Event-Generator Tuning — Overview

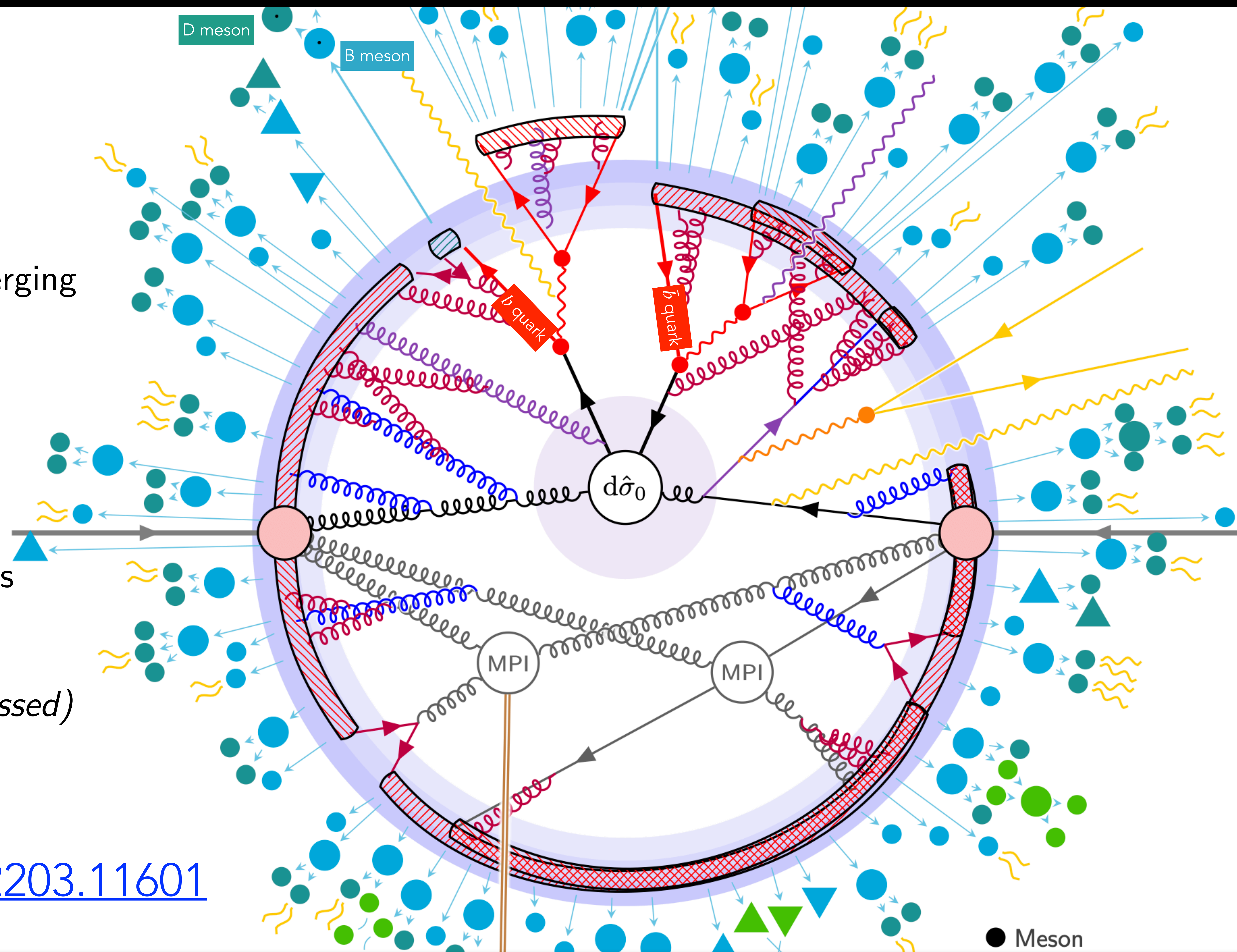
P. Skands (Monash University)

What are we tuning? Components of a Modern Monte Carlo Event Generator:

Parton Level

- Hard Interaction
- Resonance Decays
- MECs, Matching & Merging
- FSR
- ISR*
- QED
- Weak Showers
- Hard Onium
- Multiparton Interactions
- Beam Remnants*

(: incoming lines are crossed)*



Hadron Level

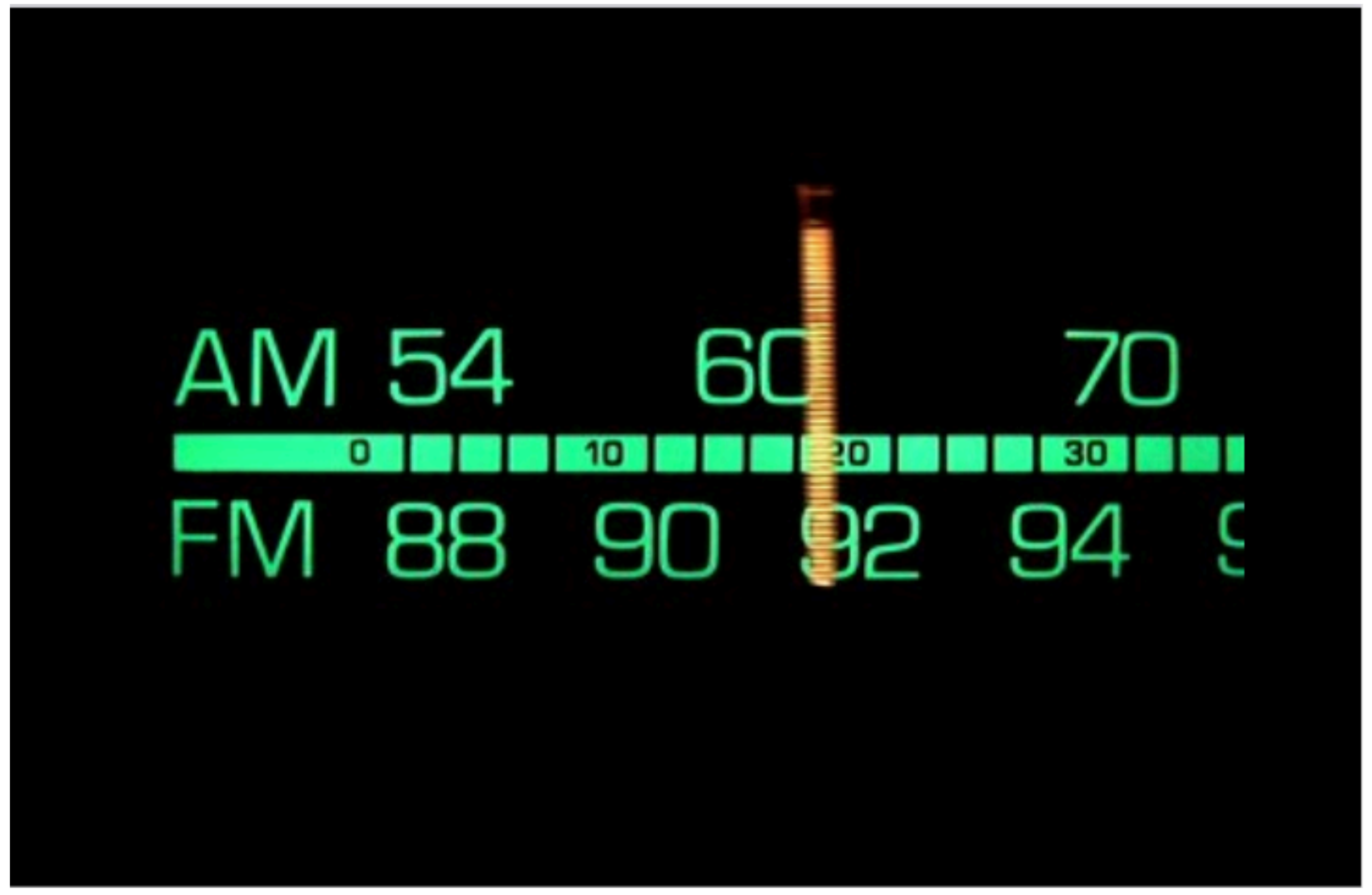
- Beam Remnants*
- Strings
- Clusters
- Colour Reconnections
- String Interactions
- Bose-Einstein & Fermi-Dirac
- Primary Hadrons
- Secondary Hadrons
- Hadronic Reinteractions
- QED in Hadron Decays

(: incoming lines are crossed)*

- Meson
- Heavy Flavour
- Baryon
- Antibaryon

Figure from arXiv:2203.11601

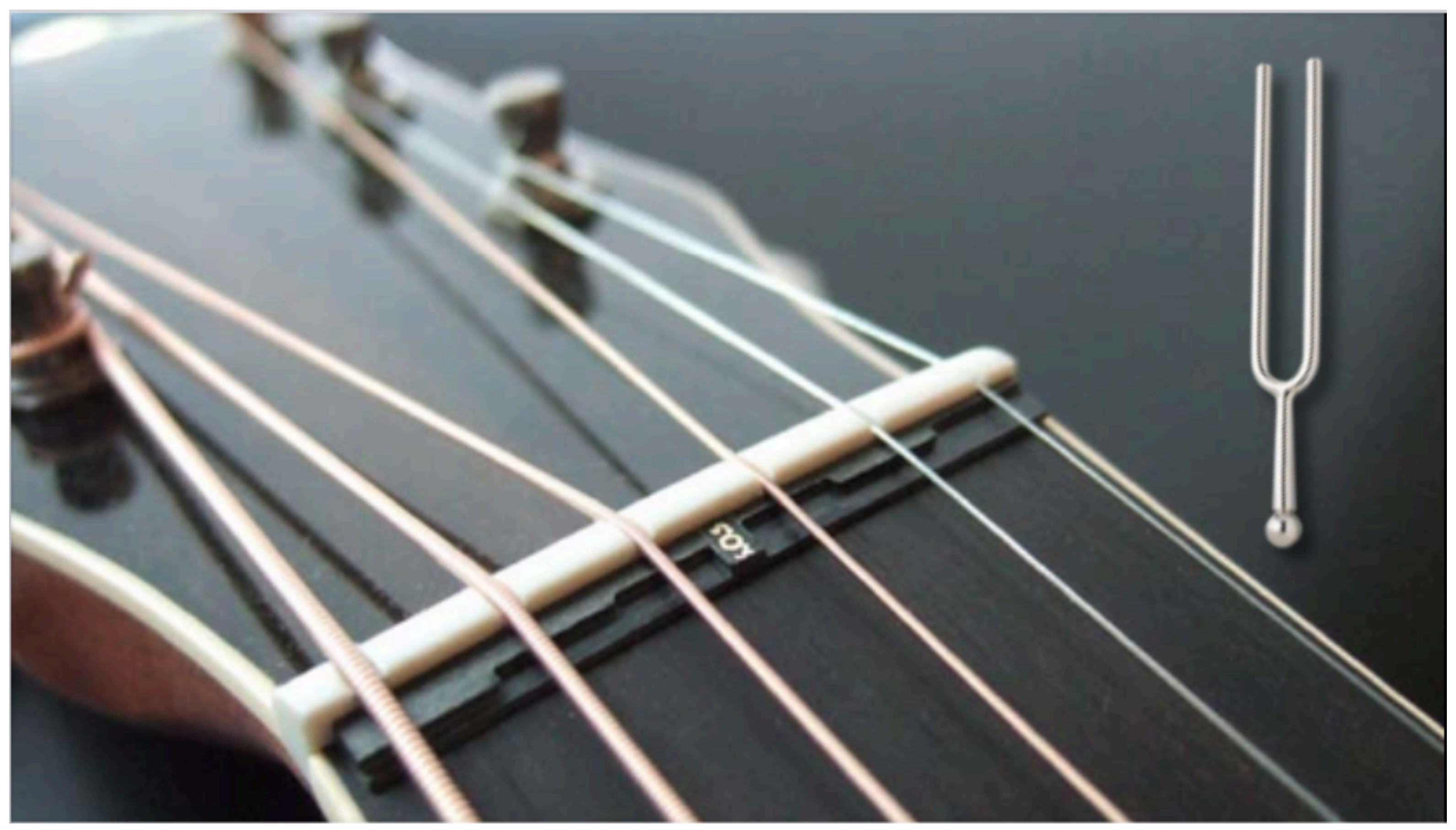
Tuning: What do you want it to be?



Sensible

A set of physically sensible central parameter values, with good universality.

What does "physically sensible" and "good universality" really mean?
Understanding MC models: hierarchies, universalities, and sensitivities.



Sophisticated

High-precision & specialised parameter sets, with reliable uncertainties

Tuning in the context of N^nLO matching & precision applications.
Theory uncertainties. Rigorous scientific analyses of parameter spaces.

Best Fit

A pure optimisation problem. The **best fit** you can get. Ask questions later.


Risky? Overfitting, oversimplification, GIGO, black-box syndrome, tunnel vision, loss of insight & scientific rigour, Tyranny of Carlo,…

"The Tyranny of Carlo" *[J. D. Bjorken, ca. 1990]*

"Another change that I find disturbing is the rising tyranny of Carlo. No, I don't mean that fellow who runs CERN [Rubbia], but the other one, with first name Monte.

The simultaneous increase in detector complexity and in computation power has made simulation techniques an essential feature of contemporary experimentation. The MC simulation has become the major means of visualization of not only detector performance but also of physics phenomena. **So far so good.**

But it often happens that the physics simulations provided by the MC generators carry the authority of data itself. They look like data and feel like data, and if one is not careful they are accepted as if they were data. **All Monte Carlo codes come with a GIGO* warning label**. But that warning label is just as easy for a physicist to ignore as that little message on a packet of cigarettes is for a chain smoker to ignore. I see nowadays experimental papers that claim agreement with QCD (translation: someone's simulation labeled QCD) and/or disagreement with an alternative piece of physics (translation: an unrealistic simulation), without much evidence of the **inputs into those simulations**."

Account for what is included in the models, parameters, pertinent cross-checks and validations. Do serious effort to estimate uncertainties, by salient MC variations.

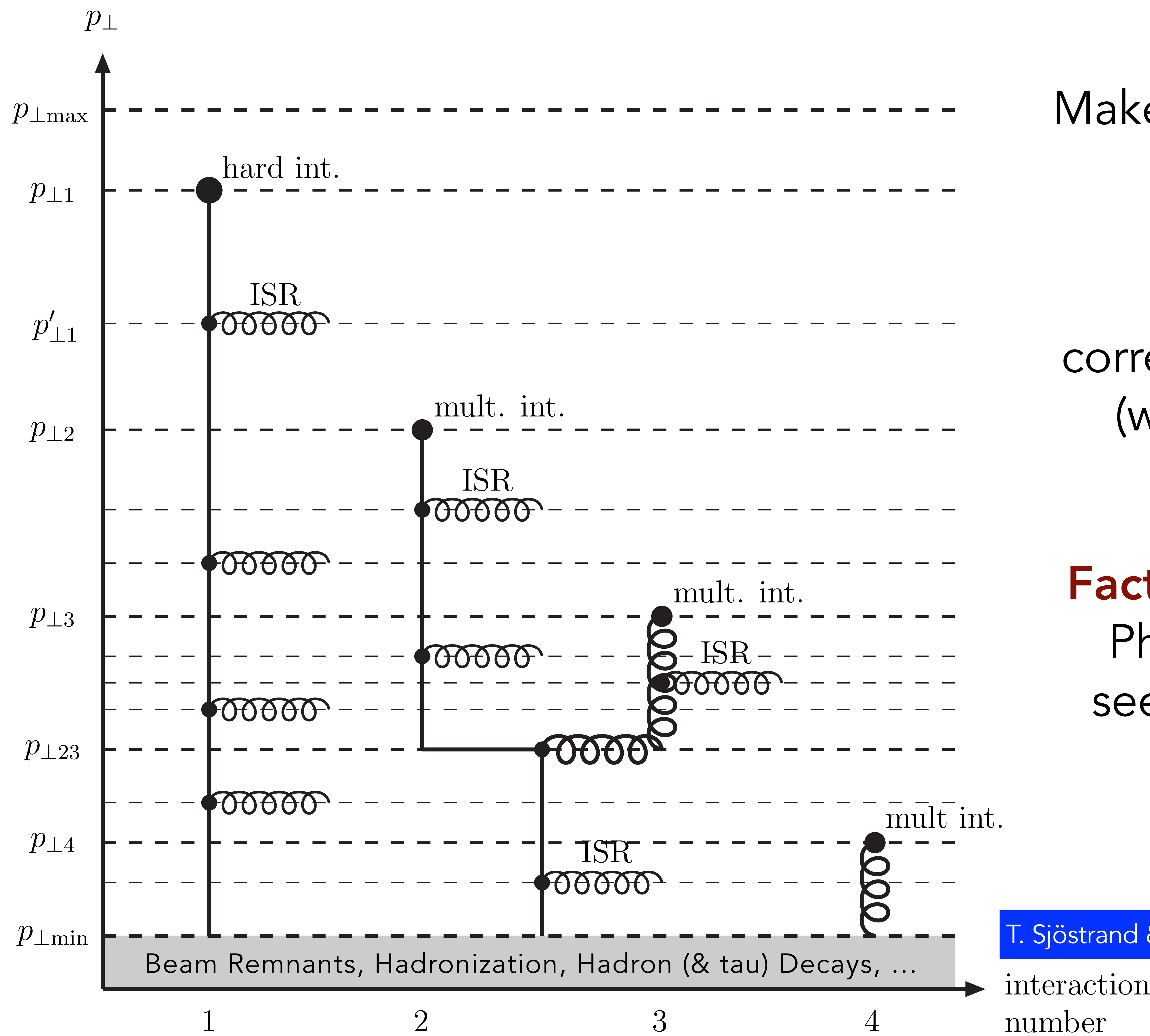
*GIGO: Garbage In, Garbage Out

Understanding MC Models: **Event Evolution**

Physics Separation of scales ➤ Factorizations **Maths**

➜ **Divide and Conquer:** $\mathcal{P}_{\mathrm{event}} = \mathcal{P}_{\mathrm{hard}} \otimes \mathcal{P}_{\mathrm{dec}} \otimes \mathcal{P}_{\mathrm{ISR}} \otimes \mathcal{P}_{\mathrm{FSR}} \otimes \mathcal{P}_{\mathrm{MPI}} \otimes \mathcal{P}_{\mathrm{Had}} \otimes \ldots$

Event Evolution ~ Increasing Detail



T. Sjöstrand & PZS, Eur.Phys.J.C 39 (2005)

Make random choices ~ as in nature

➜ **Stochastic Sampling**

Constrain / optimise the corresponding probability densities (within theoretical constraints!)

