# ML for FARICH: HSE Activity Plan

CTD Software Meeting
Jan 26 2022

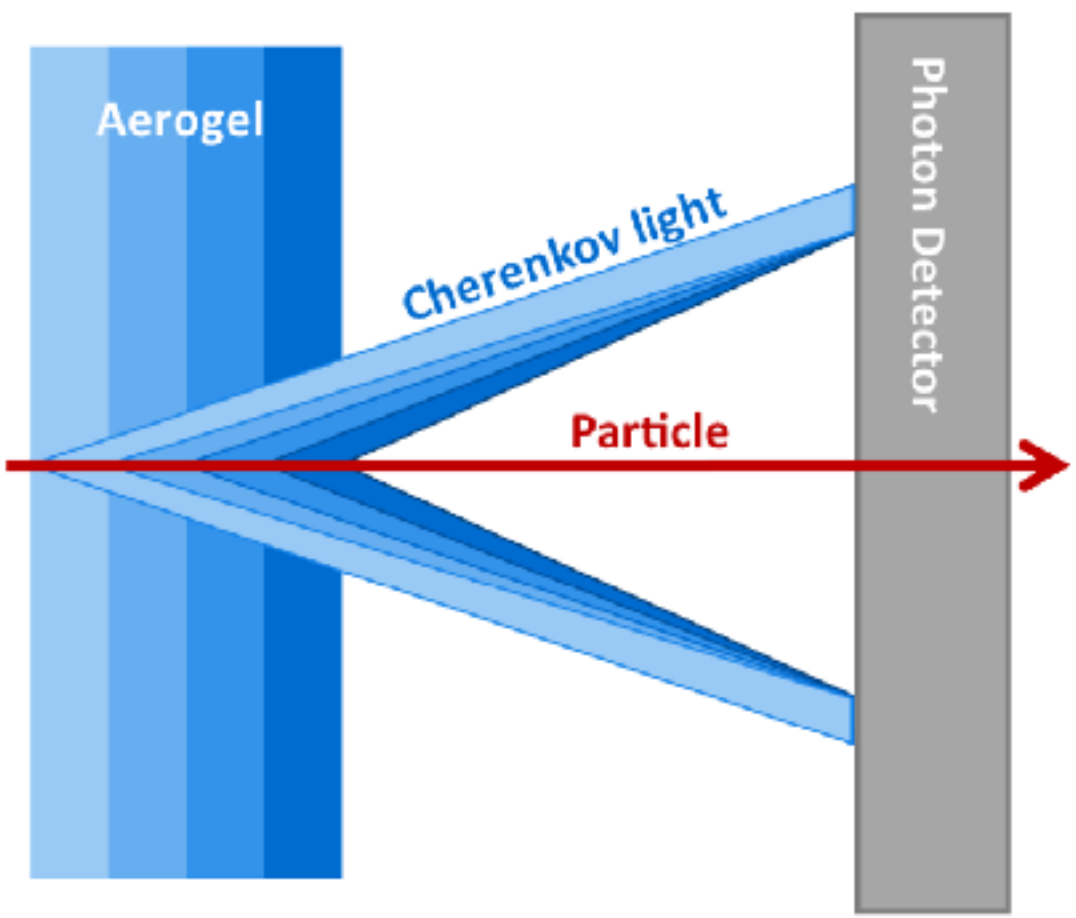
Fedor Ratnikov, Foma Shipilov
NRU Higher School of Economics

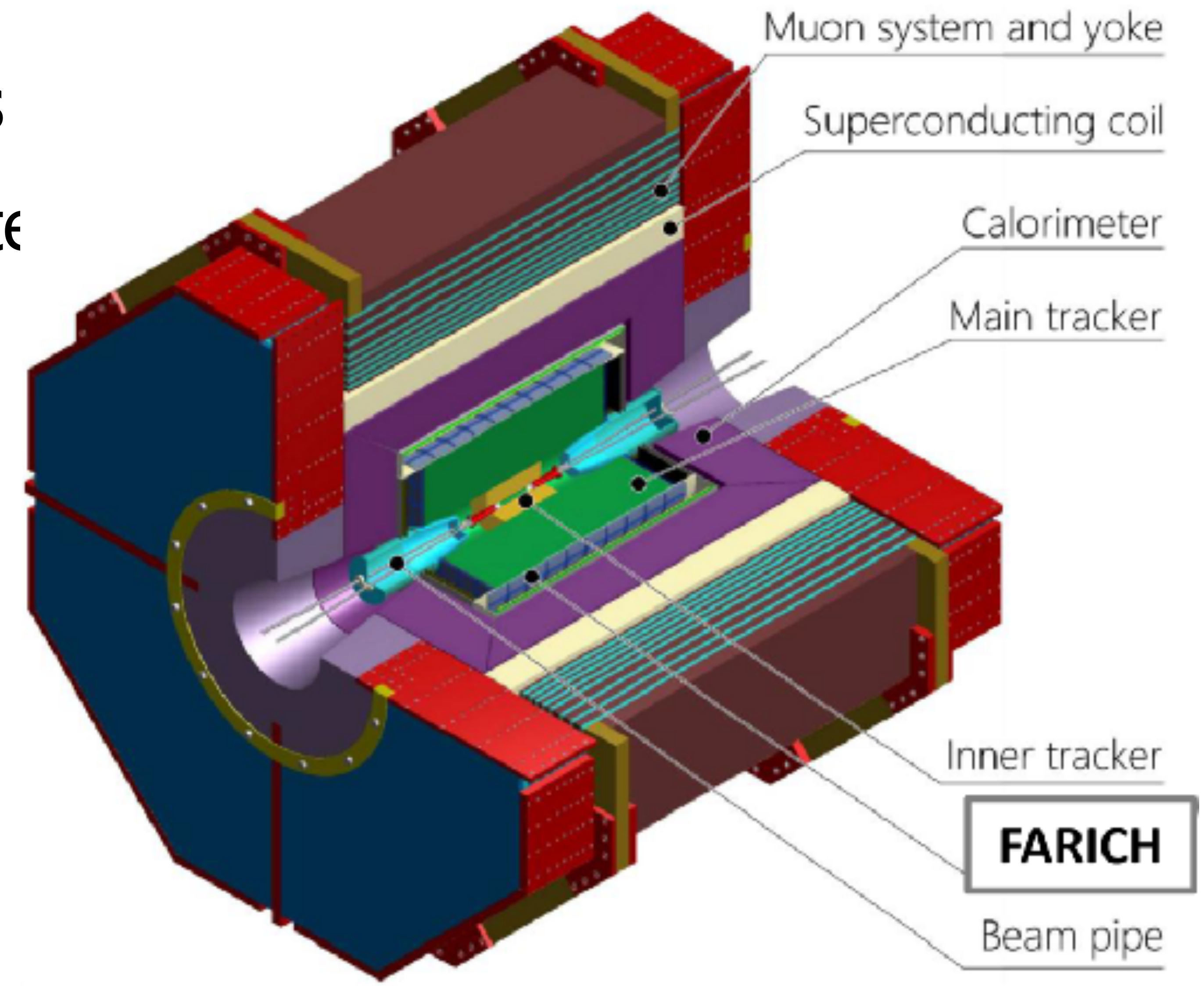
with great help from
S.Kononov, A. Barnyakov

# FARICH for CTD

## Compact particle ID detector

- trade between resolution and number of Cherenkov photons
  - improve resolution with discrete focusing capability




