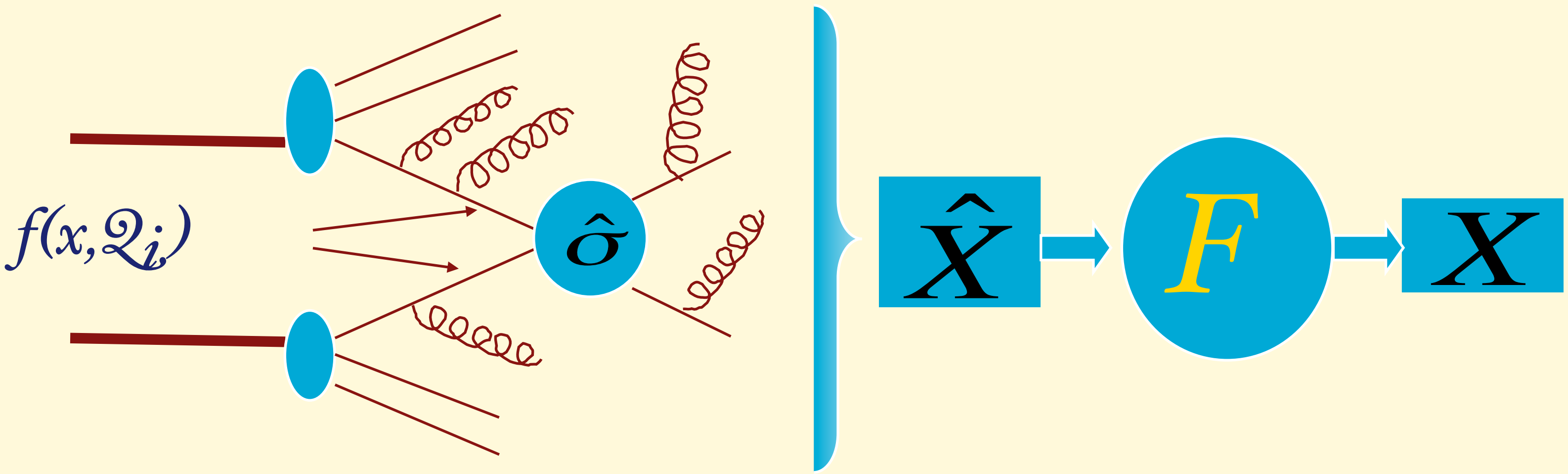
The challenge of predicting the structure of LHC final states

Workshop on LHC physics
TIFR, Mumbai, 21-27 October 2009

Michelangelo L. Mangano
Theoretical Physics Unit
Physics Department
CERN, Geneva

Factorization Theorem

$$\frac{d\sigma}{dX} = \sum_{j,k} \int_{\hat{X}} f_j(x_1,Q_i)\, f_k(x_2,Q_i)\, \frac{d\hat{\sigma}_{jk}(Q_i,Q_f)}{d\hat{X}}\, F(\hat{X} \to X; Q_i, Q_f)$$



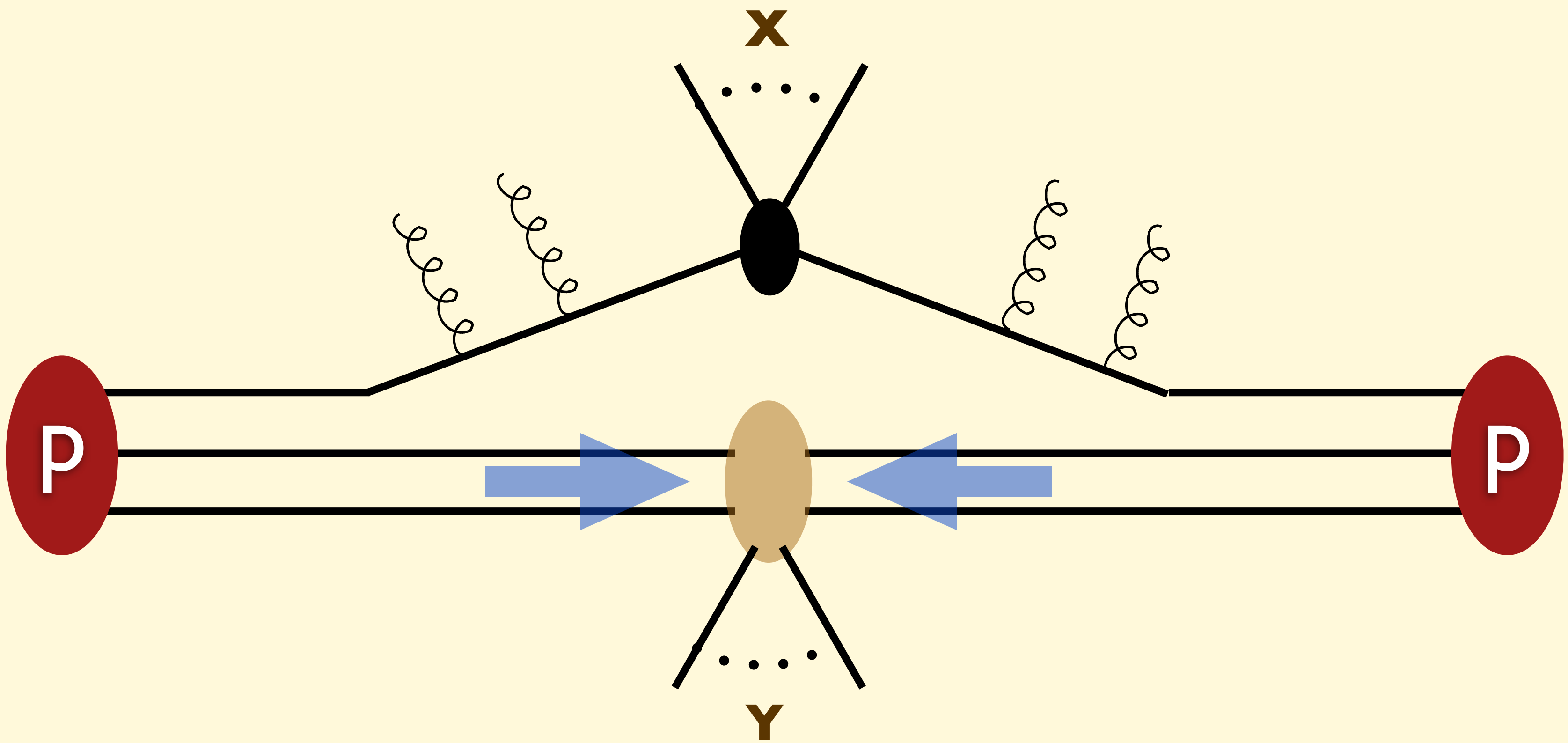
$f_j(x,Q)$ Parton distribution functions (PDF)

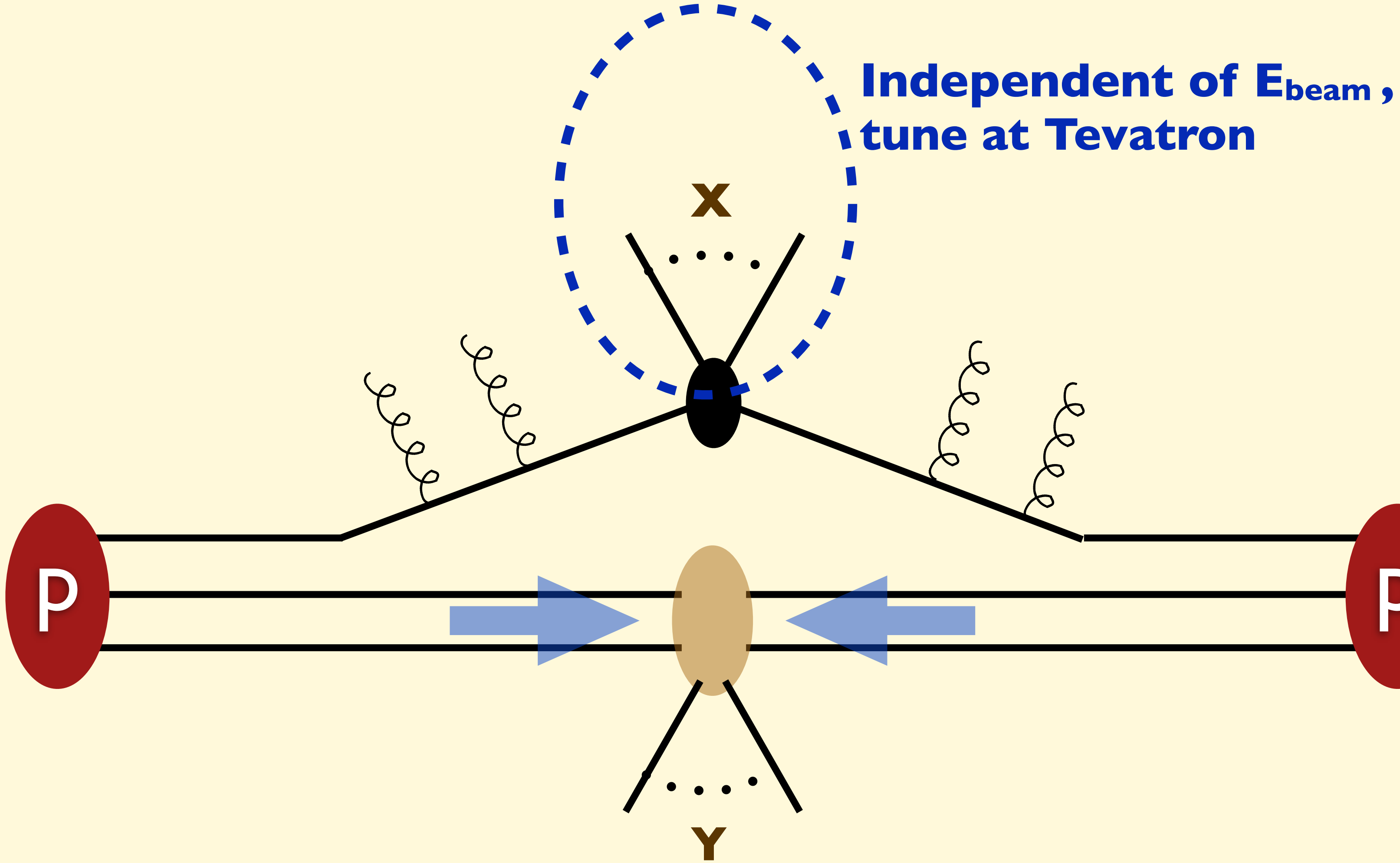
- sum over all initial state histories leading, at the scale Q, to:

