b-tagging commissionning in DØ

S. Greder

Institut Pluridisciplinaire Hubert Curien, Strasbourg

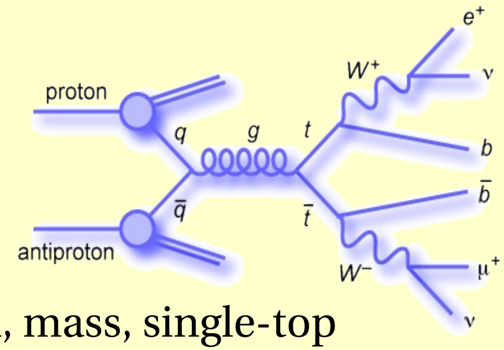
on behalf of DØ collaboration

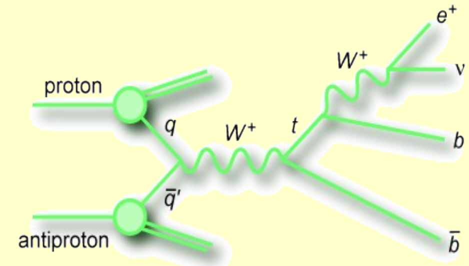
Outline

- Introduction
- The DØ detector
- Algorithms
- Performance measurement
  - Efficiency
  - Fake tag rate
- From Tevatron to LHC
- Conclusion

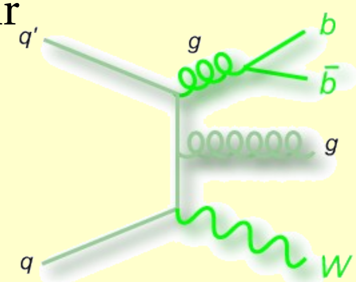
Introduction

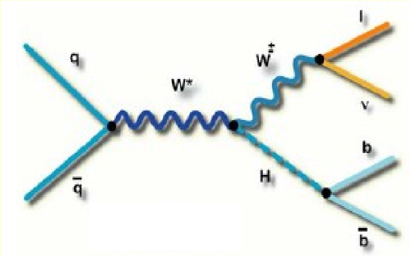
Physics

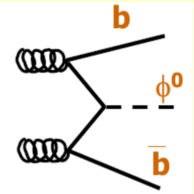




- Top physics: x-section, mass, single-top
- "Backgrounds": W/Z+heavy flavour
- Higgs searches: Low-mass, SUSY

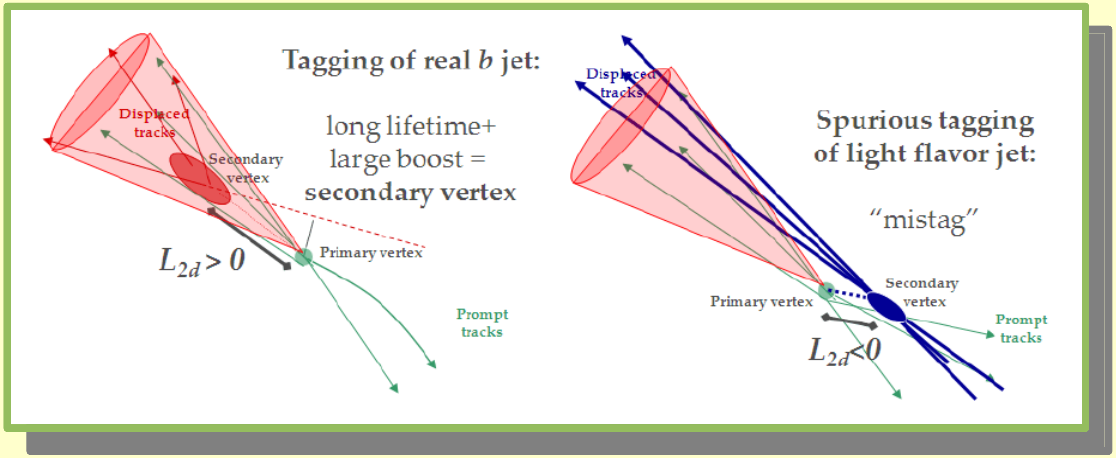






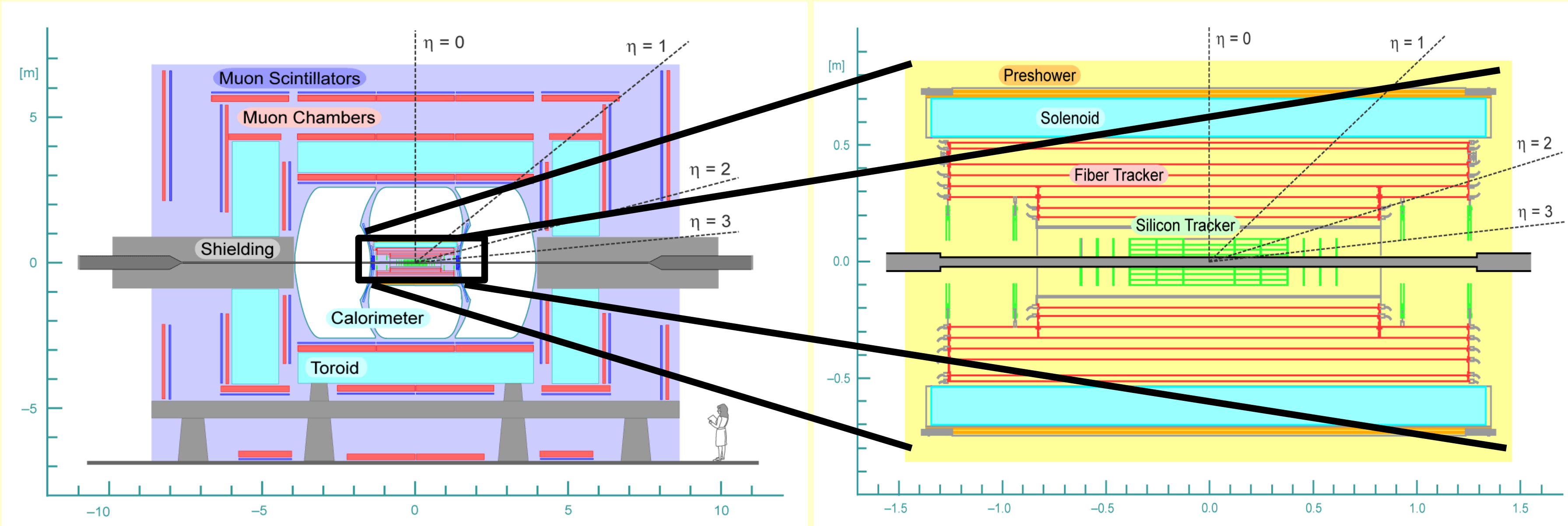
B hadrons properties

- Mass: ~5 GeV/c^2
- Decay length ~mm
- Hard fragmentation
- Semi-leptonic decays



