Jet substructure form ALEPH archived e^+e^- data

Yi Chen, Yen-Jie Lee,

Massachusetts Institute of Technology

Paoti Chang

National Taiwan University

Christopher McGinn, Dennis Perepelitsa

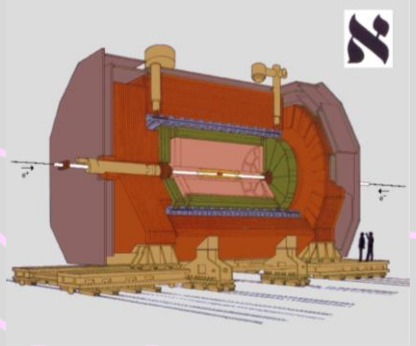
University of Colorado Boulder

Marcello Maggi

Universita degli Studi di Bari

Yang-Ting Chien

Georgia State University



Jet Physics: From RHIC/LHC to EIC

30 June, 2022

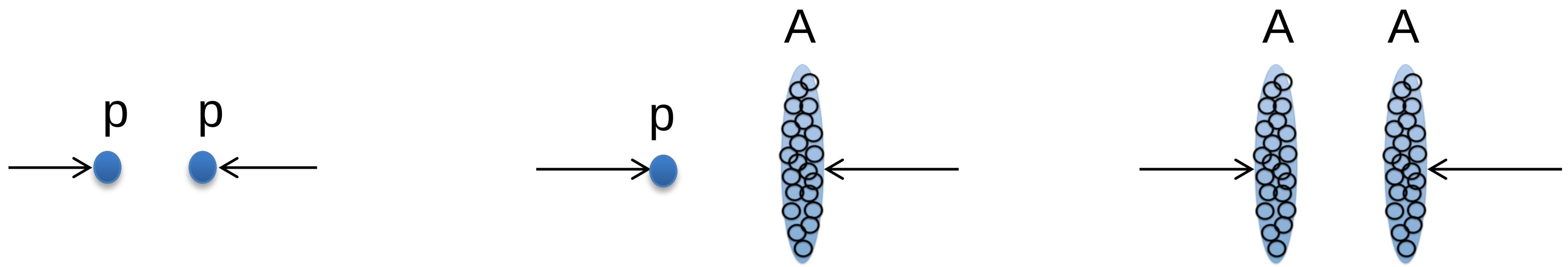
Motivation

- Jets are some of the most powerful tools for the study of Quantum Chromodynamics

- Since the end of LEP operation, significant progress has been made in jet definition and jet algorithms:
  - Jet substructure observables have been widely explored in **pp** and **HI** collisions
  - Novel tools for jet flavor identification, EW boson & top tagging and studies of QGP*
  - However, those techniques are not yet used in $e^+ e^-$ annihilation data

* H. Andrews et al.
J. Phys. G, 47(6) 065102, 2020

- Monte Carlo generators such as SHERPA, PYTHIA 6, PYTHIA 8 and HERWIG are tuned with hadronic event shape observables and hadron spectra in $e^+ e^-$
  - Then used to predict the jet spectra and substructure in more complicated hadron collisions



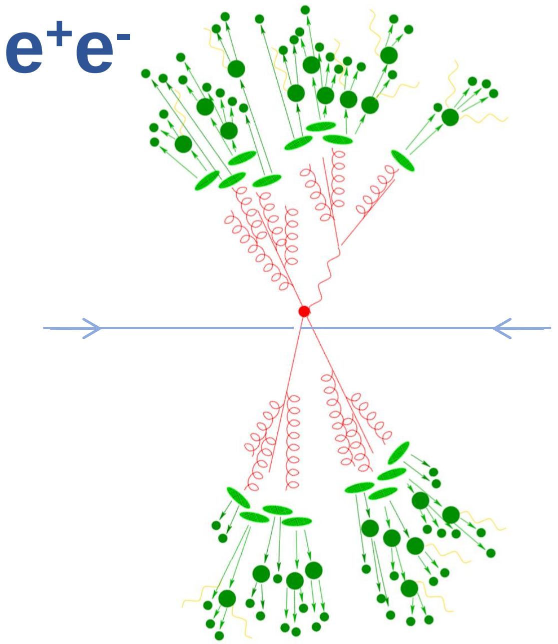
Jets in Electron-Positron Annihilation

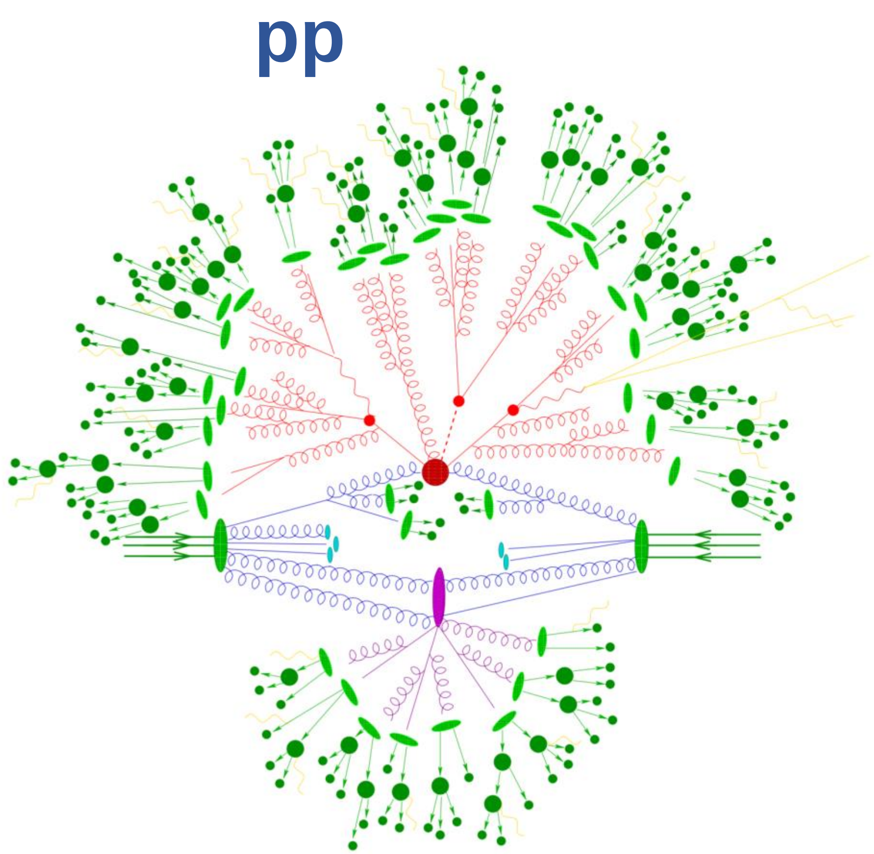
- Jets in e^+e^- with identical algorithms as those used in hadron colliders are of great interest
  - No gluonic initial state radiation
  - No complications of parton distribution functions
  - No beam remnants and multi-parton interactions

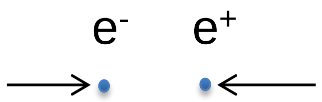
  → **Cleanest test** of pQCD and phenomenological models

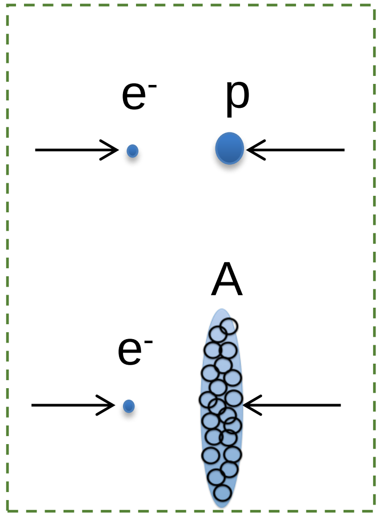
- Serve as a reference for the **pp** and future **EIC** measurements

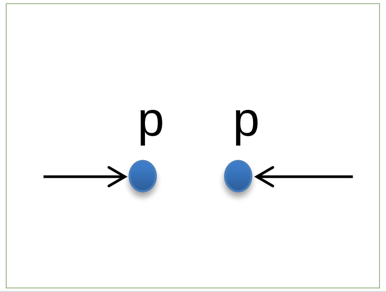
- Inform the QCD studies at the future FCC.

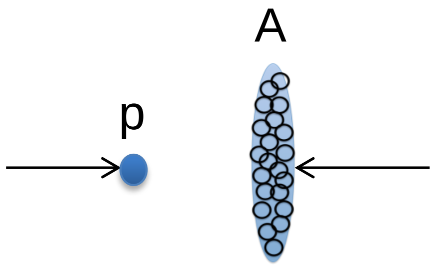


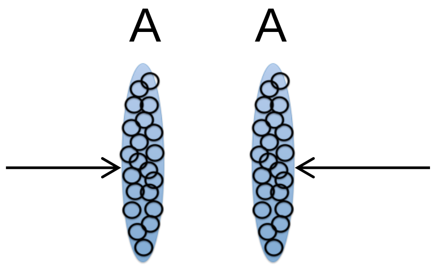




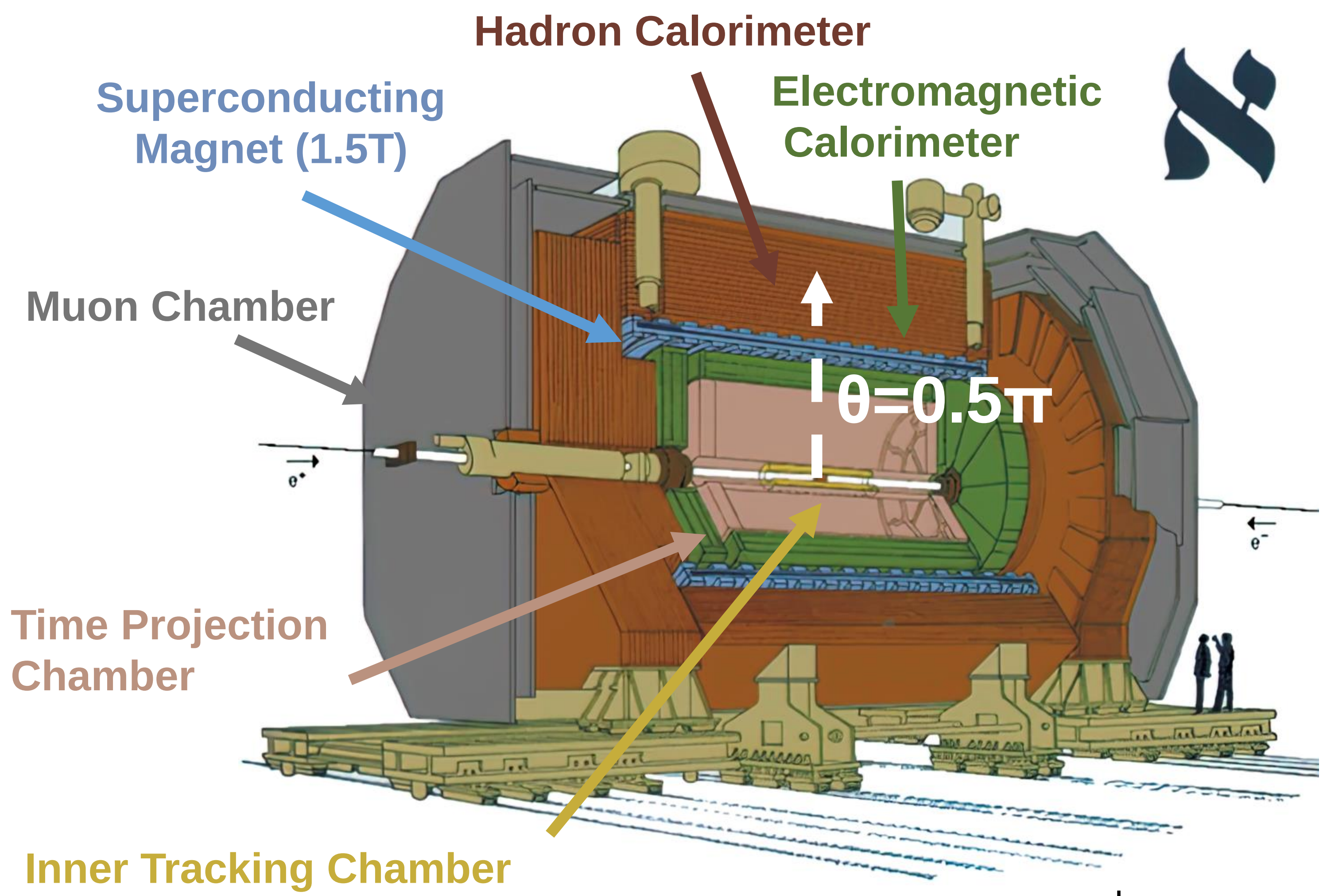


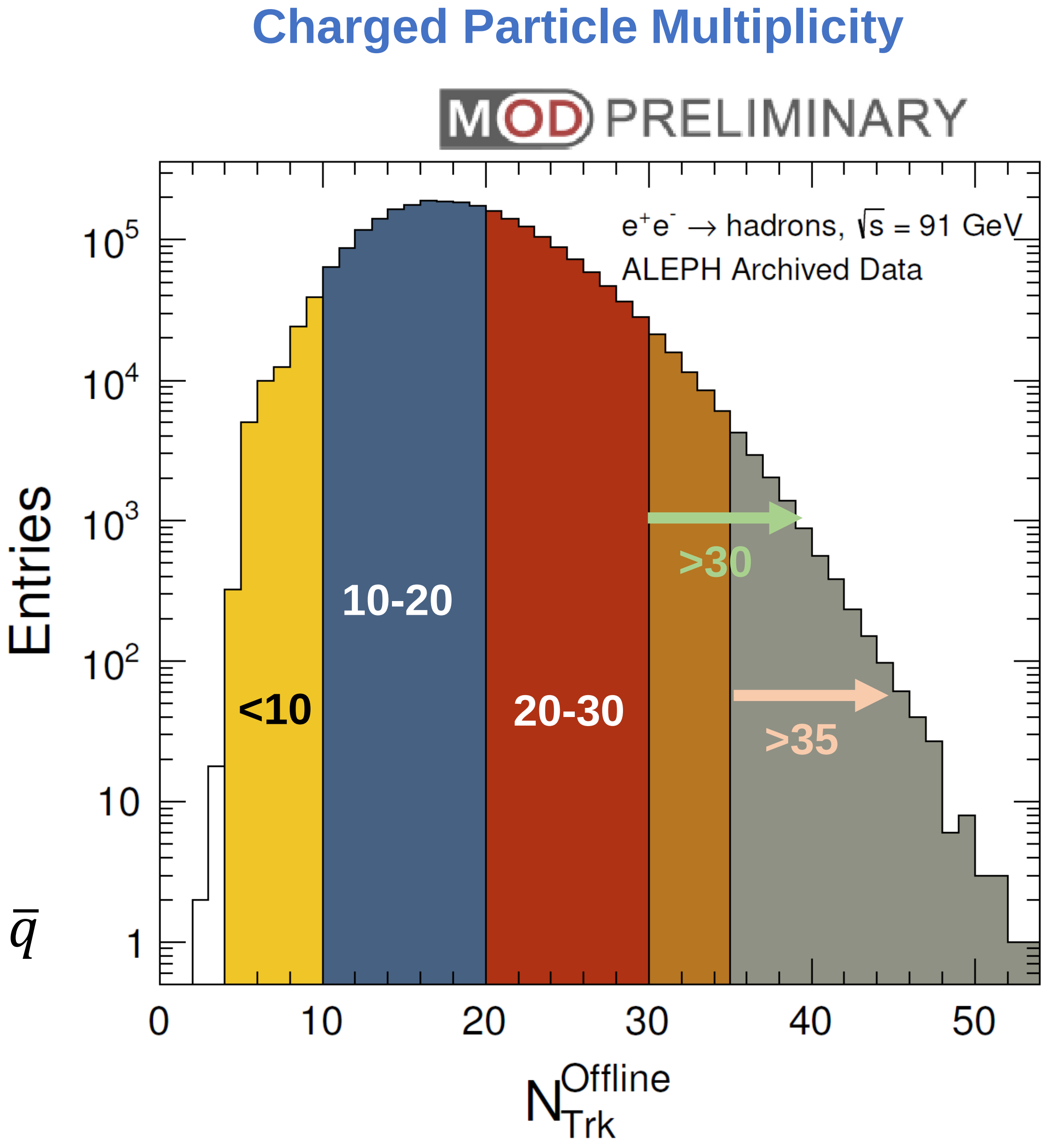




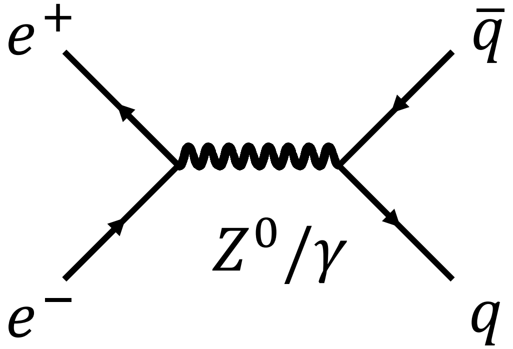


The ALEPH Detector



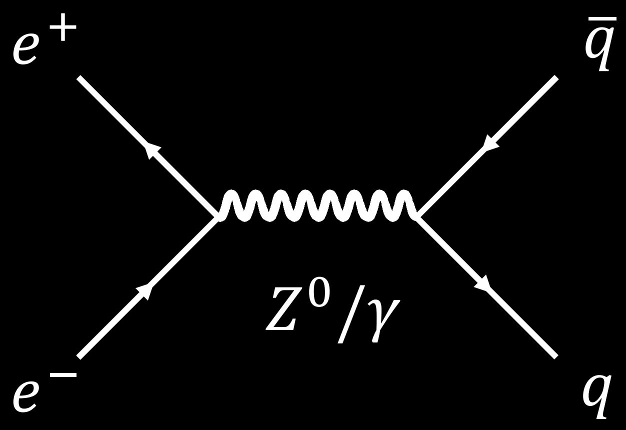


- LEP1 e^+e^- data at Z pole (91 GeV) taken between 1992-1995
- Approximately 2.5 million hadronic events are recorded



High Multiplicity Event in e^+e^- Collisions

Highest multiplicity event in ALEPH LEP1 data
Collision Energy = 91 GeV

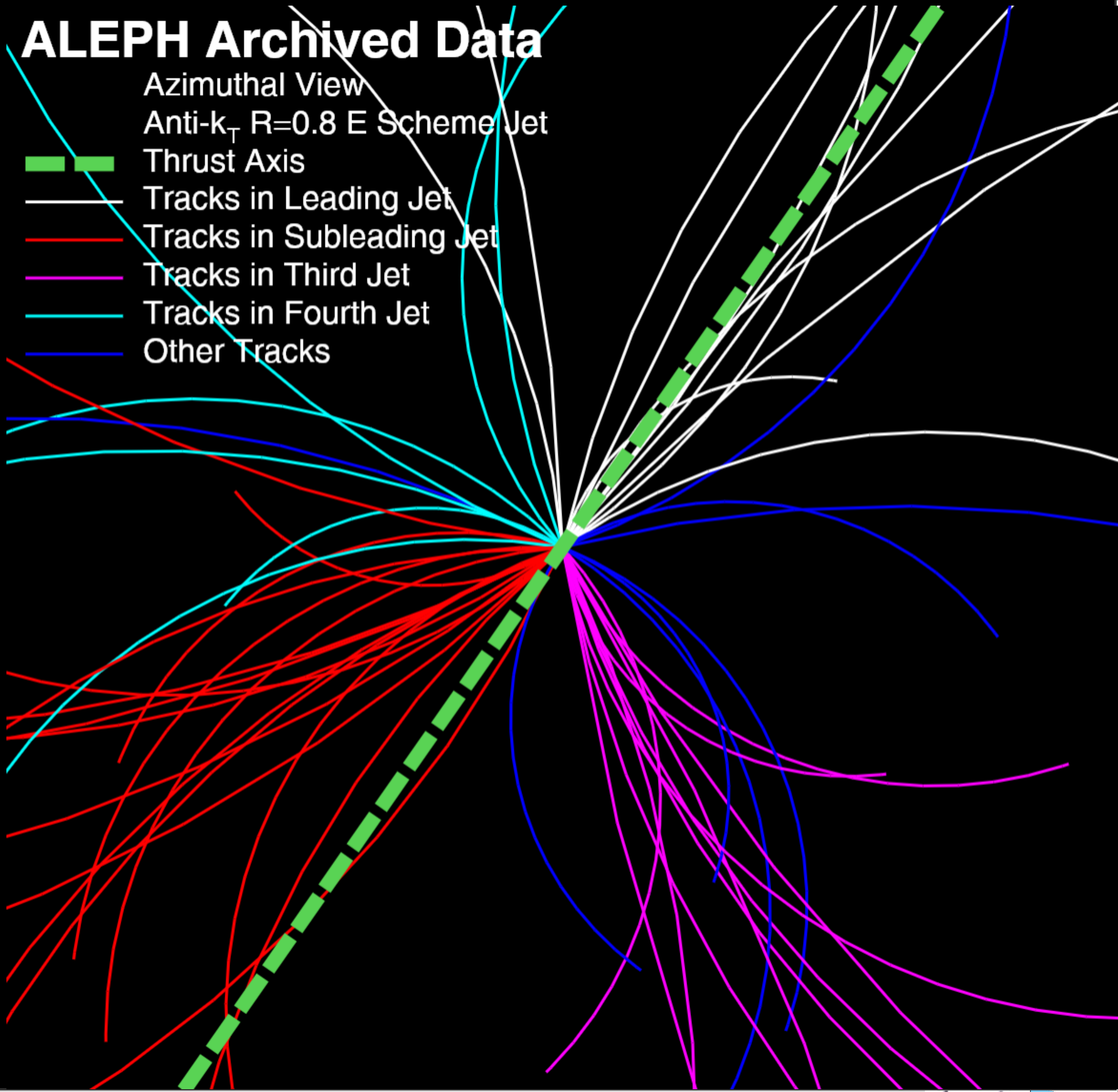


Anthony Badea, Austin Baty, Chris McGinn, Michael Peters, Jesse Thaler, Gian Michele Innocenti, Paoti Chang, Tzu-An Sheng

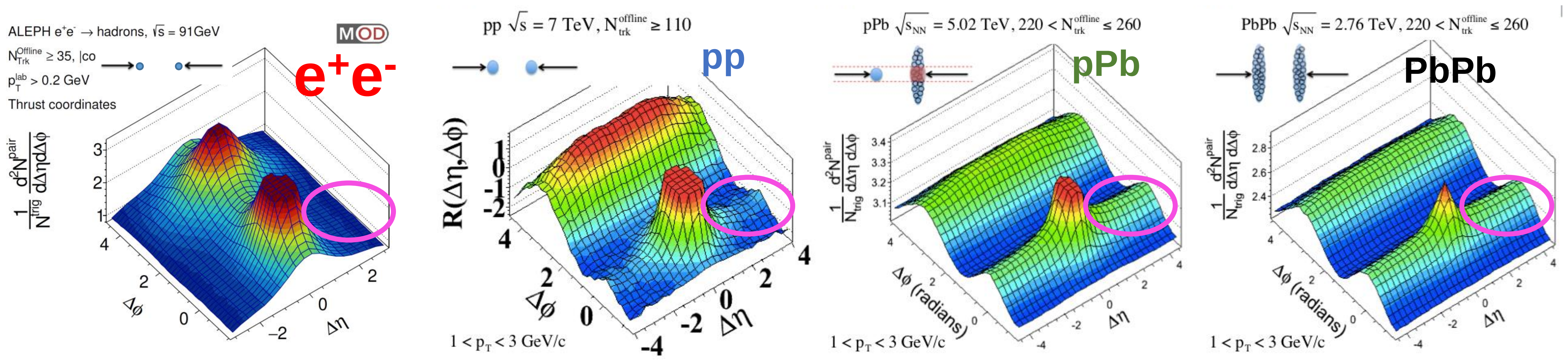
+ YJL

arXiv: 1906.00489
PRL 123, 212002 (2019)

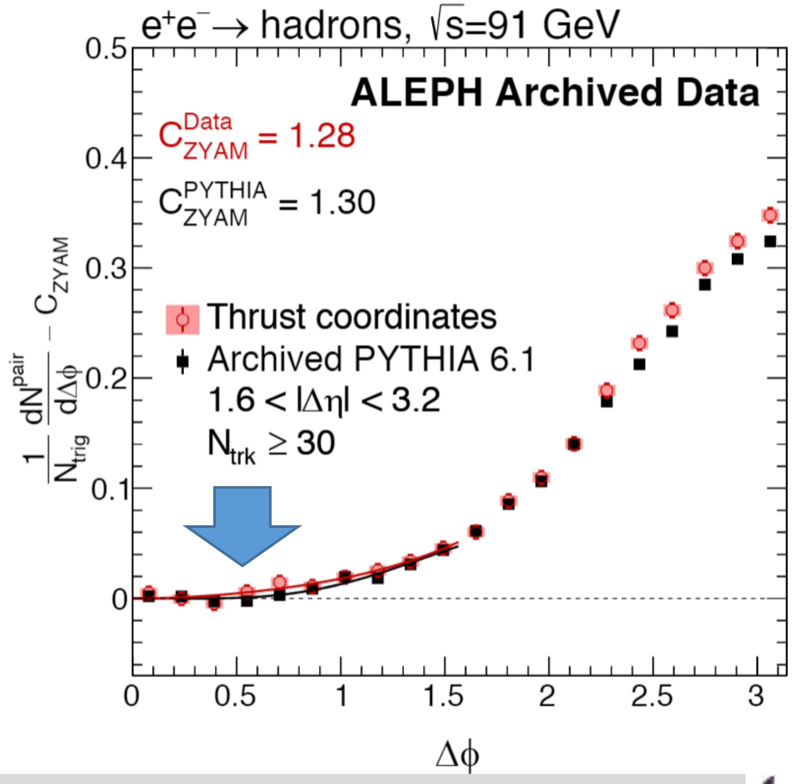
55 Charged Particles
Thrust T=0.71



Two-Particle Correlation Function from ALEPH e^+e^- at 91 GeV



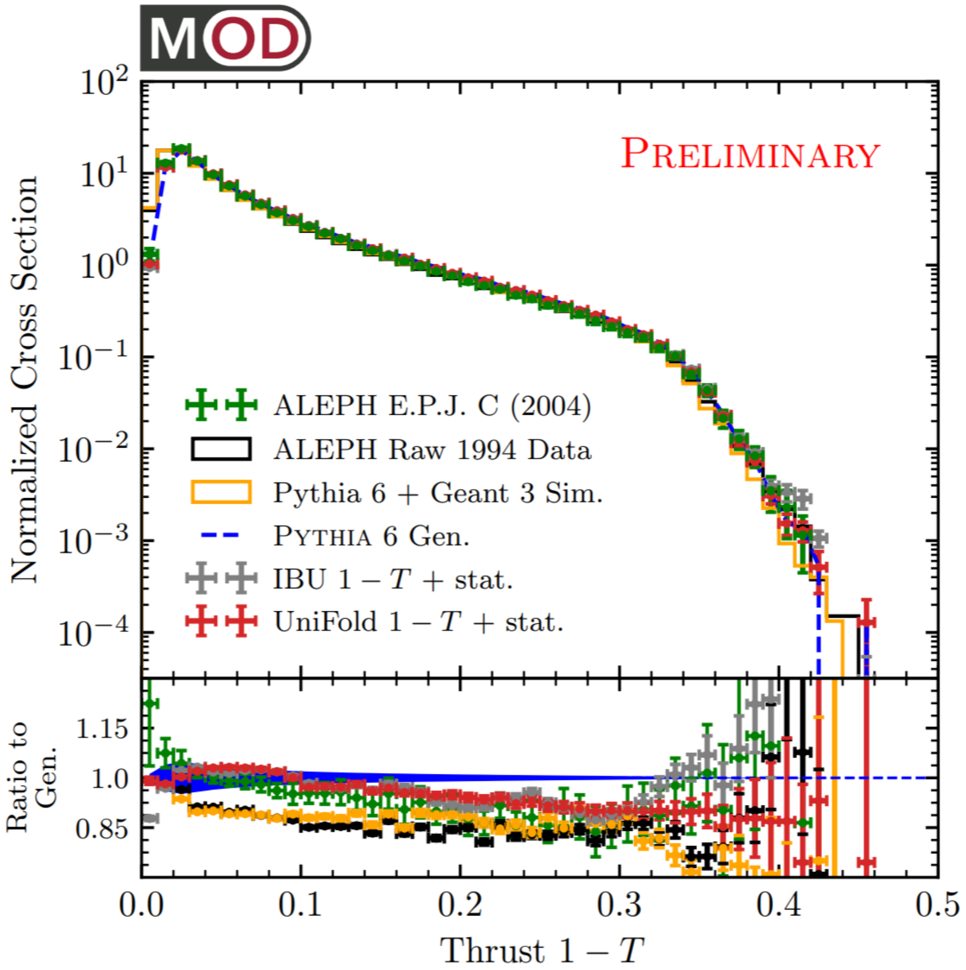
- No sign of ridge signal in electron-positron collisions up to ~ 55 charged particles per event
- New reference to the collective behavior in small collision systems!

