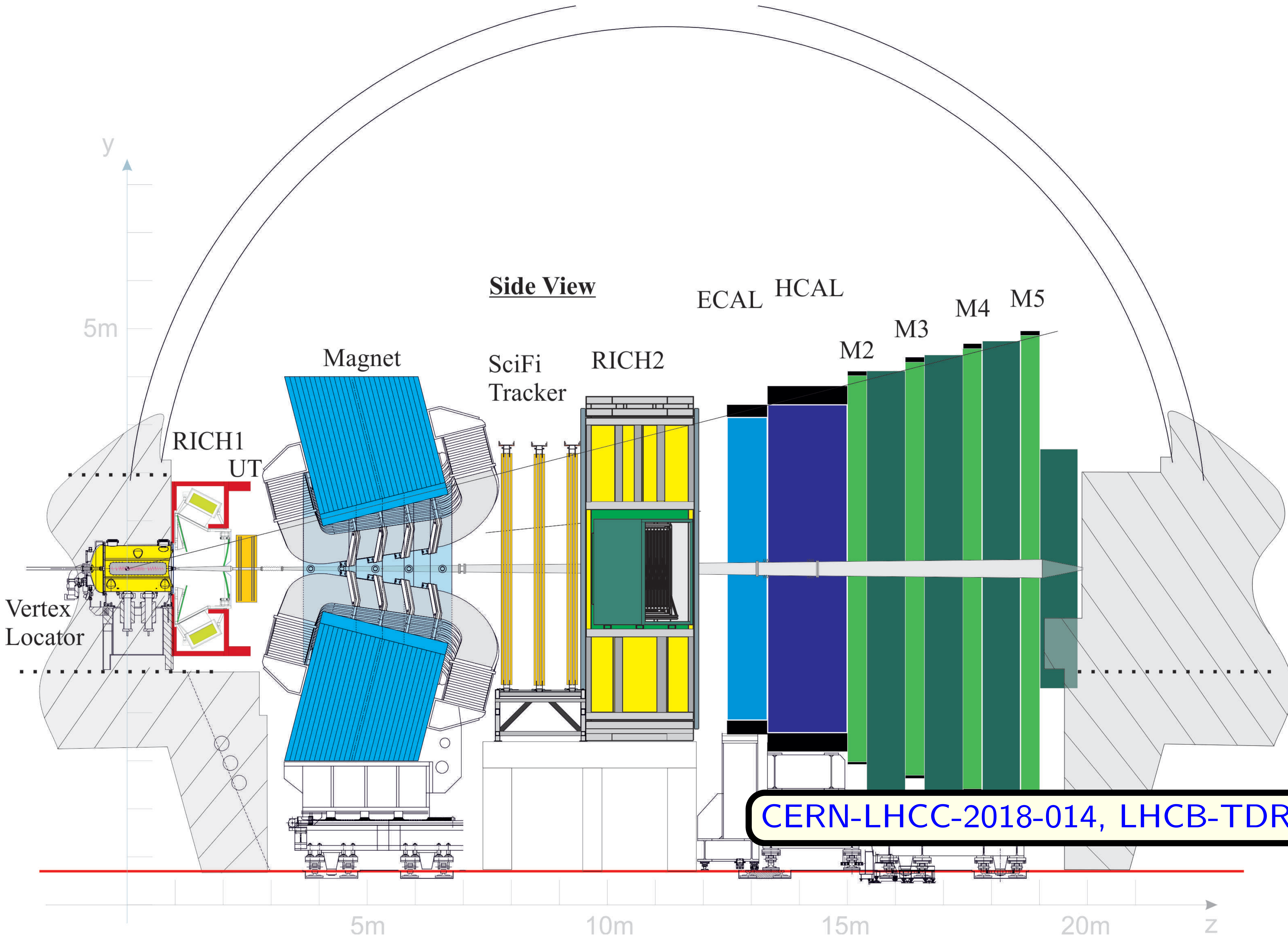
Applications of Lipschitz neural networks to the Run 3 LHCb trigger

26th International Conference on Computing in High Energy & Nuclear Physics
Norfolk, VA · May 11th, 2023