Institut Pluridisciplinaire Hubert Curien Strasbourg

The DØ detector



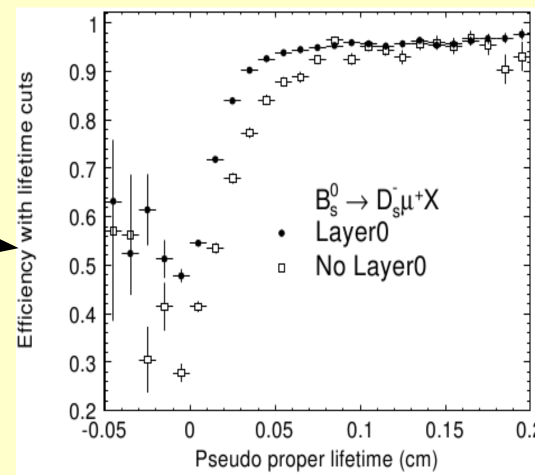
Silicon tracker (SMT)

- 6 barrels, 4 layers each, z ~1 m
- coverage |η|< 2.5
- New Layer 0 @ r =1.6 cm (RunIIb)

Central Fiber Tracker (CFT)

- 8 layers of scintillating fiber (axial and stereo)
- 20 < r < 51 cm in 2T magnetic field

Muon system covers |η|<2

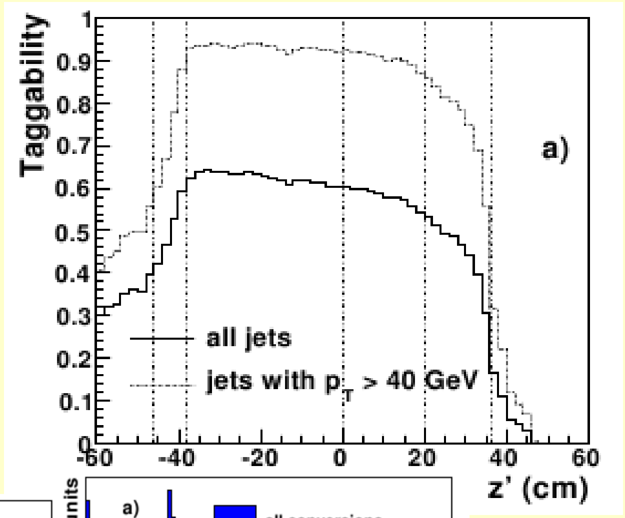


Introduction *(bis)*

Taggability

Tagging algorithm's performance evaluated on **real data** & parametrized as: F (p_T, |η|). But:

- Interaction region, $\sigma_z \approx 25$cm, + detector acceptance affect track reco. efficiency dependence on η differently for different values of the interaction point's z coordinate.
- Electronics noise -> fake jets. Track matching improves discrimination on top of calo. jet-id.
- Fraction of fake jets is ~small, but depends on **final** state. Decoupling this effect from the tagging algorithms proper allows the extraction of a tagging performance which can be assumed to be **universal**, i.e., applicable to general final states.

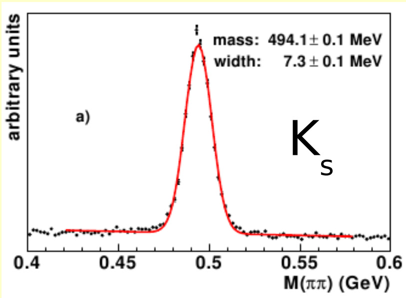


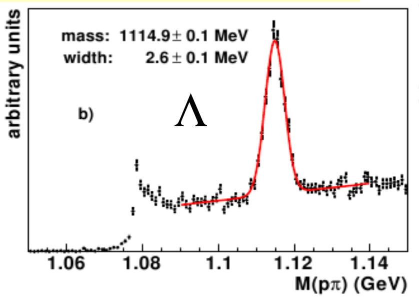
Taggable jets are thus defined as follow:

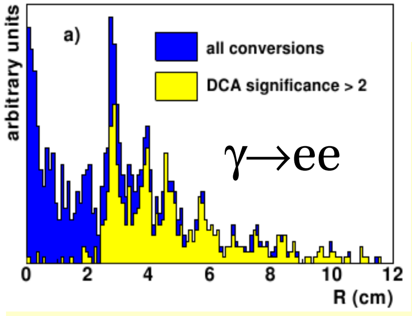
- >=1 tracks, with >= 1 hit in SMT
- **2-step clustering:** along beam axis + 0.5 cone jets (within each z-cluster) and finally require: ΔR(*calo-jet,track-jet*) < 0.5.
- **Parametrized as:** F(p_T,η,z'), with $z' \equiv |z| \cdot \mathrm{sign}(\eta \cdot z)$

Track selection

- Reco. quality
- V^0 removal



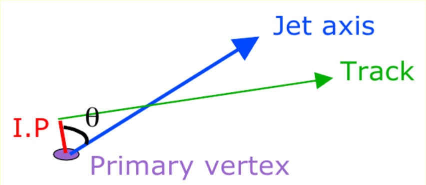


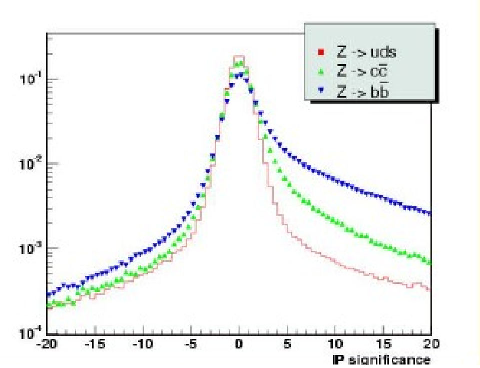


Algorithms (I)

Impact Parameter (IP) based tagger

- IP and its significance S_{IP} are **signed** w.r.t jet direction
- IP error calibrated in data and simulation for *multiple-effects* and *PV resolution dependence*





