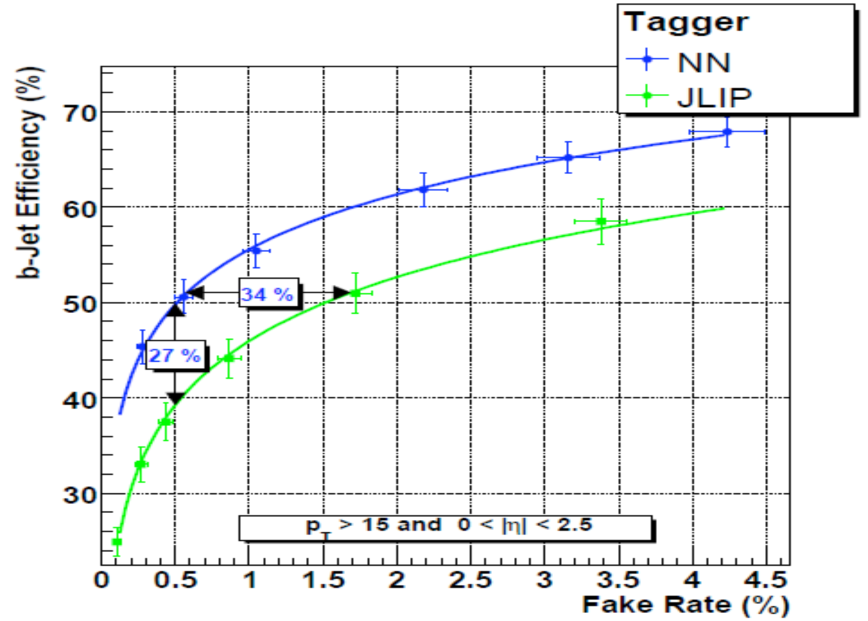
- 7-variable neural network
  - Vertex mass, L_{xy}/σ, χ^2
  - Number of vertices/jet
  - Jet total and displaced track multiplicities
  - Jet probability
- Vertex tagger uses a "super-loose" tune to get info on more jets

- Train to separate b-vs-light

- Can improve efficiency or mistag by 25%, up to a factor of two in the efficiency/purity extremes

- Efficiency and fake rates have been measured in data
- Ready for analysis use



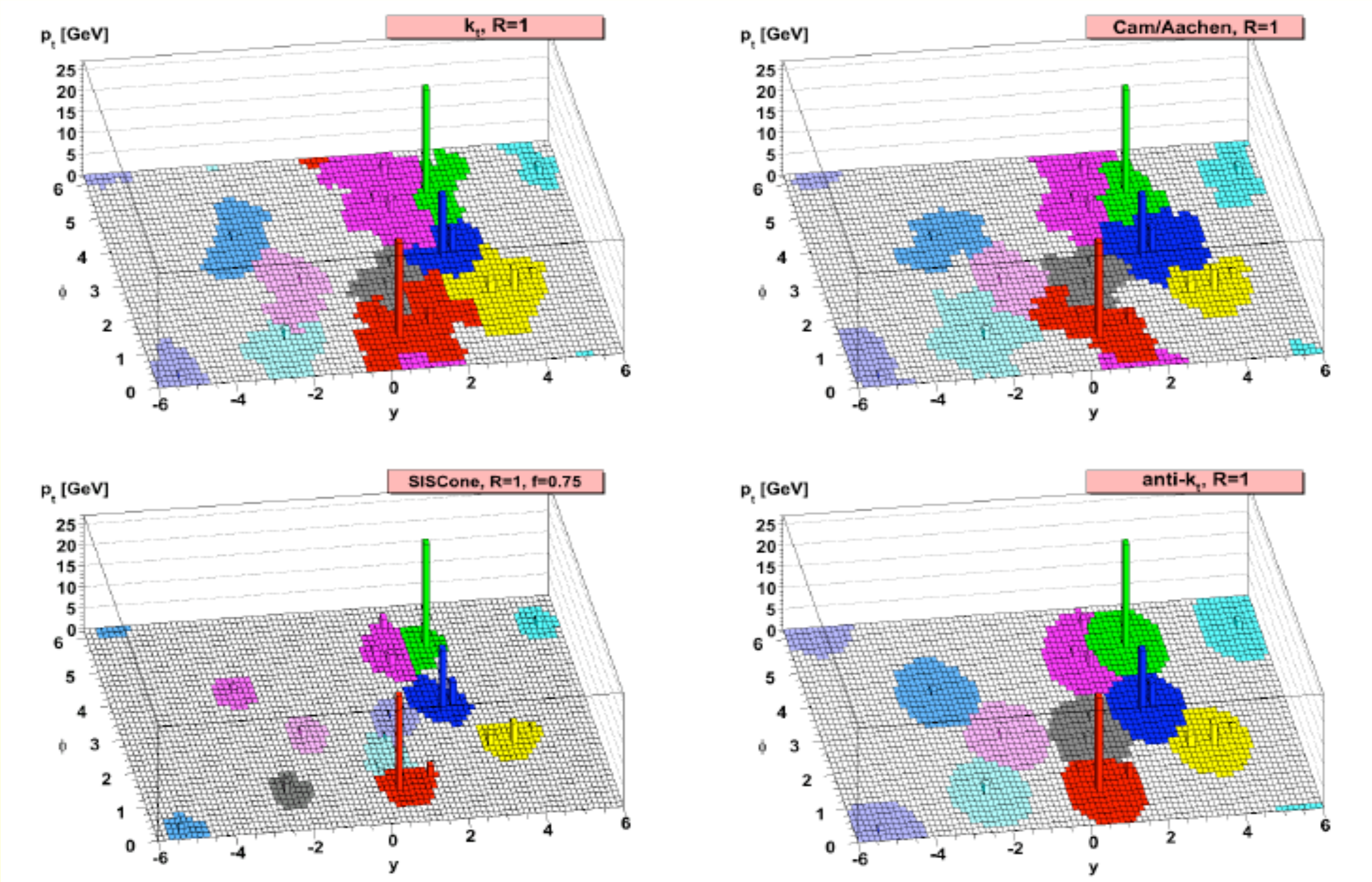


Set of Jet Algorithms

k_t SR, $d_{ij} = \min(k_{ti}^2, k_{tj}^2)\Delta R_{ij}^2/R^2$ hierarchical in rel $\perp$ momenta	Cambridge/Aachen SR, $d_{ij} = \Delta R_{ij}^2/R^2$ hierarchical in angle
anti-k_t SR, $d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2})\Delta R_{ij}^2/R^2$ gives perfectly conical jets	SISCone Seedless Infrared Safe cone +SM gives "economical" jets

This 3 Algorithms are studied in order to understand their performances under different physics processes and detector´s conditions.

Jet Areas



A sample parton-level event, together with many random soft “ghosts”, clustered with four different jets algorithms, illustrating the “active” catchment areas of the resulting hard jets. For Kt and Cam/Aachen the detailed shapes are in part determined by the specific set of ghosts used, and change when the gosts are modified.

Pileup subtraction using Jet areas

Basic Procedure:

- Use p_t/A from majority of jets (pileup jets) to get level, ρ, of pileup and UE in event
- Subtract pileup from hard jets:

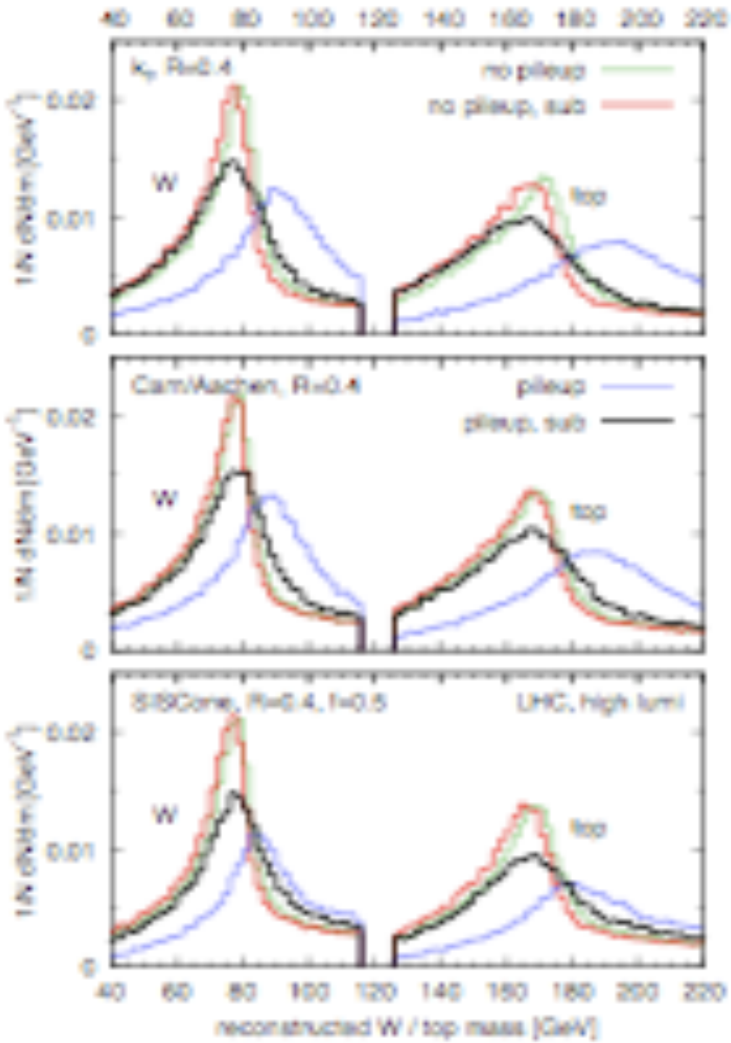
$$p_t \rightarrow p_{t,sub} = p_t - A\rho$$

