

Machine Learning at LHCb

Shanzhen Chen

18.09.2022

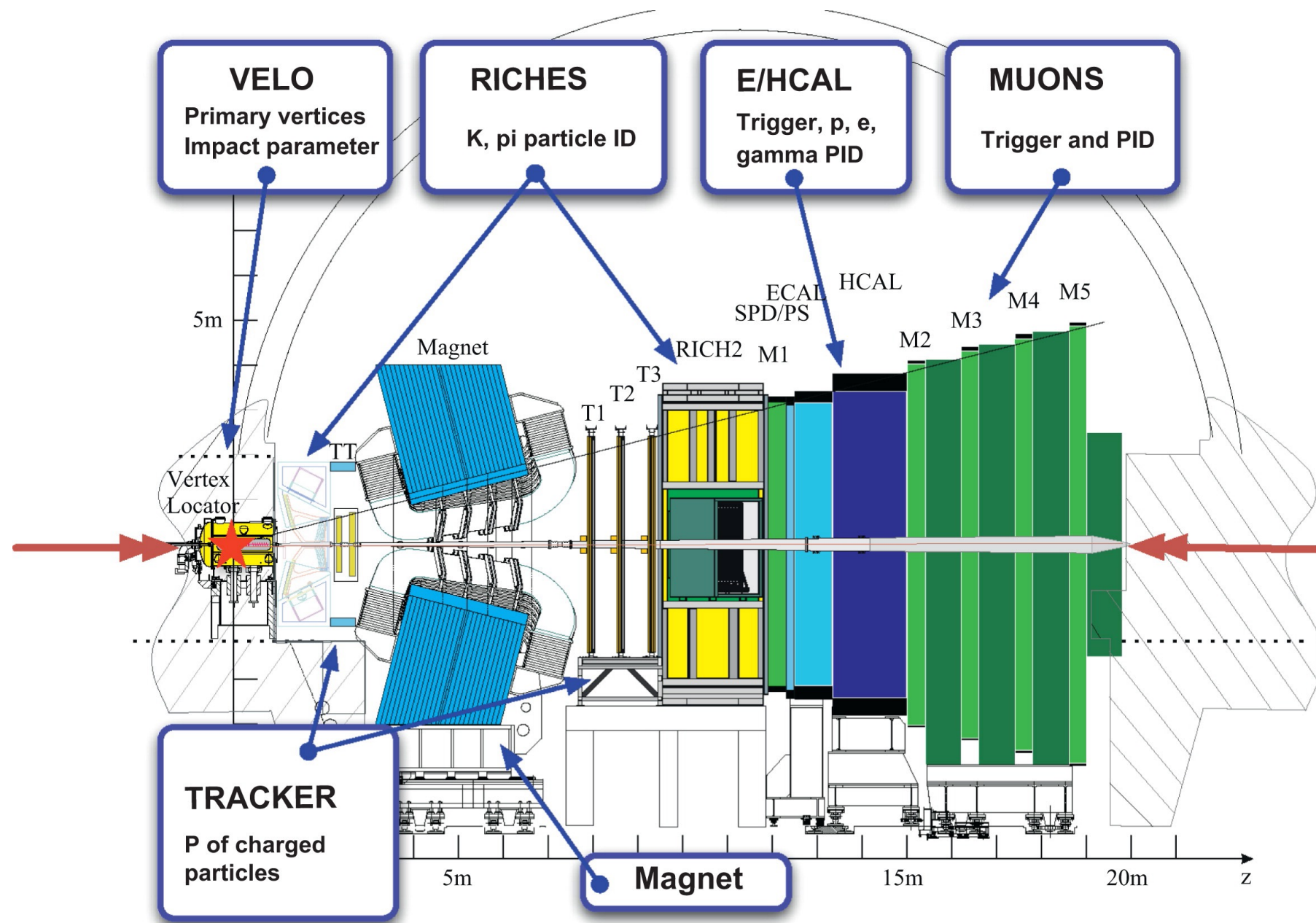
机器学习技术在高能物理中的应用研讨会

LHCb探测器

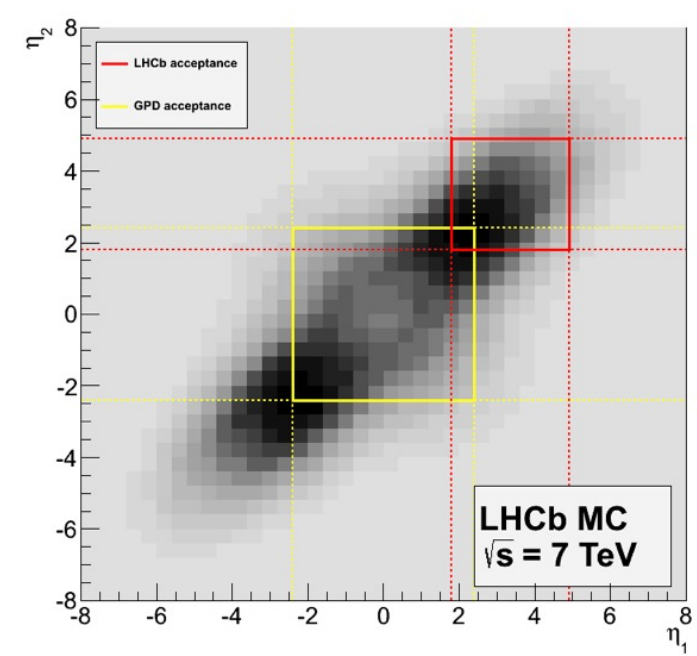
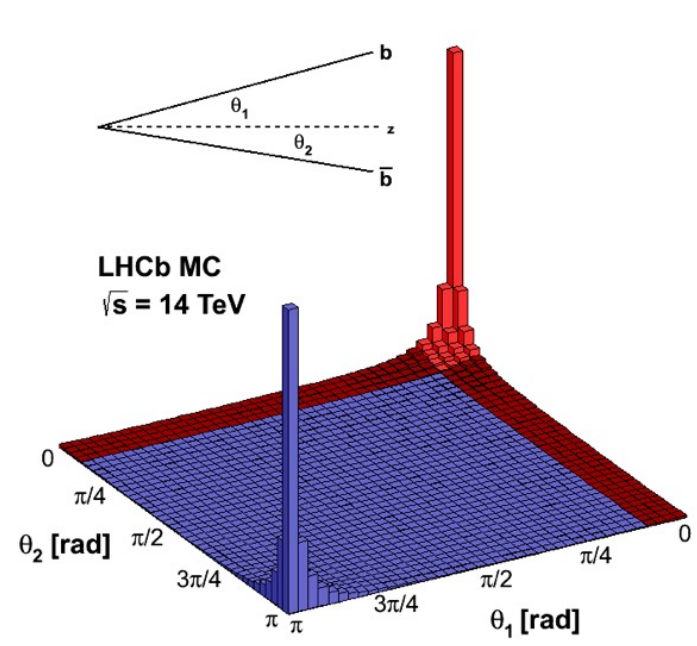
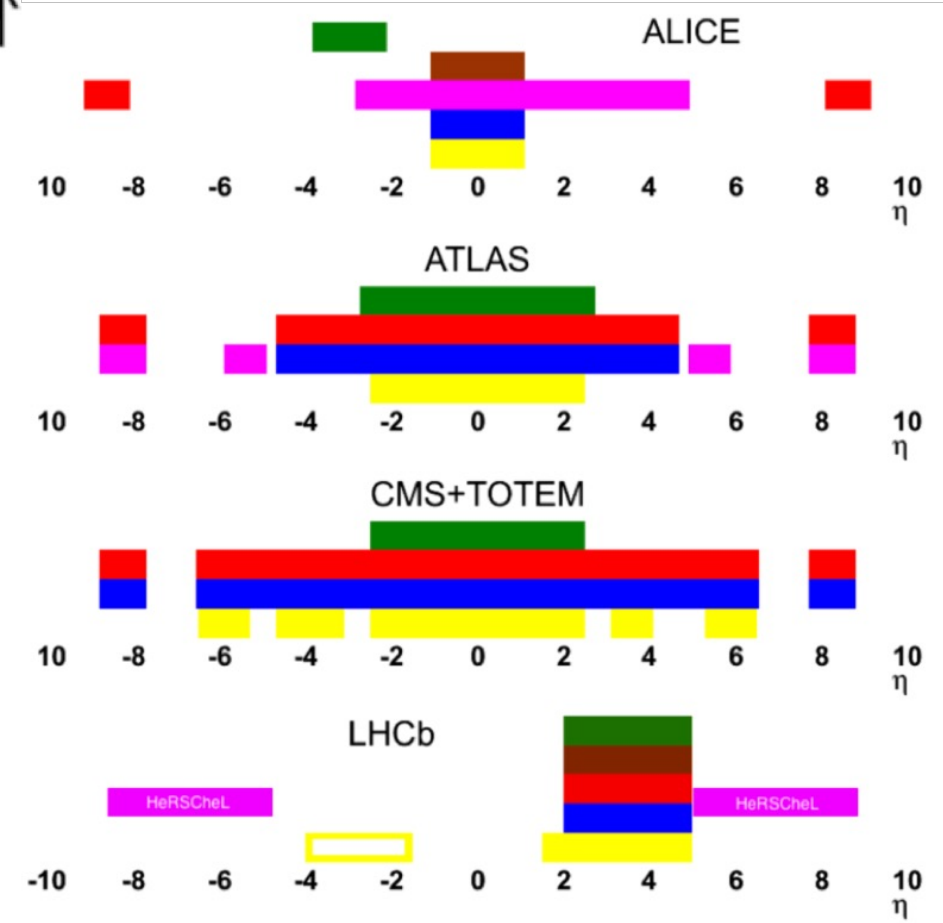
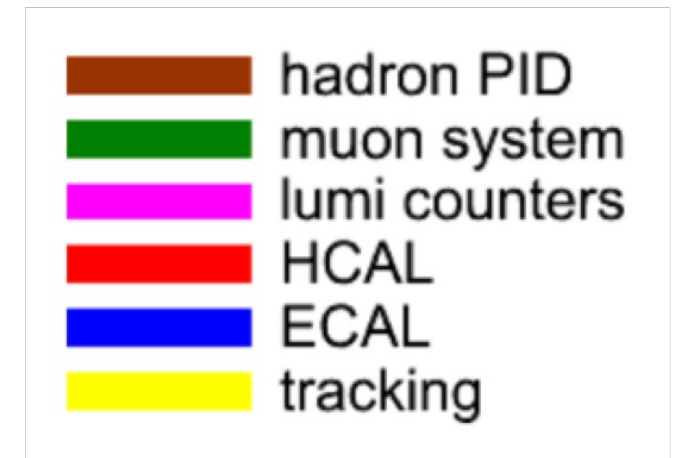
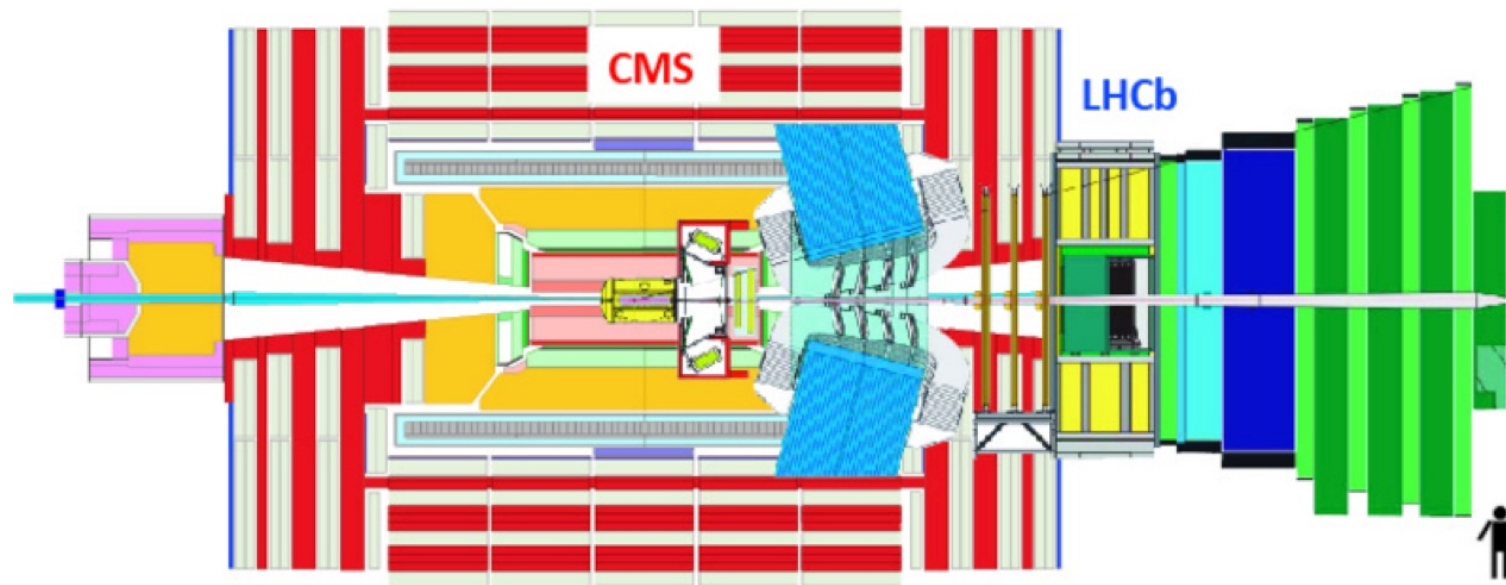
- LHCb探测器主要研究对象是重味物理，包括对含有b和c夸克的强子的CP破坏测量，稀有衰变寻找，以及新物理的搜寻。

