Exploring transverse momentum and pseudorapidity distributions of high-energy *pp* collisions with CMS Open Data

Gabriel O. Corrêa *, Arthur M. Moraes
Centro Brasileiro de Pesquisas Físicas
*gabriel.correa@cbpf.br

Introduction

Measurements of transverse momentum (p_T) and pseudorapidity (η) distributions contribute to our understanding of the hadron production in high-energy collisions and operate as a solid foundation to implement future analysis. The charged particle transverse momentum distribution also provides tests of the predictability of quantum chromodynamics. While we can describe the high-p_T distribution using perturbative quantum chromodynamics, we also need to use phenomenological methods to depict the whole spectrum. Non-extensive statistical mechanics [1] supplies one of those methods, where it provides a parametrization that fits the power-law behavior for high-p_T as well as the exponential behavior for low-p_T. Questions arise as to why a non-extensive statistical mechanical distribution successfully describes the experimental data and what are its possible physical implications. Using collision data collected by the CMS Collaboration [2] and made available through the CERN Open Data Portal [3], we reproduce p_T and η distributions of charged hadrons for *pp* collisions with $\sqrt{s} = 0.9$, 2.76, and 7 TeV. The spectra obtained were then compared to those published by the CMS Collaboration. The data were fit using the Tsallis distribution.

The CMS Open Data Portal

In November 2014, the CERN collaborations solidified their commitment to openness with the first release of LHC data to the open community. Since then, there were multiple releases from the CERN community, mainly from the CMS Collaboration. Focusing on the CMS Collaboration's released data, it is possible to access almost all of the collected data from proton-proton collisions from 2010 to 2012, with the center-of-mass energy of the collisions ranging from 0.9 TeV to 8 TeV. Among other things, the CMS Open Data provides:

- collision data
- simulated data
- documentation
- Access to the CMSSW (the collection of software needed by the simulation, reconstruction, calibration, and alignment modules)

The collided (and simulated) data provides the information detected by (or simulated through) the CMS detector, among them are

- the physical objects (electrons, muons, photons, etc) and its properties,
  - in our case, tracks, vertices, and calorimeter hits;
- trigger information (prescales and execution information);
- event information (luminosity, etc).

References

[1] C. Tsallis, Journal of Statistical Physics **52**, 479 (1988).
[2] CMS Collaboration, JINST **3**, S08004 (2008).
[3] CERN Open Data Portal, `http://opendata.cern.ch`
[4] P. D. B. Collins and A. D. Martin, Rep. Prog. Phys. **45**, 335 (1982).
[5] R. Hagedorn and J. Rafelski, Phys. Lett. B, **97**, 136 (1980).
[6] C. Wong *et al*, Phys. Rev. D **91**, 114027 (2015).
[7] CMS Collaboration (2019). MinimumBias primary dataset in RECO format from the 0.9 TeV Commissioning run of 2010. DOI:10.7483/OPENDATA.CMS.1R58.OMBD
[8] CMS Collaboration (2020). AllPhysics2760 primary dataset in RECO format from the 2.76 TeV proton-proton run of 2011. DOI:10.7483/OPENDATA.V439.NXZC
[9] CMS Collaboration (2014). MinimumBias primary dataset in AOD format from RunB of 2010. DOI:10.7483/OPENDATA.CMS.6BPY.XFRQ
[10] CMS Collaboration, J. High Energy Phys. **02**, 041 (2010).
[11] CMS Collaboration, Phys. Rev. Let. **105**, 0220022 (2010).
[12] CMS Collaboration, J. High Energy Phys. **08**, 086 (2011).

