中国科学院高能物理研究所
Institute of High Energy Physics, Chinese Academy of Sciences

HIDDEN TREASURES
Investigation of the LEP Open Data

CHE Yuzhi (车逾之)
cheyuzhi@ihep.ac.cn

Nov. 17, 2025, IHEP EPD seminar

Why LEP data?

- **News**: In 2024, the DELPHI collaboration decided to open the data and original software stack to the physics community **for public access.**

- The LEP data is:

  - The largest e^+e^- collision data above Z pole on the world.

  - A chance to extract new insights from legacy data with today's advanced theory and ideas.

  - A unique real-data for developing and validating analysis methods for CEPC.

  - A rare opportunity to test modern AI/ML-based experimental techniques on real data rather than pure simulation.

- 🤔 You may **hesitate**: the data is almost 30 years old, can we get information from the data, and do the analysis using the old data?

Outline

1 **Overview**

About LEP; Detectors; History; Data Access Policy

2 **DELPHI Open Data**

Location; Metadata; Software …

3 **Analysis Example**

Data-MC consistency; potential topics …

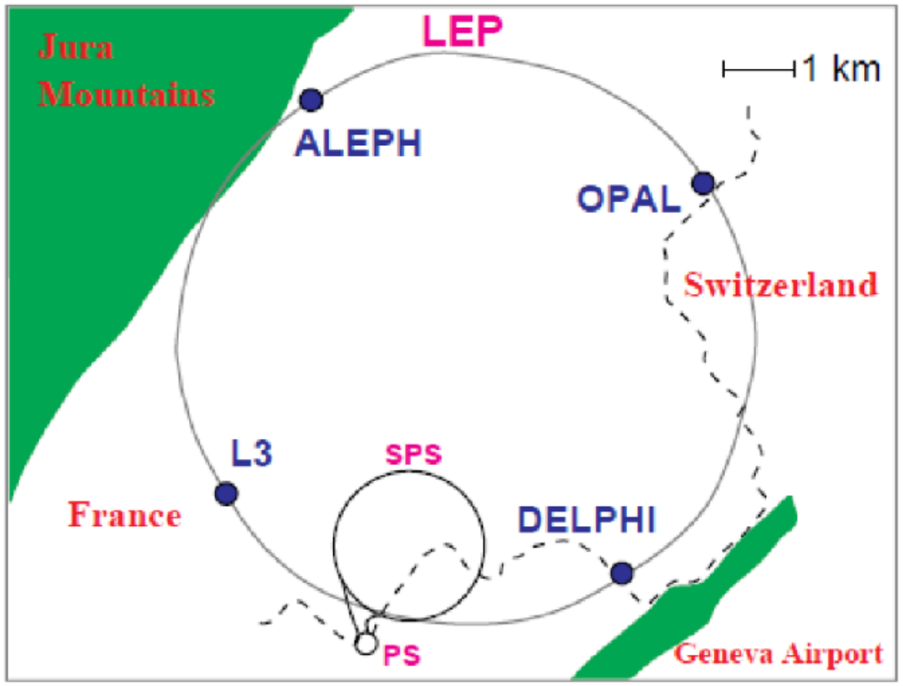
4 **other LEP experiments**

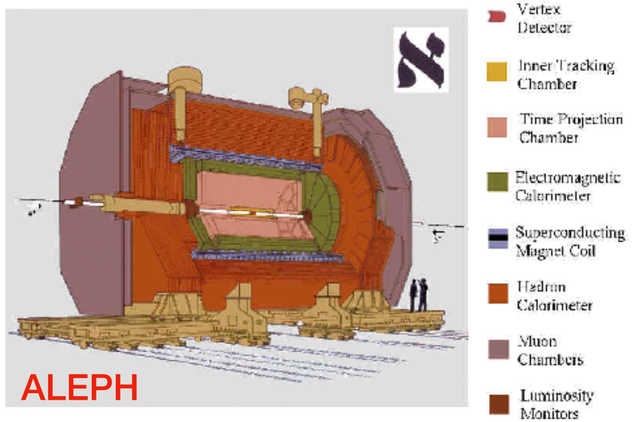
About DPHEP collaboration; old-data to EDM4hep

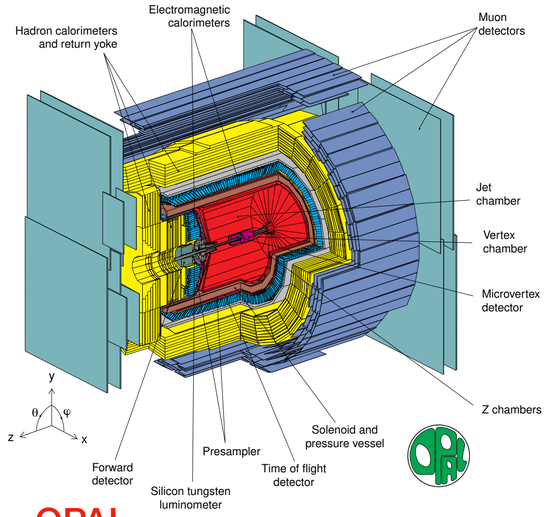
Overview: the Large Electron Positron Collider

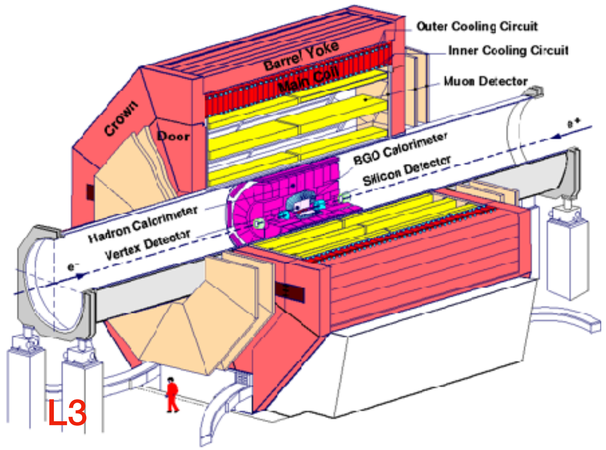
The Large Electron-Positron (LEP) collider is **the largest electron-positron accelerator** ever built.

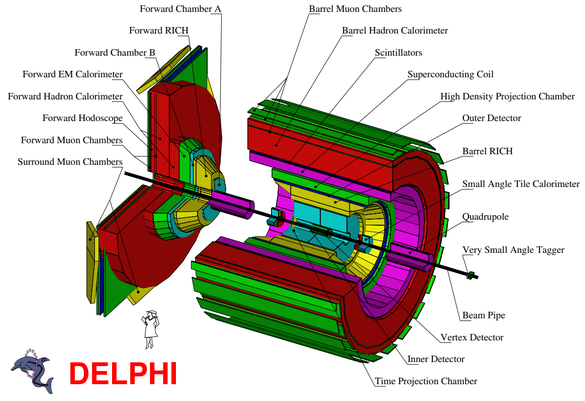
Four detectors on LEP: DELPHI, ALEPH, L3 and OPAL.











Detectors on the LEP

	Magnet Field	Tracking system	PID	ECAL	HCAL	Muon system note	Note
DELPHI	1.2 T (R ~ 2.6 m)	VTX, Inner drift, TPC, Outer drift, (~ 3%) (R: ~ 1.2 m)	dE/dx, dual-radiator RICH	HGC (~23%/$E^{0.5}$ ⊕ 1.1%)	Iron + streamer tubes		• **very new technologies** • a larger variety of techniques
ALEPH	1.5 T (R ~ 2.7 m)	VTX, Inner drift, TPC (2~3%) (R: ~ 1.80 m)	dE/dx	HGC (220 000 channel) (~18%/$E^{0.5}$ ⊕ 1%)	Iron + streamer tubes		• **Reasonably new technologies** • granularity more than energy resolution
L3	0.5 T (R ~ 5.93 m)	VTX, TEC, z-drift (~2.5%) (R: ~ 0.47 m)		-BGO (<1% above 10 GeV)	Iron + streamer tubes	Inside magnet volume	• measuring **leptons/ photons** with high resolution
OPAL	0.435 T (R ~ 2.18 m)	VTX, Large jet drift, z-drift (~3–4%) (R: ~ 1.8 m)	TOF	Lead-glass block (~6.3%/$E^{0.5}$ ⊕ 0.2%)	Iron + streamer tubes		• proven and reliable technologies • **ensure to be ready in time**

O. Callot, P. Charpentier, C. R. Physique 3 (2002) 1131–1141. and its references
B. Adeva, et al., L3 Collaboration, Nucl. Instrum. Methods Phys. Res. A 289 (1990) 35–102
Nucl.Instrum.Meth.A 294 (1990)

Overview: the Large Electron Positron Collider

- From 1990 to 2000, the LEP operation be divided into two parts:

  LEP I (1990~1995, 91.2 GeV) and **LEP II** (1996~2000, 161~205 GeV).

- The LEP terminated in 2000, while the collaboration activity continued till 2017.

CERNCourier2025SepOct

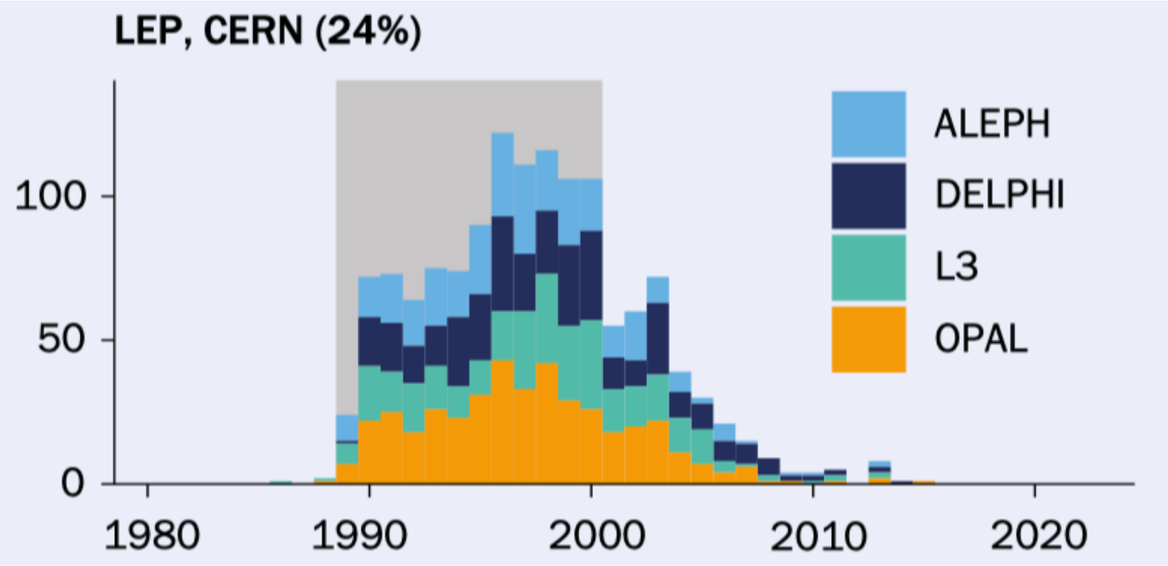


Table 2
Overview of LEP performance from 1989 to 2000. $\int \mathcal{L}dt$ is the luminosity integrated per experiment over each year and I_{tot} is the total beam current $2k_bI_b$. The luminosity $\mathcal{L}$ is given in units of $10^{30}\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$.

Year	$\int \mathcal{L}dt$ (pb^{-1})	E_b (GeV/c^2)	k_b	I_{tot} (mA)	$\mathcal{L}$
1989	1.74	45.6	4	2.6	4.3
1990	8.6	45.6	4	3.6	7
1991	18.9	45.6	4	3.7	10
1992	28.6	45.6	4/8	5.0	11.5
1993	40.0	45.6	8	5.5	19
1994	64.5	45.6	8	5.5	23.1
1995	46.1	45.6	8/12	8.4	34.1
1996	24.7	80.5 - 86	4	4.2	35.6
1997	73.4	90 - 92	4	5.2	47.0
1998	199.7	94.5	4	6.1	100
1999	253	98 - 101	4	6.2	100
2000	233.4	102 - 104	4	5.2	60

CERN-SL-2002-009-OP

Click on the figures for references or data sources.

Data preservation and policy around 2004

- Since 2004, the data have been archived to allow their use for physics analyses after the closure of the collaborations.
- Four collaborations formulated their long term data access policies.
- However, requests for data access should be approved by the collaborations.

Availability of OPAL data in future

The date for OPAL's transition to archive mode has not yet been fixed. To oversee use of the OPAL data after the transition, we will establish a "long-term OPAL editorial board" (long-term EB), to give technical assistance to people using the data, and to review publications. We expect any group wishing to use the OPAL data archive to seek the approval of the long-term EB. A group wishing to use the data should normally include former members of the collaboration. If this is not the case, but the long-term EB is convinced by the proposed analysis, then they will select one or two EB members to participate.

Authorized Users.

The use of archived Aleph data is authorized to former members of the Aleph Collaboration and their collaborators. The use of a subset of data for teaching and pedagogical purposes, under the guidance of former members of the Collaboration, is allowed.

Rules for long-term access to DELPHI archived data (01/12/03)

- The DELPHI data consist of the so-called Short DSTs and Extended Short DSTs (SDST and XSDST) stored in Castor.
- Requests for access, for both physics research and pedagogical purposes, will have to be made through the acting DELPHI spokesperson
- The acting DELPHI spokesperson has the right to veto access to DELPHI archived data if it is considered to be against the interest of DELPHI or science in general
- The acting DELPHI spokesperson will consult an "Archived data board" with respect to the involvement of former DELPHI members mentioned below
- The "Archived data board" will consist of the latest appointed members of Coordination, DEC and Research Line convenors
- Access to the archived DELPHI data will ONLY be authorised to non-former DELPHI persons in collaboration with one or more former "expert" DELPHI members
- Expert DELPHI members will be considered as those having extensively participated in the analysis of DELPHI data closely related to the newly proposed analysis
- A paper resulting from the use of archived DELPHI data will have to be co-signed by these collaborating former DELPHI members
- An "internal DELPHI referee" will have to be consulted before approval for publication; he/she will act as if a Journal referee
- This access policy will become applicable one year after the vast majority of the presently planned papers are published and will remain valid until further notice. The starting date will ultimately be defined by the Collaboration Board

Update of DELPHI's policy

- In 2024, the DELPHI collaboration decided to open the data and original software stack to the physics community **for public access** (see *DELPHI data preservation, re-use, and open access policy*) :
  - original DELPHI analysis framework
  - simulation & reconstruction
  - the event display
  - and documentation.
- Other experiments' datasets are archived on EOS awaiting similar release (see *4th DPHEP Collaboration Workshop*).

DELPHI data preservation, re-use, and open access policy

The DELPHI collaboration,
12 March 2024

run conditions are vanishing. Therefore, the collaboration strongly discourages attempts to redo high-precision analyses. Despite this, the data still holds potential for further exploration and discovery. Therefore, the data should be preserved and made accessible to the public for various purposes, such as education and citizen science. This document outlines the data preservation, re-use, and open access policy for this valuable data set.

Potential users of the data are encouraged to get in contact with DELPHI scientists to understand the limitations and possibilities. They are also encouraged to register their activity with the DELPHI data preservation board. DELPHI aims to implement the FAIR[1] principle for its data[2], but its full implementation will be subject to available person power.

Detailed and up to date information about the status and contacts will be made available via the DELPHI web page at http://delphiwww.cern.ch.

1 DELPHI Data

The data released by the DELPHI experiment consists of data sets collected during the operation of the experiments between 1989 and 2000. Additionally, various simulated data sets that simulate a large number of physics processes are also being released.

The main format of the data is called SHORT DST, which contains physics information in a compressed format. Reading of this data is supported by various software packages, as described in the documentation. In addition, the original RAW data is also available. Its main purpose is to study individual events with the event server and the display. Currently, the data is available via the CERN EOS storage system, with a backup on CERN's tape archive system, CTA. At a later point in time, the data should also be accessible via the CERN Open Data Portal.

Publications based on the DELPHI data shall give credits to the collaboration and clearly identify the data which has been used, e.g. by quoting an identifier, such as a DOI when available.

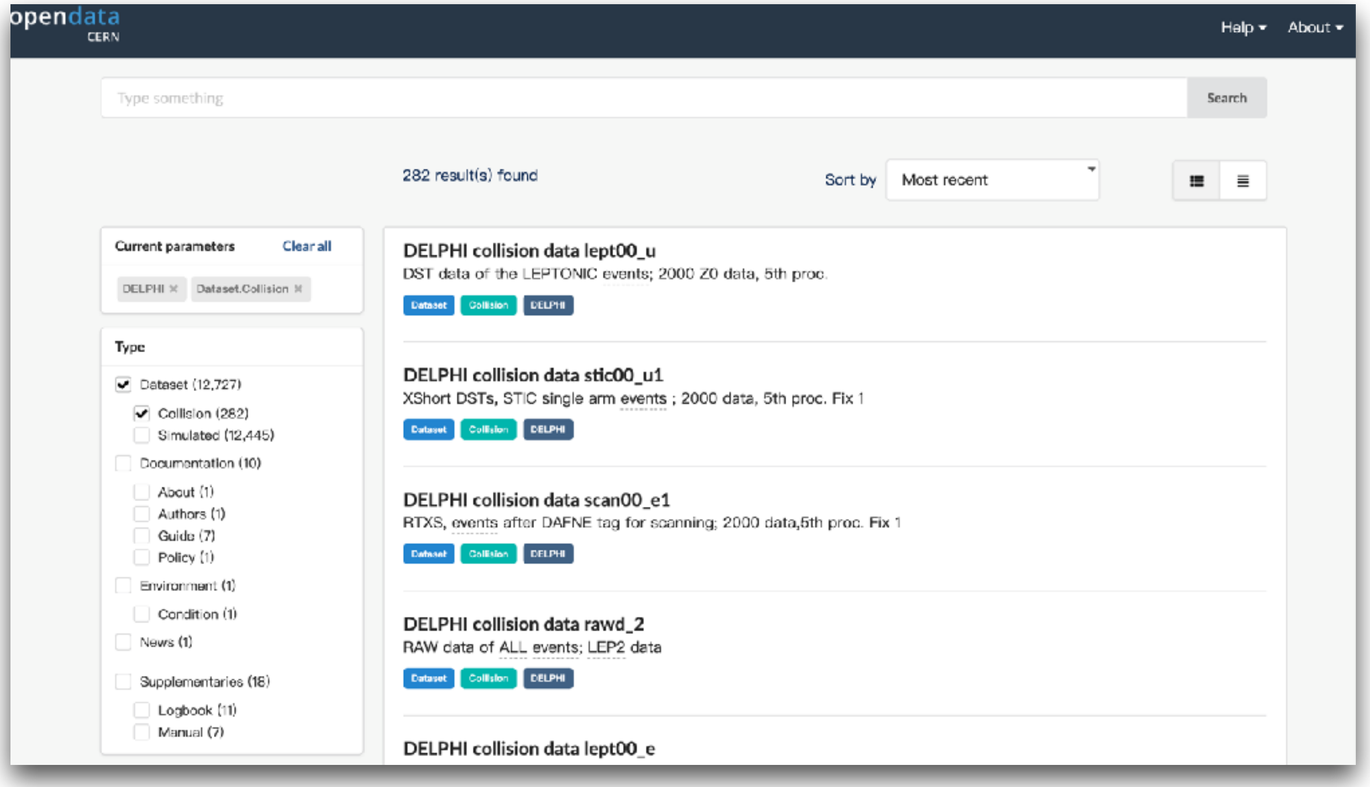
Following the convention of CERN, all metadata and data are released under the terms of the CC0 waiver.

DELPHI Open Data
Q & A

- **Where is the data stored?**
- **How is the data organized?**
  - **Format? Metadata?**
- **How to read the data?**
- **What is in the data files?**
- **What can we do using this data?**

Where is the data stored?

- The DELPHI data release can be accessed through the [CERN Open Data Portal].
- One can access them through EOS server: `/eos/opendata/delphi`




```
(base) →  delphi pwd
/eos/opendata/delphi
(base) →  delphi ls
collision-data  condition-data  documentation  logbooks  raw-data  simulated-data  upload
```

Note: Click on the figures for references.

Where is the data stored?

- The entire data and MC samples holds 36 TB.
- We can contact IHEP computing center to mount the EOS service on the IHEP cluster.

Simulation: 16 TB

Reconstructed data:
1991-2000, 14 TB

Raw Data: 1990-2000, 6 TB

```
(base) →  delphi pwd
/eos/opendata/delphi
(base) →  delphi ls
collision-data  condition-data  documentation  logbooks  raw-data  simulated-data  upload
```

How is the data organized?

The DELPHI data release covers both real and simulated datasets, including four kinds of formats (see `/cvmfs/delphi/docs/data/format.html`).

- [Full DST data]: contain information about reconstructed charged tracks and neutrals which may be used in physics analyses.
  - produced by [DELANA], the DELPHI Data Reconstruction and Analysis Program from the DELPHI RAW data.

How is the data organized?

Analyses should always use short or extended short DST format, because these generally contain fixes which were applied after the reconstruction itself.

- [Short DST data]: a more compact data set which allow faster analysis of the physics data used primarily by the hadronic teams in DELPHI.
  - produced by DSTANA, the DST Analysis and fixing package,
  - written by PHDST, the DELPHI package for DST productions, from the DELPHI DSTs.
- [Long DST data]: used primarily by the DELPHI leptonic teams.
  - produced using the same software as the Short DST's,
  - contain the Full DST information (with fixes and particle identifications), followed by the Short DST structure.

During LEP2

- [Extended Short DST data]: the main data format of Delphi collaboration for the LEP2 data analysis.

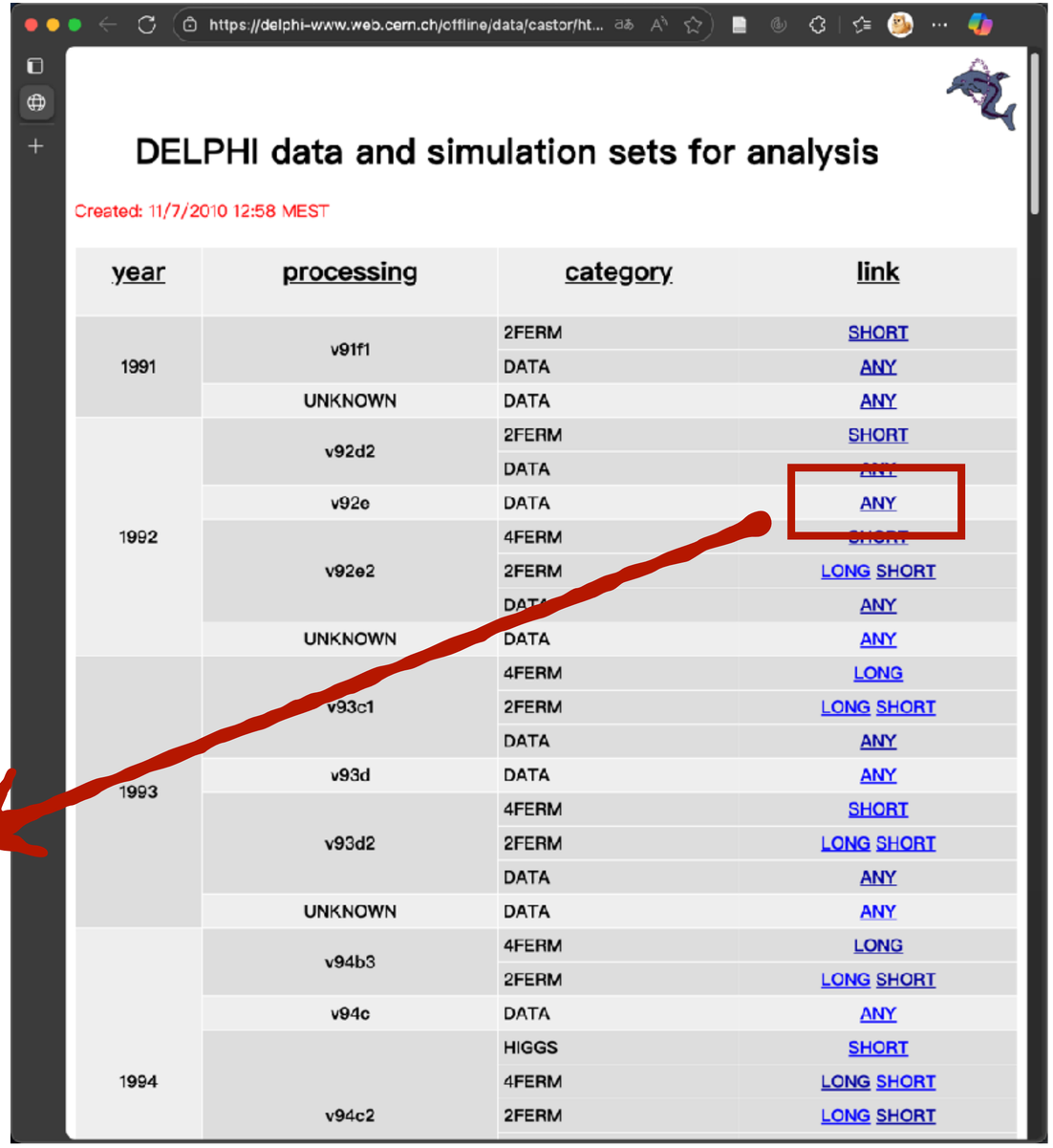
How is the data organized?

- The DELPHI data sets are organized in a dedicated `nickname` scheme, which indicates.
  - Format; Data (pre-selection) or MC (+channels)
  - Year
  - Version of the calibration software
- The nicknames and the corresponding content are listed in *DELPHI data and simulation sets for analysis*.

NOTE: This is the delana output only : --- NOT TO BE USED FOR GENERAL ANALYSIS ---

Nickname	Channel	Lab	ecm	files
dsto92_e	DSTO	CERN	91.25	344

Nickname	Channel	Lab	ecm	files
lept92_e	LEPT	CERN	91.25	64



year	processing	category	link
1991	v91f1	2FERM	SHORT
		DATA	ANY
	UNKNOWN	DATA	ANY
1992	v92d2	2FERM	SHORT
		DATA	ANY
	v92e	DATA	ANY
	v92e2	4FERM	SHORT
		2FERM	LONG SHORT
		DATA	ANY
	UNKNOWN	DATA	ANY
1993	v93c1	4FERM	LONG
		2FERM	LONG SHORT
		DATA	ANY
	v93d	DATA	ANY
	v93d2	4FERM	SHORT
		2FERM	LONG SHORT
		DATA	ANY
	UNKNOWN	DATA	ANY
1994	v94b3	4FERM	LONG
		2FERM	LONG SHORT
	v94c	DATA	ANY
	v94c2	HIGGS	SHORT
		4FERM	LONG SHORT
		2FERM	LONG SHORT

Note: Click on the figures for references.

How is the data organized?

The file lists of each nickname can be find using FATFIND.

Short description and brief metadata are also provided.

```
[root@Manjaro-B229 work]# fatfind alld00_u2
This is what I found:
        NICK : ALLD00_U2
        GNAME: /CASTOR/CERN.CH/DELPHI/TAPE/C001-126
        DESC : XShort DSTs, all events after DAFNE tag; 2000 data,5th proc. Fix 2
        COMM :    126 files, total size = 31.479 Gb, in 126 files

    1 /eos/opendata/delphi/collision-data/R07913/R07913.1.al
    2 /eos/opendata/delphi/collision-data/R07913/R07913.2.al
    3 /eos/opendata/delphi/collision-data/R07913/R07913.3.al
    4 /eos/opendata/delphi/collision-data/R07913/R07913.4.al
    5 /eos/opendata/delphi/collision-data/R07913/R07913.5.al
    6 /eos/opendata/delphi/collision-data/R07913/R07913.6.al
    7 /eos/opendata/delphi/collision-data/R07913/R07913.7.al
    8 /eos/opendata/delphi/collision-data/R07913/R07913.8.al
    9 /eos/opendata/delphi/collision-data/R07913/R07913.9.al
   10 /eos/opendata/delphi/collision-data/R07913/R07913.10.al
   11 /eos/opendata/delphi/collision-data/R07913/R07913.11.al
   12 /eos/opendata/delphi/collision-data/R07913/R07913.12.al
   13 /eos/opendata/delphi/collision-data/R07913/R07913.13.al
   14 /eos/opendata/delphi/collision-data/R07913/R07913.14.al
   15 /eos/opendata/delphi/collision-data/R07913/R07913.15.al
   16 /eos/opendata/delphi/collision-data/R07913/R07913.16.al
   17 /eos/opendata/delphi/collision-data/R07913/R07913.17.al
   18 /eos/opendata/delphi/collision-data/R07913/R07913.18.al
   19 /eos/opendata/delphi/collision-data/R07913/R07913.19.al
   20 /eos/opendata/delphi/collision-data/R07913/R07913.20.al
   21 /eos/opendata/delphi/collision-data/R07913/R07913.21.al
```

```
lxlogin003:~$ fatfind -N short
short91_f1
short92_d2
short92_e2
short93_c1
short93_d2
short94_c2
short95_d2
short96_e1
short96_f1
short_140_95b1
short_z0_93c1
short_z0_94b3
short_z0_95c2
short_z0_95d1
xshort00z_c1
xshort00z_e1
xshort00z_e2
xshort00z_s1
xshort00z_u1
xshort00z_u2
xshort96_f2
xshort96z_g1
xshort96z_g2
xshort97_e2
```

How is the data organized?

Even though physicists have access to all data at the TB-level, they often only need to work with GB-level data to complete many physics analyses.

year	center-of-mass energy	nickname	number of files	total size (GB)	Description
1991	91.2	short91_f1	45	4.003	Short DSTs,"OR" of the physics teams; 91 data 7th pr. (ANA_F) Fix 1
1992	91.2	short92_e2	172	16.911	Short DSTs,"OR" of the physics teams; 92 data 5th pr. (ANA_E) Fix 2
1993	91.2	short93_d2	181	17.233	Short DSTs,"OR" of the physics teams; 93 data 4th pr. (ANA_D) Fix 2
1994	91.2	short94_c2	429	38.339	Short DSTs,"OR" of the physics teams; 94 data 3rd pr. (ANA_C) Fix 2
1995	91.2	short95_d2	246	20.703	Short DSTs,"OR" of the physics teams; 95 data 4th pr. (ANA_D) Fix 2
1996	161	alld96_e161_g2	45	11.241	XShort DSTs, all events after DAFNE tag; 96 data, 7th proc. Fix 2
1996	172	alld96_e172_g2	28	6.802	XShort DSTs, all events after DAFNE tag; 96 data, 7th proc. Fix 2
1997	183	alld97_e183_g2	136	33.282	XShort DSTs, all events after DAFNE tag; 97 data, 7th proc. Fix 2
1998	189	alld98_e2	357	88.708	XShort DSTs, all events after DAFNE tag; 98 data, 5th proc. Fix 2
1999	192~202	alld99_e2	322	79.568	XShort DSTs, all events after DAFNE tag; 99 data, 5th proc. Fix 2
2000	204~206	alld00_u2	126	31.479	XShort DSTs, all events after DAFNE tag; 2000 data,5th proc. Fix 2

How to read the data?

- **Software:**
  - Obtain the DELPHI software through CVMFS
  - Mount /cvmfs/delphi.cern.ch on Linux
- **Data:**
  - EOS is mounted in the official container of DELPHI
- **Validated in my personal computer:**
  `Manjaro Linux 25.0.10 + container`

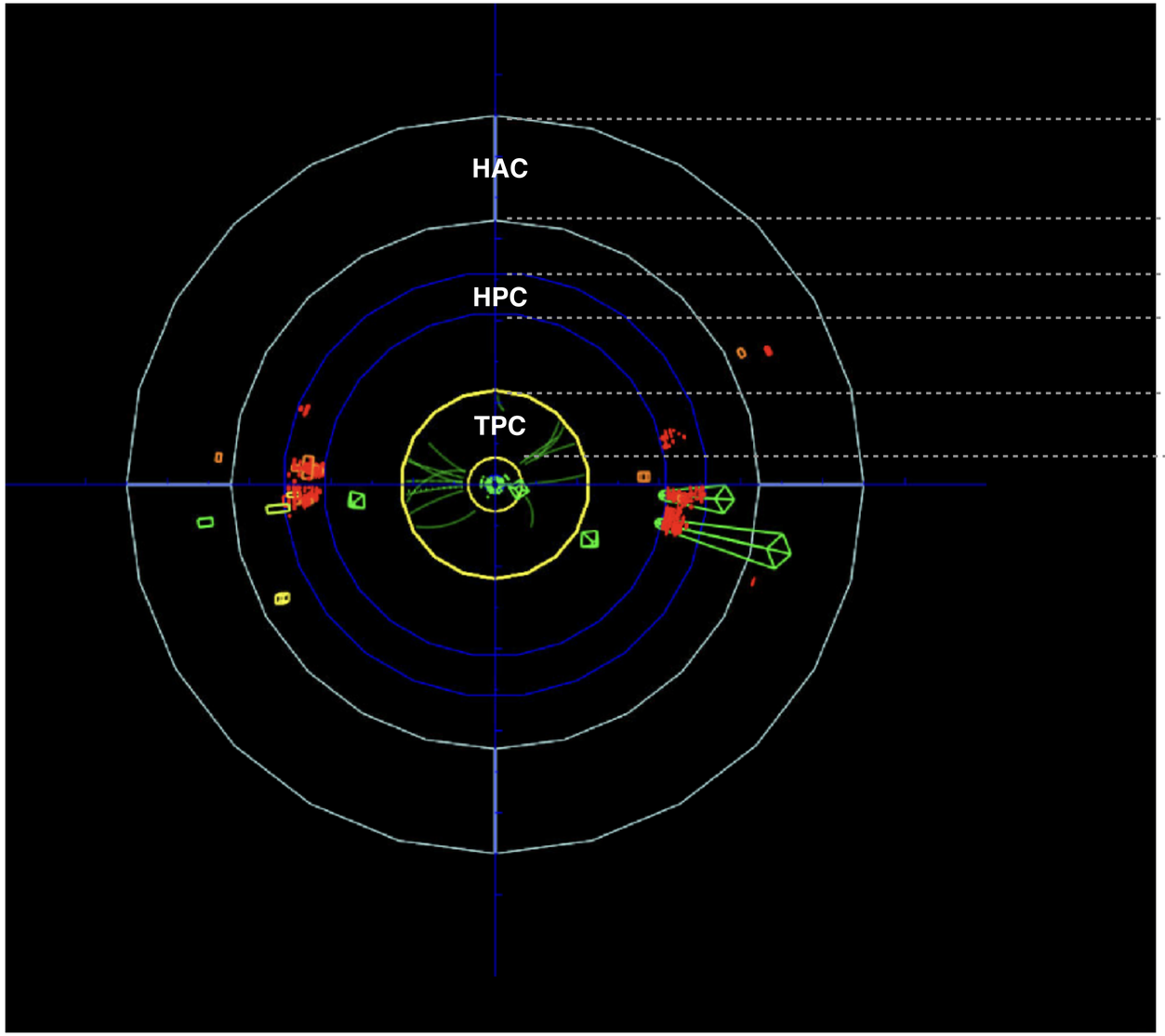
```
sudo podman run -d --name cvmfs --replace \
  --privileged \
  --network host \
  --device /dev/fuse \
  -e CVMFS_HTTP_PROXY="DIRECT" \
  -e CVMFS_CLIENT_PROFILE="single" \
  -e CVMFS_REPOSITORIES="cvmfs-
config.cern.ch,sft.cern.ch,unpacked.cern.ch,delphi.cern.ch,opendata.cern.ch" \
  -e CVMFS_QUOTA_LIMIT=8000 \
  -v /var/cvmfs:/var/lib/cvmfs \
  -v /cvmfs:/cvmfs:rshared \
  registry.cern.ch/cvmfs/service
```

```
sudo podman run --rm -it --name delphi-eos --privileged\
  --network host \
  -v /cvmfs:/cvmfs -v /eos:/eos:rshared \
  -v "$PWD":/work -w /work \
  --user root \
  gitlab-registry.cern.ch/delphi/deployment/delphi/al9_64_xrd \
  bash
```

```
source /cvmfs/delphi.cern.ch/releases/almalinux-9-x86_64_xrd/latest/profiles/delphi.sh
source /cvmfs/sft.cern.ch/lcg/views/LCG_107/x86_64-el9-gcc11-opt/setup.sh
```

Technique issues can be resolved by chatGPT 😝

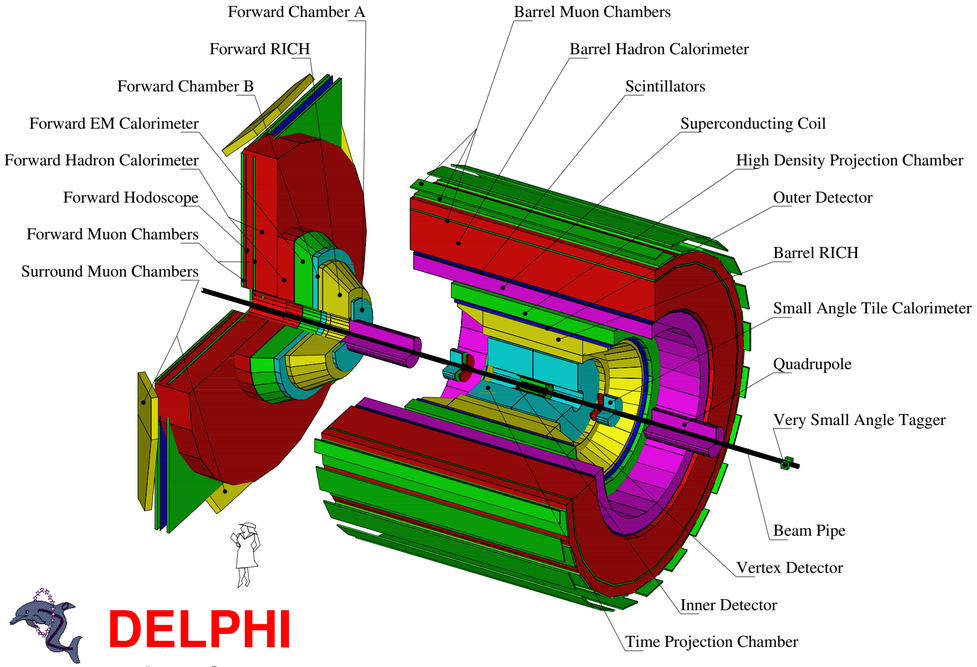
Event display: DELGRA



HAC radius: 320-479 cm

HPC radius: 208 ~ 260 cm

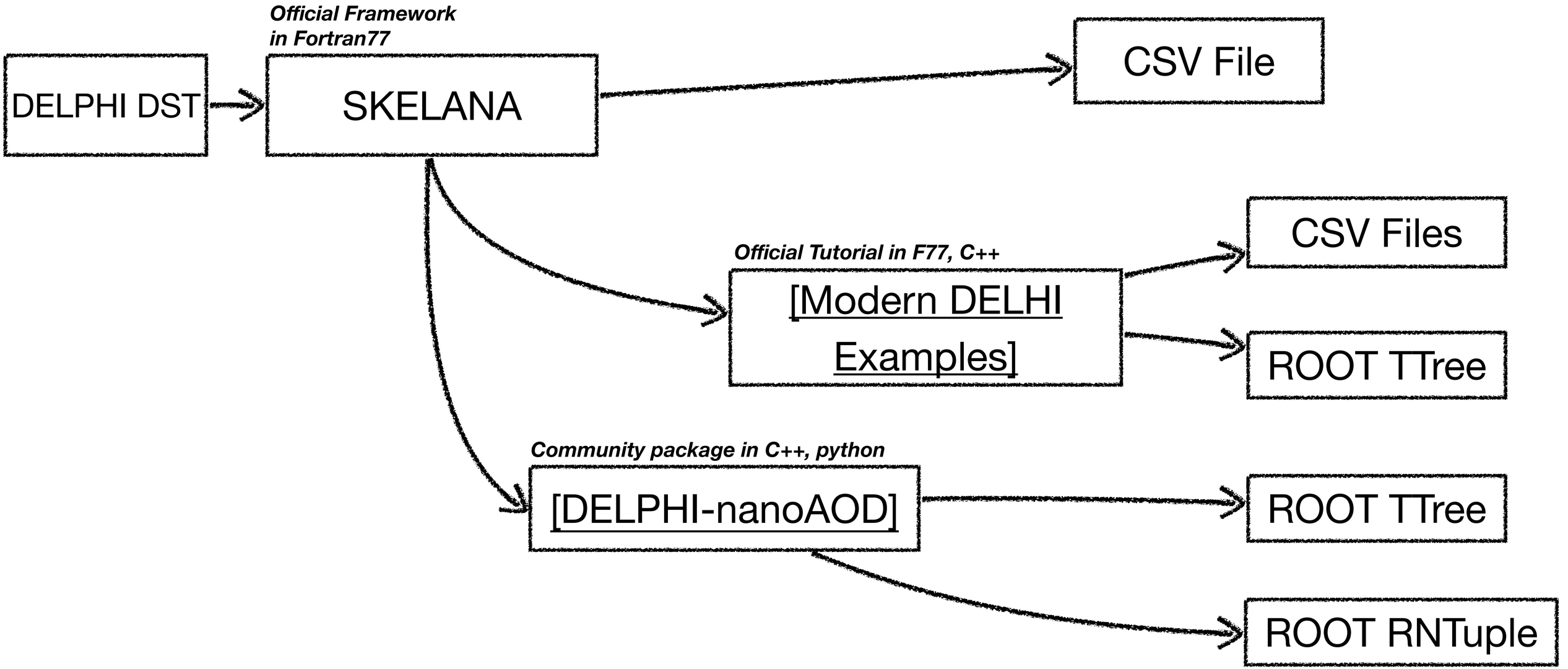
TPC outer radius: 35~111 cm



DELPHI

See more display in https://delphi-www.web.cern.ch/delphi-www/delfigs/events/events.html

How to read the data?



What is in the data files?

Several converted root files: `/cefs/higgs/cheyuzhi/data/delphi/root-files`

The number of the hadronic events agrees with the official report.

Number of Events										
	$Z \to q\bar{q}$					$Z \to \ell^+\ell^-$				
Year	A	D	L	O	LEP	A	D	L	O	LEP
1990/91	433	357	416	454	1660	53	36	39	58	186
1992	633	697	678	733	2741	77	70	59	88	294
1993	630	682	646	649	2607	78	75	64	79	296
1994	1640	1310	1359	1601	5910	202	137	127	191	657
1995	735	659	526	659	2579	90	66	54	81	291
Total	4071	3705	3625	4096	15497	500	384	343	497	1724

Table 1.2: The $q\bar{q}$ and $\ell^+\ell^-$ event statistics, in units of 10^3, used for Z analyses by the experiments ALEPH (A), DELPHI (D), L3 (L) and OPAL (O).

```
1 df.Filter("Event_hadronTagT4", "hadronTagT4").Report().Print()
```

[32] ✓ 0.1s

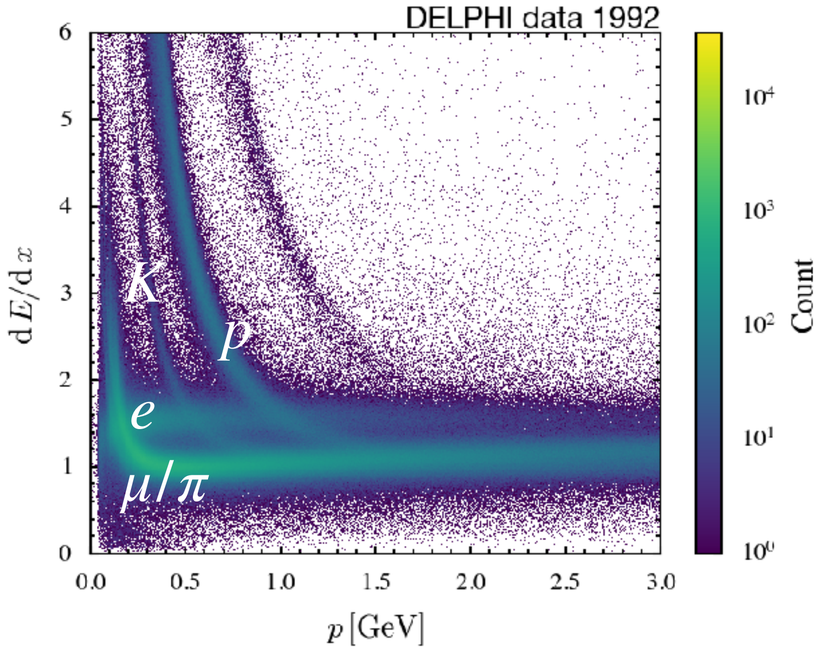
```
hadronTagT4: pass=709243    all=2482495    -- eff=28.57 % cumulative eff=28.57 %
```

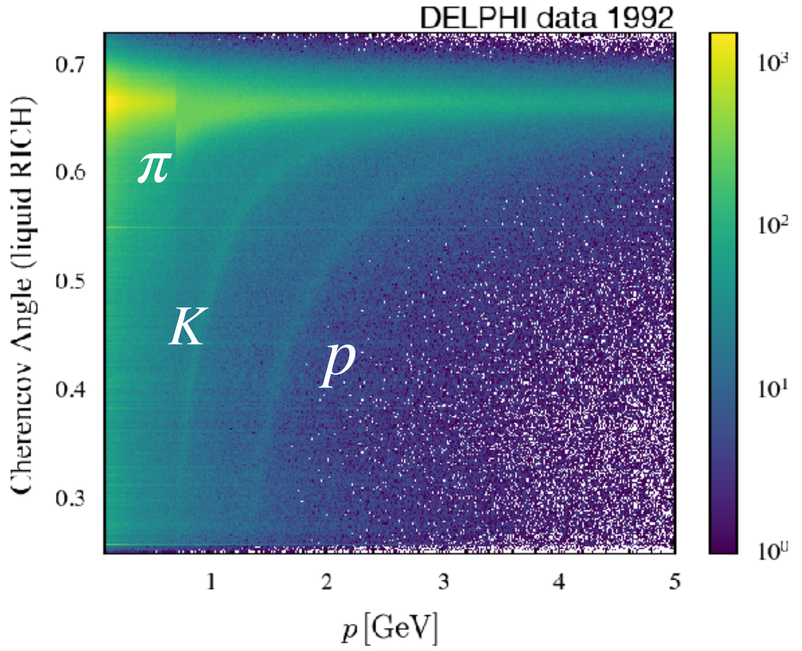
What is in the data files?

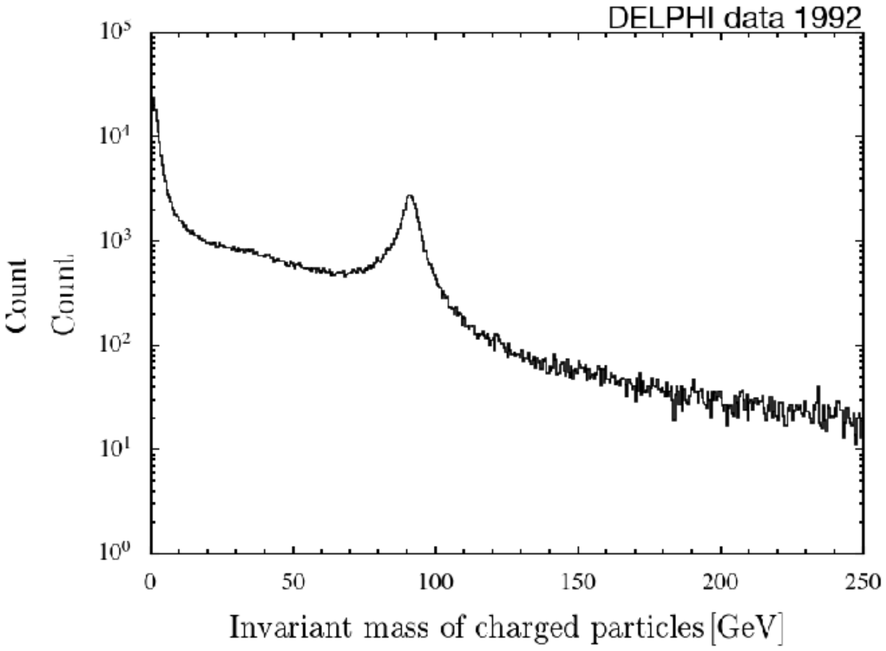
- Several converted root files: `/cefs/higgs/cheyuzhi/data/delphi/root-files`

- Information in much detail can be extracted from shortDST.

	nPart	len_Part	len_charged	len_Haid	len_Rich	len_Trac	len_Muid	len_Vtx
0	2	2	2	2	2	2	0	1
1	4	4	4	4	4	4	0	2
2	6	6	6	6	6	6	0	1
3	3	3	1	1	1	1	0	1
4	3	3	0	0	0	0	0	1
...	...	...	...	...	...	...	...	...
2482490	36	36	20	20	20	20	0	5
2482491	2	2	0	0	0	0	0	1
2482492	45	45	34	34	34	34	0	7
2482493	44	44	30	30	30	30	2	2
2482494	32	32	24	24	24	24	0	3







Analysis Example: What can we do with DELPHI data?

- An analysis example [arXiv:2510.18762] showing that:
  - There have been teams working on analyzing the DELPHI data.
  - The present resource is enough to support complete analysis procedures.
  - There are interesting topics.

About the Authors Team

The Electron–Positron Alliance

"We give old data new lives."

Members

(Please add missing people if I missed anyone)

Yen–Jie Lee

Massachusetts Institute of Technology
Fearless leader

Anthony Badea

University of Chicago

Austin Baty

University of Illinois Chicago

Christopher McGinn

Massachusetts Institute of Technology

Gian Michele Innocenti

Massachusetts Institute of Technology

Hannah Bossi

Massachusetts Institute of Technology

Jeetendra Gupta

University of Illinois Chicago

Jesse Thaler

Massachusetts Institute of Technology

Jingyu Zhang

Vanderbilt University

Note: Click on the figures for references.

About the Authors Team

The Electron–Positron Alliance

"We give old data new lives."

Energy Correlators from Partons to Hadrons: Unveiling the Dynamics of the Strong Interactions with Archival ALEPH Data

Authors: Hannah Bossi, Yi Chen, Yu–Chen Chen, Max Jaarsma, Yibei Li, Jingyu Zhang, Ian Moult, Wouter Waalewijn, Hua Xing Zhu, Anthony Badea, Austin Baty, Christopher McGinn, Gian Michele Innocenti, Marcello Maggi, Yen–Jie Lee

MITHIG–MOD–24–001

arXiv: 2511.00149

Unbinned measurement of thrust in e^+e^- collisions at $\sqrt{s}$ = 91.2 GeV with ALEPH archived data

Authors: Anthony Badea, Austin Baty, Hannah Bossi, Yu–Chen Chen, Yi Chen, Jingyu Zhang, Gian Michele Innocenti, Marcello Maggi, Chris McGinn, Michael Peters, Tzu–An Sheng, Vinicius Mikuni, Matthew Avaylon, Patrick Komiske, Eric Metodiev, Jesse Thaler, Benjamin Nachman, Yen–Jie Lee

MITHIG–MOD–25–001. Submitted to PRL

arXiv: 2510.22038

Analysis note: measurement of thrust and track energy–energy correlator in e^+e^- collisions at 91.2 GeV with DELPHI open data

Authors: Jingyu Zhang, Tzu–An Sheng, Yu–Chen Chen, Hannah Bossi, Anthony Badea, Austin Baty, Chris McGinn, Yen–Jie Lee, Yi Chen

MITHIG–MOD–NOTE–25–002

arXiv: 2510.18762

Analysis note: measurement of thrust in e^+e^- collisions at $\sqrt{s}$ = 91 GeV with archived ALEPH data

Authors: Anthony Badea, Austin Baty, Hannah Bossi, Yu–Chen Chen, Yi Chen, Jingyu Zhang, Gian Michele Innocenti, Marcello Maggi, Chris McGinn, Michael Peters, Tzu–An Sheng, Vinicius Mikuni, Matthew Avaylon, Patrick Komiske, Eric Metodiev, Jesse Thaler, Benjamin Nachman, Yen–Jie Lee

MITHIG–MOD–NOTE–25–001

arXiv: 2507.14349

Analysis note: measurement of energy–energy correlator in e^+e^- collisions at 91 GeV with archived ALEPH data

Authors: Hannah Bossi, Yu–Chen Chen, Yi Chen, Jingyu Zhang, Gian Michele Innocenti, Anthony Badea, Austin Baty, Marcello Maggi, Chris McGinn, Yen–Jie Lee

MITHIG–MOD–NOTE–24–001

arXiv: 2505.11628

Long–range near–side correlation in e^+e^- Collisions at 183–209 GeV with ALEPH Archived Data

Authors: Yu–Chen Chen, Yi Chen, Anthony Badea, Austin Baty, Gian Michele Innocenti, Marcello Maggi, Christopher McGinn, Michael Peters, Tzu–An Sheng, Jesse Thaler, Yen–Jie Lee

PLB 856 (2024) 138957, MITHIG–MOD–23–001

arXiv: 2312.05084

Analysis note: two–particle correlation in e^+e^- collisions at 91–209 GeV with archived ALEPH data

Authors: Yu–Chen Chen, Yen–Jie Lee, Yi Chen, Paoti Chang, Chris McGinn, Tzu–An Sheng, Gian Michele Innocenti, Marcello Maggi

MITHIG–MOD–NOTE–23–001

arXiv: 2309.09874

Jet energy spectrum and substructure in e^+e^- collisions at 91.2 GeV with ALEPH Archived Data

Authors: Yi Chen, Anthony Badea, Austin Baty, Paoti Chang, Yang–Ting Chien, Gian Michele Innocenti, Marcello Maggi, Christopher McGinn, Dennis V. Perepelitsa, Michael Peters, Tzu–An Sheng, Jesse Thaler, Yen–Jie Lee

JHEP 06 (2022) 008, MITHIG–MOD–21–001

Website: https://doi.org/10.1007/JHEP06(2022)008
arXiv: 2111.09914

Note: Click on the figures for references.

Analysis example: track–based EEC on DELPHI

- The general form of the definition of the energy-energy corrector (EEC) in previous results:

$$\mathrm{EEC}\left(\theta_{\mathrm{L}}\right)=\sum_{\mathrm{i>j}}^{\mathrm{n}}\int d\sigma\frac{E_{\mathrm{i}}E_{\mathrm{j}}}{E^{2}}\delta\left(\theta_{\mathrm{L}}-\theta_{\mathrm{ij}}\right)$$

  - E: center-of-mass energy
  - $E_{i,j}$: energy of particle i or j
  - $\theta_{i,j}$: open angle between particle i and j

- **In 2000:**
  - Charged & neutral particles are included. Therefore the precision limited by the calorimeter.

- **Now:**
  - EEC for only charged particles are well understood in QCD theory.
  - Resolution for EEC can be significantly improved by only considering charged tracks.

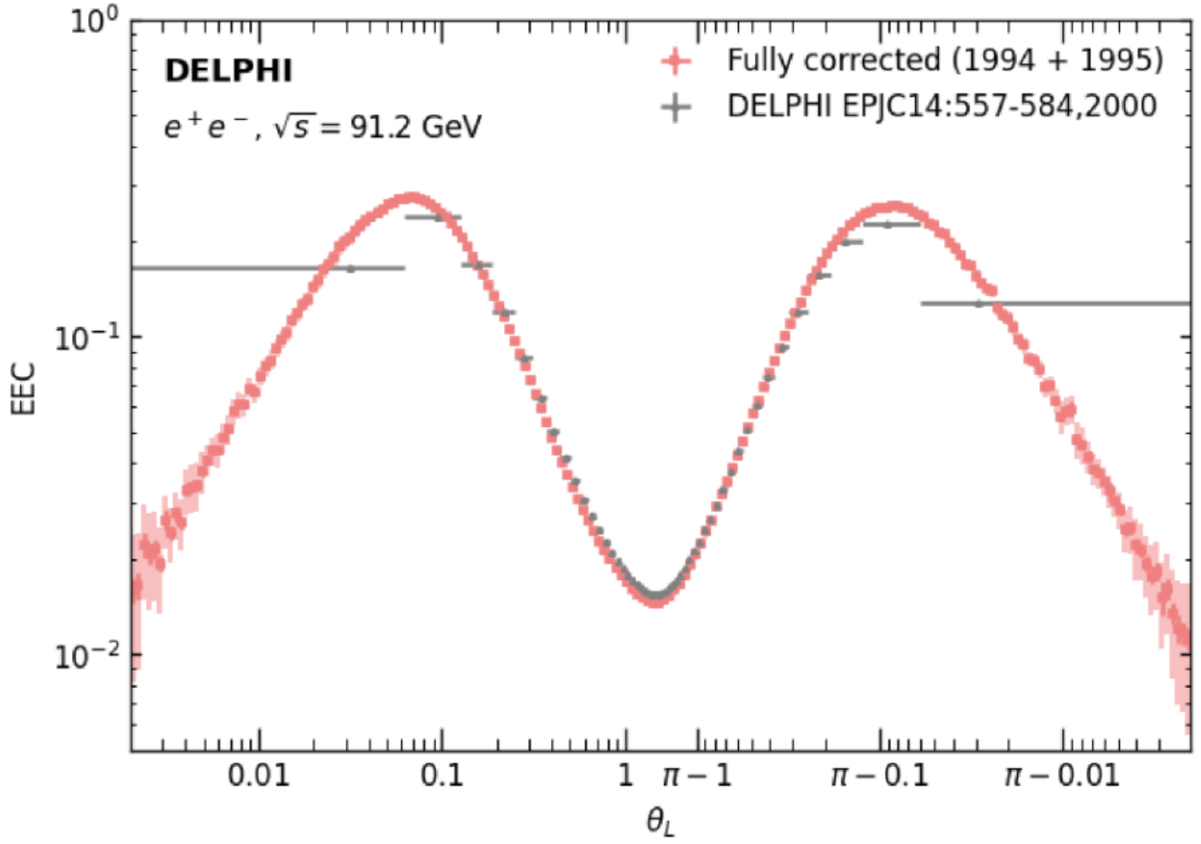


Figure 55: A comparison of the fully-corrected, track-based EEC distribution from this work (light coral) with the previous DELPHI measurement from Ref. [15] (gray). For the measurement from this work, the statistical uncertainties are shown as vertical lines, and the total systematic uncertainties are shown as boxes. For the previous measurement, the error bars represent the statistical and systematic uncertainties added in quadrature. The definition of the EEC in this work differs from that of the legacy DELPHI analysis. Specifically, the legacy measurement normalizes the energy of charged track pairs by the visible event energy, while the current work uses the beam energy.

Data, MC samples and selection

- The hadronization models in these official samples are tuned.
- The final-state hadrons are processed through the full DELPHI detector simulation using DELSIM.

Type	Sample	DOI
Data 1994	short94_c2	10.7483/OPENDATA.DELPHI.0XNE.G96F
Data 1995	short95_d2	10.7483/OPENDATA.DELPHI.K4LR.4PQ4
PYTHIA 5 1994	sh_kk2f4146qqpy_e91.25_c94_2l_c2	10.7483/OPENDATA.DELPHI.P4BN.GFID
PYTHIA 5 1995	sh_kk2f4146qqpy_e91.25_c95_1l_d2	10.7483/OPENDATA.DELPHI.0JDR.XYNG
ARIADNE 1994	sh_kk2f4146qqardcy_e91.25_r94_2l_c2	10.7483/OPENDATA.DELPHI.0VKG.S2RX

Table 1: Summary table for data and MC samples used in this note.

Charged particles	
Acceptance	$20° \leq \theta \leq 160°$
	$0.4\ \text{GeV} \leq p_\text{T} \leq 100\ \text{GeV}$
High quality tracks	measured track length ≥ 30 cm
	$\Delta p/p \leq 1.0$
Impact parameter	$d_0 \leq 4$ cm, $z_0 \leq 10$ cm
Neutral particles	
Acceptance	$20° \leq \theta \leq 160°$
	$0.5\ \text{GeV} \leq E$
Event selection	
Hadronic events	$30° \leq \theta_\text{thrust} \leq 150°$
	at least 7 good tracks
	$E_\text{tot} \geq 0.5 E_\text{cm}$

Table 2: Summary table for particle and event selections.

MC–data consistency

The DELPHI simulation reproduces the detector and describes the data well.

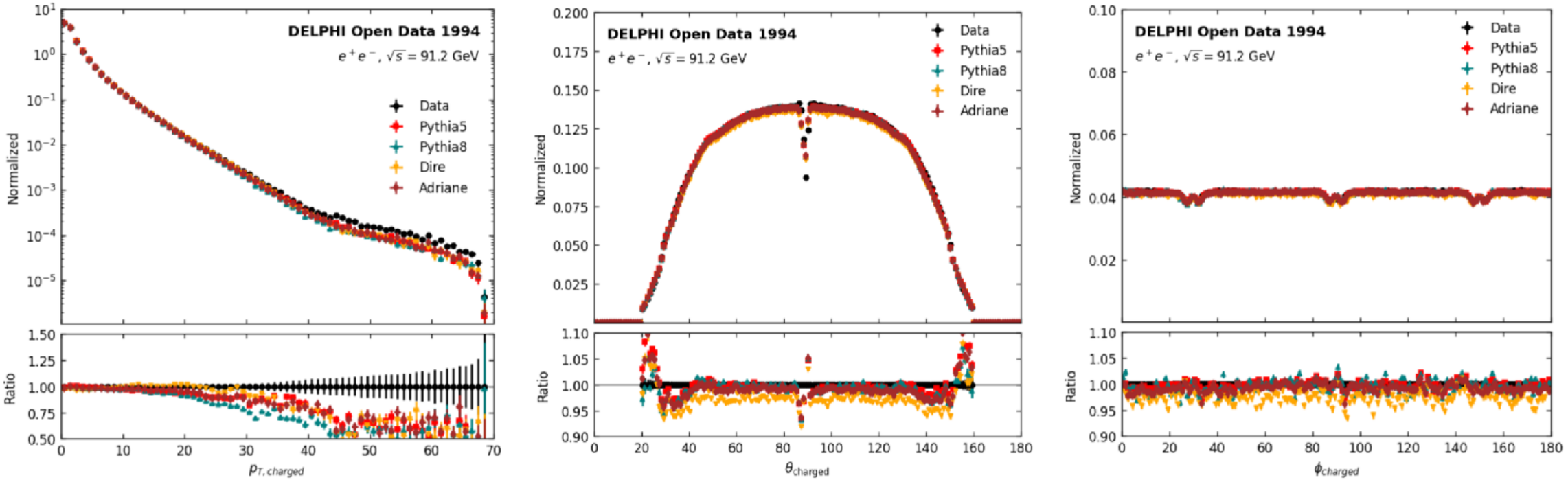


Figure 4: Comparisons of track p_{T}, θ, and ϕ in 1994 data (black), reconstructed PYTHIA 5.7/JETSET 7.4 (red), ARIADNE (green), PYTHIA 8.3 (orange), and PYTHIA 8.3 Dire (brown) samples. Note that the corresponding number of selected events of each sample normalizes the distributions.

MC–data consistency

The DELPHI simulation reproduces the detector and describes the data well.

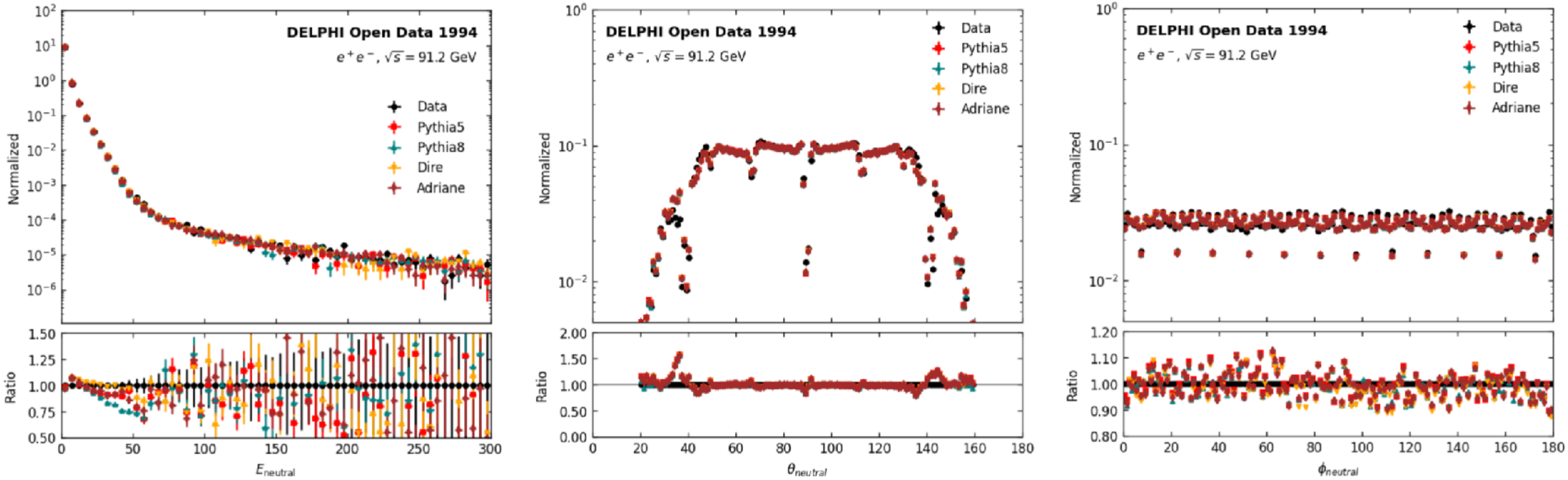


Figure 6: Comparisons of neutral particle E, θ, and ϕ in 1994 data (black), reconstructed PYTHIA 5.7/JETSET 7.4 (red), ARIADNE (green), PYTHIA 8.3 (orange), and PYTHIA 8.3 Dire (brown) samples. Note that the corresponding number of selected events of each sample normalizes the distributions.

Results & Significance

- The **first comprehensive physics study:** entire analysis procedure can be conducted.
- The **track-based EEC** measurement:
  - **Theory improvement**: well-understood of the tack-based EEC in QCD.
  - **Significantly improved resolution** over previous DELPHI results.
- **Future direction:**
  - **More observables:** the EEC's energy evolution, higher-point correlators, and flavor-tagged event shapes.
  - **New statistic tools**: unbinned measurements of event shapes.

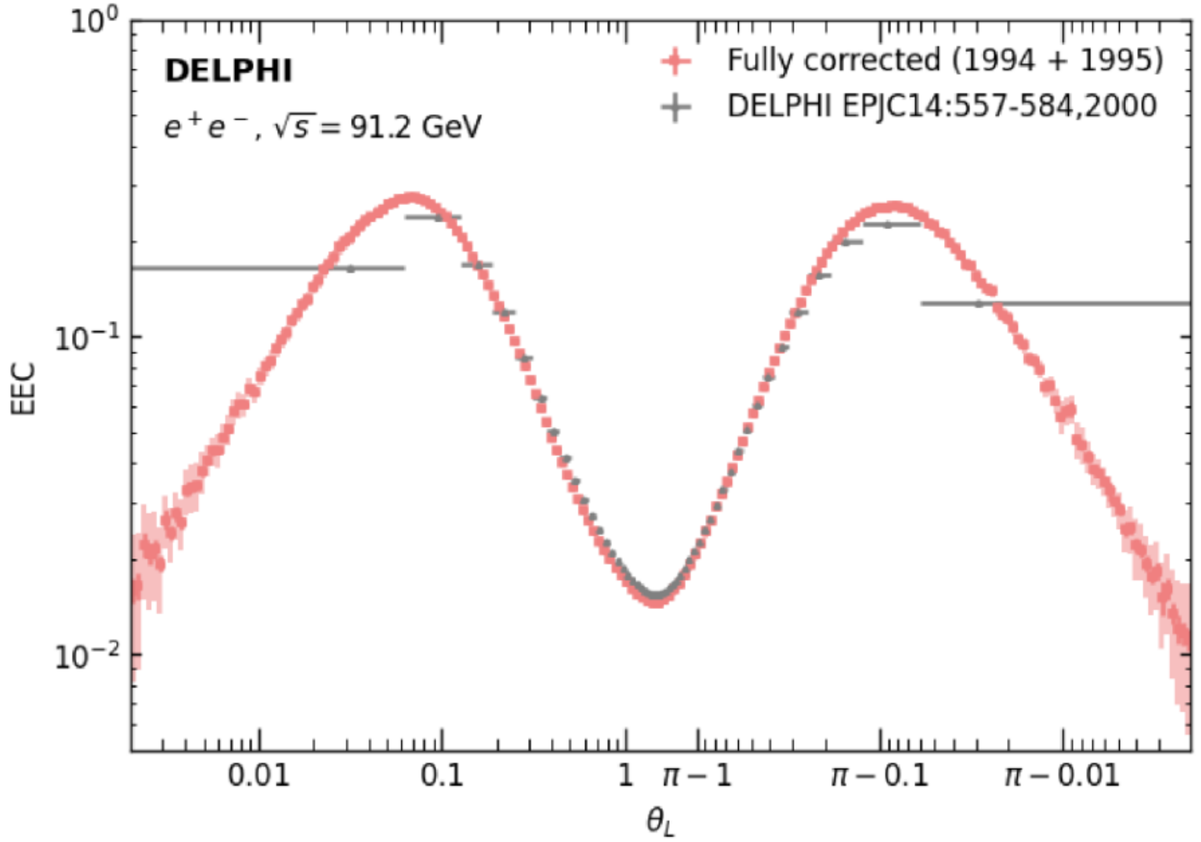


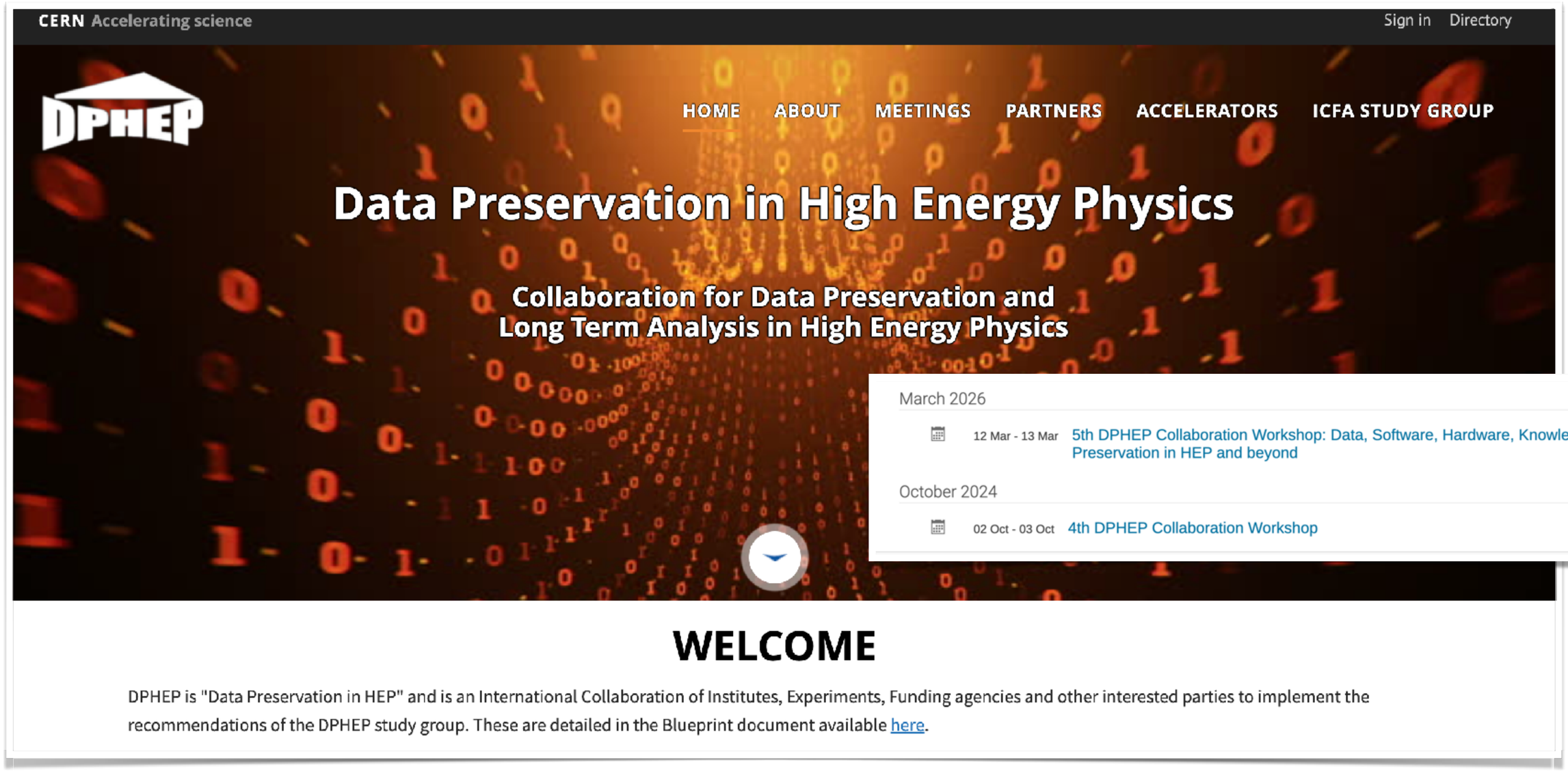
Figure 55: A comparison of the fully-corrected, track-based EEC distribution from this work (light coral) with the previous DELPHI measurement from Ref. [15] (gray). For the measurement from this work, the statistical uncertainties are shown as vertical lines, and the total systematic uncertainties are shown as boxes. For the previous measurement, the error bars represent the statistical and systematic uncertainties added in quadrature. The definition of the EEC in this work differs from that of the legacy DELPHI analysis. Specifically, the legacy measurement normalizes the energy of charged track pairs by the visible event energy, while the current work uses the beam energy.

One more thing

data preservation of other LEP experiments

One more thing: data preservation of other LEP experiments

The Data Preservation in High Energy Physics (DPHEP) collaboration.

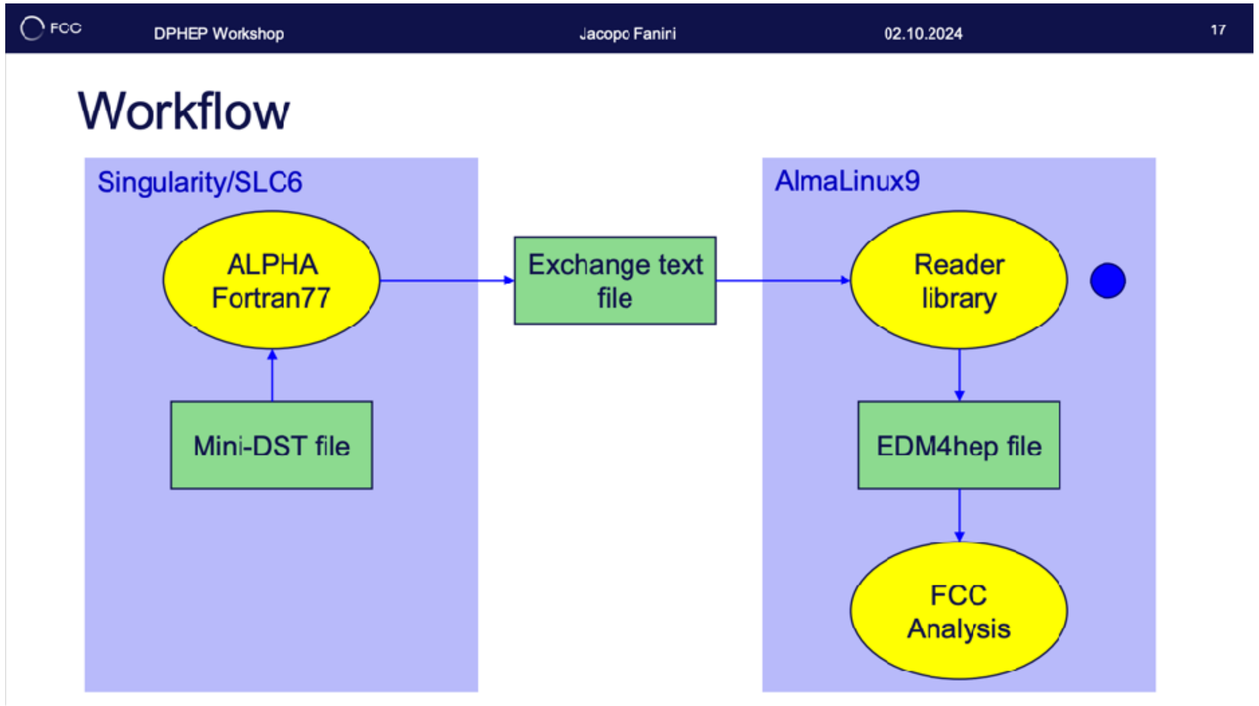


Click on the figures for references or data sources.

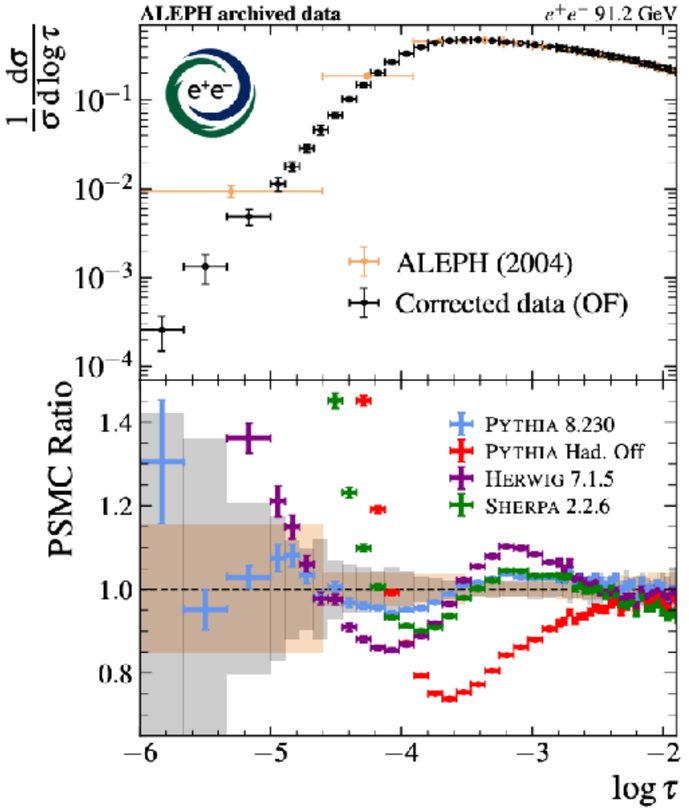
One more thing: data preservation of other LEP experiments

The data conversion tool for ALEPH data from dst to EDM4hep is under developing .

There has been studies applying unbinned unfolding technique to measure the thrust in ALEPH data.



Jacopo Fanini, DPHEP Workshop, 2024



arXiv:2510.22038

One more thing: data preservation of other LEP experiments

Open Data

- So far the OPAL data is not open
  - Interested users contact the OPAL Long Term Editorial Board
  - Users should collaborate with former OPAL member
  - See https://opal.web.cern.ch/of.pdf
- Discussion has started to make the OPAL data open
  - Feedback so far very positive
  - Will profit from DELPHIs experience with the process

Matthias Schröder, DPHEP Workshop, 2024

Click on the figures for references or data sources.

Summary

- ***Data is opened:*** The DELPHI, one of the experiments at LEP, released its full data to the public physics community.
  - Full stack of software for simulation, reconstruction, event display, usage manual and a part of internal documentations are all available.
- ***Everything still work:*** The open data is readable and analyzable.
  - Community developed packages, based on the official analysis software, enable the smooth conversion from the DST data to modern ROOT files.
  - Validated in personal linux + docker.
  - Studies using the LEP open data, incorporating new technology, are increasing.
- ***Let's collaborate!*** If you
  - have any ideas
  - ~~*interested on archaeology*~~
  - ~~*want to find some Higgs-suspect events for commemoration*~~

welcome to have more discussion with me 🤠

中国科学院高能物理研究所
Institute of High Energy Physics, Chinese Academy of Sciences

Looking forward to your participation

CHE Yuzhi (车逾之)
cheyuzhi@ihep.ac.cn

Nov. 17, 2025, IHEP EPD seminar