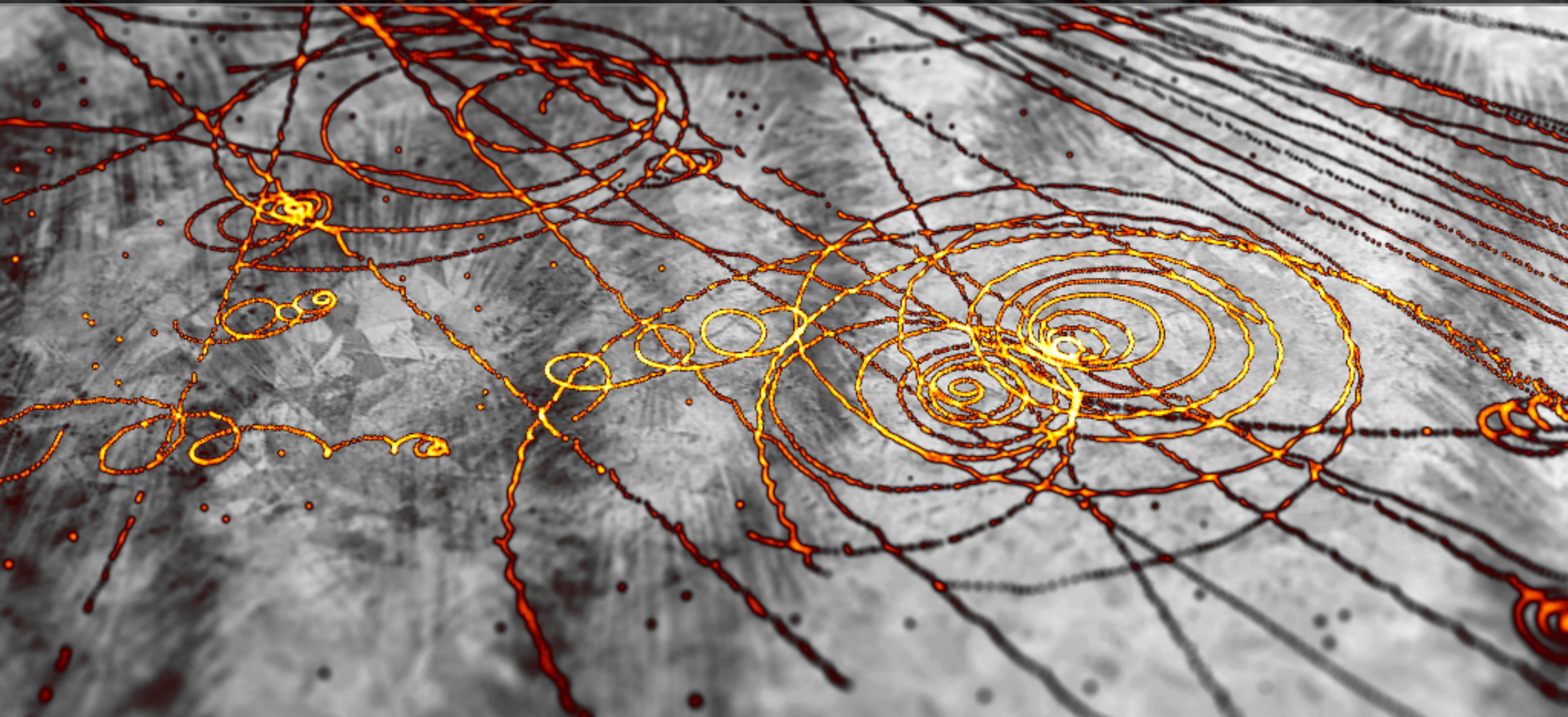
Zurich Phenomenology Workshop, January 2012

Higgs Searches - Sensitivity to Hadronization, Underlying Event, Pile-up, and MC tunings



Peter Skands (CERN)

Many plots from mcplots.cern.ch : Lot of credit to A. Karneyeu, D. Konstantinov, S. Prestel, A. Pytel

Exhortation

Final Remark from slides of M. Mangano

- Higgs-search studies are bringing in valuable information for the validation and further improvement of the tools, and further efforts should be made, alongside the discovery race, to fully exploit the potential of these data, to benefit improved tools, and further applications to studies of the Higgs once found, or other BSM searches

The SM groups in ATLAS, CMS, ALICE, and LHCb have already placed extremely valuable high-quality constraints on MC modeling. *(Thank you!)*

Main tool for propagating such constraints to MC authors and tuning:

HEPDATA and RIVET

If you look into an SM modeling aspect in the course of a Higgs or new-physics search, please consider publishing it in this form, if at all possible

Ensures that your constraints are **shared** so everyone can benefit from them

http://projects.hepforge.org/rivet/

From Partons to Pions

General-Purpose Monte Carlo models
Start from pQCD (still mostly LO). Extend towards Infrared.
HERWIG/JIMMY, PYTHIA, SHERPA, EPOS

Elastic | Min-Bias | Jets/W/Z/H/Top/…

0 ← Λ_{QCD} — 5 GeV → ∞

Elastic & Diffractive Treated as separate class No predictivity	Color Screening Regularization of pQCD Hadronization	Unitarity Showers (ISR+FSR) Multiple 2→2 (MPI)	Hard Process Perturbative 2→2 (ME) Resonance Decays

PYTHIA uses **string fragmentation**,
HERWIG, SHERPA use **cluster fragmentation**

(N)LL

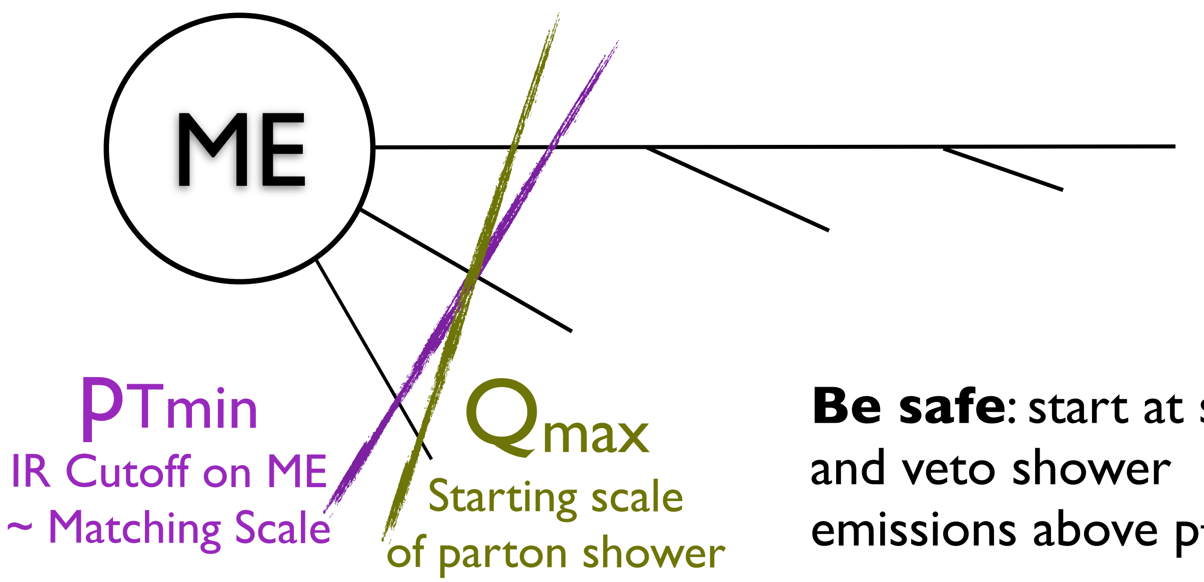
(N)LO Matching

← *Direction of this talk*

(Also possible to start from non-perturbative QCD (via optical theorem) and extend towards UV)
E.g., PHOJET, DPMJET, QGSJET, SIBYLL, … (But will not cover here)

Size of Phase Space

1) Shower Starting Scale = Matrix-element IR cutoff scale / matching scale

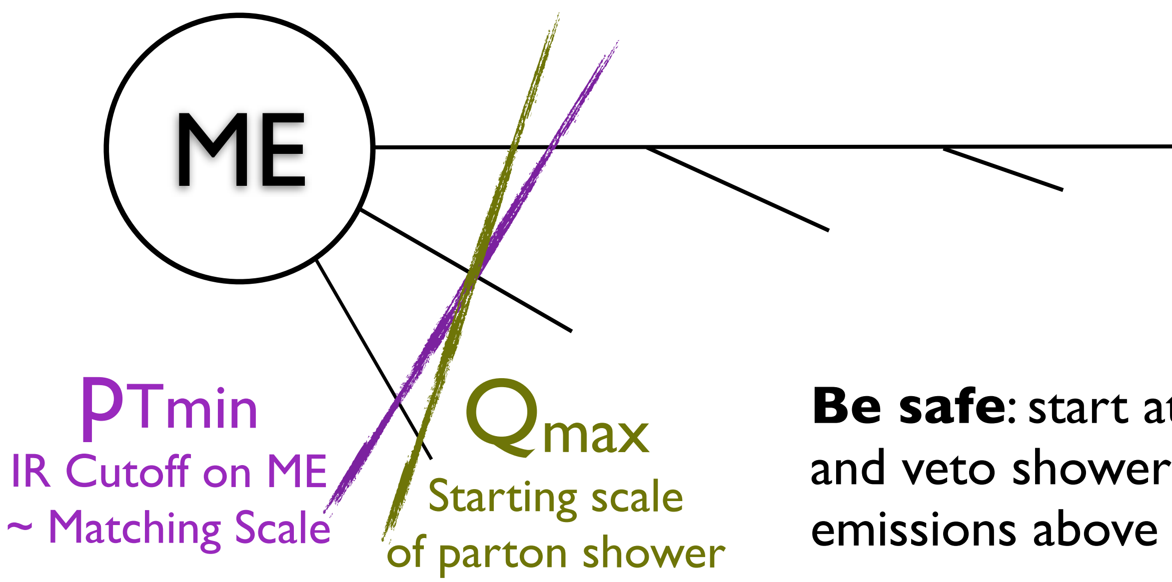


Be safe: start at s and veto shower emissions above p_{Tmin}

Size of Phase Space

1) Shower Starting Scale = Matrix-element IR cutoff scale / matching scale

Beware: multiple defiitions of p_T, Corke & Sjöstrand, EPJC69(2010)1



Be safe: start at s and veto shower emissions above p_{Tmin}

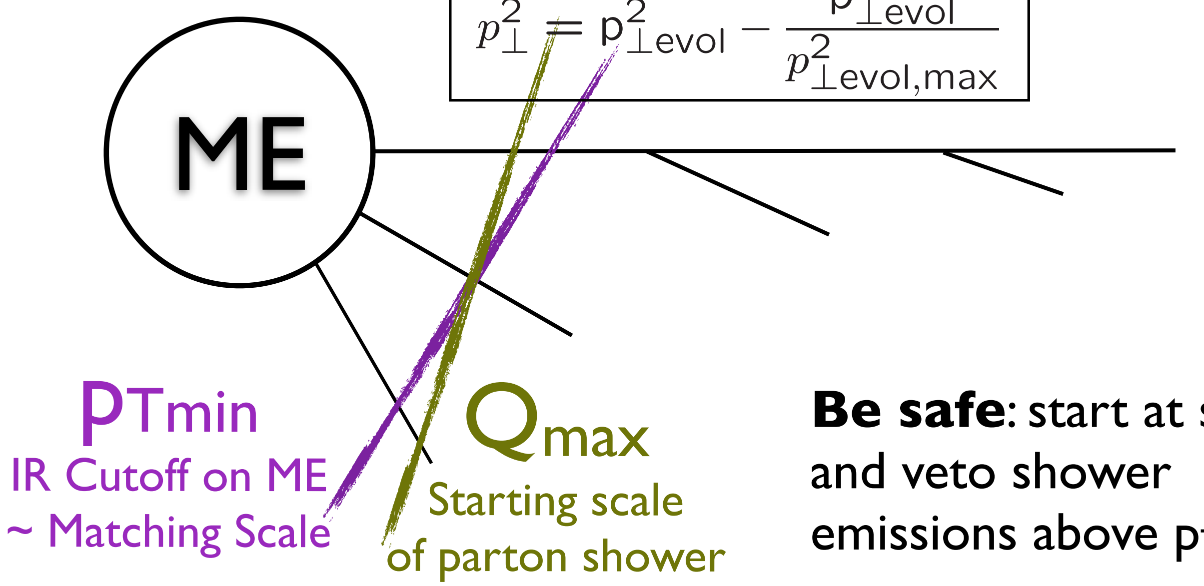
Size of Phase Space

1) Shower Starting Scale = Matrix-element IR cutoff scale / matching scale

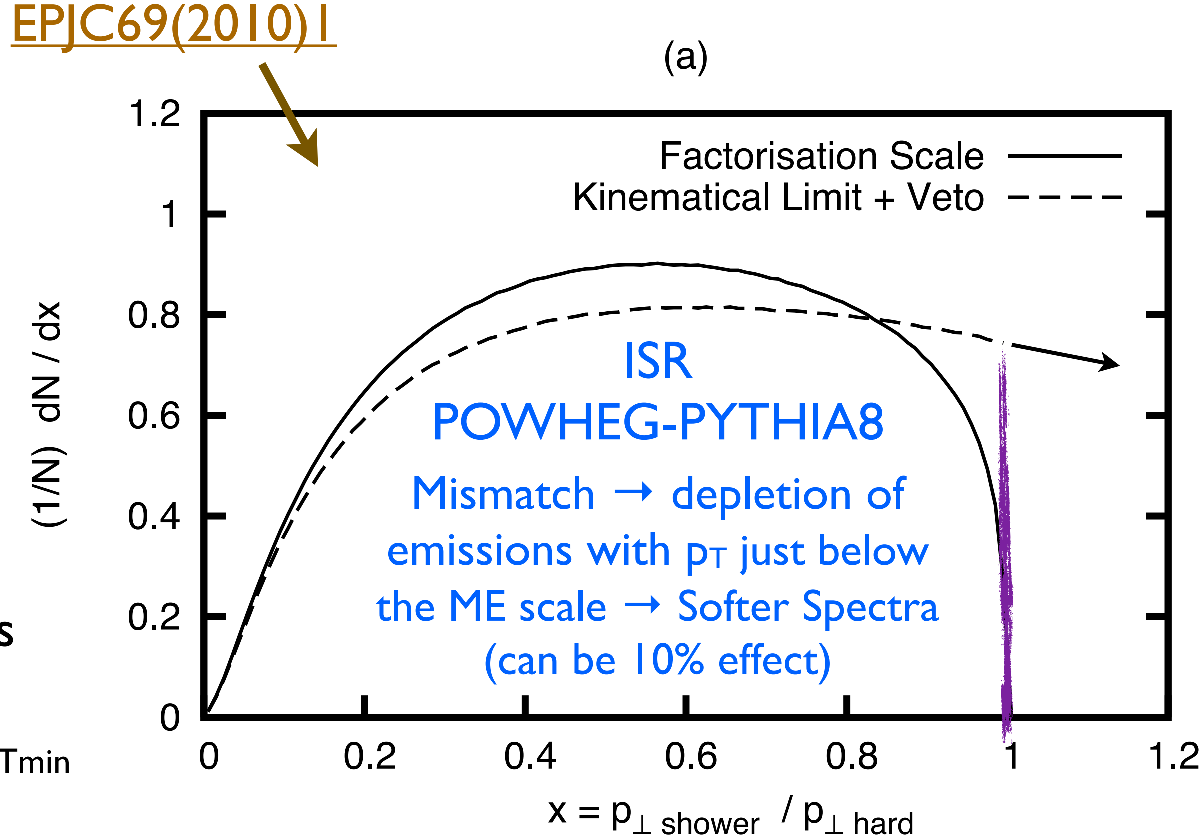
Beware: multiple defiitions of p_T, Corke & Sjöstrand, EPJC69(2010)1