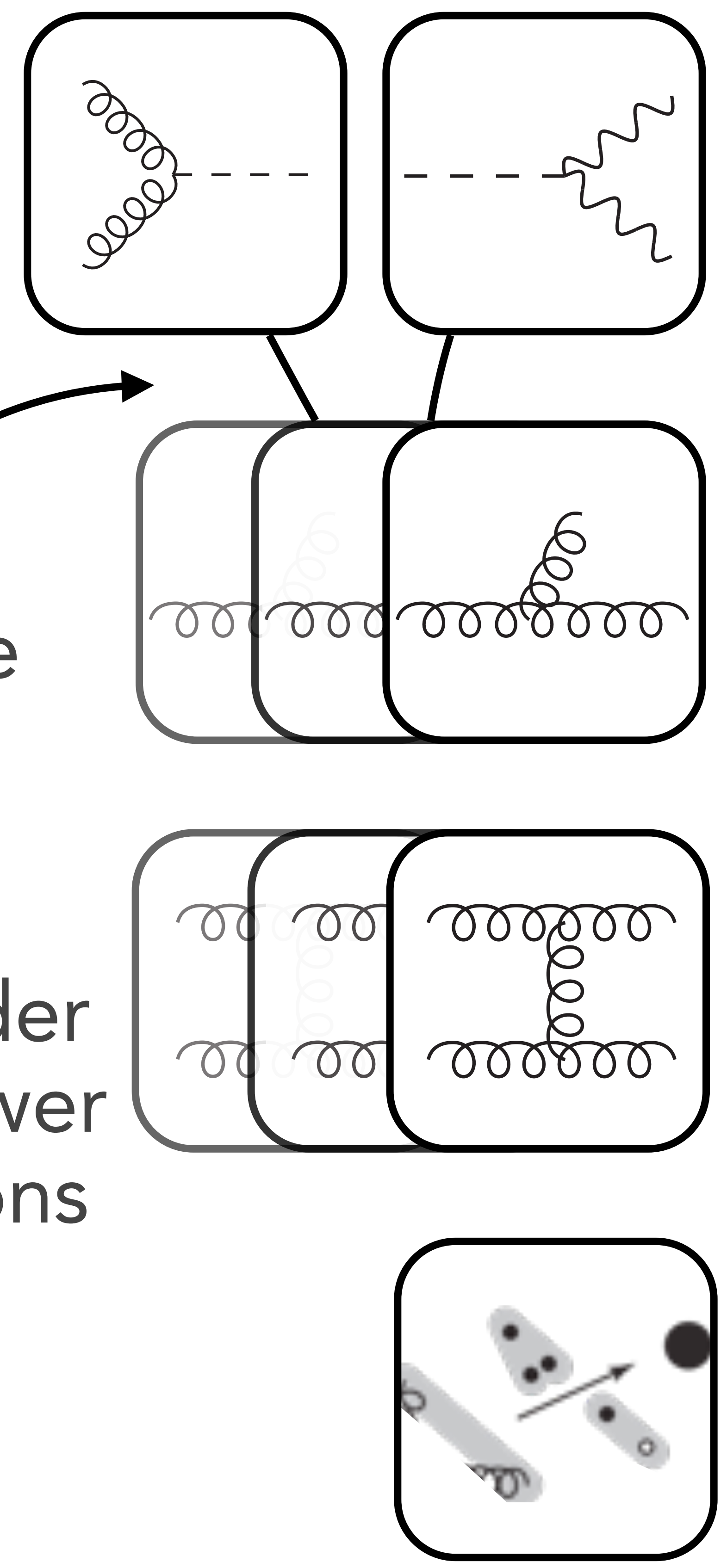
➜ **Tuning**

Factorizations & Event Evolution:

Physics at "high scales" acts as seeds for physics at "low scales"

➜ **Parameter Hierarchies**

Event Evolution ~ Increasing Detail



Merging

Eliminate double-counting between fixed-order and shower corrections

Hard Process & Decays:

Use LO / NLO / NNLO matrix elements (e.g., $gg \to H^0 \to \gamma\gamma$)

→ Sets "hard" resolution scale for process: Q_{MAX}

ISR & FSR (Initial- & Final-State Radiation):

Driven by differential (e.g., DGLAP) evolution equations, dP/dQ^2, as function of resolution scale; from Q_{MAX} to $Q_{HAD} \sim 1$ GeV

MPI (Multi-Parton Interactions)

Protons contain lots of partons → can have additional (soft) parton-parton interactions → Additional (soft) "Underlying-Event" activity

Hadronisation and Hadron Decays

Non-perturbative modeling of partons → hadrons transition

Tuning: the higher up the chain you change something, the more it will affect the large-scale event structure ➜ Start at the top, and work your way down.

Divide and Conquer: Use Ratios, Exclusivity, and Infrared Safety to exploit factorisations!

Parameters (in PYTHIA): **FSR pQCD Parameters**

Matching

Additional Matrix Elements included?

At tree level / one-loop level? Using what matching scheme?

$\alpha_s(m_Z)$

The value of the strong coupling

In PYTHIA, you set an effective value for $\alpha_s(m_Z^2)$ ⇔ choice of k in $\alpha_s(kp_\perp^2)$

α_s Running

Renormalization Scheme and Scale for α_s

1- vs 2-loop running, MSbar / CMW scheme, choice of k in $\alpha_s(kp_\perp^2)$, cf

Subleading Logs

Ordering variable, coherence treatment, effective 1→3 (or 2→4), recoil strategy, …

Branching Kinematics (z definitions, local vs global momentum conservation), hard parton starting scales / phase-space cutoffs, masses, non-singular terms, …

Hadron energy fractions

Fragmentation Function

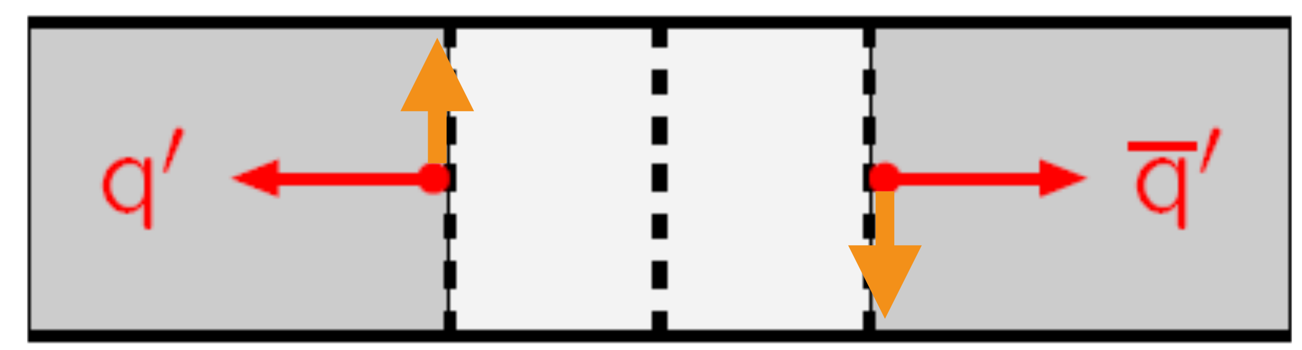
The "Lund a and b parameters" (and $\Delta a_{\mathrm{diquark}}$ for baryons)

Or use a and $\langle z \rangle$ instead (less correlated) A. Jueid et al., *JCAP* 05 (2019) 007

p_T in string breaks

Scale of string-breaking process

Shower cutoff and $\langle p_\perp \rangle$ in string breaks



Meson Multiplets

Mesons

Strangeness suppression, **Vector/Pseudoscalar**, η, η′, …

Baryon Multiplets

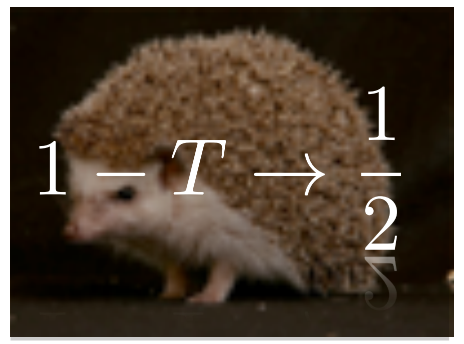
Baryons

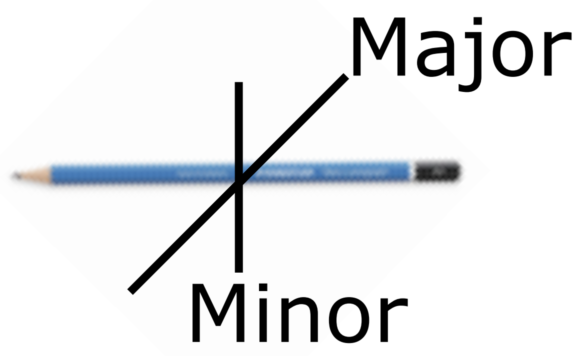
Baryon-to-meson ratios, **Spin-3/2 vs Spin-1/2**, "popcorn", colour reconnections (junctions), … ?

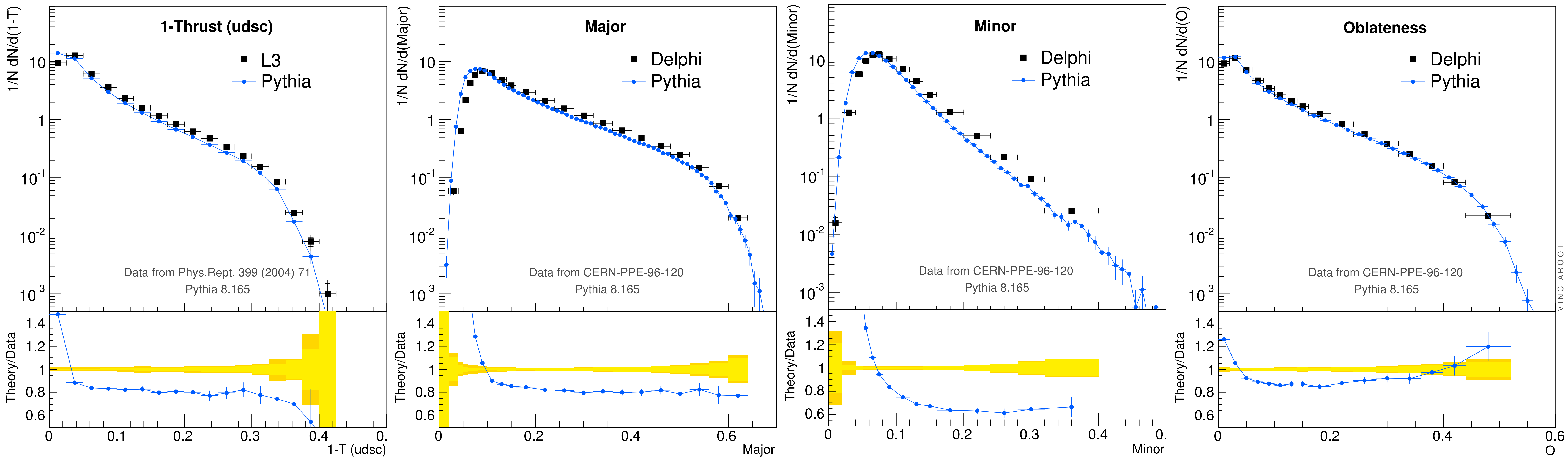
PYTHIA 8 (hadronization on) vs LEP: Thrust

$$T = \max_{\vec{n}} \left(\frac{\sum_i |\vec{p}_i \cdot \vec{n}|}{\sum_i |\vec{p}_i|} \right)$$

$1 - T \to 0$





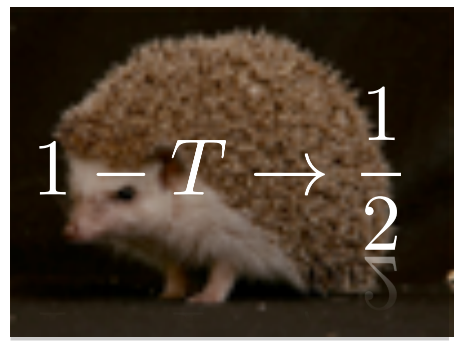


Using effective $\alpha_s(M_Z) = 0.12$

PYTHIA 8 (hadronization on) vs LEP: Thrust

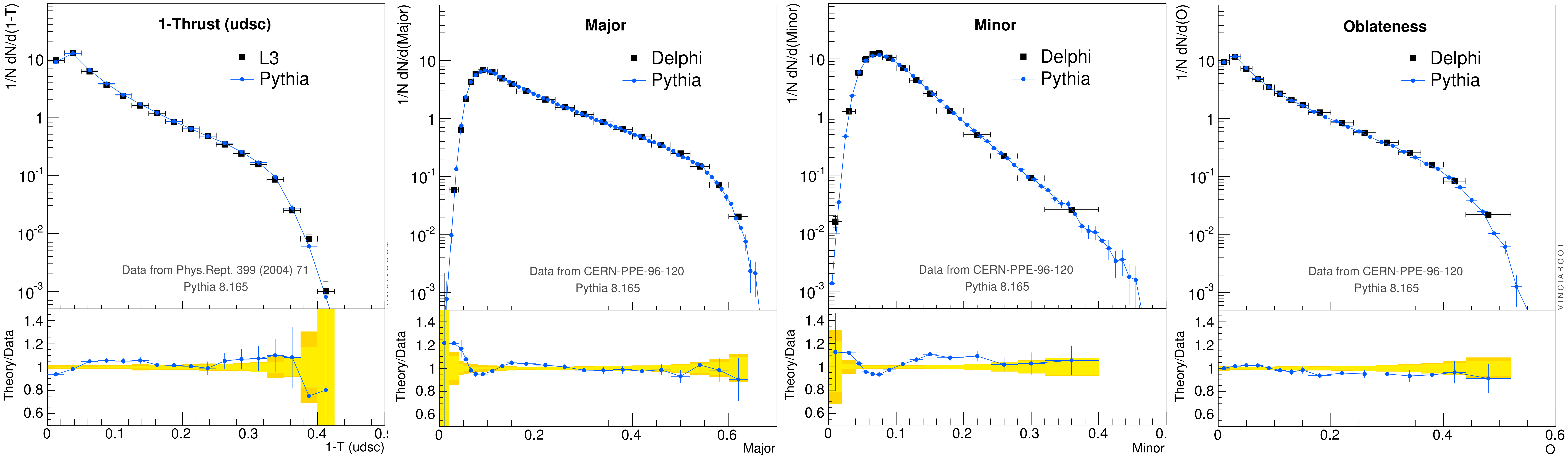
$$T = \max_{\vec{n}} \left(\frac{\sum_i |\vec{p}_i \cdot \vec{n}|}{\sum_i |\vec{p}_i|} \right)$$

$1 - T \to 0$



Major

Minor



Using effective $\alpha_s(M_Z) = 0.14$

Wait … is this Crazy?