T.Iijima et al., NIM A548 (2005) 383
A.Yu.Barnyakov et al., NIM A553 (2005) 70

# Data

many thanks to S. Kononov for data and explanations

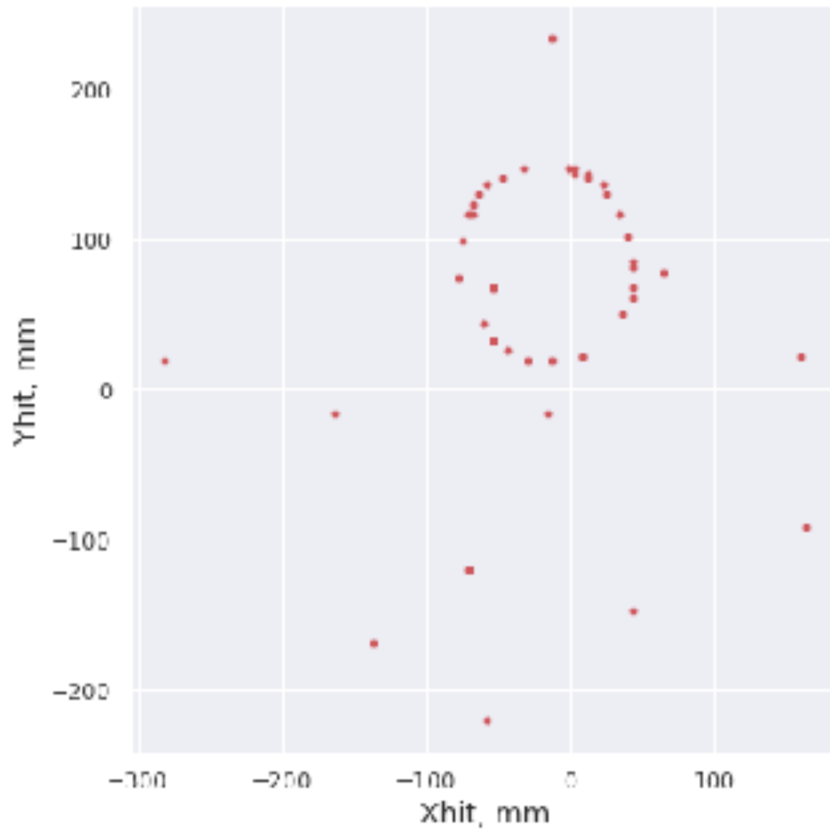
"simple" FARICH GEANT4 simulation

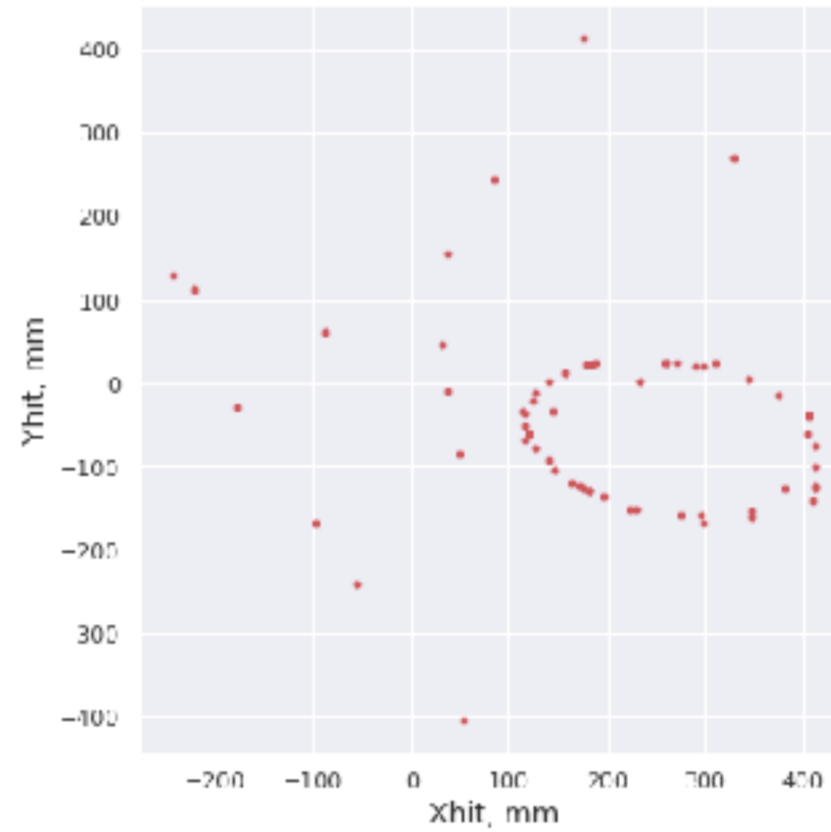
- photodetector
  - 30x30 matrix of SiPM (5760 = 30x30*8x8 total channels)
  - 3.16x3.16 $mm^2$ pixels
    - 1 mm gap between matrixes
- radiator
  - 4 layers, $n_{max}$=1.05
  - 35 mm total depth
  - 200 mm im front of the photodetector
- $\pi^-$ with varying angles $[0^\circ..45^\circ]$ and velocities $[0.957..0.999]$

