Data rates for current EIC Detector concepts

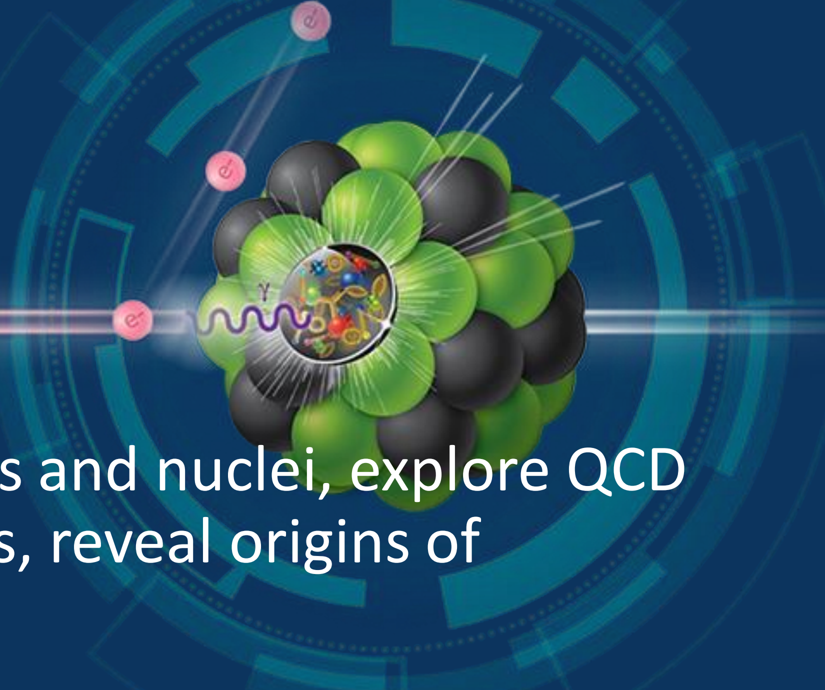
Outline: ● Uniqueness of EIC ● Data rate estimation ● Comments on DAQ strategy

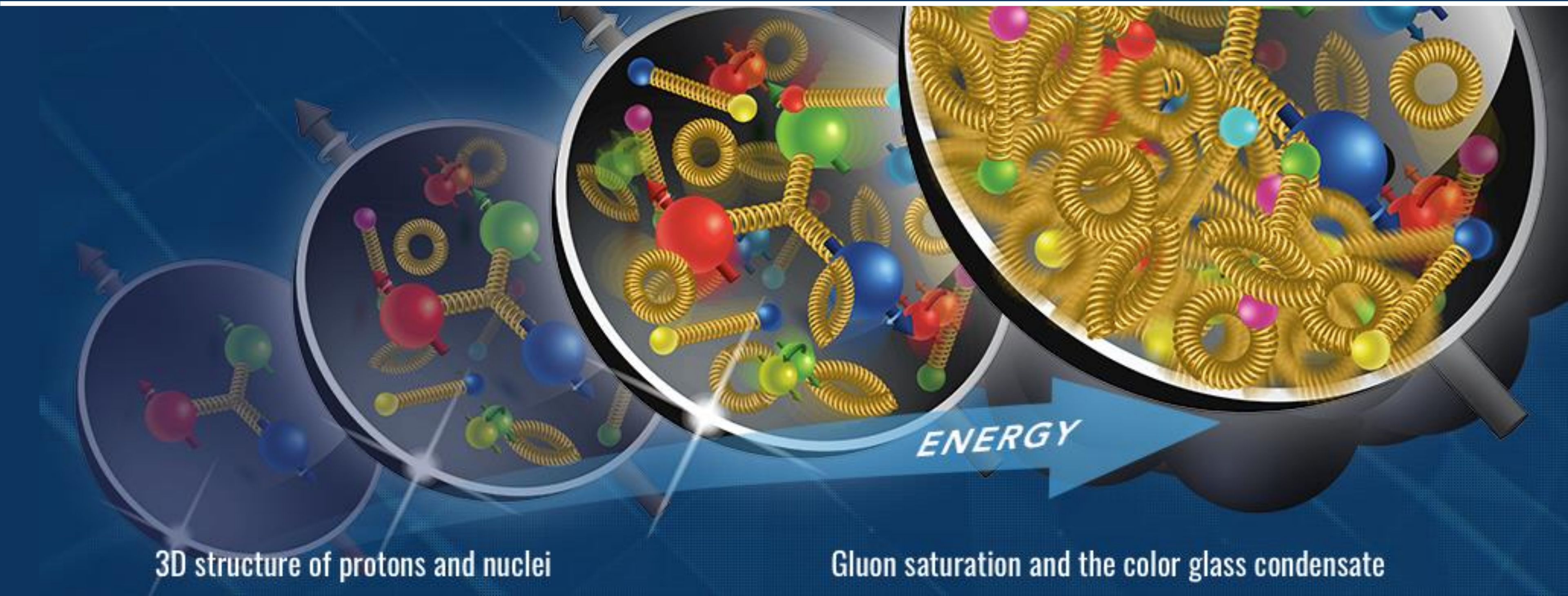
Jin Huang (BNL)

Many thanks to the inputs from
Elke **Aschenauer**, Kai **Chen**, Abhay **Deshpande**, Alexander **Kiselev**,
Tonko **Ljubicic**, David **Morrison**, Christopher **Pinkenburg**, Martin **Purschke**

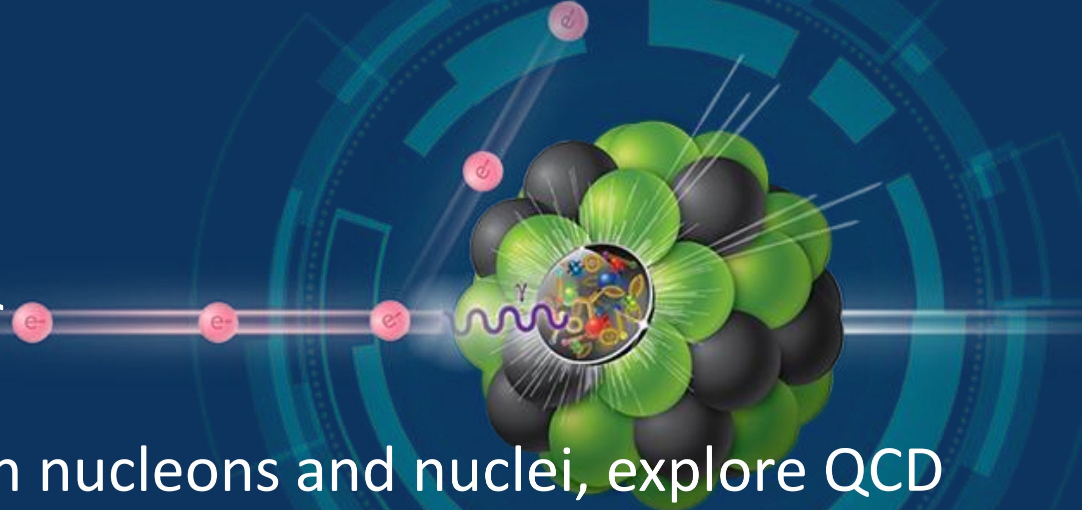
- EIC: Electron ion collider

- Precisely image gluons in nucleons and nuclei, explore QCD frontier of ultra-dense gluon fields, reveal origins of nucleon spin



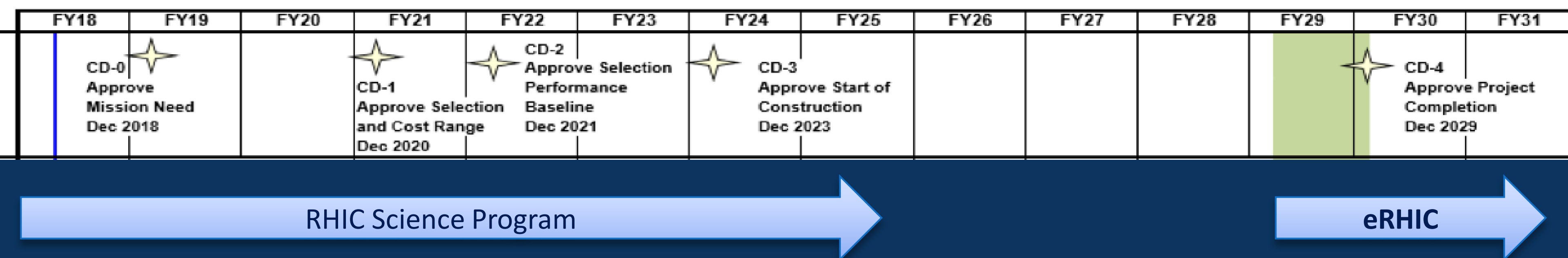


- EIC: Electron ion collider



- Precisely image gluons in nucleons and nuclei, explore QCD frontier of ultra-dense gluon fields, reveal origins of nucleon spin

BNL’s concept for timeline to EIC

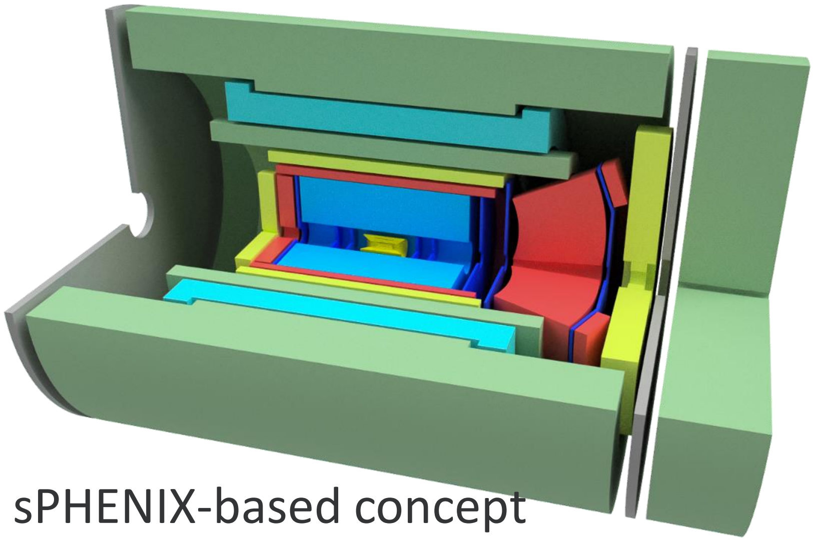


EIC: unique collider → unique real-time challenges

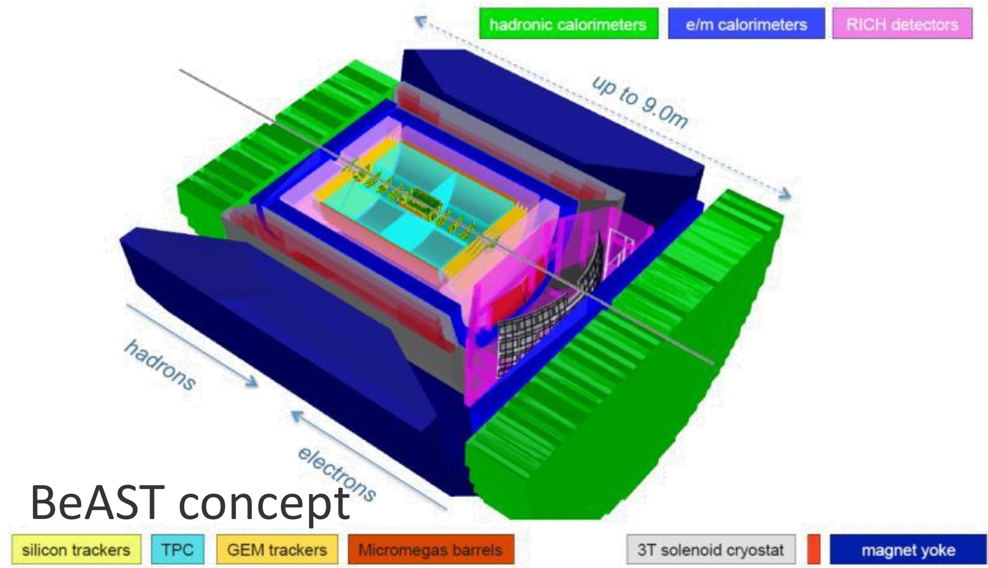
	EIC	RHIC	LHC → HL-LHC
Collision species	$\vec{e}+\vec{p}, \vec{e}+A$	$\vec{p}+\vec{p}/A, A+A$	$p+p/A, A+A$
Top x-N C.M. energy	140 GeV	510 GeV	13 TeV
Bunch spacing	2-10 ns	100 ns	25 ns
Peak x-N luminosity	10^{34} cm^{-2} s^{-1}	10^{32} cm^{-2} s^{-1}	10^{34} → 10^{35} cm^{-2} s^{-1}
x-N cross section	50 µb	40 mb	80 mb
Top collision rate	500 kHz	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	60M N_{ch}/s	30G+ N_{ch}/s

- EIC has lower collision rate and event size is small → signal data rate is low
- But events are precious and have diverse topology
- EIC luminosity is high, so background and systematic control is key

Detector concepts

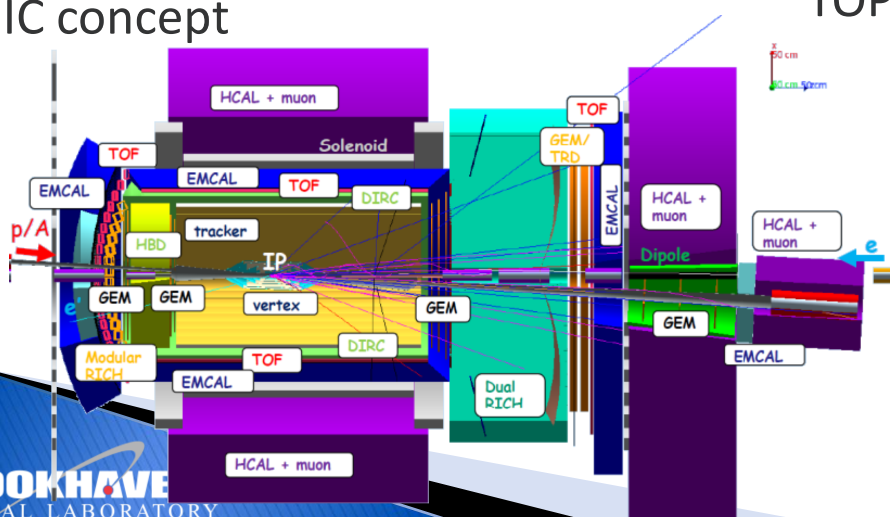


sPHENIX-based concept

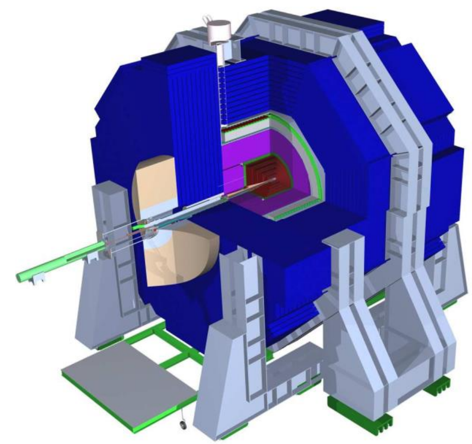


BeAST concept

JLEIC concept



TOPSiDE concept



References reports :

- ePHENIX LOI: arXiv:1402.1209
- eRHIC design report, preCDR: arXiv:1409.1633
- MEIC (JLEIC) design summary: arXiv:1504.07961
- On-going development and updates

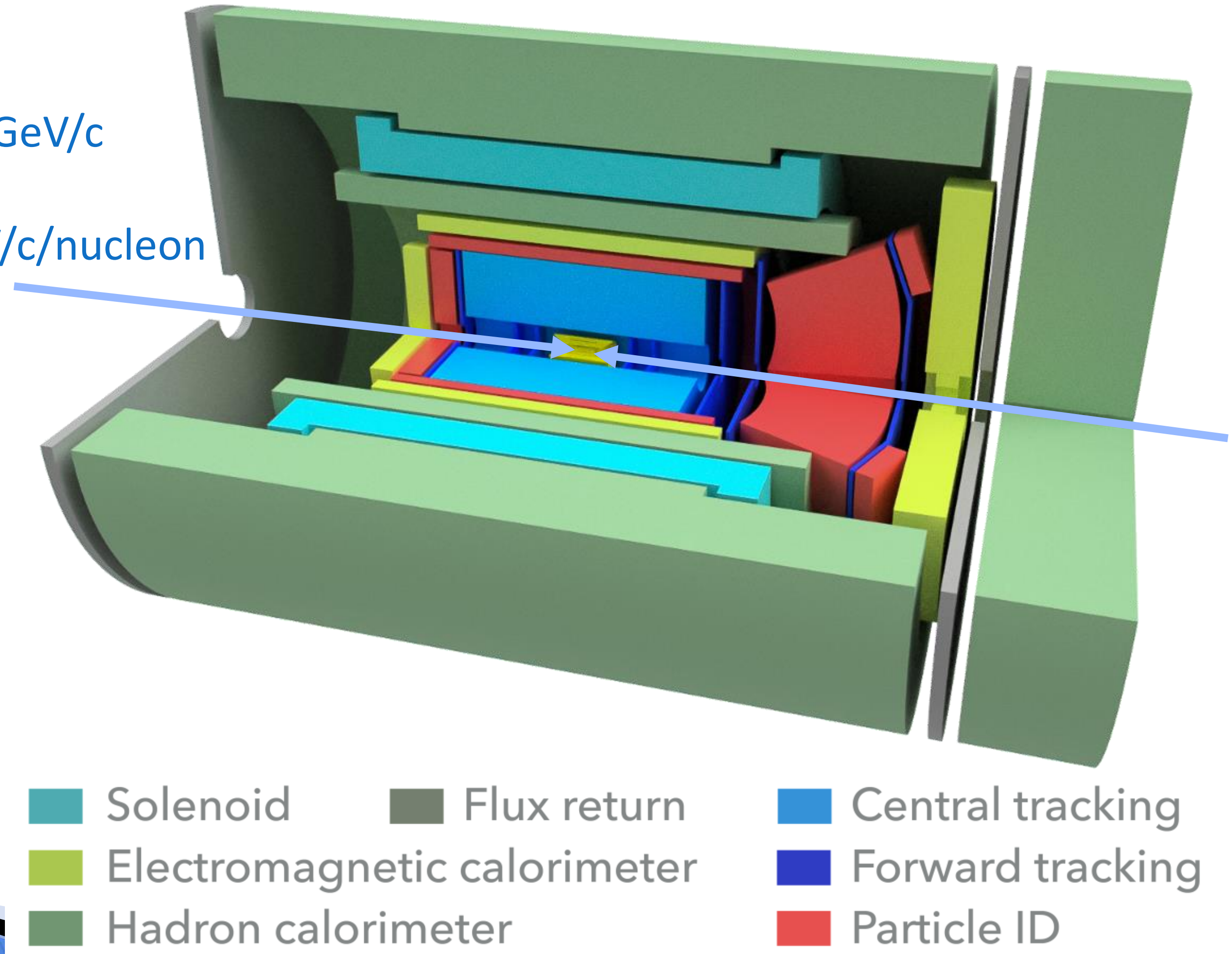
Strategy of the estimation

- Among these four concepts, three are similar: BeAST/sPHENIX/JLEIC
  - Next two sections: data rate and beam gas background estimation based on an assumptions sPHENIX digitizer data format
- TOPsiDE has much larger channel count in calorimeter (pixelated digital calorimeter) and short integration time in trackers (10-ps LGAD tracker)
  - R&D are still early (e.g. 10-ps LGAD noise rate), Readout cost is quite undefined too IMHO
  - Comments on its data rate and strategy at the end of rate estimation

Collision-related data rate estimation

Example: sPHENIX-based EIC detector

- Proton 275 GeV/c Polarized
- Ion 100 GeV/c/nucleon

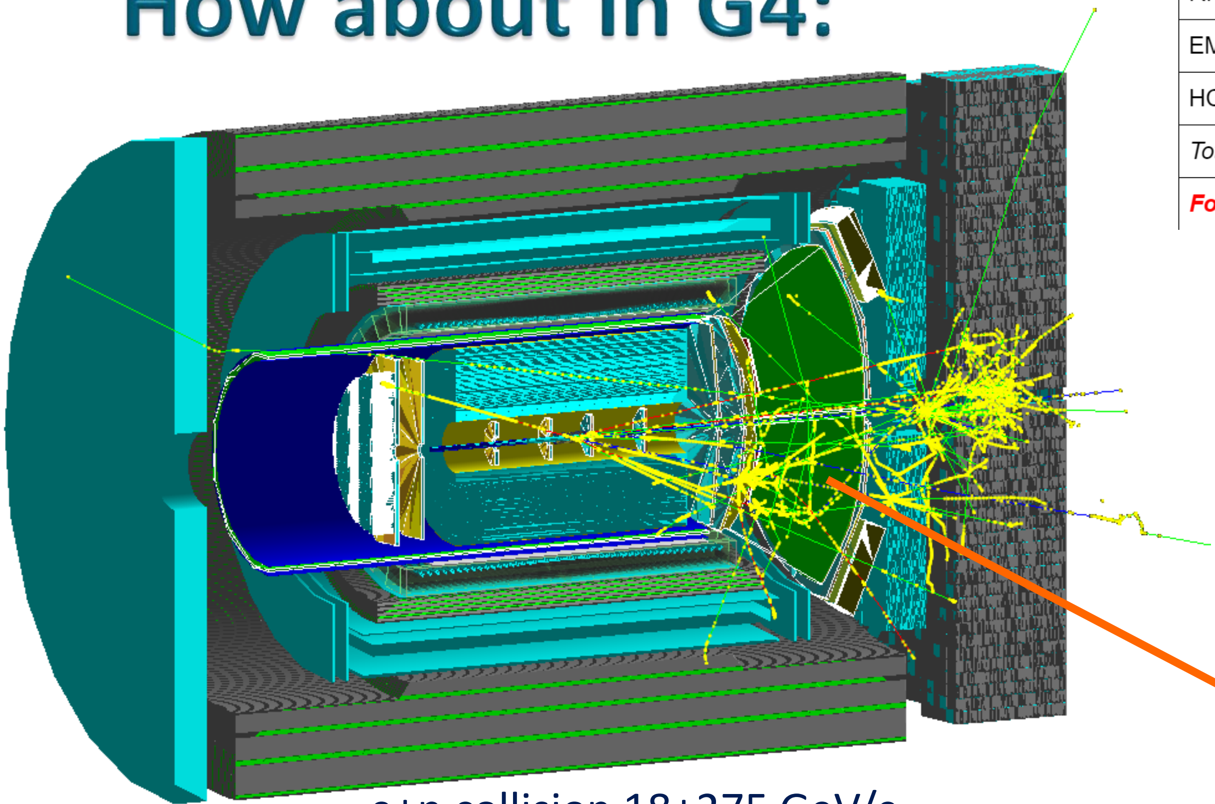


e^- 10 GeV/c Polarized

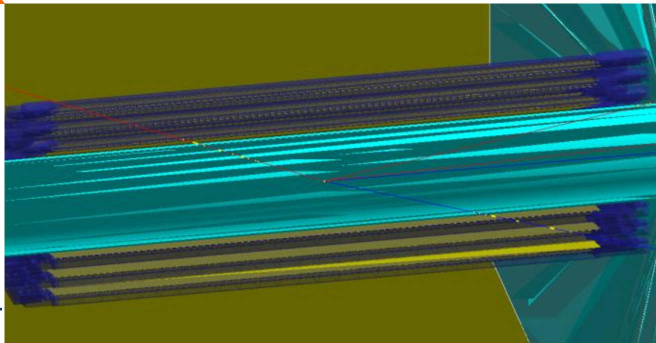
Tonko's estimation:

Signal rate = 16*8 Gbps ~ 100 Gbps @ 10^{33} cm^{-2} s^{-1}, 200kHz collision

How about in G4:



e+p collision 18+275 GeV/c
DIS @ Q^2 ~ 100 $(GeV/c)^2$



Tonko's estimation (2015)

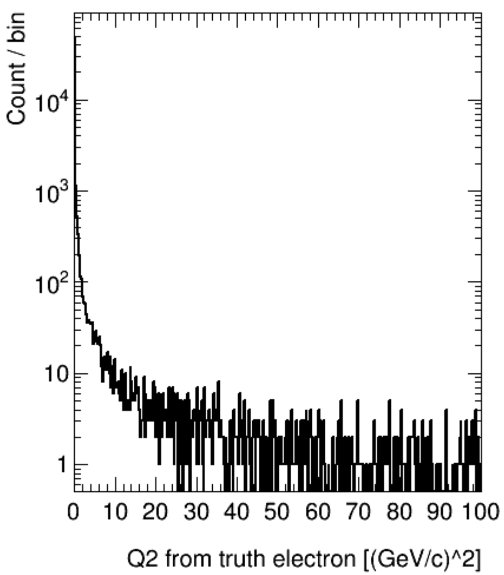
The eRHIC Detector ("BeAST") Readout Scheme

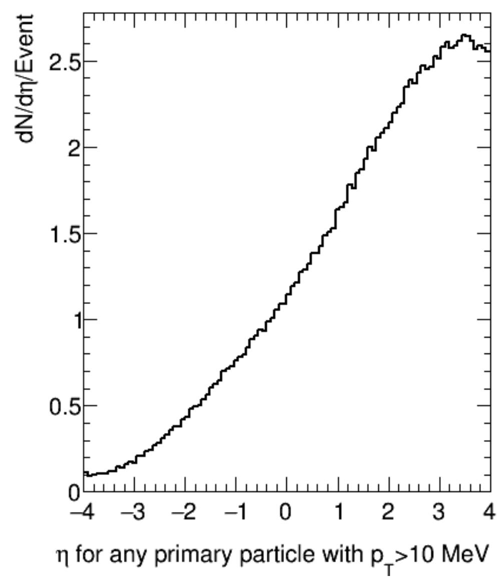
Detector	**Bytes per track**
TPC	100 x (80+4+4) ~ 9000
Silicon	7 x (4+4+4) ~ 90
RICH	20 x (4+4+4) ~ 250
EMCal	1 x (4+4+4) ~ 20
HCal	1 x (4+4+4) ~ 20
Total per track	*9.4 kB*
For 1.7M tracks/s	***(1.7M x 9.4 kB =) 16 GB/s***

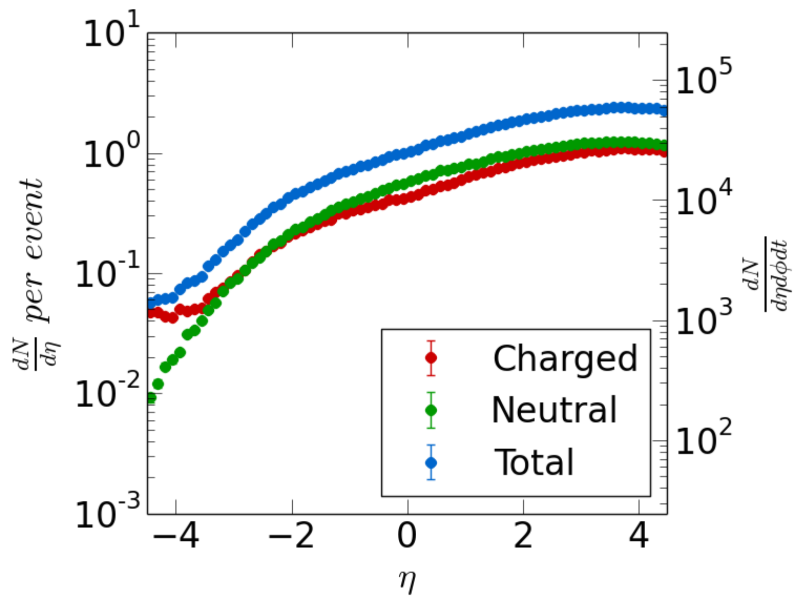
Full detector “Minimal bias” EIC events in sPHENIX framework: quick first look

Multiplicity check for all particles
Minimal bias Pythia6 e+p 20 GeV + 250 GeV
53 µb cross section

BNL EIC taskforce studies
https://wiki.bnl.gov/eic/index.php/Detector_Design_Requirements







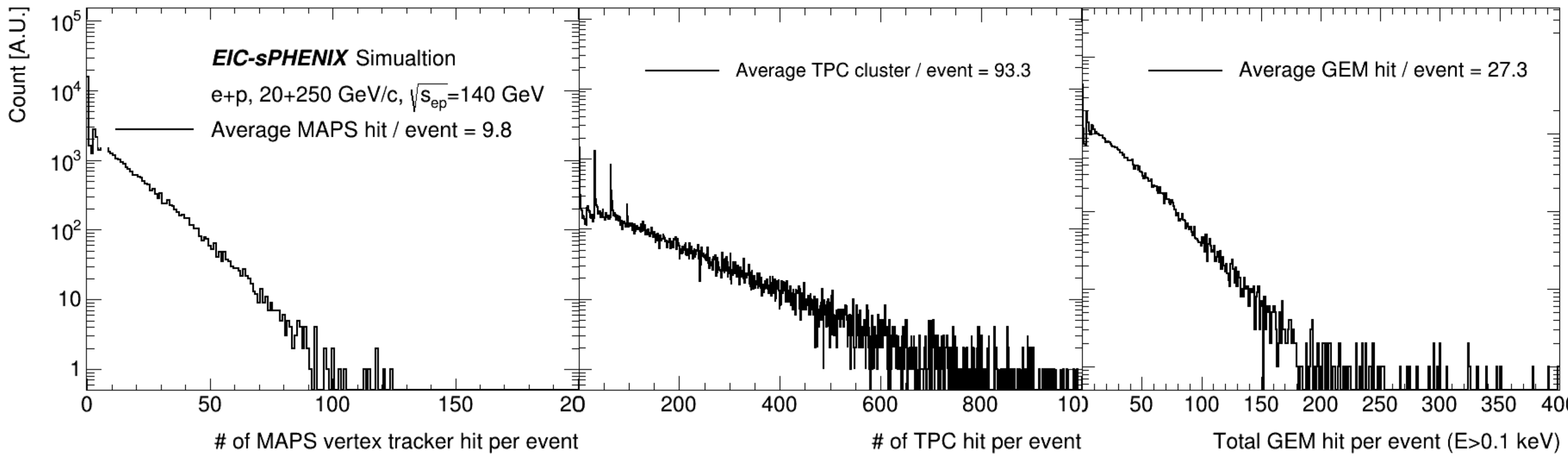
Based on BNL EIC task-force eRHIC-pythia6 55ub sample

```
pythia.ep.20x250.1Mevents.RadCor=0.root
CKIN(3)       changed from        0.00000 to        0.00000
CKIN(4)       changed from       -1.00000 to       -1.00000
```

GEANT4-based detector simulation for DAQ simulation: tracker

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

Extract mean value/collision that produces average signal data rate and tails that produce the buffer depth and latency requirements



Raw data: 16 bit / MAPS hit

3x10 signal hit / collision → 0.2 Gbps @10^{34} $cm^{-2}s^{-1}$

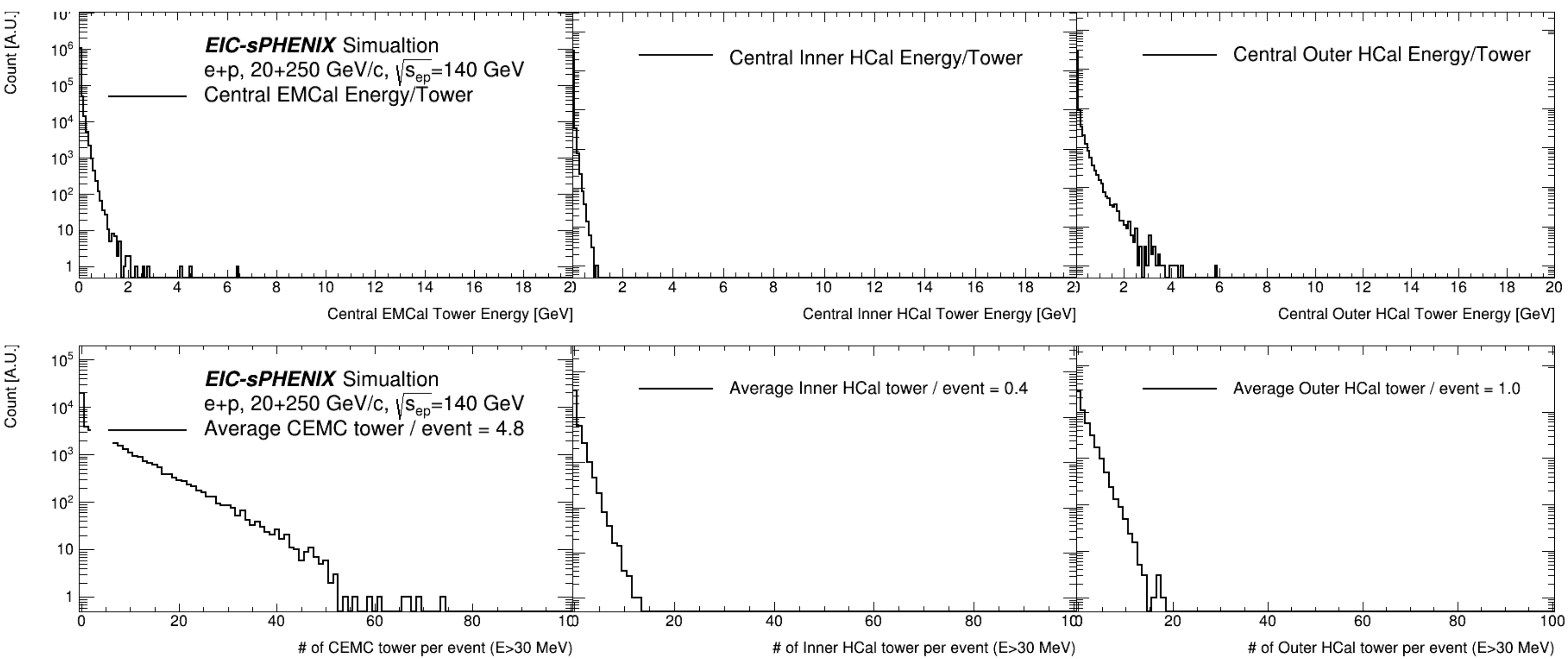
- MAPS is vulnerable to beam background see later slides
- ALPIDE MAPS noise are low, expect 10^{-6} /pixel/strobe, 200M pixel, 3us strobe → ~1Gbps

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

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

GEANT4-based detector simulation for DAQ simulation: central calorimeters

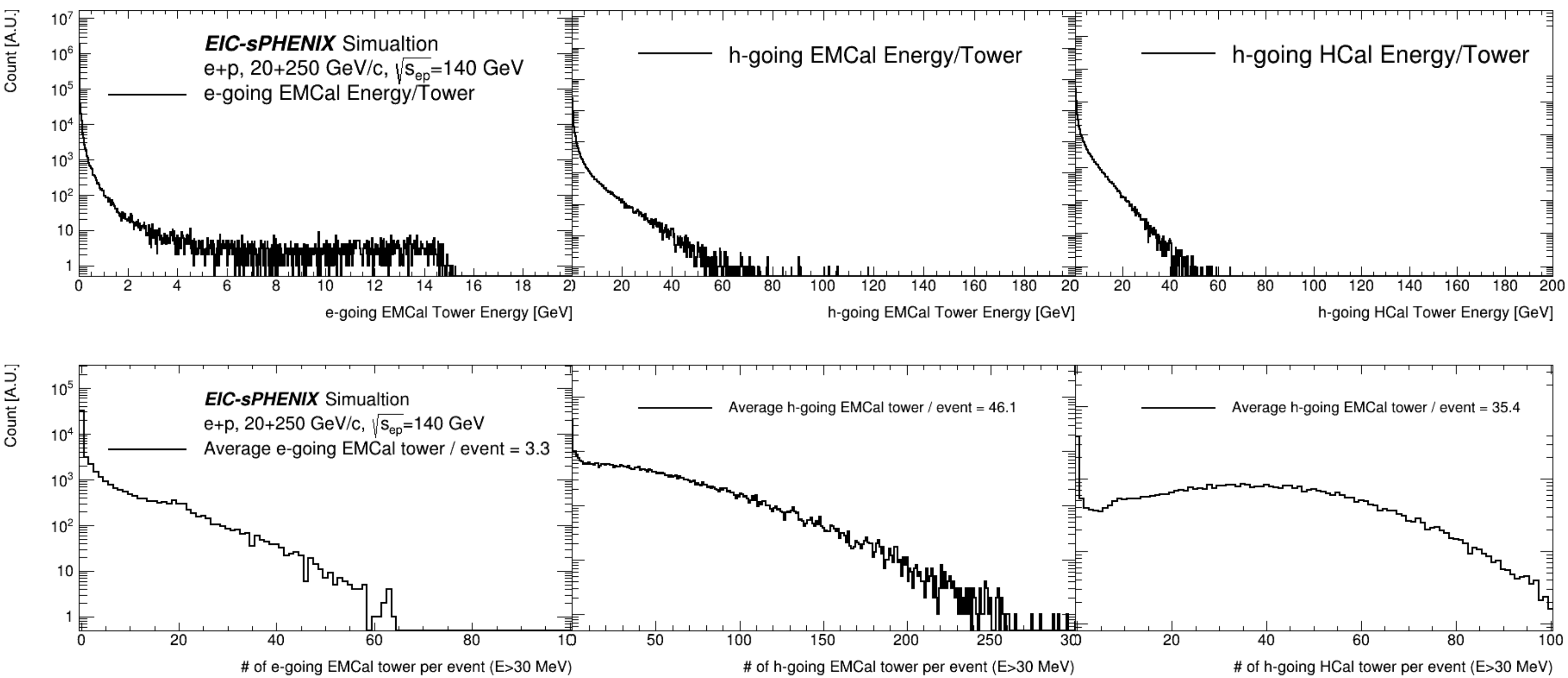
Raw data: 31x 14 bit / active tower +padding + headers ~ 512 bits / active tower



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

GEANT4-based detector simulation for DAQ simulation: forward calorimeters

Raw data: 31x 14 bit / active tower +padding + headers ~ 512 bits / active tower

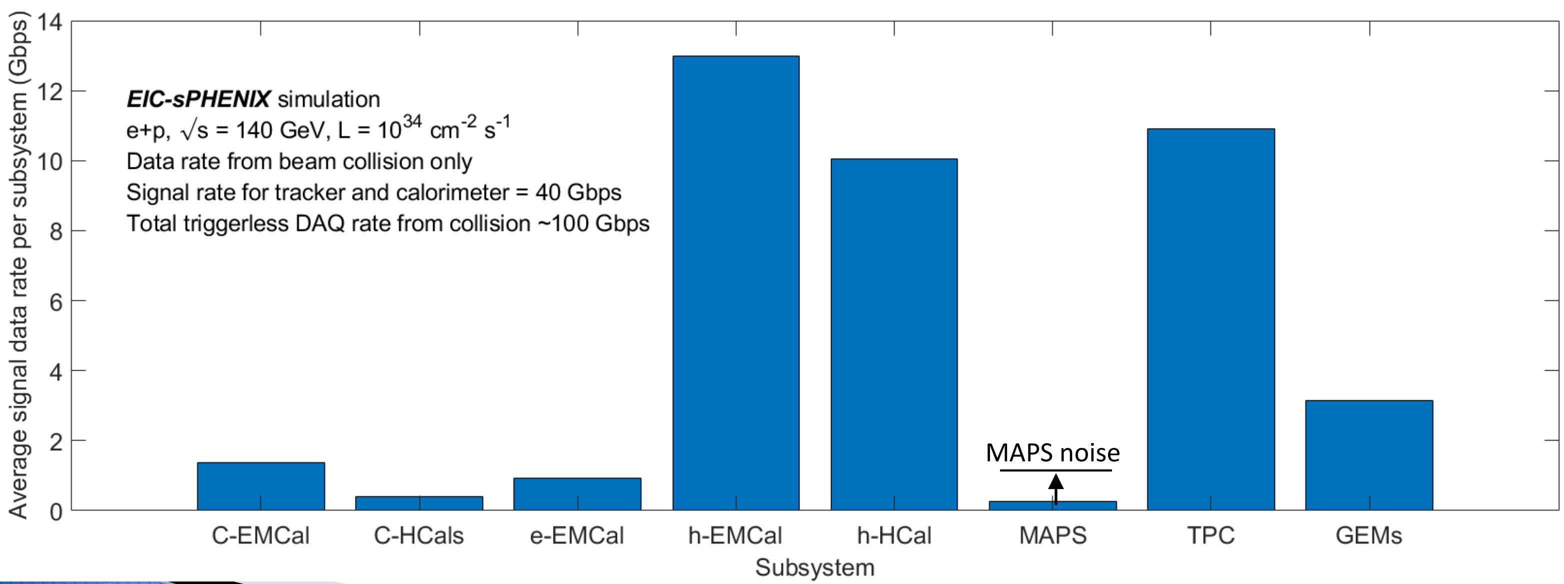


sPH-cQCD-2018-001, https://indico.bnl.gov/event/5283/

EIC preliminary data rate summary

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

- Tracker + calorimeter ~ 40 Gbps
- + PID detector + 2x for noise ~ 100 Gbps
- Signal-collision data rate of 100 Gbps seems quite manageable,
  - < sPHENIX TPC peak disk rate of 200 Gbps (See Martin's talk)
- Machine background and noise would be critical in finalizing the total data rate
  - From on-going EIC/sPHENIX R&D prototyping will show noise level from state-of-art MAPS and SAMPA ASICs, e.g. ALPIDE MAPS noise rate ~ 1 Gbps
  - Enough FPGA/CPU resource with prevision for noise filtering in EIC online system



Beam gas estimation for eRHIC detectors

- Assuming flat 10e-9 mbar vac in experimental region

Beam-gas interactions

- p + p (beam gas) cross section ~ 40 mb @ 250 GeV
- **Beam gas interaction rate** = 2.65e10(H_2/cm^2/10m) * 2(proton/H_2) * 40e-27(40mb→cm^2) * 1(A) / 1.6e-19(C/proton)
  = **13kHz / 10m beam line < 10% EIC collision rate**
- The following estimation assumes
  - HERA inspired flat 10e-9 mbar vac in experimental region of |z|<450 cm
  - 2M M.B. Pythia-8 beam gas events simulated in Geant4 full detector

Vacuum pressure	10^{-9} mbar
Beampipe temperature	Room temperature
Average atomic weight of gas	Hydrogen (H^2)
Molecular density (for 10 m pipe)	2.65×10^{10} molecules/cm^2
Luminosity (Ring-Ring)	10.05×10^{33} cm^{-2}s^{-1}
Bunch intensity (R-R) (e/p)	15.1 / 6.0×10^{10}
Beam Current (R-R) (e/p)	2.5 / 1 A
Bunch spacing (Ring-Ring)	8.7 ns → 1320 bunches
ElectronxProton beam energy	10 GeV x 275 GeV

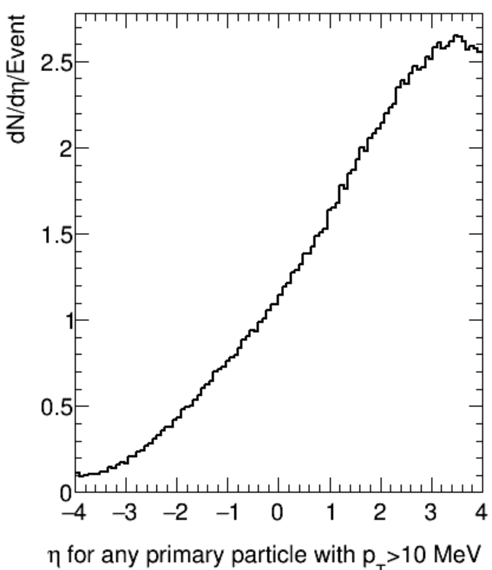
Courtesy: E.C. Aschenauer
eRHIC pre-CDR review

BROOKHAVEN NATIONAL LABORATORY

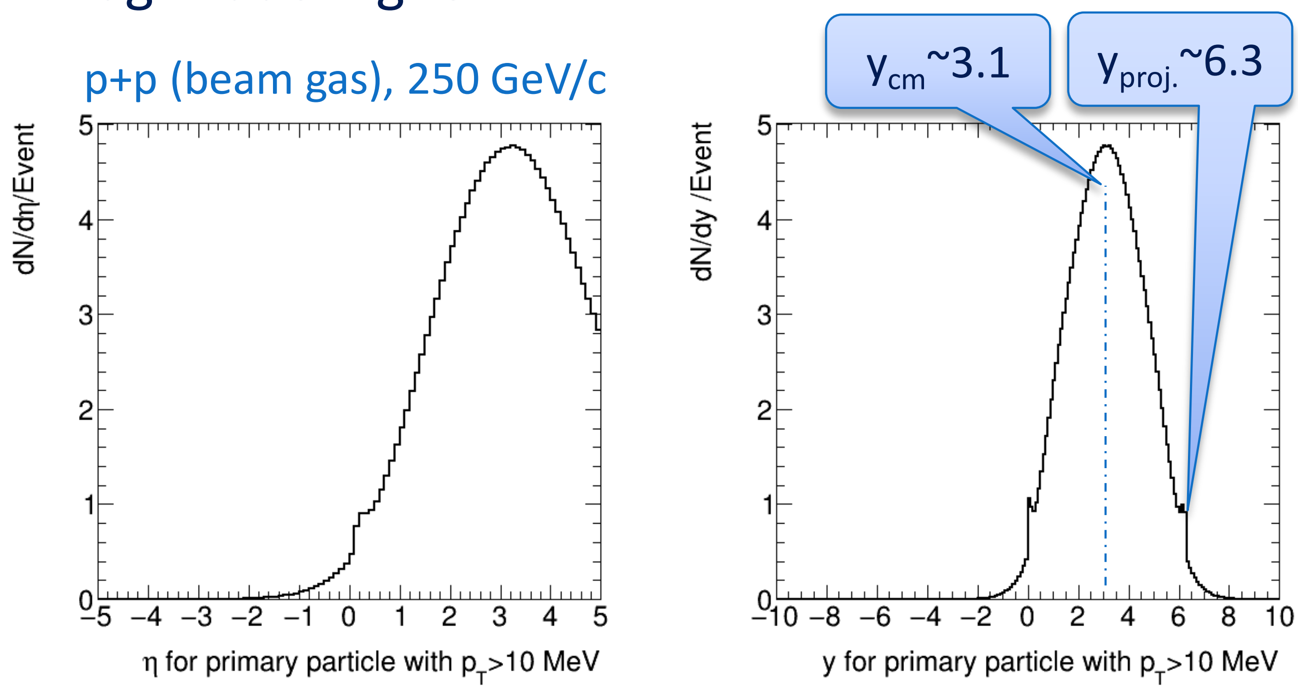
Beam gas multiplicity

- 250 GeV/c proton beam on H_2 gas target
- C.M. rapidity~3.1, sqrt[s] ~ 22 GeV, cross section~40 mb
- Lab per-pseudorapidity multiplicity is higher than e+p, but ***not*** orders of magnitude higher

e+p, 20 + 250 GeV/c

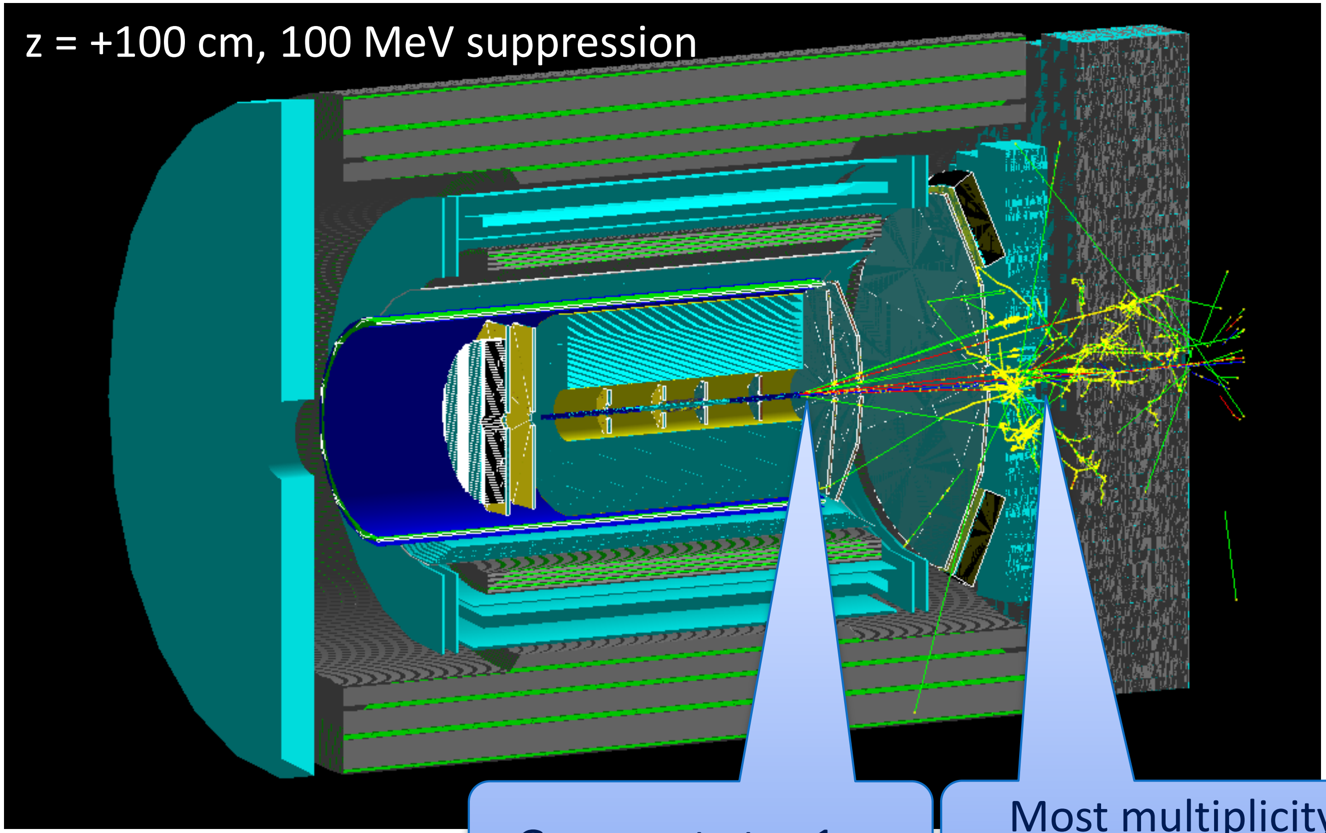


p+p (beam gas), 250 GeV/c



Beam gas event in a detector

- 250 GeV proton beam on proton beam gas, sqrt[s] ~ 22 GeV
- For this illustration, use pythia-8 very-hard interaction event (q^hat > 5 GeV/c)

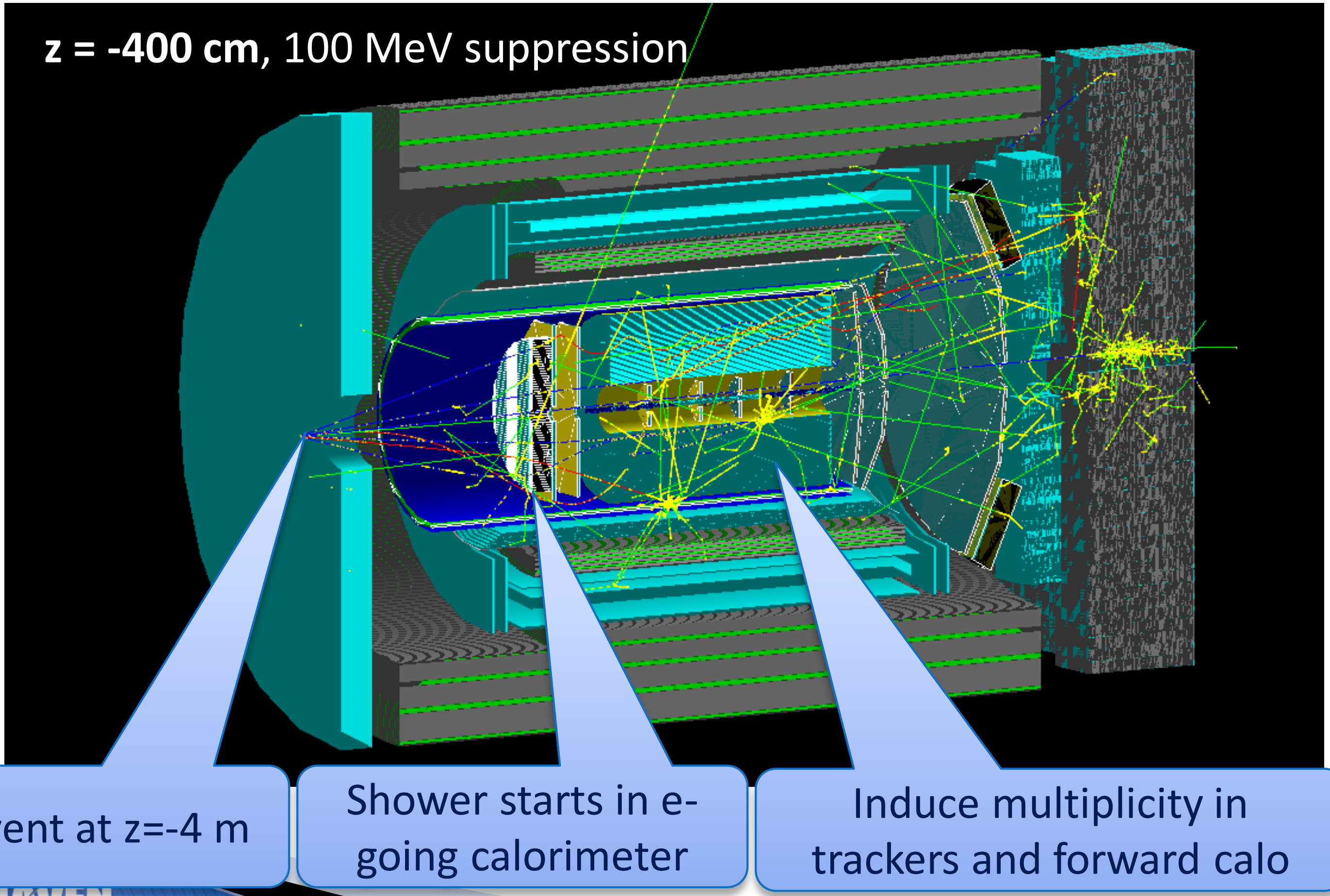


Gas event at z=1 m

Most multiplicity goes to forward calo.

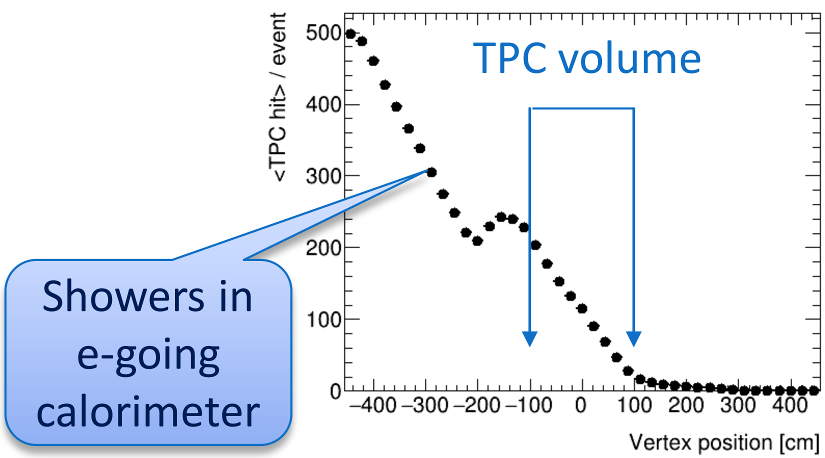
Beam gas event in a detector

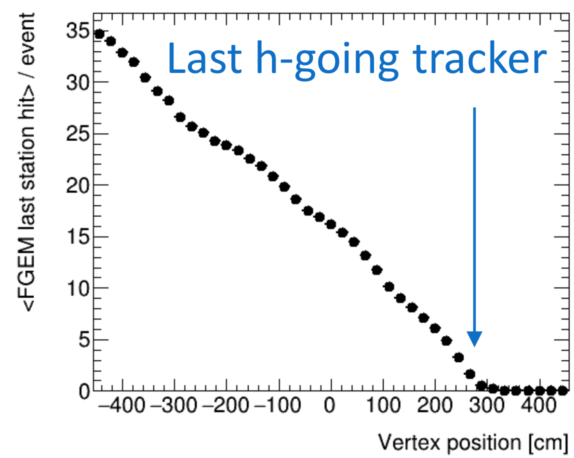
- 250 GeV proton beam on proton beam gas, sqrt[s] ~ 22 GeV
- For this illustration, use pythia-8 very-hard interaction event (q^hat > 5 GeV/c)

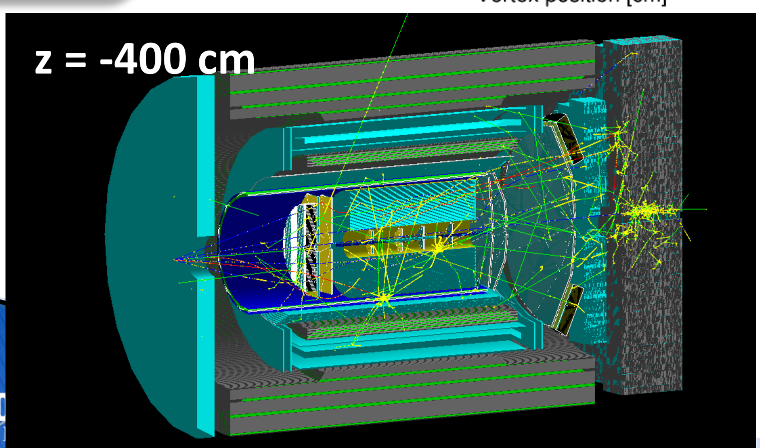


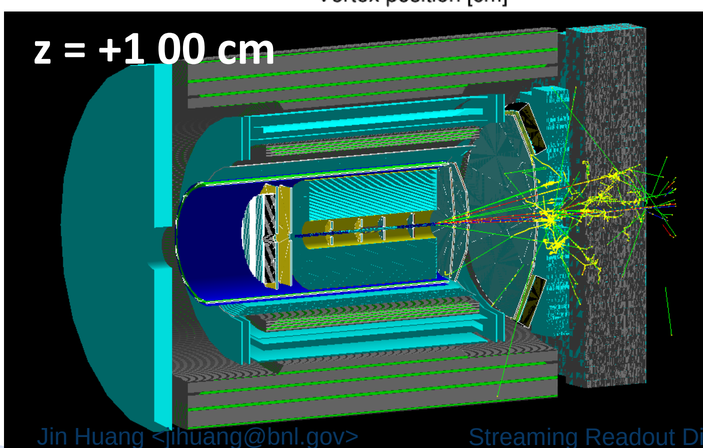
Beam gas vertex sensitivity - tracker

- Average active hit for each beam gas vertex bin
- 250 GeV proton beam on proton beam gas, Pythia-8 M.B.



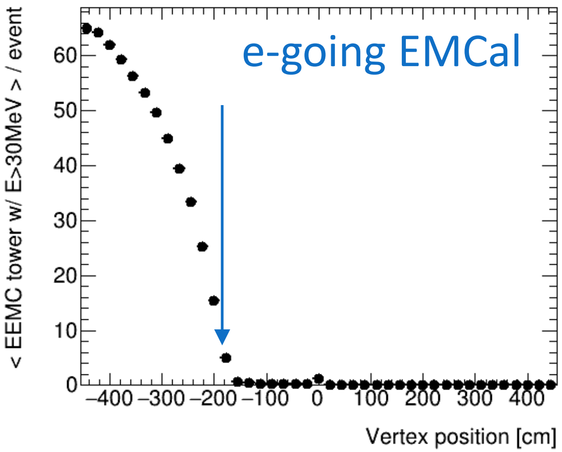


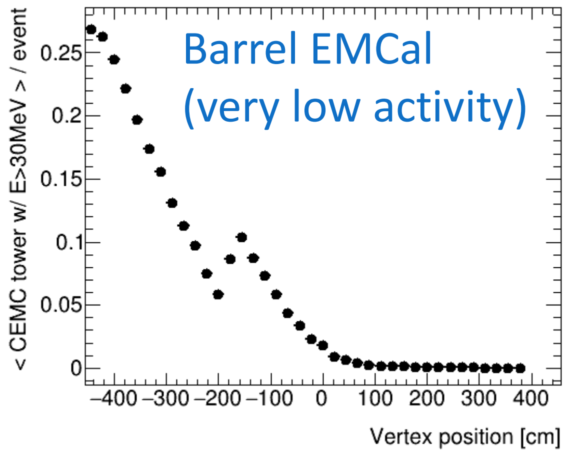


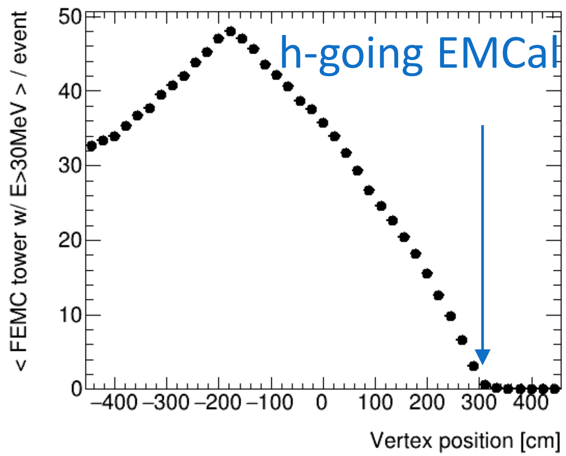


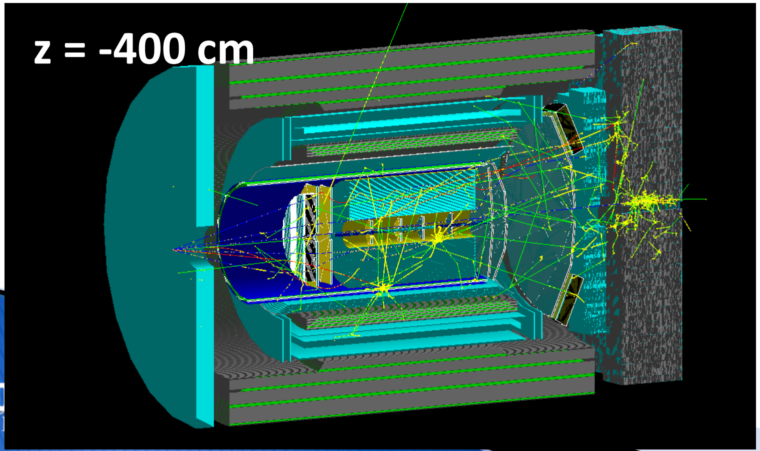
Beam gas vertex sensitivity – calo.

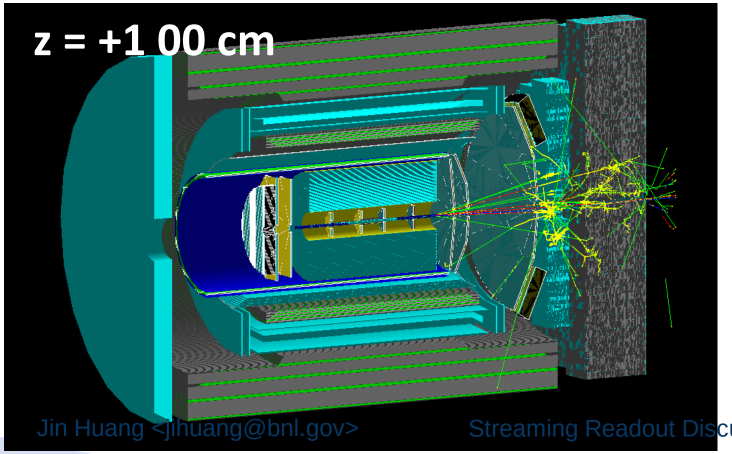
- Average active hit for each beam gas vertex bin
- 250 GeV proton beam on proton beam gas, Pythia-8 M.B.





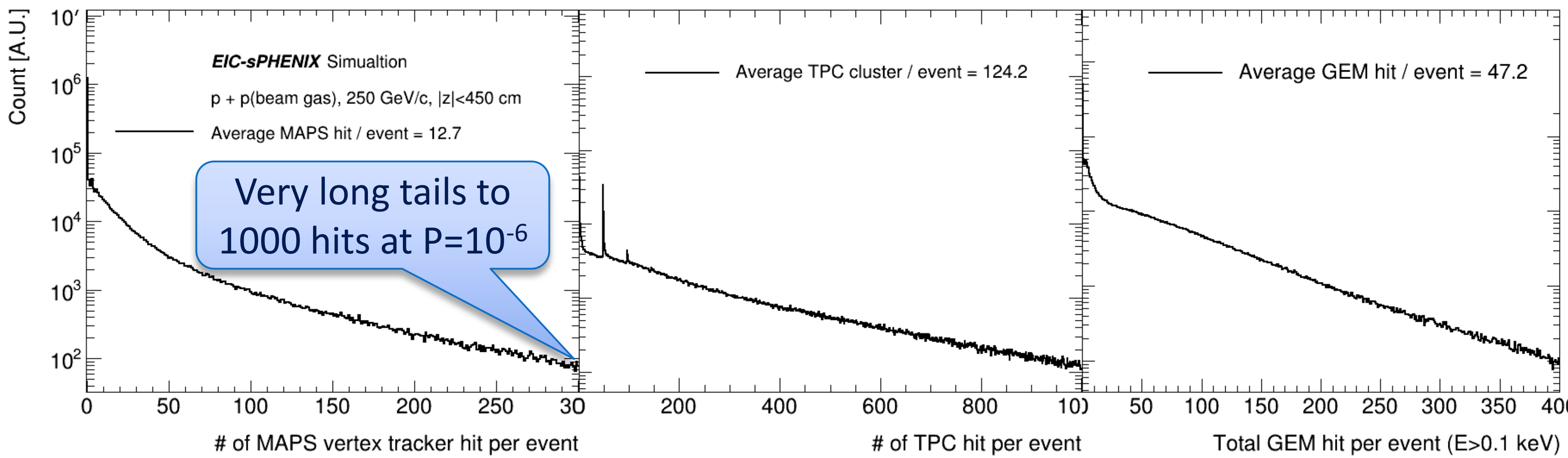






GEANT4-based detector simulation: beam gas event on tracker

Extract mean value/collision (signal data rate) and tails (relates to buffer depth requirement)



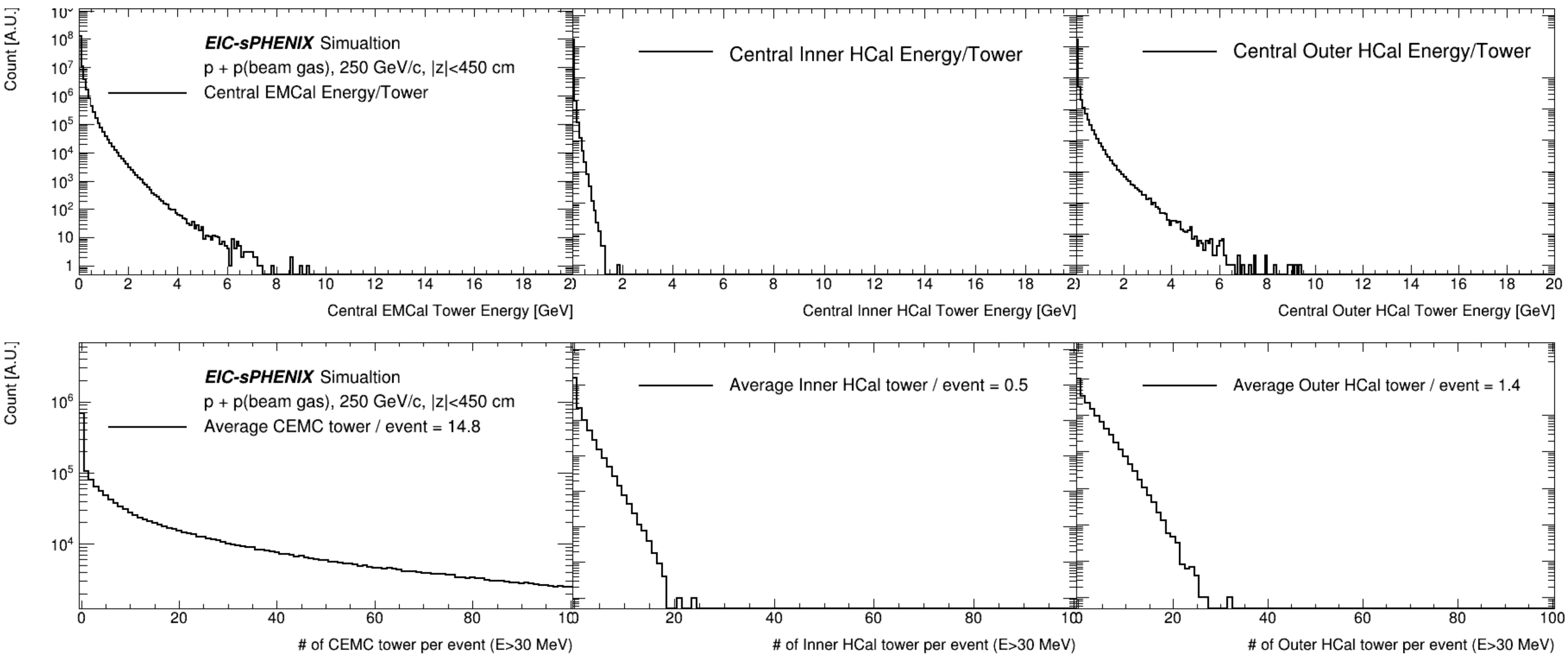
Raw data:
3 pixel x 16 bit / MAPS hit

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

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

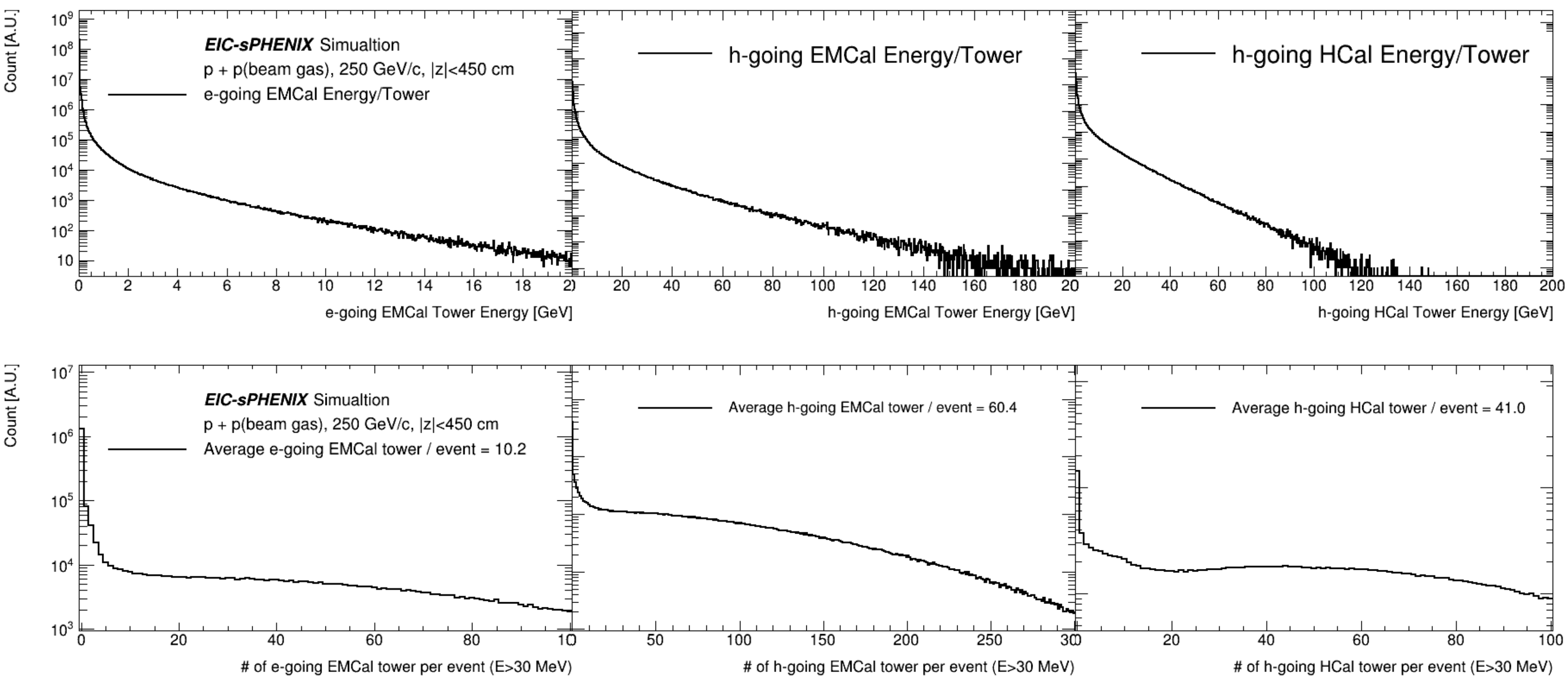
GEANT4-based detector simulation: beam gas event on central calorimeters

Raw data: 31x 14 bit / active tower +padding + headers ~ 512 bits / active tower



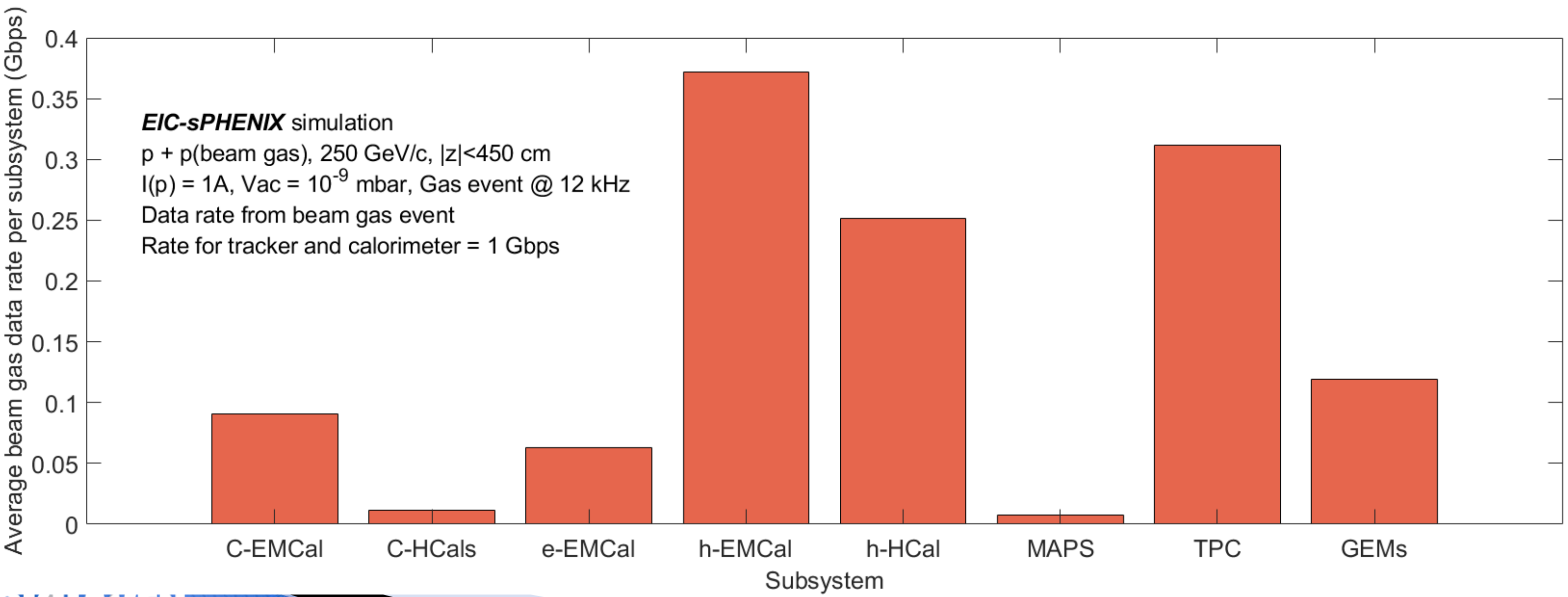
GEANT4-based detector simulation: beam gas event on forward calorimeters

Raw data: 31x 14 bit / active tower +padding + headers ~ 512 bits / active tower



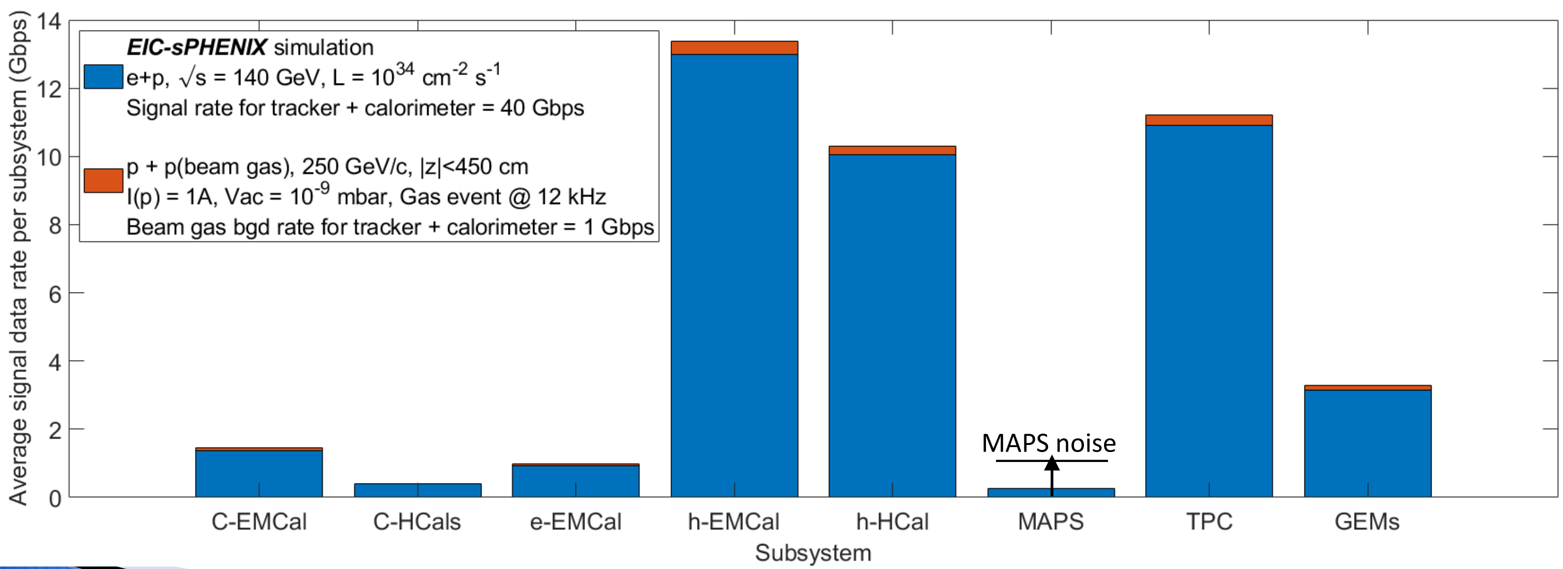
Rate summary for beam gas

- Very similar rate distribution among subsystems when compared with EIC collisions
- With an assumed vacuum profile (10^{-9} mbar flat within experiment region):
  - Overall ~ 1 Gbps @ 12kHz beam gas at 10^{-9} mbar in |z|<450 cm (detector region)
  - << EIC collision signal data rate
- Further investigation needed:
  - In the experimental region : Dynamic vac profile
  - Beyond experiment region: beam gas profile, possible passive shielding and active veto



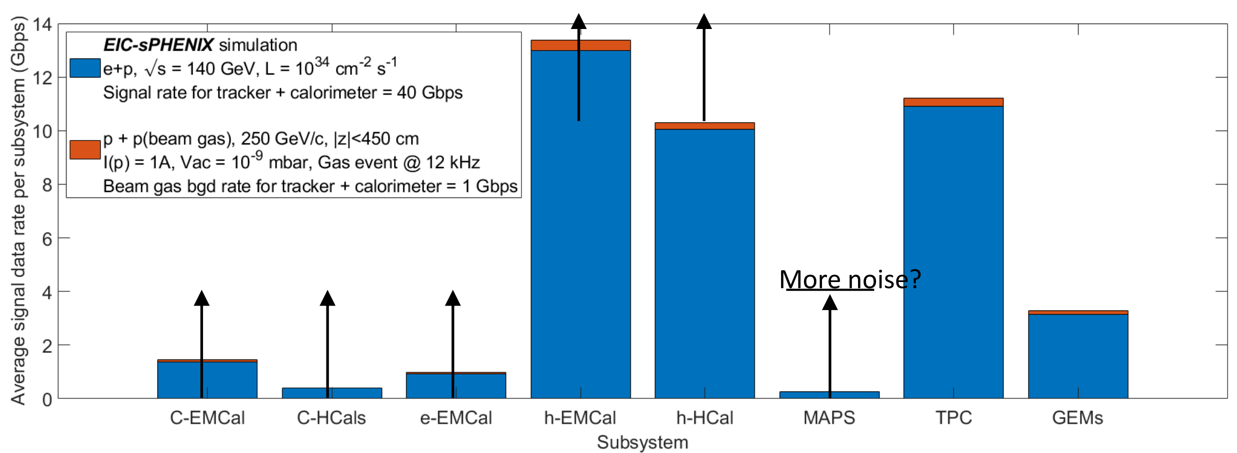
Rate summary
Sum collision + beam gas

- Total ~ 100 Gbps @ 10^{32} cm^{-2} s^{-1}
  - < sPHENIX peak disk rate
  - Beam gas data rate << collision data rate
- Further to be evaluated with more concrete detector and accelerator development:
  Beam gas profile, synchrotron radiation, detector noise



Highly segmented detectors/ TOPsiDE

- Highly spatial and/or time segmented detectors would induce higher event size
- TOPSIDE would use pixelated calorimeter and produce x O(100-1000) calorimeter data rate depending on segmentation and digitizer assumptions
- TOPSIDE propose to use LGAD tracker
  - Low Gain Amplifying Detectors (LGAD) + MAPS to enhance charge collection and timing
  - The signal data rate would not change in the leading order (e.g. 3-hit x 16bit/cluster)
  - Depending on LGAD R&D, the noise rate could be higher (i.e. higher noise/pixel, shorter integration)
- TOPSIDE may start with full streaming at 10^{33} cm^{-2} s^{-1}.
- However, assuming same data log rate (say 100Gbps), at full EIC lumi it may require global triggering to record a subset of collisions and/or real-time feature building (e.g. cluster fitting/tracking, See talk JD/TU/SY).

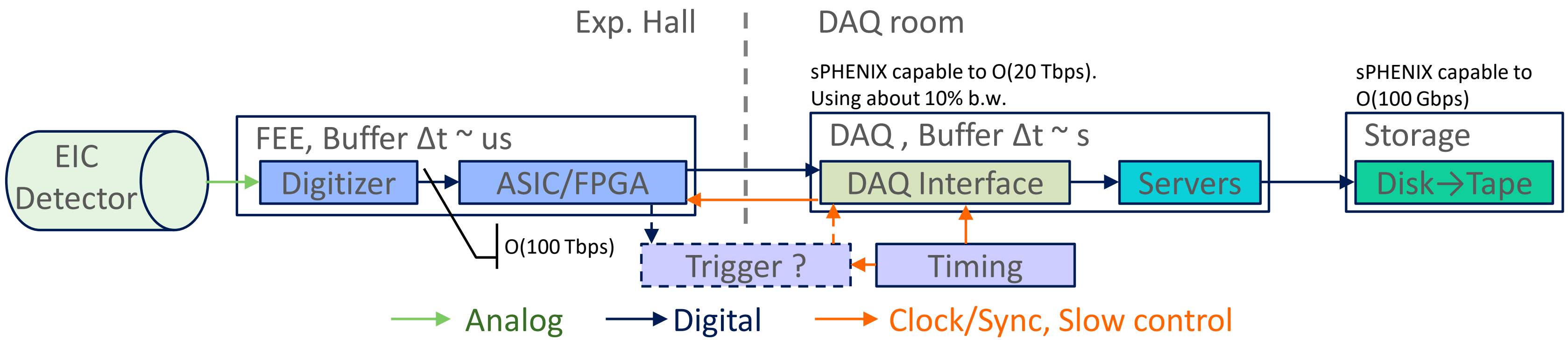


BROOKH
NATIONAL LAB

n

Discussion on real-time system for EIC

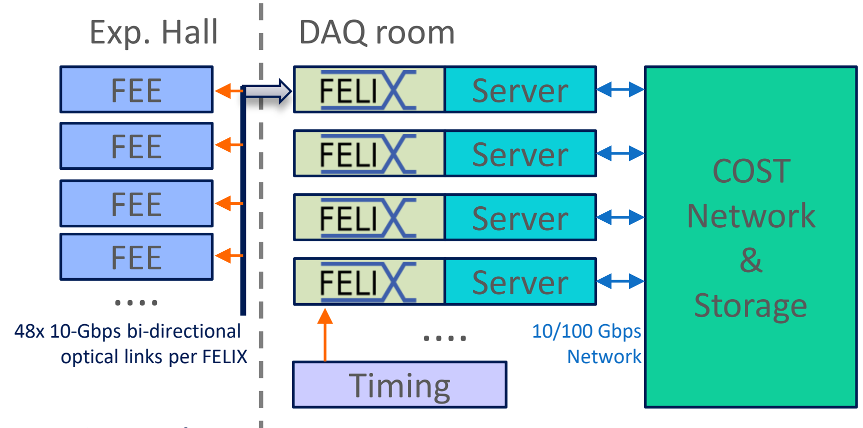
Strategy for an EIC real-time system



- In a digital pipelined real-time system, all channels are digitized at all times. Data reduction in real-time to fit within bandwidth + buffer constraints
  - A commonly used strategy in data reduction is global triggering to selectively record a small fraction of collisions.
  - However, global triggering is not required if [system throughput > rate from all collisions]
  - Data reduction beyond global triggering : e.g. zero-suppression (in ASIC/FPGA), feature building (e.g. clustering), online analysis (e.g. online tracking)
- For the signal data rate from EIC (100 Gbps), we can aim for filtering-out and streaming all collision in raw data without a hardware-based global triggering
  - One may also consider a trigger-streaming hybrid system (e.g. STAR eTOF DAQ, sPHENIX hybrid-DAQ in Martin's talk), which can quantify efficiency/bias in streaming data reduction and be resilient at high background rate

One DAQ strategy

sPH-cQCD-2018-001
https://indico.bnl.gov/event/5283/

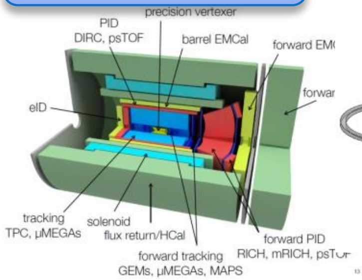


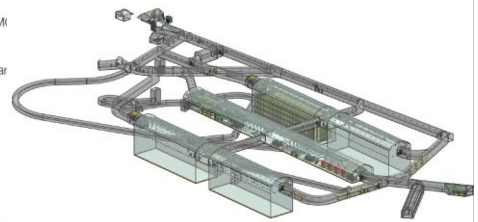
- Full streaming readout front-end (buffer length : μs)
  → DAQ interface to commodity computing via PCIe-based FPGA cards (e.g. FELIX)
  → Disk/tape storage of streaming time-framed zero-suppressed raw data (buffer length : s)
  → Collision event tagging in offline production (latency : days)
- Why time-framed streaming readout?
  - Diversity of EIC event topology. Streaming minimizing systematics by avoiding hardware trigger decision, keeping background and history
  - At 500kHz event rate, multi-μs-integration detectors would require streaming, e.g. TPC, MAPS
- Why FELIX-like DAQ interface?
  - 0.5 Tbps x bi-direction IO, bridging μs-level FEE buffer length with miliseconds + DAQ network time scale
  - Interface with commodity computing via PCIe @ ~100Gbps
  - Distribute experiment timing and synchronization cross large system
- Why keep raw data?
  - At 100 Gbps < sPHENIX rate, it is affordable to disk-write all raw signal data: If you can, always keep raw data
  - Allow time + special run needed for final calibration, followed by prompt reconstruction

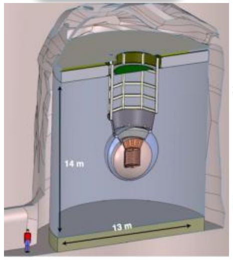
See Martin's talk on sPHENIX adaption and implementation in with sPHENIX/EIC tracker test stands

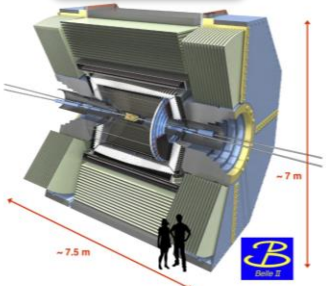
Ongoing R&D, BNL LDRD 19-026: Common development for Advanced DAQ

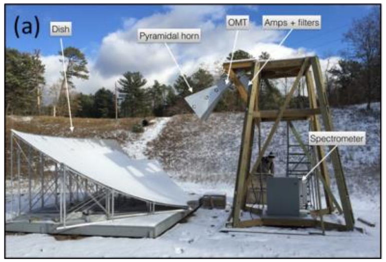
EIC detectors | DUNE far detector | nEXO | BELLE-II | 21-cm digital interferometer

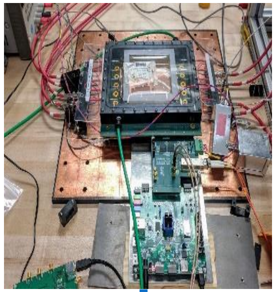


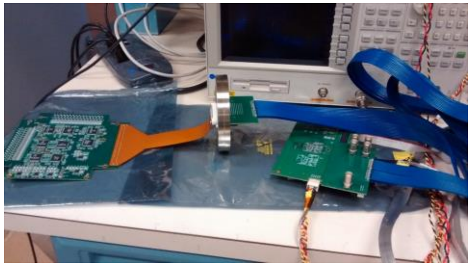


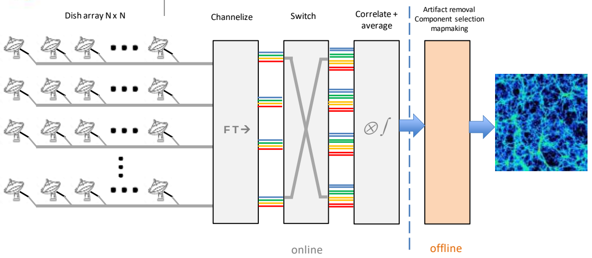












BNL 712 – series PCIe Card

Commodity Computing

Common challenge cross multi-discipline of BNL: Advanced DAQ with high throughput @ 100+ Gbps

Solution: FELIX-based DAQ, the architecture used in the LHC upgrades in the 2020s

Deliverable: test stand & firmware for each case

Co-PIs: Kai Chen (BNL/ATLAS),
Jin Huang (BNL/sPHENIX)

Summary

- EIC is a unique collider combining high luminosity and low collision cross section, diverse event topologies and vital in systematics control
- The total collision raw data rate → 100 Gbps
  - Affordable to write all to disk
  - Can be higher for highly segmented or noisy detectors (intrinsic or accelerator)
- Instead of global triggering that selects subset of event, we can choose to stream-record all collisions signals
  - Requiring all front-end to continuously digitize data or self-triggering e.g. PHENIX FVTX, STAR eTOF, all sPHENIX trackers
  - Reliably synchronize all front-ends and identify faults
  - Recording all collision data (100 Gbps if raw)
  - If needed, filtering out background with low signal loss (10^{-4}?)
  - Requiring reliable data flow → control systematics: Low data loss rate < 10^{-4}(?) and/or loss in a deterministic manor (e.g. via real-time busy signal tree to avoid correlation between event type and loss rate)
- A strategy for EIC DAQ based on sPHENIX DAQ investments: FPGA-PCIe DAQ interface, streaming all signal data to disk/tape, prompt off-line reconstruction

Extra information

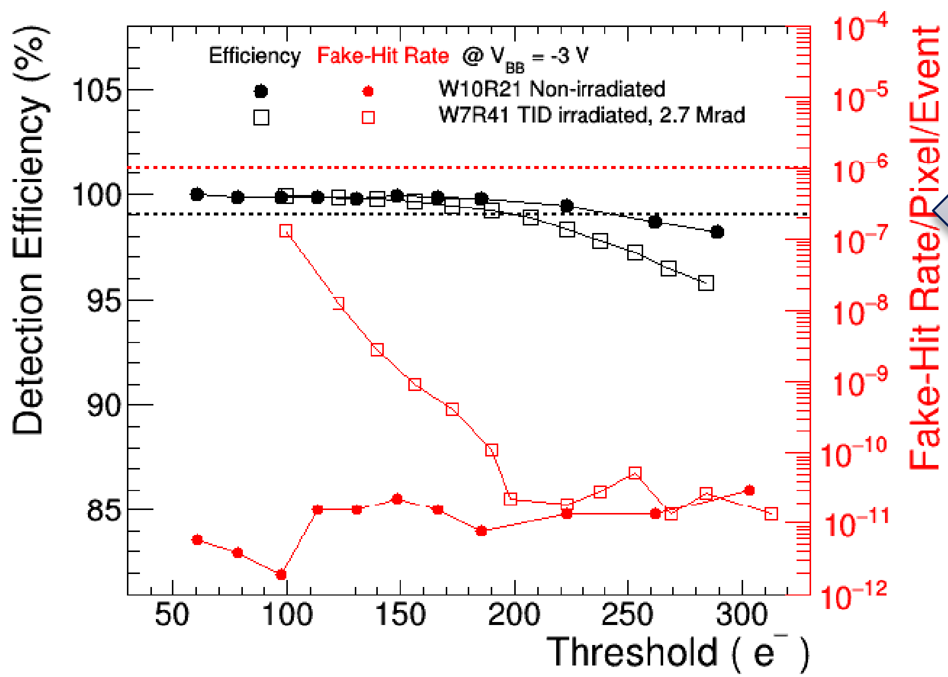
Per-strobe ALPIDE multiplicity

Four factor contributes in a MC simulation:

- Per-collision multiplicity, PDF as in last page
- Number of pile up collision, Poisson distributed
- The triggered collision, |z|<10 cm (trigger mode only)
- Number of noise, Poisson distributed

Comments received:

- Duplicated hits between strobes are not included yet (Thanks to Jo)
- UPC electron background not included (Thanks to Xin)
- Aiming for 10^{-6} noise in final detector (Many)



Bottom line: 10^{-4}

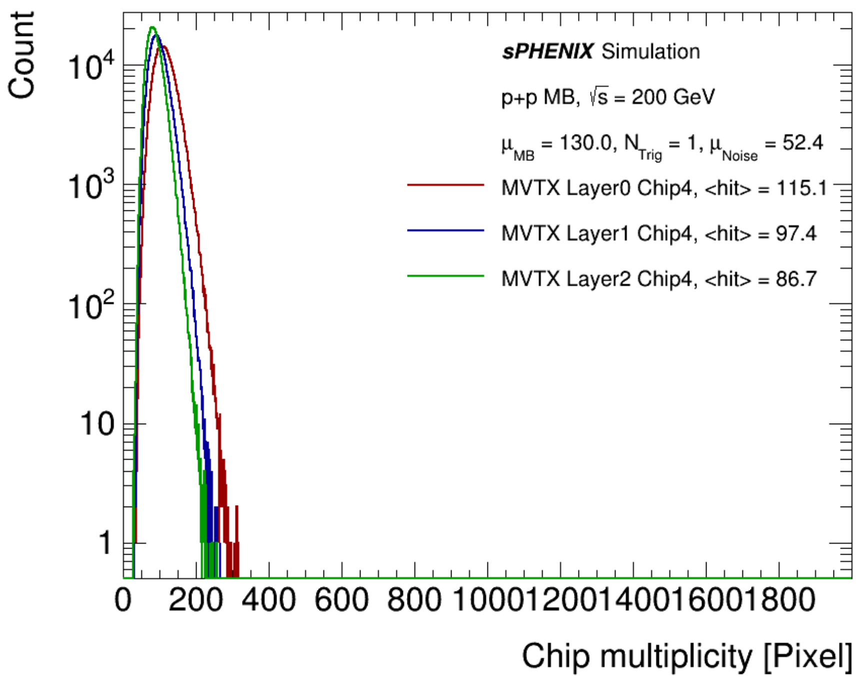
Also studied: 10^{-5}

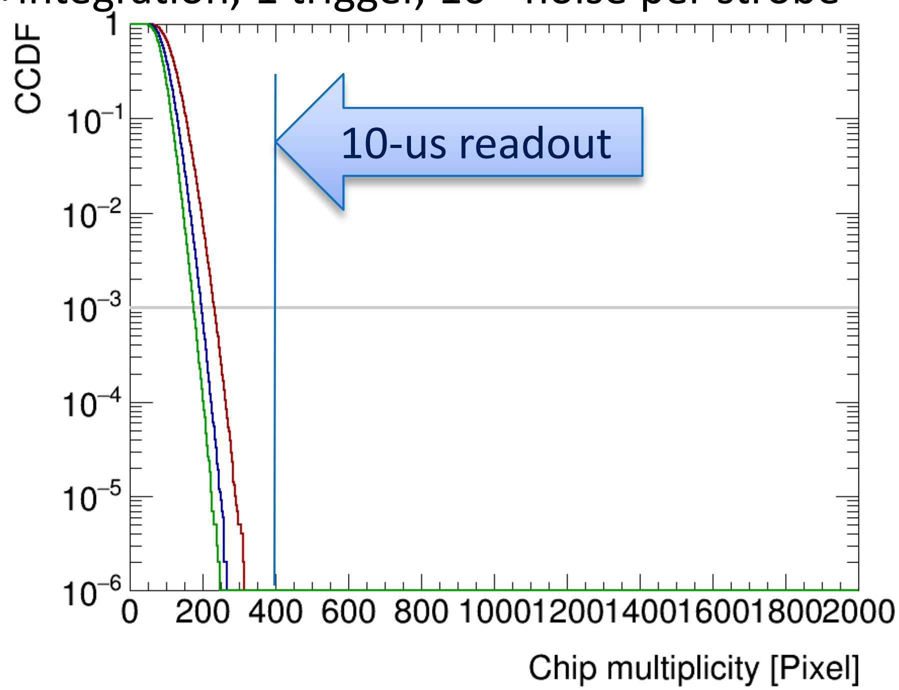
Likely in operation: 10^{-6}

p+p multiplicity, per-strobe, chip-4

- p+p collision related data is completely dominated by pile-ups
- Central limit theorem: High number of pile up → low non-Gauss high tails
- Continuous-mode is quite safe @ 10-us strobe window

13 MHz p+p collision, 10-us strobe width+integration, 1 trigger, 10^{-4} noise per strobe

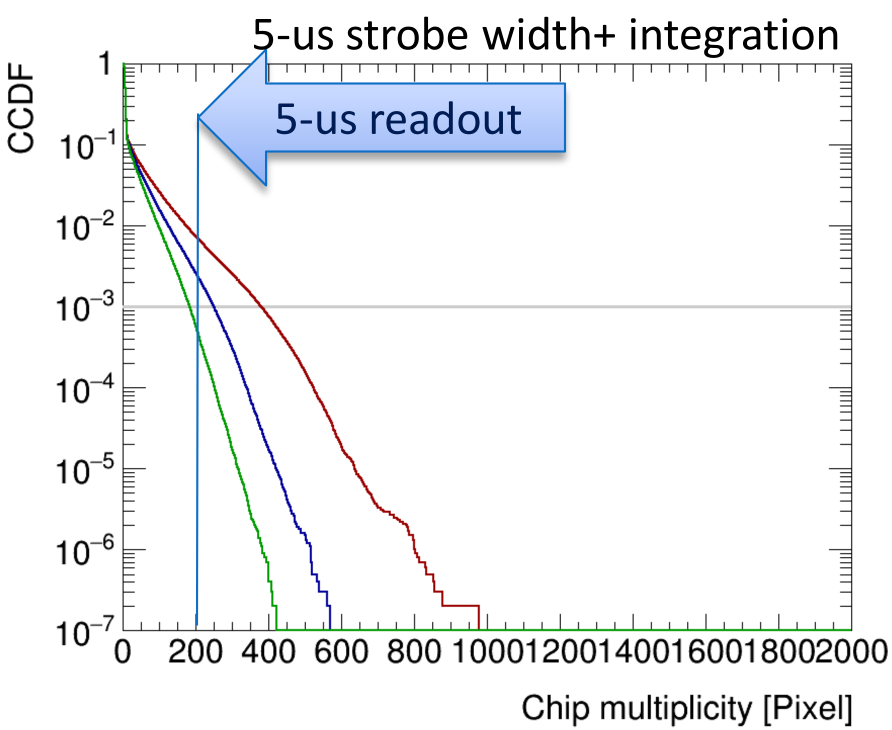


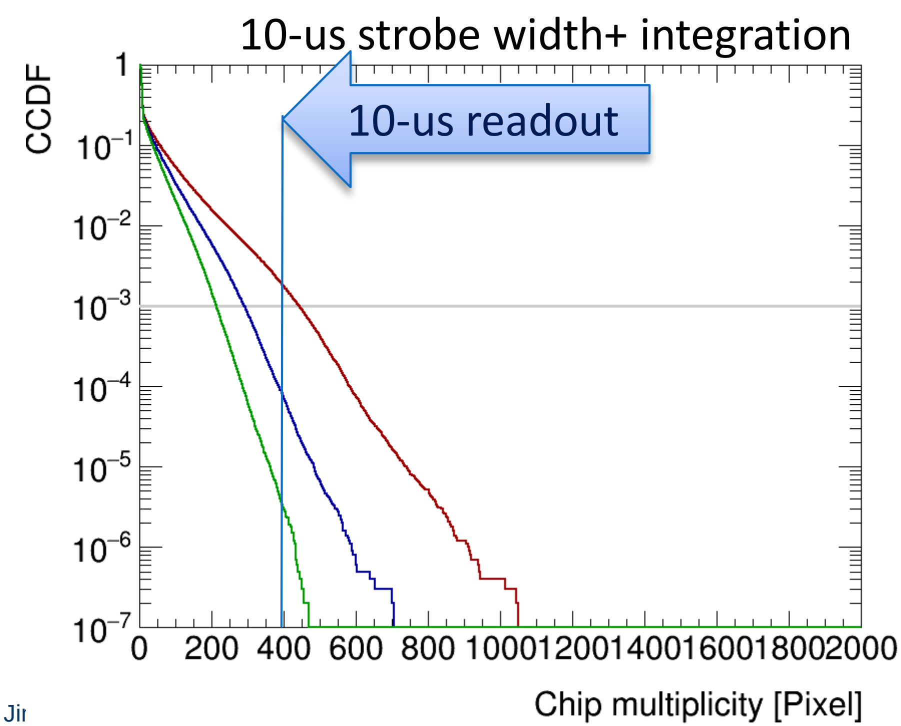


Au+Au multiplicity, per-strobe, chip-4

- Can we do better?
  - Further reducing collision rate to 50kHz by introducing a beam crossing angle
  - Reducing noise by 1/10 to 10^{-5} noise per strobe
- Still challenging for continuous , but plausible to have overflow dead-time < 0.1% further using multi-hit buffer on chip (eating the safety factor)

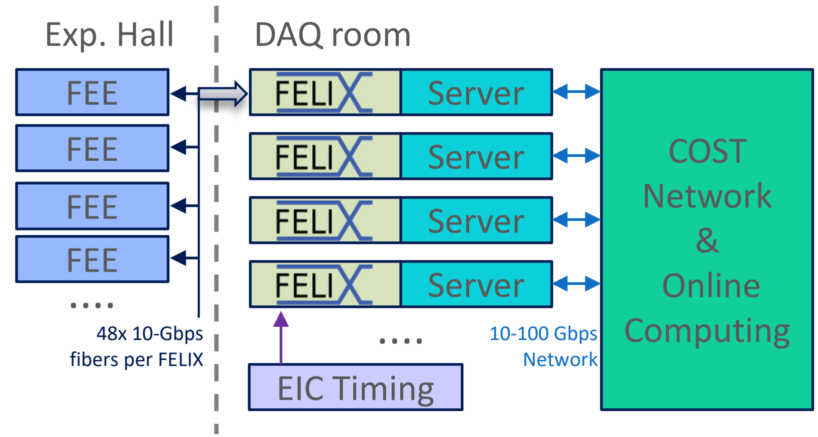
50 kHz Au+Au collision, periodic strobe, 10^{-5} noise per strobe





Timing distributions

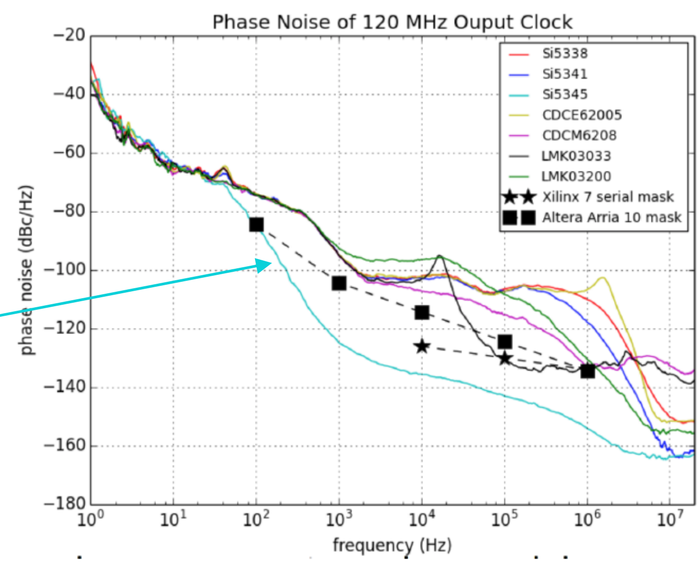
- All PHENIX/sPHENIX FEE are synced to beam clock/counter. Expecting similar for EIC detector
- BNL-712/FELIX can receive clock of multiple protocols (SPF+, White Rabbit, TTC, …) via a timing mezzanine card
- SI5345 jitter cleaner control jitter to <0.1 ps
- BNL-712/FELIX carries 48x 10 Gbps downlink fiber for control data to FEE. Beam clock and sync word can be encoded on fiber (e.g. 8b10b encoding)
- For EIC hadron beam RF, extra cautious need to be taken for hadron machine ramp from low gamma to high gamma, which leads to clock frequency variation [next slide].



timing mezzanine cards

Device	SI5338	SI5345	SI5341
Jitter (ps)	8.58	0.09	6.39
Device	CDCM6208	LMK03200	LMK03033
Jitter (ps)	2.06	5.91	2.74
Device	CDCE62005		
Jitter (ps)	8.61		

The jitter from 10 kHz to 1 MHz



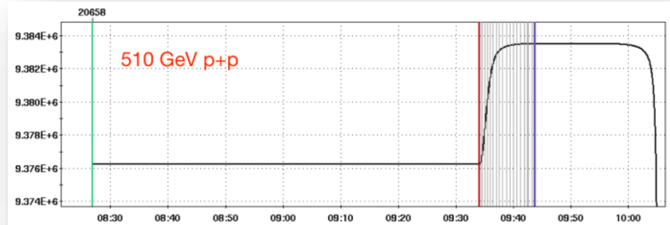
Courtesy of Kai Chen (BNL)

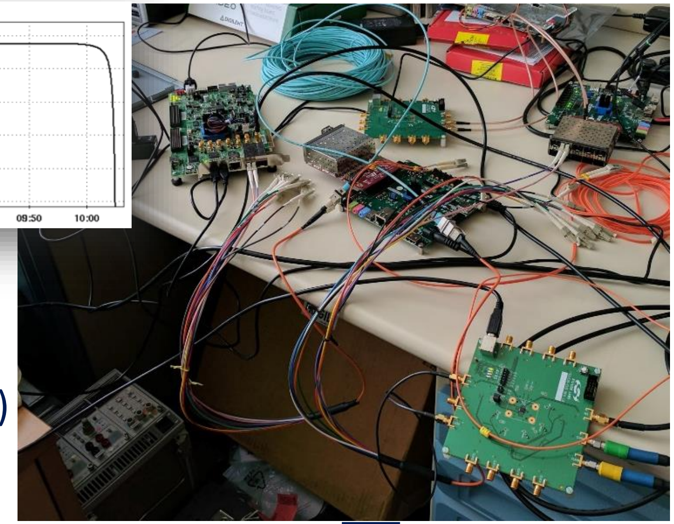
Kai Chen – FELIX Design Review

Embedded clock demo with variable beam clock frequency

Function generator mimic repeated RHIC clock ramping (triangle pattern)

RHIC frequency spread (due to ramp) is large, 9.362 MHz ±22 kHz





Demo FELIX
Kintex-7 Ultrascale

Downlink: 4.8 Gb/s
Multiples of RHIC clock (9.4 MHz)
Recover clock from 8b/10b

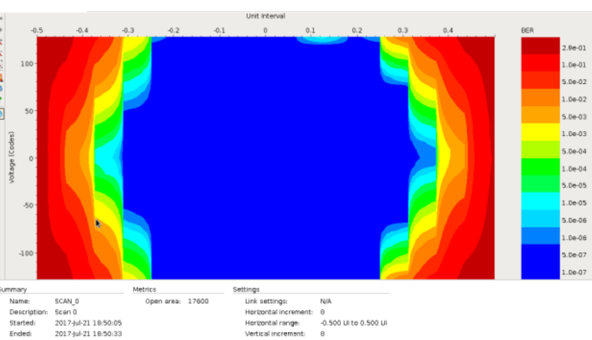
Optical Links

Demo FEE
Atrix-7

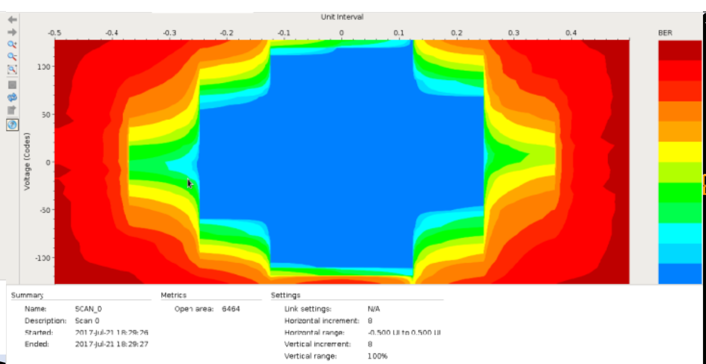
Uplink: 4.8 Gb/s, fixed clock

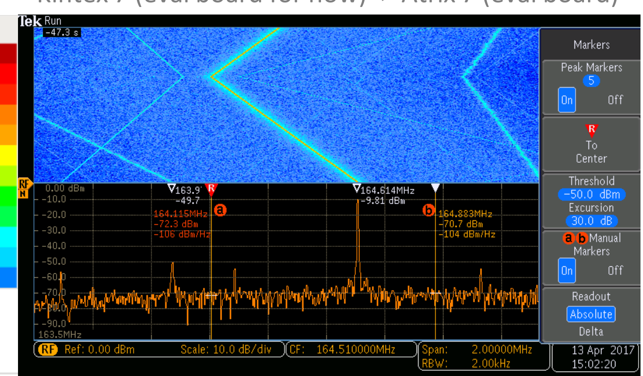
Test recovered "RHIC" clock
Kintex 7 (eval board for now) -> Atrix 7 (eval board)

Uplink iBERT @ DAM: 1.46e-13



Downlink iBERT @ FEE: 1. 023e-13





TOPSiDE: Concept of an EIC Detector

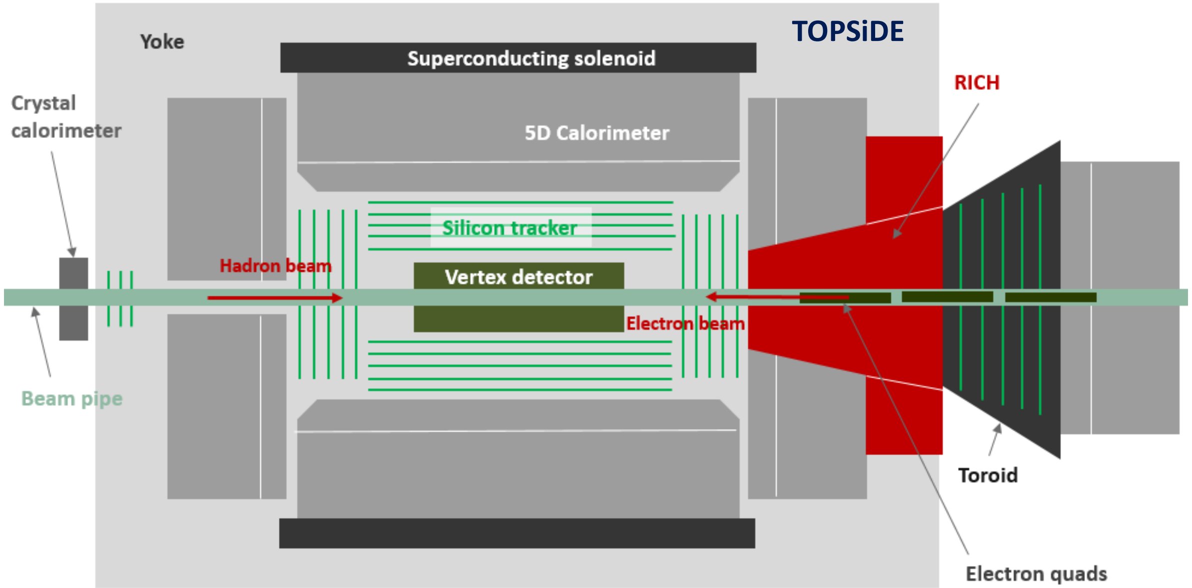
José Repond

5D Concept

Energy E

Position x,y,z

Time t



Salient features

4π detector (hermetic coverage)

Multi-purpose detector (don't need another specialized detector)

Mostly based on silicon sensors (tracker, electromagnetic calorimeter)

Each particle measured individually (optimized for Particle Flow Algorithms)

Particle identification (pion-kaon separation) performed by TOF (tracker and calorimeter)

Imaging calorimetry (tens of millions of readout channels)

Coil on the outside (not to disturb calorimetric measurements)

Dipole/Toroid in the forward direction (to obtain a momentum measurement)

Special detectors in the forward direction (Ring Imaging Cerenkov for Particle ID, debris taggers)

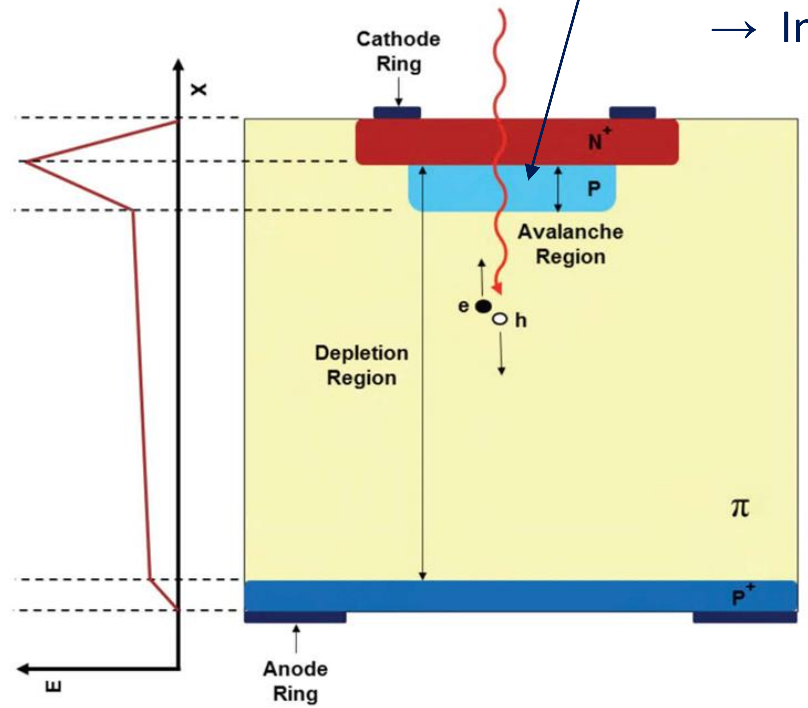
No need for additional TOFs, TRDs, Čerenkovs (in front of calorimeters), muon chambers

Low-Gain Avalanche Diode (LGAD) Sensors

José Repond

Additional thin p-layer

Through Boron/Gallium implantation
Increases E-field
Charge multiplication with moderate gain 10 – 50
Amplification of electrons close to pixel (minimal drift)
→ Improved time resolution



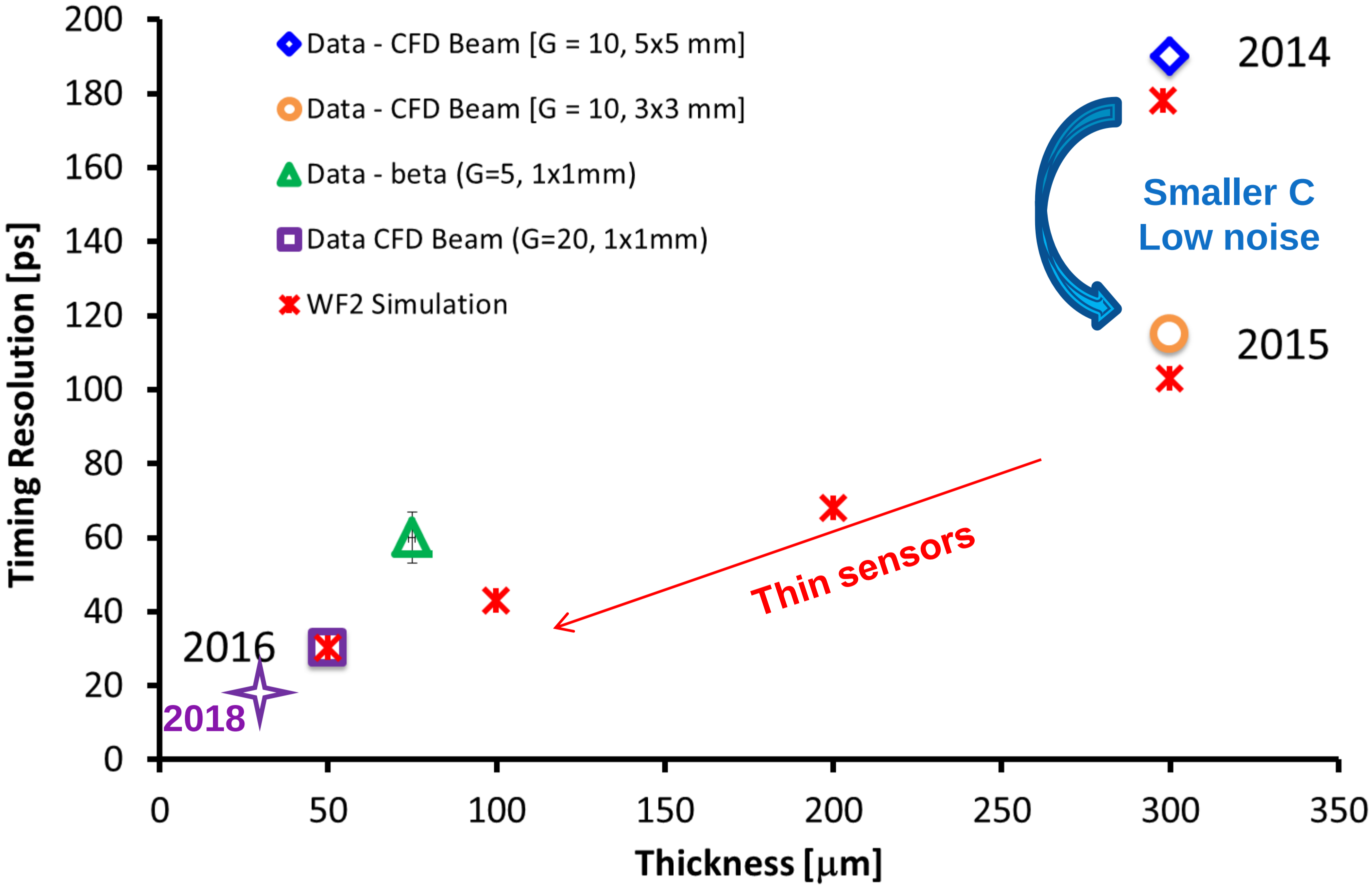
Four manufacturers

CNM Barcelona (RD50, ATLAS HGTD)
HPK Hamamatsu
FBK Trento (Italy)

Time Resolution versus Time

José Repond

H Sadrozinski at the Pico-second Timing Workshop, Torino, Italy, 2018
https://agenda.infn.it/conferenceTimeTable.py?confId=15031

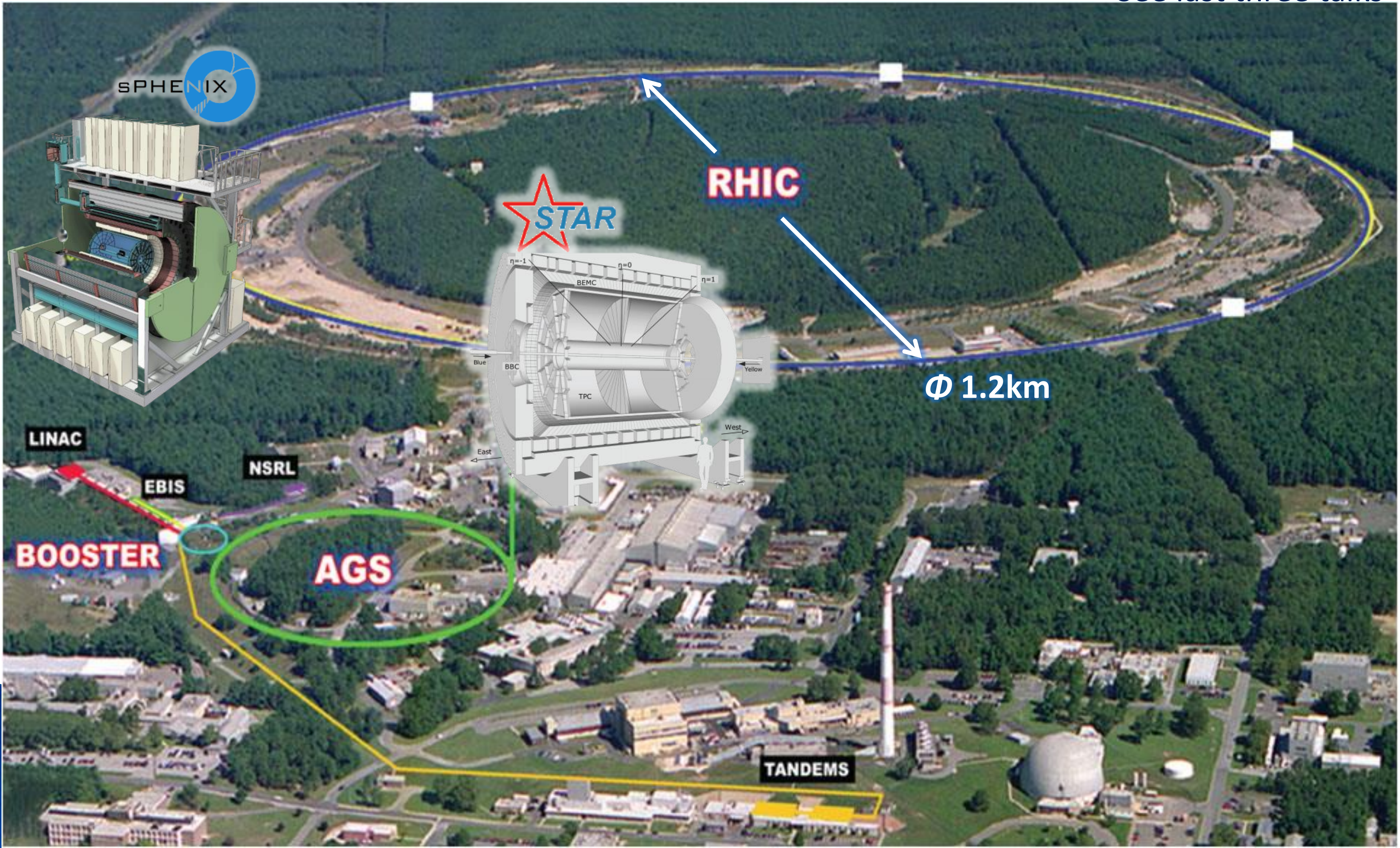


To date best results ~18 picosecond with 35 μm sensors

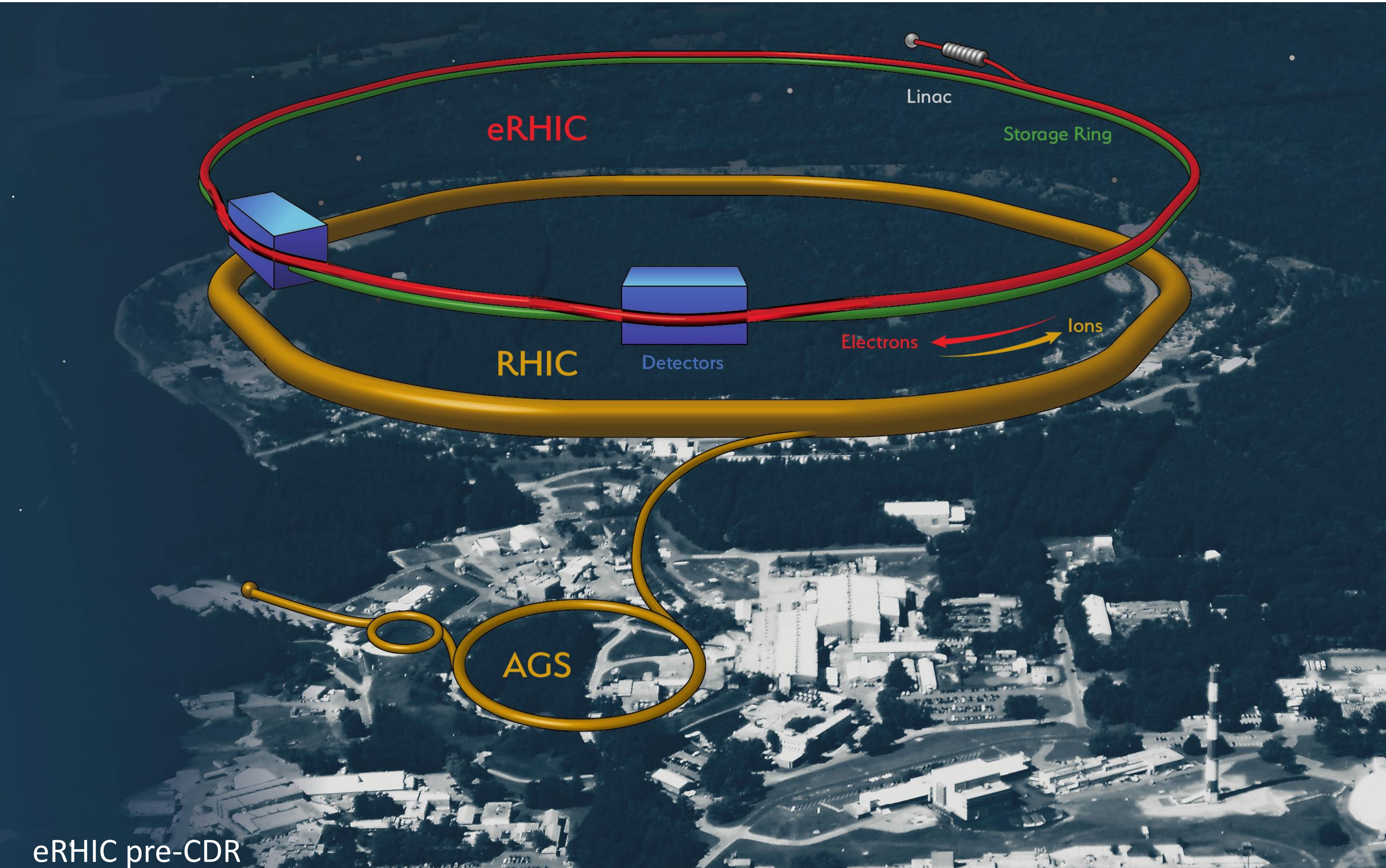
→ Results with 20 μm sensors imminent

RHIC @ mid-2020s

See last three talks



Proposed eRHIC @ end of 2020s



eRHIC pre-CDR

Potential areas of collaboration on real-time analysis

1. Real-time data reduction
2. Real-time feature building
3. Real-time/fast off-line event reconstruction
4. Large simulation sample generation that match or exceed real data event statistics
5. Data/theory-driven visualization

https://www.bnl.gov/compsci/

Computer Science and Mathematics

Computing for National Security

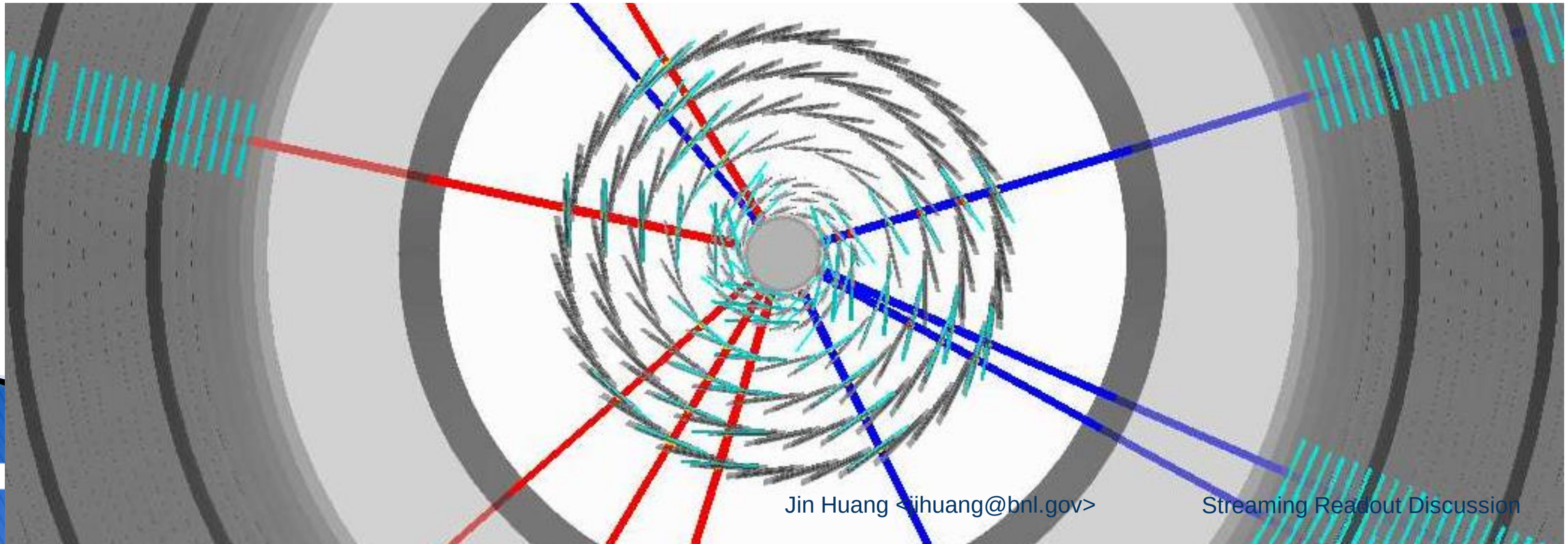
Scientific Data and Computer Center

Center for Data-Driven Discovery

Computational Science Laboratory

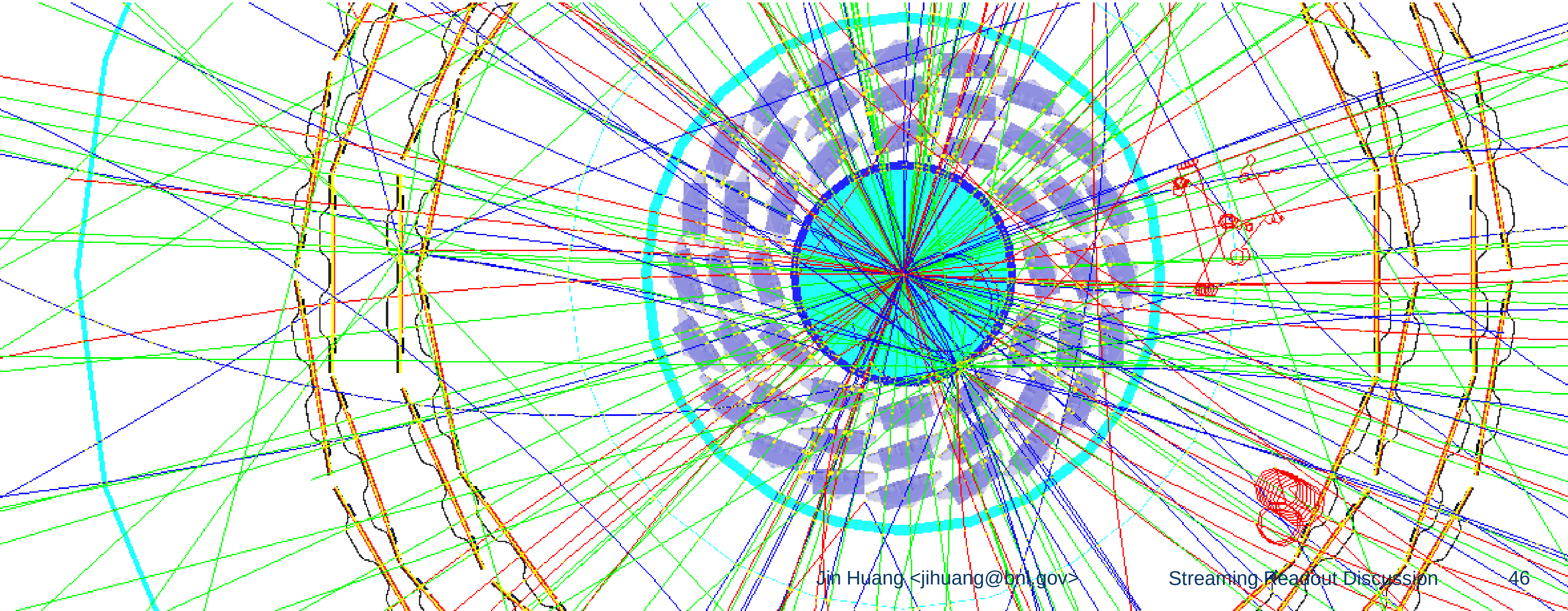
1. Suppression noise/beam background

- Considering that all tracker data is continuously recorded after zero suppression
- Half of collision-originated hits in tracker do not belong to a reconstructable particle trajectory (track)
- Electronic detector noise and collider backgrounds contribute more “noise” hits
- We could have a ML algorithm, e.g. DNN, to run in real-time on DAQ FPGA (e.g. Kintex Ultrascale) to filter out obvious hits that do not belong to a track
  - Unlike real-time triggering, this algorithm operate at low rejection, high efficiency ROC working point
- By filtering out noise hits, we could save on data storage volume, more resilient to high background operation



Publicly available dataset for ML/AI

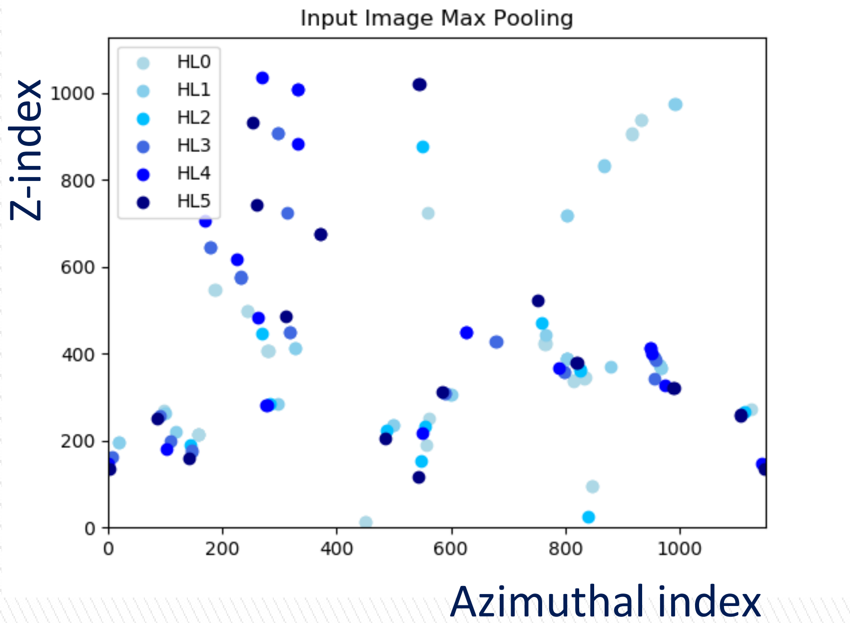
- Already have 4M simulated events in sPHENIX silicon tracking detector
  - https://github.com/sPHENIX-Collaboration/HFMLTrigger
  - JSON formatted data, self explanatory fields
- Can generate files for EIC collision + detector noise + background stream for algorithm development and performance evaluation



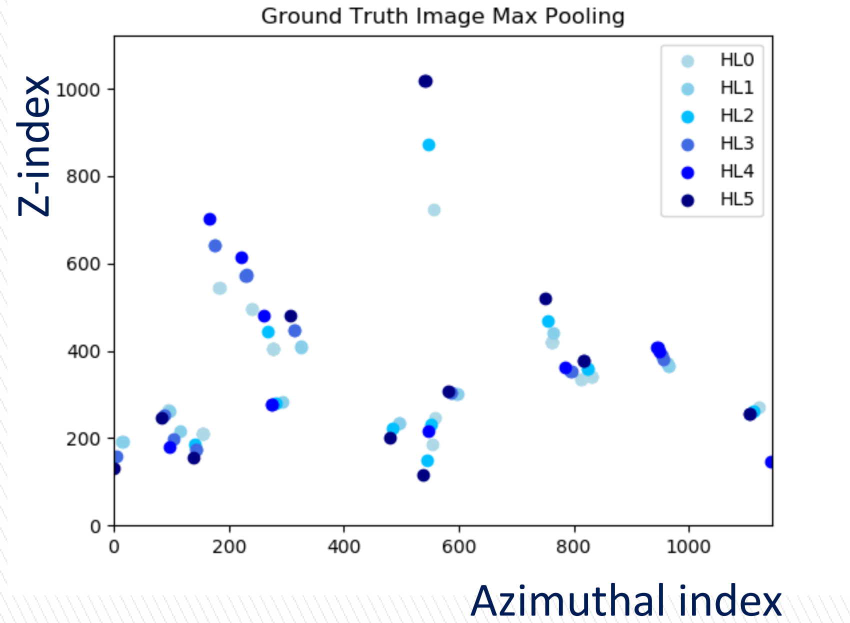
Open data used in DCNN learning

Courtesy of, Dantong Yu (NJIT), Yu Sun, Jason Chen (SBU)
Conversion of vertex tracker data (3D hits) into 2D image with color coding of layers

sPHENIX MVTX hits in a event → image
Simulated raw data. Composited picture of hits from six half layers.



sPHENIX MVTX hits in a event → image
Simulated ground truth. Hits belong to good truth track

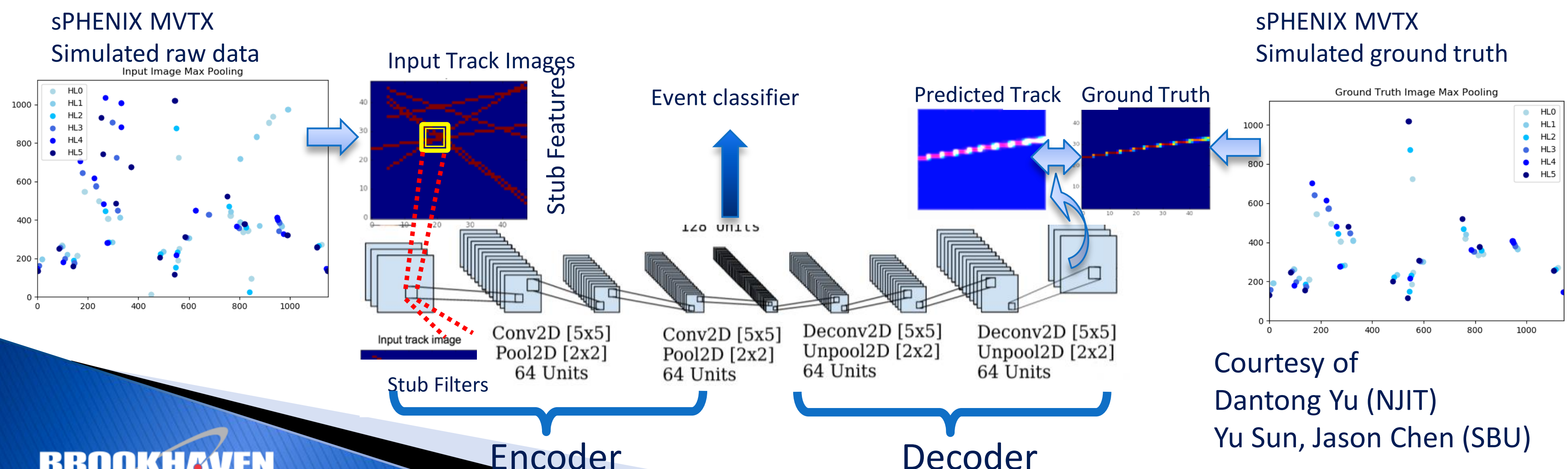


All MAPS hits
pp 200 GeV, EIC event would be similar

Reconstruct-able MVTX hits for tracks, same event

Open data used in DCNN learning

- Collaboration with computer scientists at NJ Inst. Tech and StonyBrook U.
- Exploring FPGA-HLS of Deep Convolutional Neural Network (DCNN) with capabilities of unsupervised-learning on data based on auto-encoder network
- Output of network can be used to filter out non-track hits
- “Code” level in the autoencoder may be used in another ML classifier for event tagging and event classification



Courtesy of
Dantong Yu (NJIT)
Yu Sun, Jason Chen (SBU)

2. Real-time feature building

- Real-time data reduction; improved zero-suppression

Examples :

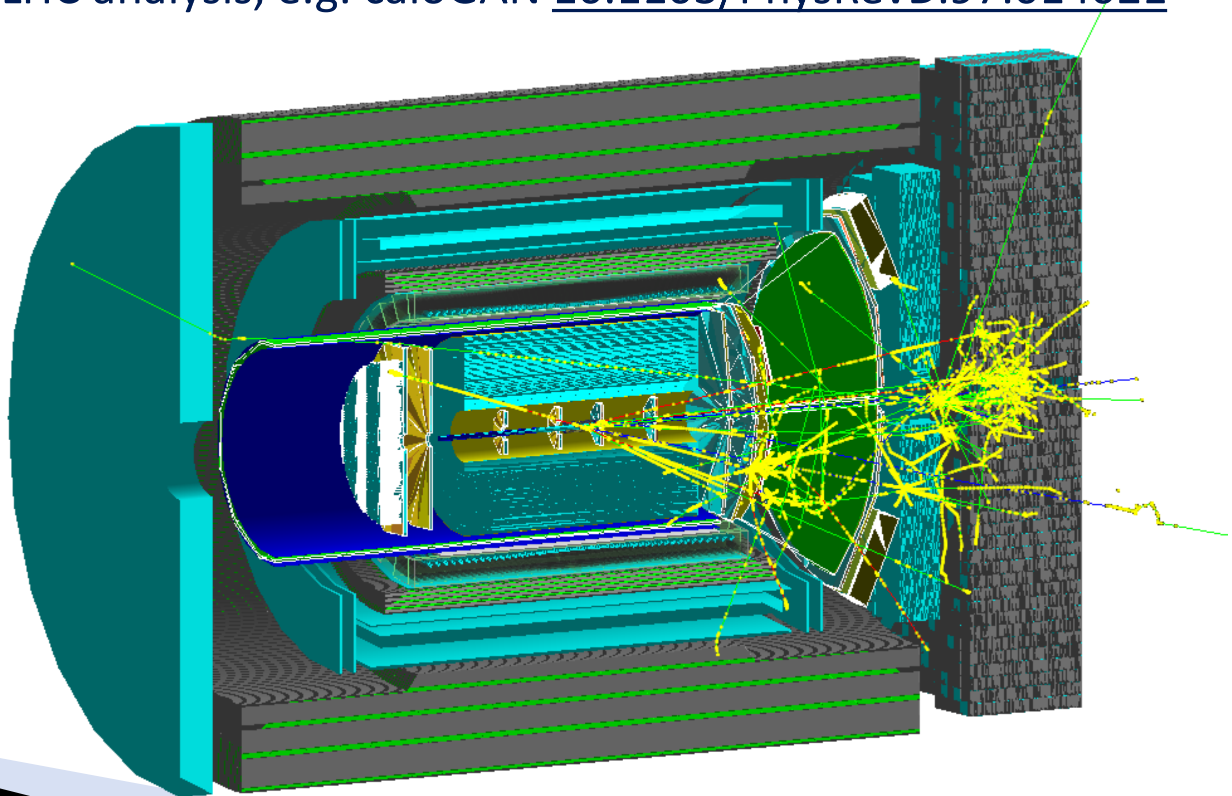
- Calorimeter
  - Multi-time sample fit → energy, shower features
  - Local clustering/triggering for improved zero-suppression. E.g. ALICE TPC: 10.1088/1742-6596/396/1/012043
- Time projection chamber
  - See detailed discussion in Martin's talk [sPHENIX]
  - Cluster building (1/2 reduction, see last talk)
  - Cluster – track matching (1/10 reduction, unlikely needed)

3. Real-time reconstruction needs

- Prompt reconstruction
  - Process data at same speed as data taking, but allow O(days) latency for final calibration
- Perform in commodity computing (CPU, GPU farms @ RCF-like facilities)
- Not a true "real-time" topic, but could use CSI's expertise. Examples:
  - Event tagging in time-framed continuous stream, event categorization
  - Pattern recognition (string hits to tracks, string track+PID+calorimeter to particle candidates)

4. Fast simulation

- Need high statistics simulations for multi-dimensional unfolding
  → Simulation speed needs to ~match data taking speed
- Detailed detector simulation currently take ~1 CPU-minute/event
- Use ML to generate event response in EIC detector
- Hot topic for ML in LHC analysis, e.g. caloGAN 10.1103/PhysRevD.97.014021

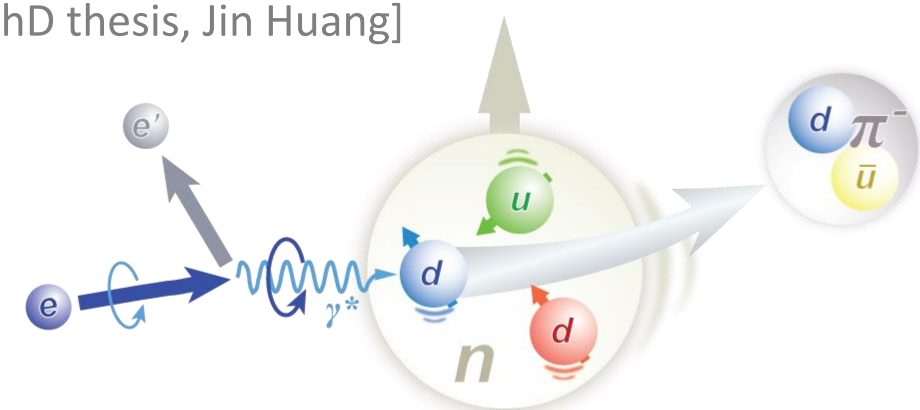


e+p collision 18+275 GeV/c
DIS @ Q^2 ~ 100 $(GeV/c)^2$

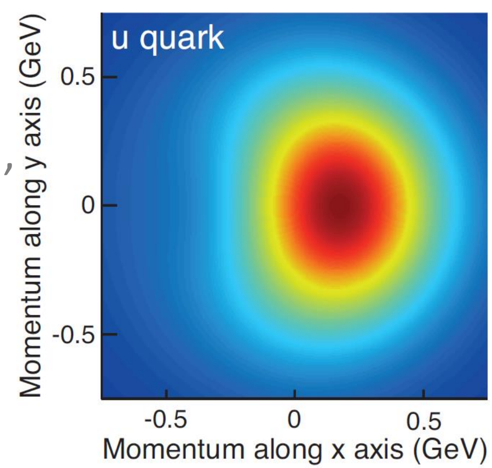
5. Visualization of EIC process @ $\Delta t = 10^{-24}s$

- E.g. data/theory driven 3-D movie of EIC collision

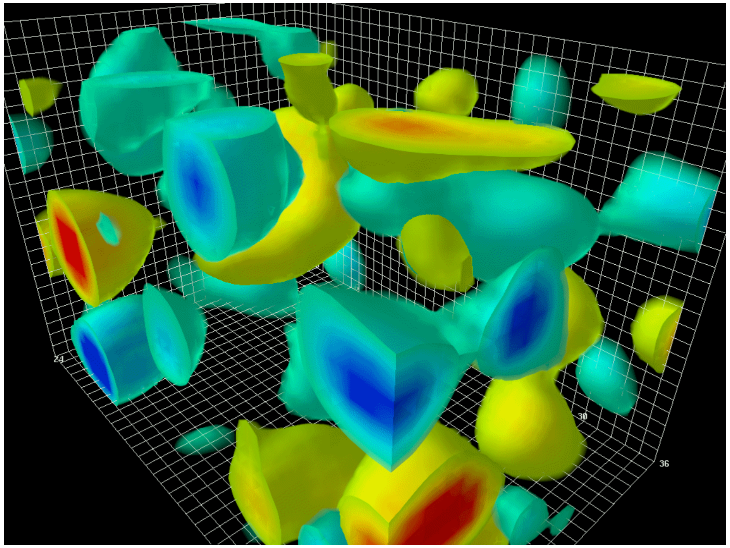
Static visualization of DIS collision
[PhD thesis, Jin Huang]



Global fit on quark distribution in proton
[u-quark Sieves function, EIC white paper]



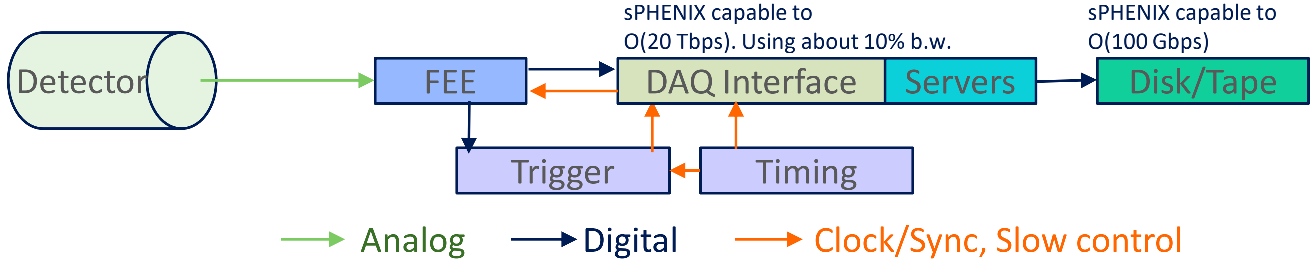
Example 3D visualization
Fluctuation of gluon field [US QCD]



Summary

- EIC is a unique collider combining high luminosity and low collision cross section, versatile events
  - The total collision signal is about 100 Gbps, plausible to write all to disk. However, background estimate is still early.
  - A real-time strategy for EIC: stream all signal data to disk/tape, real-time background hit filter, prompt off-line reconstruction
- Many areas for collaboration with CSI:
  - Real-time background reduction, feature building; fast event reconstruction and simulations; Visualization
  - We can provide open dataset extracted from EIC simulation for algorithm development and evaluation .e.g https://github.com/sPHENIX-Collaboration/HFMLTrigger

Strategy discussion



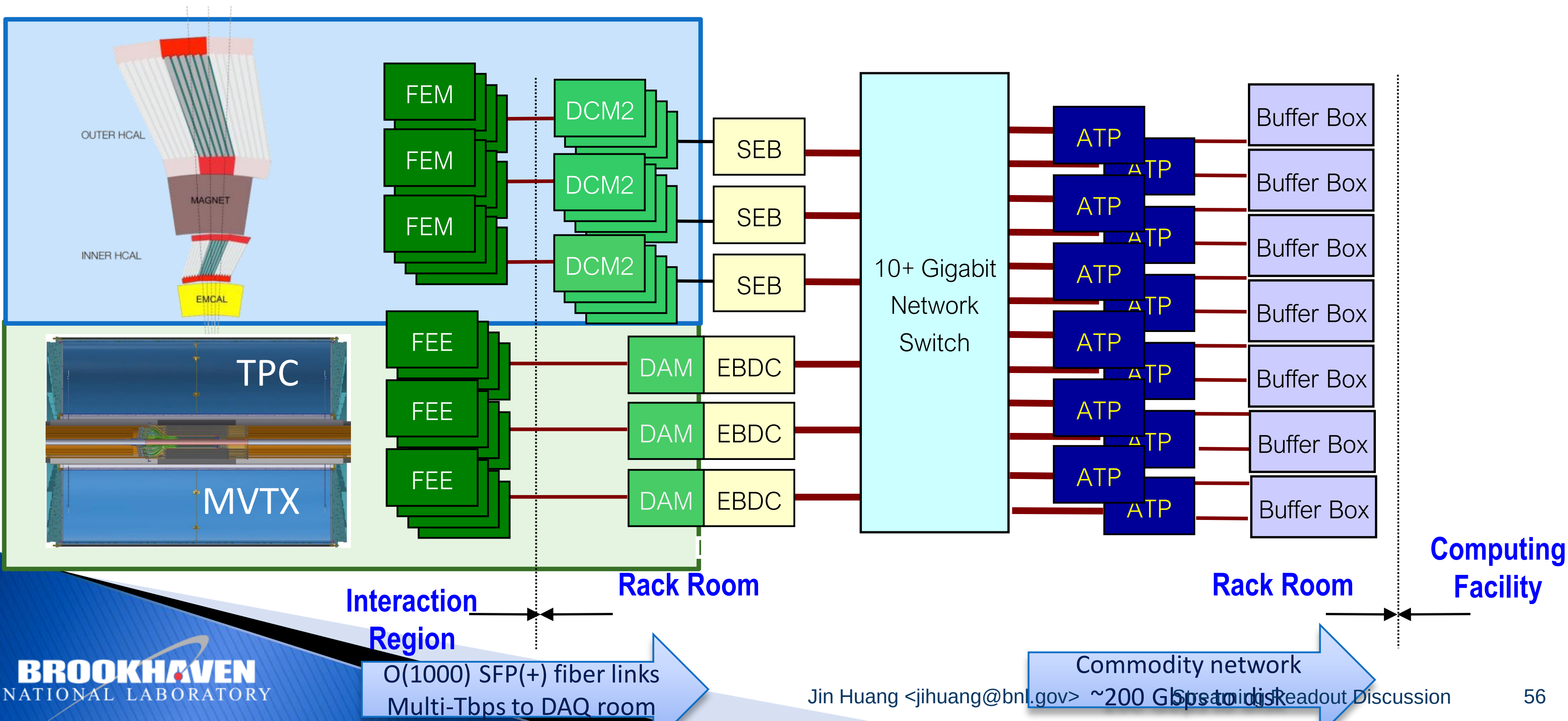
- Strategy depends on the detector, data rate and physics goals. I feel it factorizes into two choices:
- Front-end electronic (FEE) choice: digitizer + on-FEE FPGA
  - Real-time (<10us) triggered, e.g. STAR, sPHENIX calorimeter, ATLAS, CMS
  - Zero-suppressed/self triggered streaming , e.g. PHENIX FVTX, sPHENIX trackers, STAR eTOF
  - None-zero suppressed streaming , e.g. ALICE TPC, ePHENIX calorimeter?
- DAQ choice: reducing data O(<100 Tbps) -> O(~100Gbps)
  - Passthrough w/ lossless compression (PHENIX, sPHENIX)
  - Further event building and selection (HLTs)
  - Feature building in time-slices (LHCb, CBM)

PHENIX and sPHENIX Mixed-Mode DAQ, LDRD 19-026

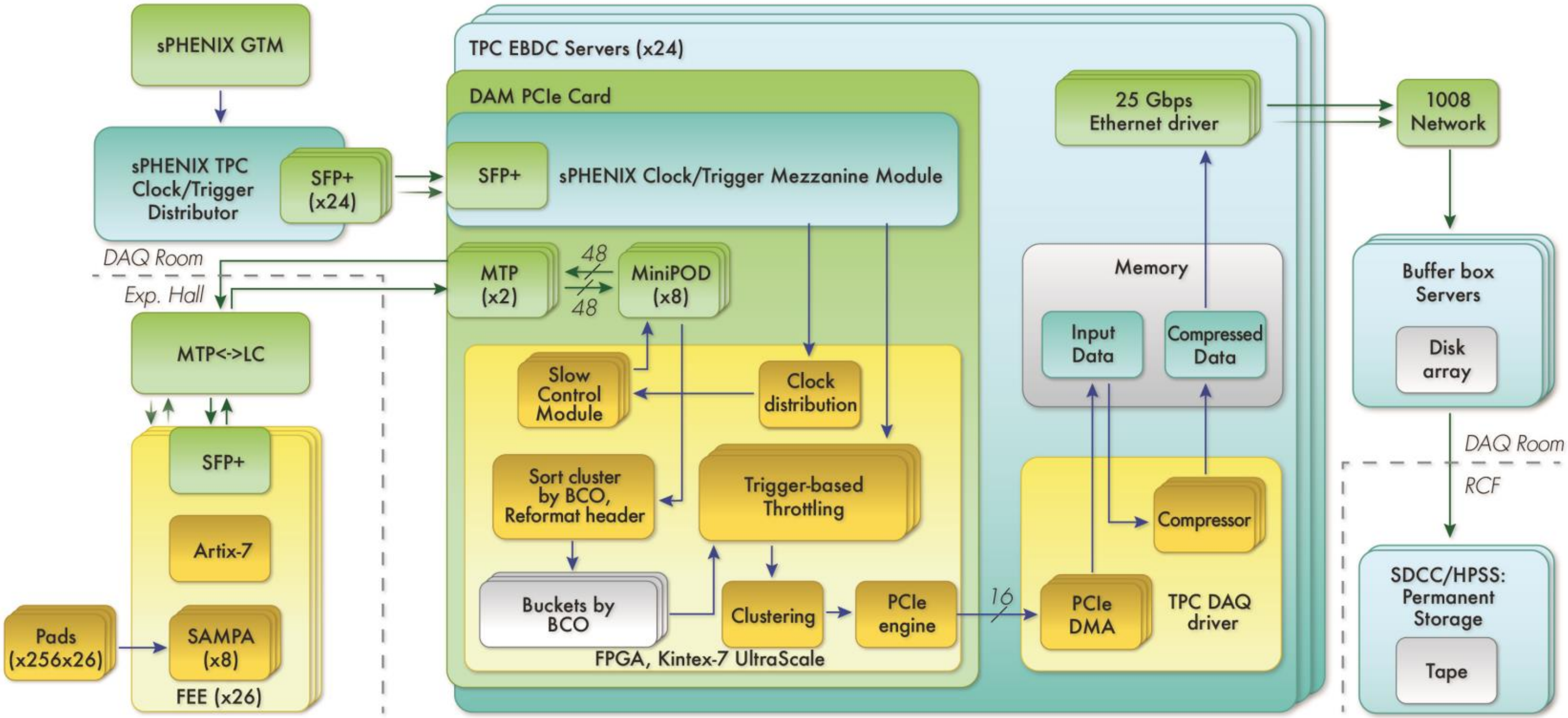
- Also in communication with CBM group as they install streaming eTOF in STAR

Selection of streaming and triggered front end

- **For calorimeter triggered FEE**,
  (signal collision rate 15kHz x signal span 200ns) << 1:
  No need for streaming readout which significantly reduce front-end transmission rate
- **For TPC and MVTX tracker FEE supports full streaming**:
  (signal collision rate 15kHz x integration time 10-20us) ~ 1:
  Streaming readout fits this scenario. Consider late stage data reduction by trigger-based filtering



TPC DAQ in streaming mode



1 sector, 26 FEEs per DAM for readout
24 sectors, 160k Pads and 624 FEEs
24DAMs total

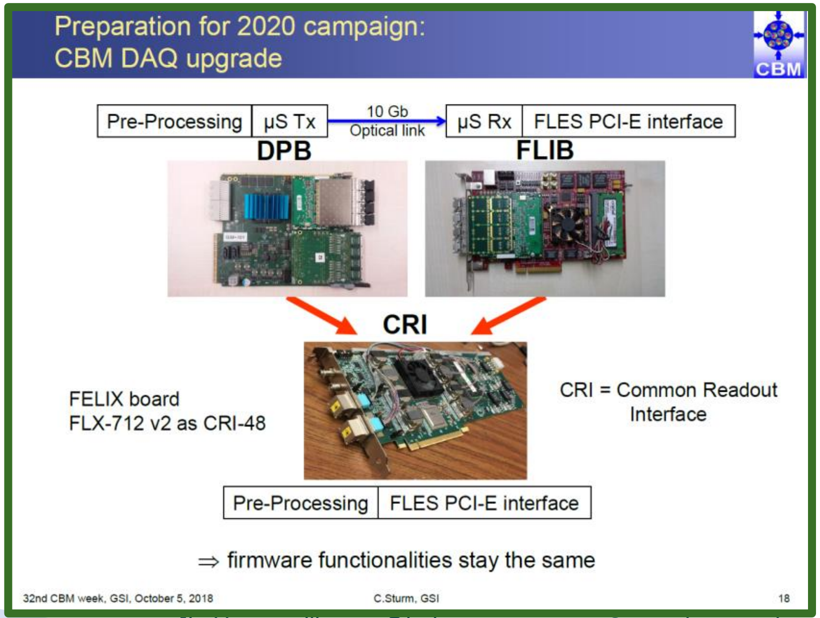
600 Fibers @ 600x 6 Gbps

Commodity networking @ 200 Gbps

Productions for BNL-712v2/FELIXv2

- Ongoing FY19 BNL-712v2/FELIXv2 card production from BNL covering sPHENIX advanced R&D
  - CBM working on joining this production and adopting this architecture for 2020 campaign too.
  - 2nd sPHENIX production planned after sPHENIX CD-3B (FY20?)
  - BNL produced 40x cards in various versions of FELIX in ATLAS pre-productions, which will continue too.
- Synergies from further EIC stream readout R&D welcomed too

Courtesy of C. Stum, D. Emschermann (GSI)

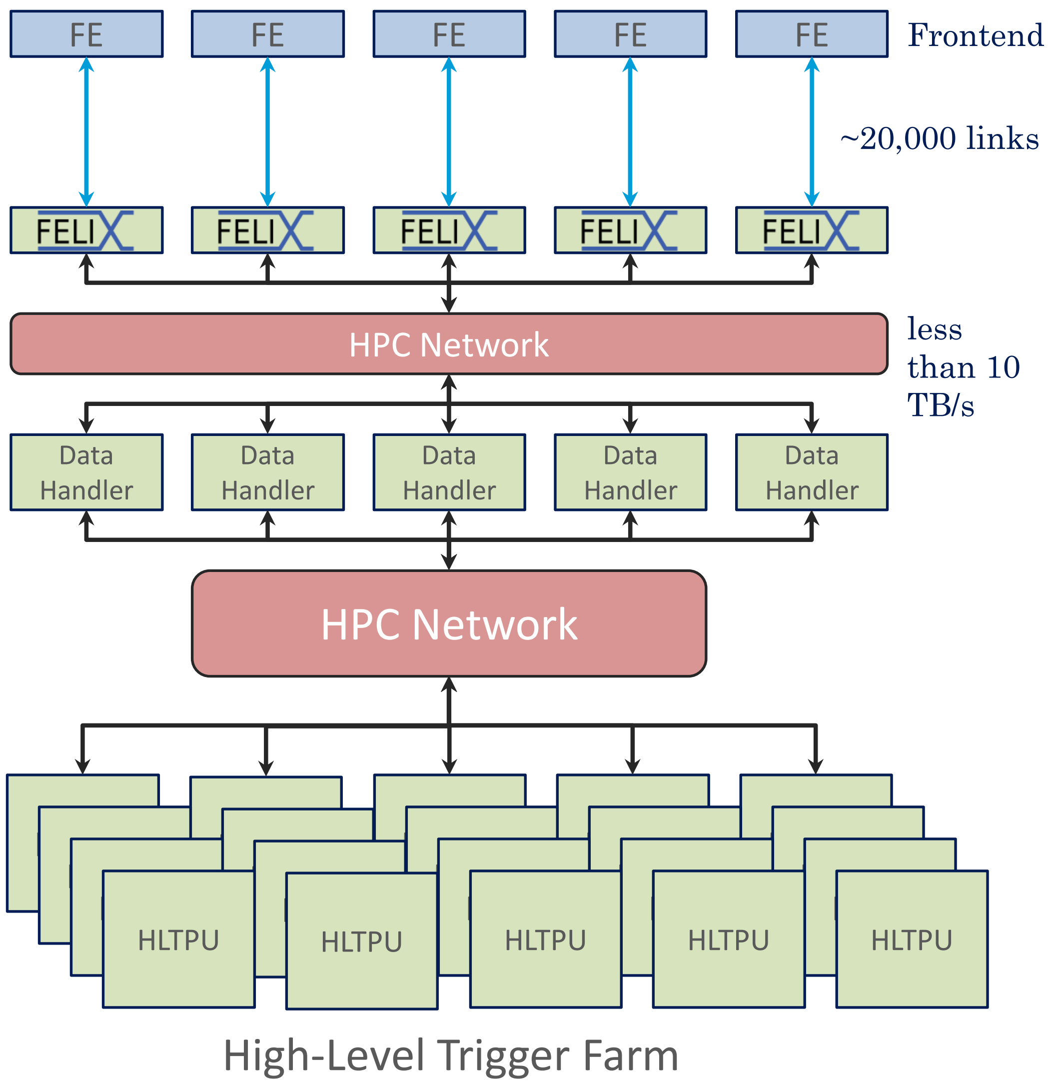


From Kai Chen (BNL)

Upgrade for HL-LHC

GBT, LpGBT* or FULL mode links

COTS network technology



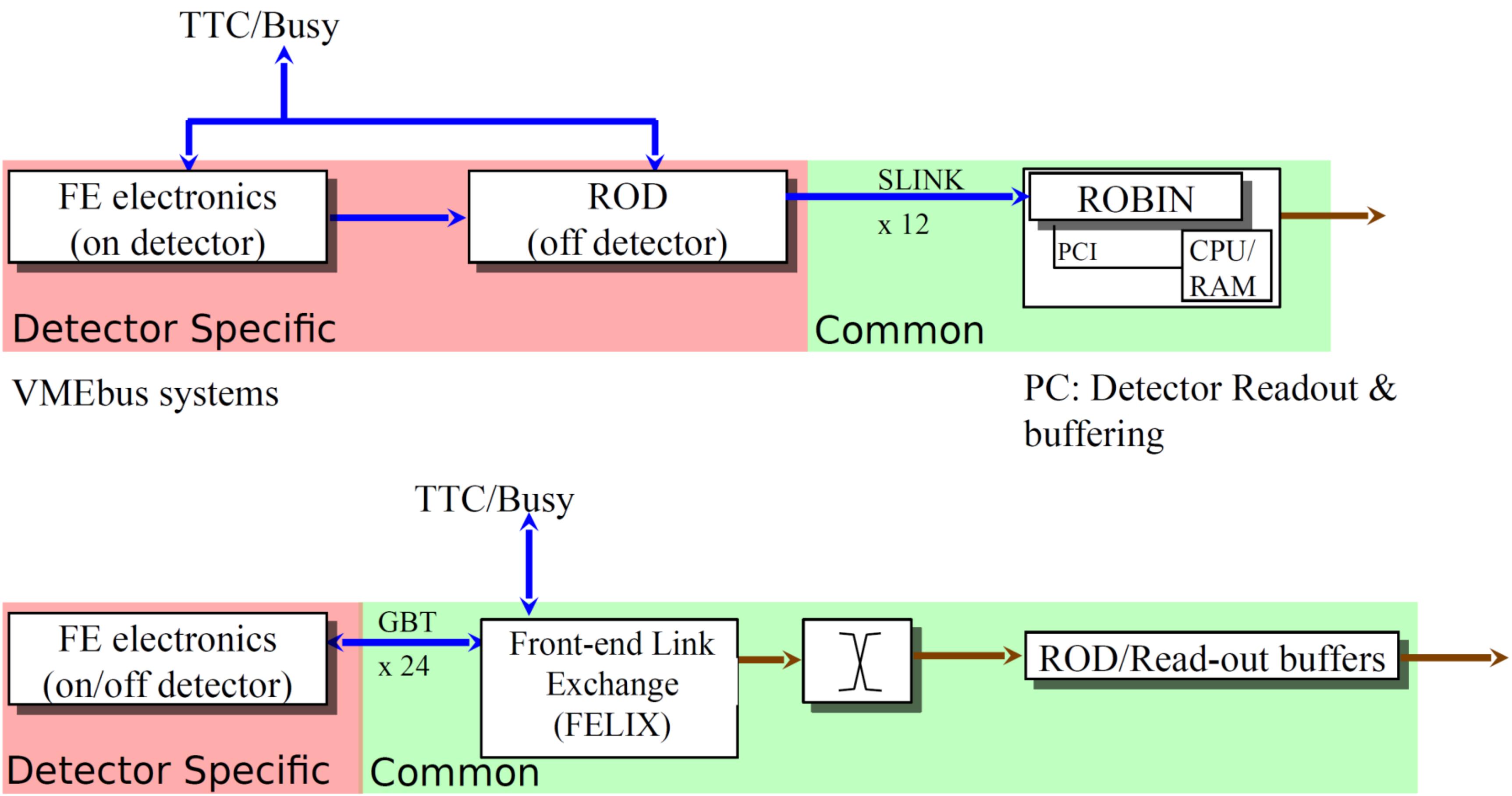
Custom electronic components including FELIX cards

PCs (COTs)

*LpGBT: Low power GBT

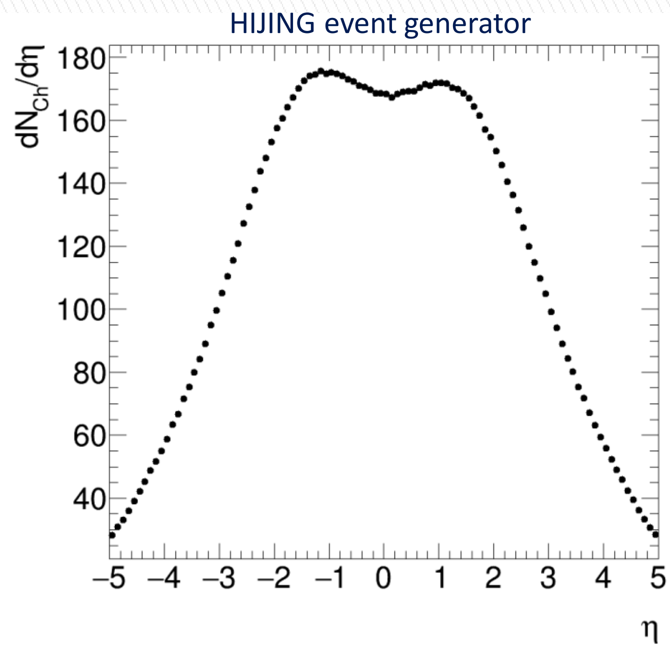
* COTs: Commercial off-the-shelf

Detector Readout Run 3



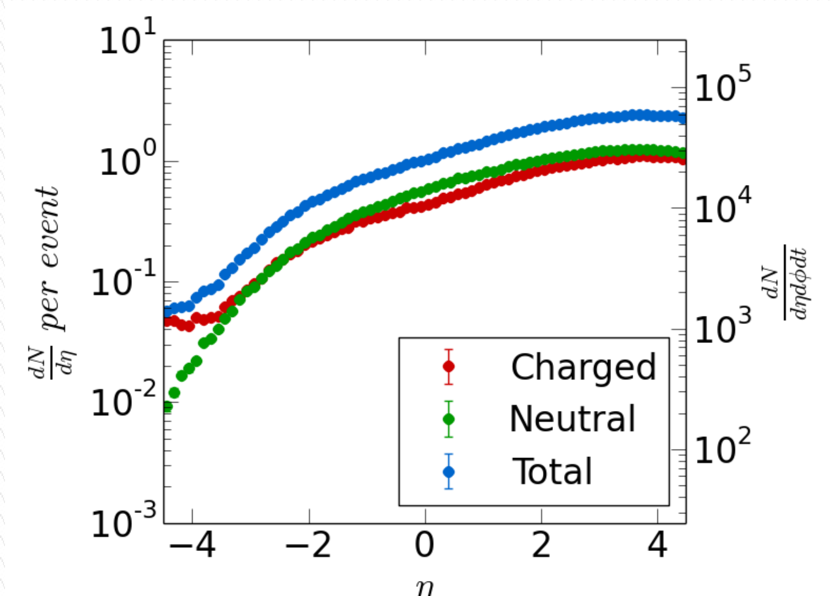
sPHENIX rate VS EIC charge track rate

Charged multiplicity, Au+Au, 100 + 100 GeV/c



Multiplicity, e+p 20+250 GeV/c, 50μb

https://wiki.bnl.gov/eic/index.php/Detector_Design_Requirements



sPHENIX AuAu $dN_{ch}/d\eta$ ~200, |η|<1

Streaming readout @ 200 kHz collision : 80 M N_{ch}/s

DAQ throughput @ trigger rate 15 kHz: 6 M N_{ch}/s + pile up

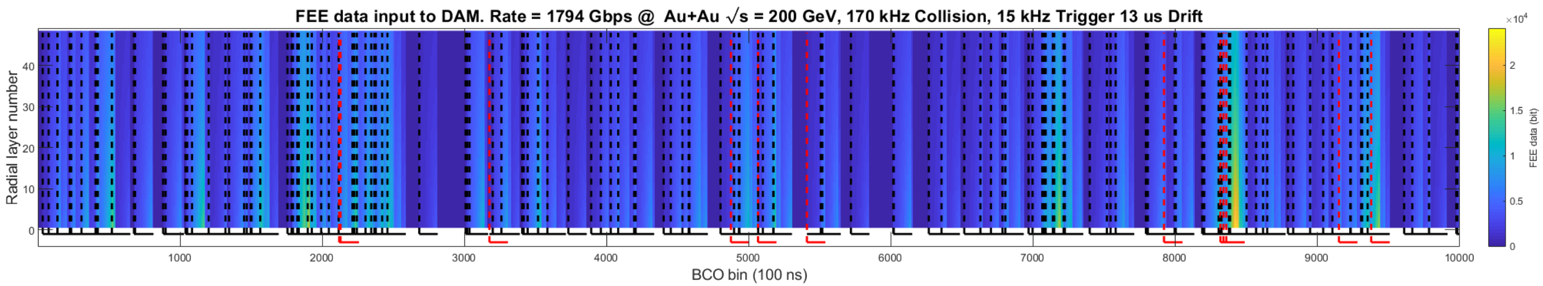
EIC 20+250 GeV/c $dN_{ch}/d\eta$ ~1, |η|<4

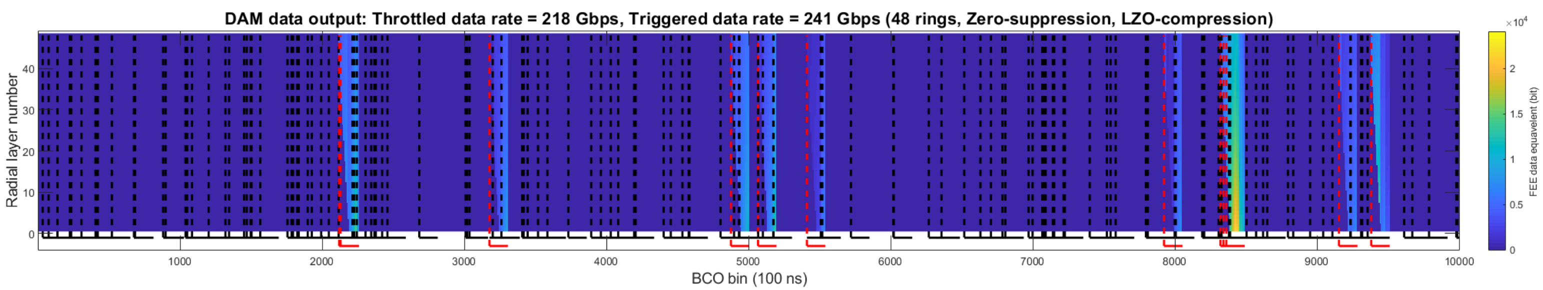
Streaming readout @ 500kHz collision (10^{34} cm^{-2} s^{-1}) : 4 M N_{ch}/s << sPHENIX

DAQ throughput, full stream: 4 M N_{ch}/s <~ sPHENIX

TPC data rate

- TPC is the dominating data contributor to sPHENIX event
- Using past <dNch/deta>x2 estimation, expect event size is
  - Single MB collision, no pile up:
    - 1.05 MB/event (before compression)
  - Year-5 average, MB + 170kHz AuAu (plots below):
    - 3.3 MB /event (before compression)
    - 240 Gbps (15kHz trigger, LZO compression)
- Now simulating the event size and data rate in Geant4 simulation.





Buffer Box hardware

We don't have to buy the off-the-shelf PCs until 2022

If we would buy today, this would be the candidate for a buffer box:

$34,000 fully equipped

84 disk slots x 8TB = 672TB raw size

Other PCs (SEB, EBDC, ATP) are standard rack-mounted PCs, too

Total Data volumes

Year 1: 47 billion events * 1.7 MB = 75 PB (Au + Au)

LTO-9: 75 * 1024 TB / 20TB = 3840 tape cartridges

LTO-10: 75 * 1024 TB / 48TB = 1600 cartridges

Year 2,4: 96 billion events * 1.6 MB = 143 PB (p + p)

LTO-9: 143 * 1024 TB / 20TB = 7300 tape cartridges

LTO-10: 143 * 1024 TB / 48TB = 3500 cartridges

Year 3, 5: 96 billion events * 2.3 MB = 205 PB (Au + Au)

LTO-10: 205 * 1024 TB / 48TB = 4400 cartridges

LTO tape vendors announce LTO-9 and LTO-10

LTO tape vendors extend the LTO roadmap to include generations 9 and 10 with increasing capacity and transfer rates.

The 2023-era tape drives ("LTO-9") can sustain about 4.5GBit/s real-world throughput (20TB capacity)

Next-gen LTO-10 has 8GBit/s throughput (48TB)

Current-generation LTO-8

HPSS

High Performance Storage System

~90 PB of data on tapes

~60K+ tapes, mix of LTO 4,5,6 and T10KD technologies

~900 TB total disk cache

BROOKHAVEN NATIONAL LABORATORY

Peak Data rates

Peak data rates determine how many tape drives will be needed

Based on a “high performance week” with 75% * 75% combined uptime in Year-1

75% * 80% combined uptime in year-2,3,5 (instead of 60% * 80%)

Year 1: 5 billion events * 1.7 MB * 8 / (7*24*3600) = 109 Gbit/s peak (14.5 weeks)

Year 2,4: 5.5 billion events * 1.6 MB * 8 / (7*24*3600) = 113 Gbit/s peak (22 weeks)

Year 3,5: 6 billion events * 2.3 MB * 8 / (7*24*3600) = 178 Gbit/s peak (22 weeks)

Year 1, 2, 4

LTO-9: 4.5 Gbit/s → 25 tape drives

LTO-10: 8 Gbit/s → 14 Tape drives

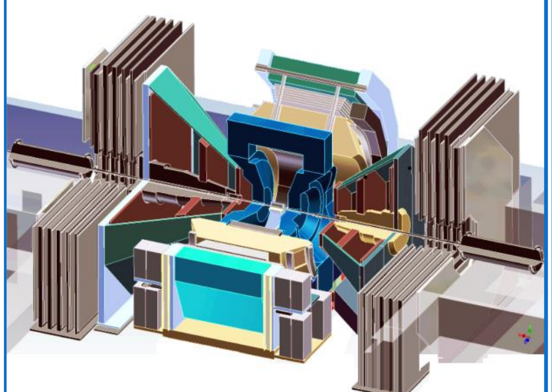
Year 3,5

LTO-10: 8 Gbit/s → 23 Tape drives

Evolution of the RHIC 1008 Interaction region

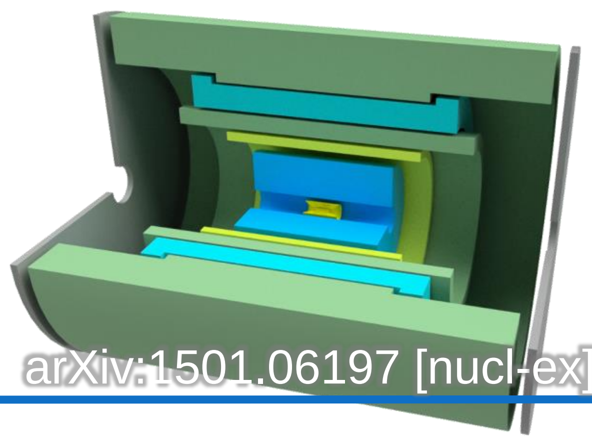
PHENIX experiment

- 16y+ operation
- Broad spectrum of physics 180+ physics papers with 25k citations
- 1.4-M channel streaming



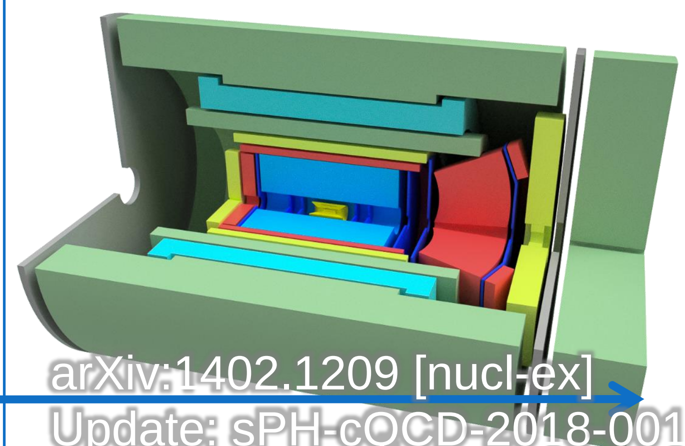
sPHENIX

- Comprehensive central upgrade base on previous BaBar magnet
- Rich jet and HF physics program
  → Microscopic nature of QGP



An EIC detector

- Path of PHENIX upgrade leads to a capable EIC detector
- Large coverage of tracking, calorimetry and PID
- Full streaming DAQ based on sPHENIX



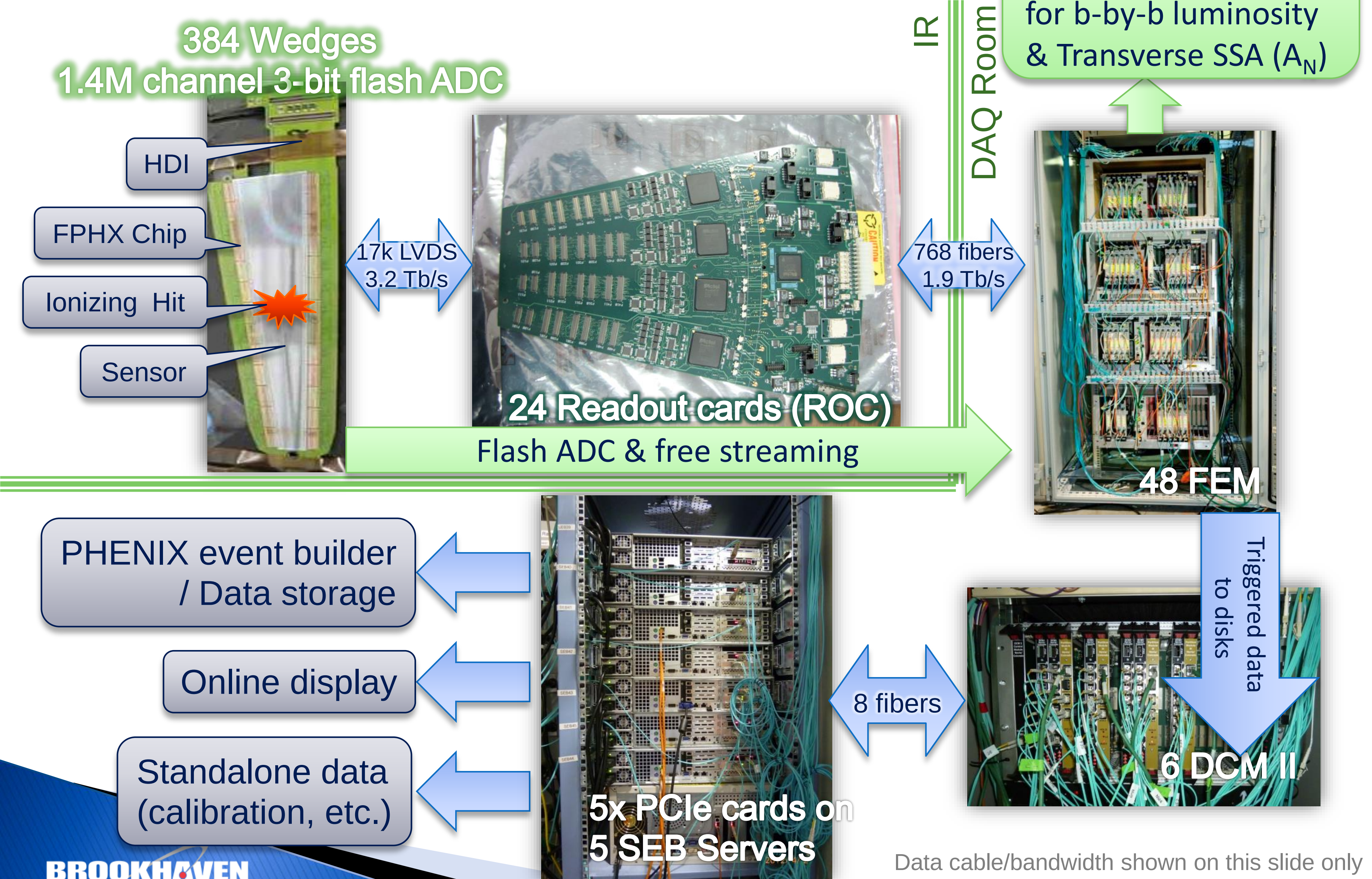
~2000 | 2017→2023, CD-1/3A Approved | >2025 | Time

RHIC: A+A, spin-polarized p+p, spin-polarized p+A

EIC: e+p, e+A

PHENIX/FVTX Streaming FEE

Streaming data processing on FPGA for b-by-b luminosity & Transverse SSA (A_N)

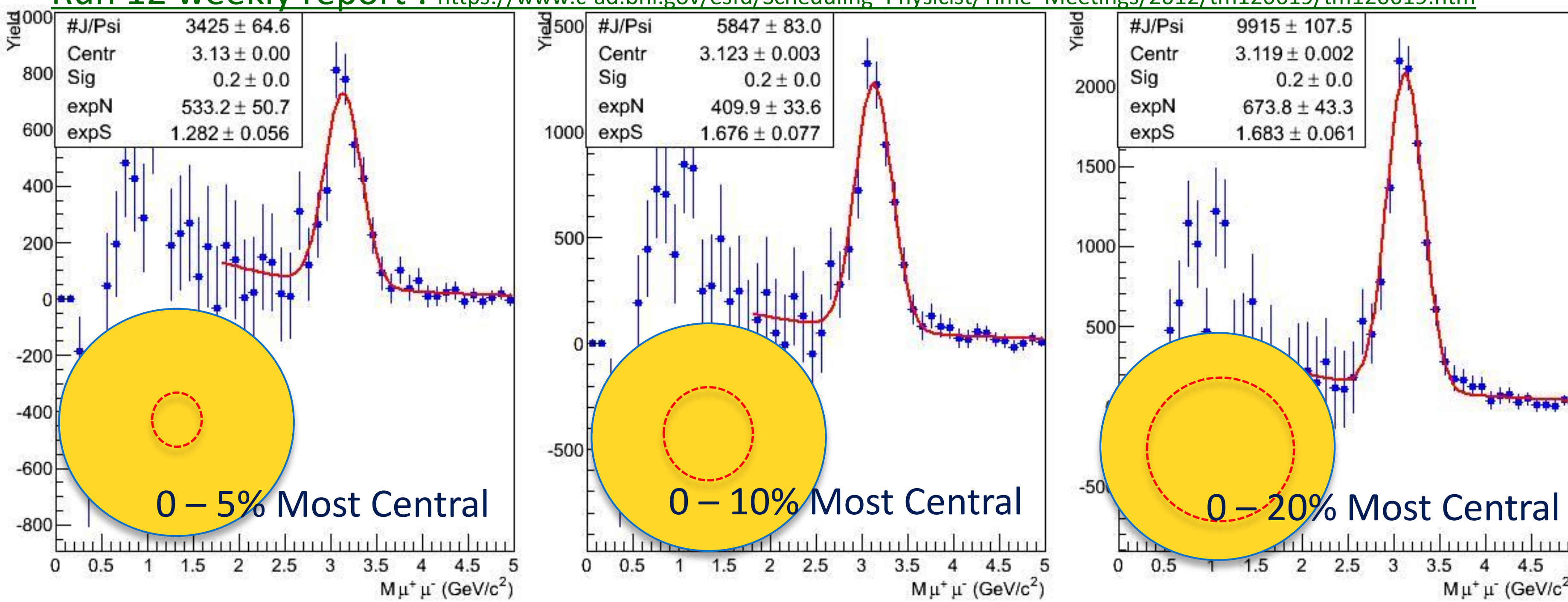


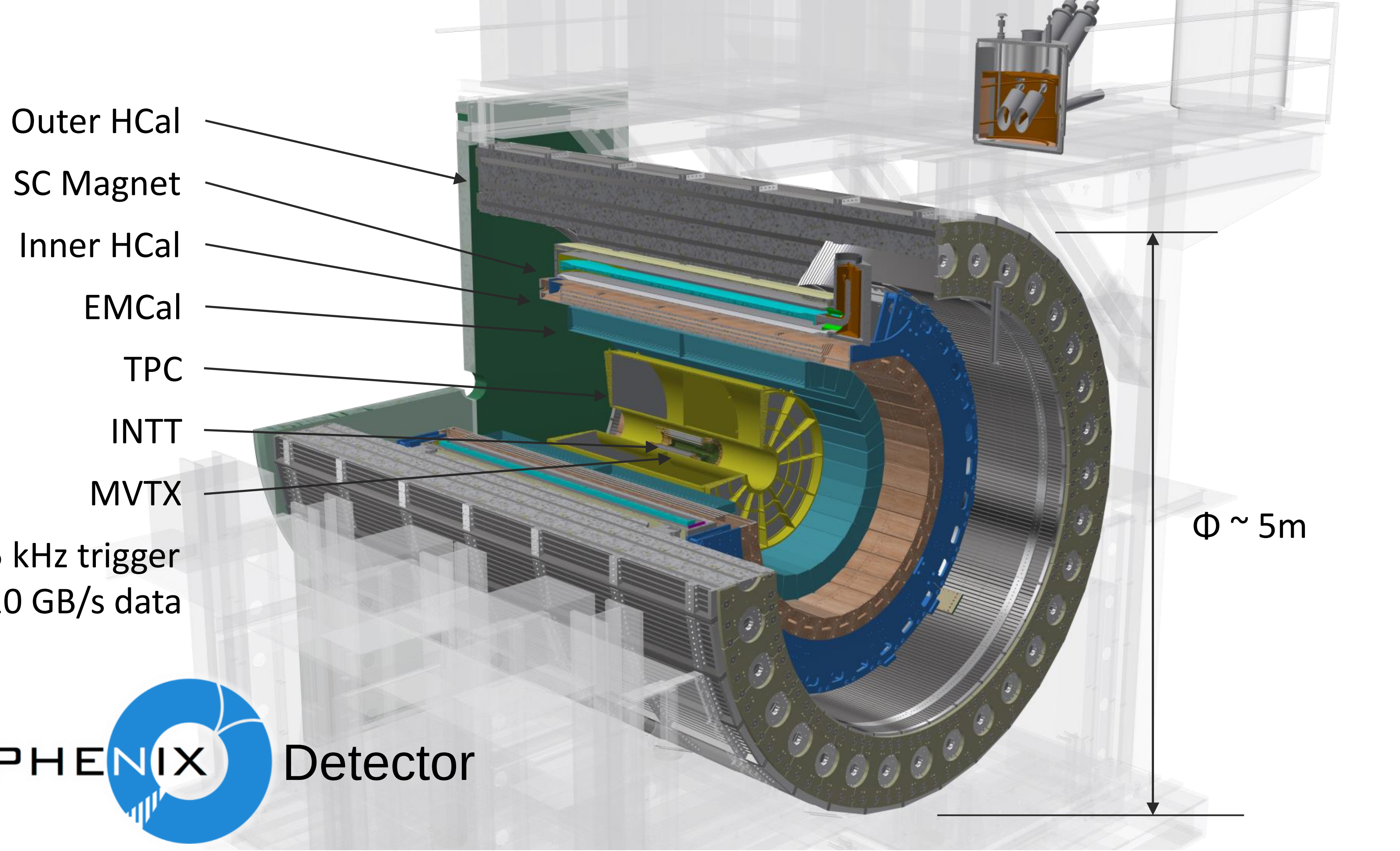
Data cable/bandwidth shown on this slide only

PHENIX Data validation & data processing in near-real-time

- PHENIX validate data and perform majority calibration in near-real-time via online system using a subset of raw data prior to disk write
- PHENIX has enough CPU to final process all data in real-time, but the limitation is usually special data need and manpower for calibration

J/Psi spectrum in Cu+Au @ sqrtS = 200 GeV via run-time data production & analysis, Run 12 weekly report : https://www.c-ad.bnl.gov/esfd/Scheduling_Physicist/Time_Meetings/2012/tm120619/tm120619.htm



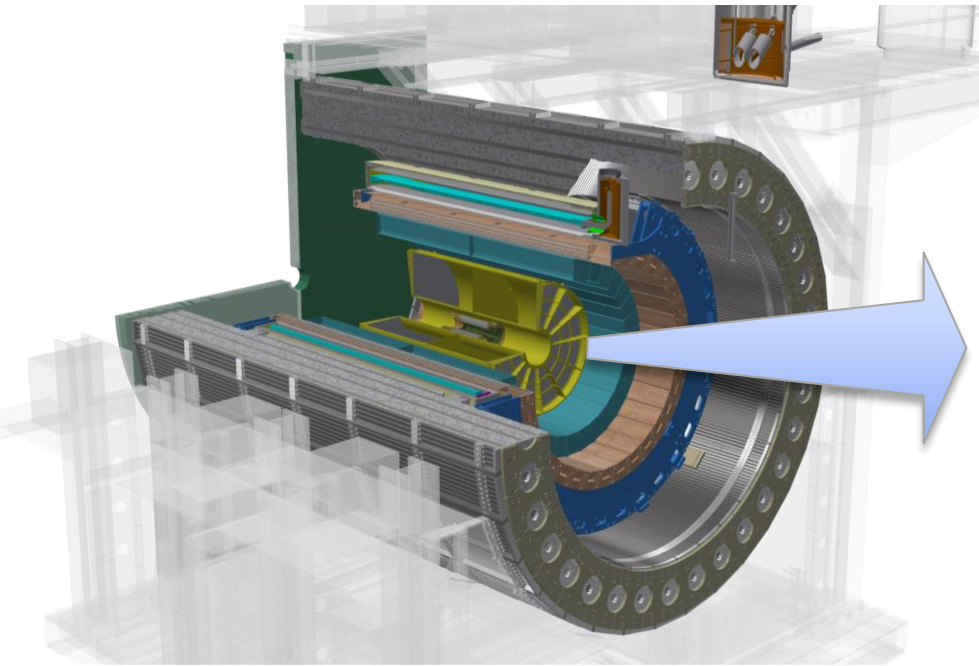


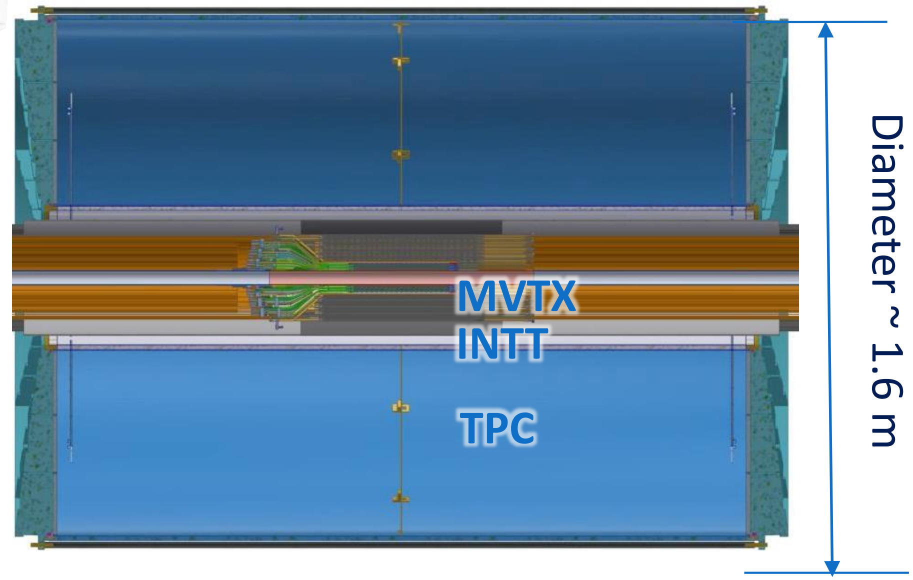
15 kHz trigger
>10 GB/s data

sPHENIX Detector

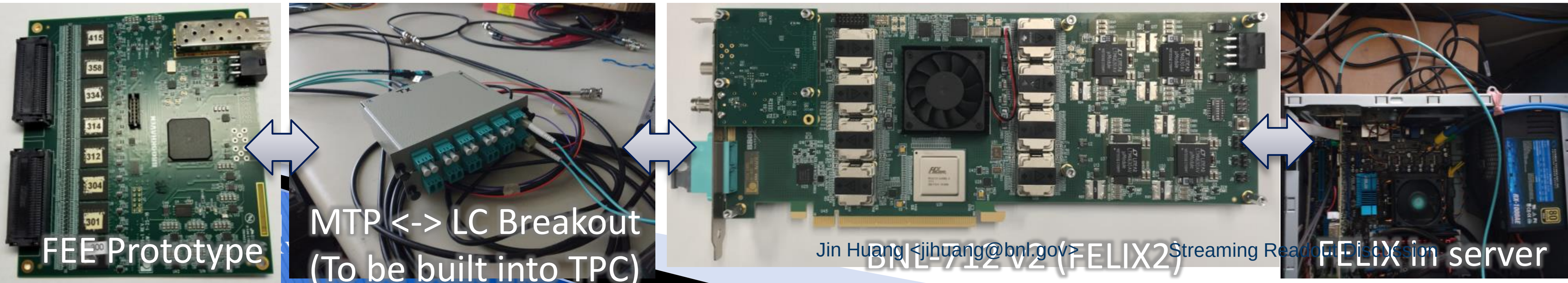
- 2016: Scientific review and DOE mission need Status (CD-0)
- 2018: Cost/schedule review and DOE approval for production start of long lead-time items (CD-1/3A)
- 2022: installation in RHIC 1008 Hall; 2023: First data
  - All tracker front end support streaming readout.
  - DAQ disk throughput for 9M particle/s + pile ups (> EIC ~4M particle/s)

sPHENIX Time projection chamber (TPC)



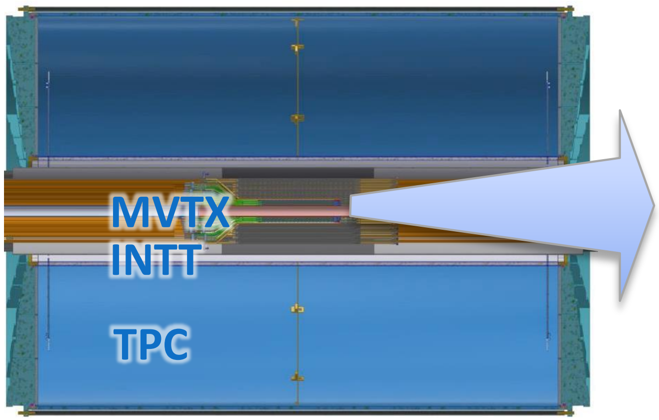


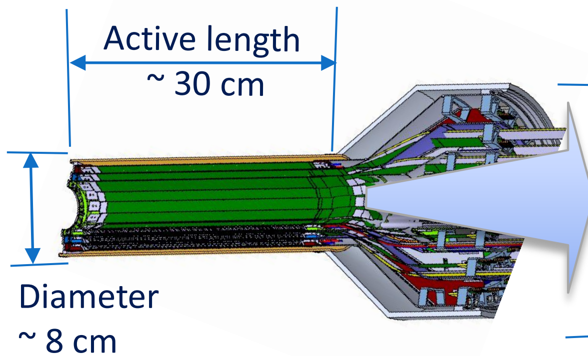
- Next-gen TPC w/ gateless and continuous readout: δp/p < 2% for p_T<10 GeV/c
- Ne-based gas for fast drift (13us). qGEM amplification and zigzag mini-pads.
- 160k channels 10b flash ADC @ 20MHz with SAMPA ASIC -> 2 Tbps stream rate.

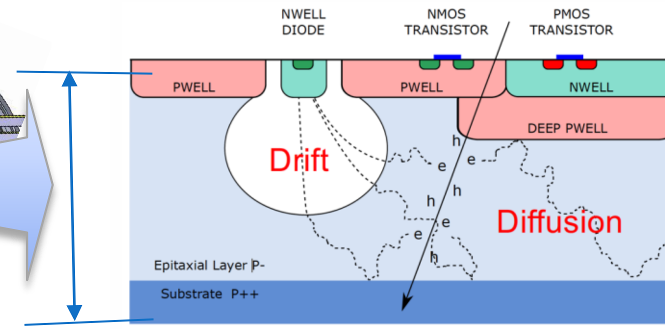


sPHENIX MVTX

- 200M pixel monolithic active pixel sensors (MAPS) vertex tracker (MVTX) → 5μm position resolution, 0.3% X0 / layer → <50 μ m DCA @ 1 GeV/c
- In close collaboration with ALICE & ATLAS phase-1 upgrades

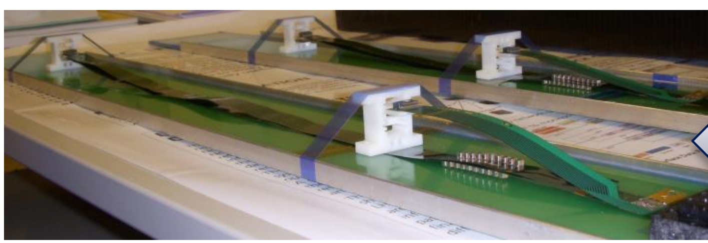


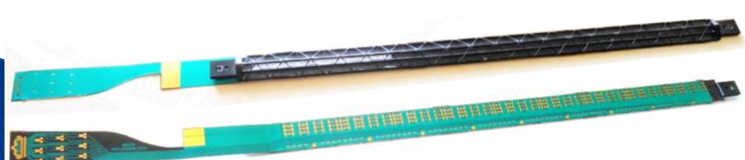




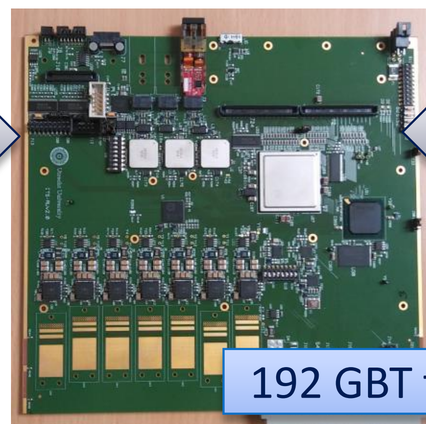
Thickness ~ 50 μm, 30x30 μm pixels

Sensor test with sPHENIX extension

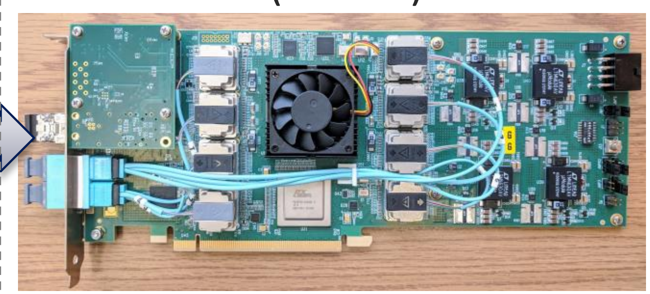




Readout Unit v2



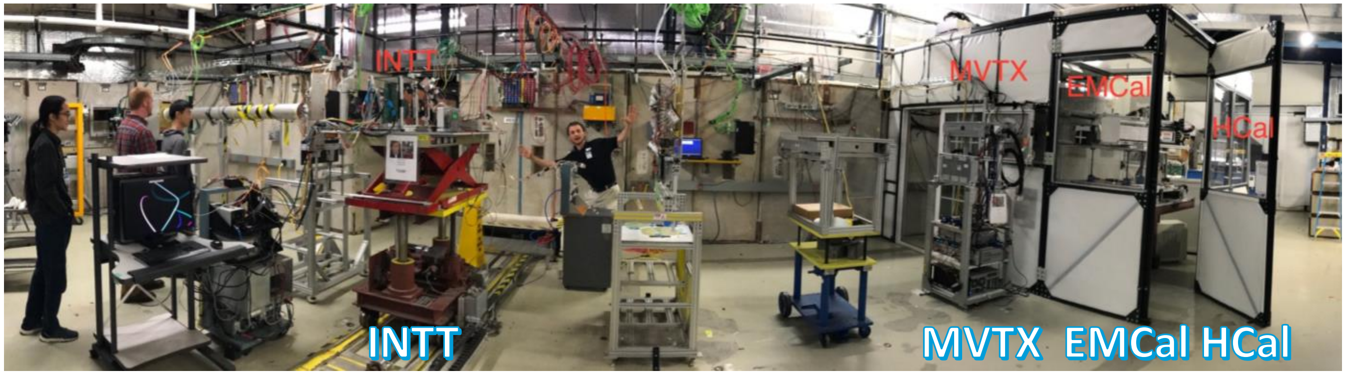
BNL-712 v2 (FELIX2)

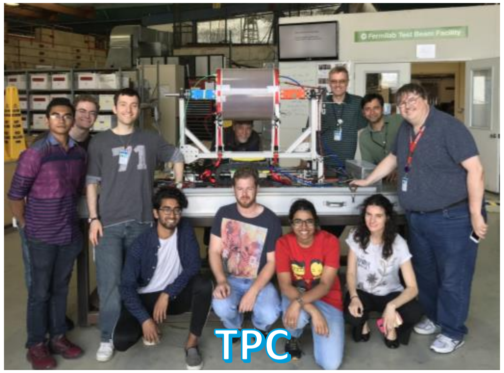


192 GBT fiber links

Exp. Hall DAQ room

Highlight of sPHENIX prototypes in action

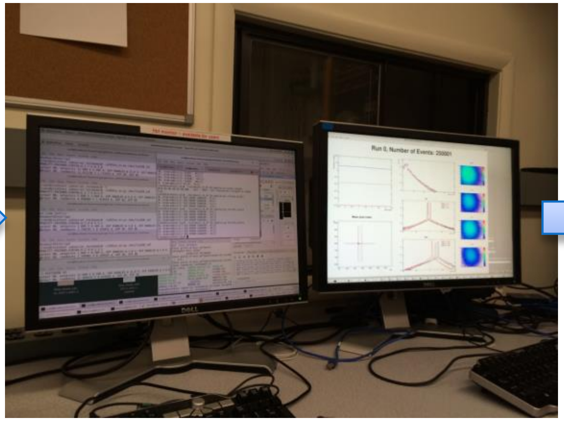




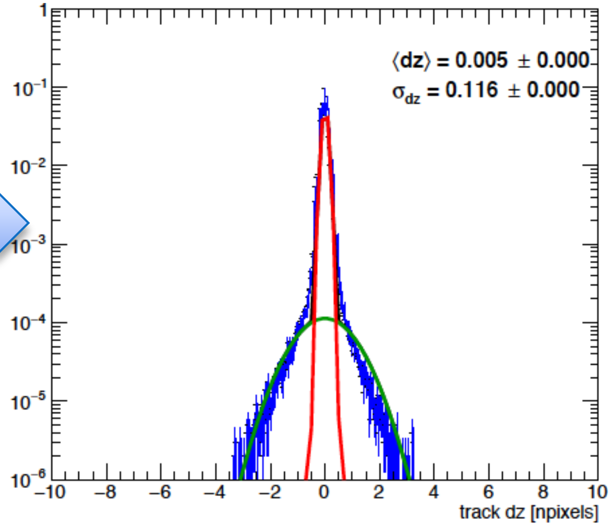
Feb-July 2018 FermiLab Test beam facility, test of each sPHENIX detector subsystem

4x MVTX sensor in beam

sPHENIX DAQ



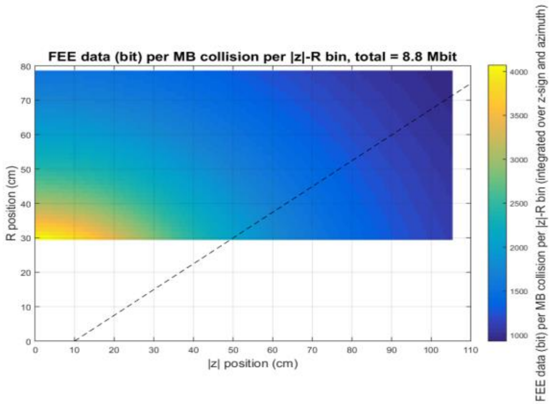
MVTX Hit Spatial Resolution: < 5 um



eRHIC and JLEIC key parameters at max Lumi points

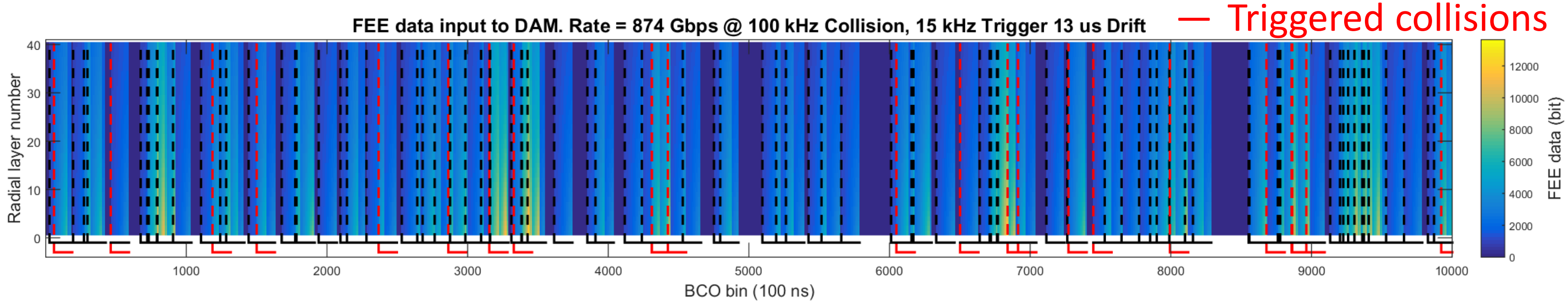
design parameter	eRHIC proton	eRHIC electron	JLEIC proton	JLEIC electron
center-of-mass energy [GeV]	105		44.7	
energy [GeV]	275	10	100	5
number of bunches	1320		3228	
particles per bunch [10^{10}]	6.0	15.1	0.98	3.7
beam current [A]	1.0	2.5	0.75	2.8
horizontal emittance [nm]	9.2	20.0	4.7	5.5
vertical emittance [nm]	1.3	1.0	0.94	1.1
β_x^* [cm]	90	42	6	5.1
β_y^* [cm]	4.0	5.0	1.2	1
tunes (Q_x, Q_y)	.315/.305	.08/.06	.081/.132	.53/.567
hor. beam-beam parameter	0.013	0.064	0.015	0.068
vert. beam-beam parameter	0.007	0.1	0.015	0.068
IBS growth time hor./long. [min]	126/120	n/a	0.7/2.3	n/a
synchrotron radiation power [MW]	n/a	9.2	n/a	2.7
bunch length [cm]	5	1.9	1	1
hourglass and crab reduction factor	0.87		0.87	
peak luminosity [10^{34} cm^{-2}s^{-1}]	1.05		2.1	
integrated luminosity/week [fb^{-1}]	4.51		9.0	

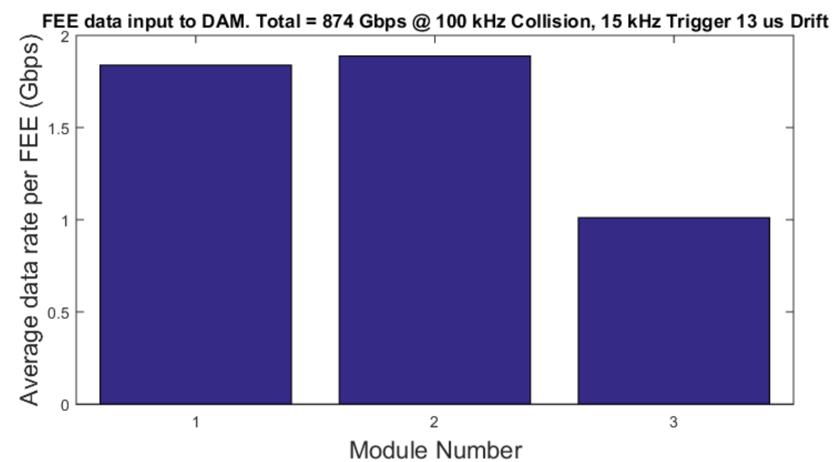
FEE data rate



100kHz collision in continuous DAQ trigger
In TPC DAQ simulation

— All collisions
— Triggered collisions





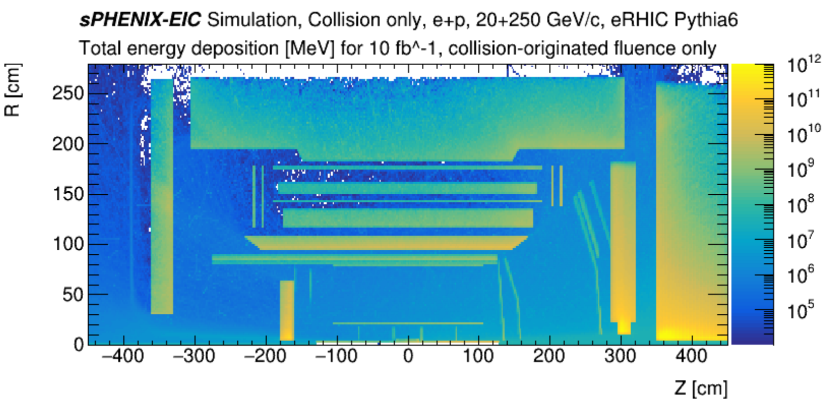
FEE -> DAM limit : 6 Gbps x 8b/10b per FEE

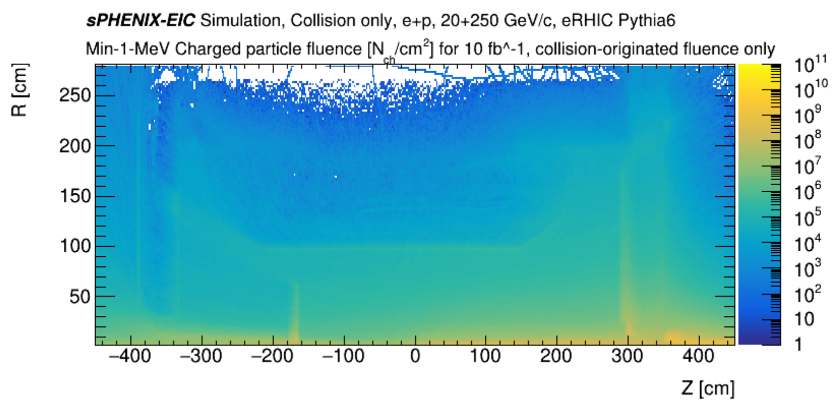
Reference design rate: 1.9 Gbps, far lower than limit

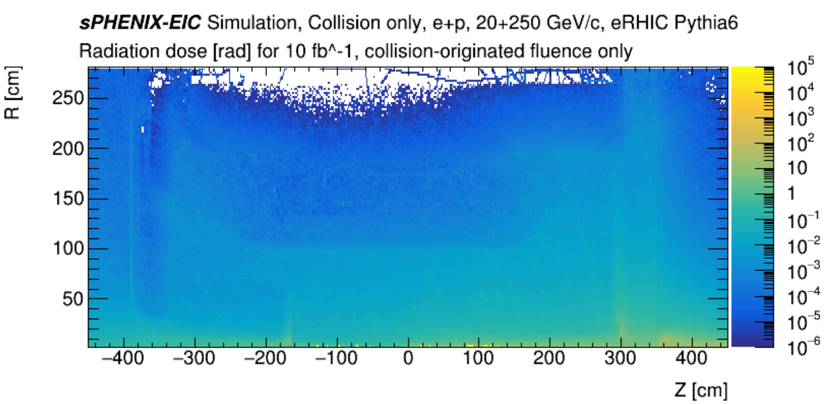
Max rate: 200kHz + 48 rings → max 7.2 Gbps @ module 1

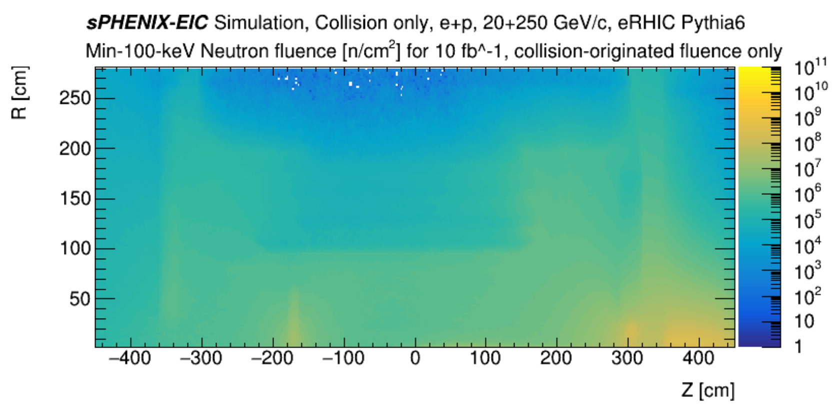
BROOKHAVEN NATIONAL LABORATORY

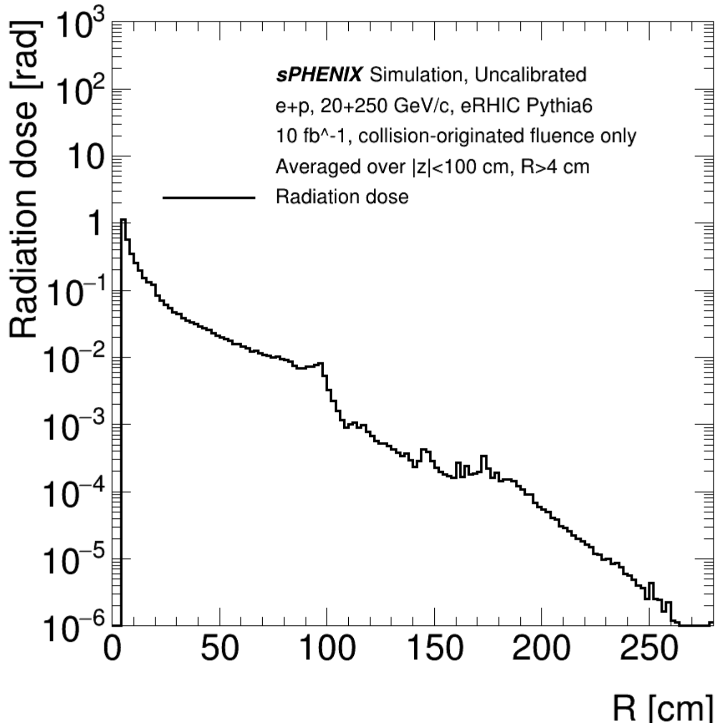
Radiation map

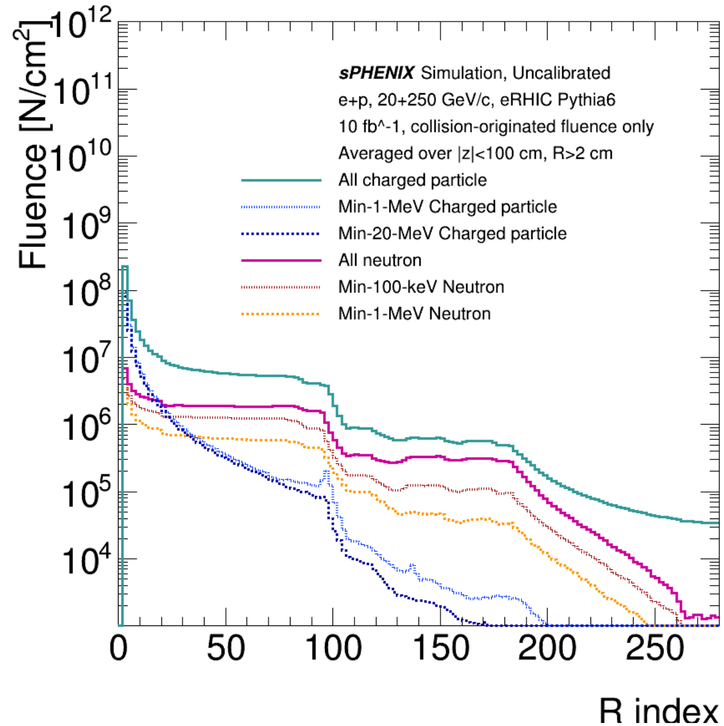


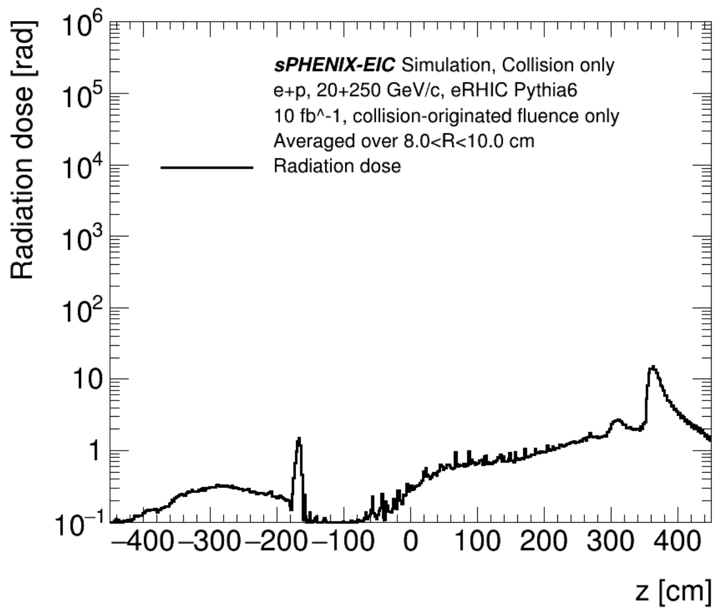


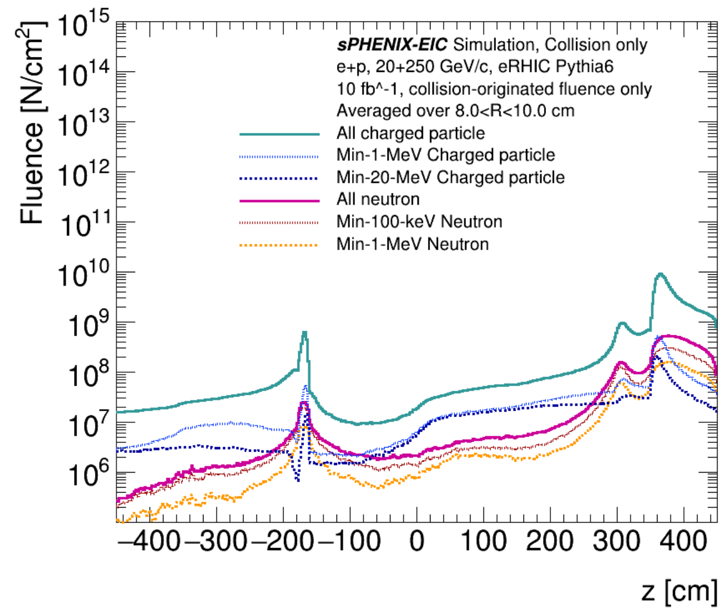





sPHENIX Simulation, Uncalibrated
e+p, 20+250 GeV/c, eRHIC Pythia6
10 fb^-1, collision-originated fluence only
Averaged over |z|<100 cm, R>4 cm
Radiation dose
Radiation dose [rad]
R [cm]


sPHENIX Simulation, Uncalibrated
e+p, 20+250 GeV/c, eRHIC Pythia6
10 fb^-1, collision-originated fluence only
Averaged over |z|<100 cm, R>2 cm
All charged particle
Min-1-MeV Charged particle
Min-20-MeV Charged particle
All neutron
Min-100-keV Neutron
Min-1-MeV Neutron
Fluence [N/cm^2]
R index


sPHENIX-EIC Simulation, Collision only
e+p, 20+250 GeV/c, eRHIC Pythia6
10 fb^-1, collision-originated fluence only
Averaged over 8.0<R<10.0 cm
Radiation dose
Radiation dose [rad]
z [cm]


sPHENIX-EIC Simulation, Collision only
e+p, 20+250 GeV/c, eRHIC Pythia6
10 fb^-1, collision-originated fluence only
Averaged over 8.0<R<10.0 cm
All charged particle
Min-1-MeV Charged particle
Min-20-MeV Charged particle
All neutron
Min-100-keV Neutron
Min-1-MeV Neutron
Fluence [N/cm^2]
z [cm]