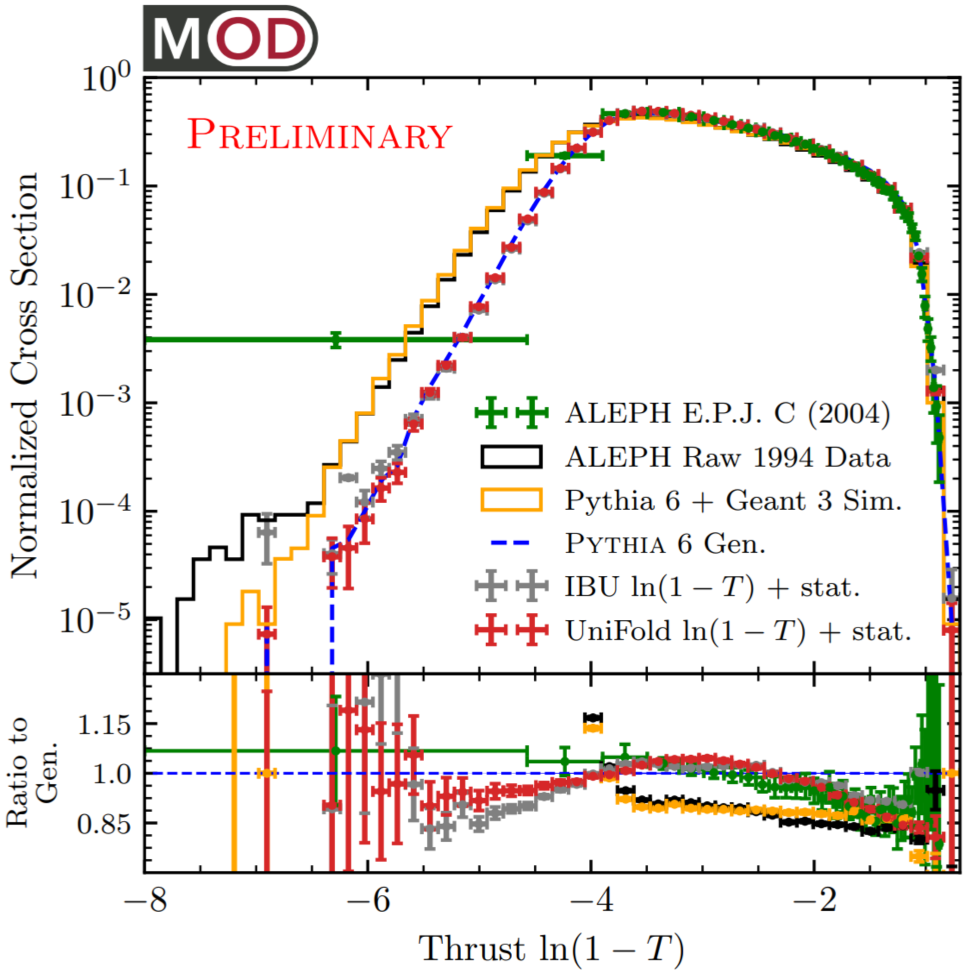


Anthony Badea | Austin Baty | Chris McGinn | Michael Peters | Jesse Thaler | Gian Michele Innocenti | Paoti Chang | Tzu-An Sheng + YJL

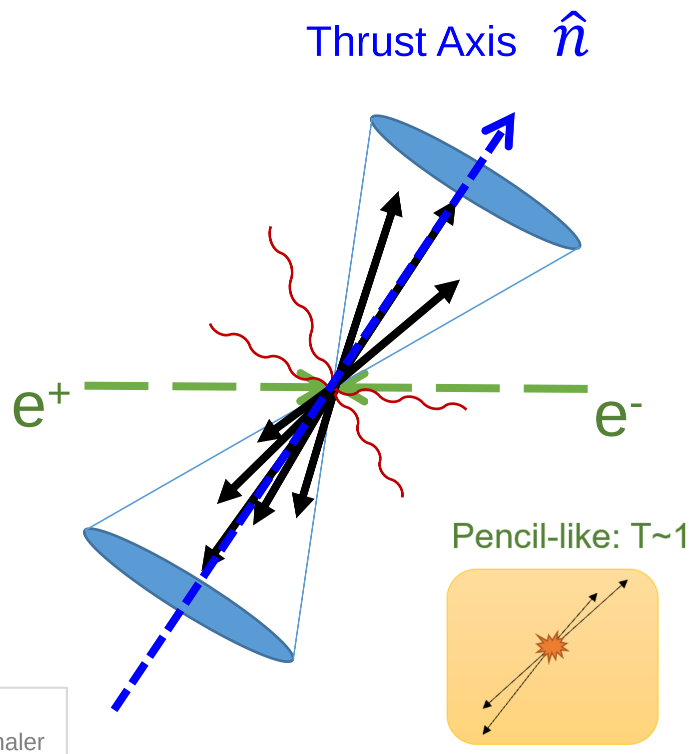
arXiv: 1906.00489
PRL 123, 212002 (2019)

Unfolded Thrust Distribution





$$T = max_{\hat{n}} \frac{\sum_i |\hat{p}_i . \hat{n}|}{\sum_i |\hat{p}_i|}$$



- Unfolded results consistent with ALEPH publication in 2004

EPJC 35 (2004) 456

- New differential measurement in the small **1-T** region (pencil-like event)

Paper in preparation

OmniFold algorithm: PRL 124, 182001
Andreassen, Komiske, Metodiev, Nachman, Thaler

Anthony Badea | Patrick T. Komiske III | Eric Metodiev | Jesse Thaler | Ben Nachman

+ YJL

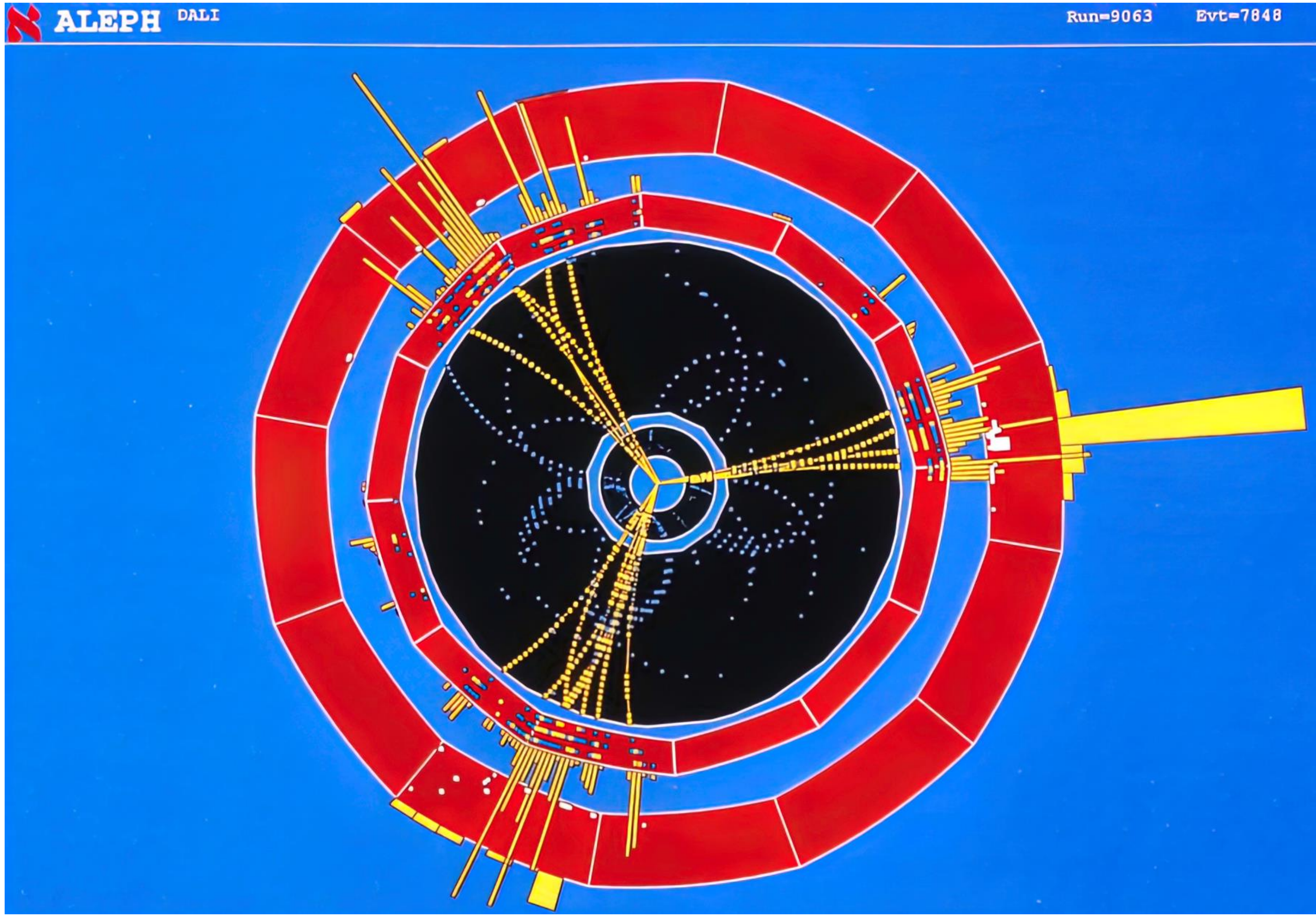
Jet Reconstruction

- ALEPH archived data and MC from **1994** is used for this analysis
- Identical hadronic Z^0 decay selection as the correlation function paper and ALEPH QCD papers
- Jets are reconstructed with an "anti-k_T" algorithm with $\mathbf{R_0=0.4}$ using energy-flow objects identified with trackers, calorimeters and muon chambers.
- Note that the distance metric is defined by jet energy (E) and jet angle (θ):

$$d_{ij} = \min\left(E_i^{-2}, E_j^{-2}\right)\frac{1-\cos\theta_{ij}}{1-\cos R_0}$$

$$d_{iB} = E_i^{-2},$$

where θ_{ij} is the **actual opening angle** between i^{th} and j^{th} pseudojets

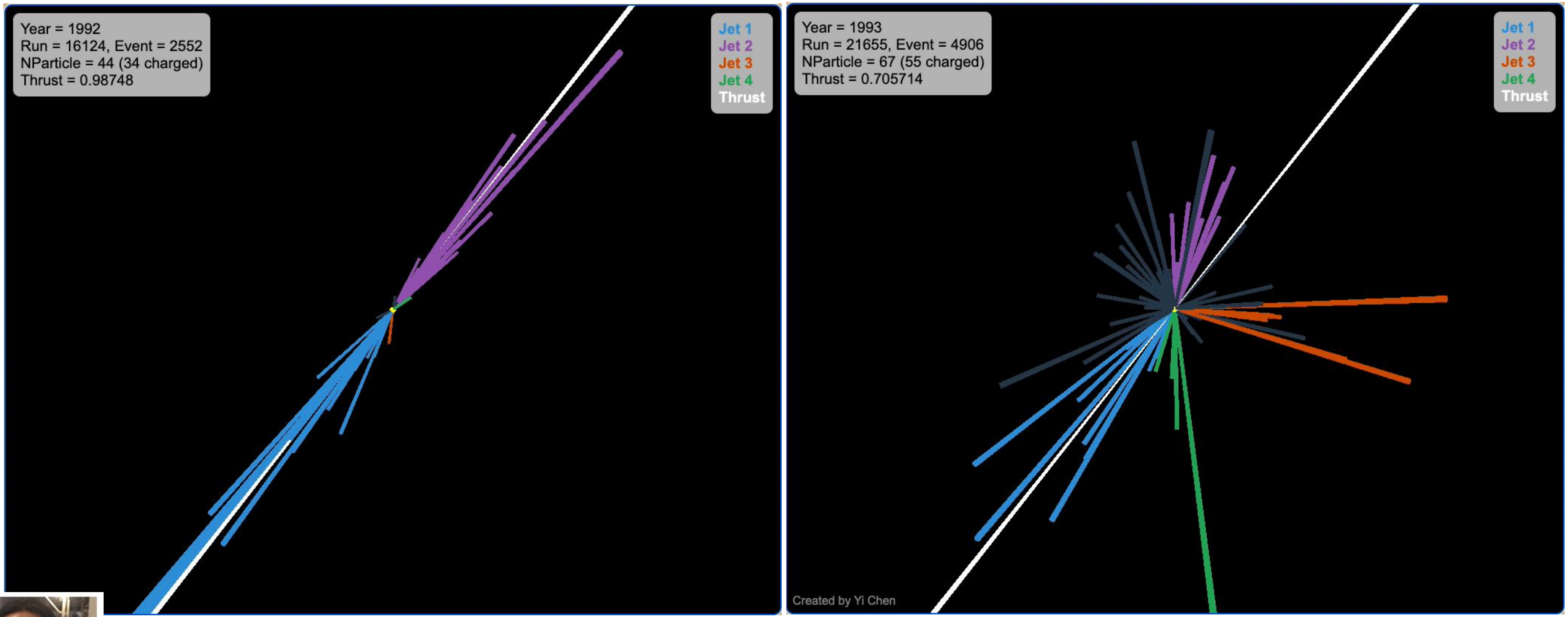


ALEPH Event Display

Anti-k_T :JHEP 04 (2008) 063

TPC paper: PRL 123, 212002 (2019)
QCD paper: EPJC 35 (2004) 456

Jet Events at LEP



Yi Chen
(MIT, CMS)

Jet Analysis with Archived ALEPH data

ALEPH Hadronic Events

Archived PYTHIA 6.1 MC

Jet Reconstruction

MC Jet Energy Correction

Resolution Unfolding, Efficiency Correction

Corrected Results

Data-based Correction: Remove +/- Side Difference

Absolute Energy Scale Correction using Multi-Jet Mass

Energy Corrected Jet Data

Event generators
PYTHIA, HERWIG, SHERPA, PYQUEN

Compare

Yi Chen
(MIT, CMS)

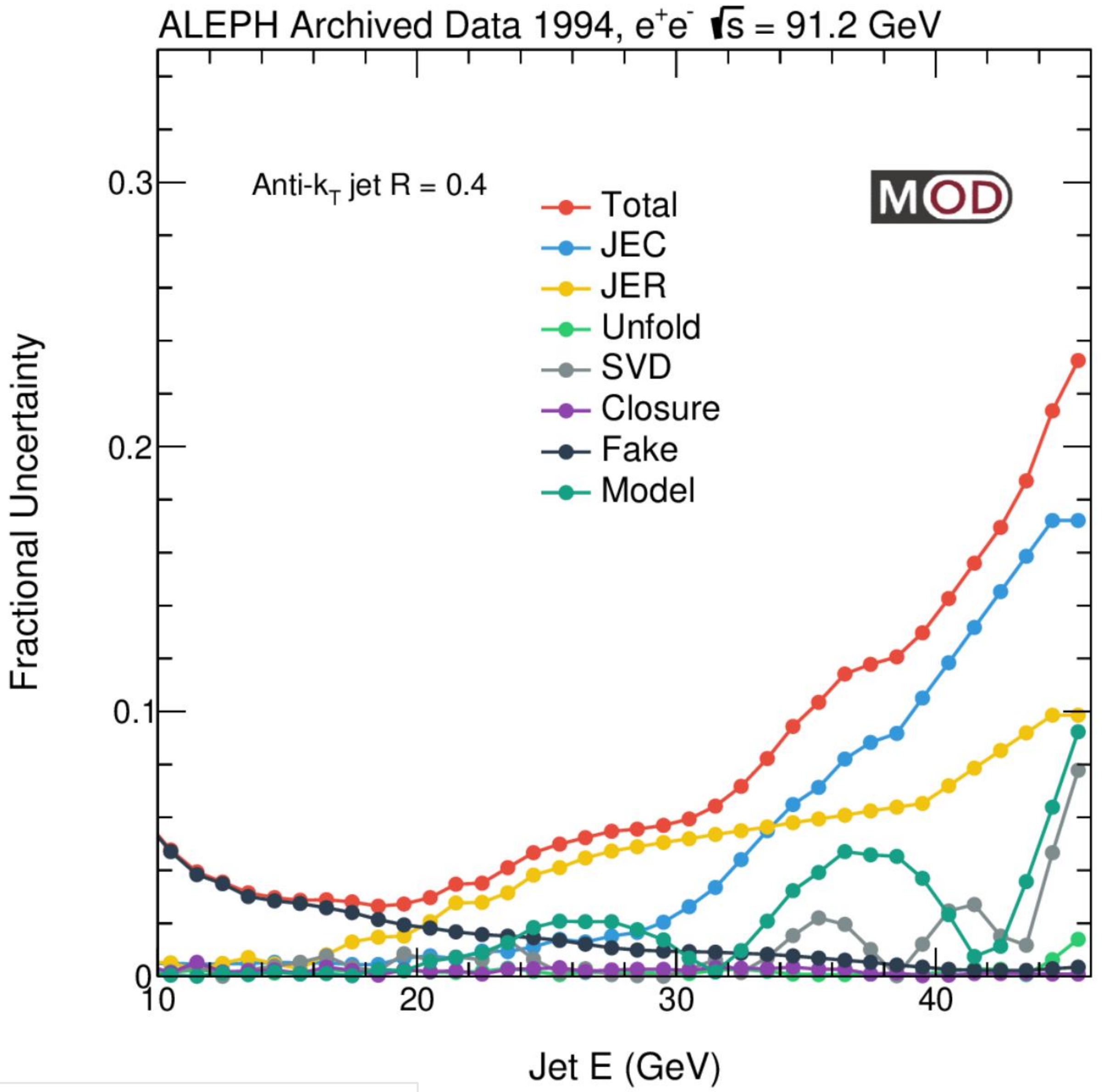
Systematic Uncertainties

For inclusive jet analyses:

- **Jet energy scale**: vary the energy scale by ±0.5%
- **Jet resolution**: vary the resolution scale factor by ± 2.5%
- "**Fake jets**": accidental clusters of energy in the final state that do not correspond to an initial high energy parton
- **Unfolding**: choice of prior, regularization and **unfolding method**
- **Variation of Models:** data-driven reweighting & subjet angle and energy smearing
- **Any non-closure** in MC studies

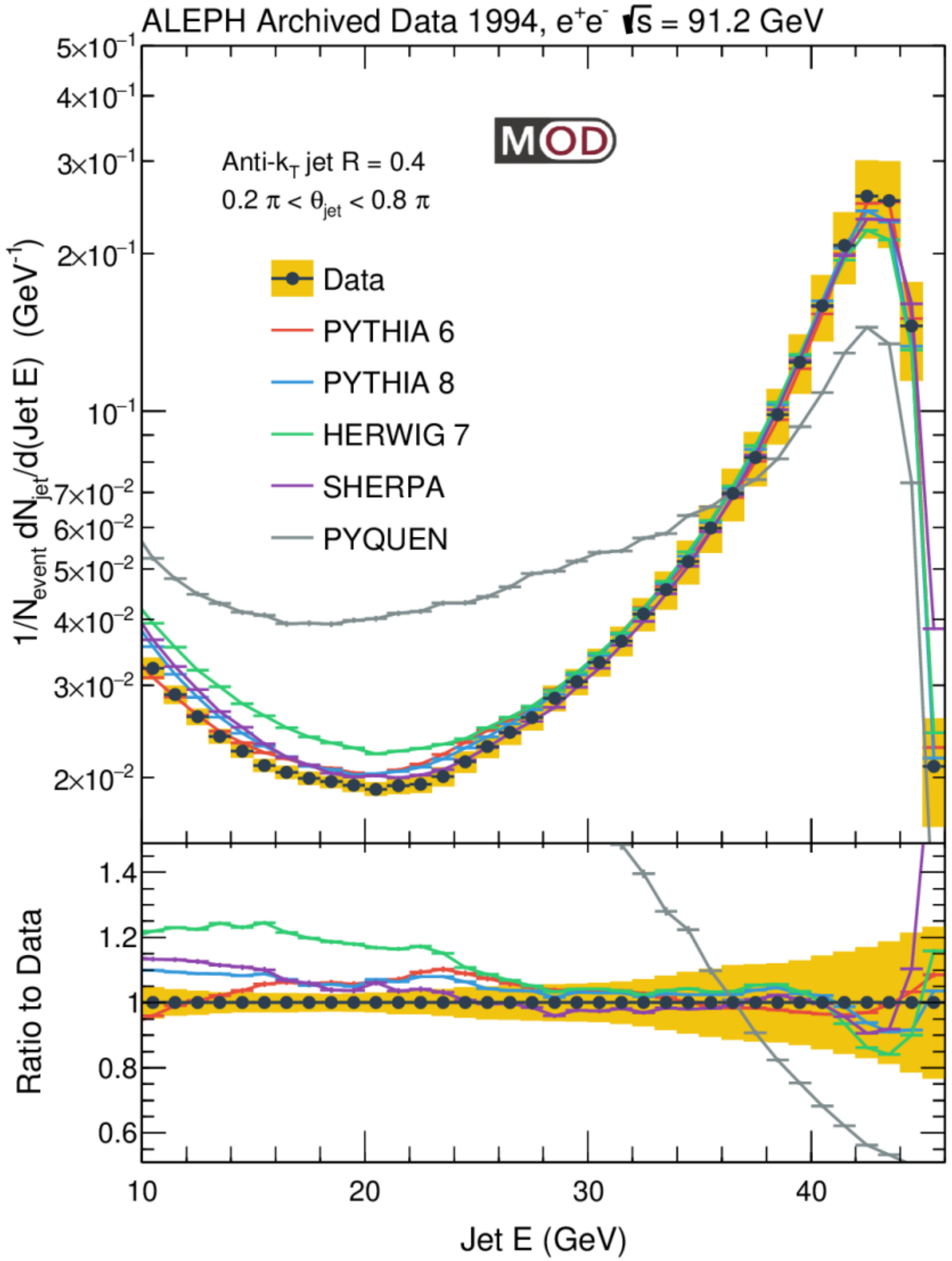
For leading dijet analysis: following additional items:

- **The total energy selection** (to ensure both jets are in the acceptance) is varied to change the purity between 98%- 99.5%.
- **Efficiency correction factor** on jet spectrum are derived using a reweighted MC and the difference is quoted as systematics

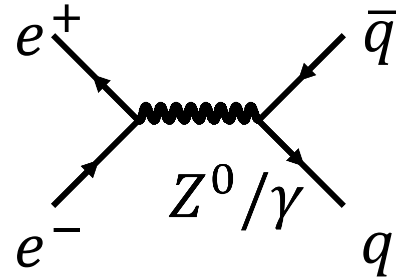


MITHIG-MOD-21-001
arXiv:2111.09914
JHEP 06 (2022) 008

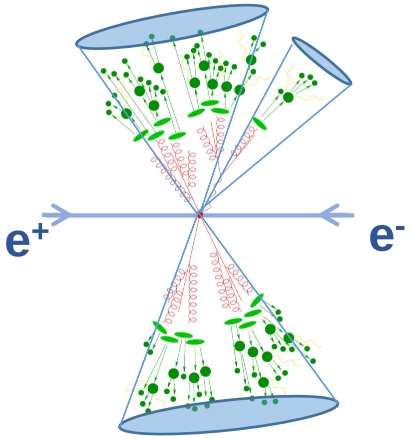
Inclusive Jet Spectrum vs. Generators



- The closest observable to the jet spectra analyses in hadron-hadron collisions!
- Peak at around 43 GeV: from $Z \rightarrow q\bar{q}$ and parton shower of the (anti-)quark **almost fully captured** by the jet clustering



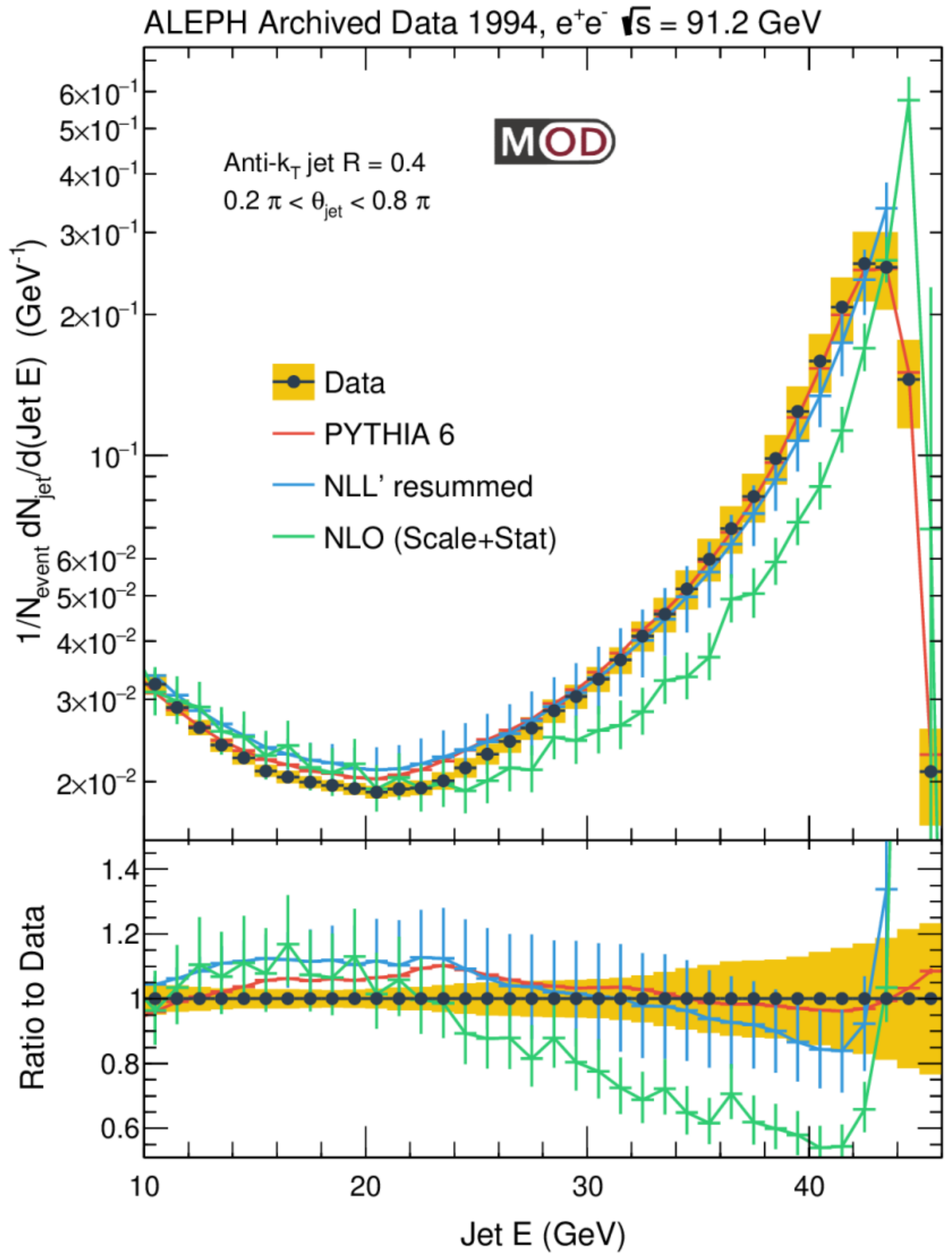
- Minimum at around 20 GeV
- At low E: increase due to a large number of jets from **soft emissions** or **combinatorial**



- SHERPA, PYTHIA 6, PYTHIA 8 and HERWIG capture those general features, overpredict the **data** at low jet E
- PYQUEN (with **energy loss** in minimum-bias PbPb at 5.02 TeV) produce a lot more low energy jets

MITHIG-MOD-21-001
arXiv:2111.09914
JHEP 06 (2022) 008

Inclusive Jet Spectrum vs. NLO Calculation



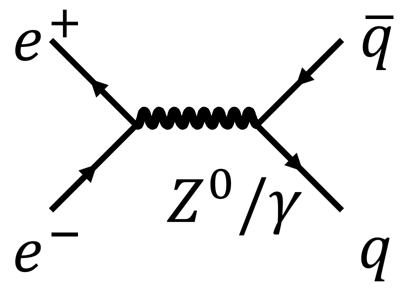
- LO calculation in parton level gives trivial result:

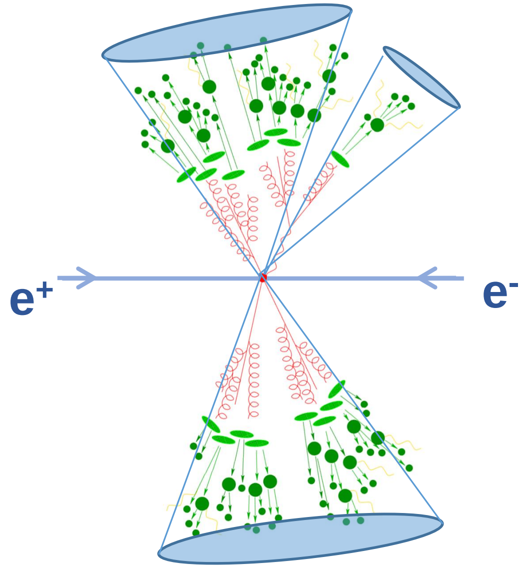
  a delta function at $E = \frac{M_Z}{2}$

- The experimental **data** is wider than predictions from an NLO calculation in **parton level***

- Next-to-leading logarithmic (NLL') threshold and jet radius resummation** give a reasonable description of data

The same data as previous slide



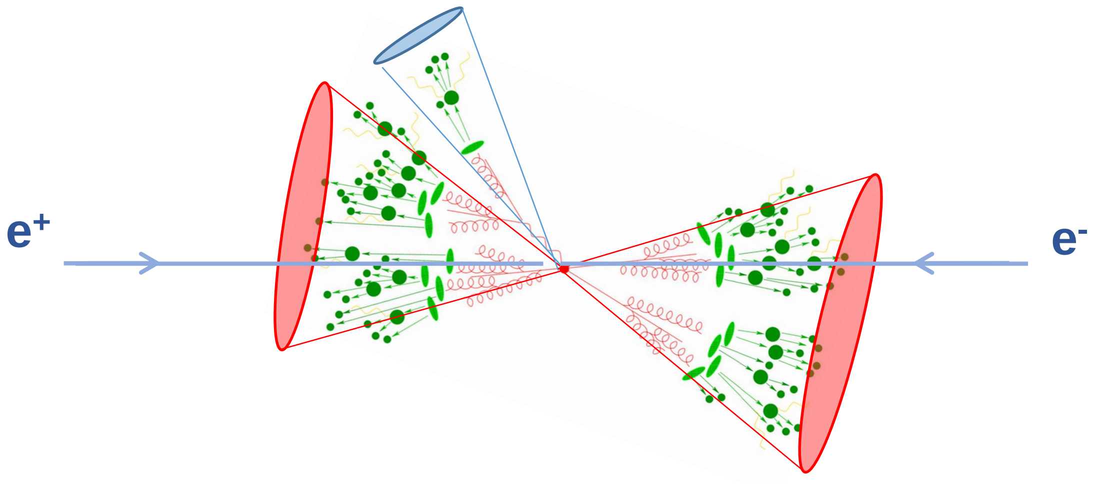


*NLO calculation from João Pires (LIP)

**NLL'+R calculation from Duff Neill, Felix Ringer, Nobuo Sato

Leading Dijet Selection

- When the global leading jets overlap with the dead region (close to beam), some of the jet energy will not be detected and they may **appear as low energy jets.**

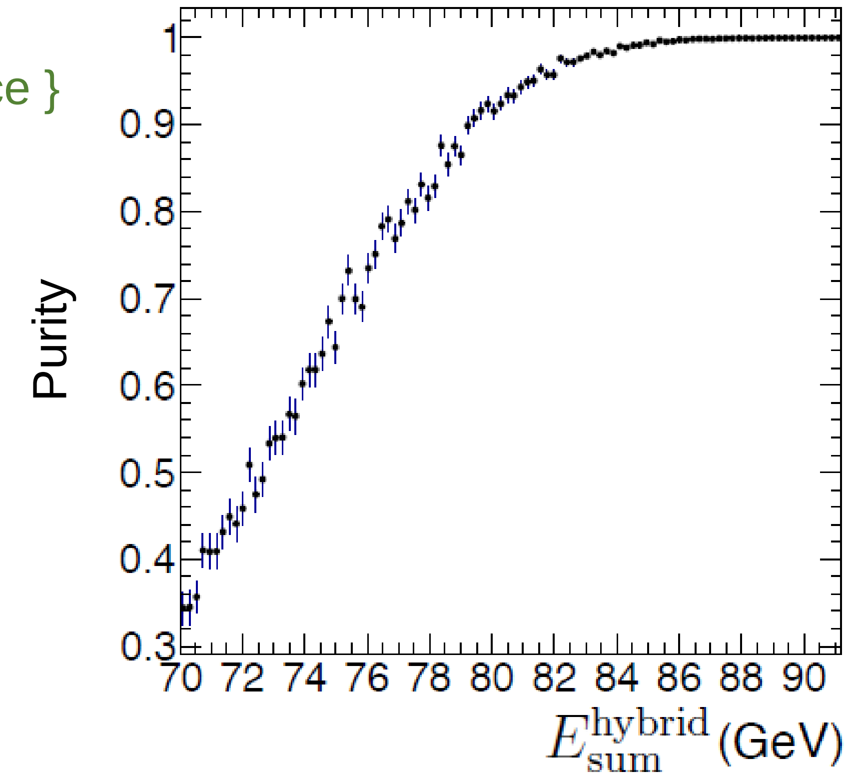


- A hybrid total energy observable is constructed to select events which have both leading jets in the acceptance:

$E_{\mathrm{sum}}^{\mathrm{hybrid}}$ = Energy sum of {Particles within acceptance 0.2π<θ<0.8π}

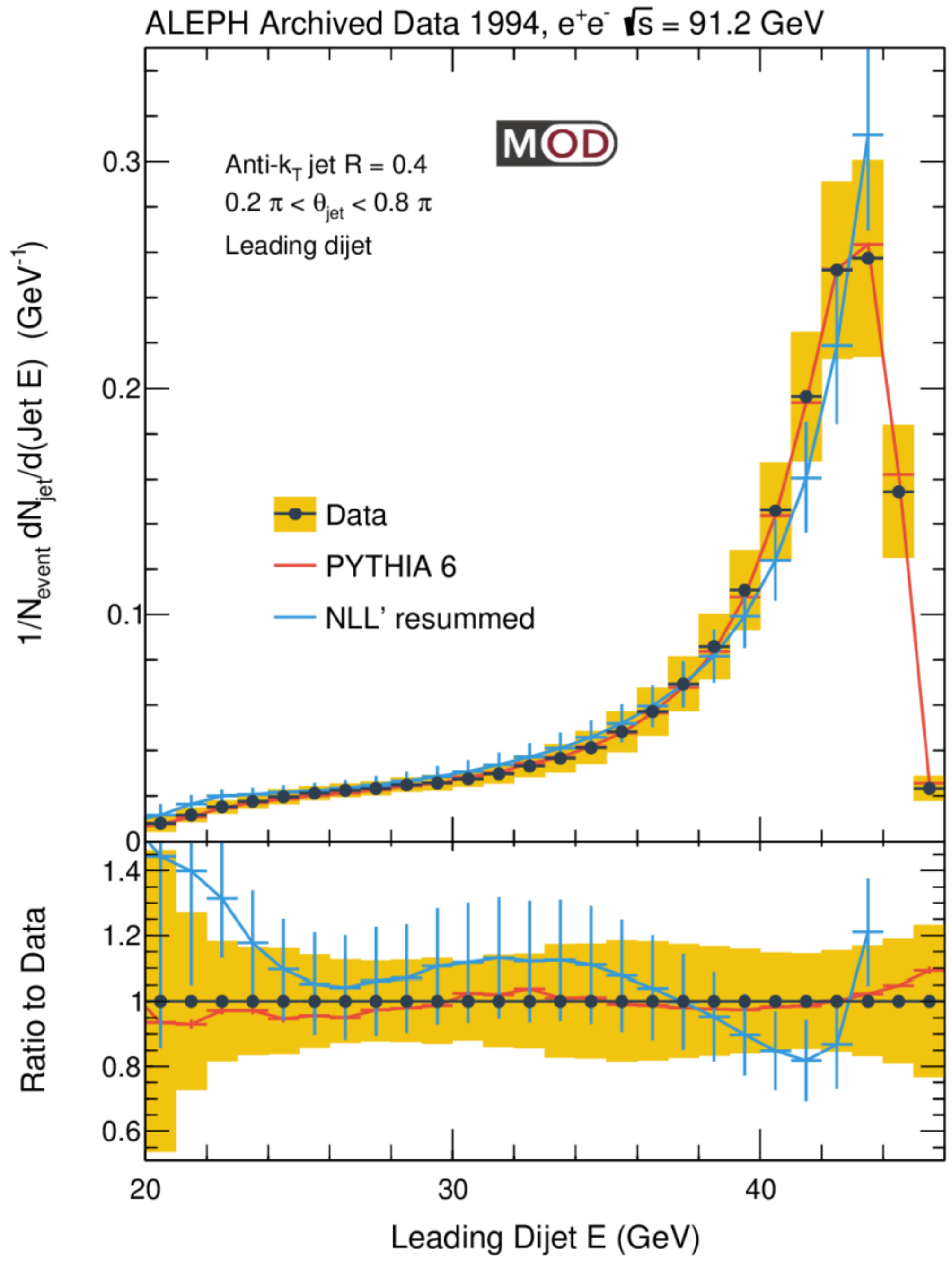
U { Particles with angle < 0.4 to axes of any jet above 5 GeV in the acceptance }

- The nominal selection cut is $E_{\mathrm{sum}}^{\mathrm{hybrid}}$ >83 GeV with a purity of 99% with both leading and subleading jet in the acceptance

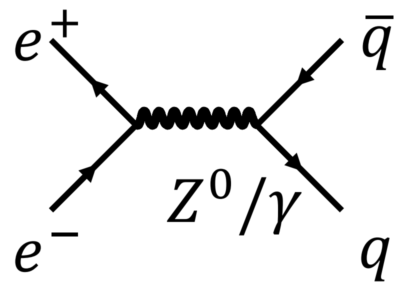


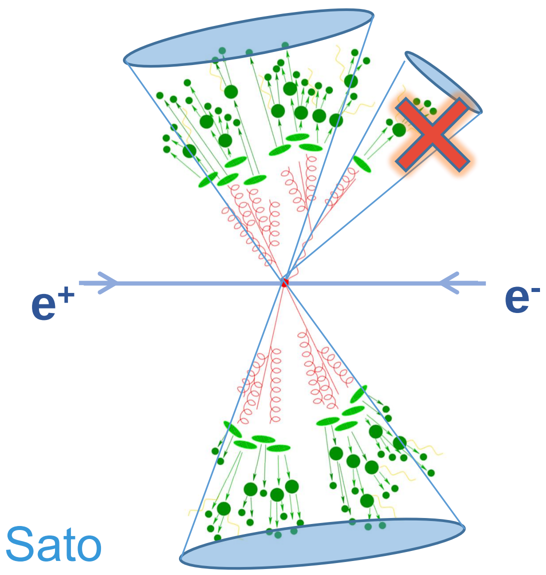
- A correction on the hybrid total energy selection is later applied to the unfolded leading dijet spectrum.

Leading Dijet Spectrum vs. Generators



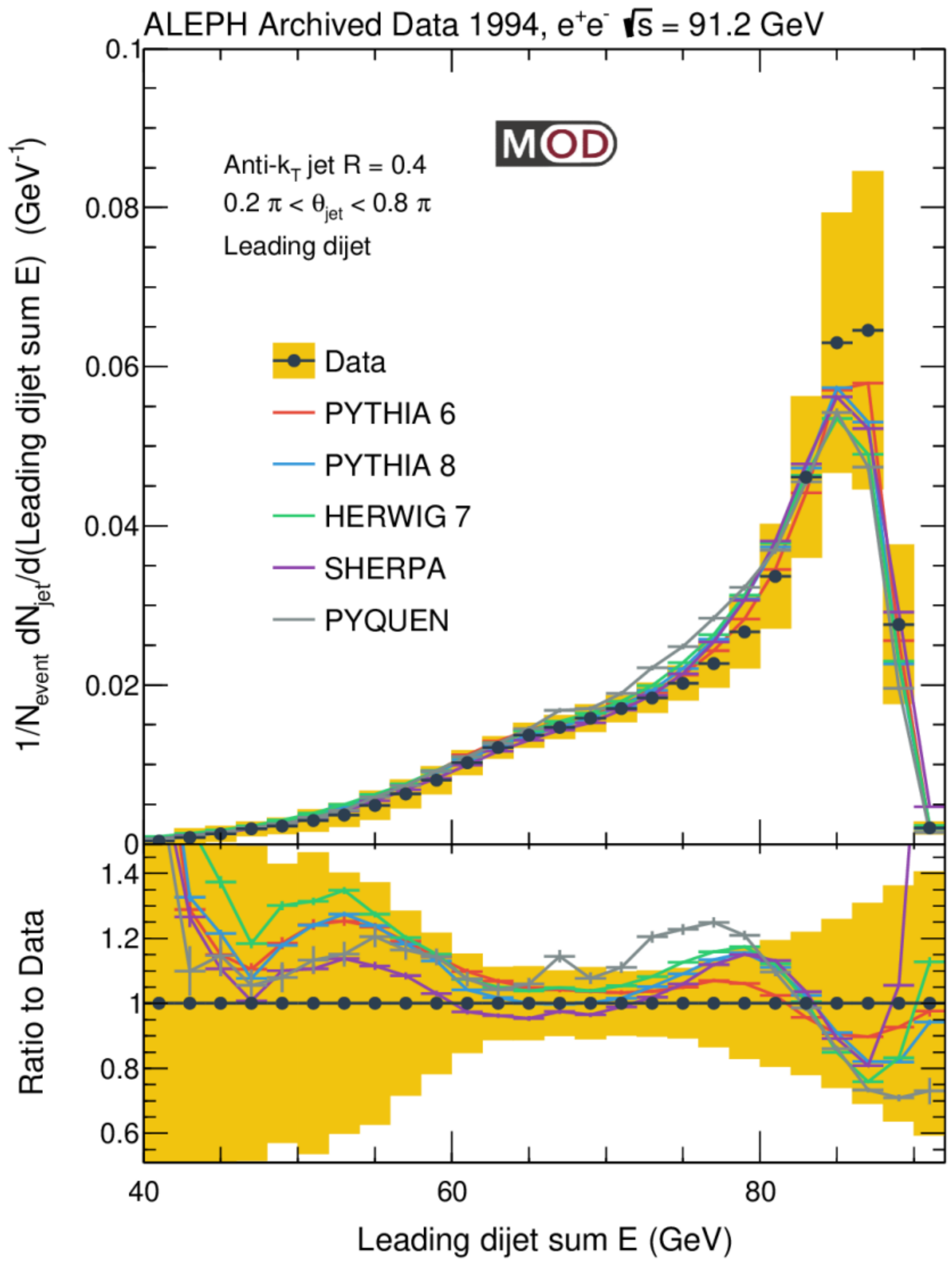
- Leading dijet energy spectrum: only include leading and subleading jets in the event (exclude other jets in the plot)
- Suppress jets from soft emissions and combinatorial jets
- A measurement of "energy loss" of (anti-)quark out of the jet cone due to parton shower
- Next-to-leading logarithmic (NLL') threshold and jet radius resummation gives a reasonable description of data



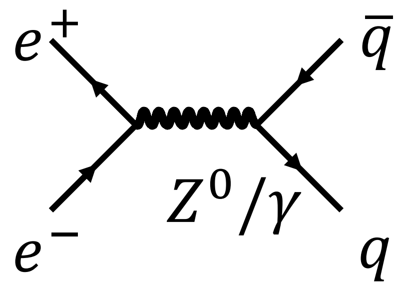


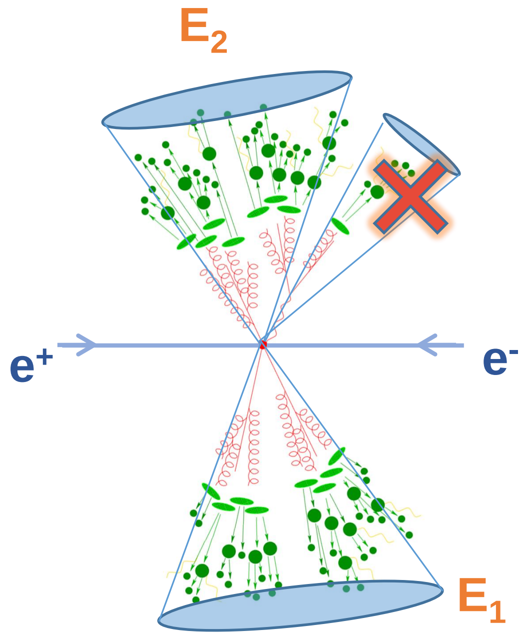
**NLL'+R calculation from Duff Neill, Felix Ringer, Nobuo Sato

Leading Dijet Energy Spectrum



- Sum of leading dijet energy (**sum E** = E_1+E_2)
- Suppress jets from soft emissions and combinatorial jets
- A measurement of "energy loss" of (anti-)quark out of the jet cone due to parton shower
- Generators capture general features of **data**

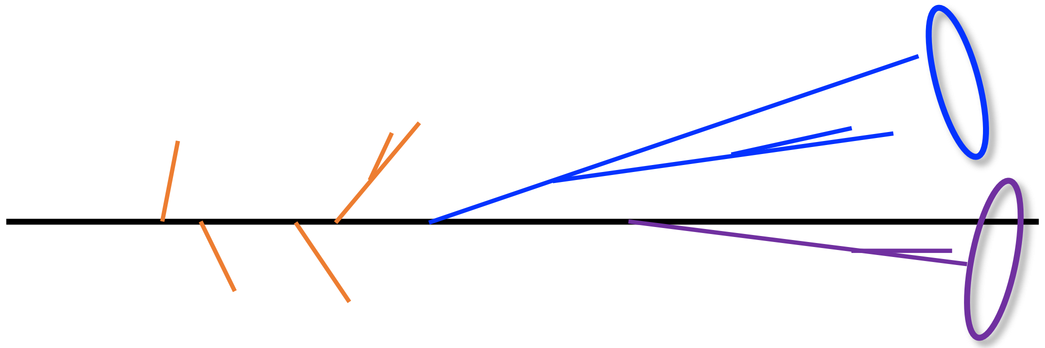




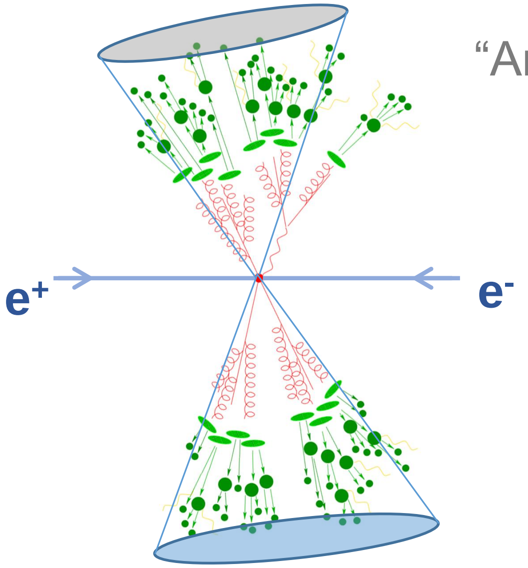
Groomed Jet Substructure with Soft Drop

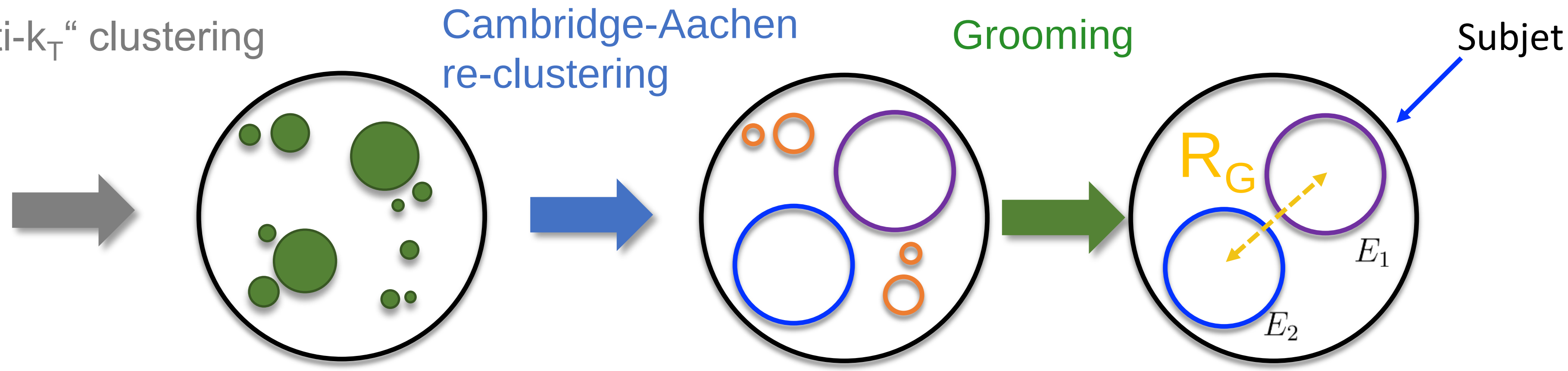
- Jet grooming: design observables **sensitive to different phase space**
  - Re-clustering with Cambridge-Aachen algorithm and grooming with a soft drop algorithm:

$$z \equiv \frac{\min(E_1, E_2)}{E_1 + E_2} \geq z_{\rm cut} \left(\frac{\theta_{12}}{R_0}\right)^\beta$$



  - θ_{12} is the **actual opening angle** between 1st and 2nd subjets
  - Soft drop setting: Z_{cut}=0.1 and β=0, identical to what are used in CMS pp and PbPb analyses
  - **β=0 → grooming independent of θ**





