Trigger and DAQ considerations for 10 TeV μ-collider

(With focus on tracker only)

Angira Rastogi[1]

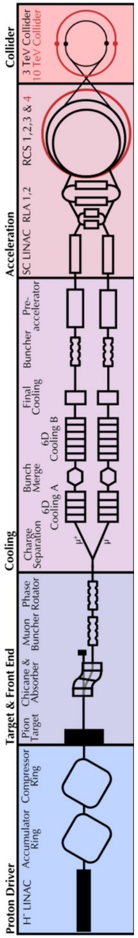
(Special thanks for inputs from Simone, Timon, Zach, Larry & Tova to “trigger” this discussion at IMCC 2025)

[1]Lawrence Berkeley National Laboratory

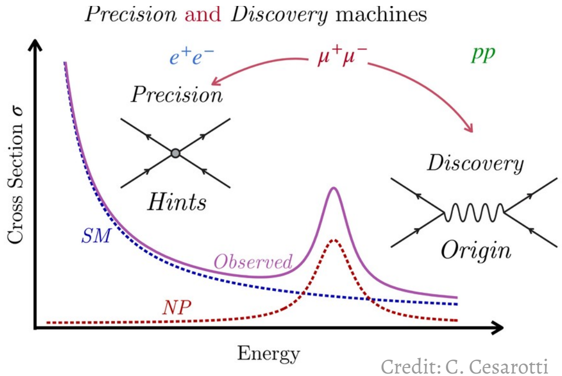
2nd Annual US Muon Collider Community Meeting

August 7-8, 2025

Introduction



- Muon collider is an attractive and exciting future collider option as:
  - Advantages of "pp" collider due to high mass or low synchrotron radiation, hence provides us <u>high energy reach</u> for "discovery".
  - Advantages of "e+e-" collider since muons are also fundamental particles, therefore <u>high precision measurements</u> for "hints" to new-physics.
  - <u>Compact machine</u> hence more "cost-effective".



Credit: C. Cesarotti

- Significant current challenges:
  - From accelerator side: proton driver beam, muon cooling before decay, high-gradient RF, magnets etc.
  - From detector side: handling extremely large muon decay backgrounds.
- Nevertheless, we need to start thinking in parallel about detector operations, most importantly - Trigger and DAQ bandwidth requirements.
  - Tracker only

Today!

What we know so far?

Parameter	LHC	HL-LHC	Muon collider
Center-of-mass energy	13/13.6 TeV	14 TeV	10 TeV
Peak inst. lumi	2 x 10^{34} $cm^{-2}s^{-1}$	7.5 x 10^{34} $cm^{-2}s^{-1}$	~ 20 x 10^{34} $cm^{-2}s^{-1}$
Intergated lumi/year	50-60 fb^{-1}/year	250 fb^{-1}/year	1 ab^{-1}/year
Target integrated lumi	300 fb^{-1} (by end of Run 3)	3000 fb^{-1} (over 10 years)	10 ab^{-1} (over 5 years)
Bunch crossing frequency	40 MHz	40 MHz	~ 30 kHz
Average PU	30-60	140-200	~0-1
L1 Trigger rate	100 KHz	750KHz – 1 MHz	**TBD**
L1 trigger latency	3-4 μs	10-12 μs	**TBD**
HLT event size	1-2 MB	3-7 MB	~ 80 MB (from ESPPU report)
HLT output rate	1 KHz	5-10 KHz	**TBD**
Recorded data volume/sec	~ upto 10 GB/s	~ up to 1 TB/s	2.4 TB/s (from BX rate)
Annual data volume	5-10 PB/year	120-160 PB/year	75 ZB/year
Total data	~100 PB	~1200-1600 PB	225 ZB (over 5 years)

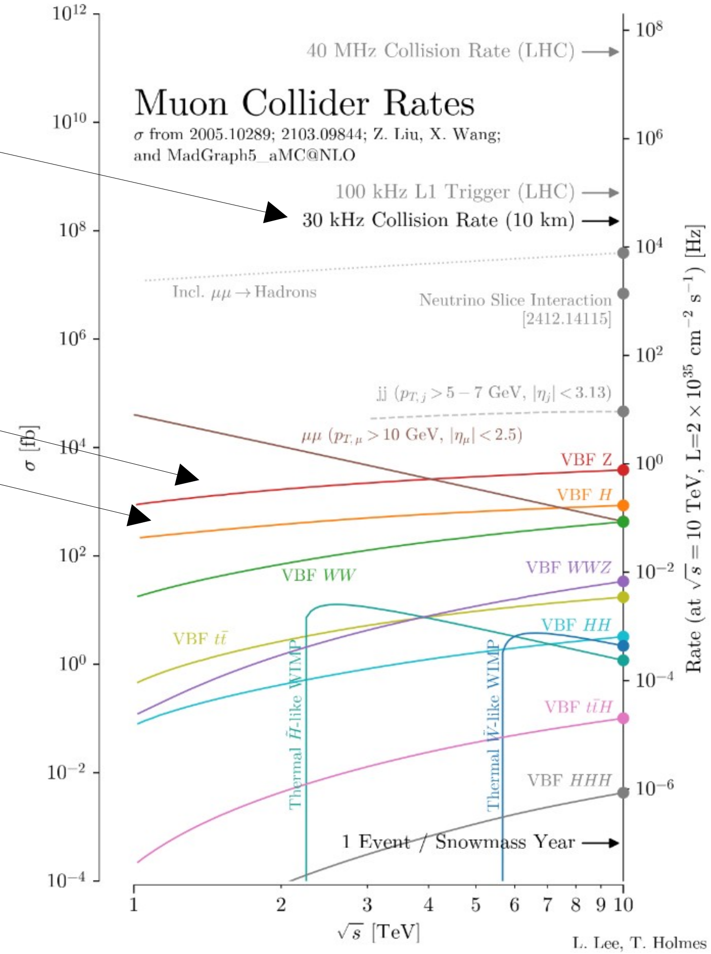
1 ZB = 1000 EB = 1000000 PB

What are the typical physics event rates at μ-collider?

As compared to beam-induced and other muon decay backgrounds in every bunch crossing

Interesting physics vs background

- **Bunch crossing (BX) rate ~ 30 kHz**
  - i.e. 30000 events/sec
- Physics processes at 10 TeV (in the order of decreasing rates):
  - **VBF Z** ~ 1 Hz i.e. 1 event/sec
  - **VBF H** ~ 0.1 Hz i.e. 1 event/ 10 secs
  - **VBF WW, μμ → μμ** ~ 0.08 Hz
  - **VBF WWZ, VBF tt** ~ 0.005 Hz
  - …
- Average number of interactions per BX (or effective pileup) = sum of rates of all processes/BX rate ~ 10^{-5} – 10^{-6} (ignoring inclusive jets production).
- That is, roughly 1-10 interesting physics events per million BXs!
- That is, most of the BXs are just … uneventful (ignoring ubiquitous muon decay backgrounds in every event).



Do we need triggers for such few physics events ?

Why we use triggers?

- Traditionally, ATLAS and CMS has been using two-tier trigger approach – hardware trigger (L1) using calo & muon detector information with low latency for faster decision (100 kHz), and software trigger (HLT) running full event reconstruction in the offline farms at lower rate (1 kHz).

- Benefits are two-fold:
  - To handle high rates of interesting processes by putting thresholds on objects and/or prescaling them.

    Highest rates at μ-collider is the collision frequency itself i.e. 30 kHz.

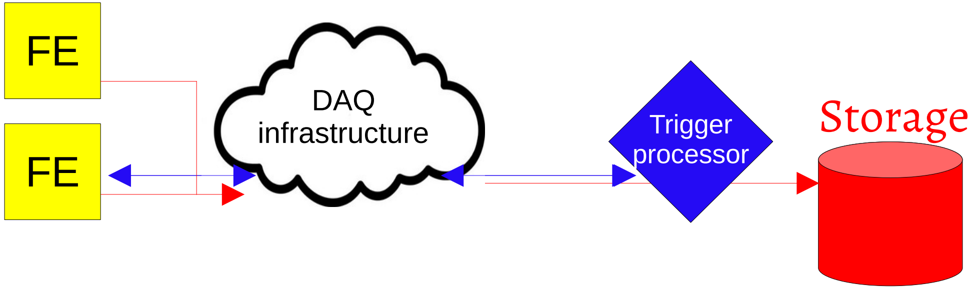
  - To discard uninteresting events to prevent bandwidth and storage resources.

    Indeed lots of bunch crossings with no interesting physics events (only muon-decay backgrounds) at μ-collider.

How much extra infrastructure (i.e. hardware, computing and cost) does it take to create trigger-based readout?

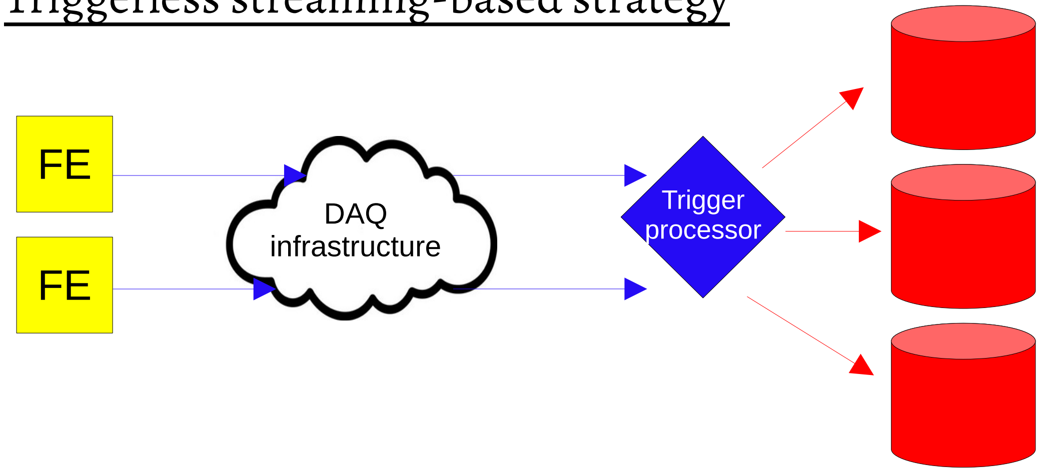
Can it be simply traded-off with triggerless streaming, by throwing away uninteresting events at the software-trigger level?

“Single-stage” trigger-based strategy



- Medium to high bandwidth, low-latency triggers, “hardware-” or “software-level”.
- Need sufficient buffers for trigger decision on the chip (FE).

