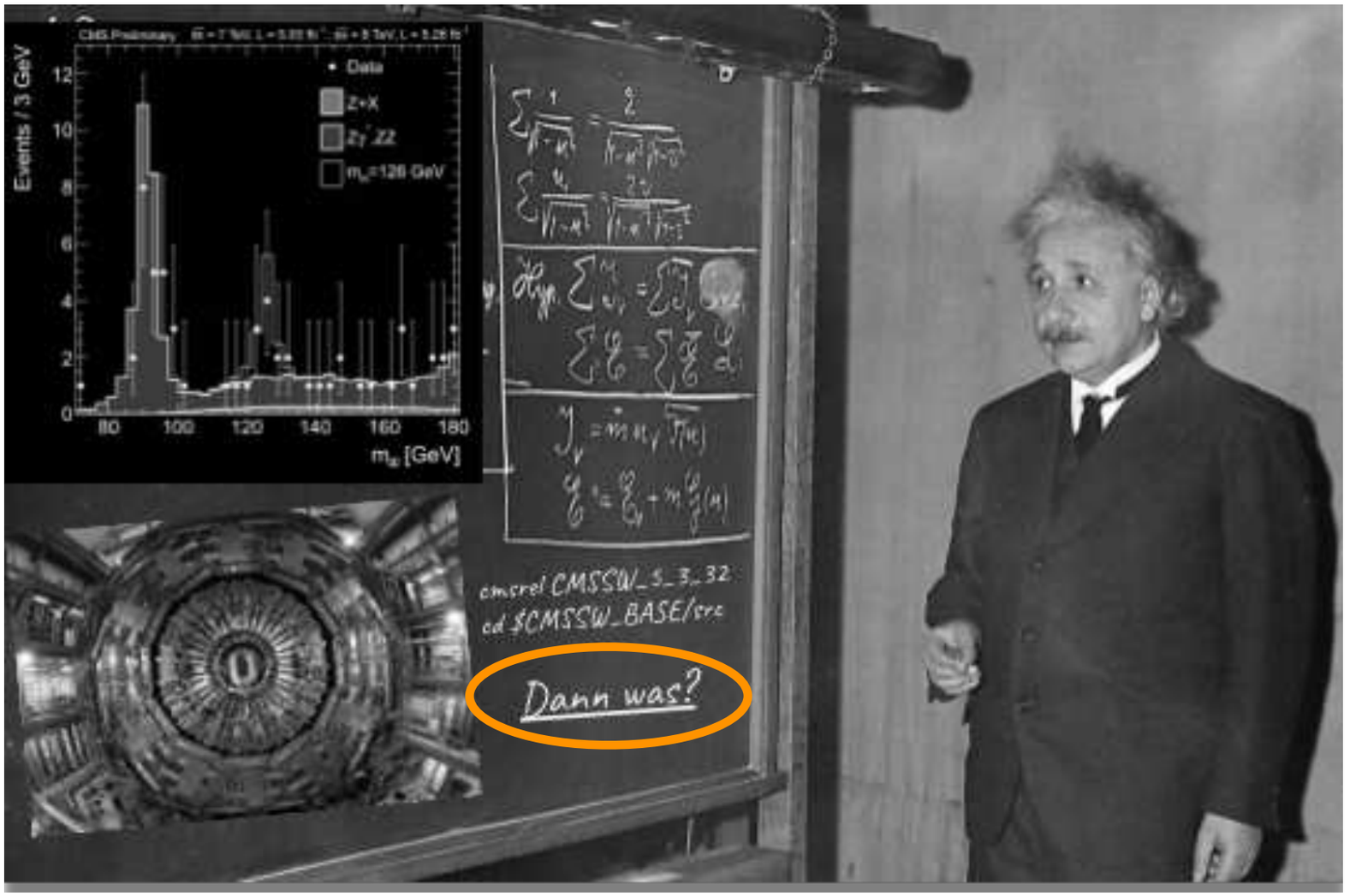
Welcome to the CMS Open Data Workshop for Theorists

Jesse Thaler

Remote Fermilab LPC Workshop — September 30, 2020

CMS Open Data Workshop for Theorists

Remote Fermilab LPC Workshop, September 30 – October 2, 2020



Organizing Committee:
- Matthew Bellis (Siena College)
- Edgar Carrera (U. San Francisco de Quito)
- Kati Lassila-Perini (U. of Helsinki)
- Jesse Thaler (MIT)

Local Organizing Committee:
- Gabriele Benelli (Brown U.)
- Christian Herwig (Fermilab)
- Julie Hogan (Bethel U. and Brown U.)
- Clemens Lange (CERN)
- Andrew Melo (Vanderbilt U.)
- Nada Mohamed (Siena College)
- Stephen Mrenna (Fermilab)
- Kevin Pedro (Fermilab)
- Emanuele Usai (Brown U.)
- David Yu (Brown U.)

