Revisiting the ALEPH Archived e^+e^- Data: Challenges, Lessons and Recent Result on an Intriguing Structure in Long Range Correlations in High Multiplicity e^+e^- Collisions

Yi Chen (Vanderbilt)

CERN EP Seminar, Mar 5, 2024

In collaboration with Yu-Chen Chen (MIT), Yen-Jie Lee (MIT), Marcello Maggi (INFN Bari), Anthony Badea (UChicago), Austin Baty (UIC), Paoti Chang (NTU), Chris McGinn (MIT), Jesse Thaler (MIT), Gian Michelle Innocenti (MIT), Michael Peters (MIT), Tzu-An Sheng (MIT)

Vanderbilt HENP's group is supported by US DOE-NP

Revisiting the ALEPH Archived e^+e^- Data: Challenges, Lessons and Recent Result on an Intriguing Structure in Long Range Correlations in High Multiplicity e^+e^- Collisions

Yi Chen (Vanderbilt)

CERN EP Seminar, Mar 5, 2024

In collaboration with Yu-Chen Chen (MIT), Yen-Jie Lee (MIT), Marcello Maggi (INFN Bari), Anthony Badea (UChicago), Austin Baty (UIC), Paoti Chang (NTU), Chris McGinn (MIT), Jesse Thaler (MIT), Gian Michelle Innocenti (MIT), Michael Peters (MIT), Tzu-An Sheng (MIT)

Vanderbilt HENP's group is supported by US DOE-NP

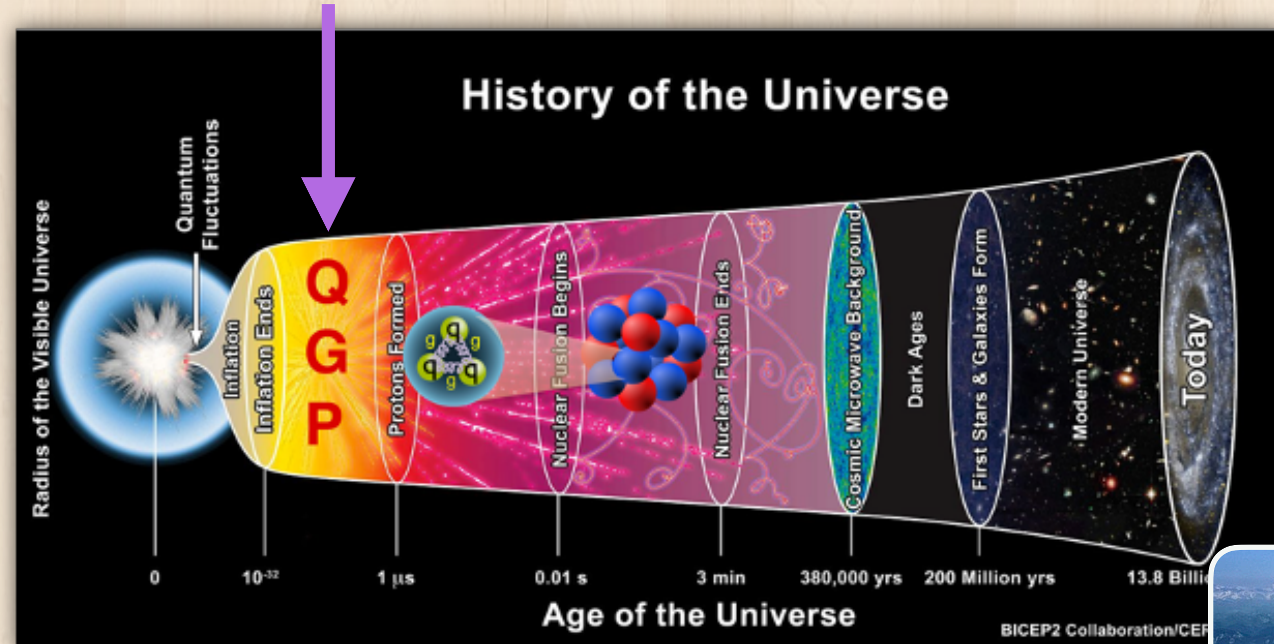
Outline

- Motivation
- Reanalysis effort on archived data
  - How it started and how it is going
  - Some lessons learned for open data
- Two-particle correlation analysis and results

Motivation:
Particle Correlations

Quark Gluon Plasma and Heavy-ion

The Quark-gluon plasma



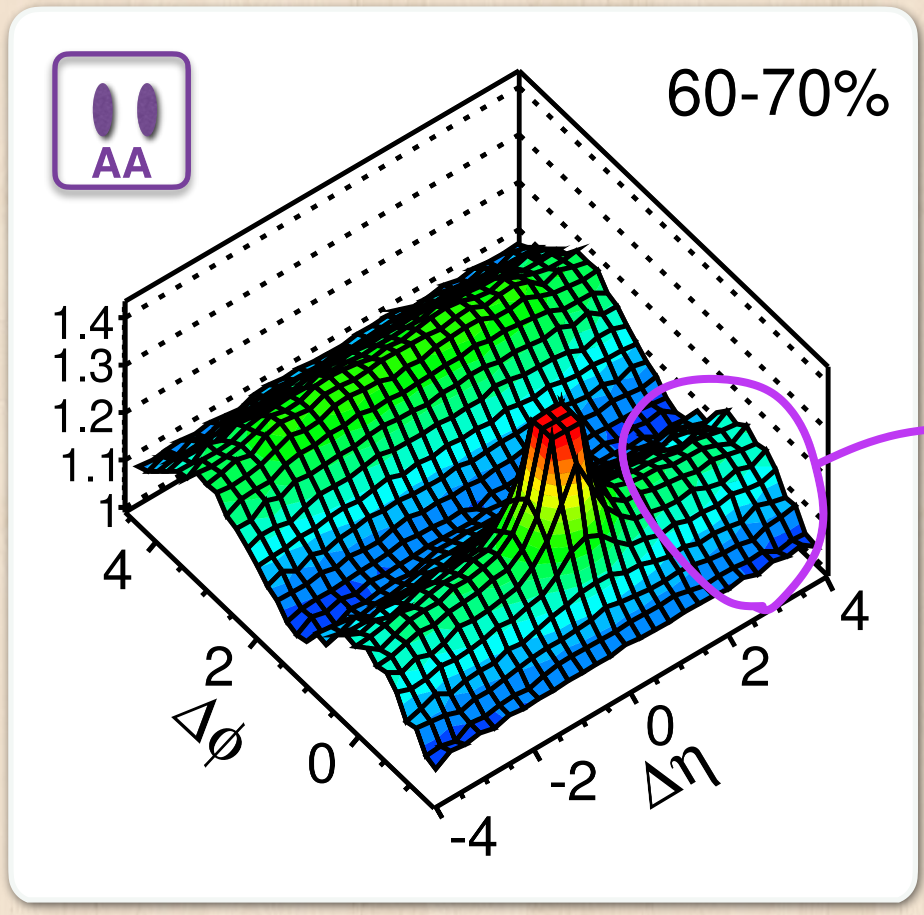
Hot! Quarks and gluons not confined in hadrons

Tiny droplets created in high energy Heavy-ion collisions (e.g. RHIC/LHC)

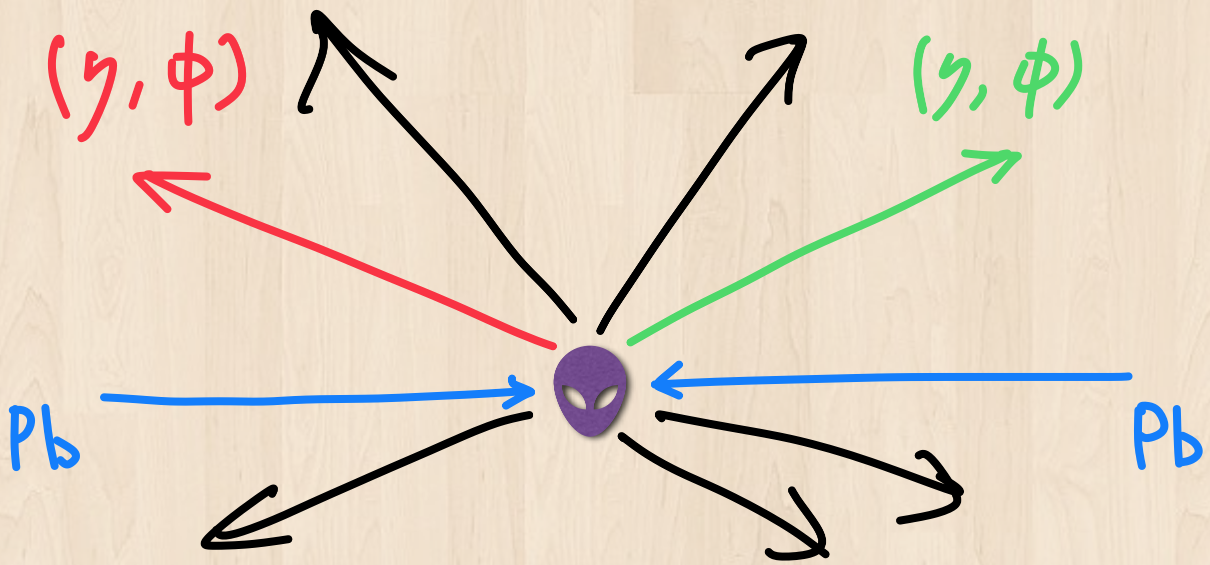
Stress test QCD under extreme conditions

Collectivity in Heavy-ion collision

2-particle correlation



PbPb collisions at LHC



Excess $\Delta\phi \sim 0$, large $\Delta\eta$

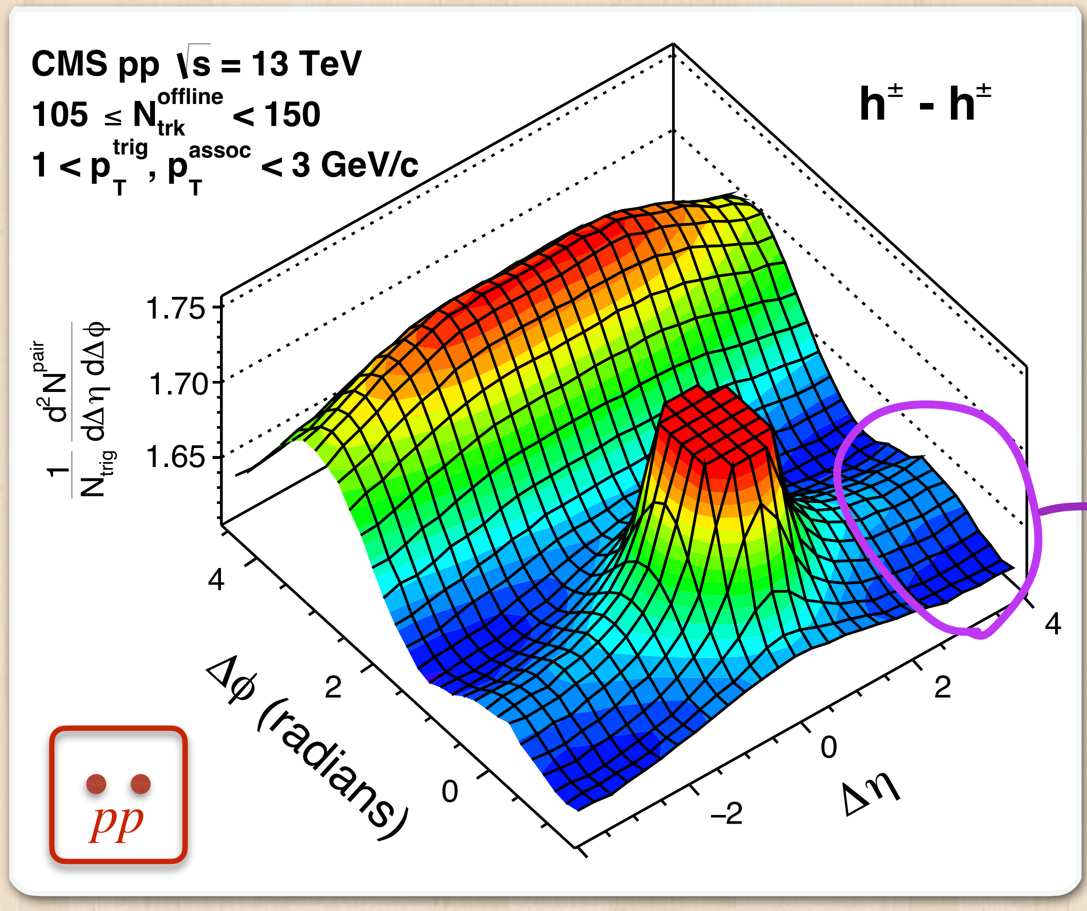
Or $\phi \sim \phi$

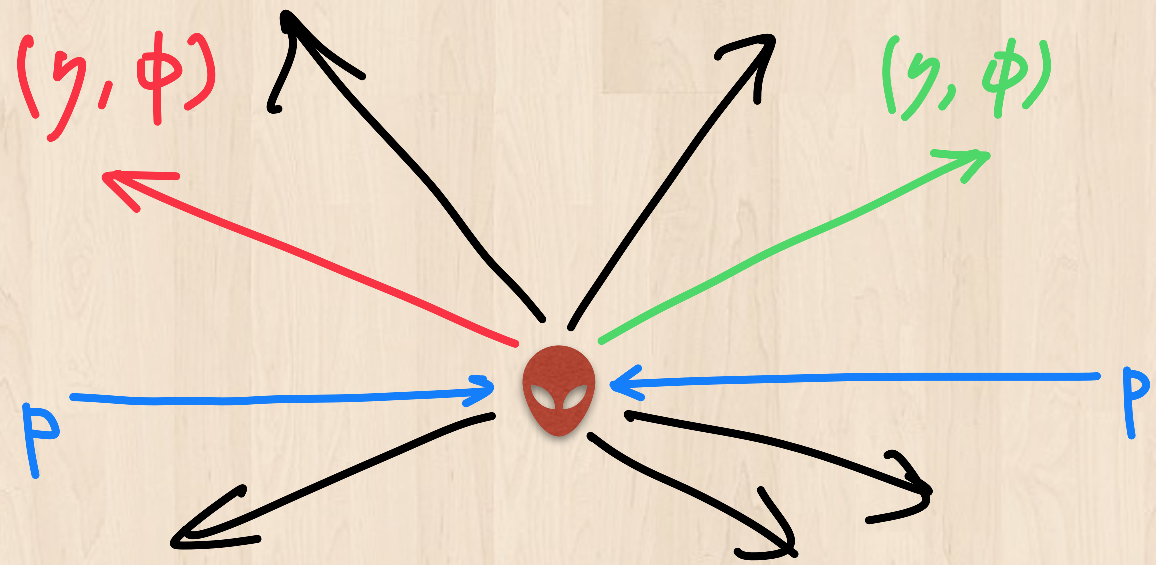
Many potential causes:
Shape of the plasma
Initial state fluctuation
…

Phys.Lett.B 724 (2013) 213

Some effects in *pp* as well

2-particle correlation





Interesting 🤔

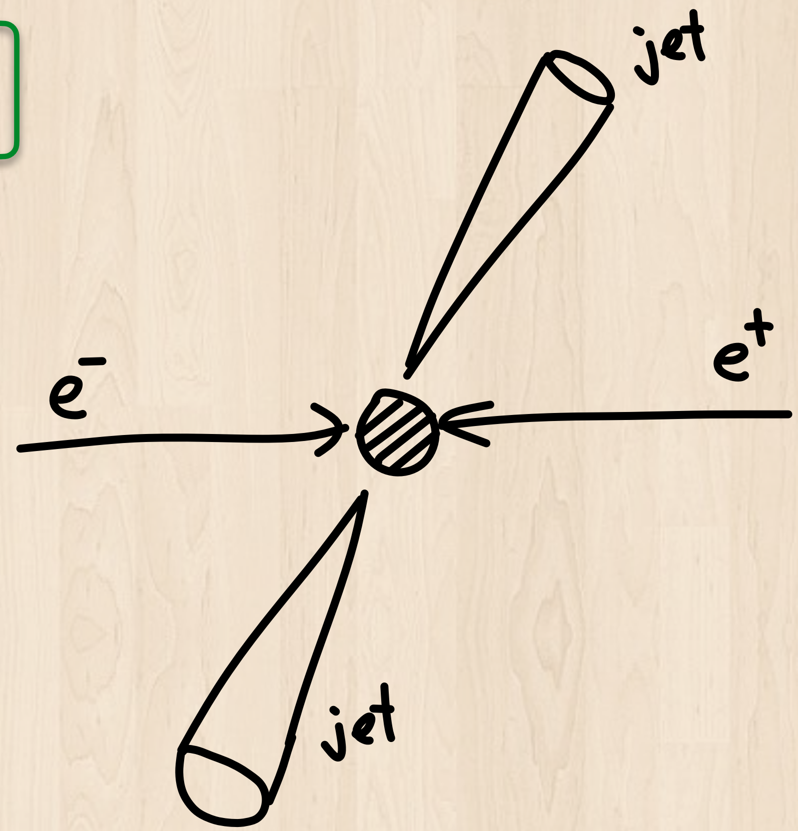
QGP droplets? Ropes? …

pp collisions at LHC

Time to go to simpler systems

Going even simpler

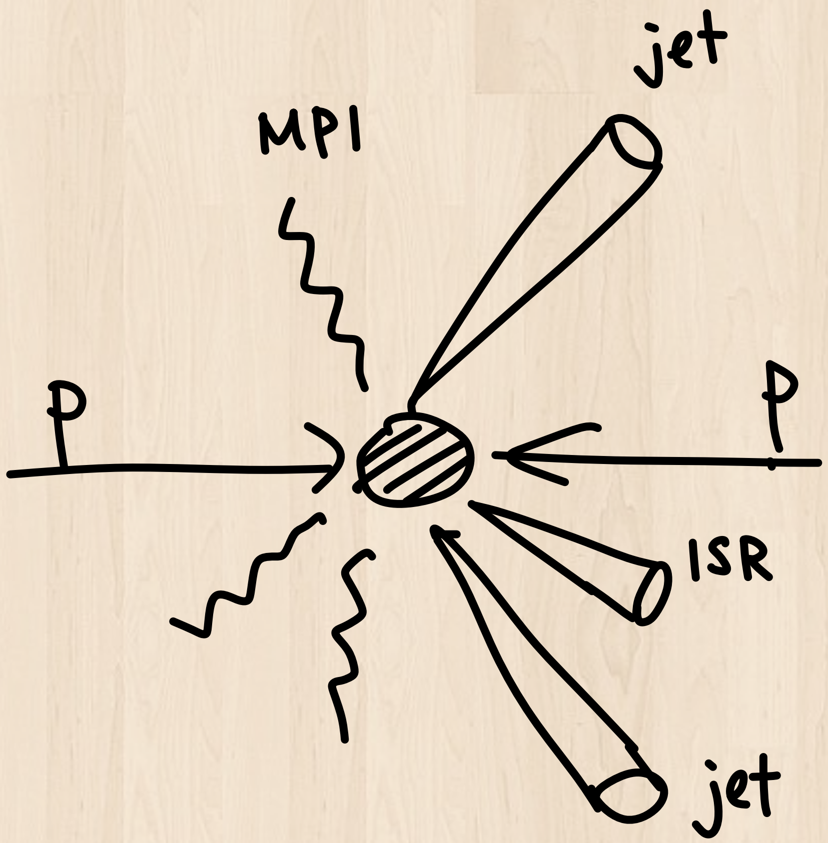
e^+e^-



Better control of event kinematics

Cleanest test of pQCD and models

pp

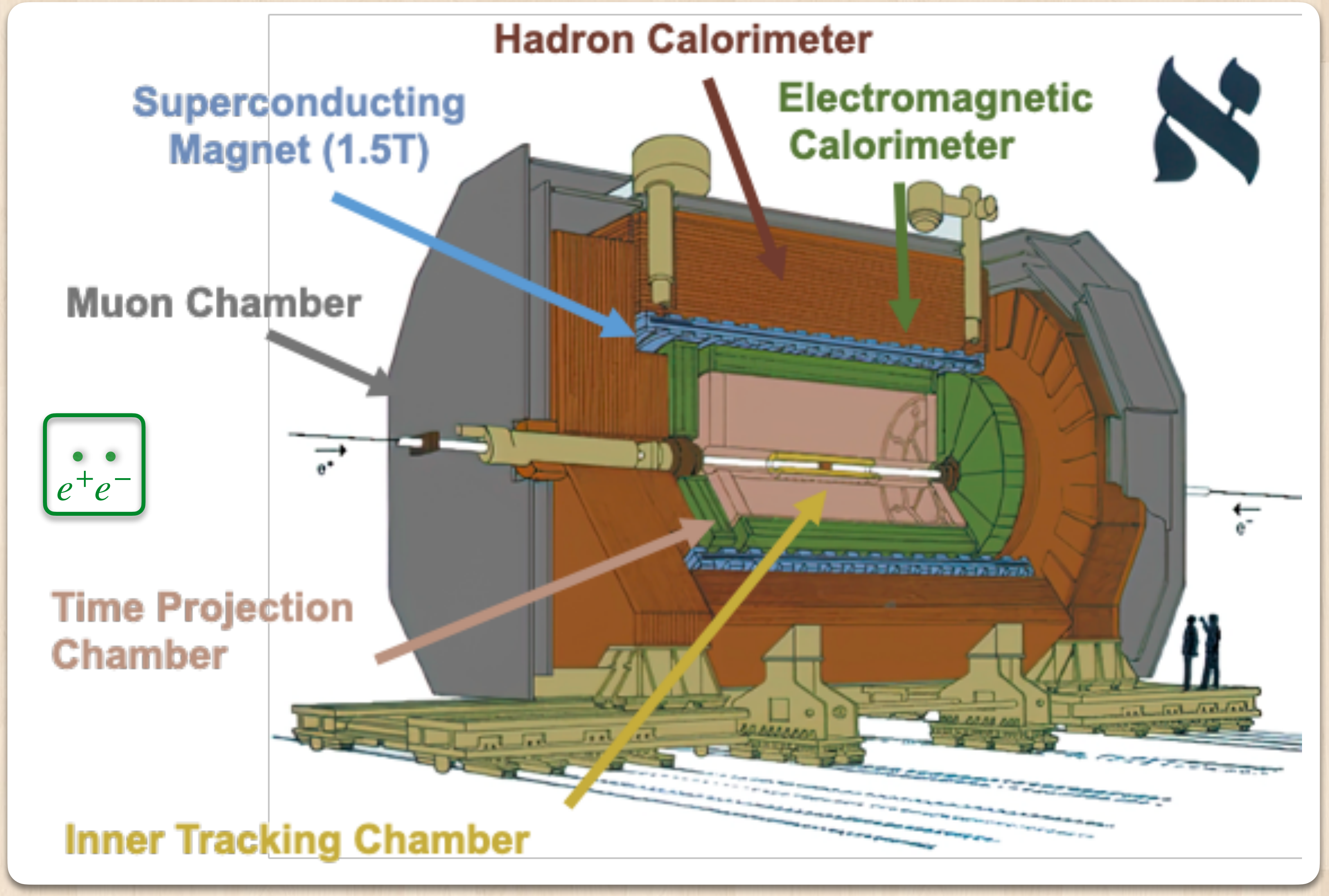


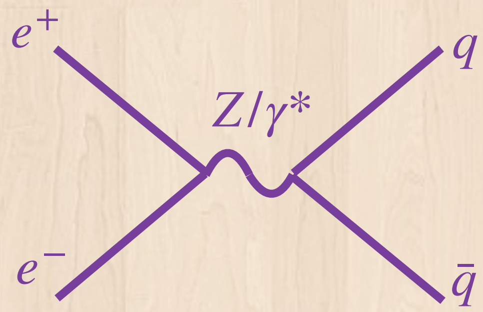
PDF convolution
No longitudinal control
More ISR
MPI

Complements well measurements from other systems

The archived data

The archived ALEPH data





Comparable with RHIC energy!

