# Reconstruction & Calibration Update

*June 3rd 2025*

Matthew Gignac & Nathan Baltzell

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

# Introduction

- Focus of reconstruction efforts has been on preparing and improving performance of 2021 dataset
  - Alignment of SVT detector (see talk tomorrow)
  - Implemented and tested data reduction scheme
  - Assessment of track reconstruction efficiency
  - Stability of track and cluster times as a function of run
  - Derived vertex (x,y) positions from 2D beam spot fits
- Throughout the process, we identified (and largely fixed) various issues with software and specific runs
- Moller mass analysis from recent reconstruction of the 10% Moller run dataset

# Data reduction

- Data volume from the 1% production over Christmas was very large (40 TB) and clearly needed to be reduced
- Approached from two directions:
  - **Reduce event size**
    - Hit containers the largest offender, but were being used in hpstr to compute hit layer and multiplicity
      - Reworked to save hit layers as a track property and unpacked when converting from LCIO to ROOT
    - Removed all other unnecessary collections
  - **Reduce number of events**
    - Skim events, based on V0s, Mollers, FEEs, etc..
    - V0 skimming has been ~validated

# Skimming validation

SLAC

- Skimmed and non-skimmed outputs were saved to allow validation of the skimming infrastructure and cuts
- Applying preselection to skimmed data returns fewer events than running on no-skim data
  - Skimming cuts are not matching preselection cuts
  - Applying the skimming cuts above preselection yields agreement up to single event differences

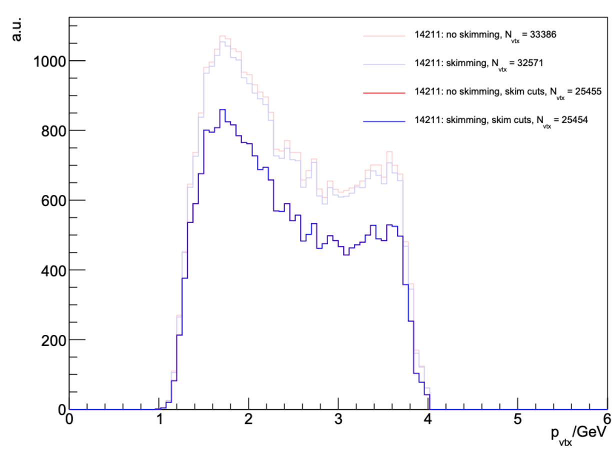
| cut | skimming | preselection |
|---|---|---|
| $E_{e^+,\mathrm{clu}}$ | – | $> 0.2$ GeV |
| $N_{\mathrm{2D\ hits}}$ | $\geq 9$ | $\geq 9$ |
| $\chi^2_{\mathrm{vtx}}$ | $< 30.0$ | $< 20.0$ |
| $p_{\mathrm{sum}}$ | $< 4.5$ GeV | $< 4.0$ GeV |
| $p_{e^-}$ | $< 4.5$ GeV | $< 2.9$ GeV |
| | – | $> 0.4$ GeV |
| $p_{e^+}$ | $< 4.5$ GeV | – |
| | – | $> 0.4$ GeV |
| $\Delta(t_{\mathrm{trk},e^-}, t_{\mathrm{trk},e^+})$ | $< 20.0$ ns | – |
| $\Delta(t_{\mathrm{trk},e^-}, t_{\mathrm{clu},e^+})$ | – | $< 6.9$ ns |
| $\Delta(t_{\mathrm{trk},e^+}, t_{\mathrm{clu},e^+})$ | – | $< 6.0$ ns |
| $\chi^2_{\mathrm{trk},e^-}$ | $< 80.0$ | – |
| $\chi^2_{\mathrm{trk},e^+}$ | $< 80.0$ | – |
| $\chi^2_{\mathrm{trk},e^-}/\mathrm{ndf}$ | – | $< 20.0$ |
| $\chi^2_{\mathrm{trk},e^+}/\mathrm{ndf}$ | – | $< 20.0$ |

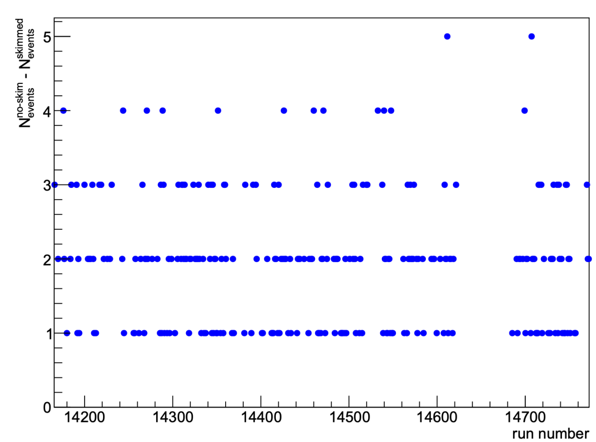
| | efficiency | | | | | |
|---|---|---|---|---|---|---|
| | 14211 | | 14487 | | 14611 | |
| cut | no-skim | skimmed | no-skim | skimmed | no-skim | skimmed |
| $\Delta(t_{\mathrm{trk},e^-}, t_{\mathrm{trk},e^+})$ | 99.83 % | 99.83 % | 99.83 % | 99.83 % | 99.82 % | 99.81 % |
| $\chi^2_{\mathrm{trk},e^-}$ | 89.00 % | 89.83 % | 89.88 % | 90.71 % | 87.92 % | 88.84 % |
| $\chi^2_{\mathrm{trk},e^+}$ | 85.93 % | 87.31 % | 87.22 % | 88.41 % | 87.80 % | 88.89 % |
| total | 76.24 % | 78.15 % | 78.24 % | 80.01 % | 77.00 % | 78.75 % |
| $N_{\mathrm{events}}$ | | | | | | |
| preselected | 33 386 | 32 571 | 33 316 | 32 578 | 32 240 | 31 520 |
| skim cuts | 25 455 | 25 454 | 26 068 | 26 066 | 24 824 | 24 822 |

# Skimming validation

- Skimmed and non-skimmed outputs were saved to allow validation of the skimming infrastructure and cuts
- Applying preselection to skimmed data returns fewer events than running on no-skim data
  - Skimming cuts are not matching preselection cuts
  - Applying the skimming cuts above preselection yields agreement up to single event differences

| cut | skimming | preselection |
|---|---|---|
| $E_{e^+,\mathrm{clu}}$ | – | $> 0.2$ GeV |
| $N_{\mathrm{2D\ hits}}$ | $\geq 9$ | $\geq 9$ |
| $\chi^2_{\mathrm{vtx}}$ | $< 30.0$ | $< 20.0$ |
| $p_{\mathrm{sum}}$ | $< 4.5$ GeV | $< 4.0$ GeV |
| $p_{e^-}$ | $< 4.5$ GeV | $< 2.9$ GeV |
| | – | $> 0.4$ GeV |
| $p_{e^+}$ | $< 4.5$ GeV | – |
| | – | $> 0.4$ GeV |
| $\Delta(t_{\mathrm{trk},e^-}, t_{\mathrm{trk},e^+})$ | $< 20.0$ ns | – |
| $\Delta(t_{\mathrm{trk},e^-}, t_{\mathrm{clu},e^+})$ | – | $< 6.9$ ns |
| $\Delta(t_{\mathrm{trk},e^+}, t_{\mathrm{clu},e^+})$ | – | $< 6.0$ ns |
| $\chi^2_{\mathrm{trk},e^-}$ | $< 80.0$ | – |
| $\chi^2_{\mathrm{trk},e^+}$ | $< 80.0$ | – |
| $\chi^2_{\mathrm{trk},e^-}/\mathrm{ndf}$ | – | $< 20.0$ |
| $\chi^2_{\mathrm{trk},e^+}/\mathrm{ndf}$ | – | $< 20.0$ |