Example: PYTHIA definition of p_T

$$p_\perp^2 = p_{\perp\mathrm{evol}}^2 - \frac{p_{\perp\mathrm{evol}}^4}{p_{\perp\mathrm{evol,max}}^2}$$



Be safe: start at s and veto shower emissions above $p_{T\min}$



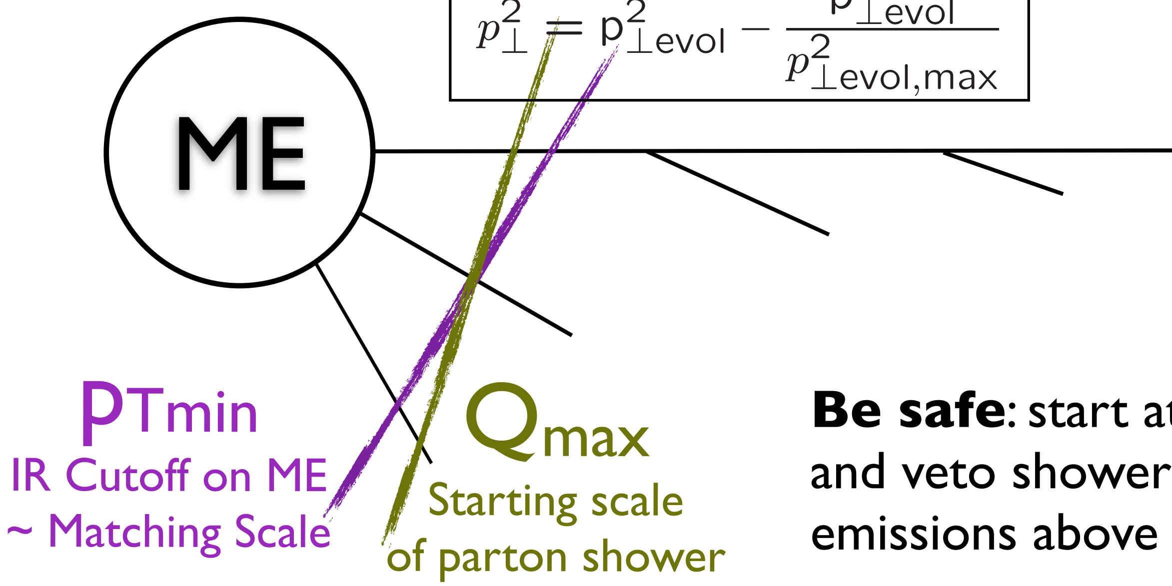
Size of Phase Space

1) Shower Starting Scale = Matrix-element IR cutoff scale / matching scale

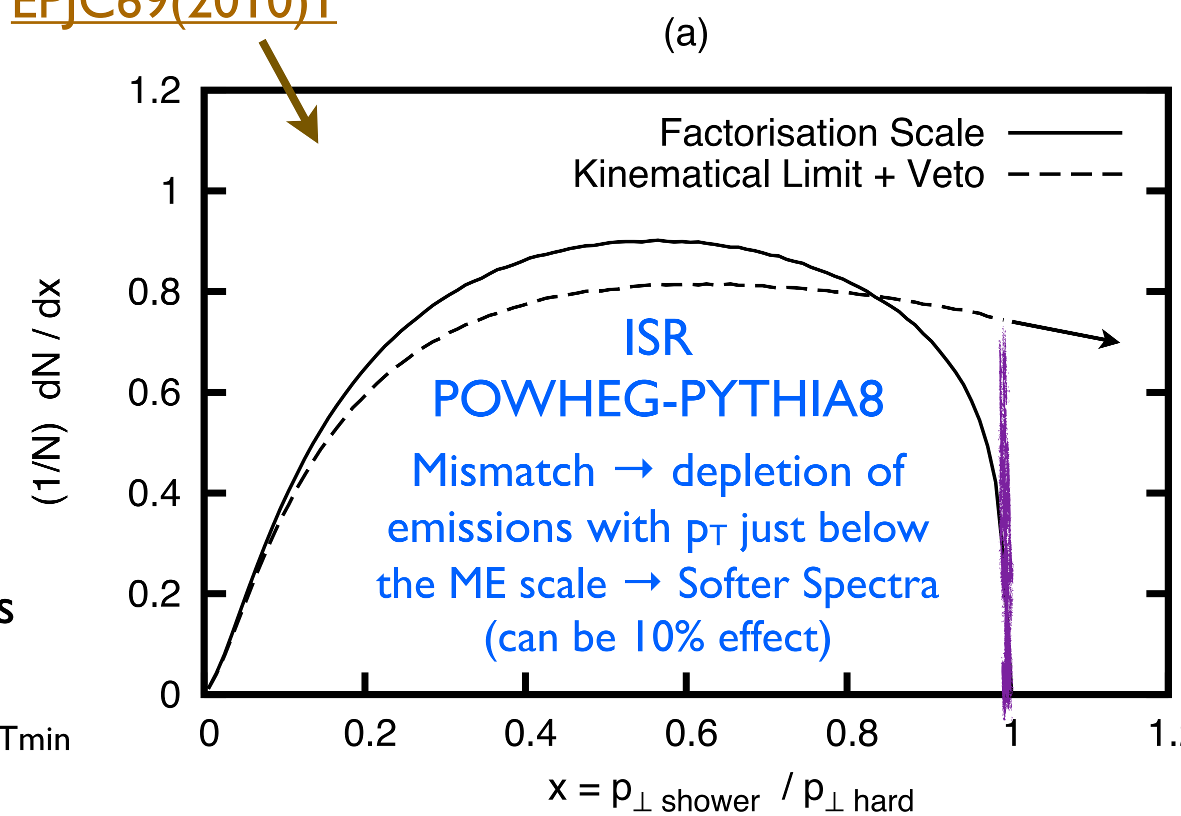
Beware: multiple defiitions of p_T, Corke & Sjöstrand, EPJC69(2010)1

Example: PYTHIA definition of p_T

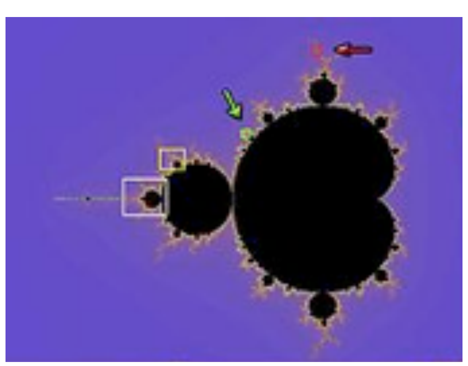
$$p_\perp^2 = p_{\perp\mathrm{evol}}^2 - \frac{p_{\perp\mathrm{evol}}^4}{p_{\perp\mathrm{evol,max}}^2}$$



Be safe: start at s and veto shower emissions above p_{Tmin}



2) Choice of matching scale



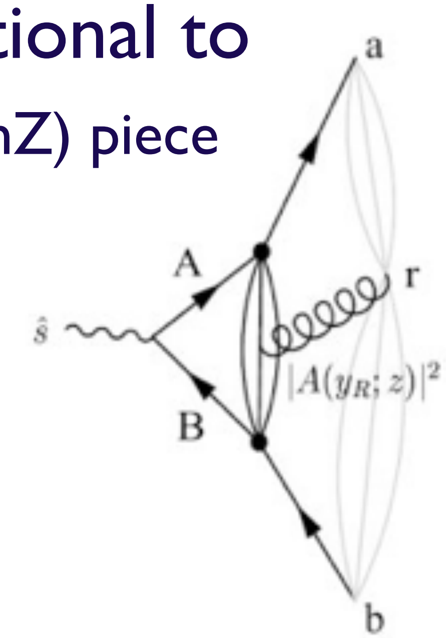
In perturbative region, QCD is approximately ***scale invariant***
→ A scale of 20 GeV for a W boson becomes 40 GeV for something weighing $2M_W$, etc … (+ adjust for C_A/C_F if g-initiated)
→ The matching scale should be written as a **ratio** (Bjorken scaling)
Using a too low matching scale → everything just becomes highest ME
Caveat emptor: showers generally do not include helicity correlations

Renormalization Scale

in Parton Showers

One-loop radiation functions contain pieces proportional to the β function (E.g.,: e+e-→3 jets, for arbitrary choice of μ_R (e.g., $\mu_R = m_Z$) piece from integrating quark loops over all of phase space

$$n_f \ A_3^0 \left(\ln\left(\frac{s_{23}}{\mu_R^2}\right) + \ln\left(\frac{s_{13}}{\mu_R^2}\right) \right)$$ + gluon loops



Proportional to the β function (b_0).

Can be absorbed by using $\mu_R^4 = s_{13}\, s_{23} = p_T^2\, s$. (~ "BLM")

In an ordered shower, quark (and gluon) loop integrals are restricted by strong-ordering condition → modified to

$\mu_R = p_T$ (but depends on ordering variable? Anyway, we're using pT here)

Additional logs induced by gluon loops can be absorbed by replacing Λ^{MS} by $\Lambda^{MC} \sim 1.5\ \Lambda^{MS}$ (with mild dependence on number of flavors)

Catani, Marchesini, Webber, NPB349 (1991) 635

Remaining ambiguity → tuning

Note: CMW not automatic in PYTHIA, has to be done by hand, by choosing effective Λ or $\alpha_s(M_Z)$ values instead of $\overline{MS}$ ones
Note 2: There are obviously still order 2 uncertainties on μ_R, but this is the background for the central choice made in showers

μ_R in a matched setting (MLM)

B. Cooper et al., arXiv:1109.5295

If using one code for MEs and another for showering

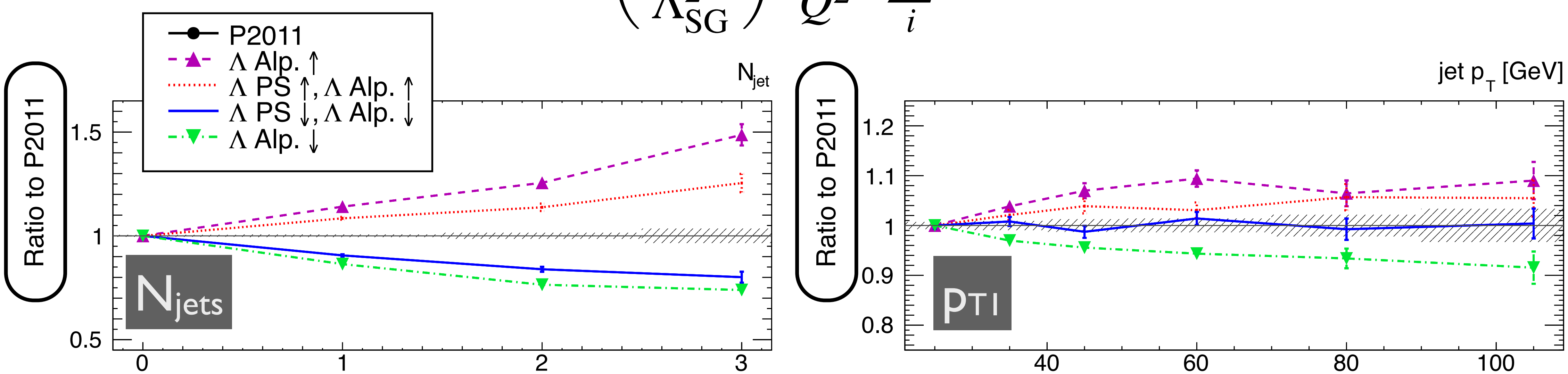
Tree-level corrections use α_s from Matrix-element Generator

Virtual corrections use α_s from Shower Generator (Sudakov)

Mismatch if the two do not use same Λ_{QCD} or $\alpha_s(m_Z)$

$$\alpha_s^2\, b_0 \ln\left(\frac{\Lambda_{\rm MG}^2}{\Lambda_{\rm SG}^2}\right) \frac{dQ^2}{Q^2} \sum_i P_i(z)\, |M_F|^2 \,.$$

note: running **order** also has a (subleading) effect



AlpGen: can set xlclu = Λ_{QCD} since v.2.14 (default remains to inherit from PDF)
Pythia 6: set common PARP(61)=PARP(72)=PARP(81) = Λ_{QCD} in Perugia 2011 tunes
Pythia 8: use TimeShower:alphaSvalue and SpaceShower:alphaSvalue

Parton Showers

Formally LL but include several important NLL aspects. A "good" shower should get close to NLL.