Blaise Delaney, on behalf of the LHCb Collaboration

The LHCb detector in Run 3 @ the LHC

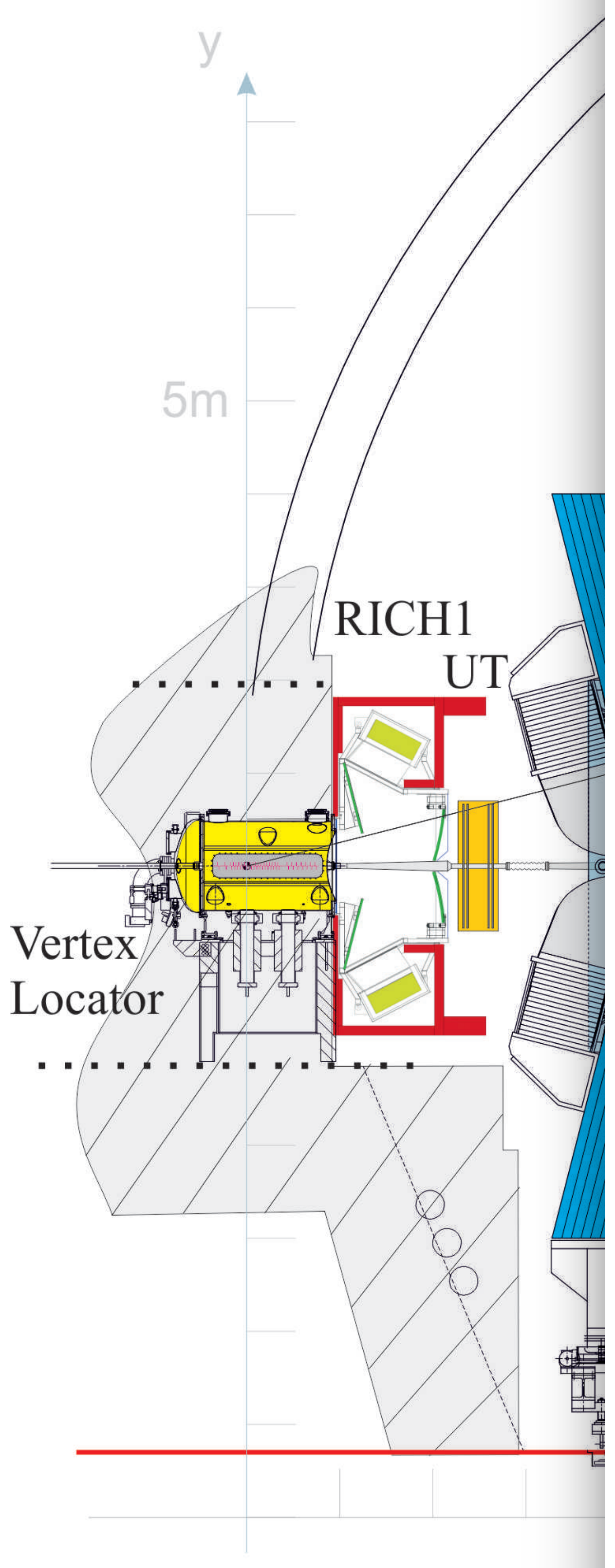


CERN-LHCC-2018-014, LHCB-TDR-018

- Forward-arm spectrometer instrumented for the study of b and c hadrons
- **Run 3**: unprecedented conditions:
  - instantaneous $\mathcal{L} = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$ $\rightarrow 5 \times$ Run 2
  - redesigned tracking & electronics @ pp bunch crossing rate of 30 MHz
  - *milestone:* fully software trigger

See talks by {LHCb colleagues}

The LHCb detector in Run 3 @ the LHC



The Run 3 LHCb trigger system

CERN-LHCC-2018-014, LHCB-TDR-018

*

LHCb raw data
15000 PB/year

500× reduction

LHCb storage capacity
30 PB/year

⇒ LHCb trigger: *real-time* data reduction: 5 TB/s → 10 GB/s

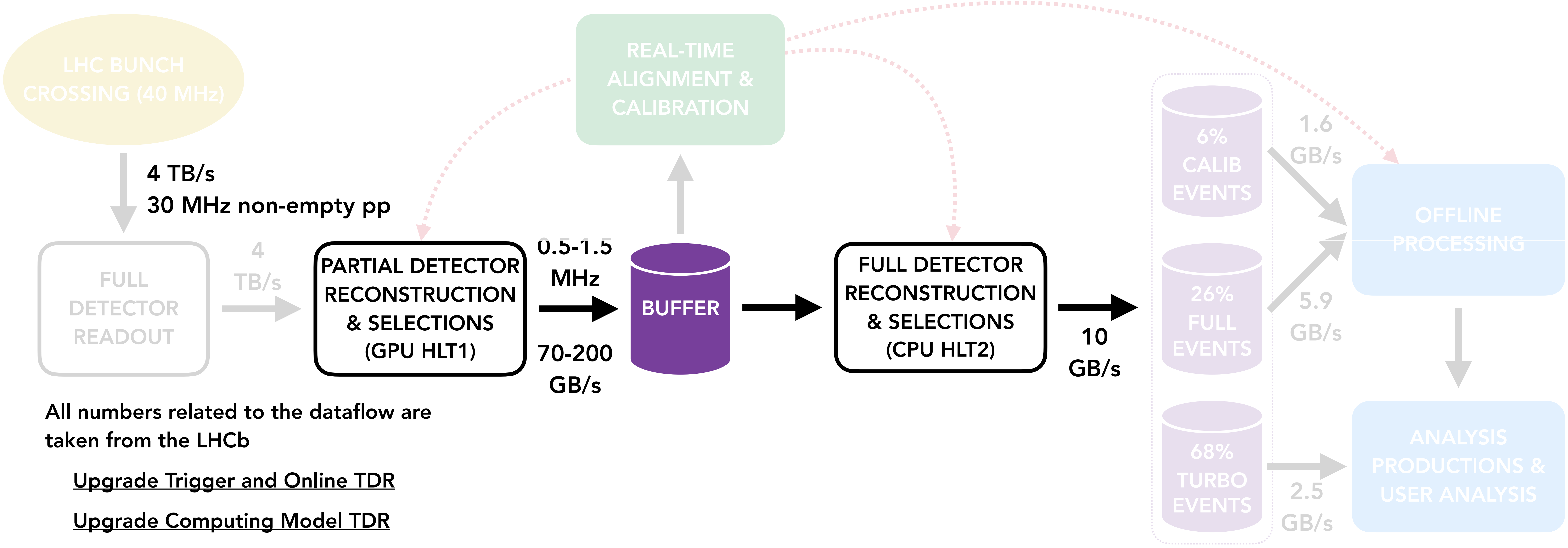
*image not to scale

$\mathrm{m}^{-2}\mathrm{s}^{-1}$

See talks by {LHCb colleagues}

The Run 3 LHCb Trigger

LHCb-FIGURE-2020-016



Fully software, two-staged trigger: select only *interesting* events

⇒ combination of *expert systems* and *machine learning*

See talks by {LHCb colleagues}

Monotonic Lipschitz Neural Networks

NeurIPS ML4PS 21
arXiv:2112.00038

No margin for error in the trigger system demands *effective* discriminators capable of

Monotonic Lipschitz Neural Networks

NeurIPS ML4PS 21
arXiv:2112.00038

No margin for error in the trigger system demands *effective* discriminators capable of

- **robustness**, *i.e.* mitigated sensitivity to
  a) experimental instabilities during data taking
  b) deficiencies in simulation
  $\Rightarrow$ constrain the Lipschitz constant, λ , of the model $g(\boldsymbol{x}) : |g(\boldsymbol{x}) - g(\boldsymbol{y})| \leq \lambda ||\boldsymbol{x} - \boldsymbol{y}||_1, \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

Monotonic Lipschitz Neural Networks

NeurIPS ML4PS 21
arXiv:2112.00038

No margin for error in the trigger system demands *effective* discriminators capable of

- **robustness**, *i.e.* mitigated sensitivity to

  a) experimental instabilities during data taking

  b) deficiencies in simulation

  ⇒ constrain the Lipschitz constant, λ , of the model $g(\boldsymbol{x}) : |g(\boldsymbol{x}) - g(\boldsymbol{y})| \leq \lambda ||\boldsymbol{x} - \boldsymbol{y}||_1, \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

