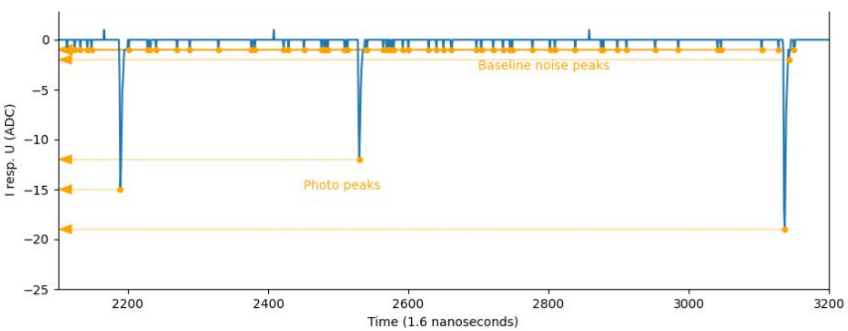
Flux calculation from PMT using Deep Learning

ERLANGEN CENTRE
FOR ASTROPARTICLE
PHYSICS

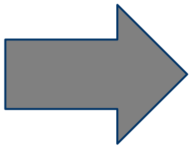
Jigar, Bhanderi
Dr. Dmitry Malyshev
Erlangen, 03.02.2022

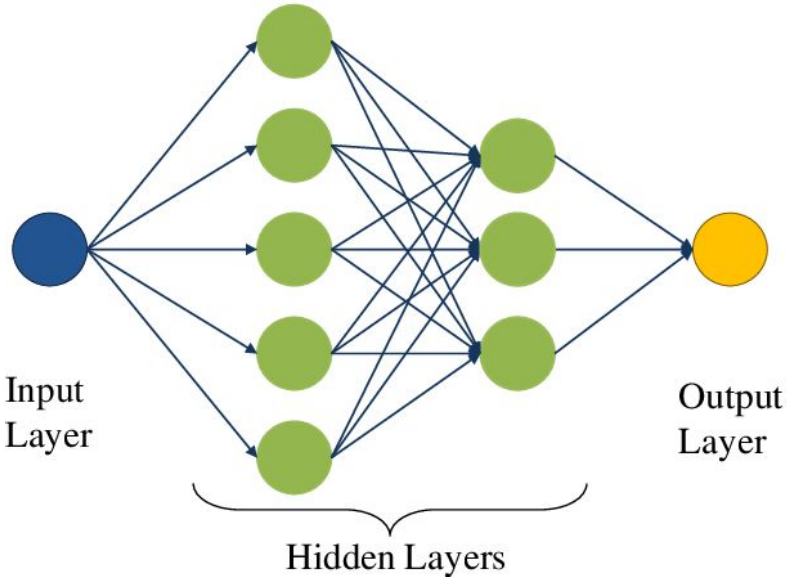
ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Motivation



Sample with photons and noise





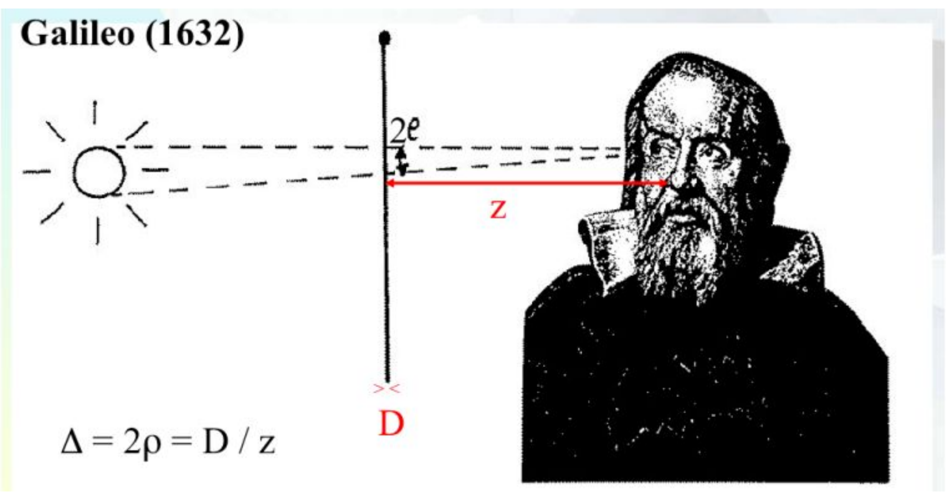
Deep Neural Network

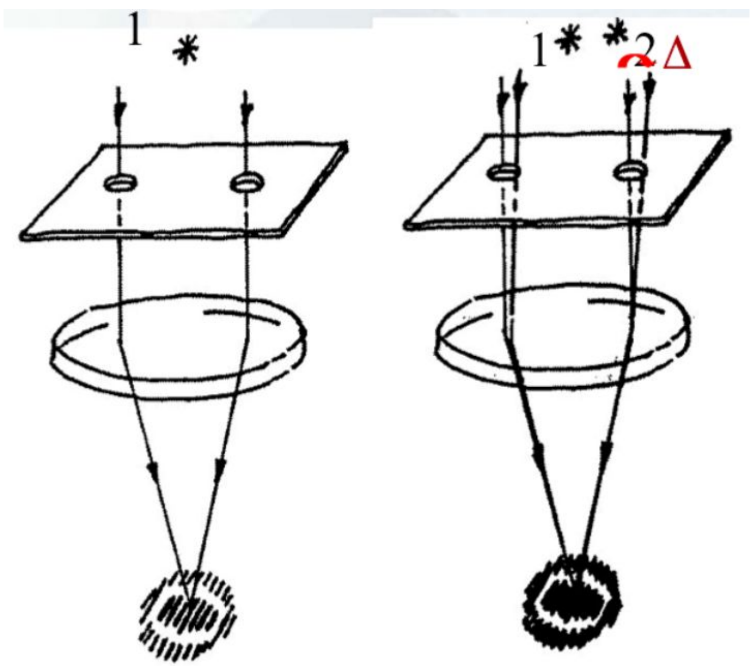
Outline

- Introduction
    - Intensity Interferometry
    - Deep Learning
- Experimental setup and data preprocessing
- Network Architectures
- Network Results

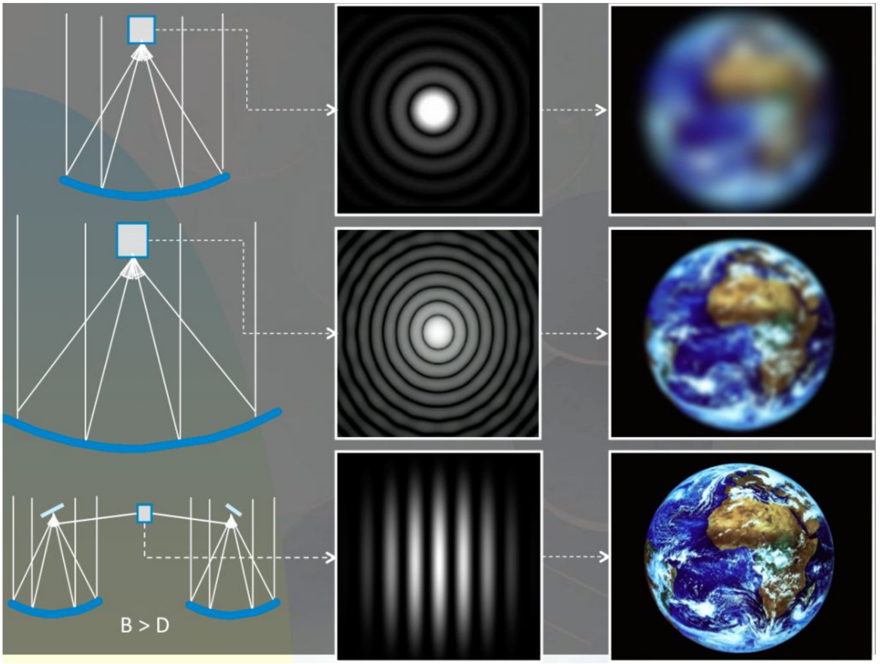
Introduction

Amplitude Interferometry





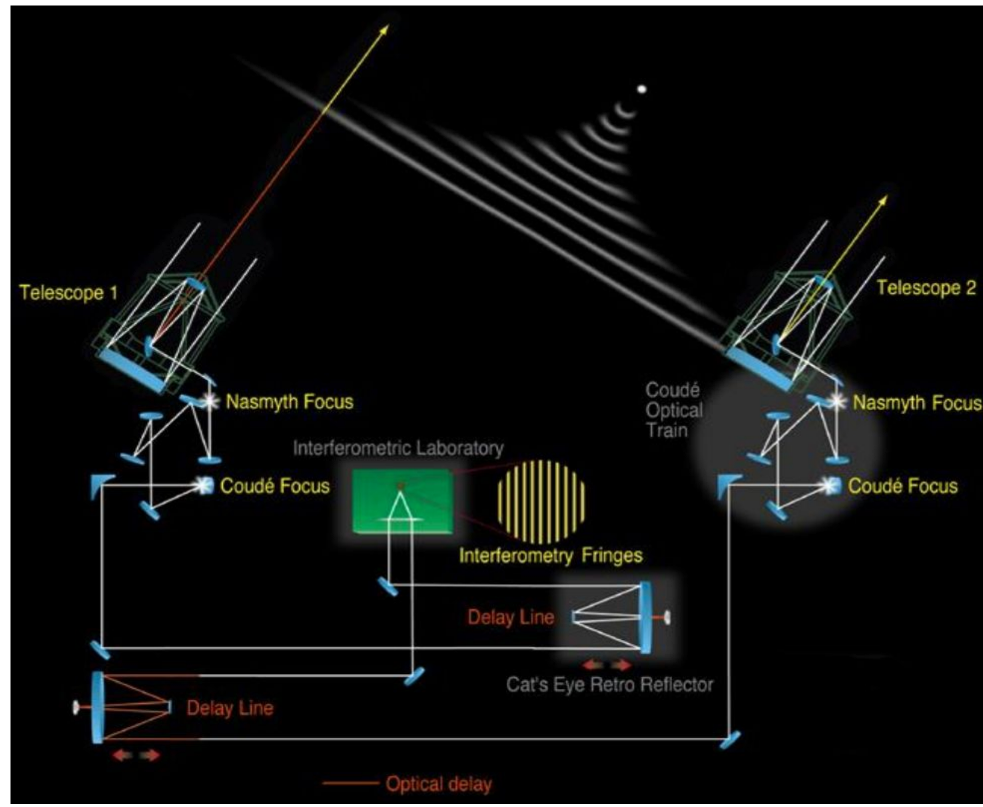
Amplitude Interferometry



Introduction to Optical/IR Interferometry: history and basic principles

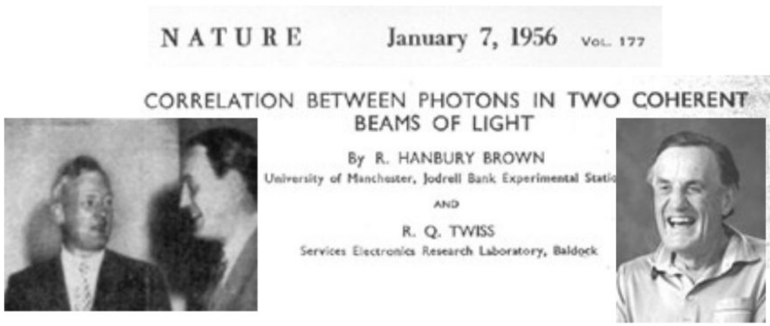
- Fringe appearance
- Increase telescope size⇒more fringes
- multiple telescopes ⇒ better resolution