Other Ambiguities and issues (all beyond LL)

Momentum Recoil Strategies (global vs local $1\to 2$ vs $2\to 3$)

Coherence (e.g., angular-ordered parton showers vs p_T-ordered dipole ones, in particular initial-final connections: FSR broadening of ISR jets, …)

Jet Substructure (e.g., DGLAP vs Dipole/Antenna radiation functions, polarization effects on brems correlations, and effective $1\to 3$ description in topologies with compressed hierarchies: high-mass substructure)

Boosted Higgs

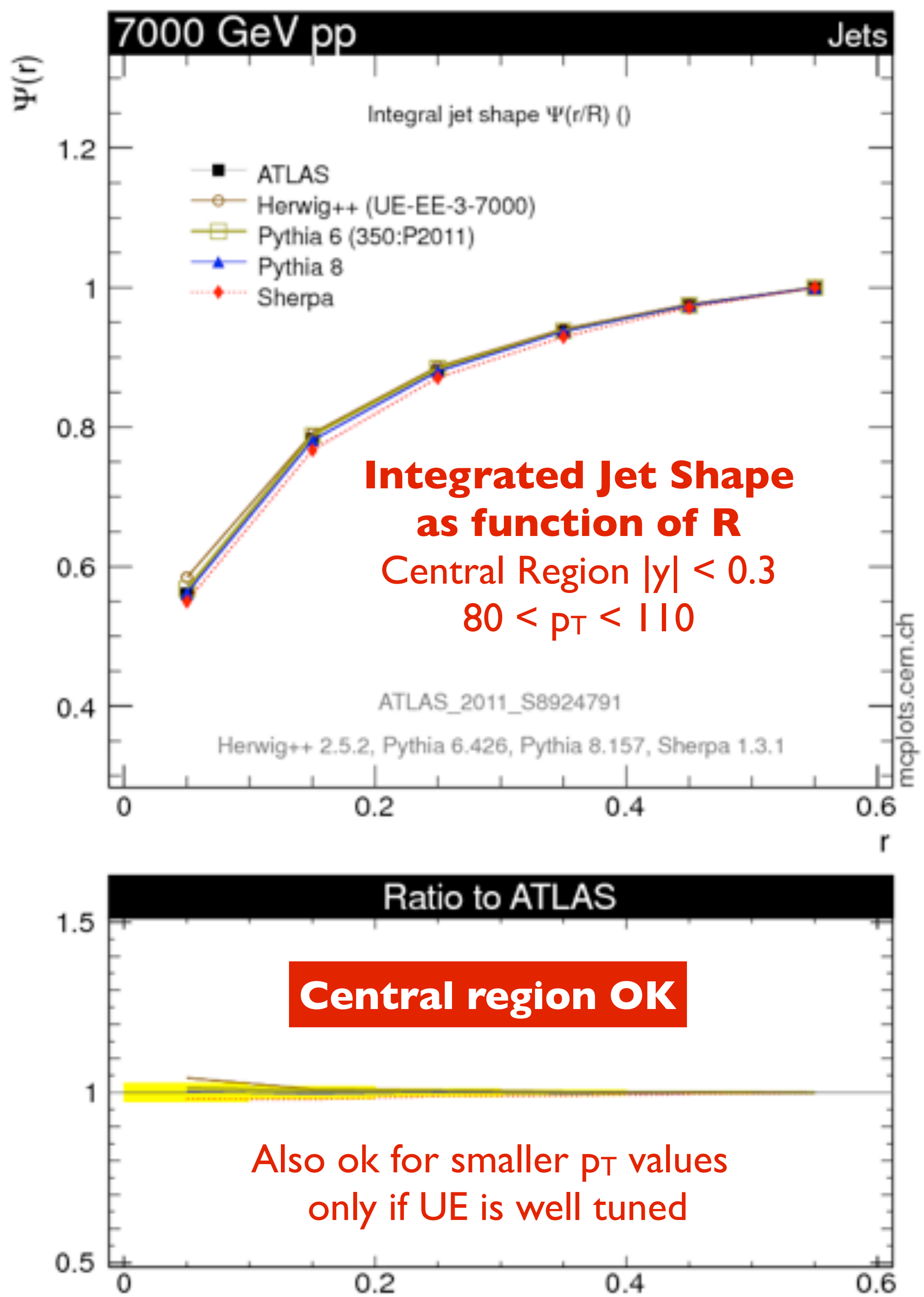
Mass Effects (b-jet calibration vs light-jet.) Description of $H\to bb$ + backgrounds

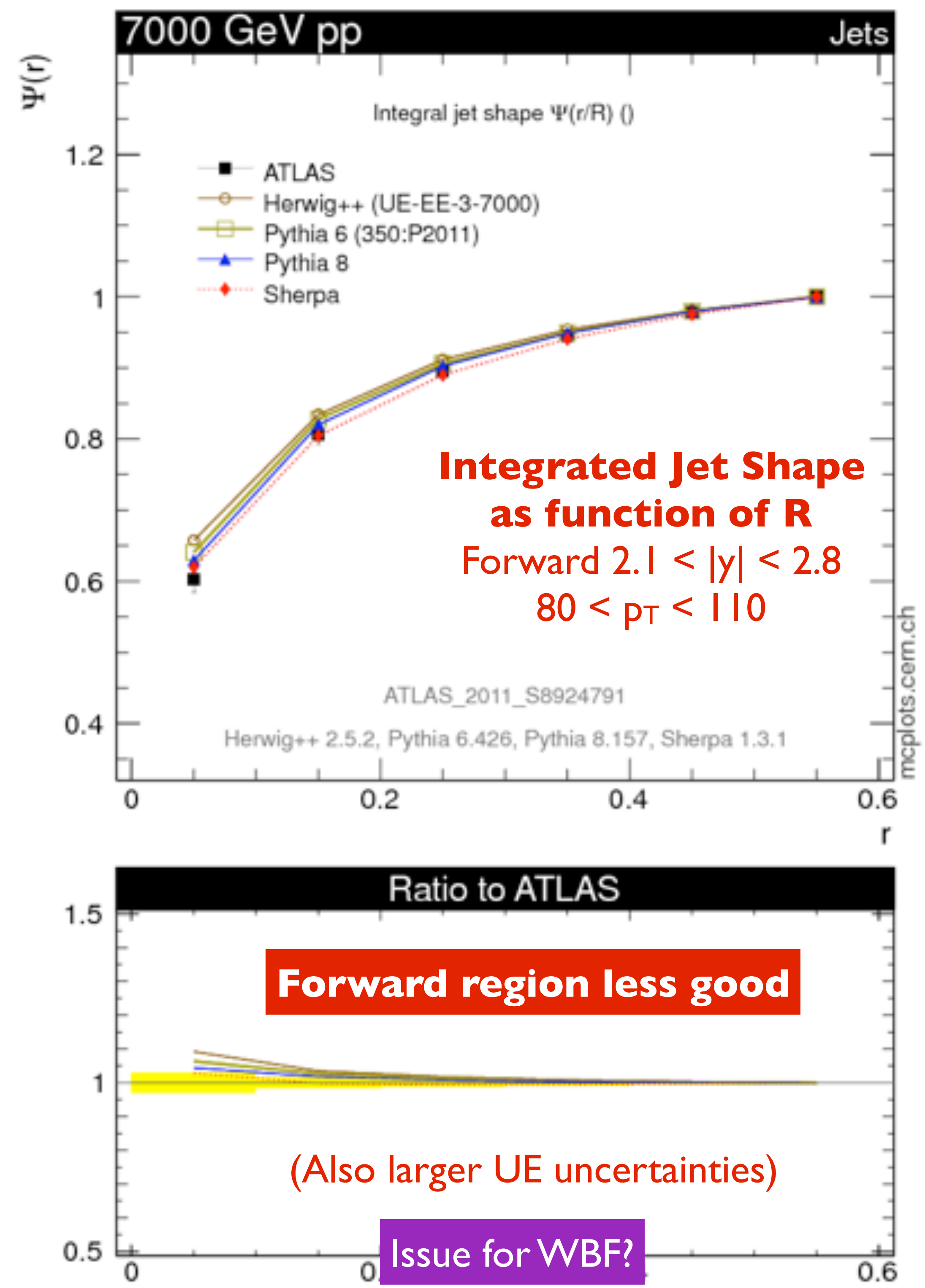
Gluon Splittings $g\to q\bar{q}$ (less well controlled even for massless quarks + not even singular for massive b quarks!) Backgrounds to $H\to bb$

Important cross-checks from comparisons to data (tuning) but also from theory (e.g., talk by Zanderighi)

Important test: LHC Jet Shapes

Dominated by FSR. For ISR, see talks by Mangano (slide 7-9), deFlorian, Zanderighi, and Higgs Working Group writeup





P. Skands

Plots from mcplots.cern.ch

b-jets (e.g., for Higgs Strahlung & backgrounds, BSM)

b mass acts as regulator → shower approximation intrinsically less good

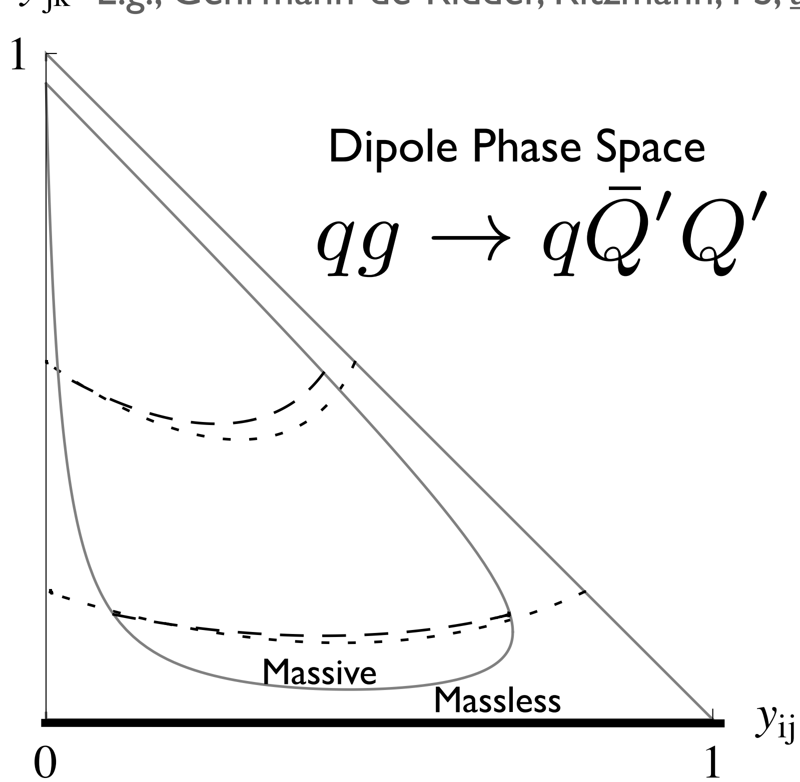
b-jets are more challenging to model

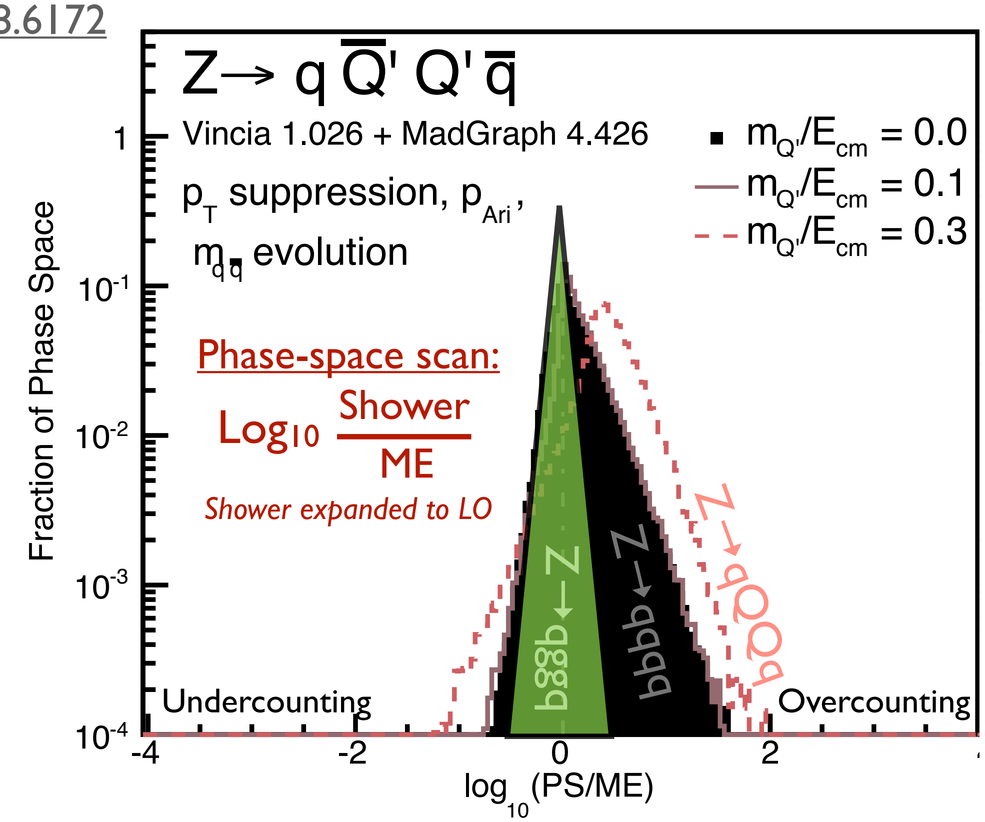
Little LEP/SLD & Tevatron in RIVET; Need more & LHC (+top?)

E.g., ATLAS arXiv:1112.6426

$g \to b\bar{b}$ is the most uncertain component of MC shower models

E.g., Gehrmann-de-Ridder, Ritzmann, PS; arXiv:1108.6172





→ Study jets with 1,2 b-quarks in them, under various combinations of jet p_T and m_{bb}.

See also talk by Mangano, slides 15-19

Underlying Event

Lots of ambiguities and issues

Interesting to get constraints on non-trivial QCD

But bottom line for High-p_T searches is UE now under good control (if using up-to-date MPI-based models/tunes)

(Though note: UE level sensitive to PDF choice!)

