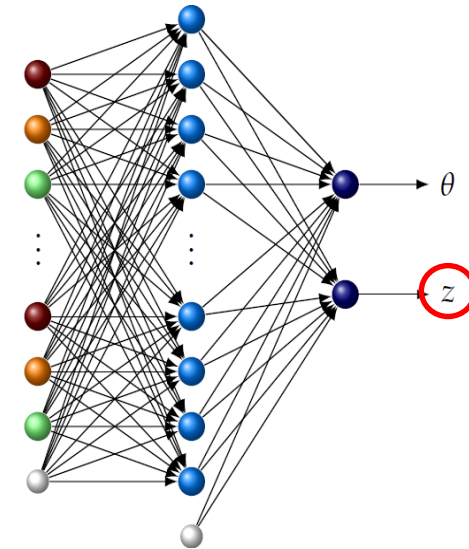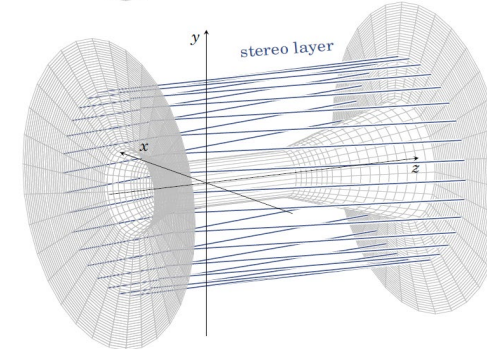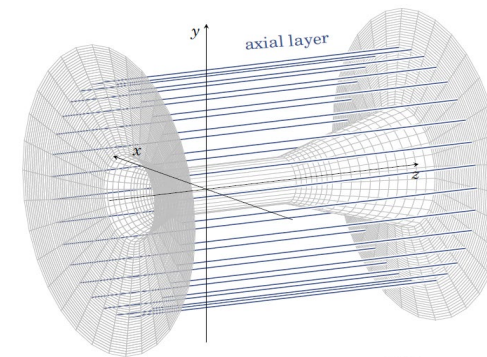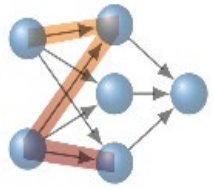# The Neural Network First-Level Hardware Track Trigger of the Belle II Experiment

Christian Kiesling
Max-Planck-Institute for Physics

**Overview:**

- SuperKEKB & Belle II's „Conventional" Track Trigger
- Principles of the Neural Approach to Track Triggers
- Physics-motivated Preprocessing of Input Variables
- Performance of the Neural Track Trigger,
  -> Launch of a Minimum Bias Single Track Trigger (STT)
- Problems and Solutions -> Upgrade program
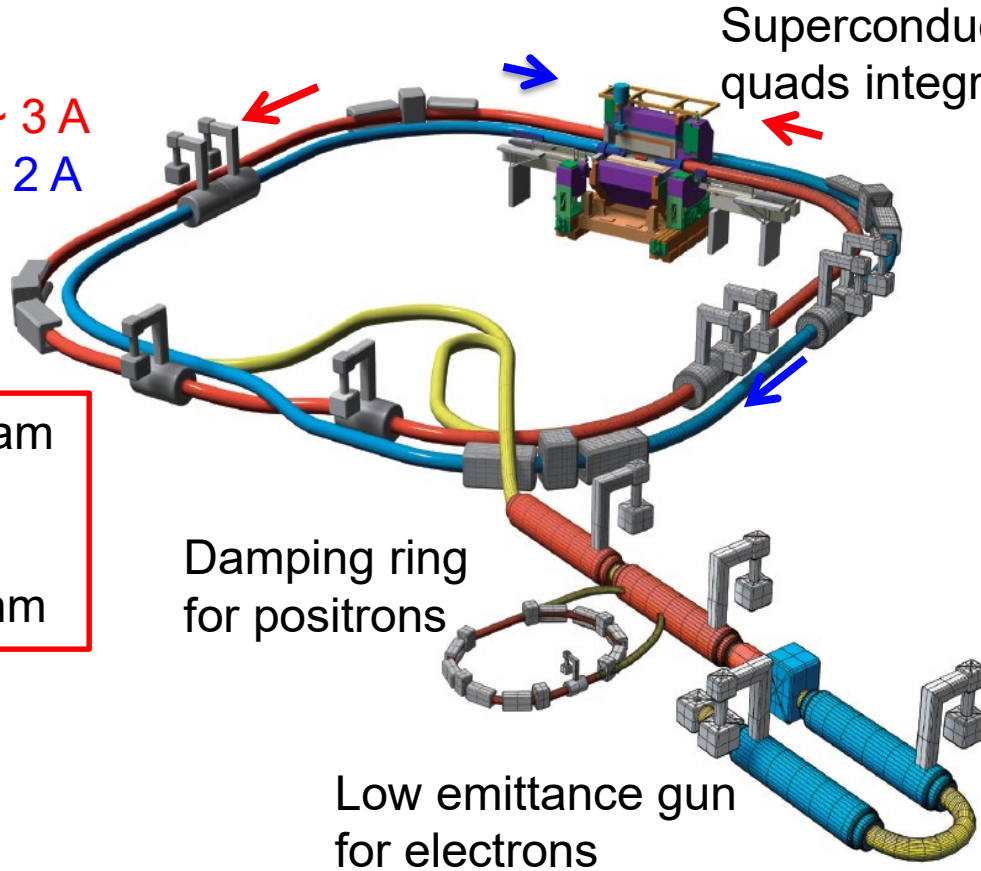- Summary and Conclusions

$e^+$ 4GeV ~ 3 A
$e^-$ 7GeV ~ 2 A

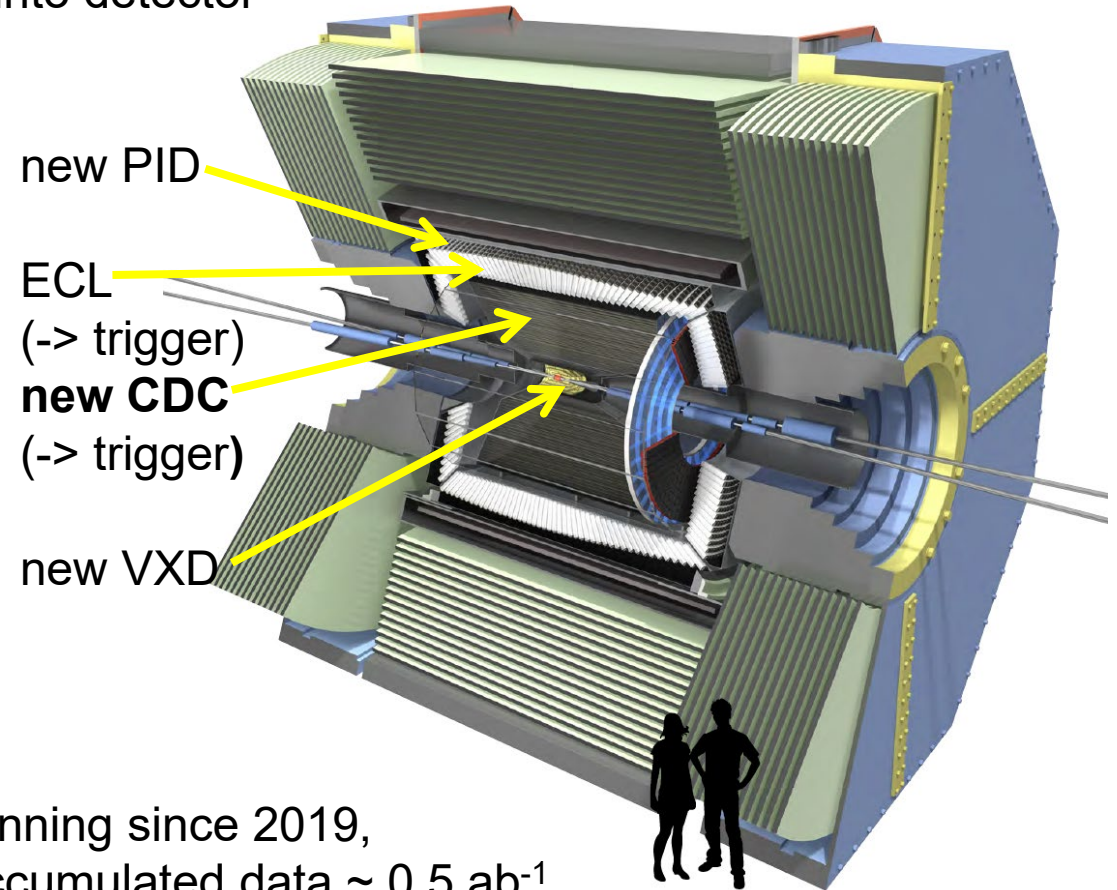Superconducting final focusing quads integrated into detector

Nano-beam scheme:

$\sigma_y$ ~ 50 nm

Damping ring for positrons

Low emittance gun for electrons

target $\mathcal{L}$ = 6x10$^{35}$/cm$^2$/s

located @ KEK, Tsukuba, Japan

new PID

ECL
(-> trigger)
**new CDC**
(-> trigger**)**

new VXD

running since 2019,
accumulated data ~ 0.5 ab$^{-1}$

peak luminosity $\mathcal{L}$ = 4.7 x 10$^{34}$ /cm$^2$/s
I($e^+$ / $e^-$) = (1.4/1.2 A) , $\beta^*$ = 1 mm
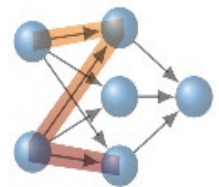
# The „Conventional" Belle II L1 Track Trigger („2D")

CDC

≈2.4 m

5 axial super layers
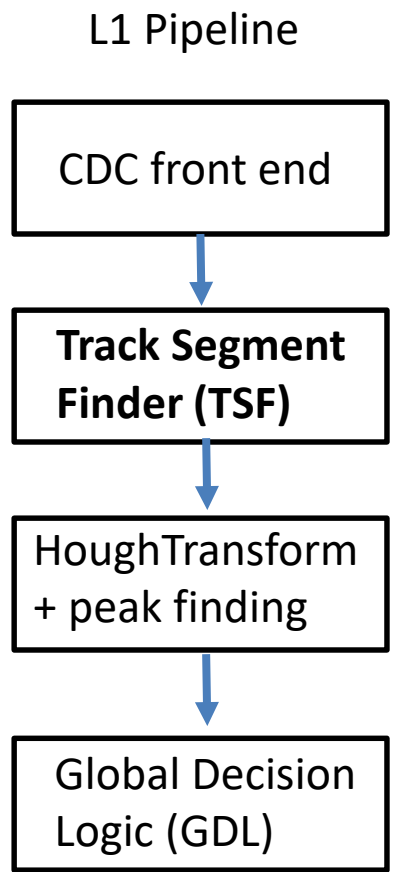
4 stereo super layers

≈1.2 m

≈16 cm

z
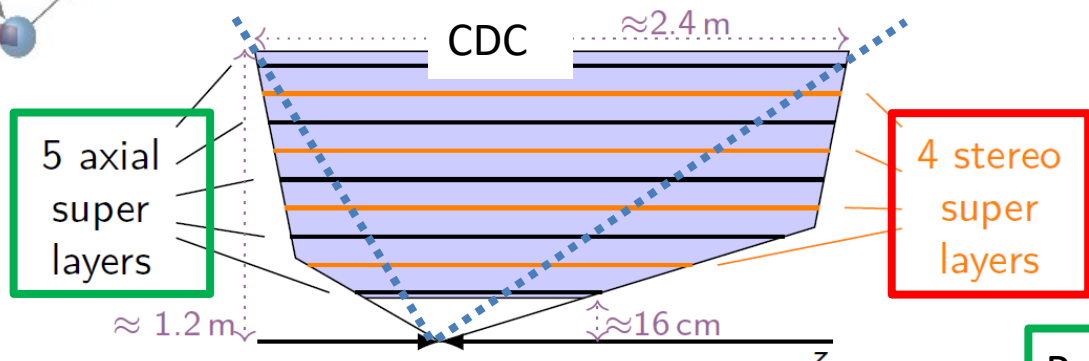
▶ 56 layers combined to 9 super layers (SL)
▶ 2336 track segments (TS) in 9 SL

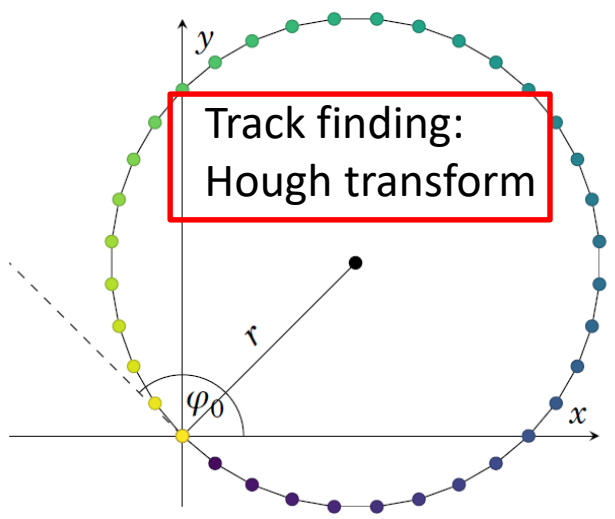**Track Segments**: Hit patterns compatible with traversing track (LUT)
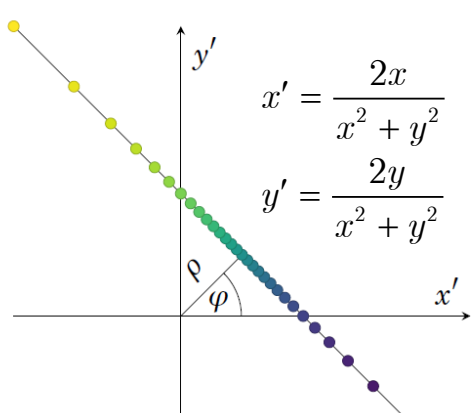
Priority wire = „hit"

5 axial SLs

L R

1$^{st}$

R/L
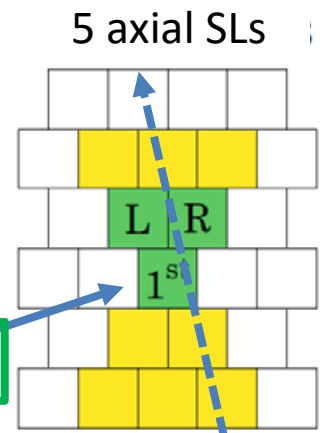
Axial track segments (ATS)

geometrical space → conformal space → parameter space

Track finding: Hough transform

$$x' = \frac{2x}{x^2 + y^2}$$

$$y' = \frac{2y}{x^2 + y^2}$$
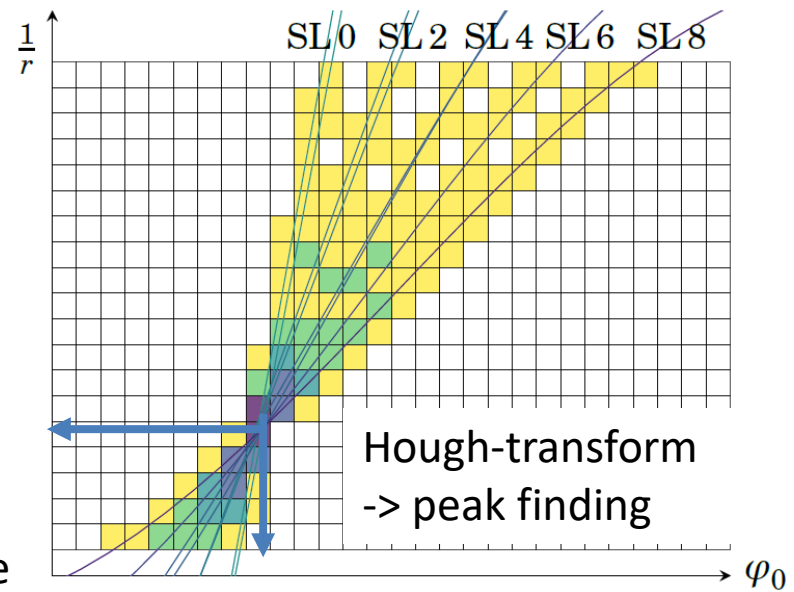
Tracks from IP by definition

Each {hit + IP} produces a set of [1/r, φ] points, basically on a straight line

SL 0  SL 2  SL 4  SL 6  SL 8

$\frac{1}{r}$

$\varphi_0$

Hough-transform -> peak finding

L1 Pipeline

CDC front end

**Track Segment Finder (TSF)**

HoughTransform + peak finding

Global Decision Logic (GDL)

5μs

algos on FPGA boards „UT(3)"
Virtex 6 XC6VHX380/565T

z-vertex distribution (offline) :

L1 Pipeline

Early lumi
running
@ Belle II

$e^+e^-$
collisions

| Entries | 367395 |
|---------|--------|
| Mean | 31.48 |
| Std Dev | 52.92 |

z [cm]

CDC front end

Track Segment
Finder (TSF)

**Hough transform
„2D tracks"**

Track Trigger derived from
Central Drift Chamber:

axial track
segments
(TS)

