Data Acquisition at High Luminosity-EIC (EIC-Phase II)

Outline ● Discussion on the specification ● DAQ strategy ● Possible novel hardware for realization

Jin Huang

Brookhaven National Lab

Brookhaven National Laboratory

Thanks to discussion with many colleagues and partly based on discussion of the past SRO workshops

The EIC collider with 10^{35} instantaneous luminosity

Accelerator implications

- What are the limitations of the current EIC and how can they be overcome?

Physics opportunities

- What novel channels would be opened up with the factor 10 jump in luminosity?

Experimental challenges

- How can we ensure we can cleanly observe the signals that are of interest?

Discussion on the specification

- Collision rate is (still) moderate
- Elephant in the room: background

EIC and HL-EIC: unique collider → unique real-time system challenges

Collision signal rate at HL-EIC looks like that from RHIC

- EIC luminosity is high, but collision cross section is small ($\propto \alpha_{EM}^2$) → low collision rate
- But events are precious and have diverse topology → hard to trigger on all process
- Background and systematic control is crucial → avoiding a trigger bias

	EIC	→ HL-EIC	RHIC	LHC → HL-LHC
Collision species	$\vec{e}+\vec{p}, \vec{e}+A$		$\vec{p}+\vec{p}/A, A+A$	$p+p/A, A+A$
Top x-N C.M. energy	140 GeV		510 GeV	13 TeV
Bunch spacing	10 ns	?	100 ns	25 ns
Peak x-N luminosity	10^{34} cm^{-2} s^{-1}	10^{35} cm^{-2} s^{-1}	10^{32} cm^{-2} s^{-1}	10^{34} → 10^{35} cm^{-2} s^{-1}
x-N cross section	**50 μb**		**40 mb**	**80 mb**
Top collision rate	500 kHz	5 MHz	10 MHz	1-6 GHz
$dN_{ch}/d\eta$ in p+p/e+p	0.1-Few		~3	~6
Charged particle rate	**4M N_{ch}/s**	**40M N_{ch}/s**	**60M N_{ch}/s**	**30G+ N_{ch}/s**

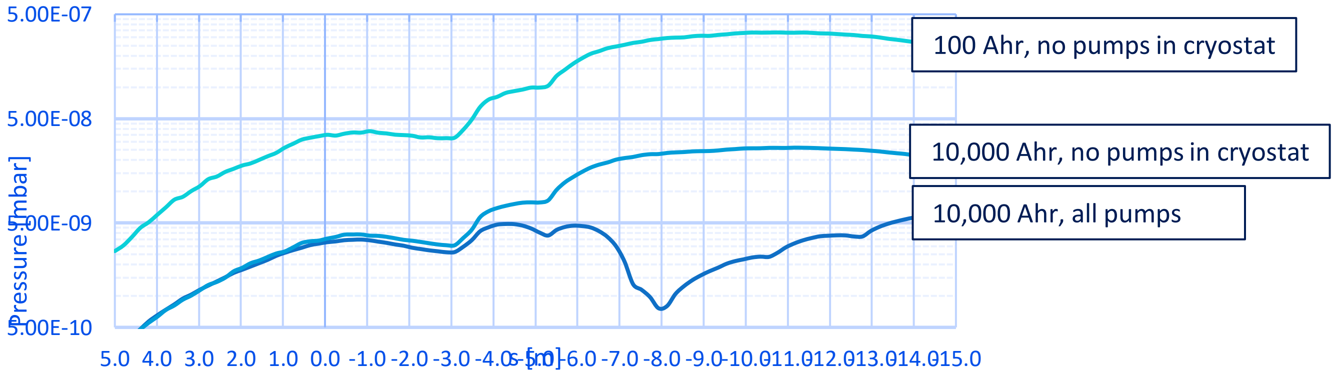
EIC x-sec : further quantification [Courtesy E. Aschenauer]

- Inelastic e+p scattering x-sec:
  - For a luminosity of 10^{35} $cm^{-2}s^{-1}$ 50ub corresponds to **5 MHz**
- Elastic e+p cross-section:
  - For EIC central barrel, elastic cross section is **small** comparing to the inclusive QCD processes
- Beam gas interaction:
  - Beam proton – beam gas fix target inelastic interactions. The pp elastic cross section is smaller (~7 mb)
  - For a vacuum of 10^{-8} to 10^{-9} mbar in the detector volume (10m) this gives a rate of **10-100 kHz**

Beam [GeV]	HERA	5 x 50	10 x 100	18 x 275
$Q^2>10^{-9}$ GeV	65.6	**29.9**	**41.4**	**54.3 ub**
$Q^2>1$ GeV	1.29	0.45	0.65	0.94 ub

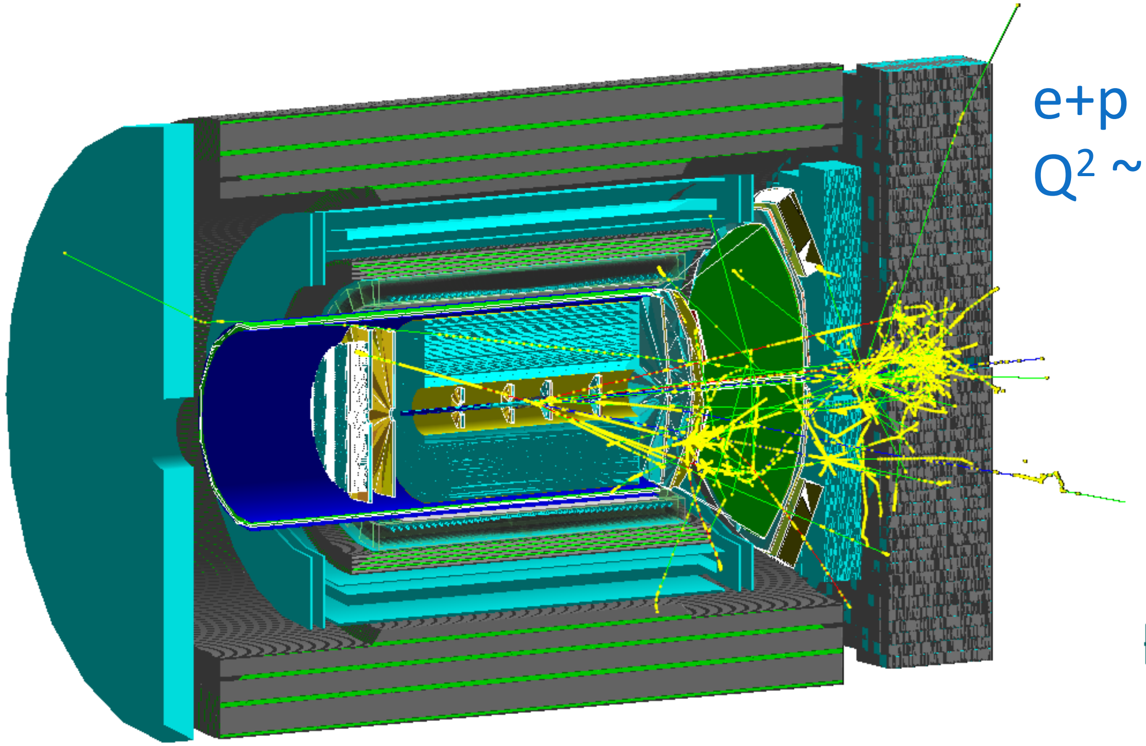
Beam [GeV]	HERA	5 x 50	10 x 100	18 x 275
σ [y_{Exp} >-4]	5 pb	**5 ub**	**0.7 ub**	**0.06 ub**
σ [y_{Exp} >-6]	11 ub	420 ub	100 ub	29 ub

E_p:	50 GeV	100 GeV	275 GeV	920 GeV
	38.4 mb	**38.4 mb**	**39.4 mb**	41.8 mb

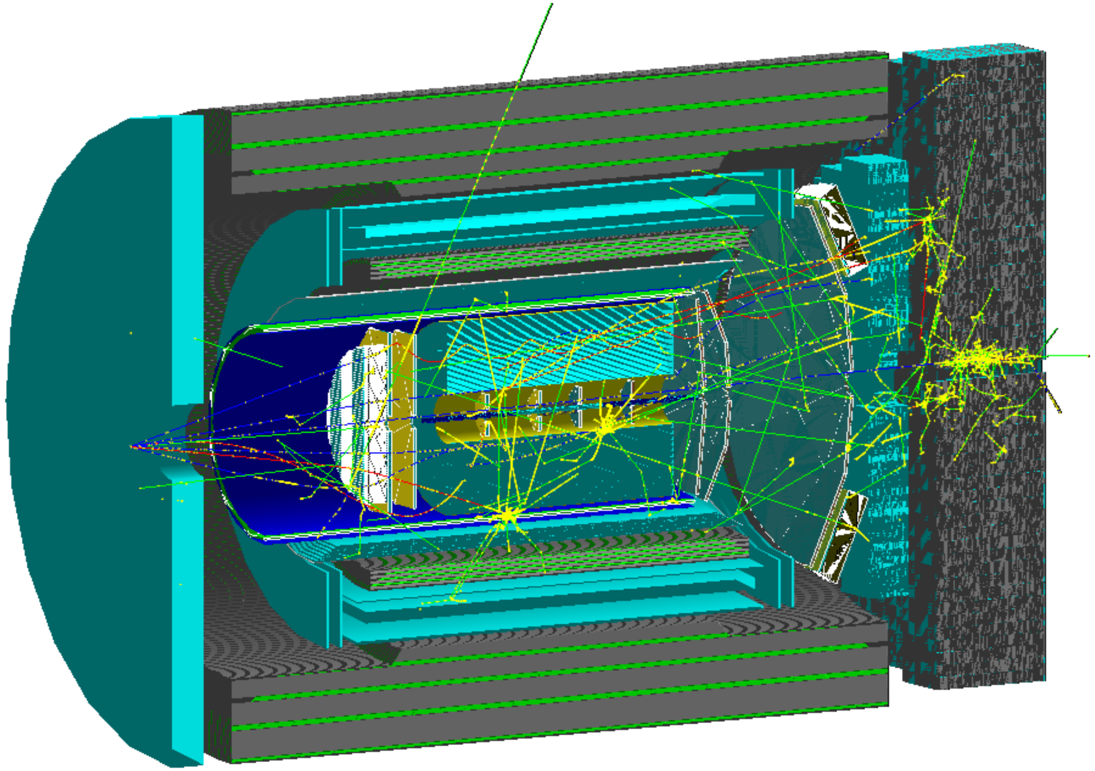


EIC DAQ in simulation

Refs: EIC CDR, sPH-cQCD-2018-001: https://indico.bnl.gov/event/5283/



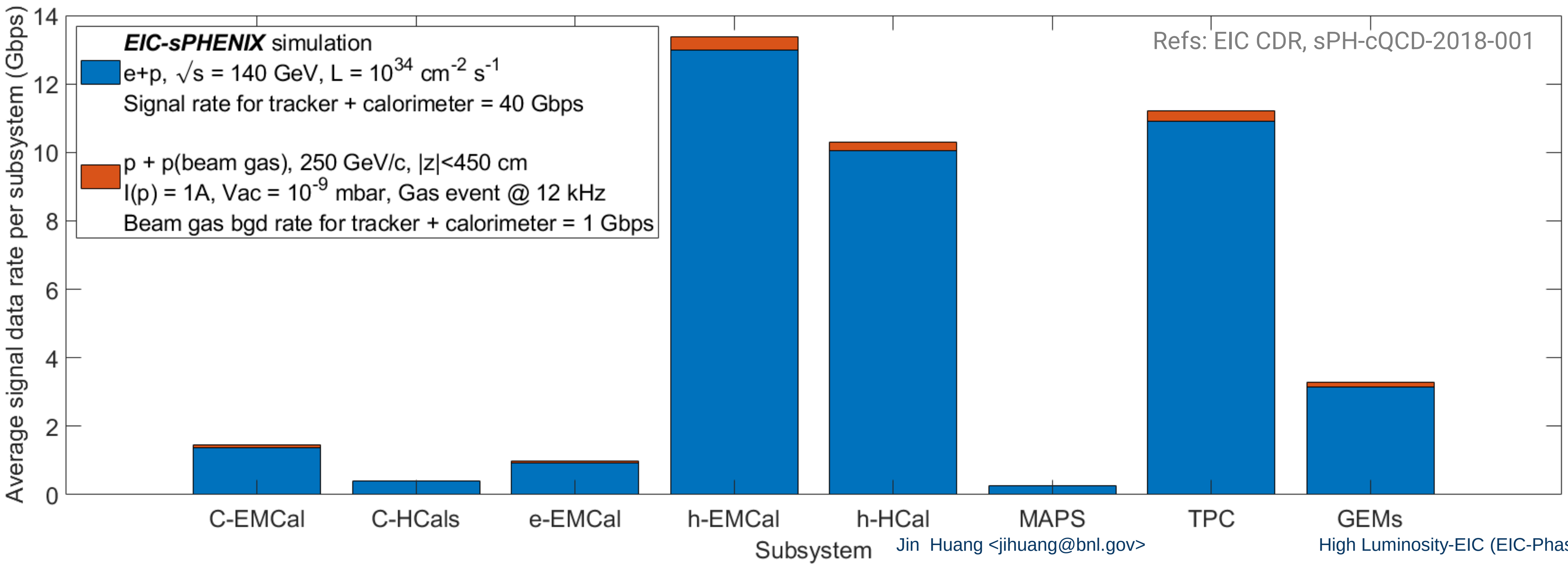
e+p DIS 18+275 GeV/c
$Q^2 \sim 100\ (\mathrm{GeV}/c)^2$



Beam gas event
p + p(gas), 275 GeV/c
at z=-4 m

Signal data rate -> DAQ strategy

- What we want to record: total collision signal ~ 1Tbps @ 10^{35} cm^{-2} s^{-1}
  - Will be high comparing today, but may in reach of recording in 2040s
- Therefore, we could (still) choose to stream out all EIC collisions data
  - In addition, DAQ may need to filter out excessive beam background and electronics noise, as they may be the dominant source.
  - (Still) very different from LHC/HL-LHC, which require saving only very small portion of collisions



Considerations for detector designs

- EIC is a high precision low interaction rate collider
  → low noise detector and low background experiment
- No L1 trigger would be sent to front-end. ASIC requires to operation in zero-suppressed data-pusher mode or continuous time-framed modes
- Synced with collider collision clock (98.5 MHz @ EIC-1)
- Factors require close attention in data rate of readout
  - Dark noise
  - Synchrotron background
  - Noise filtering

Since CDR

- Both ECCE and ATHENA proposal at EIC-1 followed similar strategy, being implemented for detector-1
- Detailed rate table also given in ATHENA proposal
  - Dominating rates are SiPM noise from RICH detectors
  - Require special filter processor (FPGA) using local trigger strategy
- Synchrotron background still under intense study [see link to detector-1 tracking meeting]

[ATHENA Proposal]

Table 2.5: Maximum data volume by detector.

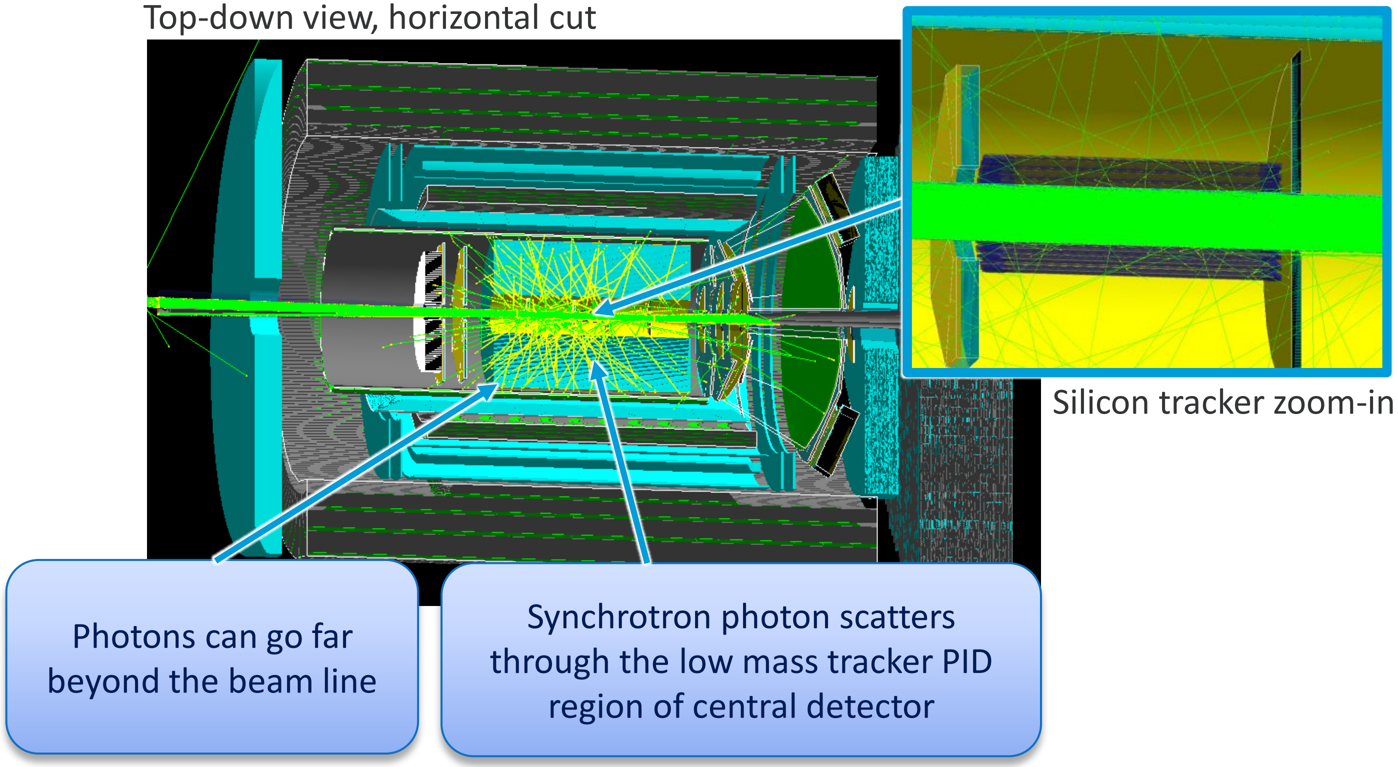
Detector	Channels	DAQ Input (Gbps)	DAQ Output (Gbps)
B0 Si	400M	<1	<1
B0 AC-LGAD	500k	<1	<1
RP+OMD+ZDC	700k	<1	<1
FB Cal	4k	80	1
ECal	34k	5	5
HCal	39k	5.5	5.5
Imaging bECal	619M	4	4
Si Tracking	60B	5	5
Micromegas Tracking	66k	2.6	.6
GEM Tracking	28k	2.4	.5
μRWELL Tracking	50k	2.4	.5
dRICH	300k	1830	14
pfRICH	225k	1380	12
DIRC	100k	11	11
TOF	332k	3	.8
Total		3334	62.9

Considerations for Synchrotron bgd.

- Synchrotron background is major challenge for high energy collider with electron beams
- Many detectors at EIC could be venerable to Synchrotron background
  - E.g. challenging for readout design, background filtering tracking, and fake large DCA for HF
- Strong emphasize on co-design of collider, IR and experiment that is low in Synchrotron background from the start:
  - General EIC R&D eRD21
  - bi-weekly IR background meeting joining accelerator and detector physicists
  - Initiative in Detector-1 collaboration to study

- 100k SynRad synchrotron photon by M. Stutzman
- Reproduce this Geant4 simulation from GitHub: macros / SynRad->HepMC reader

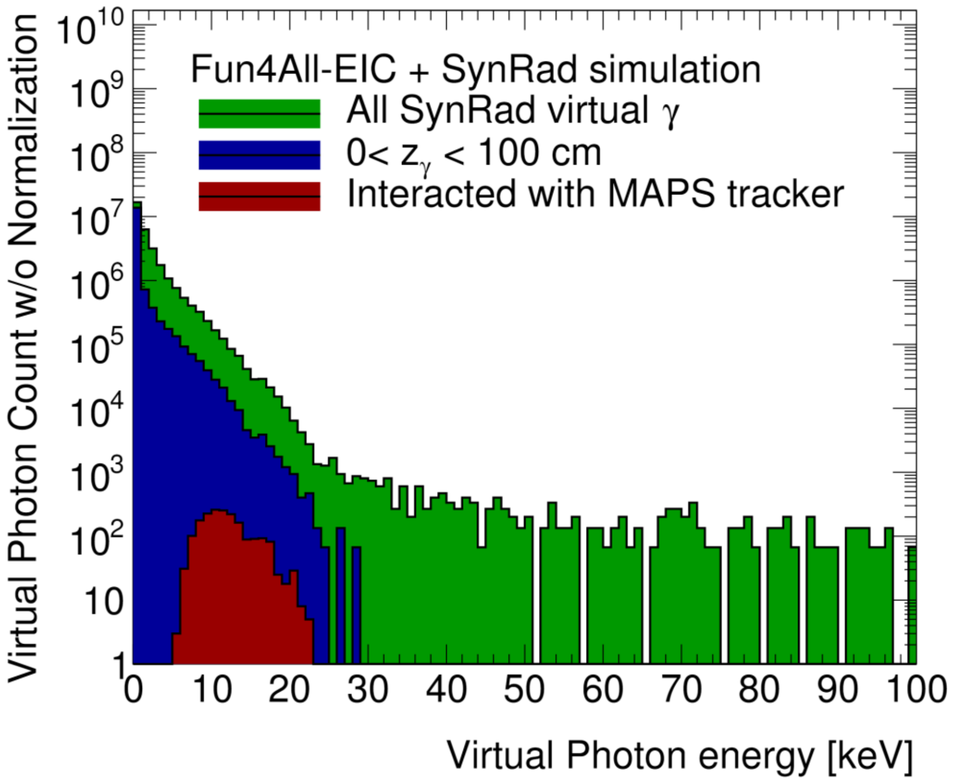
Top-down view, horizontal cut



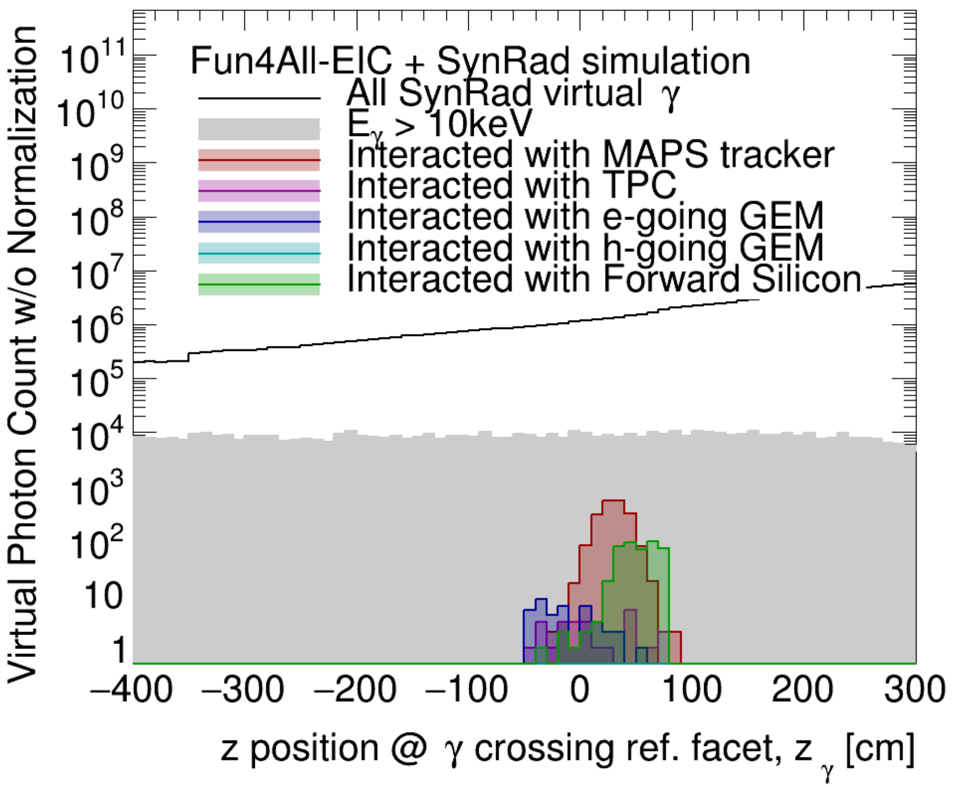
Silicon tracker zoom-in

Synchrotron background: detector response

- Iterating with accelerator design to avoid 10keV photon that exits -50 to +100cm from beam pipe



