Building Python frameworks for data analysis

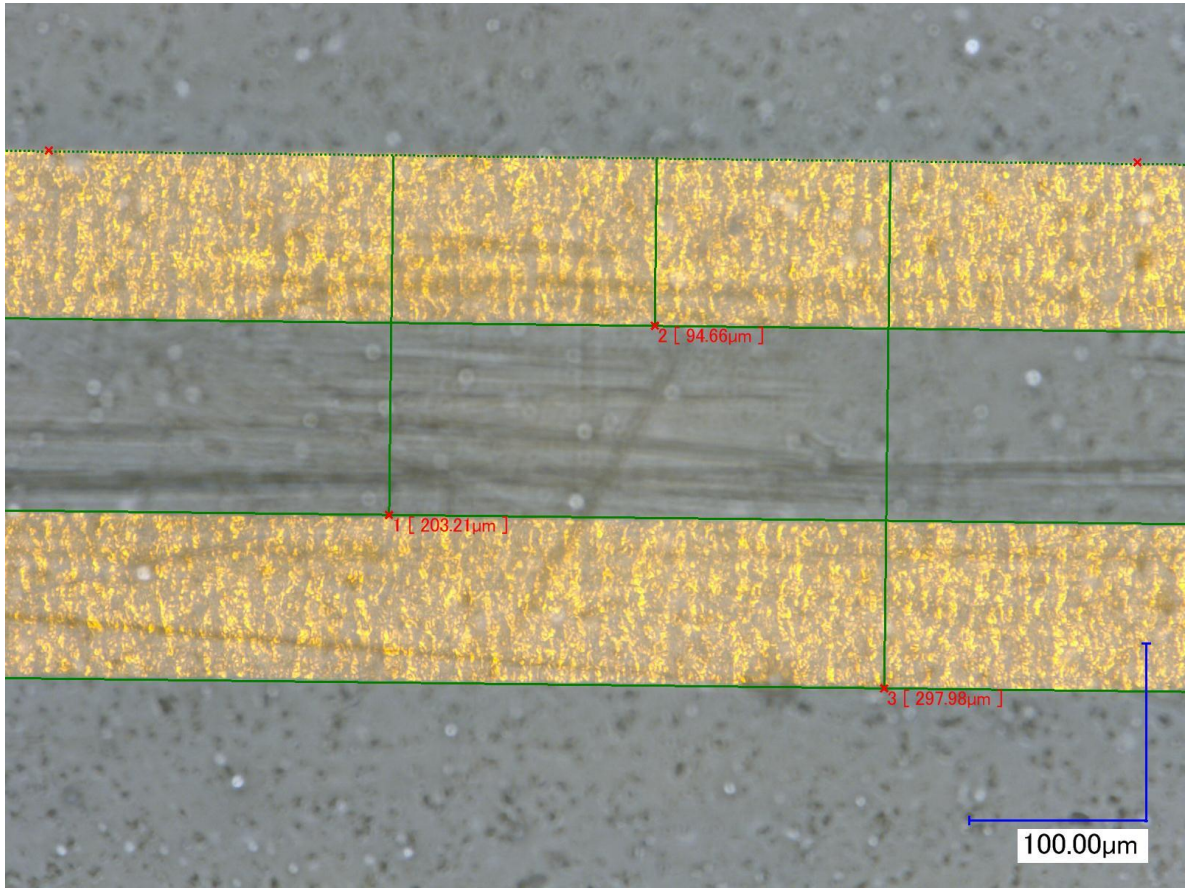
Daniel Cogbill

Two main projects

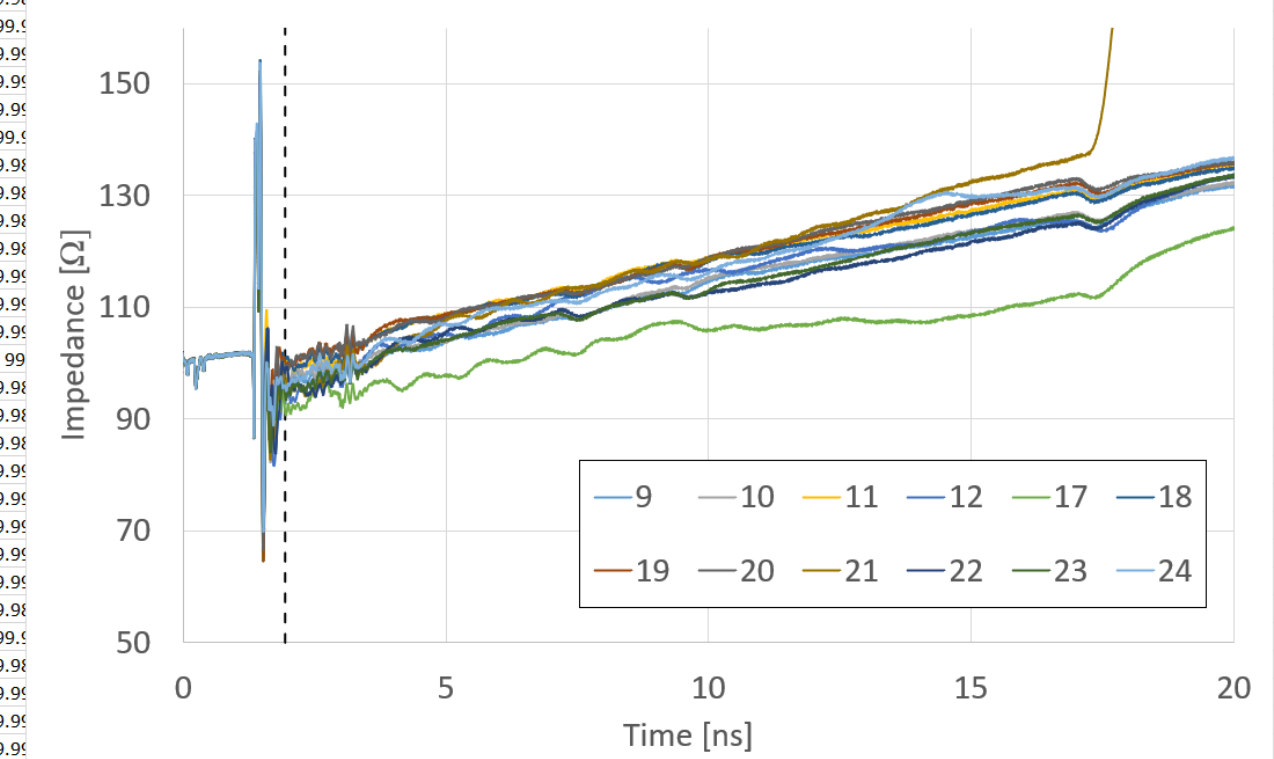
- Using Python to automate analysis of TDR tape data.
- Building a PyROOT framework of code to generate plots from testbeam data.
- Both only partially complete at the present time.

TDR tapes

- Time Domain Reflectometry is used to measure the impedance across the strip staves in the ITk modules, across long differential lines.
- Ideal initial impedance is in the region of 90-110 Ohms, with a maximum loss of 15dB at 1GHz.
- Current methods for measuring the impedance over time as the pulse propagates through the strip are very inefficient.



TDR tapes



Step Rise-T	1.80E-11	Step Rise-T	1.80E-11	Step Rise-T	1.80E-11	Step Rise-T	1.80E-11	Step Rise-T	1.80E-11	Step Rise-T	1.80E-11	Step Rise-T	1.80E-11	Step Rise-T	1.80E-11	Step Rise-T	1.80E-11	Step Rise-T	1.80E-11
Data		Data		Data		Data		Data		Data		Data		Data		Data		Data	
-3.60E-10	99.98931	-3.60E-10	99.98868	-3.60E-10	99.98828	-3.60E-10	99.9883	-3.60E-10	99.98861	-3.60E-10	99.98808	-3.60E-10	99.9884	-3.60E-10	99.98901	-3.60E-10	99.98789	-3.60E-10	99.98846
-3.58E-10	99.98939	-3.58E-10	99.9														99.98787	-3.58E-10	99.98892
-3.55E-10	99.99031	-3.55E-10	99.99														99.98854	-3.55E-10	99.99001
-3.53E-10	99.9915	-3.53E-10	99.99														99.98951	-3.53E-10	99.99105
-3.50E-10	99.9922	-3.50E-10	99.99														99.99015	-3.50E-10	99.99137
-3.48E-10	99.99194	-3.48E-10	99.9														99.99004	-3.48E-10	99.99074
-3.45E-10	99.99083	-3.45E-10	99.98														99.9892	-3.45E-10	99.98952
-3.43E-10	99.98956	-3.43E-10	99.98														99.98814	-3.43E-10	99.98848
-3.40E-10	99.98894	-3.40E-10	99.98														99.98755	-3.40E-10	99.98829
-3.38E-10	99.98941	-3.38E-10	99.98														99.98783	-3.38E-10	99.98911
-3.35E-10	99.9907	-3.35E-10	99.99														99.98884	-3.35E-10	99.99045
-3.33E-10	99.992	-3.33E-10	99.99														99.98995	-3.33E-10	99.99147
-3.30E-10	99.99247	-3.30E-10	99.99														99.99044	-3.30E-10	99.99148
-3.28E-10	99.99176	-3.28E-10	99														99.98995	-3.28E-10	99.99044
-3.25E-10	99.9903	-3.25E-10	99.98														99.98875	-3.25E-10	99.98898
-3.23E-10	99.98902	-3.23E-10	99.98														99.9876	-3.23E-10	99.98801
-3.20E-10	99.98875	-3.20E-10	99.98														99.98726	-3.20E-10	99.9882
-3.18E-10	99.98973	-3.18E-10	99.99														99.98799	-3.18E-10	99.98946
-3.15E-10	99.99137	-3.15E-10	99.99														99.98937	-3.15E-10	99.99102
-3.13E-10	99.99263	-3.13E-10	99.99														99.99054	-3.13E-10	99.9919
-3.10E-10	99.99268	-3.10E-10	99.99														99.99071	-3.10E-10	99.9915
-3.08E-10	99.99141	-3.08E-10	99.99														99.98972	-3.08E-10	99.99003
-3.05E-10	99.9896	-3.05E-10	99.98														99.98815	-3.05E-10	99.98838
-3.03E-10	99.98839	-3.03E-10	99.9														99.98699	-3.03E-10	99.98761
-3.00E-10	99.98861	-3.00E-10	99.98														99.98702	-3.00E-10	99.98826
-2.98E-10	99.99016	-2.98E-10	99.99														99.98829	-2.98E-10	99.98996
-2.95E-10	99.99209	-2.95E-10	99.99														99.99004	-2.95E-10	99.99167
-2.93E-10	99.99316	-2.93E-10	99.99														99.99117	-2.93E-10	99.99229
-2.90E-10	99.99261	-2.90E-10	99.99														99.99092	-2.90E-10	99.99138
-2.88E-10	99.99074	-2.88E-10	99.98971	-2.88E-10	99.99062	-2.88E-10	99.98996	-2.88E-10	99.99046	-2.88E-10	99.98973	-2.88E-10	99.99084	-2.88E-10	99.99053	-2.88E-10	99.98942	-2.88E-10	99.98948
-2.85E-10	99.98869	-2.85E-10	99.98784	-2.85E-10	99.98805	-2.85E-10	99.98773	-2.85E-10	99.98813	-2.85E-10	99.98762	-2.85E-10	99.98811	-2.85E-10	99.9886	-2.85E-10	99.98758	-2.85E-10	99.98781

TDR tapes

- Automatically select input data in the most convenient way possible for the user.
- Detect anomalous datasets and remove these from consideration.
- Smooth inputted data to remove errors created by collection process.
- Calculate impedance values at first and second “bumps” in the data, as well as the signal loss at 1GHz.
- Output results in colour-coded table, to see at a glance which tapes were successful.

