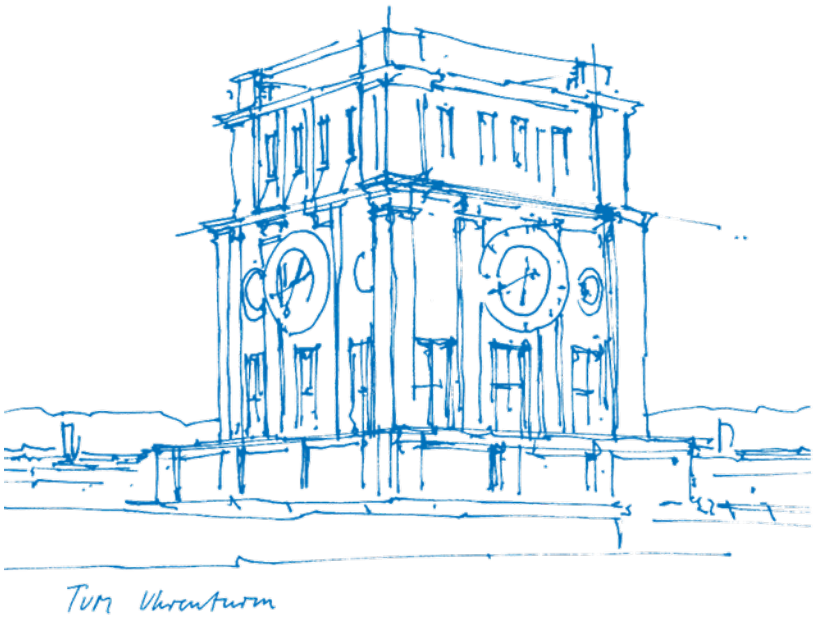
Development of a Neural Network for Online Event Reconstruction for a Radiation Monitor

Lukas Bierwirth

TU München

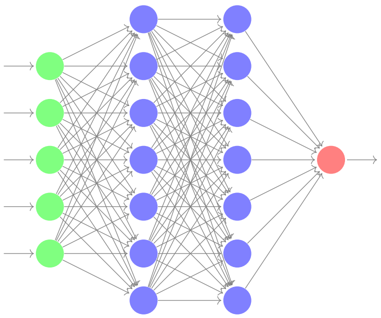
Physics Department

Munich, 3rd December 2019



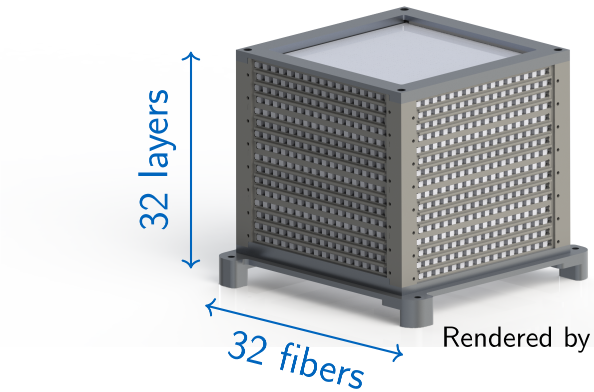
Motivation

Rendered by M. Losekamm

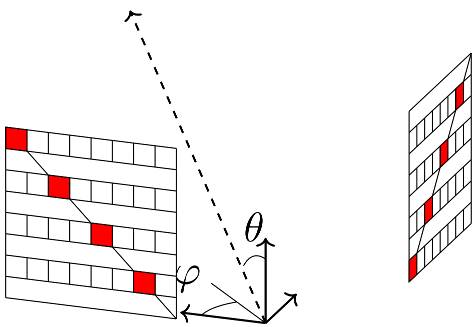


- People involved:
  - Stephan Paul
  - Martin Losekamm
  - Thomas Pöschl
- The detector:
  - Monitor radiation environment in outer space
    - High rates
    - Omnidirectional flux
- The goal:
  - Online reconstruction of particles' parameters from raw detector data:
    - Track parameters (direction)
    - Energy
    - Particle species
- The method:
  - Artificial Neural Networks (ANNs)

Radiation Monitor: MAPT



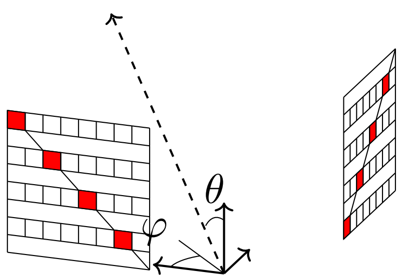
Rendered by M. Losekamm

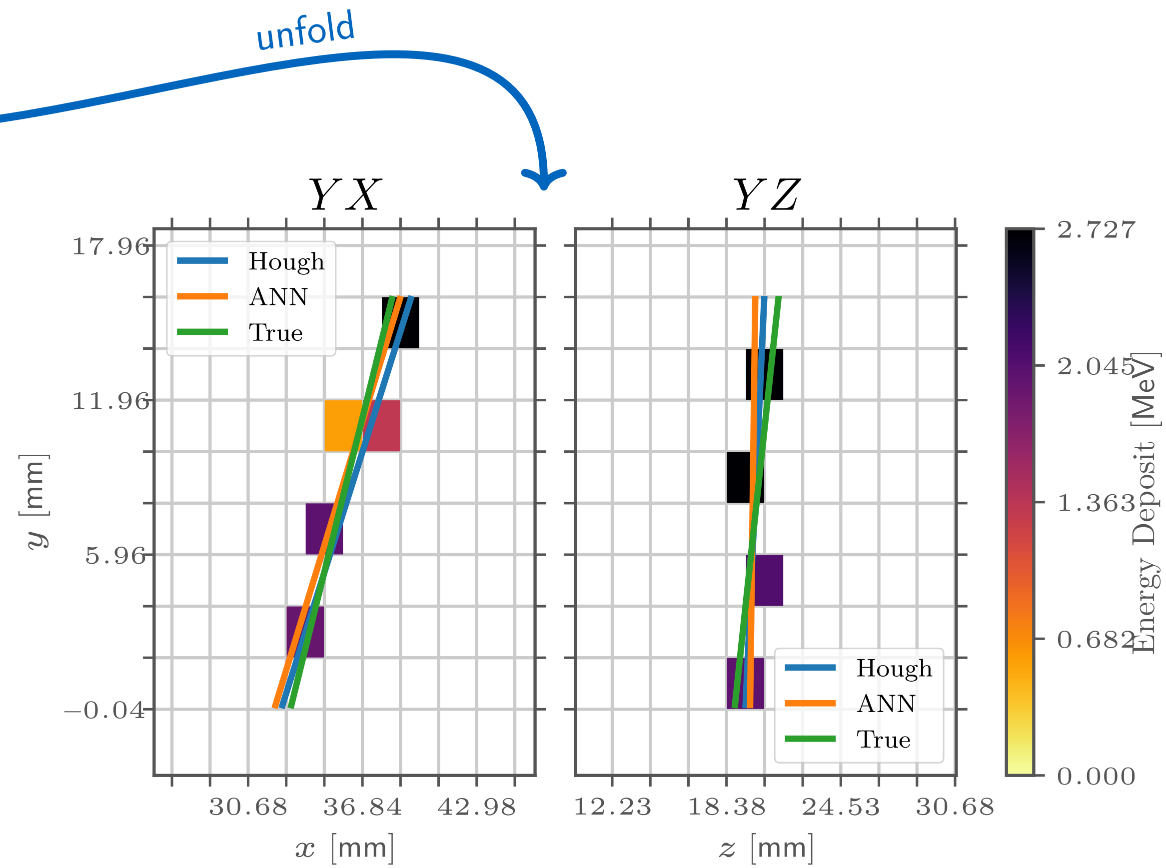


Multi-purpose Active-target Particle Telescope (MAPT):

- Segmented calorimeter
  - Omnidirectional acceptance
  - 32 Layers
- Layer:
  - 32 fibers
  - Rotated by 90° w.r.t. the previous layer
  - Shifted horizontally w.r.t. previous layer of equal orientation
- Fiber:
  - Active Core:
    - Polystyrene
    - $1.92\,\mathrm{mm} \times 1.92\,\mathrm{mm} \times 70\,\mathrm{mm}$
  - Sputtered with aluminium
- Signal:
  - Two orthogonal projections of the track

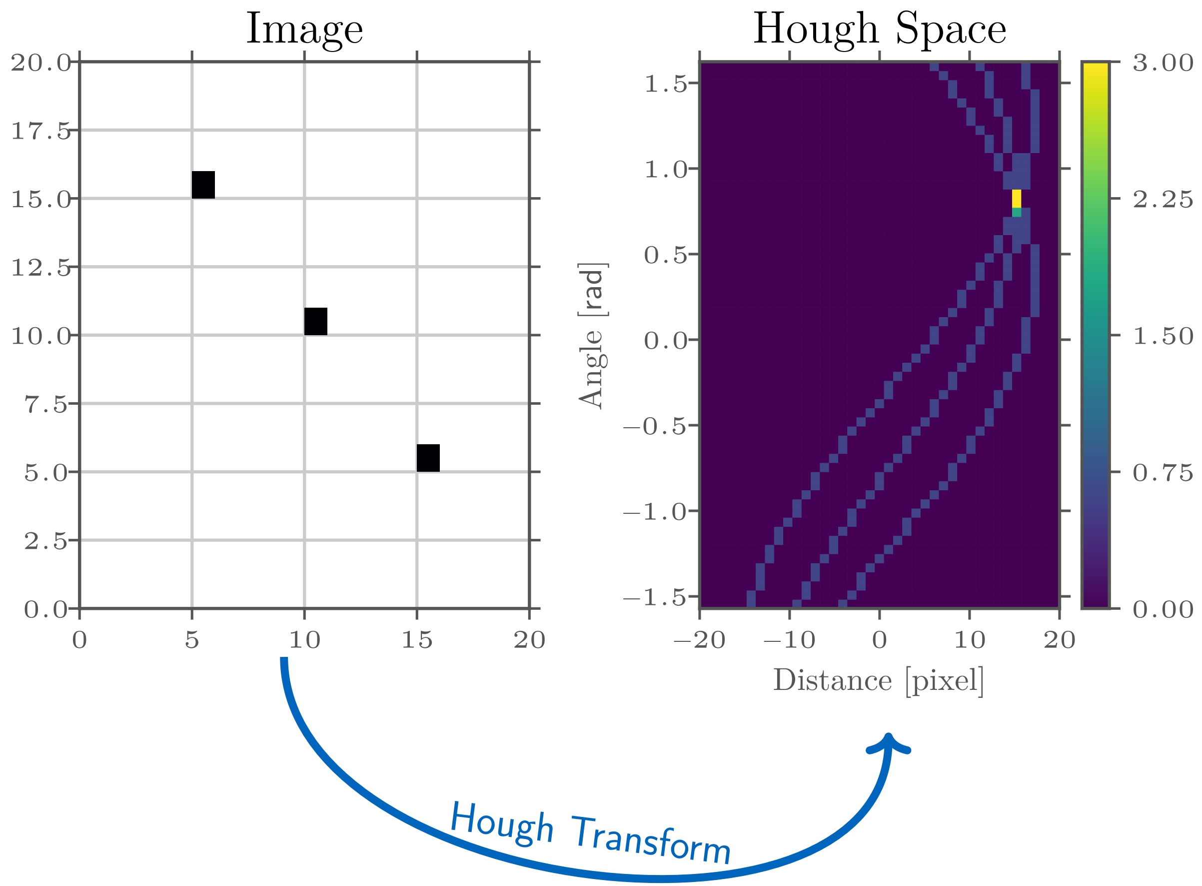
Radiation Monitor: Signal





- Particle deposits energy in active material along its path
- Active material converts energy to photons
- Photons collected by SiPMs $\Rightarrow$ electrical signal proportional to energy deposit (non-linear effects!)

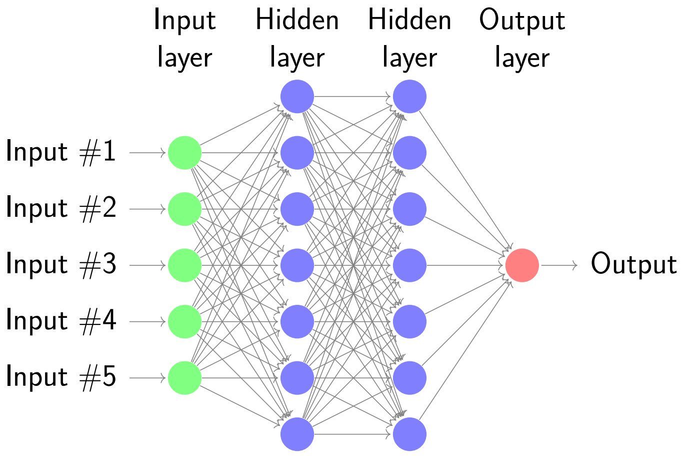
Traditional Method: Hough Transform



- Standard line-detection method in computer vision: fast and robust
- Maps every pixel of the input image to one sinusoid in the Hough space
- Iff sinusoids meet at one point, the pixels lie on one straight line in the input image
- In Hough space: line detection by peak finding

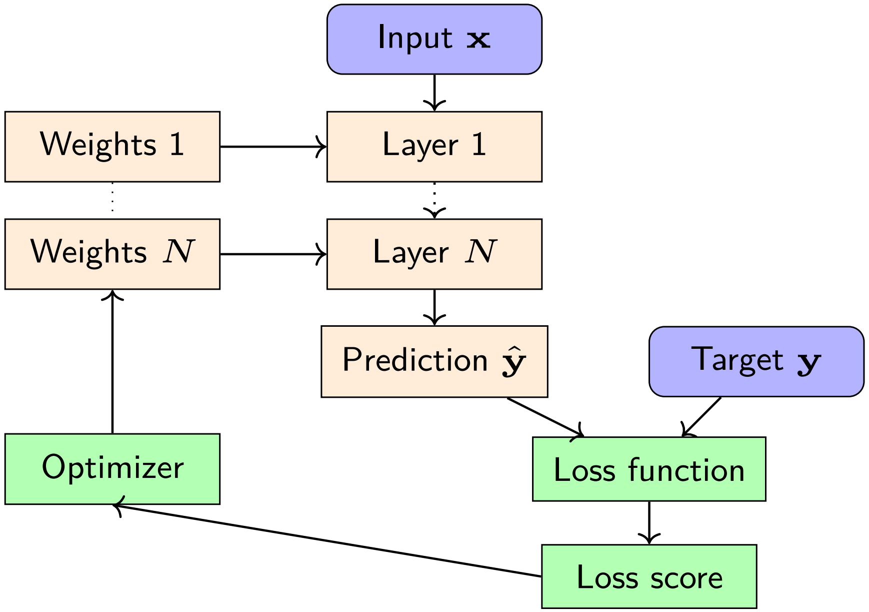
Method: ANN

- Powerful pattern-recognition algorithm
- Implementation of a (non-linear) function parametrized by a set of weights
- Practical usage: weights can be adjusted efficiently via optimization $\Rightarrow$ Supervised Learning



- Examples:
  - Online translation services,
  - snapchat filters,
  - speech recognition
- Structure:
  - Neurons form layer
  - Layers form networks
- Different types of architecture:
  - CNNs (Convolution),
  - RNNs (Recurrent $\Rightarrow$ information loops),
  - mixtures

ANN: Training

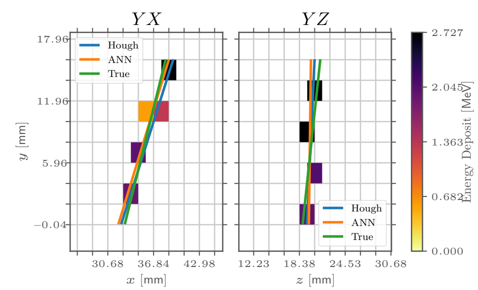


Supervised Learning:

- Given: Labeled data point (x, y)
- ANN makes prediction
- Loss function compares prediction $\hat{y}$ to target y
- Weights adapted via gradient descent to minimize loss
- Important: Sufficient amount of labeled data points!

In our case:

- Input x given by the two projections:

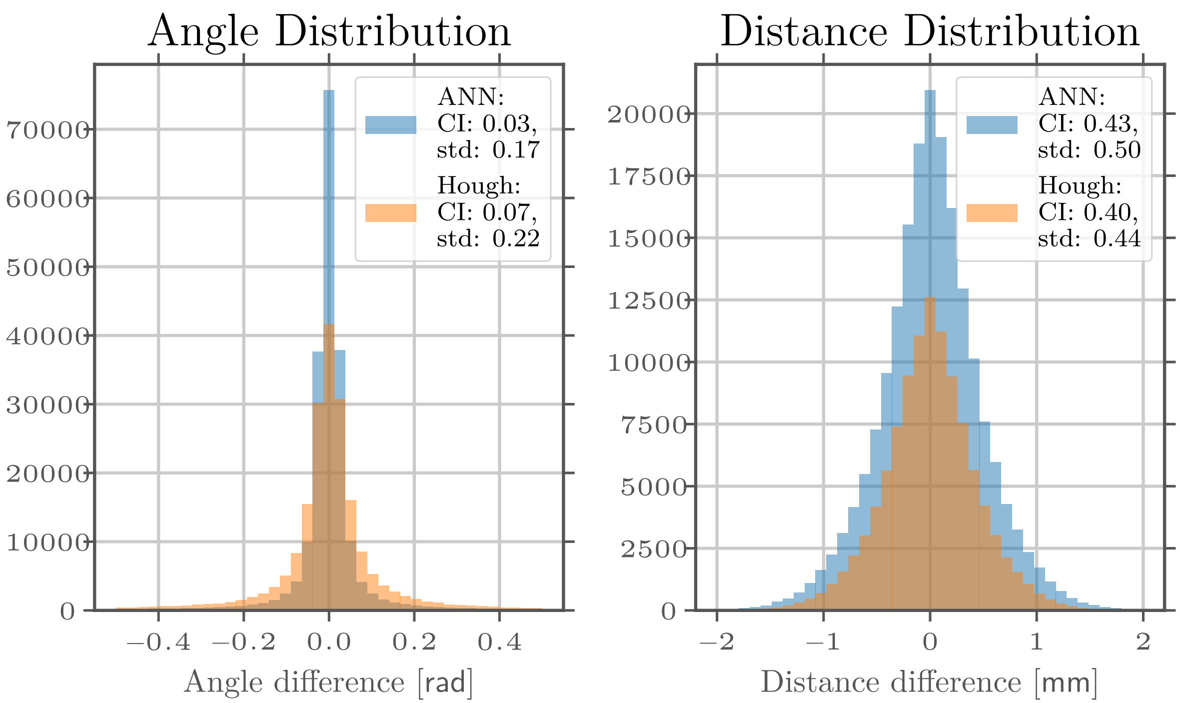


- For e.g. energy reconstruction:
  - Target $y = E_{true}$
  - Prediction $\hat{y} = E_{reco}$
  - Loss function $L \propto E_{true} - E_{reco}$

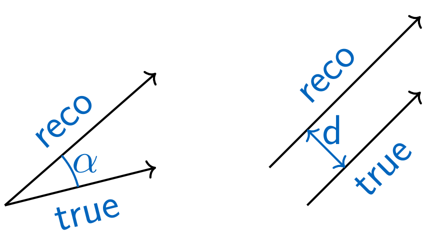
Supervised Learning with Simulated Data

- Idea:
  - Use GEANT4 to simulate training data (truth known, unlimited labeled data!)
  - Later: validate on beam-test data
- Data set:
  - 6.7 million events
  - Particle: proton
  - Energy: 30 MeV to 100 MeV
  - Angle distribution: omnidirectional
  - Ionization quenching added
  - Reject events with less than two hits per plane (line defined by two points)

Track Reconstruction: Results



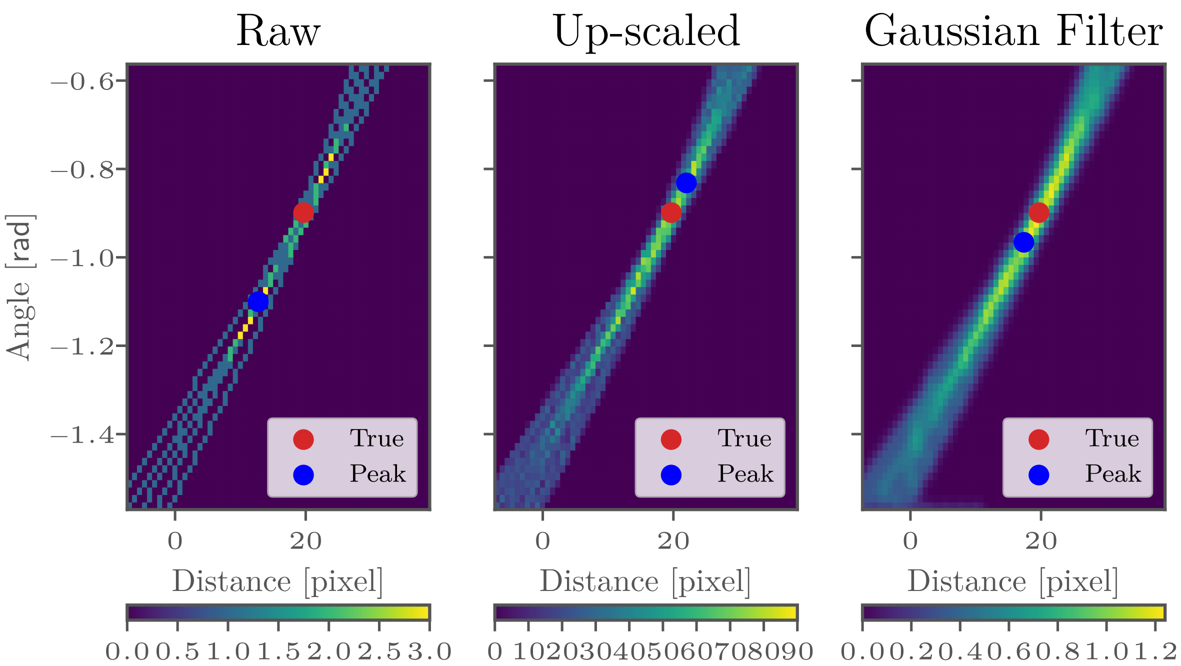
- Spatial reconstruction:
  - comparable results from Hough and ANN
- Angle reconstruction:
  - Factor two difference in 68 % Central Interval
    - ANN can capture correlations between projections and
    - use energy-deposit information



Spherical Coordinates:

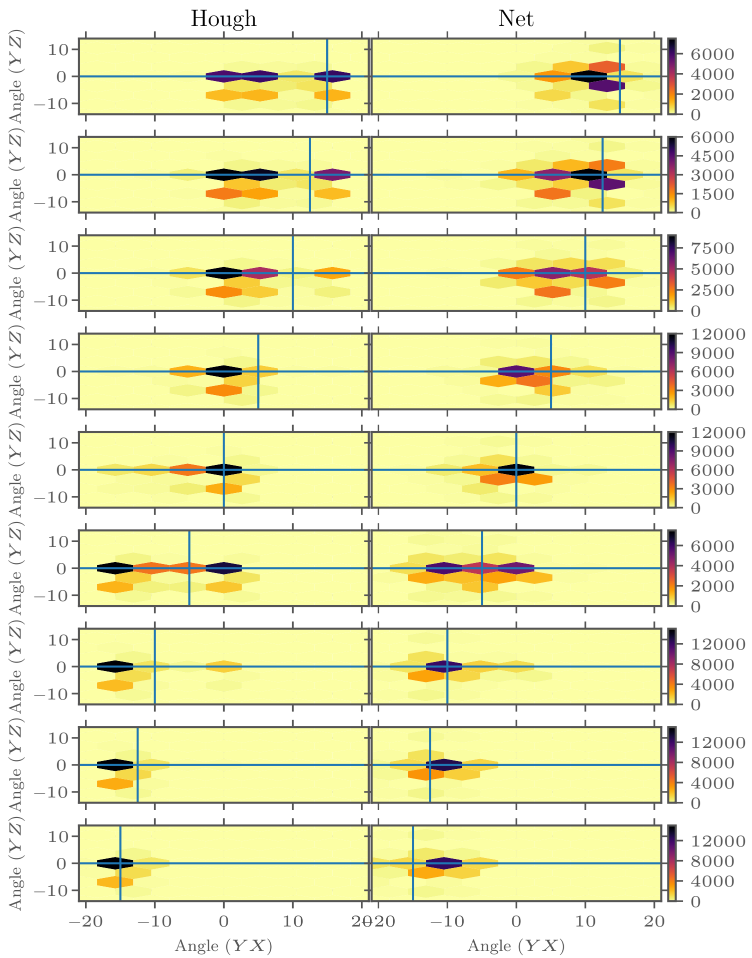
$$68\,\%\text{-CI}_{\text{Hough},\theta} = 3.29\,\text{deg}\ , \qquad 68\,\%\text{-CI}_{\text{Hough},\varphi} = 6.88\,\text{deg}\ ,$$
$$68\,\%\text{-CI}_{\text{ANN},\theta} = 1.58\,\text{deg}\ , \text{ and } \quad 68\,\%\text{-CI}_{\text{ANN},\varphi} = 3.00\,\text{deg}\ .$$

Hough Optimization

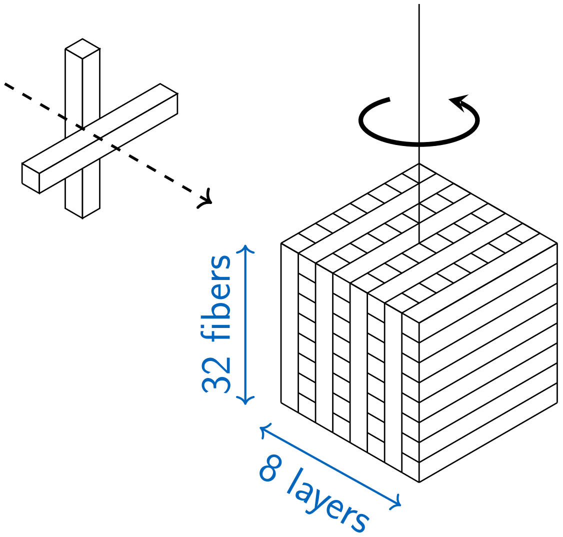


- Raw:
  - Artefacts
- Up-scaled image:
  - Few artefacts
  - Computationally expensive
- Gaussian Filter:
  - Few artefacts
  - Computationally cheap

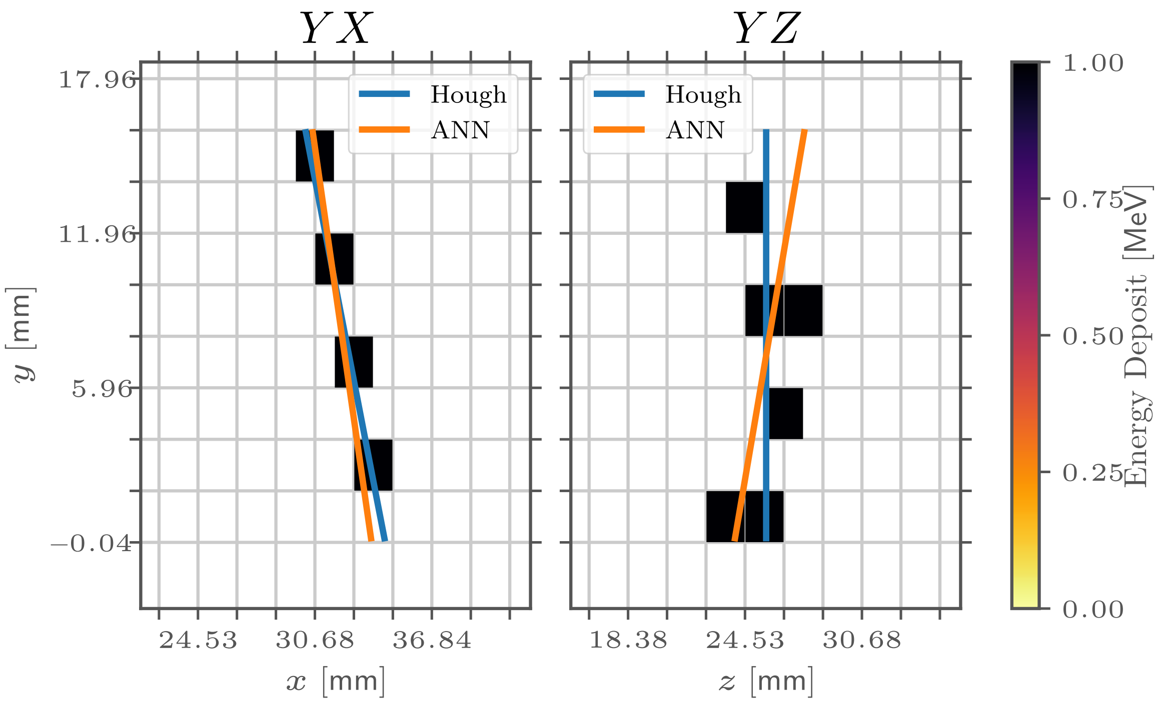
PSI Beam Test



- At Paul Scherrer Institute (πM1 beam line)
- Pion beam (450 MeV/c_0)
- Prototype module of the detector (8 layers à 32 fibers)
- Mounted on rotary table
- Upstream trigger cross



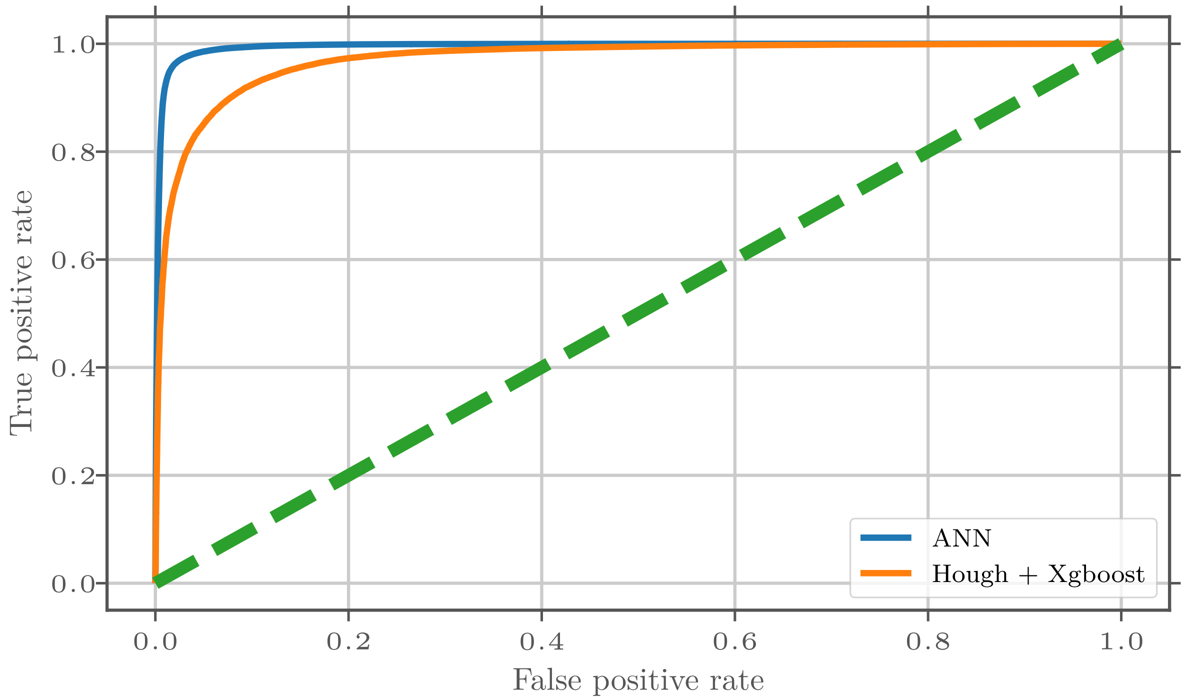
PSI Beam Test



- Crosstalk not included in simulation
  $\Rightarrow$ ANN strongly relies on 'neighboring pixel pattern'
- Hough transform: intrinsically robust

$\Rightarrow$ Build more realistic simulation and repeat

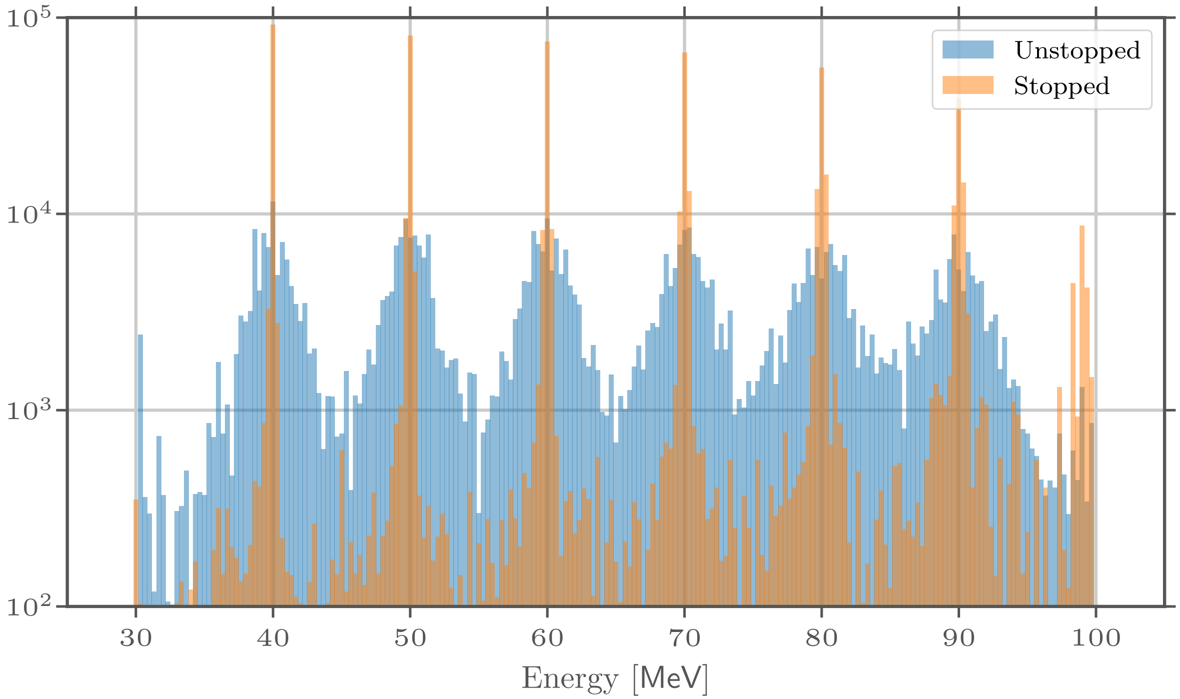
Detection of Stopped Particles in MAPT



- Boosted decision tree ('XGBoost') combines handcrafted features:
  - path-length difference
  - number of fiberhits
  - mean energy deposit
  - maximal energy deposit
- ANN:
  - raw detector images as input
  - self-learned features
  - CNN based on LeNet-5

Figure: ROC curve. Random classifier: diagonal, perfect classifier: top left corner.

Energy Reconstruction

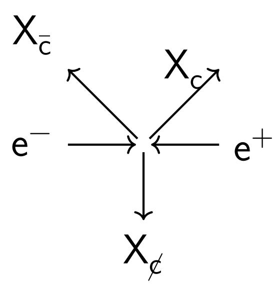


- Consider stopped and unstopped events seperately
- Validate at discrete energies: 40 MeV to 90 MeV in steps of 10 MeV
- Preliminary result:
  - Boundary effects
  - Spikes

Final Slide

Thanks for your attention!

Quark Production at B Factories



Quark production:

- $e^+e^- \longrightarrow B\bar{B}$
- $e^+e^- \longrightarrow q\bar{q}$, where $q = u, d, s$
- $\boxed{e^+e^- \longrightarrow c\bar{c} \longrightarrow X_c X_{\bar{c}} X_{\not{c}}}$

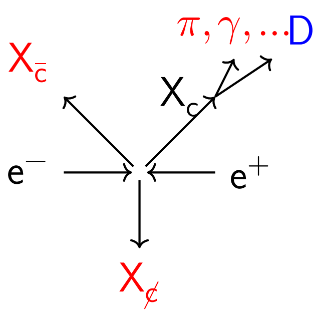
, where

X_c: open charm hadron

$X_{\bar{c}}$: open anti-charm hadron

$X_{\not{c}}$: charmless particles ('fragmentation')

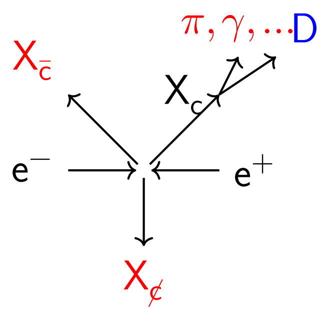
Typical Charm Analysis



Typical charm analysis:

- just reconstruct D, ignore Rest Of Event (ROE)!
  $\Rightarrow$ large combinatorial background
- suppress using vertex fitting
  - in case of no or few charged particles $\Rightarrow$ not viable
  - in case of multiple π^0s $\Rightarrow$ not viable
- solution: use information from ROE $\Rightarrow$ Charm Tagger

Charm Tagger



Charm Tagger:

- use information from Rest Of Event (ROE) $\Rightarrow$ suppress background

Algorithm:

- given particle of interest D $\Rightarrow$ find ROE
- from ROE identify $X_{\bar{c}}$, $X_{\not{c}}$ and $(X_c) = \pi, \gamma, \ldots$
- full event must obey conservation of quantum numbers:
  - charge
  - strangeness
  - baryon number
- momentum conservation: $P_D{}^{ROE} = P_{e^+} + P_{e^-} - P_{ROE}$
- compare $P_D{}^{ROE}$ to $P_D{}^{\text{reco}}$ reconstructed from decay products of D

Final Slide

Thanks for your attention!

Applications

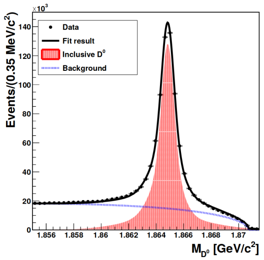


Figure: M_{D^0} distribution of inclusive D^0 sample. Taken from third paper below.

Charm tagging algorithm:

- useful for many analyses (examples below)
- high signal to background ratio
- clean sample

Similiar techniques employed at Belle:

- Measurements of branching fractions of leptonic and hadronic D_s^+ meson decays and extraction of the D_s^+ meson decay constant, A.Zupanc et al., 2013
- Search for the rare decay $\mathrm{D}_0 \longrightarrow \gamma\gamma$ at Belle, NK Nisar et al., 2016
- Search for D_0 decays to invisible final states at Belle, Y.-T. Lai et al., 2017

Data-Science Challenges

Gaols:

- improve signal to background
- increase efficiency (so far order of 0.1 %)

$\Rightarrow$ Use Neural Networks!

Main Data-Science challenges:

- $e^+e^- \longrightarrow X_c X_{\bar{c}} X_{\not{c}}$ not well measured
  $\Rightarrow$ maybe can't fully trust simulated data
  $\Rightarrow$ may need algorithm that adapts to experimental data
- Uncertainties on training sample lables
- preprocessing of Machine Learning input data (cf. next slide)

Machine Learning Input Data

Low level:

- Input: charged tracks, neutral clusters

+ all correlations intact
+ no prior interpretation of data

− large number of inputs

High level:

- Input: reconstructed particles (apply PID cuts, group tracks to composite particles)

$\Rightarrow$ input of extra knowledge

− might lose correlations

+ interpretation might be hard to learn from low level
+ smaller number of inputs

Combine low and high level information?

Final Slide

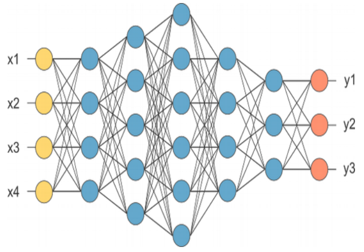
Thanks for your attention!

Development of a Neural Network for Online Event Reconstruction for a Radiation Monitor

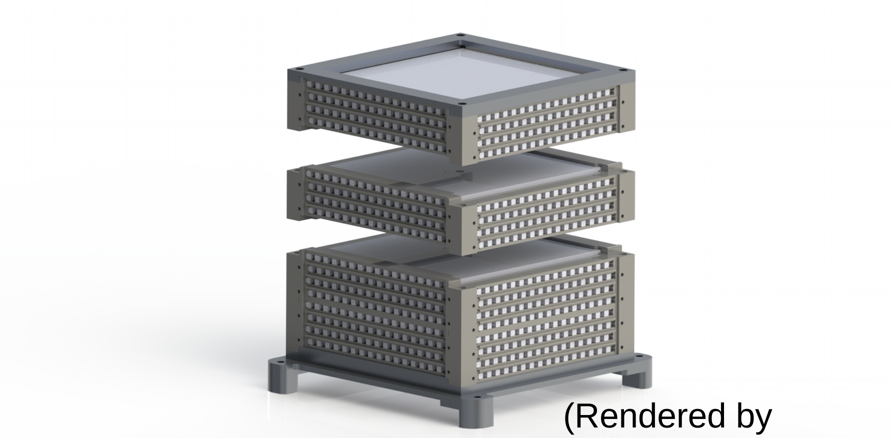
(Master’s Colloquium)

Lukas Bierwirth

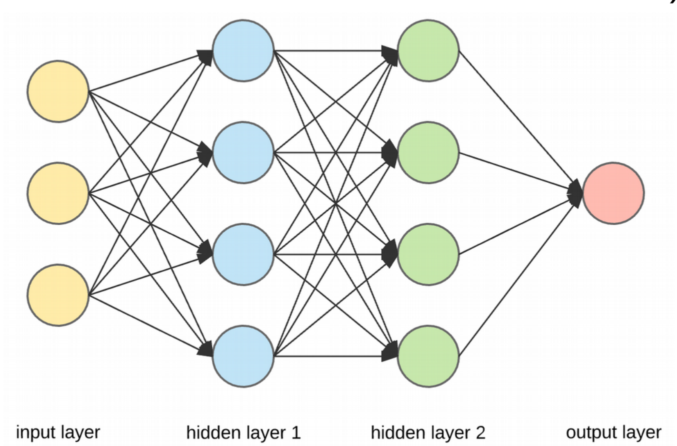
Physics Department
TU München



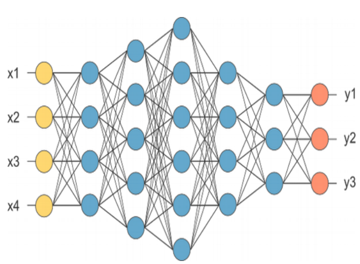
The Challenge



(Rendered by M. Losekamm)

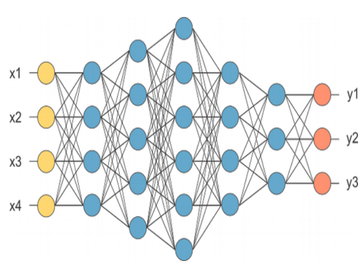


- The givens:
  - The response of the Multi-purpose Active-target Particle Telescope (MAPT) to an incoming particle
- The goal:
  - Online reconstruction of the particle´s parameters
    - Track,
    - Energy,
    - Type, ...
- The method:
  - Artificial Neural Networks (ANNs)



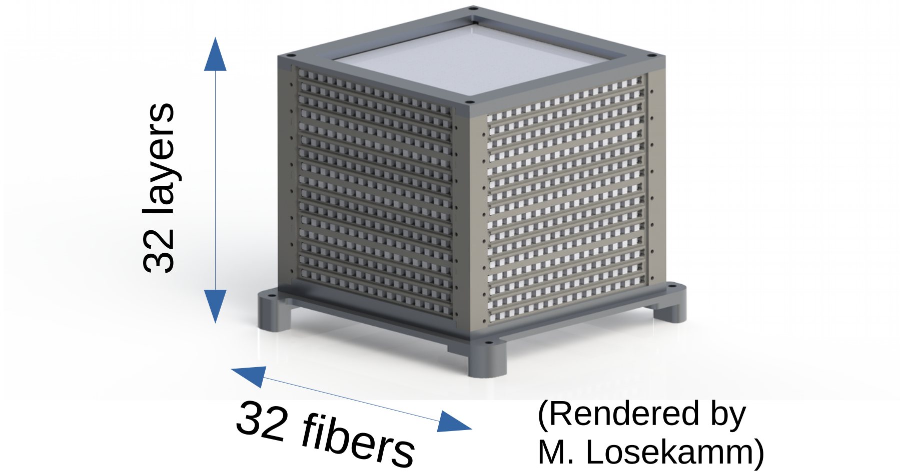
Outline

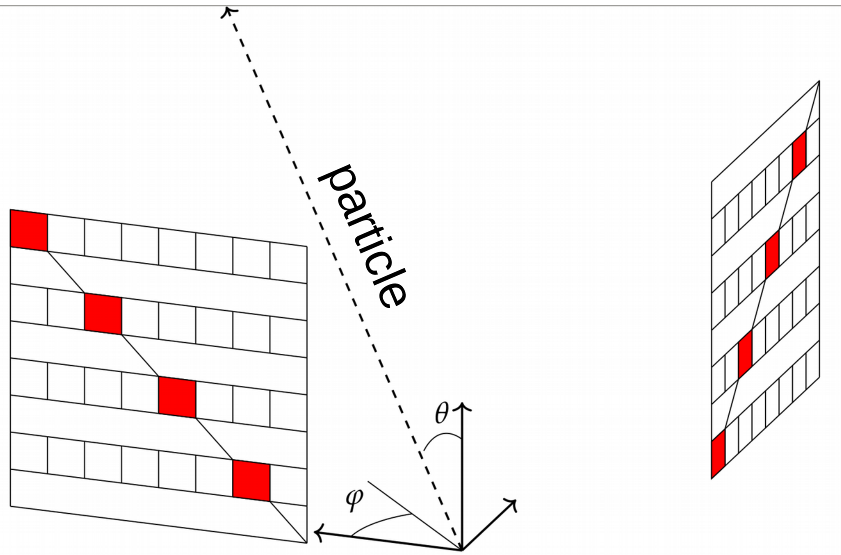
- Radiation Monitor:
  - Multi-purpose Active-target Particle Telescope (MAPT)
- Pattern-Recognition Methods:
  - Hough Transform
  - Artificial Neural Networks (ANNs)
- Central Results:
  - Simulated Data
  - Beam-Test Data
- Future Work



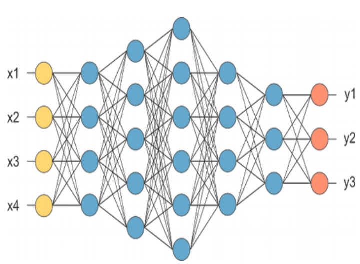
Radiation Monitor

Radiation Monitor: MAPT

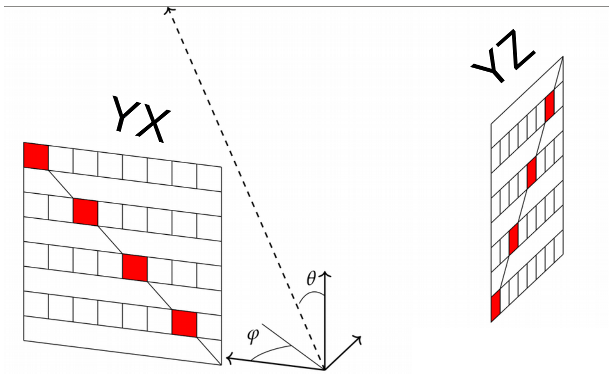


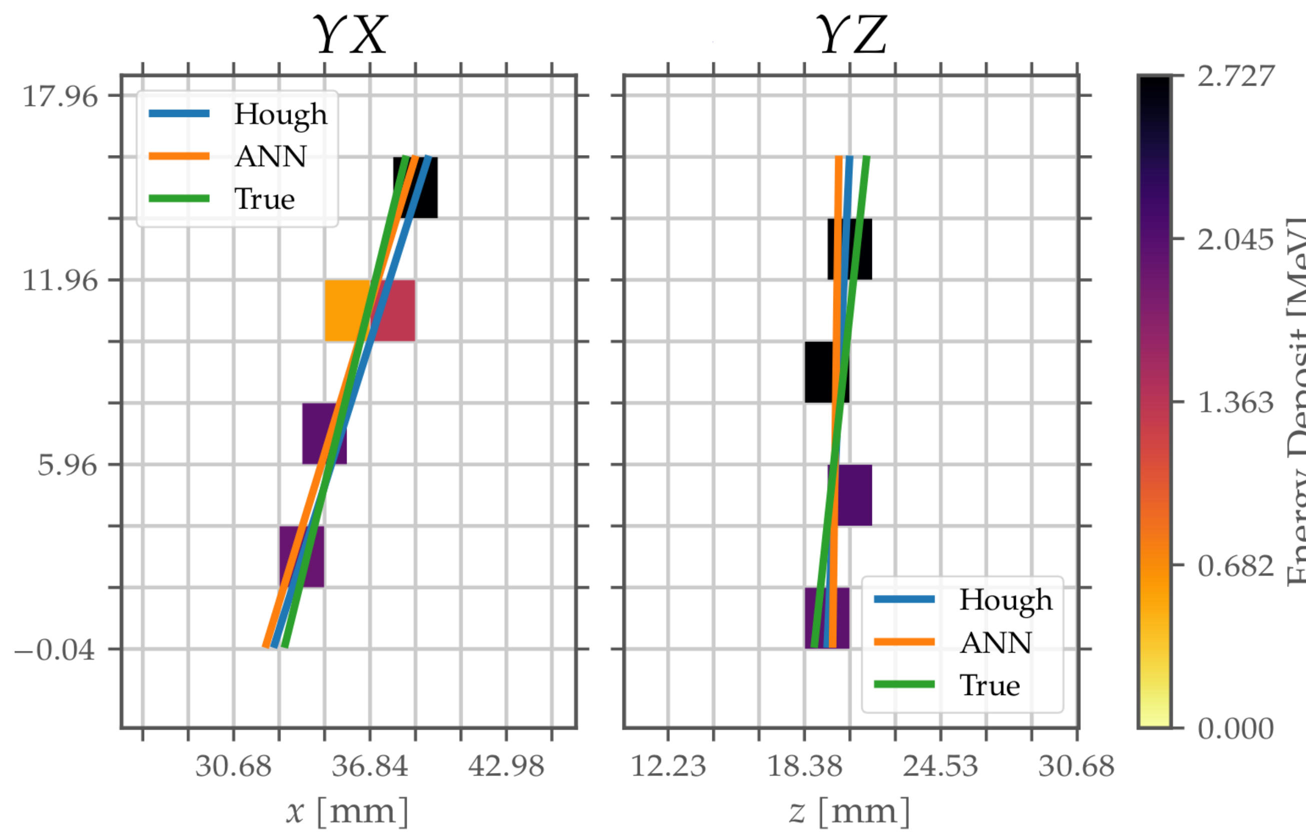


- Segmented calorimeter
  - 32 layers (2018 Design)
  - Omnidirectional acceptance
- Layer:
  - 32 fibers
  - Rotated by 90° w.r.t. previous layer
  - Shifted horizontally w.r.t previous layer of equal orientation (2018 Design)
- Fiber:
  - Active core:
    - Polystyrene
    - 1.92mm x 1.92mm x 70mm
  - Sputtered with aluminium
- Signal: Two orthogonal projections of the track

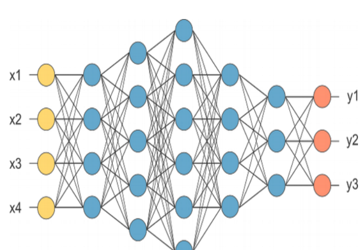


Radiation Monitor: Signal

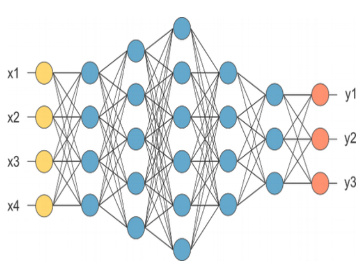




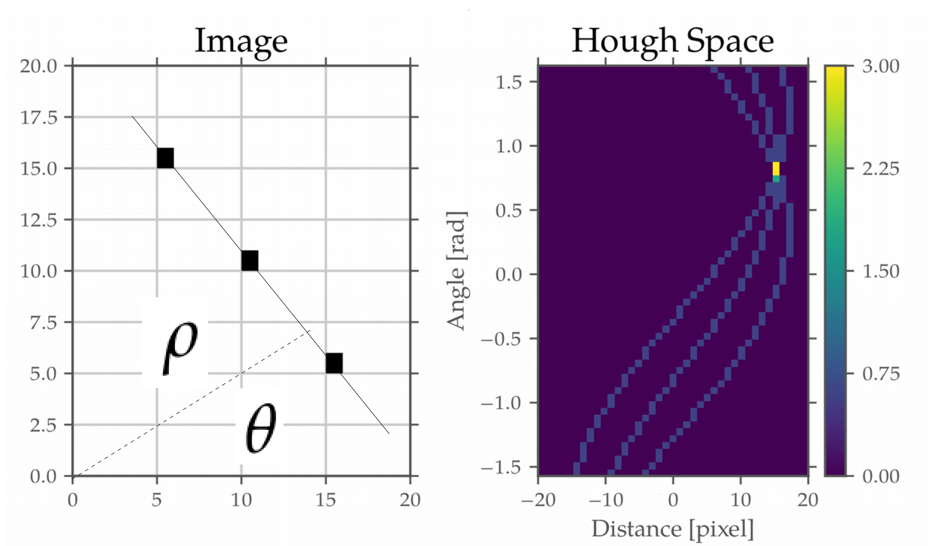
- Particle deposits energy in the active material along its path
- Active material converts energy to photons
- Photons collected by SiPMs → electronical signal proportional to energy deposit (non-linear effects!)



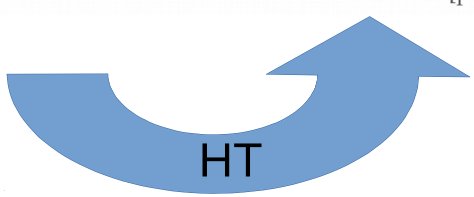
Pattern-Recognition Methods



Methods: Hough Transform



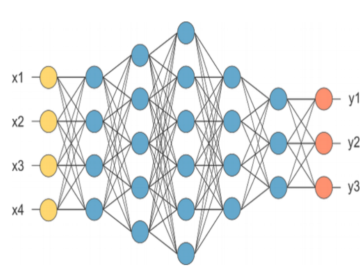
- Standard line detection method in computer vision: fast, robust
- Maps every pixel of the input image to one sinusoid in the Hough space
- Iff sinusoids meet at one point, the pixels lie on one line in the input image
- In Hough space: line detection by peak finding



$$\rho = x\cos\theta + y\sin\theta$$

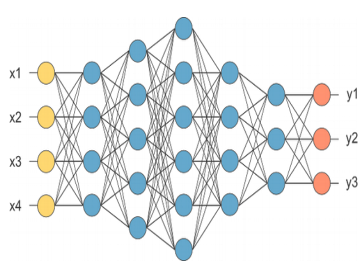
Algorithm 1 Hough transform

function HOUGH(image)
  $\rho_{\text{max}} \leftarrow \left\lceil \sqrt{\text{width}_{\text{image}}^2 + \text{height}_{\text{image}}^2} \right\rceil$
  $\rho_{\text{min}} \leftarrow -\rho_{\text{max}}$
  accumulator$[-\frac{\pi}{2} \dots \frac{\pi}{2}][\rho_{\text{min}} \cdots \rho_{\text{max}}] \leftarrow 0$
  foreach pixel $\neq 0$ **do**
    for $\theta \leftarrow -\frac{\pi}{2}$ to $\frac{\pi}{2}$ **do**
      $\rho \leftarrow \lfloor x_{\text{pixel}} \cos\theta + y_{\text{pixel}} \sin\theta \rfloor$
      accumulator$[\theta][\rho] \leftarrow$ accumulator$[\theta][\rho] + 1$
    end for
  end foreach
  return accumulator
end function



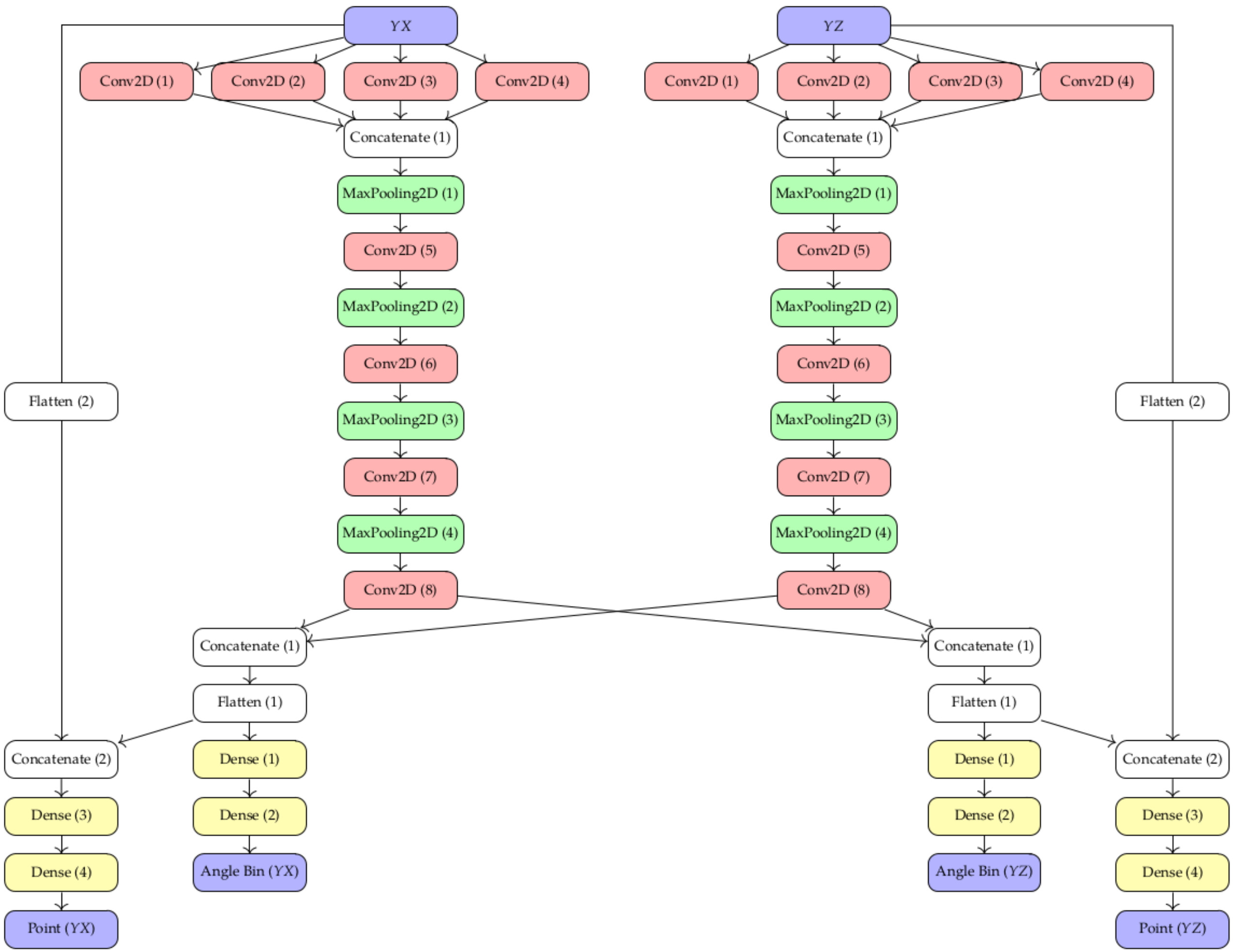
Methods: ANNs

- Powerful pattern-recognition algorithm
- Implementation of a (nonlinear) function parameterized by a set of weights
- Practical usage: weights can be adjusted efficiently via optimization → Supervised Learning

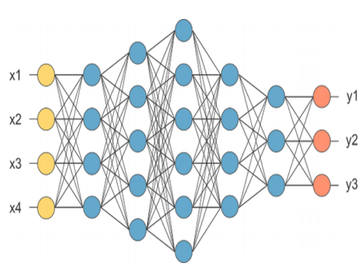


- Examples:
  - Online translation services,
  - snapchat filter,
  - speech recognition
- Structure:
  - Neurons form layers
  - Layers form networks
- Different types of architecture:
  - CNNs (Convolution),
  - RNNs (Recurrent → information loops),
  - mixtures

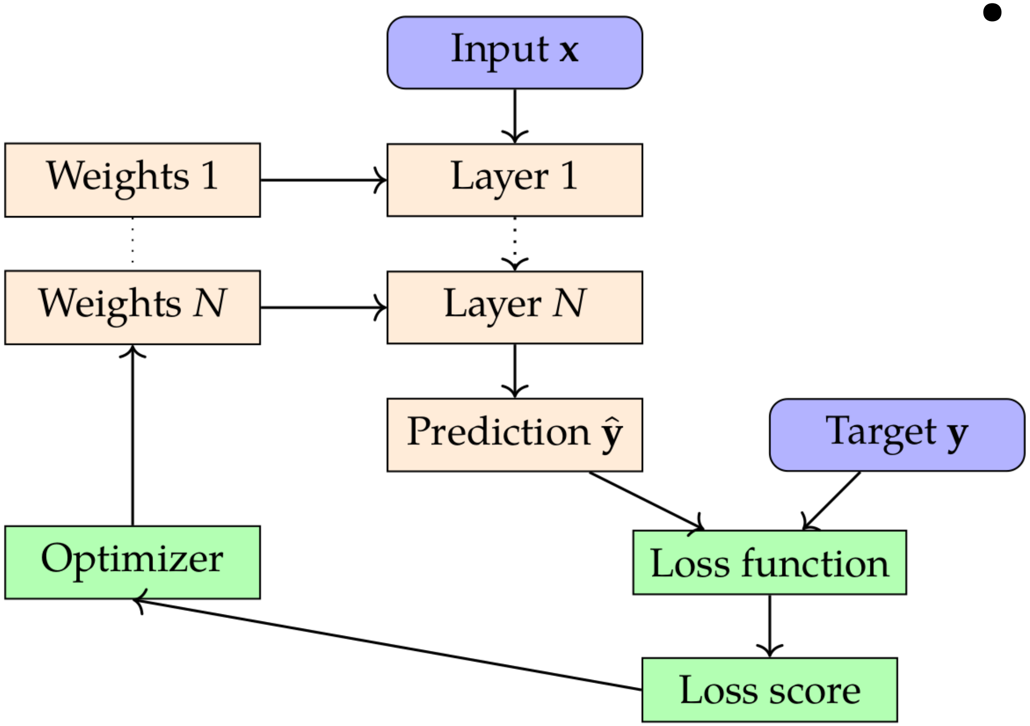
Methods: ANNs: Example



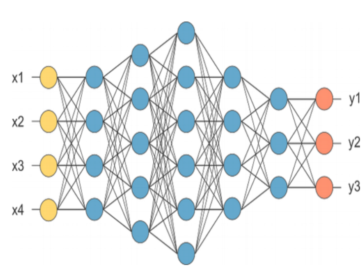
- Layers:
  - Blue: In- and Output
  - Red: Convolution
  - Green: MaxPooling (Take maximum value of certain window)
  - Yellow: Dense Layers (Neurons connected to every neuron of previous layer)
  - White: Reshaping
- Task: Track Reconstruction



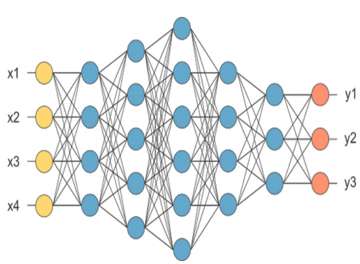
Methods: ANNs: Training



- Supervised learning:
  - Given: labeled data point (x,y)
  - ANN makes prediction
  - Loss function compares prediction to target
  - Weights adapted via gradient descent to minimize loss
  - Important: Sufficient amount of labeled data points!

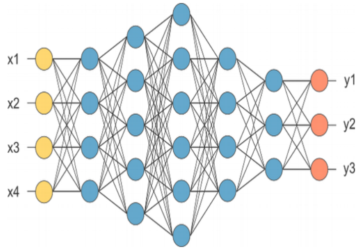


Simulated Data

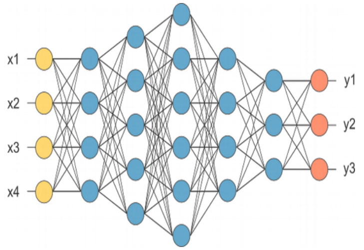


Simulation: Data Set

- Idea:
  - Use GEANT4 to simulate training data (truth known, unlimited labeled data!)
  - Later: validate on beam-test data
- Data set:
  - 6.7 million events
  - Particle: proton
  - Energy: 30 to 100 MeV (flat distribution)
  - Angle distribution: omnidirectional
  - Ionization quenching added
  - Exclude events with less than two hits per plane (line defined by two points)



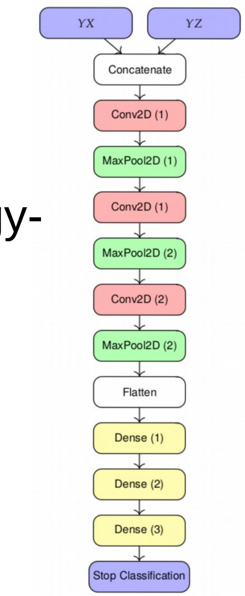
Stop Detection:
Has the particle stopped in the detector?

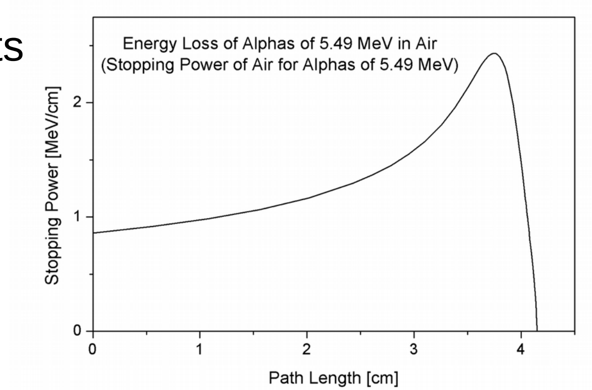


Stop Detection: Methods

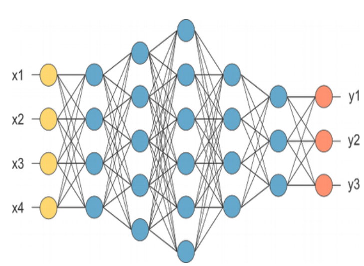
- Non-Machine Learning:
  - Handcrafted feature
  - Based on Hough transform
  - Idea:
    - Use reconstructed line to check how many fibers would be hit if the particle was unstopped --> "path-lenght difference“
    - Compare to actual number of hits
  - Improvement: Look for Bragg peak --> Decision tree (ML!)

- ANN:
  - Self-learned features
  - CNN architecture based on LeNet-5
  - Relies (supposedly) on energy-deposition characteristics as well as the track geometry



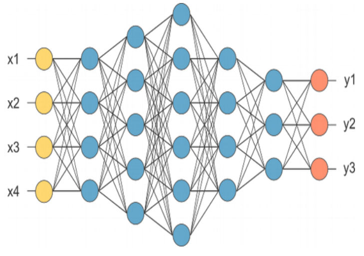


(Figure taken from Wikipedia)



Stop Detection: Binary Classification

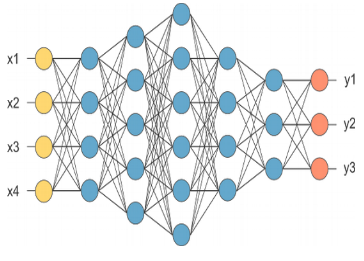
- Accuracy $\frac{TP+TN}{TP+FP+TN+FN}$
- Sensitivity or True Positive Rate $\frac{TP}{TP+FN}$
  - how good a test is at detecting the positives. A test can cheat and maximize this by always returning “positive”.
- Specificity or True Negative Rate $\frac{TN}{TN+FP}$
  - how good a test is at avoiding false alarms. A test can cheat and maximize this by always returning “negative”.
- Precision or Positive Predictive Value $\frac{TP}{TP+FP}$
  - how many of the positively classified were relevant. A test can cheat and maximize this by only returning positive on one result it’s most confident in.
- F1-Score $\frac{2}{\frac{1}{Recall}+\frac{1}{Precision}}$
  - Combines Recall and Precision into one quantity



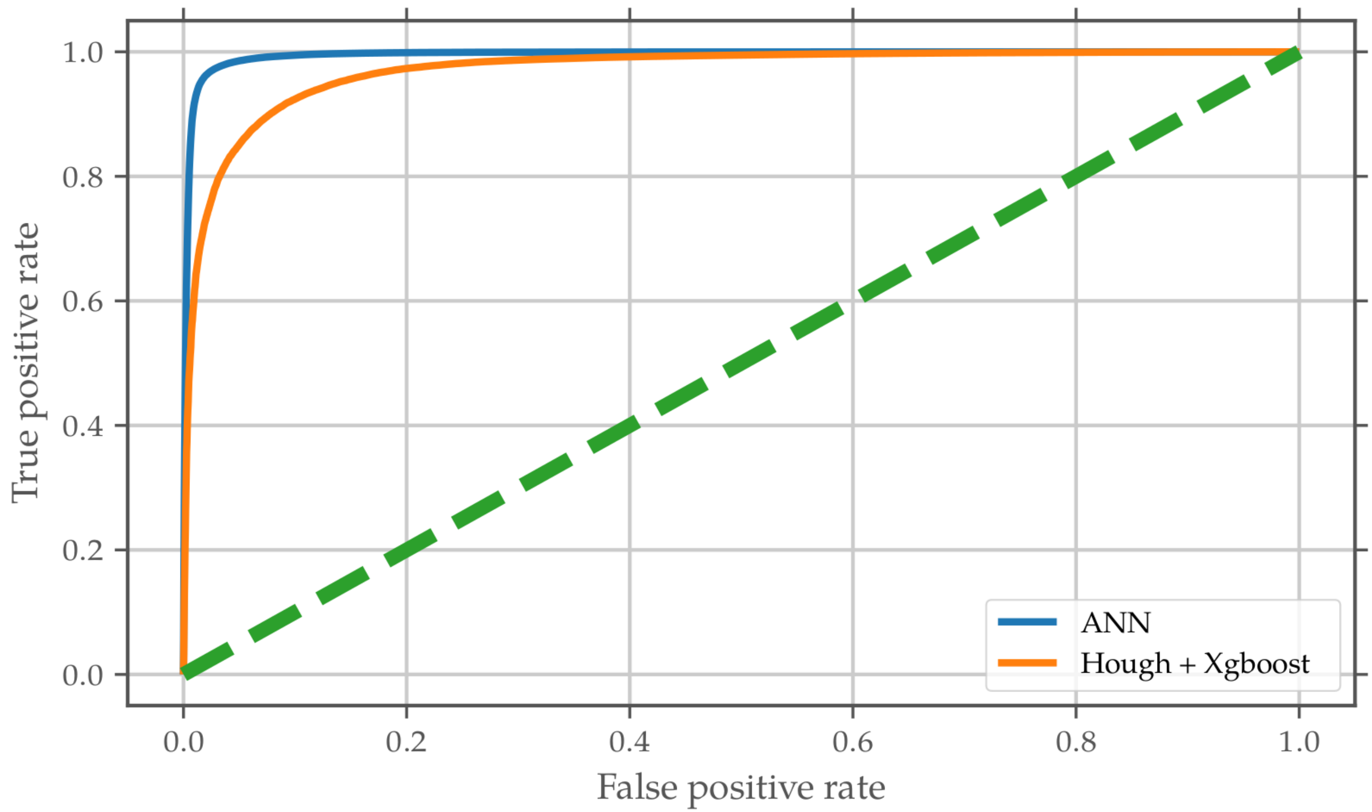
Stop Detection: Results

Method	ACC	F1	PPV	TPR	TNR	AUC	$\frac{\text{Events}}{\text{s}}$
Hough	0.81	0.83	0.82	0.85	0.75	0.89	$6.30 \cdot 10^1$
XGBoost	0.88	0.90	0.87	0.93	0.82	0.94	$5.89 \cdot 10^5$
Hough + XGBoost	0.91	0.92	0.91	0.93	0.89	0.97	$6.30 \cdot 10^1$
ANN	0.97	0.98	0.98	0.98	0.97	0.99	$1.40 \cdot 10^4$

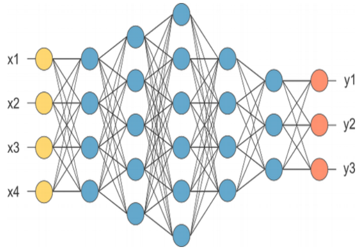
- ANN: Close to perfect score (=1)
- “Hough + XGBoost”: path-length difference (slow!), fiber hits, mean and maximum energy deposit → Bragg peak
- “XGBoost”: same inputs, except no path-length difference, i.e. only Bragg peak
- “Hough”: non-Machine Learning algorithm



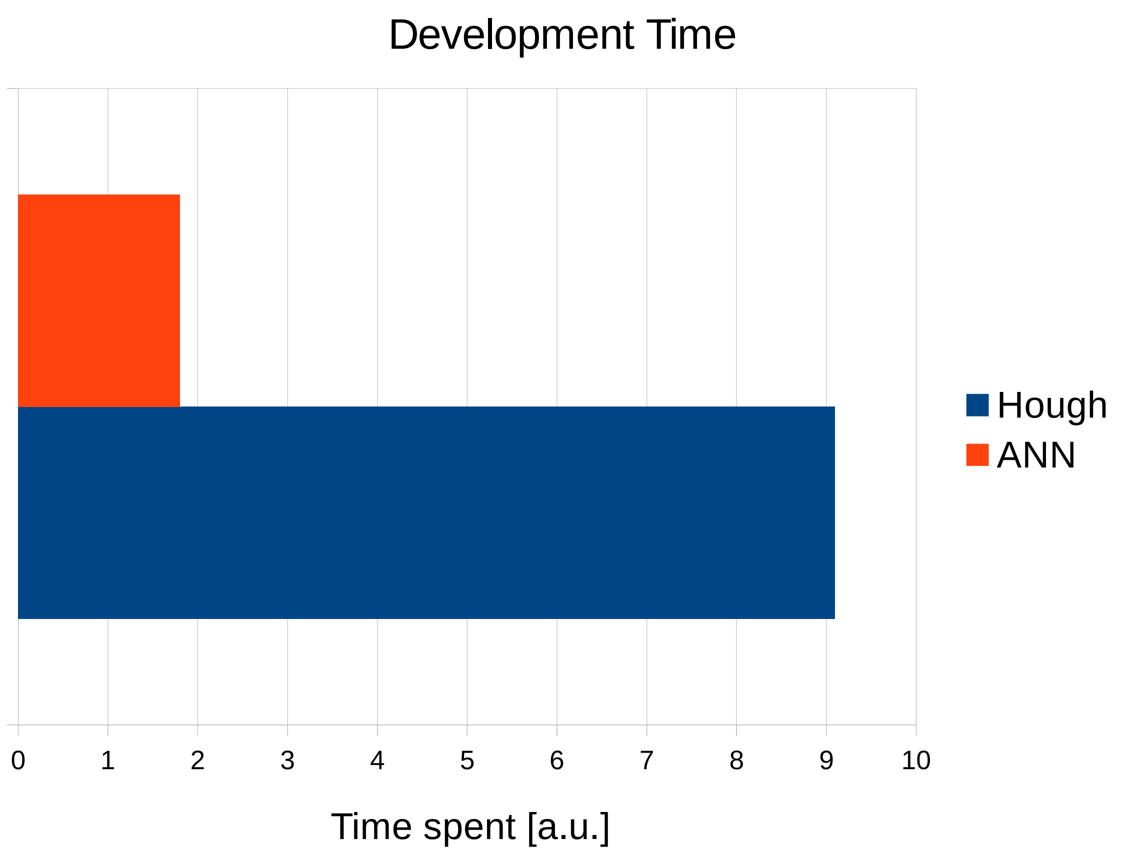
Stop Detection: ROC curve



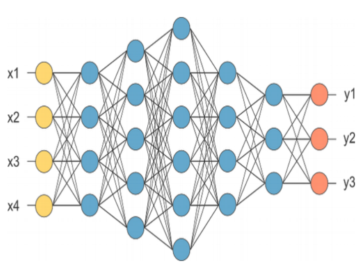
- Diagonal: random classifier
- Perfect result: top left corner



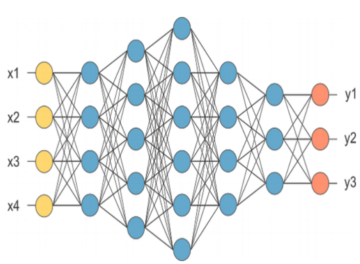
A Side Note



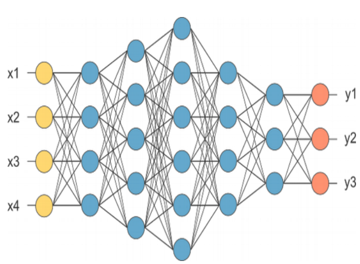
- Handcrafted features: time expensive to develop!



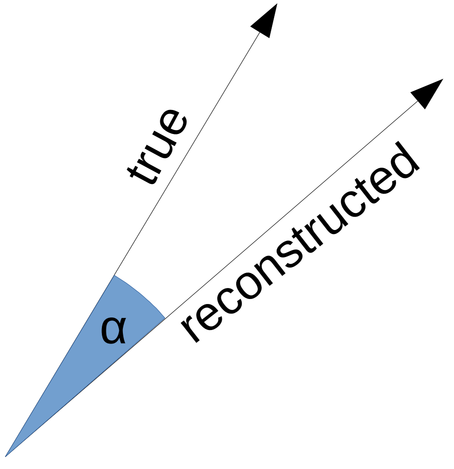
Track Reconstruction



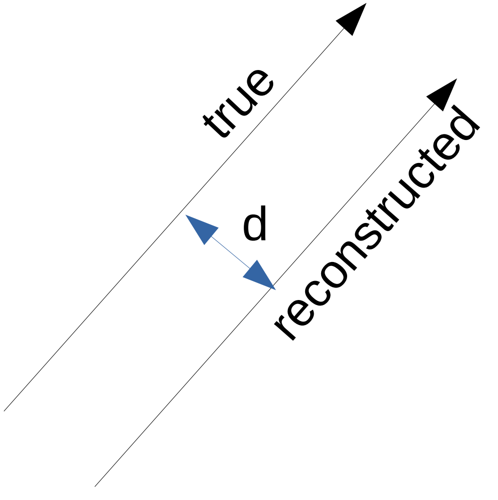
- Hough Transform
  - Output per projection:
    - angle
    - distance from origin
  - From 2-D to 3-D: math
- ANN
  - Sampe output as HT
  - Architecture inspired by Google's Inception Net
  - Can use energy-deposit information
  - Can capture correlations between projections

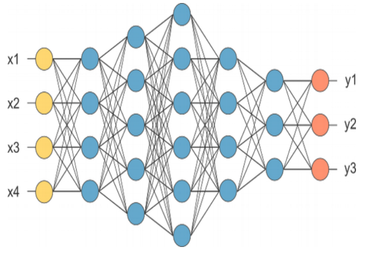


- Angle:

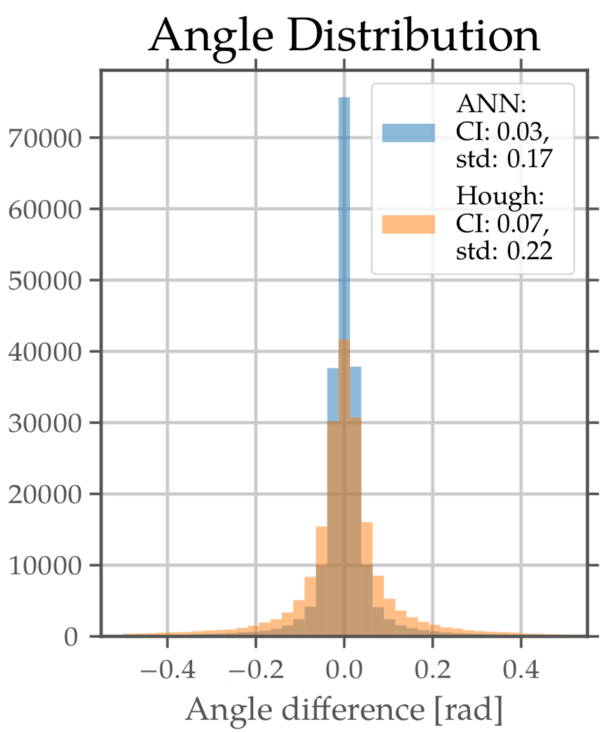


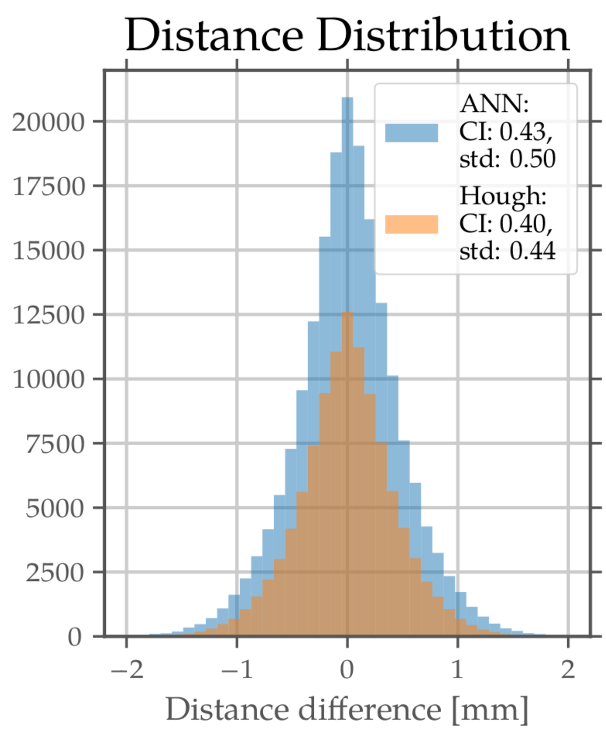
- Location (parallel lines):



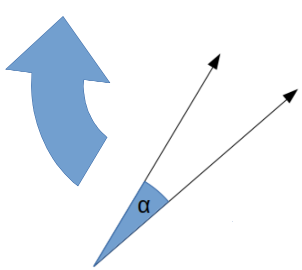


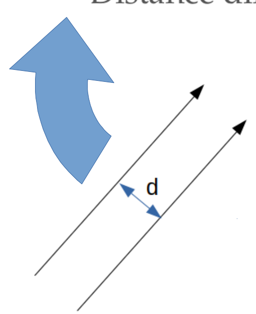
Track Reconstruction: Results





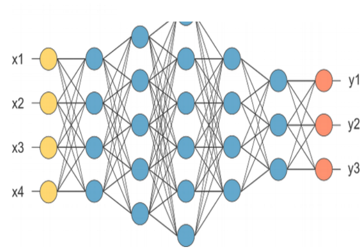
- Location extraction comparable
- Difference of factor two in 68%-Central Interval for angle reconstruction



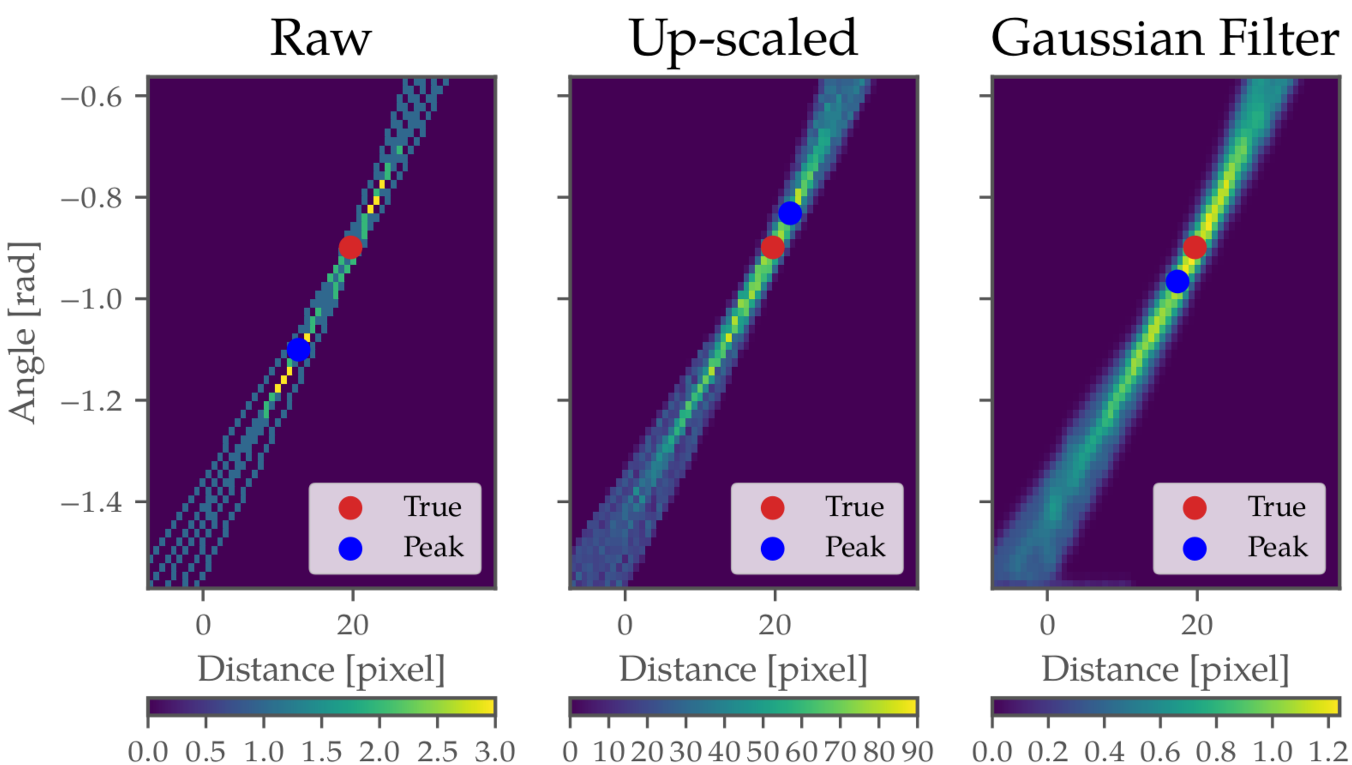


Spherical Coordinates:

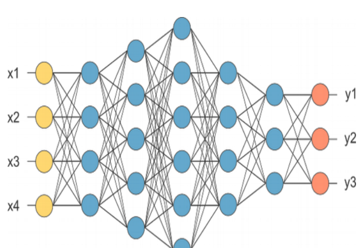
$$68\,\%\text{-CI}_{\text{Hough},\theta} = 3.29\,\text{deg}\ , \qquad 68\,\%\text{-CI}_{\text{Hough},\varphi} = 6.88\,\text{deg}\ ,$$
$$68\,\%\text{-CI}_{\text{ANN},\theta} = 1.58\,\text{deg}\ , \text{and} \qquad 68\,\%\text{-CI}_{\text{ANN},\varphi} = 3.00\,\text{deg}\ .$$



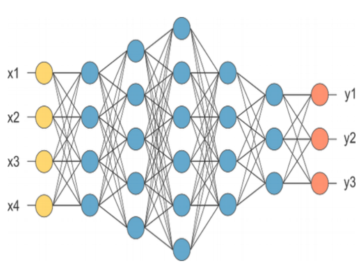
Track Reconstruction: Optimizing Hough



- Raw:
  - Noisey
- Up-scaled image:
  - Low noise
  - Computationally expensive
- Gaussian Filter:
  - Low noise
  - Computationally cheap

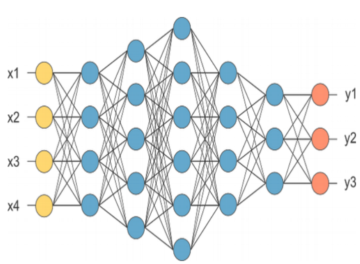


Energy Reconstruction

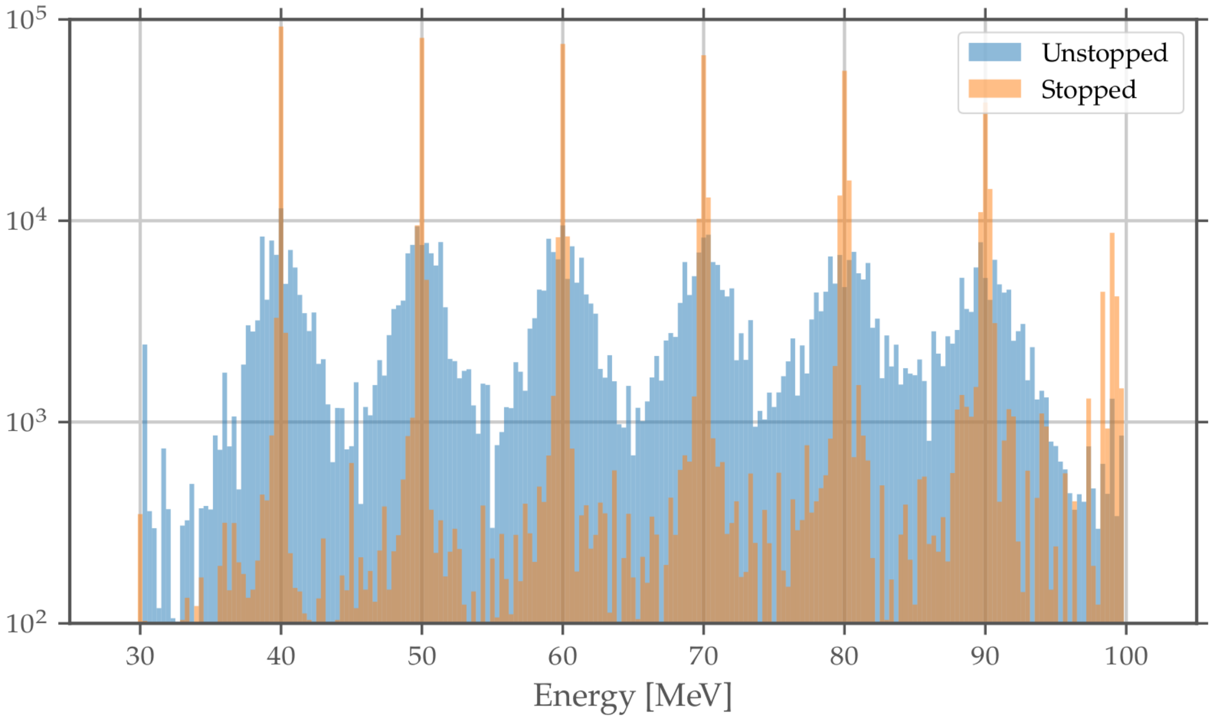


Energy Reconstruction

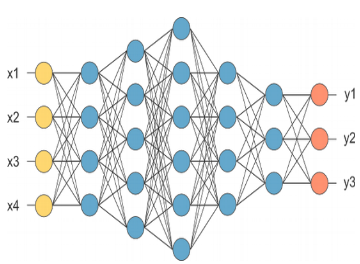
- Consider stopped and unstopped events seperately:
  - Stopped: whole energy deposited --> summation (quenching!)
  - Unstopped: Extrapolate Bragg curve
- No non-Machine Learning alternative developed as part of this work --> Compare different ANNs
- Preliminary Results!



Energy Reconstruction: CNN



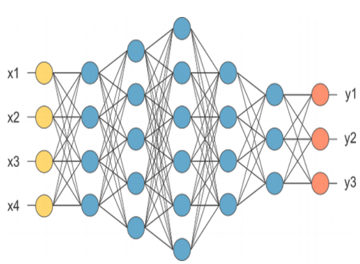
- Validate at discrete energies: 40, 50, …, 90 MeV
- Boundary effects!
- Spikes!
- Unlikely events: stopped 90 MeV protons,...



Energy Reconstruction

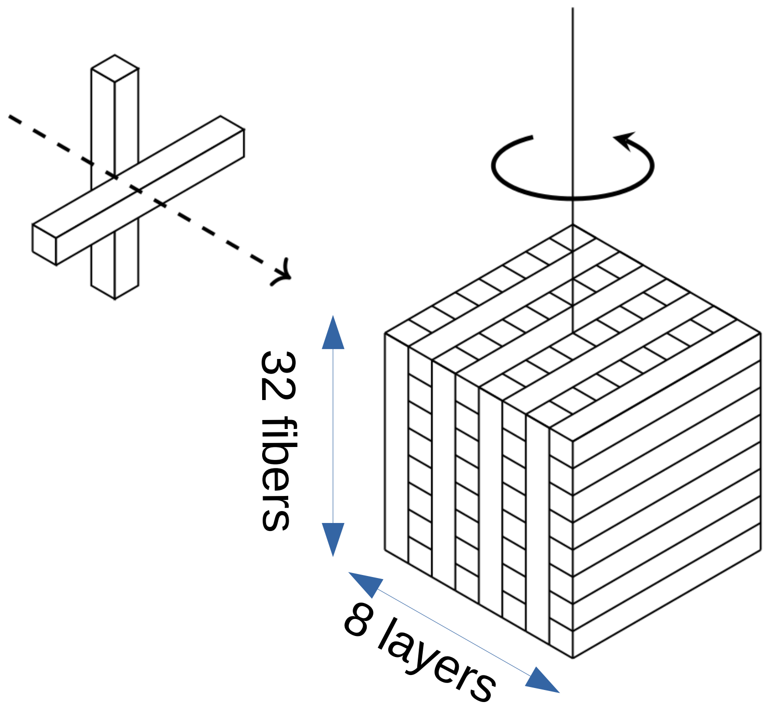
	Stopped			Unstopped		
ANN	Mean [%]	68 %-CI [%]	95 %-CI [%]	Mean [%]	68 %-CI [%]	95 %-CI [%]
CNN	−0.06	0.39	7.62	−0.07	3.53	10.94
RNN (cross entropy)	1.49	1.55	17.23	−1.38	12.45	34.44
RNN (mape)	1.80	1.98	21.82	0.74	11.62	27.75

- Flat energy distribution → “Natural energy” distribution for stopped and unstopped protons
- CNN: unbiased, highest precision
- Long tail issue

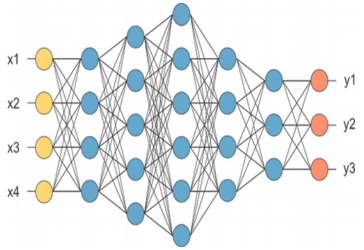


PSI Beam Test

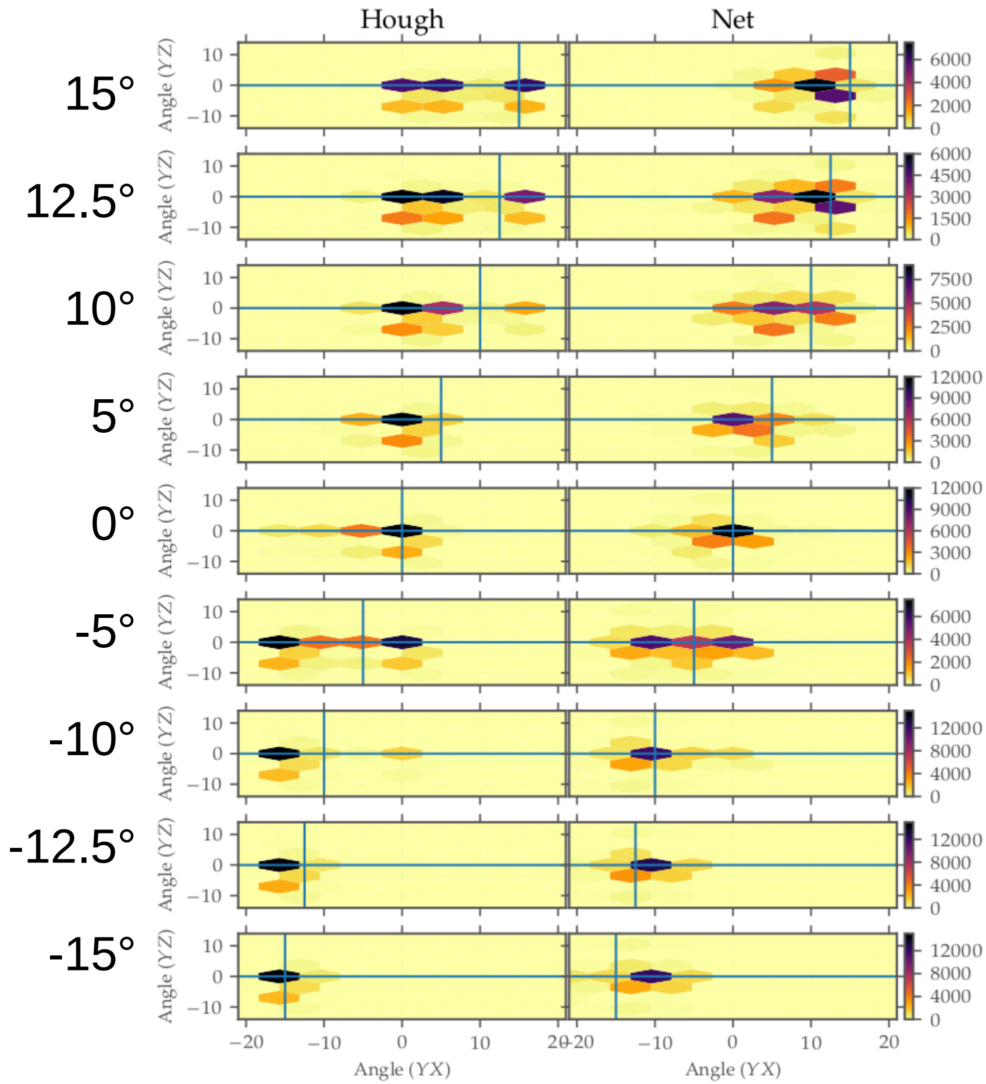
PSI Beam Test: Set-up

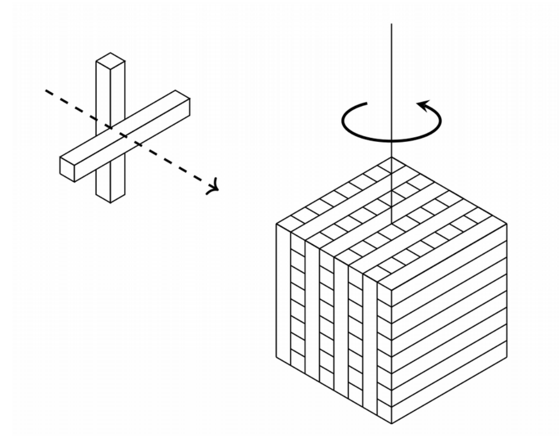


- At Paul Scherrer Institute (November, 2018)
- Small prototype of the detector (total: 8 layers)
- Mounted on rotary table
- Upstream trigger cross
- Pion beam

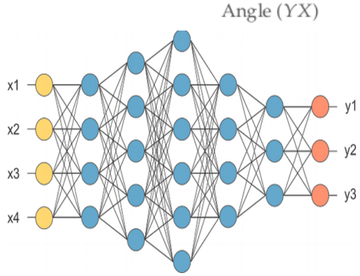


PSI Beam Test: Results

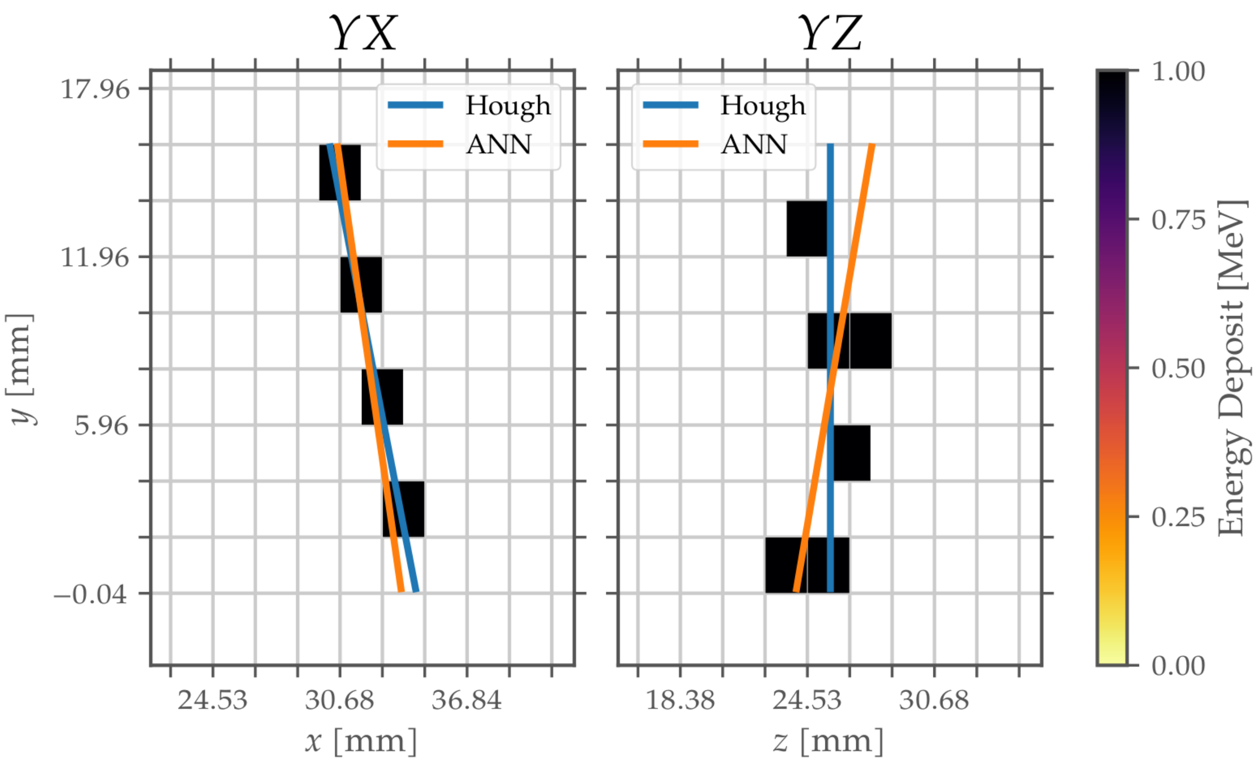




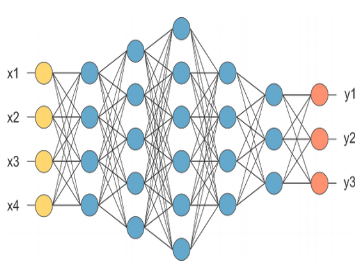
- Track rotation angle
- Dead channels!



PSI Beam Test: Generalization Problem



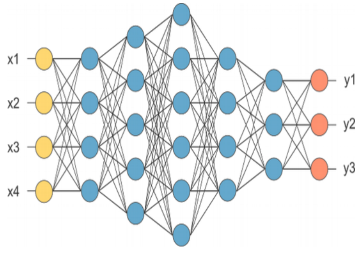
- Crosstalk not included in simulation → ANN strongly relies on „neighboring pixel pattern“ (cf. YZ-plane)
- Hough Transform: intrinsically robust



Future Work

Future Work

- More thorough analysis of PSI data (energy calibration, add more processes to simulation)
- Port ANN to small chip → integrate into experimental setup
- Explore further ANN architectures
- Put the different parts (energy, track, ...) together to form an autoencoder → capture correlations between different parameters



Thanks for your time and attention!