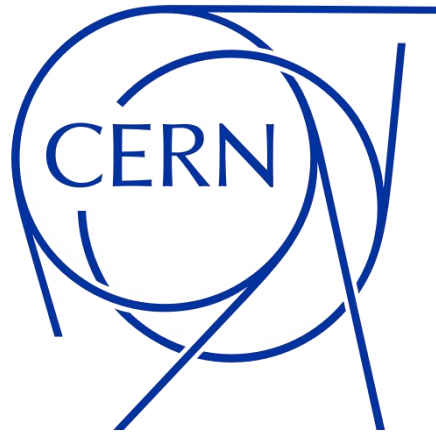




On-Chip Machine Learning for Dual-Readout Calorimeters

Liangyu Wu, Kenny Jia, Julia Gonski

SLAC

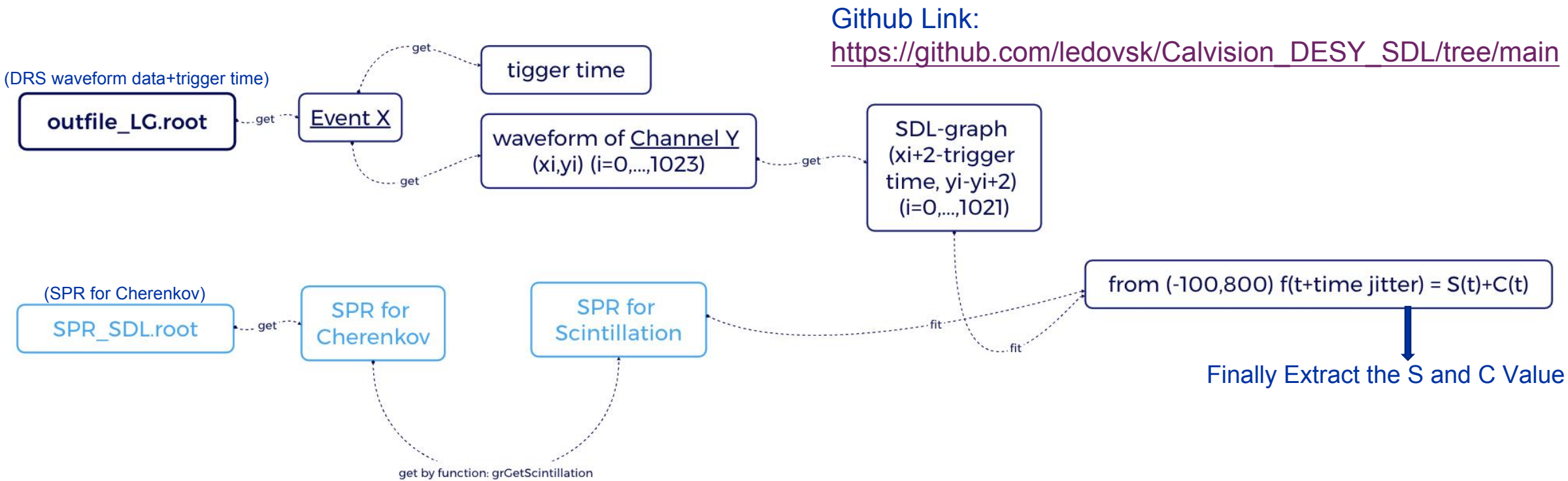
2025/02/03



NATIONAL
ACCELERATOR
LABORATORY

Overview of the Dataset

1. Workflow of the SDL (Single Delay Line) Method



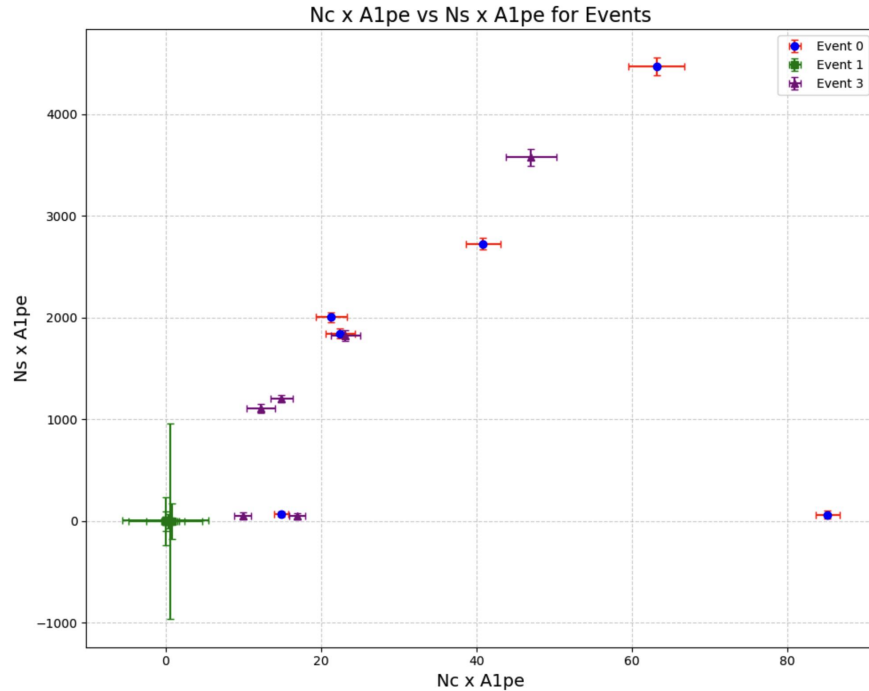
➤ Our Preliminary Dataset for ML:

	Sample 1, Sample 2, ..., Sample 1024	C	S
Event1	***	***	***
...	***	***	***
Event N	***	***	***

Overview of the Dataset

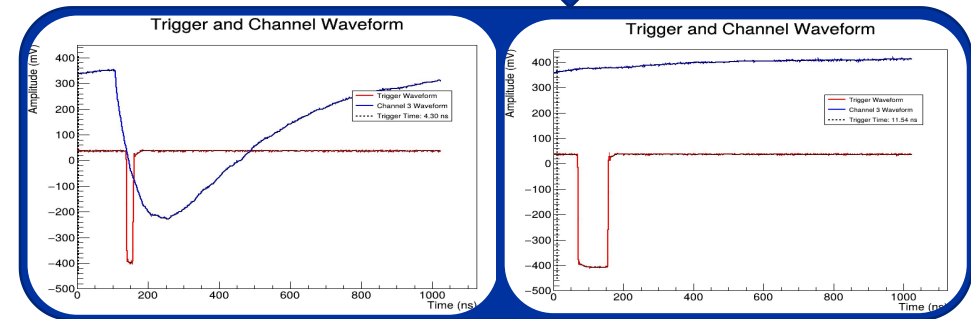
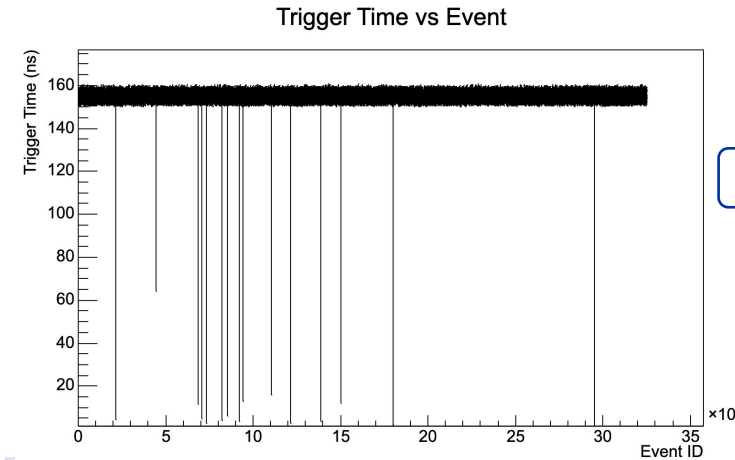
2. Remove some outlier data

➤ Some plots to get a sense:



Note:

```
* Run = 200
* Channels 0, 1, 2, and 3 are in the front
* Channels 4 and 7 are in the rear
* Channels 5 and 6 are discarded from the analysis (dead)
```



➤ Some low-signal events and events with abnormal trigger times need to be noted.

Overview of the Dataset

3. The Total Dataset Size

$$29799 = 32501 - 15 - 2687$$

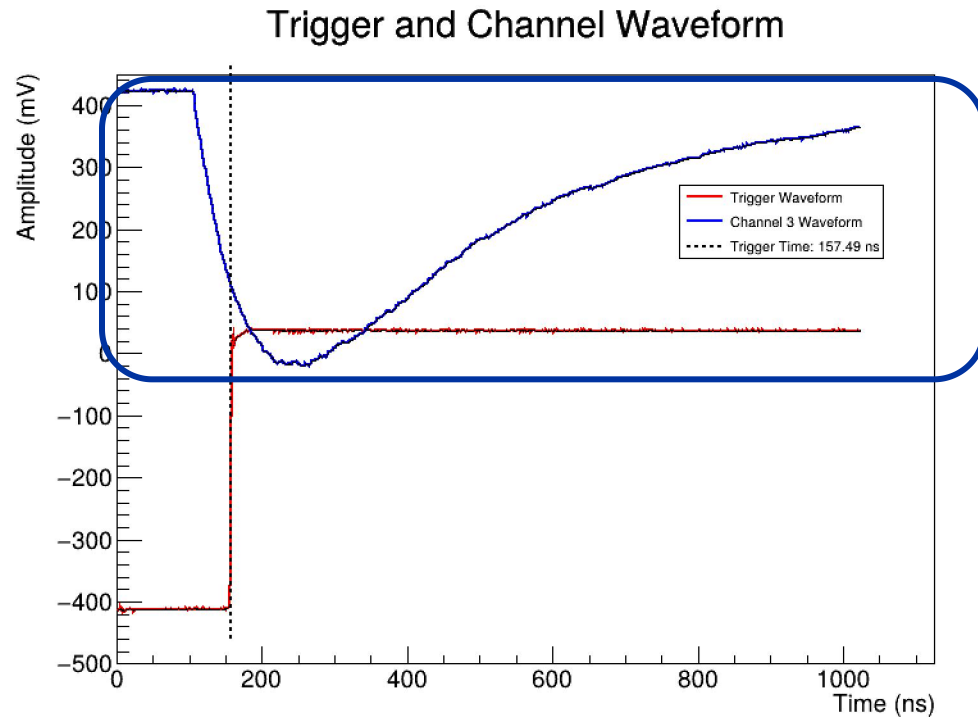
Number of events with abnormal signal triggers

Total number of events in channel 3 of run 200 data

Number of events with almost no signal, determined by the condition: $C + S < 2$

Overview of the Dataset

4. Reducing the Size of Input Data



1024 data points
sampled at 1ns intervals



52 data points
sampled at 20ns intervals

- This can reduce the number of parameters in the ML model by approximately 92%.
- The Final Dataset for ML:

	Sample 1, Sample 2, ..., Sample 52	C	S
Event1	***	***	***
...	***	***	***
Event 29799	***	***	***

Machine Learning Model

1. Basic Information About the Model

➤ Model Architecture:

- ✓ Input: A vector of 52 waveform data points.
- ✓ Hidden Layers:
Fully connected Dense layers with ReLU activation.
Neurons: 32+16+8
- ✓ Output: 2 neurons(representing the predicted values of S and C), with ReLU activation.

➤ Training Process:

- ✓ Dataset Split: 70% for training + 20% for validation + 10% for test.
- ✓ Loss Function: Mean Squared Error
- ✓ Optimizer: Adam

➤ Data Preprocessing:

- ✓ StandardScaler() was applied to the input data
- ✓ MinMaxScaler() was applied to the output data (C&S separately) to ensure that both are scaled to the same range.

