Time reconstruction in MRPC detector using deep-learning algorithms

Fuyue Wang

Content

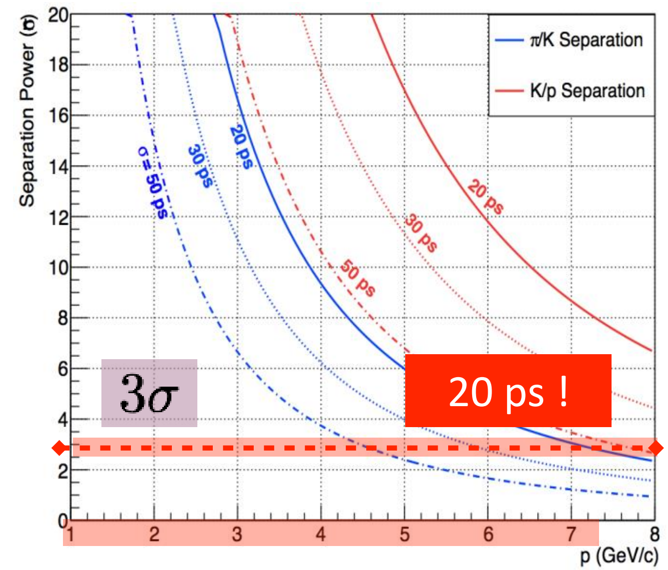
- Background
- High time resolution MRPC
- Machine Learning (ML) based reconstruction algorithm
  - ML training with simulation, testing with simulation
  - ML training with simulation, testing with experiment
  - ML training with experiment, testing with experiment
- Summary

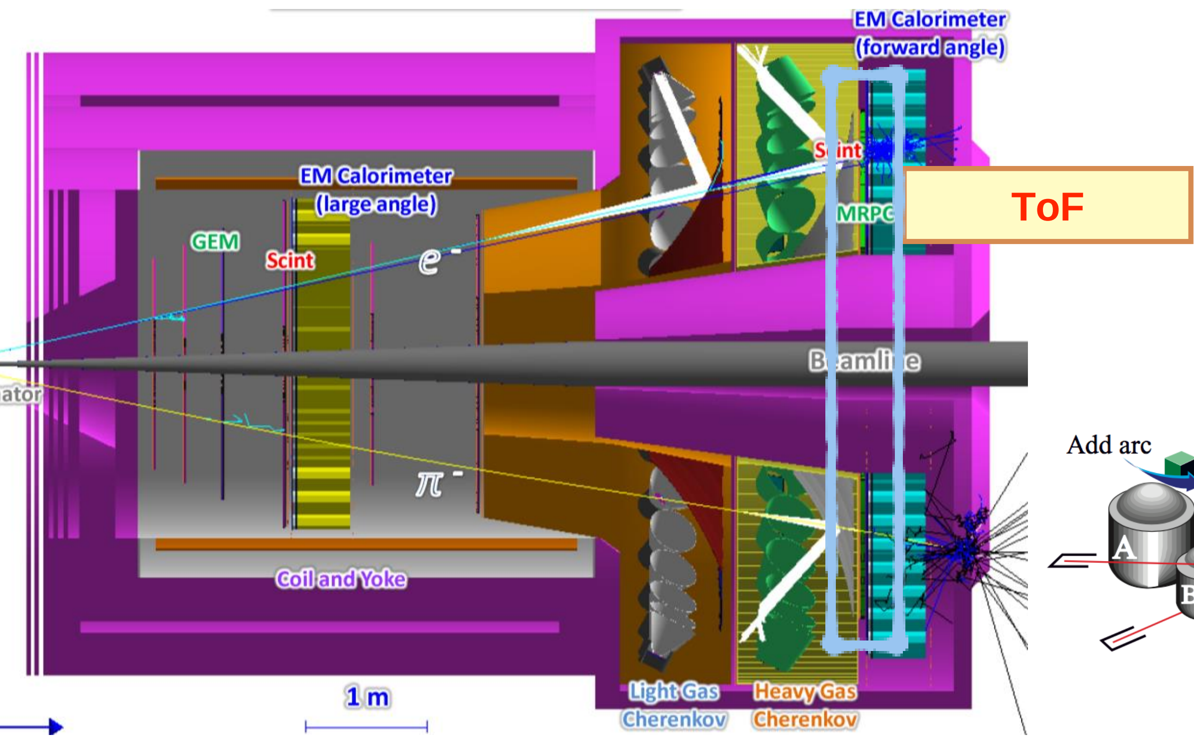
Background

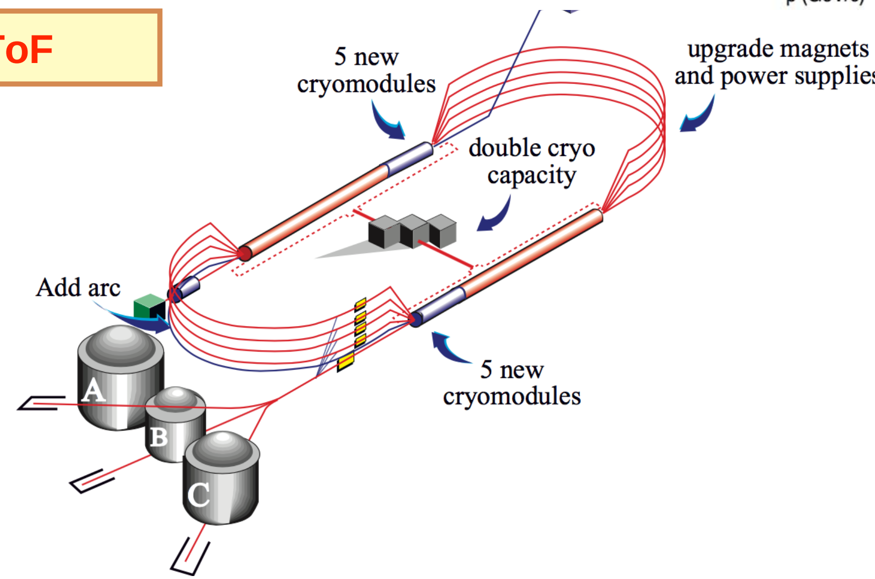
- SoLID(Solenoidal Large Intensity Device) @ JLab
- Separation between π/K up to momentum of 7GeV/c

A challenge for MRPC: 20 ps time resolution

- Time resolution of present MRPCs：
  80ps@ALICE， 70ps@STAR，60ps@CBM







Background

- How to improve ?
  1. Geometry: thinner gaps && more gaps
  2. Electronics and read out system:

From TDC to fast FEE and waveform digitizer:

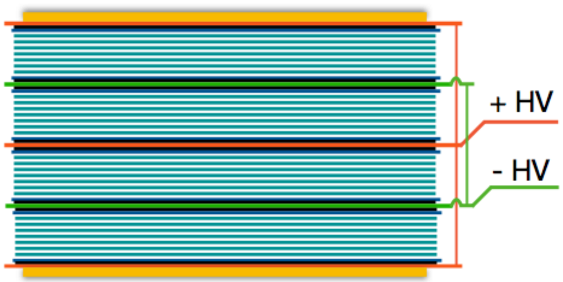
✓ Get rid of $\sigma = 20 \sim 30\mathrm{ps}$ uncertainty of every TDC channel
✓ Able to extract more information from the waveform
✓ Contribution of resolution: 15 ps from the FEE and the digitizer
15 ps from the MRPC detector