LEP1 e^+e^- data taken at 91.2 GeV from 1992-1995

LEP2 taken with higher energy up to 209 GeV

How the reanalysis effort started

2017

February: **Yen-Jie Lee** connected to **Gigi Rolandi** and later to spokesperson **Roberto Tenchini** about the use of archived data

Marcello Maggi help extract the energy flow information and archived simulation/data

Mid-2017: all samples converted to the MIT open-data format

Bibek Pandit & **Anthony Badea** (**Yen-Jie**'s undergraduate student) started working on event selection validation

Guenther Dissertori provided analysis code from the QCD paper

2018

March: Successfully reproduced unfolded thrust distribution

…

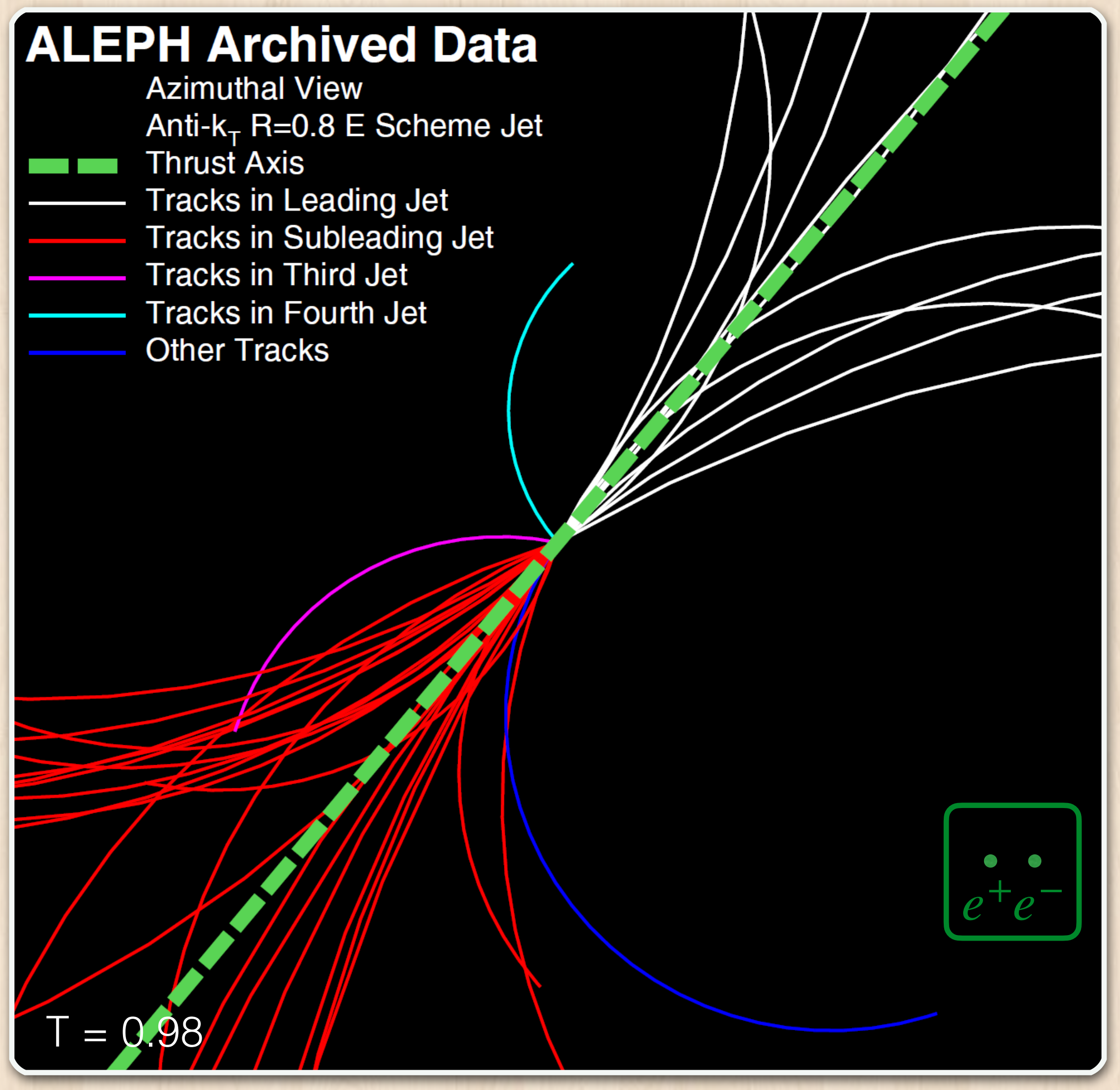
Keys to the success

- Foresight from **ALEPH collaboration** for the data archival
- Incredible support from ALEPH members **Marcello Maggi, Roberto Tenchini, Gigi Rolandi, Guenther Dissertori** on the technical aspects and knowledge
- Many **bright young students** who dug into the data collected before they were born
- **Reproduction of published physics results** using identical event selections
- Development of **data-driven checks** to understand the data

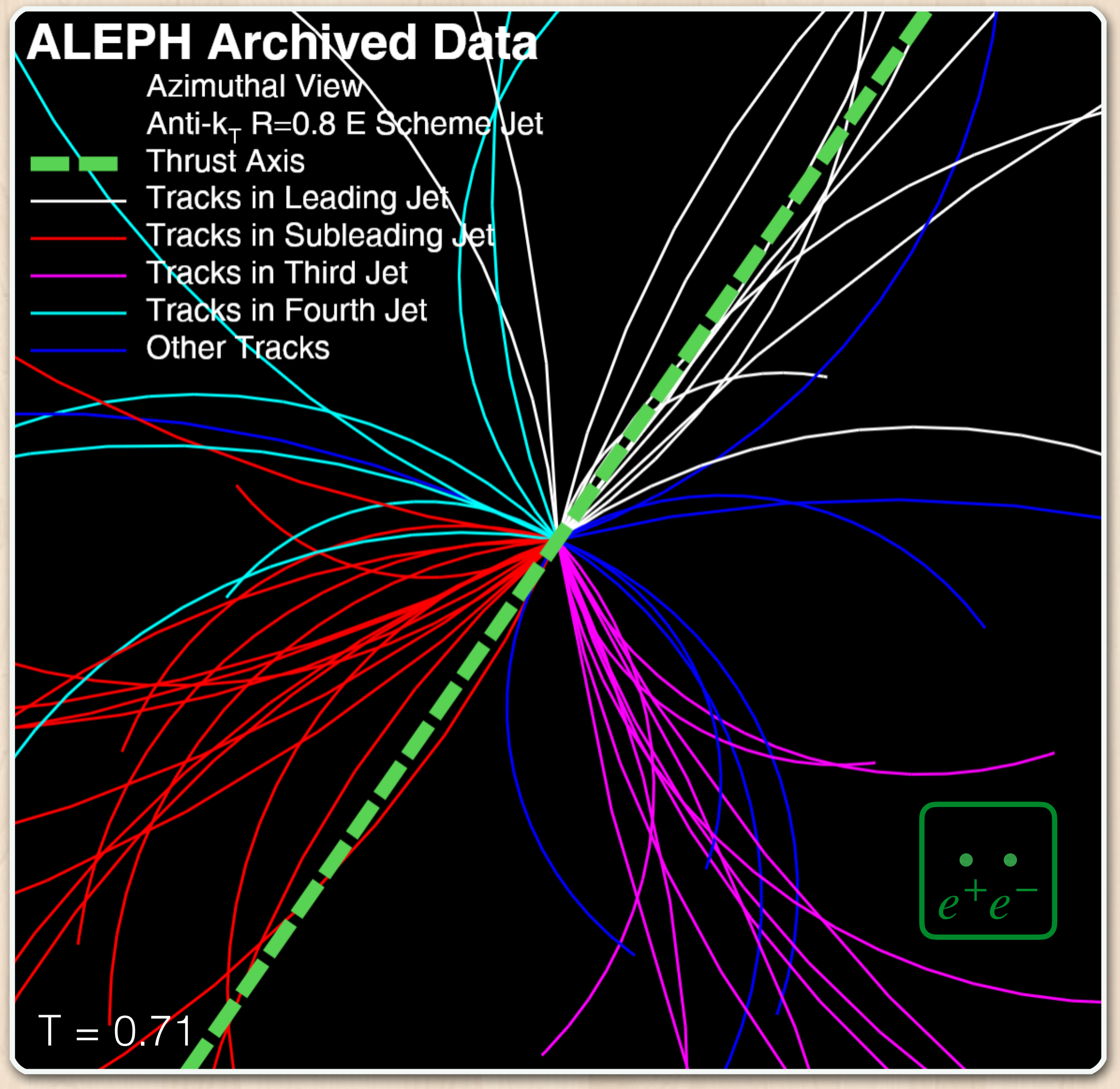
What information is available

- **Energy flow objects** — similar idea to the particle flow approach in other experimental collaborations
  - Combining information from tracker, calorimeter and muon chambers
  - Starting point of all re-analysis effort
- Some other **associated information** also available, for example PID scores, number of hits in TPC

Example high multiplicity events



$N_{\text{track}} = 39$



$N_{\text{track}} = 55$

Available simulation

- **Archived MC**: both generator level and detector level available
  - Great for deriving MC calibrations on objects
  - The only available set of MC that is fully simulated
- **More recent generators**: typically we generate things ourselves, only generator level possible
- This is the limiting factor for some observables

Data-driven checks

- Data-driven methods to study and understand the **data/MC difference**
- As a demonstration, I will go over a recent example on understanding performance of **jets**
- Dedicated calibration probing data/MC difference is developed — both jet energy scale and resolution

Jet clustering

For 1994 archived data & simulation

Energy-flow objects are used as input

In order to compare with LHC/RHIC

anti-"k_T" jet, **R = 0.4**

Hadron-hadron collider

$$d_{ij} = \min\left(p_{T,i}^{-2}, p_{T,j}^{-2}\right)\frac{\Delta R_{ij}^2}{R^2}$$

$$d_{iB} = p_{T,i}^{-2}$$

$\rightarrow$

e^+e^- distance measure

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

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

θ_{ij} = opening angle (rad.)

JHEP 0804 (2008) 063

EPJC 72 (2012) 1896

Jet calibration

		"MC calibration"		"Residual"
Jet energy scale in data	=	Jet energy scale in simulation	×	Jet energy scale in data / Jet energy scale in simulation
		Inclusive		Selection

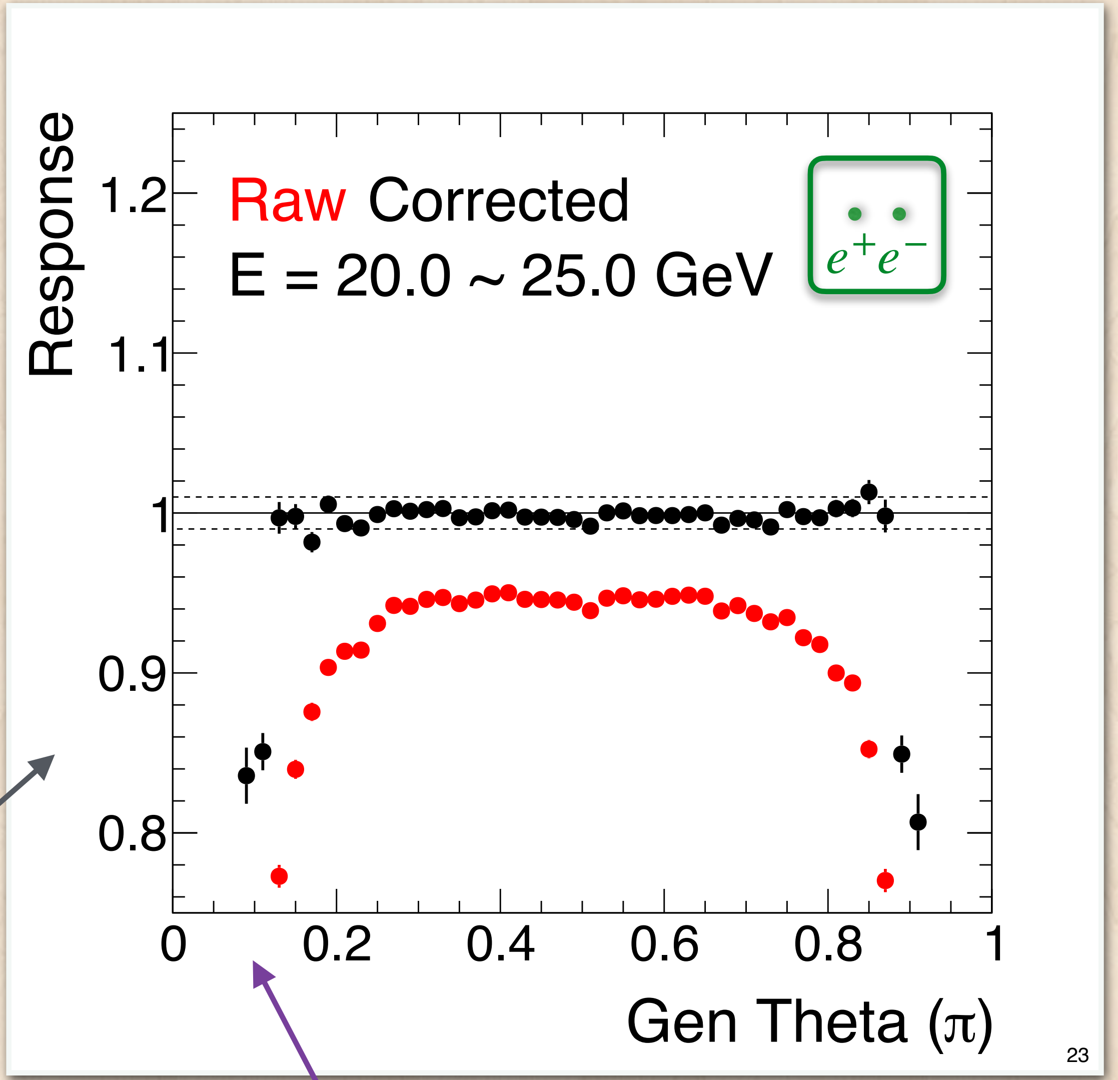
Strategy: first go 99% of the way there with simulation
Then data and MC difference in restricted phase spaces

Simulated energy scale

Correct detector jet energy in bins of jet direction (θ_{jet})

Good closure with
E > 10 GeV
$0.2\pi < \theta_{\text{jet}} < 0.8\pi$

Example raw and corrected response (= detector-level/generated)

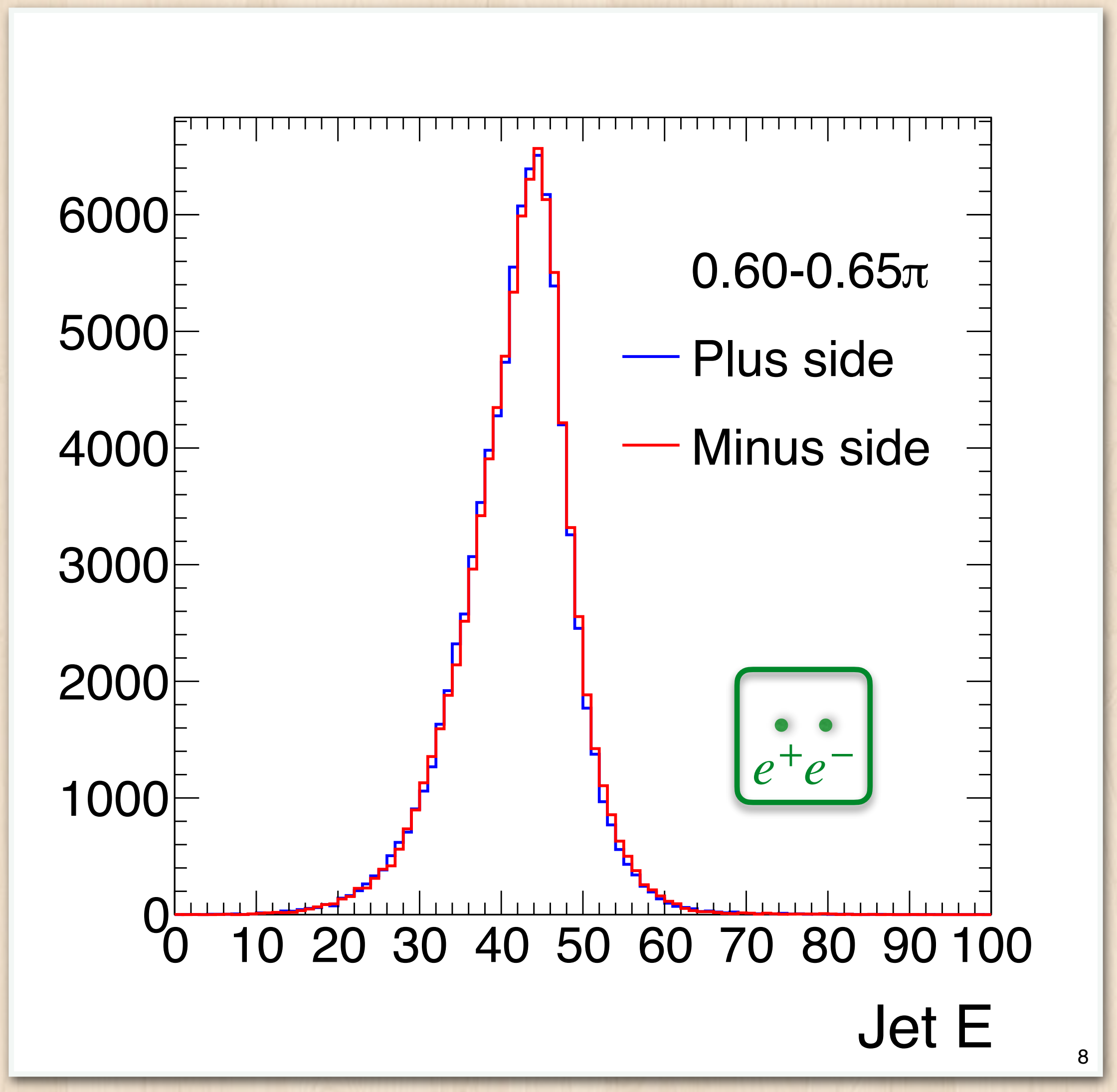


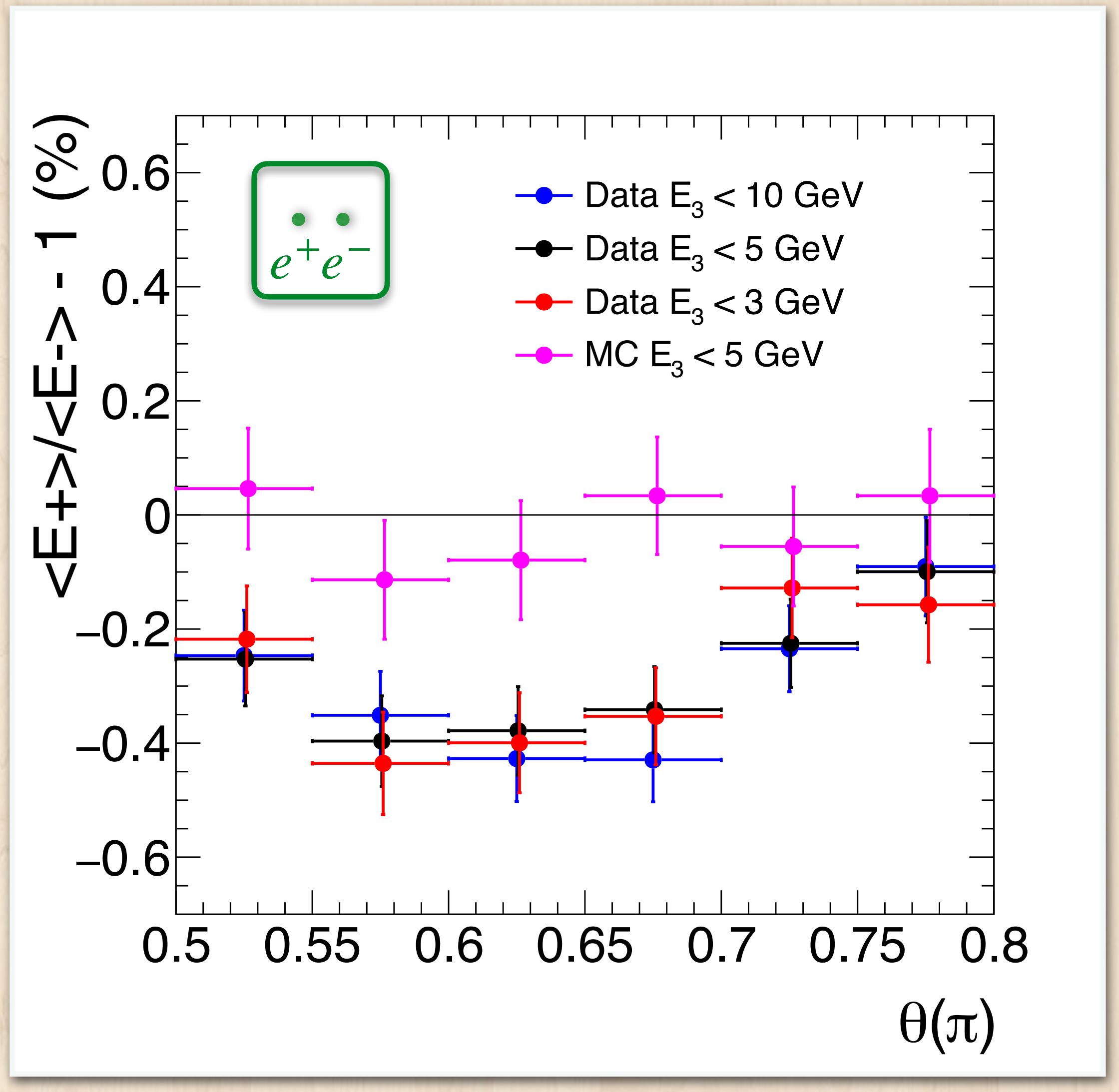
Energy leaking out around beam direction

arXiv 2108.04877

Residual calibration: step 1

Fiducial dijet, two sides of the detector



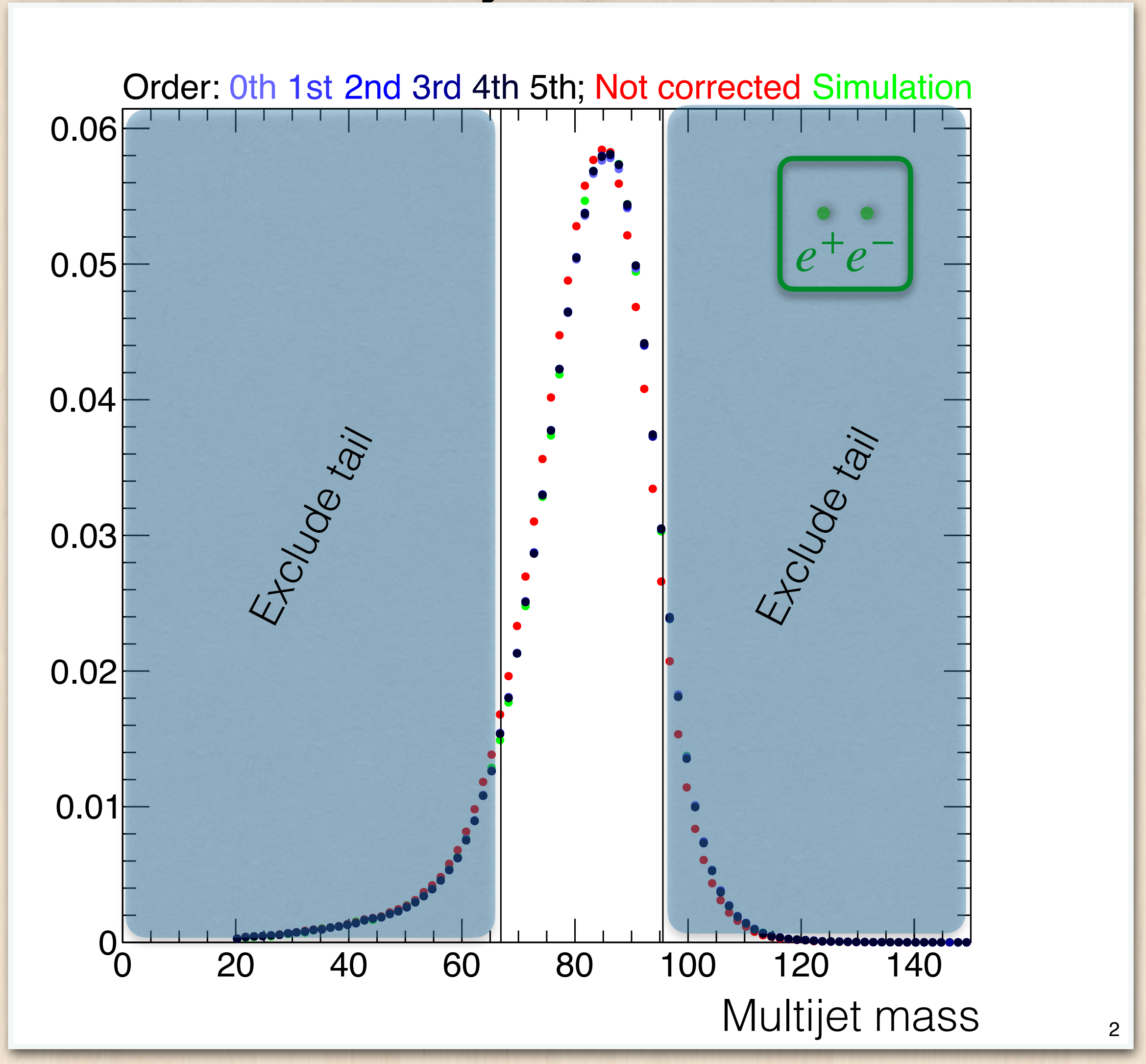


Look at data only, and calibrate out the difference between e^-- and e^+-going sides

arXiv 2108.04877

Residual calibration: step 2

Fiducial multijet invariant mass

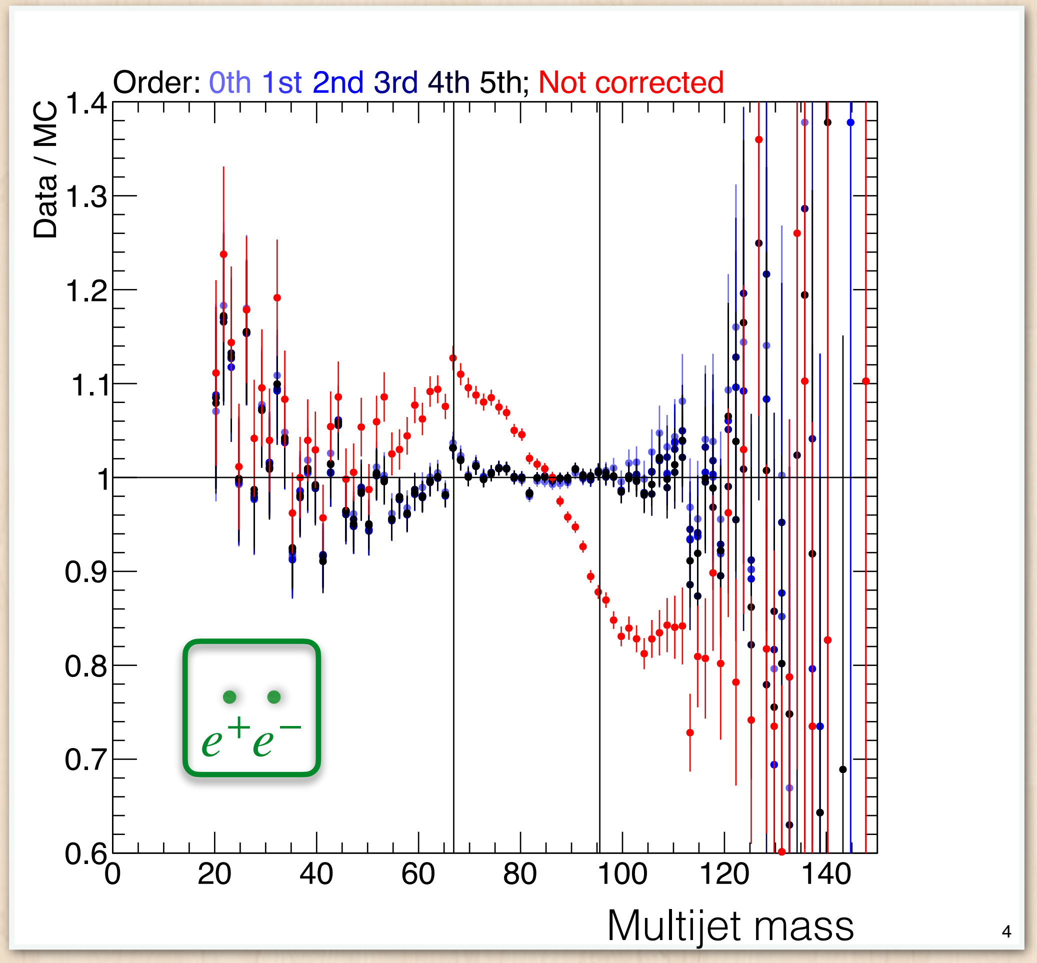


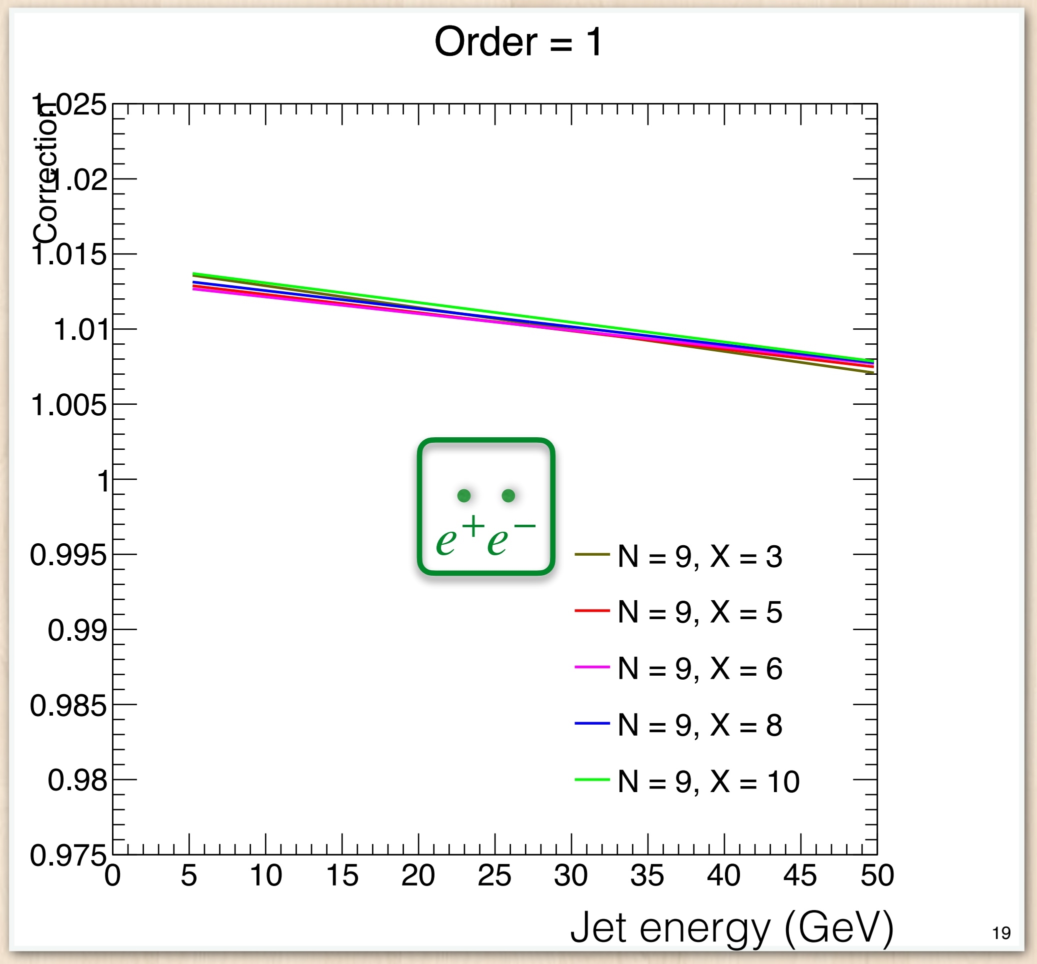
Fit jet energy correction function parameters

arXiv 2108.04877

Residual calibration: step 2

Fiducial multijet invariant mass





Take up to leading N jet above X GeV

Fit jet energy correction function parameters

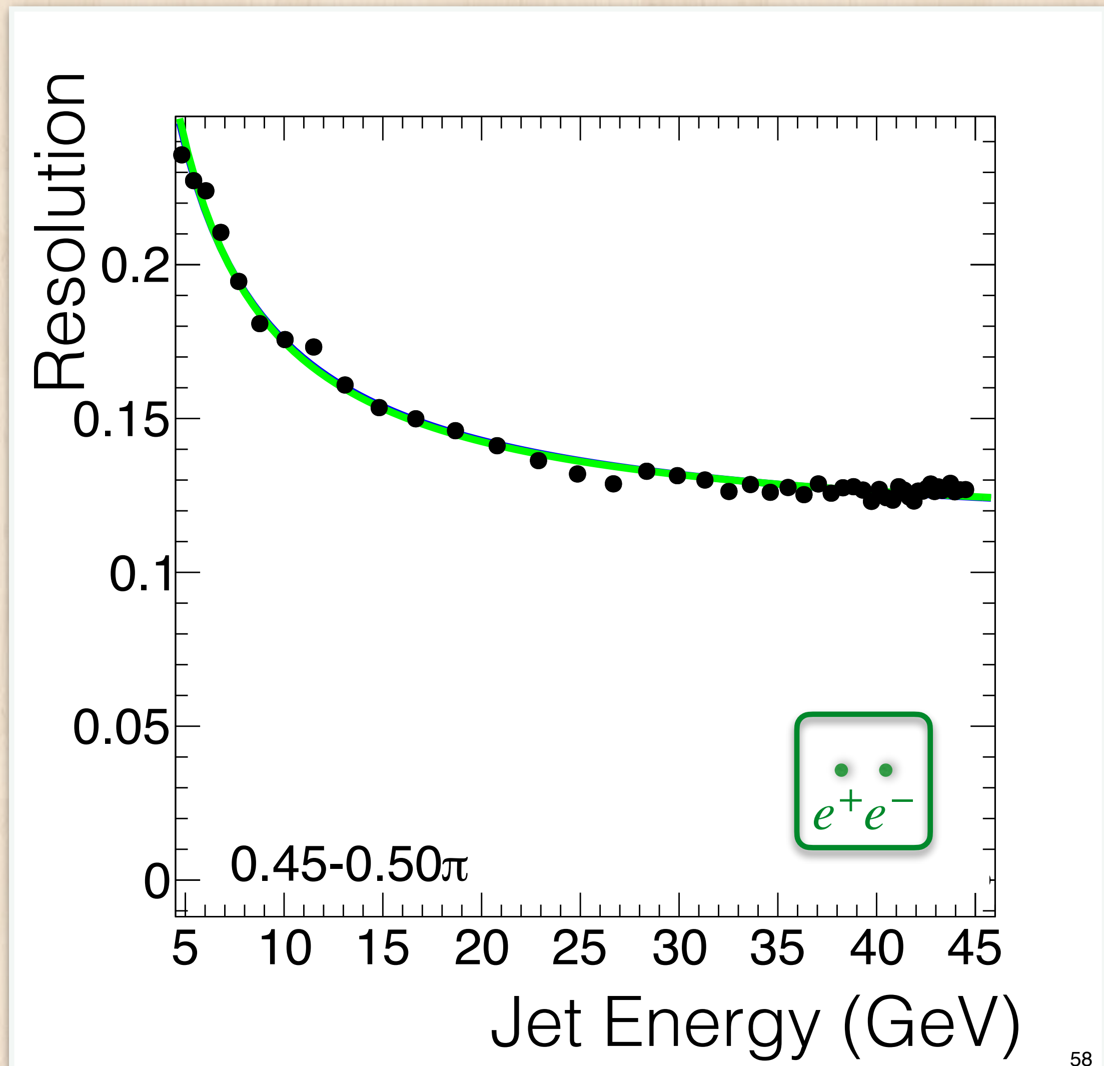
Minimize “quantile difference” (~KS) between data and MC curves

Nominal: linear correction as a function of energy

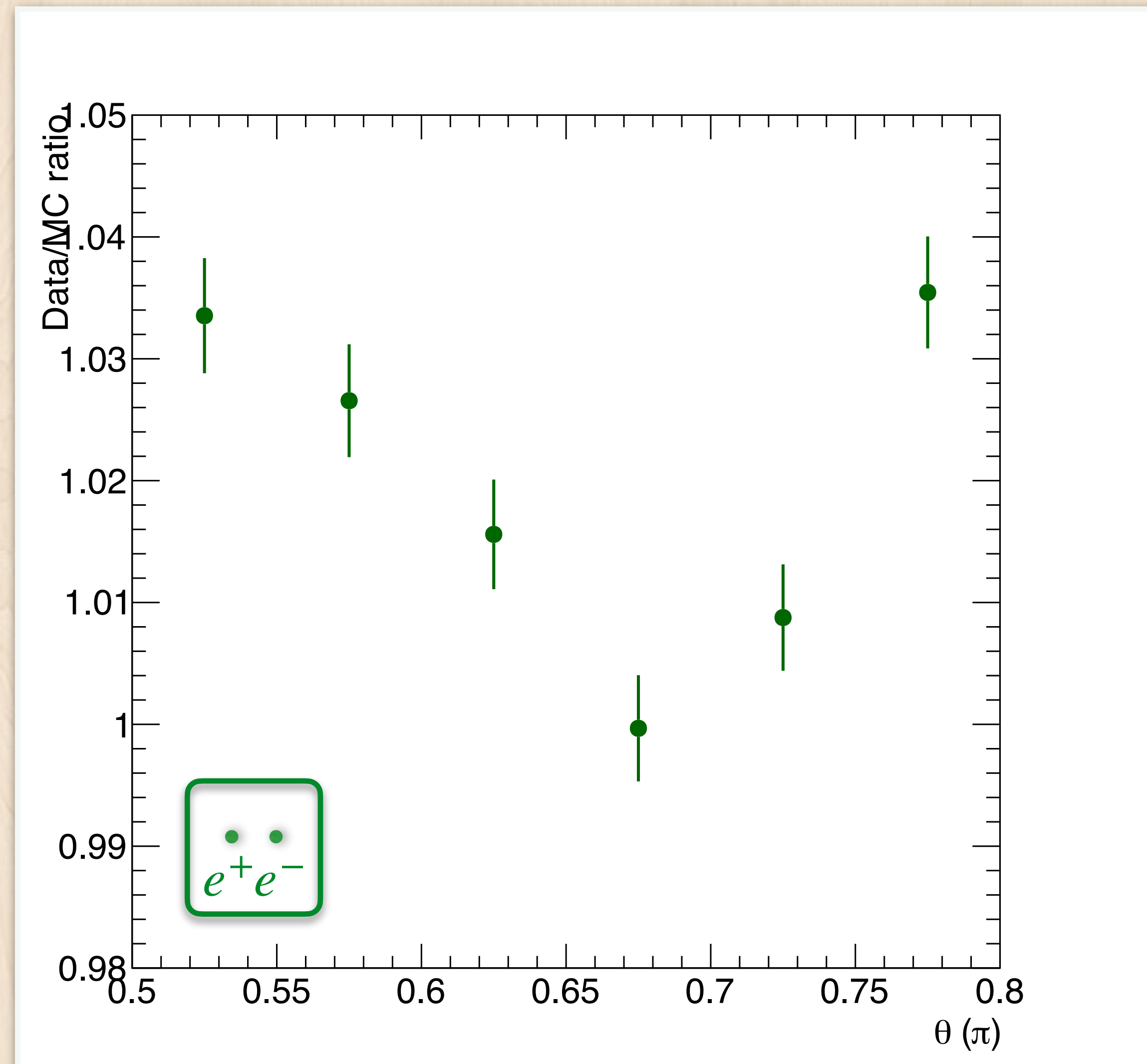
arXiv 2108.04877

Jet resolution

Jet resolution in simulation



Fiducial dijet —
vary 3rd-leading jet as systematics



Energy resolution: 10-25%

(Angular resolution: 0.01-0.05)

Up to 5% difference in energy resolution between data and MC

arXiv 2108.04877

Data-driven checks

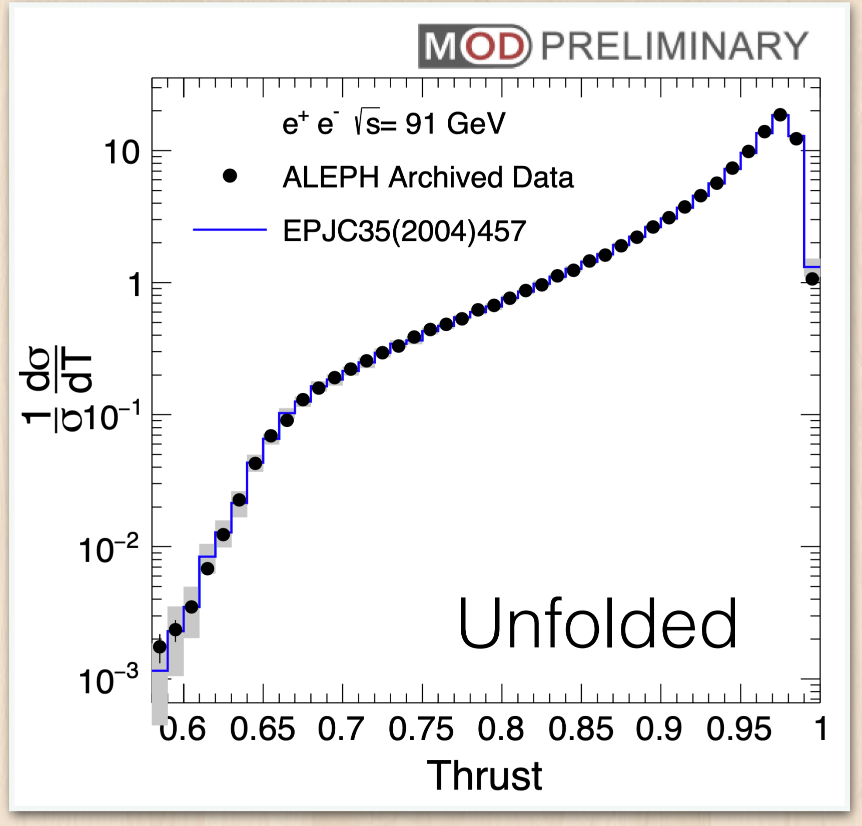
- Even though anti-kT jets did not exist during LEP, we are able to control data/MC differences with available information
  - Up to 1-2% for jet energy scale, and up to relative 4% for jet energy resolution

Some bottlenecks

- Example: PID information
- PID scores for different particle hypothesis are available: how likely it is a proton, Kaon, pion, etc
  - Supposedly one can cut and enrich particle type
- Not immediately clear how to **control data/MC differences** on the information

Reproducing published results

- Comprehensive data/MC comparisons
- Ultimate test of our understanding of the data
- Exact selection as QCD paper
- Thrust $T \equiv \max_{\hat{n}} \frac{\Sigma_i |\vec{p}_i \cdot \hat{n}|}{\Sigma_i |\vec{p}_i|}$
- Global event shape
    - Back to back dijet: T ~ 1



Lessons for future: accessing the data

- Mileage vary ***a lot*** depending on experiment (beyond ALEPH)
  - Make sense of the **format**: knowledge needed from members
  - Not easy to gain **control of stored information** — more lower-level information will be useful
  - Good to have more **sets of fully simulated MCs** available
- Many lessons for current & future experiments
  - Enough information for end-to-end measurements?
  - Best to do some "**user tests**" for open data as we go

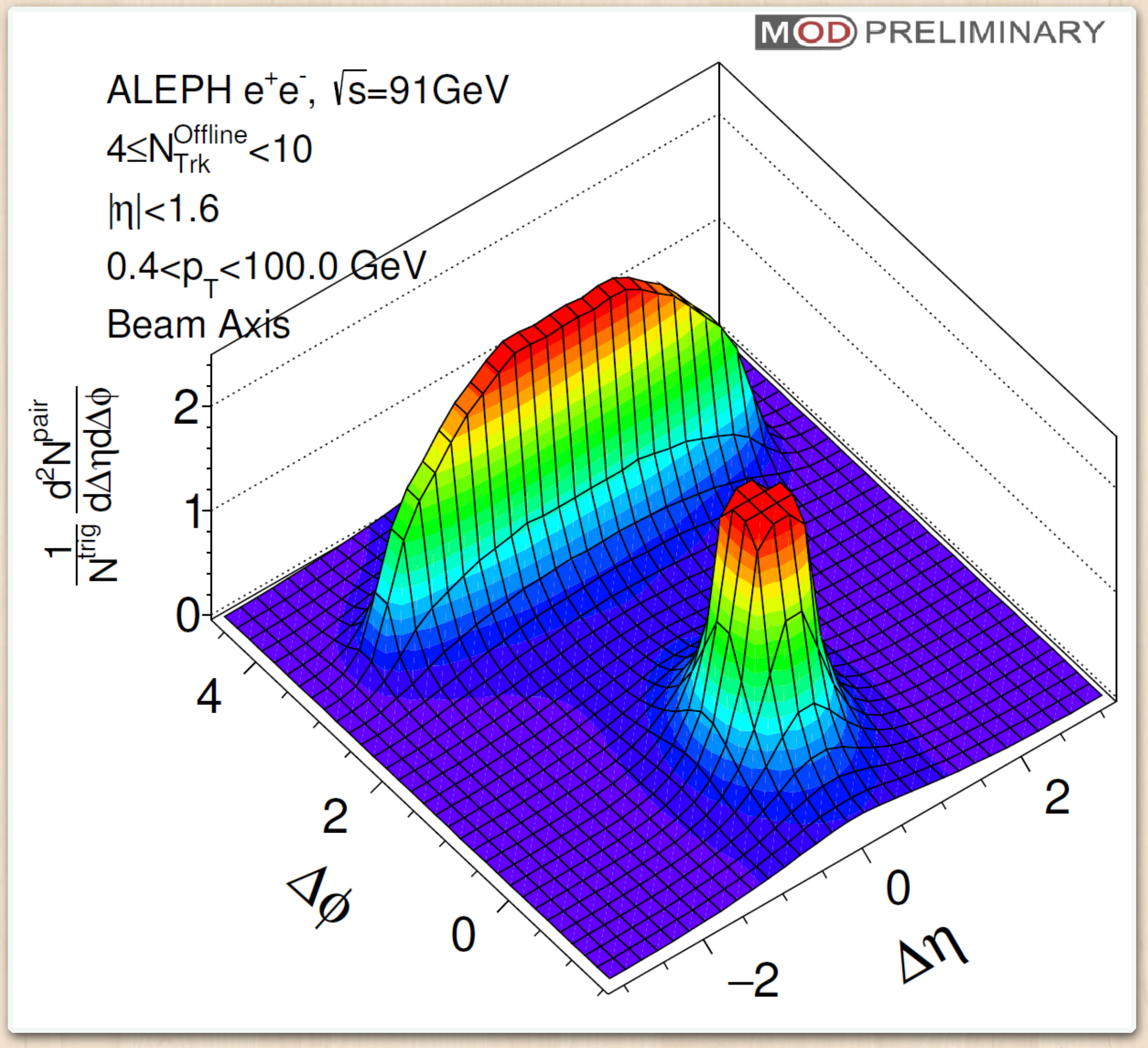
The re-analysis: 2PC

Analysis overview

Same event correlation $S(\Delta\eta, \Delta\phi)$

Correlate all pairs of particles from the same collision

However acceptance effects are in here

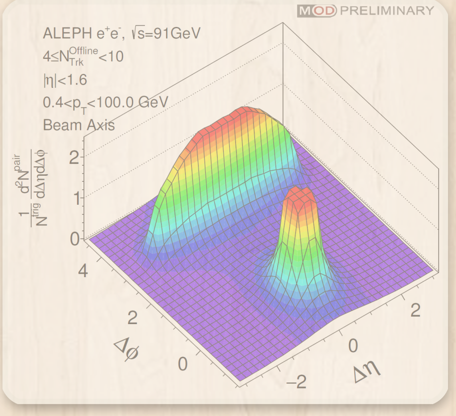


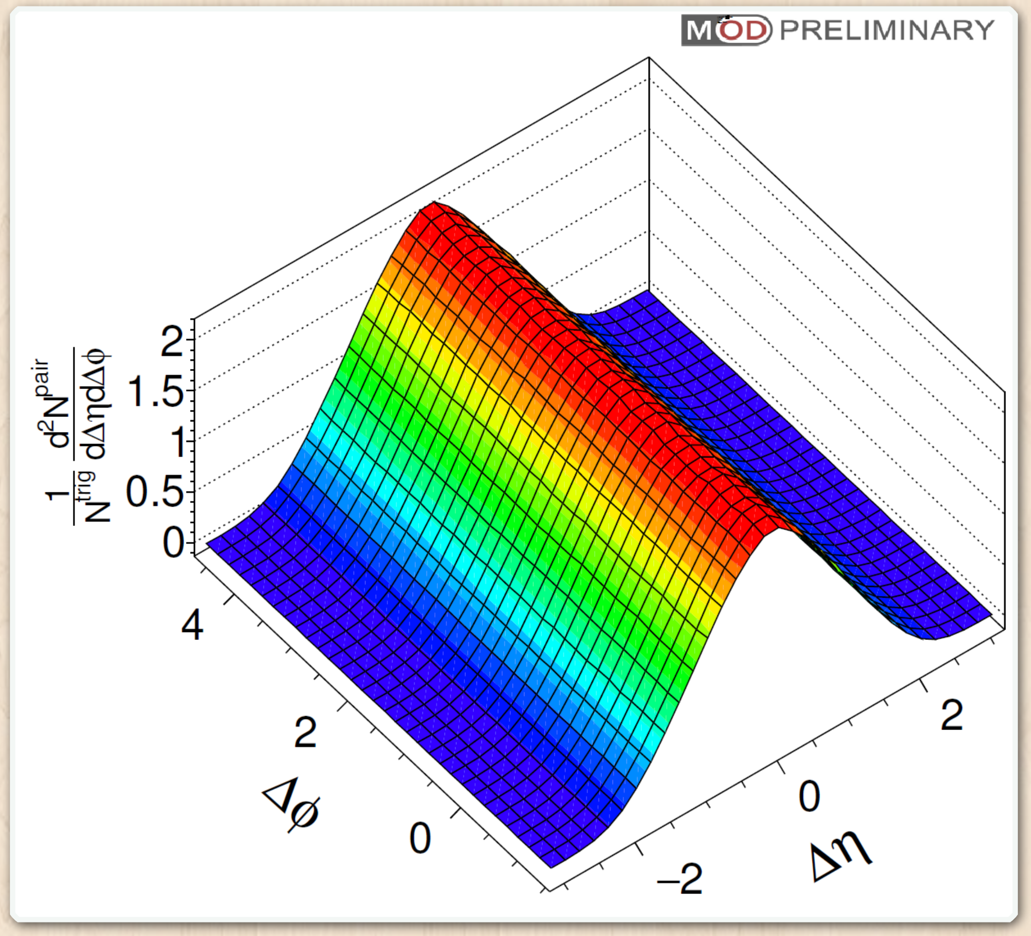
Analysis overview

Same event correlation $S(\Delta\eta, \Delta\phi)$

&

Mix event correlation $B(\Delta\eta, \Delta\phi)$





Correlate particles from different collisions

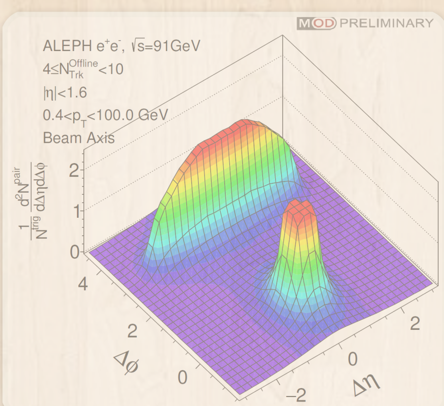
"Null hypothesis" without any physics correlations