➤ Current Params:

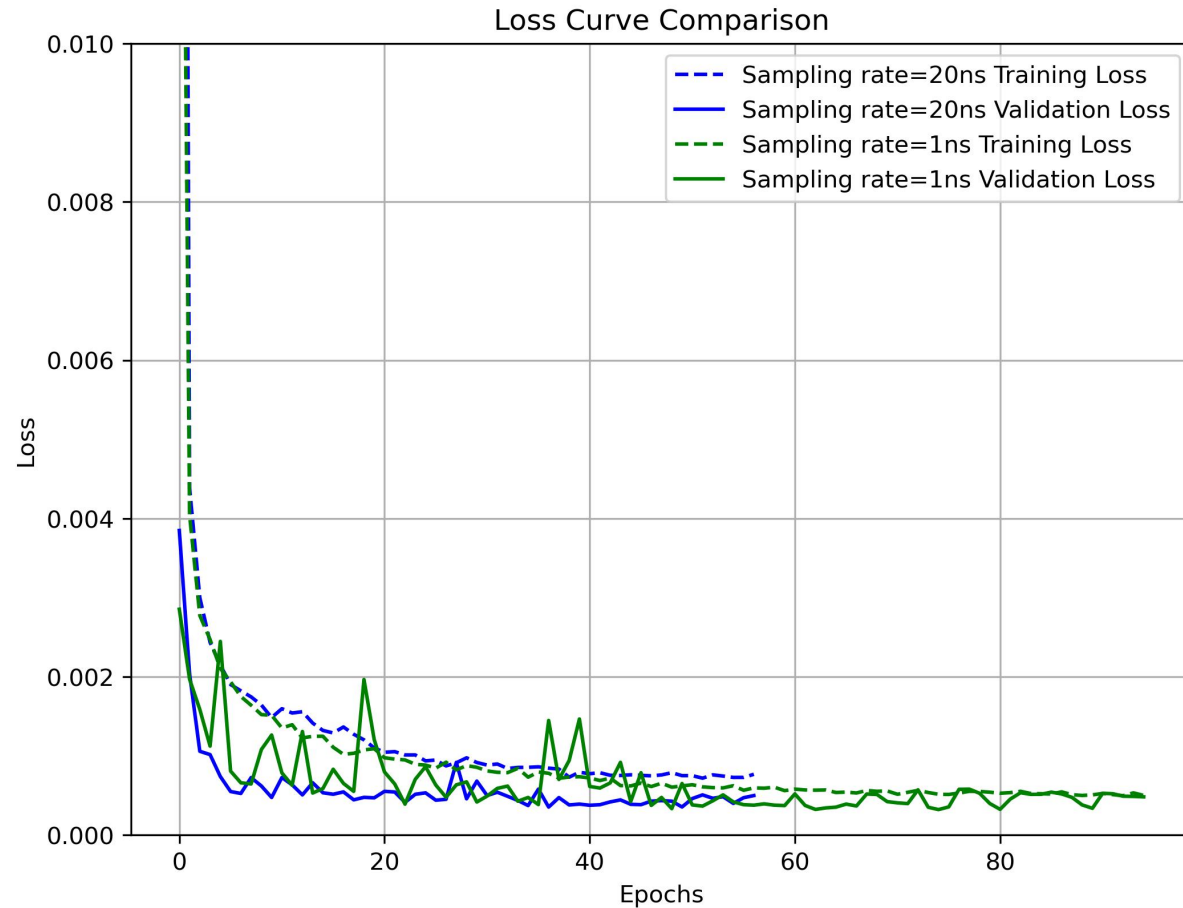
Total params: 2,602 (10.16 KB)

Trainable params: 2,490 (9.73 KB)

Non-trainable params: 112 (448.00 B)

Machine Learning Model

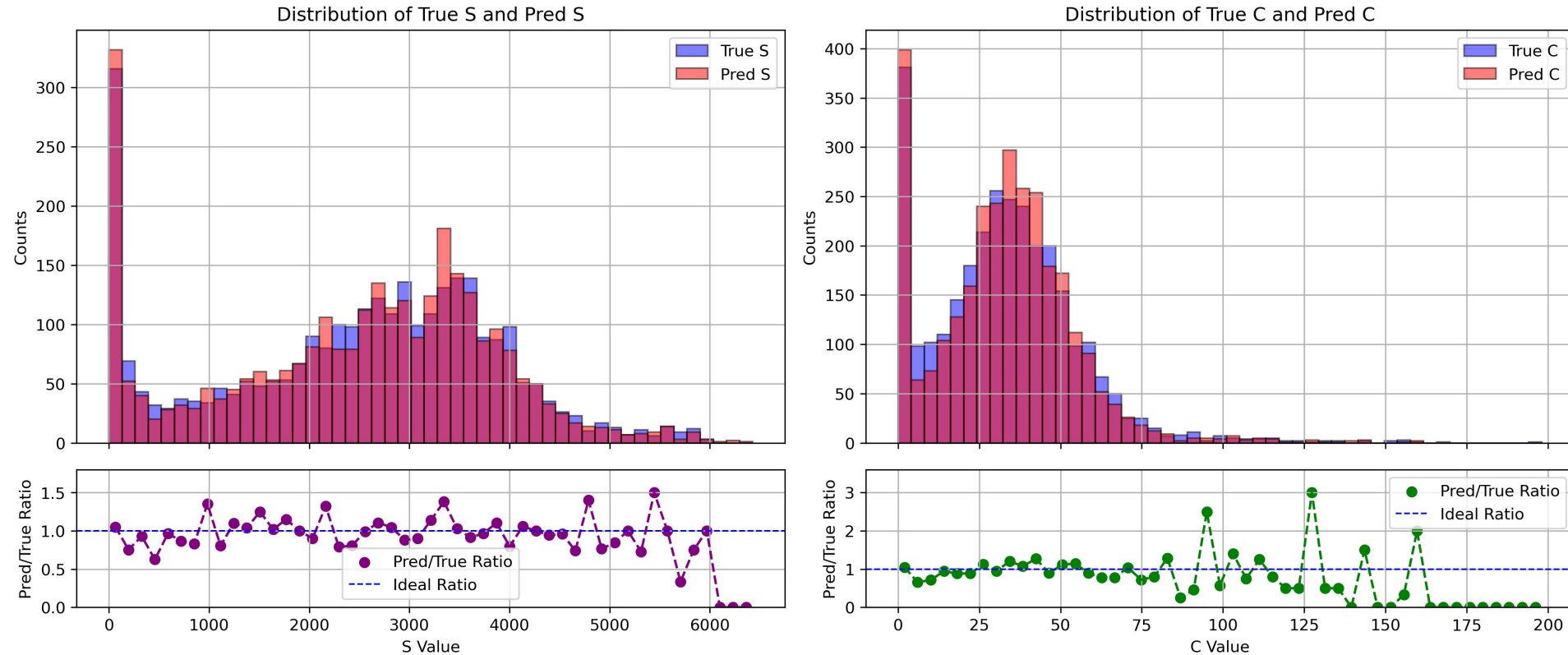
2. Inspection and Comparison of Loss Curves



- The loss curve shows rapid convergence, as validation loss closely follows training loss.
- The model generalizes well, with only slight fluctuations in validation loss.
- The loss curves before and after input data size compression both converge to stable values, and these stable values are nearly identical.

Machine Learning Model

3. Distribution of Scintillation&Cherenkov value (using 20ns dataset)



- The distributions of True and Pred value are very similar.
- The peak at low values suggests some weak-signal events remain or are inherently present, which needs to be further checked.

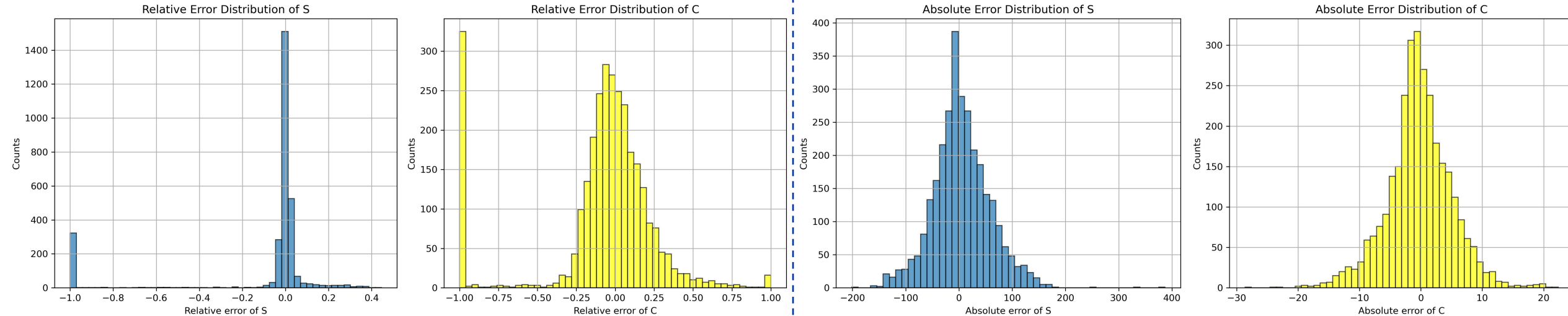
Machine Learning Model

4. Distribution of the Relative&Absolute Error

Note:

Relative Error = $\text{np.where}(\text{true} \neq 0, (\text{pred} - \text{true}) / \text{true}, 0)$

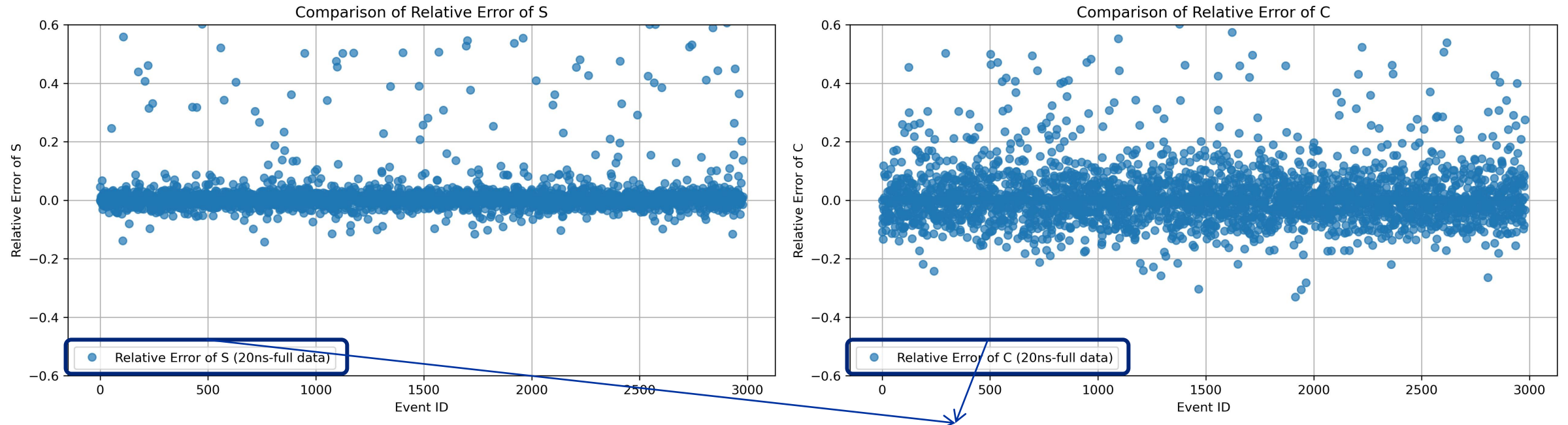
Absolute Error = $\text{pred} - \text{true}$



- Left: The distributions show that most predictions are centered around zero, indicating good overall accuracy. A few events have a relative error exceeding 100%, with the number of such events being 0 for S (0.00%), 16 for C (0.44%).
- Right: The distribution of absolute errors looks very ideal though.

Machine Learning Model

5. Relative error comparisons

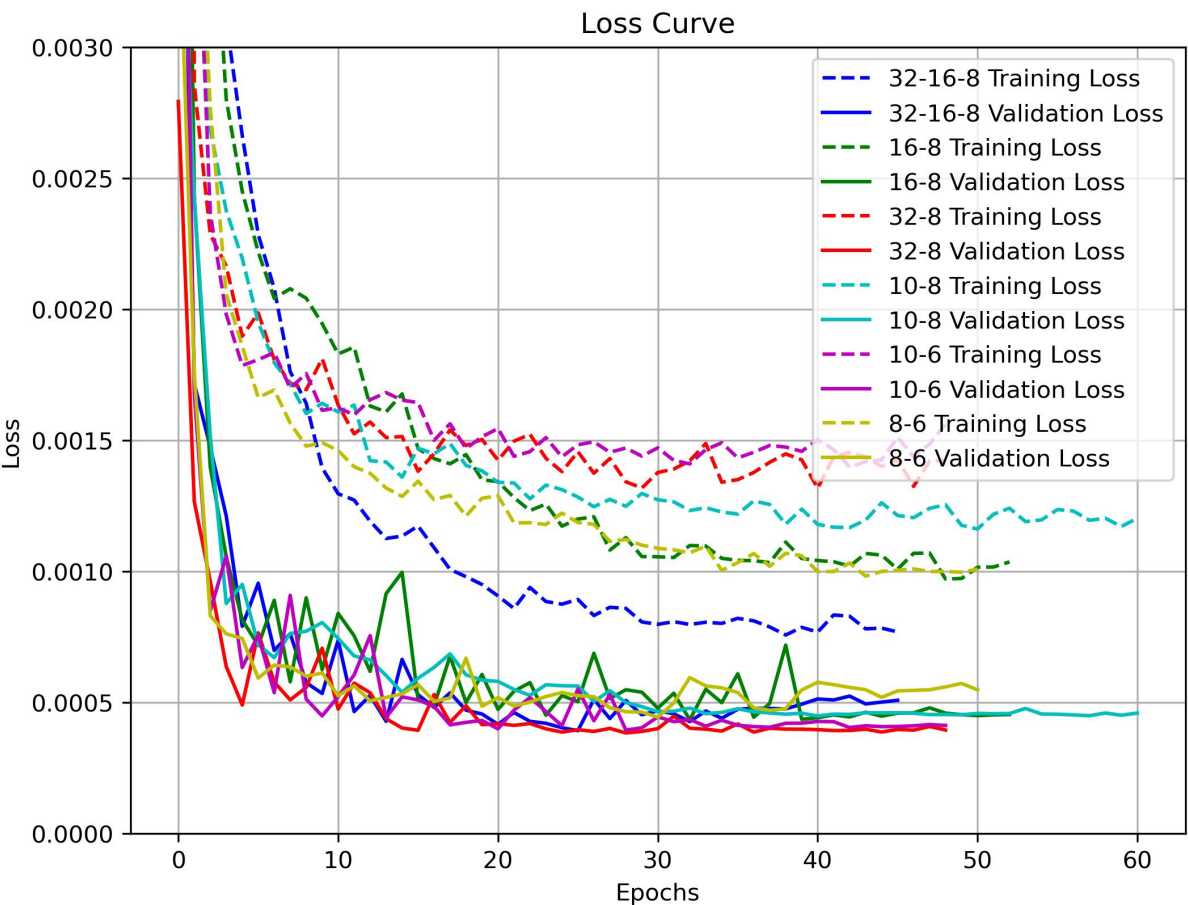


The relative error difference between models trained by 20ns and 1ns dataset

- From the comparison of the two plots, it is evident that compressing the input data size has a greater impact on the prediction accuracy of C than on S. (But note that the absolute value of S is normally larger than C.)

