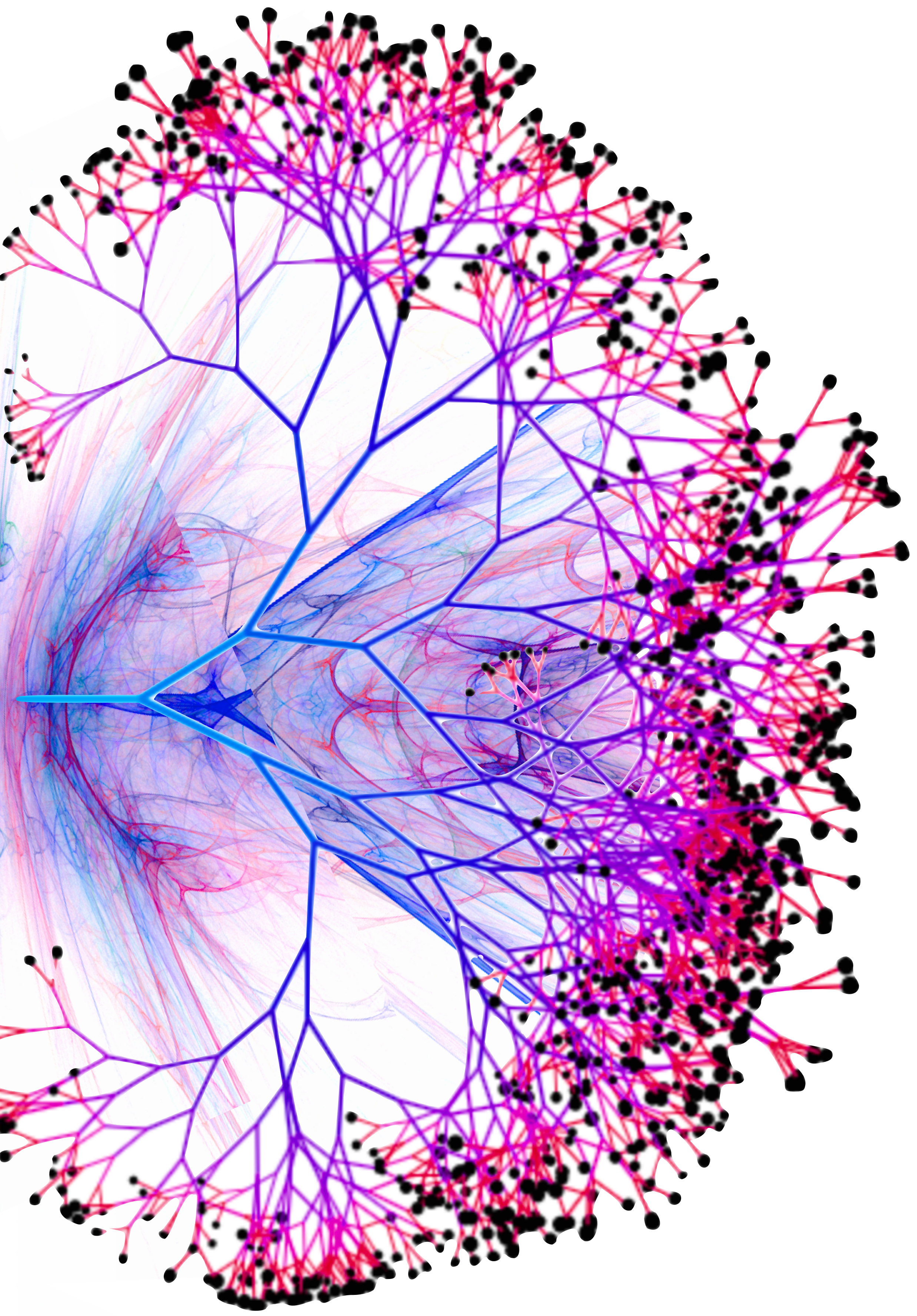
Soft QCD in MC Event Generators

(A selection of topics focusing on pp, with emphasis on Pythia)



1. **Hadronization Uncertainties** for **Precision Studies**
2. **Multiple Parton Interactions** & **PDFs**
3. **Colour Reconnections** & **Heavy-Flavour** Baryons
4. **Strangeness**, Ropes, and (Advanced) Close-Packing

Peter Skands
University of Oxford & Monash University

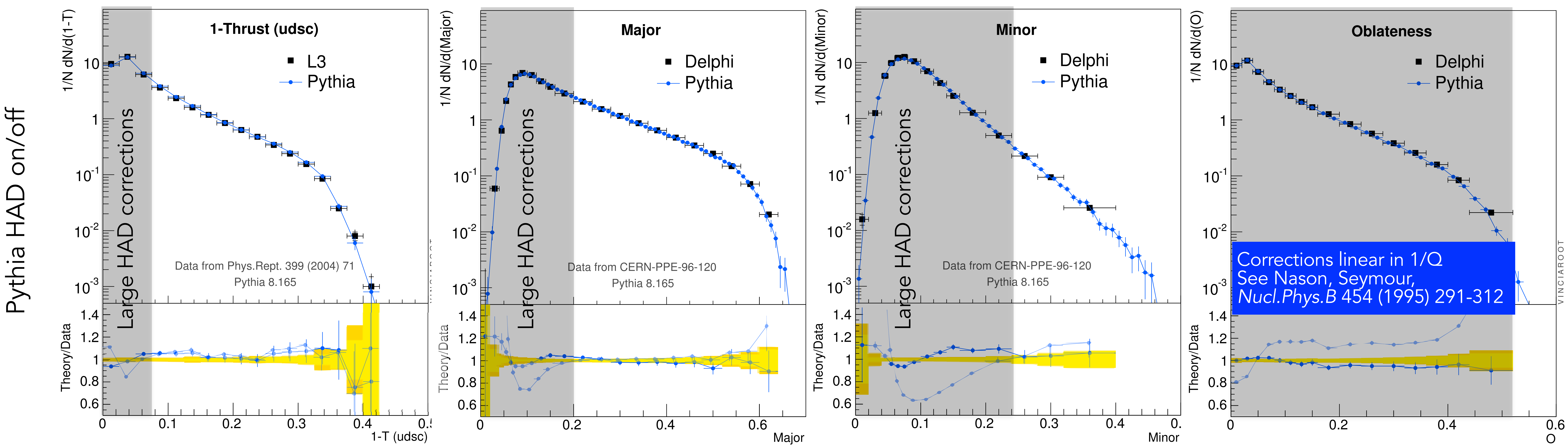
QCD@LHC
Durham — September 2023

1. **Hadronization Uncertainties** for **Precision Studies**

Map: **Partons** (defined at a low factorisation scale, after showering) → **Hadrons**

Inclusive sums

- Fully Inclusive: **Power Corrections** (to IRC Safe Observables)
- Semi-Inclusive: **Fragmentation Functions**: One hadron species at a time
- Fully Exclusive: **Dynamical Models** in MC Event Generators



Important point: even for nominally IRC safe observables, **peaks** of distributions often involve low scales where **HAD sensitivity is highest** $\Longrightarrow$ NP peak shifts.

Uncertainties

High-Precision Measurements ↔ Rigorous & Exhaustive Uncertainties

- **Expensive** to construct & perform all salient parm variations individually → GEANT …

  Not just question of CPU; also **environmental impact**, cost, inefficient duplication of man-hours & higher risk of mistakes/inconsistencies (by non-authors) + risk that lessons learned aren't perpetuated

- **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.)

Hadronization Uncertainties: More parameters and lots of subtleties

Interplay between **perturbative** (eg N_{jets}) and **nonperturbative** (eg $N_{hadrons}$) observables

Parameter correlations; for a helping hand, see AutoTunes [Bellm & Gellersen, 1908.10811]

Risk of purely data-driven methods (eg eigentunes) to **overfit** precise data points at expense of tails / asymptotics / less statistically dominant (but perhaps theoretically important) data

Tensions between different measurements

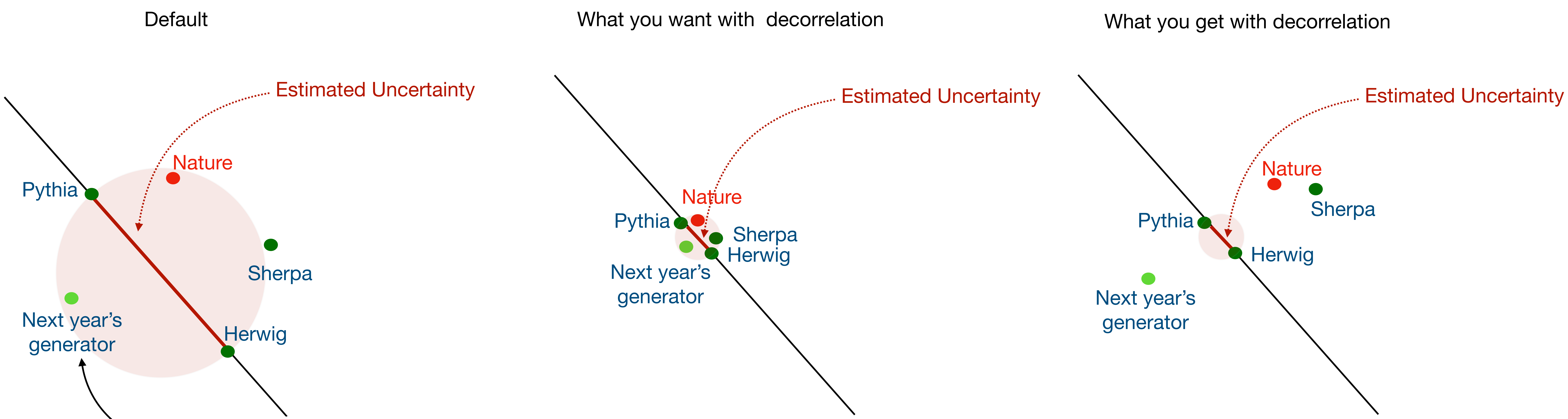
- Recent elaborate studies with PYTHIA 8, see eg: [Jueid et al., 1812.07424; 2202.11546; 2303.11363]

Another aspect of the problem

Pythia, Herwig, Sherpa all tuned to ~ same data ➤ risk central tunes being "too" similar?
No guarantee that they span the experimental uncertainties (similar issue as of old with PDFs)

Borrowed slide from A Ghosh ↔ Machine Learning of/for Theory Models

EPJC:s10052.022.10012.w: **Aishik Ghosh**, Benjamin Nachman



*Personal Comment: I would kind of hope next year's generator would be **closer** to Nature, not further from it…*

Instruction to ML: "Please shrink Pythia vs Herwig difference" **Model will learn to fool you !**

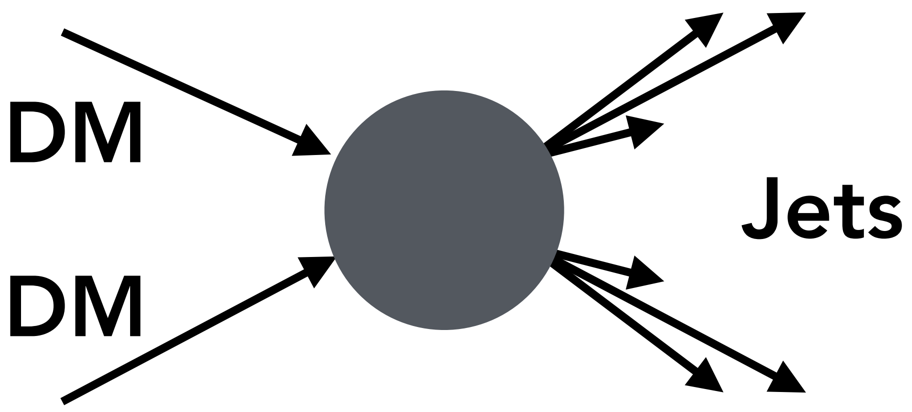
ML methods don't often generalise the way you would hope

17

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)**

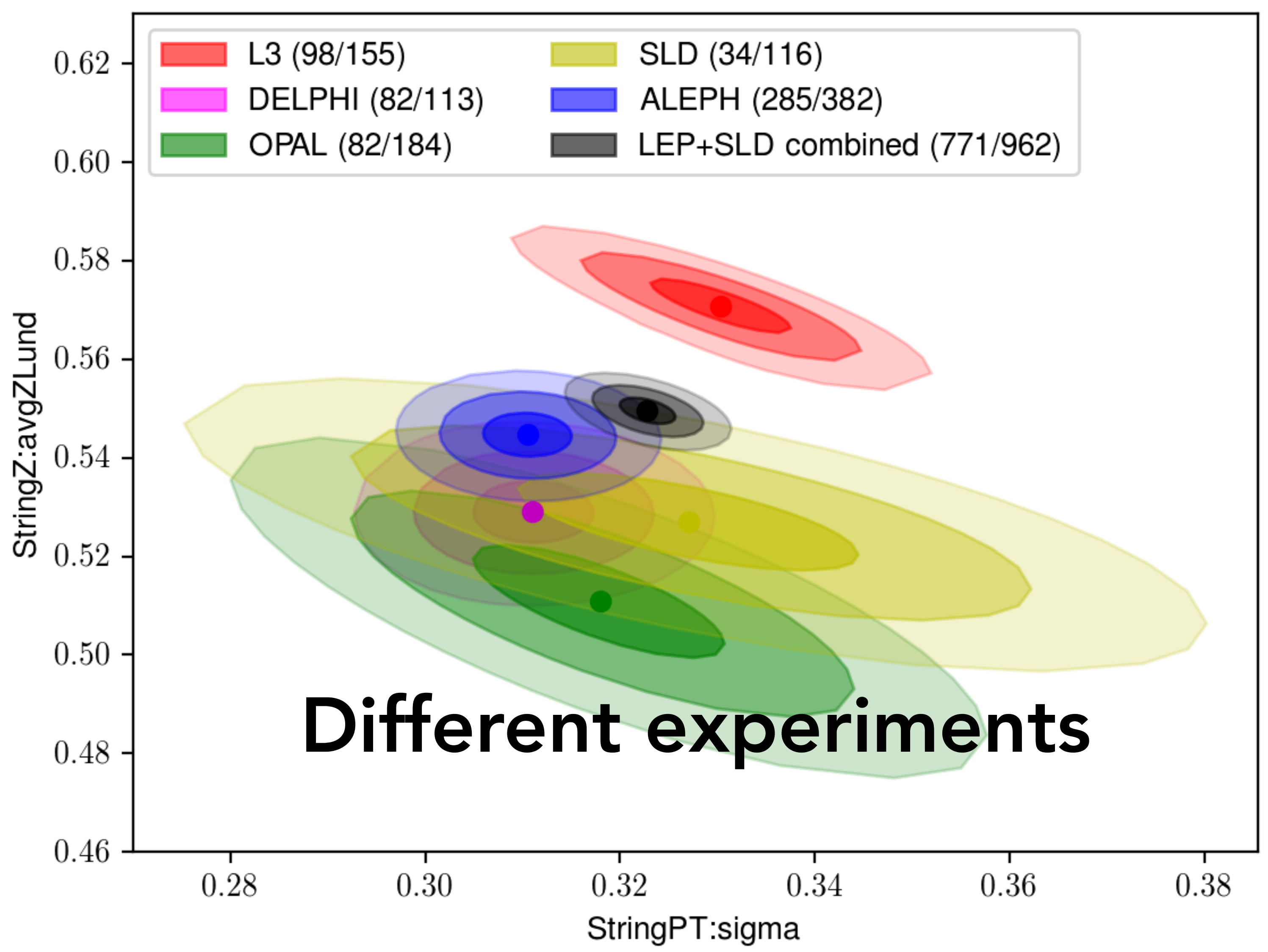
QCD uncertainties on Dark-Matter Annihilation Spectra

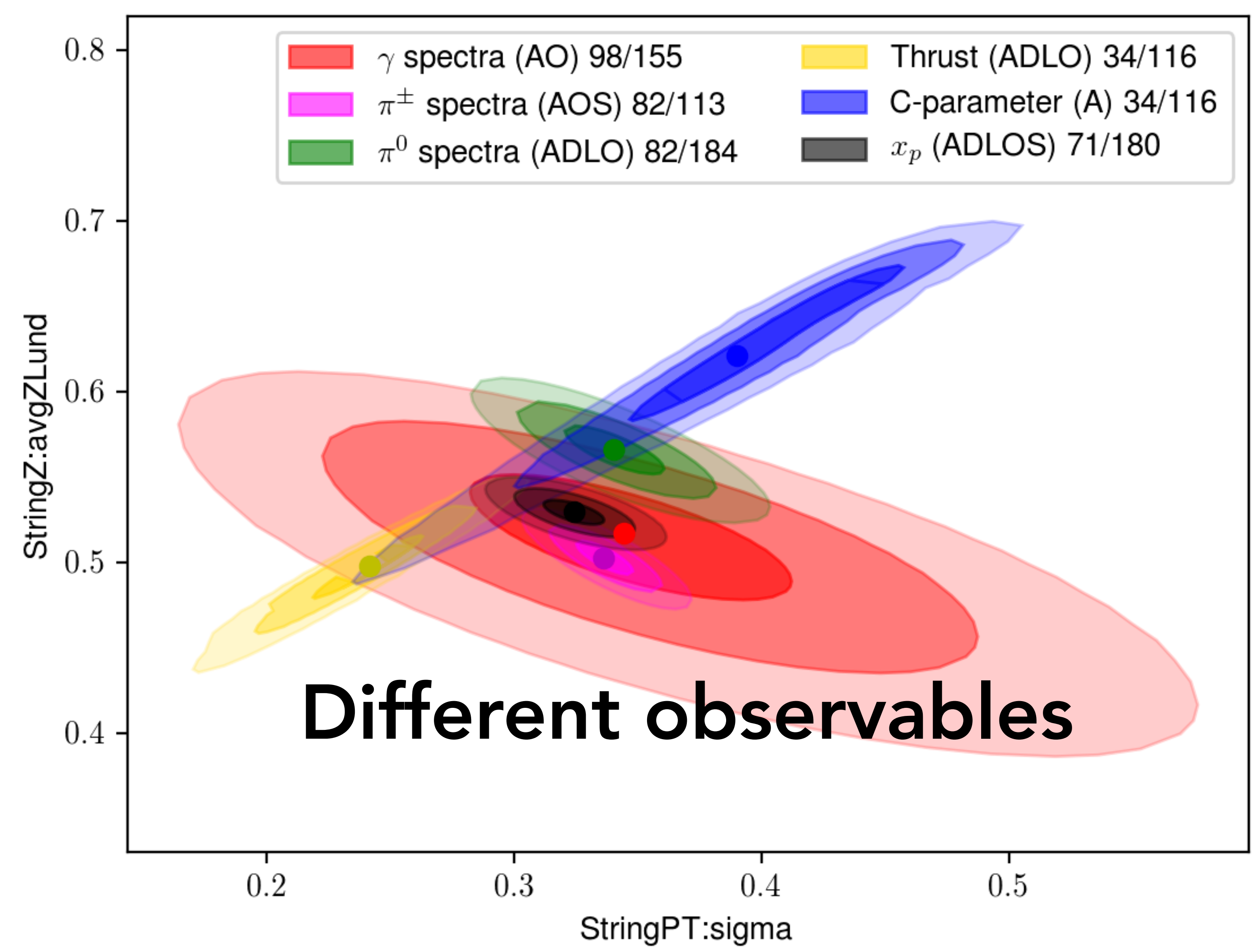


- Compare different generators? **Problem:** all tuned to ~ same data
- Instead, did **parametric refittings** of LEP data within PYTHIA's modelling

  $\langle z \rangle$, 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 ...

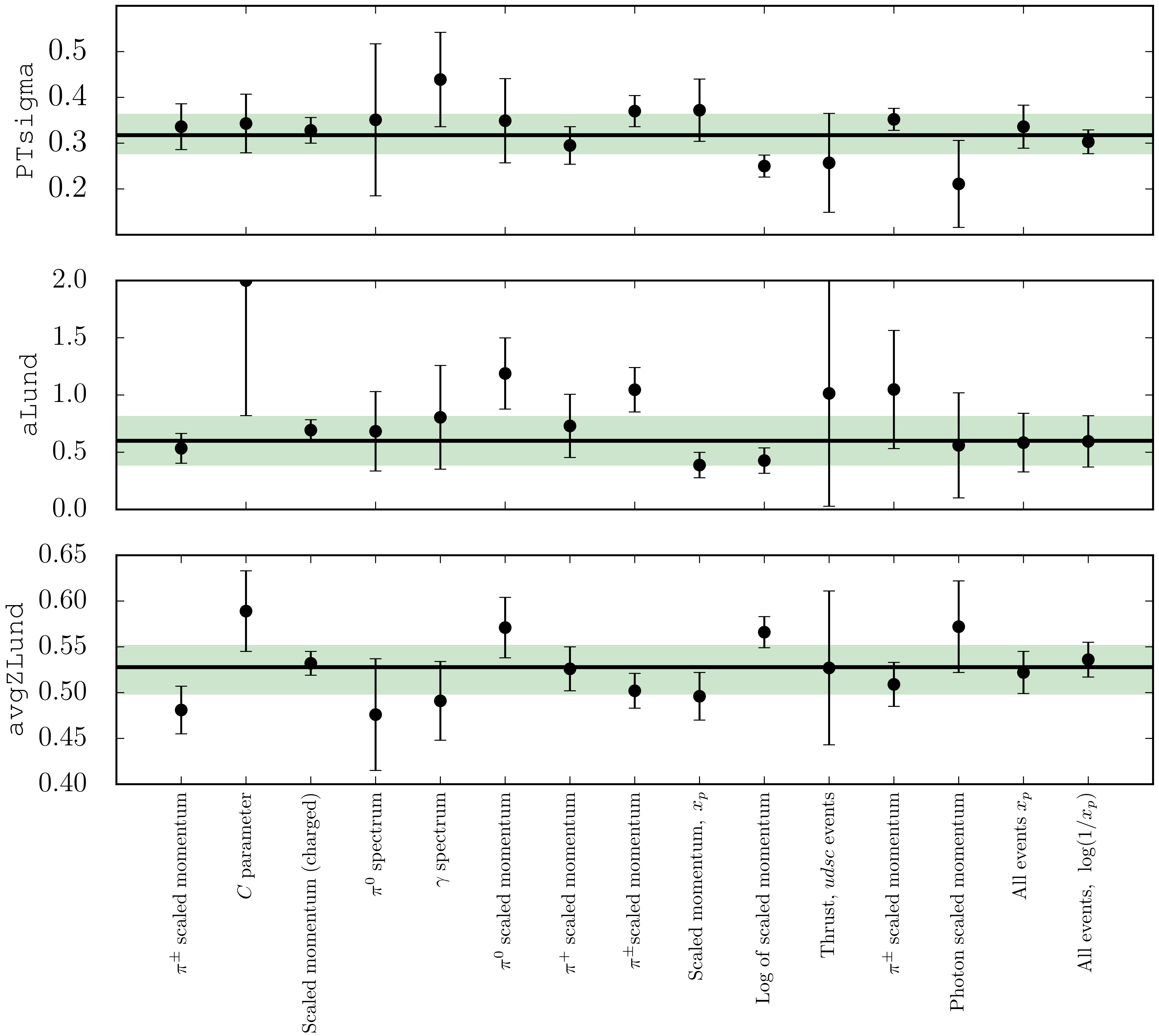
Different hard processes ...

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	$5169/963$	$778/963$

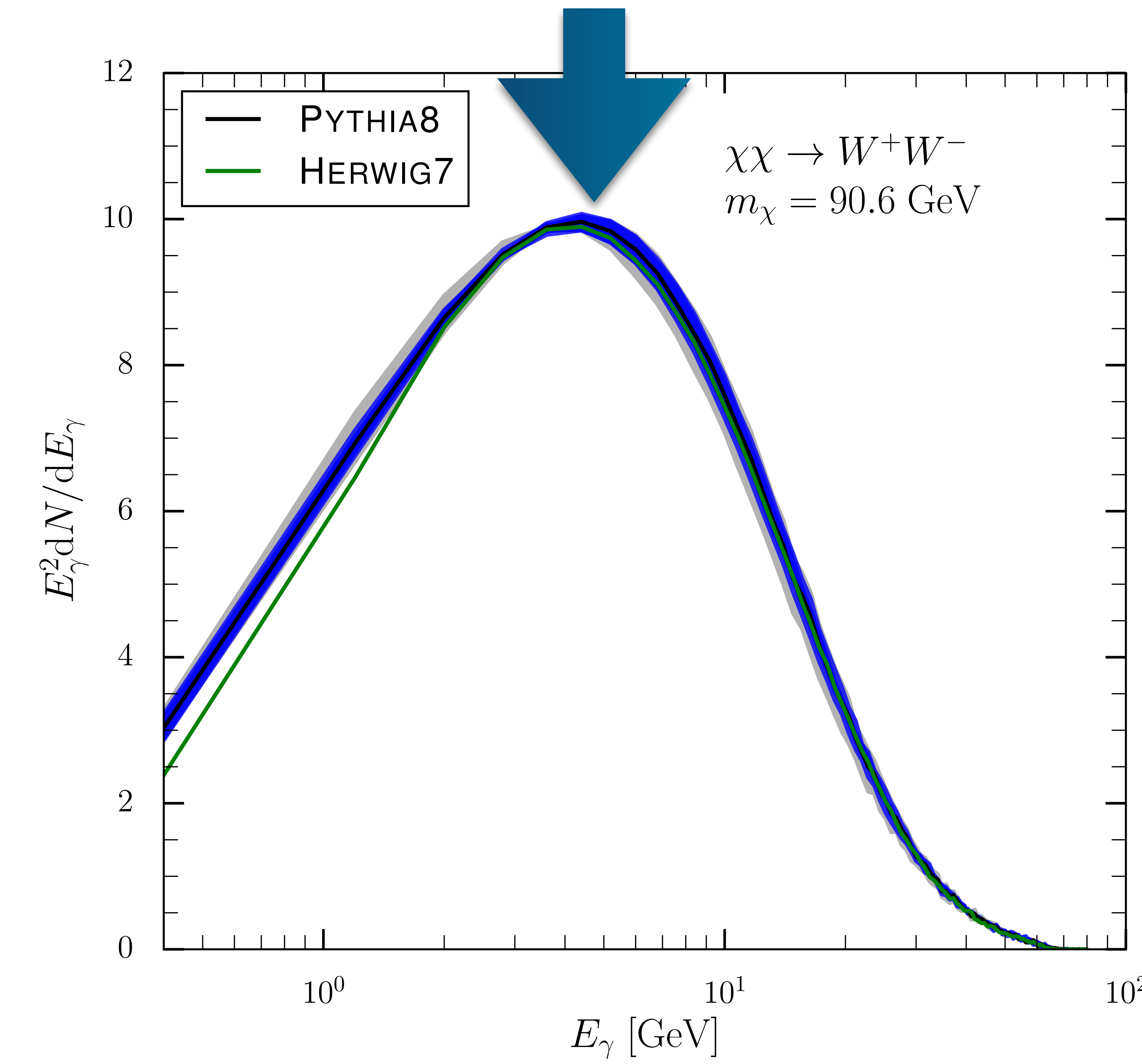
Based on A. Jueid et al., [1812.07424]() (gamma rays, *eg for GCE*) and [2202.11546]() (antiprotons, *eg for AMS*) + **[2303.11363]() (all)**



Weighted Average: good consistency across observables

Expensive?

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



Same done for antiprotons, positrons, antineutrinos

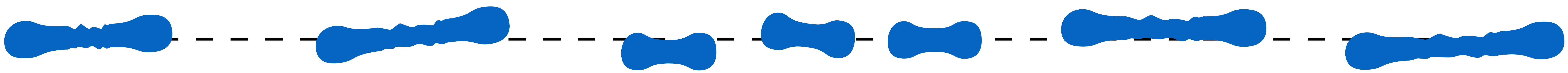
Main Contact: adil.jueid@gmail.com

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

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

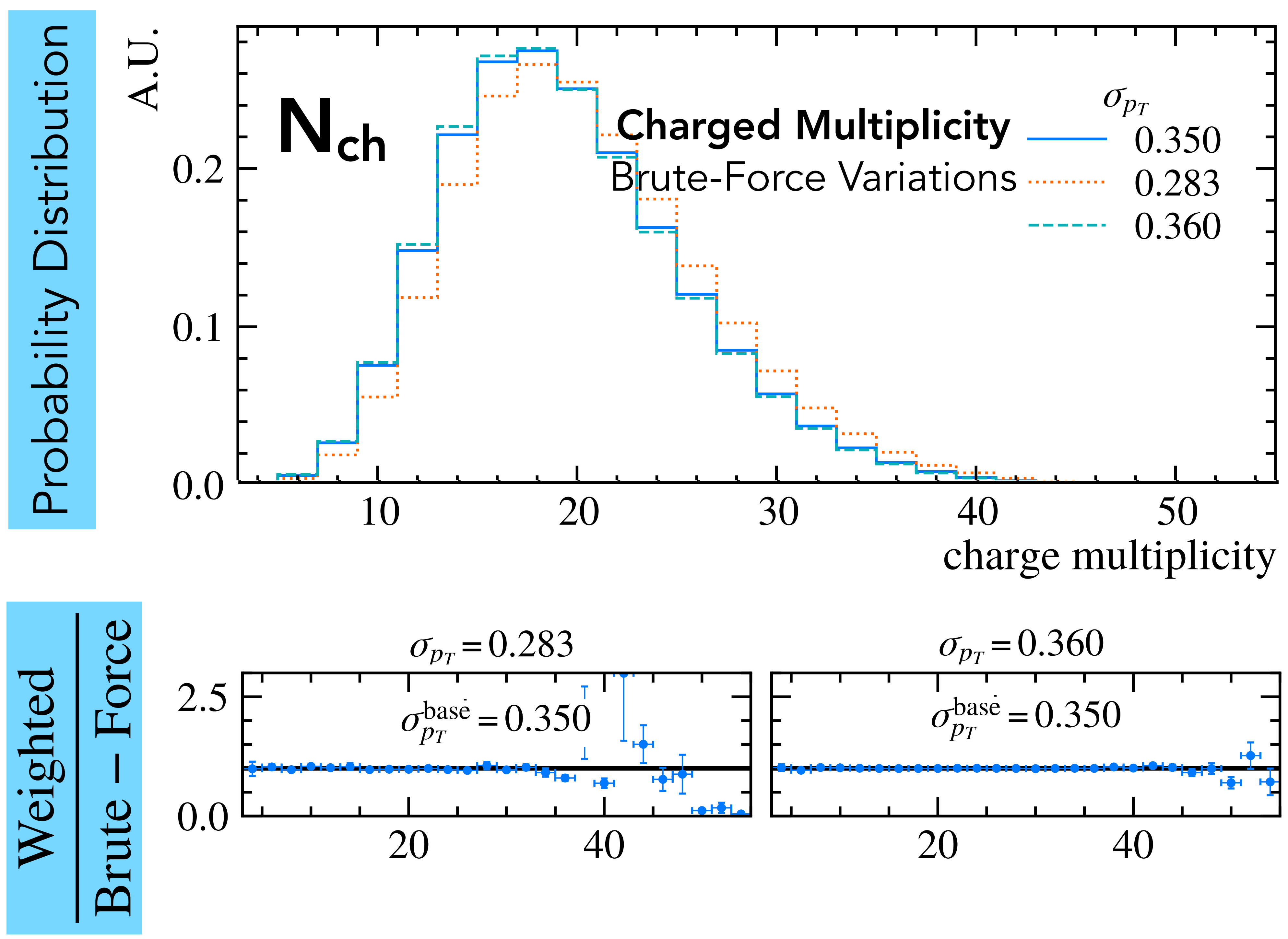
Transverse FF (Gaussian)

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

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



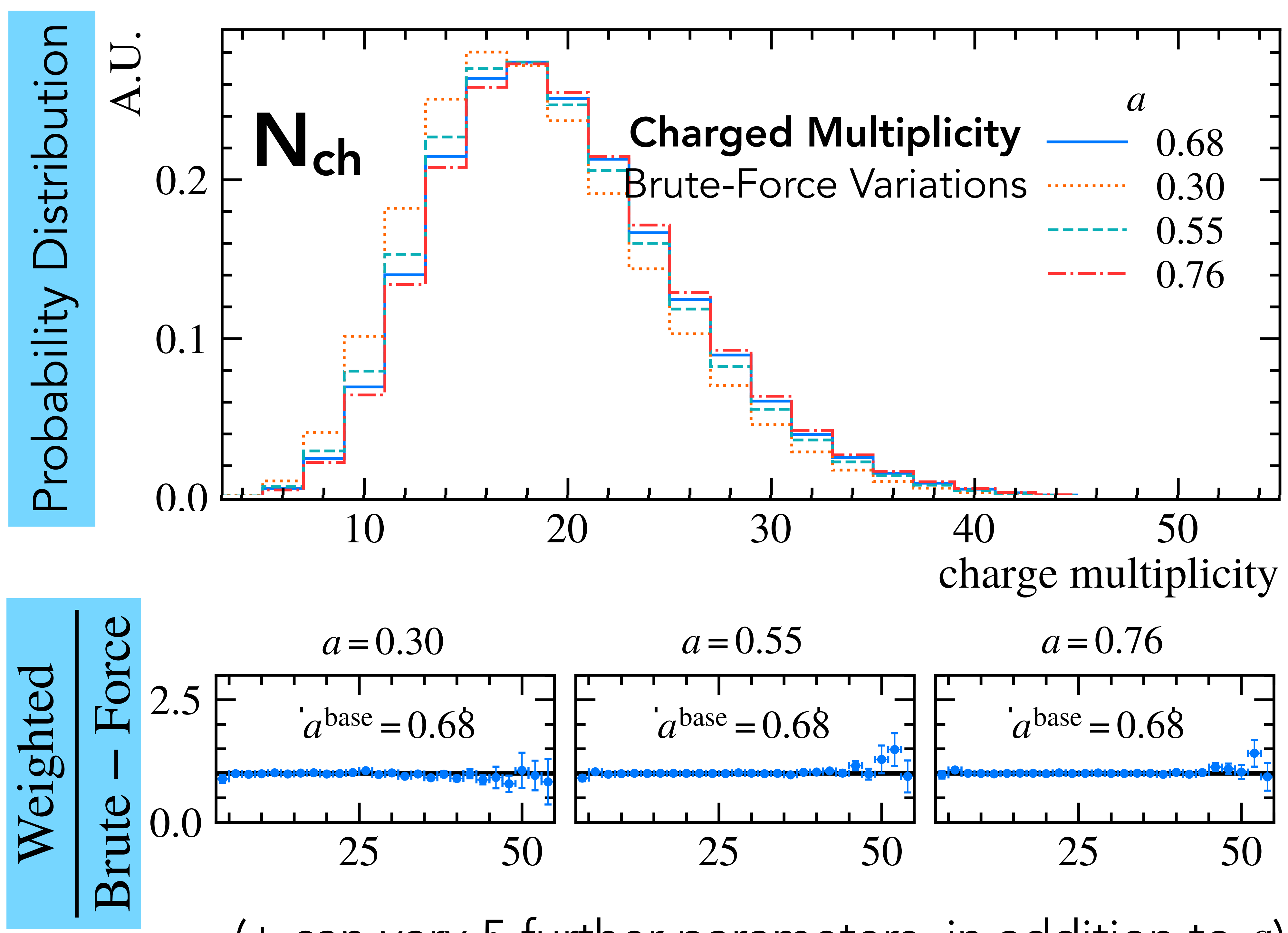
Longitudinal FF (Lund Symmetric FF)

f(z) ~ *scaled* light-cone hadron momentum fraction $\propto \frac{1}{z^{1+r_Q b m_Q^2}}(1-z)^a \exp\left(-\frac{b m_\perp^2}{z}\right)$

Accept-Reject Algorithm

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

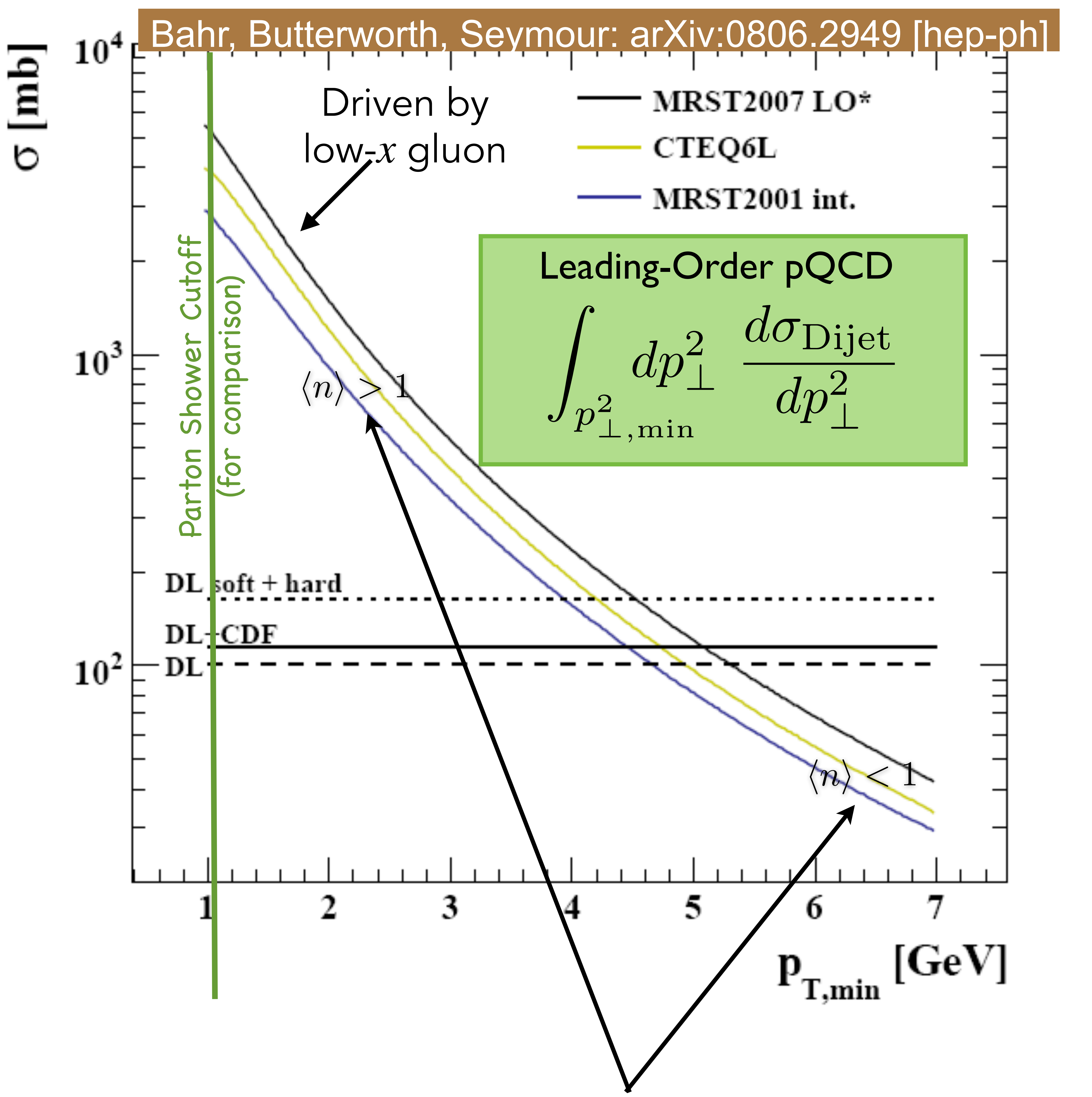
with $R'_{\mathrm{accept}}(z) = \frac{P'_{\mathrm{accept}}(z)}{P_{\mathrm{accept}}(z)}$; $R'_{\mathrm{reject}}(z) = \frac{P'_{\mathrm{reject}}(z)}{P_{\mathrm{reject}}(z)} = \frac{1-P'_{\mathrm{accept}}(z)}{1-P_{\mathrm{accept}}(z)}$



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

2. Multiple Parton Interactions & PDFs

QCD dijet cross section (cumulative)



$$\sigma_{2\to 2}(p_{\perp\min}) = \langle n \rangle (p_{\perp\min})\ \sigma_{\rm tot}$$

Parton-Parton Cross Section

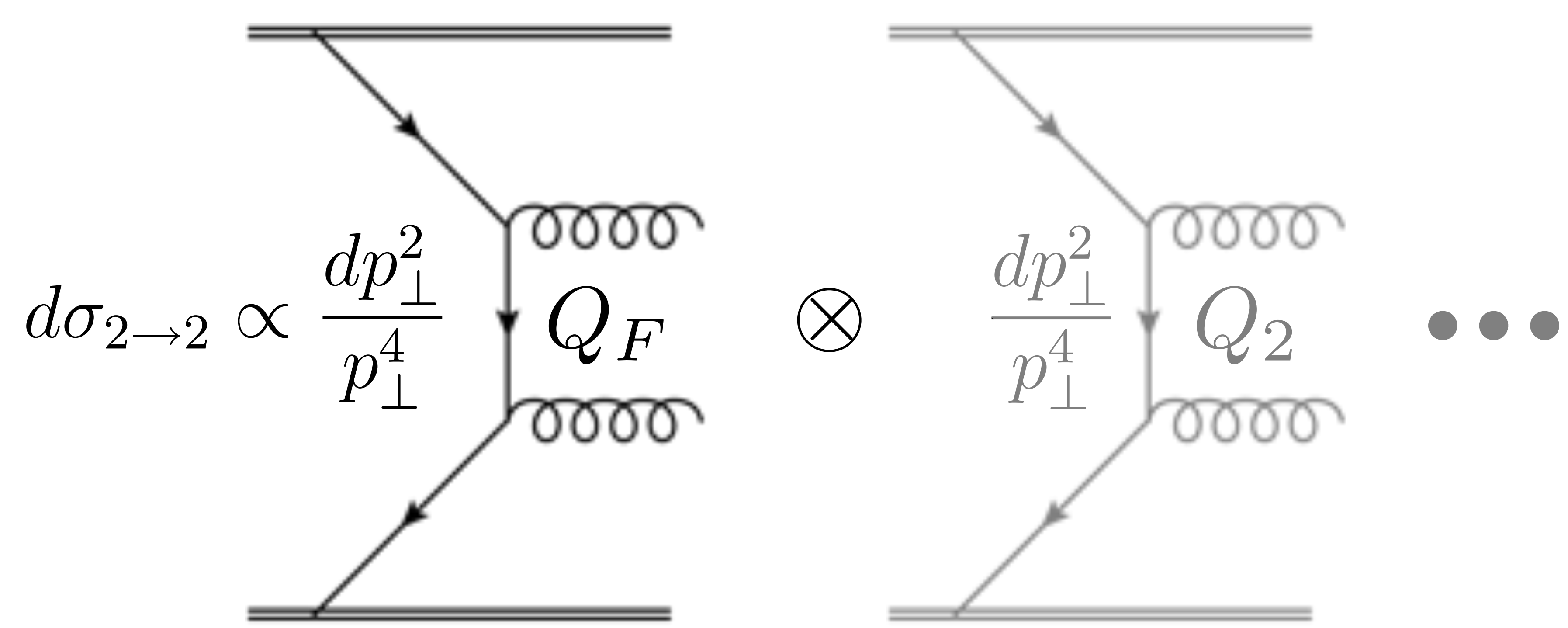
Hadron-Hadron Cross Section

Lesson from bremsstrahlung in pQCD:
Divergences → fixed-order breaks down
Perturbation theory still ok, with *resummation* (unitarity)

Unitarity: Divergent cross section for one emission reinterpreted as finite cross section for a **divergent number of emissions**

→ Resum dijets? Yes → MPI!

Interpret to mean that **every** pp collision has **more than one** $2 \to 2$ QCD scattering with $\hat{p}_\perp \lesssim 4\,\mathrm{GeV}$



MPI probe low p_T scales down to $Q \sim 1\,\mathrm{GeV}$

And very low x scales, down to $x \sim 1/s_{\rm hh}$

Earliest MC model (“old” PYTHIA 6 model) Sjöstrand, van Zijl PRD36 (1987) 2019

The issue with NLO gluons at low x

(Summary of note originally written by T. Sjöstrand, from discussions with R. Thorne though any oversimplifications or misrepresentations are our own)

Low-x gluon

Key constraint: DIS F_2

Low x: $\mathrm{d}F_2/\mathrm{d}\ln(Q^2)$ driven by $g \to q\bar{q}$

LO $P_{q/g}(z)$ ~ flat $\Longrightarrow$ x of measured quark closely correlated with x of mother gluon.

NLO $P_{q/g}(z) \propto 1/z$ for small z $\Longrightarrow$ Integral over z produces an approximate $\ln(1/x)$ factor.

➤ Effectively, the NLO gluon is probed more "non-locally" in x.

$\mathrm{d}\ln F_2/\mathrm{d}Q^2$ at small x becomes too big unless positive contribution from medium-to-high-x gluons (derived from $\mathrm{d}\ln F_2/\mathrm{d}Q^2$ in that region, and from other measurements) is combined with a **negative contribution from low-x gluons**.

Mathematically (toy NLO Calculation with just one x):

$$\frac{\mathrm{ME_{NLO}}}{\mathrm{ME_{LO}}} = 1 + \alpha_s(A_1 \ln(1/x) + A_0)$$

$\ln(1/x)$ largely compensated in def of NLO PDF:

$$\frac{\mathrm{PDF_{NLO}}}{\mathrm{PDF_{LO}}} = 1 + \alpha_s(B_1 \ln(1/x) + B_0)$$

➤ Product well-behaved at NLO if we choose $B_1 \approx A_1$

Cross term at $\mathcal{O}(\alpha_s^2)$ is beyond NLO accuracy …

For large x and small $\alpha_s(Q^2)$, e.g. $\alpha_s A_1 \ln(1/x) \sim 0.2$:

$$\frac{\mathrm{ME_{NLO}\,PDF_{NLO}}}{\mathrm{ME_{LO}\,PDF_{LO}}} = (1+0.2)(1-0.2) = 0.96$$

👍 log terms cancel

But if x and Q^2 are small, say $\alpha_s A_1 \ln(1/x) \sim 2$:

$$\frac{\mathrm{ME_{NLO}\,PDF_{NLO}}}{\mathrm{ME_{LO}\,PDF_{LO}}} = (1+2)(1-2) = -3$$

👎 Cross term dominates; The PDF becomes negative

Not so important for high-p_T processes because 1) DGLAP evolution fills up low-x region, 2) kinematics restricted to higher x, 3) smaller α_s

Some Desirable Properties for PDFs for Event Generators

General-Purpose MC Generators are used to address very diverse physics phenomena and connect (very) high and (very) low scales ➤ **Big dynamical range!**

1. Stable (& positive) evolution to **rather low Q^2 scales**, e.g. $Q_0 \lesssim 1\,\mathrm{GeV}$

   ISR shower evolution and MPI go all the way down to the MC IR cutoffs ~ 1 GeV

2. **Extrapolates sensibly to very low** $x \sim 10^{-8}$ (at LHC), especially at low $Q \sim Q_0$.

   "Sensible" ~ positive and smooth, without (spurious) structure

   Constraint for perturbative MPI: $\hat{s} \geq (1\,\mathrm{GeV})^2 \implies x_{\mathrm{LHC}} \gtrsim 10^{-8} \quad (x_{\mathrm{FCC}} \geq 10^{-10})$

   Main point: MPI can probe a **large range of** x, beyond the usual $\sim 10^{-4}$

   (Extreme limits are mainly relevant for ultra-forward / beam-remnant fragmentation)

3. **Photons** included as partons

   Bread and butter for part of the user community

4. **LO** or equivalent in some form (possibly with α_s^{eff}, relaxed momentum sum rule, …)

   Since MPI Matrix Elements are LO; ISR shower kernels also LO (so far)

5. Happy to have **N^{n}LO** ones in a similar family.

   E.g., for use with higher-order MEs for the hard process.

   Useful (but possible?) for these to satisfy the other properties too?

3. **Colour Reconnections** & **Heavy-Flavour** Baryons

Hadronization

- Map: Partons (defined at a low factorisation scale, after showering) → Hadrons
- Between which partons do the confining potentials form?

Starting point for MC generators = Leading Colour limit $N_C \to \infty$

$\Longrightarrow$ Probability for any given colour charge to accidentally be same as any other $\to 0$.

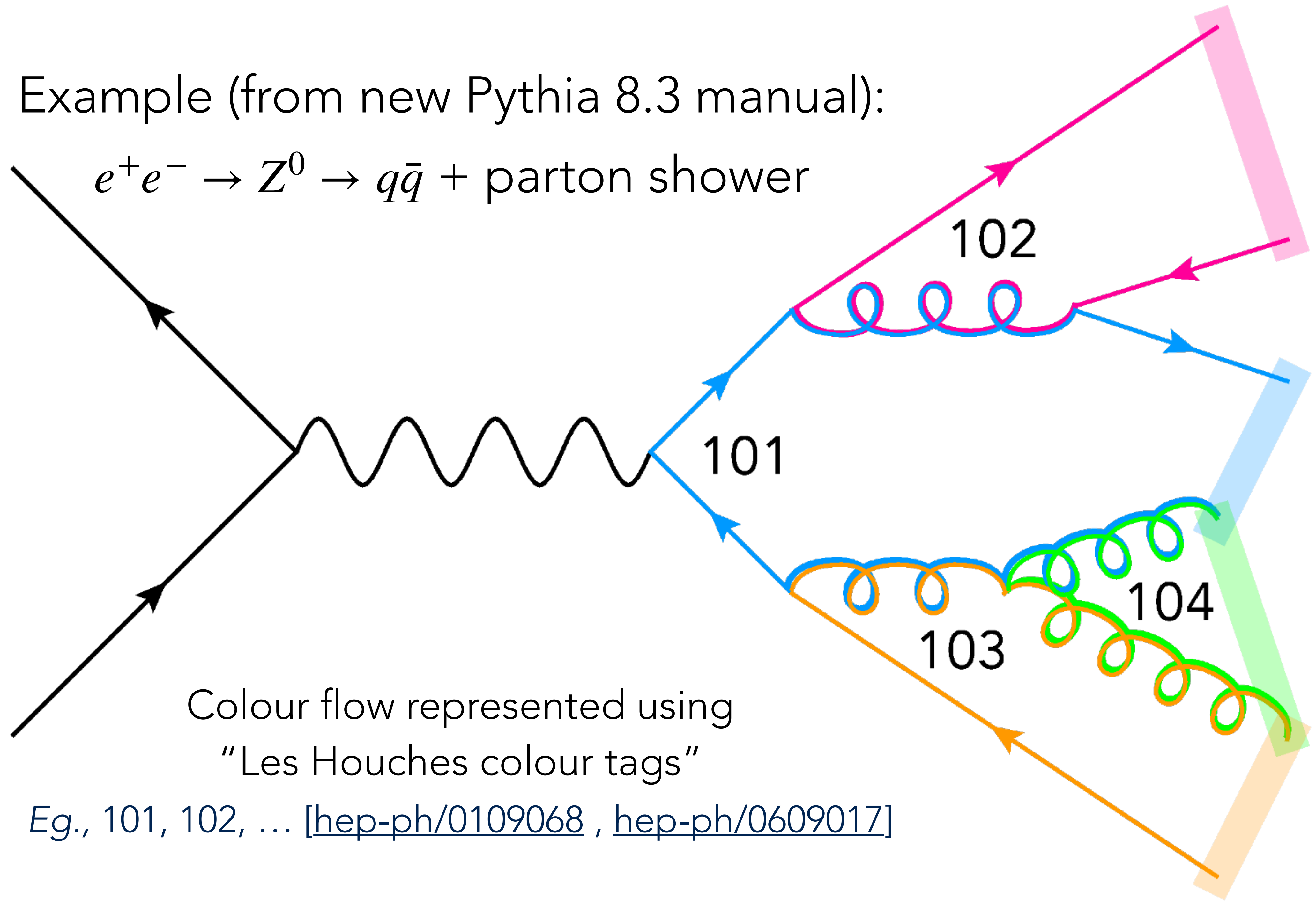
$\Longrightarrow$ Each colour appears only once & is matched by a unique anticolour.

In e^+e^- collisions (LEP):

- Corrections to the Leading-Colour picture suppressed by $1/N_C^2 \sim 10\,\%$
- Also: coherence $\Longrightarrow$ not much overlap in phase space (except in WW → 4q)

Example (from new Pythia 8.3 manual):

$e^+e^- \to Z^0 \to q\bar{q}$ + parton shower



Colour flow represented using "Les Houches colour tags"

Eg., 101, 102, … [hep-ph/0109068 , hep-ph/0609017]

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

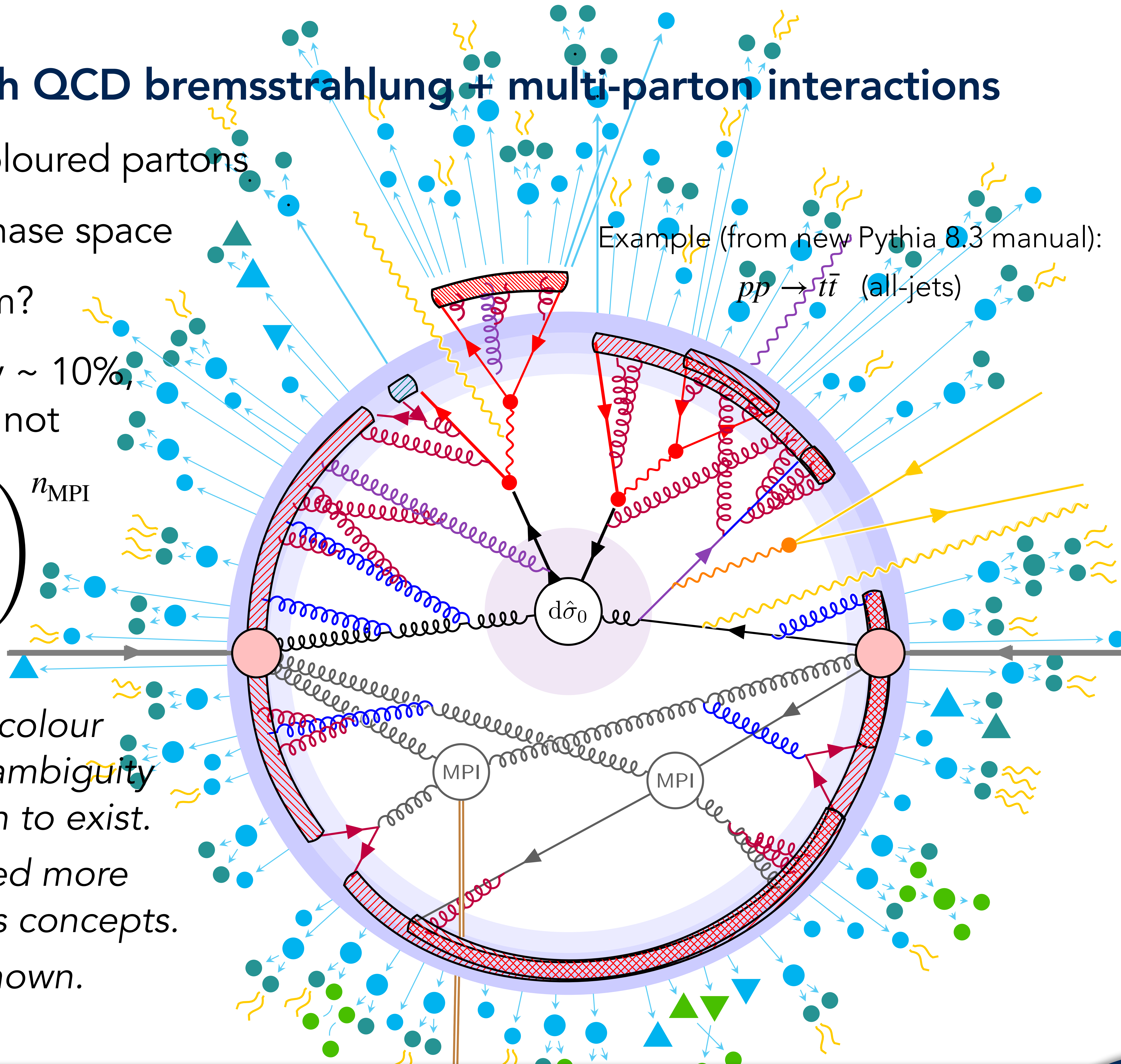
- Final states with very many coloured partons
- With significant overlaps in phase space
- Who gets confined with whom?
- If each has a colour ambiguity ~ 10%, CR becomes more likely than not

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

Note: in this context, the word "colour reconnections" simply refers to an ambiguity beyond Leading N_C, which is known to exist.

But the term "CR" can also be used more broadly to incorporate further physics concepts.

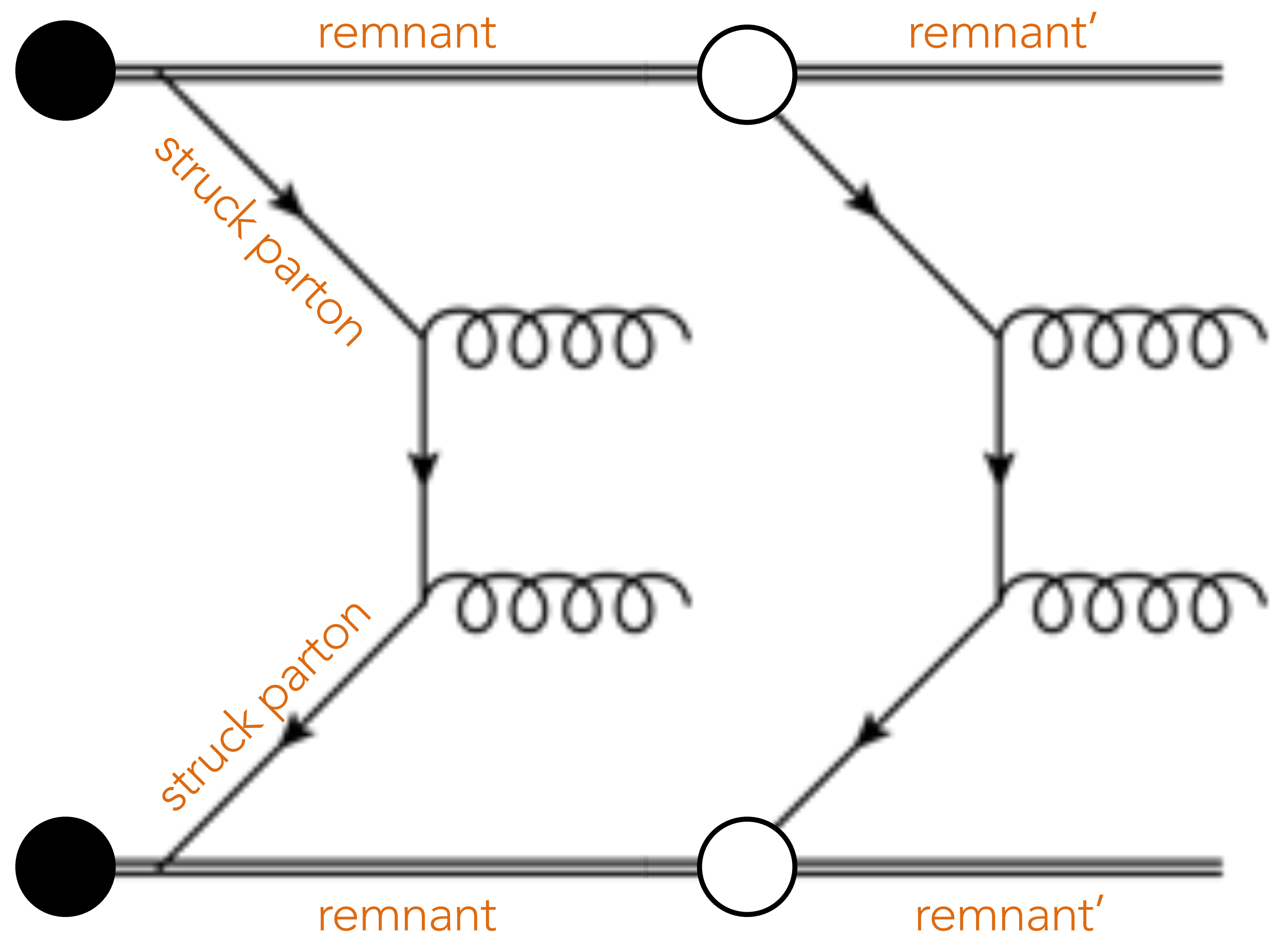
Detailed physics not yet fully known.



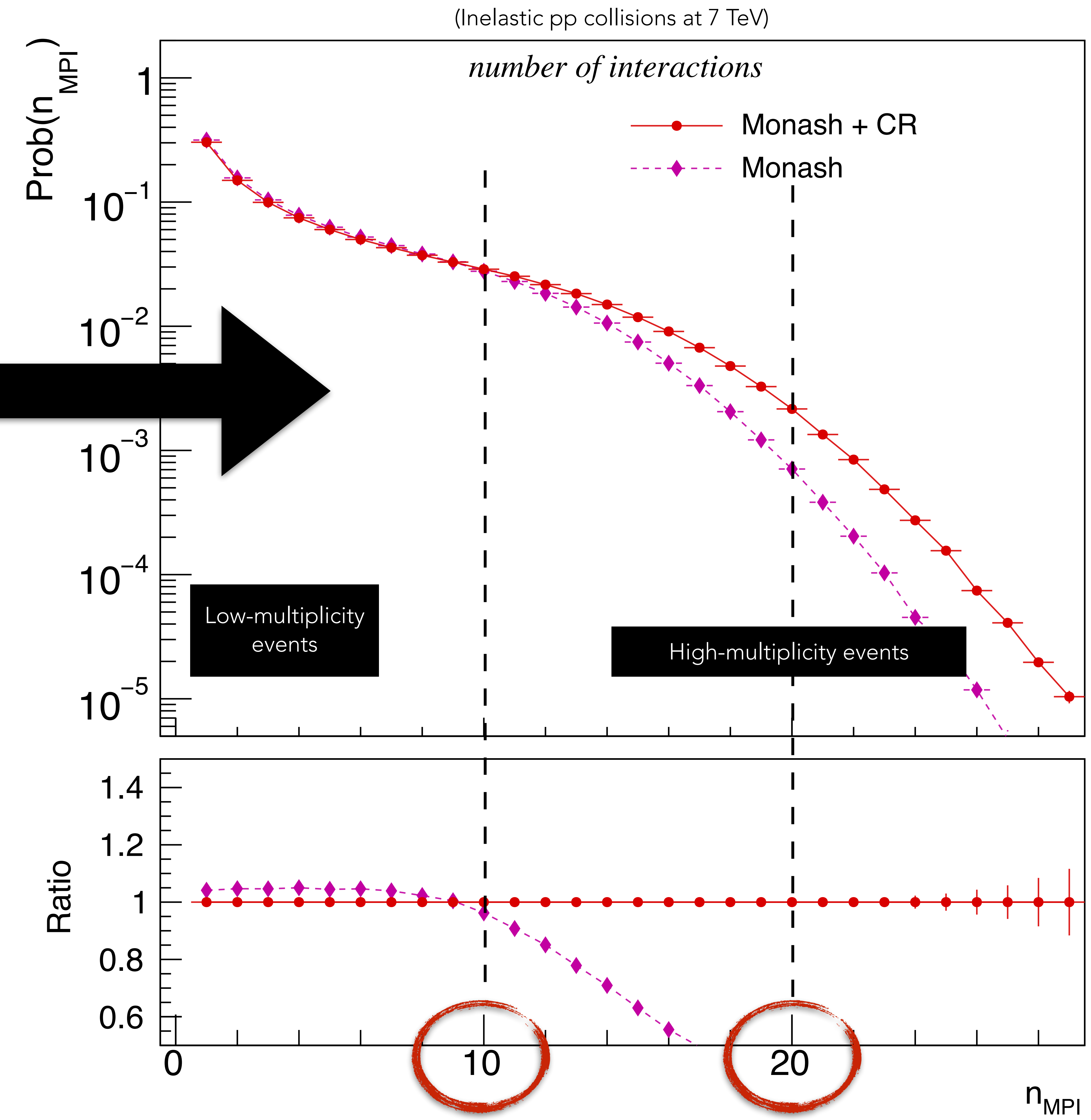
How many MPI are we talking about?

How many parton-parton systems are there in pp collisions? DPS? 3PS? ...?

Multi-Parton Interactions (MPI)



+ ...



$\Longrightarrow$ can have **very** many parton systems within a single pp collision (esp. in high-multiplicity events)

All within ~ transverse size of a proton (= right on top of each other)

Unique feature of SU(3): Y-Shaped 3-String "Junctions" ➤ Baryons

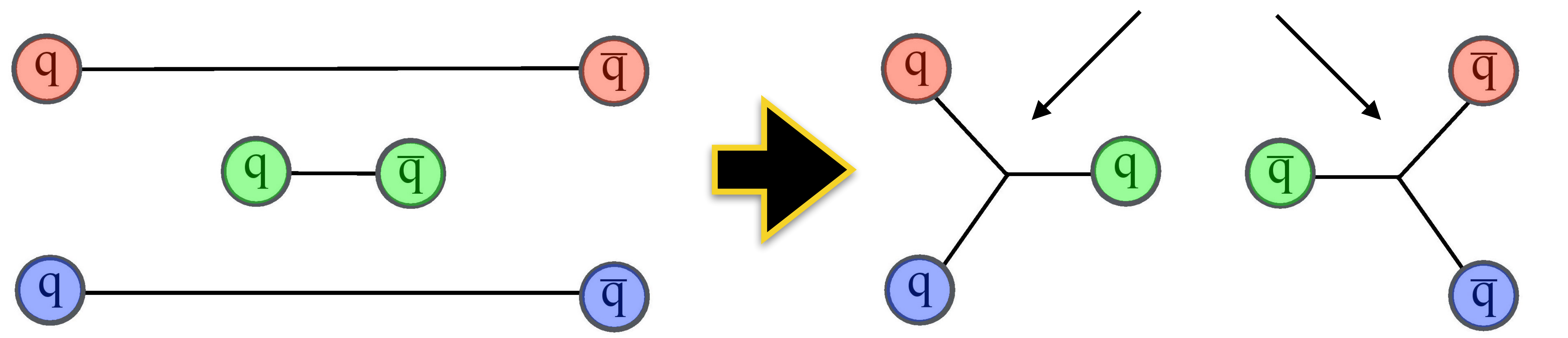
[Baryon Number Violation & String Topologies: Sjöstrand & PS hep-ph/0212264]

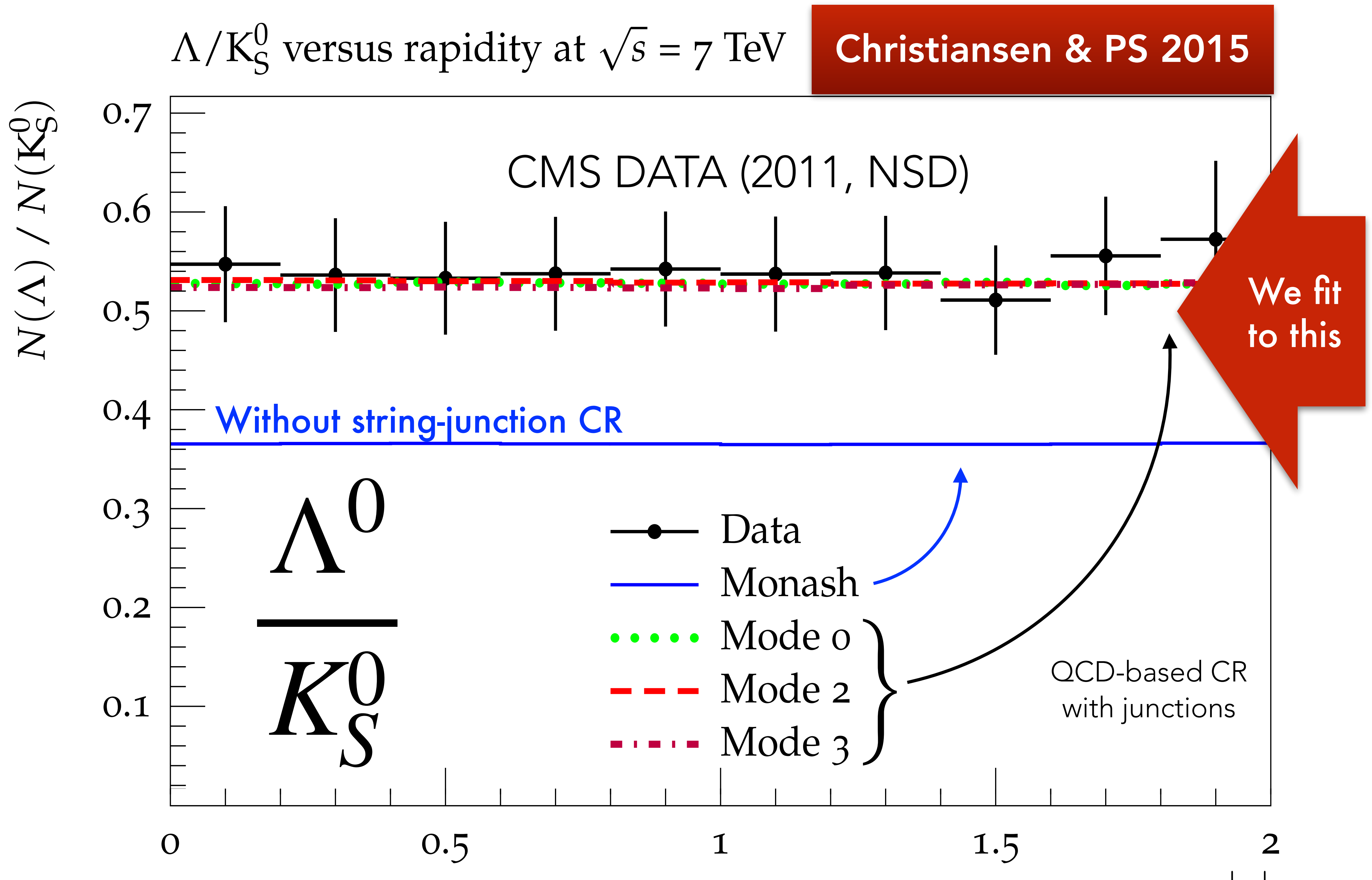
"Colour reconnection" modelling based on stochastic sampling of SU(3) group probabilities: allows for random (re)connections

[String Formation Beyond Leading Colour: Christiansen & PS 1505.01681]

For example:

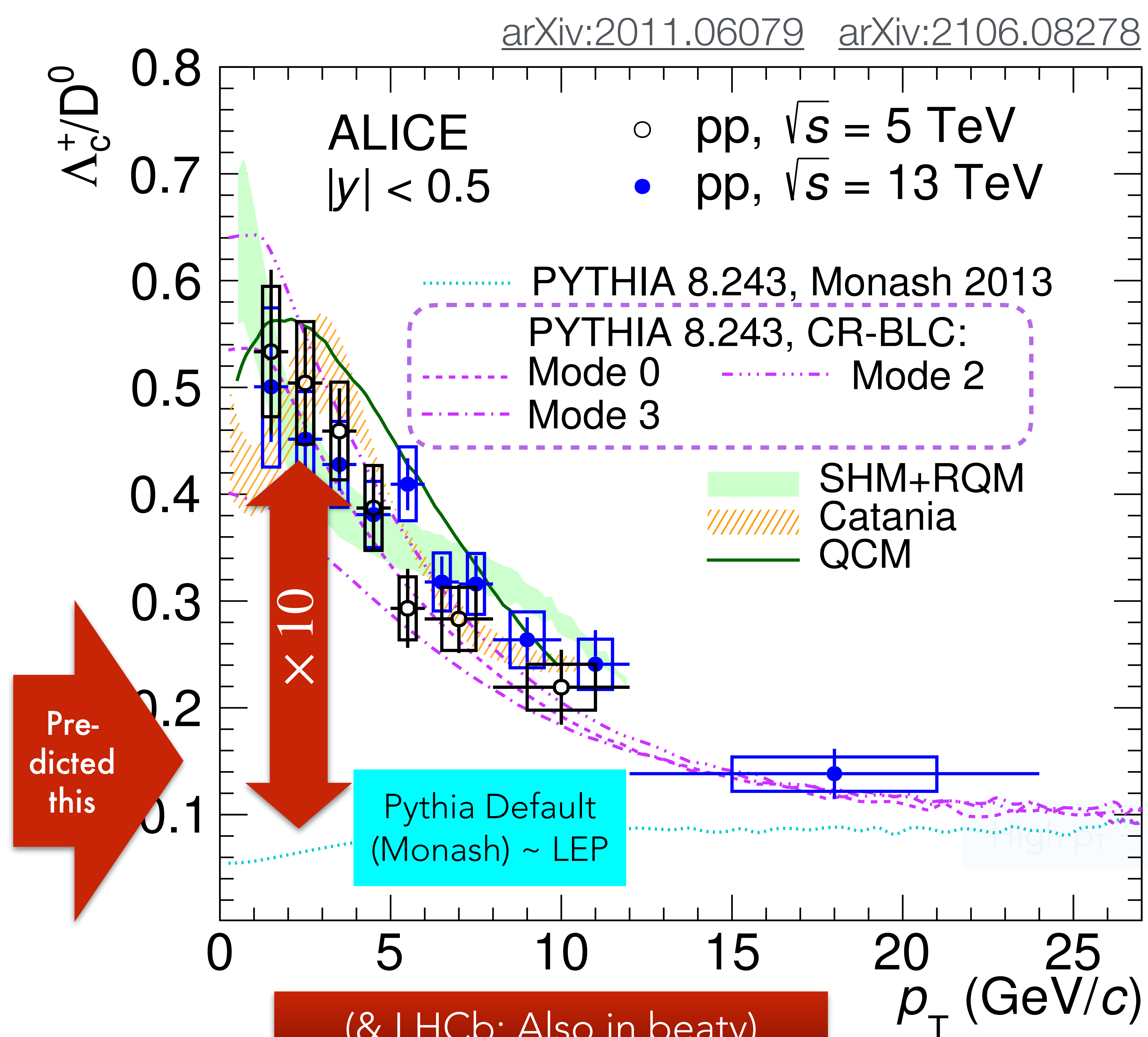
Extra baryon-antibaryon production





Mode 0, 2, 3 are different causality restrictions (0 = none)

ALICE 2021: also in charm



(& LHCb: Also in beaty)

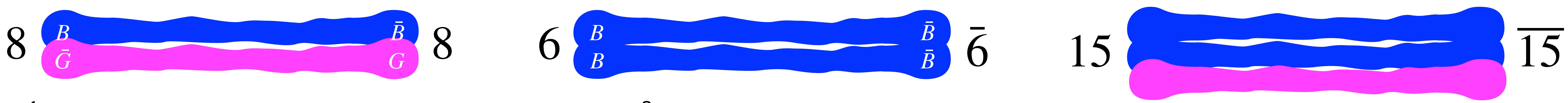
4. **Strangeness**, Ropes, and (Advanced) Close-Packing

Clear observations of strangeness enhancements in high-multiplicity pp collisions (relative to LEP and low-multiplicity pp) [e.g., ALICE Nature Phys. 13, 535 (2017)]

- Much activity to understand *dynamics* of effective breakdown of strangeness universality

In string context, MPI + Colour Ropes [e.g., Bierlich et al. 1412.6259] have been proposed:

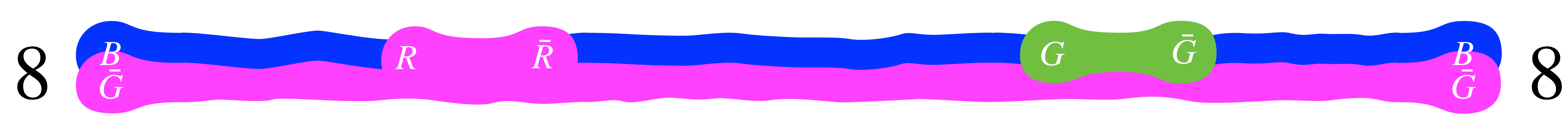
- Casimir scaling of effective string tension $\Longrightarrow$ less strangeness suppression in string breaks



p=1, q=1: $C_8 = C_A = 2.25\, C_F$ p=2, q=0: $C_6 = 2.5\, C_F$ p=2, q=1: $C_{15} = 4\, C_F$

Simplified alternative: Close-Packing [Fischer, Sjöstrand 1610.09818] string tension scales with effective background ∝ n_{MPI} (global) or $n_{strings}$ (local)

- Local version updated with Monash student J. Altmann to account for **directional colour flows** (p and q), **junction topologies**, and effective **diquark suppression** in octet-type fields (“*Altmann mechanism*”):



“Popcorn picture” in which diquark formation is viewed as a fluctuation of first one colour followed by another of a different colour

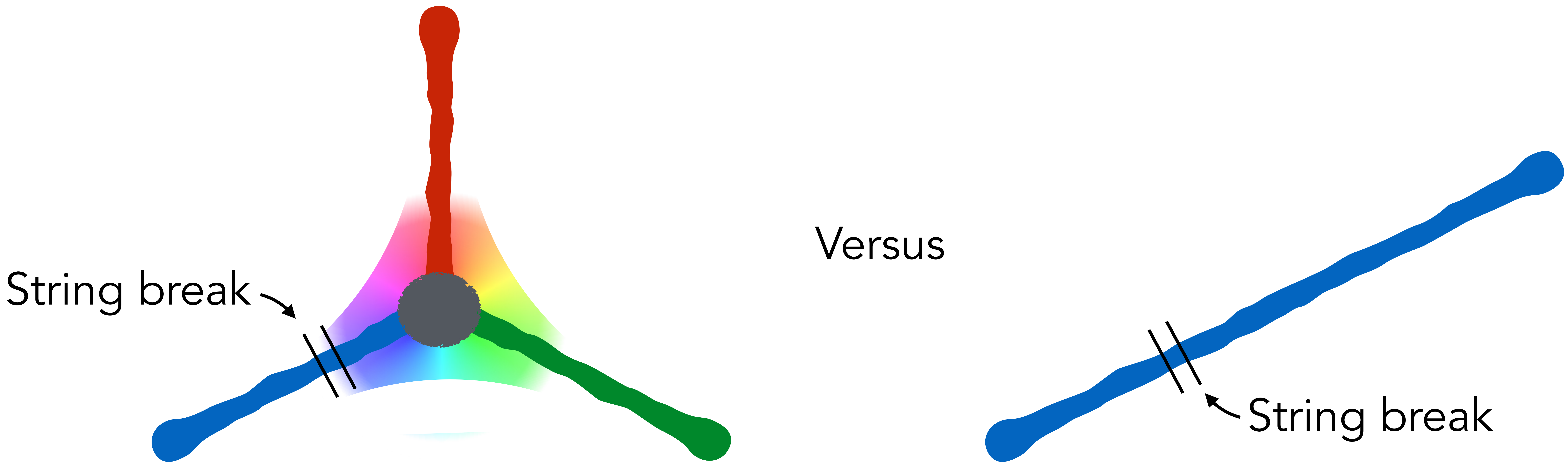
$R\bar{R}$ (or $\bar{R}R$) fluctuation increases tension from C_8 to C_6

$G\bar{G}$ (or $\bar{G}G$) fluctuation Can just break the other string

New: Strange Junctions

What do we really know about the field strength near a QCD junction?

- Probably related to baryon spectroscopy / lattice, but unaware of any specific answers



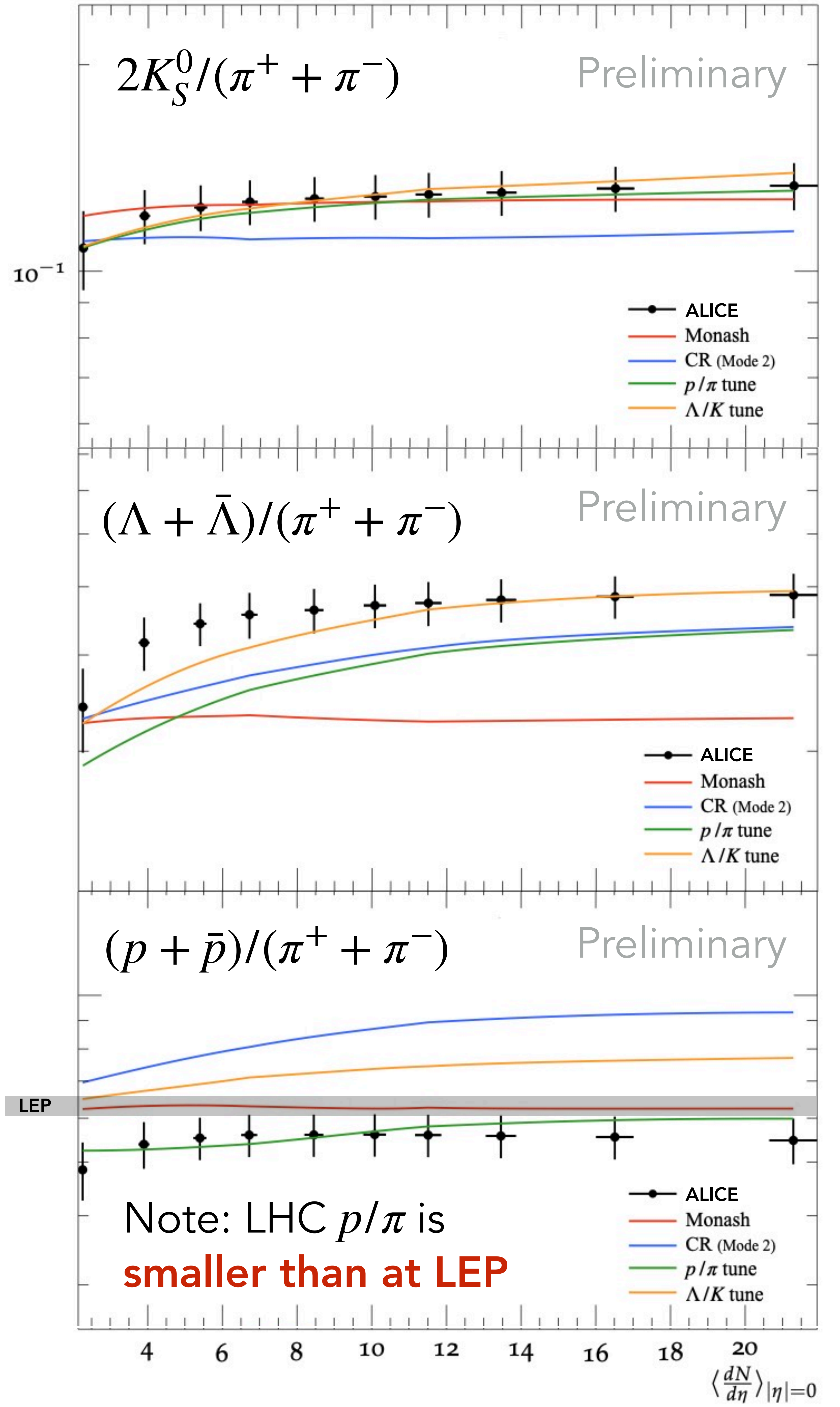
Effective energy density per unit length could be different from vacuum case near a junction?

Enhanced string tension on the string breaks closest to junction?

→ Model of **"strange junctions"** (with Monash PhD student Javira Altmann)

Mechanism for strangeness enhancement specifically for junction baryons

QCD CR + Advanced Close-Packing: First Results



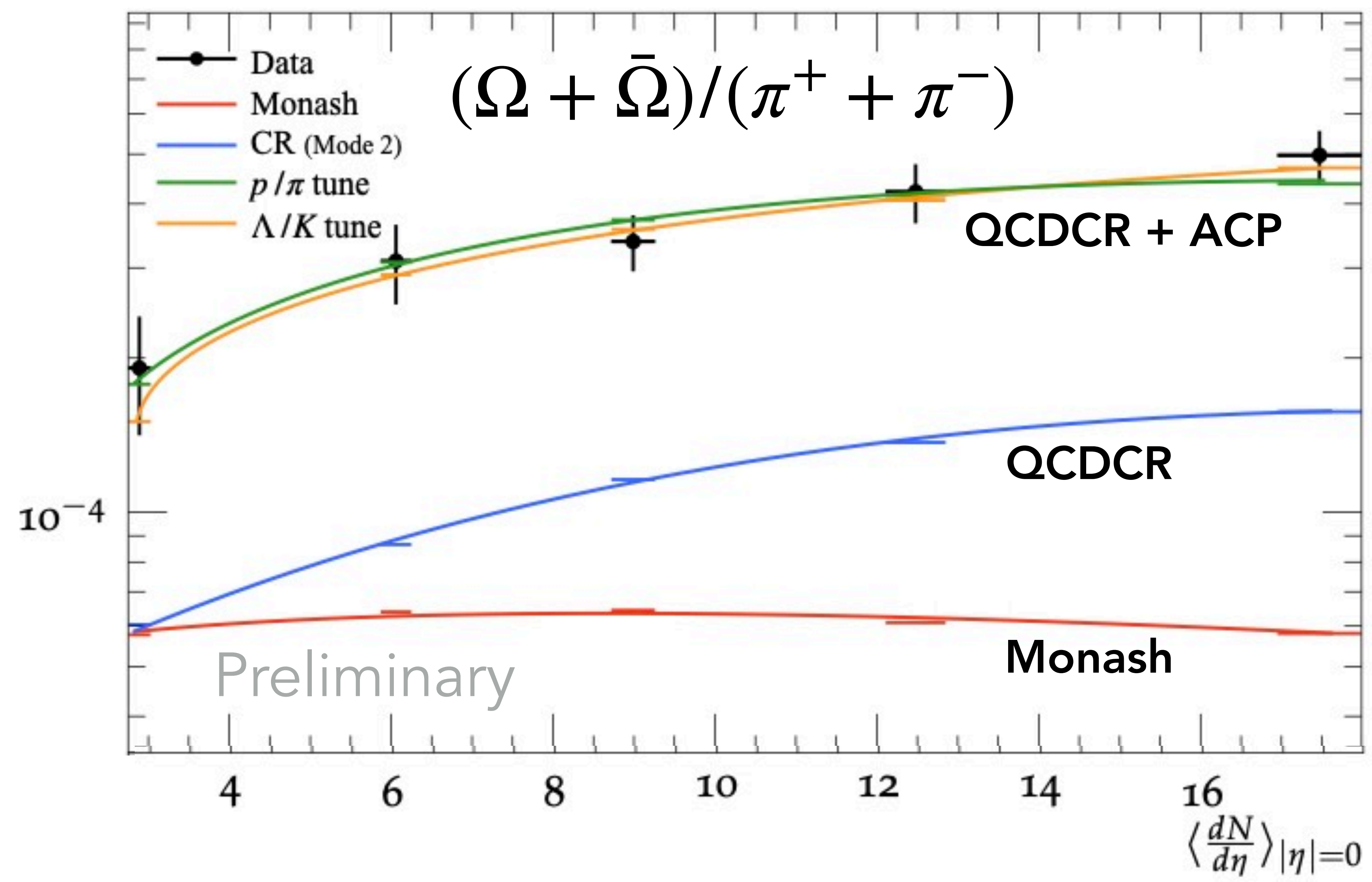
Monash (no QCD CR, no close-packing) ~ LEP

QCD CR (mode 2); no close-packing

QCD CR + ACP: p/π tune

QCD CR + ACP: Λ/K tune

		p/π	Λ/K	
`StringPT:tension`	=	0.05	0.11	← Close-packing
`StringPT:qqFac`	=	0.7	0.23	← Altmann Mechanism
`StringFlav:strangeJuncFactor`	=	0.65	0.55	← Junction Strangenesss



Being finalised now, with publication on the way.

J. Altmann, PS

5. If there is time …

Cosmic-Ray Air Showers

Single incident particle ➡ billions of final-state particles (forget about GEANT).
Recently started a collaboration with CORSIKA 8 fast/optimised air-shower tracker

New: PythiaCR [Based on Sjöstrand + Utheim, **2005.05658** & **2108.03481**]

- Provide **hadron-air cross sections** ⊕ **perform collisions** ⊕ simulate **hadron decays**
  (Air ~ ^{14}N + ^{16}O; currently also ^{40}Ar, ^{208}Pb; few hours of manual labour to add more)
- Cosmic-ray “beams” are **heterogenous** and **not mono-energetic**:
  Achieved by initialising multiple beams in energy grids + rapid beam switching
- CR (re-)interactions “fixed-target”; can probe **low CM energies** (by HEP standards)
  Standard (collider) Pythia only applies for $\sqrt{s} > 10\,\mathrm{GeV}$
  New **extensive low-energy (re)interaction models**
  ➡ **Arbitrary hadron-hadron collisions at low E, and arbitrary hadron-p/n at any energy)**
  Extend to **hadron-nucleus** using nuclear-geometry part of ANGANTYR

So far limited comparisons with data - interested in feedback

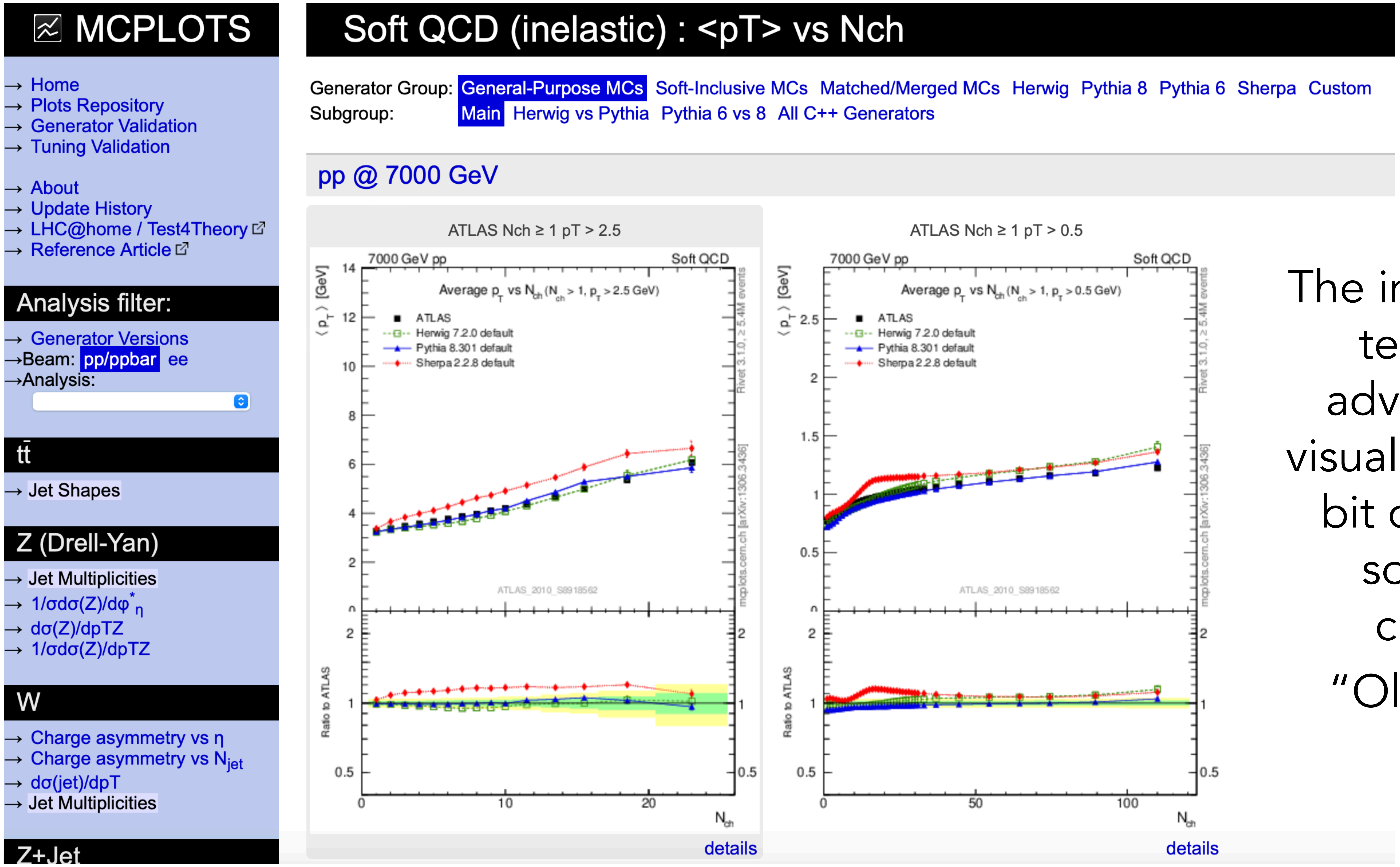
- A positive technical note: native C++ simplifies CORSIKA 8 - PYTHIA 8 interfacing

See also M. Reininghaus et al. *Pythia 8 as hadronic interaction model in air shower simulations*, **2303.02792**

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

mcplots.cern.ch started in 2010, as browsable repository of MC validations (via Rivet)

- Running continuously on ~ 1000 cores donated by BOINC LHC@home volunteers (+ Grid backfill)



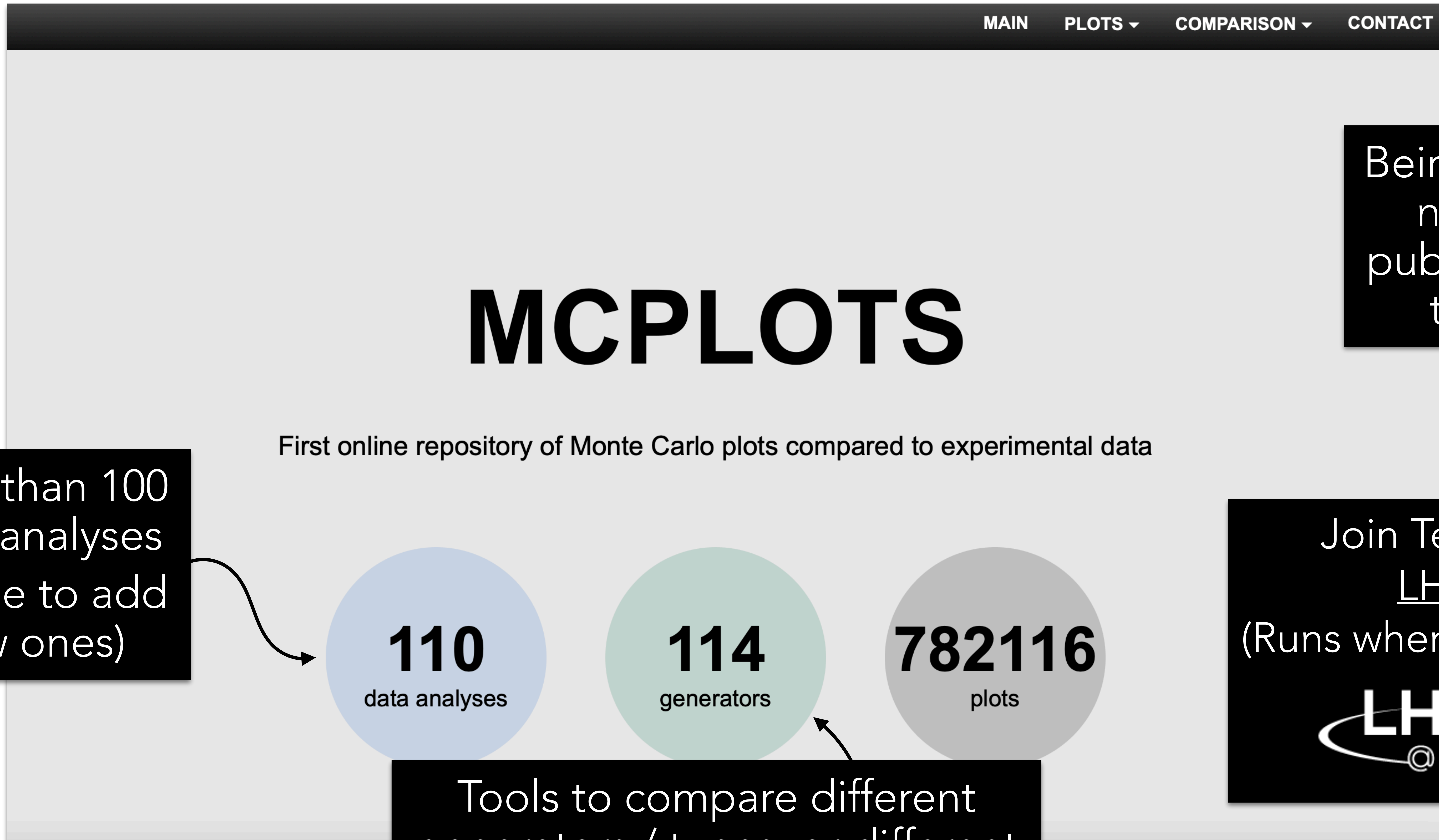
The interface was technically advanced but visually perhaps a bit dated, and somewhat cluttered

"Old School"

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

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

- Mainly driven by Natalia Korneeva, 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)

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

Volunteer computing for the LHC

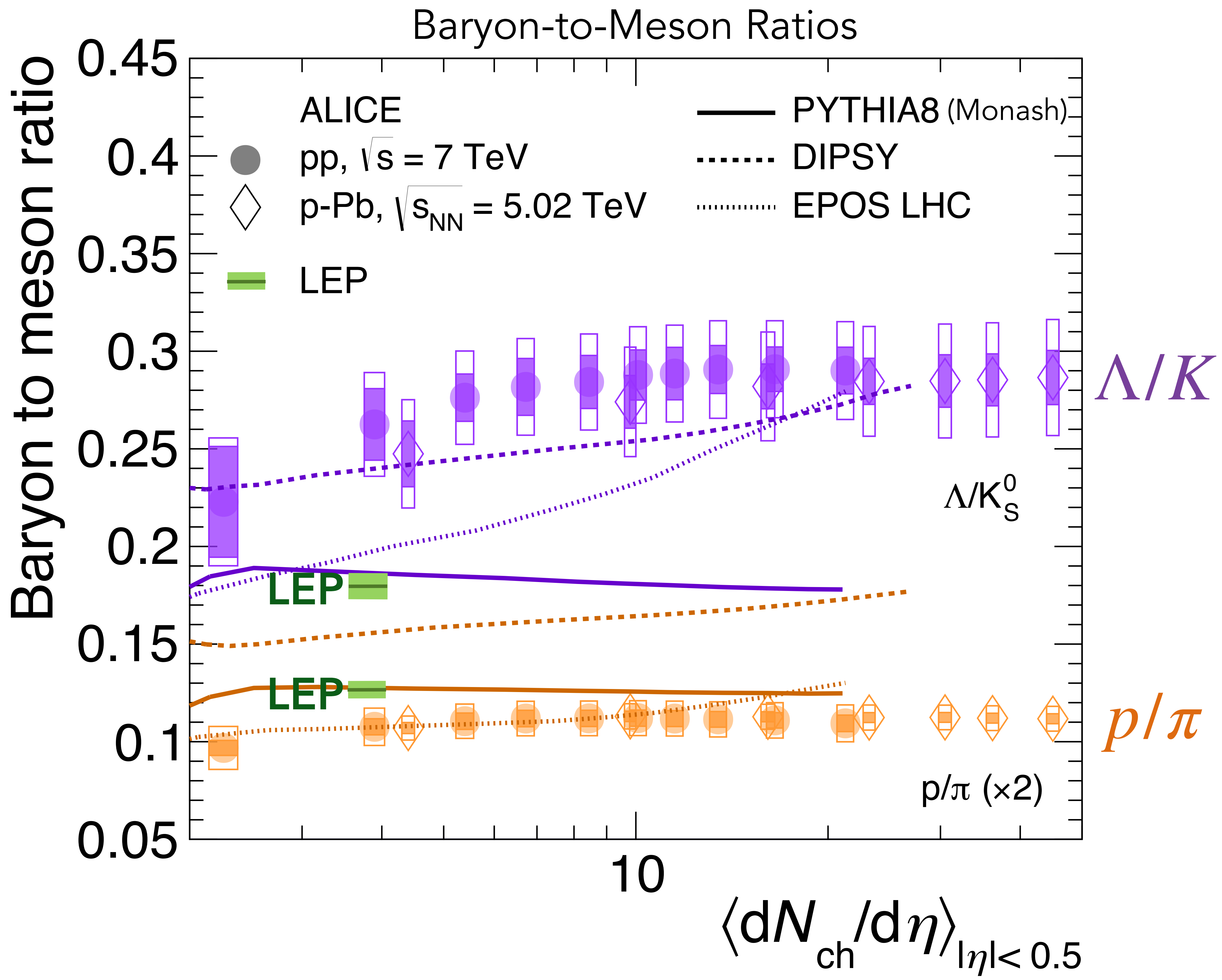
Tools to compare different generators / tunes, or different versions of same generator

Extra Slides

An Achilles' Heel? **Protons!**

So far, physics models have focused heavily on strangeness

- The original ALICE paper from 2017 also included the proton/pion ratio
- In many model setups, enhancement of strangeness is accompanied by more heavier states in general $\Longrightarrow$ **non-strange baryons also enhanced**
- Also, QCD CR model acts in colour space; junction structures are flavour-blind



Data shows that the p/π ratio at LHC is a bit *smaller* than at LEP!

With ~ no evolution with N_{ch}

Protons are the most abundant baryons!

EPOS captures this behaviour

(what about @ LEP?)

From a CR perspective, baryon enhancement appears *very* correlated with strangeness …