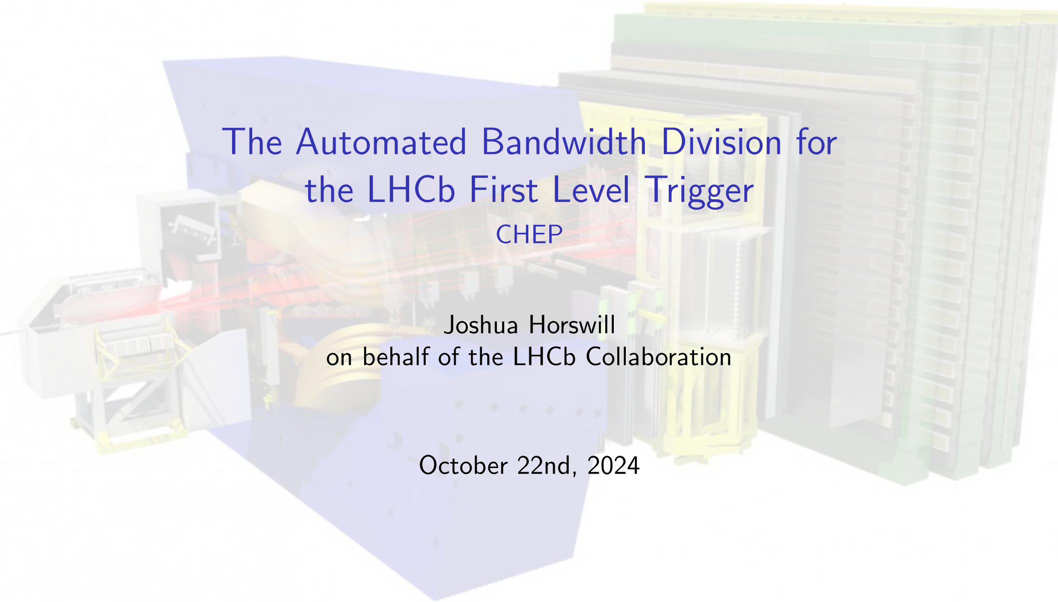


The Automated Bandwidth Division for the LHCb First Level Trigger

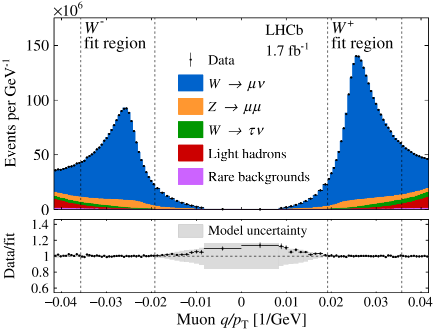
CHEP

Joshua Horswill
on behalf of the LHCb Collaboration

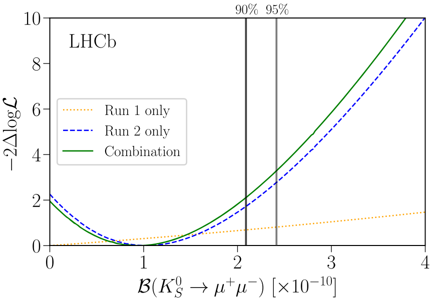
October 22nd, 2024

Introduction

- For first time ever at hadron collider experiment, LHCb uses triggerless readout and **full software** trigger - flexible
- **O(100)** HLT1 selections currently across large breadth of channels covering a huge physics reach.
- See **[this talk]** from Monday by Alessandro Scarabotto and **[this talk]** by Ross Hunter later on today for more details on the LHCb trigger.



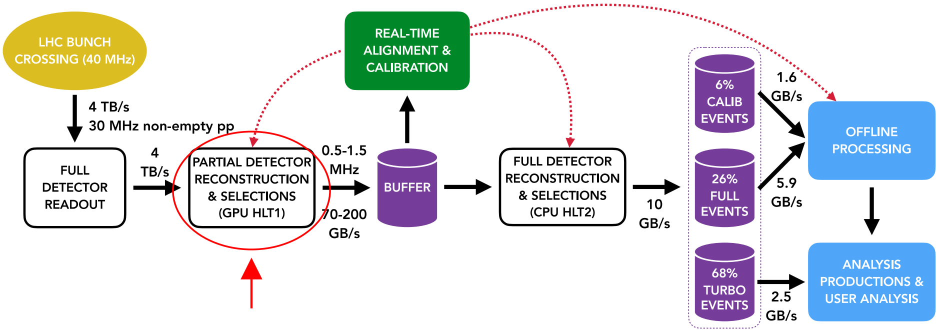
[JHEP 01 (2022) 036]



[Phys. Rev. Lett. 125, 231801]

Maximising Performance

- Introduces complexity → need to **adapt** to changing LHC conditions and modifications to trigger.
- How do we choose selection criteria that **reduces HLT1 trigger output** by a factor 30 whilst also **distributing bandwidth fairly** across all physics channels?
- **Build a tool** that automatically generates set of cuts to achieve this.



LHC BUNCH CROSSING (40 MHz)

4 TB/s
30 MHz non-empty pp

FULL DETECTOR READOUT

4 TB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)

0.5-1.5 MHz

70-200 GB/s

BUFFER

REAL-TIME ALIGNMENT & CALIBRATION

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)

10 GB/s

6% CALIB EVENTS

26% FULL EVENTS

68% TURBO EVENTS

1.6 GB/s

5.9 GB/s

2.5 GB/s

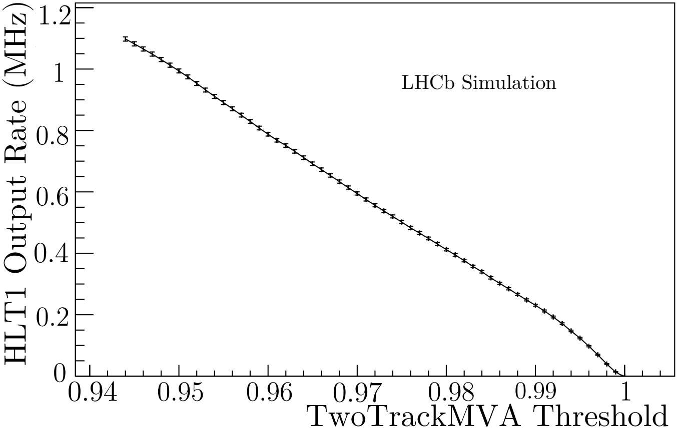
OFFLINE PROCESSING

ANALYSIS PRODUCTIONS & USER ANALYSIS

[LHCB-FIGURE-2020-016]

Most Impactful Lines

- Majority of *b* physics and total HLT1 output rate comes from inclusive hadronic multivariate selections.
- Problem more easily optimisable due to smooth threshold HLT1 output rate curves, e.g., inclusive hadronic TwoTrackMVA line cuts on MVA classifier trained to identify vertices originating from decays of long-lived particles:



Most Impactful Lines

- Some channels must be selected by **exclusive** lines to achieve good efficiencies - not possible with only the inclusive hadronic lines.

