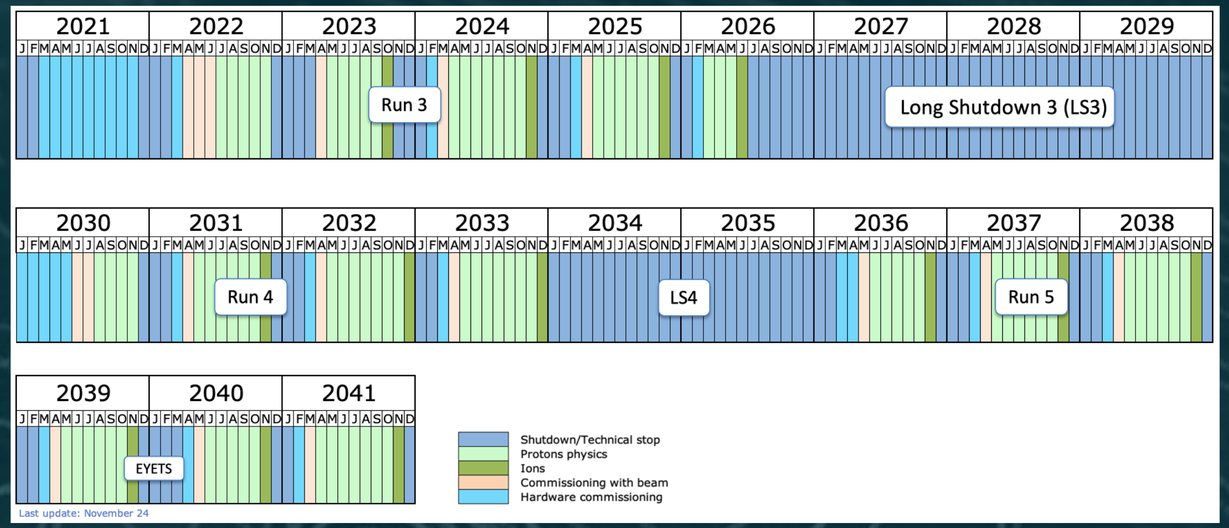
HL-LHC & Muon Collider Electronics Challenges and Requirements

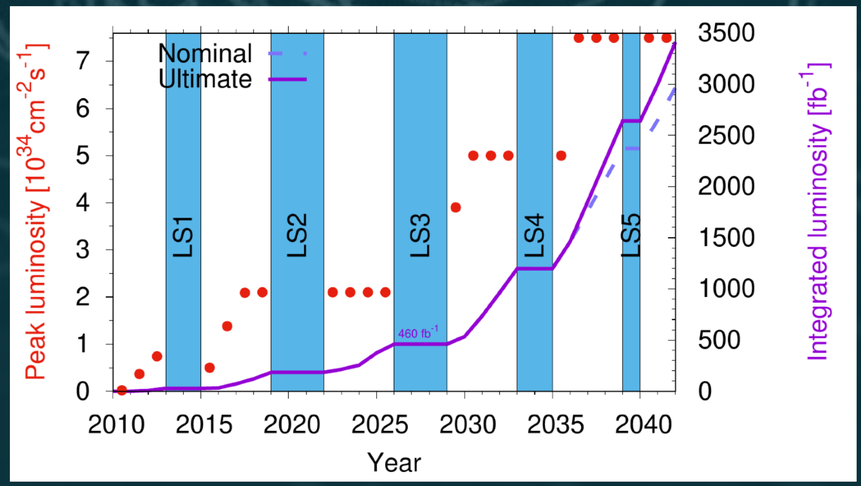
4D Tracking Workshop - 07.11.24

Timon Heim - LBNL

HL-LHC Intro

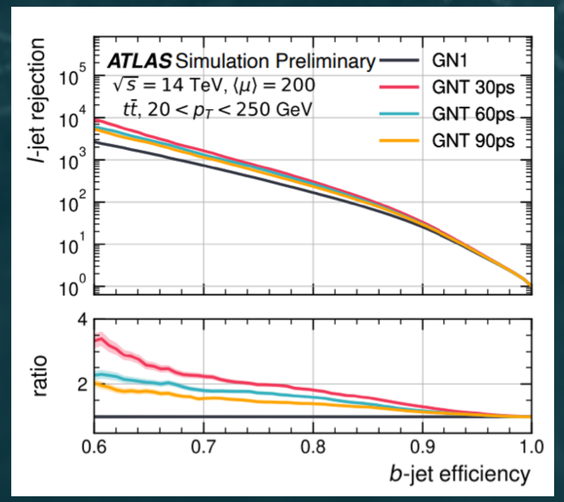
- ATLAS & CMS Innermost Layers not designed to withstand full lifetime HL-LHC radiation dose
  - Full lifetime integrated luminosity used to be 4000fb-1, now current nominal targets 2500fb-1
  - Innermost system designed for 2000fb-1
- Replacement in LS5 after 80% of runtime undesirable
  - For replacement in LS4 need strong physics motivation
    - Due to timescale would also be strongly constrained to re-use as much of the current system as possible
  - Investigation about performance impact if operated up to 2500fb-1 tbd
- No official organization within ATLAS to investigate “Phase 3” upgrade
  - Fear of doing so at the cost of ITk
- CMS has more concrete plans about a potential upgrade

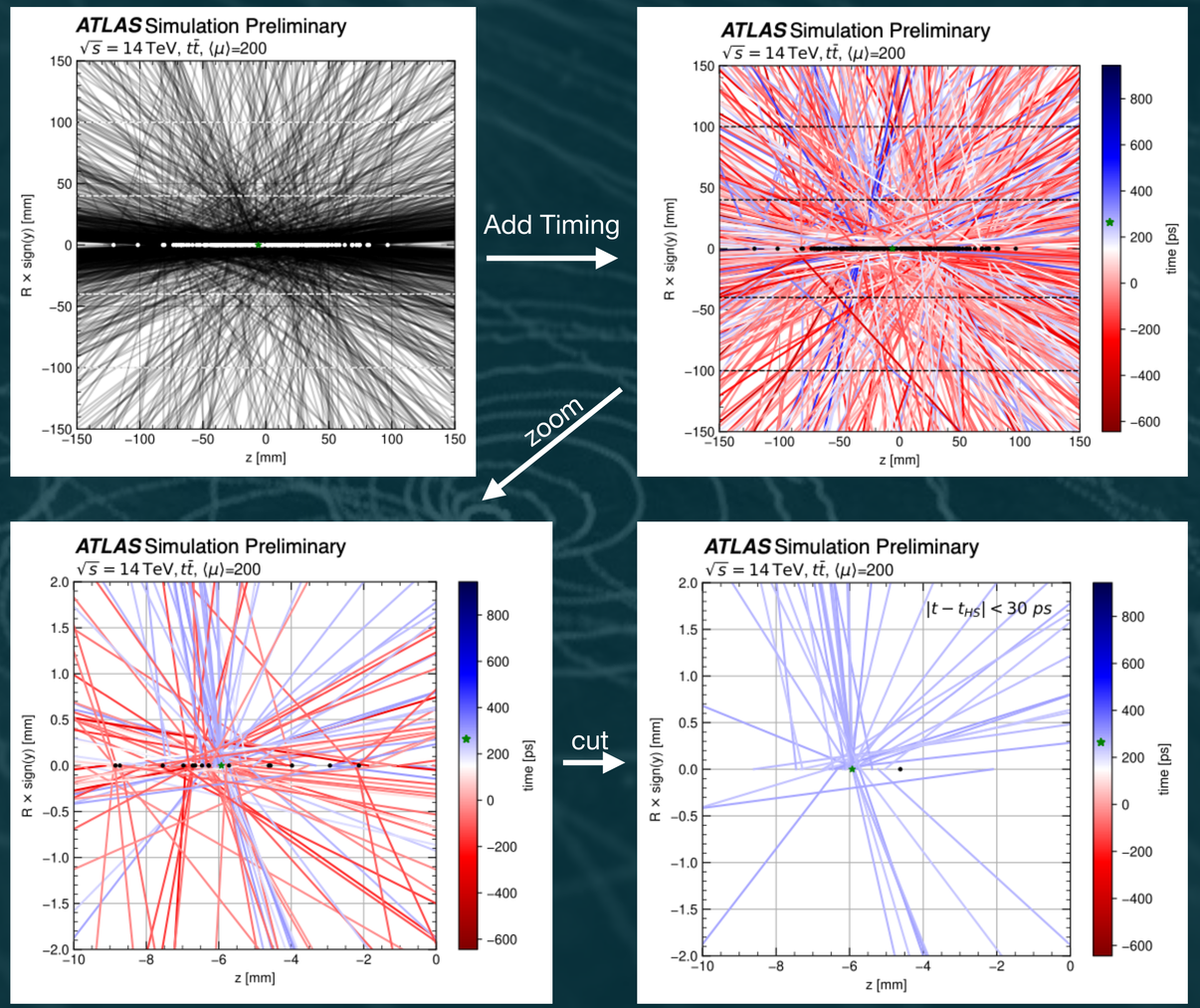




ATLAS Studies

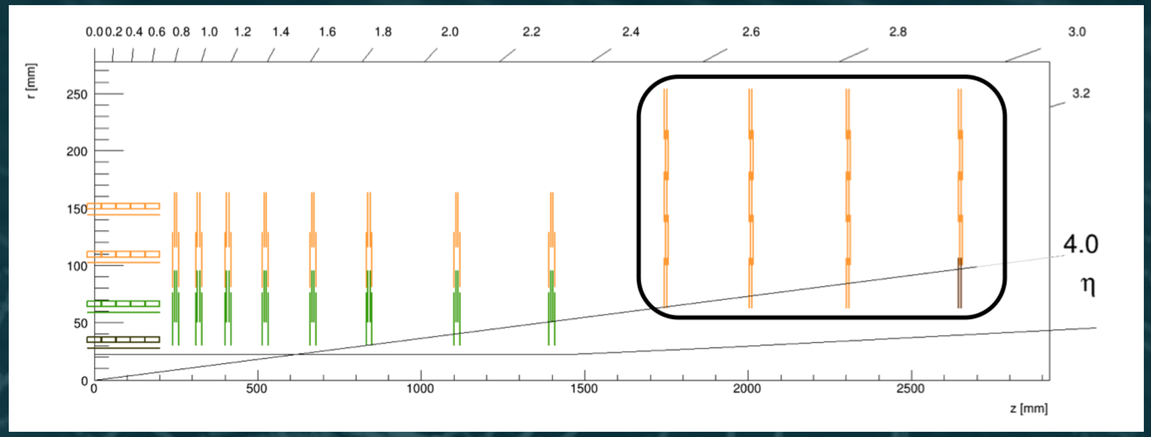
- ATLAS does not have a barrel timing layer compared to CMS -> Desire to have timing information for barrel tracks
- Timing layer hits in endcap hard to connect to tracks due to material from service in between tracker and timing detector
- Few studies to look into interesting aspects
  - Primary feature for upgrade that has been looked at is addition of precision of timing
    - Precision timing helps with vertex association in high pile-up environments
    - Reduces computational load during seeding (most computing intense part of track reco) by reducing combinatorics





CMS Studies

- CMS considers a two prong approach
  - Investigate smaller pixel pitch for barrel (25um x 25um)
    - Reducing material budget gives similar gain to smaller pixels! Could also investigate a lower power version of RD53 chip and lower mass data transmission …
  - End caps with larger pixels (100um x 100um) but high timing precision (to extend timing information beyond barrel timing detector), sigma(t) < 50ps
    - Use AC-LGADs for high efficiency and use charge sharing to improve



Inner regions	Timing disks
Pixel size 25 × 100 μm^2	Pixel size 100 × 100 μm^2
Detection threshold $\ll$ 900 e^-	Timing resolution < 50 ps
Power density $\ll$ 0.6 W/cm^2	Power density $\lesssim$ 0.6 W/cm^2

Chip size (h × w) (16.8 × ~ 21.6) mm^2

Output bandwidth $\lesssim$ 5 Gbps

Serial powering infrastructure

Trigger and latency as in phase-2

Interface to silicon photonics link

HL-LHC Challenges & Requirements

BERKELEY LAB

Most requirements translate straight from Phase 2 Upgrade

Parameter	Value
Technology	65nm
Max. hit rate	3.5 GHz/cm2
Trigger Rate	1MHz
Trigger Latency	12.5us
Pixel size (chip)	50um x 50um
Pixel array	400x384
Chip dimension	20mm x 21mm
Detector Capacitance	<100fF
Detector Leakage	<10nA
Min. threshold	<1000e
Radiation Tolerance	1Grad @ -10C
SEE Tolerance	<100Hz/chjp
Power	<1W/cm2
Readout data rate	1-4 links @ 1.28Gbps
Temp. Range	-40C to +40C

28nm would enable squeezing more logic in and or improve power, R&D already started

RD53 couldn't fit more complex trigger schemes (ROI two level readout), higher logic density could enable this. Physics case for 2 layers?

Single transistor studies show improved radiation hardness compared to 65nm

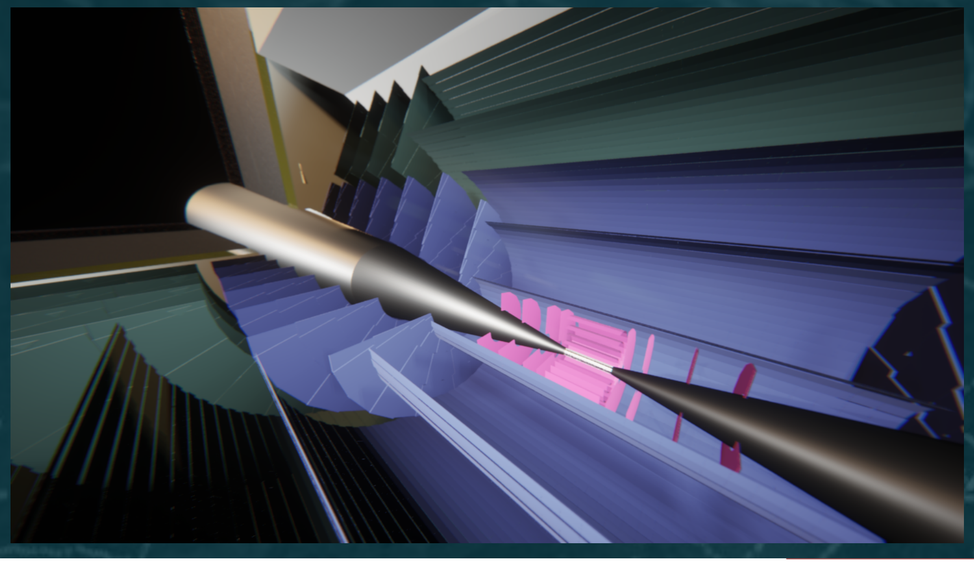
Increasing granularity or adding timing information will increase bandwidth needs. How to transmit over similar mass cables?

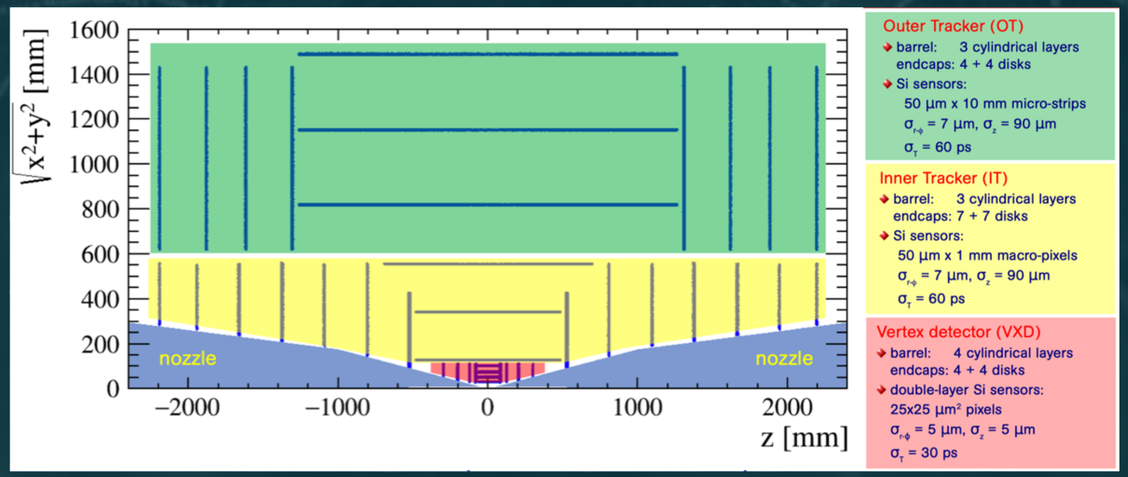
HL-LHC: Major Challenges for Electronics

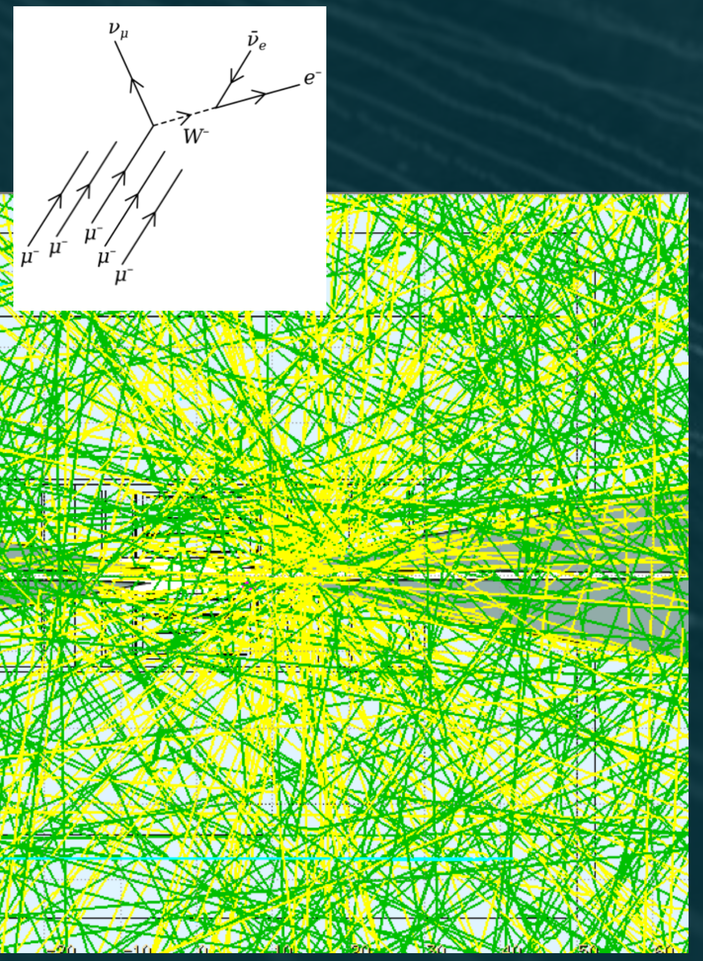
BERKELEY LAB

- Readout ASIC design:
  - Fitting a high performance chip into a low power environment -> More features and functions need to be balanced with power saving somewhere else, can not just focus on the new stuff, also need to improve the “old” stuff
  - Dealing with SEE is always troublesome and often done too late -> With high complexity comes high SEE vulnerability (and it’s usually very hard to identify what goes wrong at that point) -> Any new design needs to be designed with SEEs in mind, it might slow down initial prototyping but will pay off in the long run
  - Achieving operational conditions on the test bench is often near impossible -> invest into designing test functions that allow to emulate operational conditions
- Data transmission and Powering:
  - Serial power nearly solved material impact from powering, dominant material contribution is now data services -> Never the less data transmission at 1.28Gbps is borderline -> Need to pack more data onto similar or less mass cables
  - Optical data transmission can go fast “easily” -> I/O protocol best to designed with multiplexing in mind
- Data acquisition:
  - Data encoding schemes are highly efficient, but also highly complex -> Should also consider if decoding is feasible in suitable hardware, minor drawbacks in compression might yield huge improvements in decoding needs

Muon Collider Challenge: Beam Induced Background



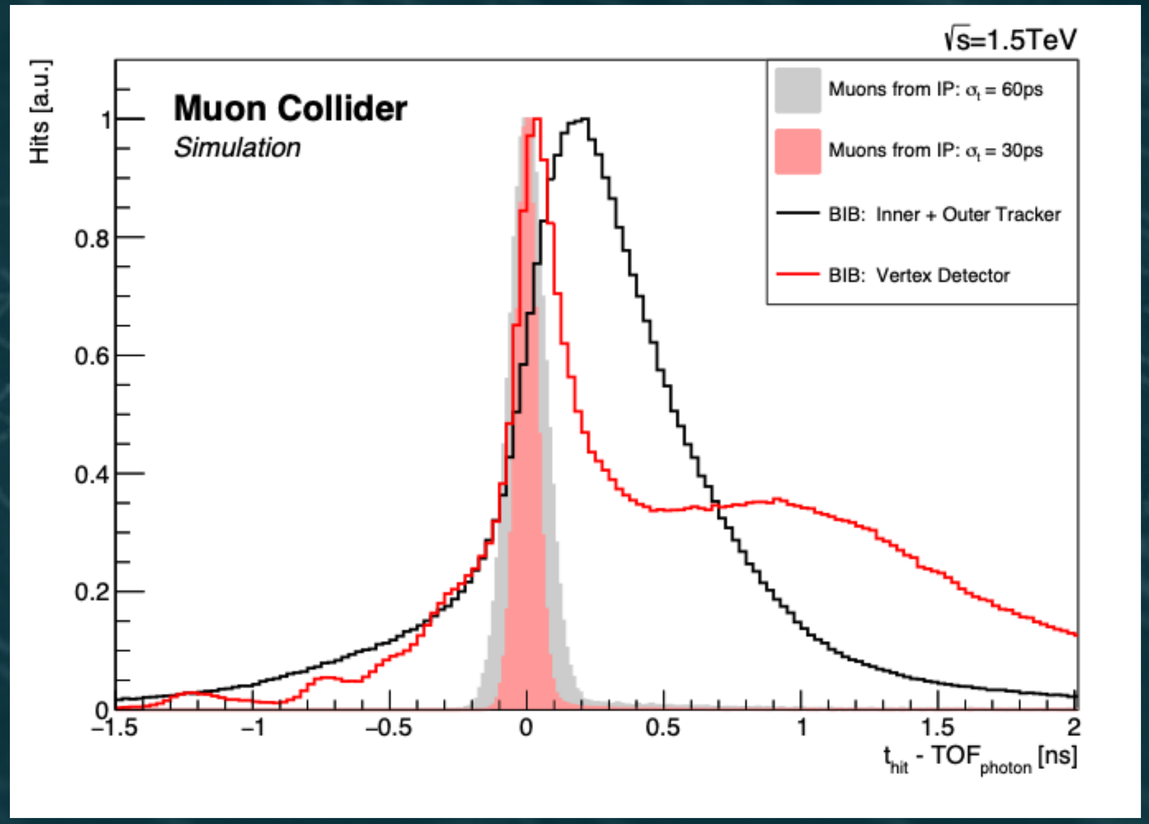




0.003% of total beam induced background

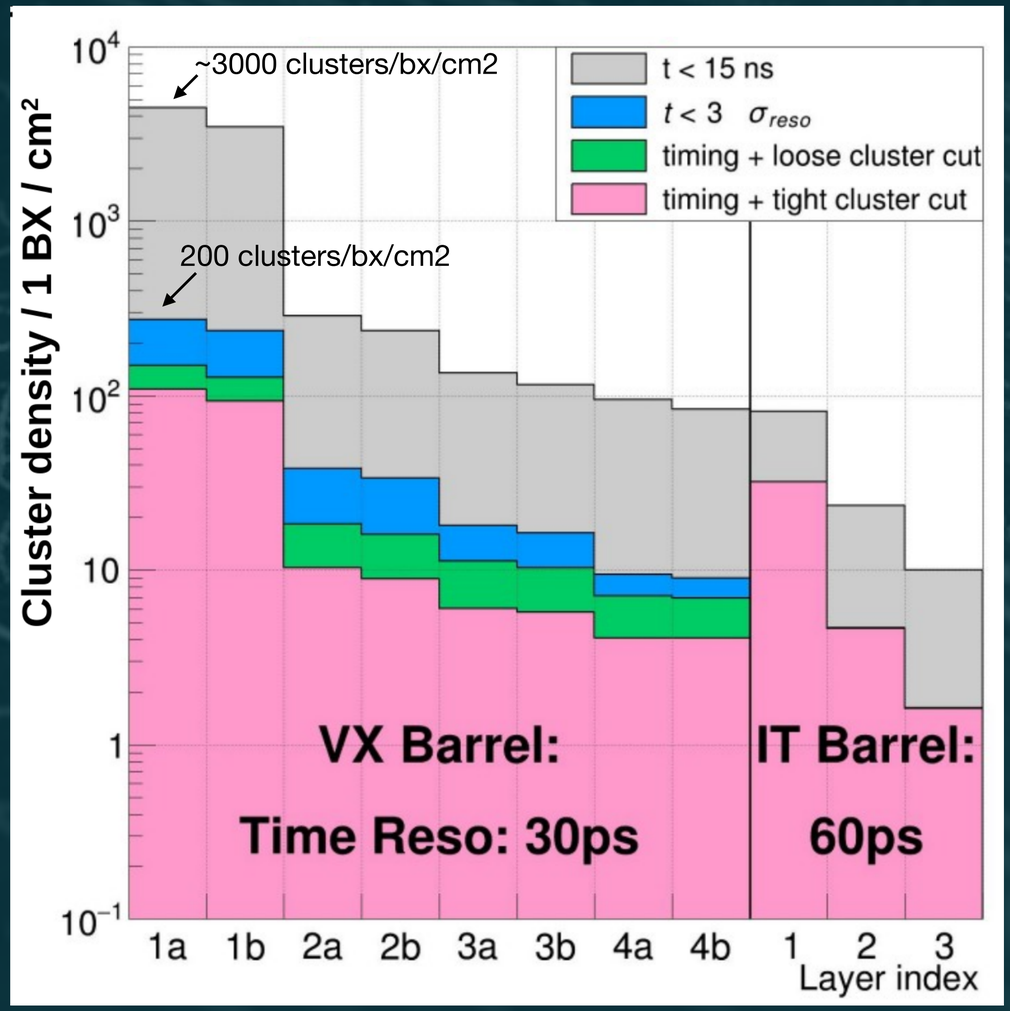
4D Tracking

- At muon collider 4D tracking is a necessity to reject beam induced background
  - Timing is primary discriminant to distinguish signal from background
  - But timing alone not sufficient, also have to check if track points back to IP (or pixel cluster shape)
    - Can potentially be done on-detector with double layers in innermost layer
- In simulation with 30ps timing information in inner tracker, reconstruction is feasible within a reasonable amount of time
- But simulation makes this look “too easy” and doesn’t tell the whole story …



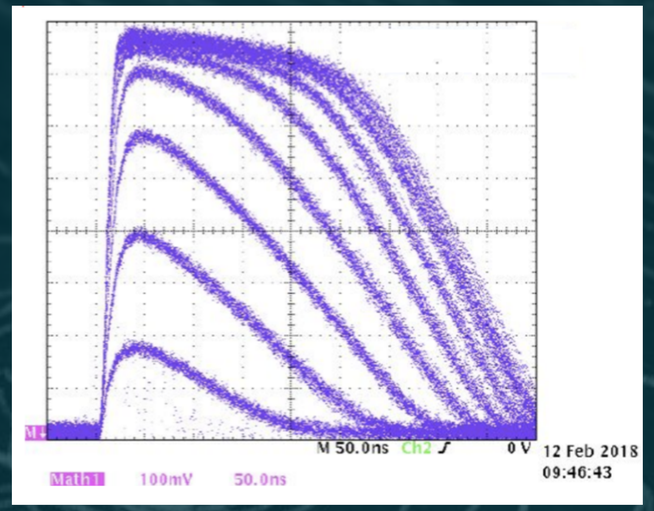
Hit Rates

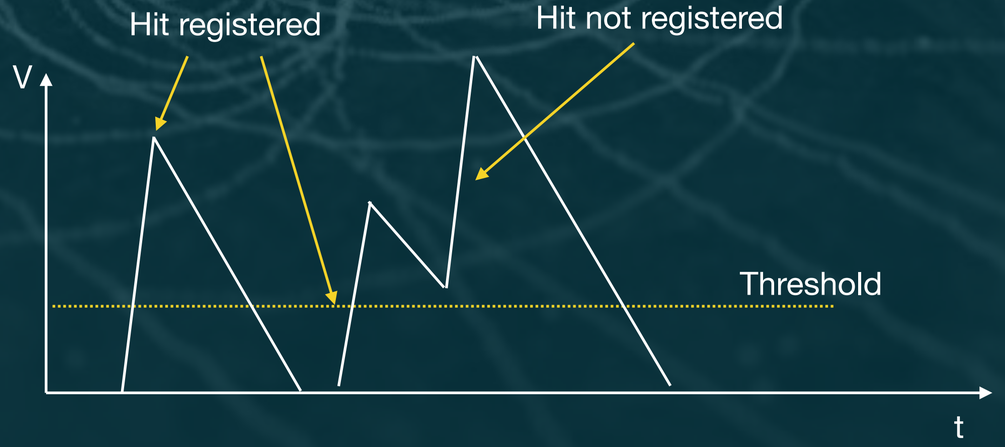
- In HL-LHC conditions hit rate is 3GHz/cm2 (innermost layer)
  - This translates to 75 hits per bunch crossing (25ns)
  - For a 50um by 50um pixel detector that results in 0.2% occupancy
  - Average return to baseline is around 150ns (6 bunch crossings) -> 1.2% of channels “busy” at any given time
- MuC conditions rather different
  - 3000 clusters/bx/cm2 (bunch crossing = 10us), average cluster size ~12 (dominated by BIB) -> 36k hits -> 3.6GHz/cm2
    - But should actually consider all 36k hits occur within 15ns!
    - For a 25um by 25um pixel detector that results in 22% occupancy!!
  - Assuming we can ignore everything except 90ps around t0 -> 200 clusters/bx/cm2 -> 2.4k hits -> 1.5% occupancy
    - Timing cuts are applied after processing, detector “sees” all hits!



In-Pixel Pileup

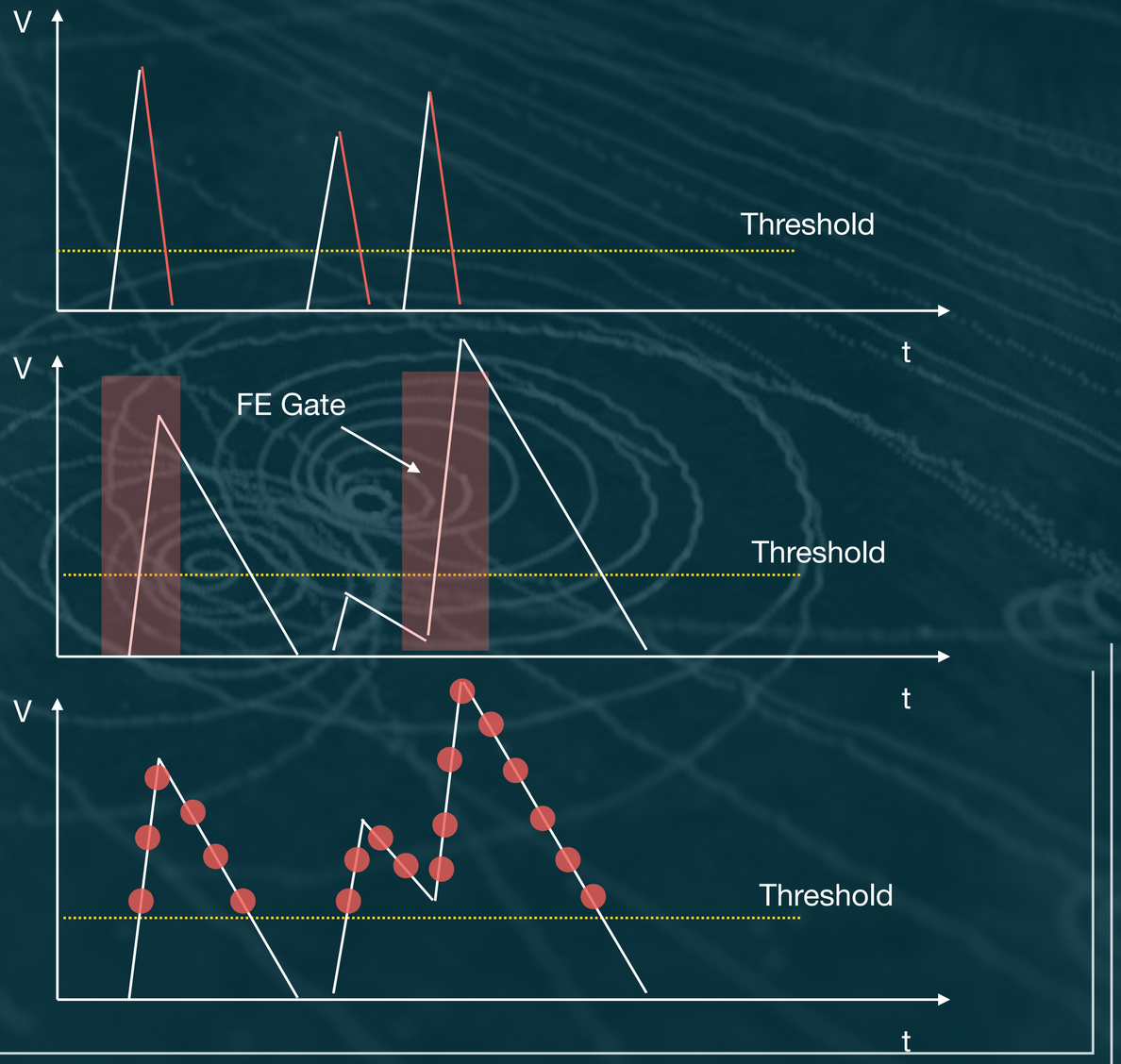
- For HL-LHC readout chip have to consider return to baseline of signal
  - Constant reset front-end architecture allows Time-over-Threshold (ToT) to be proportional to charge. ToT counted with main 40MHz clock (25ns resolution)
  - Hits are registered when they pass above threshold, hits that occur in a pixel that has not returned below threshold are lost. This is called “in-pixel pileup”
- While timing cuts in MuC conditions cut down heavily on overall occupancy, detector still sees all hits
  - In-pixel pileup will lead to reduced efficiency and could muddle timing precision we can achieve
  - I.e. MuC conditions need to be accommodated at the earliest readout chip stage!
  - The total amount of energy deposited within a very small time frame might also have adverse affects on sensor efficiency





In-Pixel Pileup II

- What to do?
  - Natural instinct might be to make return to baseline faster
    - 4D pixel detector need very fast front-ends to capture time accurately (fast rise time), this naturally shortens the ToT but not to the degree required to cut down on in-pixel pileup
    - Shorter ToT also requires faster clocks to count ToT (more power) and will likely negatively affect charge resolution
    - Time-of-Arrival (ToA) uses ToT to perform time walk correction -> worse ToT -> worse ToA?
  - Could "gate" front-end (e.g. modulate pre-amp bias)
    - Major architectural change compared to the past -> R&D necessary
  - Change digitization fundamentally?
    - Full waveform sampling -> incredibly demanding in terms of power/ noise/space, but gives all information
    - Increased information density could be processed on-chip with ML techniques -> major advancements in this field in recent decade
      - General issue with ML techniques: how to reduce risk when ML architecture is fixed in-silicon?



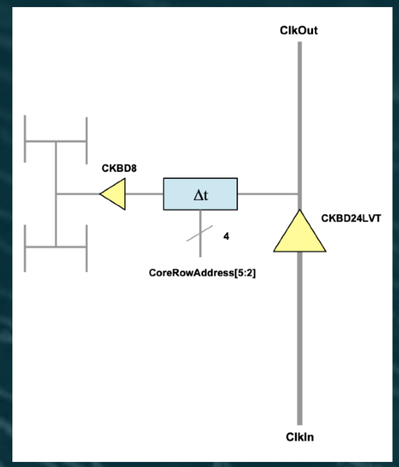
“Triggering” and Timing

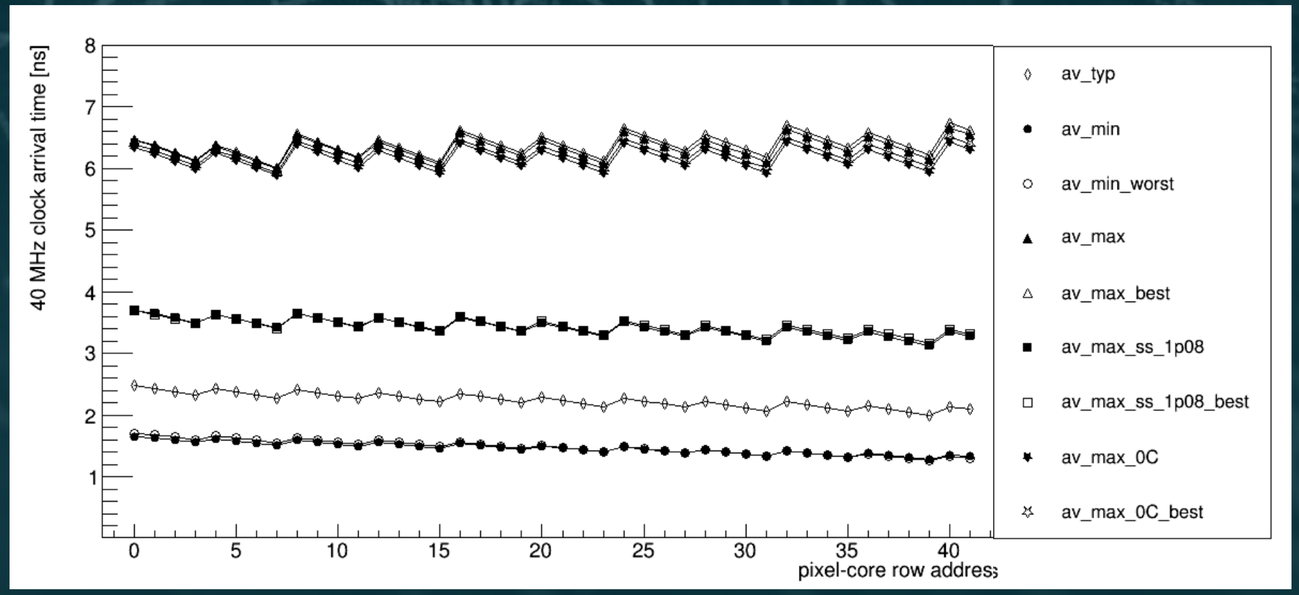
- HL-LHC has high repetition rate, so even though hits per bx are much lower, average hit rate is similar to MuC environment (~3 GHz/cm2)
  - Raw data bandwidth required to transmit everything approx 60 Gbps/cm2 (120 Gbps with timing information)-> Need to reduce data amount!
  - For HL-LHC use triggered readout at ~1/40 (1MHz) of the original bx rate -> Not suitable to MuC with 100kHz rep rate, want “streaming” readout for best physics efficiency
    - Triggering different bunch crossings (in order) is a (relatively) simple task
  - Can we read everything out in on bx? If not need some amount of memory!
- Need timing cuts to cut down on hit multiplicity … on chip picosecond level timing cuts?
  - Side effect: don’t need to transmit ToA off detector with large dynamic range
  - Assuming 90ps timing cut data rate is approx 5Gbps/cm2 -> manageable! (Factor ~5 difference to HL-LHC)
  - How to accommodate time-walk? On-chip time walk correction?
- Are on-chip timing cuts to this level realistic (or what R&D needs to be done to get there)?

On-chip Timing

- How to achieve 10 picosecond level synchronization on-chip?
  - Cannot use typical clock distribution trees due to pixel column architecture
  - RD53 type chips synchronize the 40MHz clock across all pixels via custom delays to the 100ps level
  - Critically also needs to be applied to calibration signals!
- Delay cells highly dependent on environmental conditions: temperature, radiation, powering, process corners
  - Is it possible to push this down further?
  - Do the tools allow us to do this routinely during final chip layout assembly?

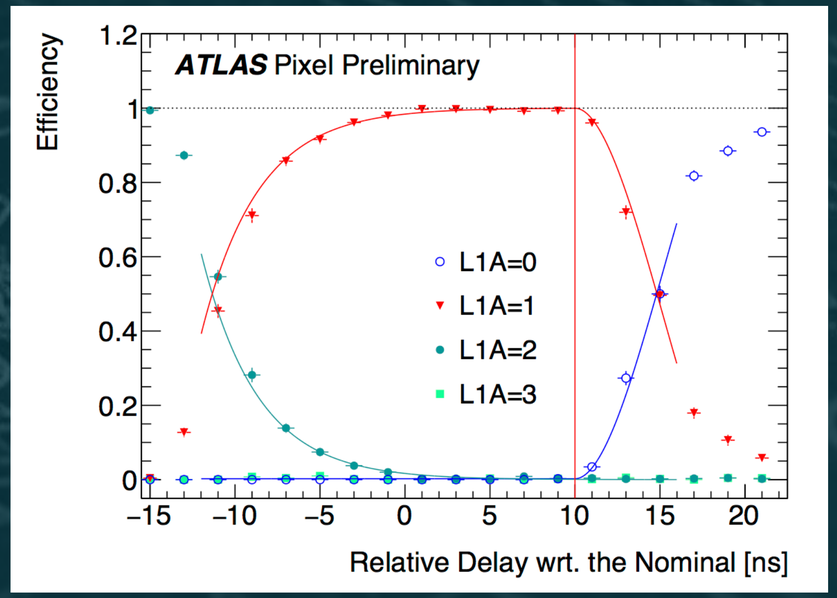
- 115 ps RMS in TYP corner
- 95 ps RMS in MIN corner
- 166 ps RMS in MAX corner





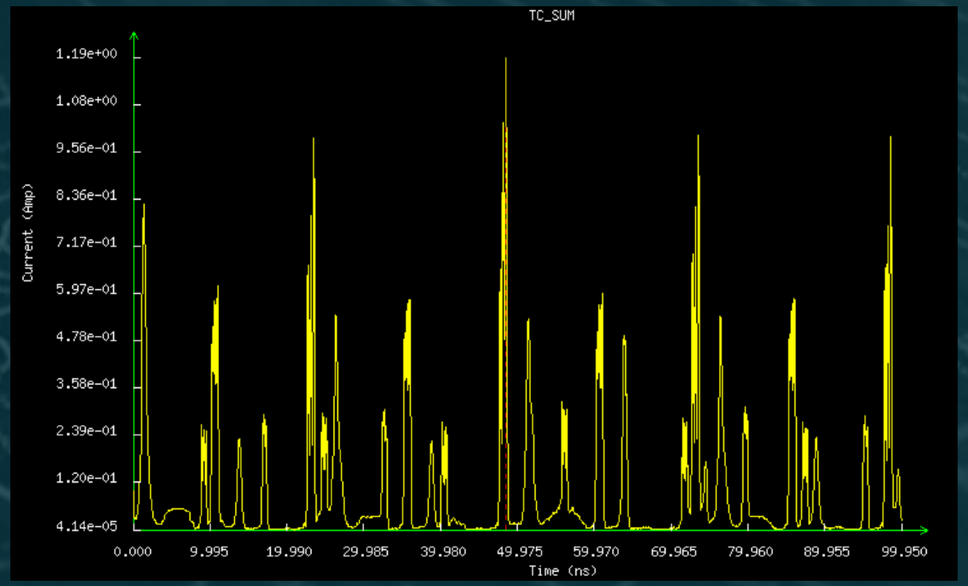
System Synchronization

- Current detectors timed in with collision events by accommodating varying clock skew of each detector element with local delays
  - Is this still possible at the MuC?
- Generally speaking though we have been avoiding “global signaling” and try and rely more on “local signaling” to improve reliability during operation
  - Example: localized trigger tagging overcomes issue of keeping bx counters synchronized
- Could we determine t0 on-chip dynamically?
  - If we have discriminators other than timing could use those to find timing window: cluster size/shape, charge, double layer stub angle, …
  - Could this be applied even within one chip to reduce on-chip timing constraints?



Timing and Power

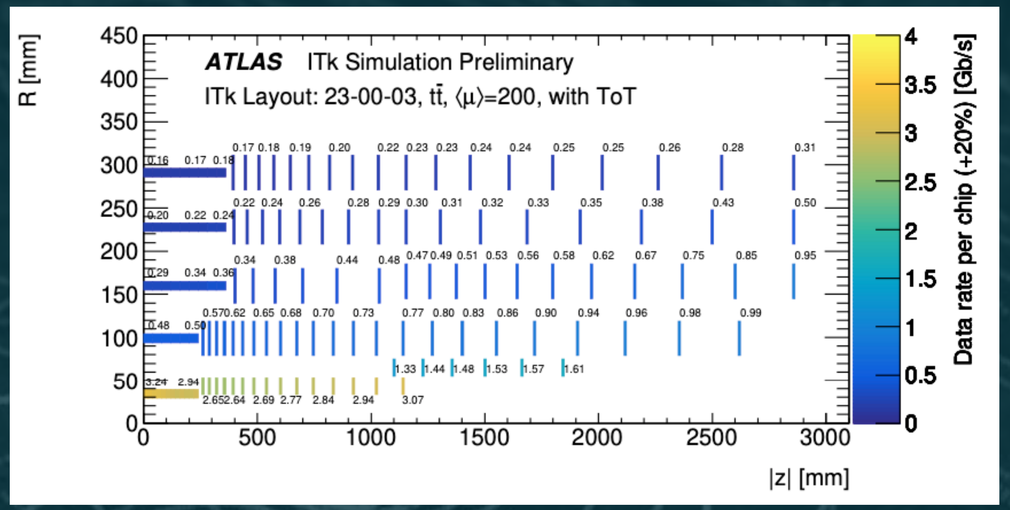
- In RD53 chips hit and trigger activity induced digital current at 3GHz/cm2 hit rate and 1MHz trigger is 23% of total current (~5uW per pixel)
  - This is by design! Total power is reduced by only clocking those parts which are actively processing!
- In MuC environment to reject hits with a timing cut, need to measure time of ALL hits
  - Low Power TDC architecture critical to achieve this!
  - For comparison ALTIROC (ATLAS timing layer front-end chip) which delivers 30ps time resolution needs 4mW per channel*
- Considering 22% occupancy in 15ns window: activity driven power reduction not sufficient anymore?
  - Cluster timing? Early abort TDC?
  - Somehow utilize long “dead” time and stager measurements?

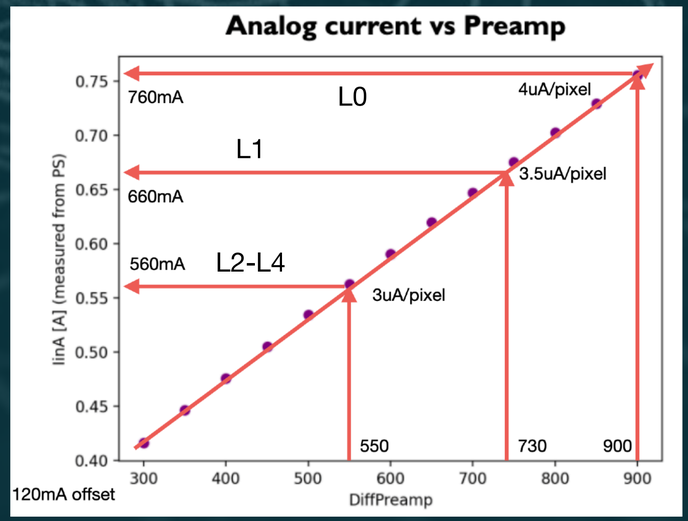


*not apples to apples comparison

Data Bandwidth

- Vertex detector has to cover large spectrum in multiplicity, at HL-LHC already and even more pronounced at MuC
  - MuD Layer 1/2 hit multiplicity ratio factor ~100
- Innermost layers covers small area and space points are important for vertexing -> worth to splurge on mass wise, can accommodate services for a high data rate, cooling for higher power
- Outer layers cover larger and larger area, sees much lower hit rates and less radiation -> want it to be as light-weight as possible
- Considering the cost and effort to develop pixel readout chip technology, not practical too have multiple flavors. Scaling needs to be build in from the start!
  - Need to foresee power scaling both in analog and digital
  - Mass dominated by data services: on-chip data link aggregation to efficiently utilize every cable/link
  - Compression algorithms need to cover all domains





Requirements Summary

BERKELEY LAB

	muC Tracker			ATLAS ITk	
	Vertex Detector	Inner Tracker	Outer Tracker	Pixel	Strips
Resolution [um x um]	25x25	50x1,000	50x10,000	50x50	75x2500
Channels	1200M	290M	170M	5000M	60M
Area [m^2]	0.75	14.5	85	13	165
Double Layer Spacing [mm]	2mm			N/A	5
Total Ionizing Dose [Mrad]*	200	10		1,000	75
Fluence [1MeV neq/cm^2]*	$3x10^{15}$	$1x10^{16}$	$<1x10^{15}$	$2x10^{16}$	$2x10^{15}$
Time resolution	30ps	60ps		25ns (1.5ns)	25ns
Hit density [mm-2]	3.7**	0.5**	0.03**	0.6	0.003
Collision rate	100kHz			40MHz	
Readout percentage	100% (trigger-less)			2.5% (triggered)	
Data Bandwidth	~30Tbps**			13.5Tbps	

*assume 10 year run-time **after ns timing cuts