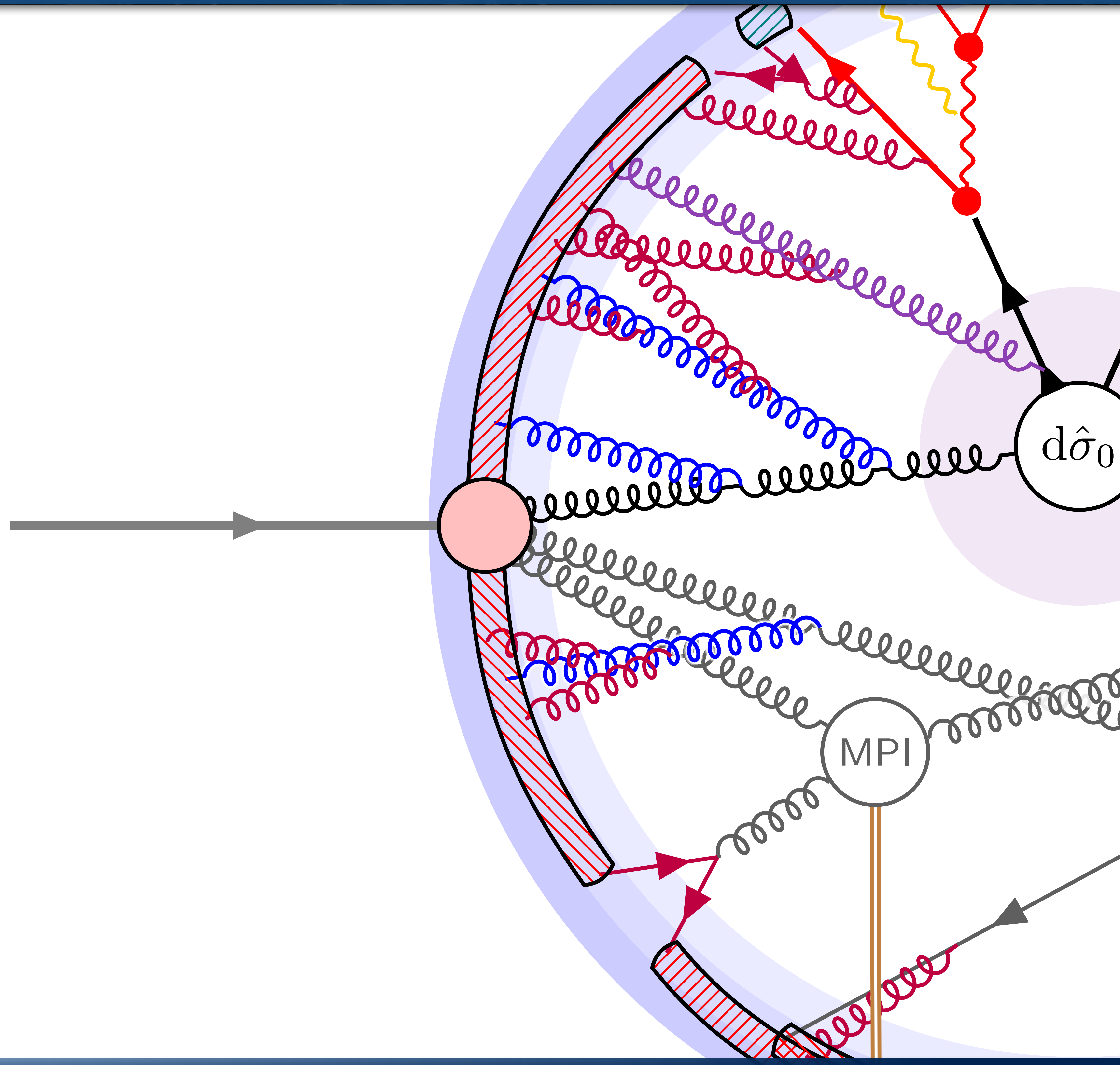
Uncertainties in Monte Carlo Event Generators

(with emphasis on Pythia 8)

1. **Perturbative Uncertainties**
2. **Hadronization Uncertainties**
3. **Tuning**
4. **Discussion …**



Peter Skands
U of Oxford & Monash U

CMS Deep Dive Into Modelling Uncertainties
September 2023

High-Precision Measurements ↔ Rigorous & Exhaustive Uncertainties

Brute Force

- Separate runs for each variation
- Construct & perform all salient variations individually
- **Expensive**
- CPU ↔ Cost
- Environmental impact
- (Duplication of) man-hours, each time
- Risk of mistakes/inconsistencies, each time
- Risk that lessons learned aren't perpetuated, each time

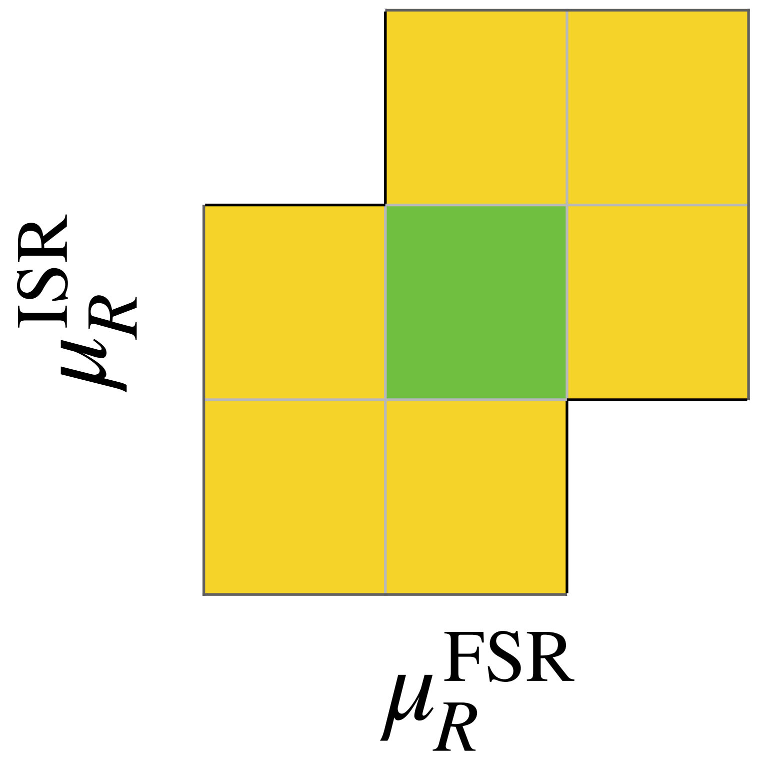
Sophisticated reweighting methods developed for Parton Showers

- Based on reinterpreting the veto algorithm's accept and reject probabilities
- [**Vincia** 1102.2126; **Sherpa** 1605.04692; **Herwig** 1605.08256; **Pythia** 1605.08352]
  (Note: reweighting of course also done for PDFs and in Fixed-Order Calculations.)

Perturbative Uncertainties

First guess: renormalisation-scale variations,

- $\mu_R^2 \to k_\mu \, \mu_R^2$, with constant $k_\mu \in [0.5, 2]$ or $[0.25, 4]$, …
  + e.g., do for ISR and FSR separately → 7-point variations*
  *See backup slides
- Induces explicit "nuisance" terms beyond controlled orders



I think most people I know actually consider this unsatisfactory and unreliable

- Problem is, little guidance on what else to do …

Big Problem 1: Multiscale Problems (e.g., a couple of bosons + a couple of jets)

- Not well captured by any variation k_μ around any **single** scale
- More of an issue for fixed-order calculations than for showers (which are intrinsically multiscale)

Big Problem 2: Terms that are not proportional to the lower orders

- **Renormalization-scale** variations $\Longrightarrow \mathrm{d}\sigma \to (1 + \Delta\alpha_s)\,\mathrm{d}\sigma$
- But in general there will also be **genuinely new terms** at each order, $\mathrm{d}\sigma \to \mathrm{d}\sigma \pm \Delta\mathrm{d}\sigma$

Vincia & Pythia 8: Finite-Term Variations

Parton Showers rely on Factorisations in Soft/Collinear Limits

$$|M_{n+1}|^2 \to \sum_{\text{radiators}} a_{\text{sing}} |M_n|^2$$

- Approximations based on universal (process-independent) singular structures of gauge theories.
- Driven by $1/Q^2$ poles from propagators, with spin-dependent numerators
- Renormalization-scale variations only produce terms proportional to these "kernels"

But genuine matrix elements also have "non-singular terms"

- Our solution [**Vincia** 1102.2126; **Pythia** 1605.08352]