Underlying Event

Lots of ambiguities and issues

Interesting to get constraints on non-trivial QCD

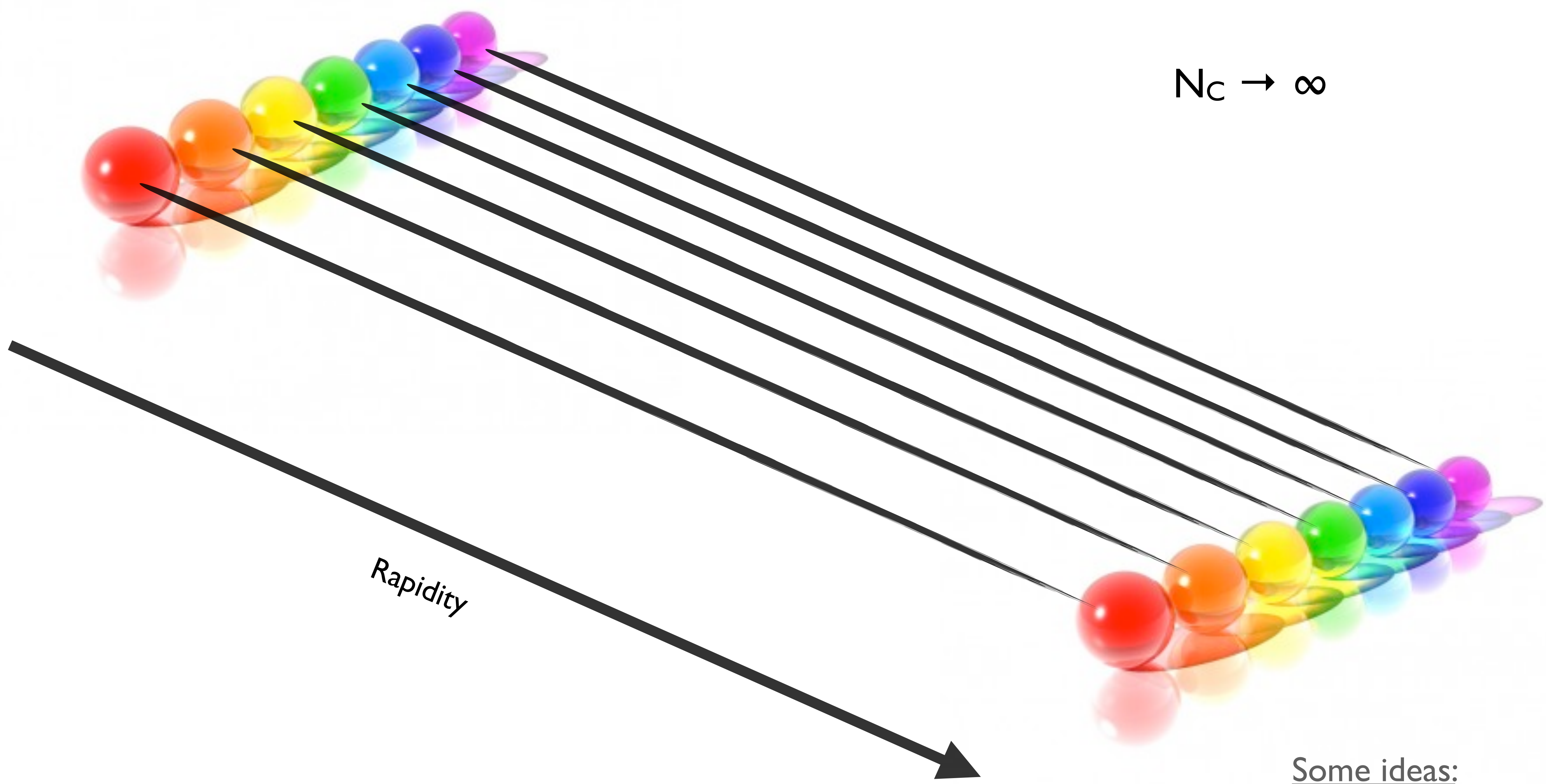
But bottom line for High-p_T searches is UE now under good control (if using up-to-date MPI-based models/tunes)

(Though note: UE level sensitive to PDF choice!)

Except

- In the **forward** region — impact on WBF?
- uncertainties on **multiple HARD** interactions (like double Drell-Yan, but small cross sections) — ?
- For **identified** particles / interplay with hadronization — $z \to 1$?
- Effects of **Color Reconnections**? — Impact on singlet taggers?

Color Connections



$N_C \to \infty$

Multiplicity $\propto N_{MPI}$

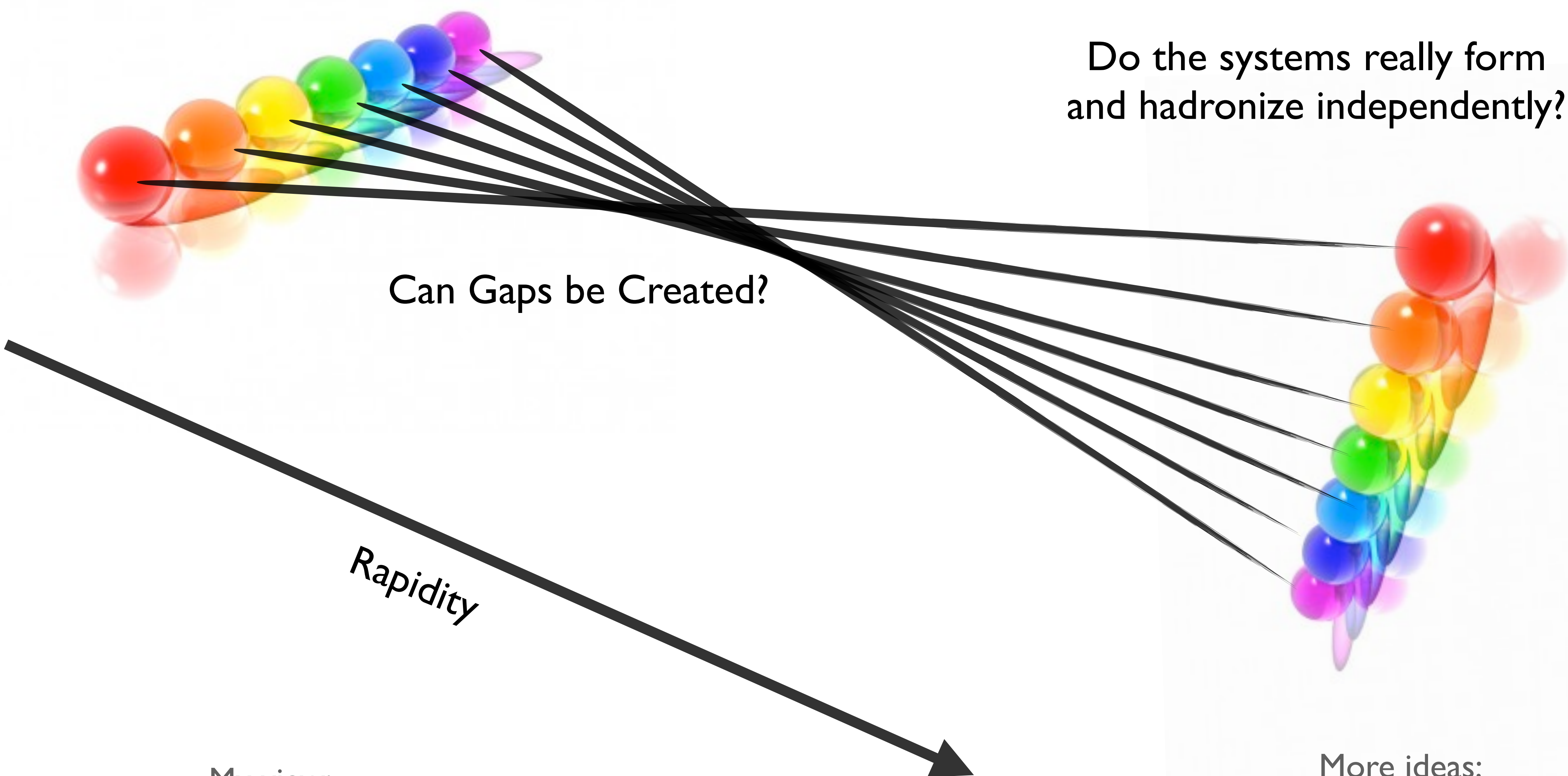
Some ideas:
Hydro? (EPOS)
E-dependent string parameters? (DPMJET)
"Color Ropes"?

Better theory models needed

Color Reconnections?

E.g.,
…
Generalized Area Law (Rathsman: Phys. Lett. B452 (1999) 364)
Color Annealing (PS, Wicke: Eur. Phys. J. C52 (2007) 133)
…

Do the systems really form
and hadronize independently?



<u>My view:</u>
Universality is ok *(a string is a string)*
Problem is $3 \neq \infty$

Multiplicity $\overset{<}{\not\propto}$ N_{MPI}

<u>More ideas:</u>
Coherent string formation?
Color reconnections?
String dynamics?

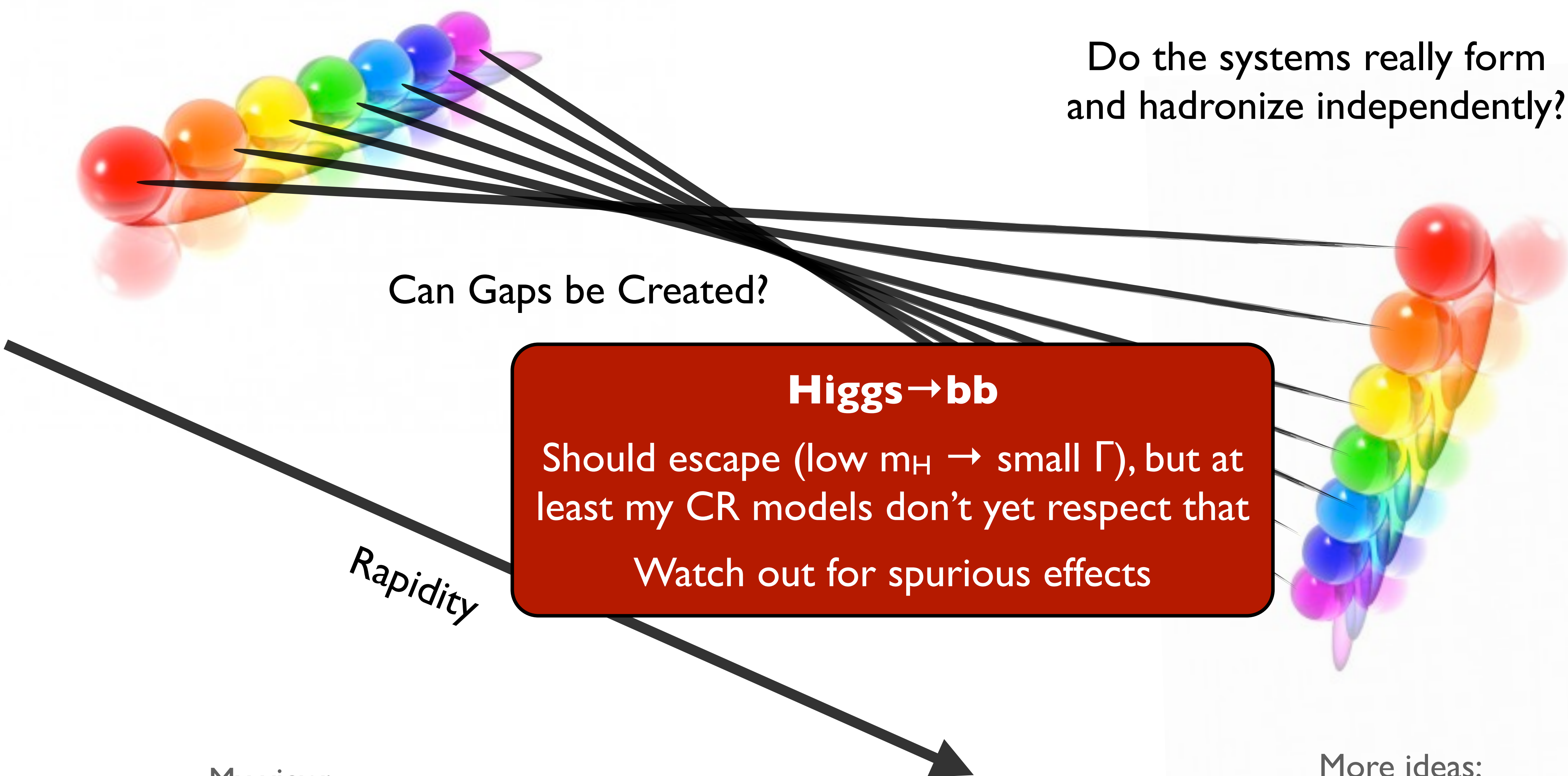
Better theory models needed

Color Reconnections?

E.g.,
…
Generalized Area Law (Rathsman: Phys. Lett. B452 (1999) 364)
Color Annealing (PS, Wicke: Eur. Phys. J. C52 (2007) 133)
…

Do the systems really form and hadronize independently?

Can Gaps be Created?



Higgs→bb

Should escape (low m_H → small Γ), but at least my CR models don't yet respect that

Watch out for spurious effects

My view:
Universality is ok *(a string is a string)*
Problem is $3 \neq \infty$

Multiplicity $\overset{<}{\not\propto}$ N_{MPI}

More ideas:
Coherent string formation?
Color reconnections?
String dynamics?

Better theory models needed

Pile-Up

= additional zero-bias interactions

Processes with *no hard scale*:

Larger uncertainties → Good Underlying Event does not imply good pile-up.

Error of 50% on a soft component → not bad. Multiply it by 60 Pile-Up interactions → bad!

Calibration & filtering good at recovering jet calibration (with loss of resolution), but missing energy and isolation sensitive to modeling.

H→WW

H→γγ? (E.g., γγ studies by ATLAS, CMS, CDF, D0)

Pile-Up

= additional zero-bias interactions

Processes with *no hard scale*:

Larger uncertainties → Good Underlying Event does not imply good pile-up.

Error of 50% on a soft component → not bad. Multiply it by 60 Pile-Up interactions → bad!

Calibration & filtering good at recovering jet calibration (with loss of resolution), but missing energy and isolation sensitive to modeling.

H→WW H→γγ? (E.g., γγ studies by ATLAS, CMS, CDF, D0)

Models

Minimum-Bias & Underlying Event

MC models so far: problems describing both MB & UE simultaneously

→ Consider using dedicated MB/diffraction model for pile-up

(UE/MB tension may be resolved in 2012 (eg. studies by R. Field), but for now must live with it)

Experimentalists advised to use unbiased data (when possible)

Pile-Up

= additional zero-bias interactions

Processes with *no hard scale*:

Larger uncertainties → Good Underlying Event does not imply good pile-up.

Error of 50% on a soft component → not bad. Multiply it by 60 Pile-Up interactions → bad!

Calibration & filtering good at recovering jet calibration (with loss of resolution), but missing energy and isolation sensitive to modeling.