Best result

Obtained with $\alpha_s(M_Z) \approx 0.14$

$\neq$ World Average ~ 0.118

Effective value of α_s depends on the order and scheme

Baseline MC ≈ Leading Order + LL resummation

Other leading-Order extractions of $\alpha_s \approx 0.13$ - 0.14

Effective scheme interpreted as “CMW” → 0.13 Catani, Marchesini, Webber, *Nucl.Phys.B* 349 (1991) 635-654

2-loop running → 0.127; NNLO Matching → 0.12 Hartgring, Laenen, PZS, *JHEP 10 (2013) 127* ; see also backup slides

Not so crazy (but does rely on “magic” mathematical accident in Z decay)

Let parameters vary to a level consistent with the (limited) formal accuracy.

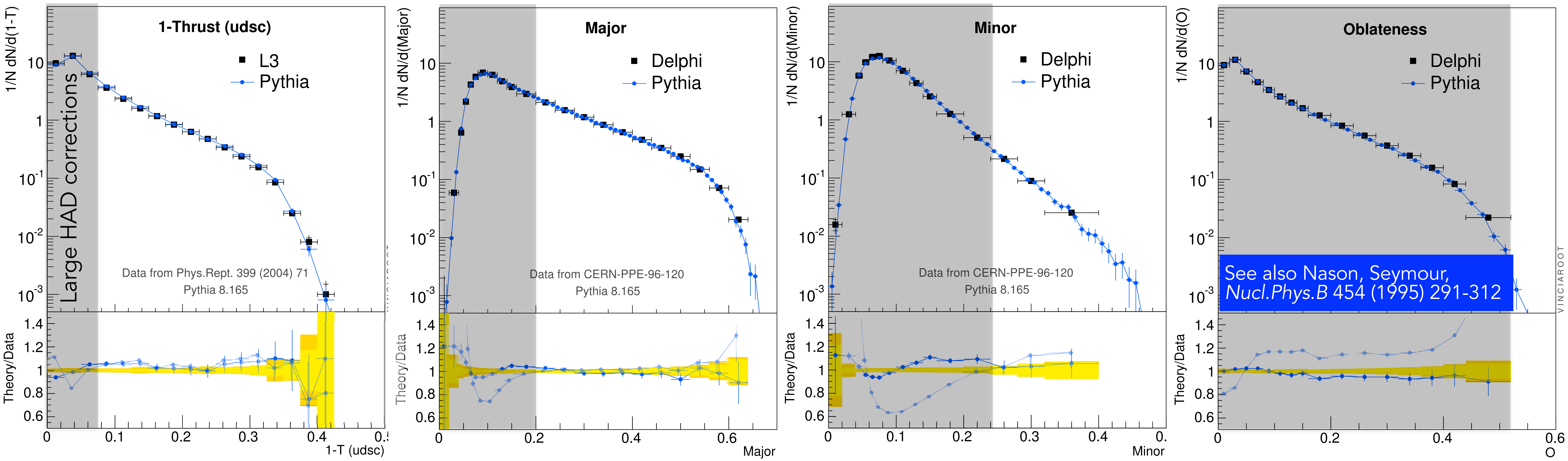
Sanity check = consistency with other determinations at a similar formal order, within the uncertainty at that order (including a CMW-like scheme redefinition to go to ‘MC scheme’)

To improve systematically → Merging at NLO

PYTHIA 8 (hadronization on) **Vs** (hadronization off)

Important point: These observables are **IR safe ➜ minimal hadronisation corrections**

Big differences in **how** sensitive each of these are to hadronisation & over what **range**

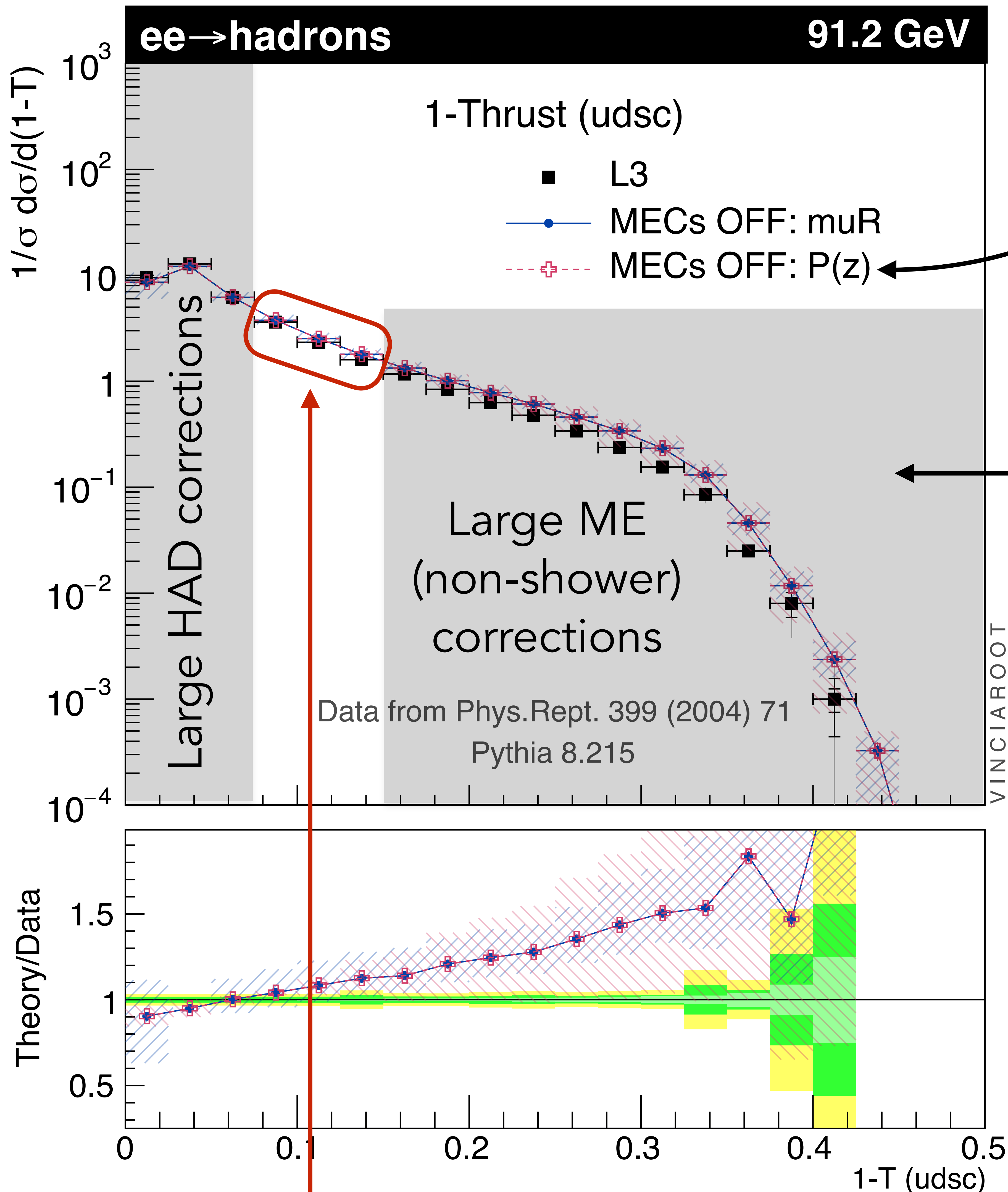


Large sensitivities to “lower” phenomena break the divide-and-conquer simplification.

Another **important** point: **peaks** of distributions are all where **HAD sensitivity is highest**!

… and **sensitivity** to **fixed-order** corrections

(Adding nuisance terms $\Delta P(z) \propto Q^2$ to the splitting kernels beyond shower accuracy)



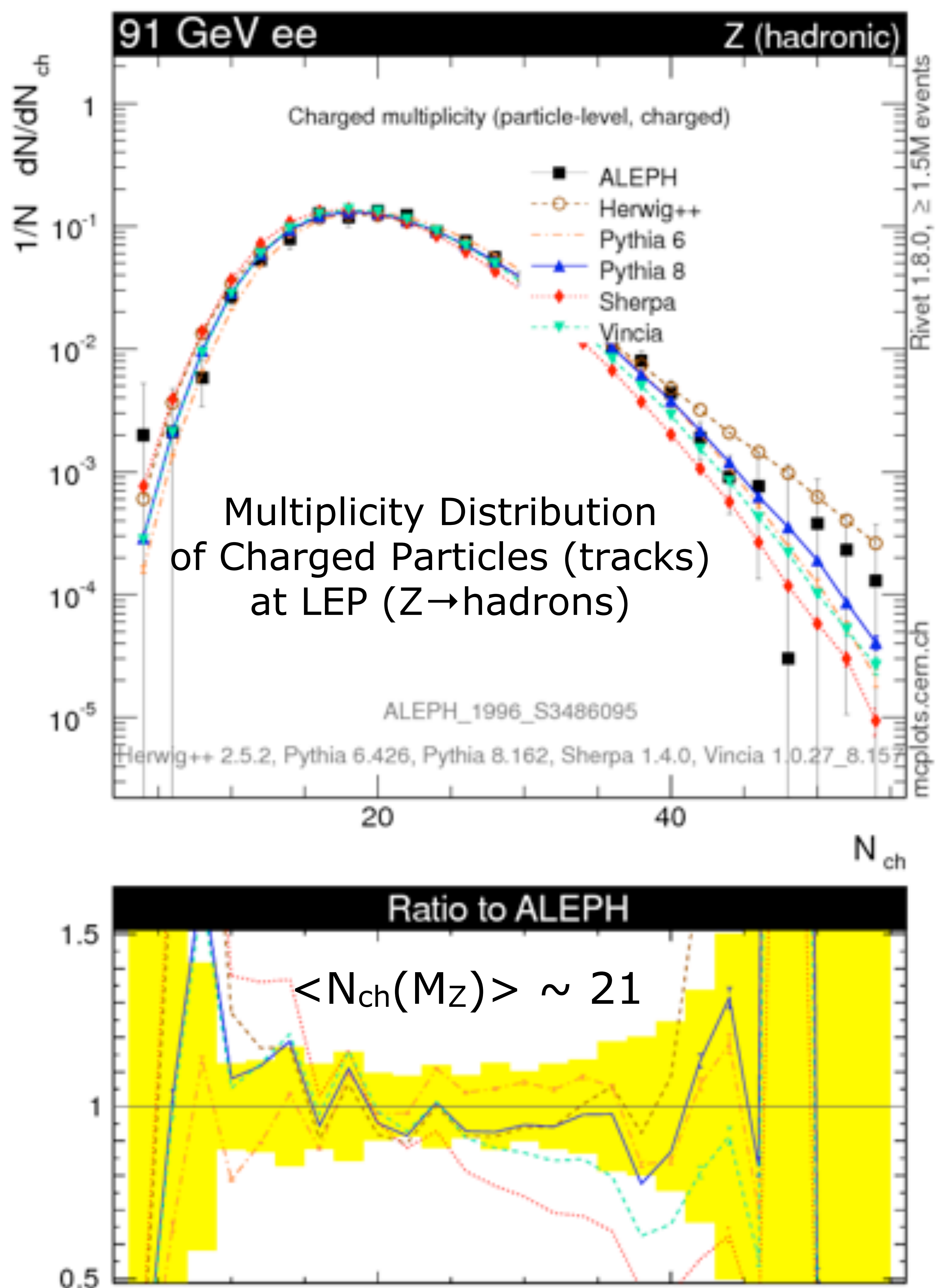
These points are quite sensitive to MECs / Matching / Merging.

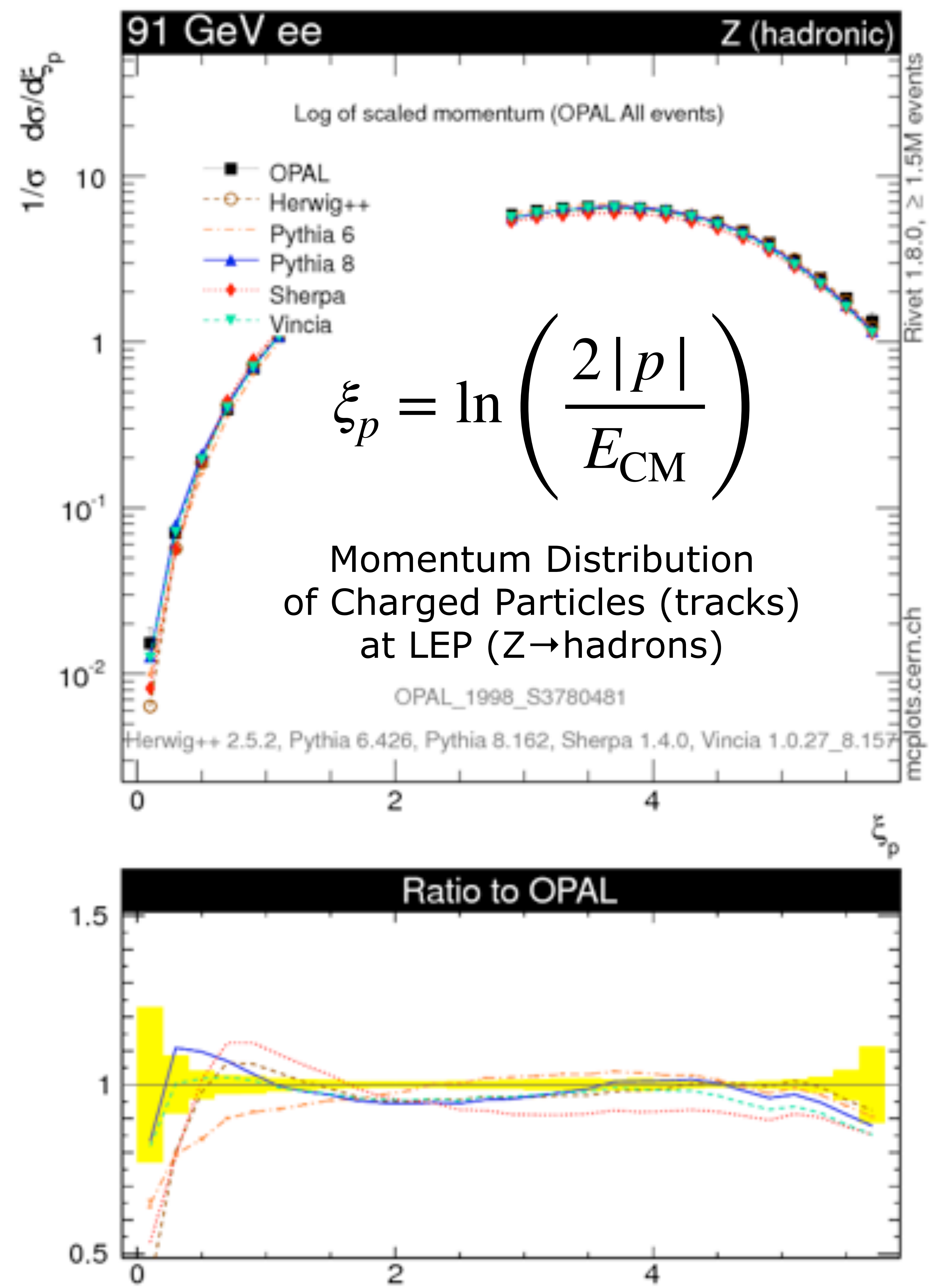
➜ we should ensure we do MECs / matching / merging if we want to use them (or something equivalent to that.)