LPC Events Committee:
- Gabriele Benelli (Brown U., Co-Chair)
- Kevin Pedro (Fermilab, Co-Chair)

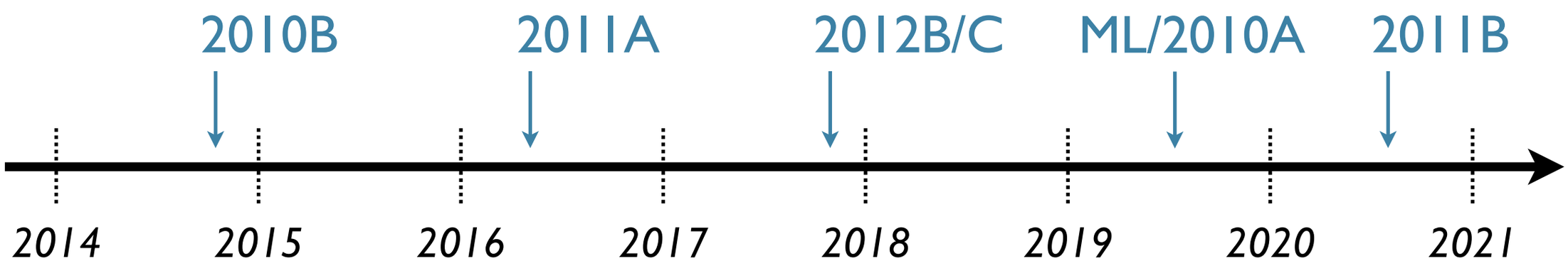
LPC Coordinators:
- Cecilia Gerber (UIC)
- Sergo Jindariani (Fermilab)

Goal: Enable theorists to use real experimental collider data in their research

[CMS Open Data Workshop, Sept-Oct 2020]

CMS is pioneering the release of research-grade public collider data

[First Release, 2014; Second Release, 2016; Third Release, 2017; Fourth Release, 2019; Fifth Release, 2020]



CMS is pioneering the release of research-grade public collider data

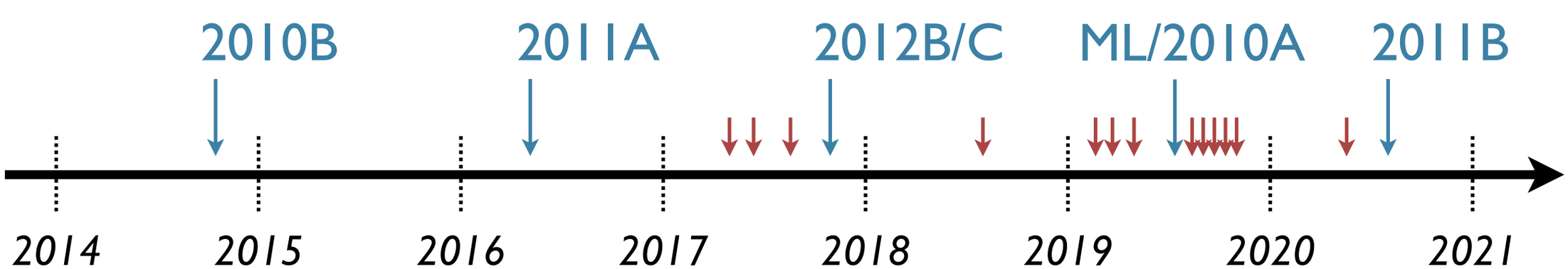
[First Release, 2014; Second Release, 2016; Third Release, 2017; Fourth Release, 2019; Fifth Release, 2020]

THEORY

External (and internal) users are taking advantage of this unique scientific resource

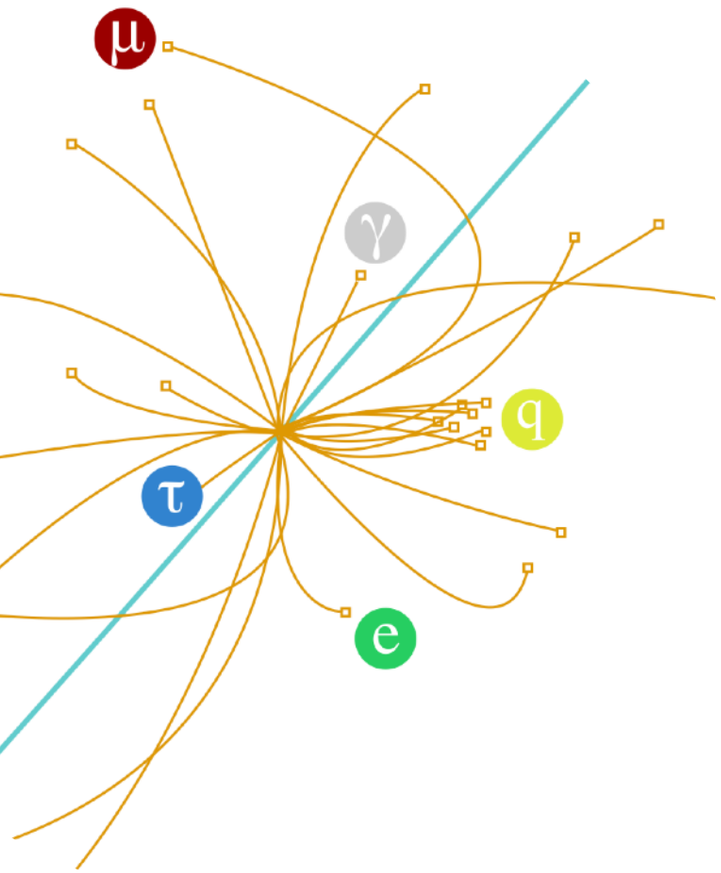
[Tripathee, Xue, Larkoski, Marzani, JDT, PRL 2017, PRD 2017]
[Fernández Madrazo, Heredia Cacha, Lloret Iglesias, Marco de Lucas, EPJWoC 2019]
[Andrews, Paulini, Gleyzer, Poczos, CSBS 2020]
[Cesarotti, Soreq, Strassler, JDT, Xue, PRD 2019]
[Andrews, Alison, An, Bryant, Burkle, Gleyzer, Narain, Paulini, Poczos, Usai, NIM 2020]
[Lester, Schott, JHEP 2019]
[Pata, Spiropulu, arXiv 2019]
[Apyan, Cuozzo, Klute, Saito, Schott, Sintayehu, JINST 2020]
[Paktinat Mehdiabadi, Fahim, JPG 2019]
[Komiske, Mastandrea, Metodiev, Naik, JDT, PRD 2020]
[Moreno, Nguyen, Vlimant, Cerri, Newman, Periwal, Spiropulu, Duarte, Pierini, PRD 2020]
[Knapp, Dissertori, Cerri, Nguyen, Vlimant, Pierini, arXiv 2020]
…