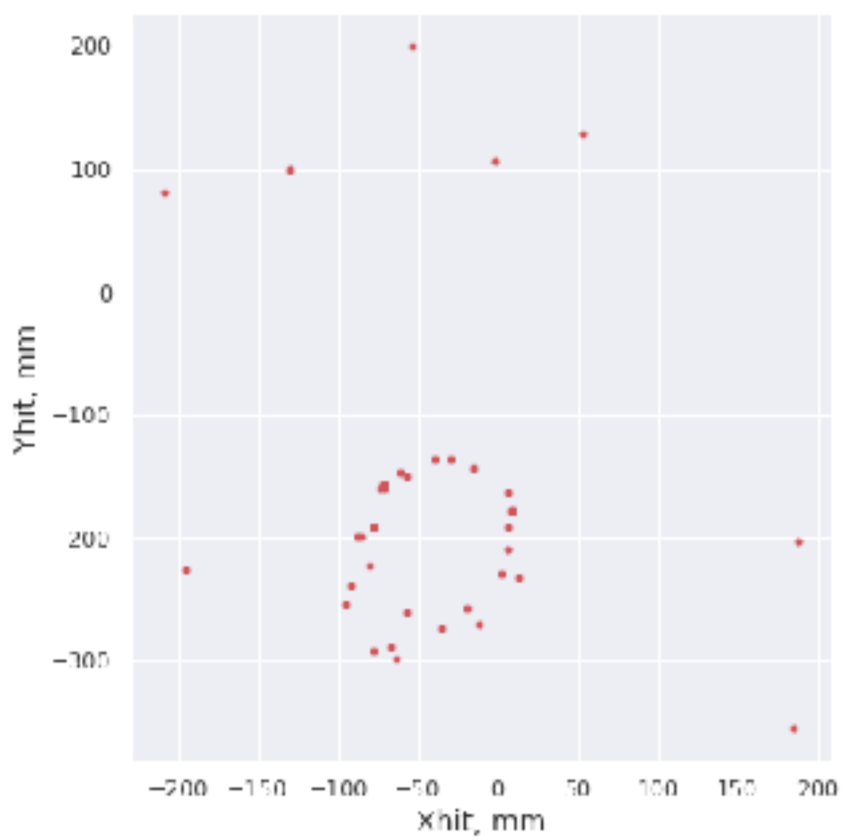
1.2M events

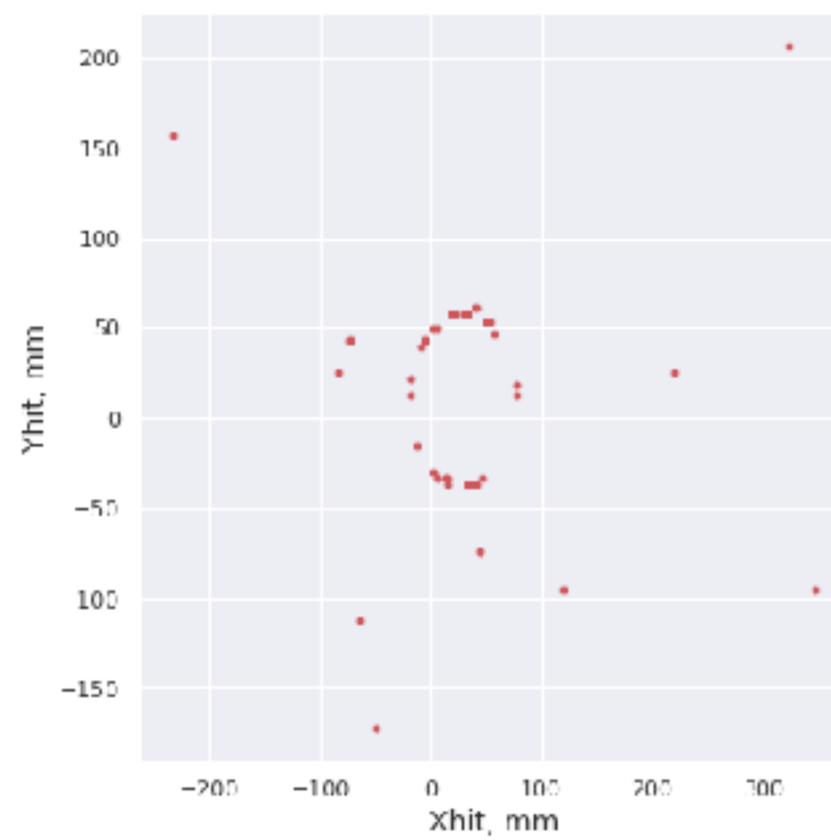
- photon hit coordinates
- hit times
- flat random noise is dynamically applied on top of events

