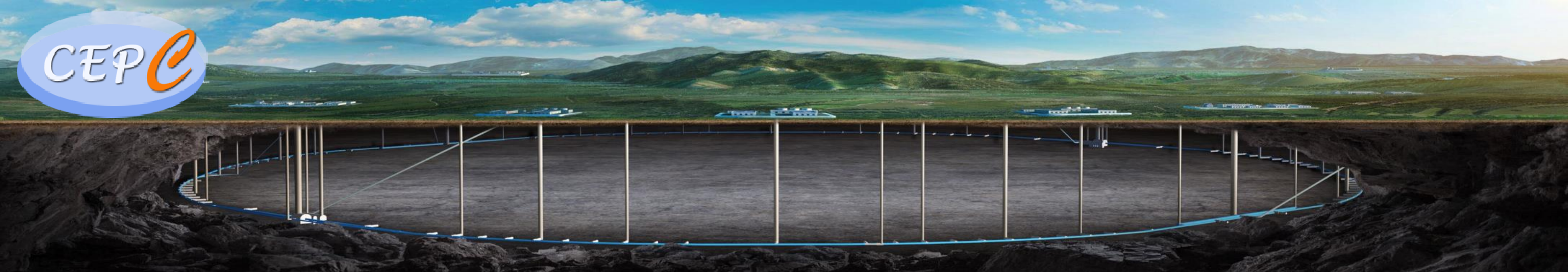


CEPC TDAQ and Online

Fei Li, Jingzhou Zhao and Xiaolu Ji
On behalf of CEPC TDAQ Group

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Content

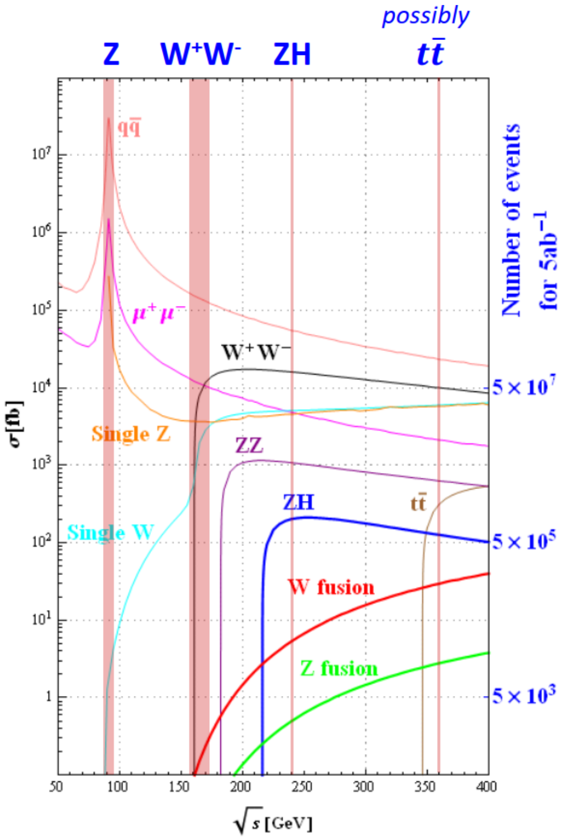
- Introduction
- Requirements
- Technology survey and our choices
- Technical challenges
- Previous experience on large facilities
- R&D efforts and results
- Detailed design
- Research team and working plan
- Summary

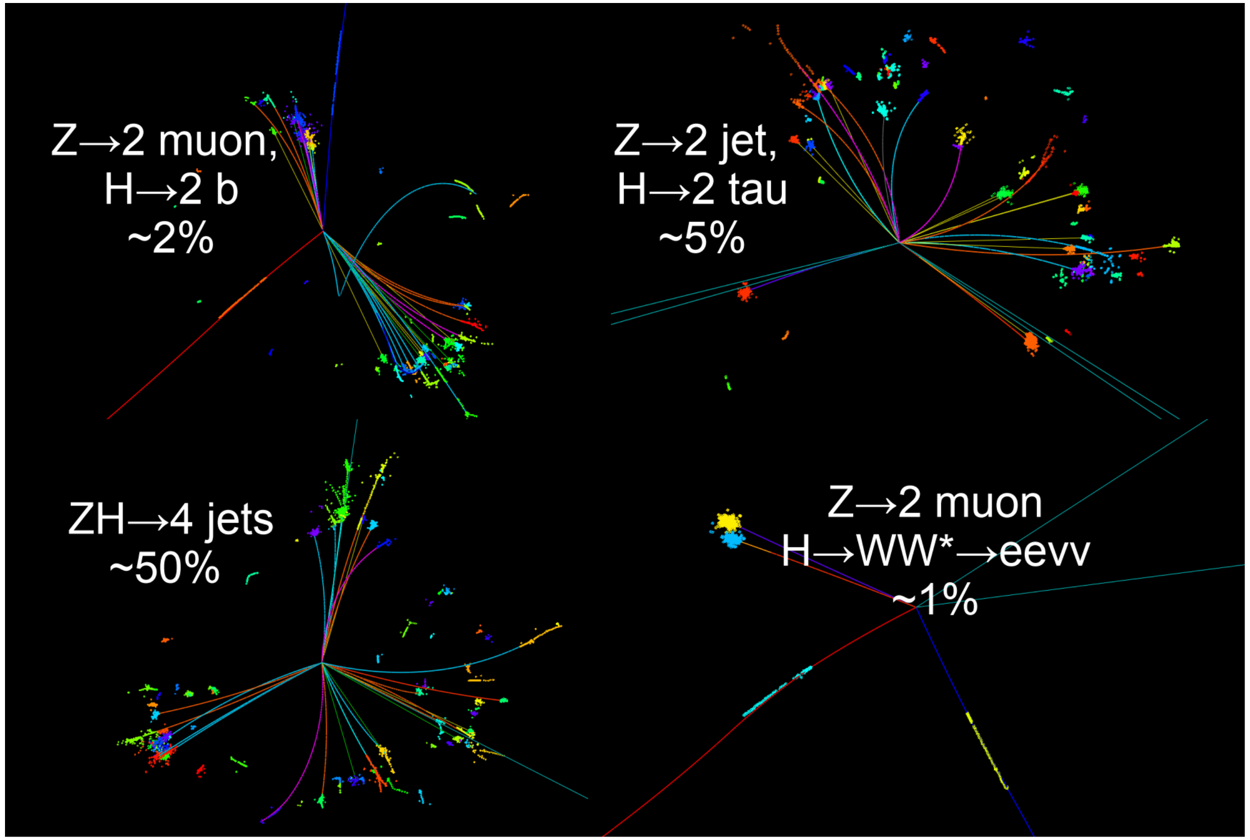
Introduction

- This talk is about the design and development of the TDAQ
- This talk relates to the Ref-TDR Ch 12.
- Questions to physics and simulation
  - What kind of events need to be saved?
  - How to identify these events?
  - What level of background?
- Questions to each detectors and electronics
  - How many raw data need to readout?
  - Whether a hardware trigger is required?
  - If hardware trigger, how fast a latency is acceptable?
  - What trigger primitive information can be provided
  - What level of noise? Signal vs noise occupancy
  - What slow control and monitoring requirements?

Introduction

- CEPC beam energy, collision period, luminosity:
  - Z(91.2 GeV, 23ns, 115E34) , W (160 GeV, 257ns, 16E34) and Higgs (240 GeV, 591ns, 5E34)
- Requirements for rough selection of the relevant objects (jet, e, muon, tau,v, ...) and combinations.





Ref: CEPC Physics at a glance, Lomonosov Conference 2021, by Manqi Ruan

TDR Outline

- **Introduction**
- **Requirements and design considerations**
  - Physics requirements for trigger
  - Trigger requirements for sub-detectors
  - Consideration on readout strategy
    - Trigger readout-on-FEE vs. Trigger readout-on-BEE
    - Main constraint on FEE triggerless readout vs. CEPC's data rate
  - Consideration of the readout-interface for electronics
  - Event rate estimation & background rate estimation
- **Technology survey and our choices**
  - Consideration on Backend Trigger strategy (Hardware Trigger vs. Software Trigger)
  - Consideration on high level trigger algorithm & resources
- **Trigger**
  - Previous experience on large facilities
  - Previous R&Ds
  - Common Electronics interface
  - Structure of the Trigger for CEPC
  - Common Trigger Board
  - Resource cost estimation

- **DAQ**
  - Previous experience on large facilities
  - Previous R&Ds
  - Platform for DAQ and computing
  - Algorithm & architecture
  - Resource cost estimation
- **Detector Control System**
  - Requirements on sub-detectors
  - On-detector monitoring consideration
  - On-detector slow control consideration
  - Electronics monitoring and control consideration
- **Experiment Control System**
  - Requirements
  - Network
  - Counting room
- **Summary**
  - Summary on data volume
  - Summary on cost

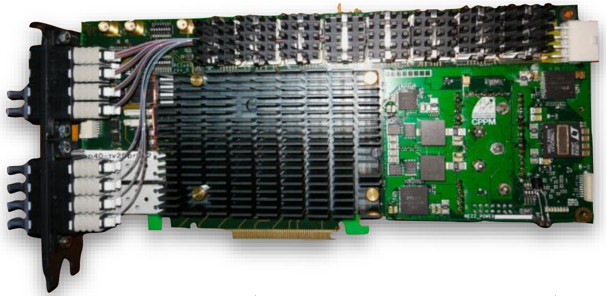
Requirement

- 7 Sub detectors
- Raw data rate before trigger
  - 4.42Tbps, 553GB/s
    - @ low Lumi Z
  - 22.1Tbps, 2.76TB/s
    - @ high Lumi Z
    - 5 times increase
  - **Key issue: FEE readout bandwidth per chip**
- Trigger and Online processing
  - Good event rate
    - <300kHz@Z
  - Hardware & software
    - Event filter
    - Data compression
  - Trigger efficiency
  - Event purity

