JOI Impacts

Kaili Zhang

zhangkl@ihep.ac.cn

Usage on PrimeTagSvc

Helper function and debug information provided;

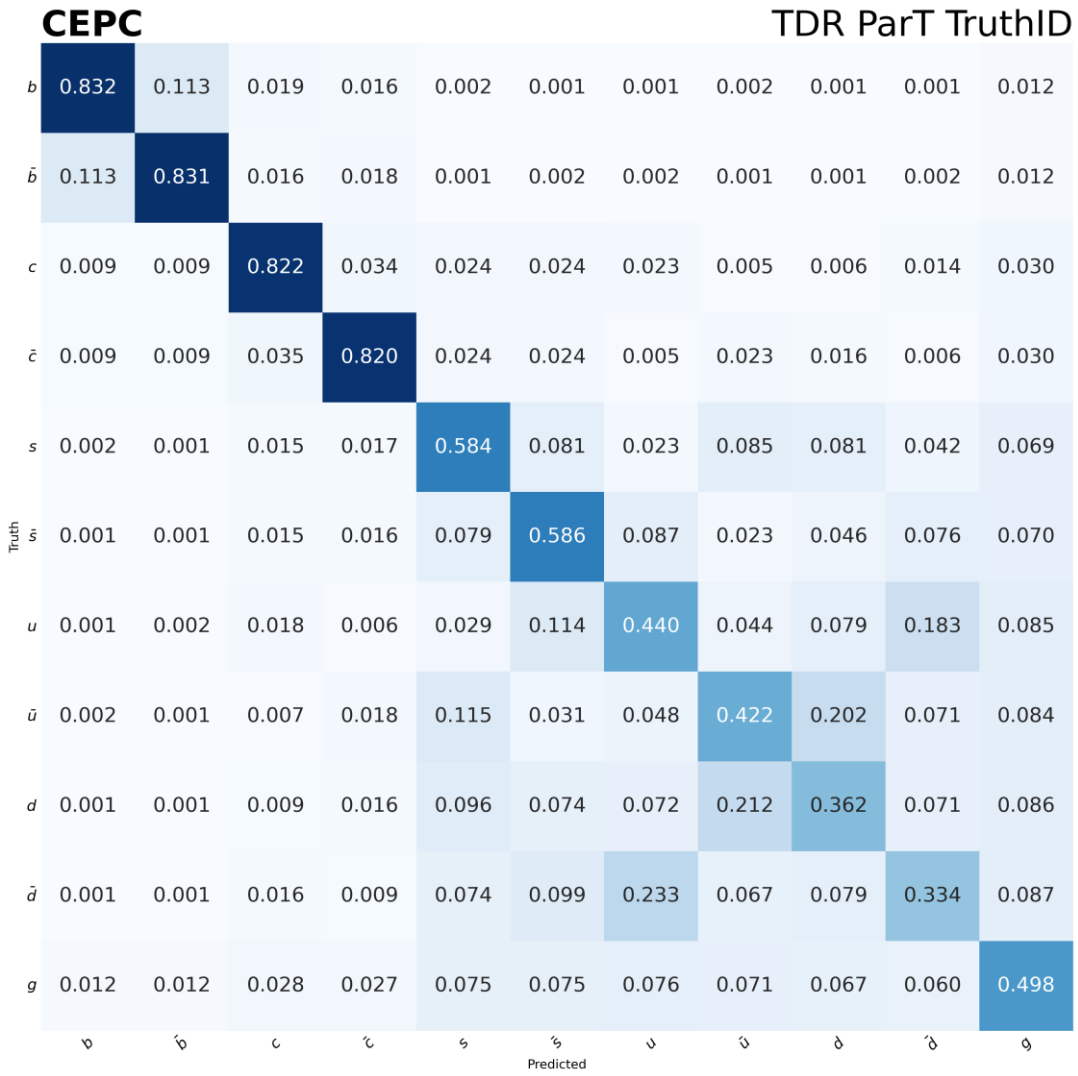
```
std::vector<float> GetProb() const override;
int get_type_M11() const override;
int get_type_M6() const override;

// M11 Boolean functions (11 total)
bool is_b_quark() const override;
bool is_b_bar_quark() const override;
bool is_c_quark() const override;
bool is_c_bar_quark() const override;
bool is_s_quark() const override;
bool is_s_bar_quark() const override;
bool is_u_quark() const override;
bool is_u_bar_quark() const override;
bool is_d_quark() const override;
bool is_d_bar_quark() const override;
bool is_gluon() const override;

// M6 Boolean functions (6 total)
bool is_b_jet() const override;
bool is_c_jet() const override;
bool is_s_jet() const override;
bool is_u_jet() const override;
bool is_d_jet() const override;
bool is_gluon_jet() const override;
```

```
============================================================
                Jet Summary Report
============================================================
Total Events Processed: 100
Total Jets Processed  : 200
------------------------------------------------------------
Jet Flavor Fractions (M6-level):
  - b-jets (b+b_bar) : 190 / 200 (95.00%)
  - c-jets (c+c_bar) : 2 / 200 (1.00%)
  - light-jets       : 8 / 200 (4.00%)
Detailed b-jet Counts (M11-level):
  - b quark jets     : 88 / 200 (44.00%)
  - b_bar quark jets : 102 / 200 (51.00%)
------------------------------------------------------------
Event Pair Fractions (for 2-jet events):
  - bb events: 93 / 100 (93.00%)
  - cc events: 0 / 100 (0.00%)
  - ss events: 0 / 100 (0.00%)
  - gg events: 3 / 100 (3.00%)
    - b + b          : 7 (7.53% of bb)
    - b_bar + b_bar  : 13 (13.98% of bb)
    - b + b_bar      : 73 (78.49% of bb)
============================================================
```

Known issue

- Transition degrading 2%-5% from H(125)qq to Z(91.2)qq and vvH(E360)qq.
  - If any cases b eff <90% please let me know.
  - Zqq /Event model available if needed.
- Removing isolated leptons/photons improves performance
  - Ensure “Jet” is exactly you referring.