- 探测器特点：

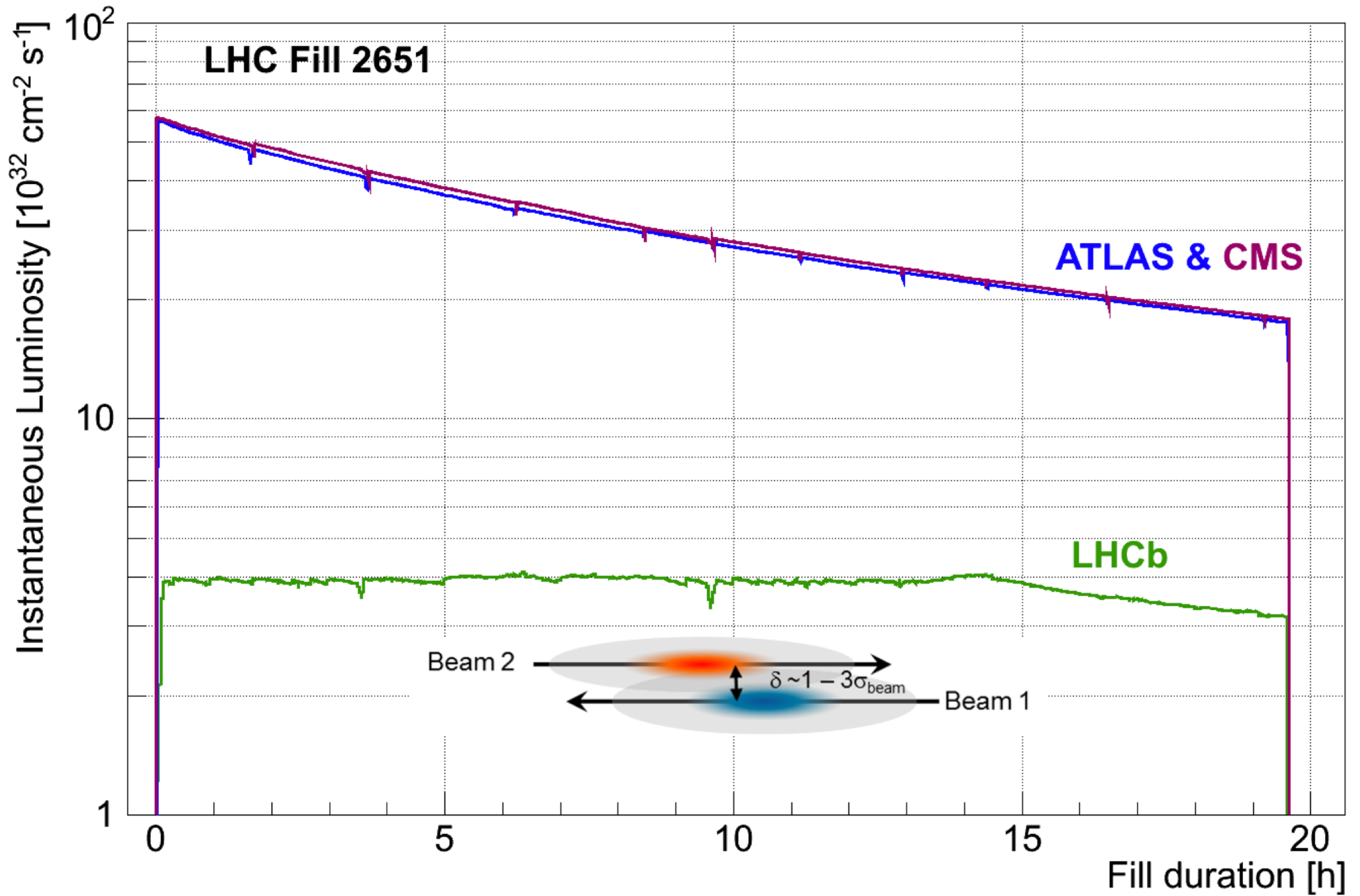
- 单臂前向探测器
- 高精度径迹系统
- 粒子鉴别系统
- 赝快度区间 $2 < \eta < 5$
- 所有系统赝快度区间重合



LHCb探测器与其它探测器的比较

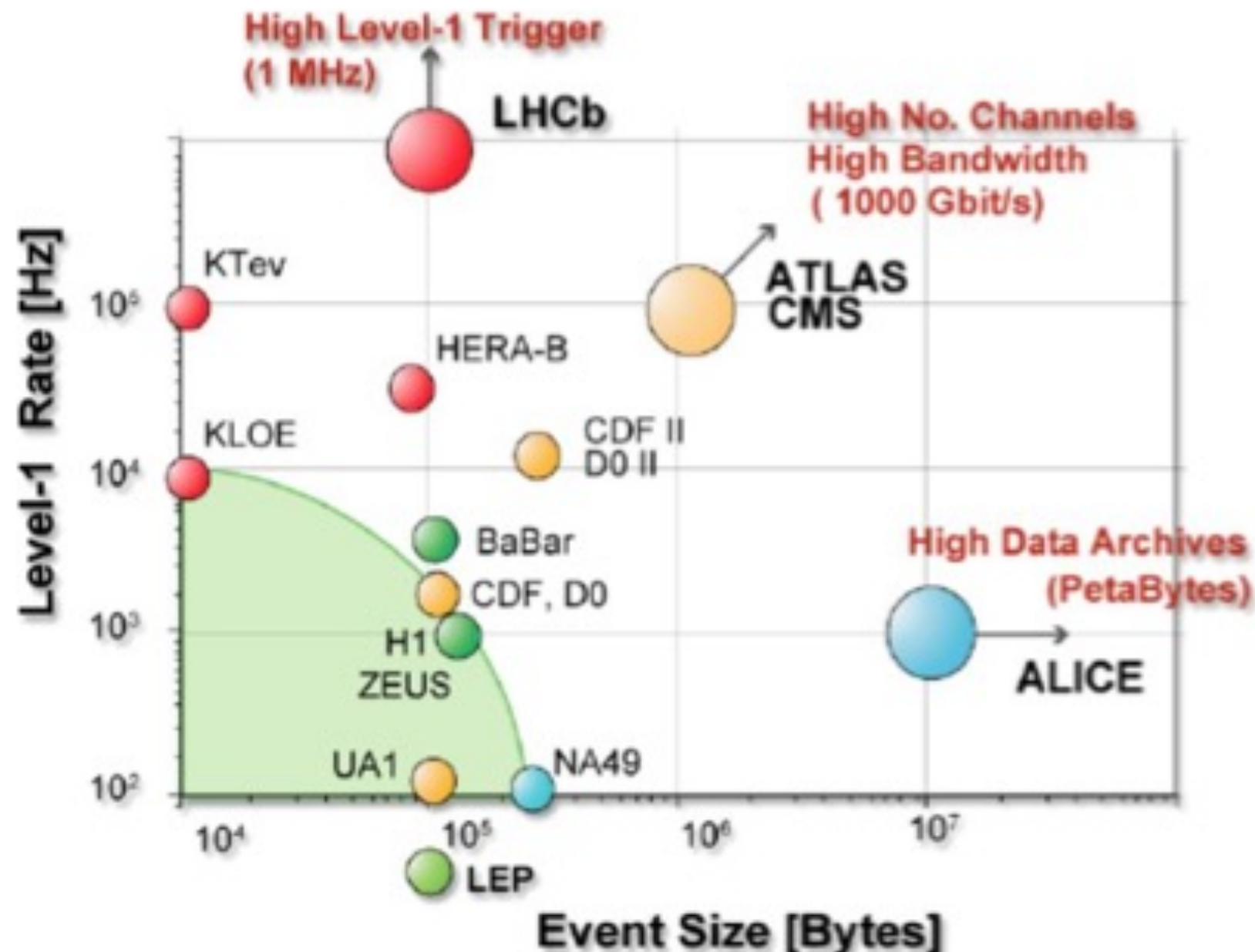


LHCb亮度leveling



触发速率与事件大小

- 不同实验研究物理目标不同，数据处理方式不同，数据大小与触发速率差异很大
- LHCb聚焦精确测量，触发速率高，统计量大

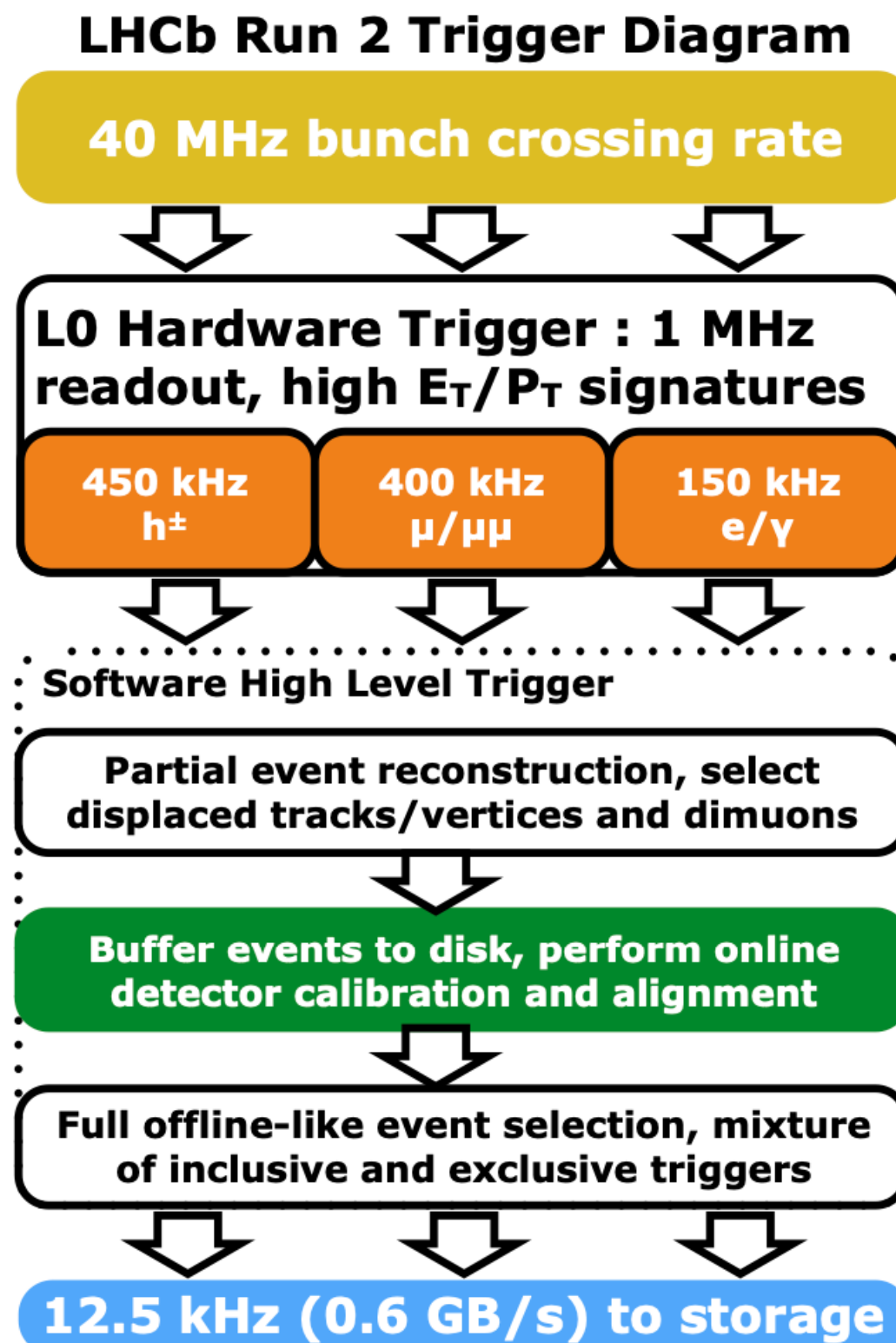


LHCb Run-II触发系统

- 重味物理事件特点：
高 p_T ，高 E_T ，衰变顶点位移
- LHCb物理需求：
高精度测量->高统计量
- Run-II：使用HLT1的数据，
对子探测器系统的实时
calibration和alignment

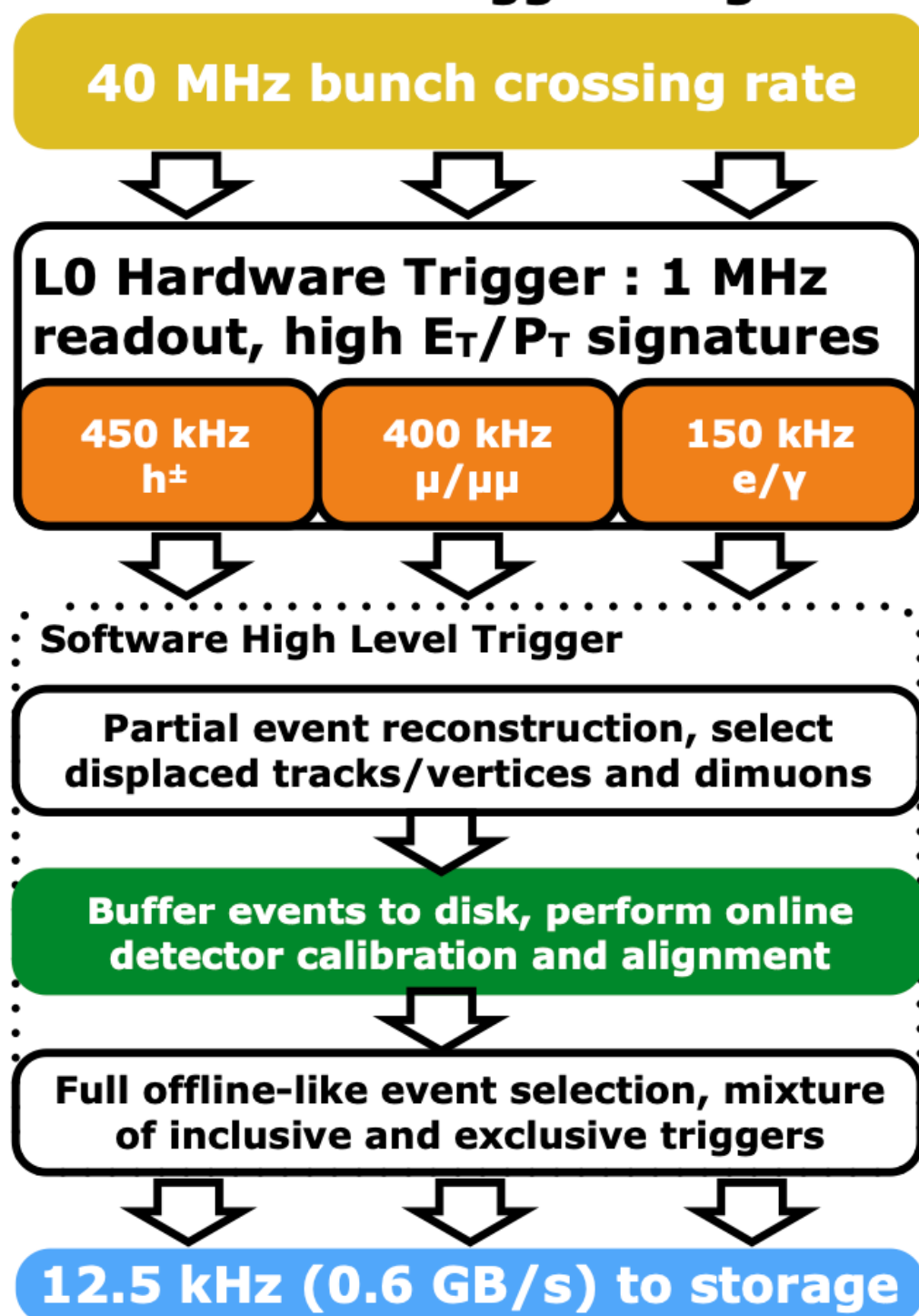
ML/DL应用

- 径迹重建
- 粒子鉴别
- High level trigger
- Alignment&Calibration
- 物理分析

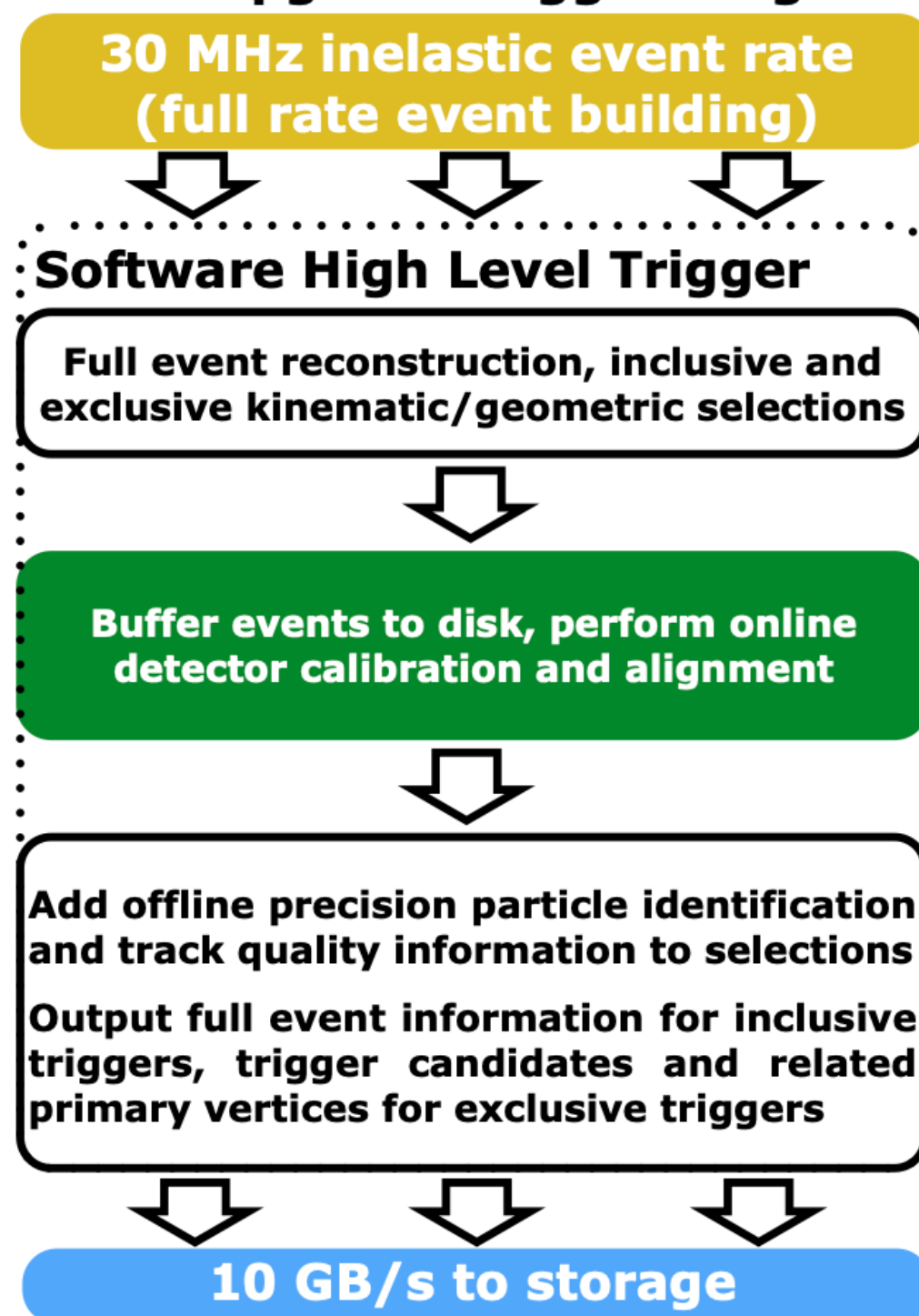


LHCb Run-II 触发系统

LHCb Run 2 Trigger Diagram

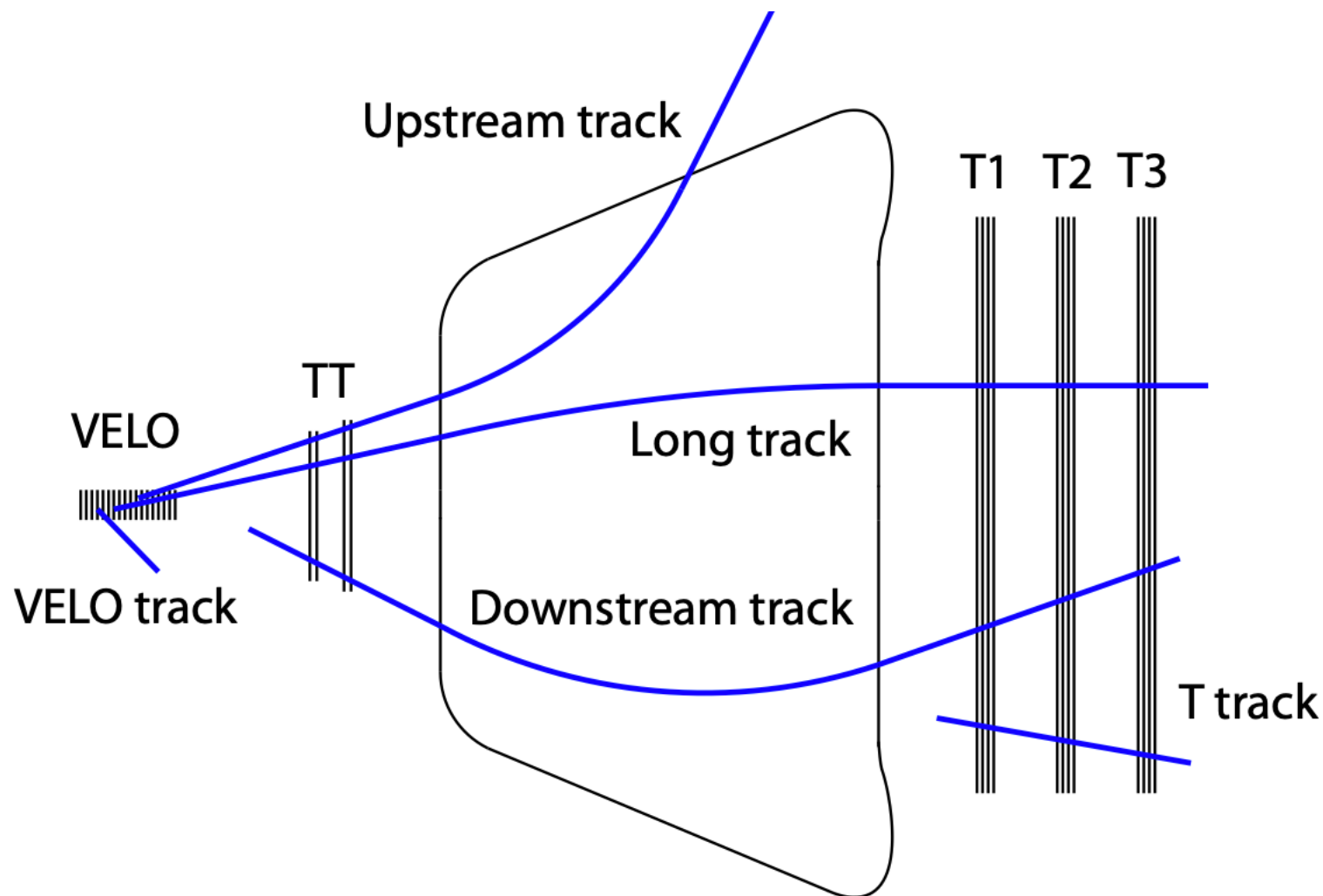


LHCb Upgrade Trigger Diagram



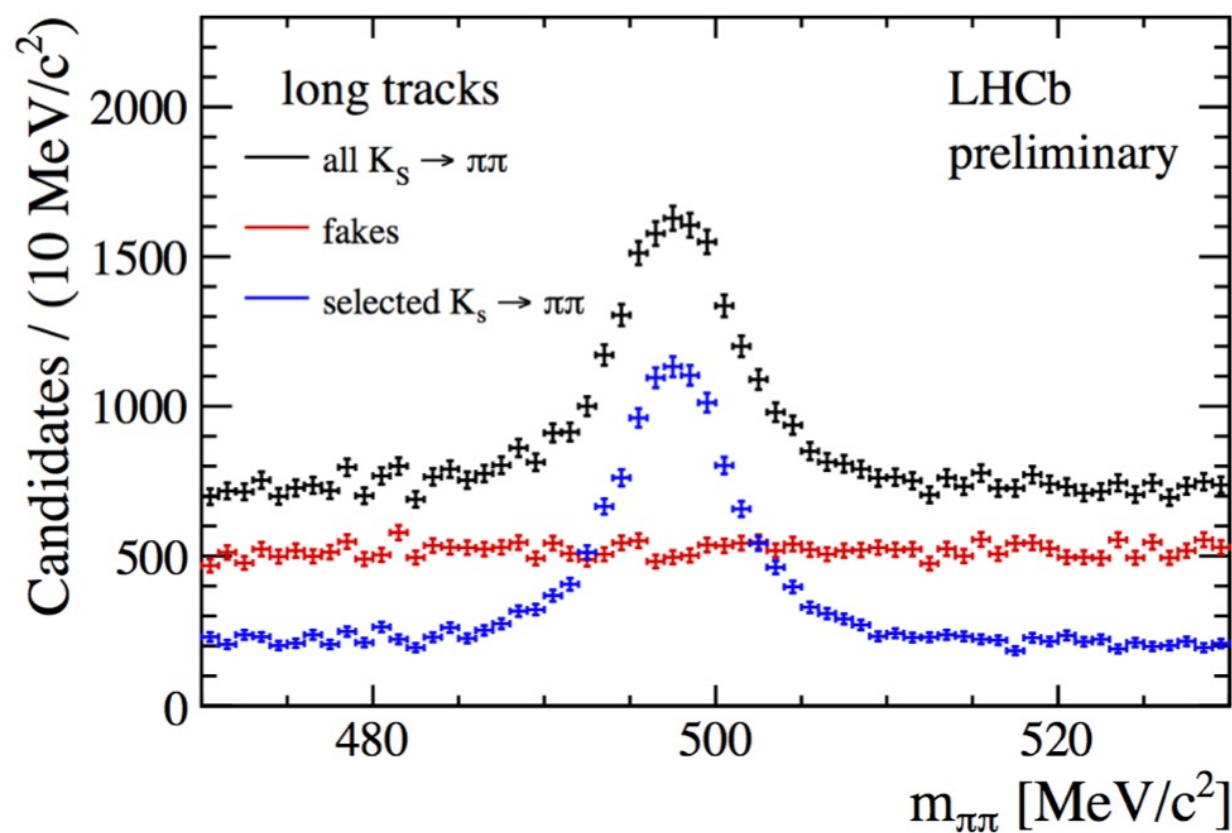
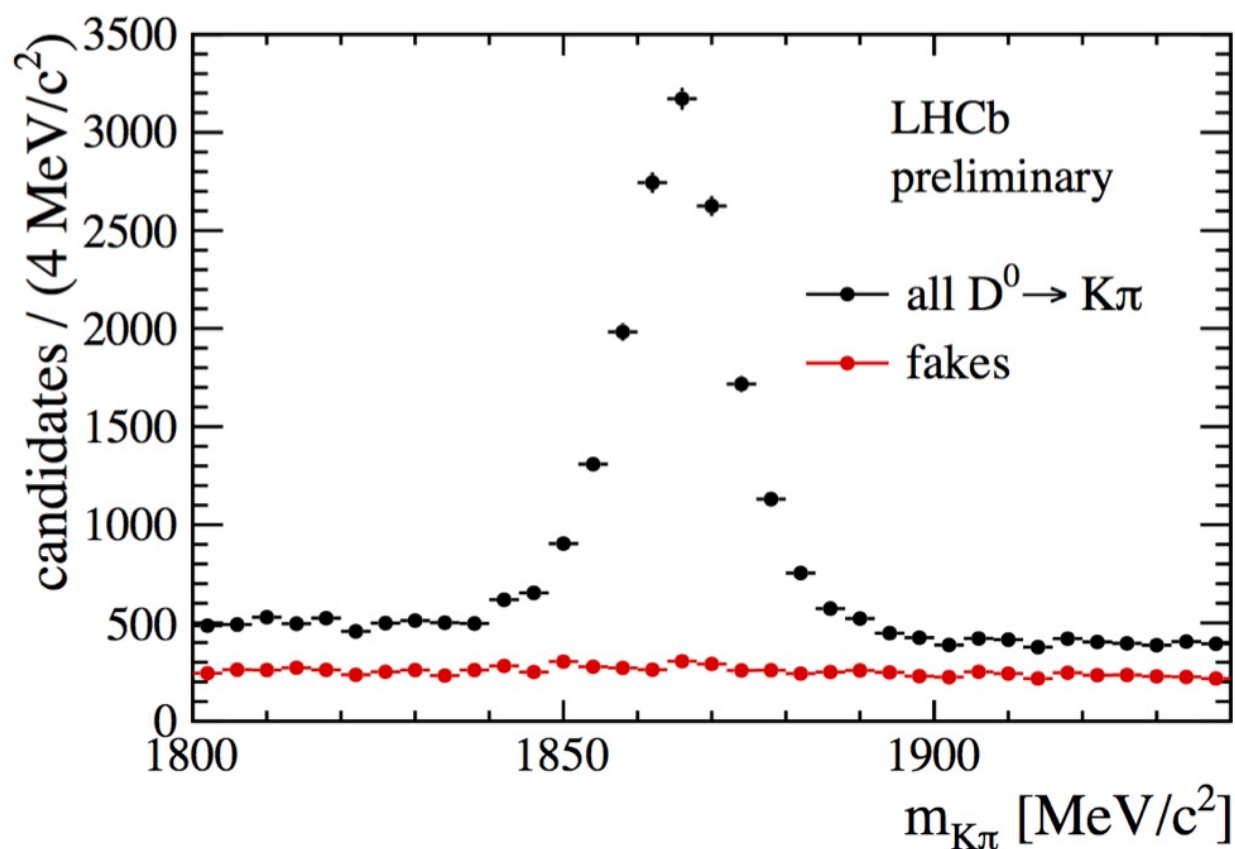
ML@径迹重建

- 径迹重建过程中ML可用在若干步骤中，包括：
Long track和Downstream track的重建
- 在排除fake track的过程中也有应用
 - 降低CPU cost，提高PID和事件重建效率



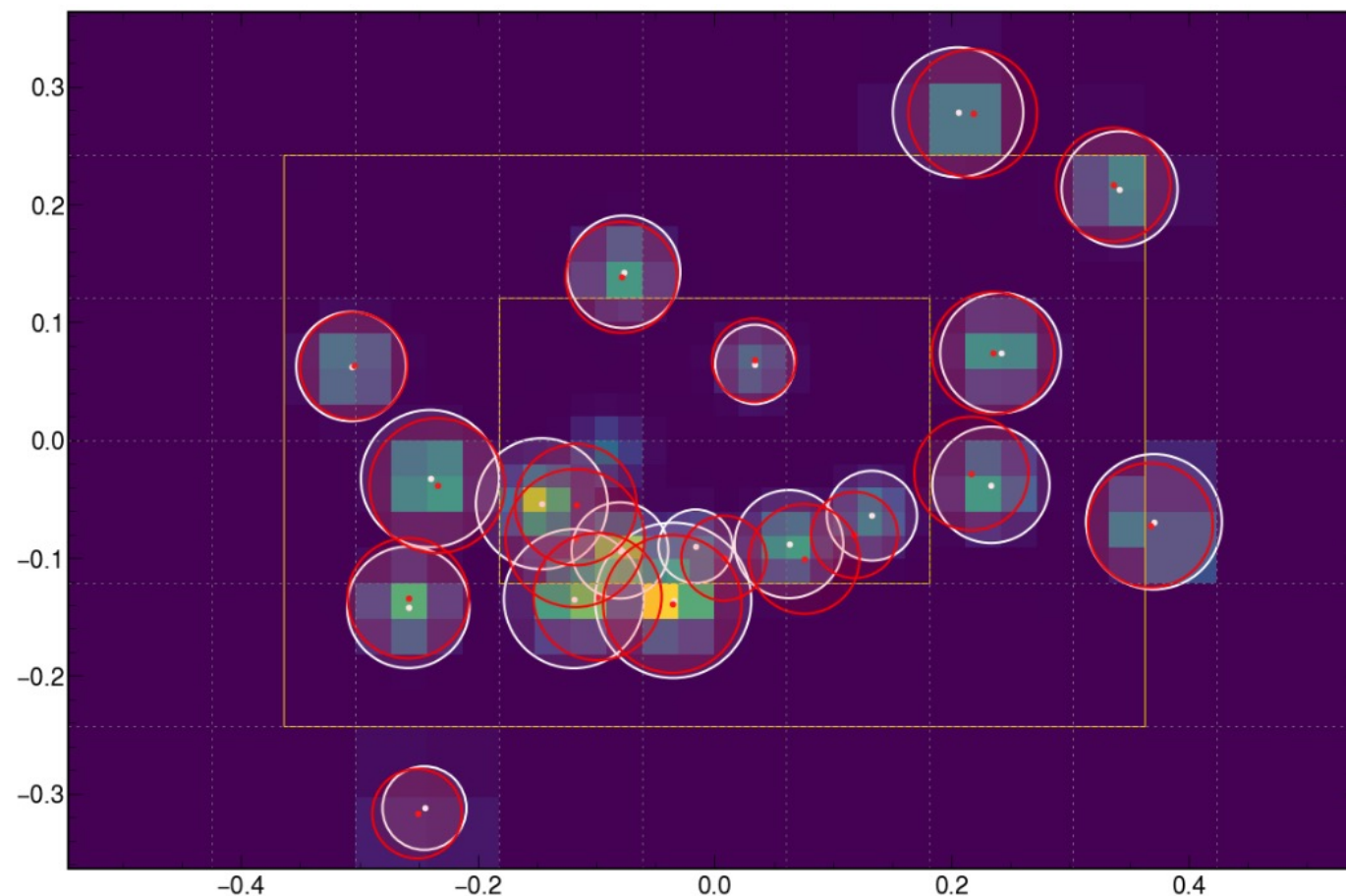
ML@径迹重建

- 通过神经网络评估Ghost probability;
- 训练变量：
21个，包括Kalman filter + hits in tracking detectors + track p_T
- 效果：排除60%的fake track，efficiency $\sim 99\%$

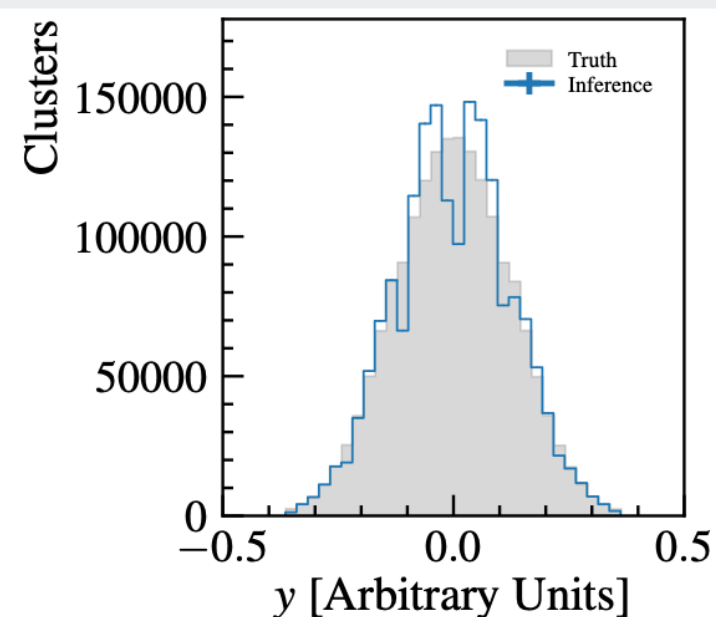
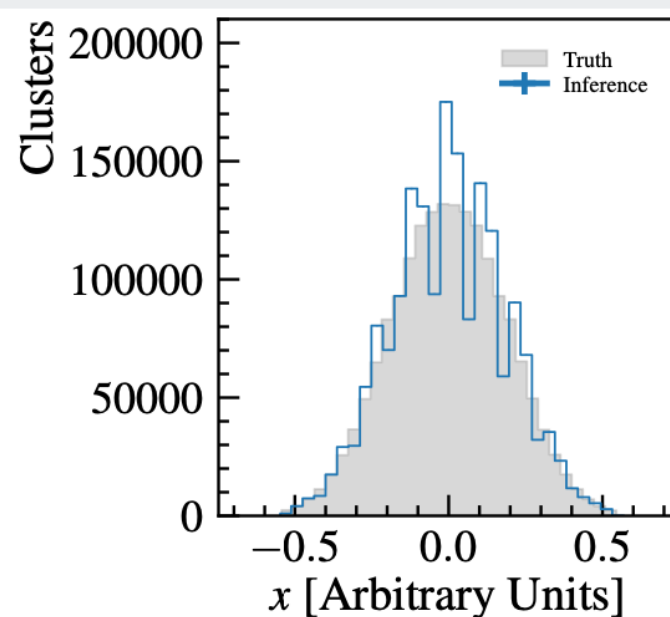
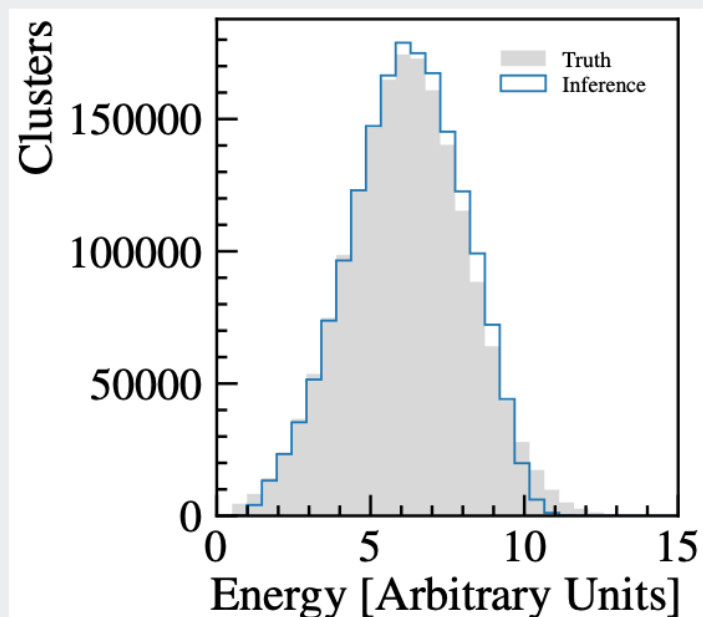


GNN在量能器上应用

- 量能器cluster重建：
2D fast-image-recognition 问题
- 对探测器几何结构使用
Graph Neural Networks



Cumulative cluster observable histograms



LHCb粒子鉴别

- 粒子鉴别 (PID) 对重味物理研究至关重要，可提高信号纯净度

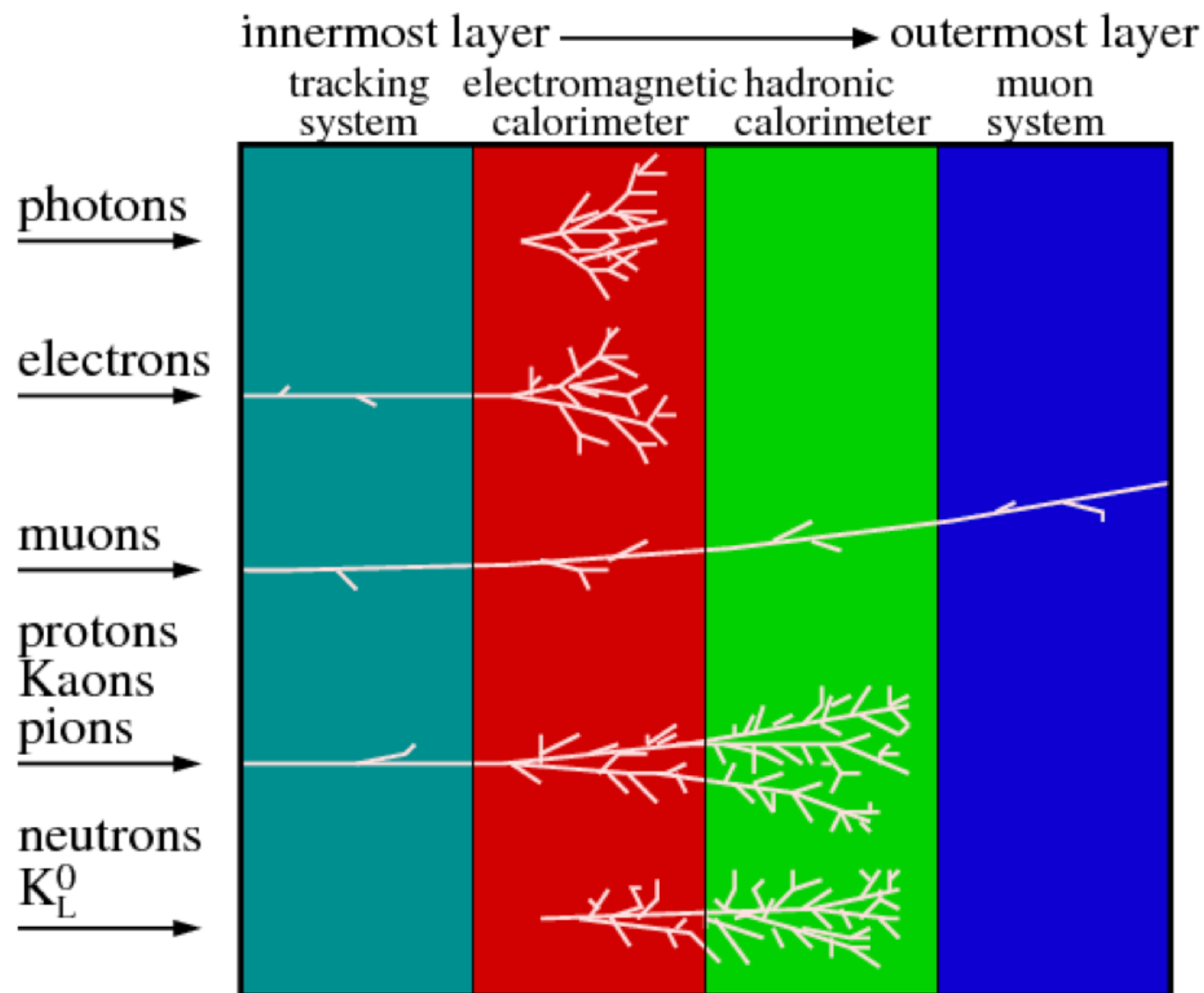
- 需鉴别粒子：

- Charged: e, μ, π, K, p

- Neutrals: π^0, γ, n

- LHCb：
除右侧PID之外，
还有Hadron PID

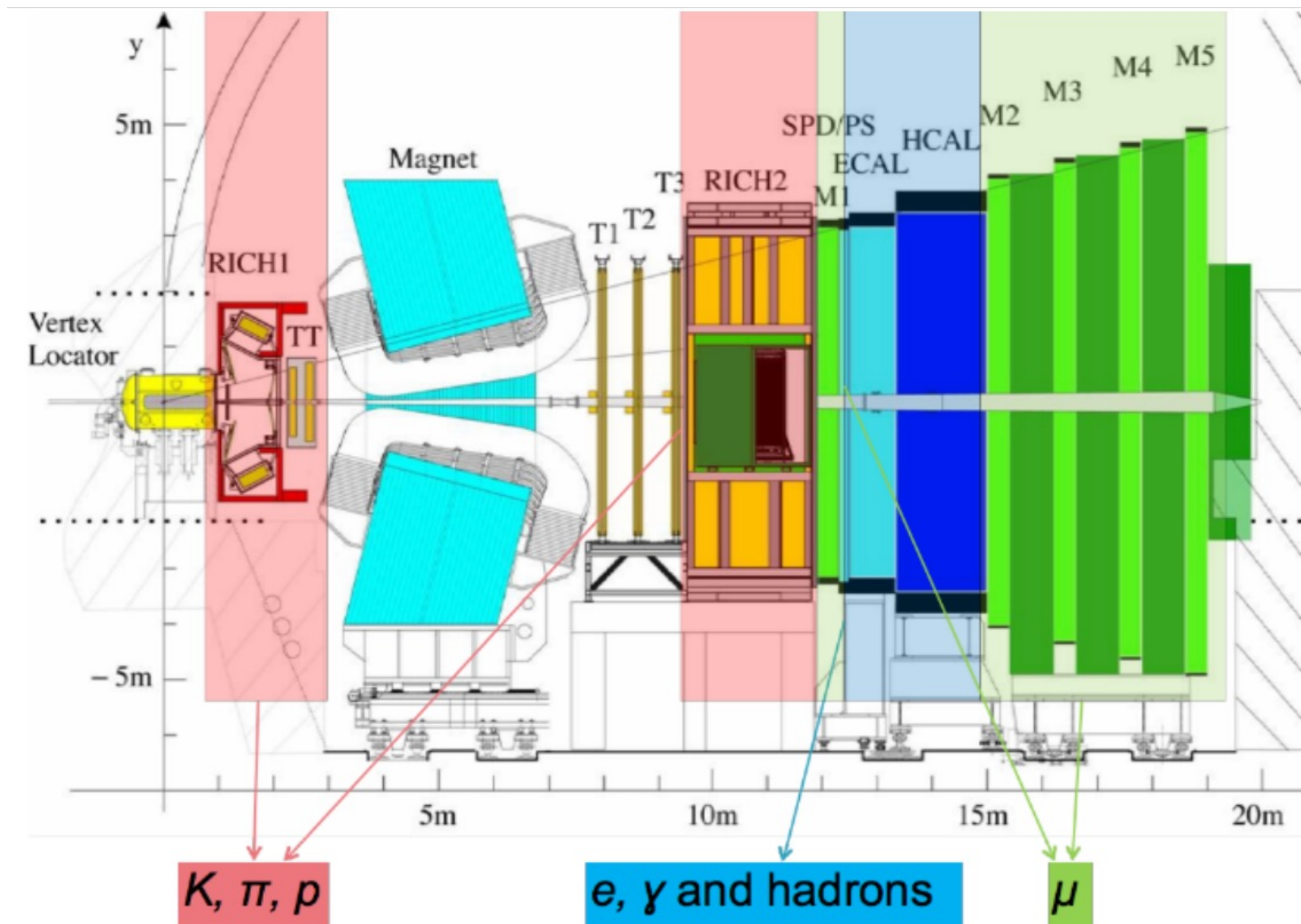
- PID在trigger
(降低带宽)和
offline阶段均有
应用



C. Lippmann – 2003

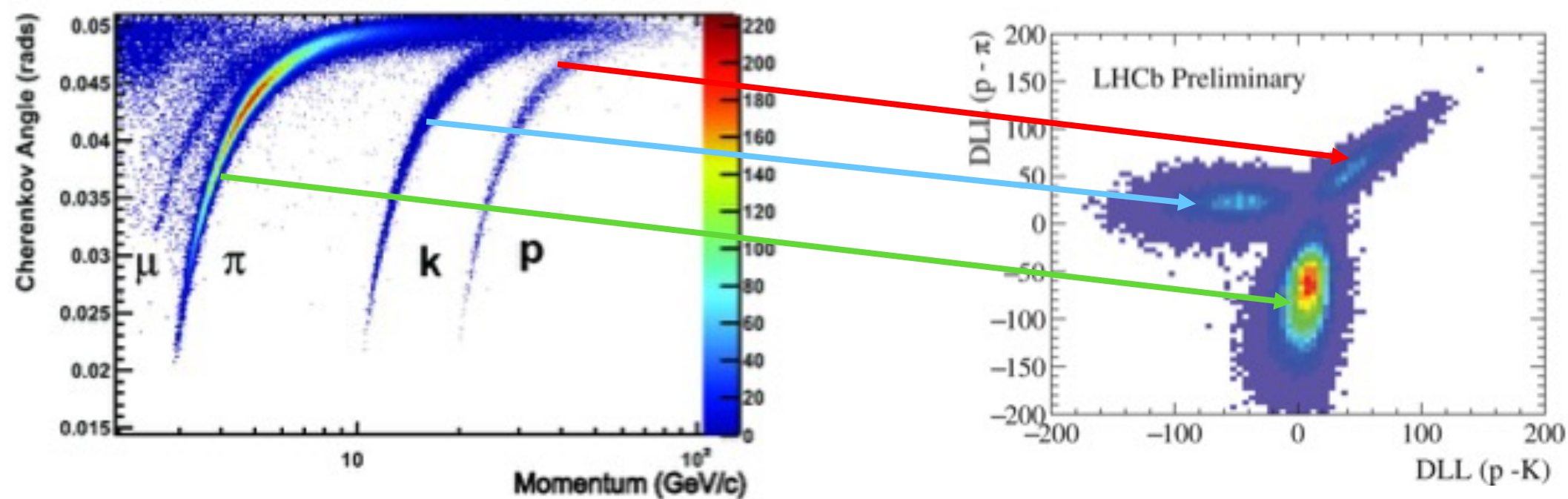
LHCb粒子鉴别探测器系统

- 切伦科夫探测器
Hadron PID
- 量能器：
 γ , e at L0 ;
PID for γ , e, h ;
测量能量和位置
- Muon探测器：
muon at L0 ;
PID for muon



ML@粒子鉴别

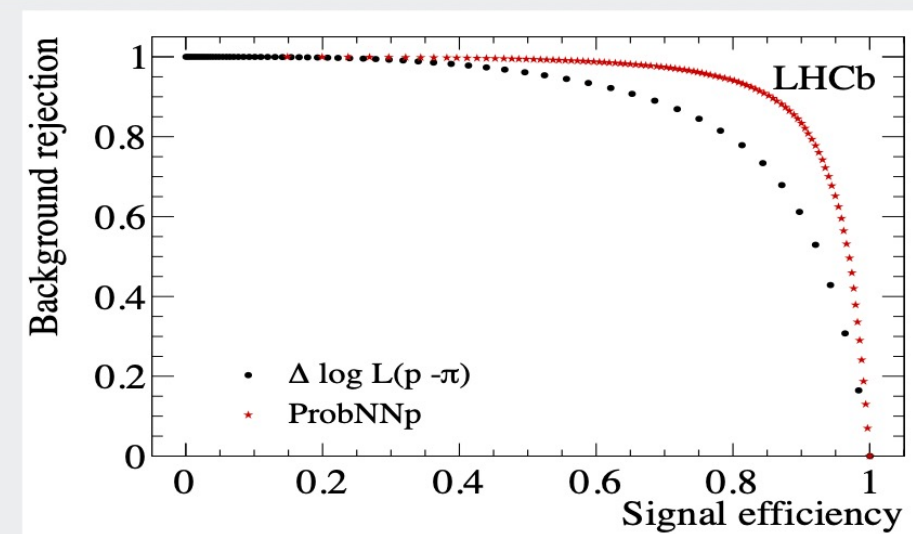
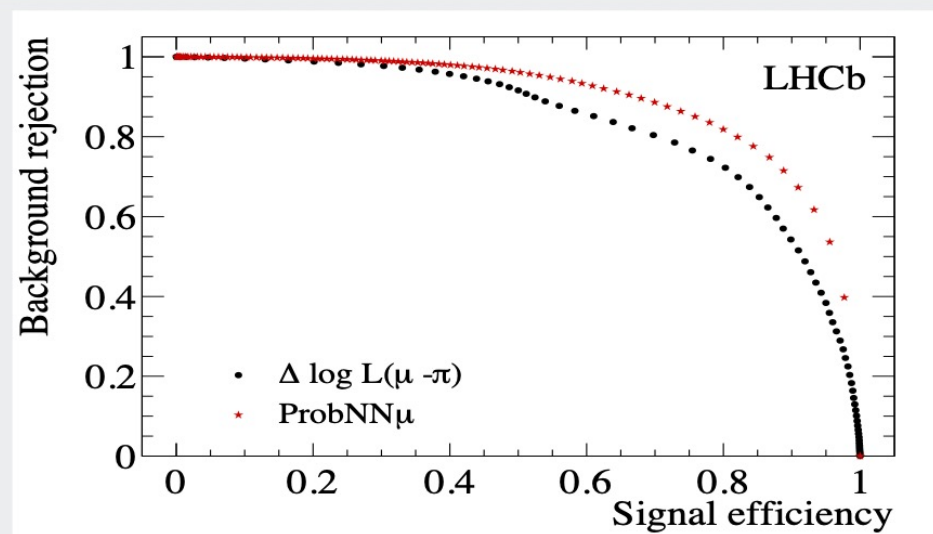
- 不同子探测器系统提供不同的重建信息
 - RICH : L_X/L_π , $X = \{e, \mu, K, p, d\}$
 - Calo : L_e/L_π
 - Muon system : L_μ/L_π
- 以上组合形成 Differential Log Likelihood (DLL) 信息



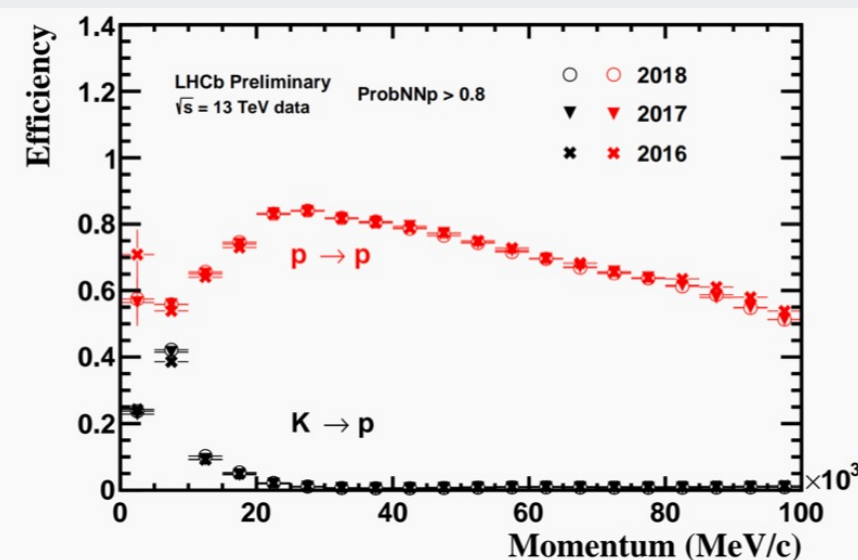
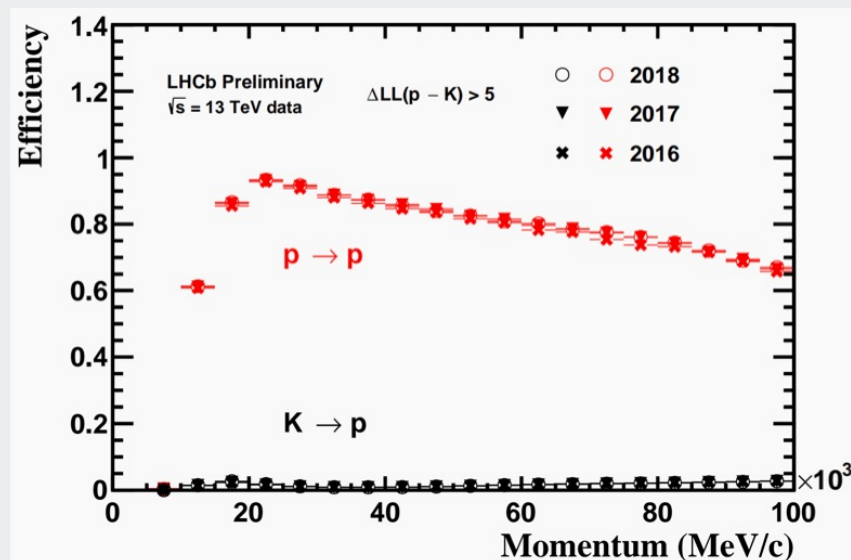
ML@粒子鉴别

- ML可提供额外的PID分辨能力
- 训练变量来源：tracking + RICH + ECAL + HCAL + Muon system

Background misidentification versus muon identification efficiency, $\Sigma^+ \rightarrow p\mu^+\mu^-$

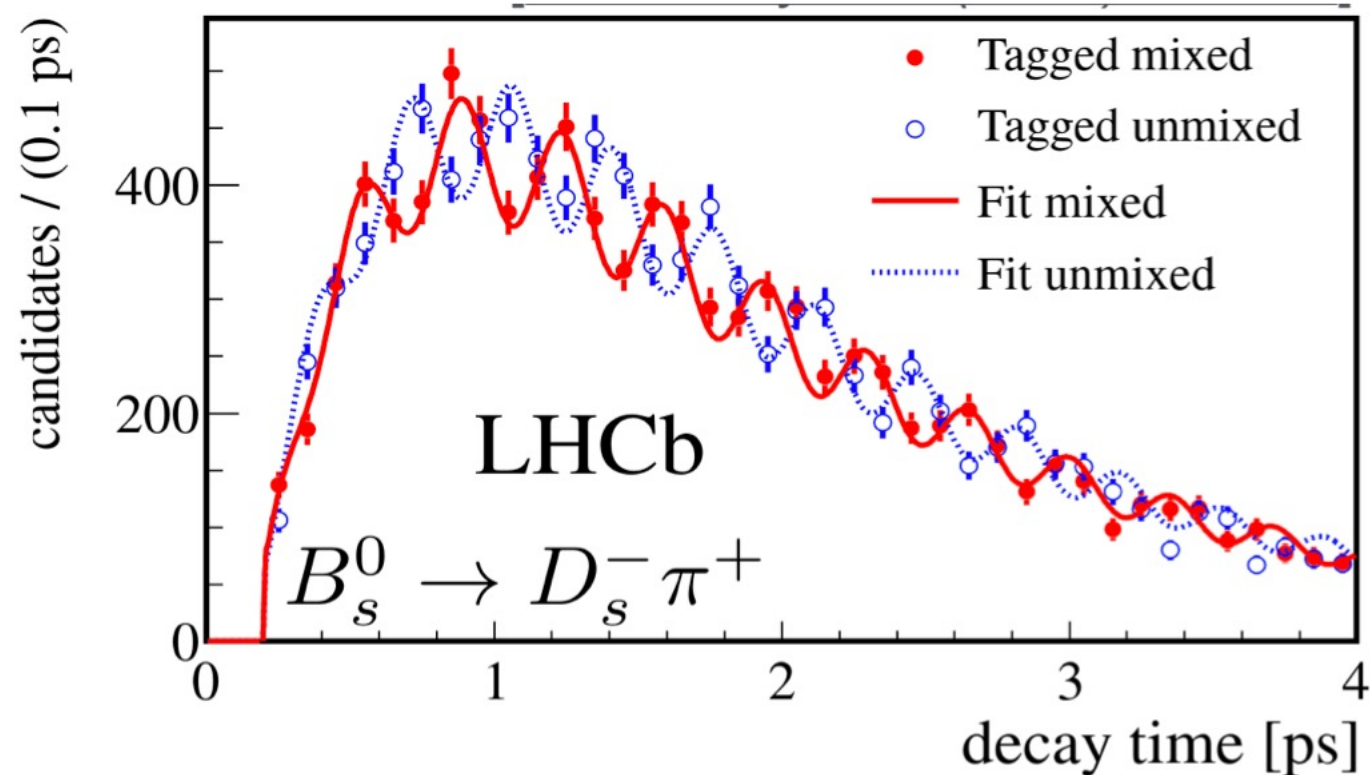


LHCb-FIGURE-2020-012, performance evaluation on calibration data



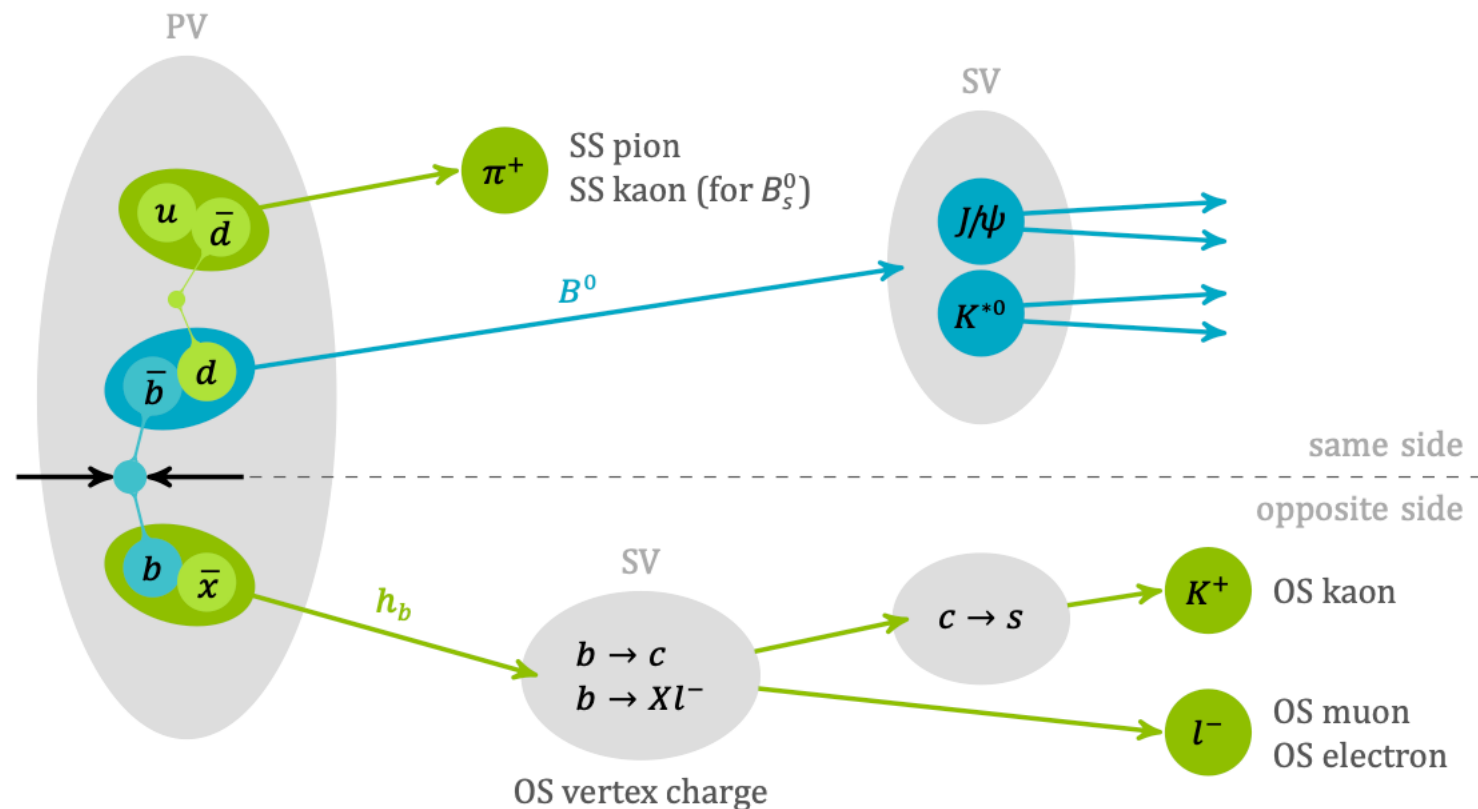
ML@flavour tagging

- 测量对时间依赖的B mixing和CPV需要知道B介子的初始 flavour
- Flavour tagging (FT) 在LHCb重要且困难
 - 不像B工厂, LHCb无法产生quantum-correlated $B\bar{B}$ pairs
 - pp对撞中的Hadronic pollution -> 降低重建效率