Non-singular variations

$$a_{\text{sing}} \to a_{\text{sing}} + \Delta a_{\text{non-sing}}$$

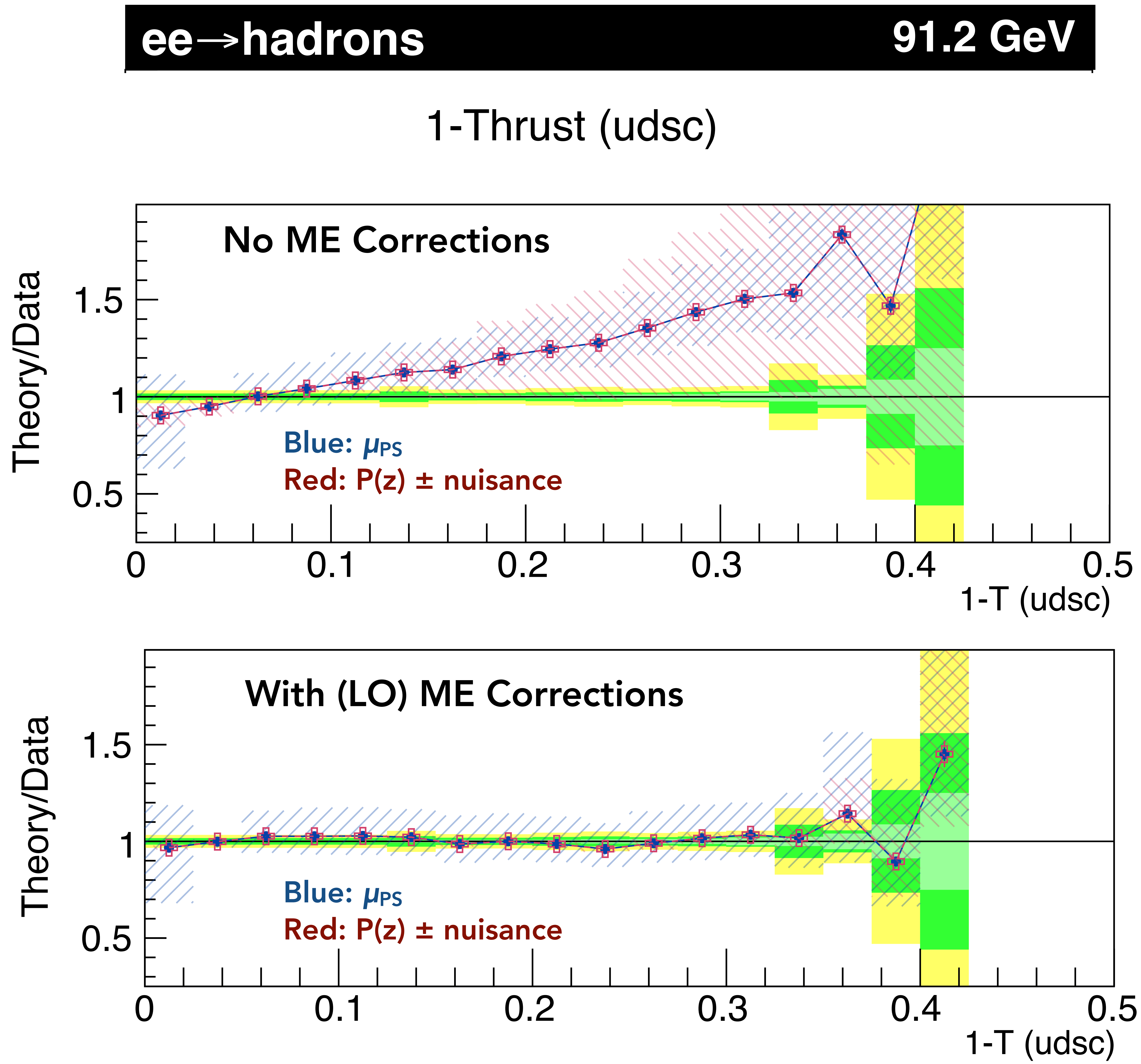
- Can also be very helpful to estimate need for higher matching/merging

Non-Singular Variations

- Add arbitrary nonsingular term to shower kernels, and **vary** to estimate sensitivity to missing ME terms
- (Reasonable size estimated by comparisons between different actual MEs)

  [**Vincia** 1102.2126; **Pythia** 1605.08352]

- The shower singularities dominate for soft and collinear radiation
- The process-specific non-singular terms dominate for hard radiation



Note: by definition, any fit of such a nuisance parameter would be process-specific

2. Hadronization Uncertainties

Hadronization: More parameters, many subtleties (ideally a coffee discussion…)

- Risk of purely data-driven methods (eg eigentunes) to overfit precise data at expense of tails / asymptotics / less statistically dominant (but perhaps theoretically important) data
- Risk of inconsistencies (breakdown of universality and/or inconsistent levels of accuracy and "tricks") between tuning context (eg LEP) and application context (eg LHC)
- Tensions between different measurements
- Interplay between perturbative (eg N_{Jets}) and nonperturbative (eg N_{Hadrons}) **observables**
- And between perturbative (α_S, merging, …) and nonperturbative (eg HAD and MPI, …) **pars**

  Parameter correlations; for a helping hand, see eg AutoTunes [Bellm & Gellersen, 1908.10811]
- Tuning, at precision level, is a challenging and very complex field.

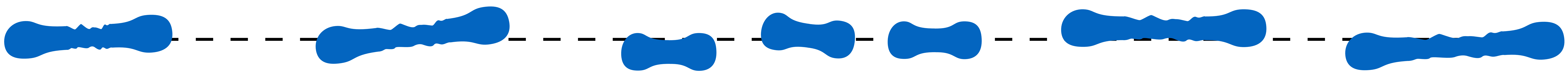
Recent elaborate studies with Pythia 8:

- Not addressing *all* of the above. Some steps/suggestions towards more systematic approaches, though by no means the final word:
- [Jueid et al., 1812.07424; 2202.11546; 2303.11363]

New: Automated Hadronization Uncertainties

Problem:

- Given a colour-singlet system that (randomly) broke up into a specific set of hadrons:



- What is the relative probability that same system would have resulted, if the fragmentation parameters had been somewhat different?
- Would this particular final state become **more likely** $(w' > 1)$? Or **less likely** $(w' < 1)$
- Crucially: **maintaining unitarity** $\Longrightarrow$ inclusive cross section remains unchanged!

Aug 25: Bierlich, Ilten, Menzo, Mrenna, Szewc, Wilkinson, Youssef, Zupan

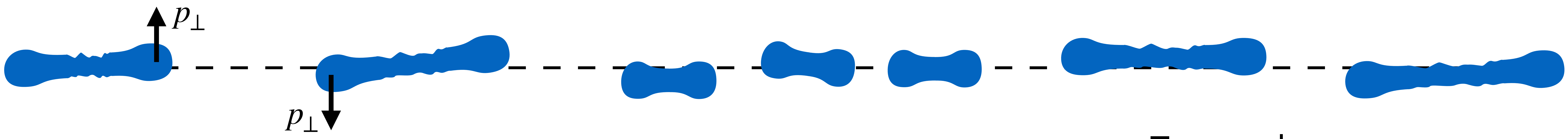
[*Reweighting MC Predictions & Automated Fragmentation Variations in Pythia 8*, 2308.13459]

Method is general; demonstrated on variations of the 7 main parameters governing longitudinal and transverse fragmentation functions in PYTHIA 8

https://gitlab.com/uchep/mlhad-weights-validation

[*Reweighting MC Predictions & Automated Fragmentation Variations in Pythia 8*, 2308.13459]