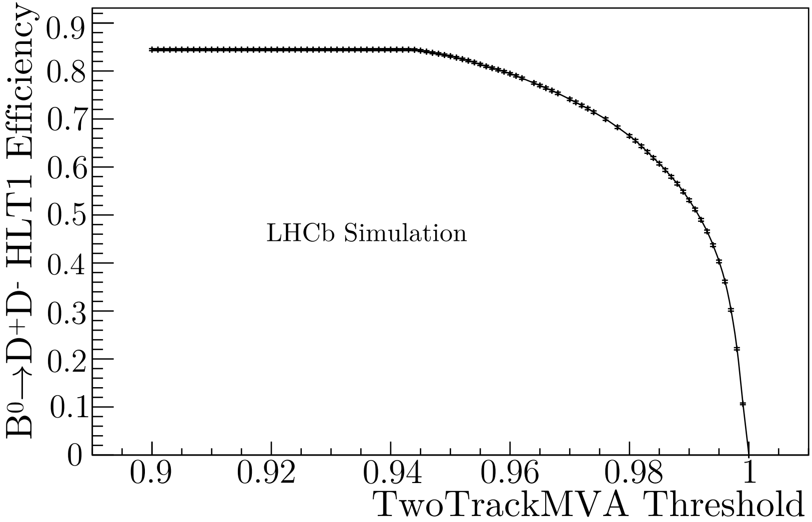


Figure of Merit

- HLT1 trigger thresholds cut on by the tool to minimise:

$$\chi^2_{\text{global}} = \sum_{i}^{\text{channels}} \omega_i \times \left(1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}}\right)^2,$$

Figure of Merit

- HLT1 trigger thresholds cut on by the tool to minimise:

$$\chi^2_{\text{global}} = \sum_{i}^{\text{channels}} \omega_i \times \left(1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}}\right)^2, \qquad \epsilon = \begin{cases} \epsilon_i & \text{BW} \leq \text{BW}_{\text{limit}}, \\ \epsilon_i \times \frac{\text{BW}_{\text{limit}}}{\text{BW}} & \text{otherwise}, \end{cases}$$

Figure of Merit

- HLT1 trigger thresholds cut on by the tool to minimise:

$$\chi^2_{\text{global}} = \sum_i^{\text{channels}} \omega_i \times (1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}})^2, \qquad \epsilon = \begin{cases} \epsilon_i & \text{BW} \leq \text{BW}_{\text{limit}}, \\ \epsilon_i \times \frac{\text{BW}_{\text{limit}}}{\text{BW}} & \text{otherwise}, \end{cases}$$

- ϵ_i^{max} calculated before global minimisation using same rate-limited efficiencies but separate $\chi^2_{\text{indiv}} = (1 - \epsilon)^2$.

Figure of Merit

- HLT1 trigger thresholds cut on by the tool to minimise:

$$\chi^2_{\text{global}} = \sum_{i}^{\text{channels}} \omega_i \times \left(1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}}\right)^2, \qquad \epsilon = \begin{cases} \epsilon_i & \text{BW} \leq \text{BW}_{\text{limit}}, \\ \epsilon_i \times \frac{\text{BW}_{\text{limit}}}{\text{BW}} & \text{otherwise,} \end{cases}$$

- ϵ_i^{max} calculated before global minimisation using same rate-limited efficiencies but separate $\chi^2_{\text{indiv}} = (1-\epsilon)^2$.
- $\epsilon_{\text{i}}^{(\text{max})}$ = number of reconstructible signal events passing fixed HLT1 thresholds, and certain tuneable thresholds chosen by the tool / number of reconstructible signal events passing fixed HLT1 line thresholds

Figure of Merit (2)

$$\chi^2_{\text{global}} = \sum_i^{\text{channels}} \omega_i \times \left(1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}}\right)^2, \qquad \epsilon = \begin{cases} \epsilon_i & \text{BW} \leq \text{BW}_{\text{limit}}, \\ \epsilon_i \times \frac{\text{BW}_{\text{limit}}}{\text{BW}} & \text{otherwise,} \end{cases}$$

- For track to be ‘reconstructible’ in this case, needs to have at least 4 hits in VELO tracker and 6 hits in the Sci-Fi tracker.

Figure of Merit (2)

$$\chi^2_{\text{global}} = \sum_i^{\text{channels}} \omega_i \times (1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}})^2, \qquad \epsilon = \begin{cases} \epsilon_i & \text{BW} \leq \text{BW}_{\text{limit}}, \\ \epsilon_i \times \frac{\text{BW}_{\text{limit}}}{\text{BW}} & \text{otherwise}, \end{cases}$$

- For track to be ‘reconstructible’ in this case, needs to have at least 4 hits in VELO tracker and 6 hits in the Sci-Fi tracker.
- 80 signal channels and 35 lines tuned by 16 thresholds that analysts are happy with to adjust output rate.

Figure of Merit (2)

$$\chi^2_{\text{global}} = \sum_i^{\text{channels}} \omega_i \times \left(1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}}\right)^2, \qquad \epsilon = \begin{cases} \epsilon_i & \text{BW} \leq \text{BW}_{\text{limit}}, \\ \epsilon_i \times \frac{\text{BW}_{\text{limit}}}{\text{BW}} & \text{otherwise}, \end{cases}$$

- For track to be ‘reconstructible’ in this case, needs to have at least 4 hits in VELO tracker and 6 hits in the Sci-Fi tracker.
- 80 signal channels and 35 lines tuned by 16 thresholds that analysts are happy with to adjust output rate.
- Current ensemble uses 9M real LHCb data events collected during 2024 with no selection criteria applied to determine the HLT1 output rate.

Figure of Merit (3)

$$\chi^2_{\text{global}} = \sum_{i}^{\text{channels}} \omega_i \times \left(1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}}\right)^2, \qquad \epsilon = \begin{cases} \epsilon_i & \text{BW} \leq \text{BW}_{\text{limit}}, \\ \epsilon_i \times \frac{\text{BW}_{\text{limit}}}{\text{BW}} & \text{otherwise}, \end{cases}$$

- Trigger is ran in special configuration allowing the modification and testing of trigger settings without having to recollect new data.

Figure of Merit (3)

$$\chi^2_{\text{global}} = \sum_i^{\text{channels}} \omega_i \times \left(1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}}\right)^2, \qquad \epsilon = \begin{cases} \epsilon_i & \text{BW} \leq \text{BW}_{\text{limit}}, \\ \epsilon_i \times \frac{\text{BW}_{\text{limit}}}{\text{BW}} & \text{otherwise}, \end{cases}$$

- Trigger is ran in special configuration allowing the modification and testing of trigger settings without having to recollect new data.
- Need to minimise this loss function in discrete multidimensional space. How are we going to do this?

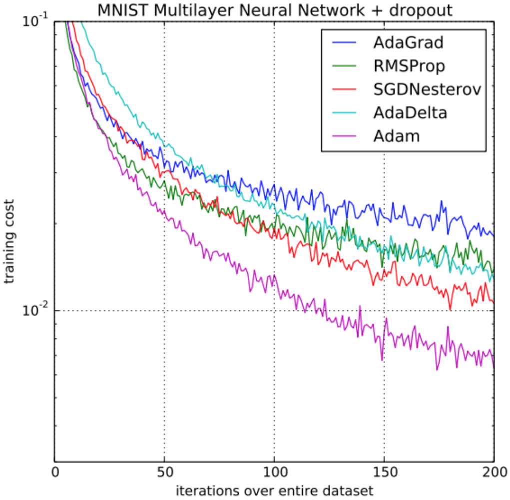
Figure of Merit (3)

$$\chi^2_{\text{global}} = \sum_{i}^{\text{channels}} \omega_i \times \left(1 - \frac{\epsilon_i}{\epsilon_i^{\text{max}}}\right)^2, \qquad \epsilon = \begin{cases} \epsilon_i & \text{BW} \leq \text{BW}_{\text{limit}}, \\ \epsilon_i \times \frac{\text{BW}_{\text{limit}}}{\text{BW}} & \text{otherwise,} \end{cases}$$

- Trigger is ran in special configuration allowing the modification and testing of trigger settings without having to recollect new data.
- Need to minimise this loss function in discrete multidimensional space. How are we going to do this?
- Minimisation performed using the Adapted Moment (Adam) algorithm, finishing with grid search in the neighbourhood of the continuous solution to find optimal discrete ensemble.

Adam Algorithm Breakdown

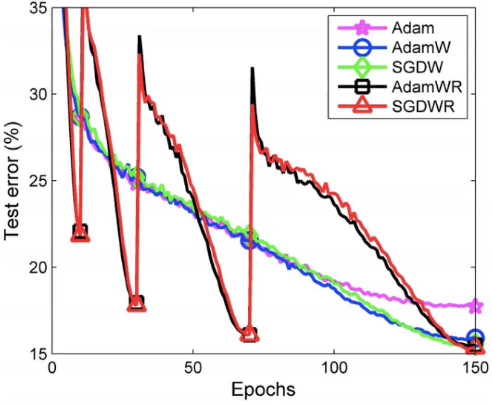
- Adam is computationally efficient and well suited to problems with large datasets, many parameters, and noisy/sparse gradients - good for this task.



ICLR 2015

Adam Algorithm Breakdown (2)

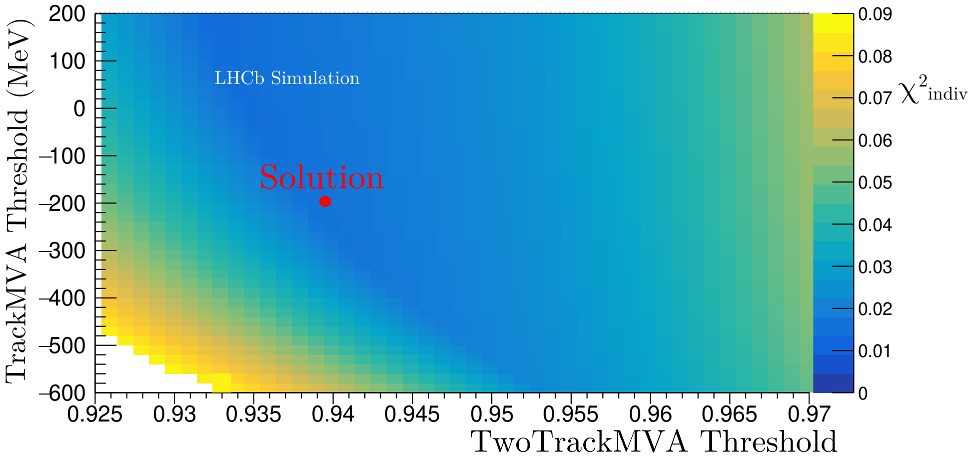
- Adam maintains learning rate (step size) for each tuneable parameter; gradient descent is accelerated/decelerated by processing weighted averages of ‘momentum’ determined by gradient history.
- In combination, these features enable minimisation path to ‘roll’ through local minima, i.e, against positive gradients to reach global minimum. ImageNet32x32 classification comparison:



[ICLR 2019]

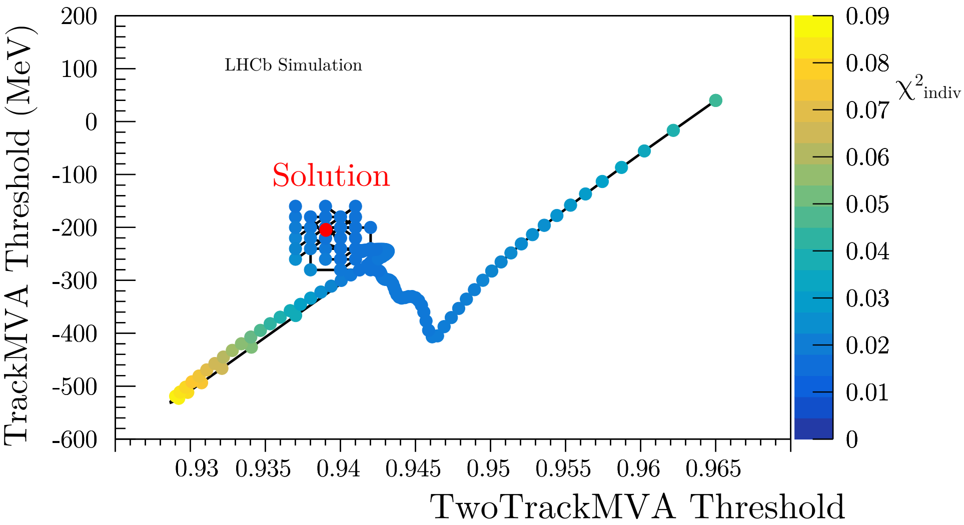
Adam at Work

- Smearing of three weighted gradient calculations per parameter over increasing widths centered on current set of thresholds used to estimate numerically stable gradient - dictates the size and direction of next step using adapted moment system.



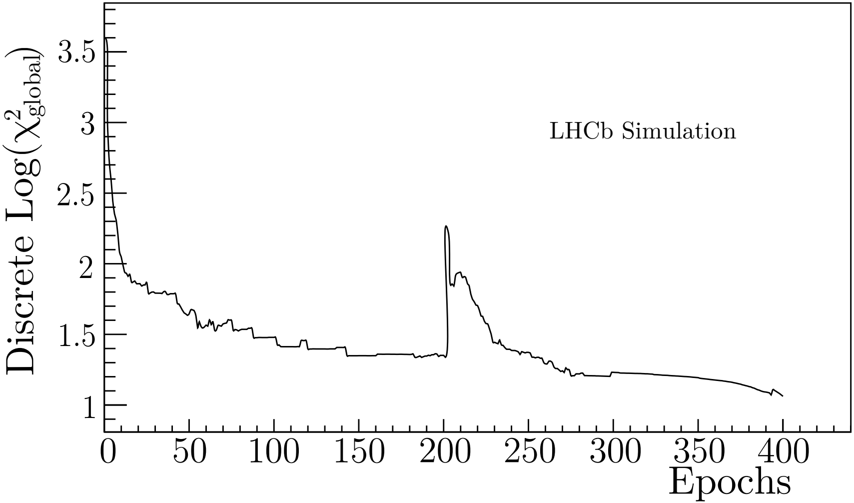
Adam at Work (2)