Exploration of Different Model Architectures

1. Loss Curves for Different Hidden Layer Structures

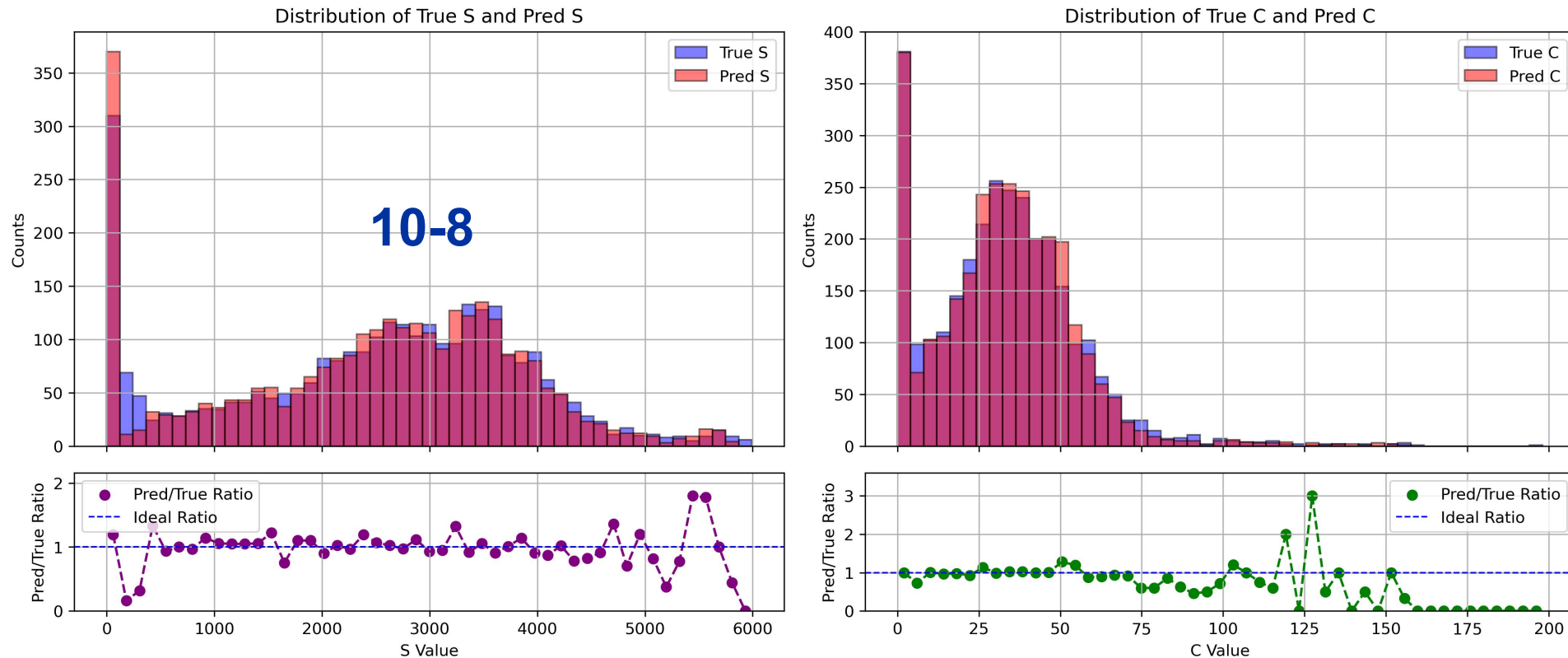


Structure	Total Params	Trainable Params	Non-trainable Params
32-16-8	2602	2490	112
32-8	2138	2058	80
16-8	1098	1050	48
10-8	708	672	36
10-6	674	642	32
8-6	548	520	28

➤ The differences in the loss curve when using hidden layers with different structures do not seem significant.

Exploration of Different Model Architectures

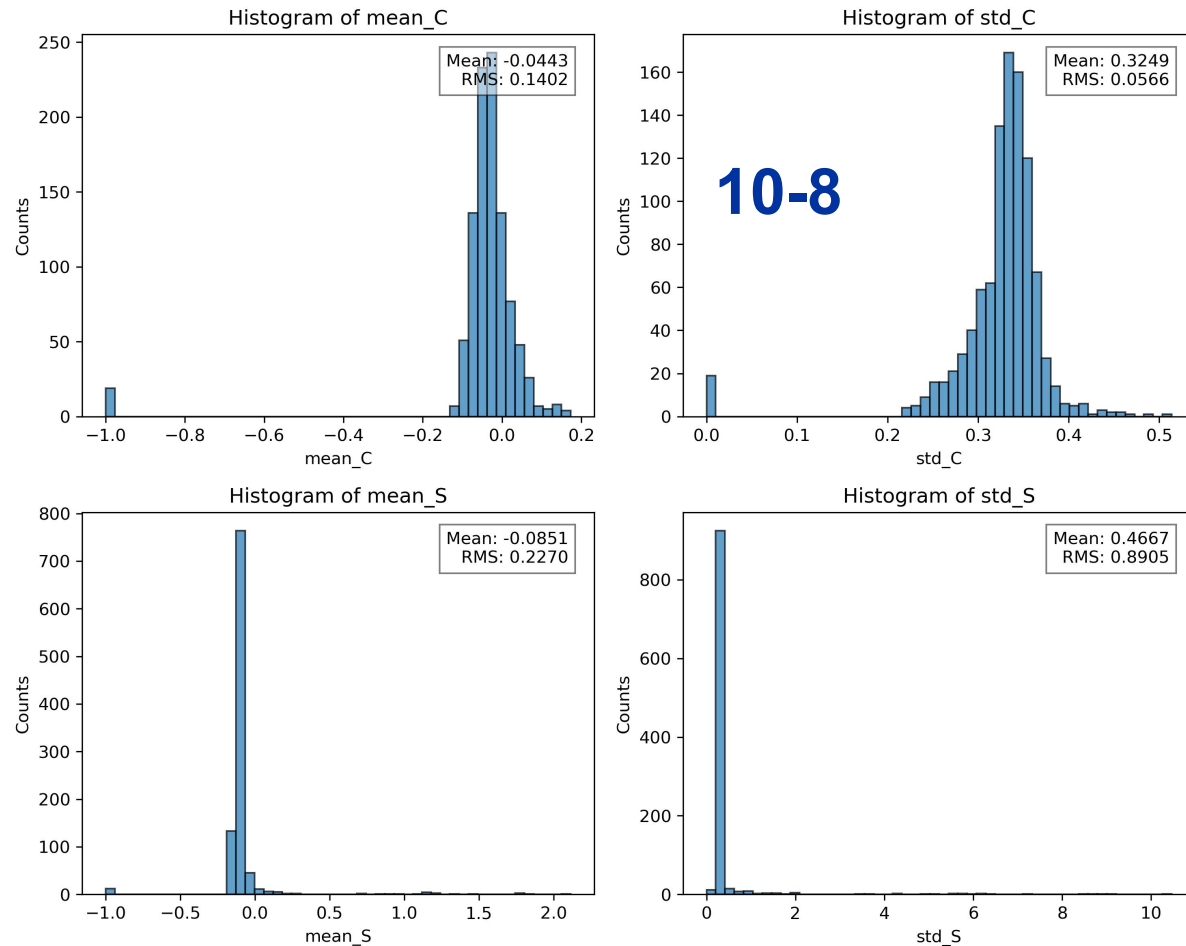
2. About 2 Hidden Layer Structures (Taking 10-8 as an example)



➤ The ratio of predicted values to true values fluctuates around 1 in both plots.

Exploration of Different Model Architectures

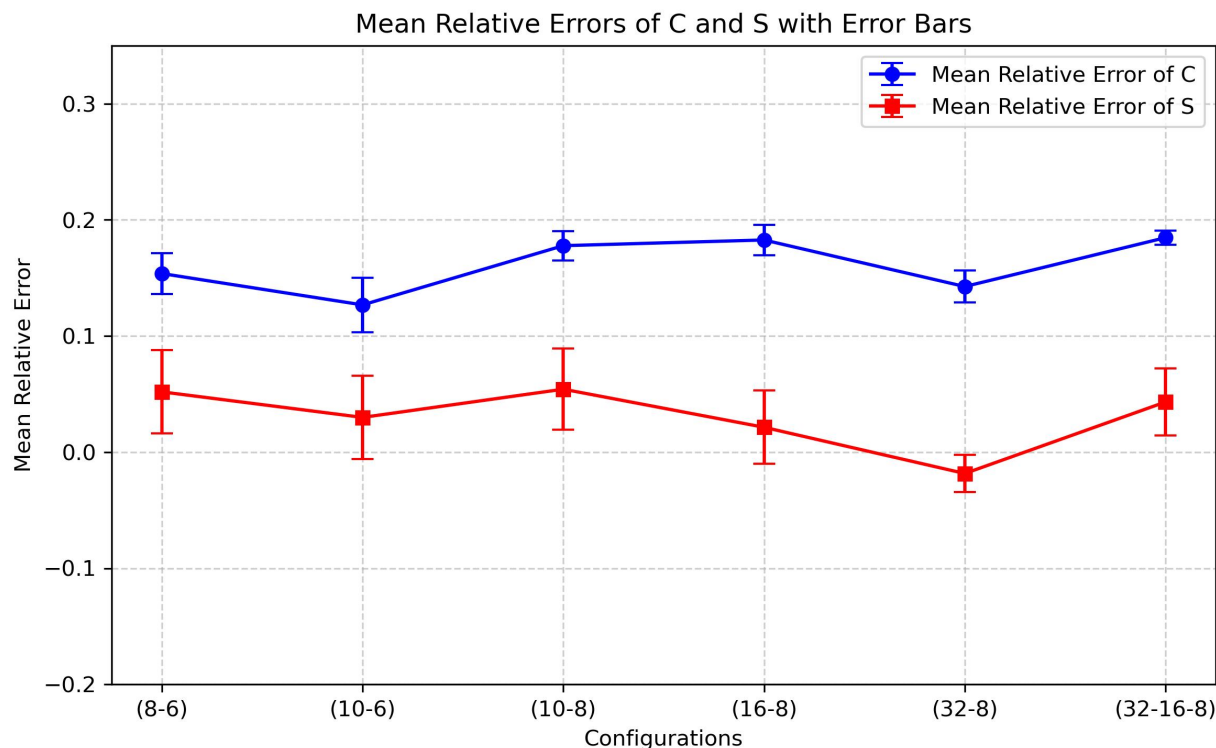
3. Analysis of the Relative Prediction Error for the 10-8 Architecture (run 1000 times)



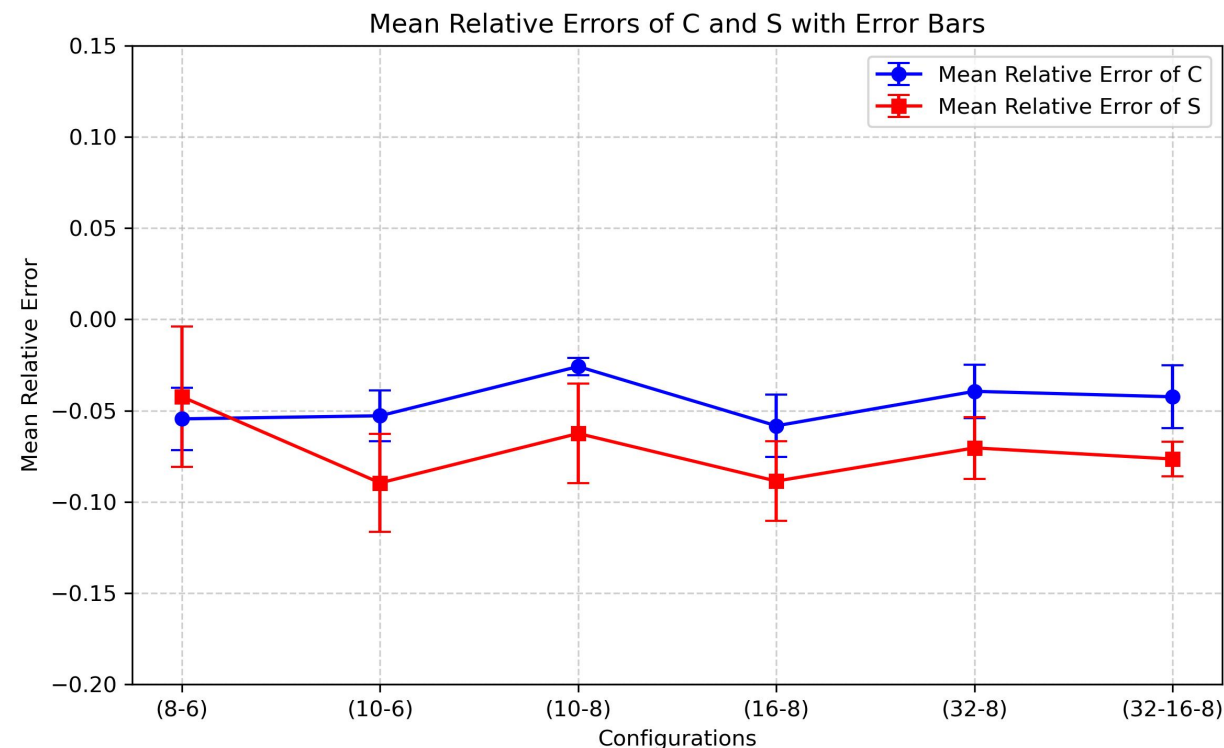
- The mean_C/S here is the mean value of the histograms for the relative error of C/S, the same for std.
- Since some true value of C/S is too small, which may bring an extremely huge relative error, in these histograms, I set:
$$\text{relativeerr} = \text{np.where}(\text{true} \geq 0.1, (\text{pred} - \text{true}) / \text{true}, 10)$$

Exploration of Different Model Architectures

4. Overview of the Relative Errors for Different Architectures (run 1000 times)



`relative_error = np.where(true >= 0.1, (pred - true) / true, 10)`

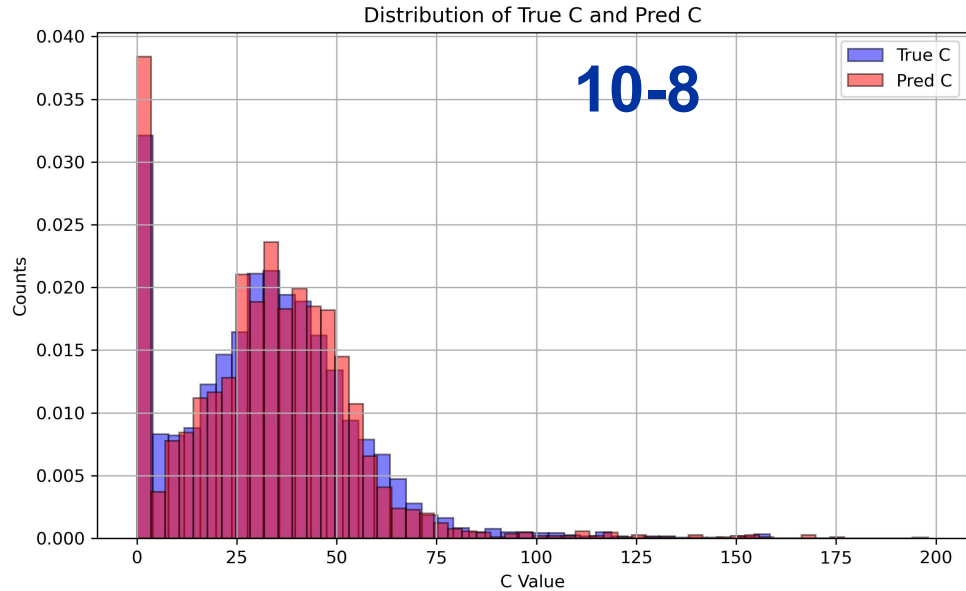


Only count when $C/S \geq 1$

- It seems that our model consistently underestimates the values of C/S, especially for S.

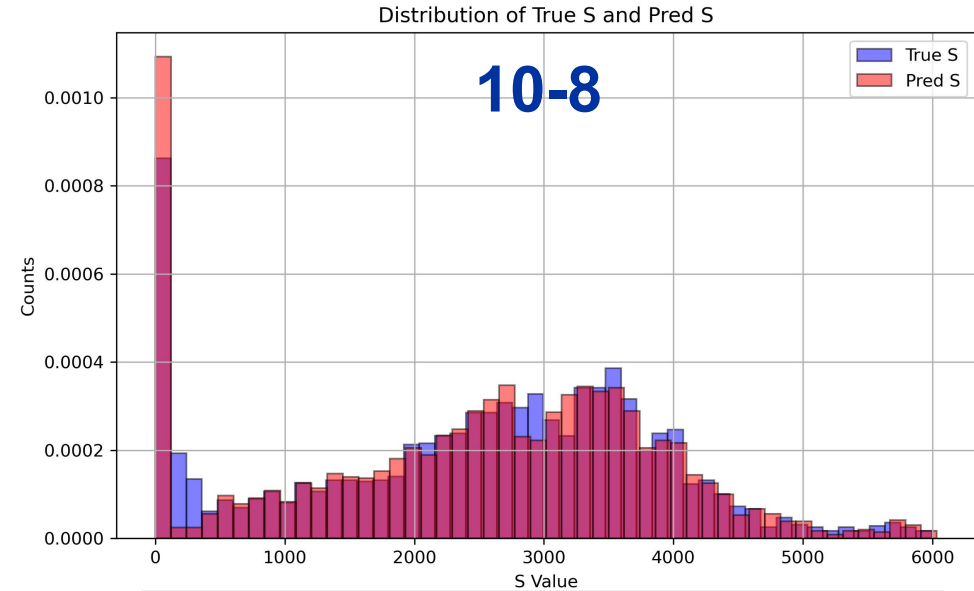
On the wierd peak near “0”

1. A Preliminary Check



Number of samples where both True C and Pred C are below 3: 329

Sample 1: True C = 2.116718292236328, Pred C = 5.204170427930421e-15
Sample 9: True C = 2.4099936485290527, Pred C = 5.204170427930421e-15
Sample 13: True C = 0.9091663956642151, Pred C = 5.204170427930421e-15
Sample 25: True C = 2.283576726913452, Pred C = 5.204170427930421e-15
Sample 45: True C = 1.8614501953125, Pred C = 5.204170427930421e-15
Sample 48: True C = 1.68816910672831e-07, Pred C = 5.204170427930421e-15
Sample 50: True C = 0.7120056748390198, Pred C = 5.204170427930421e-15
Sample 63: True C = 0.5971069931983948, Pred C = 5.204170427930421e-15
Sample 74: True C = 2.412651538848877, Pred C = 5.204170427930421e-15
Sample 107: True C = 2.1470437049865723, Pred C = 5.204170427930421e-15
Sample 108: True C = 1.722368836402893, Pred C = 5.204170427930421e-15
Sample 111: True C = 0.5397070050239563, Pred C = 5.204170427930421e-15
Sample 132: True C = 1.0415740013122559, Pred C = 5.204170427930421e-15
Sample 160: True C = 0.906544029712677, Pred C = 5.204170427930421e-15
Sample 180: True C = 0.13038106262683868, Pred C = 5.204170427930421e-15

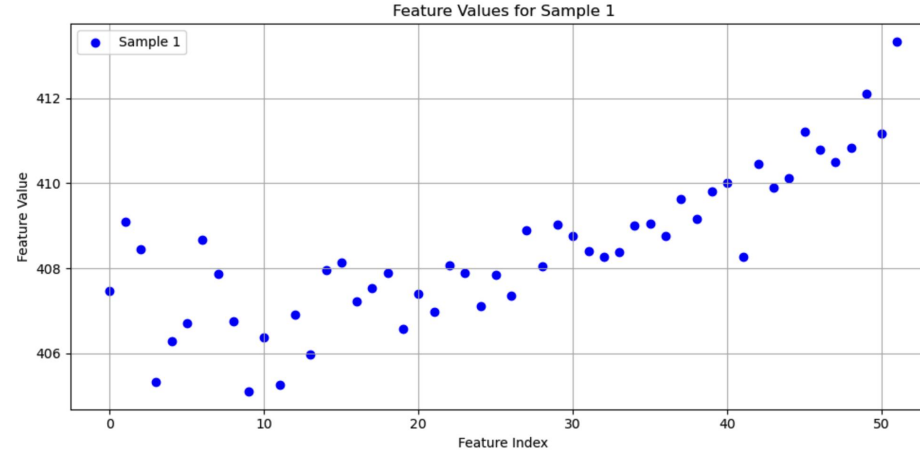
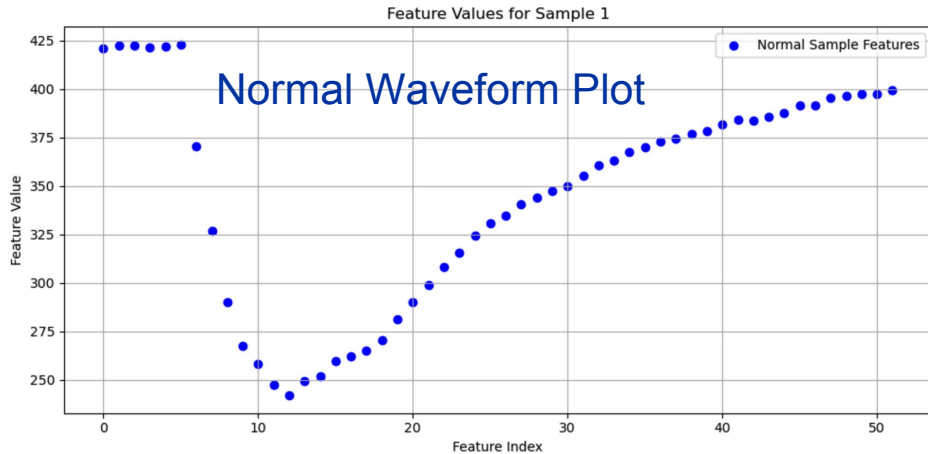


Number of samples where both True S and Pred S are below 10: 150

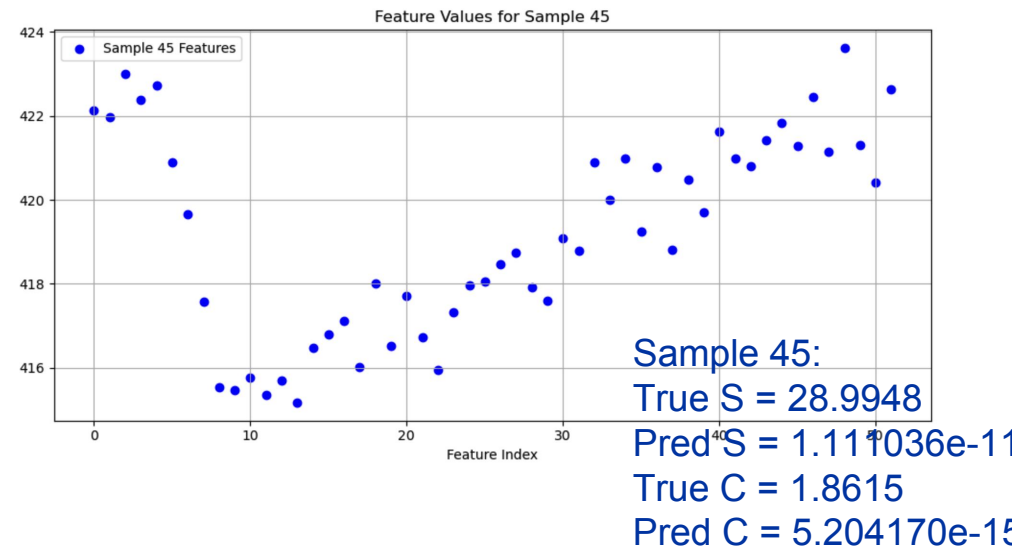
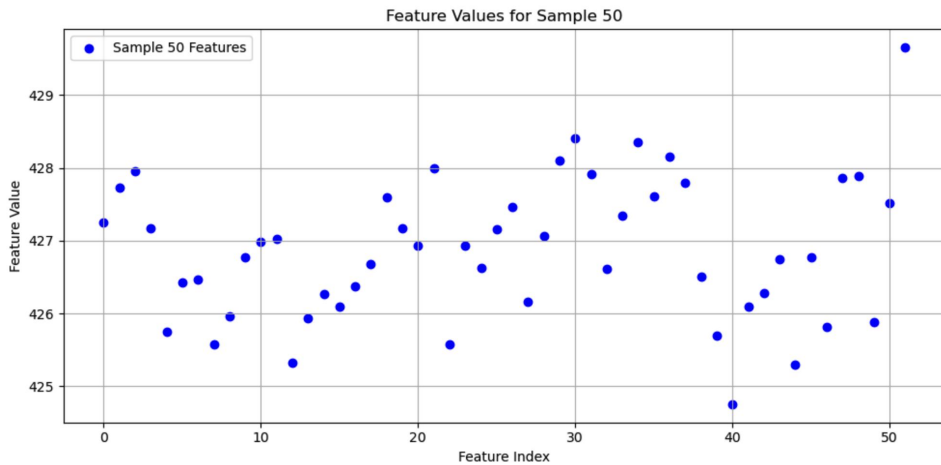
Sample 1: True S = 4.298632621765137, Pred S = 1.1110361762545207e-11
Sample 13: True S = 9.110523223876953, Pred S = 1.1110361762545207e-11
Sample 25: True S = 0.0002447253791615367, Pred S = 1.1110361762545207e-11
Sample 48: True S = 8.468006134033203, Pred S = 1.1110361762545207e-11
Sample 50: True S = 3.8178601264953613, Pred S = 1.1110361762545207e-11
Sample 74: True S = 1.9442563115035227e-09, Pred S = 1.1110361762545207e-11
Sample 111: True S = 2.3913464546203613, Pred S = 1.1110361762545207e-11
Sample 160: True S = 5.146298408508301, Pred S = 1.1110361762545207e-11
Sample 204: True S = 0.8707183003425598, Pred S = 1.1110361762545207e-11
Sample 213: True S = 2.855391263961792, Pred S = 1.1110361762545207e-11
Sample 244: True S = 5.686730861663818, Pred S = 1.1110361762545207e-11
Sample 302: True S = 5.247684478759766, Pred S = 1.1110361762545207e-11
Sample 358: True S = 4.584608554840088, Pred S = 1.1110361762545207e-11
Sample 359: True S = 4.7810959815979, Pred S = 1.1110361762545207e-11
Sample 388: True S = 3.678415536880493, Pred S = 1.1110361762545207e-11
Sample 395: True S = 4.403695106506348, Pred S = 1.1110361762545207e-11
Sample 414: True S = 4.383340358734131, Pred S = 1.1110361762545207e-11

On the wierd peak near “0”

1. A Preliminary Check



➤ Are the true values provided by SDL reliable for these waveform data?



On the wierd peak near “0”

2. Modifying the Dataset

➤ Original Total Dataset Size:

$$29799 = 32501 - 15 - 2687$$

➤ New Total Dataset Size:

$$24598 = 32501 - 15 - 2687 - 5201$$

Number of events with abnormal signal triggers

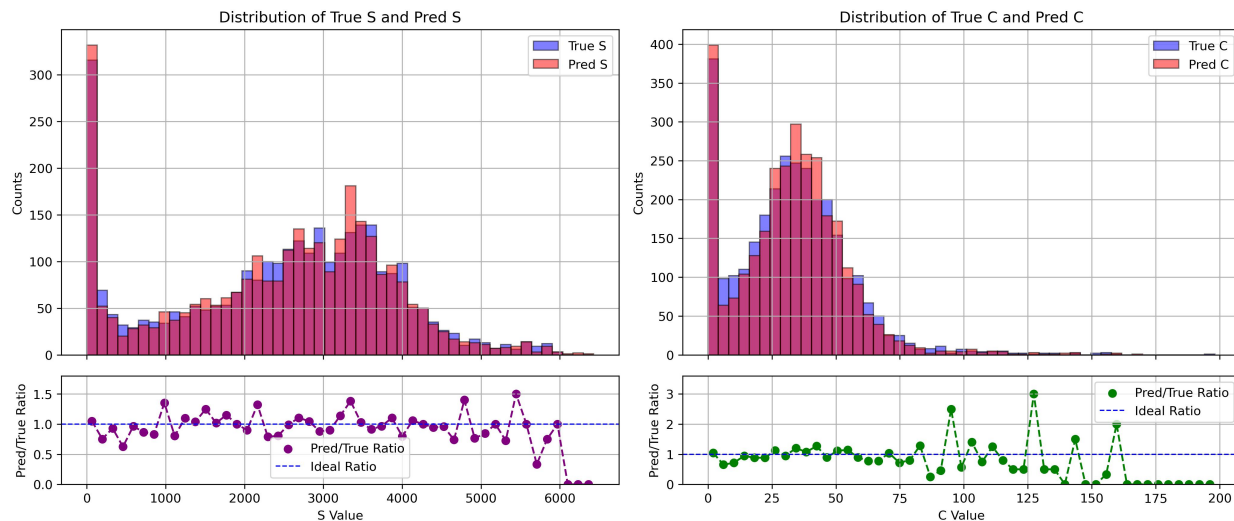
Total number of events
in channel 3 of run 200 data

Number of events with
almost no signal,
determined by the
condition: $C + S < 2$

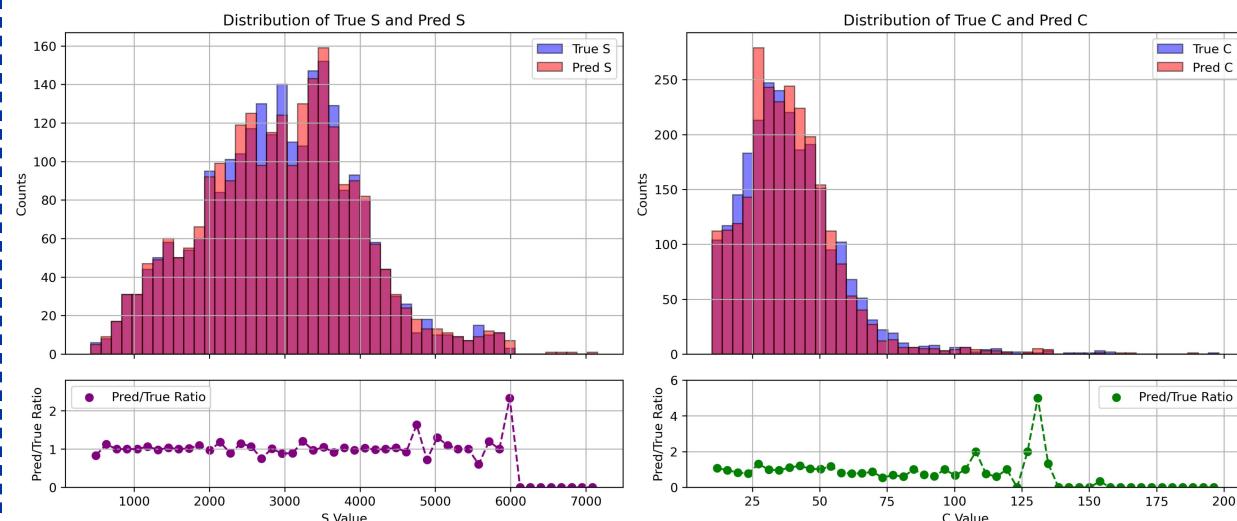
Number of events with
low signal, determined
by the condition:
 $C \geq 10 \ \& \ S \geq 10$

On the wierd peak near “0”

3. Presentation of New Prediction Results (for 32-16-8)



Original Dataset

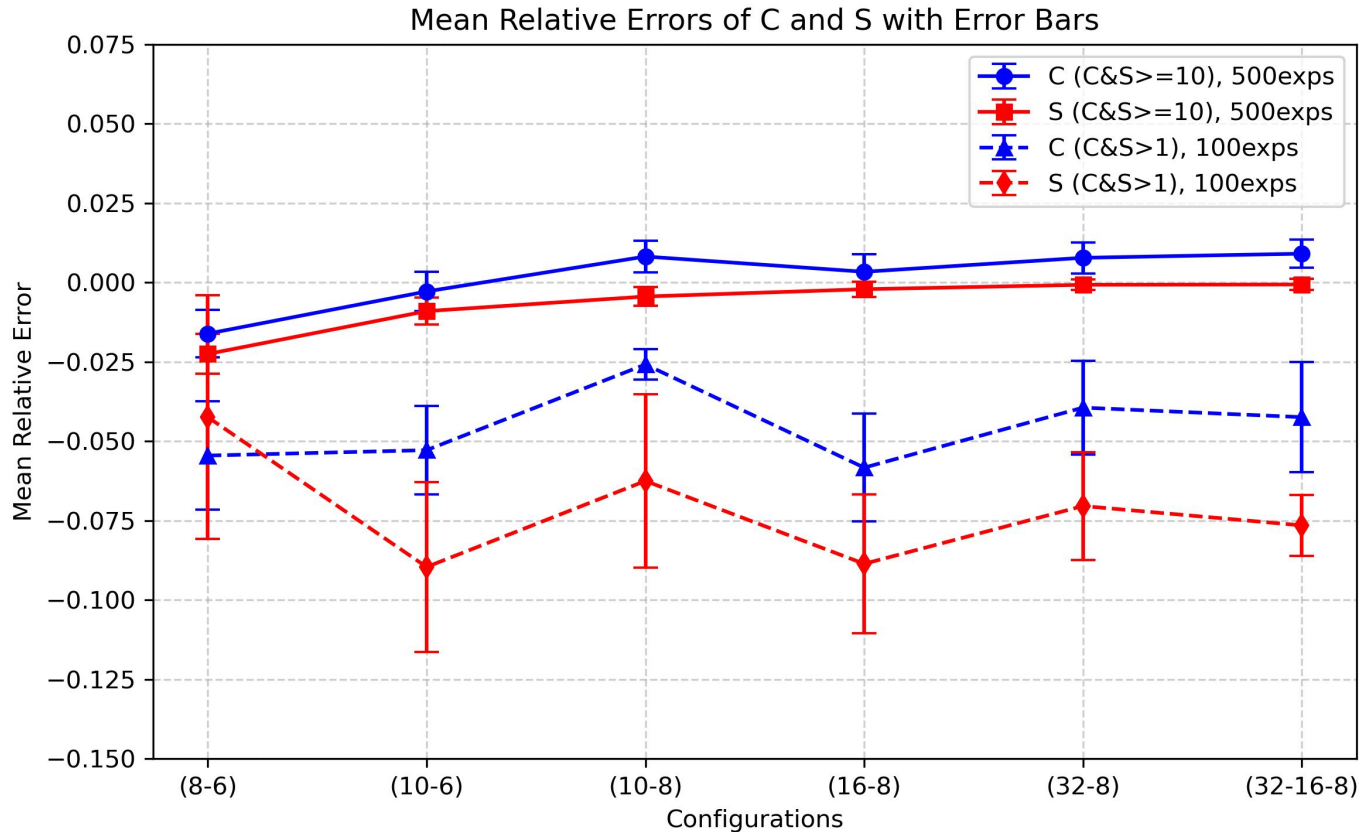


Dataset After Removing Low-signal Events

➤ The results produced by the model trained on the new dataset appear to be more stable and improved.

On the wierd peak near “0”

4. Impact on Models with Different Architectures

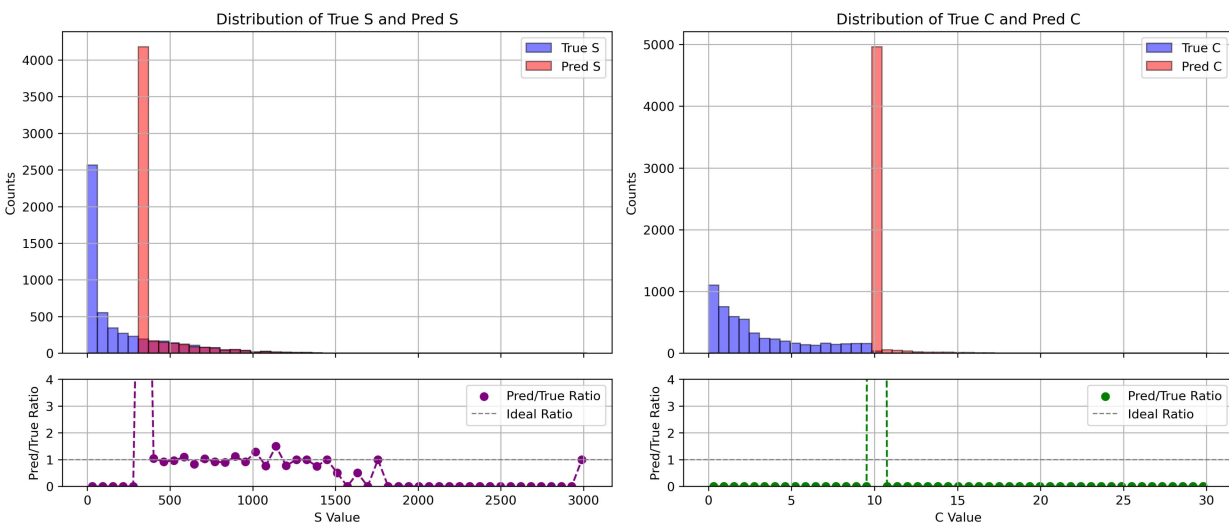


- The model trained on the new dataset clearly exhibits smaller relative errors.
- However, whether this improved accuracy can be maintained when predicting low-value signals remains a question.