Energy dependence of MAPS vertex tracker to synchrotron



Beam-pipe exit-location

Background outlook for SRO@EIC

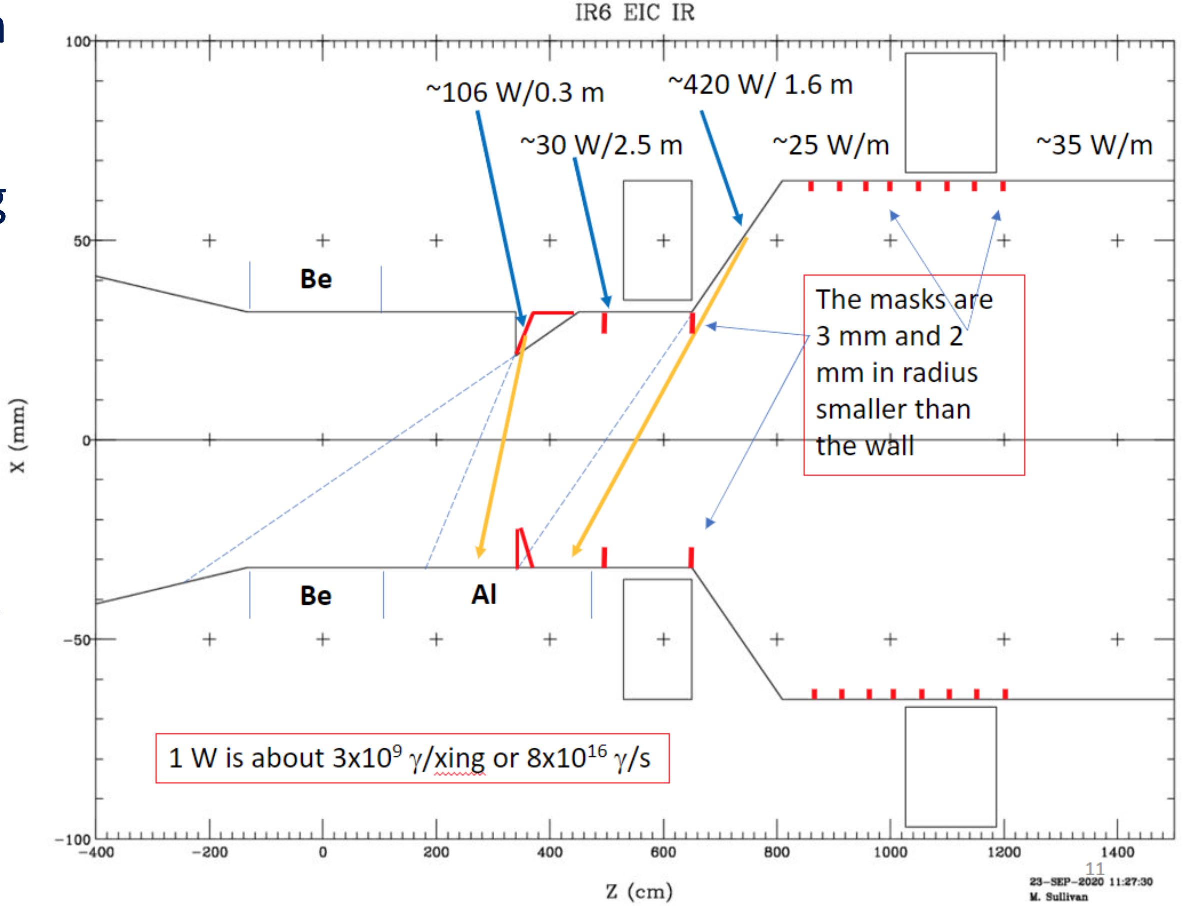
Synchrotron background is likely remaining concerning and undetermined

- As both machine and experimental region design evolves
- Prepare for the case of a large background, in particular at initial ops./after upgrade and tuning

Remedy strategies:

- Trigger-SRO hybrid:
  - e.g. use calorimeter-based fixed latency trigger, and use it to throttle SRO data
- Digital real-time background filtering:
  - e.g. building features (tracks, clusters, wavelet fits):
- Validate any data reduction with (near) real-time QA, reconstructing as much data as possible

SR Background shielding optimization
[M. Sullivan, Oct 2020 EIC SR meeting]



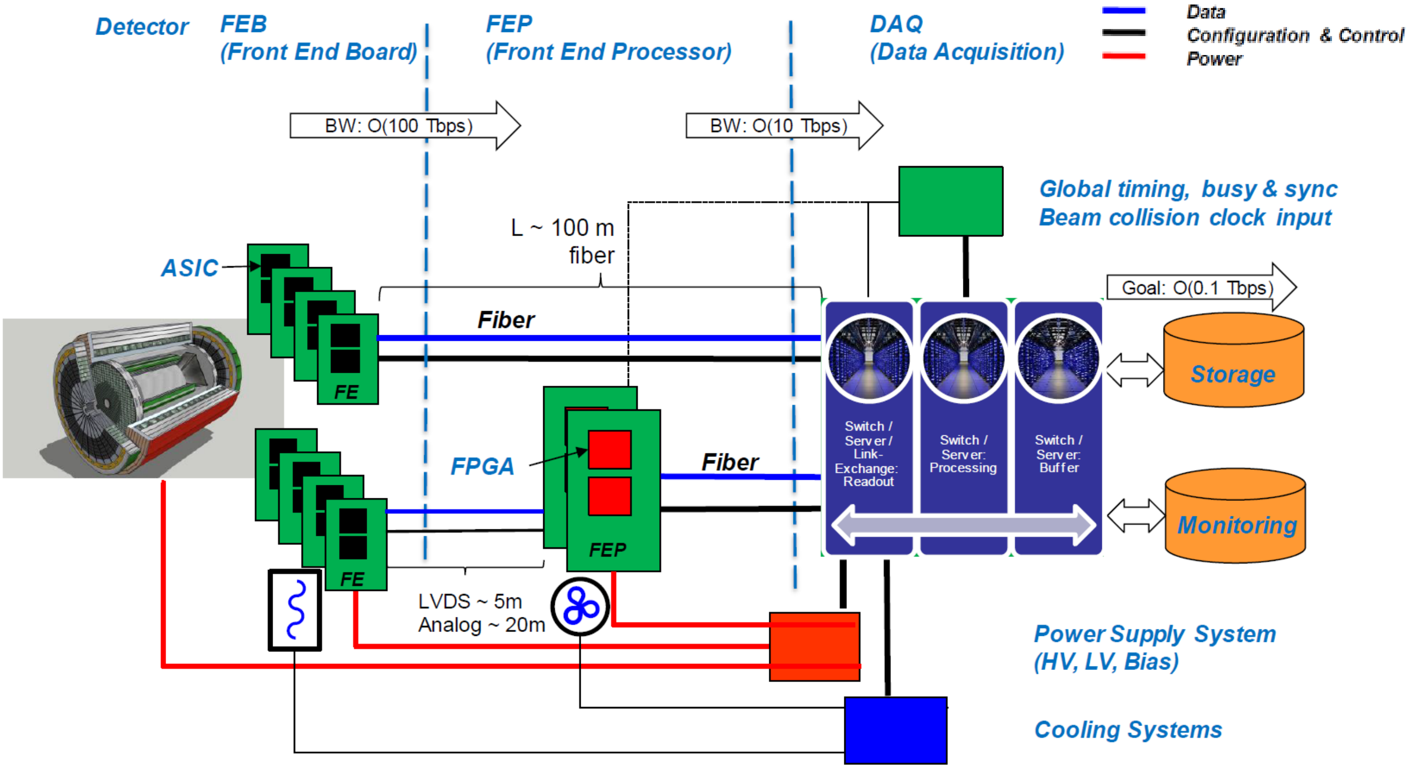
Discussion on the DAQ strategy

- In reference to considerations behind EIC-1 detectors

Brookhaven National Laboratory

Strategy for an EIC real-time system

- Triggerless readout front-end (buffer length : µs)
- DAQ interface to commodity computing (equivalent to FELIX/CRU). Background filter if excessive background rate
- Disk/tape storage of streaming time-framed zero-suppressed raw data (buffer length : s)
- Online monitoring and calibration (latency : minutes)
- Final Collision event tagging in offline production (latency : days+)



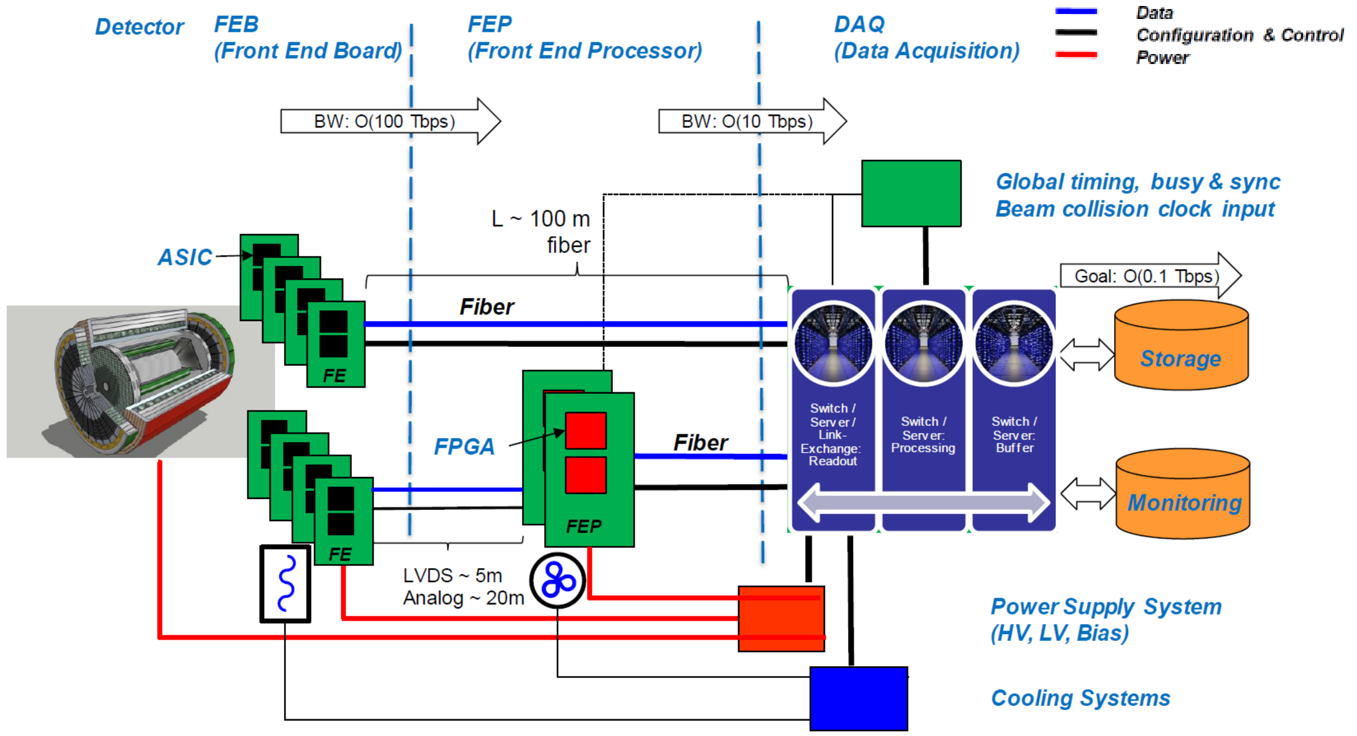
Ref: EIC-CDR, Bandwidth likely increase by x10 in EIC2

Real-time computing for streaming data pipeline

- Despite low signal rate, the raw data rate can be filled with noises and background
  - Need low background & low noise detector & electronics design
- An essential job of EIC real-time computing: reliable streaming data reduction to fit permanent storage (next topics)
- And more traditional roles for online/offline server farm:
  - Online monitoring/fault det.
  - Calibration
  - Production → Initial analysis pass

ASIC, FPGA FPGA, xPU

Digitizer BW: 10^3 Tbps → Readout BW: 10^2 Tbps → Storage: 1 Tbps

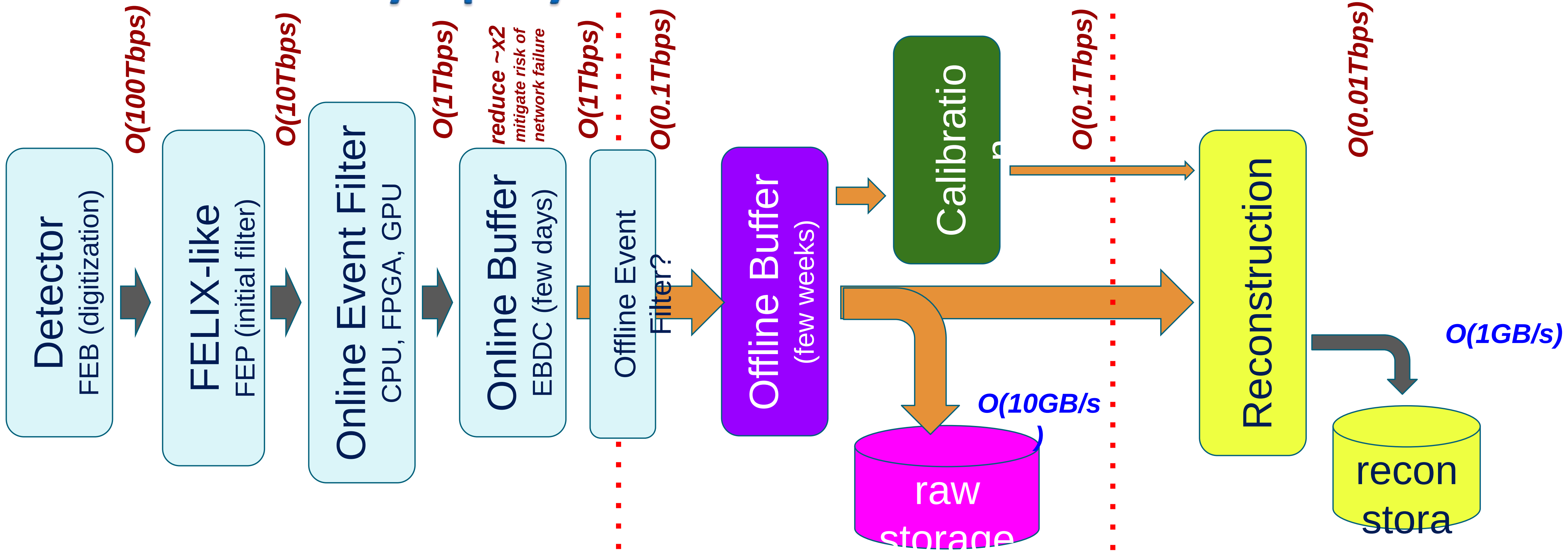


[EIC CDR]

ECCE (online) computing model

Bandwidth likely up by x10 for EIC-2

Countesy: David Lawrence
ECCE computing model [link]



Experimental Hall and Counting House (Project Funds)

Data Center(s): SDCC [,JLab, …] (Operations Funds)

HTC Compute Facilities SDCC ,JLab, … (Operations)

Selected Novel hardware opportunity on horizon for EIC-2

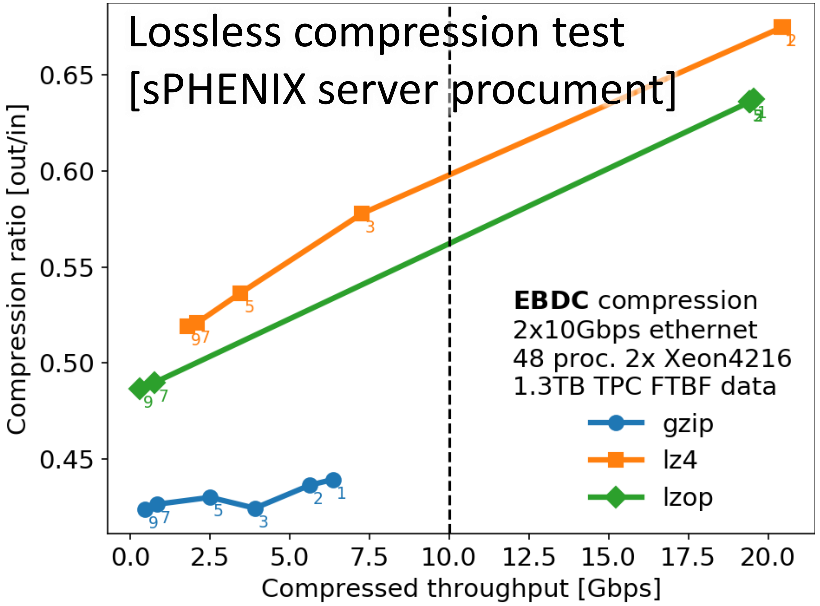
- Novel algorithms
- Novel AI-computing accelerators
- Power-efficient front-end processing

The main challenge: reliable data reduction

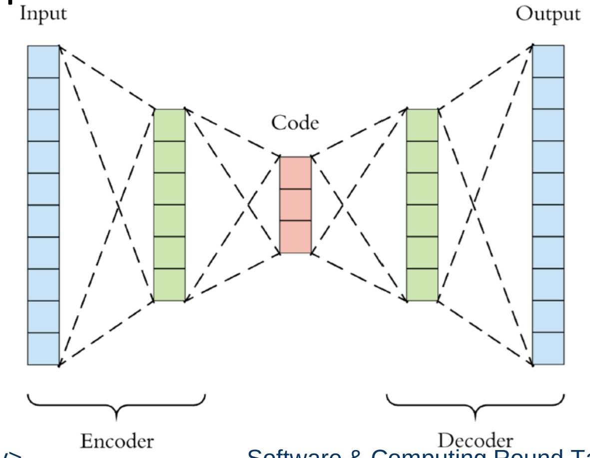
- Both EIC1&2 has much lower collision signal data rate comparing to LHC
  - But background is important and can be dominating
  - We DO NOT want to drop any event for systematic uncertainty control and broader physics interests
- Opportunities for Real-time AI application, e.g.
  - Lossy compression of data, noise filtering
    - E.g. Y. Huang @ AI4EIC workshop [link]
  - Feature extraction: Energy time extraction from ADC time-series, cluster, tracklets
    - E.g. S. Miryala @ CPAD 2021 [link]
  - High level reconstruction: full tracking, vertexing, PID, HF signal selection
    - D. Yu @ AI4EIC workshop [link]
    - S. Furletov @ streaming workshops [link]

Online computing for streaming data - compression

- Lossless compression
  - Compress by ~1/2
  - Well established fast compression algorithm
- Lossy compression/noise filtering
  - Opportunity for unsupervised machine learning based on data, e.g.
  - Auto-encoder on ASIC for HGCal @ CMS [link], CLAS12 denoise [link]
  - Bicephalous Convolutional Neural Encoder for zero-suppressed data (next)



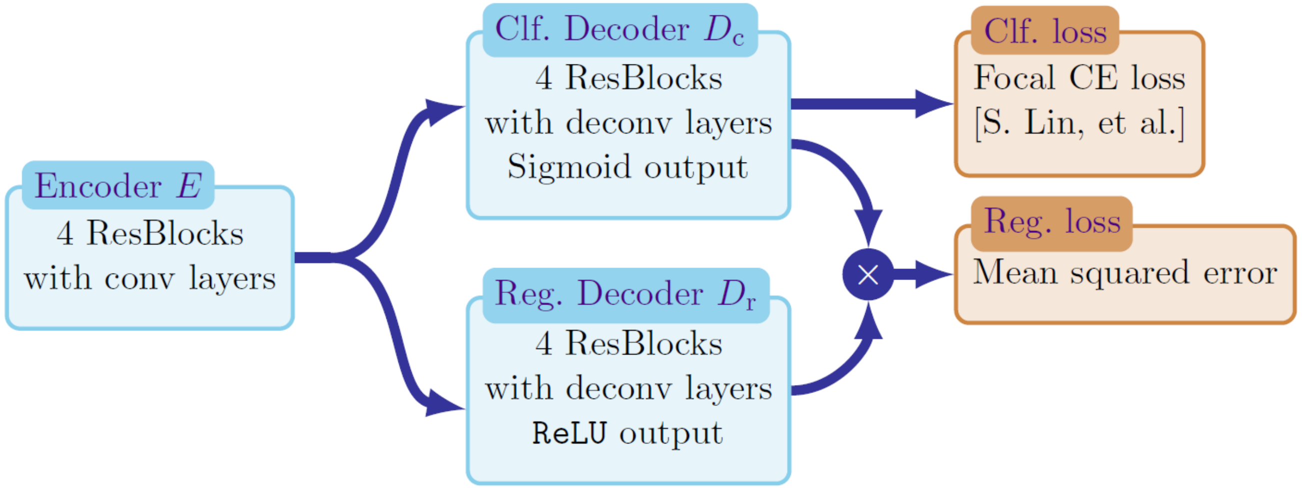
Simple auto-encode neural network



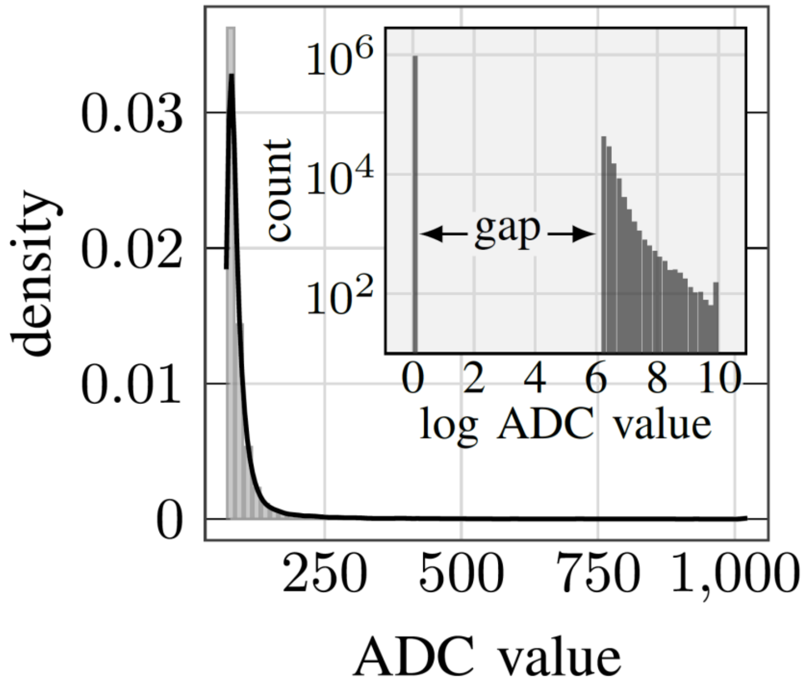
Bicephalous Convolutional Auto-Encoder for zero-suppressed data

- Some detector ADC data is challenging for Auto-Encoder, e.g. features such as zero-suppression cut off
- A dual-output auto encoder is designed to output both a region of interest and decompressed ADC. Possibility for further noise filtering (Bicephalous Convolutional AutoEncoder/BCAE)
- Ref: Y. Huang [arXiv:2111.05423, link to talk]

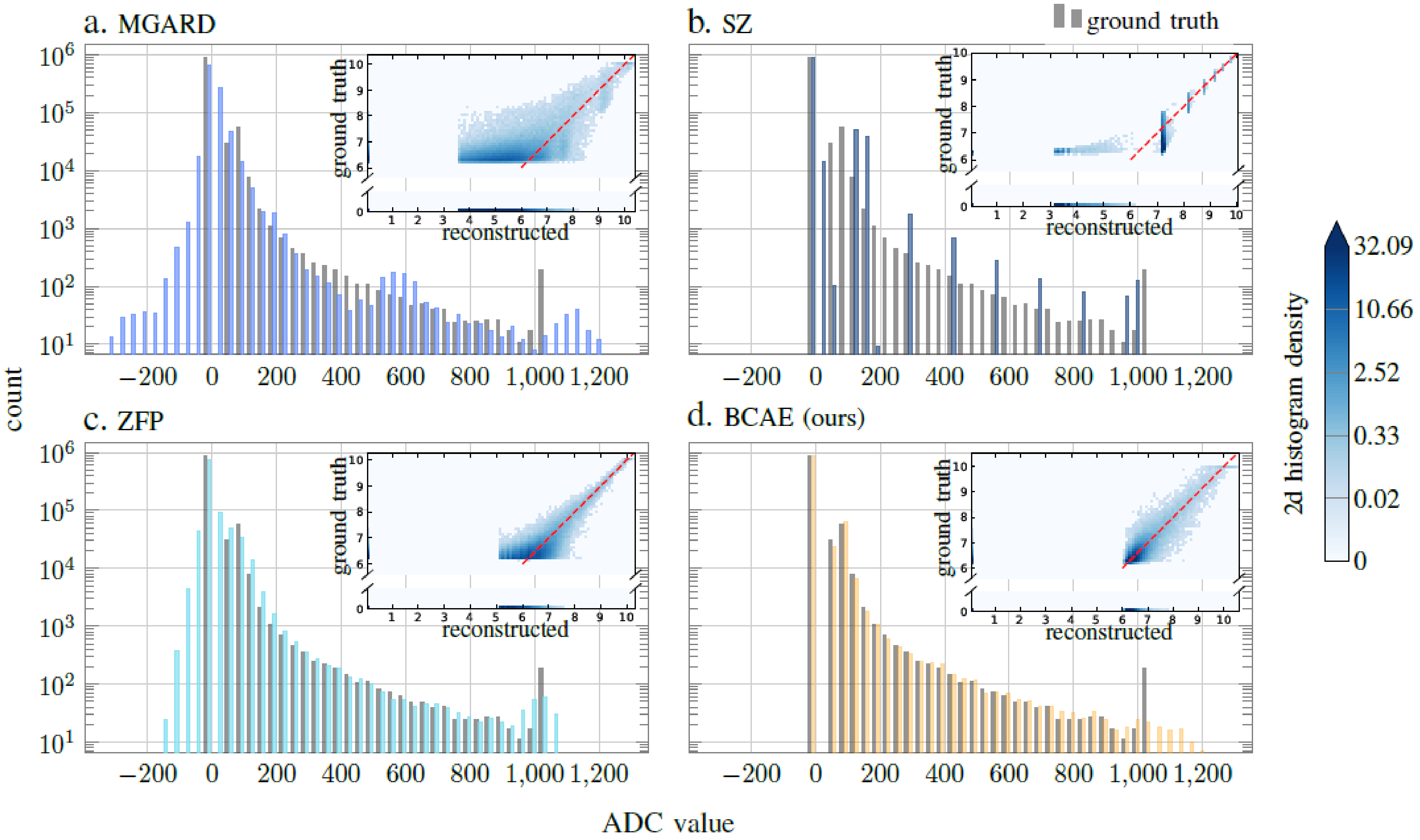
Bicephalous Convolutional AutoEncoder (BCAE)



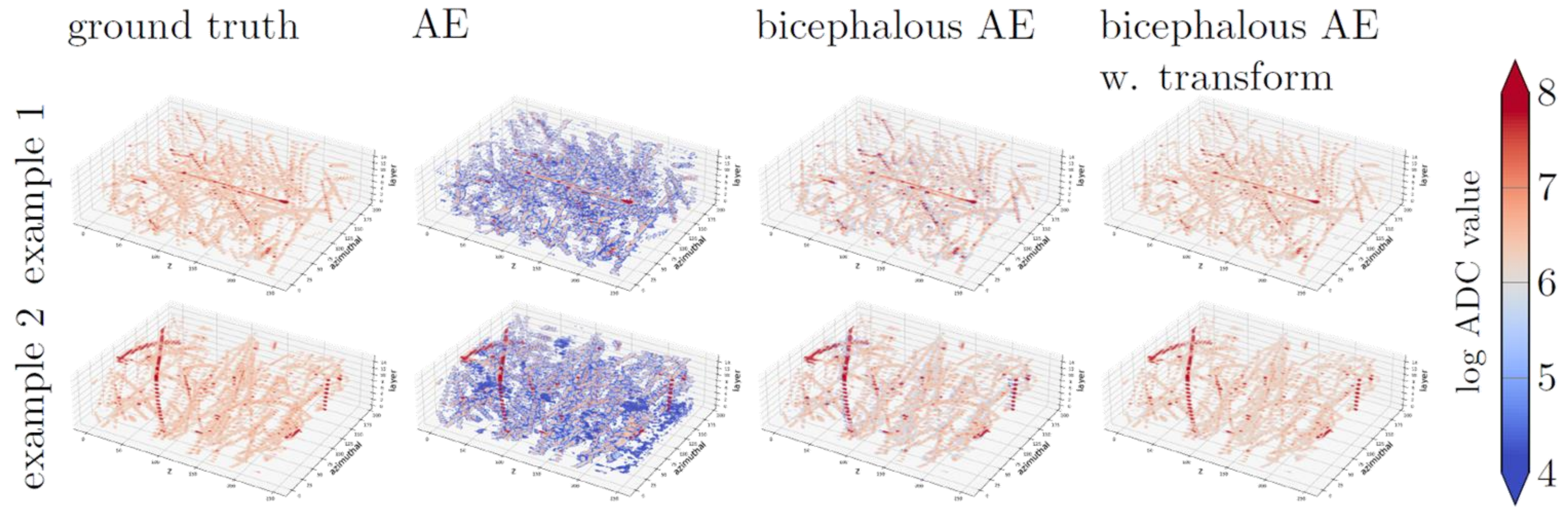
Example zero-suppression cut off on ADC
arXiv:2111.05423



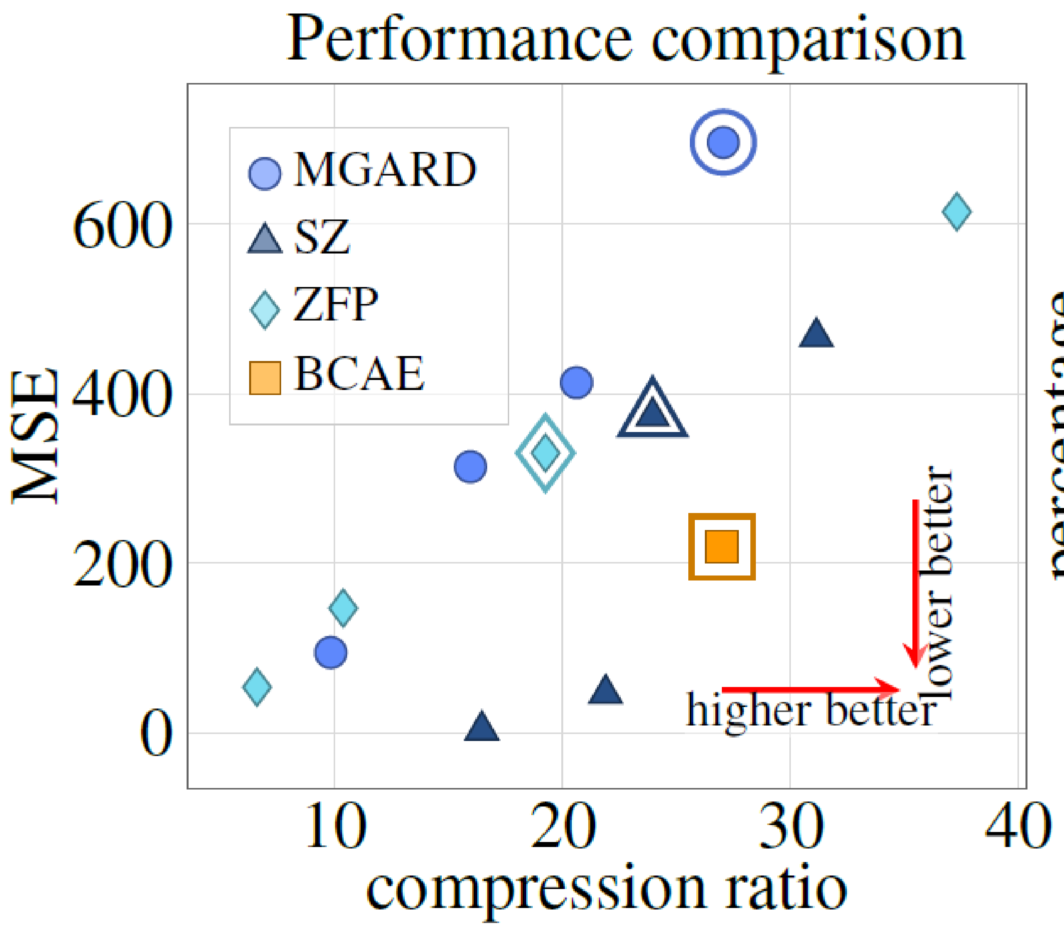
Comparison with existing algorithm [arXiv:2111.05423]

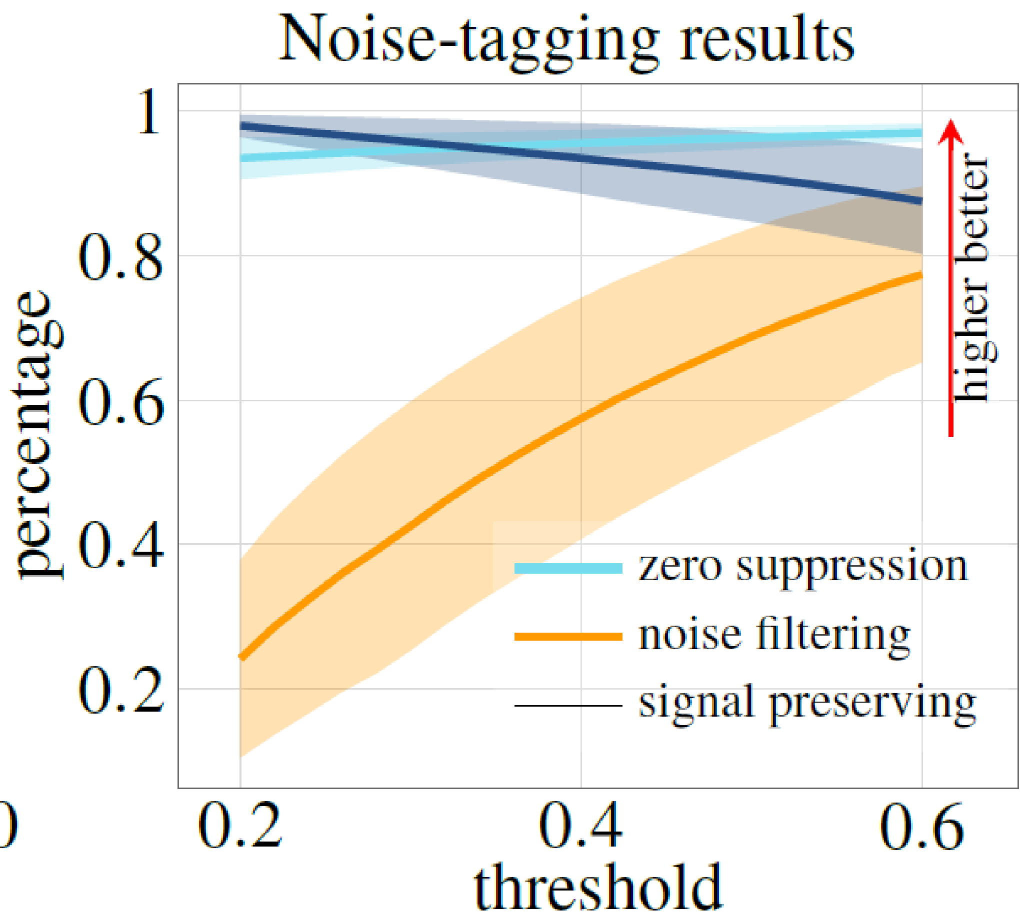


Results from Bicephalous AE with transform [arXiv:2111.05423]



BCAE: compression and noise filtering

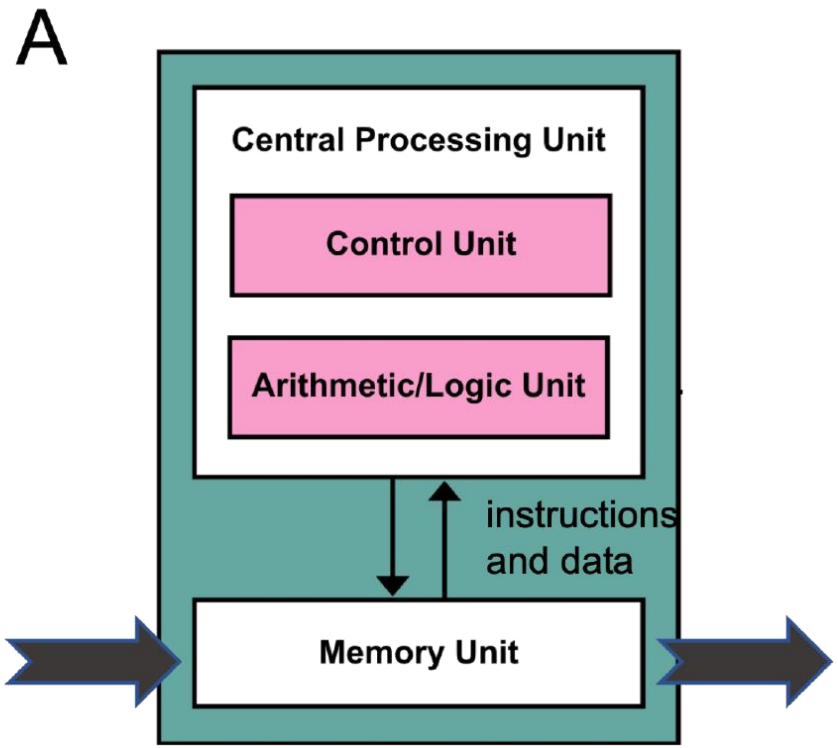


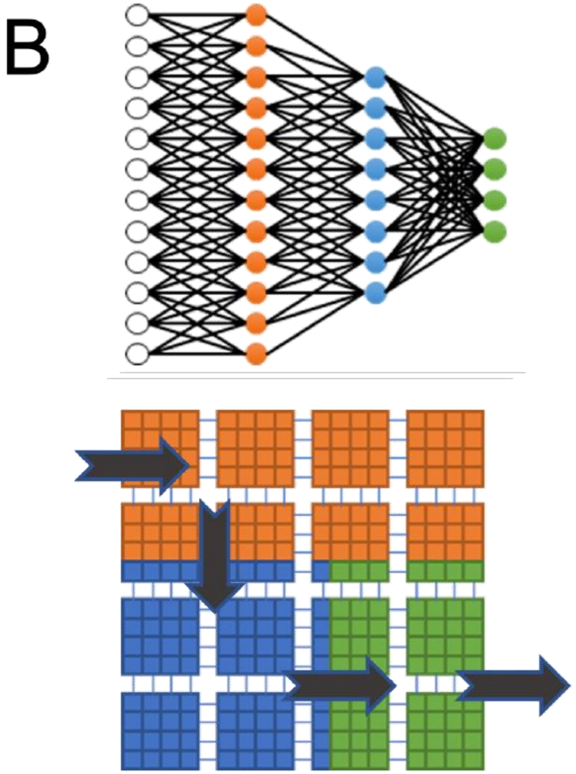


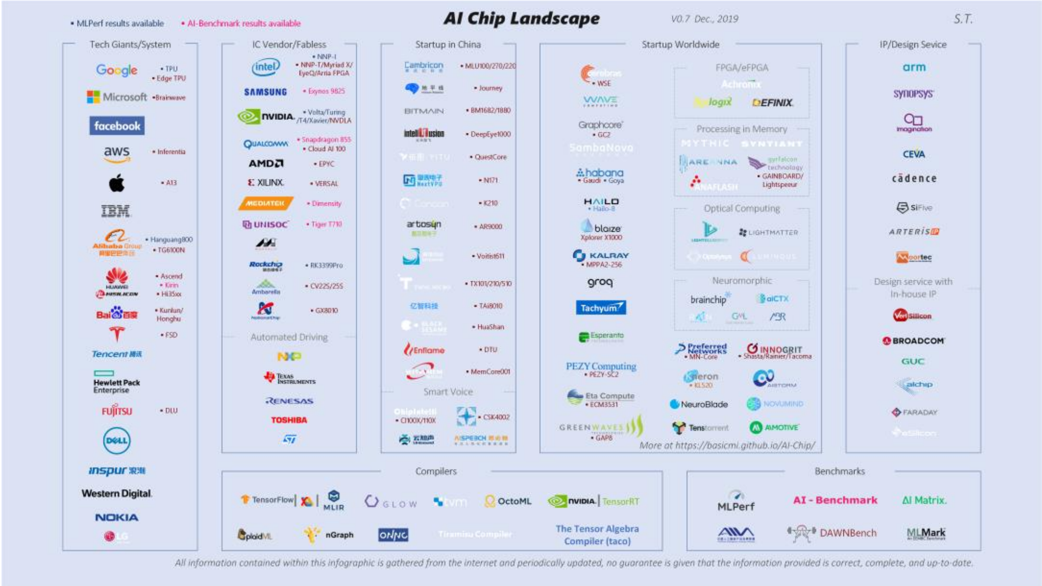
Novel processors (AI chips) for deep neural network

- Traditional processors (CPU/GPU) has built-in inefficiency for high throughput data processing for neutral network
- Addressed by a large family of AI-chips from a diverse vendors. Could be a mature solution at time of EIC-2
- Early test indicating promise processing sPHENIX streaming data (no trigger throttle, EIC2 data rate)

von Neumann Architecture. Dataflow Architecture [source]





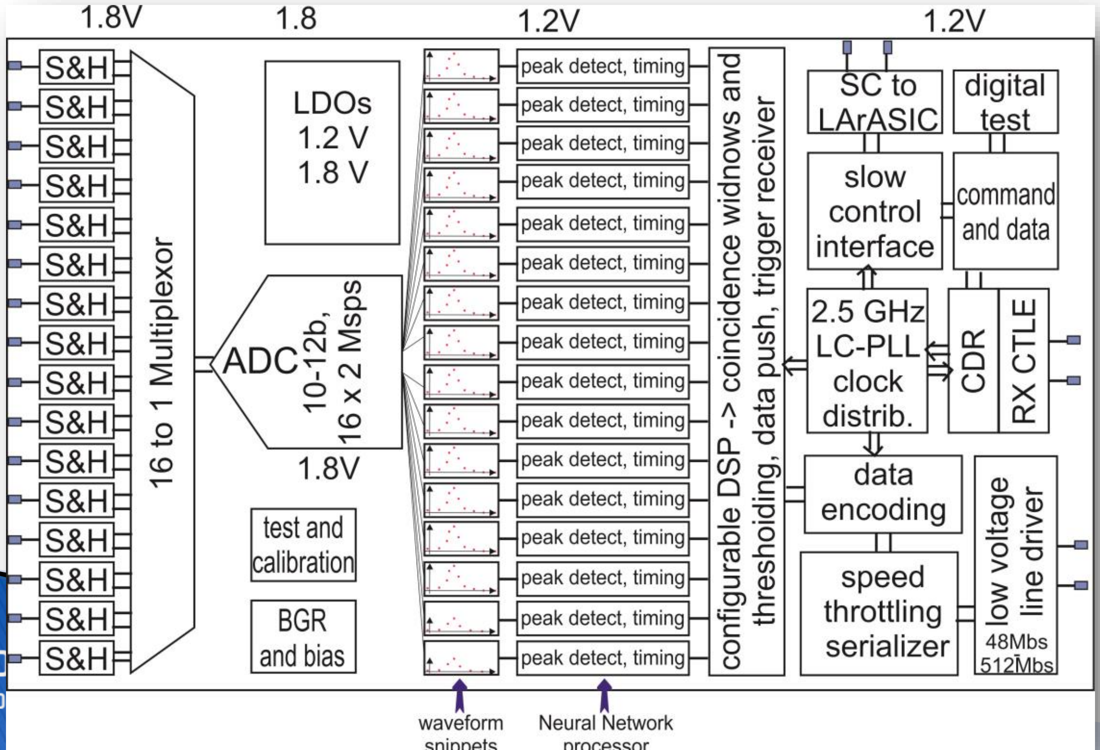


https://basicmi.github.io/AI-Chip/

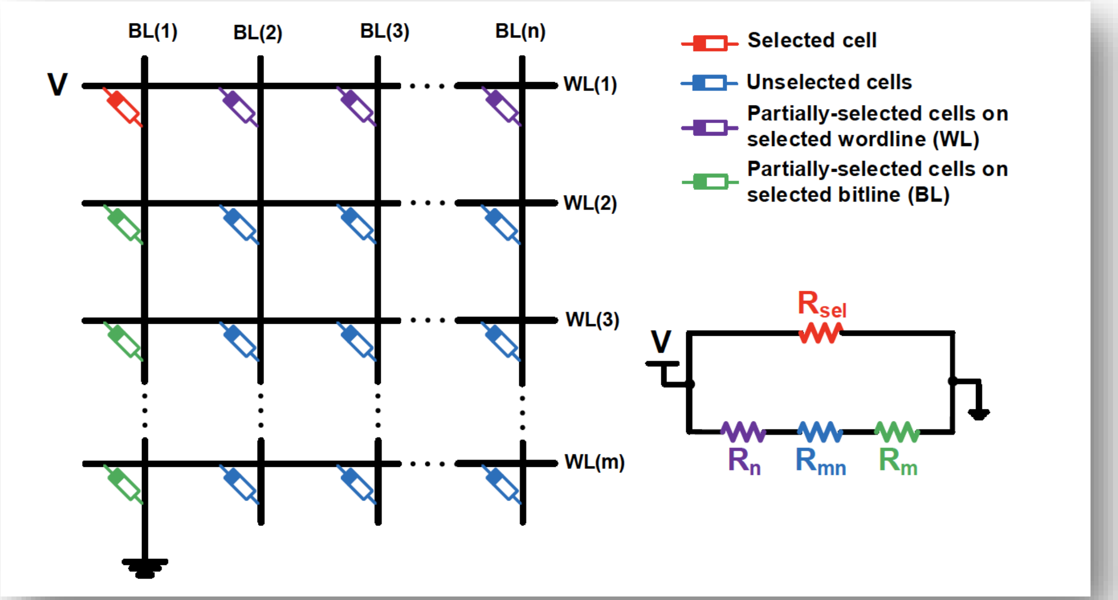
Novel front end processors for deep neural network

- To address data reduction at source (Front-end) → low power consumption → more deviation from traditional digital electronics
- Examples are at-memory analog computation, including Memristor, Electro-Photonics
- On-going explorations in both categories

S. Miryala @ CPAD 2021 [link]



Map neural networks onto memristor crossbar array [link]



Summary

- At 10^{35} cm^{-2} s^{-1} , EIC2 collision signal rate, O(1)TBps, is not far from RHIC (sPHENIX prior to trigger throttling)
- Nonetheless, for both EIC1 and EIC2, DAQ challenge lay in noises and background, e.g. synchrotron radiation → drive need on data reduction
- At horizon of EIC2, novel algorithm and hardware may be deployed at scale to efficiently process data in real-time: e.g. data flow and at-memory processor for high throughput AI inference.

Extra information

Data Rate

e+p, Pythia6 Q2>0

p+p(gas) Pythia8

MAPS silicon tracker

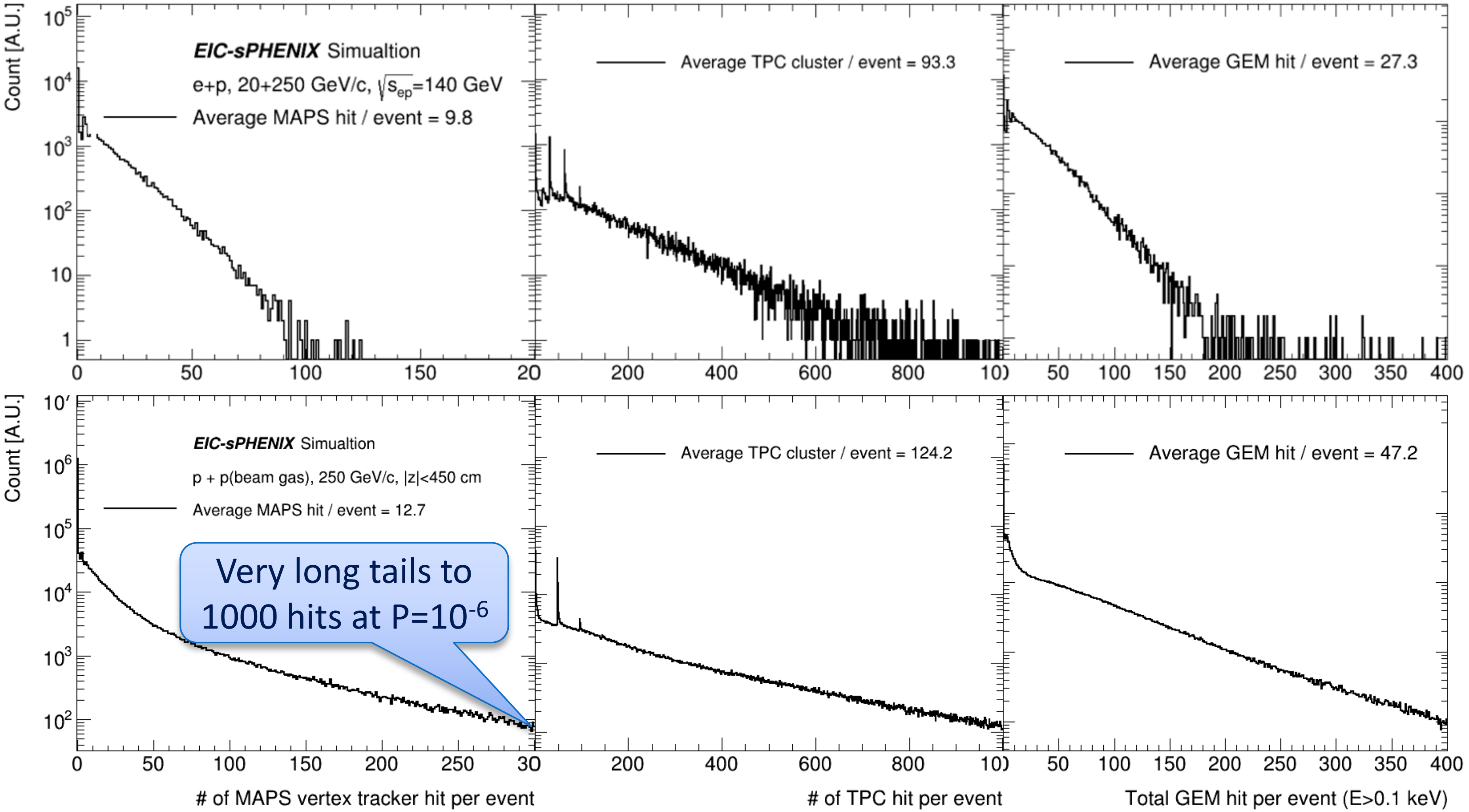
Raw data: 16-24 bit / MAPS hit
(3-layer ALPIDE model)

TPC

Raw data: 3x5 10 bit / TPC hit
+ headers (60 bits)

Forward/backward GEM

Raw data: 3x5 10 bit / GEM hit
+ headers (60 bits)

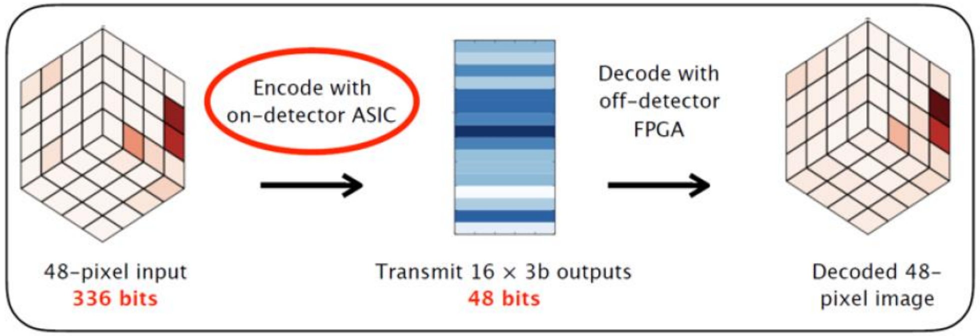


Refs: ECI CDR, sPH-cQCD-2018-001: https://indico.bnl.gov/event/5283/

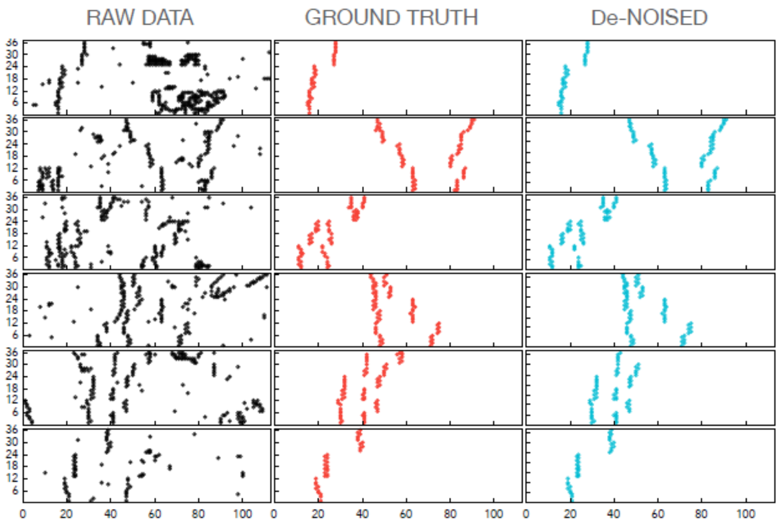
Lossy compression of data, noise filtering

- Auto-encoder (AE) is a natural choice for unsupervised learning for lossy data compression: streaming data reduction
- Same network architecture can be adopted with supervised learning to filter out noise: further data reduction, speed up reconstruction
- We are not alone in this research: see also in CMS HGCal ASIC, CLAS12 tracker offline reco.

CMS HGCal compression ASIC, [10.1109/TNS.2021.3087100]

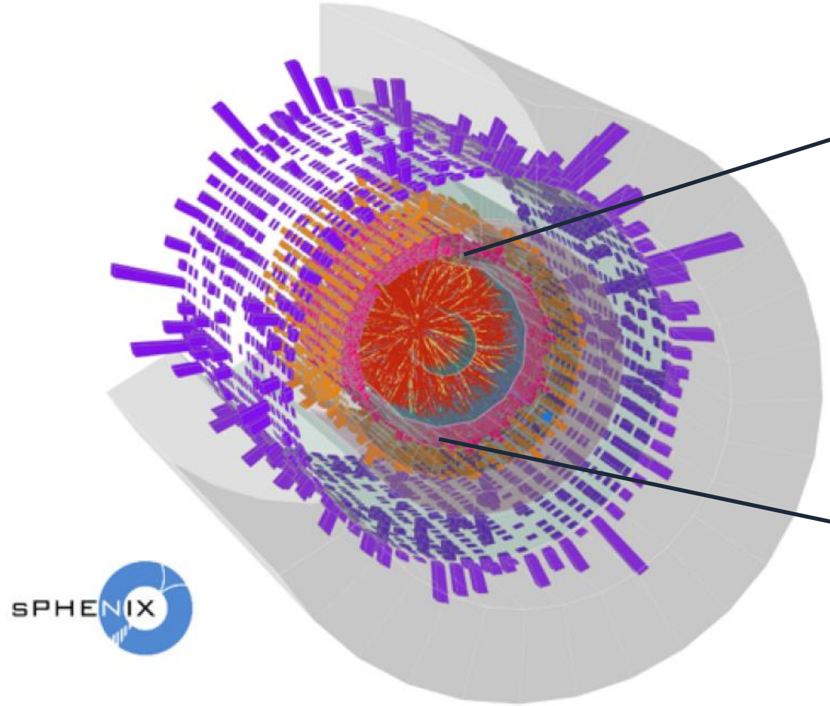


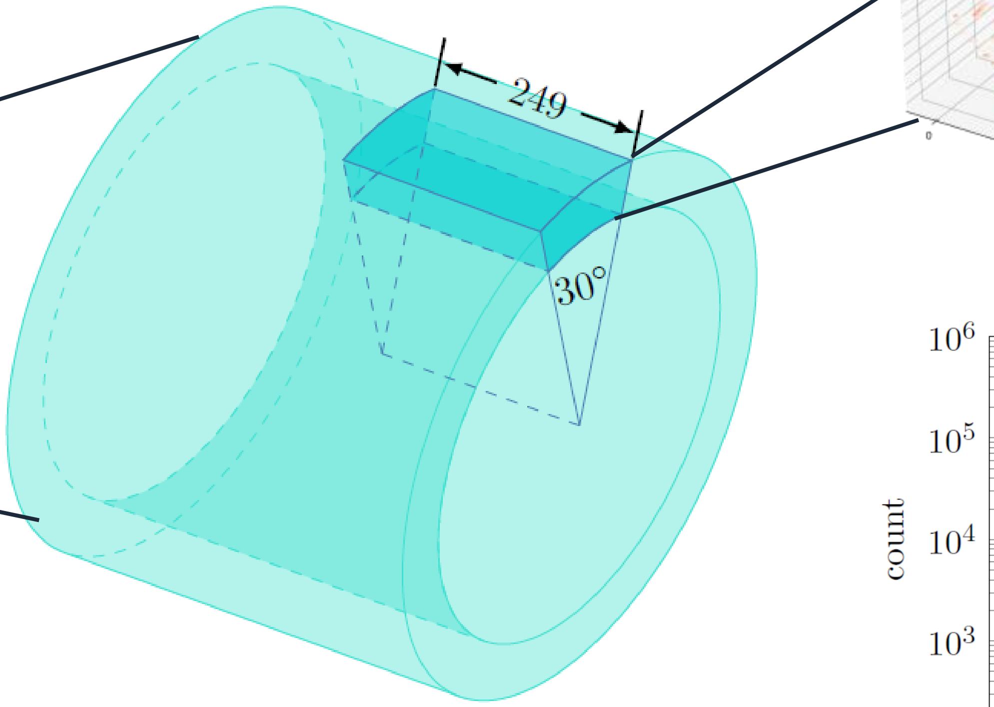
CLAS12 Drift Chamber offline AE de-noise [link]

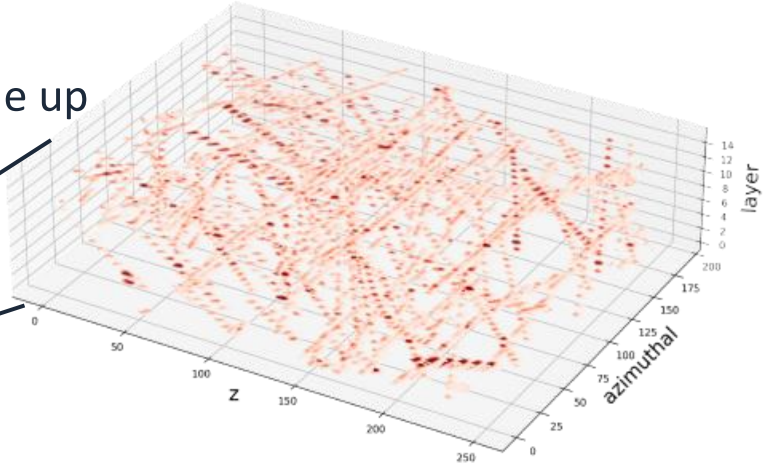


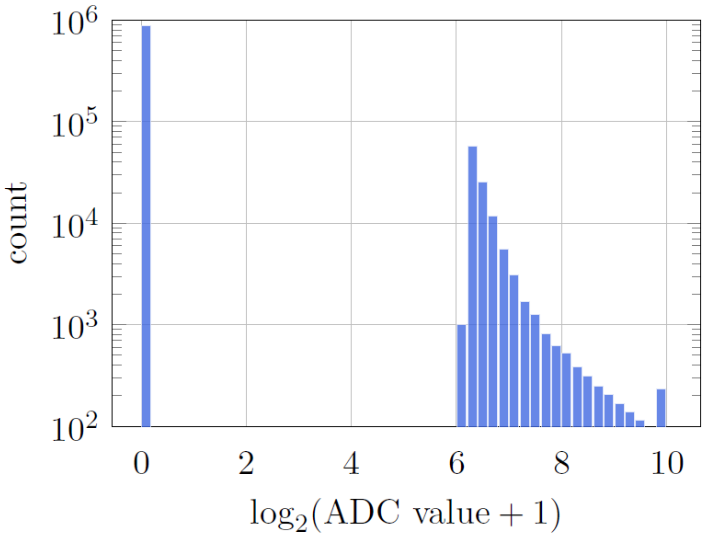
Data of time projection tracker at sPHENIX

Busiest event in sPHENIX TPC
10% central Au + Au collision with 170kHz pile up
Data frame for 1/12 azimuth sector







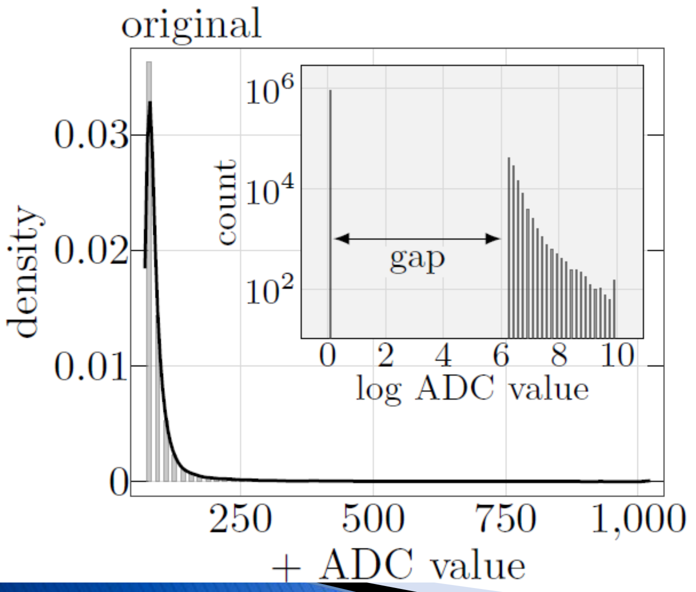


Bicephalous Convolutional Auto-Encoder (BCAE) and input transform [arXiv:2111.05423]

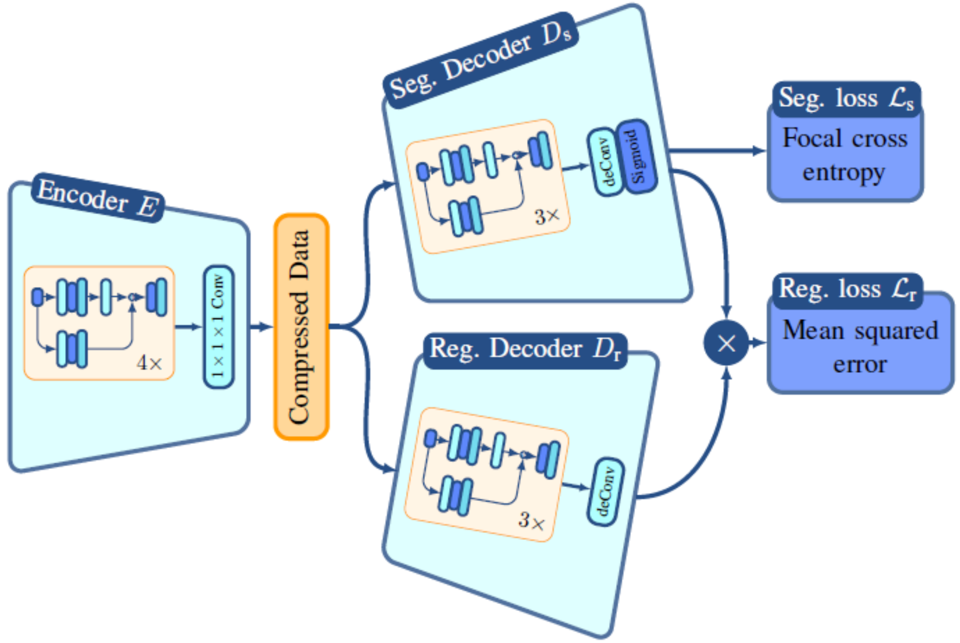
- Input transform: fill in the zero-suppression gap and make ADC distribution much less steep
- Bicephalous decoder: +classification decoder to note the zero-suppressed ADC bins (unsupervised training) and +noise bins (supervised training)

Input transform: $\mathcal{T}(x) = \log(x - 64)/6, \qquad x > 64$

Inverse transform: $\mathcal{T}^{-1}(y) = 64 + \exp(6y), \qquad x \in \mathbb{R}$



a. BCAE architecture summary



Synchronization to beam crossing

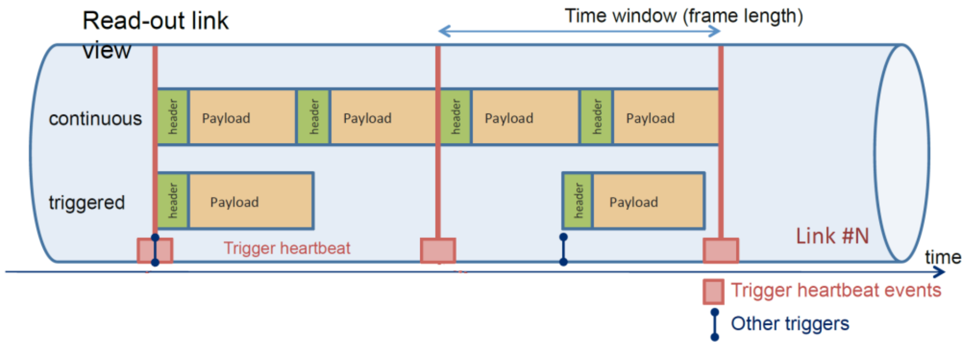
- For a collider experiment, absolute time is not directly useful (e.g picoseconds count to some reference time)
  - ms would be sufficient for calibration tracing
- What matters is relative time:
  - Hit correlation between detectors at bunch separation O(10ns) , time of flight clock distribution uncertainty requirement at O(10ps)
  - Correlation of detector hit with the origin of the beam bunch crossing at O(10ns) →
- Why tagging streaming hits with beam bunch crossing (98.5Mhz)
  - Necessity to trace back to physical quantity that could change between neighboring bunches: spin state, luminosity, polarization, etc.
  - "Free" boost of precision from 10ns counter to 100ps bunch interaction time

EIC beam crossing clock features

- 1260 RF bunch, 98.5254 MHz beam crossing clock
  - In a friendly range for many FPGA-optical transceivers base clock: e.g. 6Gbps transmit 8Byte per beam clock
  - E.g. sPHENIX beam clock was envisioned to transmit at 56.4 or 112.8MHz (56.4 was final choice)
  - Upstream signals: beam clock, revolution tick tag bunch zero
- Clock signal originating from accelerator RF source
  - High precision clock source for operation of RF cavities (rebuilding RHIC 28 MHz RF cavities as 24.5 MHz resonators), which in turn defines the bunch in time
  - High precision clock signal routed through the accelerator site, installed for sPHENIX operation
- EIC beam clock has fixed frequency without ramp variation [EIC CDR sec 3.3.2]:
  - Electron beam does not change frequency during acceleration
  - Hadron beam change *orbit* to match electron beam frequency
  - Simplification compared with hadron colliders:
    - RHIC (9.34-9.38 MHz) and LHC (40.078422 to 40.078970 MHz), tricky configuration for transceiver and PLLs
  - Caveat: possibility of change of frequency operation poitn for low sqrt-s EIC operation, TBD

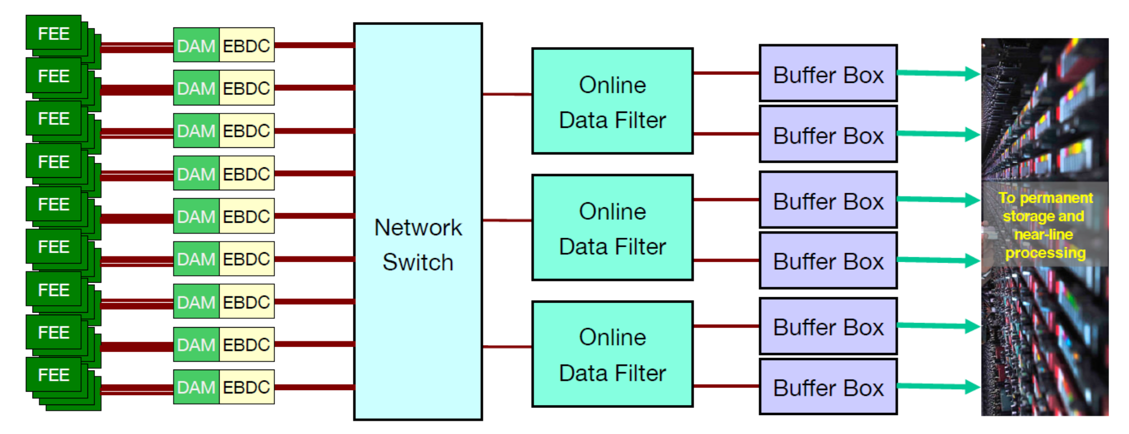
Time bucketing

Example:
ALICE heartbeat frame
O(20)ms [ALICE-TDR-019]



- Hit grouping in time period (instead of event trigger in traditional DAQ)
  - Used by many streaming system: e.g. heartbeat frame in ALICE, time slice in CBM
  - Efficient time encoding, i.e. each hit do not need to encode full beam clock counter, just relative counter inside a time bucket
  - For downstream processing, data can be processed and combined in chunks of 1-2 time bucket to obtain a complete set of event; unit for bookkeeping (calib/QA)
  - Easier handling in simulation sample preparation
- For EIC, we could *consider* a time frame of
  - 2^16 = 65536 beam crossings (→16-bit beam crossing time tag for hits within a bucket)
  - 0.7ms in absolute time
  - At top luminosity, containing about 300 collisions and 30 beam gas interactions
  - Convenient size of 8MB of raw zero-suppressed collision signal data + background data load

Envisioned system in EIC proposals



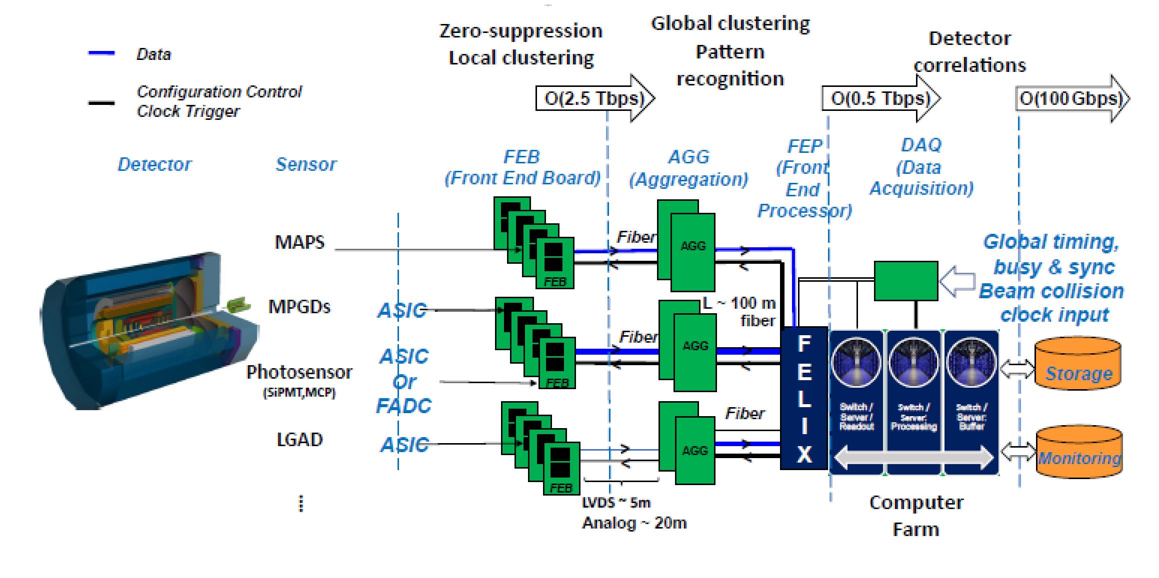


Figure 2.16: Overview of the ATHENA DAQ architecture.

ECCE DAQ architecture
[proposal, Talk by M. Purschke]

ATHENA DAQ architecture
[proposal, Talk by J. Landgraf]

FPGA-PCIe interface

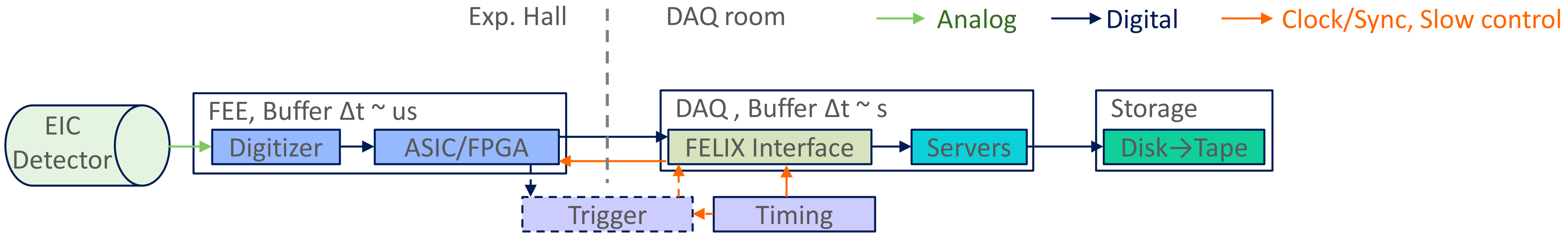
- Dedicated transceiver to allow receive timing and fast control channel and provide busy feedback
- Jitter cleaner (next slides)
- Send clock and control to FEE via optical links
  - GBT family ASIC
  - Custom SFP+ link encoding clock + data
- As EIC version of FELIX-like device develops, EIC could benefit from a large pool of current FELIX (712v2) for large scale testing after the sPHENIX operation concludes

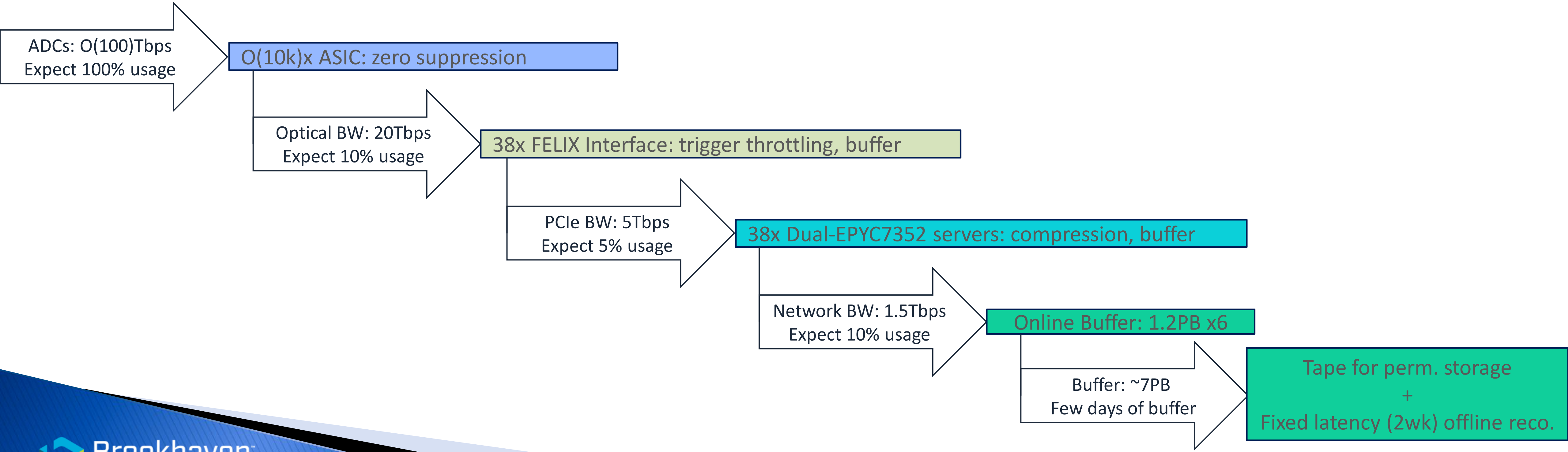
FELIX in develop [FLX-181, H. Xu SRO IX]

FELIX 712v2, ATLAS Phase-1/sPHENIX production

PCIe40, LHCb/ALICE, [N. Neufeld]

sPHENIX Streaming data flow





Streaming readout status at sPHENIX

- All three sPHENIX tracking detector uses streaming readout
- Developed plan to take streaming data for heavy flavor physics program (next slides), commended by RHIC PAC.
- Completed construction of sPHENIX FELIX DAQ interface (~50) and procurement of DAQ servers, network infrastructure and online disk buffers
- Data taking start in 2023!

RHIC PAC 2020 report

> We commend sPHENIX for developing the continuous streaming readout option for the detector, which increases the amount of data that can be collected in Run-24 by orders of magnitude. In particular in the sector of open heavy flavor, this technique will give access to a set of qualitatively novel measurements that would otherwise not be accessible. Given the tight timeline for completing the RHIC physics program before construction of the EIC begins, this is a tremendous and highly welcome achievement.

TPC data stream in sPHENIX triggered DAQ