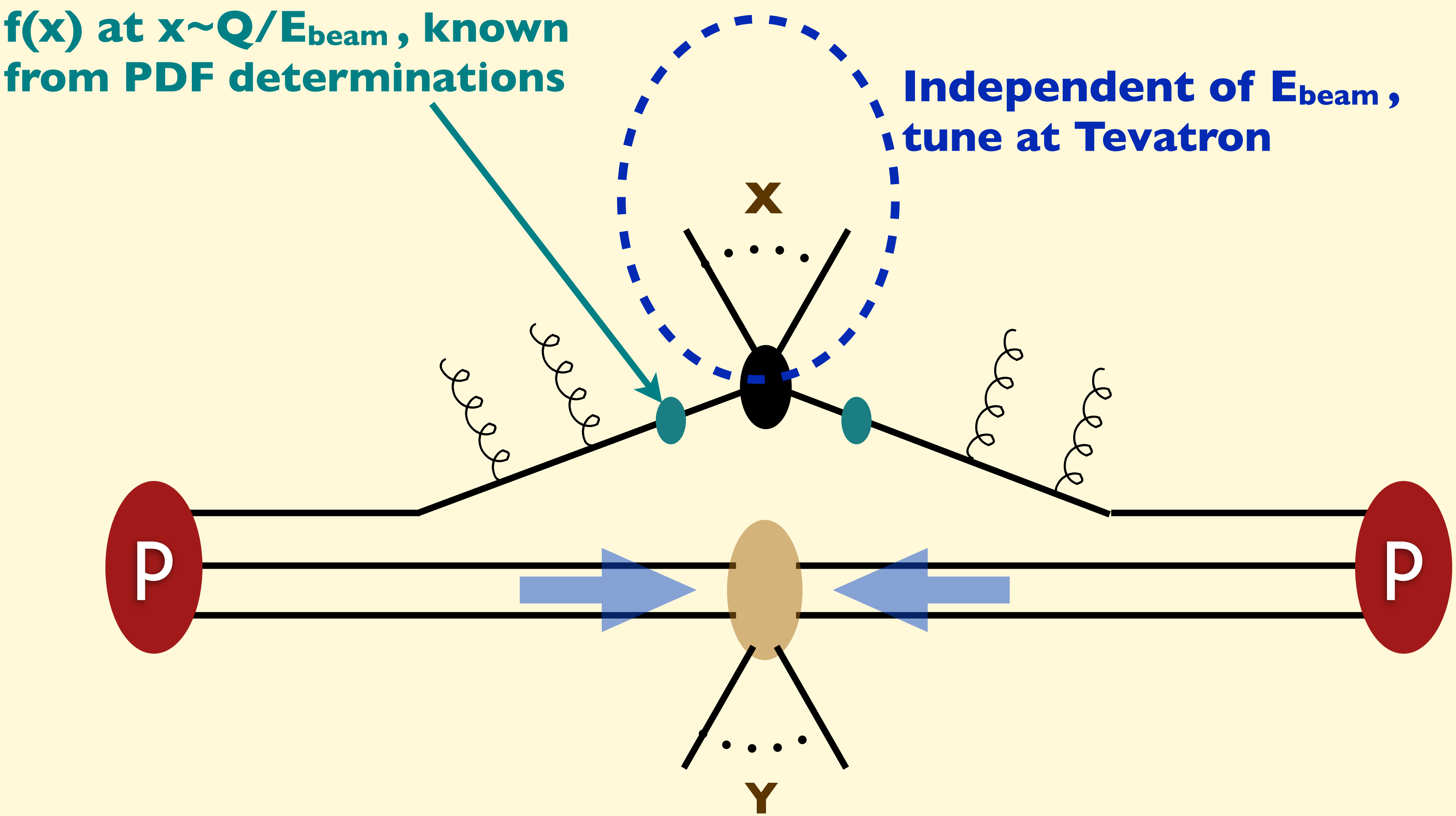
$$\vec{p}_j = x\,\vec{P}_{proton}$$

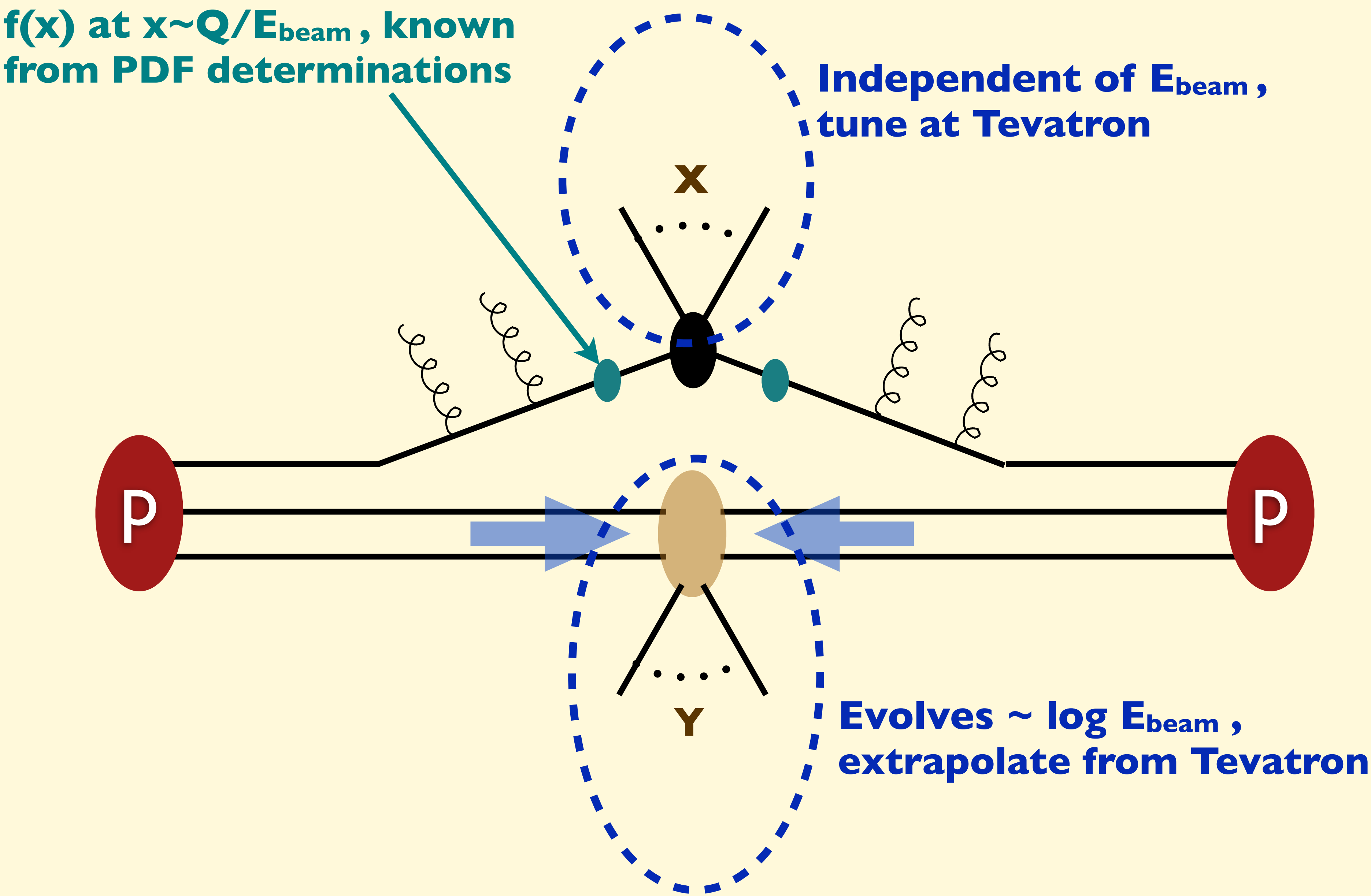
$F(\hat{X} \to X; Q_i, Q_f)$

- transition from partonic final state to the hadronic observable (hadronization, fragm. function, jet definition, etc)
- Sum over all histories with X in them

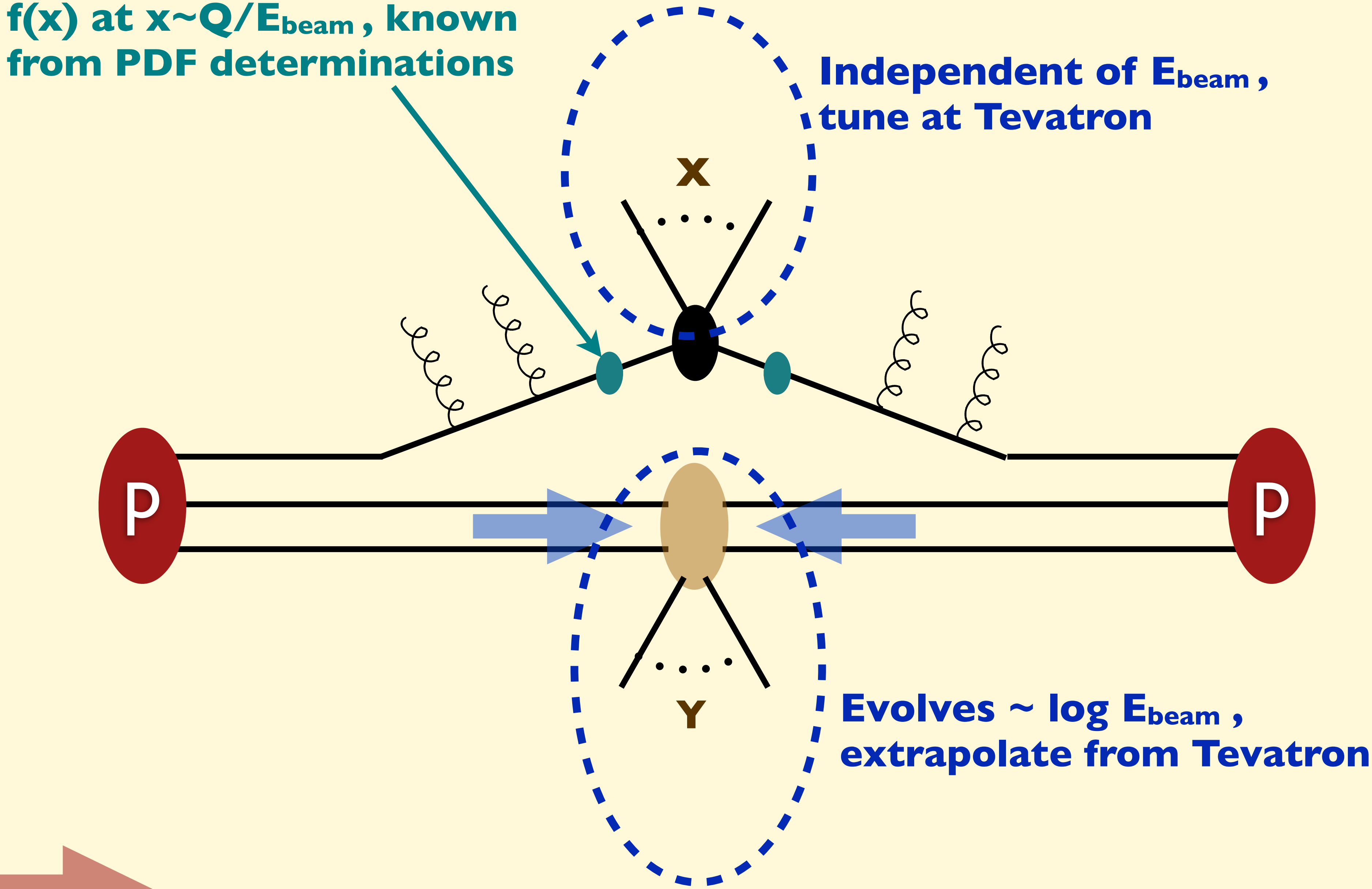

X
p
p
Y


Independent of E_{beam},
tune at Tevatron
x
p
p
Y


f(x) at x~Q/E$_{beam}$, known from PDF determinations
Independent of E$_{beam}$, tune at Tevatron
X
Y
p
p



f(x) at x~Q/Ebeam , known
from PDF determinations
Independent of Ebeam ,
tune at Tevatron
X
p
p
Y
Evolves ~ log Ebeam ,
extrapolate from Tevatron



all seems to be under control and easily predictable at the LHC

... on the other hand ...

... on the other hand ...

- the energy reach at the LHC is such that in many instances we're exploring kinematical regions never probed before

... on the other hand ...

- the energy reach at the LHC is such that in many instances we're exploring kinematical regions never probed before
- cross sections for many processes of interest at the LHC are too small at the Tevatron to give significant tests of our dynamical understanding

... on the other hand ...

- the energy reach at the LHC is such that in many instances we're exploring kinematical regions never probed before
- cross sections for many processes of interest at the LHC are too small at the Tevatron to give significant tests of our dynamical understanding
- this is particularly true of final states with many jets, especially if produced in association with gauge bosons and/or heavy quarks (such as bottom and top)

... on the other hand ...

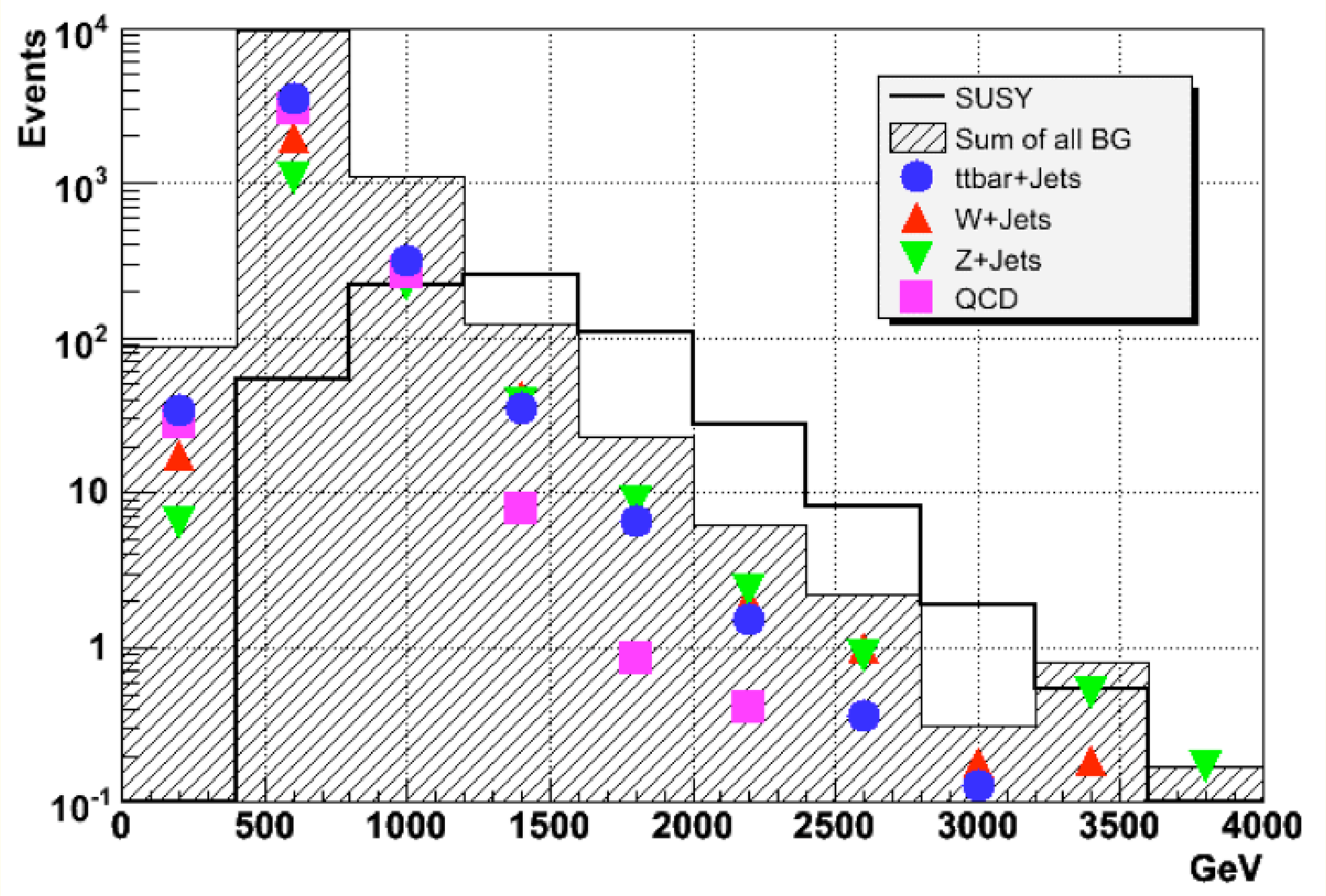
- the energy reach at the LHC is such that in many instances we're exploring kinematical regions never probed before
- cross sections for many processes of interest at the LHC are too small at the Tevatron to give significant tests of our dynamical understanding
- this is particularly true of final states with many jets, especially if produced in association with gauge bosons and/or heavy quarks (such as bottom and top)
- Example: backgrounds to Supersymmetry ....

Example of SUSY-search analysis cuts (ATLAS):

$\geq$4jets, E_T>50 GeV   leading jet E_T>100 GeV

MissET> max(100, 0.2 M_{eff}) M_{eff} = MET + $\sum_{i=1,..,4}$ E_T^i

Transverse sphericity > 0.2   no lepton with E_T>20 GeV



Given that the bg and signal shapes are almost identical, where do we get the size of these bkgs from?

Given that the bg and signal shapes are almost identical, where do we get the size of these bkgs from?

It's often said "we just measure them from the LHC data", and then go on to estimate the potential signal significance

Given that the bg and signal shapes are almost identical, where do we get the size of these bkgs from?

It's often said "we just measure them from the LHC data", and then go on to estimate the potential signal significance

But what if what we predict and what we measure can't be made to agree in a trivial way, e.g. by just including a K factor?

Given that the bg and signal shapes are almost identical, where do we get the size of these bkgs from?

It's often said "we just measure them from the LHC data", and then go on to estimate the potential signal significance

But what if what we predict and what we measure can't be made to agree in a trivial way, e.g. by just including a K factor?

I'll give here an overview of what we could find, based on our current experience at the Tevatron

Given that the bg and signal shapes are almost identical, where do we get the size of these bkgs from?

It's often said "we just measure them from the LHC data", and then go on to estimate the potential signal significance

But what if what we predict and what we measure can't be made to agree in a trivial way, e.g. by just including a K factor?

I'll give here an overview of what we could find, based on our current experience at the Tevatron

This will implicitly outline which measurements deserve fast-track status once the first data arrive

Given that the bg and signal shapes are almost identical, where do we get the size of these bkgs from?

It's often said "we just measure them from the LHC data", and then go on to estimate the potential signal significance

But what if what we predict and what we measure can't be made to agree in a trivial way, e.g. by just including a K factor?

I'll give here an overview of what we could find, based on our current experience at the Tevatron

This will implicitly outline which measurements deserve fast-track status once the first data arrive

Notice however that, at the Tevatron, these kind of measurements are among those that took the longest time to reach maturity ......