H→WW H→γγ? (E.g., γγ studies by ATLAS, CMS, CDF, D0)

Models

Minimum-Bias & Underlying Event

MC models so far: problems describing both MB & UE simultaneously

→ Consider using dedicated MB/diffraction model for pile-up

(UE/MB tension may be resolved in 2012 (eg. studies by R. Field), but for now must live with it)

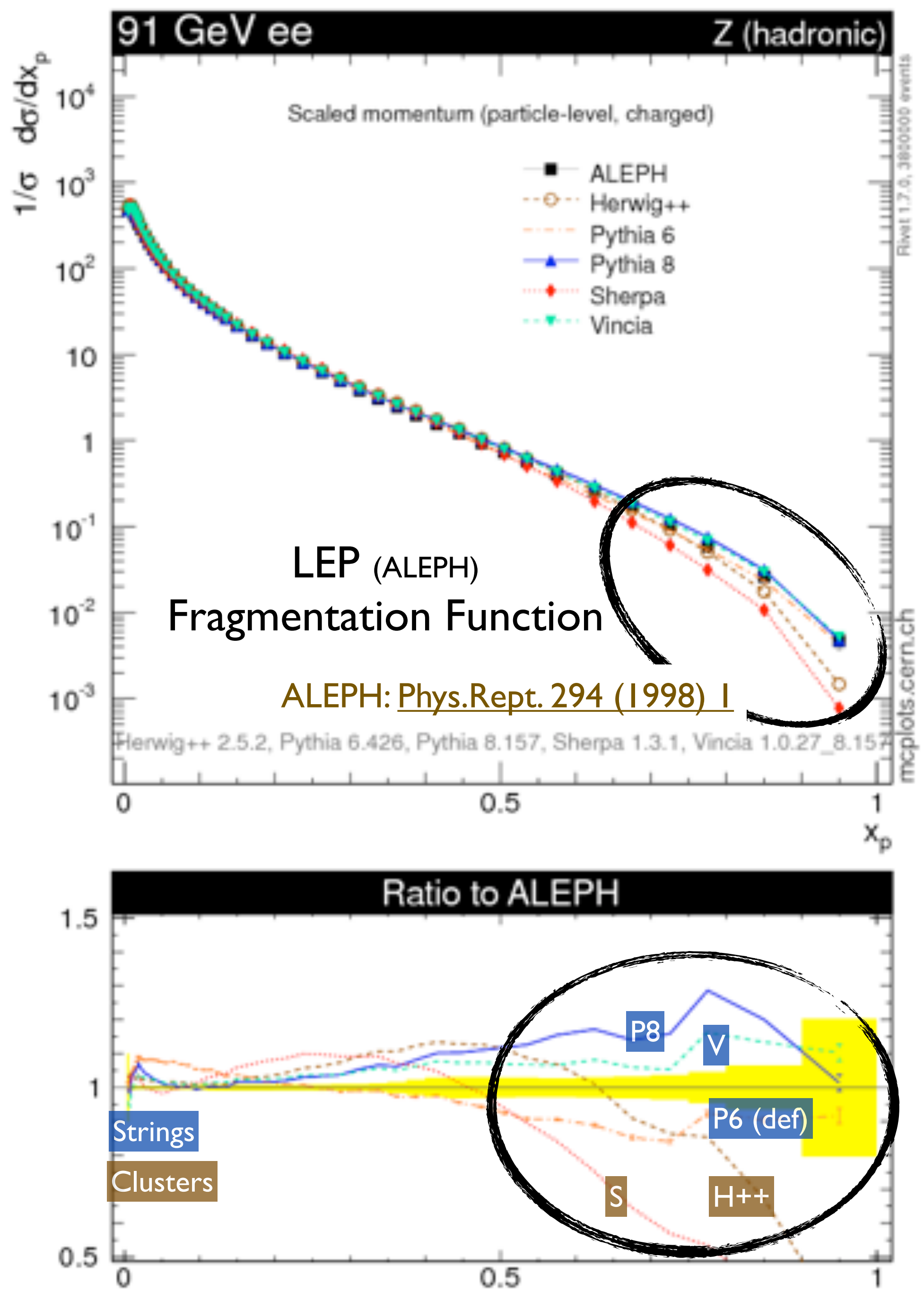
Experimentalists advised to use unbiased data (when possible)

Diffraction Warning: in forward region, Pile-up has larger diffractive component than Min-Bias (zero bias vs min-bias). Harder to reject due to lack of tracking in FWD region. Poorly described (or not at all) in current MC models → can affect ETMiss etc. An improved model has been included in PYTHIA 8, but *still needs testing and tuning*. Improved models also on their way in Herwig++ and in Sherpa. Best current description of diffraction may be PHOJET, though also not perfect.

Extreme Fragmentation

See also talk by Mangano, slides 10-12

How often does an entire jet fragment into **a single/isolated particle?** (can produce dangerous fakes)
Controlled by the behavior of the fragmentation function at z→1. Deep Sudakov region, very tough to model.
Intrinsically suppressed in cluster models. But even good string tunes probably not very reliable.

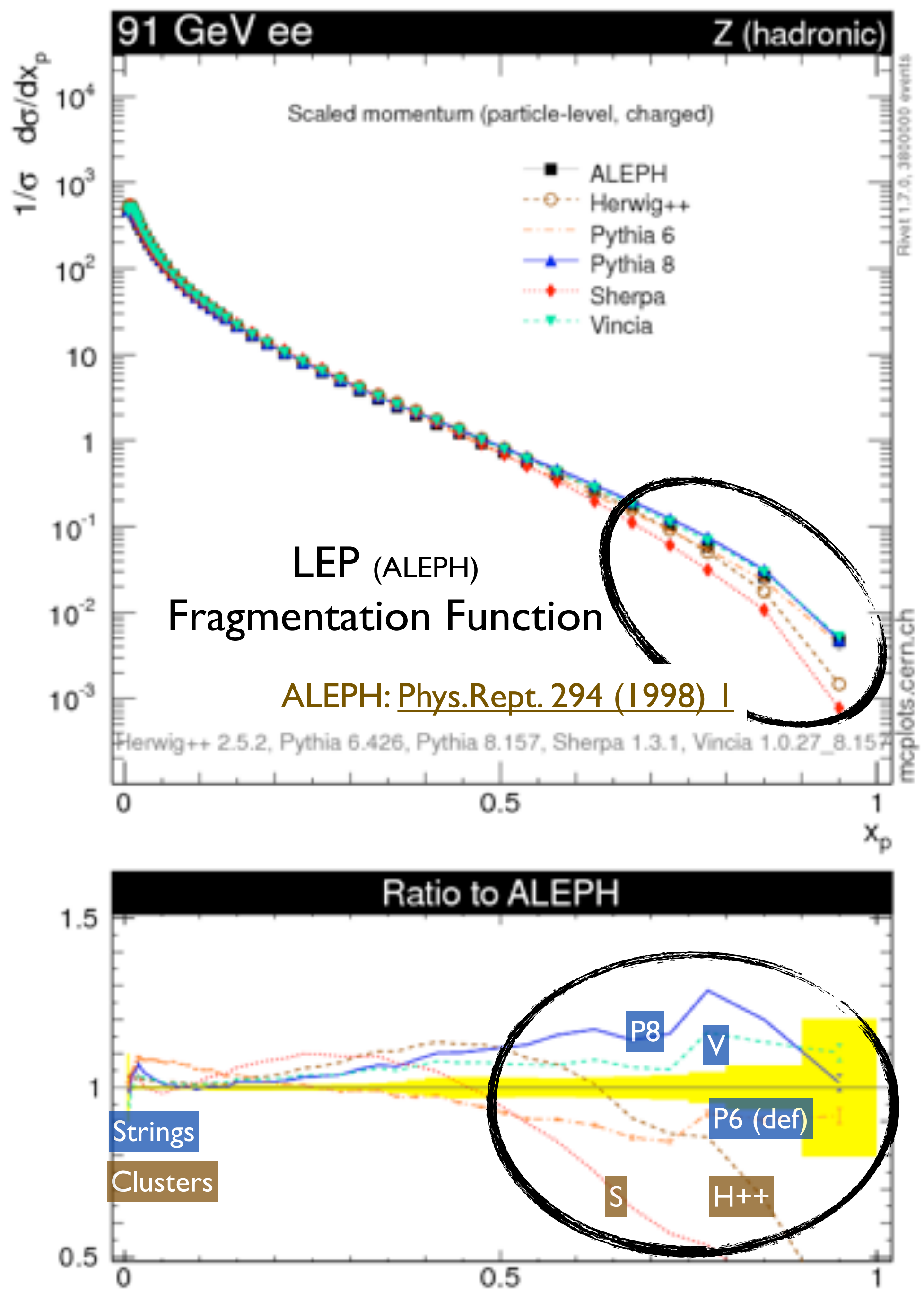


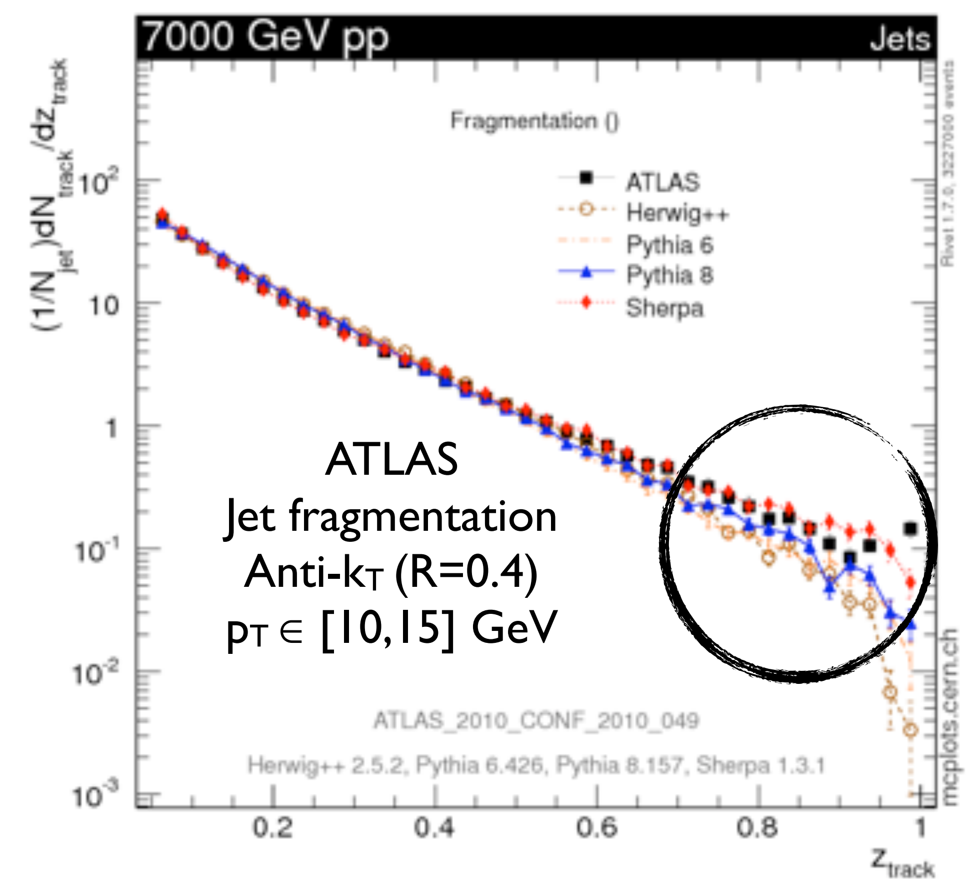
Plots from mcplots.cern.ch

Extreme Fragmentation

See also talk by Mangano, slides 10-12

How often does an entire jet fragment into **a single/isolated particle?** (can produce dangerous fakes)
Controlled by the behavior of the fragmentation function at z→1. Deep Sudakov region, very tough to model.
Intrinsically suppressed in cluster models. But even good string tunes probably not very reliable.





Pattern changes in pp jets

(though here only *inside* jets, and jets only at 10-15 GeV)

Needs to be studied in more detail if MC models to be used in z→1 region

Plots from mcplots.cern.ch

Towards automatic uncertainty estimates

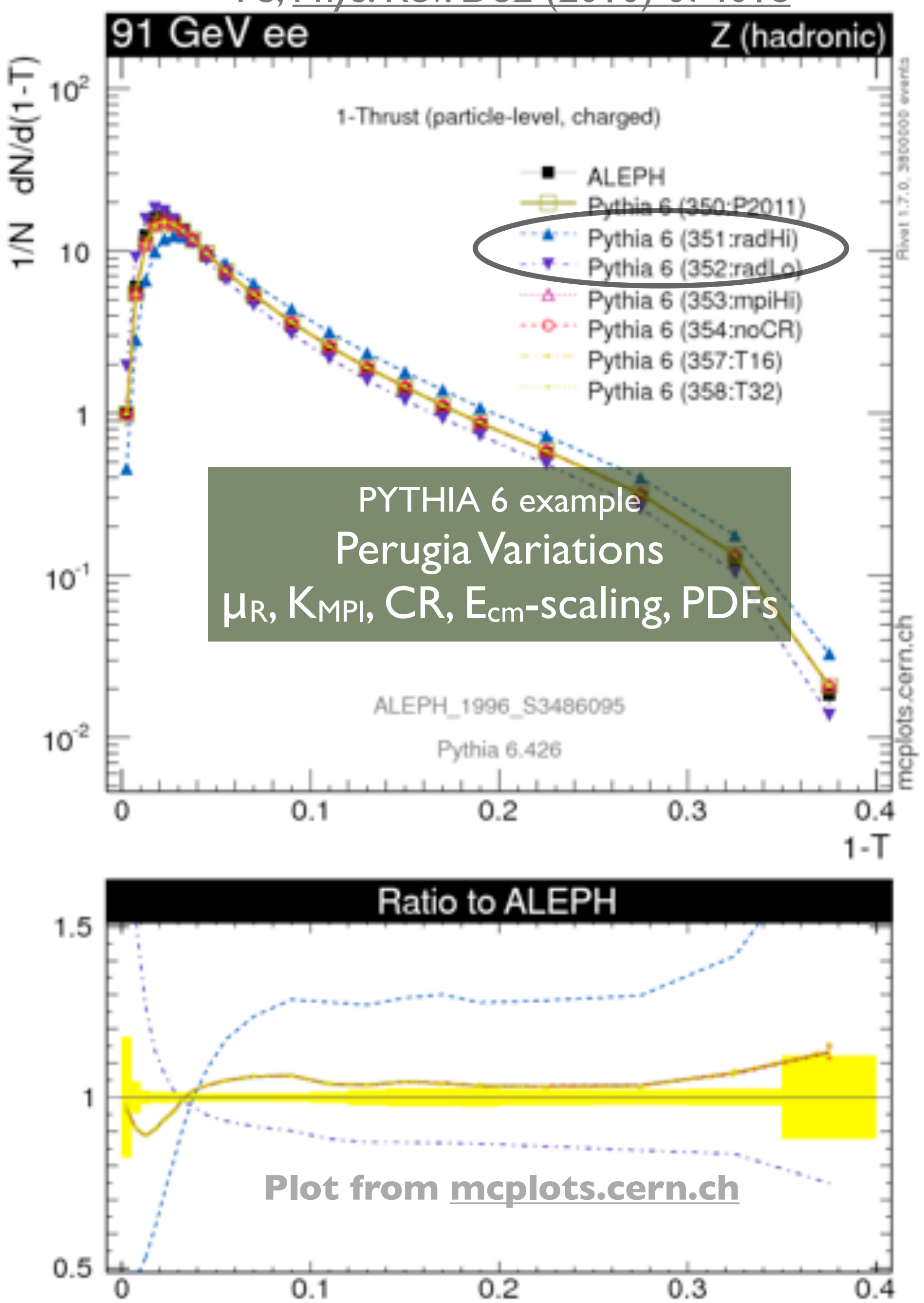
"inspired" by PDF uncertainties, see e.g., talk by J. Stirling

a) Authors provide specific "tune variations"