TDR tapes

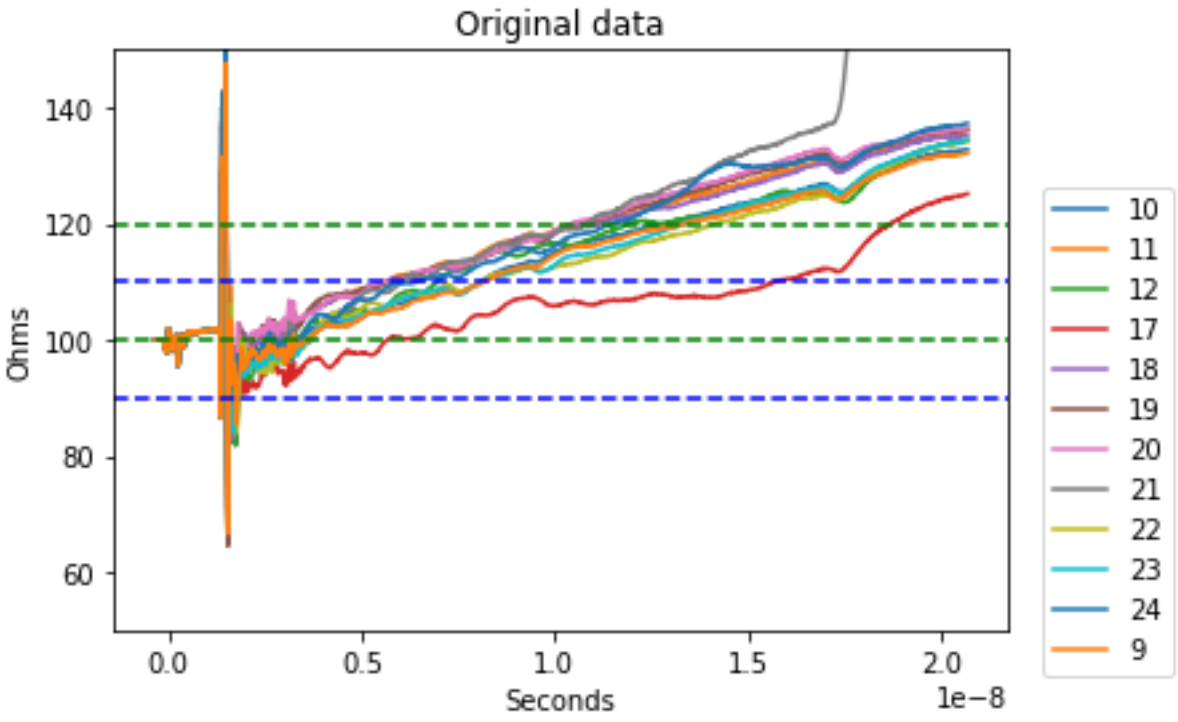
- Controlled by single function with single required argument, as opposed to an unwieldly spreadsheet.
- Data is automatically averaged with its neighbours to remove spiking behaviour inherent to the collection process.
- Non-standard datasets detected and removed from consideration.

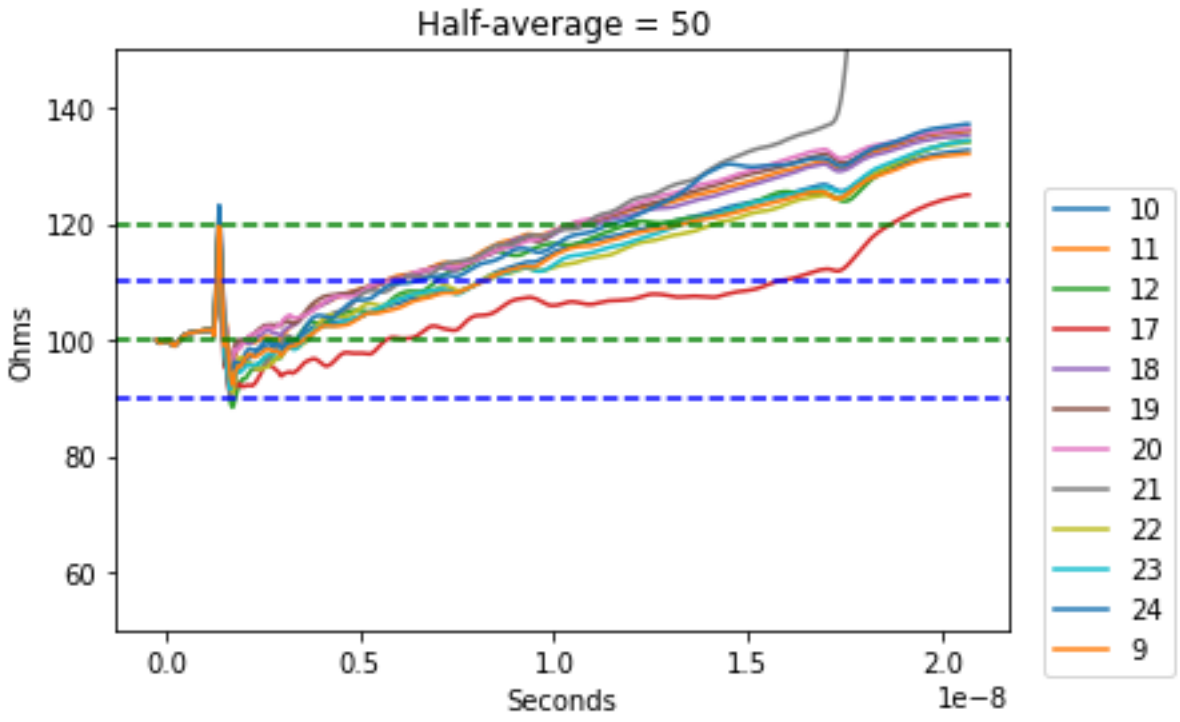
```
folder_location = "download"
folder_location_2 = "download 2"
# Set to . if programme is located in the same folder as the subfolders.

FullPlotterAuto(folder_location_2, 50, 100, "False")
#FullPlotterAuto(folder_location, 50, 100, "True")
```

TDR tapes

- Impedance values at start and end of straight-line section are calculated, and signal loss data is extracted from .txt files. Sensitivity can be tuned for non-standard data.
- Results displayed in green or red to show if they meet pre-set criteria.
- Impedance difference across length of tape is extracted and displayed.
- Currently looking into improving fitting process and automated data analysis, as well as speed and efficiency.

TDR tapes

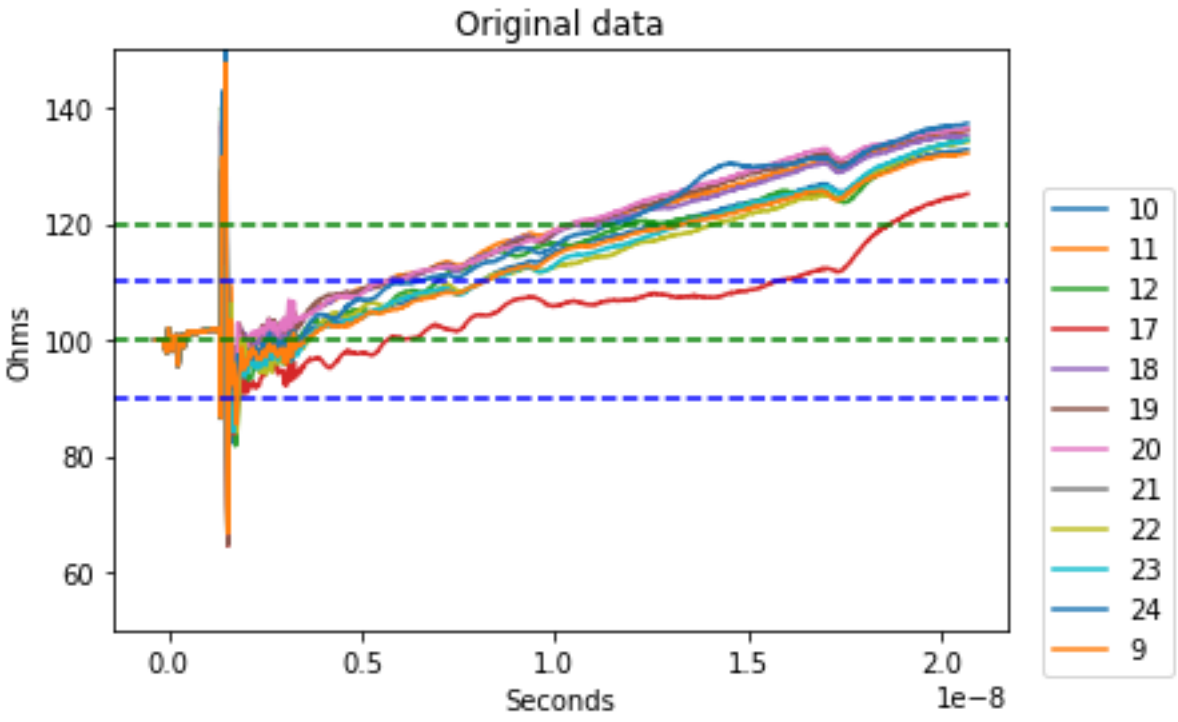

Original data
Ohms
Seconds
1e−8
140
120
100
80
60
0.0
0.5
1.0
1.5
2.0
10
11
12
17
18
19
20
21
22
23
24
9

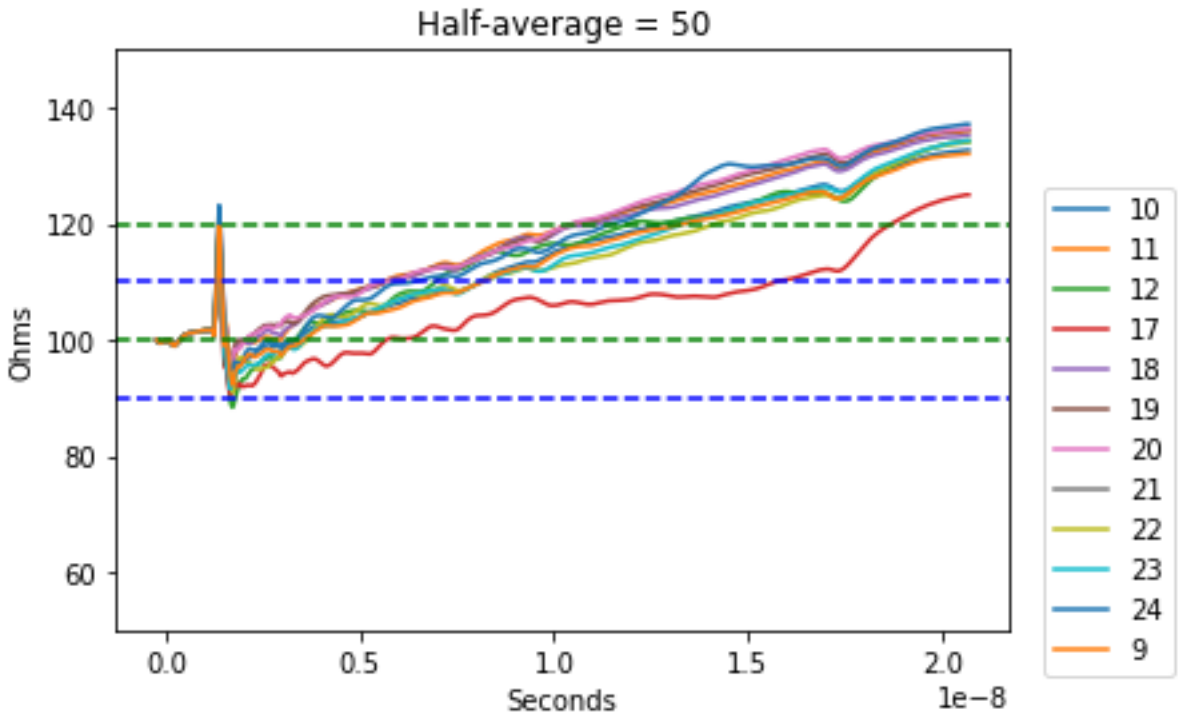

Half-average = 50
Ohms
Seconds
1e−8
140
120
100
80
60
0.0
0.5
1.0
1.5
2.0
10
11
12
17
18
19
20
21
22
23
24
9

TDR tapes

	Initial impedance (ohm)	Final impedance (ohm)	Difference (ohm)	Signal loss at 1GHz(dB)
10	92.32114804554851	125.59076541181177	33.27	-13.6553590426736
11	95.52395435956562	129.49751773012937	33.974	-13.4355286214442
12	88.10431691843945	123.87178790726293	35.767	-14.0255565549991
17	90.68524184250228	111.87337226770325	21.188	-13.6849588469586
18	96.1615444090587	129.05751993285313	32.896	-13.2851079496197
19	96.12395997451347	130.46722002757002	34.343	-13.717968176834
20	94.41459394567563	131.23519260742376	36.821	-13.3399617229489
21	90.03390766293843	119.02266662597368	28.989	-28.5692586586943
22	90.52858283088337	124.24472157183689	33.716	-13.6015238395929
23	91.48533446958854	125.41106719460336	33.926	-13.8660613304559
24	95.00716918747408	129.9261604185611	34.919	-13.4940662818923
9	92.08517972346226	124.35103713608254	32.266	-13.294267897198

TDR tapes


Original data
Ohms
Seconds
1e−8
140
120
100
80
60
0.0
0.5
1.0
1.5
2.0
10
11
12
17
18
19
20
21
22
23
24
9


Half-average = 50
Ohms
Seconds
1e−8
140
120
100
80
60
0.0
0.5
1.0
1.5
2.0
10
11
12
17
18
19
20
21
22
23
24
9

TDR tapes

	Initial impedance (ohm)	Final impedance (ohm)	Difference (ohm)	Signal loss at 1GHz(dB)
10	92.32114804554851	125.59076541181177	33.27	-13.6553590426736
11	95.52395435956562	129.49751773012937	33.974	-13.4355286214442
12	88.10431691843945	123.87178790726293	35.767	-14.0255565549991
17	90.68524184250228	111.87337226770325	21.188	-13.6849588469586
18	96.1615444090587	129.05751993285313	32.896	-13.2851079496197
19	96.12395997451347	130.46722002757002	34.343	-13.717968176834
20	94.41459394567563	131.23519260742376	36.821	-13.3399617229489
21	0.0	nan	nan	-28.569
22	90.52858283088337	124.24472157183689	33.716	-13.6015238395929
23	91.48533446958854	125.41106719460336	33.926	-13.8660613304559
24	95.00716918747408	129.9261604185611	34.919	-13.4940662818923
9	92.08517972346226	124.35103713608254	32.266	-13.294267897198

Testbeam

- DESY Hamburg operates three testbeams for the irradiation of sensitive components.
- These are used to simulate the lifetime damage sustained by components in particle accelerators and nuclear reactors, or other areas where significant irradiation is expected.
- This process generates very large amounts of data, which must be analysed as quickly and efficiently as possible

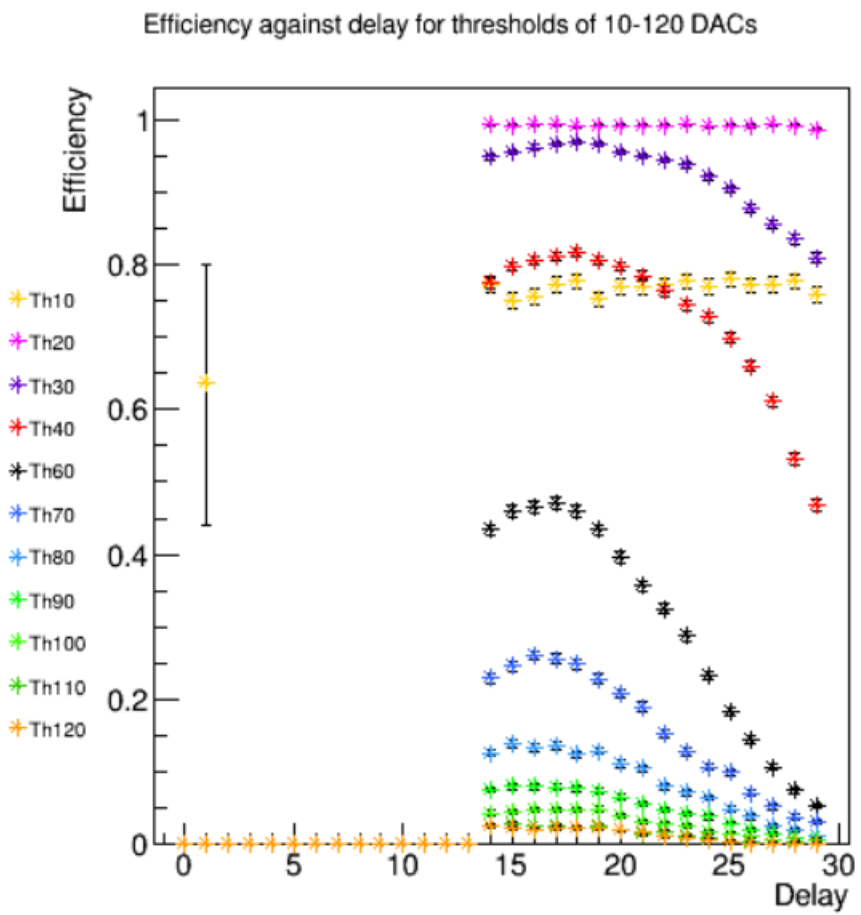
Testbeam

- Aim is to programme a robust framework to produce necessary plots from testbeam data, which can be adapted for new datasets by simply changing file locations and data inputs such as the thresholds used.
- Output produced plots to ROOT file for analysis.

binfractions.png	8.1 kB	11:37
clustermean.png	9.0 kB	11:37
efficiencyplot.png	9.7 kB	11:37
globalresidualsallmeanX.png	9.9 kB	11:37
globalresidualsallmeanY.png	10.0 kB	11:37
globalresidualsallRMSX.png	10.4 kB	11:37
globalresidualsallRMSY.png	10.7 kB	11:37
globalresidualsallUSBPixmeanX.png	11.8 kB	11:37
globalresidualsallUSBPixmeanY.png	9.7 kB	11:37
globalresidualsallUSBPixRMSX.png	10.2 kB	11:37
globalresidualsallUSBPixRMSY.png	9.6 kB	11:37
globalresidualsbyzmeanX.png	9.5 kB	11:37
globalresidualsbyzmeanY.png	9.7 kB	11:37
globalresidualsbyzRMSX.png	11.5 kB	11:37
globalresidualsbyzRMSY.png	11.5 kB	11:37
timingefficiencies.png	14.2 kB	11:37
trackdistancemean.png	10.9 kB	11:37
trackdistancemeanX.png	11.1 kB	11:37
trackdistanceRMS.png	10.2 kB	11:37
trackdistanceRMSX.png	10.5 kB	11:37

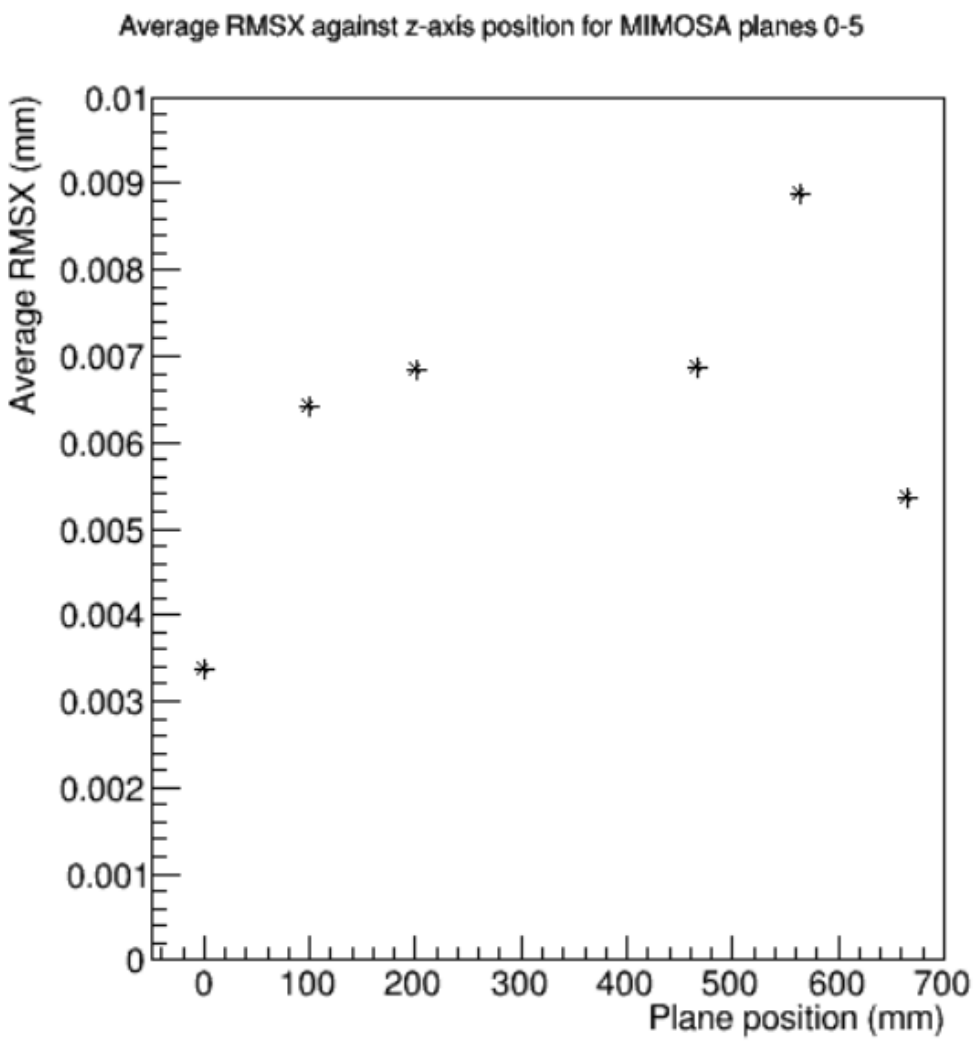
Testbeam

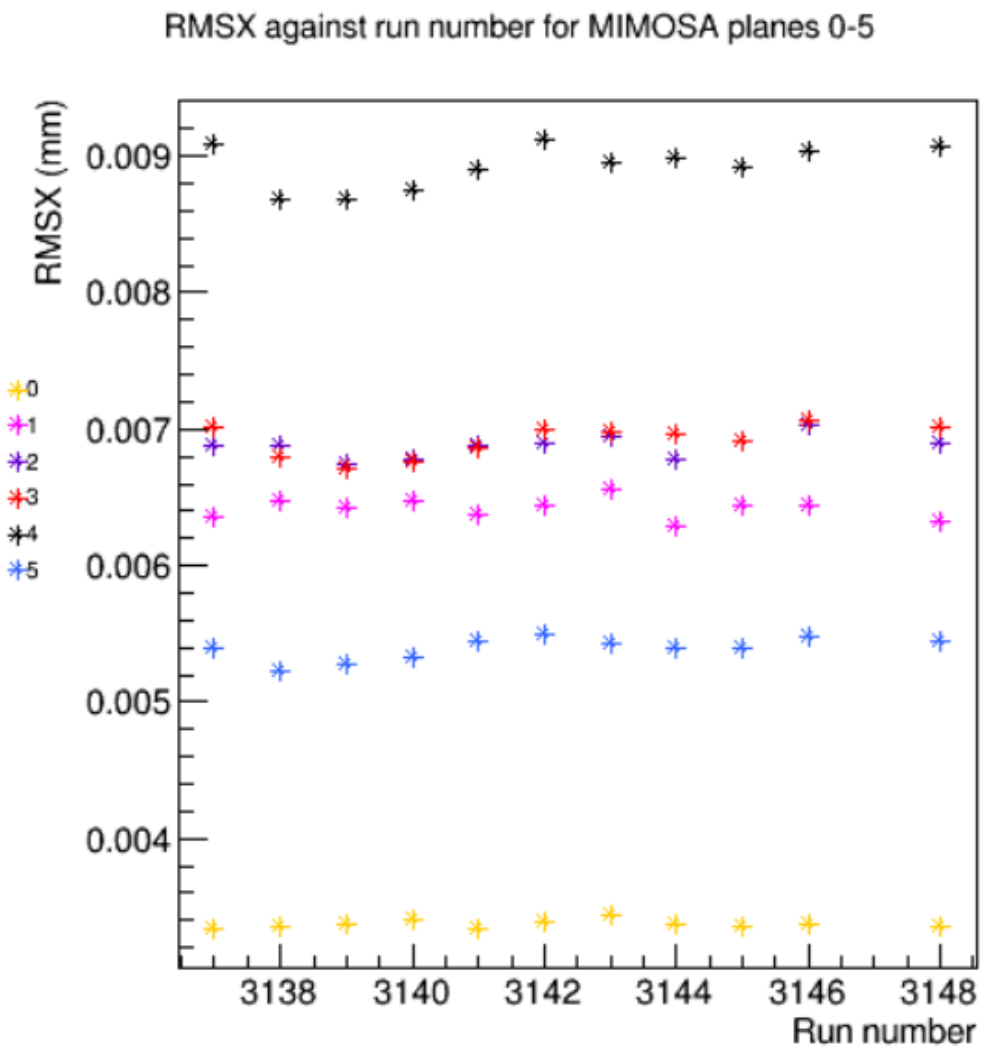
- Plotting framework is inherently customisable, and generates clear plots to specification.
- Some further work to be done in data analysis and output.



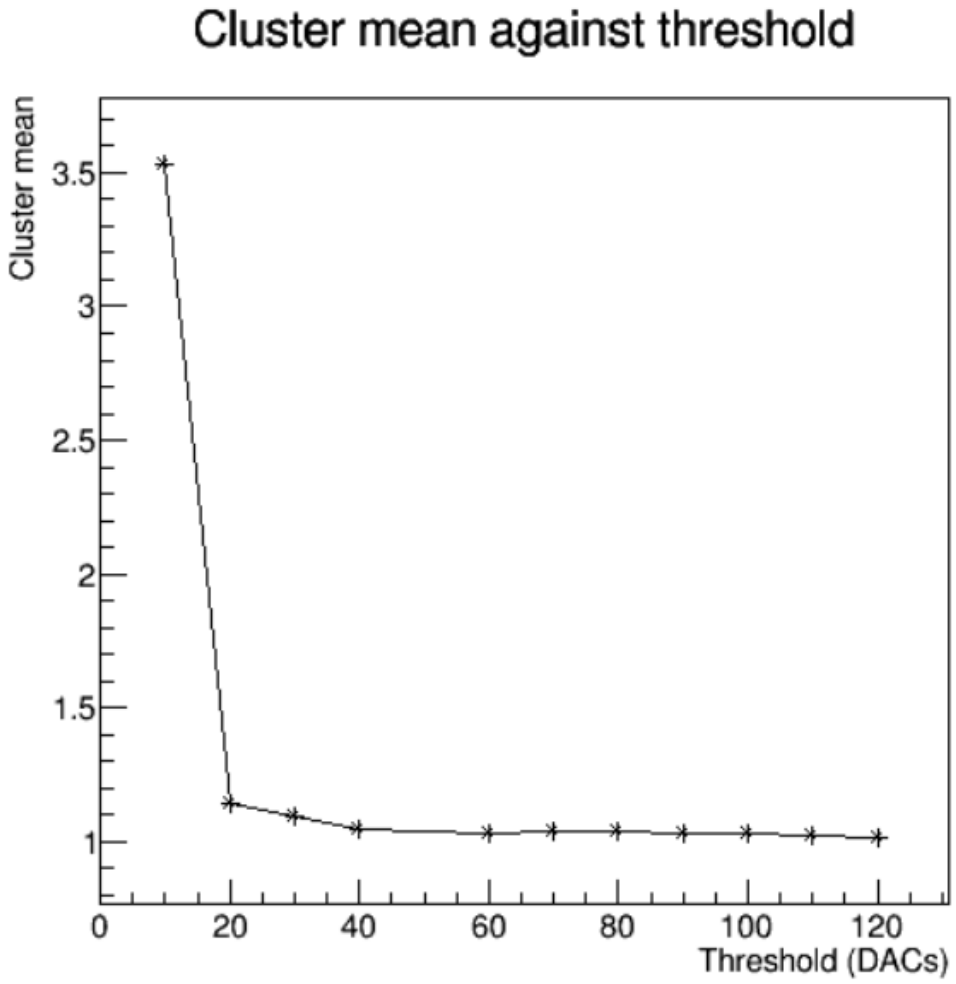

```
Threshold = [10,20,30,40,60,70,80,90,100,110,120]
run_numbers = [3137,3138,3139,3140,3141,3142,3143,3144,3145,3146,3148]
delay = np.asarray([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,2
3,24,25,26,27,28,29])
colour = [kOrange,kMagenta, 882, kRed, kBlack, 858, 861, kGreen, 820, 819, kOran
ge, 797, 802, 803, 636, 632, 634, 894, 629, 893, 921]
output_location_for_td_hists = "DCHists.root"
planes = [0,1,2,3,4,5]
z_vals = [0,100,201,466,565,666]
file_start = "/home/ppd/zdm79899/Public/forDaniel/h_ss"
file_end = "_analysis.root"
```

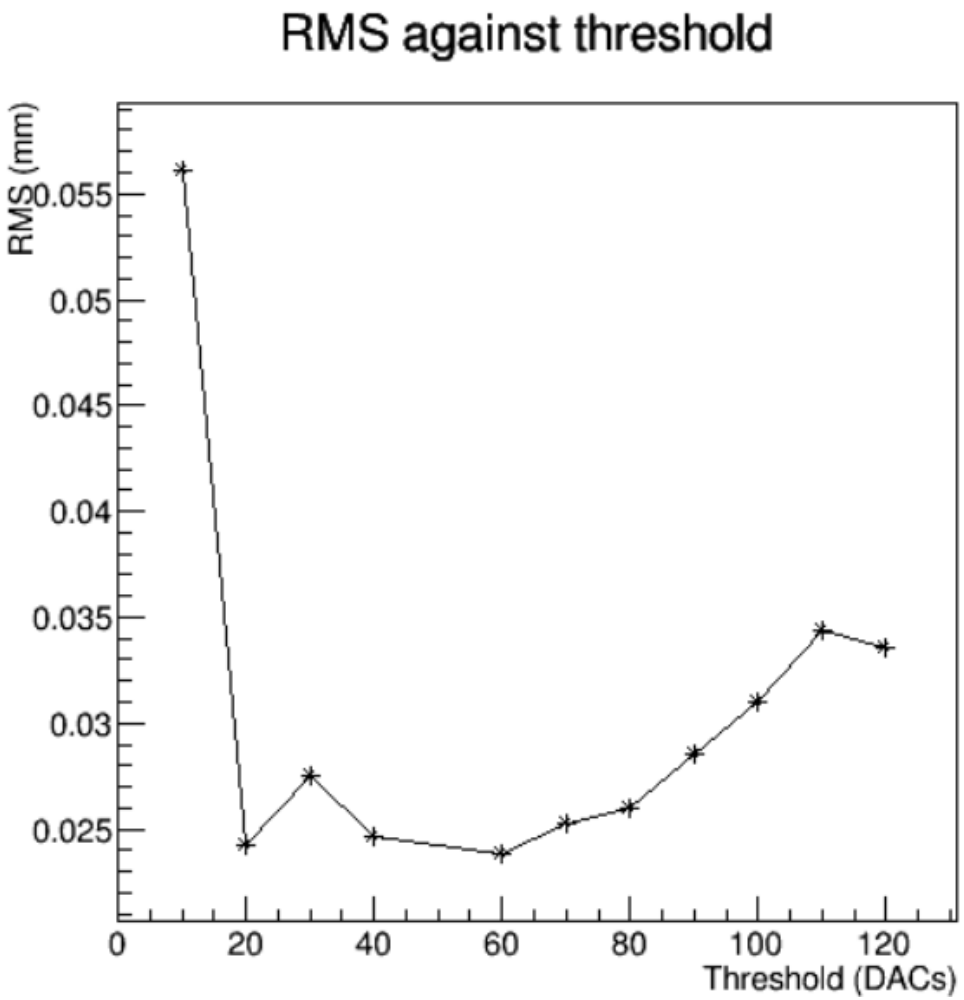
Testbeam


Average RMSX against z-axis position for MIMOSA planes 0-5
Average RMSX (mm)
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
0
100
200
300
400
500
600
700
Plane position (mm)

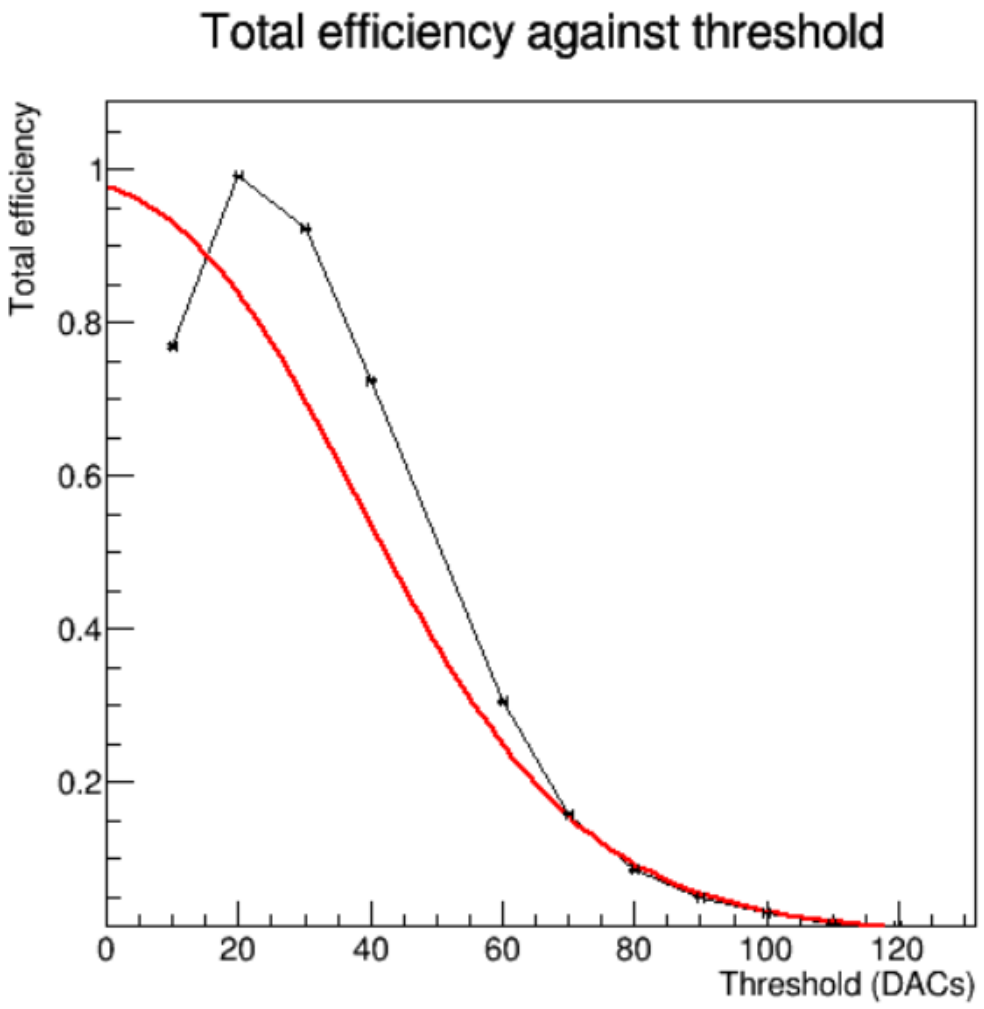

RMSX against run number for MIMOSA planes 0-5
RMSX (mm)
0.009
0.008
0.007
0.006
0.005
0.004
0
1
2
3
4
5
3138
3140
3142
3144
3146
3148
Run number

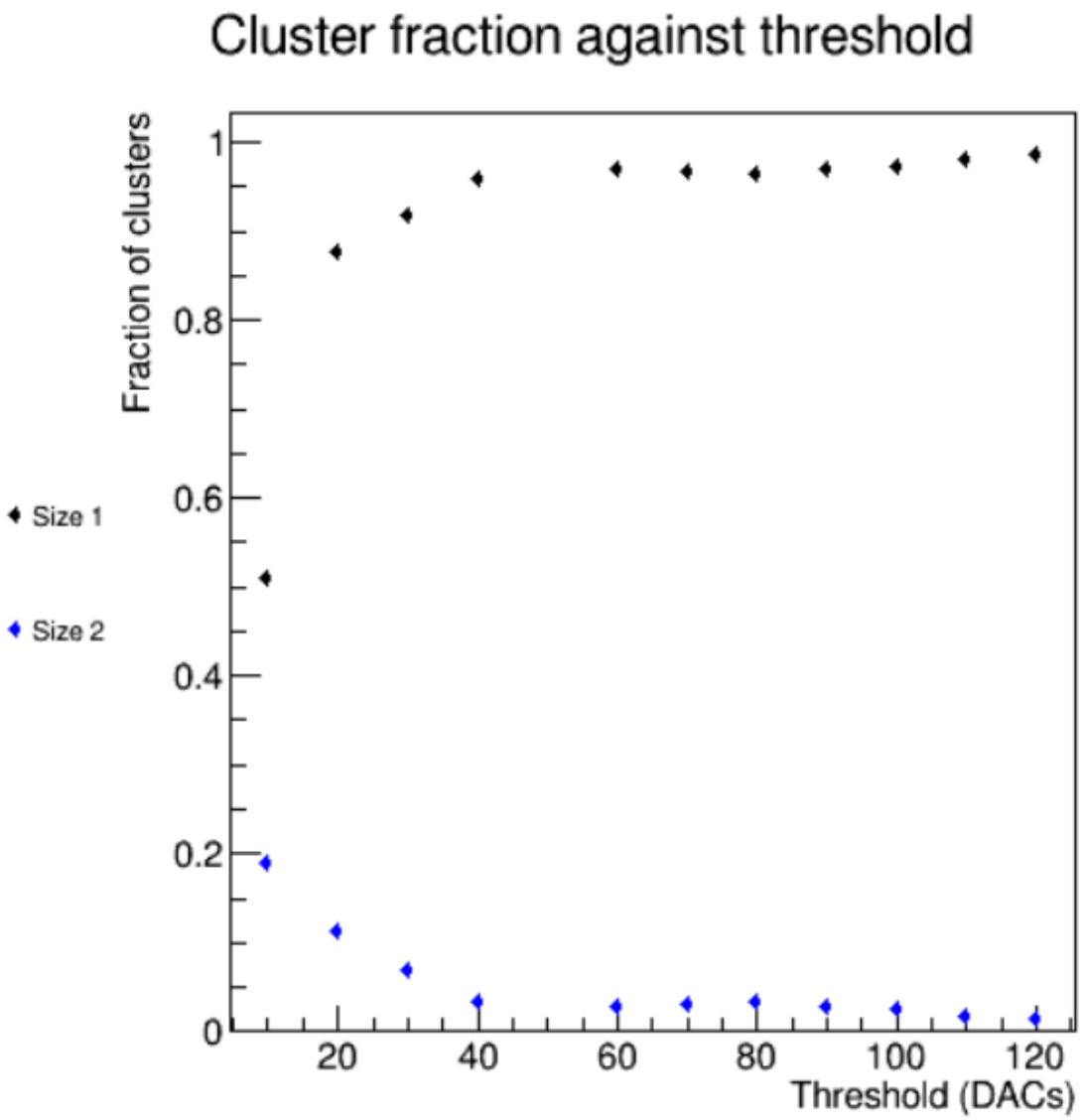
Testbeam


Cluster mean against threshold
Cluster mean
3.5
3
2.5
2
1.5
1
0
20
40
60
80
100
120
Threshold (DACs)


RMS against threshold
RMS (mm)
0.055
0.05
0.045
0.04
0.035
0.03
0.025
0
20
40
60
80
100
120
Threshold (DACs)

Testbeam


Total efficiency against threshold
Total efficiency
1
0.8
0.6
0.4
0.2
0
20
40
60
80
100
120
Threshold (DACs)


Cluster fraction against threshold
Fraction of clusters
Size 1
Size 2
1
0.8
0.6
0.4
0.2
0
20
40
60
80
100
120
Threshold (DACs)