- LHCb simulation with very similar distributions to real 2024 data used to calculate $\epsilon_{\mathrm{i}}^{\mathrm{max}}(B^0 \rightarrow D^+D^-)$ for evaluation of the χ^2_{indiv} FoM, indicated by the colour scale:



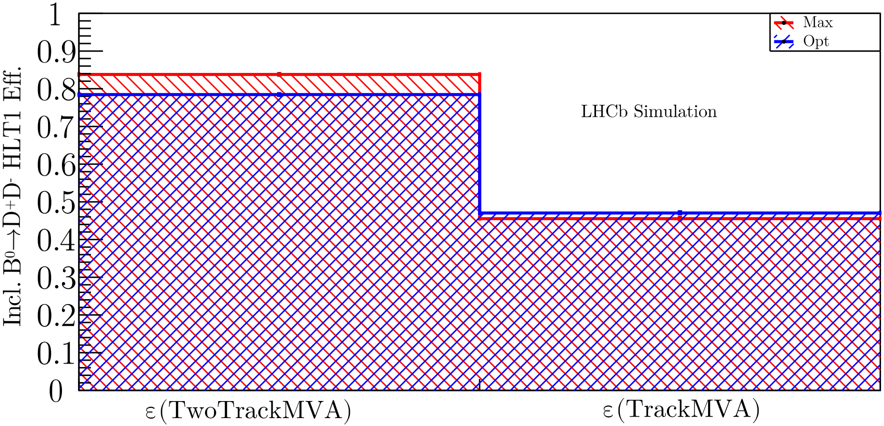
Adam Minimisation with Warm Restart

- Adam was compared with genetic algorithm and exhibited similar performance at 16 dimensions, but was **less hyperparameter-sensitive** and so Adam preferred over stochastic methods.

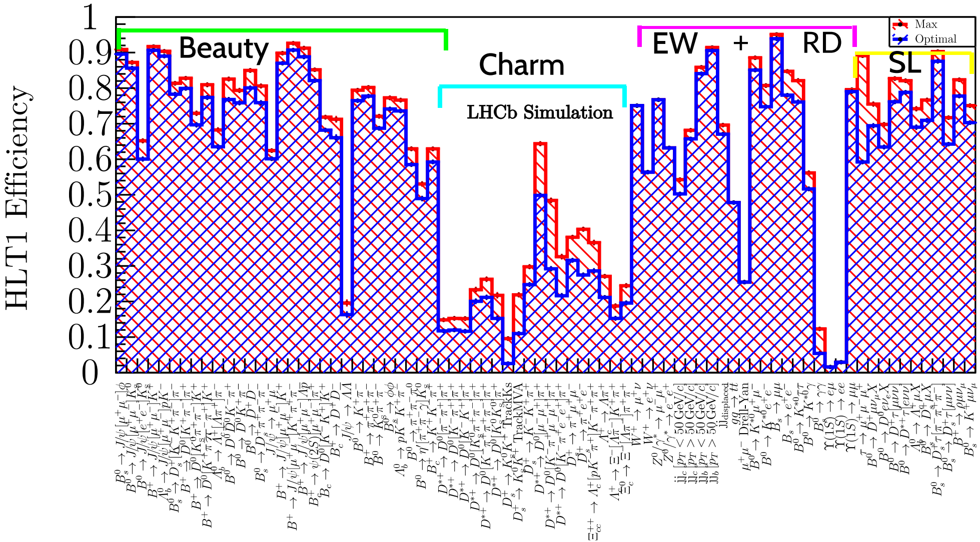


$B^0 \to D^+D^-$ Efficiencies from 2024 Thresholds

- Inclusive line HLT1 efficiencies for $B^0 \to D^+D^-$ with thresholds used to collect data in 2024 without the Upstream Tracker.
- Optimal efficiencies in blue, when bandwidth is being divided equitably; maximum efficiencies in red, where the entirety of the HLT1 bandwidth is allocated to maximising $\epsilon_i^{\max}$. The blue and red efficiencies are also calculated for total (and not just per line) HLT1 efficiency for every sample in the ensemble...

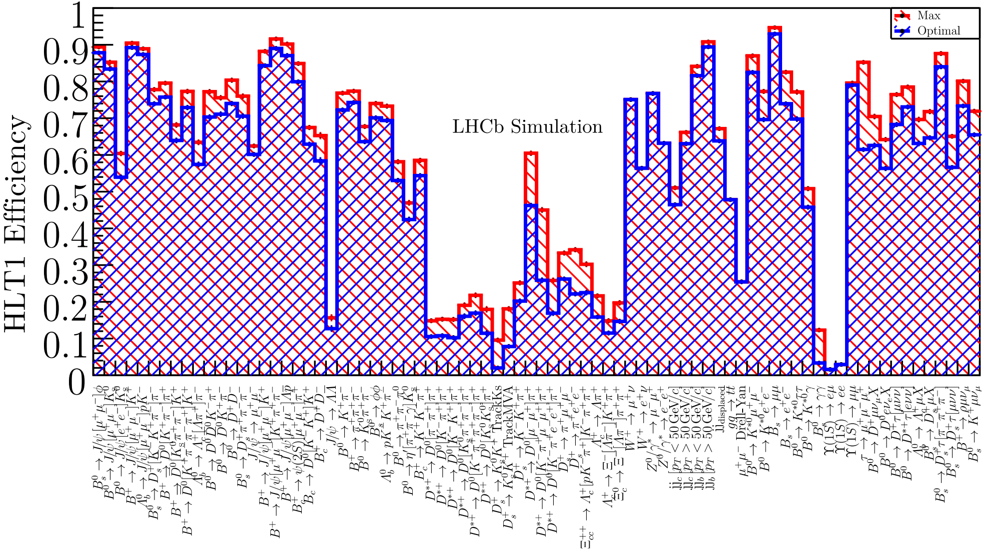


All Efficiencies from 2024 Thresholds



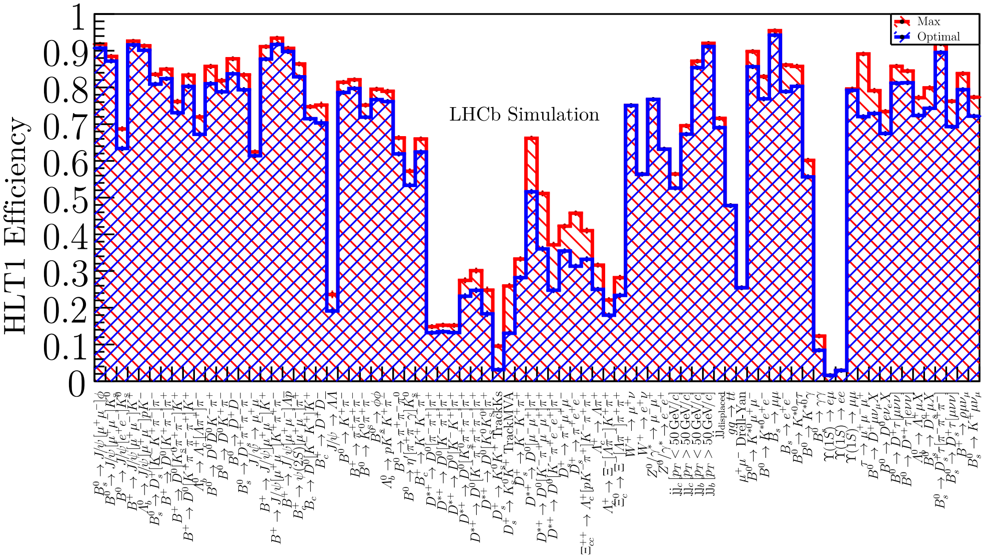
All Efficiencies at Smaller Output Rate

- Can divide bandwidth quickly under changing physics conditions



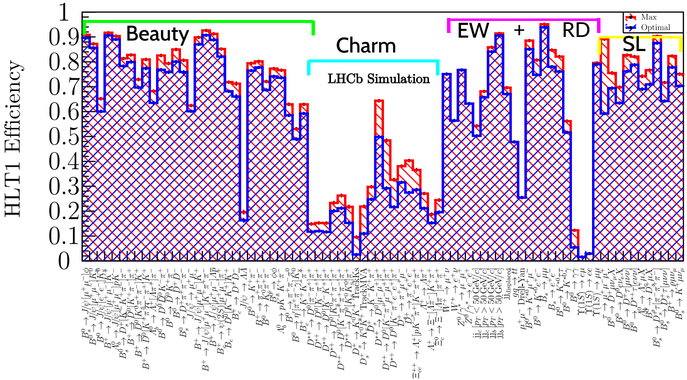
All Efficiencies at Larger Output Rate

- Can divide bandwidth quickly under changing physics conditions



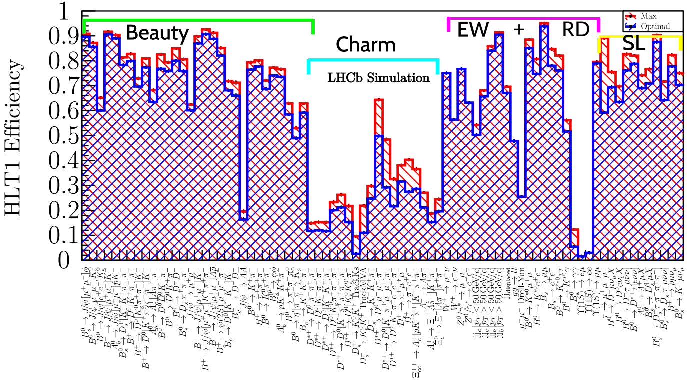
Conclusions

- Fully software trigger introduces **high-dimensional phase space** challenge that this tool enables the collaboration to adapt to quickly.
- Trigger efficiencies could be improved manually, but would be considerably time consuming; automated nature of the tool provides a simple and robust motivation for set of tunings, allowing **fast turnaround** when HLT1 needs re-optimising.



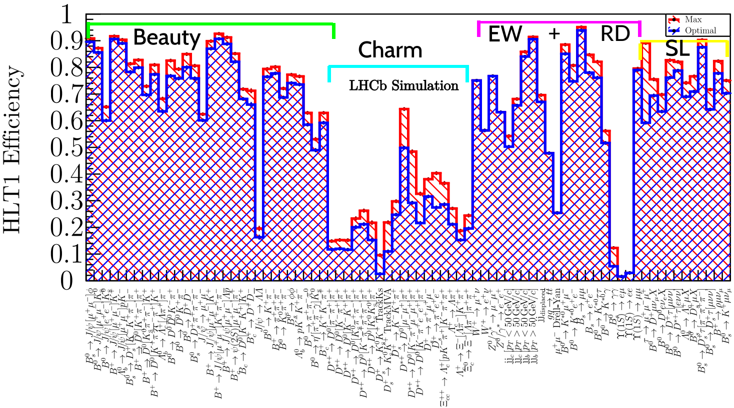
Conclusions (2)

- Comparing 2024 1 MHz ratelimit thresholds to those chosen manually before bandwidth division project began → increase in ϵ of **20% for beauty, 10% for charm and 30% for semi-leptonic channels** averaged over all samples. This translates to a factor 1.4, 2 and 1.75 × increase, respectively.
- Tool also determined thresholds and served as a useful benchmark for trigger performance in 2023.



Conclusions (3)

- For future upgrades the complexity will increase and tool will scale to deal with larger trigger menu; could make sense to build another automated tool that divides bandwidth **based on event size** rather than collision rate.
- Paper in preparation.



Backup Slides

Samples, lines, parameters used

Parameter(s)	Trigger Line(s)	Channel(s)	Mechanism
`alpha_hadron` (MeV)	`Hlt1TrackMVA`	$B_s^0 \to J/\psi[\mu^+\mu^-]\phi$, $B^0 \to J/\psi[\mu^+\mu^-]K_S^0$, $B^0 \to J/\psi[e^+e^-]K_S^0$, $B^+ \to J/\psi[\mu^+\mu^-]K^+$, $\Lambda_b^0 \to J/\psi[\mu^+\mu^-]pK^-$, $B^0 \to K^+\pi^-$, $B^0 \to D^+\mu\nu_\mu X$, $\mathrm{jj}_{\mathrm{displaced}}$ $B_s^0 \to D_s^+[K^-K^+\pi^-]\pi^-$, $B^+ \to \overline{D}^0[K^+\pi^-]\pi^+$, $B^+ \to \overline{D}^0[K_S^0\pi^+\pi^-]K^+$, $B^+ \to \overline{D}^0[K^+\pi^-\pi^-\pi^+]K^+$, $\Lambda_b^0 \to \Lambda_c^+[\Lambda\pi^+]\pi^-$, $B^0 \to D^0\overline{D}^0K^-\pi^+$, $B^0 \to \overline{D}^0K^+\pi^-$, $B^0 \to D^+D^-$, $B_s^0 \to D_s^+\pi^+\pi^-\pi^-$, $B^+ \to J/\psi[\mu^+\mu^-]K^+$, $B^+ \to J/\psi[\mu^+\mu^-]K^+\pi^+\pi^-$, $B^+ \to J/\psi[\mu^+\mu^-]\Lambda^0\overline{p}$, $B_c^+ \to \psi(2S)[\mu^+\mu^-]\pi^+$, $B_c \to D^0[K^-K^+]K^+$, $B_c^+ \to D^+D^-$, $J/\psi \to \Lambda\overline{\Lambda}$, $B^+ \to \pi^+\pi^+\pi^-$, $B^0 \to K_S^0\pi^+\pi^-$, $B_s^0 \to \phi\phi$, $\Lambda_b^0 \to pK^-K^+\pi^-$, $B^0 \to \pi^+\pi^-\pi^0$, $B^0 \to \eta'[\pi^+\pi^-\gamma]K_S^0$, $B_c^+ \to K^+K^-\pi^+$, $D^{*+} \to D^0[K^-K^+]\pi^+$, $D^{*+} \to D^0[K_S^0\pi^+\pi^-]\pi^+$, $D^{*+} \to D^0[K^-\pi^+\pi^+\pi^-]\pi^+$, $D^{*+} \to D^0[K_S^0K_S^0]\pi^+$, $D_s^+ \to K_S^0K^+$, $D_s^+ \to K^-K^+\pi^+$, $D^{*+} \to D^0[\mu^-\mu^+]\pi^+$, $D^{*+} \to D^0[K^-\pi^+\mu^-\mu^+]\pi^+$, $D^{*+} \to D^0[K^-\pi^+e^-e^+]\pi^+$, $D_s^+ \to \pi^+\mu^+\mu^-$, $D_s^+ \to \pi^+e^+e^-$, $\Xi_{cc}^{++} \to \Lambda_c^+[pK^-\pi^+]K^-\pi^+\pi^+$, $\Lambda_c^+ \to \Lambda\pi^+$, $\mathrm{jj}_c\,[p_T < 50\,\mathrm{GeV}/c]$, $\mathrm{jj}_c\,[p_T > 50\,\mathrm{GeV}/c]$, $\mathrm{jj}_b\,[p_T < 50\,\mathrm{GeV}/c]$, $\mathrm{jj}_b\,[p_T > 50\,\mathrm{GeV}/c]$, $B^0 \to K^{*0}\mu^+\mu^-$, $B^0 \to K^{*0}e^+e^-$, $B_s \to \mu\mu$, $B_s \to e^+e^-$, $B^0 \to K^{*0}\tau\tau$, $B^0 \to K^{*0}\gamma$, $B^0 \to D^0e\nu_eX$, $B^0 \to D^+\tau[\mu\nu\nu]$, $B^0 \to D^{*+}\tau[e\nu\nu]$, $\Lambda_b^0 \to \Lambda_c^+\mu X$, $B_s^0 \to D_s^+\mu X$, $B_s^0 \to D_s^+\tau[\pi^-\pi^+\pi^-]$, $B_s^0 \to D_s^+\tau[\mu\nu\nu]$, $B^+ \to \rho\mu\nu_\mu$	Requirement on tranverse momentum and impact parameter significance. Parameter shifts the transverse momentum.

