DE LA RECHERCHE À L'INDUSTRIE

# Machine learning algorithms for the gamma conversion reconstruction in the ClearMind project

Chi-Hsun Sung for the ClearMind Collaboration

26/09/2022

# Outline

- ClearMind project: Background, Objectives & Detector Principles
- Training data preparation: Geant4 simulation
- Event reconstruction
- Conclusion

# ClearMind Project & Detector Design



- Fast detector for TOF-PET:
  - – Coincidence time resolution: **<100 ps** (FWHM)
  - – 511-keV γ-ray interaction 3D resolution: **a few mm**

**Detect efficiently Cherenkov and scintillation lights**

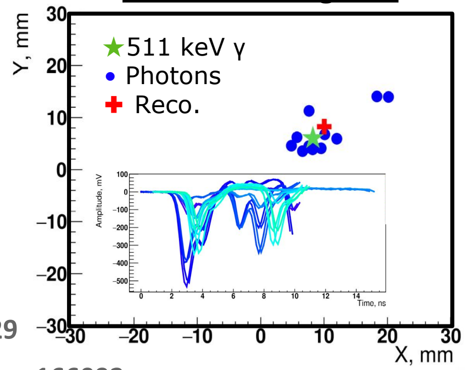
⟶ Detector with **monolithic**, **large surface, $PbWO_4$** crystal as the optical window of the **MCP-PMT**

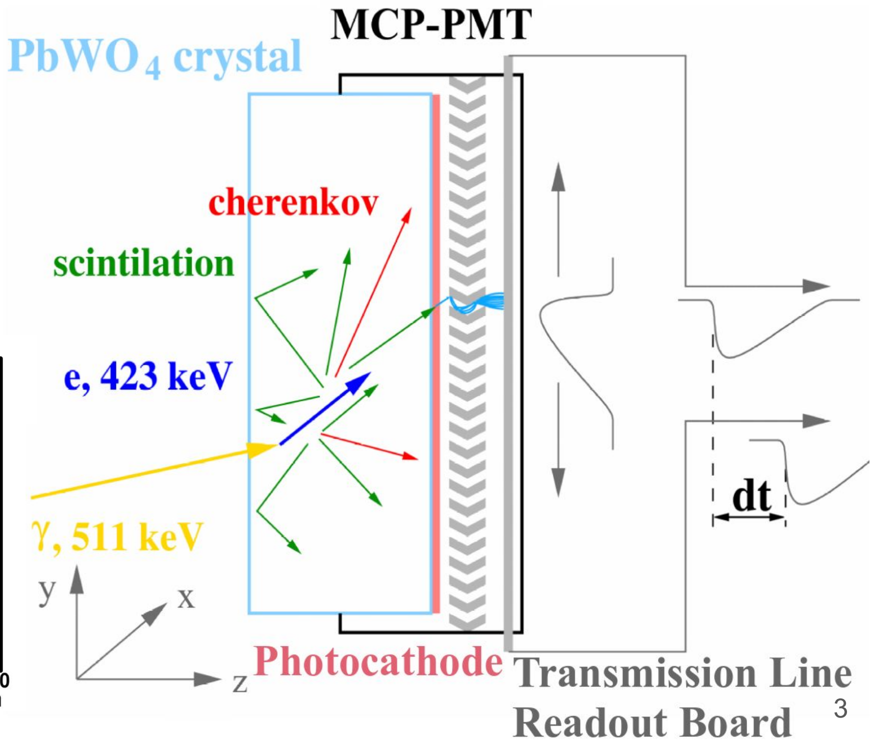
⟶ **Direct deposition** of the photocathode

⟶ **Transmission line** readout board

⟶ **SAMPIC** digitization module

**Reco. with signals**




– D. Yvon *et al.*, 2020, JINST 15 P07029
– M. Follin *et al.*, 2021, NIM A, 1027, p. 166092

# ClearMind Project & Detector Design

- Fast detector for TOF-PET:
  - – Coincidence time resolution: **<100 ps** (FWHM)
  - – 511-keV γ-ray interaction 3D resolution: **a few mm**

**Detect efficiently Cherenkov and scintillation lights**

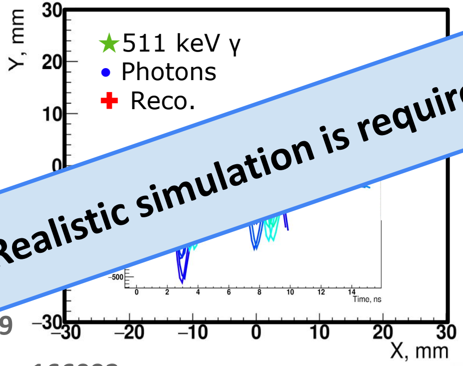
⟶ Detector with **monolithic**, **large surface, $PbWO_4$** crystal as the optical window of the **MCP-PMT**

⟶ **Direct deposition** of the photocathode

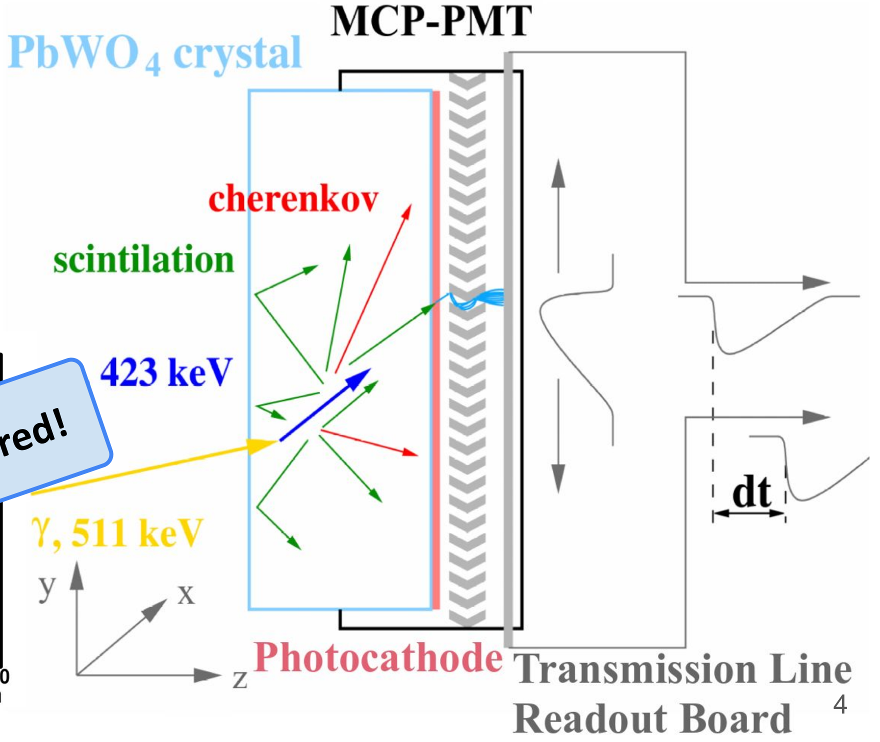
⟶ **Transmission line** readout board

⟶ **SAMPIC** digitization module

**Reco. with signals**



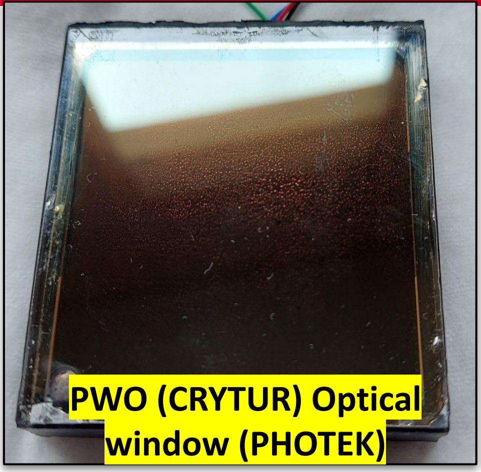

**Realistic simulation is required!**



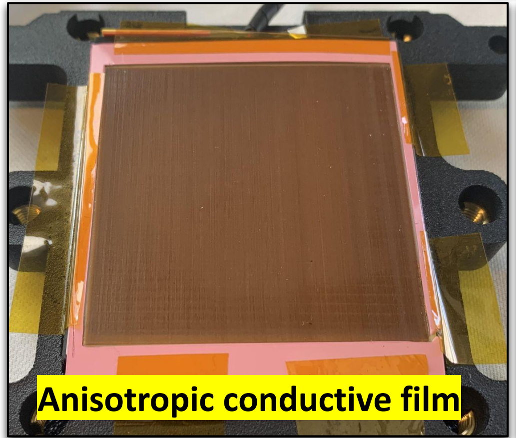

– D. Yvon *et al.*, 2020, JINST 15 P07029

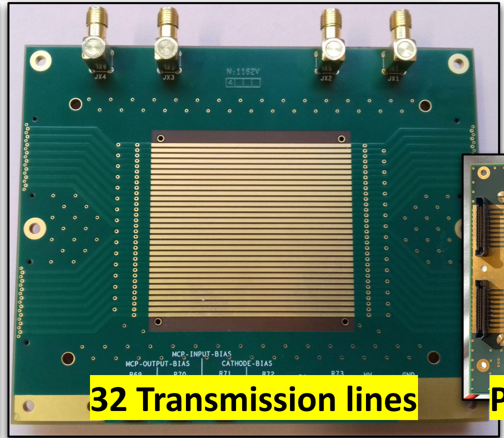
– M. Follin *et al.*, 2021, NIM A, 1027, p. 166092

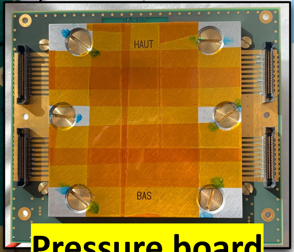
# ClearMind Prototype Components I

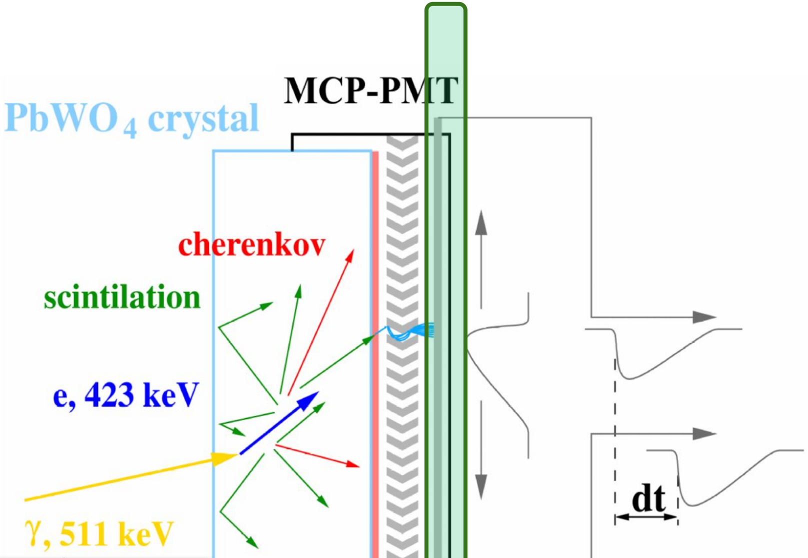






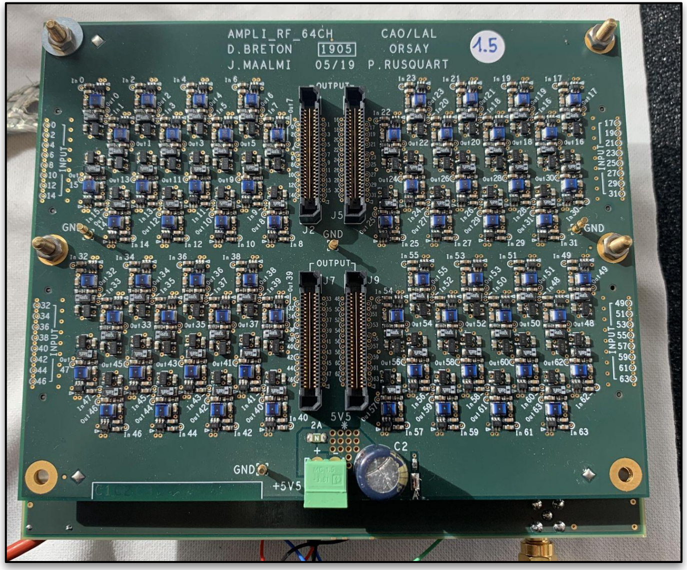



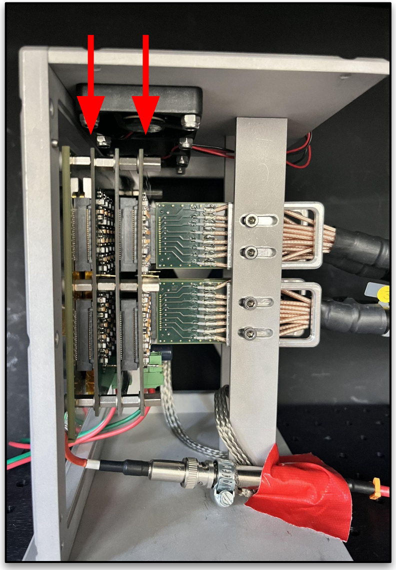



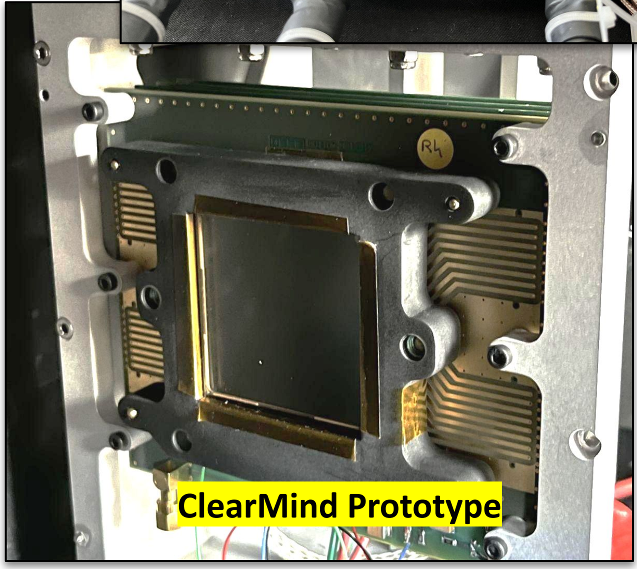
# ClearMind Prototype Components II

Amplifier 20dB*2 (IJCLab)



SAMPIC crate (IJCLab+Irfu)

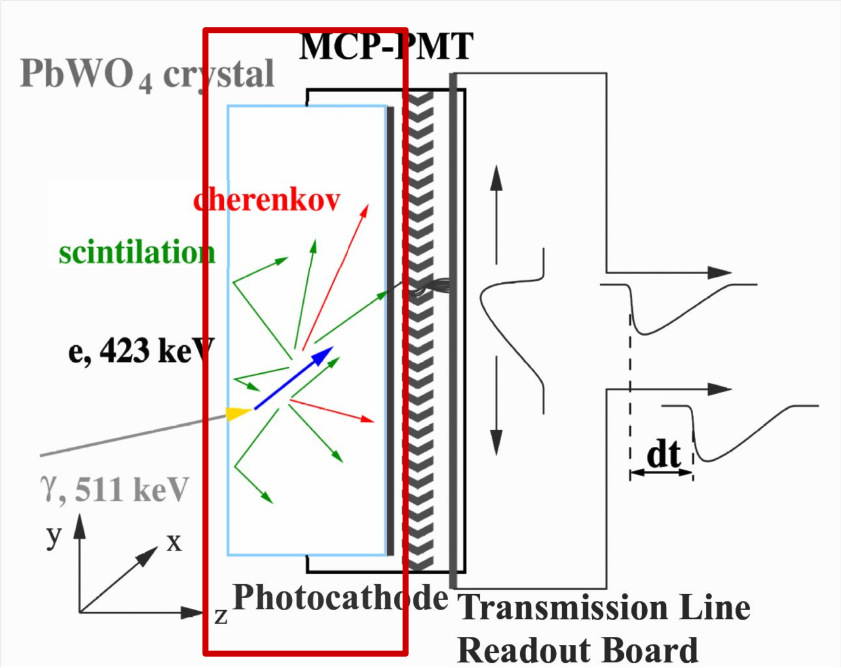


ClearMind Prototype

# Training data preparation

# Gamma Interaction Simulation

- Geant4 v.10.7
- Physics list:
  G4EmPenelopePhysics with default particle range cut
- Lead tungstate ($PbWO_4$):
  - $Z_{eff}$: 75.2
  - Density: 8.28 g/cm$^3$
  - Attenuation length: 8.7 mm
  - Thickness: 5 mm (~50% gamma interaction)
- Theoretical probability (511 keV) of
  - Photoelectric absorption: 42.7%
  - Compton scattering: 51%
  - Rayleigh scattering: 6.3% – matching NIST XCOM



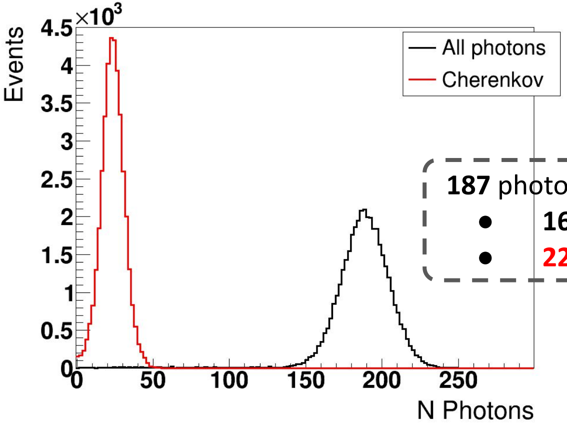
# Optical Photon and Photocathode Simulation

- Scintillation: 330 photons/MeV
  – M. Follin, *et al.* (2021)
  - Fast (58.6%): $\boldsymbol{\tau}_f$ = 1.79 ns
  - Slow (41.4%): $\boldsymbol{\tau}_s$ = 6.41 ns
  - Spectrum peak at 400 nm
  – M. Shao, *et al.* (2001)

- Optical photon propagation:
  - Absorption length of $PbWO_4$
  - Refractive indexes

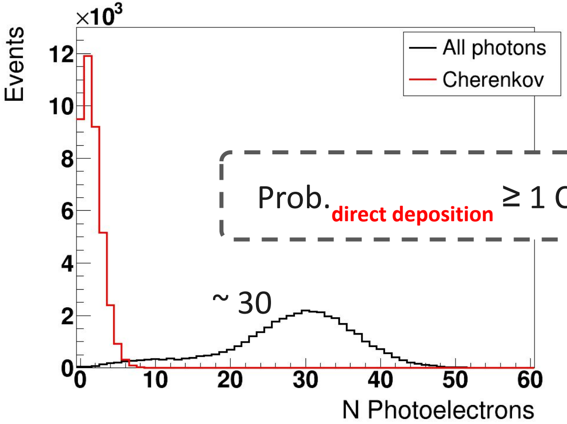
- Photocathode simulation → **QE = 30%**
  - <u>Photon absorption probability</u> in Bialkali
  - <u>Photoelectron extraction probability</u>
  – Motta *et al.* (2005)
  – Sung *et al.* (2022)

**Photon production**



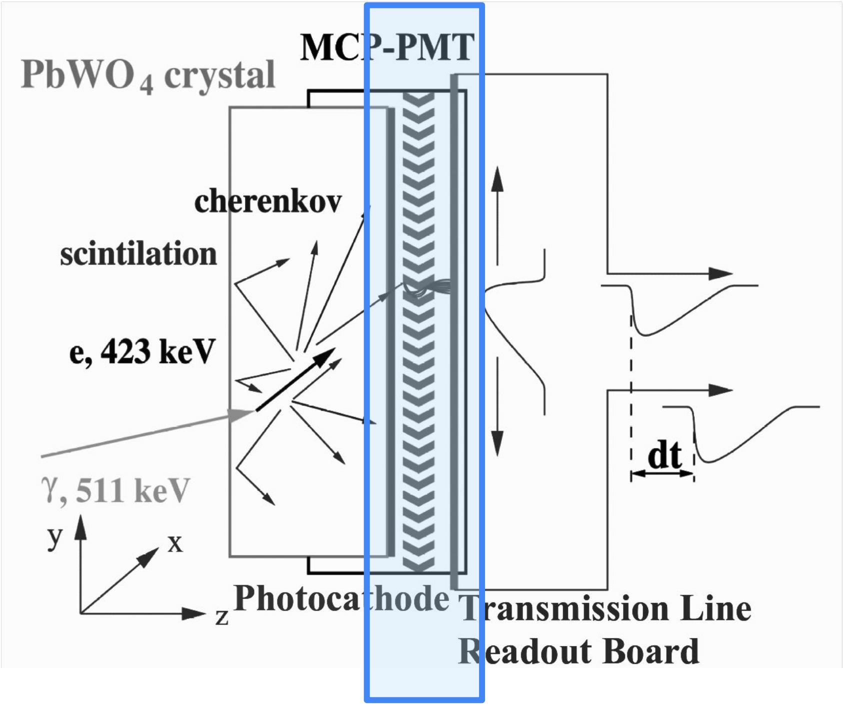

**187** photons (511 keV deposition)
- **165** scintillations
- **22** Cherenkov photons

**Photon detection**



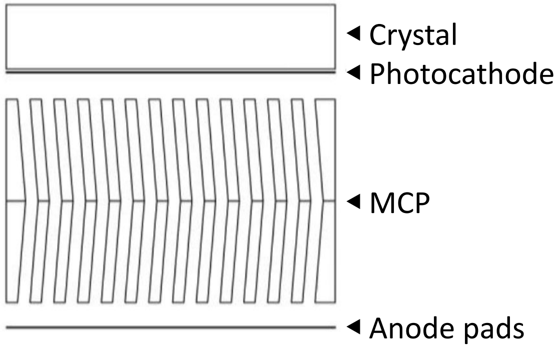

Prob.$_{\text{direct deposition}}$ ≥ 1 Che./event **> 75%**

# MCP-PMT Simulation

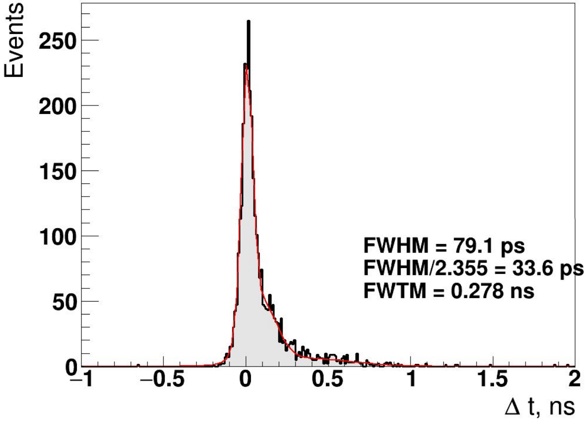


# MCP-PMT Simulation

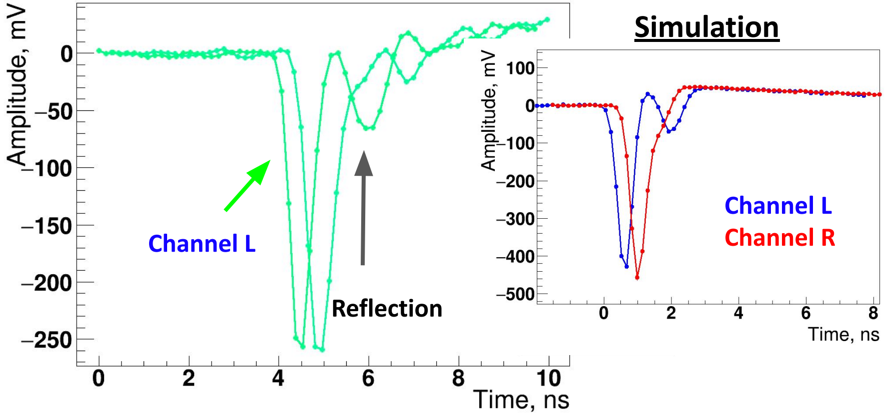
- MCP-PMT process calculation
  1. Time response (time transit spread)
  2. Gain and fluctuation
  3. Charge sharing effect through charge induction
  4. Signal propagation & amplification
  5. Signal (shape) digitization
- PMT model tuned on the experimental data




**Time response of MCP-PMT**

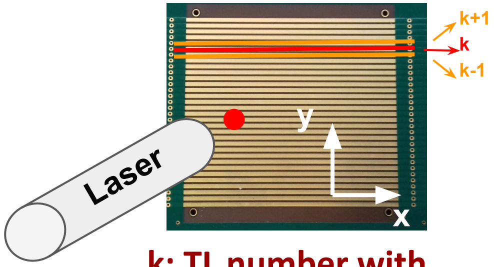



**Position dependent signal shape**



# Detector Simulated Performance

**MAPMT253**




**k: TL number with maximum amplitude**

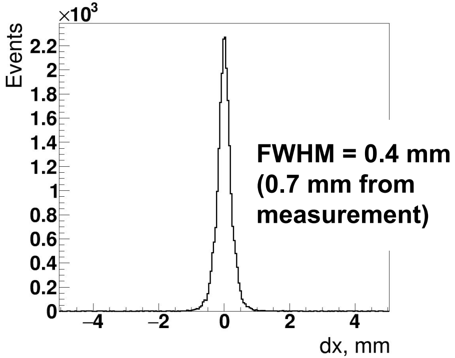
$$x_r = \frac{(t_R - t_L) \times s}{2}$$

$$y_r = \frac{\sum_{k=i-1}^{i+1} y_k C_k}{\sum_{k=i-1}^{i+1} C_k}$$

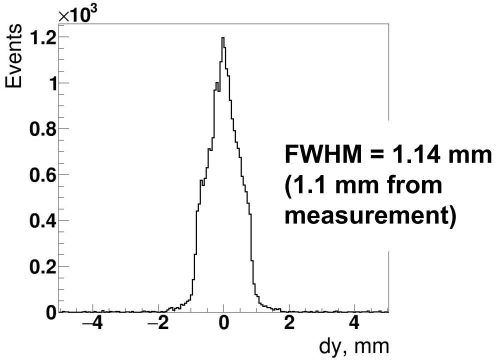
- $t_R$: time from signal of channel R
- $t_L$: time from signal of channel L
- s: propagation speed (35% c)
- $y_k$: coordinate of $TL_k$
- C: charge

## One-photon resolustion

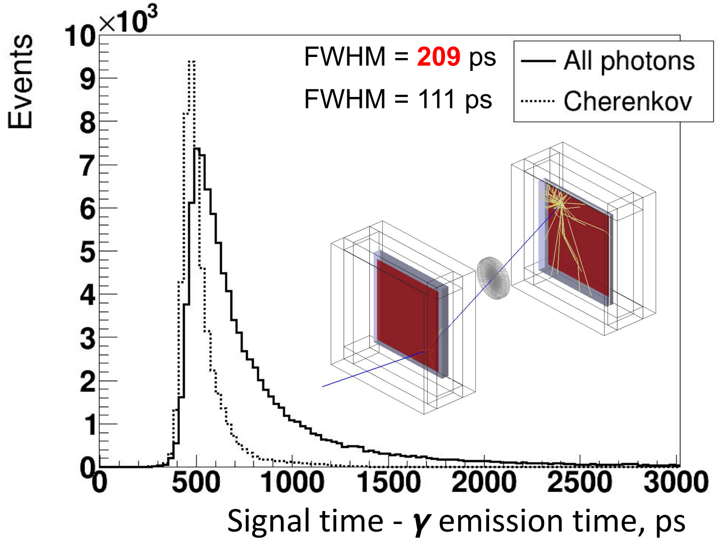
### X-coordinate Error




### Y-coordinate Error

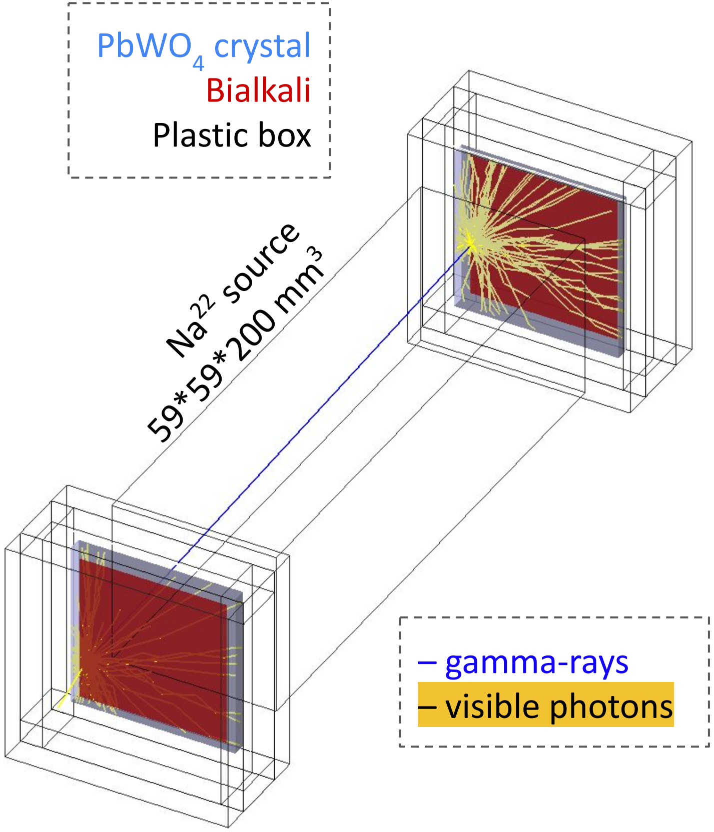



## Simulated Time Resolution of gamma detection

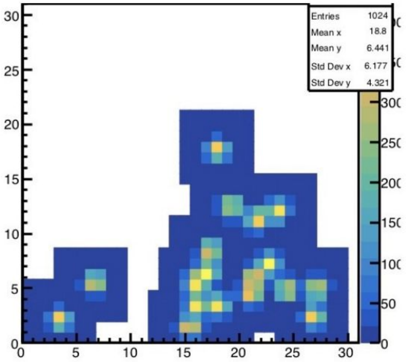


# Gamma Conversion Position Reconstruction

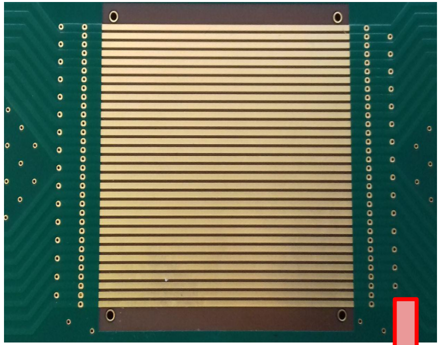
# Training Data Preparation – Simulation



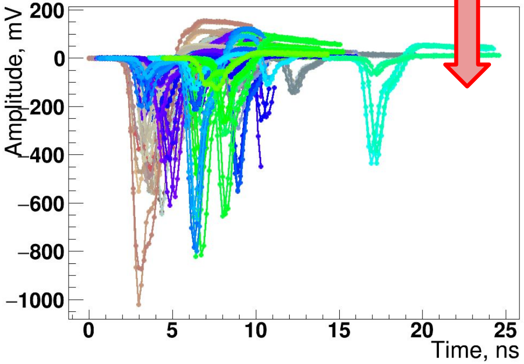

**Photon distribution & detection**



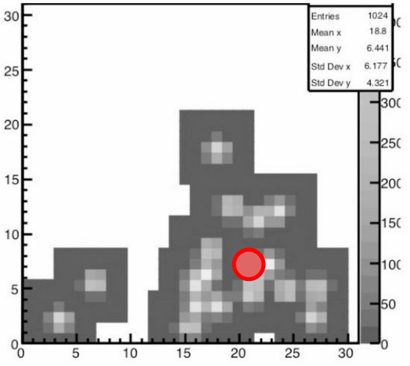

**Transmission line readout**



**64-channel signal registration**

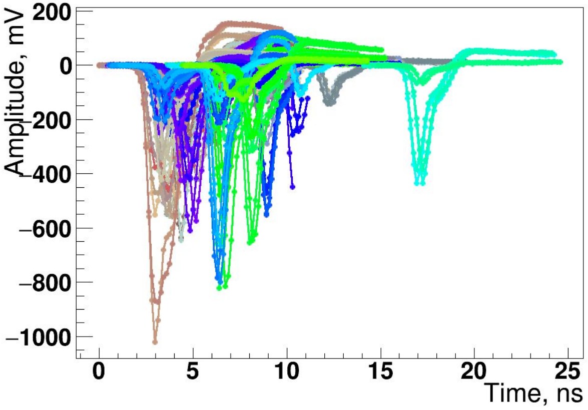


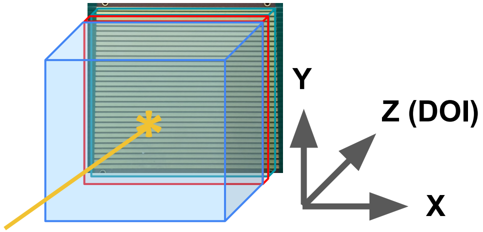

**Reconstructed $\gamma$ position**



# Training Data Preparation – Input & Target

**64-channel signal registration**




**Input**

**Only Pre-processed Input:**
- **Signal time difference in line (*32)**
- **Charge (*32)**
- **DOI estimators**
- **Statistical reconstruction**

**Unprocessed + Pre-processed Input:**
- ***Signals* (*64)**
- **Signal time (*32)**
- **Signal start time (*32)**

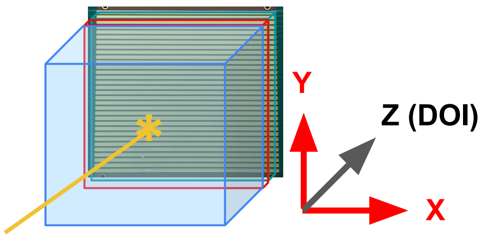
DM2S

**Prediction**

**- 3D coordinates**
of gamma conversion

**- 2D coordinates + σ**
- DOI **+ σ**
- Time **+ σ**
- Energy **+ σ**
of gamma conversion

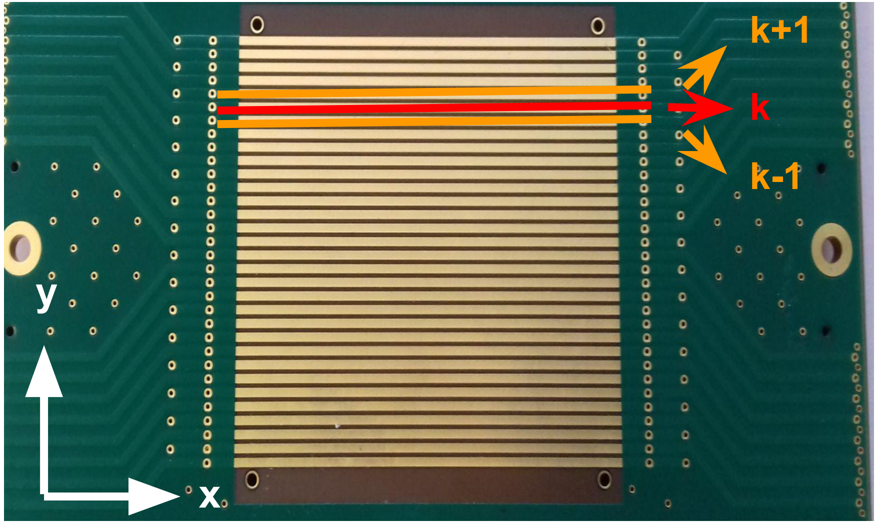
σ: uncertainty of each output




# 2D-Coordinate Reconstruction

- Statistical method
- Machine learning
  - Pre-processed input
  - Un-processed + pre-processed input

# Gamma 2D Reconstruction Statistical Method




$$x_r = \frac{(t_{R,k} - t_{L,k})}{2} \times s$$

$$y_r = \frac{\sum_{k=i-1}^{i+1} y_k C_k}{\sum_{k=i-1}^{i+1} C_k}$$

- ***k*: TL number with maximum amplitude**
- $t_R$: time from signal of channel R
- $t_L$: time from signal of channel L
- *s*: propagation speed
- $y_k$: coordinate of $TL_k$
- *C*: charge

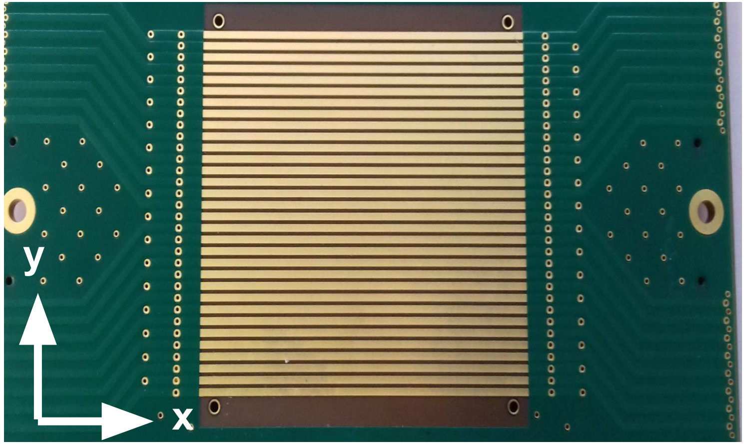
# Gamma 2D Reconstruction Algorithms with **Pre-processed** Input

- Algorithms: **Decision tree**, **Neural Network**
- Package: ROOT TMVA v.6.18/04
- Training Samples: 100k events
- Test Samples: 100k events
- Loss function: **Mean squared error**

| | X | Y |
|---|---|---|
| ClearMind Prototype | $C_i$, $x_i$, $x_r$ | $C_i$, $y_r$ |
| Variable transform | Gaussian | Normalization |

- $C_i$: charge on triggered lines
- $x_i$: reconstructed x position on each line
- $x_r$, $y_r$: reconstructed position using statistical method
- one 511 keV gamma-conversion selection

**Transmission lines**




$$x_i = \frac{(t_{R,i} - t_{L,i})}{2} \times s$$

$$x_r = \frac{(t_{R,k} - t_{L,k})}{2} \times s$$

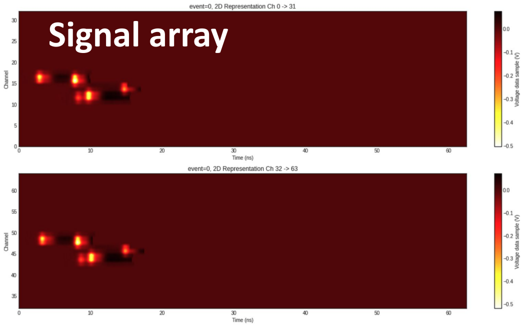
$$y_r = \frac{\sum_{k=i-1}^{i+1} y_k C_k}{\sum_{k=i-1}^{i+1} C_k}$$

- Algorithms: **Convolutional Neural Network**
- Package: Tensorflow/Tensorflow probability
- Training Samples: >1M events
- Test Samples: >1M events

- Loss function **Geoffrey Daniel – 10h50 (Tuesday)**

$$l_w = \log\left(\sqrt{2\pi} \cdot \sigma_i\right) + \frac{1}{2}\frac{(y_i - \hat{y})^2}{\sigma_i^2}$$

**INPUT**

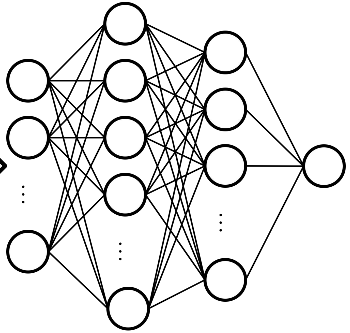



Time of the signals

⇒ **1D-convolution** ⇒

⇒ **Dense NN** ⇒

Flattening ⇒



**PREDICTION**

( $y_i$ , $\sigma_i$ )

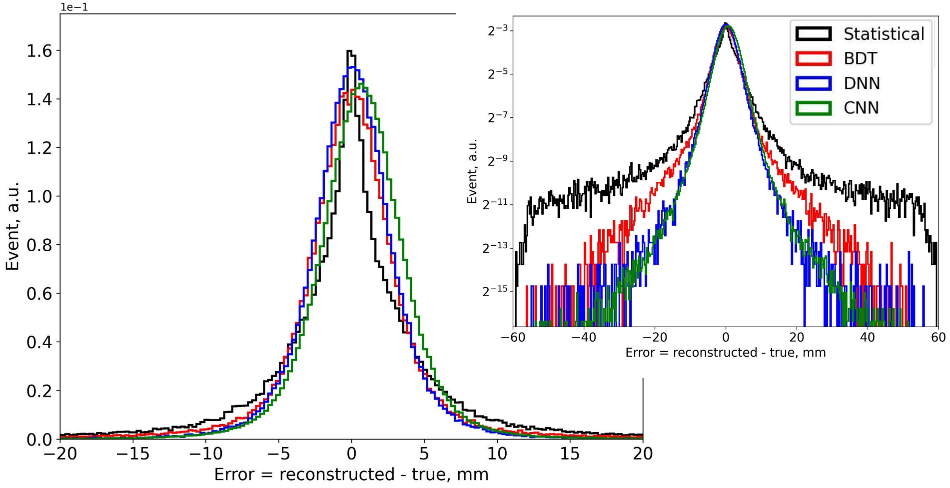
μ: 3D positions, etc

σ: learning uncertainty

DE LA RECHERCHE À L'INDUSTRIE

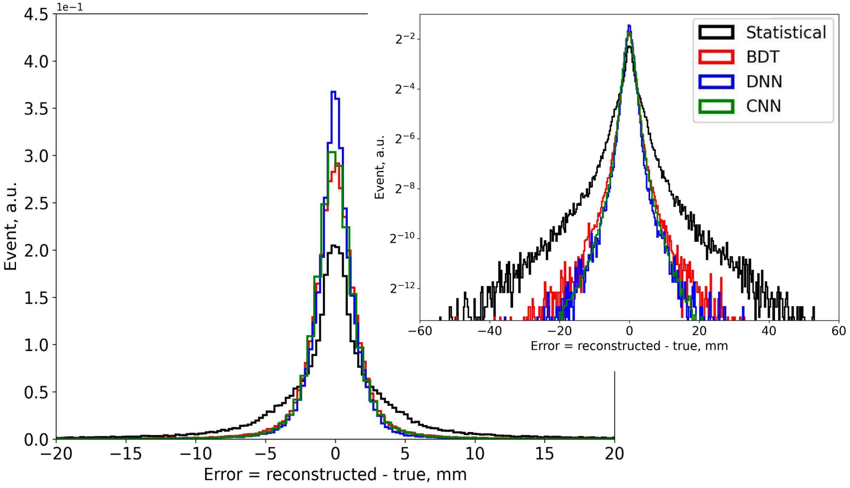
# 2D Reconstruction Results

## X-coordinate (along lines) error

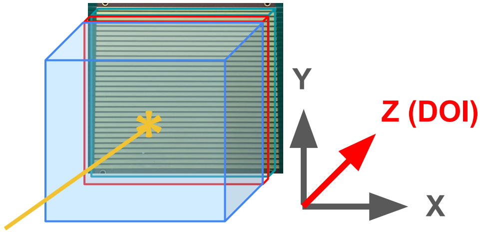



| | FWHM, mm | SD, mm | Fraction of outlier |
|---|---|---|---|
| Statistical | 2.9 | 5.1 | 23% |
| BDT | 5.8 | 4.0 | 14% |
| DNN | 5.5 | 3.3 | 10% |
| CNN | 5.7 | 3.7 | 11% |

## Y-coordinate (across lines) error




| | FWHM, mm | SD, mm | Fraction of outlier |
|---|---|---|---|
| Statistical | 2.5 | 4.5 | 17% |
| BDT | 2.7 | 2.5 | 4% |
| DNN | 2.0 | 2.0 | 3% |
| CNN | 2.5 | 2.4 | 3% |




# Depth-of-Interaction Reconstruction

- Statistical method
- Machine learning
  - Pre-processed input

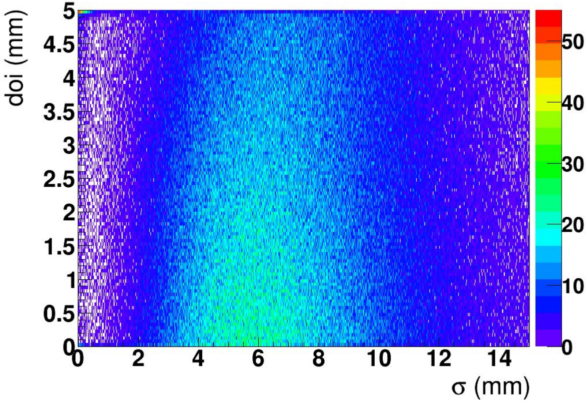
## Statistical Method
(spread of photons)

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{32} (x_i - \bar{x})^2 \cdot C_i}{\alpha \cdot \sum_{i=1}^{32} C_i}} \quad \sigma_y = \sqrt{\frac{\sum_{i=1}^{32} (y_i - \bar{y})^2 \cdot C_i}{\alpha \cdot \sum_{i=1}^{32} C_i}}$$

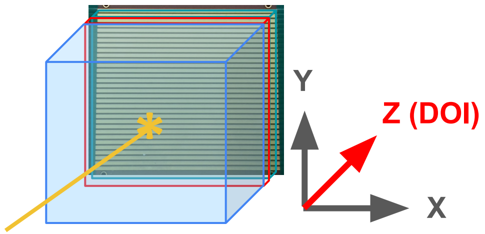
$$\alpha = 1 - \frac{\sum_{i=1}^{32} C_i^2}{\left(\sum_{i=1}^{32} C_i\right)^2}$$

- $i$: triggered TL number (1-32)
- $x_i$: reconstructed x per TL
- $y_i$: coordinate of $TL_i$
- $\bar{x}, \bar{y}$: weighted average x- & y-coordinate

## DOI vs. σ




## Machine Learning Method



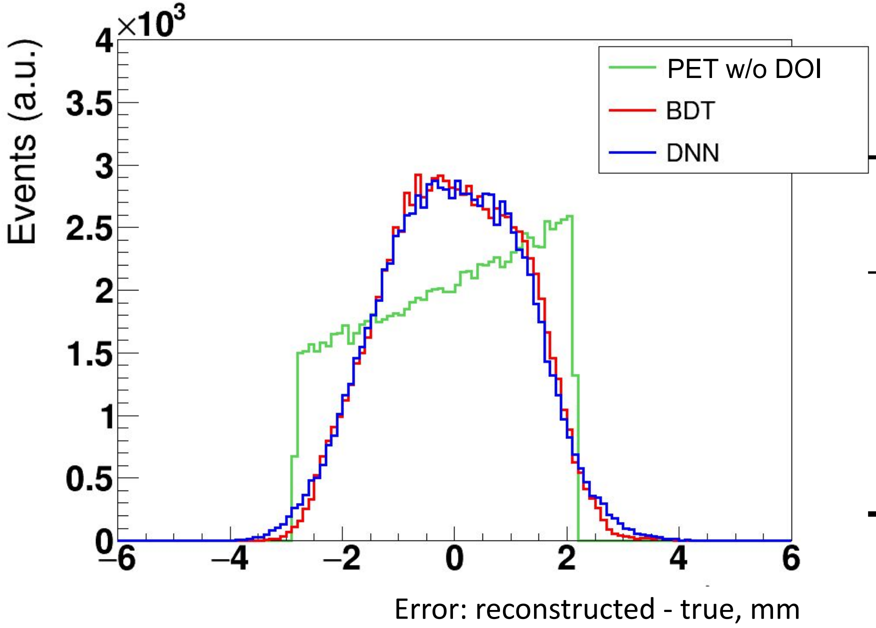

- Algorithms: **Decision tree**, **Neural Network**
- Package: ROOT TMVA v.6.18/04
- Training Samples: 97k events
- Test Samples: 97k events
- Loss function: **Mean squared error**

| | DOI |
|---|---|
| Variables | $C_i$, $\sigma_x$, $\sigma_y$ |
| Variable transform | Normalization |

# DOI Reconstruction Results

511 keV gamma

DOI error




| | FWHM, mm | SD, mm | Fraction of outlier |
|---|---|---|---|
| Conventional | 4.9 | 1.4 | 36% |
| BDT | 3.5 | 1.2 | 21% |
| DNN | 3.4 | 1.2 | 23% |

# Conclusion

- 3D spatial resolution in FWHM **To be improved**

| X (along the lines) | Y (across the lines) | DOI |
|---|---|---|
| 5.5 mm | 2.0 mm | 3.5 mm |

– Sung *et al.* (2022) arxiv:2209.11587 [physics.ins-det]


- Energy resolution → **Work in progress**
- Time resolution

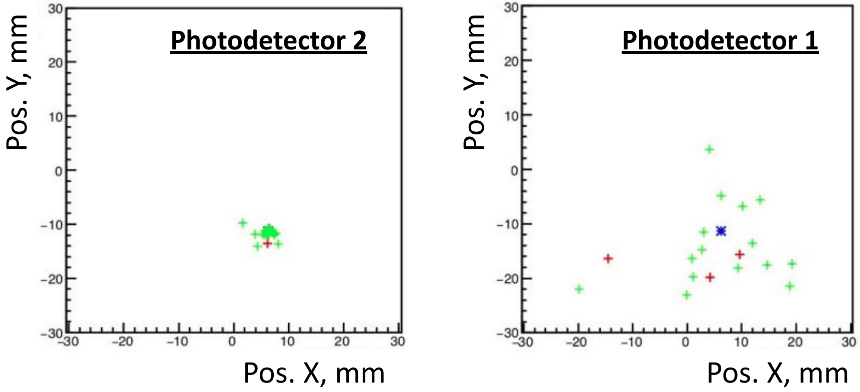
→ For future detector configuration

1) 10-mm $PbWO_4$→ gamma detection efficiency

2) 2 photodetectors → Cherenkov det. efficiency

**Better DOI estimation & Time resolution**

**Detected photon distribution**

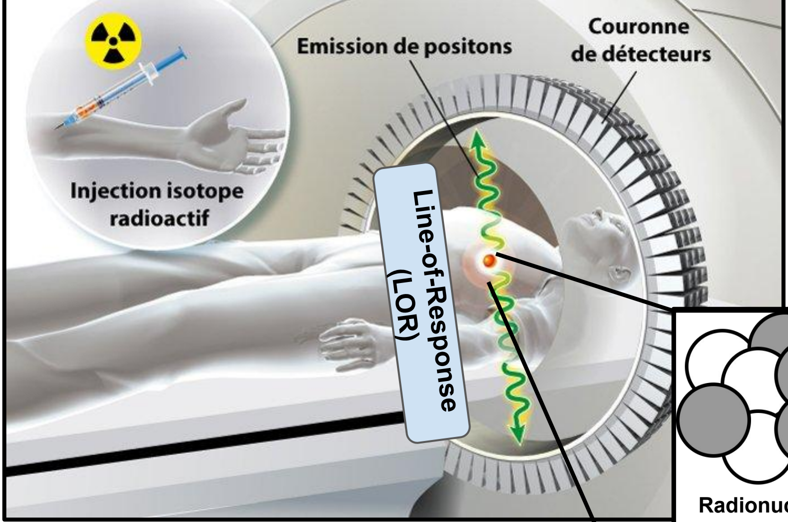


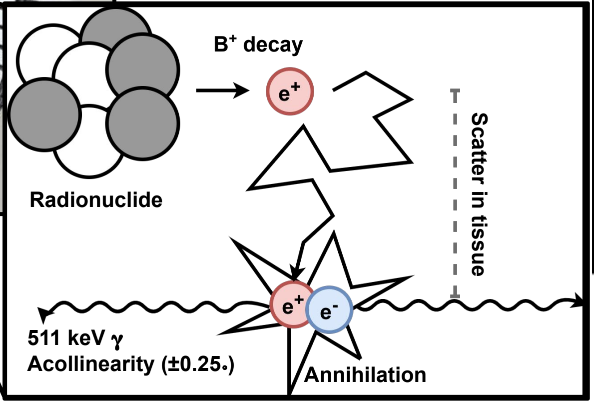
Back up

# Positron Emission Tomography

## PET Scan protocal

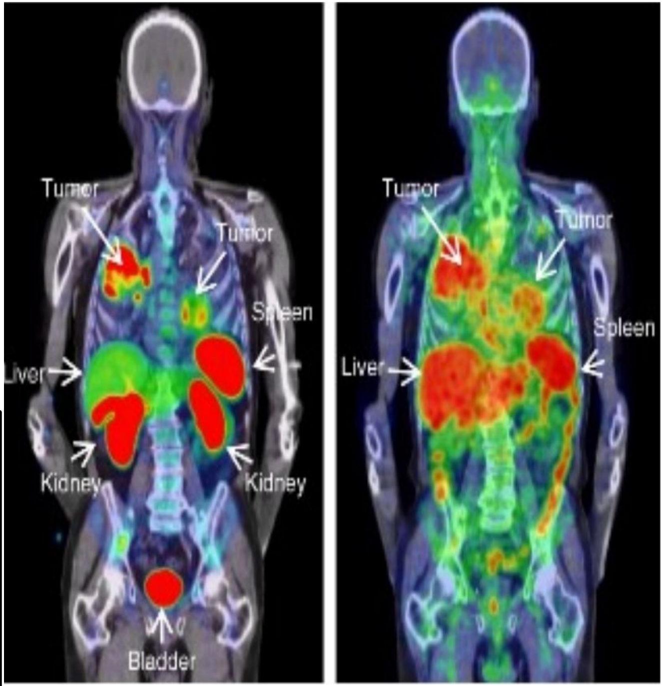
Tracer (positron emitter with short half-life): [$^{18}$F]FDG ..




+ **Image Reconstruction** =

## PET-CT Image





Ramsay Santé (https://ramsaysante.fr/vous-etes-patient-en-savoir-plus-sur-ma-pathologie/pet-scan)
M. Conti, B. Bendriem, The new opportunities for high time resolution clinical TOF PET, Clin. Transl. Imaging. (2019) 139–147. https://doi.org/10.1007/s40336-019-00316-5.

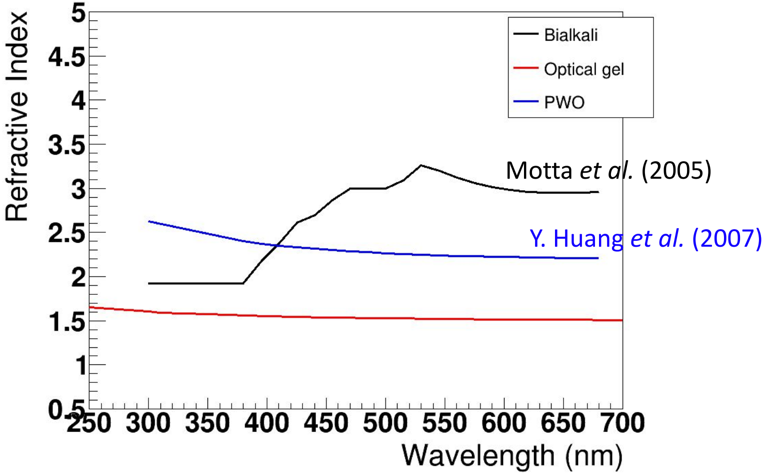
# Photon Propagation Process

**Absorption** → Absorption Length ($PbWO_4$)

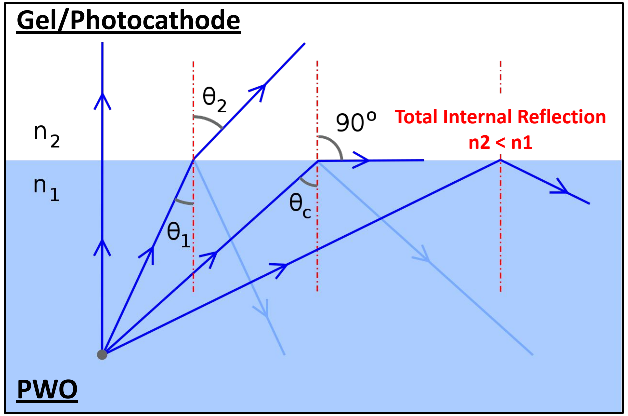
Annenkov *et al.* (2002)

**Reflection**

**Refraction (Transmission)**




**Optical Contact of PWO & Gel/Photocathode**




https://en.wikipedia.org/wiki/Total_internal_reflection#cite_note-10

- Optical gel → Total Internal Reflection (TIR) for all wavelength: $n_1 > n_2$ (with a critical $\theta$)
- Direct deposition → TIR reduction in all wavelength (including < 400 nm)

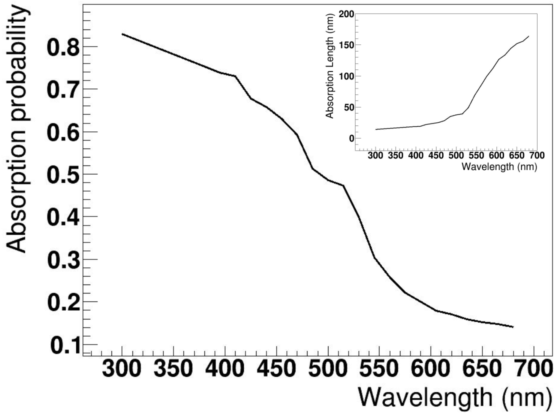
# Photocathode Simulation Input

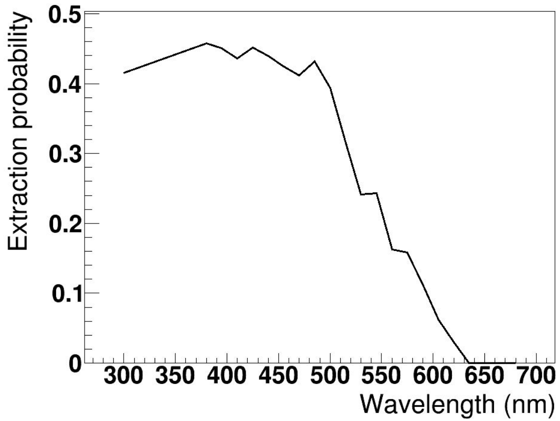
**Simple Simulation**
- Complete photon absorption
- Quantum efficiency function

**Realistic Simulation**

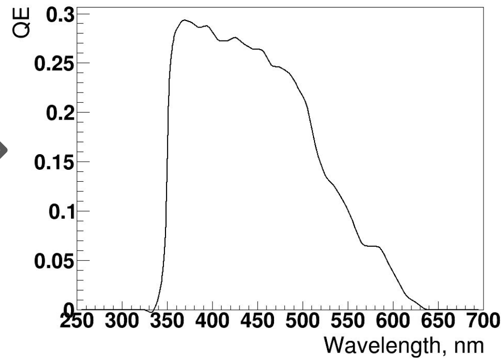
Bialkali photocathode
- Absorption probability
- Extraction Probability
- Refractive index

– Motta *et al.* (2005)




Quantum Efficiency (90∘) = 30%



# Reconstruction - Machine Learning Method

**Decision Tree**

| | | | |
|---|---|---|---|
| Train samples | 50k - 592k | | |
| Test samples | 50k - 592k | | |
| Learning rate | 0.005 - 0.1 | | |
| | X | Y | DOI |
| Var. transform | Gaussian | Normalization | |
| Variables | $C_i$, $x_i$ | $C_i$ | $C_i$, $\sigma_x$, $\sigma_y$ |

- $C_i$: charge on triggered lines
- $x_i$: reconstructed x position on each line $x_i = \frac{(t_{R,i} - t_{L,i})}{2} \times s$
- $\sigma_x$, $\sigma_y$: photon spread in x/y direction
- $x_r$, $y_r$: reconstructed position using statistical method $x_r = \frac{(t_{R,k} - t_{L,k})}{2} \times s$, $y_r = \frac{\sum_{k=i-1}^{i+1} y_k C_k}{\sum_{k=i-1}^{i+1} C_k}$
- one 511 keV gamma-conversion selection

**Neural Network**

| | | | |
|---|---|---|---|
| Train samples | 50k - 592k | | |
| Test samples | 50k - 592k | | |
| Hidden layers | 4 - 6 | | |
| Neurons/layer | 100 - 500 | | |
| Activation func. | RELU | | |
| | X | Y | DOI |
| Var. transform | Gaussian | Normalization | |
| Variables | $C_i$, $x_i$, $x_r$ | $C_i$, $y_r$ | $C_i$, $\sigma_x$, $\sigma_y$ |

# Reconstruction - Machine Learning Method

Table 4.2: Training parameters of DNN model

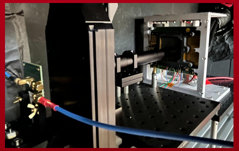
| Configuration | Simplified Detector | | | CM Prototype | | |
|---|---|---|---|---|---|---|
| Coordinates | X | Y | DOI | X | Y | DOI |
| **Parameters** | | | | | | |
| Train samples (events) | 592k | 50k | 119k | 100k | 100k | 97k |
| Test samples (events) | 592k | 50k | 119k | 100k | 100k | 97k |
| Hidden layers | 6 | 5 | 4 | 6 | 4 | 4 |
| Neurons per layer | 300 | 500 | 300 | 100 | 300 | 100 |
| Activation function | RELU | | | | | |
| Batch size | 10 | 64 | 32 | 10 | 10 | 10 |
| Variable transform | Gaussian | Normalization | | Gaussian | Normalization | |
| Variables | $Charge_i$ $x_i$ | $Charge_i$ | $Charge_i$ $\sigma_x$ $\sigma_y$ | $Charge_i$ $x_i$ $x_R$ | $Charge_i$ $y_R$ | $Charge_i$ $\sigma_x$ $\sigma_y$ |
| Strategy I | | | | | | |
| Learning rate | 5.e-4 | 1.e-3 | 1.e-3 | 5.e-4 | 5.e-4 | 5.e-4 |
| Convergence steps | 34 | 15 | 15 | 34 | 9 | 9 |
| Regularization | L2 | None | None | L2 | None | None |
| Weight decay | $5\times10^{-6}$ | 0 | 0 | $5\times10^{-6}$ | $1\times10^{-6}$ | $1\times10^{-6}$ |
| Momentum | 0.5 | 0 | 0 | 0.5 | 0 | 0 |
| Dropout | 10% | 0 | 0 | 10% | 0 | 0 |
| Strategy II | | | | | | |
| Learning rate | 2.e-5 | 1.e-4 | 1.e-4 | 2.e-5 | 2.e-5 | 2.e-5 |
| Convergence steps | 34 | 20 | 20 | 34 | 14 | 9 |
| Regularization | L2 | None | None | L2 | None | None |
| Weight decay | $1\times10^{-6}$ | 0 | 0 | $1\times10^{-6}$ | $1\times10^{-6}$ | $1\times10^{-6}$ |
| Momentum | 0.5 | 0 | 0 | 0.5 | 0 | 0 |
| Dropout | 10% | 0 | 0 | 10% | 1% | 1% |
| Strategy III | | | | | | |
| Learning rate | 9.e-6 | 1.e-5 | 1.e-5 | 9.e-6 | 1.e-6 | 1.e-6 |
| Convergence steps | 24 | 35 | 40 | 24 | 19 | 14 |
| Regularization | L2 | None | None | L2 | None | None |
| Weight decay | $1\times10^{-6}$ | 0 | 0 | $1\times10^{-6}$ | $1\times10^{-6}$ | $1\times10^{-6}$ |
| Momentum | 0.5 | 0 | 0 | 0.5 | 0 | 0 |
| Dropout | 2% | 0 | 0 | 2% | 2% | 2% |
| Strategy IV | | | | | | |
| Learning rate | 1.e-6 | | | 1.e-6 | 5.e-7 | |
| Convergence steps | 24 | | | 24 | 49 | |
| Regularization | L2 | | | L2 | None | |
| Weight decay | $1\times10^{-6}$ | | | $1\times10^{-6}$ | | |
| Momentum | 0.5 | | | 0.5 | 0 | |
| Dropout | 2% | | | 2% | 0 | |
| Strategy V | | | | | | |
| Learning rate | | | | | 1.e-7 | |
| Convergence steps | | | | | 49 | |
| Regularization | | | | | None | |
| Weight decay | | | | | $1\times10^{-6}$ | |
| Momentum | | | | | 0 | |
| Dropout | | | | | 0 | |

Table 4.1: Training parameters of GBDT model

| Configuration | Simplified Detector | | | CM Prototype | | |
|---|---|---|---|---|---|---|
| Parameters | X | Y | DOI | X | Y | DOI |
| Train samples (events) | 592k | 50k | 75k | 100k | 100k | 97k |
| Test samples (events) | 592k | 50k | 75k | 100k | 100k | 97k |
| Maximum trees | 2000 | 2000 | 500 | 3000 | 2500 | 2000 |
| Maximum tree depth | 100 | 30 | 10 | 10 | 1000 | 100 |
| Seperation type | RegressionVariance | | GiniIndex | RegressionVariance | | |
| Shrinkage factor | 0.01 | 0.01 | 0.1 | 0.005 | 0.01 | 0.01 |
| Tree pruning method | CostComplexity | | | | | |
| Pruning strength | 50 | 30 | 20 | 80 | 500 | 300 |
| Variable transform | Gaussian | Normalization | | Gaussian | Normalization | |
| Variables* | $Charge_i$ $x_i$ | $Charge_i$ | $Charge_i$ $\sigma_x$ $\sigma_y$ | $Charge_i$ $x_i$ $x_R$ | $Charge_i$ $y_R$ | $Charge_i$ $\sigma_x$ $\sigma_y$ |

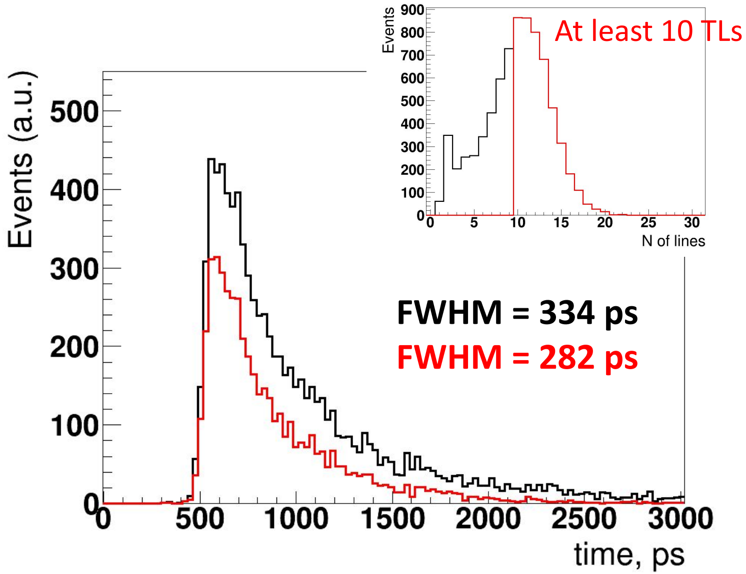
* $i$ indicates the all TL numbers

# Time Resolution



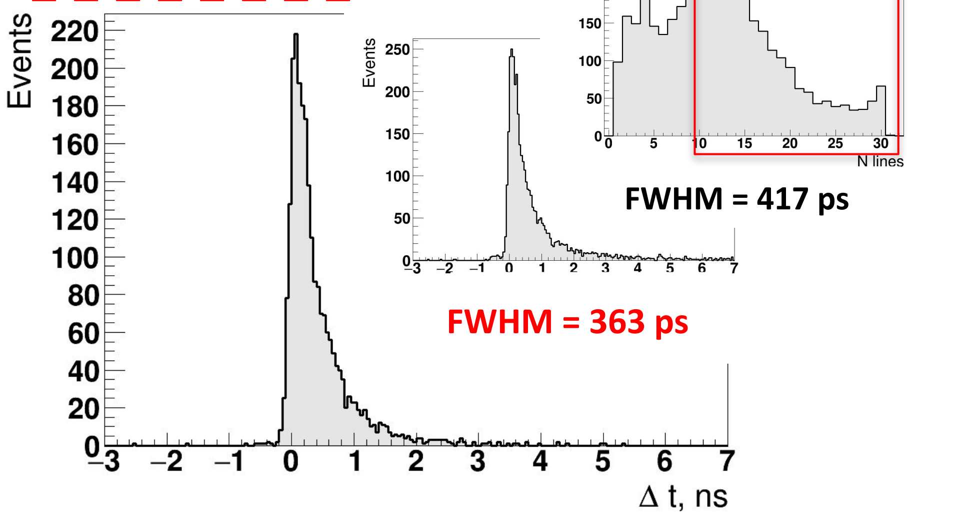
Preliminary

## Simulation 40%QE




## Measurement
## CTR of CM prototype & SiPM (131 ps)

$$TR_{CM} = \sqrt{CTR^2 - TR_{ref}^2}$$




- Better time resolution with event selection → still can’t distinguish the events with Cherenkov
- The difference in time resolution → to be investigated (events with many triggered TLs)