ESO Very Large Telescope Interferometer



Introduction

Intensity Interferometry



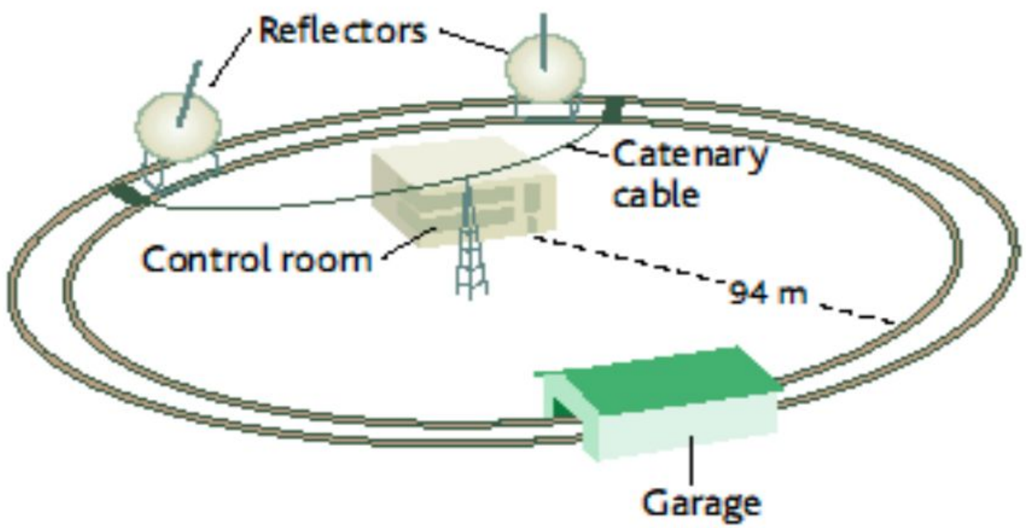
NATURE January 7, 1956 Vol. 177

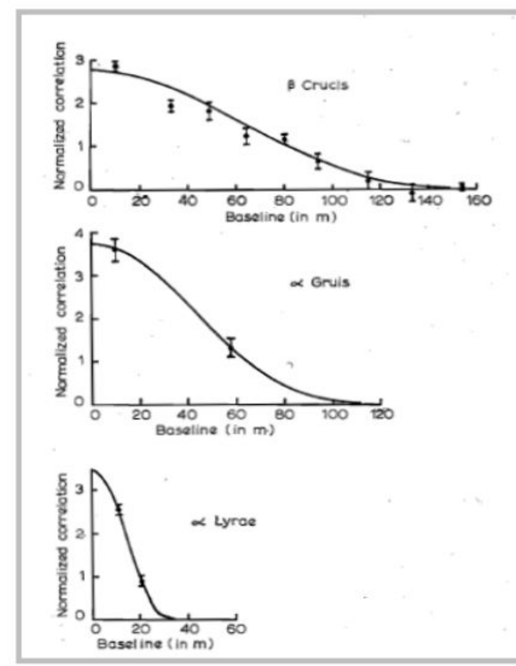
CORRELATION BETWEEN PHOTONS IN TWO COHERENT BEAMS OF LIGHT

By R. HANBURY BROWN
University of Manchester, Jodrell Bank Experimental Statio

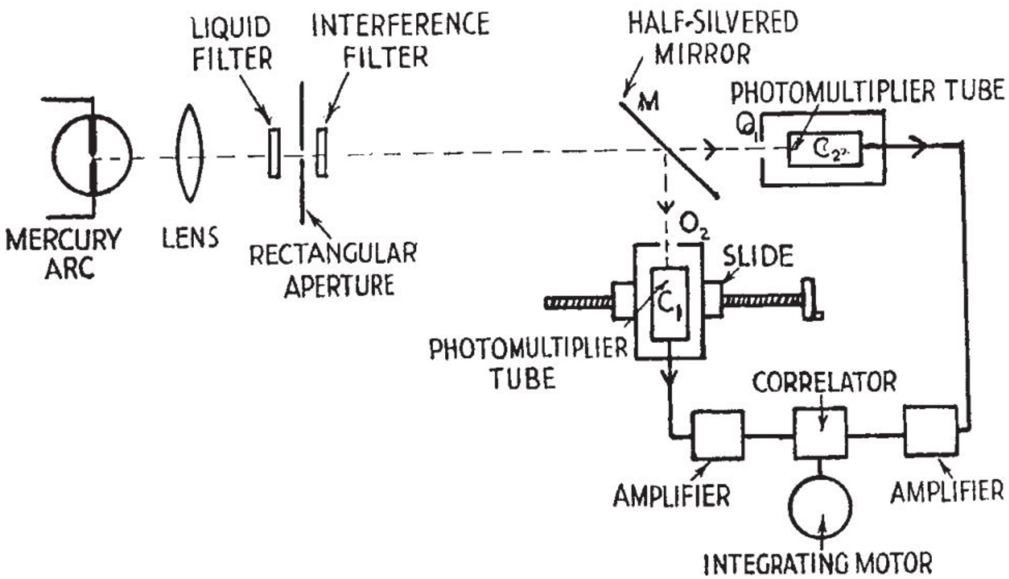
AND

R. Q. TWISS
Services Electronics Research Laboratory, Baldock





Intensity Interferometry



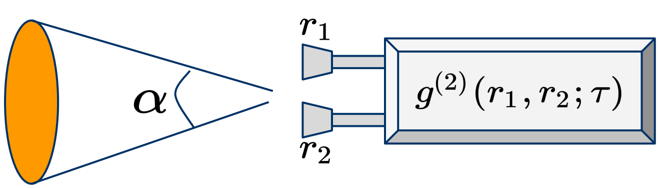
Second order correlation function

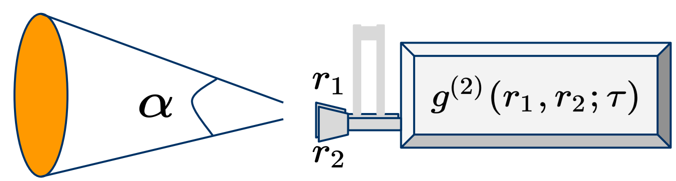
$$g^{(2)} = \frac{\langle I(r_1,t) I(r_2,t+\tau) \rangle}{\langle I(r_1,t) \rangle \langle I(r_2,t+\tau) \rangle} \Rightarrow L_c$$

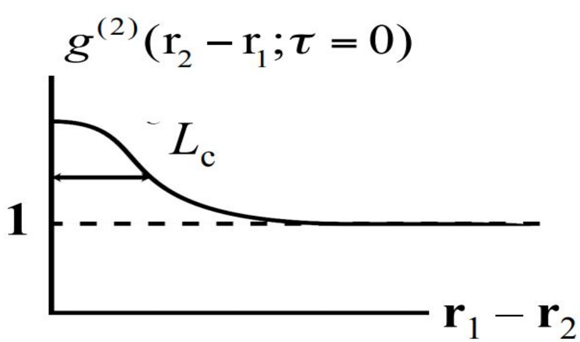
Coherence area ⟹ Angular diameter of star

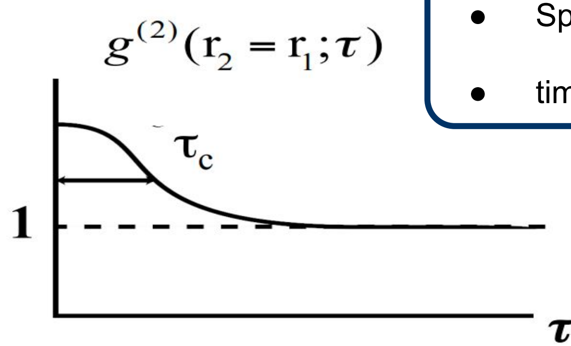
Intensity pattern in the observation plane⇒Speckle

- correlation radius $L_c \propto \frac{\lambda}{\alpha}$
- pattern change in each $\tau_c \propto \frac{1}{\Delta\omega}$









- Space coherence $L_c \propto \frac{\lambda}{\alpha}$
- time coherence $\tau_c \propto \frac{1}{\Delta\omega}$

Intensity Interferometry

Time integrated signal of temporal correlation

$$S = \int_{-t/2}^{t/2} g^{(2)}(\Delta t) d\Delta t$$

Source flux

$$\frac{S}{N} = C \cdot \epsilon \cdot n(\nu) \cdot \sqrt{\frac{T}{\tau_e}}$$

Observation time

System resolution

$$N = \frac{1}{\sqrt{R_1 R_2}} \sqrt{\frac{\tau_e}{T}}$$

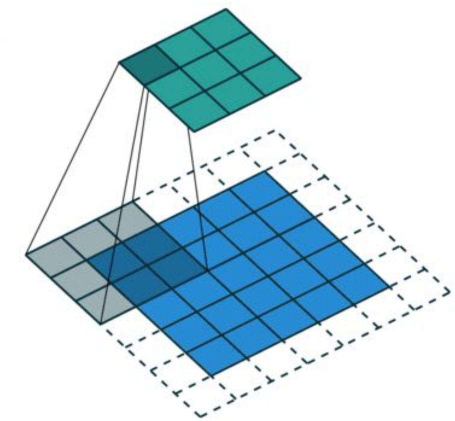
Telescope collection area

Photon detection efficiency of the system

Photon rates

LED as laboratory test source for astronomical intensity interferometry

Different type of Neural Networks



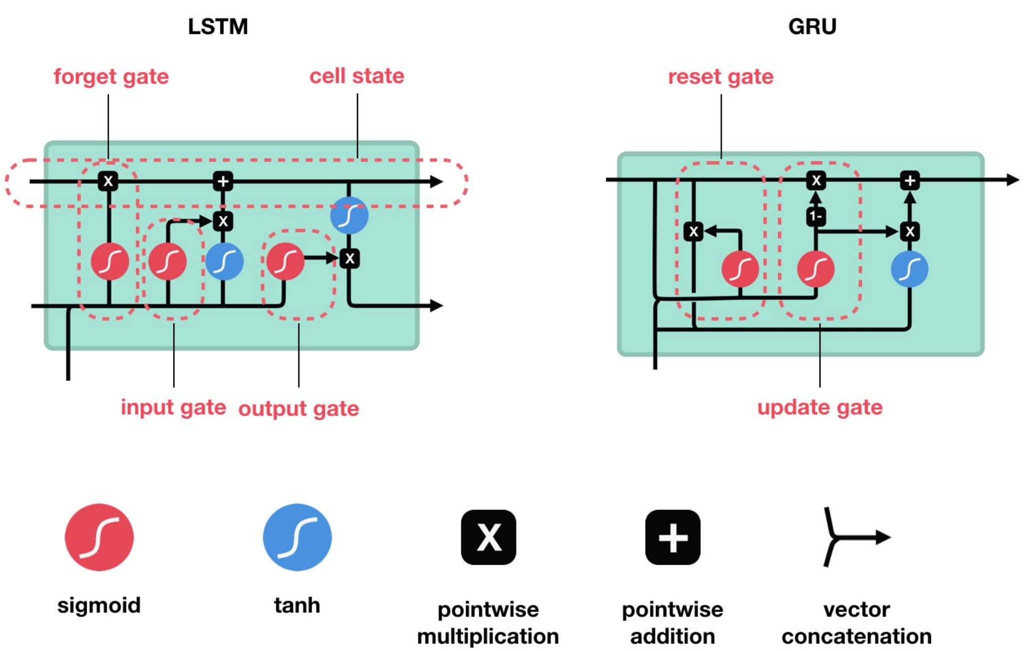
A Comprehensive Guide to Convolutional Neural Networks

0	50	0	29
0	80	31	2
33	90	0	75
0	9	0	95

80	?
?	?