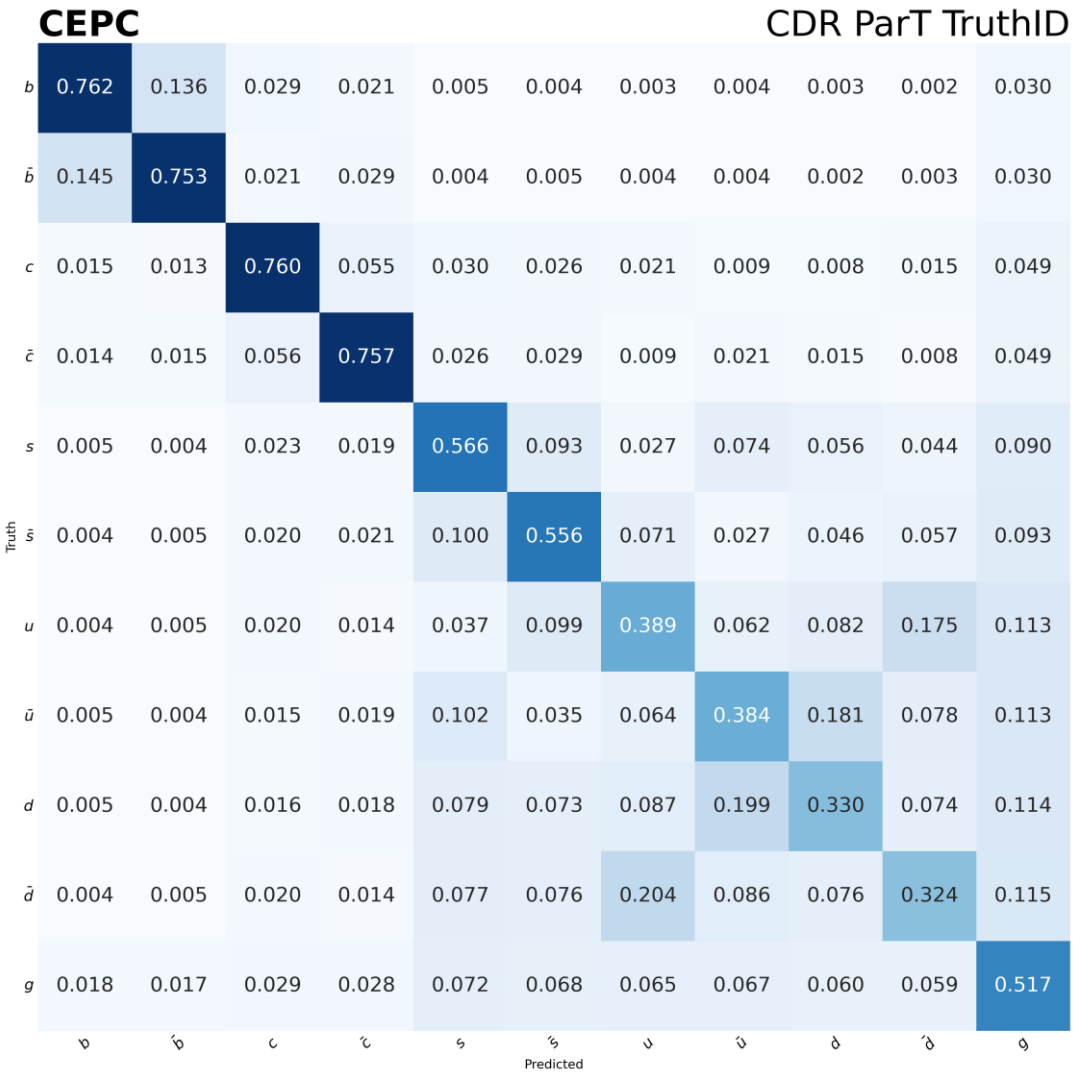
Baseline: TDR, Full, ParT, Perfect ID:

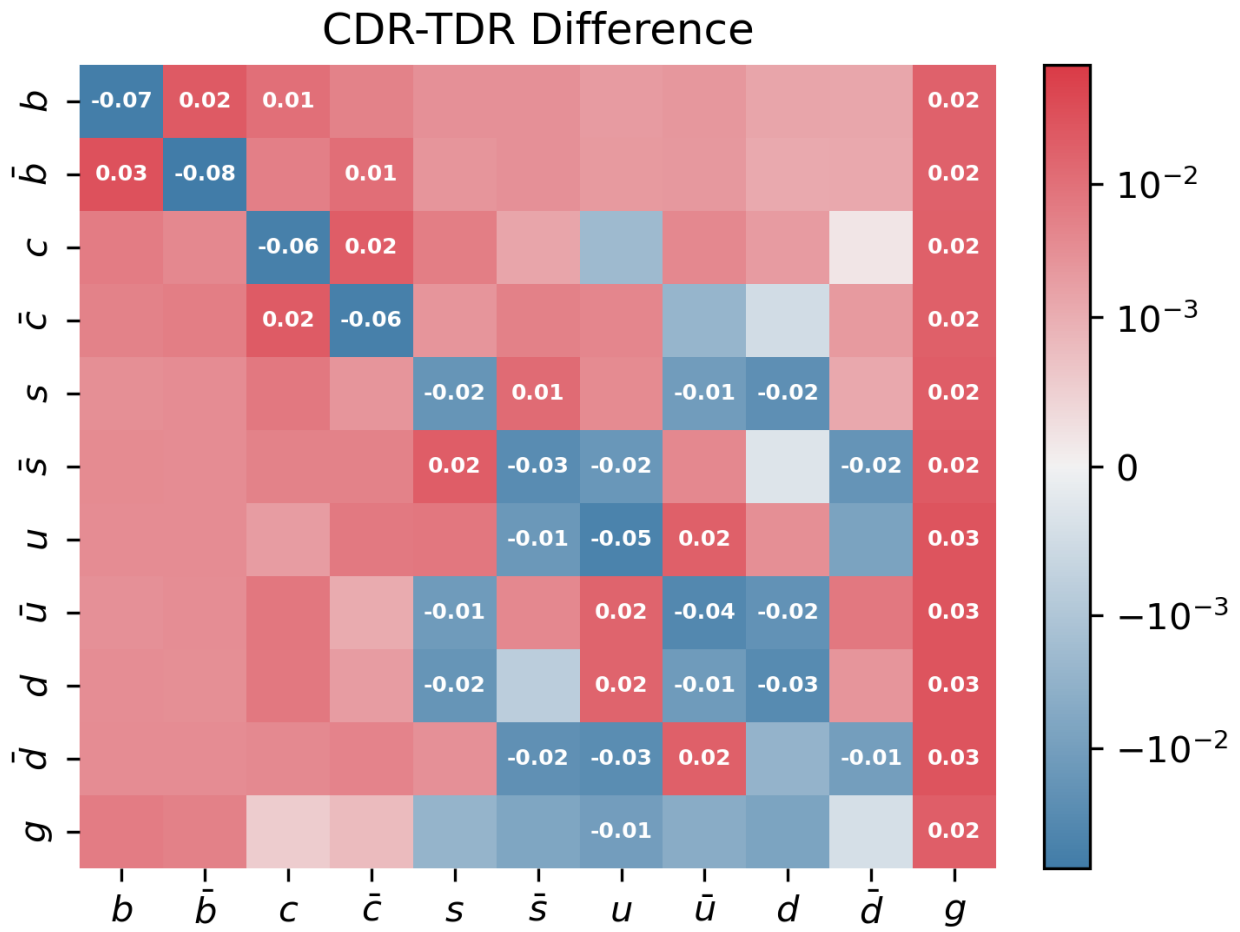


Geometry	Algo	PID	Simu	Metric
TDR	ParT	Truth	full	0.5936
TDR	ParT	Reco	full	0.5562
TDR	PN	Reco	full	0.5401
CDR	ParT	Truth	fast	0.5543
CDR	PN	Truth	fast	0.5435
CDR	PN	Truth	full	0.5285*

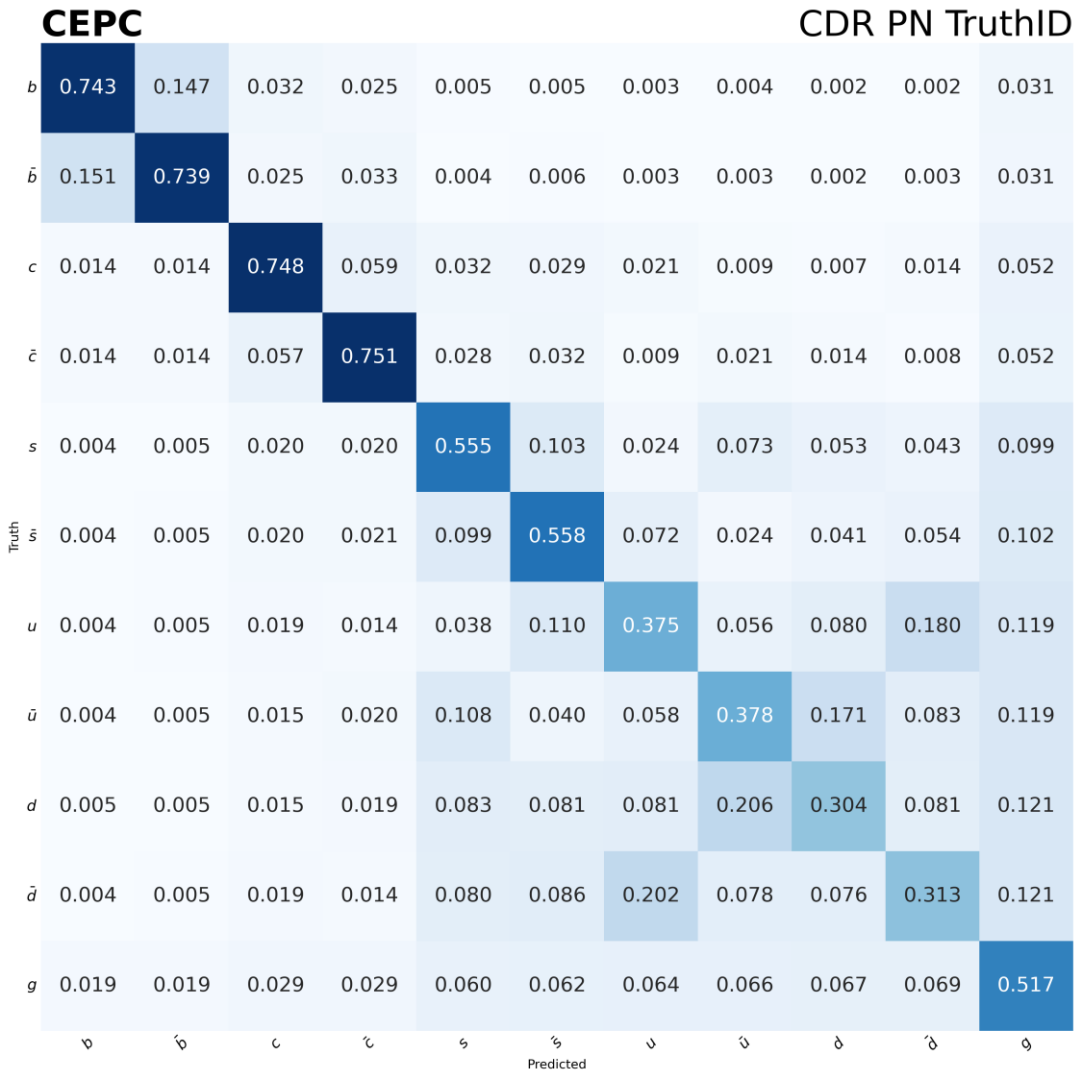
*:From YF.

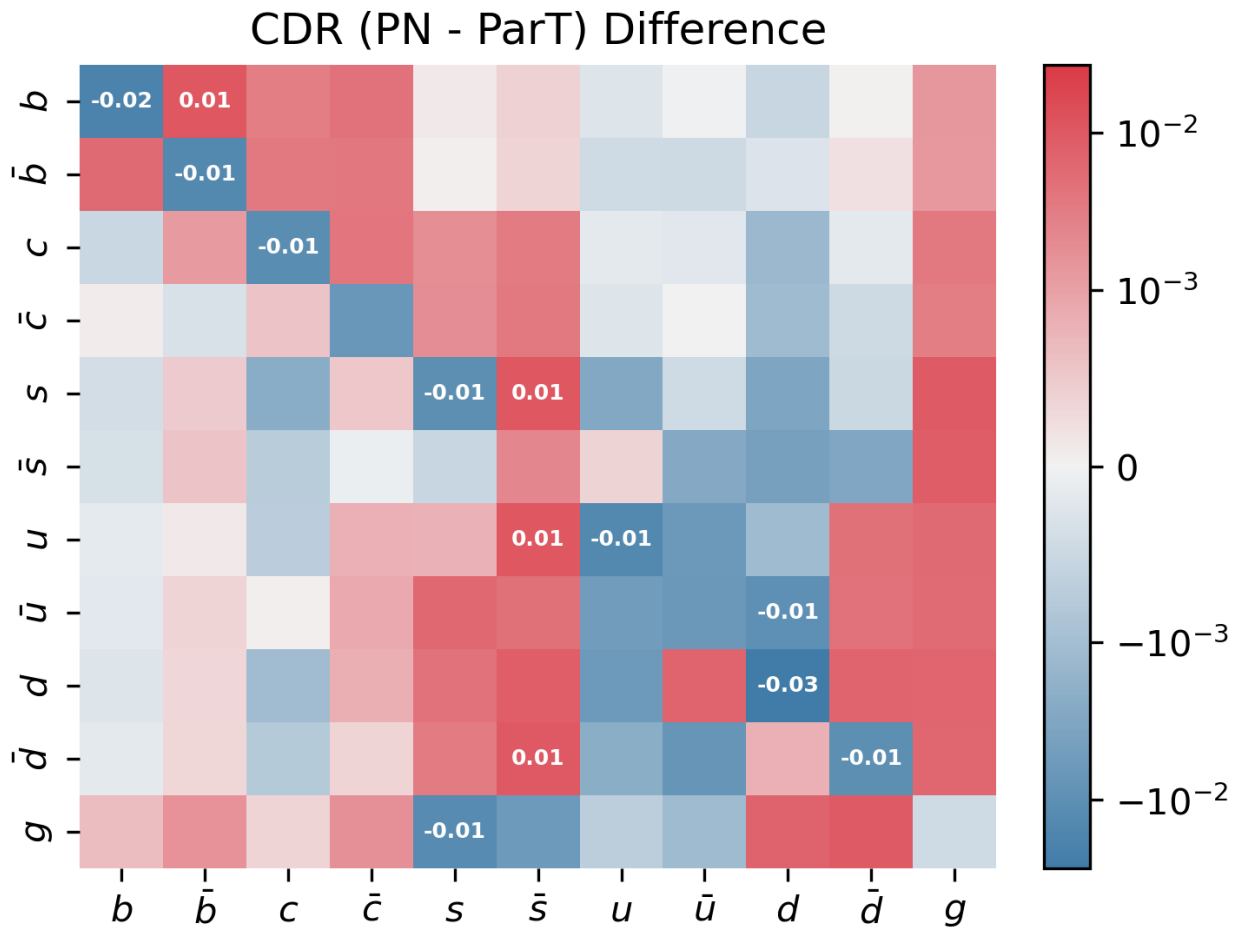
Difference in CDR(Fast)/TDR:



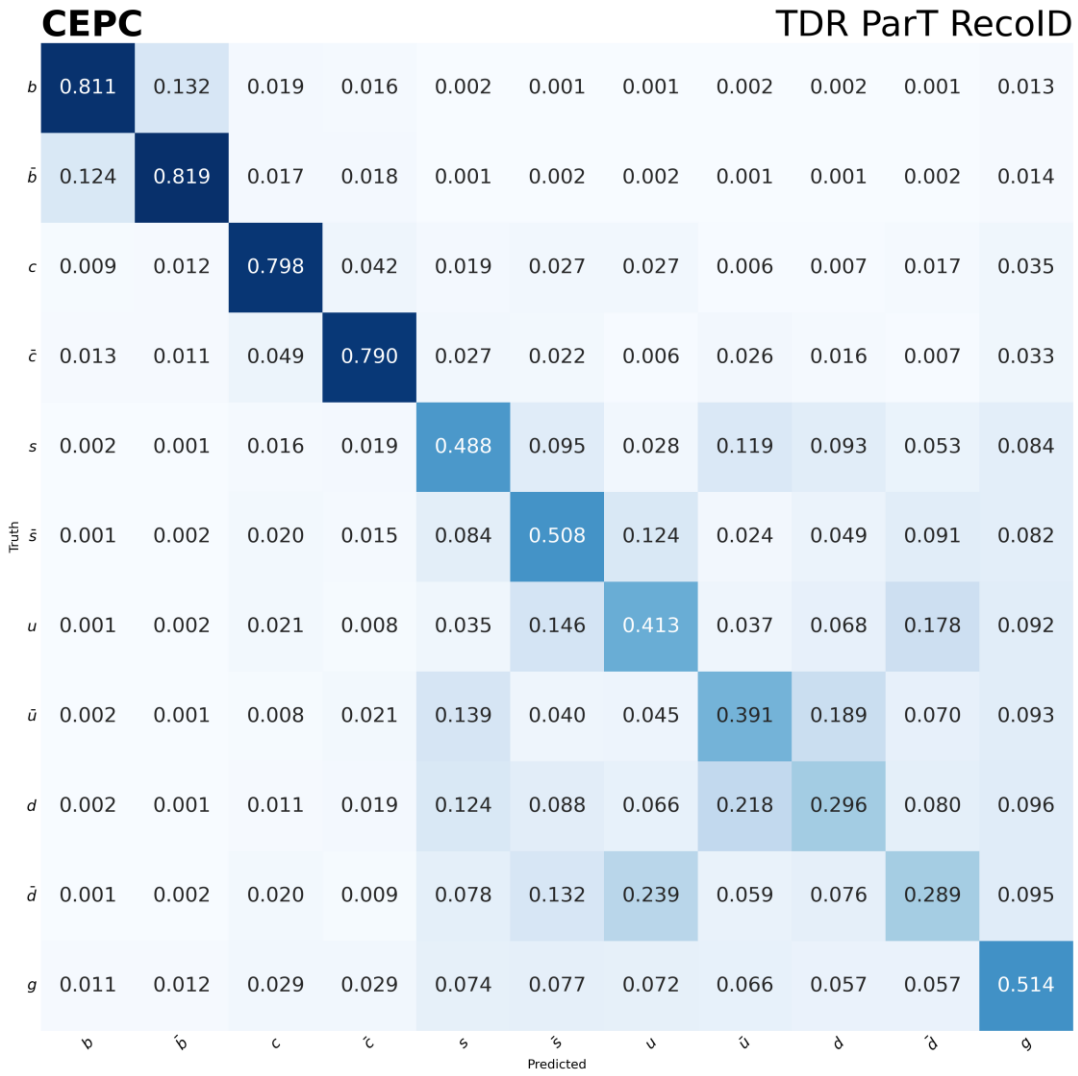


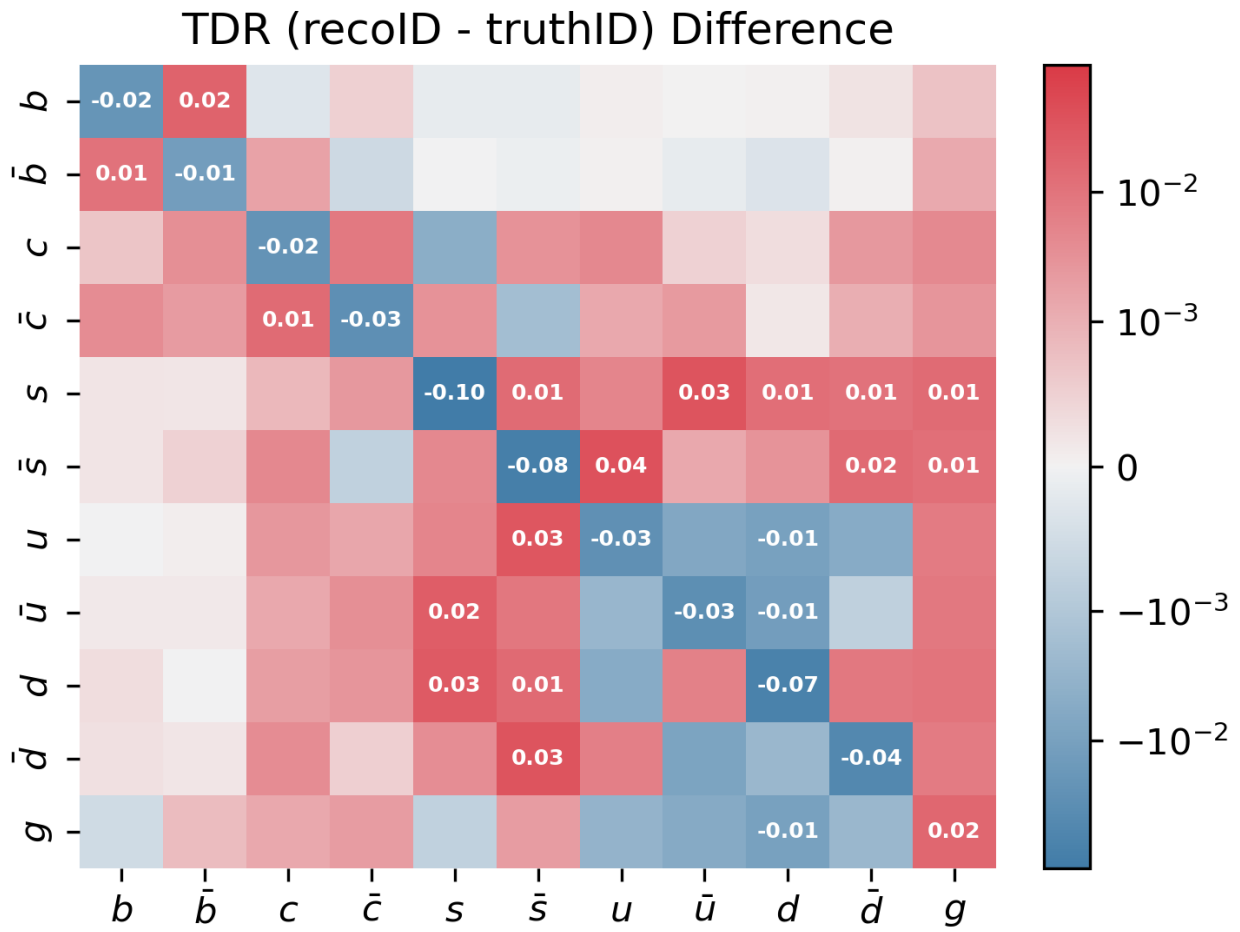
Difference in ParT/PN, in CDR:



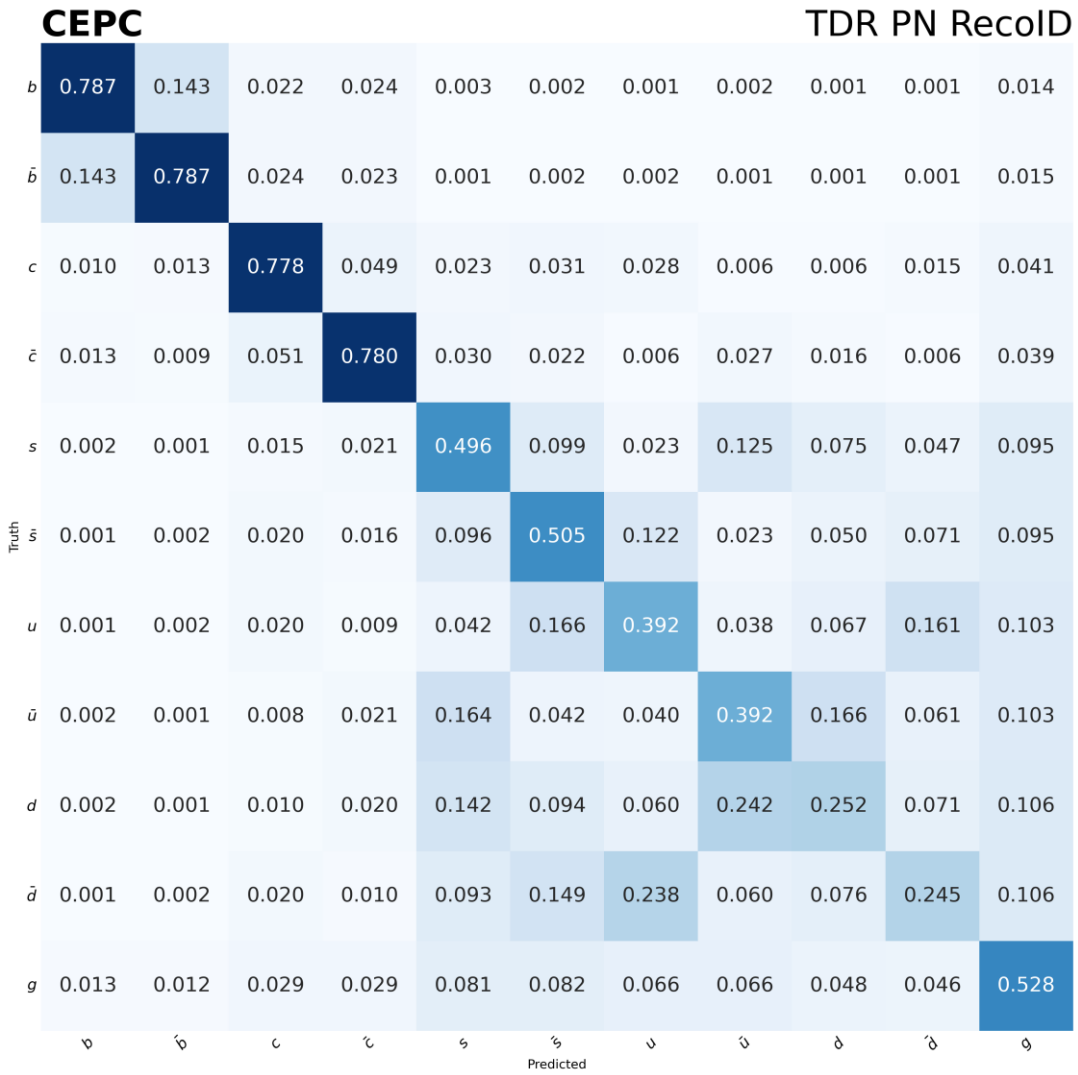


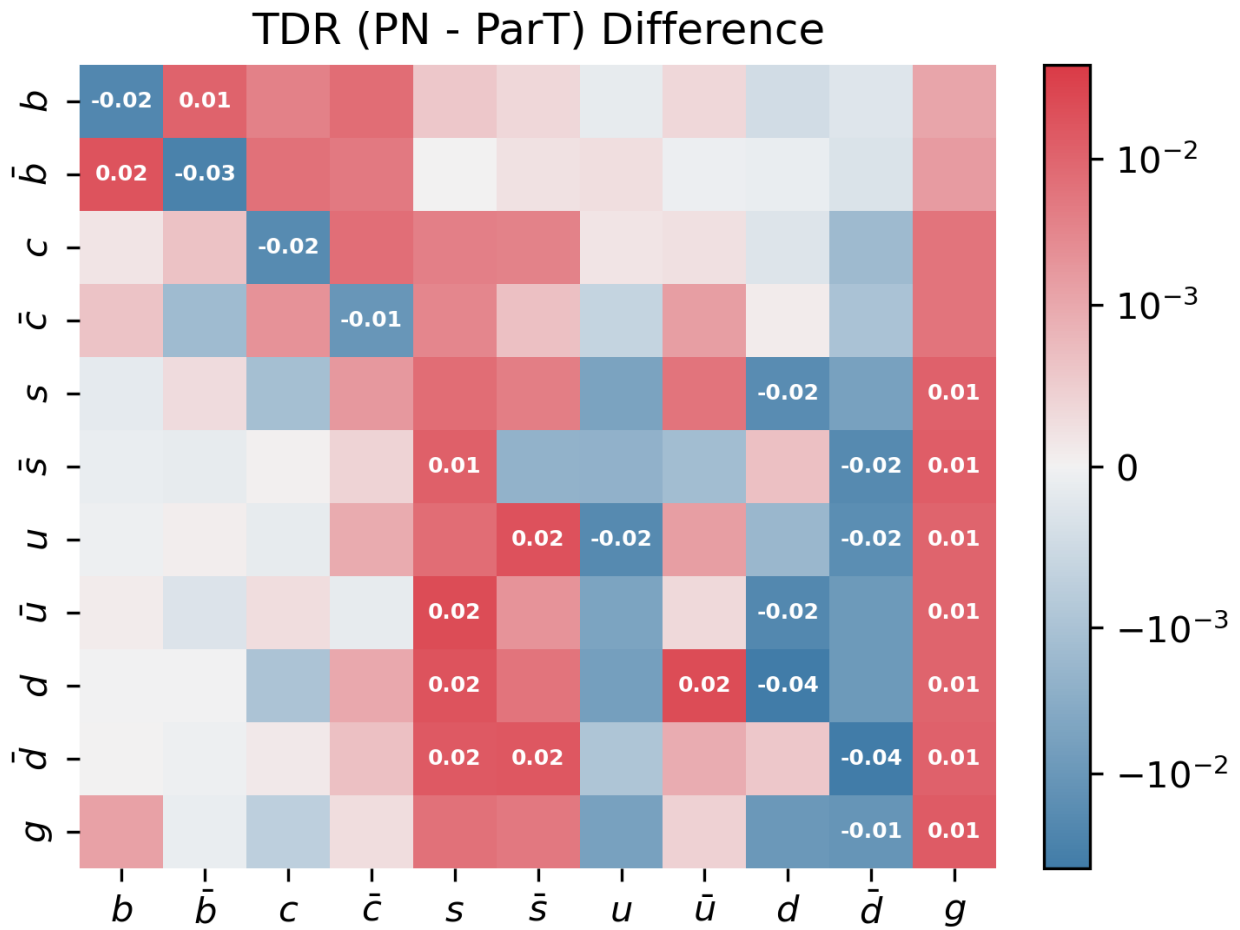
Difference in Reco ID/Perfect ID, TDR:





Difference in ParT/PN, in TDR:



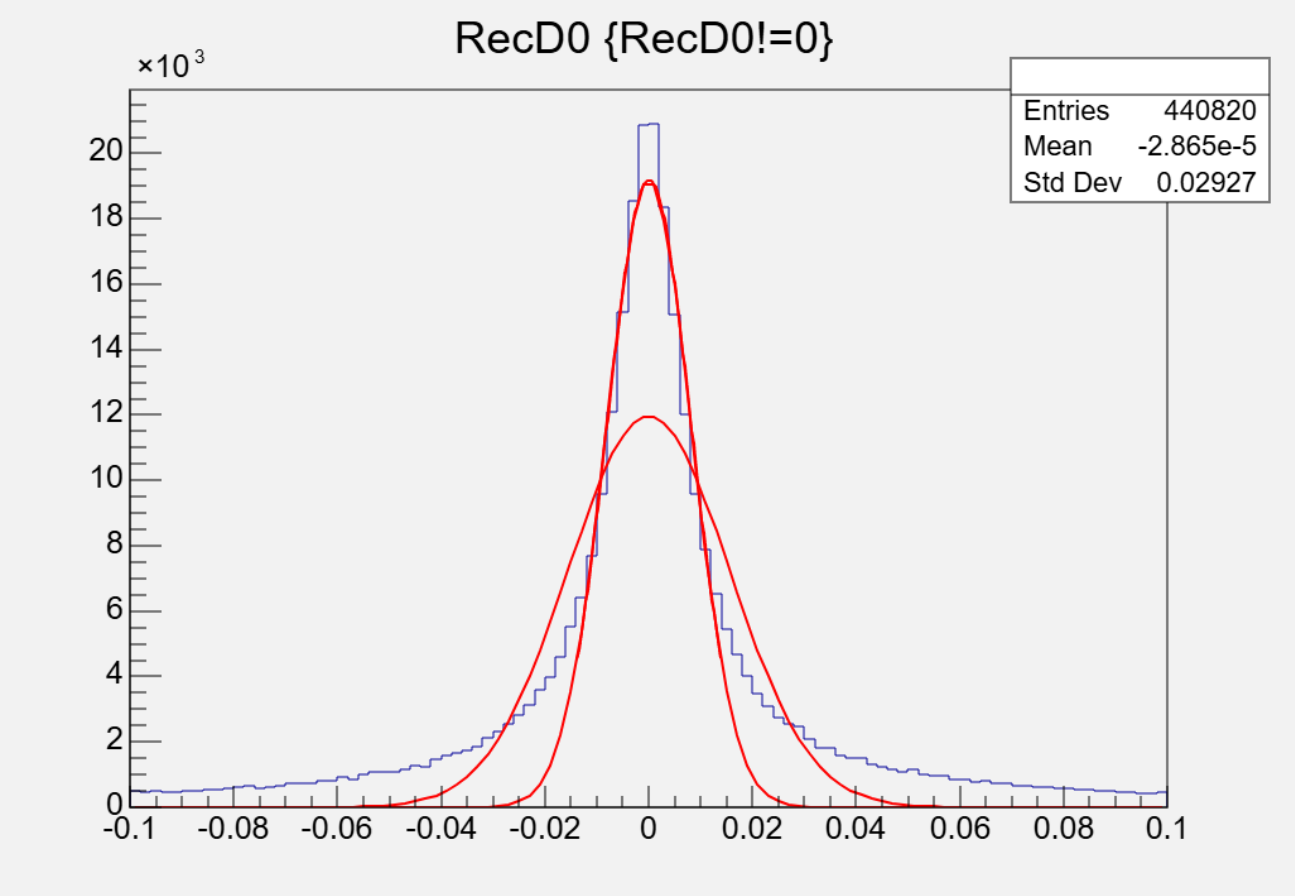


Conclusion for JOI comparison:

Use matrix metric as model performance benchmark.

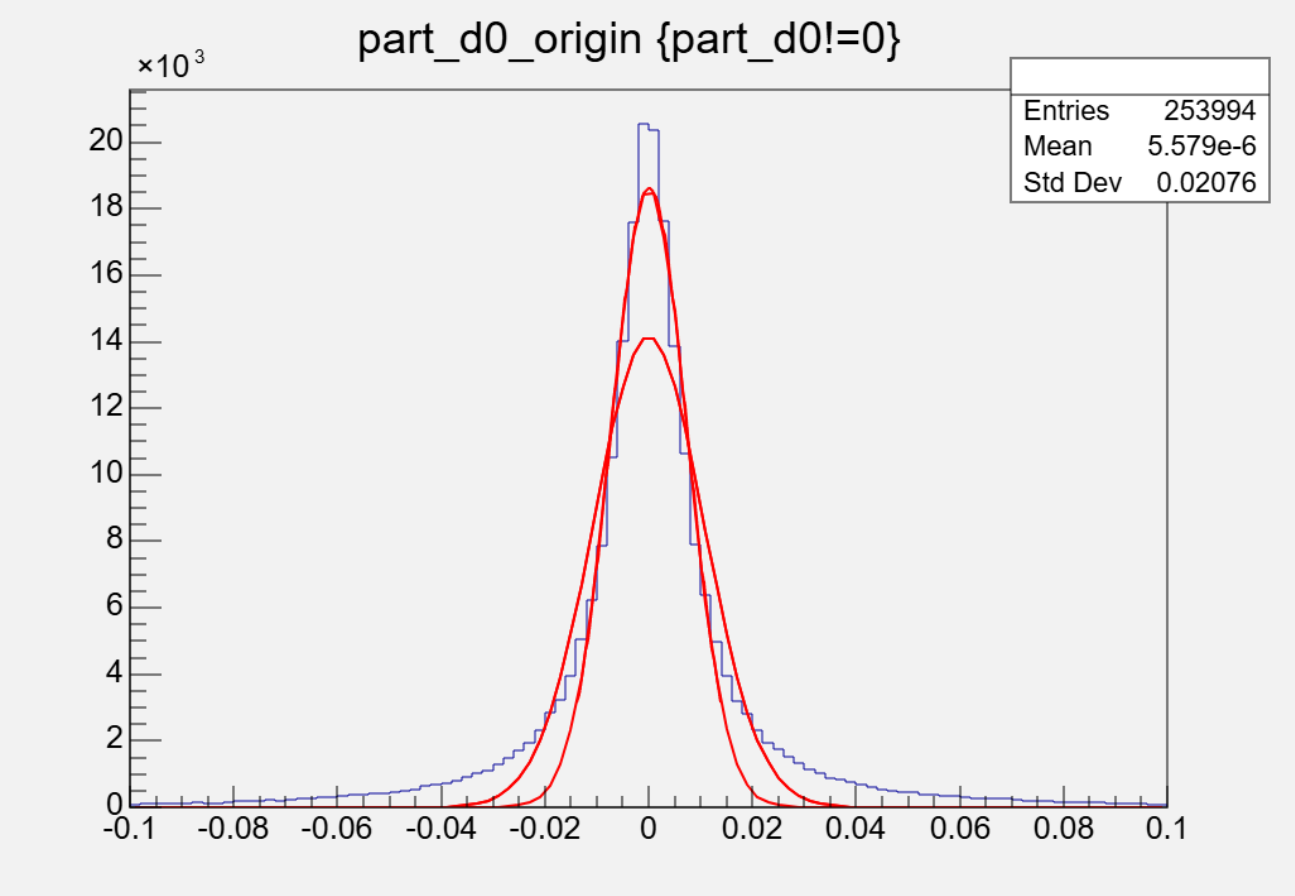
- ParT > PN in 3.0%
- TDR > CDR(Full) in 12.3%; (PRL result)
- TDR > CDR(Fast) in 7.1%;
- Truth ID > Reco ID in 6.7%

CDR Full D0



- Std: 29um
- Gauss (-0.1, 0.1): 15.53um
- Gauss(-0.014,0.014): 8.16um

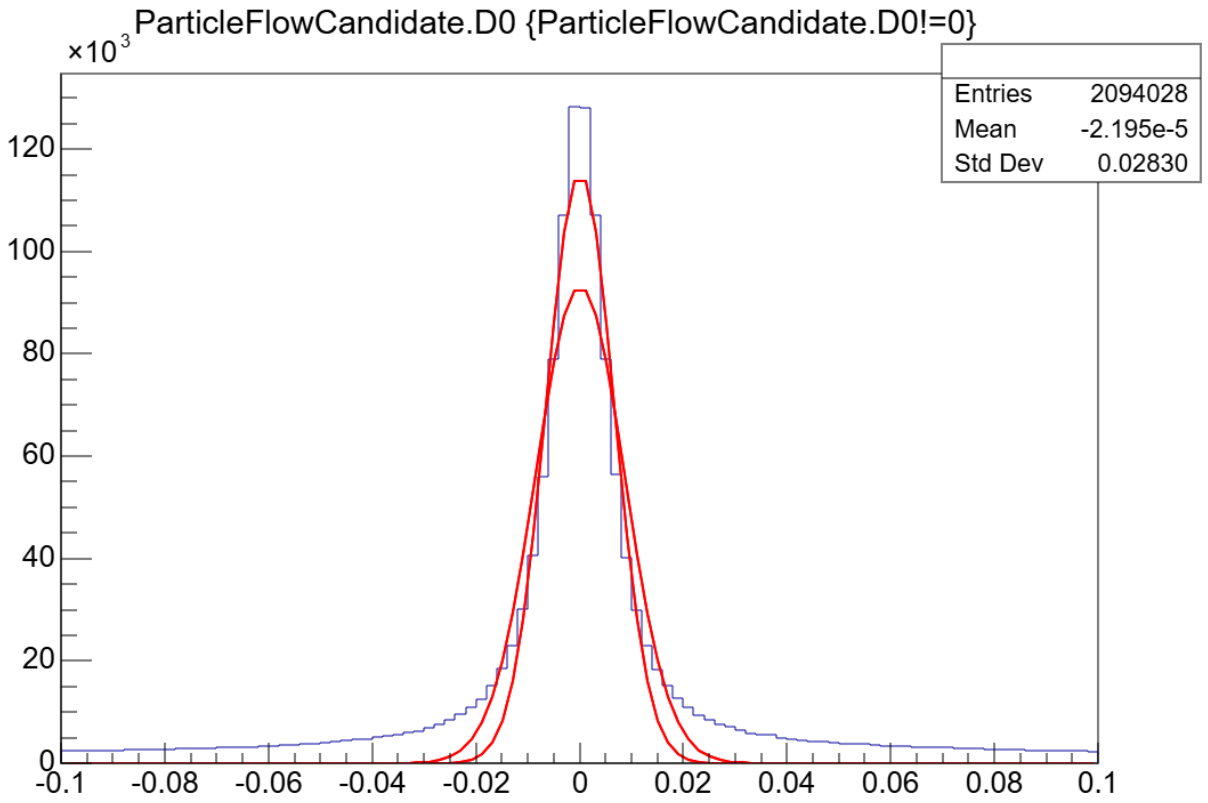
CDR Fast D0



- Std: 20um
- Gauss (-0.1, 0.1): 10.61um
- Gauss(-0.014,0.014): 7.36um

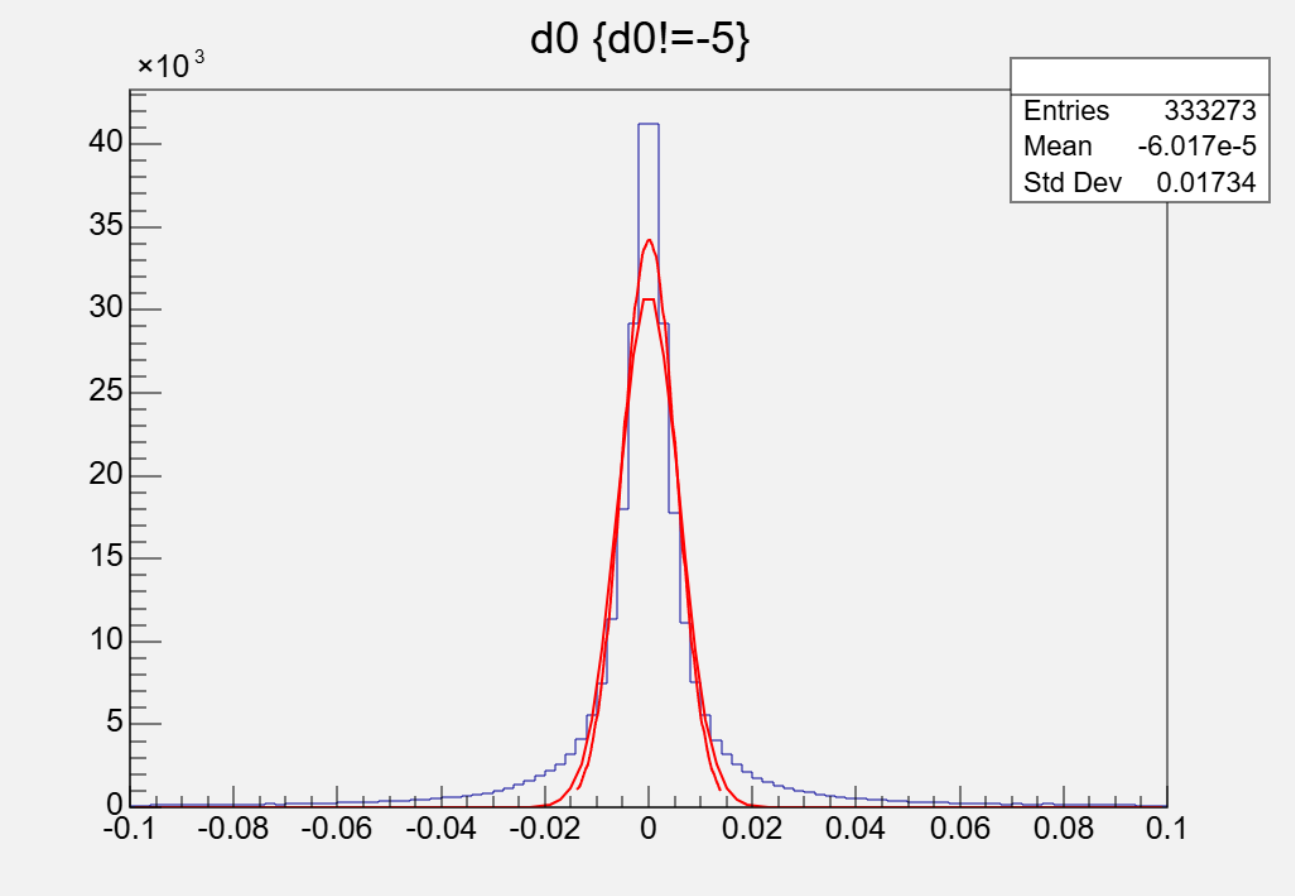
TDR Fast D0

https://code.ihep.ac.cn/zhangkl/delphes_cepc/-/blob/master/cards/delphes_card_CEPC_RefTDR.tcl?ref_type=heads Updated in 2025.01.



- Std: 28um
- Gauss (-0.1, 0.1): 8.56um
- Gauss(-0.014,0.014): 6.54um
- In Delphes vertex precision already set to 3um. Need better value to simulate the vertex behavior.

TDR Full D0



- Std: 17um
- Gauss (-0.1, 0.1): 5.83um
- Gauss(-0.014,0.014): 5.23um

Geometry conclusion

- Different geometry/simulation gives vertex 5-8um precision.
  - d0/z0 in TDR 2-3um corresponding to high energy tracks. For tracks in Jet (E<1GeV) 5um as benchmark.
- Overall: TDR Full > (TDR Fast) > CDR Fast > CDR Full.
- Vertex precision dominants in b jet tagging
  - Though different layout, B eff can be ~90% for all cases.

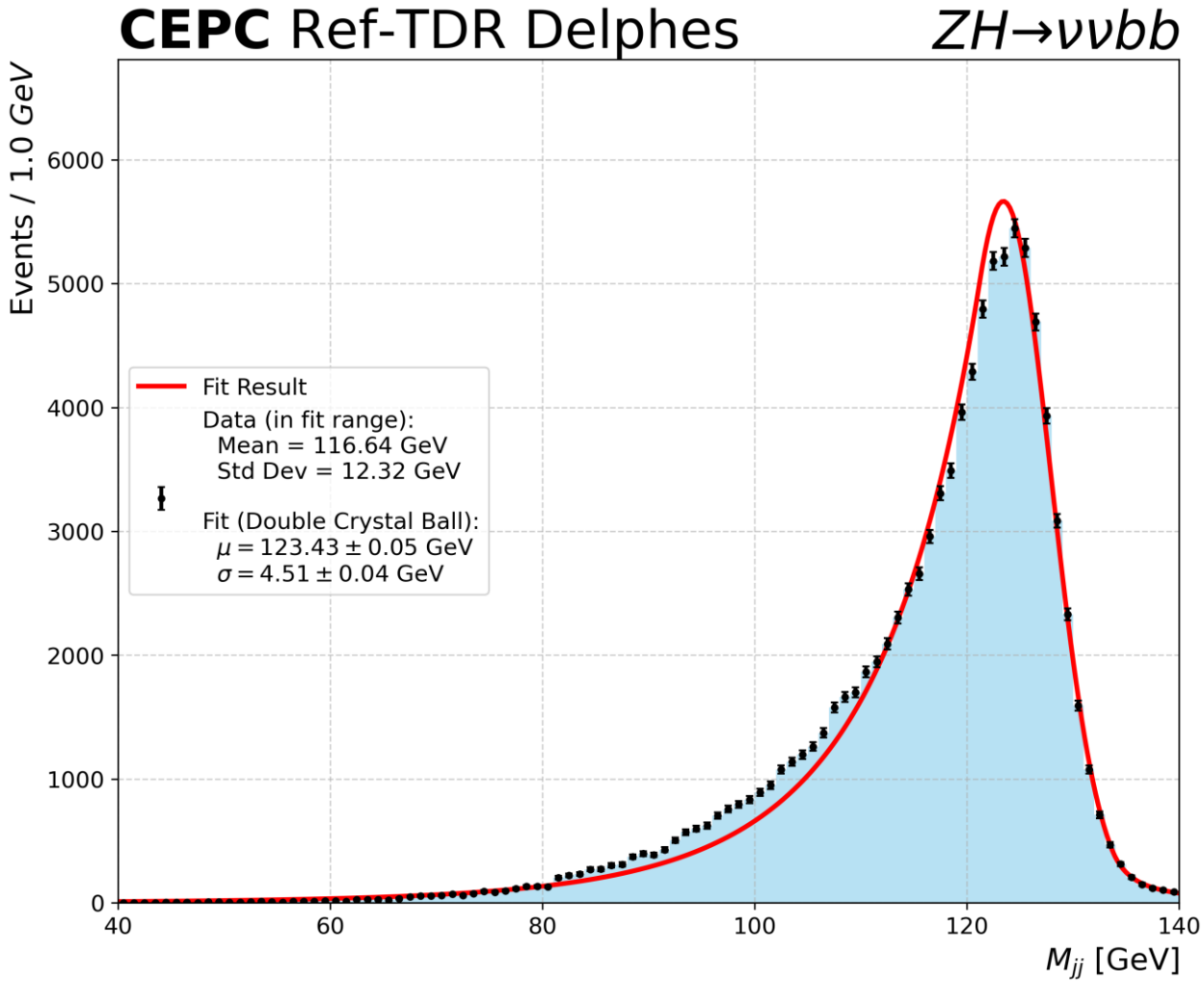
Geometry	Algo	PID	Simu	WP_B eff
TDR	ParT	Truth	full	94.5%
TDR	ParT	Reco	full	94.3%
TDR	PN	Reco	full	93.1%
CDR	ParT	Truth	fast	89.8%
CDR	PN	Truth	fast	88.0%
CDR	PN	Truth	full	89.9%

Delphes Vertex

Delphes with same dimension, material budget and better vertex (3um, full 5um) compared to full, (CEPCSW 25037)

```
# barrel  name     zmin     zmax   r        w (m)      X0        n_meas  th_up (rad) th_down (rad)    reso_up (m)   reso_down (m)  flag
# base geometry: https://code.ihep.ac.cn/cepc/CEPCSW/-/blob/master/Detector/DetCRD/compact/TDR_o1_v01/TDR_Dimensions_v01_01.xml
# For material budget, use 0.00162X0(pipe), 0.000834(vtx1) and , 0.0009(2), 0.00676(3) for scaling. (changing the thickness of Silicon).
# 140um silicon ~0.0015X0.
# Pipe inner r 10mm;
1       PIPE  -3.0   3.0     0.010         0.000151      0.0937     0     0     0          0             0           0
# Vertex geometry defined in: https://code.ihep.ac.cn/cepc/CEPCSW/-/blob/master/Detector/DetCRD/compact/CRD_common_v02/VXD_Composite_v01_02.xml
1       VTX1A       -0.255       0.255         0.011063 0.0000389 0.0937     2     0     1.5708     3e-006        3e-006      1
1       VTX1B       -0.255       0.255         0.016562 0.0000389 0.0937     2     0     1.5708     3e-006        3e-006      1
1       VTX2A       -0.255  0.255       0.022061 0.000042  0.0937     2     0     1.5708     3e-006        3e-006      1
1       VTX2B       -0.255  0.255       0.027561 0.000042  0.0937     2     0     1.5708     3e-006        3e-006      1
1       VTX3A       -0.255  0.255       0.043792 0.000316  0.0937     2     0     1.5708     3e-006        3e-006      1
1       VTX3B       -0.255  0.255       0.044543 0.000316  0.0937     2     0     1.5708     3e-006        3e-006      1
# Shell as Carbon fiber-reinforced plastic(CFRP). L=29.1cm.    0.000
1       SHELL       -0.51 0.51  0.07      0.0025  0.291       0     0     0          0             0           0
# vertex end at 72.5mm.
# Silicon Inner Tracker (SIT for barrel and EST, endcap silicon tracker for Endcap).
# Silicon tracker precision ~10um.
# 3 SIT layers 1.5% X0. ->1.4mm Si. Each 470mm.
1       SIT01       -0.5005      0.5005        0.24      0.00047    0.0937     2     0     1.5708     1e-005        9e-005      1
1       SIT02       -0.715       0.715          0.35     0.00047    0.0937     2     0     1.5708     1e-005        9e-005      1
1       SIT03       -1.001       1.001          0.57     0.00047    0.0937     2     0     1.5708     1e-005        9e-005      1
```

Current TDR Fast



Current sigma: 4.51 GeV; BMR: 3.65%.

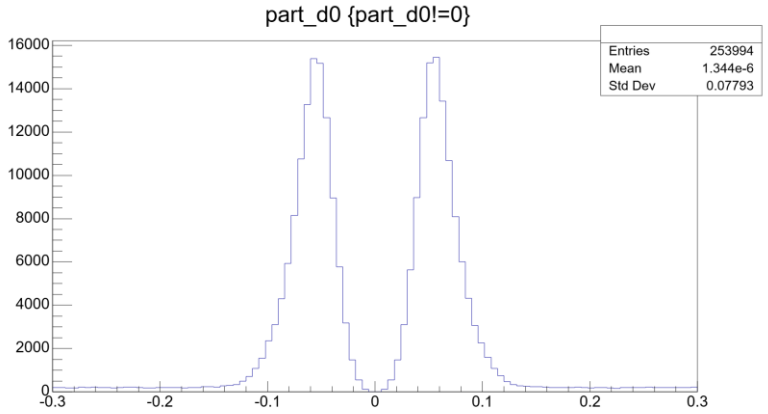
So Jet performance roughly consistent (Not exactly same) with current TDR full simulation.

For further agreement: need dig into delphes source code.

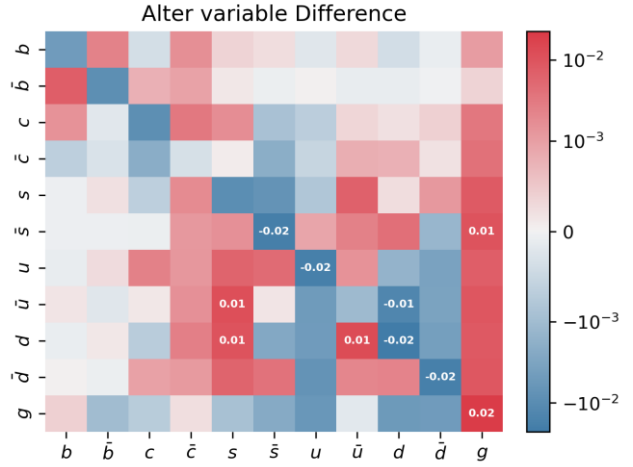
Variables Feature Engineering

- D0, Z0 transformation in previous work

```
d0 = 0.2 * ((d0 > 0) ? 1 : -1) * std::pow(std::abs(d0), 0.25);
d0err = 0.2 * ((d0err > 0) ? 1 : -1) * std::pow(std::abs(d0err), 0.25);
dz = 0.2 * ((dz > 0) ? 1 : -1) * std::pow(std::abs(dz), 0.25);
dzerr = 0.2 * ((dzerr > 0) ? 1 : -1) * std::pow(std::abs(dzerr), 0.25);
```



- Compared to tanh(d0), this proves to reduce 2%.
  - RecoID (0.5562) use same variable setting as YF: 0.546.



JOI with Data Scale size

- Same Variable, Hyperparameter, Validation/Test Set
- 11 categories, each $10^{2,3,4,5\ldots}$

Each Category Scale	Metric
10^2	0.2830
10^3	0.3939
10^4	0.4484
10^5	0.5137
$8 * 10^5$	0.5562

Feature Engineering

- Information like Calo energy deposition, only in Full simu, not in Fast simu.
- Including serval features, metric: 0.576.
  - Asymmetric response for b/bbar, s/sbar.
  - $uds/\ \bar{u}\bar{d}\bar{s}$ distribution not agree with TruthID.
  - Need further check