Belle II:
Initially, track trigger
only in 2D, using Hough
transforms
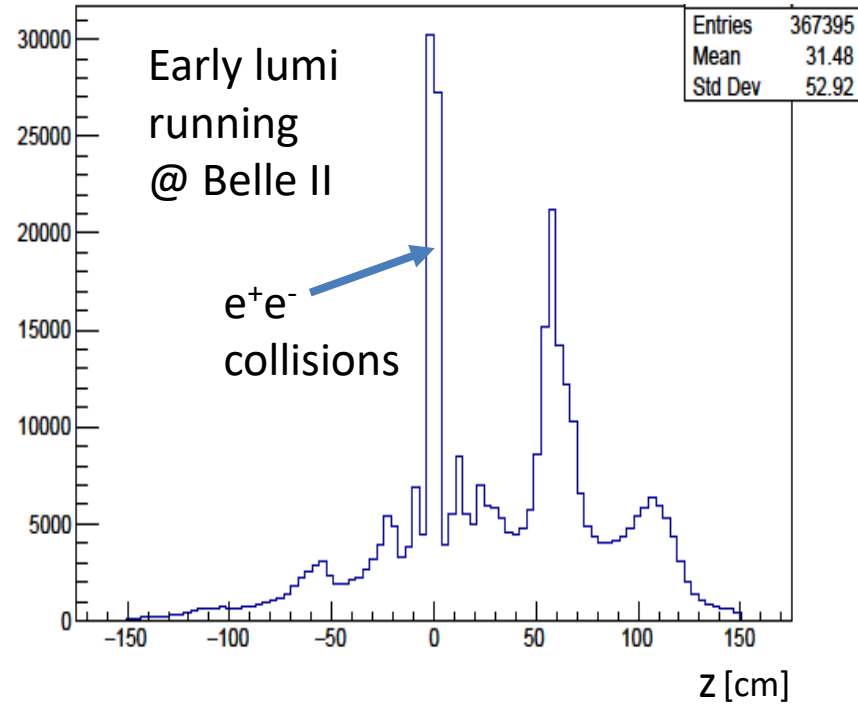
Majority of tracks from
„obstacles" outside of the
interaction region (IP)
(|z| >> 1 cm): only ~10% from IP

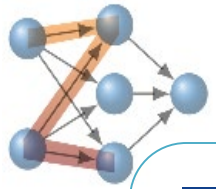# 2D tracks >= 2

„z-vertex" trigger mandatory
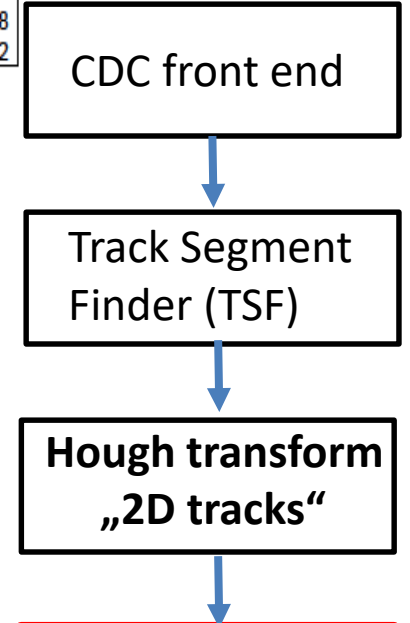
3D track fit:
iterative process
- Latency ??
- precision ??

-> Machine
Learning

5µs

# AI Trigger Group at Belle II
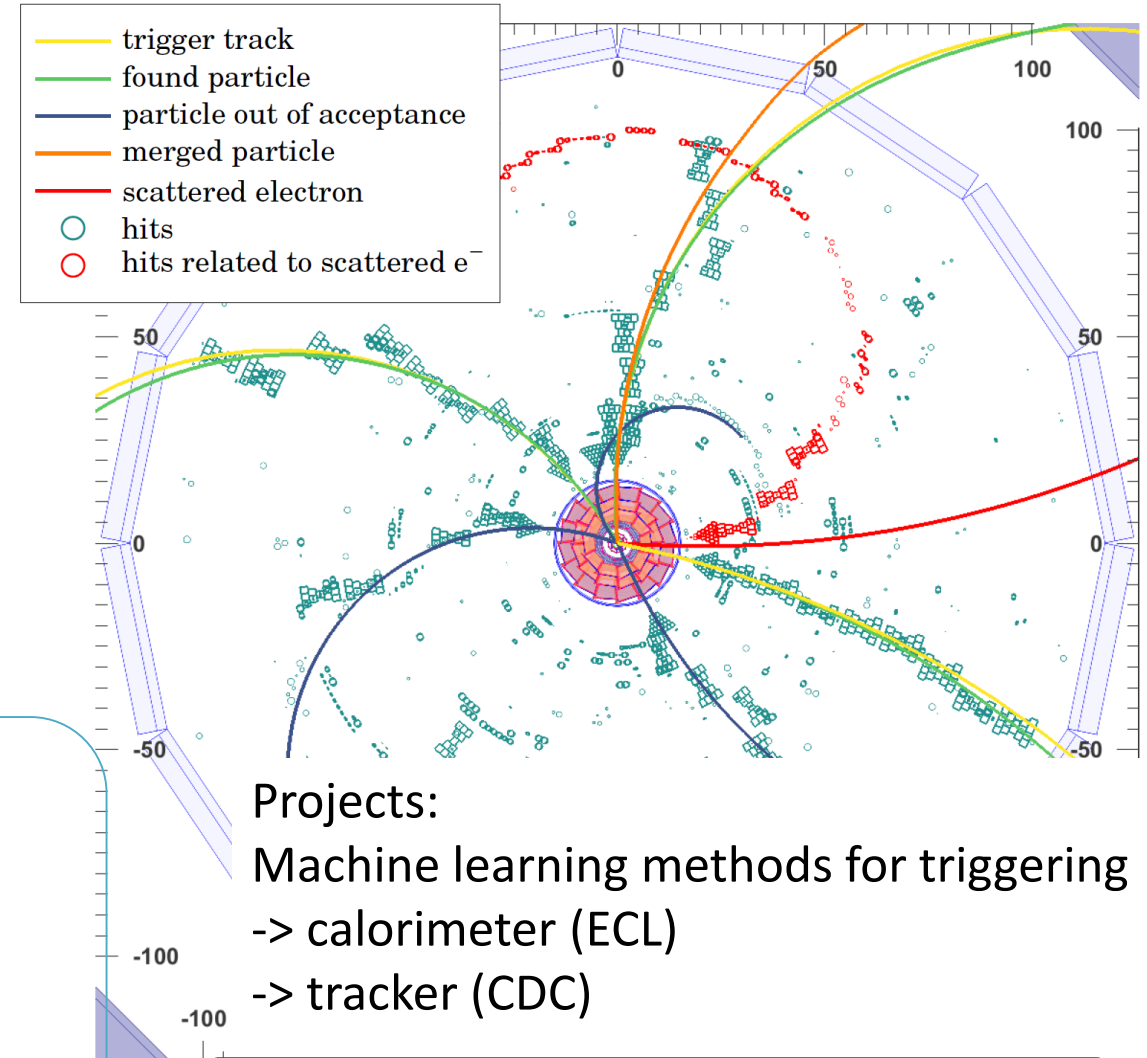
## KIT ITIV
- Marc Neu
- Kai Unger
- **Jürgen Becker**

## KIT ETP
- Lea Reuter
- Greta Heine
- Slavomira Stefkova
- **Torben Ferber**

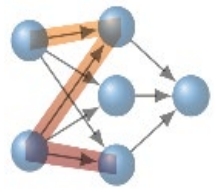## MPI / LMU / TUM
- Felix Meggendorfer
- Simon Hiesl
- Timo Forsthofer
- **Christian Kiesling**
- Alois Knoll



Legend:
- trigger track
- found particle
- particle out of acceptance
- merged particle
- scattered electron
- ○ hits
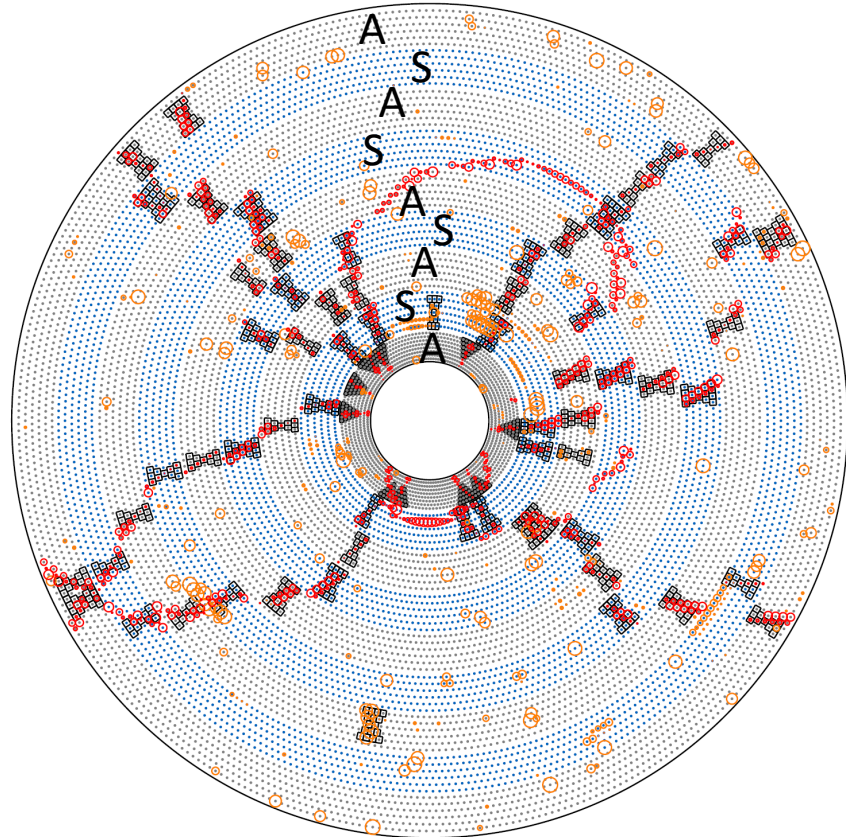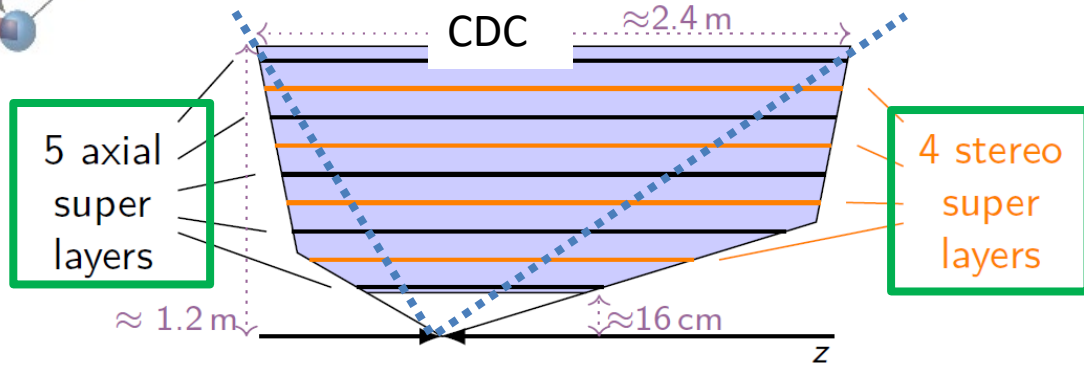- ○ hits related to scattered e⁻

Projects:
Machine learning methods for triggering
-> calorimeter (ECL)
-> tracker (CDC)

here: Neural Network „z" Trigger @L1

# Principle of the Neural L1 Track Trigger

CDC  ≈2.4 m

5 axial super layers

4 stereo super layers

≈ 1.2 m   ≈16 cm

z

A S A S A S A S A

Architecture for each track candidate
(networks to solve a regression task)

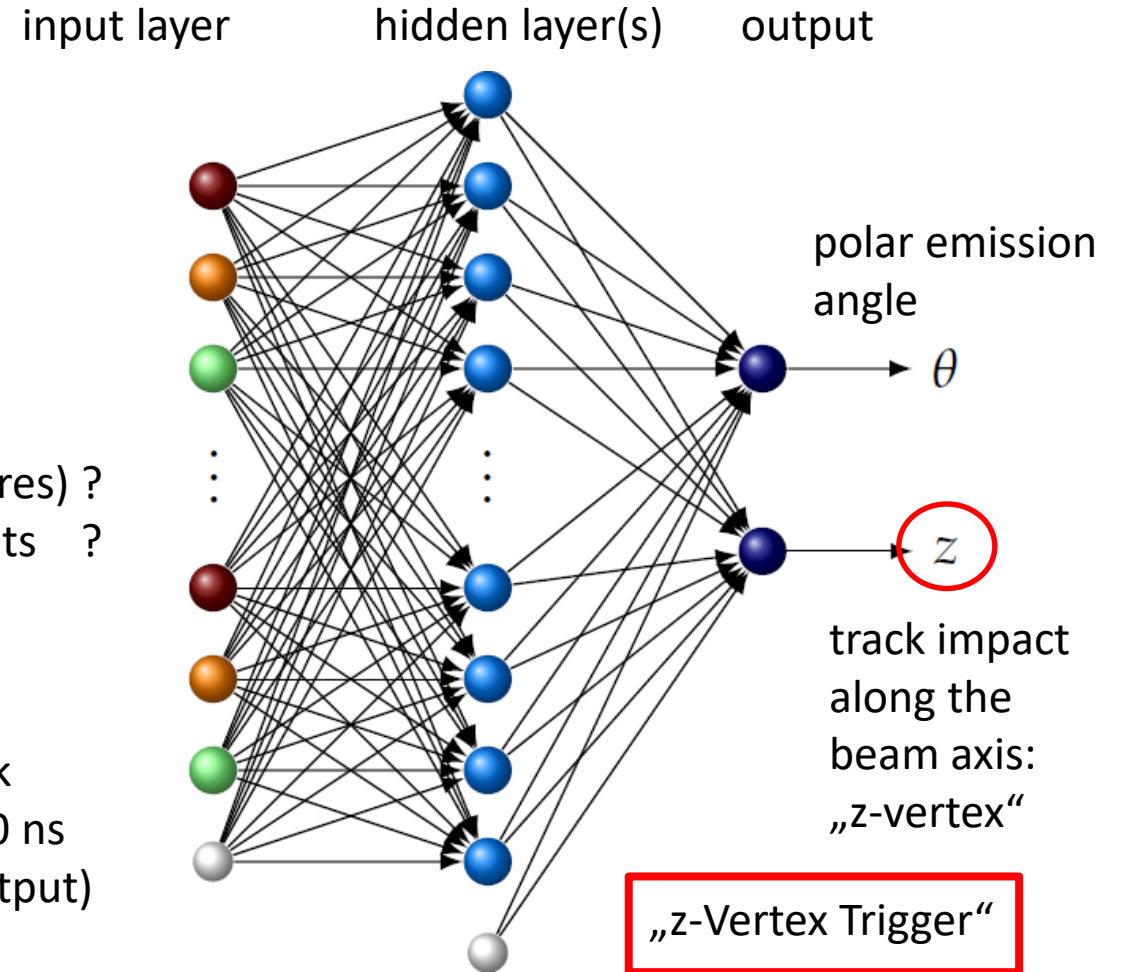input layer    hidden layer(s)    output
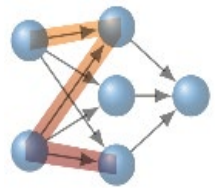
Central question:

What are the input variables:

- entire „picture" (wires) ?
- set of track segments    ?
- ?

Note:
total latency for track reconstruction  ~ 700 ns
(starting with TSF output)

polar emission angle
$\theta$

$z$

track impact along the beam axis:
„z-vertex"

„z-Vertex Trigger"

# Input Preprocessing & Neural Networks

CDC

$\approx 2.4$ m

5 axial super layers

4 stereo super layers

$\approx 1.2$ m

$\approx 16$ cm

$z$

$1^{st}$ priority, passage left

$2^{nd}$ priority left, passage right

$2^{nd}$ priority right, passage undecided

LUT

from all 5 aSLs / 4 sSLs

$\varphi$
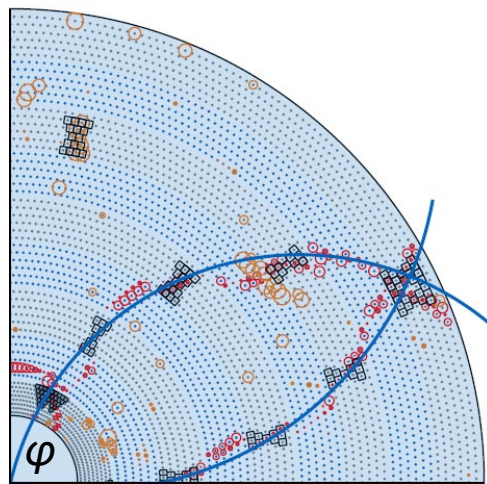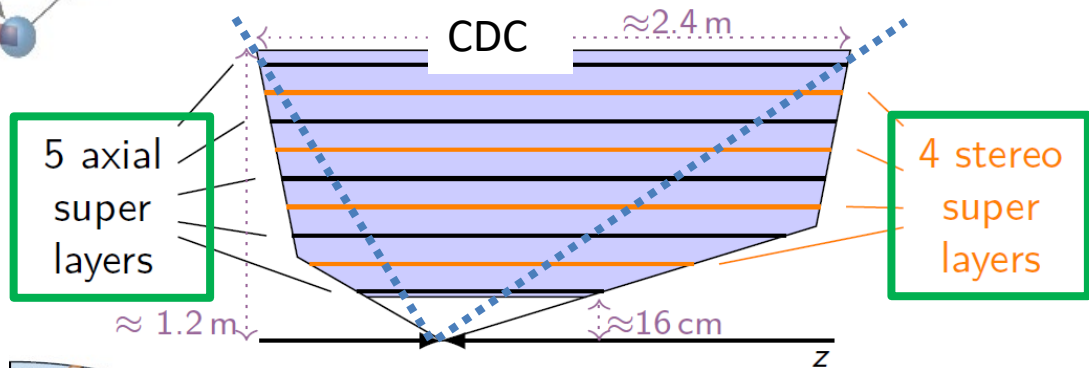
Target: tracks = well-known geometrical objects

patterns in known B-field:
- helices in space
- circles in transverse plane

3 preprocessed inputs per TS in each of the 9 SLs:
- crossing angle $\alpha$ (calculation)
- signed dirfttime (LUT)
- stereo wires selected from predef. range $\Delta\phi$ (LUT)

limited by exec time (300 ns on Virtex 6)

81

27

$\alpha$

$\pm t_{drift}$

$\Delta\varphi$

$9\times$

SL 0

SL 1 – 7

SL 8

$\theta$

$z$

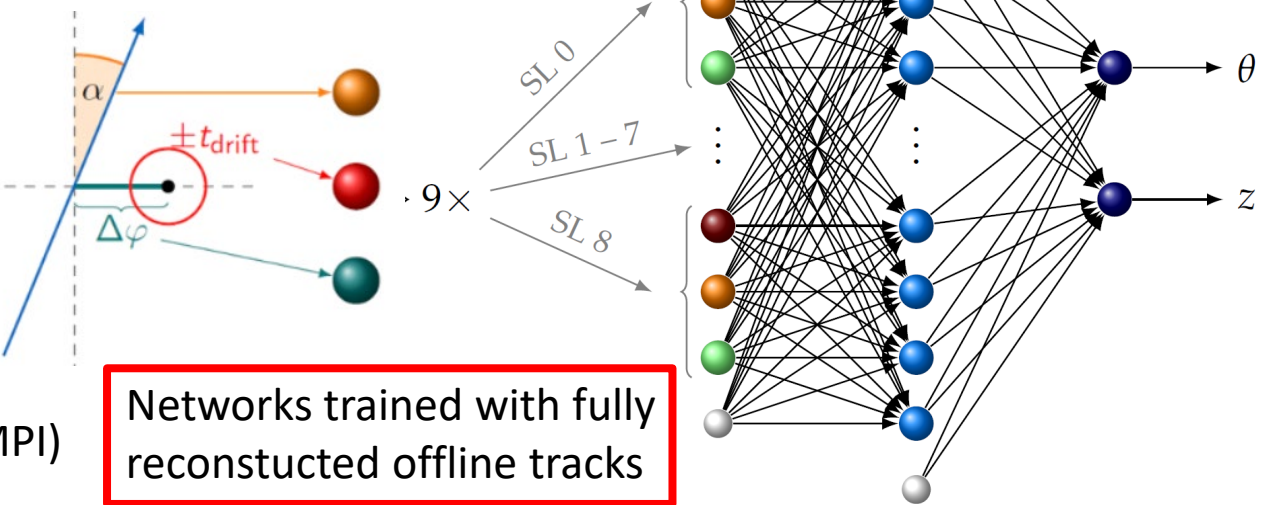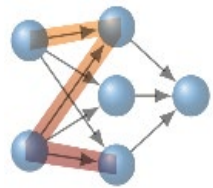„Natural input": 2D track candidates in each of 4 quadrants from Hough transforms (-> azimuth $\varphi$ and $1/R = 1/p_T$)
- calculate crossing angle $\alpha$ through TS
- determine „sign" of drifttime (from wire pattern in TS)

Felix Meggendorfer (TUM/MPI)

Networks trained with fully reconstucted offline tracks

*Belle II*

**recoZ0**

| | |
|---|---|
| Entries | 2824647 |
| Mean | 16.13 |
| Std Dev | 33.2 |

Exp. 16, L ~ 0.7 x $10^{34}$

large bg also in the „IP" peak

(vacuum worse in small diam. beam pipe)

Offline reconstructed tracks

Crotch of bp

QCS tip

**nnhwZ0**

| | |
|---|---|
| Entries | 2223686 |
| Mean | 23.98 |
| Std Dev | 32.27 |

nice match of z-structure seen in neural tracks
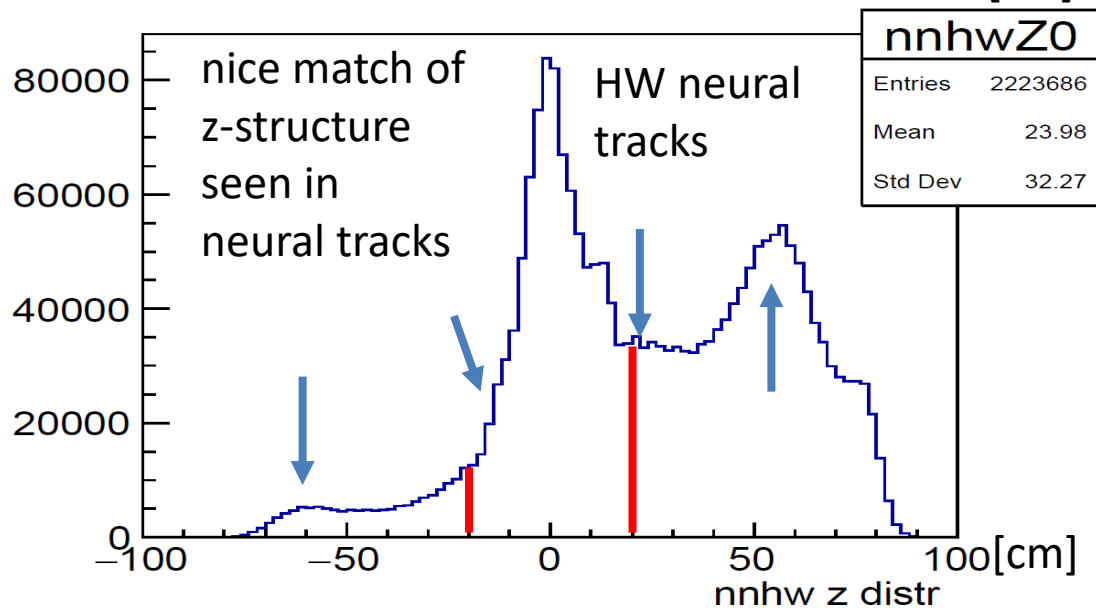
HW neural tracks

nnhw z distr
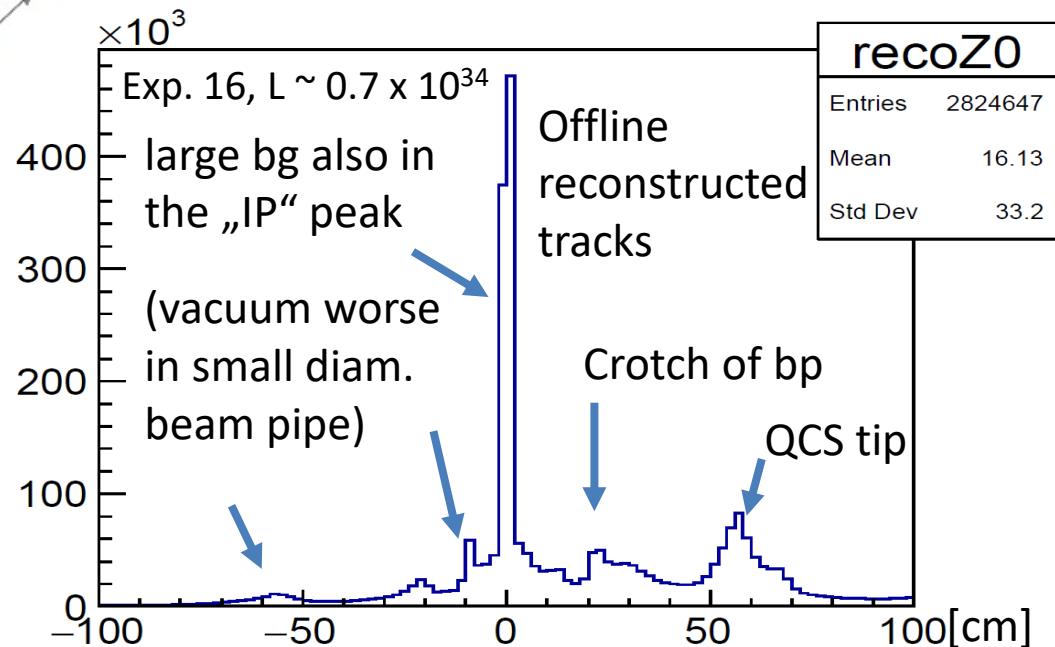
Fall 2020 running

Reco tracks:
z-distribution after full off-line reconstruction, including VXD space points

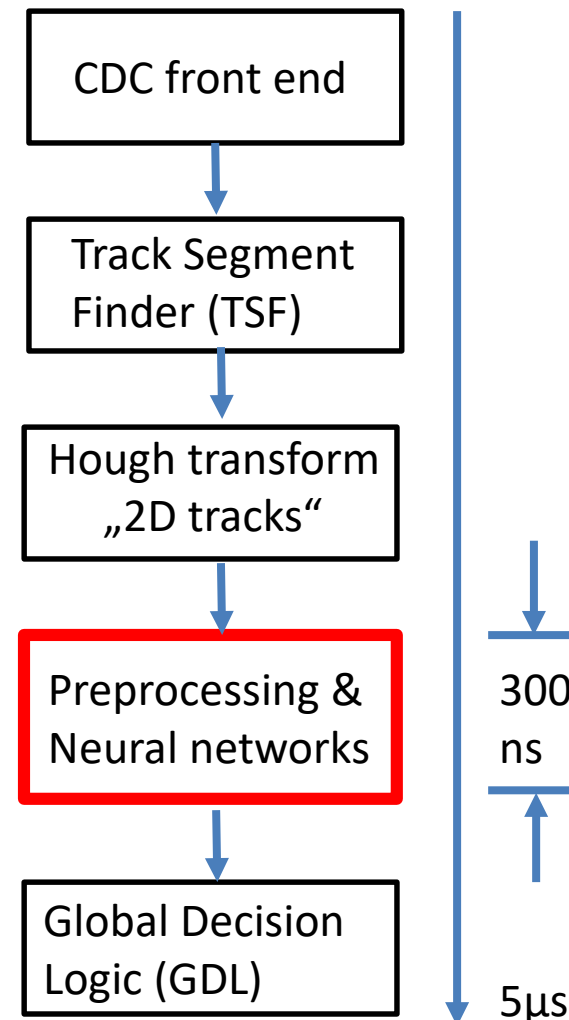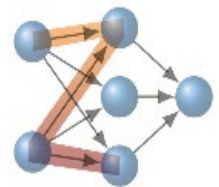Networks trained with real data from May-June 2020

The „Expert Networks":

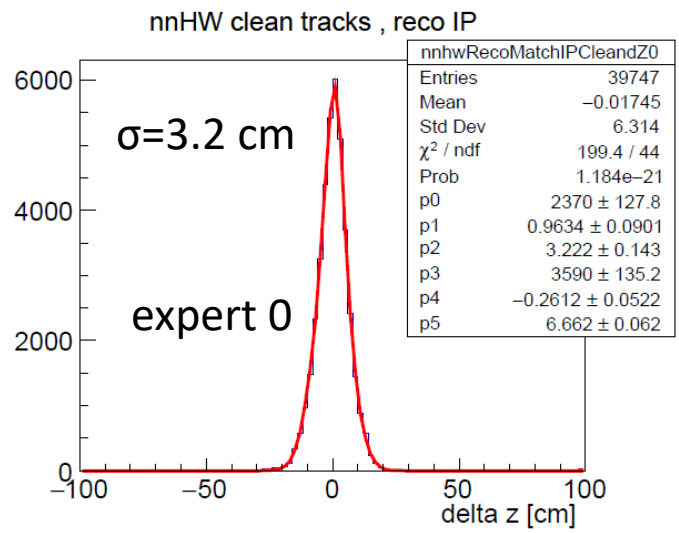5 different networks trained, depending on the number of available stereo TS

Expert 0: all 4 stereo TS
Expert 1-4: one of the stereo TS missing

L1 Pipeline

CDC front end

↓

Track Segment Finder (TSF)

↓

Hough transform „2D tracks"

↓

Preprocessing & Neural networks

↓

Global Decision Logic (GDL)

300 ns

5µs

nnHW tracks , reco IP

| nnhwRecoMatchIPdZ0 | |
| --- | --- |
| Entries | 82033 |
| Mean | −0.1881 |
| Std Dev | 7.127 |
| $\chi^2$ / ndf | 620.5 / 59 |
| Prob | 0 |
| p0 | 9850 ± 93.4 |
| p1 | 0.2917 ± 0.0341 |
| p2 | 5.645 ± 0.057 |
| p3 | 816.4 ± 111.0 |
| p4 | −2.557 ± 0.240 |
| p5 | 11.47 ± 0.38 |

$\sigma$=5.6 cm

all experts

nnHW clean tracks , reco IP

| nnhwRecoMatchIPCleandZ0 | |
| --- | --- |
| Entries | 39747 |
| Mean | −0.01745 |
| Std Dev | 6.314 |
| $\chi^2$ / ndf | 199.4 / 44 |
| Prob | 1.184e−21 |
| p0 | 2370 ± 127.8 |
| p1 | 0.9634 ± 0.0901 |
| p2 | 3.222 ± 0.143 |
| p3 | 3590 ± 135.2 |
| p4 | −0.2612 ± 0.0522 |
| p5 | 6.662 ± 0.062 |

$\sigma$=3.2 cm

expert 0

nnHW Err2 tracks , reco IP

| nnhwRecoMatchIPErr2dZ0 | |
| --- | --- |
| Entries | 42286 |
| Mean | −0.3485 |
| Std Dev | 7.811 |
| $\chi^2$ / ndf | 324 / 56 |
| Prob | 2.202e−39 |
| p0 | 4728 ± 53.8 |
| p1 | 0.3355 ± 0.0477 |
| p2 | 5.912 ± 0.073 |
| p3 | 449.9 ± 61.1 |
| p4 | −3.184 ± 0.314 |
| p5 | 12.29 ± 0.43 |

$\sigma$=5.9 cm

experts 1-4

Belle II Track Trigger 2021 running

Large 2D trigger rate in 2021 -> „y" bit: ≥ 1 track,|z| < 20 [cm], require ≥ 2 (2D) tracks && y=1
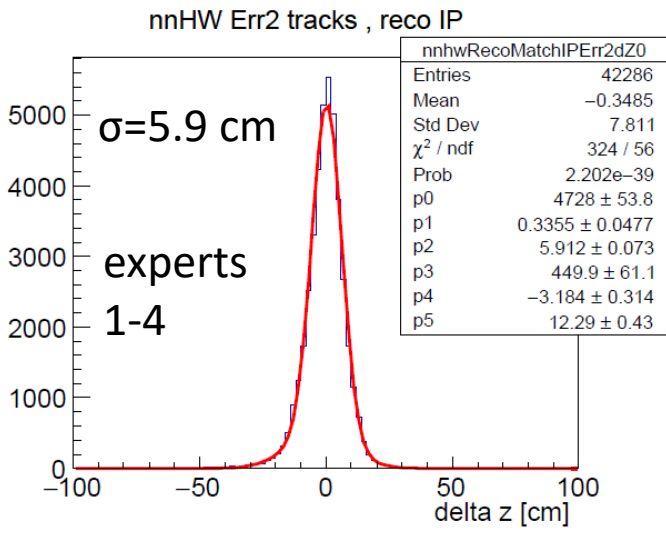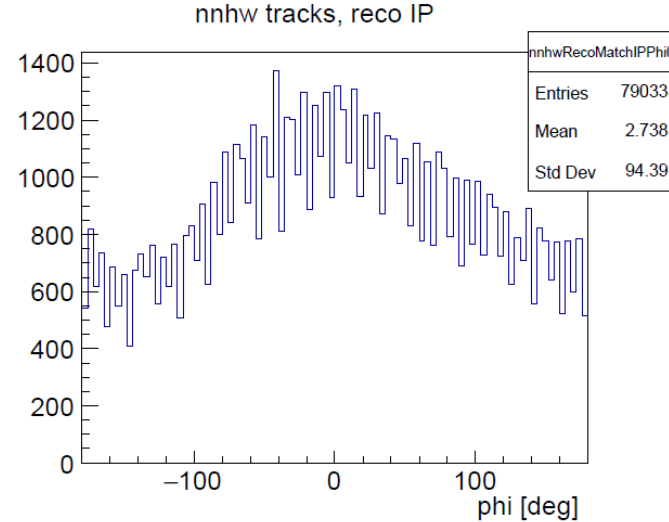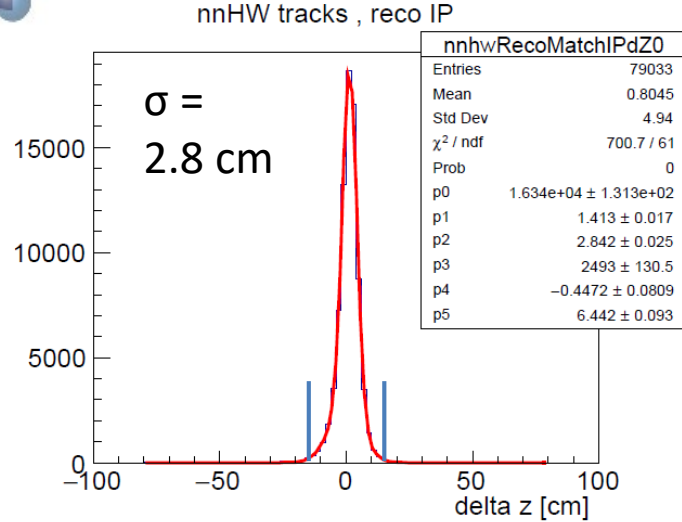
Instantaneous lumi = (3.8 x $10^{34}$) end of 2021, background rising with luminosity in 2021

NN resolution of IP tracks very stable, proving robustness of the neural network technique against changing conditons

**Fundamental change at Belle II wrt Track Triggers: due to overwhelming BG, all 2-Track Triggers require at least one neural track: „y=1"**
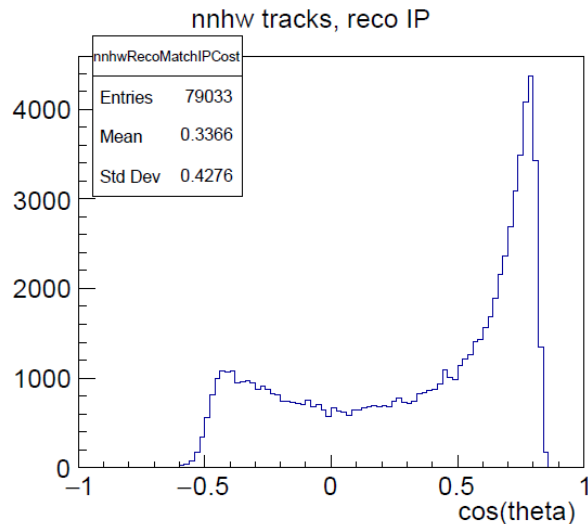
**Belle II**



nnHW tracks , reco IP

| nnhwRecoMatchIPdZ0 | |
|---|---|
| Entries | 79033 |
| Mean | 0.8045 |
| Std Dev | 4.94 |
| $\chi^2$ / ndf | 700.7 / 61 |
| Prob | 0 |
| p0 | $1.634e+04 \pm 1.313e+02$ |
| p1 | $1.413 \pm 0.017$ |
| p2 | $2.842 \pm 0.025$ |
| p3 | $2493 \pm 130.5$ |
| p4 | $-0.4472 \pm 0.0809$ |
| p5 | $6.442 \pm 0.093$ |

$\sigma$ = 2.8 cm



nnhw tracks, reco IP

| nnhwRecoMatchIPPhi0 | |
|---|---|
| Entries | 79033 |
| Mean | 2.738 |
| Std Dev | 94.39 |



nnhw tracks, reco IP

| nnhwRecoMatchIPCost | |
|---|---|
| Entries | 79033 |
| Mean | 0.3366 |
| Std Dev | 0.4276 |

Exp 24, runs 790 - 890, New FW
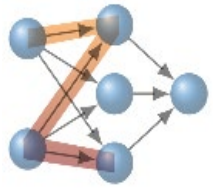
Retrainig of neural networks with data from the end of 2021 (high background data)

Use modern training library **PyTorch** (previously used FANN, integrated into Belle II software library)

Results from improved training

Gaussian fits to neuro tracks associated with reco tracks from IP ($|z|<1$ cm, d < 1.5cm)

Central Gauss: $\sigma$ = 2.8 cm

2nd Gauss: $\sigma$ = 6.4 cm (13.2 %)

**2020 training:**
**central Gauss    $\sigma$ =   5.6 cm**
**2nd Gauss        $\sigma$ = 11.5 cm**

**factor 2 improvement !!**
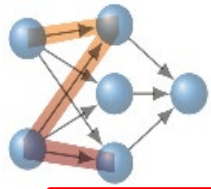
Can we launch a track trigger requiring only one track?

Sources of Background:

Collisions of electrons/positrons with elements of the beam guide system, mostly producing protons from nuclear spallation

(momentum of particles outside of IP mostly below 1 GeV !!)
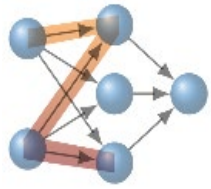
from IP (!!): QED events

### Distribution of Reco Track momentum vs z



recoPVsZ0
Entries 738153

Annihilation events

„beam-wall" events

Irreducible IP bg from
$e^+e^- \rightarrow e^+e^-e^+e^-$
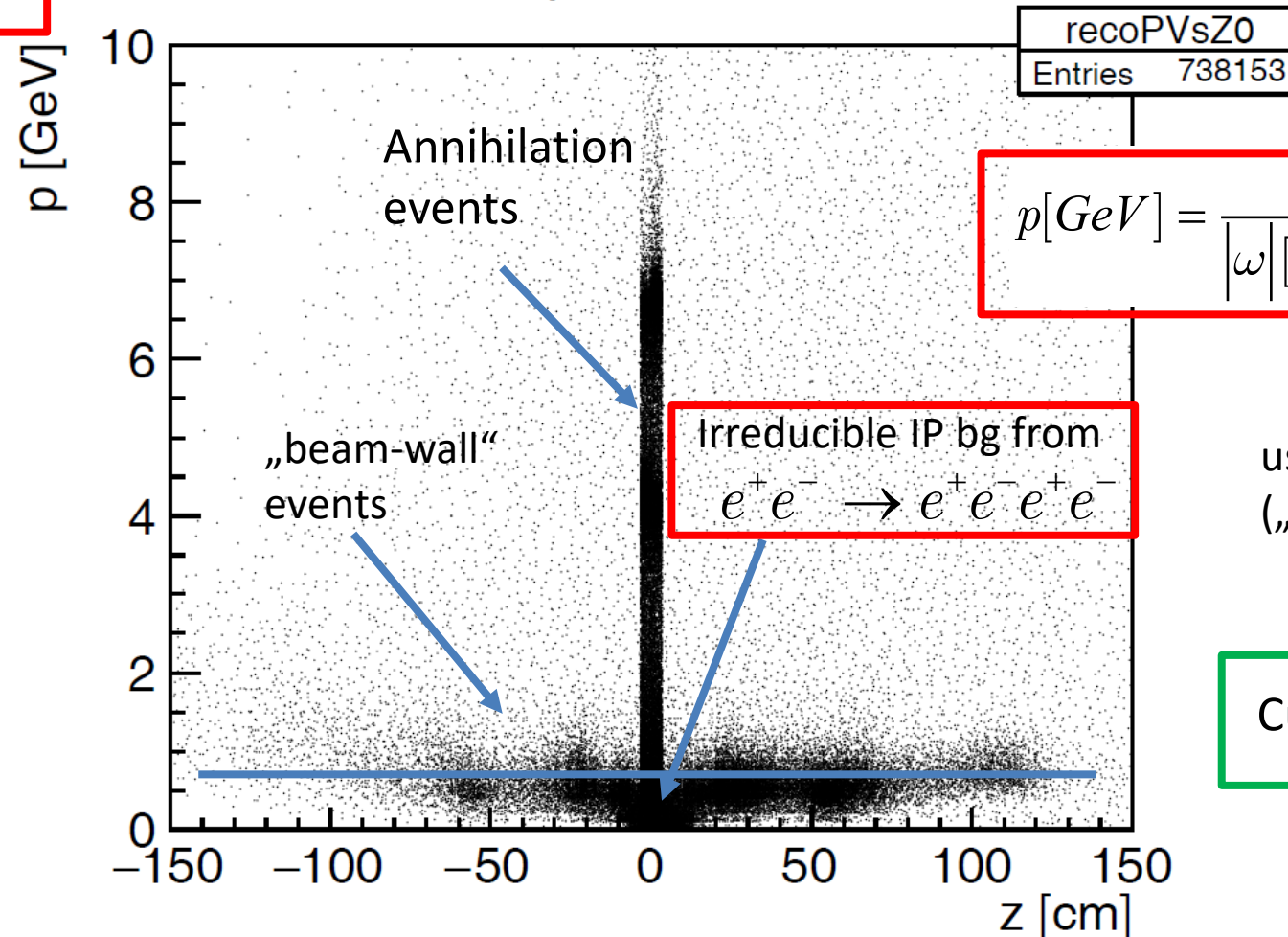
$$p[GeV] = \frac{1}{|\omega|[1/m]\sin(\theta)} 0.3 B[T]$$

use the second output ("θ") of the networks

Cut on p > 0.7 GeV

## reco mom vs HW mom, STT cut

momentum
correlation of
neuro tracks
and reco tracks

| recoPvsnnhwPMomCutOnlySTTcut | |
|---|---|
| Entries | 16076 |
| Mean x | 0.5593 |
| Mean y | 0.5716 |
| Std Dev x | 0.1781 |
| Std Dev y | 0.1902 |

Resolution in p  ~ 80 MeV

2D trigger limit
in pt

**Neuro track momentum [GeV]** (y-axis)

**Reco track momentum [GeV]** (x-axis)

calculate the
neuro track
momentum:

$$p[GeV] = \frac{1}{|\omega|[1/m]\sin(\theta)} 0.3 B[T]$$
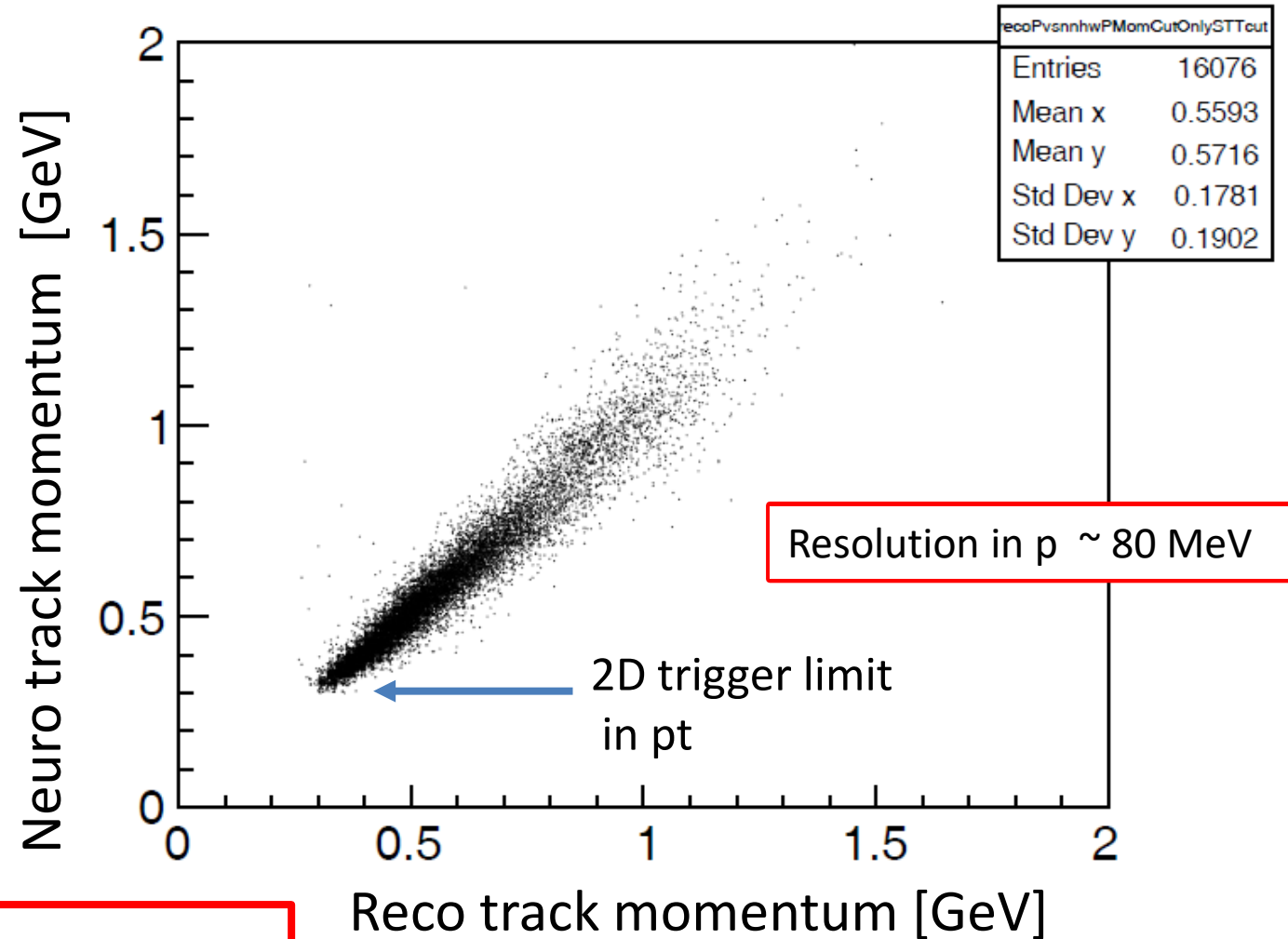
ω ~ transverse momentum (from 2D track)
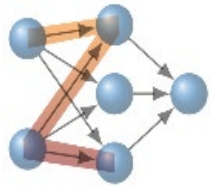
# STT Triggers ONLY
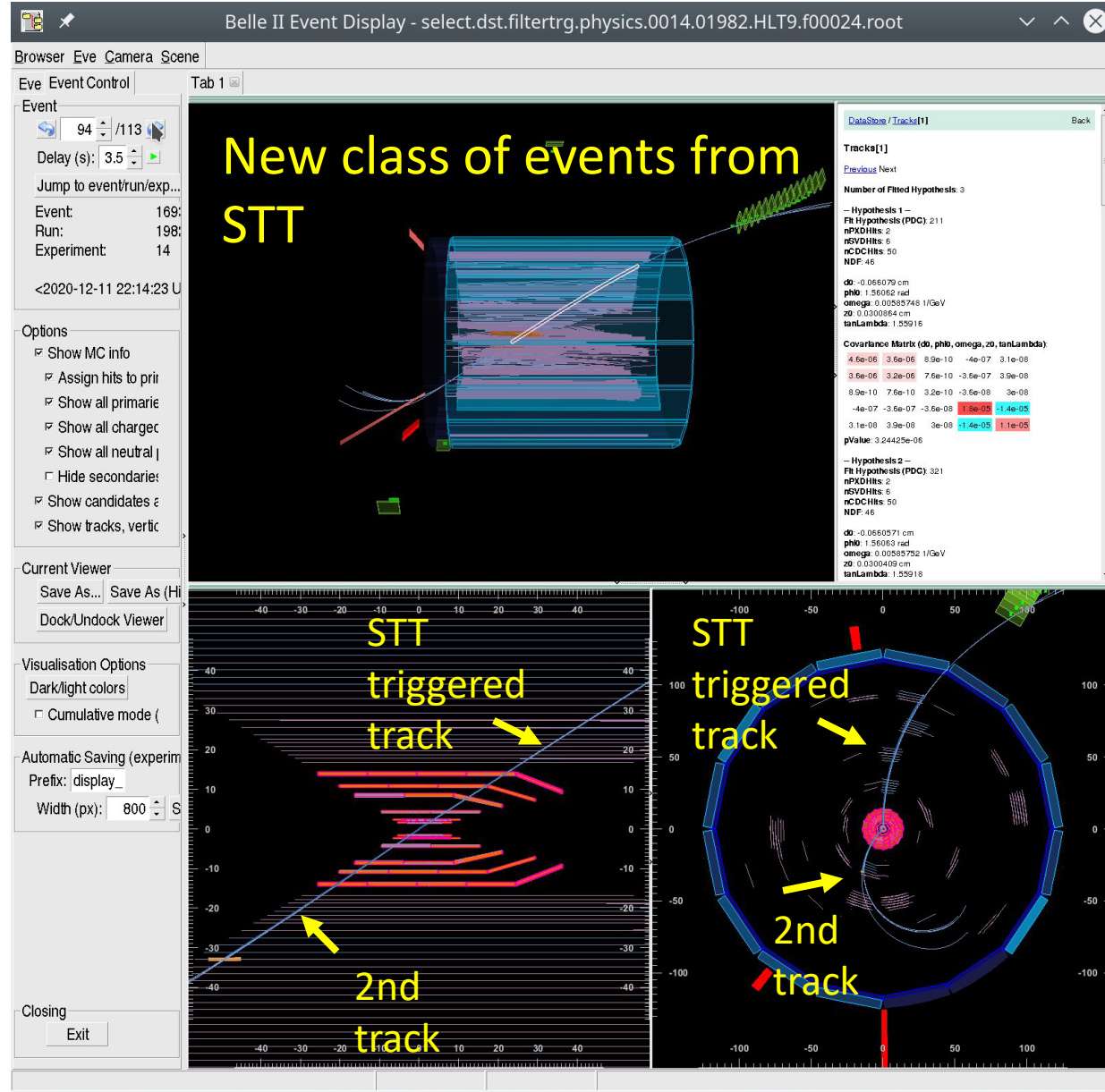


Event display shows the reco tracks

2nd track at shallow Ѳ cannot be seen by CDC trigger

Note:
2nd track is unbiased
(can be anywhere in the detector, usually reconstructed by silicon tracker)

Event class only triggered by STT (~12% of STT events)

*Belle II*
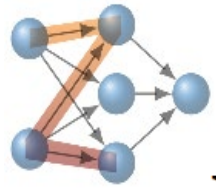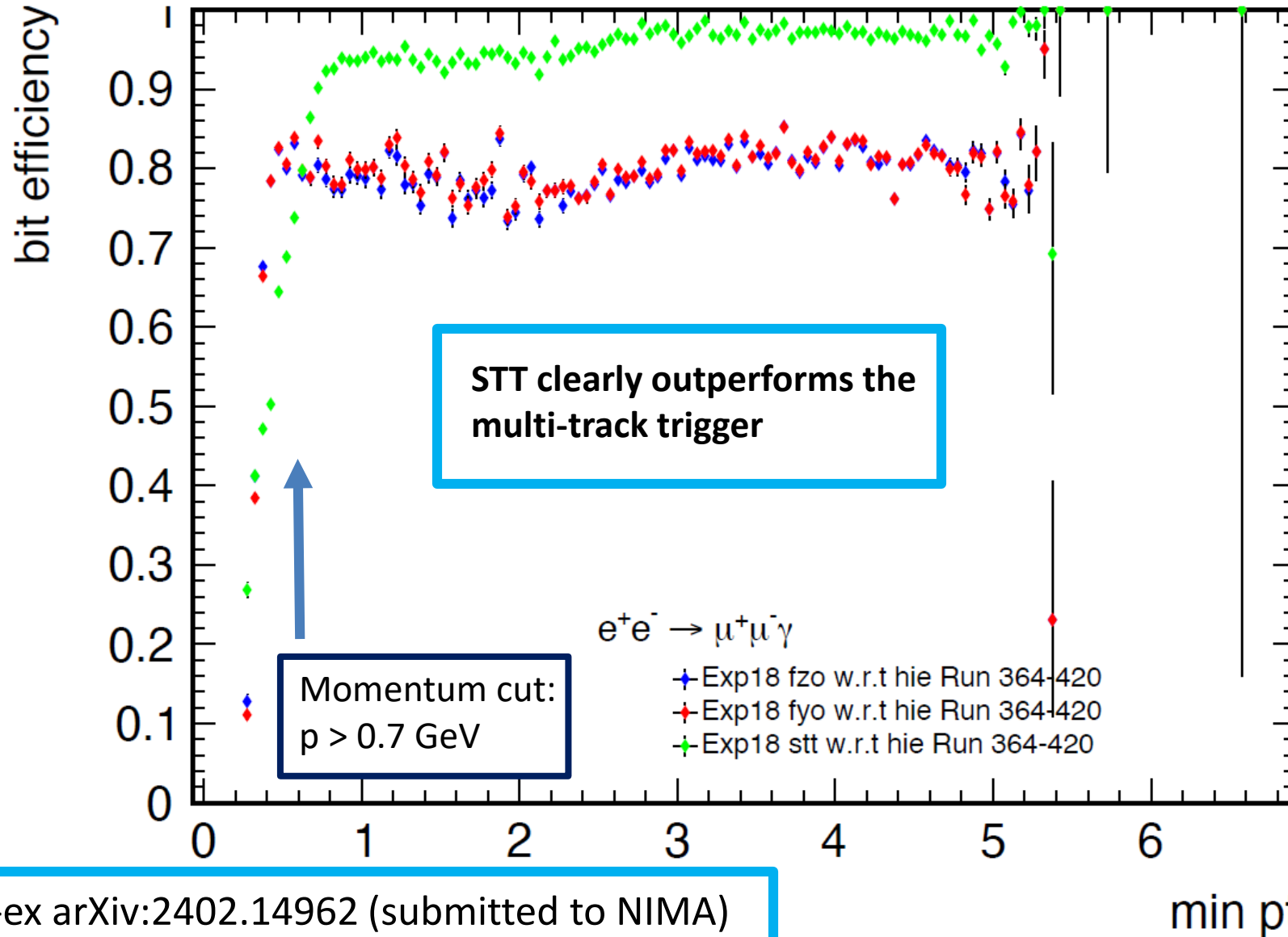


Trigger rate of the STT
~ 20% -25% of total rate budget

-> acceptable

First minimum bias track
trigger in HEP

BUT: some problems during
summer 2022 running:
rate rising to ~50% of total
budget -> ??

**STT clearly outperforms the
multi-track trigger**

Momentum cut:
p > 0.7 GeV

$e^+e^- \to \mu^+\mu^-\gamma$

- Exp18 fzo w.r.t hie Run 364-420
- Exp18 fyo w.r.t hie Run 364-420
- Exp18 stt w.r.t hie Run 364-420

hep-ex arXiv:2402.14962 (submitted to NIMA)

Increase of machine-induced background
with rising luminosity

-> increase of STT trigger rate
observed
(but efficiency stable !!)

Excess rate may saturate DAQ
and increase deadtime

## z0 reco vs z0 nnhw

| nnhwZ0vsRecoZ0 | |
| --- | --- |
| Entries | 2563960 |
| Mean x | 37.98 |
| Mean y | 26.63 |
| Std Dev x | 45.28 |
| Std Dev y | 32.66 |

these NN tracks are
problematic
(„feed down")
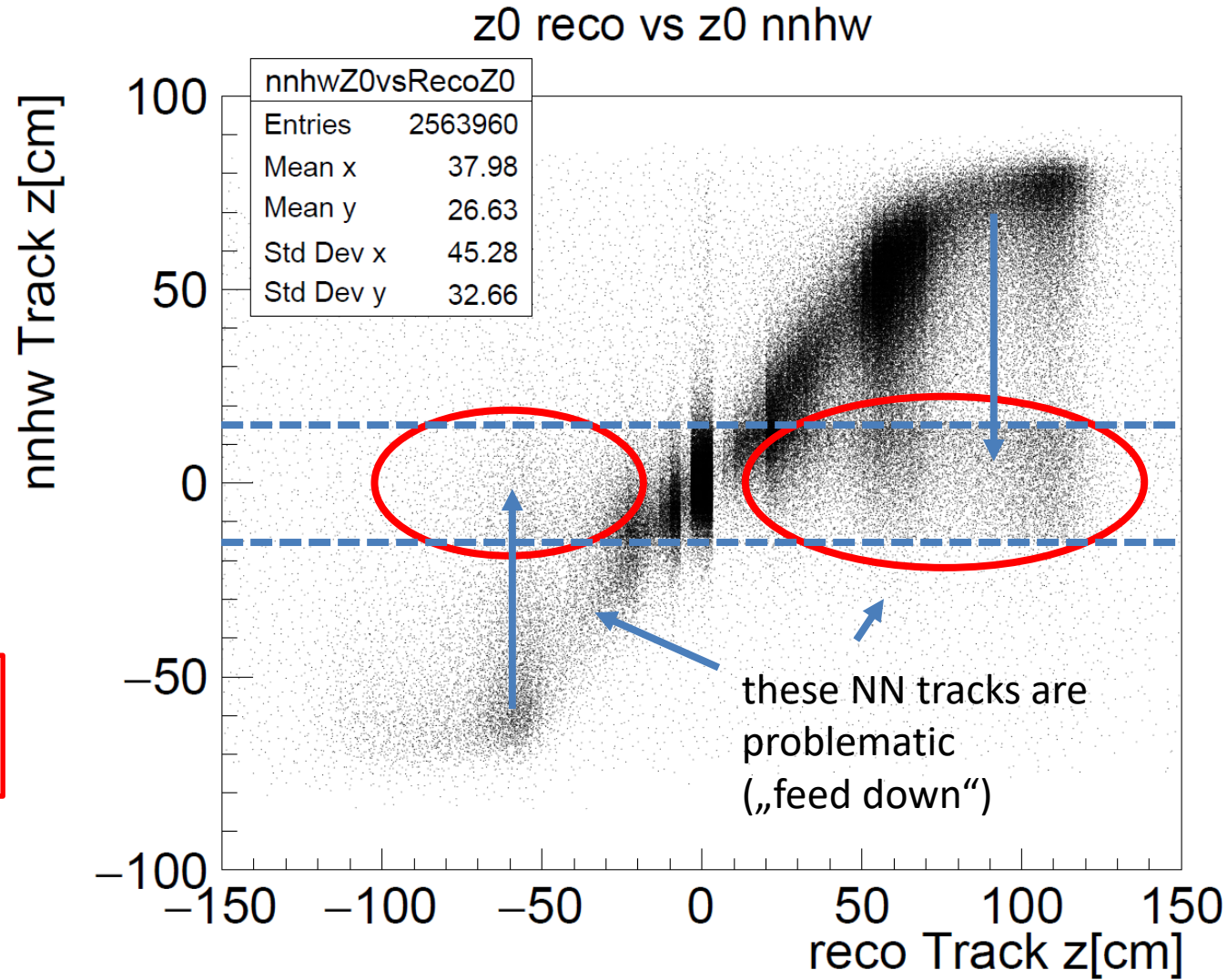
Band at |z|<15cm:
acceptance for a
valid neural track

Large |z|:
a certain fraction of
tracks shifted into
IP region
-> increase of rate

Why are tracks
predicted around
IP while coming
from large |z| ?

feed-down especially strong for expert 4 (inner stereo SL missing)

Belle II

3D track model

-> keep efficiency of STT & low rate budget with rising luminosity (BG)

**Physics goals**: low charged multiplicity, e.g. $\tau$ 1-prong decays ( - > $\tau$ EDM, LFV),
- $e^+e^- \rightarrow \pi^+\pi^-(\gamma)$  for g-2 (hadronic vaccum polarization) etc.
- quite generally: determination of lepton ID, tracking efficiency for the „other track"
- **STT is a minimum bias track trigger**

**New FPGA Hardware now available:** „UT4 Board" with Virtex Ultrascale 160/190

**Improved track model for neural input  / training algorithms**:
- **track finding in 3D Hough space** -> this is **really new** (S. Skambraks, S. Hiesl)
- network architecture: „deep-learning" + additional inputs (T. Forsthofer)
- -> improve resolutions @ IP and for larger |z|
- -> reduce feed-down & fake tracks
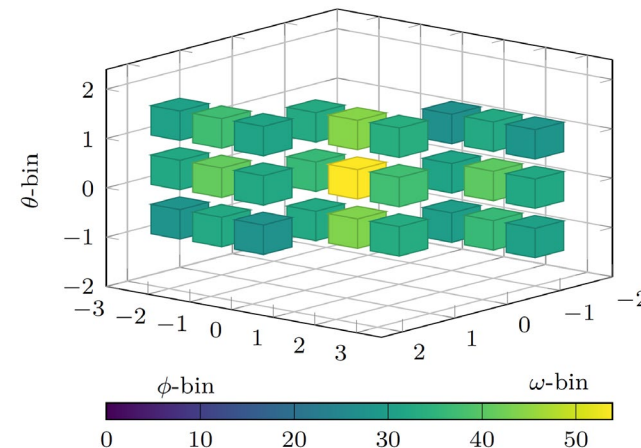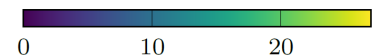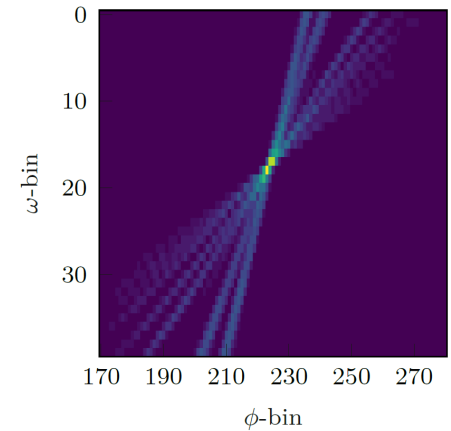
**FPGA Implementation** (Kai Unger):
- new algorithms on new UT4-Boards using hls4ml
- optimize latency: e.g. move STT decision to NN



(d) $\theta$-bin 3



(g) $\theta$-bin 6

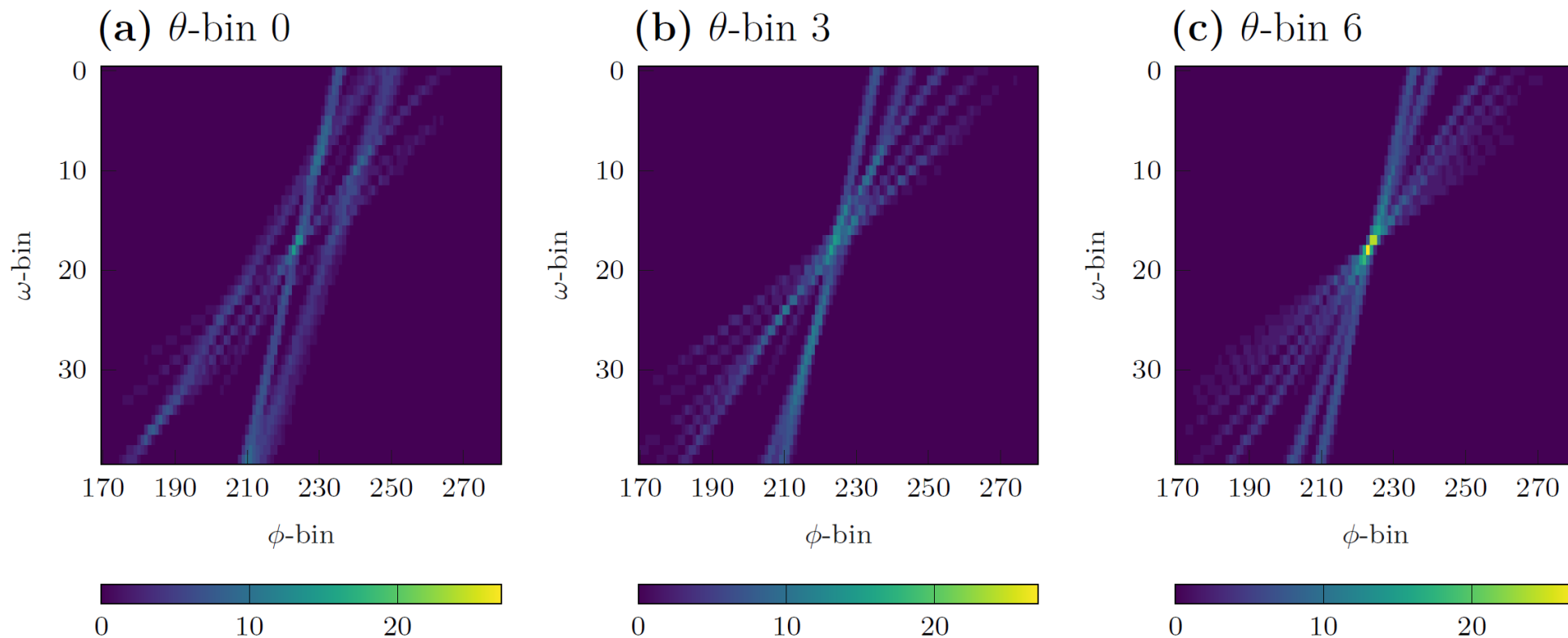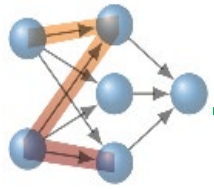New curve parameter: Polar angle $\theta \implies$ 3D-Hough space

- 9 bins in $\theta \in [19, 140]°$, 384 bins in $\phi \in [0, 360]°$, 40 bins in $\omega \propto q \cdot p_\mathrm{T}^{-1}$, $p_\mathrm{T} \in [0.25, 10]\,\mathrm{GeV}/c$

Vertex assumption: The track originates from $(x, y, z) = (0, 0, 0)$ (IP)



**(a)** $\theta$-bin 0     **(b)** $\theta$-bin 3     **(c)** $\theta$-bin 6

$\implies$ Intersection point yields $\omega$, $\phi$ and $\theta$
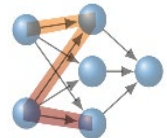
Simon Hiesl (MPI & LMU)

# Hough Clustering in 3 Dimensions

Original algorithm: DBSCAN $\rightarrow$ Difficult to implement on an FPGA (non-deterministic length $\implies$ latency not fixed)

Update: **Fixed Clustering**

Three steps, repeated `iterations` times:

- Step 1: Global maximum search on Hough space

- Step 2: A fixed shape is put around the maximum

  ▶ The weights in this shape are added up (total weight)

  ▶ If total weight $\geq$ `mintotalweight` and peak weight $\geq$ `minpeakweight` the cluster is saved

  ▶ All hits (TS) are extracted and have to pass two TS cuts

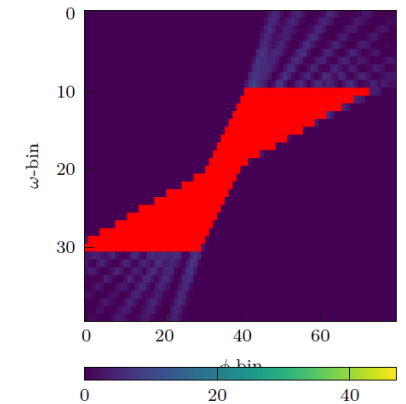- Step 3: Cells around the global maximum are set to zero ("Butterfly-Shape" cutout)
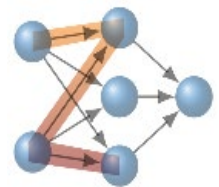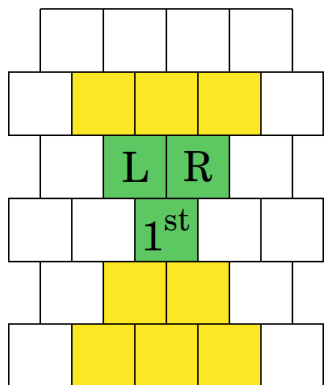
Fixed shape:
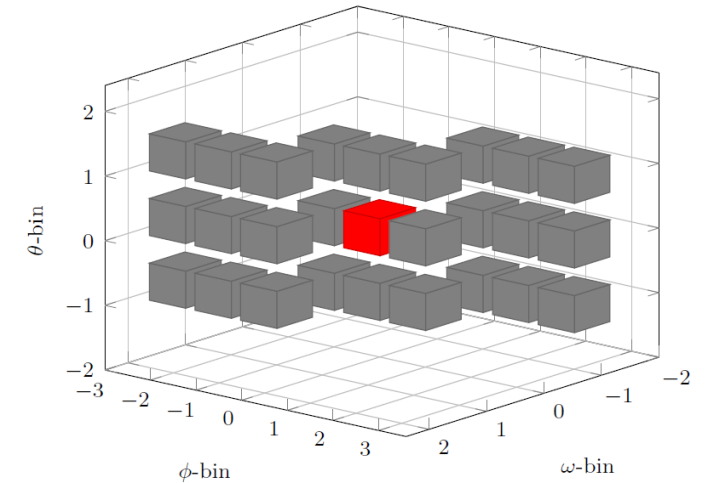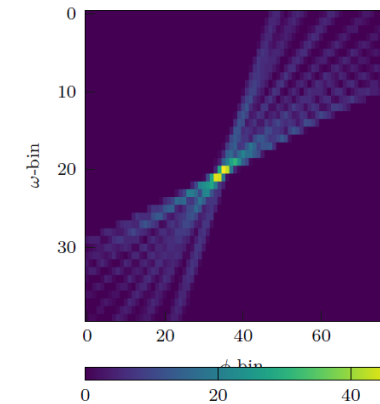


(a) Complete Cluster    (b) Cutout

Simon Hiesl (MPI & LMU)

superlayer 1 – 8



**Extended inputs** to network
- „standard" priority wire (Δφ, α, DT)
- plus entire wire pattern in TS
  (10 additional binary inputs)
- with condition: ADC count > min
(remove bg from electronic
 cross-talk & synchr. photons)

electronic cross-talk &
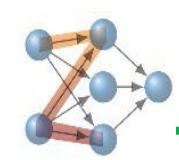synchrotron photons



Num_bg: 2188402
Num_sig: 5394082
Cut_bg: 424537
Cut_sig: 219626
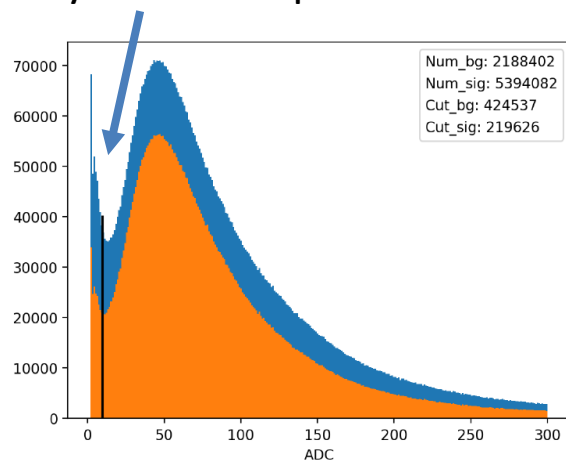
**Multi-hidden**-layer network
(„deep learning")

- Several architectures
  investigated
- „optimal" configuration with
  4 hidden layers and fewer (!)
  nodes
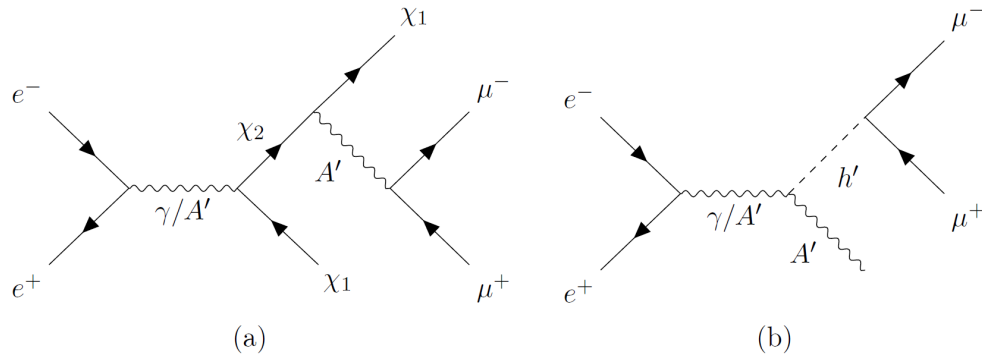  when using 3D track model
  (easier to implement in HW)
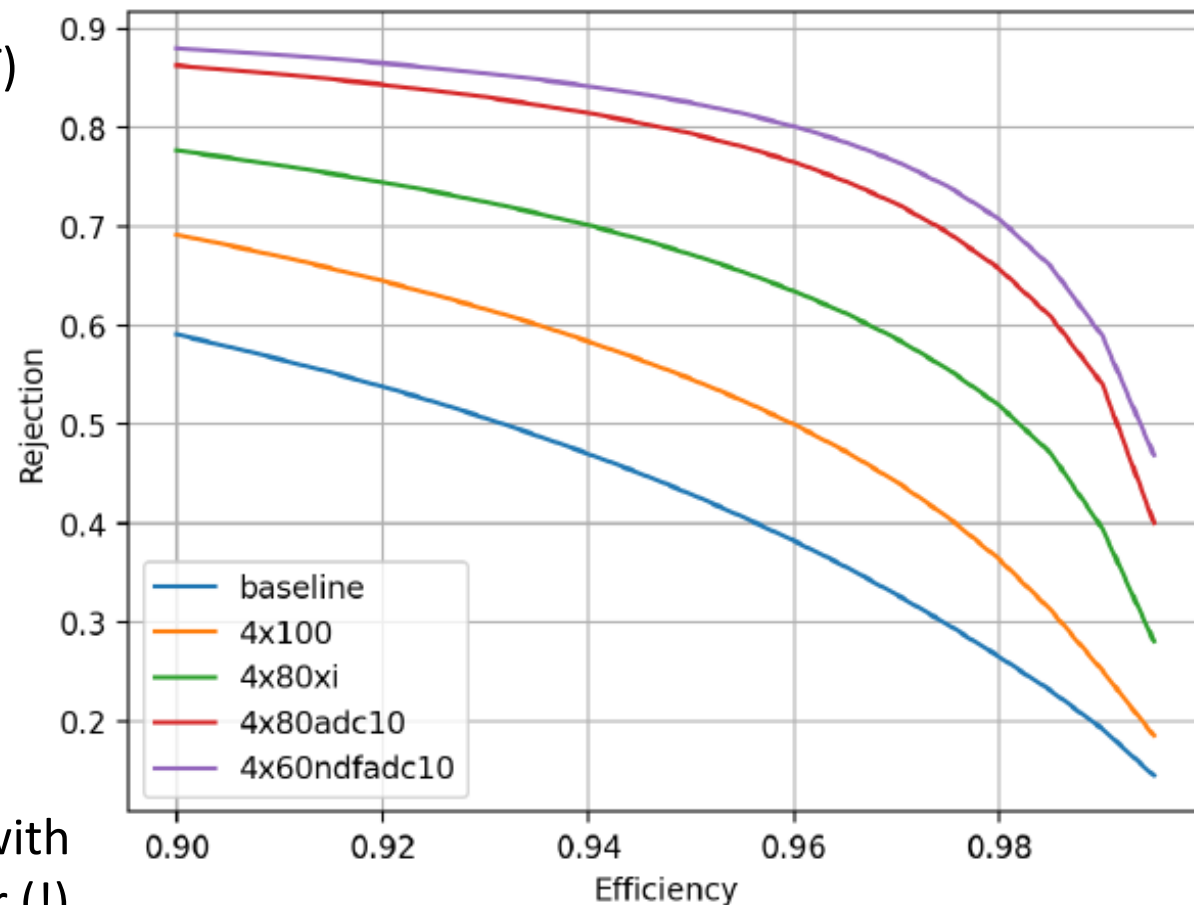


Timo Forsthofer
(MPP & LMU)

Example: Inelastic Dark Matter production
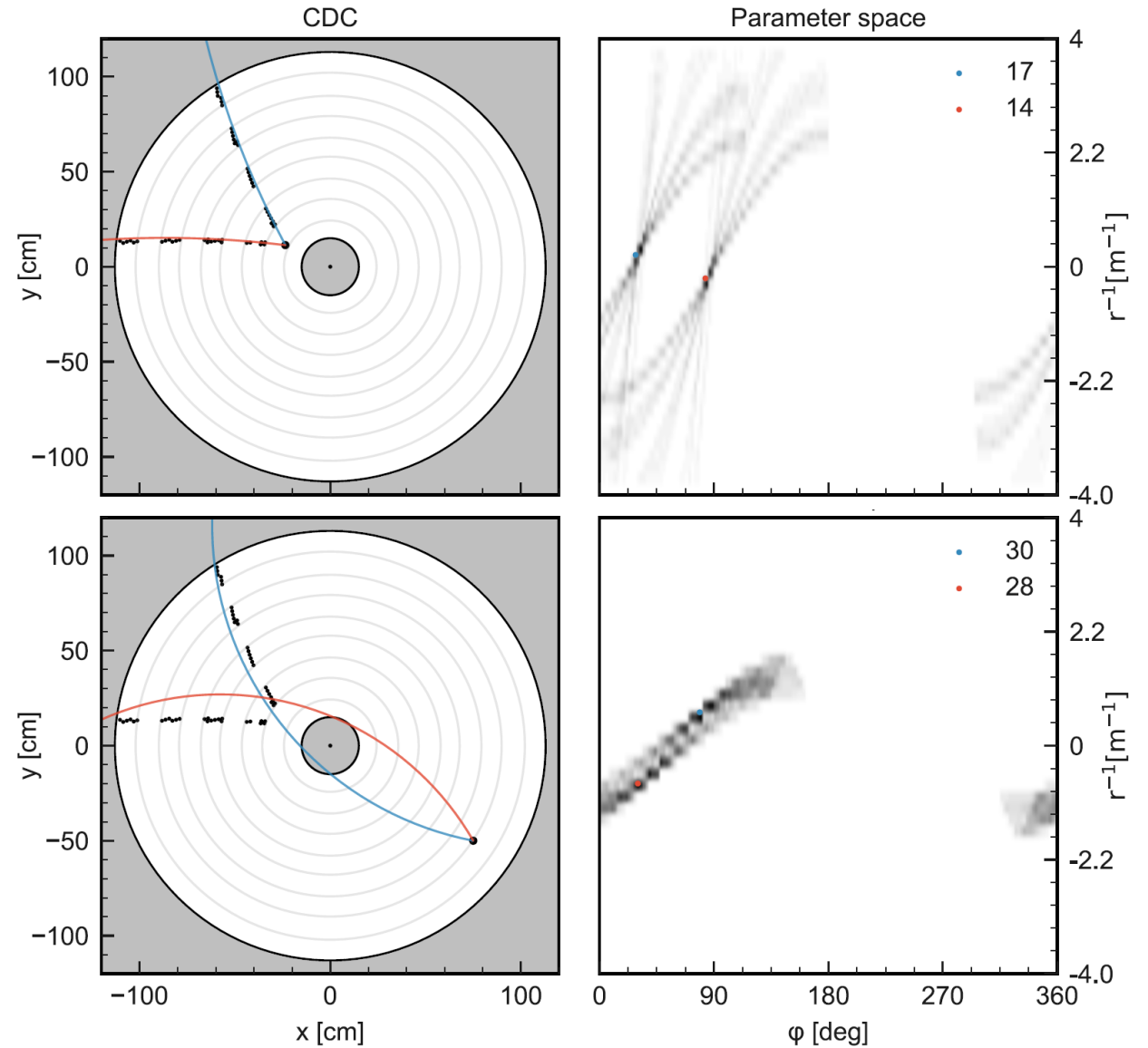DM particles expected to be quite long-lived
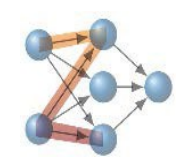


Basic idea:

Divide the CDC axial wire planes into a set of „Macro Cells", serving as origins for the Hough transforms

FPGA:

-> execute all Hough transforms with origins in each of the Mcro cells in parallel (typically of O(100))
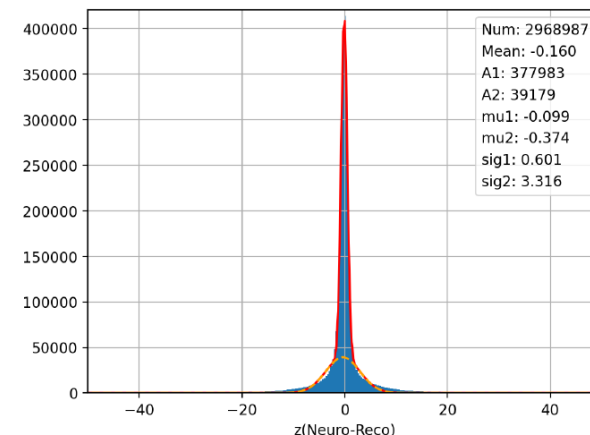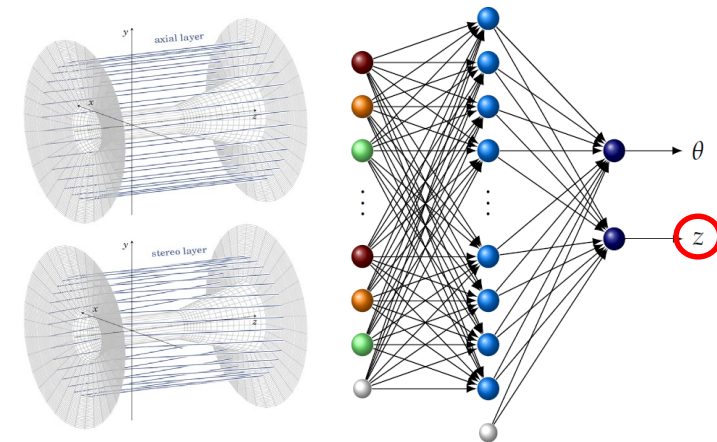
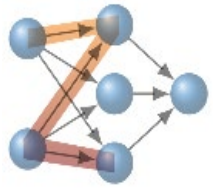-> use neural networks to determine „correct" vertex
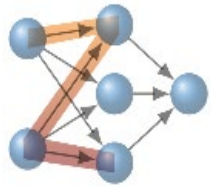
# Summary and Conclusions

First Level-1 Neural Network Track Trigger in HEP, operational @Belle II since Jan. 2021

- At least one Neural Track required to assert a track trigger (using CDC)
        -> Neural Nets: „working horses" for Belle II track trigger system
- Novel feature: **Minimum Bias Single Track Trigger (STT)**
  - excellent performance even under severe background conditions,
  - however, "Feed-down" and "Fakes" need attention with rising luminosity

- Upgrade: More powerful FPGA boards now available (Virtex UltraScale 7 XCVU160)
  - track finding via optimized 3D Hough cluster algorithm (novel method!)
  - additional inputs from all wires within the TSs (117 inputs total)
    (coarse analog thresholds for CDC wire signals to suppress background)
  - deep-learning neural network architectures (4 x 60 hidden nodes)
  - superior performance both in resolution and background suppression

- Commissioning by summer 2024, launch planned for the fall 2024 data taking

- Neural **Displaced Vertex Trigger** on the horizon, aiming at long-lived new particles, commissioning planned end of 2024
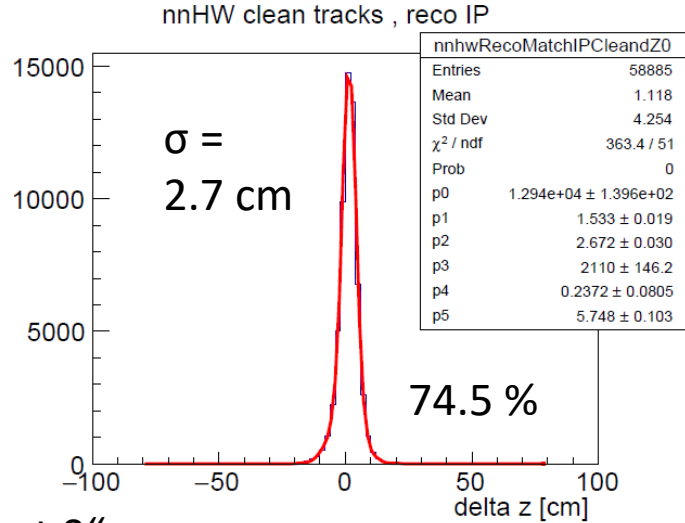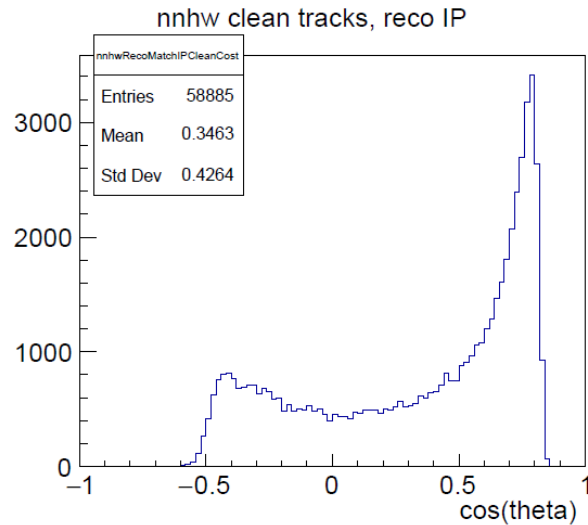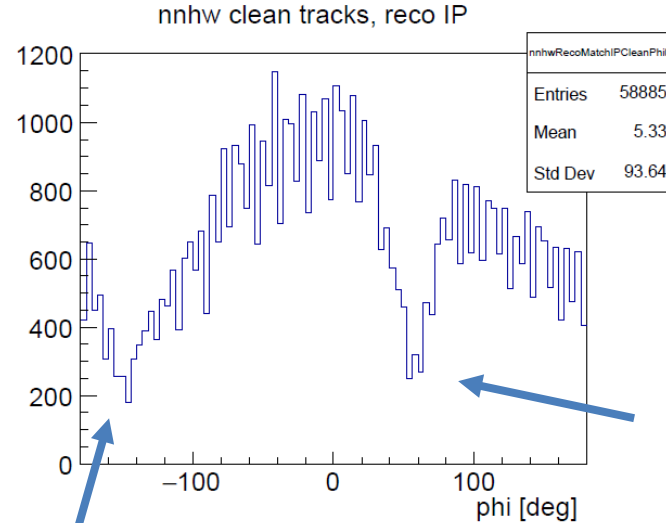
# BACKUP
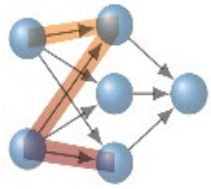
nnHW clean tracks , reco IP

| nnhwRecoMatchIPCleandZ0 | |
|---|---|
| Entries | 58885 |
| Mean | 1.118 |
| Std Dev | 4.254 |
| $\chi^2$ / ndf | 363.4 / 51 |
| Prob | 0 |
| p0 | 1.294e+04 ± 1.396e+02 |
| p1 | 1.533 ± 0.019 |
| p2 | 2.672 ± 0.030 |
| p3 | 2110 ± 146.2 |
| p4 | 0.2372 ± 0.0805 |
| p5 | 5.748 ± 0.103 |

$\sigma$ = 2.7 cm

74.5 %

„expert 0"

nnhw clean tracks, reco IP

| nnhwRecoMatchIPCleanPhi0 | |
|---|---|
| Entries | 58885 |
| Mean | 5.33 |
| Std Dev | 93.64 |

known problem

new CDC inefficiency

nnhw clean tracks, reco IP

| nnhwRecoMatchIPCleanCost | |
|---|---|
| Entries | 58885 |
| Mean | 0.3463 |
| Std Dev | 0.4264 |

Gaussian fits to neuro tracks associated with reco tracks from IP
(|z|<1 cm, d < 1.5cm)

Central Gauss: $\sigma$ = 2.7 cm

2nd Gauss: $\sigma$ = 5.7 cm (14.1 %)
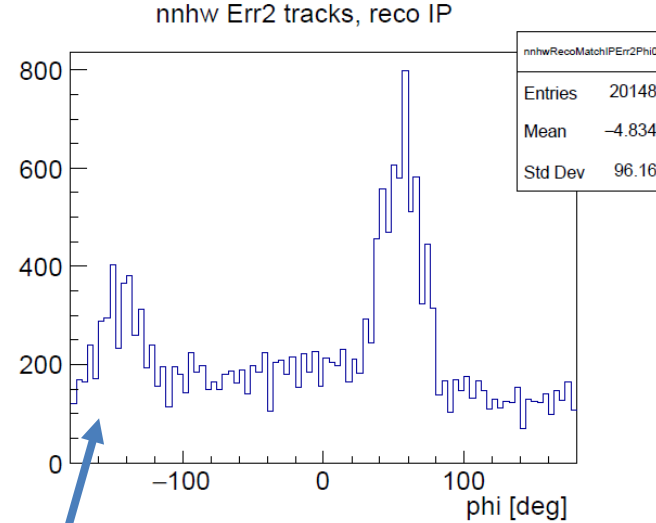
Exp 24, runs 790 - 890, New FW          STT active, zcutTrig = 20 cm          19/03/2022  50
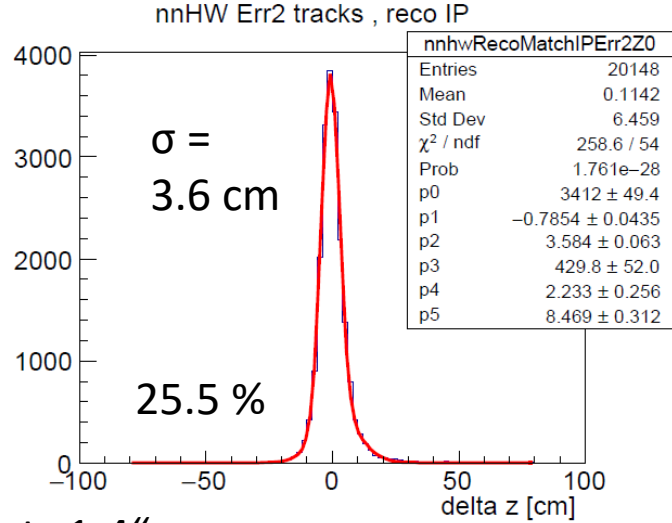
„experts 1-4"

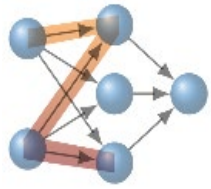new CDC inefficiency
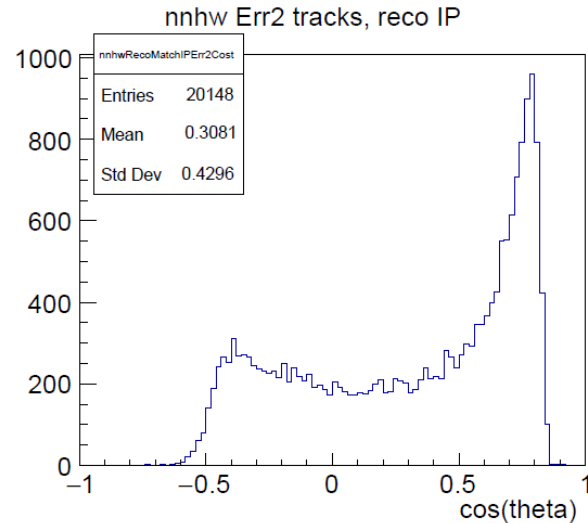
Gaussian fits to neuro tracks associated
with reco tracks from IP
($|z|$ <1 cm, d < 1.5cm)

Central Gauss: $\sigma$ = 3.6 cm

2nd Gauss:        $\sigma$ = 8.5 cm (11.2 %)

New class of events from STT

STT triggered track

2nd track

STT triggered track

2nd track

2nd track reconstructed only in PXD/SVD

Caution:

efficiency of STT not easy to calculate from data since no other orthogonal trigger (e.g. ECL) available

Hope for New Physics in low multiplicity final states ?
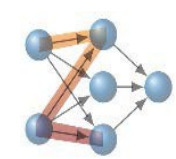
- Extend traditional 2D ($\omega = 1/p_T$, $\phi$=azimuth angle)
  Hough space by a third dimension, the (binned) polar angle $\theta$
- For track finding use axial and stereo track segments (->3D)
- Peak finding in 3D Hough space
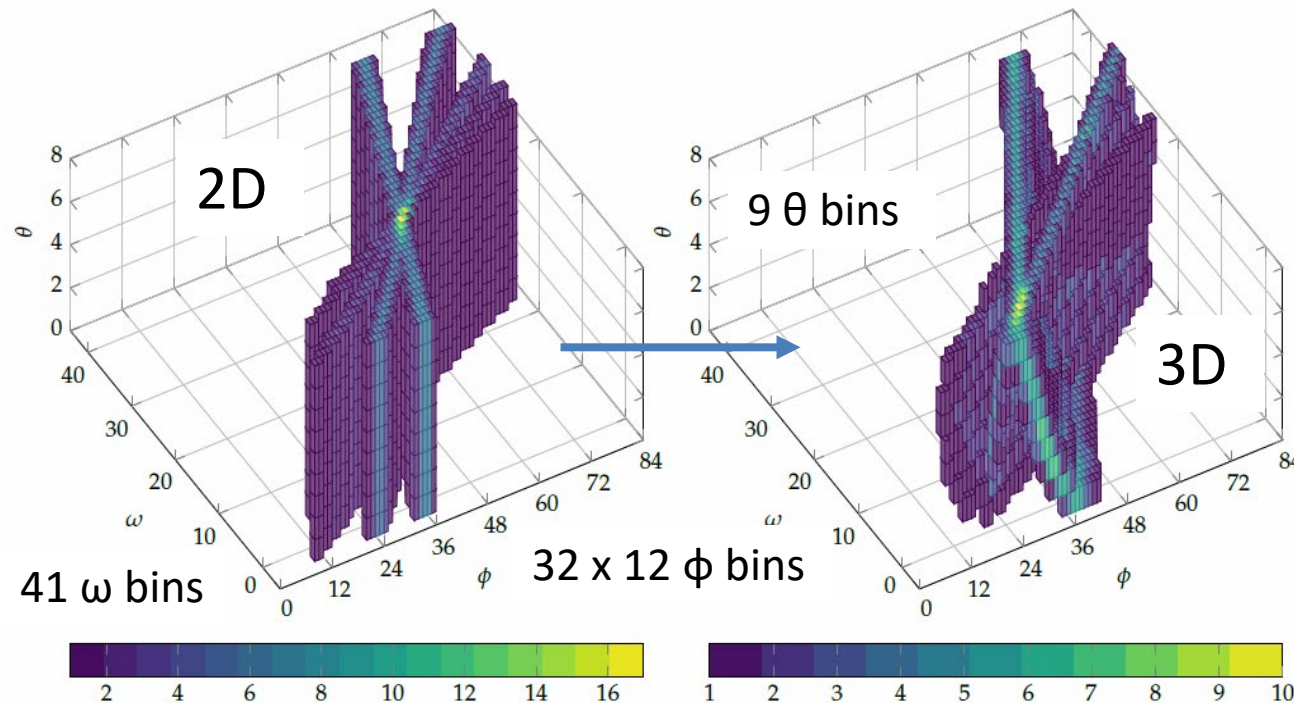
Main advantages:

- more TS (9 vs 5)
  -> suppress fakes
- No need to choose STS
  by min drift time
  -> find „correct" STS
- Force track model to
  originate from IP
  -> suppress candiates
      far from IP
- 3D track candidates
  come with $\theta$ estimate,
  -> improve z resolution



2D — 41 $\omega$ bins — 32 x 12 $\phi$ bins

3D — 9 $\theta$ bins
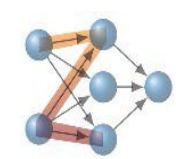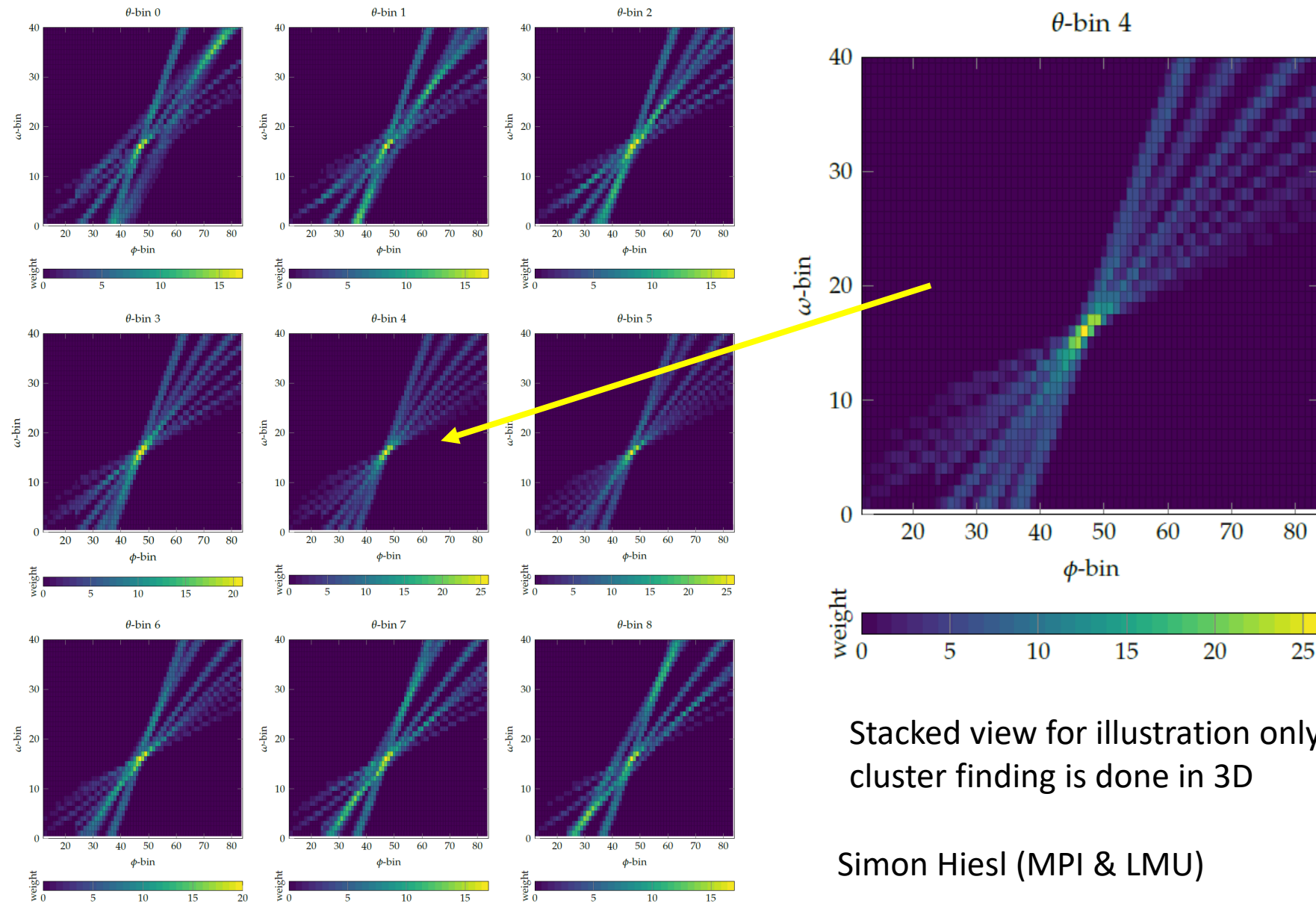
Stacked view for illustration only, cluster finding is done in 3D

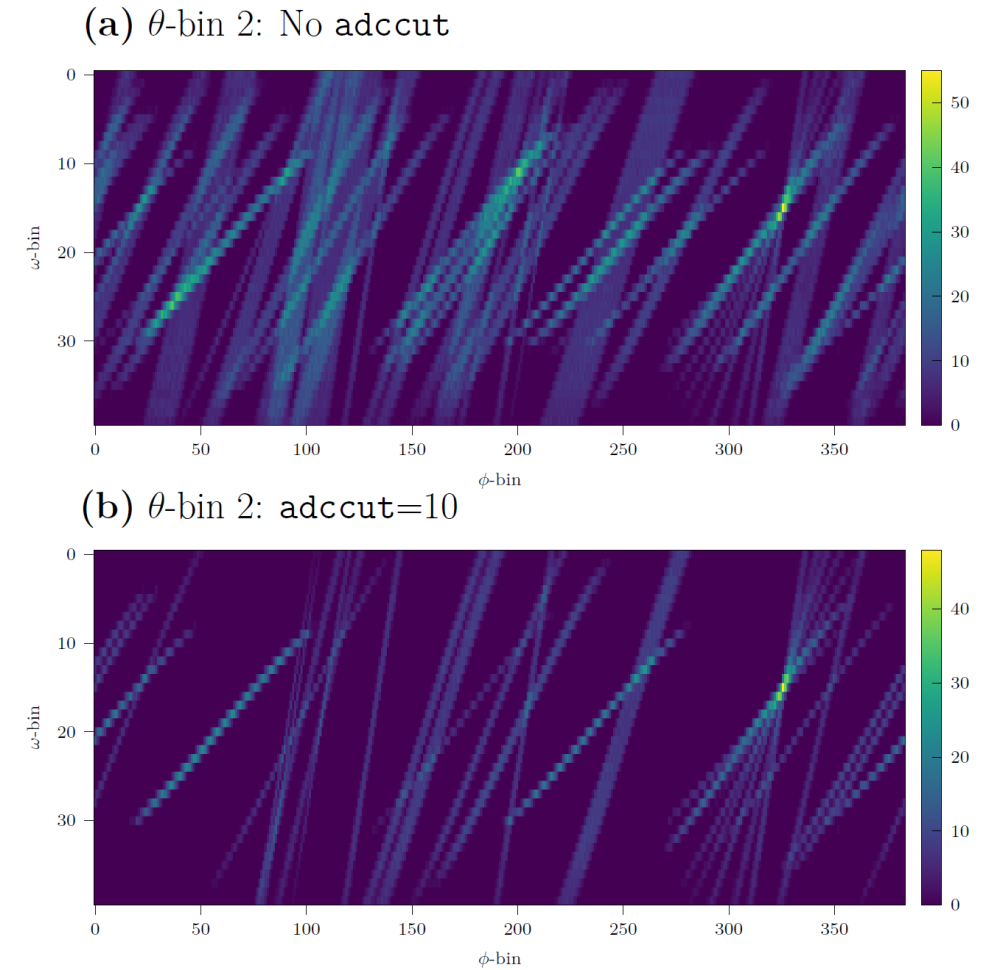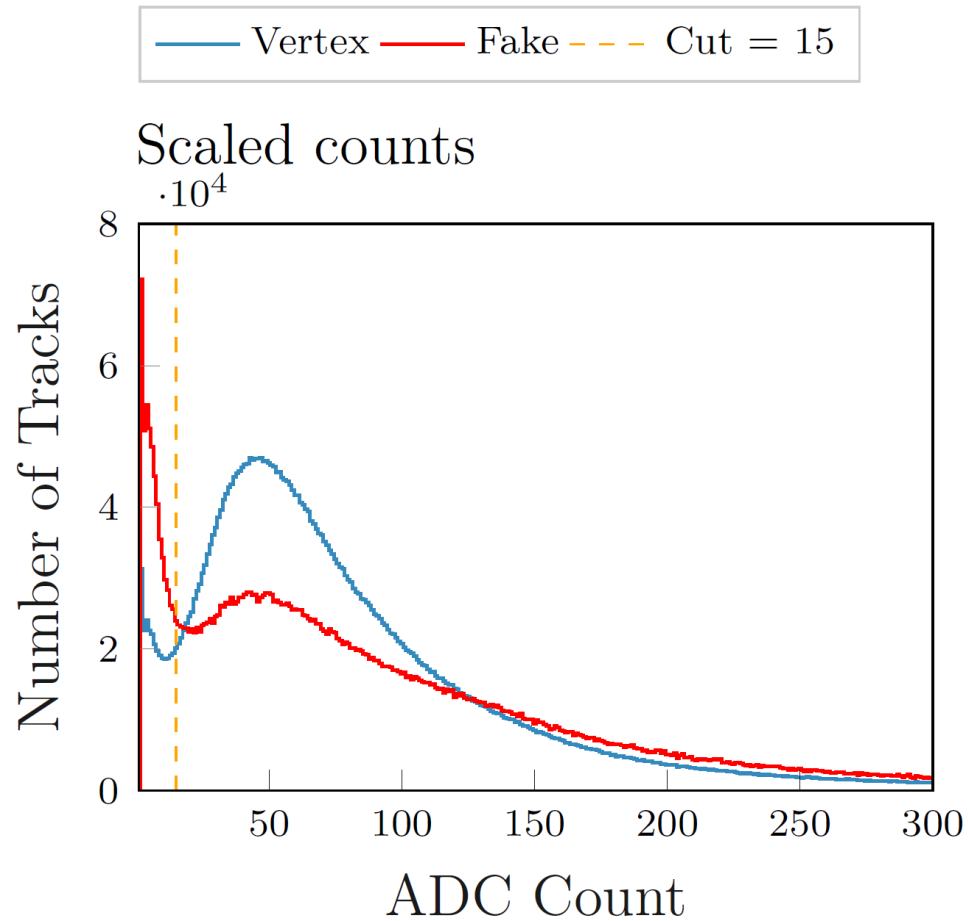Simon Hiesl (MPI & LMU)

# Real Data Analysis

- Very high backgrounds were observed in the last experiment (due to high luminosity)

- The Hough spaces contain a lot of background track segments



$\Longrightarrow$ Reduction of noise using a cut on the ADC count

Simon Hiesl (MPI & LMU)

# Efficiency on Real Single Track Events

Simon Hiesl (MPI & LMU)

- Hit to cluster relation:
  - ▶ All hits in a cluster are considered
  - ▶ The largest weight distribution for each SL is used

- Cut on the number of axial and stereo SL hits (for background reduction)

Efficiency for single track events: Cut at $\pm 10\,\mathrm{cm}$

| adccut | Efficiency 3D | Efficiency 2D |
|--------|---------------|---------------|
| No Count | 94.1% | 94.0% |
| 10 Counts | 96.3% | 95.3% |

Fake-Rate for all found tracks:

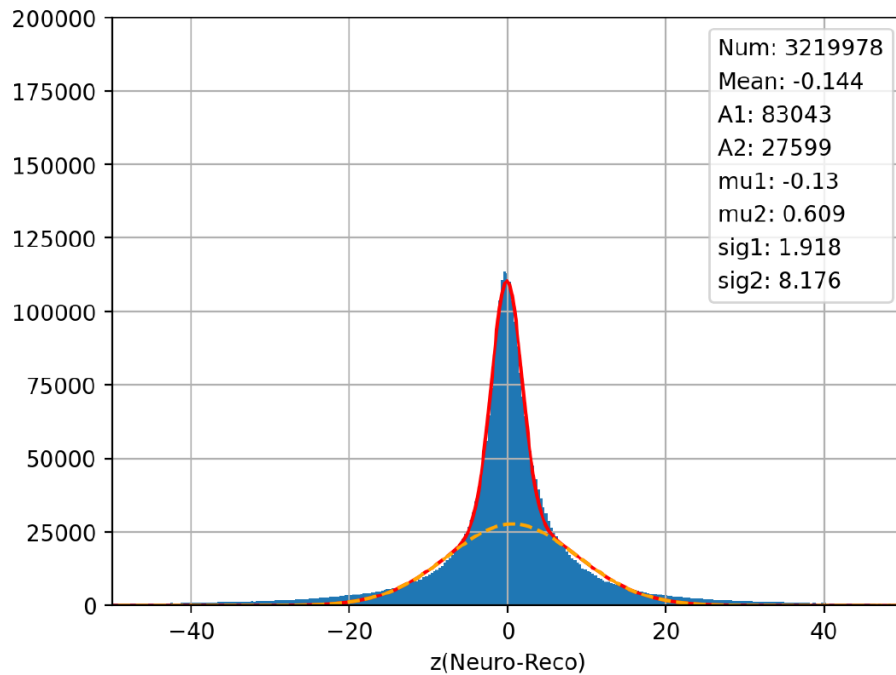| adccut | Fake-Rate 3D | Fake-Rate 2D |
|--------|--------------|--------------|
| No Count | 13.1% | 31.6% |
| 10 Counts | 5.8% | 13.5% |

But: Neural network not trained for 3D candidates at the moment (see presentation by Timo Forsthofer)
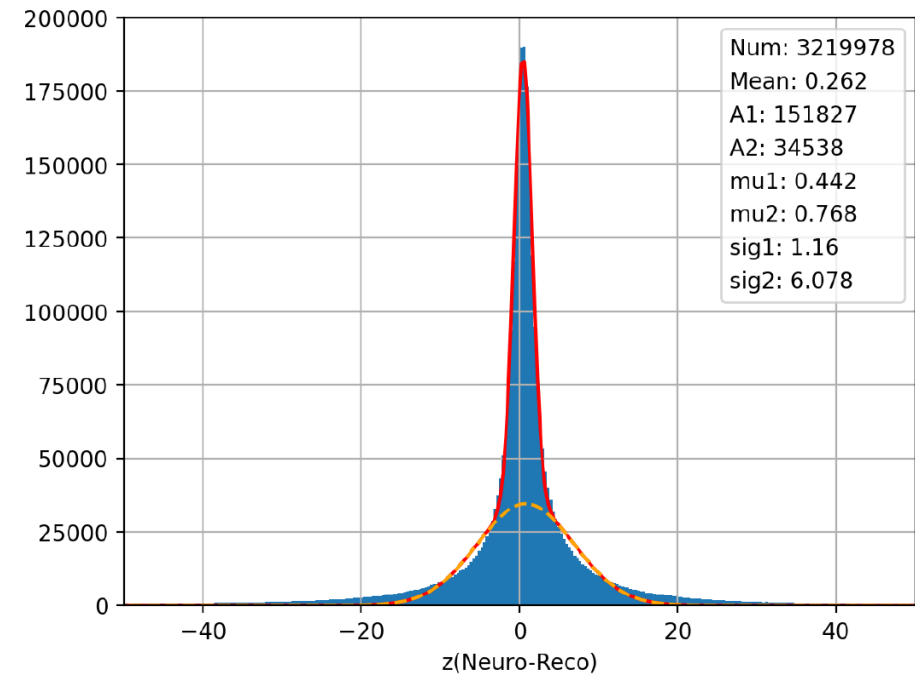
# Deep Learning Architectures

- New, more powerful FPGAs allow for bigger networks

- Three or four hidden layers beneficial for resolution

Timo Forsthofer
(MPP & LMU)

- More hidden layers better than more nodes per layer



Num: 3219978
Mean: -0.144
A1: 83043
A2: 27599
mu1: -0.13
mu2: 0.609
sig1: 1.918
sig2: 8.176

z(Neuro-Reco)

1HL with 81 Nodes



Num: 3219978
Mean: 0.262
A1: 151827
A2: 34538
mu1: 0.442
mu2: 0.768
sig1: 1.16
sig2: 6.078

z(Neuro-Reco)

4HL with 100 Nodes per HL
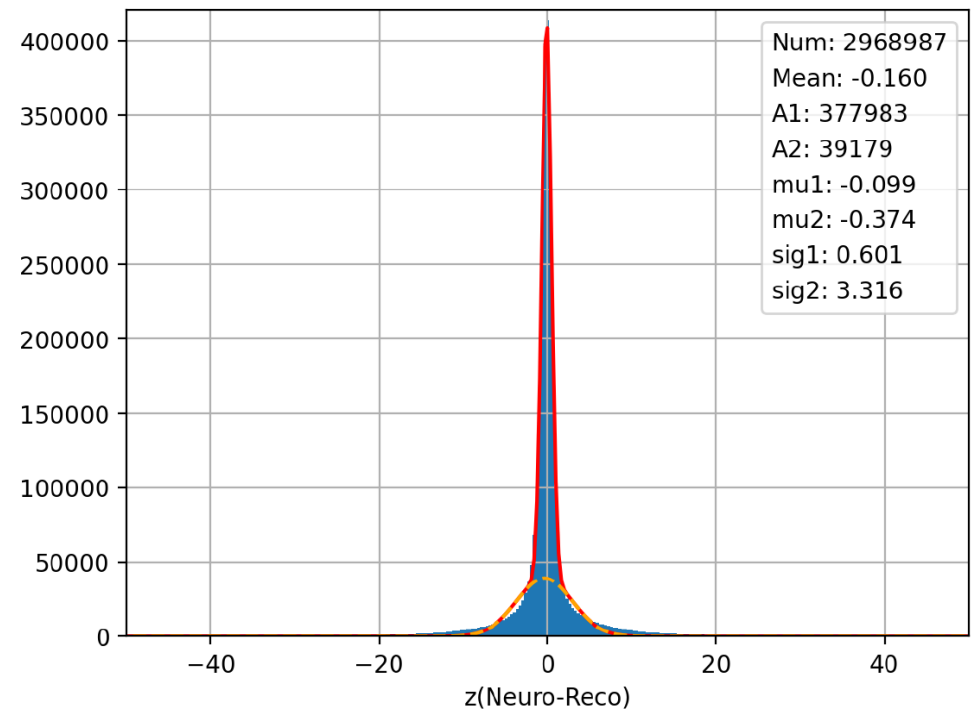
- Combination of all advances leads to increase in accuracy by almost a factor of three
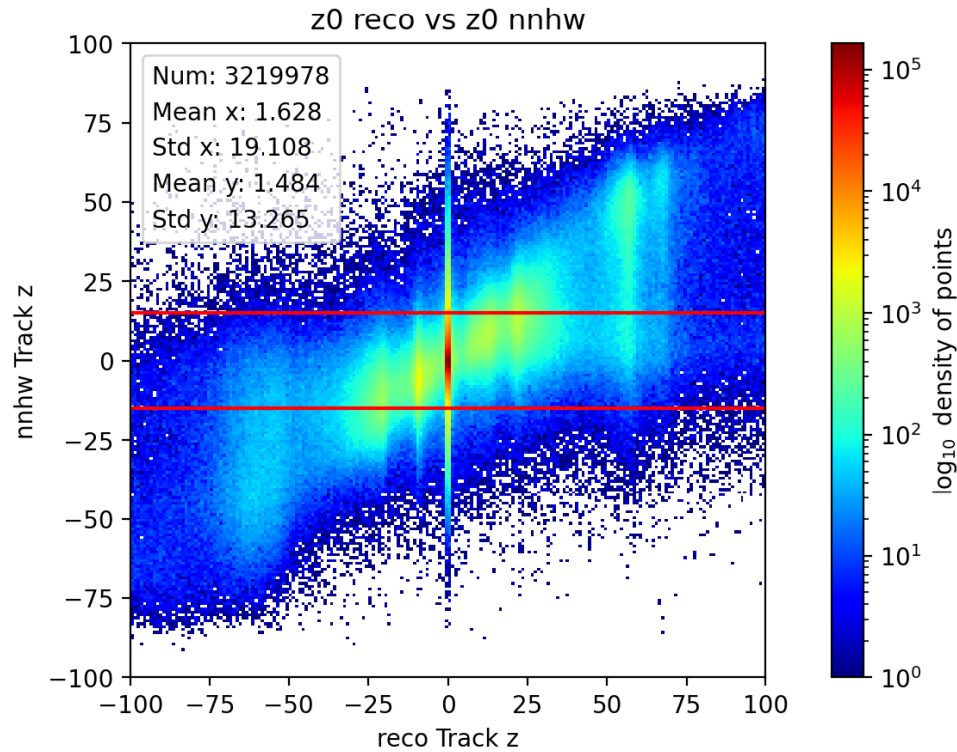
- z-Cut can be reduced from 15cm to under 10cm



| | |
|---|---|
| Num: 3219978 | |
| Mean: -0.144 | |
| A1: 83043 | |
| A2: 27599 | |
| mu1: -0.13 | |
| mu2: 0.609 | |
| sig1: 1.918 | |
| sig2: 8.176 | |

Present Network Architecture

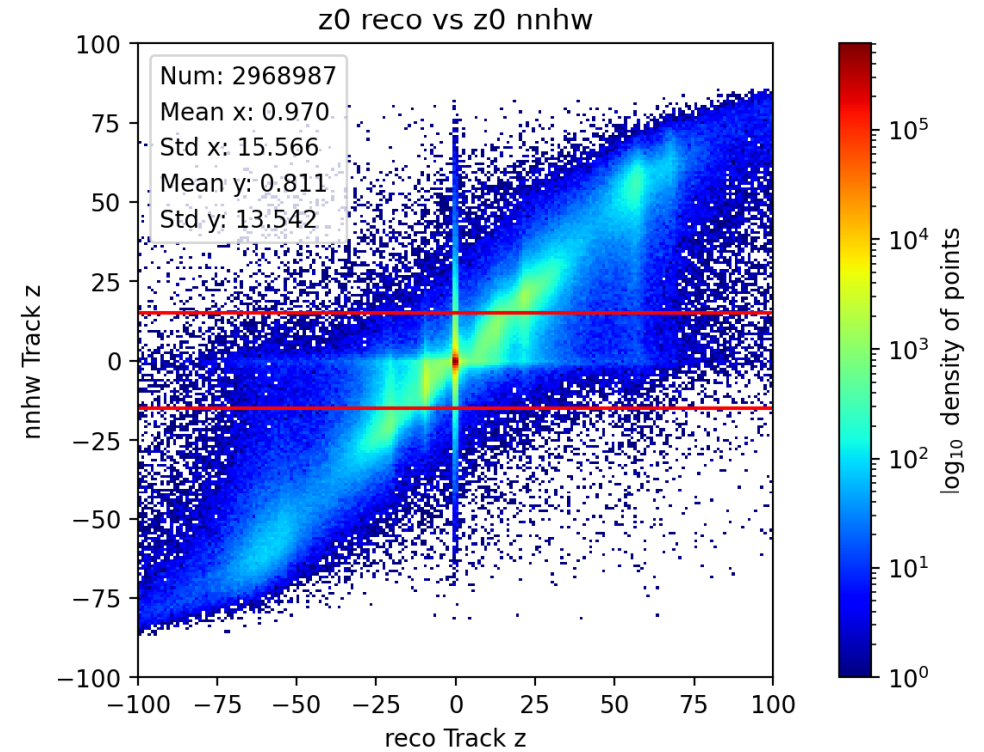| | |
|---|---|
| Num: 2968987 | |
| Mean: -0.160 | |
| A1: 377983 | |
| A2: 39179 | |
| mu1: -0.099 | |
| mu2: -0.374 | |
| sig1: 0.601 | |
| sig2: 3.316 | |

Deep Neural Network with Extended Input, ADC-cut and 3D-Input

- Especially extended input helpful in reducing Feed-Up and Feed-Down



Present Network Architecture



Deep Neural Network with Extended Input, ADC-cut and 3D-Input