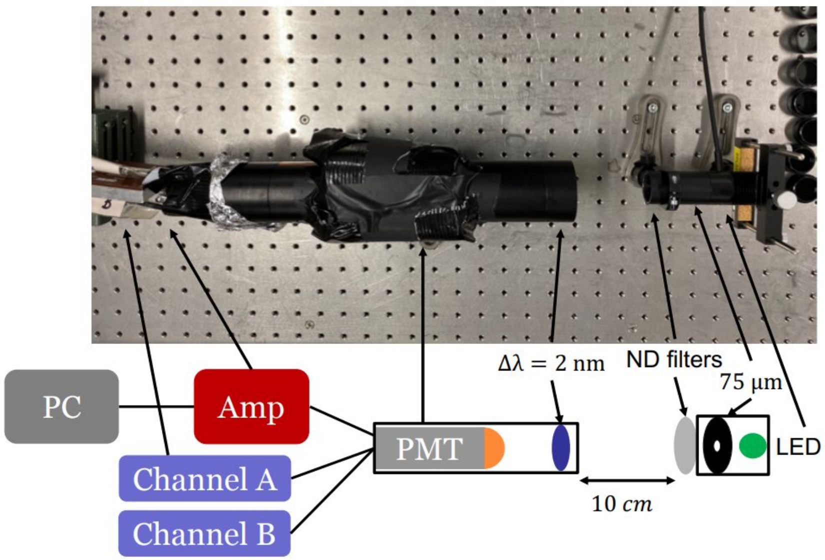
CNNs, Part 1: An Introduction to Convolutional Neural Networks

- Convolutional Neural Network (CNN)
  - Weight sharing(use same weights all over)
  - Hierarchy of features(eyes+nose+ears ➪ face)
  - Connect pixels in a neighbourhood ➪ spatial structure
- Maxpooling
  - Summary of region
  - Less computational cost and overfitting
- LSTM and GRU
  - Hidden state
    - information storage
    - pass information to next cell
  - GRU is simpler than LSTM



Illustrated Guide to LSTM's and GRU's: A step by step explanation

Experimental Setup and data preprocessing



Calibration of Photomultiplier Tubes for Intensity Interferometry at H.E.S.S.

$$\frac{S}{N} = C \cdot \epsilon \cdot n(\nu) \cdot \sqrt{\frac{T}{\tau_e}}$$

Source flux

Observation time

Telescope collection area

Photon detection efficiency of the system

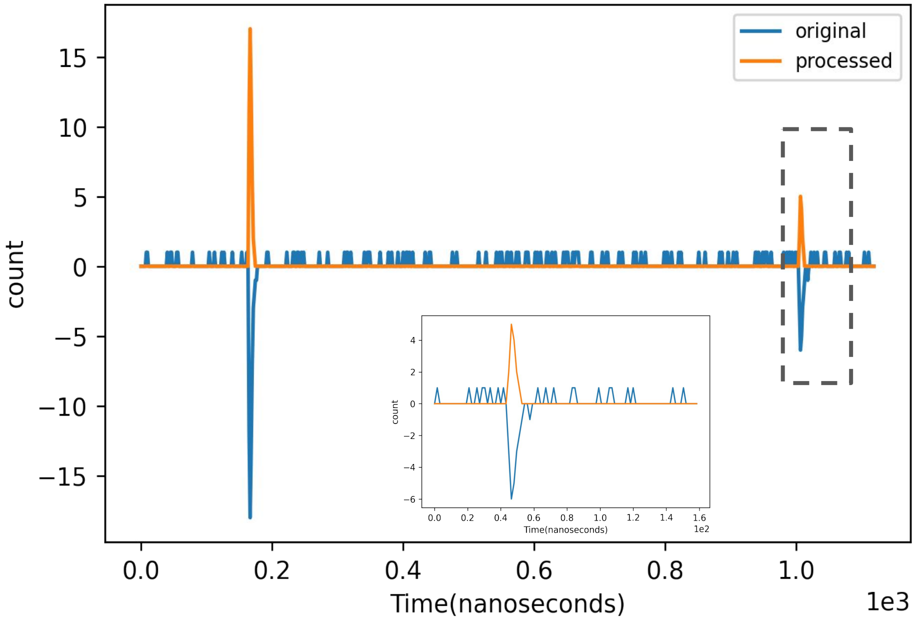
Optical bandwidth

System resolution

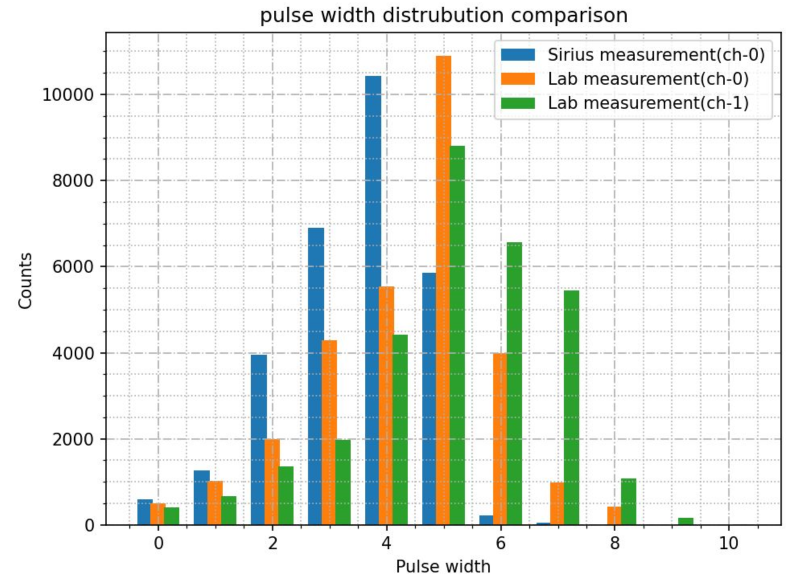
$$N = \frac{1}{\sqrt{R_1 R_2}} \sqrt{\frac{\tau_e}{T}}$$

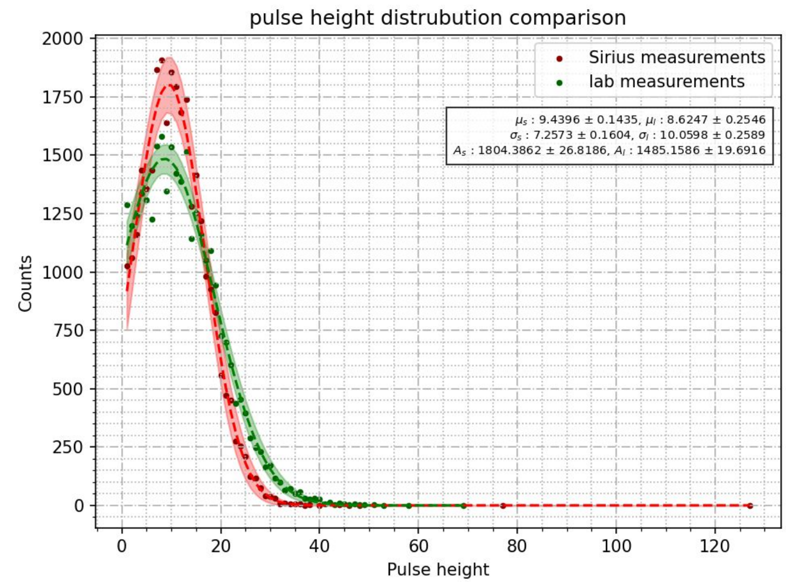
Photon rates

Data preprocessing



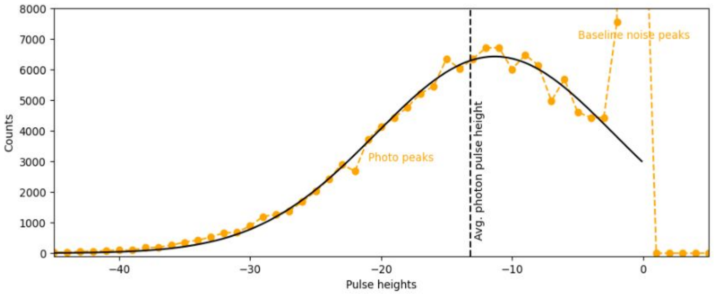
- Extraction of photon shapes after preprocessing with sample size 100
- Shape set splitting for training testing and validation by 20%
- MC-data creation for selected sample size
- Input: 1D-waveform, output: number of photon peaks in sample
- Train selected neural network
- Evaluate network(test dataset)
- Testing of network
  - Lab : waveform with different transmission
  - Sirius : measurement at different instances



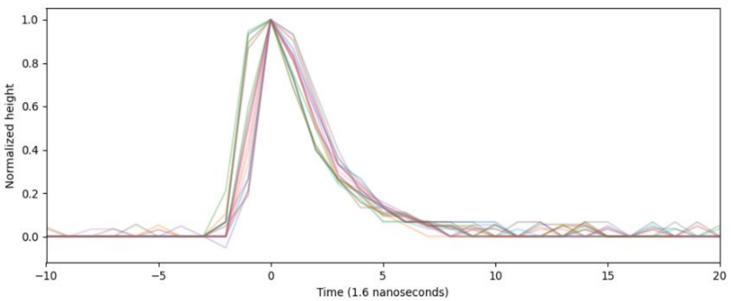


Charge Integration

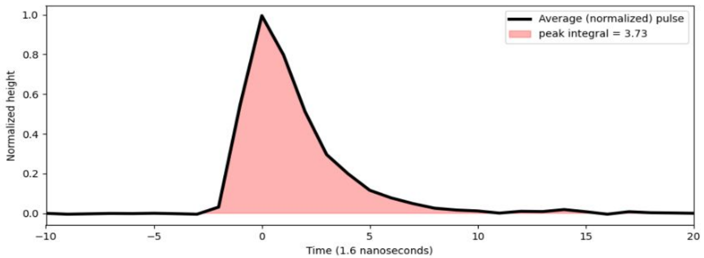
Calculate average photon pulse height



Normalize each pulse by overlaying each at same position



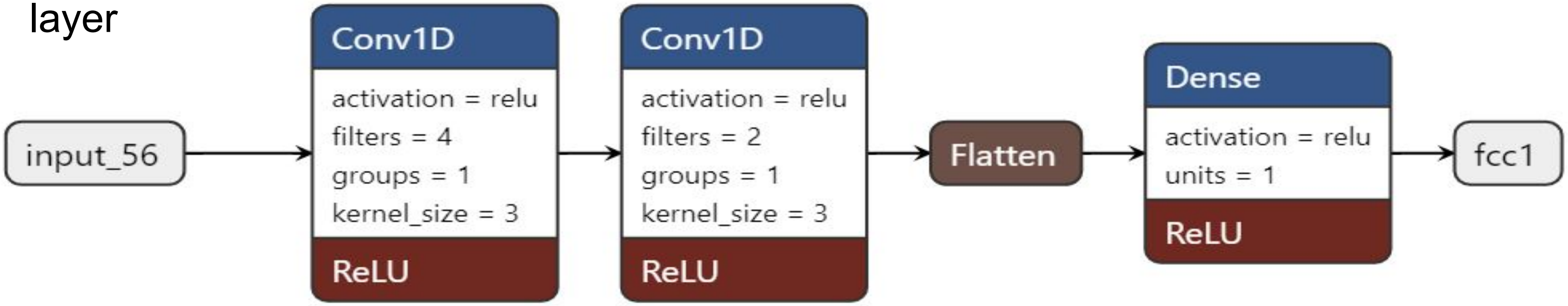
Average charge = Average pulse height × Average normalized pulse sum