3. Time reconstruction algorithms
   From time over threshold (ToT) into deep learning based algorithms

MRPC detector

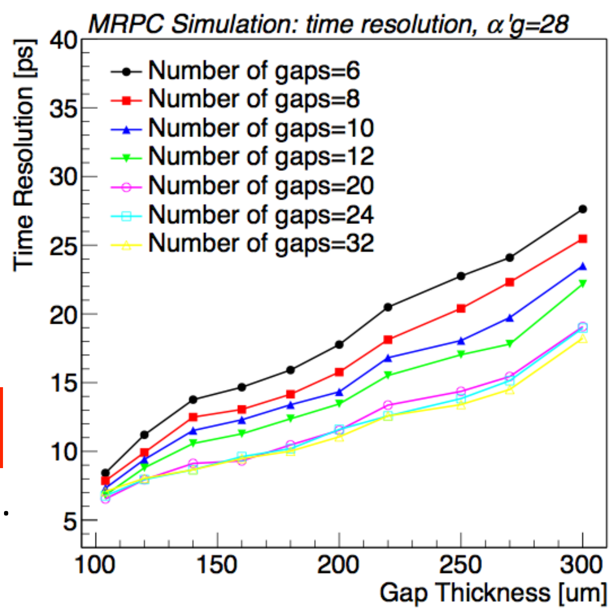
- High time resolution MRPC: selected according to simulation
- 4 x 8 gaps, 0.104 mm thickness

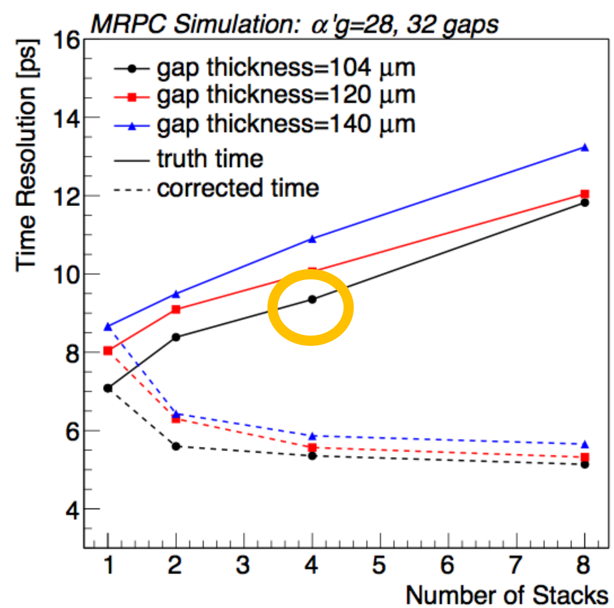
	Length（mm）	width（mm）	thickness（mm）
gap	258	80	0.104
glass	258	80	0.5
mylar	268	90	0.25
Mid PCB	328	120	1.2
PCB	298	120	0.6
Honeycomb	265	90	7.5



Intrinsic time resolution

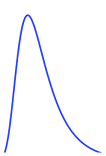
Wang F, et al. NIMA 950: 162932.





Machine Learning

- Machine learning algorithms can acquire knowledge from the data through feature extraction and representation learning
- Deep neural network is one of the most important machine learning algorithms that solves problems with significant non–linearities
- Widely used in high energy physics
- For MRPCs:
  networks are trained to reconstruct the time related to each event



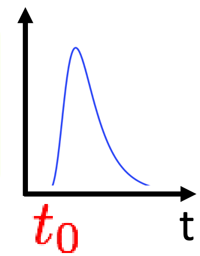
Input
MRPC waveform

→

Network
MLP, LSTM, CNN

→

Output
MRPC related Time



feature

label

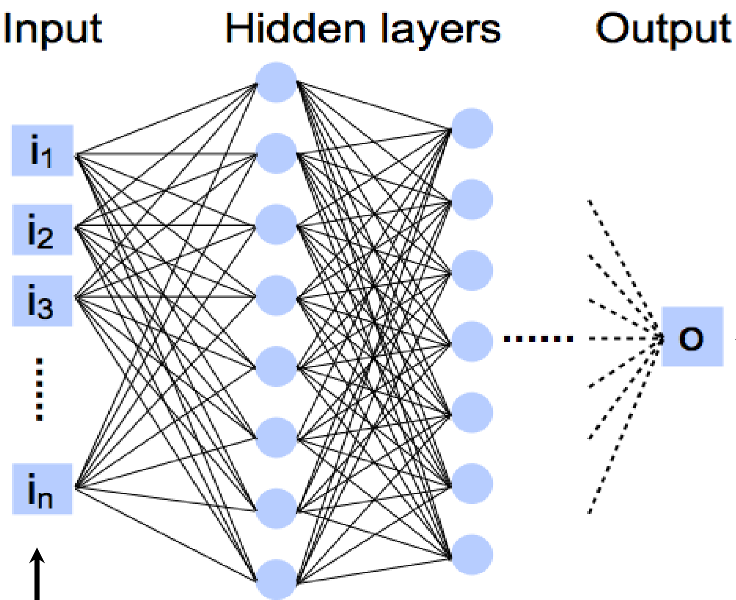
✓ Generated by simulation
✓ Generated by experiment

Tsinghua University

Multilayer perceptron (MLP)

- Multilayer perceptron (MLP)

$$\underbrace{F_i(\vec{x})}_{\text{Output}} = h(\sum_j(\omega_{ij}^2 g(\ldots g(\sum_k(\omega_{jk}^1 g(\sum_l(\omega_{kl}^0 \underbrace{x_l}_{\text{Input}} + \chi_k^0)\ldots + \chi_j^1) + \chi_i^2)$$



- Activation function: g and h —— tanh
- Weights: $\omega^{0,1\ldots}$, $\chi^{0,1\ldots}$
- "Dropout": avoid overfitting

The length of the leading edge t_r

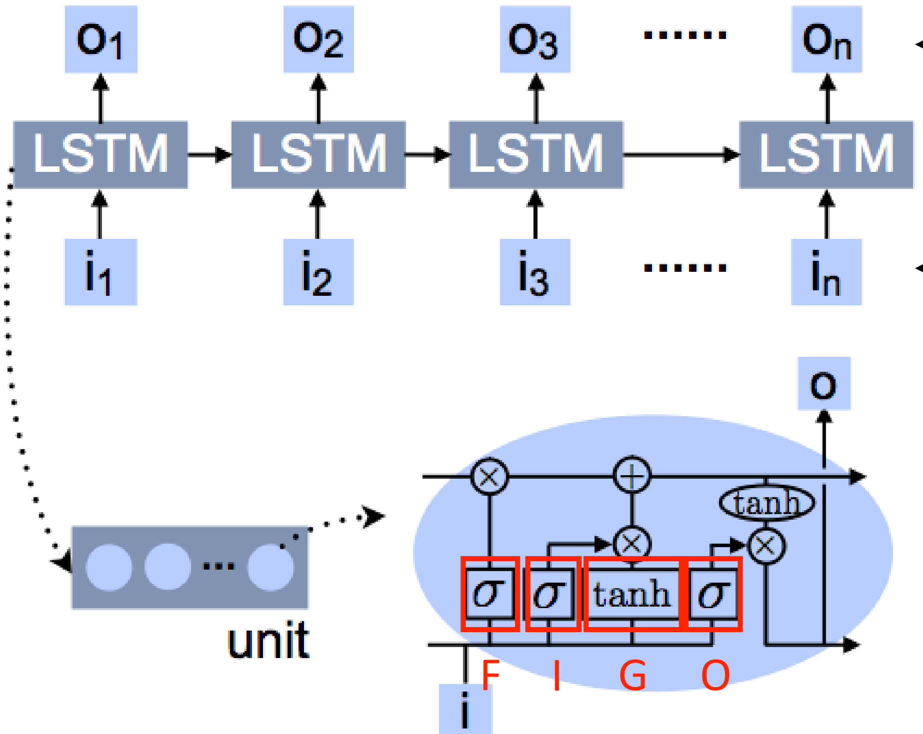
Time of the very first interaction: $t_0 = t_p - t_r$

Several uniformly distributed points along the leading edge

- Train/validate/test: 60k, 10k,6k
- Tensorflow & GPU: GTX 1080 Ti
- ~ 10 mins for training

LSTM-420

- Recurrent neural networks(RNN): Long Short Term Memory network(LSTM)



The length of the leading edge t_r

Several uniformly distributed points along the leading edge

f : forget gate: Whether to erase

I : Input gate, whether to write

g: gate gate, How much to write

o: output gate, How much to reveal

- Every "LSTM" box has 420 units
- Tensorflow & GPU: GTX 1080 Ti

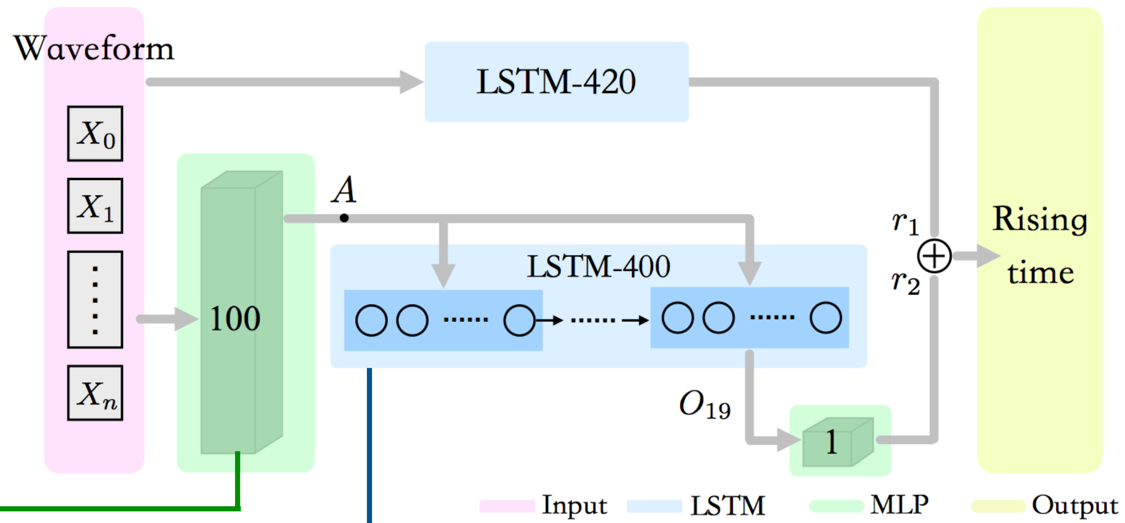
> 30 mins for training

Tsinghua University

ML with simulation

- Advanced: ComLSTM——combination of LSTM-420 and MLP
- Inputs are divided into 2 paths
  1. Go directly into LSTM-420
  2. Go into a MLP with 100 nodes, and then go into a LSTM-400
- The outputs of the 2 paths are added together with ratio r_1, r_2

$$o = r_1 \cdot o_1 + r_2 \cdot o_2$$



Expand the input feature and extract the details

Learn patterns from the magnified details

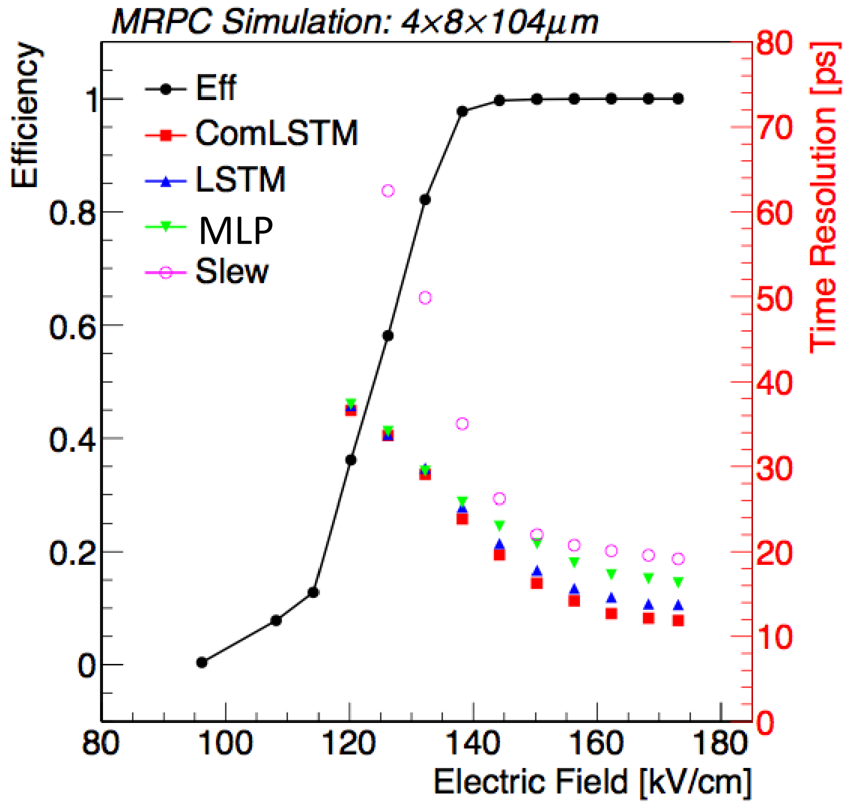
Performance is improved when NN structure is more reasonable

Simulation results

- Simulation data used for both training and testing

Leading edge (simulation) → Neural network → Rising time (simulation)

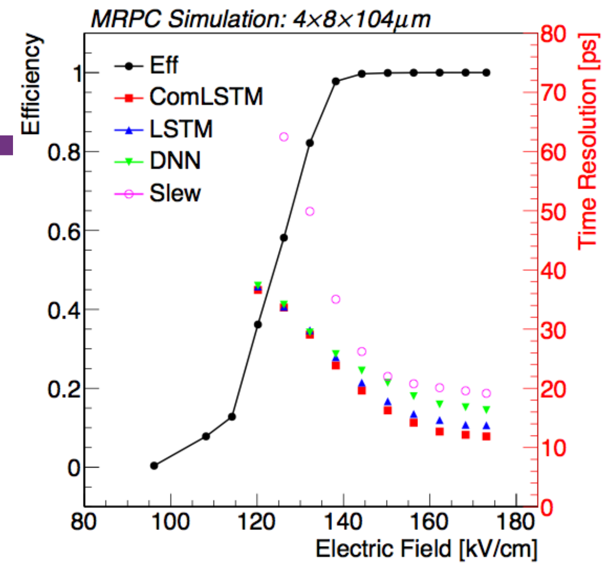
- Input: 10 sampling points around leading edge
- Output: rising time t_r
- Find the time when waveform reaches the peak t_p
- $t_0 = t_p - t_r$: first interaction time
- Time resolution: $\sigma(t_0 - t_{0true})$
- ComLSTM:
  - ✓ the best results
  - ✓ differences become apparent at high electric field.