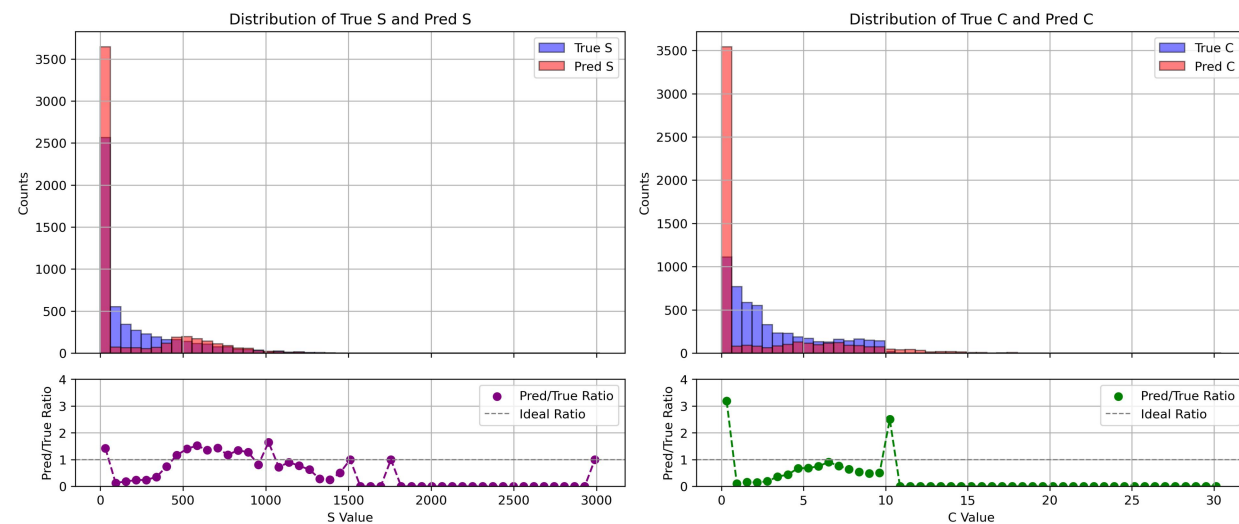
On the wierd peak near “0”

5. Evaluation of Prediction Performance on Low-Signal Events

- Create a test set that includes only the data where $C < 10$ or $S < 10$:



Model Trained on the Dataset with $C \& S \geq 10$

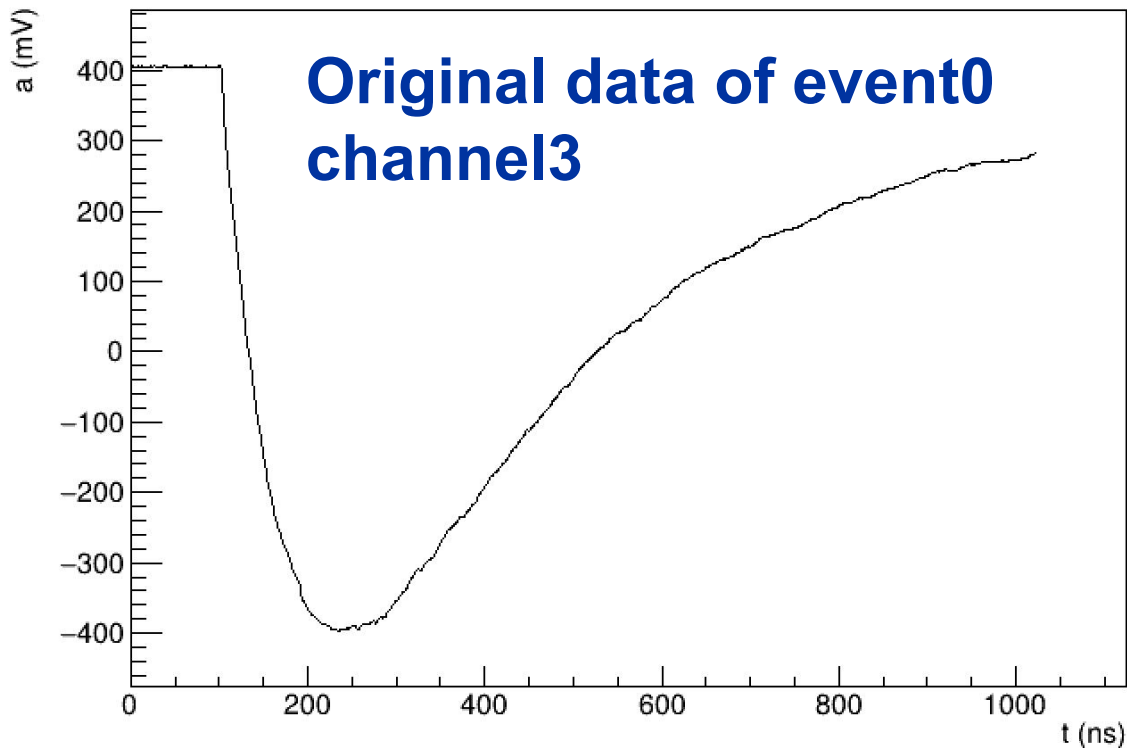


Model Trained on the Full Dataset

Constructing a Synthetic Waveform

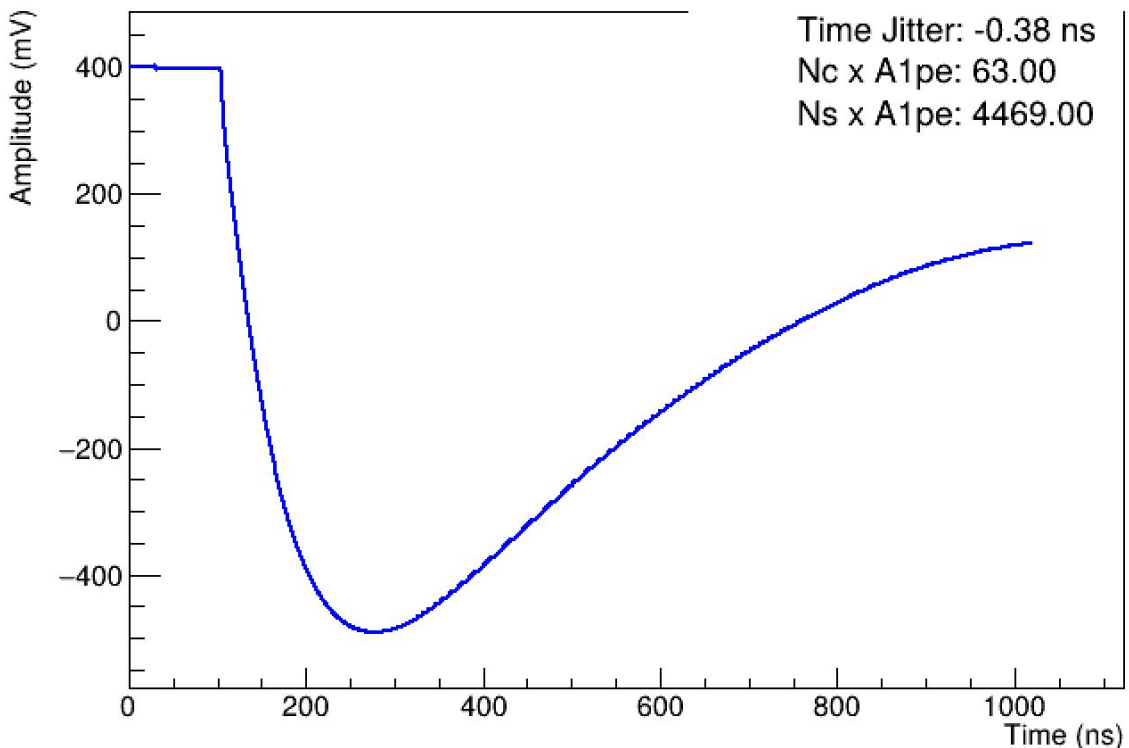
1. Evaluation of Prediction Performance on Low-Signal Events

Graph



time jitter	=	-0.386282	+/-	0.0783419	(limited)
Nc x A1pe	=	63.1718	+/-	3.58578	(limited)
Ns x A1pe	=	4469.01	+/-	89.502	(limited)

Reconstructed Original Waveform



time jitter	=	4.5799e-05	+/-	1.07602e-05	(limited)
Nc x A1pe	=	63.0008	+/-	0.000491394	(limited)
Ns x A1pe	=	4468.98	+/-	0.0122766	(limited)

Resource Report

➤ The 32-16-8 model:

```
=====
== Performance Estimates
=====
+ Timing:
  * Summary:
  +-----+-----+-----+-----+
  | Clock | Target | Estimated | Uncertainty |
  +-----+-----+-----+-----+
  | ap_clk | 5.00 ns | 3.617 ns | 1.35 ns |
  +-----+-----+-----+-----+

+ Latency:
  * Summary:
  +-----+-----+-----+-----+-----+-----+
  | Latency (cycles) | Latency (absolute) | Interval | Pipeline |
  | min | max | min | max | min | max | Type |
  +-----+-----+-----+-----+-----+-----+
  | 14 | 14 | 70.000 ns | 70.000 ns | 1 | 1 | yes |
  +-----+-----+-----+-----+-----+-----+
```

```
=====
== Utilization Estimates
=====
* Summary:
+-----+-----+-----+-----+-----+-----+
| Name | BRAM_18K | DSP | FF | LUT | URAM |
+-----+-----+-----+-----+-----+-----+
| DSP | - | - | - | - | - |
| Expression | - | - | 0 | 12 | - |
| FIFO | - | - | - | - | - |
| Instance | - | 1864 | 12209 | 68690 | - |
| Memory | - | - | - | - | - |
| Multiplexer | - | - | 0 | 36 | - |
| Register | - | - | 3571 | - | - |
+-----+-----+-----+-----+-----+-----+
| Total | 0 | 1864 | 15780 | 68738 | 0 |
+-----+-----+-----+-----+-----+-----+
| Available SLR | 1344 | 3072 | 864000 | 432000 | 320 |
+-----+-----+-----+-----+-----+-----+
| Utilization SLR (%) | 0 | 60 | 1 | 15 | 0 |
+-----+-----+-----+-----+-----+-----+
| Available | 5376 | 12288 | 3456000 | 1728000 | 1280 |
+-----+-----+-----+-----+-----+-----+
| Utilization (%) | 0 | 15 | ~0 | 3 | 0 |
+-----+-----+-----+-----+-----+-----+
```

Resource Report

➤ The 10-8 model:

```
=====
== Performance Estimates
=====
+ Timing:
  * Summary:
    +-----+-----+-----+-----+
    | Clock | Target | Estimated | Uncertainty |
    +-----+-----+-----+-----+
    | ap_clk | 5.00 ns | 3.490 ns | 1.35 ns |
    +-----+-----+-----+-----+

+ Latency:
  * Summary:
    +-----+-----+-----+-----+-----+
    | Latency (cycles) | Latency (absolute) | Interval | Pipeline |
    | min | max | min | max | min | max | Type |
    +-----+-----+-----+-----+-----+
    | 10 | 10 | 50.000 ns | 50.000 ns | 1 | 1 | yes |
    +-----+-----+-----+-----+-----+
```

```
=====
== Utilization Estimates
=====
* Summary:
+-----+-----+-----+-----+-----+
| Name | BRAM_18K | DSP | FF | LUT | URAM |
+-----+-----+-----+-----+-----+
| DSP | - | - | - | - | - |
| Expression | - | - | 0 | 10 | - |
| FIFO | - | - | - | - | - |
| Instance | - | 508 | 3107 | 18314 | - |
| Memory | - | - | - | - | - |
| Multiplexer | - | - | 0 | 36 | - |
| Register | - | - | 1742 | - | - |
+-----+-----+-----+-----+-----+
| Total | 0 | 508 | 4849 | 18360 | 0 |
+-----+-----+-----+-----+-----+
| Available SLR | 1344 | 3072 | 864000 | 432000 | 320 |
+-----+-----+-----+-----+-----+
| Utilization SLR (%) | 0 | 16 | ~0 | 4 | 0 |
+-----+-----+-----+-----+-----+
| Available | 5376 | 12288 | 3456000 | 1728000 | 1280 |
+-----+-----+-----+-----+-----+
| Utilization (%) | 0 | 4 | ~0 | 1 | 0 |
+-----+-----+-----+-----+-----+
```

Thank you !



home.cern

Backups

Try to extract some data from “outfile_LG.root”

- Some raw data:

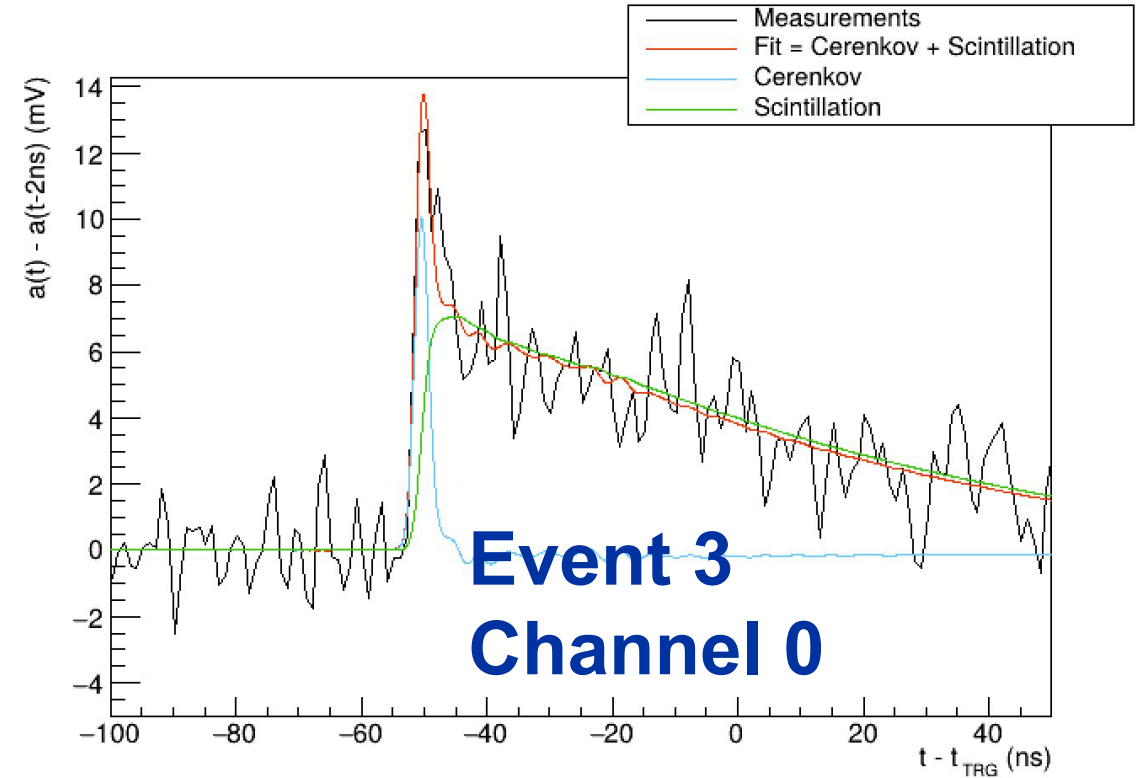
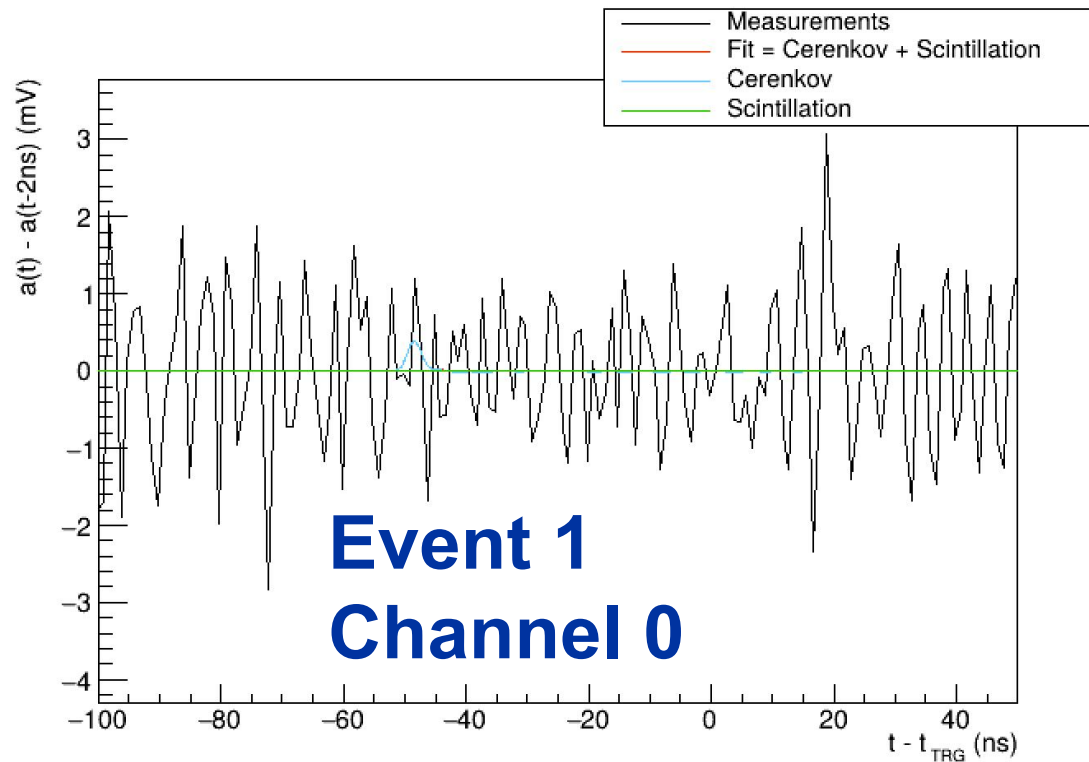
Note:

```
* Run = 200
* Channels 0, 1, 2, and 3 are in the front
* Channels 4 and 7 are in the rear
* Channels 5 and 6 are discarded from the analysis (dead)
```

	Event 0	Event 1	Event 3
CH0	(21.35±1.96, 9.2% 2005.27±48.13 2.4%)	(0.58±0.96, 0.15±963.30)	(14.93±1.44, 1201.10±36.12)
CH1	(40.85±2.22, 5.4% 2725.51±55.79 2.0%)	(1.63e-5±4.75, 0.49±234.02)	(23.17±1.93, 1823.87±50.66)
CH2	(22.49±1.84, 8.2% 1845.23±45.39 2.5%)	(8.11e-8±2.44274, 1.82e-5±96.25)	(12.27±1.83, 1107.07±43.9472)
CH3	(63.17±3.59, 5.7% 4469.01±89.50 2.0%)	(0.81±0.95, 3.90e-5±176.82)	(47.05±3.22, 3576.87±82.47)
CH4	(14.89±0.96, 6.4% 66.23±26.44 39.9%)	(0.38± 0.91, 2.52±69.94)	(9.94±1.13, 55.21±31.13)
CH7	(85.18±1.52, 1.8% 62.48±40.96 65.6%)	(1.31e-6±5.57, 4.26±39.81)	(16.94±1.04, 47.86±27.97)

→ (Nc x A1pe, Ns x A1pe)

Some Fitting Plots (SDL waveforms)

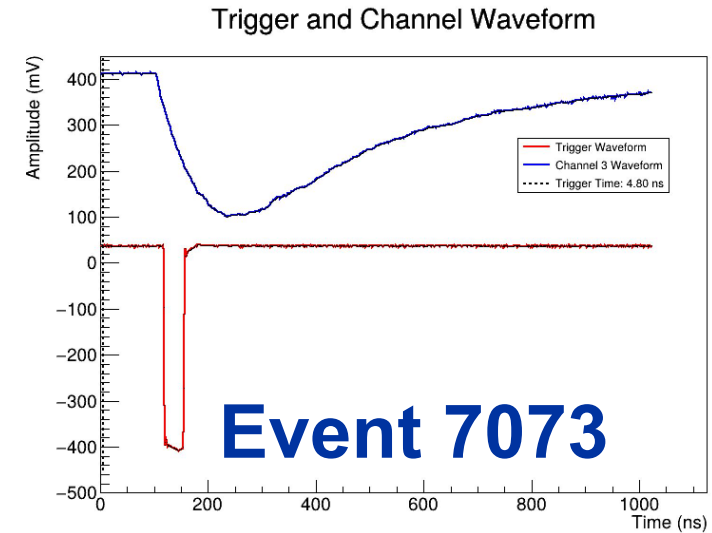
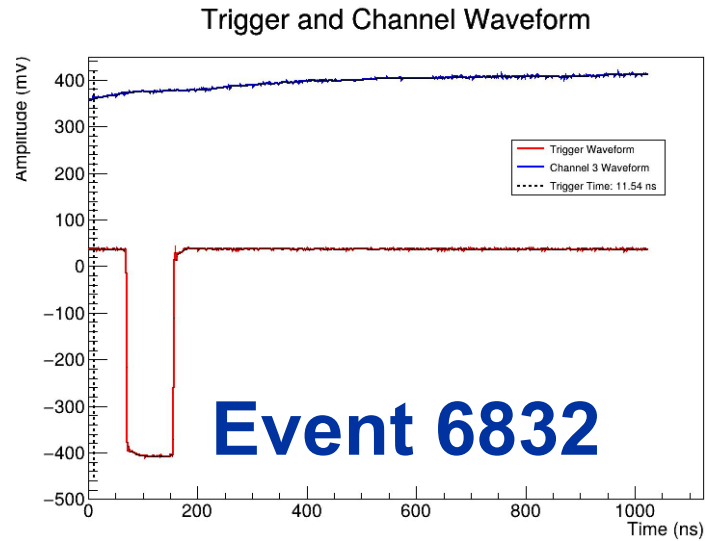
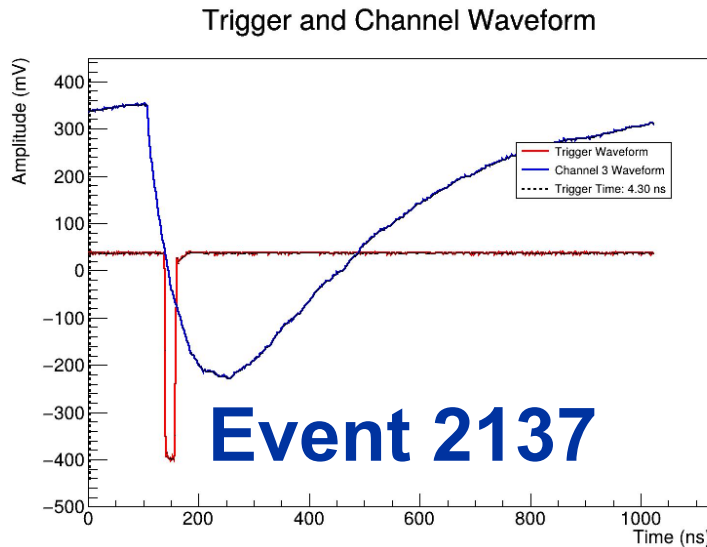


Check of Trigger Time

About the “bad” events

The count of bad events: 15

Bad Event IDs: 2137 4467 6832 7073 7319 8242 8524 9200 9424 11066 12134 13892 15019 17990 29539

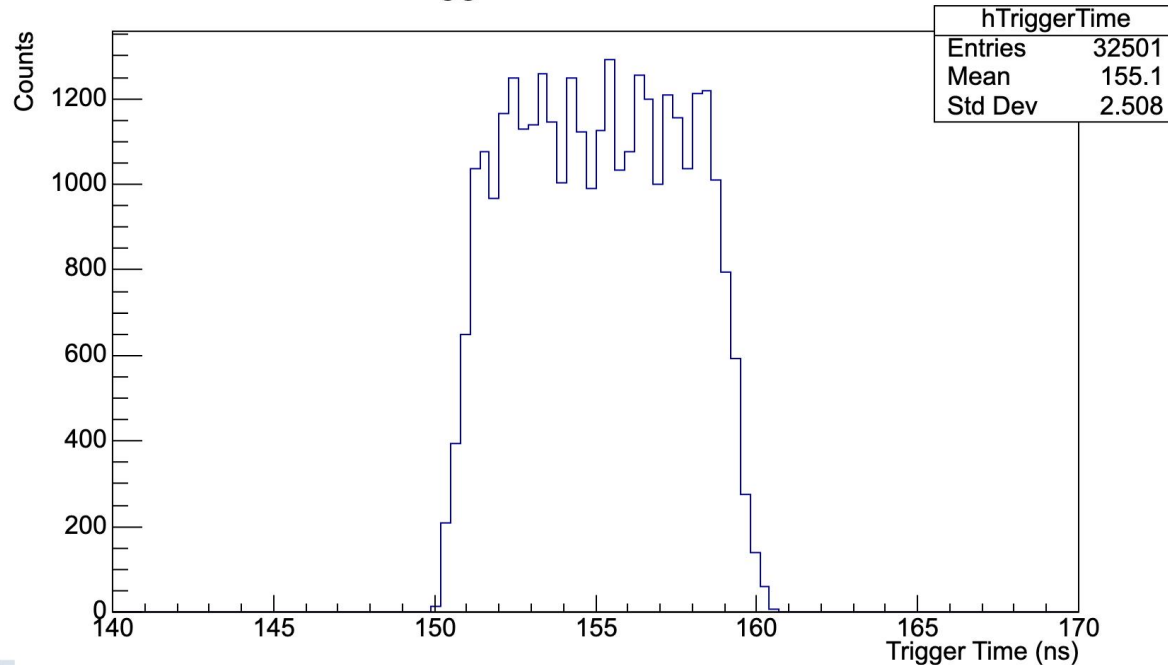


- These events have very small trigger times because the threshold is set as 0.5 times the voltage difference between the 0-100 ns and 400-1000 ns signal regions, beyond the 0-100 ns range.
- These events cannot be used for ML analysis, but we may need to understand what happened to their triggers.

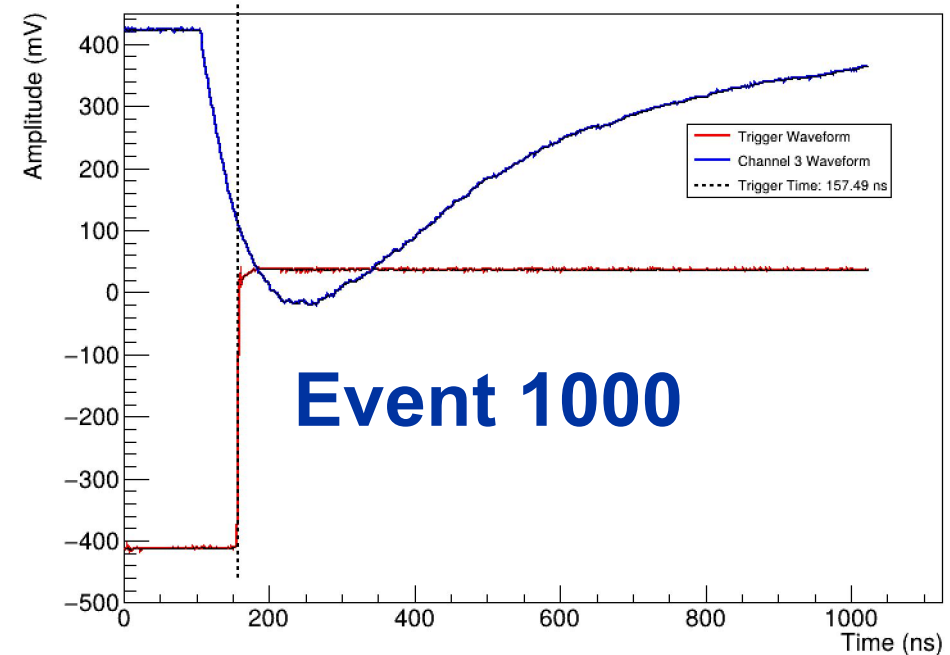
Check of Trigger Time

About the “good” events

Trigger Time Distribution



Trigger and Channel Waveform

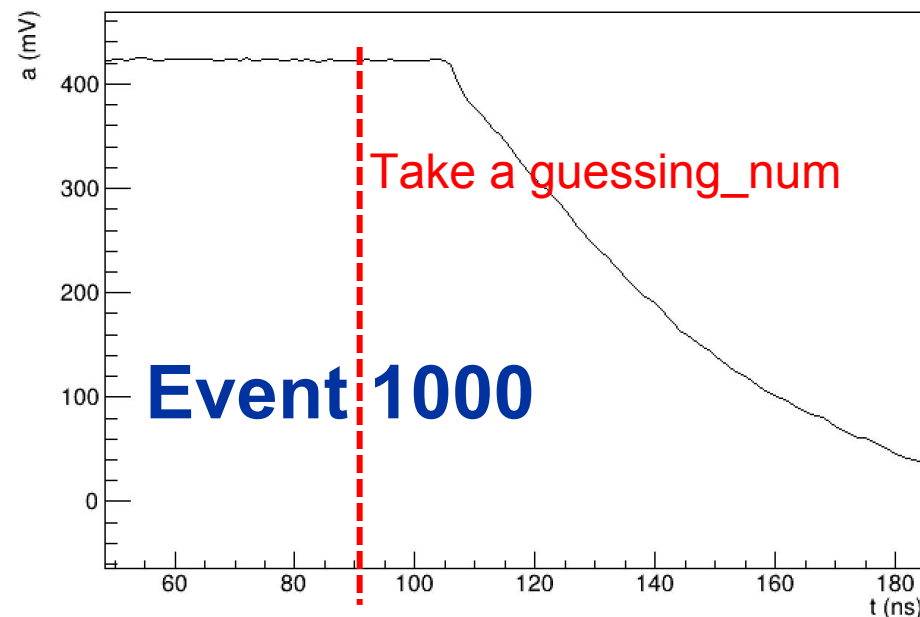


- The left graph shows that the trigger time distribution is not very Gaussian but strictly within 145-165 ns.
- The right graph suggests the trigger time doesn't match the desired exclusion point. Maybe the electronics are configured to save the waveform for a specific time before the trigger.

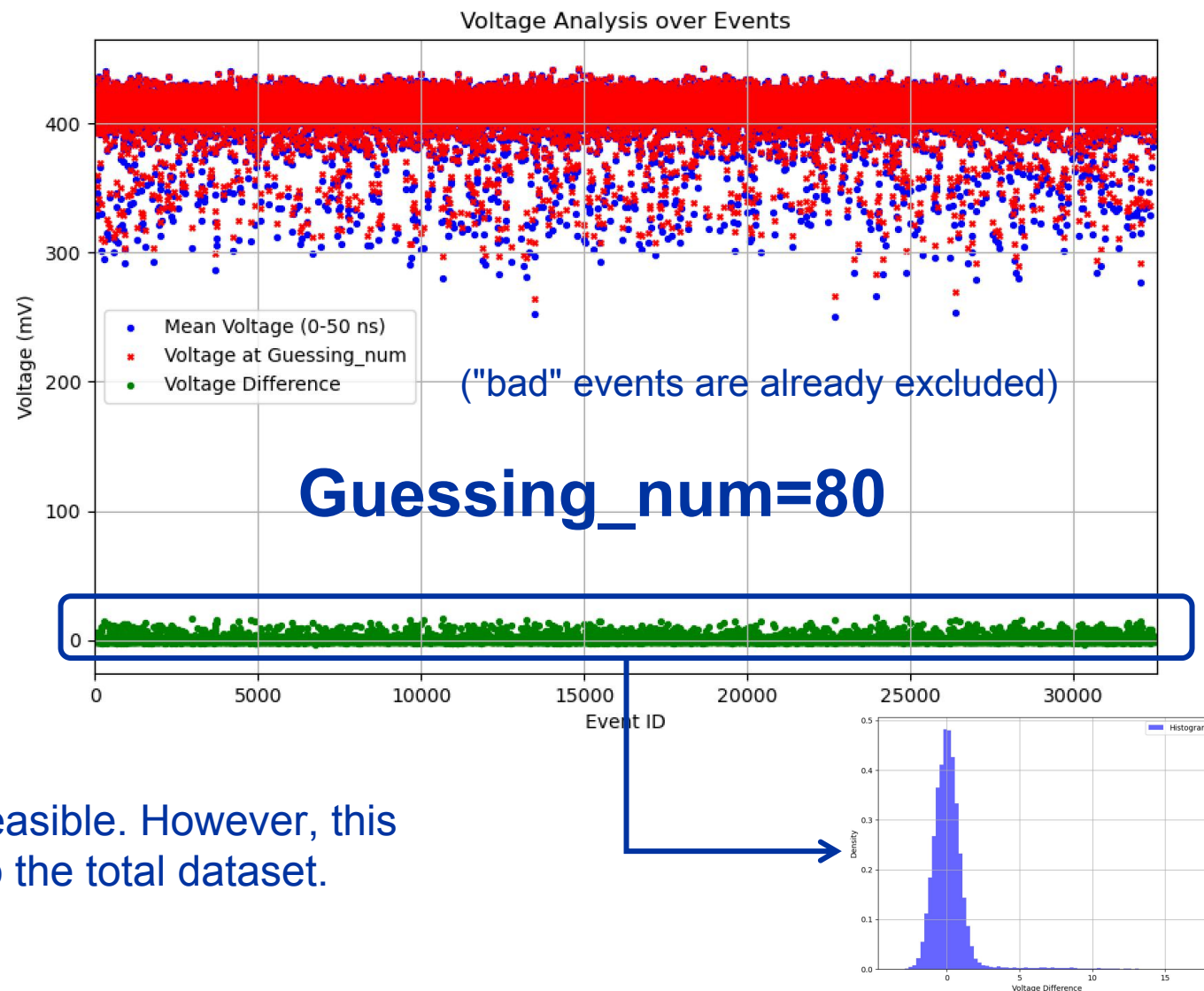
Check of Trigger Time

About the “good” events

Locally magnified channel waveform:



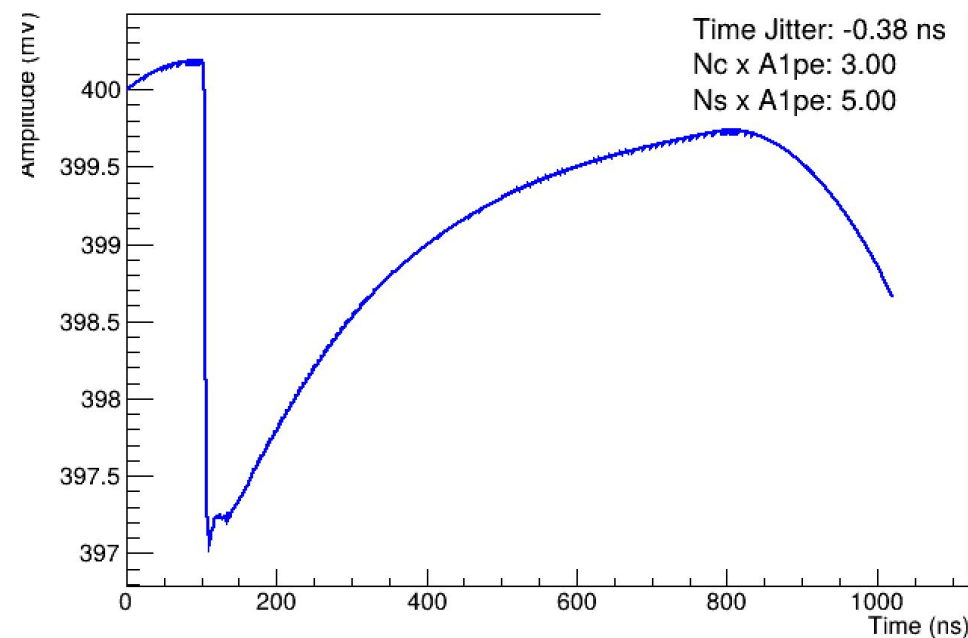
- Excluding data points before 80 ns seems feasible. However, this portion of data is not significant compared to the total dataset.



Constructing a Synthetic Waveform

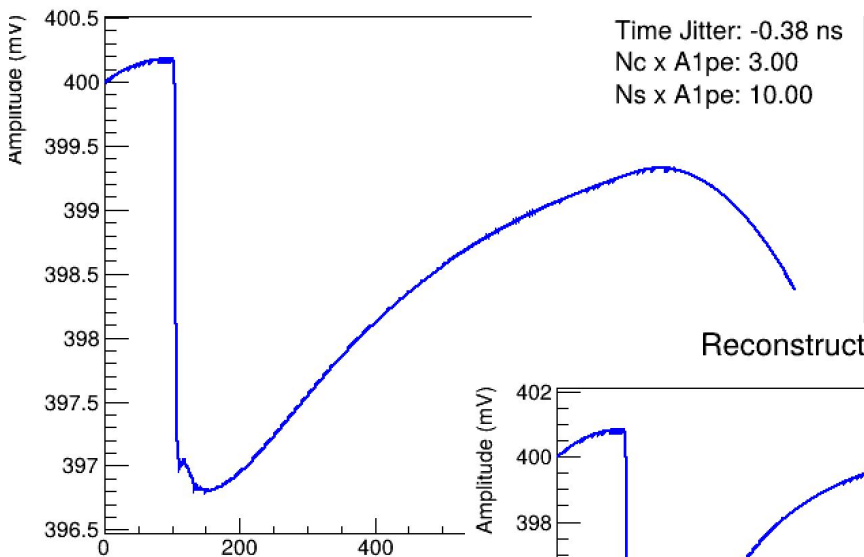
1. Evaluation of Prediction Performance on Low-Signal Events

Reconstructed Original Waveform

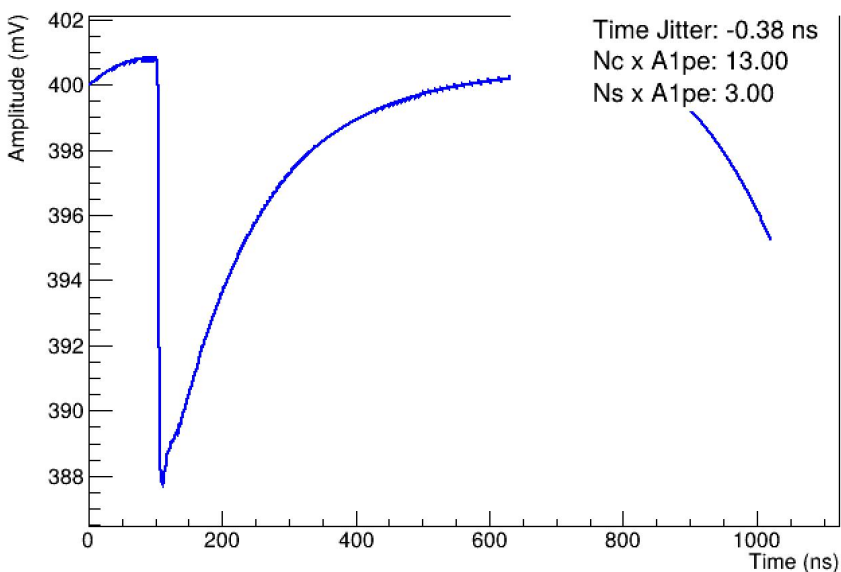


time jitter	=	-0.0004792	+/-	0.000221048	(limited)
Nc x A1pe	=	2.99921	+/-	0.00045633	(limited)
Ns x A1pe	=	4.95219	+/-	0.0168701	(limited)

Reconstructed Original Waveform



Reconstructed Original Waveform



Normalization of the C/S values

➤ The Current Data Preprocessing:

- ✓ StandardScaler() to the input data
- ✓ MinMaxScaler() to the output data (C&S separately)

➤ In the DRO simulations:

- ✓ Calibration Procedure:

The total numbers of C or S photons for a given incoming particle were divided by the average number of photons obtained for single electrons with the same energy.