Goals of theoretical predictions:

- **accurate absolute predictions for inclusive quantities:**
  - E.g. W/Z total cross sections ⇒ luminosity determination, PDF measurements
  - E.g. Higgs and other new particles cross sections ⇒ extract couplings, BRs
  - **⇨ require N(N)LO for reduced scale dependence**
- **complete description of final states**
  - complete description of SM processes with, e.g. ,
    - large jet multiplicities
    - associated production of multiple EW and QCD objects (t,b,g,H,W,...)
  - *Goal is not necessarily first-principle predictability, but good agreement with data after tuning*
  - **⇨ require full MC generators, flexibility in the input param's for accurate tuning**

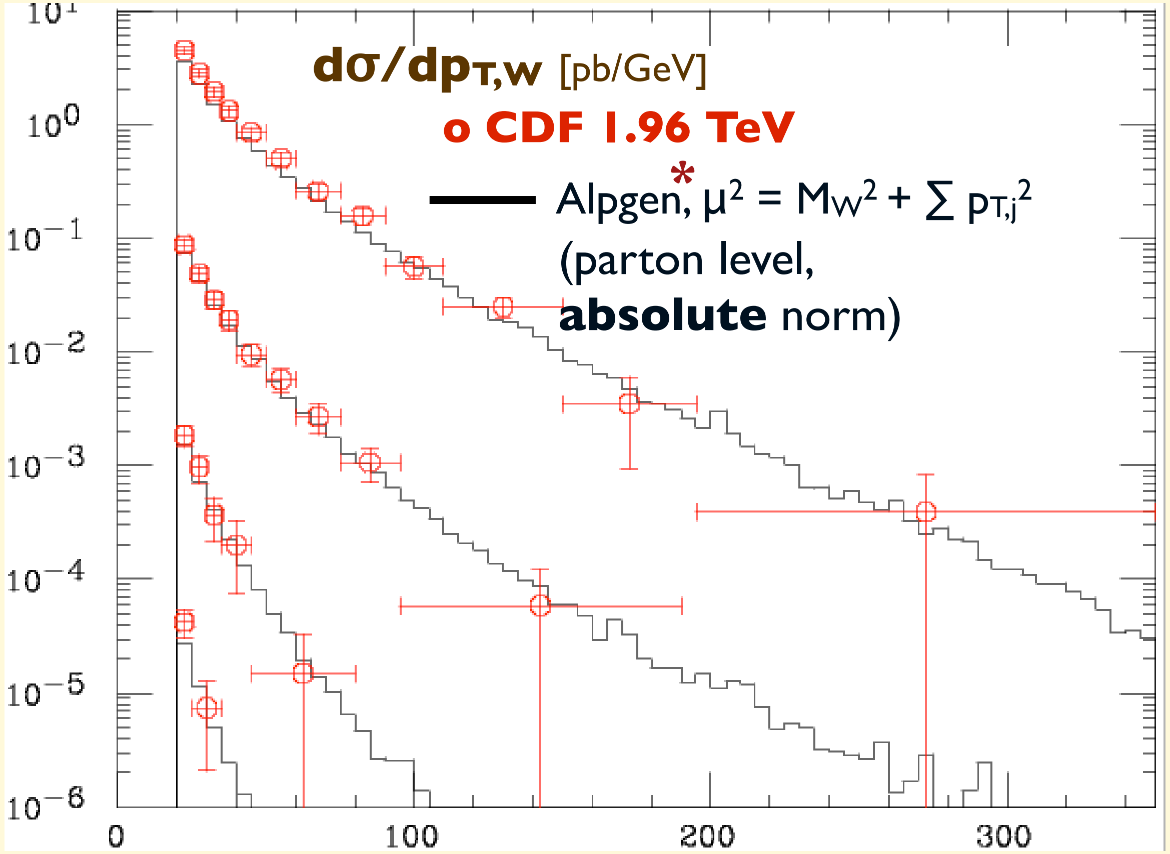
Tools: examples for Z/W/γ+jets

- Parton-level LO matrix element generators
- LO ME + shower MCs, with **merging of different jet multiplicities** (up to 4–6 jets, depending on code)**:**
  - ALPGEN (MLM merging scheme)
  - ARIADNE (Lonnblad merging)
  - HELAC, MadEvent (MLM merging)
  - SHERPA (CKKW merging)
- NLO PL matrix element generators:
  - DYRAD (up to 1 jet @ NLO)
  - MCFM (up to 2 jets @NLO)
- MC@NLO/POWHEG (inclusive W @NLO)
- Resummed inclusive W pt spectra (RESBOS)

Validation: comparison of jet Et spectra in data and LO ME calculations of [W→e/μ ν]+multijet events.

Example: Tevatron data

* any other PL ME generator (Vecbos, Madgraph, etc) would give the same result



$p_{T,W}$ (GeV)

Key experimental issue:
at large jet multiplicity and MET, the non-W bg to [W→e/μ ν]+multijet is very very large*! So the control sample itself is dominated by backgrounds yet harder to estimate .....

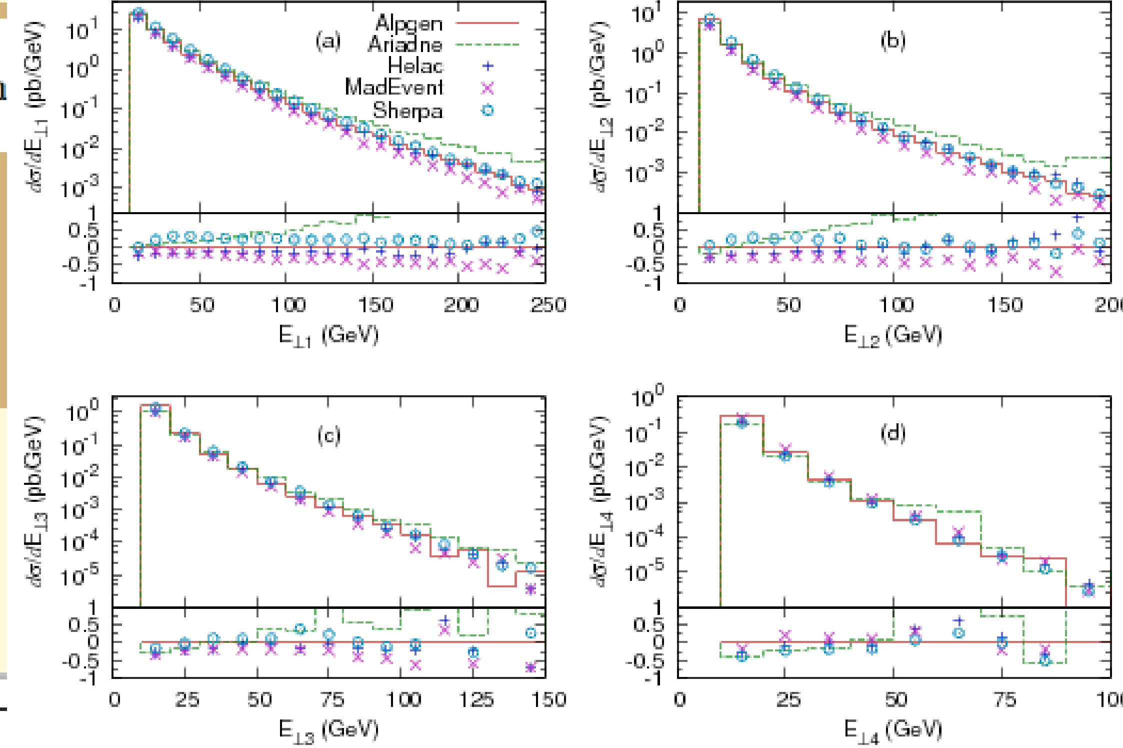
* Mostly b/c SL decays, together with mismeasurement of jet ET, but also t-tbar

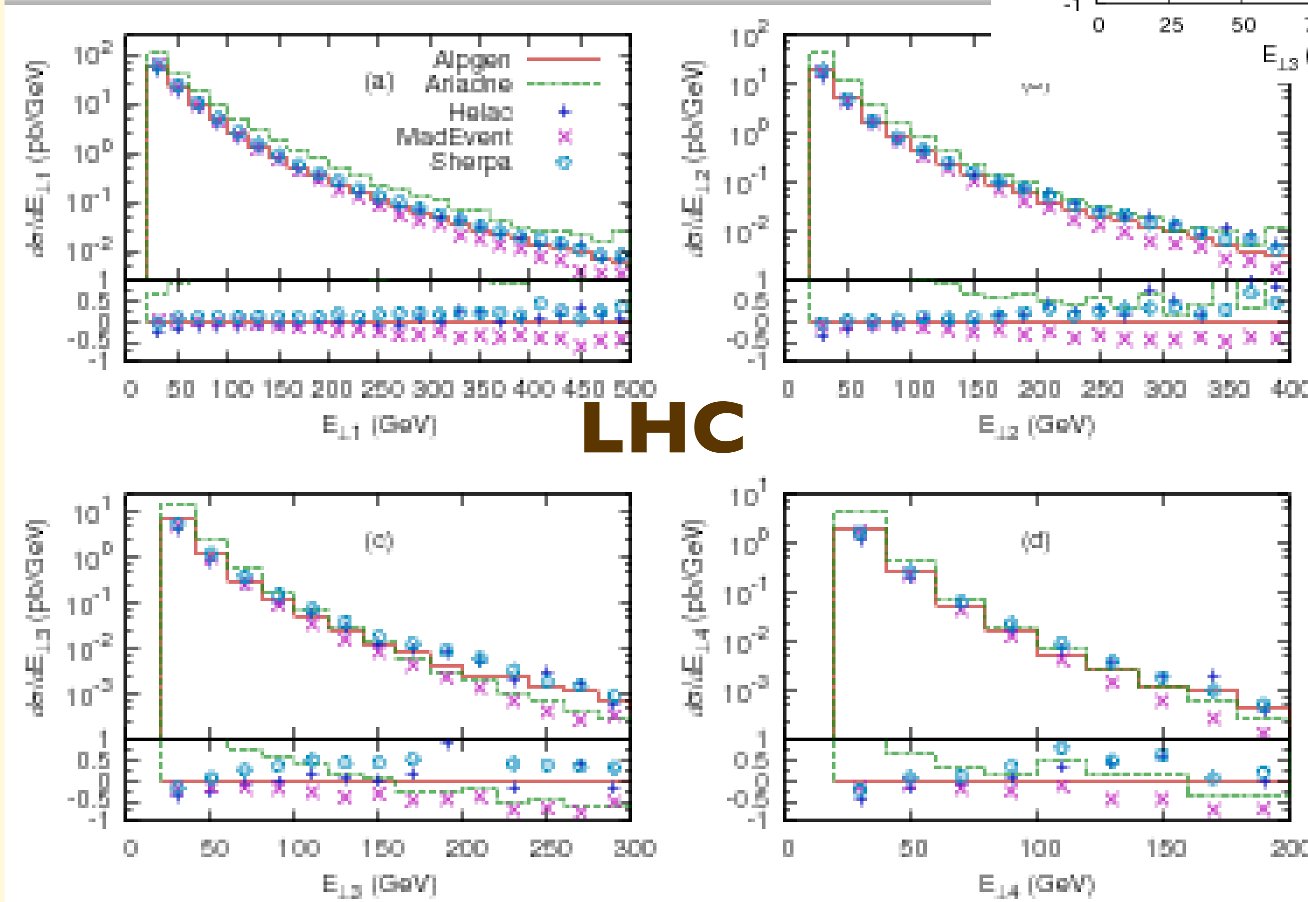
Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions *

Alpgen, Ariadne, Helac, MadEvent, Sherpa

J. Alwall[1], S. Höche[2], F. Krauss[2], N. Lavesson[3], L. Lönnblad[3], F. Maltoni[4], M.L. Mangano[5], M. Moretti[6], C.G. Papadopoulos[7], F. Piccinini[8], S. Schumann[9], M. Treccani[6], J. Winter[9], M. Worek[10,11]

