Progress on the peak finding algorithm

Guang Zhao

zhaog@ihep.ac.cn

February 18, 2022

Outline

- **Peak finding with derivatives**
- **Peak finding with deep learning (preliminary)**

Peak finding with derivatives

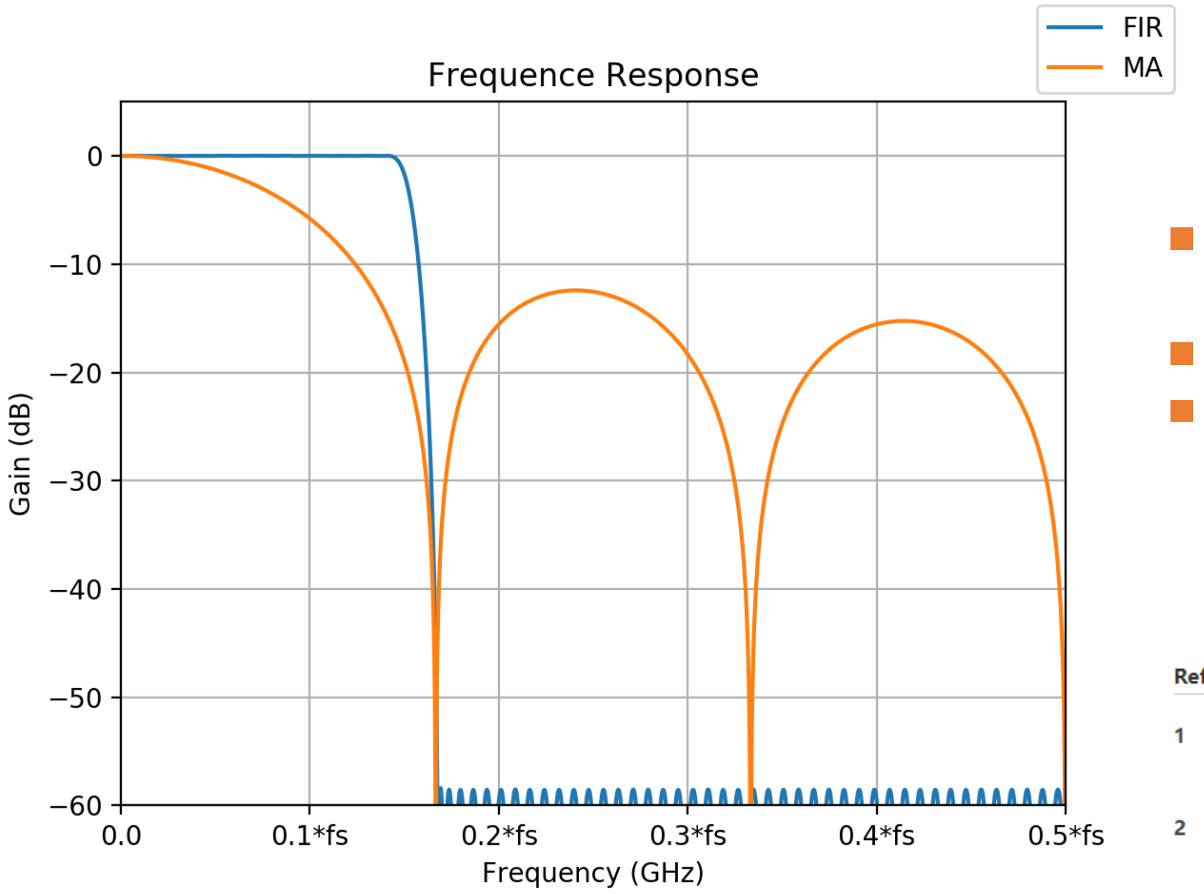
Peak finding with derivatives

- **Low pass filter (smoothing)**
  - Filter out high frequency noises in the waveforms in order to improve the S/N ratio
  - Finite impulse response (**FIR**) filter with a cutoff frequency (**F**): FIR[i] = a0*x[n] + a1*x[n-1] + ⋯

- **Derivative (peak detection)**
  - First derivative (**D1**): D1[i] = FIR[i] – FIR[i – **G**]
  - Second derivative (**D2**): D2[i] = D1[i] – D1[i – **G**]
  - Hit detection: threshold passing (**T**)

Parameters:
- Cutoff frequency
- Derivative step
- Threshold

Smoothing with FIR



- **Digital filter:** Reduce impact from noises with high frequencies
- **Moving average:** Poor frequency response
- **Optimal filter** with Remez exchange algorithm
  - Fast roll-off
  - Good stopband attenuation

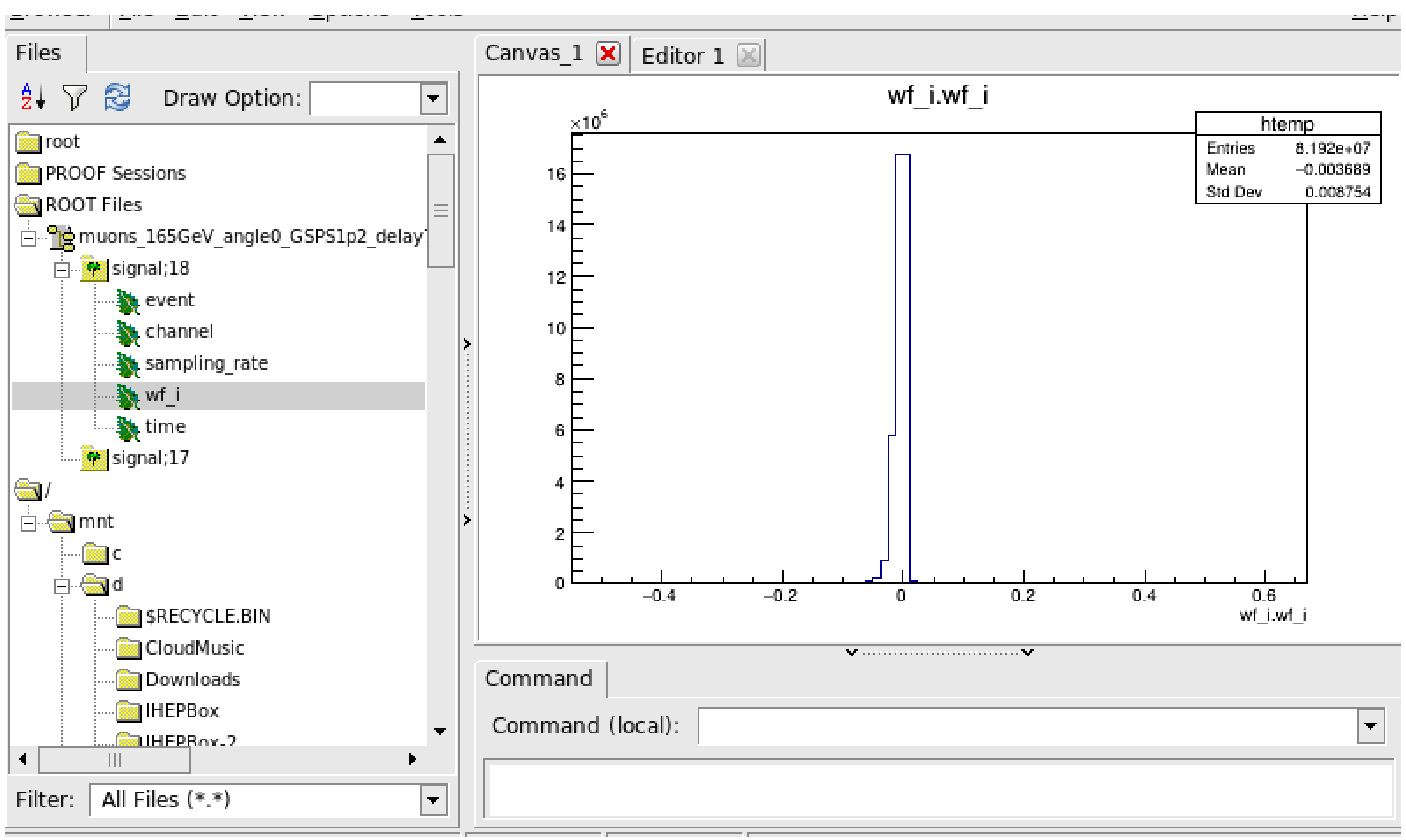
References

1 J. H. McClellan and T. W. Parks, "A unified approach to the design of optimum FIR linear phase digital filters", IEEE Trans. Circuit Theory, vol. CT-20, pp. 697-701, 1973.

2 J. H. McClellan, T. W. Parks and L. R. Rabiner, "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters", IEEE Trans. Audio Electroacoust., vol. AU-21, pp. 506-525, 1973.

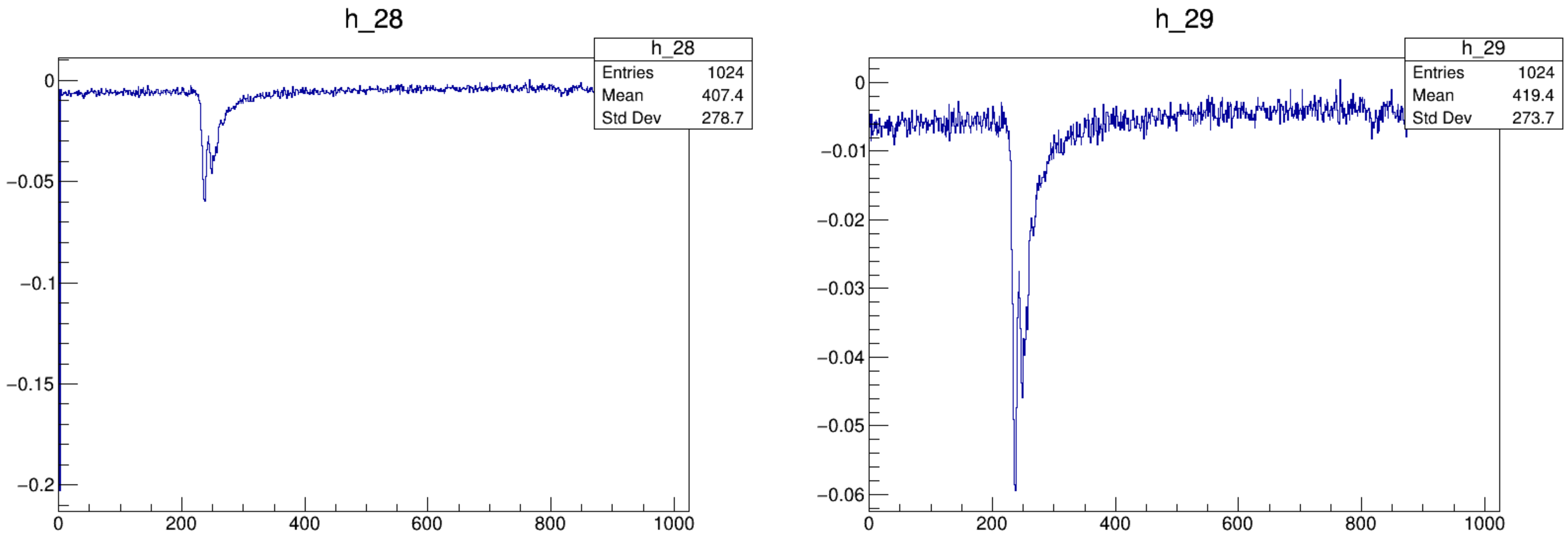
Data preprocessing: converter

Word	Byte 0	Byte 1	Byte 2	Byte 3	Contents
0	'D'	'R'	'S'	'8'	File header, Byte 3 = version
1	'T'	'I'	'M'	'E'	Time Header
2	'B'	'#'	Board number		Board serial number
3	'C'	'0'	'0'	'0'	Channel 0 header
4	Time Bin Width #0				Effective time bin width in ns for channel 0 encoded in 4-Byte floating point format
5	Time Bin Width #1				
...	...				
1027	Time Bin Width #1023				
1028	'C'	'0'	'0'	'1'	Channel 1 header
1029	Time Bin Width #0				Effective time bin width in ns for channel 1 encoded in 4-Byte floating point format
1030	Time Bin Width #1				
...	...				
2052	Time Bin Width #1023				
2053	'E'	'H'	'D'	'R'	Event Header
2054	Event Serial Number				Serial number starting with 1
2055	Year		Month		Event date/time 16-bit values
2056	Day		Hour		
2057	Minute		Second		
2058	Millisecond		Range		Range center (RC) in mV
2059	'B'	'#'	Board number		Board serial number
2060	'C'	'0'	'0'	'0'	Channel 0 header
2061	Scaler #1				Scaler for channel 0 in Hz
2062	'T'	'#'	Trigger cell		Channel 0 first readout cell
2063	Voltage Bin #0		Voltage Bin #1		Channel 0 waveform data encoded in 2-Byte integers. 0=RC-0.5V and 65535=RC+0.5V. RC see header.
2064	Voltage Bin #2		Voltage Bin #3		
...	...		...		
2574	Voltage Bin #1022		Voltage Bin #1023		
2575	'C'	'0'	'0'	'1'	Channel 1 header
2576	Scaler #2				Scaler for channel 1 in Hz
2077	'T'	'#'	Trigger cell		Channel 1 first readout cell
2578	Voltage Bin #0		Voltage Bin #1		Channel 1 waveform data encoded in 2-Byte integers. 0=RC-0.5V and 65535=RC+0.5V. RC see header.
2579	Voltage Bin #2		Voltage Bin #3		
...	...		...		
3089	Voltage Bin #1022		Voltage Bin #1023		
3090	'E'	'H'	'D'	'R'	Next Event Header
...					



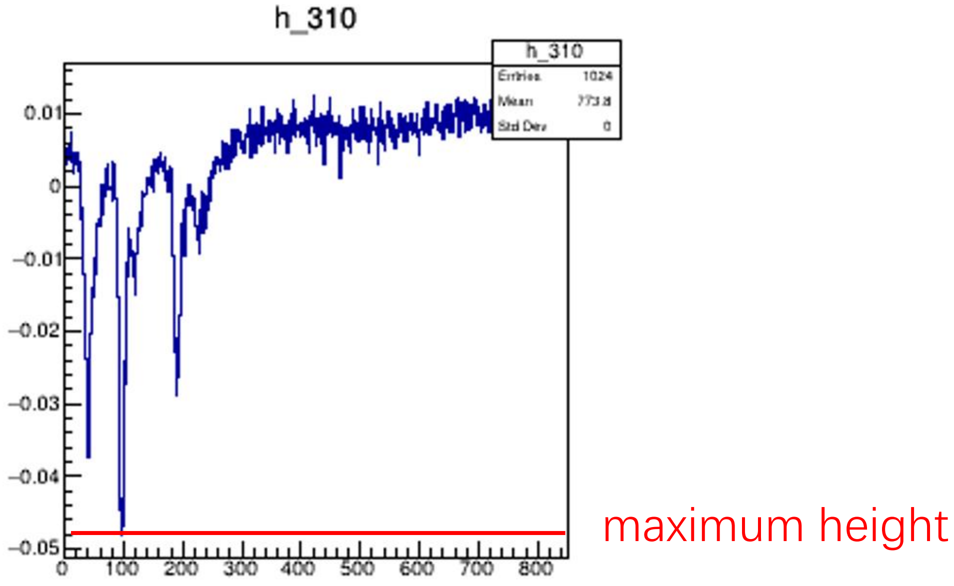
Convert the binaries to ROOT format that compatible to the counting algorithm

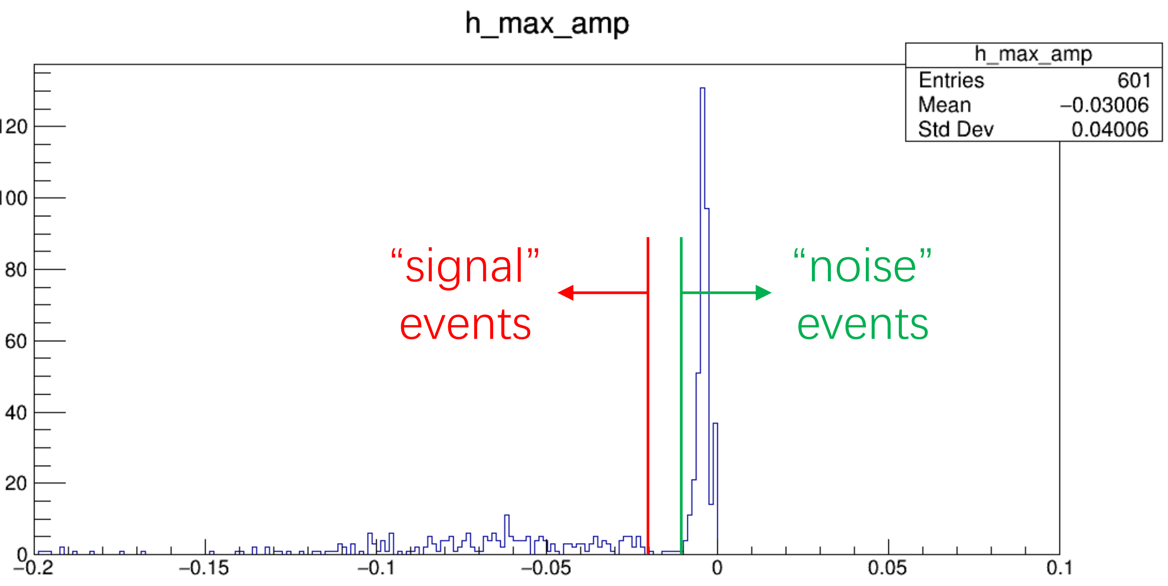
Data preprocessing: Outlier removal



Some outliers appear at the very beginning or ending of the waveform.
Detect and remove them automatically by an algorithm.

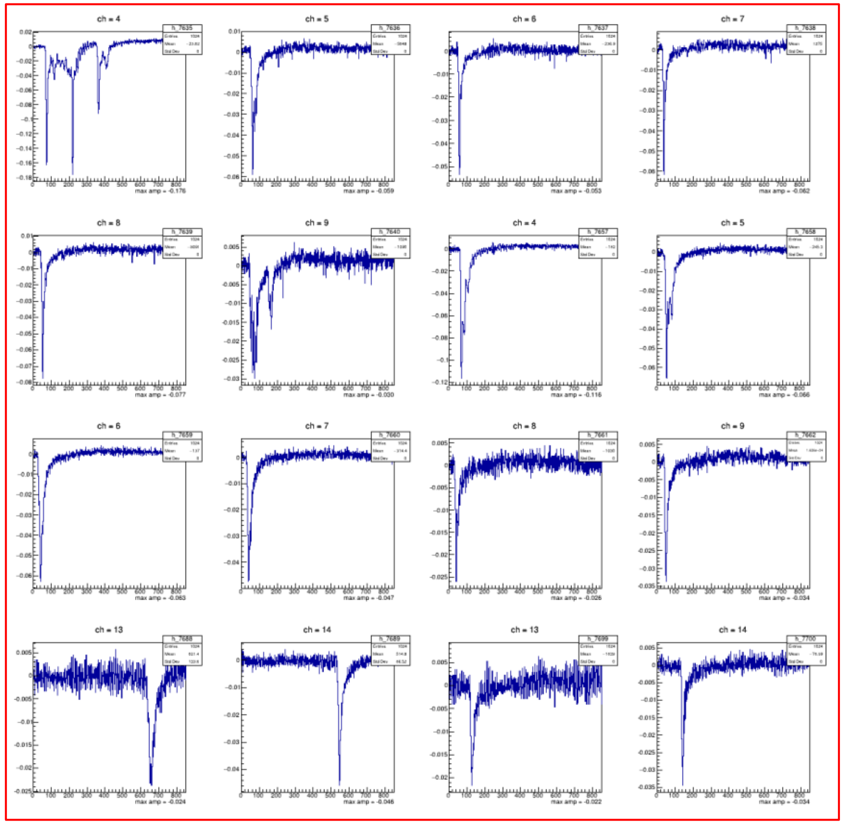
Data preprocessing: classification

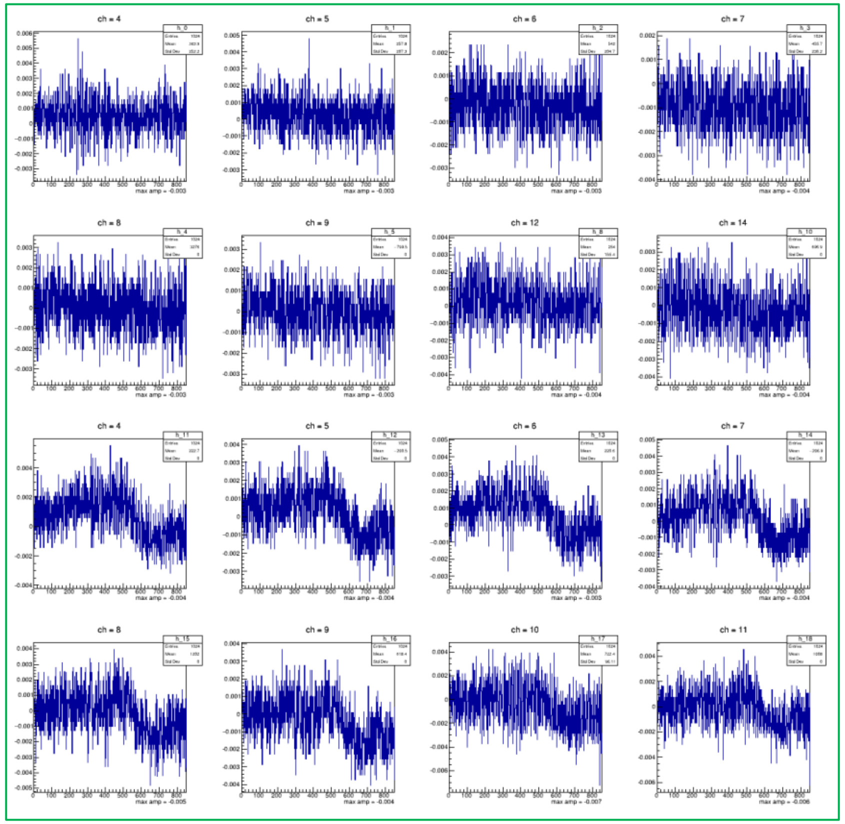




Classify waveforms to "signals" and "noises"

Data preprocessing: classification (II)

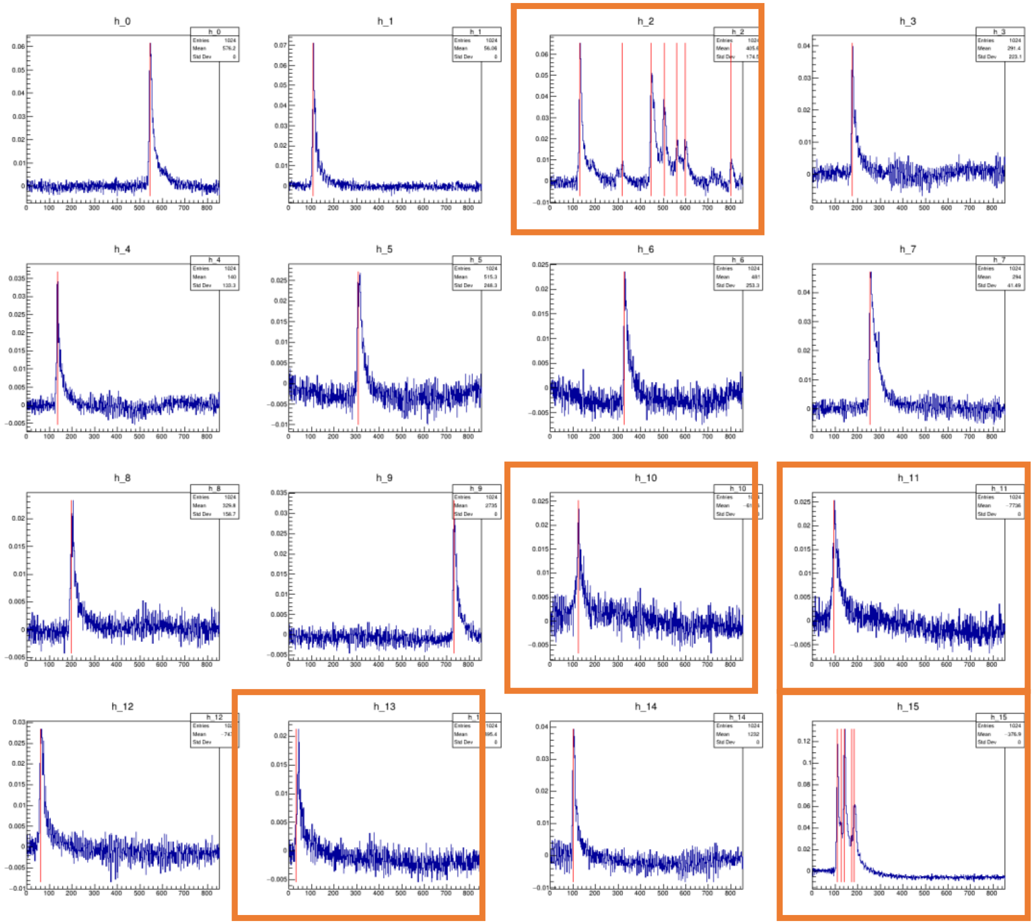




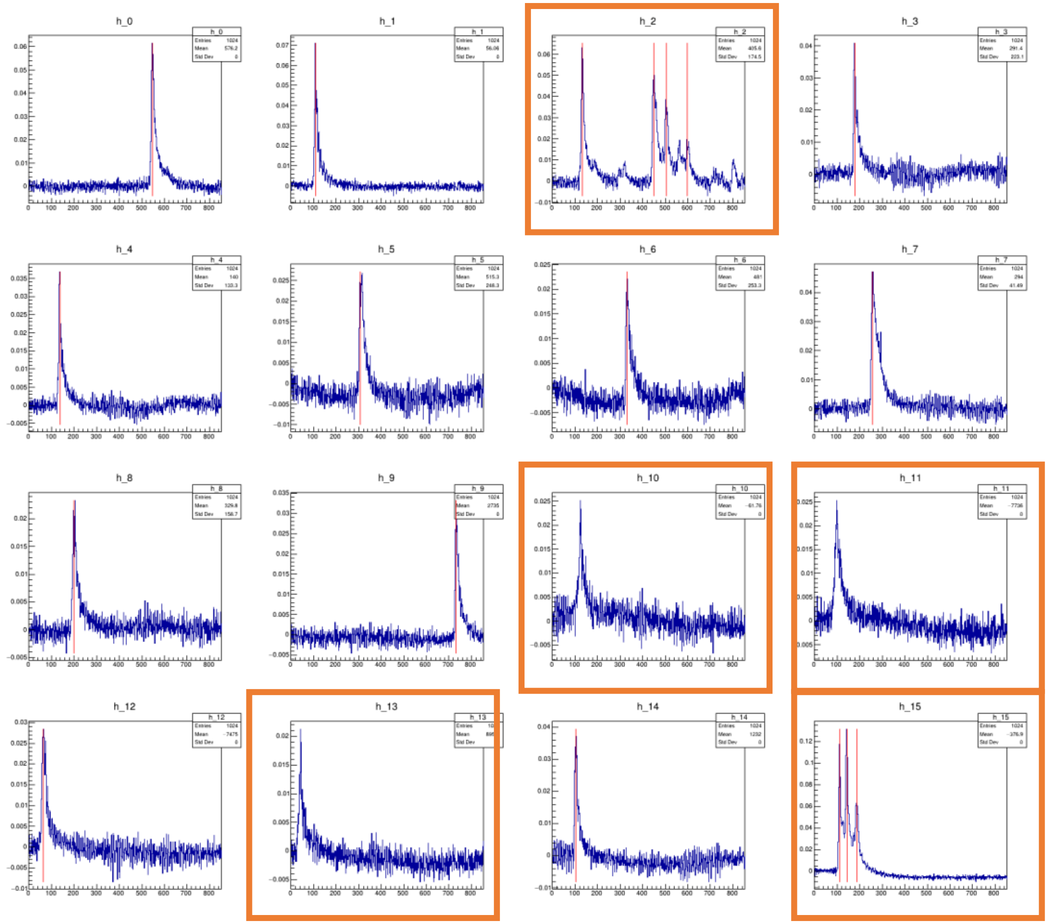
Further optimize the peak finding algorithm by using these classified events. Scan the parameters and control the fake rate with the “noise” events. Look at the peak finding with the “signal” events

Peak finding

FIR



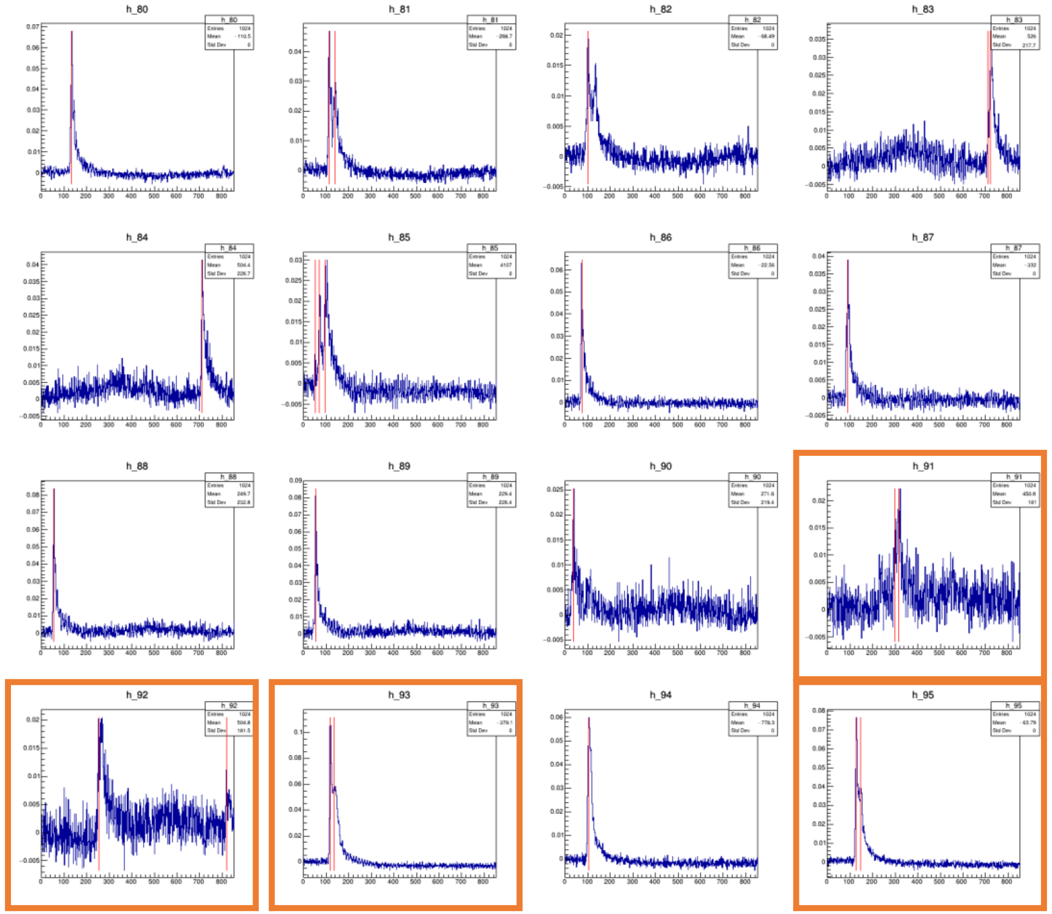
MA



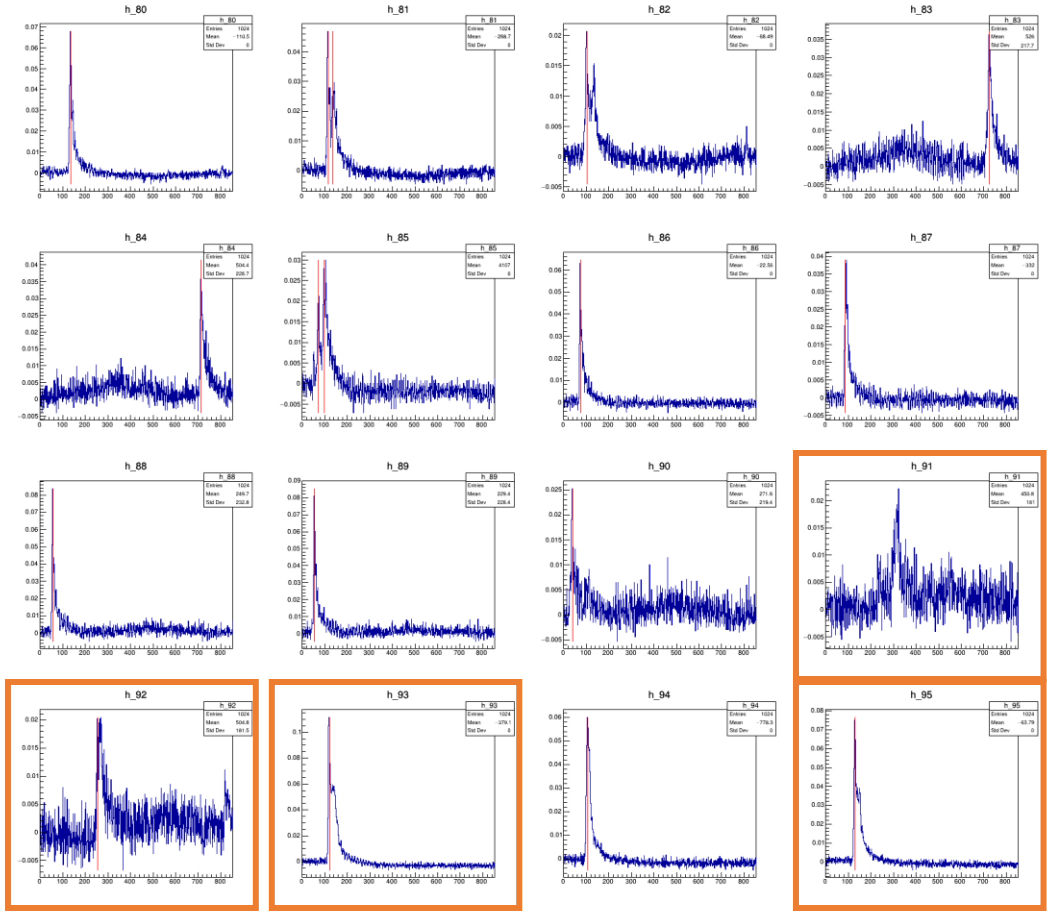
muons_165GeV_angle0_GSPS1p2_delay725ns_7Nov_0321

Peak finding (II)

FIR



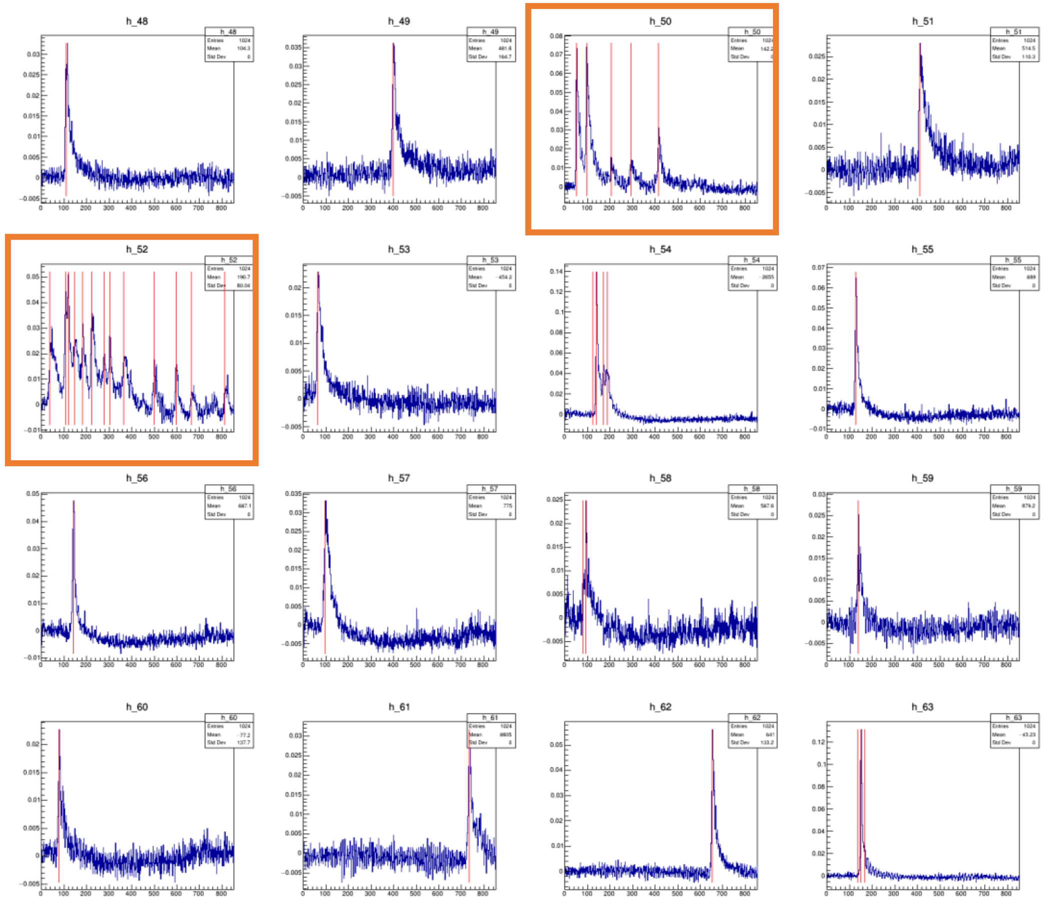
MA



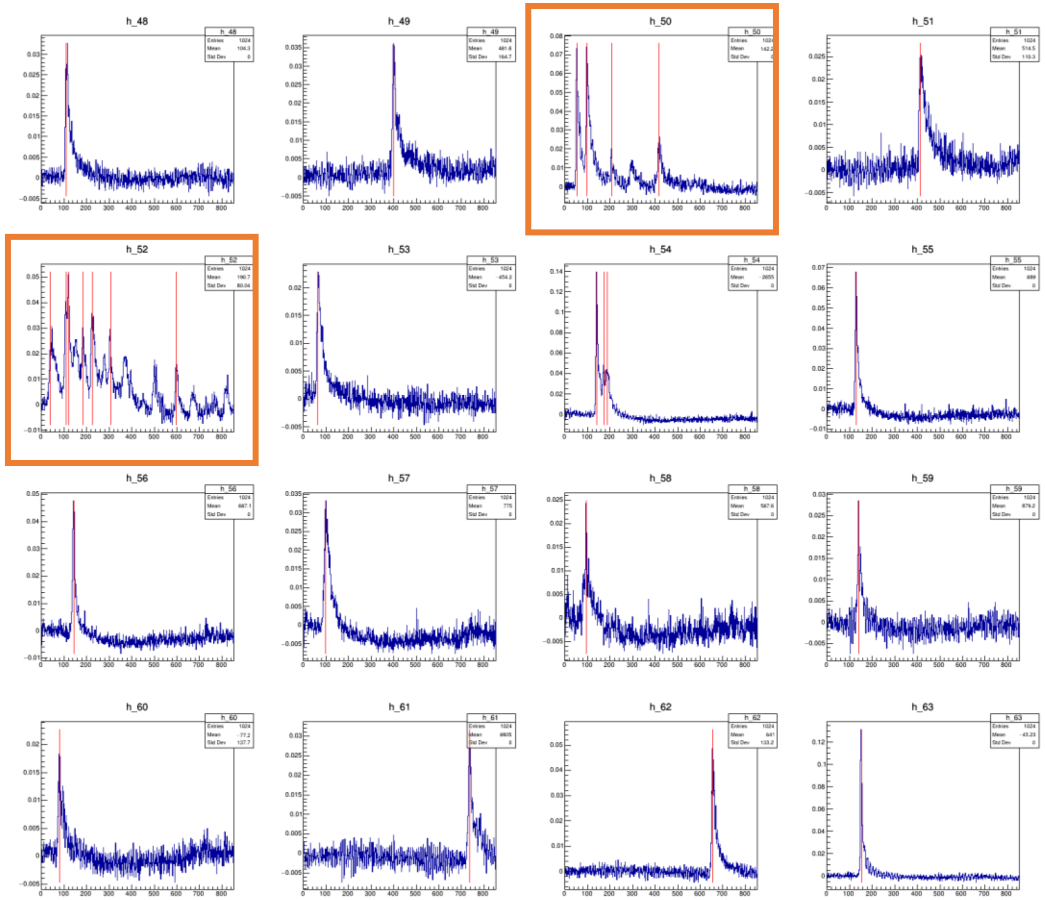
muons_165GeV_angle0_GSPS1p2_delay725ns_7Nov_0321

Peak finding (III)

FIR



MA



Conclusion

- **Using FIR filter instead of MA, as the former one gives much better frequency response.**

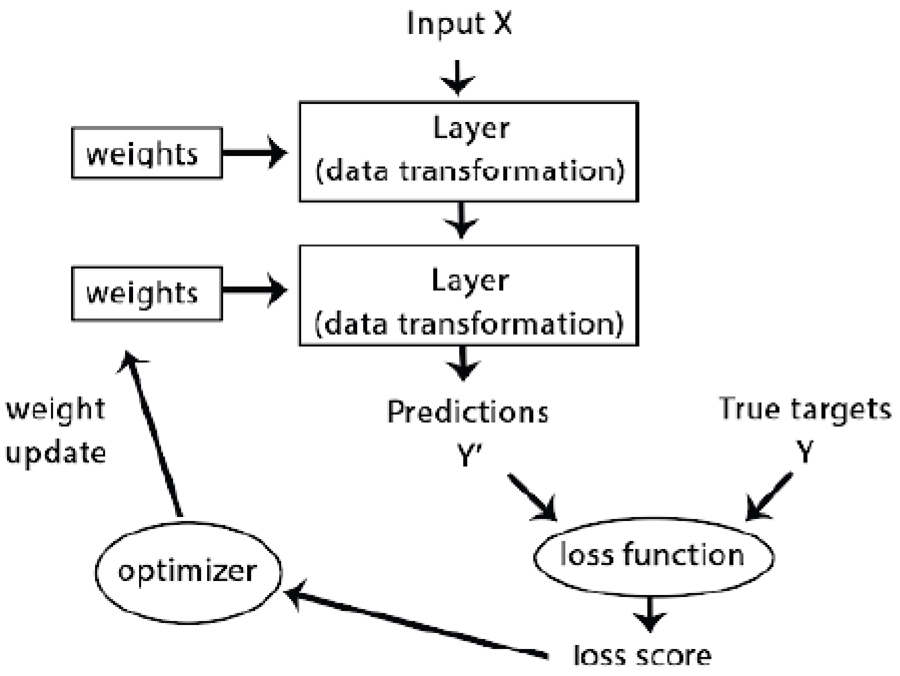
- **Apply the peak finding algorithms with derivatives to the beam test data**
  - Can detect most of the major peaks
  - Will focus on the small peaks. Need to discriminate them from noises

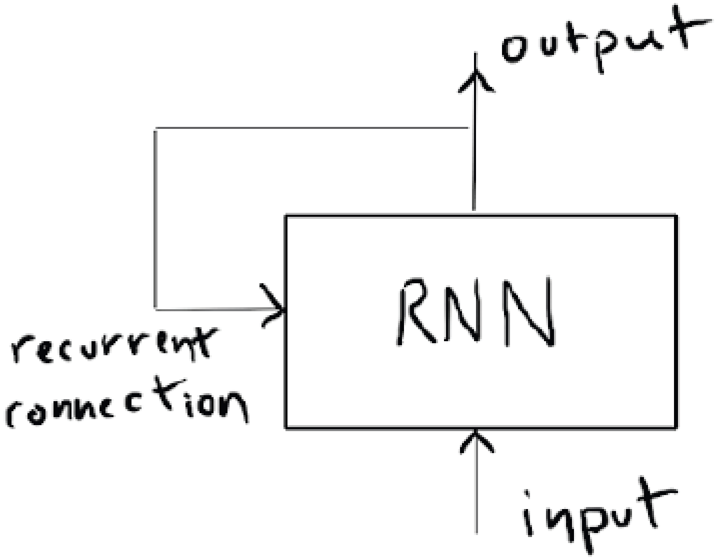
- **Open question:**
  - Why most waveforms only has one “big” peak?

Peak finding with deep learning

Peak finding with deep learning

For time-sequence problem





Densely-connected network:

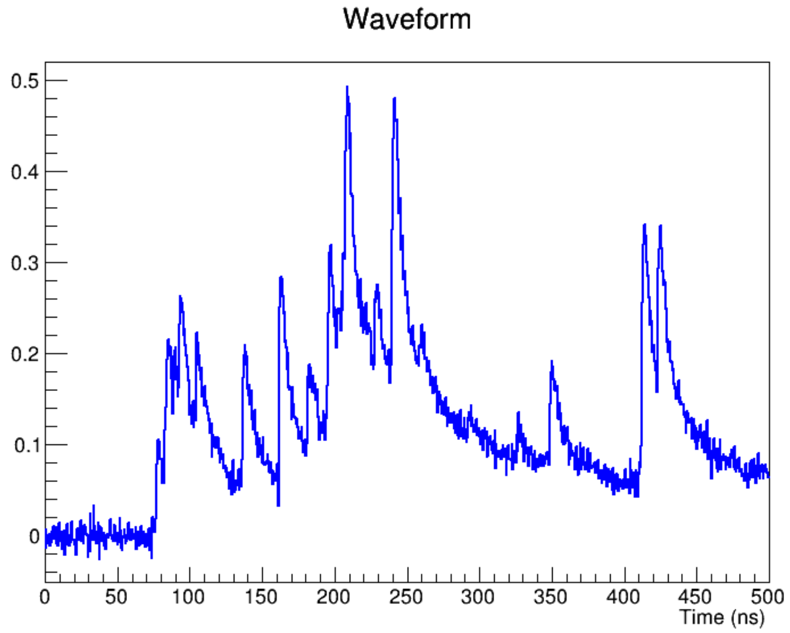
- No memory among sequence elements

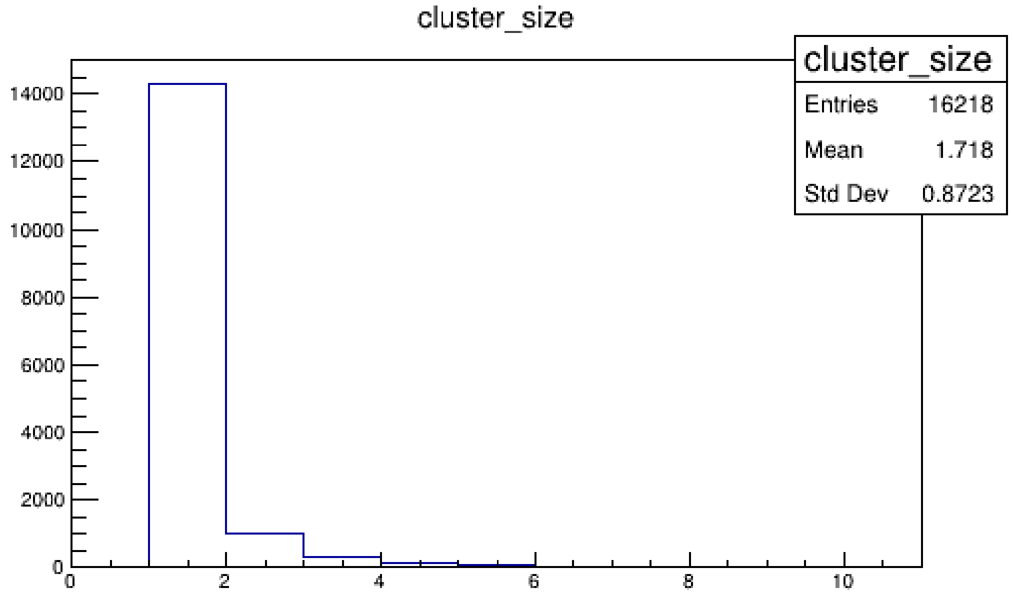
Recurrent Neural Network (RNN):

- Internal loops over sequence elements. Has "memories"
- Powerful to handle time-sequence problems

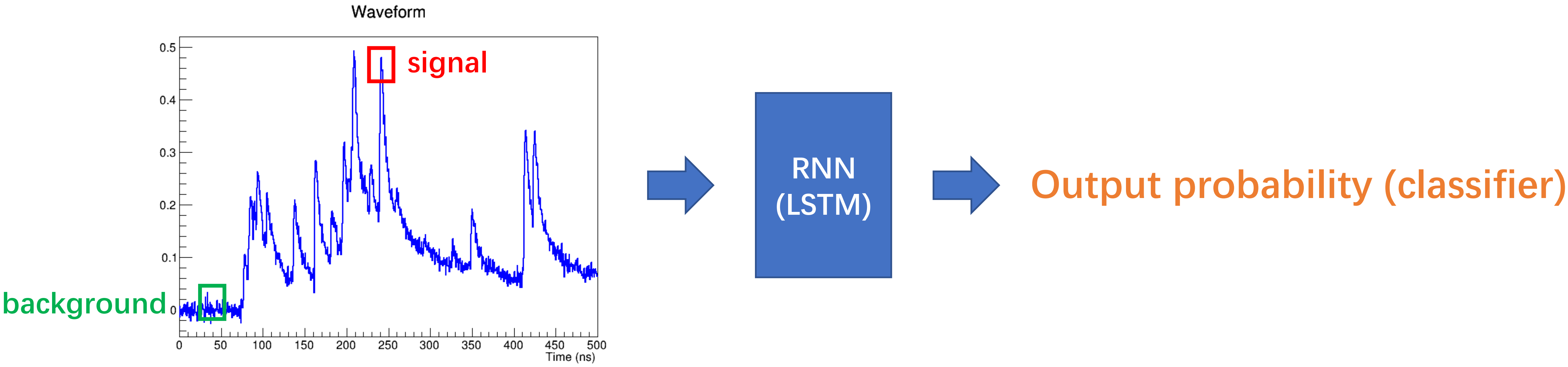
Dataset

- **Toy Monte Carlo samples**
  - ~20 primary ionizations per waveform (~90% cluster size == 1)
  - ~10% noise level
  - 10000 samples





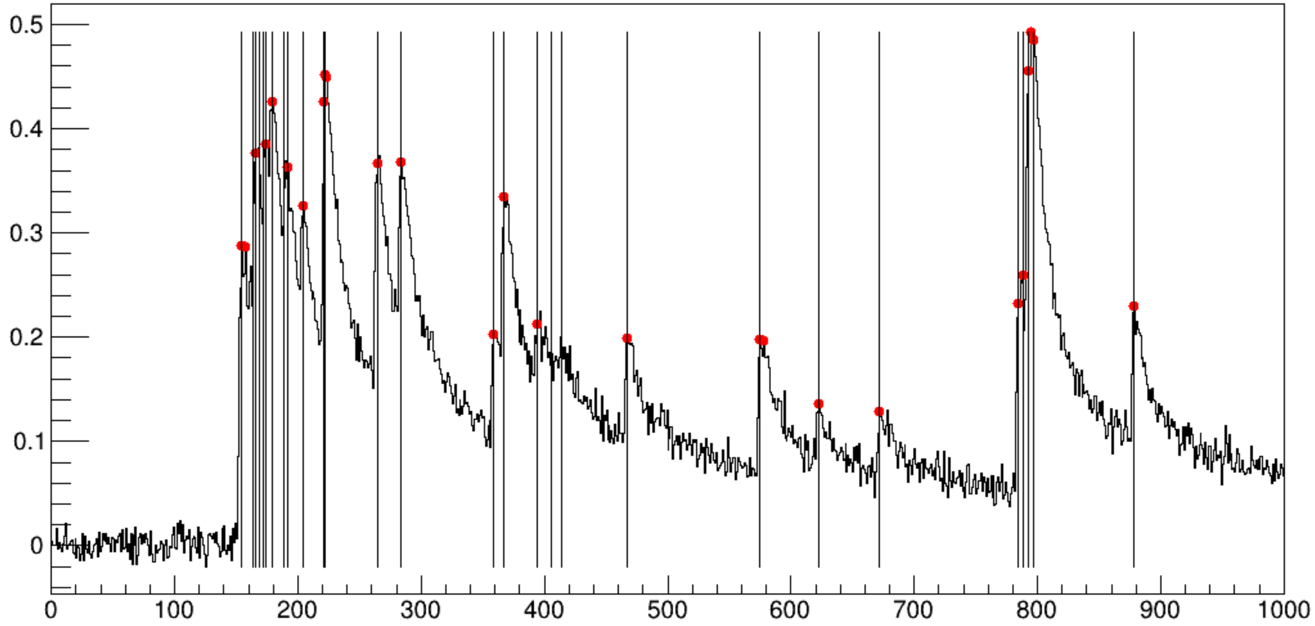
Model



- Using slices of the waveform as the inputs
- Using LSTM (a variation of RNN) to train the dataset
- Preliminary results show good performances (training has not been fully optimized yet)

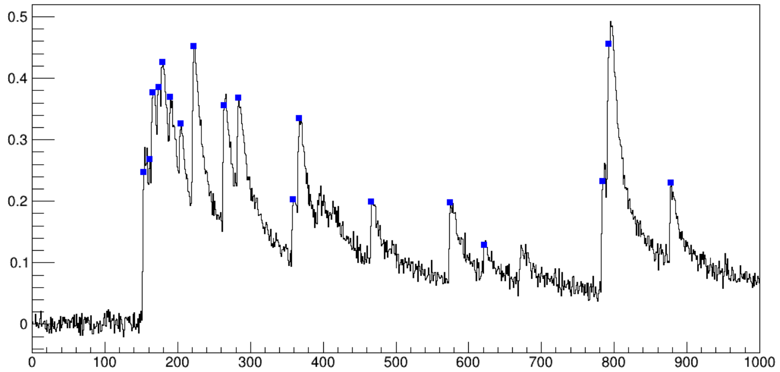
Peak finding

RNN (LSTM)



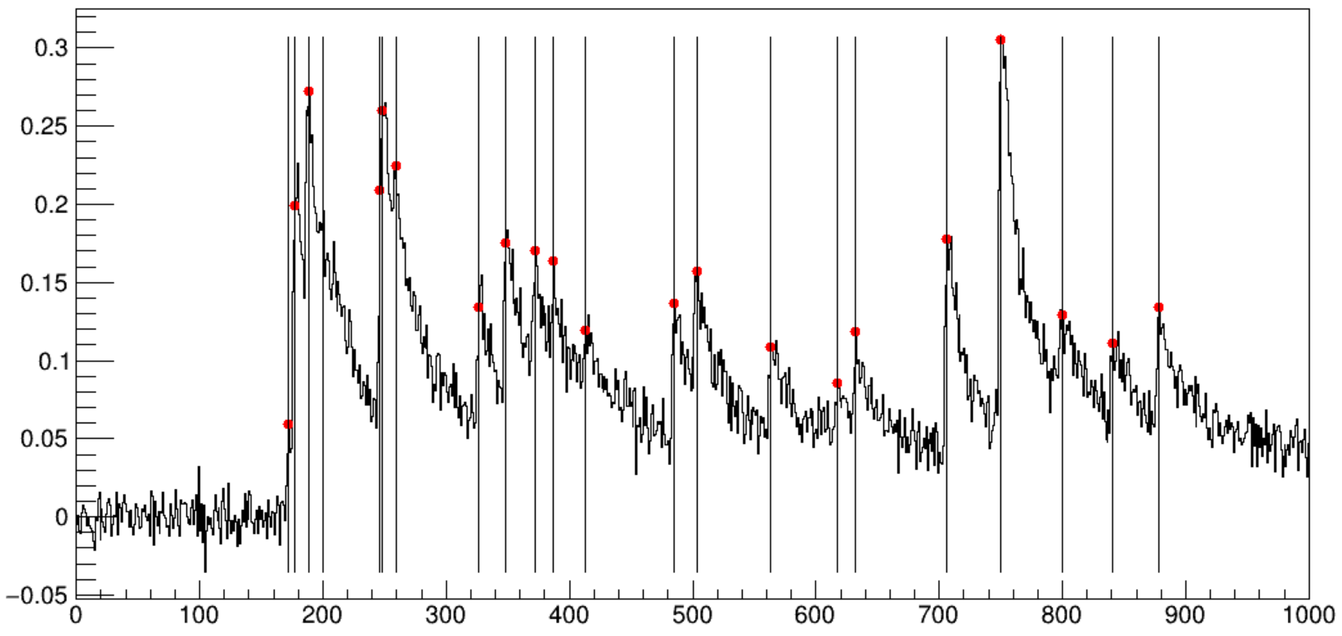
Black line: truth (primaries and secondaries)

Derivative



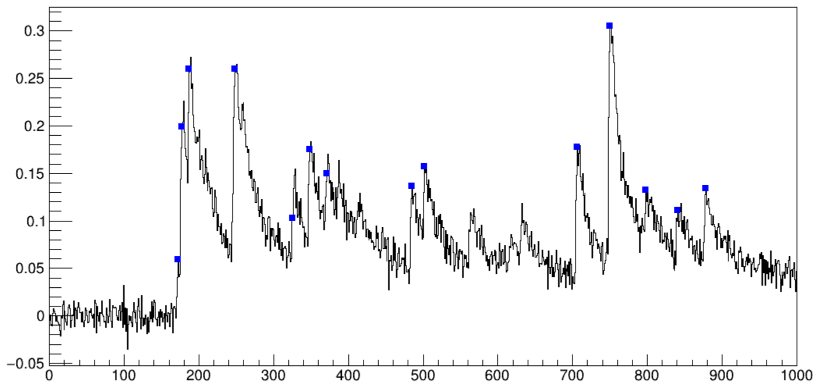
Peak finding (II)

RNN (LSTM)



Black line: truth (primaries and secondaries)

Derivative



Conclusion

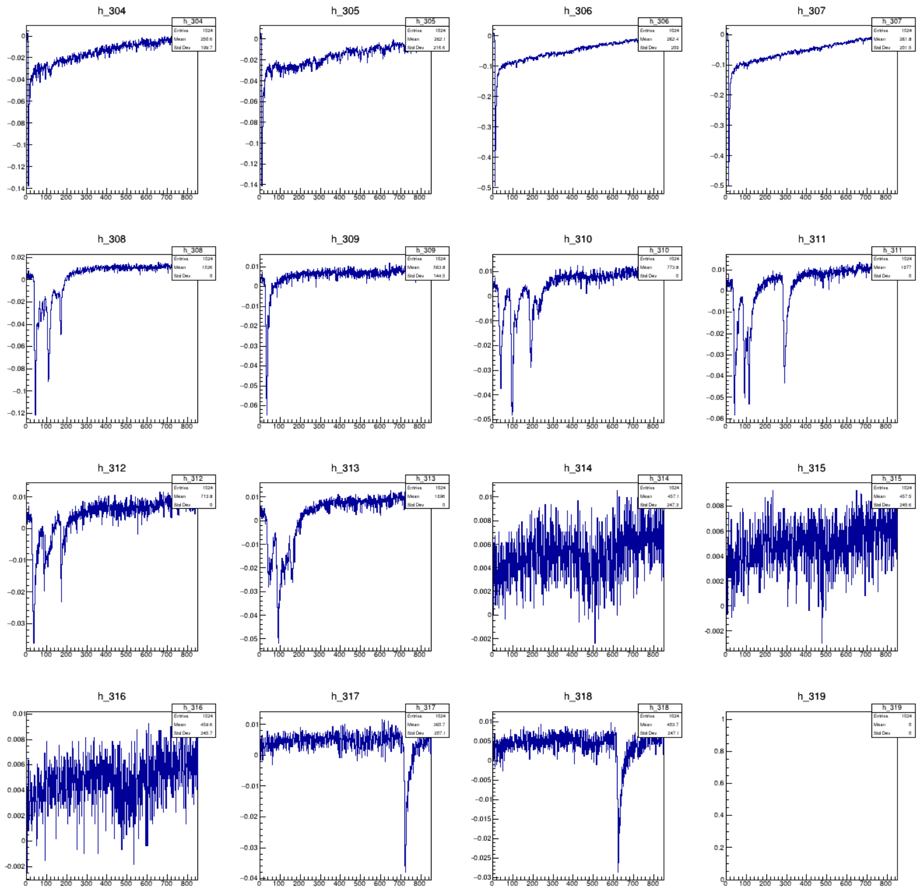
- **Peak finding with the LSTM model show good performance for MC samples**

- **Next to do**
  - Need further tunning the model
  - Will make more detailed comparisons to the derivative-method
  - Consider to classify primary peaks and secondary peaks

- **Open question:**
  - How to apply the model to data? As differences exist between data and MC.

Backup

Waveform display: An example from an event



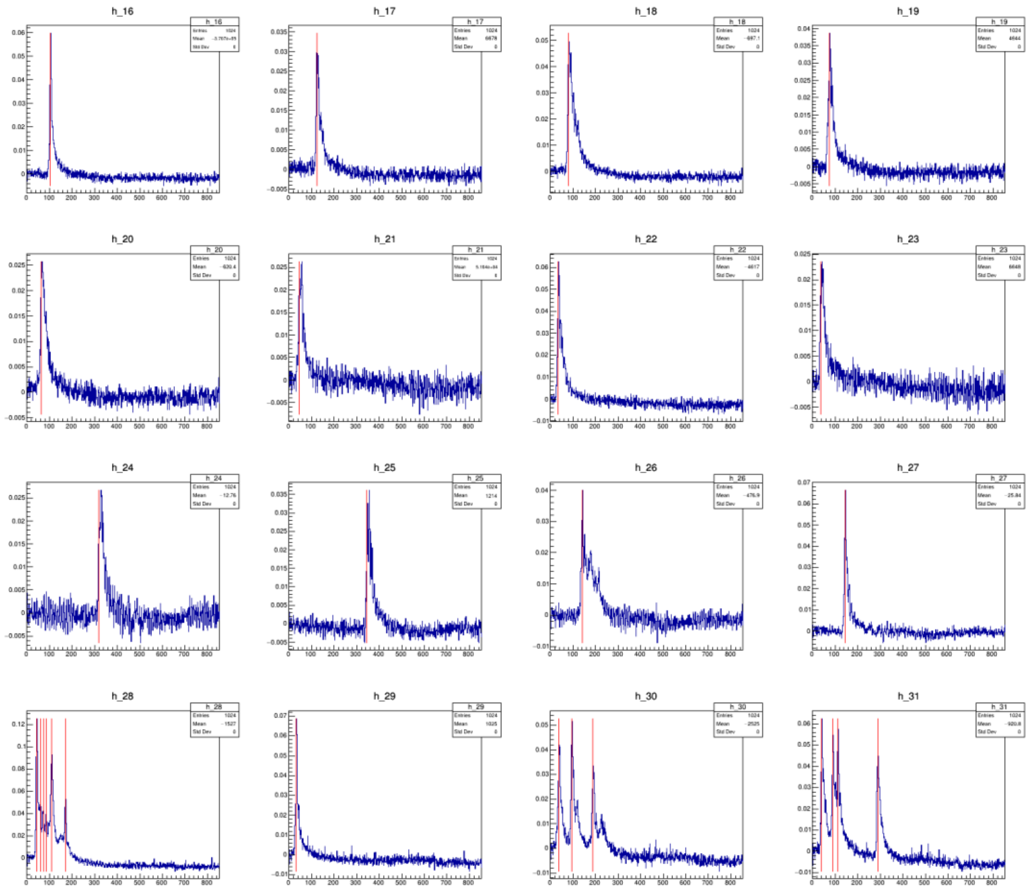
Channel ID

0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

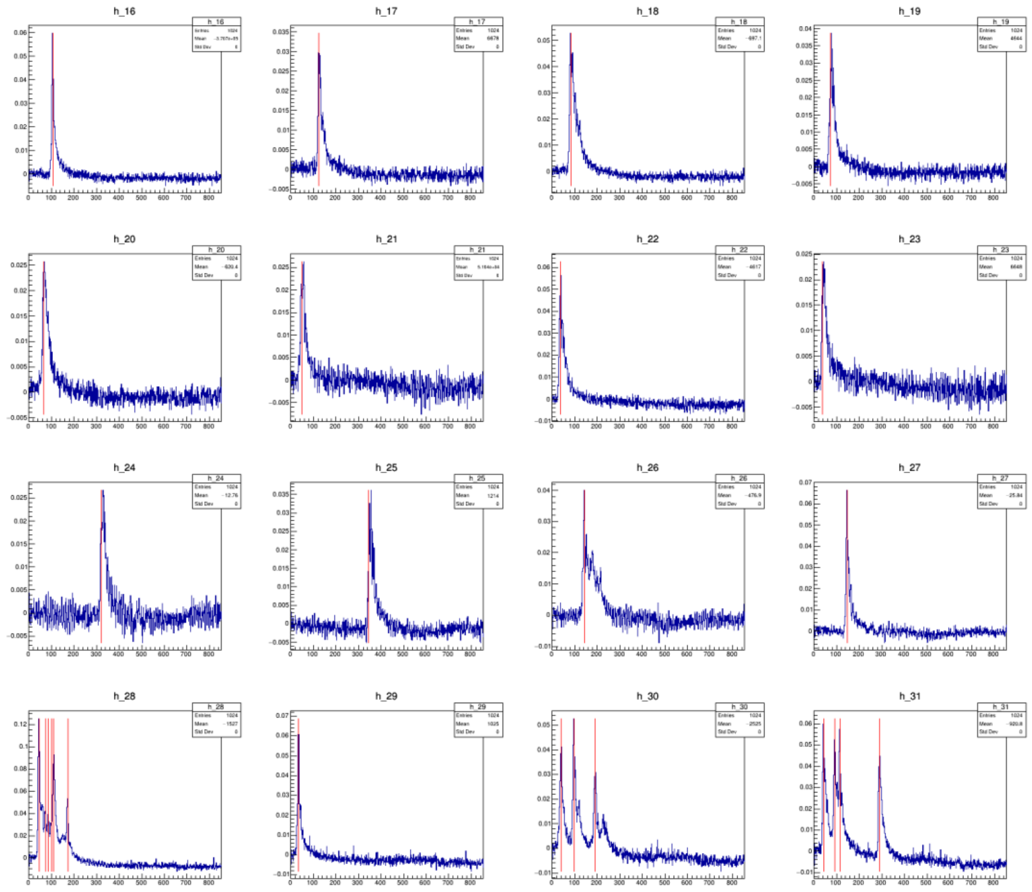
ch0-3: trigger
ch4-9: 1 cm
ch10-12: 2 cm
ch13-14: 3 cm
ch15: NULL

Waveforms

FIR

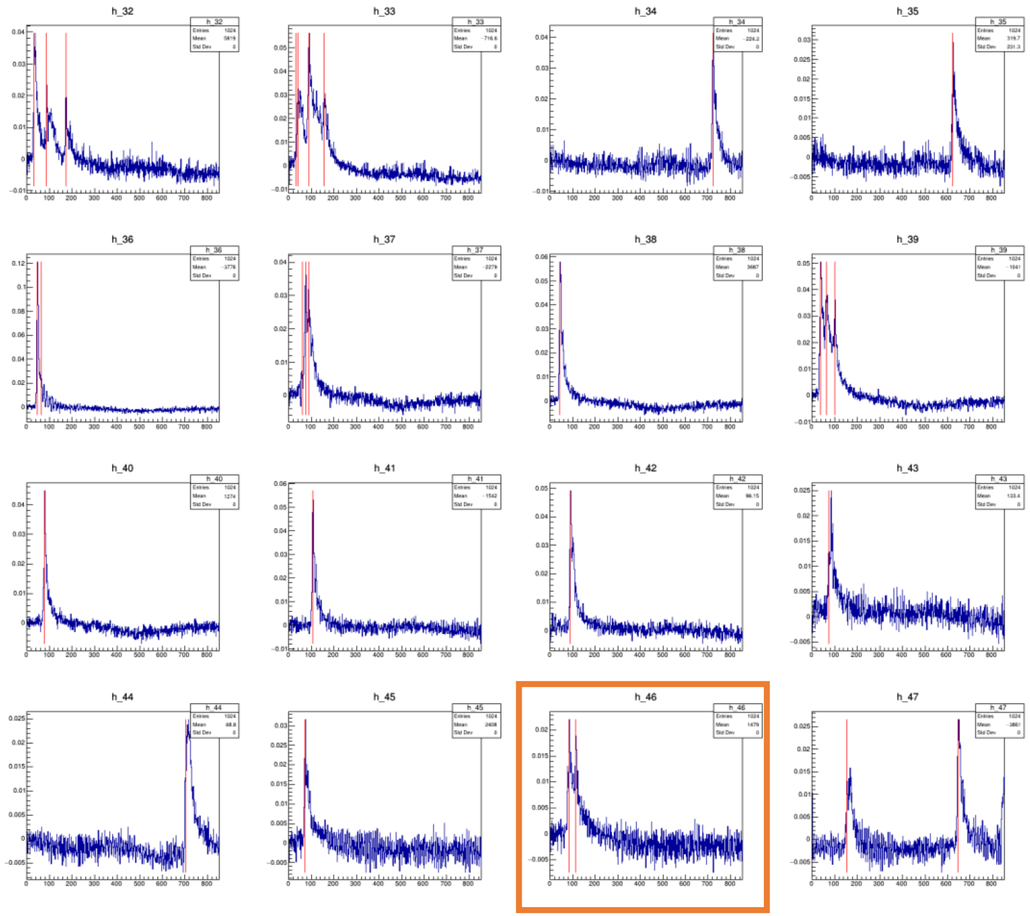


MA 8

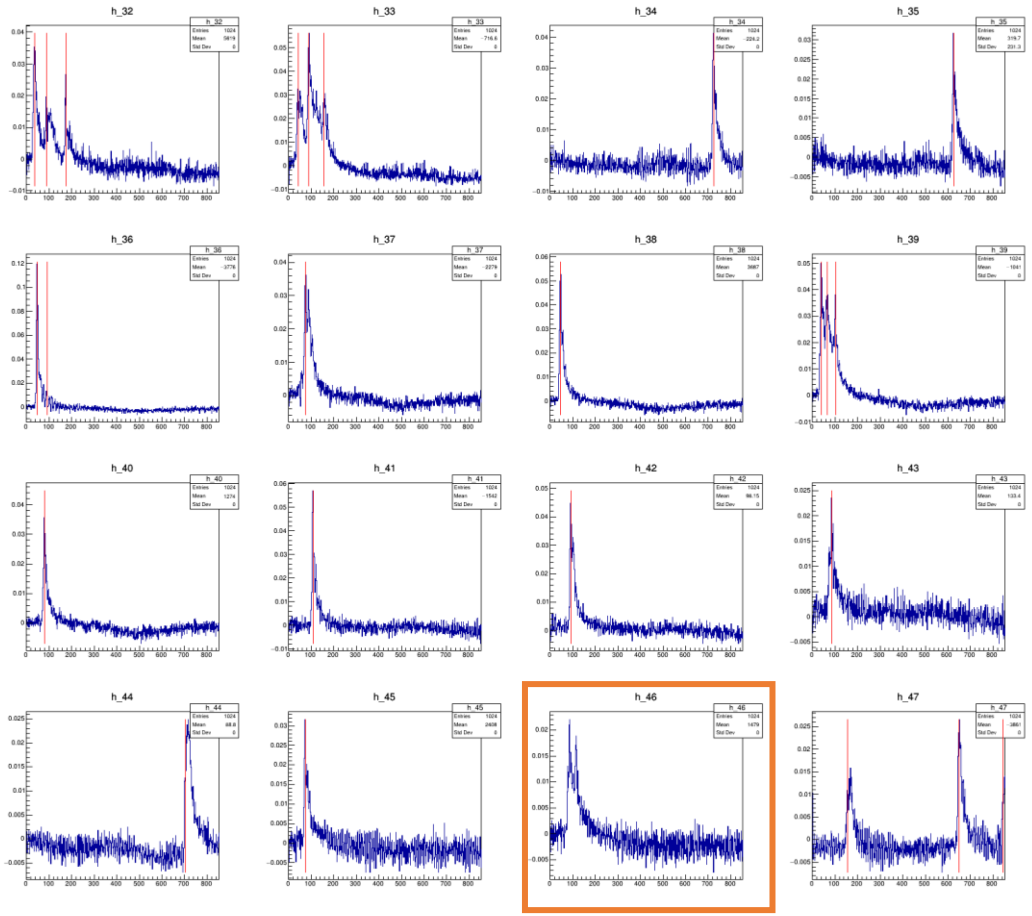


Waveforms

FIR

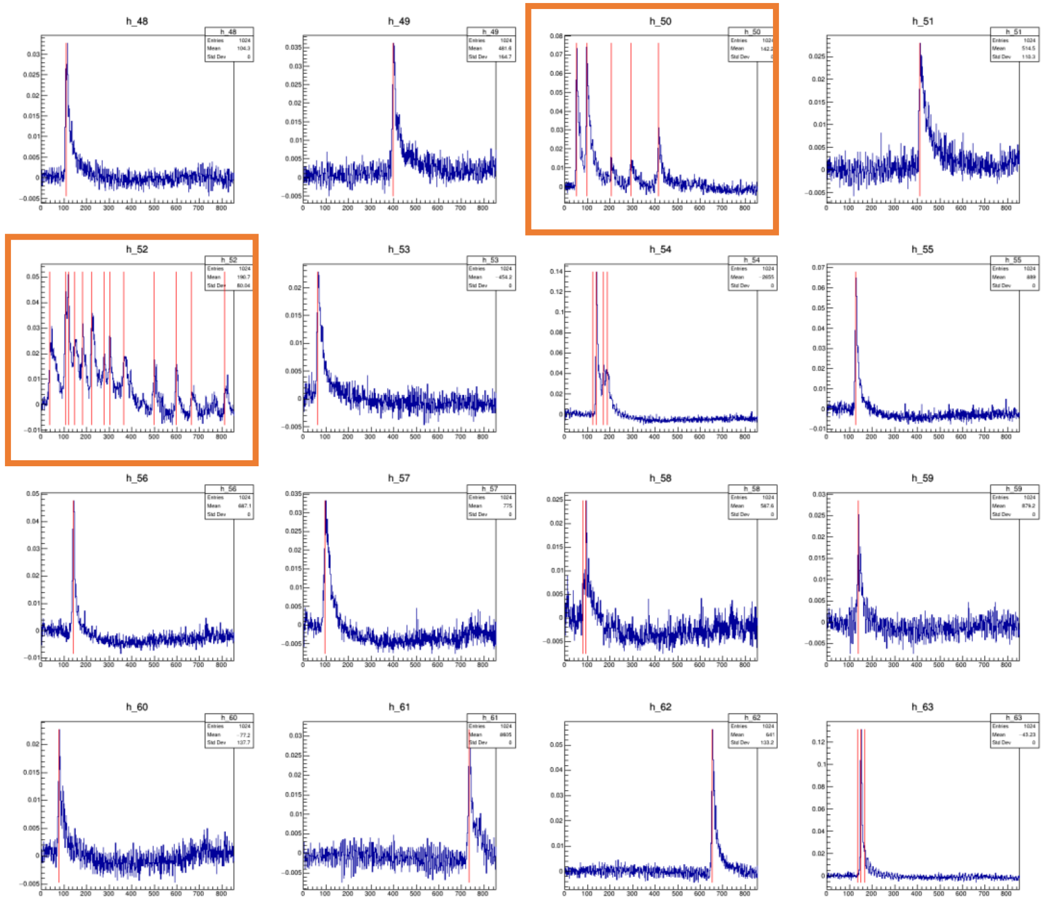


MA 8

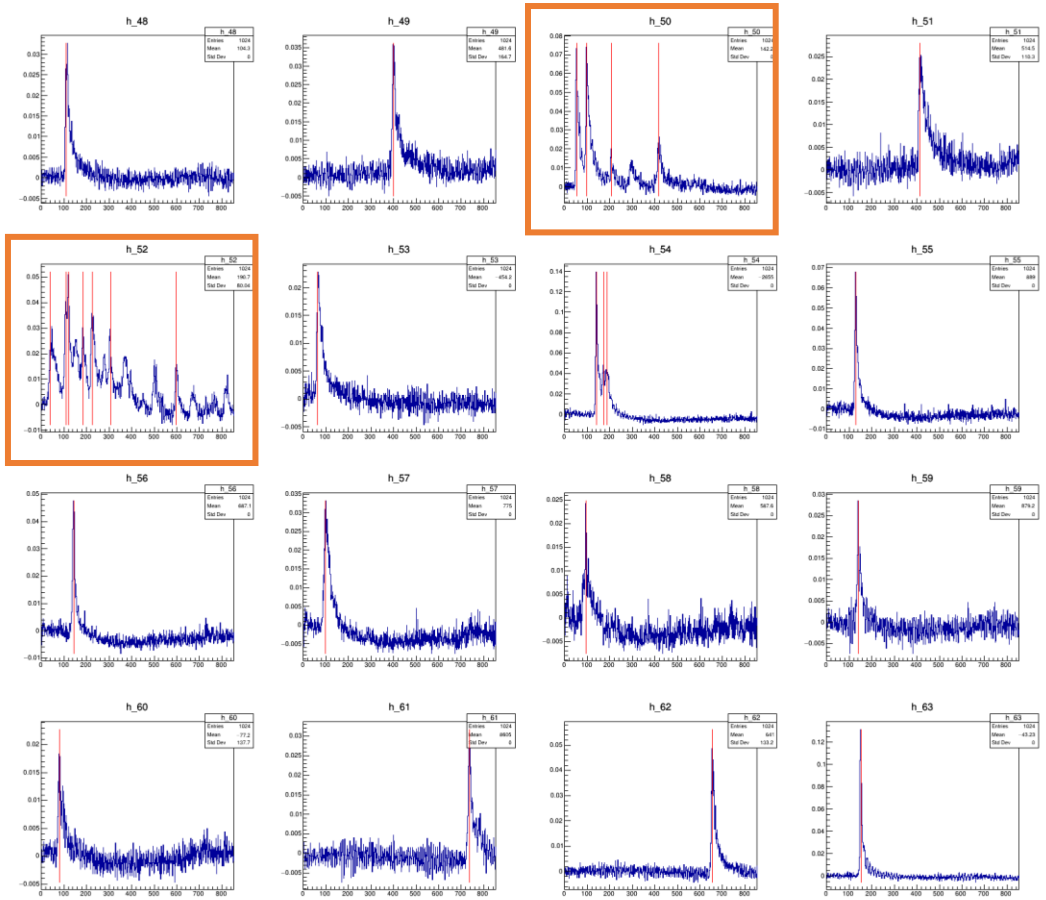


Waveforms

FIR

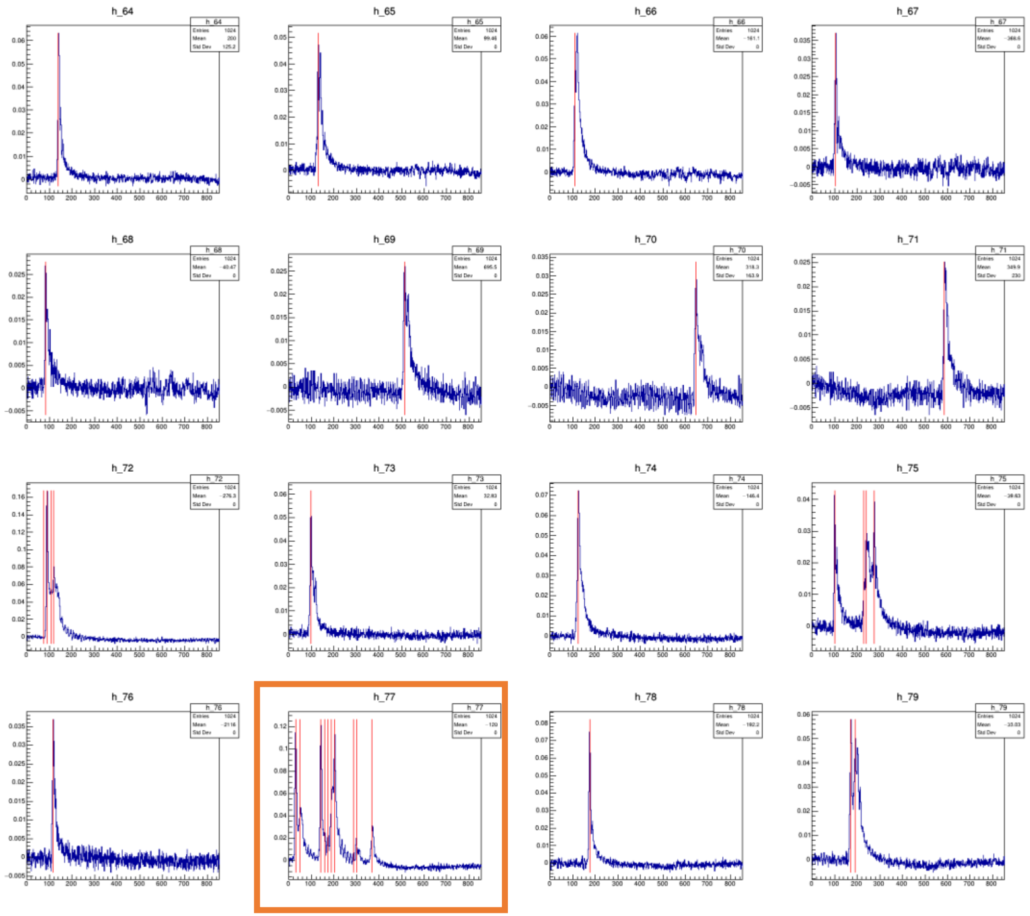


MA 8

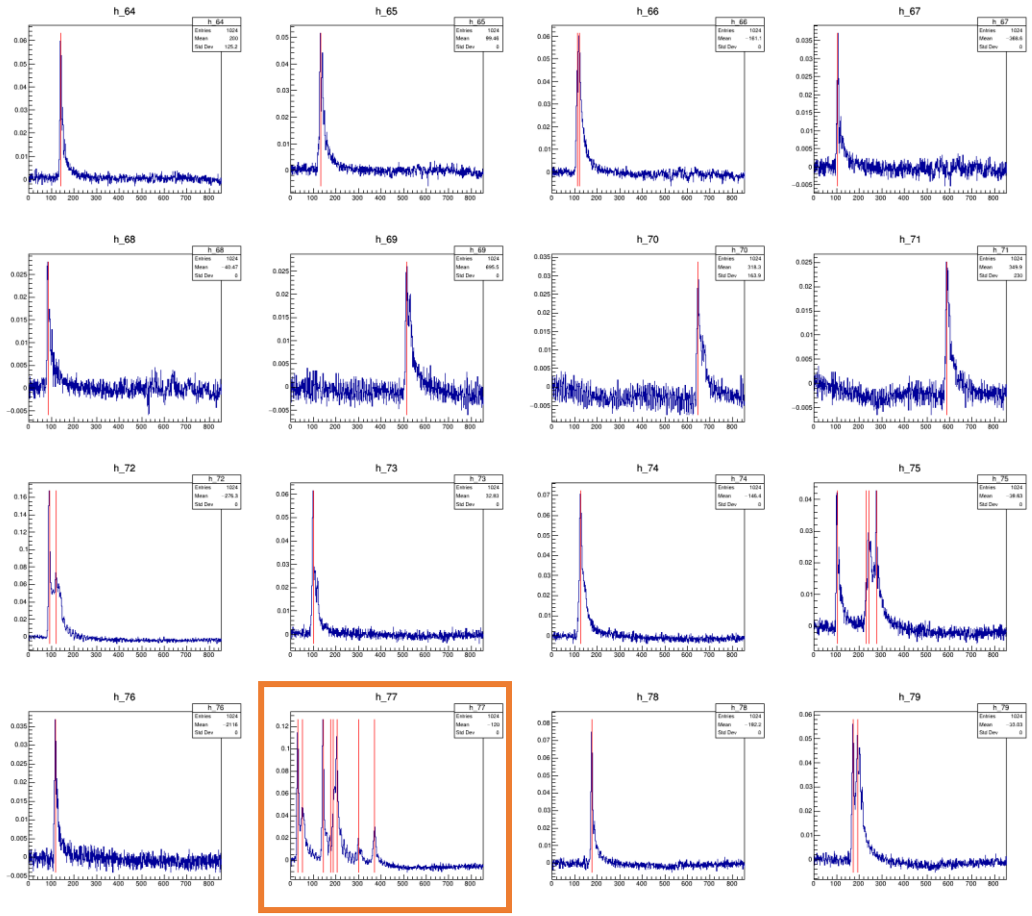


Waveforms

FIR

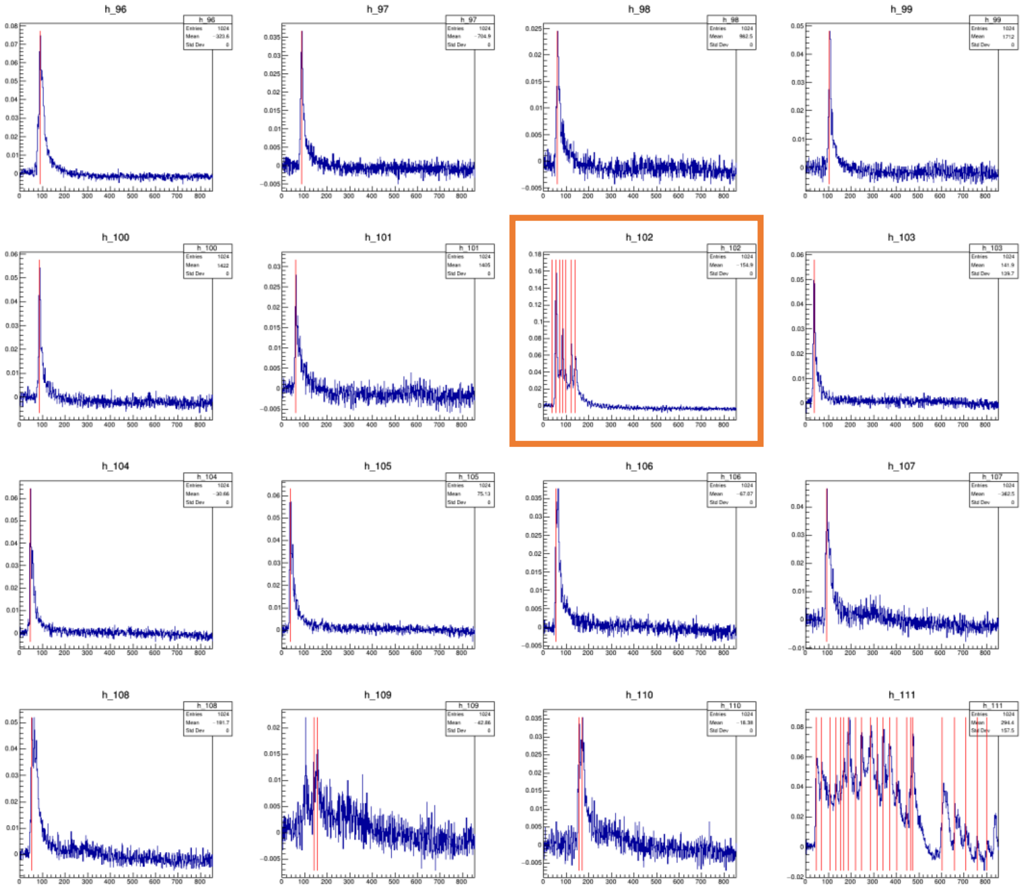


MA 8

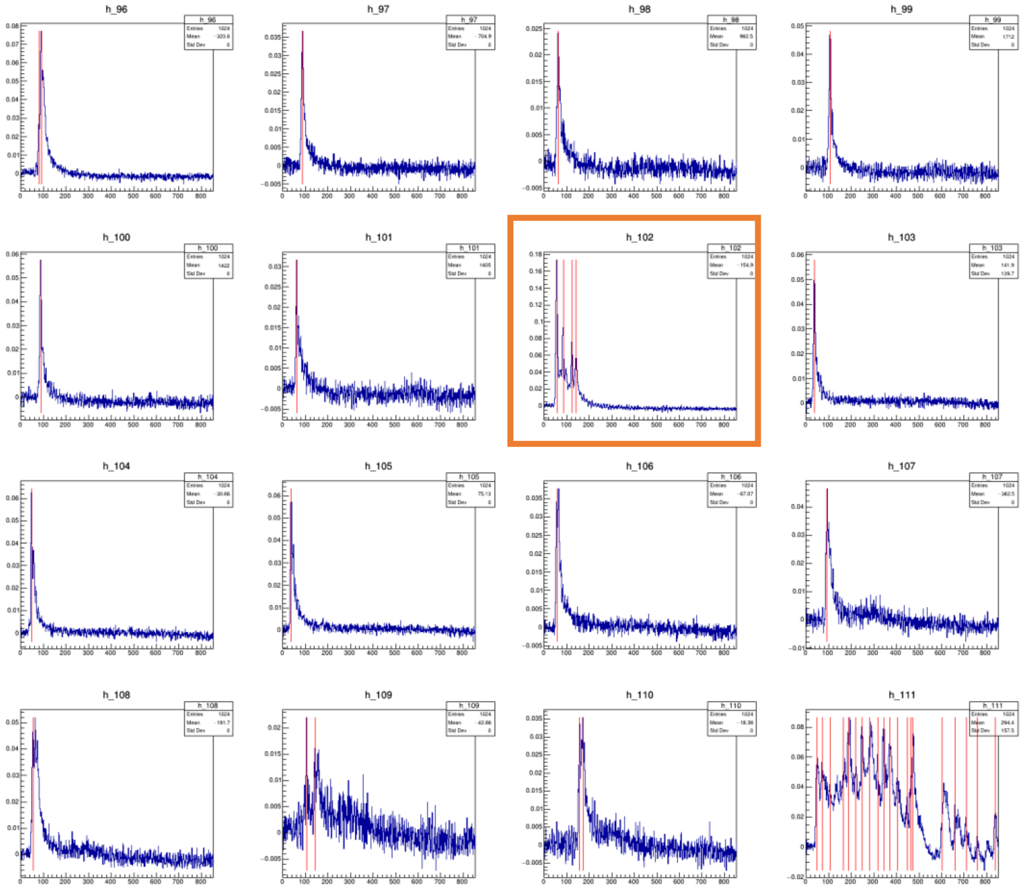


Waveforms

FIR

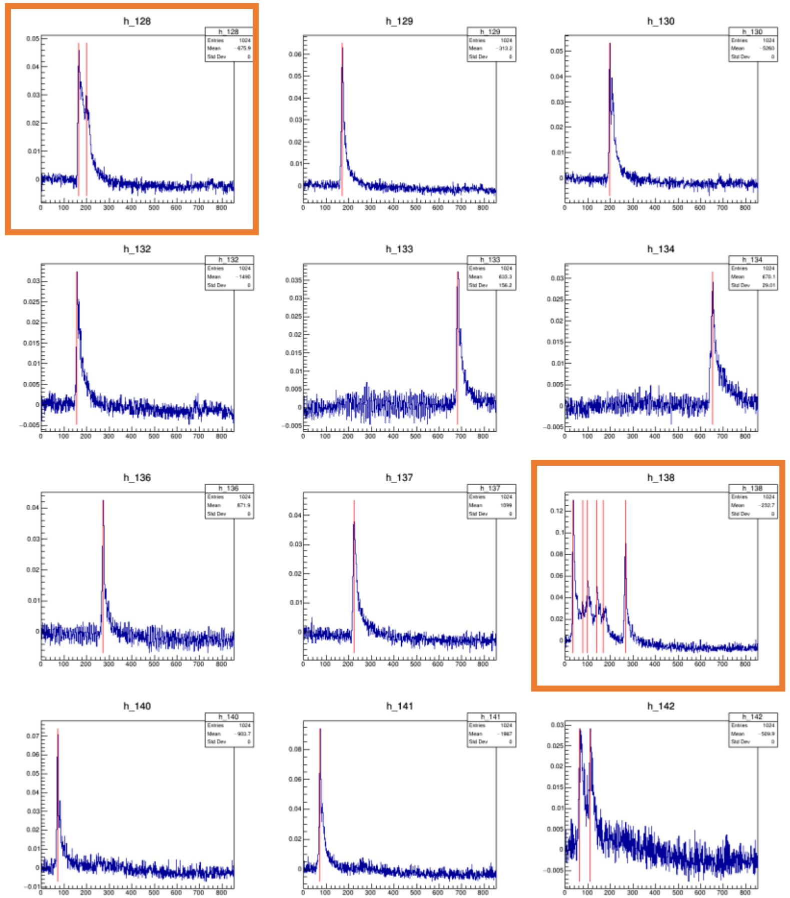


MA 8

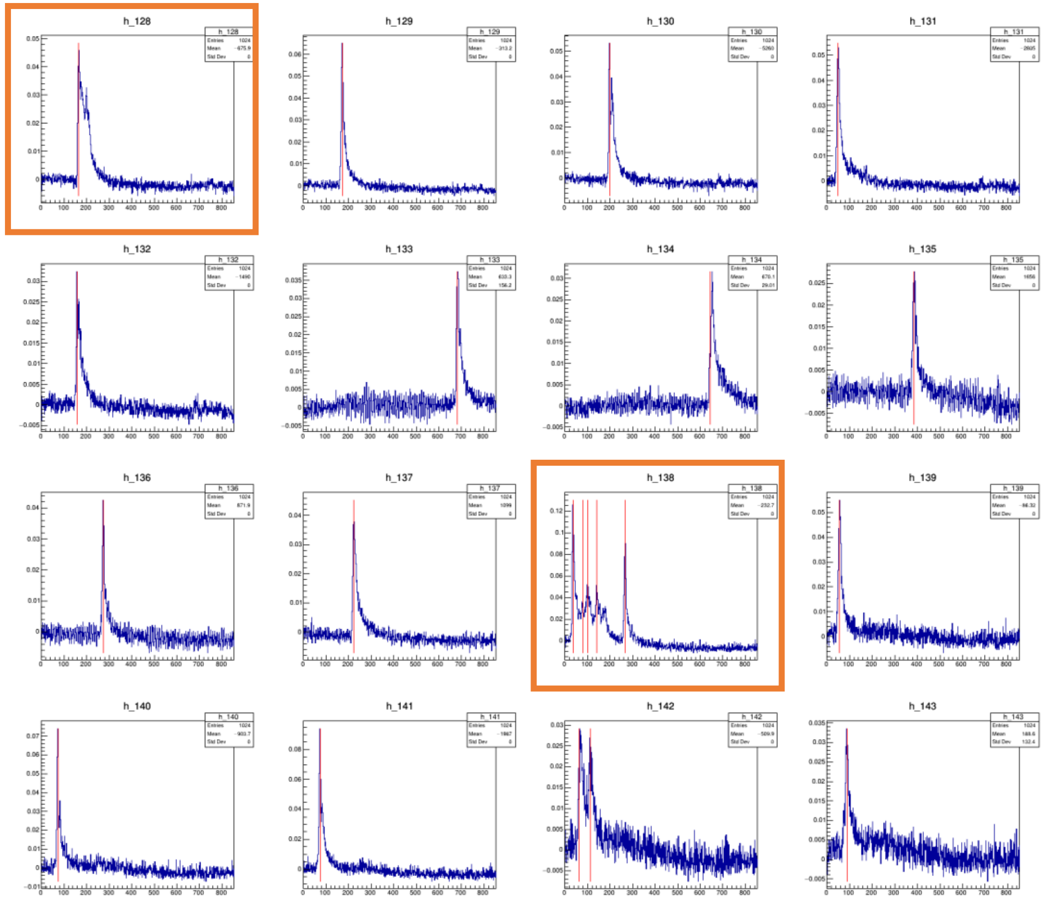


Waveforms

FIR

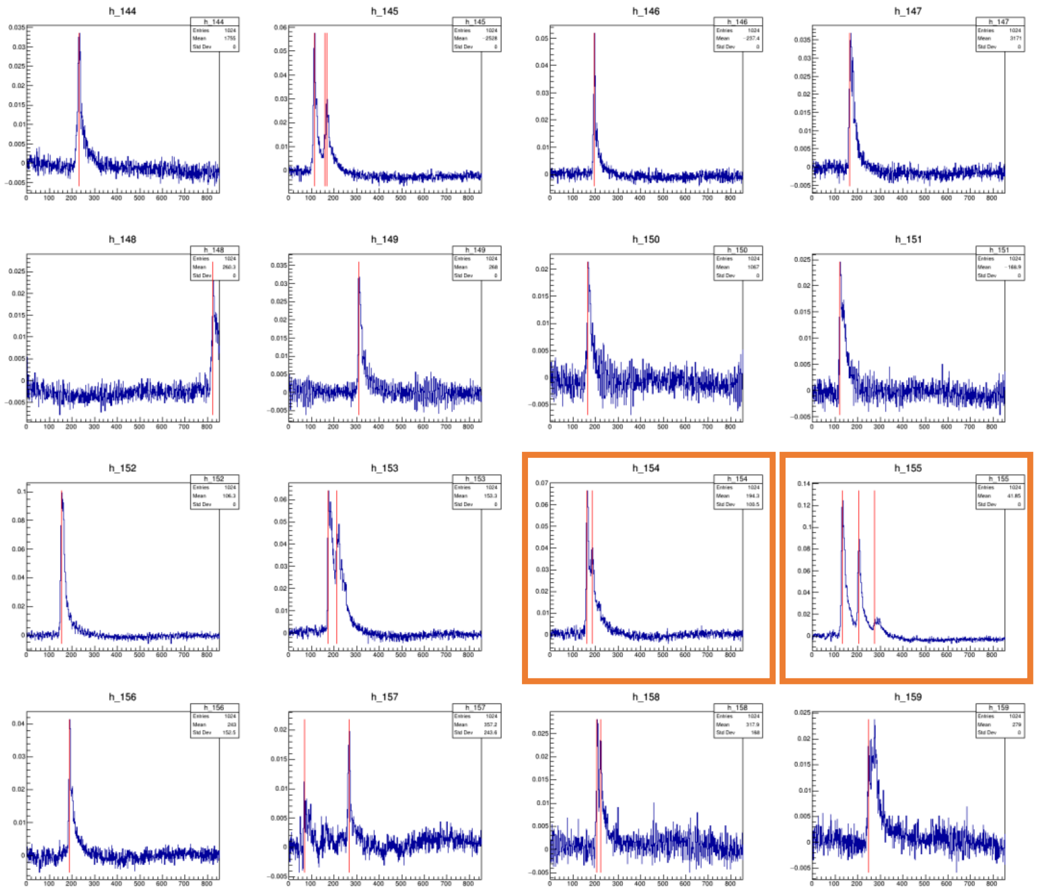


MA 8



Waveforms

FIR



MA 8