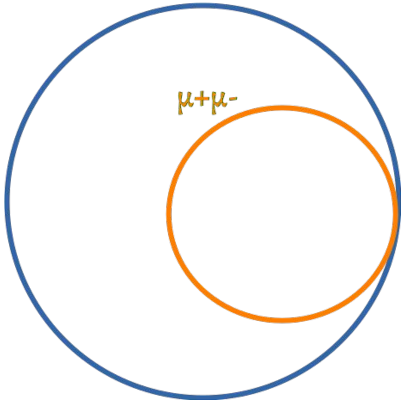
Lecture (Part I) + Hands-on Exercise (Part II)
on



The Muon Collider “problem”

Samuel Ferraro[1], Angira Rastogi[2]

[1]Harvard University
[2]Lawrence Berkeley National Laboratory

June 4th, 2026

Part I

Introduction to μ-Col
Physics benchmarks
Key design parameters
Challenges

Why a μ-Col?

A paradigm-shifting machine: first collider to combine the energy reach of a hadron collider with the precision of a lepton collider.

200x

Heavier than e-

$\sim 10^9\times$ less synchrotron energy than electrons, enabling TeV-scale circular lepton collisions

Point-like

Fundamental lepton

No quark substructure. Collisions have a well-defined initial state, no PDF uncertainties

Compact

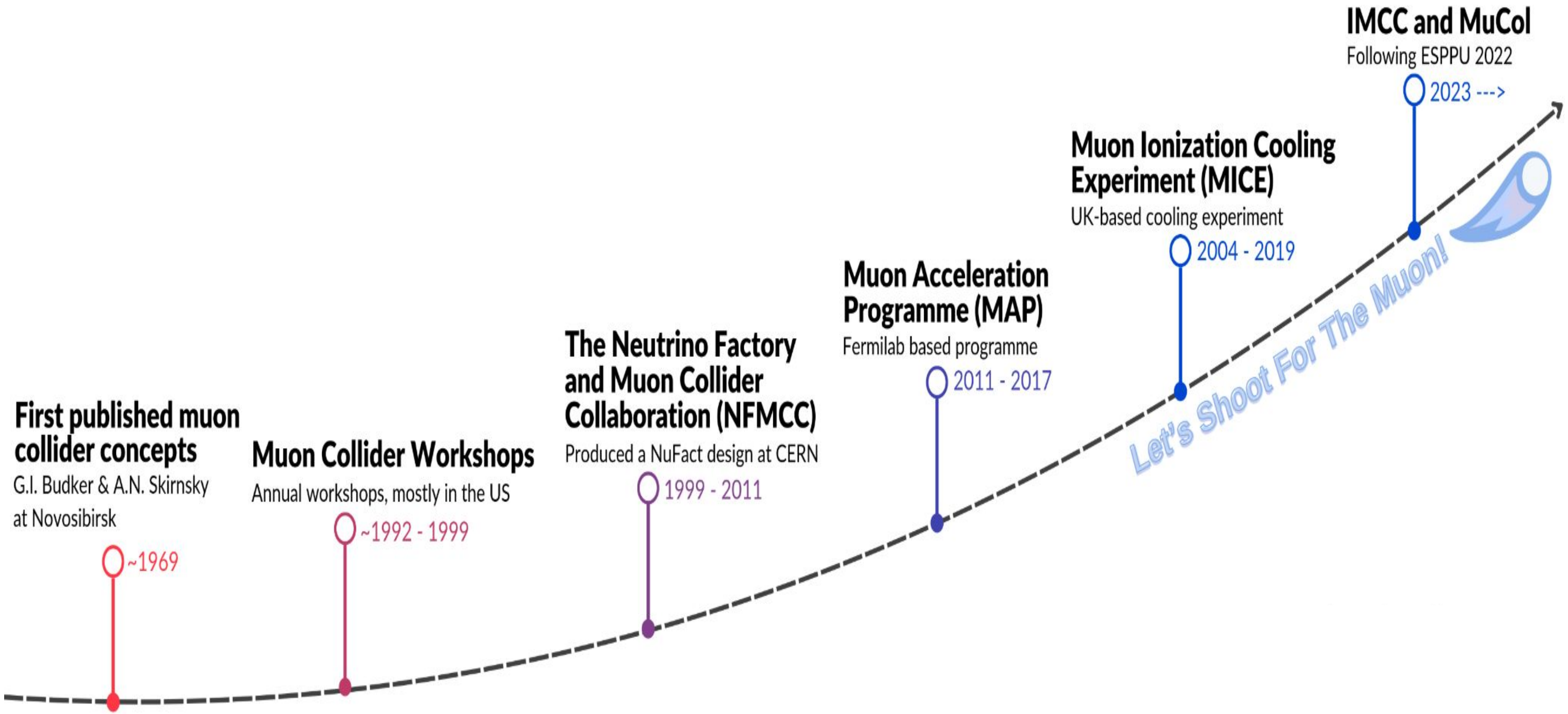
10 km ring @ 10 TeV

Equivalent physics reach to a 100 km pp collider, cost- and energy-efficient

Time dilation

$\gamma \sim O(10^4)$ @ $v=c$

Lifetime long enough to undergo several rotations and collide



First published muon collider concepts
G.I. Budker & A.N. Skirnsky
at Novosibirsk
~1969

Muon Collider Workshops
Annual workshops, mostly in the US
~1992 - 1999

The Neutrino Factory and Muon Collider Collaboration (NFMCC)
Produced a NuFact design at CERN
1999 - 2011

Muon Acceleration Programme (MAP)
Fermilab based programme
2011 - 2017

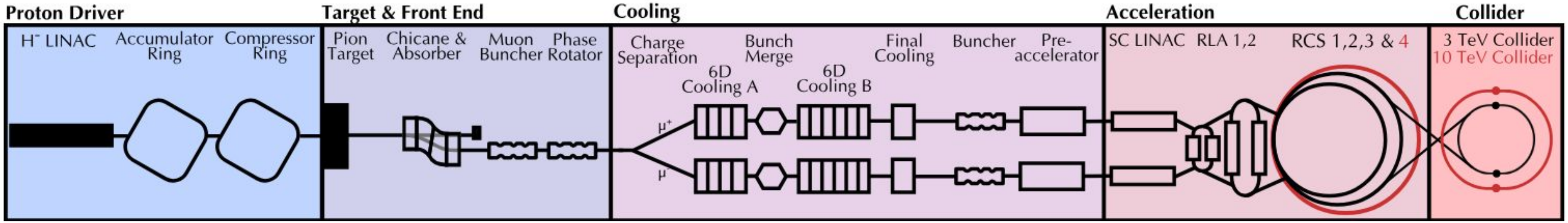
Muon Ionization Cooling Experiment (MICE)
UK-based cooling experiment
2004 - 2019

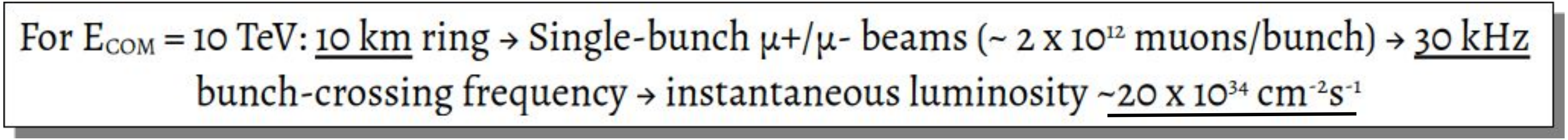
IMCC and MuCol
Following ESPPU 2022
2023 --->

Let's Shoot For The Muon!

EUS26

Conceptual layout of μ-Col

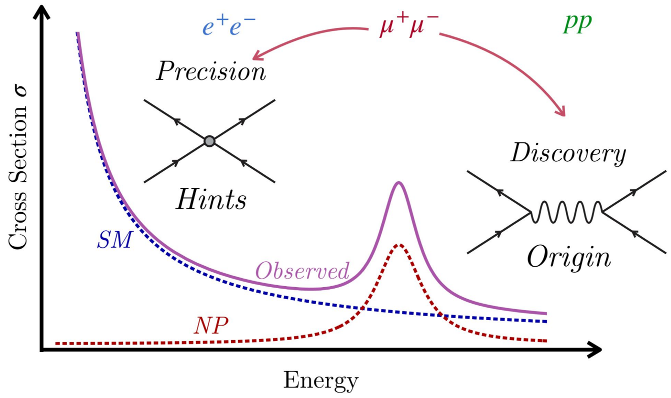



For E_{COM} = 10 TeV: 10 km ring → Single-bunch μ+/μ- beams (~ 2 x 10^{12} muons/bunch) → 30 kHz bunch-crossing frequency → instantaneous luminosity ~20 x 10^{34} cm^{-2}s^{-1}

10x inst. Lumi. as LHC
~3x inst. Lumi. as HL-LHC
In total 5 years of operation, projected to collect 10 ab^{-1}

Physics potential

The ability to do both is what makes muon colliders the most scientifically interesting future collider option.

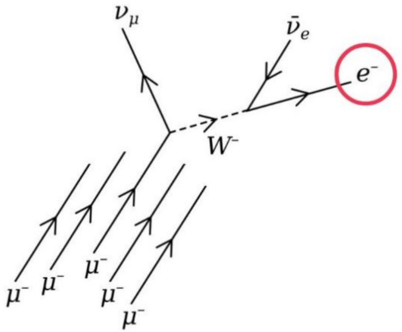


- Higgs s-channel production, direct access to higgs total width @ MeV precision.
- Weak boson fusion (WBF) dominant process, study electroweak precision at sub-permille level
- Top yukawa coupling to <1% via tH production
- HH production via VBF ~ 10x more events than HL-LHC
- Higgs trilinear coupling λHHH at 5% precision
- Quartic coupling (H^4) accessible for the first time at 10 TeV
- New heavy gauge boson (or SUSY particles pair) production up to 10 (5) TeV in mass
- Couplings to muon can probe unique leptophilic scenarios
- …

Also possible to carry out auxiliary experiments with highly collinear and energetic neutrino beam from muons decay - complimentary to future long baseline experiments.

Beam-Induced Backgrounds (BIB)

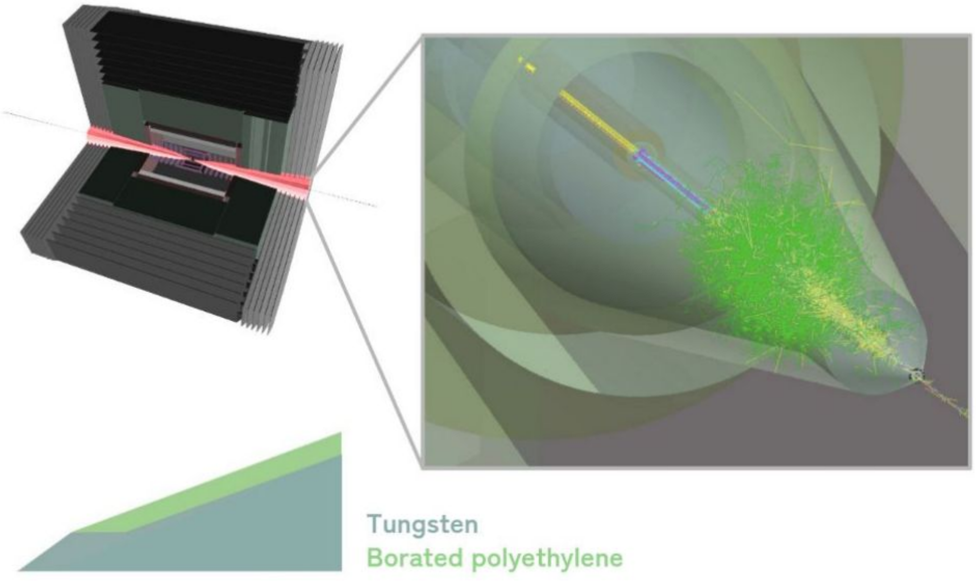
Multi-TeV muon decays produce TeV-scale electrons!
(Going straight into the detector volume if not stopped)
In addition, other charged & neutral secondary particles also produced from interaction with the surrounding material, all traveling towards the interaction point.



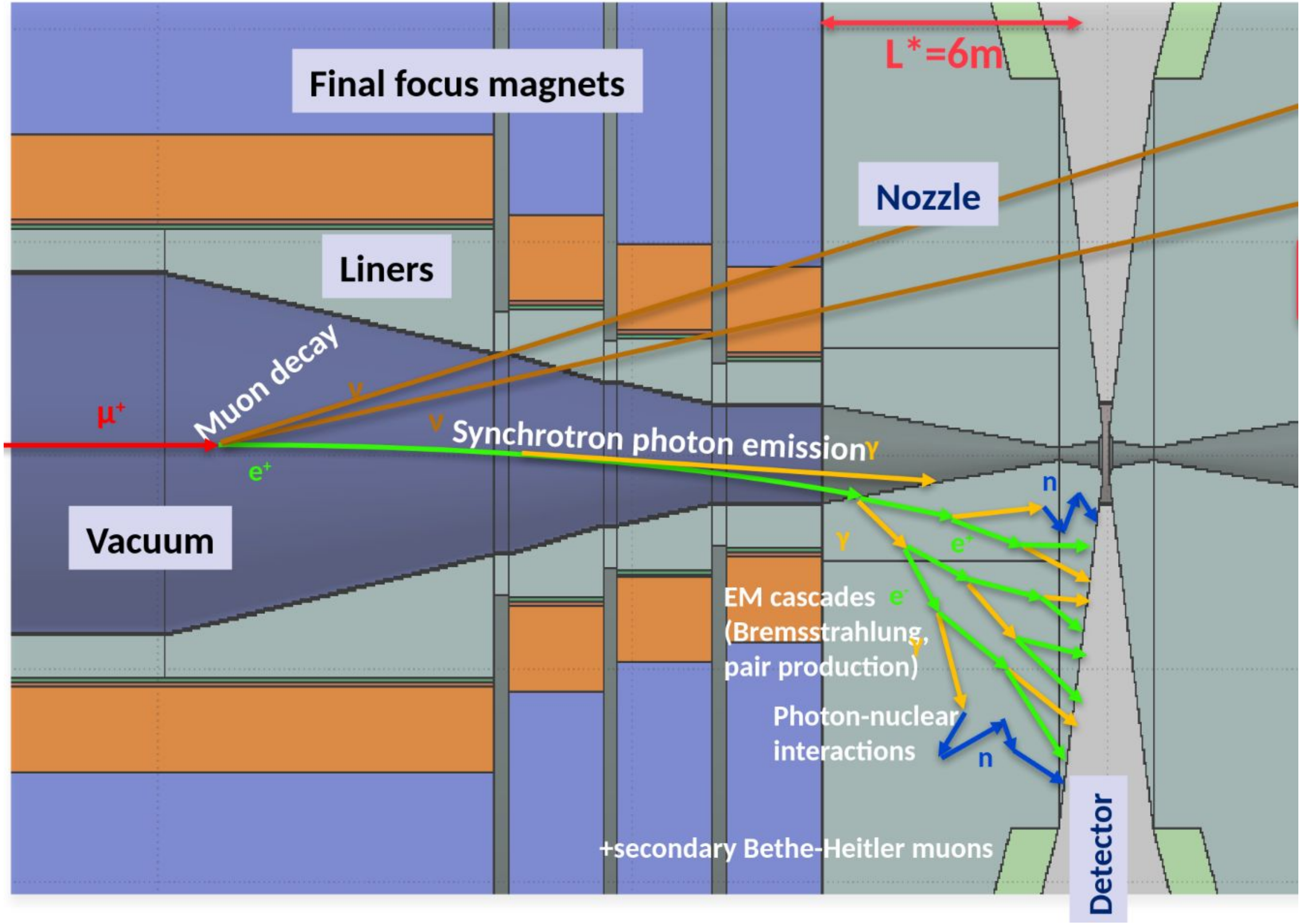
This is bad because:

- Damages the detector components over time due to high radiation
- Early onset of detector electronics, even before the muon collisions - resulting in loss of signal
- Pollutes collision environment - harder to cleanly reconstruct and identify the physics objects

Shielding from nozzles

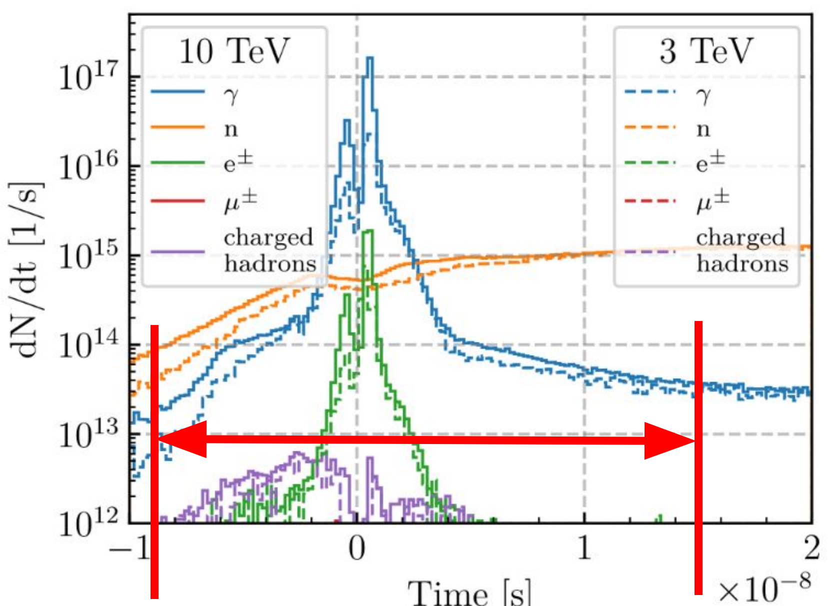


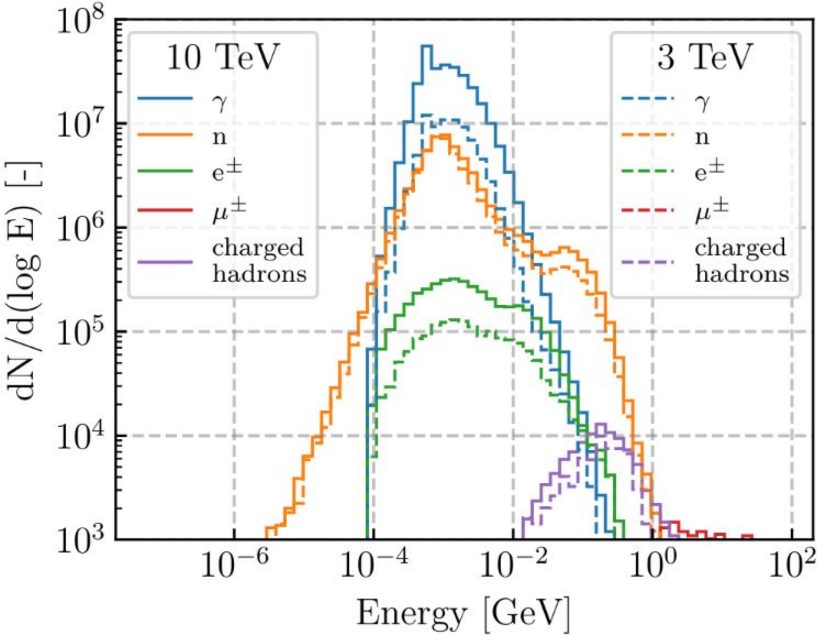
- Integrated in the MDI design, very close up to the IR.
- Reduces BIB by several orders of magnitude.
- Changes BIB composition: highly energetic particles diffuse into soft secondaries.
- In addition, collider lattice design (magnets placement for final focusing) also updated to reduce BIB, e.g. using shorter straight section.


L*=6m
Final focus magnets
Nozzle
Liners
Muon decay
ν
ν
μ^+
Synchrotron photon emission
γ
e^+
Vacuum
n
γ
e^+
EM cascades
e^-
(Bremsstrahlung,
pair production)
γ
Photon-nuclear
interactions
n
+secondary Bethe-Heitler muons
Detector

BIB characteristics

Long out-of-time tail of soft particles without much dependence on collider energy.

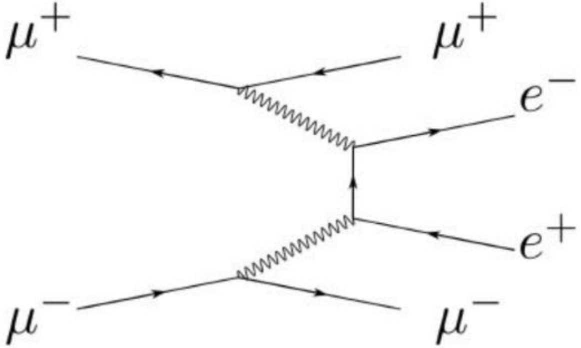


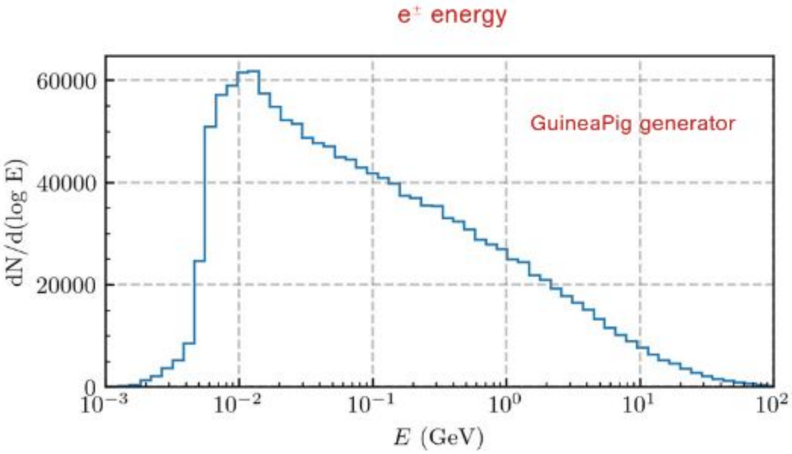


O(10^8) >100 KeV photons, O(10^7) >10^{-5} eV neutrons and O(10^6) >100 KeV e+/e-

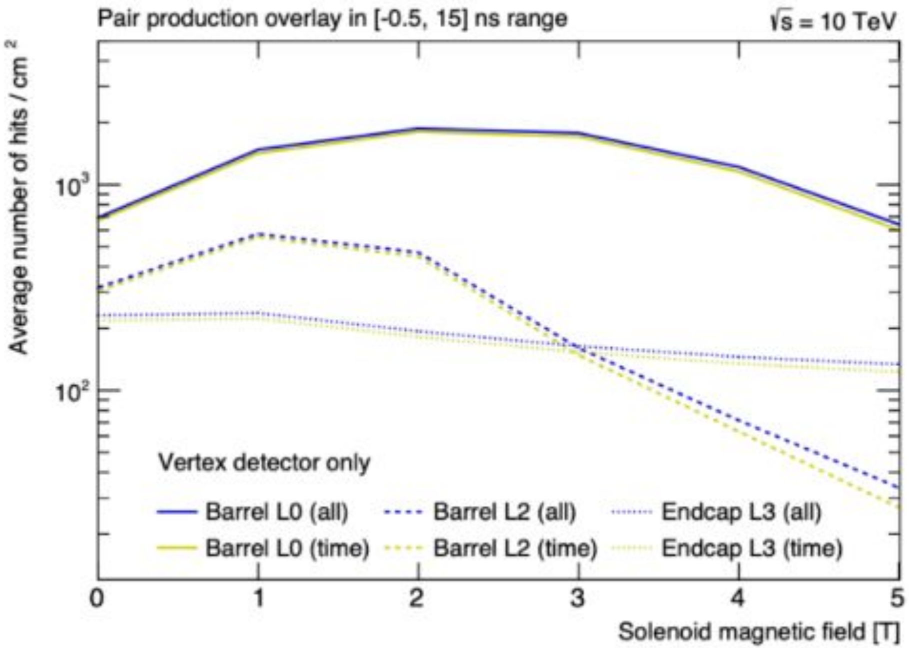
Incoherent Pair Production (IPP)

- Low energy e+e- pair production from real or virtual photons emitted by muons in the counter-rotating bunches.
- Produced at the interaction point, hence higher time-coincidence with signal.
- Higher energy tail as compared to BIB.





IPP mitigation

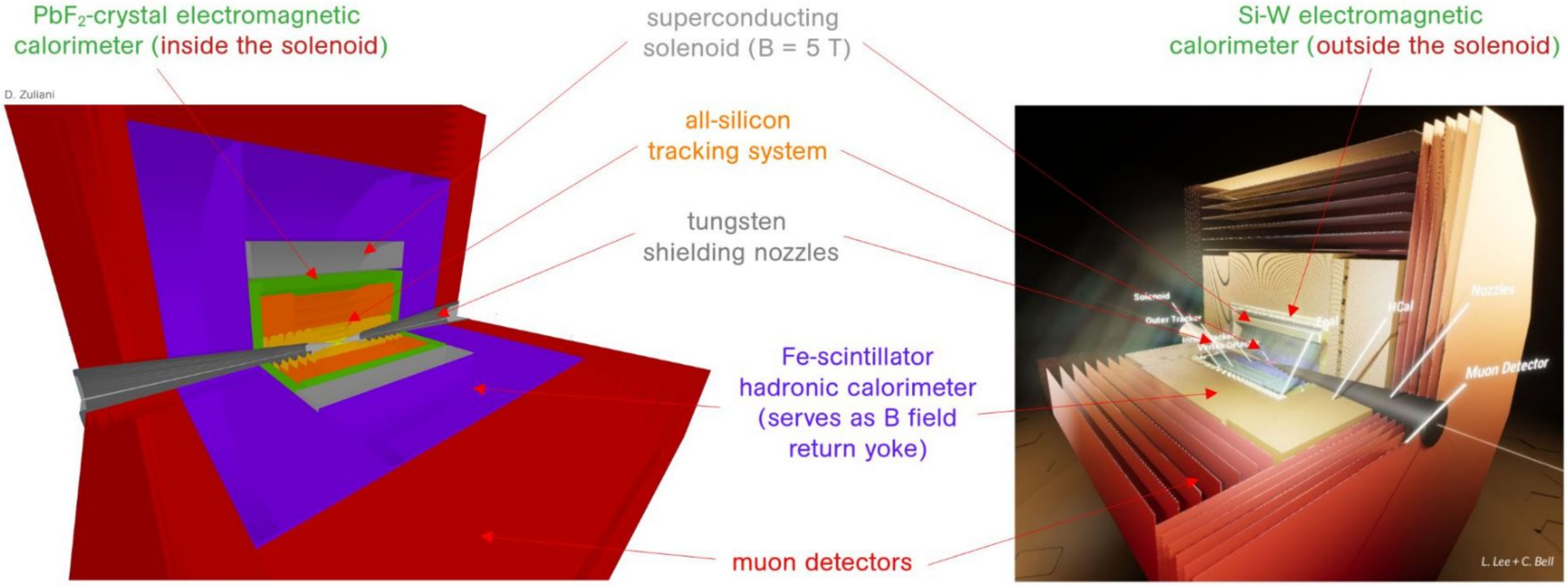


Increasing the magnetic field (e.g. 5T) in the tracking volume will sweep the e+e- from IPP out of the detector, to a large extent.

Only small flux of particles will affect the physics of the muon collision event.

(Left out of today's exercise)

10 TeV detectors design



PbF$_2$-crystal electromagnetic calorimeter (inside the solenoid)

superconducting solenoid (B = 5 T)

Si-W electromagnetic calorimeter (outside the solenoid)

all-silicon tracking system

tungsten shielding nozzles

Fe-scintillator hadronic calorimeter (serves as B field return yoke)

muon detectors

D. Zuliani

L. Lee + C. Bell

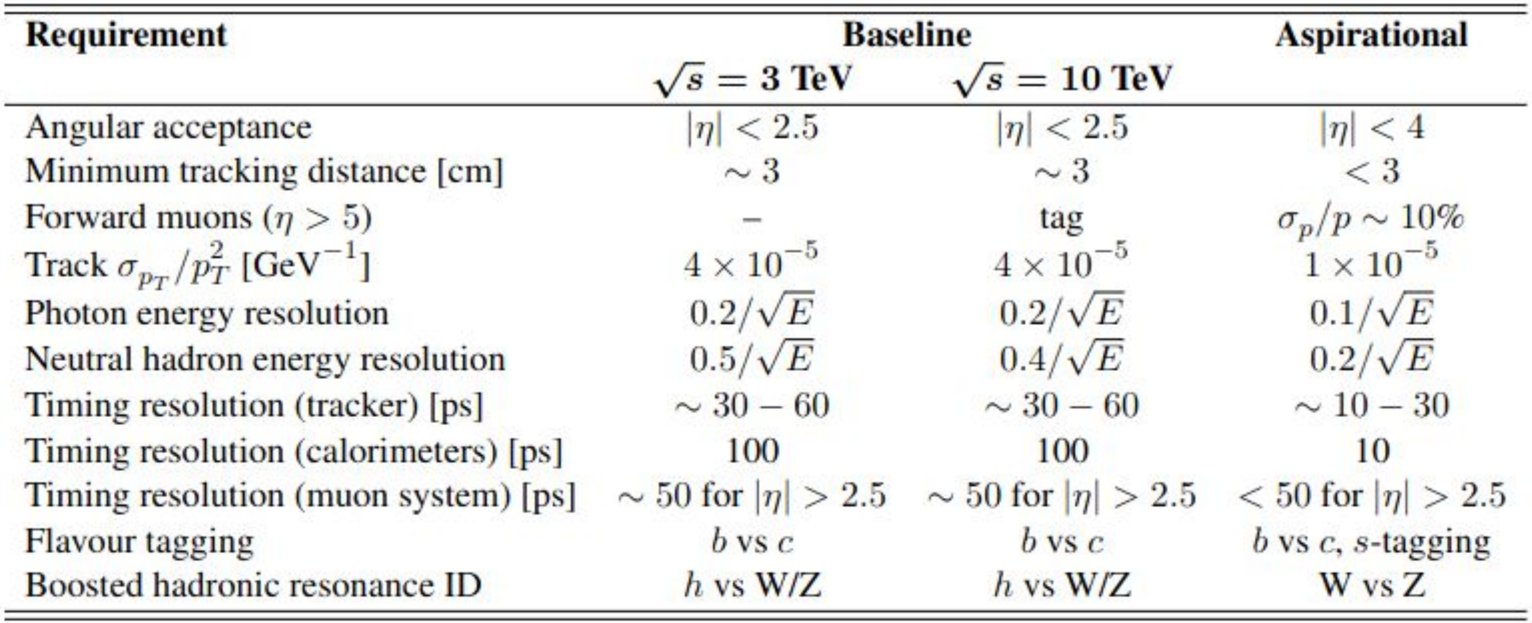
MUSIC

(MUon System for Interesting Collisions)

MAIA

(Muon Accelerator Instrumented Apparatus)

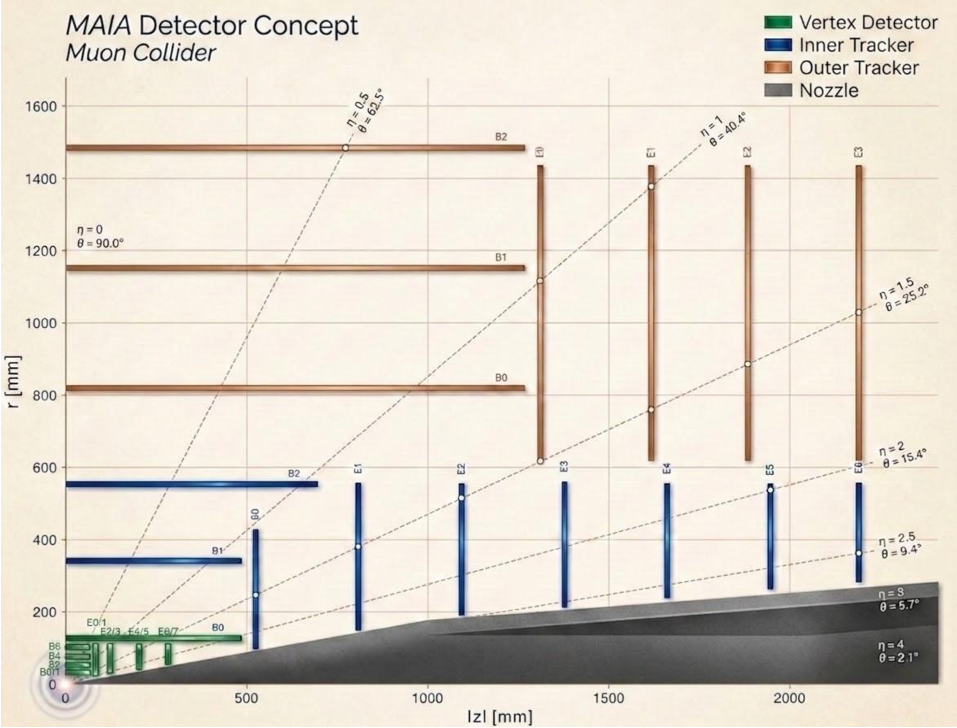
Key design parameters

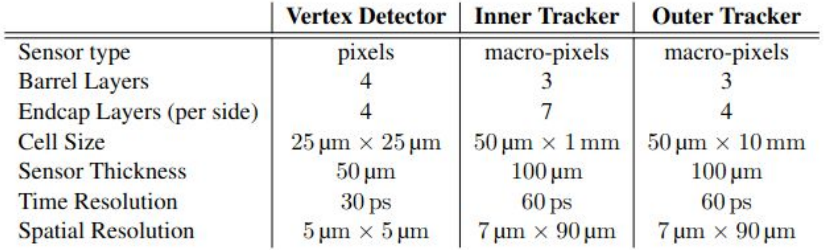


Requirement	Baseline $\sqrt{s} = 3$ TeV	Baseline $\sqrt{s} = 10$ TeV	Aspirational
Angular acceptance	$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 4$
Minimum tracking distance [cm]	~ 3	~ 3	< 3
Forward muons ($\eta > 5$)	–	tag	$\sigma_p/p \sim 10\%$
Track σ_{p_T}/p_T^2 [GeV^{-1}]	4×10^{-5}	4×10^{-5}	1×10^{-5}
Photon energy resolution	$0.2/\sqrt{E}$	$0.2/\sqrt{E}$	$0.1/\sqrt{E}$
Neutral hadron energy resolution	$0.5/\sqrt{E}$	$0.4/\sqrt{E}$	$0.2/\sqrt{E}$
Timing resolution (tracker) [ps]	$\sim 30 - 60$	$\sim 30 - 60$	$\sim 10 - 30$
Timing resolution (calorimeters) [ps]	100	100	10
Timing resolution (muon system) [ps]	~ 50 for $\lvert\eta\rvert > 2.5$	~ 50 for $\lvert\eta\rvert > 2.5$	< 50 for $\lvert\eta\rvert > 2.5$
Flavour tagging	b vs c	b vs c	b vs c, s-tagging
Boosted hadronic resonance ID	h vs W/Z	h vs W/Z	W vs Z

Tracker design

All silicon-based, high spatial precision, fast timing, high radiation tolerance, low material budget

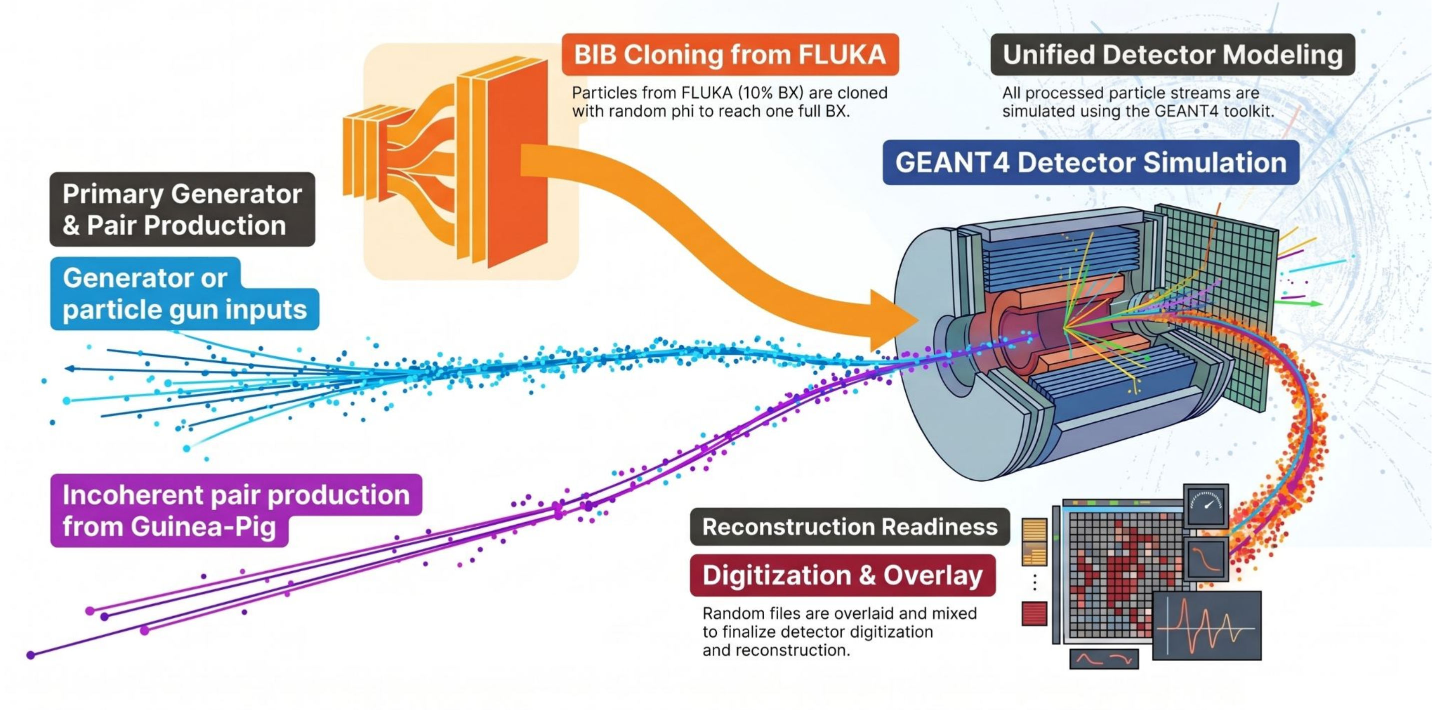




	Vertex Detector	Inner Tracker	Outer Tracker
Sensor type	pixels	macro-pixels	macro-pixels
Barrel Layers	4	3	3
Endcap Layers (per side)	4	7	4
Cell Size	25 µm × 25 µm	50 µm × 1 mm	50 µm × 10 mm
Sensor Thickness	50 µm	100 µm	100 µm
Time Resolution	30 ps	60 ps	60 ps
Spatial Resolution	5 µm × 5 µm	7 µm × 90 µm	7 µm × 90 µm

- First vertex barrel layer and all vertex endcap layers are implemented as double-layers, with 2 mm gap.
- Currently, assuming silicon planar sensors but future technology choice based on 3D, LGADs, or MAPS sensors R&D.

Simulating μ-Collider environment



BIB Cloning from FLUKA

Particles from FLUKA (10% BX) are cloned with random phi to reach one full BX.

Unified Detector Modeling

All processed particle streams are simulated using the GEANT4 toolkit.

GEANT4 Detector Simulation

Primary Generator & Pair Production

Generator or particle gun inputs

Incoherent pair production from Guinea-Pig

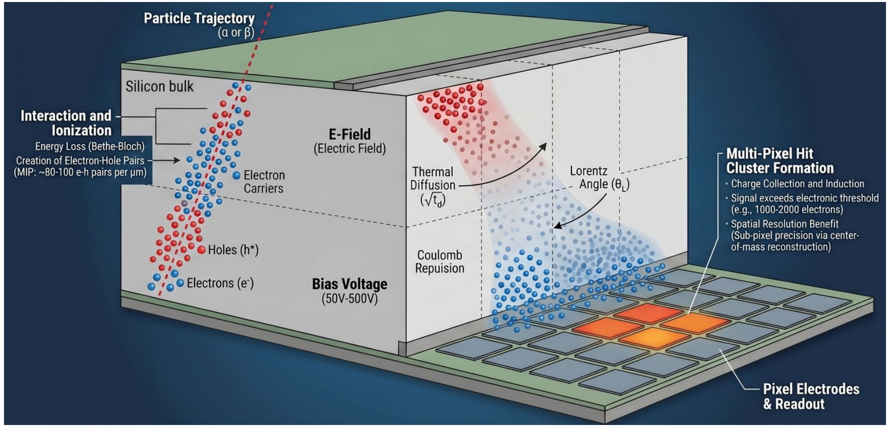
Reconstruction Readiness

Digitization & Overlay

Random files are overlaid and mixed to finalize detector digitization and reconstruction.

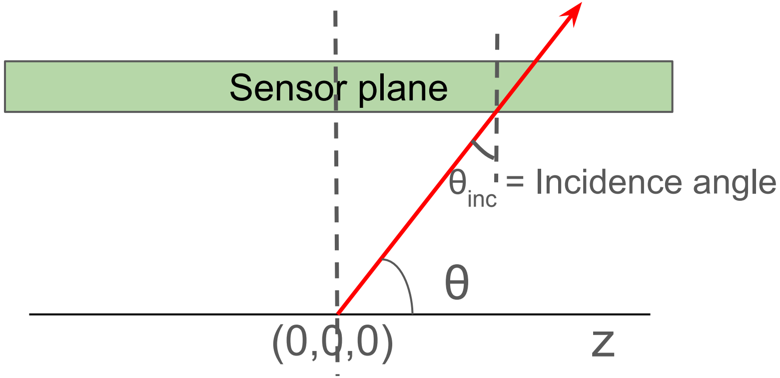
Tracker detector digitization

- **Particle Ionization:** When a charged particle travels through the silicon planar sensor, it ionizes the material and generates electron-hole pairs along its trajectory.
- **Charge Drift:** A high bias voltage applied creates an electric field in the depleted zone. This field separates the electron-hole pairs and forces them to drift toward the collection electrodes (pixel “hits”).
- **Cloud Expansion:** As the charge cloud drifts toward the pixels, it spreads out and grows in size. This expansion is driven by thermal diffusion and Coulomb repulsion.
- **Cluster Formation:** A "hit cluster" forms when the charge cloud is split and collected by two or more adjacent pixels.

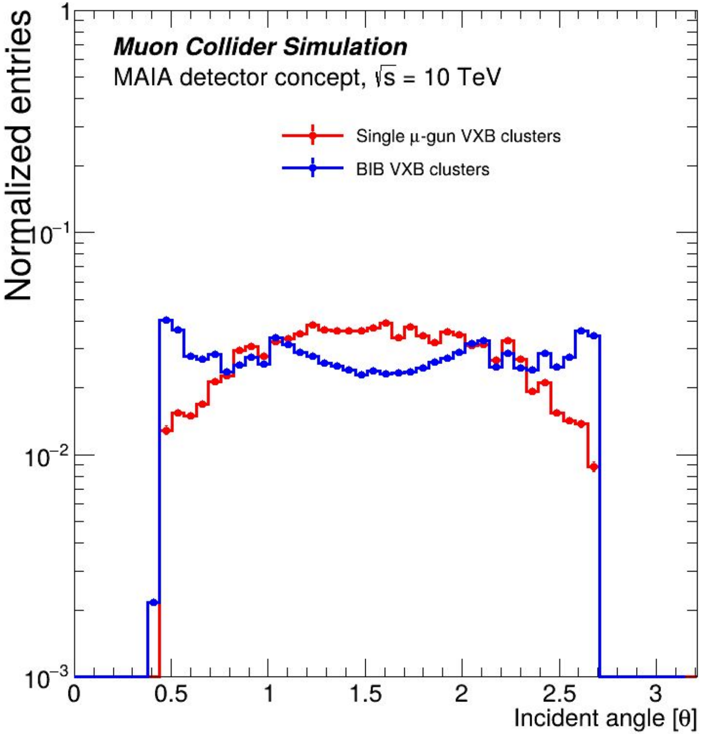


Signal vs BIB

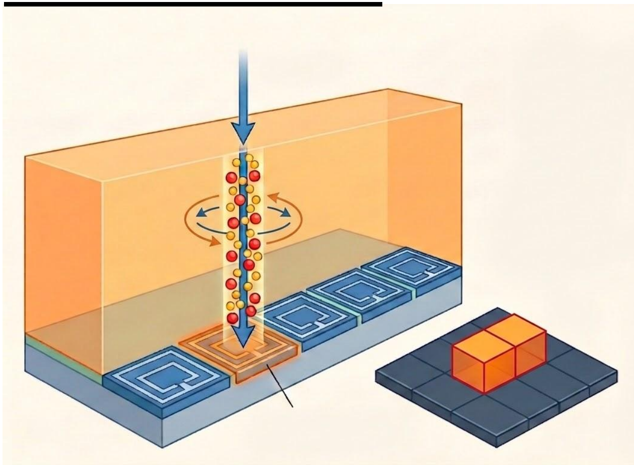
- Single muon particle gun with pT = 0 to 5 TeV and |η|<2.4 (signal) & BIB



- Given signal has normal incidence (since high pt particles), and BIB has shallow incidence, how does the pixel hit cluster shape change?

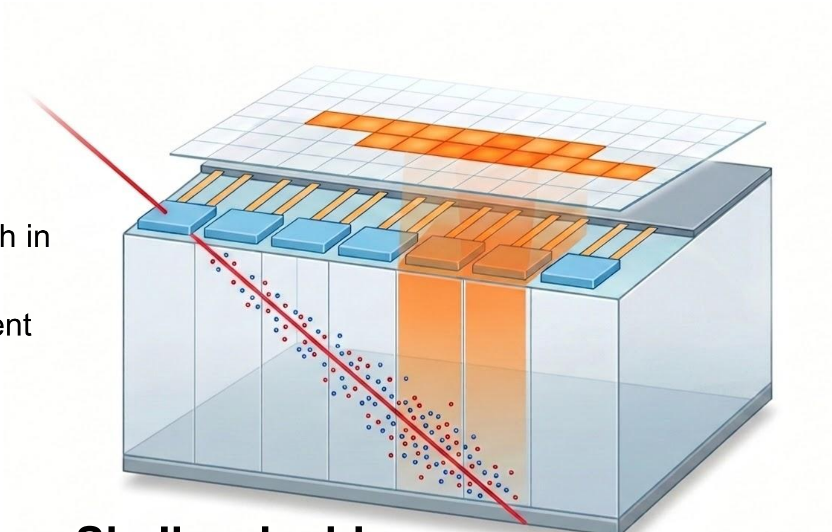


Normal incidence



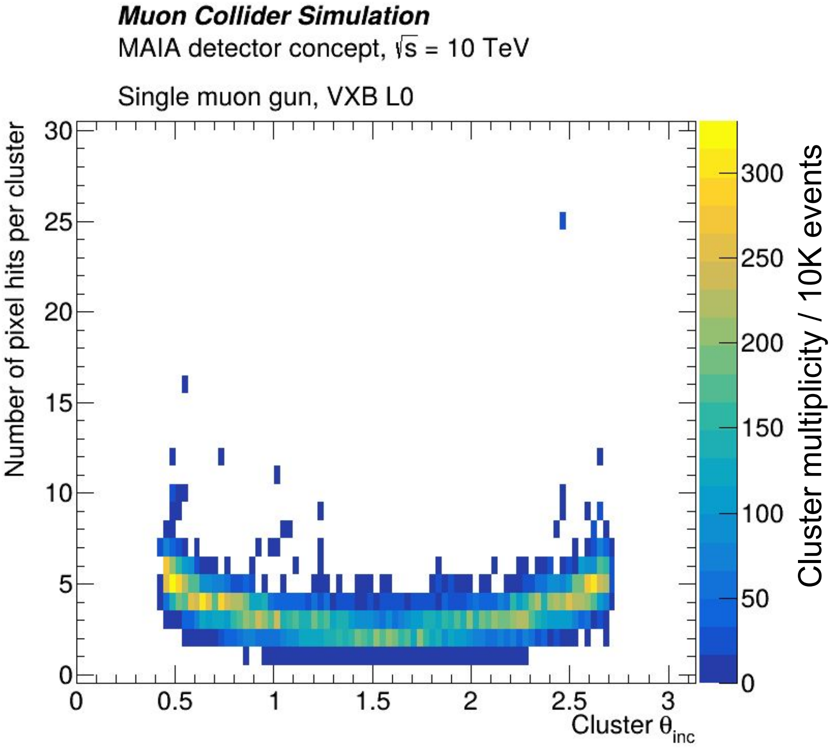
- Short, straight vertical column of electron-hole pairs through the silicon.
- Charge sharing only due to lateral expansion of the charge cloud (thermal diffusion and Coulomb repulsion) or lateral deflection from a magnetic field (Lorentz drift).
- Resulting hit cluster is highly compact, typically triggering only 1-2 pixels.

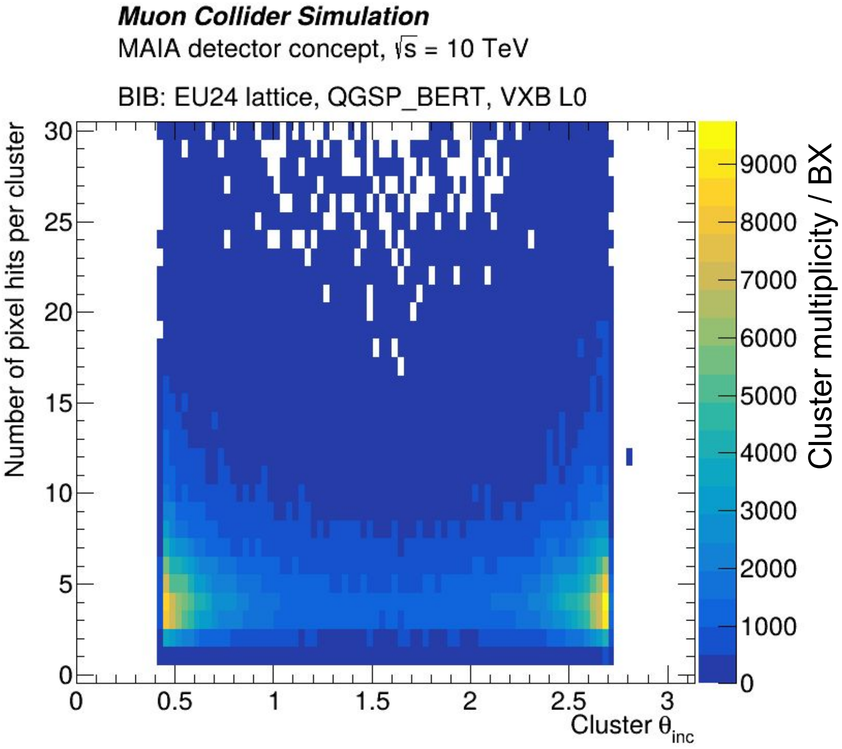
- Traveling diagonally and passing through a longer physical path in the silicon before exiting.
- Electron-hole pairs directly into the territories of multiple adjacent pixels (irrespective of diffusion and drift).
- Overall size of the cluster grows rapidly with angle.
- Elongated or multi-hit pixel cluster.



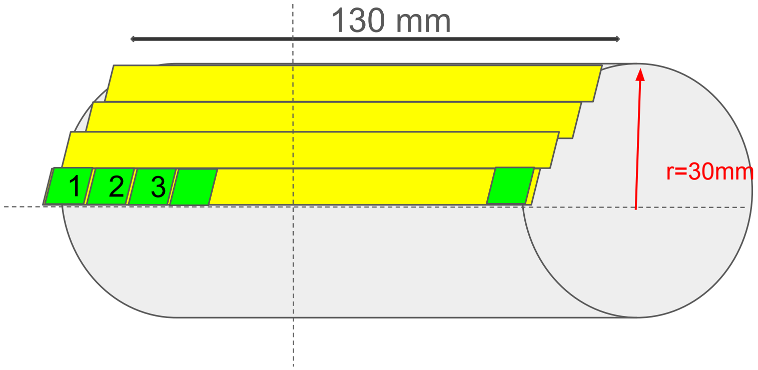
Shallow incidence

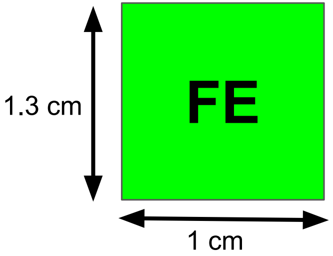
Simulation results: Innermost tracker layer



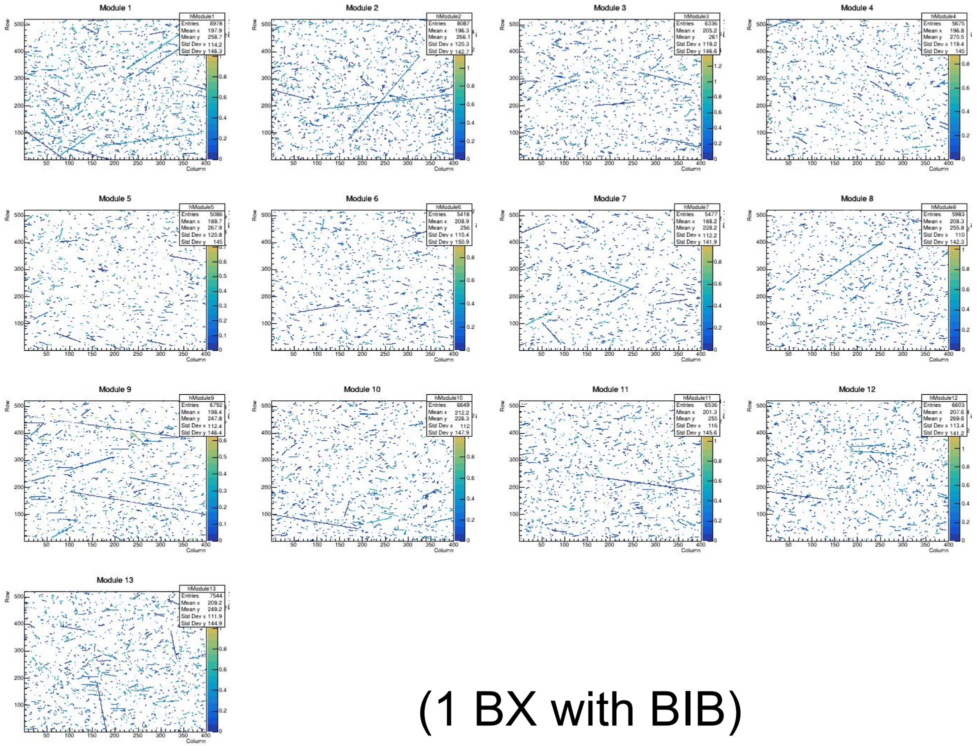


Pixel Hit-map of VXB L0





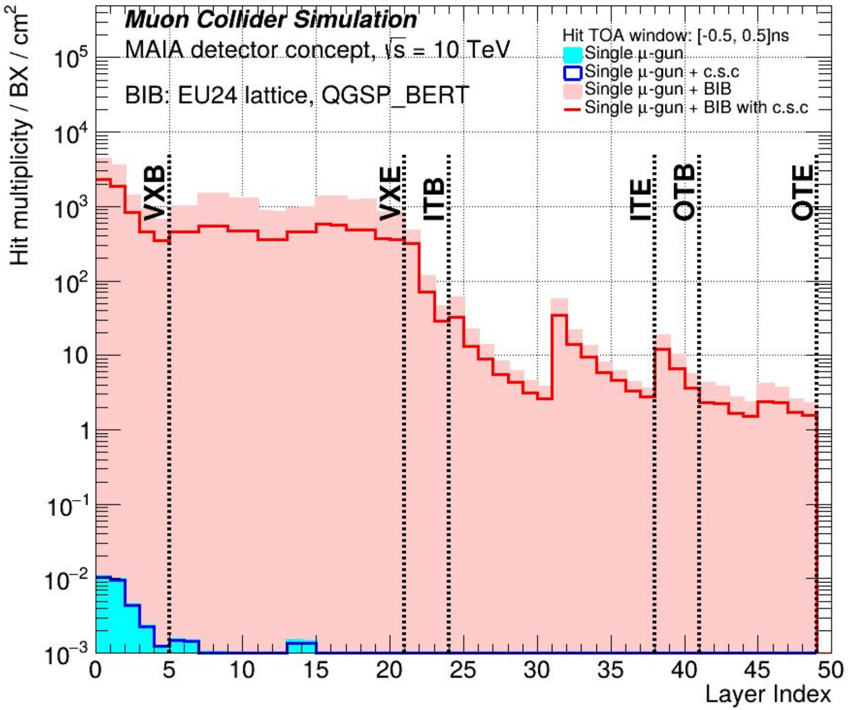
Total pixels ~ 520 x 400 pixels



(1 BX with BIB)

Tracker hit occupancy

Very high hit multiplicities per BX in the innermost layers of the detectors.

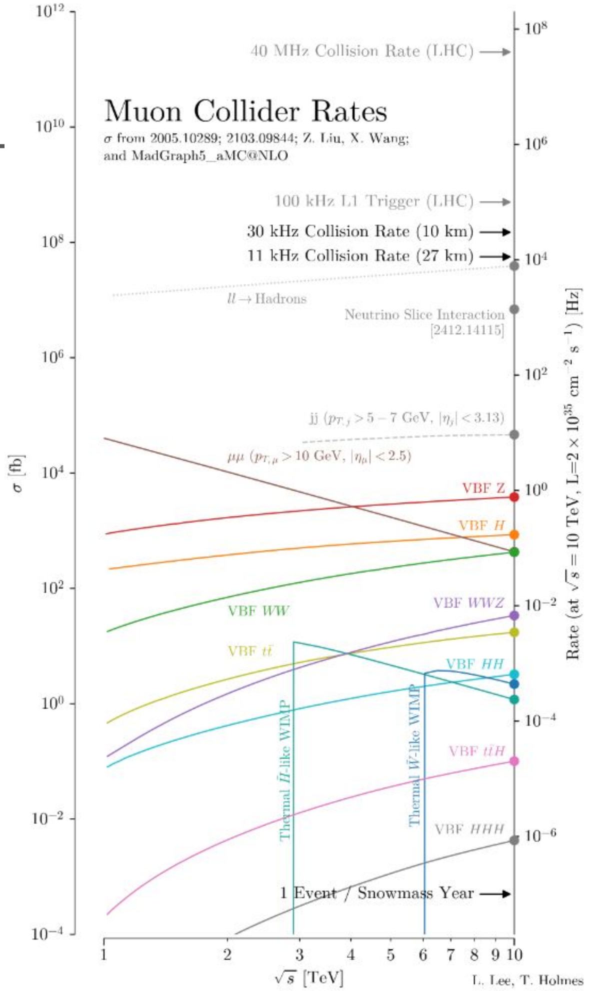


- ~4000 hits/BX/cm^2 in VXB L0
- ~16000 pixel hits on FE (with RD53-like design, 153600 pixels)
- ~10% tracker occupancy!!

- Compared to ATLAS ITk requirement for HL-LHC = 1%

Event rates

- 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} to 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, but full of BIB.



What do we want to do?

i.e. keep only interesting events (triggered-readout)?

Or stream all data (triggerless readout) but reducing readout data volume?

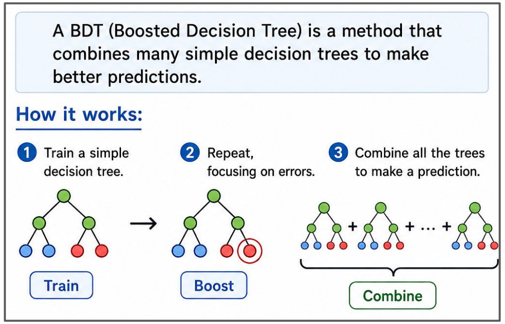
Can we classify between signal and BIB clusters in an event? (Online or offline)

Part II

Signal vs BIB pixel-hit clusters (offline design)

The BDT vs. The Transformer

Apply a Boosted Decision Tree to Particle Gun Samples



A BDT (Boosted Decision Tree) is a method that combines many simple decision trees to make better predictions.

How it works:

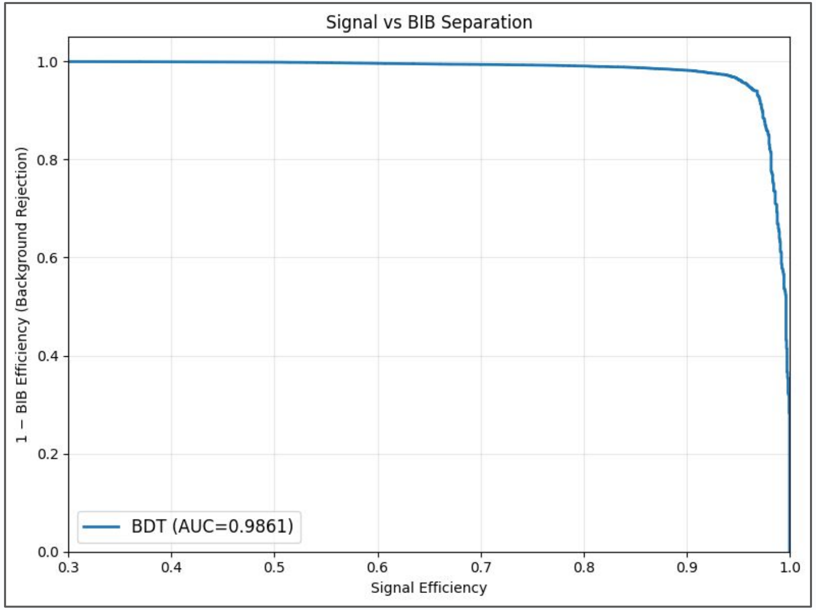
1. Train a simple decision tree.
2. Repeat, focusing on errors.
3. Combine all the trees to make a prediction.

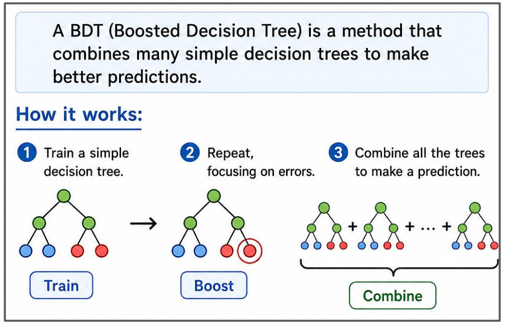
Cluster Features

Energy	Time	R
Size X	Size Y	Total Size
RMS X	RMS Y	
Skew X	Skew Y	
Incident Angle	Aspect	Eccentricity

Apply a Boosted Decision Tree to Particle Gun Samples

Super good! No room for improvement




A BDT (Boosted Decision Tree) is a method that combines many simple decision trees to make better predictions.

How it works:

1. Train a simple decision tree.
2. Repeat, focusing on errors.
3. Combine all the trees to make a prediction.

Cluster Features

Energy	Time	R
Size X	Size Y	Total Size
RMS X	RMS Y	
Skew X	Skew Y	
Incident Angle	Aspect	Eccentricity

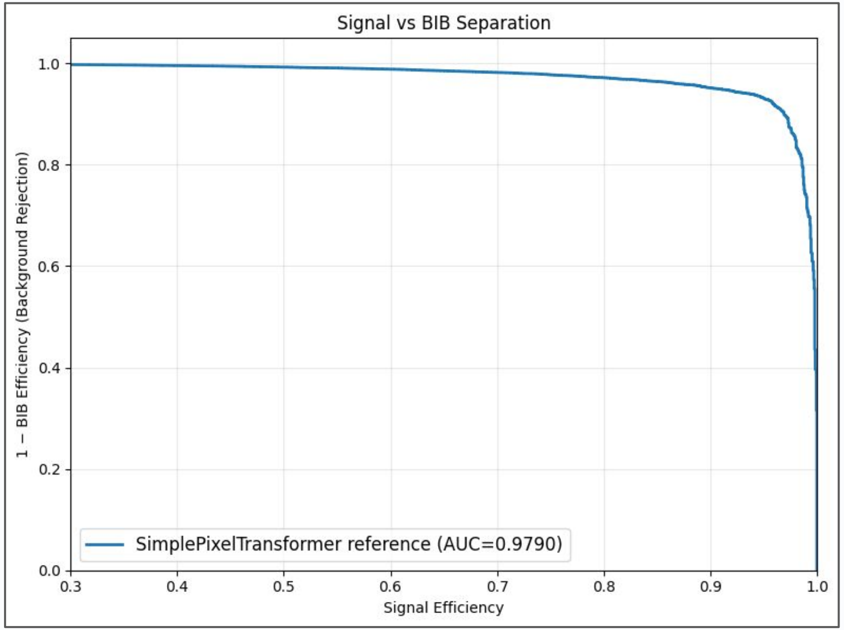
Apply the Transformer to Particle Gun Samples

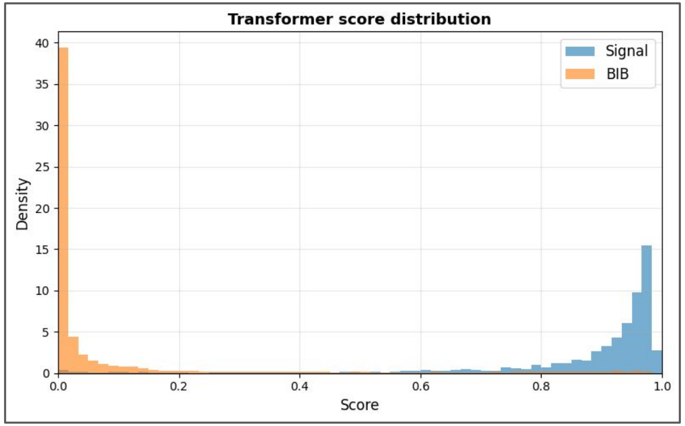
Raw Hit Features

Energy	Time
X Position	Y Position

Apply the Transformer to Particle Gun Samples

Super good! But slightly worse than the BDT

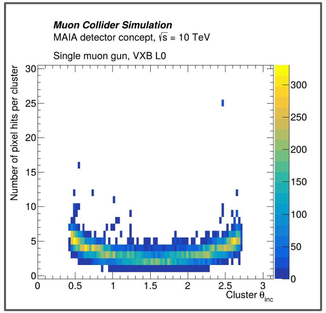




Raw Hit Features

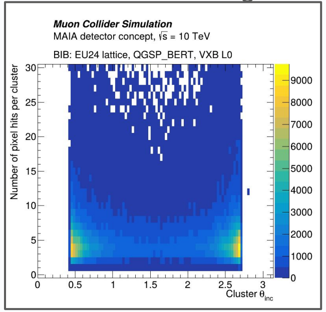
Energy	Time
X Position	Y Position

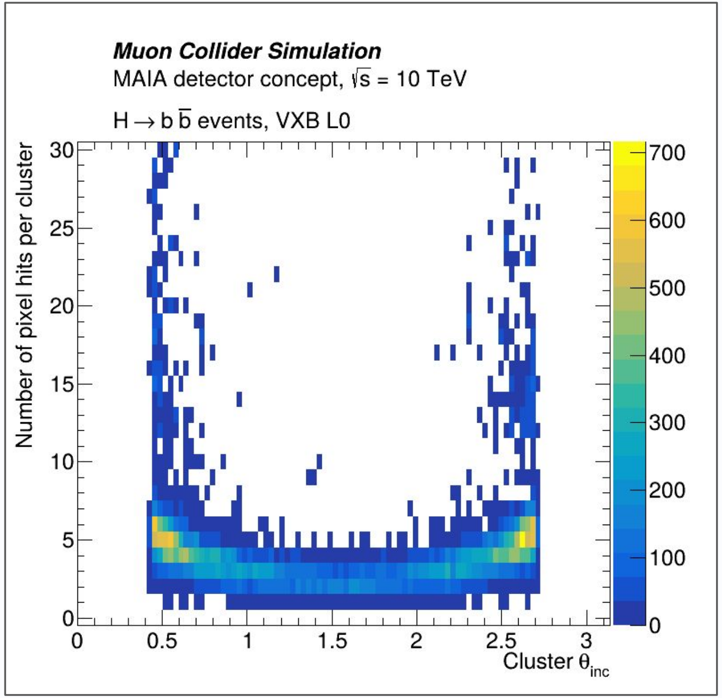
Switching to H → bb Signal Events

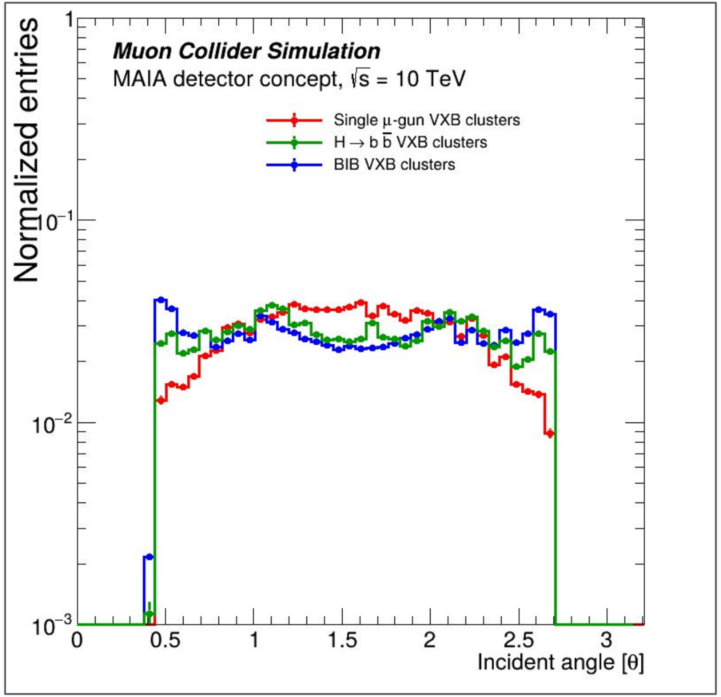


Particle Gun

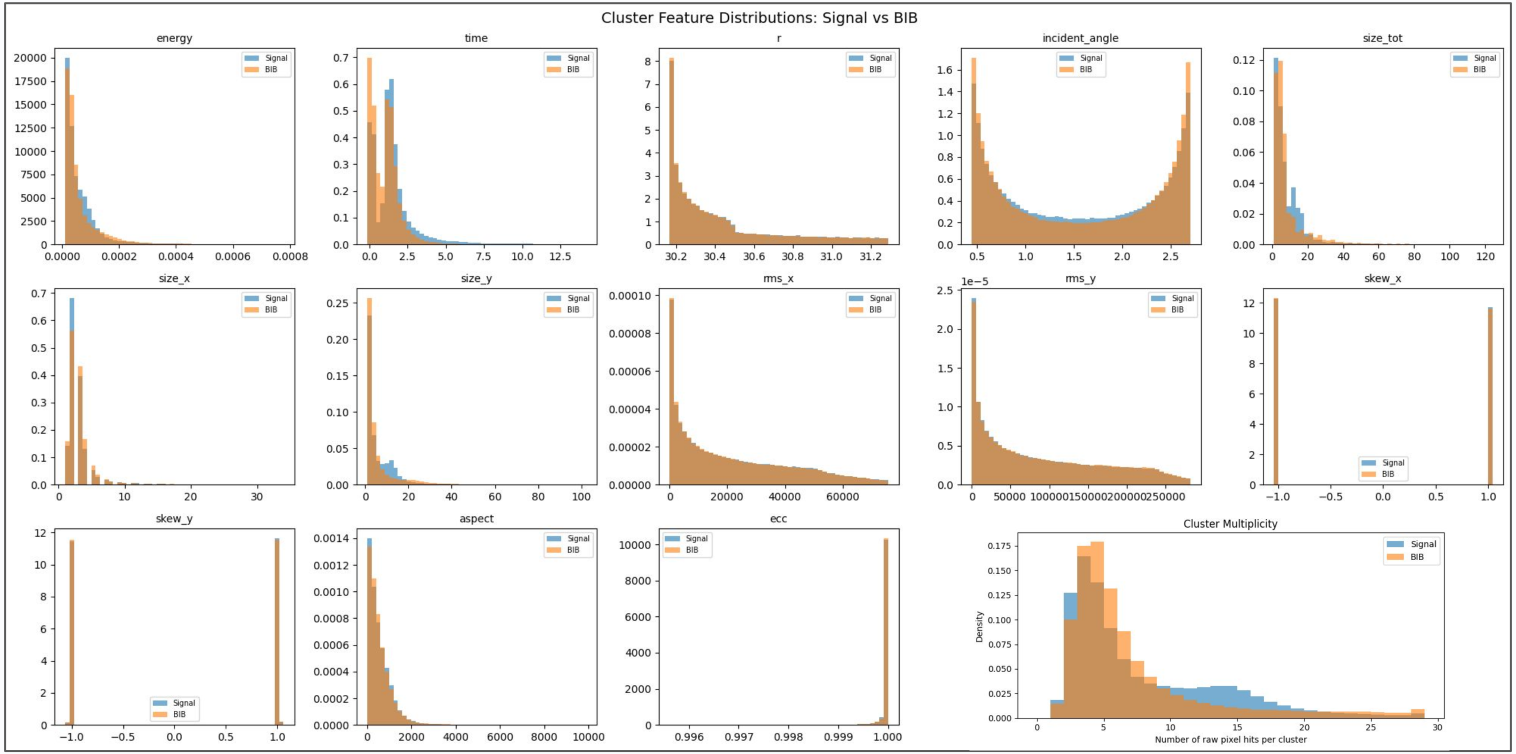
Beam-Induced Background





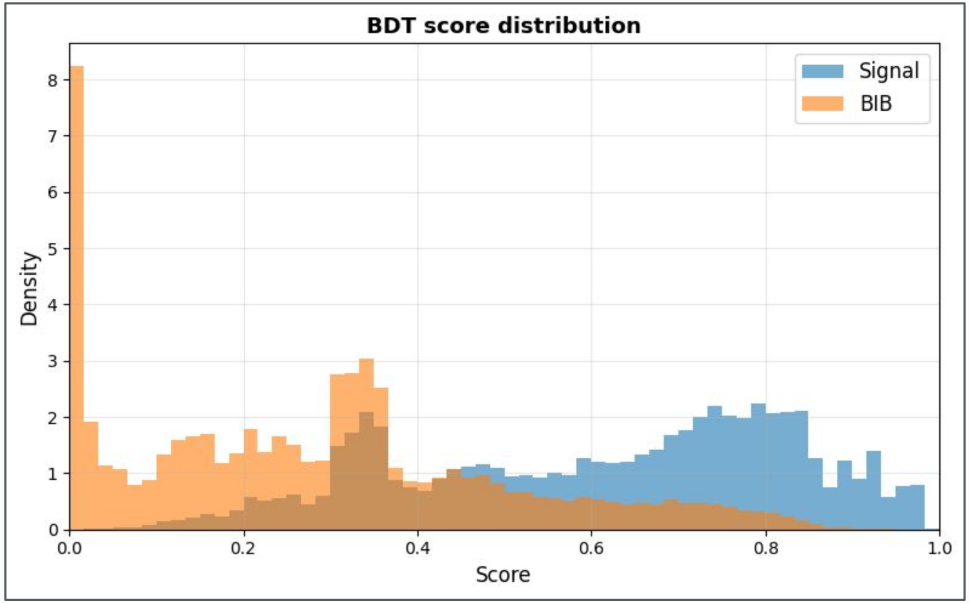


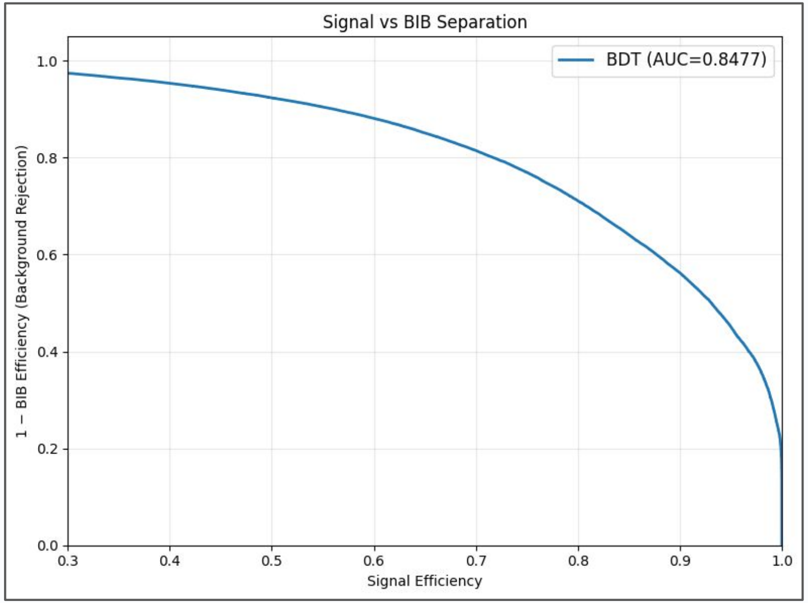
Feature Distributions



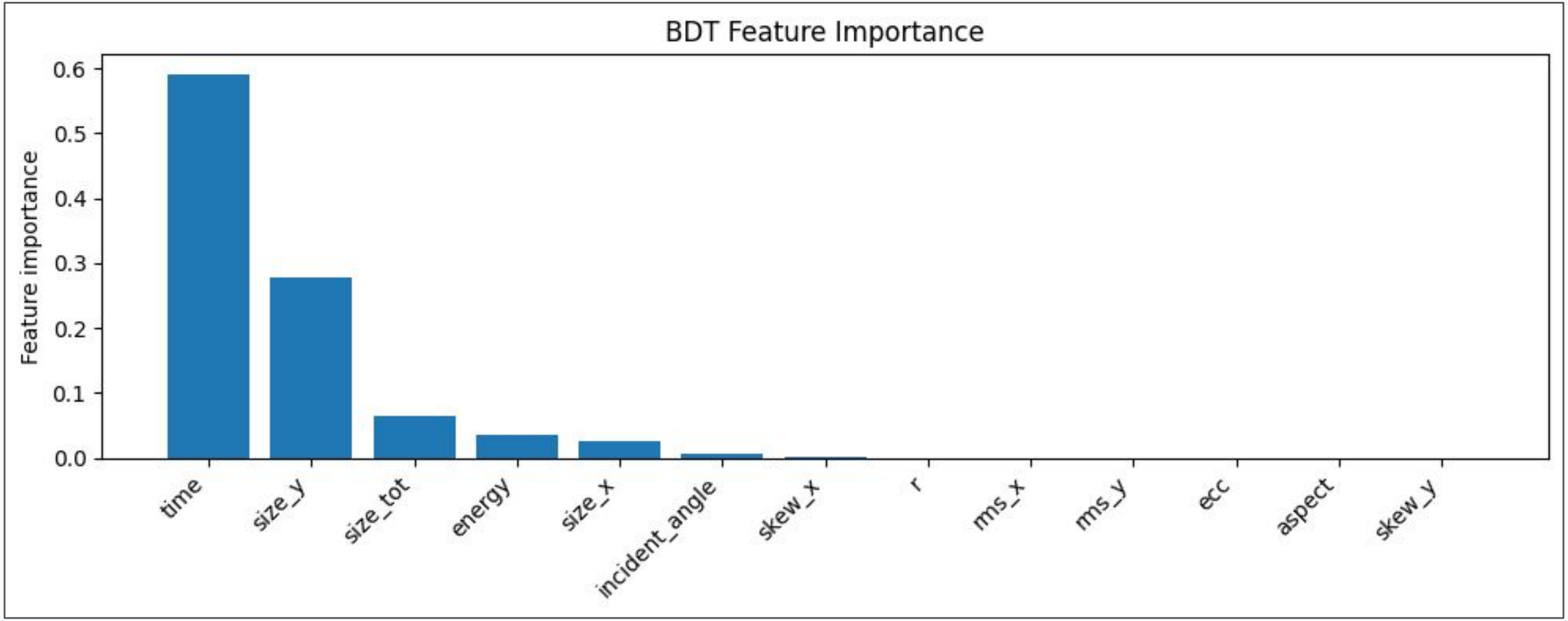
Apply BDT to H → bb Samples

Lots of room for improvement!





BDT Tells Us About Important Features



Why Would a Transformer Help Here?

Raw Hit Features

Energy	Time
X Position	Y Position

Cluster Features

Energy	Time	R
Size X	Size Y	Total Size
RMS X	RMS Y	
Skew X	Skew Y	
Incident Angle	Aspect	Eccentricity

Can the Transformer learn these?

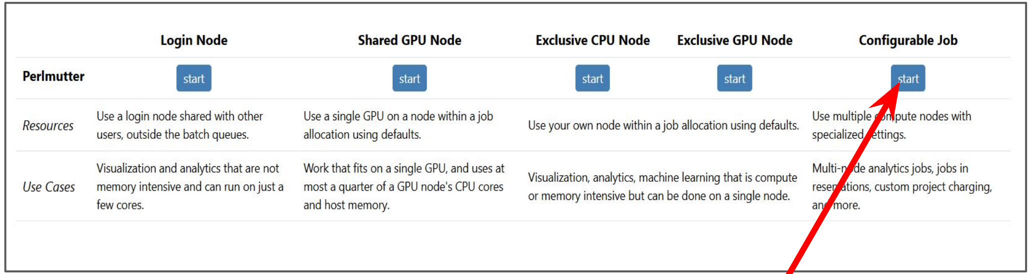
Or maybe even more important quantities?

Raw Hits are a ragged data type. (Zero padded)

↓

We can mask entries in the Transformer

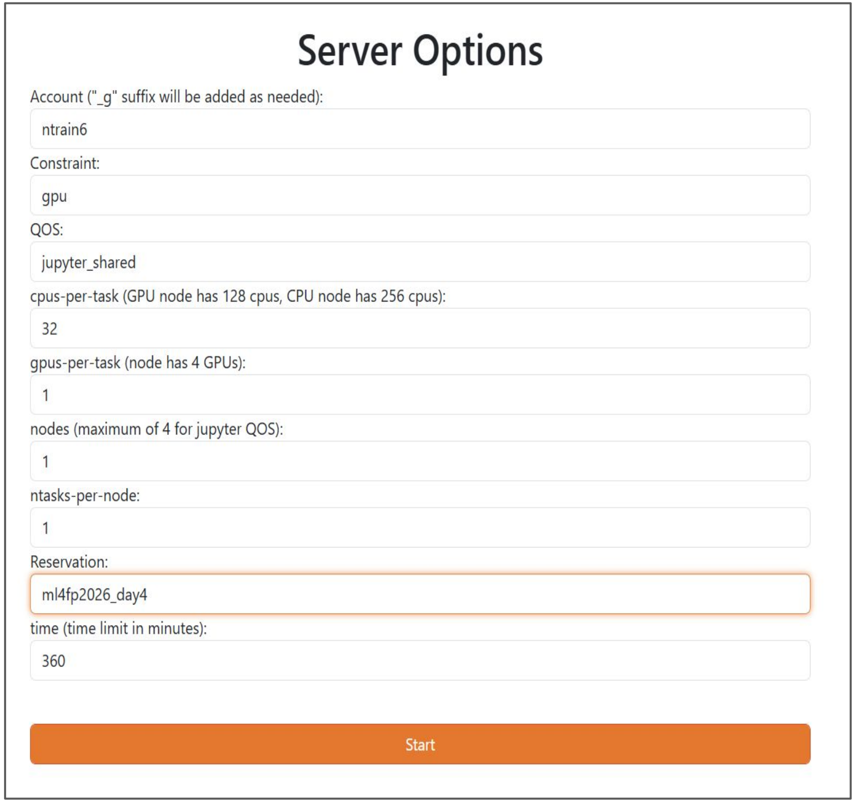
Setting Up the Tutorial



	Login Node	Shared GPU Node	Exclusive CPU Node	Exclusive GPU Node	Configurable Job
Perlmutter	start	start	start	start	start
Resources	Use a login node shared with other users, outside the batch queues.	Use a single GPU on a node within a job allocation using defaults.	Use your own node within a job allocation using defaults.		Use multiple compute nodes with specialized settings.
Use Cases	Visualization and analytics that are not memory intensive and can run on just a few cores.	Work that fits on a single GPU, and uses at most a quarter of a GPU node's CPU cores and host memory.	Visualization, analytics, machine learning that is compute or memory intensive but can be done on a single node.		Multi-node analytics jobs, jobs in reservations, custom project charging, and more.

https://jupyter.nersc.gov

Configurable Job



Server Options

Account ("_g" suffix will be added as needed): ntrain6

Constraint: gpu

QOS: jupyter_shared

cpus-per-task (GPU node has 128 cpus, CPU node has 256 cpus): 32

gpus-per-task (node has 4 GPUs): 1

nodes (maximum of 4 for jupyter QOS): 1

ntasks-per-node: 1

Reservation: ml4fp2026_day4

time (time limit in minutes): 360

Setting Up the Tutorial

1. Go to the Terminal
2. Go to the Repo:
   ```
   cd ~/2026-gatech
   ```
3. Update the Repo:
   ```
   git add -u
   git commit -m”past tutorials”
   git pull
   ```
4. Set up the environment:
   ```
   cd sessions/04_muon_col/tutorial
   source setupTutorialEnv.sh
   ```
5. Open the notebook:
   `02_simple_transformer_tutorial.ipynb`
6. Set the Kernel to `pytorch-2.6.0`

Please record the AUC value of your best Transformer training at the end of the session to this spreadsheet - https://docs.google.com/spreadsheets/d/1Q2ISRGGdE4QYXJ82086jcRxQP5nhCUqv9Kh0Ku8BKNQ/edit?usp=sharing

Also, give brief details about the exact method you tried!