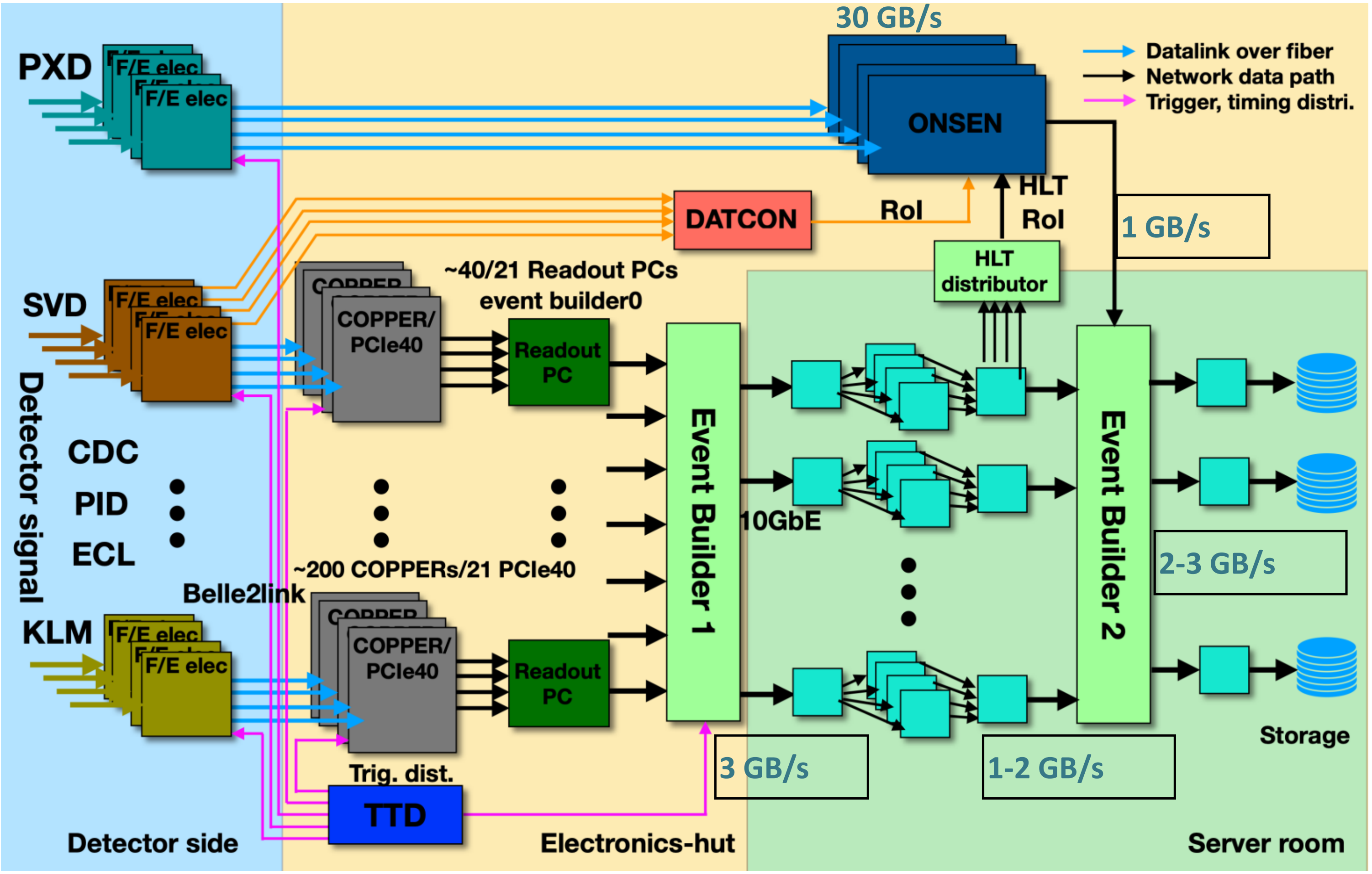
	Vertex	Pix (ITKB)	Strip (ITKE)	TOF (OTK)	TPC	ECAL	HCAL
Channels per chip	512*1024 Pixelized	512*128 (2cm*2cm@34um*150um)	512	128	128	8~16	8~16
Ref. Signal processing	XY addr + BX ID	XY addr + timing	Hit + TOT + timing	ADC+TDC/TOT+TOA	ADC + BX ID	TOT + TOA/ ADC + TDC	TOT + TOA/ ADC + TDC
Data Width /hit	32bit	48bit	32bit	40~48bit	48bit	48bit	48bit
Data rate / chip	1Gbps/chip @Triggerless @Low LumiZ Innermost	640Mbps/chip Innermost	Avg. 1.01MHz/chip Max. 100MHz/chip	Avg：26kHz/chip @ z pole Max：210kHz/chip @z pole	~70Mbps/ modu Inmost	<4.8Gbps/module	<4.8Gbps/module
Data aggregation	10~20:1, @1Gbps	1. 1-2:1 @Gbps; 2. 10:1@O(10Gbps)	1. 10:1 @Gbps 2. 10:1 @O(10Gbps)	1. 10:1 @1Mbps 2. 10:1 @O(10Mbps)	1. 279:1 FEE-0 2. 4:1 Module	1. 4~5:1 side brd 2. 7*4 / 14*4 back brd @ O(10Mbps)	< 10:1 (40cm*40cm PCB – 4cm*4cm tile – 16chn ASIC)
Detector Channel/ module	2218 chips @long barrel	30,856 chips 2204 modules	22720 chips 1696 modules	41580 chips 1890 modules	258 Module	1.1M chn	6.7M chn
Data Volume before Trg	2.2Tbps	2 Tbps	22.4 Gbps	1 Gbps	18Gbps	164.8Gbps	14.4Gbps

Technology survey

- Belle II TDAQ
  - 30 kHz level 1 trigger
  - 4.5us L1 latency
  - PCIe card readout (except for PXD)
  - Buffer PXD data at ONSEN
    - Read out by HLT RoI
    - Gen. RoI by SVD track



PCIe40(LHCb, ALICE)



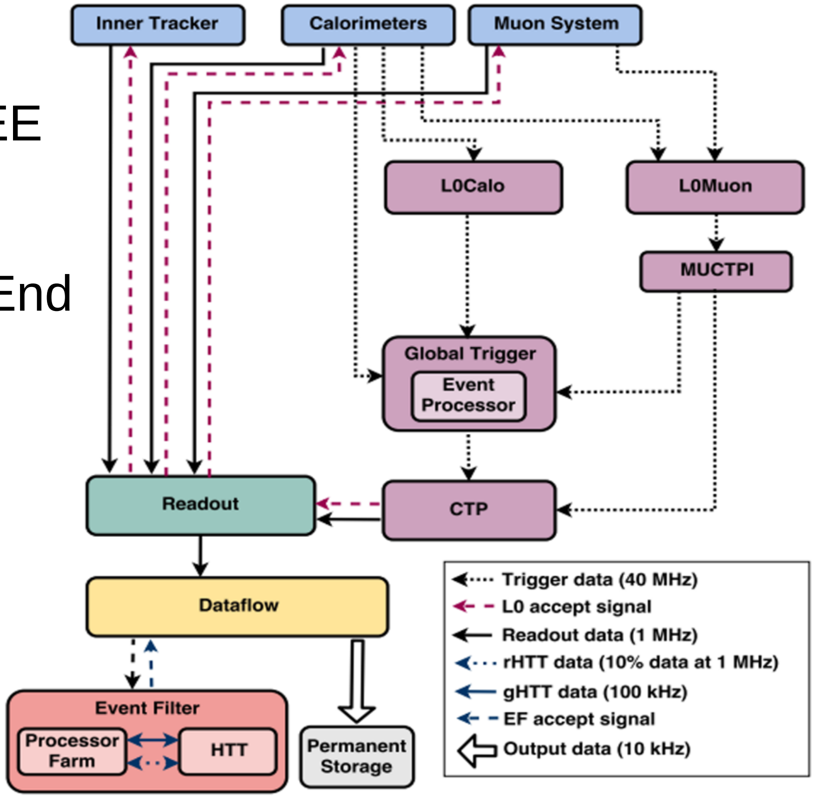
Ref: Belle II DAQ system talk by Qidong Zhou

Conventional hardware trigger + software HLT RoI for PXD readout

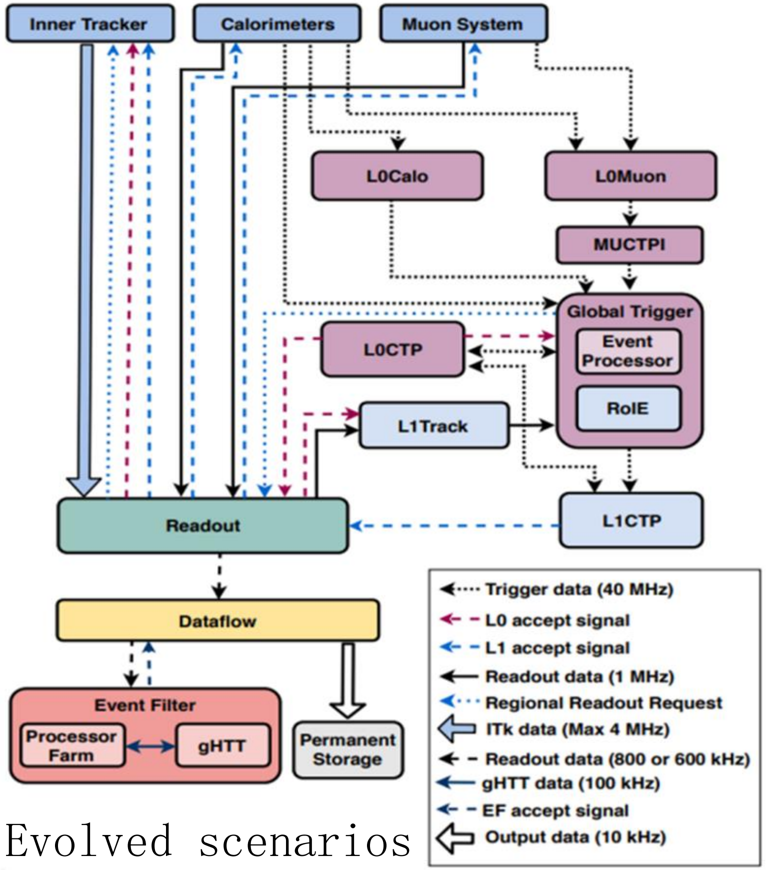
Technology survey

- Atlas TDAQ(Phase II)
  - Data rate 4.6 TB/s
  - Collect trigger primitive from BEE (Back-End Electronics)
  - Fast L0(3us) + L1(10us) + HLT
  - HW trigger sent to FEE (Front-End Electronics)
  - Common PCI card BEE
  - Global HTT(Hardware Track Trigger)
    - FPGA based

FELIX-155, PCIe gen5x16 482Gb/s, 48 fiber link



Baseline



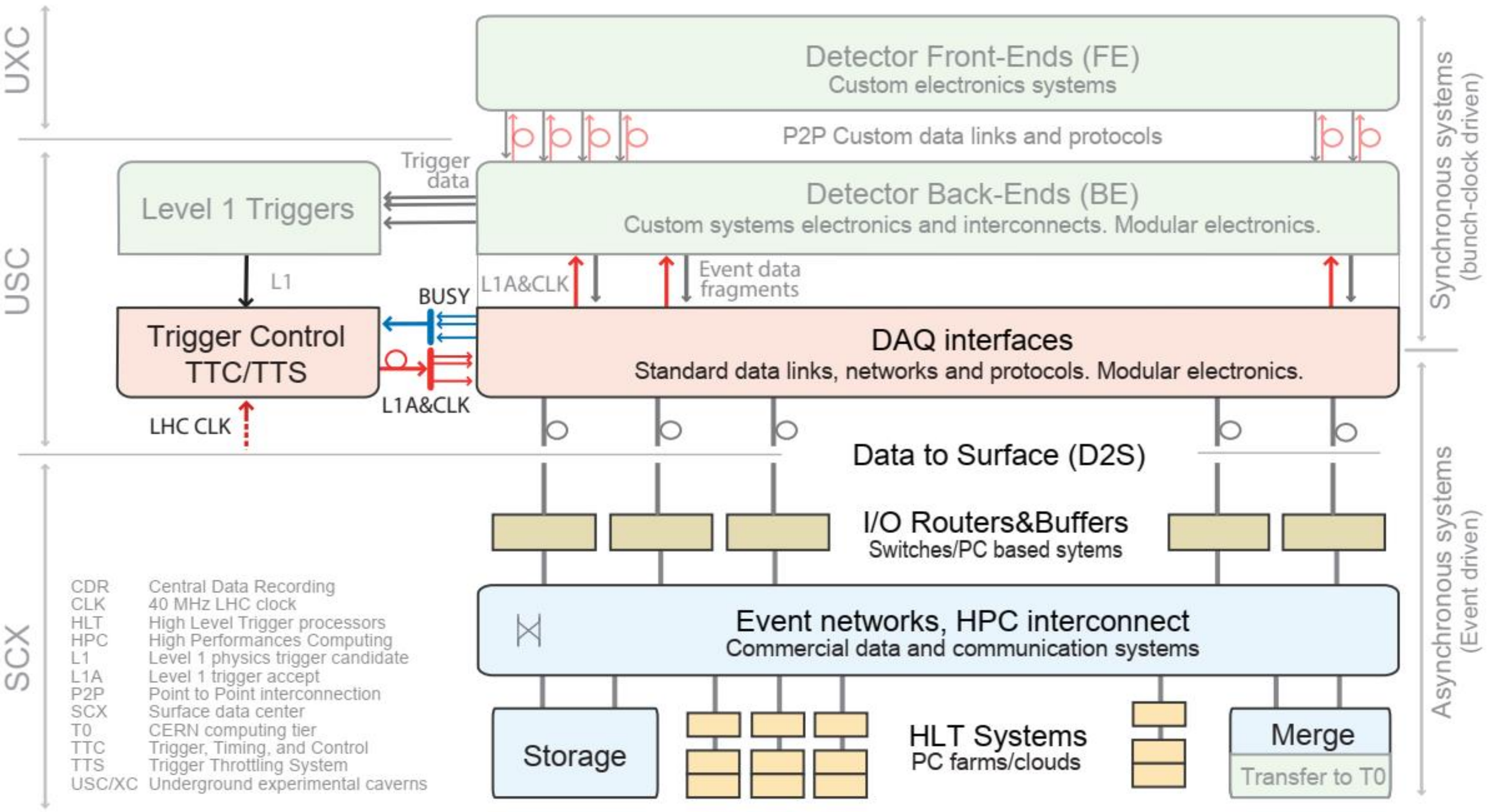
Evolved scenarios

Ref: ATLAS Trigger and Data Acquisition Upgrades for the High Luminosity LHC, LeptonPhoton2019