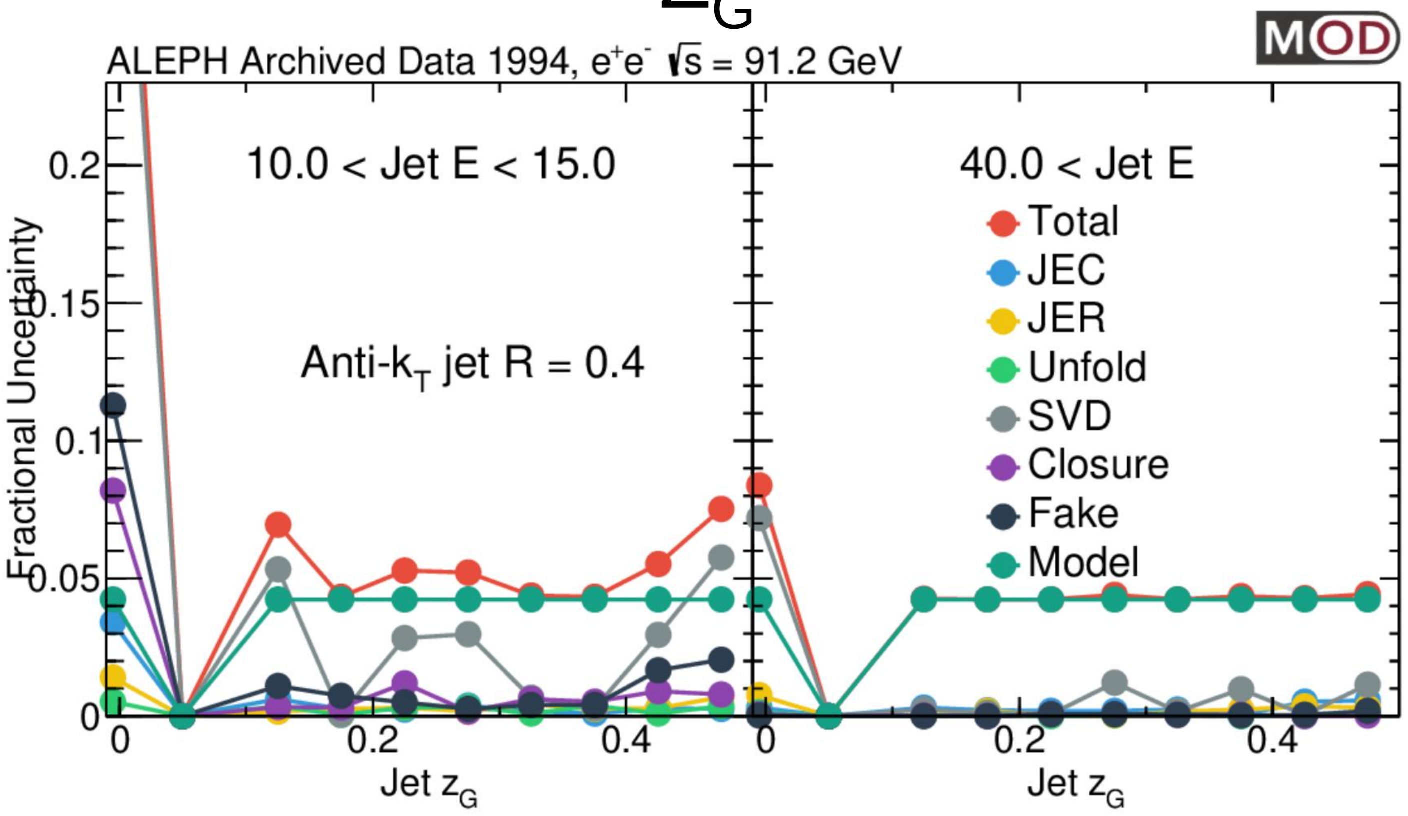
Anti-k_T jet

$$Z_G = \frac{E_1}{E_1 + E_2}$$

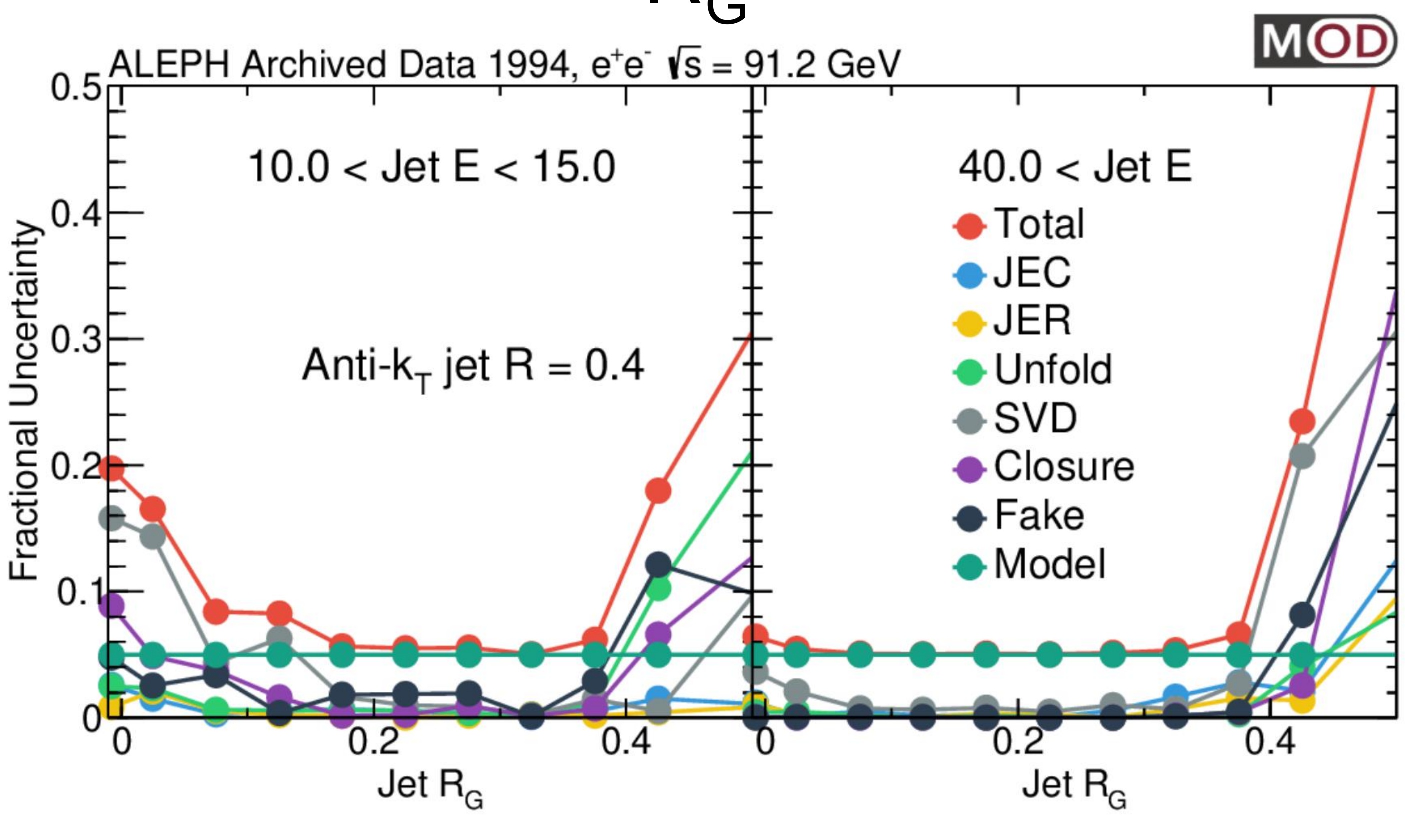
Soft Drop:
JHEP 1405 (2014) 146

Systematics: Jet Substructure

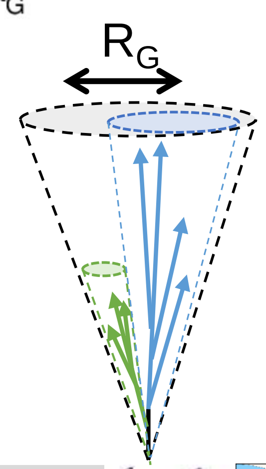
Z_G



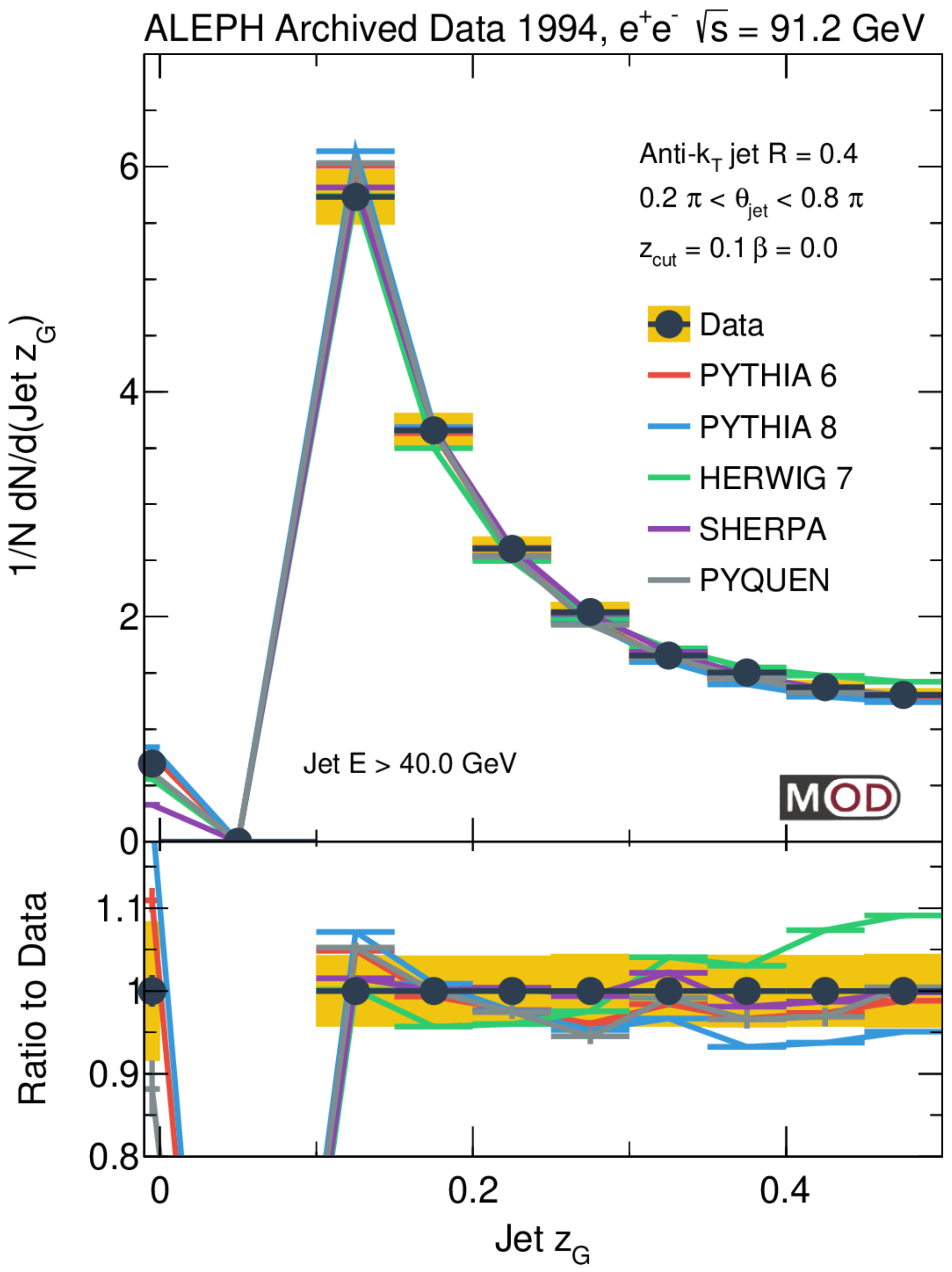
R_G

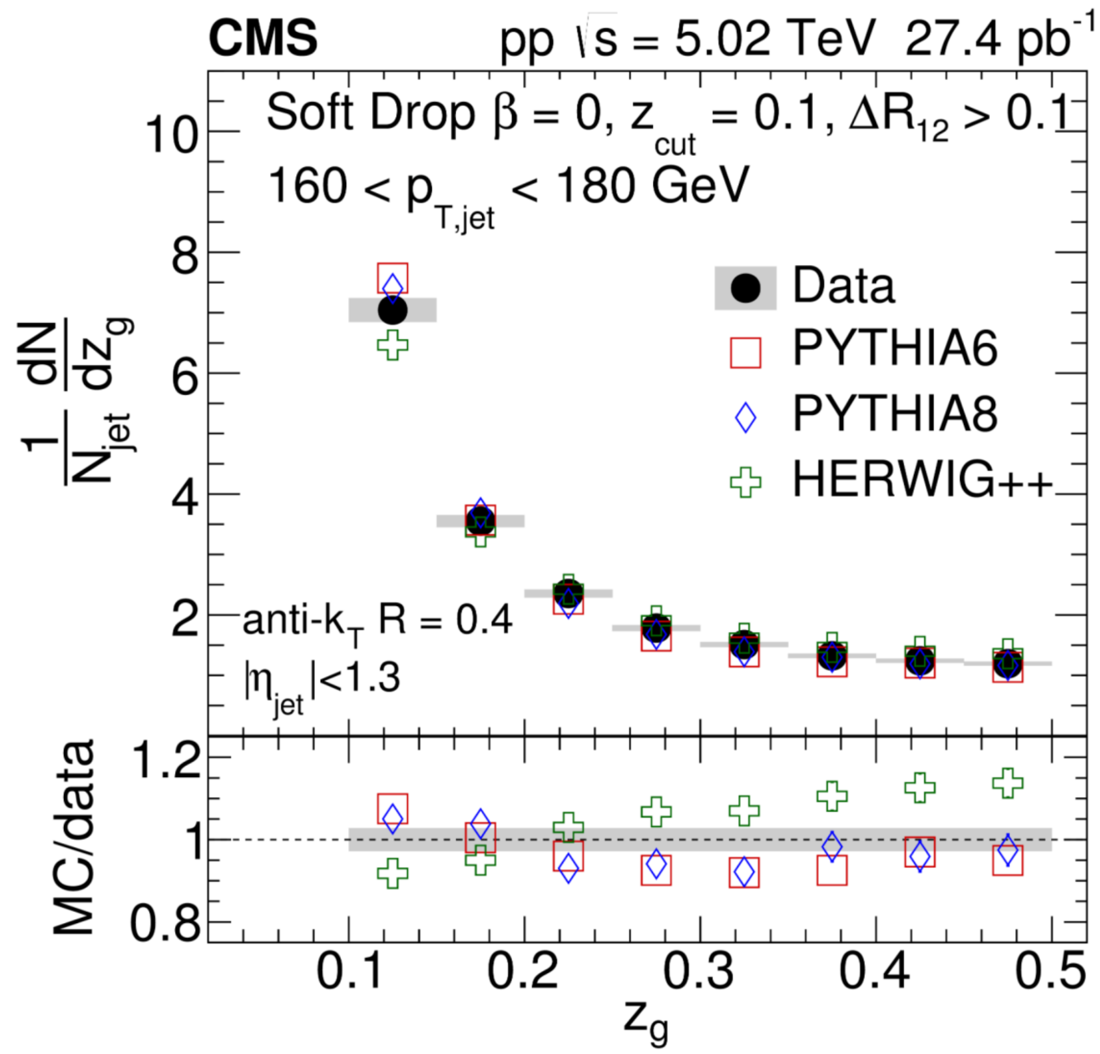


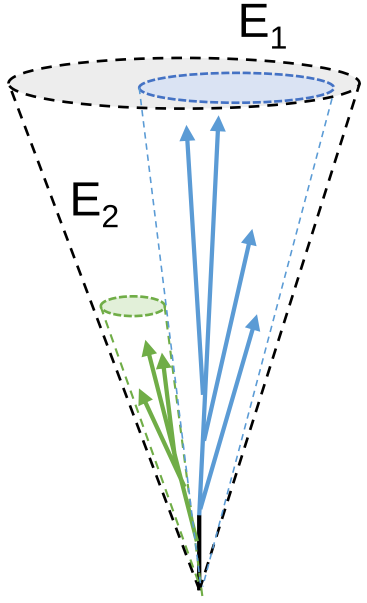
- Main uncertainty: Model dependence
  - Compare unfolded data with nominal and reweighted MC
  - Smear the subjet energy (3% on leading subjet and 6% on subleading subjet)
  - Smear the angle of subjets (by 0.002)



Groomed Momentum Sharing Z_G







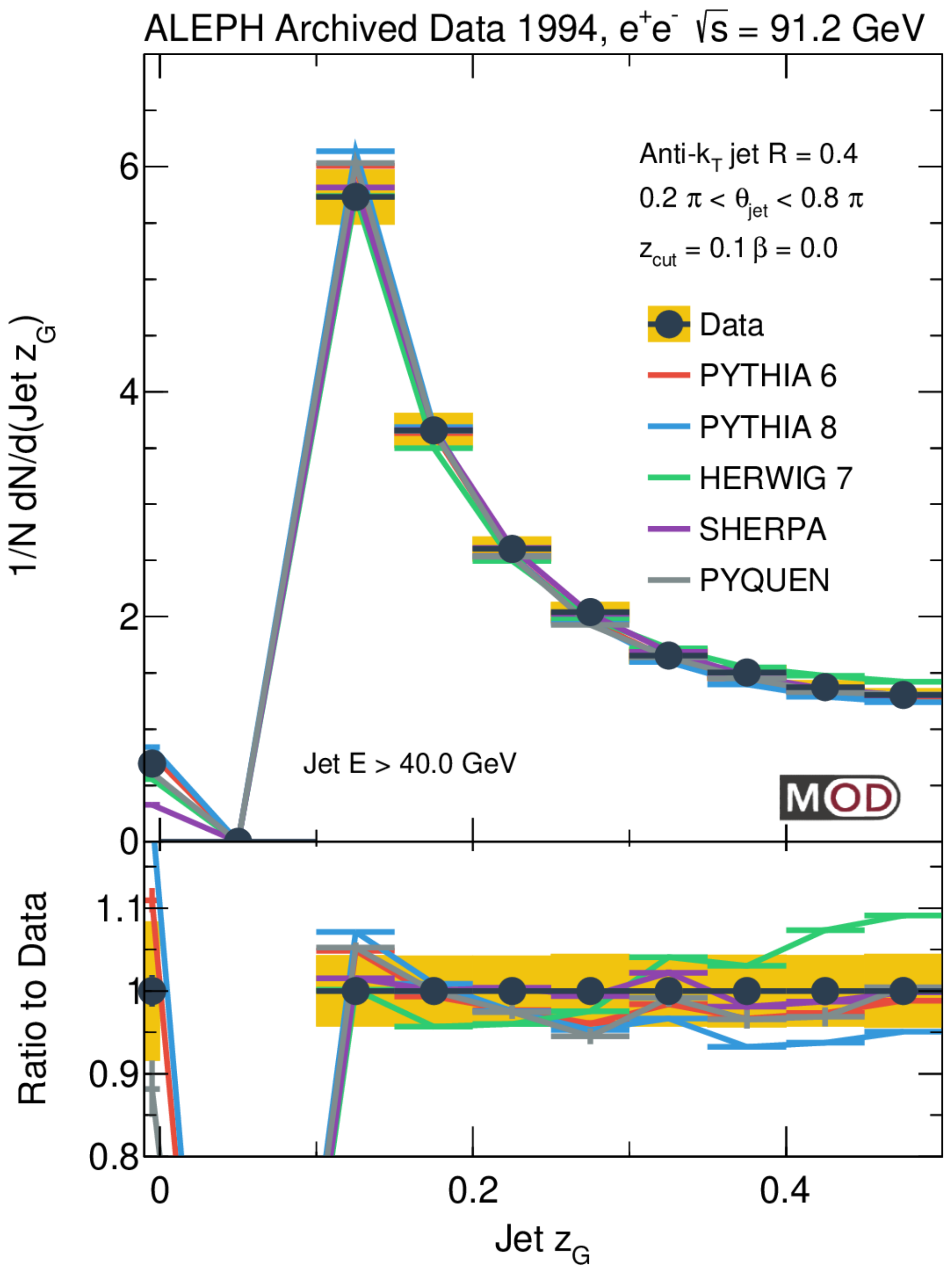
$$Z_G = \frac{E_1}{E_1 + E_2}$$

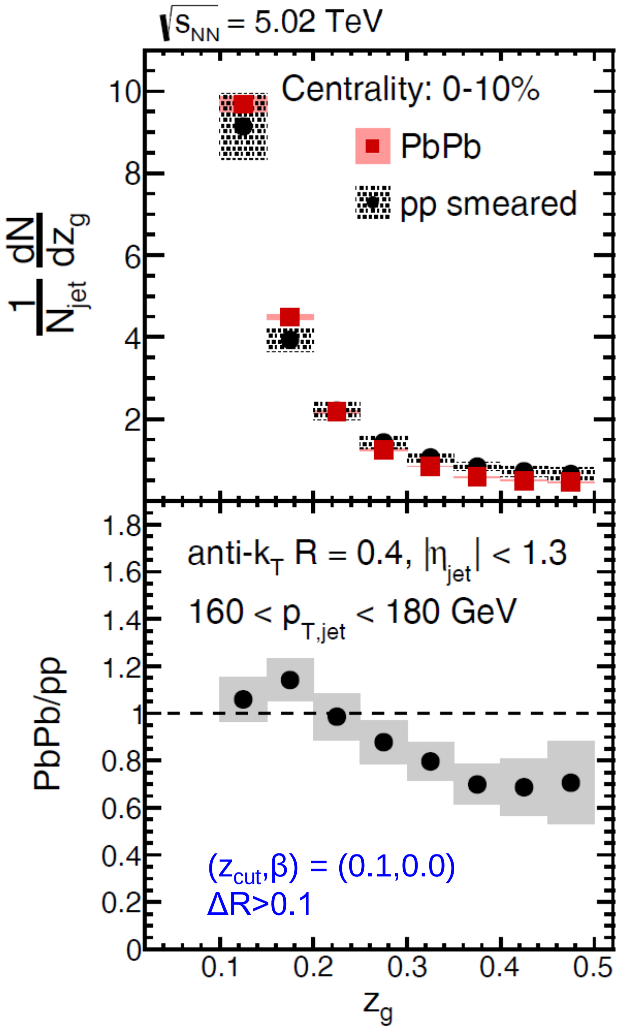
PRL 120 (2018) 14, 142302

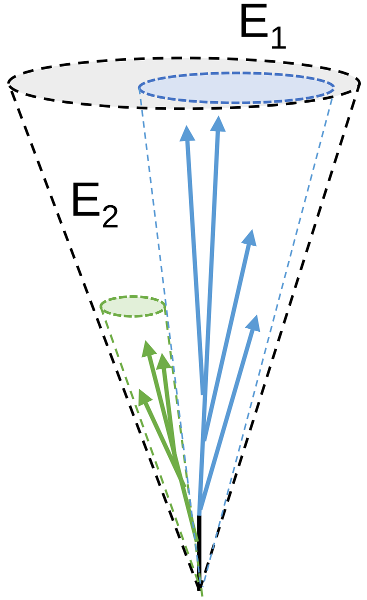
- At high jet energy, data wider than **PYTHIA 8**, narrower than **HERWIG**
- **Similar to the conclusion from pp data**

MITHIG-MOD-21-001
arXiv:2111.09914
JHEP 06 (2022) 008

Groomed Momentum Sharing Z_G







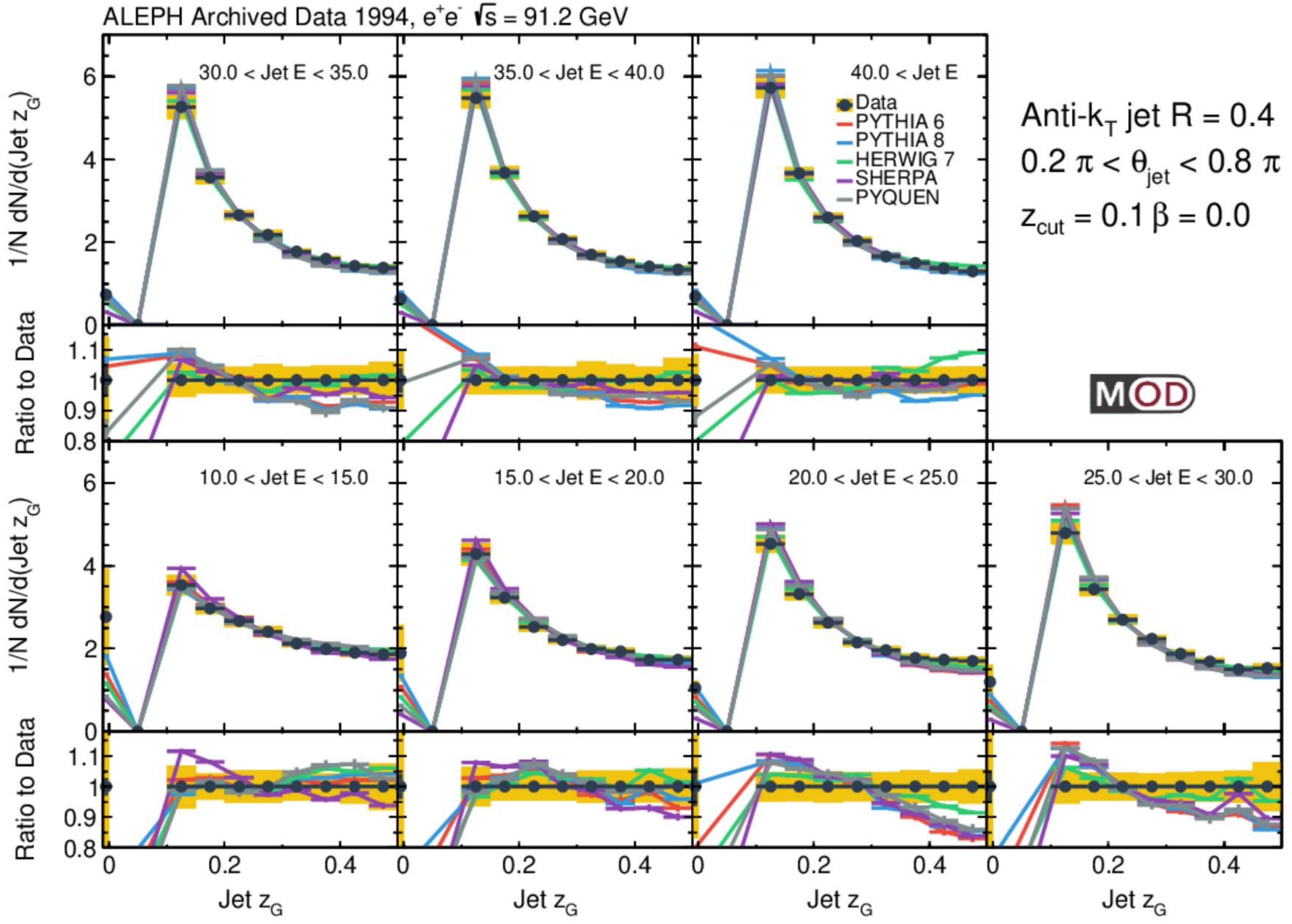
$$Z_G = \frac{E_1}{E_1 + E_2}$$

PRL 120 (2018) 14, 142302

MITHIG-MOD-21-001
arXiv:2111.09914
JHEP 06 (2022) 008

- In PbPb collisions, modification of Z_g for jets with R_g>0.1
- Size of the medium effect ~ discrepancy between pp (ee) data and event generators

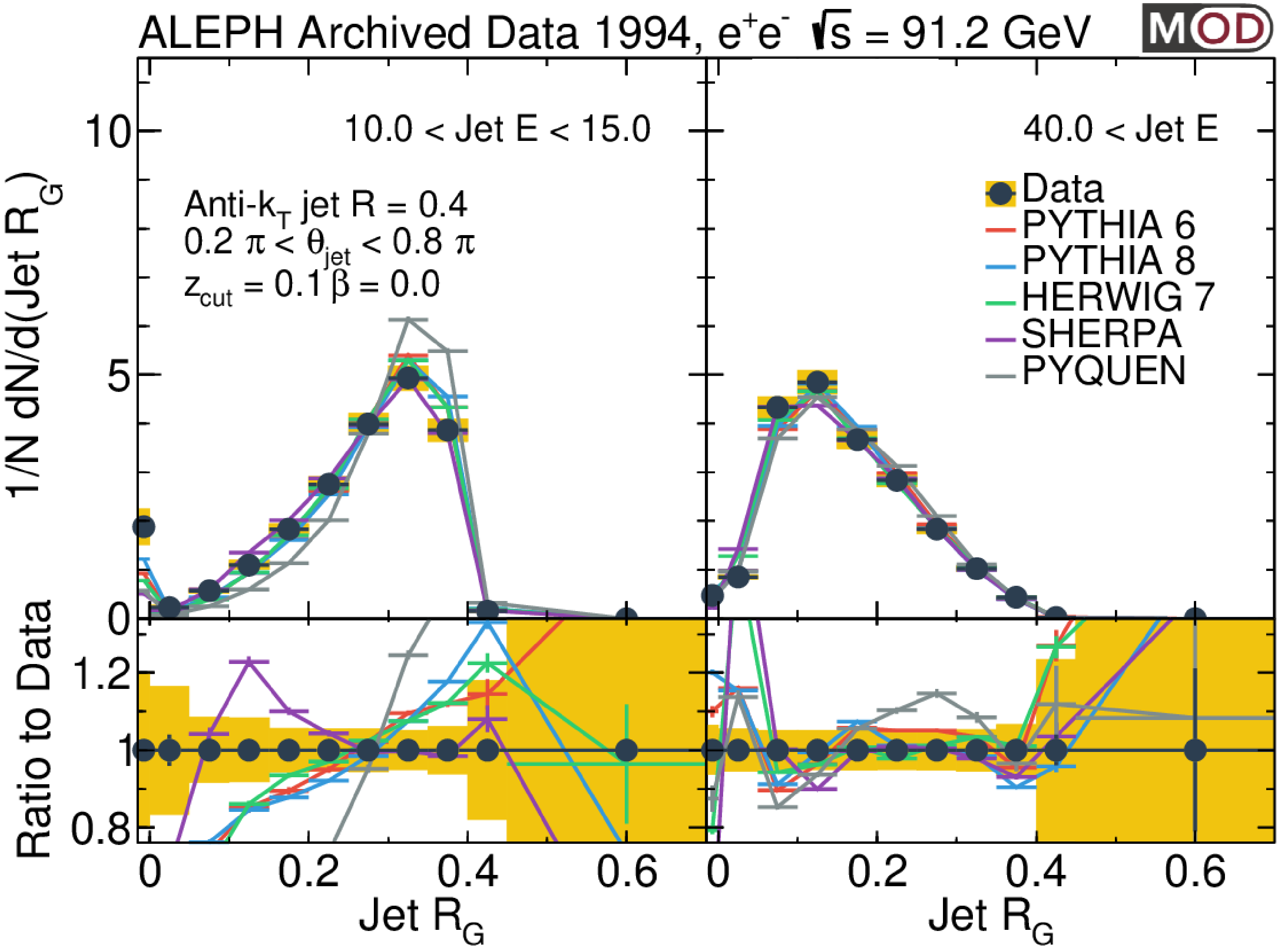
Groomed Momentum Sharing Z_G vs Jet E



- The results are compared to SHERPA, PYTHIA 6, PYTHIA 8, HERWIG, and PYQUEN
- Most of the event generators describe the general trend (wider Z_g spectra as we go to lower E)
- None of the event generators provide full description of data (discrepancy up to around 15%)

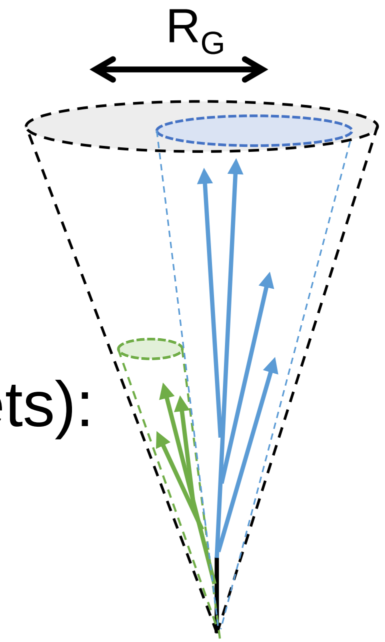
MITHIG-MOD-21-001
arXiv:2111.09914
JHEP 06 (2022) 008

Groomed Jet Radius R_G vs. Event Generators



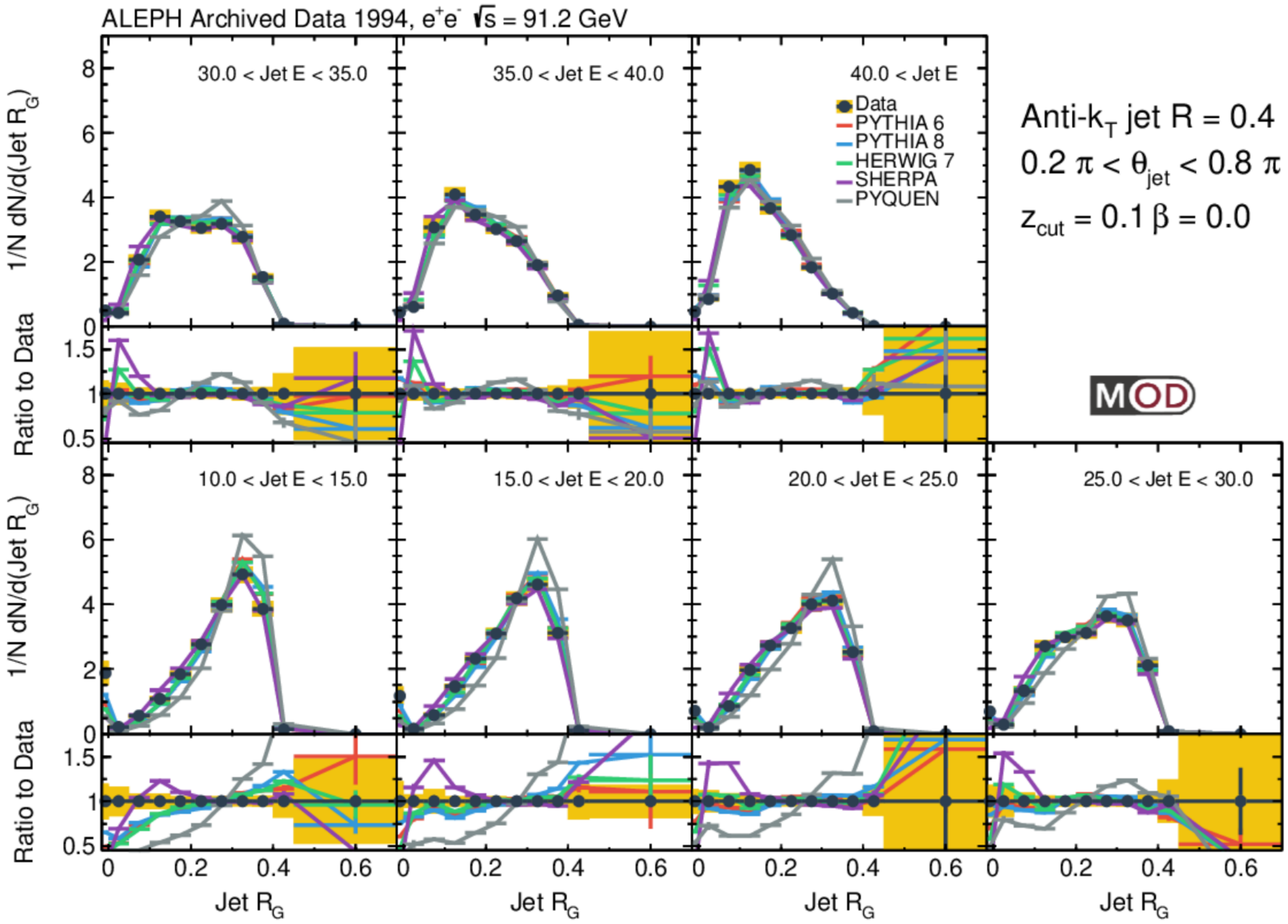
Low jet E **High jet E**

MITHIG-MOD-21-001
arXiv:2111.09914
JHEP 06 (2022) 008

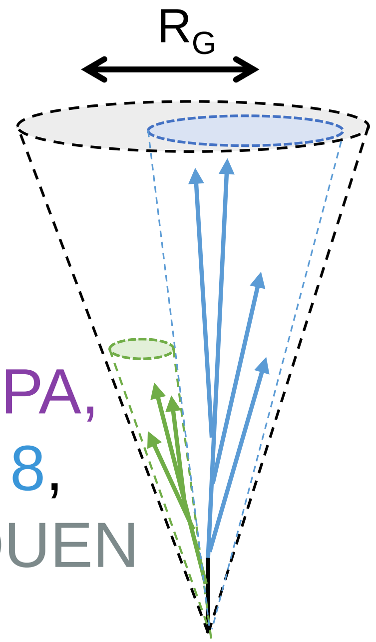


- **High jet E** (mainly quark jets):
  - Peak at smaller R_G value
  - Generators give a better description of the data
- **Low jet E** (mainly from soft emissions and combinatorial):
  - Peak at larger R_G value as one would expect
  - SHERPA gives a better description of the data
  - PYTHIA 6, PYTHIA 8, HERWIG, and PYQUEN overpredict the R_G

Groomed Jet Radius R_G vs. Event Generators

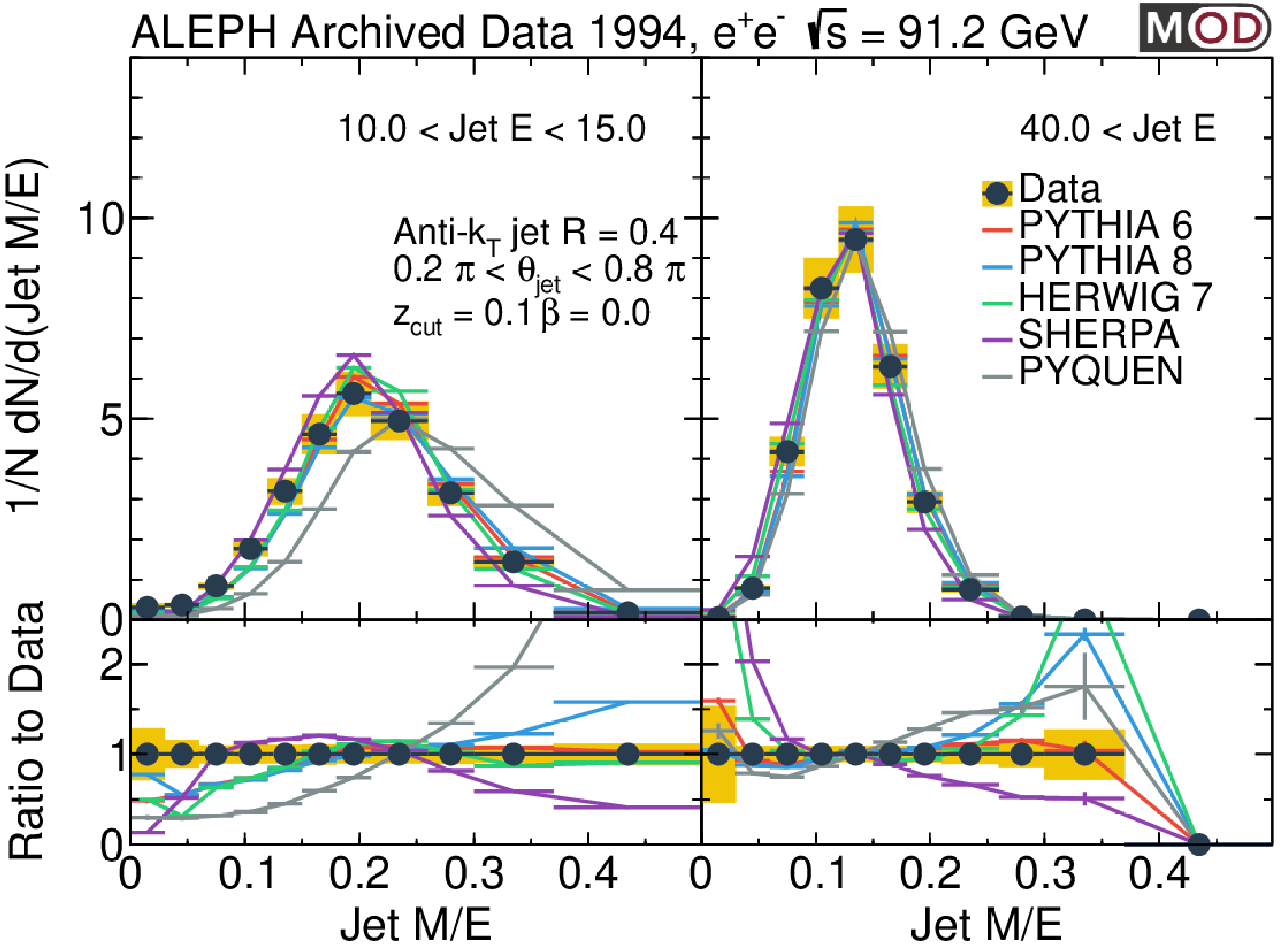


MITHIG-MOD-21-001
arXiv:2111.09914
JHEP 06 (2022) 008



- The results are compared to SHERPA, PYTHIA 6, PYTHIA 8, HERWIG, and PYQUEN
- Groomed jet radius decreases as a function of jet energy
- Archived **PYTHIA 6.1** is closer to the data
- None of the event generators provide full description of data (discrepancy up to ~50%)

Jet Mass / Energy Ratio



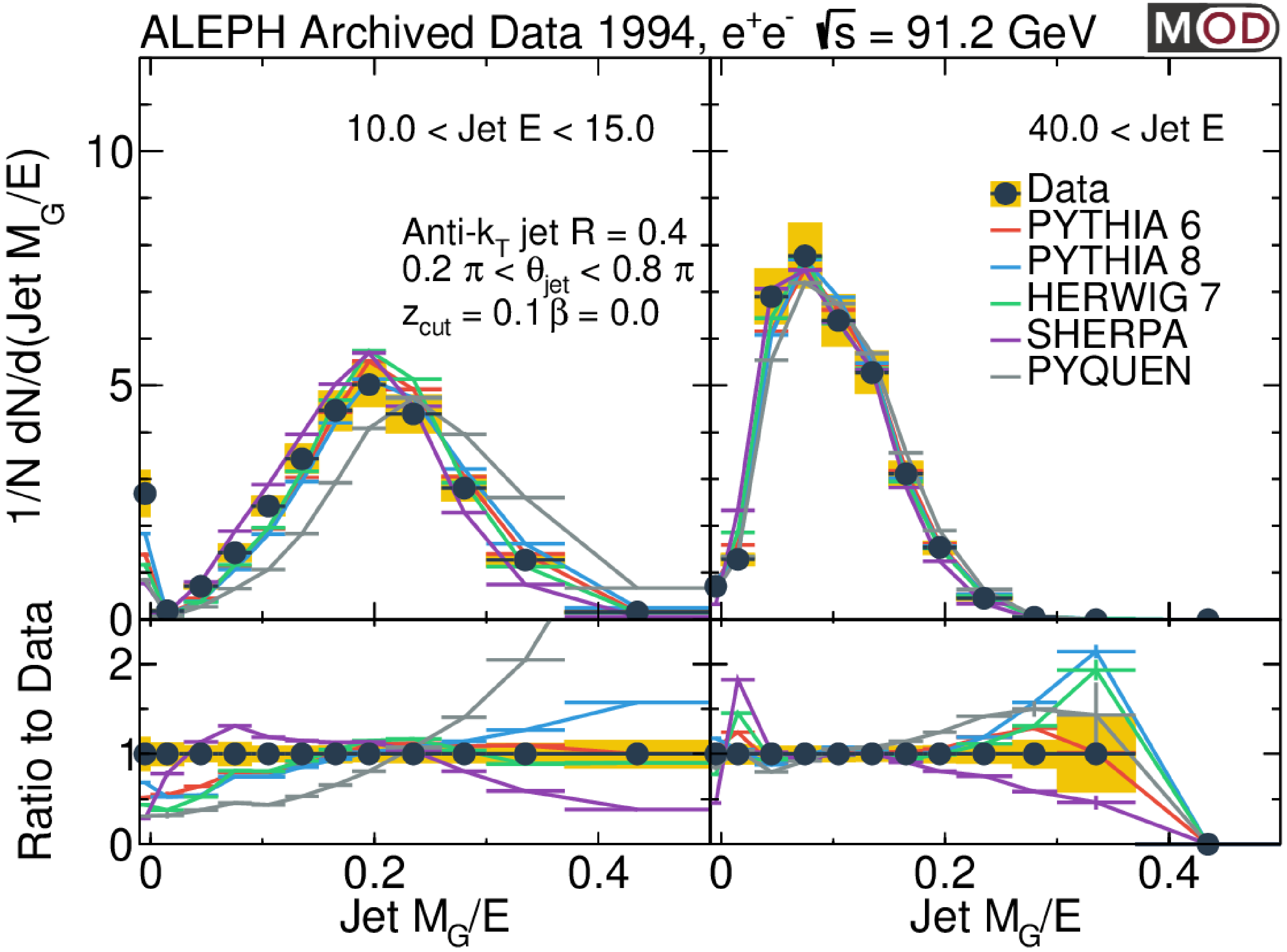
Low jet E **High jet E**

Jet M/E:

- SHERPA, PYTHIA 8, HERWIG, and PYQUEN don't describe the distribution well.
- Archived **PYTHIA 6.1** is closer to the data
- SHERPA: lower mass than other generators

MITHIG-MOD-21-001
arXiv:2111.09914
JHEP 06 (2022) 008

Groomed Jet Mass / Energy Ratio



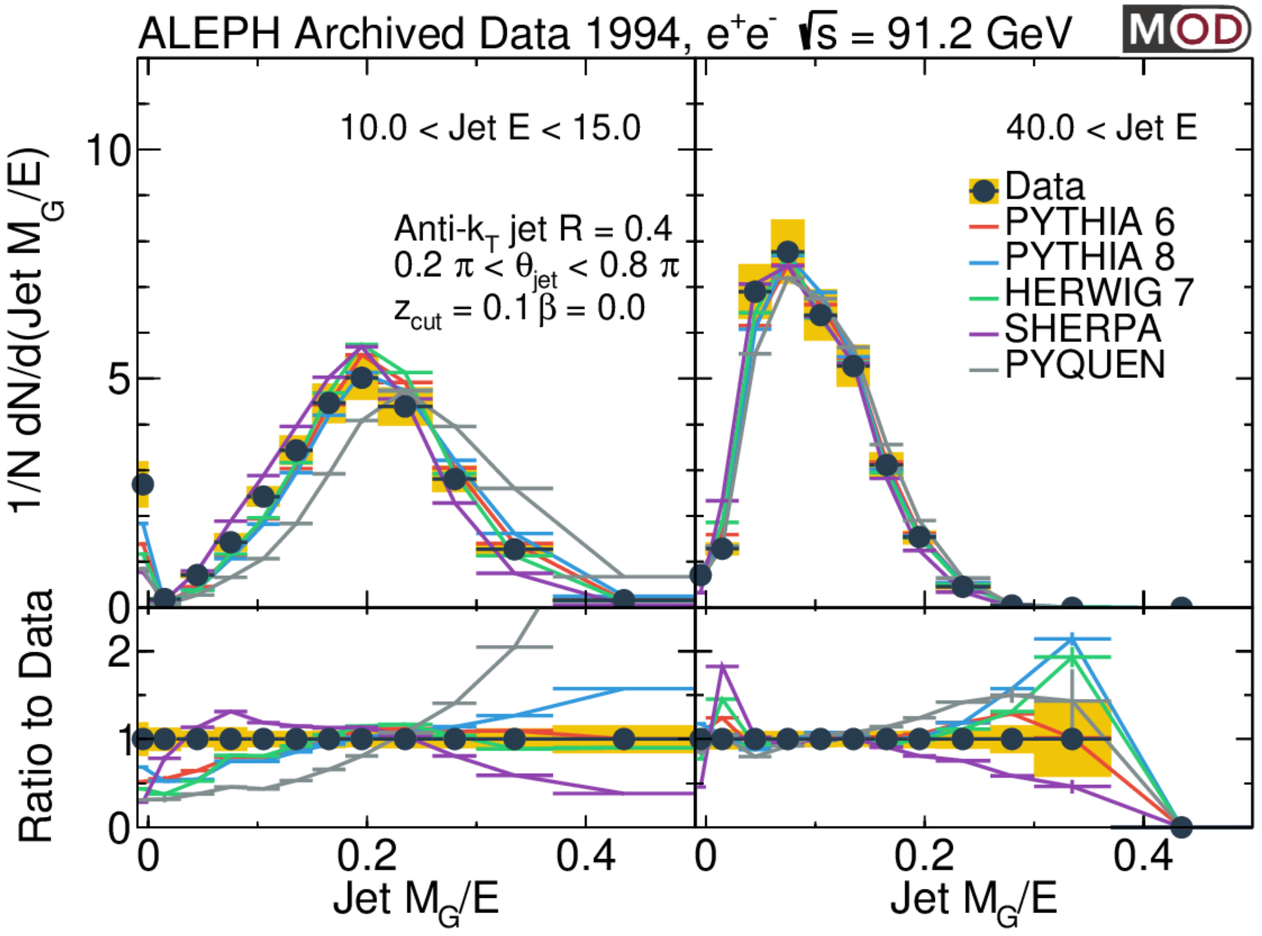
Low jet E **High jet E**

Groomed Jet M_G/E:

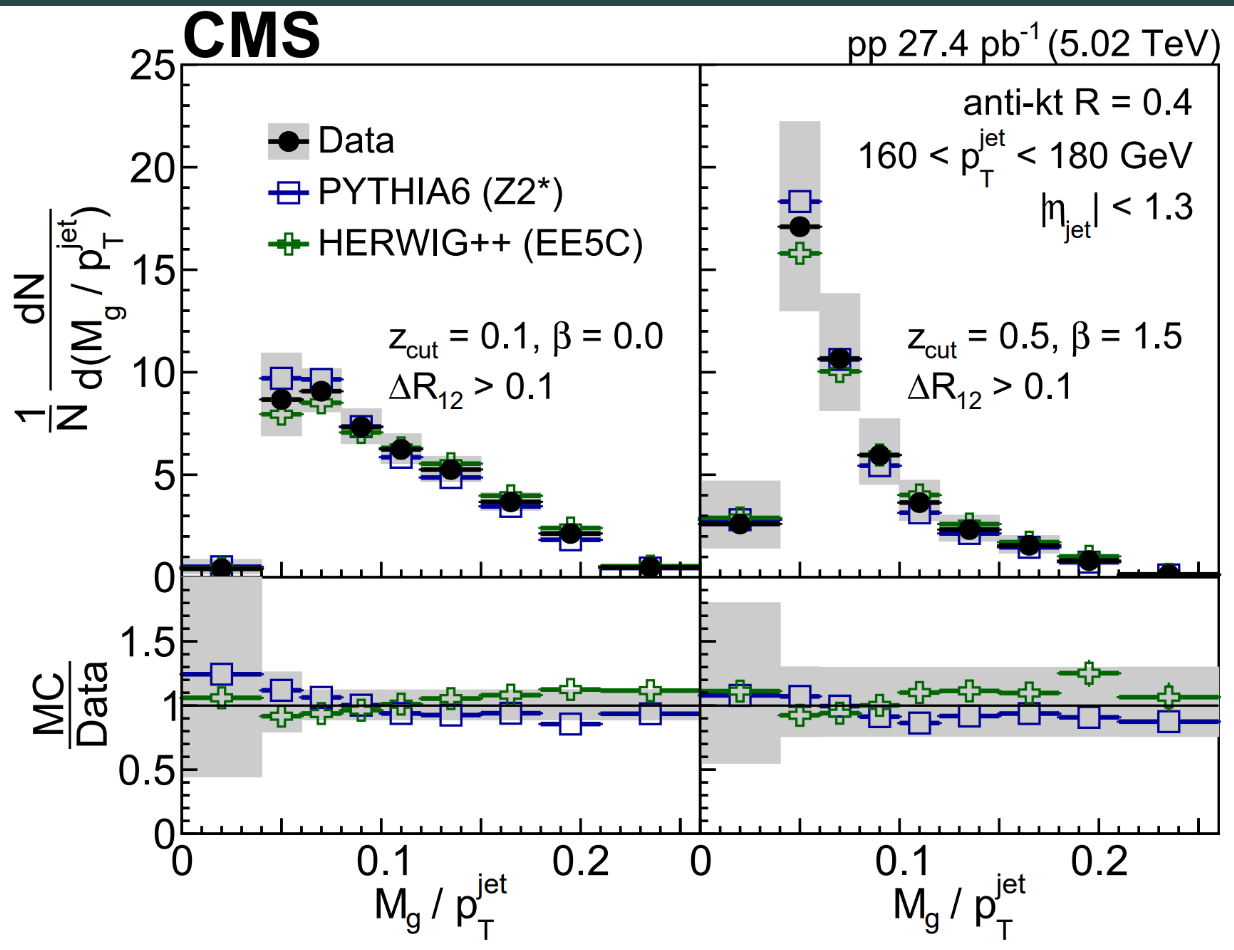
- Reduce the impact of soft radiation in the jet cone
- Jet mass becomes smaller after grooming
- Similar discrepancies seen between data and generators as the ungroomed mass
- To enhance the sensitivity to soft emissions: Follow up with **(M-M_G)/E** and collinear drop observables in the future

MITHIG-MOD-21-001
arXiv:2111.09914
JHEP 06 (2022) 008

Groomed Jet Mass in e^+e^- and pp



MITHIG-MOD-21-001
[arXiv:2111.09914](https://arxiv.org/abs/2111.09914)
JHEP 06 (2022) 008

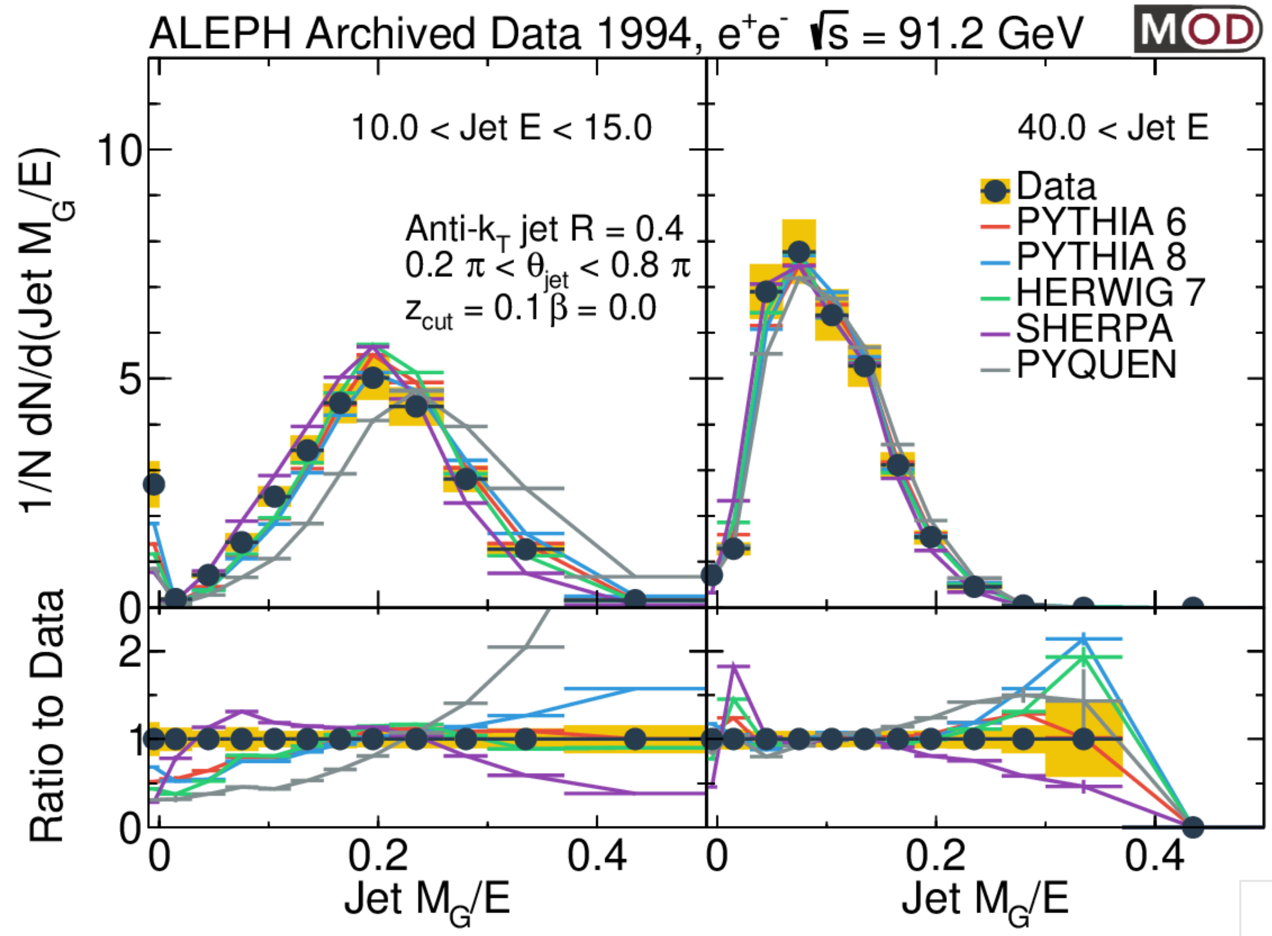


JHEP 10 (2018) 161

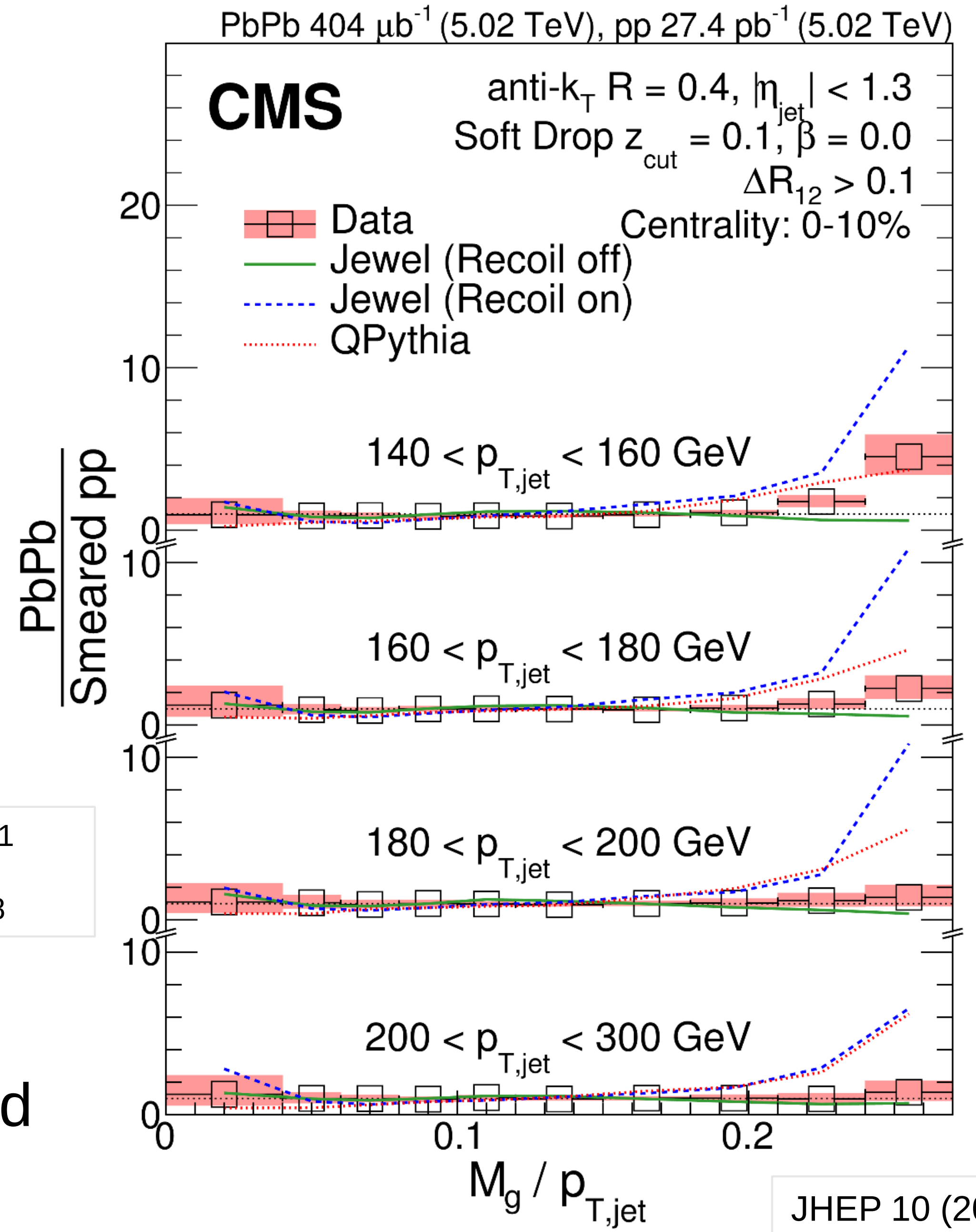
- Large difference between e^+e^- data and event generators

- Within the experimental uncertainties, event generators agree with pp data
- Cancellation of effects from MPI and shower?

Groomed Jet Mass in e^+e^- and PbPb



MITHIG-MOD-21-001
[arXiv:2111.09914](https://arxiv.org/abs/2111.09914)
JHEP 06 (2022) 008

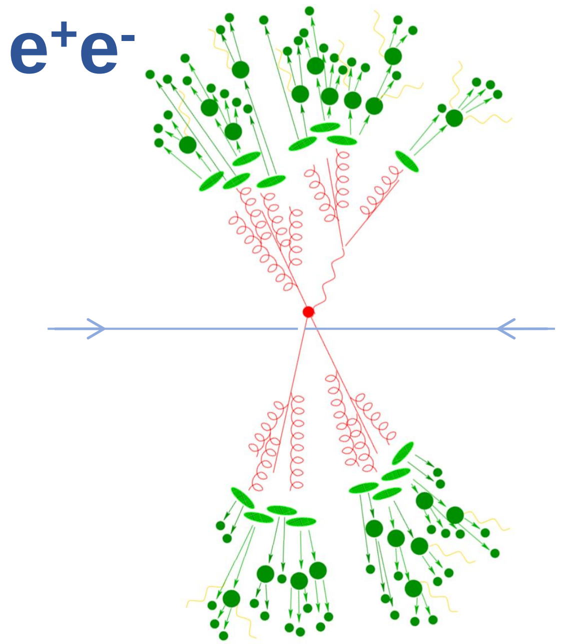


JHEP 10 (2018) 161

- PbPb: modification mass spectra
  - **In particular, at high and low M_g/p_T regions**
- Also the phase space where e^+e^- data is poorly described by event generators

From e^+e^- to EIC

- e^+e^- jet data provide opportunities to validate theoretical approaches and tuning of event generator
  - High accuracy pQCD calculations
  - Hadronization
- Provide tests on EIC-inspired jet algorithms
  - For instance, Centauro
- Provide tests on the new observables
  - Such as EC and CF
- Reference for ep and eA data at EIC



Summary

- The first measurement of anti-k_T jet energy spectra and substructure in hadronic e^+e^- collisions

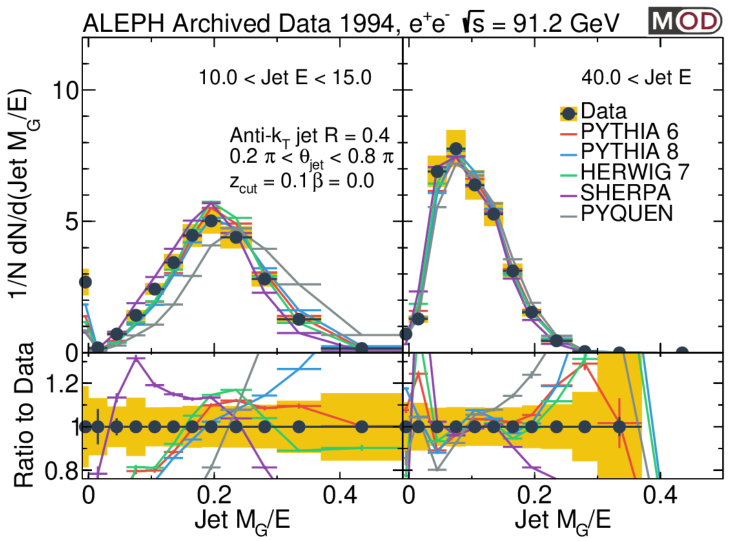
- None of the event generators gives full description of the e^+e^- data (discrepancy as large as 10-50%)

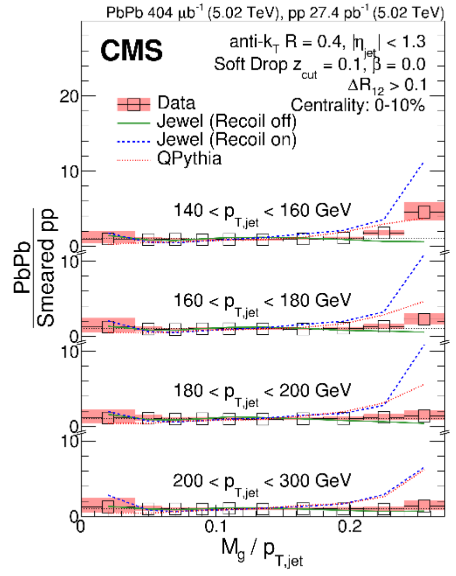
- Discrepancy between data and generators in e^+e^- is as large as the difference between pp and PbPb in places

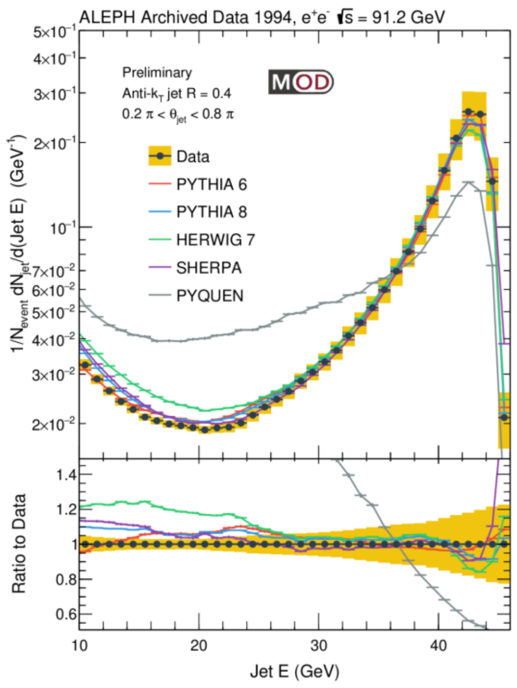
- Provide new inputs to event generators and hadronization models

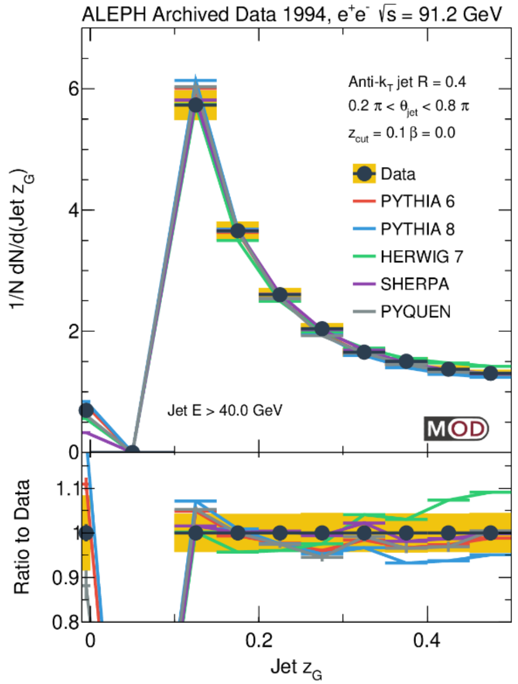
- We look forward to more future analyses with this dataset such as:
  - Jet substructure vs. event multiplicity
  - Jet fragmentation function
  - Strangeness content
  - EIC jet algorithm and observables

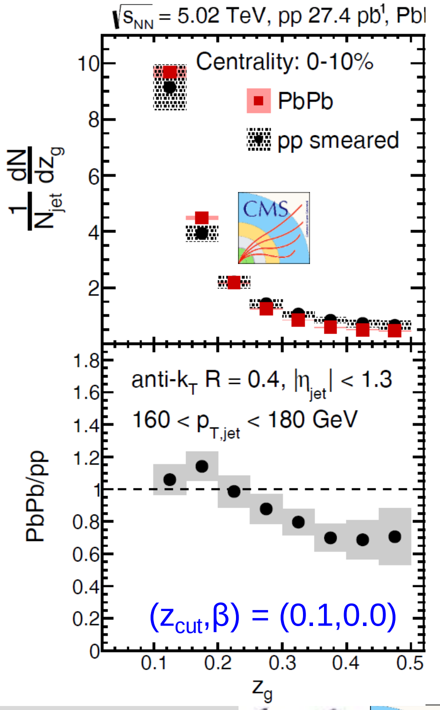
… stay tuned!











CMS Z_G PRL 120 (2018) 14, 142302

ALICE R_G arXiv:2107.12984

Acknowledgement

We would like to thank **Roberto Tenchini** and **Guenther Dissertori** from the ALEPH collaboration for the useful comments and suggestions on the use of ALEPH archived data.

We would like to thank **Felix Ringer**, **Jesse Thaler, Andrew Lakoski, Liliana Apolinário, Ben Nachman, Camelia Mironov, Jing Wang** for the useful discussions on the analysis.

The MIT group's work was supported by US DOE-NP

Thank you!

Anthony Badea
(Harvard, ATLAS)

Austin Baty
(Rice, CMS)

Chris McGinn
(CU Boulder, ATLAS)

Michael Peters
(MIT, CMS)

Janice Chen
(MIT, BELLE/CMS)

Cheng-Wei Lin
(NTU, BELLE)

Dennis Perepelitsa
(CU Boulder, ATLAS)

Tzu-An Sheng
(MIT, CMS)

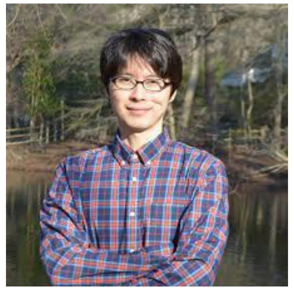
Patrick T. Komiske III
(MIT, CTP)

Eric Metodiev
(MIT, CTP)

Jing Wang
(MIT, CMS)

Ben Nachman
(LBNL, ATLAS)

Yi Chen
(MIT, CMS)

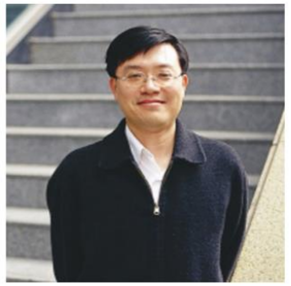


Yang-Ting Chien
(GSU)

Yen-Jie Lee
(MIT, CMS)

Gian Innocenti
(CERN, ALICE)

Jesse Thaler
(MIT, CTP)

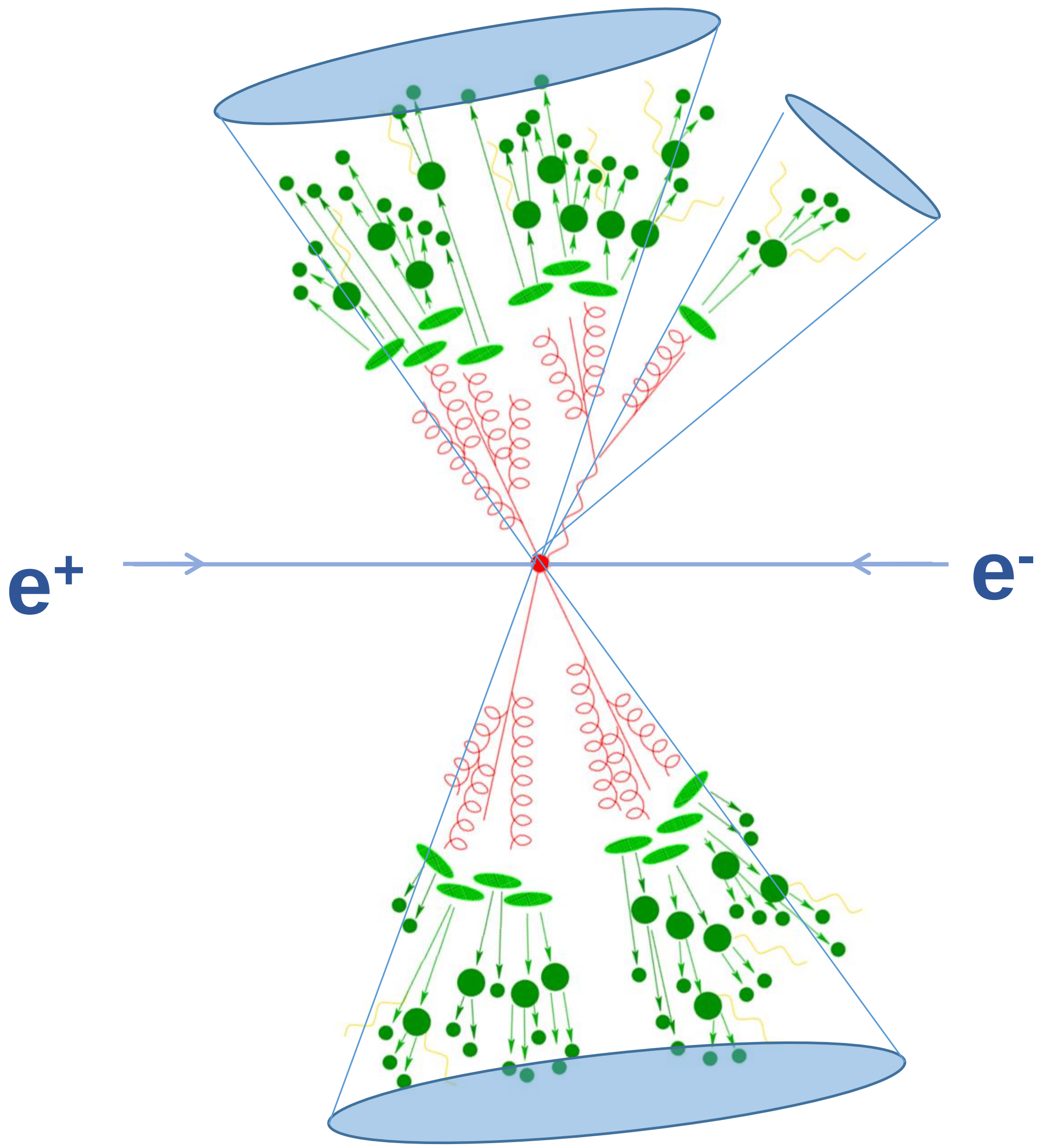


Paoti Chang
(NTU, BELLE/CMS)

Marcello Maggi
(INFN, CMS)

Günther Dissertori
(ETH Zürich, CMS)

Backup Slides



Z^0 Hadronic Decay Event Selection

- **Track Selection:**
  - Particle Flow Candidate 0, 1, 2
  - Number of TPC hits for a charged tracks >= 4
  - $|d0| < 2$ cm
  - $|z0| < 10$ cm
  - $|\cos\theta|<0.94$
  - $p_T > 0.2$ GeV (transverse momentum with respect to beam axis)
  - $N_{TPC} >=4$
  - $x^2/ndf < 1000.$

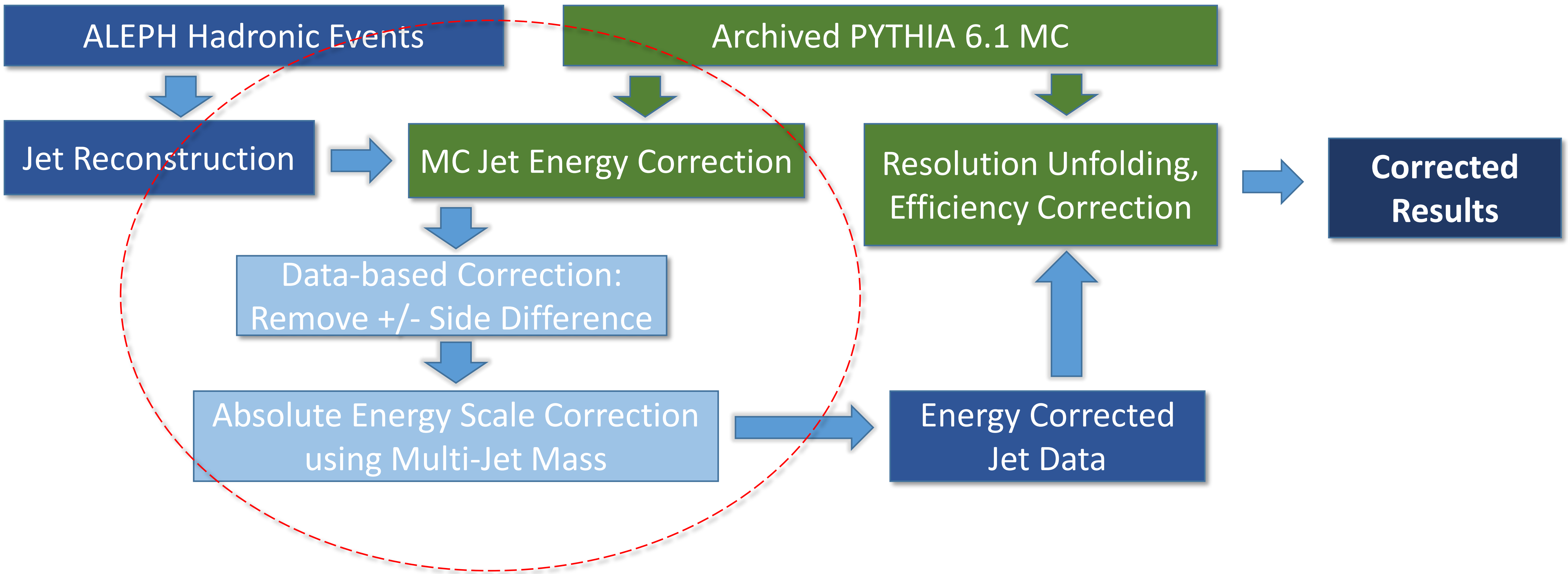
- **Neutral Hadron Selection:**
  - Particle Flow Candidate 4, 5 (ECAL / HCAL object)
  - $E > 0.4$ GeV
  - $|\cos\theta|<0.98$

- **Event Selection:**
  - Number of good charged particles >= 5 (including charged hadrons and leptons)
  - Number of good ch+neu. Particles >= 13
  - $E_{charged} > 15$ GeV
  - $|\cos(\theta_{sphericity})|<0.82$

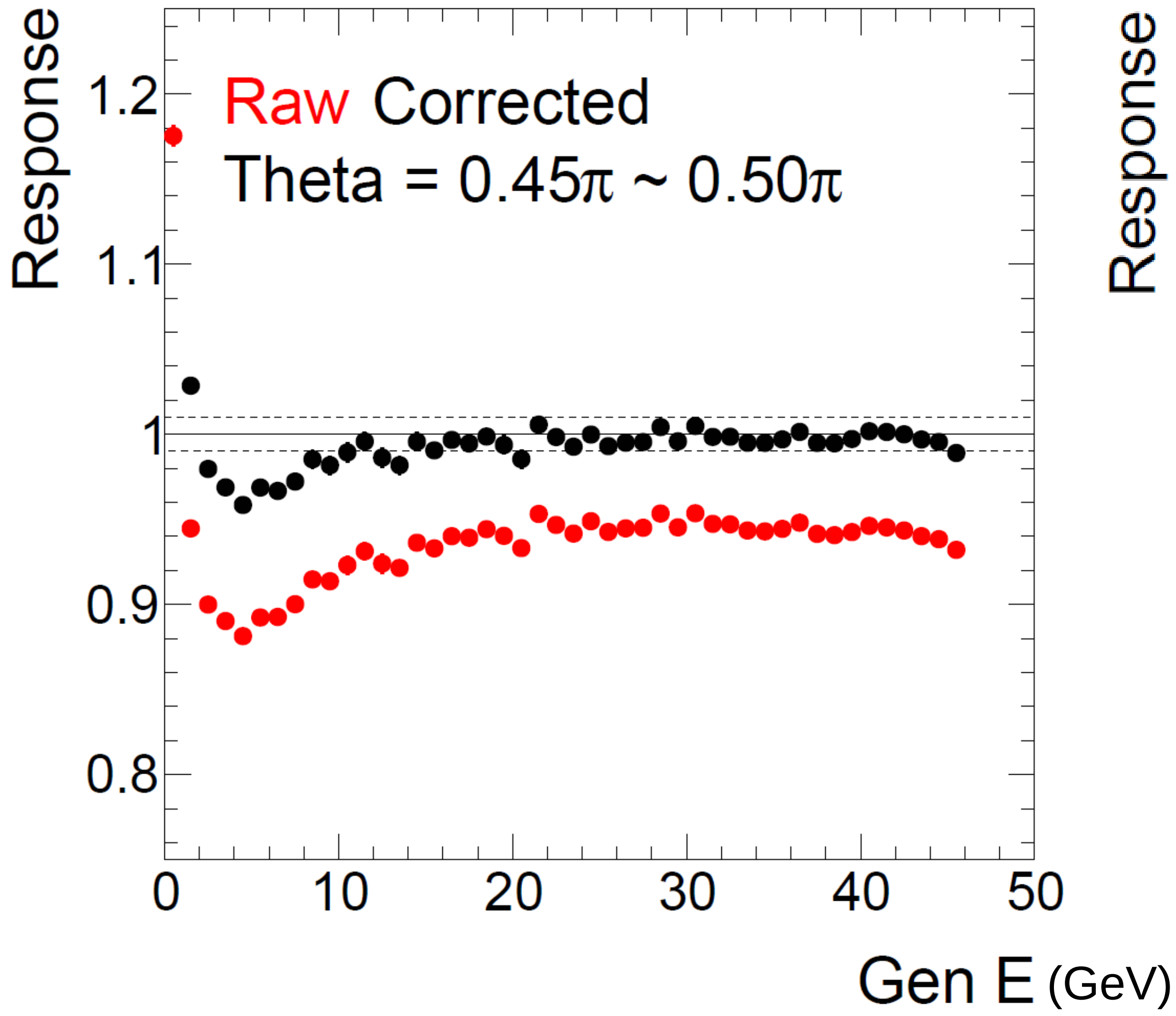
Jet Energy Correction Steps

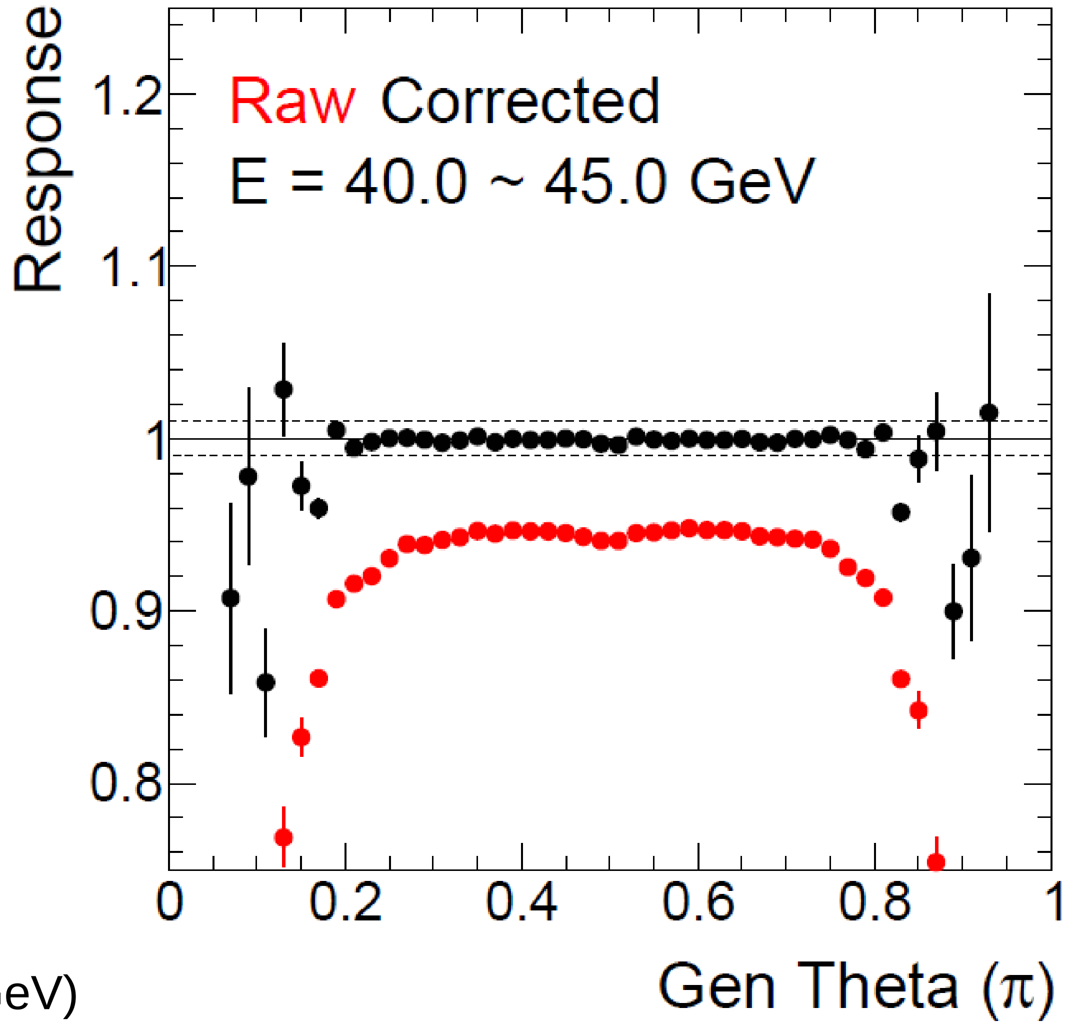
In order to inspect the event in a shower-by-shower basis, we need to perform jet reconstruction

Jet analysis workflow:



MC Jet Energy Correction

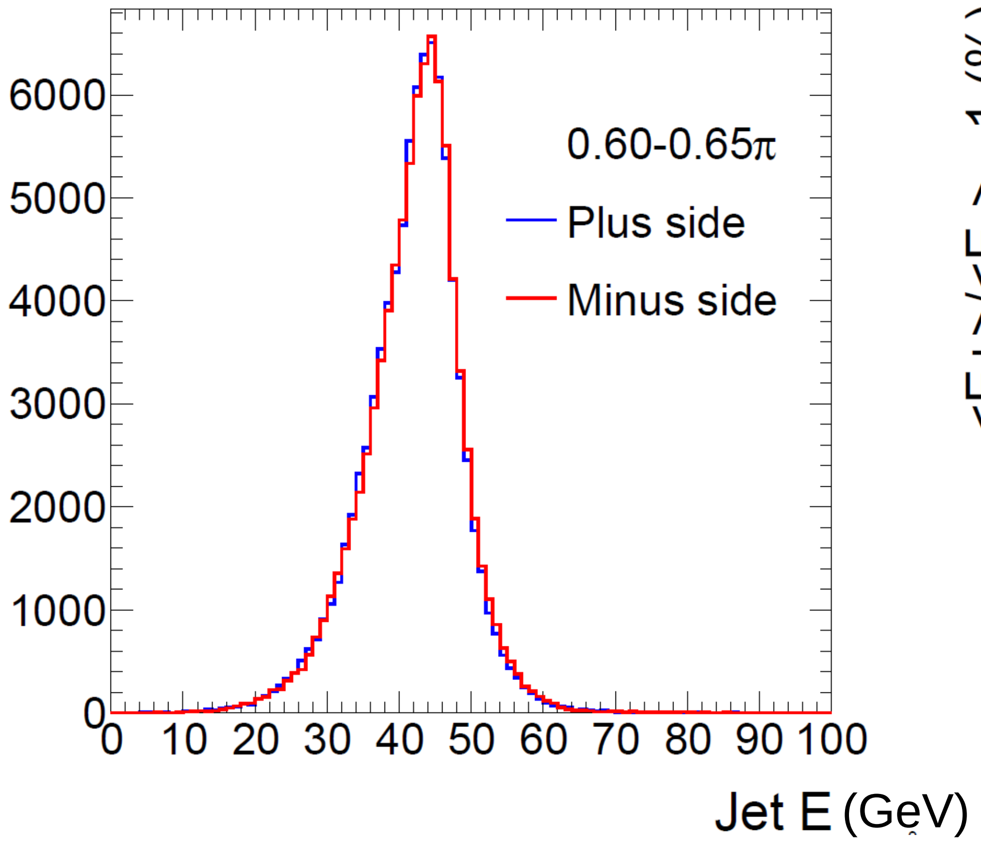


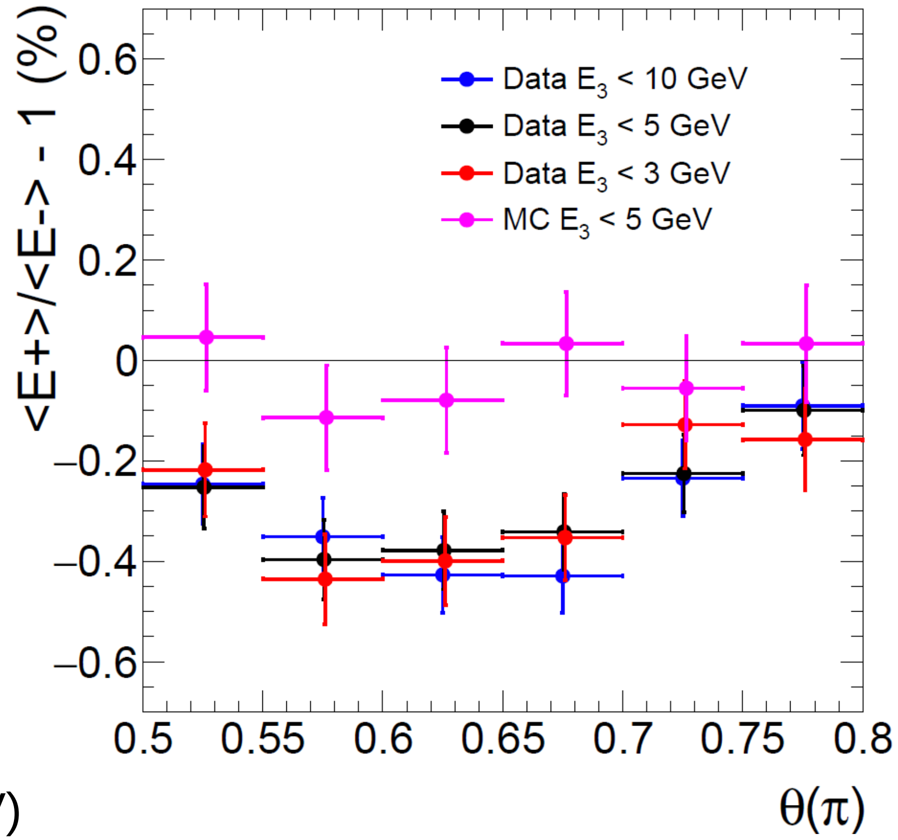


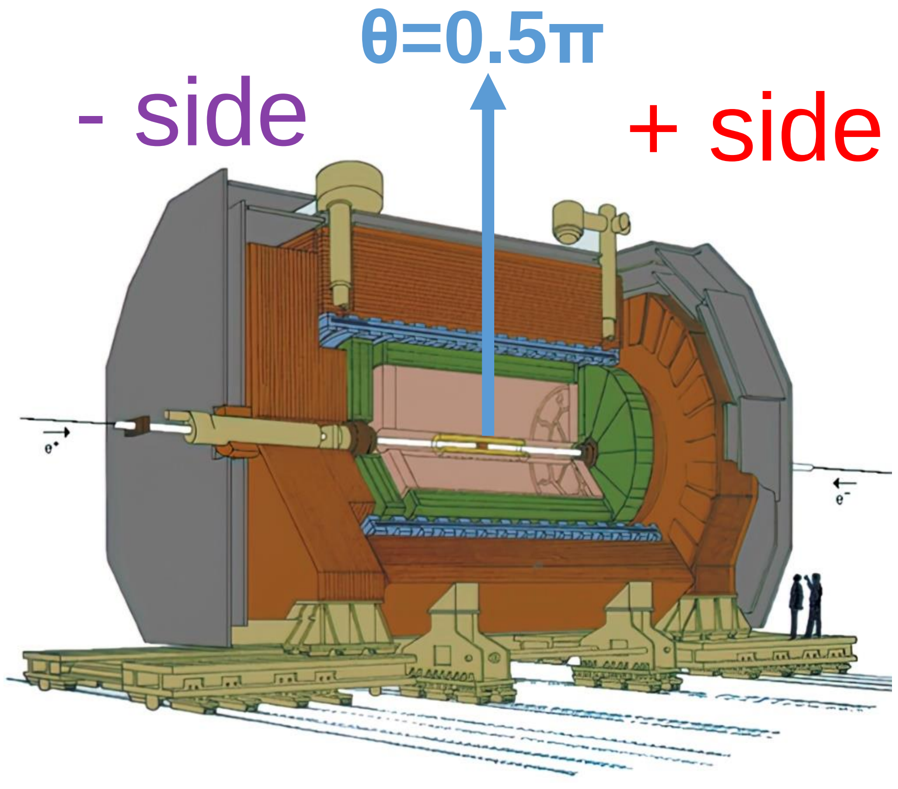
Yi Chen
(MIT, CMS)

- **Archived PYTHIA 6.1 MC** (tuned to describe ALEPH hadron spectra) was used for jet energy calibration
- **Response** = (**Raw** or **Corrected**) Reconstructed jet energy **/** “Generated jet energy” in truth level
- Good closure was achieved for **E>10 GeV** and **0.2π<Jet θ<0.8π**

Data-based Calibration: Relative Plus/Minus Difference



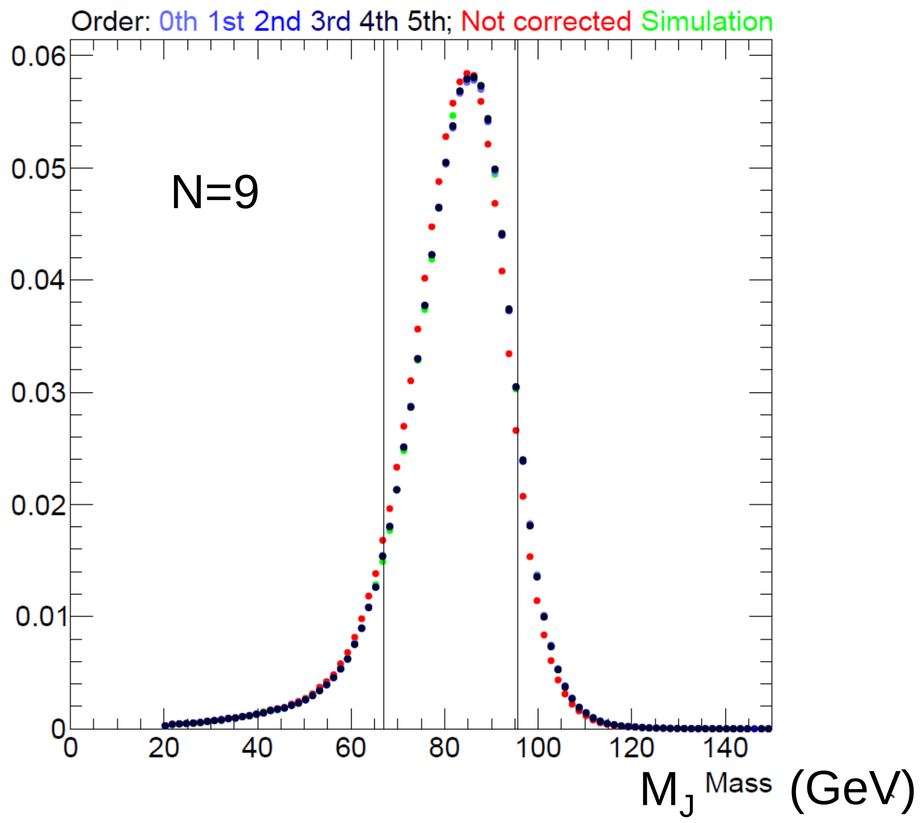


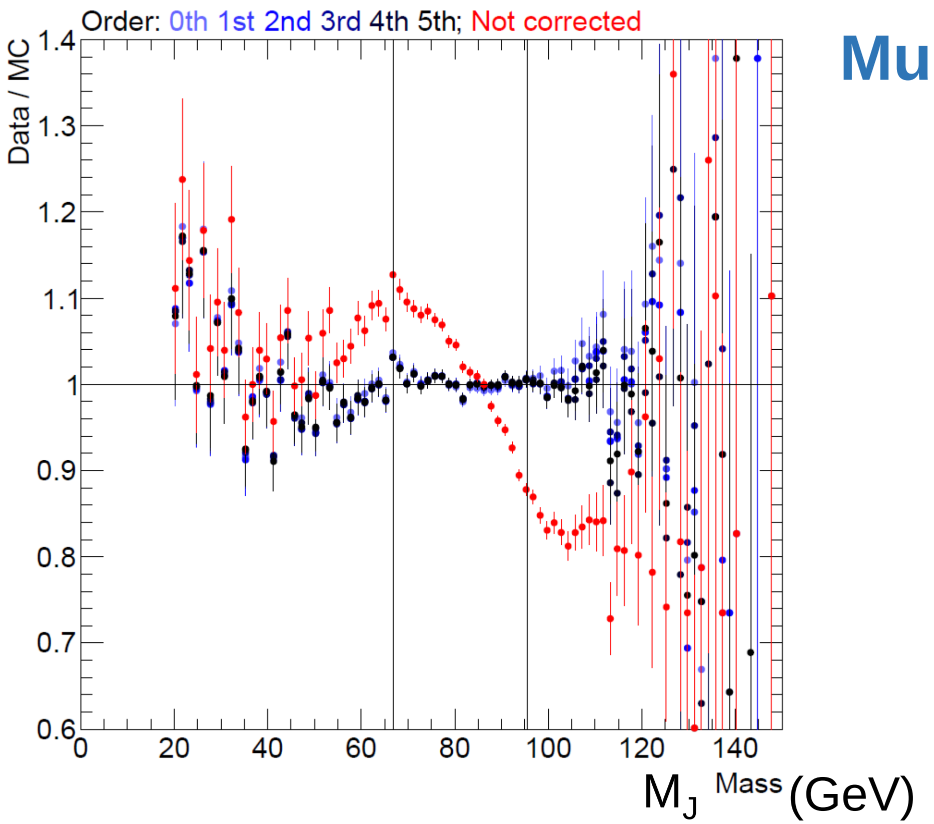


- The second step is to remove the difference in jet energy scale between the **+ side** and **– side**
- Differences in average jet energy in the **+ side <E+>** and **– side <E->** is corrected

Yi Chen
(MIT, CMS)

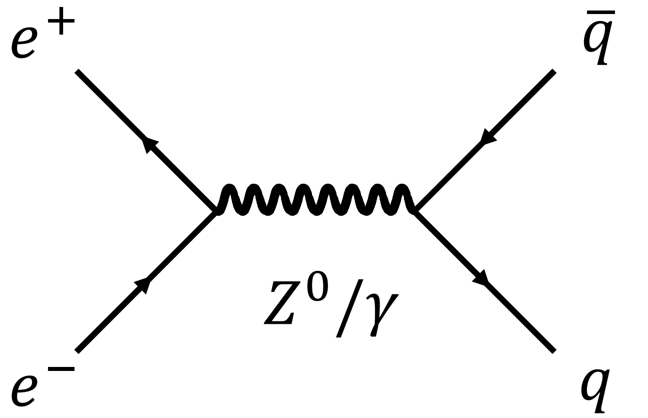
Data-based Calibration: Absolute Scale





Multi-jet invariant mass M_J method

Peak at $M_J = Z^0$ mass

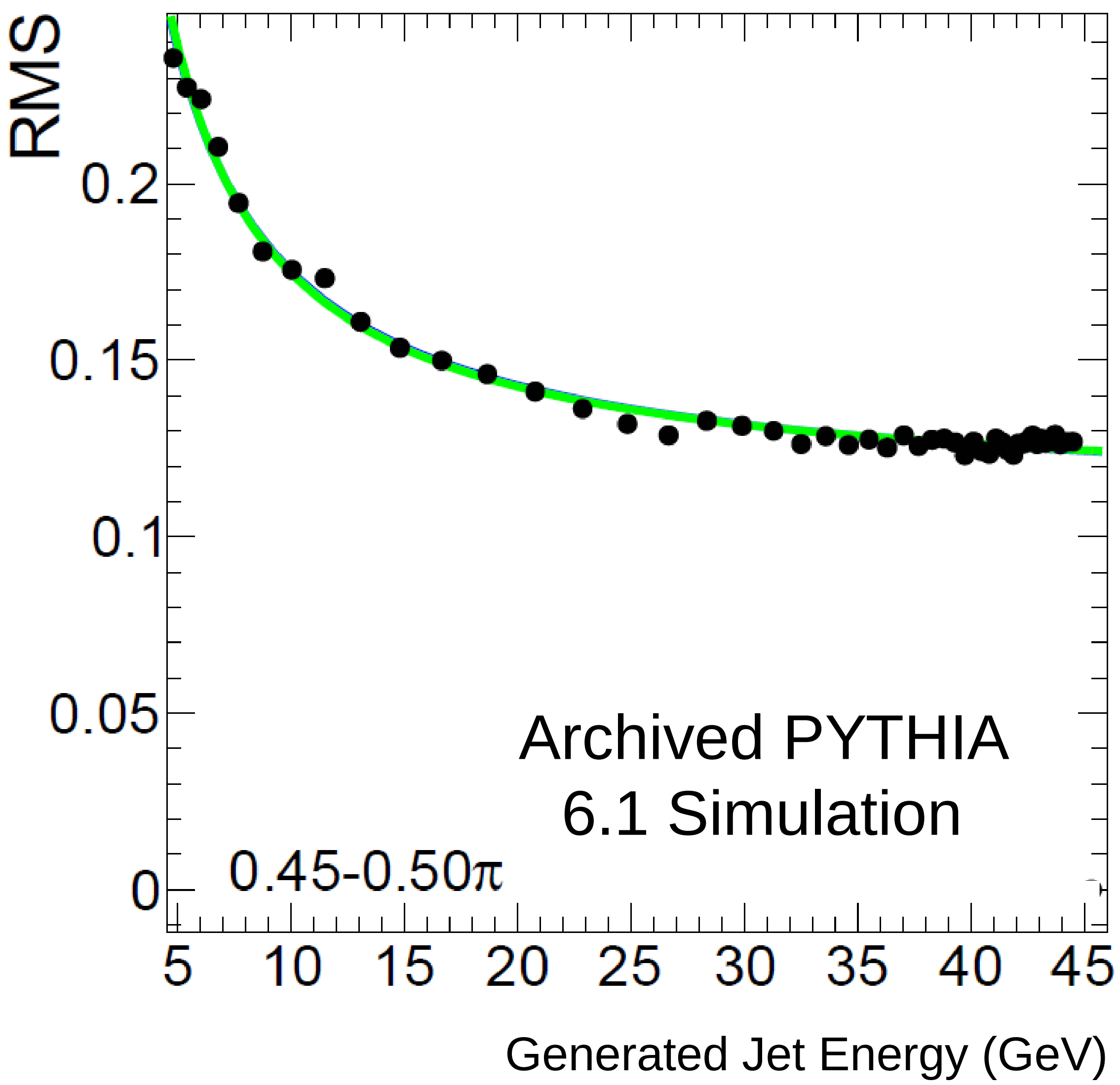


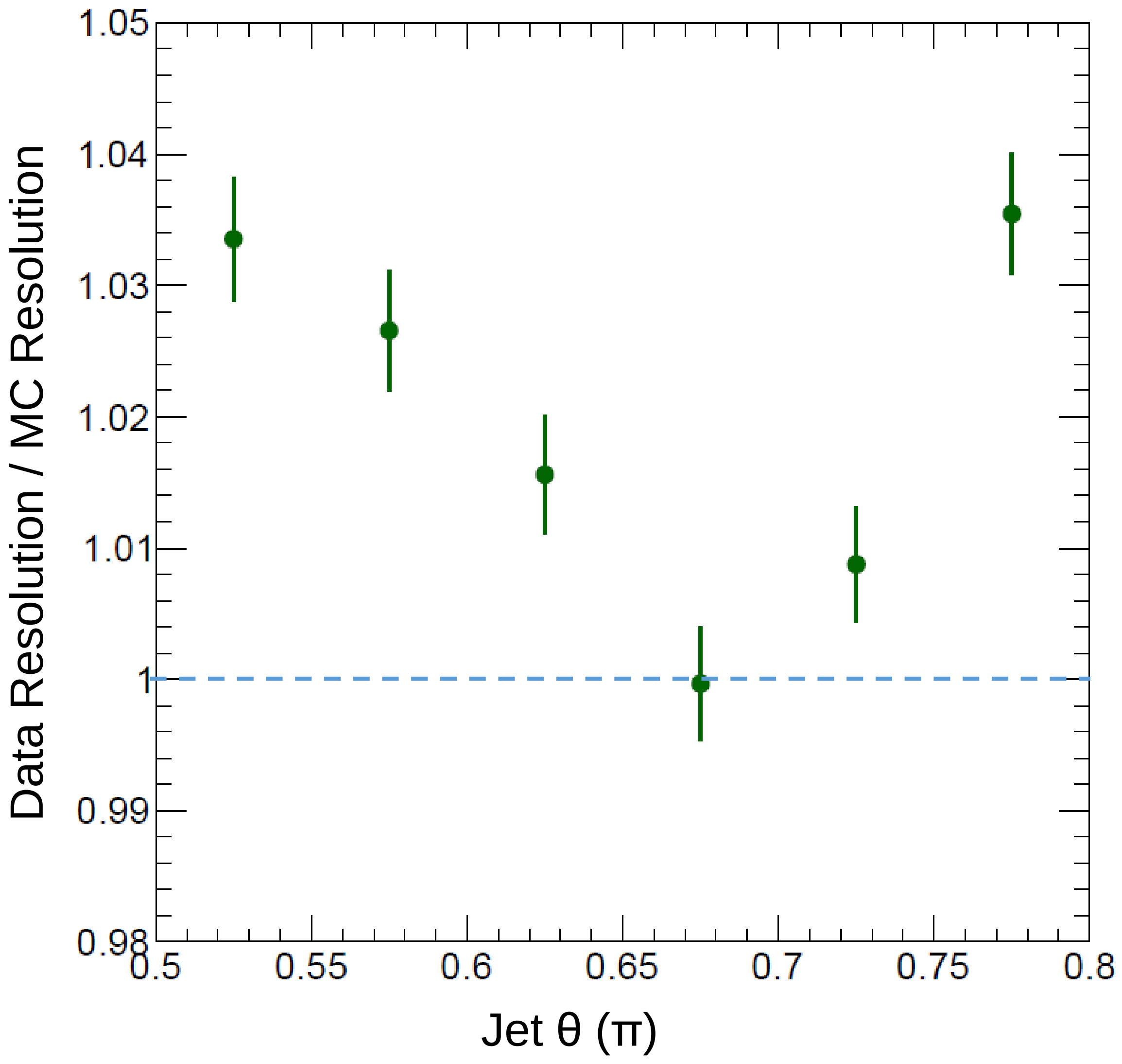
Yi Chen
(MIT, CMS)

- Performance with up to **N**=9 leading jets is presented (<u>ranked by energy</u>)
  - Require (**N**+1)th jet energy < **E$_{cut}$** to reduce the sensitivity to soft jets, where **E$_{cut}$** is set to 3 GeV
  - **N** and E_{cut} are varied for systematical uncertainty checks
- Jet energy correction is modeled by **nth** order polynomial of the jet energy **f(E$_{jet}$)**
- Minimize the χ^2 difference in mean mass between MC and data in 2% quantile ranges (slicing the mass spectrum)
  - 0-10% and 90-100% are removed to minimize impact of outliers
  - **n=1 chosen as the nominal correction**

MITHIG-MOD-NOTE-21-001
arXiv:2108.04877

Jet Energy Resolution





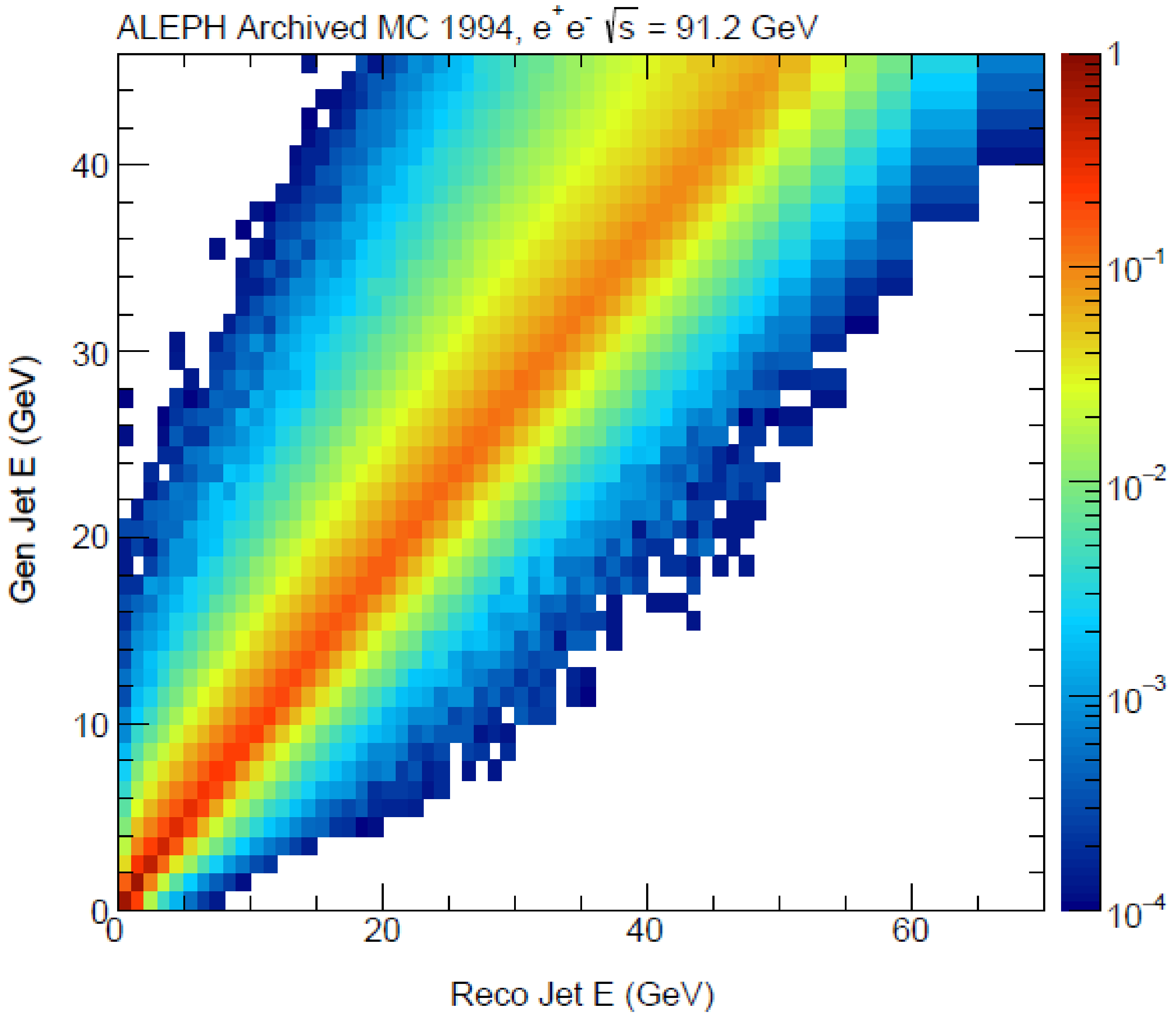
Yi Chen
(MIT, CMS)

- Jet energy resolution is around 15-25%
- Difference between data and MC are studied by leading dijet (0% ~ 4%)

MITHIG-MOD-NOTE-21-001
arXiv:2108.04877

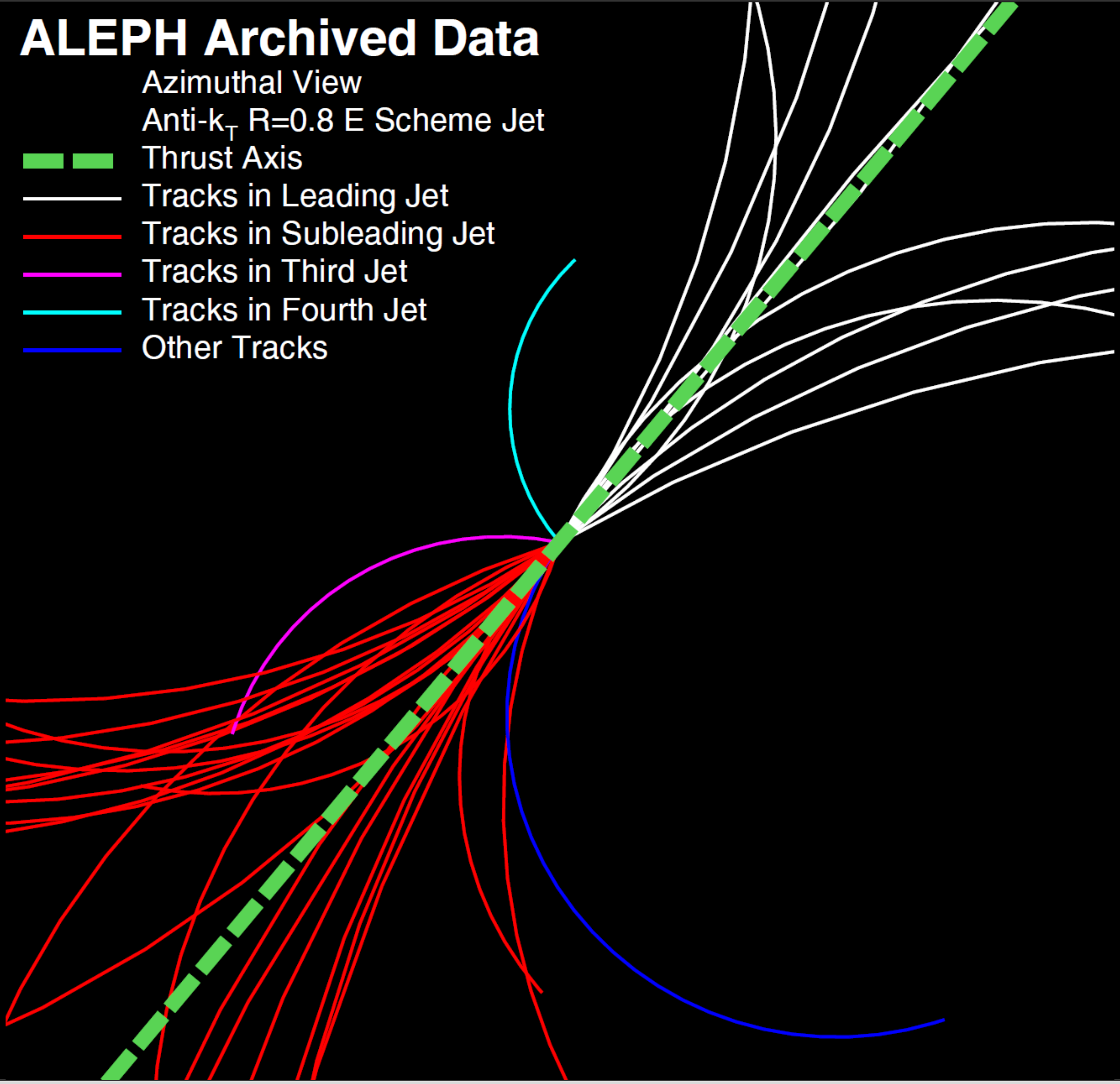
Unfolding

- Performed using the ROOUNFOLD package (v2.0.0). **Bayes Unfold method as the nominal result** and SVD as systematics check
- A flat prior is used in Bayes Unfold (PYTHIA 6.1 MC spectra used as prior for systematic check)
- 1D unfolding is performed for jet energy, leading dijet energy spectra.
- For jet mass, Z_G, R_G and groomed mass, 2D unfolding of the observables in bins of jet energy is performed



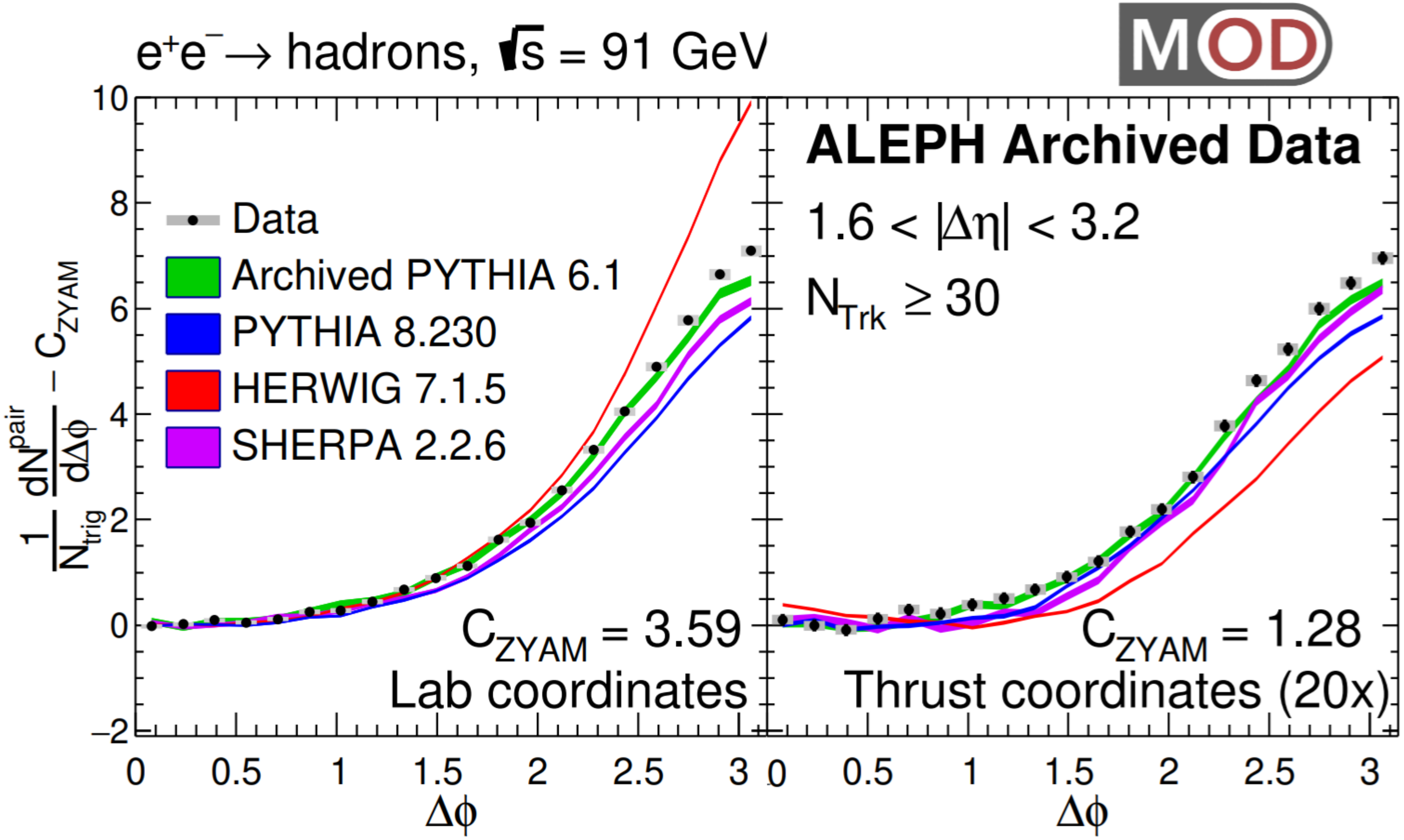
MITHIG-MOD-NOTE-21-001
arXiv:2108.04877

High Multiplicity Event in e^+e^- Collisions

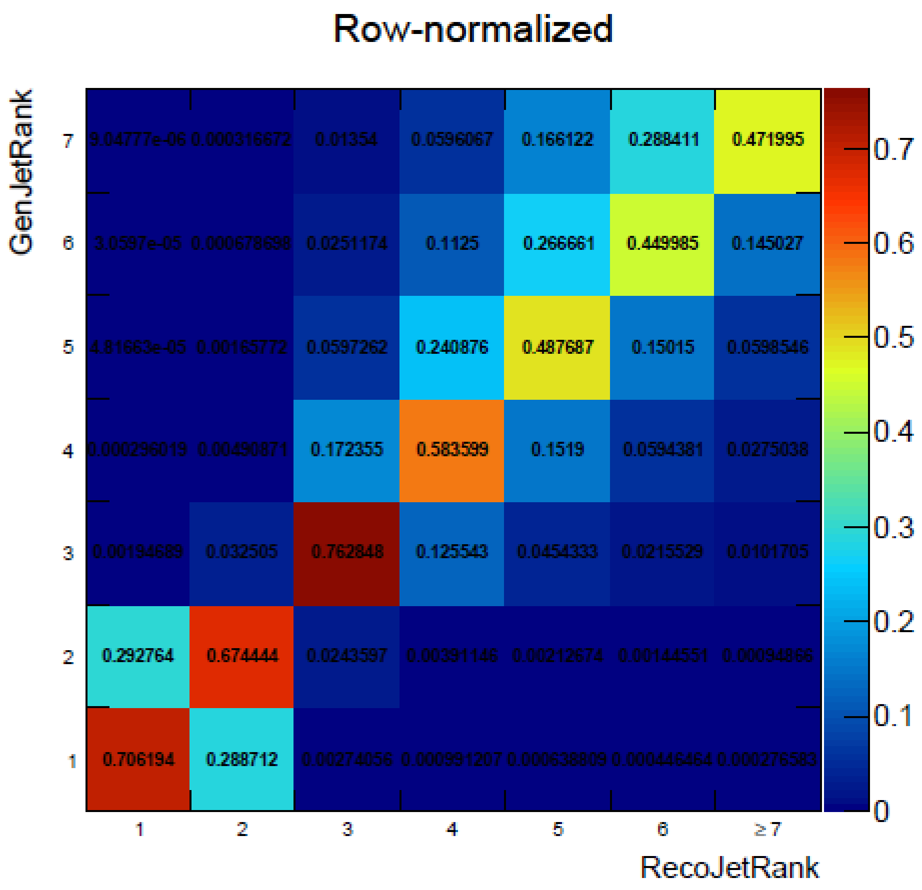


39 Tracks
T=0.98

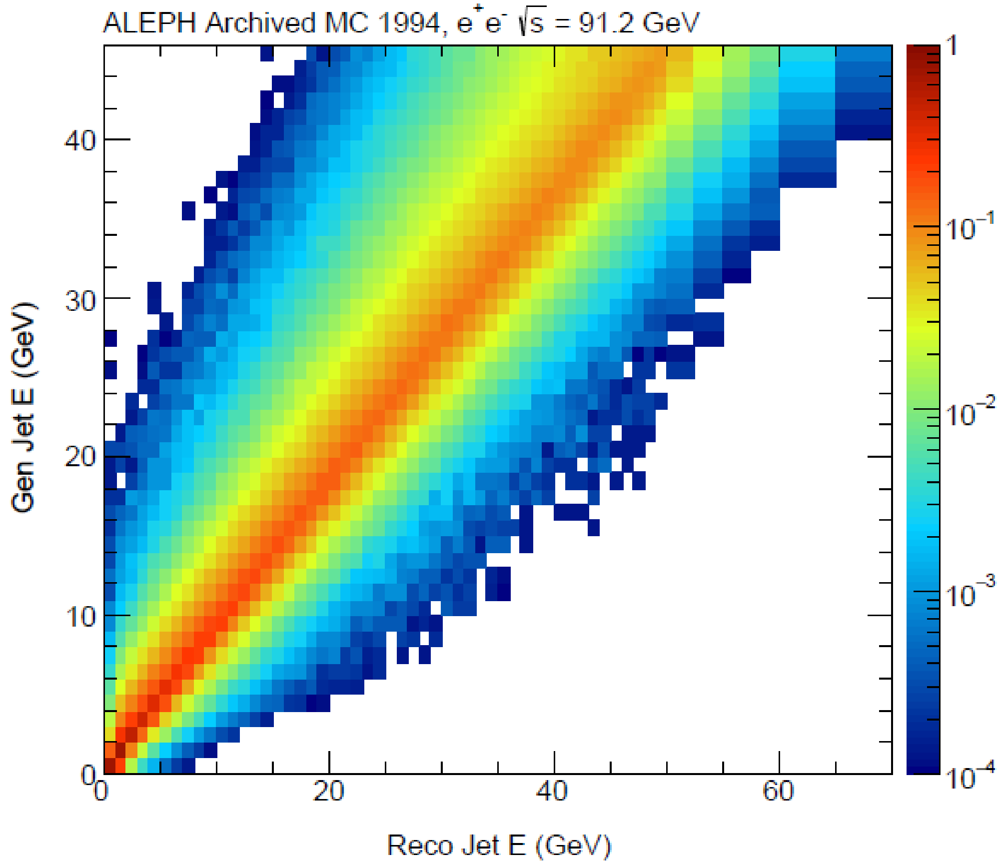
Correlation Function vs. Generators



Unfolding

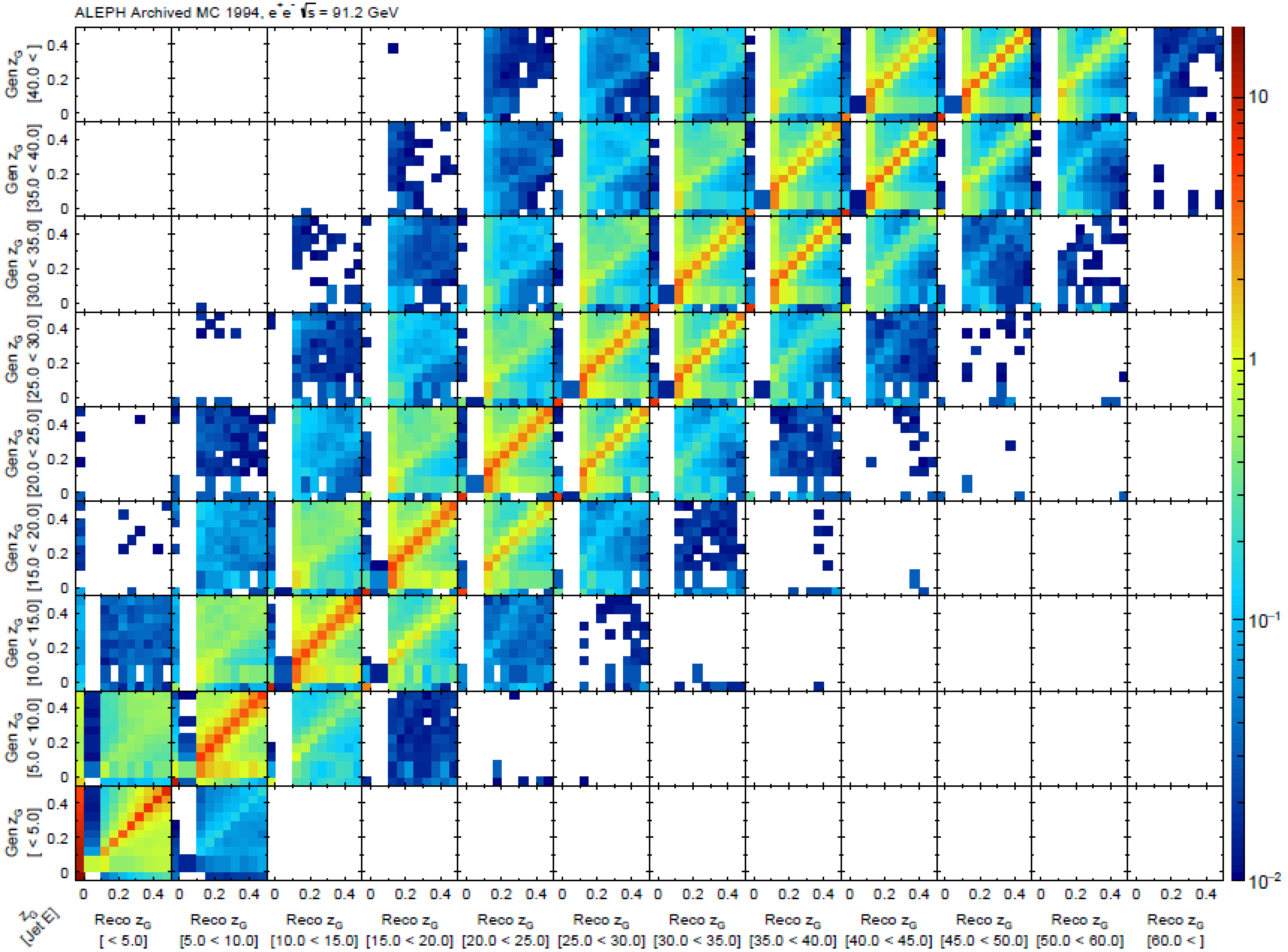


Row-normalized
GenJetRank
RecoJetRank
7 9.04777e-06 0.000316672 0.01354 0.0596067 0.166122 0.288411 0.471995
6 3.0597e-05 0.000678698 0.0251174 0.1125 0.266661 0.449985 0.145027
5 4.81663e-05 0.00165772 0.0597262 0.240876 0.487687 0.15015 0.0598546
4 0.000296019 0.00490871 0.172355 0.583599 0.1519 0.0594381 0.0275038
3 0.00194689 0.032505 0.762848 0.125543 0.0454333 0.0215529 0.0101705
2 0.292764 0.674444 0.0243597 0.00391146 0.00212674 0.00144551 0.00094866
1 0.706194 0.288712 0.00274056 0.000991207 0.000638809 0.000446464 0.000276583
1 2 3 4 5 6 ≥7
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7

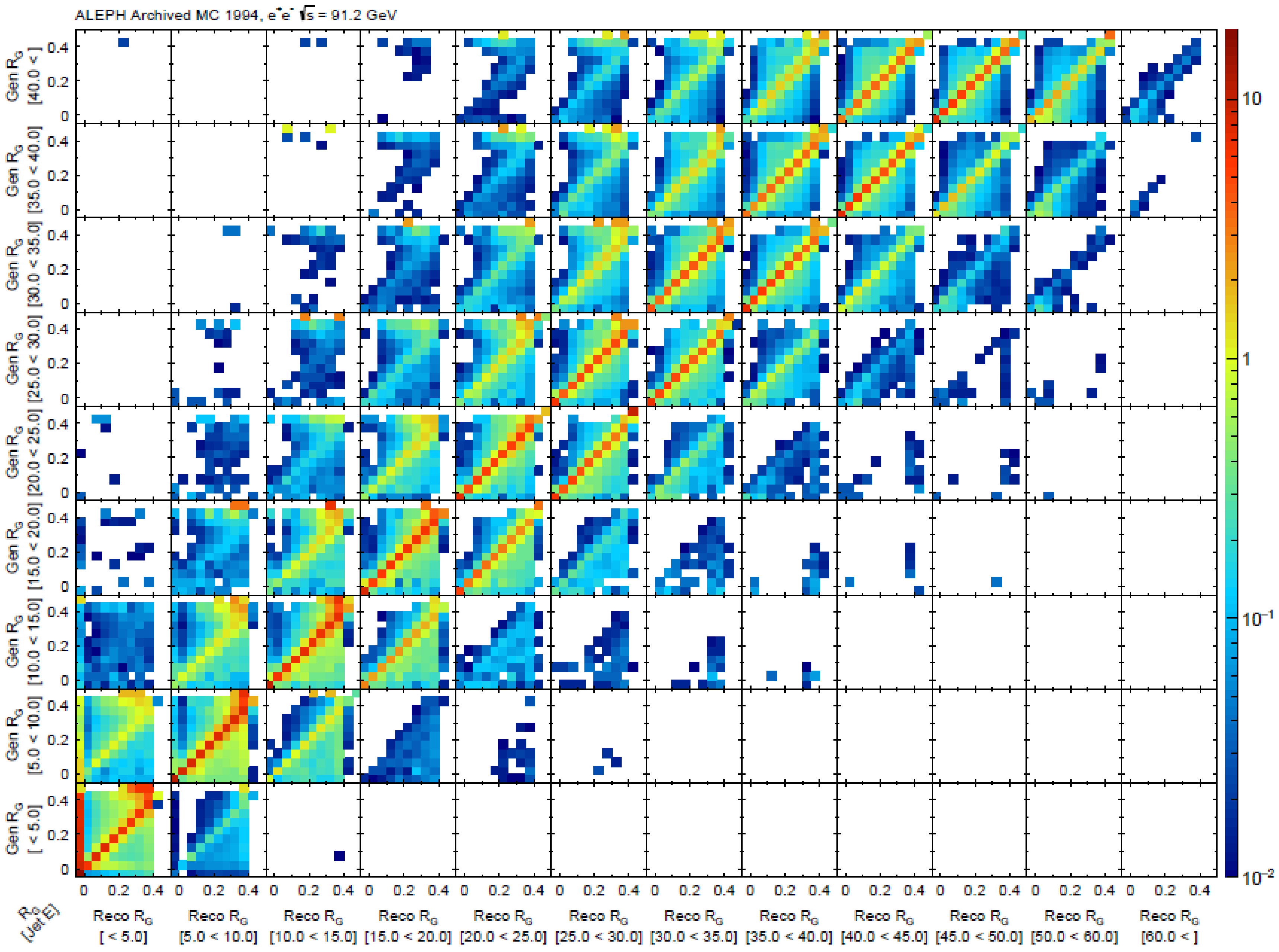


ALEPH Archived MC 1994, e^+e^- $\sqrt{s}$ = 91.2 GeV
Gen Jet E (GeV)
Reco Jet E (GeV)
0 10 20 30 40
0 20 40 60
10^{-4} 10^{-3} 10^{-2} 10^{-1} 1

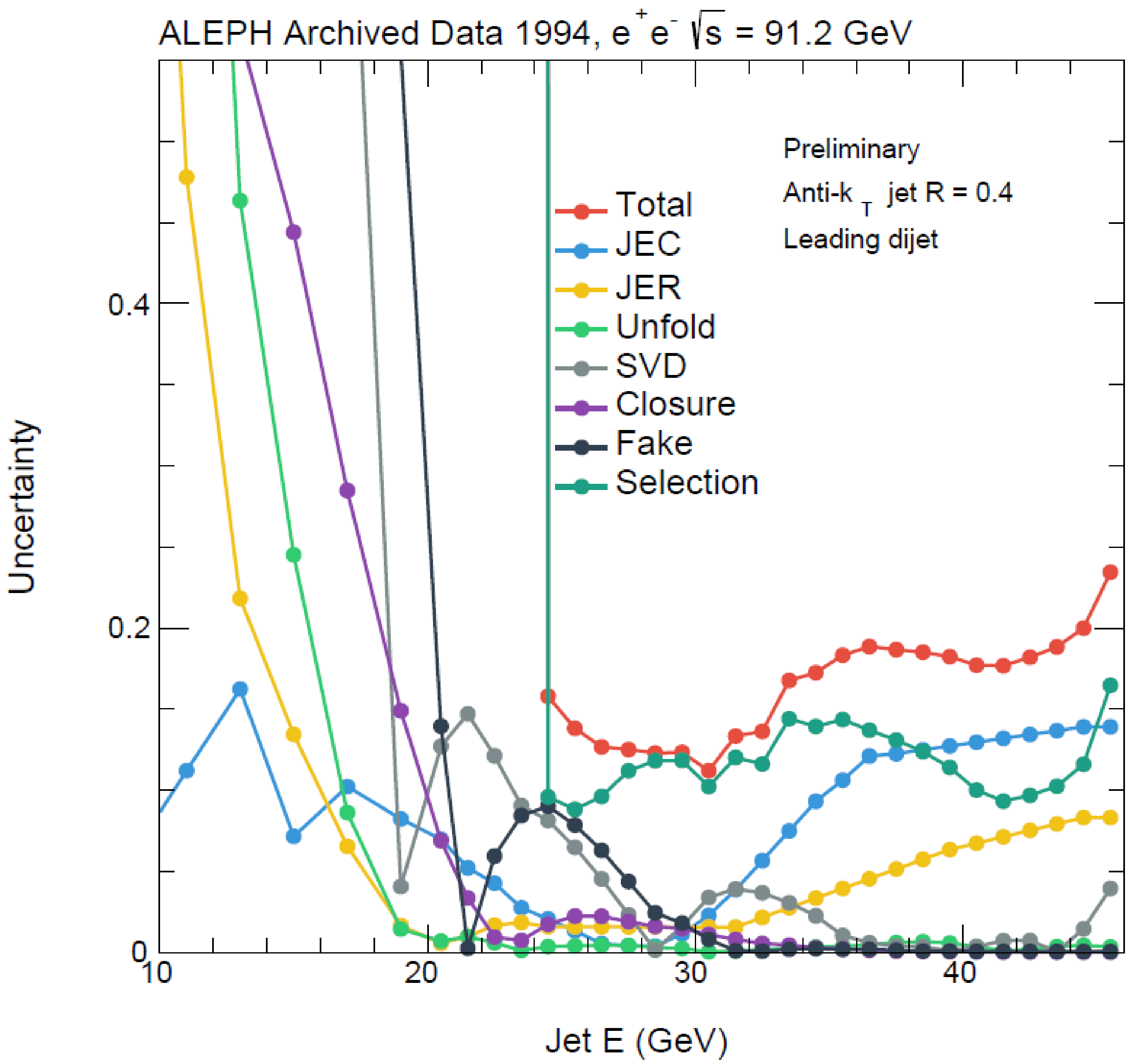
Response Matrix for Z_G



Response Matrix for R_G



Systematics: Leading Dijet Energy



Summary of M_J studies

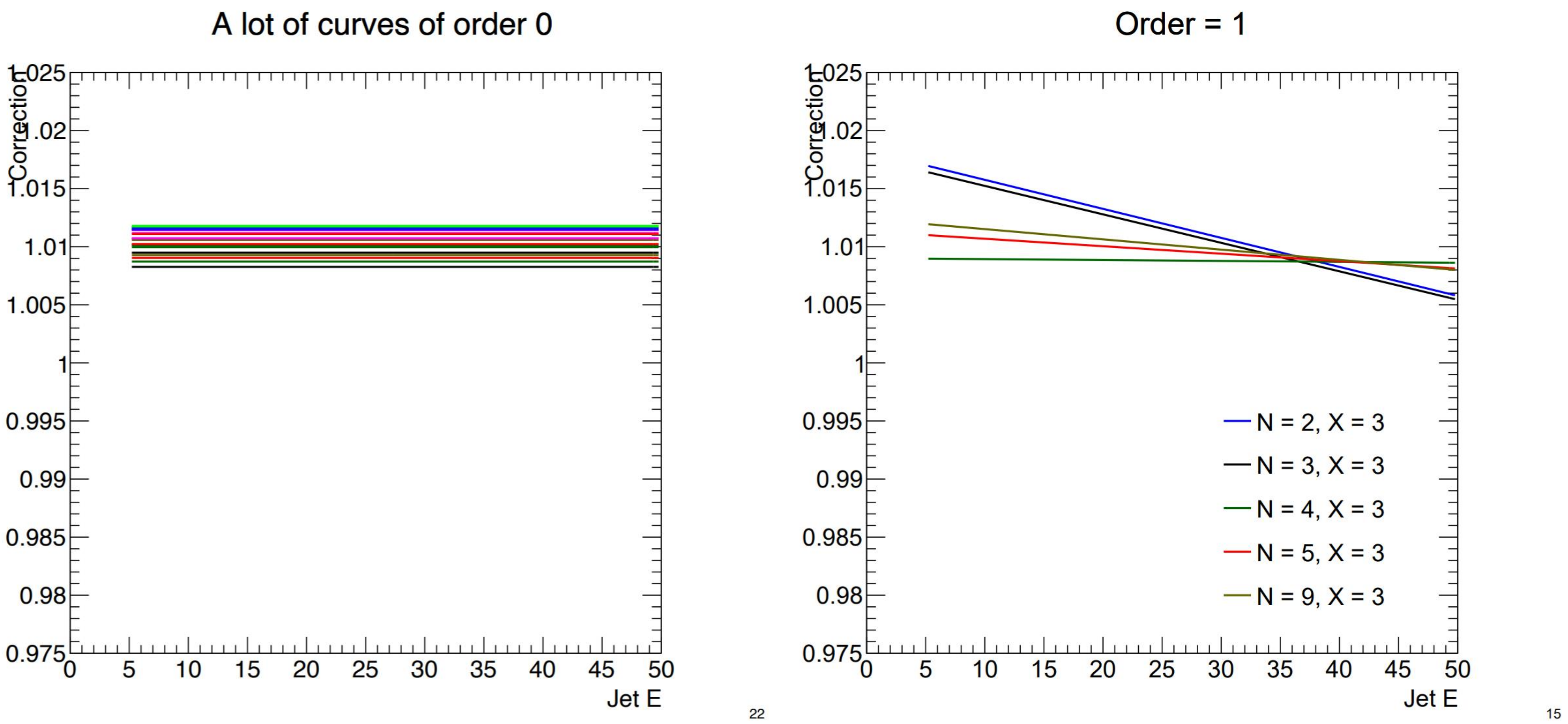


Figure 11: Example of fitted corrections with order 0 (left) and 1 (right). All variations of N and X agree for 0-th order fits, and there is a dependence on number of jets N at low energy.