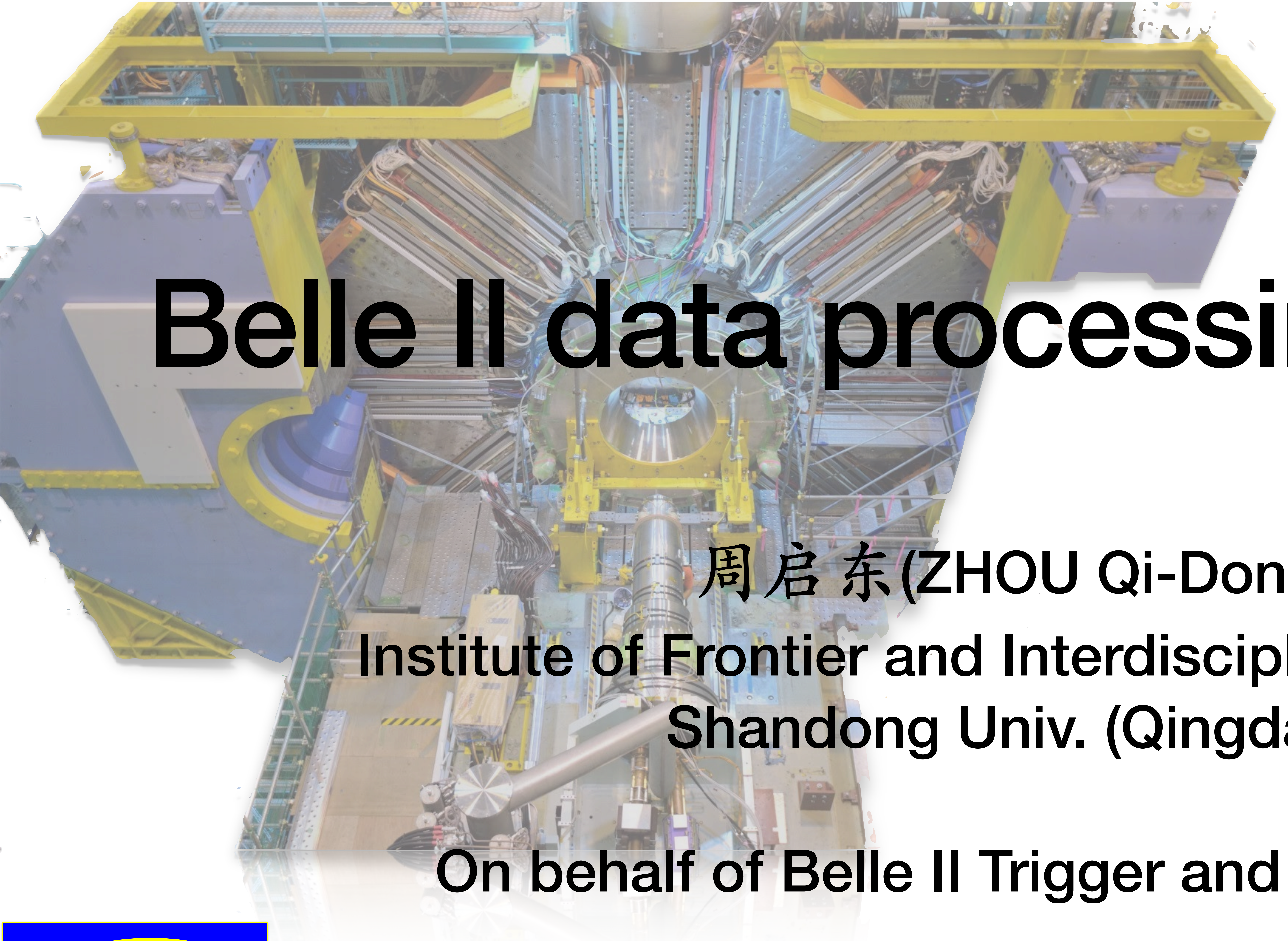


Belle II data processing with ML

周启东(ZHOU Qi-Dong)

Institute of Frontier and Interdisciplinary Science,
Shandong Univ. (Qingdao)

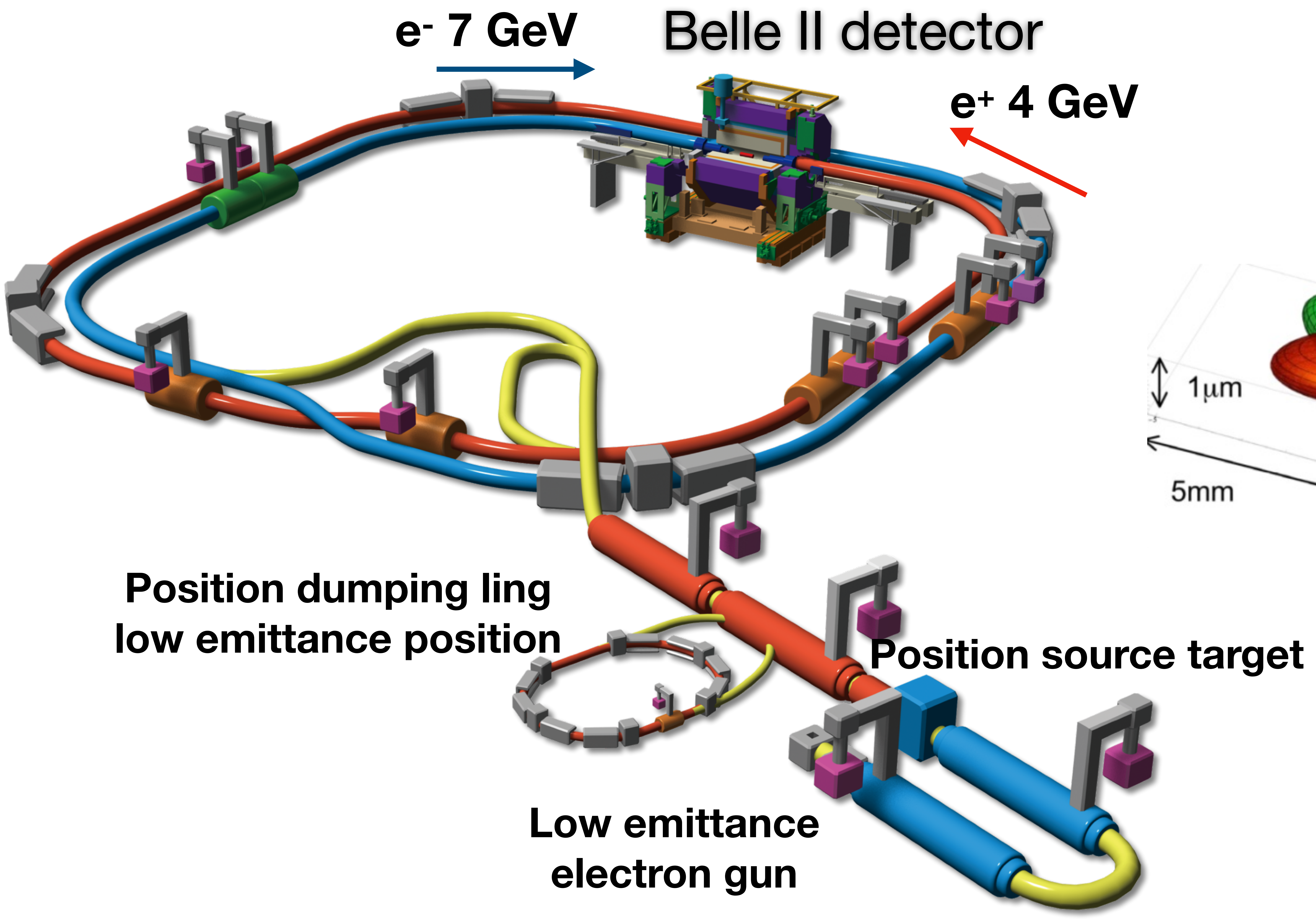
On behalf of Belle II Trigger and DAQ group

18-21 Dec. 2024, Guangzhou
The 6th International Workshop on Future Tau Charm Facilities

Luminosity frontier: SuperKEKB

Luminosity frontier: SuperKEKB

- Asymmetric e^+e^- collider
  - $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$
  - very clean and well-known initial state

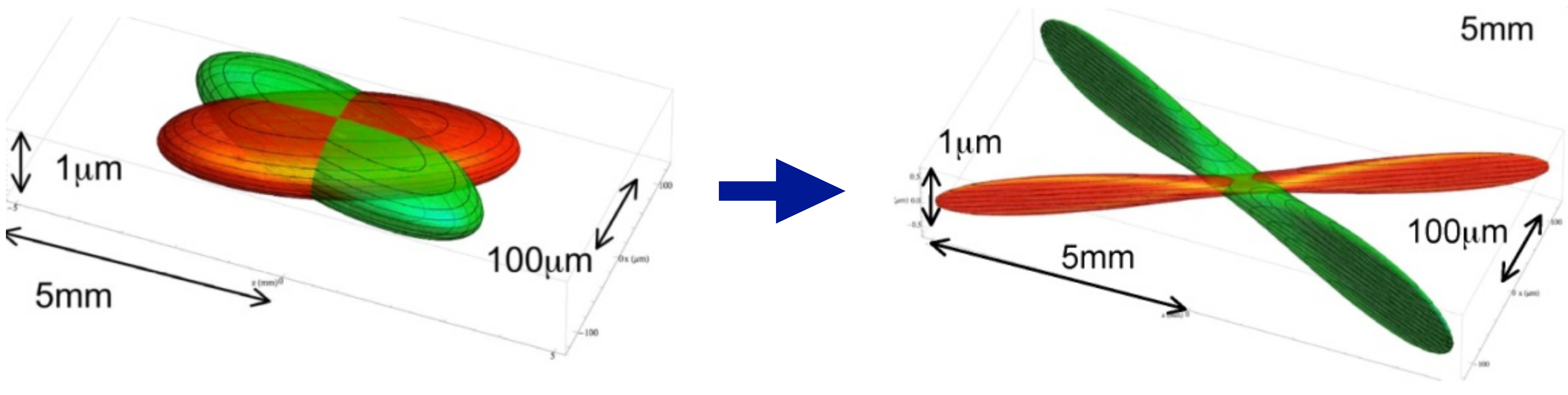


Beam current: KEKB x ~1.5

$$L = \frac{\gamma_\pm}{2er_e}\left(1+\frac{\sigma_y^*}{\sigma_x^*}\right)\frac{I_\pm \xi_{\pm y}}{\beta_y^*}\left(\frac{R_L}{R_y}\right)$$

Beam squeeze: KEKB / ~20

<u>Nano beam scheme</u>



Target: L = 60 x 10^{34} cm^{-2} s^{-1}

Achieved : 4.7 x 10^{34} cm^{-2} s^{-1} (Record)

- Data:
  - **531** fb^{-1} (Belle II) <-> 980 fb^{-1} (Belle)

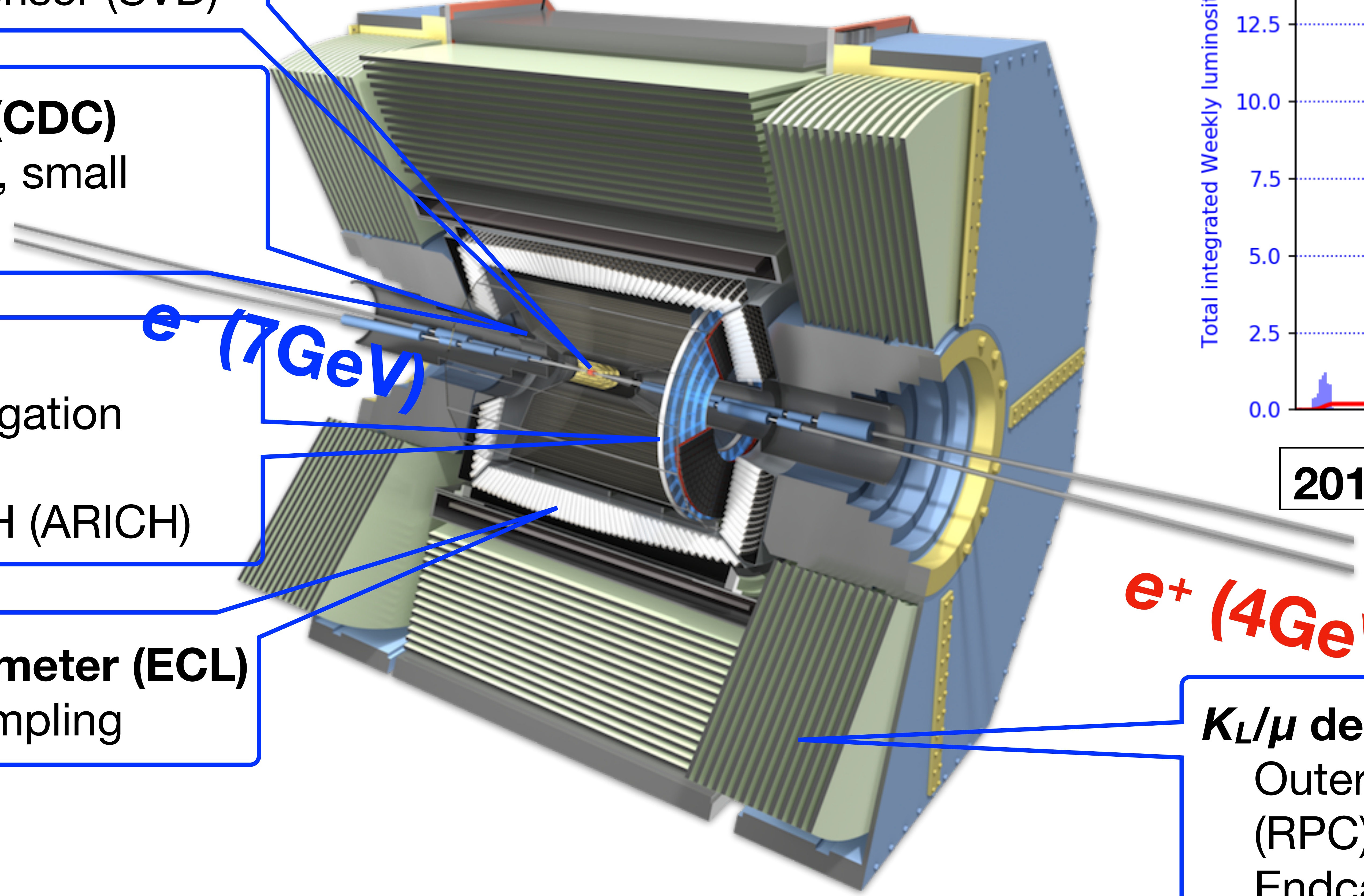
Belle II detector and dataset

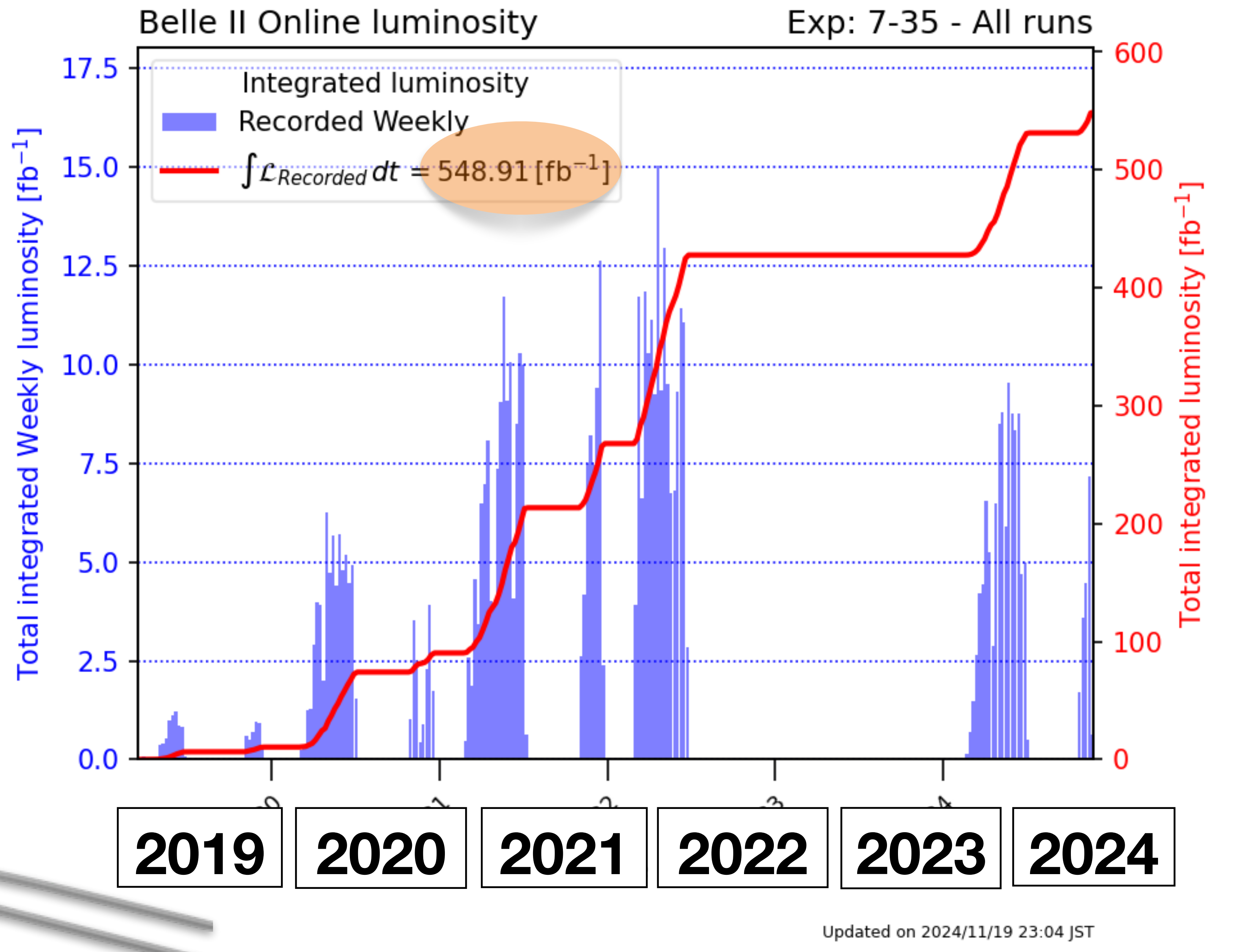
Vertex detector (VXD)
Inner 2 layers: pixel detector (PXD)
Outer 4 layers: strip sensor (SVD)

Central Drift Chamber (CDC)
He (50%), C_2H_6 (50%), small cells, long lever arm

Particle Identification
Barrel: Time-Of-Propagation counters (TOP)
Forward: Aerogel RICH (ARICH)

ElectroMagnetic Calorimeter (ECL)
CsI(Tl) + waveform sampling





***K_L/μ* detector (KLM)**
Outer barrel: Resistive Plate Counter (RPC)
Endcap/inner barrel: Scintillator

- Features:
  - Near-hermetic detector
  - Vertexing and tracking: σ vertex ~ 15μm, CDC spatial res. 100μm $\sigma(P_T)/P_T$ ~ 0.4%
  - Good at measuring neutrals, π^0, γ, K_L… $\sigma(E)/E$ ~ 2-4%

Belle II trigger strategy

- Design requirements: ~100% for $\Upsilon(4S)$->BB(hadronic decay), Tau/Charm, Exotics
  - No dead-time -> pipeline
  - Single photon trigger
  - Single track trigger
- Max. trigger rate: 30 kHz @ 6 x 10^{35} cm^{-2} s^{-1}
  - Physics trigger ~15 kHz
- Latency limit: ~5 usec (SVD APV25 buffer structure)
  - A fixed latency of about 4.4 usec
- Event timing resolution: 10 nsec

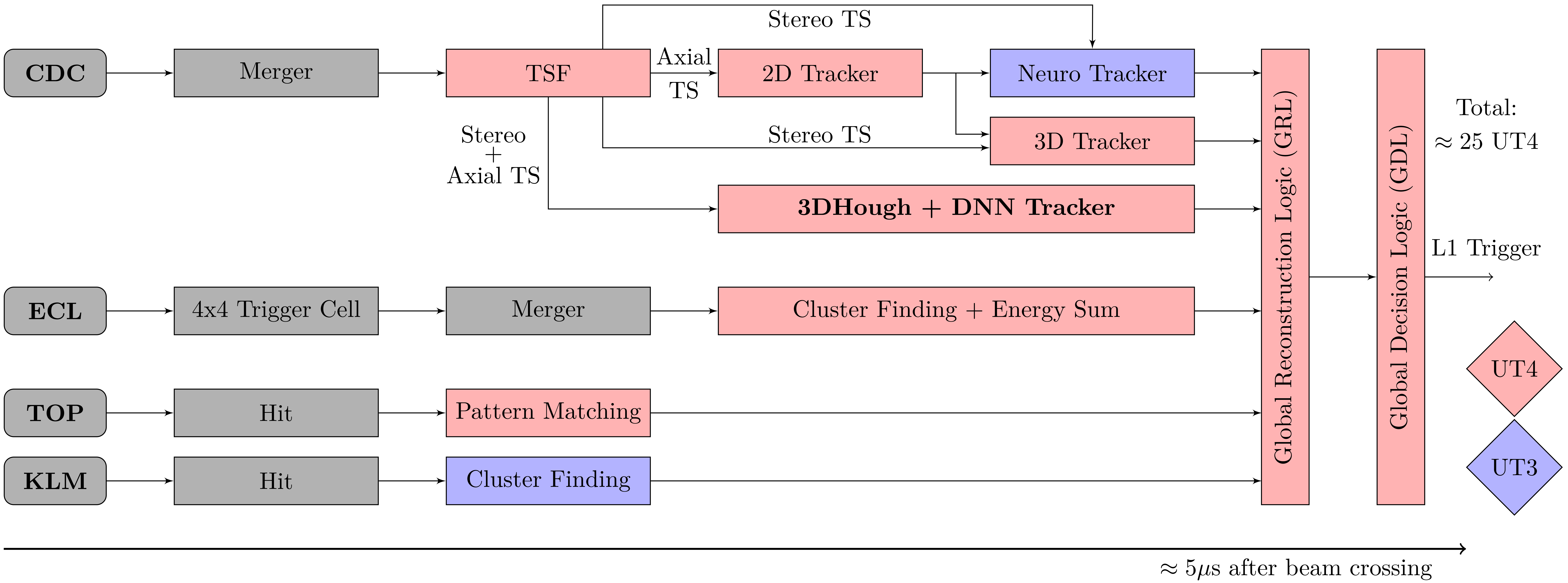
Process	σ(nb)	Rate@L=$6x10^{35}$ (kHz)
Bunch. cross.	-	$2x10^5$
Beam bkg	-	300-600
Bhabha	44	50
Total->L1	-	200350->~15

Process	σ(nb)	L1@L=$6x10^{35}$ (kHz)
Bhabha	44	0.35*
Two photon	13	10
Upsilon(4S)	1.2	0.96
Continuum	2.8	2.2
μμ	0.8	0.64
ττ	0.8	0.64
γ-γ	2.4	0.019*
Total	67	~15

Belle trigger system

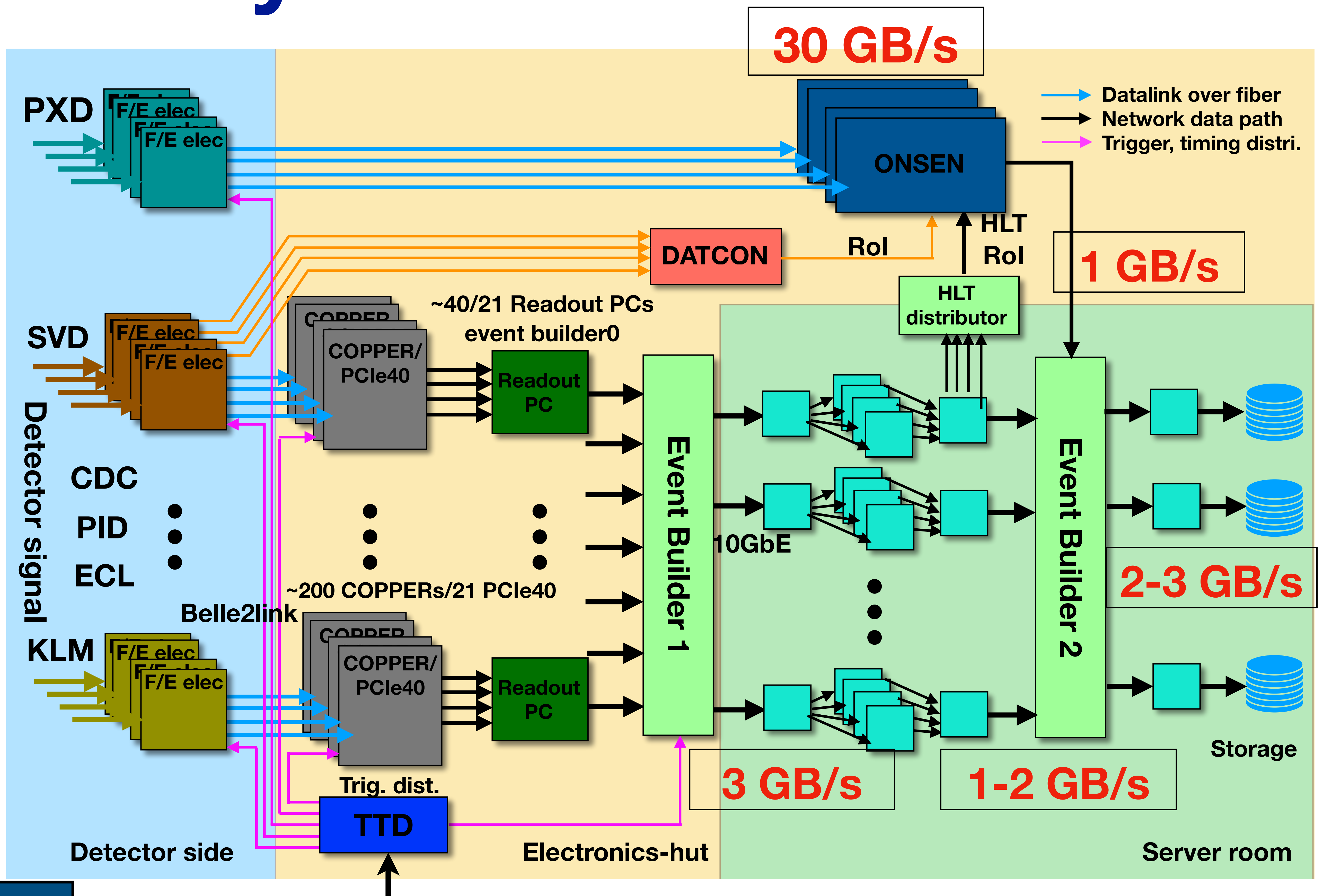
- CDC, ECL: main triggers for tracks and clusters
- KLM: trigger muon
- TOP: event timing
- GRL: matching of sub-triggers
- GDL: final trigger decision

- Challenges:
  - low multiplicity trigger vs. background
  - High track trigger vs. crosstalk
  - Drawback of track trigger at endcap
  - Latency budget vs. transmission and logics
  - …



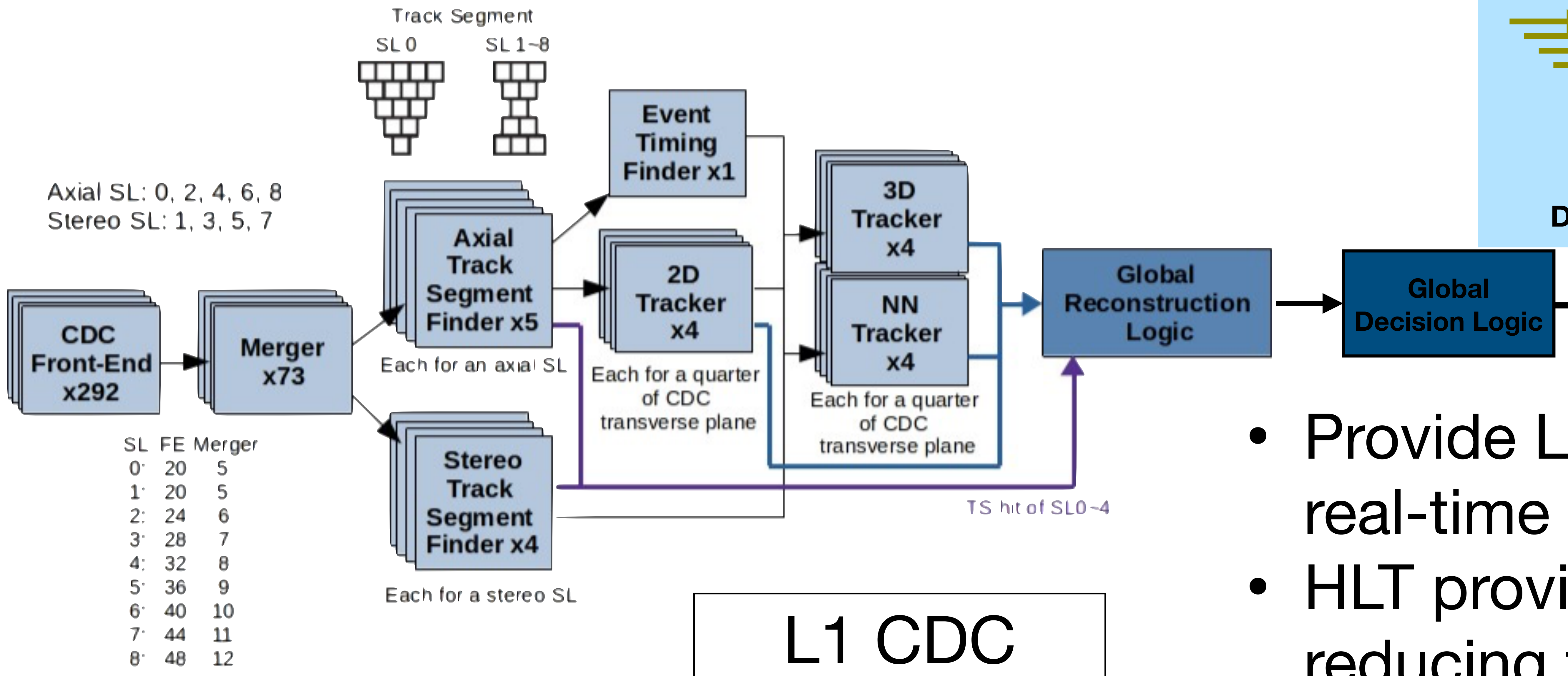
Belle II TDAQ system

- Unified common readout system (except for PXD)
- Unified timing and trigger distribution (TTD) system
- A pipeline readout
- To handle 30 kHz level 1(L1) trigger with O 1% dead time under raw event size of 1 MB



High level trigger(HLT)

Example: CDC

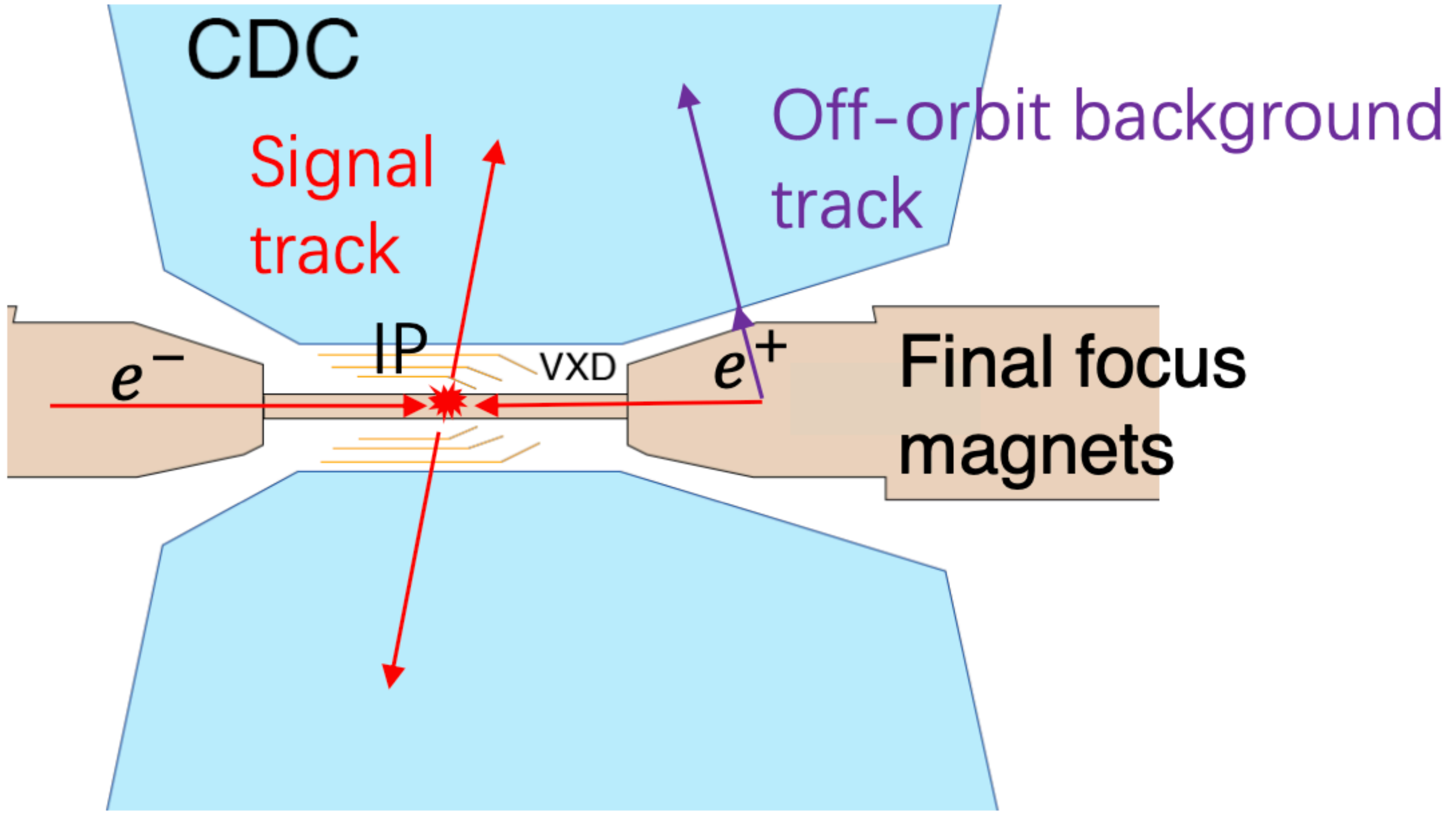


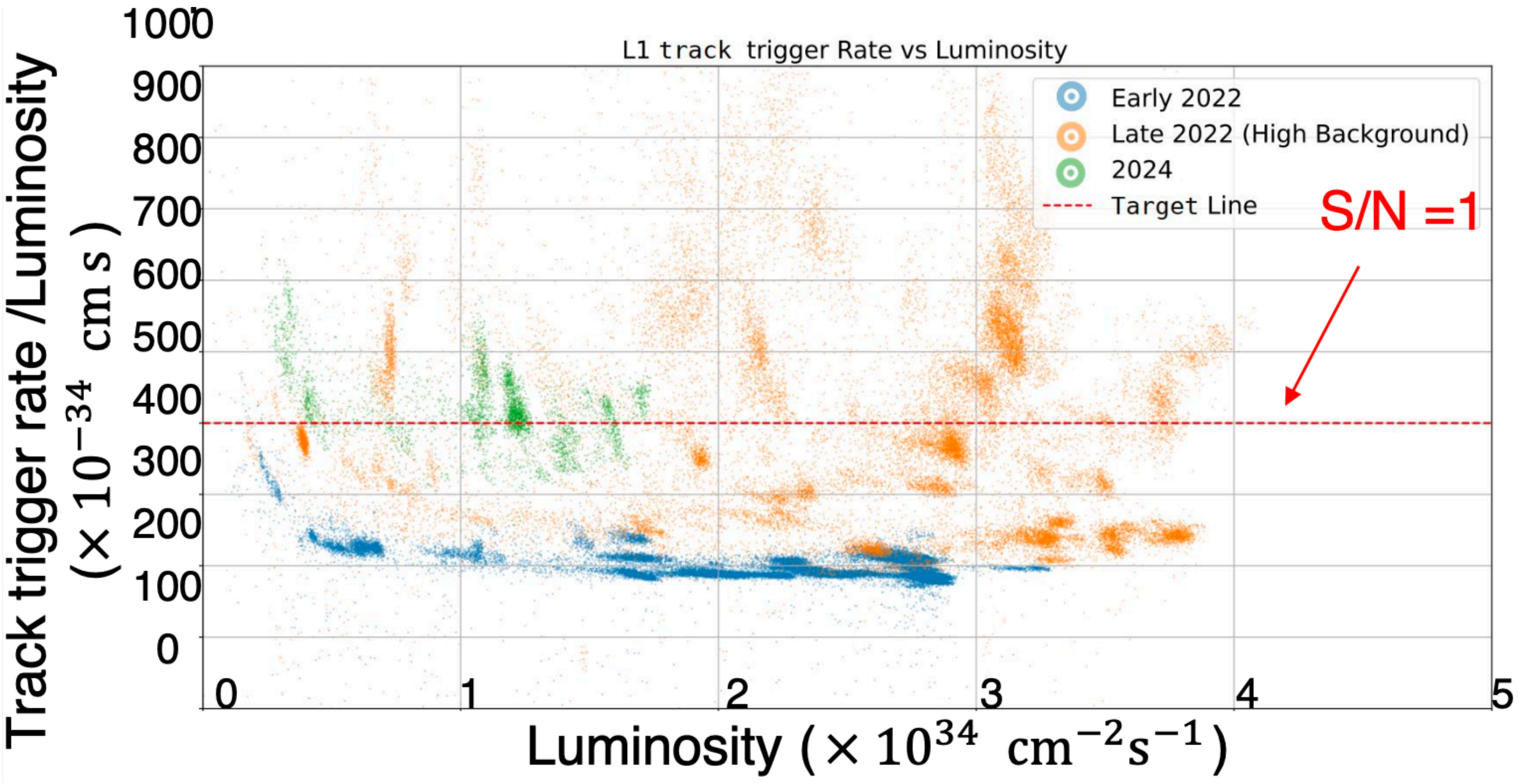
L1 CDC trigger logic

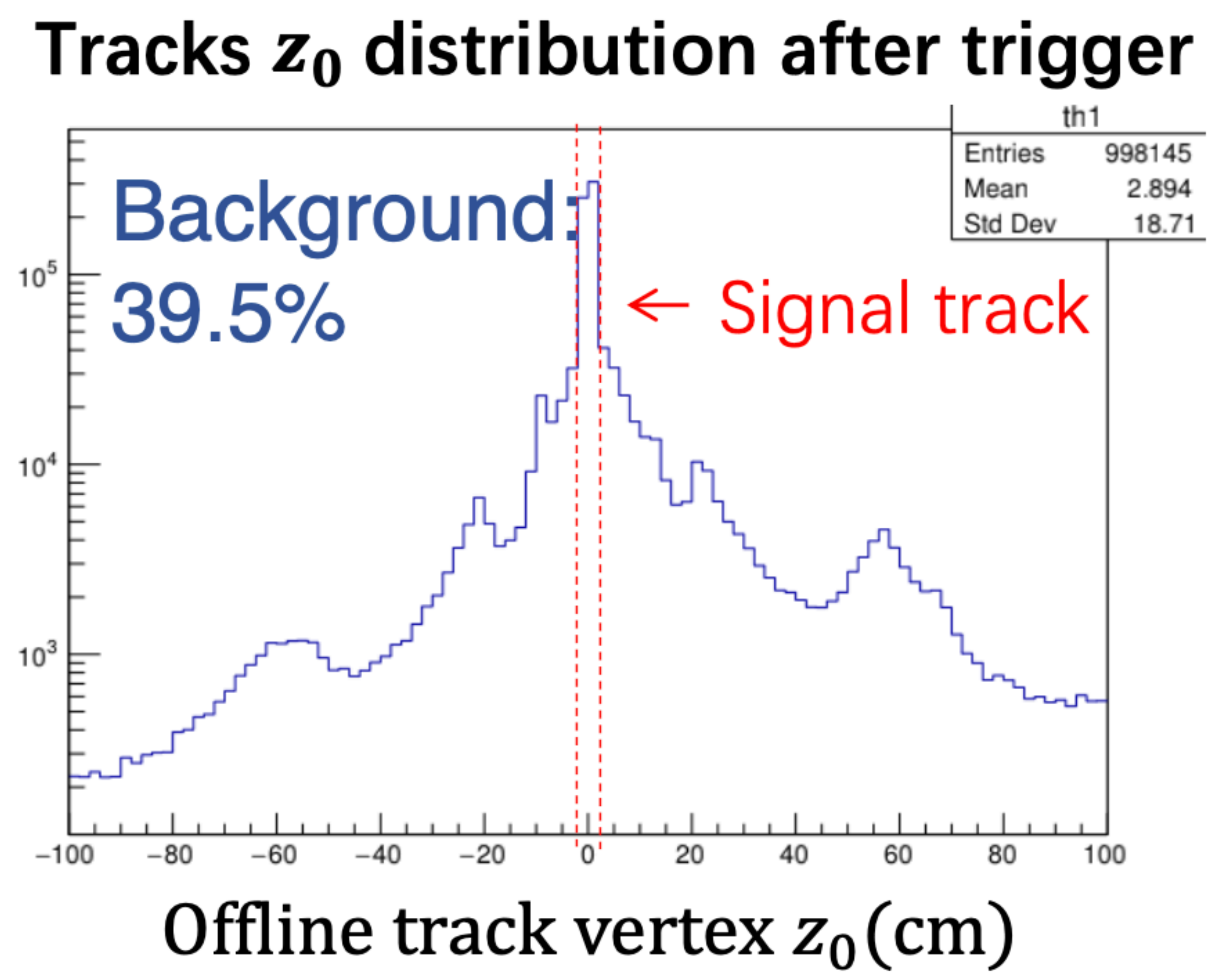
- Provide L1 trigger signal to DAQ using FPGA chips for real-time processing on detector raw data.
- HLT provide Region of Interest (RoI) to PXD for significantly reducing the data size.
  - Latency O sec.

Motivation of DNN for L1 trigger

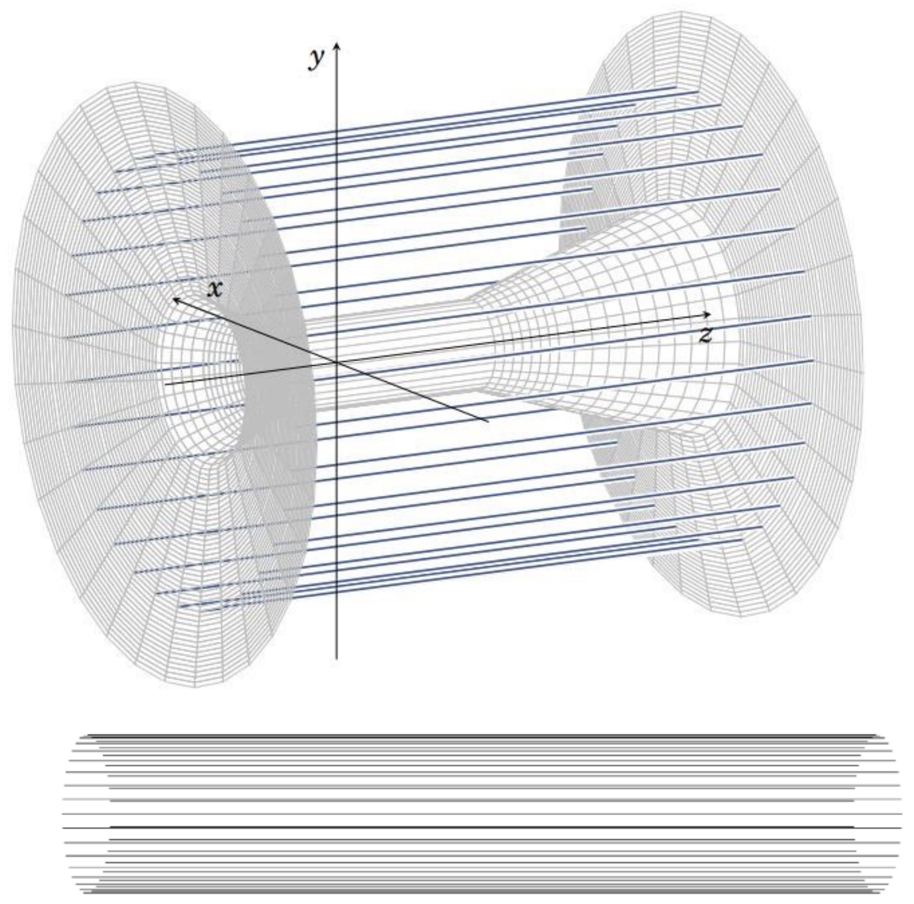
- DAQ system is designed to handle 30 kHz
  - Physical trigger ~15 kHz, require S/N = 1
- L1 trigger rate depends significant on background condition
- Advanced CDC algorithm to further suppress background
- A fixed latency of about 4.4 usec



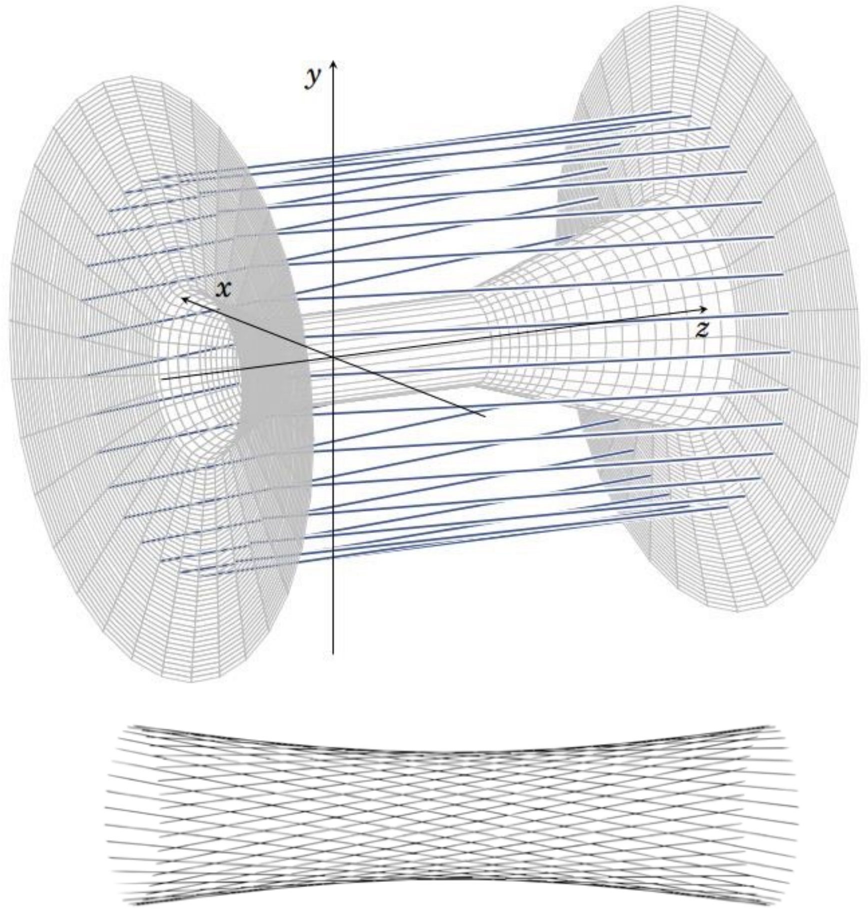




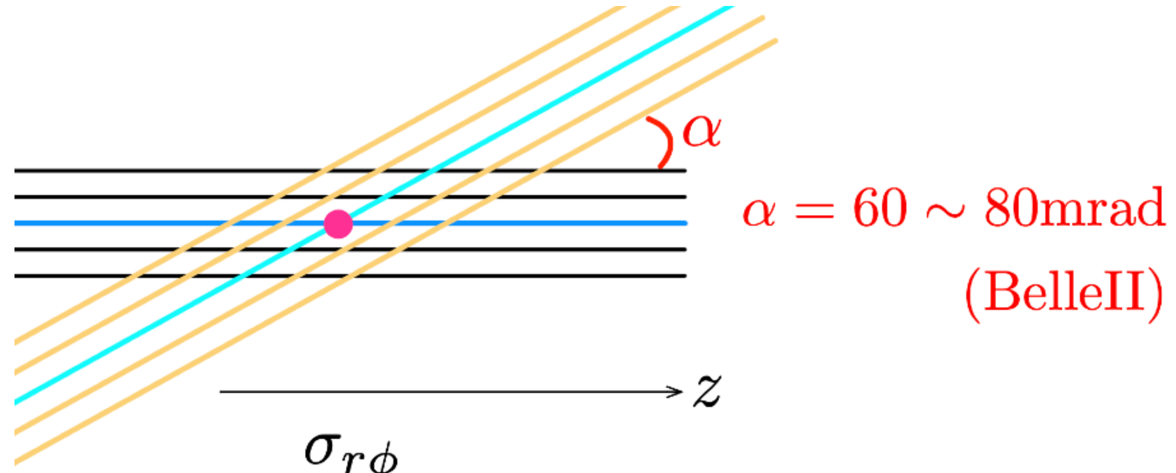
Basics of L1 CDC trigger



Axial wire

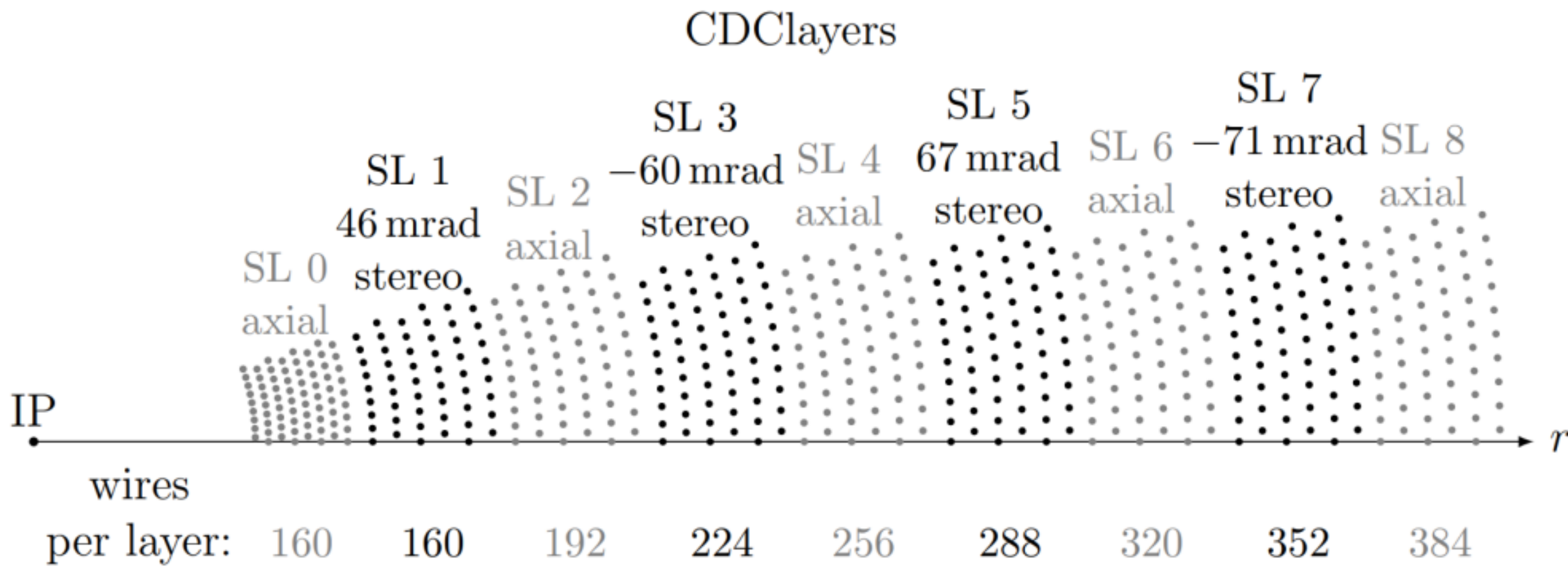


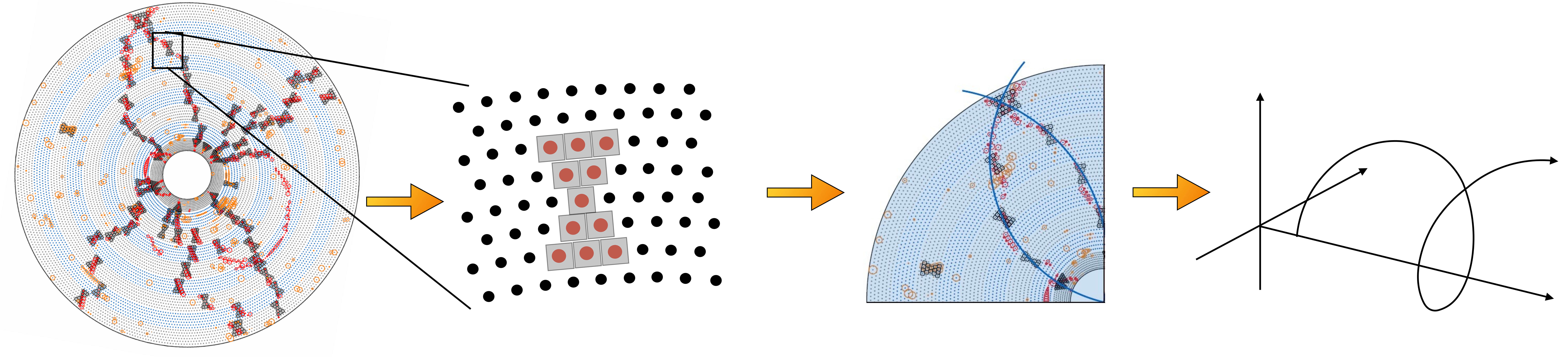
Stereo wire



$$\sigma_z = \frac{\sigma_{r\phi}}{\sin\alpha}$$

$$\sigma_z = 1.9 \sim 2.5\text{mm}(\sigma_{r\phi} = 0.15\text{mm})$$





CDC 2D hit map

Track Segment (TS) Finder in every super layer

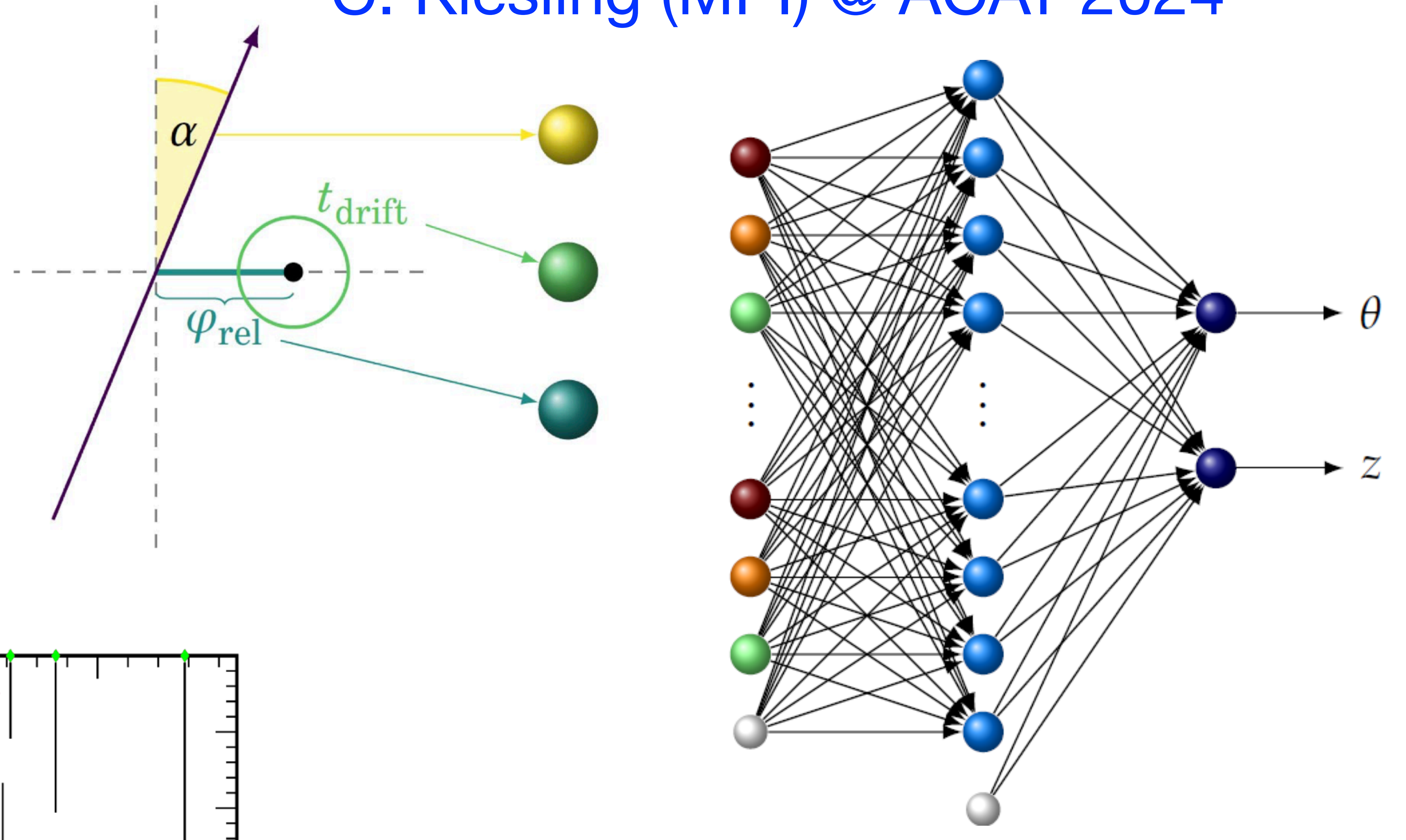
Build 2D track with axial hits using Hough transformation

3D Track trigger base on 2D track and Stereo TS

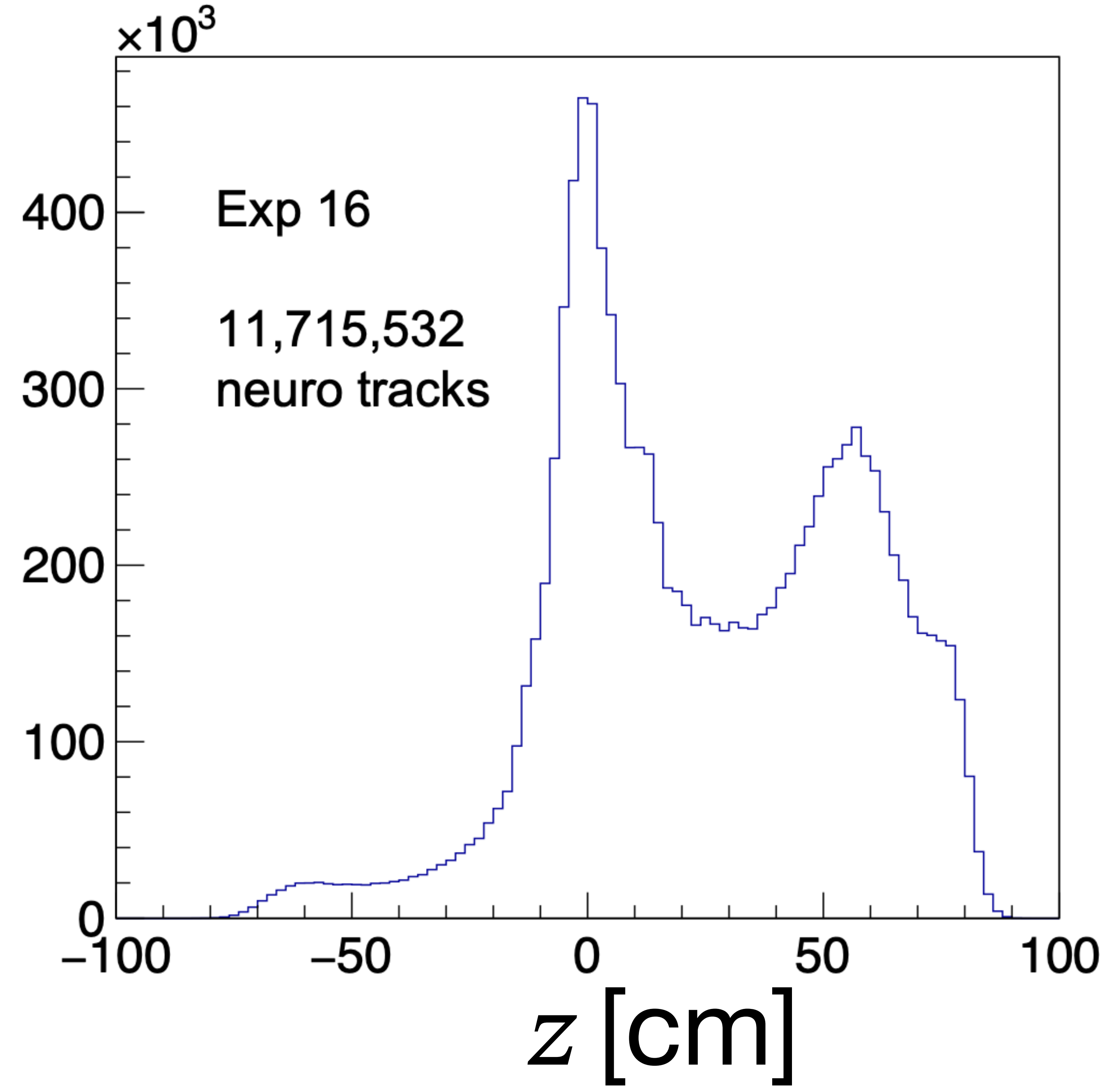
Neutral Network z-trigger

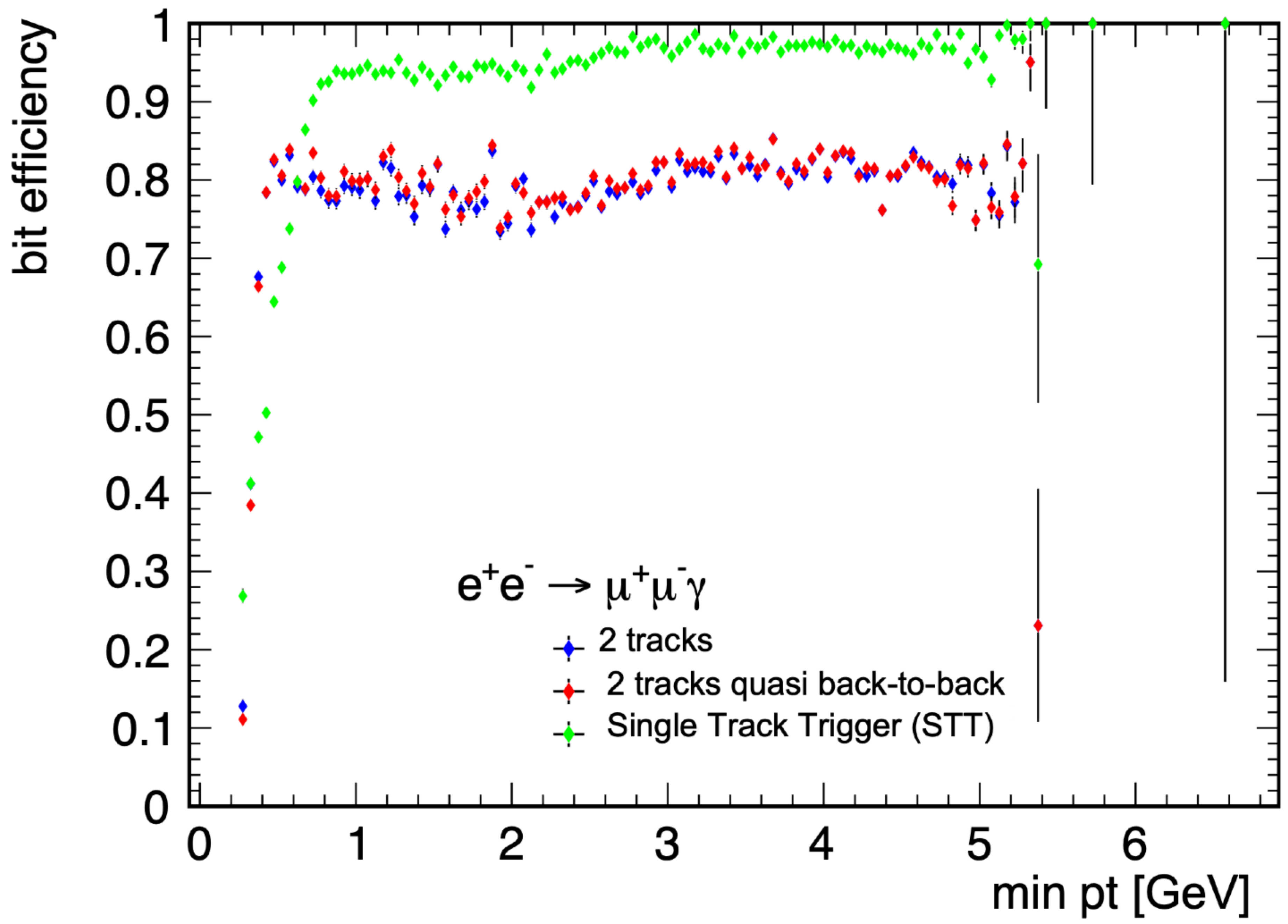
- Pioneer of NN on **HARDWARE** for Belle II trigger
- Inputs: **Drift time t_{drift}**, **wires relative location ϕ_{rel}**, **Crossing angle α** for priority wires
- Outputs: track vertex z_0, track θ
- Selected 1 Track Segment per one Super Layer
- Networks trained with real data from May-June 2020

C. Kiesling (MPI) @ ACAT 2024



arXiv:2402.14962 (submitted to NIMA)



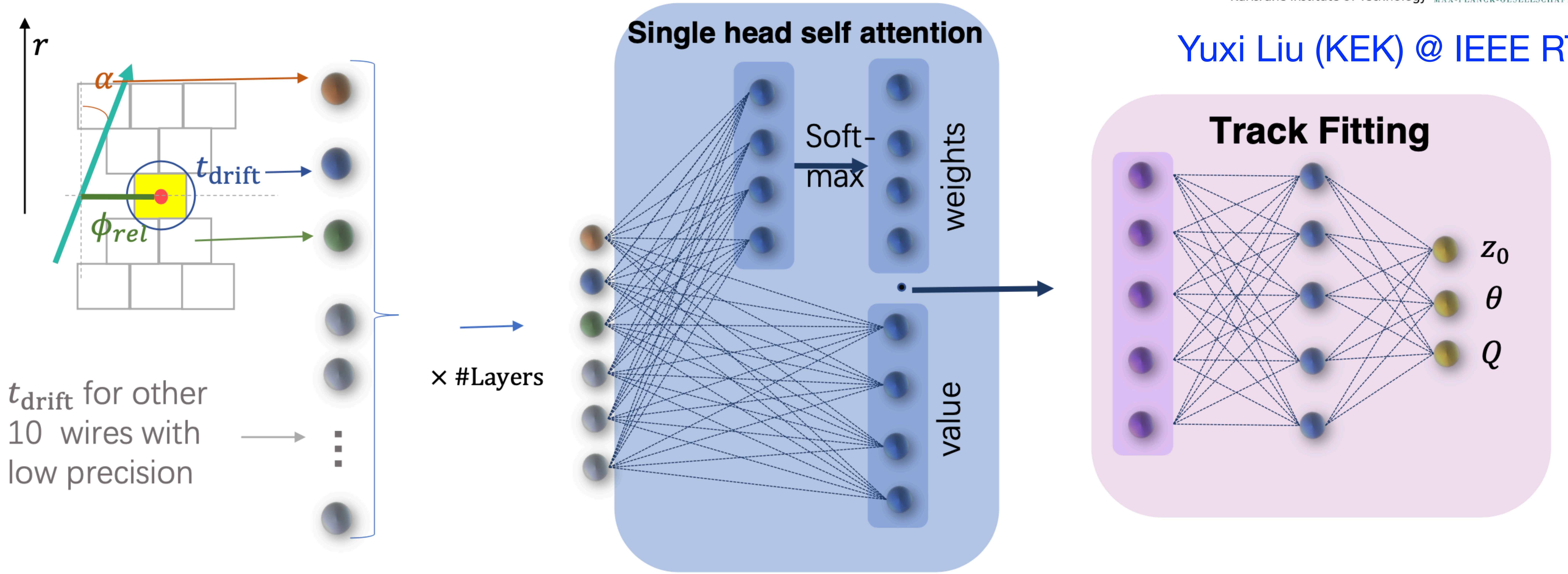


- Backgrounds still high
- Fake background

- Upgrade plan in trigger group:
- Track finding in **3D Hough space**
- network architecture: “deep-learning” + additional inputs

Deep Neutral Network

Yuxi Liu (KEK) @ IEEE RT 2024



- Inputs: **Drift time t_{drift}**, **wires relative location ϕ_{rel}**, **Crossing angle α** for priority wires + **Drift time for all other wires**
- Introduce the **self-attention architecture** to “focus” on certain inputs
- Output track vertex z_0, track θ and **signal/ background classifier output** (Q)

Parameter	#Attention value	#hidden nodes	#hidden layer	activate	precision	Total multiplier
Values	**27**	**27**	**2**	**Leaky Relu**	**Float 16**	**4,185**

Development flow of DNN on FPGA

Software → Hardware → FPGA

Keras + TensorFlow
PyTorch

- Machine Learning model
- Parameter

hls4ml

- C/C++ transition

Vivado™ HLS

- Translate into Verilog/VHDL FPGA language

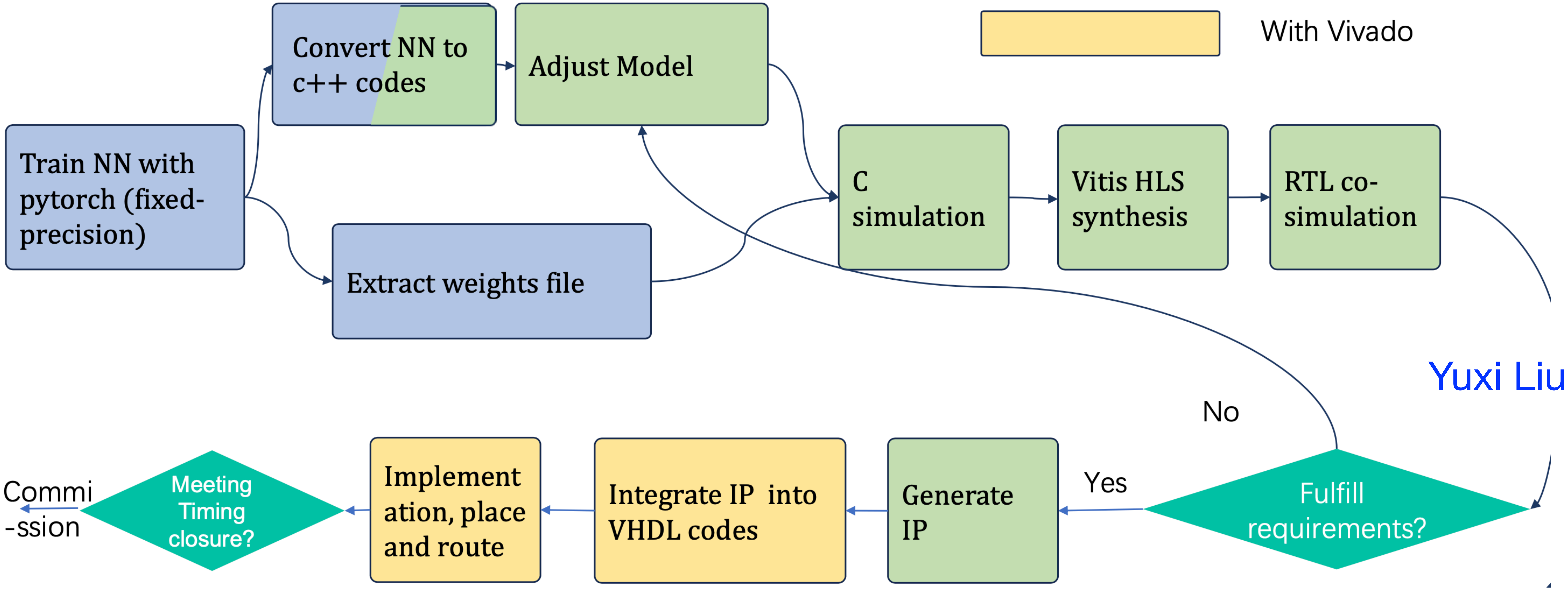
VIVADO

- Start fitter

- Evaluation

*include some function from hls4ml lib

Blue	With Python
Green	With Vitis HLS
Yellow	With Vivado



Belle II UT4

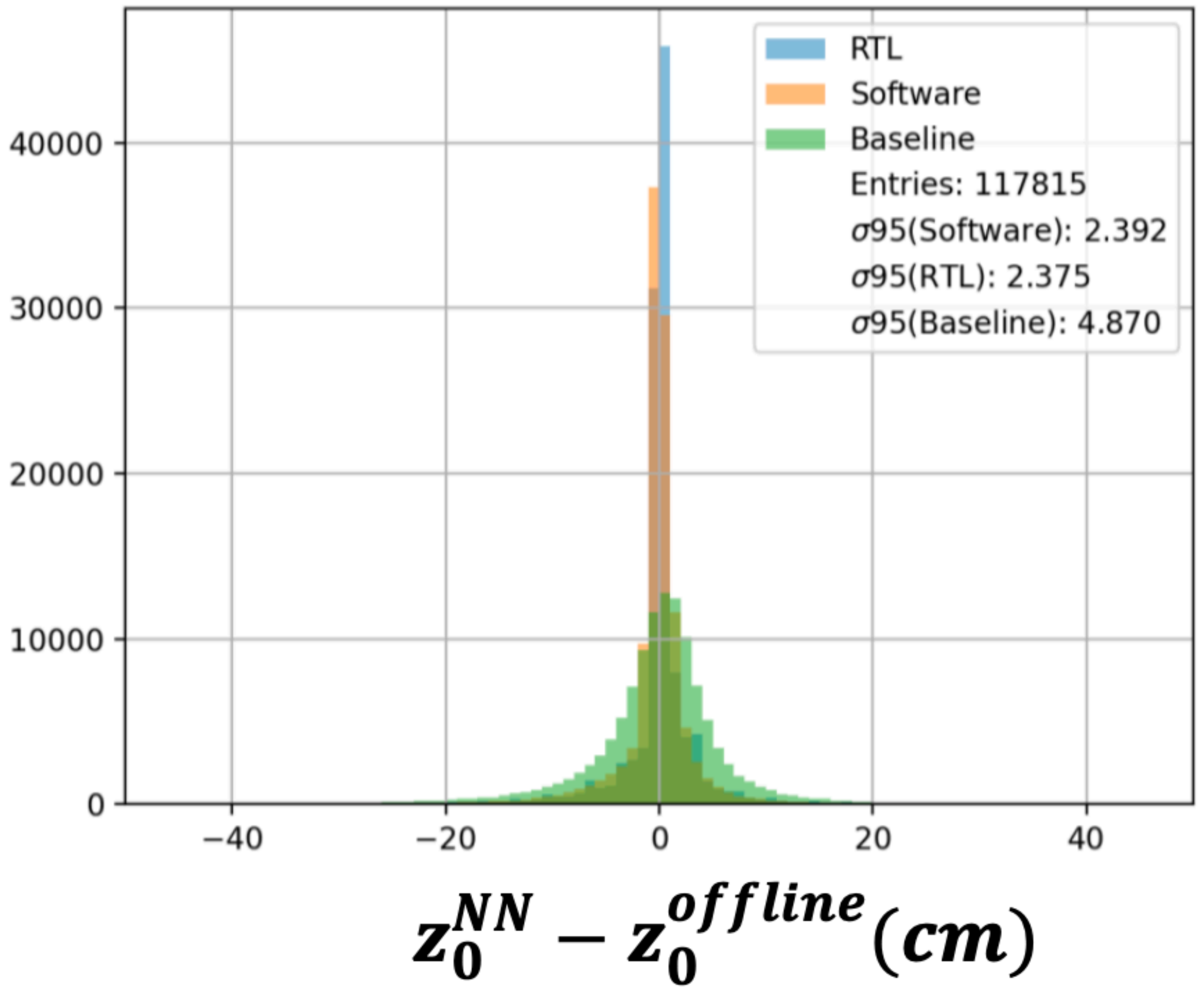
Xilinx UltraScale
XCVU080, XCVU160
25 Gbps with 64B/66B

Yuxi Liu (KEK) @ IEEE RT 2024

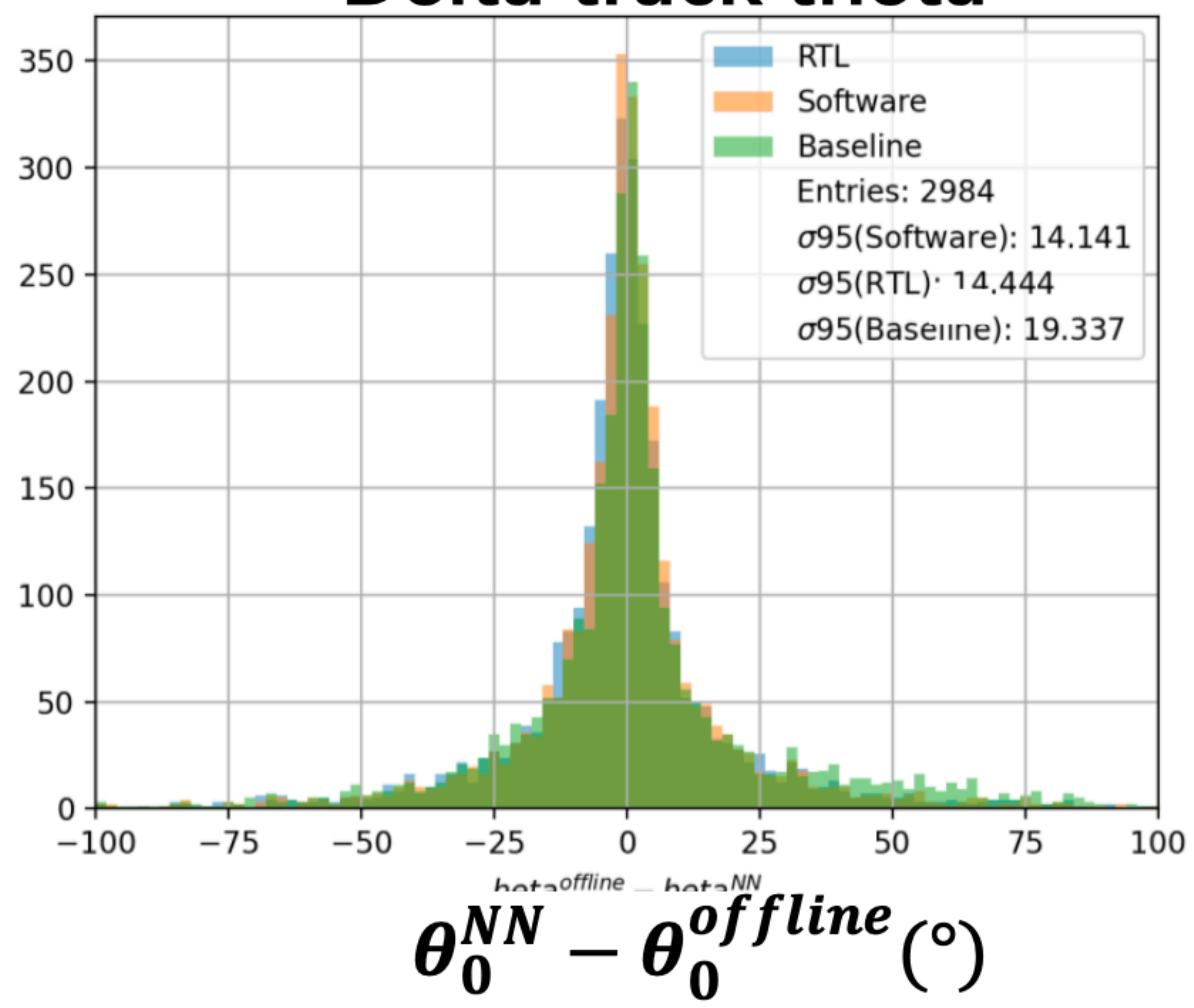
Simulation performance of DNN

Yuxi Liu (KEK) @ IEEE RT 2024

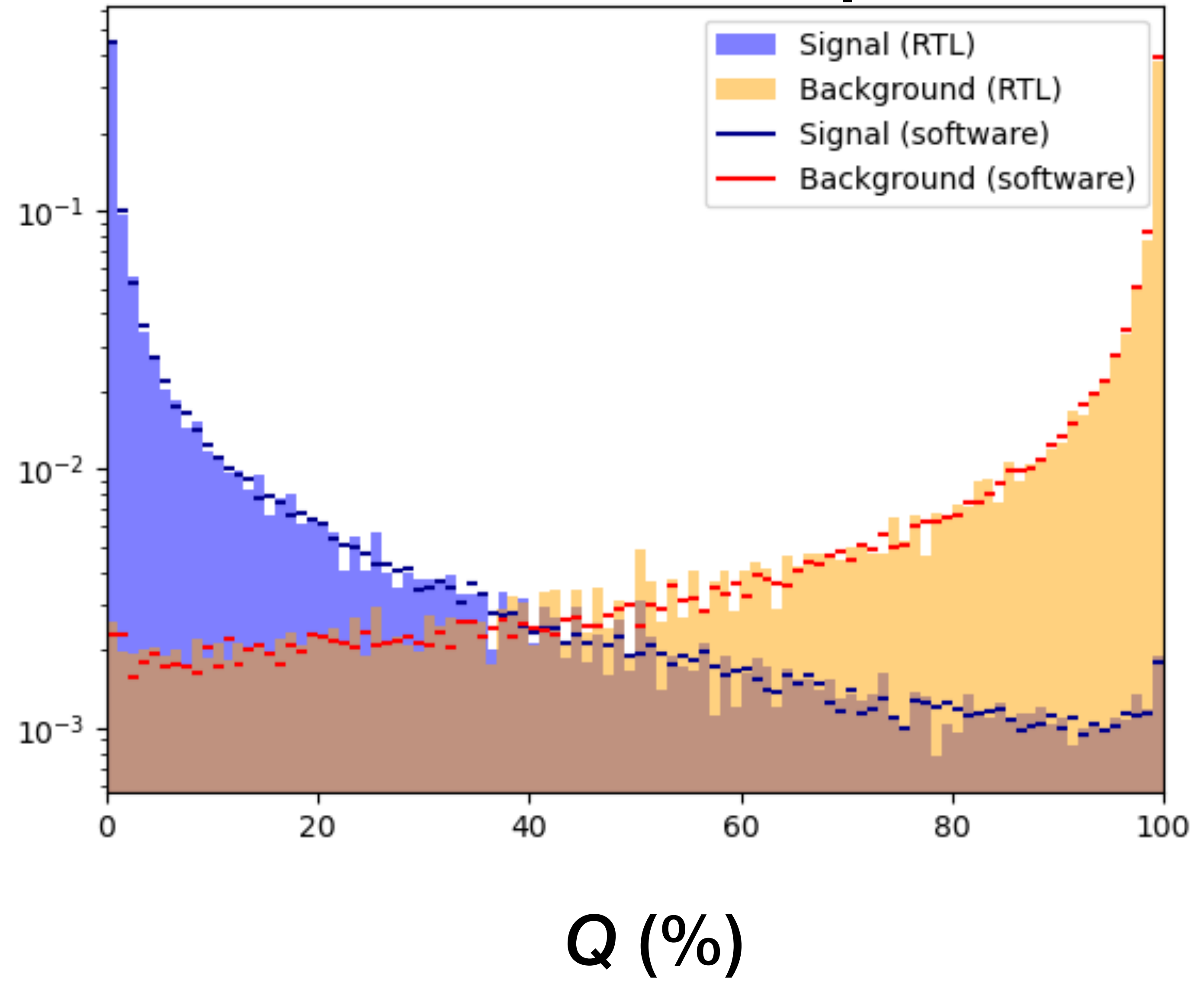
Delta track z

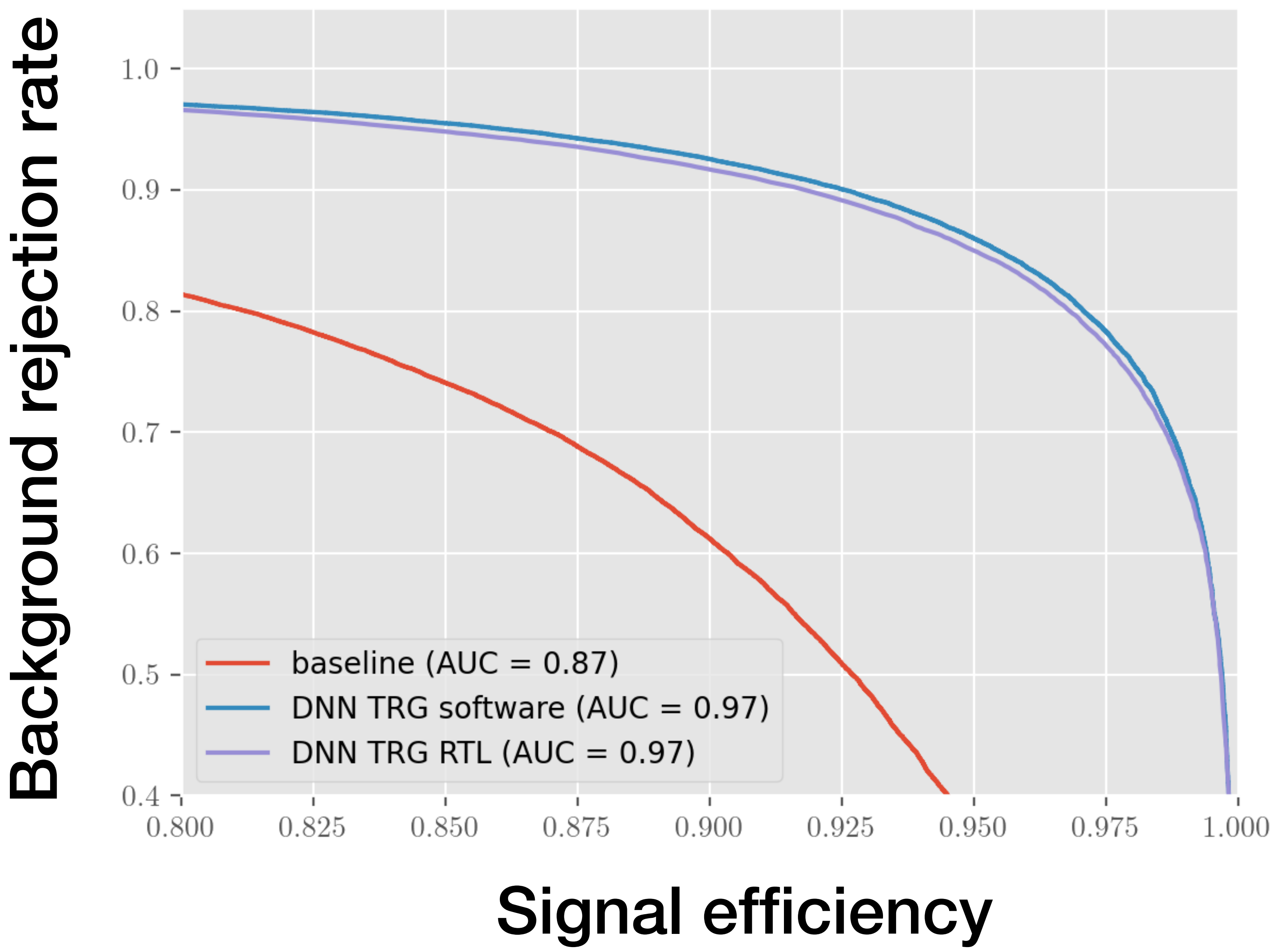


Delta track theta



Classifier output



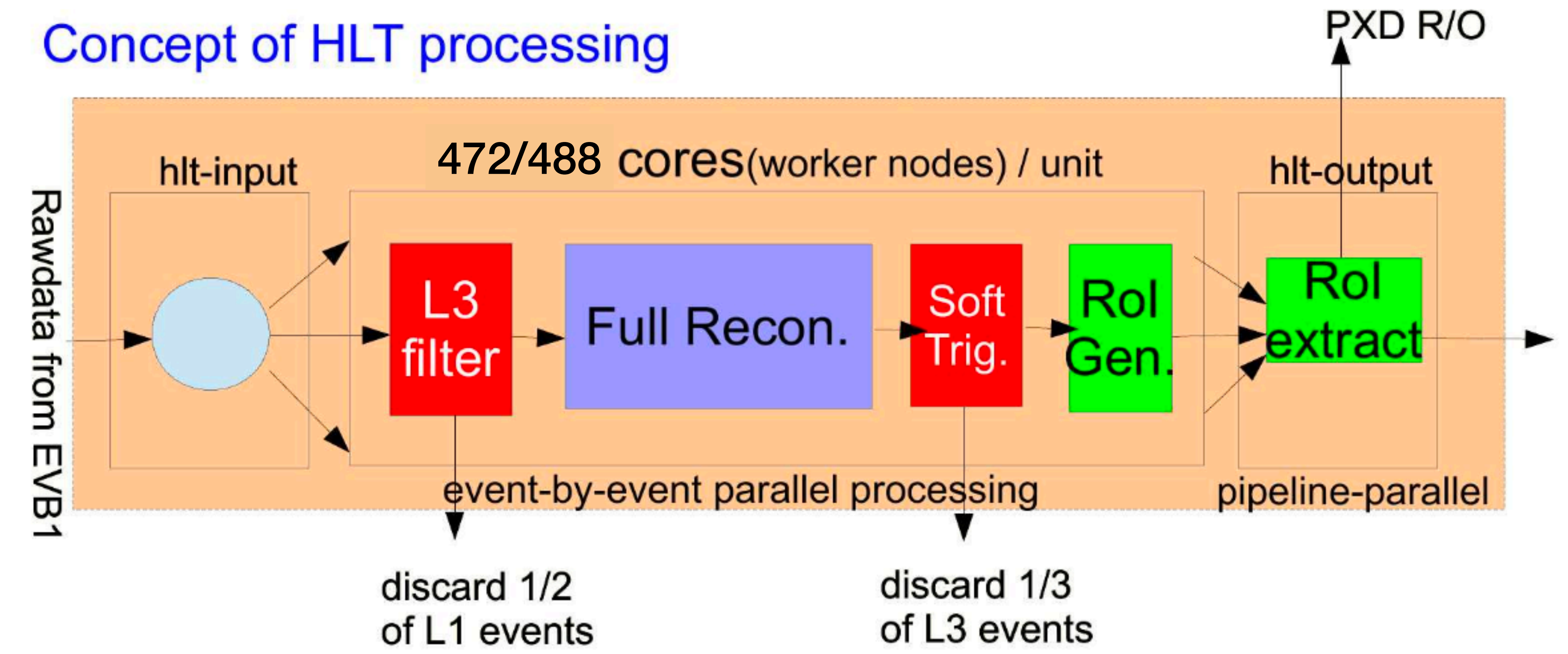


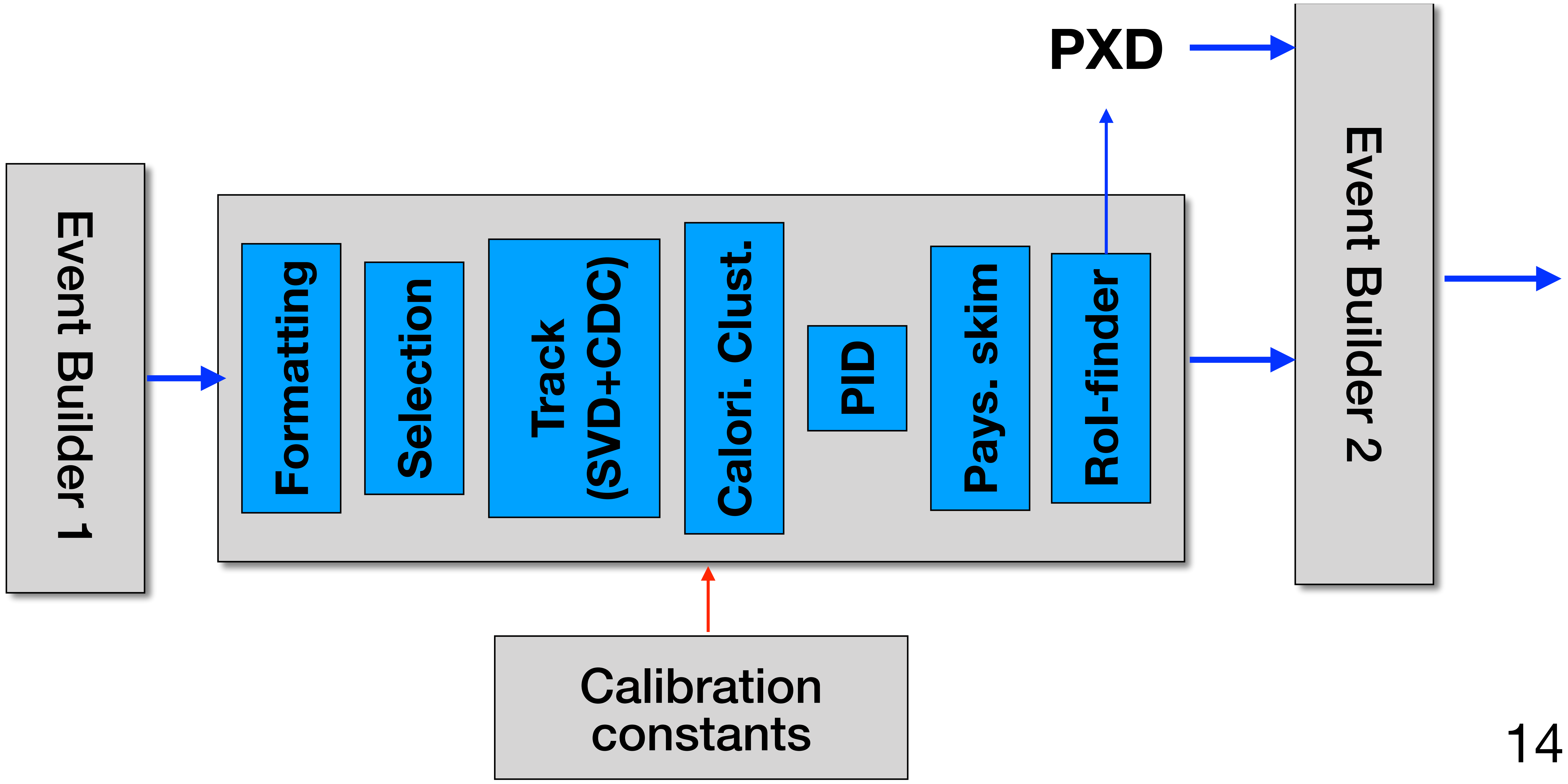
- Latency : 76 clock = 592.8 ns ;require: < 600ns
- FPGA resource (UT4: Virtex UltraScale XCVU160) usage:
  - DSP: ~70%, LUT: ~50%, others <30%
- AUC do not get large drop comparing RTL and software simulation
- At signal efficiency ~95%
  - Background rejection rate ~85%

- **DNN trigger with HARDWARE under commissioning, close to operate**

Overview of high level trigger system at Belle II

- Full event reconstruction (same as offline processing)
- Crude calibration constant
- 13 HLT units, in total ~6200 CPU cores (design: 7000 cores)
- Data processing: ~ 2.1kHz/ HLT unit w/ hyper-threading
- Event size at HLT in the last run period: ~150 kB/event
- PXD event size = 1MB/event, 10 times larger than the rest of detectors
- Region of interest (RoI) method is effective to reduce the data size
- ROI
  - **Tracking software** running on HLT nodes

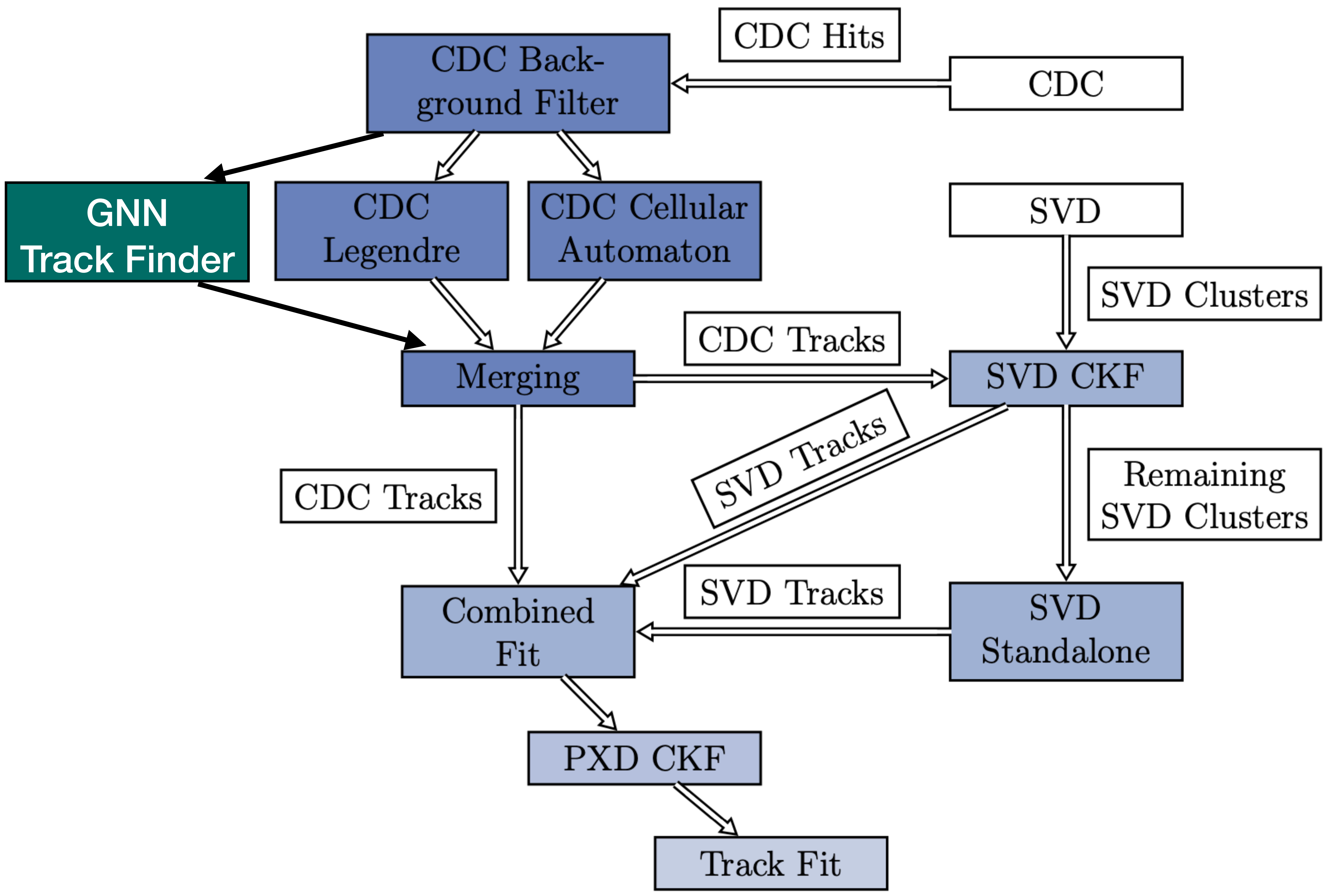




GNN based CDC track finder

- <u>Motivations of introducing a GNN track finder (**SOFTWARE**)</u>
- Low efficiency for displaced vertices
  - Efficiency decrease as displacement increase
  - Important signature for new physics search
- Higher background
- CDC wire inefficiencies
  - Bad wires or electrics
  - Decreased efficiency

Comput.Phys.Commun. 259 (2021) 107610



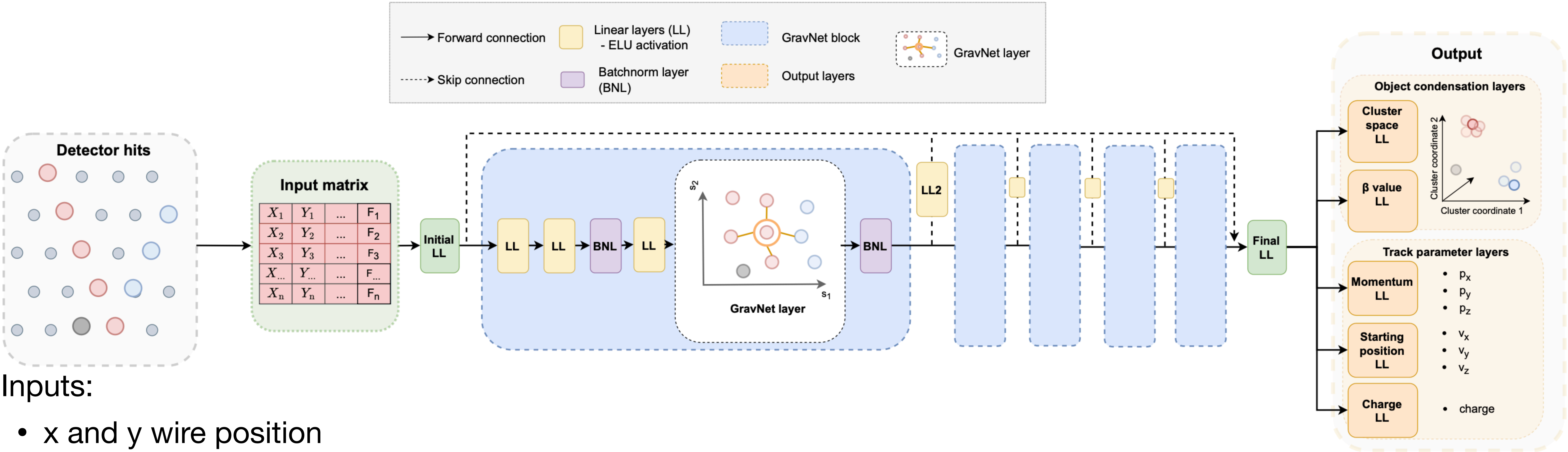
- Modular structure for track finding, with flexible of reconstruction sequence

GNN for offline track finding

- Find track parameters: momentum, starting position and charge
- Find unknown number of tracks → Object Condensation (arXiv:2002.03605)
- Computing resource and time constraint may reducible

Noise filtering → Clustering → Fitting

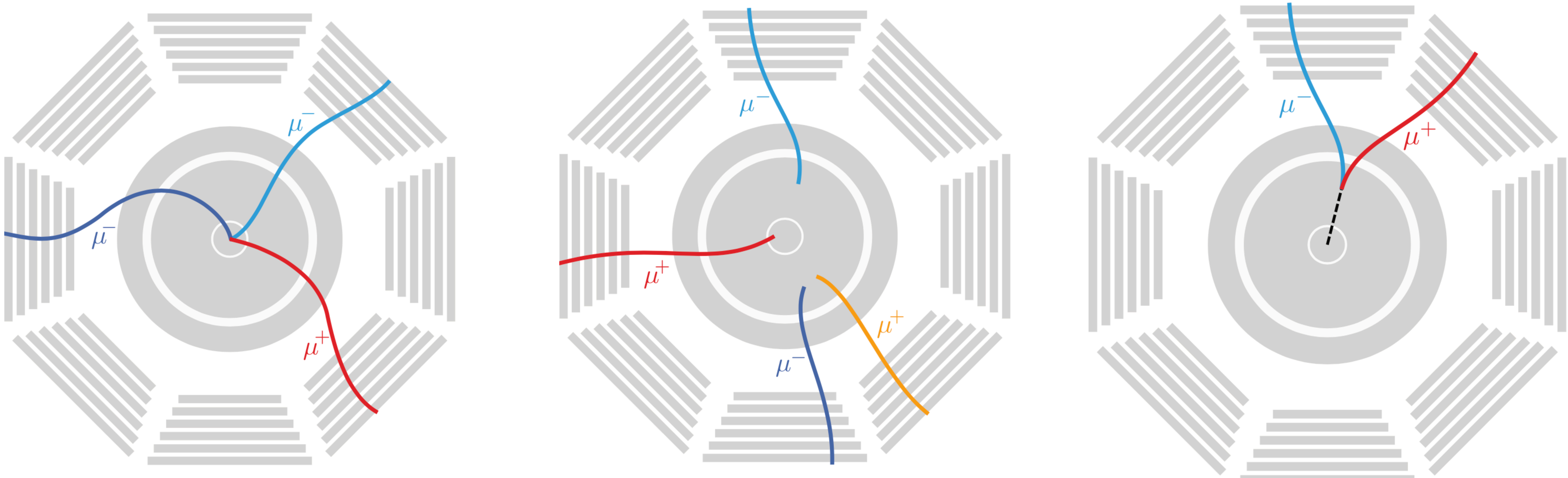
L. Reuter et. al. (KIT) BELLE2-NOTE-TE-2024-008



- Inputs:
  - x and y wire position
  - TDC and ADC of signal information
  - layer, superlayer, and layer info. with suprlayer
- Adjustable Parameters
  - 797,812 trainable parameters (3MB weight files)

Training of GNN

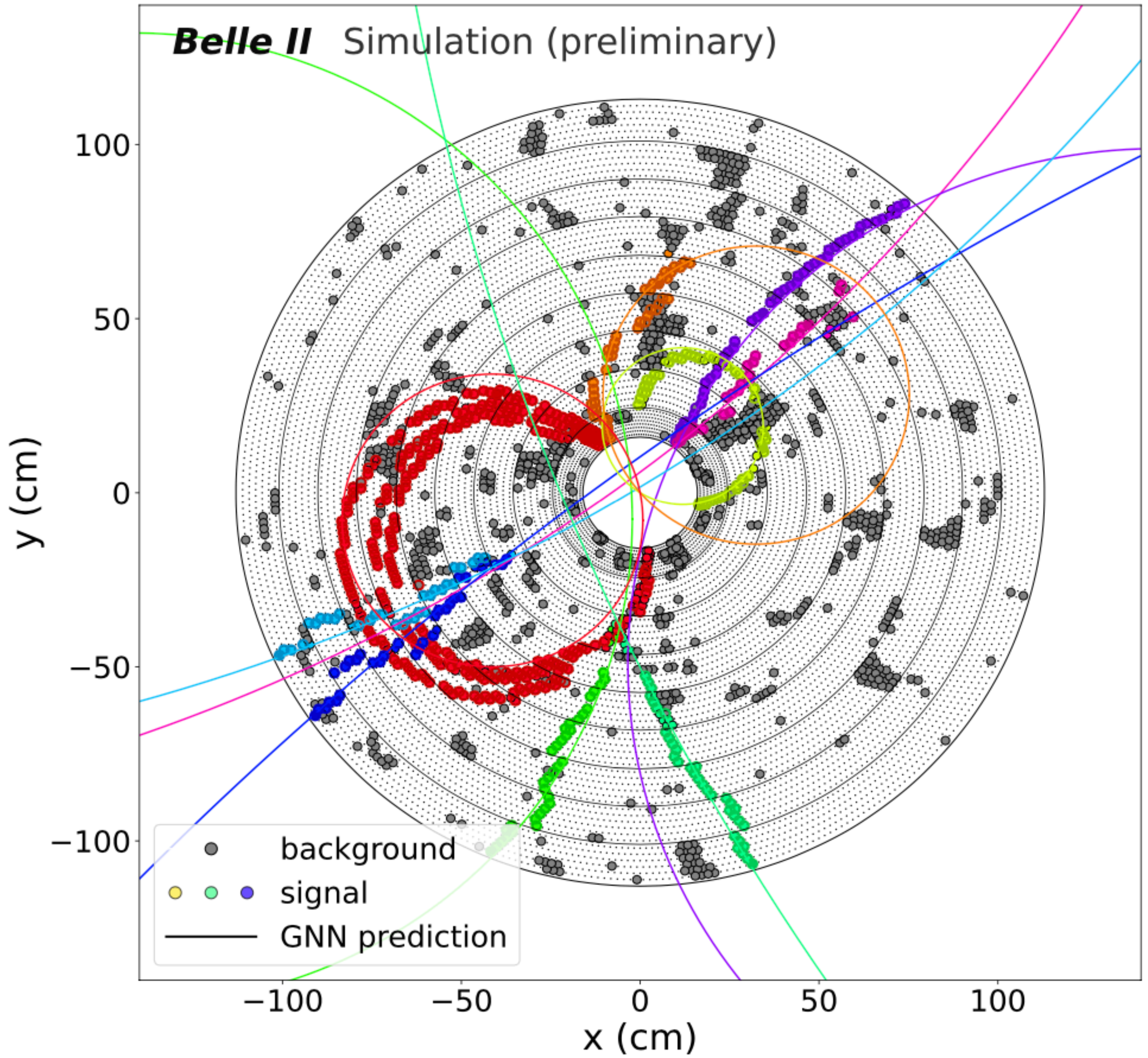
- Simulate 1 million events with over 4 million tracks
  - Train: Validation = 4 :1
- Training samples contain different topologies that cover all interested event features, to not bias the model, **no conservation laws involved here!**
  → crucial step to be agnostic about the physics processes
- Sample features
  - Low momentum tracks forming circles in the CDC ($P_t < 0.4$ GeV) <-> High momentum tracks
  - Short tracks <-> tracks penetrate all CDC layers
  - Small opening angle <-> well isolated two tracks
  - …



Performance of GNN

- Efficiency of displaced vertex tracks improved from 85.4% with a fake rate of 2.5%, compared to 52.2% and 4.1%
  - The other performance similar as original algorithm
- Momentum p_x , p_y , p_z starting position v_x , v_y , v_z ,charge
  - Provide initial inputs for GENFIT
- GNN prediction is drawn according to the track parameters predicted by the GNN
- Plan to added as additional track finder for Belle II

L. Reuter et. al. (KIT) BELLE2-NOTE-TE-2024-008



Motivations of trigger-DAQ upgrade

Physics

- Tau trigger efficiency now is >95% (to be pre-scaled if luminosity is high)
- Low multiplicity trigger efficiency (to be pre-scaled pre-scaled if luminosity is high)
- Low-momentum track trigger efficiency
- “Anomaly” trigger
  - Design a special trigger line for some specific physics channel
- Trigger efficiency of displaced vertex

Current hardware limitation:

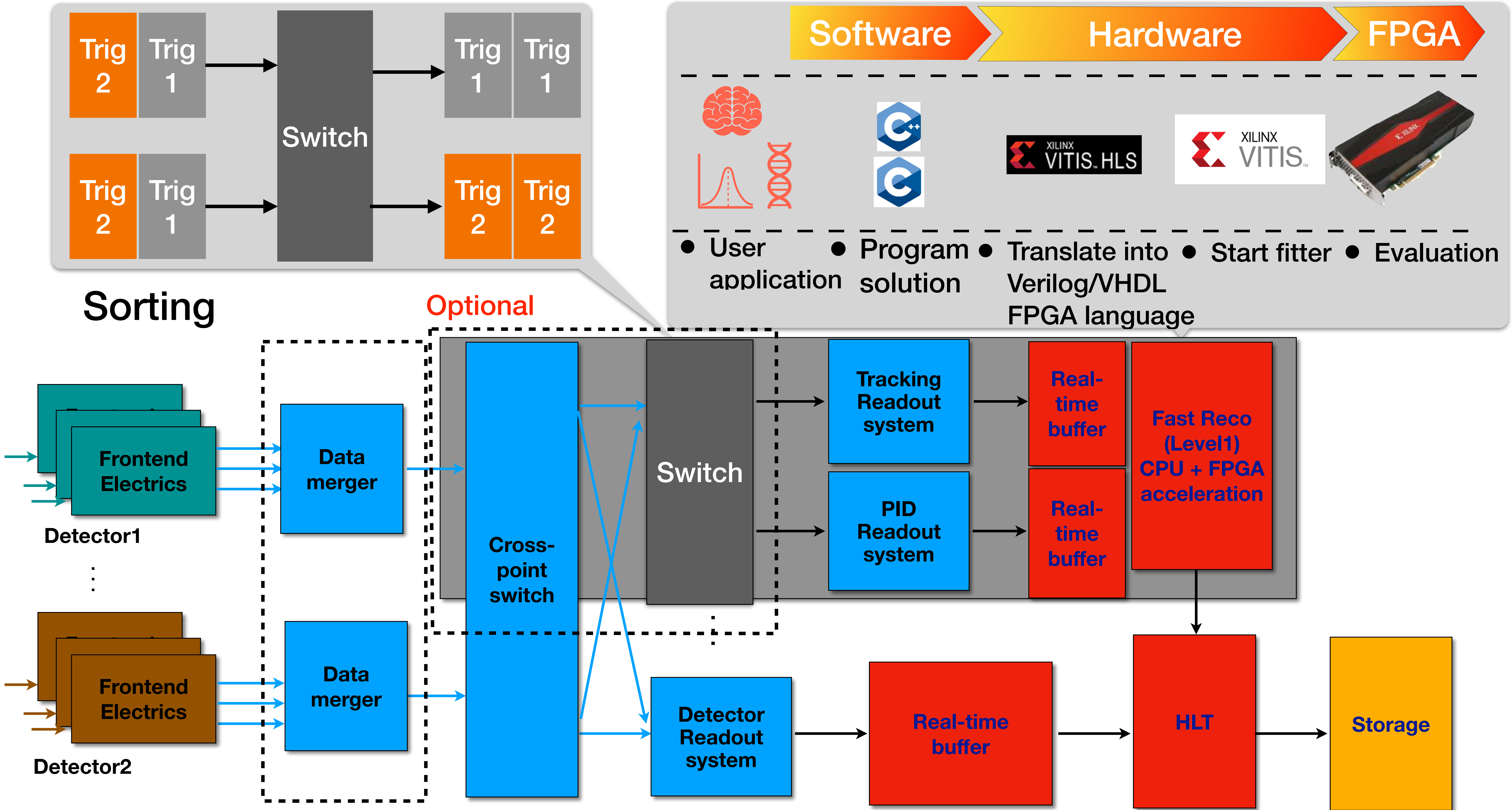
- DAQ system is designed to handle 30 kHz
  - L1 latency 4.4 us (SVD APV25 buffer)
    - CDC DNN trigger latency ~500 ns, latency already limited more large model
- L1 trigger rate will reach to ~20 kHz at 0.9×10^{-35} cm^{-2} s^{-1} (13 HLT units, w/o hyperthreading), planed full HLT: 15 units (7000 CPU cores)
- TTD system: VME bus limit, no more than 3 triggers within 80 clock (624ns)

Vertex detector is planed to be upgraded during long shutdown 2 (after 2028)

- Latency limit target: 5 us -> 10 us (5.2 us KLM, 9 us TOP, considering upgrade)
- New TTD hardware: VME bus -> Ethernet
- New trigger board (UT5): Versal ACAP

Idea of upgrade trigger and DAQ system

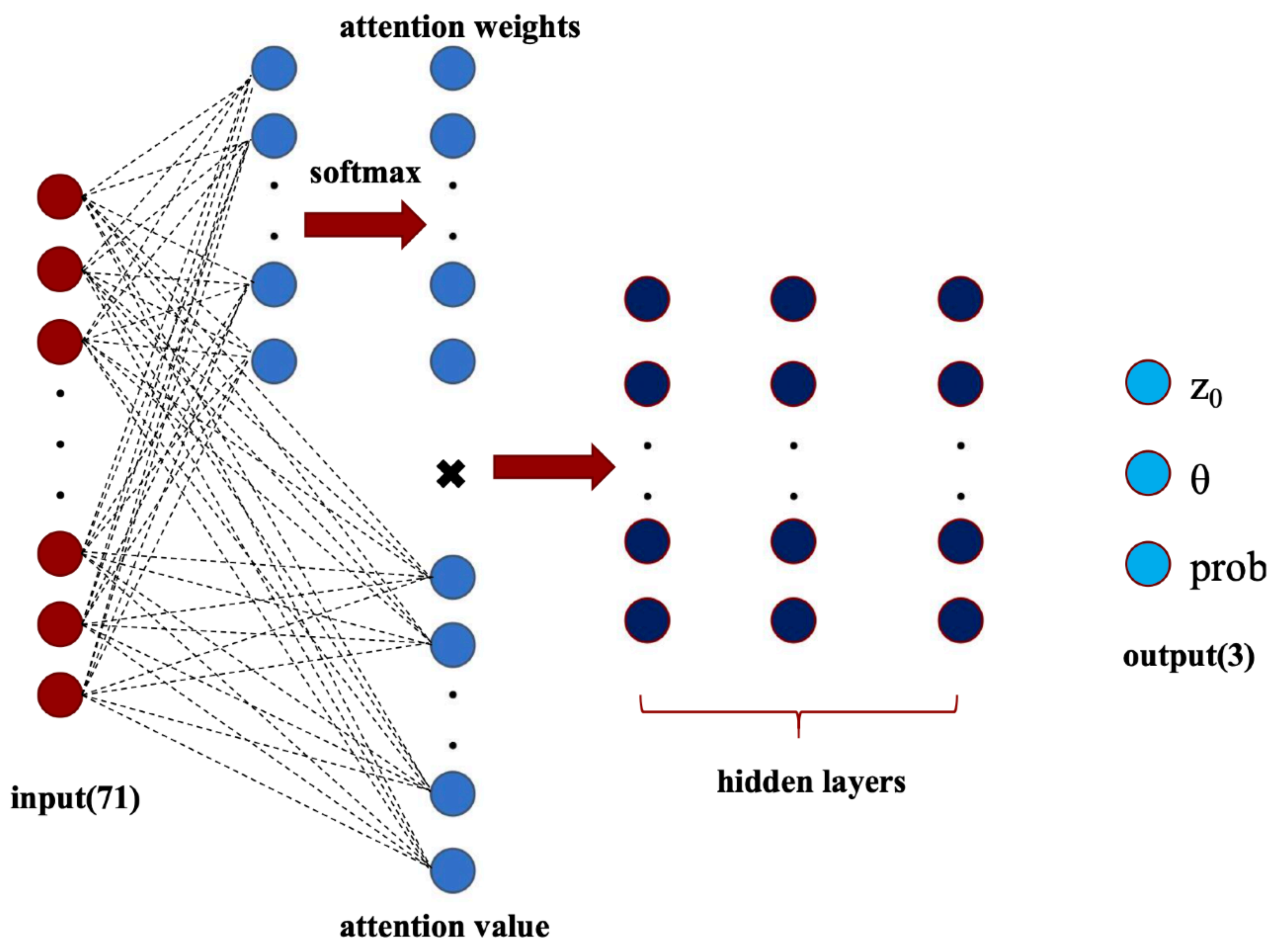
Proposed on Belle II trigger DAQ workshop 2022

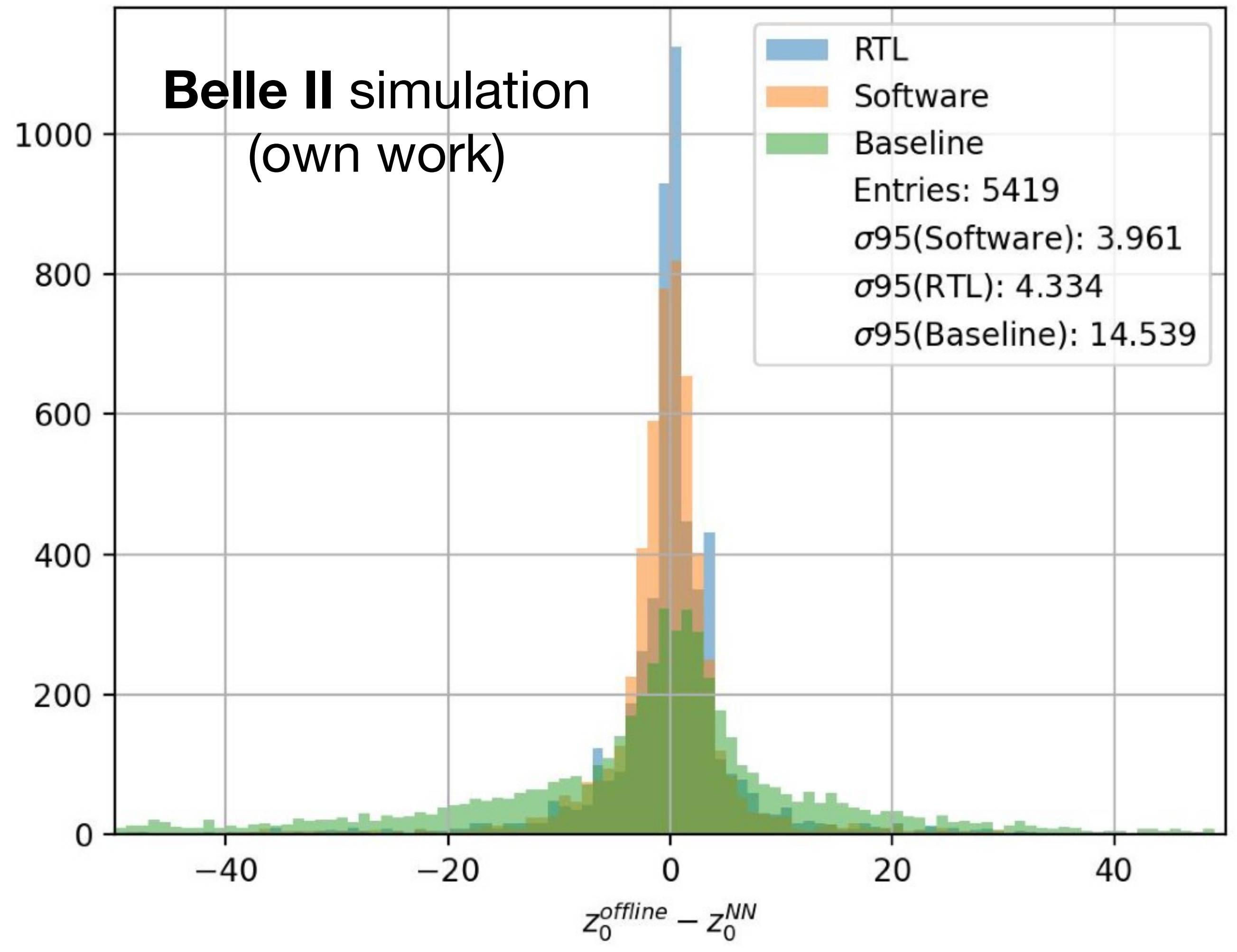


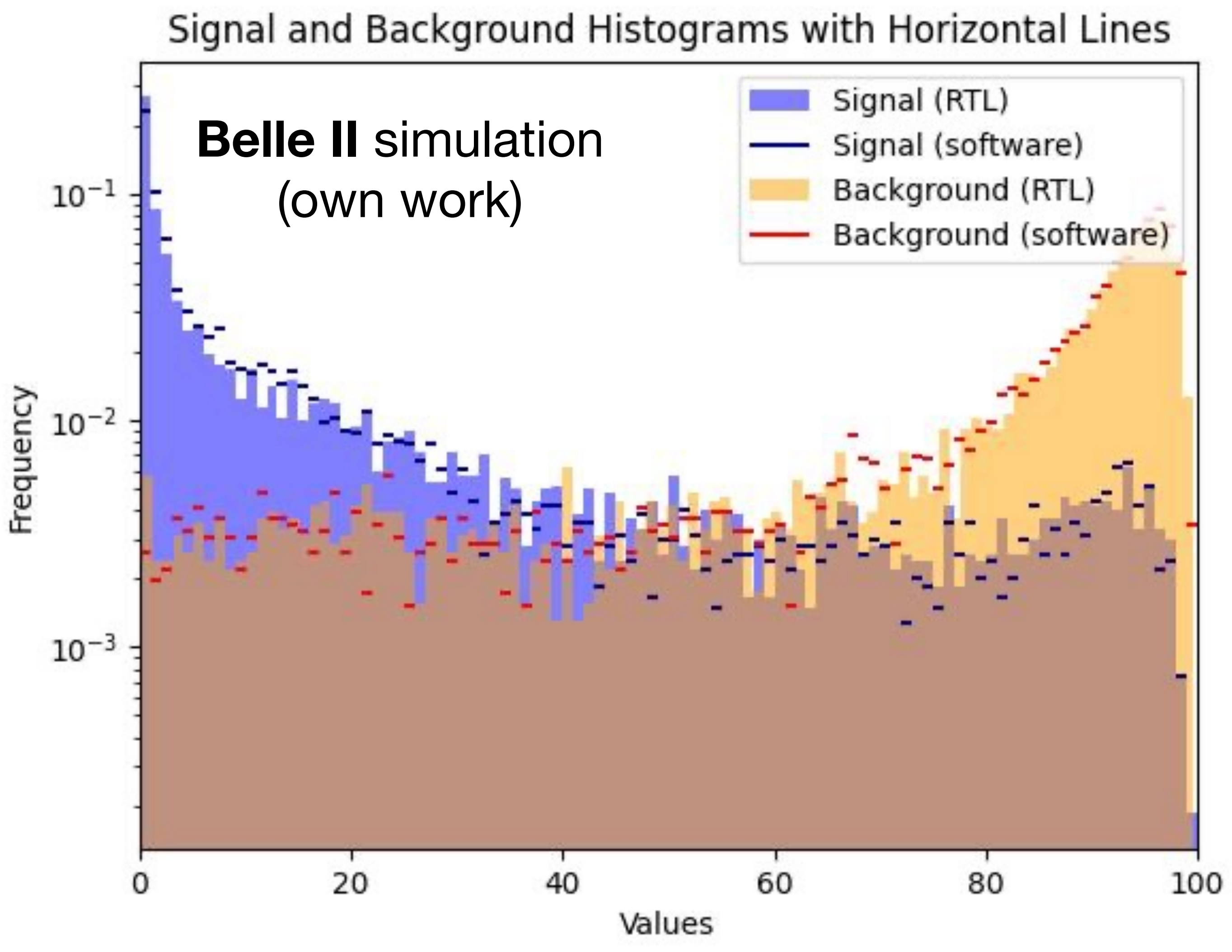
Improvement try for CDC track trigger

- Develop a algorithm improve the performance for the upgrade (10 usec latency)
  - Start from optimization of DNN model
- Modify the number of hidden layers and learning rate
  - Hidden layer: 2 -> 4, learning rate: $1e^{-2}$ -> $1e^{-3}$
  - Others keep the same
  - No improvement
  - Latency: 76 clock (592.8 ns) -> 82 clocks (640 ns)
- Next step, change the inputs (CDC hits info.), instead of 2D track parameters

Shangshang Zhang (SDU)



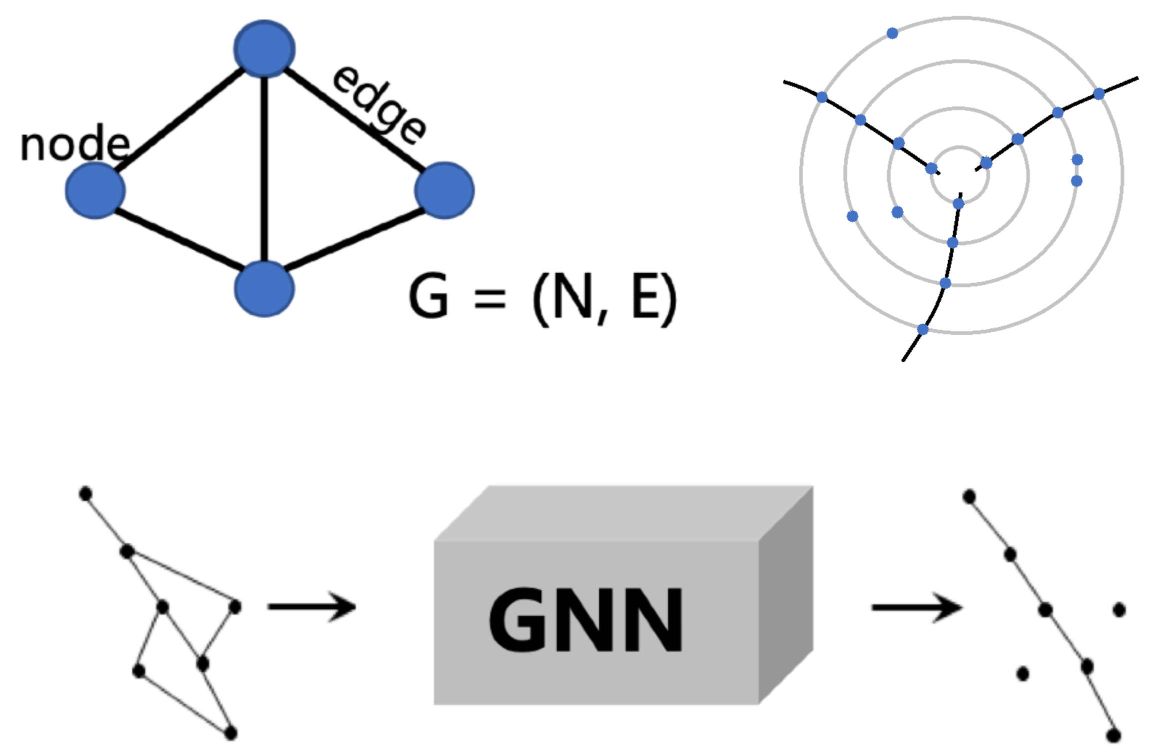


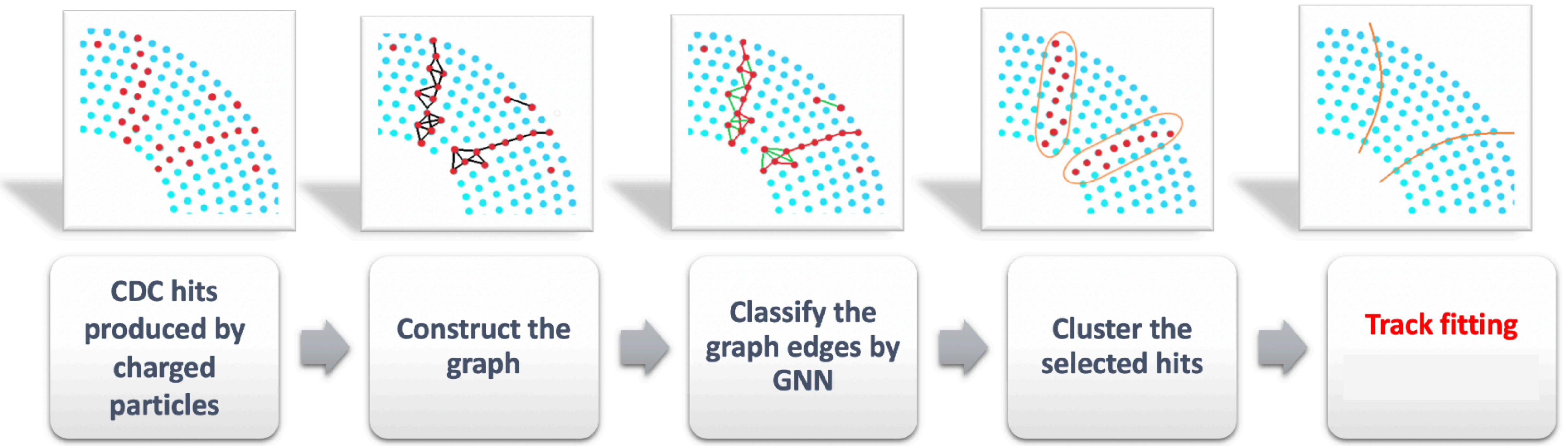


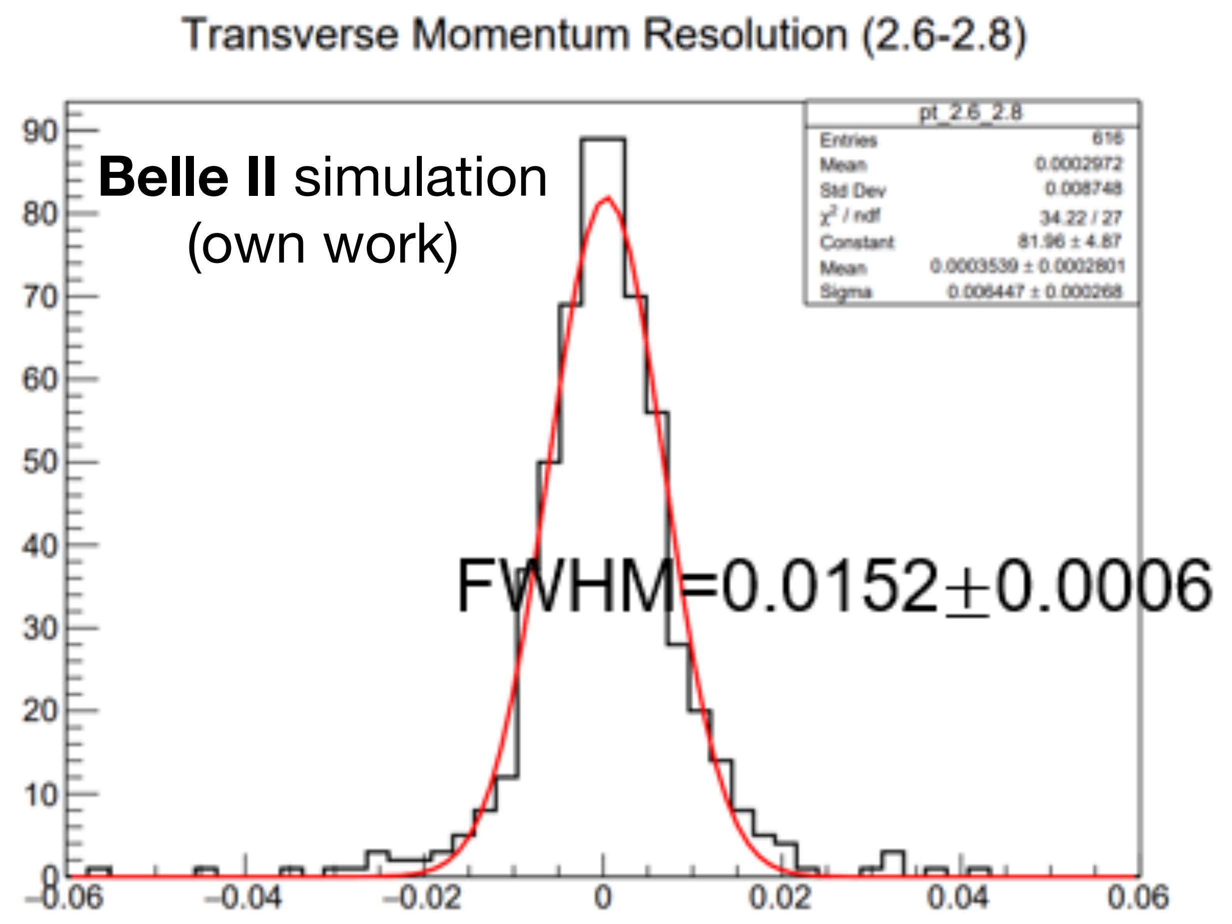
GNN for CDC track background filtering

- Developed a GNN algorithm (based on X. Q. Jia (SDU) et al. BESIII's algorithm) for Belle II CDC hits clean up
  - Inputs: TDC, position coordinates r, ϕ

Xiaoqian Hu (SDU)

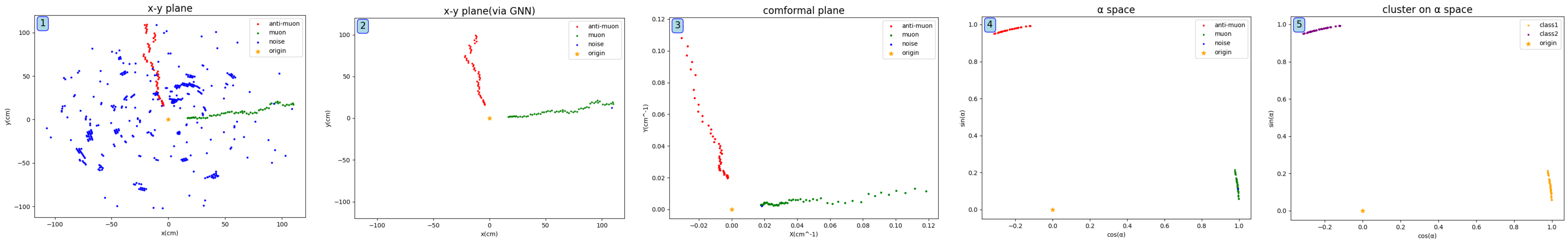






$(P_t^{pred} - P_t^{MC})/P_t^{MC}$

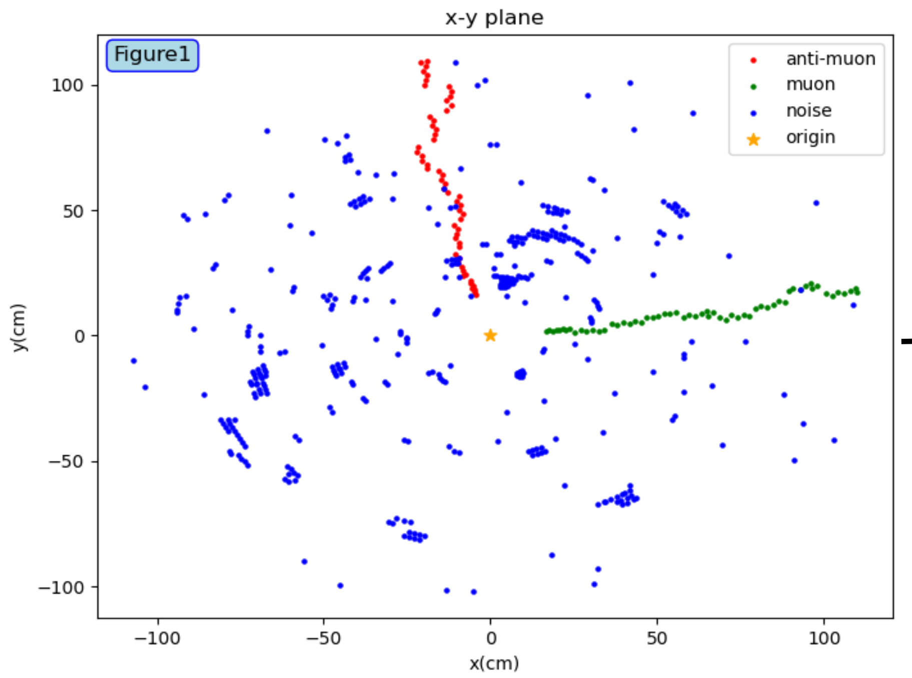
Belle II simulation (own work)



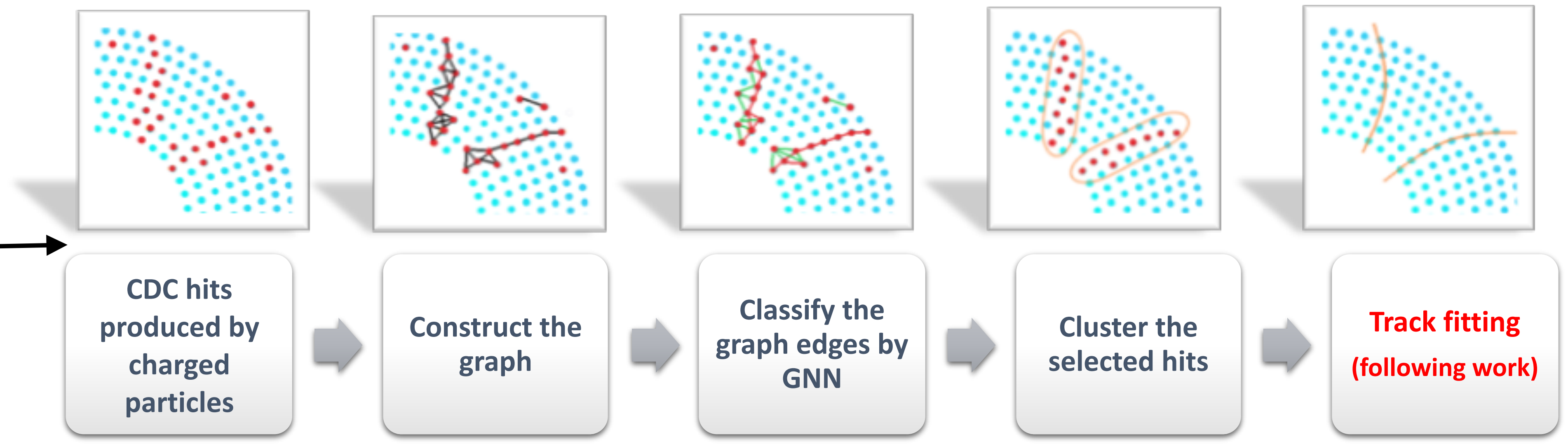
μ+ μ- (particle gun) | GNN noise filtering | Transform space | Transform α space | DBSCAN clustering

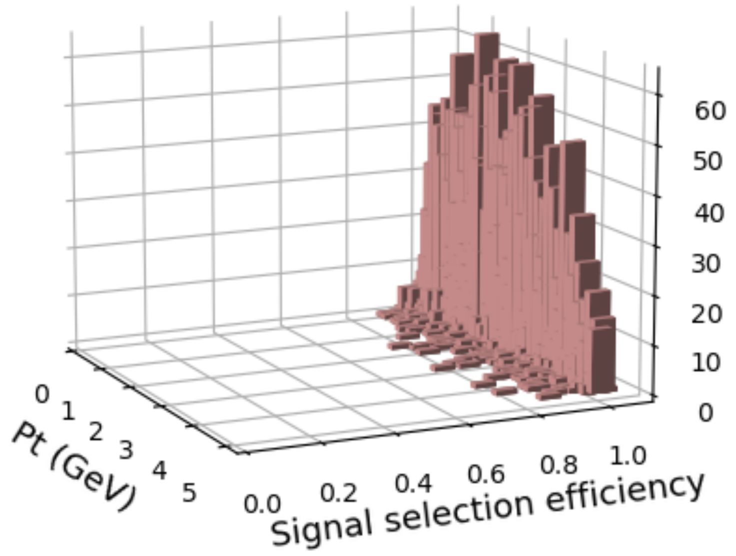
Acceleration on Versal ACAP platform

μ+ μ- (particle gun)



Belle II simulation (own work)



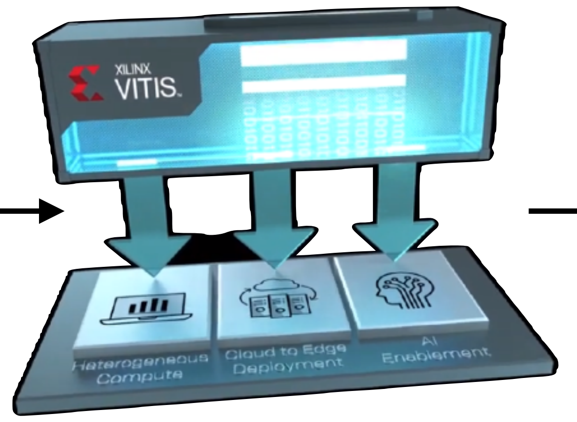


Hit selection efficiency: 98.4%

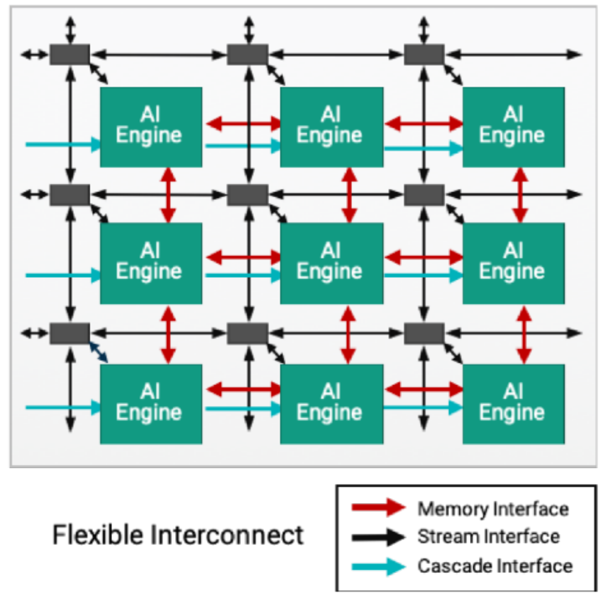
Hit selection purity : 97.9%

AI engine

C++ software



VIVADO



DPU

ML

TensorFlow

QKeras

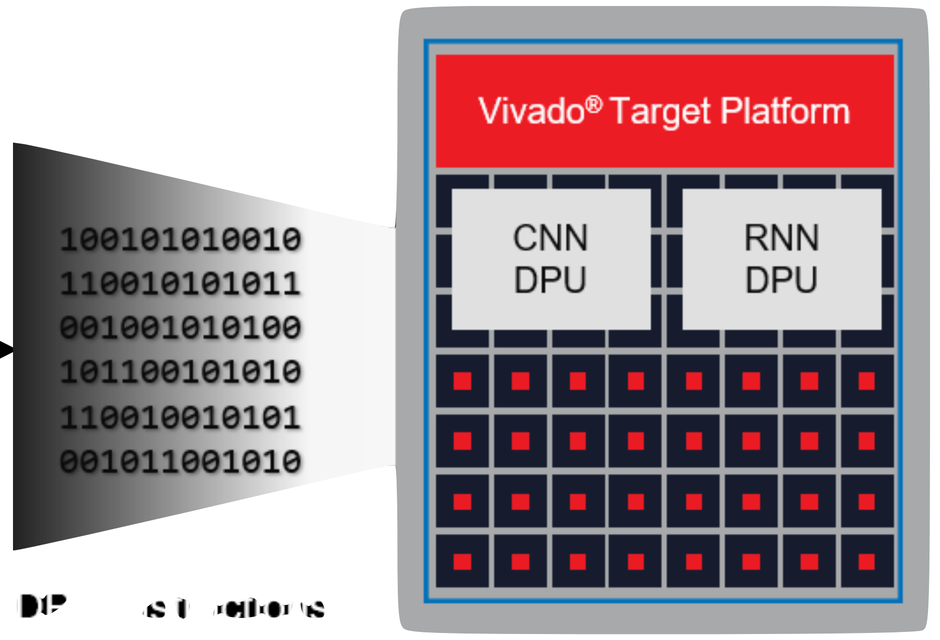
Keras

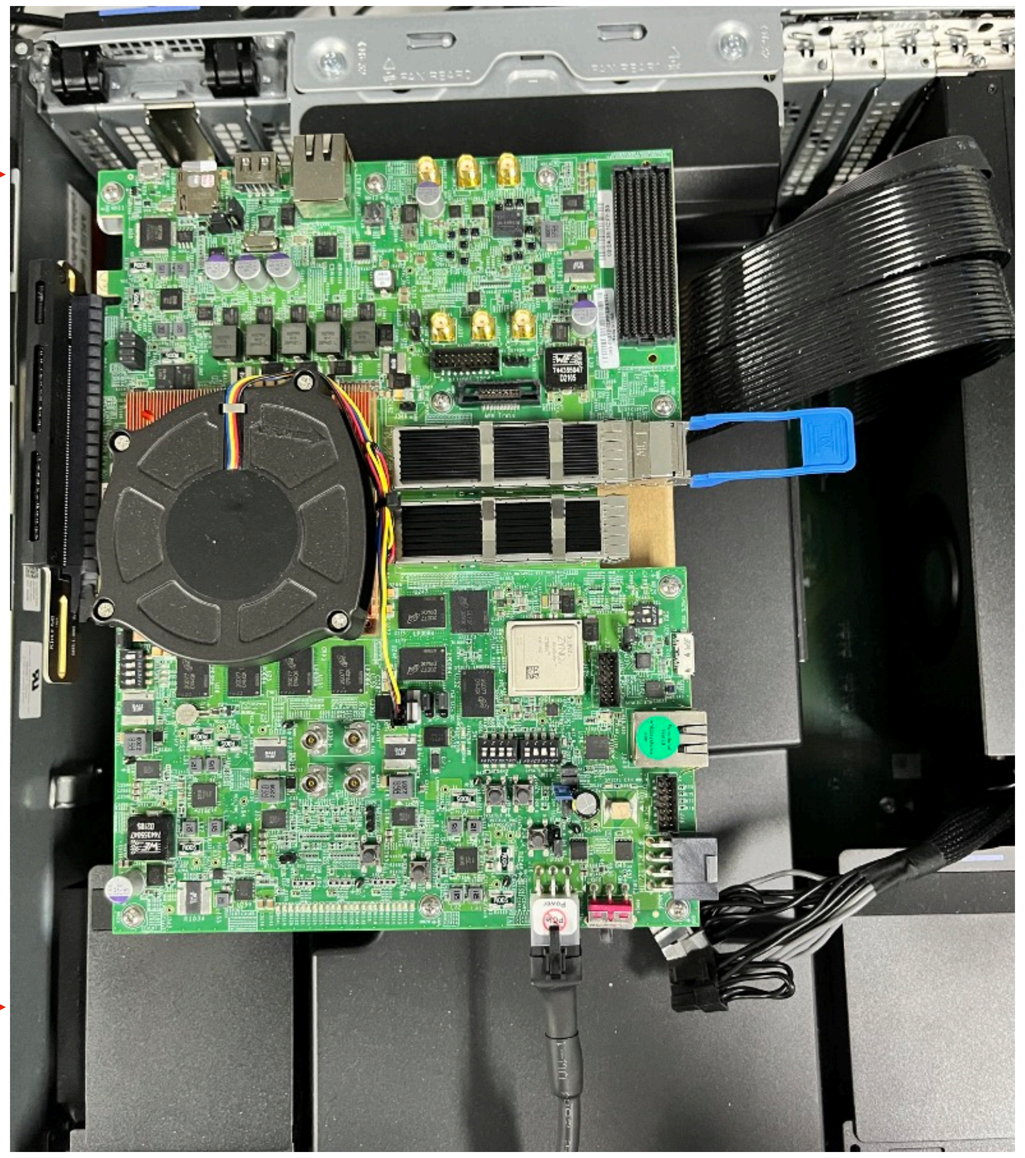
PyTorch

C++ software

Vitis AI

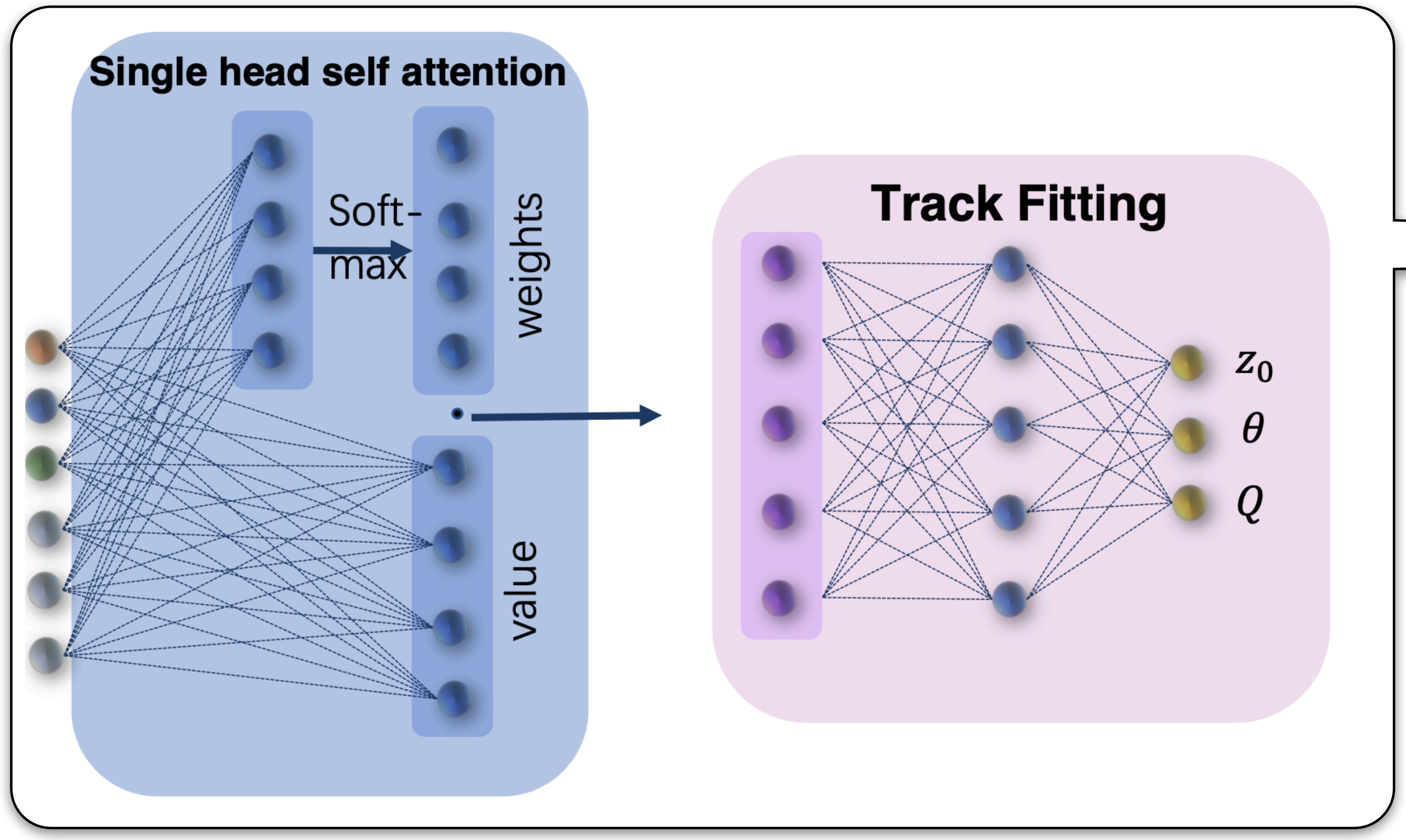
XILINX VITIS AI

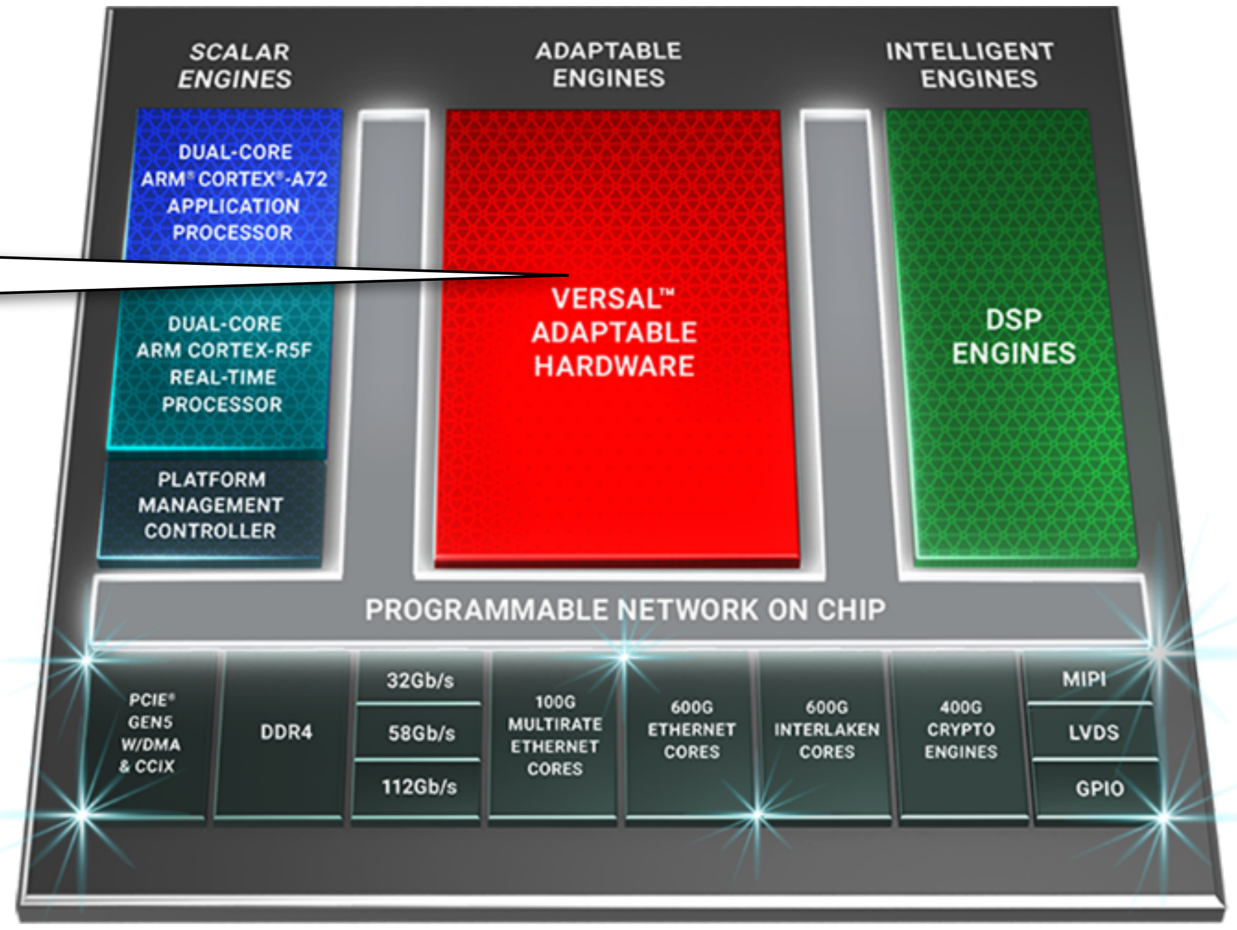


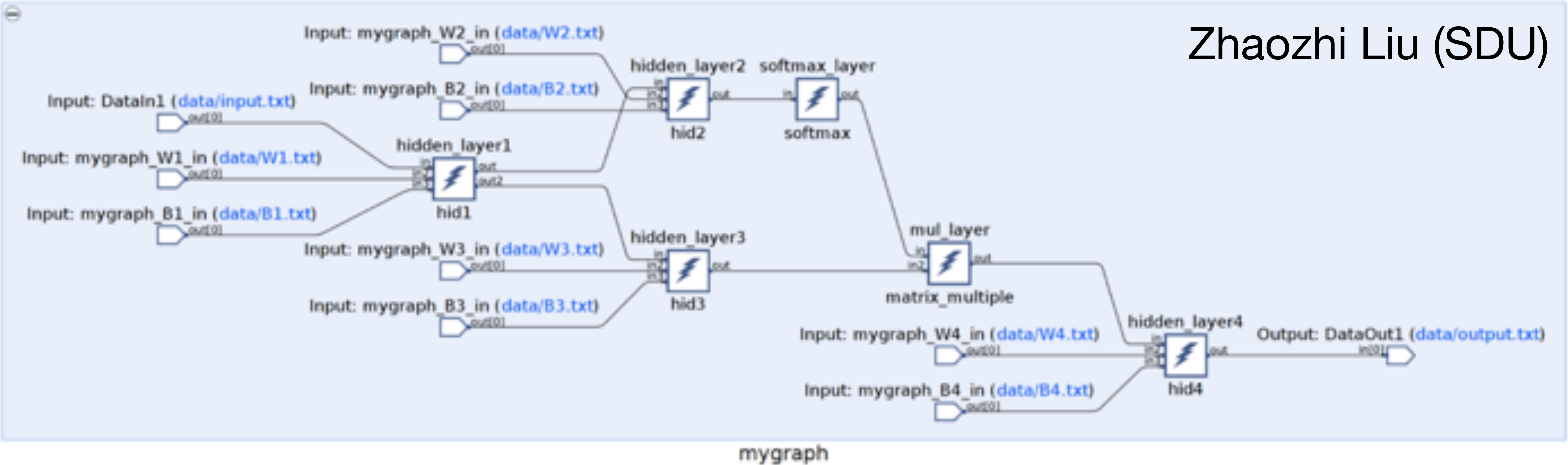


NN acceleration on Versal ACAP

- Real-time graph building algorithm enables GNN implementation on FPGA for Belle II
  M. Neu et al. Comp. Soft. BigSci. 8, 8(2024)
- R&D of a new general FPGA device using the Versal ACAP
  - Heterogenous acceleration (VCK190, VCK5000 evaluation kit)
    - AI engine, DPU







Zhaozhi Liu (SDU)

Summary and prospects

- Belle II TDAQ system was designed to handle 30 kHz level 1 trigger
- NN and DNN with hardware based CDC L1 track trigger to improve background rejection
- GNN with software based offline CDC track finder to improve the efficiency of displaced vertex tracks
- Not covered in the talk: GNN with hardware based clustering trigger for Belle II is under commissioning
- Upgrade of vertex detector, trigger board, possible detector electrics during LS2
    - Idea about upgrade of trigger and DAQ system
    - ML based developments on going for the upgraded system