Fast L0 trigger for inner tracker readout, full PCIe bus readout

Technology survey

- CMS TDAQ(Phase II)
  - Data rate 5.5TB/s
  - L1(12.5us) + HLT
  - Part FEE trigger less readout
  - Common ATCA BEE
    - Serenity
  - ATCA readout board with Ethernet
    - DTH-400Gb/s
    - DAQ-800Gb/s

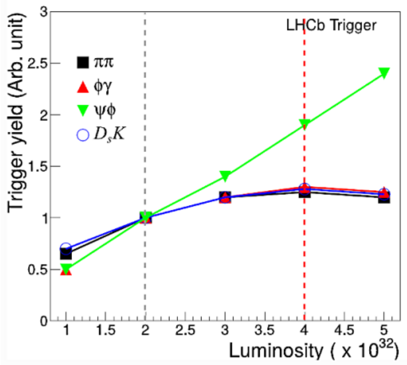


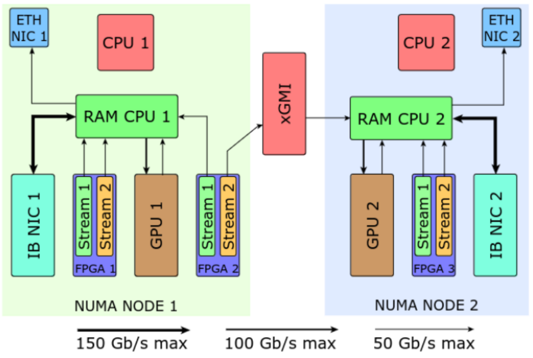
Ref: The Phase-2 Upgrade of the CMS DAQ Interim Technical Design Report

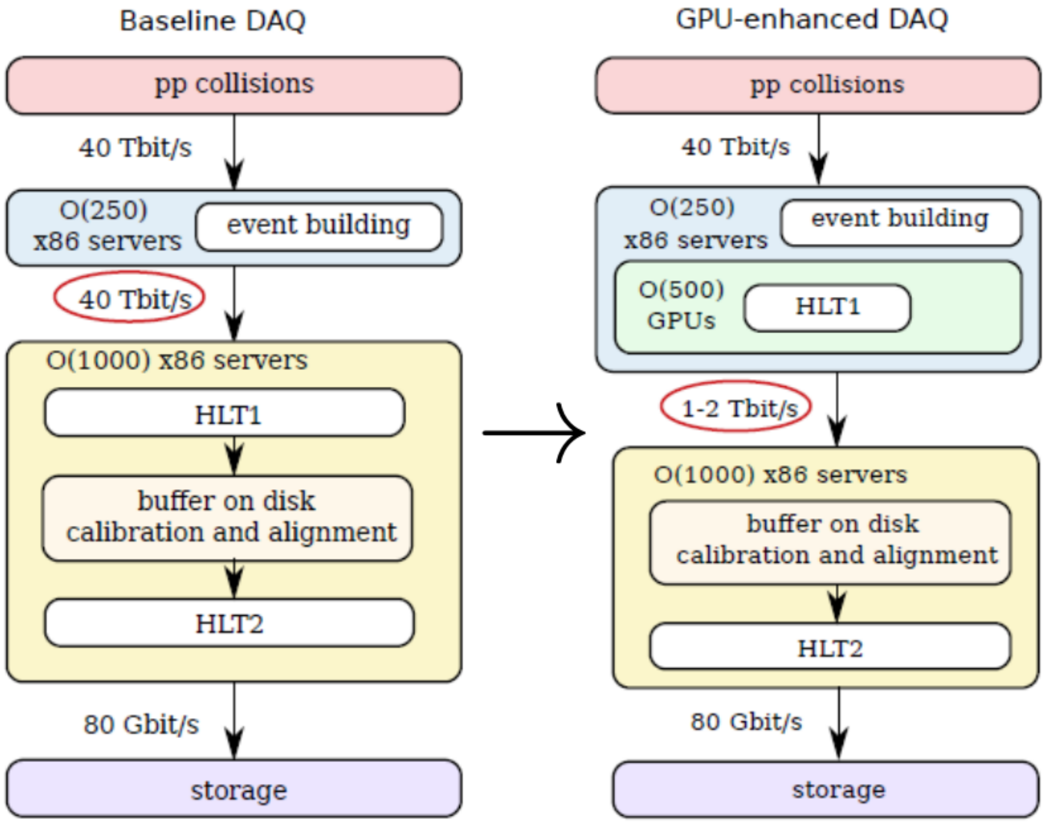
Collect trigger primitive from BEE, full network readout

Technology survey

- LHCb run3
  - Run 1–2 trigger: hardware L0 (40→1 MHz)
  - Read full event at bunch-crossing rate(4TB/s)
    - Cope with higher occupancy.
    - Faster/higher precision tracking
  - Design characteristic:
    - Use disk as a buffer between HLT1 and HLT2.
    - Compute at HLT1 level using GPUs.
    - Event Building using Smart NICs.





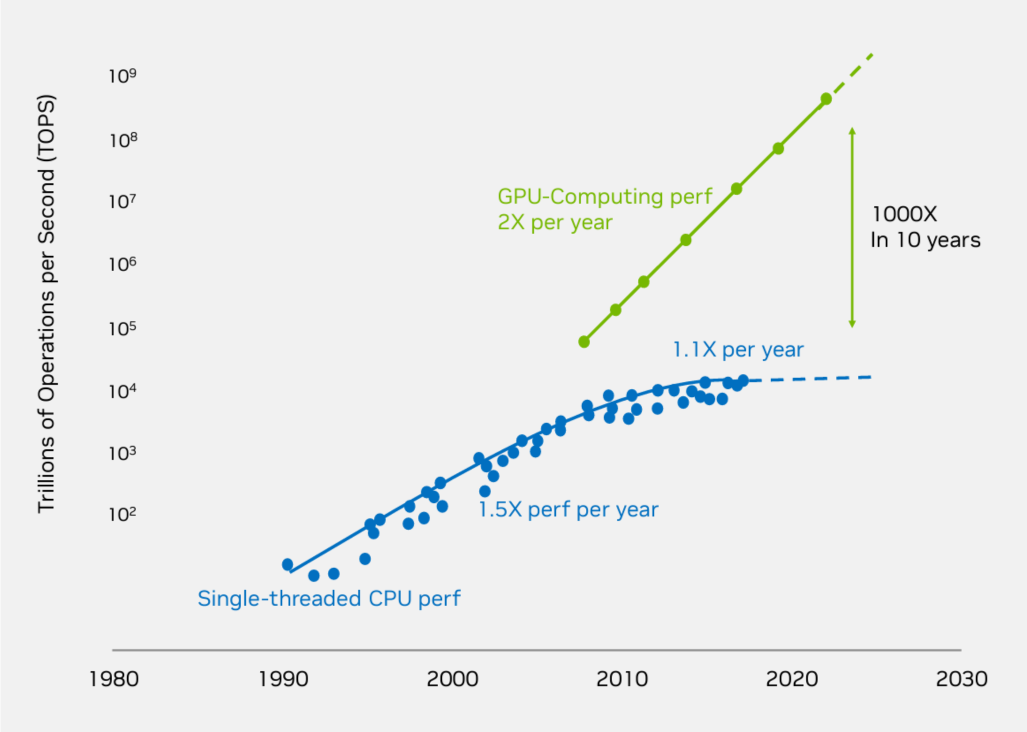


Ref: GPU-based software trigger for LHCb experiment talk by Anton Poluektov.

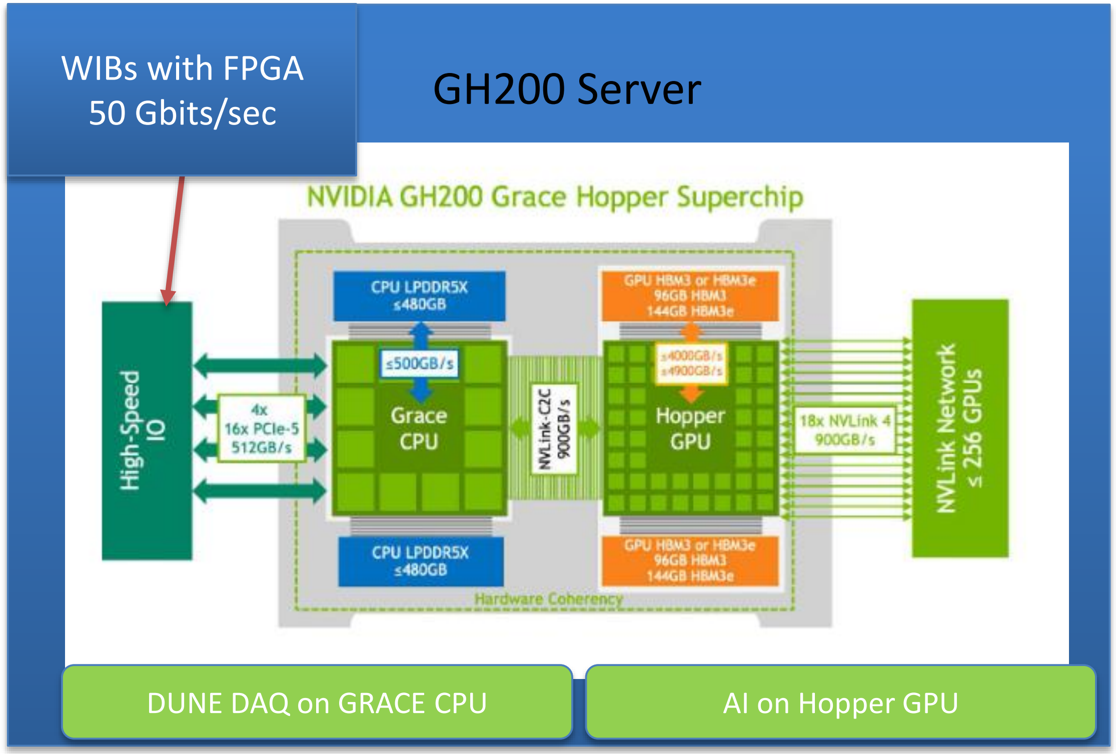
Full software trigger, PCIe bus readout, GPU acceleration

Technology survey

- 800Gbps network is commercially available
- Huang's law: computational power of GPU increase 1000 times in past 10 years
- NVIDIA GH200 server: Arm CPU + GPU, IO > 500GByte/s