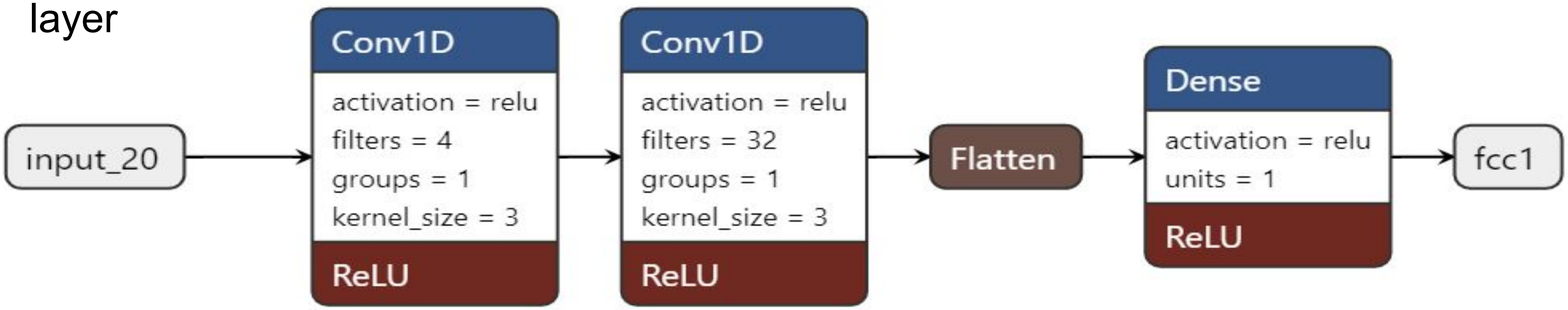
$$Rate = \frac{waveform\ mean}{Average\ Charge \cdot 1.6ns}$$

Network Architectures

Model-1: First layer(thick)⇒second layer(thin)⇒Dense layer



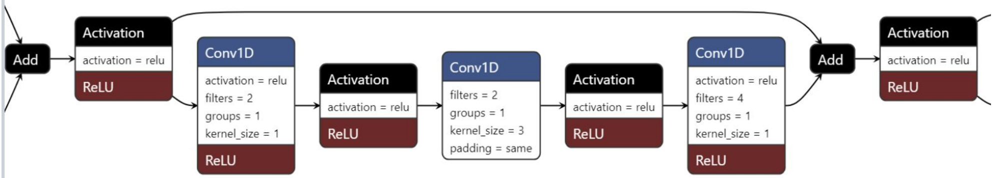
Model-2: First layer(thin)⇒second layer(thick)⇒Dense layer



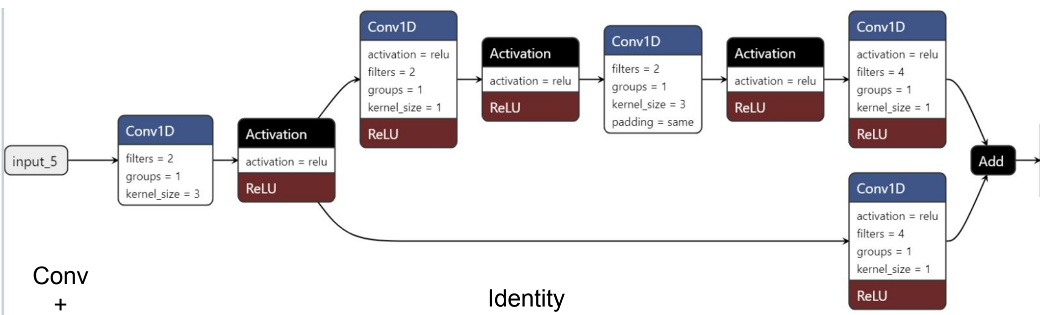
Model architectures

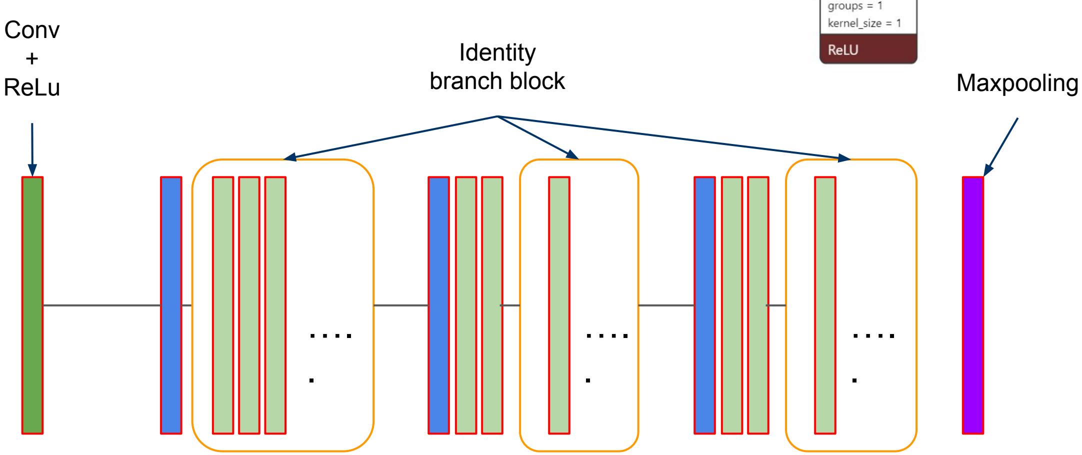
Model-3: Residual Neural Network(ResNet)

Identity branch: skip connection



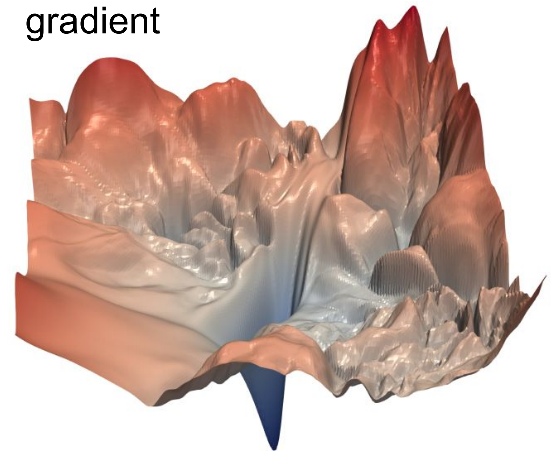
Convolution branch: modified skip connection



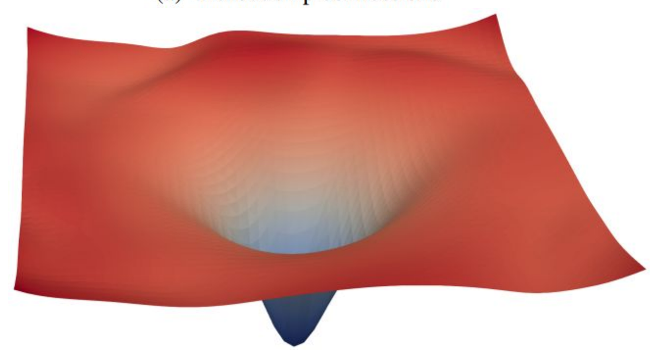


Why ResNet?

Deeper the network smaller the gradient



(a) without skip connections



(b) with skip connections

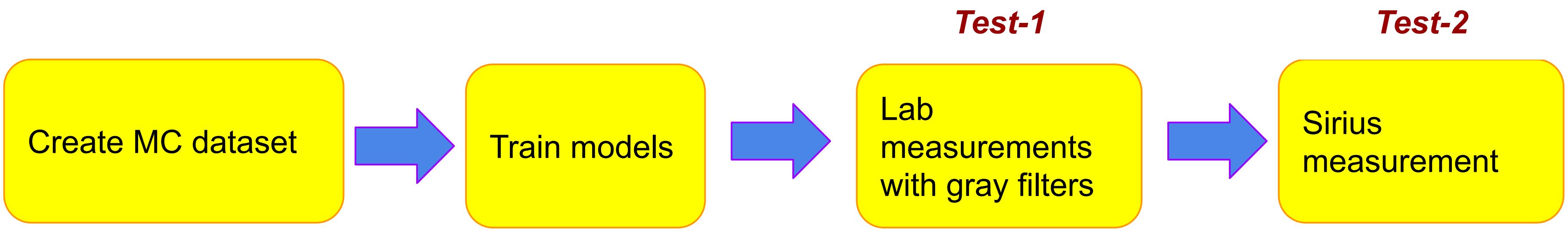
["Visualizing Loss Landscape of Neural Nets"]

Model complexity

- Weight sharing in CNN makes it less complex, even with ResNet architecture, but gated computations makes LSTMs and GRUs more complex
- More flops(Floating point operations)⇒computationally costly training

Network Results

General process

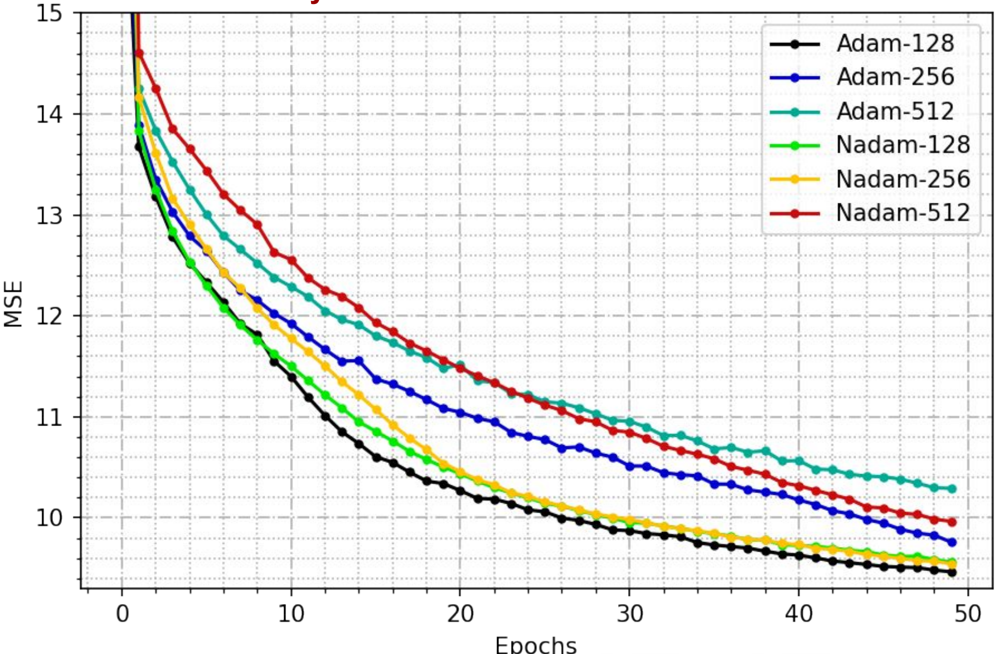


Hyper Parameters

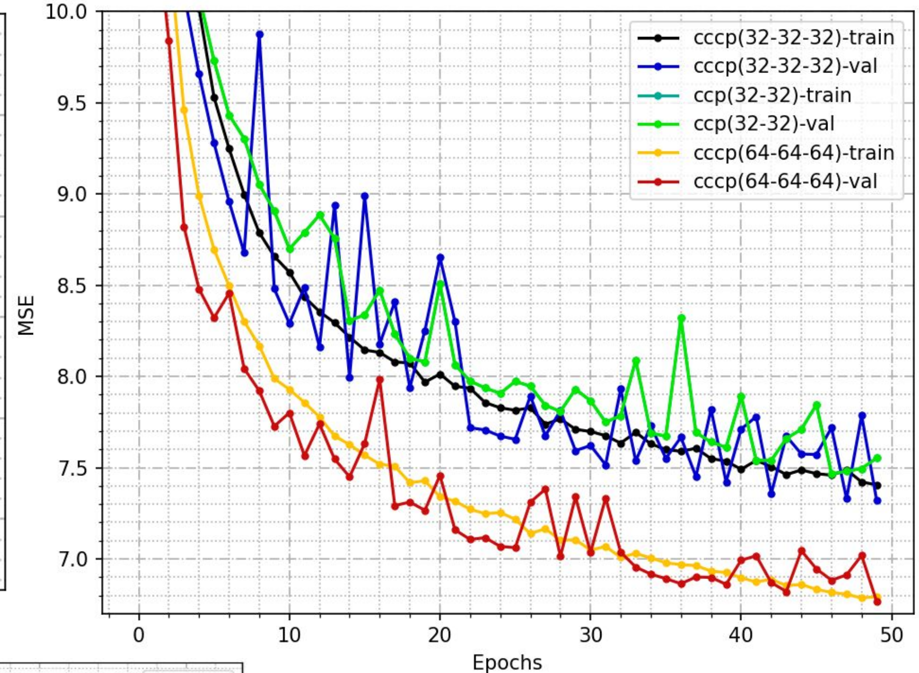
- Number of layers
- Layer thickness
- Learning rate
- Batch size