- **interpretability**: built in inductive bias "*the higher the momentum & longer the lifetime, the better*"

  ⇒ NN architecture monotonic wrt a (sub)set of input features

Monotonic Lipschitz Neural Networks

NeurIPS ML4PS 21
arXiv:2112.00038

No margin for error in the trigger system demands *effective* discriminators capable of

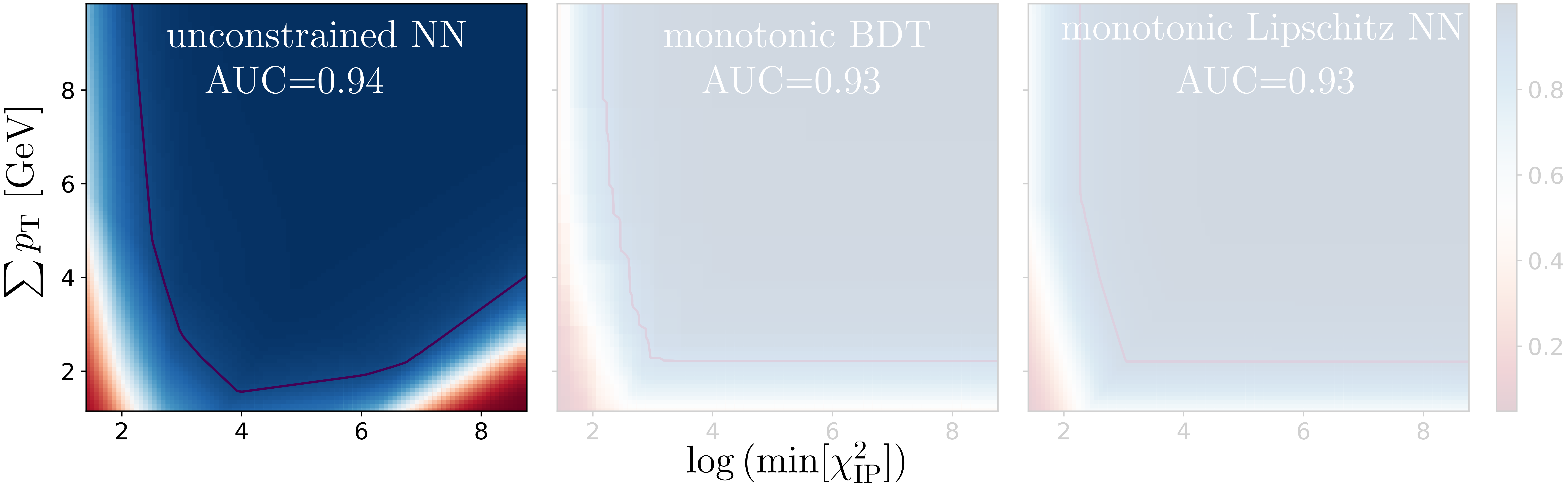
- **robustness**, *i.e*
  a) experimenta
  b) deficiencies
  ⇒ constrain th $\boldsymbol{y}||_1, \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$
- **interpretabili** *time, the better"*
  ⇒ NN architecture monotonic wrt a (sub)set of input features

∴ Expressive *shallow* architectures meeting memory and compute requirements of the LHCb trigger

Monotonic Lipschitz Neural Networks

Exempli gratia: simplified HLT1 inclusive heavy-flavor trigger

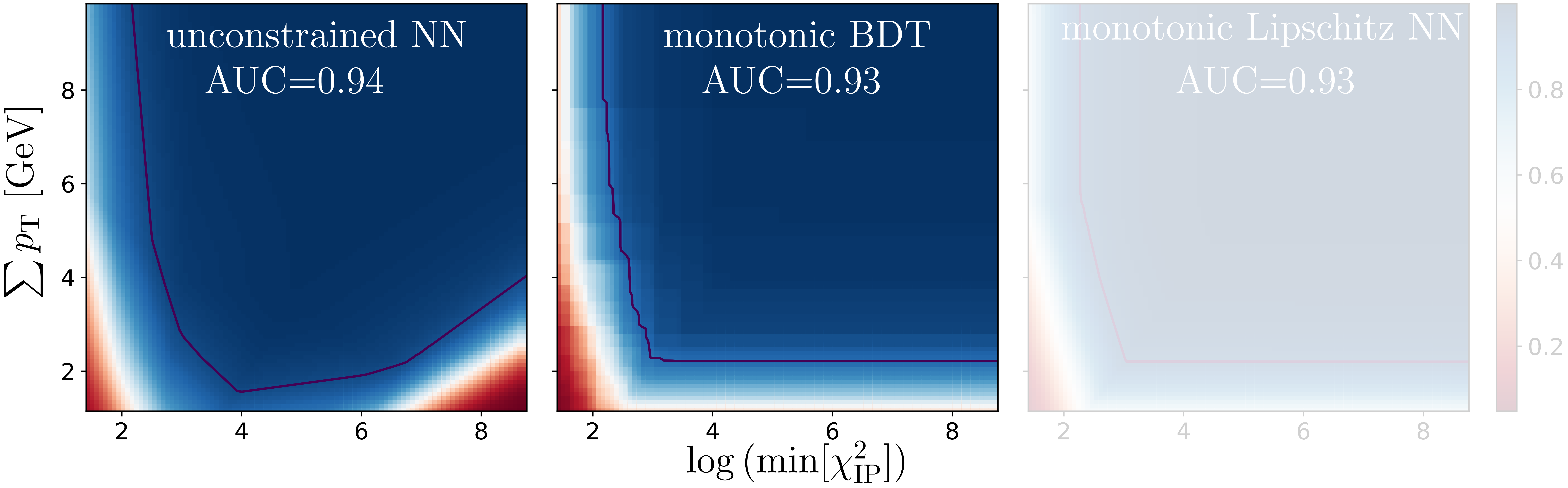
NeurIPS ML4PS 21 arXiv:2112.00038



Monotonic Lipschitz Neural Networks

Exempli gratia: simplified HLT1 inclusive heavy-flavor trigger

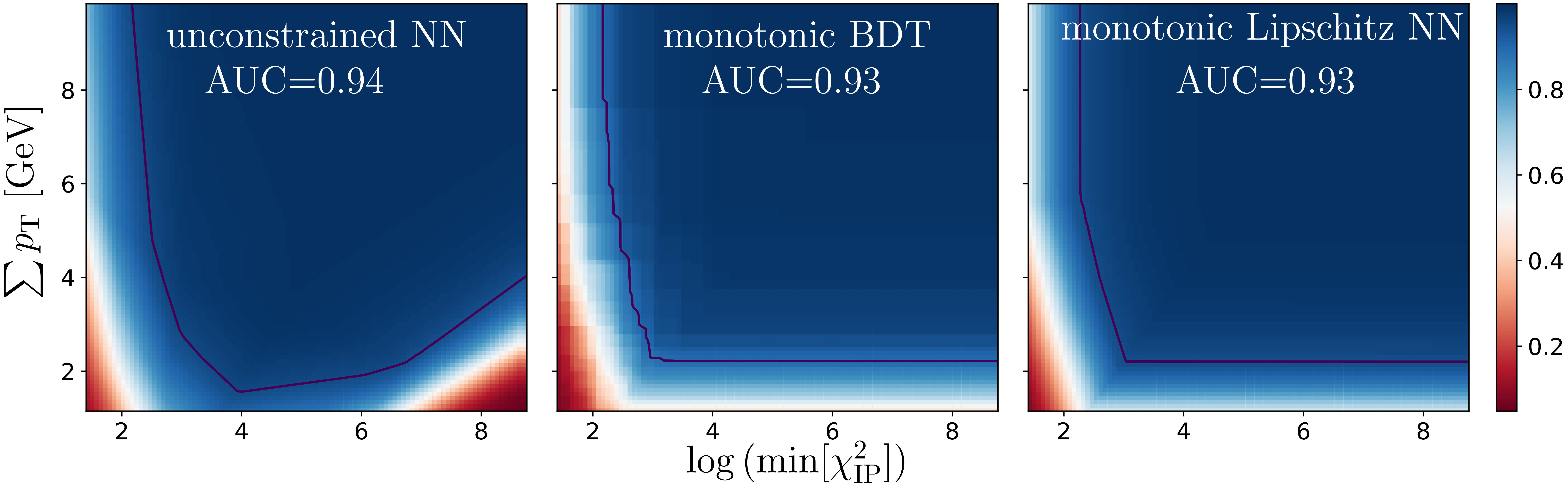
NeurIPS ML4PS 21 arXiv:2112.00038



Monotonic Lipschitz Neural Networks

Exempli gratia: simplified HLT1 inclusive heavy-flavor trigger

NeurIPS ML4PS 21 arXiv:2112.00038

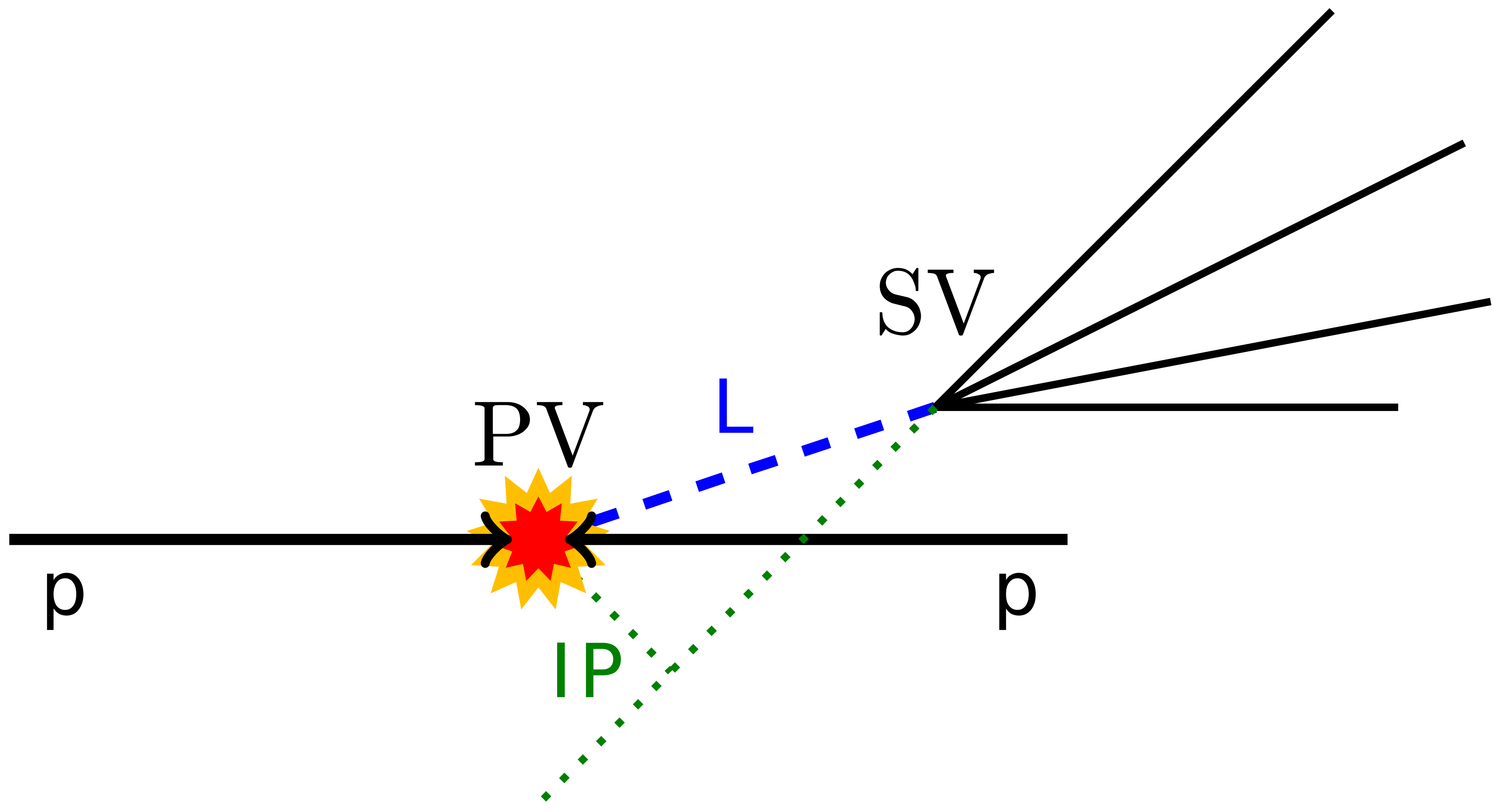


The LHCb Topological Triggers

- Higher-level (HLT2) trigger for *inclusive* selection of B decays

$\Rightarrow$ select multi-body candidates with b-hadron decay topologies:

- B mass $\mathcal{O}(5\text{ GeV}) \rightarrow$ high transverse momentum, p_T
- Lifetime of $\mathcal{O}(1\text{ ps}) \rightarrow$ displaced decay vertex
- Boosted in forward direction $\rightarrow$ $\mathcal{O}(1\text{ cm})$ before decay vertex (DV)



Beauty decay topology

Credit: Vom Bruch, Vistas on Detector Physics, Heidelberg

The HLT2 Topological Triggers in Run 3

Run 3: Monotonic Lipschitz NN ⇒ identify 2- and 3-body *b*-candidates using

- Kinematics
- Decay topology
- Increasing monotonic wrt

  a) candidate p_T

  b) candidate flight distance

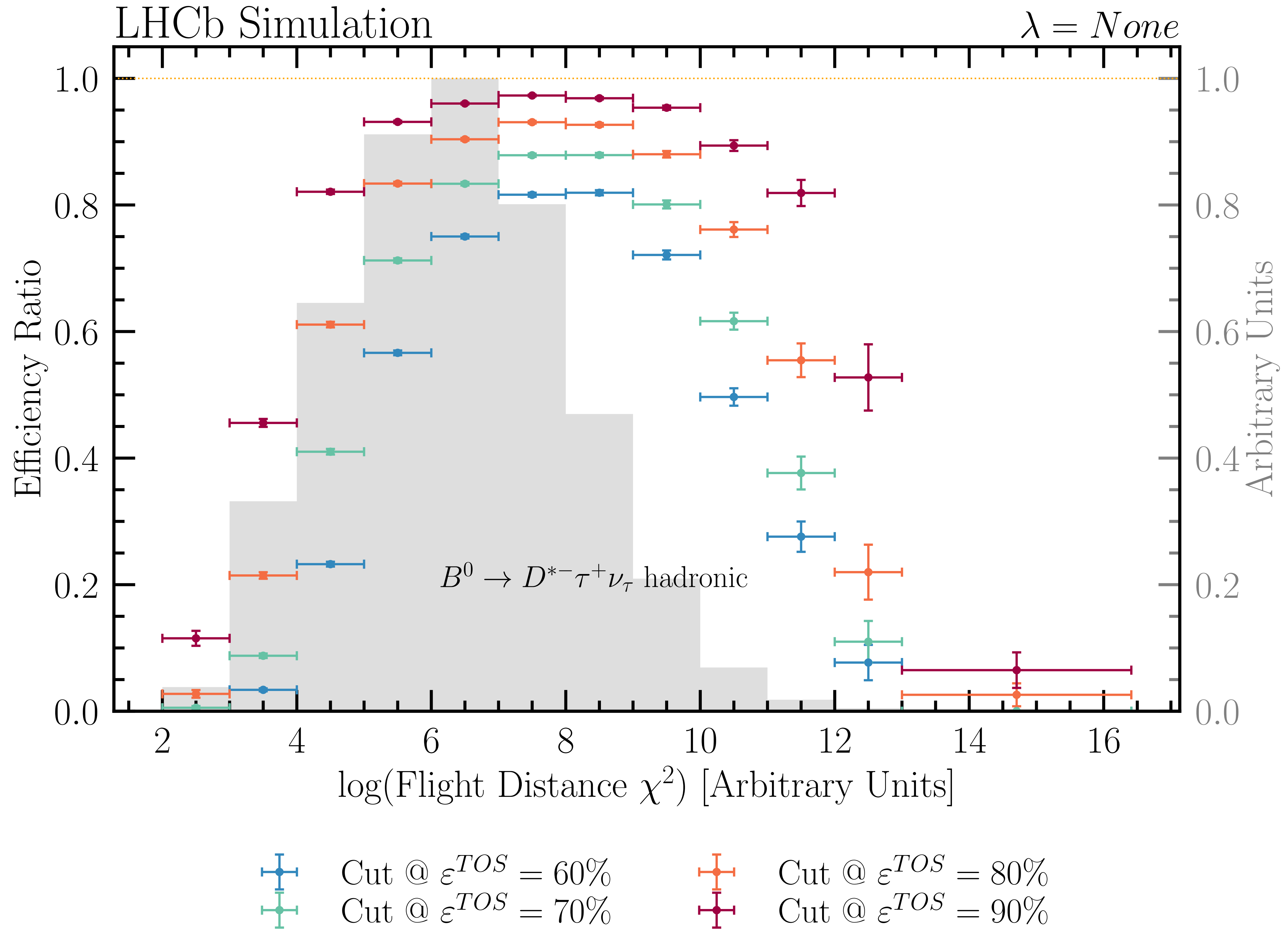
  c) χ^2 of the impact parameter (IP)

Sensitivity to:

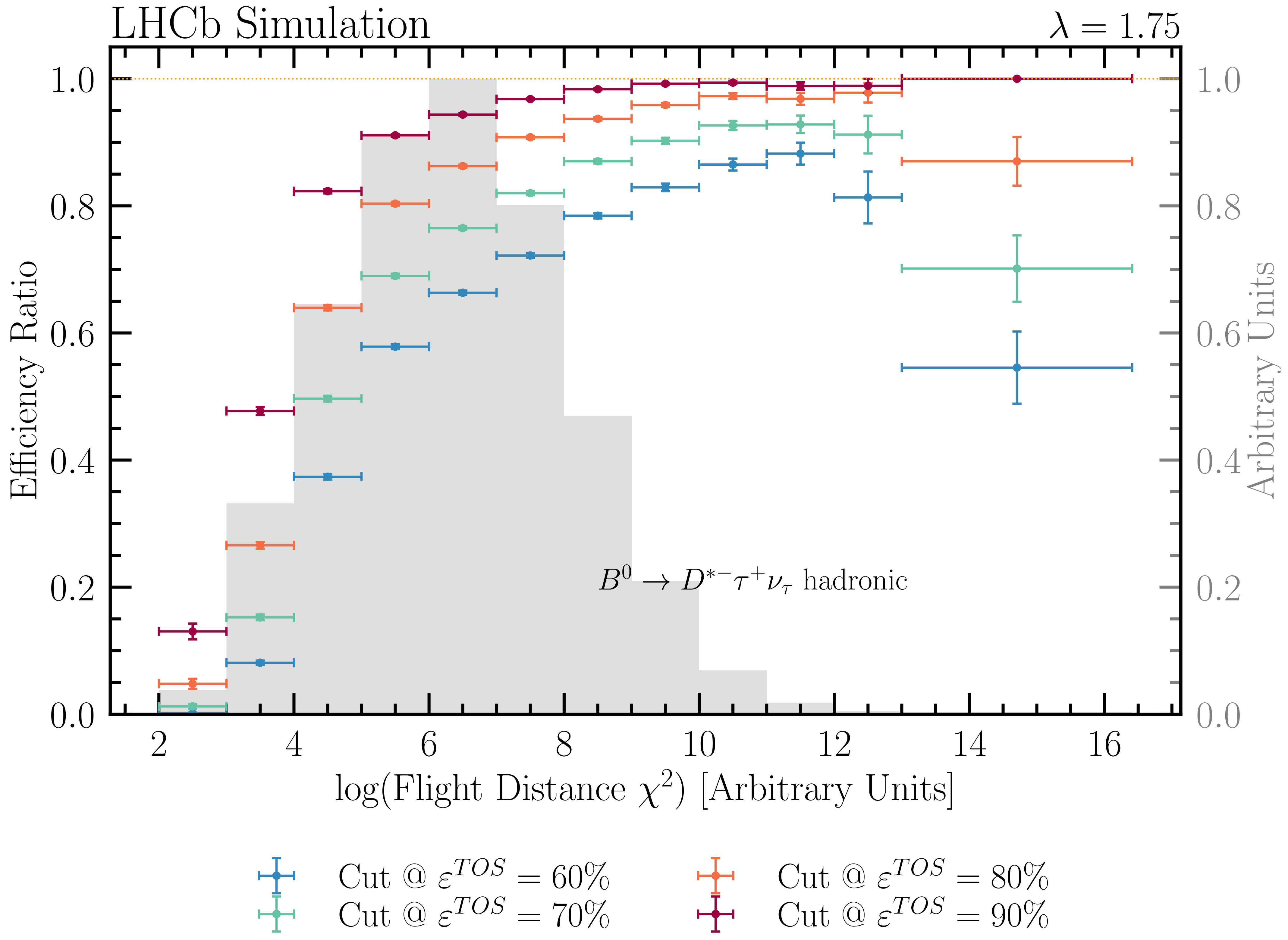
- Beauty candidates
- Potential feebly interacting BSM

The HLT2 Topological Triggers in Run 3

N. Schulte, J. Albrecht, BD, N. Nolte, M. Williams @ ACAT 2022



Unconstrained NN



Lipschitz monotonic NN

Summary & Outlook

Extensive application of Lipschitz monotonic NNs in the Run 3 LHCb trigger:

- Select inclusively heavy-flavor decays
- Enhanced sensitivity to long-lived feebly interacting BSM particles
- Planned applications to tracking, electron ID & ghost rejection

Beyond LHCb:

- Applications to medicine, criminal justice [ICLR 23] & collider phenomenology [arXiv:2209.15624v1]

Appendix

Monotonic Lipschitz Neural Networks

NeurIPS ML4PS 21
arXiv:2112.00038

- Goal: NN representing $f(\boldsymbol{x}) \ : \ \mathbb{R}^n \rightarrow \mathbb{R}$ monotonic wrt feature (sub)set and with bound gradient wrt inputs in any direction
- Lip[1] model $g(\boldsymbol{x}) \ : \ |g(\boldsymbol{x}) - g(\boldsymbol{y})| \leq \lambda ||\boldsymbol{x} - \boldsymbol{y}||_1, \ \forall \ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$
- Monotonicity wrt to features x_i via residual connection $f(\boldsymbol{x}) = g(\boldsymbol{x}) + \lambda \sum_{i \in I} x_i$

  ⇒ monotonicity defined via partial derivative

$$\frac{\partial f}{\partial x_i} = \frac{\partial g}{\partial x_i} + \lambda \geq 0, \ \forall \ i \in I$$

while keeping $x_{\neq i}$ *constant*

Monotonic Lipschitz Neural Networks

NeurIPS ML4PS 21
arXiv:2112.00038

- Goal: $g(\boldsymbol{x})$ a universal approximator of Lip[1] functions
- Given the fully connected NN

$$g(\boldsymbol{x}) = W^m \sigma(W^{m-1}\sigma(...\sigma(W^1 \boldsymbol{x} + b^1)...) + b^{m-1}) + b^m,$$

robustness (aka bounded λ) achieved if

$$\prod_{i=0}^{m} \|W^i\|_1 \leq \lambda$$

and σ has Lipschitz constant less than or equal to 1

Monotonic Lipschitz Neural Networks

NeurIPS ML4PS 21
arXiv:2112.00038

- Goal: $g(\boldsymbol{x})$ a universal approximator of Lip[1] functions
- Given the fully connected NN

$$g(\boldsymbol{x}) = W^m \sigma(W^{m-1} \sigma(\ldots \sigma(W^1 \boldsymbol{x} + b^1)\ldots) + b^{m-1}) + b^m,$$

robustness (aka bounded λ) achieved if

$$\prod_{i=0}^{m} \|W^i\|_1 \leq \lambda$$

$$W^i \to W'^i = W^i \text{diag}\left(\frac{1}{\max\left(1, \lambda^{-1/m} \sum_j |W^i_{jk}|\right)}\right).$$

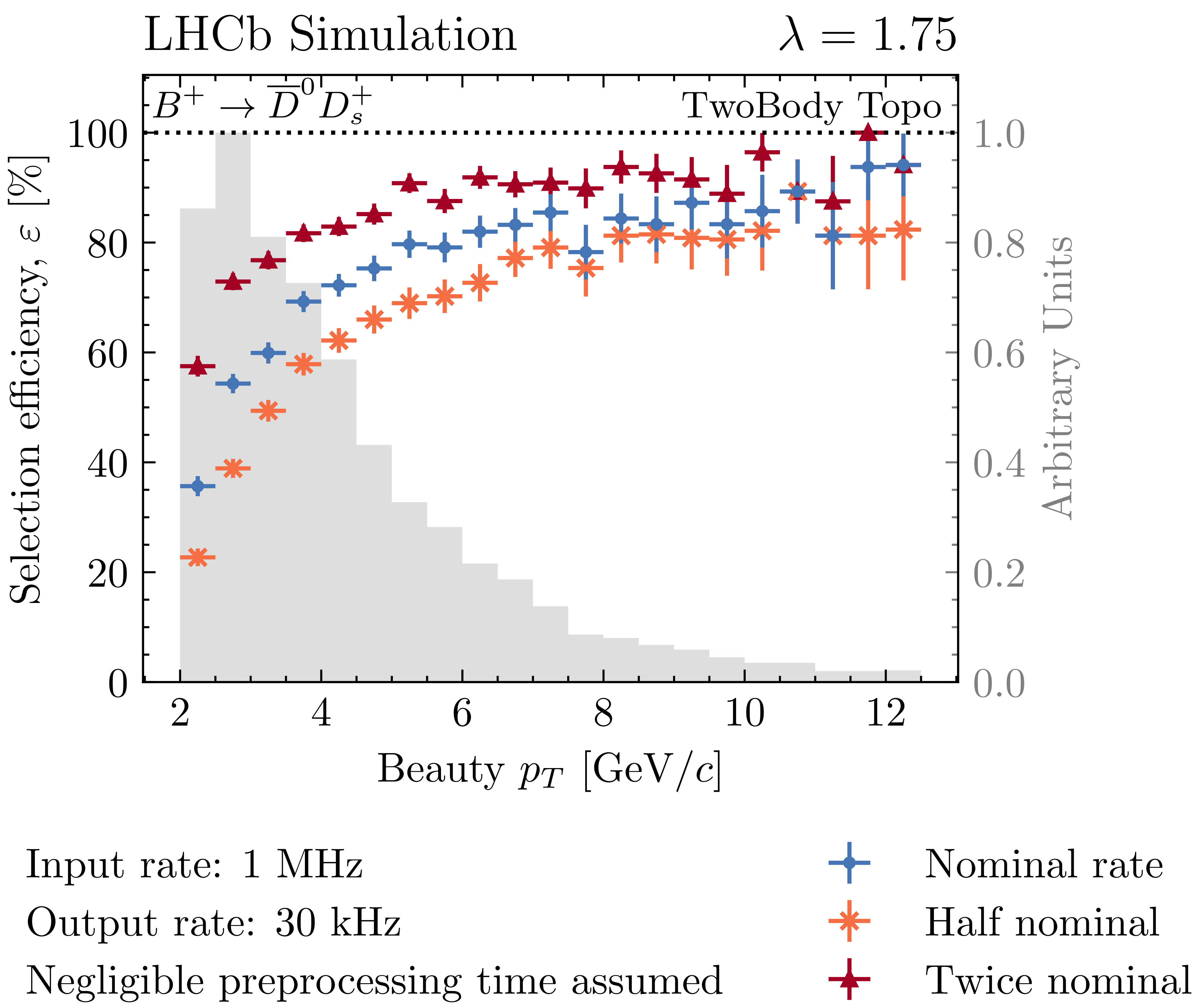
and σ has Lipschitz constant less than or equal to 1

