MARIETTA BLAU
INSTITUTE FOR
PARTICLE PHYSICS

Metadata for DELPHI FCCAnalysis

Dietrich Liko

MBI, Vienna

Metadata and FCCAnalysis

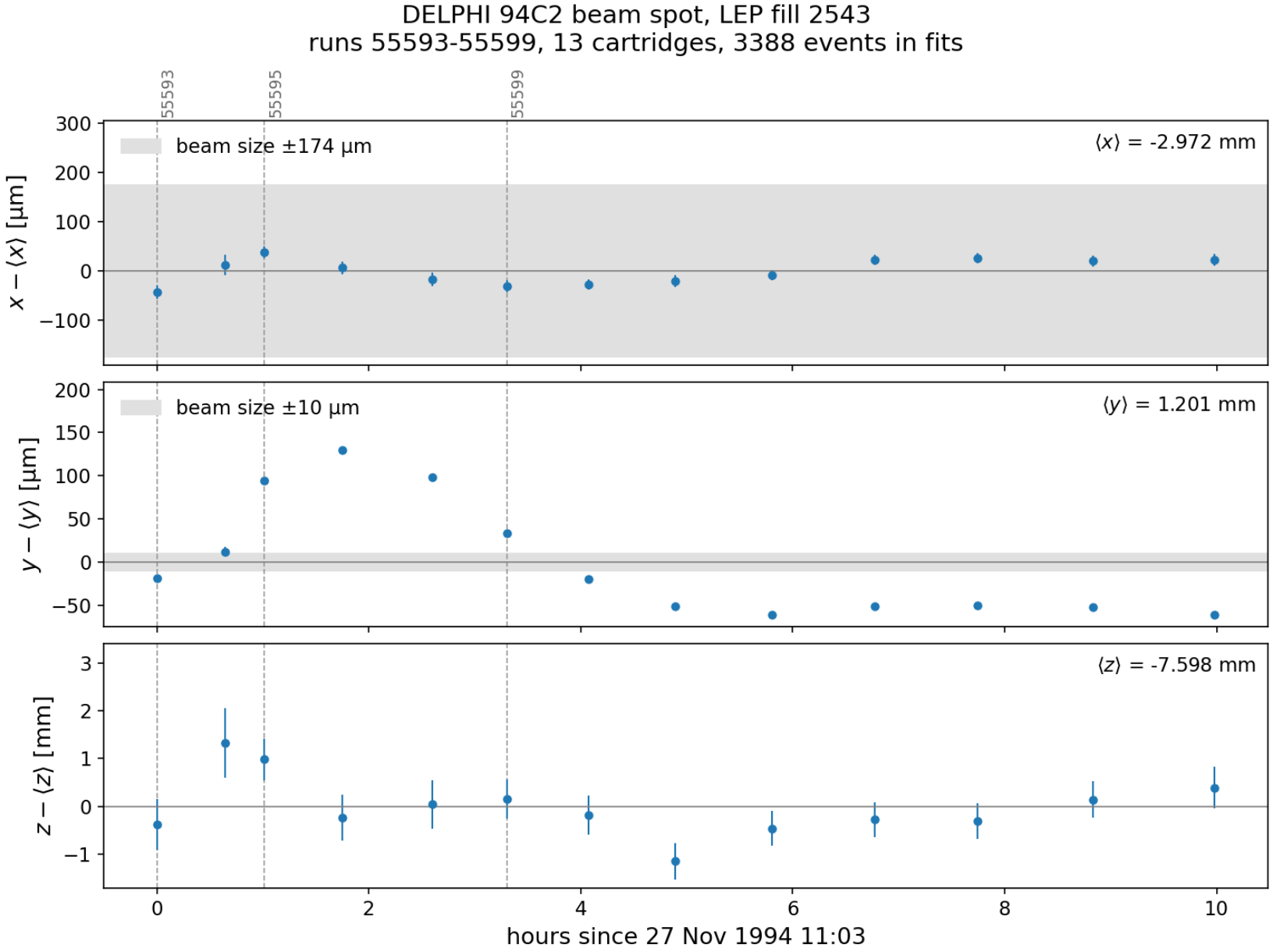
- All known metadata is part of the DELPHI SW release
  - Some in ASCII files in $DELPHI_DAT
  - Some info included into FORTRAN
- Dataset metadata is just in the FATMEN nickname
- Support routines are in FORTRAN and do not support FCCAnalysis (RDataFrame)
- Small C++ library and python support routines
  - SQLITE as backend

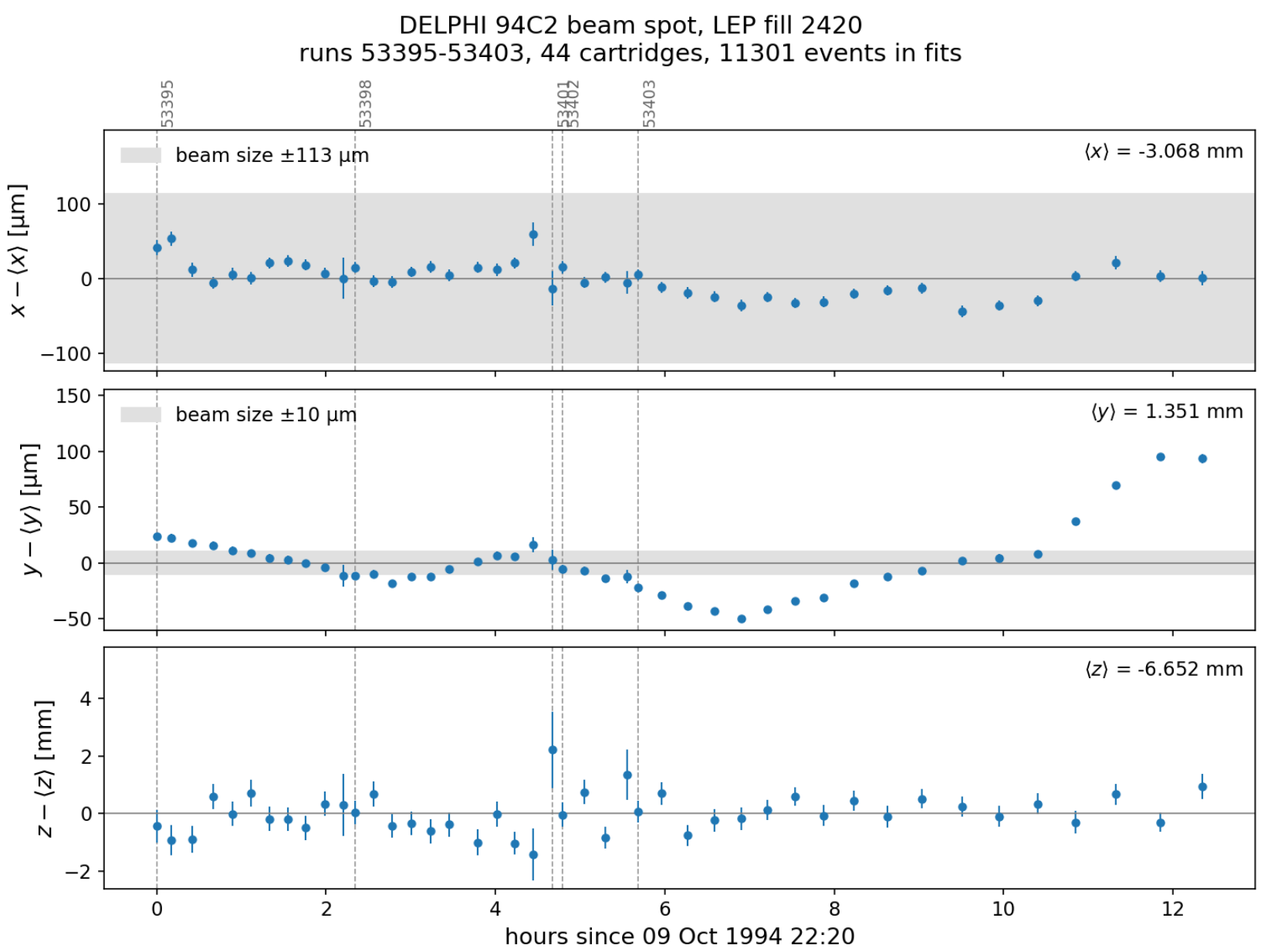
Categories

- Data from $DELPHI_DAT
  - Beam spot
  - Run selection (Subdetector status)
  - Beam energy
  - Luminosity
- Data embedded in source code
  - TPC dE/dx for particle hypotheses
  - RICH refractive indices and Cherenkov angles for particle hypotheses
- Dataset selection
  - Correspondence to OpenData portal
- Utility routines

Beam spot

- Beam spot on the DST has been provided online from the BOMs
  - It is overwritten by the EDM4HEP conversion
  - No further action are required
- Beam spot has been measured over the years in several ways
- It has been determined by several techniques over the years
  - D. Johnson, D. Reid and W. Trischuk, *"A Beamspot Database for Lifetime Measurements"*, DELPHI 92-36 PHYS 168.
  - W.J. Murray, *"Measurement of the beam position in DELPHI"*, DELPHI 96-6 PHYS 590.
- It has been found useful to average periods of several hundred hadronic events


DELPHI 94C2 beam spot, LEP fill 2543
runs 55593-55599, 13 cartridges, 3388 events in fits
55593
55595
55599
beam size ±174 µm
⟨x⟩ = -2.972 mm
x − ⟨x⟩ [µm]
beam size ±10 µm
⟨y⟩ = 1.201 mm
y − ⟨y⟩ [µm]
⟨z⟩ = -7.598 mm
z − ⟨z⟩ [mm]
hours since 27 Nov 1994 11:03


DELPHI 94C2 beam spot, LEP fill 2420
runs 53395-53403, 44 cartridges, 11301 events in fits
53395
53398
53401
53402
53403
beam size ±113 µm
⟨x⟩ = -3.068 mm
x − ⟨x⟩ [µm]
100
0
−100
beam size ±10 µm
⟨y⟩ = 1.351 mm
y − ⟨y⟩ [µm]
150
100
50
0
−50
⟨z⟩ = -6.652 mm
z − ⟨z⟩ [mm]
4
2
0
−2
0
2
4
6
8
10
12
hours since 09 Oct 1994 22:20

Usage

In the Analysis loop

```
class RDFanalysis:


    def analysers(df, data=True):
        df = df.Define("run", 'gpInt(GPIntKeys, GPIntValues, "sDST_EVT_runNumber")')
                .Define("cartridge", 'gpInt(GPIntKeys, GPIntValues, "sDST_EVT_fileSeq")'
        )
        if data:
            df = df.Define("BS", "beamspot.getData(run, cartridge)")
        else:
            df = df.Define("BS", "beamspot.getMC(sDST_LUJ_GenParticles)")
        return (
            df.Define("BS_x", "BS.position.x")
            .Define("BS_y", "BS.position.y")
            .Define("BS_z", "BS.position.z")
        )
```