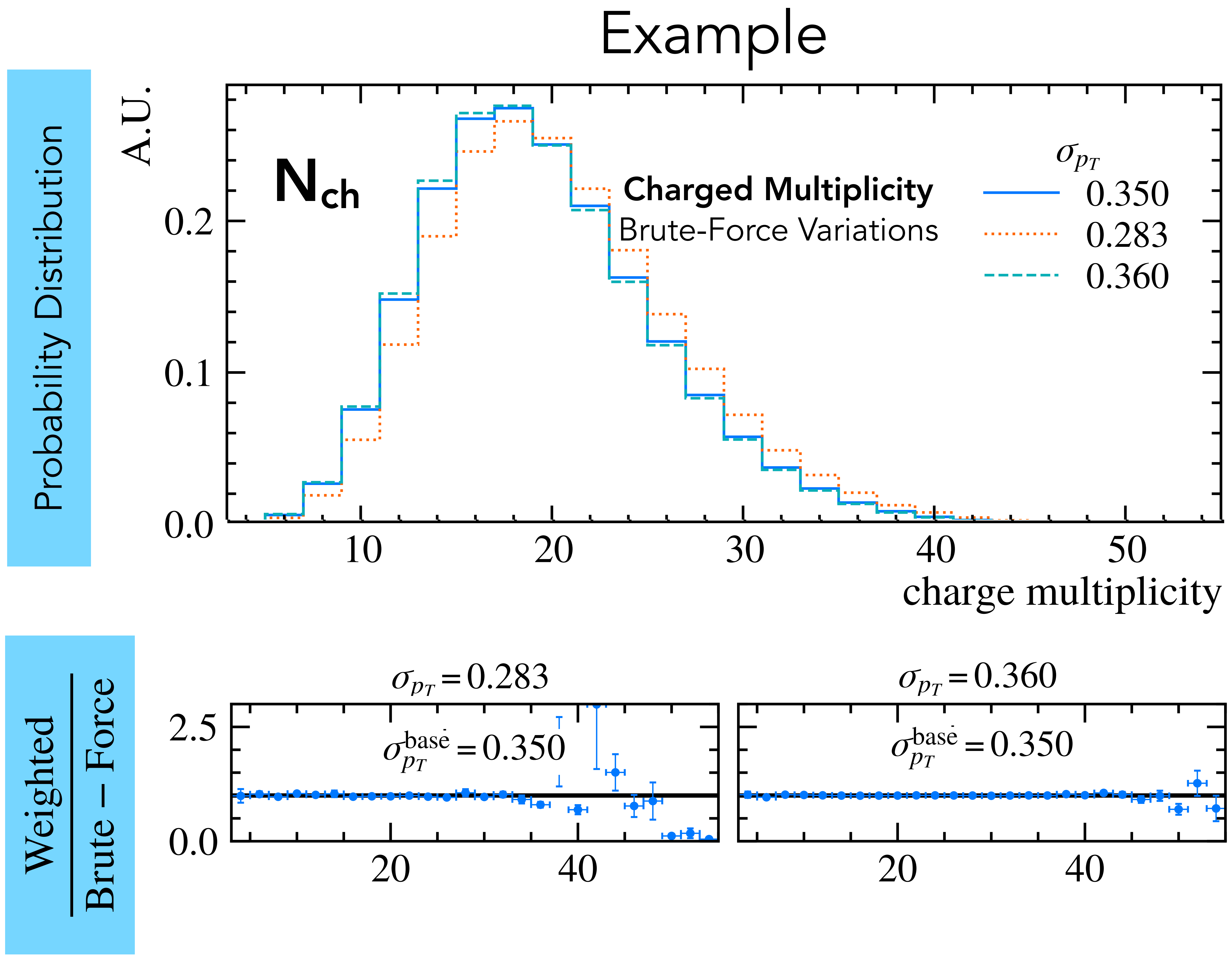
Transverse Fragmentation Function (Gaussian)



$$\frac{1}{2\pi\sigma_{p_T}^2}\exp\left(-\frac{(\Delta p_x)^2+(\Delta p_y)^2}{2\sigma_{p_T}^2}\right)$$

`StringPT:sigma`

variations



Reweighting Methodology:

For each p_T (Box-Muller transform):

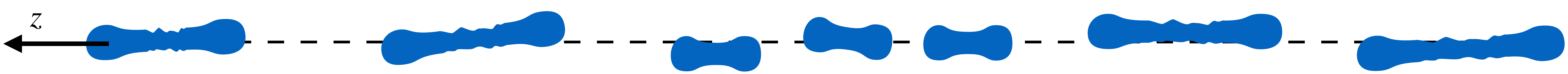
$$w' = \frac{\sigma^2}{\sigma'^2}\exp\left(-\kappa\left(\frac{\sigma^2}{\sigma'^2}-1\right)\right)$$

$\kappa = (n_1^2 + n_2^2)/2$ and n_i are normally distributed random variates

Examples with Pythia 8

[*Reweighting MC Predictions & Automated Fragmentation Variations in Pythia 8*, 2308.13459]

Longitudinal FF (Lund Symmetric FF)

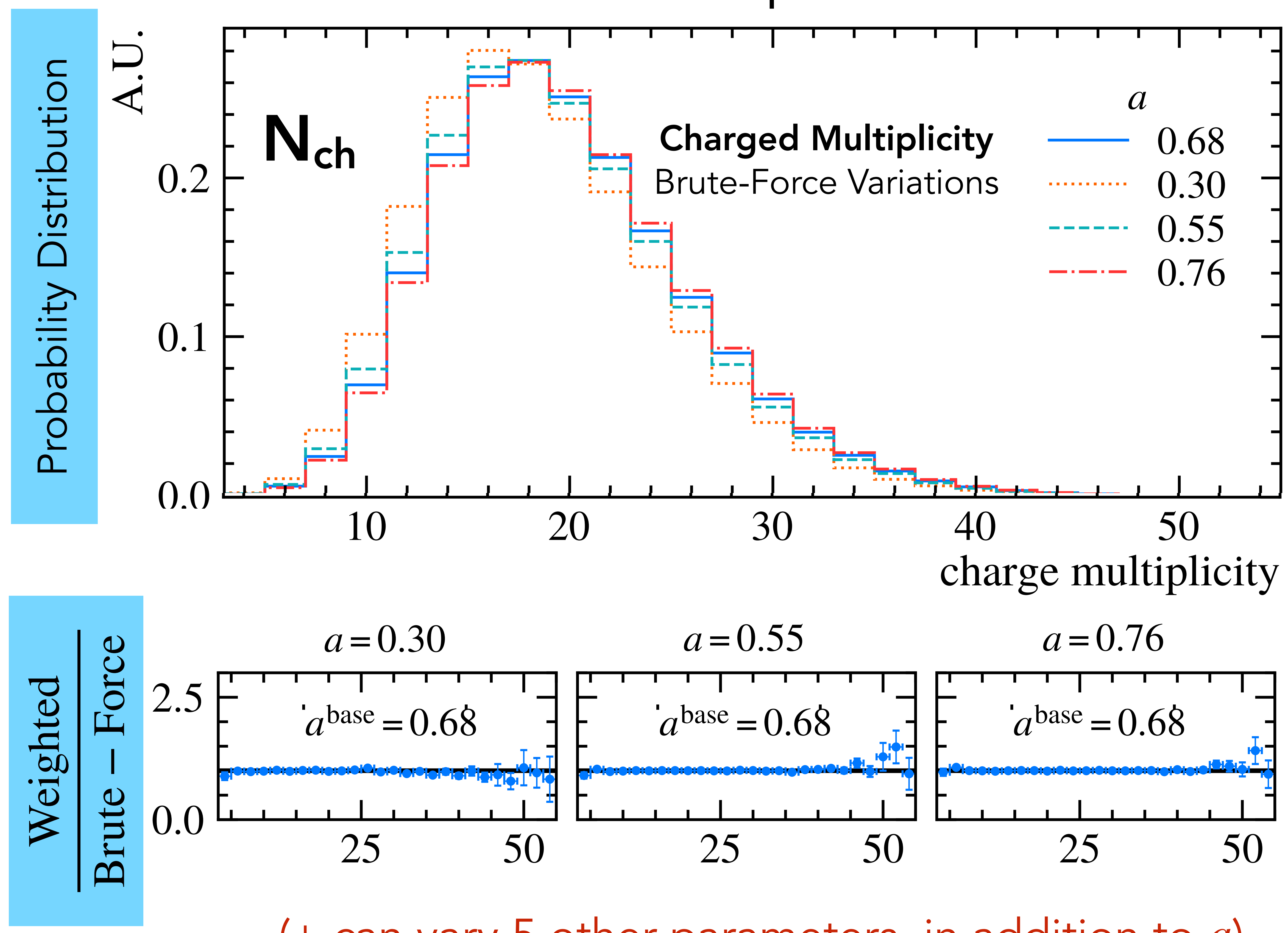


f(z) ~ *scaled* light-cone hadron momentum fraction

$$\propto \frac{1}{z^{1+r_Q b m_Q^2}} (1-z)^a \exp\left(-\frac{b m_\perp^2}{z}\right)$$

variations

Example



(+ can vary 5 other parameters, in addition to a)

Reweighting Methodology:

Accept-Reject Algorithm (analogous to shower variations):

$$w' = w \prod_{i \in \text{accepted}} R'_{i,\text{accept}}(z) \prod_{j \in \text{rejected}} R'_{j,\text{reject}}(z),$$