Discrete (CSIP)

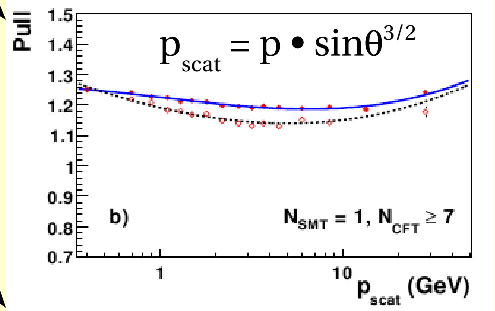
- counts tracks with: $|S_{IP}|$ > cut *(2 >3 || 3 > 2)*

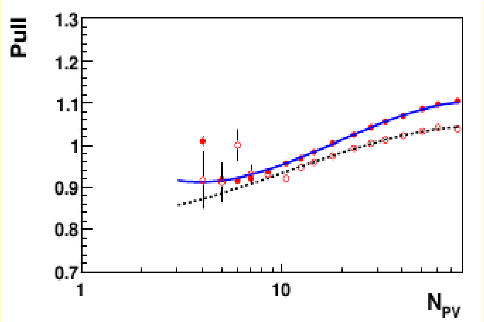
Continuous (JLIP)

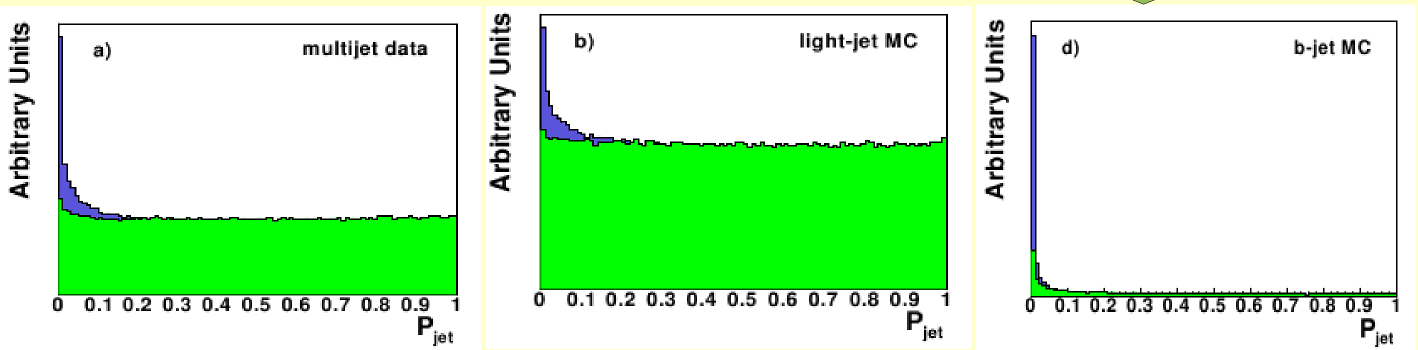
- p.d.f from negative IP resolution function, ***R(s)***

$$\mathcal{P}_{trk}(\mathcal{S}_{IP}) = \frac{\int_{-50}^{-|\mathcal{S}_{IP}|} \mathcal{R}(s)ds}{\int_{-50}^{0} \mathcal{R}(s)ds}$$

$$\mathcal{P}^{\pm}_{jet} = \Pi^{\pm} \times \sum_{j=0}^{N^{\pm}_{trk}-1} \frac{(-\log \Pi^{\pm})^j}{j!} \quad \text{with} \quad \Pi^{\pm} = \prod_{i=1}^{N^{\pm}_{trk}} \mathcal{P}_{trk}(\mathcal{S}^{IP>0}_{IP<0})$$

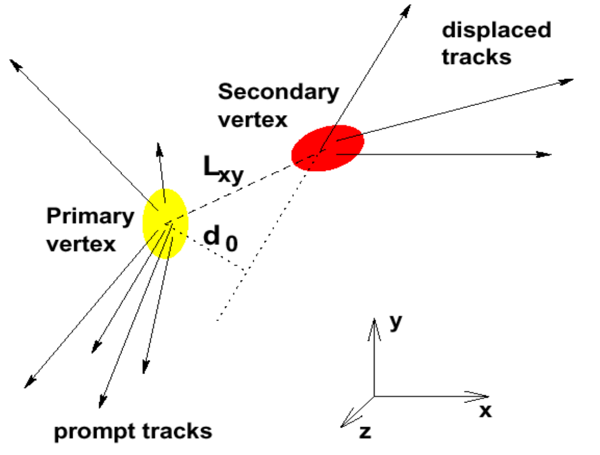




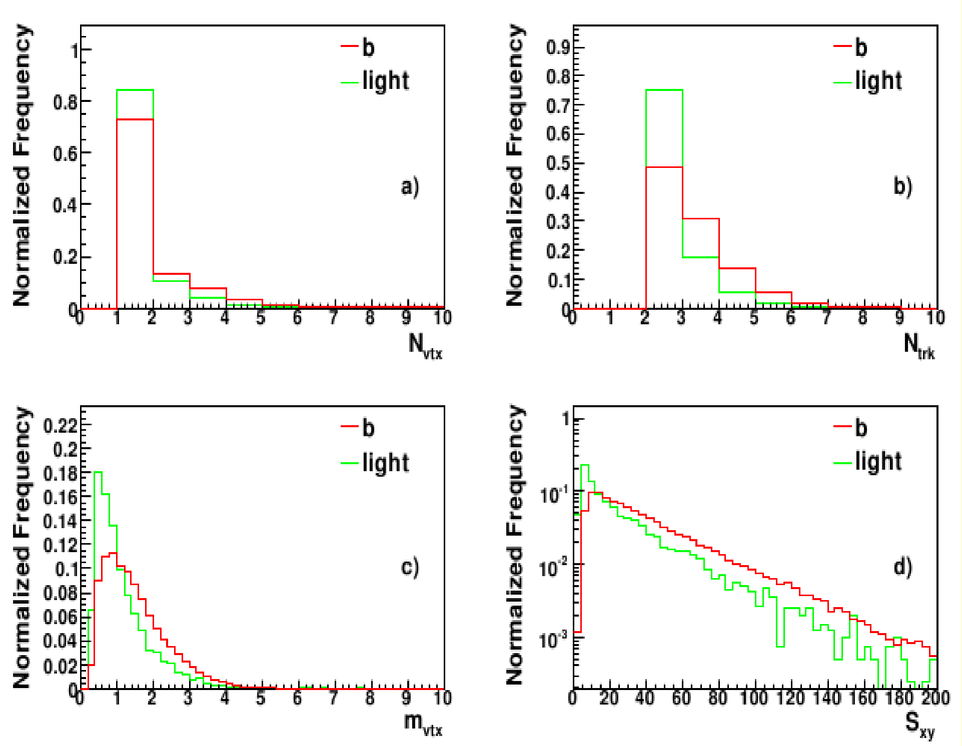


Algorithms (II)