+6 more!



Preview of Today's Colloquium

The CMS Open Data is a fantastic resource, with many exciting applications



Educating future scientists

Stress-testing archival data strategies

Enabling exploratory/proof-of-principle studies

Facilitating dialogue between theory and experiment

Researching physics in and beyond the standard model

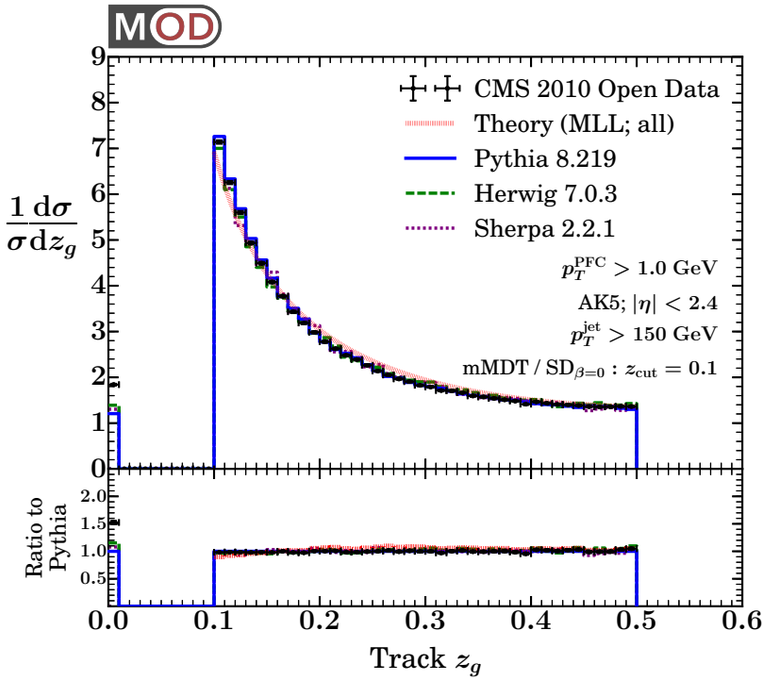
These are only possible with sustained investment in public data initiatives

[JDT, CMS Blog 2017]

Colloquium Backup Slides

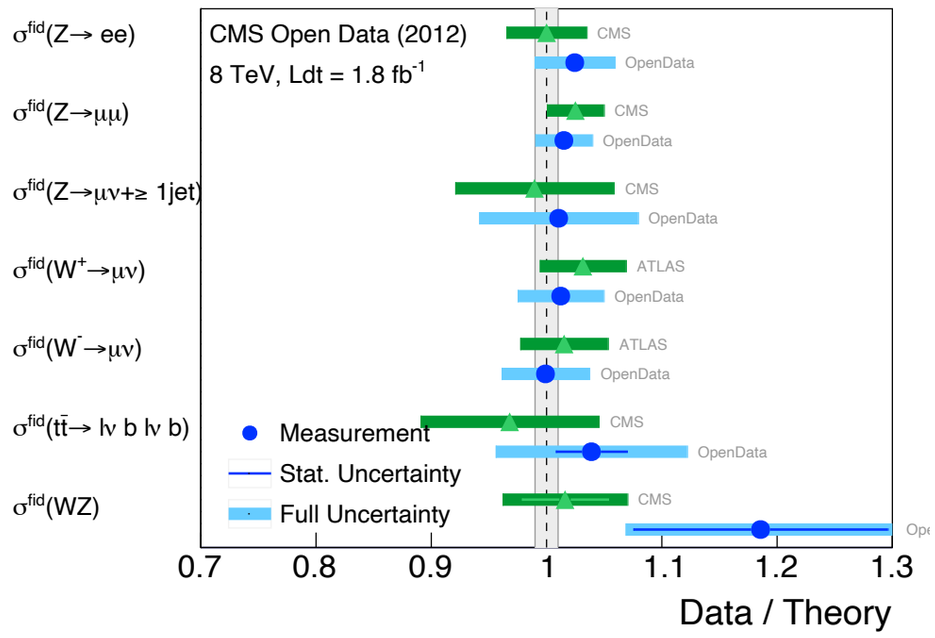
Standard Model Analyses

Jet Substructure Studies



[Tripathee, Xue, Larkoski, Marzani, JDT, PRL 2017, PRD 2017; based on Larkoski, Marzani, JDT, PRD 2015]

Electroweak Benchmarks

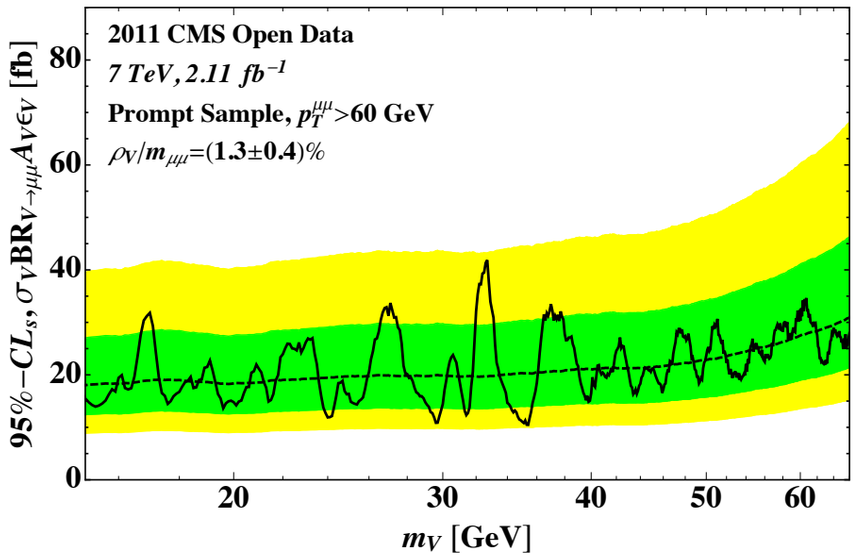


[Apyan, Cuozzo, Klute, Saito, Schott, Sintayehu, JINST 2020]

2010 data ⇒ 2014 release ⇒ 2015 idea ⇒ 2017 analysis

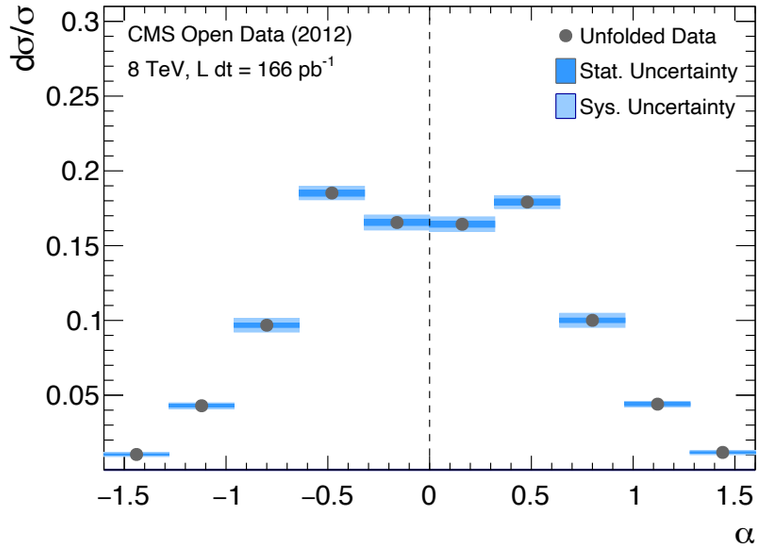
BSM Searches

Dimuon Resonance Hunt



[Cesarotti, Soreq, Strassler, JDT, Xue, PRD 2019]

Non-Standard Parity Violation

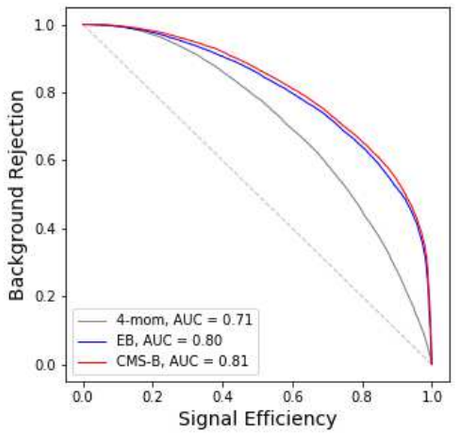


[Lester, Schott, JHEP 2019]

"It is hard to imagine any reason why every possible attempt should not be made to test and re-test the fundamental symmetries of nature every time a door opens onto a new energy range."

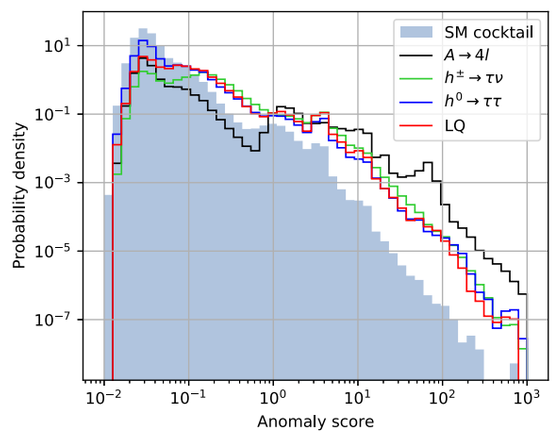
Machine Learning Studies

End-to-End Classification



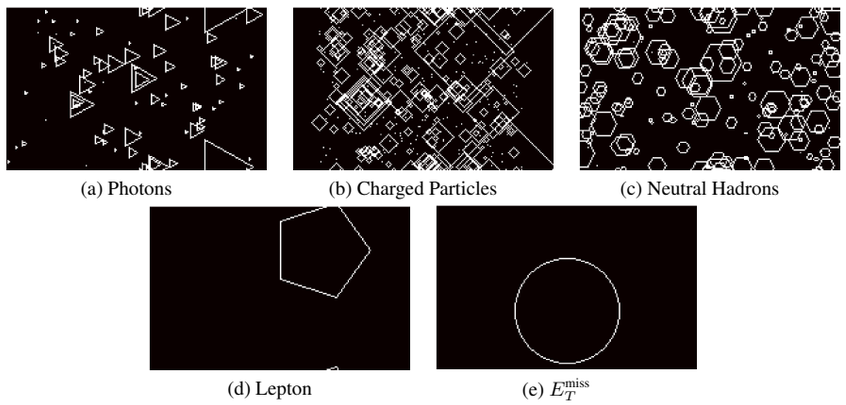
[Andrews, Paulini, Gleyzer, Poczos, CSBS 2020; see also Andrews, Alison, An, Bryant, Burkle, Gleyzer, Narain, Paulini, Poczos, Usai, NIM 2020]

Anomaly Detection



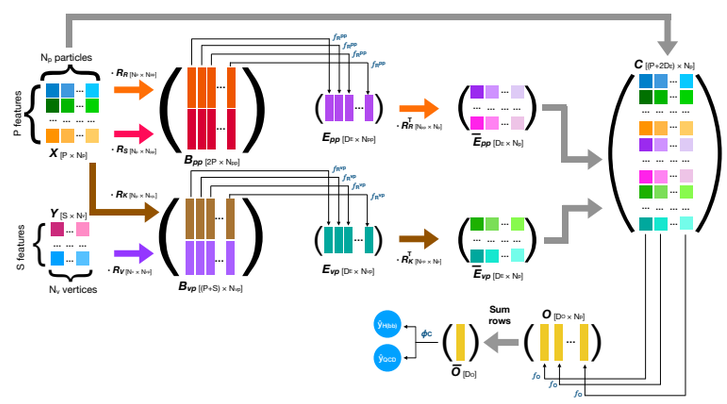
[Knapp, Dissertori, Cerri, Nguyen, Vlimant, Pierini, arXiv 2020]

Computer Vision Techniques



[Fernández Madrazo, Heredia Cacha, Lloret Iglesias, Marco de Lucas, EPJWoC 2019; figure from Nguyen, Weitekamp, Anderson, Castello, Cerri, Pierini, Spiropulu, Vlimant, CSBC 2019]

Interaction Networks



[Moreno, Nguyen, Vlimant, Cerri, Newman, Periwal, Spiropulu, Duarte, Pierini, PRD 2020]

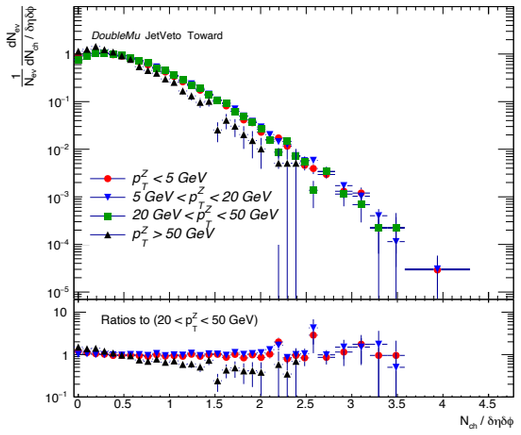
Machine Learning Studies

The NSF AI Institute for Artificial Intelligence and Fundamental Interactions

Postdoctoral fellowship opportunity (Oct 20 deadline):
http://iaifi.org/fellows.html

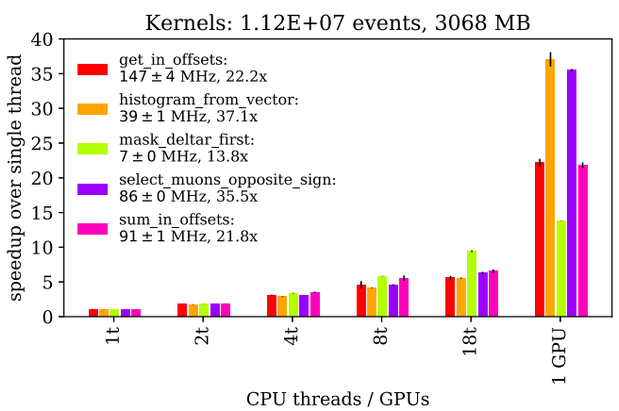
And More!

Underlying Event Studies



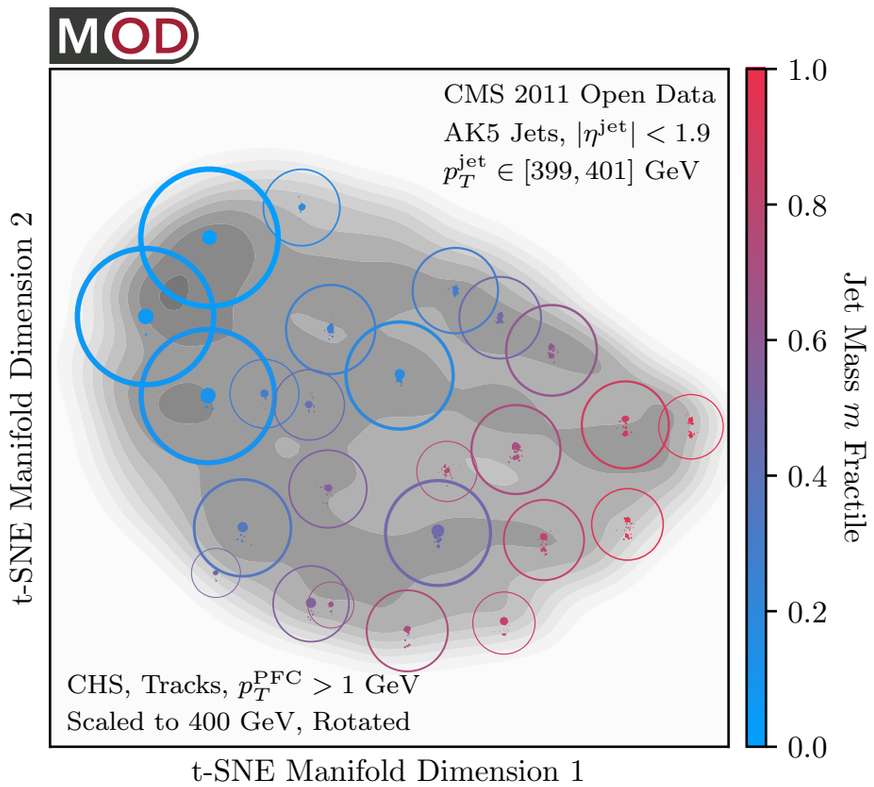
[Paktinat Mehdiabadi, Fahim, JPG 2019]

GPU Acceleration



[Pata, Spiropulu, arXiv 2019]

Event Space Geometry



[Komiske, Mastandrea, Metodiev, Naik, JDT, PRD 2020]

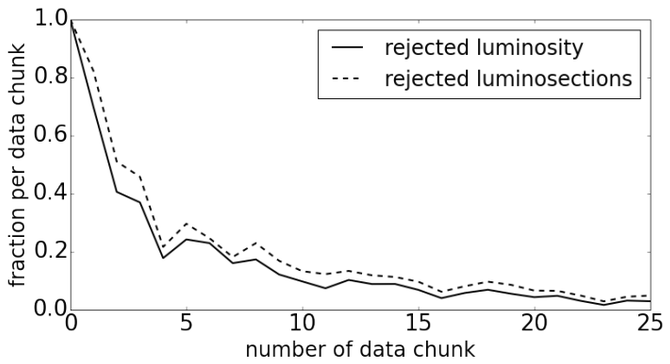
Please contact me if I missed your CMS Open Data study!

And Even More!

Addendum

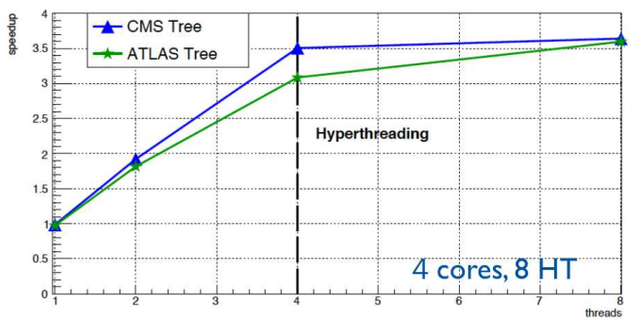
Thanks to Achim Geiser for the bibliography!

Data Quality



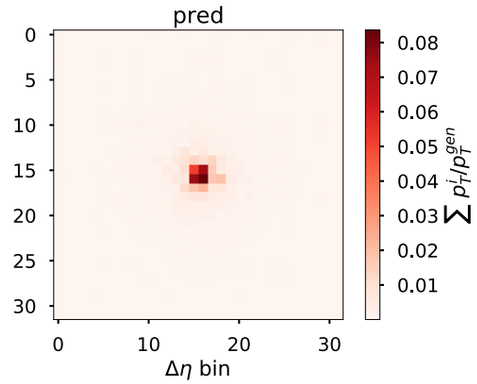
[Borisyak, Ratnikov, Derkach, Ustyuzhanin, JPCS 2017]

Parallel Processing



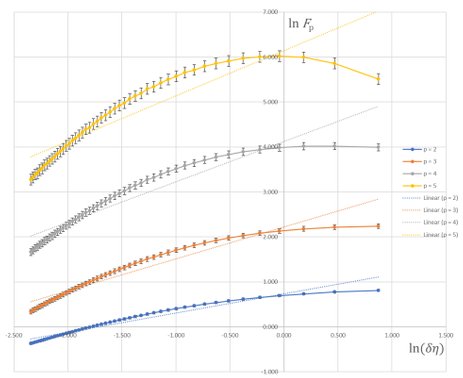
[Piparo, Tejedor, Guiraud, Ganis, Mato, Moneta, Valls Pla, Canal, JPCS 2017]

Generative Models



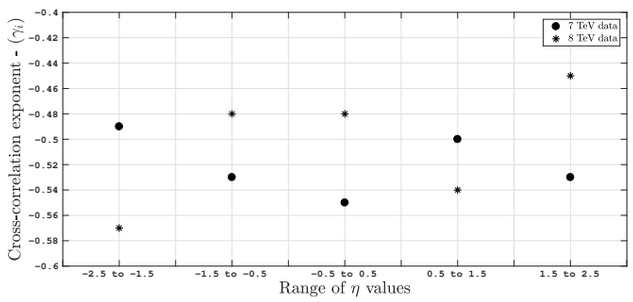
[Musella, Pandolfi, CSBS 2018]

Rapidity Fluctuations



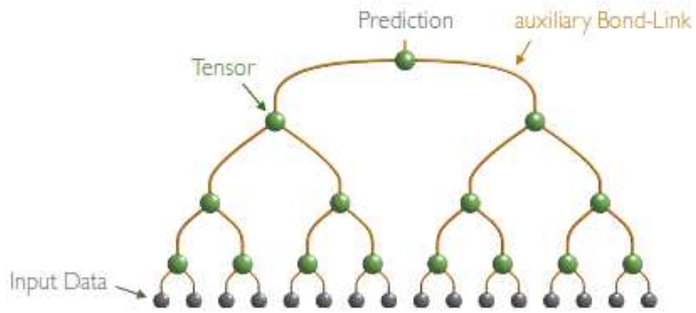
[Ong, Yuen, Ang, Chan, Oh, EPJWoC 2019]

Symmetry Scalings



[Bhaduri, Bhaduri, Ghosh, AHEP 2020]

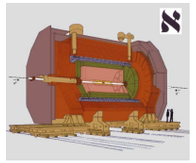
Tensor Networks (LHCb)

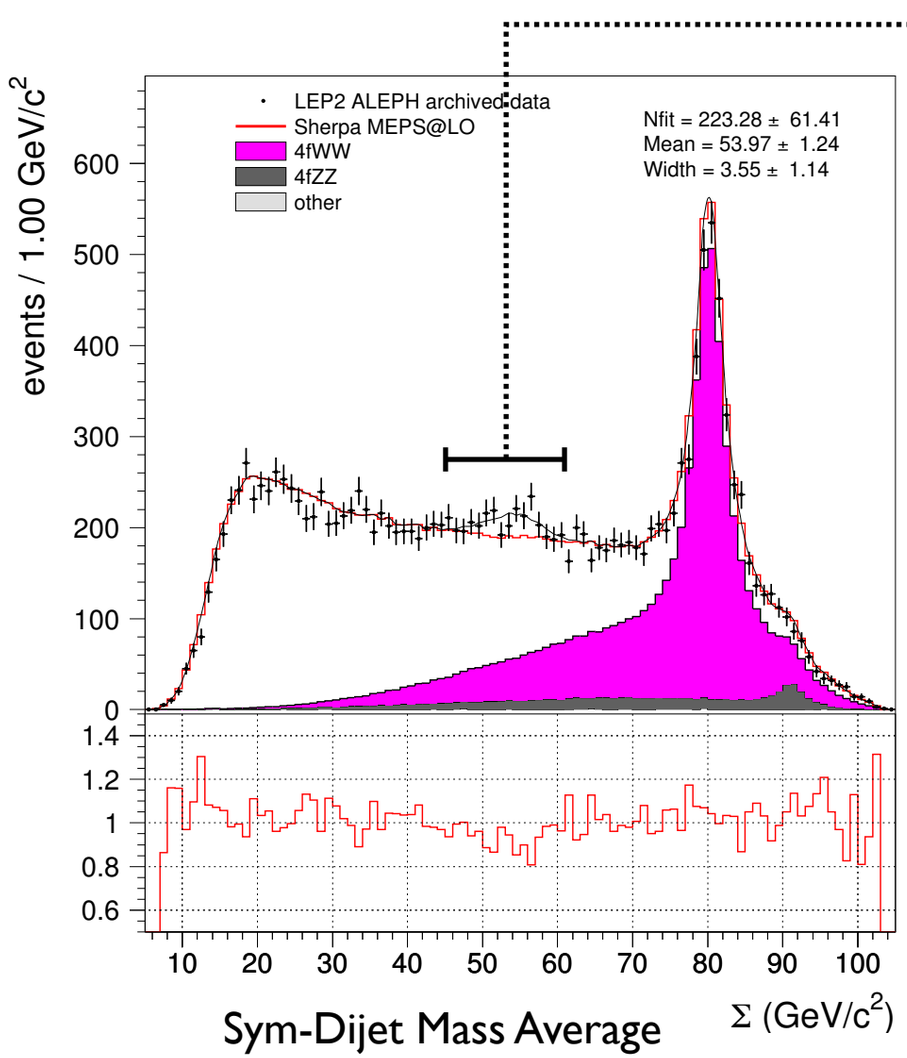


[Trenti, Sestini, Gianelle, Zuliani, Felser, Lucchesi, Montangero, arXiv 2020]

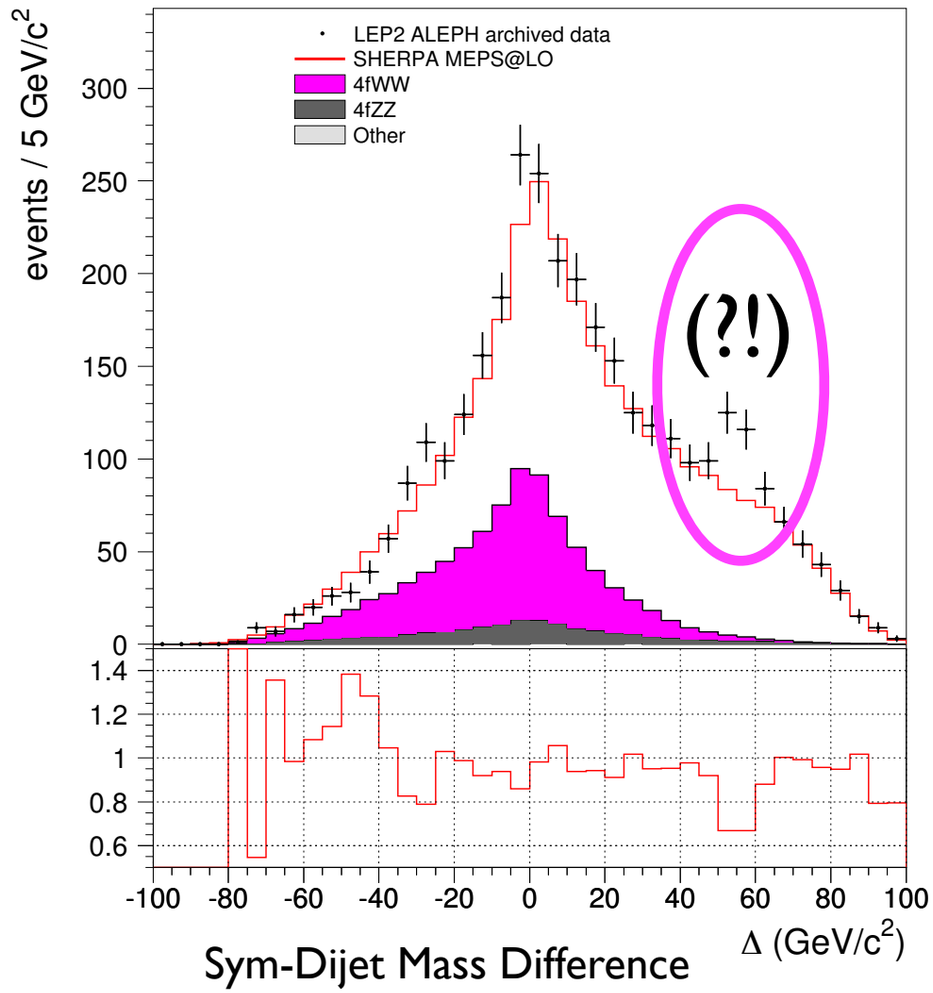
The Other 27 Kilometer Circular Collider

ALEPH Puzzle in Quad-Jet Kinematics





Δ for 45 GeV $< \Sigma <$ 61 GeV



"Whether the excesses described here ultimately are explained by QCD or physics beyond the Standard Model, our results demonstrate the lasting utility of the archived LEP data."

[Kile, von Wimmersperg-Toeller, arXiv 2017, JHEP 2018, arXiv 2017]

Data preservation and outside analyses require significant resources:

People, time, ideas, and money

Data preservation and outside analyses require significant resources:

People, time, ideas, and money

Thank you (both organizers and participants) for investing in this workshop!