Triggerless streaming-based strategy



- Very-high bandwidth for continuous readout from each detector link.
- No buffers (hit-memories) needed on chip.
- Can decide to split streams into different data-types based on certain “software-level” selections (more on this later).

Is triggerless streaming possible at μ-collider?

Depends on how fast we can process the data and move it out of FE, such that we keep up with the BX rate.

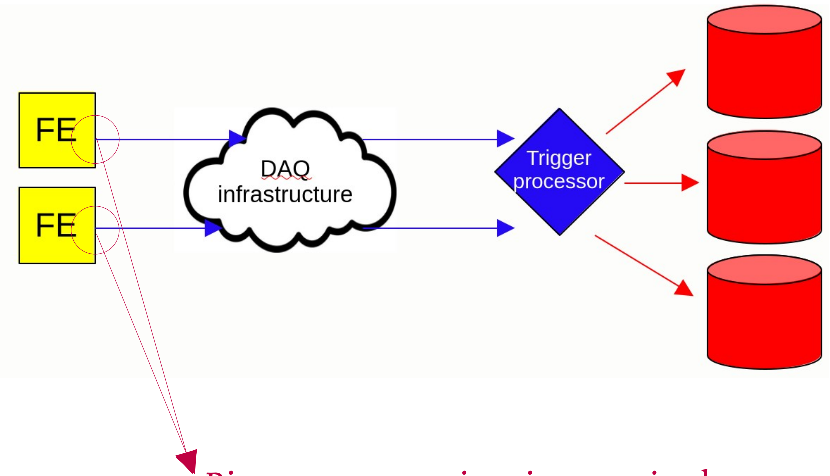
N.B. This is the baseline choice for all detector sub-systems as of now.

Triggerless streaming

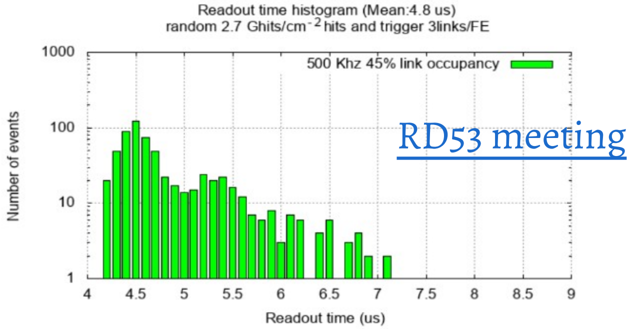
To answer this, we need some preliminary estimate of the readout data volume. To do that, we'll use:

- Detector hit-occupancy driven by beam-induced backgrounds (BIB) and incoherent-pair production (IPP), since signal is very small comparatively.
- Data volume calculated as: hit-size x hit-rate
  - Let's calculate this for various tracker subdetector layers.

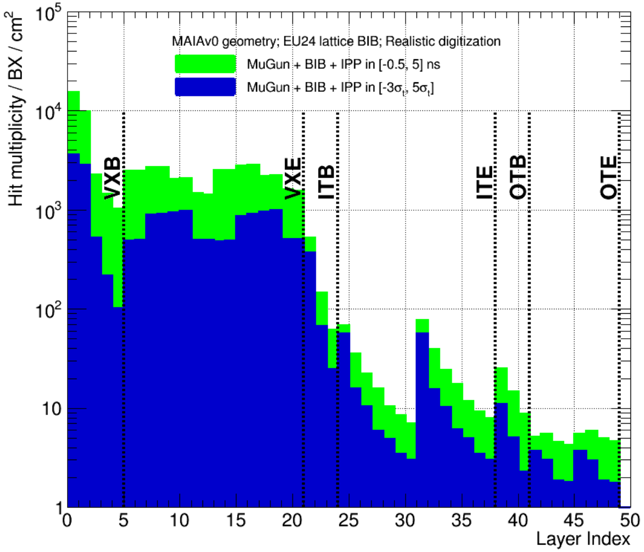
    Using hit occupancy calculated from realistic digitization implementation in mucol SW (link).
- How much time does it take for the chip to prepare this hit-data packet?
  - E.g. for 2.7 Ghit/cm² hit-rate in the innermost 25 µm x 25 µm pixels at 500 kHz rate (15% occupancy), RD53B chip can take up to 7 µs with 3 links @ 1.28 Gbps to fully readout.



Biggest stress points i.e. required readout bandwidth.



Readout data volume



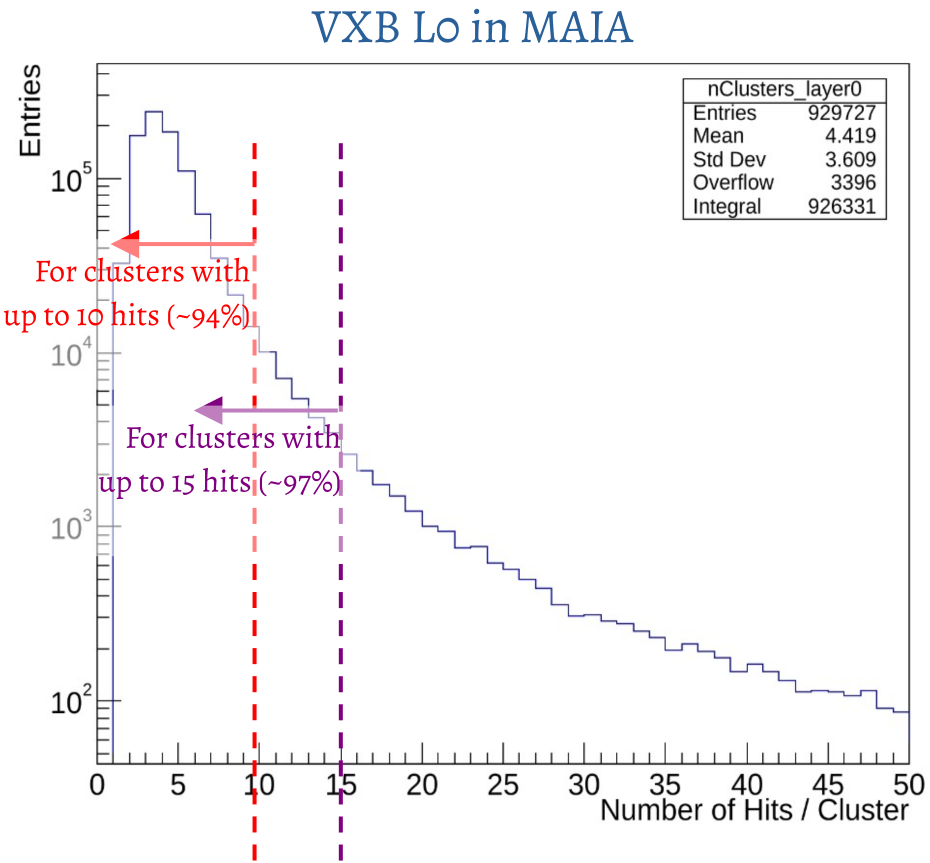
- Bunch crossing frequency, f_{BX} = 30 kHz
- Maximum hit data size ~ 40 bits = 5 Bytes
  (e.g. 8-bit pixel row within core, 8-bit pixel column within core, 10-bit BCID or timing info, 4-bit ToT, 8-bit pixel cluster pattern or hit-map, 2-bit core address)
- Let's take FE size of 2 cm x 2 cm (RD53B chip).
- Data rate / FE (Bytes per sec) = Occupancy x FE_area (cm^2) x f_{BX} x hit-size (Bytes)

Loose hit-time window Tight hit-time window

Sub-detector layer	Occupancy ($hits/BX/cm^2$) with [-0.5, 5] ns	Occupancy ($hits/BX/cm^2$) with [$-3\sigma_T$, $5\sigma_T$]	RAW data size (/FE/BX) (kB) with [-0.5, 5] ns	Data rate/FE (Gbps) with [-0.5, 5] ns	Data rate/FE (Gbps) with [$-3\sigma_T$, $5\sigma_T$]
VXB L0 (σ_T=30 ps)	15422 (40%)	3600 (10%)	300	73.6	16.8
VXE L5 (σ_T=30 ps)	5541	1979	110	26.4	9.6
ITB L0 (σ_T=60 ps)	528	373	10	2.4	1.6
ITE L0 (σ_T=60 ps)	145	114	2.9	0.7	0.5
OTB L0 (σ_T=60 ps)	25	11	0.5	0.1	0.05
OTE L0 (σ_T=60 ps)	10	7.4	0.2	0.05	0.03

Not realistic for detector to maintain DAQ synchronization at picosecond-level (i.e. with tight hit-time window).

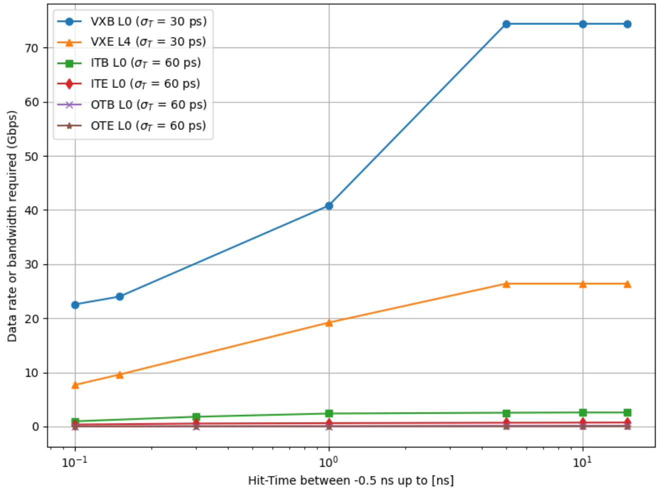
Let's focus on innermost layer…



- BIB has extremely long clusters currently (due to straight-line extrapolation ignoring multiple scattering in the digitization code).
- But, to a good approximation, we can estimate bandwidth for hit-rates coming from clusters with up to 10 hits and allow a margin for clusters with up to 15 hits.
- That is, for loose time-window selection (-0.5 to 5 ns):
  - Data rate/FE (Gbps) for clusters with up to 10 hits = 12278 hits/event/cm² x 4 cm² x 30 kHz x 40 bits
    ~ 60 Gbps
  - Data rate/FE (Gbps) for clusters with up to 15 hits = 13571 hits/event/cm² x 4 cm² x 30 kHz x 40 bits
    ~ 65 Gbps
    Ignoring the variation from num. of clusters since it is small (~930K ± 900).
  - Compared to 3.2 Gbps bandwidth with tight hit-time window for clusters with up to 15 hits!

Bandwidth requirements

- Some on-chip processing (e.g. some hit time-of-arrival or TOA selection) is necessary, especially for vertex detector layers for a manageable readout with future technology (20 years from now).
  - E.g. ATLAS Phase-2 readout board (FELIX) with 16-lane PCIe Gen 4 interface can offer readout with 24-channels @ 25 Gbps currently.



- Can think of progressive rejection on-chip (e.g. TOA → TOT → Cluster selection) since it gets easier to handle smaller and smaller data.
- Based on hit efficiency from on-chip processing vs timing and power constraints, can also think of a hybrid of on-chip vs off-detector implementation, after a sizeable reduction has been achieved.
  - E.g. cluster rejection on FPGA to further slim down hit-data from chip or even quick tracking at L0-trigger for event selection.

How can we bring down the hit-rates to reduce overall data volume?

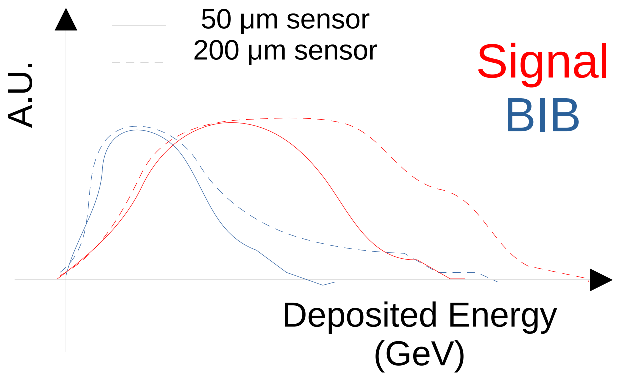
Pixel-based BIB suppression

- Two studies:
  - Varying sensor thickness for better signal vs BIB separation
  - Cluster shape analysis for cluster rejection

Sensor thickness studies

Juliet Wright, Laura Jeanty (U. Oregon)
S.P. Griso, A. Rastogi (LBL)

- Improve pixels cluster selection efficiency by enlarging the signal to beam induced background (BIB) separation.
- One way to do this is by increasing active sensor thickness, since e-h pair generation from high-energy signal particles increases linearly with sensor thickness while BIB loses energy after certain depth.
- To do these studies, we are testing five different thicknesses of silicon planar sensors: {50 μm (nominal), 75 μm, 100 μm, 200 μm, 400 μm} in VXB L0 in the MAIA detector geometry.
  - study alternate technologies (e.g. 3D sensors, LGADs) for better separation without compromising on timing resolution.
- Comparing single-muon particle gun (signal) with BIB from output of realistic digitization with increasing sensor thickness.



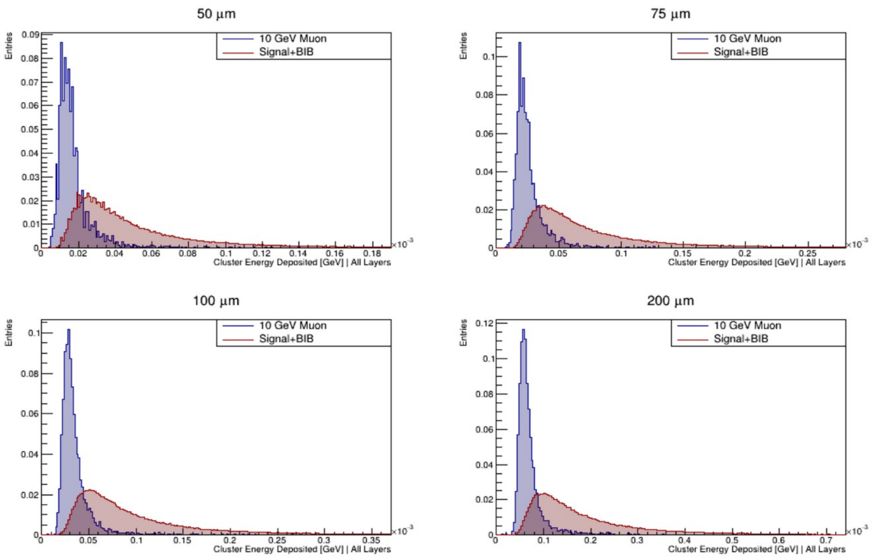
Example illustration

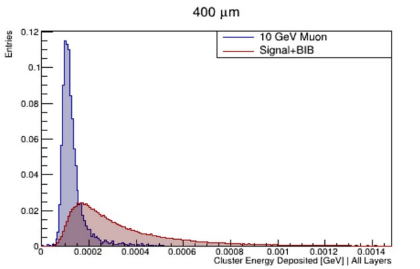
Preliminary results!

WORK IN PROGRESS

Juliet Wright, Laura Jeanty (U. Oregon)
S.P. Griso, A. Rastogi (LBL)

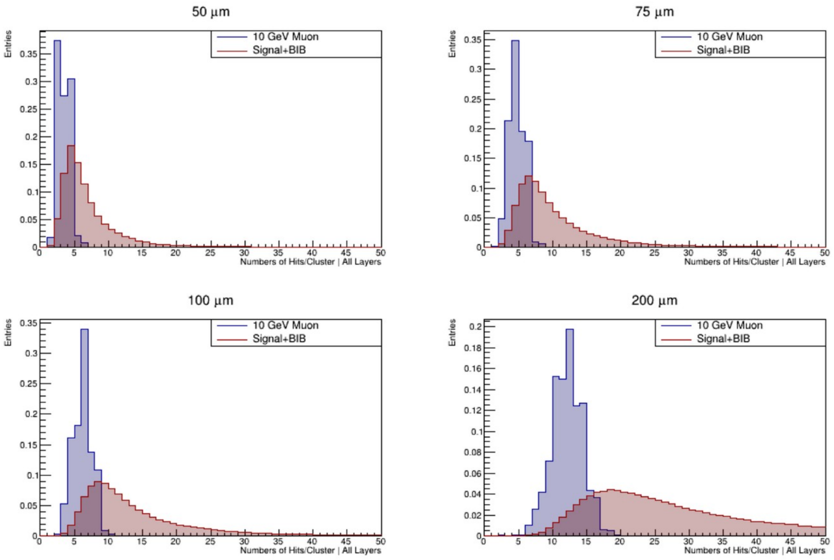
Cluster energy deposition

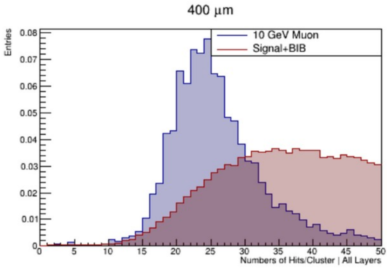




- Most probable energy loss in silicon, scaled to the mean loss of a minimum ionizing particle, is 388 eV/µm
- MPV≈ 0.95·388 = 370 eV/µm
- MPV can be scaled by:
  MPV(x) = 370 eV/µm · x [µm]
- The x-range is from 0-10 × MPV

Hits/cluster



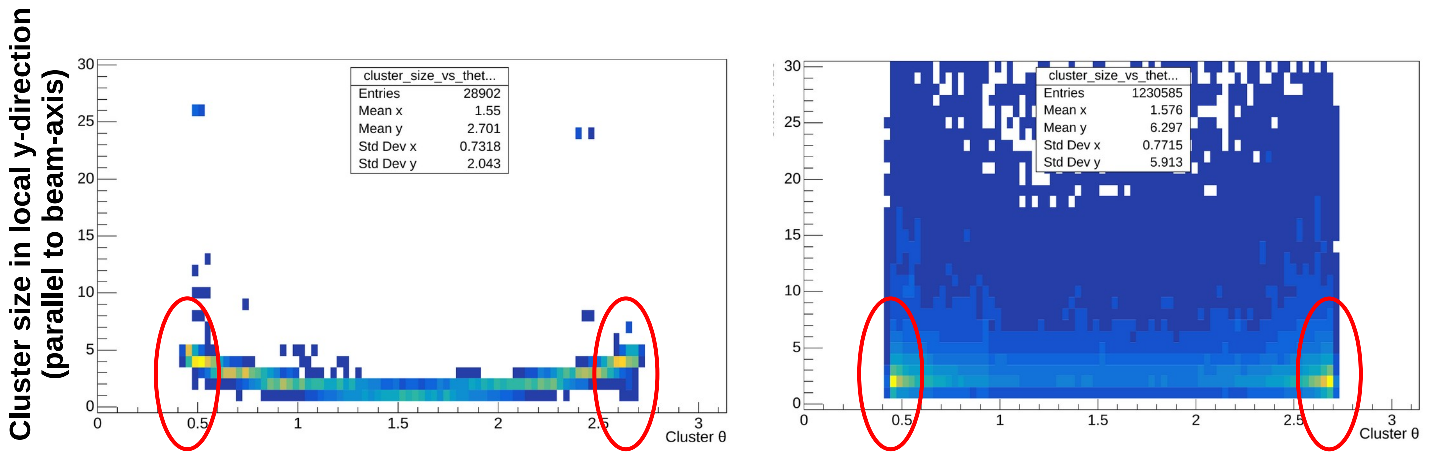


- Need to implement multiple scattering.

Cluster shape analysis

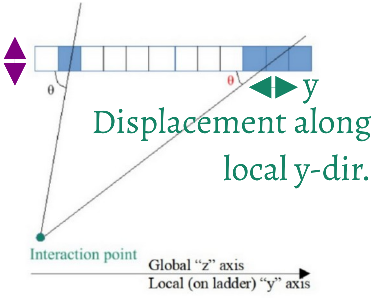
S.P. Griso, A. Rastogi (LBL)

Using correlation between incidence angle and number of pixel hits per cluster to reject long clusters – characteristic of BIB particles from muon decays.



μ-gun μ-gun + BIB

(e.g. MAIA_v0 geometry, VXB Layer 1, sensor thickness = 50μm, y-pitch = 25μm)



$\tan\theta = 50\mu m / y$

$\tan(0.5) \sim 0.5 = 50\mu m / y$

$y = 100\mu m$

$y/\text{pitch} \sim 4$ pixels

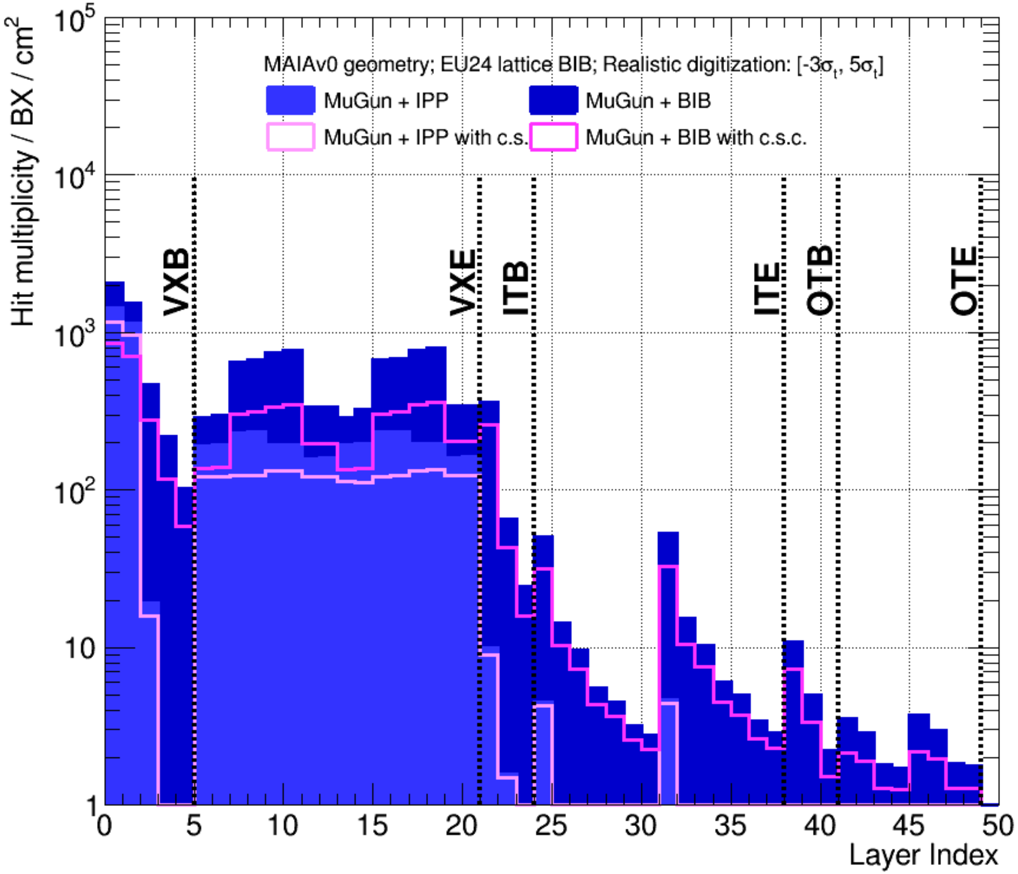
BIB particles either have very short clusters at same angles as signal (due to low-pT particles) or excessively long clusters (due to shallow incidence).

WORK IN PROGRESS

Preliminary results!

S.P. Griso, A. Rastogi (LBL)

IMCC talk

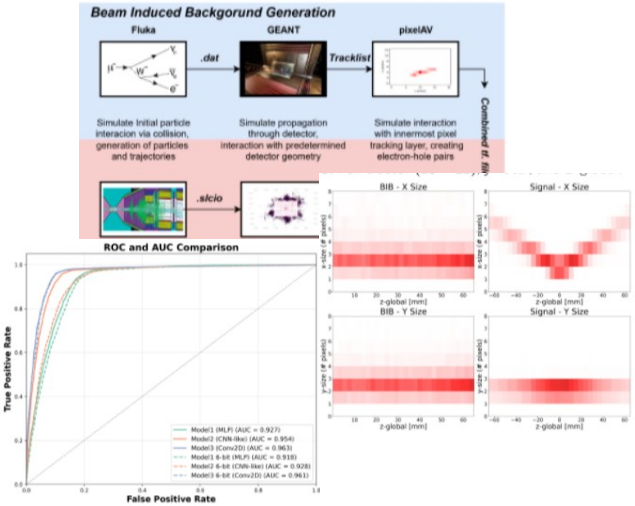


- From first principles, we can cut down up to 20-30% BIB clusters (with <5% loss for signal clusters) from each layer of subdetector!
- For ex, for signal+BIB clusters in VXB L0, occupancy down from 2000 to 850 hits/BX/cm^2 → 50% reduction in bandwidth!

- Starting to explore MVA-based methods for better background rejection over signal.

Other blue sky R&D

- SmartPixels: reducing silicon data via in-pixel intelligence such as discarding low pT tracks (<2 GeV) or feature extraction from charge (angle, position). Can save bandwidths up to 50-75%!
  - More in Abhijith's talk tomorrow



(Snippets from Eric's poster today)

- Embedded FGPAs for reconfigurable digital logic on 28nm CMOS ASIC, can be applied to variety of subsytems (ESPPU #95).
- ASIC compression with autoencoders (ESPPU #11).

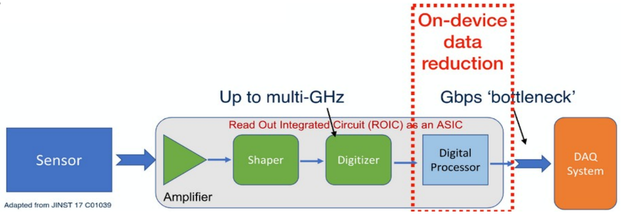


Image Credit: T. Arrestead

Other blue sky R&D

- Wireless data transmission (DRD-7.1c): Sending single signal to several receivers saves cabling, cost reduction, simplified installation/repair, reduction in dead material especially important for future tracking detectors. Few Gbps possible with 802.11 ac/ad WiFi!

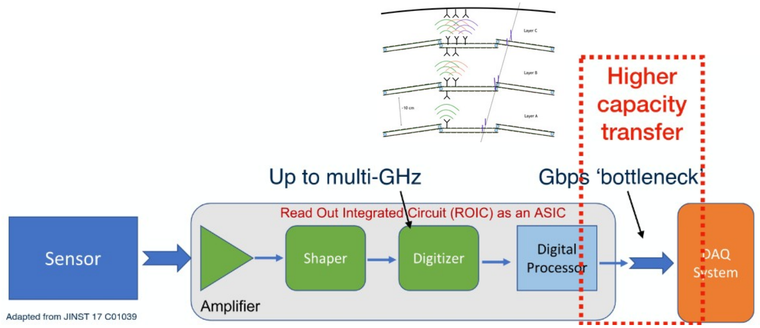


Image Credit: T. Arrestead

Some other open ideas/points to ponder on:

- Adding a loose off-detector "L0-trigger" e.g. using wireless network to a system next to the tracker to do a quick fitting and send trigger accept.

    - How does this lower rate readout with trigger affect the overall power budget?

What about offline computing & storage requirements?

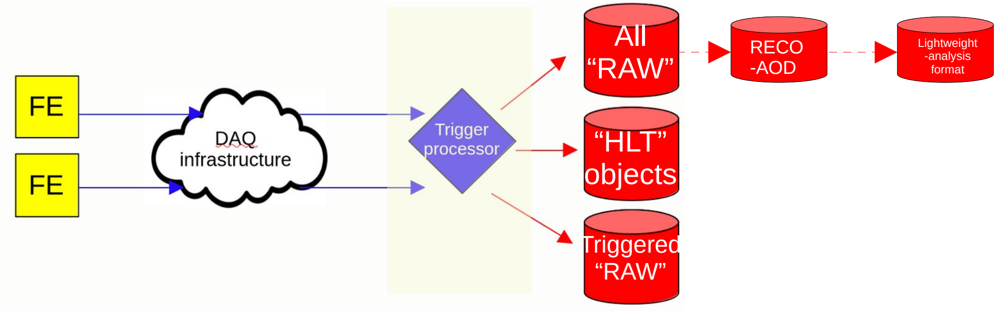
More on this tomorrow from K. Pedro

Computing and storage

From ESPPU report

	HEPscore[1] s / event	events	total CPU [kHEPscore[1]-years]
Generation[2]	640		$2.03 \cdot 10^3$
Simulation[3]	550	10^{11}	$1.74 \cdot 10^3$
Reconstruction (sim)[4]	1130		$3.58 \cdot 10^3$
Reconstruction (data)[4 5]		$5 \cdot 10^{12}$	$1.79 \cdot 10^5$
Re-processing (sim)	1680	$2 \cdot 10^{11}$	$1.07 \cdot 10^4$
Re-processing (data)	1130	$2 \cdot 5 \cdot 10^9$	$7.17 \cdot 10^2$
Total			$\mathbf{1.98 \cdot 10^5}$

Projections based on ATLAS & CMS for HL-LHC, since experimental conditions at PU=200 come close to presence of muon BIB ~O(1M) hits in tracker.



Software-based event selection using reconstruction algorithms, ML and parallelized filtering (FPGAs, GPUs)

From ESPPU report

	size [MB] / event	events	total size [PB]
RAW[6]	80		400
RECO / AOD[7 8]	20	$5 \cdot 10^9$	100
analysis[9]	0.005		0.03
SIM[10]	250		25000
SIM RECO[11]	40	10^{11}	4000
SIM analysis[11]	0.01		1.0
Total			**29501**

Tier-type	Data-type	Storage for…
"All RAW" data	RAW	All RAW data from detector on tape. Further used for processing RECO-AOD data offline, by applying object-level thresholds for high-level physics analysis. For MC, only RECO-AOD.
"Trigger-level Analysis or TLA"	HLT objects	All events but with limited information, especially for high-statistics lower thresholds analysis.
"Triggered RAW" data	RAW	Randomly chosen for min-bias studies

Conclusion

- Exciting challenges ahead to make sure TDAQ is not the precision bottleneck.
- Rates of interesting physics events are below O(1 Hz) for instantaneous luminosity of 2e35 $cm^{-2}s^{-1}$ at 10 TeV, with average pileup interactions per BX of 10^{-5}-10^{-6}.
- Triggerless streaming can be a workable overall choice for tracker detector, if we can do some on-chip processing for hit time-of-arrival and cluster sorting without significantly increasing the timing and power budget.
- Need to systematically study the efficiency of progressive reduction of data volume, from on- vs off-detector electronics to decide how much can we afford on the front-end.
  - Can also study if we need a quick "L0-trigger" accept for event based on tracking.
- AI/ML tools should be employed wherever necessary to meeting these challenges e.g. in cluster sorting.
- Simultaneously, we must develop and maintain versatile heterogeneous frameworks and platforms for faster and higher-capacity readout, including new technologies such as wireless network to reduce material budget.
- Finally, for data storage, we can have proper "software-level" trigger selection and dedicated data-streams for different applications.