Secondary vertex, SVT



- Starts from track-based jets *(simple cone algo.)*
- Kalman-filter based vertex finder
- Track pruning w.r.t χ^2 contribution to vertex
- **Tag is defined if:** ΔR(vertex,jet) < 0.5 and if decay length significance, S_{Lxy} > cut

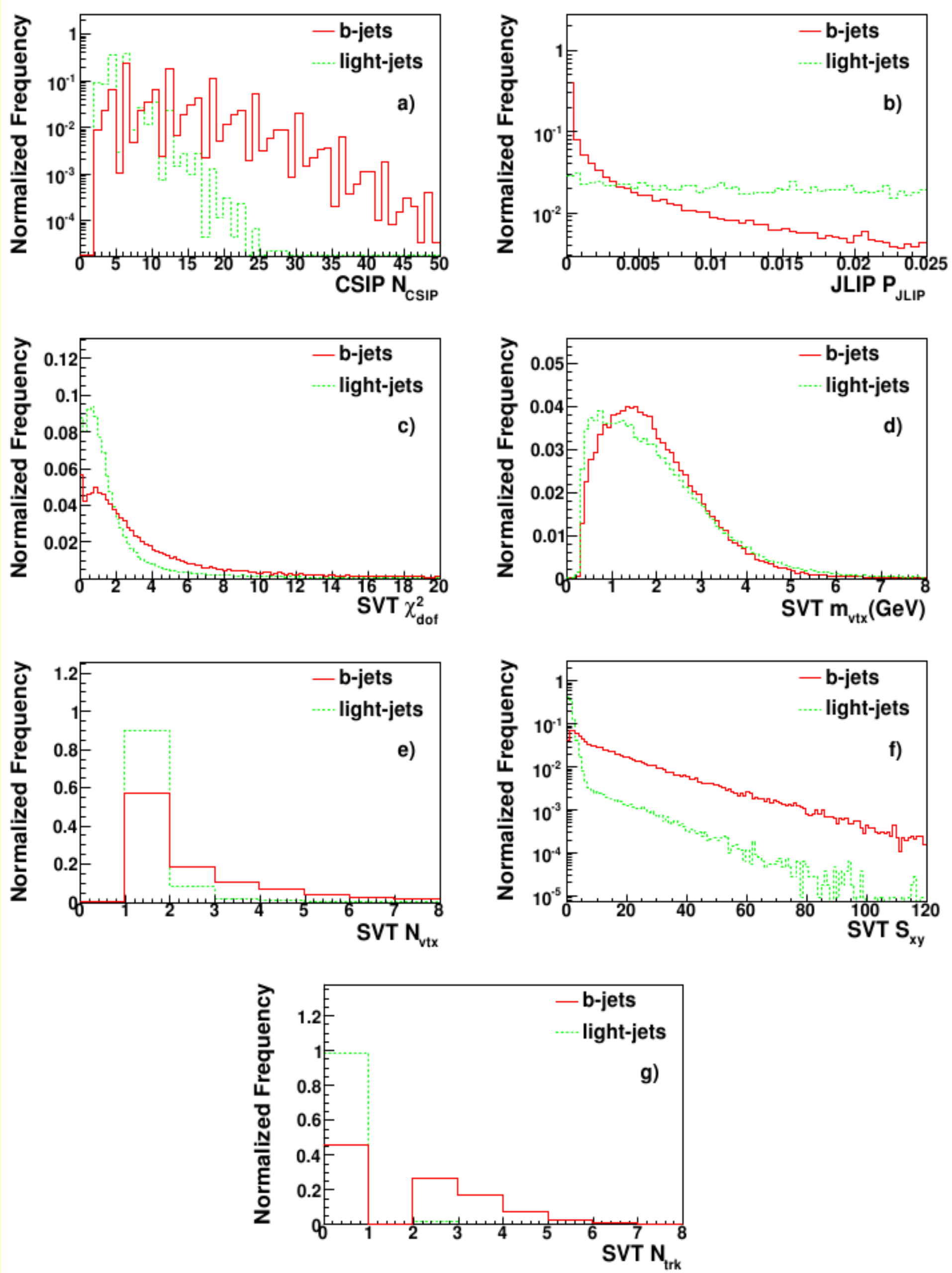


Soft Lepton Tagger, SLT

- 40% of top/anti-top pairs contain a lepton
- Look for a soft (> 4GeV/c) muon/electron in require ΔR(<0.5)
- More details later

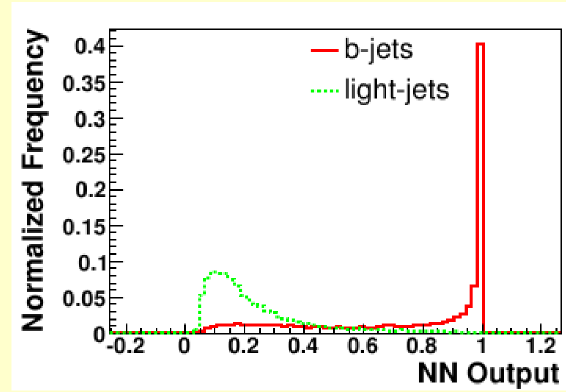
IPHC Institut Pluridisciplinaire Hubert Curien STRASBOURG

All in one: Neural Network tagger

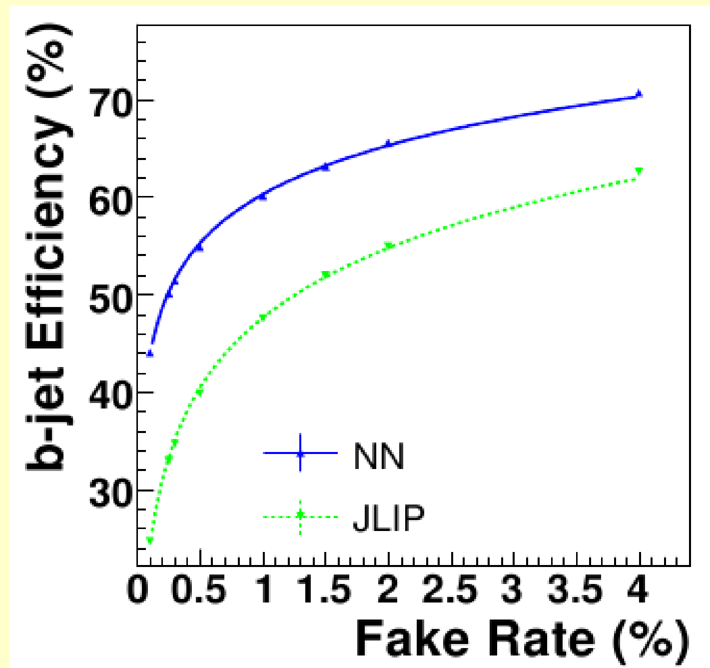


Optimized selection of inputs:

CSIP, JLIP & 5 SVT properties



... can lead to significant improvement:



Sébastien Greder

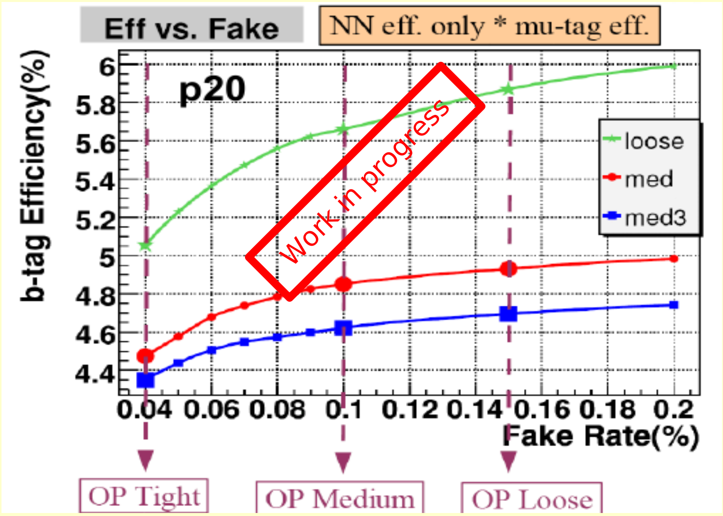
IPHC Institut Pluridisciplinaire Hubert Curien Strasbourg

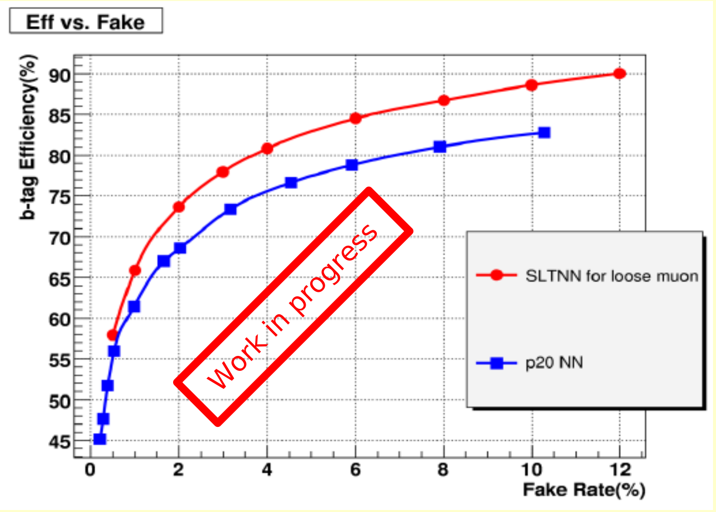
SLT NN

Improved tagging in dedicated topologies

SLTNN with muons

- SLT variables (p_T^{rel}, χ^2, ΔR(jet), ...) can be combined with lifetime variables in a dedicated NN to improve identification performance for semi-leptonic b decays
- Up to **10%** relative increase of signal efficiency @ same fake rate level





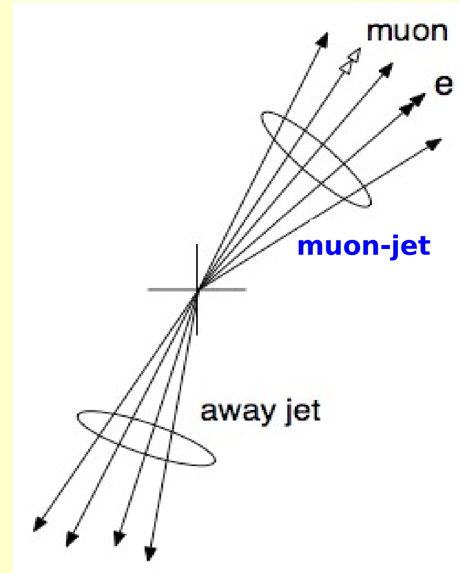
SLT with electrons

- Reconstruction of (low-p_T) electrons in jets is more challenging
- b->eX ~25% identification efficiency for 1% fake rate

Performance: efficiency

Measured in data

- Using <u>b-enriched</u> data samples:
  - Di-jet sample & require $\Delta R(<0.5)$ matched soft (> 4GeV/c) muon in jet

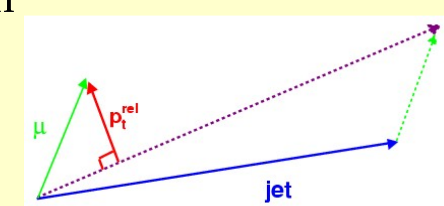


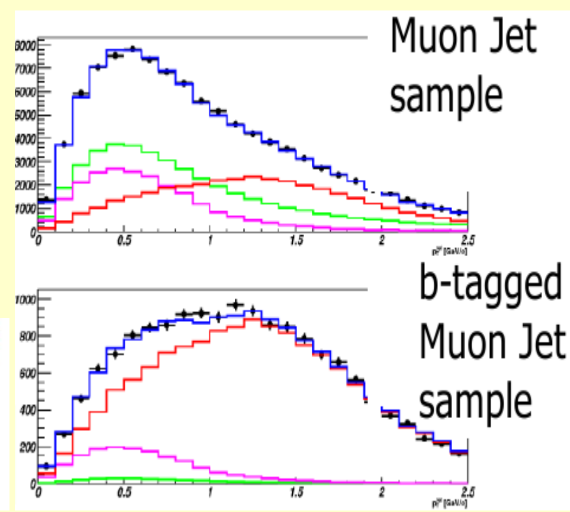
"SystemD" *(bother B. Clement for details ;) !)*

- 2 ~uncorrelated taggers: NN / SLT
- 2 data samples w/ different flavour content:
  - Muon-jet / Muon-jet+away tag
- Apply 2 taggers separately / simultaneously on 2 samples
- 8 equations / 8 unknowns among which ε_b(NN)
- Simulation only used to *corrections factors*

p_T^{rel} (single/double tag)

- Fit p_T^{rel} with templates from simulation
- Used only as x-checks





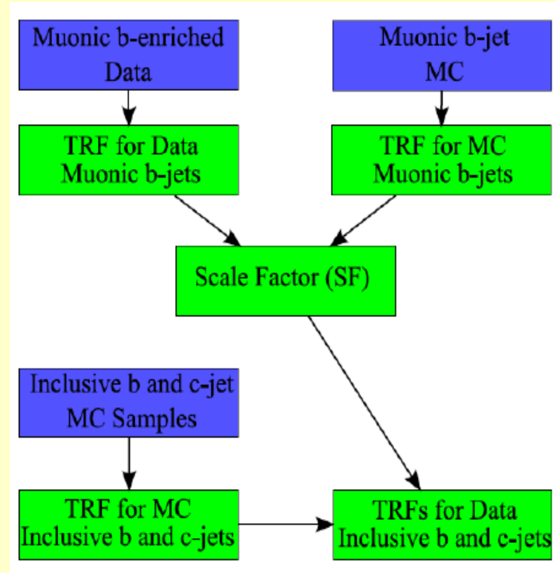
IPHC
Institut Pluridisciplinaire Hubert Curien
STRASBOURG

Scale factors

Tag Rate Functions (TRF)

- Parametrized as a function of p_T,η:

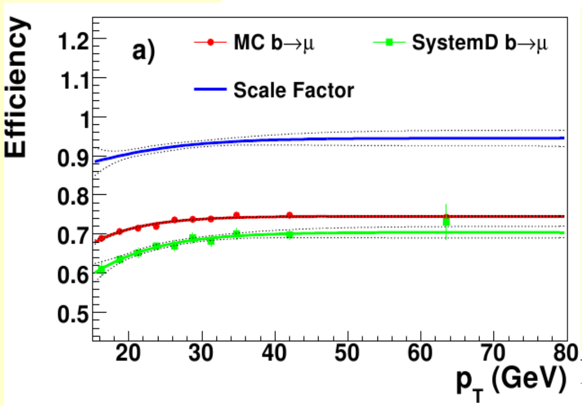
$$\varepsilon(p_T,\eta) = \frac{1}{\varepsilon_{\mathrm{all}}} \cdot f(p_T) \cdot g(|\eta|)$$

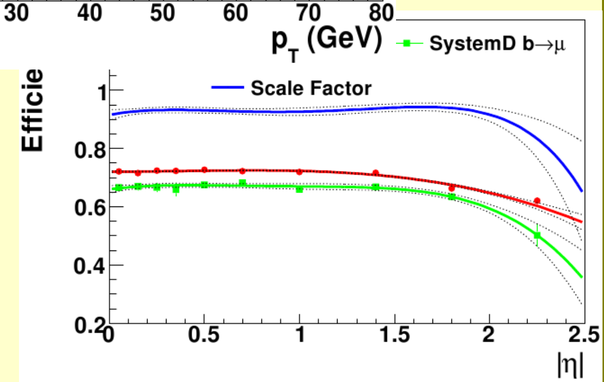


SystemD allows to estimate b-tagging efficiency for semi-leptonic b decays

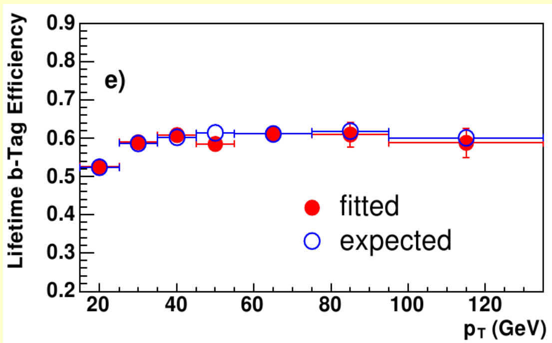
- Inclusive efficiency requires a **Scale Factor** correction:

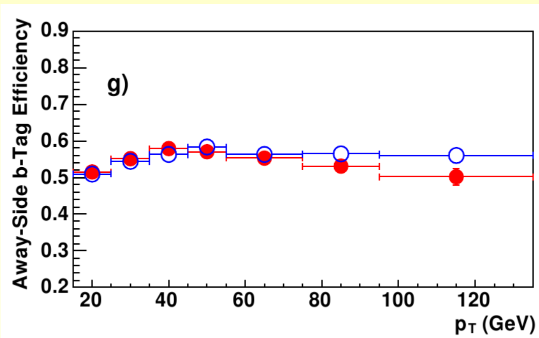
$$\varepsilon_b^{\mathrm{data}} = \frac{\varepsilon_{b\to\mu X}^{\mathrm{data}} \cdot \varepsilon_b^{\mathrm{MC}}}{\varepsilon_{b\to\mu X}^{\mathrm{MC}}} = \mathrm{SF}_b \cdot \varepsilon_b^{\mathrm{MC}}$$





Method validated in simulation:



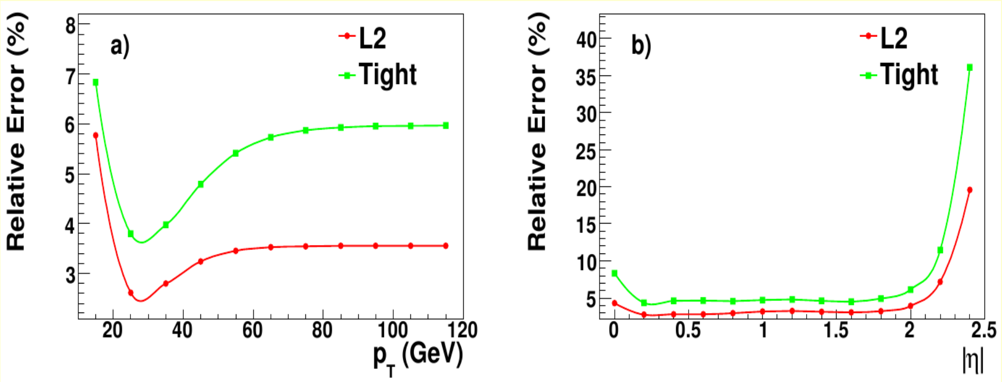


IPHC Institut Pluridisciplinaire Hubert Curien Strasbourg

Systematic uncertainties

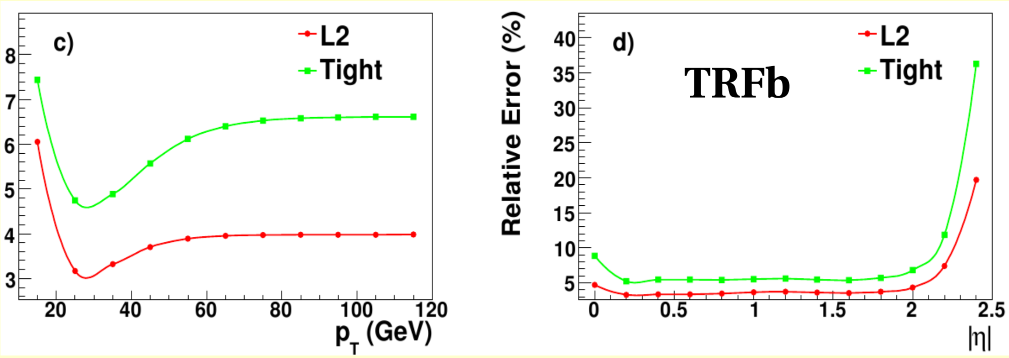
SF_b , SystemD

- Choice of p_T^{rel} cut
- Tagger correlations (factorization, PV)
- Fit uncertainty

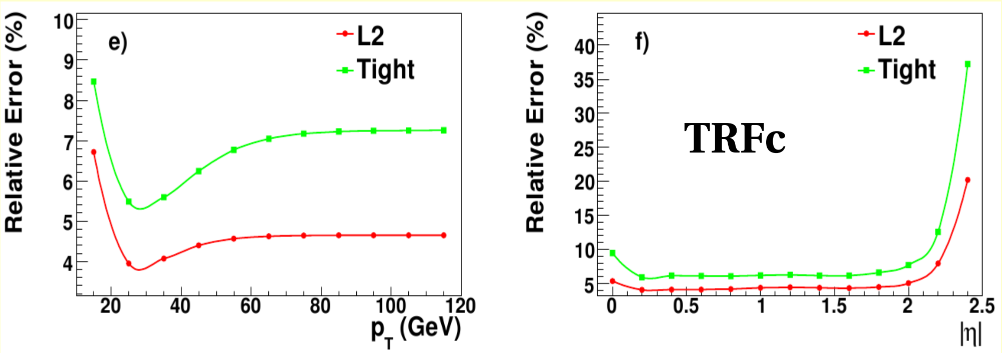


TRF b/c

- Parametrization
- MC sample dependance
- Residuals from closure plot predicted and observed tags



Total uncertainties added in quadrature



Performance: Fake tag rate

Goal

- Estimate $\boldsymbol{\varepsilon}_{\mathbf{light}}$ where light = u, d, s and gluon

Measured in data

- Various samples: jet / EM triggered multijet data ("QCD")

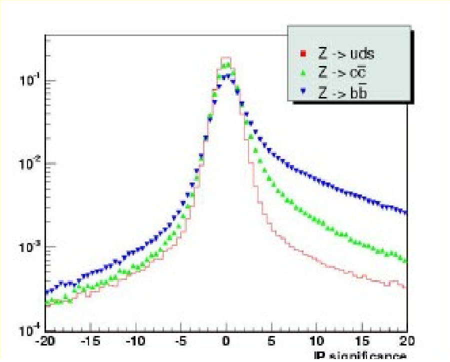
Estimated from negative tags

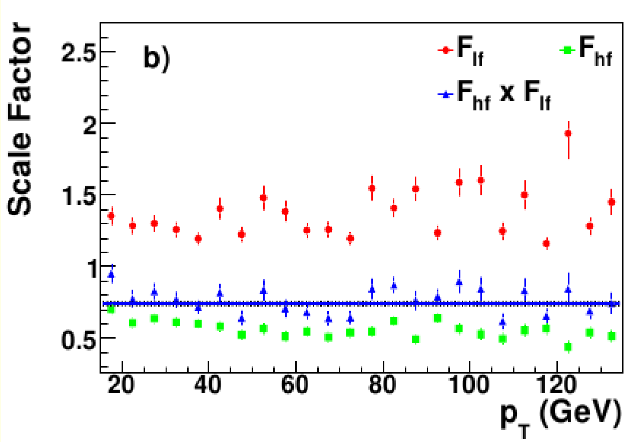
- **corrected for:**
  - HF contamination $F_{\rm hf} = \varepsilon^-_{\rm QCD,light}/\varepsilon^-_{\rm QCD,all}$
  - neg./pos. asymmetry $F_{\rm lf} = \varepsilon^+_{\rm QCD,light}/\varepsilon^-_{\rm QCD,light}$
- Parametrized as F(p_T,η)

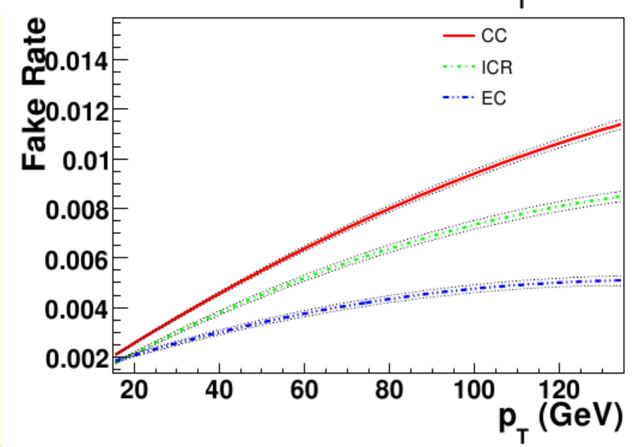
$$\varepsilon_{\rm light} = \varepsilon^-_{\rm data} \cdot F_{\rm hf} \cdot F_{\rm lf}$$