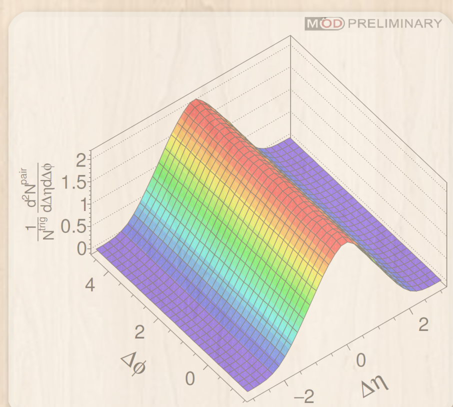
PRL 123 (2019) 212002

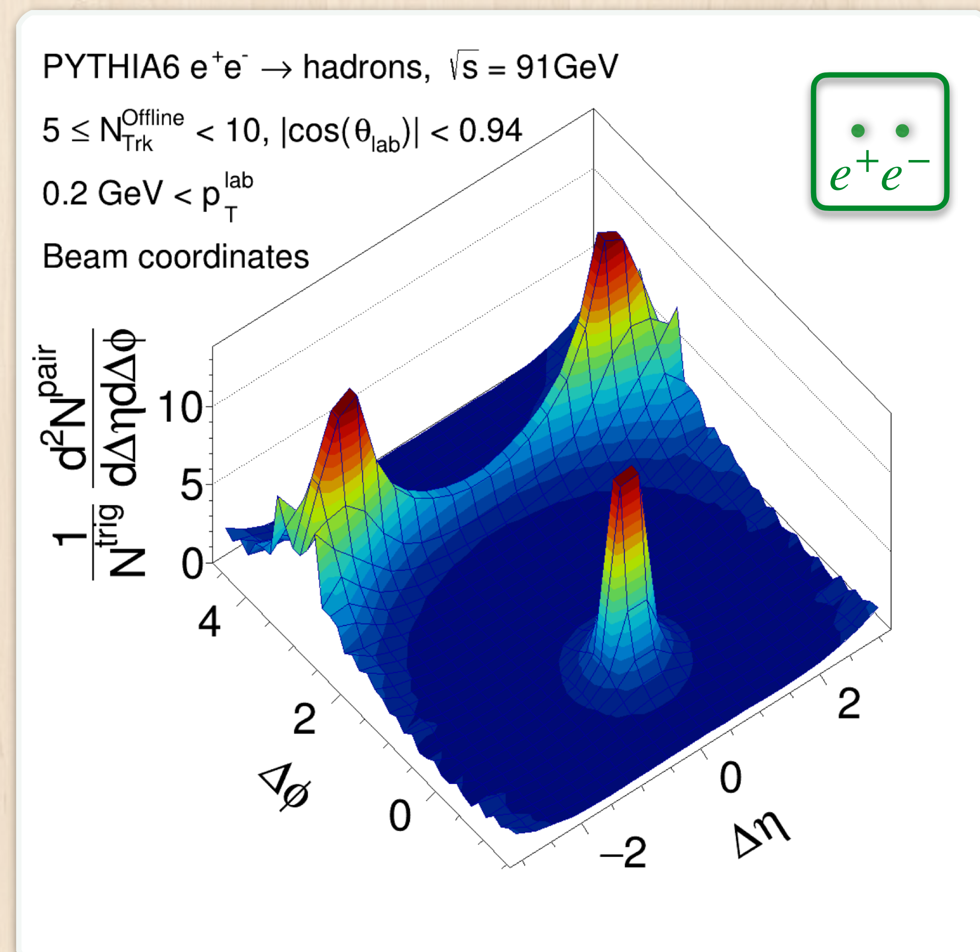
Analysis overview

Same event correlation $S(\Delta\eta, \Delta\phi)$ & Mix event correlation $B(\Delta\eta, \Delta\phi)$ = Acceptance-corrected signal

$$S(\Delta\eta, \Delta\phi) \times \frac{B(0,0)}{B(\Delta\eta, \Delta\phi)}$$



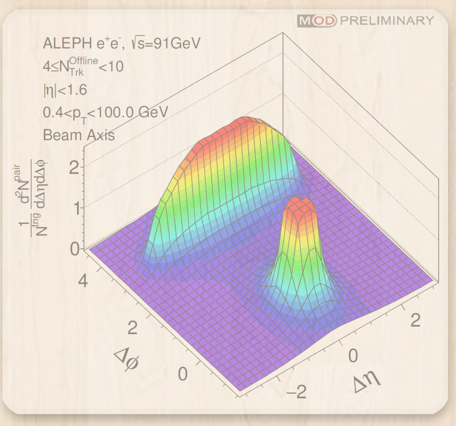


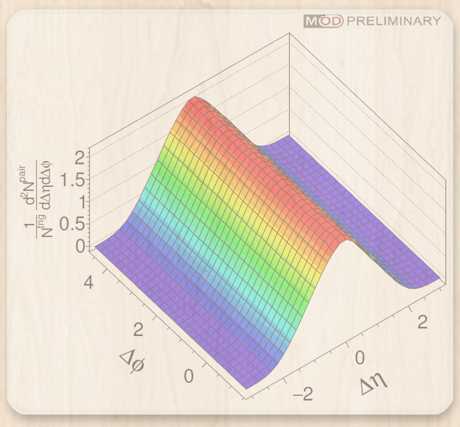


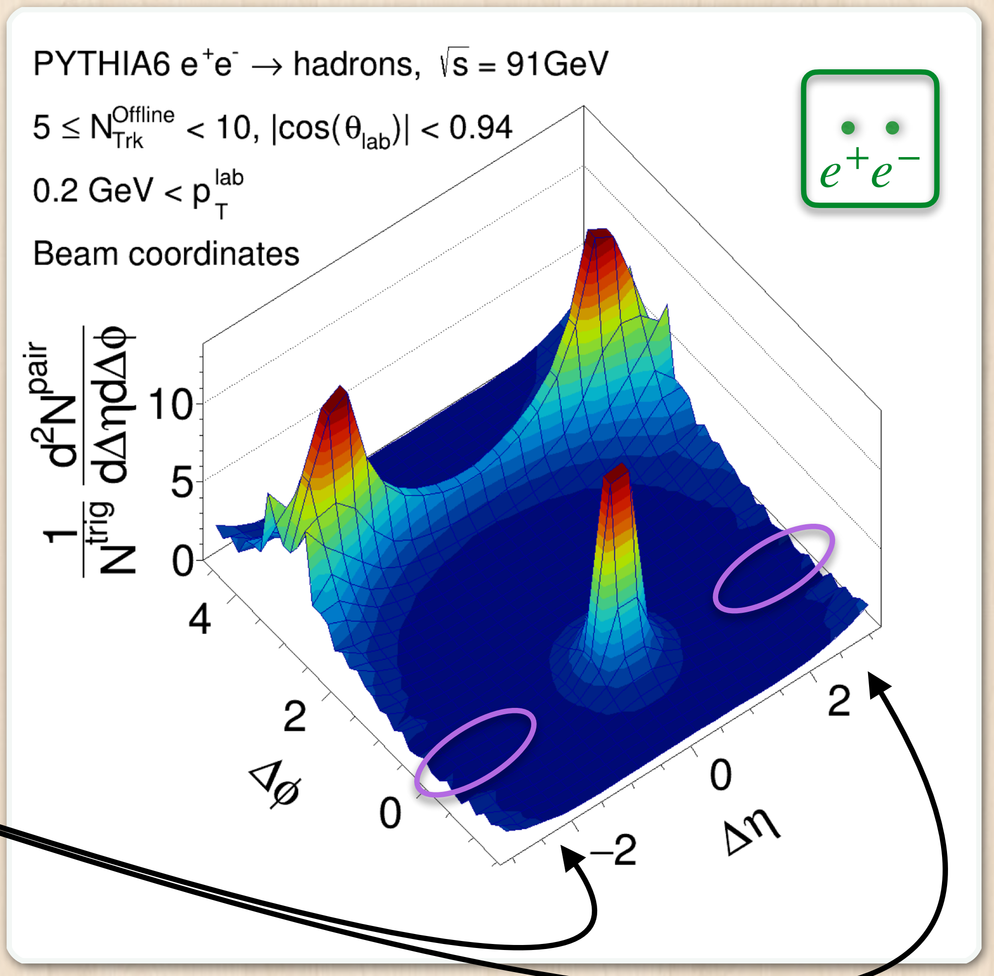
Keep only correlation from physics processes

Analysis overview

Same event correlation $S(\Delta\eta, \Delta\phi)$ & Mix event correlation $B(\Delta\eta, \Delta\phi)$ = Acceptance-corrected signal →



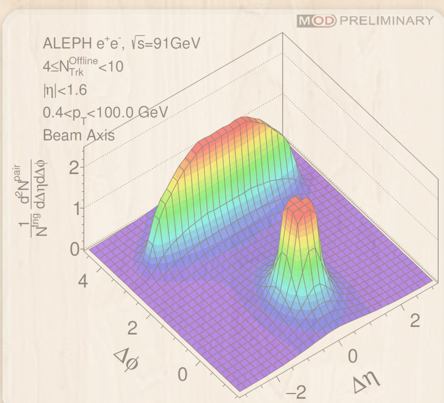


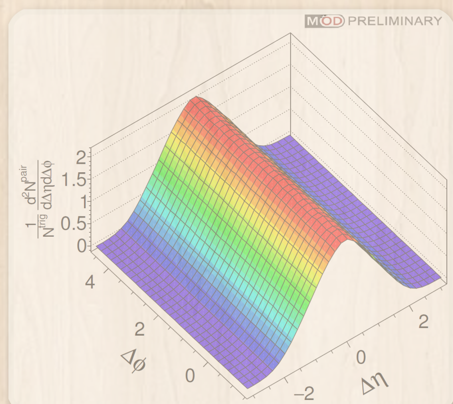


Look at near-side long-range correlation in (η, ϕ) space

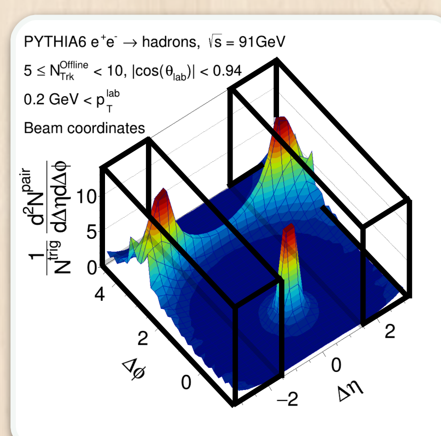
Analysis overview

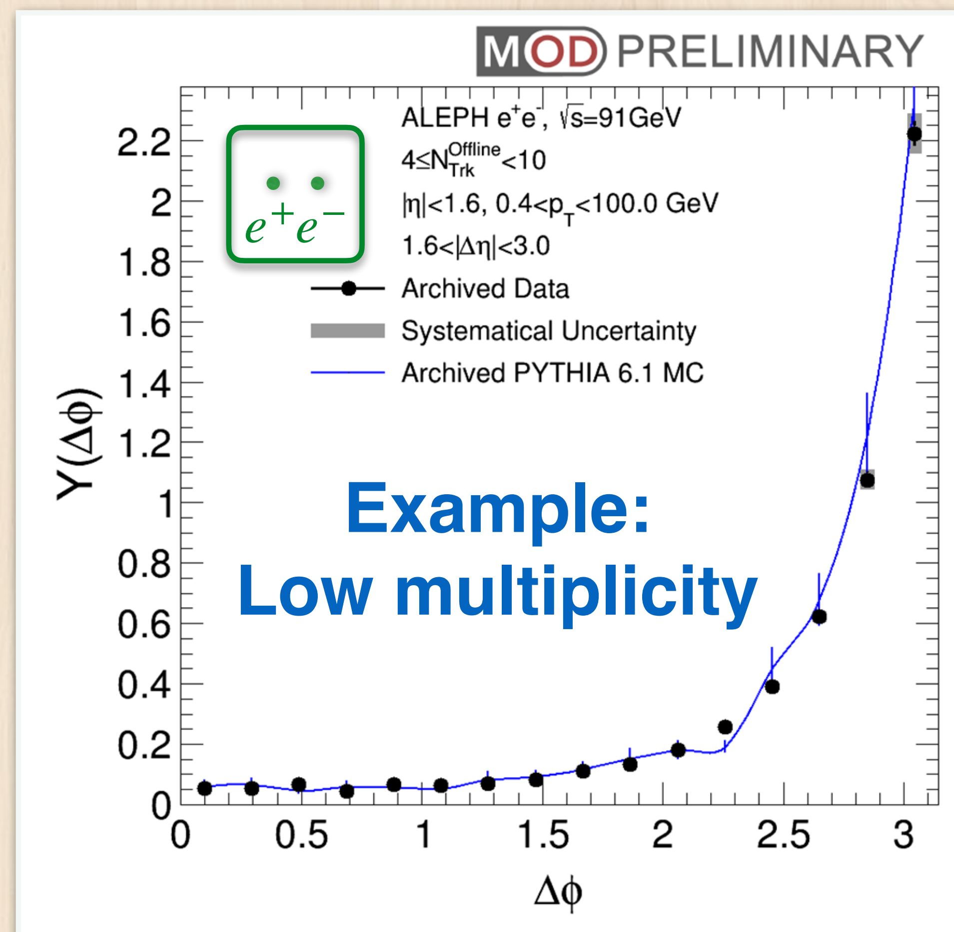
Same event correlation $S(\Delta\eta, \Delta\phi)$ & Mix event correlation $B(\Delta\eta, \Delta\phi)$ = Acceptance-corrected signal → Projection





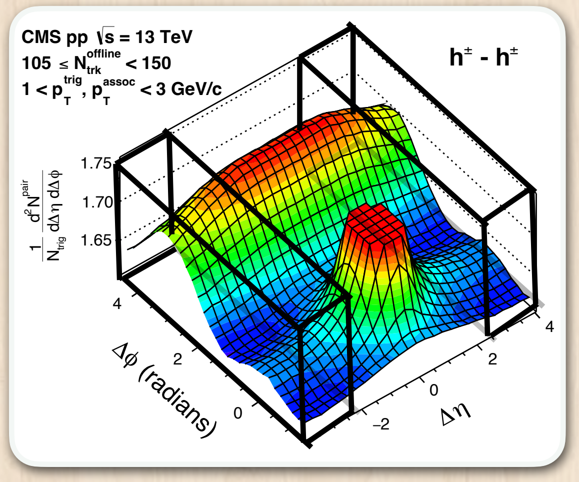
Project onto $\Delta\phi$ for further studies



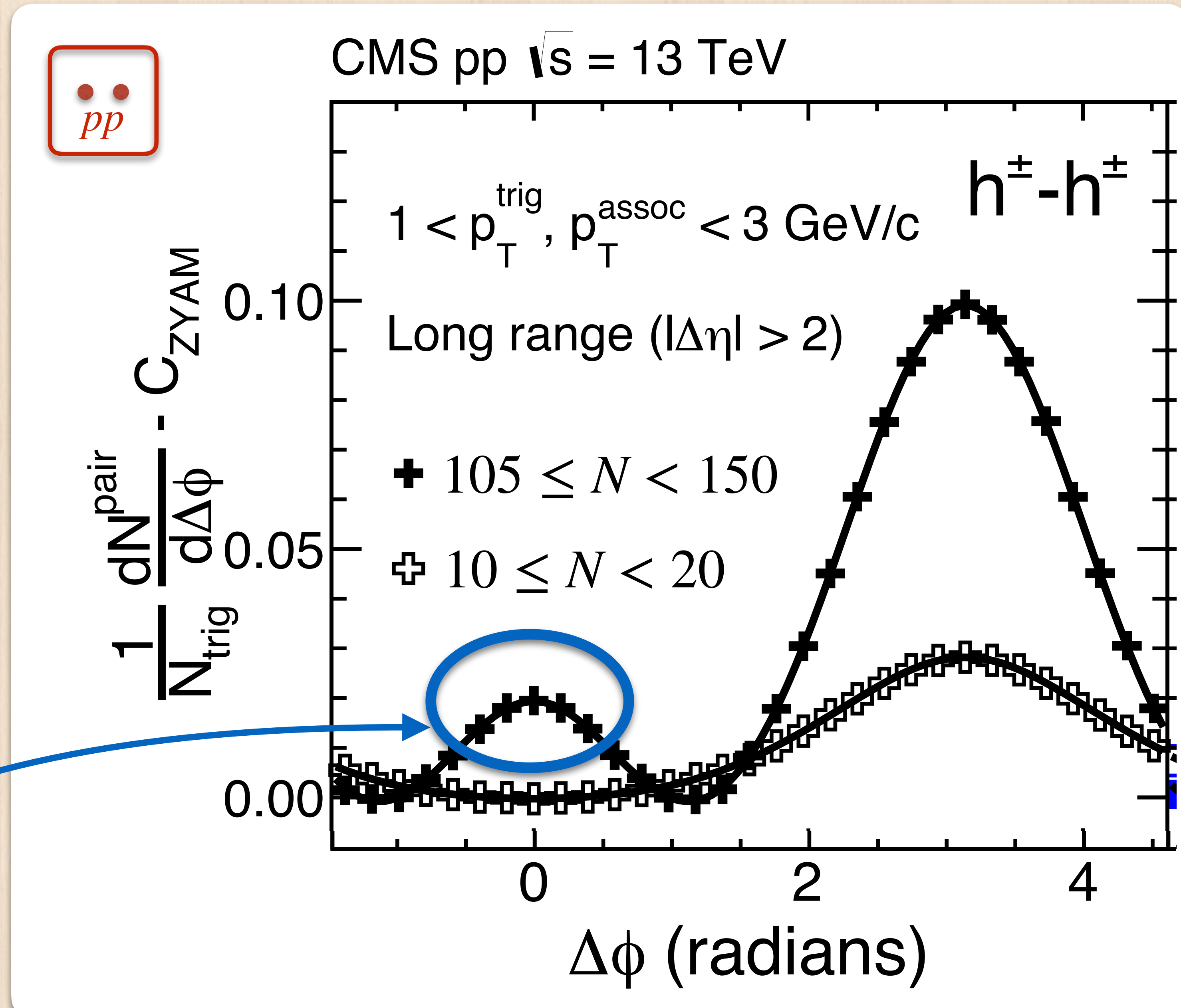


PRL 123 (2019) 212002

How ridge looks like in pp



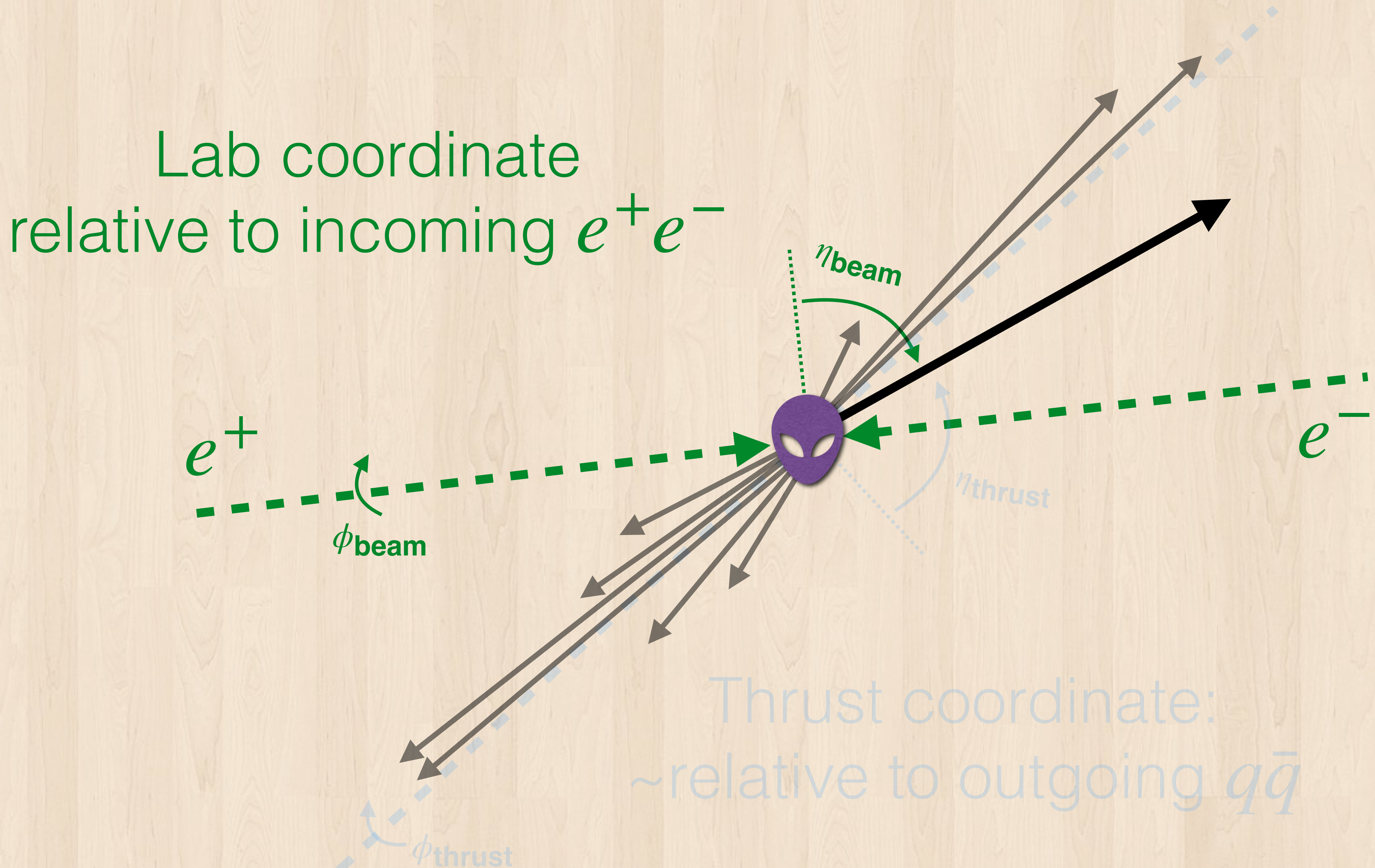
pp



Excess at small $\Delta\phi$

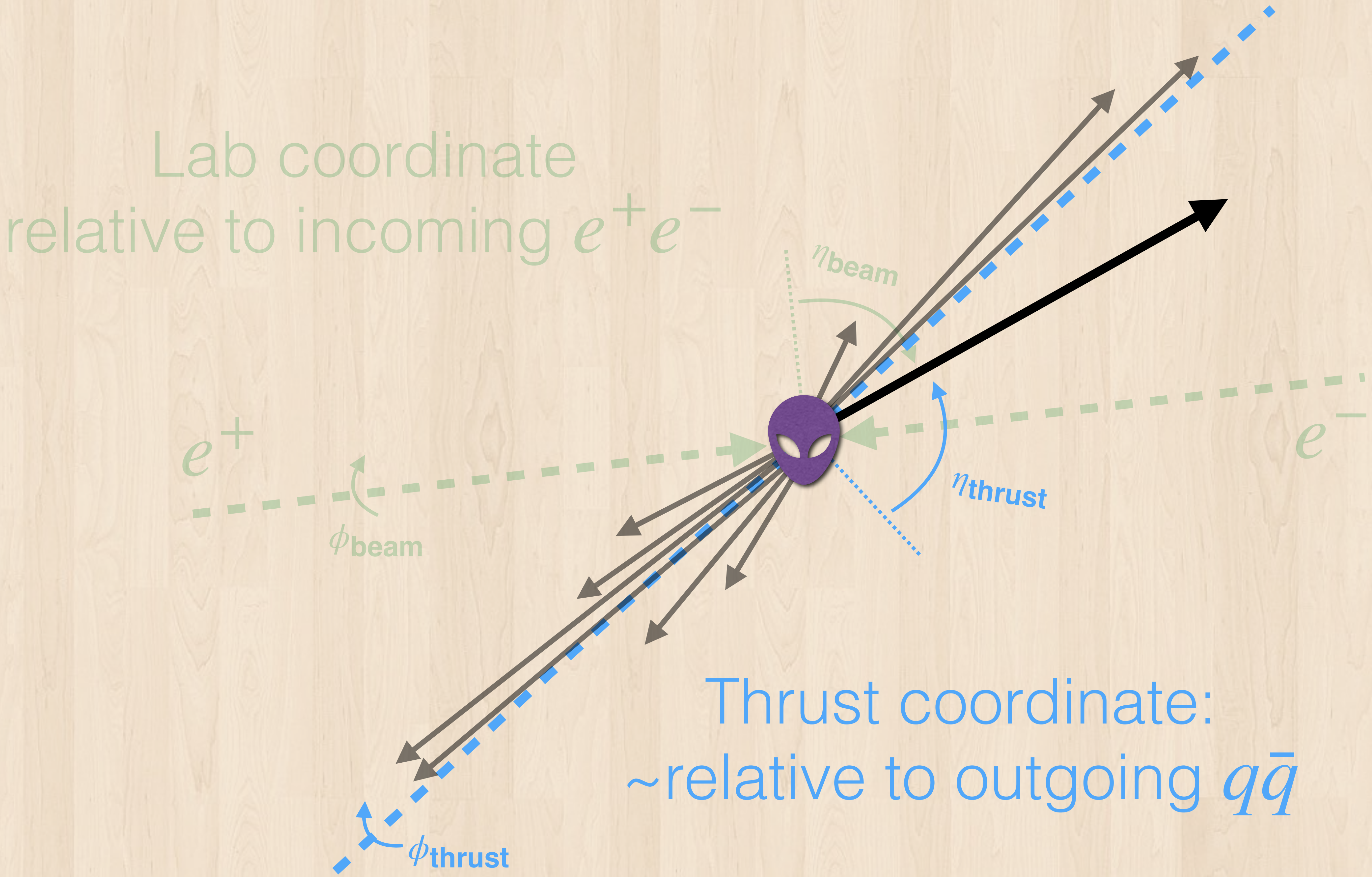
PLB 765 (2017) 193

The two coordinate systems



Analogous to hadron collider setup

The two coordinate systems



~direction of color-string in $q\bar{q}$ topology

Baseline event selections

- Following methods from previous ALEPH publication
- Select **hadronic** events
  - Number of good charged particles >= 5
  - Number of good particles >= 13
  - $E_{charged} > 15$ GeV
  - $|\cos(\theta_{sphericity})| < 0.82$: ensure collision well-contained in the detector

Phys. Rep. 294 (1998)

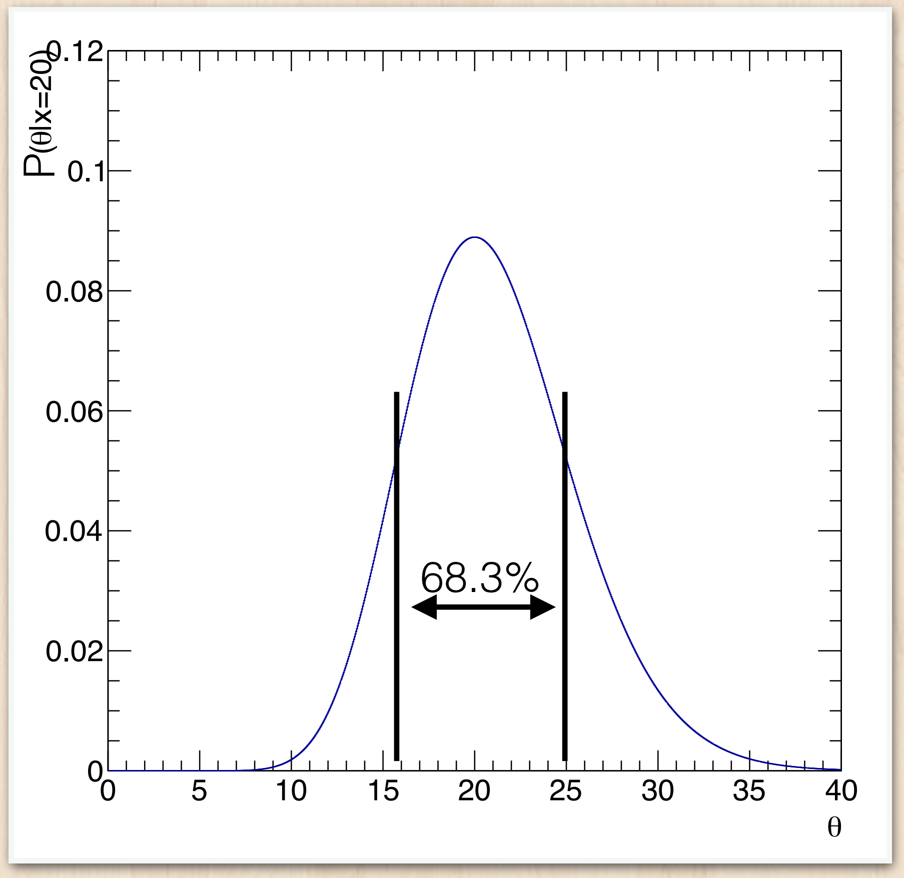
Baseline event selections: LEP2

- In LEP2, **initial state QED radiation** is significant
  - Reject collisions with a lot of QED radiation
- Method from previous ALEPH publication
  - First, ignore “QED jets” and examine the rest
  - Effective center-of-mass energy $\sqrt{s'} > 0.9\sqrt{s}$
  - Visible invariant mass $M_{vis} > 0.7\sqrt{s}$

Eur. Phys. J. C 35 (2004) 457

Uncertainty: Bayesian approach

- Things not always Gaussian: **Bayesian approach**
- Construct posterior $P(\theta|x)$ using Bayes' theorem
  - $P(\theta|x) = P(x|\theta)P(\theta)/P(x)$
  - Probability of some value θ to be true given observed data x
- Example: counting experiment
- Then we quote central value and uncertainty (68% most likely interval)



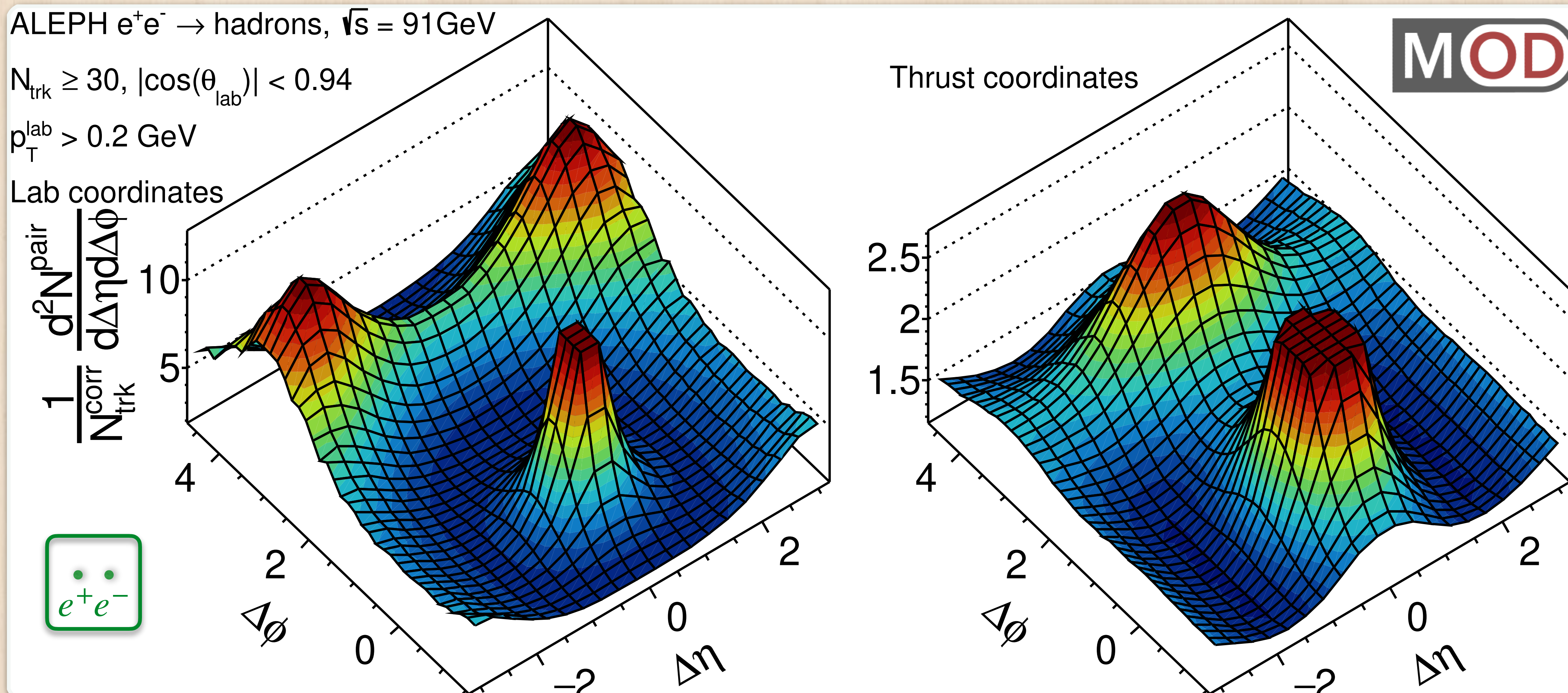
Propagating uncertainty

- In the Bayesian paradigm everything has a **probability density** interpretation
- **Monte-Carlo technique** can be used to **propagate uncertainty** to nontrivial observables
  - For example, associated yield across many bins
- Extensive internal studies show that this approach is reasonable and robust

Results

LEP1: High multiplicity

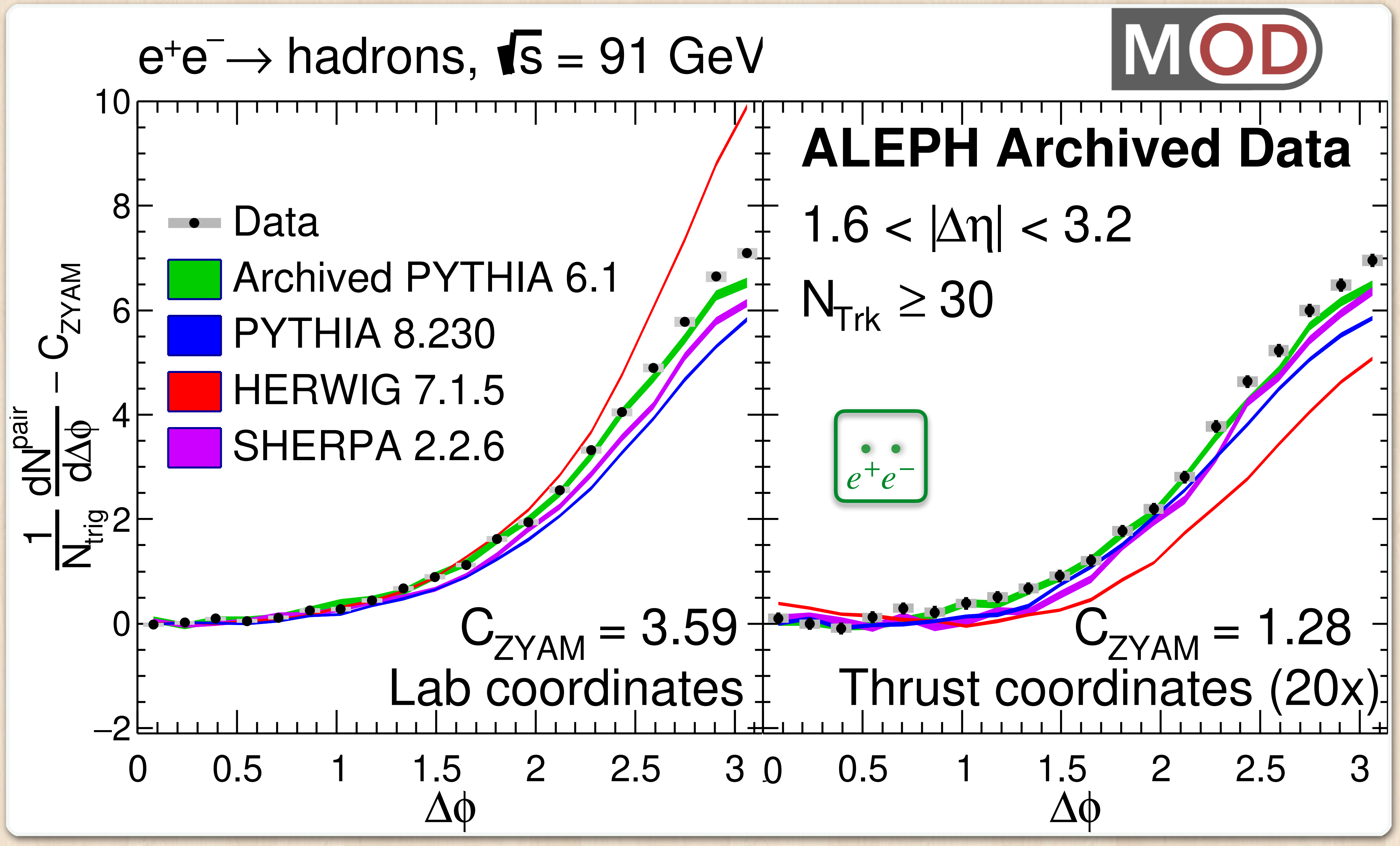
Focus on high multiplicity here: $N_{trk} \geq 30$



Many interesting features!

For now let's focus on the ridge search

No sign of ridge 😣 (LEP1)

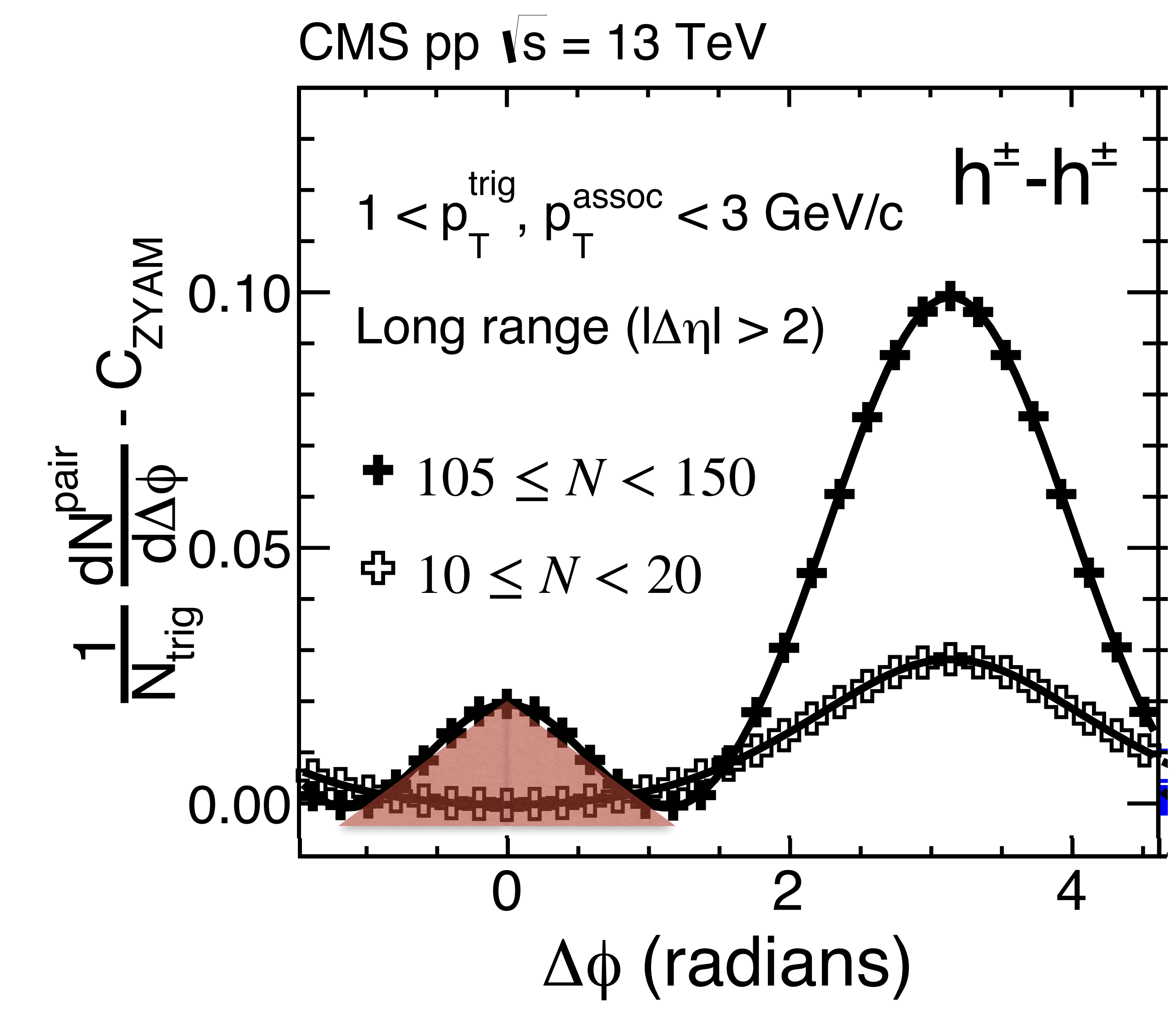


Also no ridge-like excess in other multiplicity bins

PRL 123 (2019) 212002

Ridge-like yield extraction

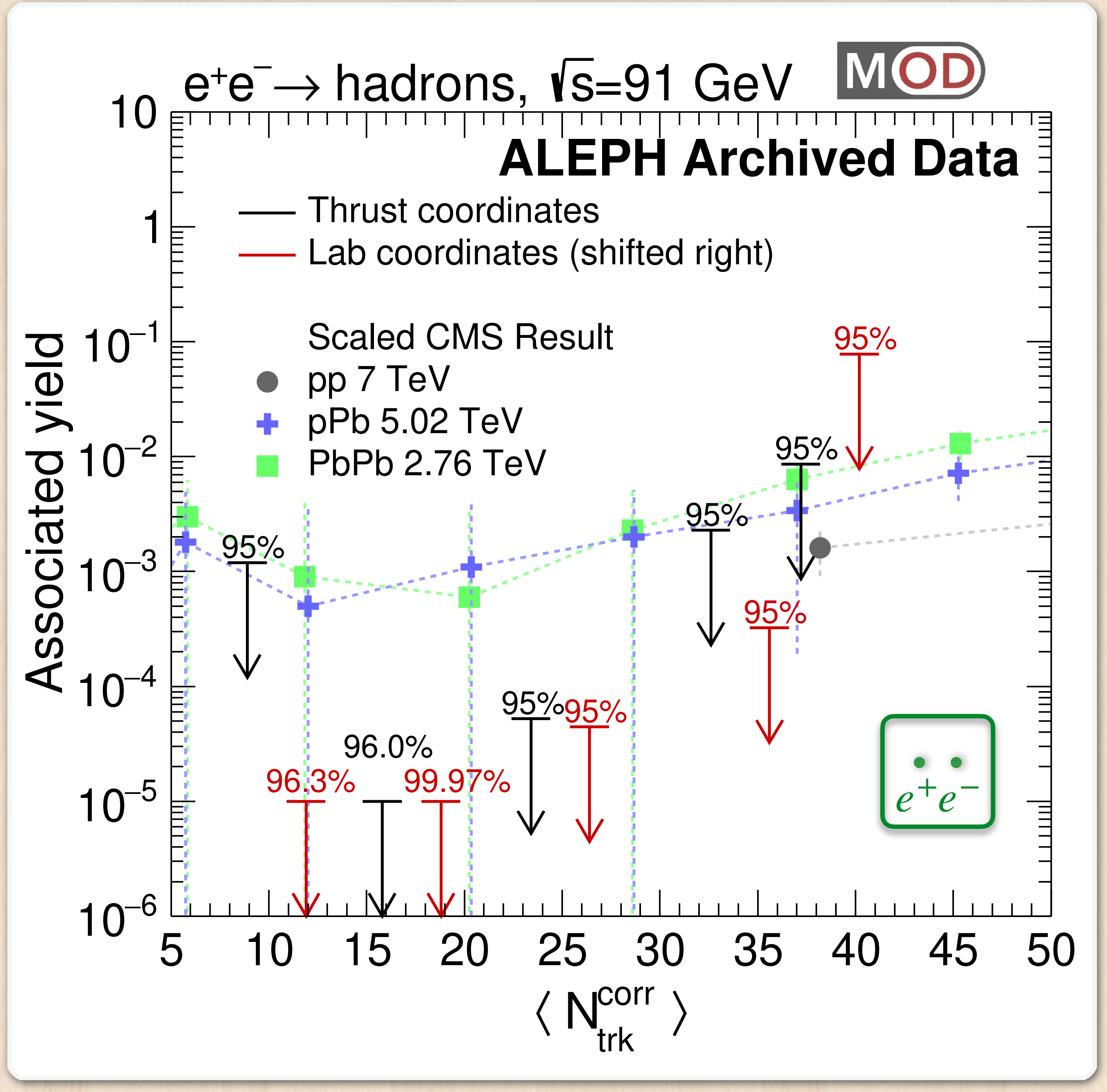
pp



1. Find minimum point
2. Integrate area under it

Phys.Lett.B 765 (2017) 193

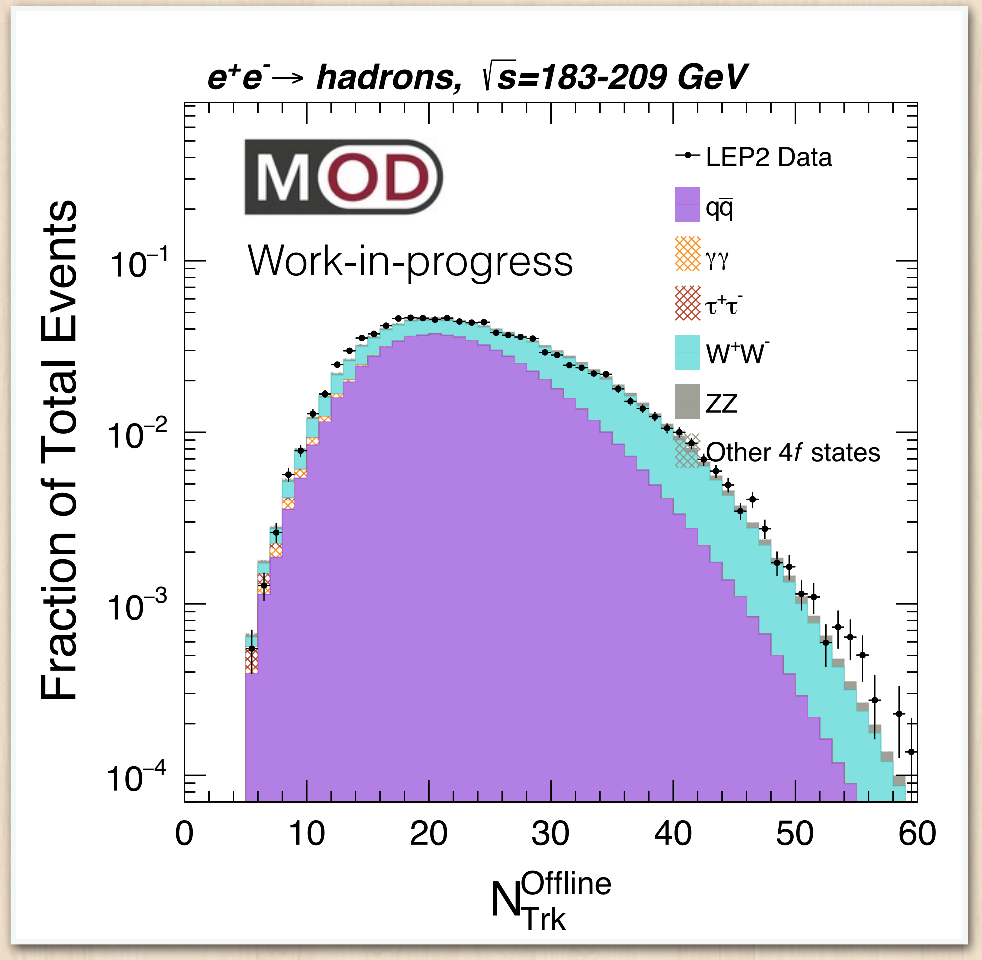
Limits on ridge yield: LEP1



No significant ridge observed — we proceed to set limit

What about higher energy e^+e^- data?

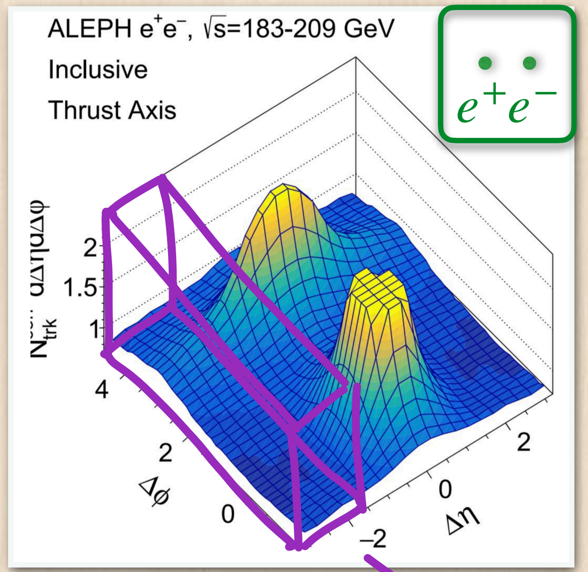
Higher energy LEP2 data

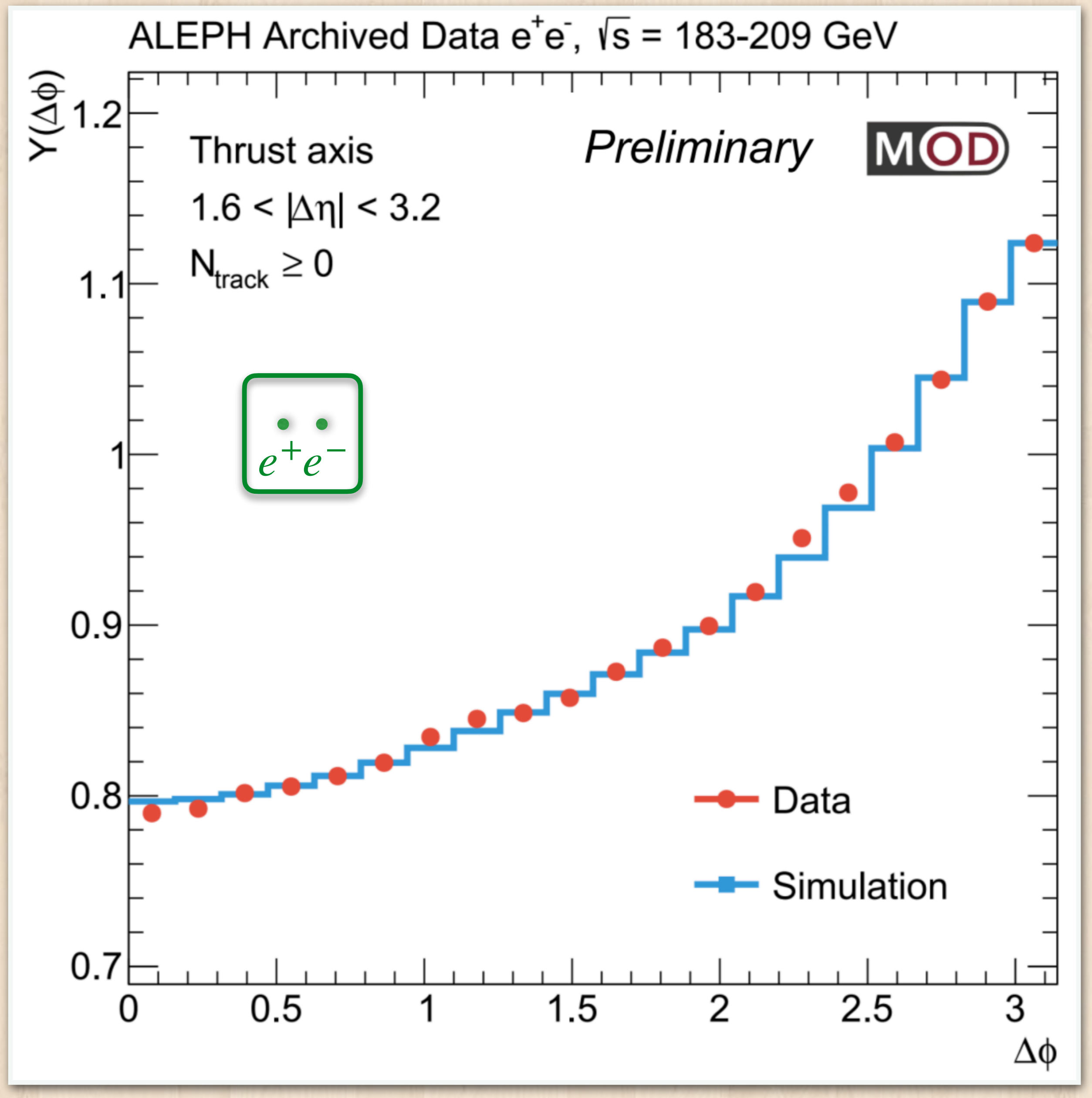


Integrated over higher energy datasets

Generally decent data-MC agreement

Inclusive correlation function

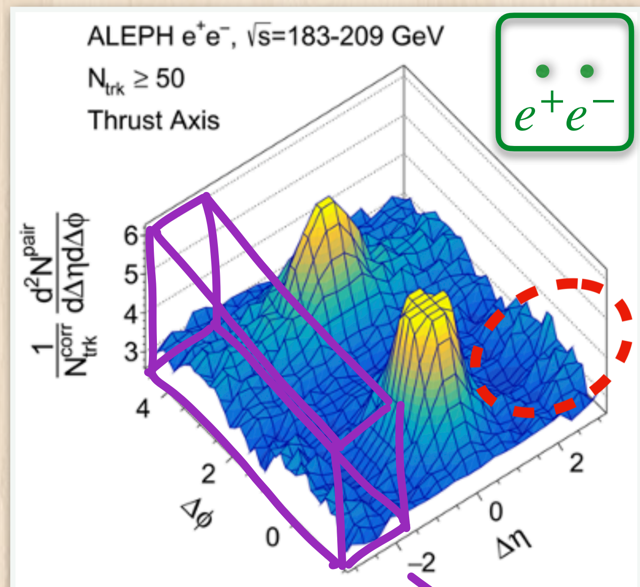


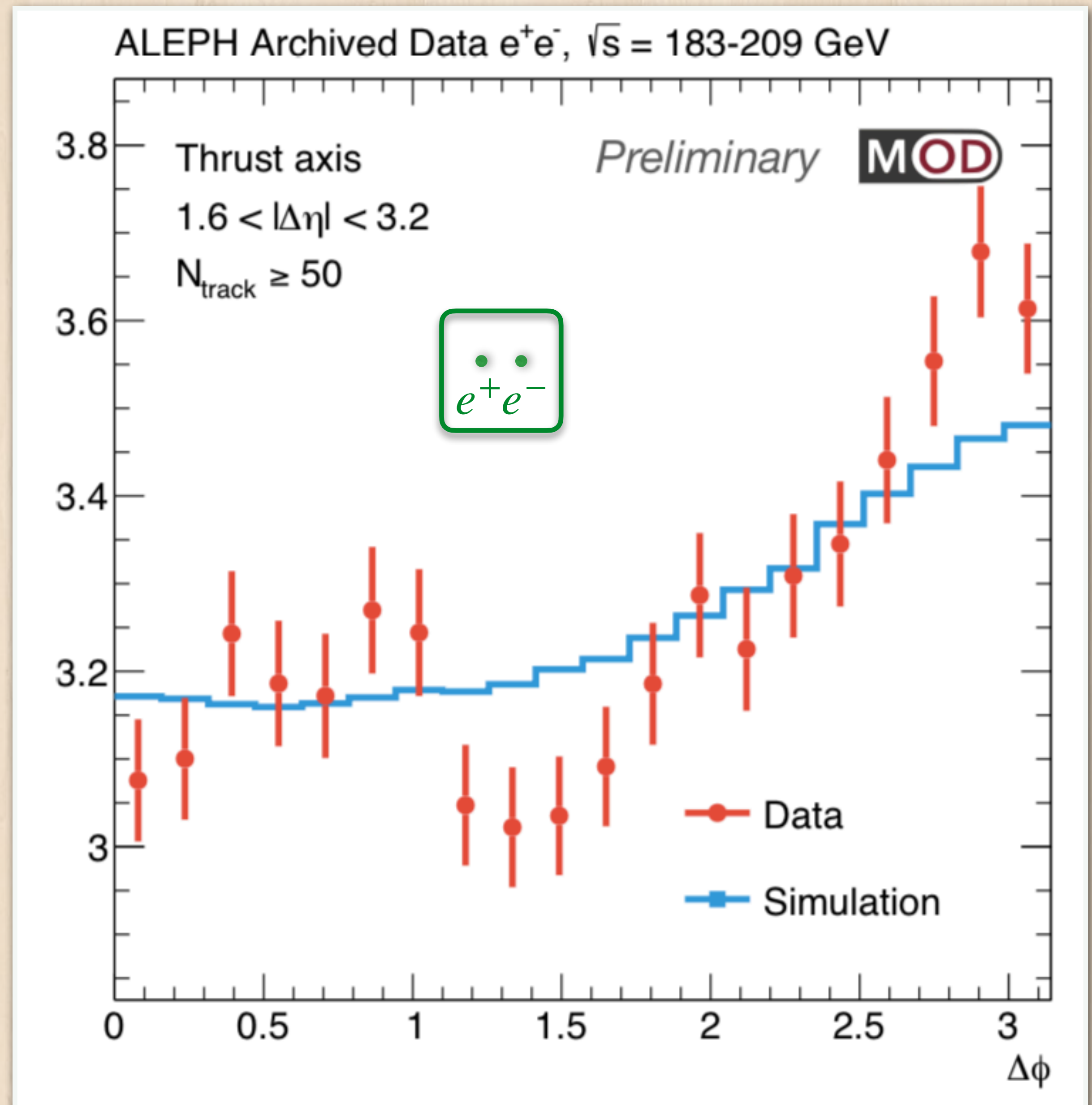


Decent data-MC agreement

Overall things match quite well

Focus on high multiplicity





Intriguing structure!

Work mostly by Y.-C. Chen

Quantify with Fourier components

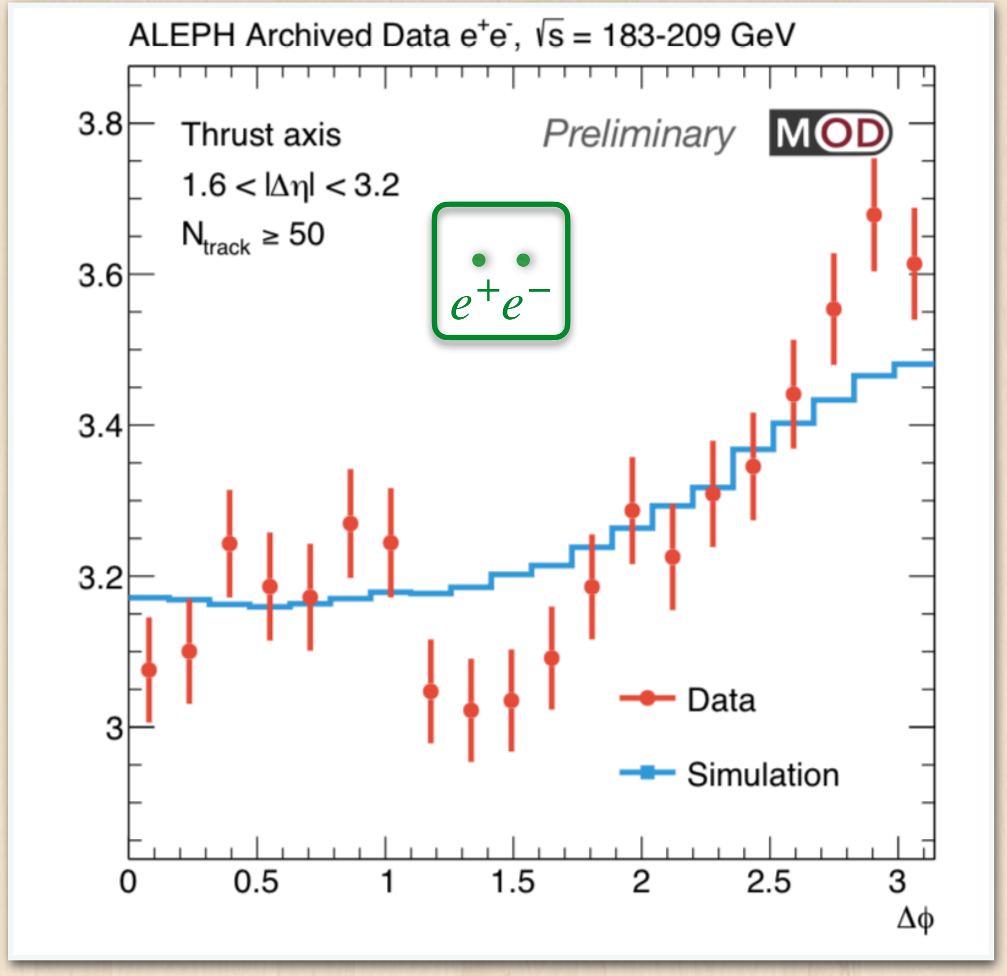
Fourier decomposition of the 2-particle distribution:

$$Y(\Delta\phi) \propto 1 + \sum_n 2V_n \cos(n\Delta\phi)$$

Coefficients of "single particle":

$$v_n \equiv \text{sign}(V_n)\sqrt{|V_n|}$$

Assumes factorization

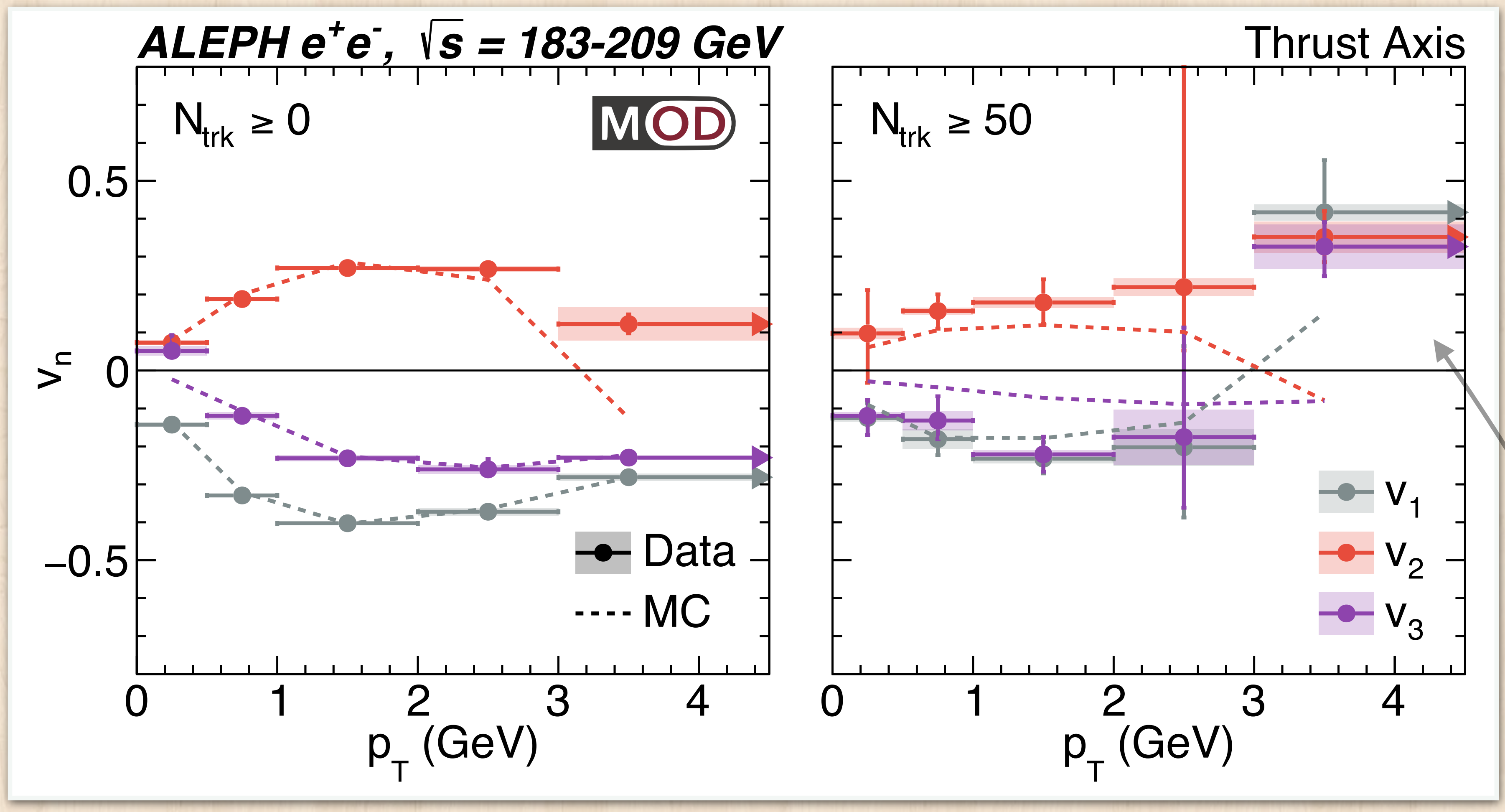


arXiv 2312.05084

Work mostly by Y.-C. Chen

Quantify with Fourier components

Inclusive: decent data-MC agreement



Smaller magnitude for large multiplicity

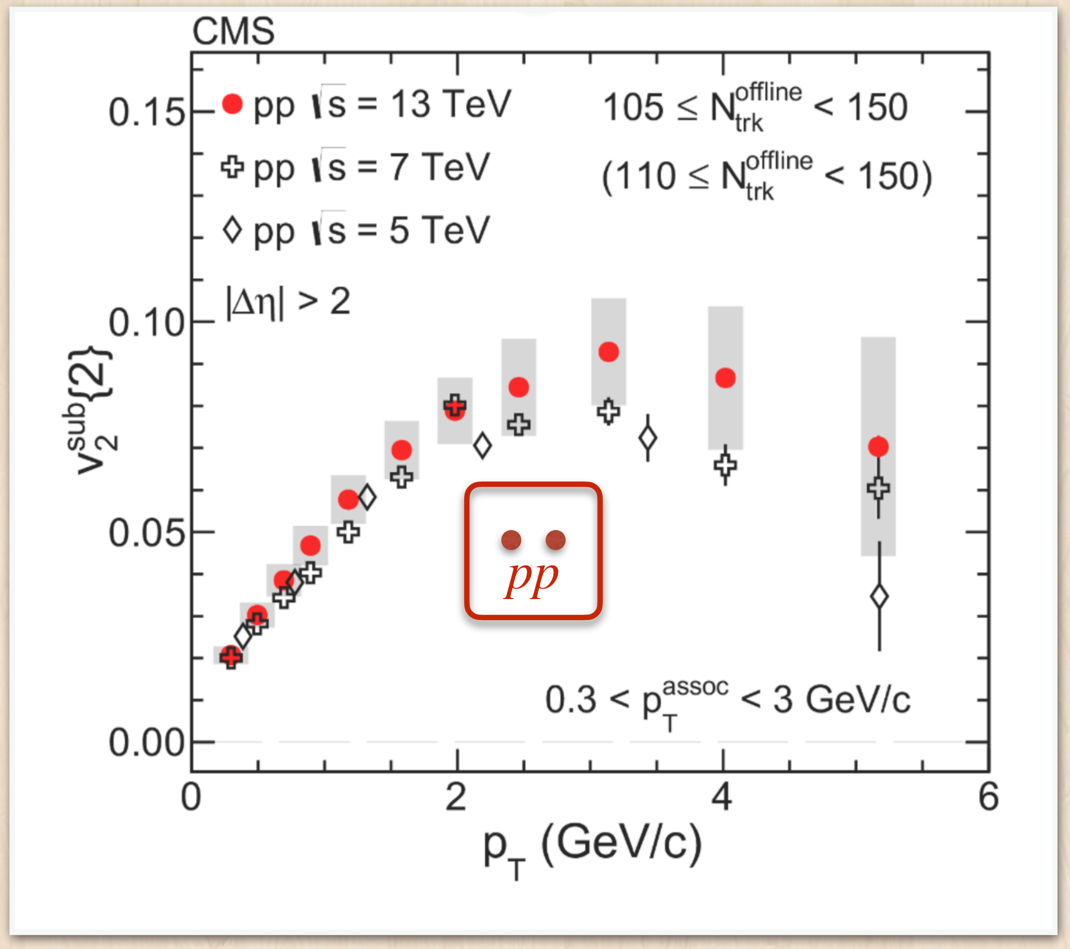
High multiplicity: magnitude larger in data

Sign change for v_1 and v_3

arXiv 2312.05084

Work mostly by Y.-C. Chen

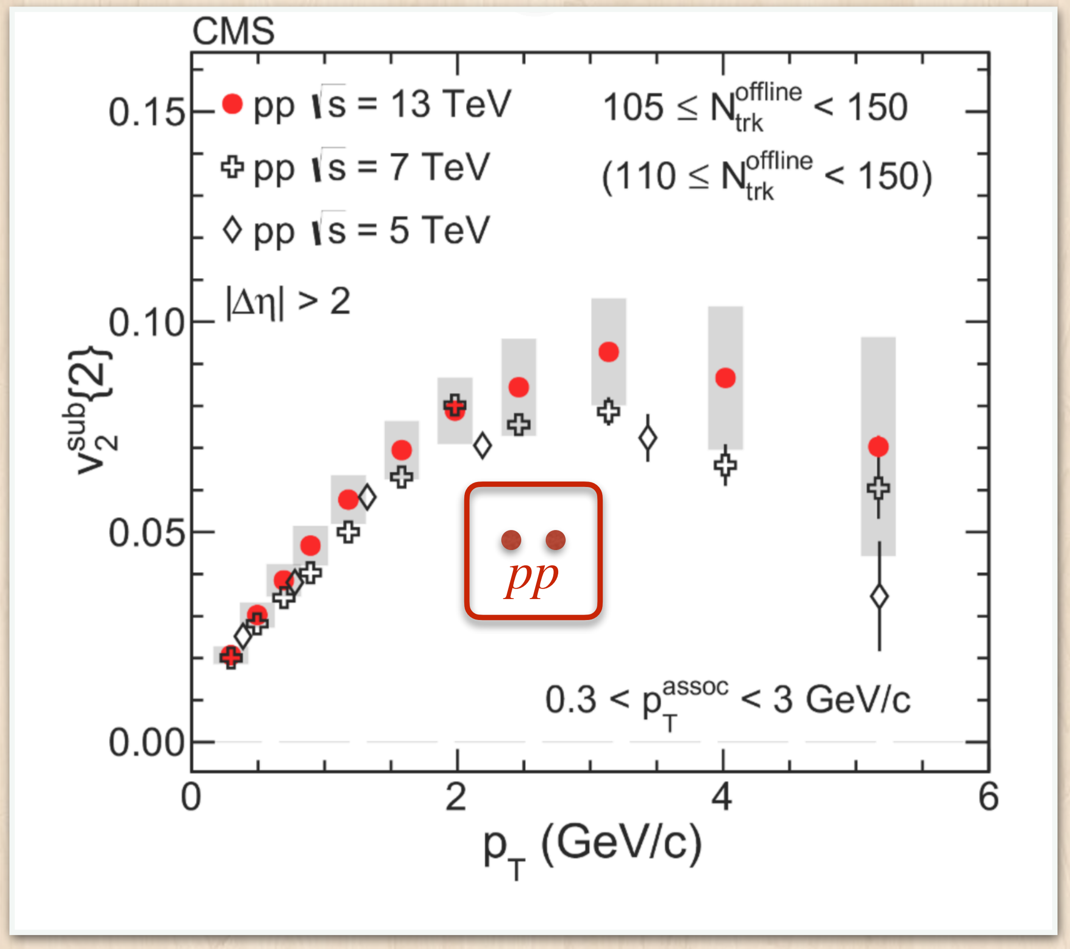
Comparisons to LHC

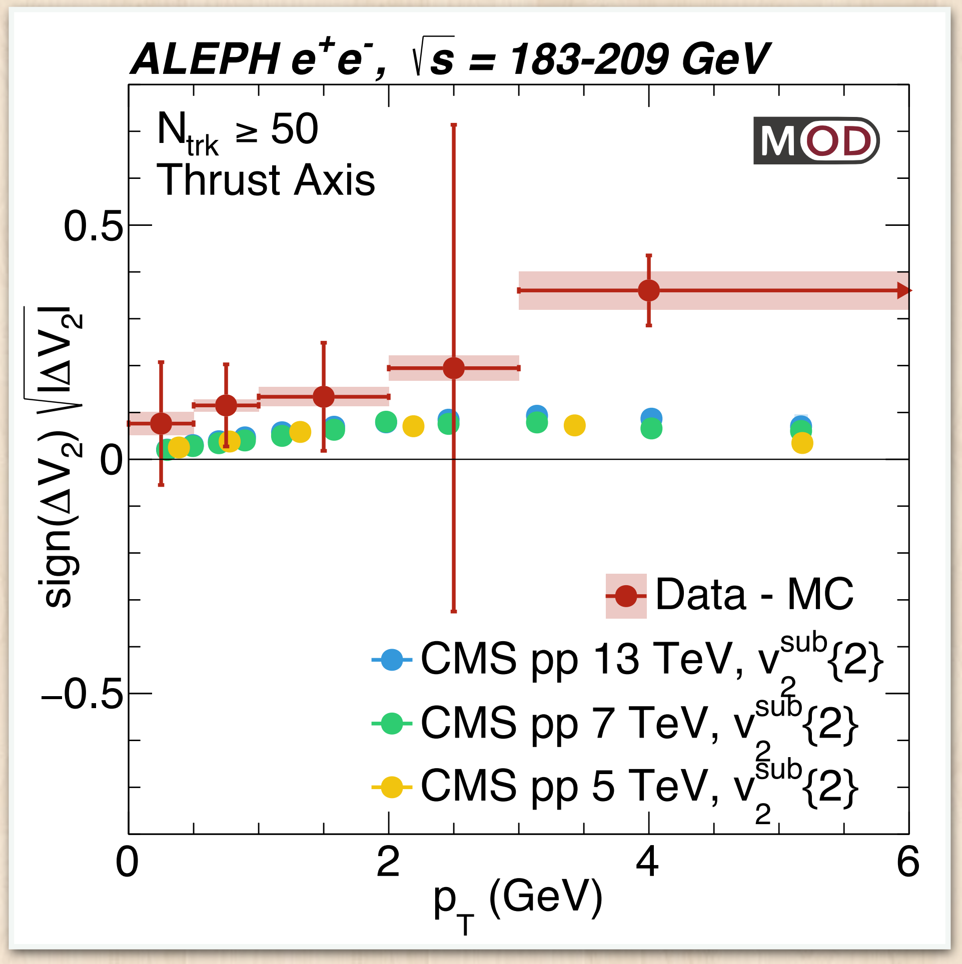


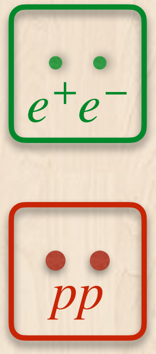
High multiplicity *pp*

Non-zero v_2

Comparisons to LHC







High multiplicity pp

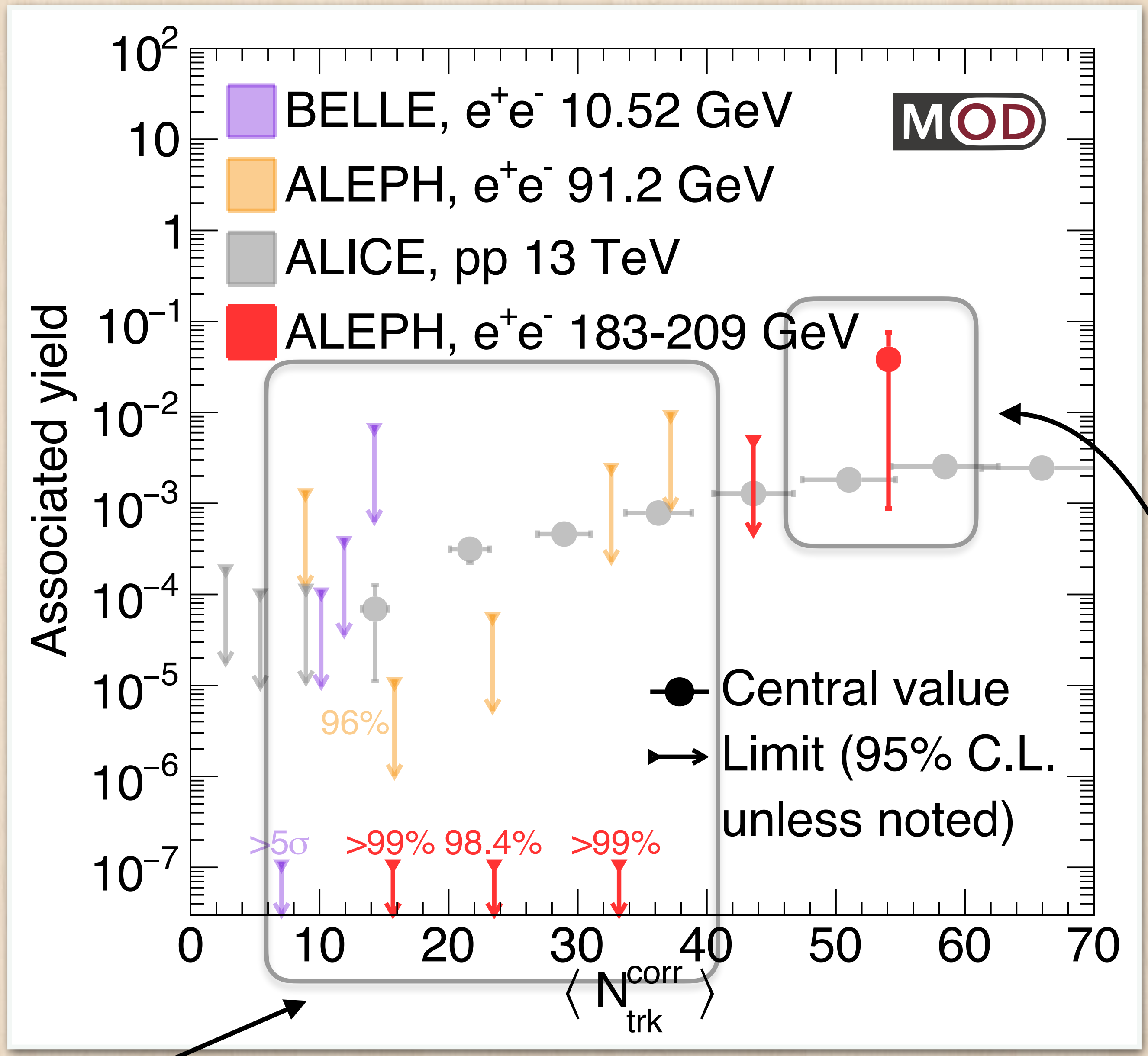
Non-zero v_2

e^+e^-: v_2(data) − v_2(MC)

Interesting trend 🤔

=> final state effect?

Quantifying the result: ridge yield



Associated yield lower than what ALICE found in pp

Uncertainty still large but interesting trend

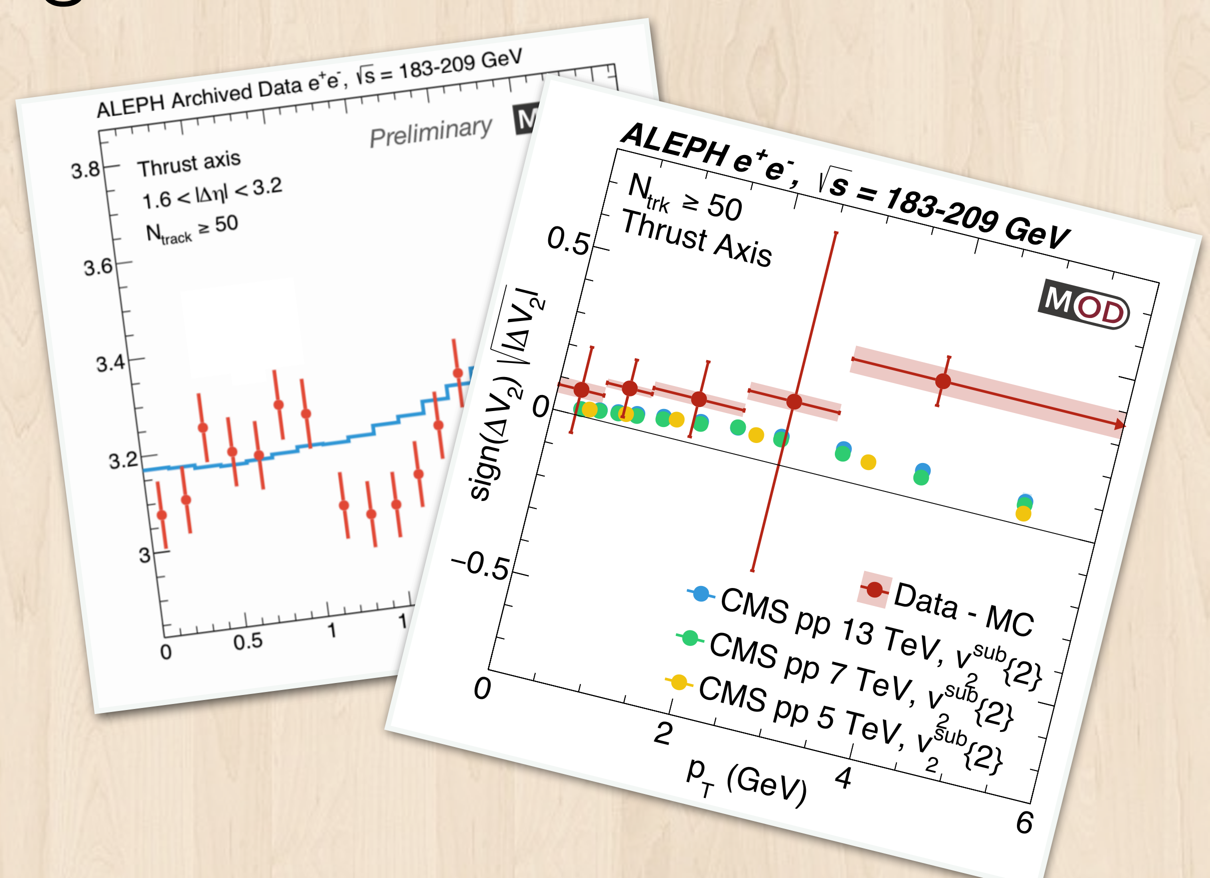
arXiv 2312.05084, 2306.04808

PRL128(2022)14; PRL123(2019)21

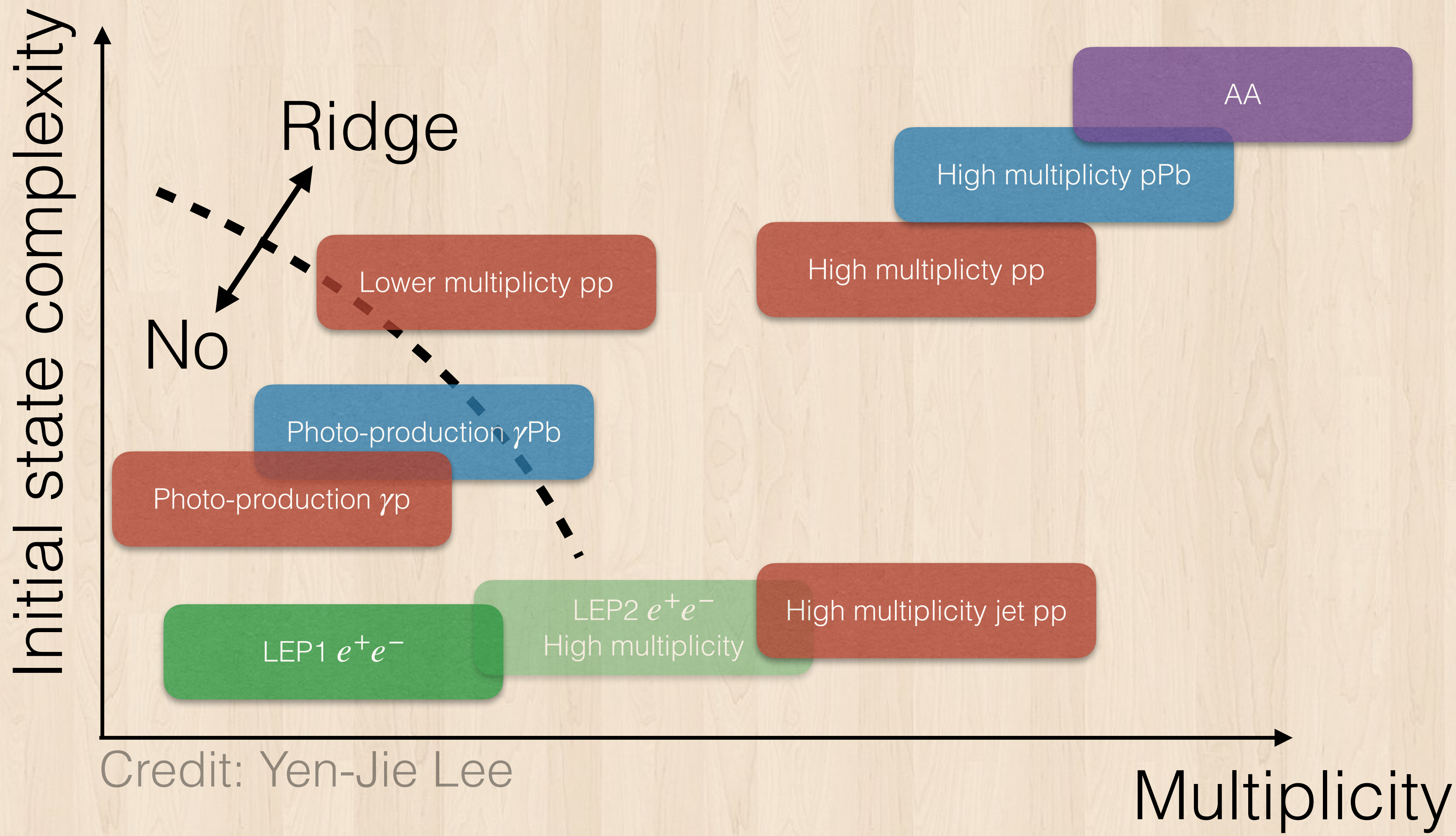
Concluding Remarks

Summary: two-particle correlation

- First measurement of two-particle correlation function for e^+e^- collisions up to 209 GeV at LEP
- No significant ridge-like signal at 91 GeV
- LEP2 with thrust axis: interesting structure in high multiplicity events



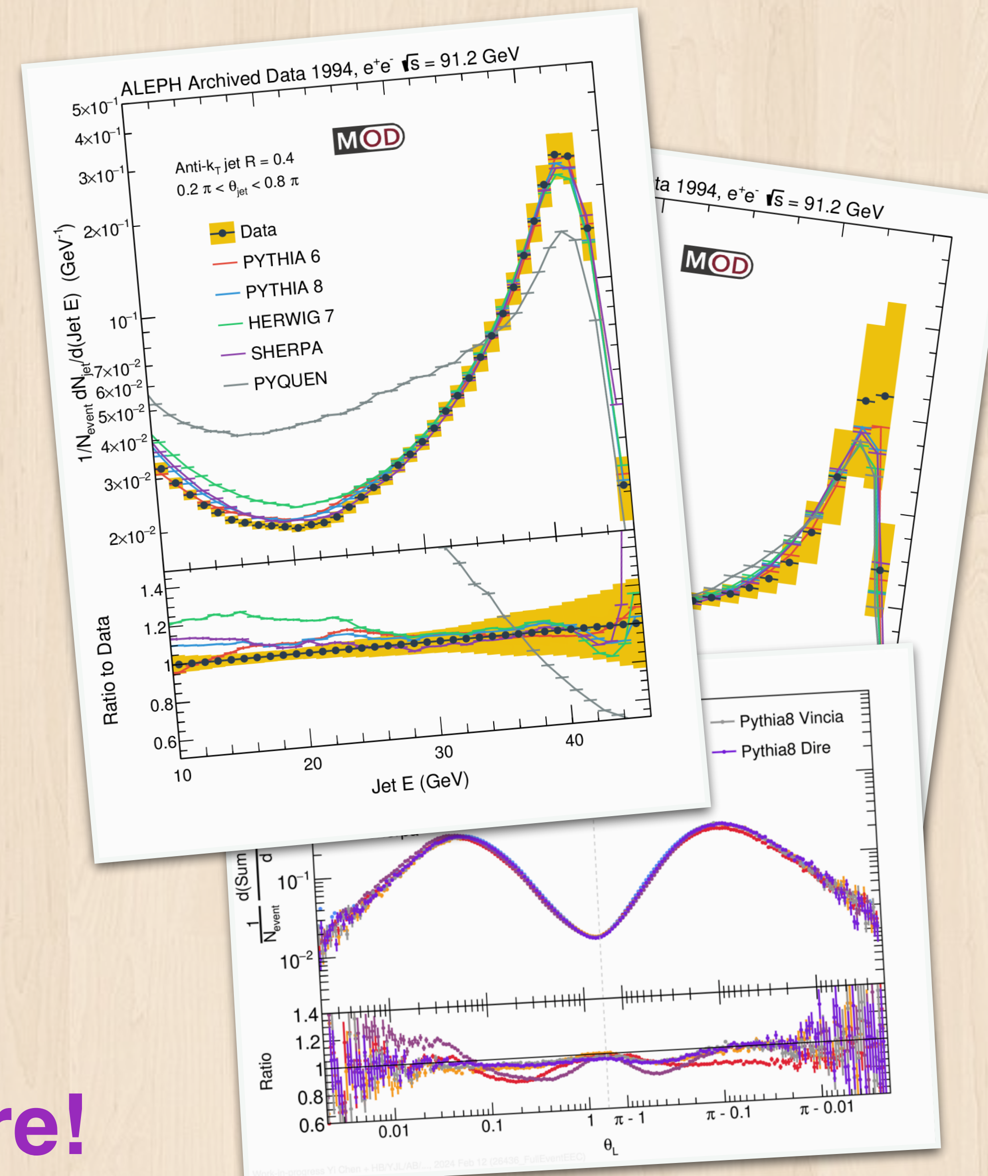
Putting into bigger picture



More places to explore! e.g. topologies in e^+e^-

What comes next for the effort?

- Two-particle correlation: more to explore with selections focusing on **different event topologies**
- Other efforts
  - First measurement of **jet spectra and substructure** [1]
  - Energy-energy correlator, etc…
- **Testing ground** for new algorithm developments (e.g. EIC)
- **Huge amount of things to explore!**



[1] JHEP 06 (2022) 008

Archived data

- Re-analysis of archived ALEPH data — multi-year process
  - A lot of effort in making sure we understand the data
  - Huge amount of help from ALEPH members
- Food for thought for ongoing experiments: preservation of knowledge, multiple MC samples, ability to rerun key software, low-level information, …
- **User test** while experiment ongoing would be best
- Allows **new ideas** popped up long after end of data-taking

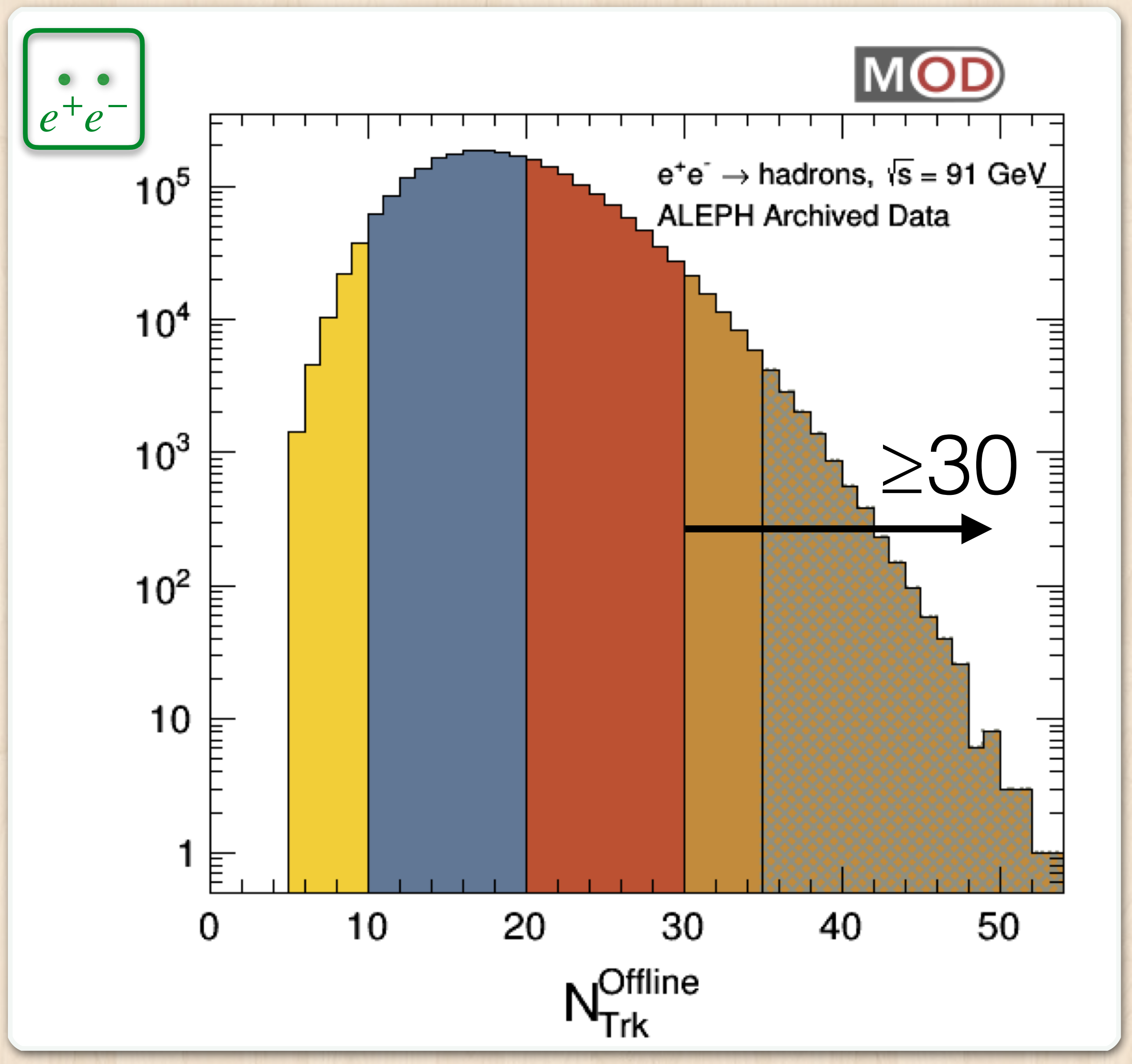
Thank you!

- 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 Larkoski, Liliana Apolinário, Ben Nachman, Camelia Mironov, Wei Li, Wit Busza, Yang-Ting Chien, Jamie Nagle, Maxime Guilbaud, Jing Wang for the useful discussions on the analysis

Backup Slides Ahead

Particle multiplicity (LEP1)

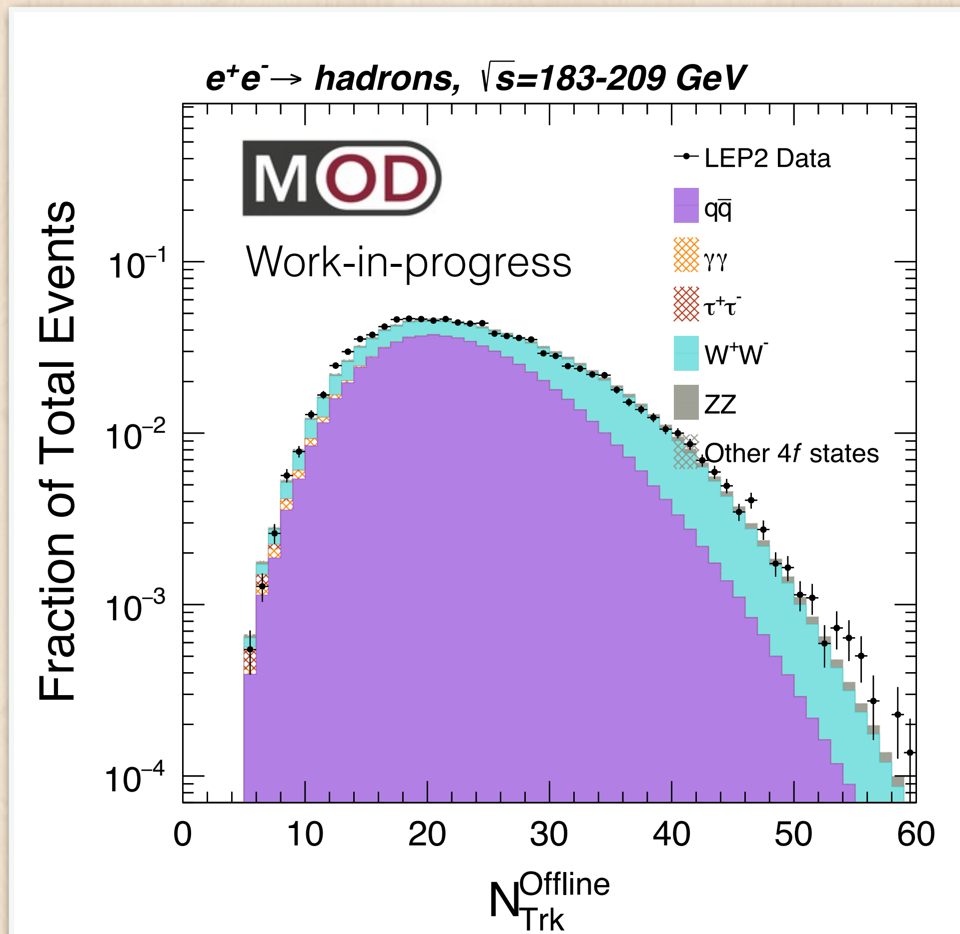


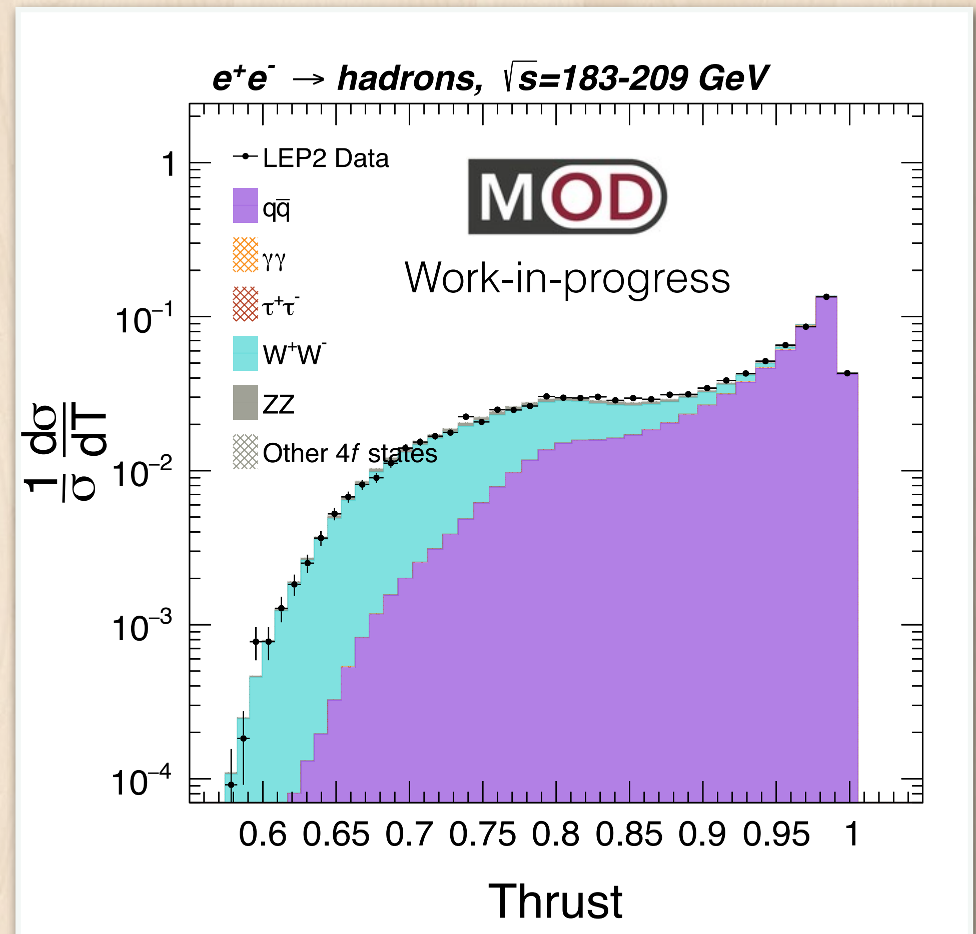
Inspired by the pp experience, look at correlations in bins of multiplicity

We focus on the high multiplicity events in this talk

PRL 123 (2019) 212002

Higher energy LEP2 data

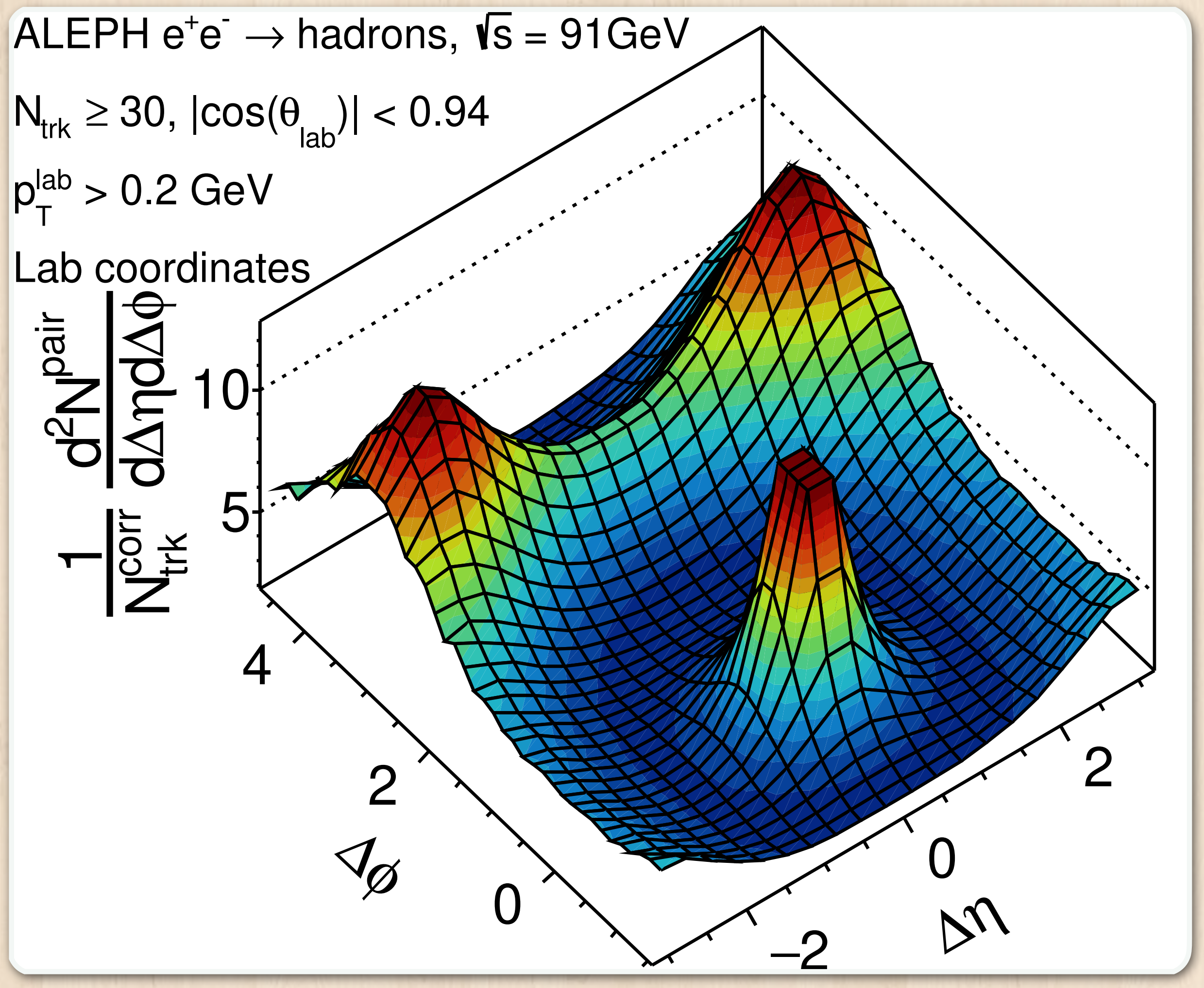


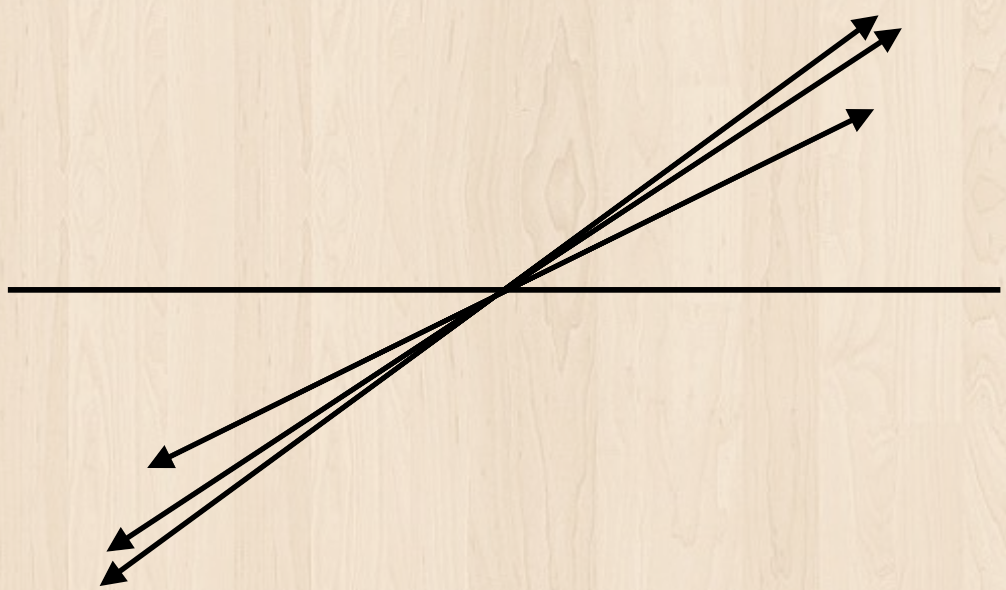


Integrated over higher energy datasets

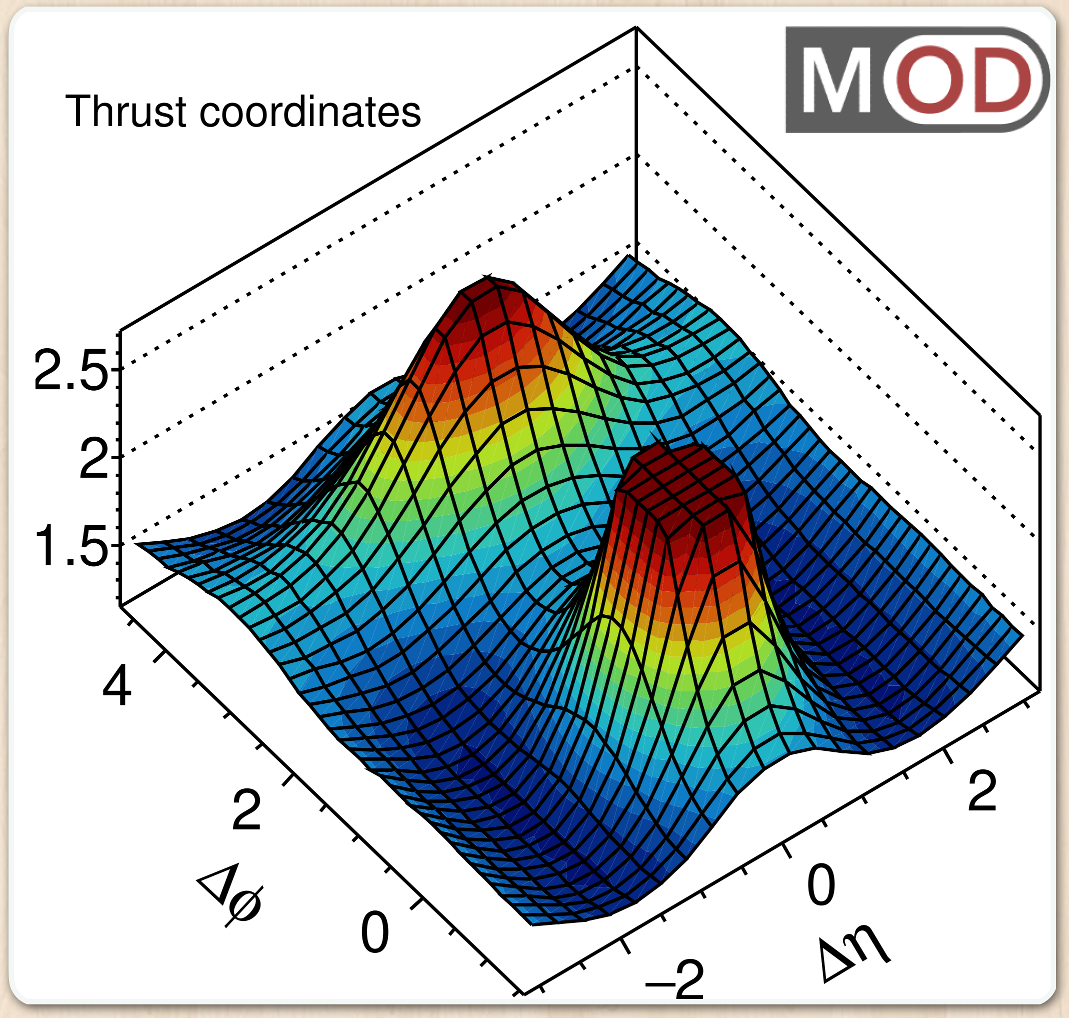
Generally decent data-MC agreement

Understanding beam axis

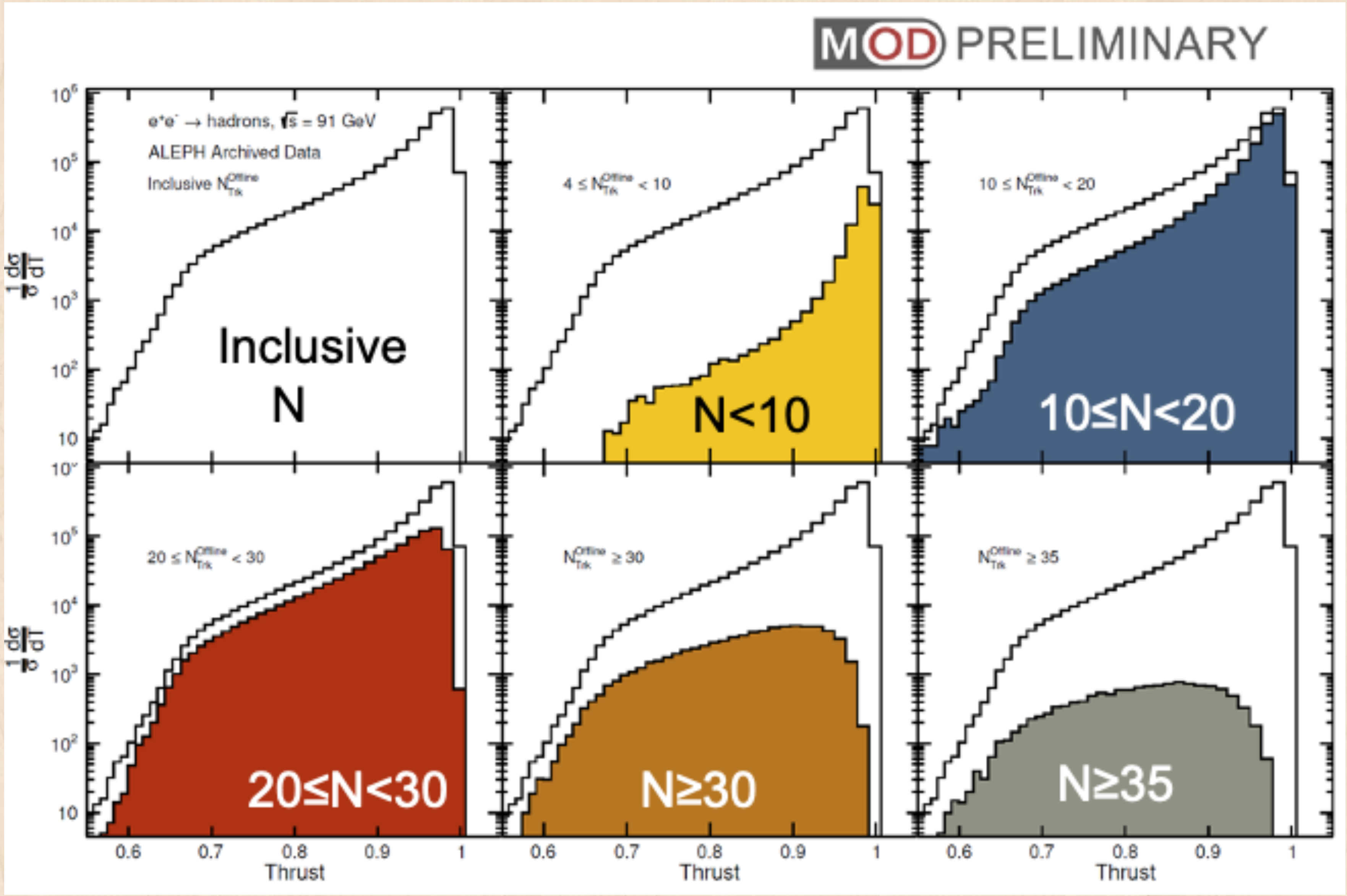




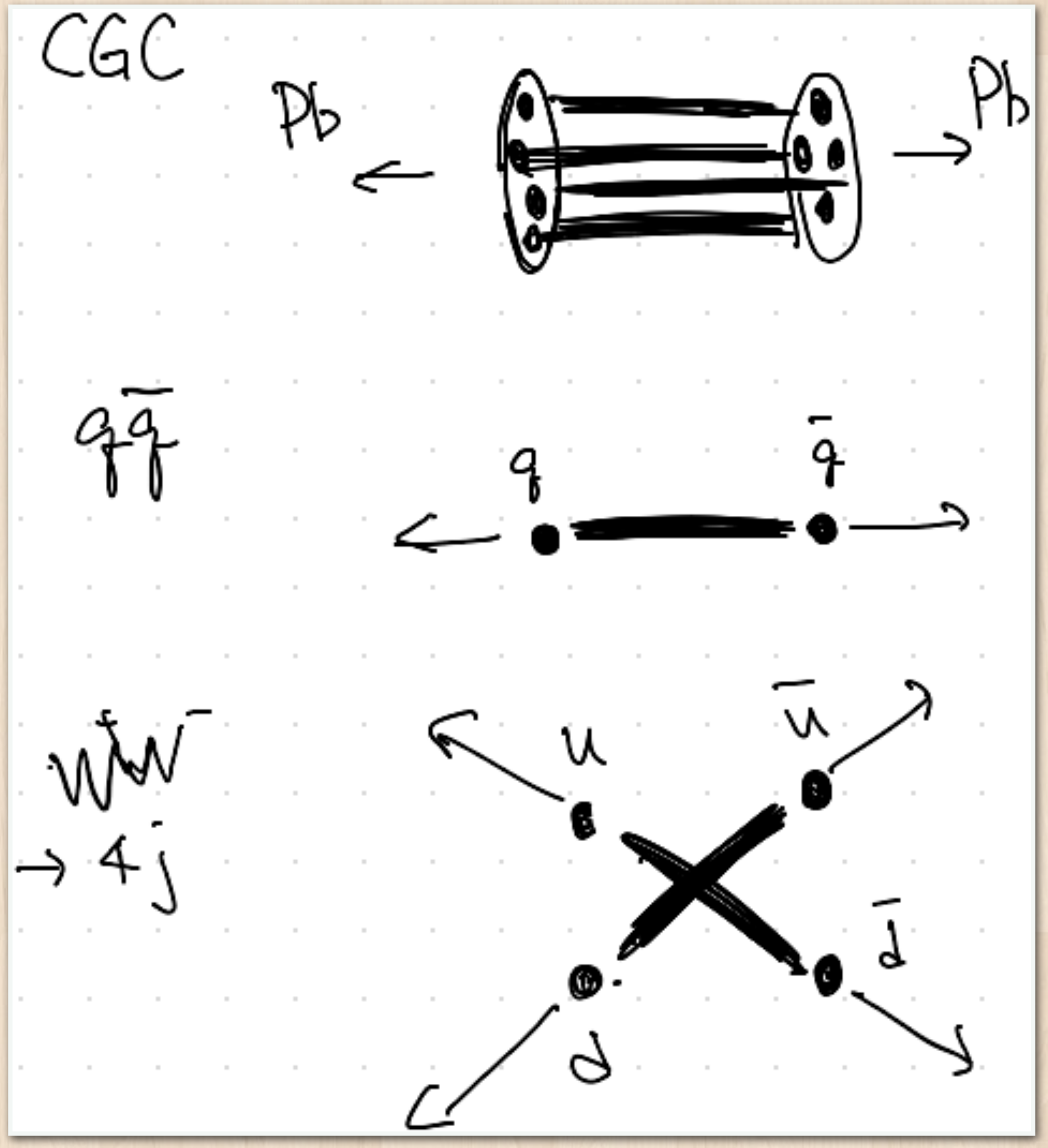
Thrust: what is what



Backup: thrust distributions



Color string configuration



Systematic uncertainties

- Event and track cuts
- Number of hits in the TPC detector
- Overall normalization
- Residual MC correction

Lab coordinate

$N_{Trk}^{Offline}$	TPC hits	Event energy and track cuts	$B(0,0)$	Residual MC correction
		LEP-I		
$[5, 10)$	0.7	0.6	0.11	10.3
$[10, 20)$	0.7	0.0	0.015	2.3
$[20, 30)$	0.7	0.0	0.013	0.2
$[30, \infty)$	0.7	0.0	0.027	1.2
$[35, \infty)$	0.7	0.0	0.057	4.4
		LEP-II		
$[10, 20)$	0.28	6.84	0.10	1.52
$[20, 30)$	1.99	2.97	0.06	0.61
$[30, 40)$	1.13	0.64	0.06	1.10
$[40, 50)$	0.45	0.10	0.09	1.50
$[50, \infty)$	2.52	0.21	0.17	1.74

Thrust coordinate

$N_{Trk}^{Offline}$	TPC hits	Event energy and track cuts	$B(0,0)$	Residual MC correction
		LEP-I		
$[5, 10)$	0.3	3.4	0.88	0.50
$[10, 20)$	0.3	0.0	0.09	0.21
$[20, 30)$	0.3	0.0	0.05	0.06
$[30, \infty)$	0.3	0.0	0.06	0.21
$[35, \infty)$	0.3	0.0	0.13	0.21
		LEP-II		
$[10, 20)$	1.09	0.39	0.44	1.17
$[20, 30)$	0.68	0.44	0.21	0.11
$[30, 40)$	0.65	0.05	0.12	0.10
$[40, 50)$	0.73	0.04	0.16	0.13
$[50, \infty)$	1.60	0.50	0.27	0.02