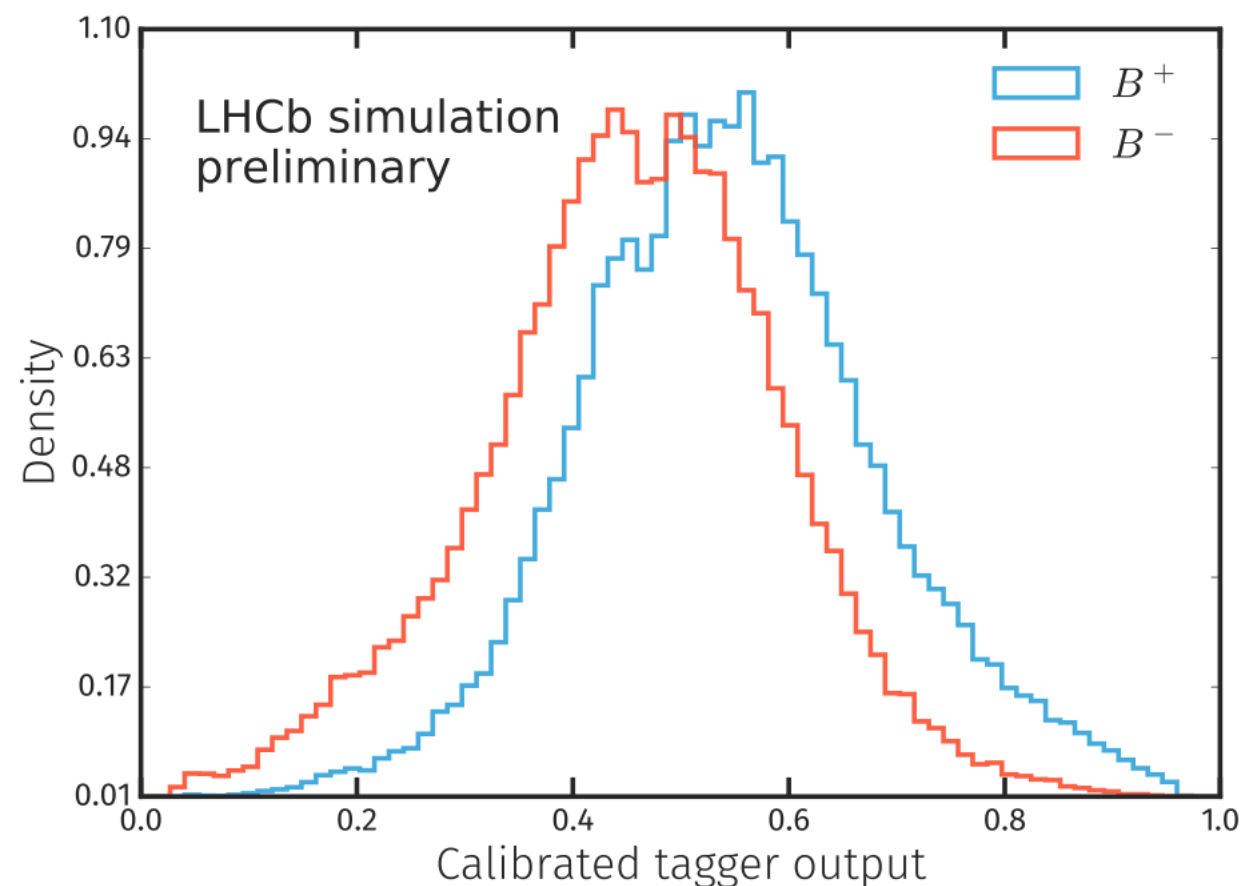
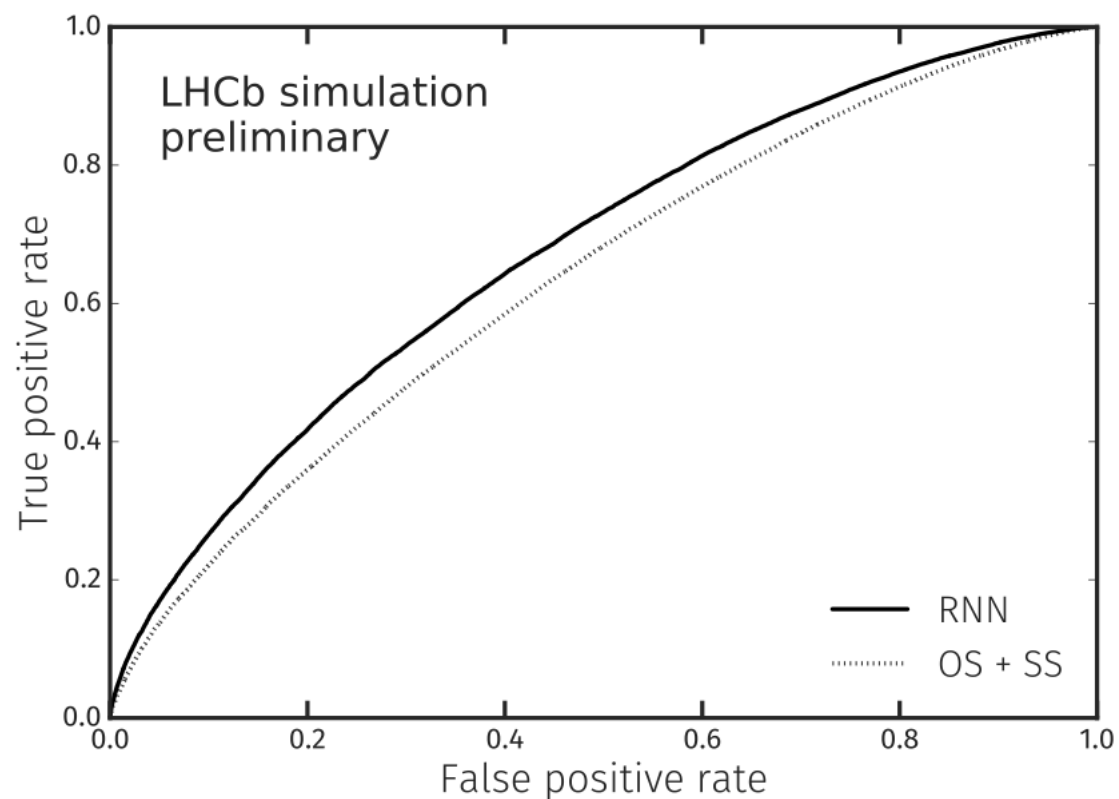
ML@flavour tagging

- Flavour tagger使用衰变事件的几何信息和运动学信息进行MVA训练
- Run-I & II , 两类tagger :
 - Opposite side (OS) taggers: identify (quasi-)flavour-specific decay other B in the event;
 - Same side (SS) taggers: identify particles in the signal b fragmentation.



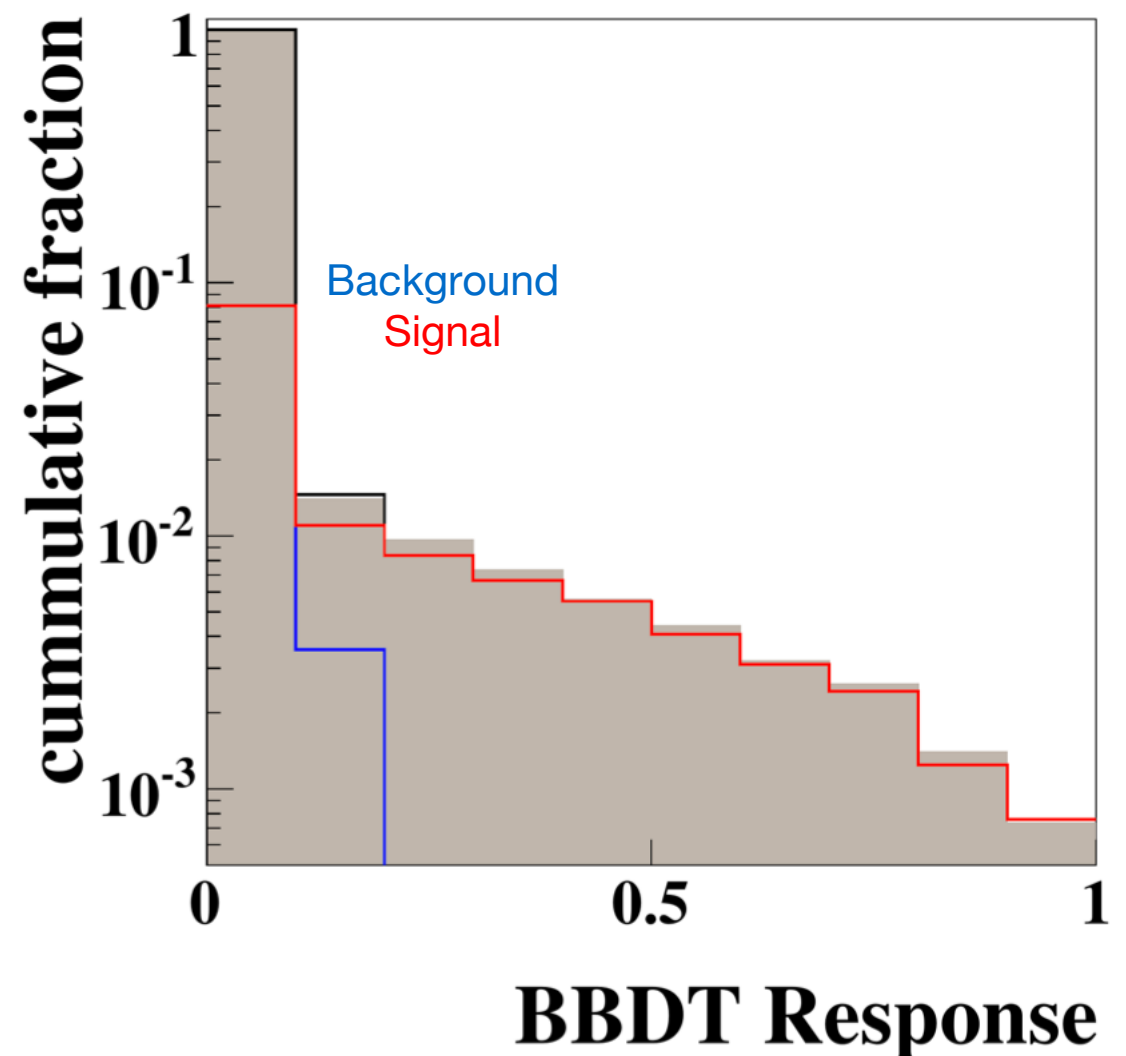
ML@flavour tagging

- Run-III : 平均每bunch crossing的对撞数由 ~ 1.1 提升至 ~ 5
- 对Run-I&II的tagger进行tuning十分困难
- Run-III : 将使用来自一个Primary Vertex的全部信息构建新tagger
- 利用运动学信息、几何信息以及PID信息对RNN进行训练



HLT topo B trigger

- 在LHCb HLT阶段，使用Bonsai BDT可大量reject背景事件
- 训练变量： p_T ，IP χ^2 wrt PV，track χ^2/ndf ，FD χ^2 等拓扑结构变量
- BBDT response > 0.2 区间高信号纯净度
- 目前，LHCb 40%的HLT2 带宽是 inclusive n-body topologic trigger



LHCb上的其它ML应用

- Run-II运行：使用ML/DL辅助实现对探测器的实时alignment
- Run-III运行：在更高亮度和数据量条件下达到同样性能
- Data quality monitoring
- 大量使用TMVA进行数据筛选的物理分析
-

总结

- ML在优化LHCb数据处理过程方面有重要体现
- 纯软件触发系统的巨大数据量会是Run-III阶段的巨大挑战，ML将协助解决这一问题



backup

LHCb顶点探测器

- 探测模组包围对撞点，提供精准对撞/衰变顶点位置信息
- 是LHC上距离对撞点最近的探测系统
- 有效探测单元距对撞点仅8mm（2010-2018）或5mm（2022-）
- 在束流注入阶段必须左右打开待束流稳定后再关闭