Run once for each variation

(= separate samples) → envelope

PS, Phys. Rev. D82 (2010) 074018

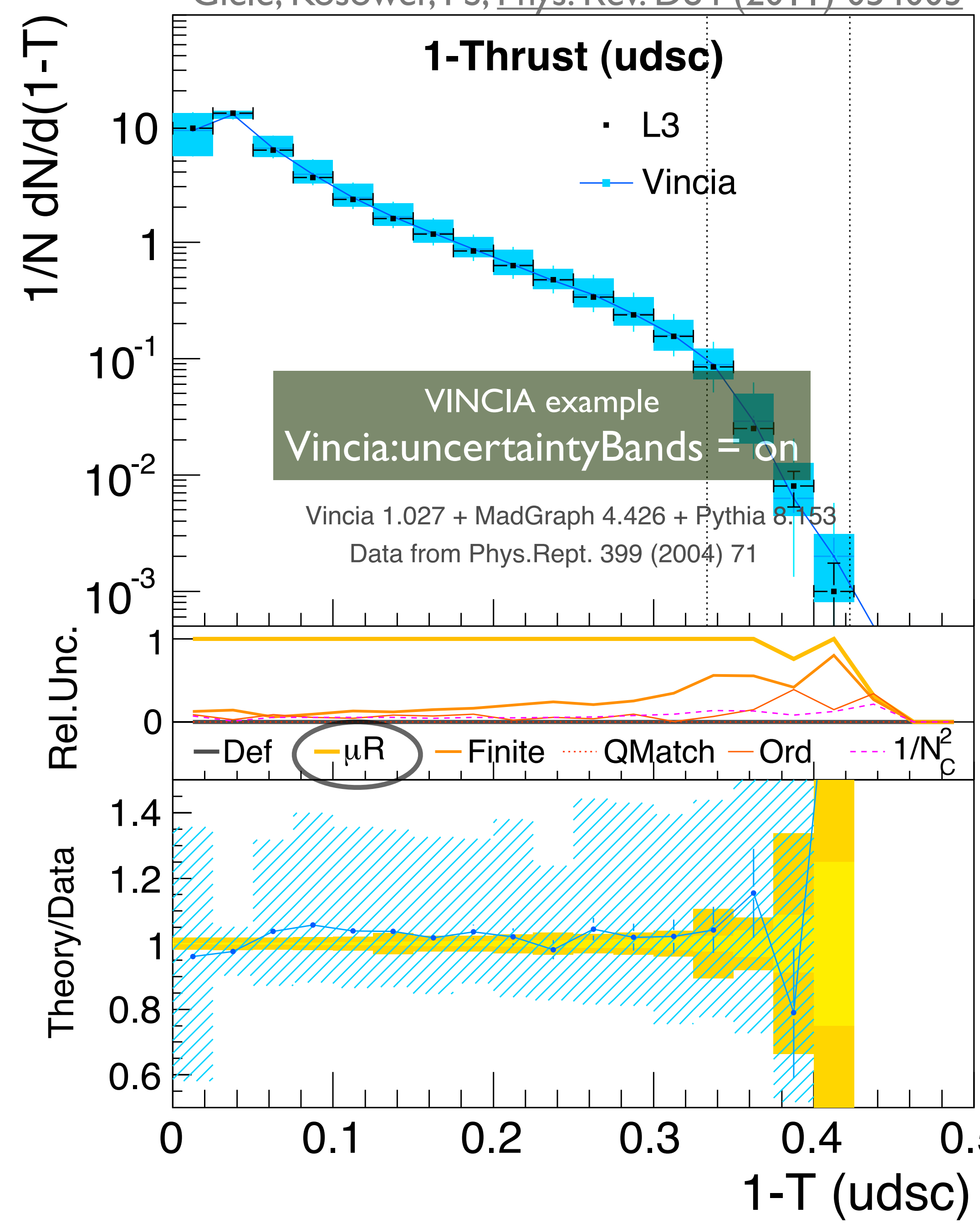


b) **One** shower run (unweighted)

+ unitarity-based uncertainties

(= sets of weights) → envelope

Giele, Kosower, PS; Phys. Rev. D84 (2011) 054003



Note: not done yet for hadronization parameters

Towards automatic uncertainty estimates

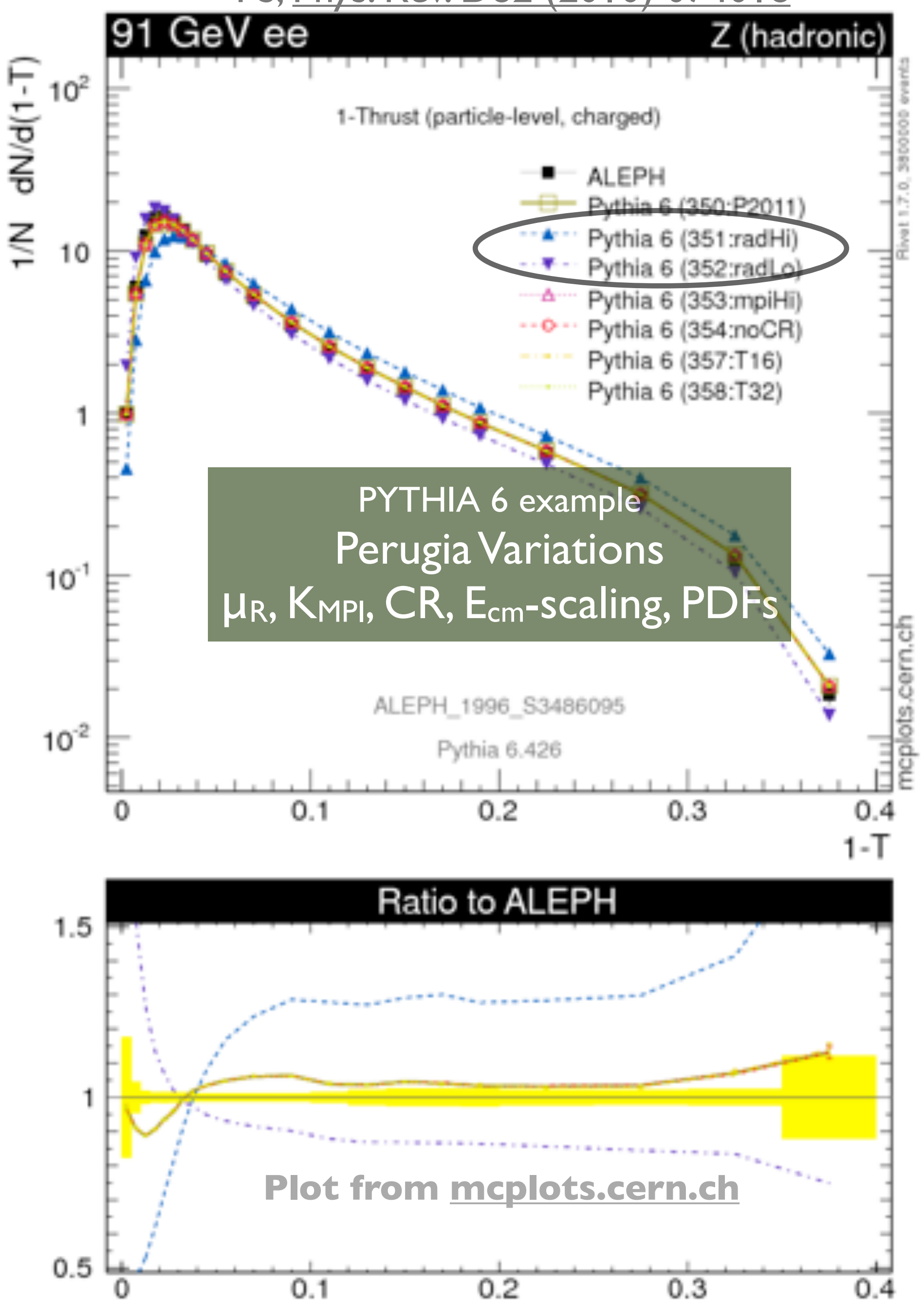
"inspired" by PDF uncertainties, see e.g., talk by J. Stirling

a) Authors provide specific "tune variations"

Run once for each variation

(= separate samples) → envelope

PS, Phys. Rev. D82 (2010) 074018

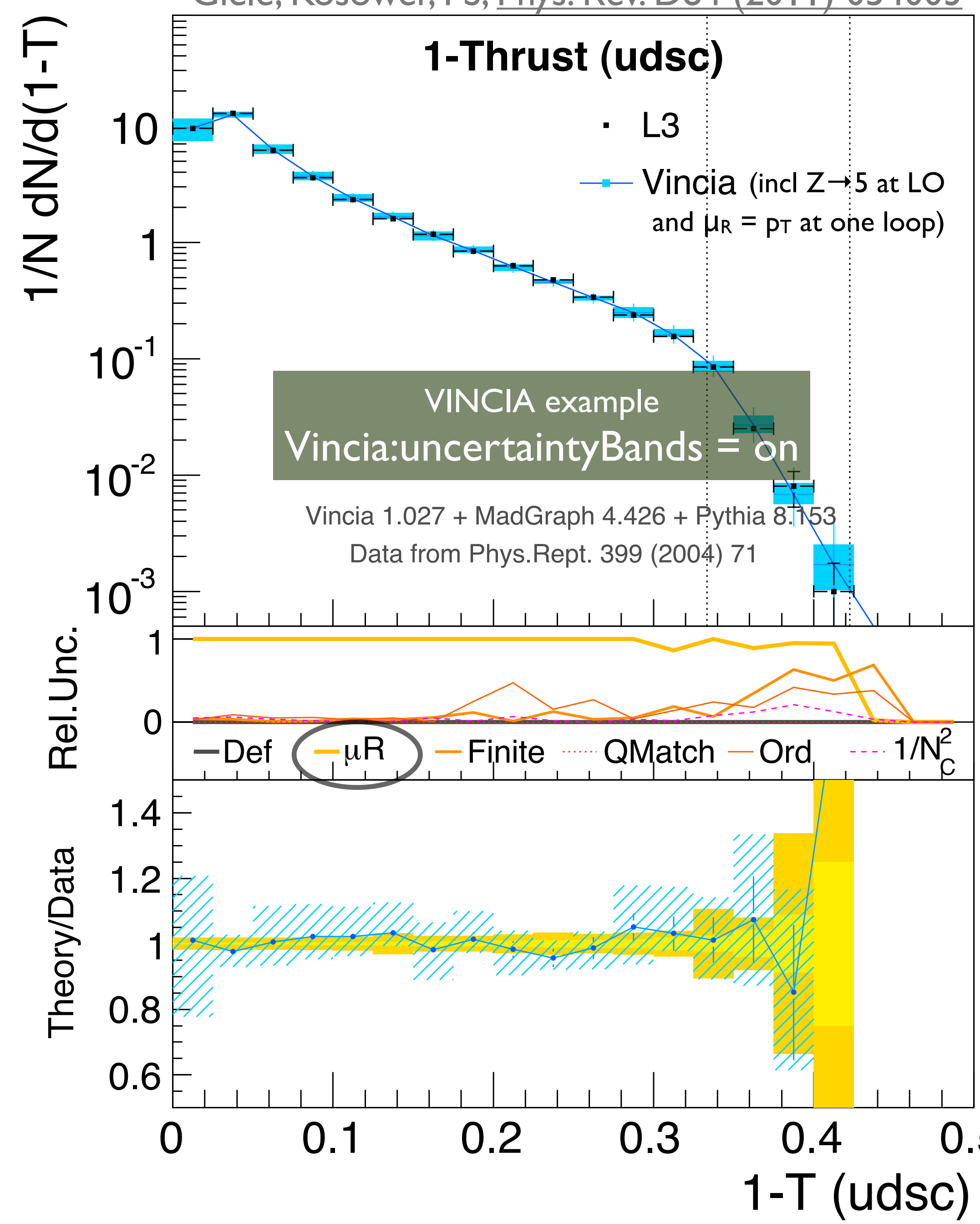


b) **One** shower run (unweighted)

+ unitarity-based uncertainties

(= sets of weights) → envelope

Giele, Kosower, PS; Phys. Rev. D84 (2011) 054003



Note: not done yet for hadronization parameters

Towards automatic uncertainty estimates

"inspired" by PDF uncertainties, see e.g., talk by J. Stirling

a) Authors provide specific "tune variations"
Run once for each variation
(= separate samples) → envelope

b) **One** shower run (unweighted)
+ unitarity-based uncertainties
(= sets of weights) → envelope