Systematics:

- Sample dependance: jet vs. EM QCD
- Relative heavy flavour fractions
- Parametrizations
- **Total:** ~3 - 8.5%





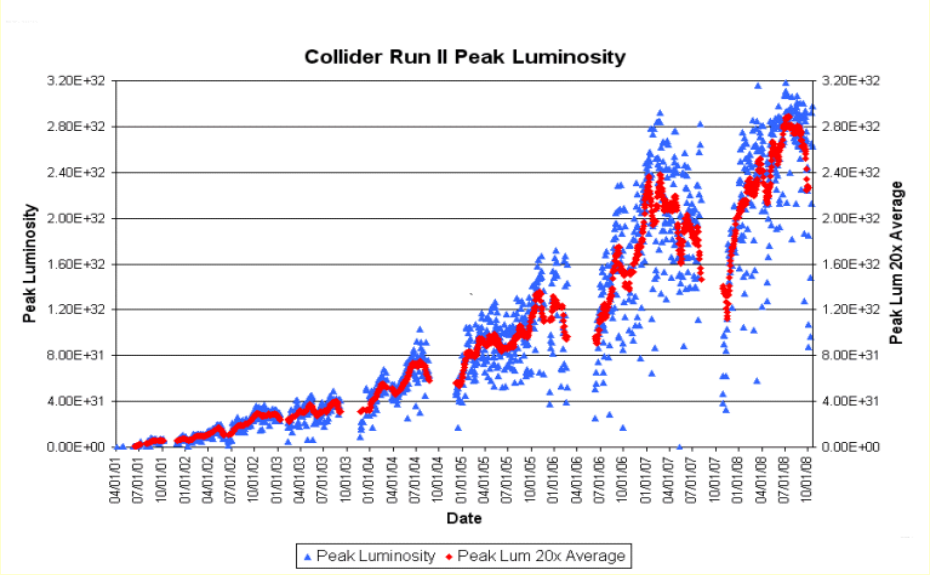


IPHC Institut Pluridisciplinaire Hubert Curien Strasbourg

From Tevatron to LHC

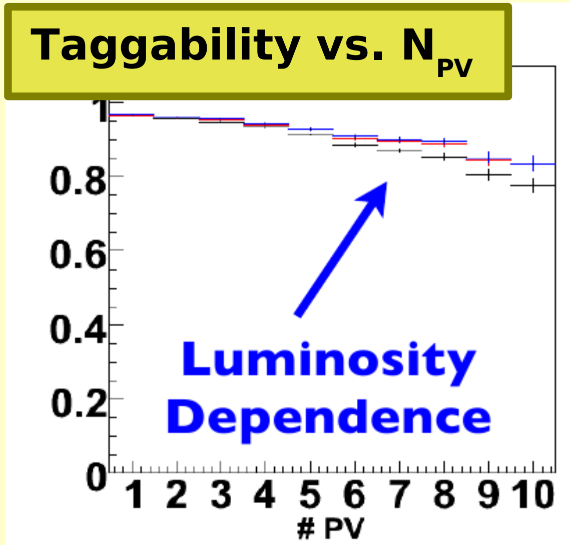
RunIIa -> RunIIb -> LHC

- Increase of instantaneous luminosity
- ~ 4 “PV” @ $3x10^{32}$



Performance sensitive to

- Number of multiple interaction
- Tracker occupancy
- Hits merge
- Track selection vs. luminosity ?
- ....



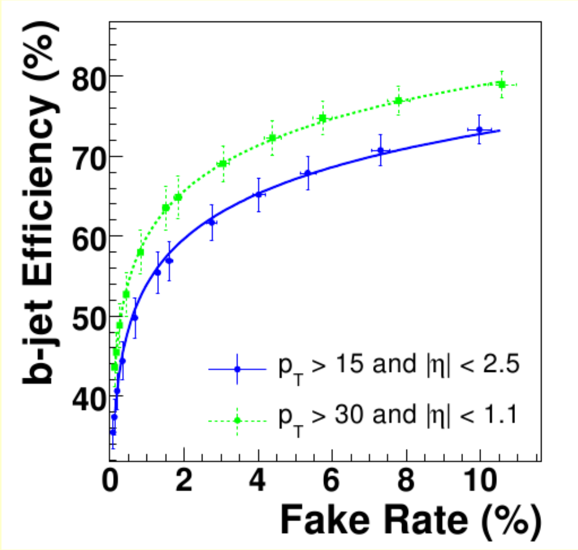
Sébastien Greder

Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

Conclusion

B-tagging is a necessary tool for many forefront analyses

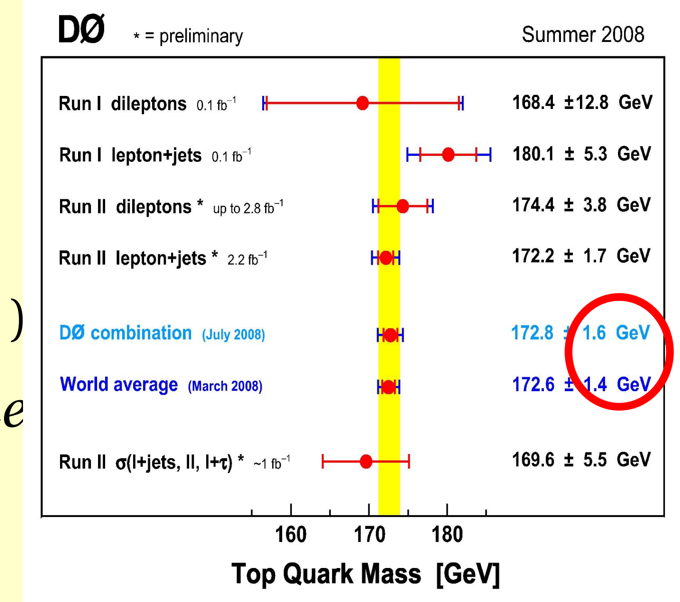
- Very good performance despite of *complex* and *busy* hadronic environment
- Advanced multi-variate tools are an **asset** to keep high signal efficiency/low fake rates
  - And **simplify** procedures: reduce complexity to 1 variable



Increasing luminosity must be carefully handled

Very good control of systematic errors is compulsory

- they enter in top-3 errors for top mass, single-top and low-mass Higgs searches, ...)



Smart techniques + very good understanding of the detector require a lot of effort .... but worth it: