BELLE II PIXELDETECTOR CLUSTER ANALYSES USING NEURAL NETWORK ALGORITHMS

STEPHANIE KÄS

ON THE BEHALF OF THE BELLE II GROUP AT THE UNIVERSITY OF GIESSEN

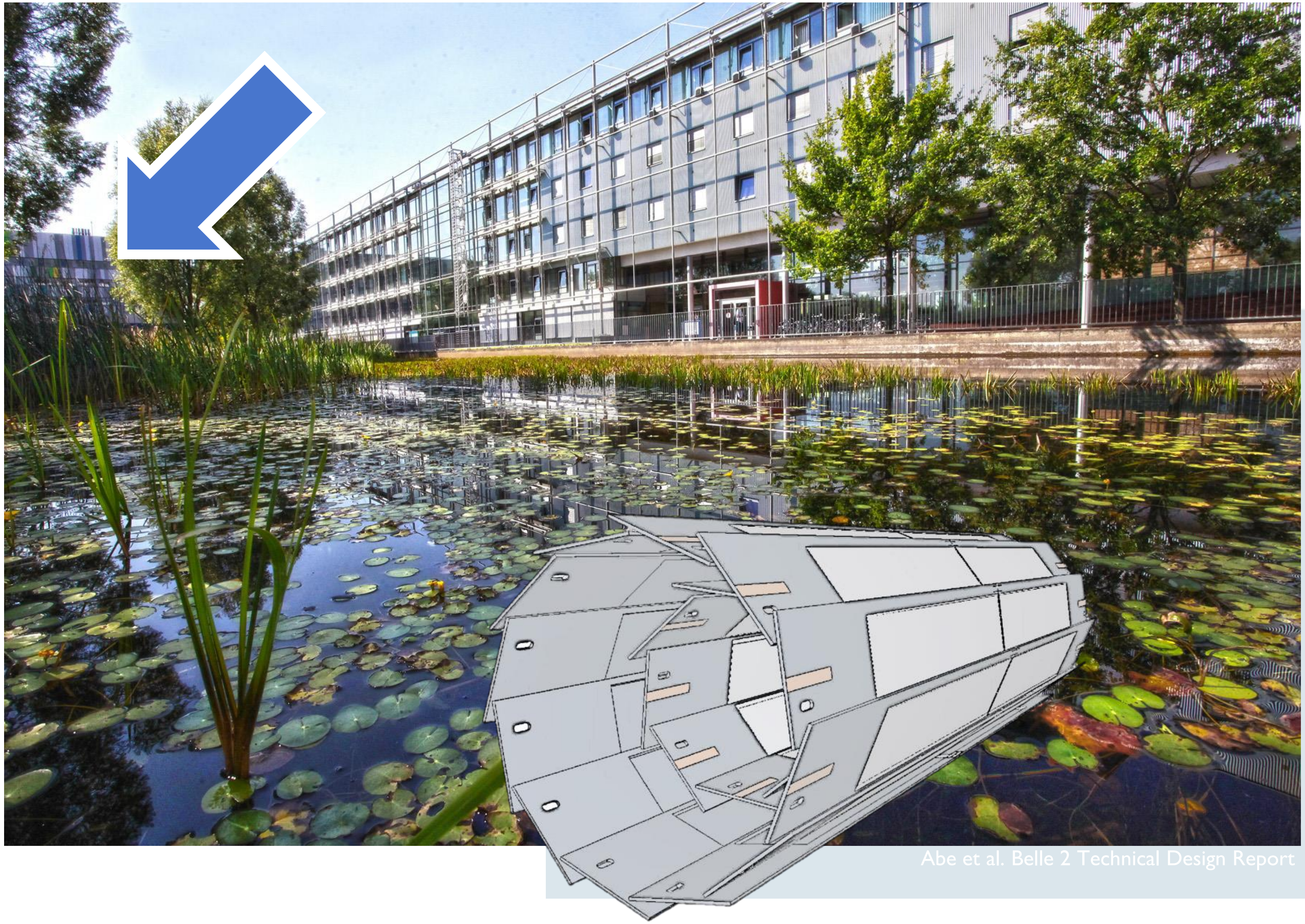
Results by: apl. Prof. Dr. J. S. Lange, K. Dort, S. Käs, M. Peter, I. Heinz, J. Bilk, J. Budak, P. Lehnhardt , F. Zorn

Bundesministerium für Bildung und Forschung

JENNIFER2
EU grant n.822070

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Belle II



Abe et al. Belle 2 Technical Design Report

JLU

NEUE WEGE. SEIT 1607.

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

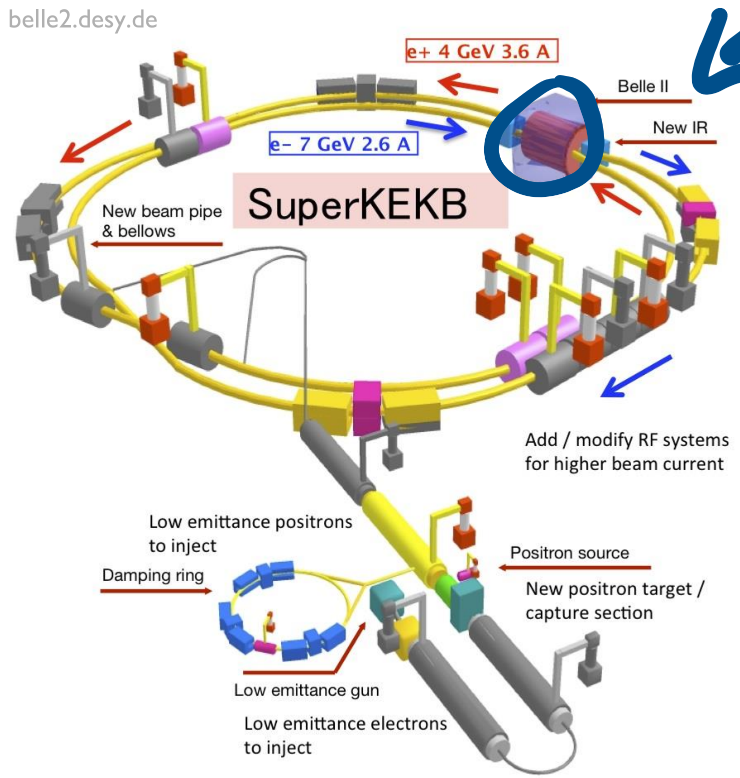
Classic MLP & Kohonen Maps

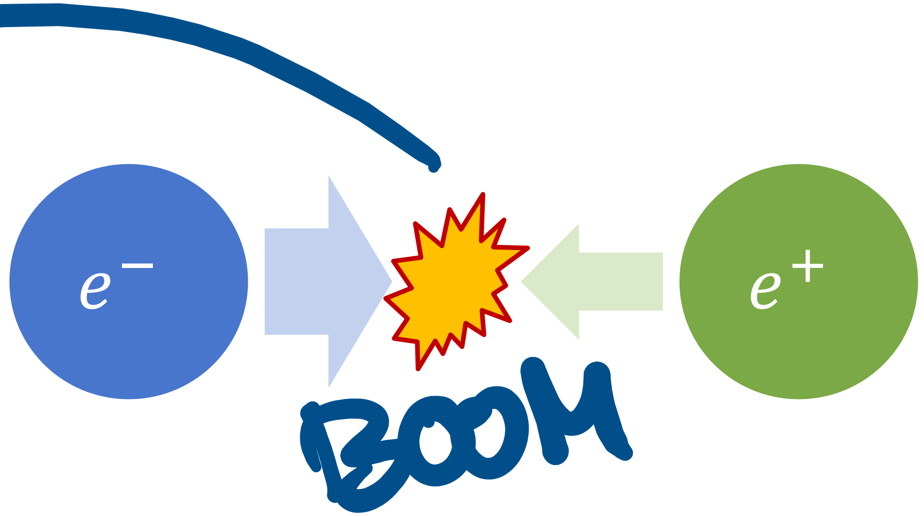
Hopfield Networks, Voxels, Autoencoder…

Networks on FPGA

PROJECTS RELATED TO DATA SCIENCES AND AI….

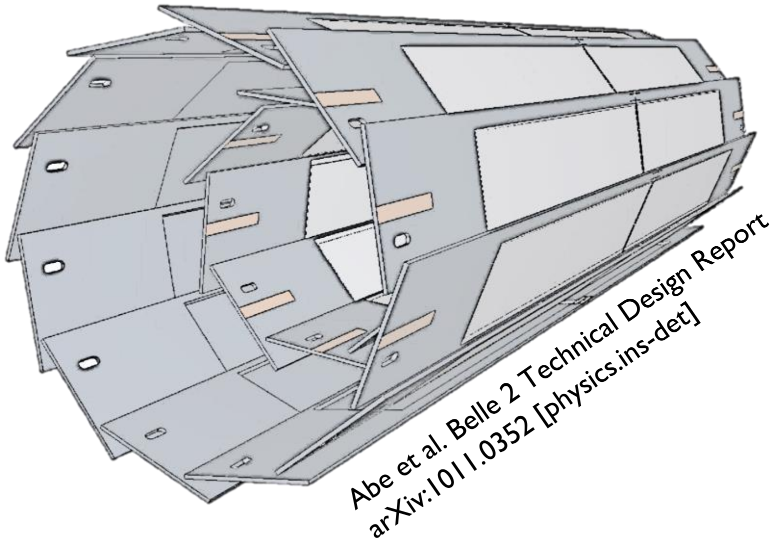
BELLE II EXPERIMENT

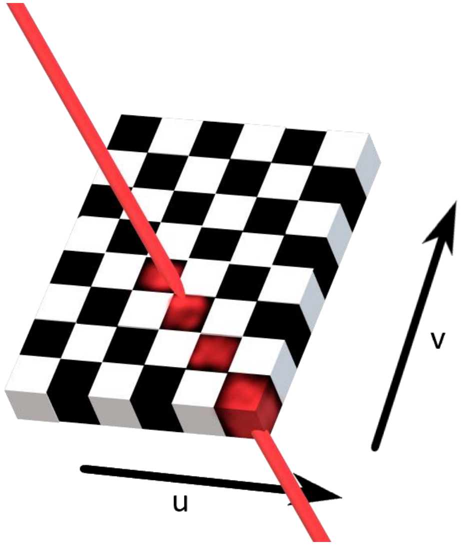




- Location: Tsukuba, Japan
- High luminosity

BELLE II PIXELDETECTOR



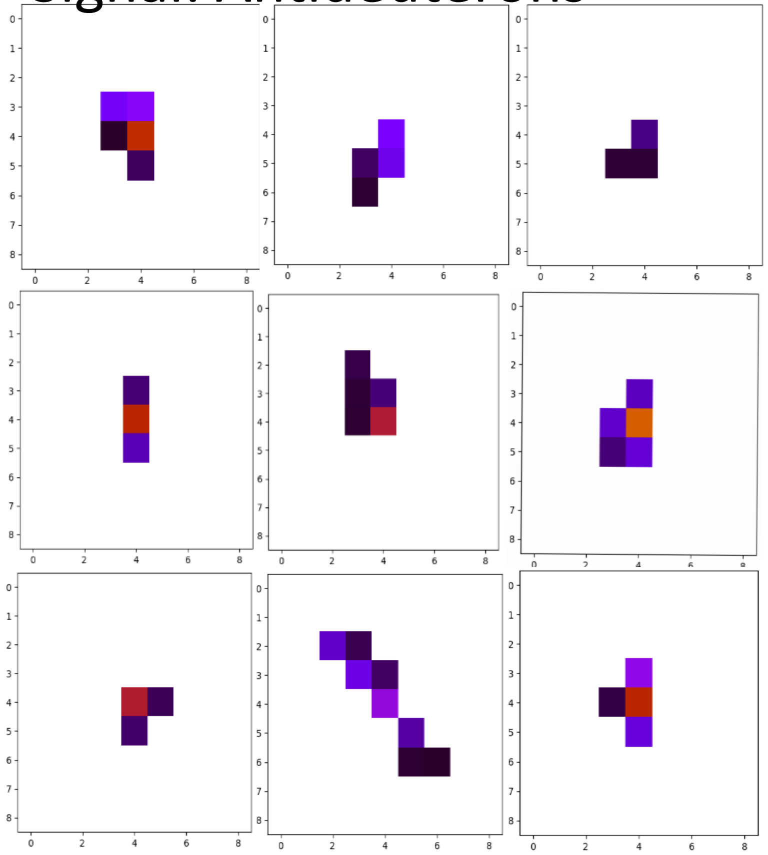


- Innermost detector
- Pixelated silicon sensors (PXD)
- 2 layers of 40 sensors each
- 8 M pixels

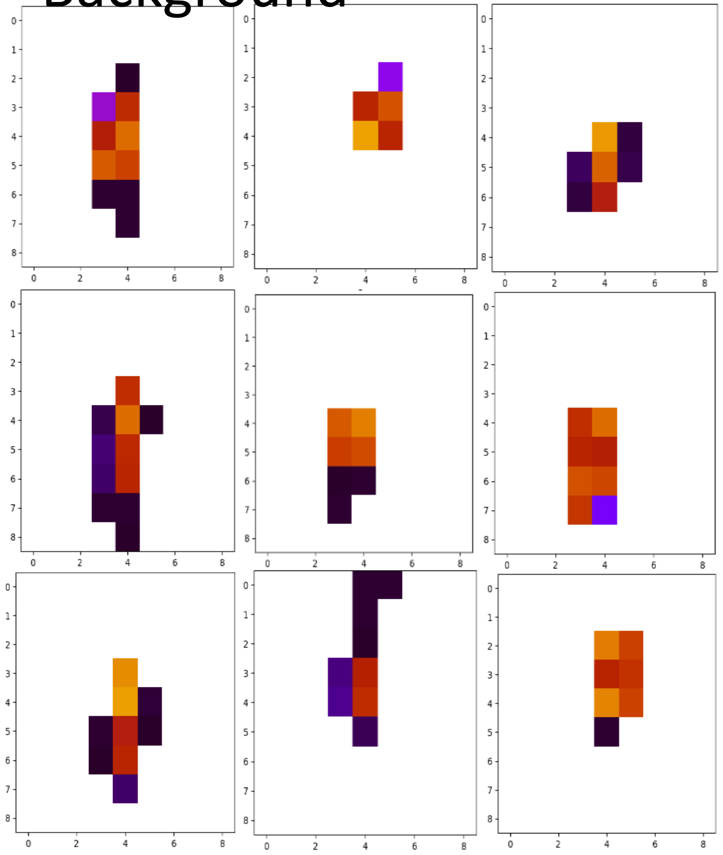
Captures highly ionizing particles.

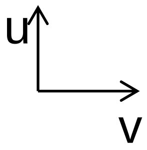
PXD clusters

Signal: Antideuterons

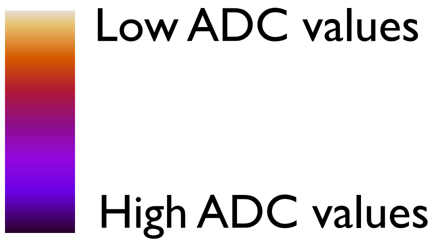


Background





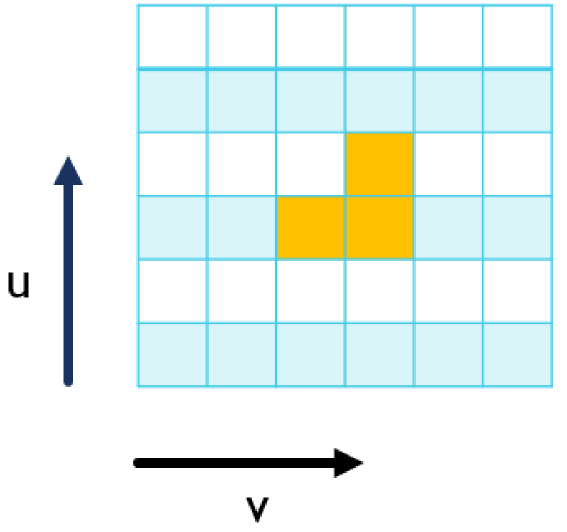
9x9 matrix
ADC values



M. Peter, Unpublished

BELLE II PIXELDETECTOR

Cluster properties



Total charge	Minimum charge	Maximum charge
Total size	Size in u	Size in v

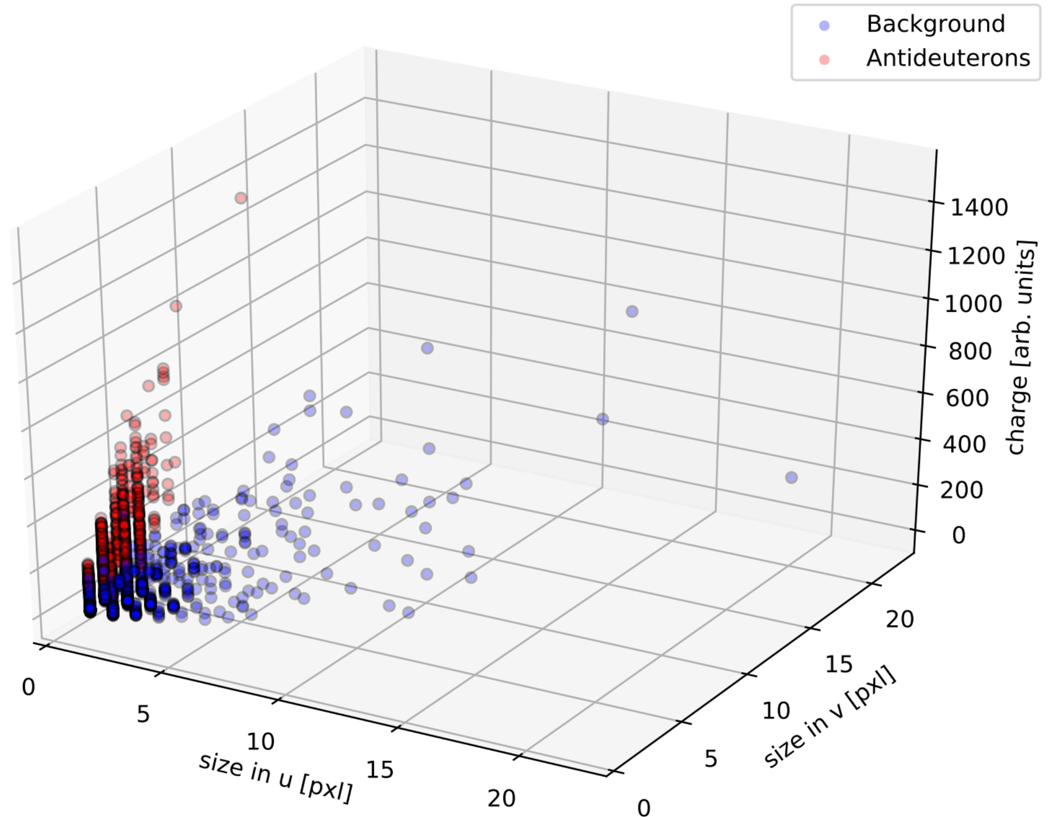
ANTIDEUTERON DATASET

Goal

Differenciate between….

$\bar{d}$

Back-ground



S. Käs, Multiparameter Analysis of the Belle II Pixeldetector's Data.

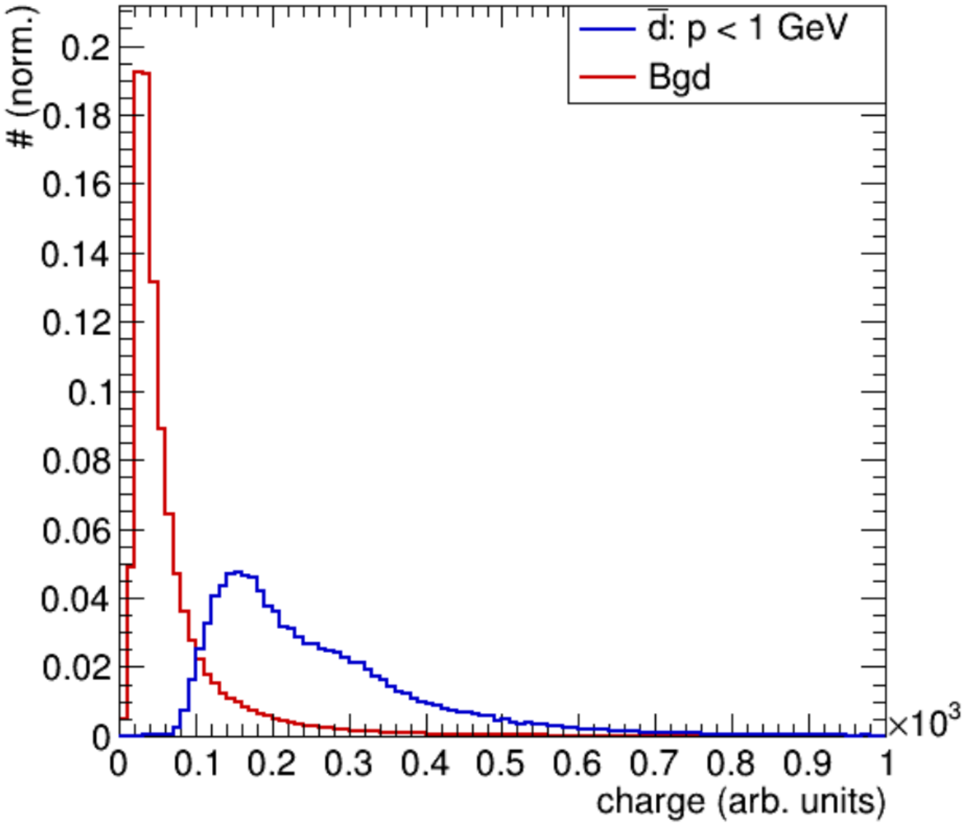
ANTIDEUTERON DATASET

Goal

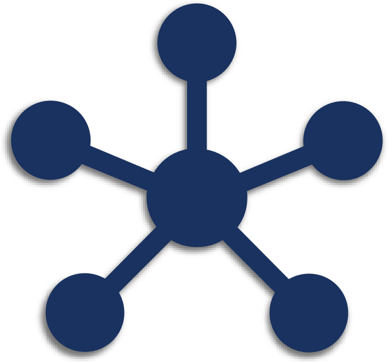
Differenciate between….

$\bar{d}$

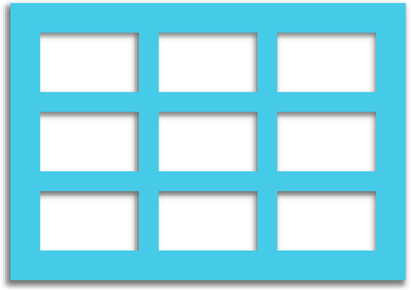
Back-ground



K. Dort, Search for Highly Ionizing Particles with the Pixel Detector in the Belle II Experiment.



Classic MLP & Kohonen Maps



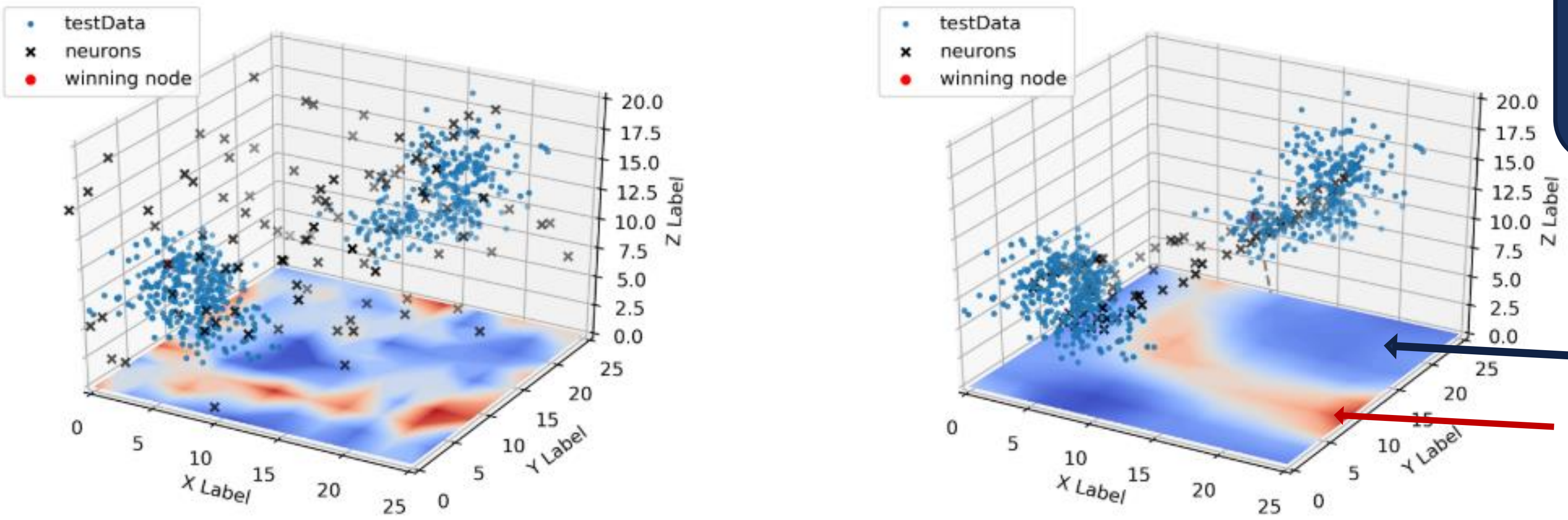
Hopfield Networks, Voxels, Autoencoder…

Networks on FPGA

PROJECTS RELATED TO DATA SCIENCES AND AI….

KOHONEN MAPS

- Unsupervised learning
- Self-organizing



Low distance between nodes

High distance between nodes

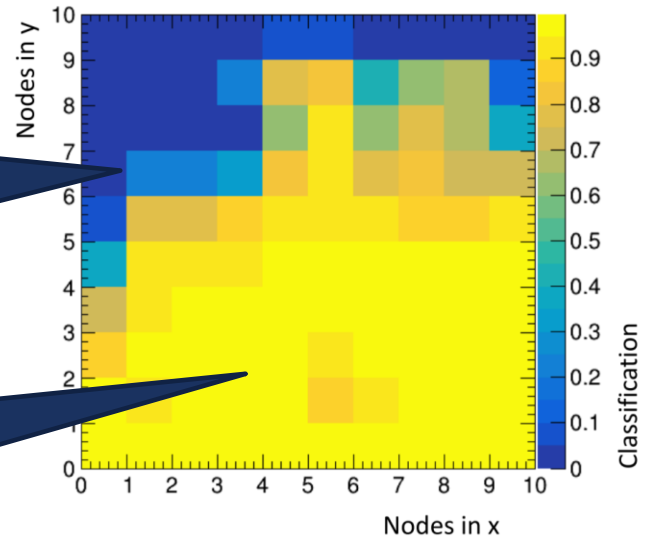
J. Bilk & J. Budak, Detecting Clusters in Highdimensional Data

Figure 6: The first step of a self-organizing map. One can see the data clouds in blue, the vectors of each neuron as black x's and on the floor the U-matrix.

KOHONEN MAPS

Background-like nodes

Antideuteron - like nodes



S. Käs, Multiparameter Analysis of the Belle II Pixeldetector's Data.

Self-Organizing Maps Parameters

15 x 15 Nodes

Neighborhood function: Gaussian

Width of Gaussian: 7

Learning Rate: 0.01

KOHONEN MAPS

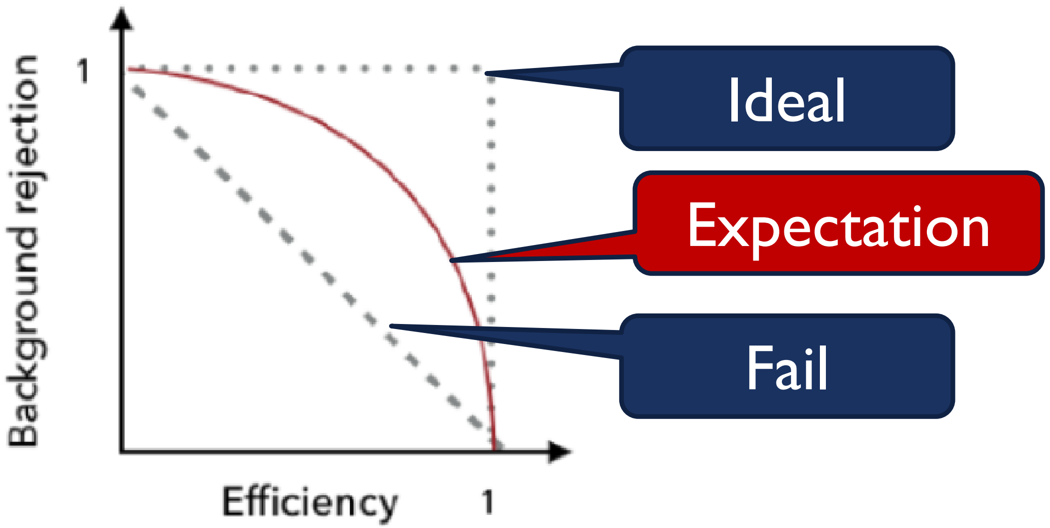
Efficiency

$$\frac{\text{number of correctly identified signals}}{\text{total number of signals}}$$

Background rejection

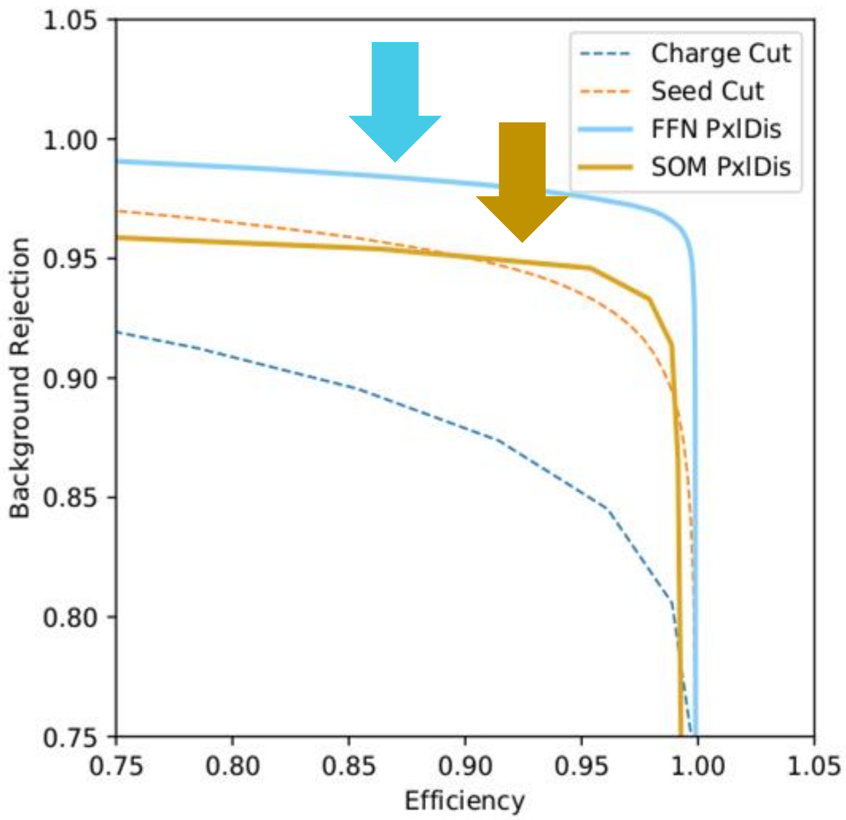
$$\frac{\text{number of correctly identified BG data}}{\text{total number of BG data}}$$

Reminder: ROC-Curves



K. Dort, Search for Highly Ionizing Particles with the Pixel Detector in the Belle II Experiment.

COMPARISON: KOHONEN MAP & MLP



MLP performed better.

K. Dort, Search for Highly Ionizing Particles with the Pixel Detector in the Belle II Experiment.

Classic MLP & Kohonen Maps

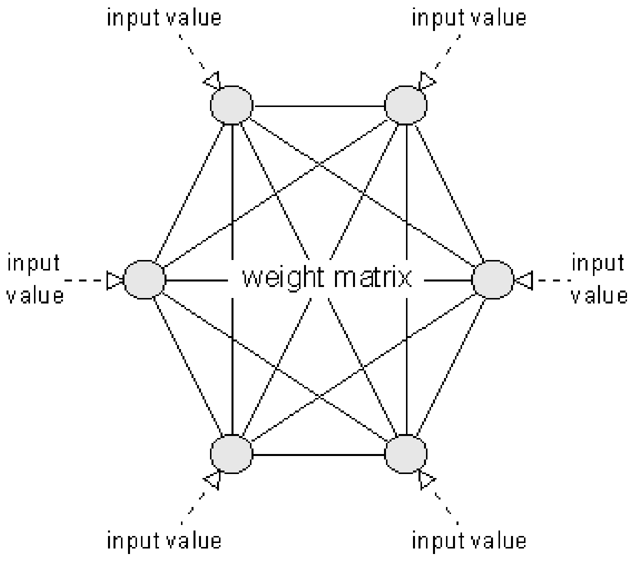
Hopfield Networks, Voxels, Autoencoder…

Networks on FPGA

PROJECTS RELATED TO DATA SCIENCES AND AI….

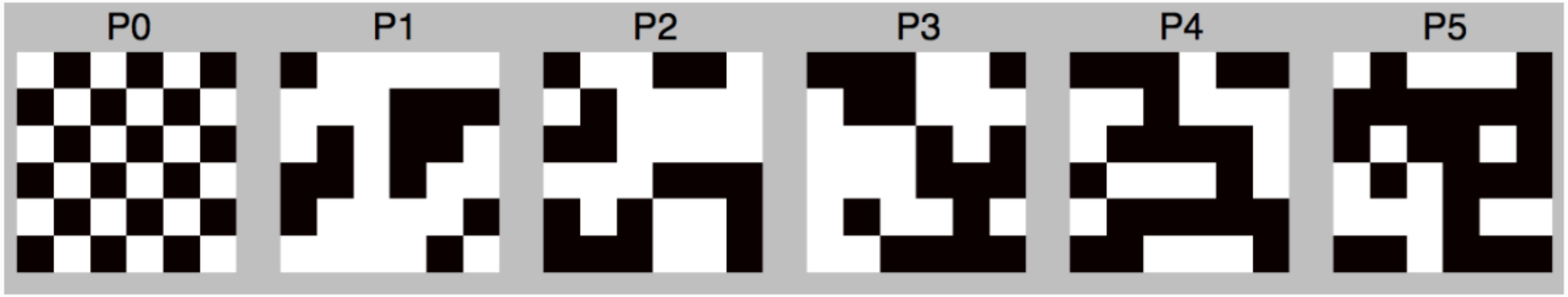
HOPFIELD NETWORK

- There is only 1 layer
  input layer = output layer
- Each node is connected to each node



nnwj.de

neuronaldynamics-exercises.html



Six patterns are stored in a Hopfield network.

HOPFIELD NETWORK

Hopfield networks require binary data:

ADC value

0	0	0	0	0
0	149	30	0	0
0	67	7	0	0
0	10	0	0	0
0	0	0	0	0

$\longrightarrow$

0	0	0	0	0
0	1	1	0	0
0	1	1	0	0
0	1	0	0	0
0	0	0	0	0

I. Heinz, Hopfield Network for Cluster PID at the PXD

How can we represent other cluster properties (*charge, shape, …*)?

HOPFIELD NETWORK

Hopfield networks require binary data:

0	0	0	0	0
0	149	30	0	0
0	67	7	0	0
0	10	0	0	0
0	0	0	0	0

$\longrightarrow$

0	0	0	0	0
0	1	1	0	0
0	1	1	0	0
0	1	0	0	0
0	0	0	0	0

I. Heinz, Hopfield Network for Cluster PID at the PXD

How can we represent other cluster properties (*charge*, *shape*, …)?

Customized activation function

HOPFIELD NETWORK – CUSTOMIZED ACTIVATION FUNCTION

I. Heinz, Hopfield Network for Cluster PID at the PXD

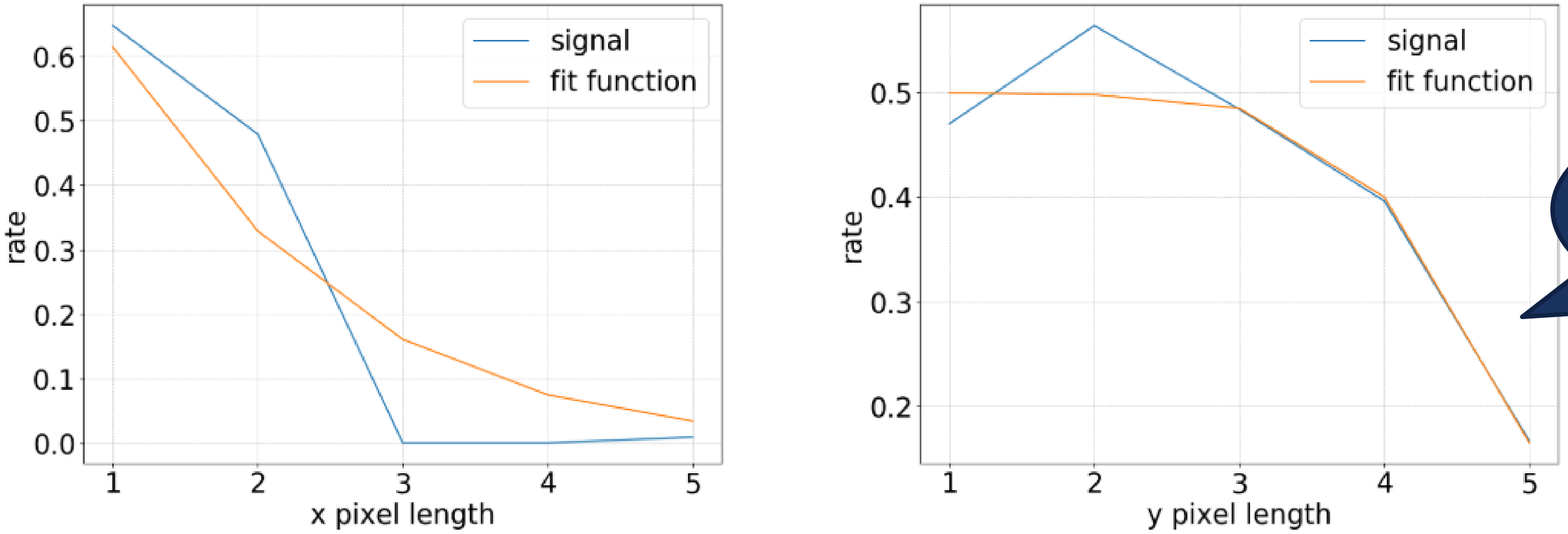
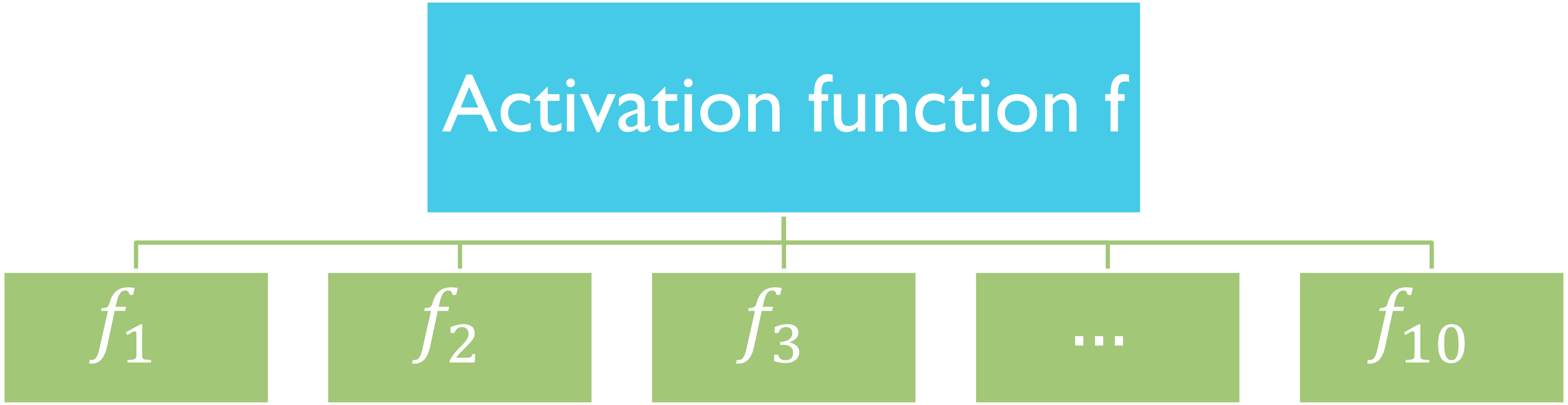


FIG. 11: Fit function for signal rate depending on **(a)** x-pixel-length n_x and **(b)** y-pixel-length n_y

Fits were made for each parameter.

Function f_i

HOPFIELD NETWORK – CUSTOMIZED ACTIVATION FUNCTION



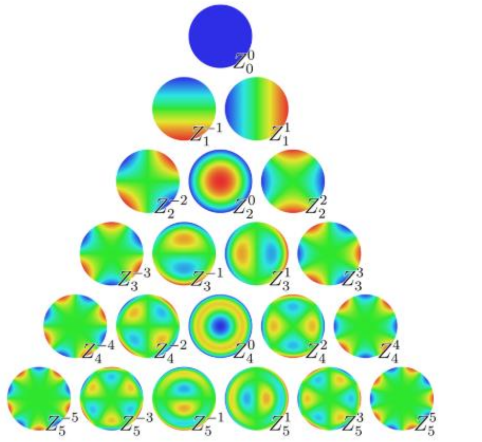
Individual functions f_i are weighted by their rate range of possible domain.

HOPFIELD NETWORK – CUSTOMIZED ACTIVATION FUNCTION

Activation function f

f_1 | f_2 | f_3 | … | f_{10}

Individual functions f_i are weighted by their rate range of possible domain.



I. Heinz, Hopfield Network for Cluster PID at the PXD

Cluster shape added via Zernicke moments

HOPFIELD NETWORK – RESULTS

Efficiency

$$\frac{\text{number of correctly identified signals}}{\text{total number of signals}}$$

Background rejection

$$\frac{\text{number of correctly identified background}}{\text{total number of background}}$$

Efficiency
96.83%

BG rejection
98.49%

HOPFIELD NETWORK – RESULTS

Efficiency

$$\frac{\text{number of correctly identified signals}}{\text{total number of signals}}$$

Background rejection

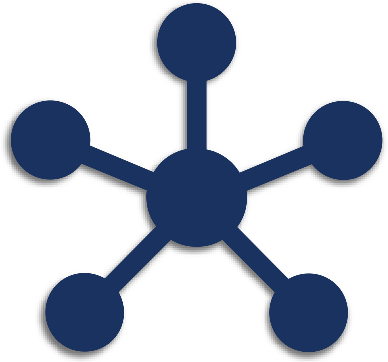
$$\frac{\text{number of correctly identified background}}{\text{total number of background}}$$

Efficiency
96.83%

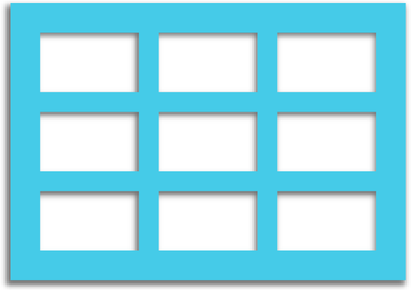
BG rejection
98.49%

Downsides?

- Requires preprocessing
- Static / not versatile



Classic MLP & Kohonen Maps



Hopfield Networks, Voxels, Autoencoder…

Networks on FPGA

PROJECTS RELATED TO DATA SCIENCES AND AI….

VOXELS - SPHERICAL

- Ideal radius?
- Curse of dimensionality

92.38%

coverage of data

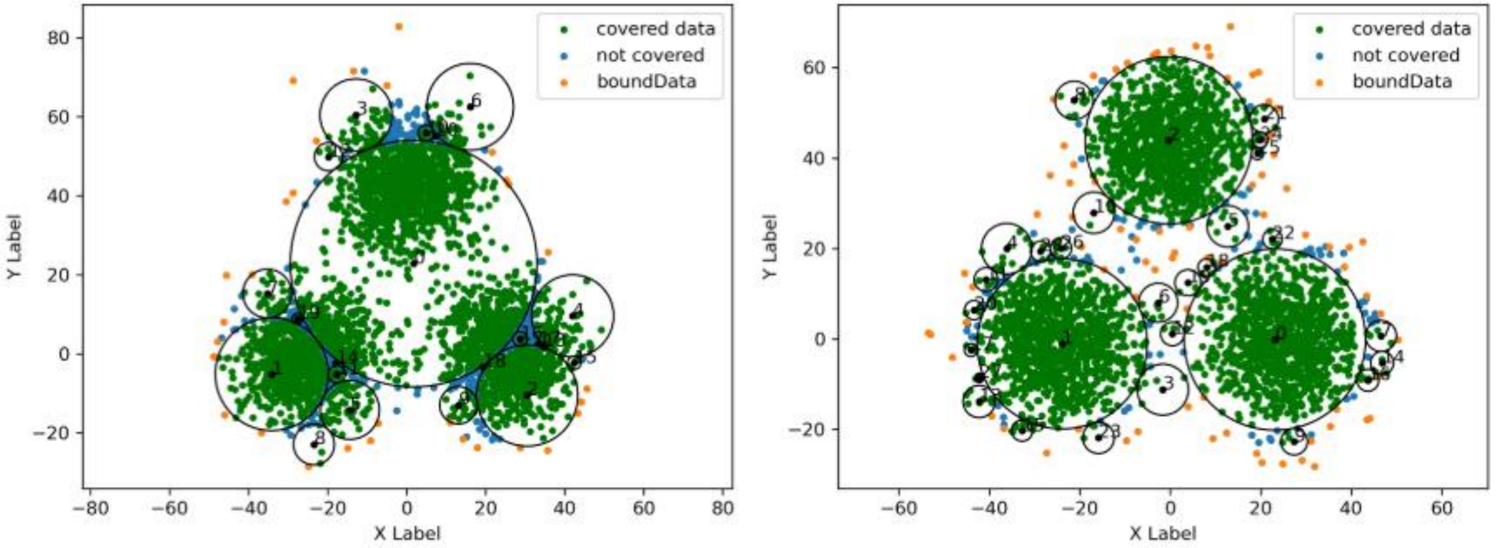


Figure 19: Demonstrating what impact the input parameters have on finding the boundary and thus on the end result. J. Bilk & J. Budak, Detecting Clusters in Highdimensional Data

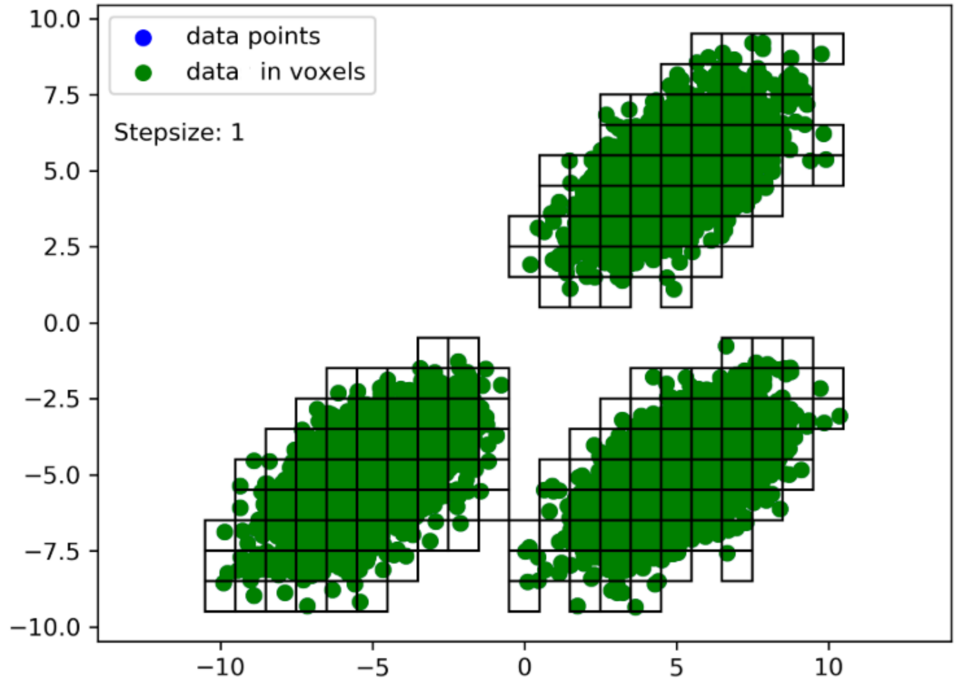
VOXELS

Spherical voxels

92.38%
coverage of data

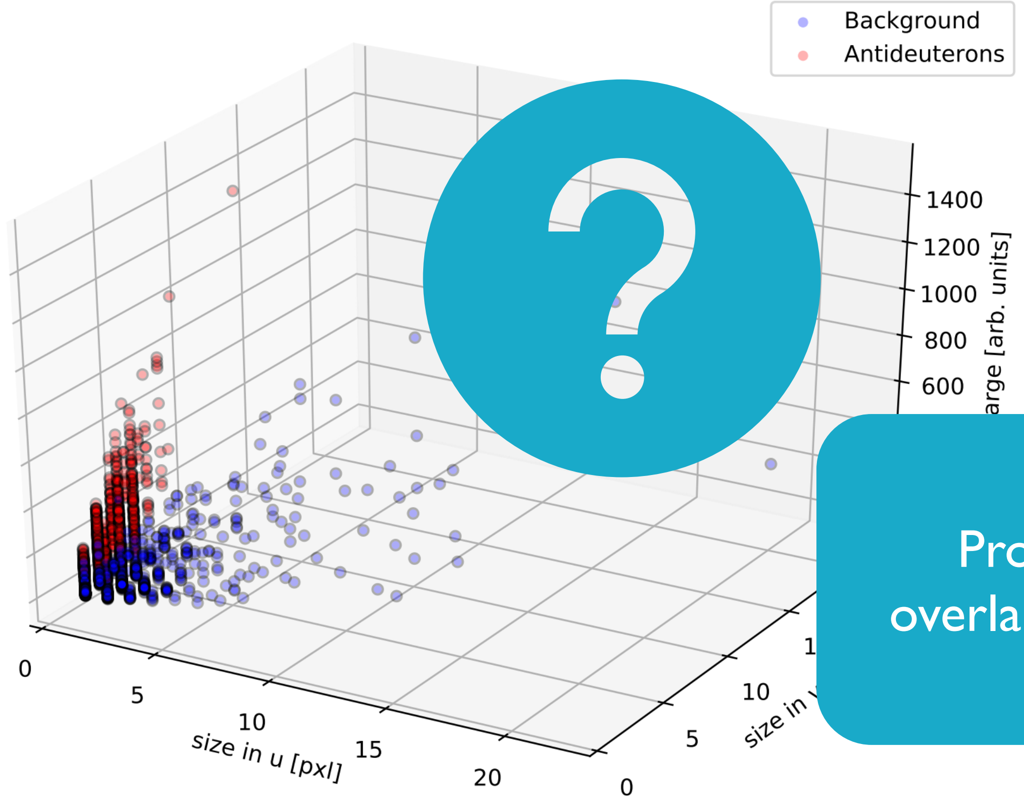
Cubic voxels

>99%
coverage of data



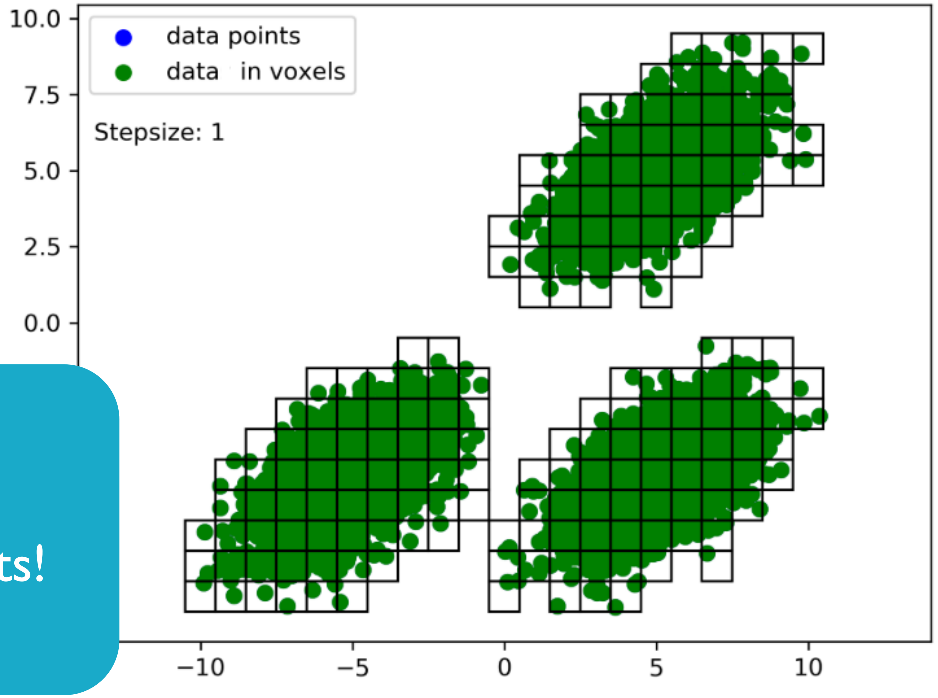
J. Bilk & J. Budak, Detecting Clusters in Highdimensional Data

VOXELS



S. Käs, Multiparameter Analysis of the Belle II Pixeldetector's Data.

Problematic on overlapping data sets!



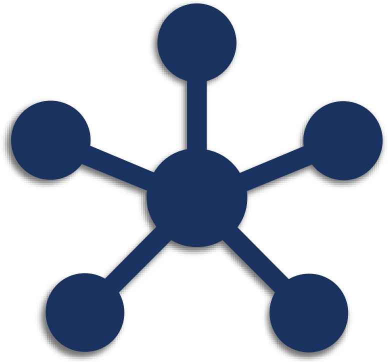
J. Bilk & J. Budak, Detecting Clusters in Highdimensional Data

ONGOING PROJECTS

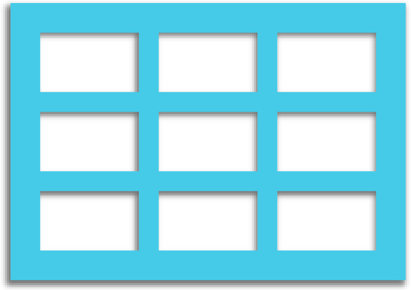
Elastic Matching

CNNs

…will be presented in 2022!



Classic MLP & Kohonen Maps



Hopfield Networks, Voxels, Autoencoder…

Networks on FPGA

PROJECTS RELATED TO DATA SCIENCES AND AI….

FPGA HARDWARE IMPLEMENTATION

Two approaches:

Matrixmult. on FPGA
(DSP Slices)

High Level Sythesis
(C++ on FPGA)

FOR FAST MATRIX
OPERATIONS

FPGA Data Acquisition System of the PXD

Ongoing projects by

Peter Lehnhardt &
Falk Zorn

Classic MLP & Kohonen Maps

Hopfield Networks, Voxels, Autoencoder…

Networks on FPGA

PROJECTS RELATED TO DATA SCIENCES AND AI….

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
Belle II

WE HAVE A DATE TOMORROW!

18:00 -18:15

Topic: Autoencoder

Identification of exotic highly ionising particles at the Belle II pixel detector using unsupervised autoencoders

By Katharina Dort

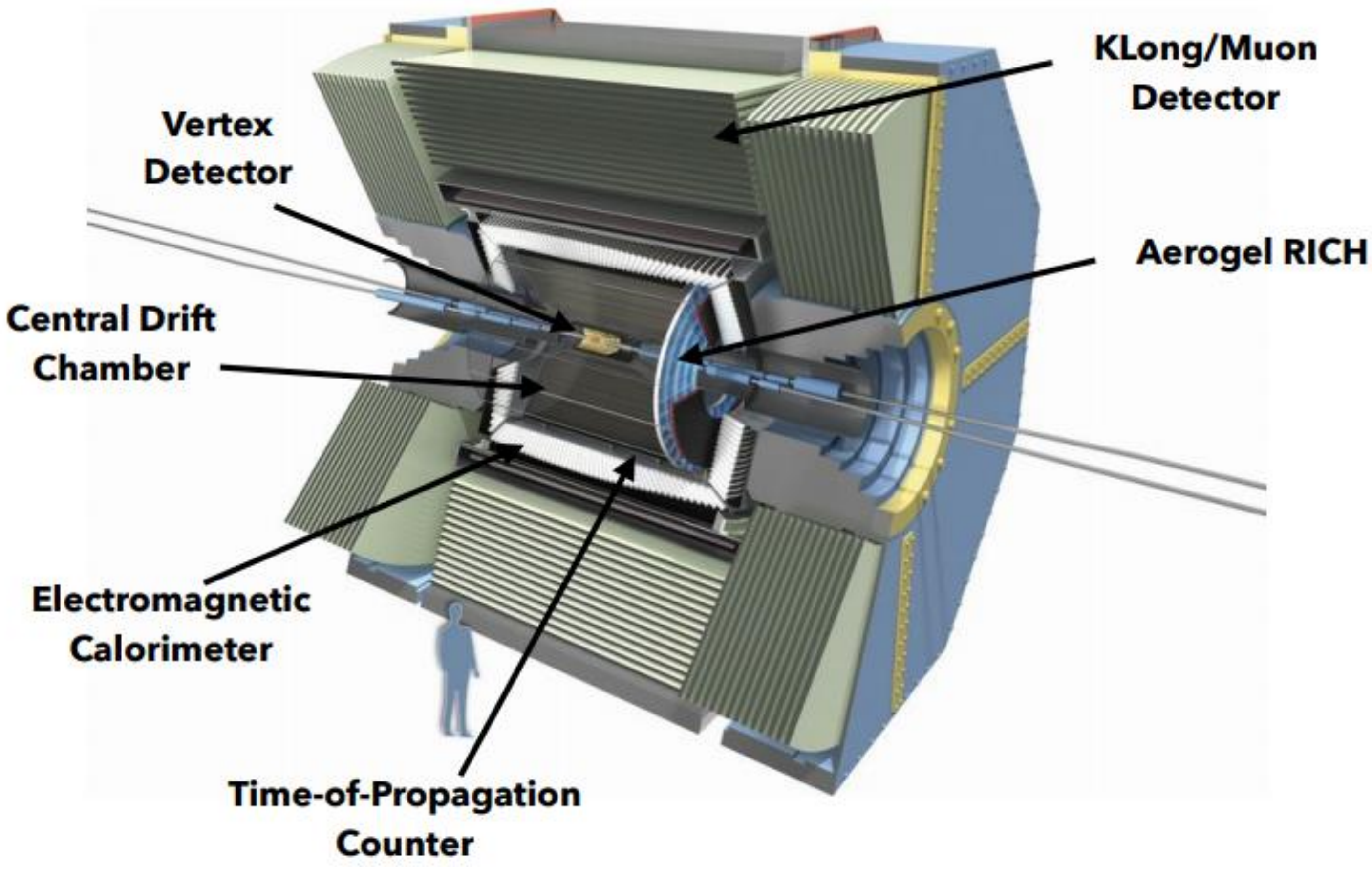
THANK YOU!

QUESTIONS?

STEPHANIE.KAES@PHYSIK.UNI-GIESSEN.DE

BACKUP

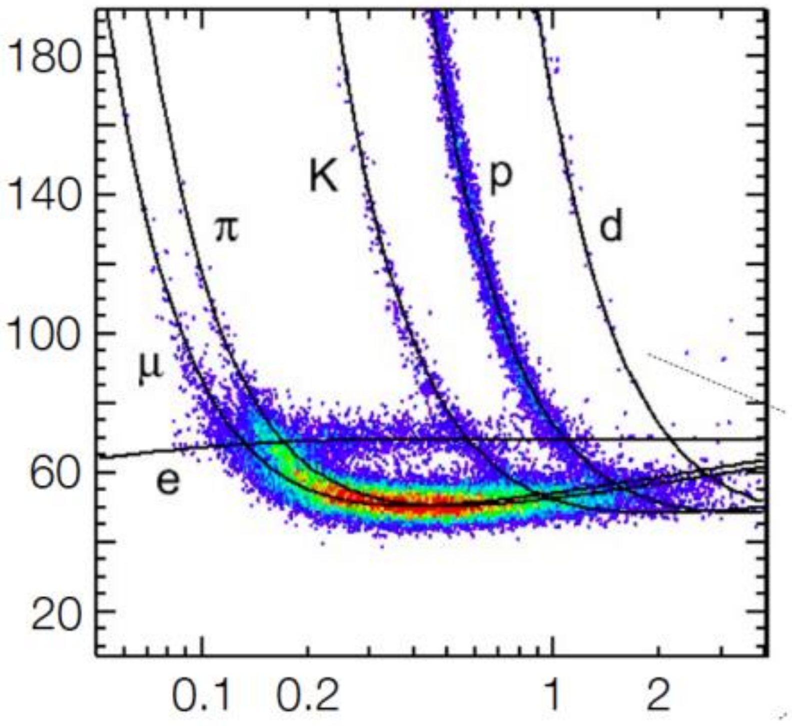
BELLE II DETECTOR



K. Dort, Belle II Pixeldetector Cluster Analysis Using Neural Network Algorithms, Talk @ Vienna 13/01/20

WHAT ARE HIGHLY IONIZING PARTICLES?

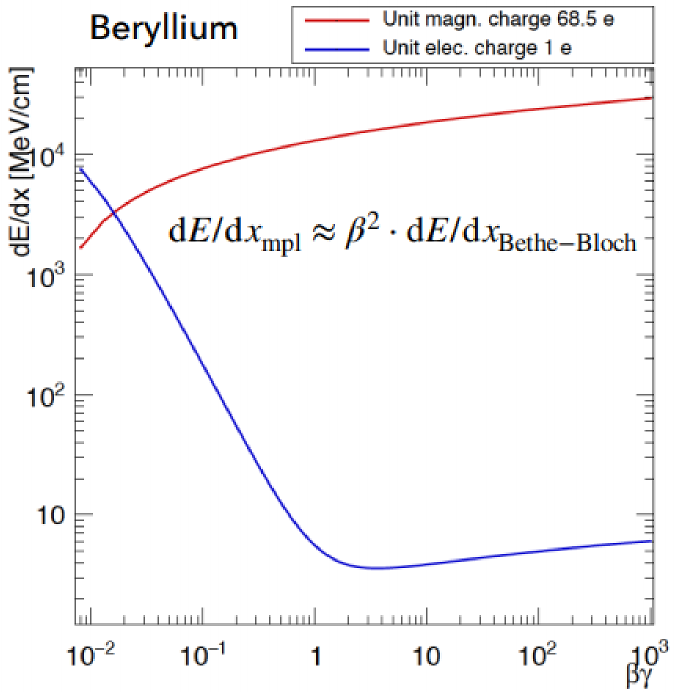
- Energy deposition of particles in matter increases with
  - decreasing momentum of projectile
    Example: slow pions
  - increasing mass of projectile
    Example: deuterons, heavy exotic particles



K. Dort, Belle II Pixeldetector Cluster Analysis Using Neural Network Algorithms, Talk @ Vienna 13/01/20

WHAT ARE HIGHLY IONIZING PARTICLES?

- Energy deposition of particles in matter increases with
  - decreasing momentum of projectile
    Example: slow pions
  - increasing mass of projectile
    Example: deuterons, heavy exotic particles
  - non-Bethe-Bloch energy loss
    Example: magnetically charged particles



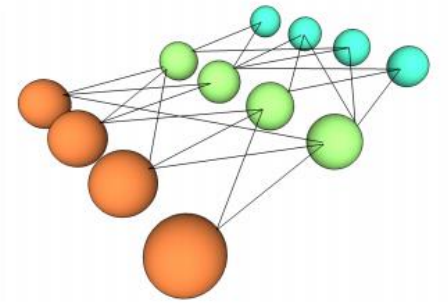
K. Dort, Belle II Pixeldetector Cluster Analysis Using Neural Network Algorithms, Talk @ Vienna 13/01/20

FEED-FORWARD NEURAL NETWORKS

Feed-Forward Network Parameters

4 layers / 2 hidden
> 50 nodes per layer
ReLu Activation Function
CrossEntropy Loss Function
Stochastic Gradient Descent (SGD) Optimizer
Batch Size: 256
Learning Rate: 0.0001
Momentum: 0.9

- Supervised learning in order to separate HIPs from beam background*
- Implemented with PyTorch and trained on CPU and GPU
- Loss and accuracy is monitored during training (~8h)
- Cut on classification axis determines accuracy of neural network

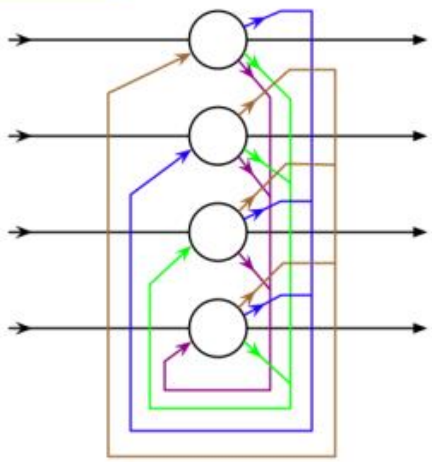


*beam background

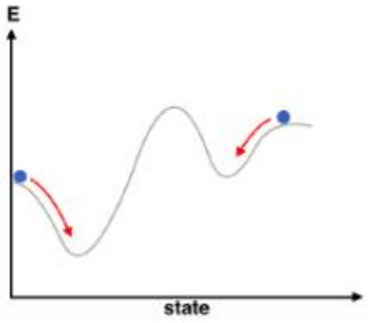
- Official mixed MC beam background
- Includes luminosity-dependent and beam-induced background

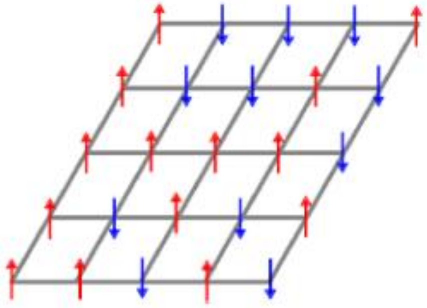
K. Dort, Belle II Pixeldetector Cluster Analysis Using Neural Network Algorithms, Talk @ Vienna 13/01/20

HOPFIELD NETWORKS



- Recurrent binary neural network (*associative* memory)
- Network learns *pattern* associated with background/signal input vector -> stored in weight matrix
- Weight between neurons determines *energy* of the entire network
- Stable state is reached when energy of network is minimized (similar to spin-spin interaction in quantum mechanical many-body systems)
- Incomplete or distorted patterns are recognized





K. Dort, Belle II Pixeldetector Cluster Analysis Using Neural Network Algorithms, Talk @ Vienna 13/01/20

HOPFIELD NETWORK – USE OF ACTIVATION FUNCTION

Signal weight matrix

Background weight matrix

$$\begin{pmatrix} w_{11}^{signal} & \cdots & w_{n1}^{signal} \\ \vdots & \ddots & \vdots \\ w_{1n}^{signal} & \cdots & w_{nn}^{signal} \end{pmatrix}$$

Seperately calculated

$$\begin{pmatrix} w_{11}^{background} & \cdots & w_{n1}^{background} \\ \vdots & \ddots & \vdots \\ w_{1n}^{background} & \cdots & w_{nn}^{background} \end{pmatrix}$$

$f \qquad 1 - f$

$$\begin{pmatrix} w_{11} & \cdots & w_{n1} \\ \vdots & \ddots & \vdots \\ w_{1n} & \cdots & w_{nn} \end{pmatrix}$$

weighted sum = full weight matrix

HOPFIELD NETWORKS

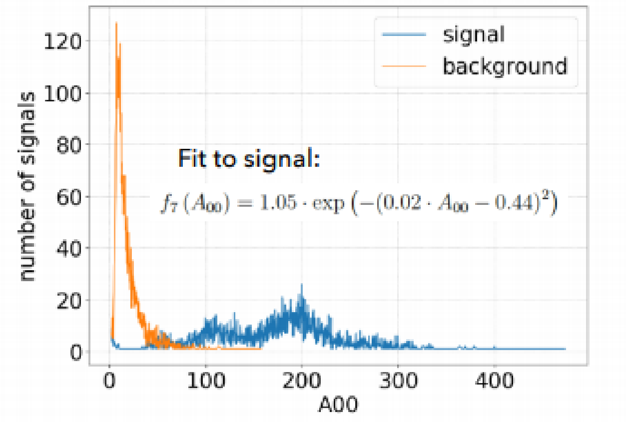
- So far, Hopfield network is trained to store 4 patterns (2 background, 2 signal)
- Custom activation function (instead of binary activation) is used to feed additional information into the network
- Local, global properties and Zernike moments used in final analysis

$$\begin{pmatrix} w_{11}^{signal} & \cdots & w_{n1}^{signal} \\ \vdots & \ddots & \vdots \\ w_{1n}^{signal} & \cdots & w_{nn}^{signal} \end{pmatrix} \qquad \begin{pmatrix} w_{11}^{background} & \cdots & w_{n1}^{background} \\ \vdots & \ddots & \vdots \\ w_{1n}^{background} & \cdots & w_{nn}^{background} \end{pmatrix}$$

$$f \searrow \quad \swarrow 1-f$$

$$\begin{pmatrix} w_{11} & \cdots & w_{n1} \\ \vdots & \ddots & \vdots \\ w_{1n} & \cdots & w_{nn} \end{pmatrix}$$

Zernike Moment A00



K. Dort, Belle II Pixeldetector Cluster Analysis Using Neural Network Algorithms, Talk @ Vienna 13/01/20

Separation of magnetic monopoles from beam background

3D	local prop.	global prop.	Zernike mom.	accuracy
✓	✗	✗	✗	63.0 %
✓	✓	✗	✗	76.5 %
✓	✓	✓	✗	95.7 %
✓	✓	✓	✓	97.7 %

Local prop - cluster size + cluster size in x/y + max. charge + cluster charge

Global prop. - local prop + global position

Zernike mom. - global prop + Zernike moments

NEURAL NETWORKS ON FPGA

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Field programmable gate arrays (FPGAs) have prominent role in data acquisition for Belle II

- Highly parallel processing architecture make FPGAs ideal candidates for machine learning tasks
- DSP slices can be used for saving resources and speed up computation

Parallel Computation Using DSP Slices in FPGA, S. Unnikrishnan et al., Procedia Technology 24 (2016) 1127-1134

- Communication with DSP slices and adaptation of neural network to FPGA architecture currently tested

- In future, neural networks could be directly integrated into the ONSEN which would also require online cluster finding on DHH or ONSEN (not implemented yet)

K. Dort, Belle II Pixeldetector Cluster Analysis Using Neural Network Algorithms, Talk @ Vienna 13/01/20

Hardware-Accelerated Neural Networks

- More performant and efficient neural network inference
- Accelerated AI application development on Xilinx FPGAs via High-Level Synthesis (HLS), FINN framework and Vitis AI development environment
- Attempt to allow real-time AI applications in time critical environments

Slide by P. Lehnhard.