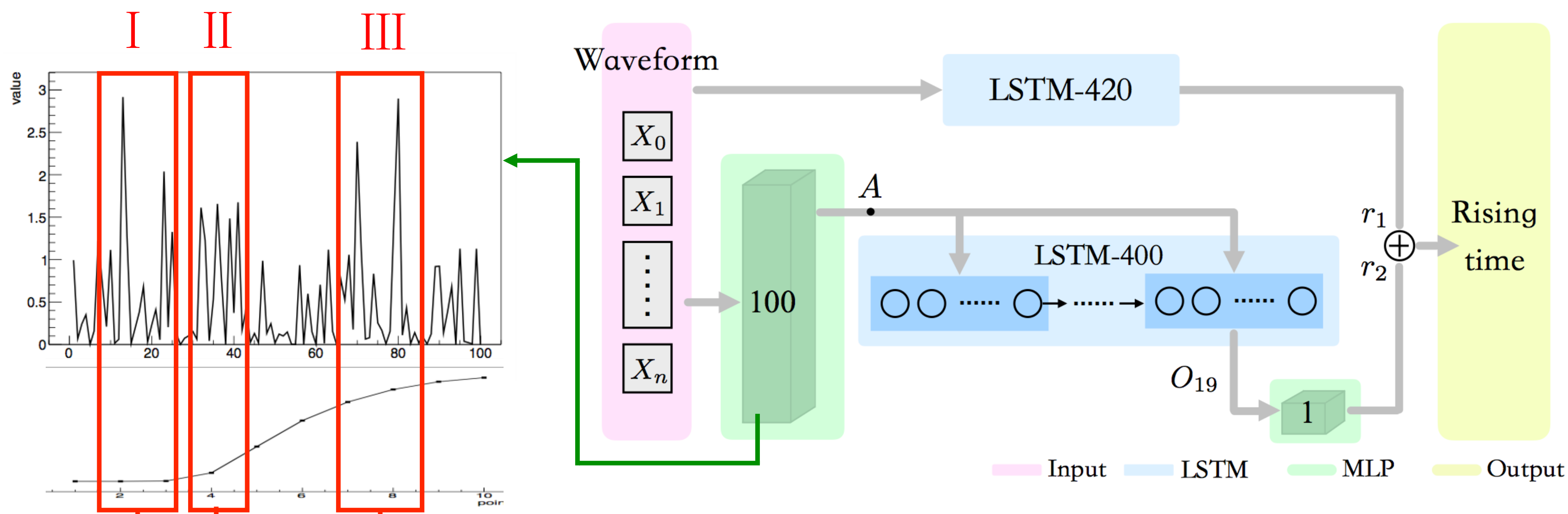


Simulation results

- Simulation data used for both training and testing

Leading edge (simulation) → Neural network → Rising time (simulation)

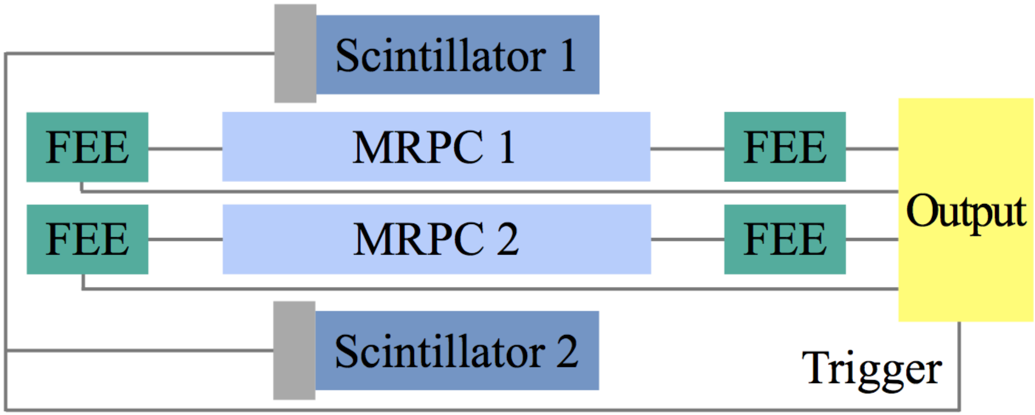


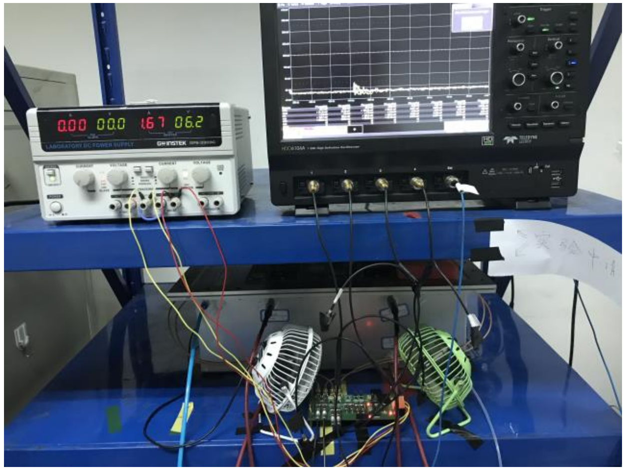


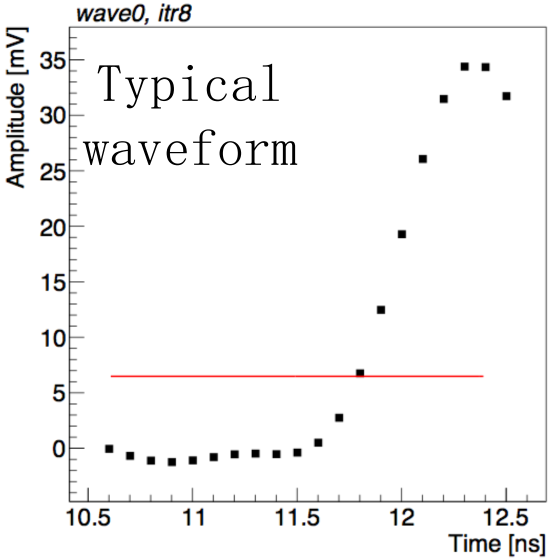
Details are magnified!

Experiment

- 2 identical MRPC: 4x8 gaps, 0.104 mm
- Cosmic ray experiment







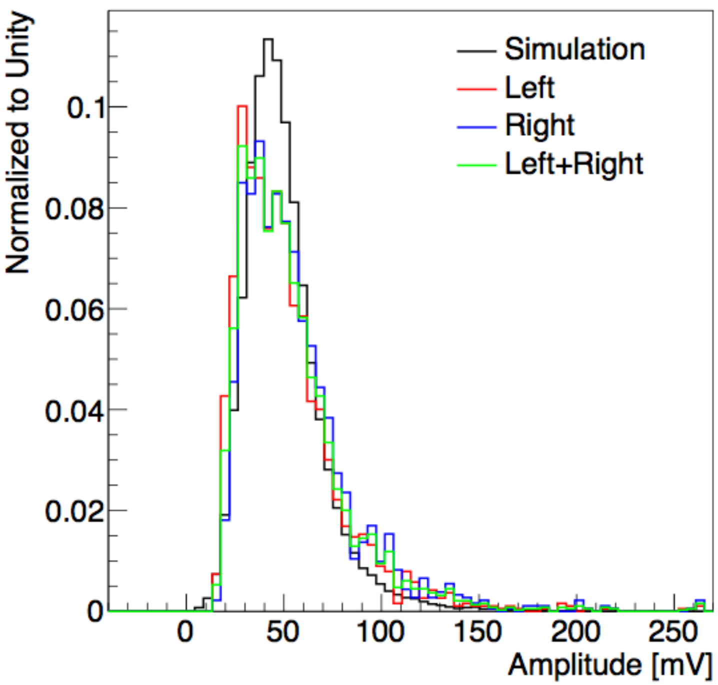
- High performance FEE from USTC: NIMA 925 (2019) 53–59
- Waveform digitizer: Lecroy HDO6104A Oscilloscope, 1GHz bandwidth, 10Gs/s sampling rate
- MRPC waveform rising time around 1 ns, 10 points

Experiment results

- Simulation data used for training, experiment data for testing

Training	simulation			simulation
	Leading edge	→ Neural network →		Rising time
Testing	experiment			experiment

- Success of the algorithm: simulation data match experiment data!



Histogram of peak point value in simulation & experiment data

- KL divergence: difference of 2 distributions

$$D(p,q) = \sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)}$$

True: experiment

Esti.: simulation

- Information lost when approximate p(x) with q(x)
- Smaller D(p,q): similar p(x) and q(x)

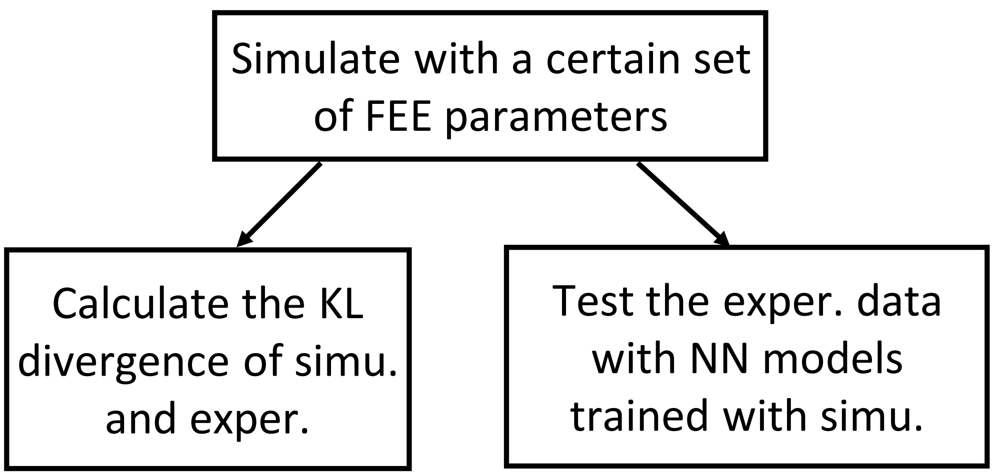
KL divergence

- Evaluate how similar the simulation data are with the experiment:

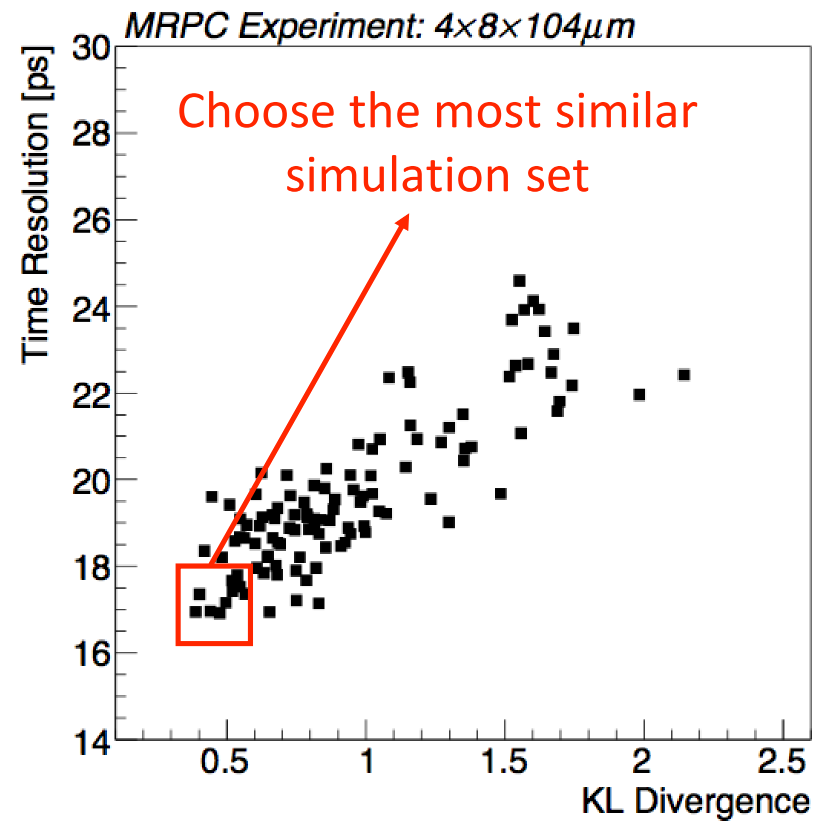
$$D_s = \frac{1}{n}\sum_{i=0}^{n} D_i(p,q)$$

i: ID of point along the leading edge

- Take the average of KL divergence for every point along the leading edge

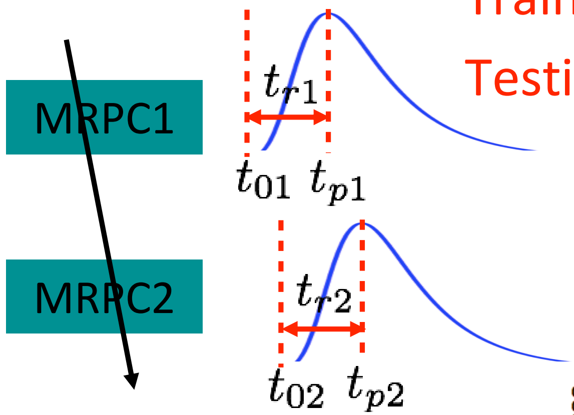


D_{si} $\sigma(t)$



Experiment results

- Simulation data used for training, experiment data for testing



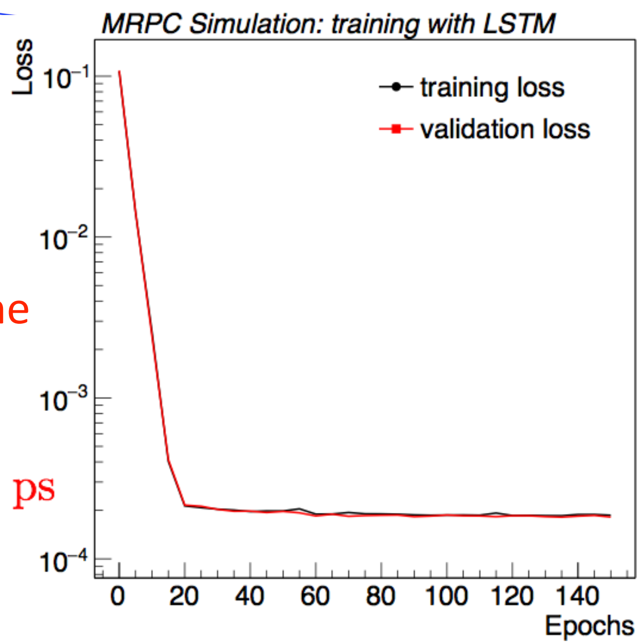
Training/evaluating: 60,000/10,000 simulation events

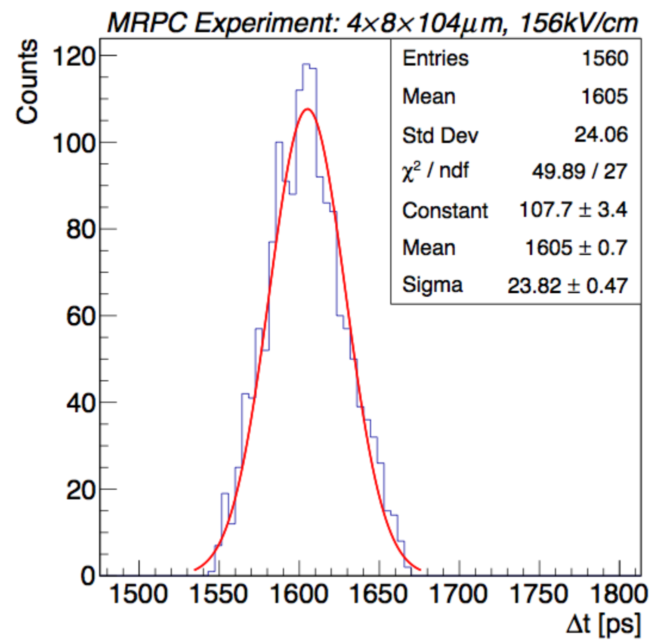
Testing: Leading edge (experiment) → Neural network → Rising time t_{r1} t_{r2} (experiment)

- @156 kV/cm, the network converges quickly

1. Find t_{p1}, t_{p2}
2. t_{r1}, t_{r2} from NN
3. Get t_{01}, t_{02} ← MRPC time
4. Define: $\Delta t = t_{01} - t_{02}$

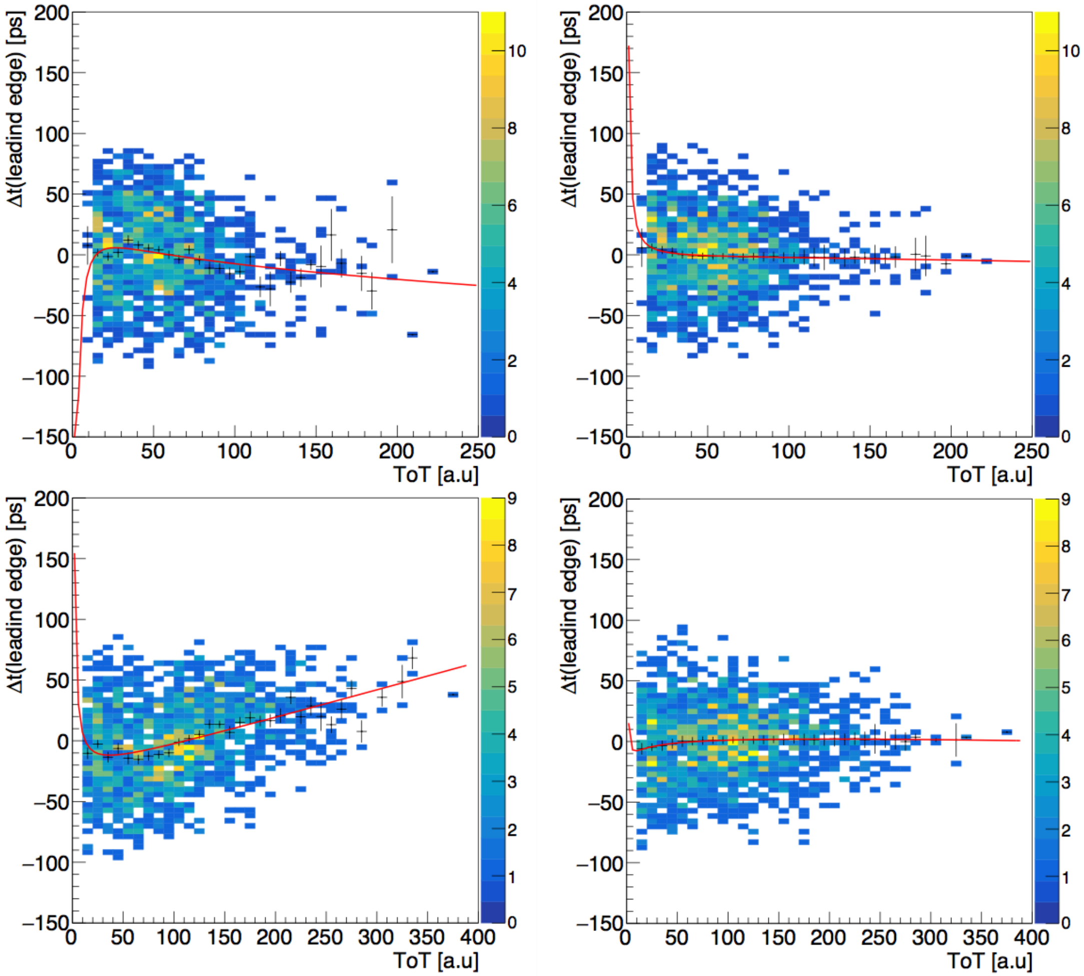
$$\sigma = \frac{\sigma(\Delta t)}{\sqrt{2}} = \frac{23.82}{\sqrt{2}} = 16.8 \text{ ps}$$

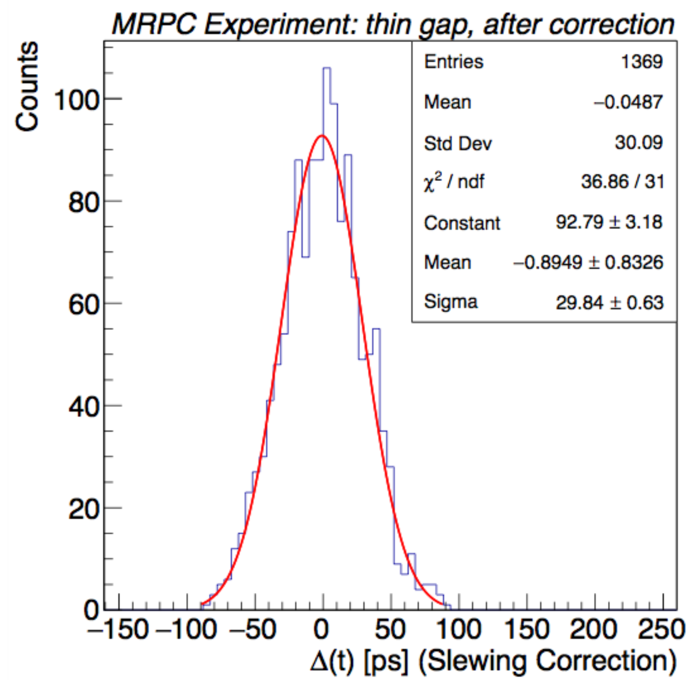




Experiment results

- ToT method: discriminate the threshold crossing time, and correct it with waveform peak.

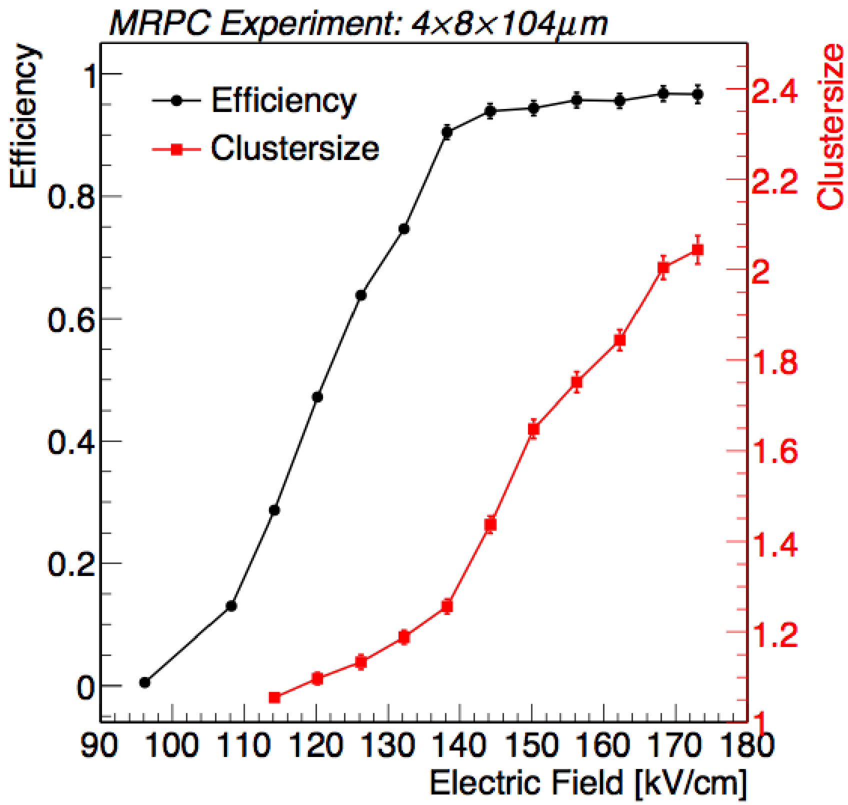


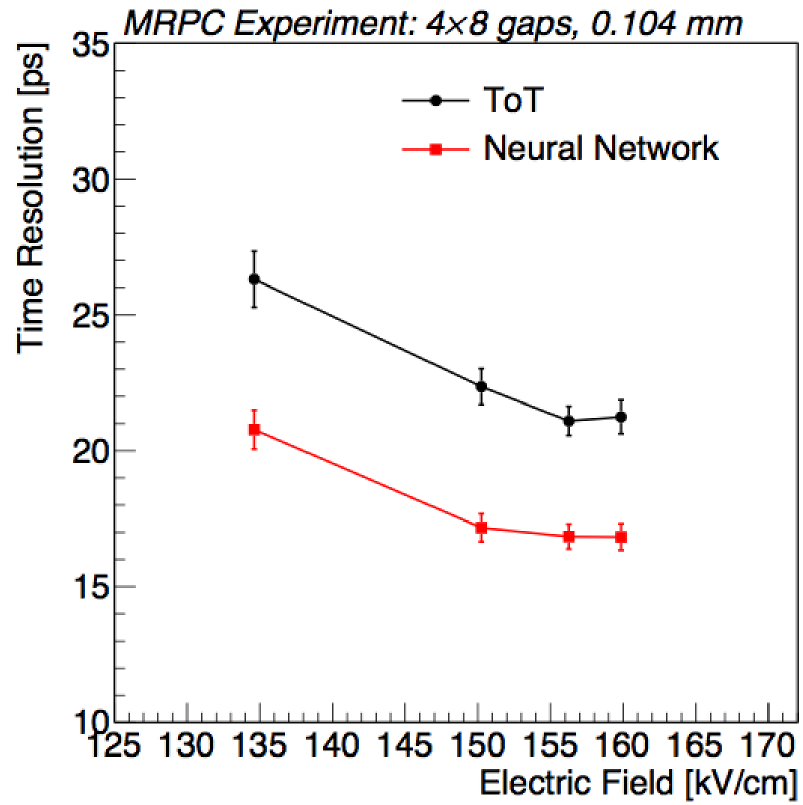


$$\sigma = \frac{\sigma(\Delta t)}{\sqrt{2}} = \frac{29.84}{\sqrt{2}} = 21.1 \text{ ps}$$

Experiment results

- Electric field (E) was scanned in the experiment
- Neural network is better than ToT method no matter what E is





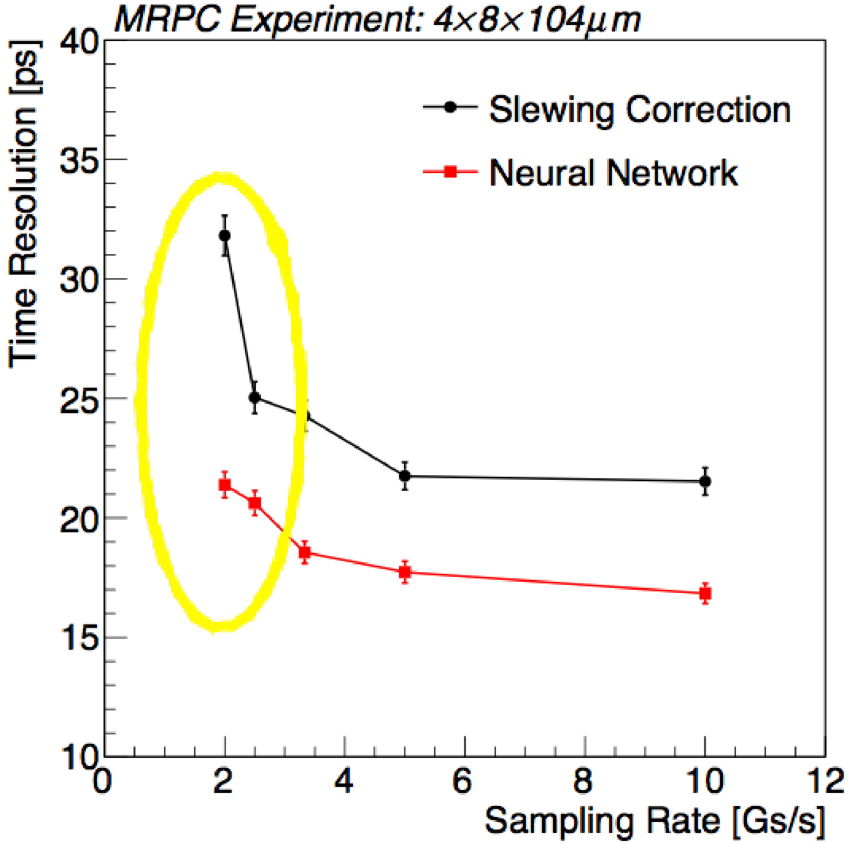
Experiment results

Tsinghua University

- Time resolution VS sampling rate
- Lecroy HDO6104A Oscilloscope: 10Gs/s sampling rate
- Down sampling manually to 5G, 3.33G, 2.5G and 2G
- Time resolutions are given by both tot and comLSTM
- In the real application, waveform digitizer is designed to be 5G

The resolution improves much at low sampling rate, while keeps stable at high rate!

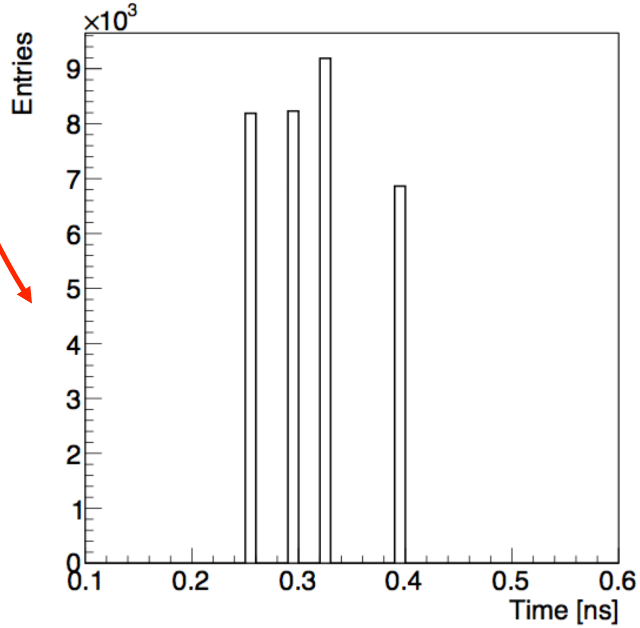
Differences between 2 methods are huge when sampling rate is low !

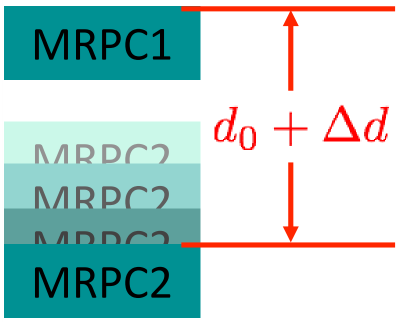


Experiment results

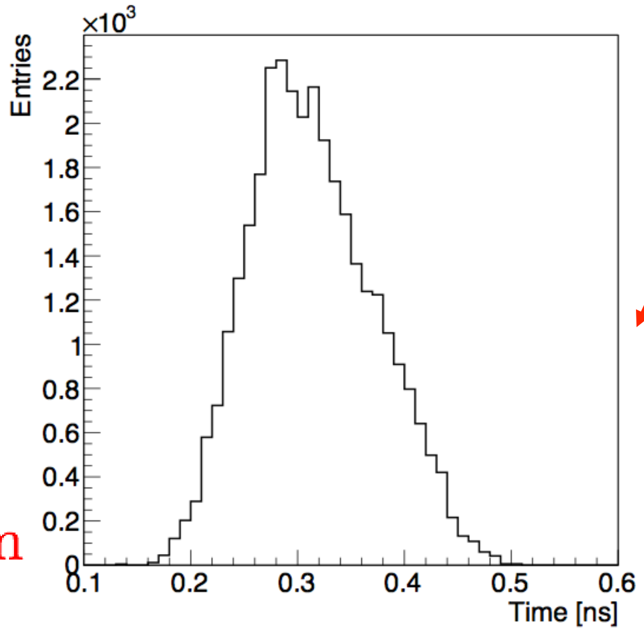
- Train and test the ComLSTM network with experiment data

	Train Input	Test input	Output
1	Exper. waveform + reference time		Time difference of 2 MRPCs give by distance/c
2	Exper. waveform + reference time		Time difference given by ToT

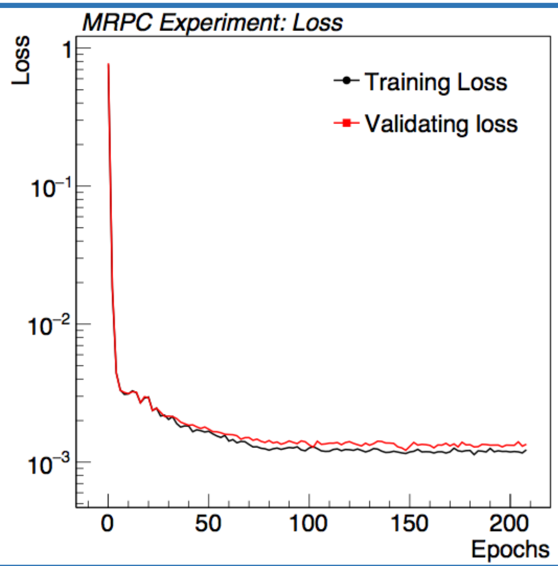


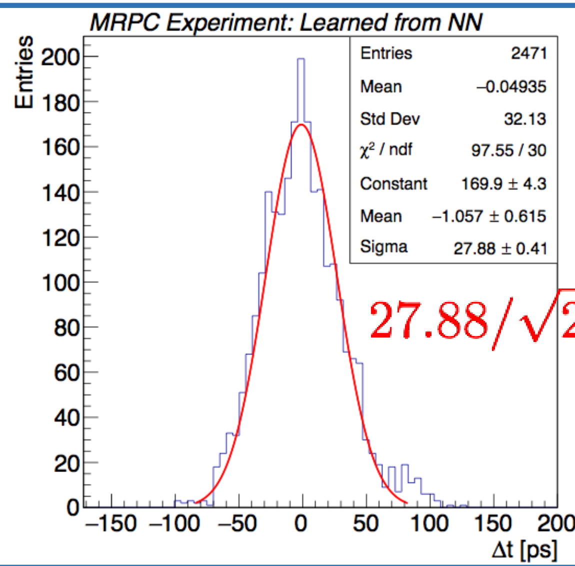


$\Delta d = 0, 1, 2, 4$ cm



Experiment results

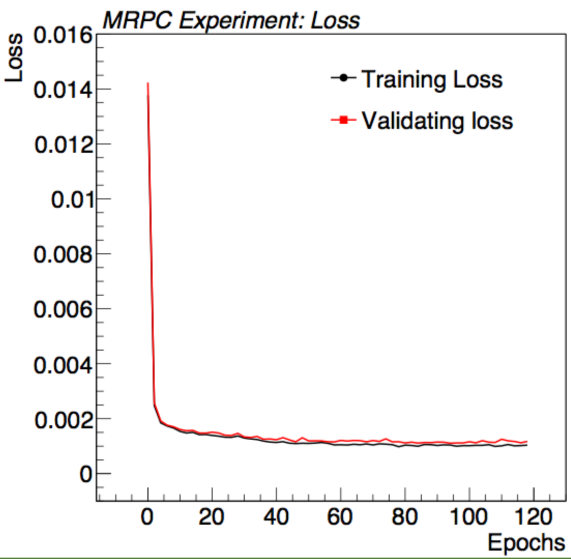


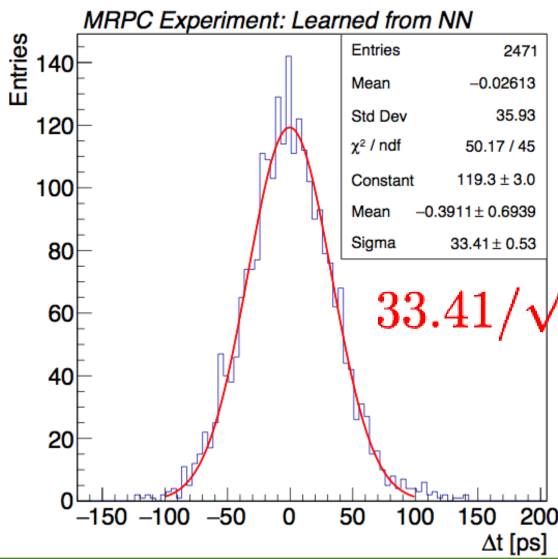


$$27.88/\sqrt{2} = 19.71\ ps$$

Time difference given by distance/c

- Network converges fast
- E=156 kV/cm





$$33.41/\sqrt{2} = 23.62\ ps$$

Time difference given by ToT+slewing correction

Results summary

	Train Input	Test input	output	Reso
1	Exper. waveform + reference time		Time difference give by distance/c	19.71 ps
2	Exper. waveform + reference time		Time difference given by ToT	23.62 ps
3	Simu. waveform	Exper. waveform	Rising time	16.84 ps

- Learn from the simulation: true information of waveform and time, therefore resolution is the best when simulation matches the experiment data.
- Learn from the ToT time: performance highly depends on the accuracy of ToT method. Time resolution is not so good as the other 2.
- Learn from the time given by distance/c: closely related when selection of perpendicular events are made. Time resolution is relatively good!

The 1st and 2nd methods prove that the learning based algorithms can be designed only with experiment data.

Summary

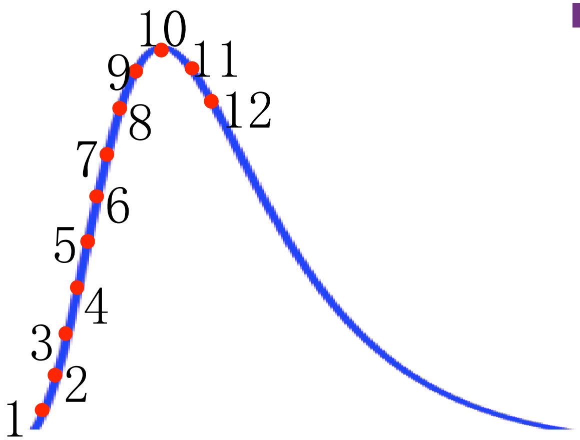
- A deep learning based algorithm has been developed to reconstruct the time of the MRPC detectors
- Different kinds of neural networks, different structures of the networks, and different logic of the algorithms are designed.
- The best time resolution of the MRPC detector with 4x8 gaps (0.104 mm thick) achieves 16.84 ps with ComLSTM network.
- The networks are also trained with the experiment data and a resolution of 19.71 ps is achieved.
- It is really hopeful for the success of sub-20 ps MRPC detector, and hopeful for the implementation of neural networks in analyzing the time detected by MRPCs.

Thank you

Backup

Data augmentation

- If experiment data are used to train the network, it needs to be augmented, because the volume of the data collected is not enough.



- Data augmentation, for every waveform one can choose the following to be network features:
  1. PointID 0-9
  2. PointID 1-10
  3. PointID 2-11
  4. …

Data can be augmented by 3~5 times !