These points are relatively insensitive to **both** hadronization **and** matching/merging

Hadronization Corrections: Fragmentation Tuning

Now use infrared **sensitive** observables - sensitive to hadronization + first few bins of previous (IR safe) ones



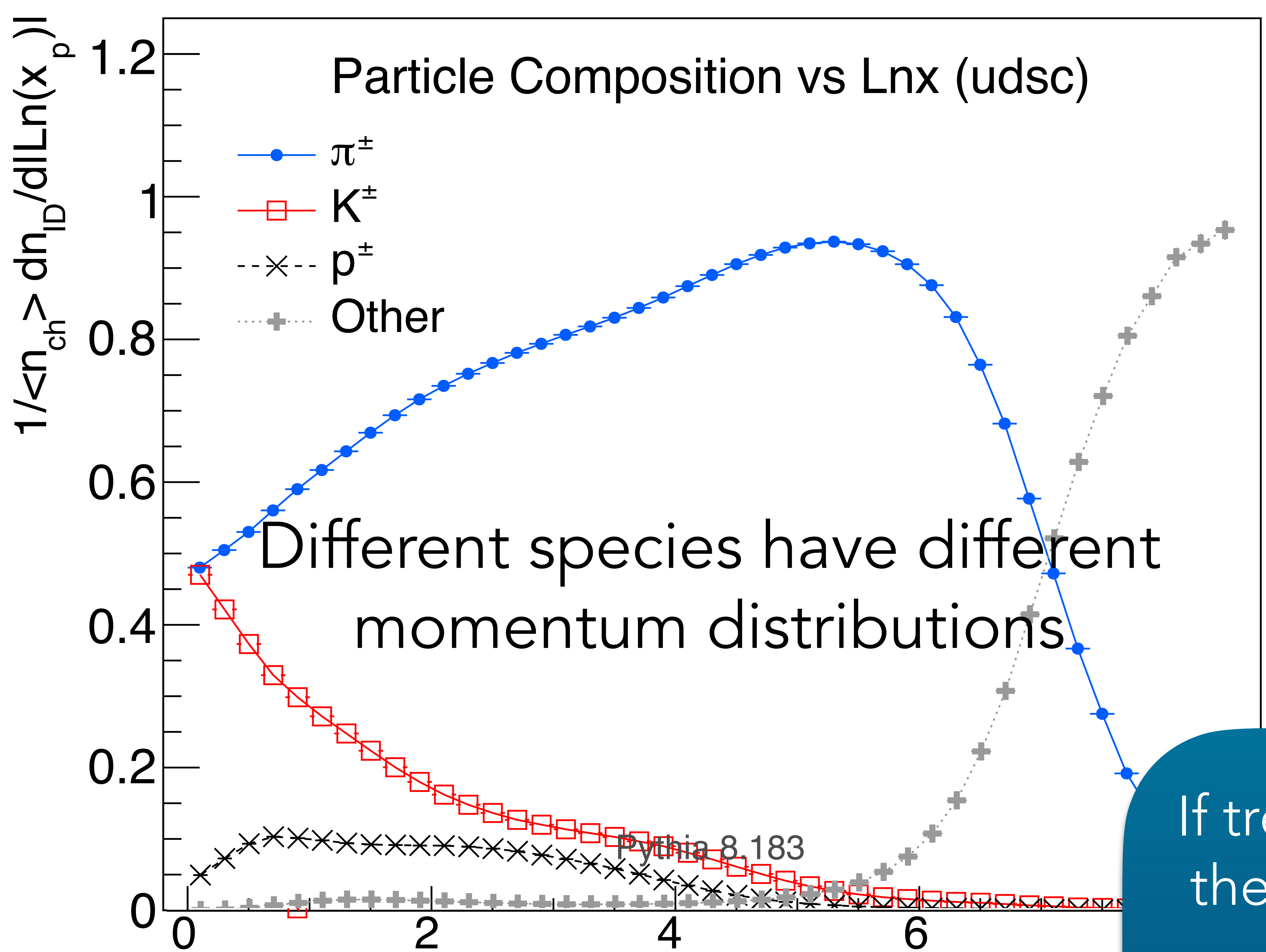


→ Tutorial

Fragmentation Tuning

Note: use infrared-**unsafe** observables - sensitive to hadronization (example)

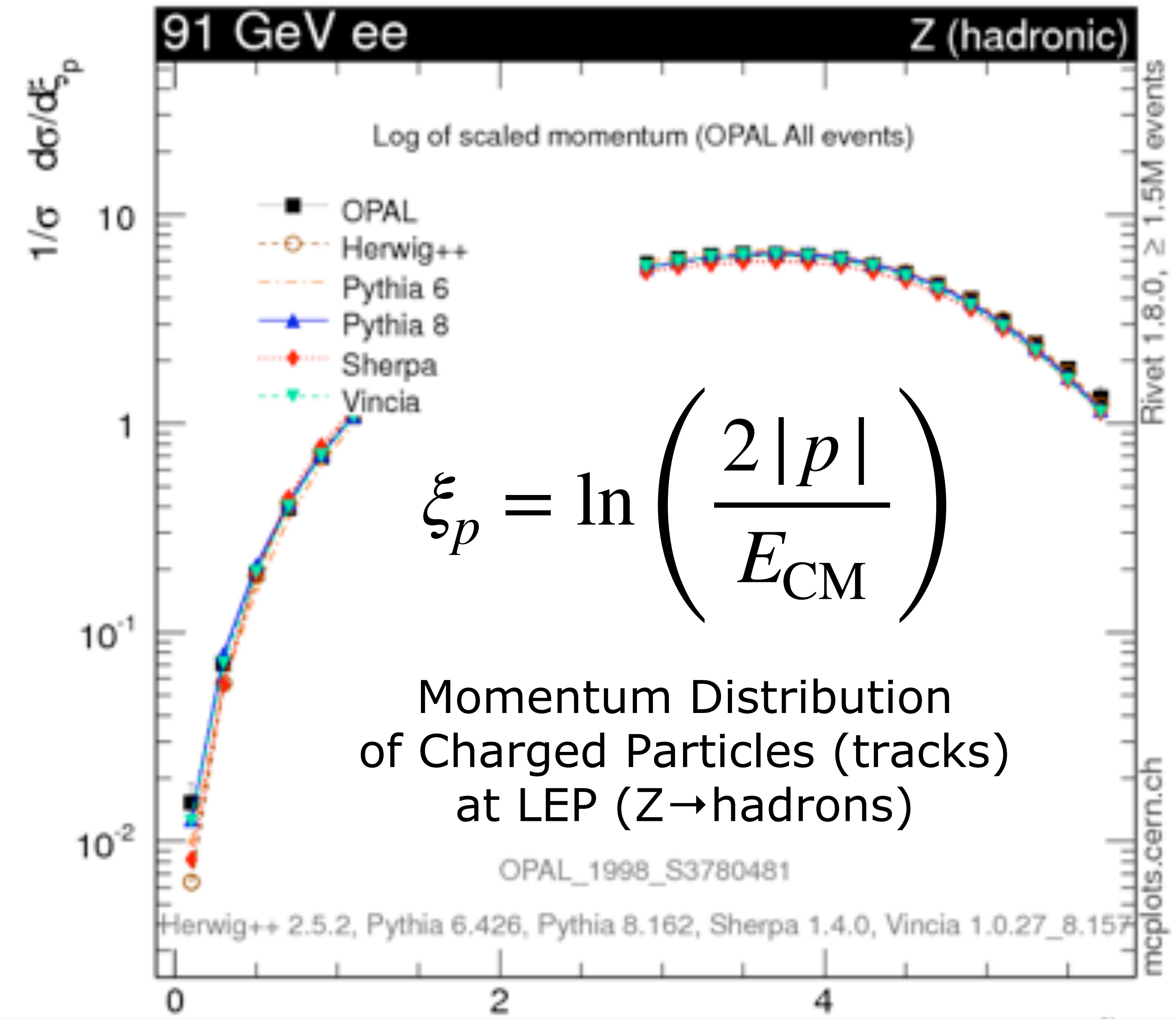
Know what **physics** goes in



Different species have different momentum distributions

+ effects of feed-down!

(e.g., $\rho \to \pi\pi$, $K^* \to K\pi$, $\eta \to \pi\pi\pi$,



$$\xi_p = \ln\left(\frac{2|p|}{E_{\mathrm{CM}}}\right)$$

Momentum Distribution
of Charged Particles (tracks)
at LEP (Z→hadrons)

If treated like a black box, we could tune the shape of the momentum spectrum solely by modifying eg the relative amounts of strangeness! Bad idea?

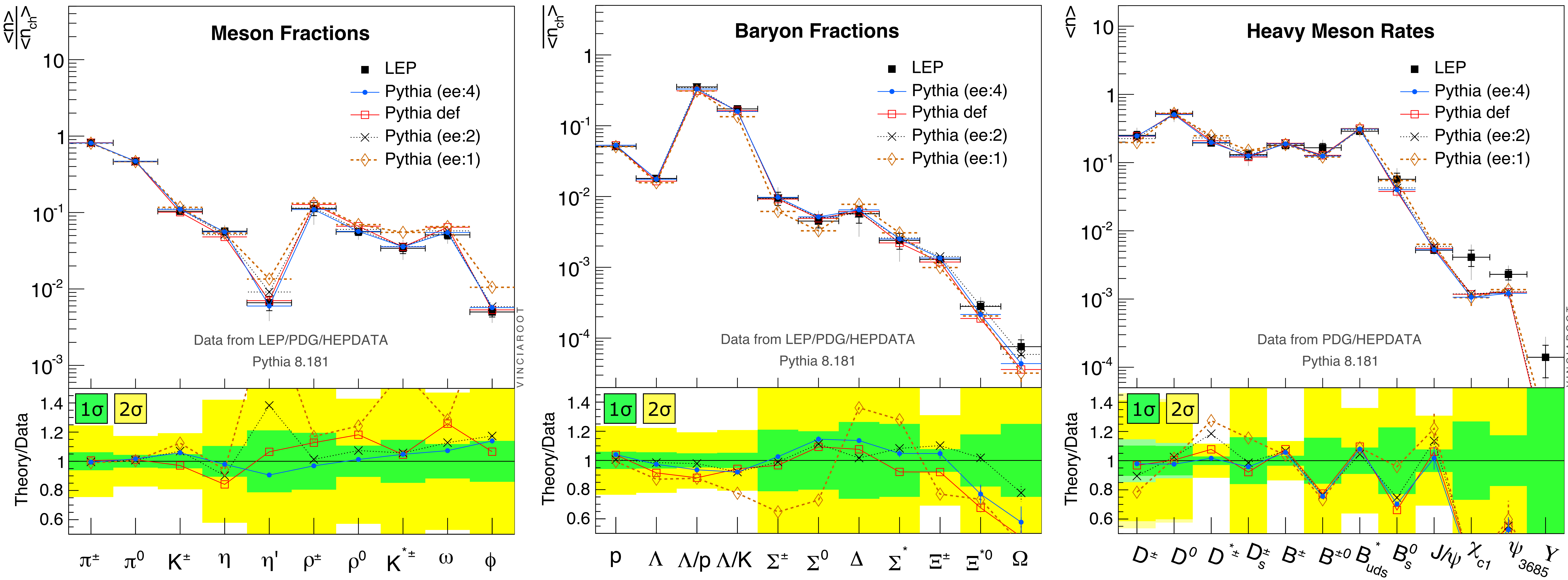
Will get back to that

Need **direct sensitivity to parameters**

Identified Particles

Plenty of observables have **direct** sensitivity to strangeness (& other PID) fractions

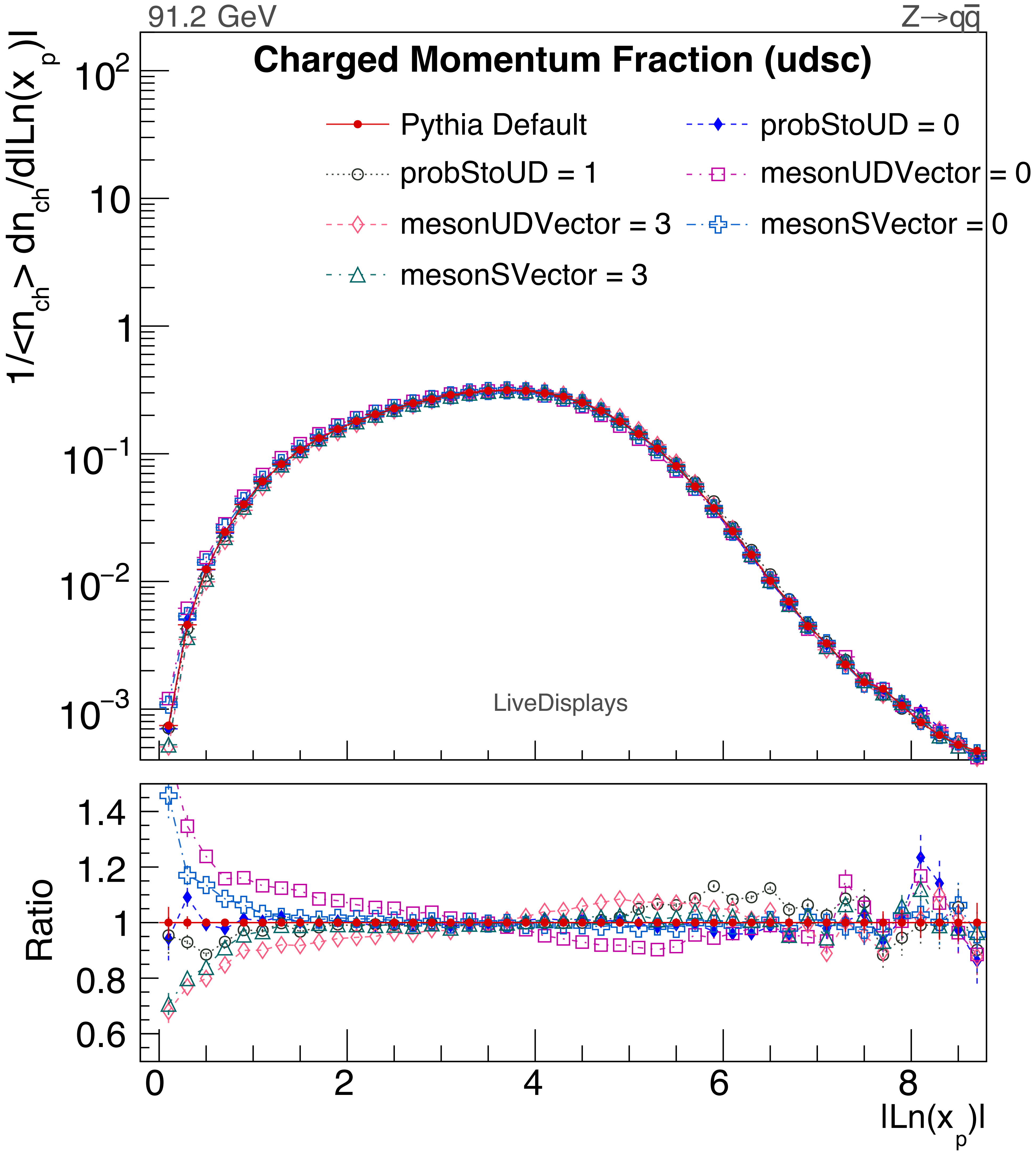
V/P, B/M, $B_{3/2}/B_{1/2}$, strange/unstrange, Heavy, …



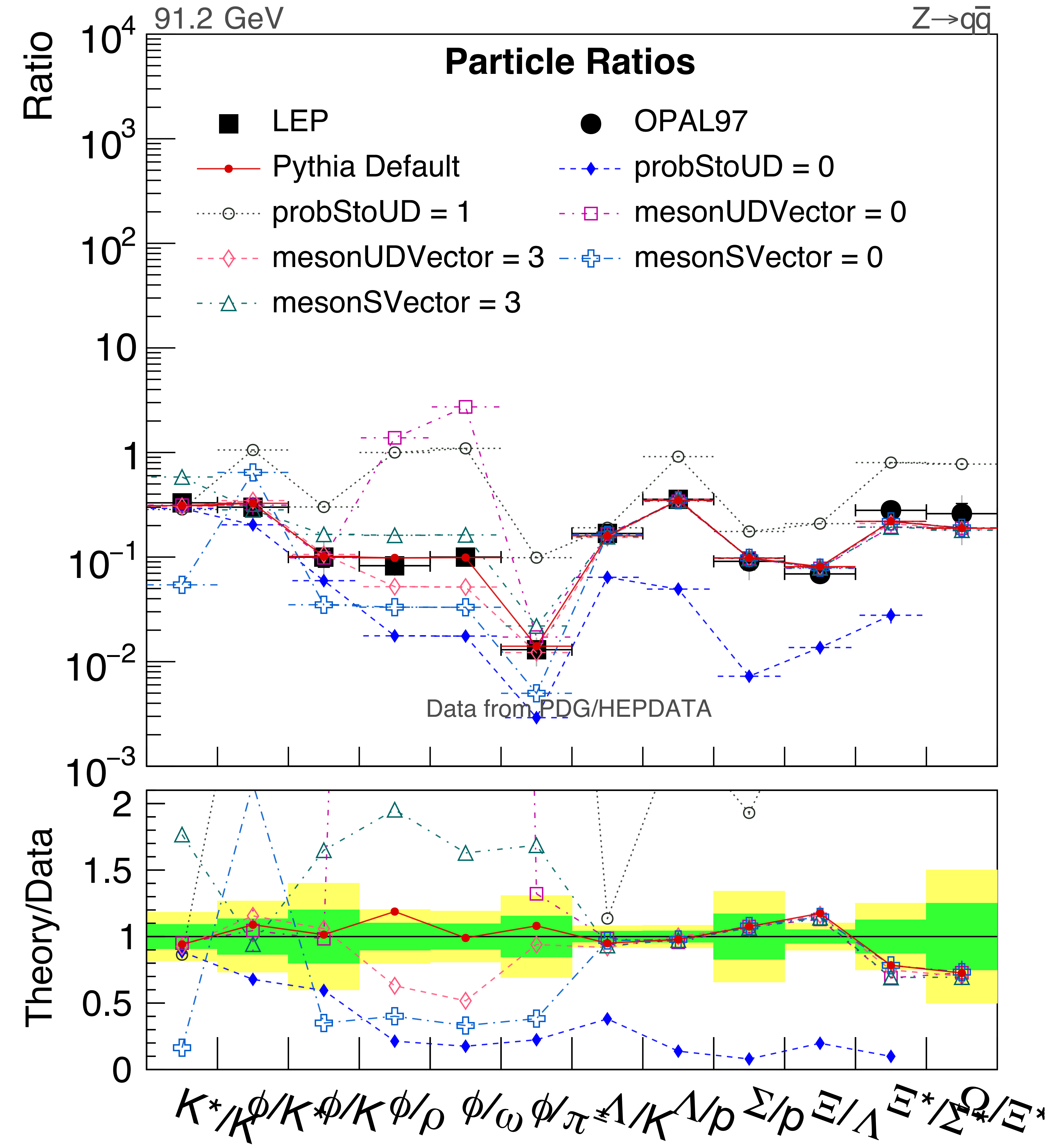
Could be completely mistuned if looking **only** at inclusive charged $\ln(x)$ spectrum

Point: include observables with **direct** sensitivity to each parameter you include.

GIGO



Large changes in strangeness or vector/pseudoscalar ratios **do** modify the momentum spectrum



At the cost of totally destroying agreement with observables that are **directly sensitive** to those parameters

Parameters (in PYTHIA): Initial-State Radiaton

Matching & Merging

Additional Matrix Elements included?

At tree level / one-loop level? What matching scheme?

+ PDF Choice

Size of Phase Space

Starting scale

Relation between Q_{PS} and Q_F (Vetoed showers? Suppressed? cf matching)

Coherence

Initial-Final interference

I-F colour-flow interference effects (eg VBF & Tevatron $t\bar{t}$ asym) & interleaving

α_s

Value and running of the strong coupling

Governs overall amount of radiation (cf FSR)

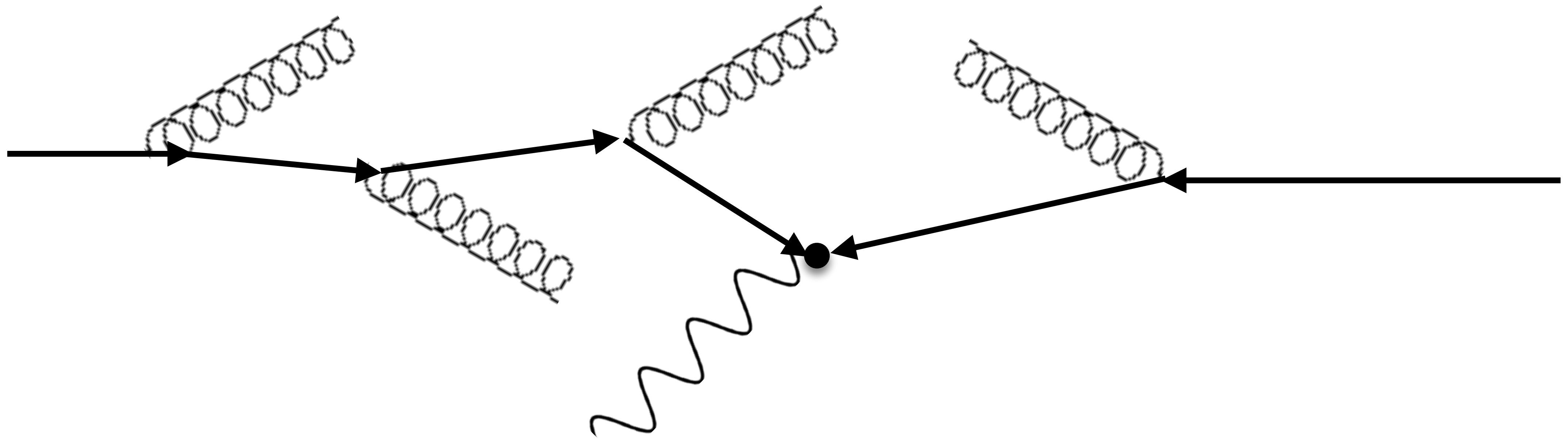
"Primordial kT"

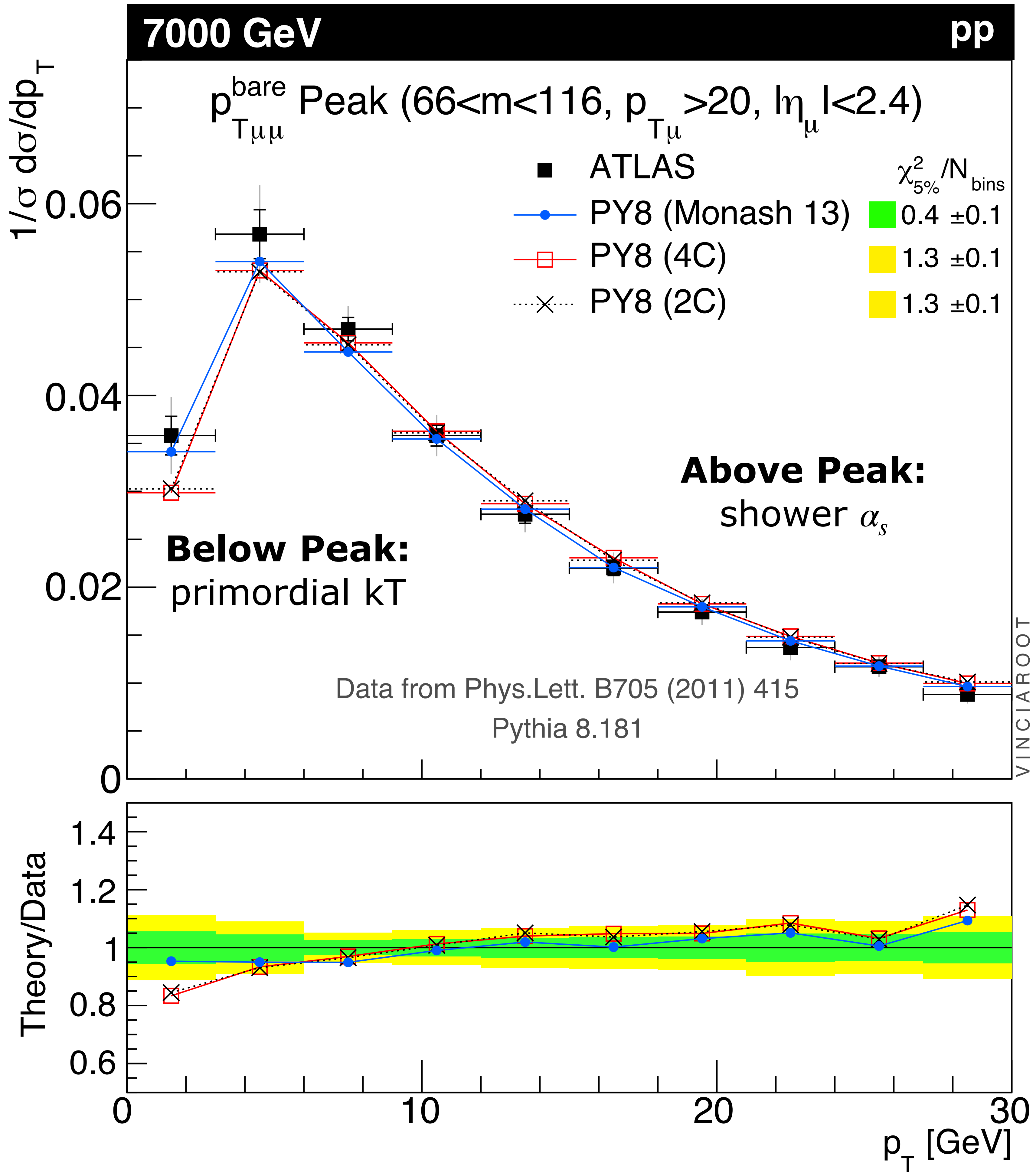
A small additional amount of "unresolved" kT

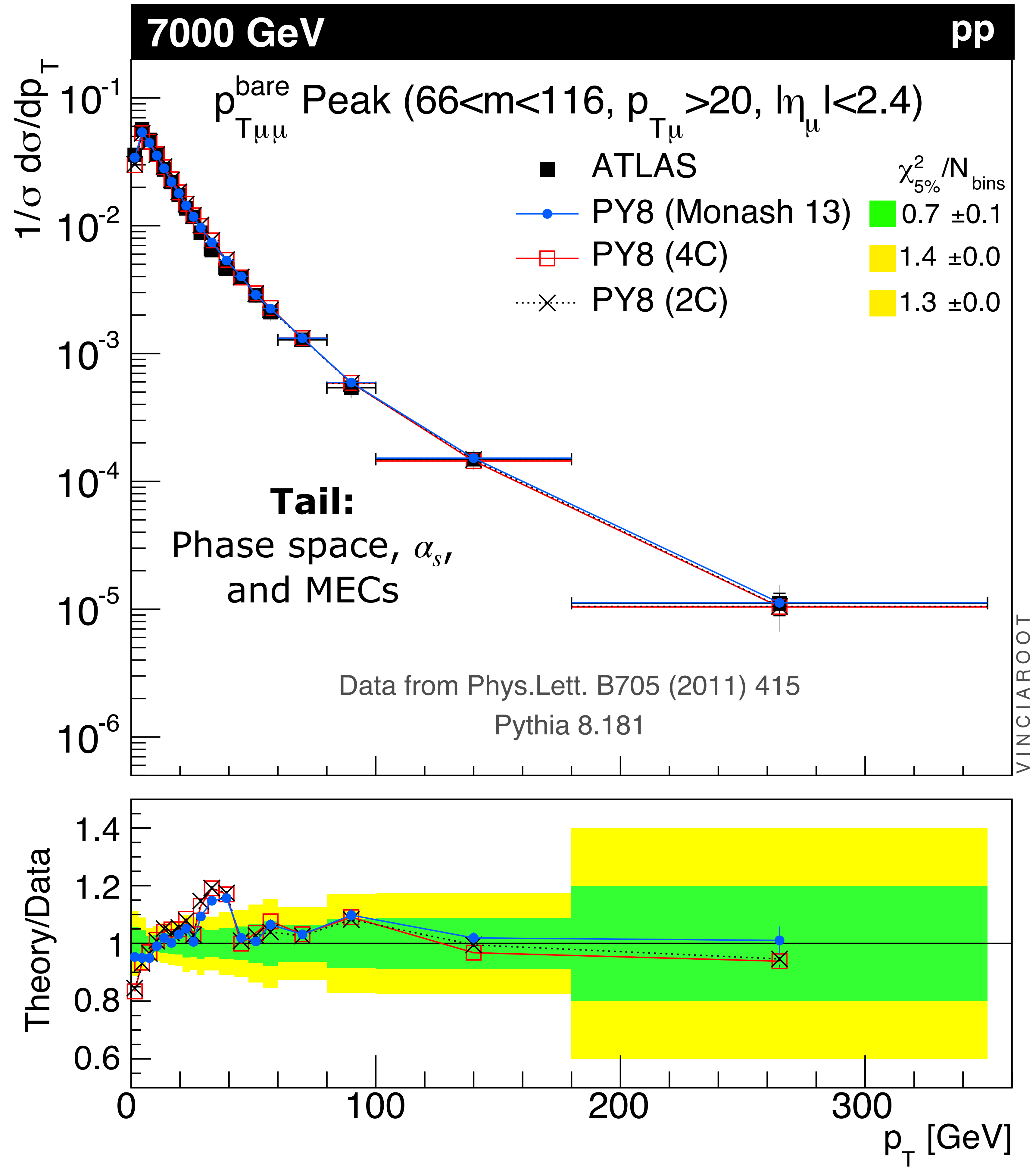
Fermi motion + unresolved ISR emissions + low-x effects?

ISR + Primordial kT

Drell-Yan pT distribution

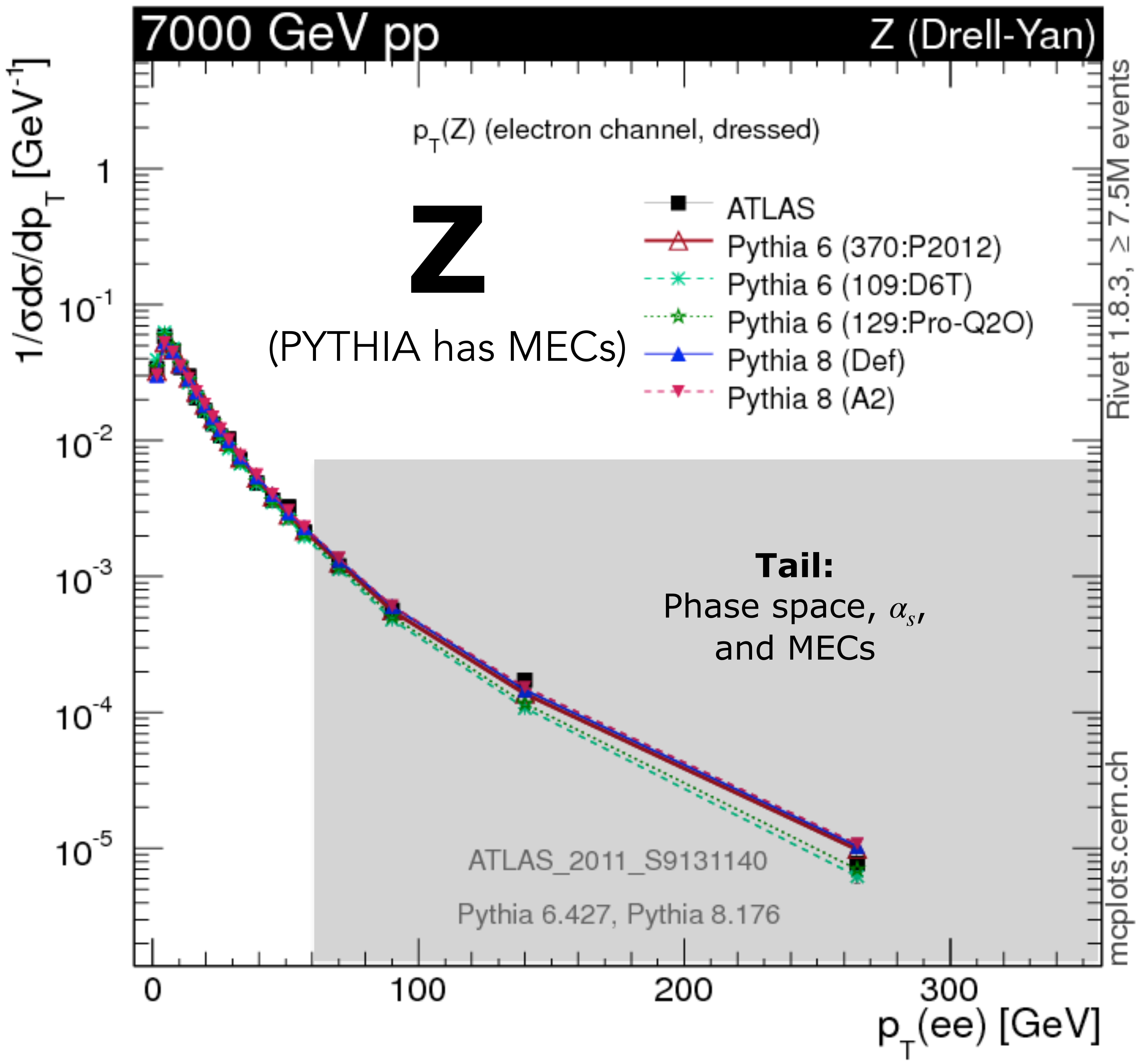




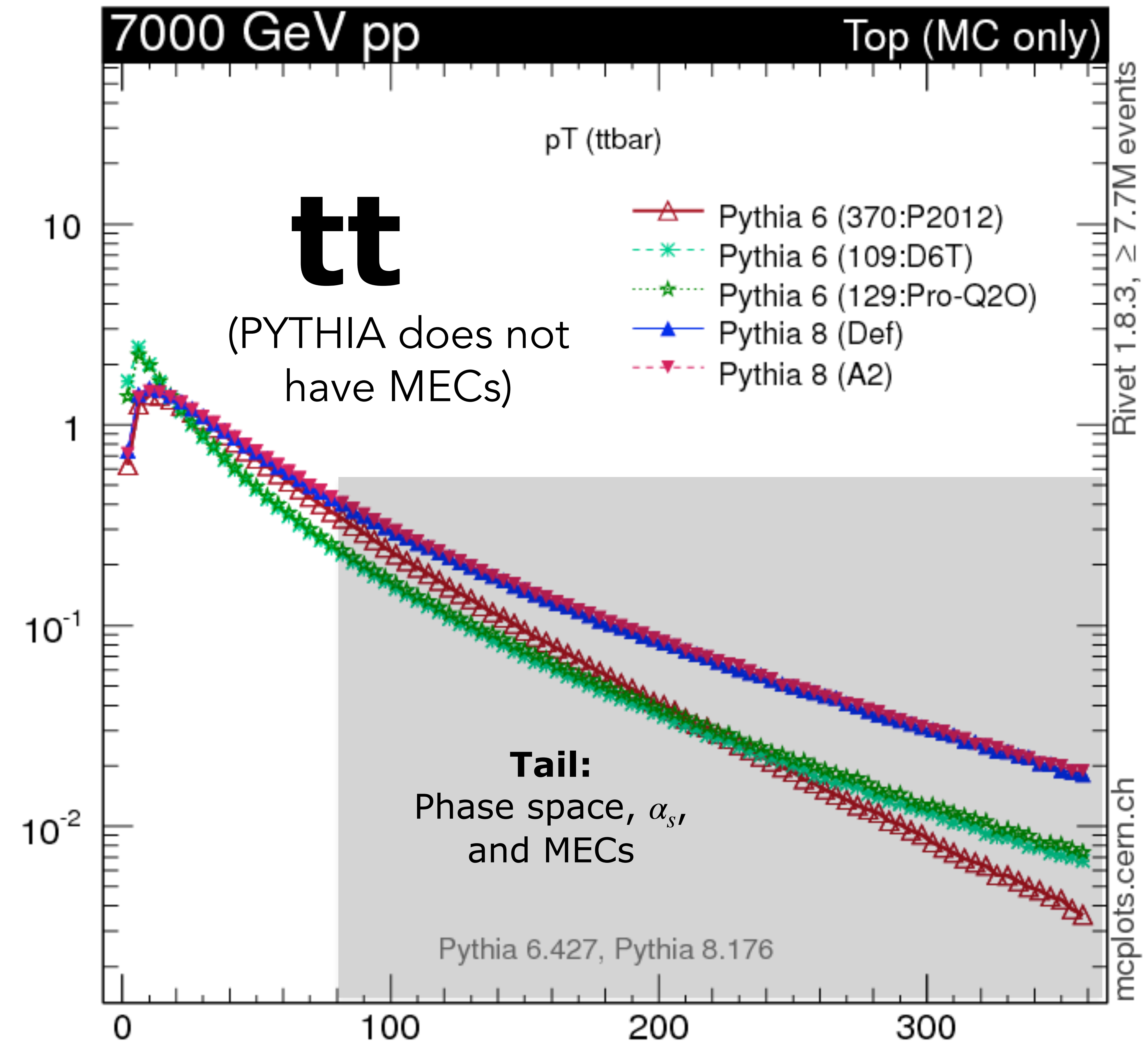


Note: Q.M. requires physical observable!

Beware Process Dependence!



These points are quite sensitive to MECs / Matching / Merging.



➜ we should ensure we do MECs / matching / merging if we want to use them (or something equivalent to that.)

Minimum-Bias & Underlying Event

Number of MPI

Infrared Regularization scale $p_{\perp 0}$ for the QCD 2→2 (Rutherford) scatterings used for multiple parton interactions → size of overall activity

Note: strongly correlated with choice of PDF set! (low-x gluon)

Pedestal Rise

Proton transverse mass distribution → difference between central (more active) **vs peripheral** (less active) **collisions**

Strings per Interaction

Color correlations between multiple-parton-interaction systems (aka colour *reconnections* — relative to LC)

→ shorter or longer strings → less or more hadrons per interaction

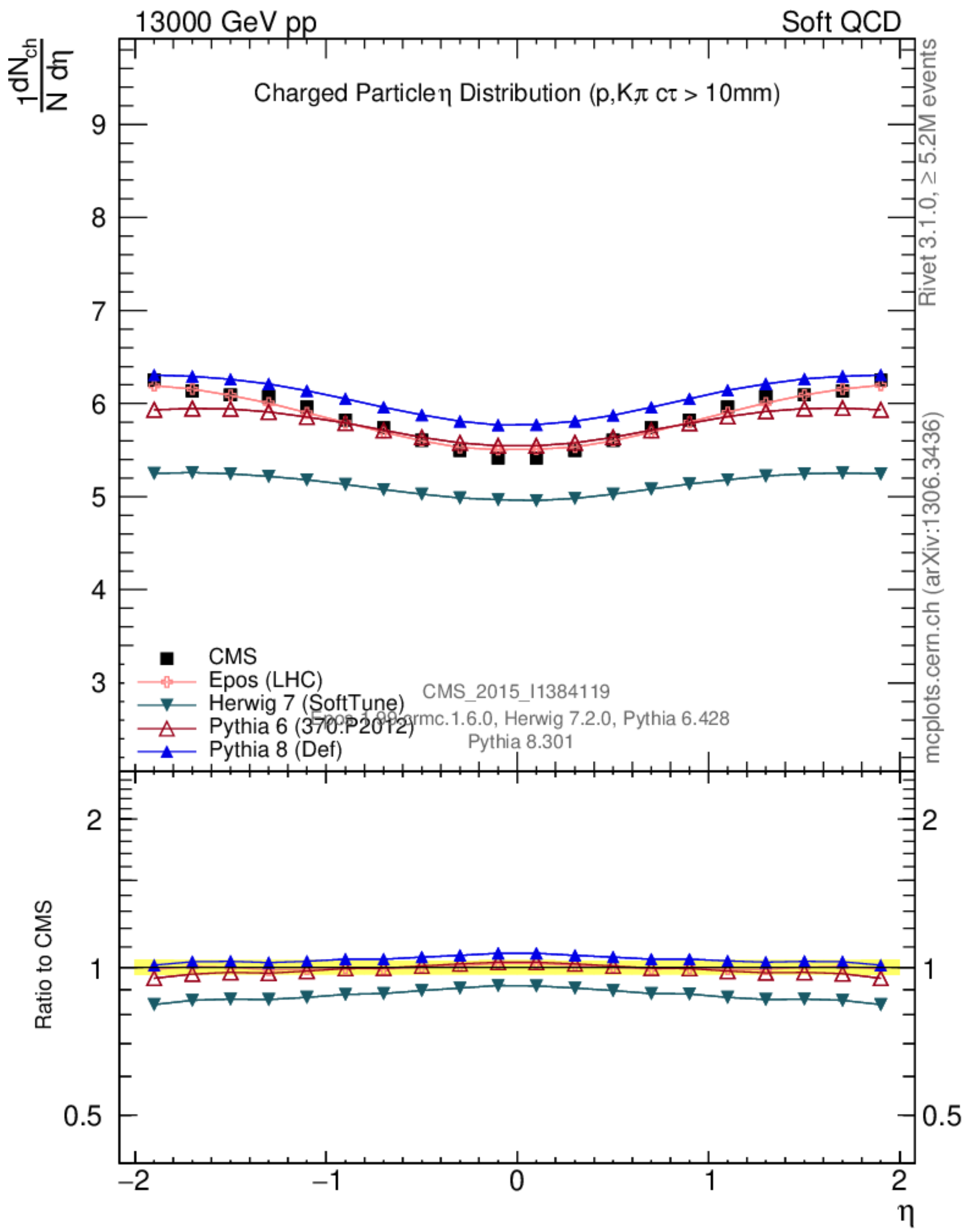
$\sqrt{s}$ scaling

Evolution of UE, $\langle \mathrm{d}N/\mathrm{d}\eta \rangle$, … with collider CM energy

Cast as energy evolution of p_{T0} parameter.

Bad Example: Why dN/dη is useless (by itself)

$\langle dN_{\mathrm{ch}}/d\eta \rangle$ often used as main constraint on models of minimum-bias physics



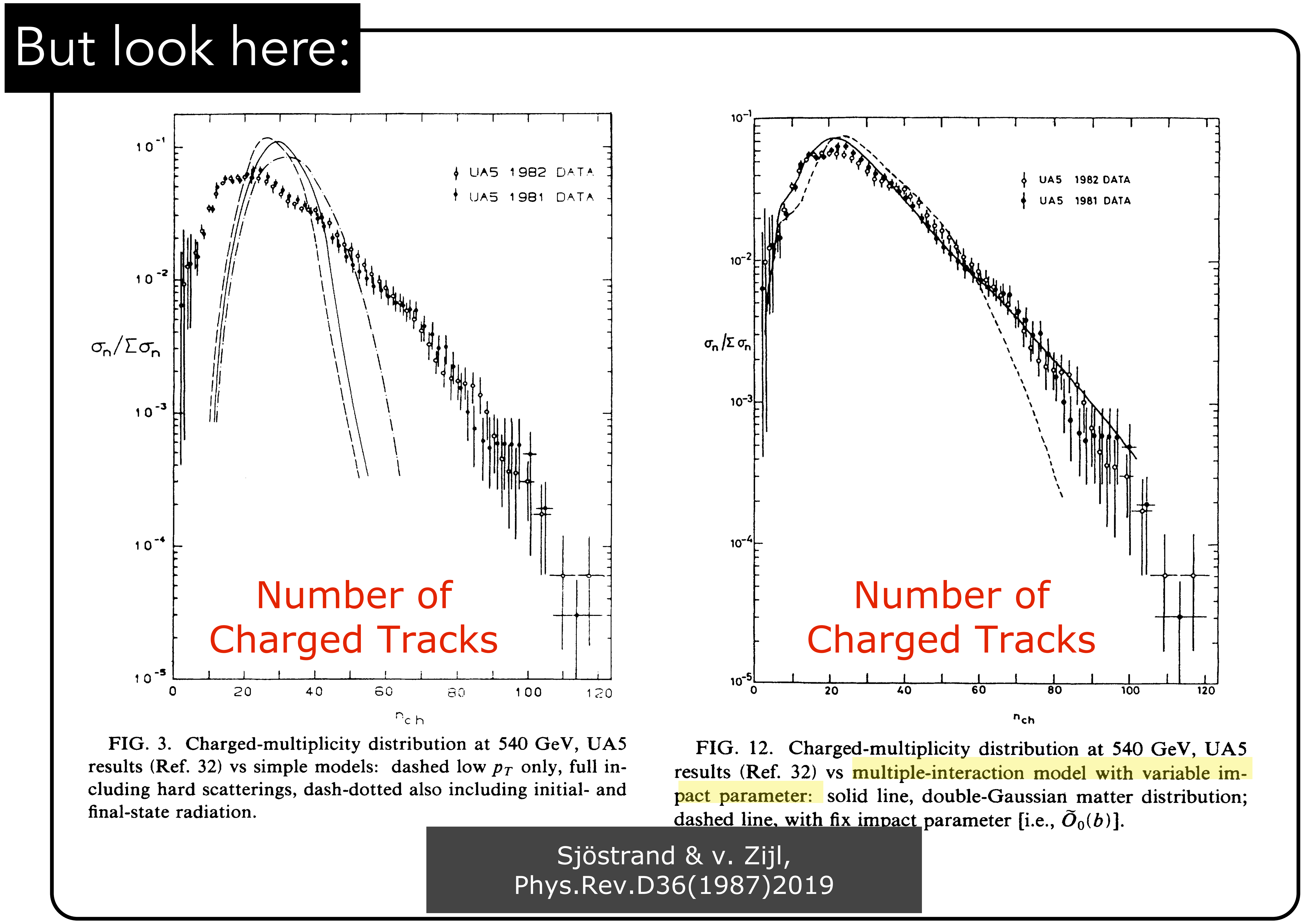
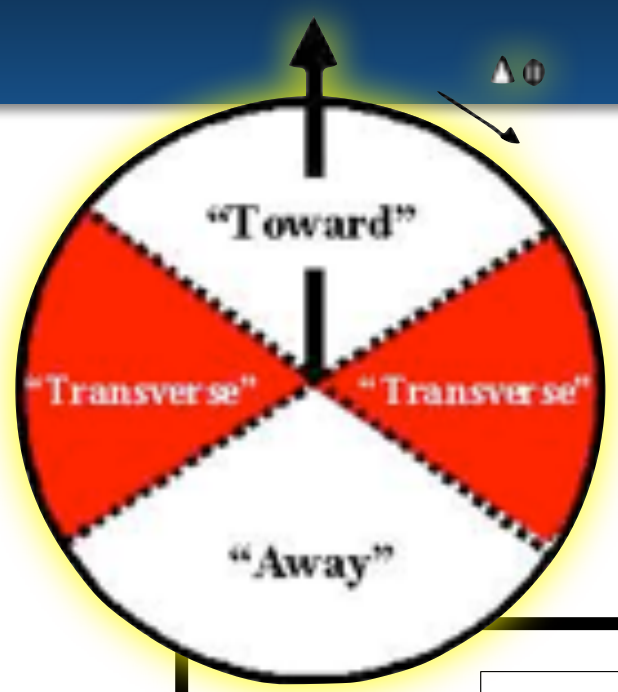


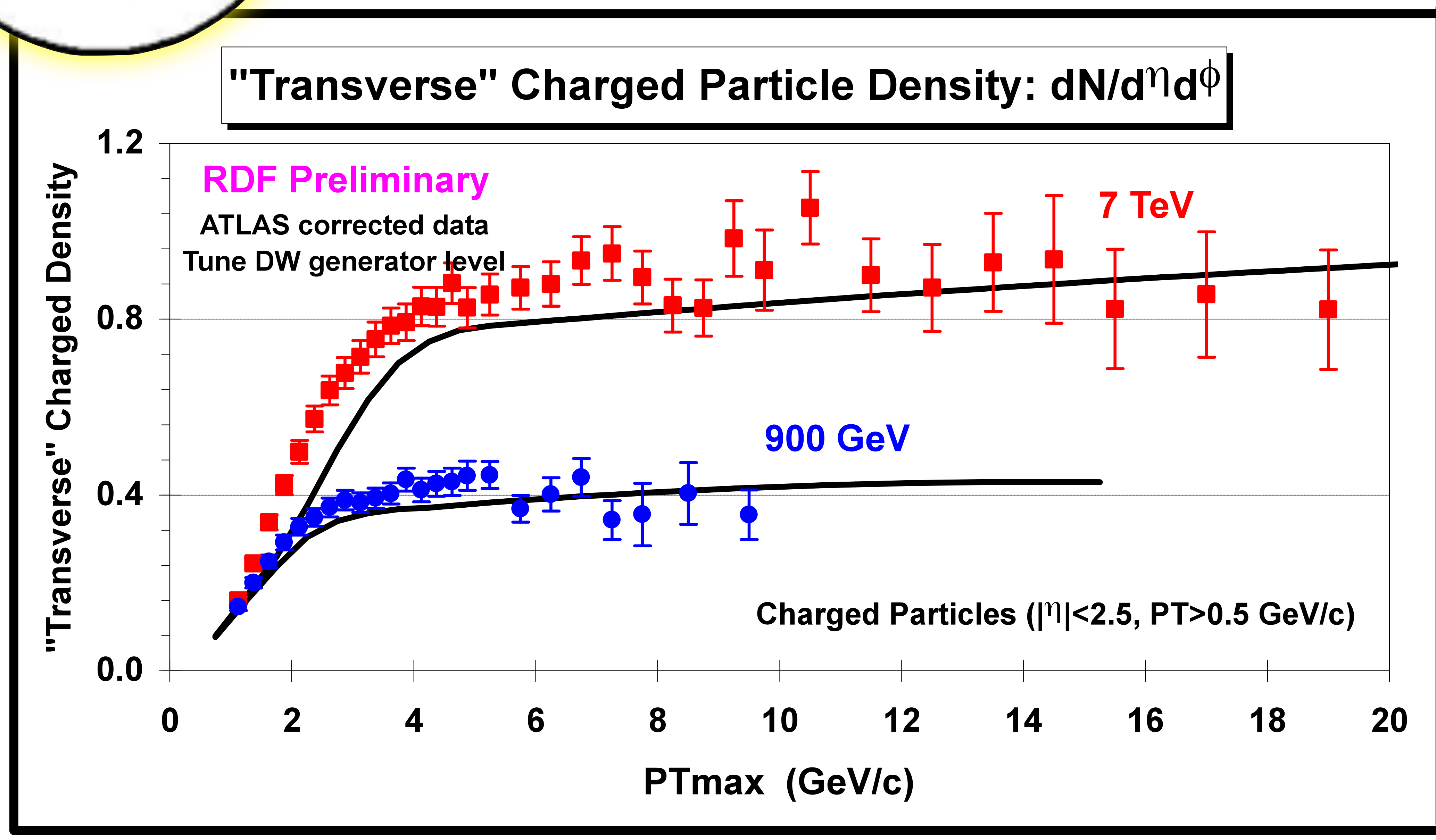
FIG. 3. Charged-multiplicity distribution at 540 GeV, UA5 results (Ref. 32) vs simple models: dashed low p_T only, full including hard scatterings, dash-dotted also including initial- and final-state radiation.

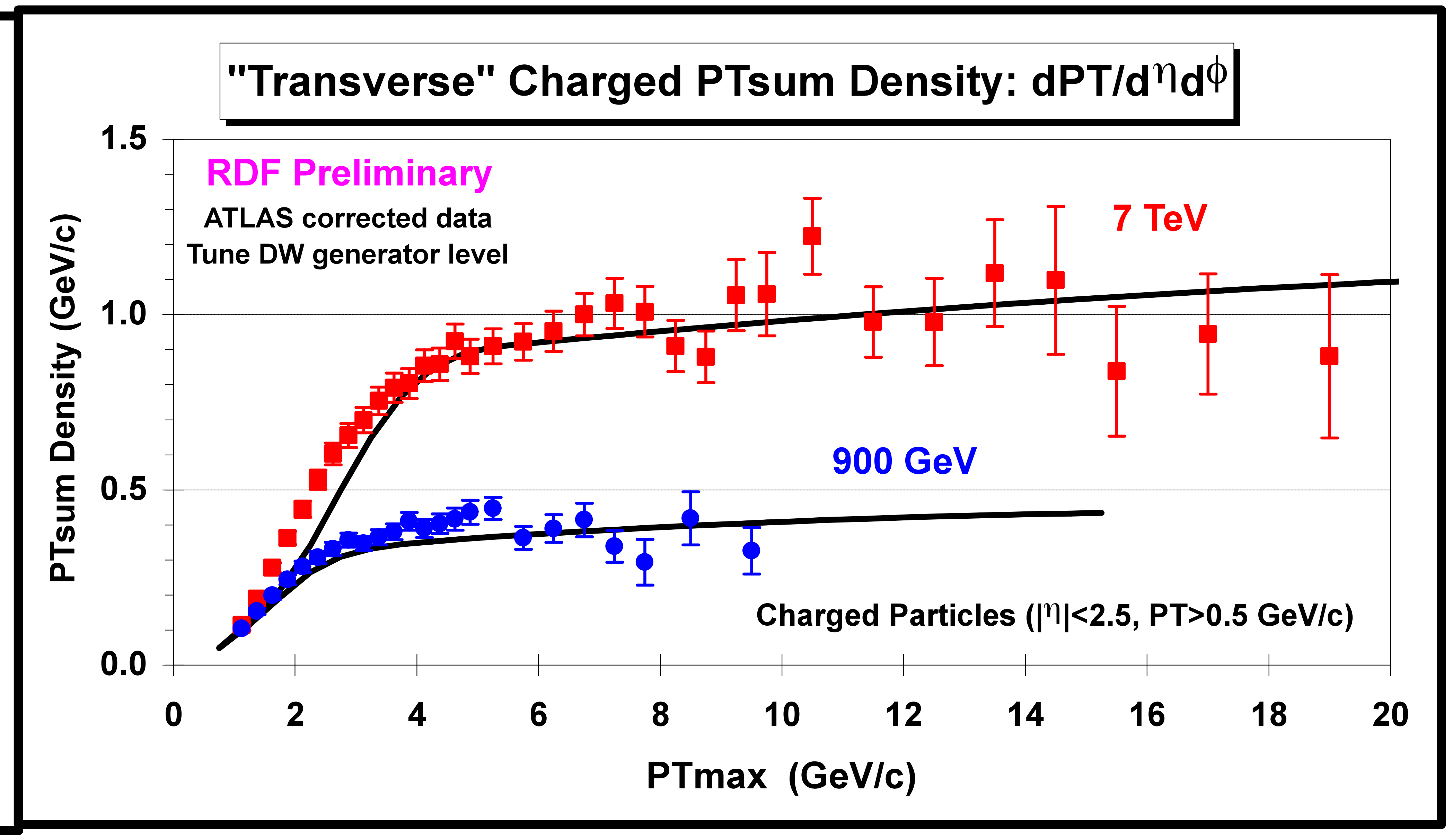
FIG. 12. Charged-multiplicity distribution at 540 GeV, UA5 results (Ref. 32) vs multiple-interaction model with variable impact parameter: solid line, double-Gaussian matter distribution; dashed line, with fix impact parameter [i.e., $\tilde{O}_0(b)$].

Can get right $\langle N_{\mathrm{ch}} \rangle$ with completely wrong models. (Need a few more moments at least.)



UE - LHC from 900 to 7000 GeV - ATLAS





As you trigger on progressively higher p_T, the entire event increases …

… until you reach a plateau ("max-bias") also called the "jet pedestal" effect

Interpreted as impact-parameter effect

Qualitatively reproduced by MPI models

Relative size of this plateau / min-bias depends on pT0, PDF, *and* b-profile

(More Specialised Parameters)

Hadron decay tables

Branching fractions and decay modelling

Collective Effects (in pp)

Colour Reconnections (& effects on precision measurements like m_{top})

Strangeness Enhancements (eg close-packing, ropes, …)

Flow-like effects (eg close-packing, string shoving, …)

Forward Physics

Beam-Remnant Handling

Diffractive Modelling (incl hard diffraction, Pomeron substructure)

Total and Elastic Cross-section parametrisatons

…

Parameter Hierarchies: An Example

Tuning: the higher up the chain you change something, the more it will affect the large-scale event structure ➜ Start at the top, and work your way down.

Divide and Conquer: Use Infrared Safety, **Exclusivity**, and **Ratios** to exploit factorisations!

3-jet events have a larger $\langle N_{\rm ch} \rangle$ than 2-jet events

So if you don't get the relative mixture of 2- to 3-jet events right, then you would be in unsafe territory trying to fit your **lower-scale** non-perturbative parameters to an inclusive measurement of $\langle N_{\rm ch} \rangle$.

What can you do? Adjust shower α_s, or use NNLO merging, or use reweighting, or use $\langle N_{\rm ch} \rangle$ in an **exclusive 2-jet sample** that does not depend on the relative 2-to-3-jet ratio. **But don't do nothing.**

Similarly, the total number of particles is different

But **relative ratios** like $\langle N_{\rm K} \rangle / \langle N_{\pi} \rangle$ should be more universal

Wouldn't it be nice if there was a tool:

That could automatically detect correlations between parameters and observables.

And tell you which "groups" they fall into naturally : which parameter sets you should ideally tune together, and which are more nicely factorised.

This is (at least partly) what the tool AutoTunes does

I won't have time to discuss that today, but I think it looks promising

I encourage you to study it and use it: Bellm, Gellersen, Eur.Phys.J.C 80 (2020)

You may also be interested in Apprentice Krishnamoorthy et al., EPJ Web Conf. 251 (2021) 03060

Variance reduction to semi-automate how to weight observables & bins

Systematic Tests of Universality

Systematic Universality Tests + characterisation of any deviations.

Do independent tunes for different CM energies find universal parameters?

Do independent tunes for different processes find universal parameters?

Do independent tunes for different experiments find universal parameters?

Do independent tunes for different obervables find universal parameters?

I experimented a bit with that so far only in specific contexts, but I would say good experiences, increasing faith in robustness and universality

E.g., arXiv:1103.3649 tested MB universality across different CM energies; found good universality except for CR strength. Further explored in arXiv:1808.07224.

arXiv:1812.07424 tuned independently to ALEPH, DELPHI, OPAL, L3, with/without event shapes ➜ rejected a few extreme "outliers" which were inconsistent with bulk of tunes, defined envelope of uncertainties from rest.

Another example: FSR in $t\bar{t}$ at LHC prefers lower $\alpha_s(M_Z)$ than FSR in Z decays (presumably due to non-universal ME corrections and/or coherence issues.)

Reliable Uncertainties and Preventing Overfitting

Monash Tune: 5% flat sanity-limit Theory Uncertainty to prevent overfitting

Can this be improved on? Using better theory uncertainty estimates? & sensitivities?

Would like TH uncertainties to get to ~ $\chi^2_{\rm red}$ ~ 1. Not well-defined across multiple distributions with unknown correlations.

(Monash Tune was done by eye, so this was simply a matter of judgement.)

Use Pythia to map correlations between observables and incorporate in tuning?

Professor's eigentunes may be prone to artifacts of overtuning

E.g., **well-measured peak will dominate**, with arbitrarily tiny uncertainties, at price of not spanning range in tails/asymptotics. Unclear interplay with genuine theory uncertainties.

See eg arXiv:1812.07424 for examples (and slightly more elaborate way to address issue but still fundamentally based on the flat 5% sanity limit)

There is still a need to develop reliable well-motivated uncertainty variations

Beyond "eigentunes" (Perugia had simple ones, Monash had none)

Ideally also propose ***method*** for how to obtain them, and justify or improve on the 5% approach.

Resources

Data Preservation: HEPDATA

Online database of experimental measurement results

Please make sure all published results make it there

Analysis Preservation: RIVET

Large library of encoded analyses + data comparisons

Main analysis & constraint package for event generators

All your analysis are belong to RIVET

Updated validation plots: MCPLOTS.CERN.CH

Online plots made from Rivet analyses

Want to help? **Connect to LHC@home project Test4Theory**

Reproducible tuning: PROFESSOR, AUTOTUNES, APPRENTICE (& more?)

Automated tuning (& more)

mcplots.cern.ch

Menu

→ Front Page
→ LHC@home 2.0
→ Generator Versions
→ Generator Validation
→ Update History
→ User Manual and Reference

Analysis filter:

→ ALL pp/ppbar
→ ALL ee
Specific analysis:
→
→ Latest analyses

Z (Drell-Yan)

→ Jet Multiplicities
→ $1/\sigma d\sigma(Z)/d\varphi^*_\eta$
→ dσ(Z)/dpTZ
→ 1/σdσ(Z)/dpTZ

W

→ Charge asymmetry vs η
→ Charge asymmetry vs N_{jet}
→ dσ(jet)/dpT
→ Jet Multiplicities

Top (MC only)

→ Δφ (ttbar)
→ Δy (ttbar)
→ |Δy| (ttbar)
→ M (ttbar)
→ pT (ttbar)
→ Cross sections
→ y (ttbar)
→ Asymmetry
→ Individual tops

Bottom

→ η Distributions
→ pT Distributions
→ Cross sections

Jets

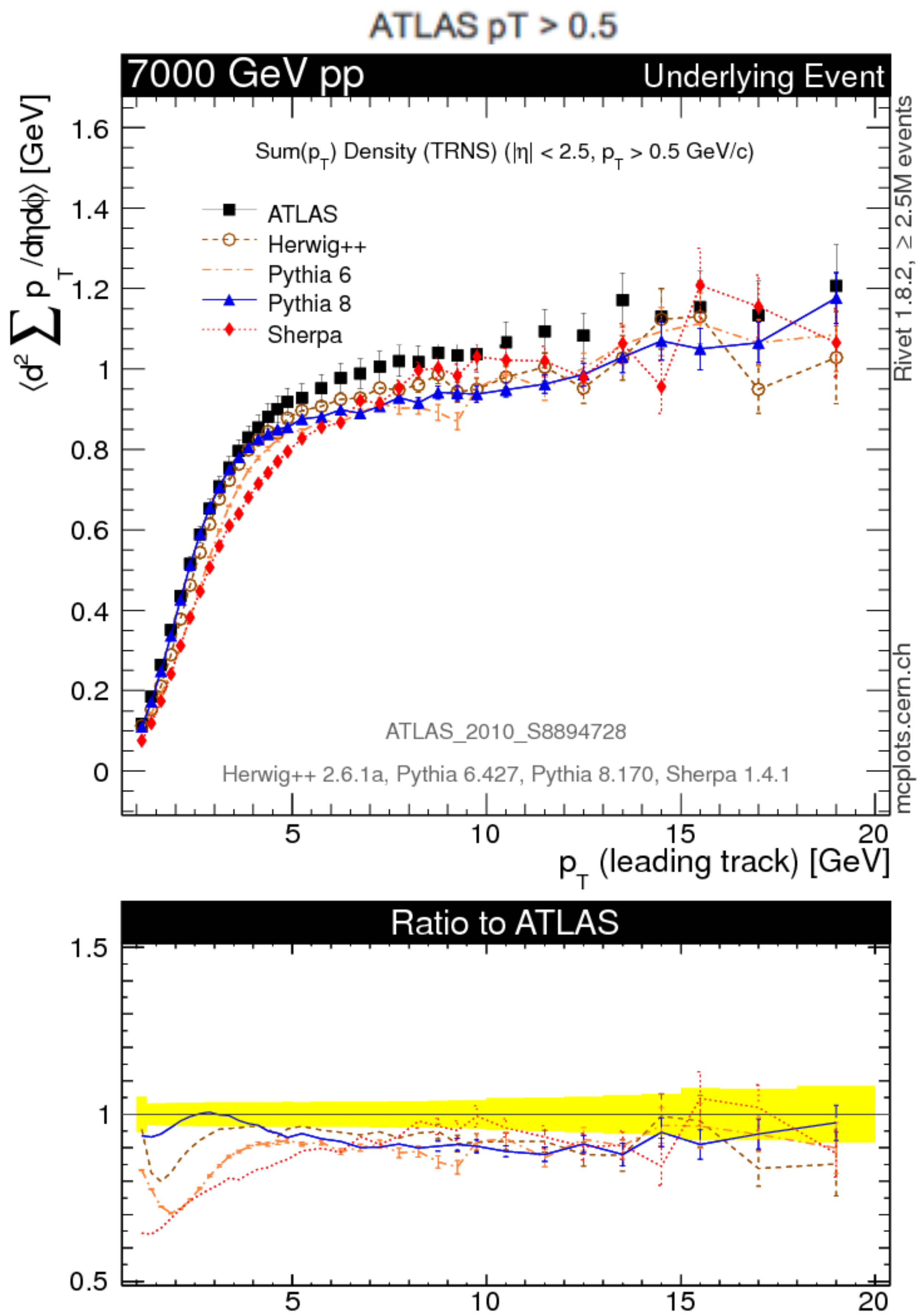
Underlying Event : TRNS : Σ(pT) vs pT1

Generator Group: General-Purpose MCs Soft-Inclusive MCs Alpgen Herwig++ Pythia 6 Pythia 8 Sherpa Vincia Epos Phojet Custom

Subgroup: Defaults LHC Tunes C++ Generators Tevatron vs LHC tunes

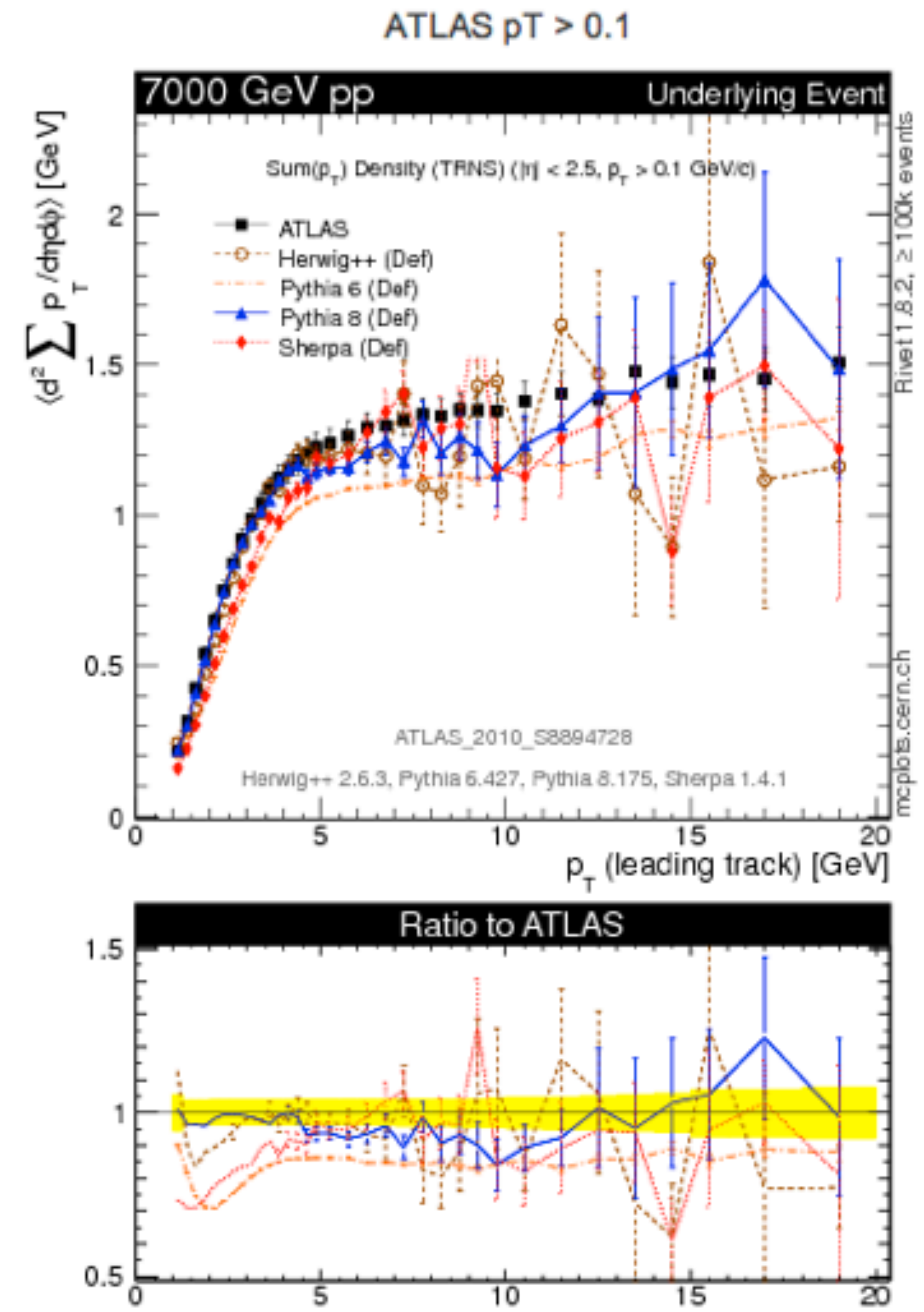
pp @ 7000 GeV

ATLAS pT > 0.5



[pdf] [eps] [png] hide details ←
[ATLAS] reference
[Herwig++ (Def)] param
[Pythia 6 (Def)] param
[Pythia 8 (Def)] param
[Sherpa (Def)] param
[steer]

ATLAS pT > 0.1

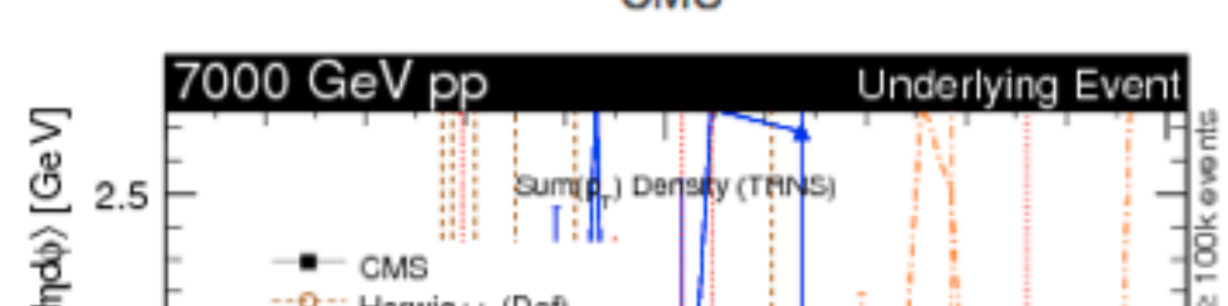


[pdf] [eps] [png] show details →

A. Karneyeu et al., Eur.Phys.J. C74 (2014) 1

- Explicit tables of data & MC points
- Run cards for each generator
- Link to experimental reference paper
- Steering file for plotting program
- (Will also add link to RIVET analysis)

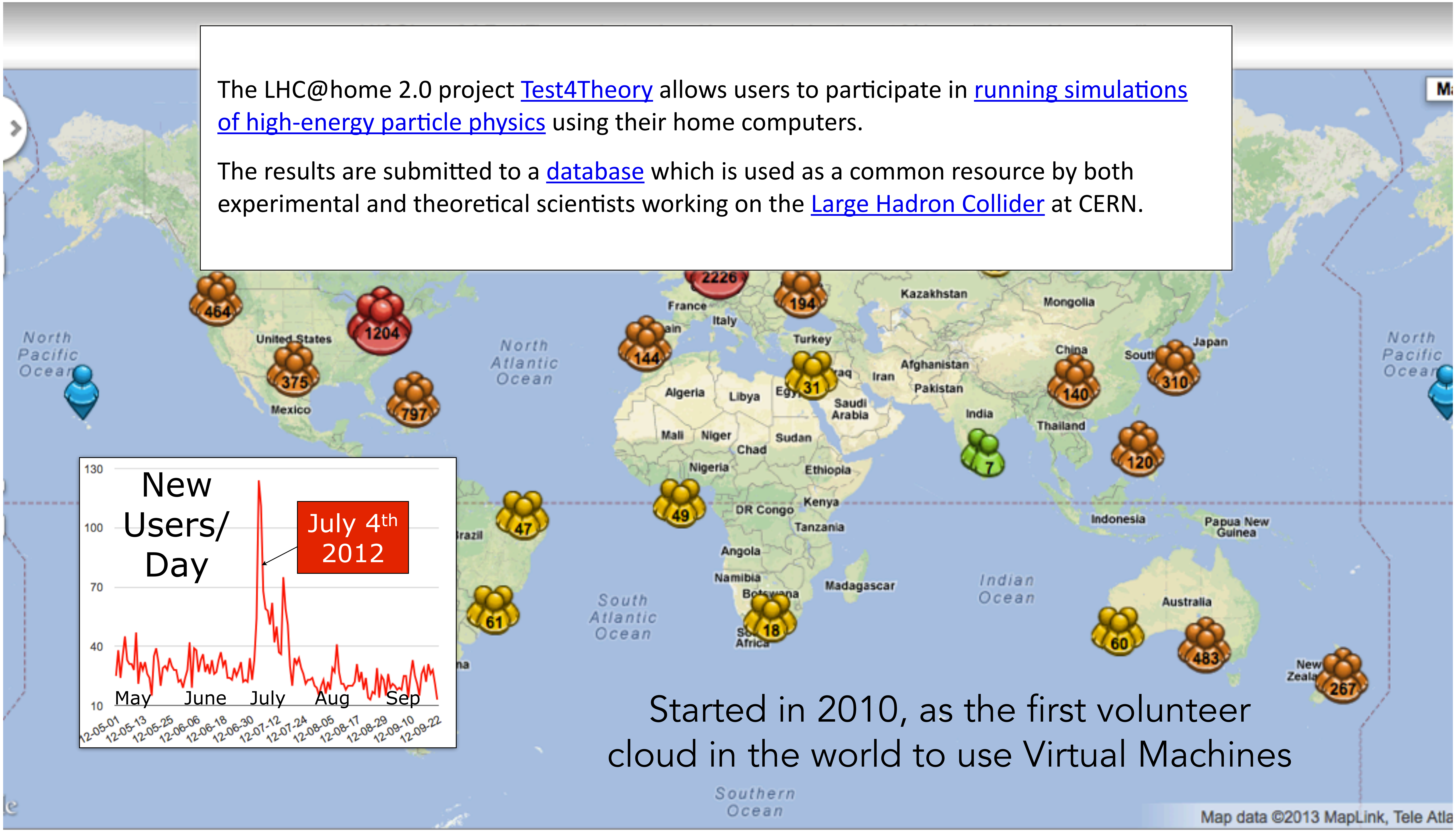
CMS



Join us at LHC@home Test4Theory

The LHC@home 2.0 project Test4Theory allows users to participate in running simulations of high-energy particle physics using their home computers.

The results are submitted to a database which is used as a common resource by both experimental and theoretical scientists working on the Large Hadron Collider at CERN.



Started in 2010, as the first volunteer cloud in the world to use Virtual Machines

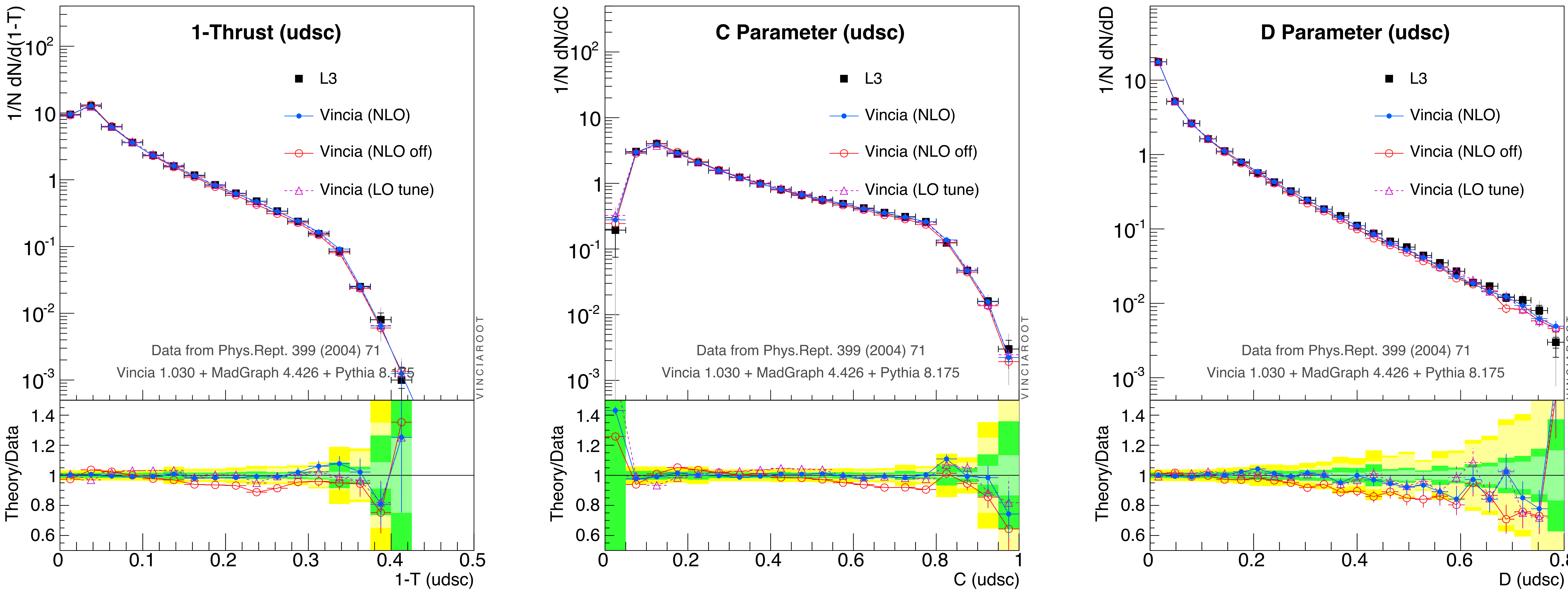
Backup Slides

First LEP tune with NLO 3-jet corrections

Hartgring, Laenen, Skands, arXiv:1303.4974

LO tune: $\alpha_s(M_Z) = 0.139$ (1-loop running, MSbar)

NLO tune: $\alpha_s(M_Z) = 0.122$ (2-loop running, CMW)



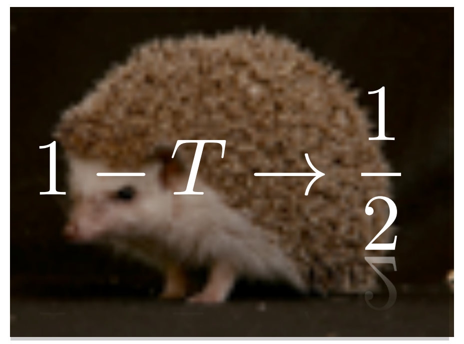
Need IR Corrections?

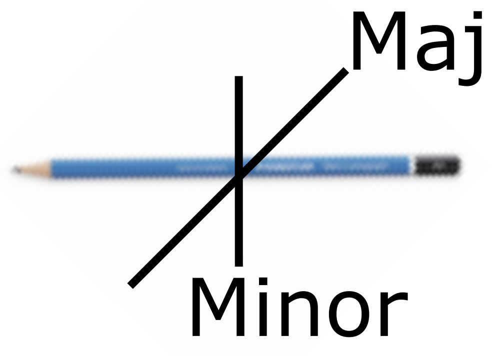
PYTHIA 8 (hadronization off)

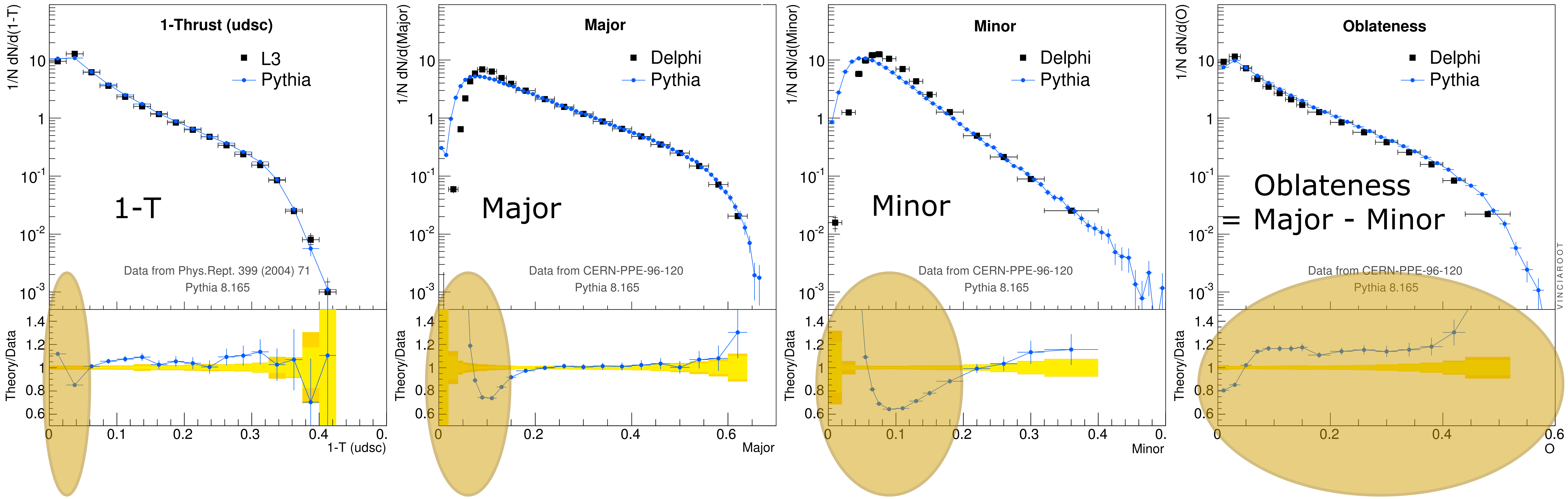
vs LEP: Thrust

$$T = \max_{\vec{n}} \left(\frac{\sum_i |\vec{p}_i \cdot \vec{n}|}{\sum_i |\vec{p}_i|} \right)$$

$1 - T \to 0$







Significant Discrepancies (>10%)

for T < 0.05, Major < 0.15, Minor < 0.2, and for all values of Oblateness

+ cross checks: different eCM energies (HAD and FSR scale differently)

Standalone LO/LL vs Merging

One for LO applications, starting from best fit standalone.

Introduce LO merging as cross check on universality, ensuring good all-round performance for LO applications with/without MECs and merging.

Another with highest achievable level of NLO merging

Need NLO merging for all tuning samples.

Not totally clear if this is realistically doable.

+ Eg merging in e^+e^- not well developed.

Could presumably have $\alpha_s(M_Z) \sim 0.12$ while maintaining a good fit.

Subtlety: interplay between α_s values in shower and in ME.

Or … could they be one and the same?

Happiest if hadronisation parameters were universal

Possible to settle on a **single choice** of non-perturbative parameters that would give good fits both **with and without** (N)LO merging?

True for many hadronisation parameters (eg strangeness fractions)

Also eg for MPI: p_{T0} mainly depends on PDF; would use same for MPI here.

Main differences are # of hard jets and IR limit of shower (Q_{cut} and α_s)

Could address # of hard jets by **reweighting** event samples?

Choose α_s : eg 1-loop for LO, 2-loop for NLO, with **similar** Λ_{QCD}

+ can experiment with smooth dampening (similar to MPI) to make behaviour near cutoff less extreme? (Done in Vincia.)

➜ Could operate with lower cutoffs (though we do still want an absolute cutoff, with $O(\Lambda)$ crinkles absorbed in string).

Possible to get ~ universality by allowing Q_{cut} to float a bit?

And/or carefully ensure IR limits near cutoff are ~ same.

➜ Universal hadronisation tuning?

Universal hadronisation tuning?

Independent of perturbative order (as discussed) would be a major step

Would require some dedicated thought. Physics of universality (shower behaviour near boundary) and mathematical formulation.

Reweighting techniques to bring LO and NLO jet rates into agreement ➜ similar initial conditions for HAD; needed to tackle the many constraints which are sensitive to a mixture of high and low scales.

+ Propose observables (eg hadronisation in exclusive 2-jet events) less sensitive to high-scale corrections?

Universality of MPI under PDF swapping?

Let the reference value of pT0 be a derived parameter, from a given <nMPI> ~ sigmaQCD(pT0)/sigmaINEL, so that the UE level is more stable against the sometimes huge changes in the low-x gluon.

Ilkka emphasised that NLO evolution is faster, so probably want to do something similar with the energy scaling, eg by looking at <nMPI> at two different ECM values.