Giele, Kosower, PS; Phys. Rev. D84 (2011) 054003

```
*-------  PYTHIA Event and Cross Section Statistics  -------------------------------------------------------------*
|                                                                                                                  |
| Subprocess                                    Code |            Number of events       |      sigma +- delta    |
|                                                    |       Tried   Selected   Accepted |     (estimated) (mb)   |
|                                                    |                                   |                        |
|------------------------------------------------------------------------------------------------------------------|
|                                                    |                                   |                        |
| f fbar -> gamma*/Z0                            221 |       10511      10000      10000 |   4.143e-05  0.000e+00 |
|                                                    |                                   |                        |
| sum                                                |       10511      10000      10000 |   4.143e-05  0.000e+00 |
|                                                                                                                  |
*-------  End PYTHIA Event and Cross Section Statistics ------------------------------------------------------------*

*-------  VINCIA Statistics  --------------------------------------------------------------------------------------*
|                                                                                                                  |
|                                                                                                                  |
| Number of nonunity-weight events                                 =      none                                    |
| Number of negative-weight events                                 =      none                                    |
|                                                                                                                  |
|                                  weight(i)         Avg Wt   Avg Dev  rms(dev)      kUnwt     Expected effUnw     |
| This run                            i =     IsUnw    <w>    <w-1>                 1/<w>      Max Wt  <w>/MaxWt   |
|    User settings                       0     yes   1.000    0.000          -     1.000          -          -     |
|    Var : VINCIA defaults               1     yes   1.000    0.000          -     1.000      1.000      1.000     |
|    Var : AlphaStr-Max                  2      no   0.996 -3.89e-03         -     1.004     22.414   4.44e-02     |
|    Var : AlphaStr-Min                  3      no   1.020  1.99e-02         -     0.981     43.099   2.37e-02     |
|    Var : Antennae-Max                  4      no   1.000  2.61e-04         -     1.000      5.417      0.185     |
|    Var : Antennae-Min                  5      no   0.996 -4.33e-03         -     1.004     10.753   9.26e-02     |
|    Var : RESERVED                      6     yes   1.000    0.000          -     1.000      1.000      1.000     |
|    Var : RESERVED                      7     yes   1.000    0.000          -     1.000      1.000      1.000     |
|    Var : Ordering-Stronger             8      no   1.004  4.48e-03         -     0.996     14.225   7.06e-02     |
|    Var : Ordering-mDaughter            9      no   1.033  3.25e-02         -     0.968     55.954   1.85e-02     |
|    Var : ColorNLC-Max                 10      no   1.001  7.37e-04         -     0.999      1.505      0.665     |
|    Var : ColorNLC-Min                 11      no   1.006  6.44e-03         -     0.994      5.283      0.191     |
|                                                                                                                  |
*-------  End VINCIA Statistics -----------------------------------------------------------------------------------*
```

Summary

Underlying Event and Jet Shapes: ok (for high-p_T physics)

If in doubt check **mcplots.cern.ch** ISR: include p_{TZ} , $p_{Tt\bar{t}}$, p_{Tjj} (EXP) & p_{TH} , jet vetos (TH)

WARNING: UE tuning depends **explicitly** on the PDF it was tuned with !!!

Pile-Up: Mismodeling can impact missing energy (and isolation?) estimates

No hard scale → more challenging for pQCD-based models (only PYTHIA and PHOJET so far include diffraction. HERWIG++ and SHERPA models on their way)

Especially soft & diffractive aspects need more study/constraints/modeling

Other Modeling & Tuning Aspects

μ_R : Fixing μ_R to its $\overline{\text{MS}}$ value without accounting for known physics (e.g., CMW) and remaining ambiguities is too naive (in shower context)

Matching: remember Bjorken + ensure consistency between ME and PS sides, especially when combining different codes (e.g., ALPGEN/MADGRAPH + PYTHIA/HERWIG)

Color Reconnections: coherence not well understood *between* MPI chains. Affects hadronization in busy pp events. Can alter IR sensitive properties*, like color-flow-variables, particle momentum spectra, and isolation.

Hadronization: depends on color connections.
Extreme tails ($z\to 1$) already difficult at LEP, important to be checked in situ (not just in min-bias)

*Sometimes unintentionally


MC4BSM-2012
Monte Carlo Tools
for Physics
Beyond the Standard Model
March 22-24, 2012
Cornell University
Ithaca, New York USA
www.lepp.cornell.edu/events/MC4BSM/
Organizing Committee
Kyle Cranmer, NYU
Csaba Csaki, Cornell
Christophe Grojean, CERN
Yuval Grossman, Cornell
Jay Hubisz, Syracuse
Konstantin Matchev, Florida
Stephen Mrenna, Fermilab
Maxim Perelstein (Chair), Cornell
Anders Ryd, Cornell
Peter Skands, CERN
Workshop Secretary
Elizabeth Gustafson, Cornell
Organizer Email
mc4bsm@physics.syr.edu
Cornell University
Founded A.D. 1865
MC4BSM
2012

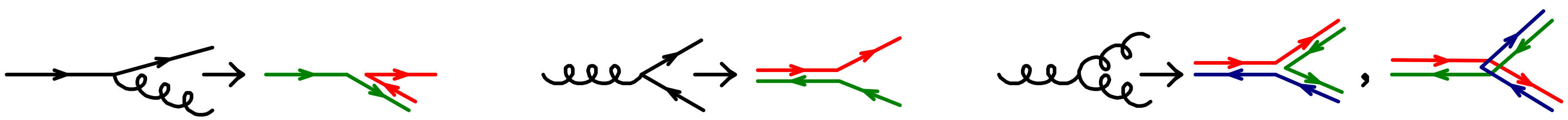
Backup Slides

(Color Flow in MC Models)

"Planar Limit"

Equivalent to $N_C \to \infty$: no color interference*

Rules for color flow:

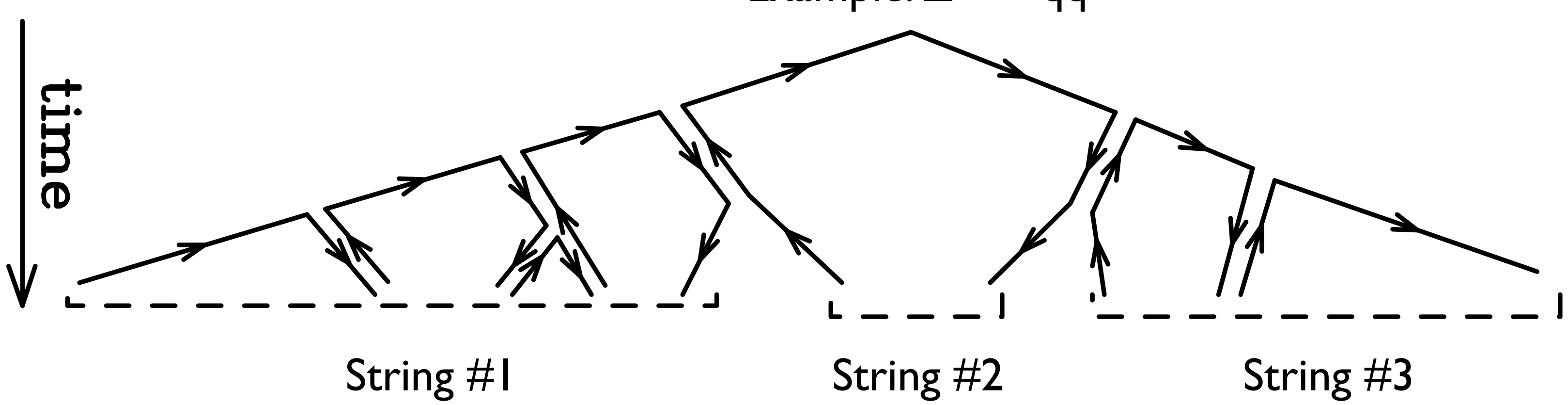


*) except as reflected by the implementation of QCD coherence effects in the Monte Carlos via angular or dipole ordering

For an entire cascade:

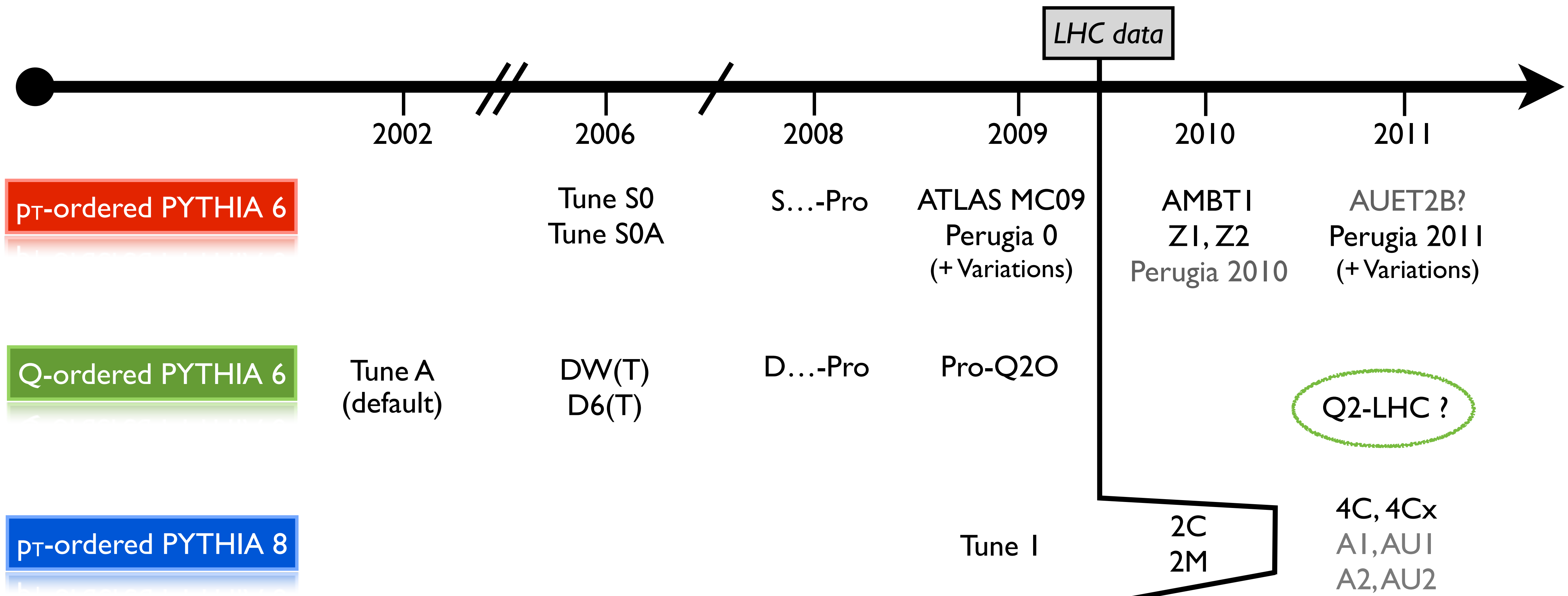
Illustrations from: Nason + PS, PDG Review on *MC Event Generators*, 2012

Example: $Z^0 \to q\bar{q}$



Coherence of pQCD cascades → not much "overlap" between strings
→ planar approx pretty good
LEP measurements in WW confirm this (at least to order 10% ~ $1/N_c^2$)

PYTHIA Models



Main Data Sets included in each Tune (no guarantee that all subsets ok)

	A	DW, D6, …	S0, S0A	MC09(c)	Pro-…, Perugia 0, Tune 1, 2C, 2M	AMBT1	Perugia 2010	Perugia 2011	Z1, Z2	4C, 4Cx	AUET2B, A2, AU2
LEP					✔		✔	✔		✔	✔
TeV MB			✔	✔	✔		✔	✔		(✔)	?
TeV UE	✔	✔		✔	✔		✔	✔		(✔)	✔?
TeV DY		✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
LHC MB						✔	✔	✔		✔	?
LHC UE								✔	✔		✔

Pythia 6: The Perugia Variations

"Tuning MC Generators: The Perugia Tunes" - PRD82 (2010) 074018

Central Tune + 9 variations

Note: no variation of hadronization parameters! (sorry, ten was already a lot)

Perugia 2011 Tune Set

(350)	Perugia 2011	Central Perugia 2011 tune (CTEQ5L)	
(351)	Perugia 2011 radHi	Variation using $\alpha_s(\frac{1}{2}p_\perp)$ for ISR and FSR	Harder radiation
(352)	Perugia 2011 radLo	Variation using $\alpha_s(2p_\perp)$ for ISR and FSR	Softer radiation
(353)	Perugia 2011 mpiHi	Variation using $\Lambda_{\mathrm{QCD}} = 0.26\,\mathrm{GeV}$ also for MPI	UE more "jetty"
(354)	Perugia 2011 noCR	Variation without color reconnections	Softer hadrons
(355)	Perugia 2011 M	Variation using MRST LO** PDFs	UE more "jetty"
(356)	Perugia 2011 C	Variation using CTEQ 6L1 PDFs	Recommended
(357)	Perugia 2011 T16	Variation using `PARP(90)=0.16` scaling away from 7 TeV	
(358)	Perugia 2011 T32	Variation using `PARP(90)=0.32` scaling away from 7 TeV	
(359)	Perugia 2011 Tevatron	Variation optimized for Tevatron	~ low at LHC

Can be obtained in standalone Pythia from 6.4.25+

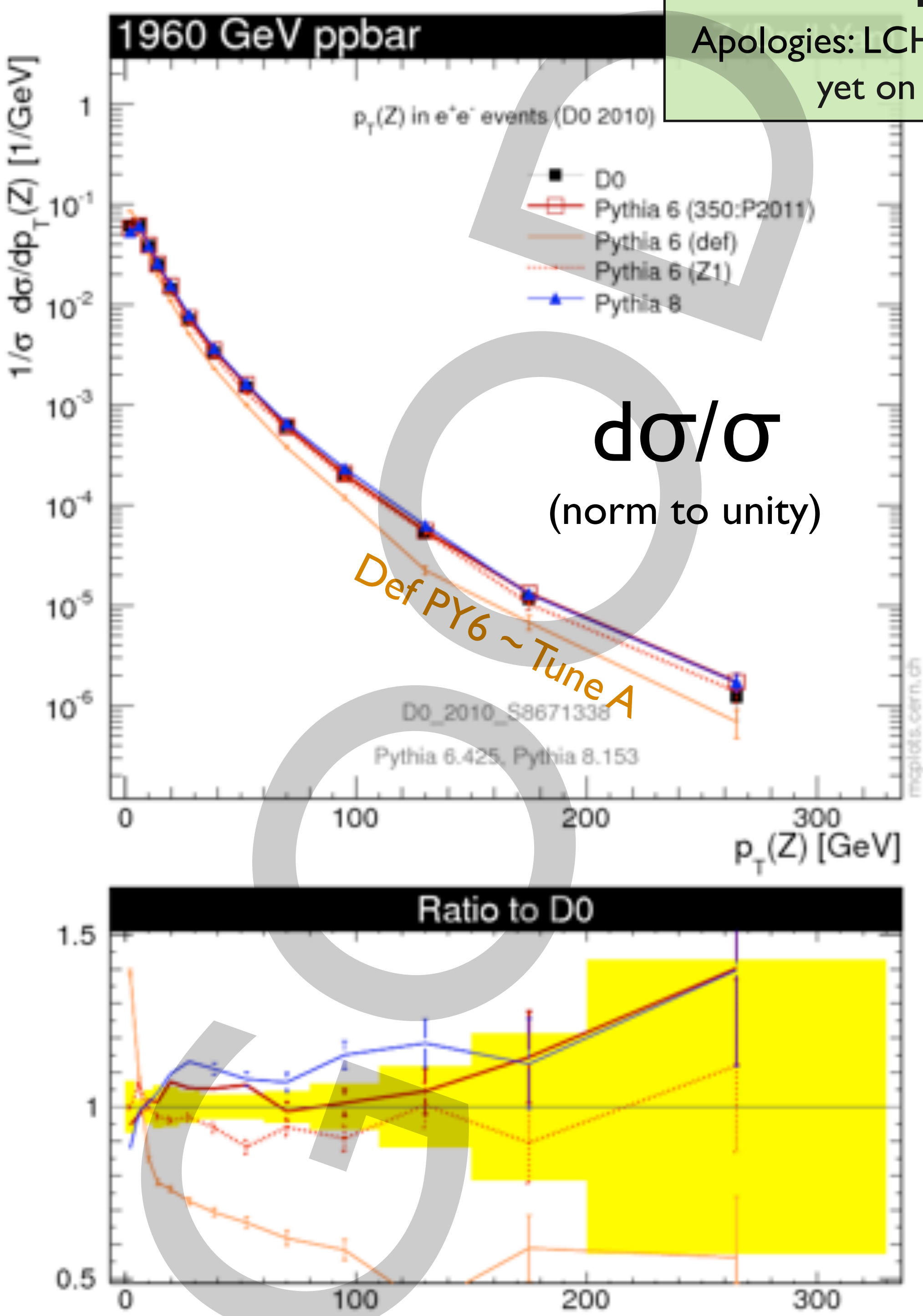
MSTP(5) = 350	MSTP(5) = 351	MSTP(5) = 352	MSTP(5) = …
Perugia 2011	Perugia 2011 radHi	Perugia 2011 radLo	...

Tunes of PYTHIA 8 : Corke & Sjöstrand - JHEP 03 (2011) 032 & JHEP 05 (2011) 009

(Important test: Drell-Yan p_T spectrum)

Tevatron
Apologies: LCH DY measurements not yet on mcplots.cern.ch

ATLAS: arXiv:1107.2381
CMS: arXiv:1110.4973



qq→Z

Oldest Tevatron tunes fail (e.g., default Pythia 6, Tune A)

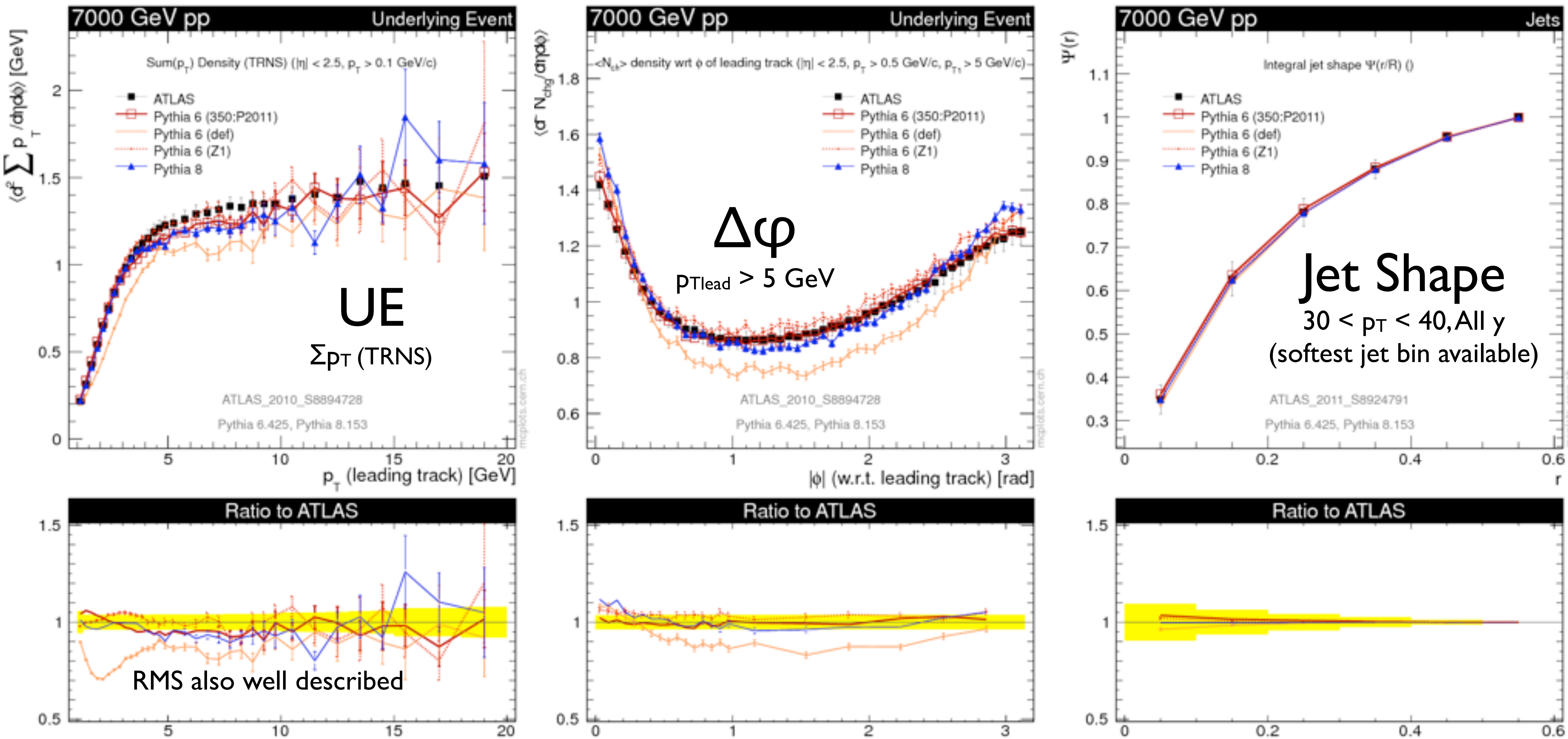
Basically all other models (including more recent Pythia ones) do fine.

gg→Higgs

Need additional cross-checks sensitive to gg-initiated processes:

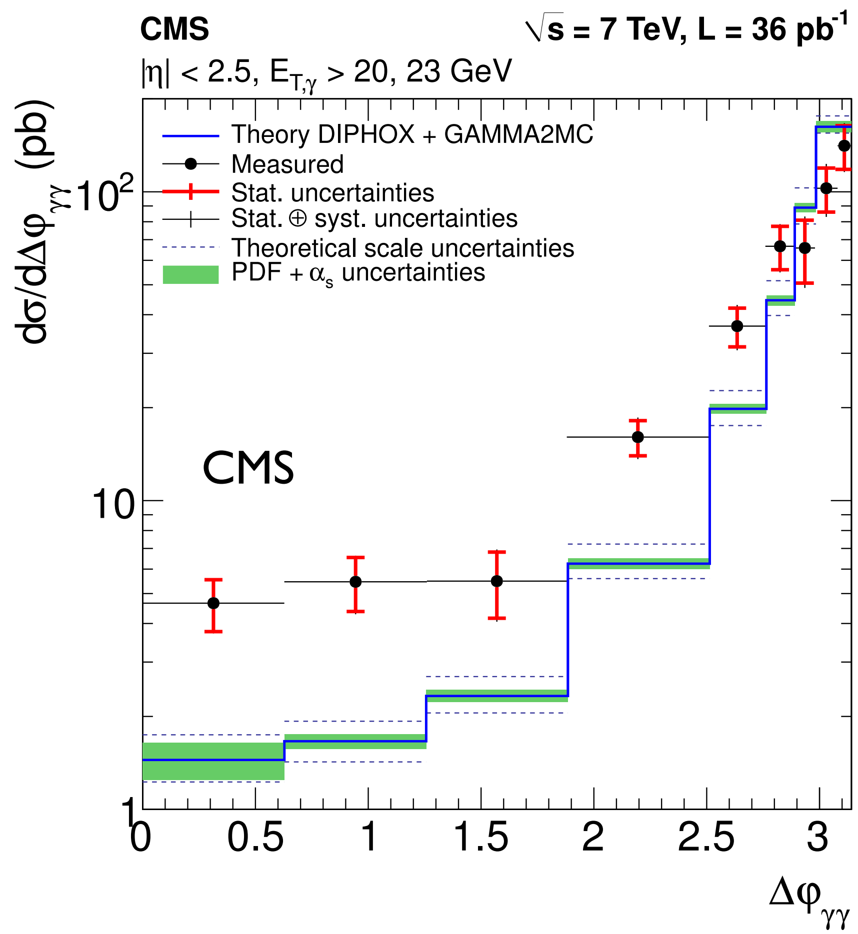
Dijets with $2p_T \sim m_H \sim$ acceptable + $p_T(tt)$ in top events
(though note: different color structures)

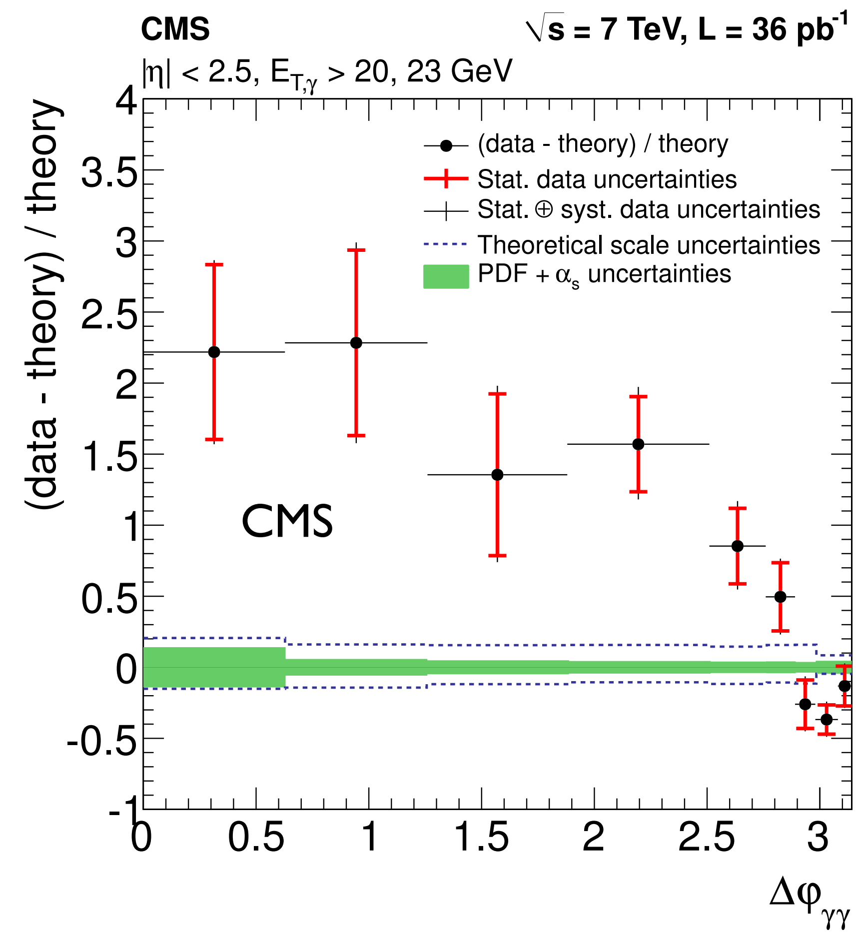
(Underlying Event Tuning)

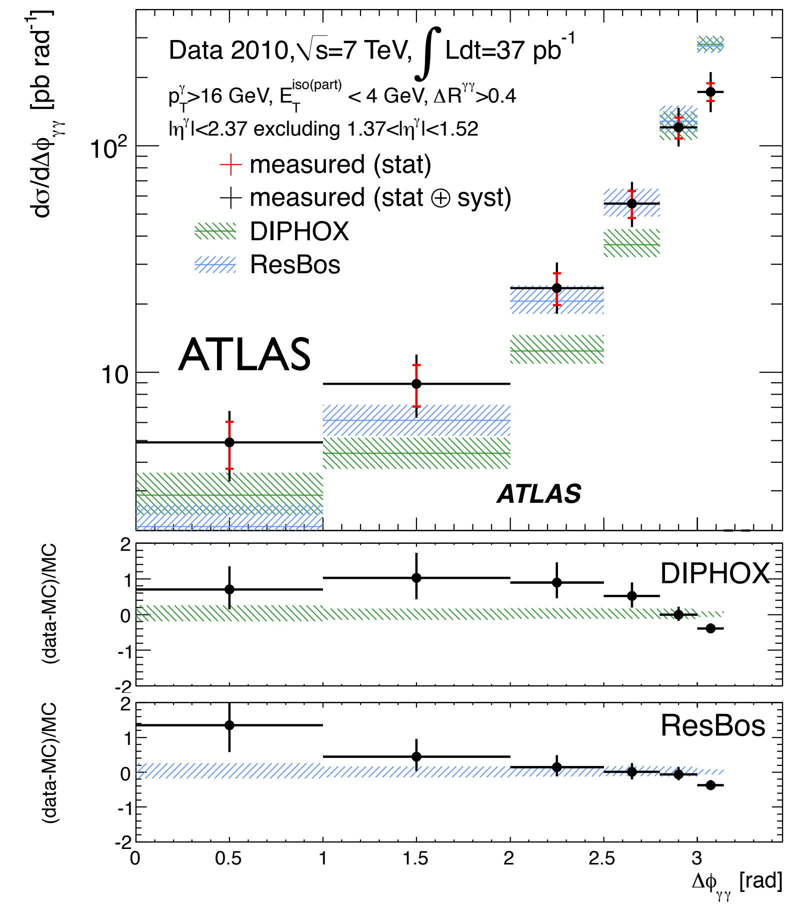


PS: yes, we **should** update the PYTHIA 6 defaults ...

γγ







Disagreement much smaller in ATLAS study arXiv:1107.0581

See also Mangano's from this workshop, slide 12
+ Daniel's talk? (slides not posted when writing this)

Disclaimer

Not an expert on H searches

How well do we know Theory?

Executive summary of issues and ambiguities

How well do we describe LHC?

Hadronization, Underlying Event (UE) and Pile-Up

→ MC Modeling and Constraints (tuning)

For Discussion

Areas of improvement with importance for Higgs Searches?