Samples, Lines, Parameters used

Parameter(s)	Trigger Line(s)	Channel(s)	Mechanism
`twotrackmva`	`Hlt1TwoTrackMVA`	$B_s^0 \to J/\psi[\mu^+\mu^-]\phi,\ B^0 \to J/\psi[\mu^+\mu^-]K_{\mathrm{S}}^0,$ $B^0 \to J/\psi[e^+e^-]K_{\mathrm{S}}^0,\ B^+ \to J/\psi[\mu^+\mu^-]K^+,$ $\Lambda_b^0 \to J/\psi[\mu^+\mu^-]pK^-,\ B^0 \to K^+\pi^-$ $B^0 \to D^+\mu\nu_\mu X,\ \mathrm{jj}_{\mathrm{displaced}}$ $B_s^0 \to D_s^+[K^-K^+\pi^-]\pi^-,\ B^+ \to \overline{D}^0[K^+\pi^-]\pi^+,$ $B^+ \to \overline{D}^0[K_{\mathrm{S}}^0\pi^+\pi^-]K^+,\ B^+ \to \overline{D}^0[K^+\pi^-\pi^-\pi^+]K^+,$ $\Lambda_b^0 \to \Lambda_c^+[\Lambda\pi^+]\pi^-,\ B^0 \to D^0\overline{D}^0K^-\pi^+,$ $B^0 \to \overline{D}^0K^+\pi^-,\ B^0 \to D^+D^-,$ $B_s^0 \to D_s^+\pi^+\pi^-\pi^-,\ B^+ \to J/\psi[\mu^+\mu^-]K^+\pi^+\pi^-,$ $B^+ \to J/\psi[\mu^+\mu^-]\Lambda^0\overline{p},\ B_c \to D^0[K^-K^+]K^+,$ $B_c^+ \to D^+D^-,\ J/\psi \to \Lambda\bar{\Lambda},$ $B^+ \to \pi^+\pi^+\pi^-,\ B^0 \to K_{\mathrm{S}}^0\pi^+\pi^-,$ $B_s^0 \to \phi\phi,\ \Lambda_b^0 \to pK^-K^+\pi^-,$ $B^0 \to \pi^+\pi^-\pi^0,\ B^0 \to \eta'[\pi^+\pi^-\gamma]K_{\mathrm{S}}^0,$ $B_c^+ \to K^+K^-\pi^+,\ D^{*+} \to D^0[K^-K^+]\pi^+,$ $D^{*+} \to D^0[K_{\mathrm{S}}^0\pi^+\pi^-]\pi^+,\ D^{*+} \to D^0[K^-\pi^+\pi^+\pi^-]\pi^+,$ $D^{*+} \to D^0[K_{\mathrm{S}}^0K_{\mathrm{S}}^0]\pi^+,\ D_s^+ \to K^-K^+\pi^+,$ $D^{*+} \to D^0[\mu^-\mu^+]\pi^+,\ D^{*+} \to D^0[K^-\pi^+\mu^-\mu^+]\pi^+,$ $D^{*+} \to D^0[K^-\pi^+e^-e^+]\pi^+,\ D_s^+ \to \pi^+\mu^+\mu^-,$ $D_s^+ \to \pi^+e^+e^-,\ \Xi_{cc}^{++} \to \Lambda_c^+[pK^-\pi^+]K^-\pi^+\pi^+,$ $\Lambda_c^+ \to \Lambda\pi^+,$ $\mathrm{jj}_c\,[p_T < 50\,\mathrm{GeV}/c],\ \mathrm{jj}_c\,[p_T > 50\,\mathrm{GeV}/c],$ $\mathrm{jj}_b\,[p_T < 50\,\mathrm{GeV}/c],\ \mathrm{jj}_b\,[p_T > 50\,\mathrm{GeV}/c],$ $B_s \to \mu\mu,\ B_s \to e^+e^-,$ $B^0 \to K^{*0}\tau\tau,\ B^0 \to K^{*0}\gamma,$ $B^0 \to D^0e\nu_eX,\ B^0 \to D^+\tau[\mu\nu\nu],$ $B^0 \to D^{*+}\tau[e\nu\nu],\ \Lambda_b^0 \to \Lambda_c^+\mu X,$ $B_s^0 \to D_s^+\mu X,$ $B_s^0 \to D_s^+\tau[\pi^-\pi^+\pi^-],\ B_s^0 \to D_s^+\tau[\mu\nu\nu],$ $B^+ \to \rho\mu\nu_\mu$	MVA classifier trained to identify vertices originating from decay of long-lived particle.

Samples, Lines, Parameters Used

Parameter(s)	Trigger Line(s)	Channel(s)	Mechanism
`alpha_electron` (MeV)	`Hlt1TrackElectronMVA`	$D^{*+} \to D^0[K^-\pi^+e^-e^+]\pi^+,\ D_s^+ \to \pi^+e^+e^-,$ $B^0 \to K^{*0}e^+e^-,\ B_s \to e^+e^-,$ $B^0 \to D^0 e\nu_e X,\ B^0 \to D^{*+}\tau[e\nu\nu]$	Same mechanism as `alpha_hadron`, but the electronic OneTrack line uses a shift in corrected transverse momentum due to bremsstrahlung radiation.
`alpha_muon` (MeV)	`Hlt1TrackMuonMVA`	$B^+ \to J/\psi[\mu^+\mu^-]K^+, ,\ B^+ \to J/\psi[\mu^+\mu^-]K^+\pi^+\pi^-,$ $B^+ \to J/\psi[\mu^+\mu^-]\Lambda^0\overline{p},\ B_c^+ \to \psi(2S)[\mu^+\mu^-]\pi^+,$ $D^{*+} \to D^0[\mu^-\mu^+]\pi^+,\ D^{*+} \to D^0[K^-\pi^+\mu^-\mu^+]\pi^+,$ $D_s^+ \to \pi^+\mu^+\mu^-,\ B^0 \to K^{*0}\mu^+\mu^-,$ $B_s \to \mu\mu,\ \tau \to \mu^+\mu^-\mu^+,$ $B^0 \to D^+\mu\nu_\mu X,$ $\Lambda_b^0 \to \Lambda_c^+\mu X,\ B_s^0 \to D_s^+\mu X,$ $B_s^0 \to D_s^+\tau[\mu\nu\nu],$ $B^+ \to \rho\mu\nu_\mu$	Same mechanism as `alpha_hadron`
`twotrack_ks`	`Hlt1TwoTrackKs`	$B^0 \to J/\psi[\mu^+\mu^-]K_{\mathrm{S}}^0,\ B^0 \to J/\psi[e^+e^-]K_{\mathrm{S}}^0,$ $B^+ \to \overline{D}^0[K_{\mathrm{S}}^0\pi^+\pi^-]K^+,\ B^0 \to K_{\mathrm{S}}^0\pi^+\pi^-,$ $B^0 \to \eta'[\pi^+\pi^-\gamma]K_{\mathrm{S}}^0,\ D^{*+} \to D^0[K_{\mathrm{S}}^0\pi^+\pi^-]\pi^+,$ $D^{*+} \to D^0[K_{\mathrm{S}}^0K_{\mathrm{S}}^0]\pi^+,\ D_s^+ \to K_{\mathrm{S}}^0K^+$	Exclusive K_s line with threshold requirements on transverse momentum, impact parameter significance, K_s transverse momentum, K_s pseudorapidity and the combined impact parameter.
`single_highpt` (MeV)	`Hlt1SingleHighPt Electron` `Hlt1SingleHighPtMuon`	$W^+ \to e^+\nu,\ Z^0/\gamma^* \to e^-e^+,$ $gg \to t\bar{t},\ \Upsilon(1S) \to e\mu,$ $\Upsilon(1S) \to ee,\ B_s^0 \to J/\psi[\mu^+\mu^-]\phi,$ $W^+ \to \mu^+\nu,\ Z^0/\gamma^* \to \mu^-\mu^+$	Direct leptonic high p_t requirements.

Samples, Lines, Parameters Used

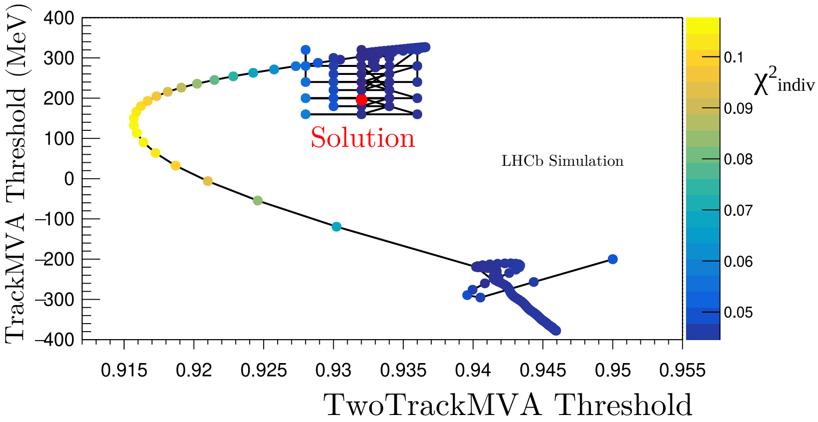
Parameter(s)	Trigger Line(s)	Channel(s)	Mechanism

`charm_track_pt` (MeV)
`charm_track_ip` (mm)
	`Hlt1D2PiPi`
	`Hlt1D2KPi`
	`Hlt1D2KK`
		$D^{*+} \to D^0[\pi^-\pi^+]\pi^+$,
		$D^{*+} \to D^0[K^-\pi^+]\pi^+$,
		$D^{*+} \to D^0[K^-K^+]\pi^+$
			Direct requirements on track transverse momentum, track impact parameter and $D^0{}_{\mathrm{ct}}$.

`highmass_dimuon_pt` (MeV)	`Hlt1DiMuonHighMass`	$B_s^0 \to J/\psi[\mu^+\mu^-]\phi,\ B^0 \to J/\psi[\mu^+\mu^-]K_\mathrm{S}^0$, $B^+ \to J/\psi[\mu^+\mu^-]K^+$, $\Lambda_b^0 \to J/\psi[\mu^+\mu^-]pK^-,\ J/\psi \to \mu^+\mu^-$, $B^+ \to J/\psi[\mu^+\mu^-]K^+,\ B^+ \to J/\psi[\mu^+\mu^-]K^+\pi^+\pi^-$, $B^+ \to J/\psi[\mu^+\mu^-]\Lambda^0\overline{p},\ B_c^+ \to \psi(2S)[\mu^+\mu^-]\pi^+$, $B^0 \to K^{*0}\mu^+\mu^-,\ B_s \to \mu\mu\Upsilon(1S) \to \mu\mu$,	Direct requirement on high-mass dimuon transverse momentum.
`displaced_dielectron_pt` `displaced_dielectron_ipchi2`	`Hlt1DisplacedDielectron`	$B^0 \to J/\psi[e^+e^-]K_\mathrm{S}^0,\ D^{*+} \to D^0[K^-\pi^+e^-e^+]\pi^+$, $D_s^+ \to \pi^+e^+e^-,\ B^0 \to K^{*0}e^+e^-$, $B_s \to e^+e^-$,	Direct requirements on displaced dielectron transverse momentum, and the corresponding impact parameter significance
`displaced_dimuon_pt` `displaced_dimuon_ipchi2`	`Hlt1DiMuonLowMass`	$B^0 \to K^{*0}\mu^+\mu^-,\ B_s \to \mu\mu$, $\Upsilon(1S) \to \mu\mu,\ \tau \to \mu^+\mu^-\mu^+$	Direct requirements on displaced dimuon transverse momentum, and the corresponding impact parameter significance
`highmass_diphoton_et`	`Hlt1DiPhotonHighMass`	$B_s^0 \to \gamma\gamma$	Requirement on minimum tranverse photon energy.
`lambda_detached_track_mipchi2` `lambda_detached_combination_bpvfd` (mm)	`Hlt1LambdaLLDetachedTrack`	$\Lambda_c^+ \to \Xi^-[\Lambda\pi^-]K^+\pi^+$ $\Xi_c^0 \to \Xi^-[\Lambda\pi^-]\pi^+\pi^+$	Direct requirements on minimum impact parameter significance and B meson primary vertex flight-distance.
`xi_lll_ipchi2`	`Hlt1XiOmegaLLL`	$\Lambda_c^+ \to \Xi^-[\Lambda\pi^-]K^+\pi^+$ $\Xi_c^0 \to \Xi^-[\Lambda\pi^-]\pi^+\pi^+$	Direct requirement on long track Ξ baryon impact parameter significance.

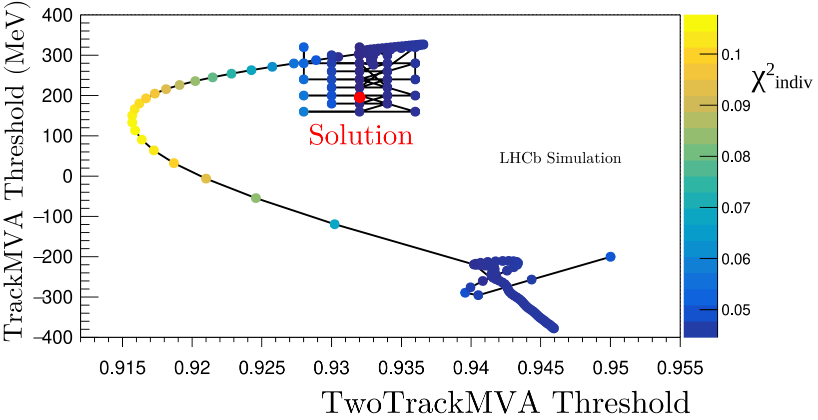
Convergence Rate

- Initially, with out-of-the-box Adam, a significant percentage of minimisations failed to converge - evident when the tool produced values of $\epsilon_i^{\mathbf{max}} < \epsilon_i^{\mathbf{optimal}}$
- Convergence rate depended on **stability of lines** selecting a signal channel, **resolution** chosen for thresholds and **statistics limitations** of each dataset.
- **Artificially multiplying moments** at first epoch and warm restart, **reflecting Adam moment at parameter boundaries**, and use of grid search mechanism increased both the parameter space searched and convergence rate.



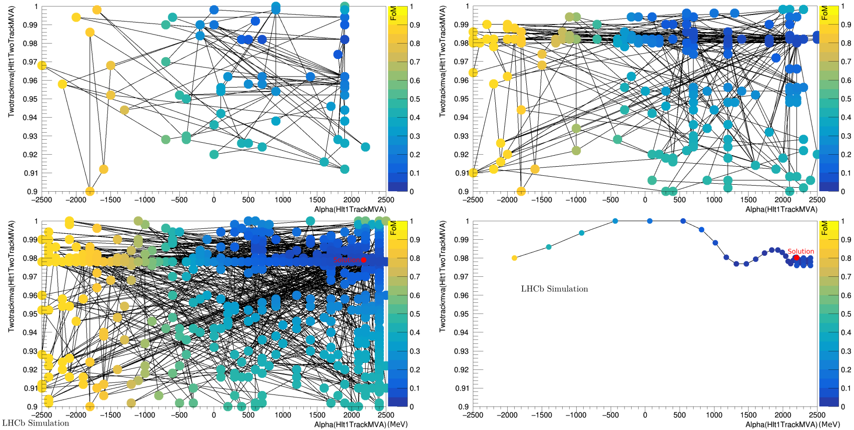
Convergence Rate (2)

- After running many minimisations of χ^2_{indiv} from random seeds, $\Delta\epsilon$ between ϵ^{best} and $\bar{\epsilon}$, averaged over all samples was **0.9%**.
- Analogous procedure for χ^2_{global} resulted in average efficiency loss of 0.3% between $\chi^{2,\text{best}}_{\text{global}}$ and $\bar{\chi}^2_{\text{global}}$, which was generally $\sim \boldsymbol{\sigma(\epsilon)}$.



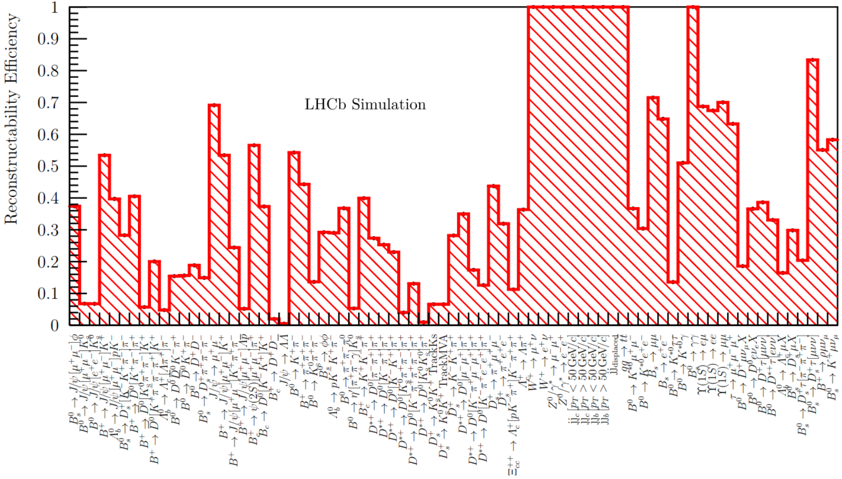
Genetic Algorithm vs. Adam Minimisation

- For larger number of dimensions (χ^2_{global}), performance of Adam and GA was similar, due to extra time discretising continuous thresholds in Adam.
- Adam preferred to GA since GA requires careful choice of hyperparameters to balance runtime vs. convergence rate.
- Adam also produces more monotonic thresholds when varying the ratelimit between divisions.



2024 Reconstructability Efficiencies

- The efficiency shown in this plot is the ratio of total reconstructible events, over the total number of **reconstructible and non-reconstructible** events entering HLT1.
- For electroweak samples, jets, or samples triggered only by the calorimeter, there is no existing method to determine the reconstructability of tracks and events. Therefore, there has been **no requirement** applied to these samples, and the reconstructability efficiency is 100%.



New Features and Challenges

A more complex problem:

- During 2011-2018 data taking period hardware trigger cut on five **simple ADC thresholds** to reject events and distribute the bandwidth.
- During 2023-2030 HLT1 employs **multivariate selections** to reject **candidates**, with multiple cuts chained to make a decision on an event.
- Each channel can be tuned by **multiple parameters**.
- Parameters can be shared across **multiple lines** to ensure selections are consistent between control/signal modes.

New Features and Challenges (2)

- Many more lines to optimise over more signal modes of interest.
- Larger **computational complexity** - requires a more efficient program. Need to modernise and optimise code.
- Often **multiple candidates** per event - need to collate candidates and store unique events.
- This might not be suited to ROOT's data structure due to **row dependence** for multiple-candidate events. Current plan is to use ROOT for fast I/O, but for efficiency calculations use **customised** data structure.