# Data

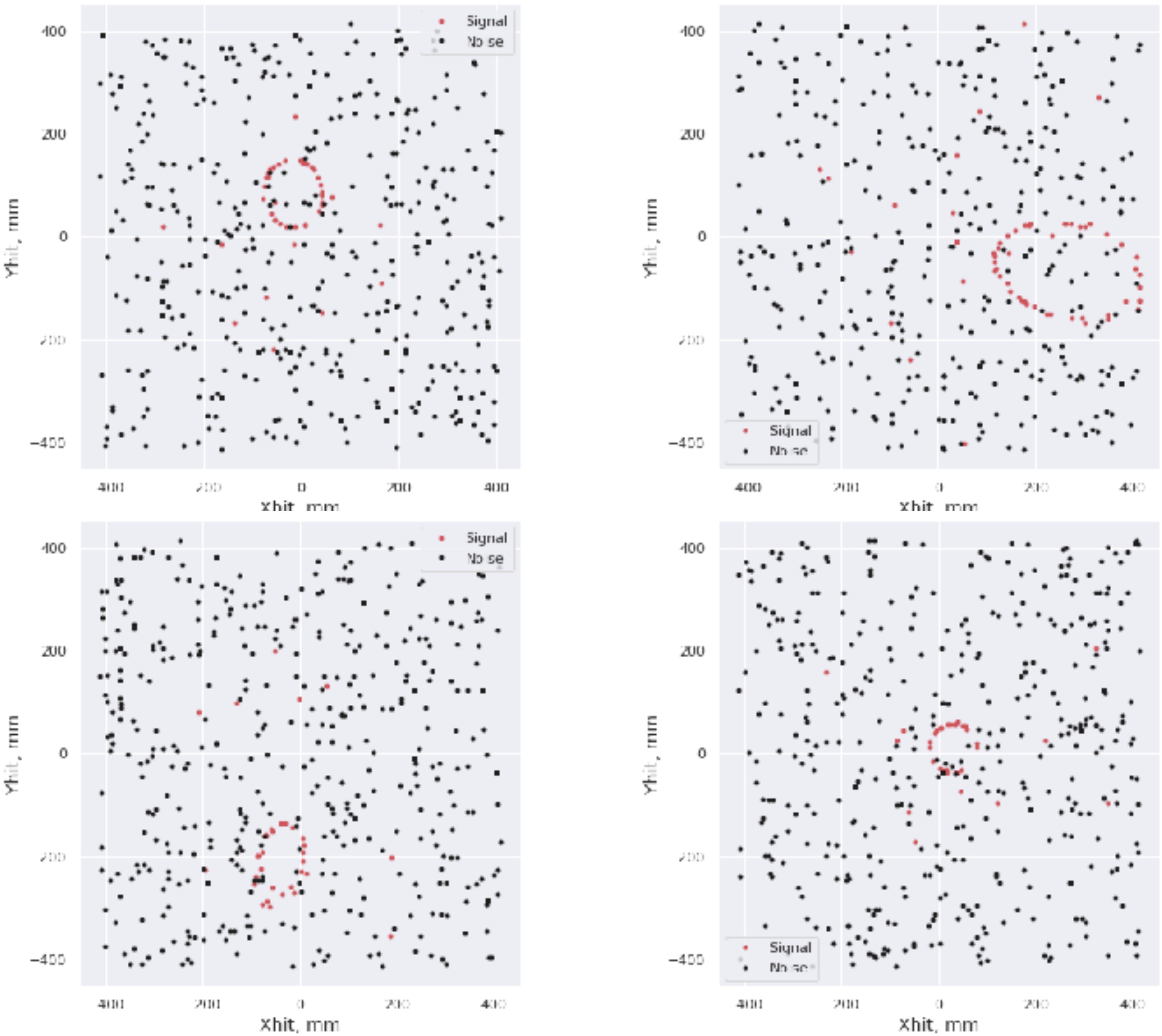


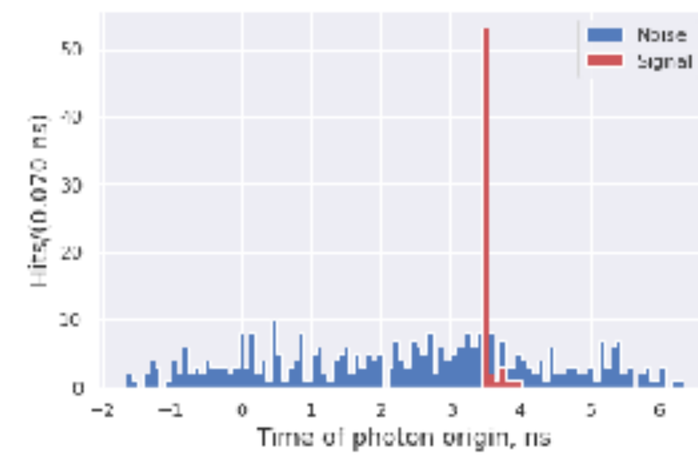



# Data (Noise 100KHz/mm$^2$)



# Big Goals

## Generative model for RECO response

- distribution $p(\beta \mid \vec{r}, \vec{p})$

  FOM: consistency between (GEANT+RECO) and GEN

- PRO: simple, CON: hard to account for explicit correlations

## Generative model for detector response

- distribution $p(\{hits(x, y, z, t)\} \mid \vec{r}, \vec{p})$

  FOM: consistency between (GEANT+RECO) and GEN+RECO

- PRO: account for correlations CON: complicated model

## RECO baseline (with known track)

- high fidelity
- necessary for GEN models evaluation

## RECO standalone (without track prior)

- can work in trigger if made fast enough

# Procedures

- not regular matrix structure

  due to gaps between SiPM

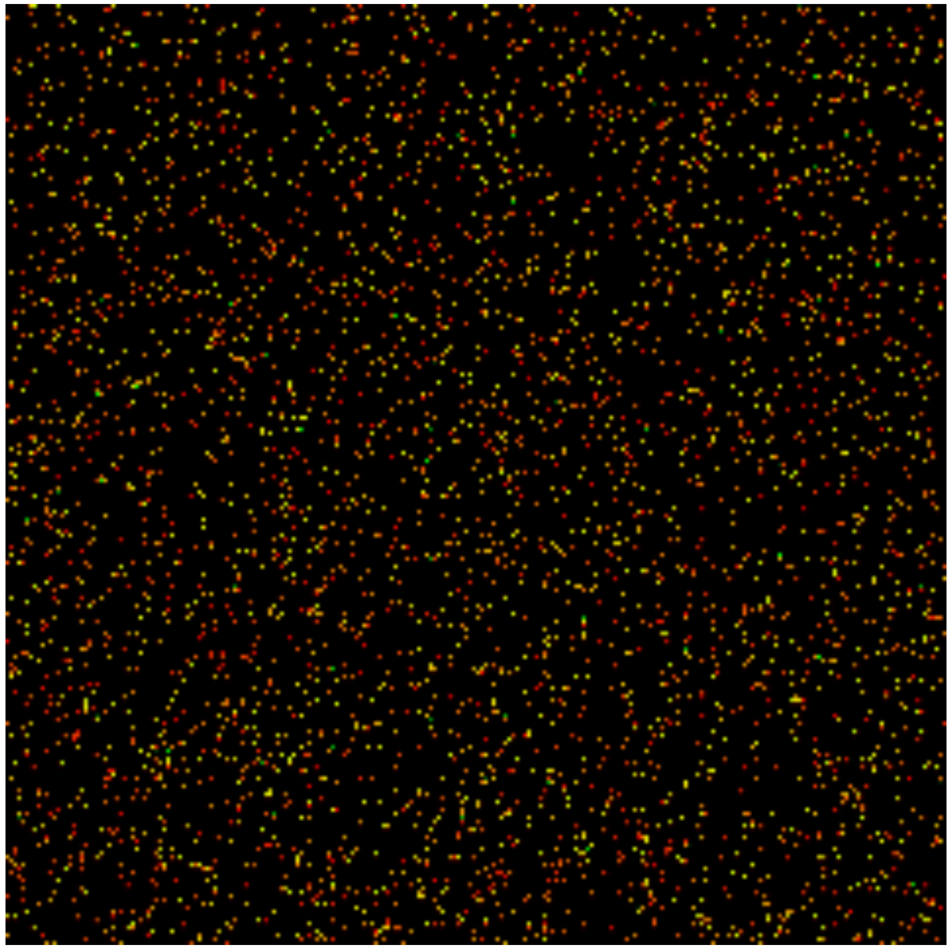
  - project hits (times) to a regular grid
- train Convolutional Neural Network to extract information

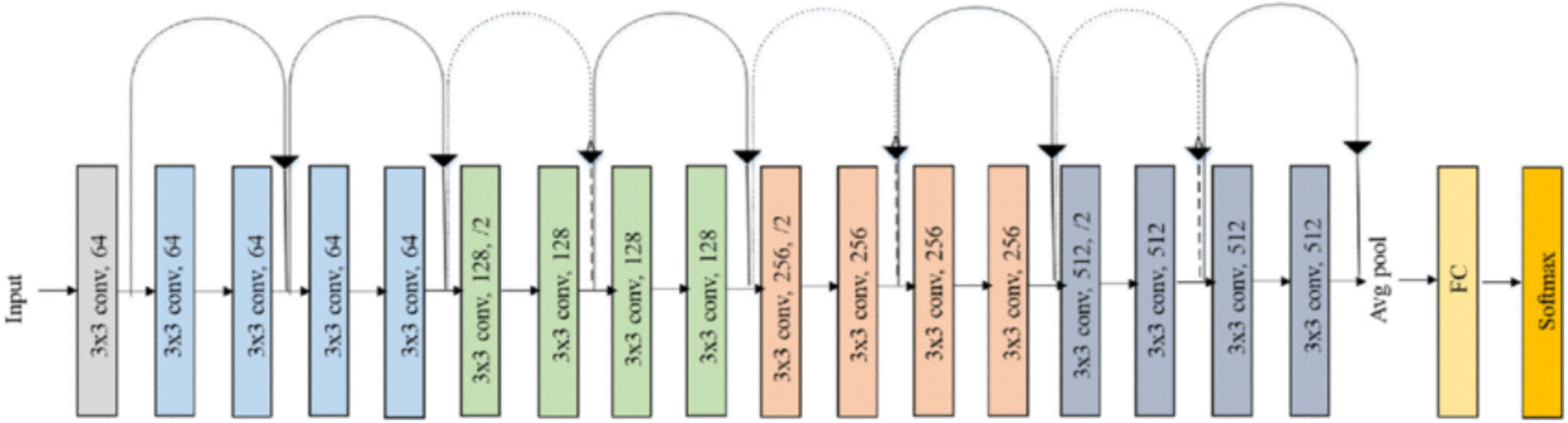
  $$\beta: \ [0.957..0.999] \ \rightarrow \ [0..1]$$

- use ResNet-18 CNN architecture

  tune first and last layers to accommodate input and output data formats

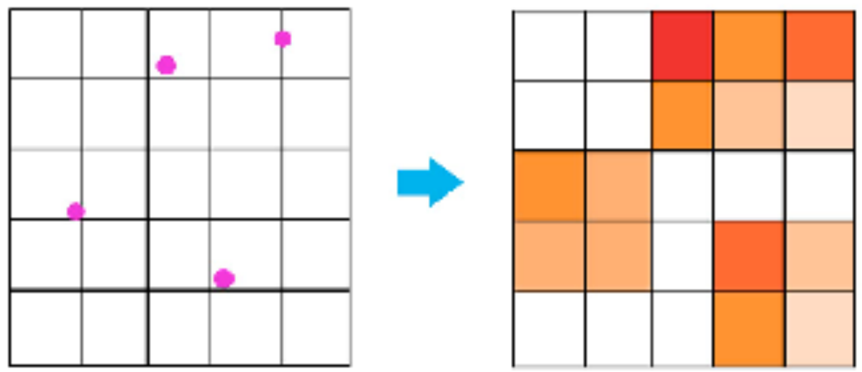
Map example. Red – hits, green – times





# Procedures

- not regular matrix structure



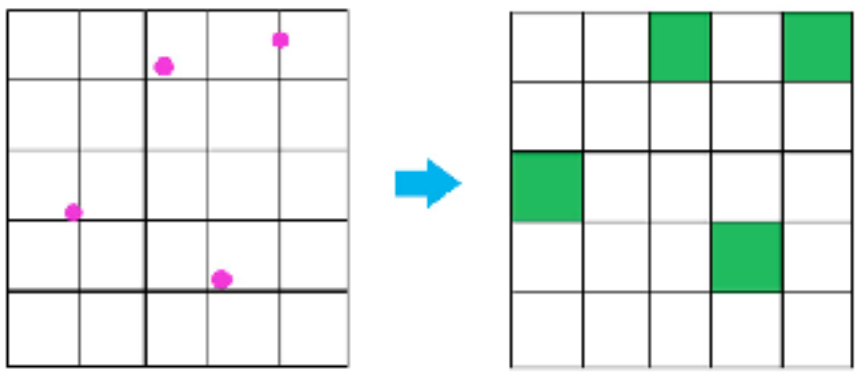
hits mapping

egular grid

Network to

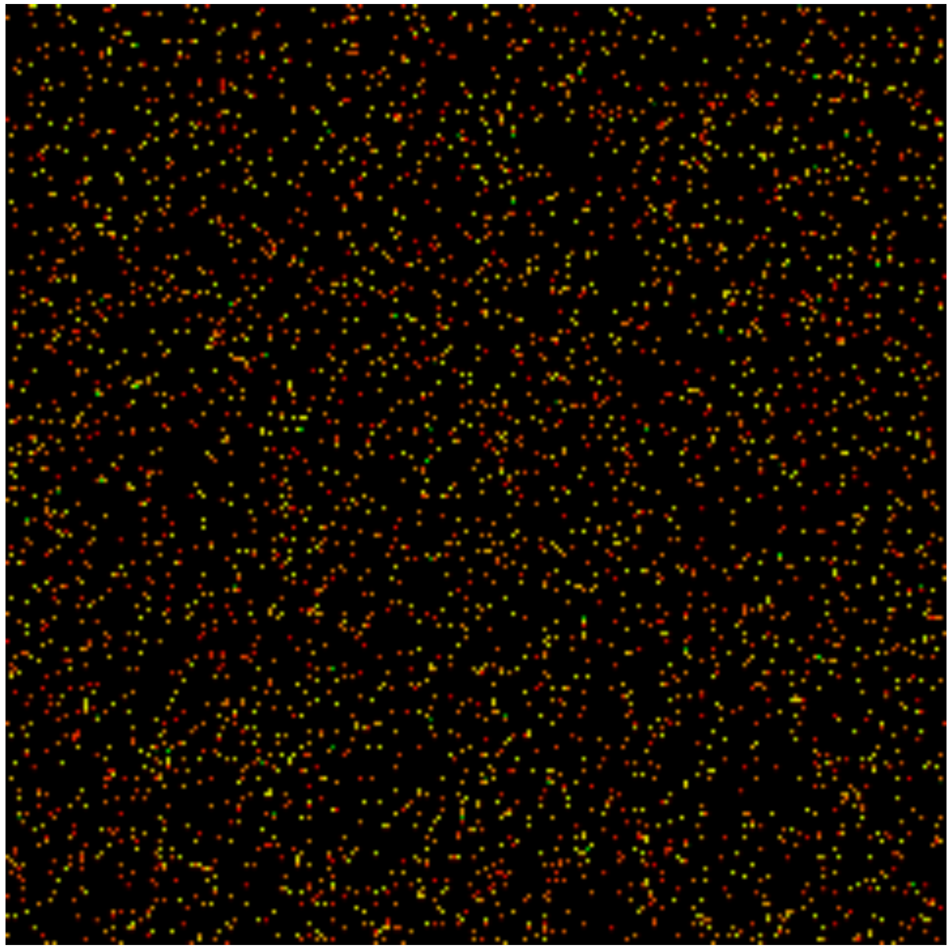
- use ResNet-1

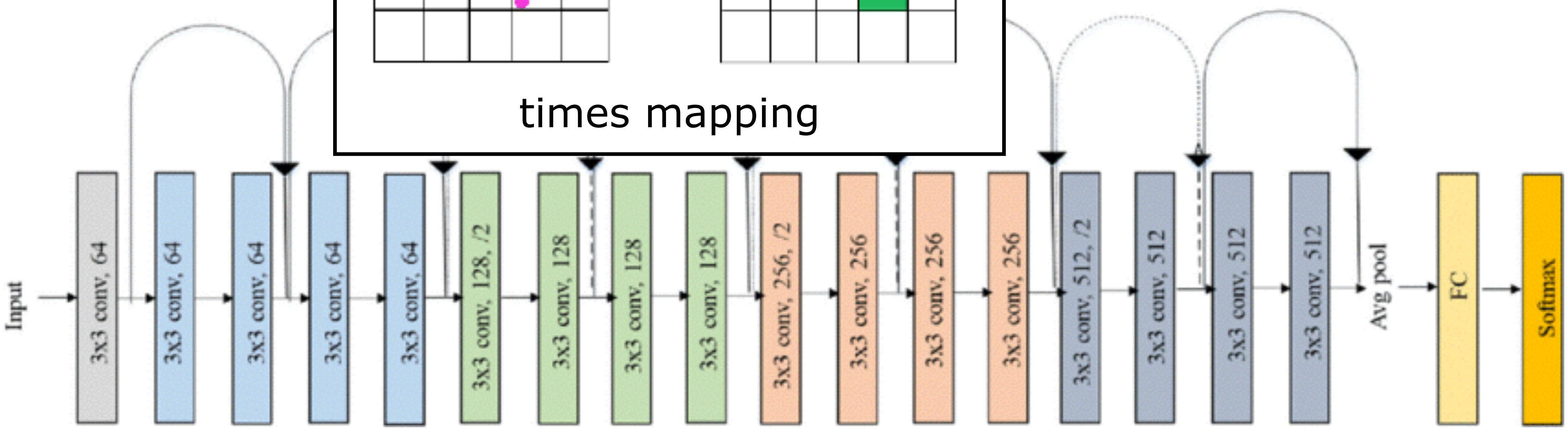
tune first an
input and ou



times mapping

Map example. Red – hits, green – times





# Disclaimer

The following are first attempts to understand and analyze data

No special tuning is applied

- used tools are out of the box

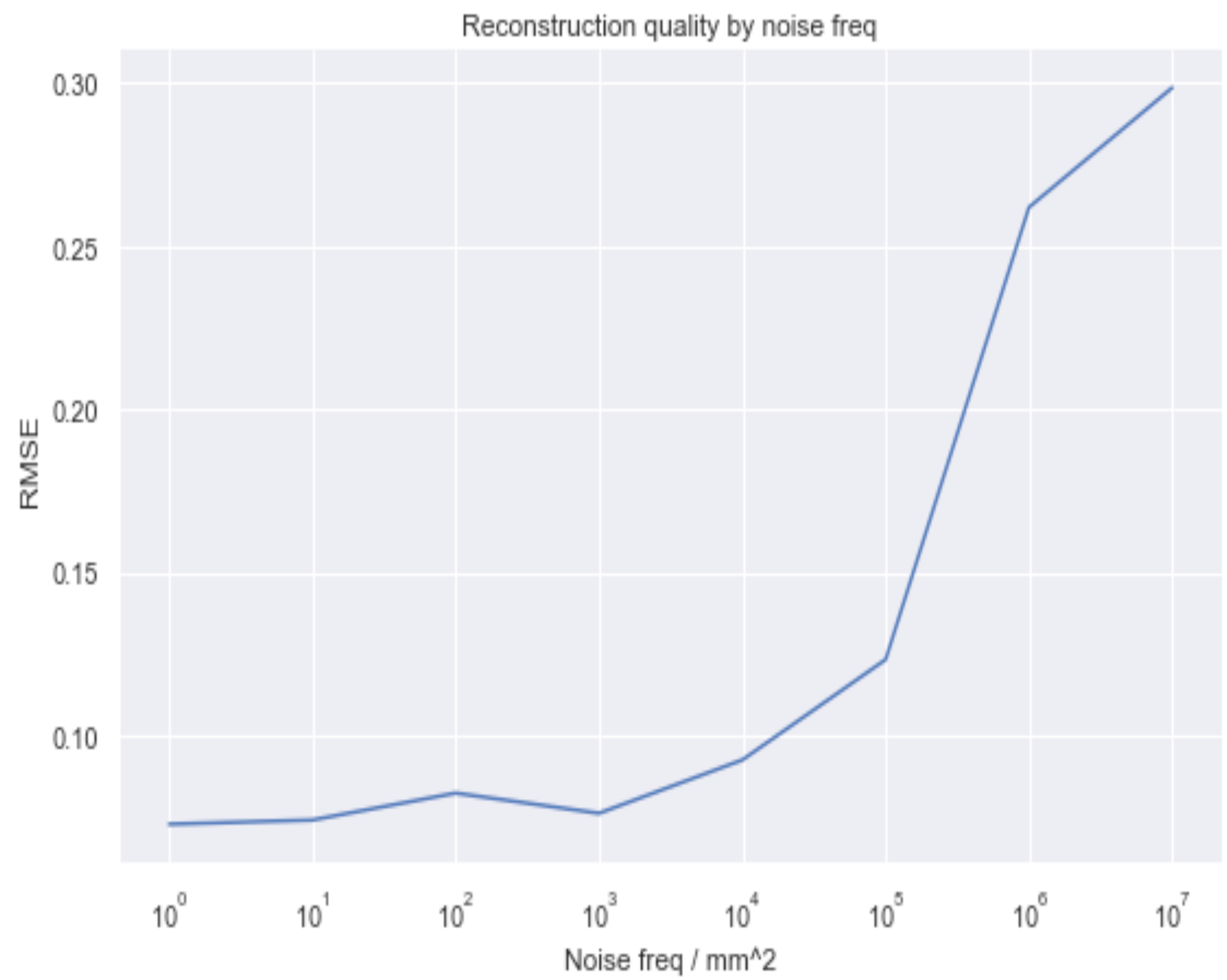
The goal is not presenting results but rather discussing approaches and directions for attack

# Seedless RECO Quality

- $\sigma(\beta_{scaled})$
- $\beta: \; [0.957..0.999] \; \rightarrow \; [0..1]$
- no track prior is used

  however, signal existence prior

- RMSE ~0.3 for [0,1] means no sensitivity

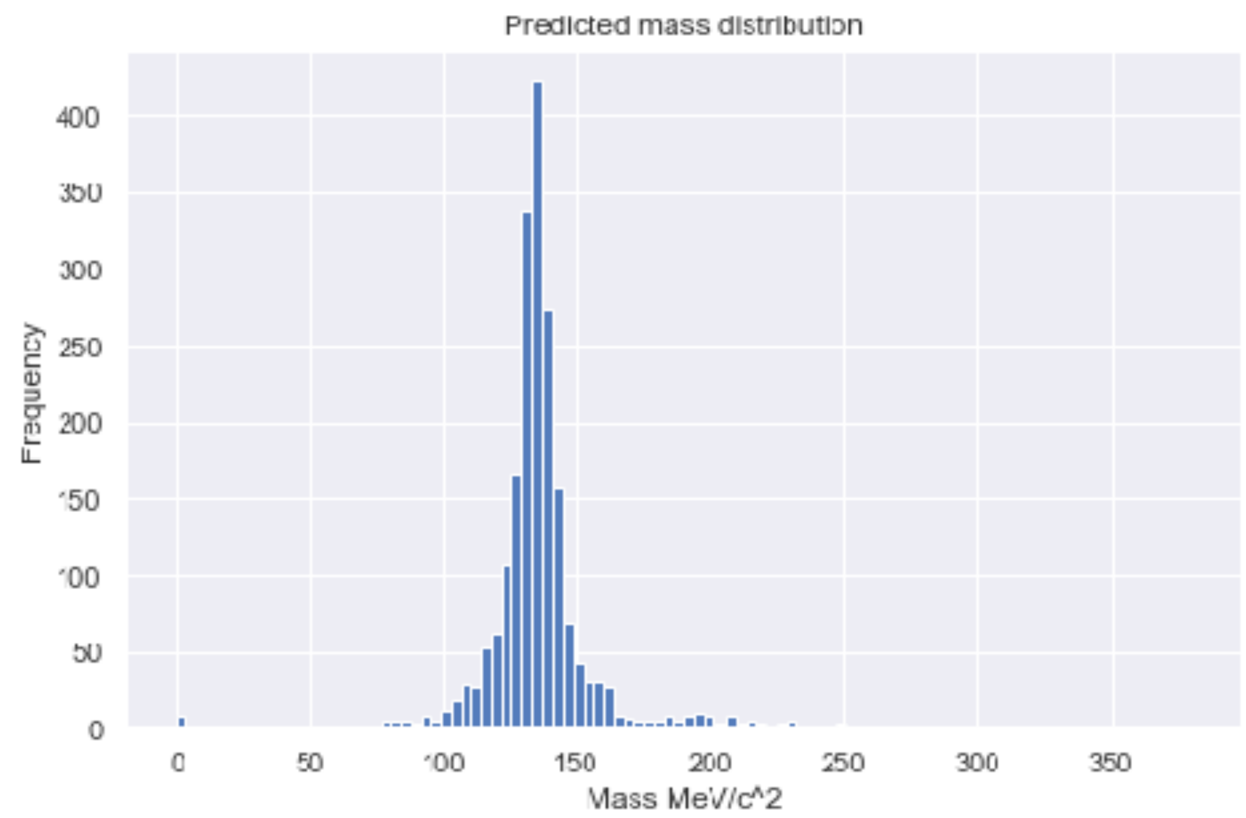


# Convert $\beta$ to Mass

Use known pion MC momentum to convert measured velocity to measured mass

Evaluate particle ID

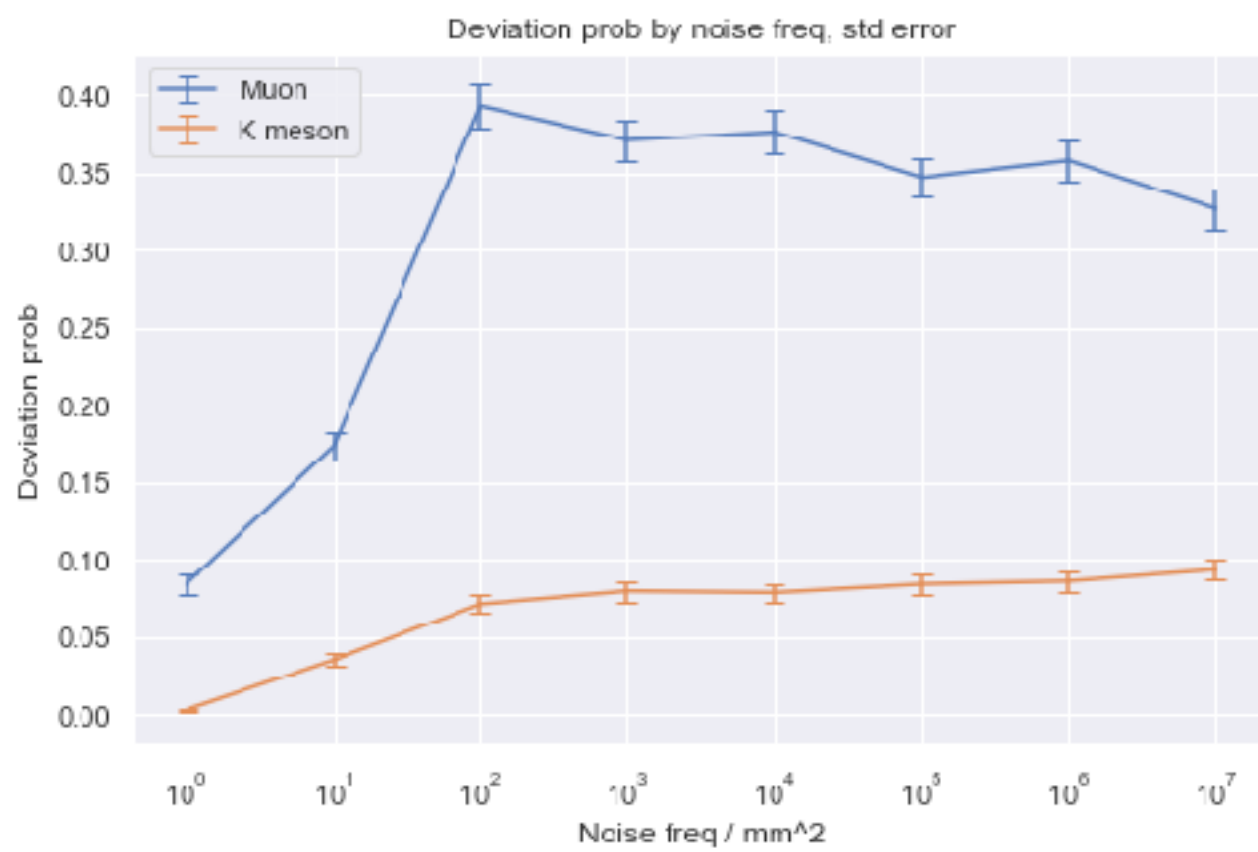
- consider identified as $\mu$ or K if reconstructed mass is more than half mass difference off




NB: due to original momentum spectra corresponding to $\pi$ the reconstructed mass is grouped around $m(\pi)$ even with no reconstruction ability

- need other particle type samples for fair comparison

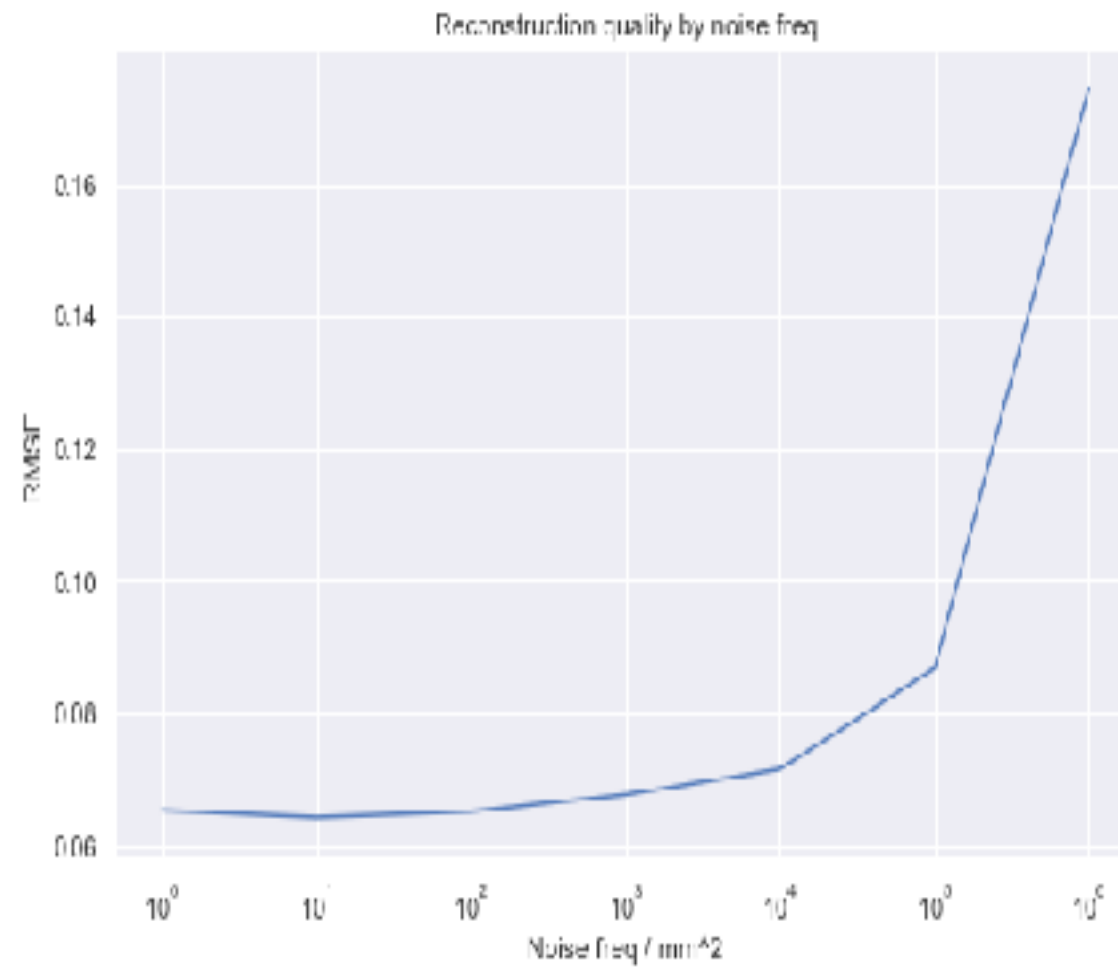
better with smooth distribution in mass



# Seeded RECO Quality

Add track point at photodetector plane to inputs

- $\sigma(\beta_{scaled})$
- $\beta : \; [0.957..0.999] \; \rightarrow \; [0..1]$
- RMSE ~0.3 for [0,1] means no sensitivity



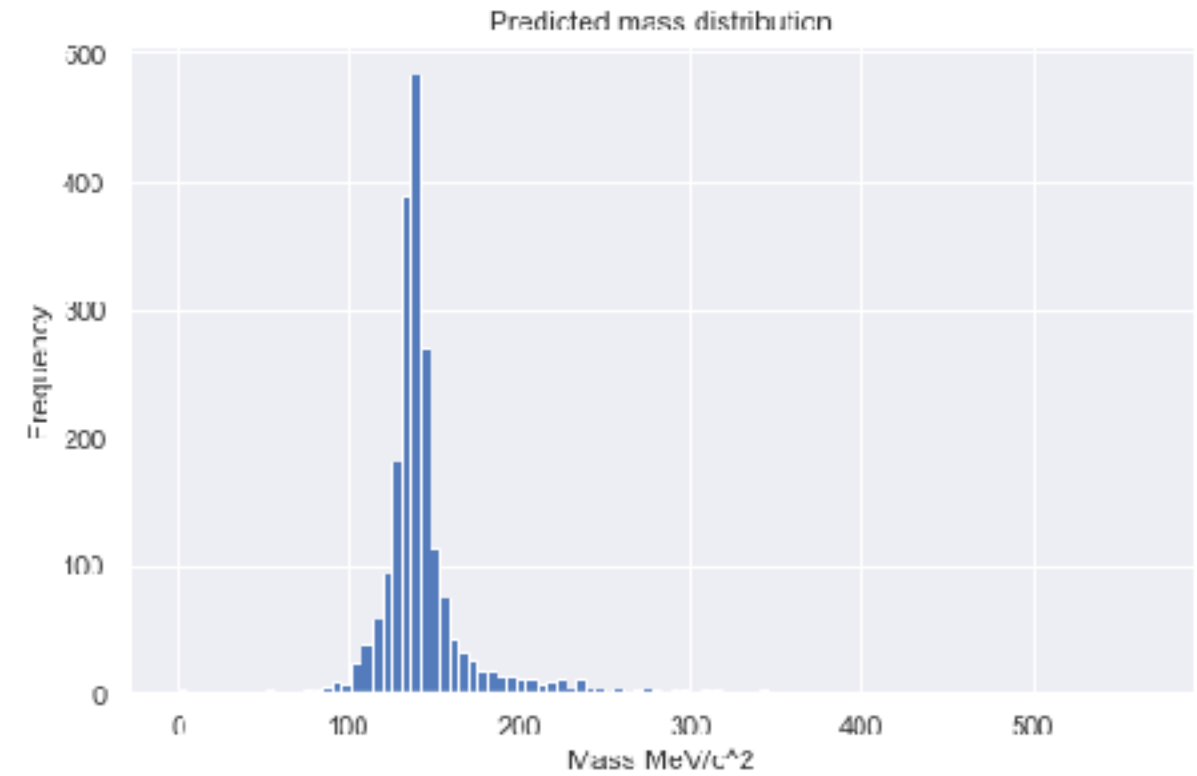

Significantly better resolution

# Convert $\beta$ to Mass

Use known pion MC momentum to convert measured velocity to measured mass

Evaluate particle ID

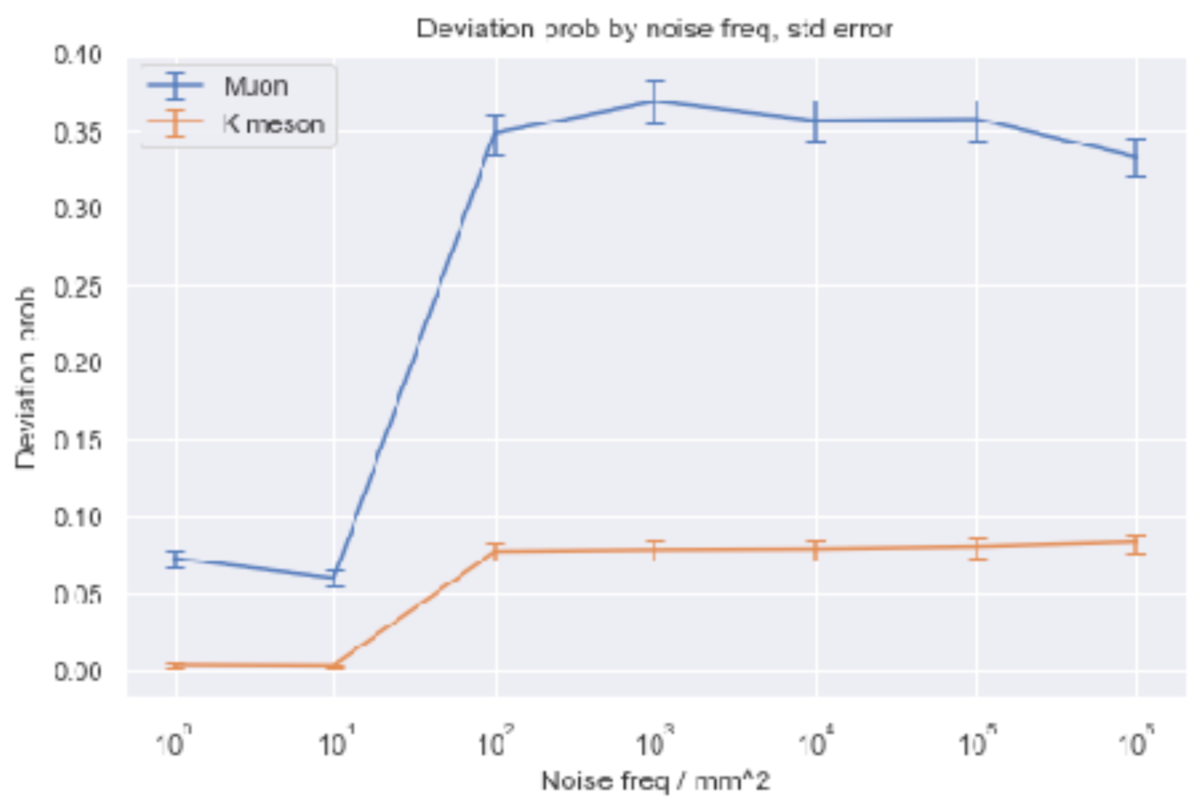
- consider identified as $\mu$ or K if reconstructed mass is more than half mass difference off




NB: due to original momentum spectra corresponding to $\pi$ the reconstructed mass is grouped around $m(\pi)$ even with no reconstruction ability

- need other particle type samples for fair comparison

  better with smooth distribution in mass



# Questions

- Is baseline FARICH RECO available?
- What is realistic noise level?

  to which extend it may be considered uniform?
- Is signal time window known in advance?
- Do we need to reconstruct $\beta$?

  could variables like $\log \frac{p(K)}{p(\pi)},\ \log \frac{p(\mu)}{p(\pi)},\ \ldots$ be more appropriate?
- Would GEN model for $\beta$ with average correlation effects be enough?
- What are quantitive FOM for GEN models?
- Which physics information we want from seedless online RECO (just event multiplicity or track kinematics)?

# Conclusions

Analysis data are good

- will coordinate few more samples to avoid ML train biases

First approaches to plain CNN based RECO are presented

- is some RECO baseline available?

To be continued…

# Параметры "простого" моделирования ФАРИЧ для СЧТФ в Geant4

**Фотодетектор (ФД)**

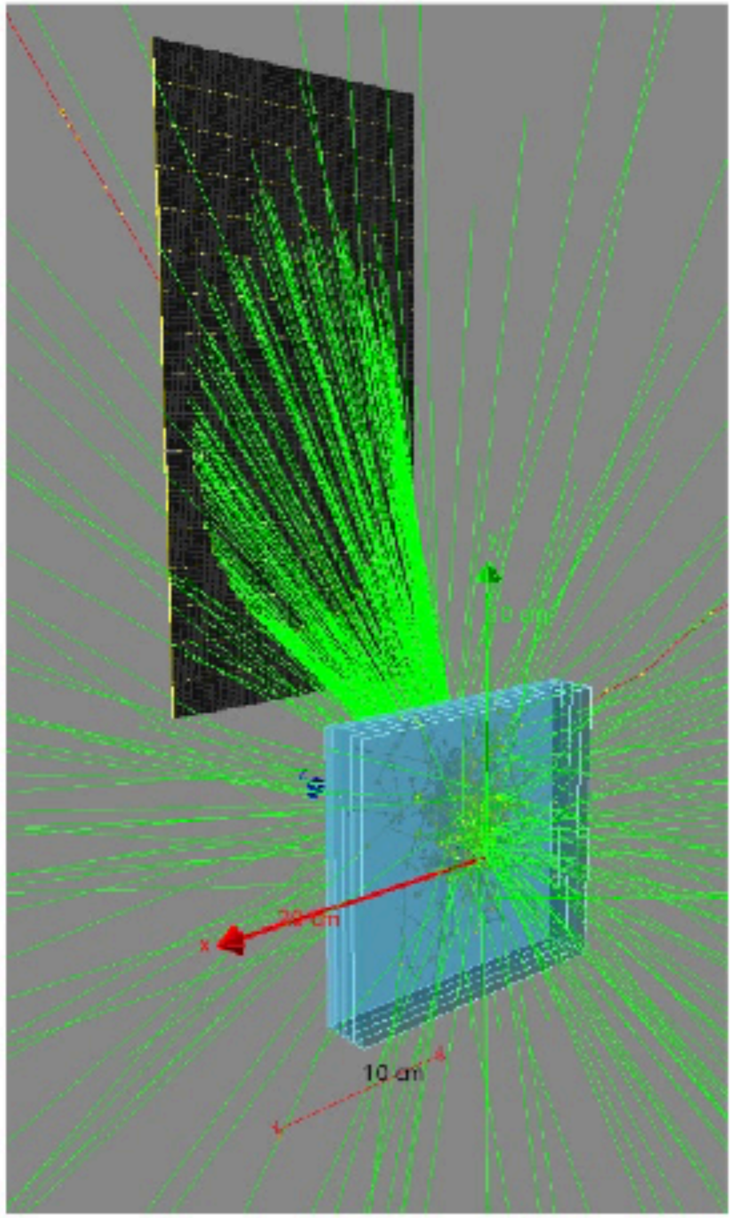
- 30×30 матриц кремниевых фотоумножителей (КФУ) с 8×8 пикселями
  Модель КФУ: **ON Semiconductor ArrayJ-30020-64P-PCB**
- Размер ФД: 830×830 мм,
- Полное число пикселей: $(30*8)^2 = 5760$
- Размер пикселя 3.16×3.16 мм
- Шаг пикселя в матрице 3.36 мм
- Зазор между матрицами: 1 мм
- Средняя плотность пикселей ≈ 88%
- $U_{смещ.} = 2.5$ В
- $\lambda_{max} \approx 400$ нм, $PDE_{max} \approx 38\%$

**Радиатор**

- Чертырехслойный фокусирующий радиатор
- Макс. показатель преломления $n_{max} = 1.05$
- Полная толщина 35 мм

**Расстояние ФД-радиатор:** 200 мм

Частицы запускаются прямо перед радиатором с заданными углами и скоростью



# RECO Masses

NB: due to original momentum spectra corresponding to the reconstructed mass is grouped around even with no reconstruction ability

- need other particle type samples for fair comparison

  better with smooth distribution in mass