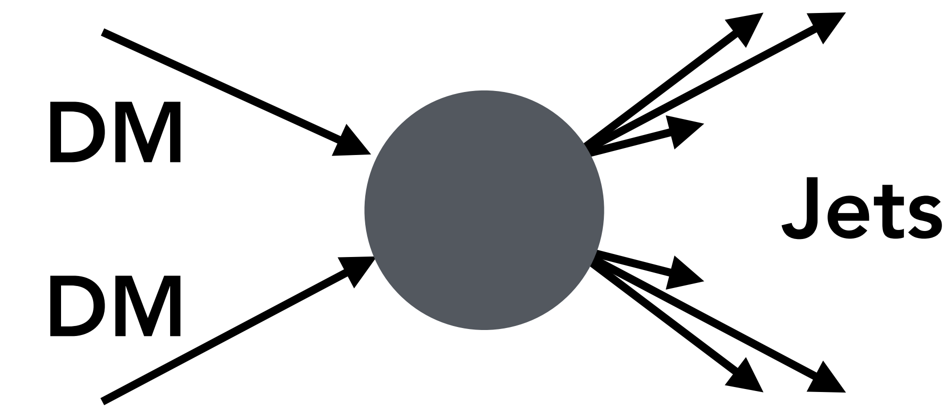
with

$$R'_{\text{accept}}(z) = \frac{P'_{\text{accept}}(z)}{P_{\text{accept}}(z)} \qquad R'_{\text{reject}}(z) = \frac{P'_{\text{reject}}(z)}{P_{\text{reject}}(z)} = \frac{1-P'_{\text{accept}}(z)}{1-P_{\text{accept}}(z)}$$

Example: The Strong Force Meets the Dark Sector

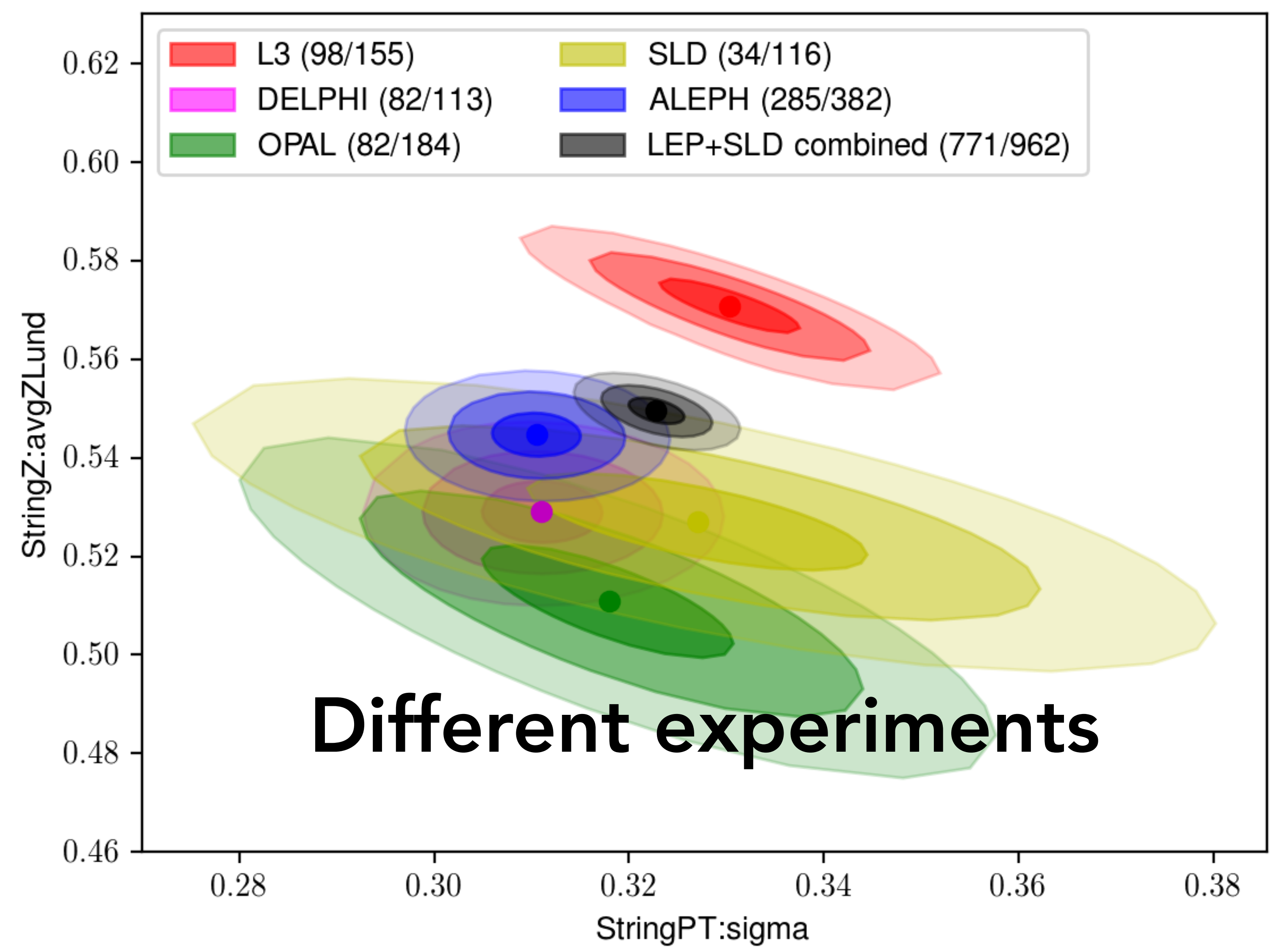
Based on A. Jueid et al., 1812.07424 (gamma rays, *eg for GCE*) and 2202.11546 (antiprotons, *eg for AMS*) + **2303.11363 (all)**

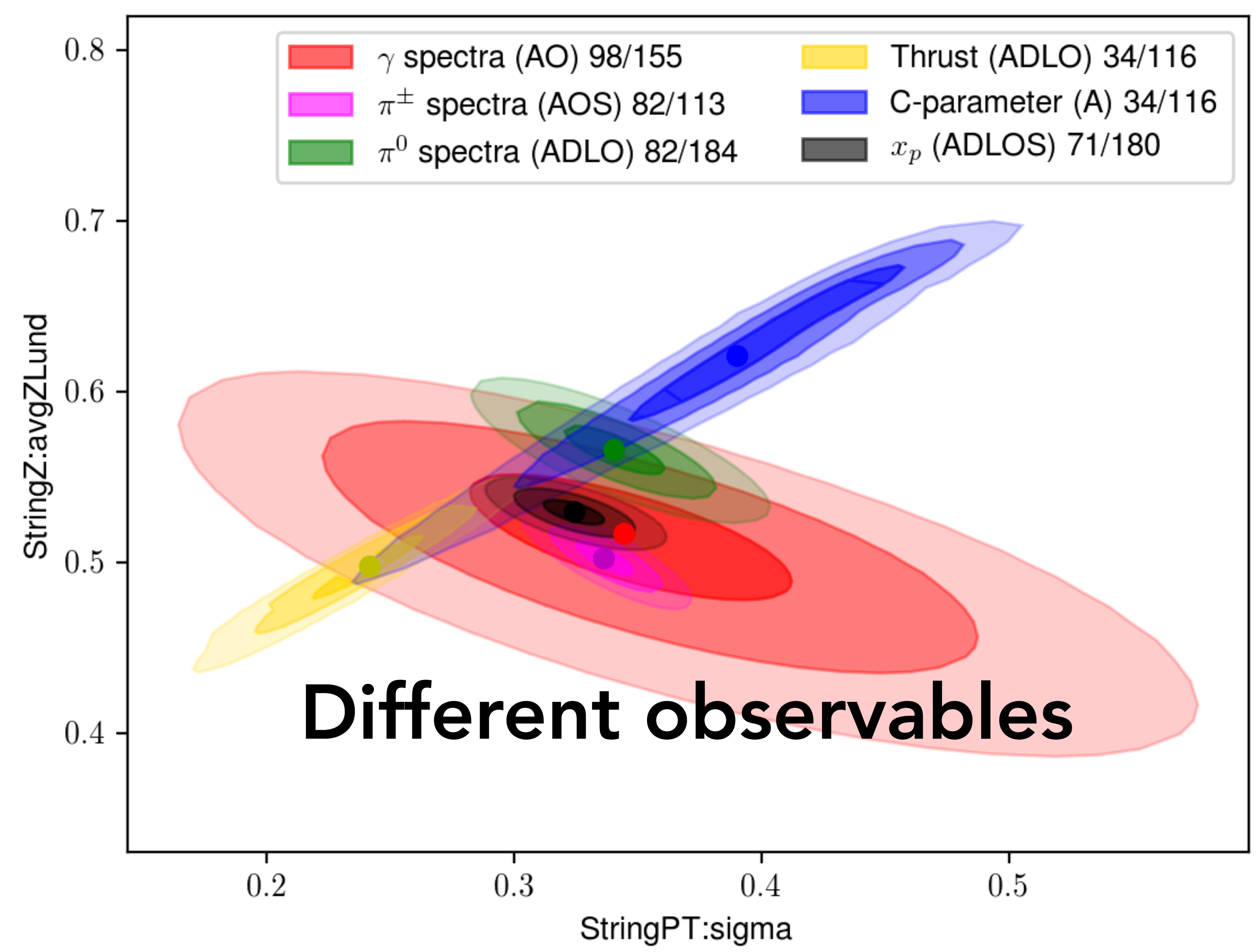
QCD uncertainties on Dark-Matter Annihilation Spectra