Loss comparison

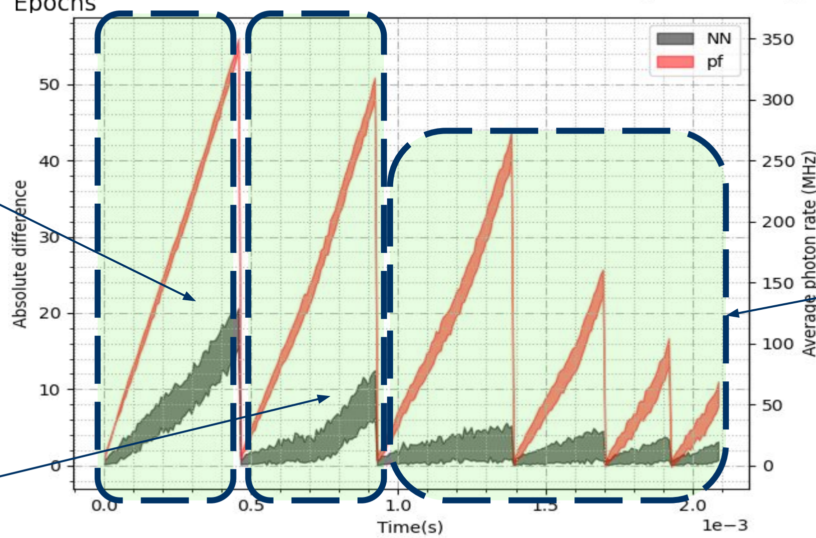
Fully Connected Neural Network



Convolutional Neural Network



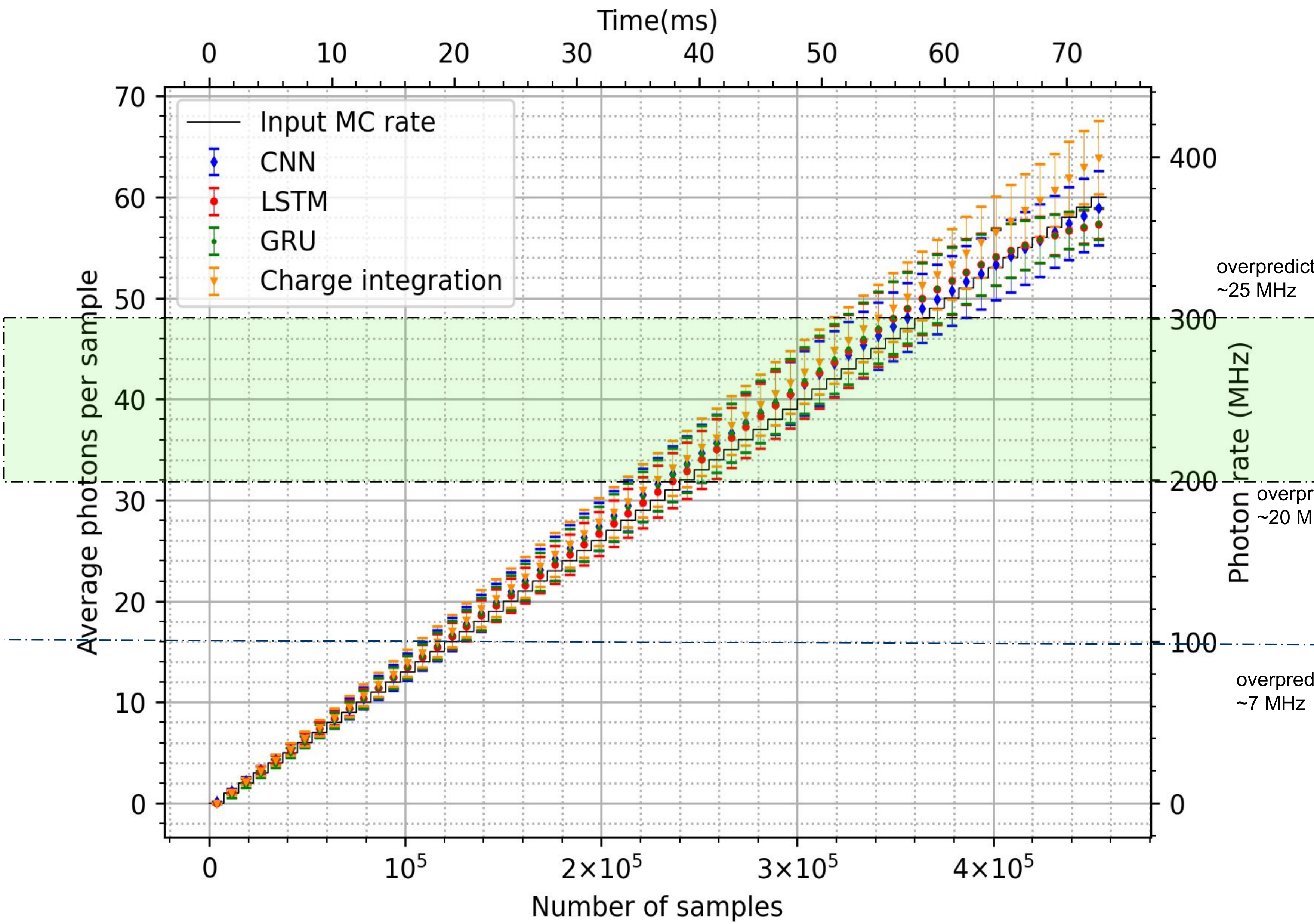
Better prediction by factor ~10



Very small improvement in predictions

Better prediction by factor ~5

Model Evaluation



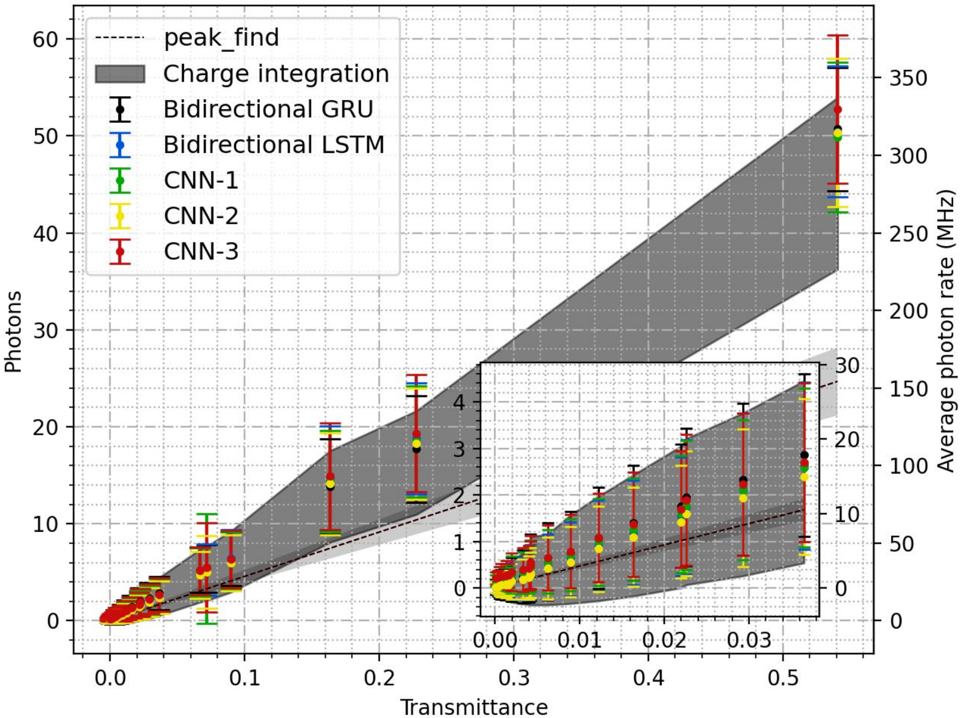
overprediction by ~25 MHz

overprediction by ~20 MHz

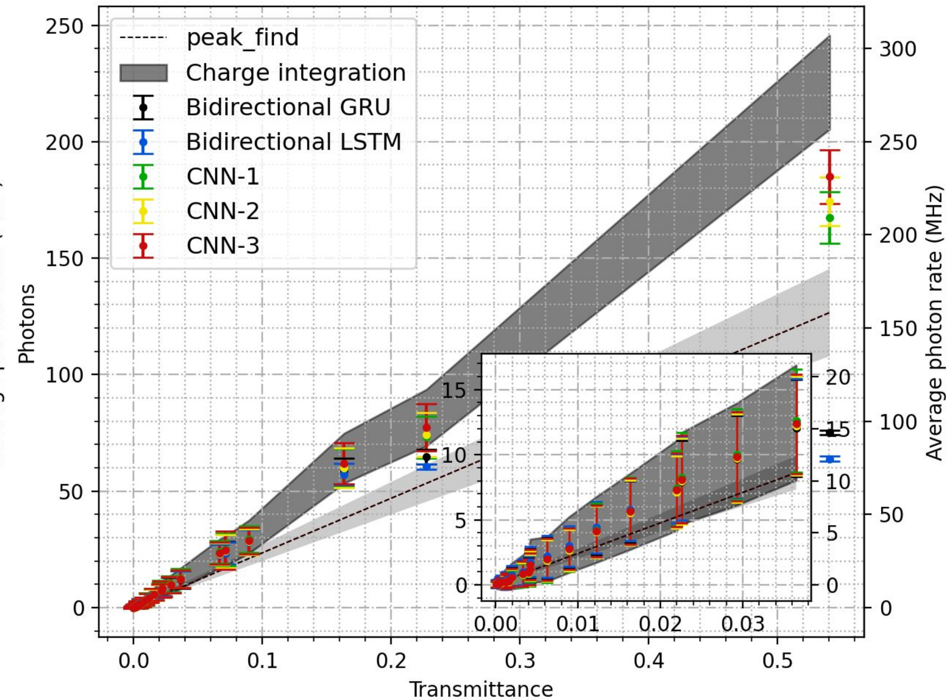
overprediction by ~7 MHz

Prediction comparison (Lab)

Sample size 100



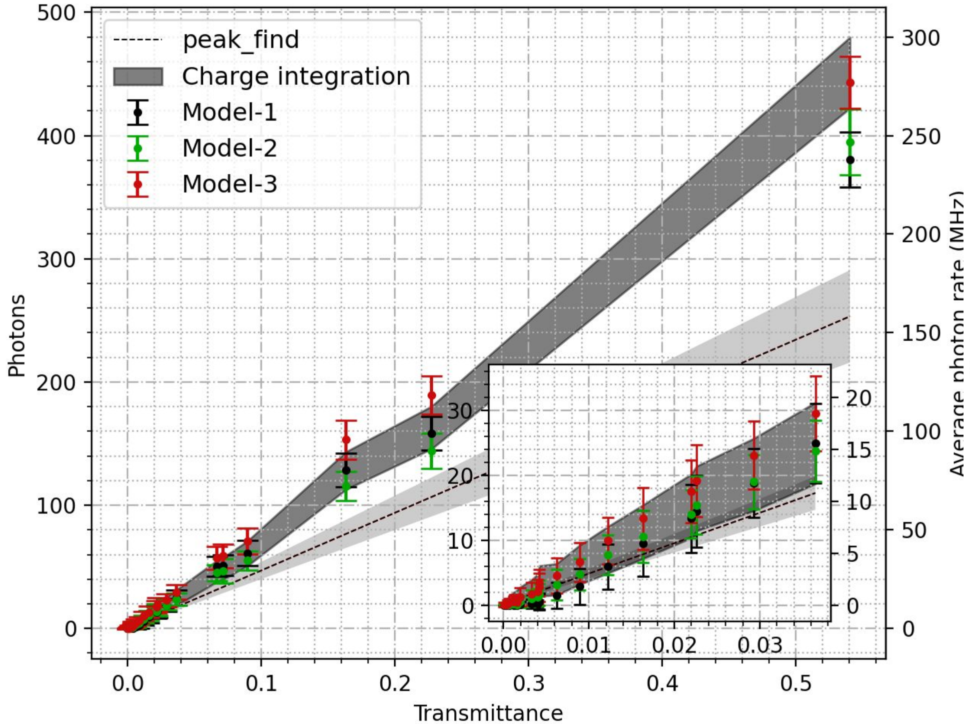
Sample size 500

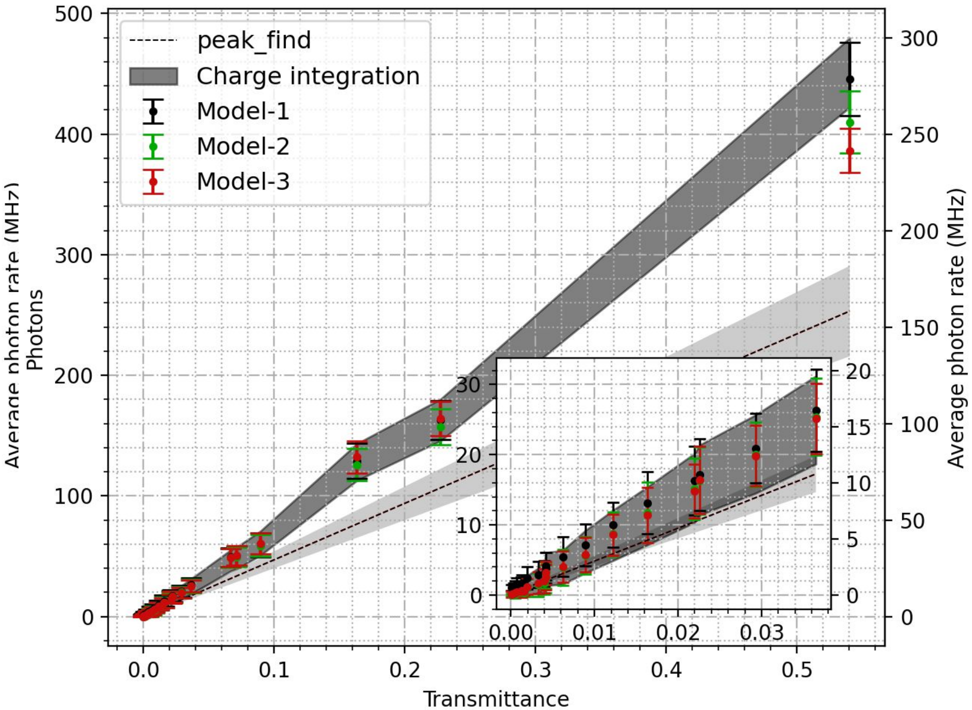


Different points correspond to measurements with different grey filters, and corresponding transmittances are plotted on the x-axis

Prediction comparison(Lab)

Sample size 1000





- Learning rate : 0.0001
- Batch size : 128

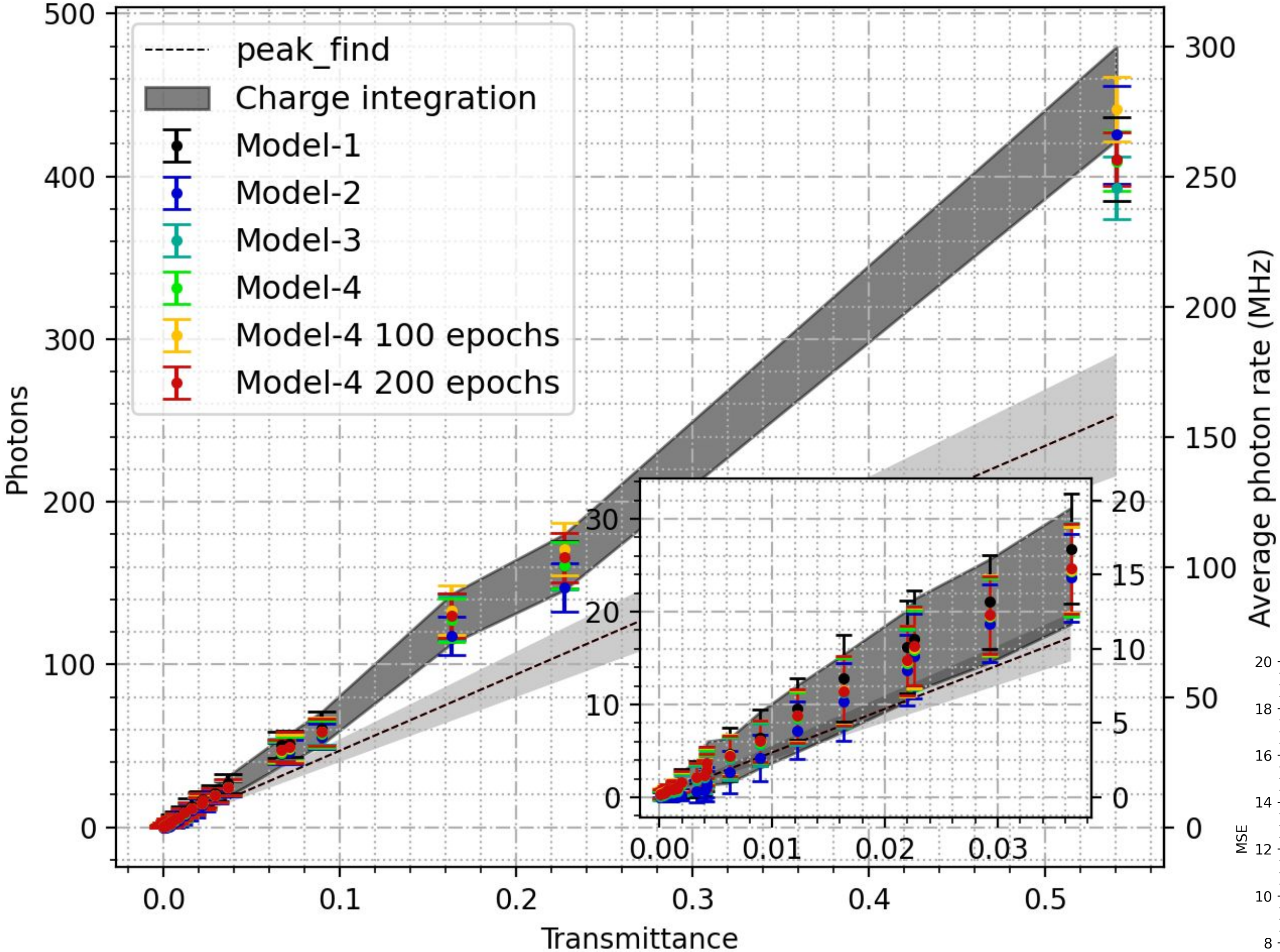
- Learning rate : 0.0001
- Batch size : 512

All model predictions are different from each other even at smaller rate

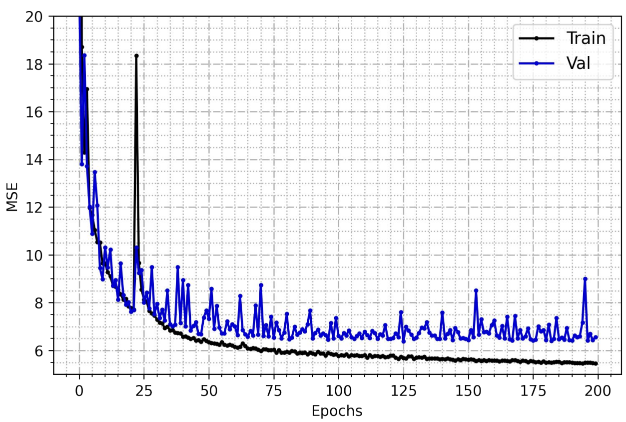
At very lower rate all three models are in agreement with Charge Integration method, but they predict differently as rate increases

Prediction comparison(Lab)

Sample size 1000



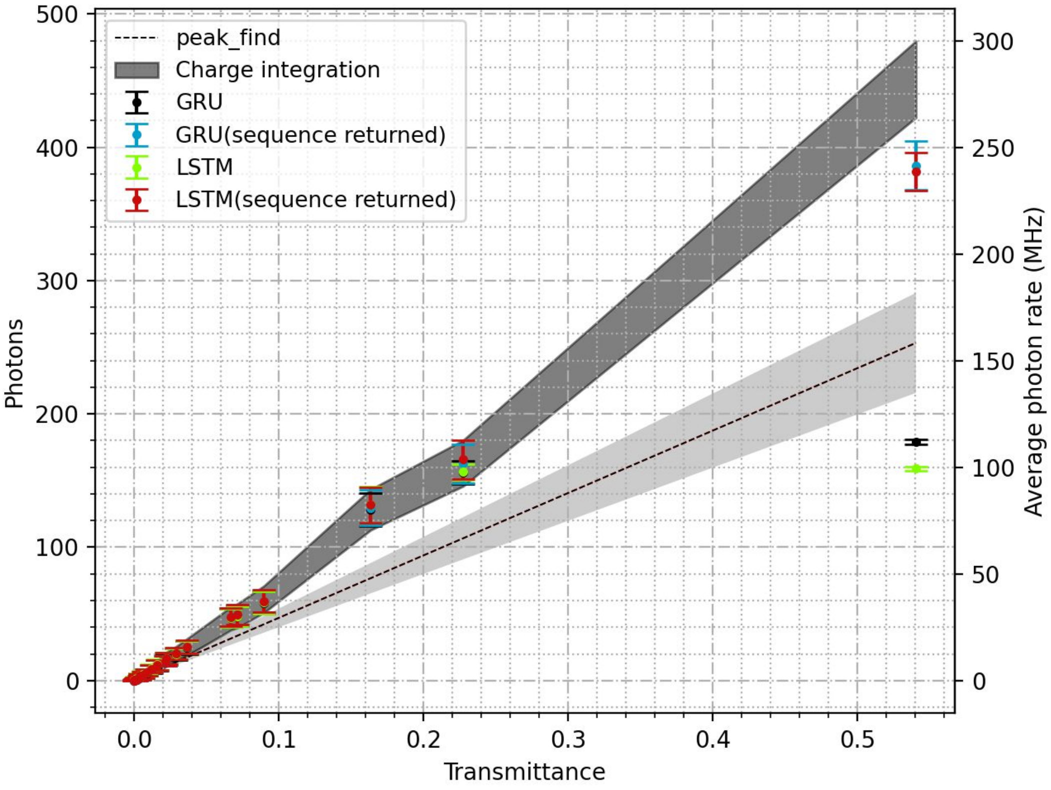
- Model-4: 5 times deeper than model-3
- Learning rate : 0.001
- Batch size : 128

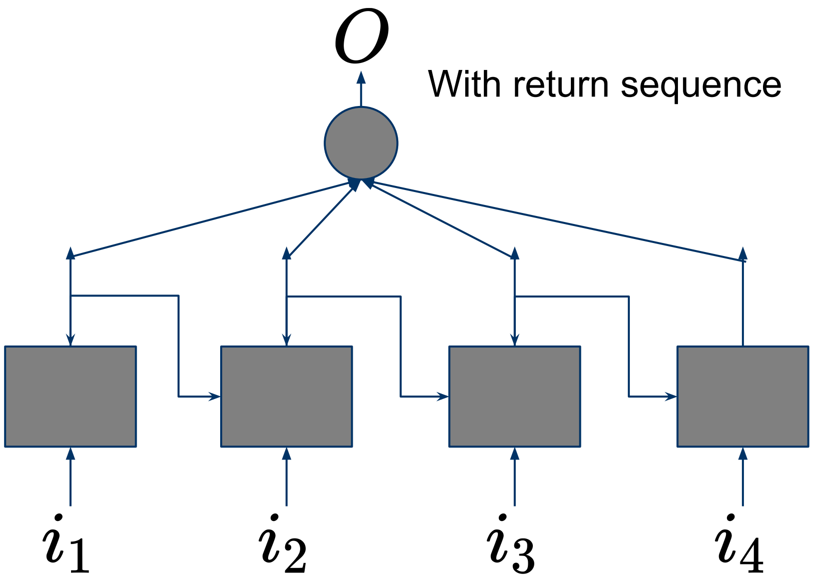


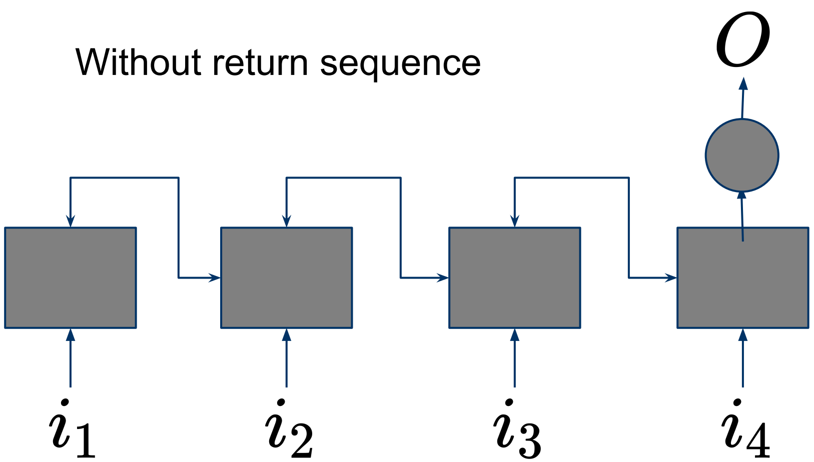
Overfitting problem occurs when training duration is longer and deeper models are used

Prediction comparison(Lab)

Prediction comparison between two different configurations of LSTM and GRU

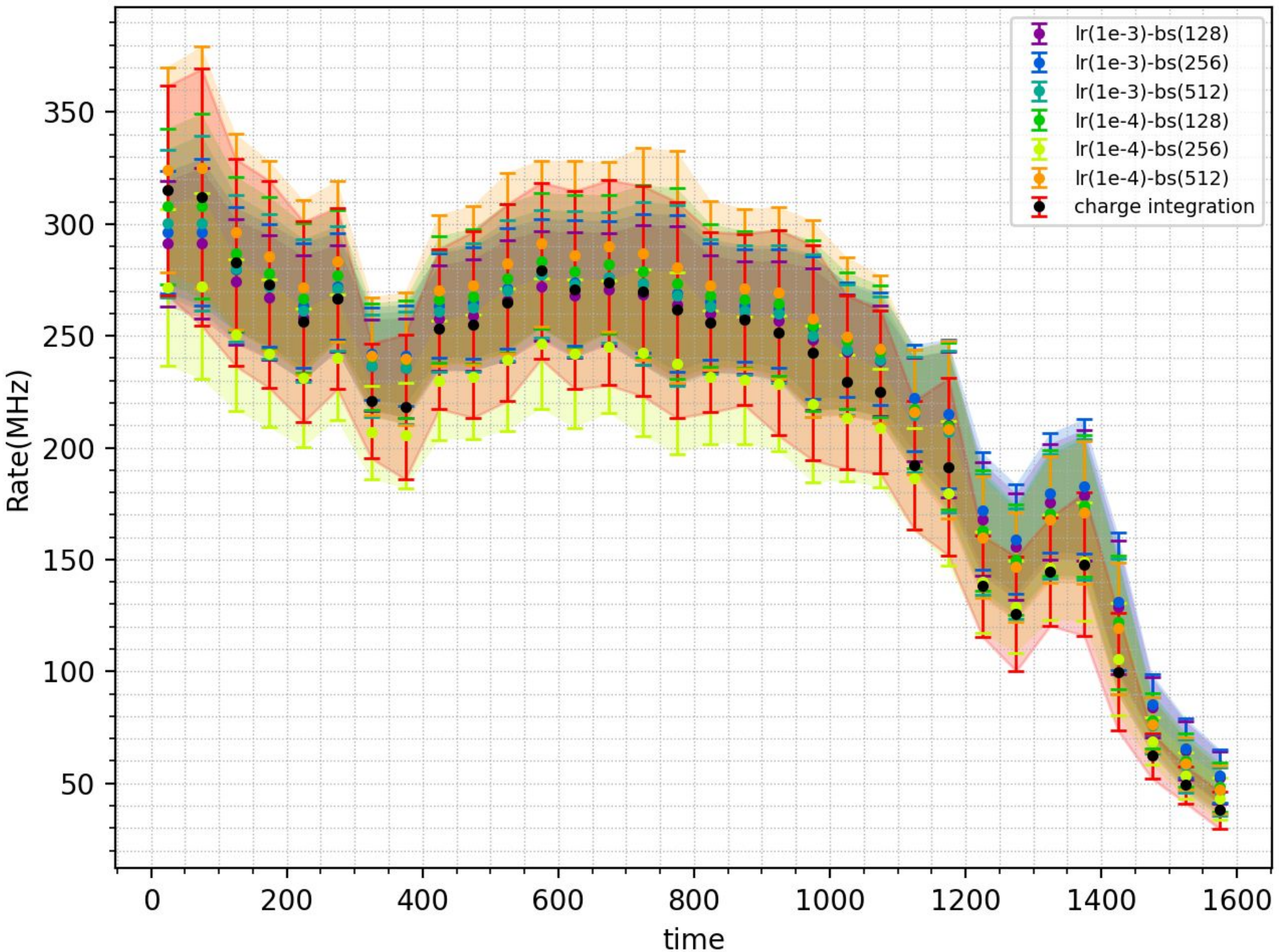




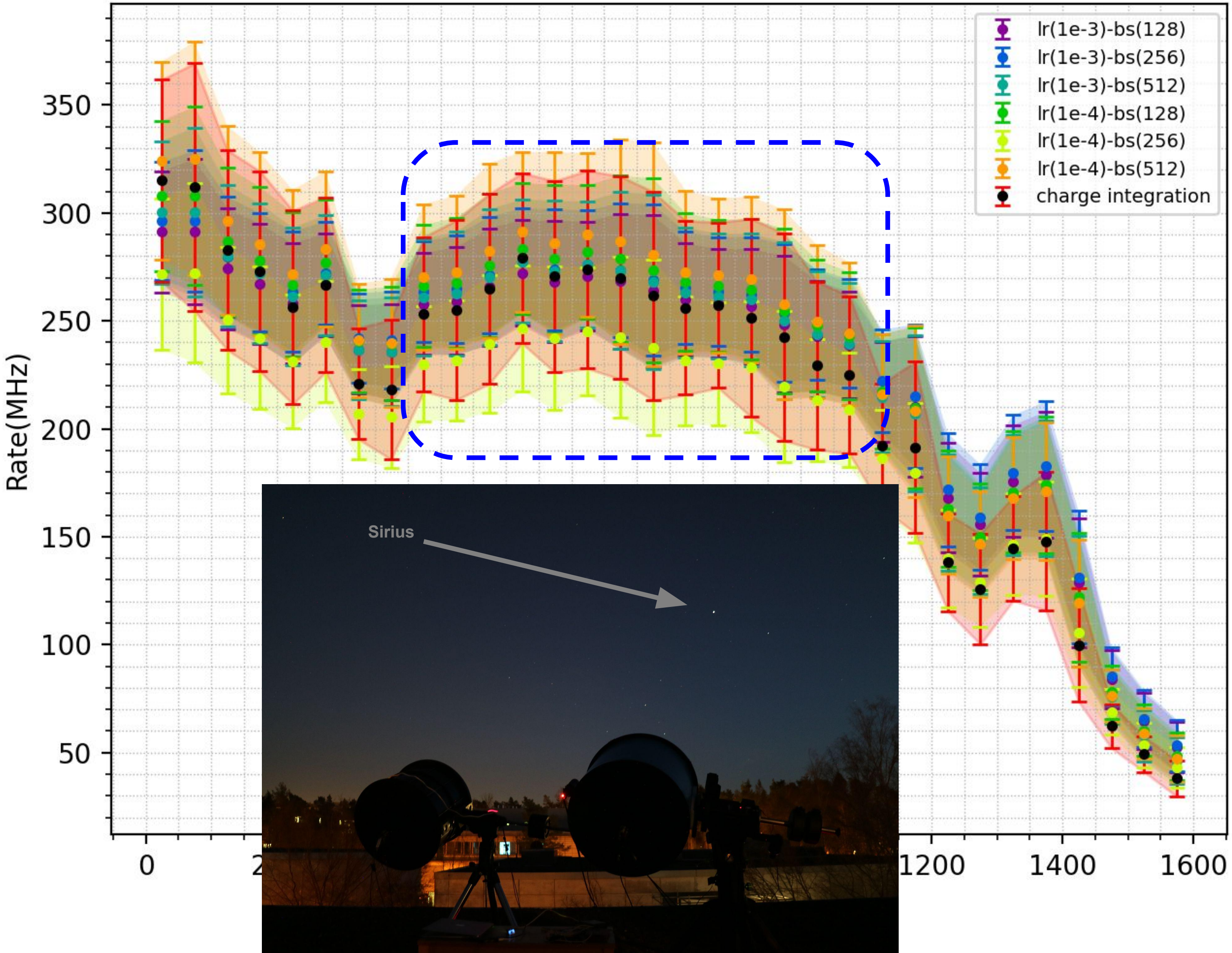


LSTM and GRU predicting similarly from 0 to 100 MHz in agreement with Charge Integration

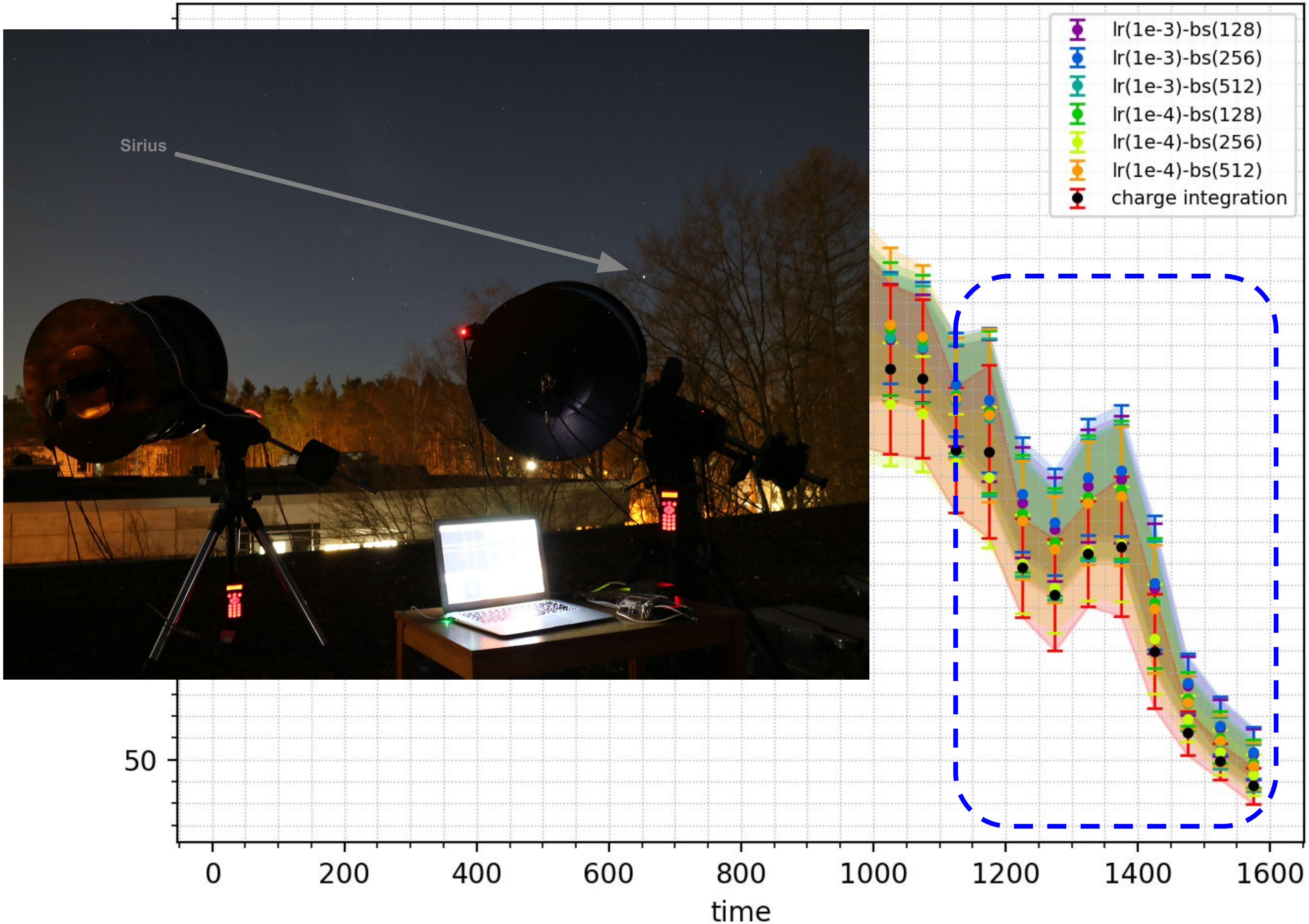
Prediction comparison(Sirius)



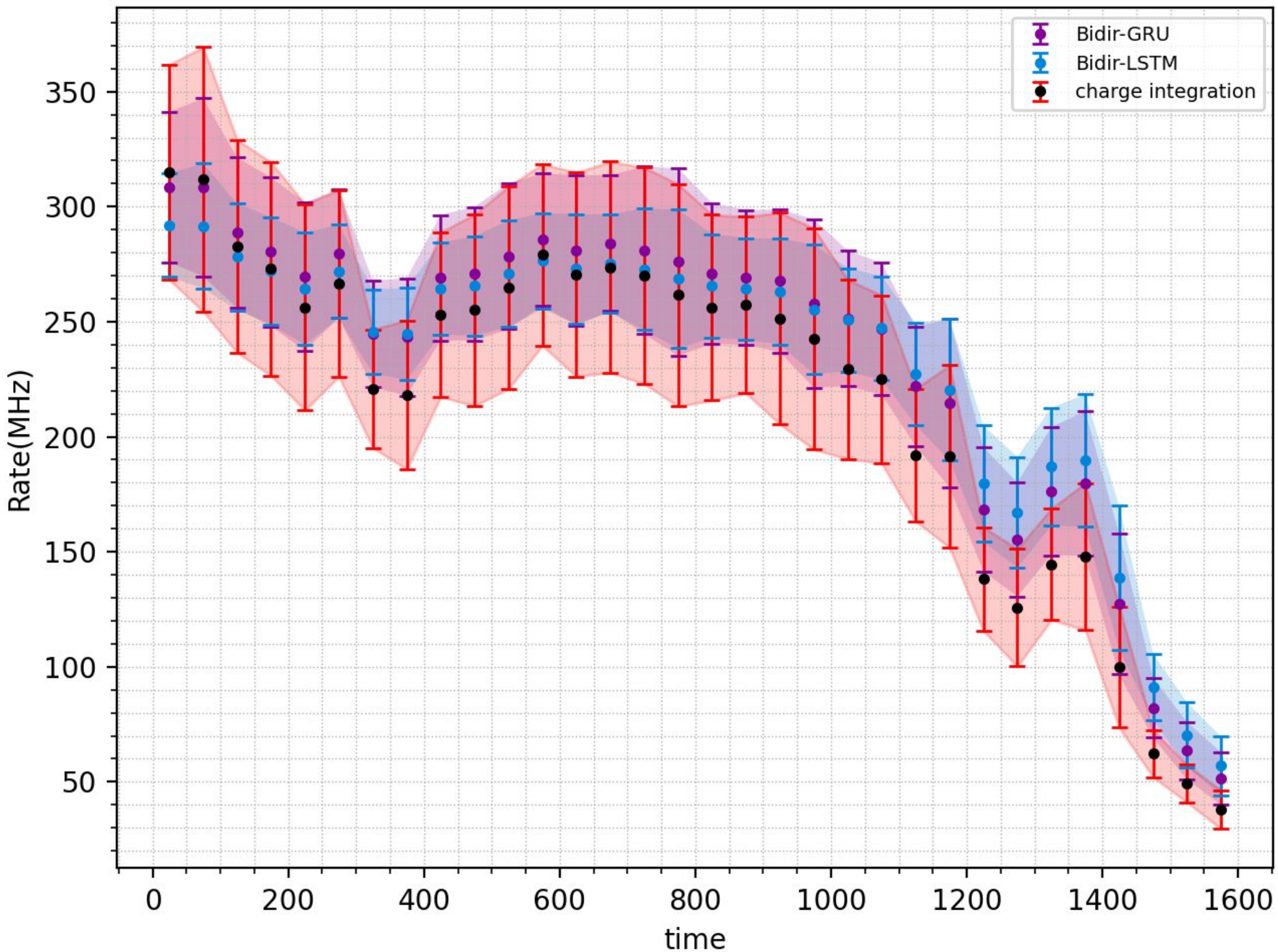
Prediction comparison(Sirius)



Prediction comparison(Sirius)



Prediction comparison(Sirius)



Conclusions

- Compare to charge integration prediction
  - NN predictions have smaller uncertainties for MC test data
  - Predictions closer to ground truth
- Comparison among different models
  - CNN is better choice than other NN models since it is predicting close to MC truth
  - Deeper CNN provides more precise predictions
    - Overfitting in extended training ⇒ requirement of regularization
  - LSTM and GRU predicting similarly from 0 to 100 MHz in agreement with Charge Integration
- Larger sample size ⇒ smaller prediction uncertainties
- Predictions are slightly different from each other

Thank you for your attention

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG