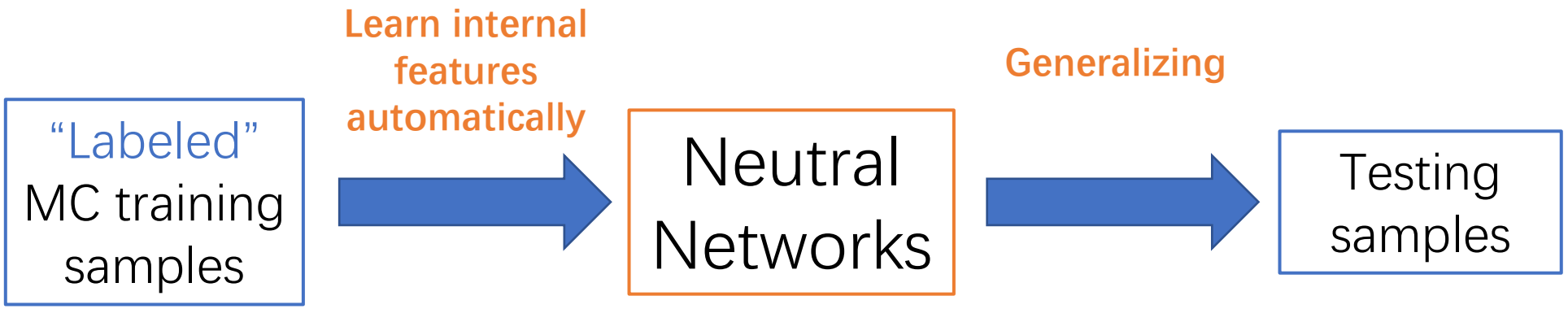
Plan of the peak finding algorithm development for beam test data

Guang Zhao

zhaog@ihep.ac.cn

Recap: NN algorithm



Price need to be paid:

A. The performance relies on the quality of the labels
B. The training sample and the testing sample should have quite similar distributions

This not an issue for MC simulations ➔ Both A and B can be satisfied
But always a big challenge for data analysis.

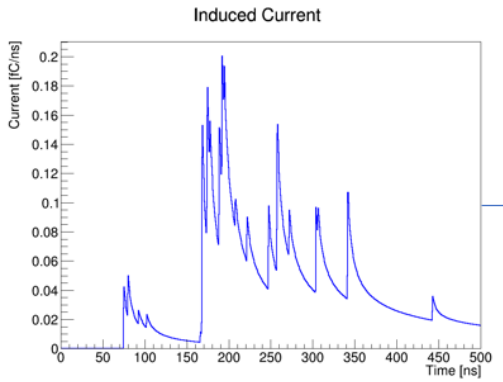
Strategy for data analysis

- **Strategy 1: Train with MC, and apply to data**
  - Advantage: Full labels for a supervised learning (A √)
  - Challenge: Require excellent data/MC consistency (B ×)
    - Need to tune the MC in several aspects:
      - Electronics responses
      - Noises
      - Space charge effect
      - …

- **Strategy 2: Train with data, and apply to data**
  - Advantage: Do NOT rely on data/MC consistency (B √)
  - Challenge: Need to add reliable labels (A ×)

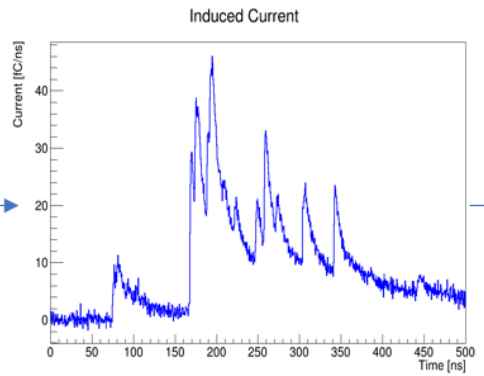
Strategy 1: Train NN with MC

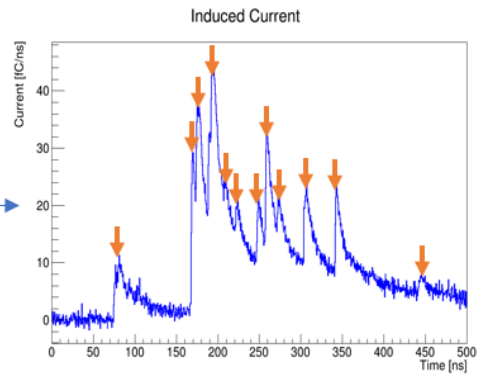
Induced current



Electronics

Realistic waveform





Signal generator (Garfield++):

- Heed: ionization process
- Magboltz: gas properties (drift/diffusion)

Electronics:

- Preamplifier
- Noises
- ADC

Peak finding algorithm:

- Second derivative

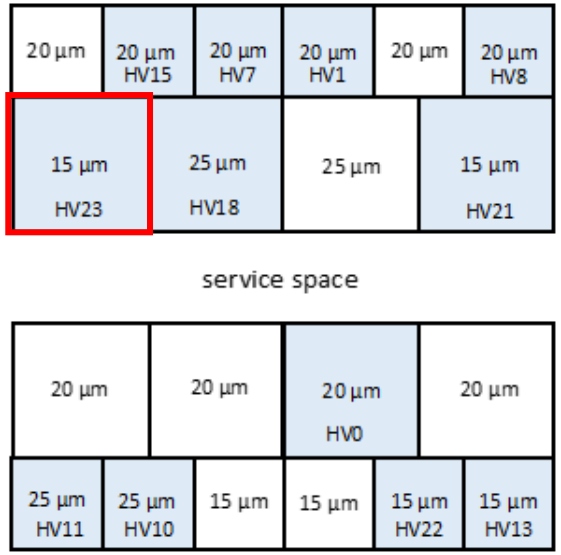
Signal generation | **Digitization** | **Reconstruction**

- **Space charge effect:** need to measure or calculate the local charge density near the wire?

- **Noises:** can be extracted by doing FFT analysis
- **Electronics responses:** require transfer function (in time or Laplace domain)
- …

Strategy 2: Train NN with data

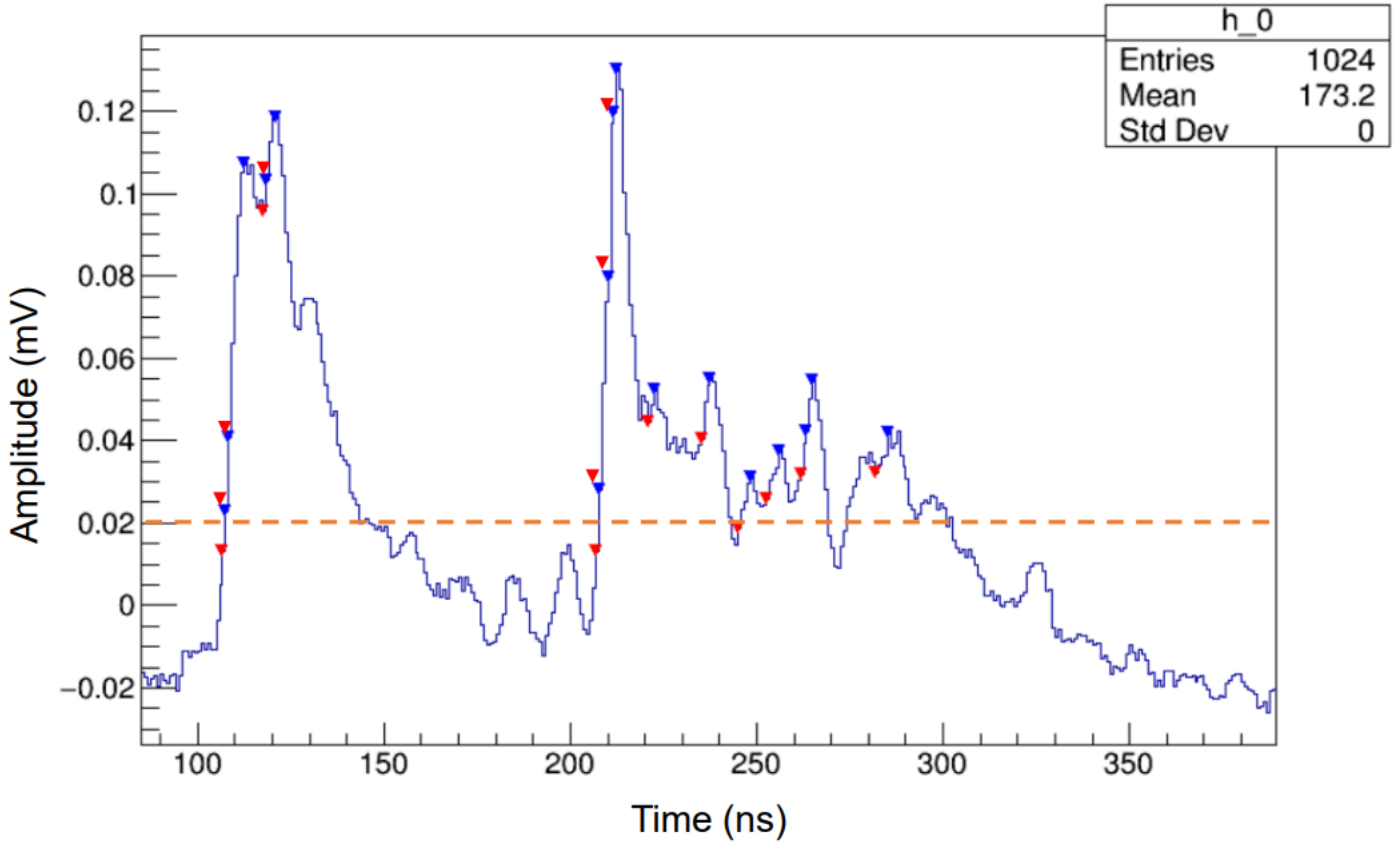
- **Test beam data file # 29, ch 11**
  - Gas mixture: He/iC4H10 = 80/20
  - Sampling rate: 1.5 GHz
  - HV tag: 10
  - Angle: 0



DRS channels	HV channels	Tubes
0	0	1.5cm-20μm
1	15	1.0cm-20μm
2	22	1.0cm-15μm
3	13	1.0cm-15μm
4	18	1.5cm-25μm
5	7	1.0cm-20μm
6	1	1.0cm-20μm
7	21	1.5cm-15μm
8	8	1.0cm-20μm
9	11	1.0cm-25μm
10	10	1.0cm-25μm
11	23	1.5cm-15μm
12		Upstream scintillator up
13		Upstream scintillator down
14		Downstream scintillator up
15		Downstream scintillator down

		HV tag	HV tag	HV tag	HV tag	HV tag	HV tag	HV tag
		5	6	7*	8	9	10	11(+10)
Tubes	HV channels	Volt (V)	Volt (V)	Volt (V)	Volt (V)	Volt (V)	Volt (V)	Volt (V)
1cm-15mum	13.22	1500	1580	1580	1580	1580	1580	1590
1cm-20mum	1,7,8,15	1550	1620	1620	1680	1670	1670	1680
1cm-25mum	10,11	1600	1690	1690	1750	1740	1740	1750
1.5cm-15mum	21.23	1650	1650	1650	1650	1830	1830	1840
1.5cm-20mum	0	1700	1760	1760	1780	1750	1770	1780
1.5cm-25mum	18	1830	1830	1830	1830	1830	1830	1840

Pre-selection



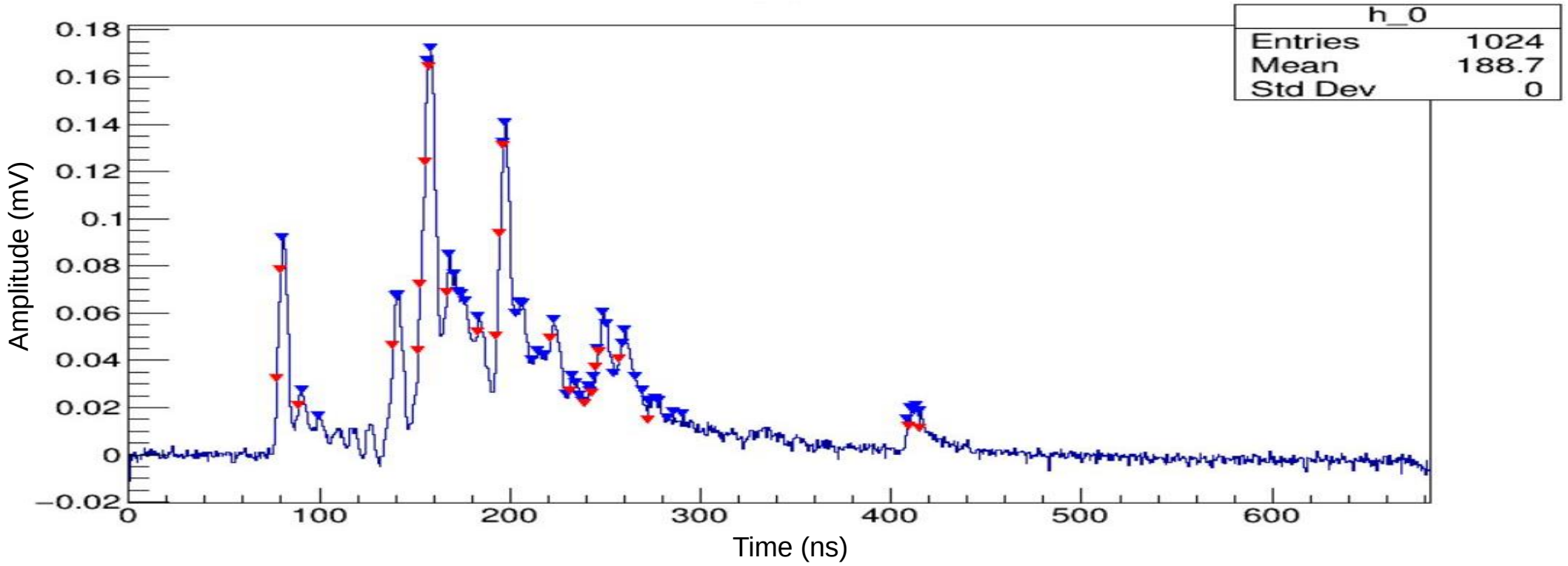
Rising-edge amplitude calculation:

1. Find all the local maximum points of the waveform (>0.015mV)
2. For each local maximum points, find the nearest early point, whose slope starts to decrease (inflection point)
3. The Amplitude distance between the local maximum and the inflection points is recorded as the rising edge.

The points marked in the figure are local maximum points and inflection points

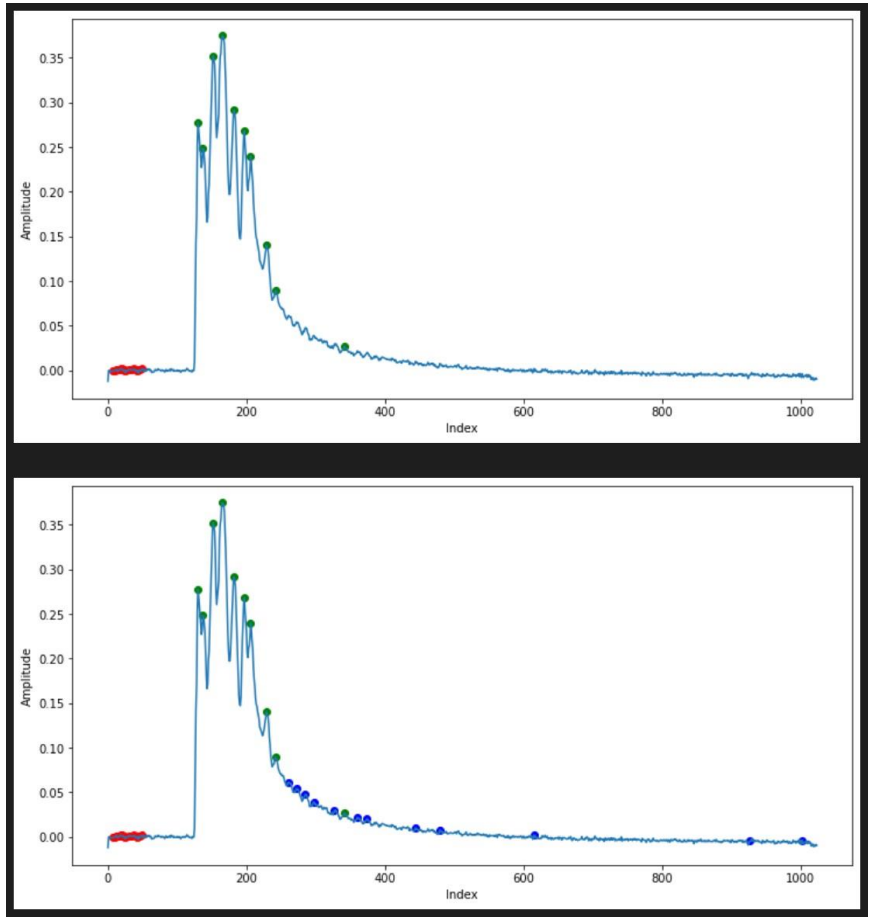
Pre-selection

- **For better demonstrate the algorithm:**
  - **Require at least 6 rising edges whose amplitudes are within 0.01 and 0.03 mV**



Labelling the data

Intuition: Using derivative algorithm to guide the labelling



<u>Tight signal:</u> D1 threshold > 6E-2

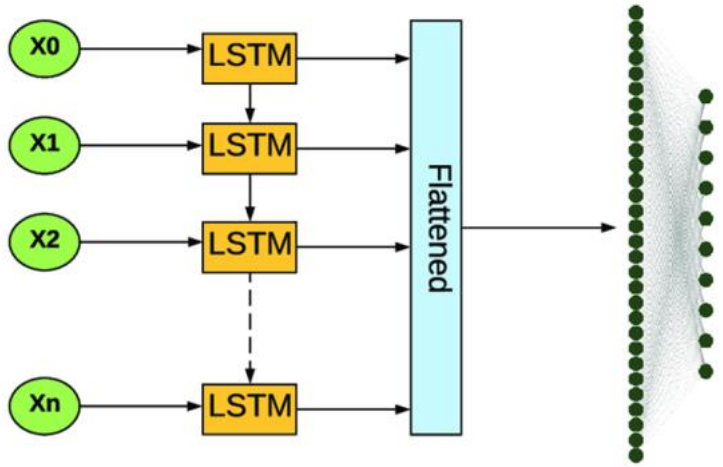
<u>Loose signal:</u> D1 threshold > 4E-2

<u>Labelling with large confidence:</u>

- **Signal: Tight signal**
- **Background: COMPLEMENTARY of loose signal**

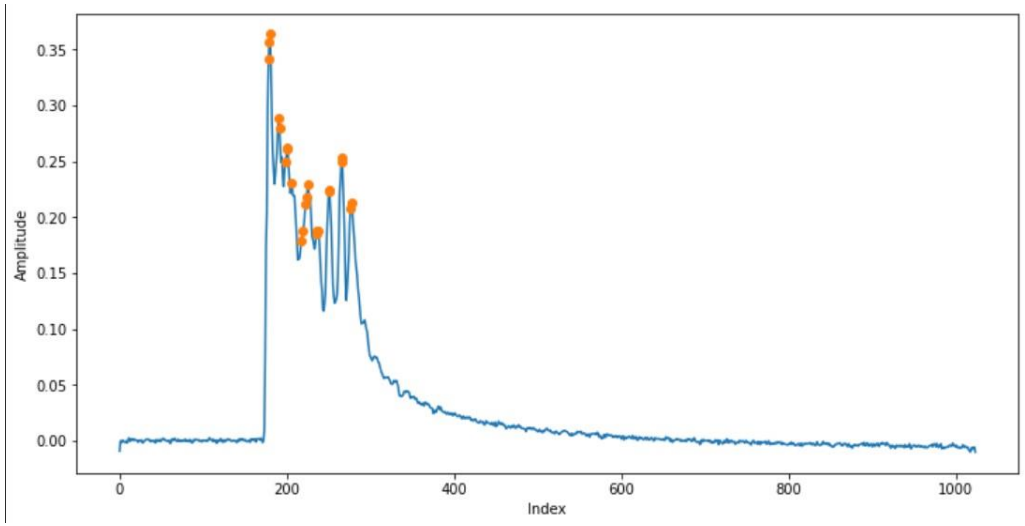
Model and training

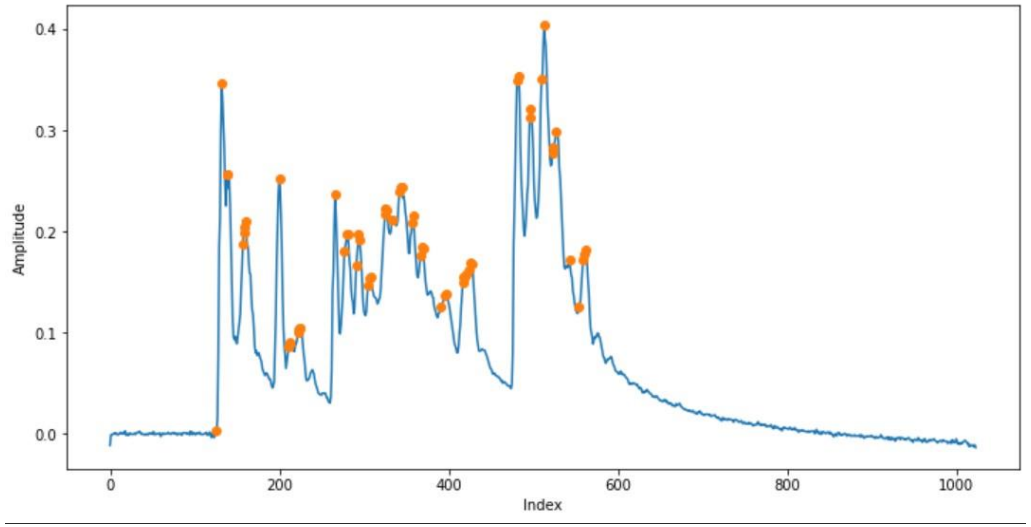
- **NN model:**
  - LSTM layers + full-connected layer + sigmoid output



- **Training:**
  - First 800 waveforms in file # 29, channel 11
  - After 20 epochs: 98.4 % accuracy

Testing





- Very preliminary results by apply the NN to waveforms after index 800 in the same file
- Need more investigations

Conclusion

- **Strategy 1: Train NN with MC**
  - Need to tune the MC ➔ input from data
- **Strategy 2: Train NN with data**
  - Find a way to labelling the data
  - Investigate more possibilities with the NN models
- **Should work in both directions**