Ref: https://www.hangyan.co/charts/3351671202081932642



Ref:DUNE Cold Electronics R&D at ICEBERG, ICHEP-2024, Prague

Our choices

- Trigger solutions
  - 1.Full FEE trigger less readout + L1 + HLT
    - Baseline option
    - Simplified FEE design, extract trigger primitive from BEE
    - No high demand for low L1 latency
  - 2.Full software trigger
    - Preferred option
    - Simplified BEE and trigger design
    - When L1 compression ratio is low
  - 3.Fast L0 + L1 + HLT
    - Backup option
    - When not enough readout bandwidth for part FEE
    - Fast L0 means low trigger latency requirement for part FEE
    - Extract L0 trigger primitive from part FEE

Main Technical Challenges

- 1. Full FEE trigger less readout + L1 + HLT
  - – FPGA algorithm: high data compression ratio
- 2. Full software trigger
  - – Resource requirement
  - – High data throughput and online processing efficiency
- 3. Fast L0 + L1 + HLT
  - – Low L0 latency
  - – Trigger efficiency
  - – Synchronize control
  - – Compression ratio

Previous experience of TDAQ Hardware

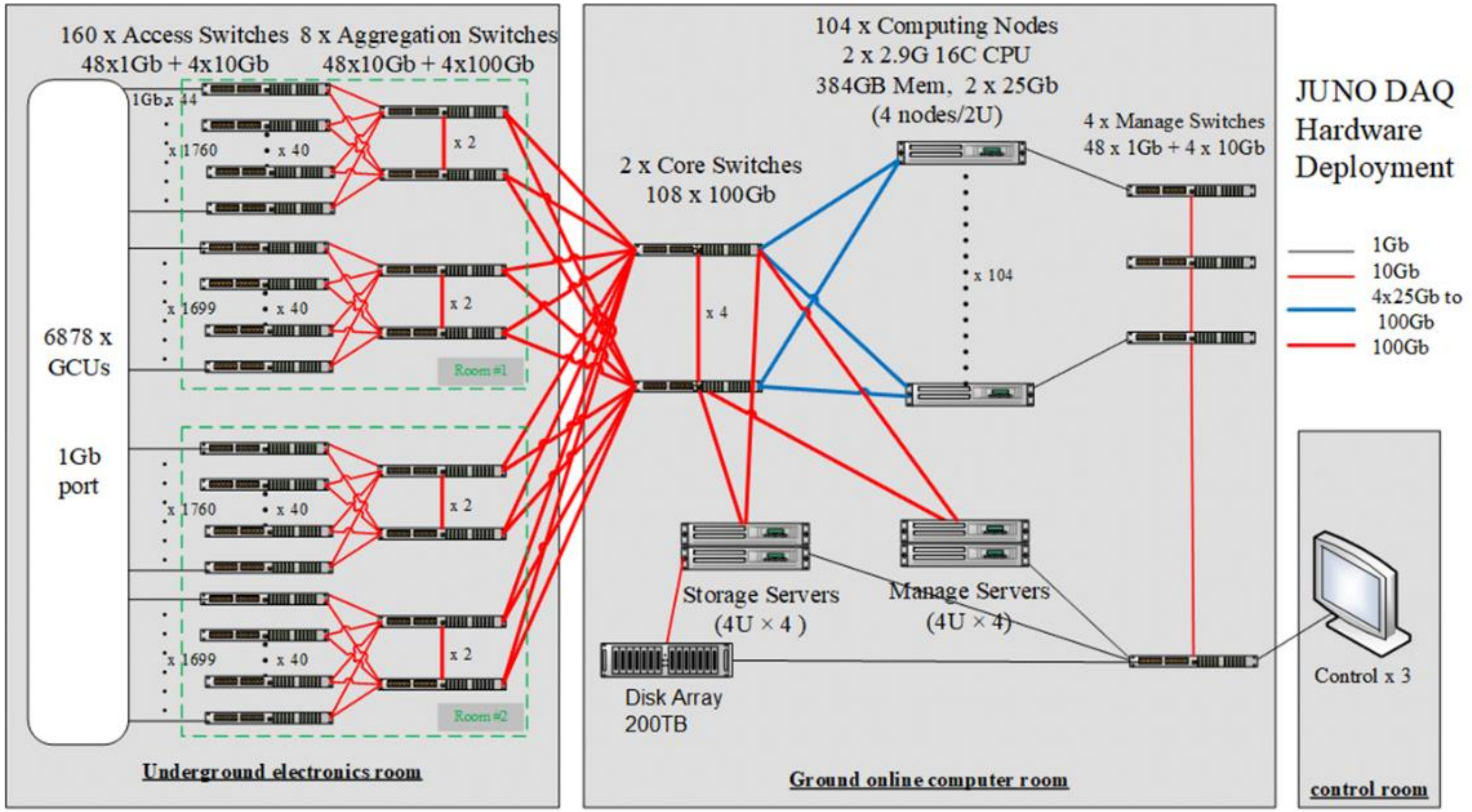
- Designed BESIII trigger system
  - Trigger simulation/hardware design/core trigger firmware development
  - Common trigger board design for upgrade
  - Share link for data readout, data, fast/slow control and clock transmission
- GSI PANDA TDAQ R&D
  - Proposed concept of triggerless readout in TDAQ
  - Designed HPCN board for TDAQ/EMC trigger algorithm development
- Designed Belle2Link and HPCN V3 as ONSEN for Belle II
- Designed CPPF system for CMS Phase-I
  - Design MTCA board, Cluster finding and fanout to EMTF/OMTF
- Designing iRPC/RPC Backend/Trigger for CMS Phase-II
  - Proposed iRPC Backend system scheme, cluster finding firmware
  - ATCA common Backend and trigger board

FAIR — Facility for Antiproton and Ion Research in Europe

Extensive experience in TDAQ system design, algorithm and hardware development

Previous experience of DAQ&DCS

- BESIII DAQ & DCS
  - Running since 2008
- Dayabay experiment DAQ&DCS
  - Running from 2011 to 2020
- LHAASO DAQ
  - Running since 2019,
  - 7k channels, TCP readout
  - Full software trigger
- JUNO DAQ&DCS under developing
  - 40GB/s, 45k channels, TCP
  - Two type data stream
    - HW trigger for waveform
    - Software trigger for TQ hits
  - Online event classification
    - Integrated offline algorithms, compress waveform data to 60MB/s.

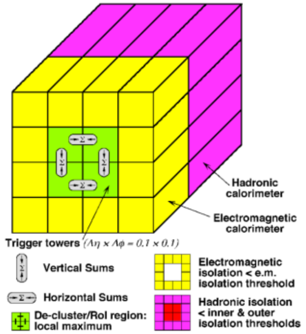


Extensive experience in DAQ&DCS development and operation, some in software trigger

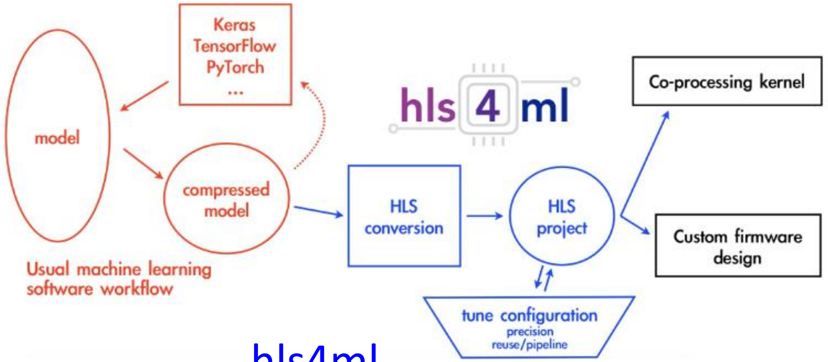
Previous experience of Advance algorithm

Neural network at ATLAS global trigger (Boping Chen)

- Example: tau reconstruction at the hardware trigger level
  - Seed with local maximum algorithm
  - Cluster around the seed to build region of interest (ROI)
- Train the neural network (NN) with ROI
  - To distinguish tau from jet
  - Same as image identification: CNN/DNN…
- Use hls4ml to convert NN model to hls project

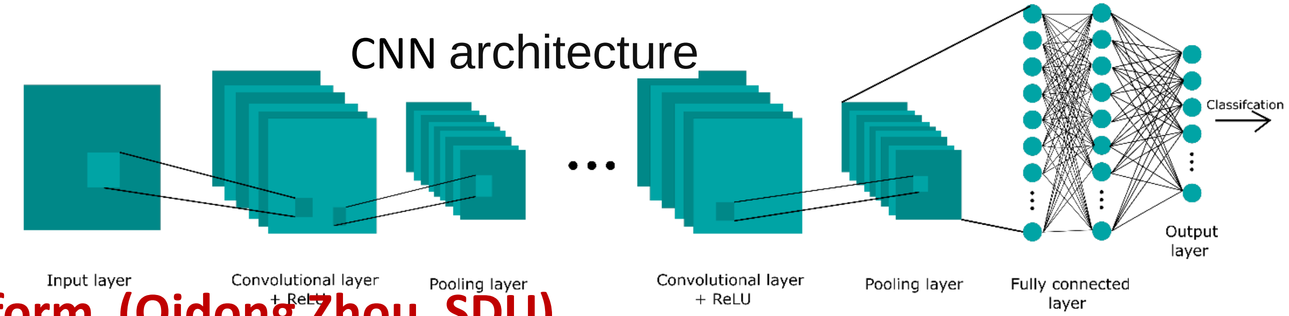


Seed and cluster



hls4ml

CNN architecture

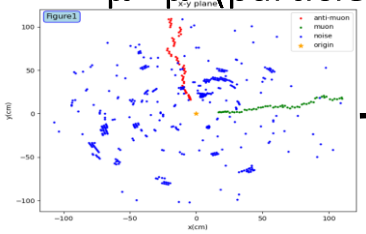


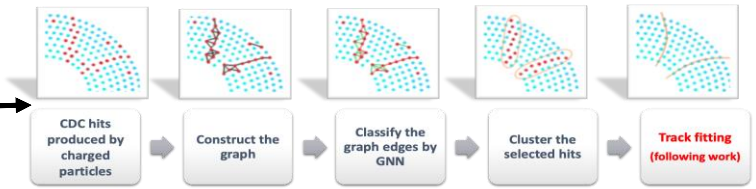
VCK190 test bench

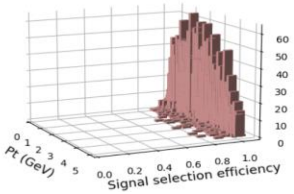
HLT Acceleration on FPGA platform (Qidong Zhou, SDU)

GNN based tracking algorithm for BESIII/Belle II

μ+ μ- (particle gan)







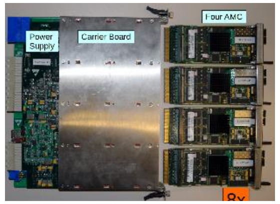
Hit selection efficiency: 98.4%

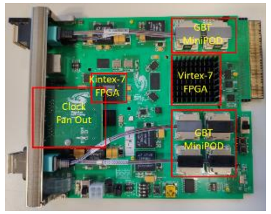
Hit selection purity : 97.9%

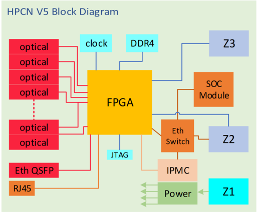
Some experience in L1 NN application and HLT acceleration on FPGA

R&D efforts and results

- Start to design ATCA TDAQ board for CEPC
  - Based on ATCA standard, designed series of ATCA boards
  - Already used in PANDA, Belle II and CMS experiment





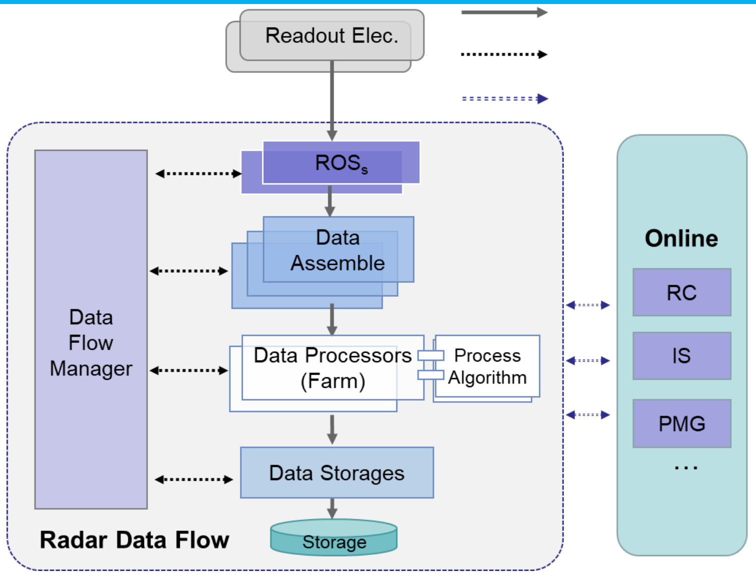


Version	HPCN V1	HPCN V2	HPCN V3	CPPF	HPCN V4	Serenity	HPCN V5
Line rate	3Gbps	3Gbps	6Gbps	10Gbps	16Gbps	25Gbps	10-25Gbps
Num. of ch	8 ch	8ch	8-16ch	36ch	48ch	124ch	36-80 ch
Buffer	2GB	2GB	4GB	—	48GB	—	TBD
Ethernet	1GbE			1/10GbE	10GbE	100GbE	40-100GbE
Aplication	PANDA		Belle II/PXD	CMS Phase-I	CMS Phase-II		CEPC R&D
Time	2006-		2009-	2013-	2017-		2024-

Streaming Readout Framework – RADAR

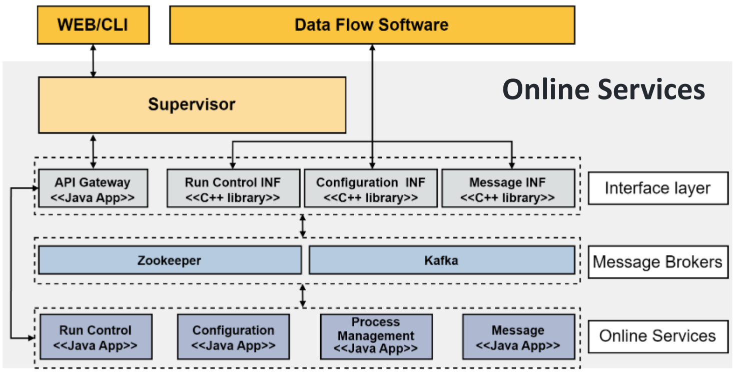
heteRogeneous Architecture of Data Acquisition and pRocessing

- **V1:** deployed in LHAASO (3.5 GB/s data rate), *software trigger mode*
- **V2:** upgraded for JUNO (40 GB/s data rate), *mix trigger mode*
  - Containerized running
  - High availability support
- **V3: CEPC-oriented** (~ TB/s data rate) , under development



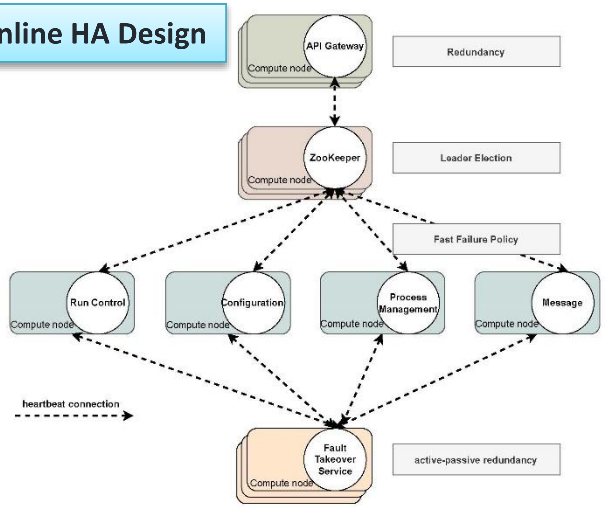
- **Motivation:**
  - High-throughput data acquisition and processing
- **Current Status:**
  - Over a decade of work led to significant progress, tested through experiments
- **Recent Focus:**
  - **Heterogeneous online processing platforms with GPU**
  - **Real-time data processing acceleration solutions**
- **Expansion:**
  - **Application across various domains** (DAQ, triggering, control, etc.)
  - **Integration of AI technologies** (ML, NLP, expert systems, etc.)

- **General-purpose distributed framework**
- **Lightweight structure**
- **Plug-in modules design**
- **Microservices architecture**

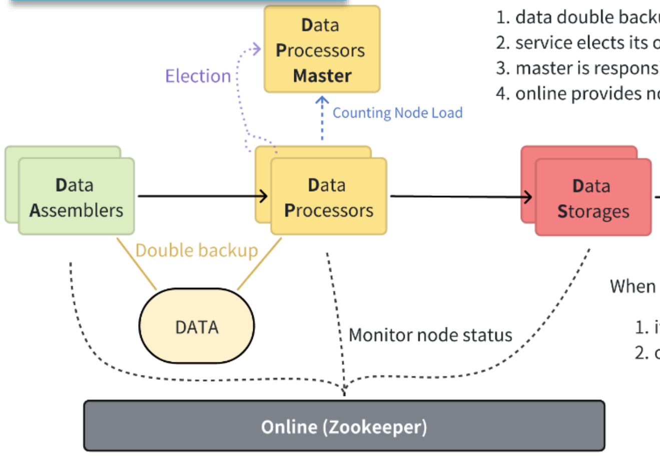


R&D efforts and results

Online HA Design



Dataflow HA Design



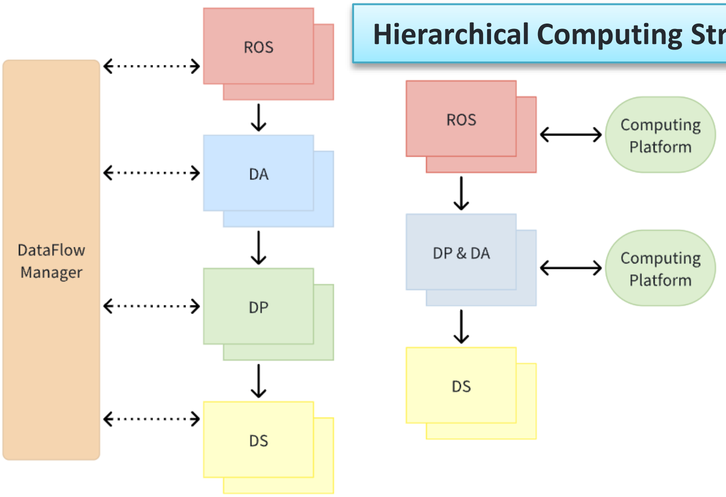
High Availability (HA):

1. data double backup
2. service elects its own master
3. master is responsible for load balancing
4. online provides node status monitoring services

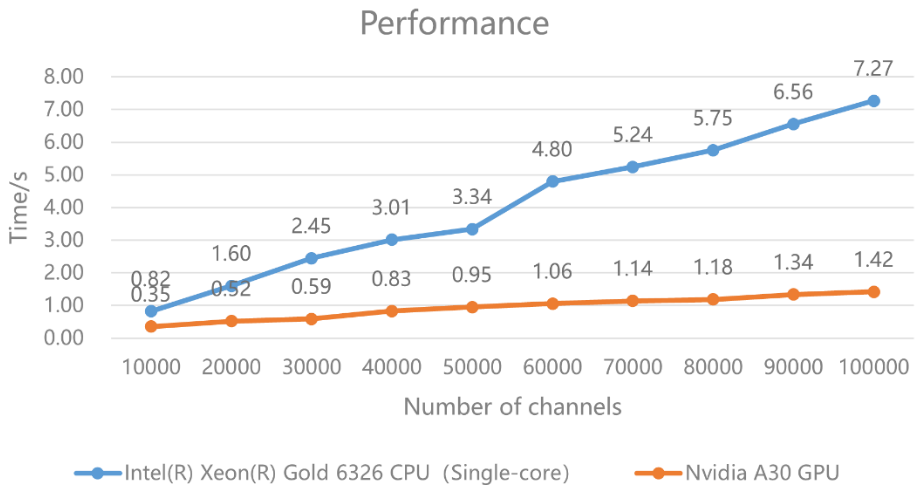
When one **DP** crash:

1. if master node, re-election
2. or, re-transferring backed up data

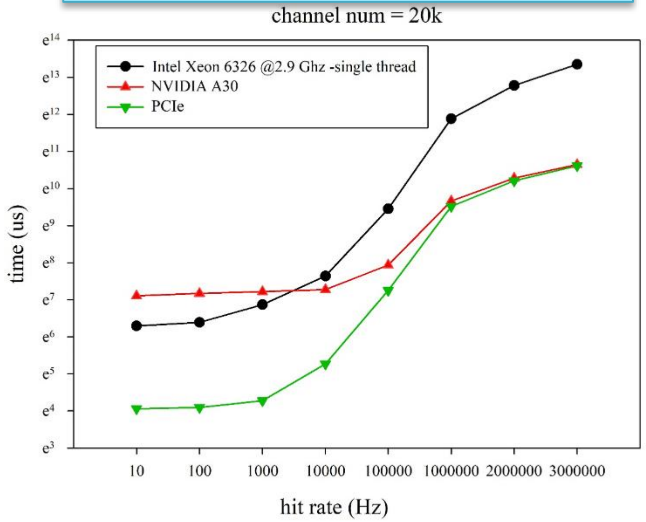
Hierarchical Computing Strategy



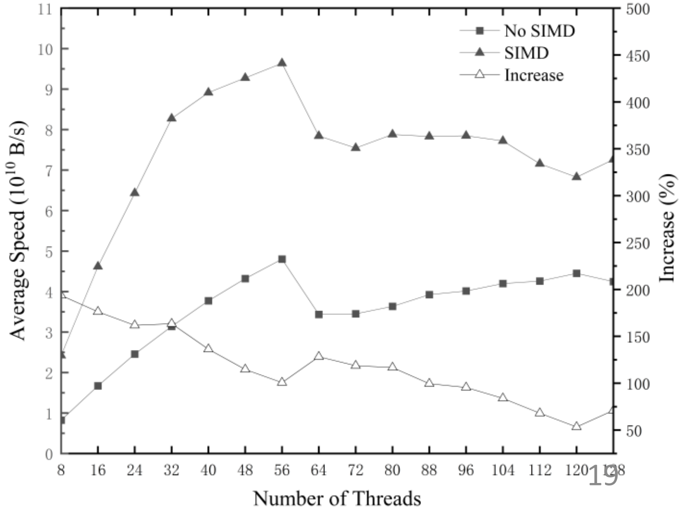
GPU based online waveform reconstruction in JUNO