# Skimming validation

- Skimmed and non-skimmed outputs were saved to allow validation of the skimming infrastructure and cuts
- Applying preselection to skimmed data returns fewer events than running on no-skim data
  - Skimming cuts are not matching preselection cuts
  - Applying the skimming cuts above preselection yields agreement up to single event differences

| cut | skimming | preselection |
|---|---|---|
| $E_{e^+,\mathrm{clu}}$ | – | $> 0.2$ GeV |
| $N_{\mathrm{2D\ hits}}$ | $\geq 9$ | $\geq 9$ |
| $\chi^2_{\mathrm{vtx}}$ | $< 30.0$ | $< 20.0$ |
| $p_{\mathrm{sum}}$ | $< 4.5$ GeV | $< 4.0$ GeV |
| $p_{e^-}$ | $< 4.5$ GeV | $< 2.9$ GeV |
| | – | $> 0.4$ GeV |
| $p_{e^+}$ | $< 4.5$ GeV | – |
| | – | $> 0.4$ GeV |
| $\Delta(t_{\mathrm{trk},e^-}, t_{\mathrm{trk},e^+})$ | $< 20.0$ ns | – |
| $\Delta(t_{\mathrm{trk},e^-}, t_{\mathrm{clu},e^+})$ | – | $< 6.9$ ns |
| $\Delta(t_{\mathrm{trk},e^+}, t_{\mathrm{clu},e^+})$ | – | $< 6.0$ ns |
| $\chi^2_{\mathrm{trk},e^-}$ | $< 80.0$ | – |
| $\chi^2_{\mathrm{trk},e^+}$ | $< 80.0$ | – |
| $\chi^2_{\mathrm{trk},e^-}/\mathrm{ndf}$ | – | $< 20.0$ |
| $\chi^2_{\mathrm{trk},e^+}/\mathrm{ndf}$ | – | $< 20.0$ |

- Changed definition of the chi2 cut in the skimming to divide through by nDOF: https://github.com/JeffersonLab/hps-java/pull/1098

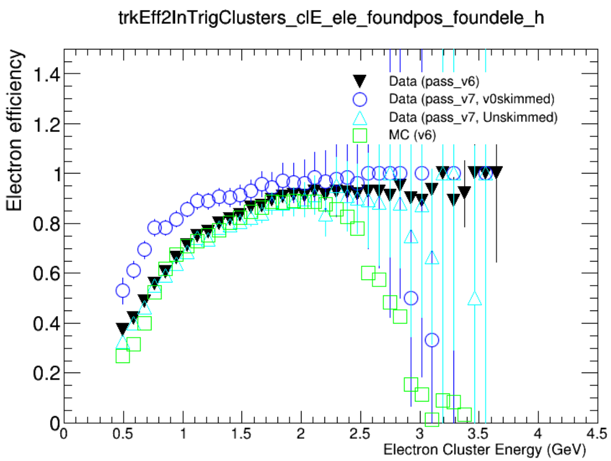
# Track reconstruction efficiency

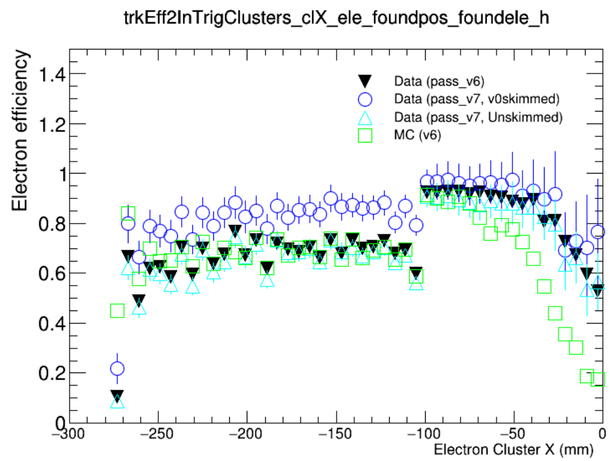
- Method developed and documented by M. Graham in the 2016 Physics Run studies (link)
- Uses the ECal to select events that look consistent with a 2-prong (e+e−) trident event and has at least one track pointing to a the cluster in the ECal.
- Track matched in the ECal to ”tag” the event as a likely (e+e−) event and then use the other ECal cluster to “probe” the track efficiency on the other side.
- Efficiency defined as:

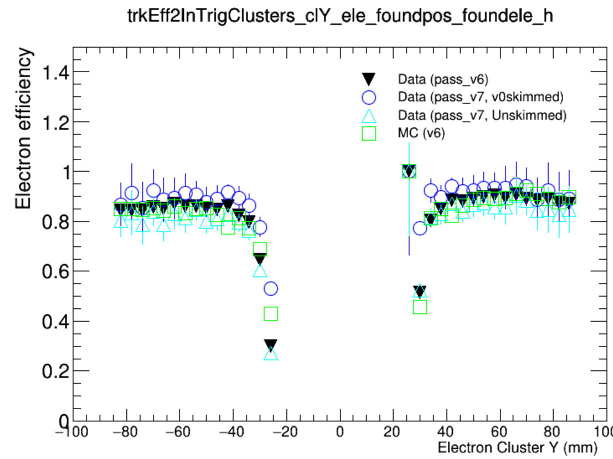
$$\epsilon(E/X/Y) = \frac{N(matched\ probe\ track)}{N(tag\ events)}$$

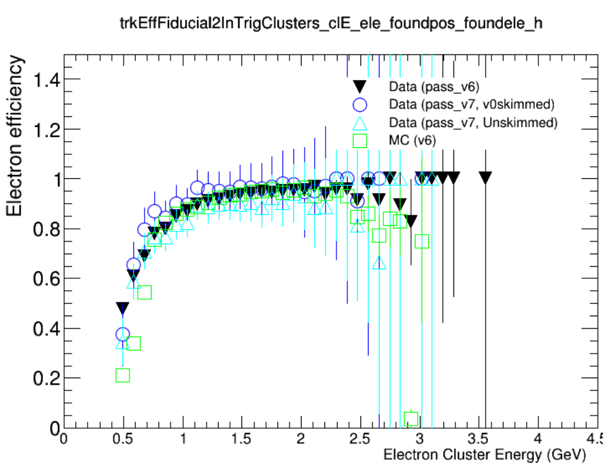
# Track reconstruction efficiency in v7

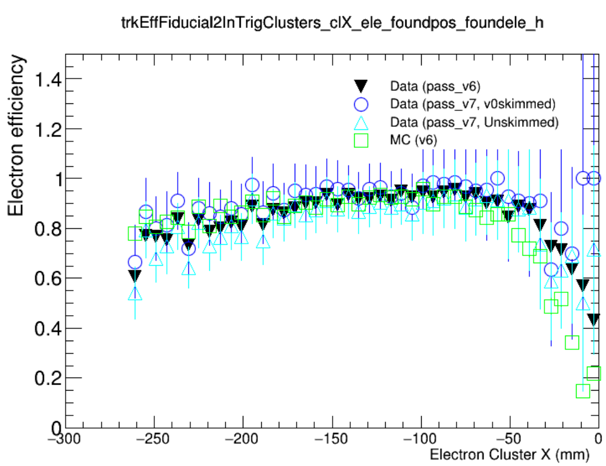
- Comparable to v6 — good! Higher efficiency for skimmed outputs, but agree well after fiducial selections (bottom)