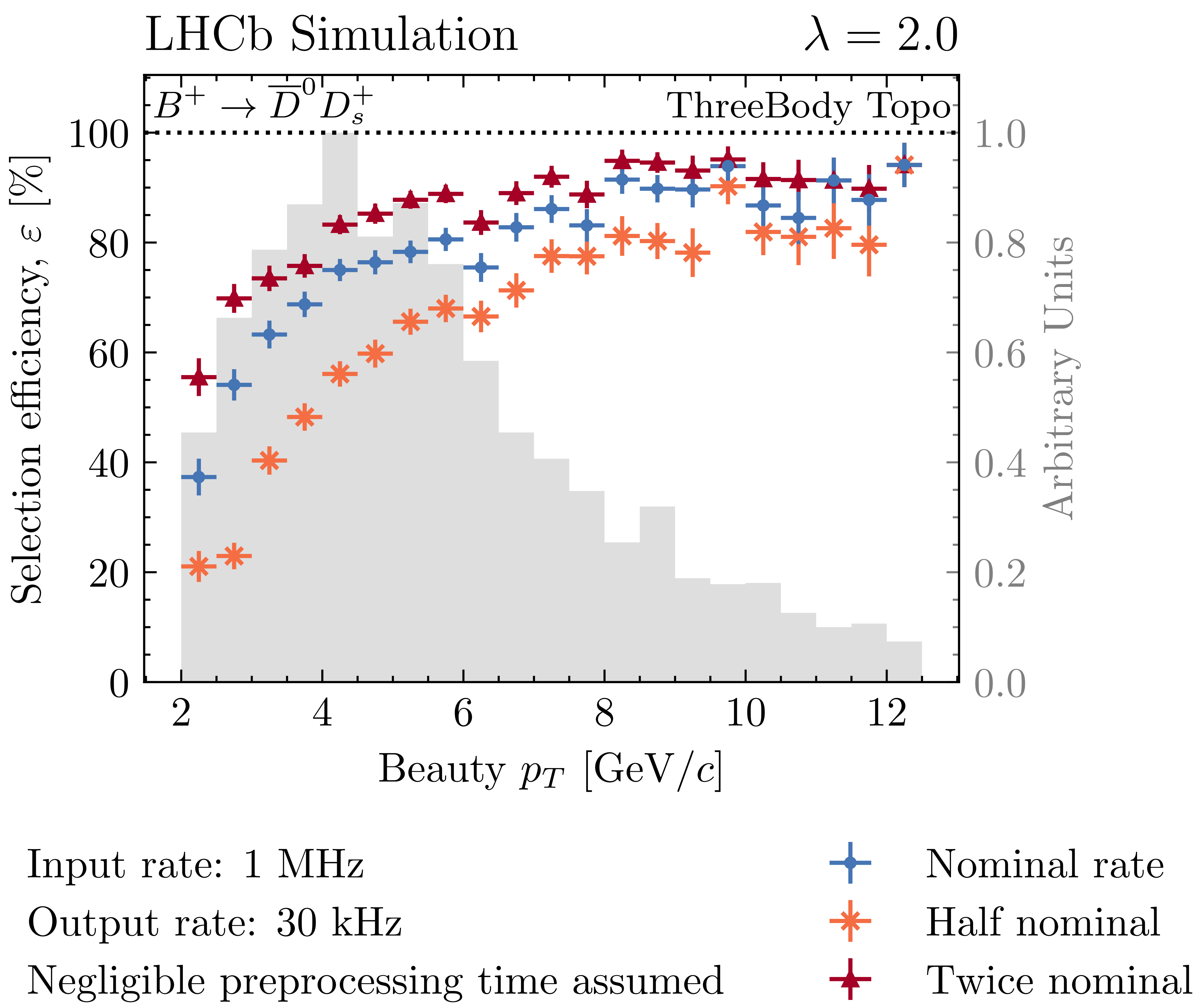
$$\sigma = \textbf{GroupSort}$$

Anil, Cem, James Lucas, and Roger Grosse. "Sorting out Lipschitz function approximation." *International Conference on Machine Learning*. PMLR, 2019.

The HLT2 Topological Triggers in Run 3

N. Schulte, J. Albrecht, BD, N. Nolte, M. Williams @ ACAT 2022





The HLT2 Topological Triggers in Run 3

N. Schulte, J. Albrecht, BD, N. Nolte, M. Williams @ ACAT 2022

TwoBody Features	ThreeBody Features
$\min\left(p_{\text{T, FS particles (1,2)}}\right)$	$\min\left(p_{\text{T, FS particles (1,2,3)}}\right)$
$\text{sum}\left(p_{\text{T, FS particles (1,2)}}\right)$	$\text{sum}\left(p_{\text{T, FS particles (1,2,3)}}\right)$
$p_{\text{T, B-Hadron}}$	$p_{\text{T, B-Hadron}}$
$\log\left(\min(\chi^2_{\text{IP, FS particles (1,2)}})\right)$	$\log\left(\min(\chi^2_{\text{IP, FS particles (1,2,3)}})\right)$
$\log\left(\max(\chi^2_{\text{IP, FS particles (1,2)}})\right)$	$\log\left(\max(\chi^2_{\text{IP, FS particles (1,2,3)}})\right)$
$\log\left(\chi^2_{\text{FD, B-Hadron}}\right)$	$\log\left(\chi^2_{\text{FD, B-Hadron}}\right)$
$\log\left(\chi^2_{\text{Vertex, B-Hadron}}\right)$	$\log\left(\chi^2_{\text{Vertex, B-Hadron}}\right)$
$\text{DOCA}\,(\text{B-Hadron})$	$\text{DOCA}\,(\text{B-Hadron})$
	$\min\left(p_{\text{T, FS particles (1,2)}}\right)$
	$\text{sum}\left(p_{\text{T, FS particles (1,2)}}\right)$
	$\text{DOCA}\,(\text{TwoBody})$
	$\log\left(\chi^2_{\text{FD, TwoBody}}\right)$
	$\log\left(\chi^2_{\text{Vertex, TwoBody}}\right)$
	$\log\left(\chi^2_{\text{IP, TwoBody}}\right)$
	$p_{\text{T, TwoBody}}$

The HLT2 Topological Triggers in Run 3

N. Schulte, J. Albrecht, BD, N. Nolte, M. Williams @ ACAT 2022

Training Samples

$\Lambda_b^0 \to (\Lambda_c(2625)^+ \to (\Lambda_c^+ \to p^+ K^- \pi^+)\pi^+\pi^-)\mu^-\overline{\nu}_\mu$	$B^+ \to \mu^+\mu^-\mu^+\nu_\mu$
$B+ \to (\overline{D}^0 \to K^+\pi^-)(K_S^0 \to \pi^+\pi^-)\pi^+$	$\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$
$B^0 \to (D^0 \to K^-\pi^+\pi^+\pi^-)(\overline{D}^0 \to K^+\pi^-)$	$B^+ \to K^+\mu^+\mu^-\mu^+\mu^-\mu^+\mu^-$
$B^+ \to (\overline{D}^0 \to K^+\pi^-)\pi^+\pi^-\pi^+$	$B_c^+ \to (J/\psi(1S) \to \mu^+\mu^-)\mu^+\nu_\mu$
$B_s^0 \to (D_s- \to K^+K^-\pi^-)\nu_\mu\mu^+$	$B_s^0 \to K^-\nu_\mu\mu^+$
$B^+ \to (D^{\star +}(2010) \to (D^0 \to K^-\pi^+)\pi^+)(\overline{D}^0 \to K^+\pi^-)$	$B_s^0 \to K^-(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$
$B^0 \to (D^{\star -} \to \pi^-(\overline{D}^0 \to K^+\pi^-))(\tau^+ \to \pi^+\pi^+\pi^-\overline{\nu}_\tau)\nu_\tau$	$B^+ \to \pi^+\pi^-K^+$
$B_c^+ \to (J/\psi(1S) \to \mu^+\mu^-)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	$B^0 \to (K^{\star 0}(892) \to K^+\pi^-)\gamma$
$B^- \to (D^0 \to K^-\pi^+)(\tau^- \to \mu^-\nu_\tau\overline{\nu}_\mu)\overline{\nu}_\tau$	$B^- \to p^+\overline{p}^-(\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau)\overline{\nu}_\tau$
$B^+ \to (\overline{D}^{\star 0} \to \pi^0((\overline{D}^0 \to K^+\pi^-))(\tau^+ \to \pi^+\pi^-\pi^+\overline{\nu}_\tau)\nu_\tau$	$B^- \to p^+\overline{p}^-\mu^-\overline{\nu}_\mu$