Cacciari & GPS '07

Illustration:

- semi-leptonic $t\bar{t}$ production at LHC
- high-lumi pileup ($\sim$ 20 ev/bunch-X)

Same simple procedure works for a range of algorithms



Invariant mass vs Jet size

$$Q^{1/f}_{w=x\sqrt{M}} \equiv \left(\frac{\text{Max \# reco. massive objects in window of width } w = x\sqrt{M}}{\text{Total \# generated massive objects}} \right)^{-1}$$

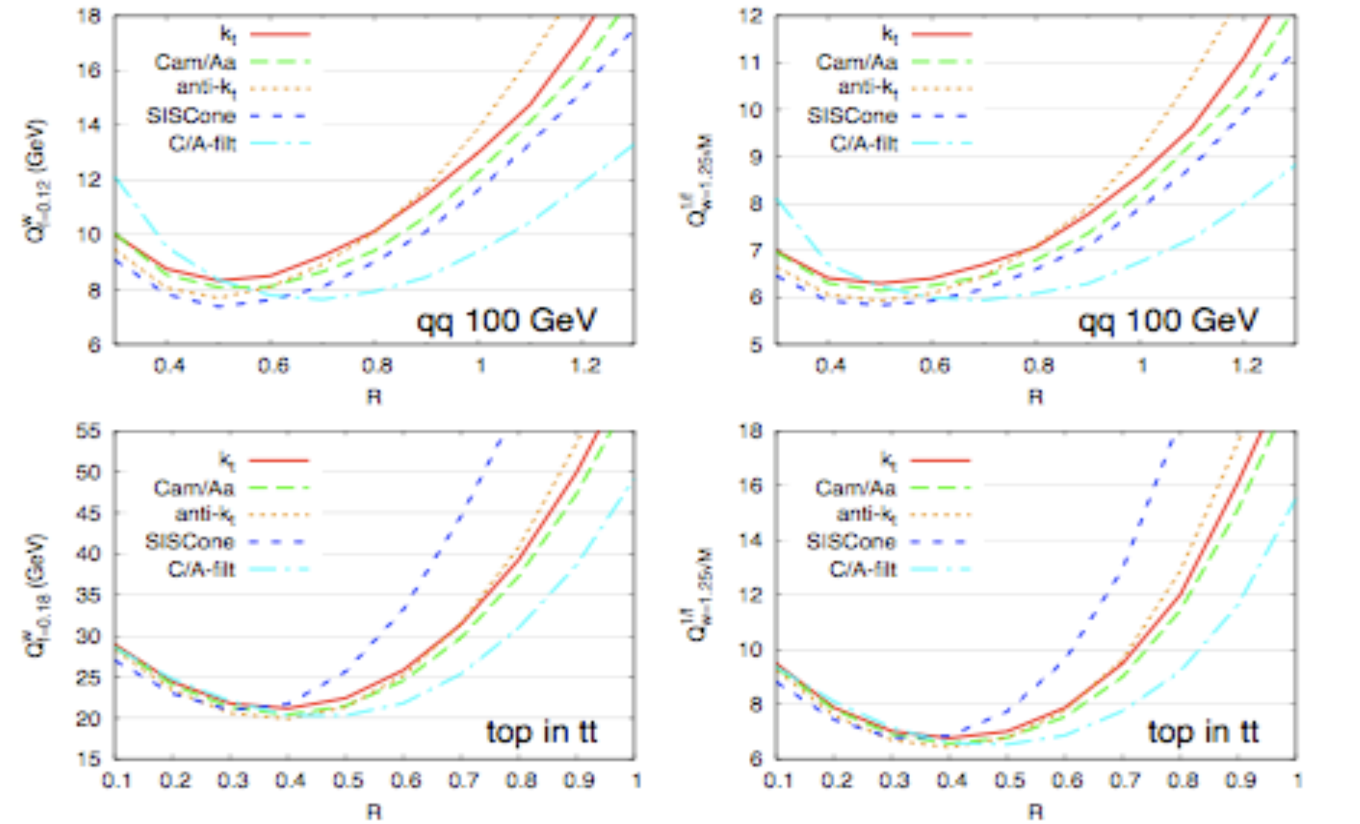


Figure 3: The quality measures $Q^{w}_{f=z}$ (left) and $Q^{1/f}_{w=1.25\sqrt{M}}$ (right), for different jet algorithms as a function of R, for the 100 GeV $q\bar{q}$ case (top row), 2 TeV gg (middle row) and top reconstruction in $t\bar{t}$ events (bottom row).