Training clusters

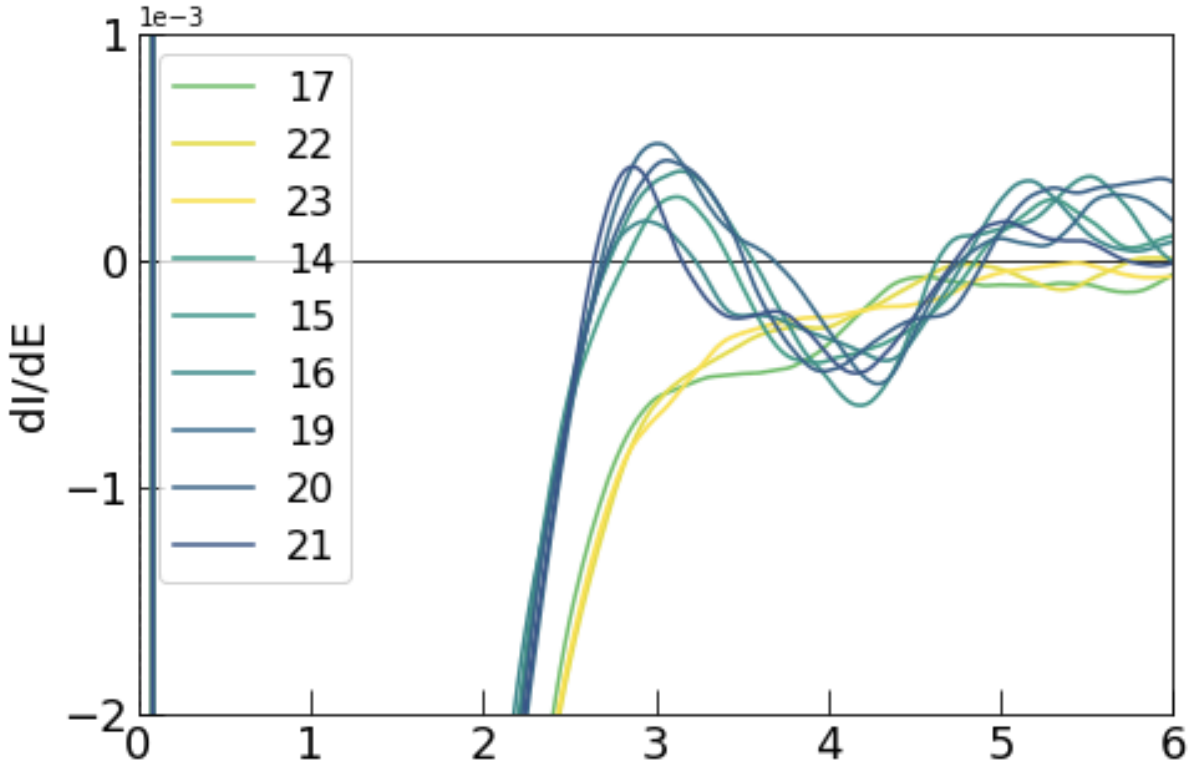
Data in clusters: filter outliers

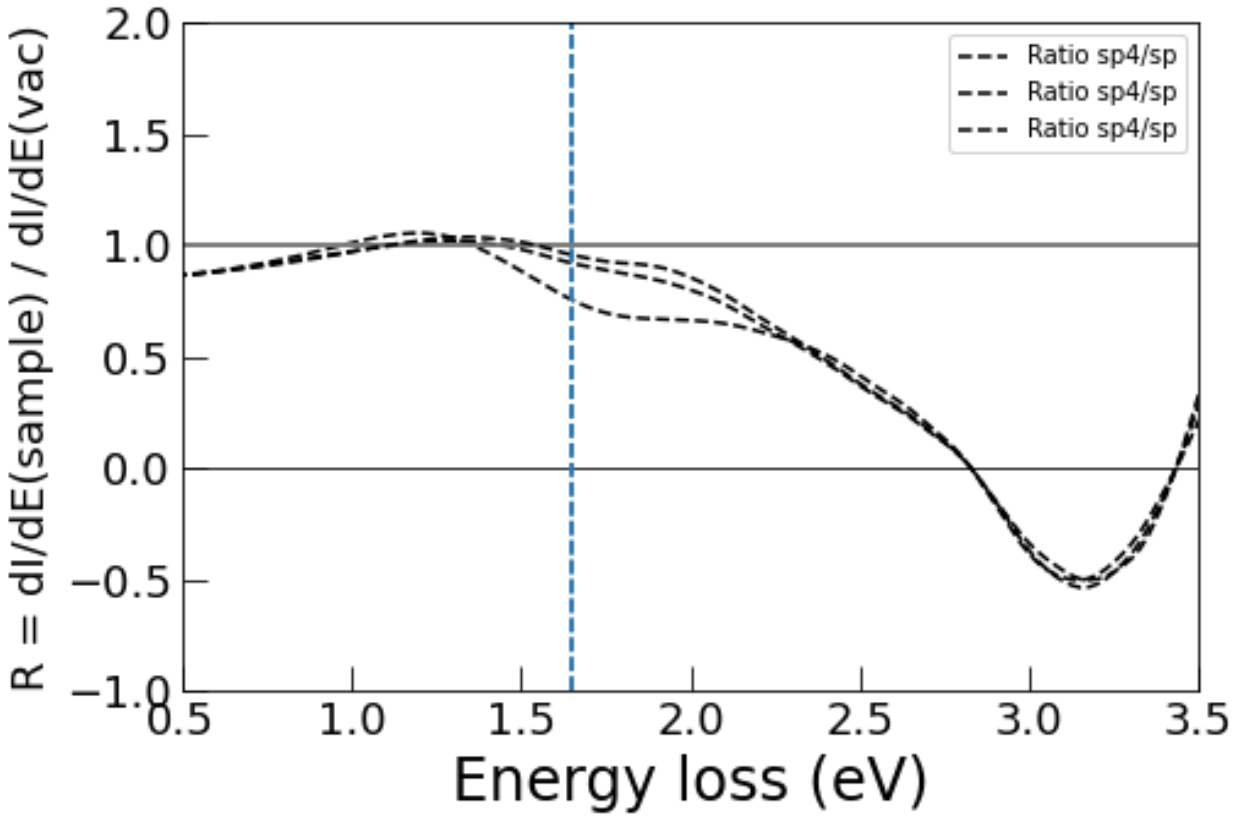
- We take the mean integrated intensity for each spectrum in cluster
  - → filter outliers

- Just use trimmed mean & trim cluster? (i.e. discard equal number of top and bottom values):
  - CI = round((0.05 * n_cluster)/2)
  - Cluster = cluster[CI:-CI]

- Use mean shift outlier model to determine outliers? (Check influence of each value on mean cluster, discard ones with most influence, i.e. not equal number top and bottom)

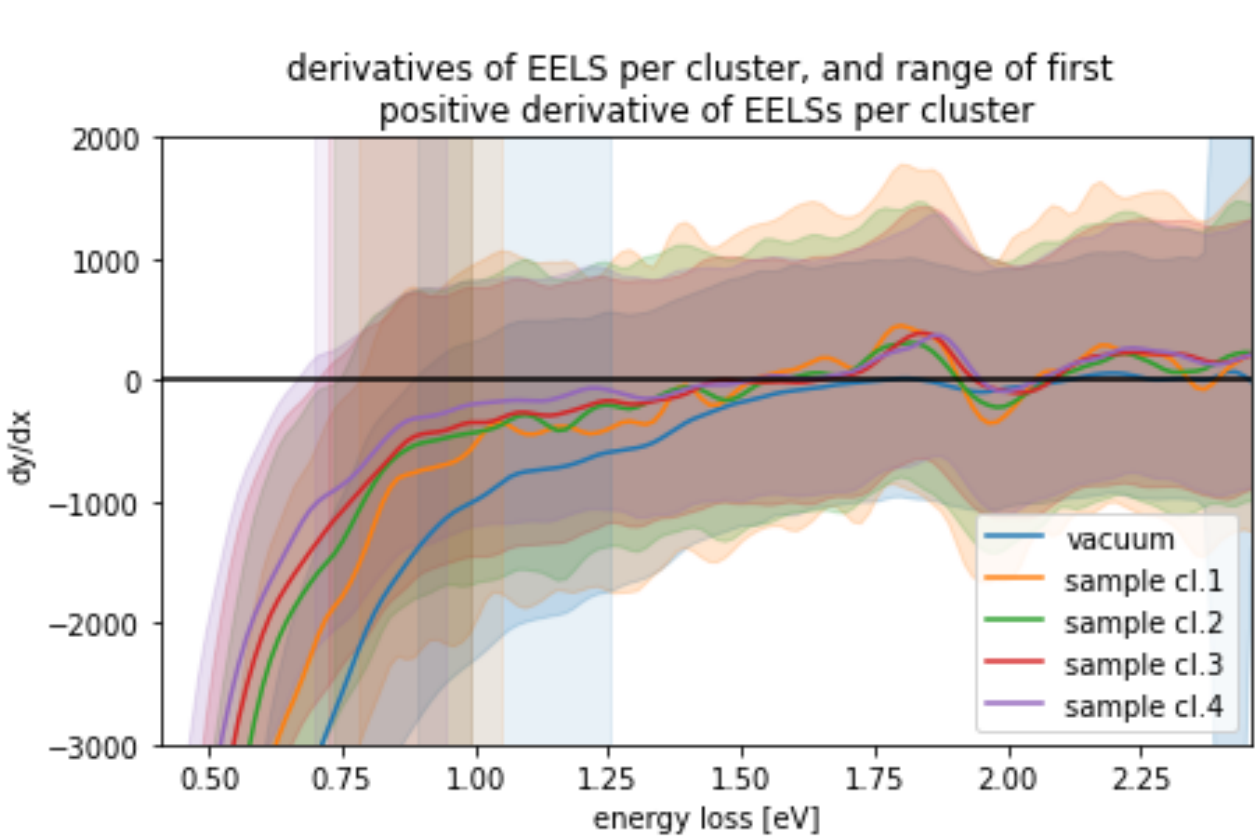
Finding dE1 -- Laurien

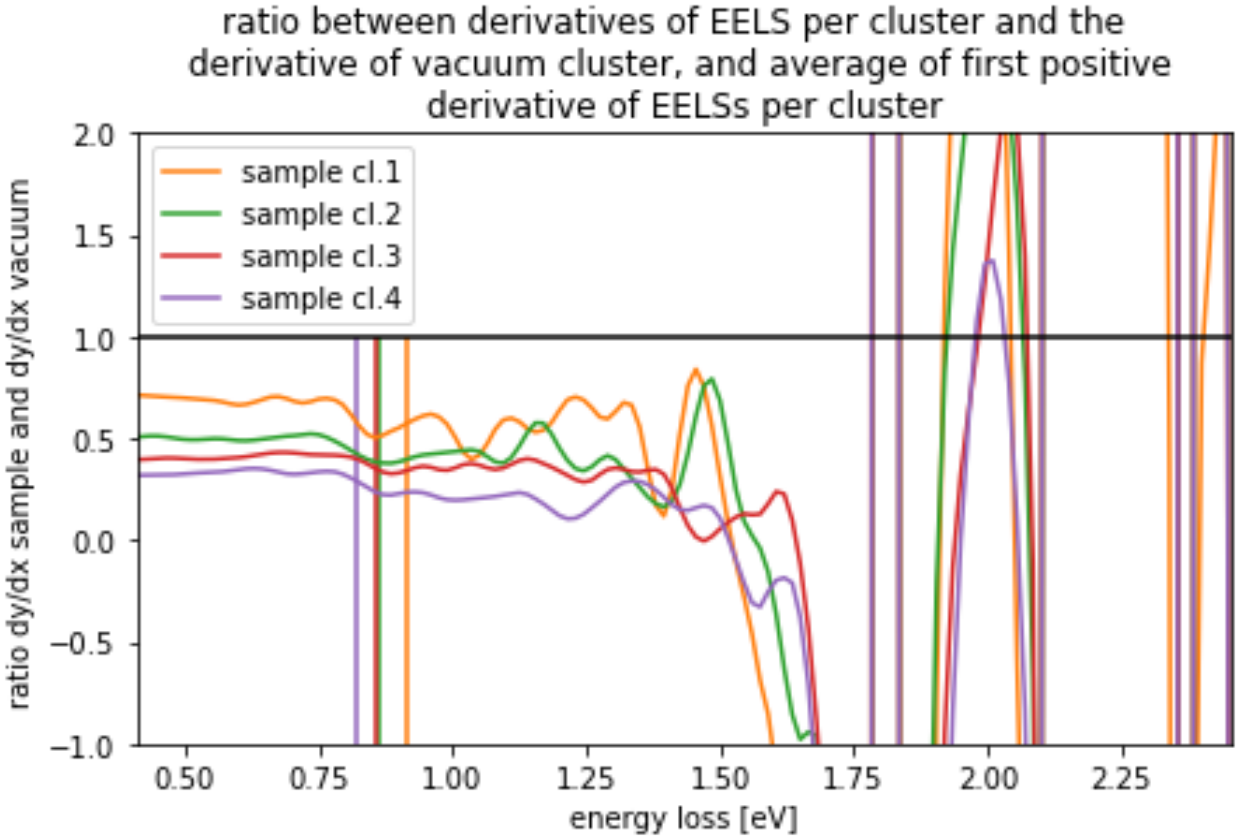




Finding dE1

Note: not 1!



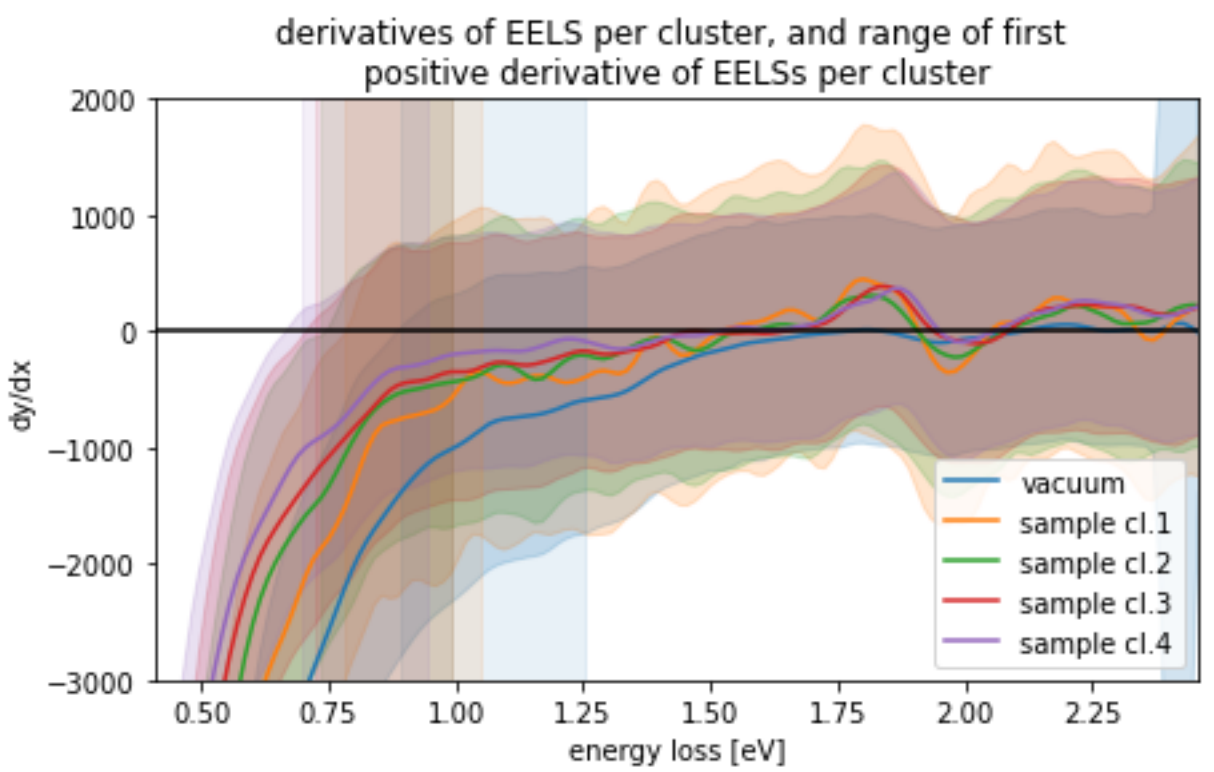


Values

- dE1 cluster 0 avg: 1.072 , std: 0.1827 , min: 0.615
- dE1 cluster 1 avg: 0.9167 , std: 0.1337 , min: 0.585
- dE1 cluster 2 avg: 0.8621 , std: 0.1302 , min: 0.54
- dE1 cluster 3 avg: 0.8554 , std: 0.1337 , min: 0.525
- dE1 cluster 4 avg: 0.8188 , std: 0.1224 , min: 0.51

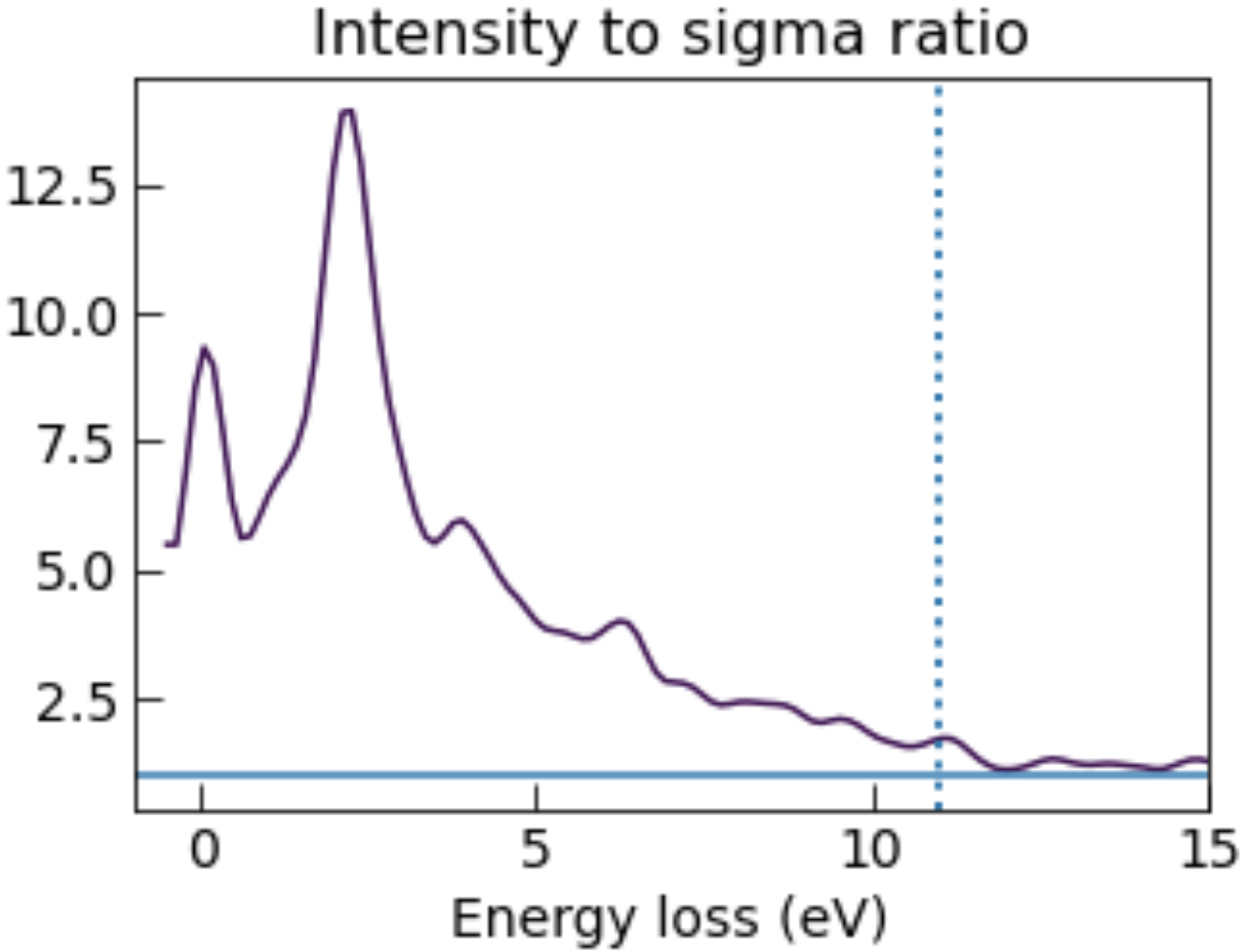
- Thicker sample → lower dE1:
  - Makes sense: thicker sample → higher overall influence → sooner visable

How to determine dE1?

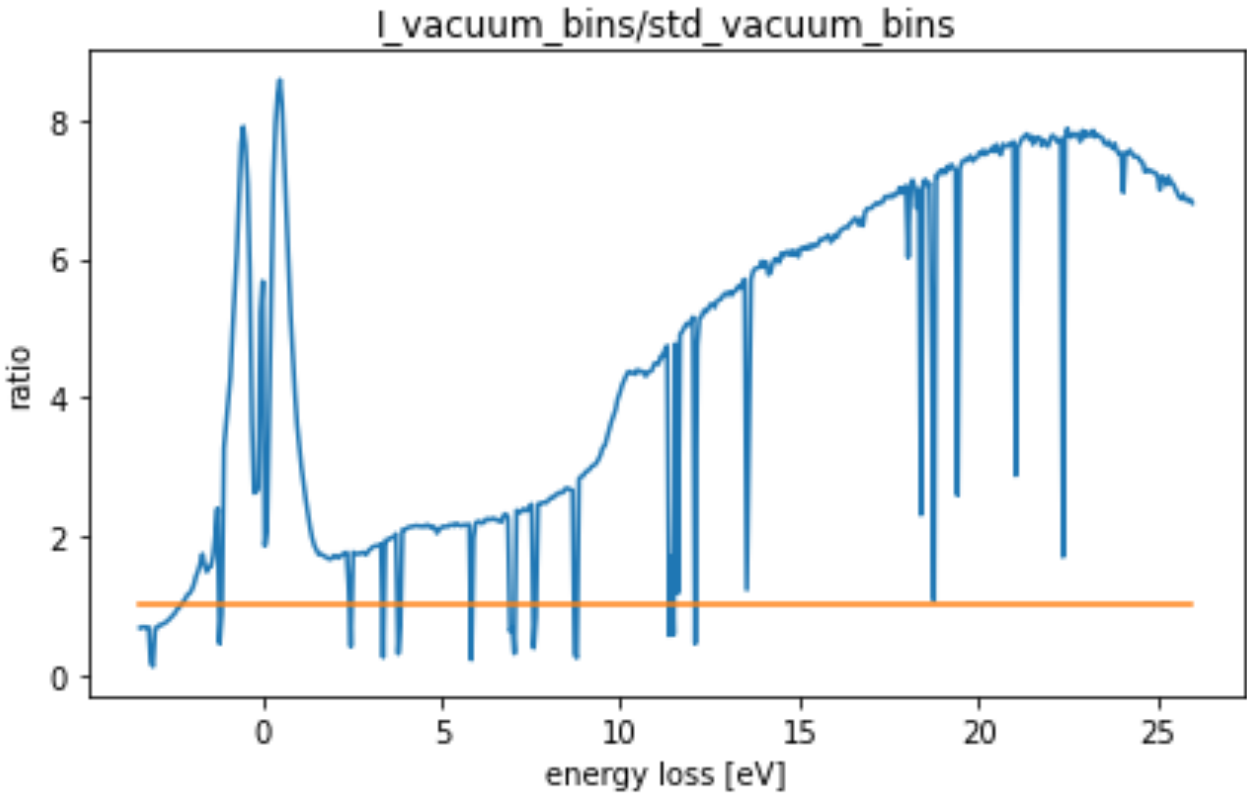


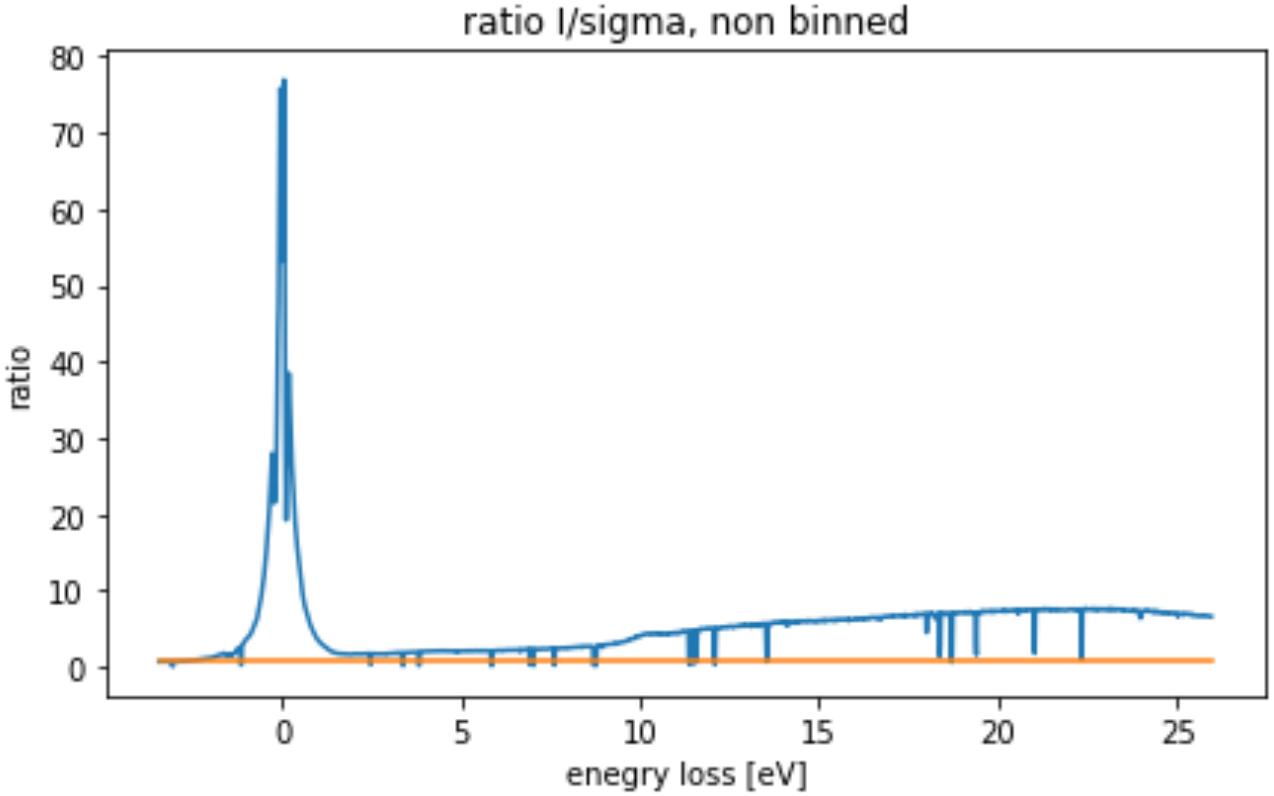
- Lowest of all spectra?
- Lowest of cluster averages?
- Point where cluster derivative average crosses 0?

Finding dE2 -- Laurien

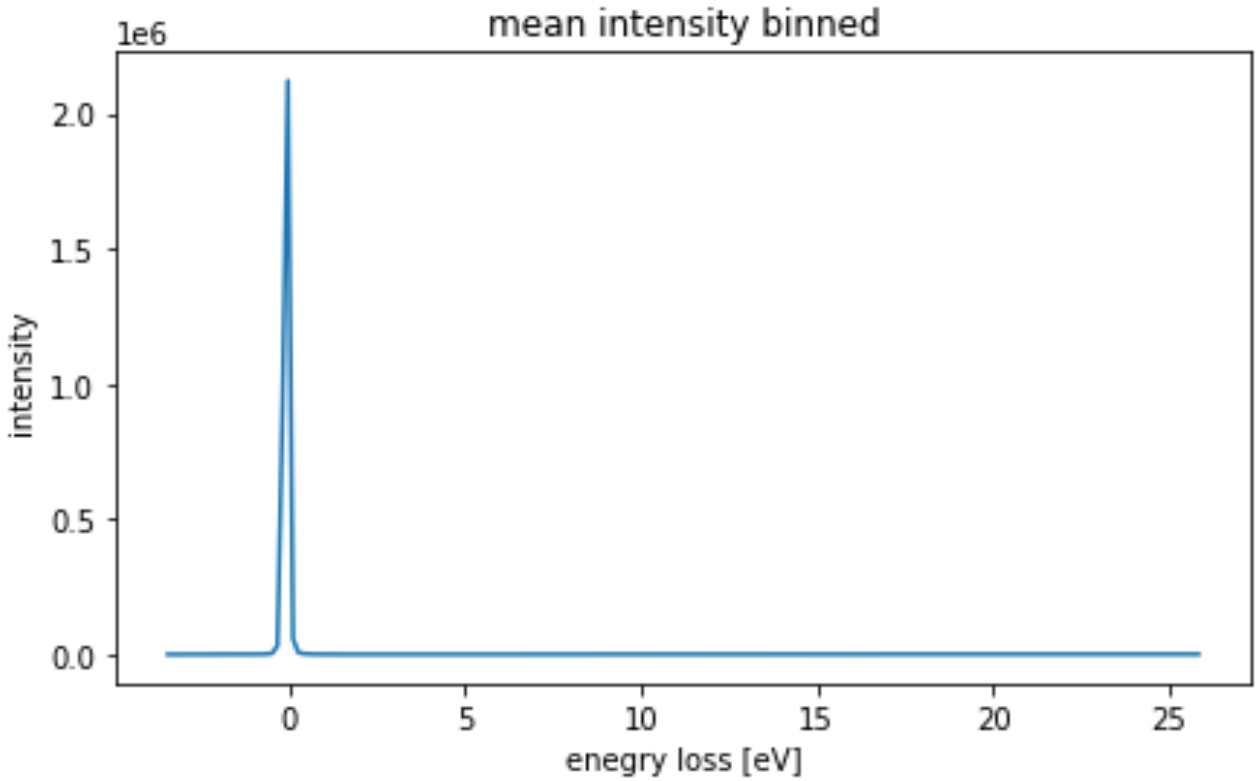


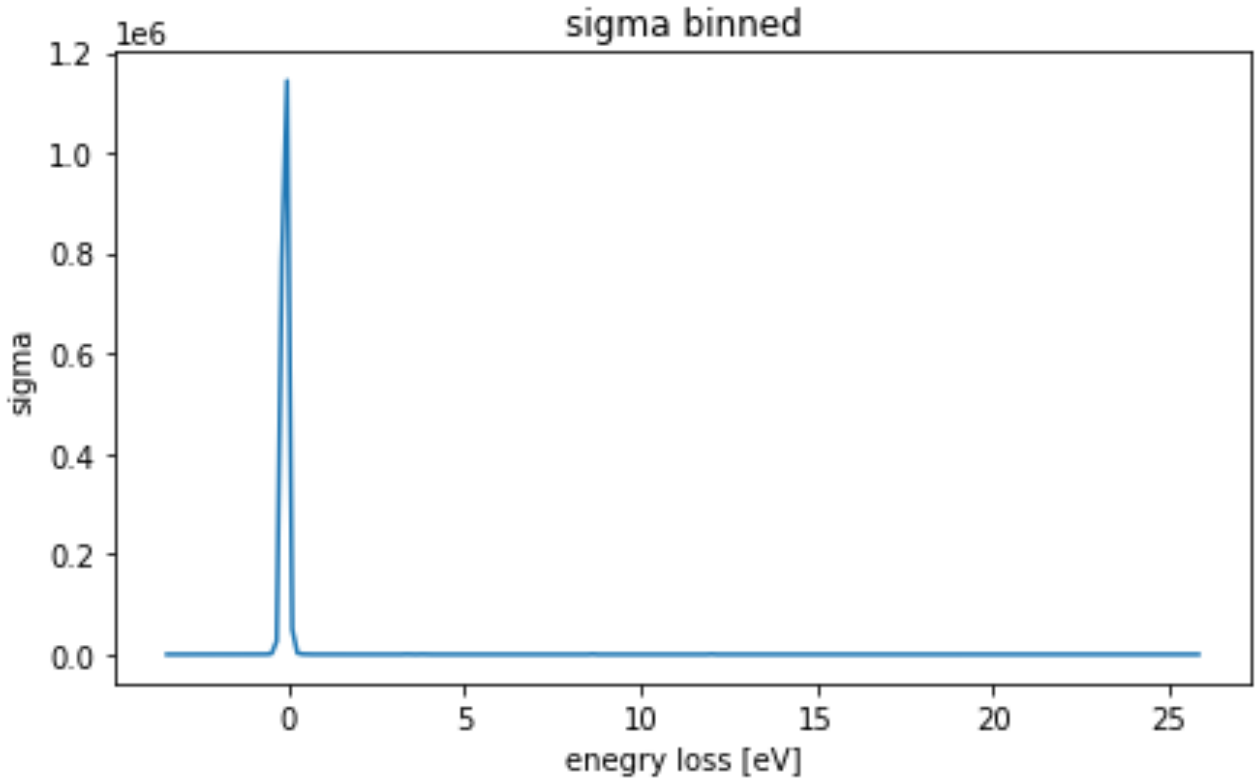
Finding dE2


I_vacuum_bins/std_vacuum_bins
ratio
energy loss [eV]
0
2
4
6
8
0
5
10
15
20
25

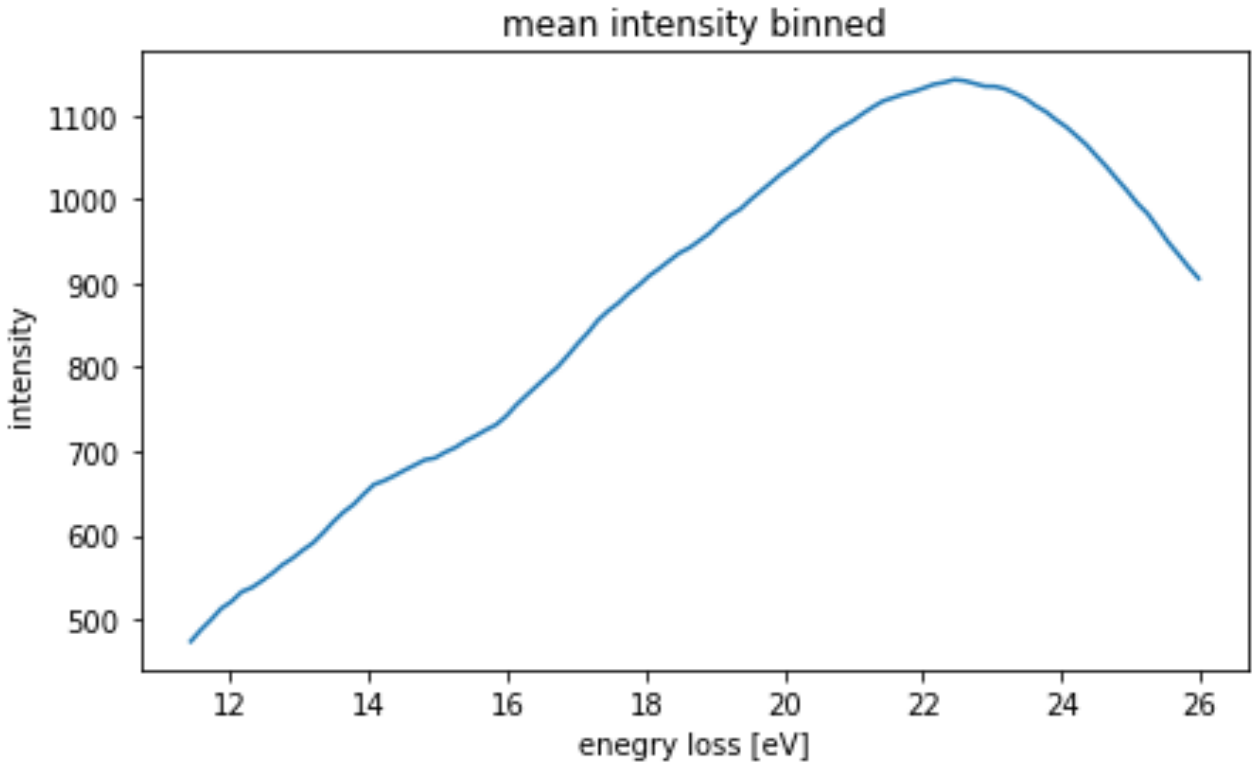

ratio I/sigma, non binned
ratio
enegry loss [eV]
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25

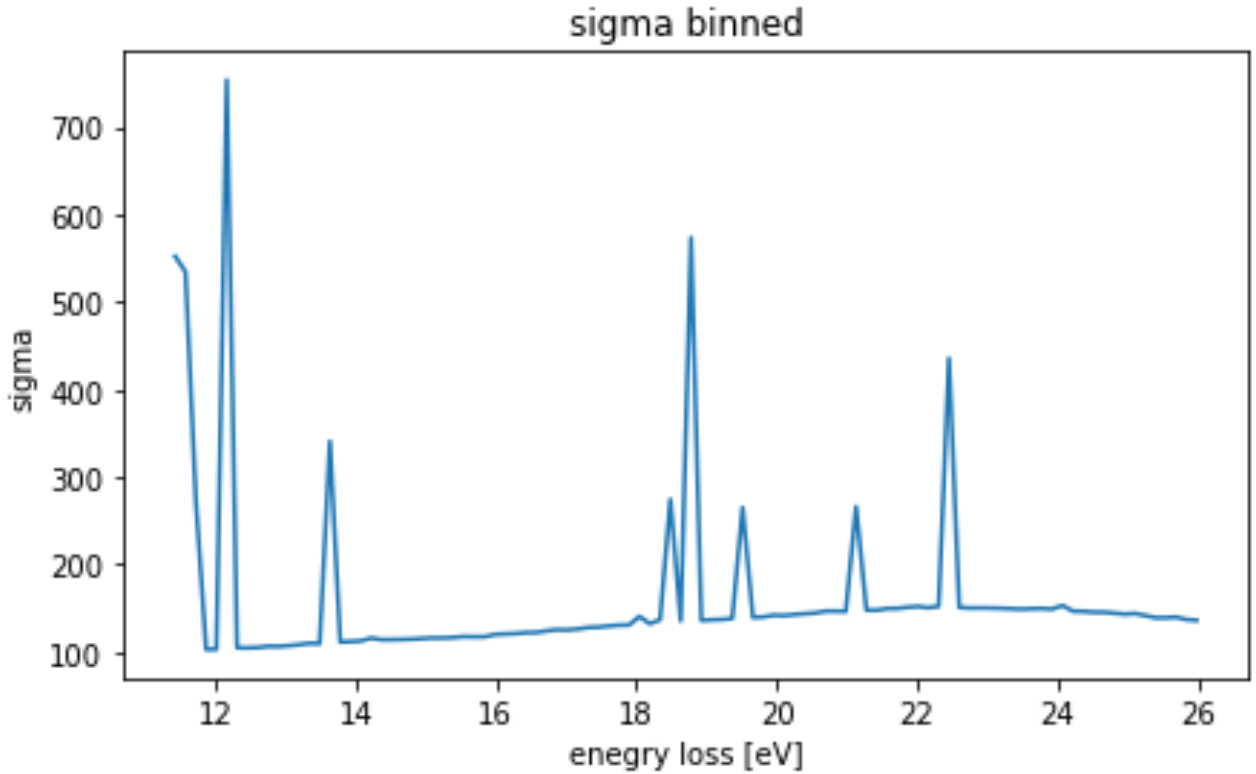
Mean intensity and sigma of bins





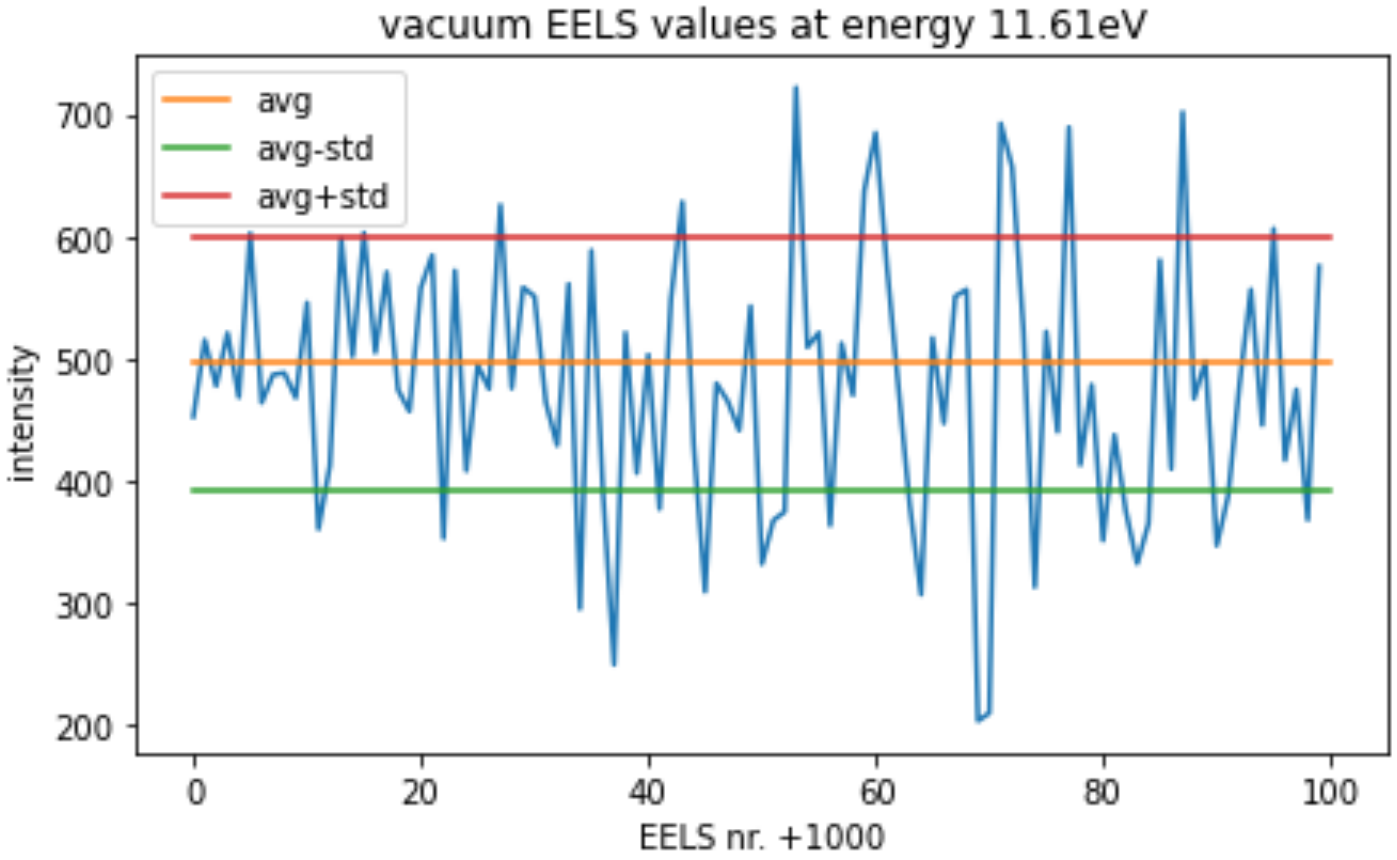
Zoom in on end



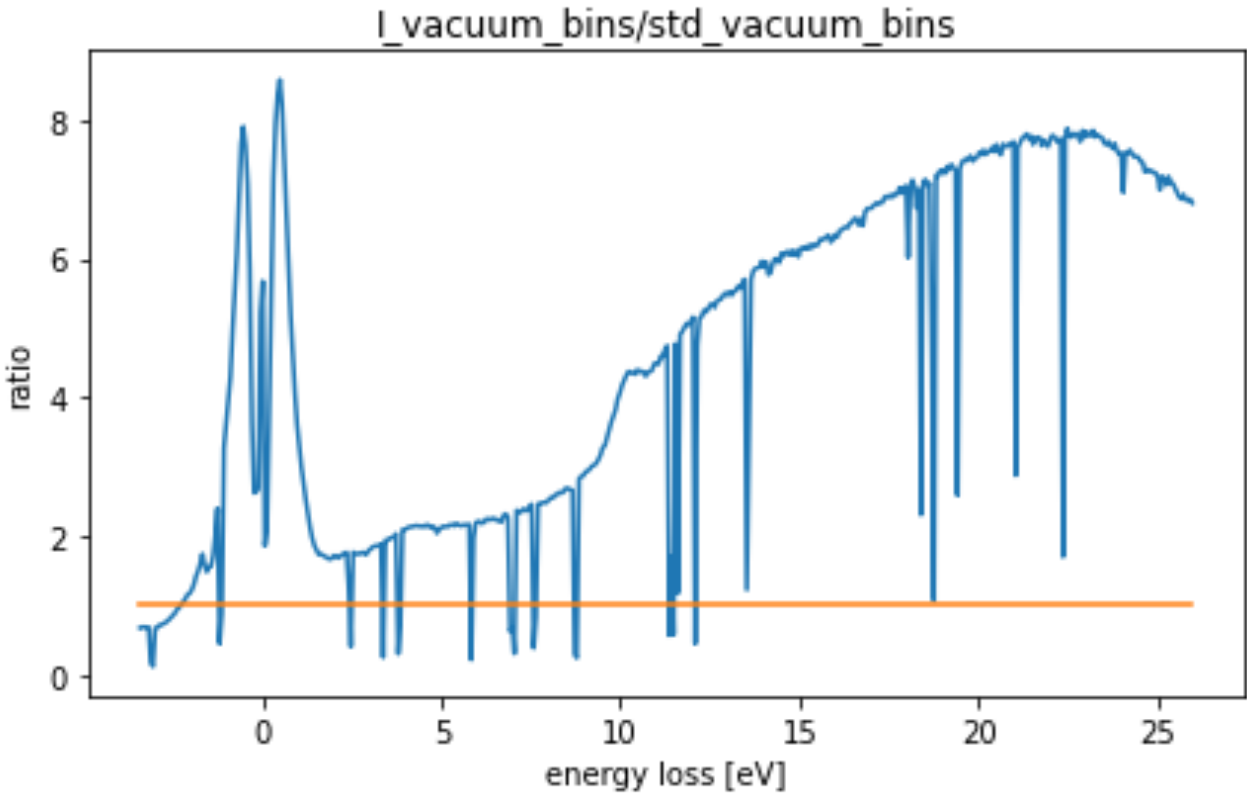


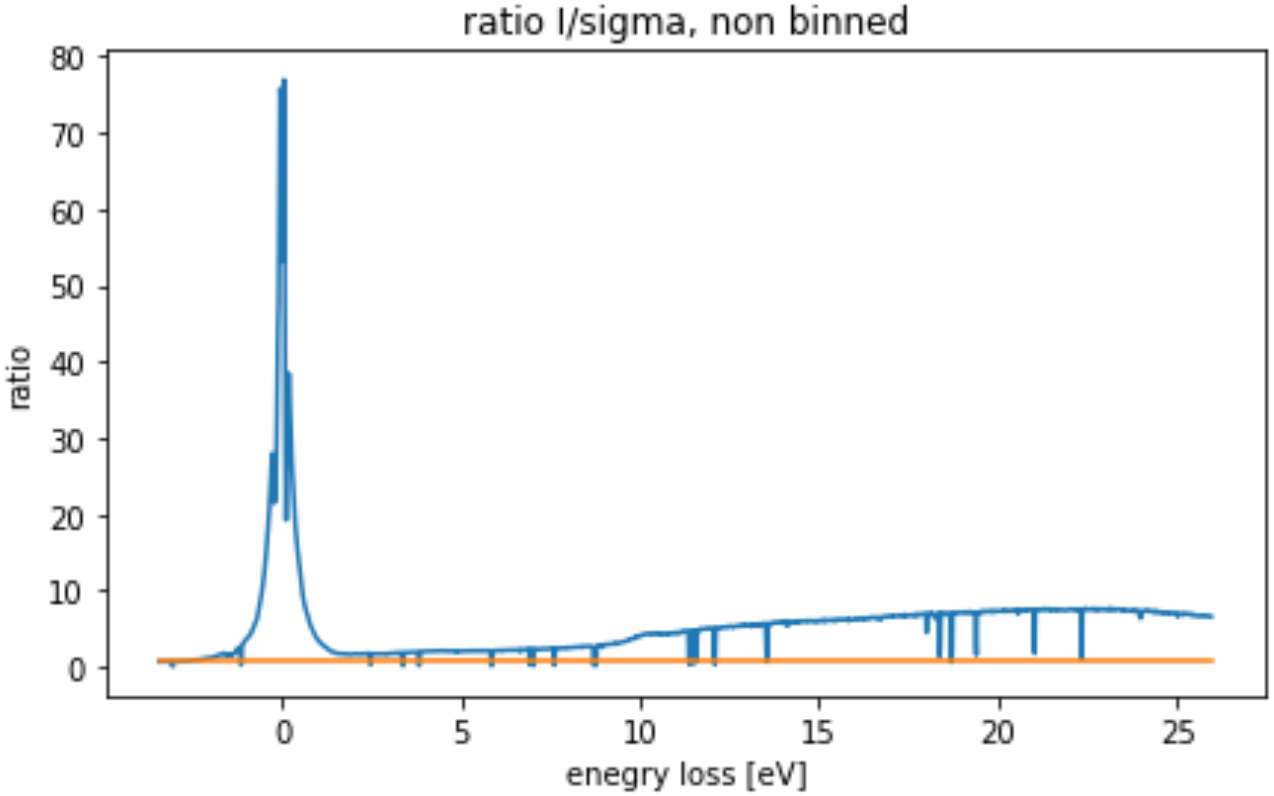
Why not converging to 1?

Values of all spectra at single energy

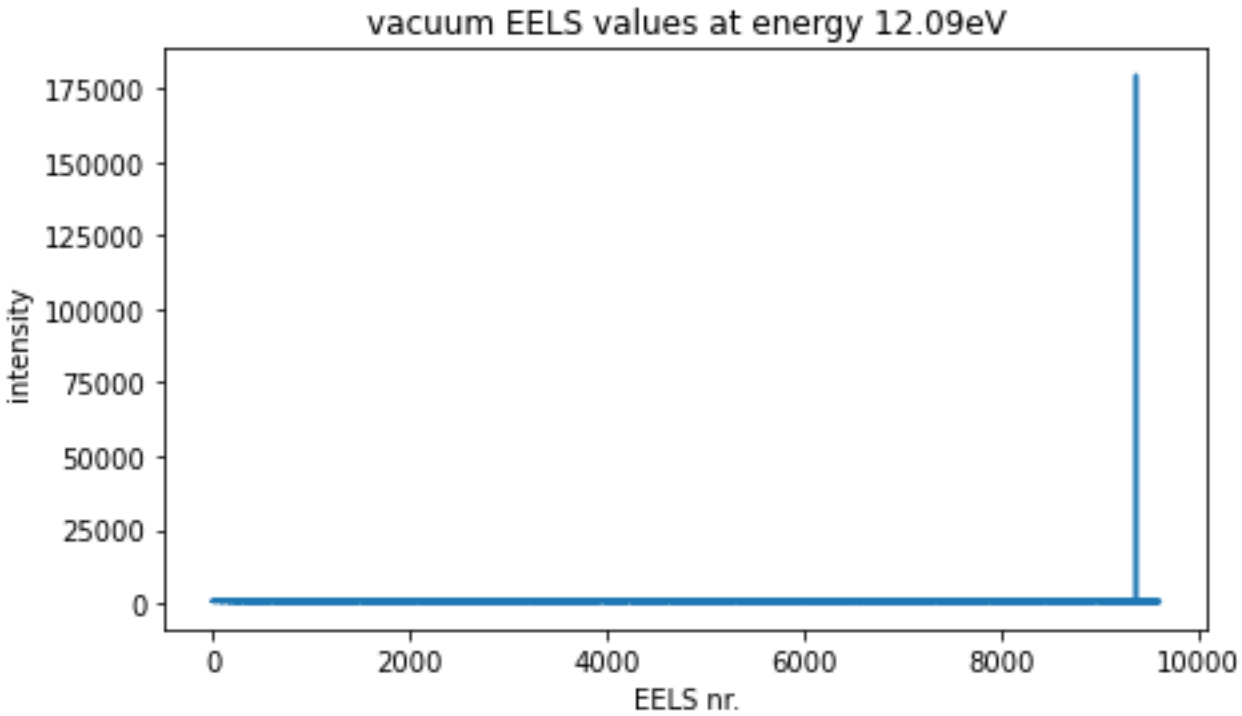


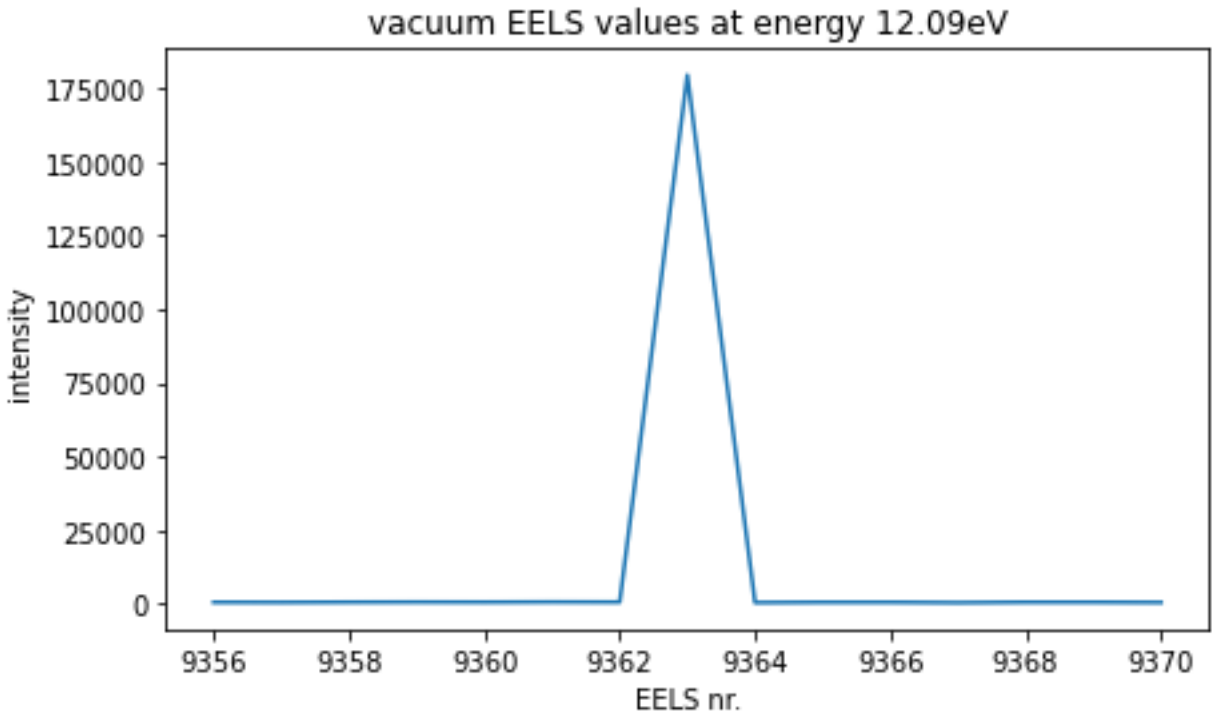
Finding dE2


I_vacuum_bins/std_vacuum_bins
ratio
energy loss [eV]
0
2
4
6
8
0
5
10
15
20
25

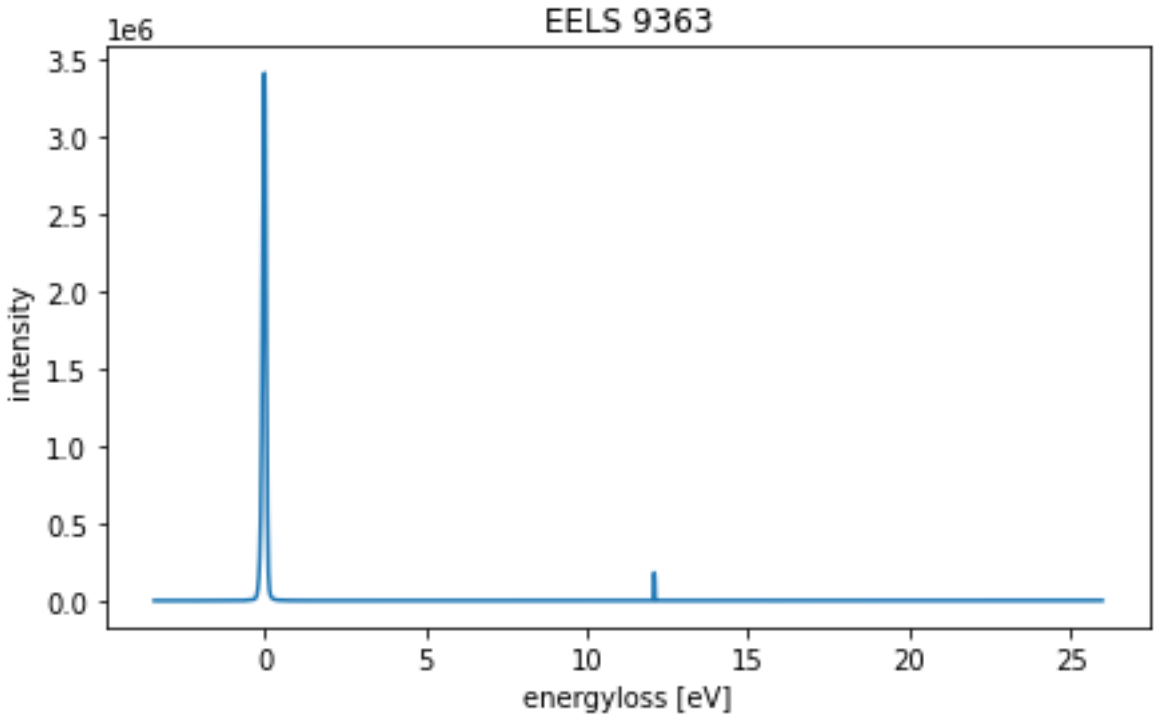

ratio I/sigma, non binned
ratio
enegry loss [eV]
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25

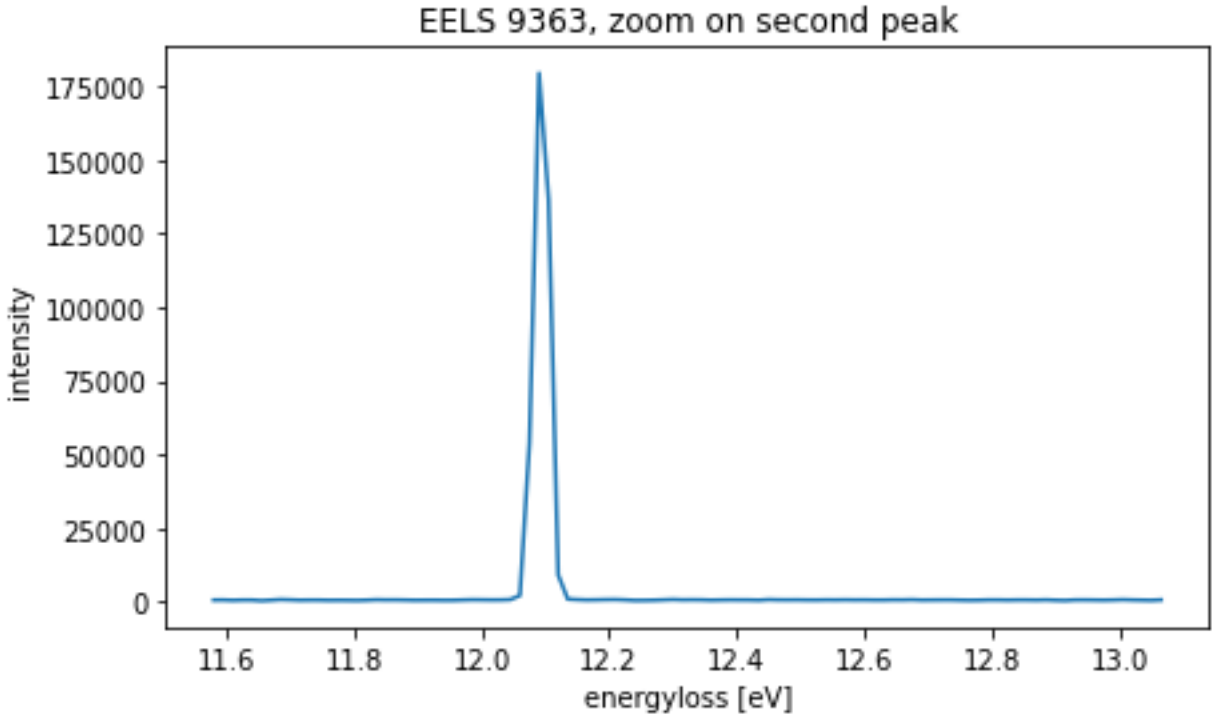
Why the spikes?
all vacuum values at 12.09eV





Why the spikes?



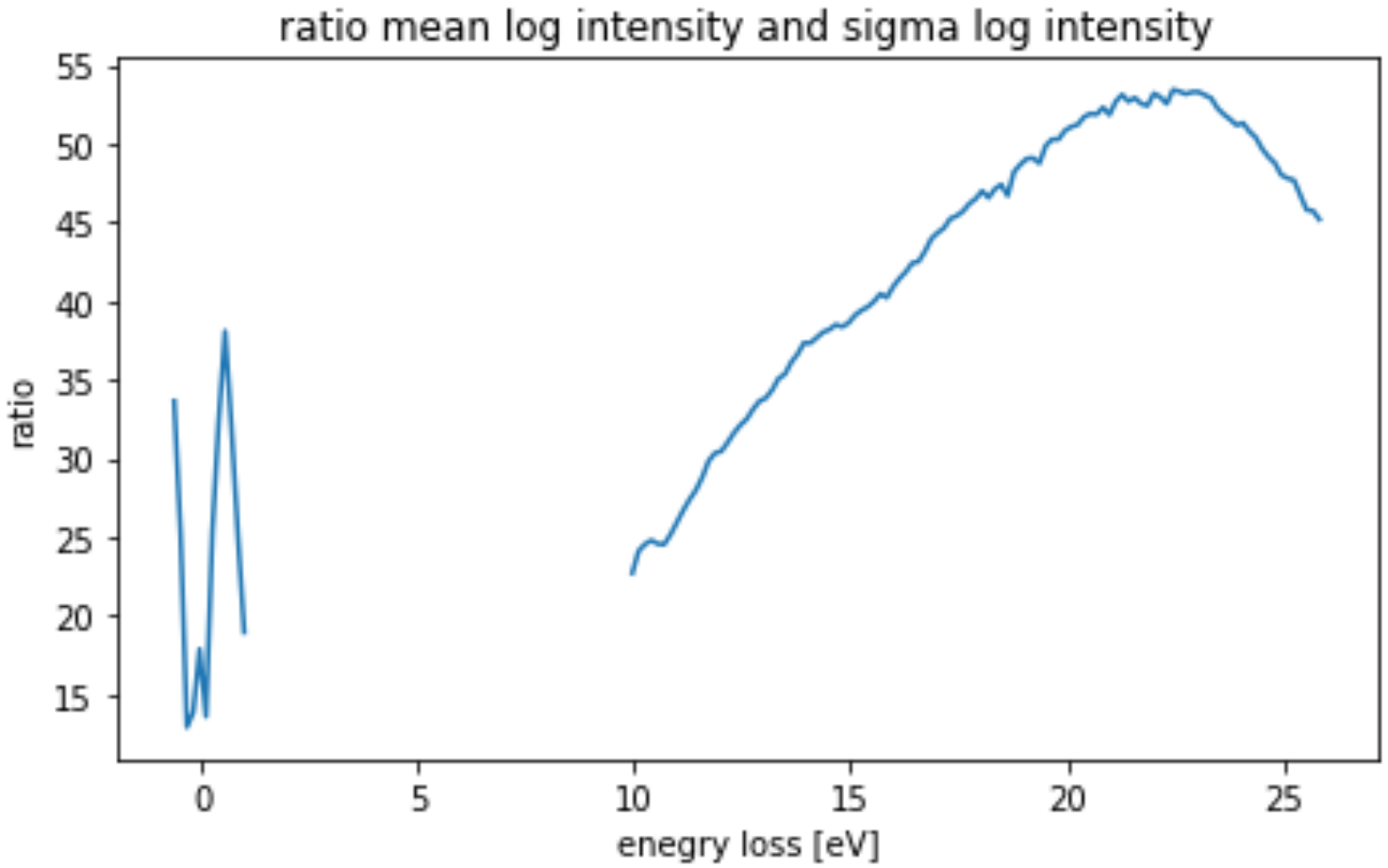


- Sonia????

Solutions

- Spikes should be filterable (some comparison to neighbours)
- Value not converting to one more difficult I think:
  - Take higher threshold?
  - Assume ZLP indeed still has influence and set higher dE2?

PS: log helps for the spikes, not for the convergence



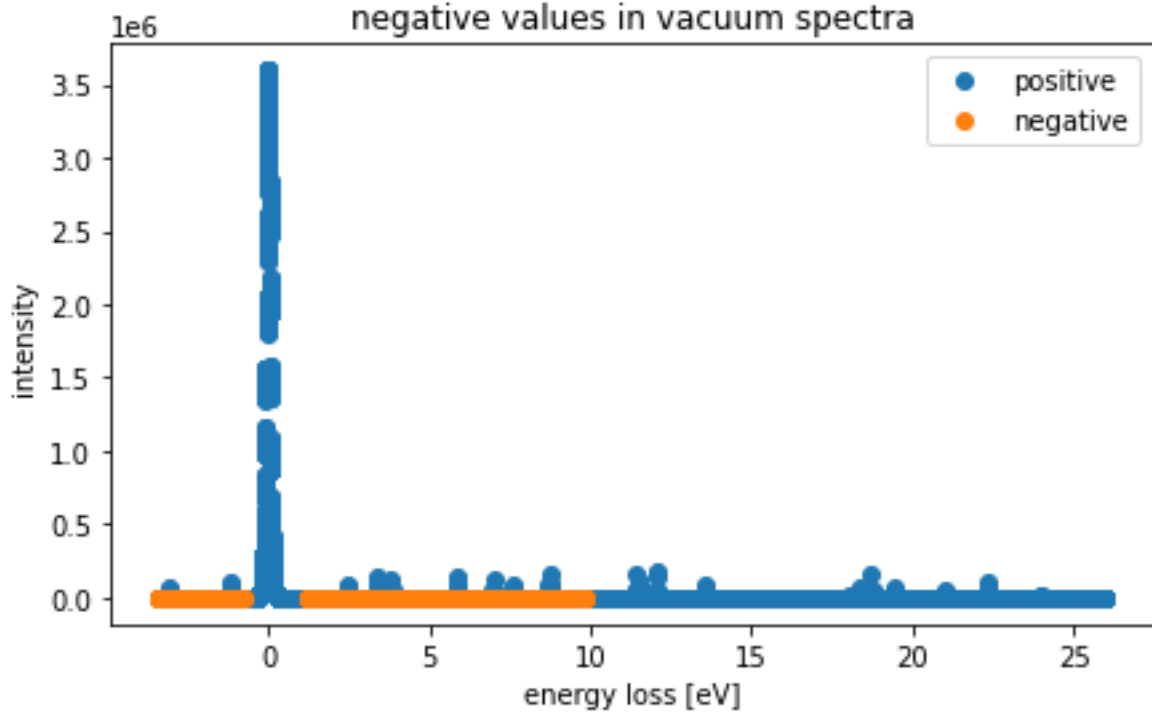
For log: how to handle negatives (and 0)?

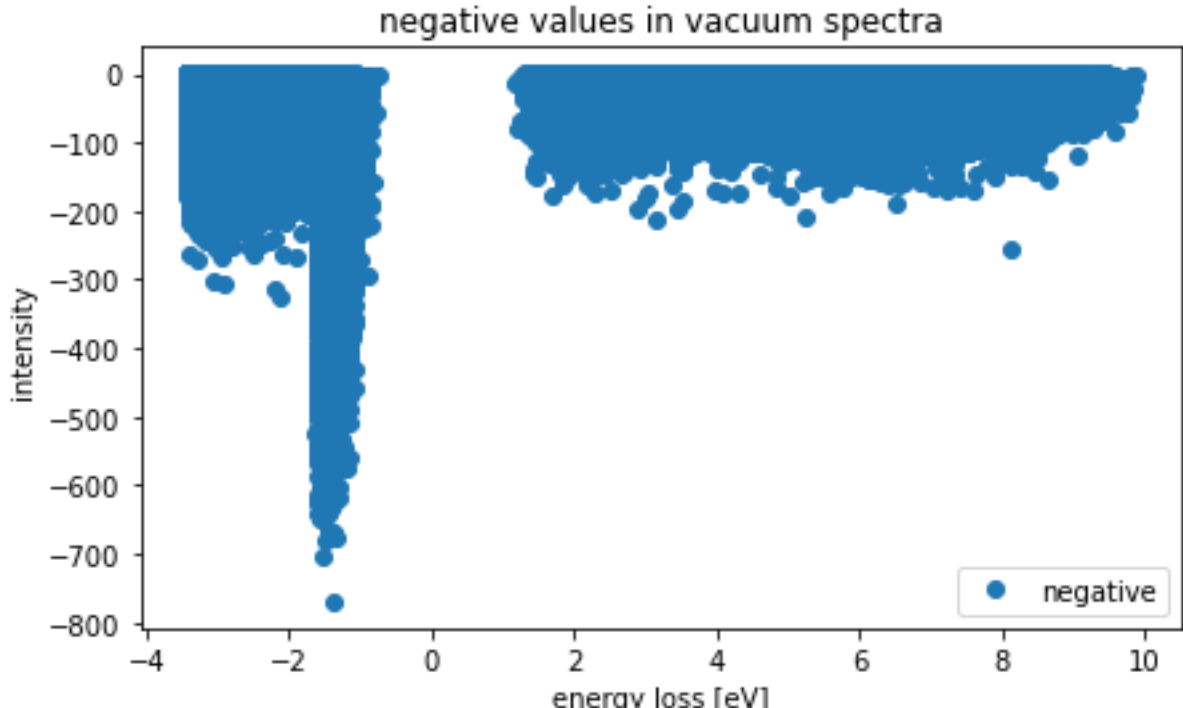
Problem:

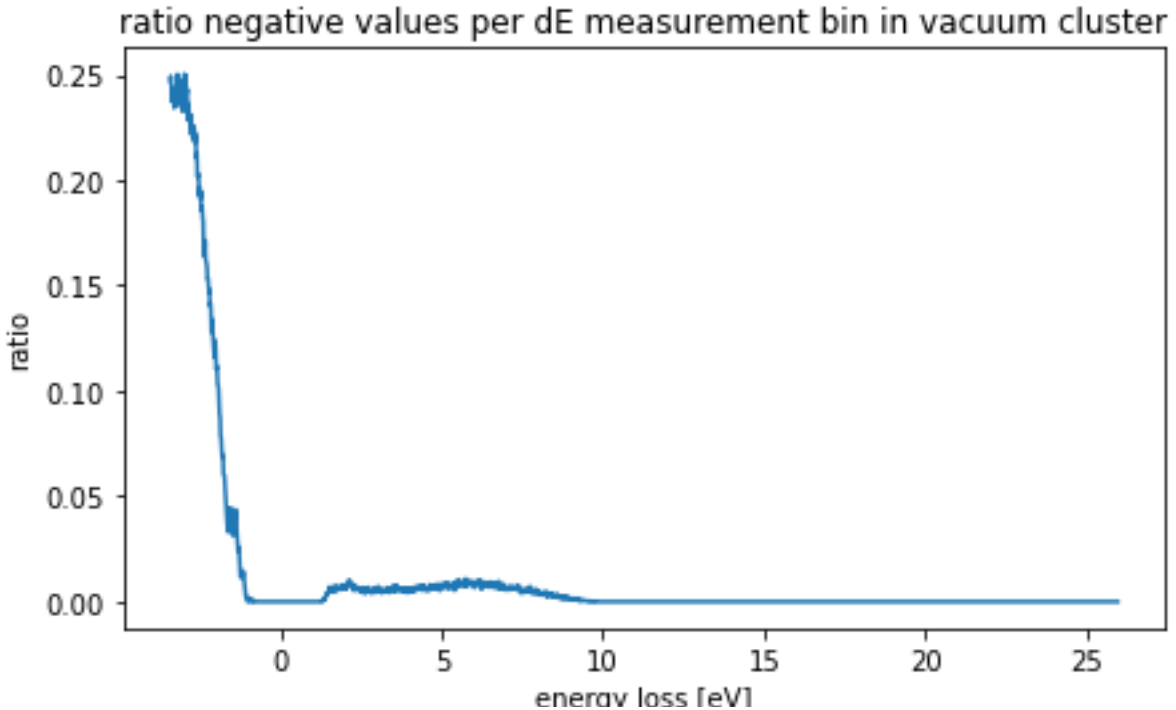
- For vacuum cluster:
  - Ratio negative/pos: 2.5E5/1.86E7 = 0.013
  - Avg negative: -39
  - Min negative: -770
  - Avg positive: 16900

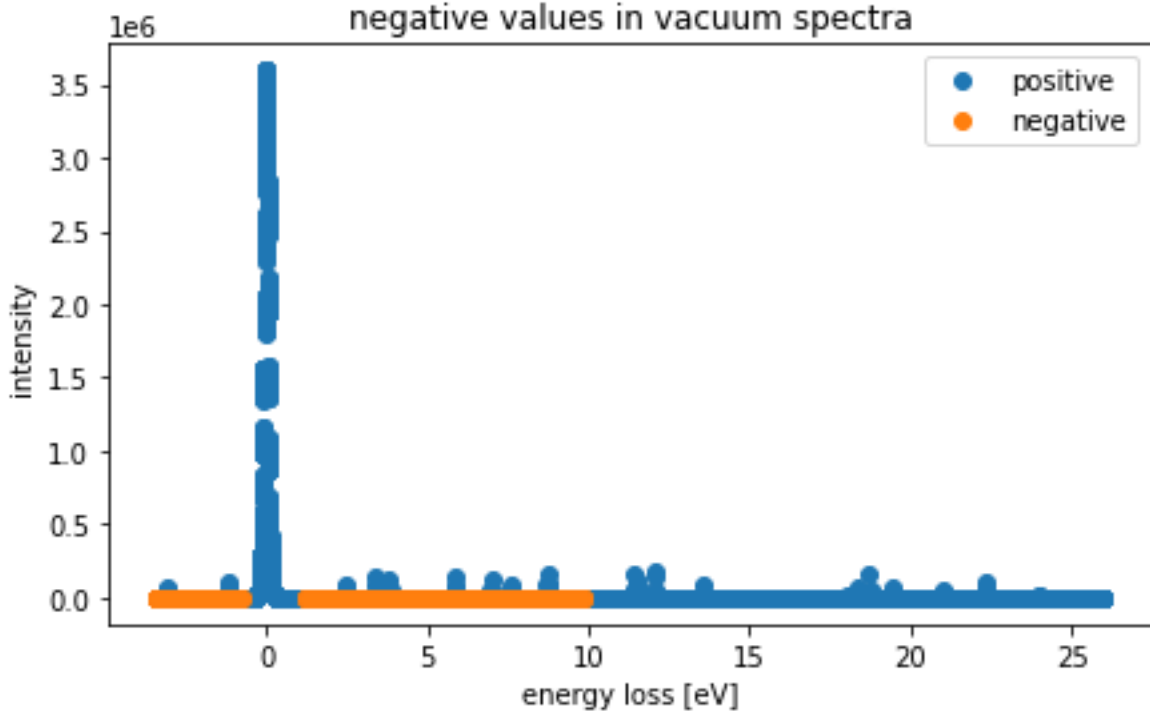
Solution?

- I = I – min(I)
  - Per spectrum?
  - Per cluster?
- I[I<1] = 1 ?

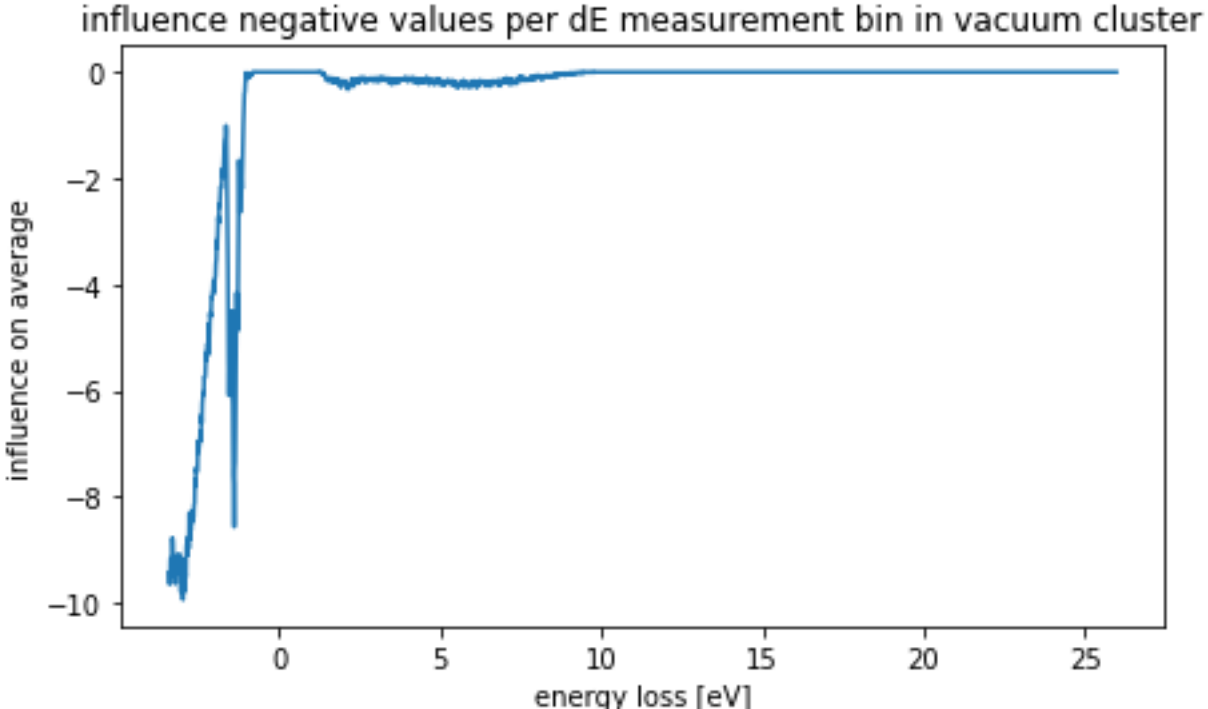


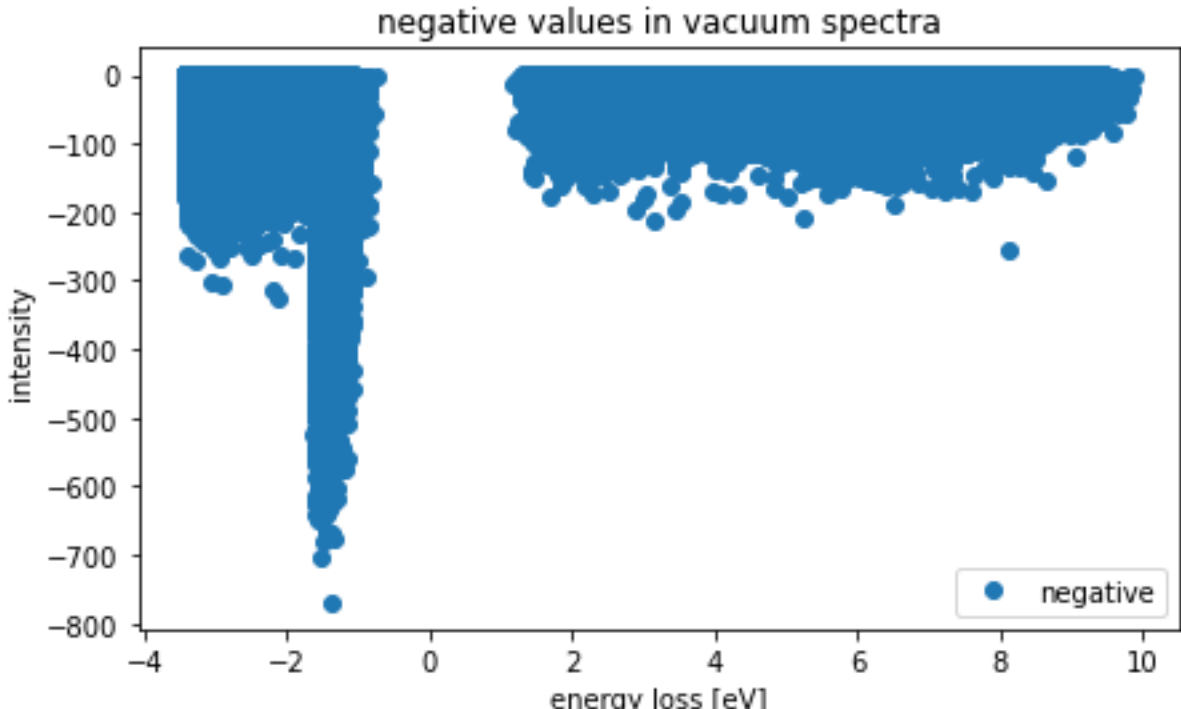


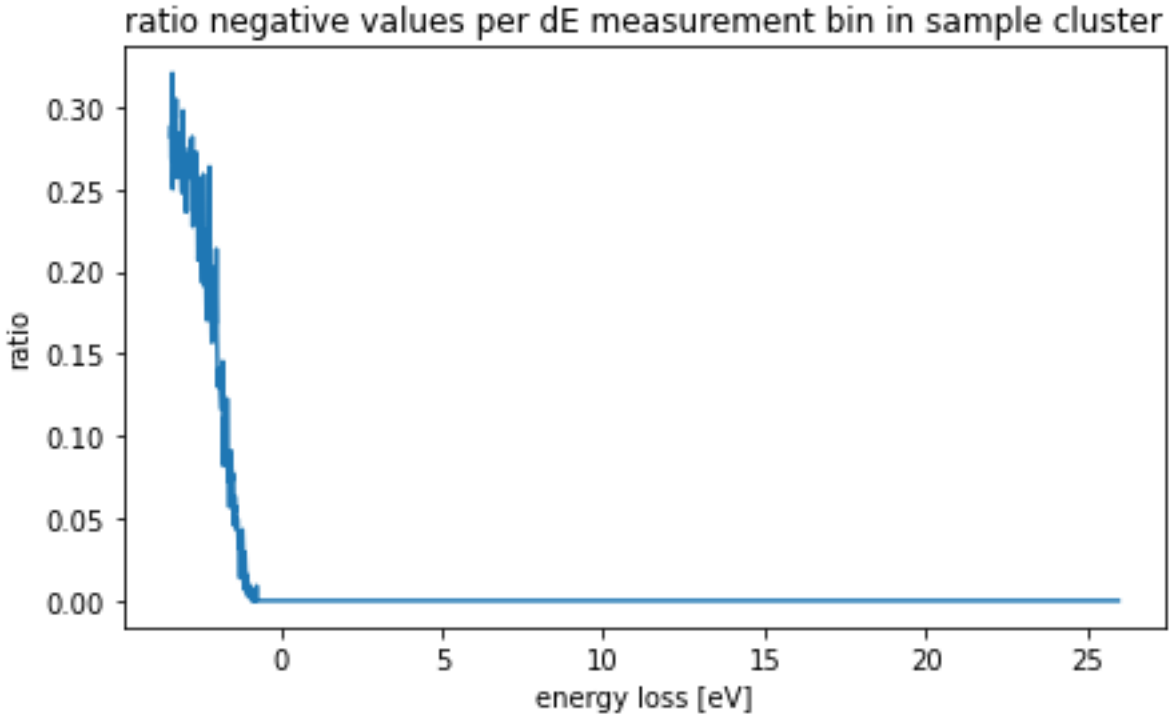

ratio negative values per dE measurement bin in vacuum cluster
ratio
energy loss [eV]

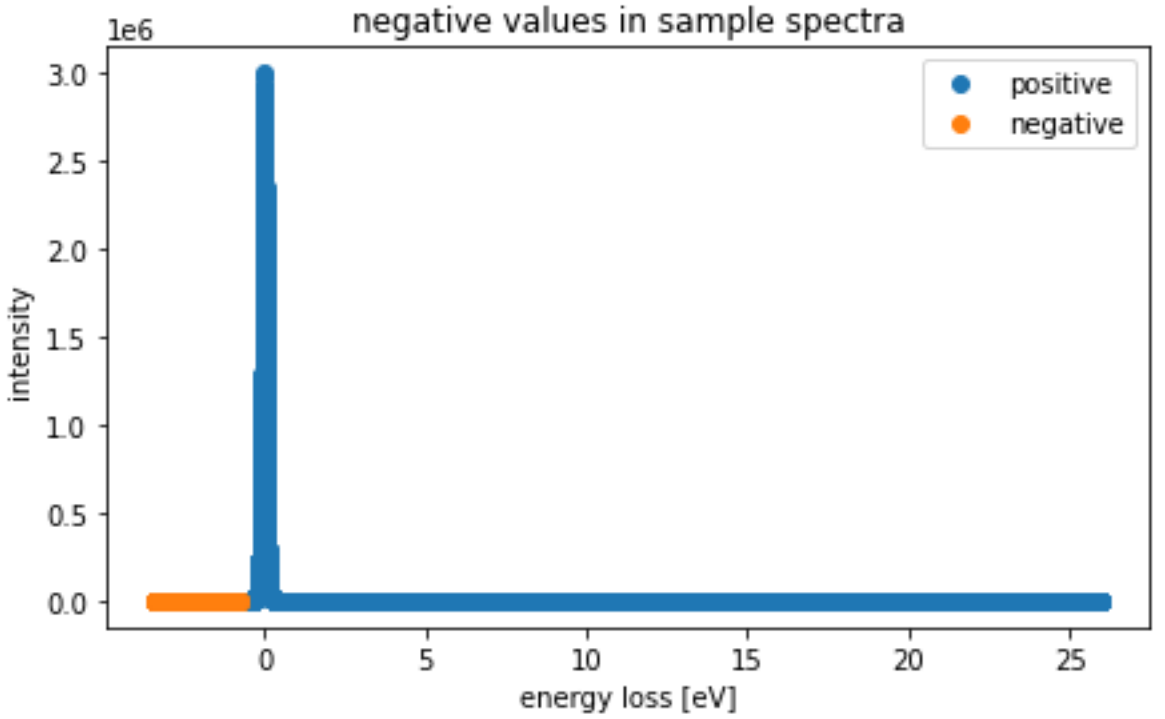

negative values in vacuum spectra
1e6
intensity
energy loss [eV]
positive
negative

VACUUM

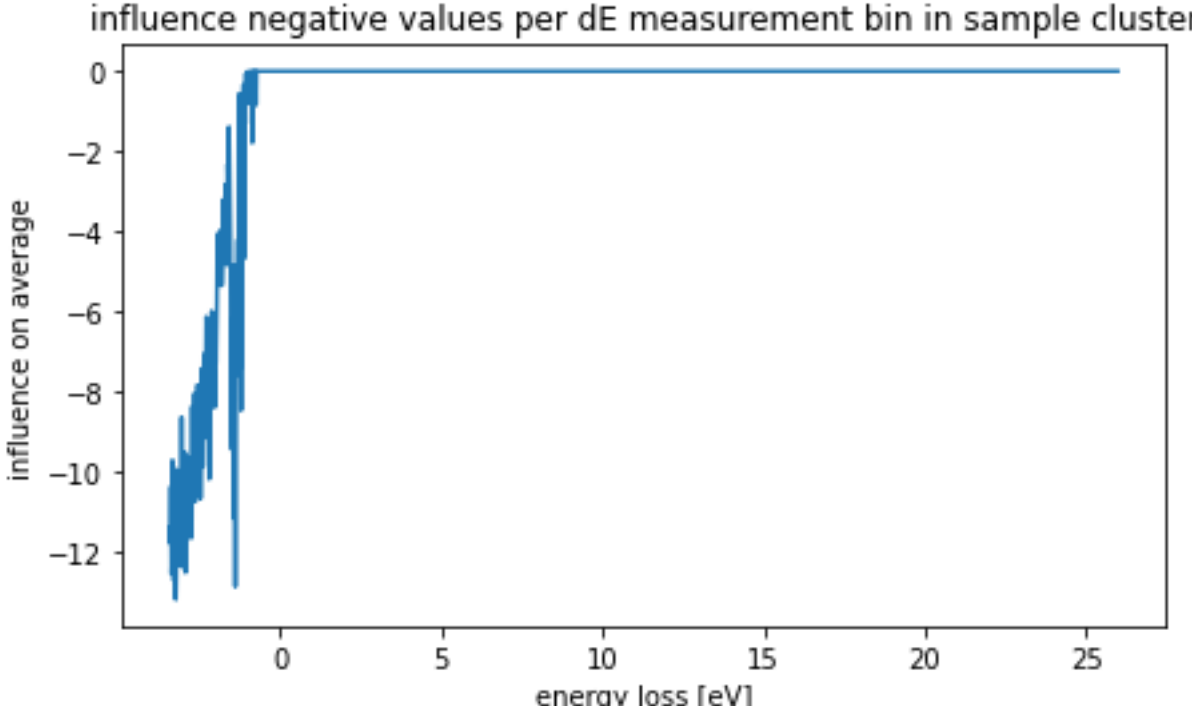

influence negative values per dE measurement bin in vacuum cluster
influence on average
energy loss [eV]

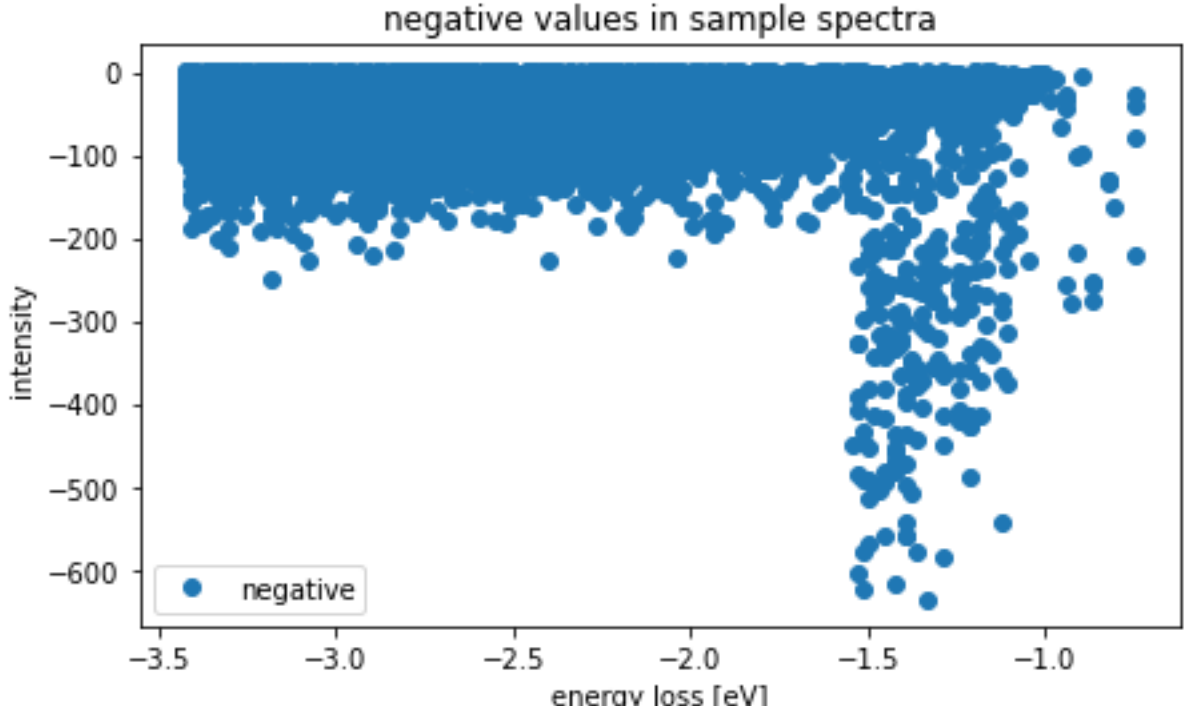

negative values in vacuum spectra
intensity
energy loss [eV]
negative

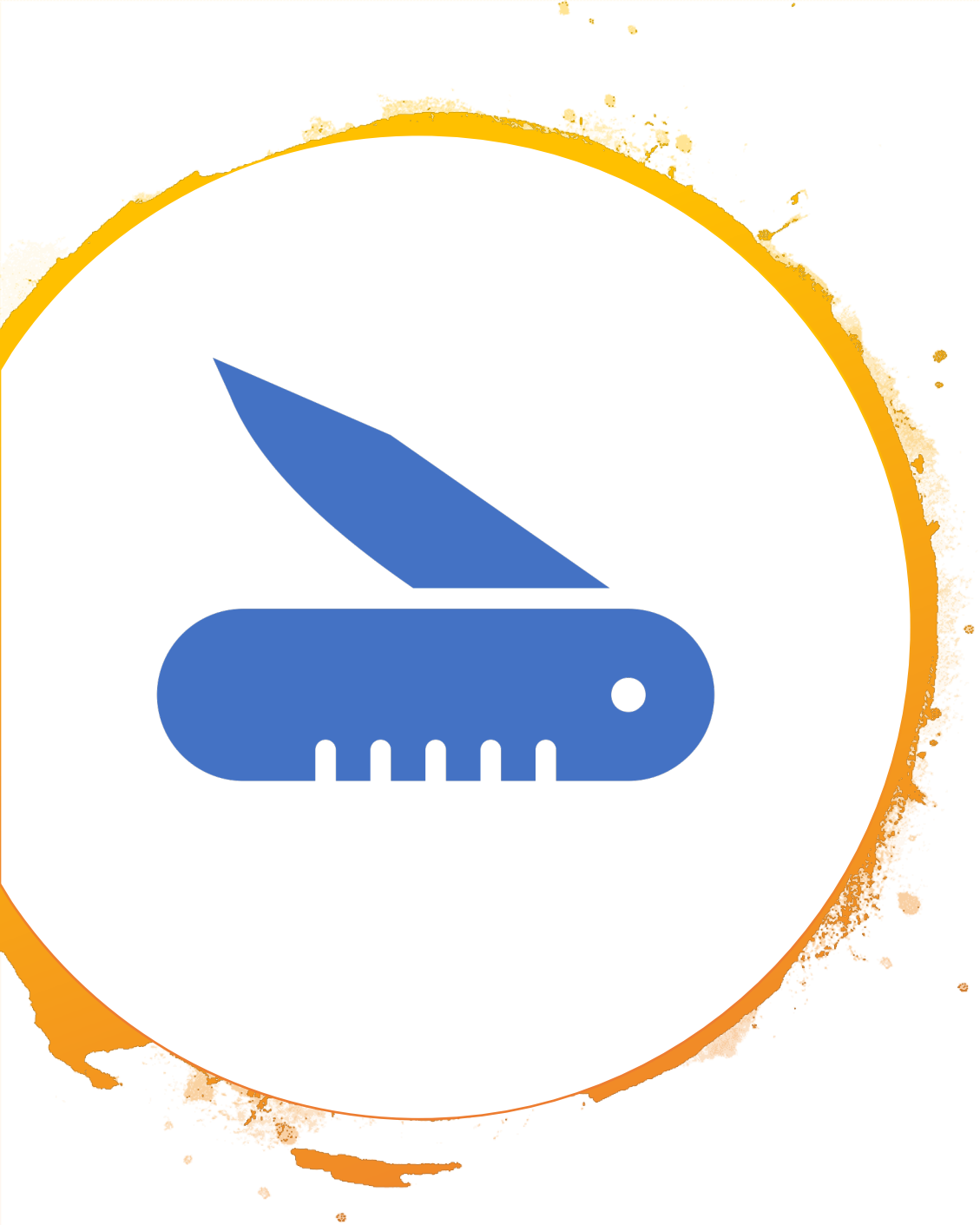

ratio negative values per dE measurement bin in sample cluster
ratio
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
25
energy loss [eV]


negative values in sample spectra
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
intensity
positive
negative
0
5
10
15
20
25
energy loss [eV]

SAMPLE


influence negative values per dE measurement bin in sample cluster
influence on average
0
−2
−4
−6
−8
−10
−12
0
5
10
15
20
25
energy loss [eV]


negative values in sample spectra
0
−100
−200
−300
−400
−500
−600
intensity
negative
−3.5
−3.0
−2.5
−2.0
−1.5
−1.0
energy loss [eV]



BTW: Lau only trained -0.5eV to 20eV

- I figured I’d just take the range of the sample?