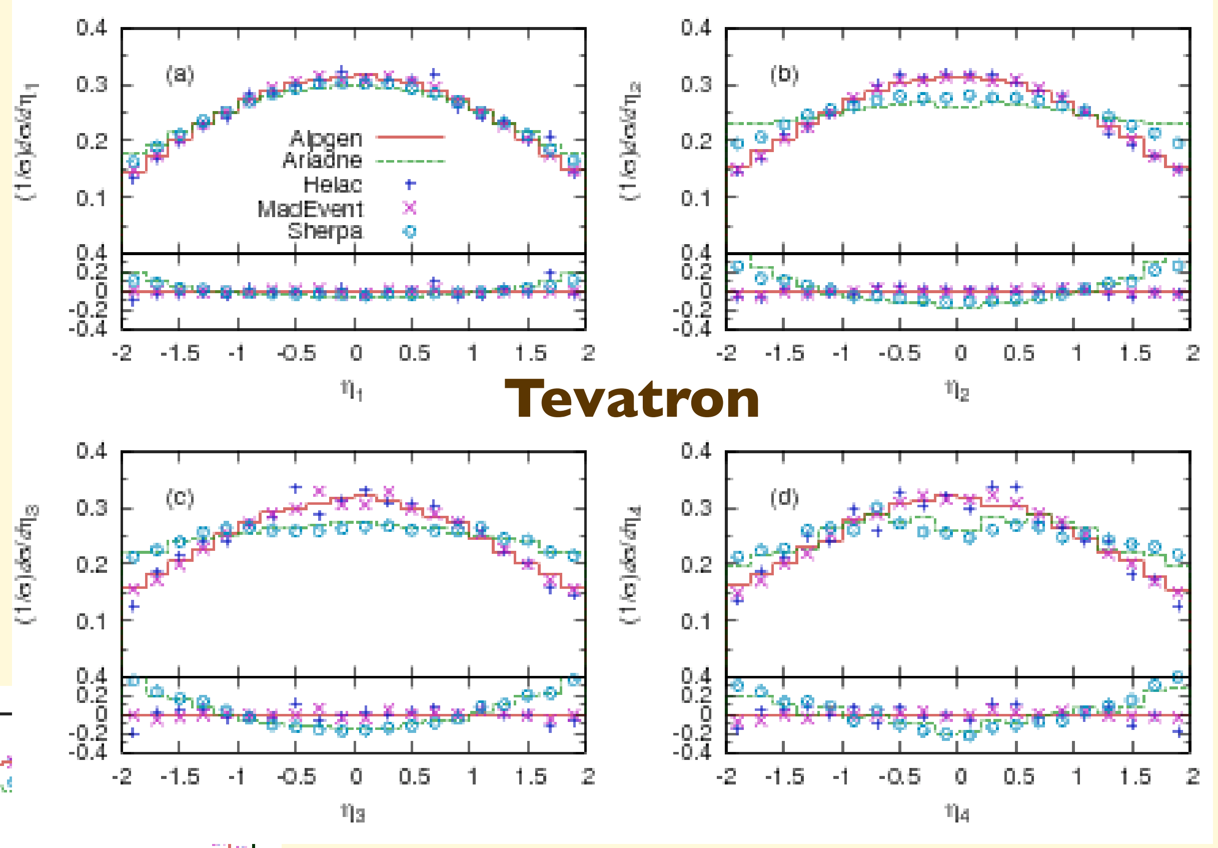
W+multijet, jet E_T spectra



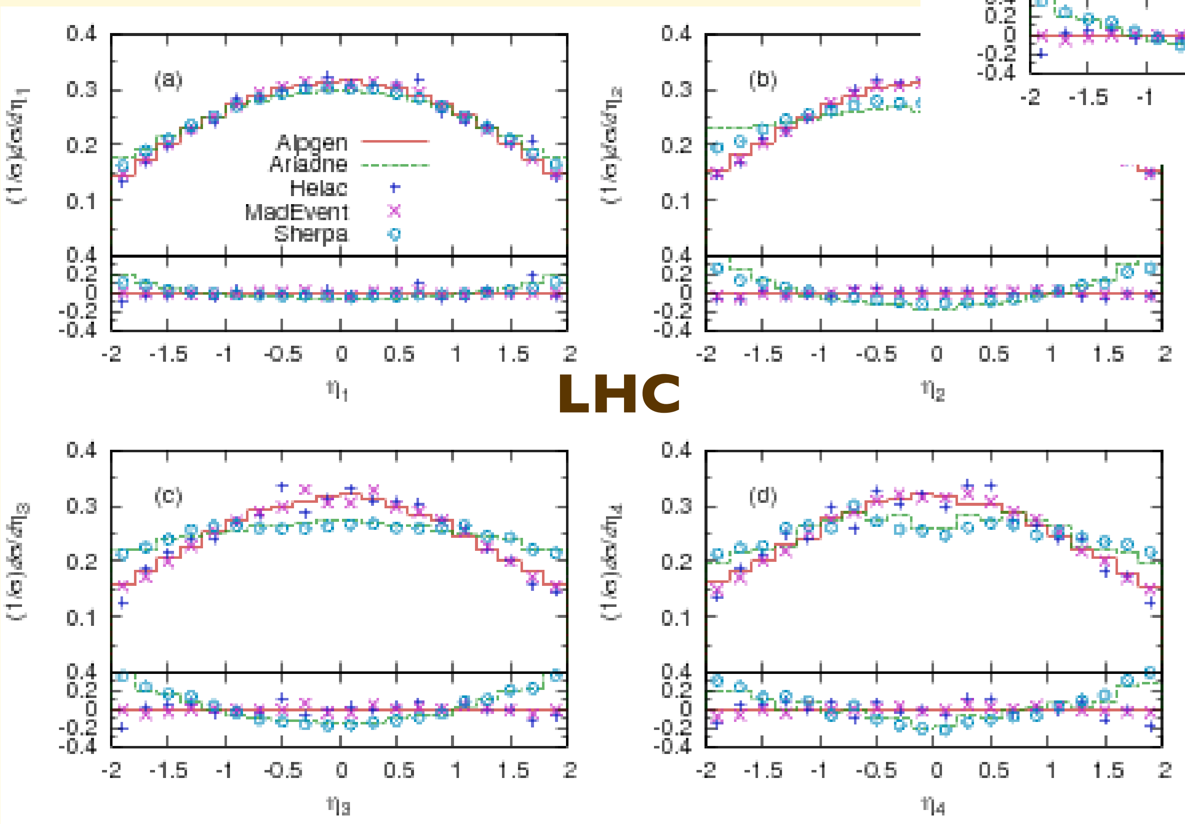


LHC

Jet rapidity distributions



Tevatron

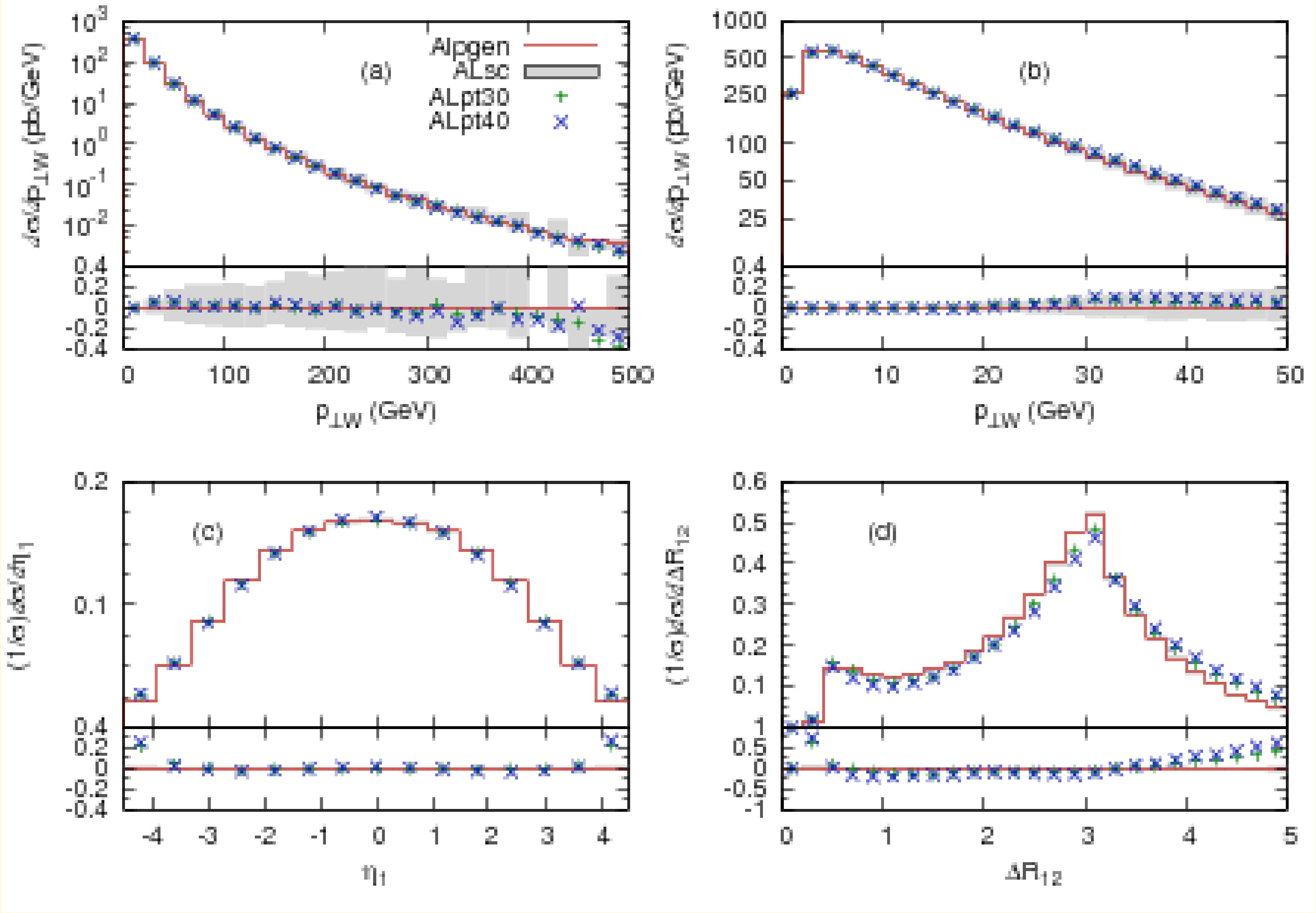


LHC

Examples of systematics studies (LHC Energy)

Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions *

J. Alwall[1], S. Höche[2], F. Krauss[2], N. Lavesson[3], L. Lönnblad[3], F. Maltoni[4], M.L. Mangano[5], M. Moretti[6], C.G. Papadopoulos[7], F. Piccinini[8], S. Schumann[9], M. Treccani[6], J. Winter[9], M. Worek[10,11]

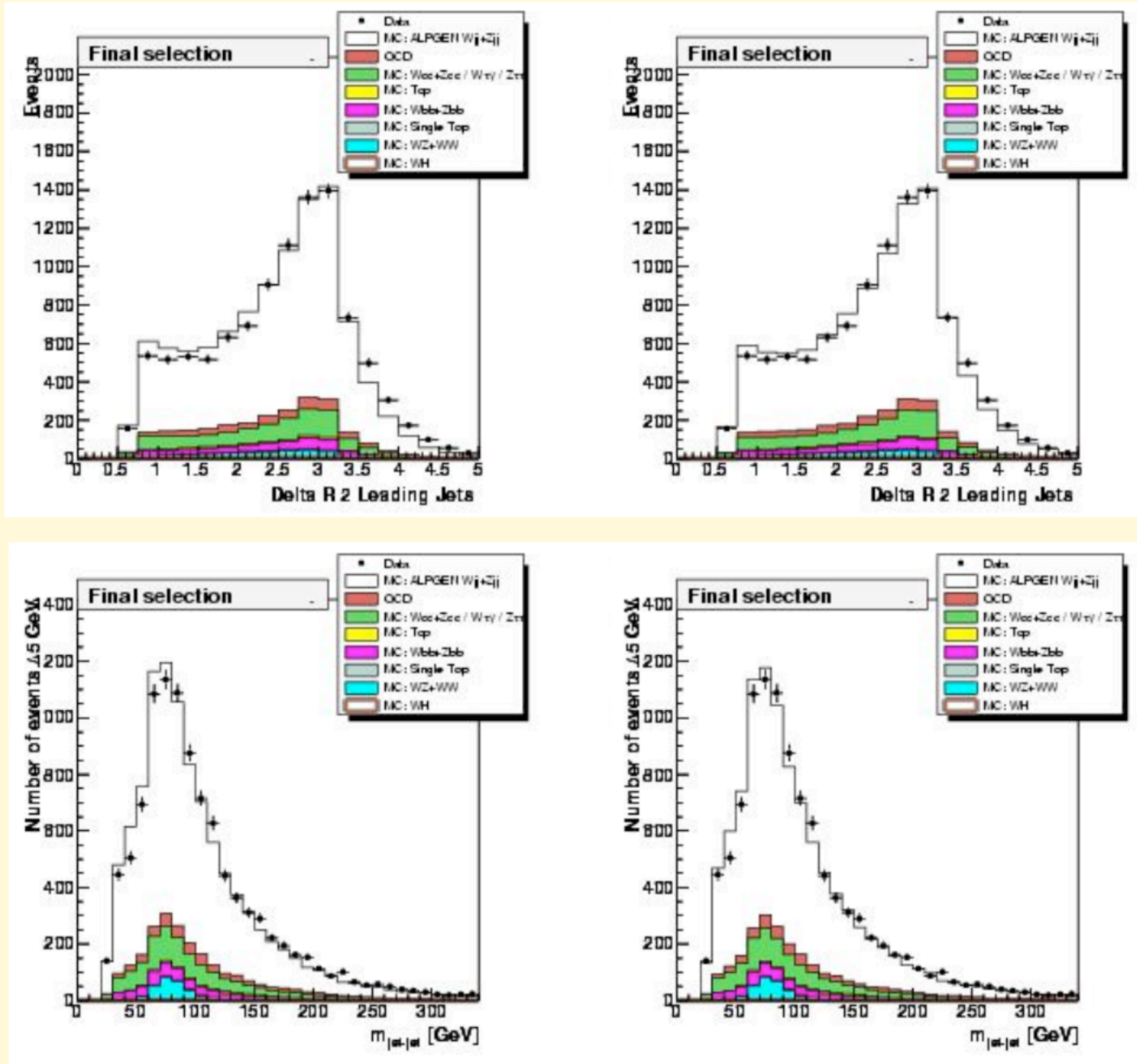


Different E_T matching thresholds: **20**, **30**, **40** GeV

Different renorm. scale factor [0.5–2]=

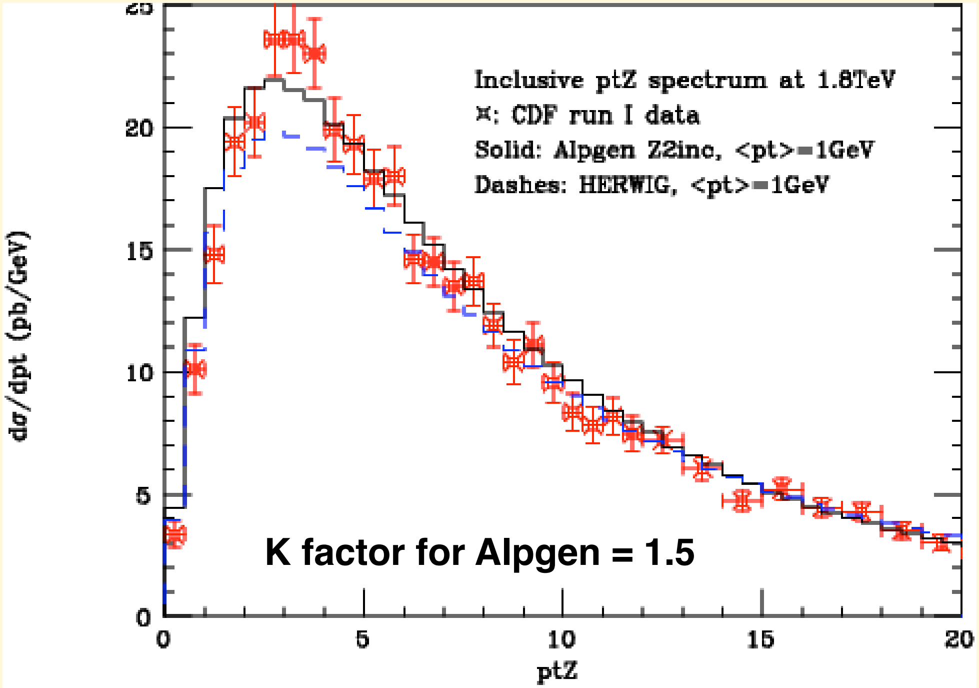
D0, kinematical reweigthing in Wjj events

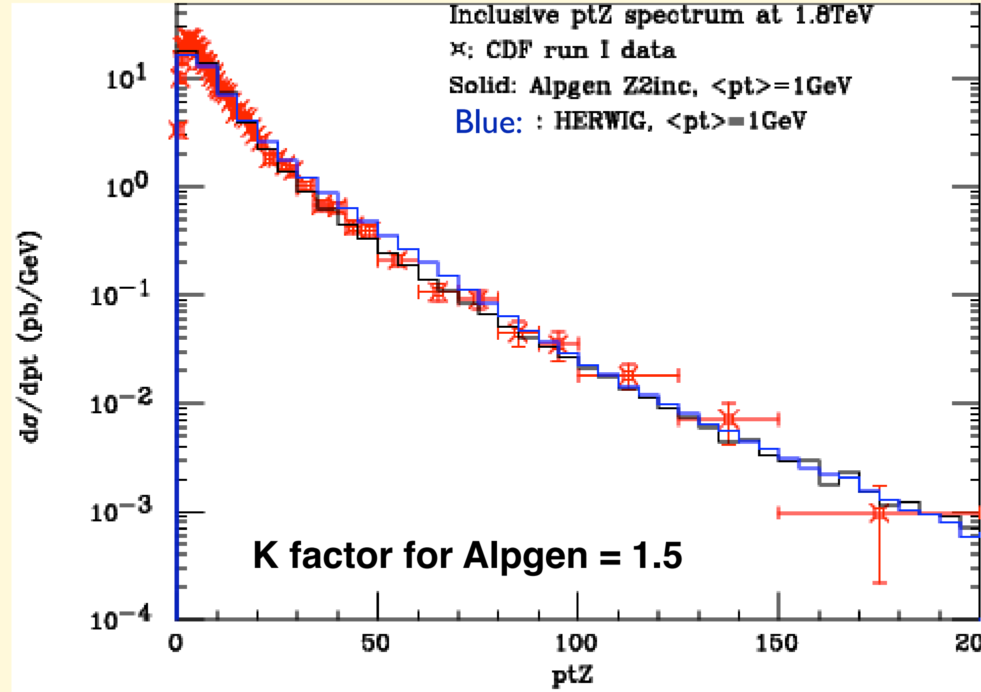
J. Lellouch, PhD thesis

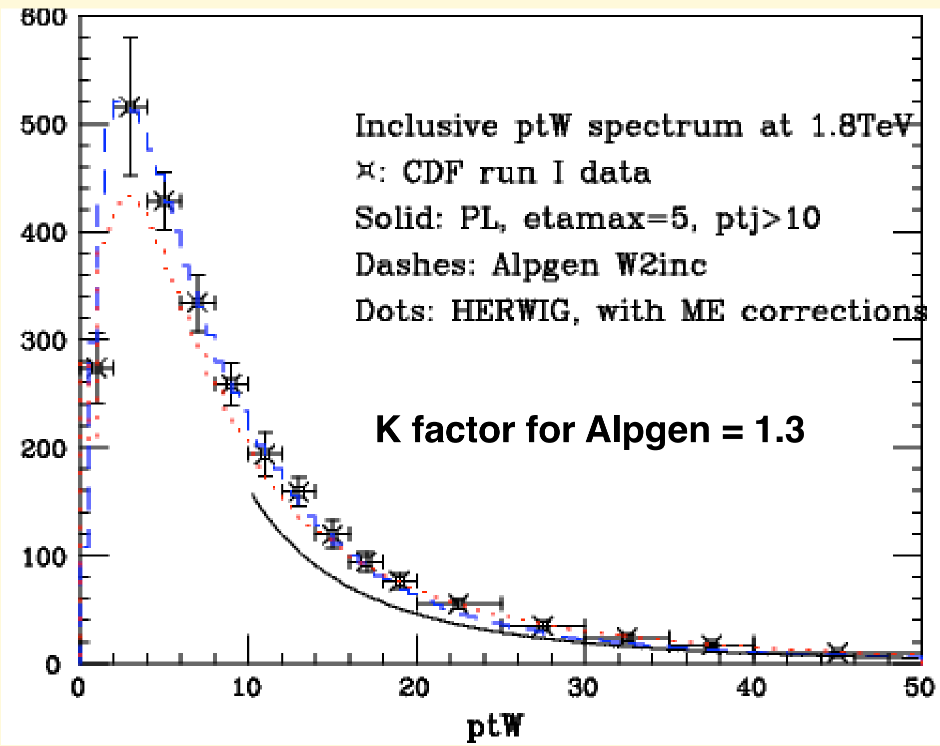


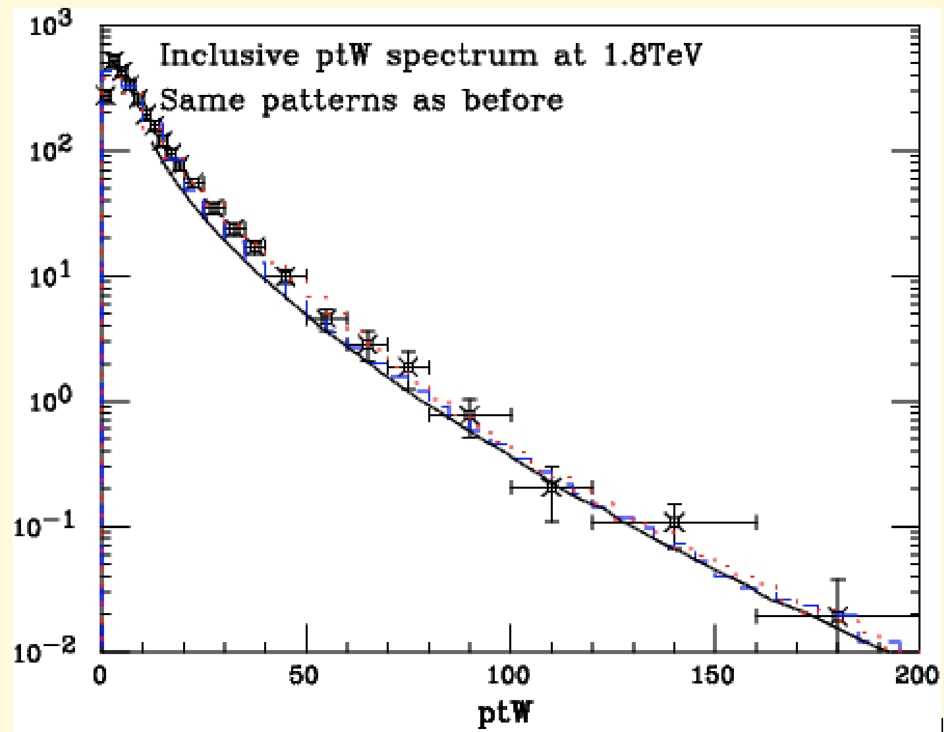
M[jj]	**M[jj]**
Before rewgt	After rewgt

Comparisons with data: Inclusive Z/W pt spectrum at 1.8 TeV (CDF data)



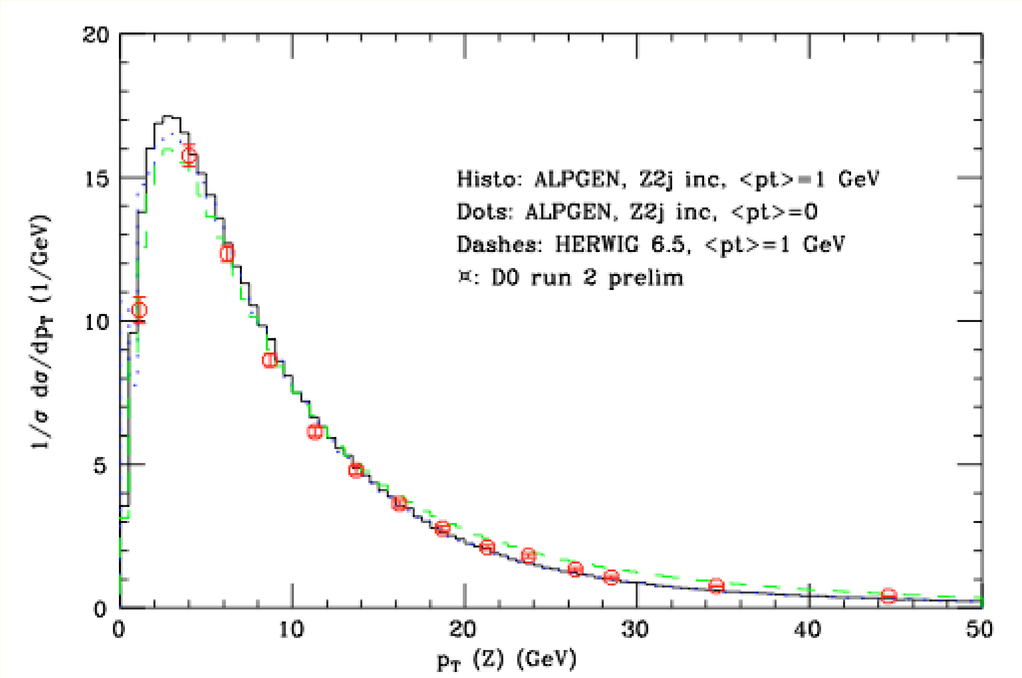


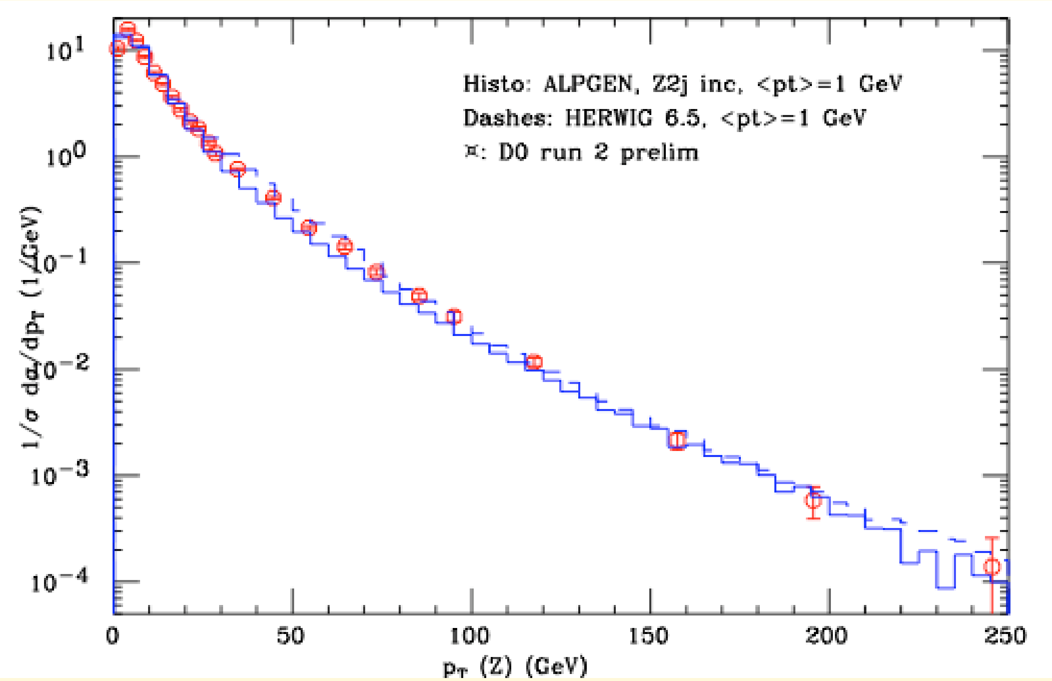




Comparisons with D0 data: Inclusive Z pt spectrum at 1.96 TeV

(D0 data: http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/EW/E22/E22.pdf)

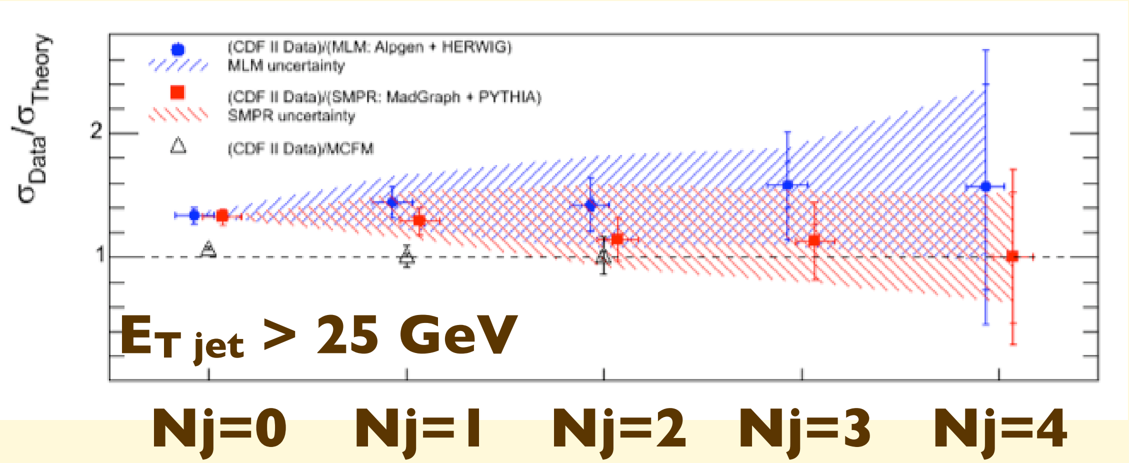


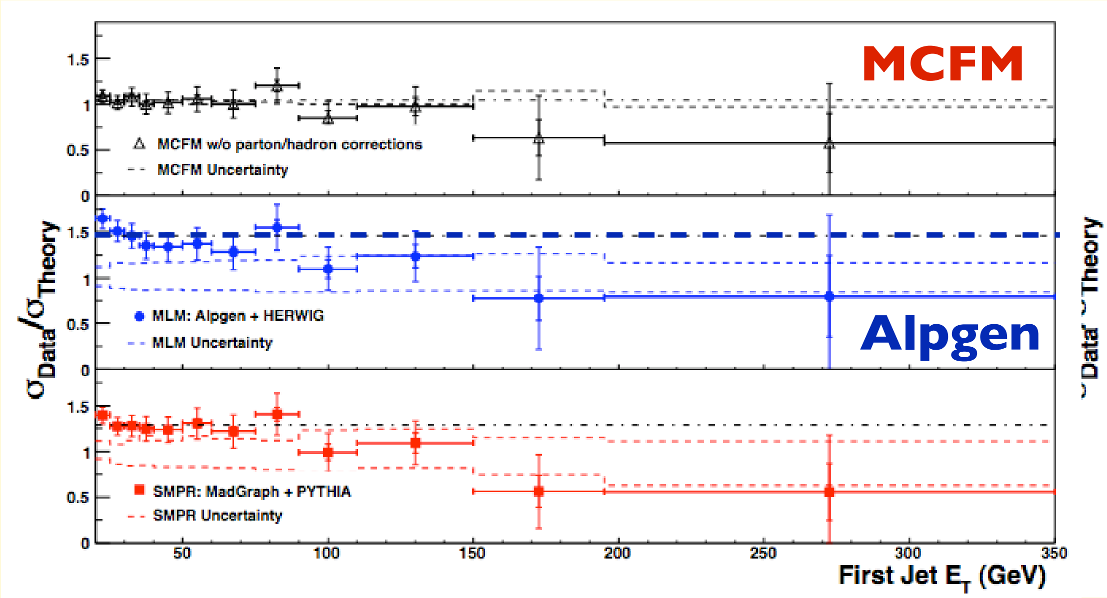


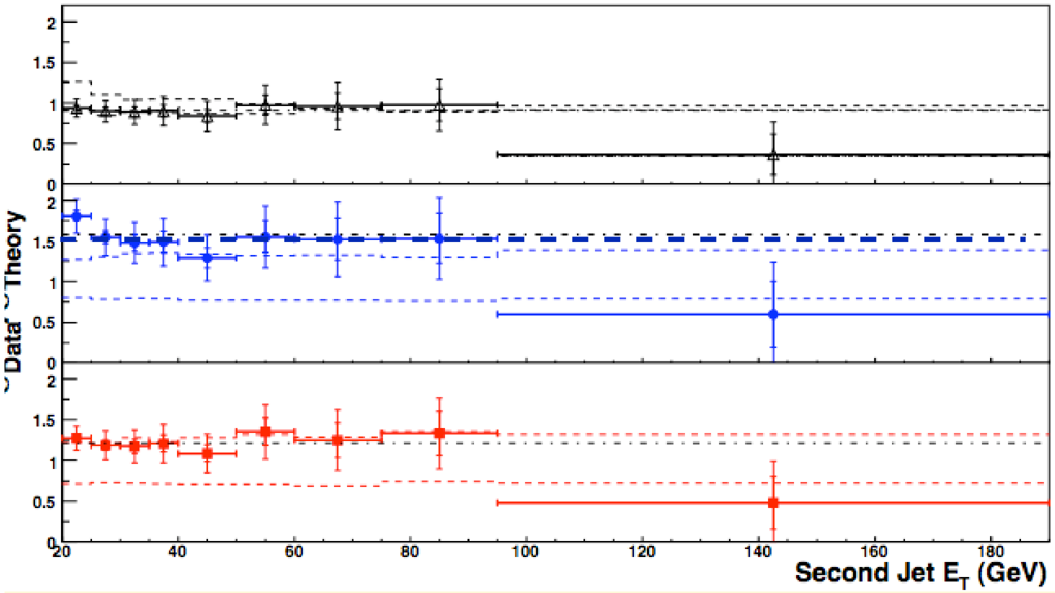
NB: once we allow the Z to decay to neutrinos, this distribution corresponds to the SM expectation for the missing E_T spectrum (after inclusion of the appropriate BR($Z \to \nu\nu$))

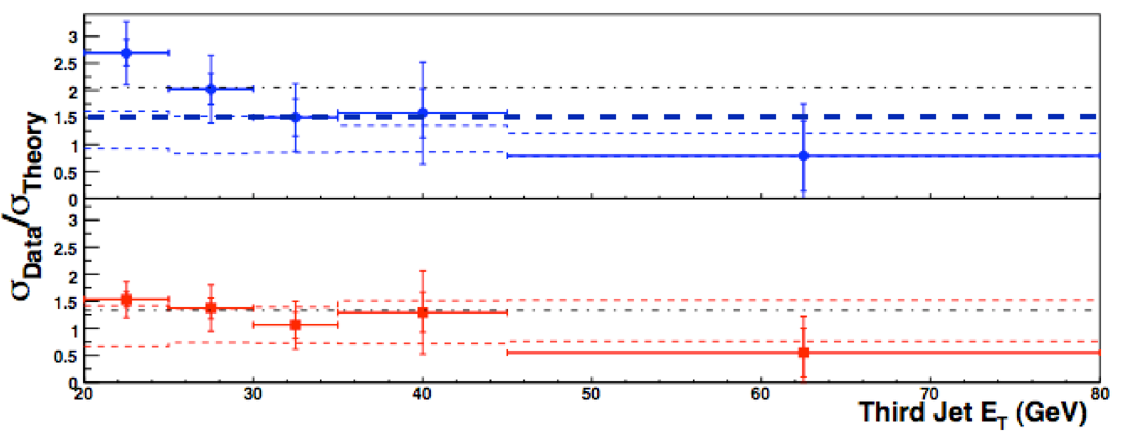
Jet spectra in W+jets CDF, 380pb^{-1}

PRD 77, 011108 (2008)



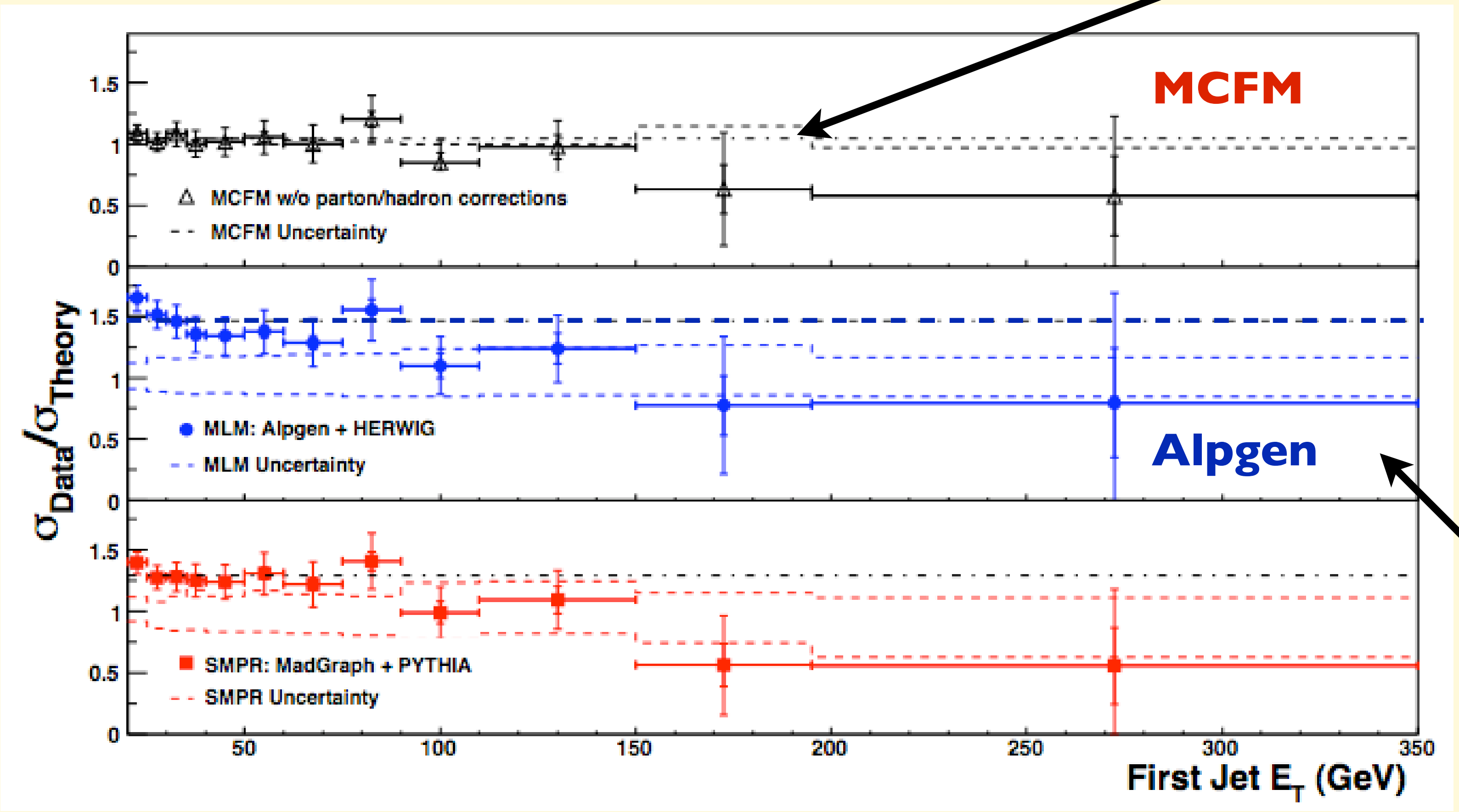






NB

This comparison takes yrs to do: you need to measure a **cross section**



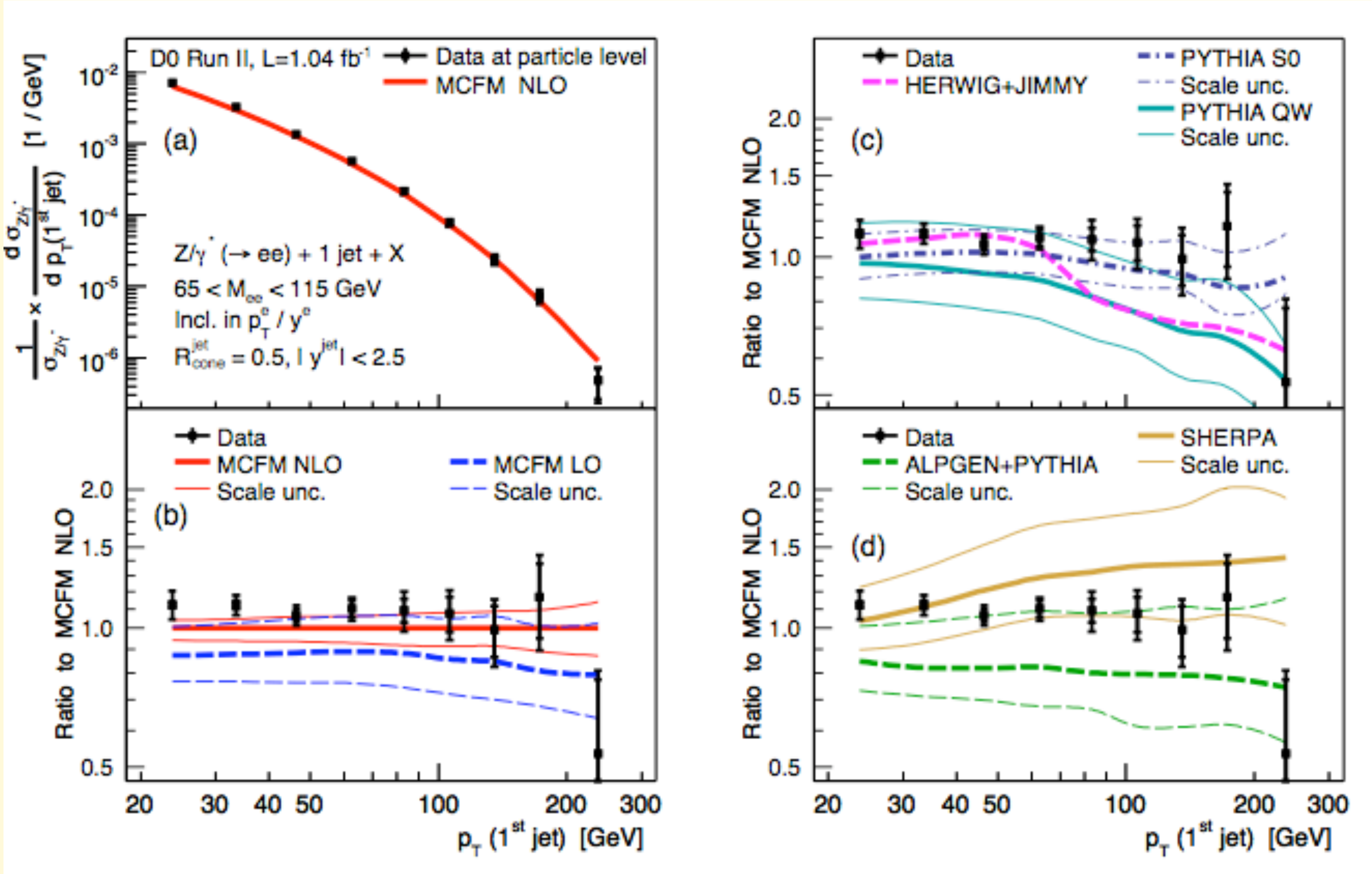
This comparison can be obtained rather soon -- just run the MC event output through the detector simulation

D0 Z+jets, 1fb^{-1}

arXiv:0903.1748

- $E_{T\ jet}$ > 20 GeV , |$\eta_{\ jet}$ | < 2.8 , R=0.5

Jet 1

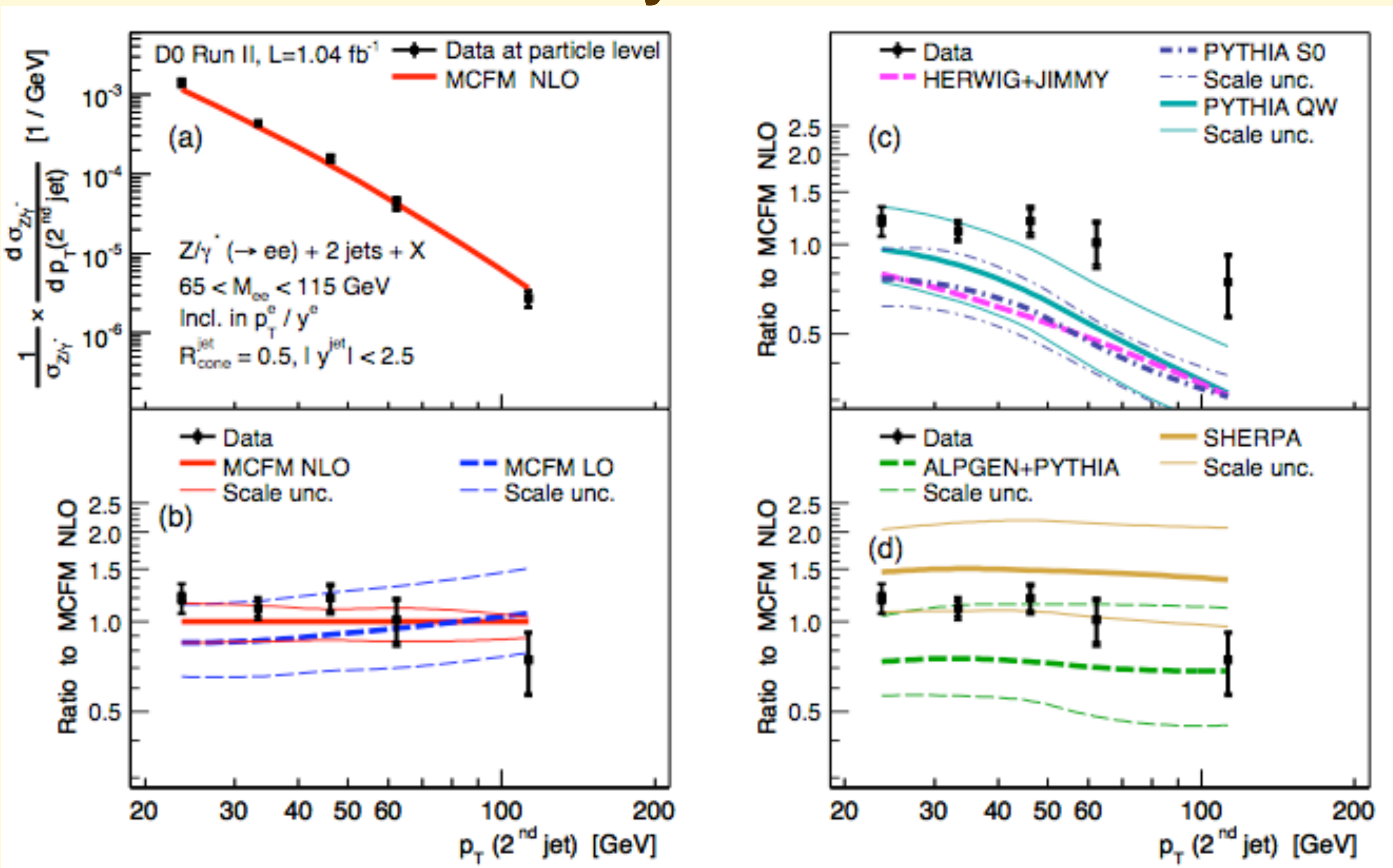


D0 Z+jets, 1fb^{-1}

arXiv:0903.1748

- $E_{T\,jet} > 20$ GeV , $|\eta_{jet}| < 2.8$, R=0.5

Jet 2

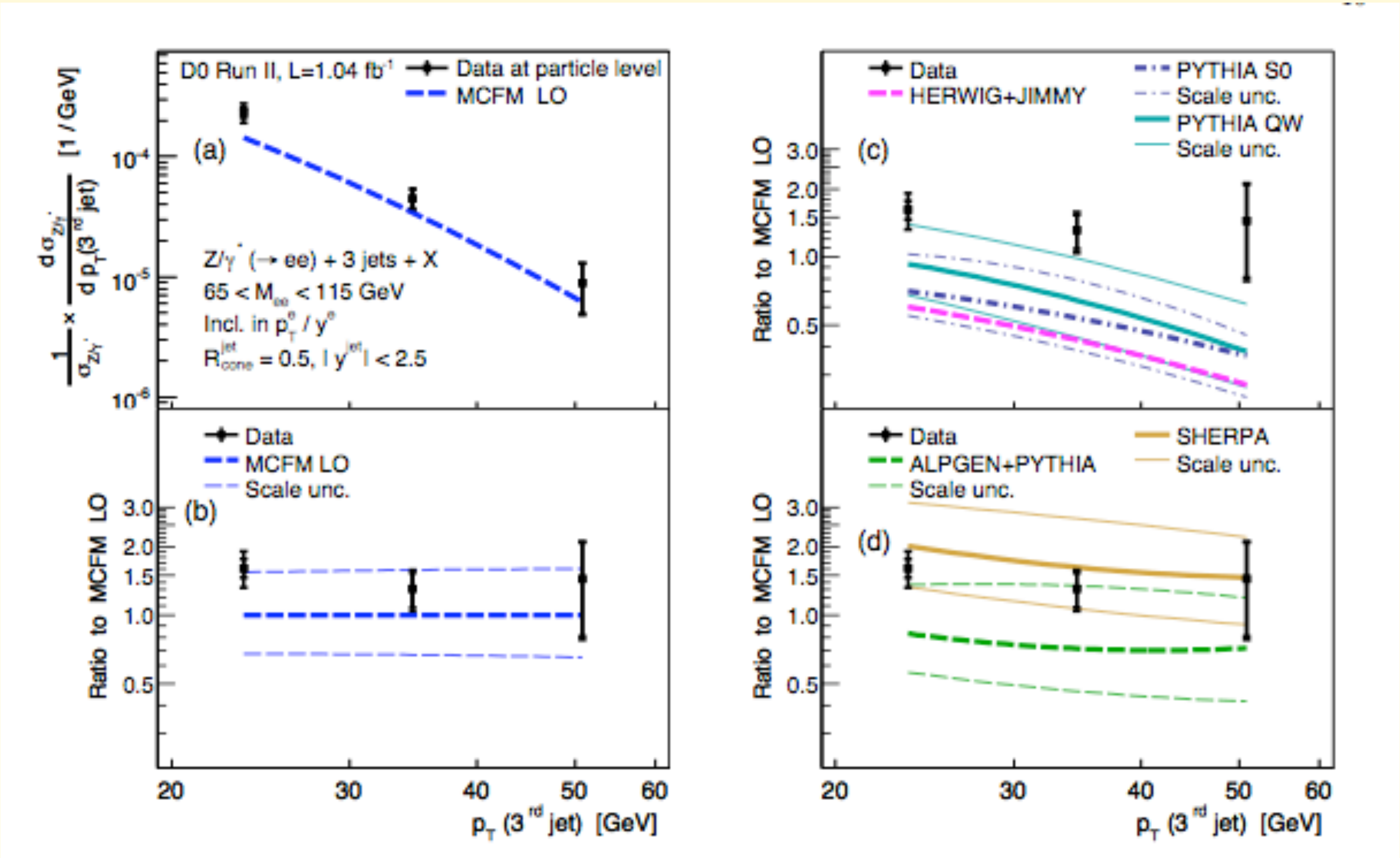


D0 Z+jets, 1fb^{-1}

arXiv:0903.1748

- $E_{T\ jet}$ > 20 GeV , $|\eta_{\ jet}|$ < 2.8 , R=0.5

Jet 3



W/Z+jets, bottom line

- NLO OK (but available only up to 2 jets, and partially up to 3*).
- LO+showers:
  - need K factor
  - shapes ~OK to within some 20-30%
  - Fine tuning at the level of 20-30% of the matching algorithms, scale choices, etc to achieve better agreement, without the need of reweigthing
  - Use NLO as a benchmark to validate, tune and normalize the LO +shower predictions, before experimental analyses are ready?

* R.K.Ellis etal, arXiv:0901.4101; C.Berger et al, arXiv:0902.2760,

W/Z+jets, bottom line

- NLO OK (but available only up to 2 jets, and partially up to 3*).
- LO+showers:
  - need K factor
  - shapes ~OK to within some 20-30%
  - Fine tuning at the level of 20-30% of the matching algorithms, scale choices, etc to achieve better agreement, without the need of reweigthing
  - Use NLO as a benchmark to validate, tune and normalize the LO +shower predictions, before experimental analyses are ready?

Things start getting a bit more complex when introducing heavy quarks ....

* R.K.Ellis etal, arXiv:0901.4101; C.Berger et al, arXiv:0902.2760,

W+charm

CDF analysis

- $p_{T charm} > 20$ GeV
- $|\eta_{charm}| < 1.5$

	σ_{Wc} x BR(W→e nu) [pb]
CDF	9.8±3.2
LO ($Q^2=M_W^2+p_T^2$)	6.80
LO ($Q^2=p_T^2$)	8.75

W+charm

CDF analysis

- p_{Tcharm} > 20 GeV
- $|\eta_{charm}|$ < 1.5

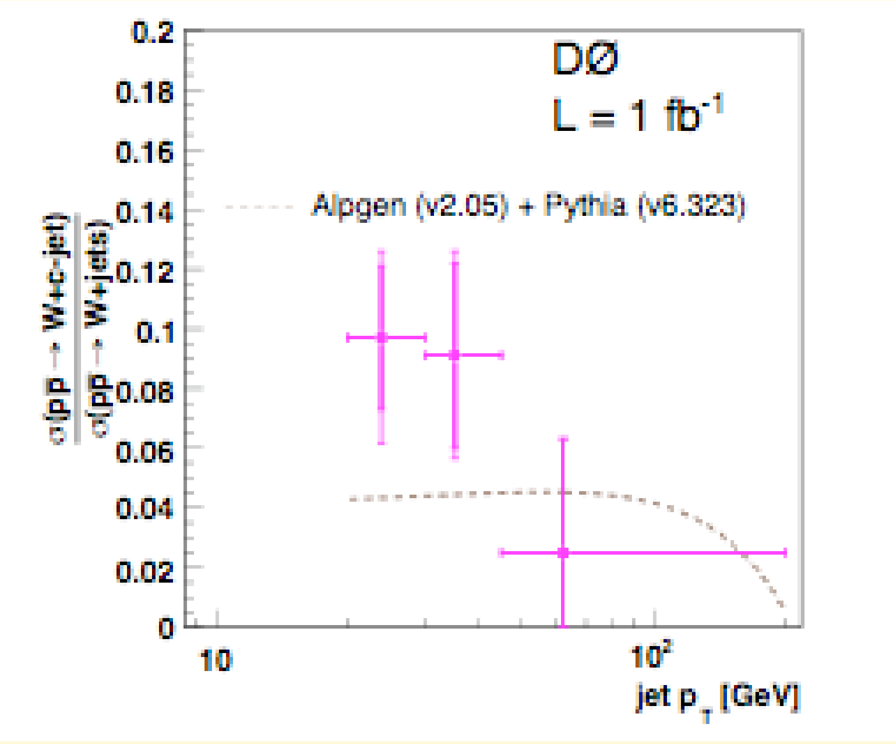
	σ_{Wc} x BR(W→e nu) [pb]
CDF	9.8±3.2
LO ($Q^2=M_W^2+p_T^2$)	6.80
LO ($Q^2=p_T^2$)	8.75

D0 analysis

- E_{Tjet} > 20 GeV (typically p_{Tcharm} << 20 GeV)
- $|\eta_{charm}|$ < 2.5

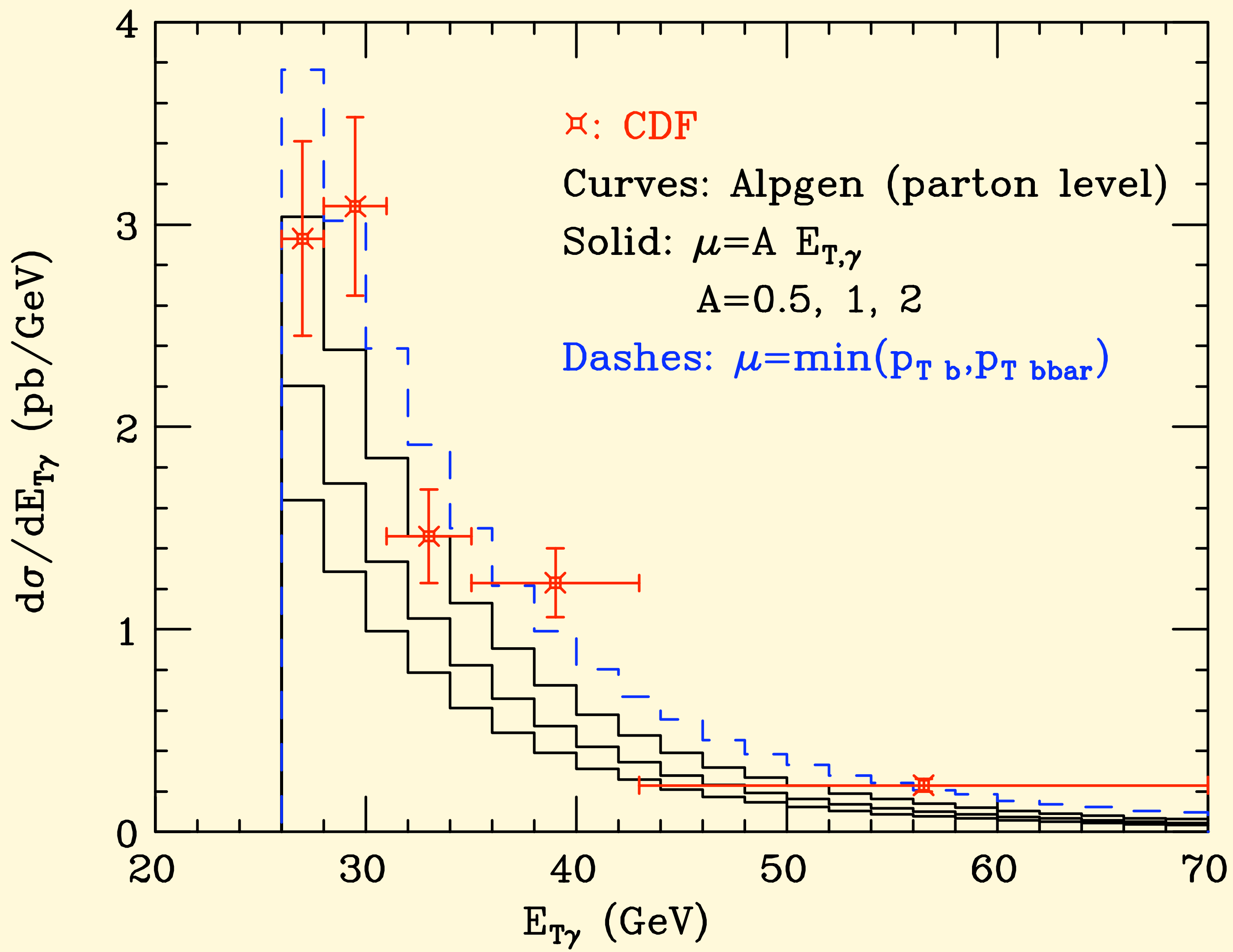
D0: $$\frac{\sigma[W+c-\text{jet}]}{\sigma[W+\text{jet}]} = 0.074 \pm 0.019 \pm 0.013$$

Alpgen: $$\frac{\sigma[W+c-\text{jet}]}{\sigma[W+\text{jet}]} = 0.044 \pm 0.003$$



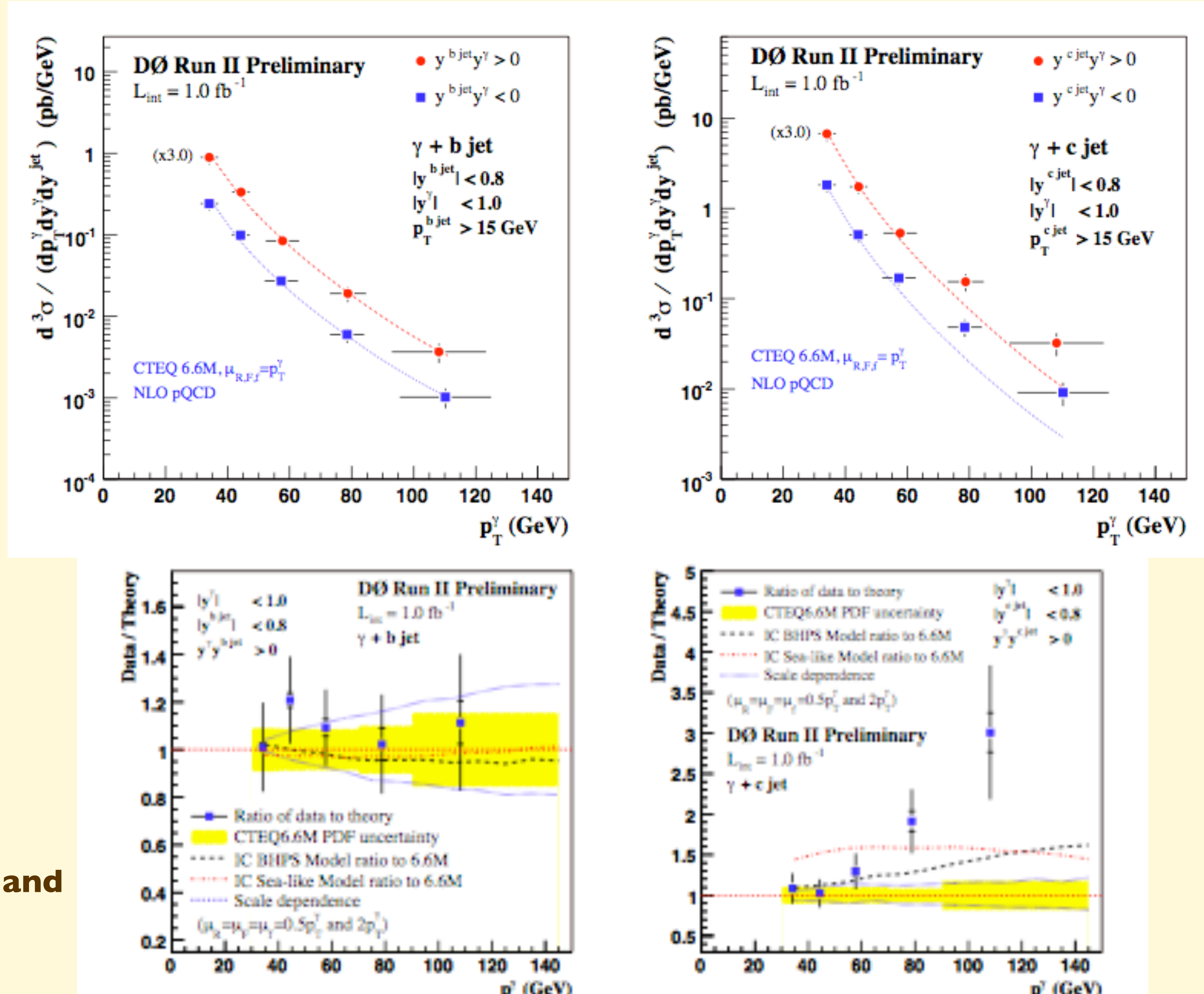
CDF: γ+b-jet analysis

- $E_{T\gamma} > 26$ GeV , $|\eta_\gamma| < 1.1$
- $E_{T\,jet} > 20$ GeV , $|\eta_{jet}| < 1.5$, R=0.7



D0: γ+b/c-jet analysis, 1 fb^{-1} ICHEP08, arXiv:0810.3754

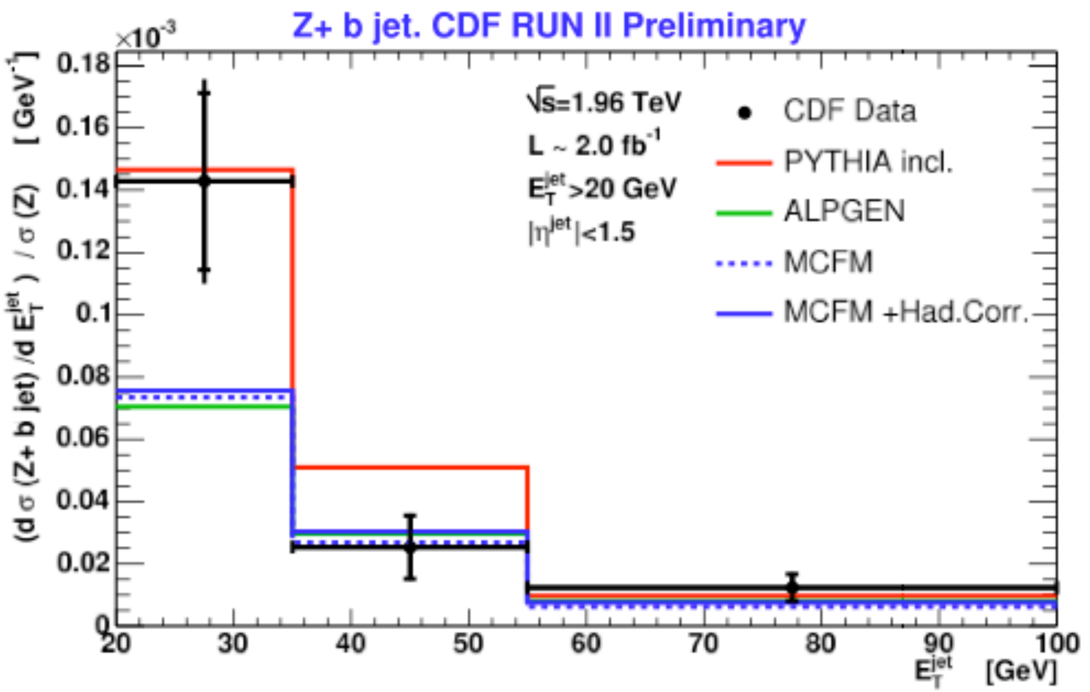
- $E_{T\gamma} > 30$ GeV , $|\eta_\gamma| < 1$, $E_{T\,jet} > 15$ GeV , $|\eta_{jet}| < 0.8$, R=0.5



NLO QCD: T.Stavreva and J.Owens

Z+b-jet , CDF 2fb^{-1}

ICHEP08, arXiv:0810.2914



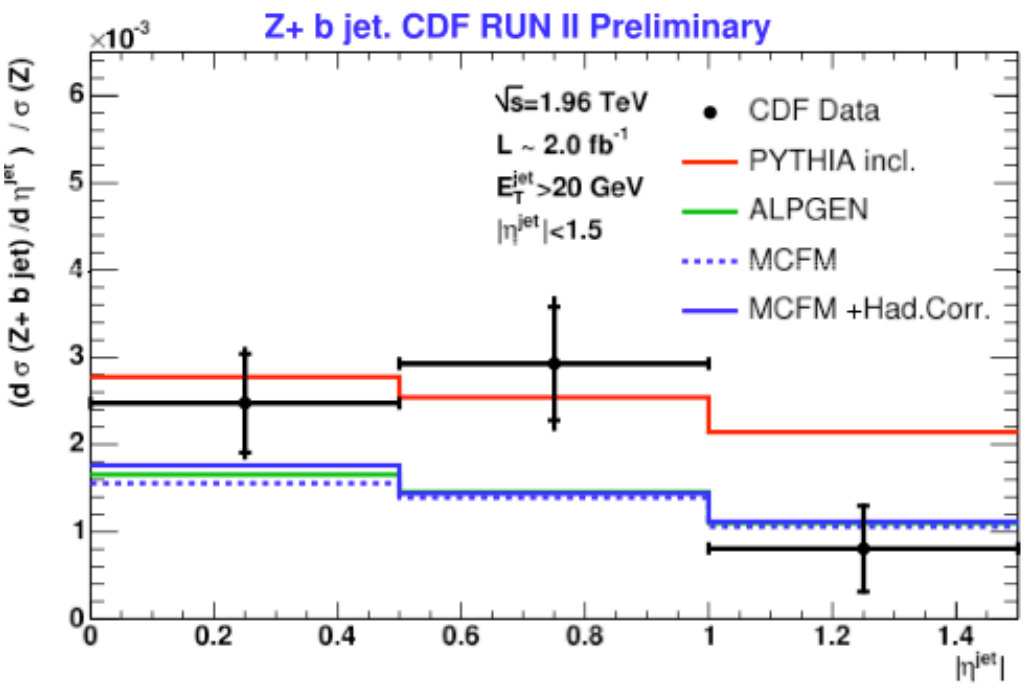


Table I: Results on the $Z + b$-jets production.

	CDF Data	Pythia	Alpgen	MCFM NLO	MCFM NLO+UE +Hadronization
$\sigma(Z+b\text{-jet})$	$0.86 \pm 0.14 \pm 0.12$ pb	–	–	0.51 pb	0.53 pb
$\sigma(Z+b\text{-jet})/\sigma(Z)$	$0.336 \pm 0.053 \pm 0.041\%$	0.35%	0.21%	0.21%	0.23%
$\sigma(Z+b\text{-jet})/\sigma(Z+\text{jet})$	$2.11 \pm 0.33 \pm 0.34$ %	2.18%	1.45%	1.88%	1.77%

ALPGEN and NLO agree with each other, but neither agrees well with data ....

Z+b-jet

CDF analysis

ICHEP08, arXiv:0810.2914

- $E_{T\ jet}$ > 20 GeV
- $|\eta_{\ jet}|$ < 1.5
- R=0.7

$$\left[\frac{\sigma[Z+b-\text{jet}]}{\sigma[Z+\text{jet}]}\right]_{CDF} = 2.35 \pm 0.6\%$$

$$\left[\frac{\sigma[Z+b-\text{jet}]}{\sigma[Z+\text{jet}]}\right]_{Alp-PL} = 1.6\% \qquad Q^2 = M_Z^2 + \sum_{i=b,\bar{b}} m_{i,T}^2$$

$$\left[\frac{\sigma[Z+b-\text{jet}]}{\sigma[Z+\text{jet}]}\right]_{Alp-PL} = 2.3\% \qquad (Q^2 = \sum_{b,\bar{b}} p_T^2)$$

In the numerator only

D0 analysis

PRL 94, 161801(2005)

- $E_{T\ jet}$ > 20 GeV
- $|\eta_{\ jet}|$ < 2.5
- R=0.5

$$\left[\frac{\sigma[Z+b-\text{jet}]}{\sigma[Z+\text{jet}]}\right]_{D0} = 2.3 \pm 0.4\%$$

$$\left[\frac{\sigma[Z+b-\text{jet}]}{\sigma[Z+\text{jet}]}\right]_{Alp-PL} = 1.5\% \qquad Q^2 = M_Z^2 + \sum_{i=b,\bar{b}} m_{i,T}^2$$

$$\left[\frac{\sigma[Z+b-\text{jet}]}{\sigma[Z+\text{jet}]}\right]_{Alp-PL} = 2.3\% \qquad (Q^2 = \sum_{b,\bar{b}} p_T^2)$$

In the numerator only

W+b-jet

CDF analysis

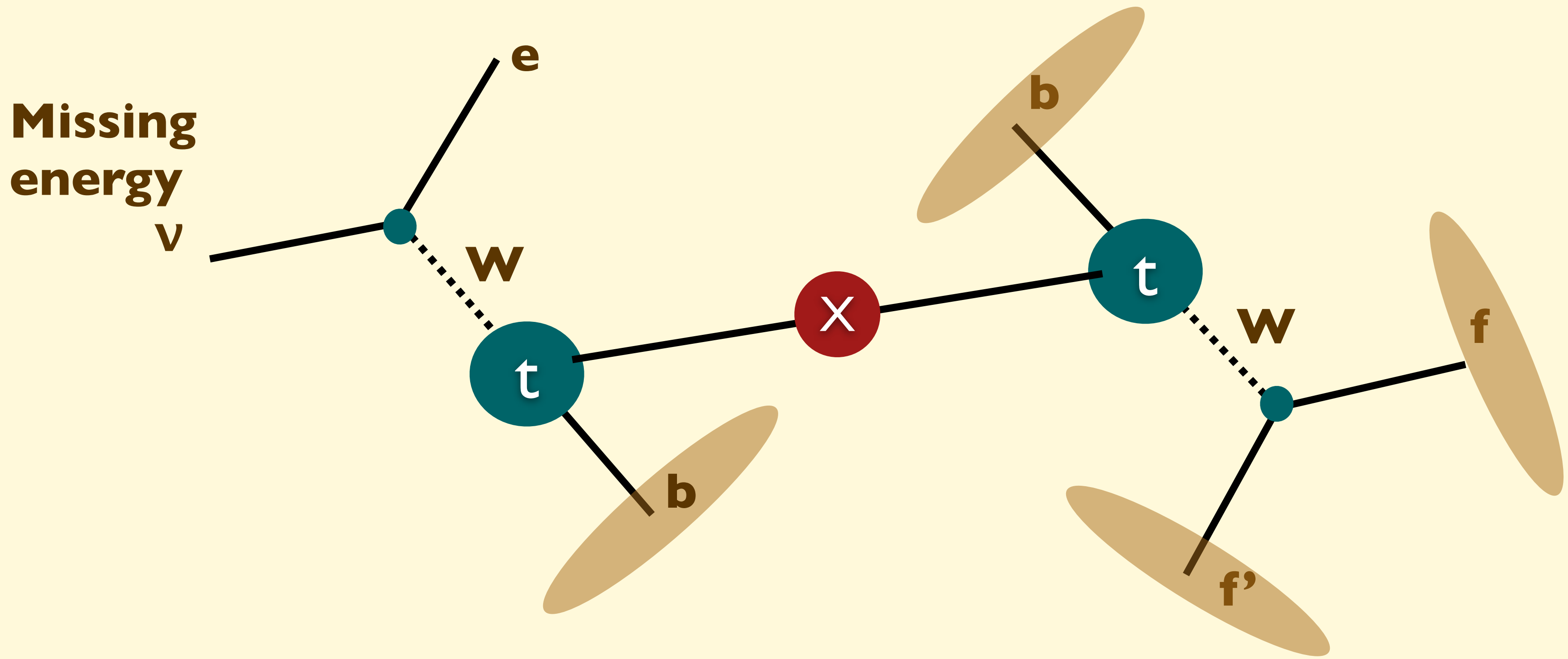
- $p_{T\ lepton}$ > 20 GeV , $|\eta_{lepton}|$ < 1.1 MET>25 GeV
- $p_{T\ jet}$ > 20 GeV , $|\eta_{jet}|$ < 2 , R=0.4

	σ_{Wb} x BR(W→e nu) [pb]
CDF	2.74 ± 0.27 (stat) ± 0.42 (syst)
PL LO, Wbb ($Q^2=M_W^2+p_T^2$)	0.78
Wbb+ Wbb1jet MLM matching with Herwig	$[0.504]_{Wbb}$ +$[0.126]_{Wbbj}$ =0.73

C. Neu, ICHEP08, arXiv: 0809.1407

Data/Theory ~ 3.5 !!

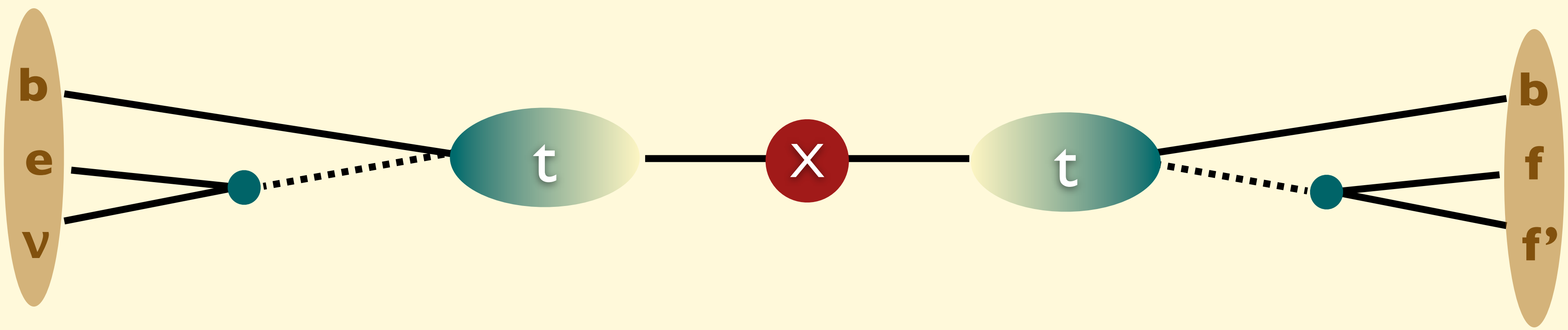
Top final states, from signal at the Tevatron to bg at the LHC



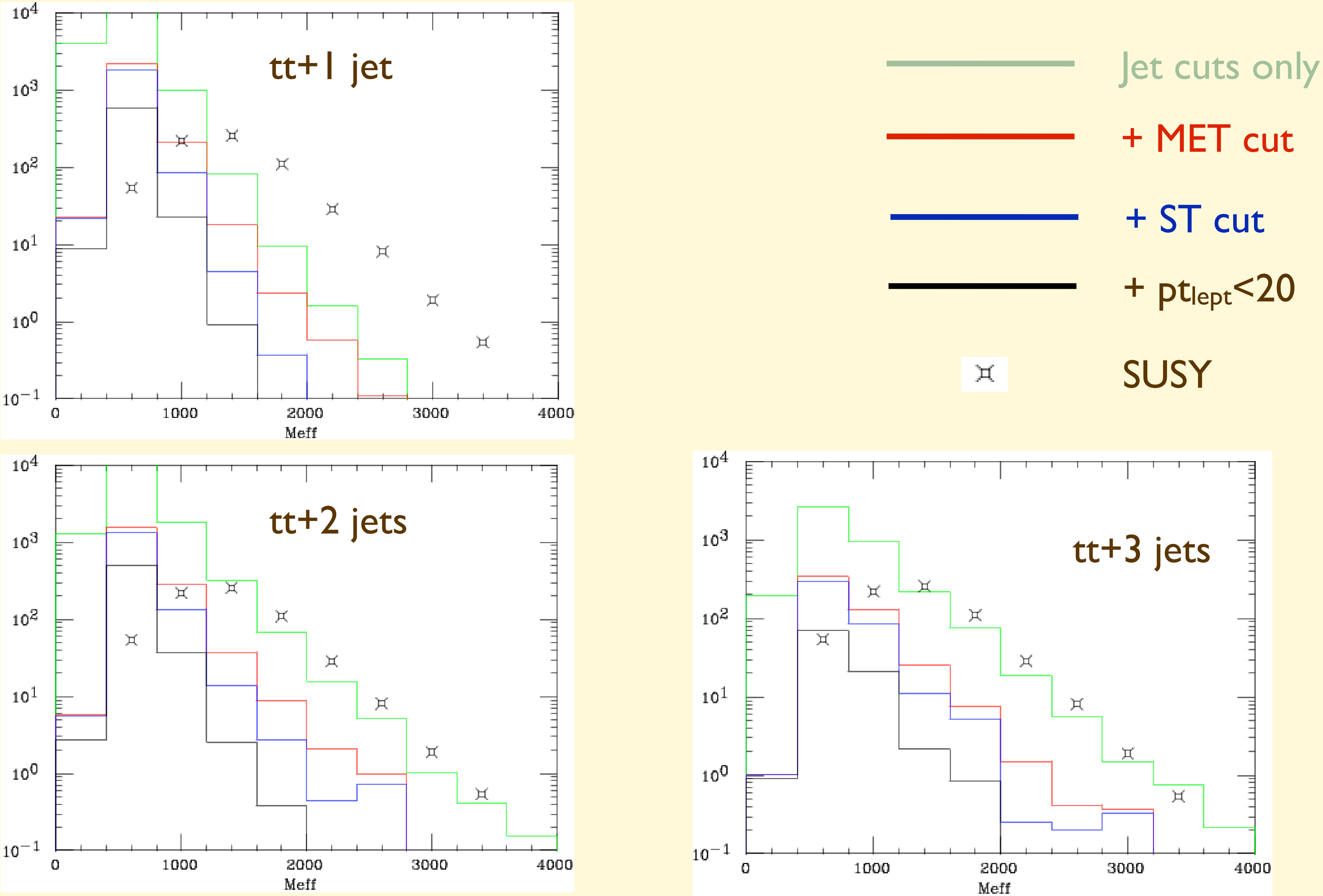
Top final states

Large MET / Meff in top events => large pt[top]

=> highly collimated final states



In SUSY searches, sphericity and multi-jet cuts very effective against the **leading-order** t-tbar contribution!



All jet multiplicities contribute at approximately the same level!!

No validation of predictions for such final states is available from the Tevatron, due to low rates

Things I didn't cover

- Top mass systematics
  - non-PT effects (colour recombination, hadronization, etc)
  - start having now enough lum at Tevatron for first studies
- Limiting systematics for precision EW measurements (M_W, $\sin^2\theta_W$):
  - EW/QED corrections to initial and final states
  - PDF systematics
- Vector-boson-fusion final states, and relative bg's
  - not enough energy/statistics at the Tevatron: whole new terrain of exploration at the LHC
- Diffractive hard processes (e.g. Higgs production)
- ....

Conclusions

- A better picture of associated production of gauge bosons and jets is emerging from the Tevatron data, and the tools are becoming mature
- This picture is however still incomplete
  - statistics still limited for quantitative studies of
    - heavy quark content
    - highest jet multiplicities, particularly in the Z case
  - no global analysis (e.g. study of consistency of the data vs MC comparisons over different channels)
- Within the limited statistics the pattern of (dis)agreements between theory (LO+parton showers, NLO) and data is still unclear

Conclusions

- Need to address more quantitatively the "portability" of tunings from one set of final states to another (e.g. from Z+jets to W +jets)

- The definition of an overall and coherent campaign of MC testing, validation and tuning at the LHC will probably happen only once the data are available, and the first comparisons will give us an idea of how far off we are.