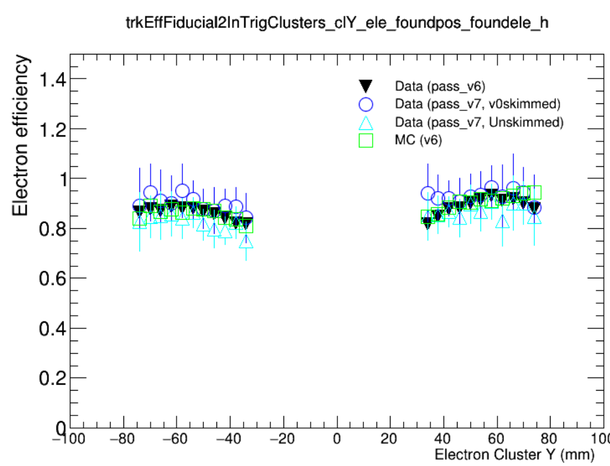


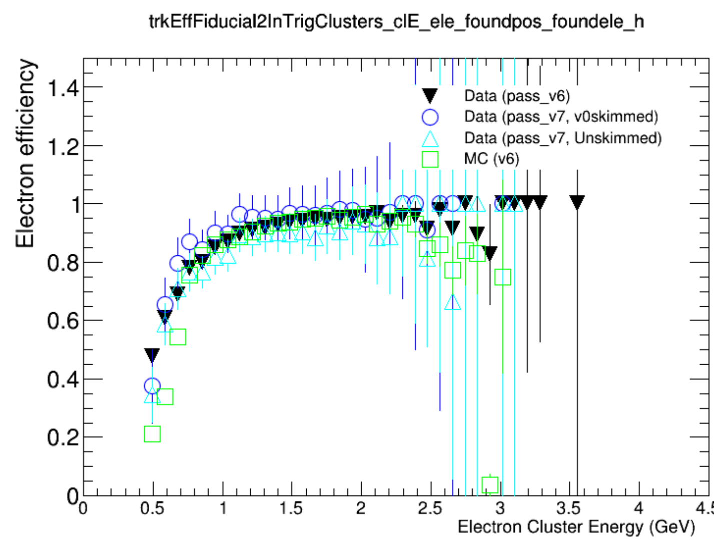



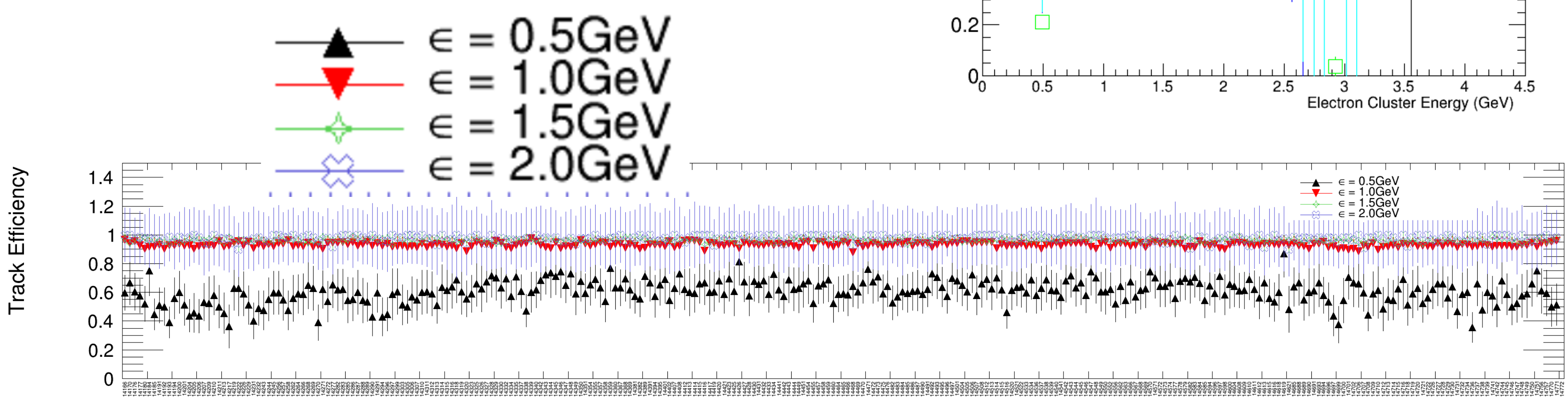



# Track reconstruction efficiency in v7

- Comparable to v6 — good! Higher efficiency for skimmed outputs, but agree well after fiducial selections (bottom
- Stable as a function of run!
  - Efficiency plotted for various track momentum benchmarks
  - For p>1 GeV, maintain efficiency close to 1 across entire datasets

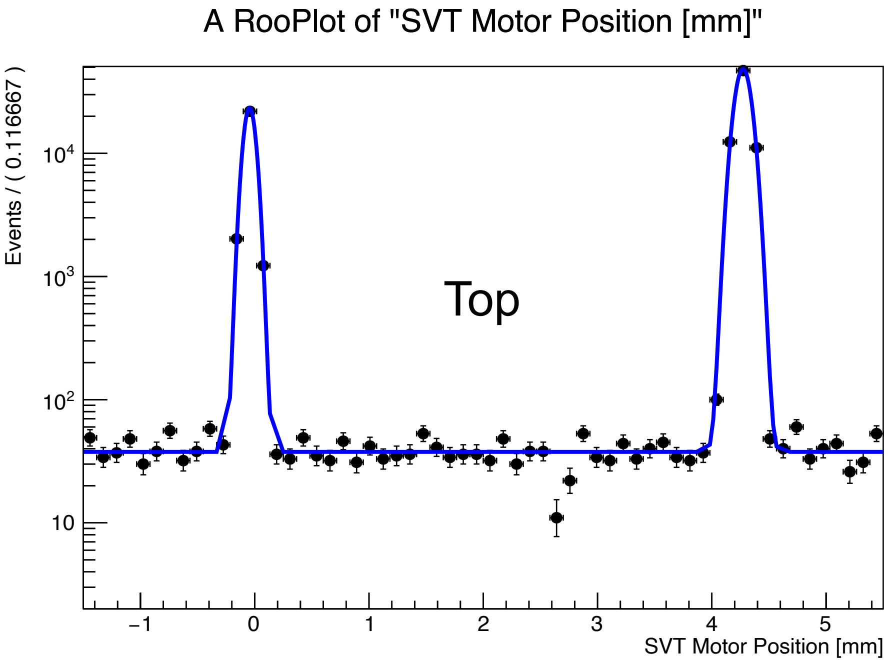


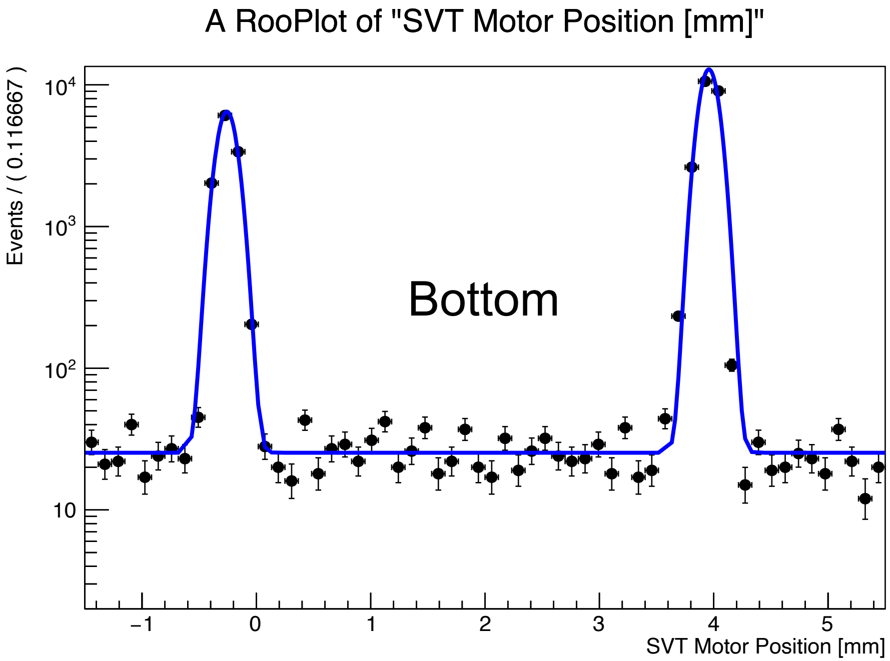




*All runs in 2021 dataset*

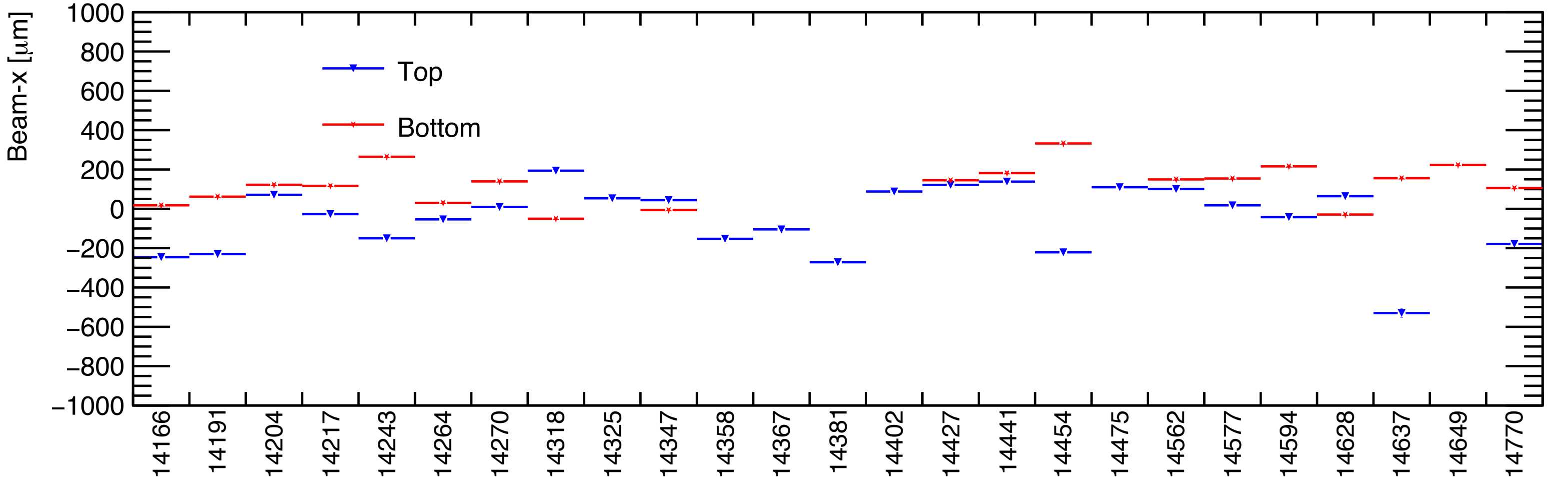
# Determining the (x,y) beam position

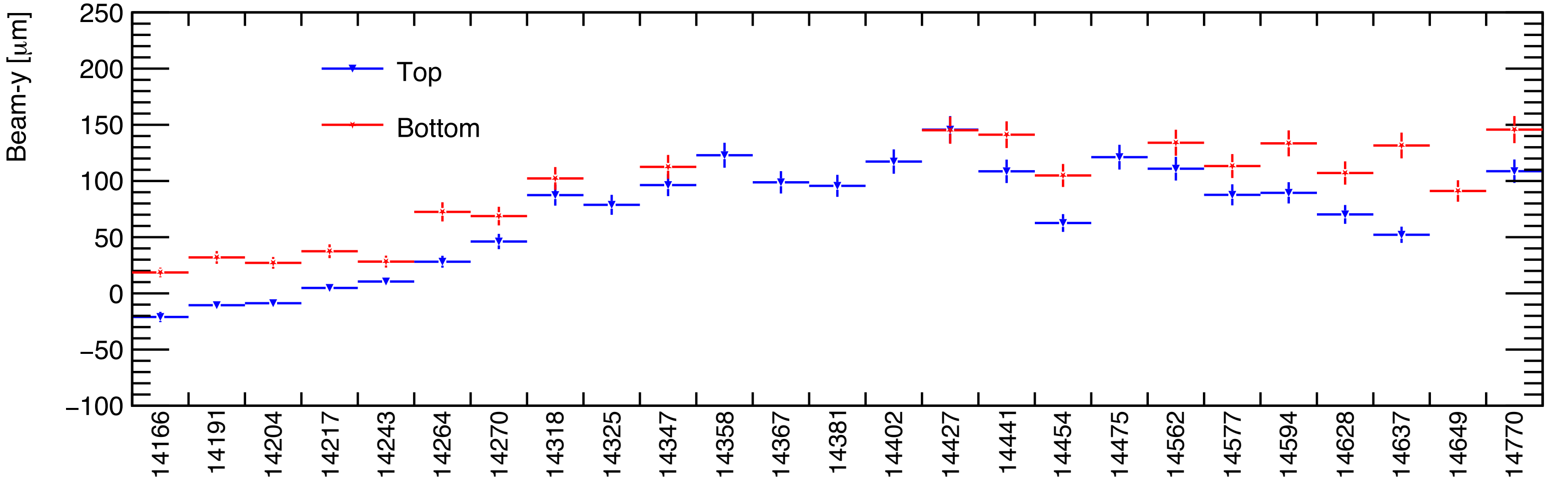
- SVT wire scan data was taken sporadically throughout the 2021 dataset: useful to determine beam position
- Fits to counters in "HPS_SC" as a function of the SVT motor position, separately for top & bottom



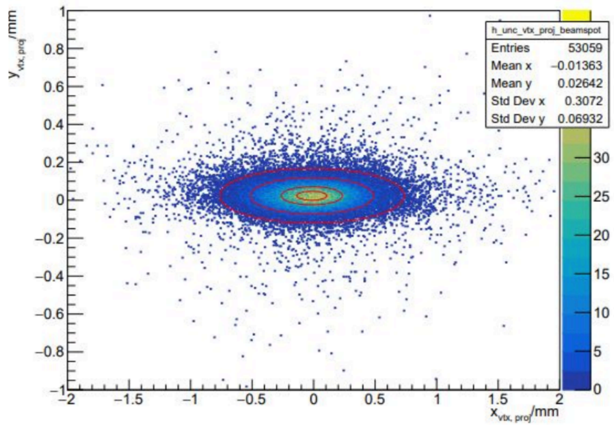
# Beam position by physics run number

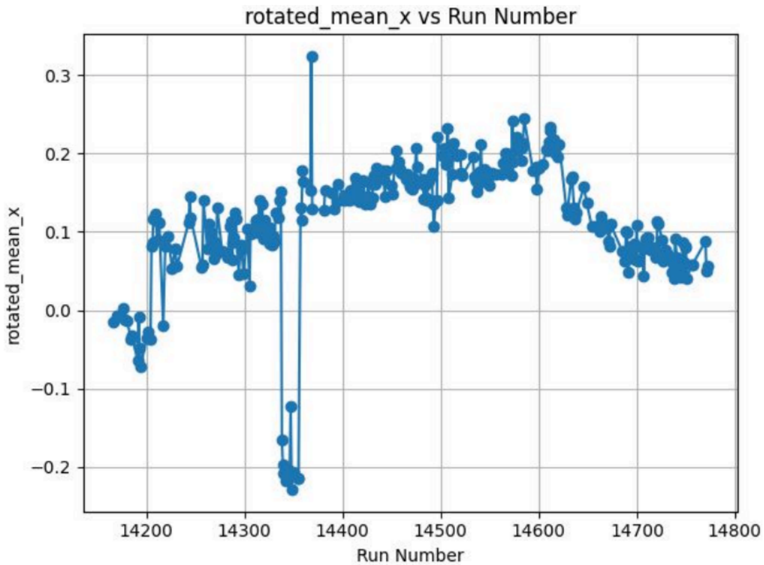


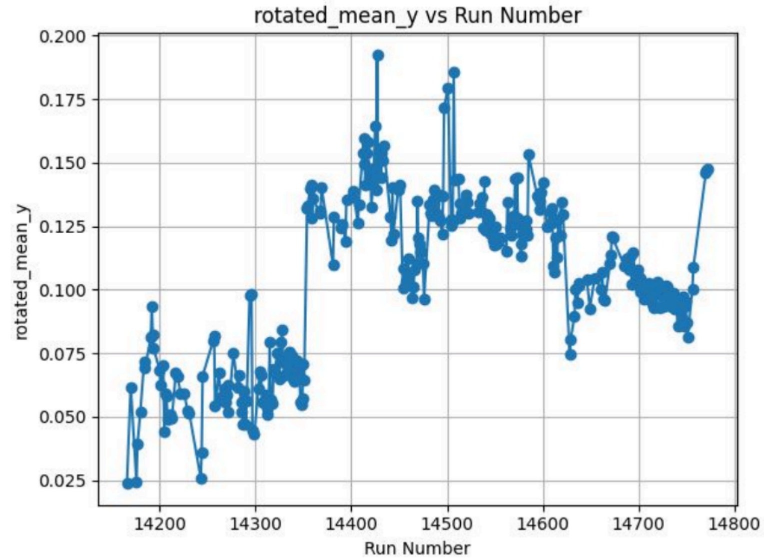



# X-Y beam spot determination

- Beam spot in (x,y) determined from 2D fit
  - Summarized in json format and used as input for analysis quantities
  - Added to (local) conditions database, which will be used as input for next pass for BSC and TC vertex fits

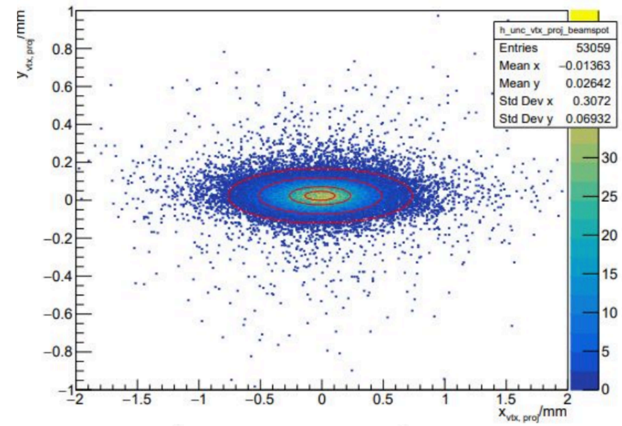


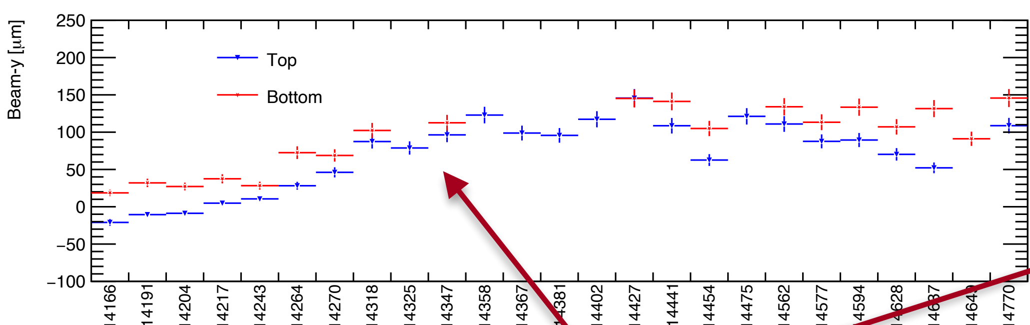



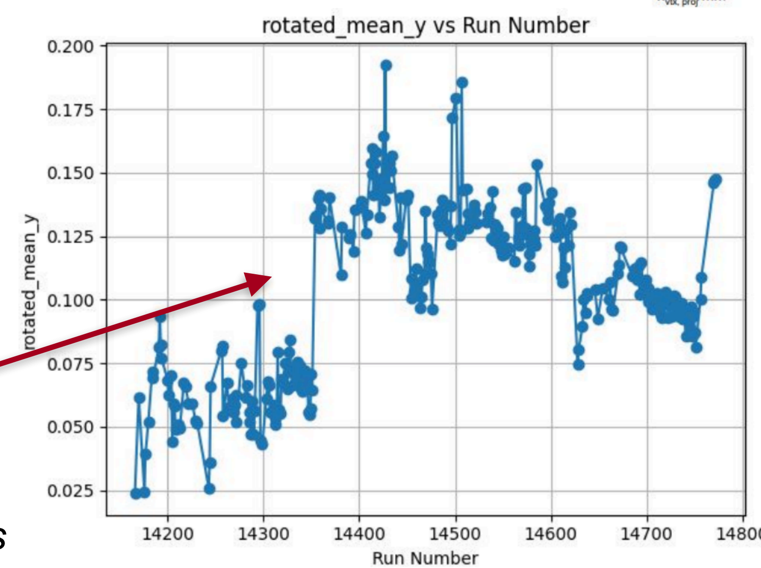



# X-Y beam spot determination

- Beam spot in (x,y) determined from 2D fit
  - Summarized in json format and used as input for analysis quantities
  - Added to (local) conditions database, which will be used as input for next pass for BSC and TC vertex fits



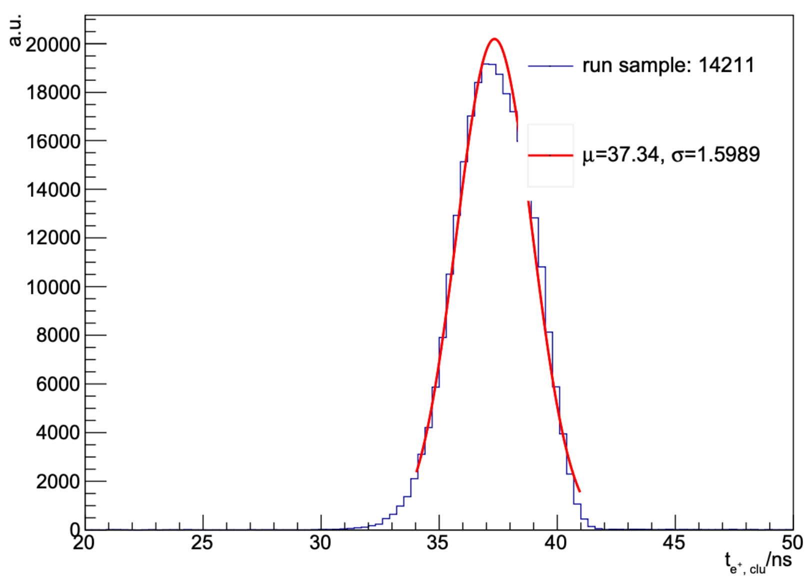



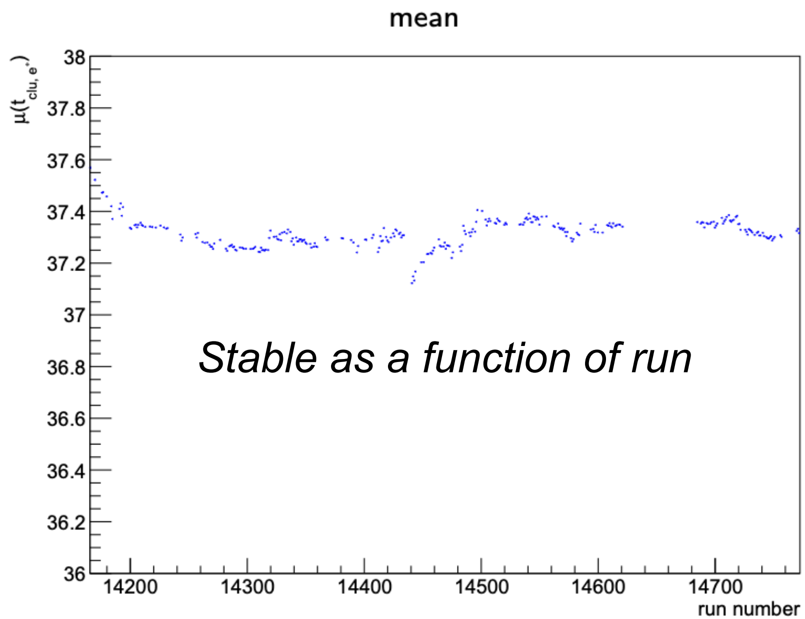




*Similar trends seen in wire-scan vs physics vertex positions*

# Track & cluster times

- Track and cluster times very important quantity at analysis level — studied as a function of run number
  - Fit a gaussian to the data to extract mean and sigma, and used to adjust mean value to zero

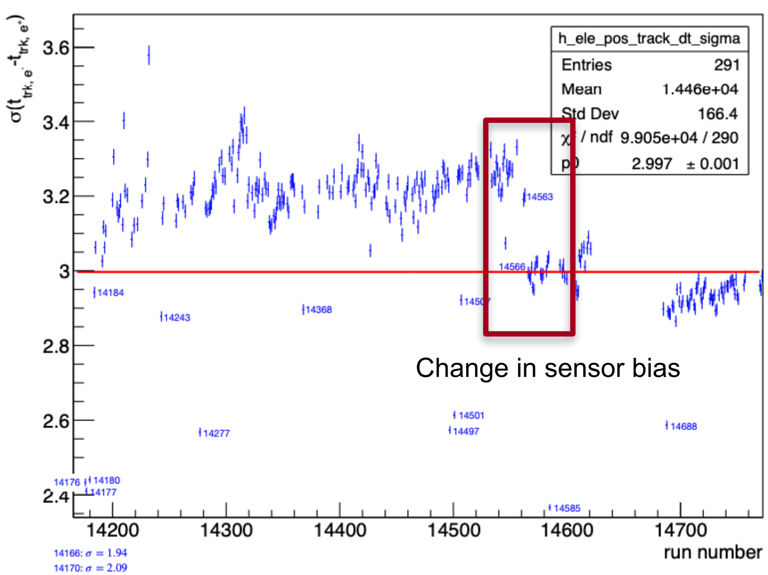


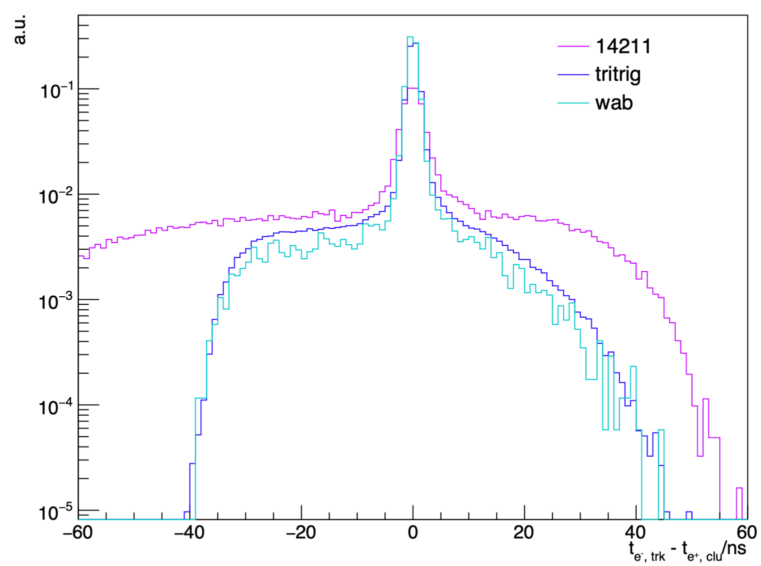




*Example showing positron cluster times*

# Track time

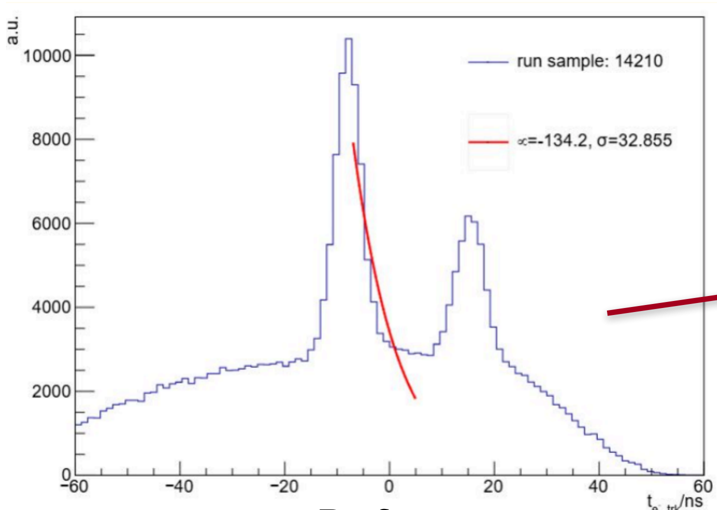
- Electron minus positron track times used in skimming and should be stable as a function of run number
- Observe much broader tails in data than MC for electron track to positron cluster time difference
  - Need to devise strategy to ideally reduce these contributions and/or estimate (see Sarah's talk)



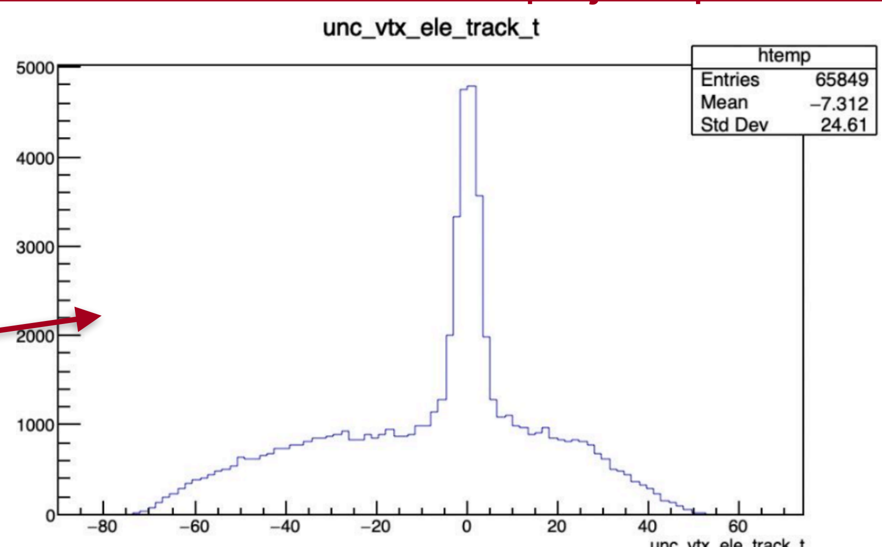



# Phase issue

- In validating pass2, uncovered a bimodal peak in the time distribution for a few runs: 14210 and 14232.
- Suspected phase offset problem, where phase and layer dependent time shifts are needed for each layer
  - This was a recurrence of a problem that is ~2-3 yr old
  - Promptly fixed: https://github.com/JeffersonLab/hps-java/pull/1099

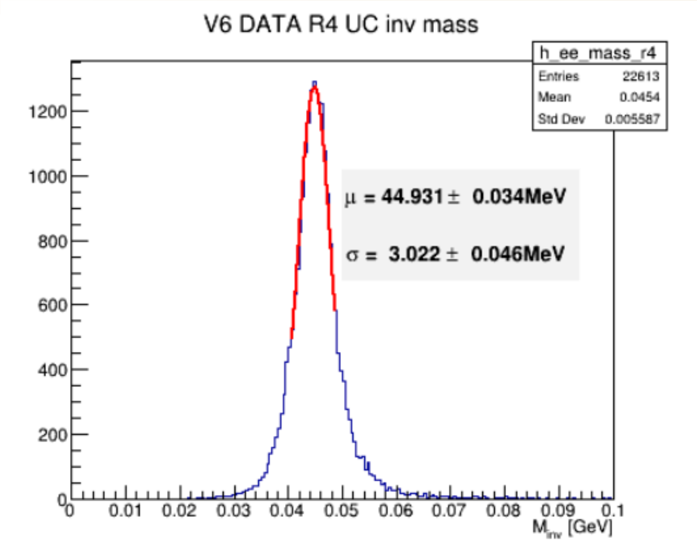



*Before*



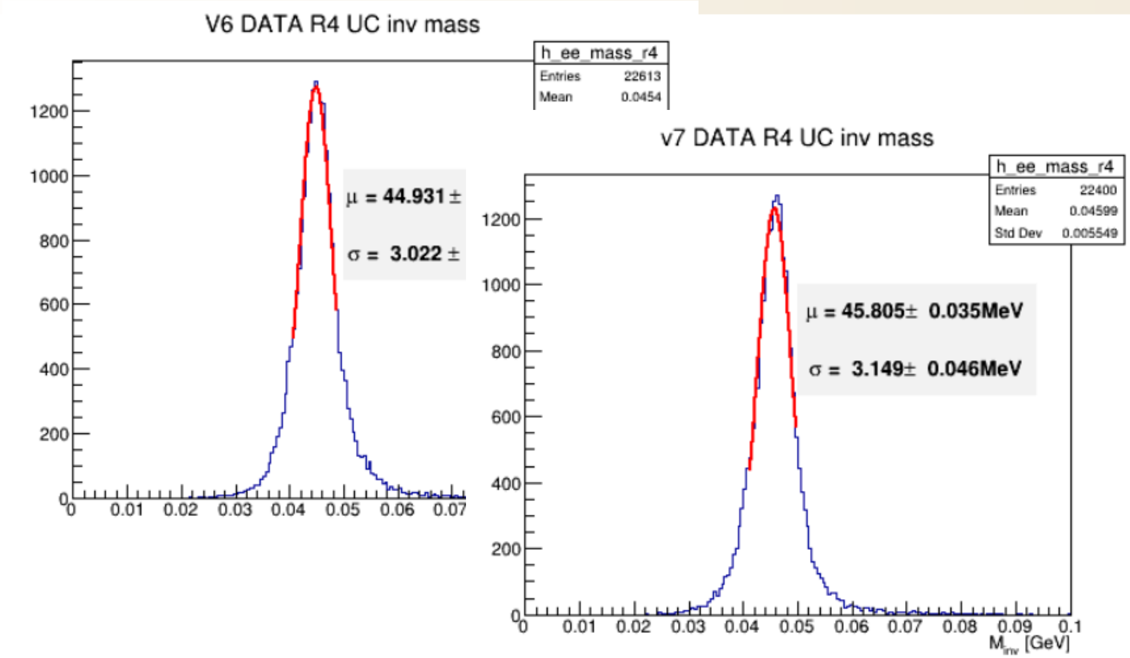

*After fix*

# Moller mass: through (a subset of) the versions



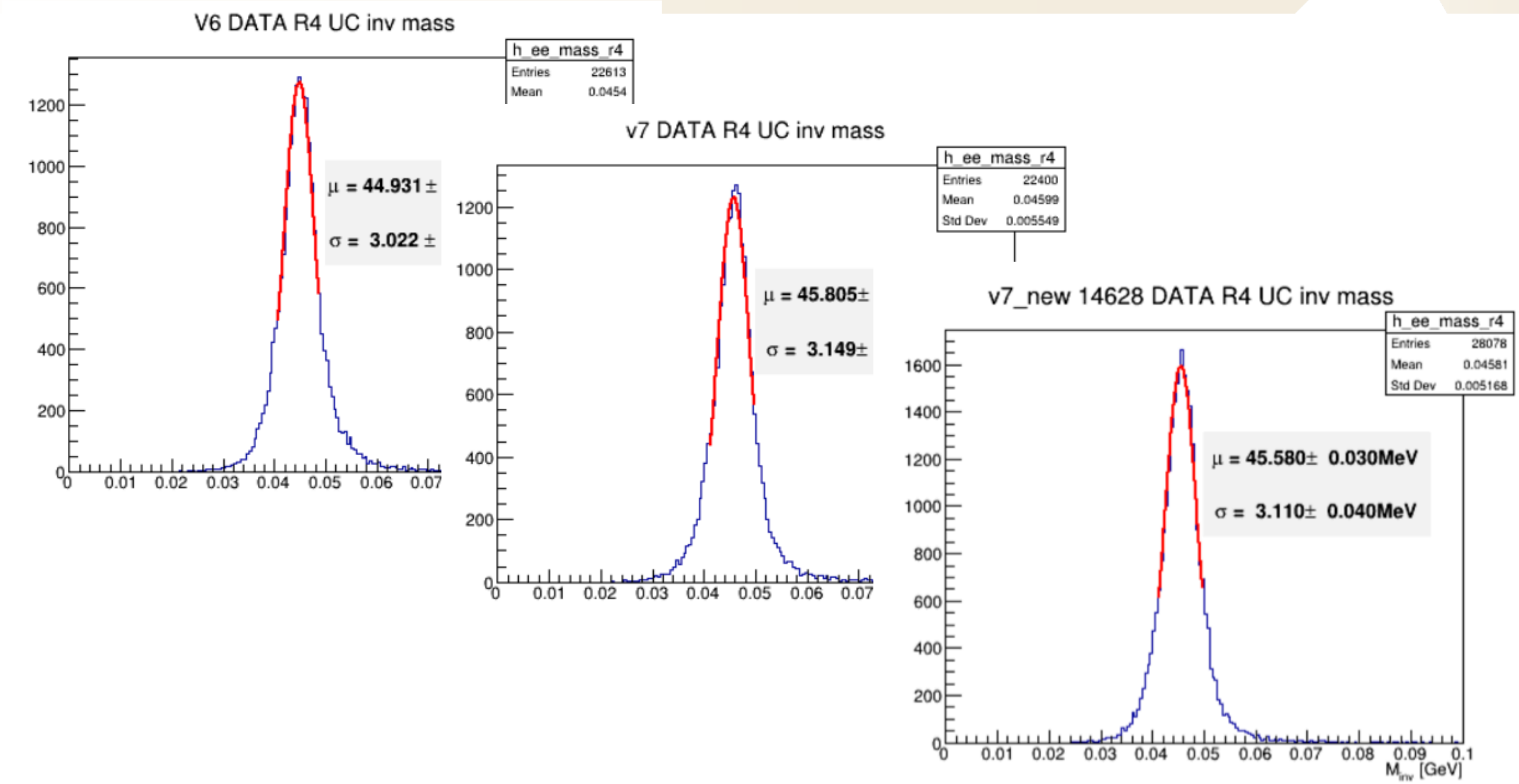

v6 alignment model
1% pass1 (Jan. 2025)

# Moller mass: through (a subset of) the versions



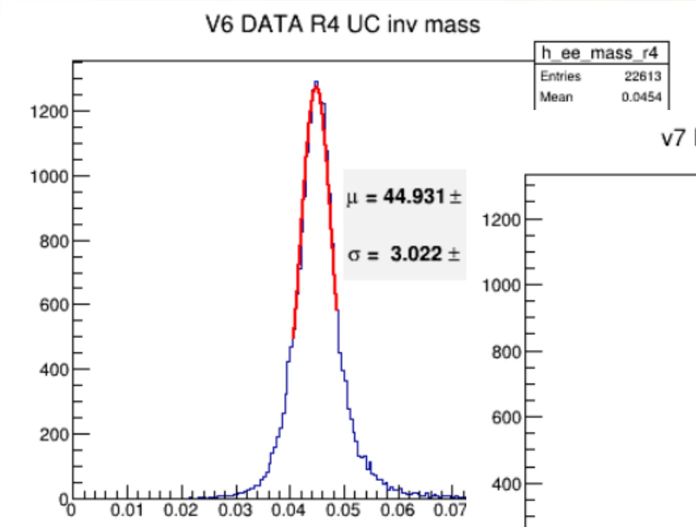

v7beta alignment model
Preparations for pass2 (March. 2025)

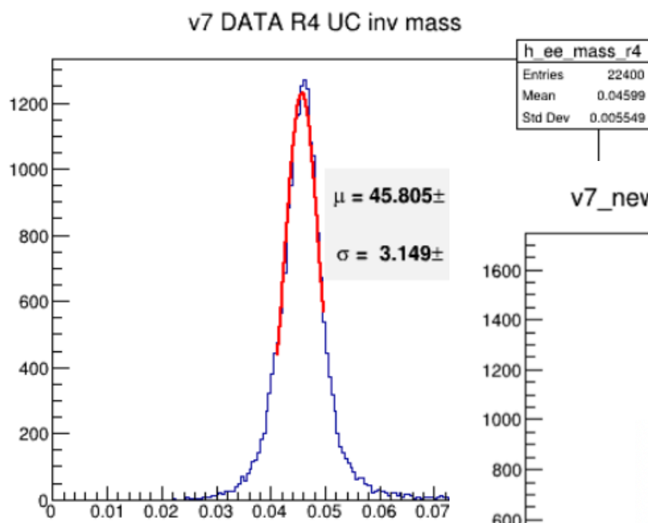
# Moller mass: through (a subset of) the versions

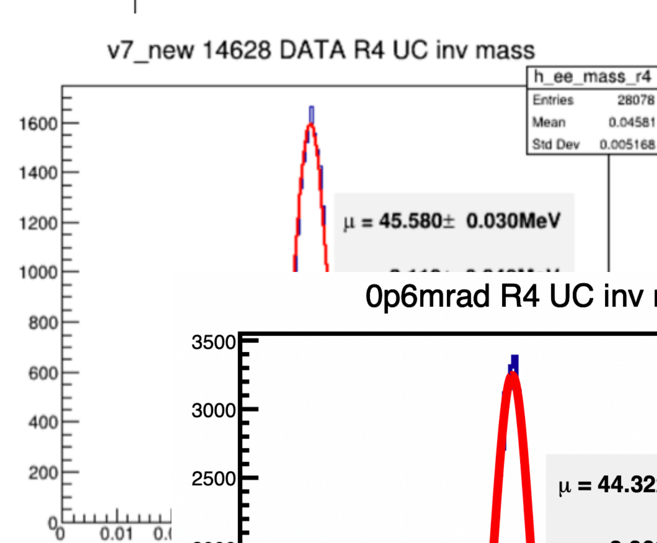


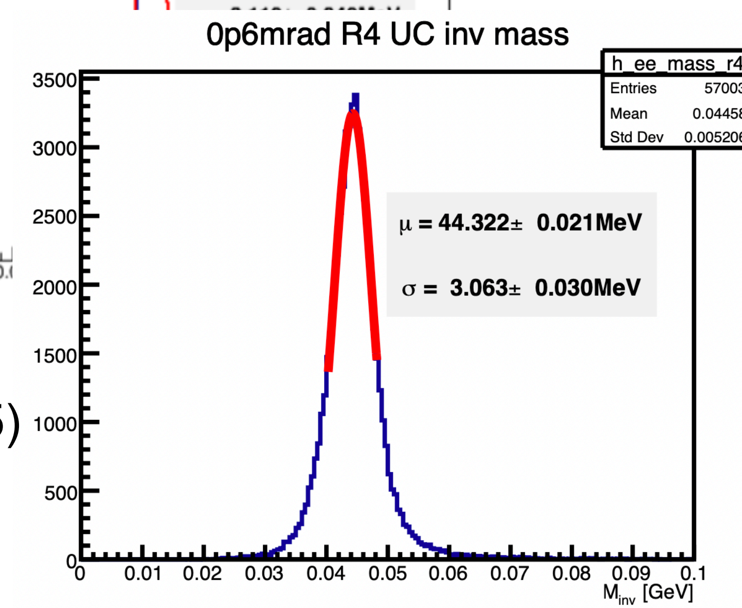

v7beta-2 alignment model
Preparations for pass2 (March. 2025)

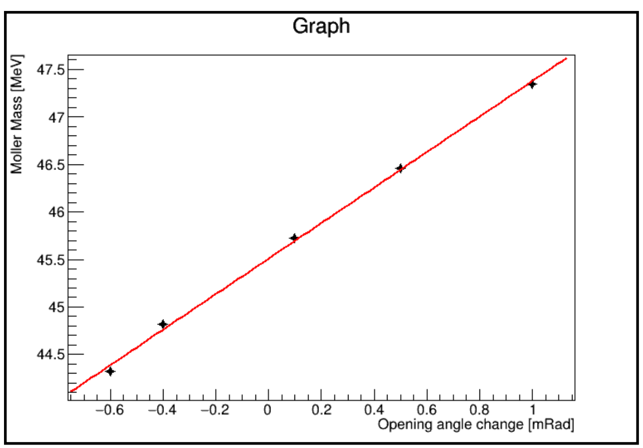
# Moller mass: through (a subset of) the versions



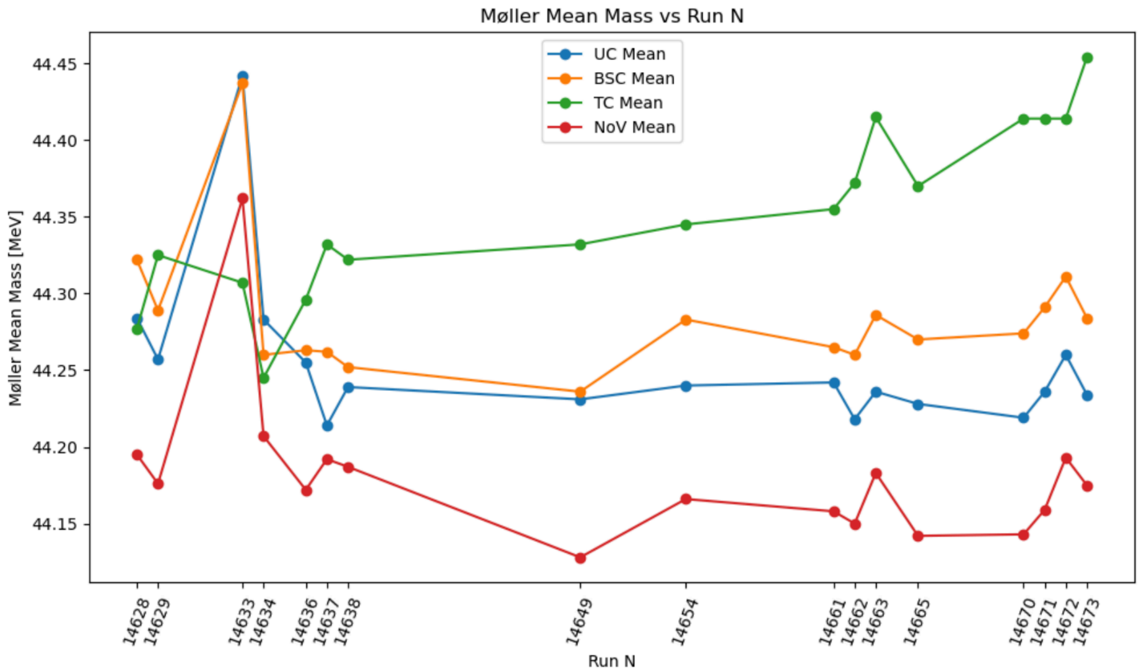




v7 alignment model
Pass2 (March. 2025)

# Moller mass studies: v7 alignment model

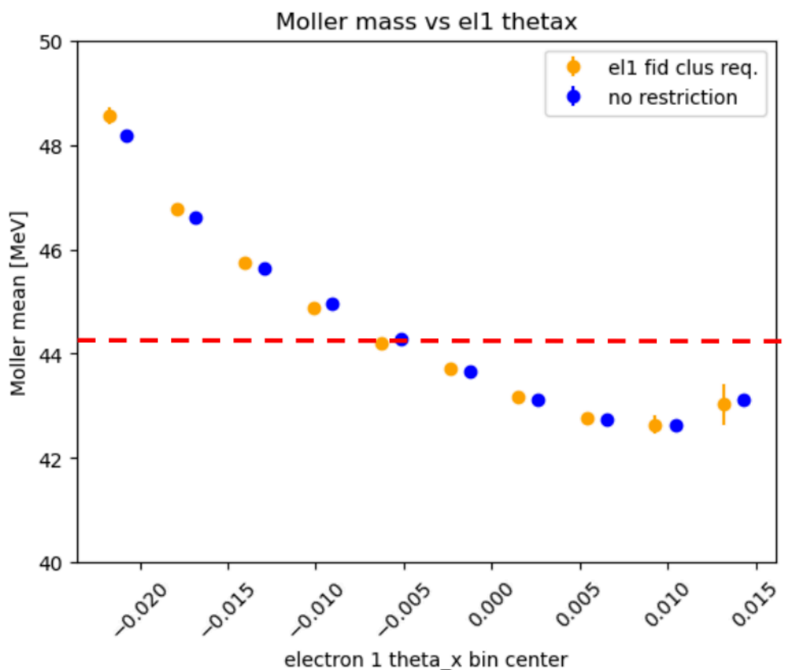
- Reconstructed 10% of Moller run dataset @1.92 GeV
- Average Moller mass extracted using BSC and TC vertices, showing good agreement UC vertex fits



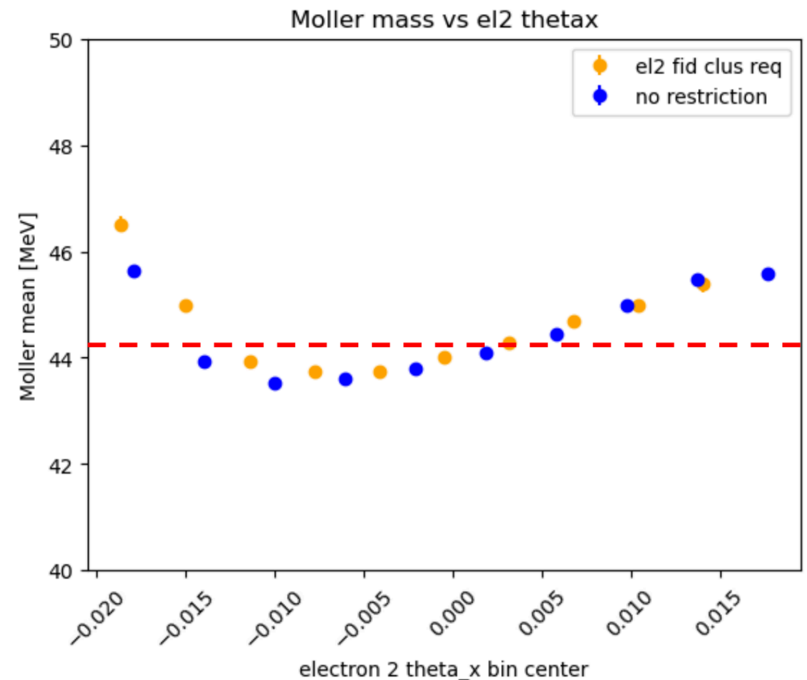

- Note: vertex (x,y,z) positions were taken from a single Moller run; small evolution in (x,y) already known. The z-vertex position disagrees between different approaches (z0 vs tanL, e+e-, e-e-, and multitrack give different results spanning ~1mm)

# Moller mass studies

- Reconstructed 10% of Moller run dataset @1.92 GeV
- Differential analysis: Moller mass vs theta_x shows strong dependency, especially in the top detector
  - More details in Lewis' talk on Wednesday morning




*Top*




*Bottom*

# Conclusions

- Lots of studies on the reconstructed ~0.3% of 2021 dataset and great progress towards the next pass
- Alignment work ongoing for v8
  - Lots of studies performed to understand origins of momentum scale as a function of phi0 and tanL
  - Beam spot constrained alignment technically working, but work needed to better understand vertex z-position
  - It may make sense to move to v8 for the next pass
    - <1% pass, also enabling flags needed for hit killing and smearing studies Elizabeth is working on
- Efforts starting on calibration work for 2019 dataset

# Questions

# Momentum scale: through the versions…