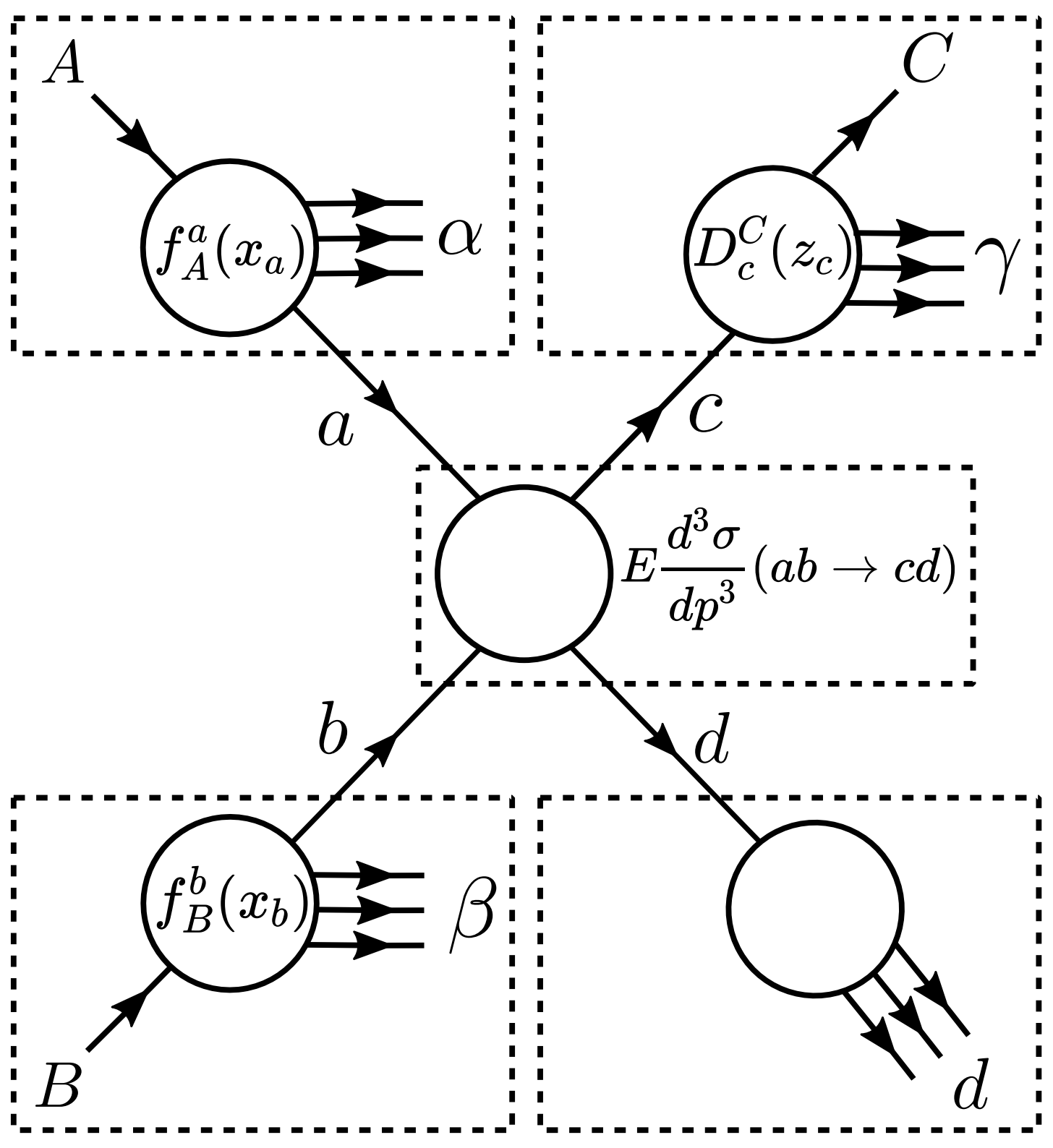
Hadron collisions: the pQCD approach

A perturbative expansion of a QCD process results in higher-order terms of the strong coupling constant (α_s). α_s is a function of the exchanged particle's energy Q, where α_s decreases with its increase [4]:

$$\alpha_s(Q^2) \approx \frac{1}{\frac{b_0}{4\pi} \ln\left(\frac{Q^2}{\Lambda^2}\right)}.$$

This property is called asymptotic freedom. Therefore, only high-p_T can be described by pQCD. The pQCD model describes a high-p_T process where the structure functions $f(x)$, the fragmentation functions $D(z)$, and the partonic subprocess a+b→c+d are known. This model predicts that the high-p_T yield behaves as a power law:

$$E\frac{d^3N}{dp^3} \propto \frac{1}{p_T^n}.$$



Hadron collisions: the phenomenological approach

Phenomenological methods to explain the hadron production use a statistical approach. Hagedorn proposed an empirical formula to describe the hadron production [5]

$$E\frac{d^3N}{dp^3} = C\left(1 + \frac{p_T}{p_0}\right)^{-n},$$

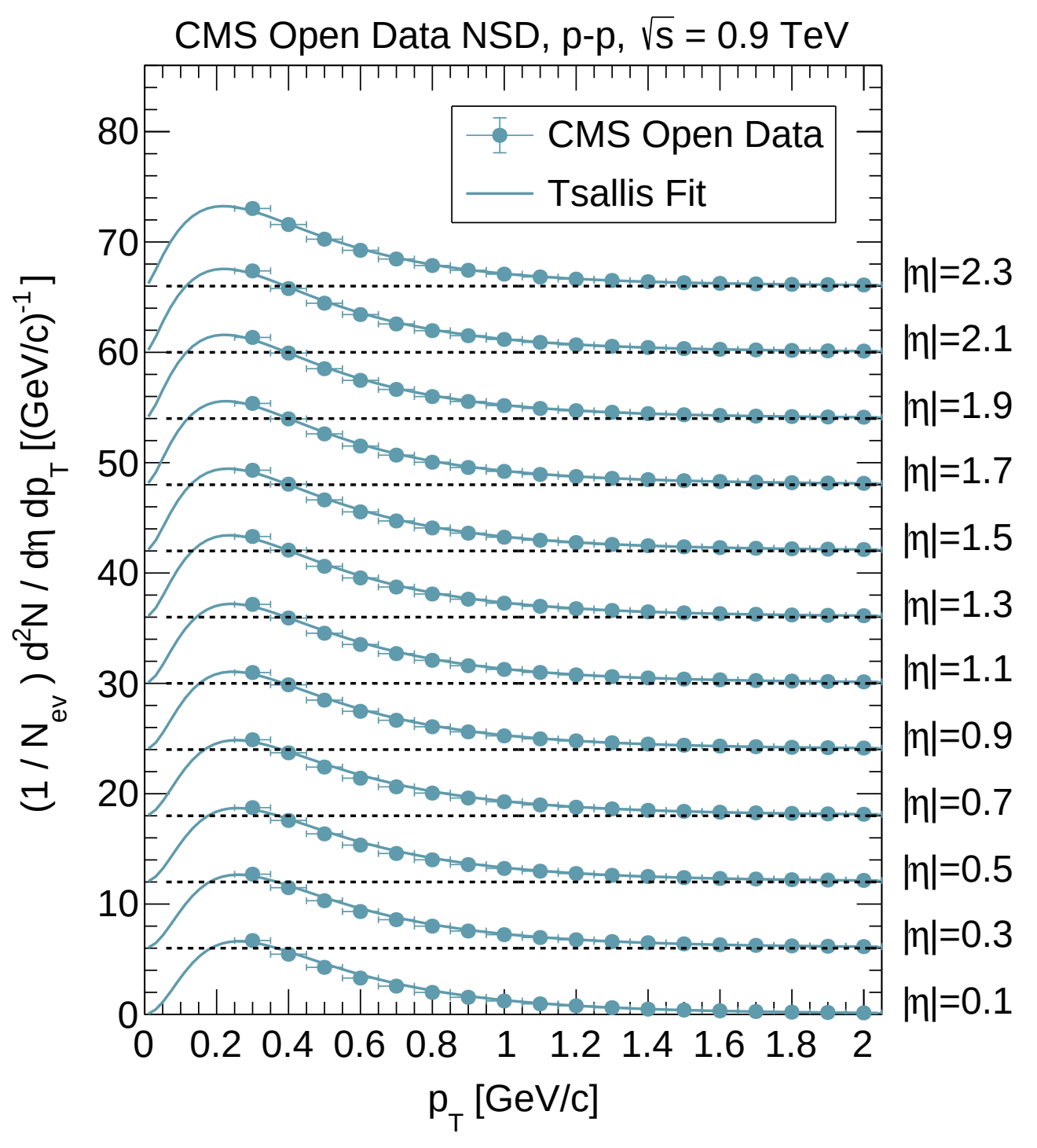
where p_0 and n are fitting parameters. For $p_T \to 0$ we have an exponential behavior, and for $p_T \to \infty$ the yield behaves as a power law.

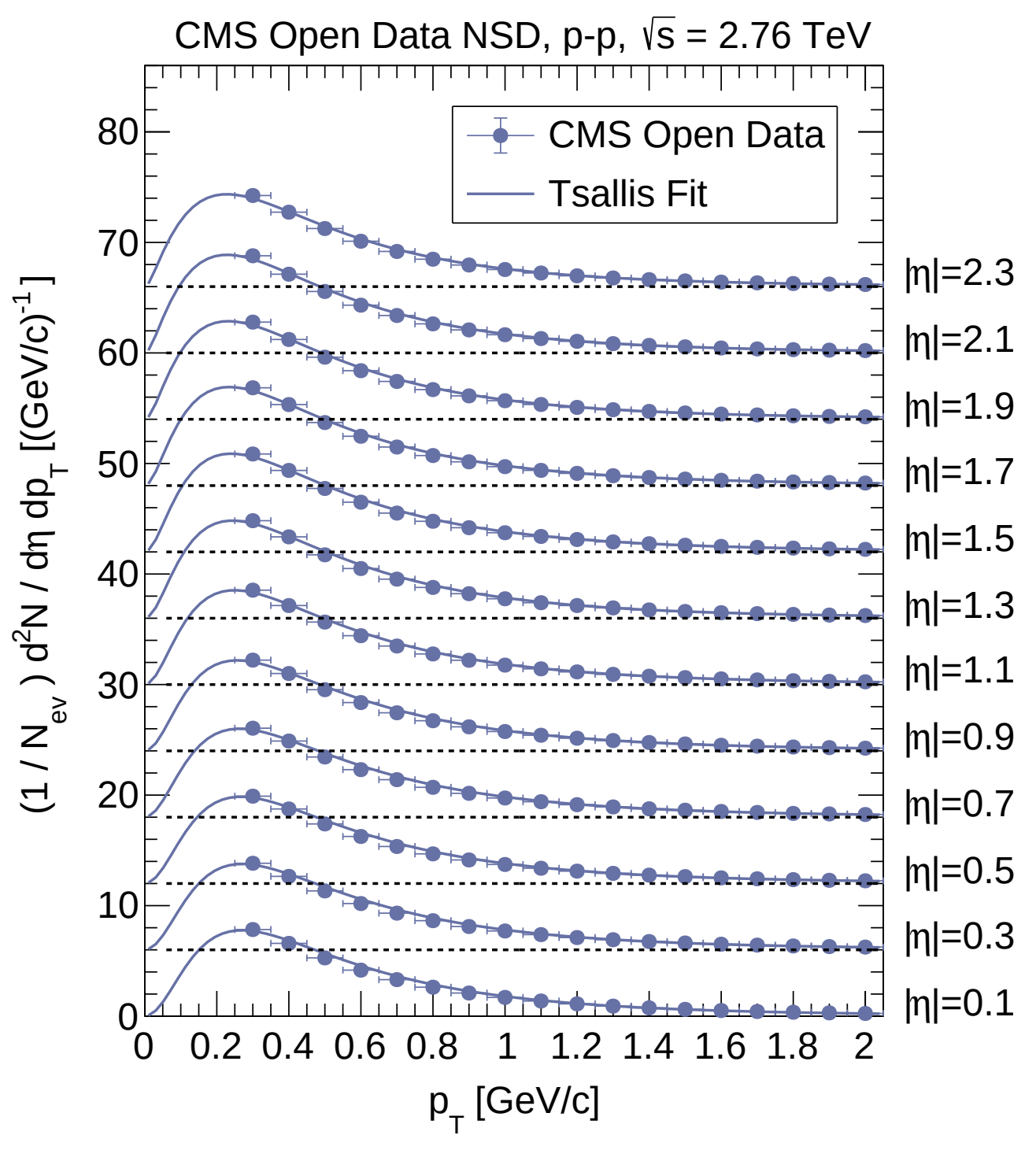
Another way to describe the hadron production data is through the nonextensive statistical mechanics. This approach is a nonextensive generalization of the BG statistics. This method gives a Tsallis distribution given by [6]

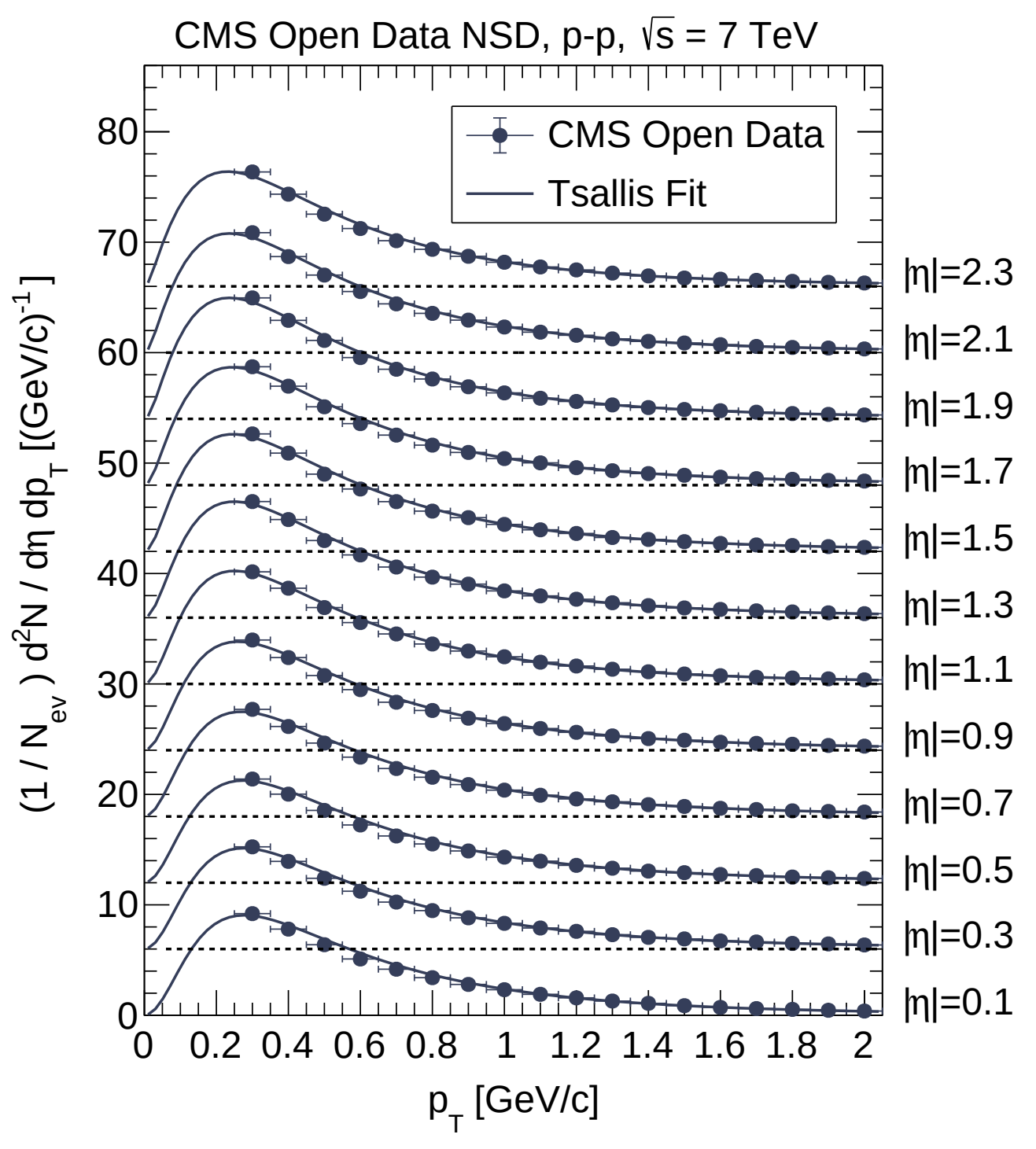
$$E\frac{d^3N}{dp^3} = A\left[1 - (1-q)\frac{E_T}{T}\right]^{\frac{1}{(1-q)}},$$

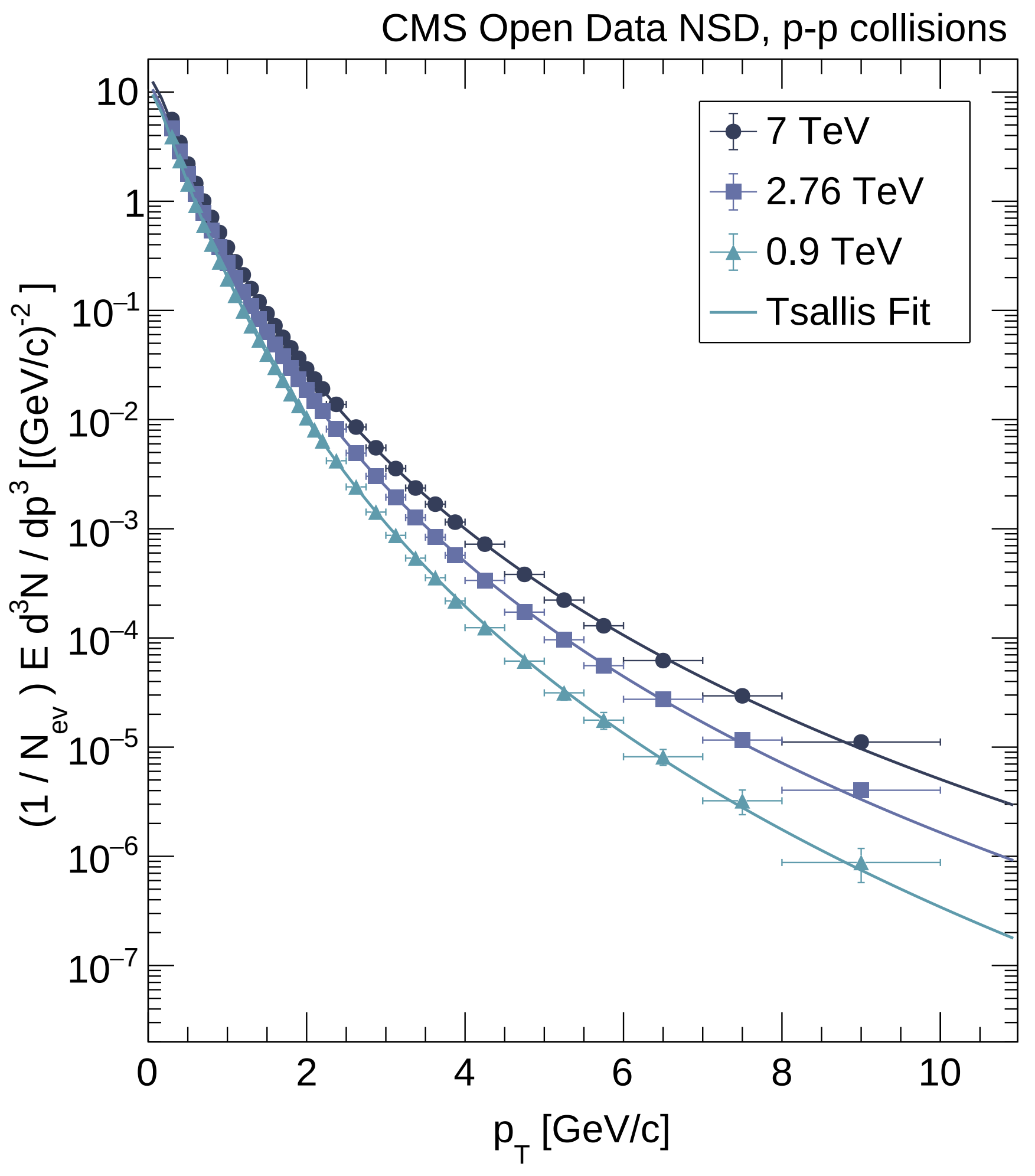
where $E_T = \sqrt{m^2 + p_T^2} - m$. T and q are fitting parameters.

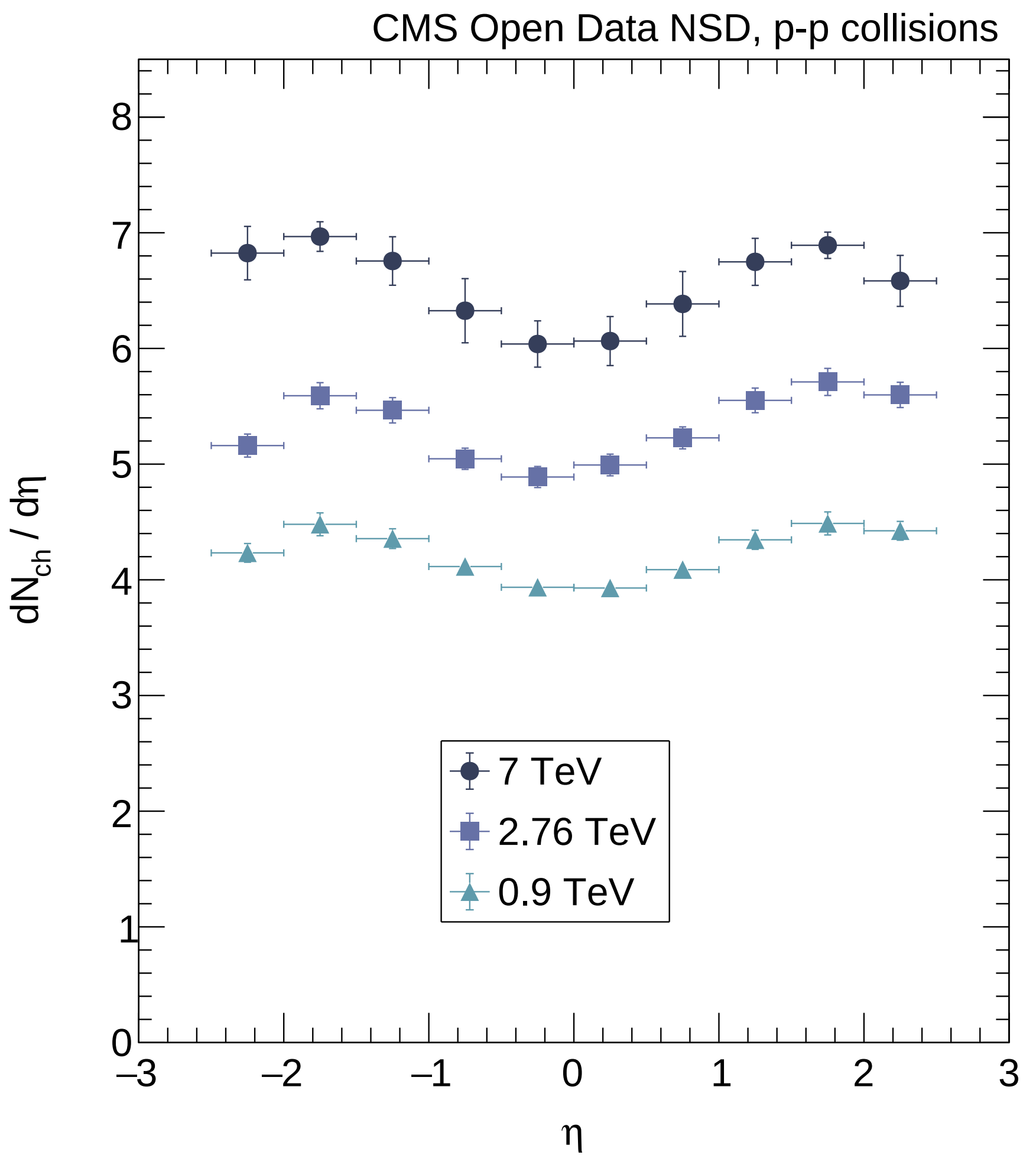
p_T and η distributions for non-single diffractive (NSD) events











- Collision datasets used: [7, 8, 9]
- $\frac{dN}{d\eta}$ higher than published data [10, 11, 12];
- *q* (and *n*) consistent with published data;
- *T* higher than published data;

Outlook: Even though we do not perfectly reproduce the published results, the distributions are close enough to indicate that we are in the right direction. Now, we will look into the fine-tuning necessary to obtain datasets as reliable as we can. After gaining confidence in our distributions, we will look into the parameters obtained through the Tsallis fit and analyze their possible physical implications.