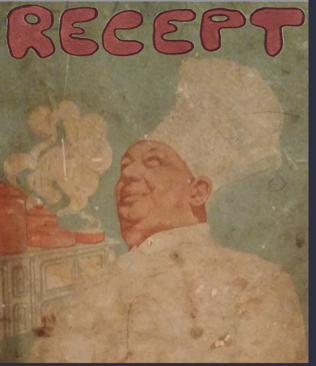


Future LHCb upgrades
Scenarios for Run 5

Vladimir V. Gligorov, CNRS/LPNHE

6th Workshop on LHCb Upgrade 2

Barcelona, 31.03.2023

Scoping is important

Steve Jobs at the 1997 Apple WWDC meeting

“We had to decide what were the fundamental directions we were going in, what makes sense and what doesn’t. And there were a bunch of things that didn’t. Microcosmically they might have made sense, macrocosmically they made no sense. And the hardest thing is, when you think about focusing… focusing is about saying no. And when you say no you piss off people.”

Scoping is important

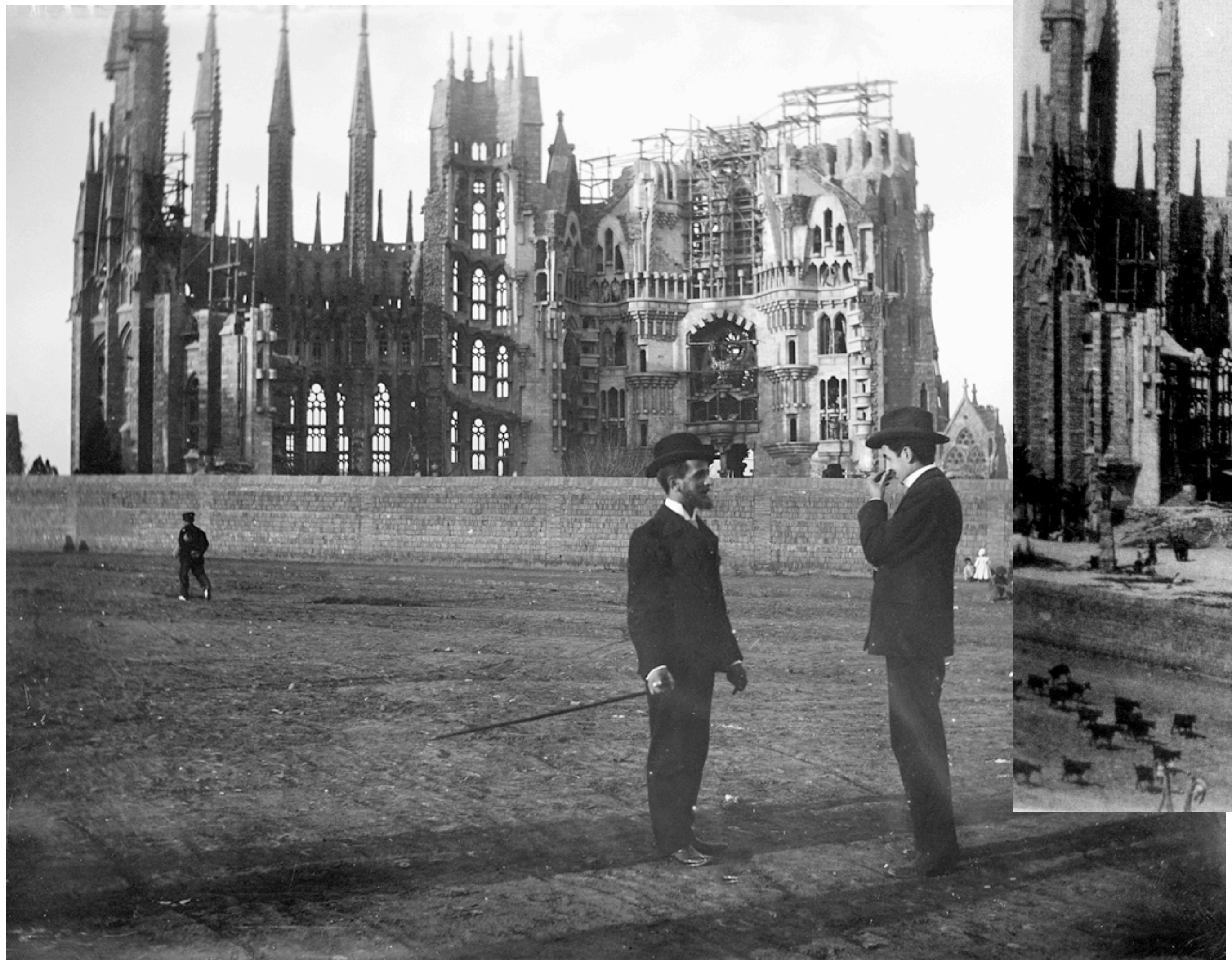
Steve Jobs at the 1997 Apple WWDC meeting

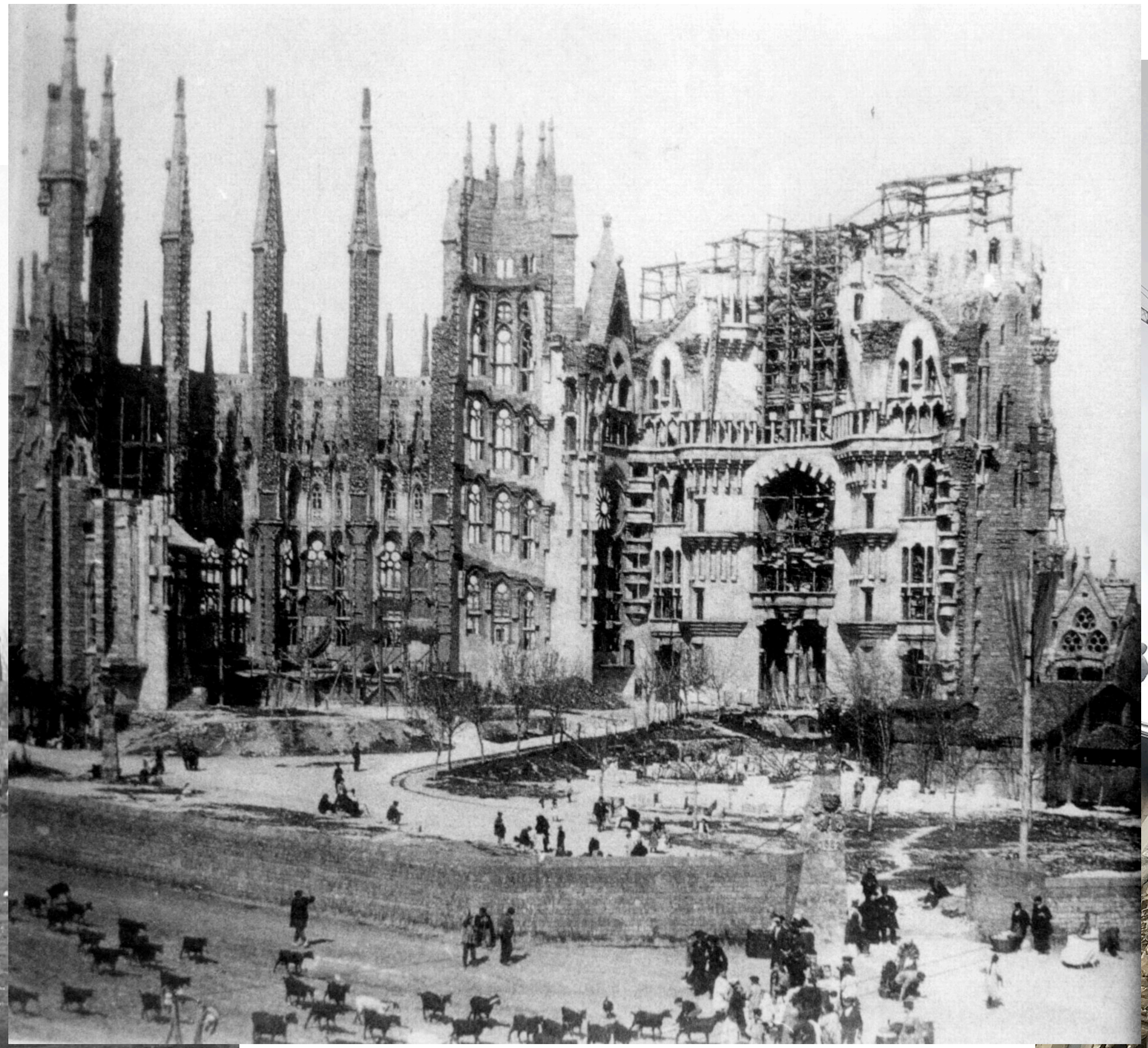
"We had to decide what were the fundamental directions we were going in, what makes sense and what doesn't. And there were a bunch of things that didn't. Microcosmically they might have made sense, macrocosmically they made no sense. And the hardest thing is, when you think about focusing… focusing is about saying no. And when you say no you piss off people."

Scoping is important

My reinterpretation of Steve Jobs's quote in our context: we need to decide not only whether we are giving the right answers, but whether we are asking the right questions.

Focusing means not discussing money





Sagrada Familia consumed Gaudi's life and became a constant struggle, which perhaps he welcomed (atonement requires pain and struggle, after all). Gaudi's funding from patrons for Sagrada Familia was limited and as the project progressed, Gaudi literally went door to door trying to raise money for his project. Apparently there are members of the Catholic Church on the Sagrada Familia board now, but Gaudi got no funds from the Church for construction.

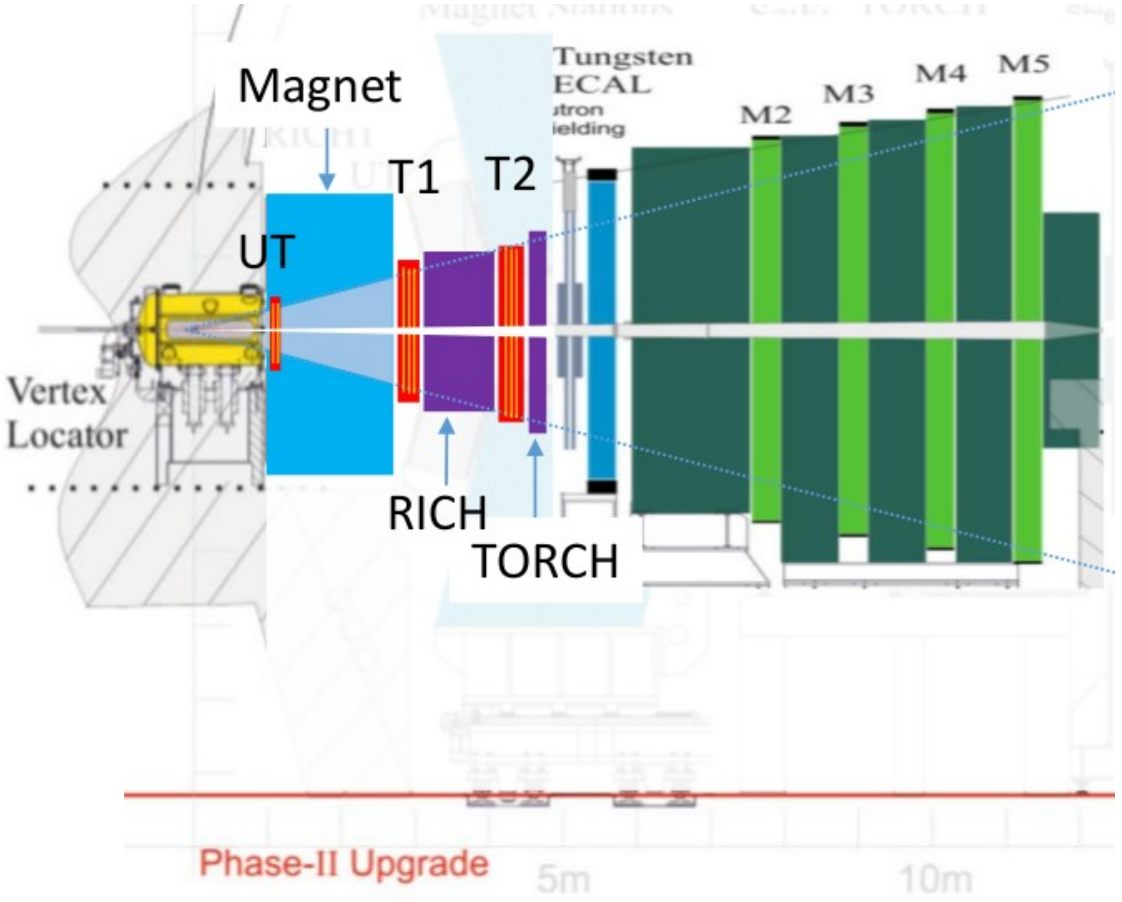
Nobody cultured would ask how much the Segrada Familia costs. When it will be finished is more relevant…

Do we believe in our physics case?

Observable	Current LHCb	Upgrade I		Upgrade II
	(up to 9 fb^{-1})	(23 fb^{-1})	(50 fb^{-1})	(300 fb^{-1})
CKM tests				
γ ($B \to DK$, *etc.*)	4° [9,10]	1.5°	1°	0.35°
ϕ_s ($B^0_s \to J/\psi\phi$)	32 mrad [8]	14 mrad	10 mrad	4 mrad
$\lvert V_{ub}\rvert/\lvert V_{cb}\rvert$ ($\Lambda^0_b \to p\mu^-\overline{\nu}_\mu$, *etc.*)	6% [29,30]	3%	2%	1%
a^d_{sl} ($B^0 \to D^-\mu^+\nu_\mu$)	36×10^{-4} [34]	8×10^{-4}	5×10^{-4}	2×10^{-4}
a^s_{sl} ($B^0_s \to D^-_s\mu^+\nu_\mu$)	33×10^{-4} [35]	10×10^{-4}	7×10^{-4}	3×10^{-4}
Charm				
ΔA_{CP} ($D^0 \to K^+K^-, \pi^+\pi^-$)	29×10^{-5} [5]	13×10^{-5}	8×10^{-5}	3.3×10^{-5}
A_Γ ($D^0 \to K^+K^-, \pi^+\pi^-$)	11×10^{-5} [38]	5×10^{-5}	3.2×10^{-5}	1.2×10^{-5}
Δx ($D^0 \to K^0_S\pi^+\pi^-$)	18×10^{-5} [37]	6.3×10^{-5}	4.1×10^{-5}	1.6×10^{-5}
Rare Decays				
$\mathcal{B}(B^0 \to \mu^+\mu^-)/\mathcal{B}(B^0_s \to \mu^+\mu^-)$	69% [40,41]	41%	27%	11%
$S_{\mu\mu}$ ($B^0_s \to \mu^+\mu^-$)	—	—	—	0.2
$A^{(2)}_{\mathrm{T}}$ ($B^0 \to K^{*0}e^+e^-$)	0.10 [52]	0.060	0.043	0.016
$A^{\mathrm{Im}}_{\mathrm{T}}$ ($B^0 \to K^{*0}e^+e^-$)	0.10 [52]	0.060	0.043	0.016
$\mathcal{A}^{\Delta\Gamma}_{\phi\gamma}(B^0_s \to \phi\gamma)$	$^{+0.41}_{-0.44}$ [51]	0.124	0.083	0.033
$S_{\phi\gamma}(B^0_s \to \phi\gamma)$	0.32 [51]	0.093	0.062	0.025
$\alpha_\gamma(\Lambda^0_b \to \Lambda\gamma)$	$^{+0.17}_{-0.29}$ [53]	0.148	0.097	0.038
Lepton Universality Tests				
R_K ($B^+ \to K^+\ell^+\ell^-$)	0.044 [12]	0.025	0.017	0.007
R_{K^*} ($B^0 \to K^{*0}\ell^+\ell^-$)	0.12 [61]	0.034	0.022	0.009
$R(D^*)$ ($B^0 \to D^{*-}\ell^+\nu_\ell$)	0.026 [62,64]	0.007	0.005	0.002

On paper the best physics case of any detector likely to be built or operated in my working life
But we are proposing to go to precisions nobody has ever gone to, across a broader range of observables, experimental techniques, and in a tougher environment than any experiment in the history of our field
This is a 25+ year commitment — requires full effort from all, *and* convincing colleagues to join

Are we building the right detector?



Factor 6 improvement from luminosity scaling is at a psychological limit

If people keep wondering whether we can do better, there is a reason

We are scientists — conduct a rapid, concrete, and tough evaluation of whether this is feasible and what the physics selling point is to justify it

In parallel work to quantify and prove potential gains beyond lumi scaling to increase our collective motivation in the baseline layout

Timescale for these evaluations should be some months, not more

Can we install and commission it?

- U1: 5 years planned from 2014 TDR→ Needed 7 + 1 Covid year delay in reality

SciFi

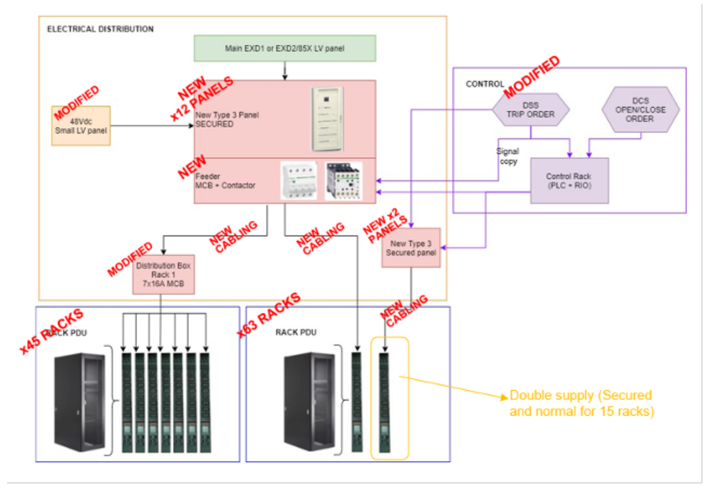
- Fibre and Mat R&D 2023-2025, needed for descoping decision
  - Irradiation studies and Test-beams in 2024 and 2025 on any new mats.
- Cryo-feasibility review end of 2023
- Cryo-box demonstrator and fibre feedthrough in 2025 for descoping/TDR

- Mat production to start 2027 until mid of 2029
  - Winding machines and tooling to be designed and ordered end of 2025
  - possible also produce clear fibre ribbons
  - need a final fibre-envelope and interface defined within the module
  - Prototypes of mats and tooling end-2026
- Module production early 2028 to end of 2029
  - Need a final design by mid 2026 to place tenders for components. Most will require EU+UK wide tenders.
- Detector Frames assembled in 2029 prototype to end of 2031 production
  - C-Frame Mechanics design by mid 2028 for a prototype in early 2029, tendering and delivery of production late 2029.

- Assembly should be complete at the beginning of LS4

Infrastructure consolidation in LS3

- Backend of the Detector Safety System (DSS) to be replaced (common project for all LHC experiments).
- SNIFFER system to be upgraded/replaced. Different technical solutions being investigated.
- Obsolete LV distribution boards (Hazemeyer TDMs) feeding racks in D1 – D3 and B1 (inherited from LEP) need to be replaced.
- At the same time, improve the granularity of the distribution downstream of the TDM panels (one feeder per rack) and adapt to new D1/D2 rack layout.
- Also on our wish list: dedicated Diesel generator set for critical loads (*e. g.* cooling systems).

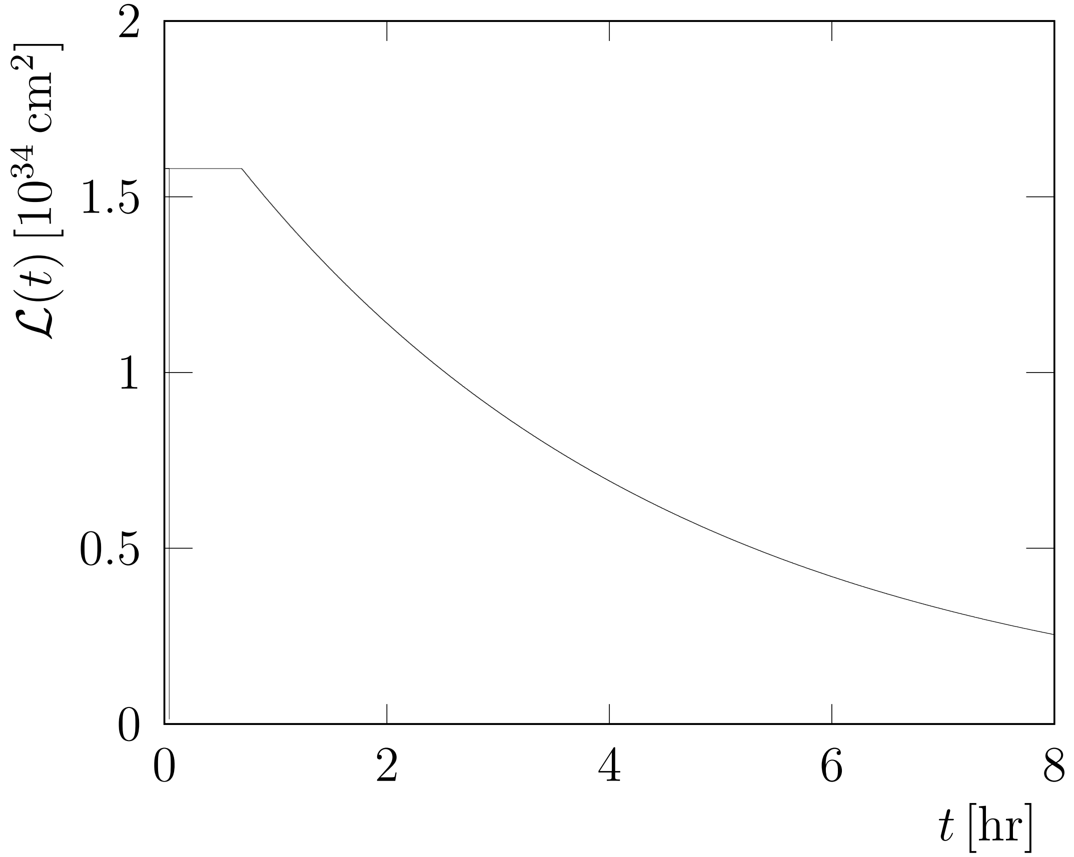


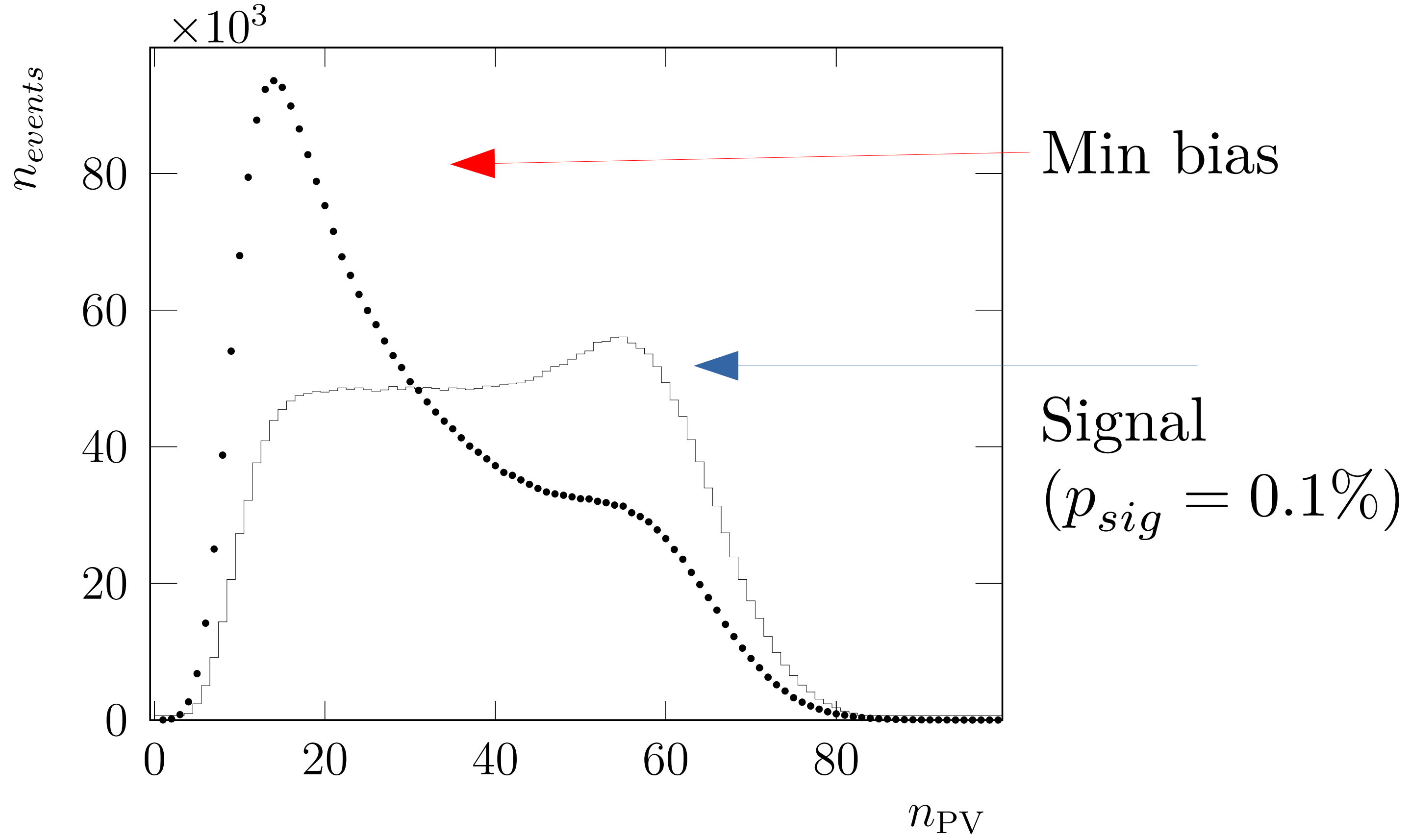
2022 showed us how difficult it is to start up a new detector, despite a tremendous effort by the subdetector, online, and software projects

We cannot afford to lose even 1 year of Run 5 for commissioning if we want to achieve our physics aims. We must be internally honest on this.

Frontloading infrastructure work seems critical to give a chance of success. Insist on proper contingencies for commissioning *before* Run 5!

Can we really avoid systematic limits?



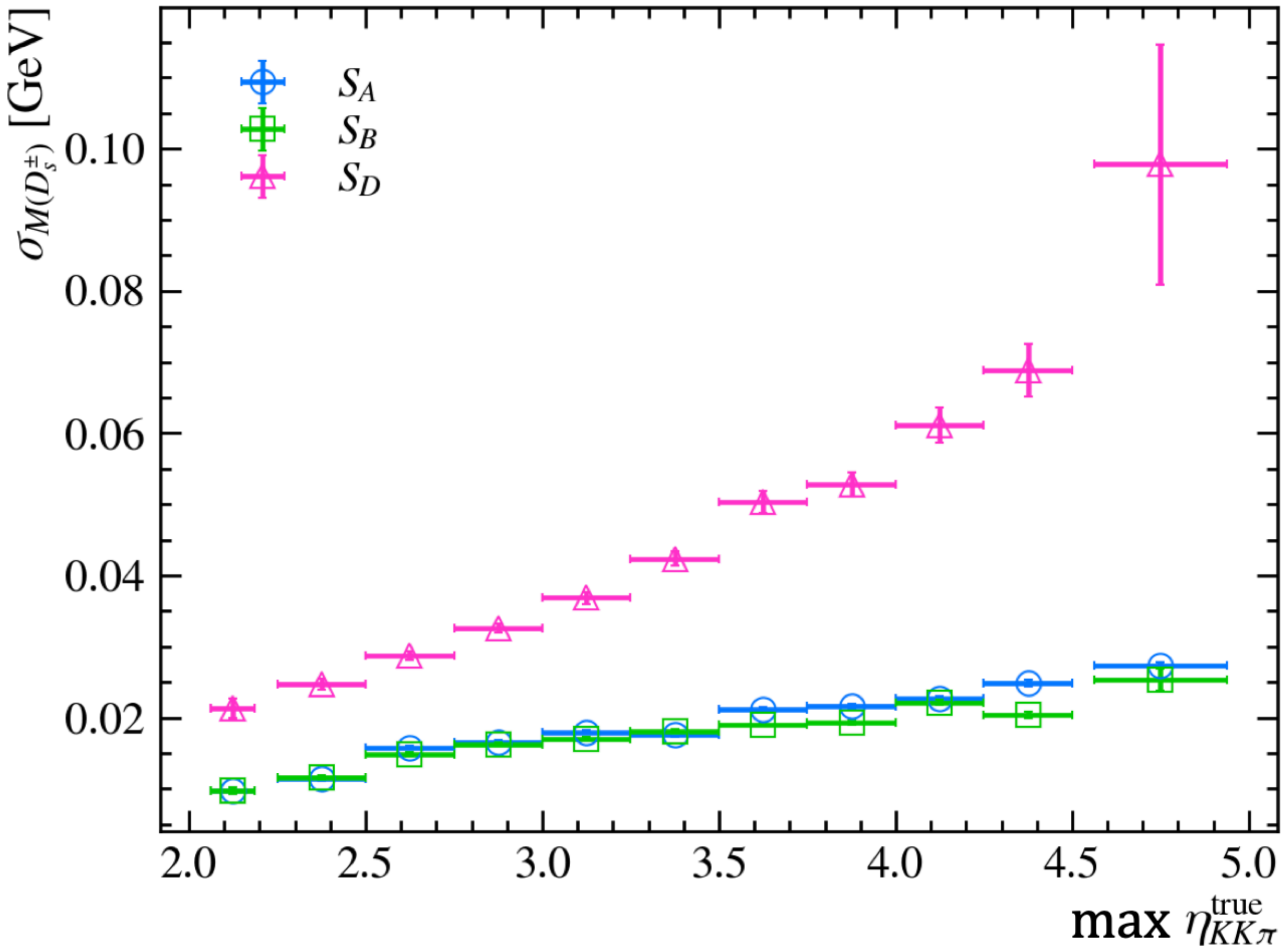


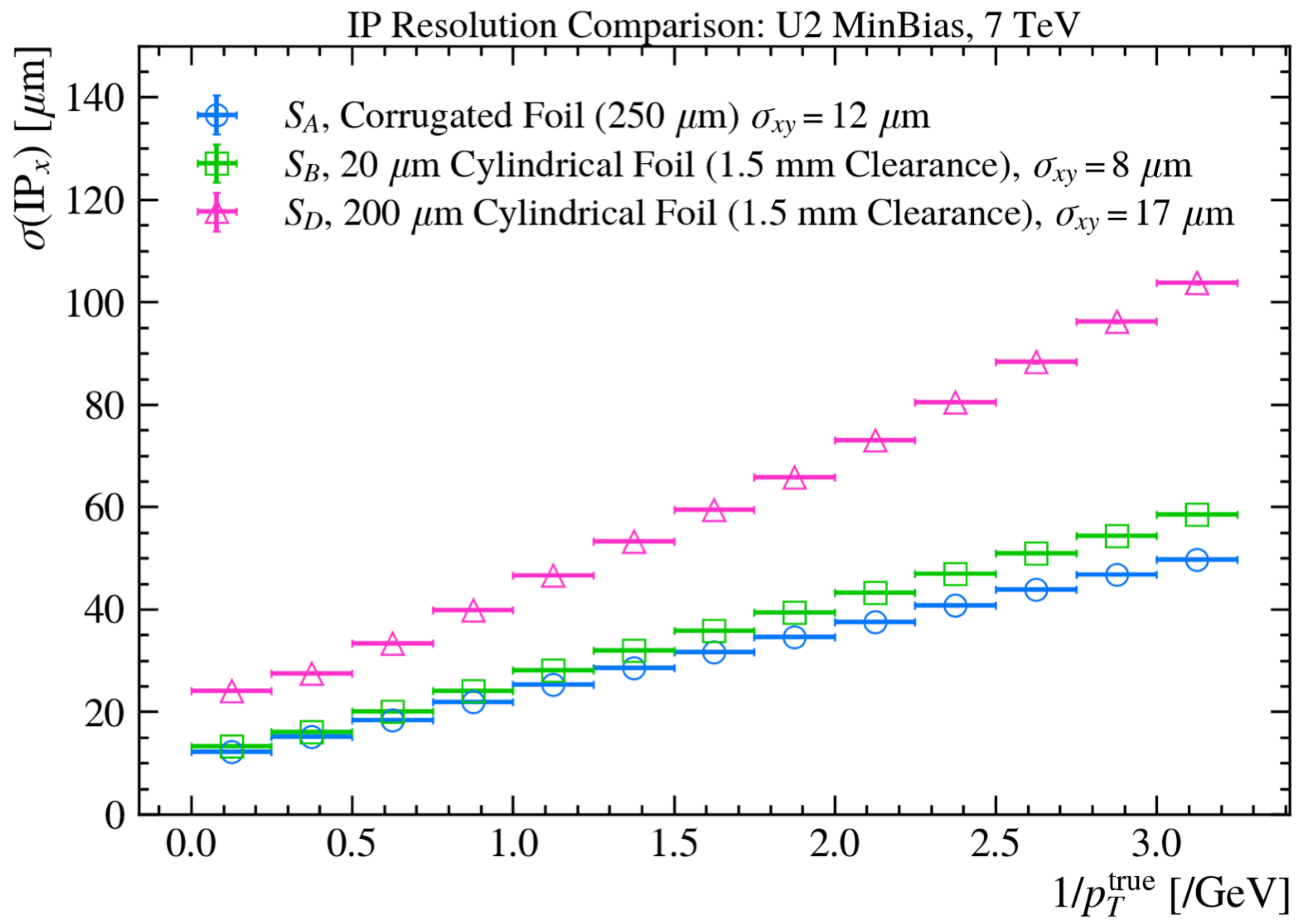
We know our data-simulation corrections only work in the limit of small differences

Time dependence of detector performance with varying occupancy is not trivial

We should do the equivalent of pre-Run-1 misaligned simulation studies (also for PID and so on) to prove we can correct these effects with the U2 precisions needed

We should build the best VELO possible



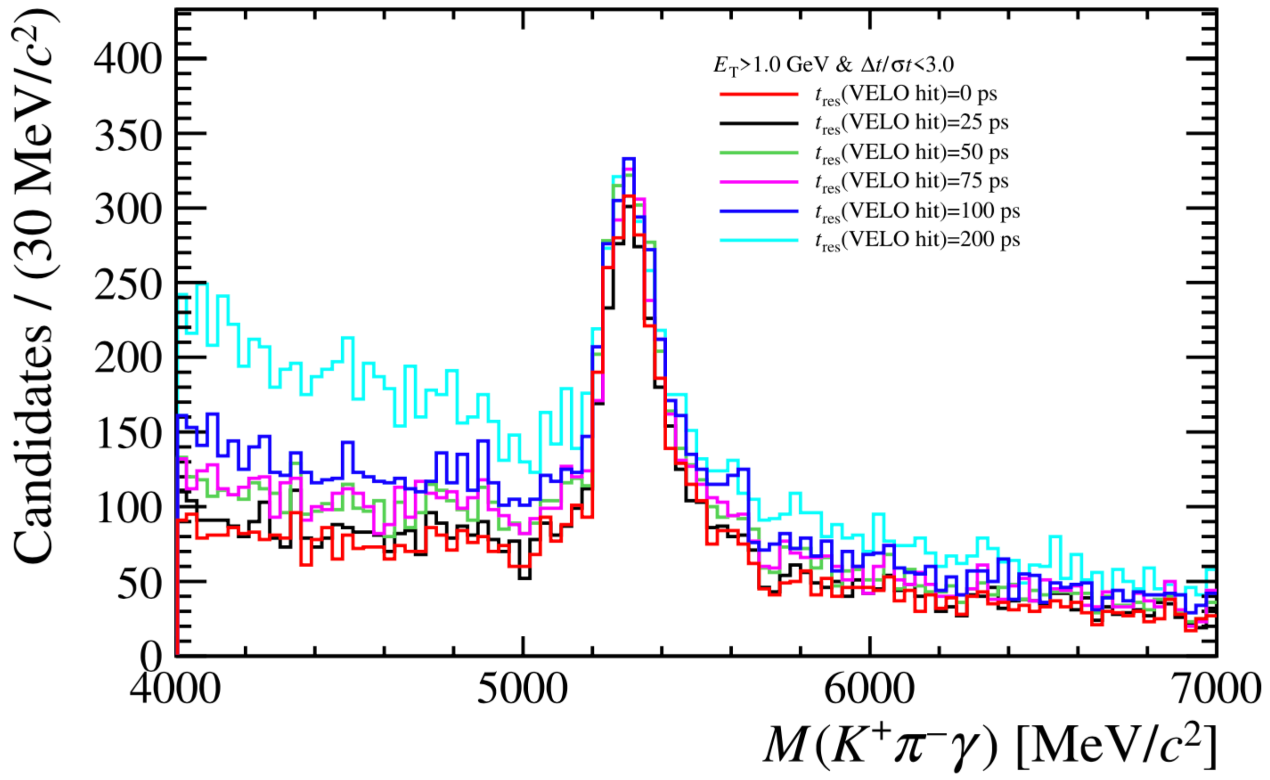


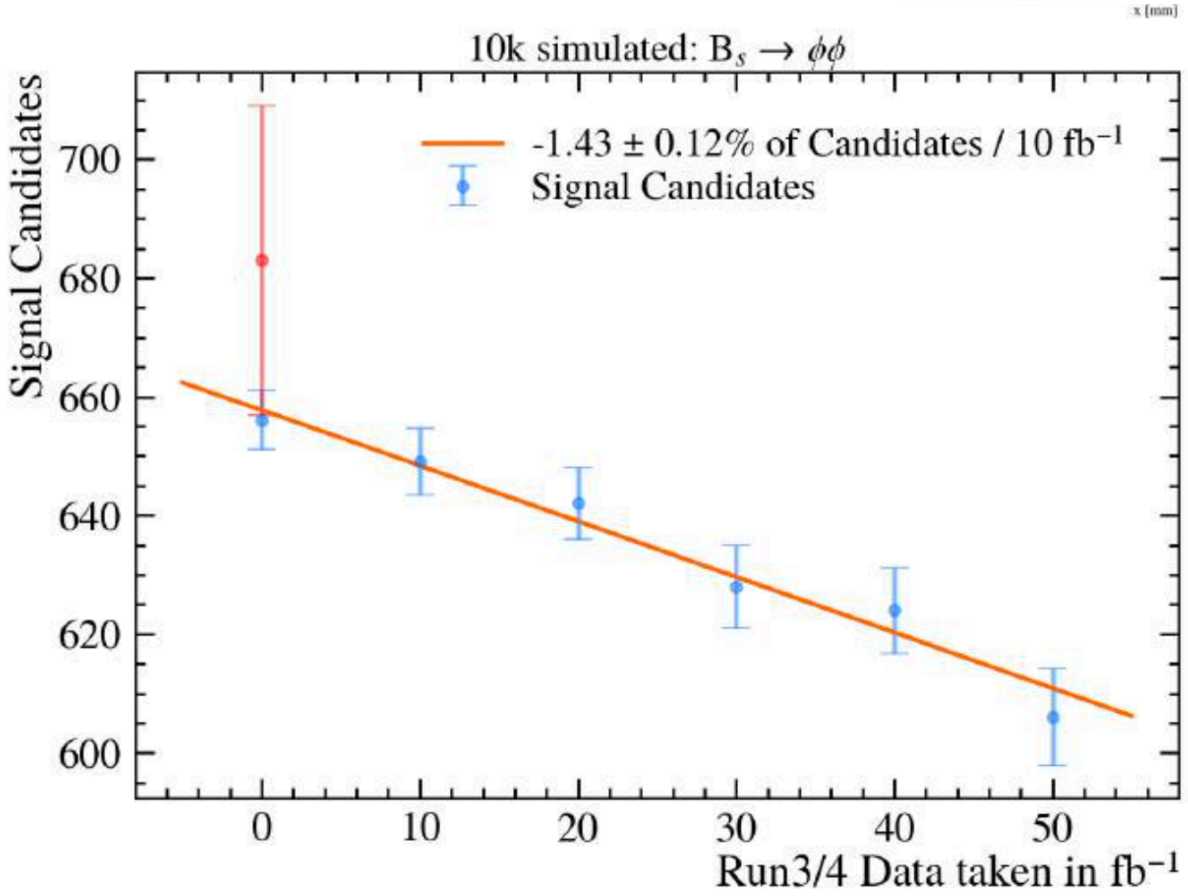
Huge work done by the VELO team to study descoped scenarios

However our vertex detector is the single most important driver of both statistical sensitivity and systematic uncertainties: we should not talk about descoping it

Parallel R&D on the various technological aspects should allow to go beyond FTDR performances (see Scenario X), maybe in another few years can do even better?

The tracker is a coherent object





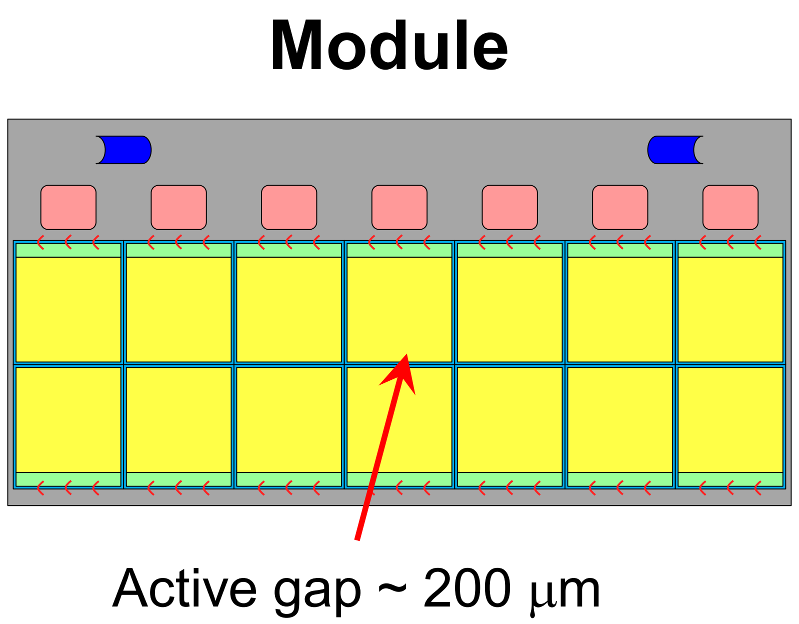
Many important studies across different areas of the tracking system: UT, MT, MS

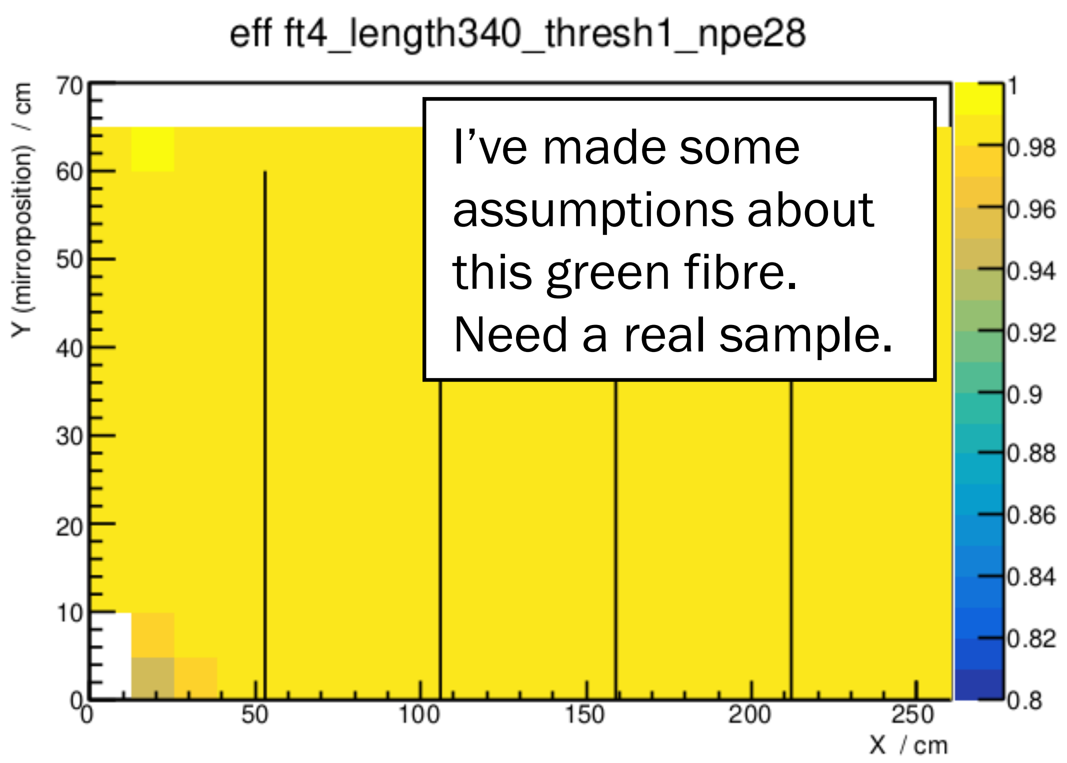
Crucial to move towards a coherent global optimization of the tracking system

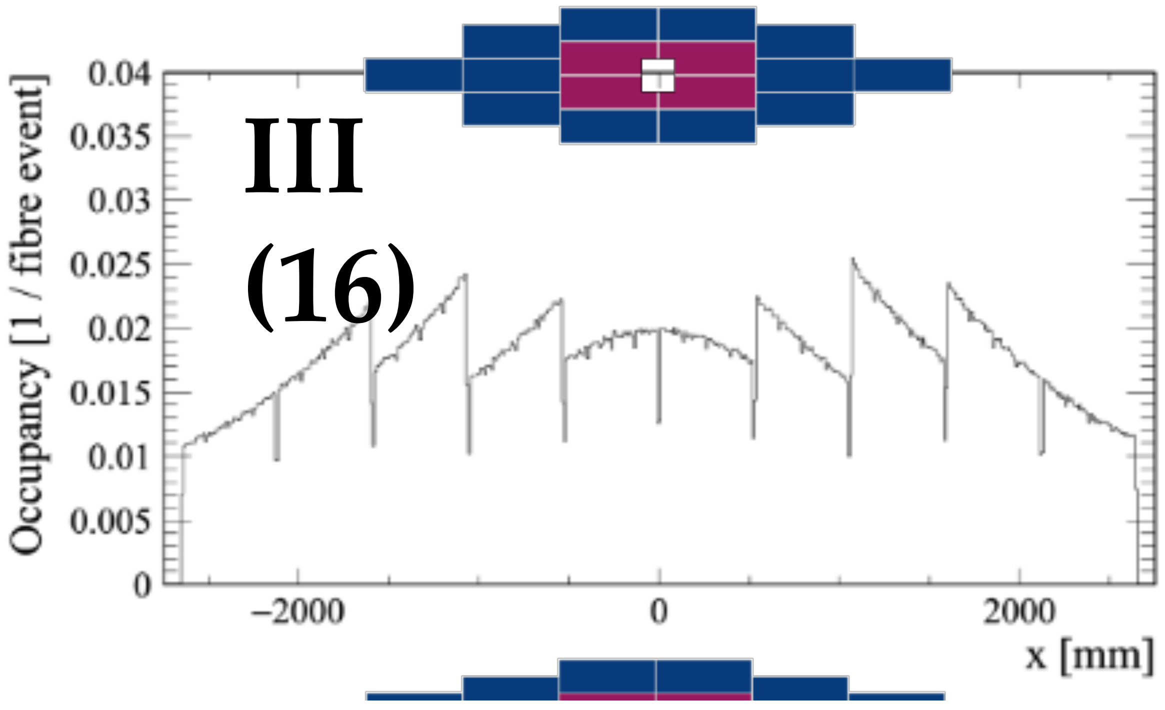
Optimize for both pp, SMOG, and Ion physics programmes from the beginning

Evaluate both statistical reach and how well we will be able to understand and calibrate this object after it has absorbed 300 fb^{-1} of radiation damage

How many hits is enough hits?







A naïve calculation

- ~1.4% inefficiency / plane / track
- ~4.2% for 3 hits out of 3 planes
- ~25% for all 6 tracks
- ~21% for all 5 tracks

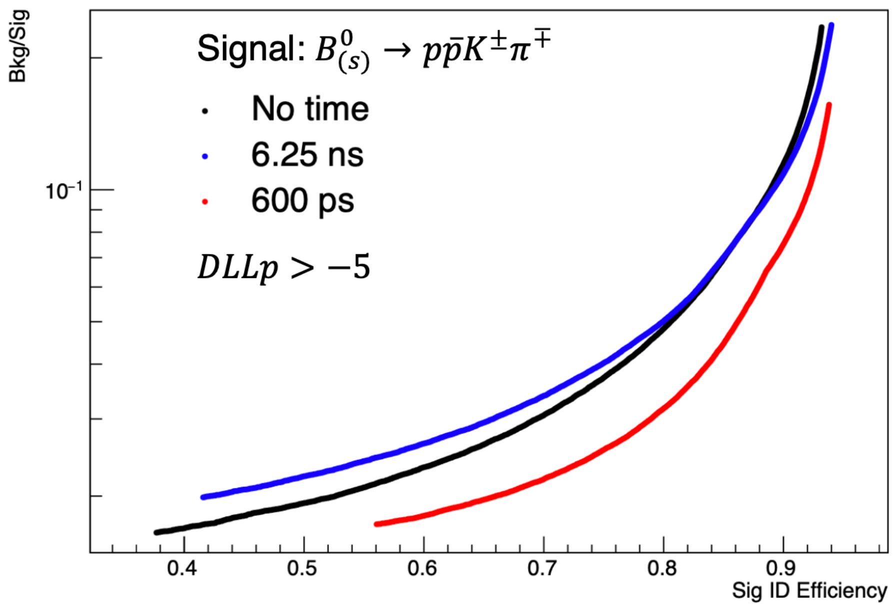
Fewer tracking layers = less material = less cost = good thing

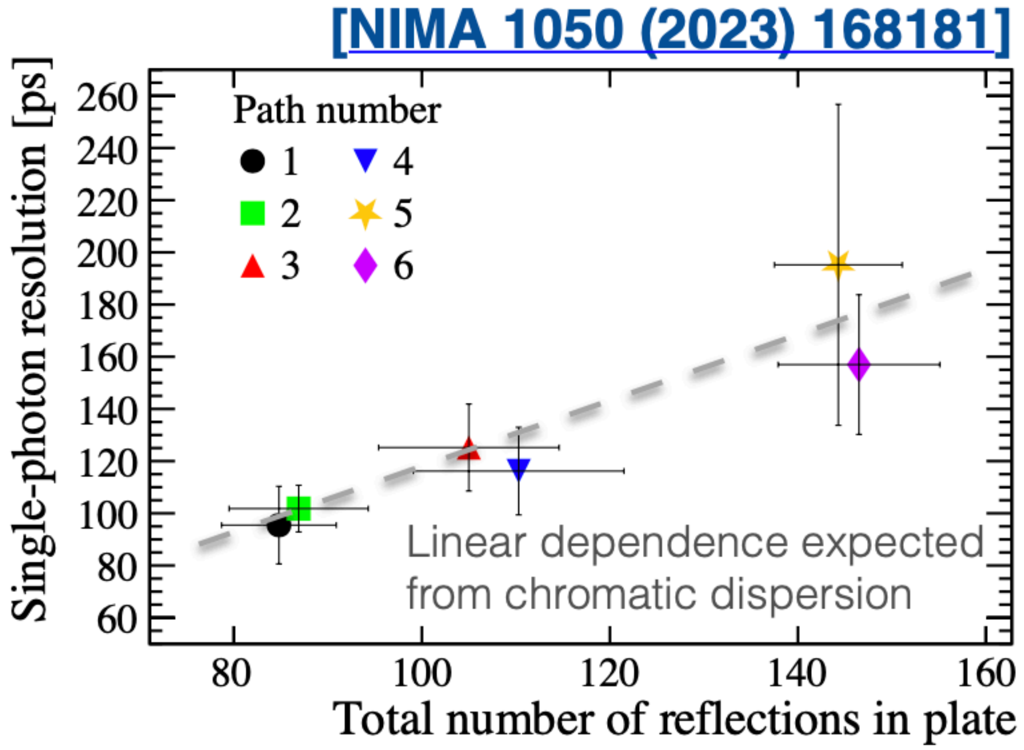
Fewer tracking layers = less robustness/redundancy = bad thing

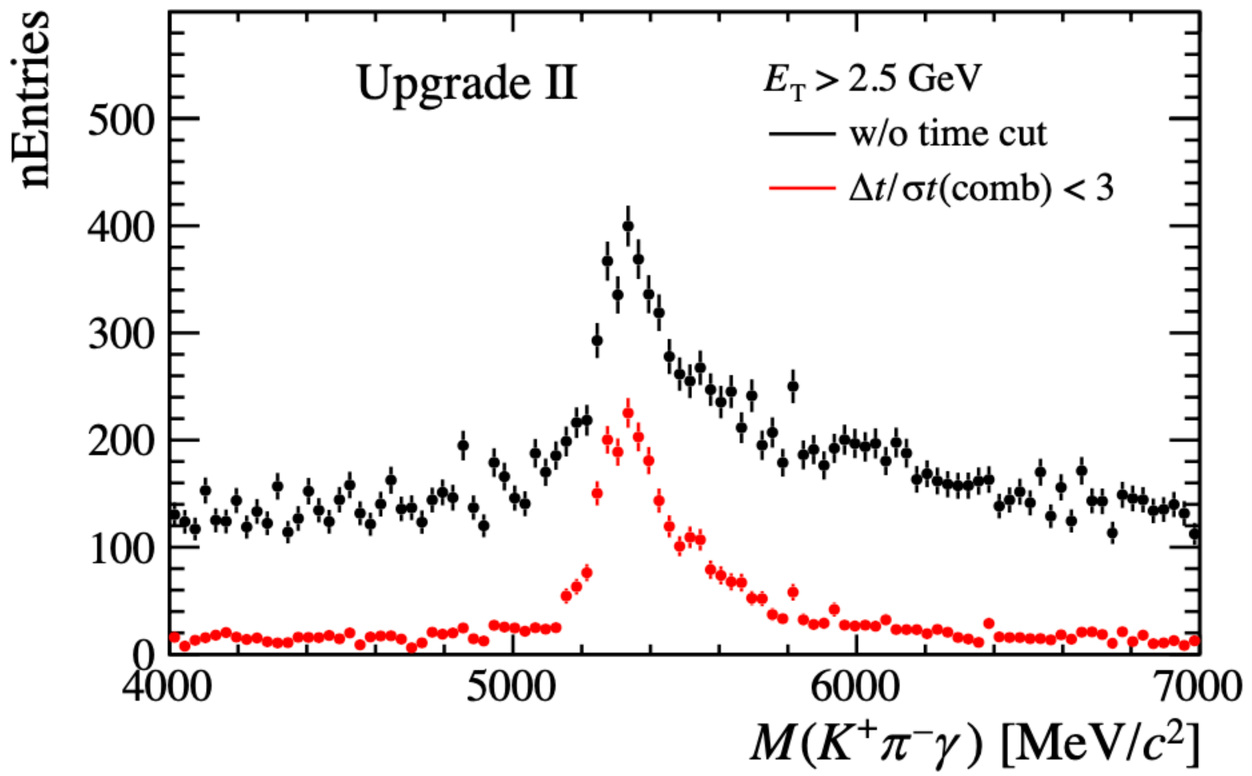
Go beyond hit efficiency and occupancy and understand impact of choices on the resolution of our track parameters for all track types. Resolution of track slopes at CALO/RICH is vital to our overall performance.

If we build the magnet stations then VELO + UT + MS and UT + MS (?) become new track categories to optimize for as well.

7 (pico)seconds away





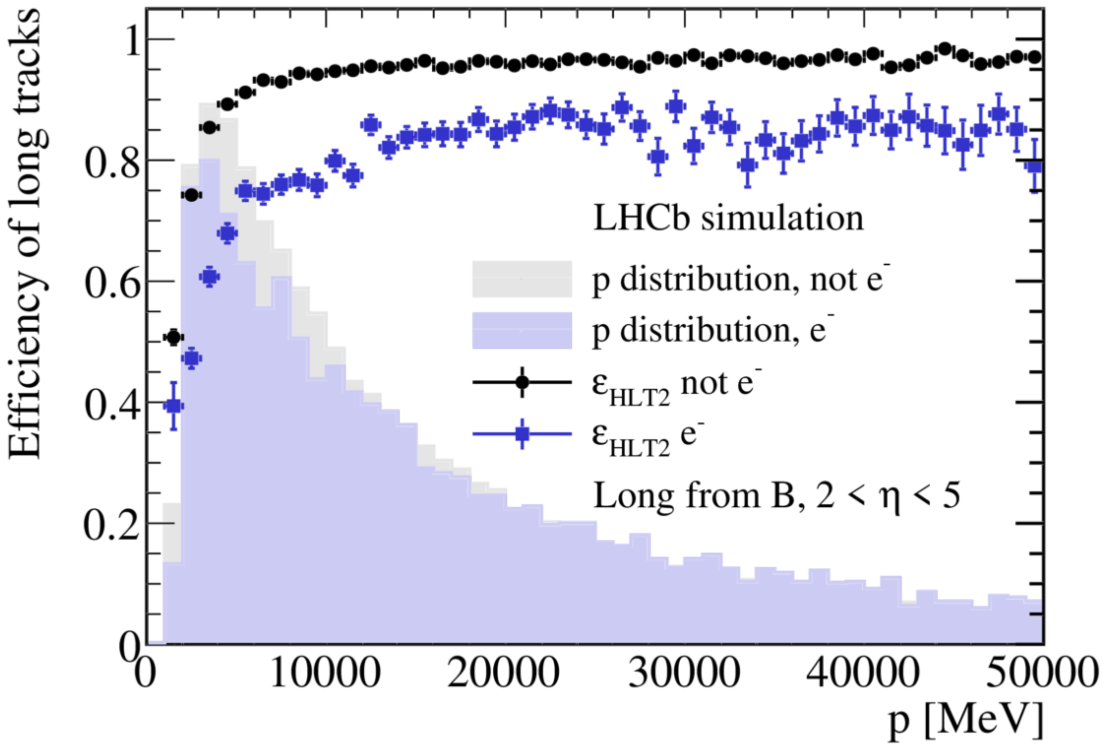


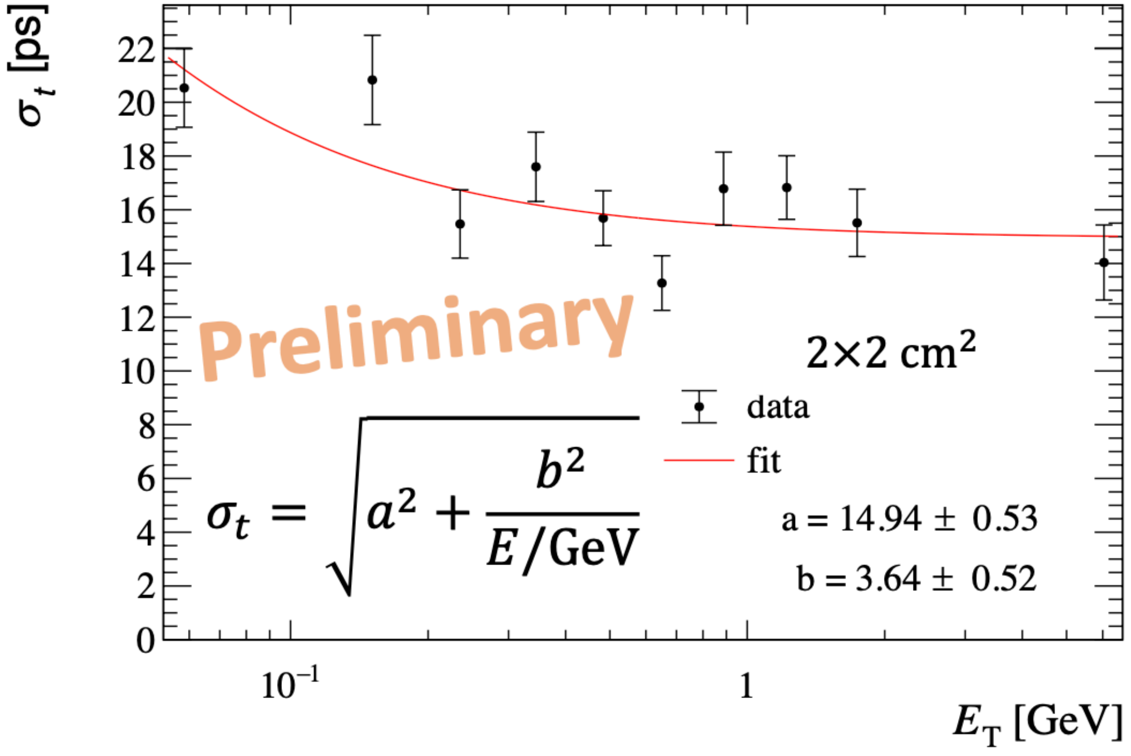
Many studies show isolated benefits of using timing information, particularly combining assumptions about the VELO with individual particle ID detectors

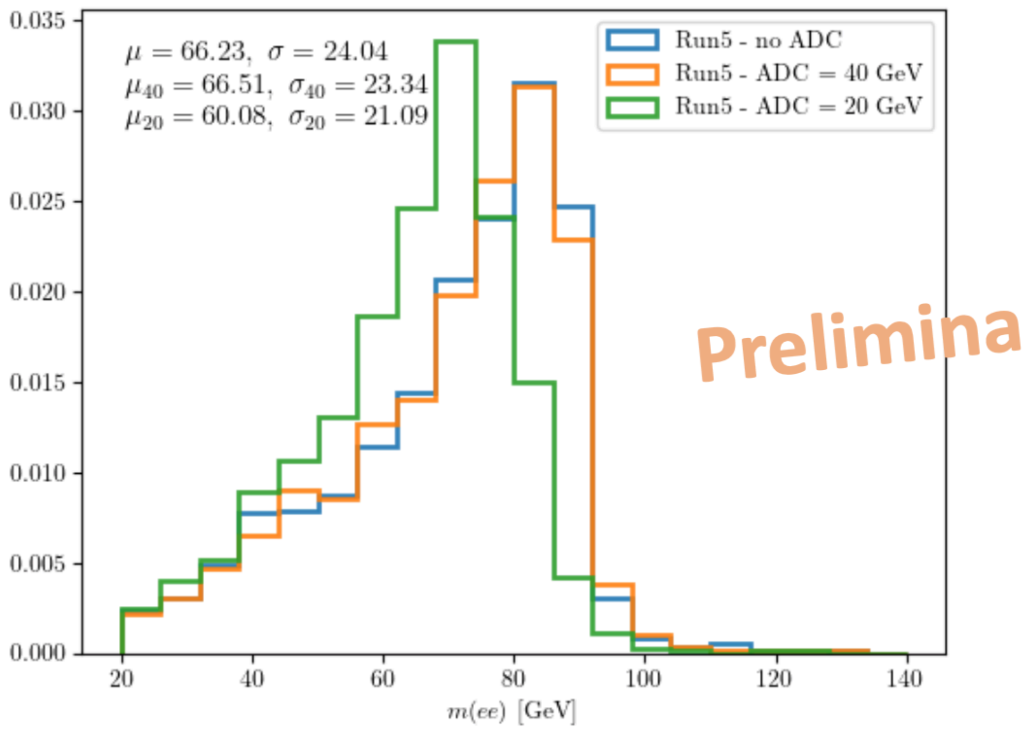
However I would claim we are still missing a take home plot which makes it compellingly obvious to a funder why we must build precise timing throughout

As with tracking PID studies should move towards a coherent evaluation of the performance, particularly considering the known gains when combining information from different PID detectors — which timing may enhance further!

Electrons are tracks too





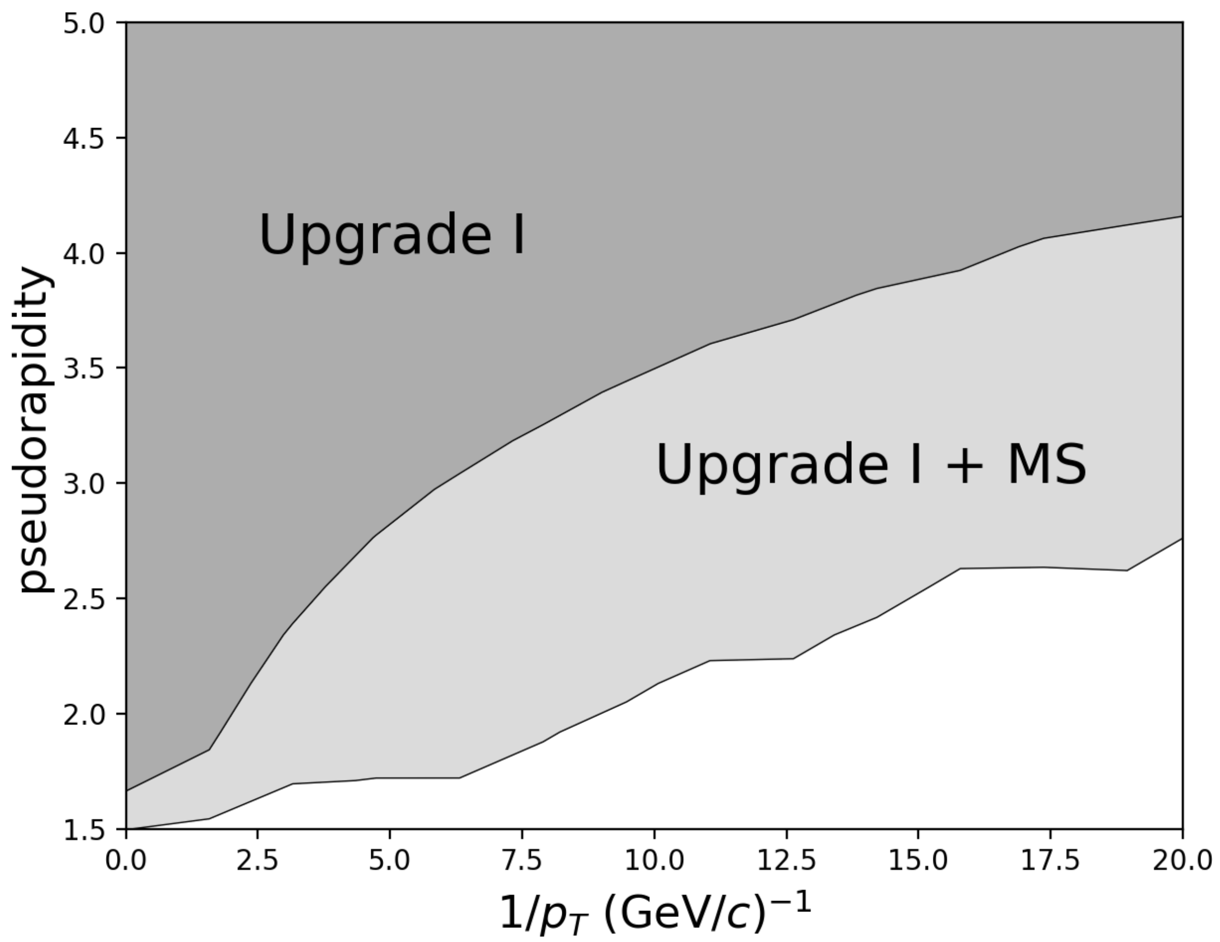


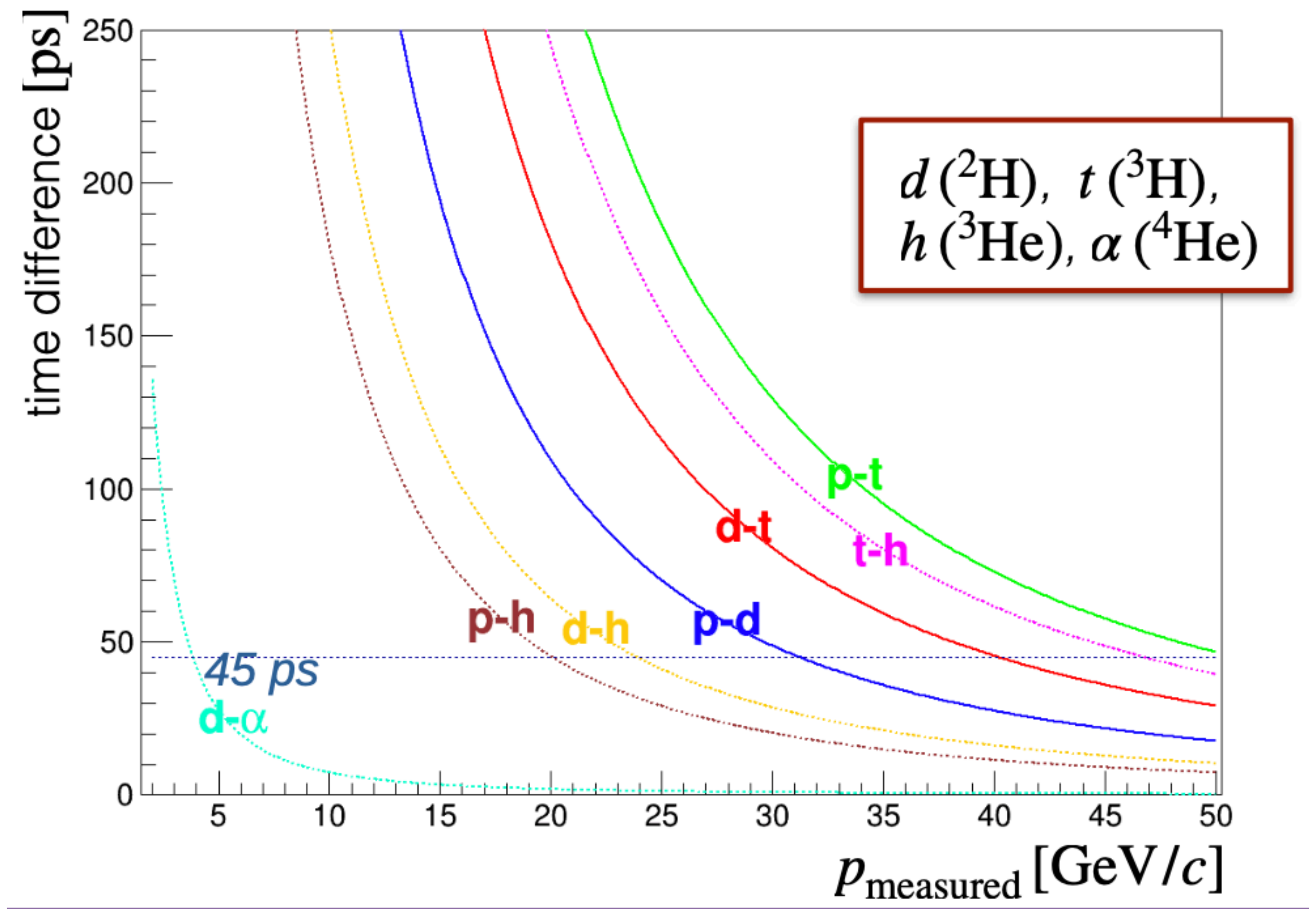
Electron tracking performance is rather poor in U1 conditions, but can U2 tracker choices help?

Important to study the track finding, brems recovery, and CALO energy measurement together in a coherent way and understand if these can guide some of the detector choices we need to make

Performance for high energy electrons is also important — good to see progress on studies

More acceptance, better acceptance





Compared to U1, two projects in particular offer to increase the detector acceptance: TORCH and the magnet stations

We have also heard in the context of the short detector about moving acceptance from the hottest regions near the beampipe to lower eta, where there are less backgrounds

An improved acceptance is clearly desirable but these claims must now stand up to the scrutiny of more extensive and once again more global simulation studies

Vetoing pions and accepting muons

	$B_s^0 \to \mu^+\mu^-$		$D^0 \to \mu^+\mu^-$		$K_s^0 \to \mu^+\mu^-$		$B_s^0 \to J/\psi(\mu^+\mu^-)\ \phi(K^+K^-)$	
Scenario	1-ε (MWPC)	1-ε (μRwell)	1-ε (MWPC)	1-ε (μRwell)	1-ε (MWPC)	1-ε (μRwell)	1-ε (MWPC)	1-ε (μRwell)
HCAL	24.7 %	10.3 %	25.9 %	9.4 %	20.0 %	8.4 %	24.9 %	9.5 %
SHIELD	19.0 %	8.6 %	19.4 %	7.6 %	13.9 %	6.3 %	18.7 %	7.8 %
w/o M2	13.4 %	6.0 %	13.7 %	5.3 %	8.3 %	3.2 %	13.1 %	5.3 %

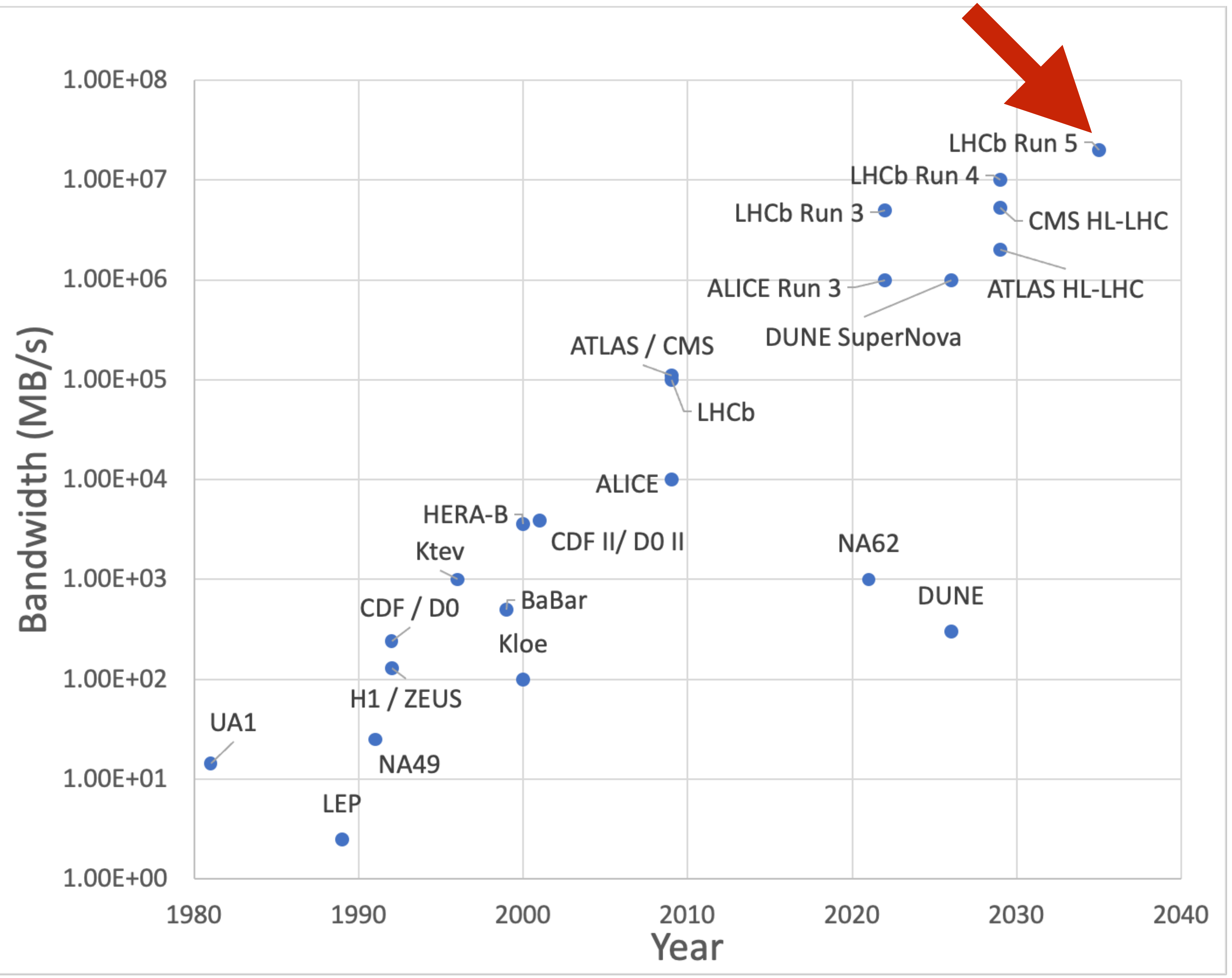
- Superimposing 20 events: <N>= 3700

χ2<	1	2	3	4	5	6	7	8	9	nocut
Muon efficiency	37.0	67.8	79.7	85.4	88.3	89.5	90.5	91.1	91.5	93.1
Pion misID	0.36	1.2	2.2	3.3	4.3	5.4	6.4	7.3	8.1	14.3

In our rush to build exciting new detectors let us not forget that many of our key physics drivers involve muons, and muon performance will therefore remain fundamental

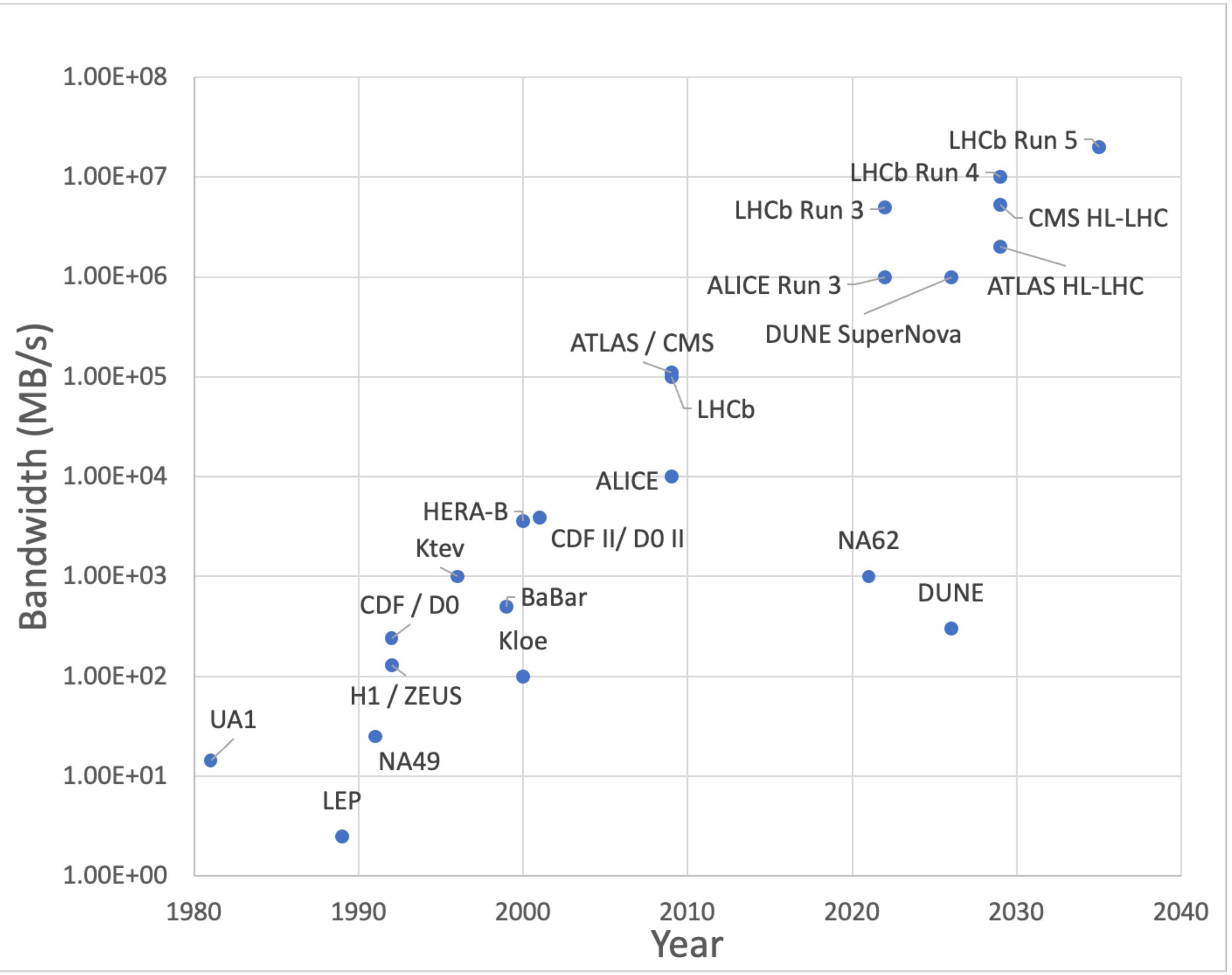
Clearly a very challenging environment in U2 — lots of studies have been performed but must now converge on the optimal shielding and geometry. Can global studies help?

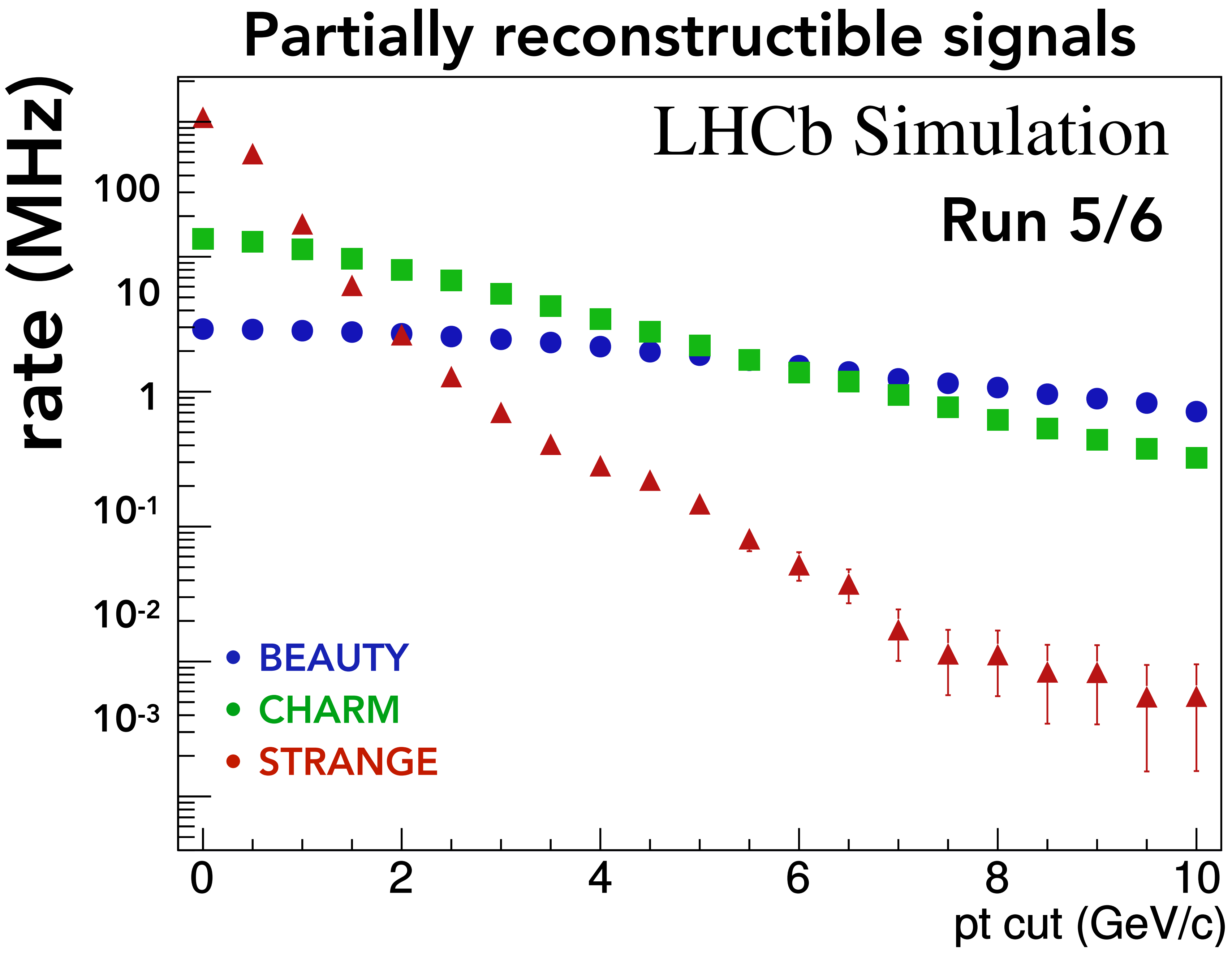
The biggest data challenge in HEP



An order of magnitude more than the HL-LHC ATLAS trigger!

The biggest data challenge in HEP





And we are signal saturated already at the first-level trigger!

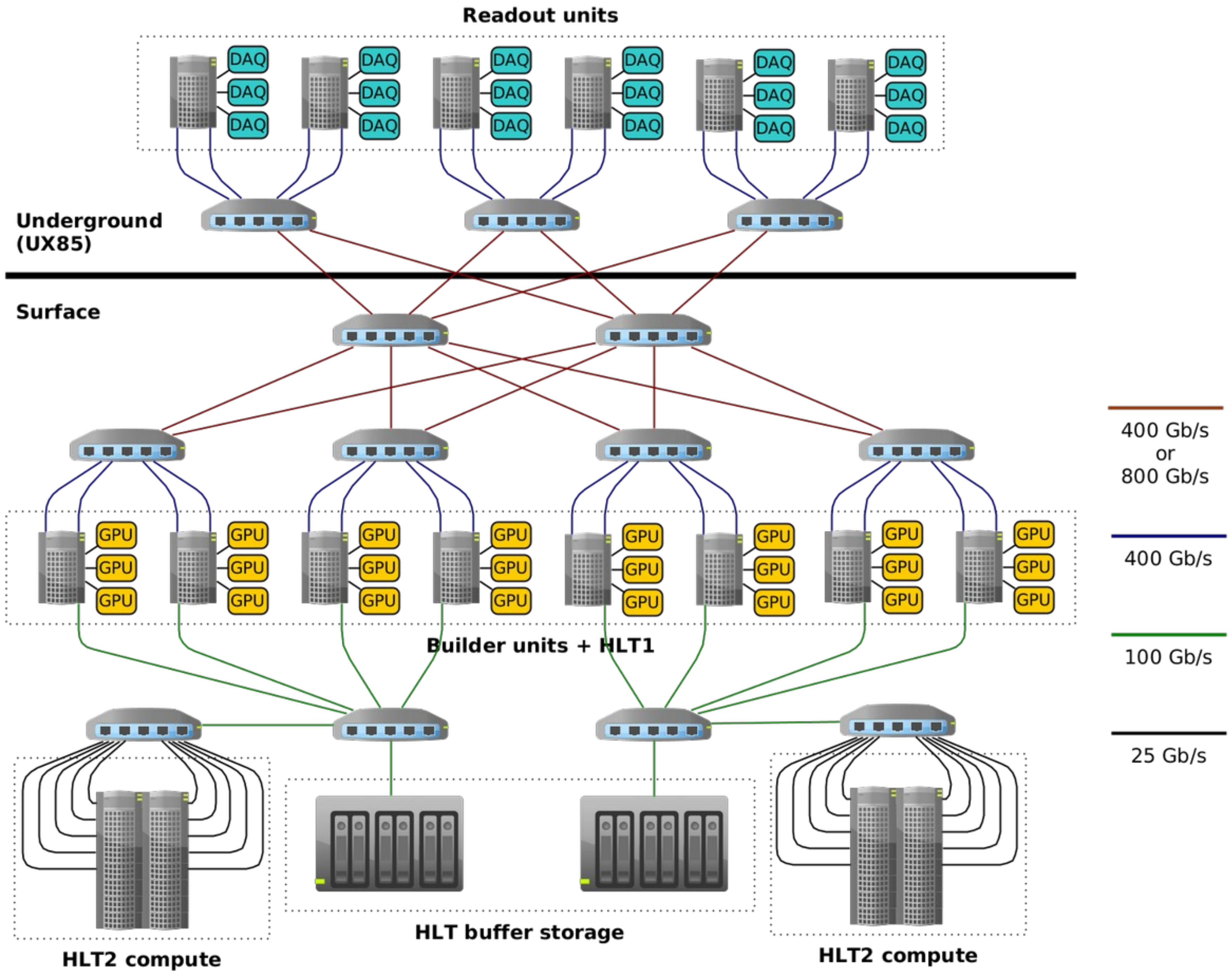
Evolving the detector readout

An extremely challenging but evolutionary path forward

Our triggerless design remains correct and scalable

Aim to have more flexibility in the first level processing units

Understand what compute can be done on the readout boards themselves (and the servers hosting them!)



Try to study as much of this as possible already in Run 4 with the PCIe400 boards

Biggest analysis challenge in HEP

This is not only a trigger challenge!

Unprecedented volume of data and signals to offline analysts

300 fb^{-1} means 1e10 $D_s \to \pi\pi\pi$ *signal events* to measure CP violation.

Can we extend the power and reproducibility of analysis productions to support full pipelines, not just ntupling? AKA "Analysis facilities".

Will this help small institutes to not be locked out of complex analyses?

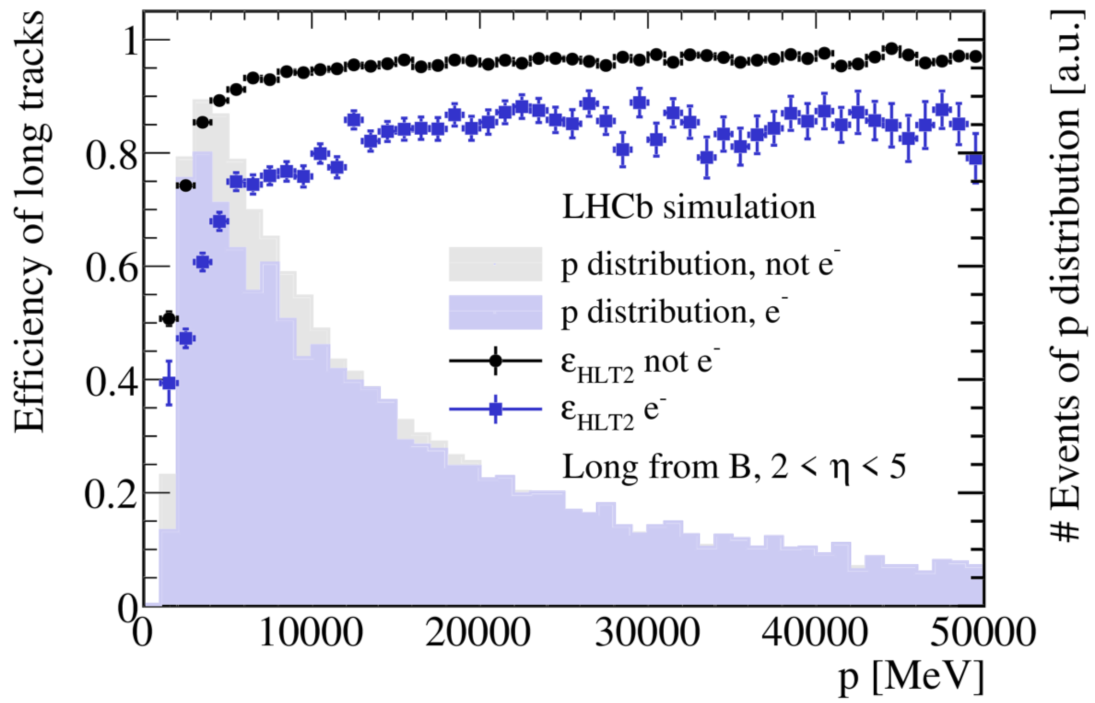
On which architectures will it run? For which computing languages?

How do we maintain it over 20+ years?

How do we efficiently train newcomers in best practices without shutting down new ideas? How do we ensure rapid prototyping remains possible?

Huge opportunities for collaboration with Data Science institutes in other disciplines. Can we take them?

Biggest simulation challenge in HEP



LHCb simulation faces a unique combination of challenges compared to the rest of HEP

1. Control per-particle reconstruction and identification efficiencies at the permille level if not better
2. Deal with an efficiency which is rapidly varying as a function of kinematics for the bulk of our signals
3. In U2 learn to live without lumi levelling

What mixture of detailed, parametric, and/or machine-learned simulation is sufficient to ensure our key physics objectives are not systematics limited?

Can we exploit the available computing resources? Including HPC?

Huge opportunity to lead the development of HEP simulation tools *and* get jobs for this. Can we take it?

Biggest framework challenge in HEP

LHCb's core software, online, and RTA teams built two scalable frameworks to efficiently exploit parallel architectures: Gaudi and Allen

Begin R&D to build on this success with a workshop later this year

- Define production use-cases and their end-to-end requirements
- Define focused technology demonstrators to guide discussions

Scope is very important: what are the jobs of a modern framework?

- Some areas like configuration, persistency, provenance are clear.
- I/O, scheduling, and monitoring may be more application specific.

How do we minimize maintenance while remaining flexible to integrate new architectures and languages (?) as they emerge?

We already lead HEP in the development of high-performance applications. Can we get paid for it?

Evolutionary path to Run 5

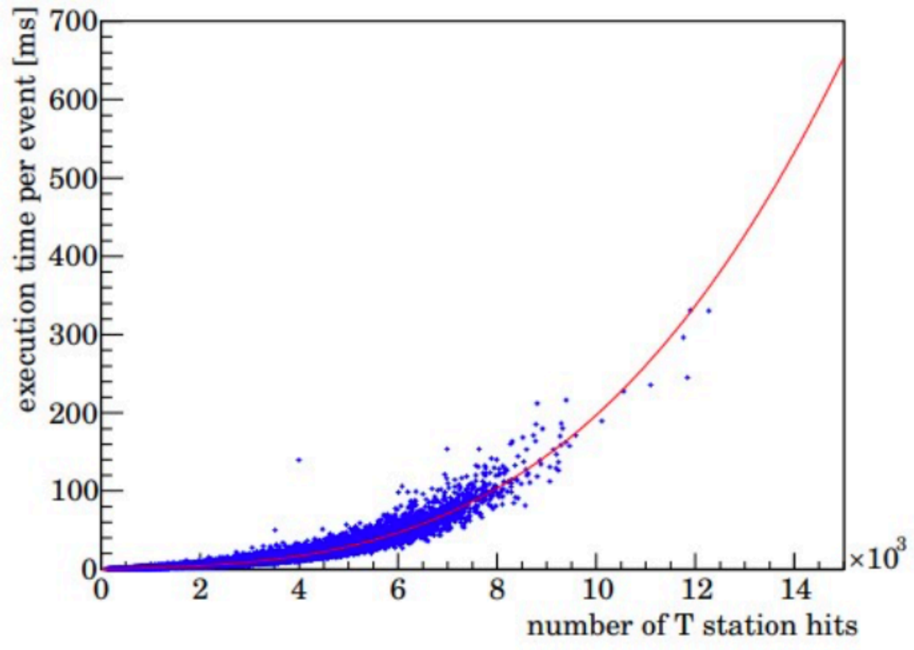
Scale up the Run 3 processing model. Full detector reconstruction (and selections?) on GPUs.

HLT1 inclusive output rate will go to O(10) MHz — use exclusive selections for charm/strange at HLT1.

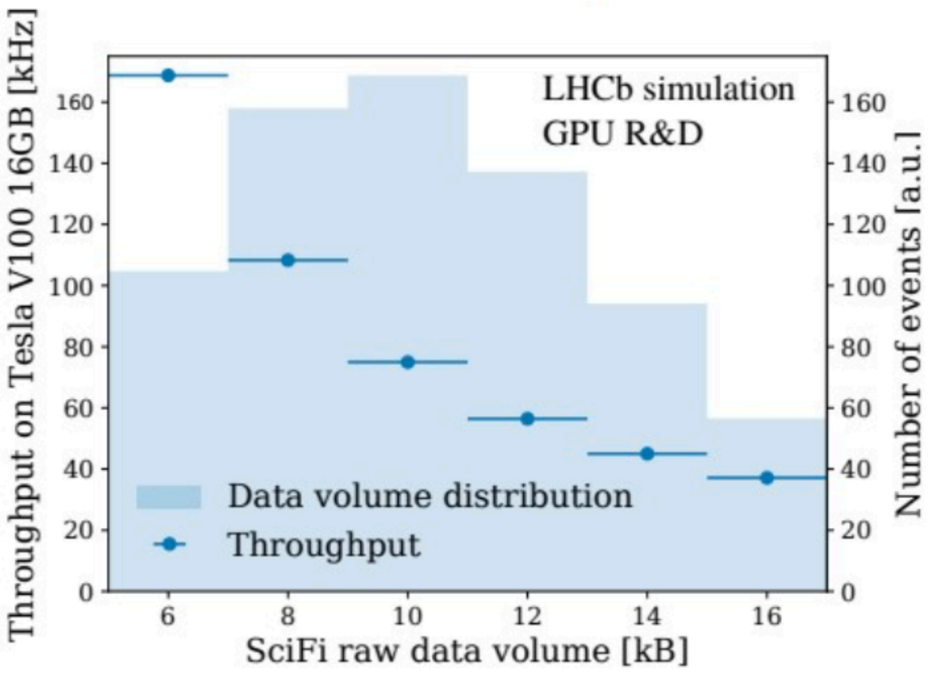
How much will improved detector granularity recover a linear (or better) scaling of cost with luminosity.

Does timing actually help make reconstruction or selections faster?

Critical feedback loop to detector optimizations!

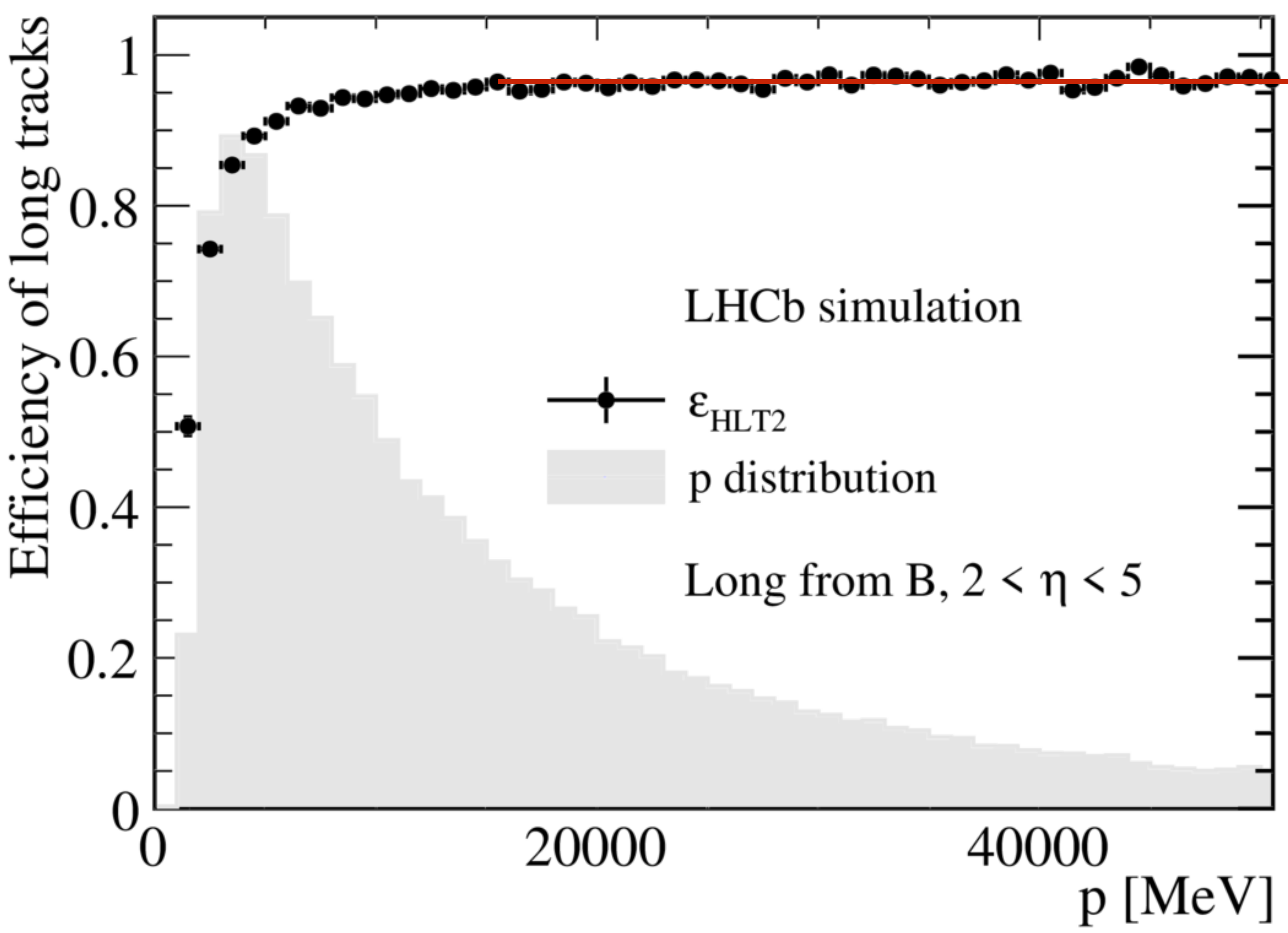


(a) *PatSeeding*

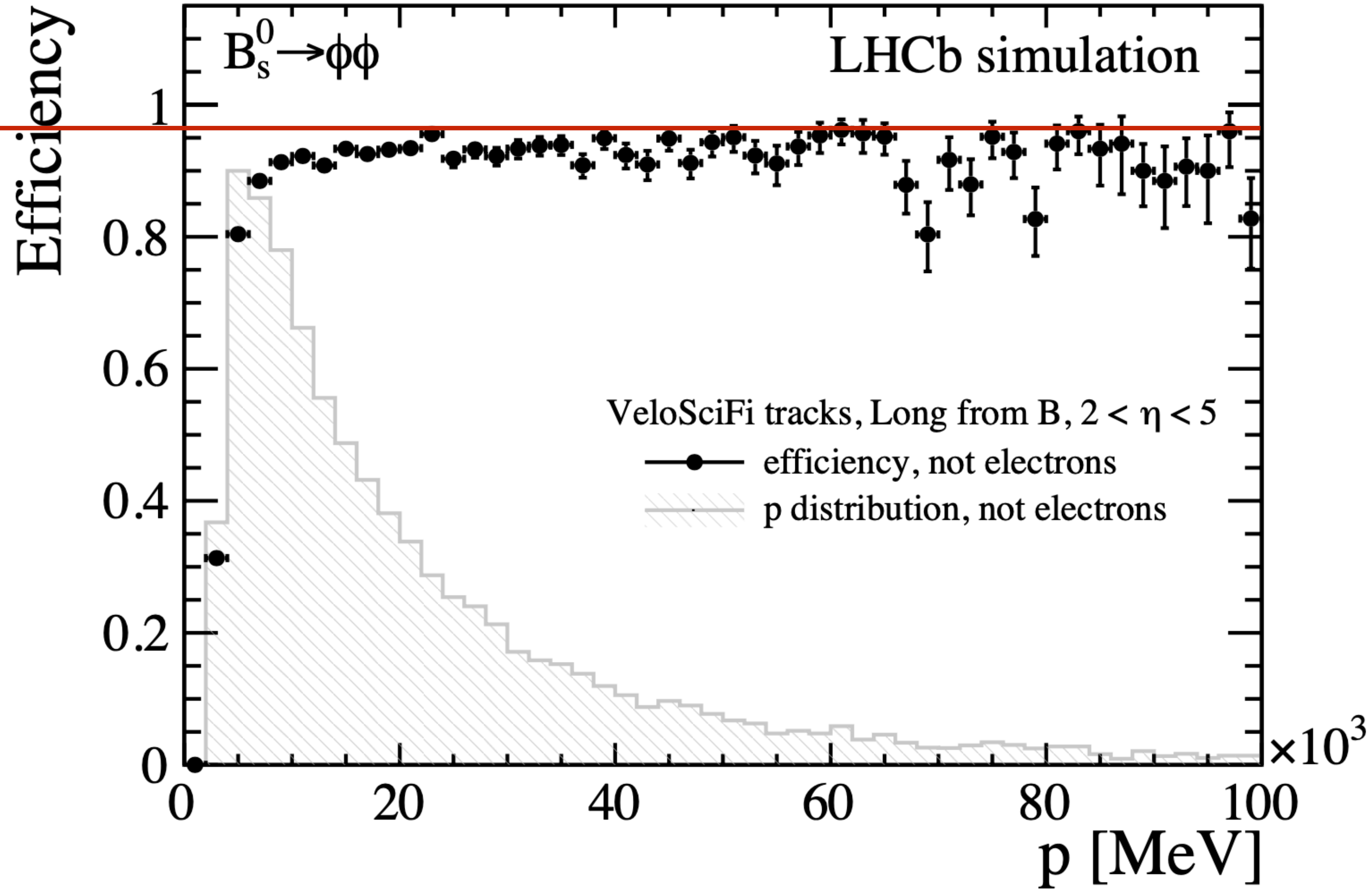


Where do we spend our resources?

HLT2

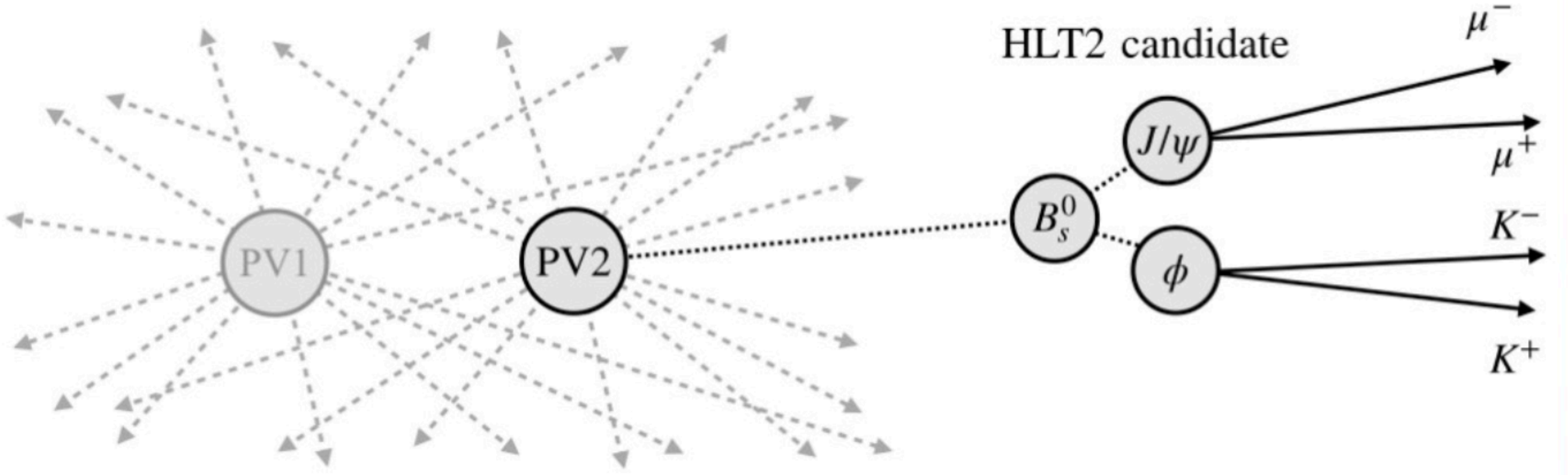


HLT1



We spend roughly 30x more computing resources in HLT2 tracking + fitting than HLT1 tracking.
It gains us the last few percent of tracks and the best covariance estimate.
Is this proportionate? Are we spending our resources answering the right questions?

Can we have inclusive pileup suppression?



R&D ongoing right now for Run 3 and will guide U2 design

At what level of granularity can we associate objects to PVs? Hits, primitives, particles… ?
To what extent does timing information help with this? What biases are introduced for physics analysis?
How do we calibrate identification efficiencies and misidentification rates?

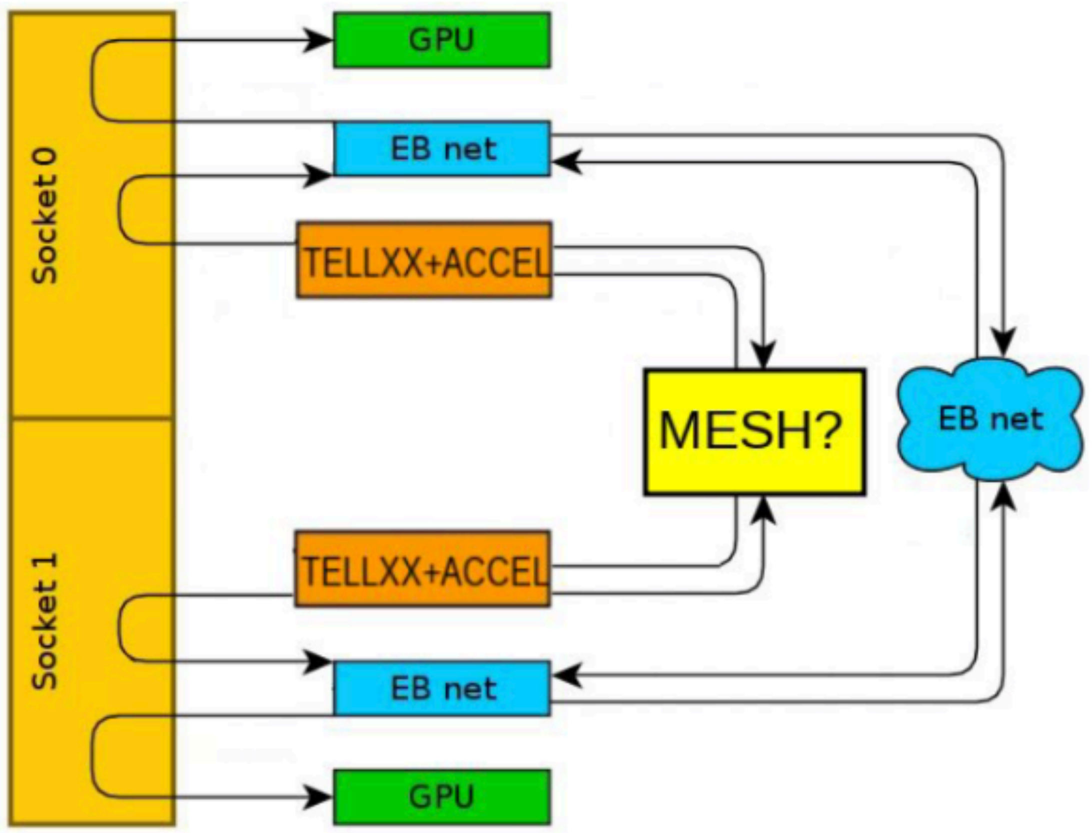
(R)evolutionary path to Run 5

R&D focused on two main areas

1. Finding subdetector primitives, for example tracks or calorimeter clusters, on FPGAs.
2. Exploiting beyond-GPU architectures such as the IPU or even more exotic hardware

Can clever choice of detector geometry make local reconstruction of primitives easier?

Ensure critical mass of skills across all architectures. Decide the architecture mix for Run 5 at the right moment to maximize both technology and long-term support for people from the funding agencies.



Discussions around long-term support for people now beginning in RTA — these are critical for success!

All roads however lead to the pit

« ‹ › »

May 2023

today

Mon	Tue	Wed	Thu	Fri	Sat	Sun
1 06:30 Qiuchan Lu 14:30 Afternoon Shif 22:30 Evening Shift	2 06:30 Vladimir Gligo 14:30 Vsevolod Yero 22:30 Evening Shift	3 06:30 Paula Collins 14:30 Andrea Villa 22:30 Evening Shift	4 06:30 Paula Collins 14:30 Elena DallOcc 22:30 Matteo Paluta	5 06:30 Vladimir Gligo 14:30 Paula Collins 22:30 Alberto Lusiar	6 06:30 Andrea Villa 14:30 Igor Diachkov 22:30 Alberto Lusiar	7 06:30 Qiuchan Lu 14:30 Afternoon Shif 22:30 Alberto Lusiar
8 06:30 Igor Diachkov 14:30 Giacomo Graz 22:30 Evening Shift	9 06:30 Vladimir Gligo 14:30 Giacomo Graz 22:30 Murilo Santan	10 06:30 Giacomo Graz 14:30 Wojciech Krup 22:30 Murilo Santan	11 06:30 Antonis Papar 14:30 Wojciech Krup 22:30 Igor Diachkov	12 06:30 Vladimir Gligo 14:30 Antonis Papan 22:30 Evening Shift	13 06:30 Giacomo Graz 14:30 Antonis Papan 22:30 Evening Shift	14 06:30 Giacomo Graz 14:30 Paula Collins 22:30 Evening Shift
15 06:30 Antonis Papar 14:30 Cristian Pirghi 22:30 Evening Shift	16 06:30 Igor Diachkov 14:30 Cristian Pirghi 22:30 Evening Shift	17 06:30 Igor Diachkov 14:30 Cristian Pirghi 22:30 Evening Shift	18 06:30 Igor Diachkov 14:30 Cristian Pirghi 22:30 Evening Shift	19 06:30 Vladimir Gligo 14:30 Cristian Pirghi 22:30 Evening Shift	20 06:30 Morning Shift 14:30 Cristian Pirghi 22:30 Evening Shift	21 06:30 Morning Shift 14:30 Cristian Pirghi 22:30 Evening Shift
22 06:30 Morning Shift 14:30 Regis Lefevre 22:30 Cristian Pirghi	23 06:30 Vladimir Gligo 14:30 Regis Lefevre 22:30 Evening Shift	24 06:30 Cristian Pirghi 14:30 Regis Lefevre 22:30 Evening Shift	25 06:30 Cristian Pirghi 14:30 Regis Lefevre 22:30 Evening Shift	26 06:30 Vladimir Gligo 14:30 Igor Diachkov 22:30 Evening Shift	27 06:30 Morning Shift 14:30 Afternoon Shif 22:30 Evening Shift	28 06:30 Morning Shift 14:30 Afternoon Shif 22:30 Evening Shift

Discussing U2 is nice but unless we make U1 work, and work well, there won't be any U2
Hoping to see many of you in the pit over the coming months — let's get that data!

Backup