- Compare different generators? **Problem:** all tuned to ~ same data
- Instead, did **parametric refittings** of LEP data within PYTHIA's modelling
  - $\langle z \rangle$, bLund, σ_{p_T} : **also useful for collider studies of hadronization uncertainties**

+ universality tests: identifying and addressing tensions, overfitting & universality/consistency





Other possible universality tests (eg in pp):

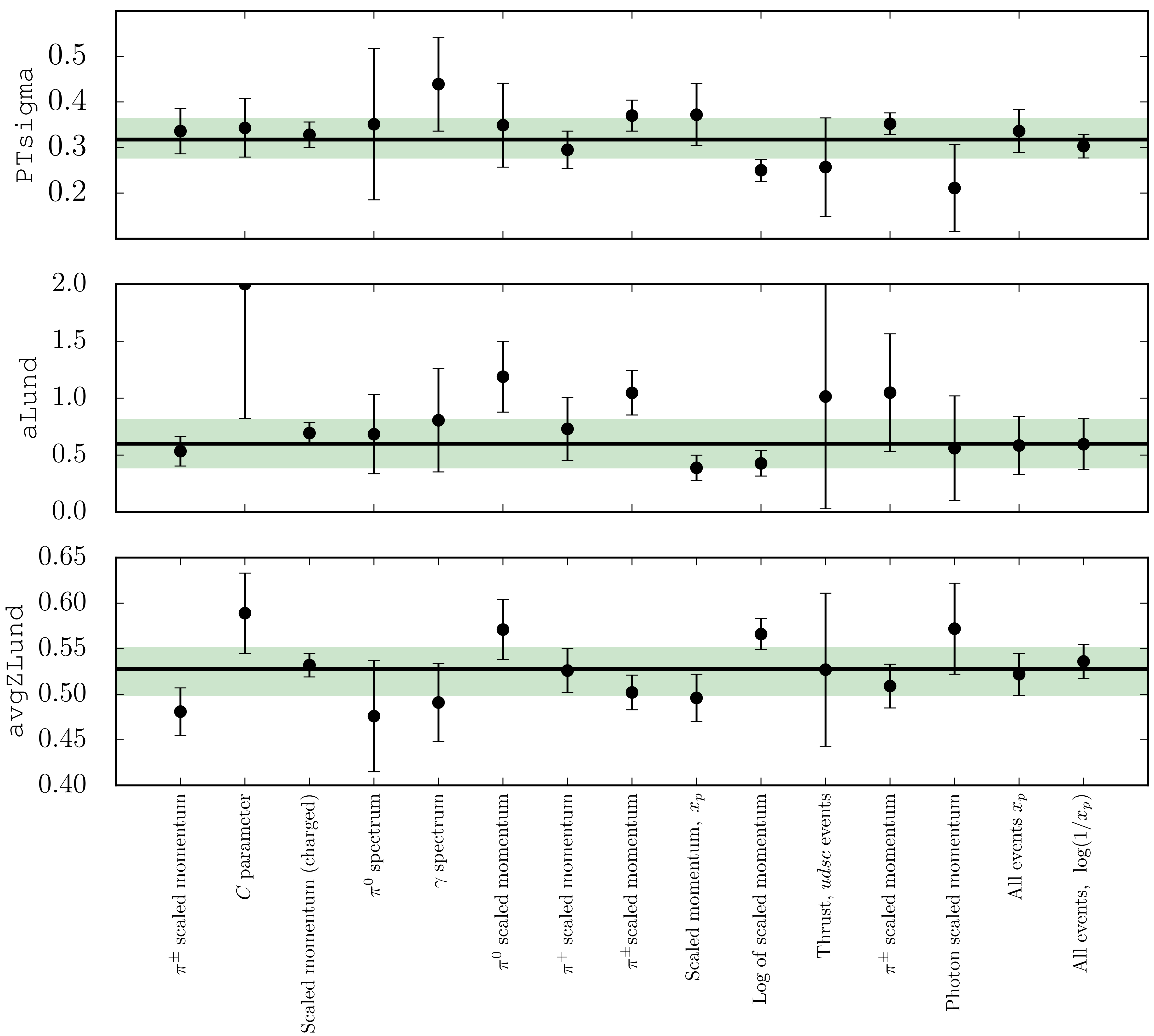
Different CM energies …
Different fiducial windows …
Different hard processes …
Quarks vs Gluons …

Simple sanity limit / overfit protection / tension resolution:
add blanket 5% baseline TH uncertainty
(+ exclude superseded measurements)

Parameter	without 5%	with 5%
StringPT:Sigma	$0.3151^{+0.0010}_{-0.00010}$	$0.3227^{+0.0028}_{-0.0028}$
StringZ:aLund	$1.028^{+0.031}_{-0.031}$	$0.976^{+0.054}_{-0.052}$
StringZ:avgZLund	$0.5534^{+0.0010}_{-0.0010}$	$0.5496^{+0.0026}_{-0.0026}$
χ^2/ndf	5169/963	778/963

Example: The Strong Force Meets the Dark Sector

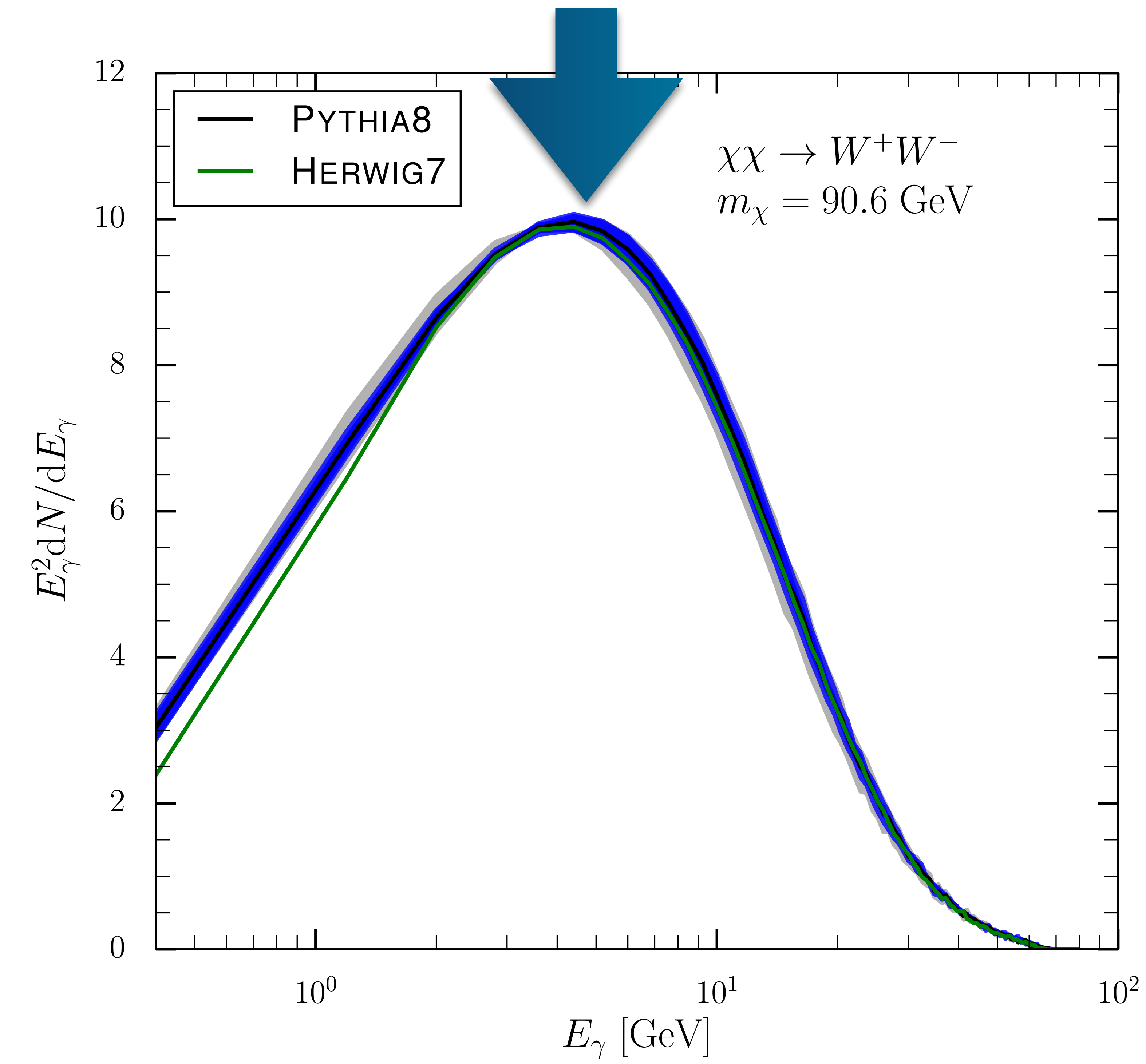
Based on A. Jueid et al., 1812.07424 (gamma rays, *eg for GCE*) and 2202.11546 (antiprotons, *eg for AMS*) + **2303.11363 (all)**



Weighted Average: good consistency across observables

Expensive?

10-point variations ➤ **Fairly convincing** uncertainty bands?



Same done for antiprotons, positrons, antineutrinos

Main Contact: adil.jueid@gmail.com

- Tables with uncertainties available on request. Also the spanning tune parameters of course.

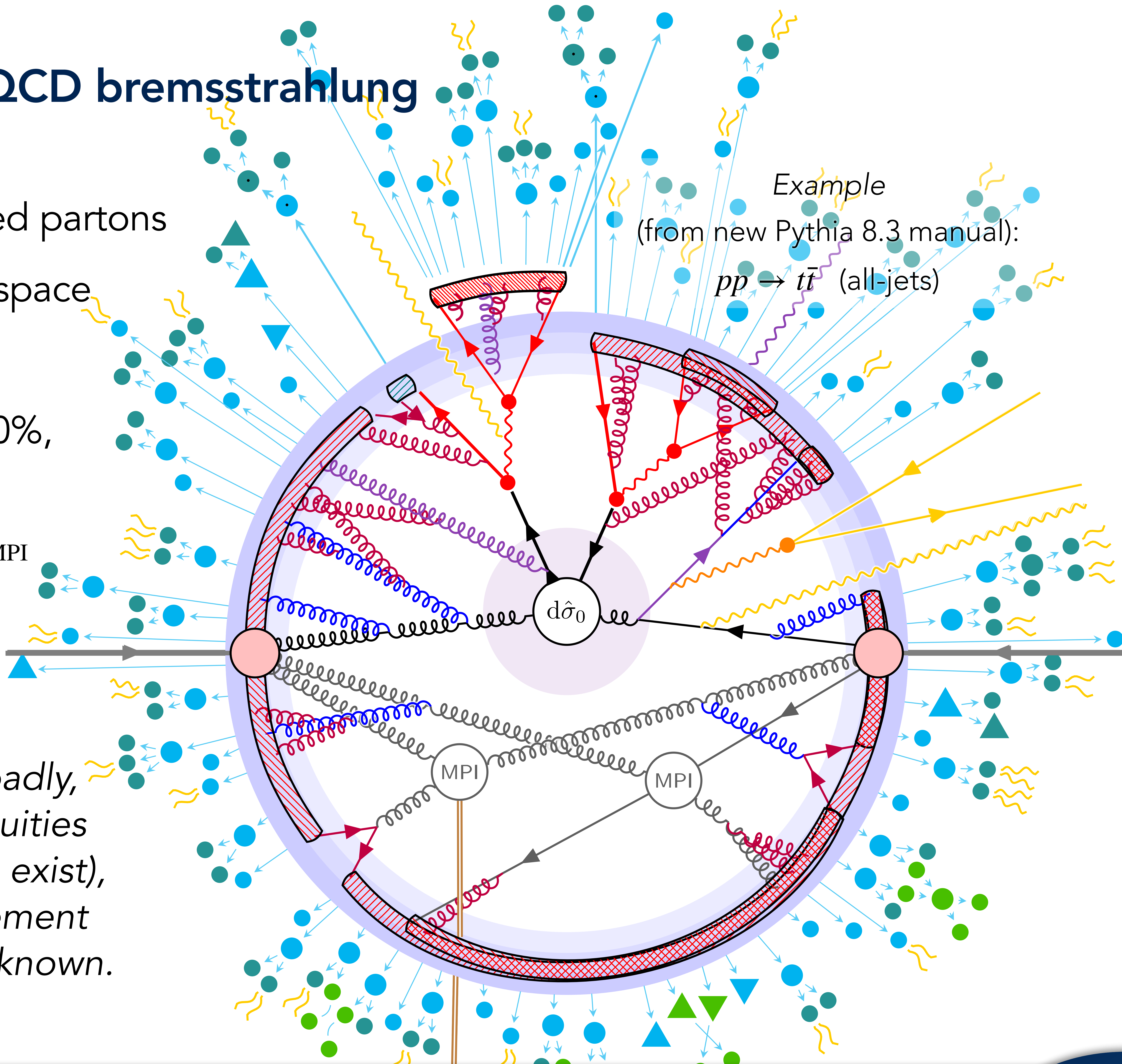
Reminder: **Colour Reconnections**

High-energy pp collisions with QCD bremsstrahlung + multi-parton interactions

- Final states with very many coloured partons
- With significant overlaps in phase space
- **Who gets confined with whom?**
- If each has a colour ambiguity ~ 10%, CR becomes more likely than not

$$\text{Prob(no CR)} \propto \left(1 - \frac{1}{N_C^2}\right)^{n_{\text{MPI}}}$$

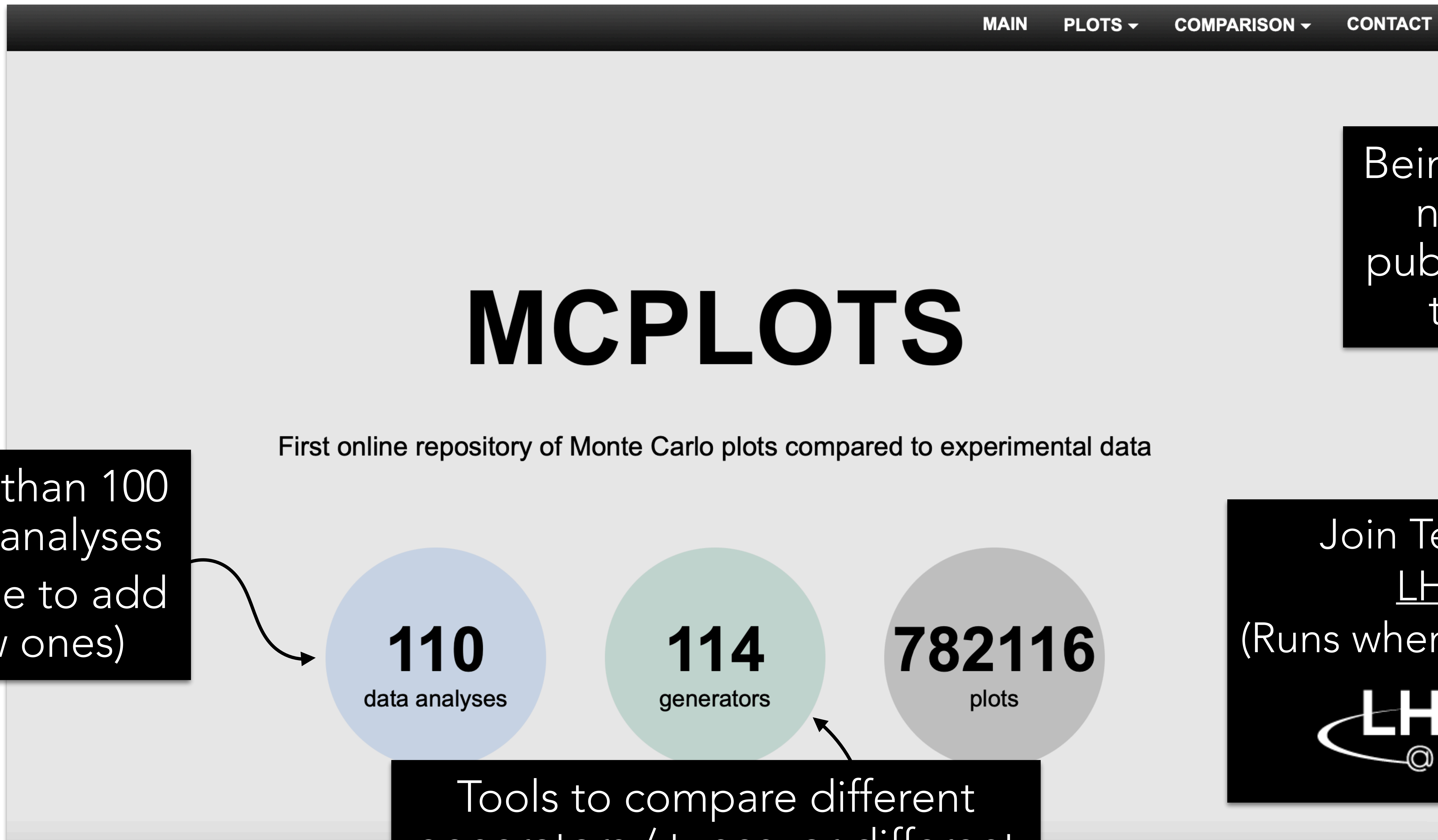
Note: the term "CR" is often used broadly, to cover everything from colour ambiguities beyond leading N_C (which are known to exist), to more speculative soft-gluon/confinement dynamics. Detailed physics not yet fully known.



mcplots.cern.ch — New and Updated coming soon!

Modern clean interface developed through 2023 (+ many improvements under the hood)

- Mainly driven by Natalia Korneeva (CMS), now an adjoint at Monash U (with support from LPCC)



Being finalised now, with publication on the way.

More than 100 Rivet analyses (simple to add new ones)

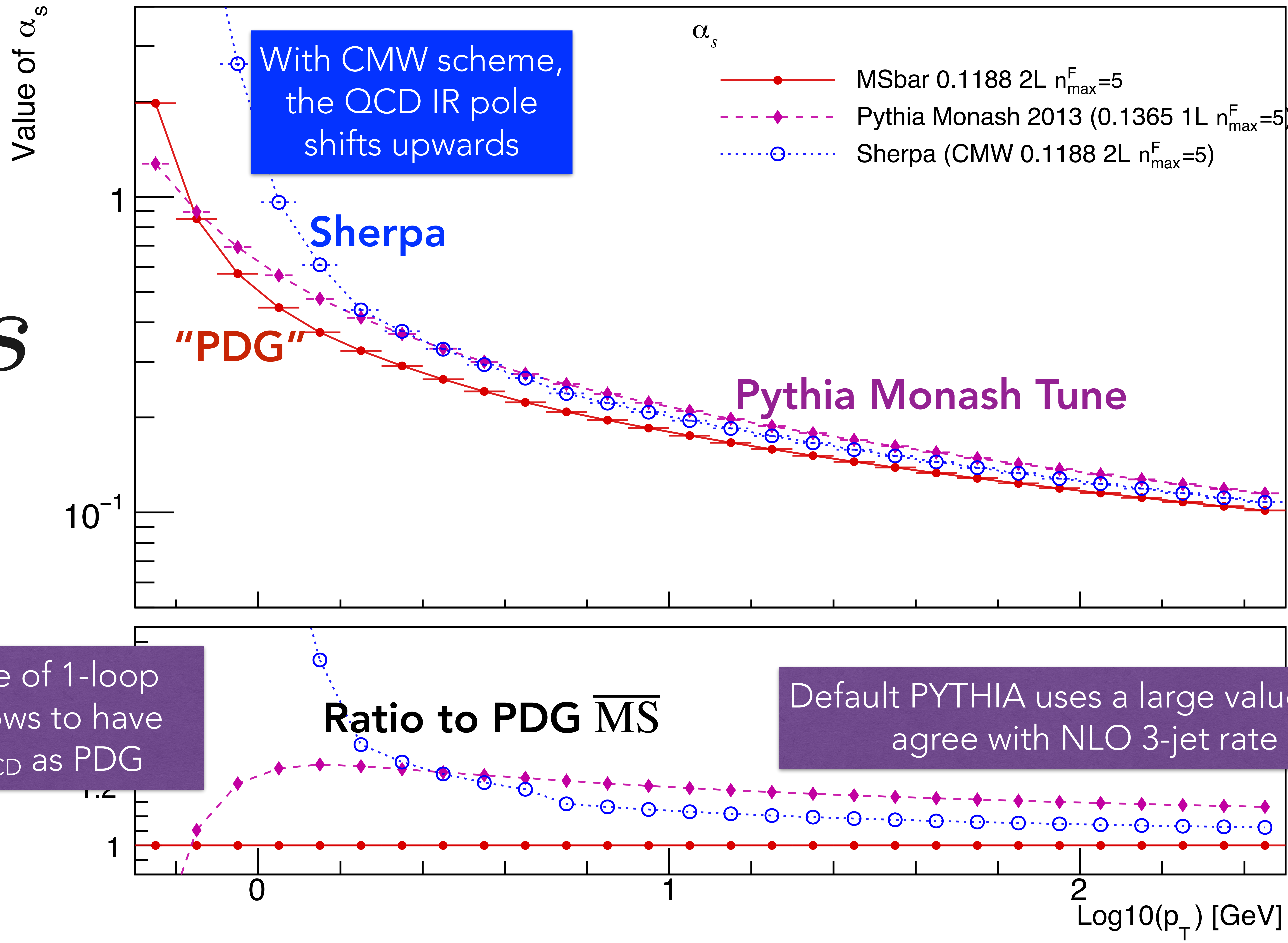
Tools to compare different generators / tunes, or different versions of same generator

Join Test4Theory on LHC@home (Runs when computer is idle)

LHC@home Volunteer computing for the LHC

Extra Slides

Note on Different alpha(S) Choices



α_s

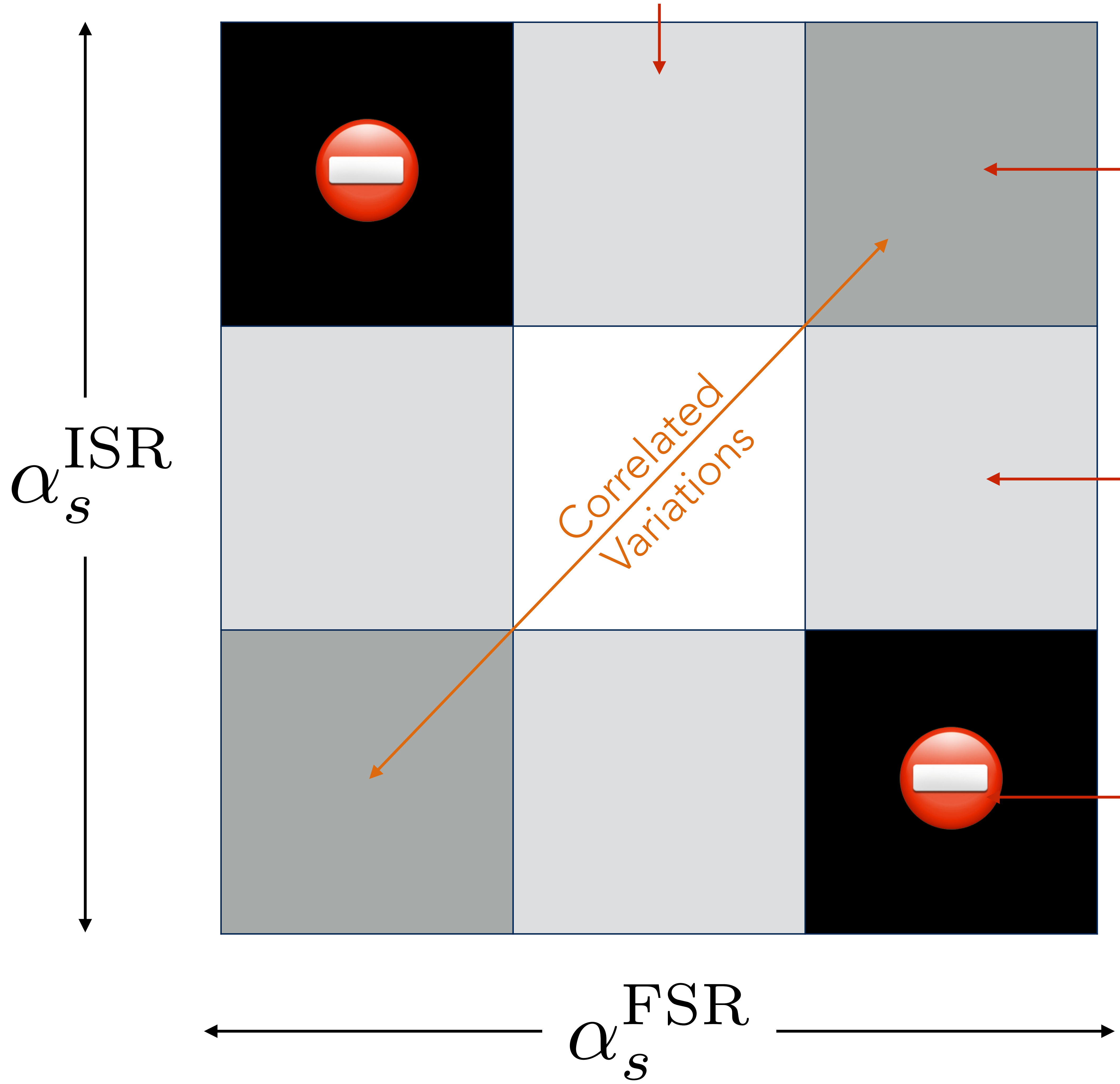
Slower pace of 1-loop running allows to have similar Λ_{QCD} as PDG

Default PYTHIA uses a large value of $\alpha_s(M_Z)$ to agree with NLO 3-jet rate at LEP

Correlated or Uncorrelated?

What I would do: **7-point variation** (resources permitting → use the automated bands?)

Increasing only ISR
➠ More H_T and N_{jets}; similar core jet shapes



Increasing both ISR and FSR
➠ More H_T in the events.
➠ More OOC loss (from FSR) but also more H_T and more hard ISR jet seeds → partial cancellation in N_{jets}?

Increasing only FSR
➠ More OOC loss (FSR jet broadening), acting on similar number of seed partons (no increase in ISR).
➠ Similar H_T

Increasing FSR, Decreasing ISR -> Exclude?
➠ Double counting? Fewer ISR partons, **and** more smearing of those that remain. *(Easy to rule out?)*
➠ Also from theoretical/mathematical point of view, the artificially induced discrepancy is now proportional to ln(16) = 2.8 instead of ln(4) = 1.4.

Scale Variations: How Big?

Scale variations induce 'artificial' terms beyond truncated order in QFT ~ Allow the calculation to float by (1+O(α_s)).

$$\frac{\alpha_s(k_1^2\mu^2)}{\alpha_s(k_2^2\mu^2)} \sim 1 - b_0 \ln(k_1^2/k_2^2)\alpha_s(\mu^2)$$

Flavour-dependent slope of order 1
$b_0 \sim 0.65 \pm 0.07$

Expansion around μ only sensible if this stays ≲ 1

Proportionality to $\alpha_s(\mu)$ ⟹ can get a (misleadingly?) small band if you choose central μ scale very large.

E.g., some calculations use μ ~ H_T ~ largest scale in event ?!

Worth keeping in mind when considering (uncertainty on) central μ choice

Mainstream view:

- Regard scale dependence as unphysical / leftover artefact of our mathematical procedure to perform the calculations.
- Dependence on it has to vanish in the 'ultimate solution' to QFT
- → Terms beyond calculated orders must sum up to at least kill μ dependence
- Such variations are thus regarded as a useful indication of the size of uncalculated terms. (Strictly speaking, only a lower bound!)

Typical choice (in fixed-order calculations): k ~ [0.5,1,2]

Note: In PYTHIA you specify k^2
TimeShower:renormMultFac
SpaceShower:renormMultFac

Scale Variations: How big?

What do parton showers do?

- In principle, LO shower kernels proportional to α_s

  Naively: do the analogous factor-2 variations of μ_{PS}.

- There are at least 3 reasons this could be **too** conservative

1. **For soft gluon emissions,** we know what the NLO term is

   → even if you do not use explicit NLO kernels, you are effectively NLO (in the soft gluon limit) **if** you are coherent and use $\mu_{PS} = (k_{CMW}\ p_T)$, with 2-loop running and $k_{CMW} \sim 0.65$ (somewhat n_f-dependent). *[Though there are many ways to skin that cat; see next slides.]*

   Ignoring this, a **brute-force** scale variation **destroys** the NLO-level agreement.

2. Although hard to quantify, showers typically achieve better-than-LL accuracy by accounting for **further physical effects** like (E,p) conservation

3. We see empirically that (well-tuned) showers tend to stay inside the envelope spanned by factor-2 variations in **comparison to data**

Scale variations: How Big?

Poor man's recipe: Use $\sqrt{2}$ instead?

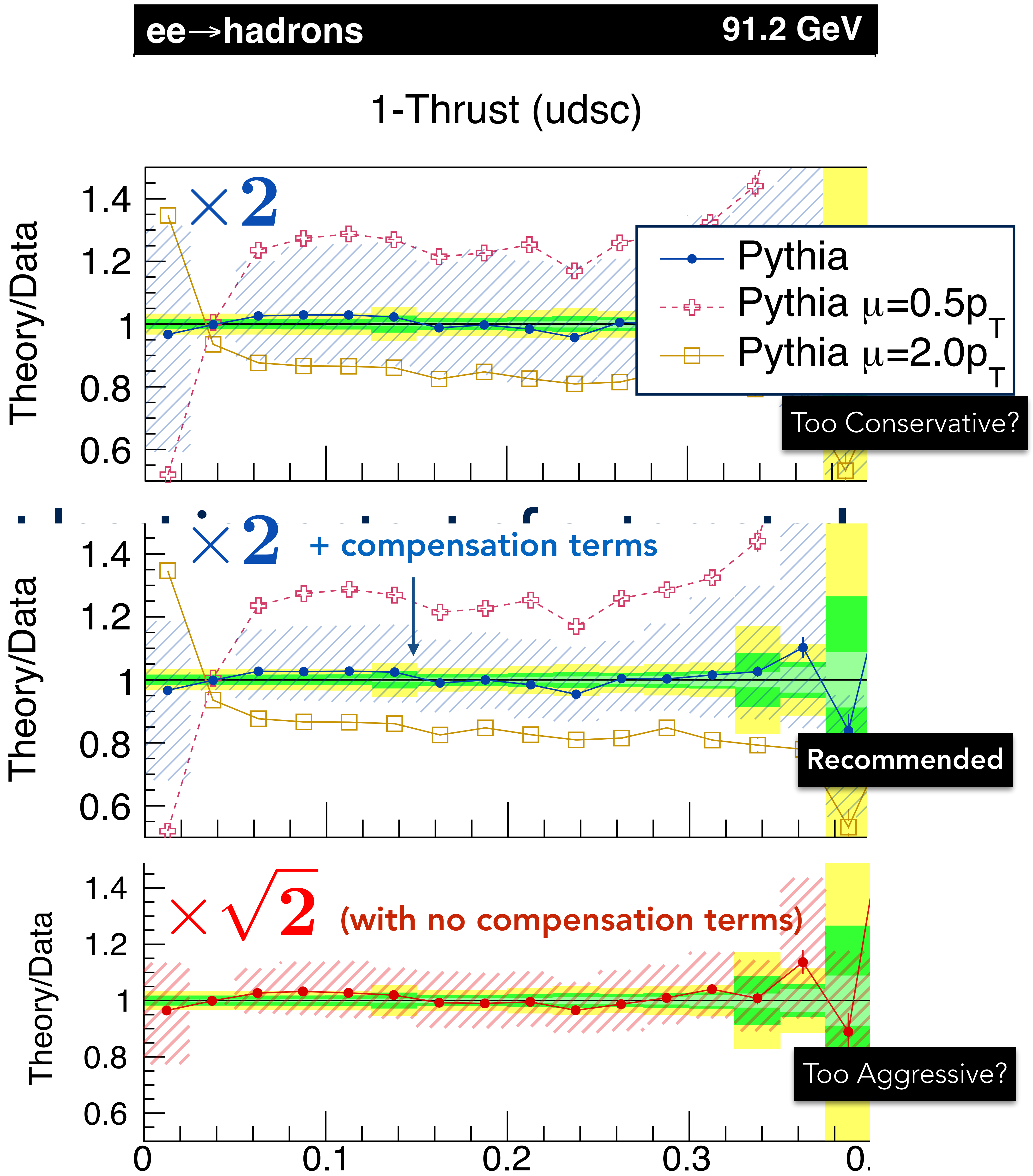
- Sure … but still somewhat arbitrary

Instead: add compensation term to preserv

- Still allowing full factor-2 outside that limit.

Pythia includes such a compensation term, a uncertainty bands *(next slides)*.

- Since aggressive definitions can lead to overcom predictions → very small uncertainty bands, we c PYTHIA → larger bands.



$$P'(t,z) = \frac{\alpha_s(kp_\perp)}{2\pi}\left(1 + (1-\zeta)\frac{\alpha_s(\mu_{\max})}{2\pi}\beta_0 \ln k\right)\frac{P(z)}{t}$$

Kills the compensation outside the soft limit

Small absolute size of compensation

$$\zeta = \begin{cases} z & \text{for splittings with a } 1/z \text{ singularity} \\ 1-z & \text{for splittings with a } 1/(1-z) \text{ singularity} \\ \min(z, 1-z) & \text{for splittings with a } 1/(z(1-z)) \text{ singularity} \end{cases}$$

S. Mrenna & PS: PRD94(2016)074005; arXiv:1605.08352