GPU based nHit trigger algorithm study

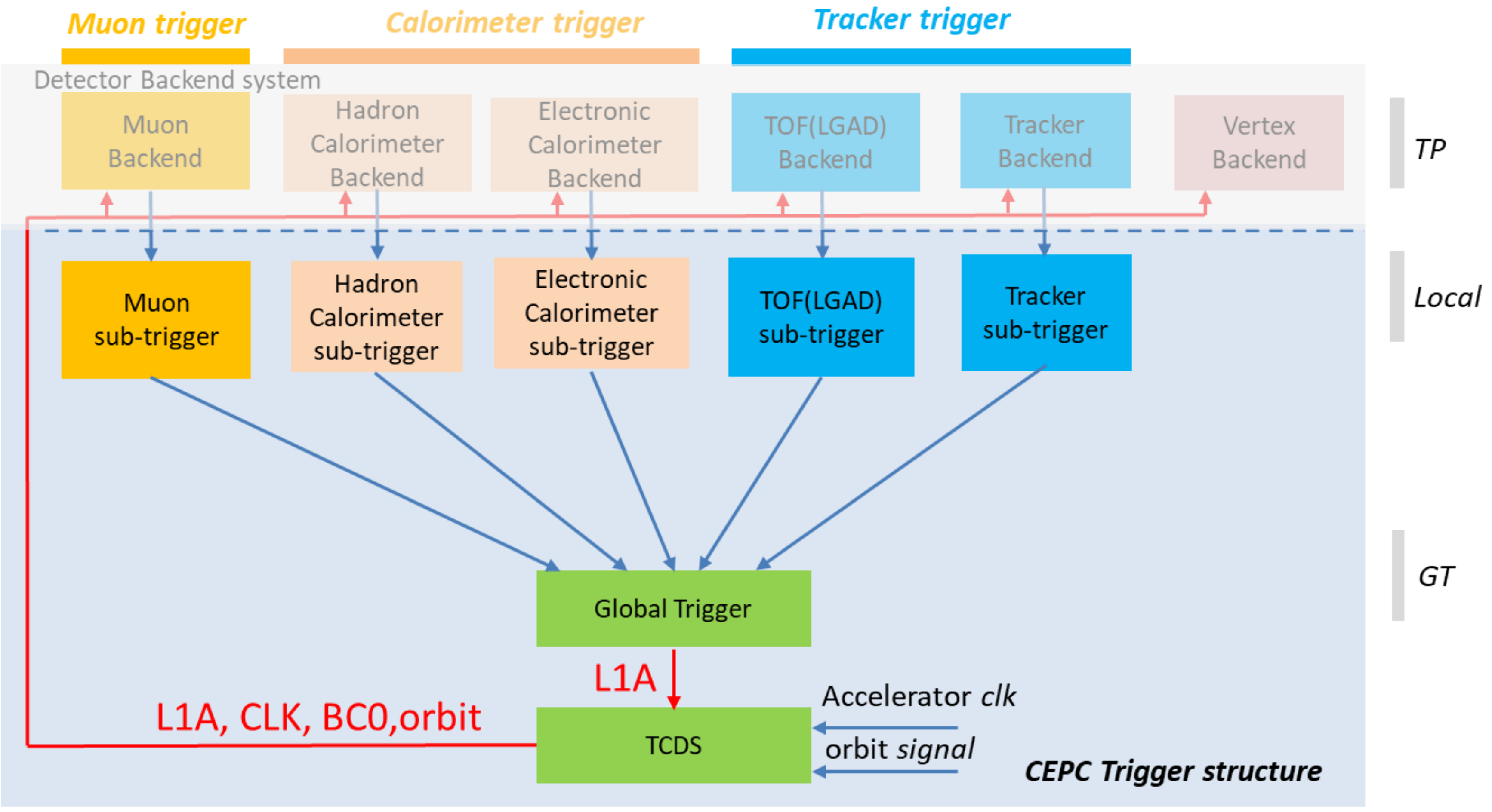


SIMD based CPU processing acceleration



Preliminary design of hardware trigger

- HW Trigger structure
  - Baseline option
    - HW trigger sent to BEE
    - L1 at back-end
  - ➢ Backend of each detector generate Trigger Primitive(TP)
  - ➢ Sub trigger of generate local detector trigger information(energy, track…)
  - ➢ Global trigger generate L1A according to physical requirement.
  - ➢ TCDS distribute clock and fast control signals to BEE.



Preliminary design of hardware trigger

- Trigger structure
  - Backup option
    - HW trigger sent to FEE
    - Fast L0 + L1

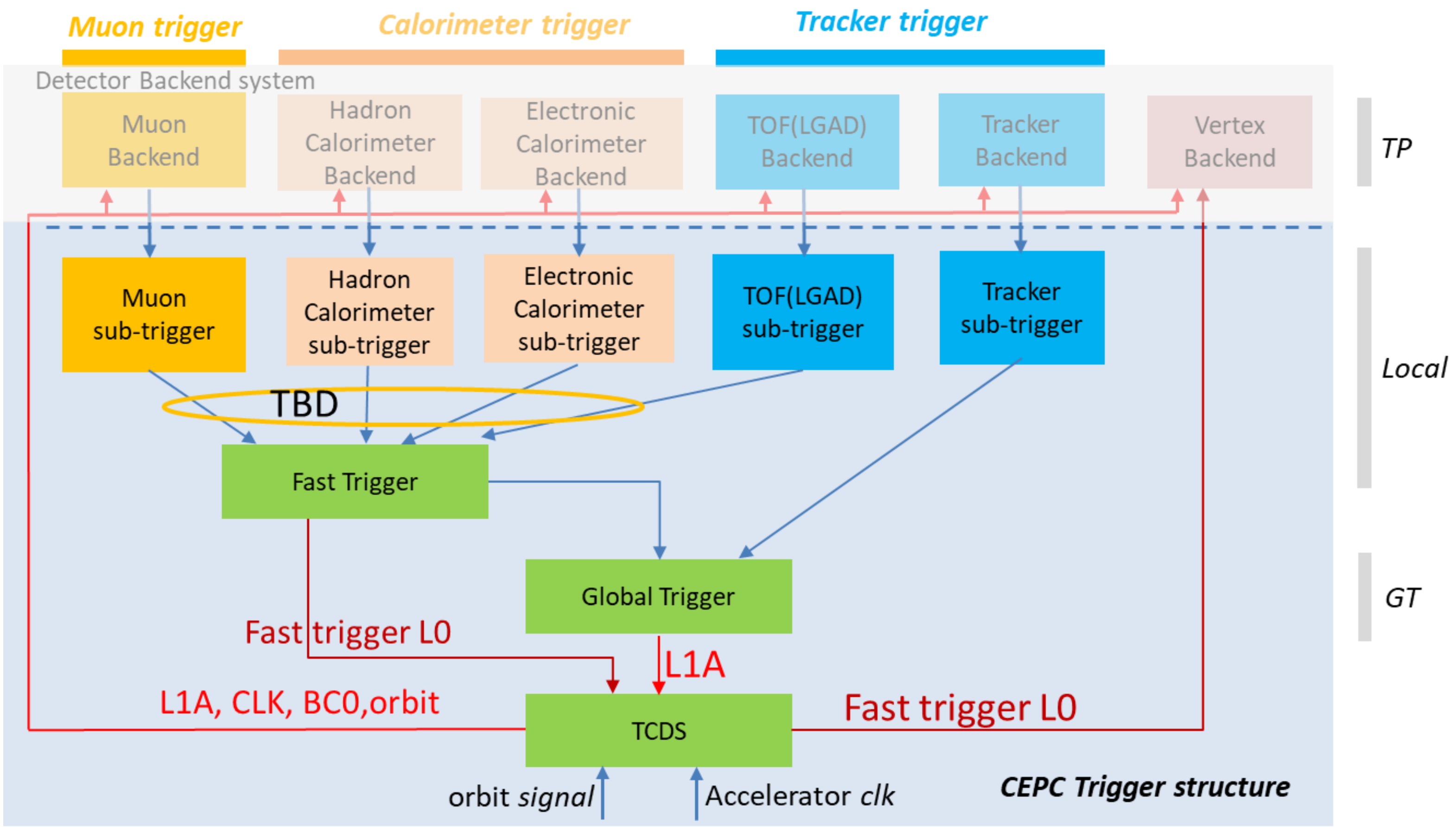
➢ Backend of each detector generate Trigger Primitive(TP)

➢ Sub trigger of generate local detector trigger information(energy, track…)

➢ Fast trigger generate local low latency L0A for Vertex to reduce data. Which detectors join this trigger need to be discussed.

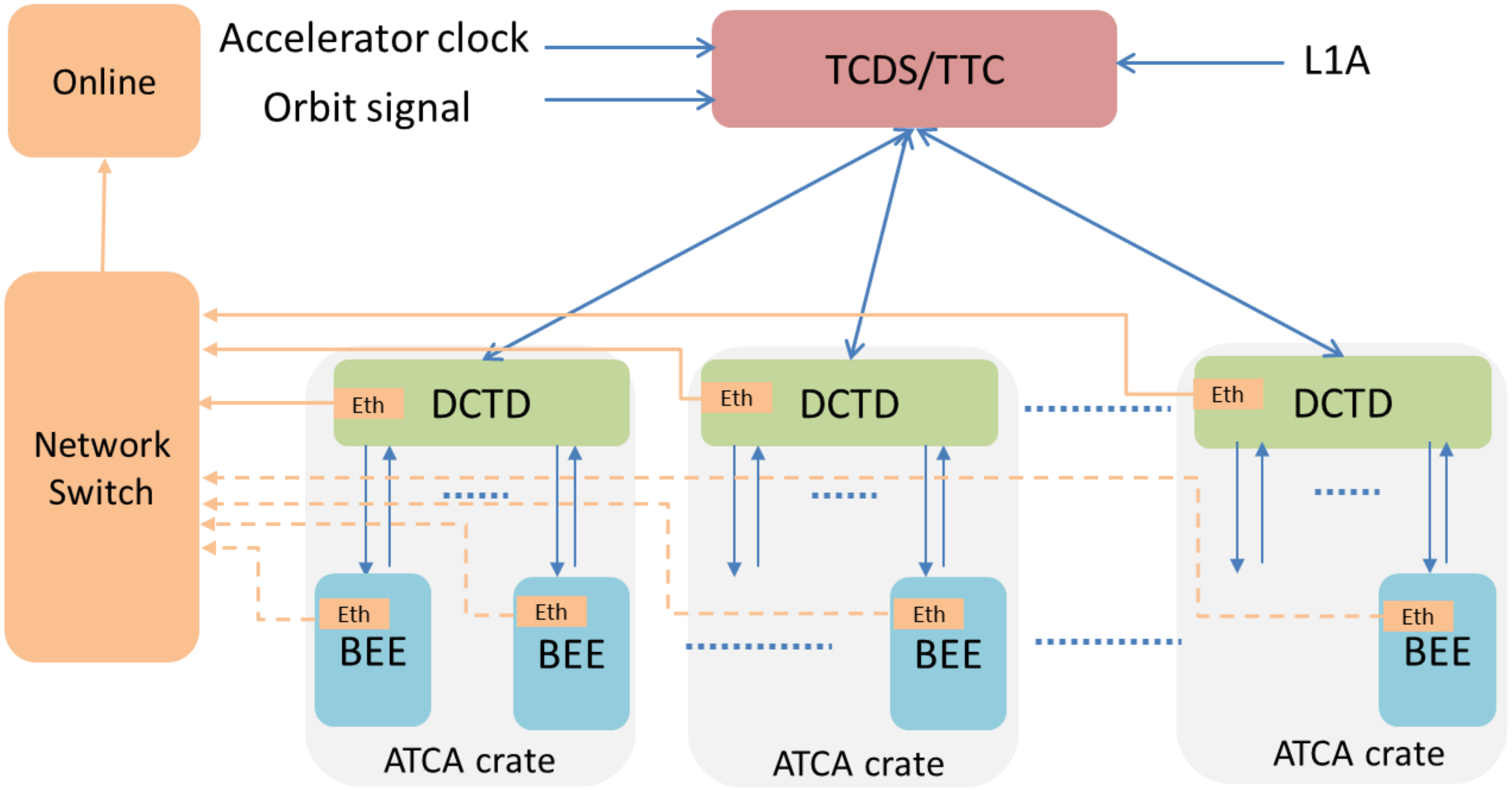
➢ Global trigger generate L1A according to physical requirement.

➢ TCDS distribute clock and fast control signals to BEE.



Preliminary design of TCDS/TTC and readout

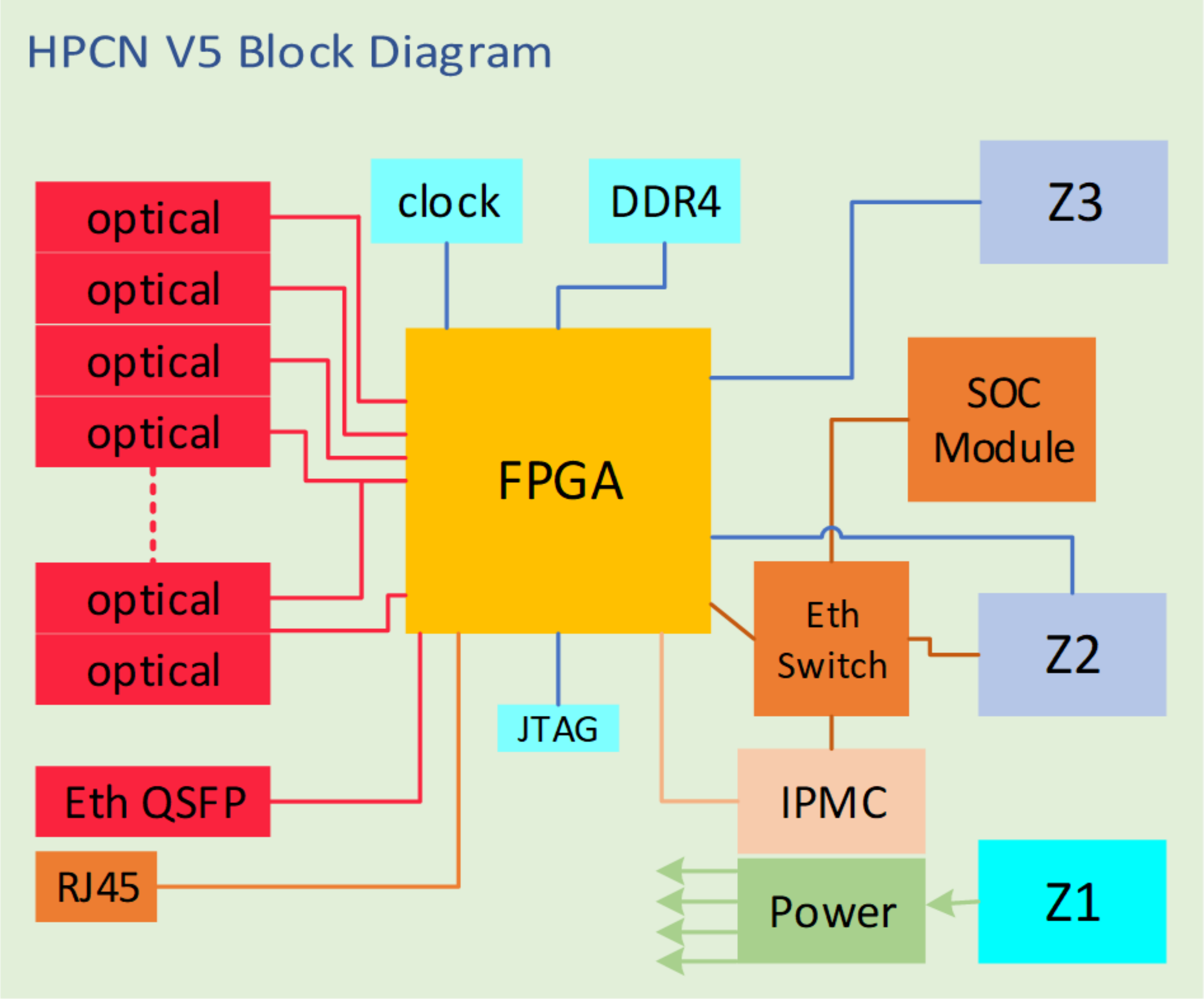
- TCDS/TTC
  - Clock, BC0, Trigger, orbit start signal distribution
  - Full, ERR signal feed back to TCDS/TTC and mask or stop L1A
- DAQ readout
  - Option1：BEE Data collected and packaged by DCTD board, and sent to Online via network switch.
  - Option2： BEE Data sent to Online via network switch.



- TCDS-Tigger Clock Distribution System
- TTC- Trigger, Timing and Control
- DCTD-Data Concentrator and Timing Distribution
- BEE-Backend board Electronic

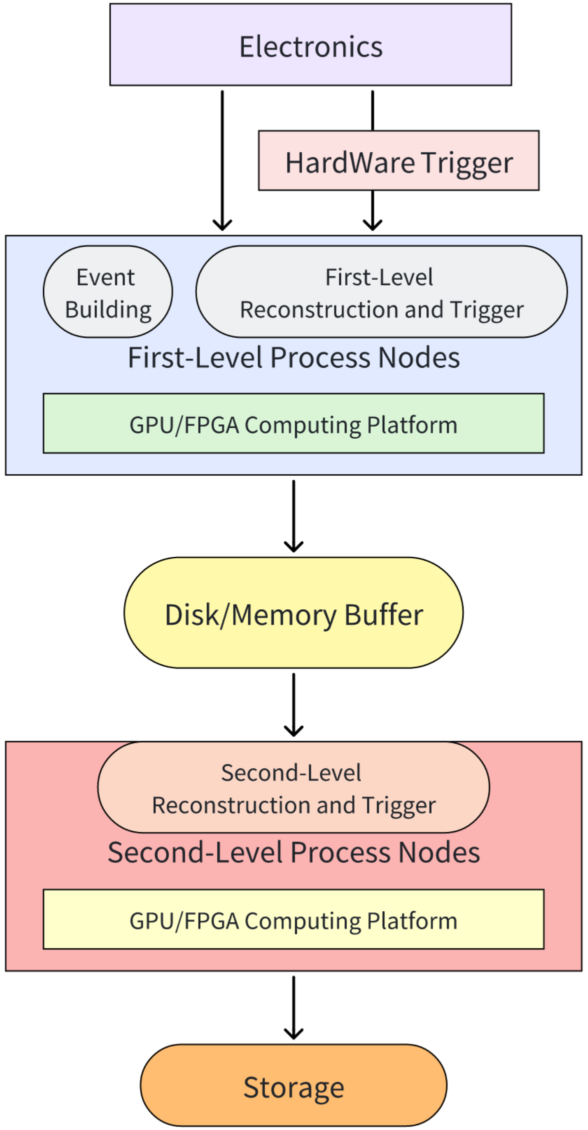
Preliminary design of the common Trigger Board

- Common Trigger board function list
  - ATCA standard
  - Virtex-7 FPGA
  - Optical channel: 10-25 Gbps/ch
  - Channel number:36-80 channels
  - Optical Ethernet port: 40-100GbE
  - DDR4 for mass data buffering
  - SoC module for board management
  - IPMC module for Power management



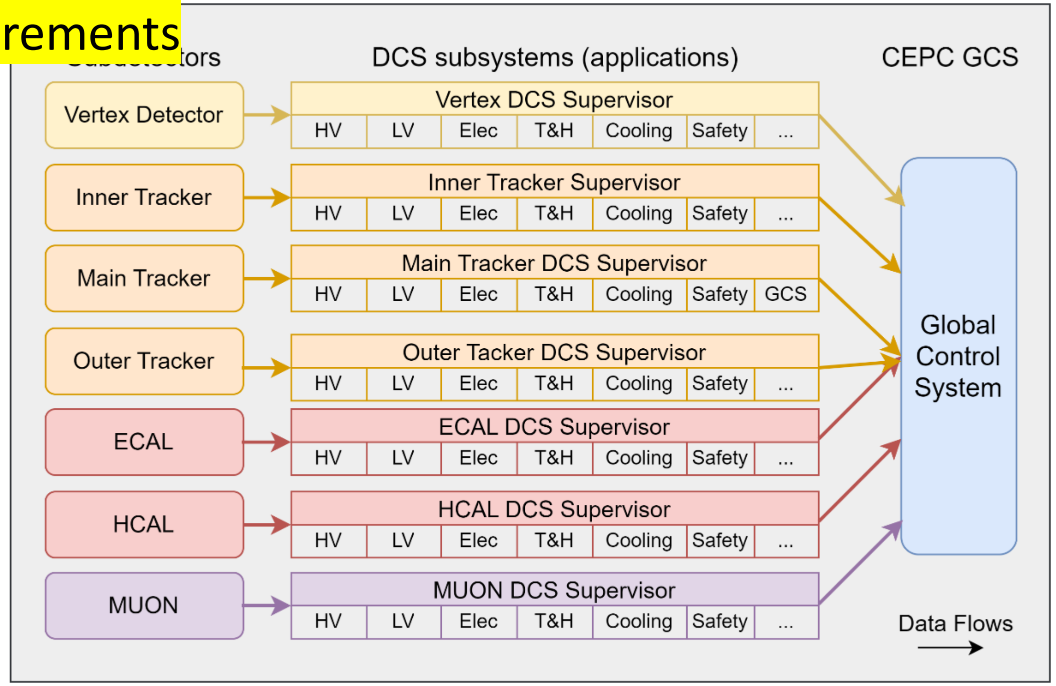
Preliminary design of DAQ

- Same with or without hardware trigger
- Readout interface and protocol
  - Ethernet X*100Gbps/ TCP or RDMA
  - PCIe optional
- GPU acceleration at HLT1 & HLT2
  - FPGA optional
- Memory vs disk buffer for HLT2
  - Better IO performance but smaller volume size
- RADAR software framework
  - Heterogeneous computing cluster

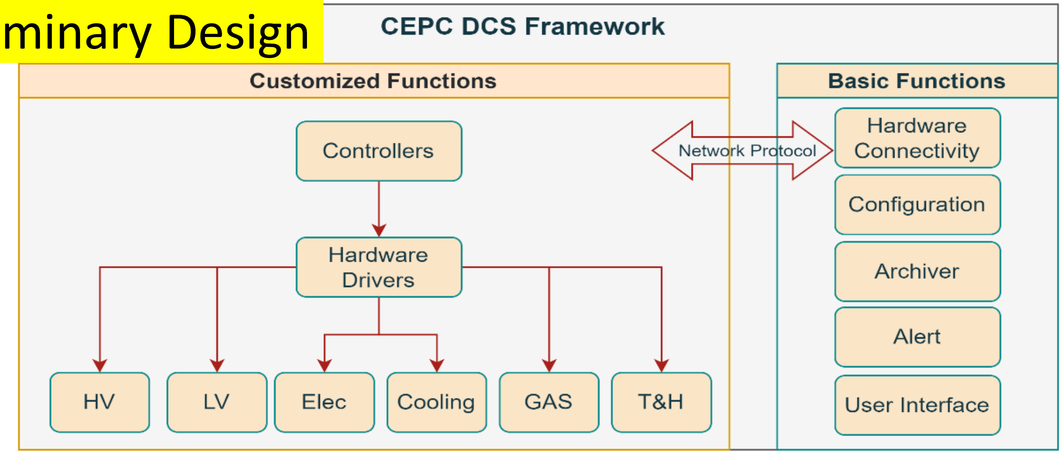


Preliminary design of DCS

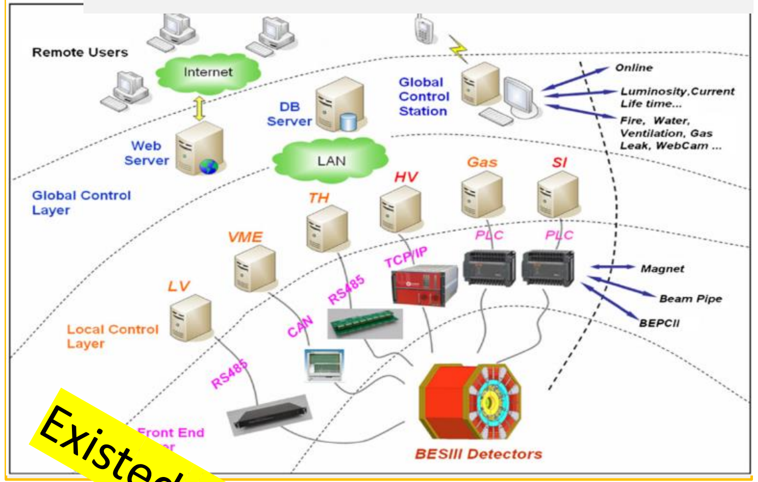
Requirements



Preliminary Design



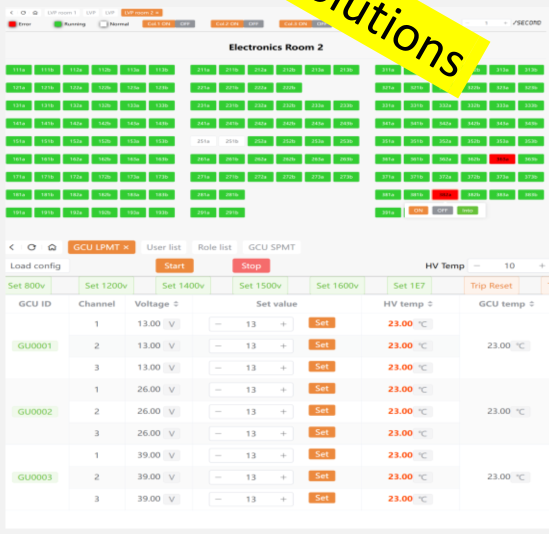
BESIII DCS ***Based on LabVIEW***



Existed Solutions

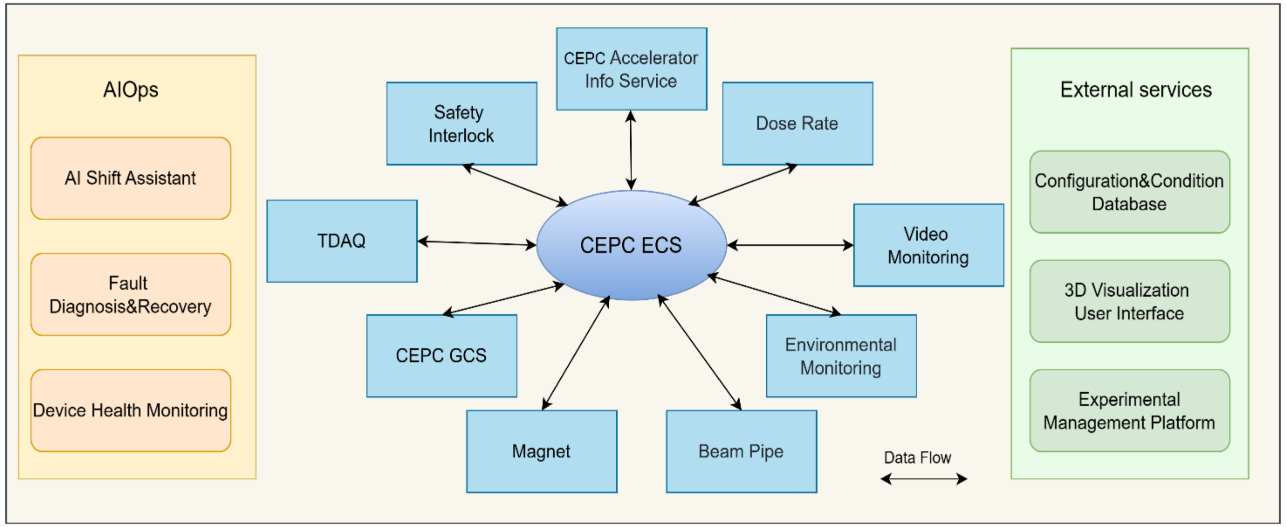
JUNO DCS

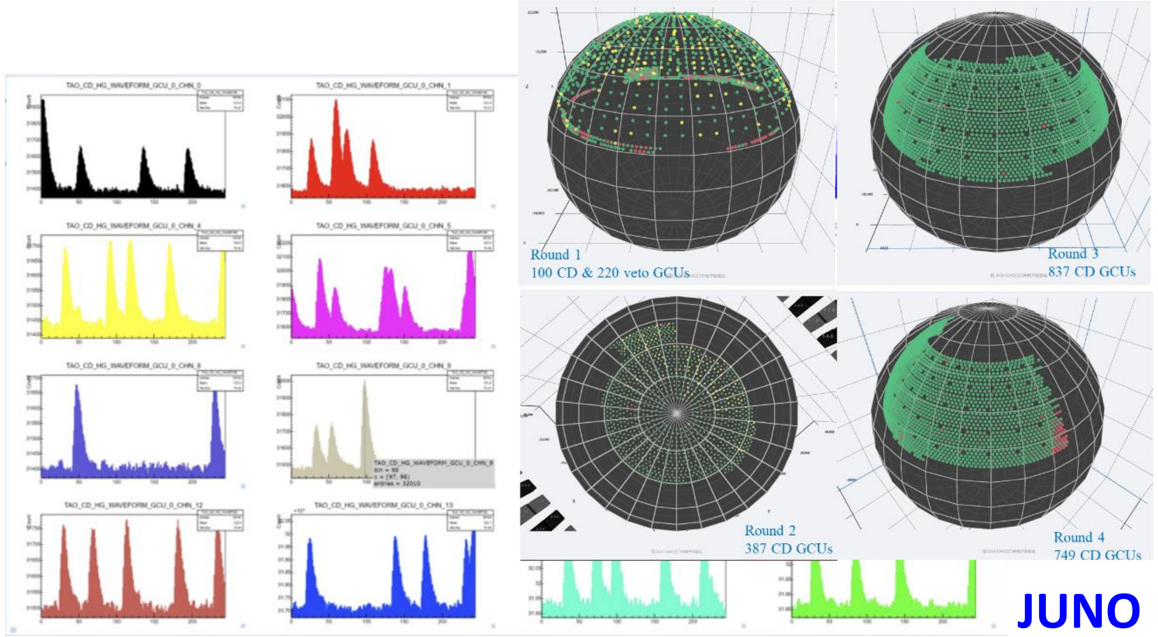
Based on EPICS



Preliminary design of ECS

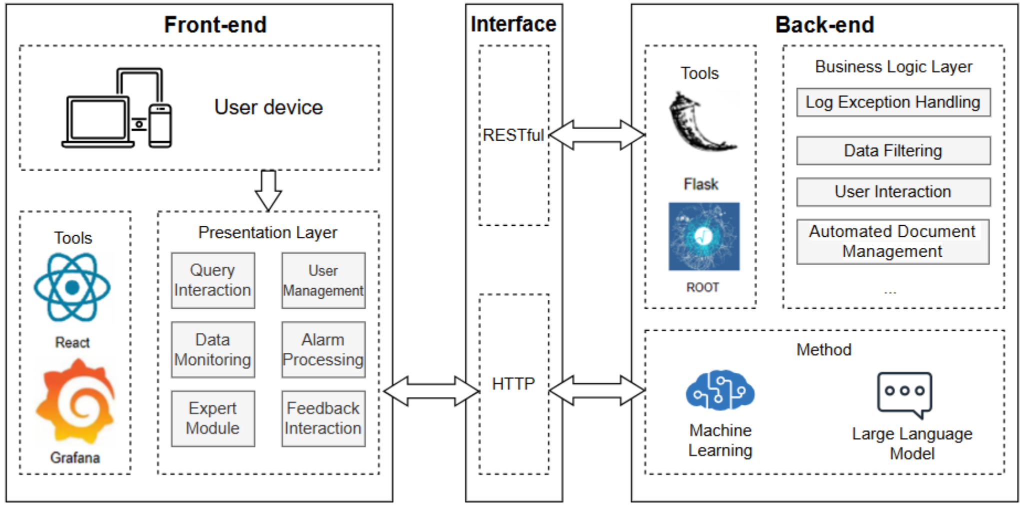
Main components of the ECS





JUNO

- Existed Solutions:
  - 3D Visualization Monitoring
  - AI shift assistant research based on LLM+RAG (TAOChat)
  - Fault root analysis method based on directed acyclic graph
  - A ROOT-based Online Data Visualization System (ROBOT)
- Unified control and monitoring for TDAQ, DCS and others
- AI operation and maintenance



Research Team

- 15 staff of IHEP TDAQ group
  - Kejun Zhu (team director)

Born in	number
1960s	3
1970s	4
1980s	4
1990s	4

- DAQ (4 of 6)
  - Hongyu Zhang (readout)
  - Fei Li (DAQ, team manager)
  - Xiaolu Ji (online processing)
  - Minhao Gu (software architecture)
- Trigger (4 of 5)
  - Zhenan Liu (trigger director)
  - Jingzhou Zhao (hardware trigger)
  - Boping Chen (simulation/algorithm)
  - Sheng Dong (firmware/DCS)
- DCS/ECS (1 of 4)
  - Si Ma

- Collaborators
  - Qidong Zhou (HLT, SDU)
  - Yi Liu (HLT, ZZU)
  - Junhao Yin(HLT, NKU)

Born in	number
1980s	2
1990s	1

- IHEP Students(20 totally)
  - 2 Phd and 4 master
- New member need
  - 2 staff next year
  - 2 postdoc
- Looking for more collaborators

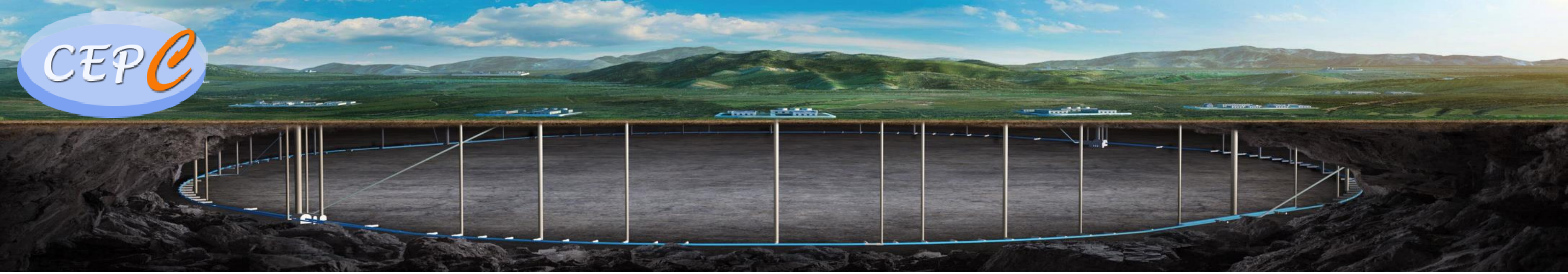
Beginning to gather manpower for R&D, most involved only small part of the time

Working plan

- TDR related
  - Basic Trigger simulation and algorithm study
    - Event rate and basic algorithm scheme for each detector
  - Hardware trigger and interface design
  - Finalize TDAQ design scheme
- R&D
  - Trigger simulation and algorithm
  - Hardware trigger, fast control and clock distribution
  - ROCE/RDMA readout protocol and smart NIC
  - TB/s high throughput software framework(RADAR)
    - FPGA/GPU acceleration and heterogeneous computing
    - Memory based distributed buffer
  - ML/AI algorithm application
    - Trigger/data compression/ AI operation and maintenance

Summary

- Completed preliminary design of TDAQ and online
  - Mix hardware and software trigger could be adapted solution currently
  - Full software trigger will be best one if no IO and computational power constraints
- No show-stopper found for hardware and software trigger scheme
  - But fast L0 trigger algorithm and handling TB/s data at manageable hardware scale remain challenges.
- Much R&D effort still needed from design to implementation



Thank you for your attention!

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences