CVN Status

Nitish Nayak

20th Nov 2023

Introduction

- Retraining HD & VD with latest samples from production
- 4 Trainings : (HD, VD) x (FHC, RHC) :
  - Trainings are all done
  - Weight files : /dune/app/users/bnayak/larsoft_cvn2023/interactive/model_trainings/*/*.pb
  - Ready to be integrated
- Training cuts :
  - Require true neutrino vertex to be inside FV
  - Total nhits > 100
- Number of training samples :
  - HD-FHC : 2785330
  - VD-FHC : 3733161
  - HD-RHC : 2942065
  - VD-RHC : 3352880

```
bool fApplyFidVol = true;
bool isFid = true;
// If outside the fiducial volume don't waste any time filling other variables
if(fApplyFidVol){
  // Get the interaction vertex from the end point of the neutrino. This is
  // because the start point of the lepton doesn't make sense for taus as they
  // are decayed by the generator and not GEANT
  TVector3 vtx = true_neutrino.Nu().EndPosition().Vect();
  if(fIsVD)
    isFid = (fabs(vtx.X())<300 && fabs(vtx.Y())<680 && vtx.Z()>40 && vtx.Z()<850); // vd
  else
    isFid = (fabs(vtx.X())<310 && fabs(vtx.Y())<550 && vtx.Z()>50 && vtx.Z()<1244); // hd
  if(!isFid) return;
}
```

Training Statistics

FHC-Nueswap

```
/dune/app/users/bnayak/larsoft_cvn2023/interactive/model_trainings bnayak@dunegpvm05
❯ defs=("hd" "vd"); for d in "${defs[@]}"; do box $d; samweb list-files --summary 'defname: bnayak_cvn'"$d"'_nue'; done
======
= hd =
======
File count:     20346
Total size:     33513533259186
Event count:    2034501
======
= vd =
======
File count:     23233
Total size:     29217549601523
Event count:    2323102
```

RHC-Nueswap

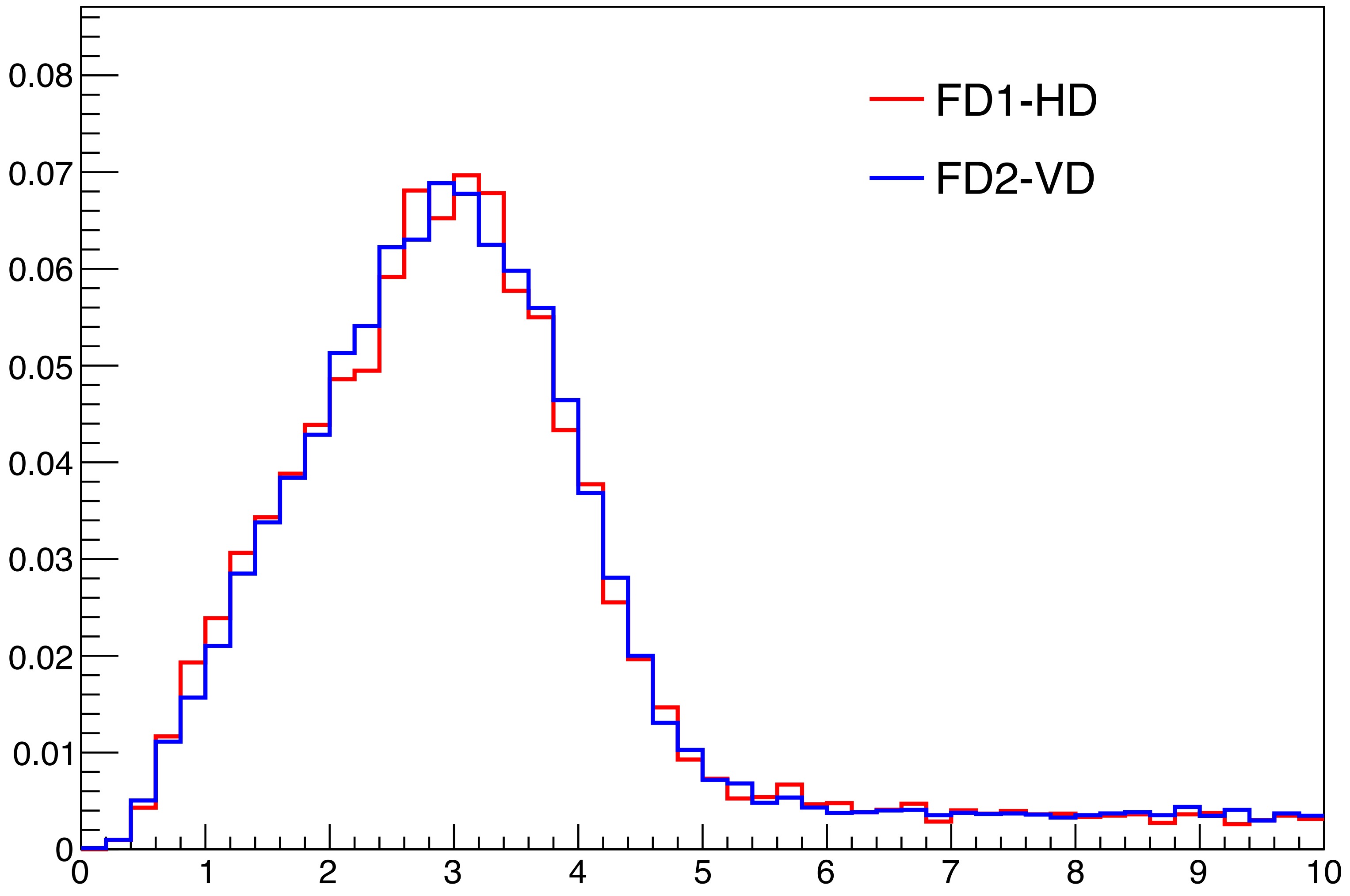
```
/dune/app/users/bnayak/larsoft_cvn2023/interactive/model_trainings bnayak@dunegpvm05
❯ defs=("hd" "vd"); for d in "${defs[@]}"; do box $d; samweb list-files --summary 'defname: bnayak_cvn'"$d"'_anue'; done
======
= hd =
======
File count:     19588
Total size:     40009768659437
Event count:    1958800
======
= vd =
======
File count:     20653
Total size:     33333291053807
Event count:    2065102
```

- FHC-VD started off with ~15% more generated statistics
  - Final training sample ~30% higher
- RHC-VD started off ~similar generated statistics
  - Final training sample ~15% higher
- VD preselections fetching ~15% extra events

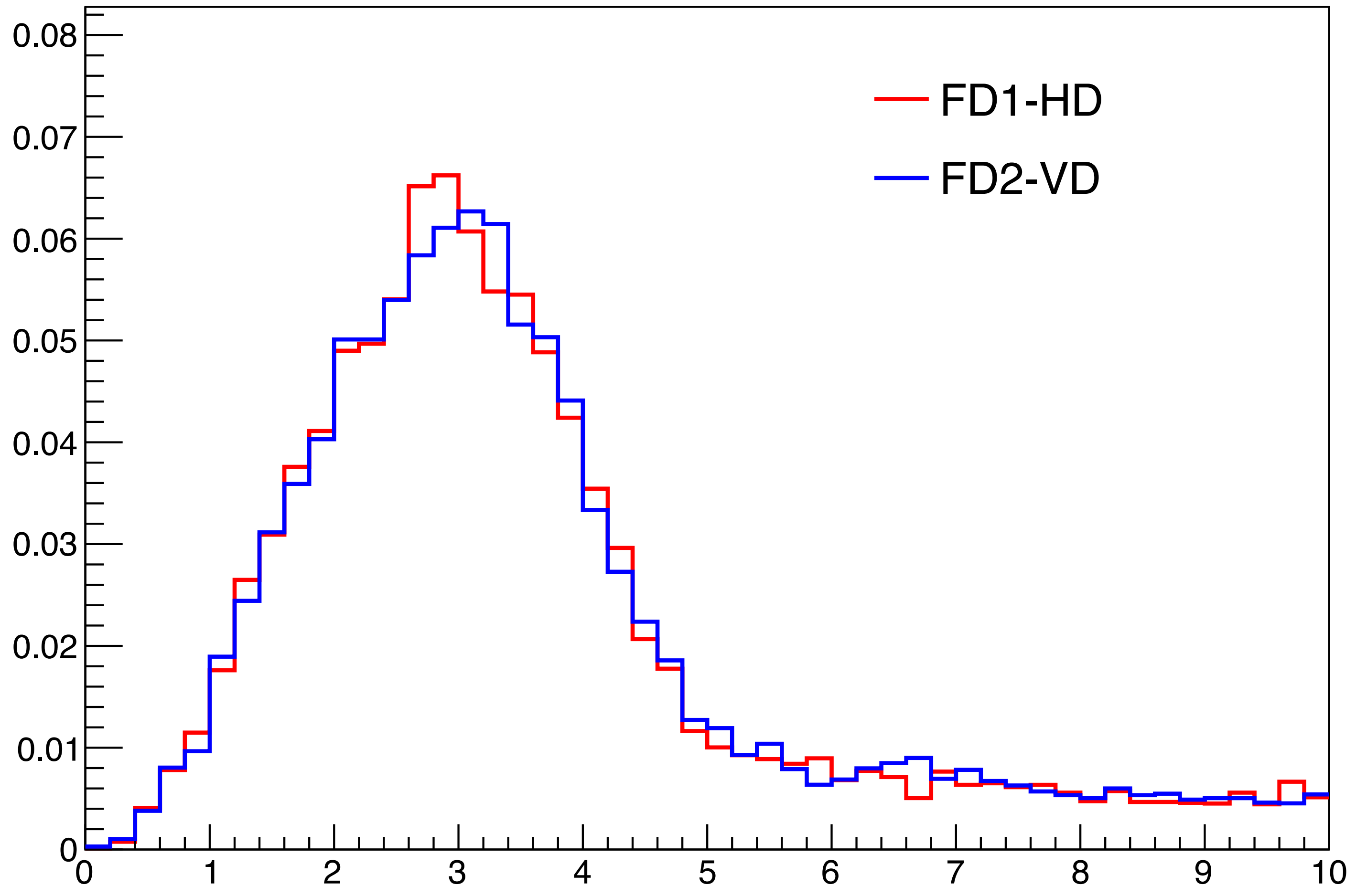
Training Statistics

Neutrino Energy

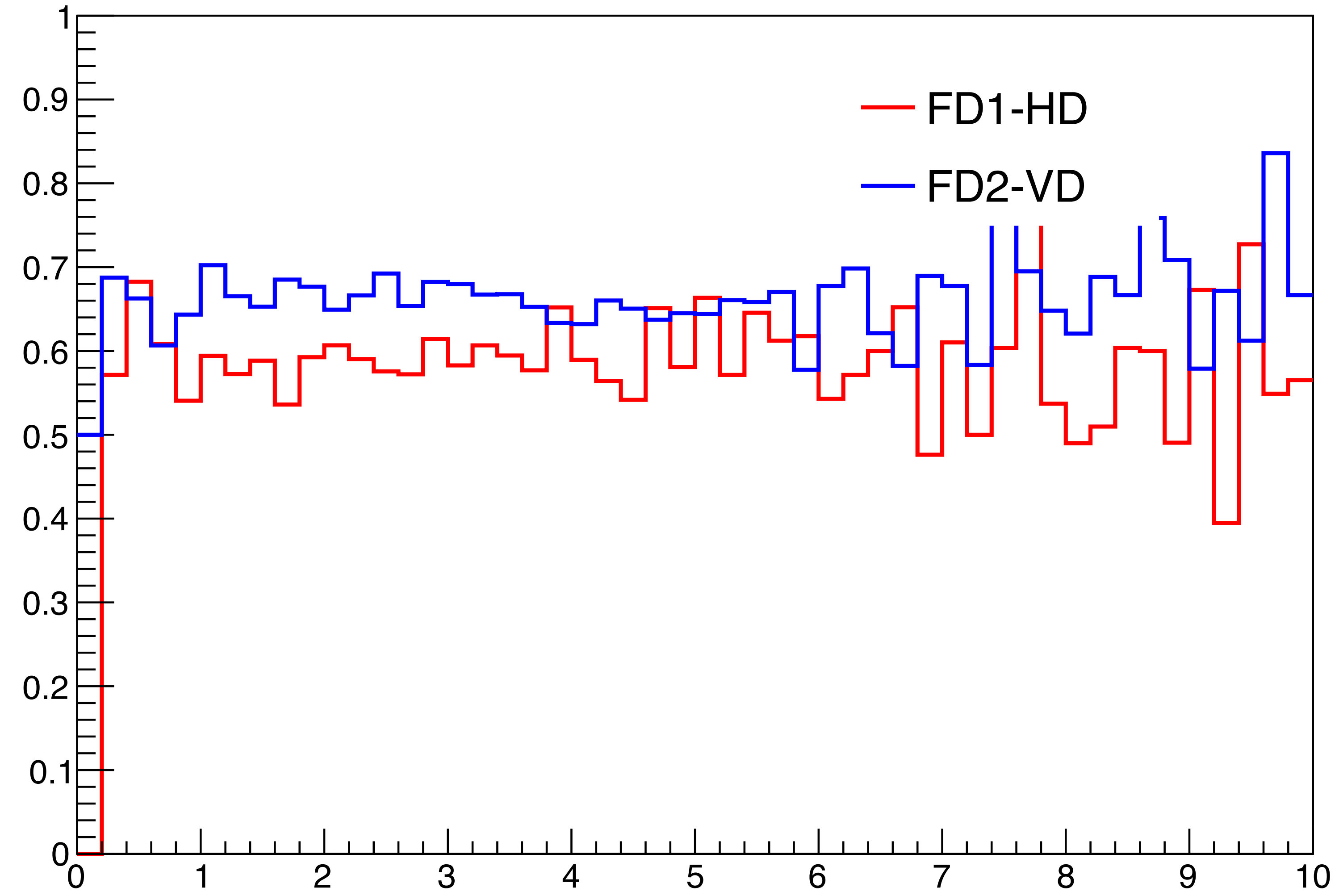
FHC-Nominal (Out of the Box)



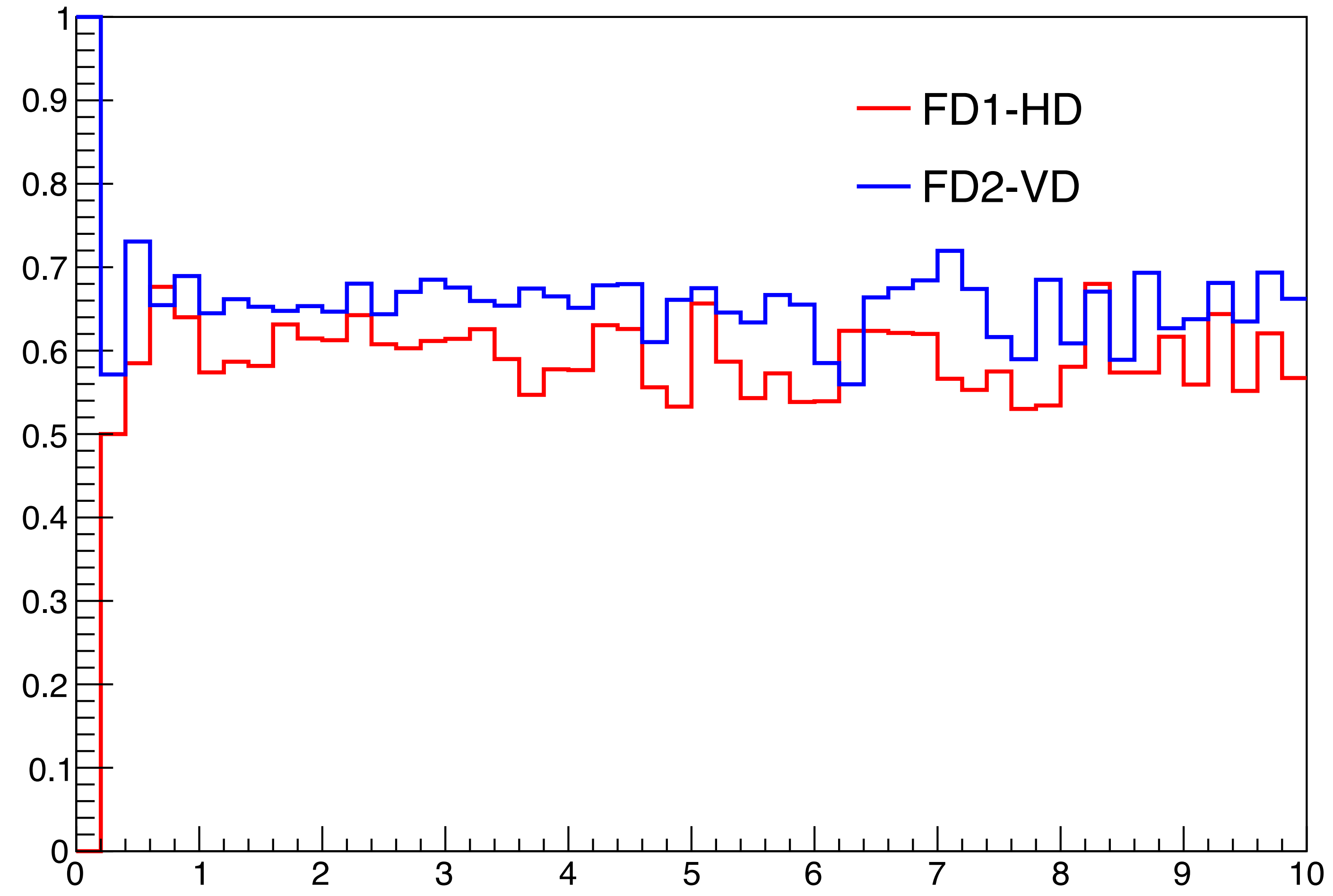
RHC-Nominal (Out of the Box)



Efficiency after applying FV cuts



Efficiency after applying FV cuts

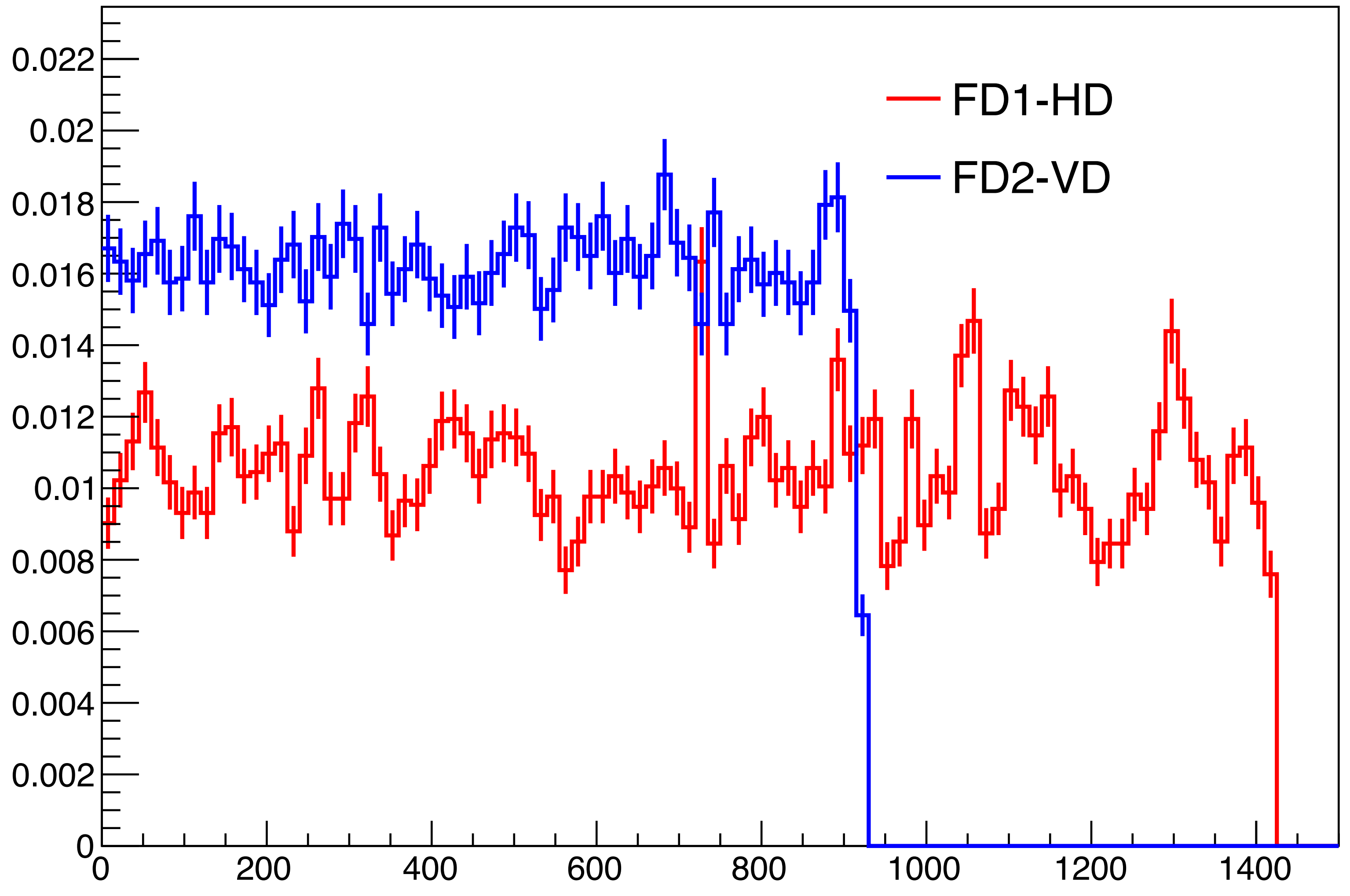


- Seems like VD has more fraction of events simulated within FV?
- Explains why we end up getting more VD training statistics

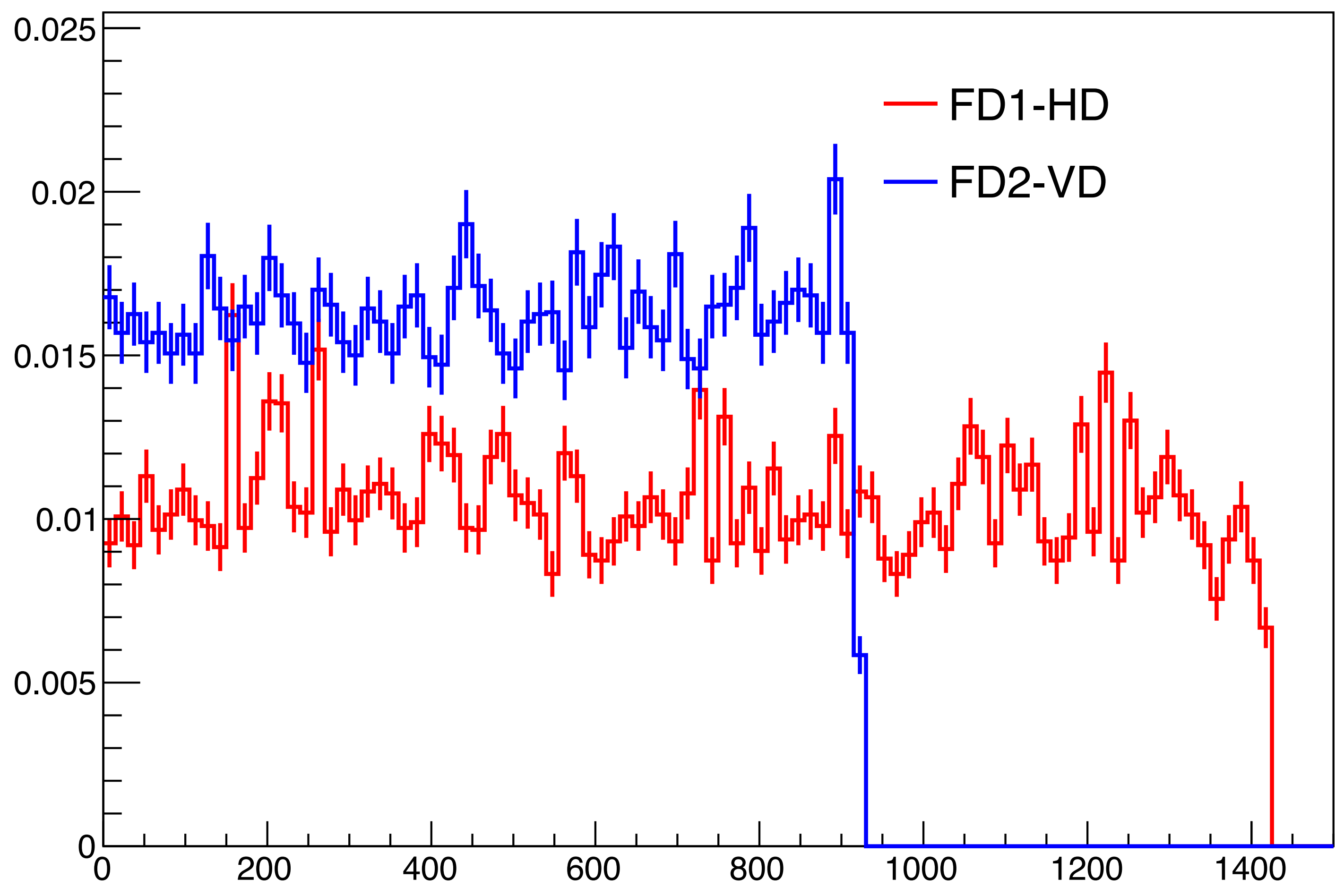
Training Statistics

Vtx Z

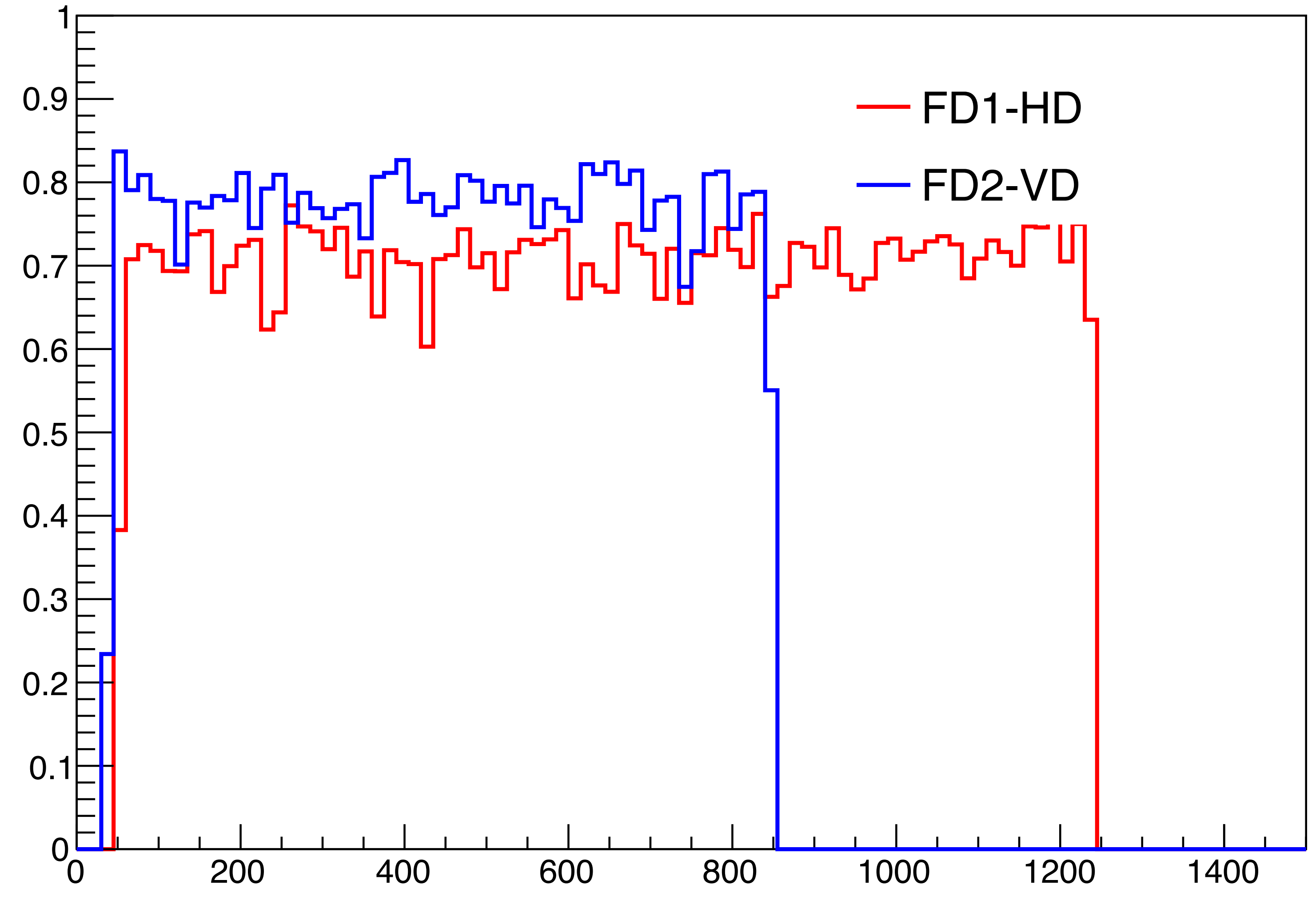
FHC-Nominal (Out of the Box)



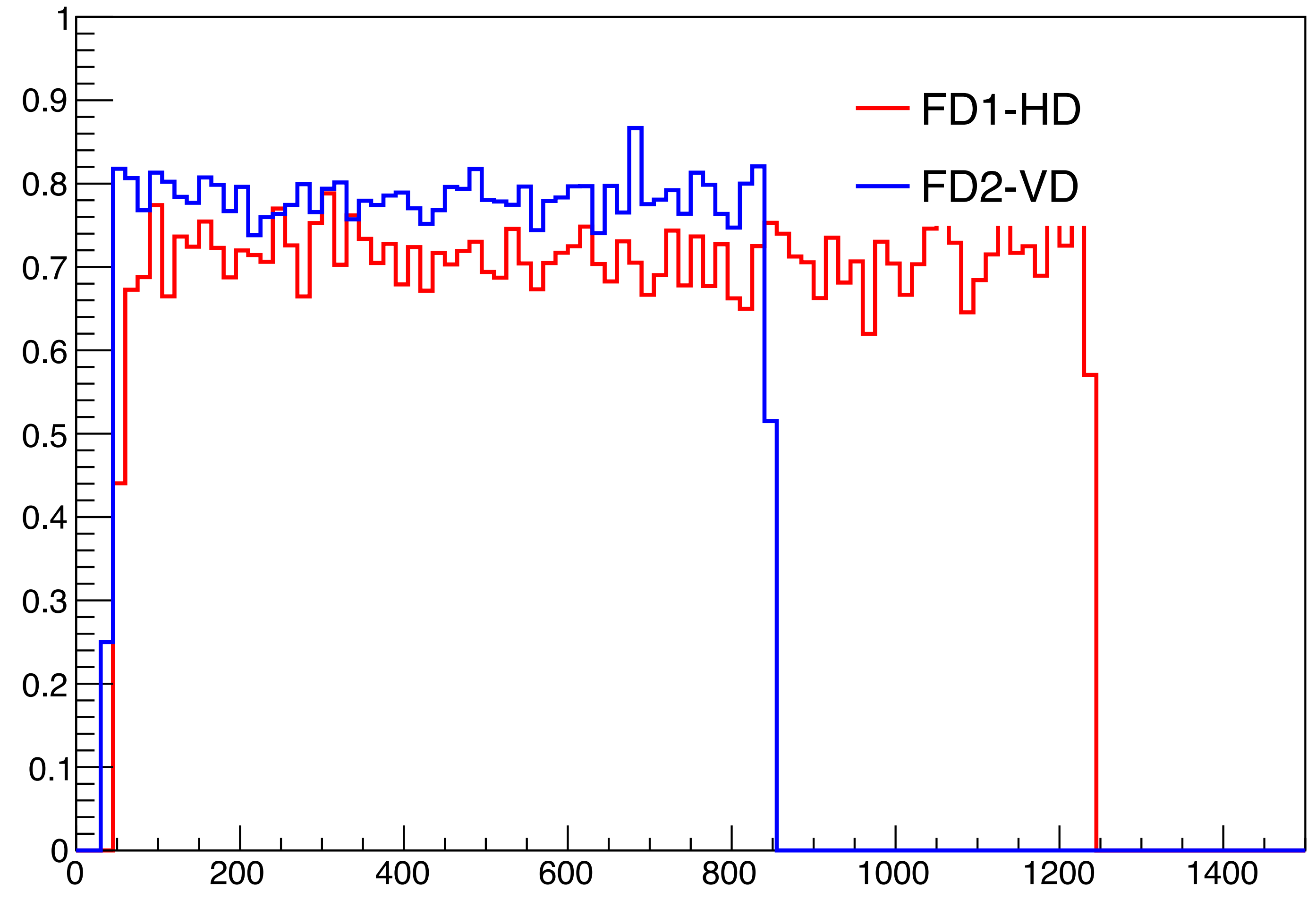
RHC-Nominal (Out of the Box)



Efficiency after applying FV cuts



Efficiency after applying FV cuts

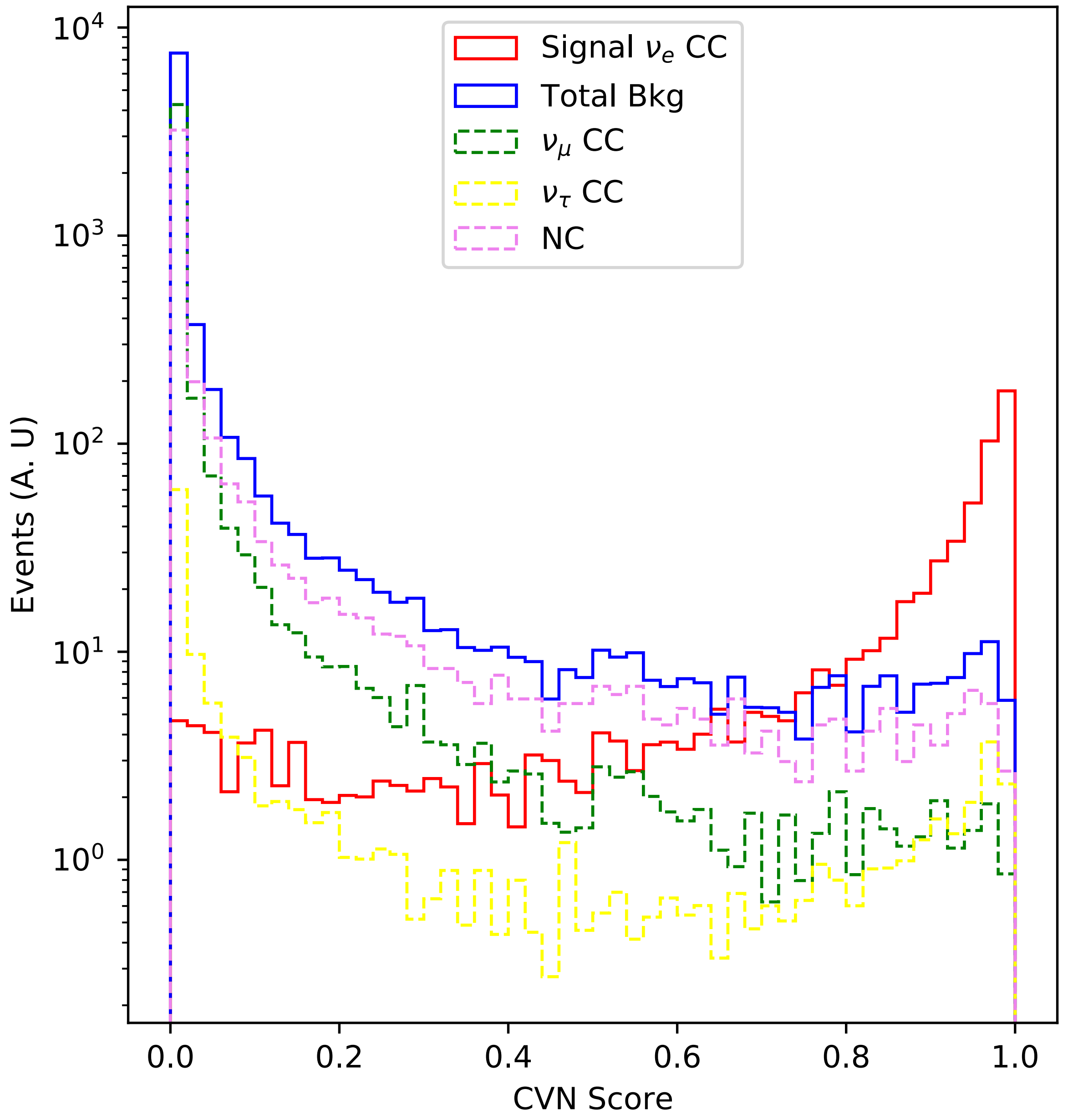


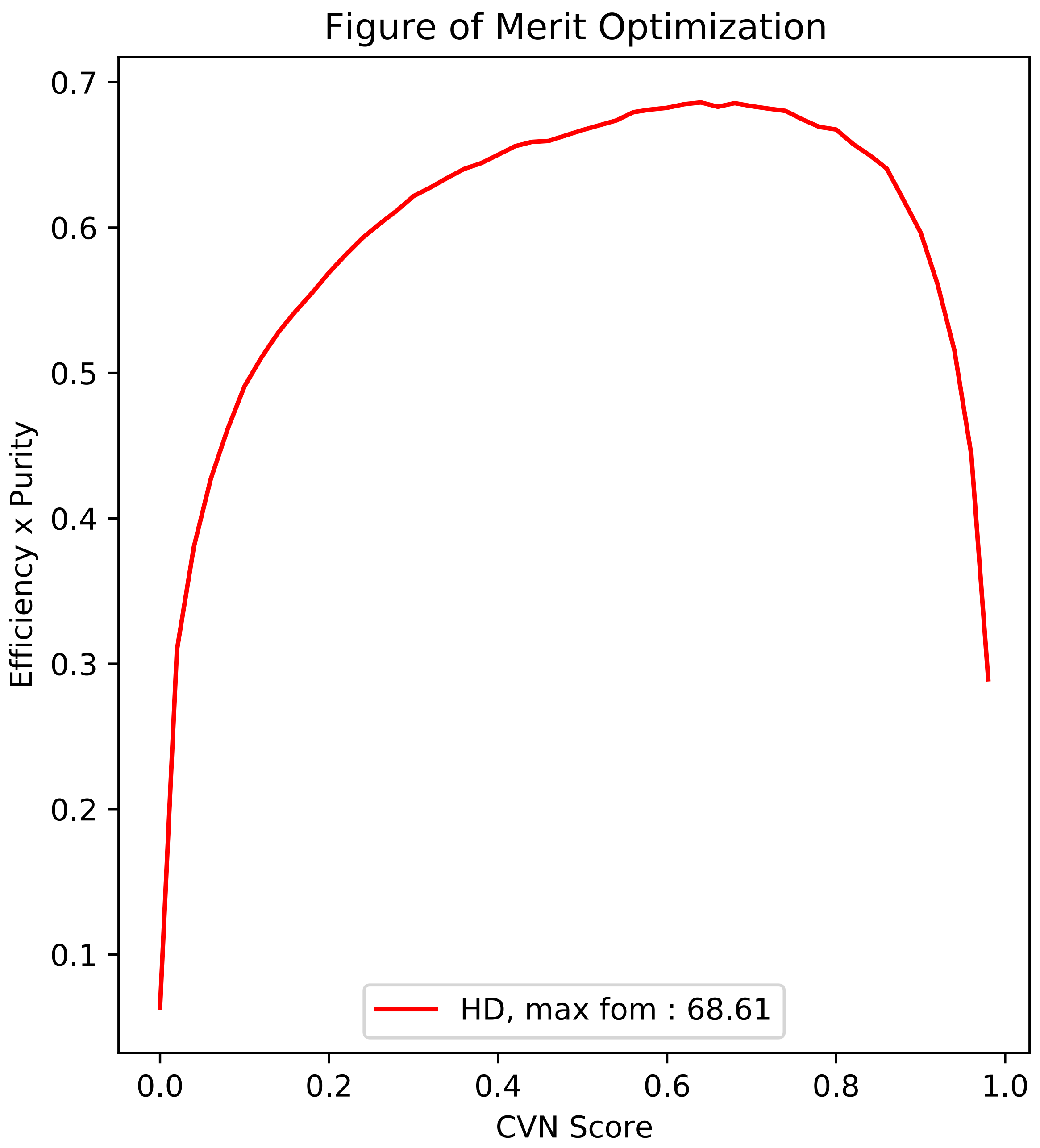
- Seems like VD has more fraction of events simulated within FV?
- Explains why we end up getting more VD training statistics

Training Results

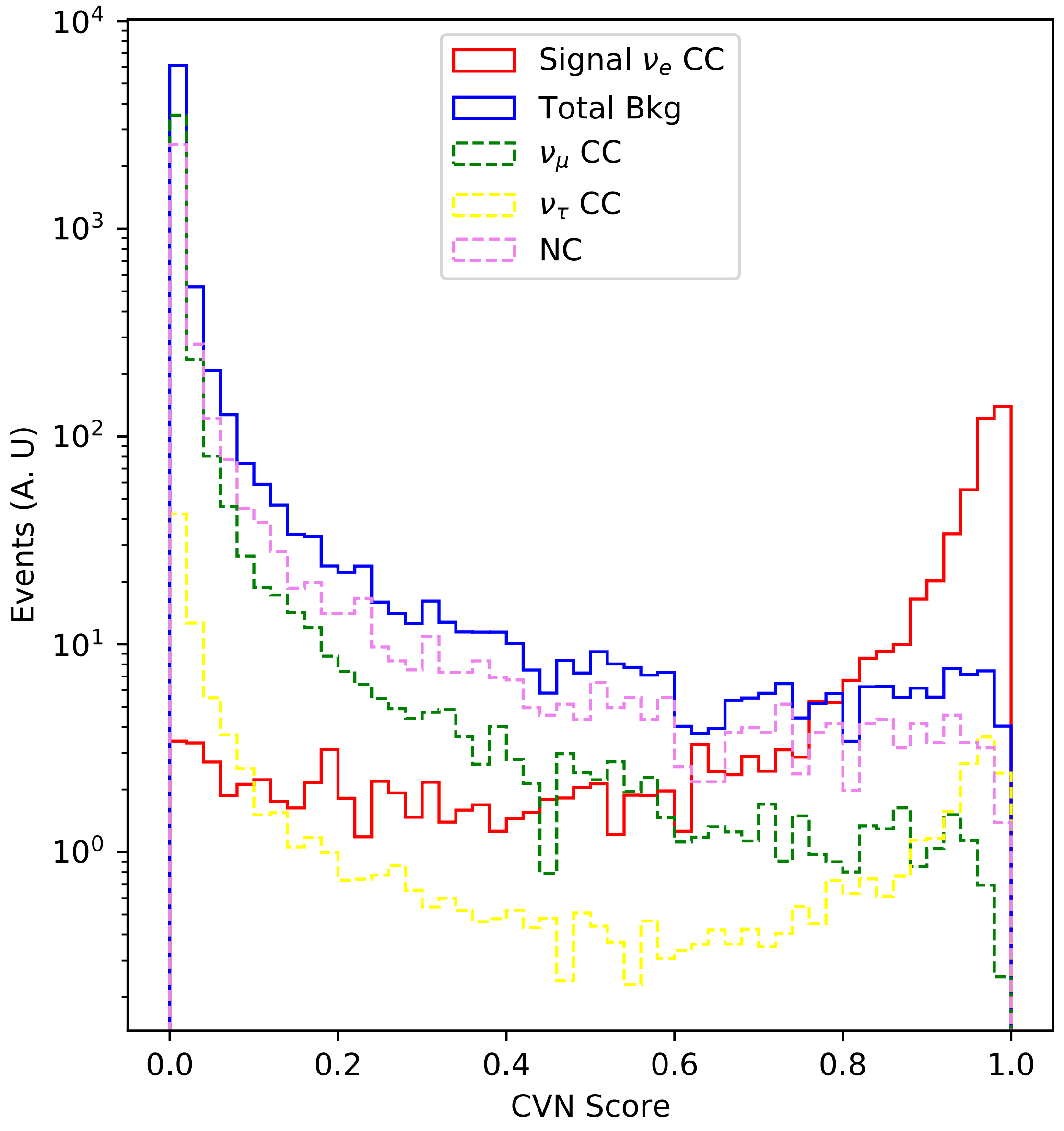
FHC (Shown previously)

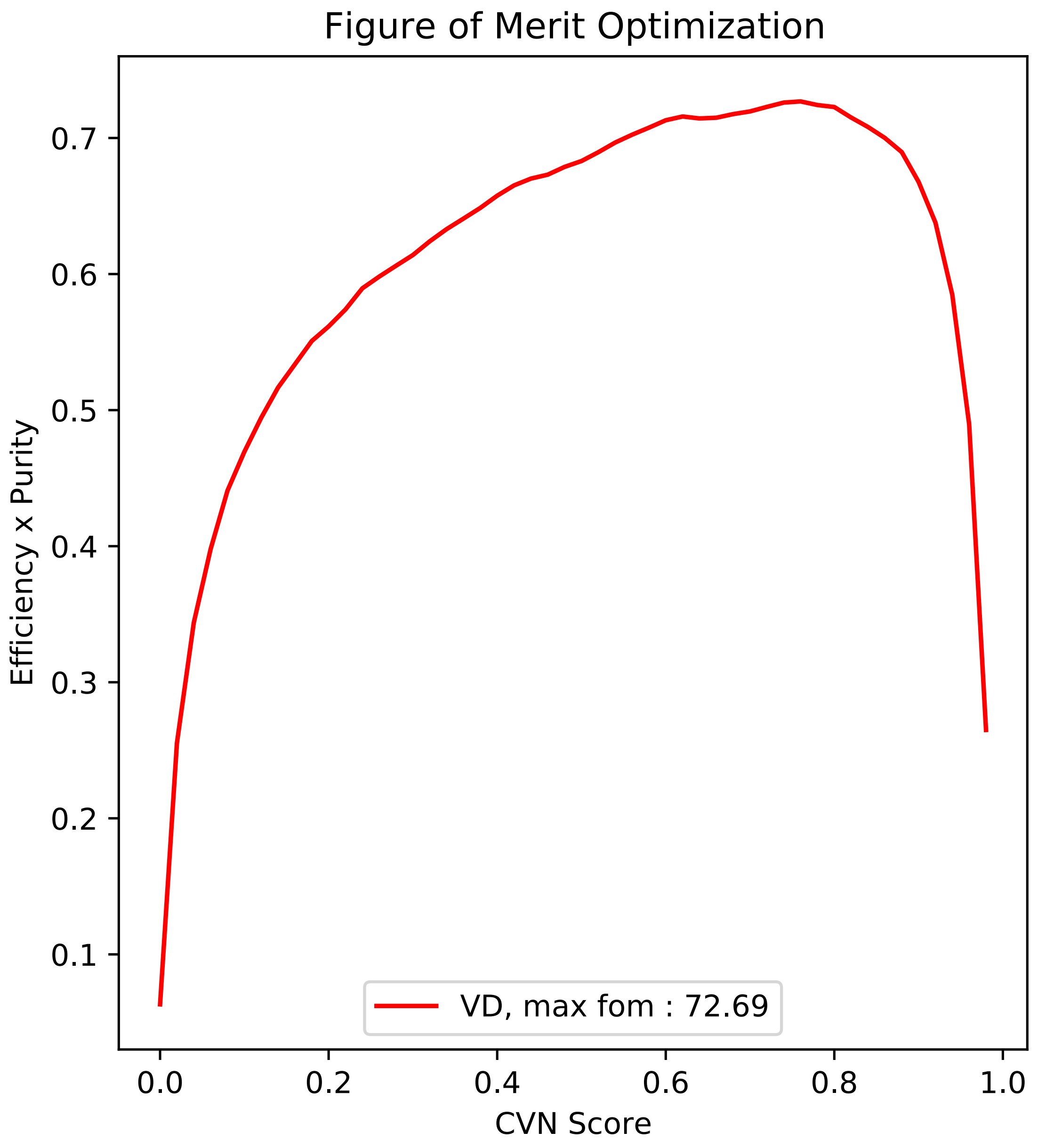
HD





VD

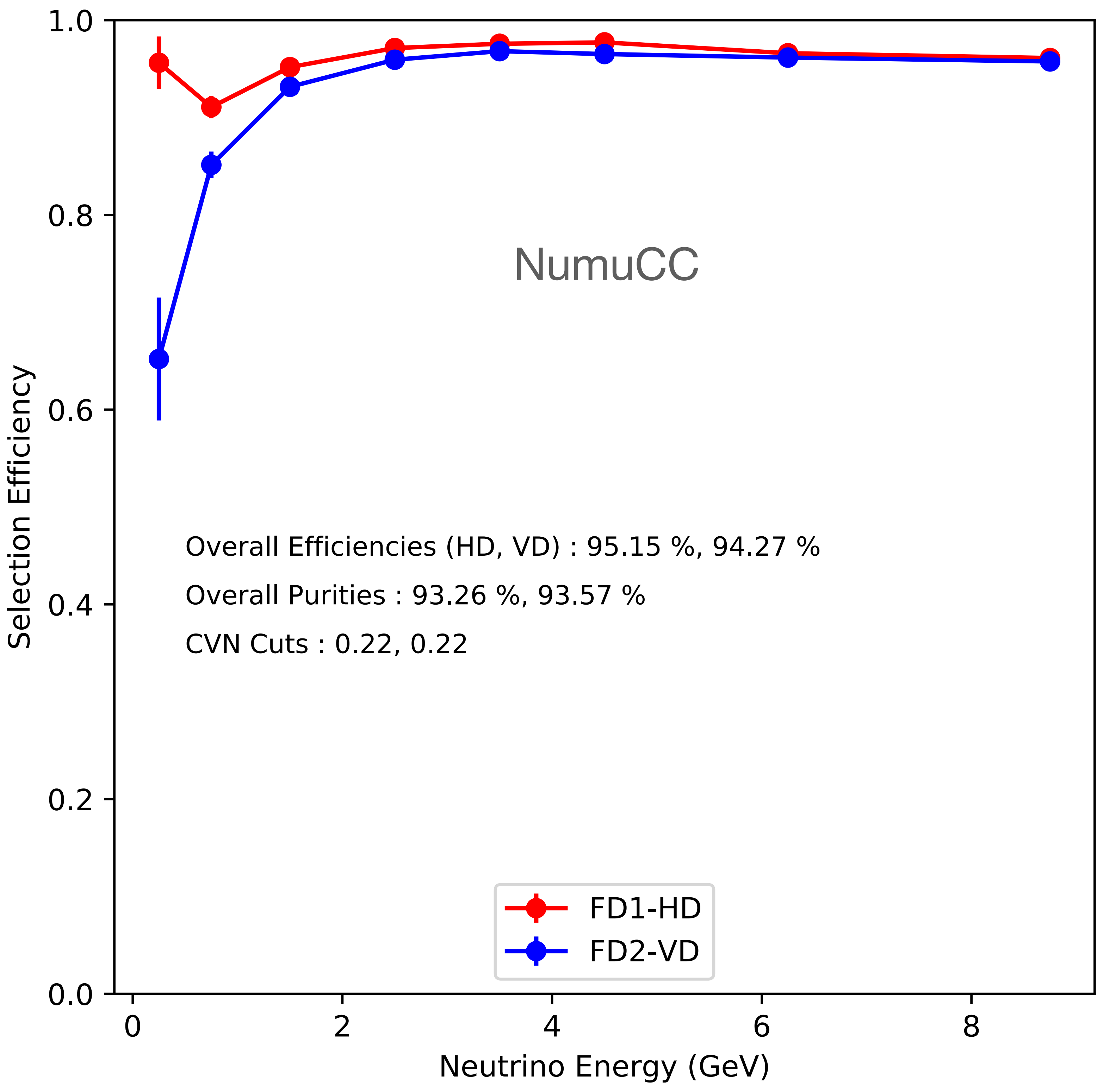


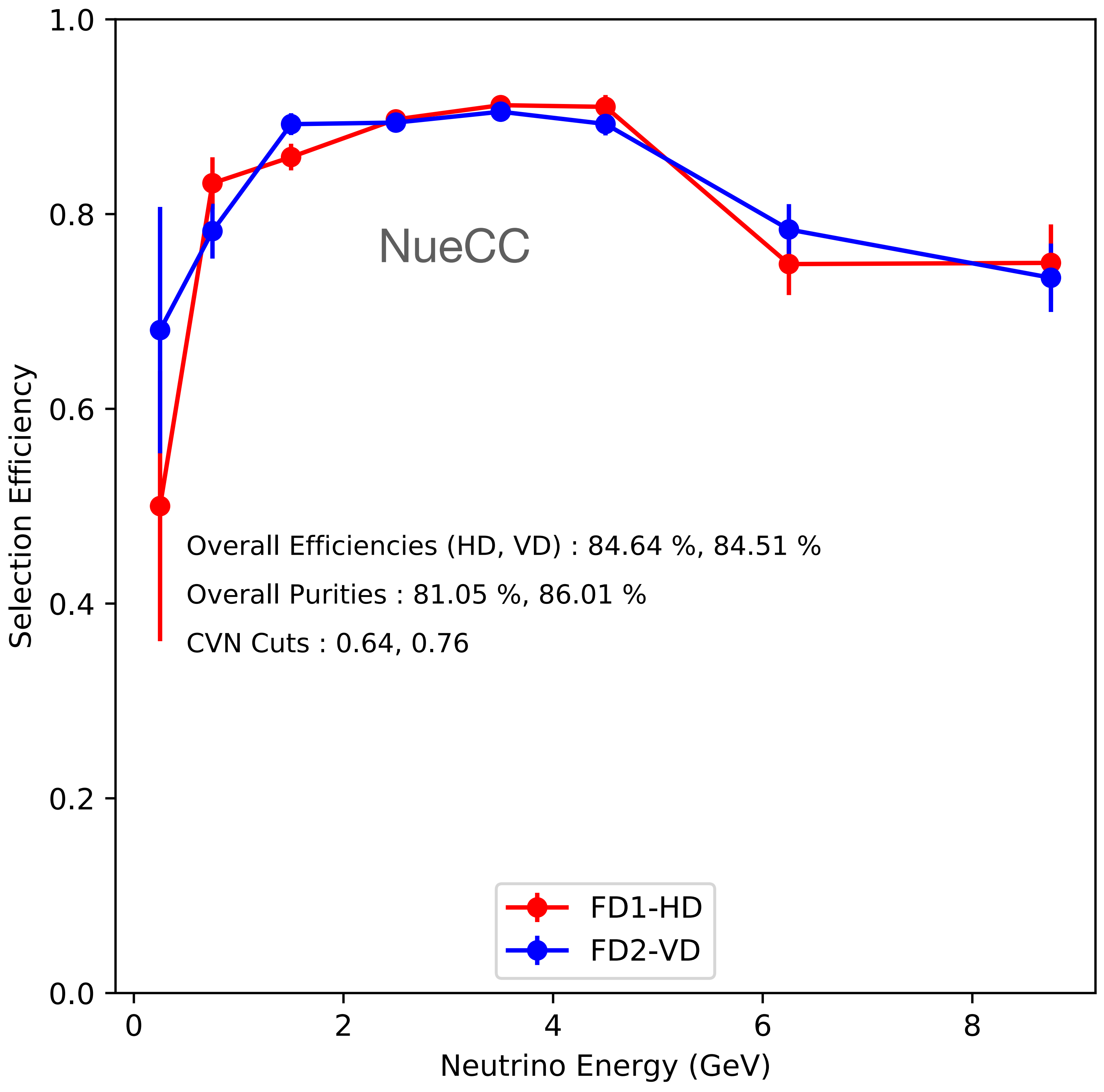


- Distributions look sensible. VD FOM values are a bit higher
- Optimizing CVN cut for Efficiency * Purity (FOM)
  - Distributions are oscillated and POT-weighted
  - simple oscillations ~ 2-flavor approximations used, maximal mixing, +2.4e-3 for Δm^2

HD vs VD - Neutrino Energy

FHC (Shown previously)



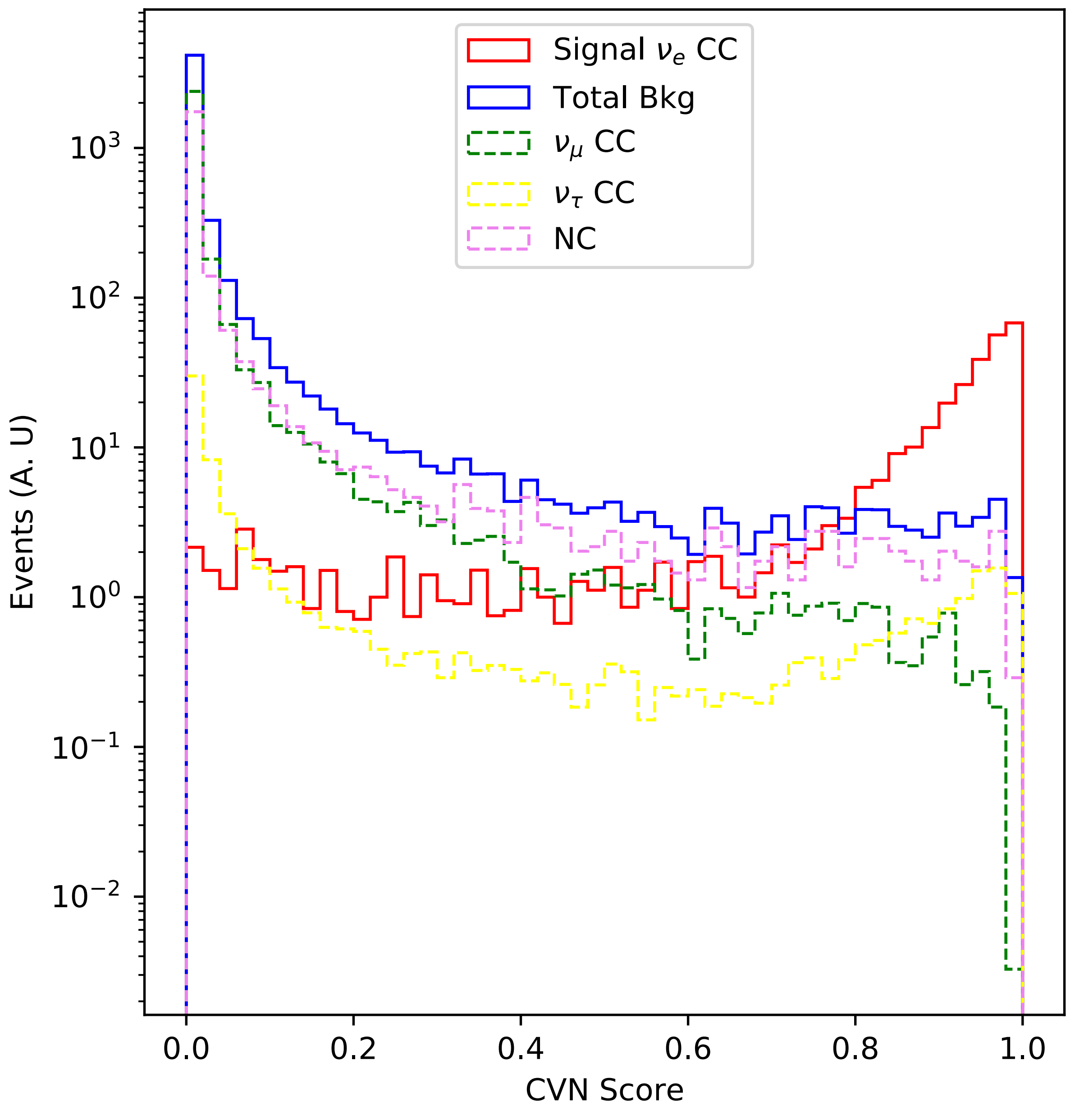


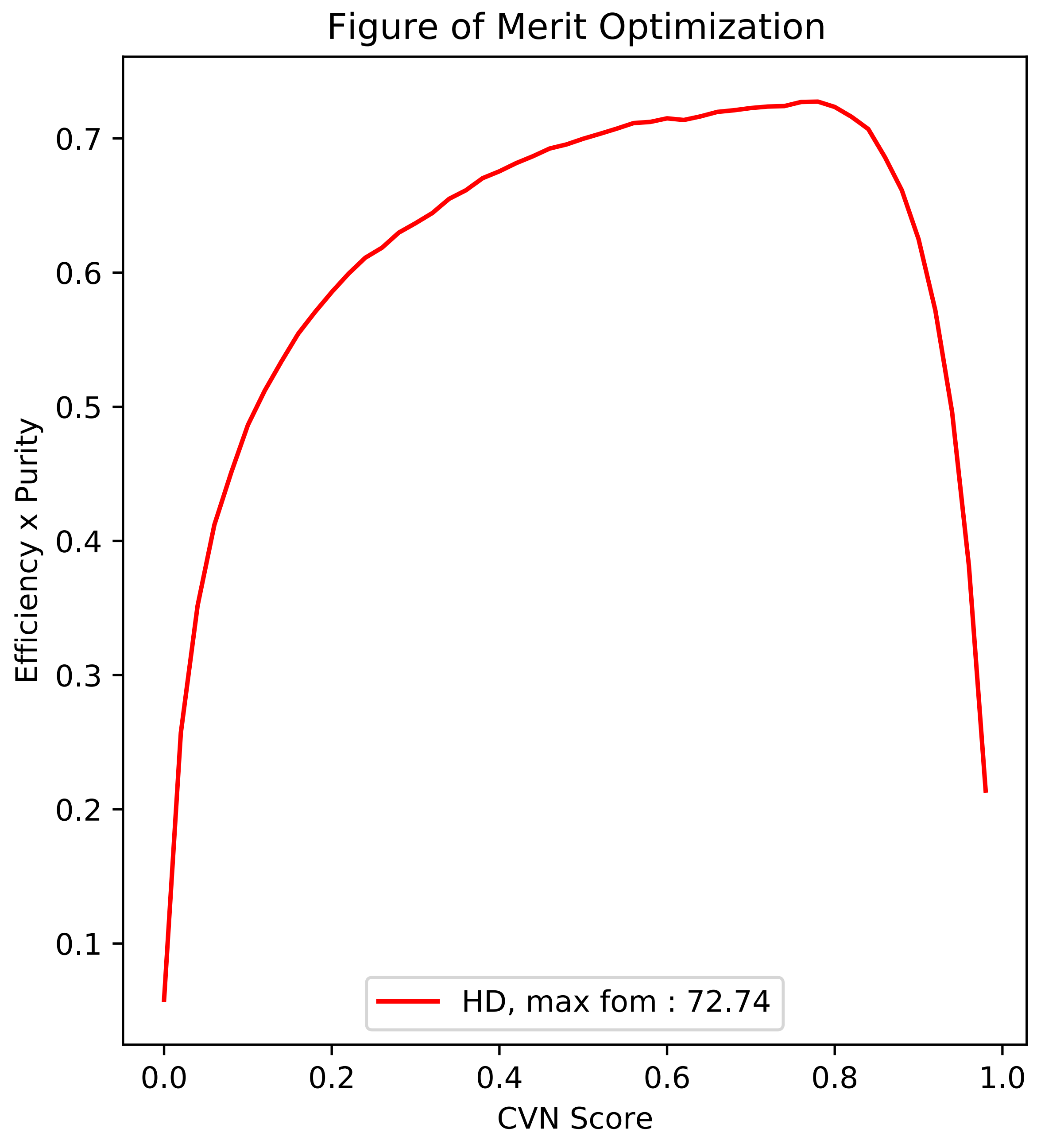
- NumuCC differences at low Enu - HD better than VD
  - HD also maybe slightly better at other Enu
- NueCC differences - VD performs better
- Performance gap closer than before

Training Results

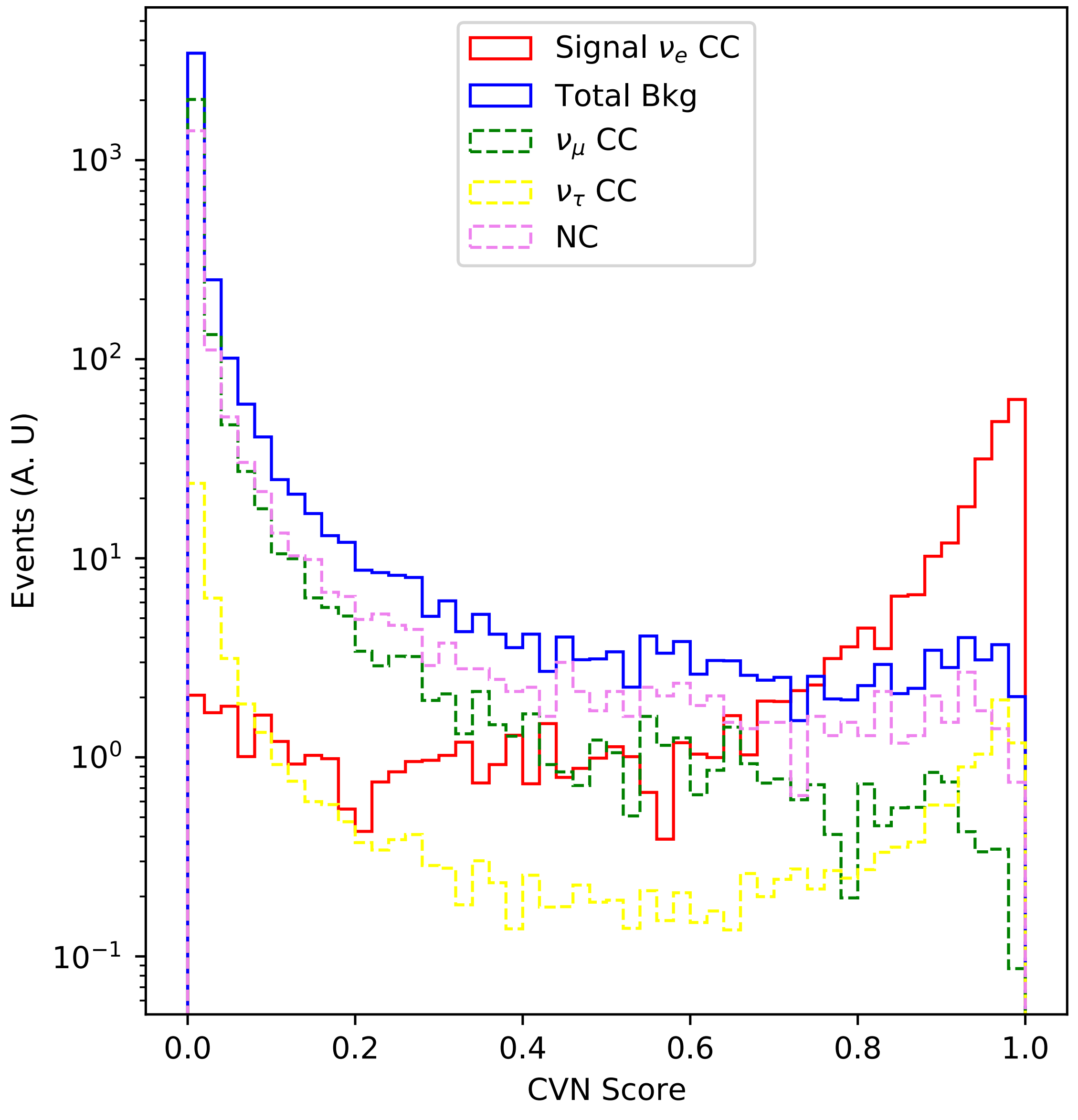
RHC (New)

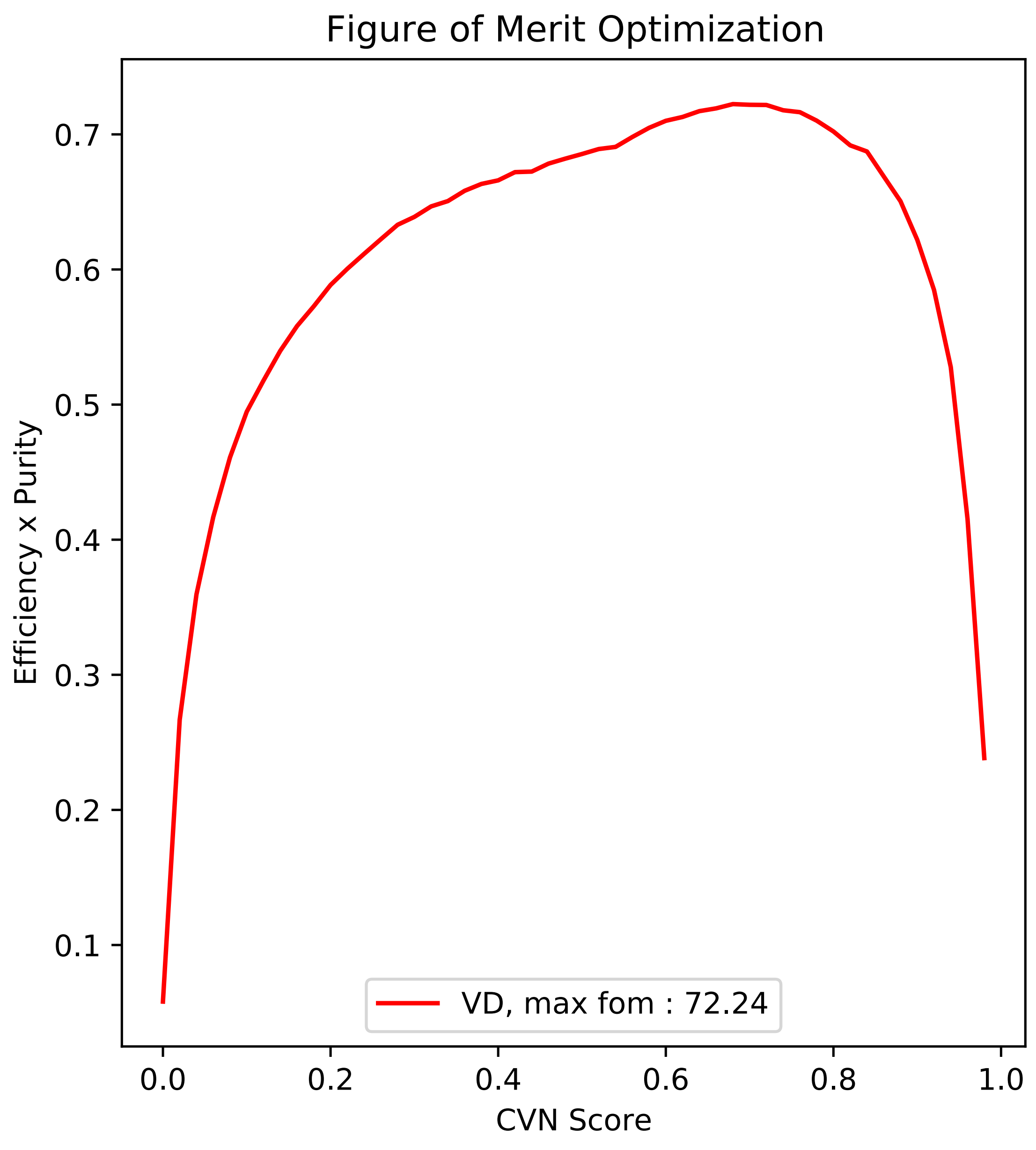
HD





VD

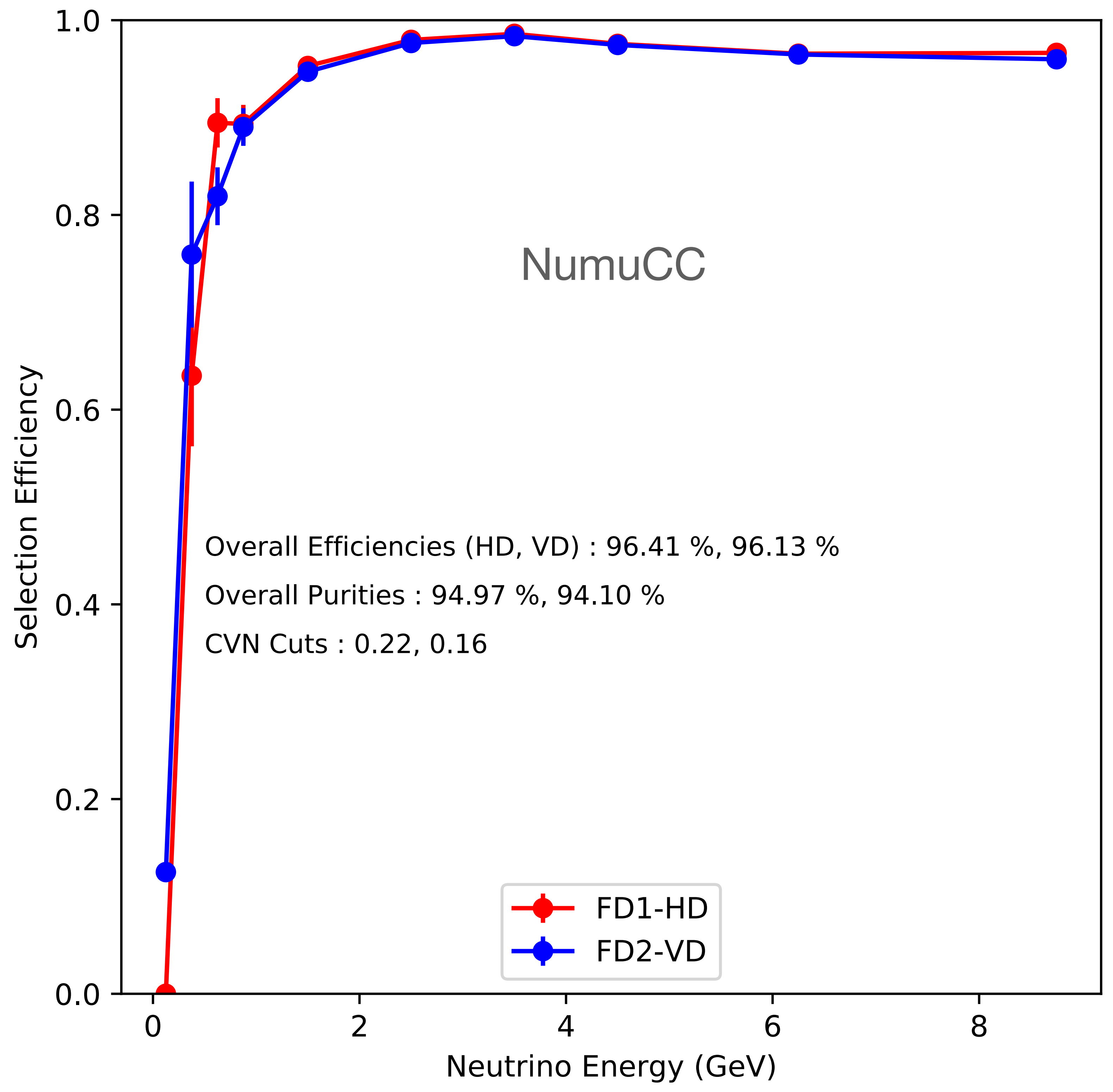


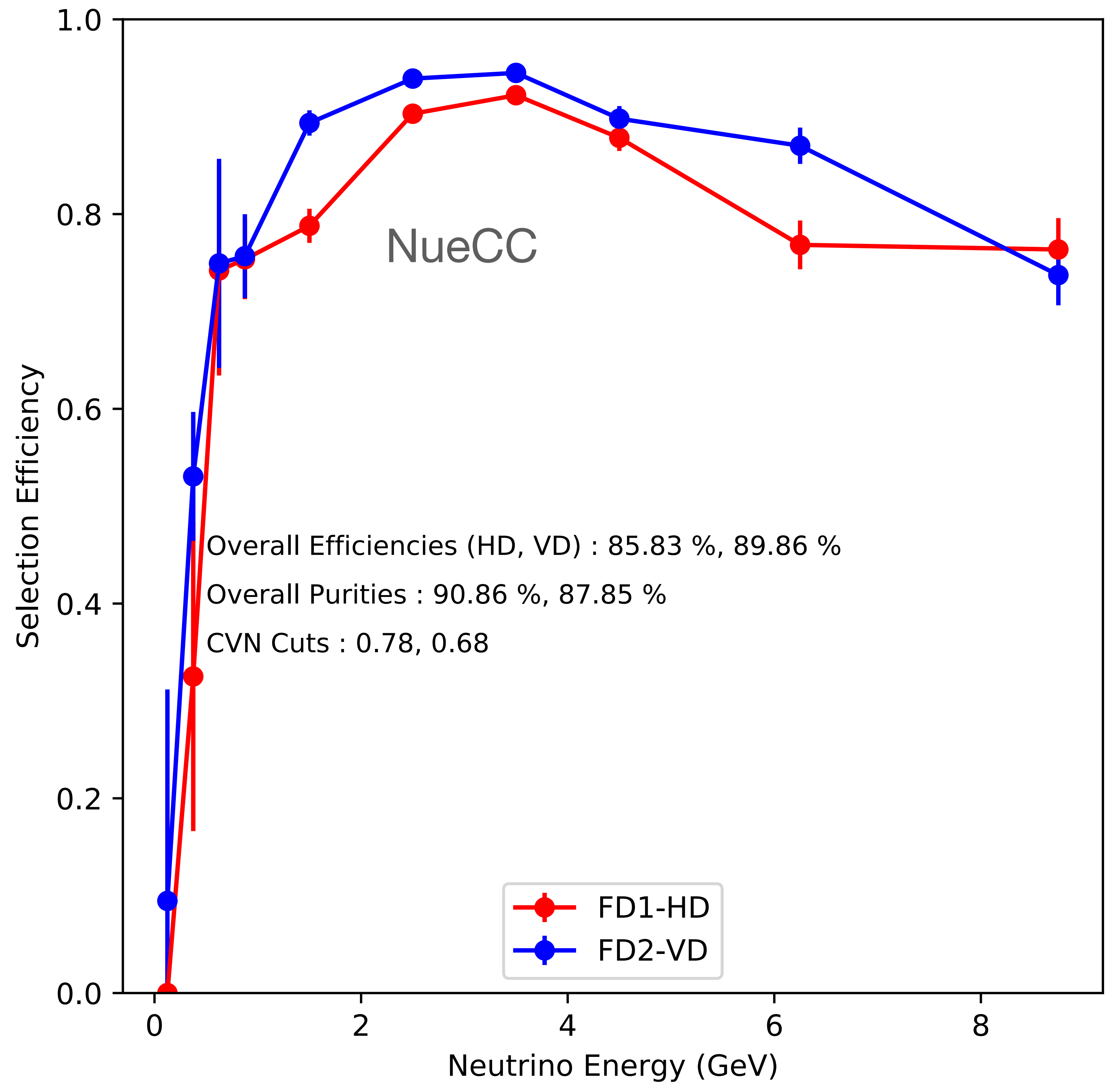


- Distributions look sensible. FOM values are pretty close (< 1% difference)
- Optimizing CVN cut for Efficiency * Purity (FOM)
  - Distributions are oscillated and POT-weighted
  - simple oscillations ~ 2-flavor approximations used, maximal mixing, +2.4e-3 for Δm^2

HD vs VD - Neutrino Energy

RHC (New)

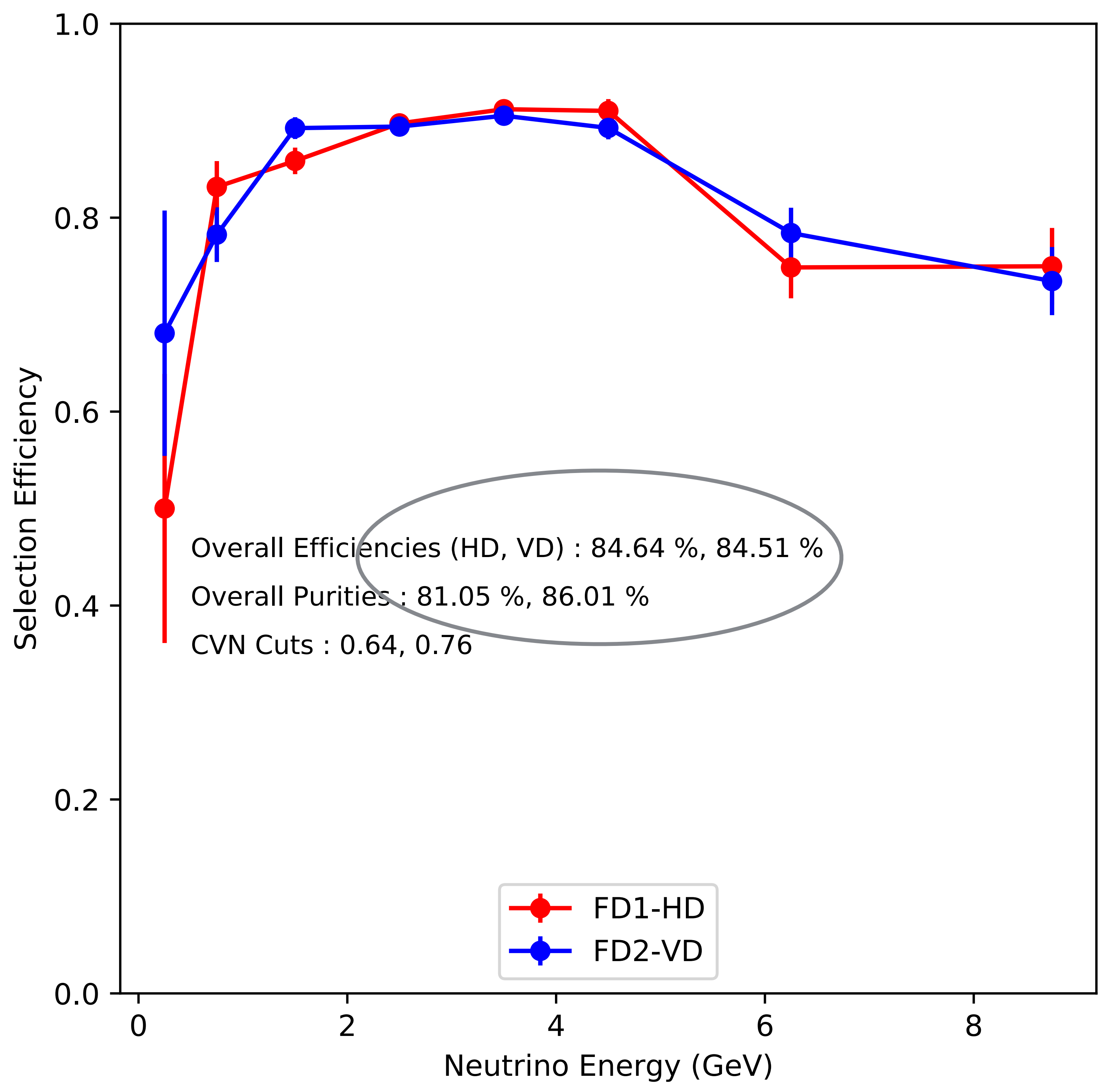


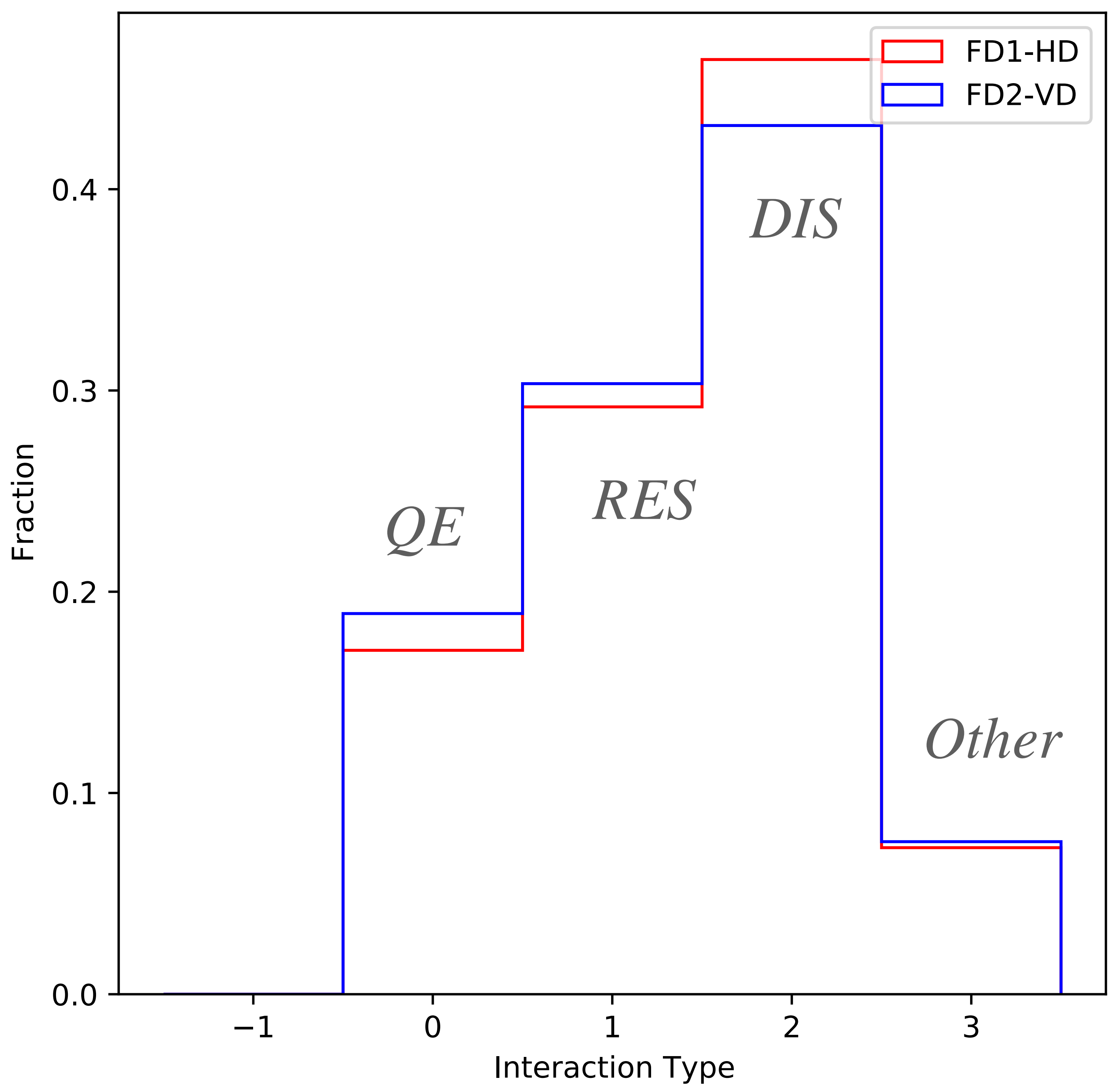


- NumuCC performance is better than FHC (96% eff vs 94-95% for FHC — expected since RHC has more forward going leptons)
  - No visible differences in HD vs VD here
- NueCC performance is better than FHC
  - No major differences in HD vs VD (VD eff comes out higher but at lower purity — prev FOM was comparable)

Training Composition

FHC - ν_eCC



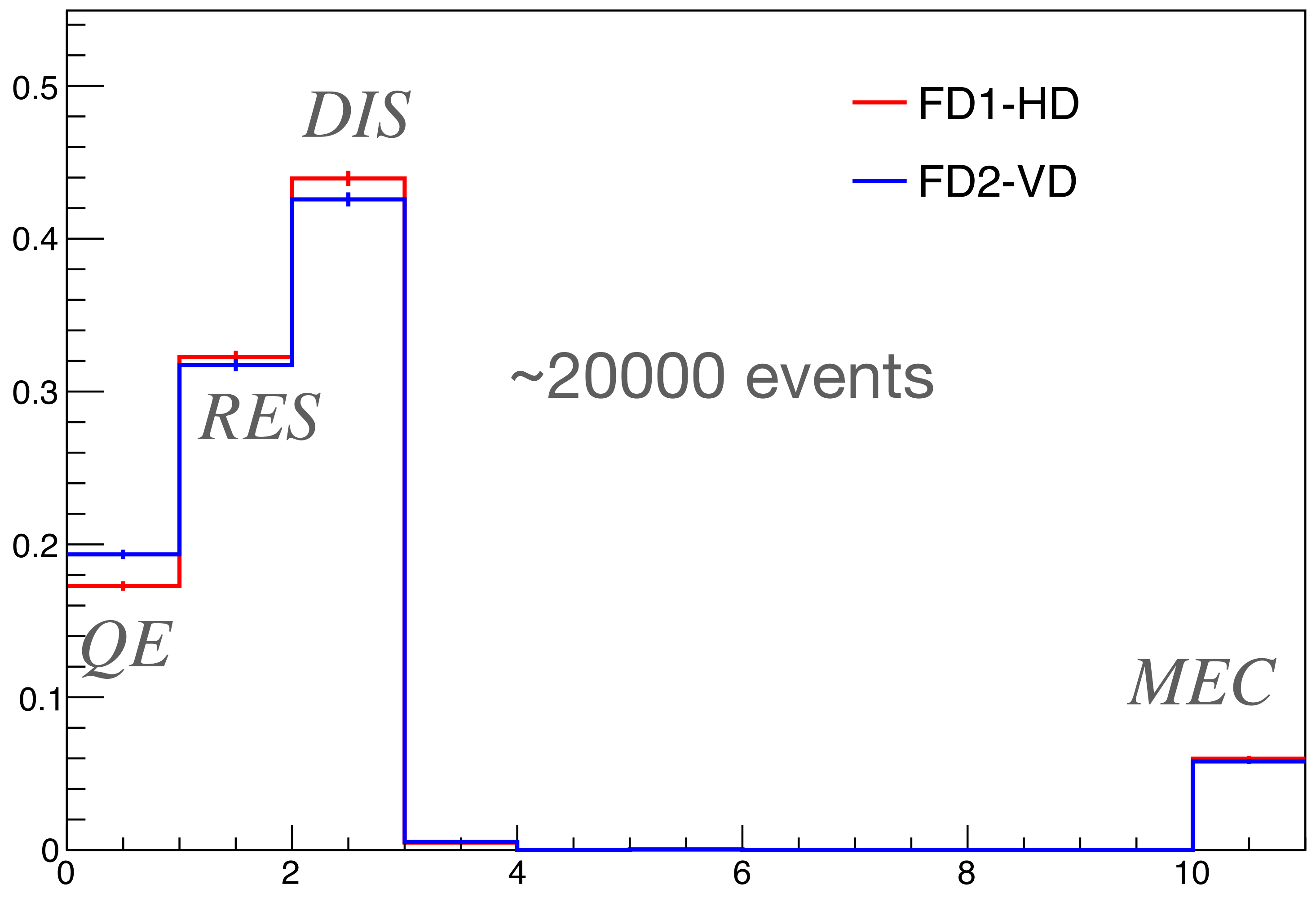


- Previously, we saw some differences in composition of interaction modes for ν_eCCs
  - More DIS in HD, could cause drop in purity since DIS performance is weaker (QE is higher in VD as well)
  - Same GENIE version used, so differences are a bit odd/may be coming from preselection

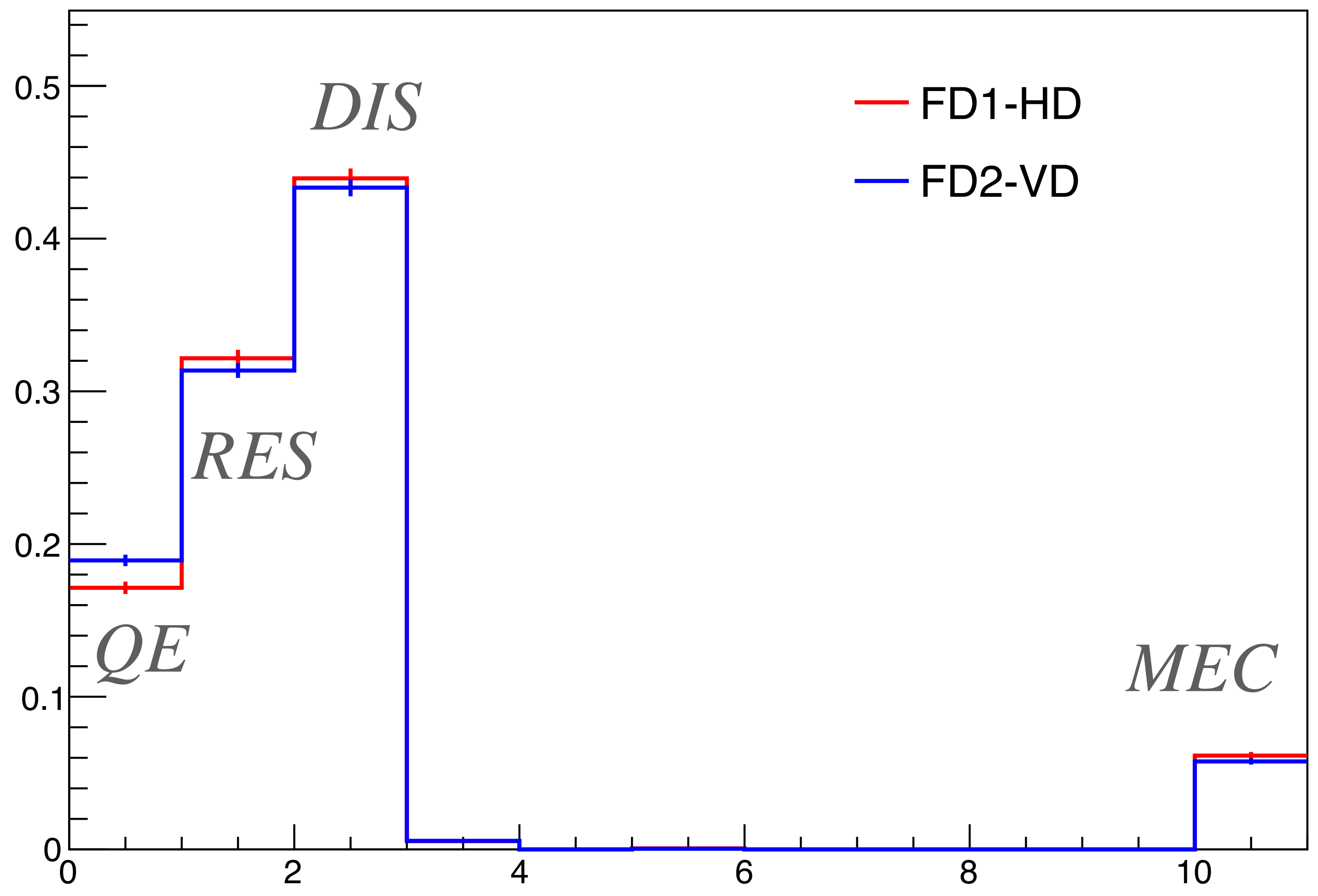
Training Composition

FHC - ν_eCC

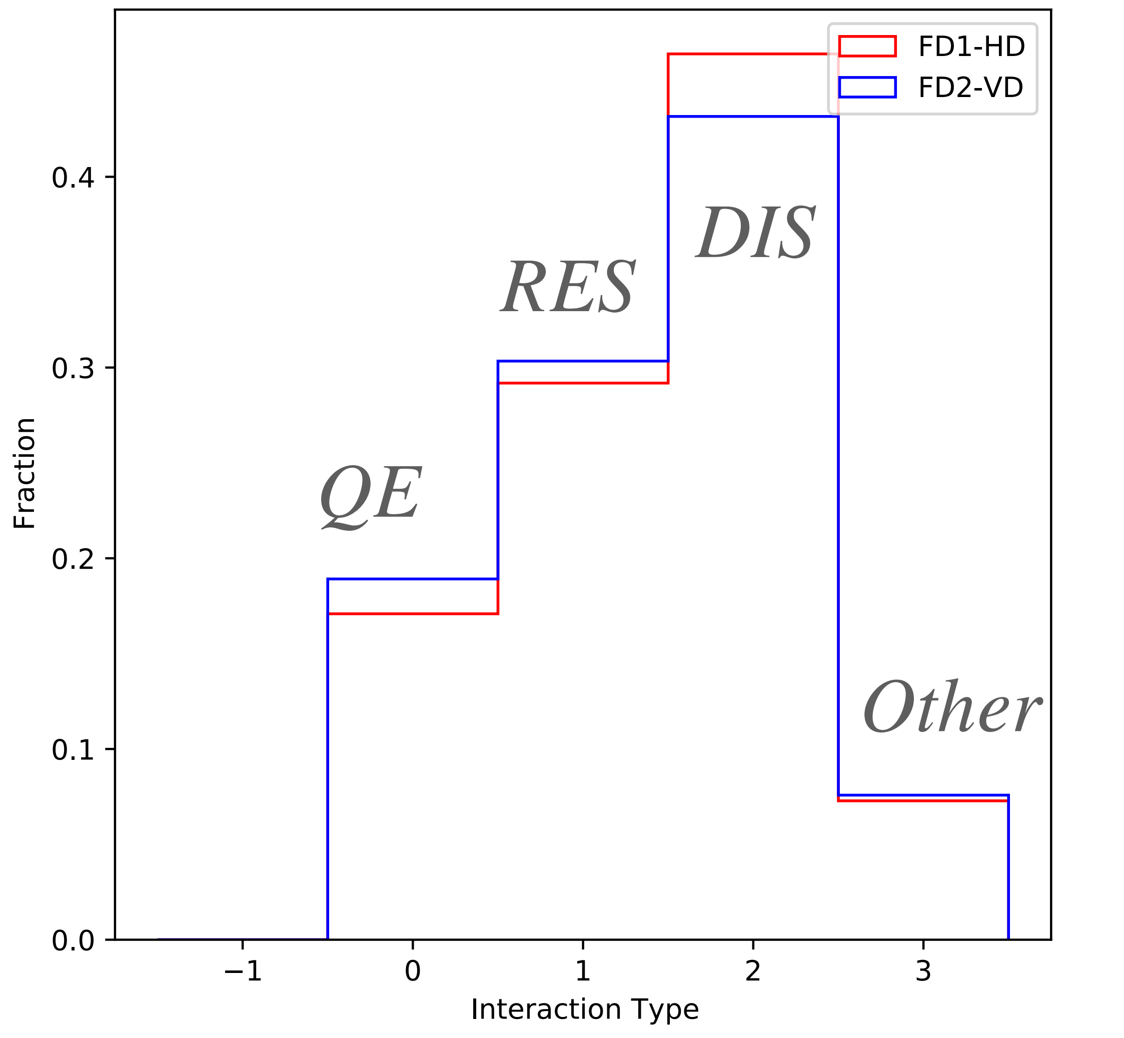
No Cuts (Out of the Box)



FV Cuts



Training sample (FV + Nhits > 100)

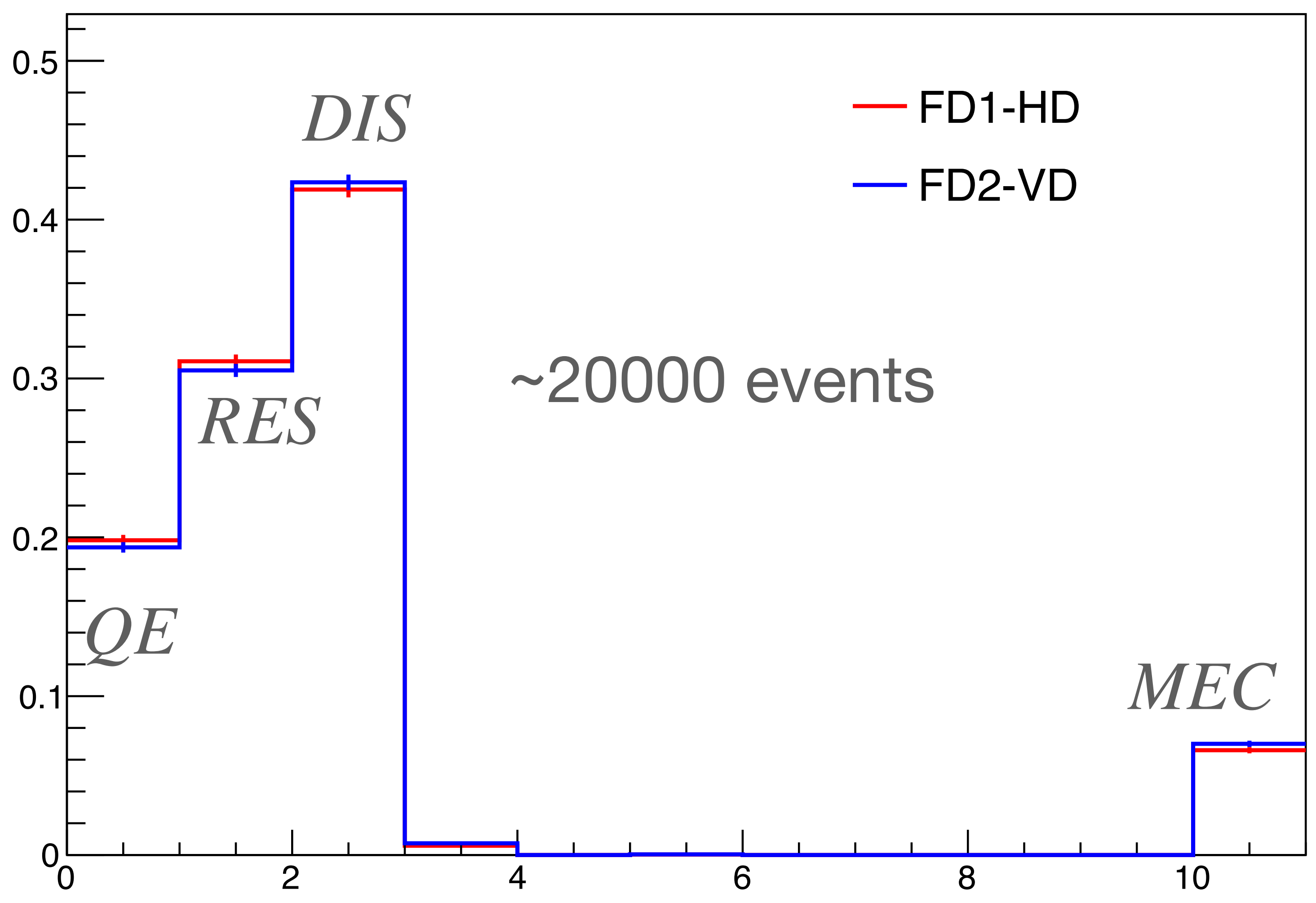


- Some differences in QE maybe for no cuts
    - But in general DIS etc are fairly close (within stats error)
    - Could be some differences in hit formation too for VD vs HD, so maybe that has an impact as well
- This + extra training stats could explain differences
- More plots : https://www.phy.bnl.gov/~nitish/random/cvnpresel/nue/

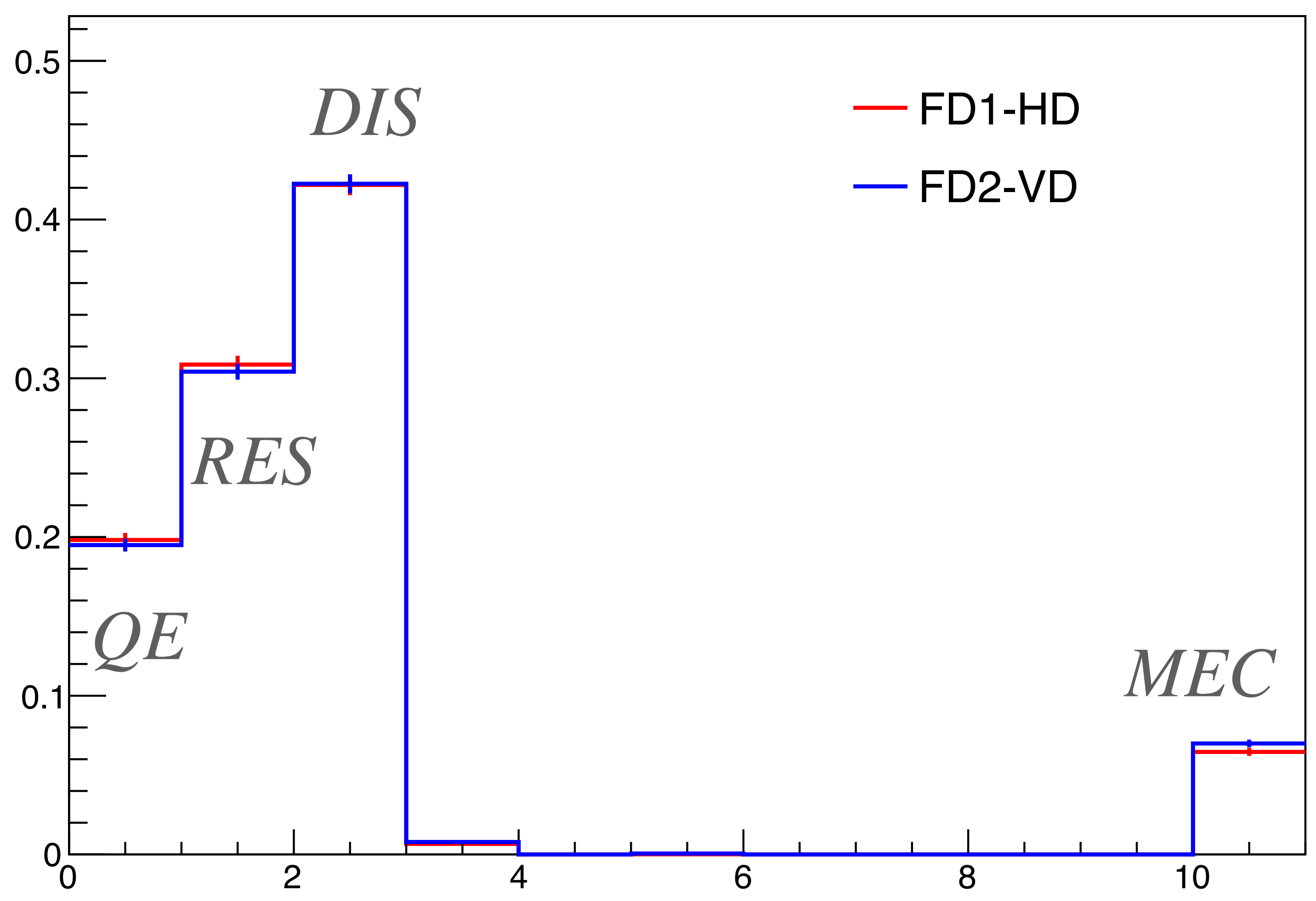
Training Composition

RHC - ν_eCC

No Cuts (Out of the Box)

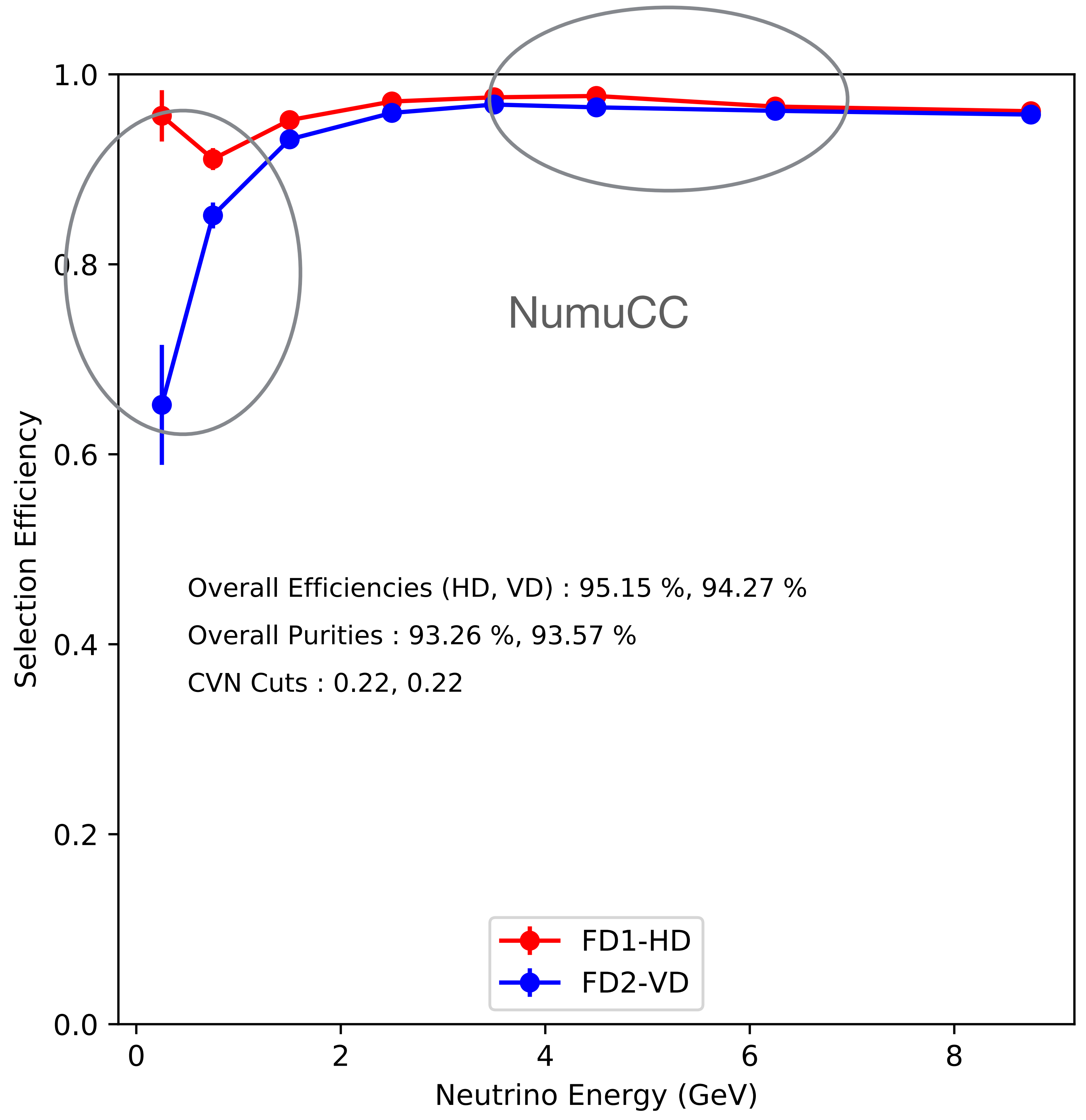


FV Cuts

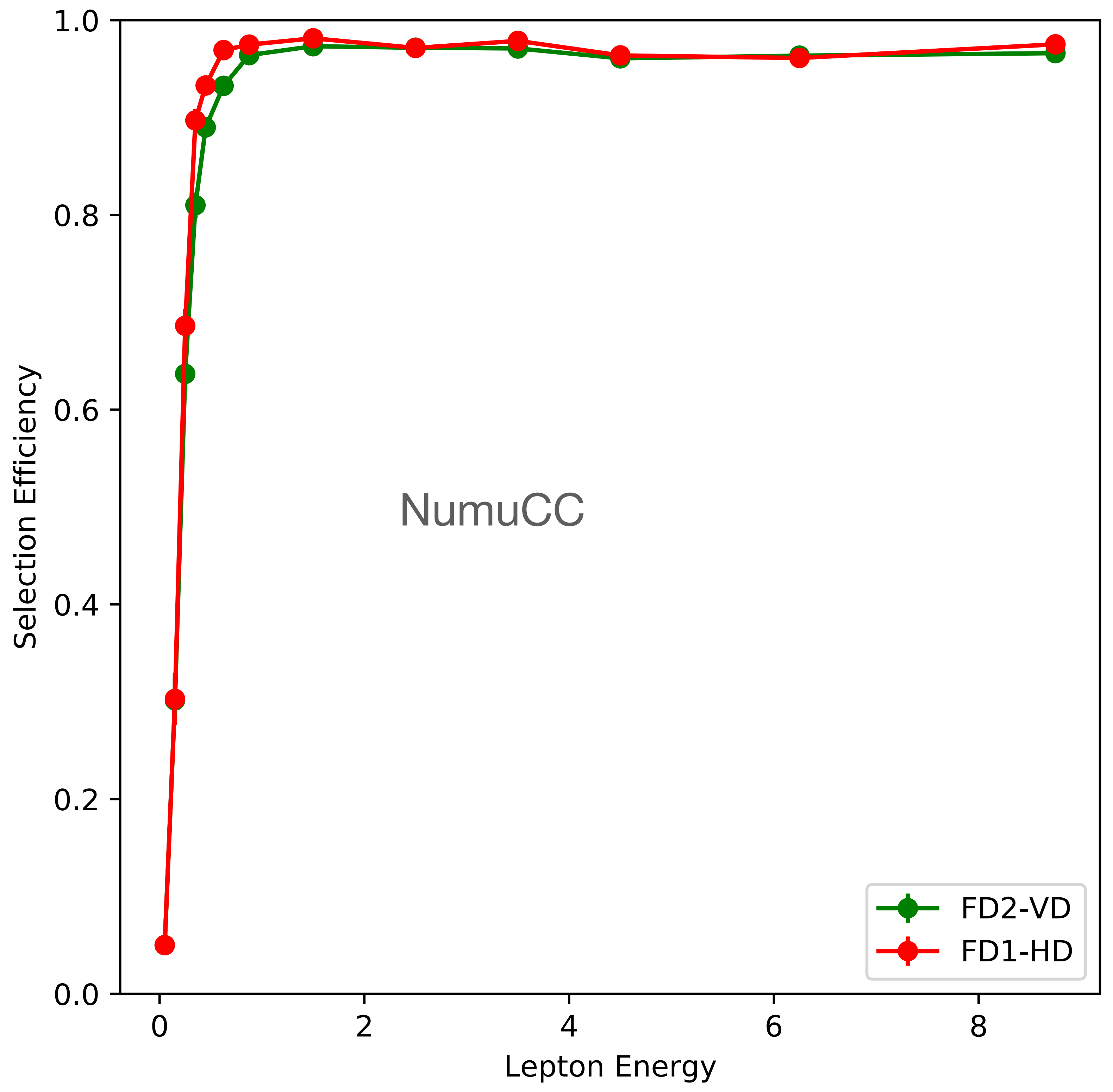


- No real differences in composition for RHC ν_eCC
- More plots : https://www.phy.bnl.gov/~nitish/random/cvnpresel/anue/

HD vs VD - Muon Energy



FHC - Understanding $\nu_\mu CC$ Differences

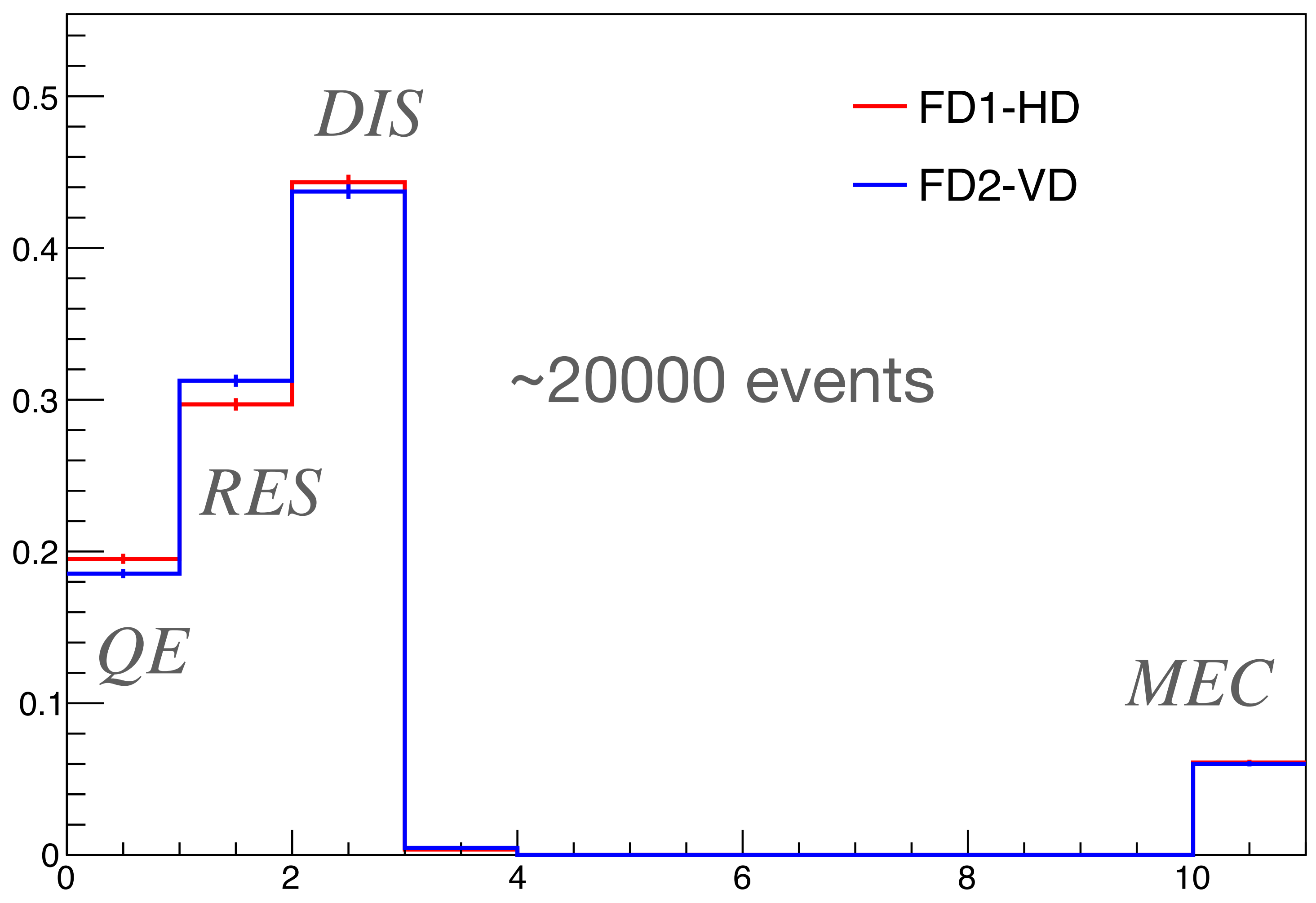


- Working on getting plots vs muon length (esp for muons) to probe differences
- For $\nu_\mu CC$ - lepton energy is a good proxy. See all difference ~concentrated at < 1GeV muons
    - Low neutrino energy or DIS interactions

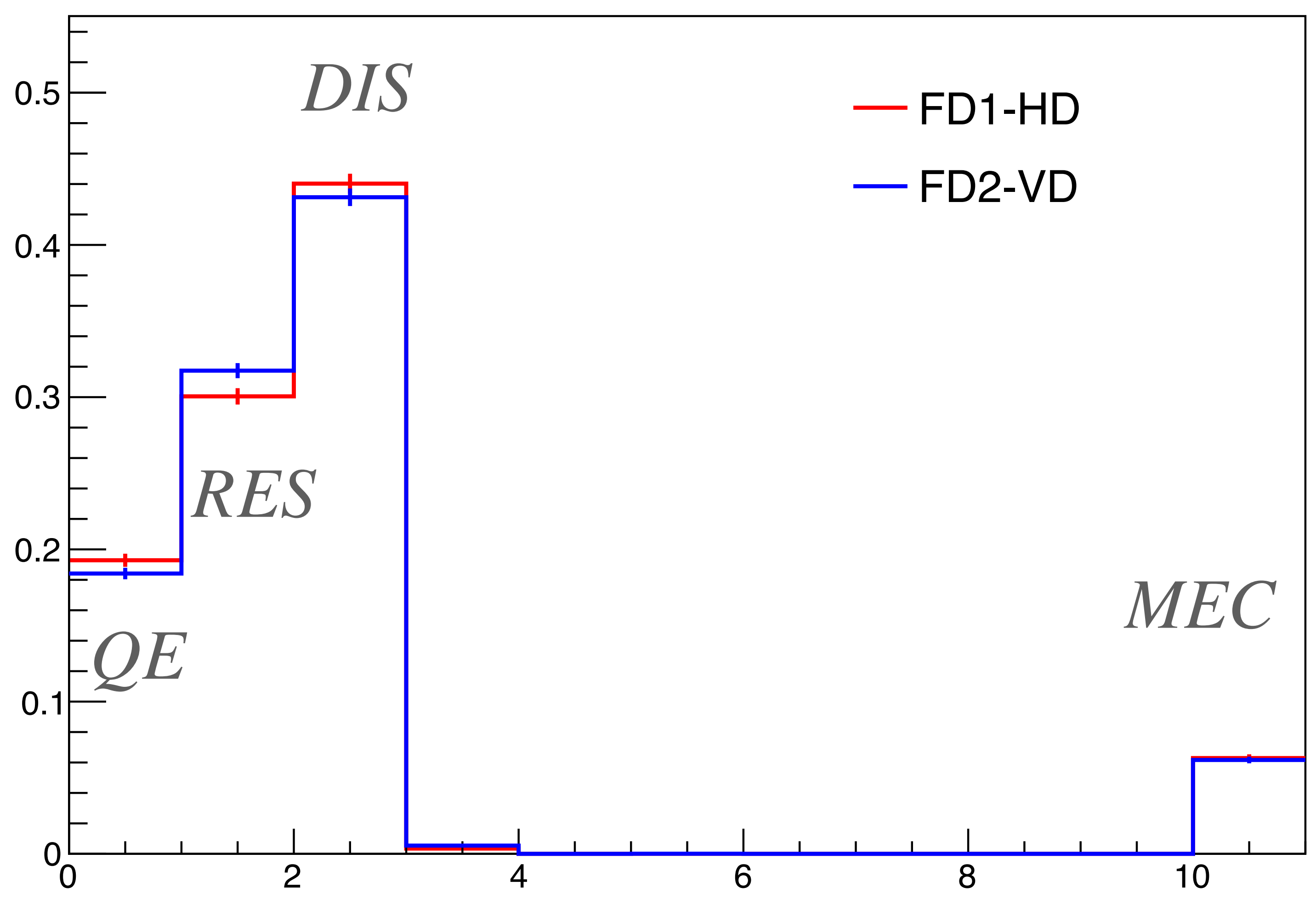
Training Composition

FHC - ν_μCC

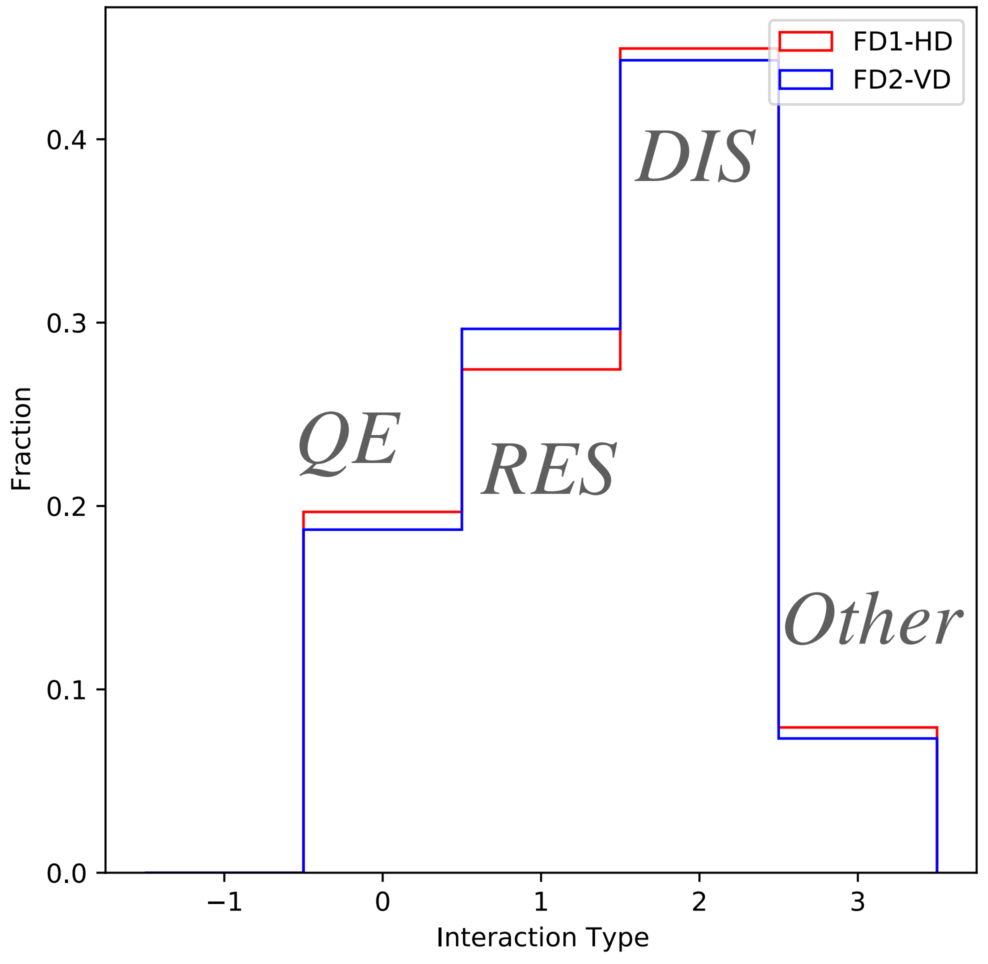
No Cuts (Out of the Box)



FV Cuts



Training sample (FV + Nhits > 100)



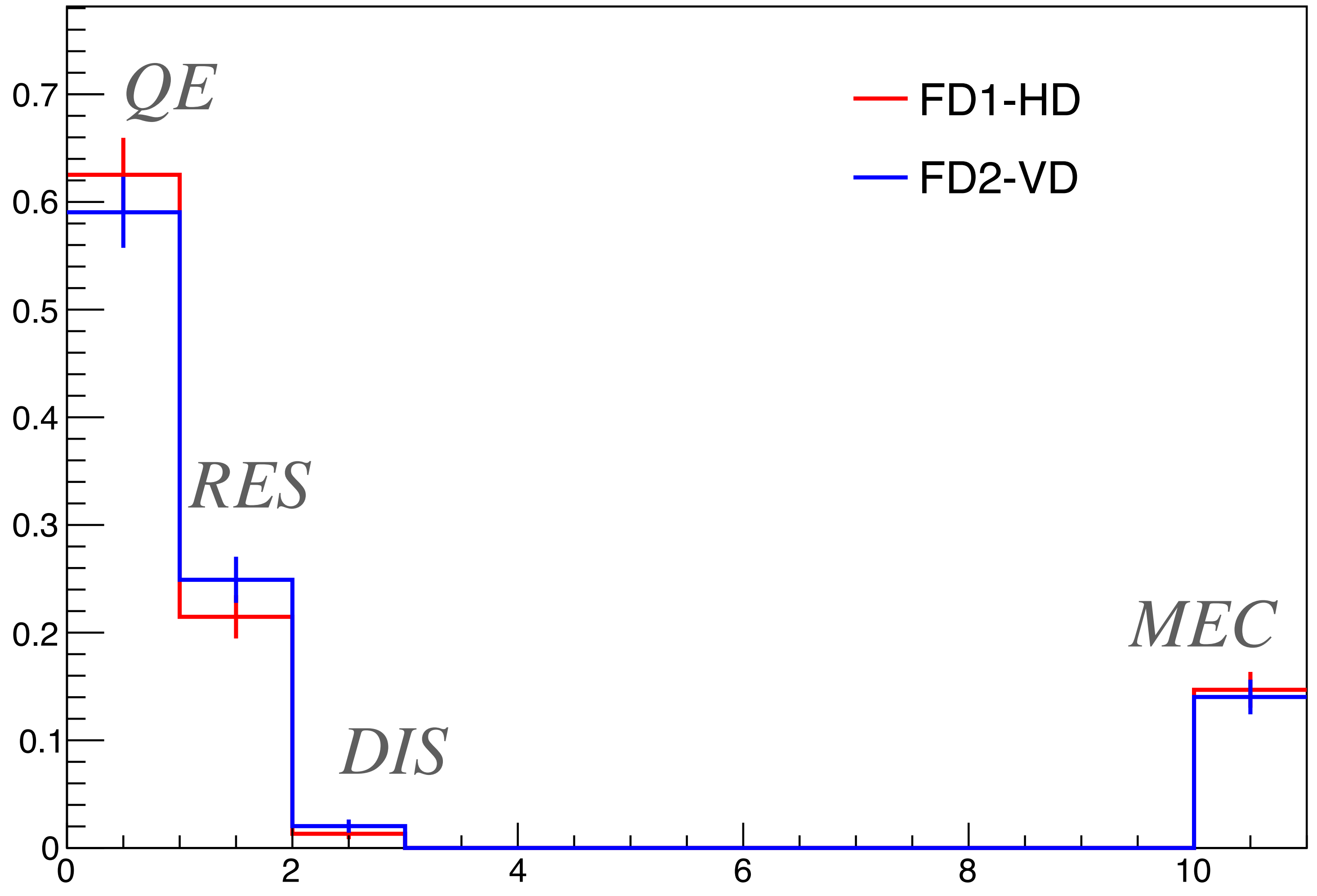
- No real differences in training composition
- More plots : https://www.phy.bnl.gov/~nitish/random/cvnpresel/nu/

Training Composition

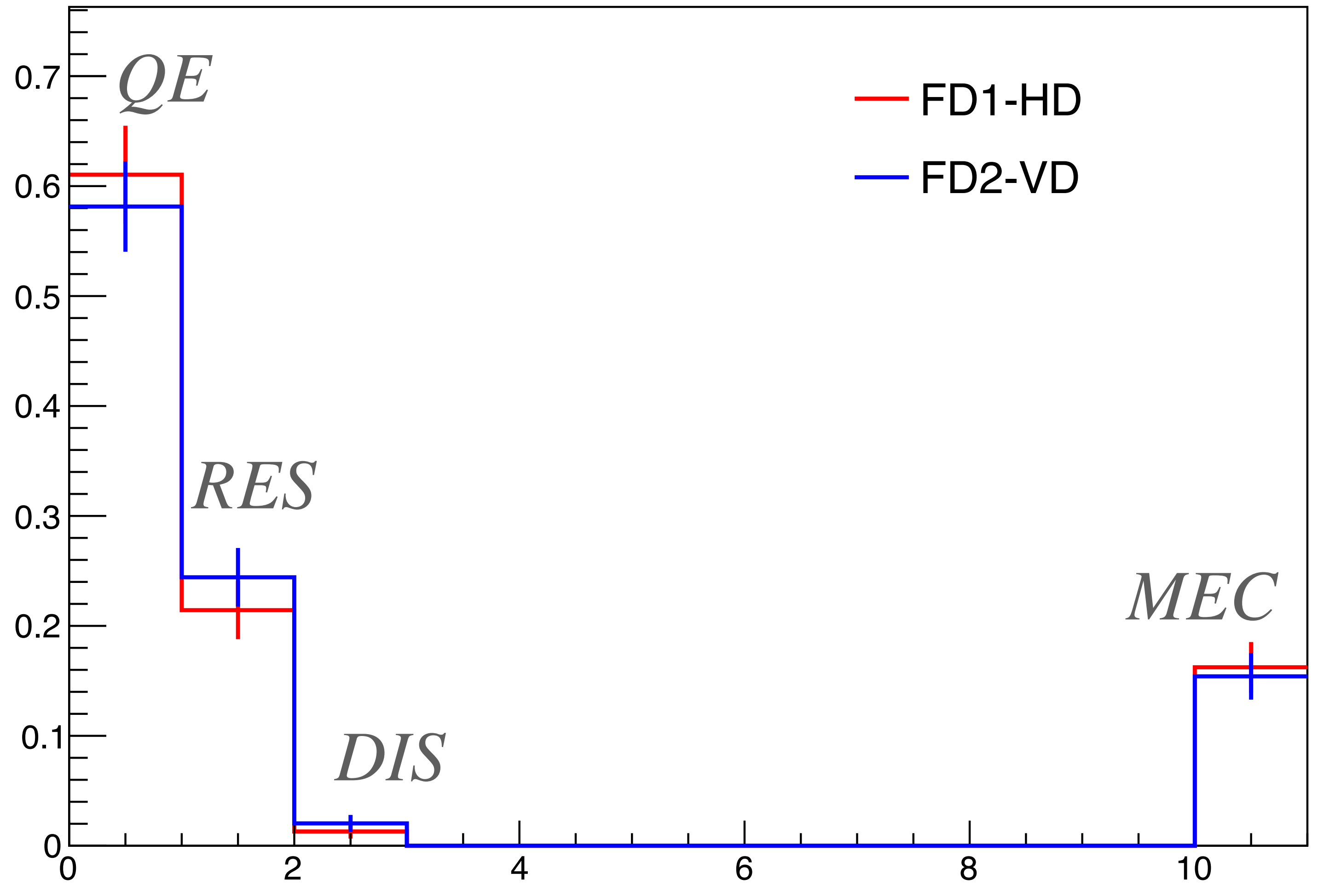
FHC - ν_μCC

E_ν < 1 GeV

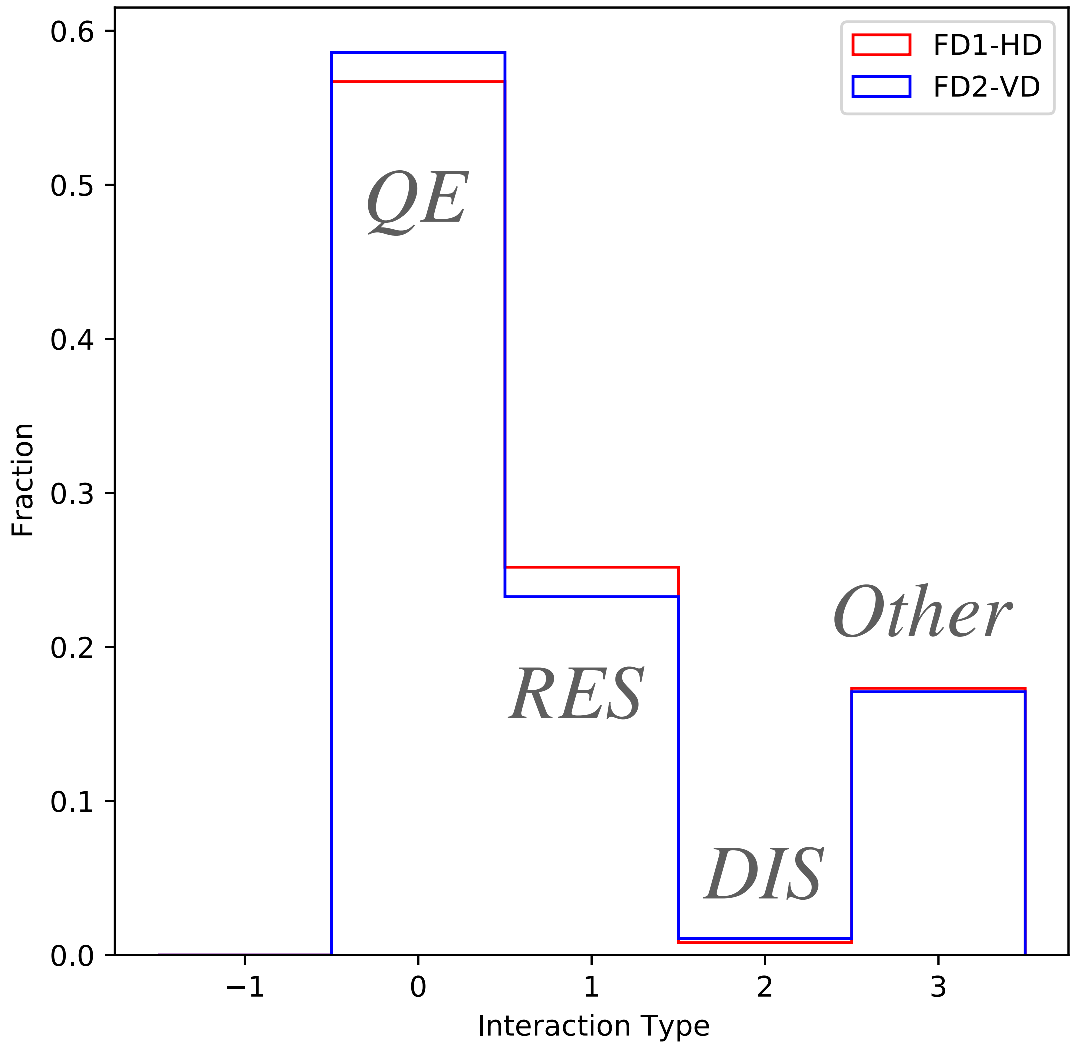
No Cuts (Out of the Box)



FV Cuts



Training sample (FV + Nhits > 100)



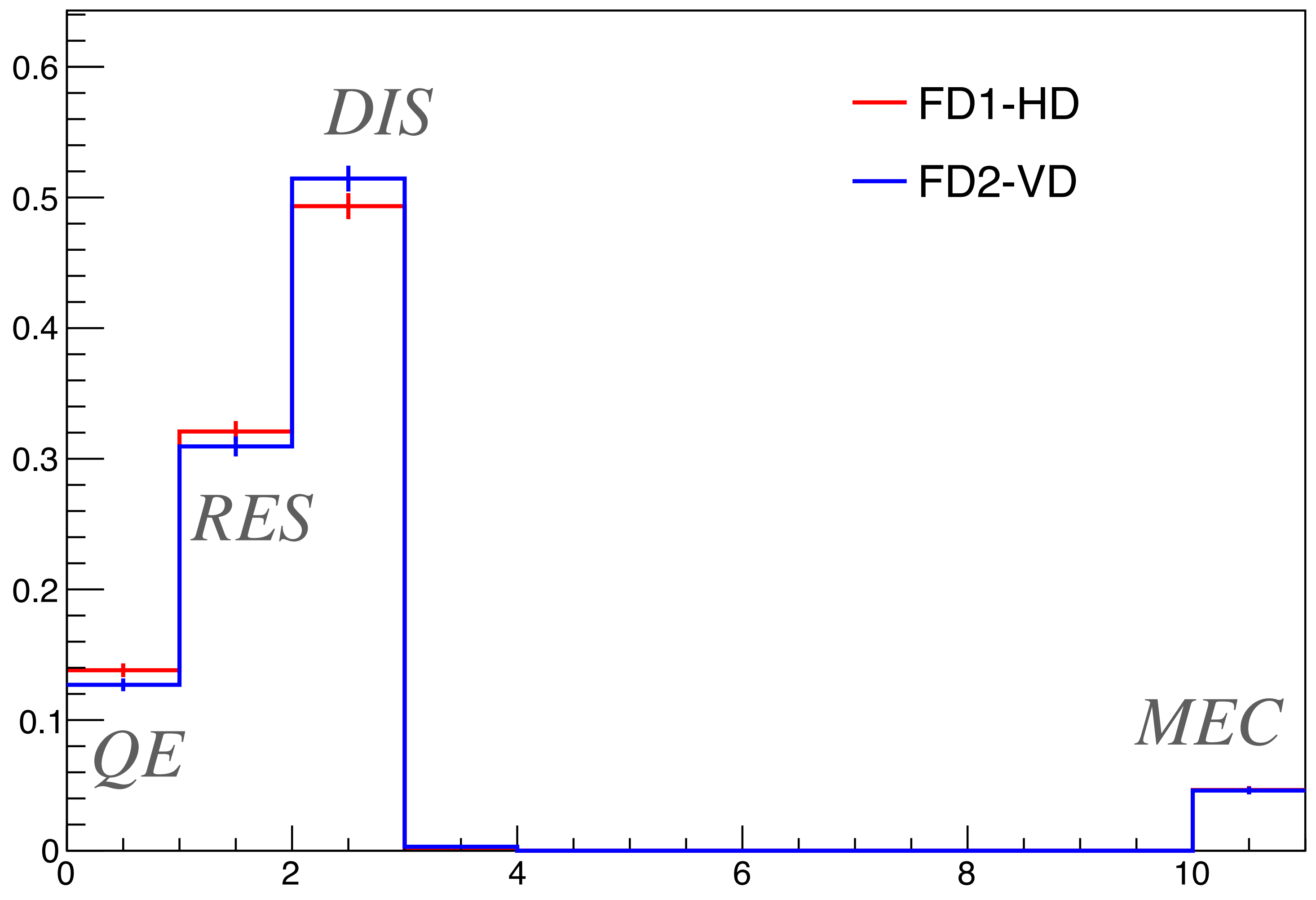
- No real differences in training composition for low energy events
  - Doesn't explain HD performing better
- More plots : https://www.phy.bnl.gov/~nitish/random/cvnpresel/nu/

Training Composition

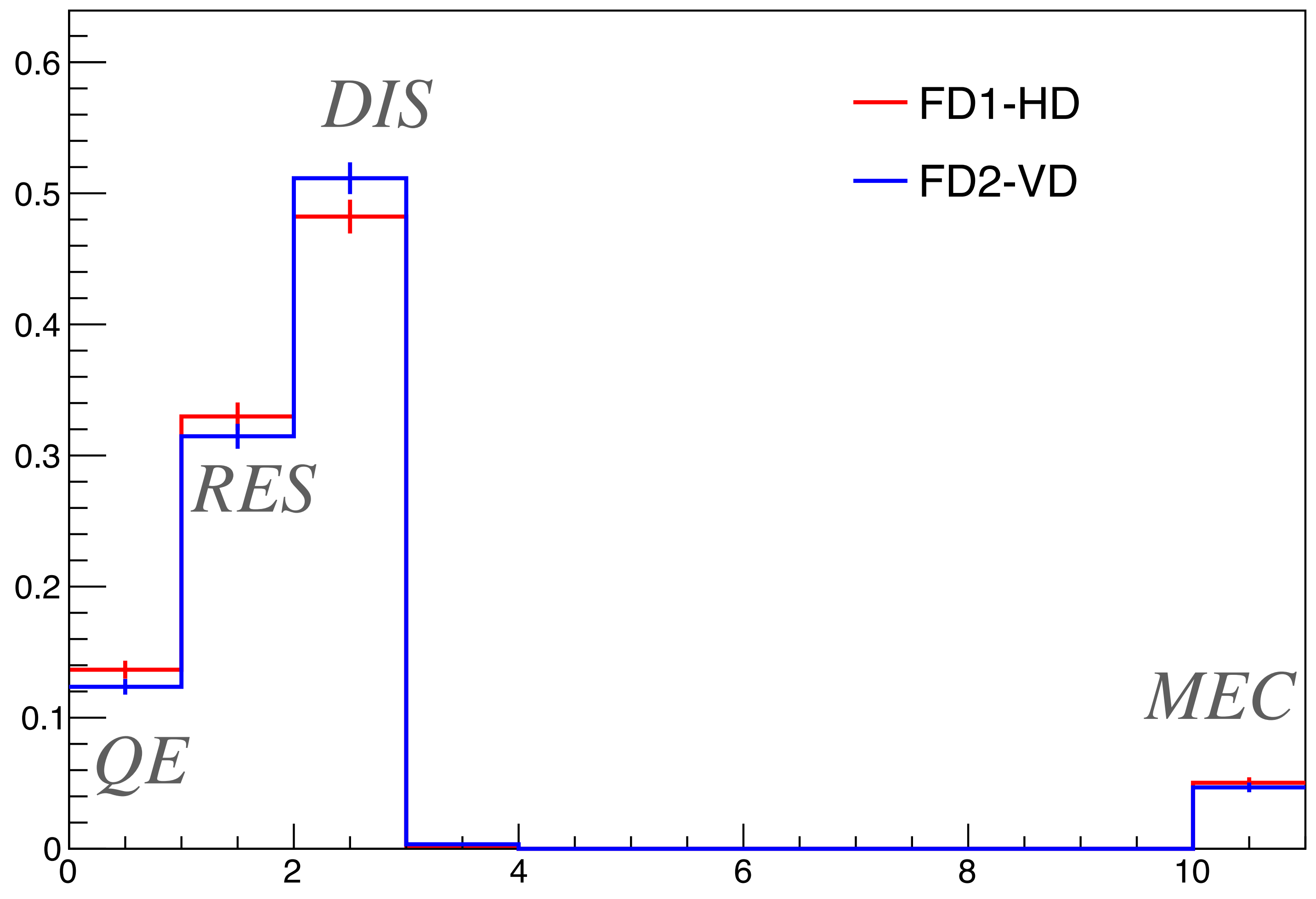
FHC - ν_μCC

E_μ < 1 GeV

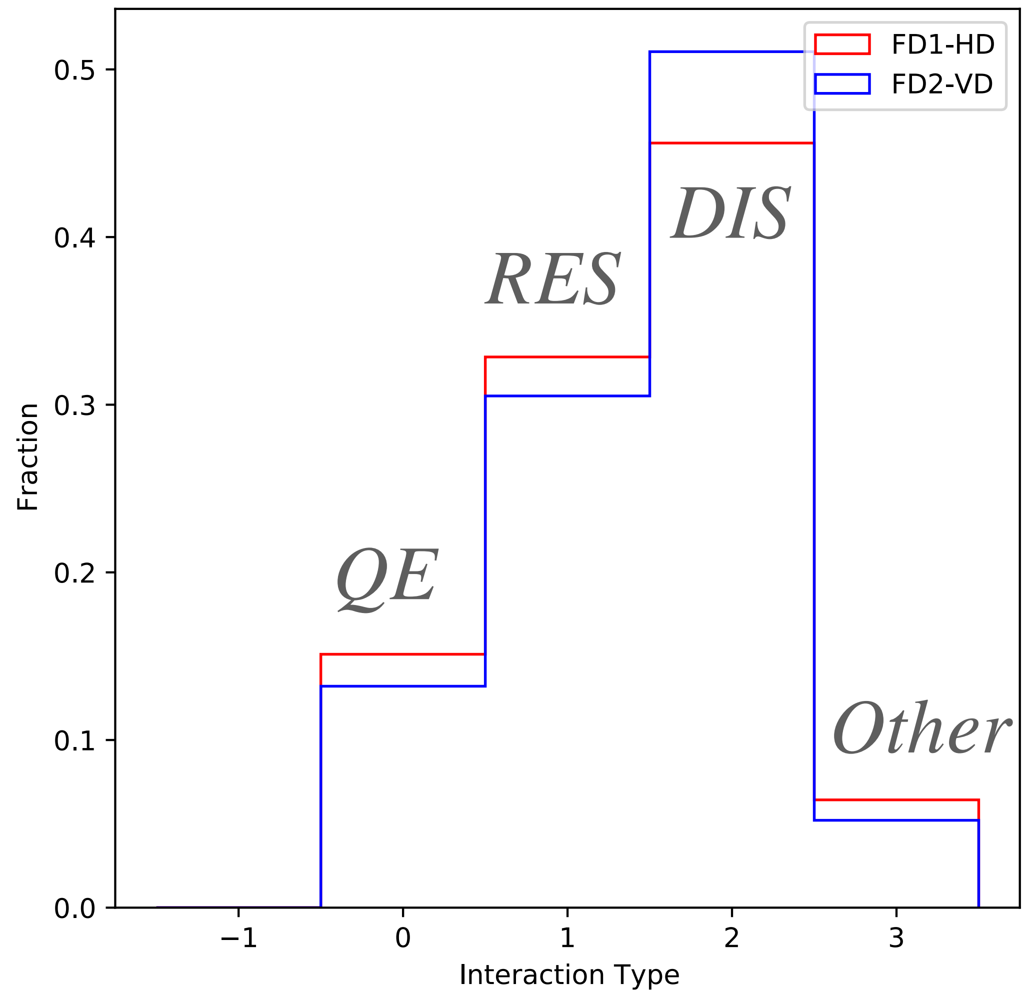
No Cuts (Out of the Box)



FV Cuts



Training sample (FV + Nhits > 100)

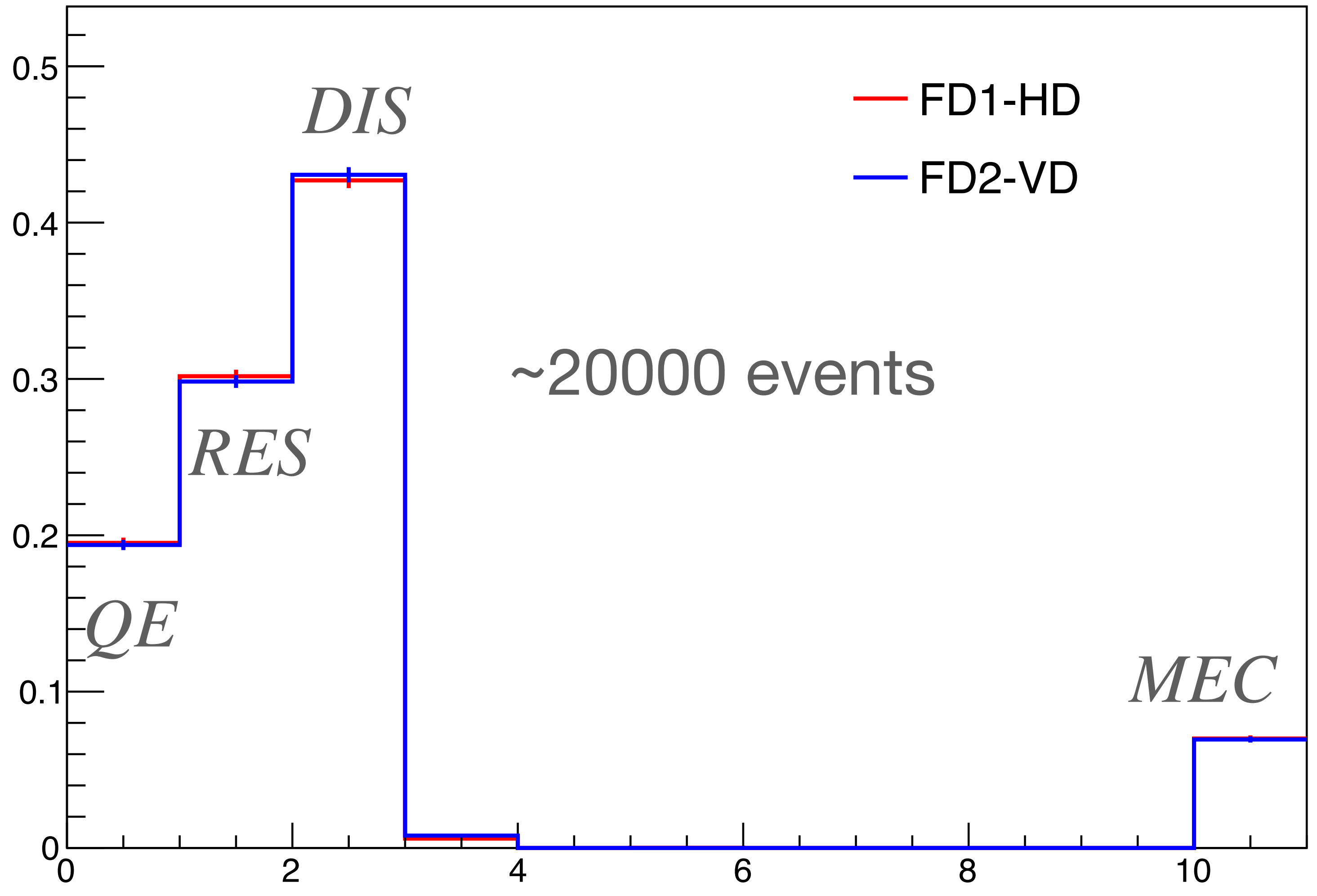


- Lot more DIS events if I just apply E_μ < 1 GeV cut instead of E_ν
- No real differences in DIS out of the box or when I apply FV cuts
  - But do notice it at the final stage, could explain part of the gap but not all
- More plots : https://www.phy.bnl.gov/~nitish/random/cvnpresel/nu/

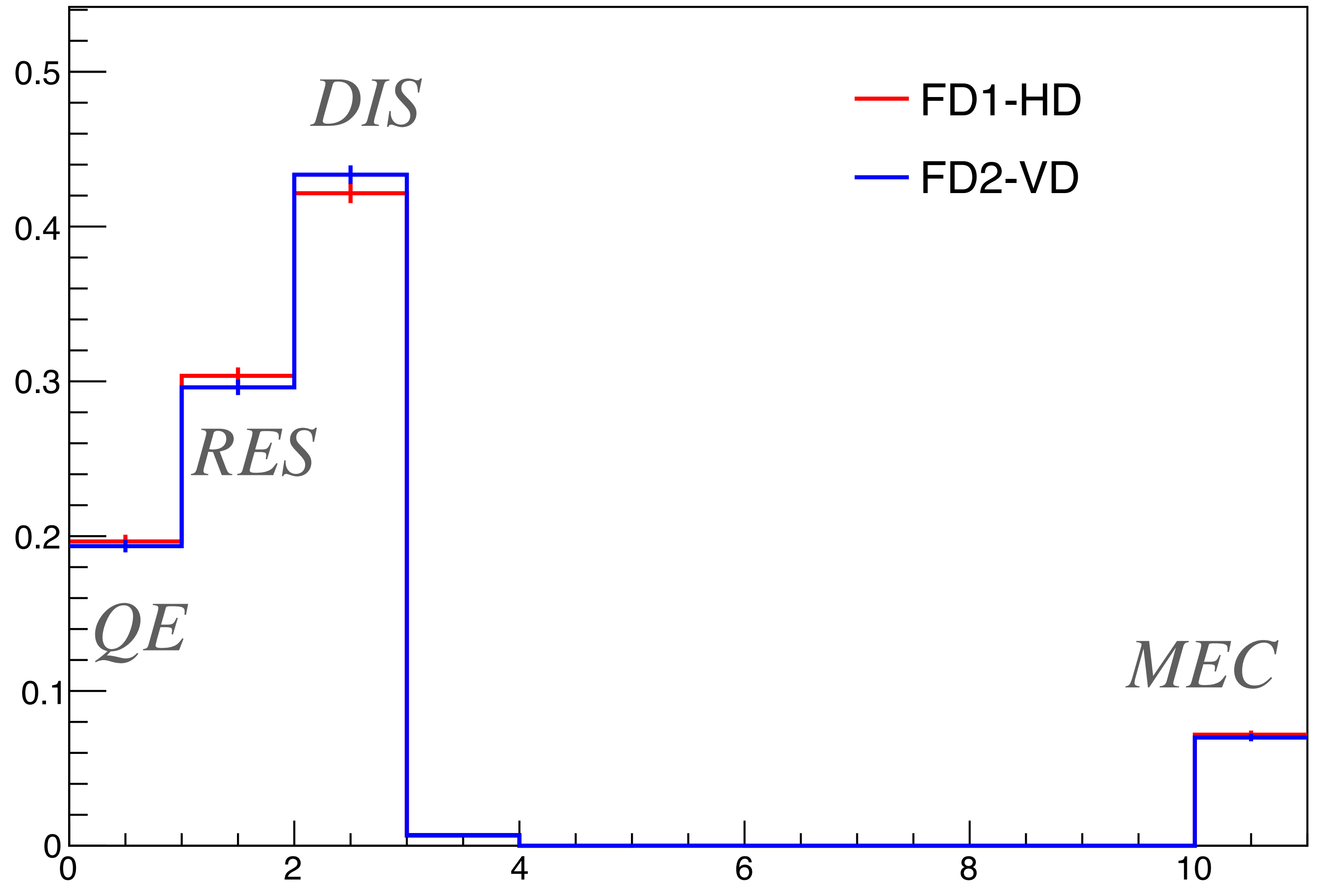
Training Composition

RHC - ν_μCC

No Cuts (Out of the Box)



FV Cuts



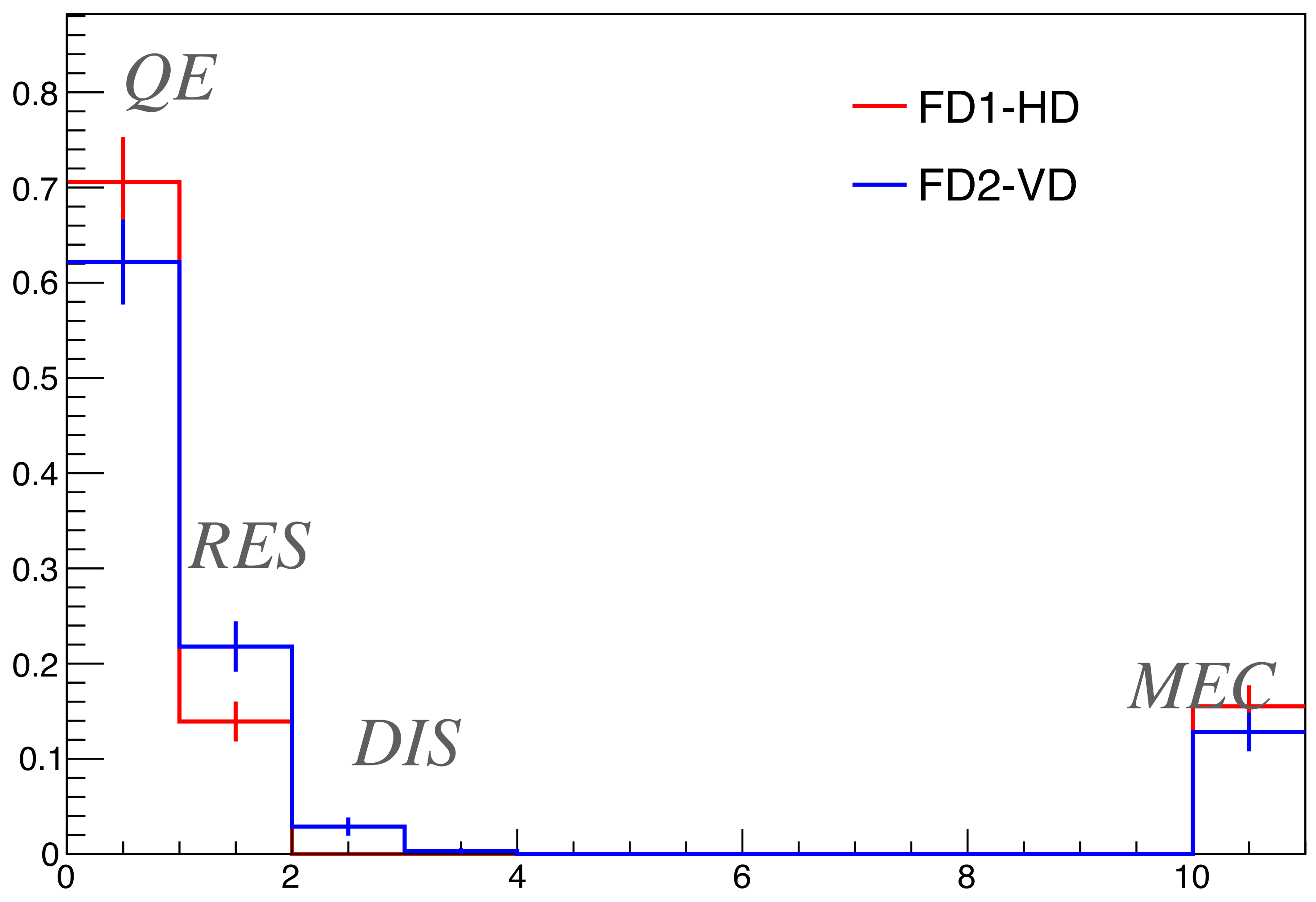
- No real differences in composition for RHC ν_μCC
- More plots : https://www.phy.bnl.gov/~nitish/random/cvnpresel/anu/

Training Composition

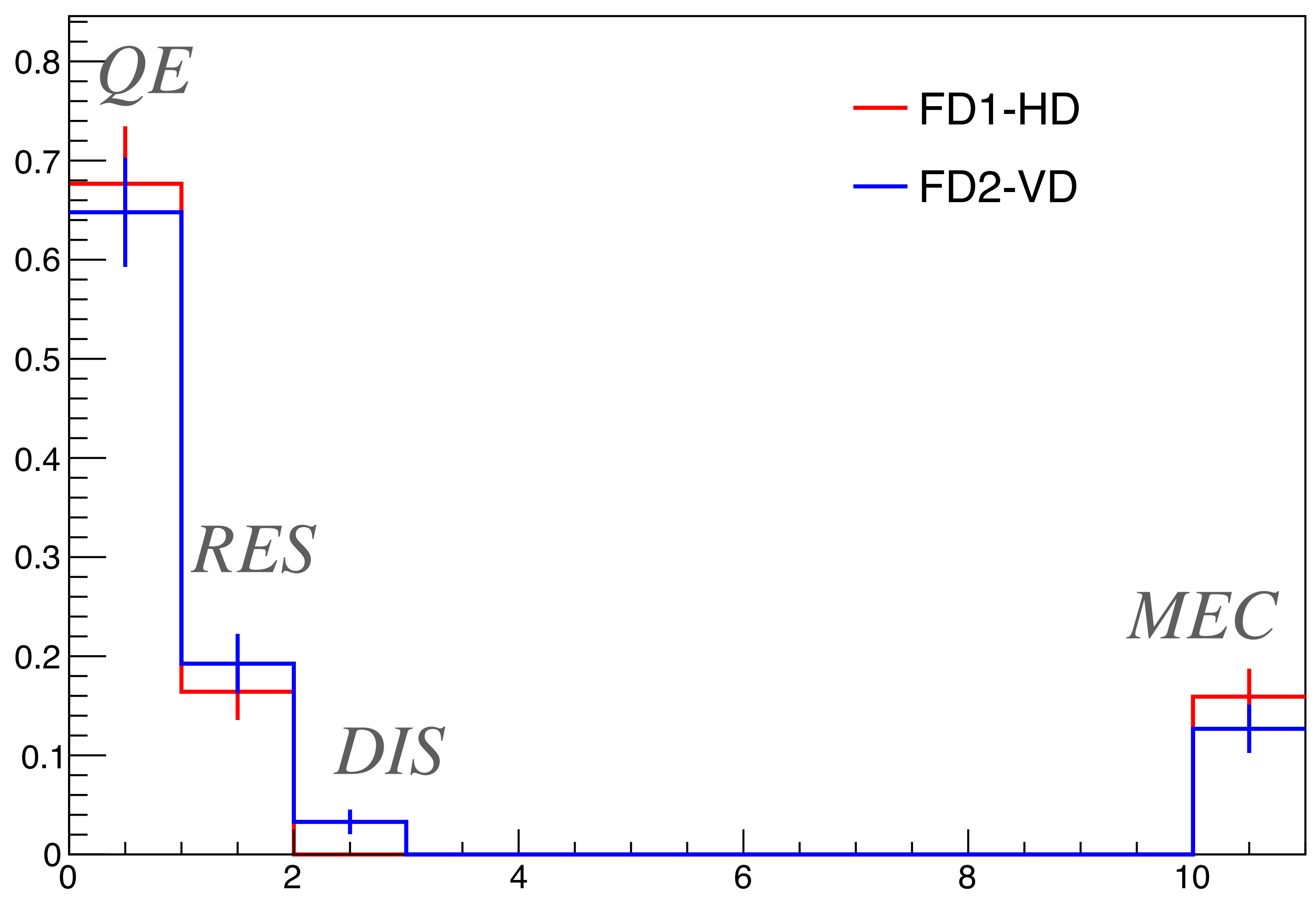
RHC - ν_μCC

E_ν < 1 GeV

No Cuts (Out of the Box)



FV Cuts



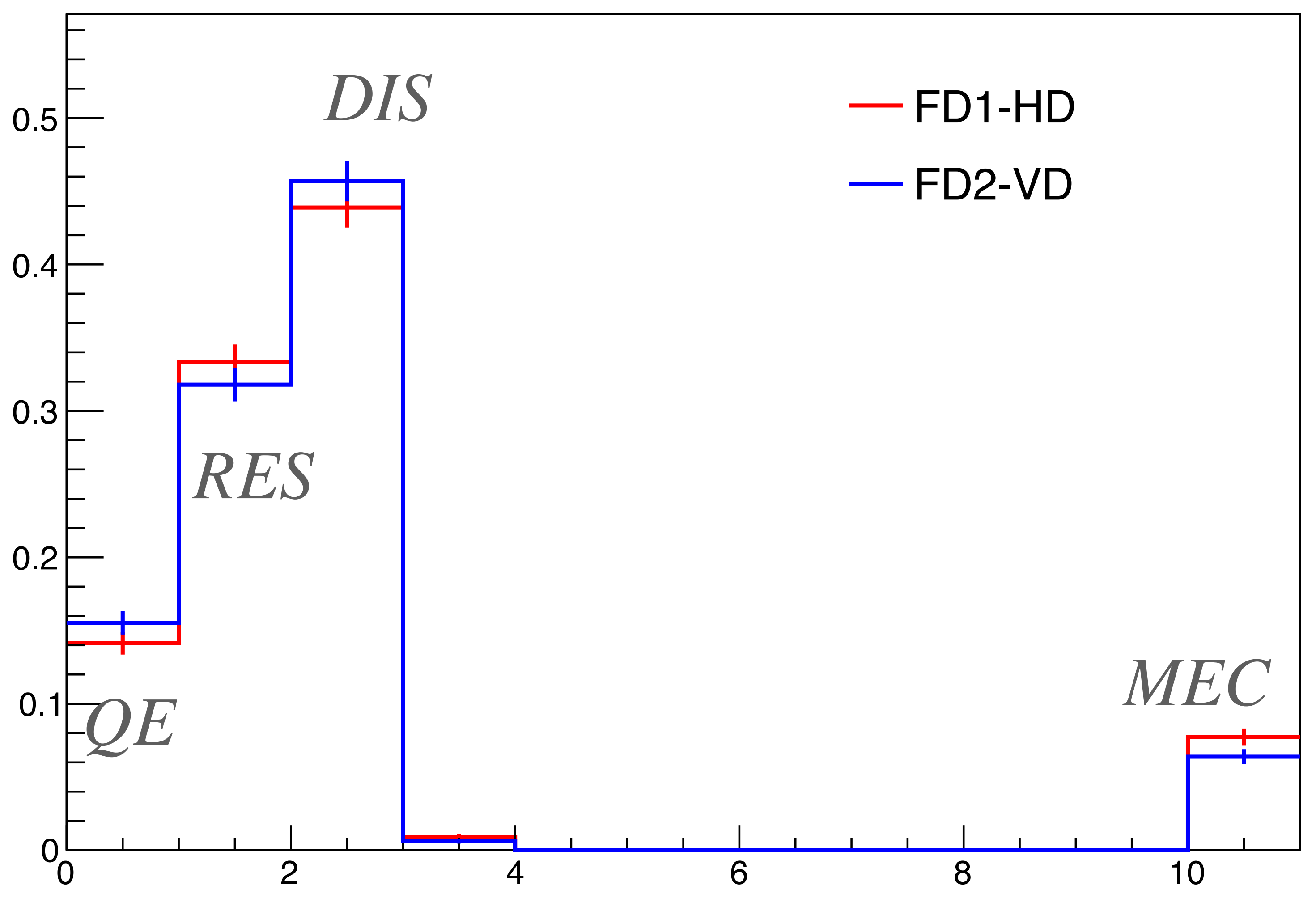
- No real differences in composition for RHC ν_μCC
- More plots : https://www.phy.bnl.gov/~nitish/random/cvnpresel/anu/

Training Composition

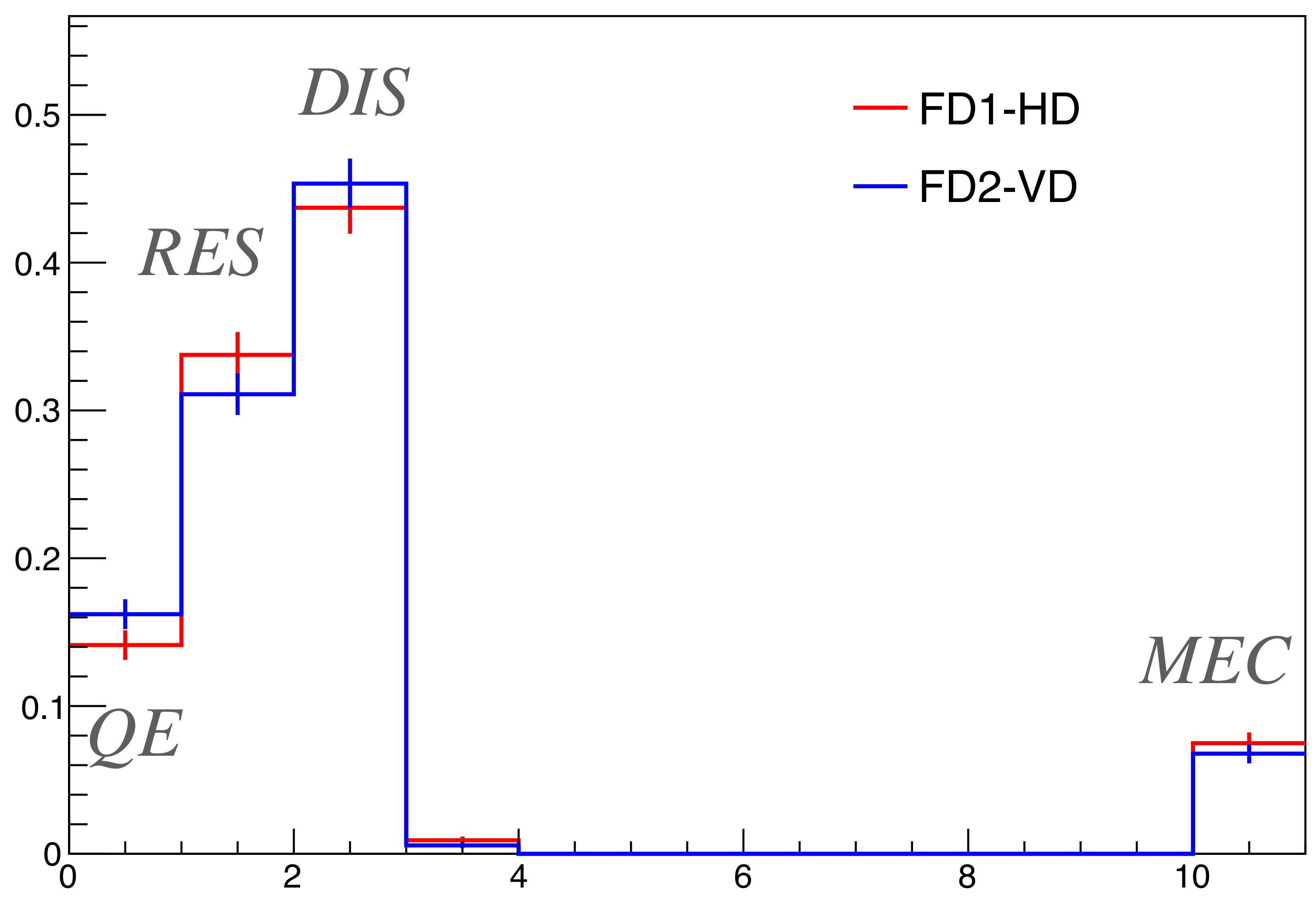
RHC - ν_μCC

E_μ < 1 GeV

No Cuts (Out of the Box)



FV Cuts



- No real differences in composition for RHC ν_μCC
- More plots : https://www.phy.bnl.gov/~nitish/random/cvnpresel/anu/

Conclusions

- HD/VD performance finally mostly in line with each other — relative to previous comparisons
- Performance gap exists in FHC :
  - HD better for NumuCCs, especially at Enu < 1 GeV
  - VD better for NueCCs
- Performance gap much smaller in RHC
- RHC performs better than FHC as expected
- Looking into training sample :
  - VD fiducial volume cuts fetch more fraction of neutrinos, explains difference in training stats
  - In terms of differences in actual composition :
    - NueCC - HD does have a bit more DIS, less QE (even before FV cuts) in FHC. Not sure why.. RHC is much closer
      - Could explain performance differences
    - Rest are more or less similar between HD and VD. Cannot explain difference between HD and VD for NumuCCs in FHC

Next Steps

- Plots vs muon track length :
  - Added this information to preprocessing step and ran jobs
  - CVN framework produces event by event text files containing this information
  - Have to match previous FHC validation sample to new text files (using run/subrun/event information) to get efficiency vs muon length for those samples
  - Cumbersome for ~million events, didn’t converge yet
  - Hopefully will have them ~this week
- Model trainings are in a good enough state to integrate I think
  - Would be good to understand remaining differences but don’t think its a sign of any major issues