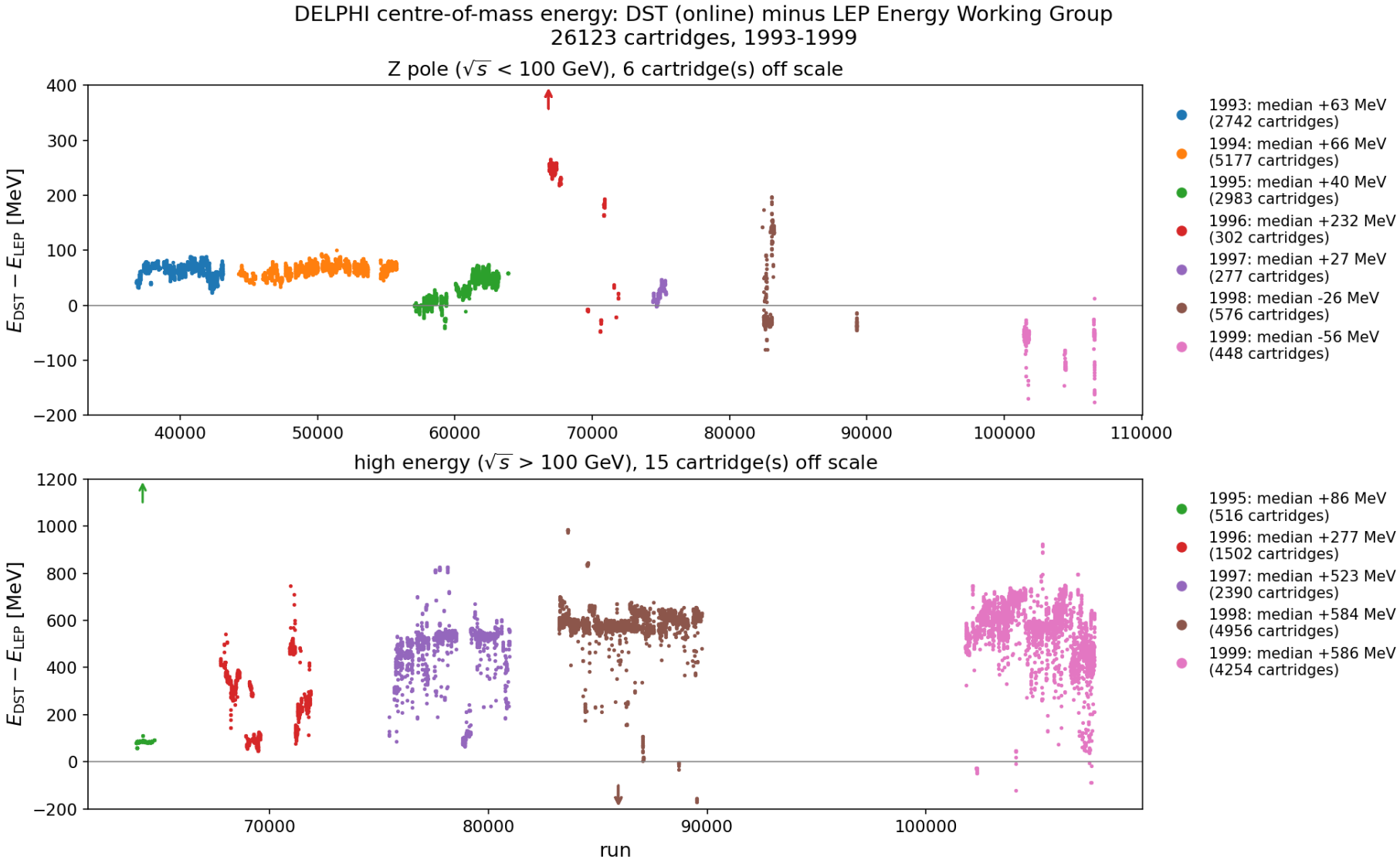
Initialization

```
from delphi_metadata.rdf import bind, load


load()


db = ROOT.DelphiMetadata.Db(samples.DB_PATH)
beamspot = ROOT.DelphiMetadata.Beamspot(db, "94C2")
bind("beamspot", beamspot)
```

LEP Energy

- The LEP Energy on the DST is the online LEP energy
- It was later corrected

- Origin: Notes by the LEP Energy Working Group
  - LEP ECAL 96-05, "LEP Energy Calibration in 1993, 1994 and 1995”
  - LEP ECAL 96-08, "RF Corrections to the LEP Energy for the '94 Running"


DELPHI centre-of-mass energy: DST (online) minus LEP Energy Working Group
26123 cartridges, 1993-1999
Z pole ($\sqrt{s}$ < 100 GeV), 6 cartridge(s) off scale
$E_{DST} - E_{LEP}$ [MeV]
1993: median +63 MeV (2742 cartridges)
1994: median +66 MeV (5177 cartridges)
1995: median +40 MeV (2983 cartridges)
1996: median +232 MeV (302 cartridges)
1997: median +27 MeV (277 cartridges)
1998: median -26 MeV (576 cartridges)
1999: median -56 MeV (448 cartridges)
high energy ($\sqrt{s}$ > 100 GeV), 15 cartridge(s) off scale
$E_{DST} - E_{LEP}$ [MeV]
1995: median +86 MeV (516 cartridges)
1996: median +277 MeV (1502 cartridges)
1997: median +523 MeV (2390 cartridges)
1998: median +584 MeV (4956 cartridges)
1999: median +586 MeV (4254 cartridges)
run

Run quality as recorded by the detectors

- Each detector had its own procedure
- Each analysis would have its own selection
- No “golden” JSON

- Simulation has fixed effects, but not temporary effects like voltage trips etc. …
  - In function of the analysis MC/data has to be studied

- Bad HPC event rejection (DAQ readout problem during 1997)

```
 dliko  …/delphi-metadata  main !?  16:08  loaded/allowed
 dmd runquality breakdown --year 1994
```

detector	0	1	2	3	4	5	6	7	8	9
mvx_a	1.1	0.0	0.1	0.0	0.1	0.1	0.1	98.5	0.0	0.0
mvx_c	1.1	0.1	0.1	0.0	0.2	0.1	0.1	98.4	0.0	0.0
id_jet	2.1	0.3	0.2	0.0	0.4	0.9	0.3	95.7	0.0	0.0
id_trig	2.2	0.1	1.1	0.0	1.1	0.2	0.3	95.0	0.0	0.0
tpc_0	0.4	0.1	0.0	0.0	0.0	0.2	0.0	99.2	0.0	0.0
tpc_1	0.5	0.0	0.0	0.0	0.0	0.2	0.0	99.3	0.0	0.0
brich_l	4.4	0.1	0.1	0.1	0.6	0.0	0.0	94.7	0.0	0.0
brich_g	3.2	0.1	0.2	0.3	0.5	0.0	0.0	95.7	0.0	0.0
od_b	1.4	0.3	0.5	0.1	0.3	0.0	29.3	68.2	0.0	0.0
od_d	1.4	0.2	0.4	0.1	0.3	0.0	29.3	68.2	0.0	0.0
hpc_0	1.0	0.2	0.2	0.1	0.3	0.0	0.0	98.2	0.0	0.0
hpc_1	1.2	1.0	0.2	0.1	0.3	0.0	0.0	97.2	0.0	0.0
hcab_a	1.6	0.1	1.0	0.1	0.1	0.0	0.0	97.1	0.0	0.0
hcab_c	1.7	0.2	1.2	2.7	0.3	0.0	0.0	93.8	0.0	0.0
mub_b	3.0	0.1	0.0	0.1	3.8	13.5	34.6	44.9	0.0	0.0
mub_d	4.2	0.3	0.0	2.3	5.3	11.5	36.3	40.0	0.0	0.0
fca_a	15.0	0.5	0.5	0.0	0.0	5.9	21.4	56.6	0.0	0.0
fca_c	18.5	0.6	0.0	39.3	0.1	0.1	41.2	0.3	0.0	0.0
rif_a	0.5	0.0	0.1	0.0	0.0	0.0	0.0	99.3	0.0	0.0
rif_c	0.6	0.0	0.0	0.1	0.0	0.0	0.0	99.4	0.0	0.0
fcb_a	4.2	0.2	0.8	0.1	0.4	0.8	1.0	92.6	0.0	0.0
fcb_c	3.4	0.2	0.8	0.1	0.2	0.3	1.7	93.3	0.0	0.0
emf_a	0.5	0.1	0.2	0.1	0.0	6.5	90.8	1.8	0.0	0.0
emf_c	0.5	0.1	0.2	3.9	0.3	10.6	84.5	0.0	0.0	0.0
hcaf_a	1.7	0.1	0.1	0.0	0.0	0.0	0.0	98.1	0.0	0.0
hcaf_c	1.6	0.2	0.1	0.1	0.0	0.0	0.0	98.1	0.0	0.0
muf_a	2.3	0.4	0.2	0.0	0.0	0.2	8.6	88.3	0.0	0.0
muf_c	2.4	1.0	0.2	0.0	0.1	0.1	2.6	93.6	0.0	0.0
sat_cal	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sat_tra	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
vsat	1.3	0.0	0.0	0.0	0.0	0.0	0.0	98.7	0.0	0.0

```
 dmd runquality legend
```

status	meaning
0	unusable
1	< 50% of nominal
2	50–65% nominal
3	65–80% nominal
4	80–90% nominal
5	90–95% nominal
6	95–99...% nominal
7	> 99...% nominal
8	efficiency varies during run
9	status unknown

detector	meaning
mvx_a	Vertex Detector (Microvertex, silicon strips), A-side half
mvx_c	Vertex Detector (Microvertex, silicon strips), C-side half
id_jet	Inner Detector, jet drift-chamber part
id_trig	Inner Detector, trigger proportional-chamber part
tpc_0	Time Projection Chamber (main barrel tracker), drift half 0
tpc_1	Time Projection Chamber (main barrel tracker), drift half 1
brich_l	Barrel RICH (Ring Imaging CHerenkov), liquid radiator
brich_g	Barrel RICH (Ring Imaging CHerenkov), gas radiator
od_b	Outer Detector (barrel drift-tube trigger layers), group B

Integration into FCCAnalysis

- Filter with run, cartridge, event as parameters
- Statistic at the end (in python atexit handler)
- Can write CMS inspired JSON file

Luminosity

SAT (Small Angle Tagger), main luminometer 1989–1993

- Lead and scintillating-fiber calorimeter, fibers parallel to the beam
- Covers 2.5°–8°; a lead mask defined the acceptance

STIC (Small angle Tile Calorimeter), main luminometer 1994–2000

- Lead–scintillator shashlik calorimeter at ±2.2 m, covering 29–185 mrad
- Silicon pad planes for shower position, a scintillator veto for e/γ separation, and a tungsten mask defining the acceptance

VSAT (Very Small Angle Tagger), 1990s through LEP2

- Four silicon–tungsten modules at about 7.7 m, covering about 5–7 mrad
- Very high Bhabha rate gave a precise relative luminosity and monitoring of beam background and beam parameters

Available values

```
dliko …/delphi-metadata main !?
 dmd luminosity list
year  detector  lumi_nb     entries
----  --------  ----------  -------
1991  VSAT      11139.132   3807
1992  SAT       24143.071   4298
1992  VSAT      23253.896   4323
1993  COMBINED  36371.667   2755
1993  SAT       36336.008   2741
1994  STIC      46301.362   4925
1995  STIC      31715.650   13245
1996  STIC      21198.914   1766
1997  STIC      61684.864   2623
1998  STIC      159245.384  5365
1999  STIC      226281.296  4248
2000  STIC      224230.210  7846
```

```
dliko …/delphi-metadata main !? 16:0
 dmd luminosity list --year 2000 --summary
year  detector  point     lumi_nb     entries
----  --------  --------  ----------  -------
2000  STIC      LEP2-200  637.341     45
2000  STIC      LEP2-202  2284.225    96
2000  STIC      LEP2-205  71850.604   2182
2000  STIC      LEP2-207  134442.119  4458
```

Some remarks

- The CLI supports asking for detector conditions
- JSON file or detector selection can be used to calculate the effective Luminosity
- 2000 had miniramps – I did not look into that.
- All considerations assume an efficient trigger. Delphi had also many triggers and also trigger scaling …. I did not look into that. Cannot be analyzed with the tools considered at this point. Trigger MC is also unclear to me.

TPC dE/dx

- It has been studied by various authors.
- Has been fitted on the data

Expected TPC $\mathrm{d}E/\mathrm{d}x$ (GETDNEW, F. Parodi & A. Stocchi 1996; PXDST version ≥ 333)

$$\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = \frac{1}{N}\left[\frac{P_1}{\beta^2}\Big(\ln\beta\gamma - \delta(y)\Big) + R(\beta\gamma)\right], \qquad x = \beta\gamma = \frac{p}{m}, \quad y = \ln x$$

Density effect:

$$\delta(y) = \begin{cases} 0 & y < y_0 \\ A_2\,(y - y_a) + A_1\,(y_1 - y)^3 & y_0 \leq y < y_1 \\ A_2\,(y - y_a) & y \geq y_1 \end{cases} \qquad \begin{aligned} y_0 &= y_a - \tfrac{A_1}{A_2}\big(\tfrac{A_2}{3A_1}\big)^{3/2} = 4.26 \;\; (\beta\gamma \approx 71) \\ y_1 &= y_0 + \big(\tfrac{A_2}{3A_1}\big)^{1/2} = 8.98 \;\; (\beta\gamma \approx 7900) \end{aligned}$$

Low-$\beta\gamma$ term:

$$R(x) = \begin{cases} P_2 + P_3/x^2 & x > x_0 \\ R_0 + P_8/x^{P_7} & x \leq x_0 \end{cases} \qquad \begin{aligned} x_0 &= \Big(\tfrac{2P_3}{P_7 P_8}\Big)^{3/(P_7+1)} = 0.80 \\ R_0 &= P_2 + P_3/x_0^2 - P_8/x_0^{P_7} \;\; \text{(continuous at } x_0) \end{aligned}$$

P_1	P_2	P_3	A_2	A_1	y_a	P_7	P_8	N
0.13914	0.65853	0.93177	1.0	0.015	5.835	1.5953	1.4166	0.921605

In MIP units: minimum $= 1$ at $\beta\gamma \approx 3.9$; plateau $= (P_1\,y_a + P_2)/N = 1.5955$. Same curve for data and MC.

Error already stored in EDM4HEP

No need to recalculate it

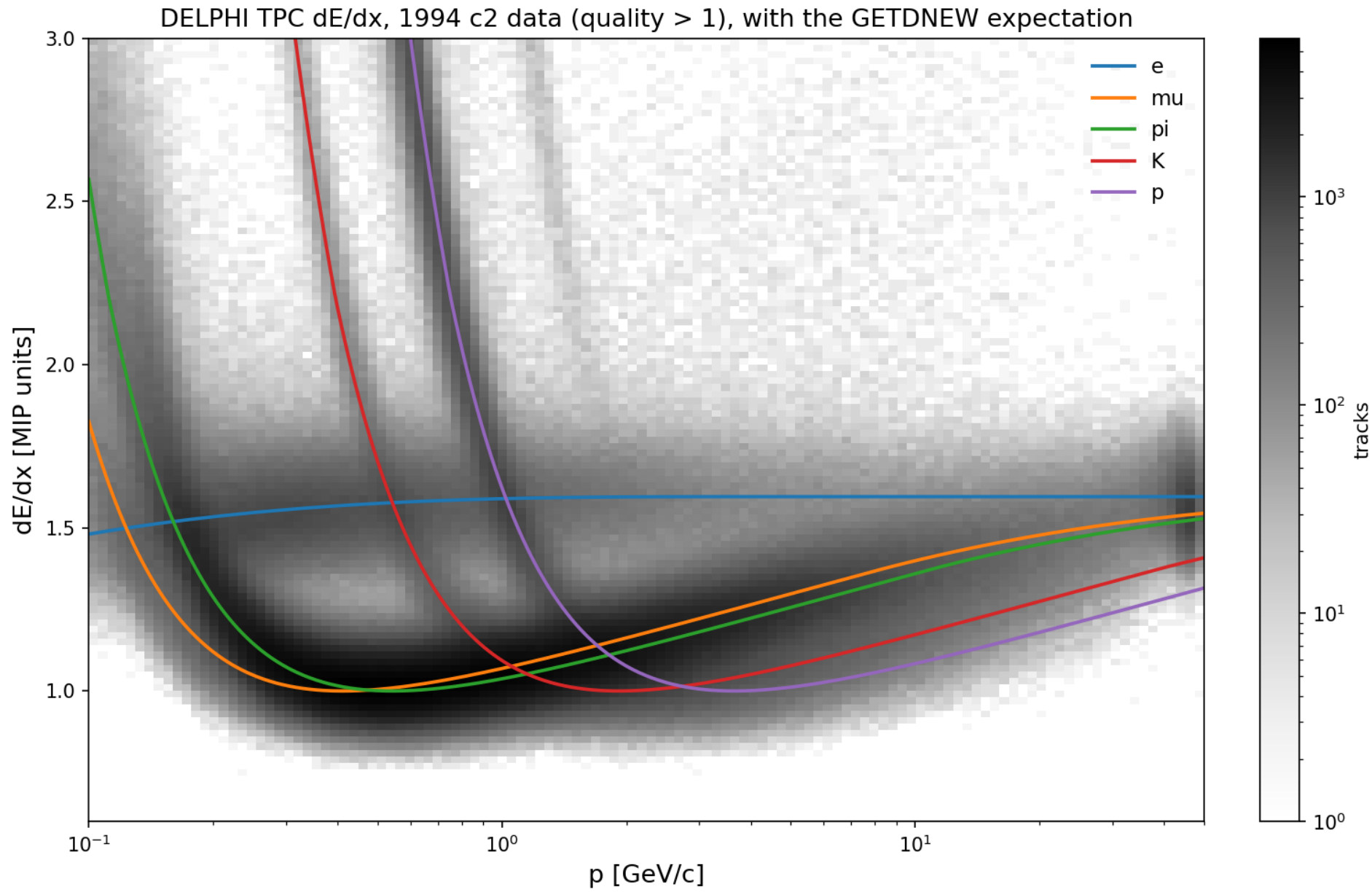
Expected $\mathrm{d}E/\mathrm{d}x$ **error** (BBDXER, P. Antilogus 1996; PXDST version ≥ 333)

$$\sigma = 1.0286\;E\sqrt{\frac{1.0757/N_w + 0.823/N_w^2}{E^{1.29}\;g^{0.926}}}, \qquad g = \min\Big(\frac{\text{gap}}{0.4},\, 1.71\Big), \qquad \sigma \to 1.33\,\sigma \text{ (truncation method 2)}$$

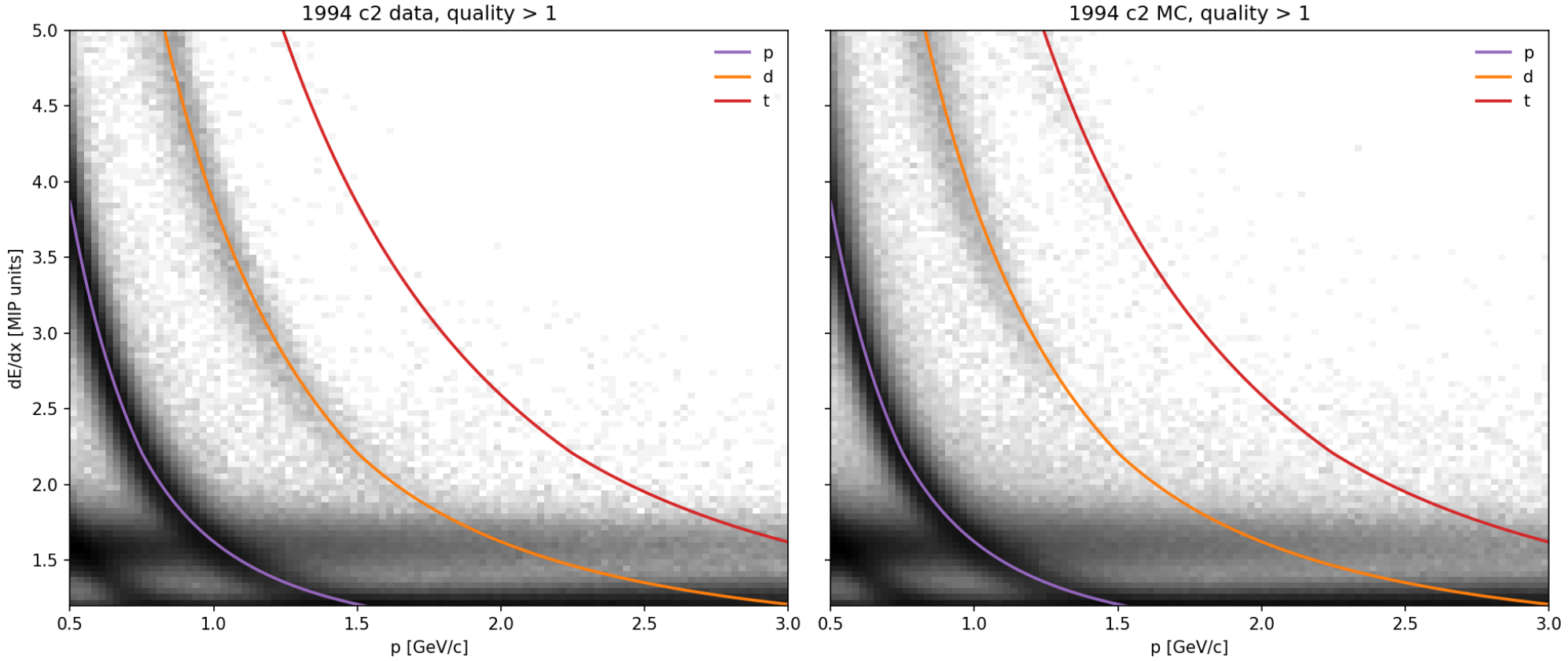
with $E = \langle \mathrm{d}E/\mathrm{d}x \rangle$ of the hypothesis and N_w the number of wires.

Pull and probability per hypothesis $i \in \{e, \mu, \pi, K, p\}$ (RPRODE):

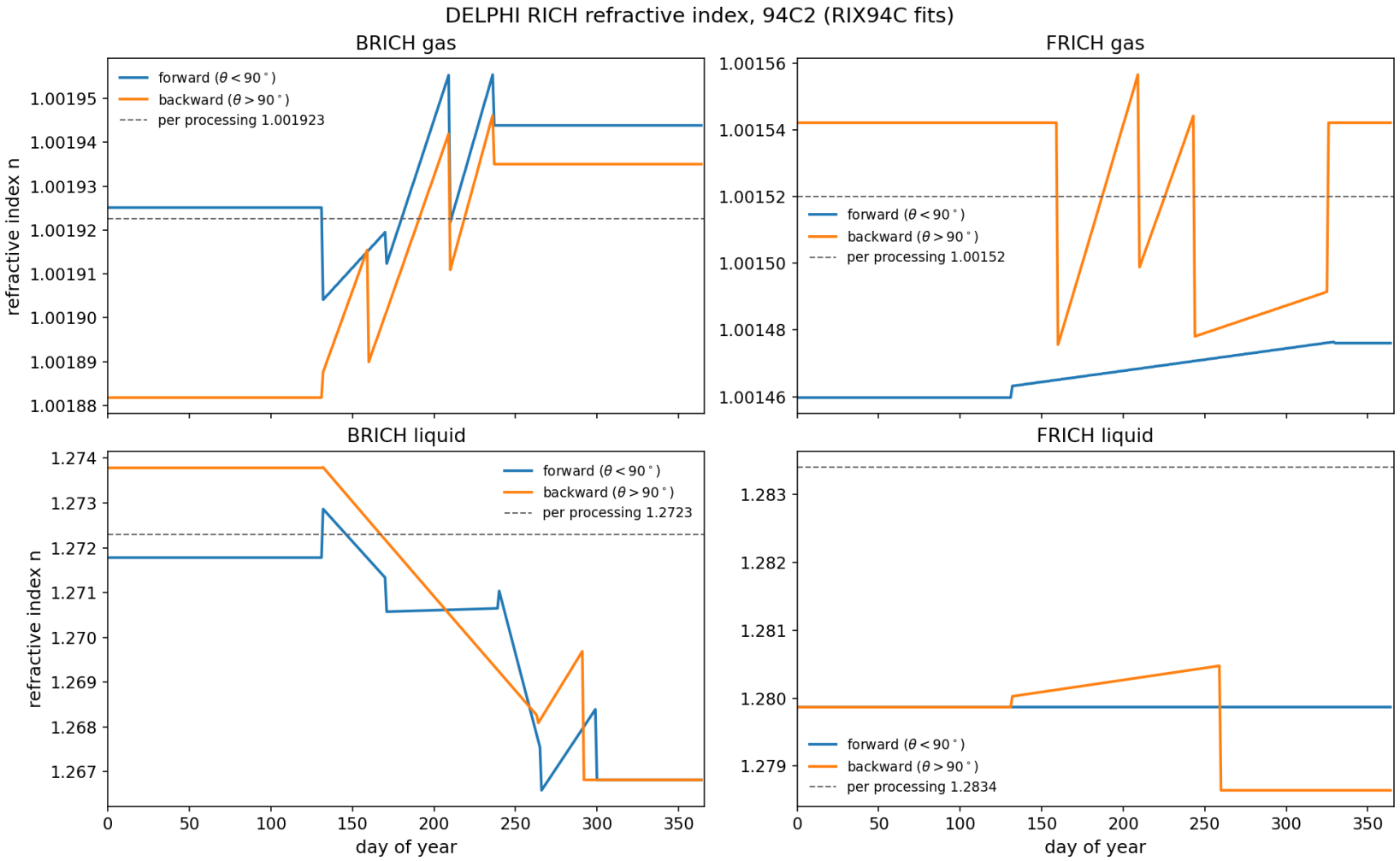
$$r_i = \frac{(\mathrm{d}E/\mathrm{d}x)_{\text{meas}} - E_i}{\sigma_i}, \qquad P_i = \frac{\max\big(e^{-r_i^2/2},\, 0.02\big)}{\sum_j \max\big(e^{-r_j^2/2},\, 0.02\big)} \qquad (N_w > 30)$$


DELPHI TPC dE/dx, 1994 c2 data (quality > 1), with the GETDNEW expectation
dE/dx [MIP units]
p [GeV/c]
tracks
e
mu
pi
K
p
3.0
2.5
2.0
1.5
1.0
10^{-1}
10^{0}
10^{1}
10^{3}
10^{2}
10^{1}
10^{0}

Deuterons clearly visible



RICH Refractive Index



In summary

- I consider it save if you use the tags.
  - Basically, they are flags, that allow to tag various particles
  - Use the heavy tag for efficiency

- Underlying quantities need more understanding ….
  - Special data samples like D* or phi statistically limited

What is the best PID strategy ?

For particles

- Tags from evaluations of likelihoods
- Likelihoods with expected values
- Use underlying quantities (e.g. predictions from MC)

For hadron flavors

- Resolve particle pid separate from a hadron flavor tag
- Integrated particle pid in flavor tag

Datasets

- In total about 17K datasets
  - The number is driven by MC samples

- Metadata is encoded in the nickname

Example for dataset search

fatfind –N “short.*”

Example for content listing

fatfind short94_c2

What is xs_yfsww16lpaamvud_e188.6_m80.4_c98_1l_e1

Simulated WW events from YFSWW3 1.16 at 188.6 GeV with M_W = 80.4 GeV. They use the LPA_a treatment of the O(α) corrections and the "MV" setting of the second switch, and were run through the 1998 detector simulation and e1 reconstruction.

(Approx before the the full one-loop four fermion calculation was available.)

DELPHI dataset nickname: xs_yfsww16lpaamvud_e188.6_m80.4_c98_1l_e1

Token	Meaning	Confidence
`xs`	Extended Short DST (XSDST)	Certain
`yfsww16`	YFSWW3 generator v1.16: WW production with O(α) electroweak corrections	High
`lpaa`	Leading Pole Approximation, variant a (LPA_a); siblings use lpab	High
`mvud`	Value of a second generator switch; alternative value is scud	Structure certain, meaning uncertain
`e188.6`	Centre-of-mass energy 188.6 GeV (1998 "189 GeV" run)	Certain
`m80.4`	Generated W mass, 80.4 GeV	High
`c98`	1998 detector conditions in simulation	Certain
`1l`	Production/list tag of newer naming scheme; constant across samples	Medium
`e1`	DELANA reconstruction processing e1, matching 1998 data	Certain

Thanks Claude!

Extract Metadata

- Should be identical information to the “facets” of the Open Data Portal
  - Format, statistic, energy, processing versions
  - MC should include process and cross-section, if available

delphi-metadata datasets find –year <year> --format short --data

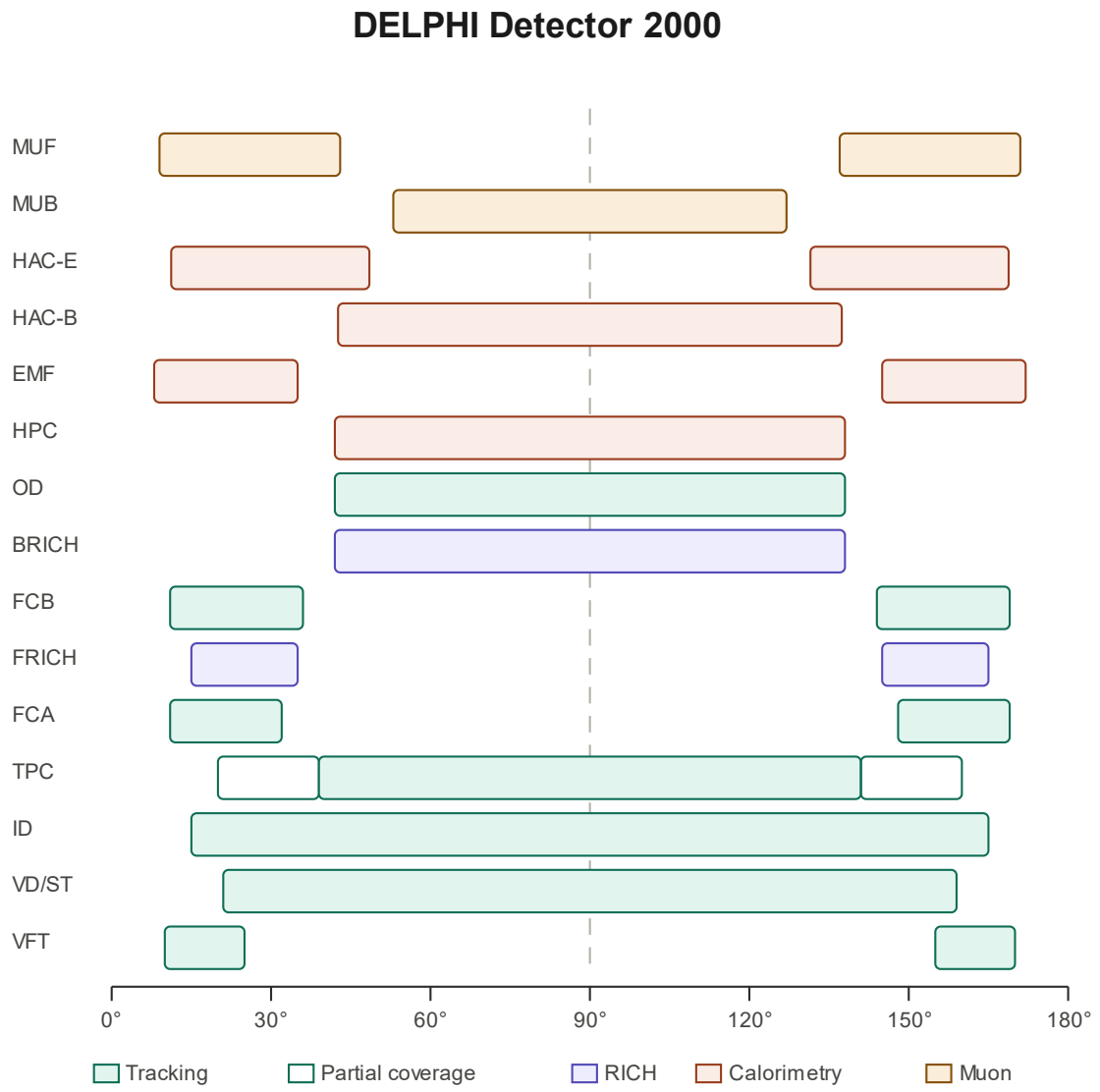
- For MC
  - Some datasets are described on the websites
  - Some datasets have a freeform text associated containing useful info
  - Some datasets have just the name

- Trying to use CERN AI Gateway to extract information at scale (Qwen3.8-27B). Needs some rules collected by Claude

- Needs still a bit more work ….

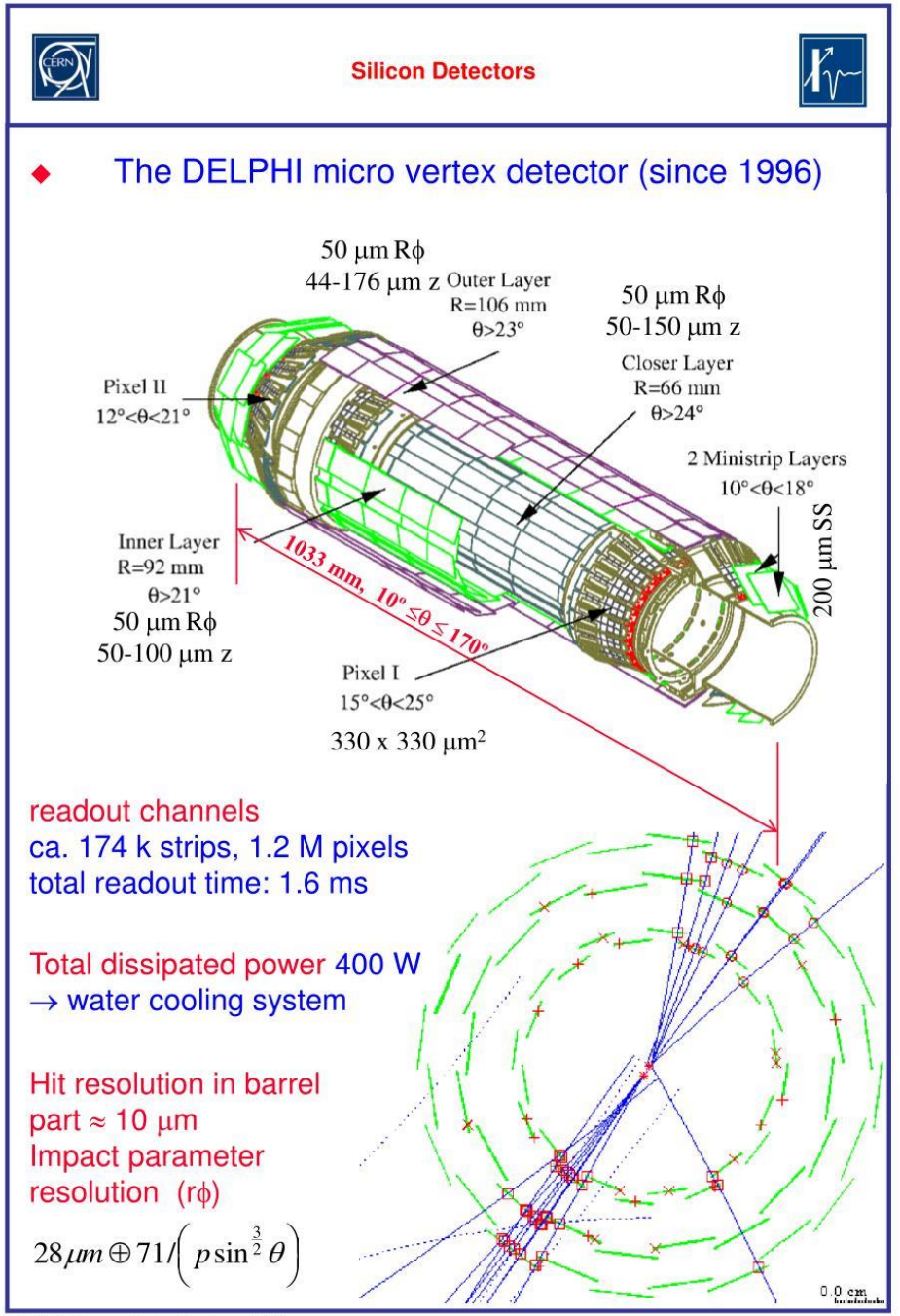
Utilities

- DELPHI had some standard particle selections. They require some pre-calculations
- Bitted word for detector selection
  - It's easy as word is present as attribute in the frame
  - C++ help routine for selections and printout



VD Layers

- VD hit pattern as bitted layer map
  - Can be derived from the VD Hits stored in EDM4HEP
- Non trivial (Merci Claude!)



Plan

- http://gitlab.cern.ch/liko/delphi-metadata
- Some checks still required
- Installable tar file available at the gitlab release tab
  - For a specific Key4hep version
- Not a general solution for metadata. Some information could be also integrated into the data
  - To me not obvious if one can find a generic solution