

# Triggers and Machine Learning

Mika Vesterinen

27<sup>th</sup> August 2014

1. Introduction
  1. LHC/trigger generalities
  2. Low level triggers
  3. High level triggers
2. MLHEP and triggers (LHCb HLT)

Unterstützt von / Supported by

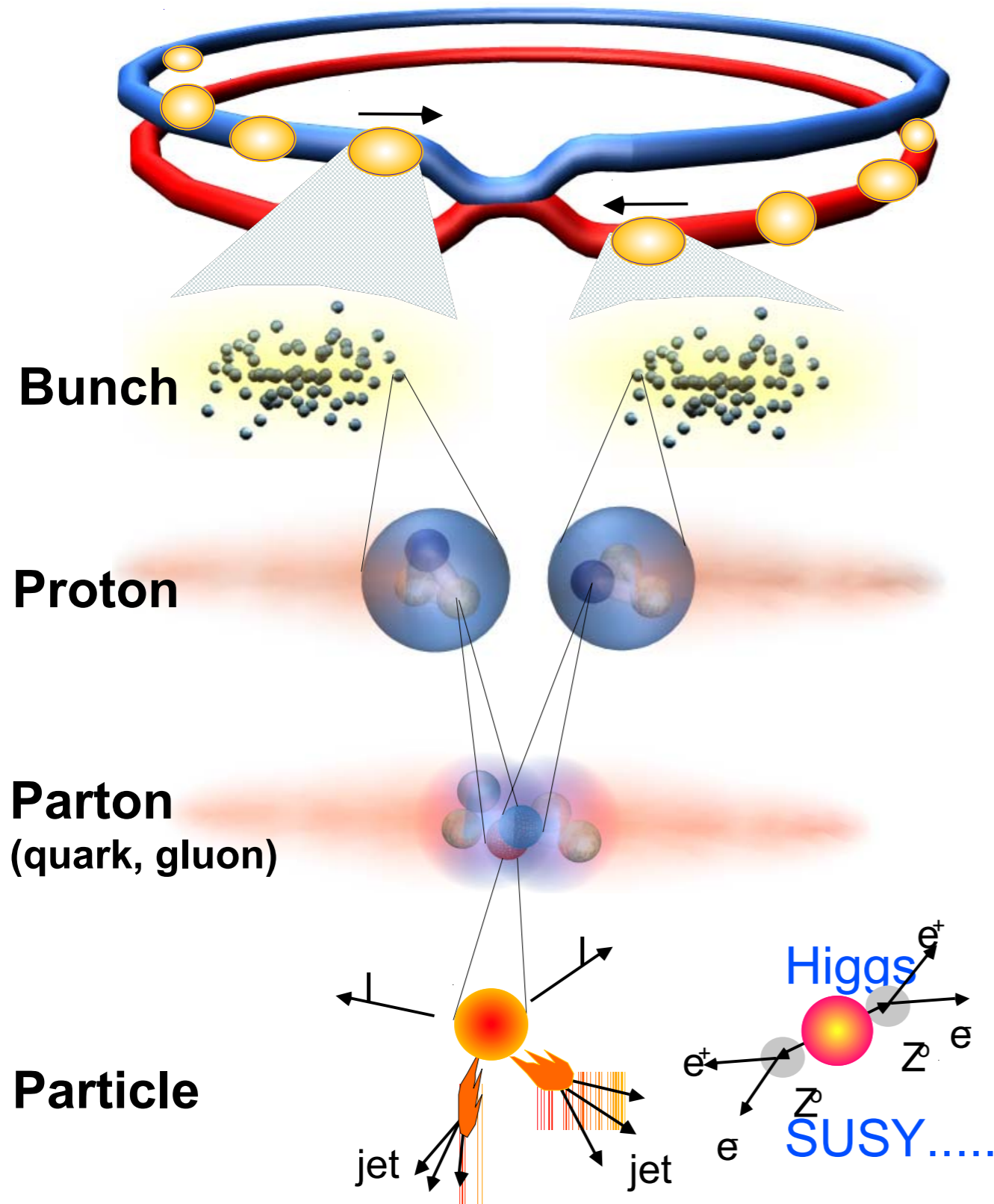


**Alexander von Humboldt**  
Stiftung / Foundation



**UNIVERSITÄT  
HEIDELBERG**  
ZUKUNFT  
SEIT 1386

# The LHC



<b>Proton - Proton</b>	<b>2804 bunch/beam</b>
<b>Protons/bunch</b>	<b><math>10^{11}</math></b>
<b>Beam energy</b>	<b>7 TeV (<math>7 \times 10^{12}</math> eV)</b>
<b>Luminosity</b>	<b><math>10^{34} \text{cm}^{-2} \text{s}^{-1}</math></b>

<b>Crossing rate</b>	<b>40 MHz</b>
----------------------	---------------

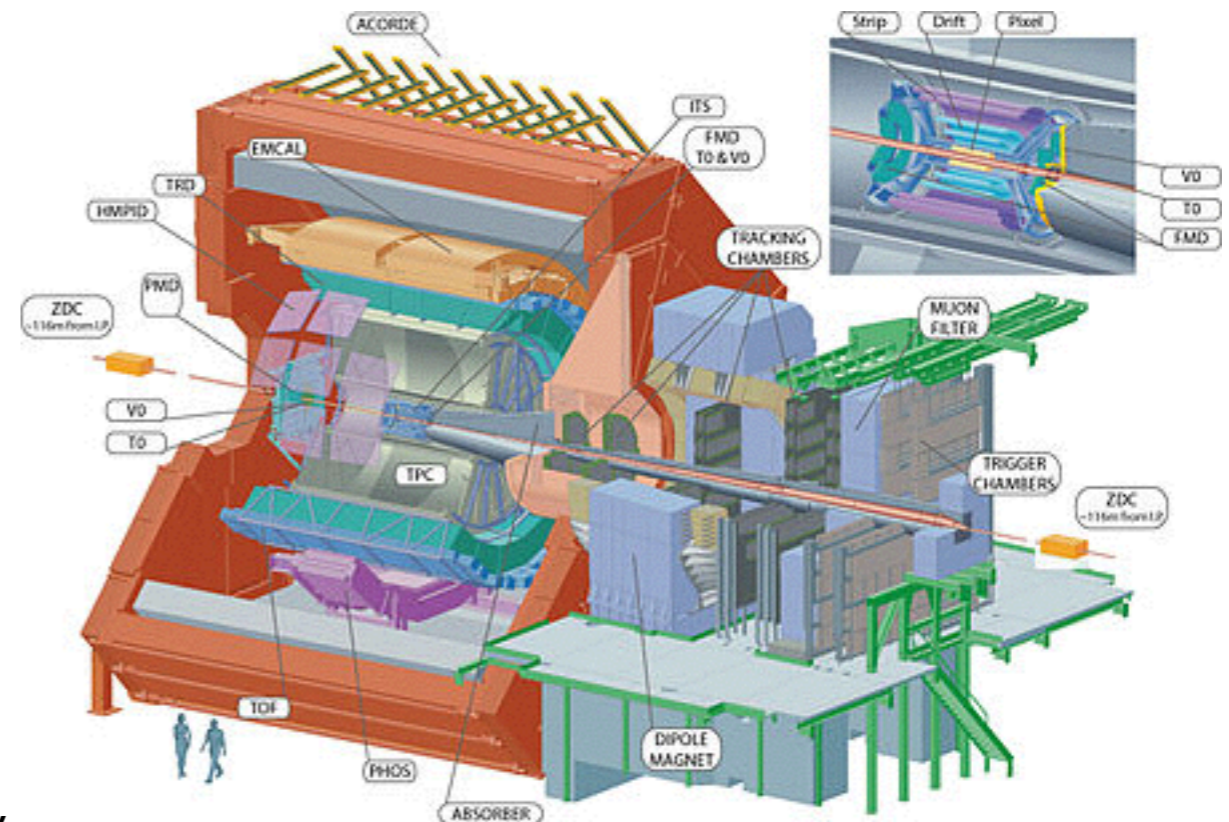
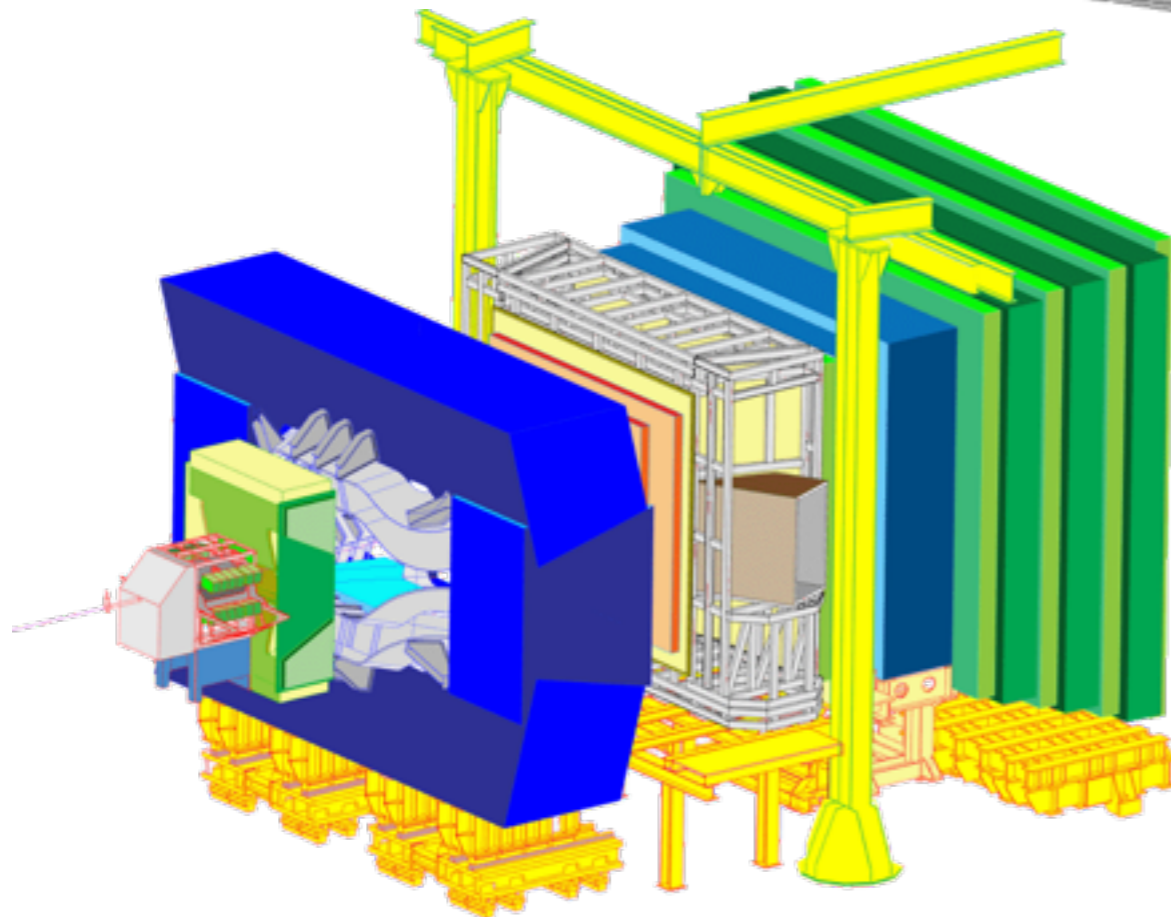
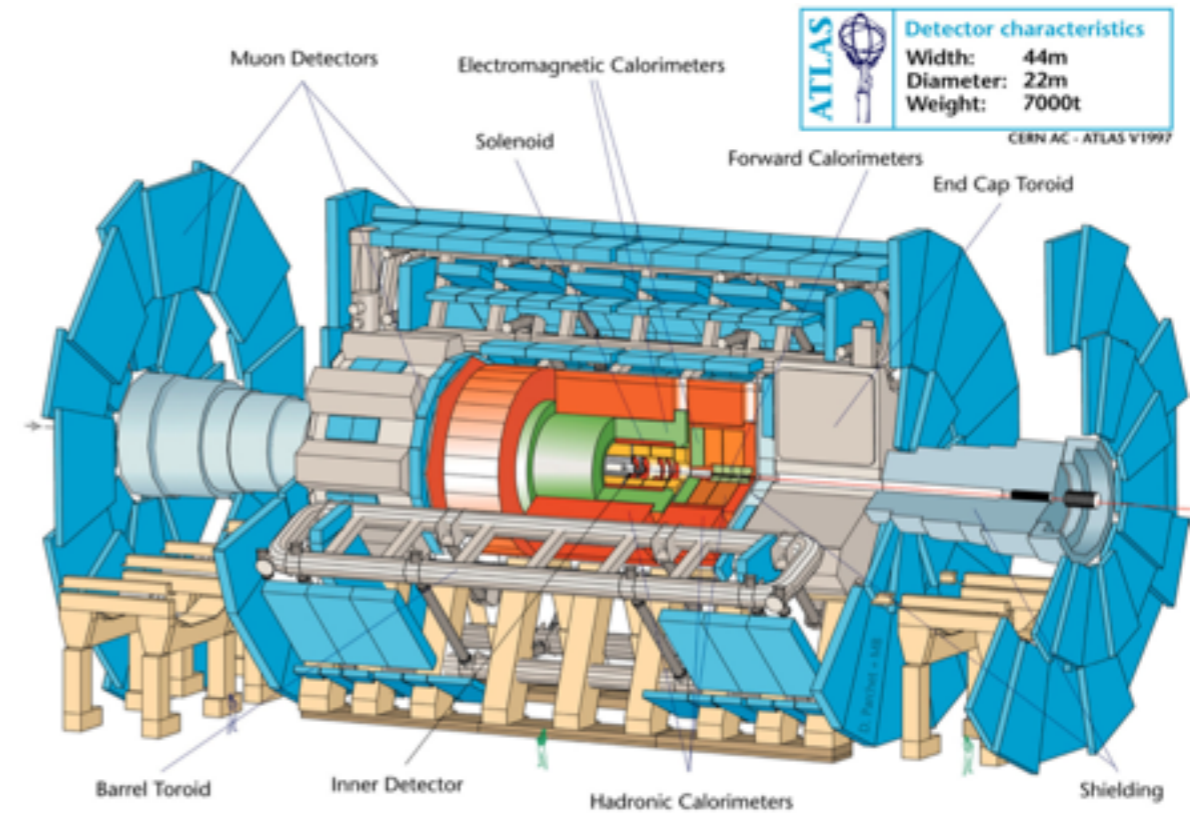
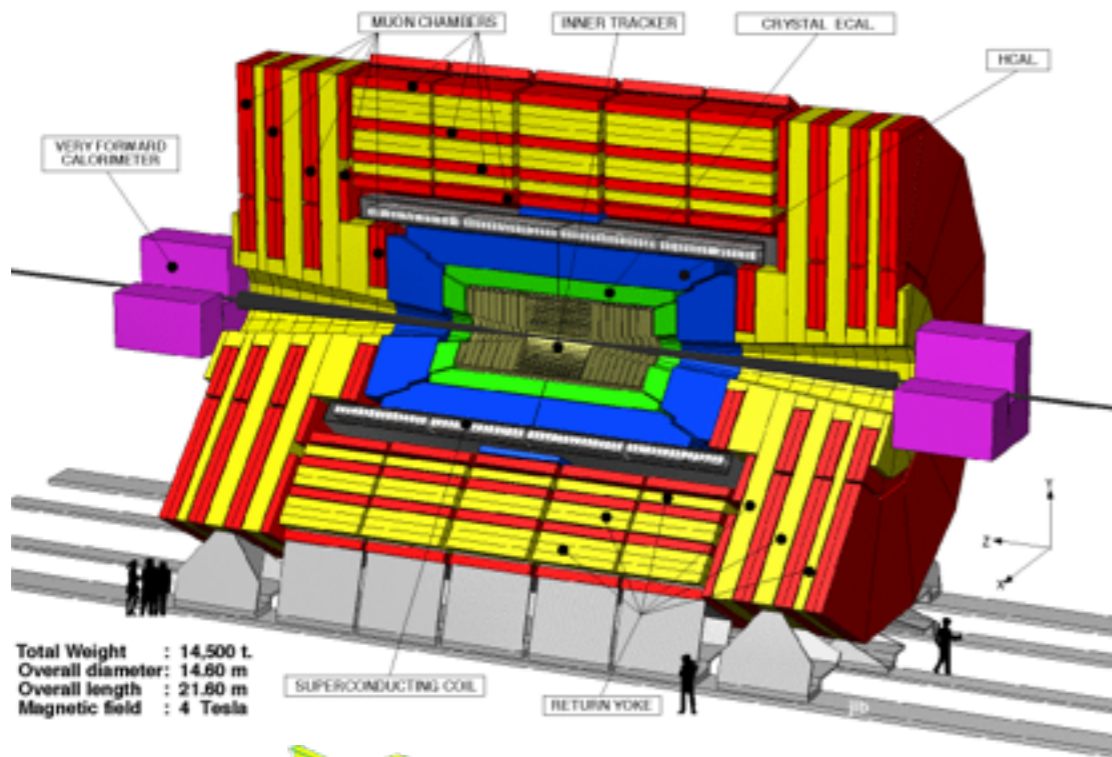
<b>Collision rate <math>\approx</math></b>	<b><math>10^7 - 10^9</math></b>
--	---------------------------------

**New physics rate  $\approx$  .00001 Hz**

<b>Event selection:</b>	<b><math>10^{13}</math></b>
<b>1 in</b>	<b>10,000,000,000,000</b>



# The detectors



# Nomenclature

**Event: A single bunch crossing (40 MHz)**  
**(Versus single proton-proton interaction)**



# Nomenclature

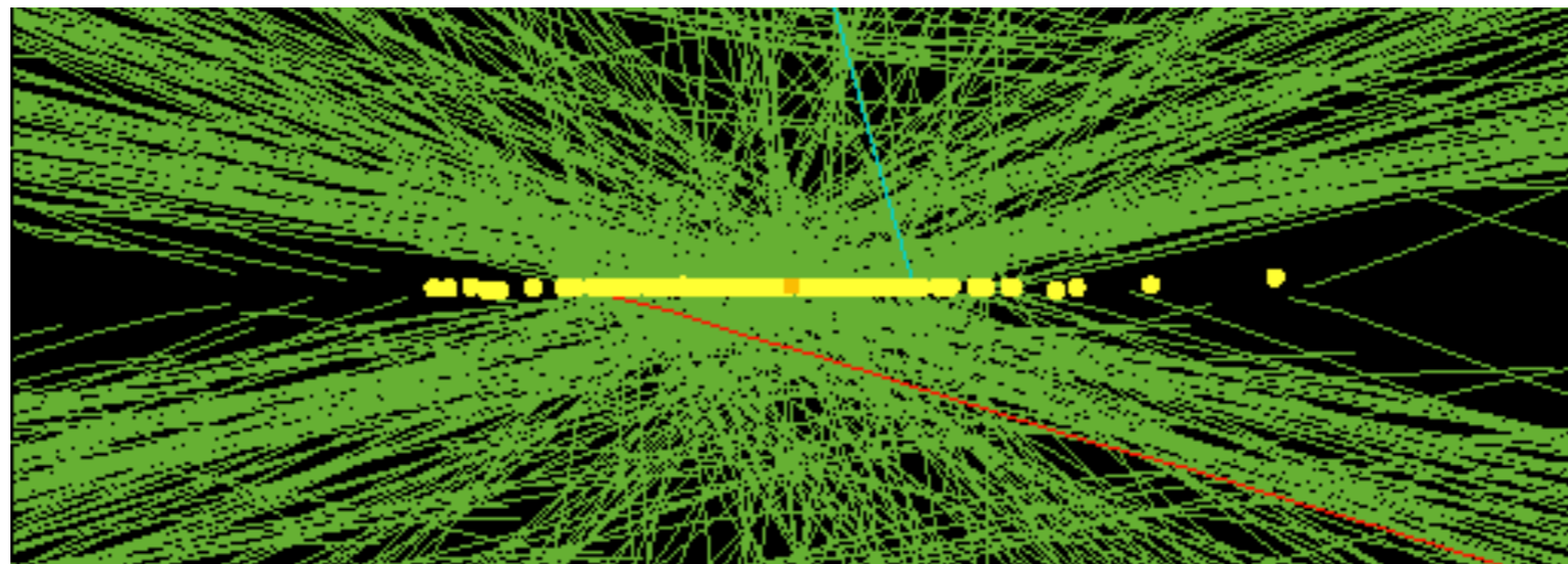
**Event: A single bunch crossing (40 MHz)**  
**(Versus single proton-proton interaction)**

$$\sigma \sim 70 \text{ mb}$$

$$L_{\text{ATLAS/CMS}} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$$

$$\text{Interaction rate} = L \times \sigma \sim 700 \text{ MHz}$$

**~20 “pileup” interactions per event@ 40 MHz**



**Q: How much would it cost to have a trigger less detector?**





# ATLAS...



**$10^8$  channels: 1 MB /event**

**30 MHz bunch crossing rate**

**LHC year:  $5 \times 10^6$  seconds**

# ATLAS...



**$10^8$  channels: 1 MB /event**

**30 MHz bunch crossing rate**

**LHC year:  $5 \times 10^6$  seconds**

**$0(10^5)$  pB per year**

**Storage cost: €100B**

**Total budget of LHC €10B**

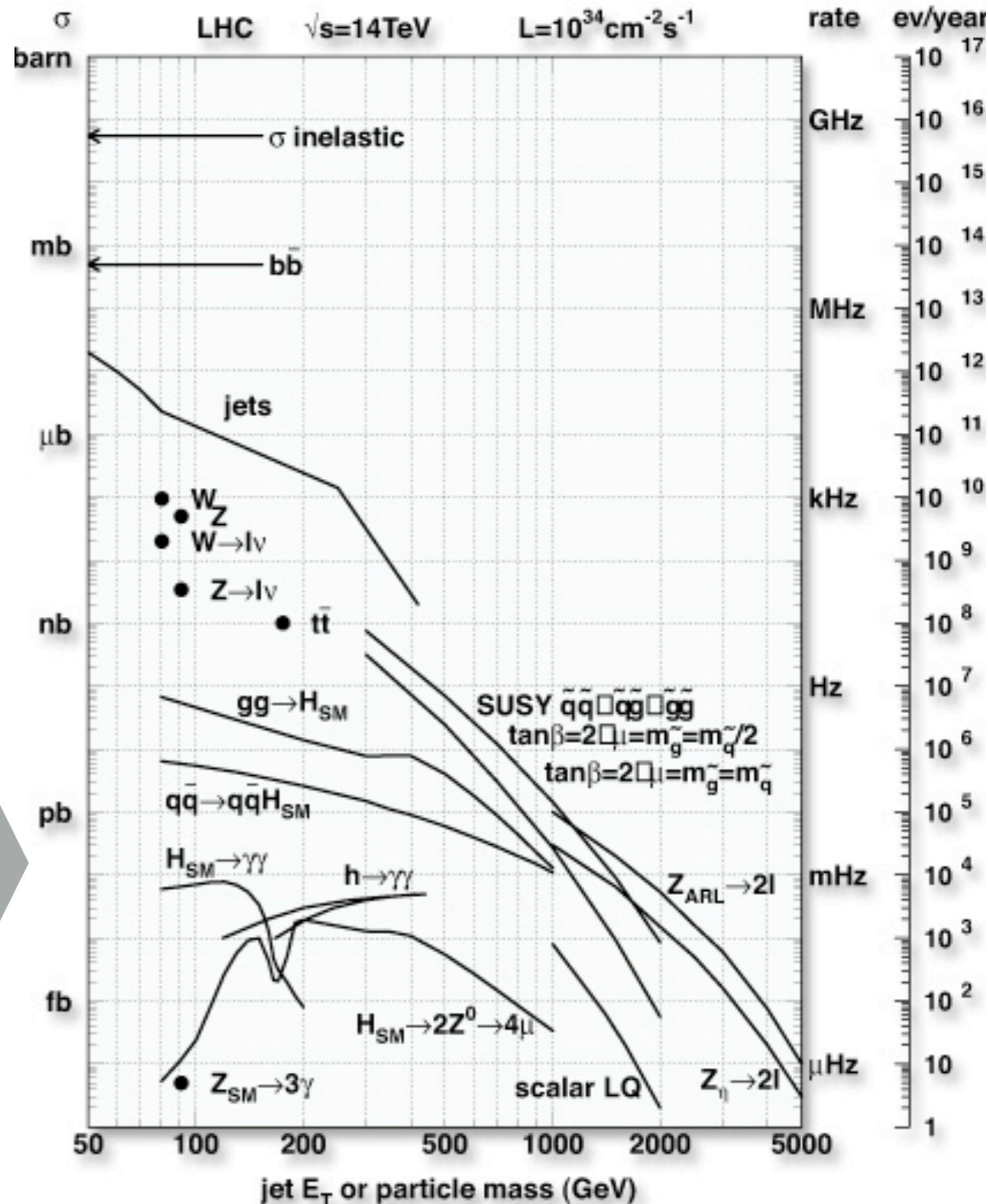
**Trigger: reduce rate by 4-5 orders  
of magnitude in real time.**



**Roughly the total global internet traffic!**

# Selectivity

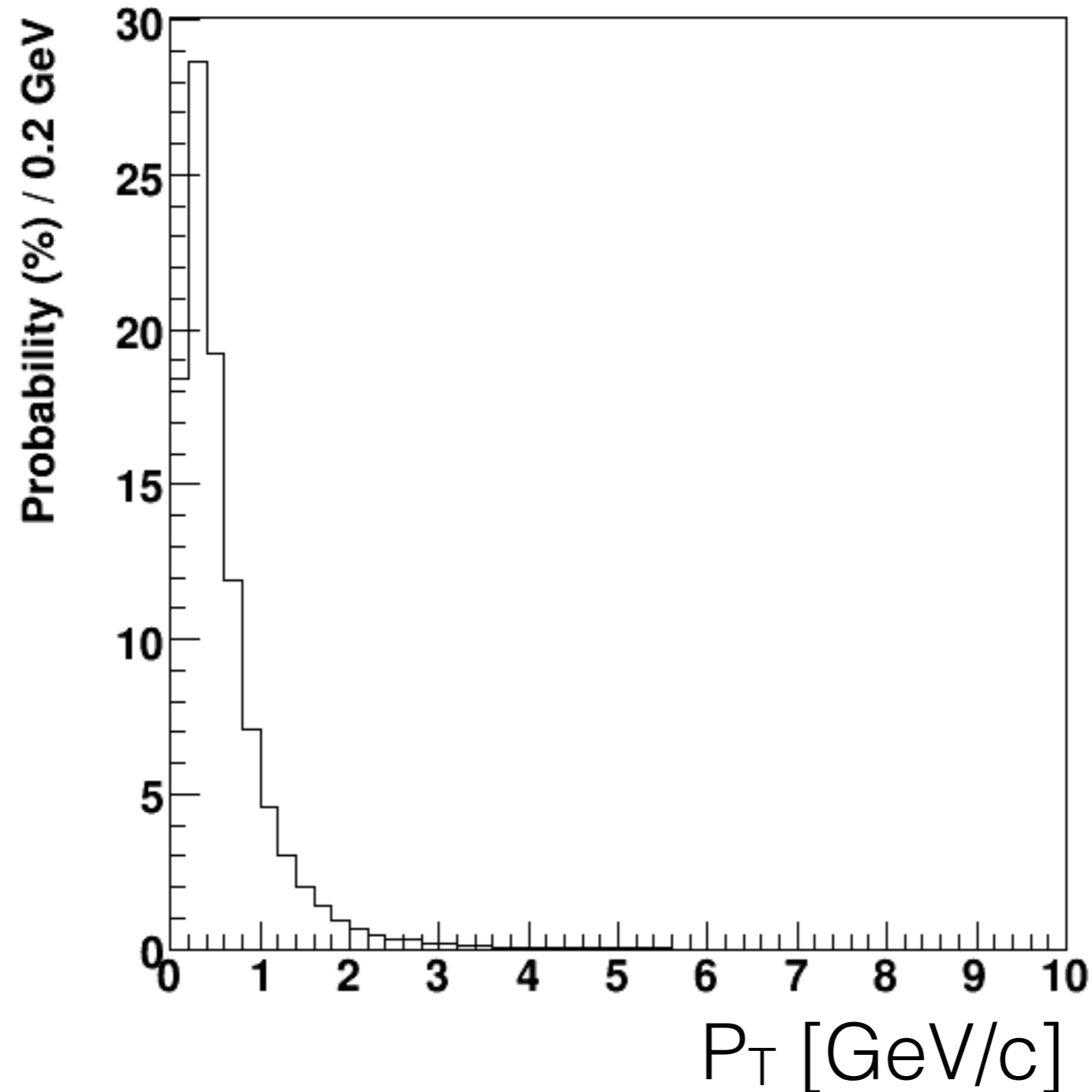
The really interesting stuff is  $\sim 13$  orders of magnitude smaller.





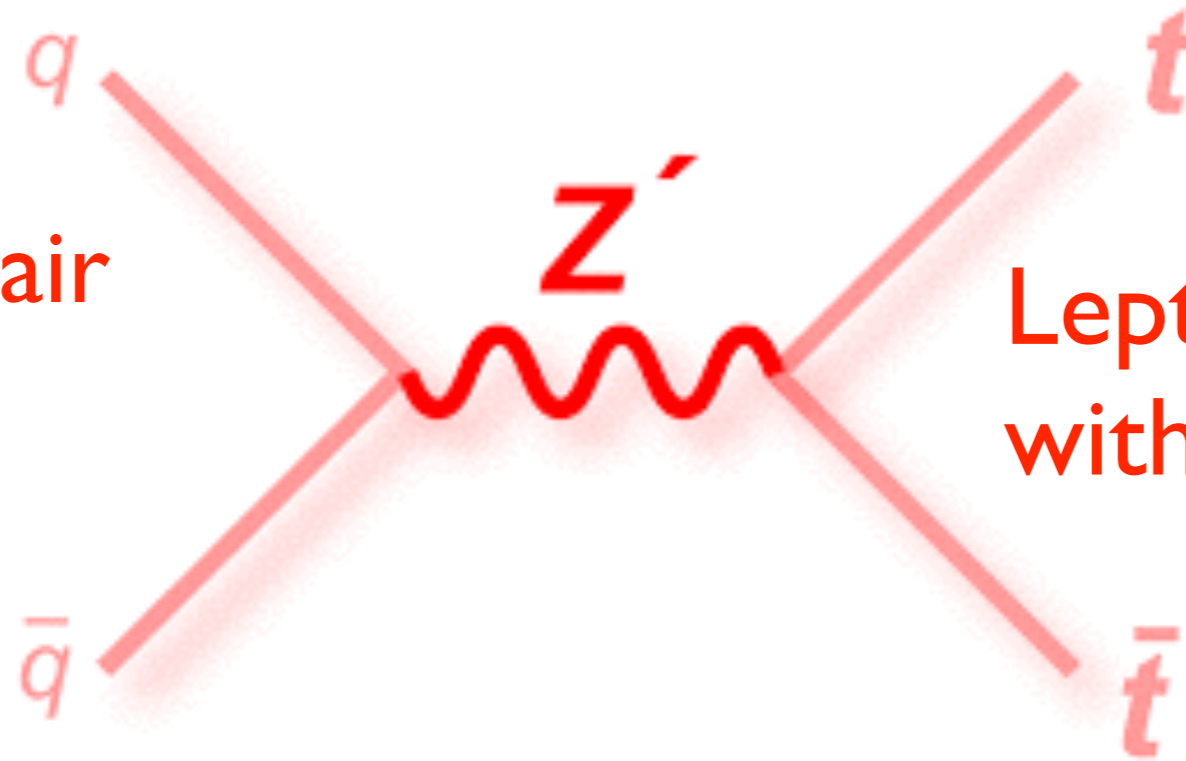
# Background is low $p_T$

$P_T$  tracks in minimum bias events



# Signal is high $p_T$

E.g., a 3 TeV  $Z'$   
decay to top pair



Leptons, jets  
with  $p_T > 100$  GeV,

Can trigger **inclusively** on single high  $p_T$  leptons, jets, missing  $E_T$ , or simple combinations of these

Only part of the decay is needed to trigger the event, so large physics programme with small set of triggers.

# How to discover new physics at the LHC

## Direct approach

$$E=MC^2$$



Works if  $M \ll 14 \text{ TeV}$

## Indirect approach



With high enough precision  
(theory and experiment), can  
probe any mass scale



# How to discover new physics at the LHC

## Direct approach

$$E=MC^2$$



Works if  $M \ll 14 \text{ TeV}$

**Inclusive triggering  
on high  $p_T$  particles**

## Indirect approach



With high enough precision  
(theory and experiment), can  
probe any mass scale

**Mixture of inclusive  
and exclusives...**

# How to discover new physics at the LHC

## Direct approach

$$E=MC^2$$



Works if  $M \ll 14 \text{ TeV}$

Inclusive triggering  
on high  $p_T$  particles

MLHEP in final analysis

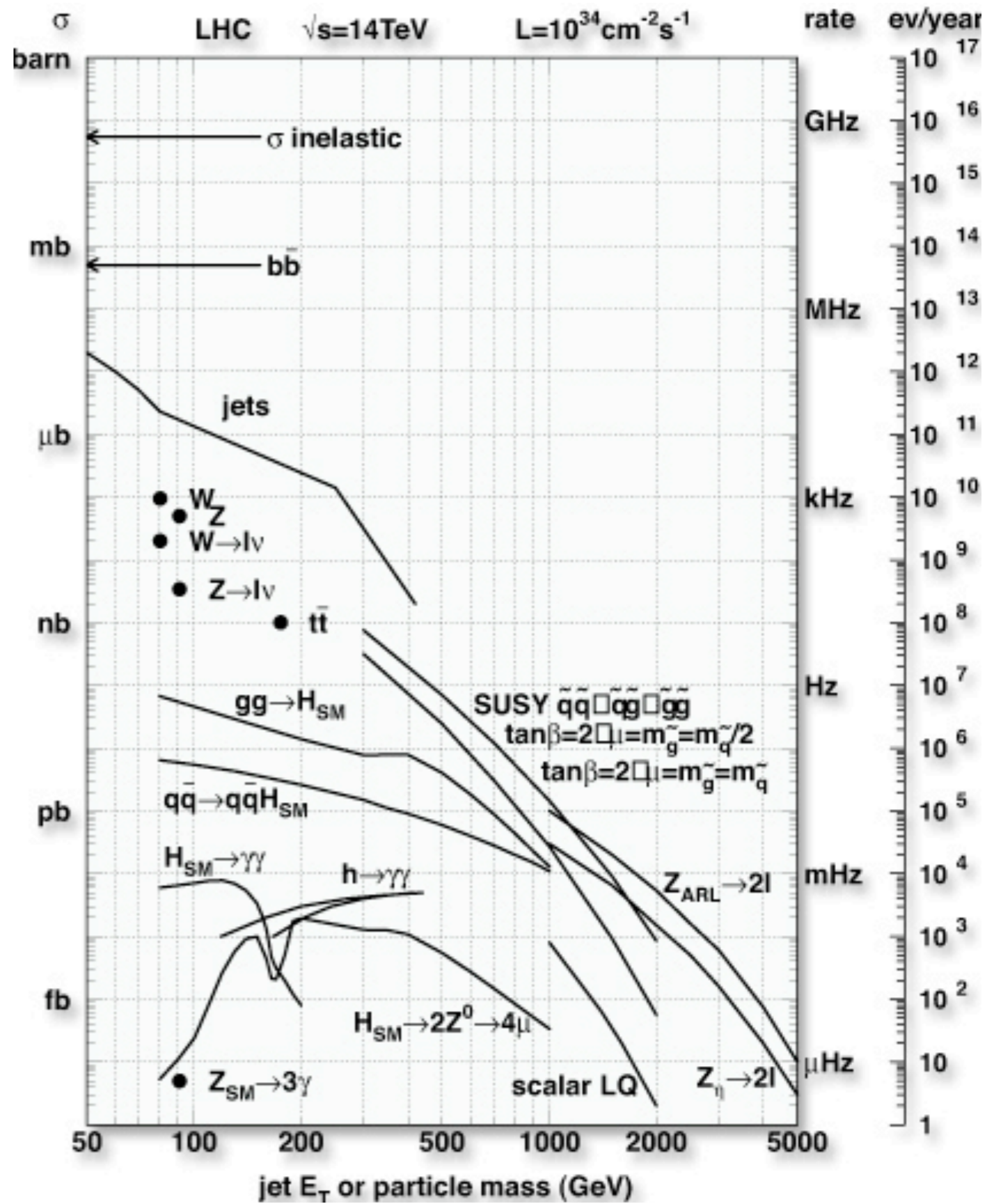
## Indirect approach



With high enough precision  
(theory and experiment), can  
probe any mass scale

Mixture of inclusive  
and exclusives...

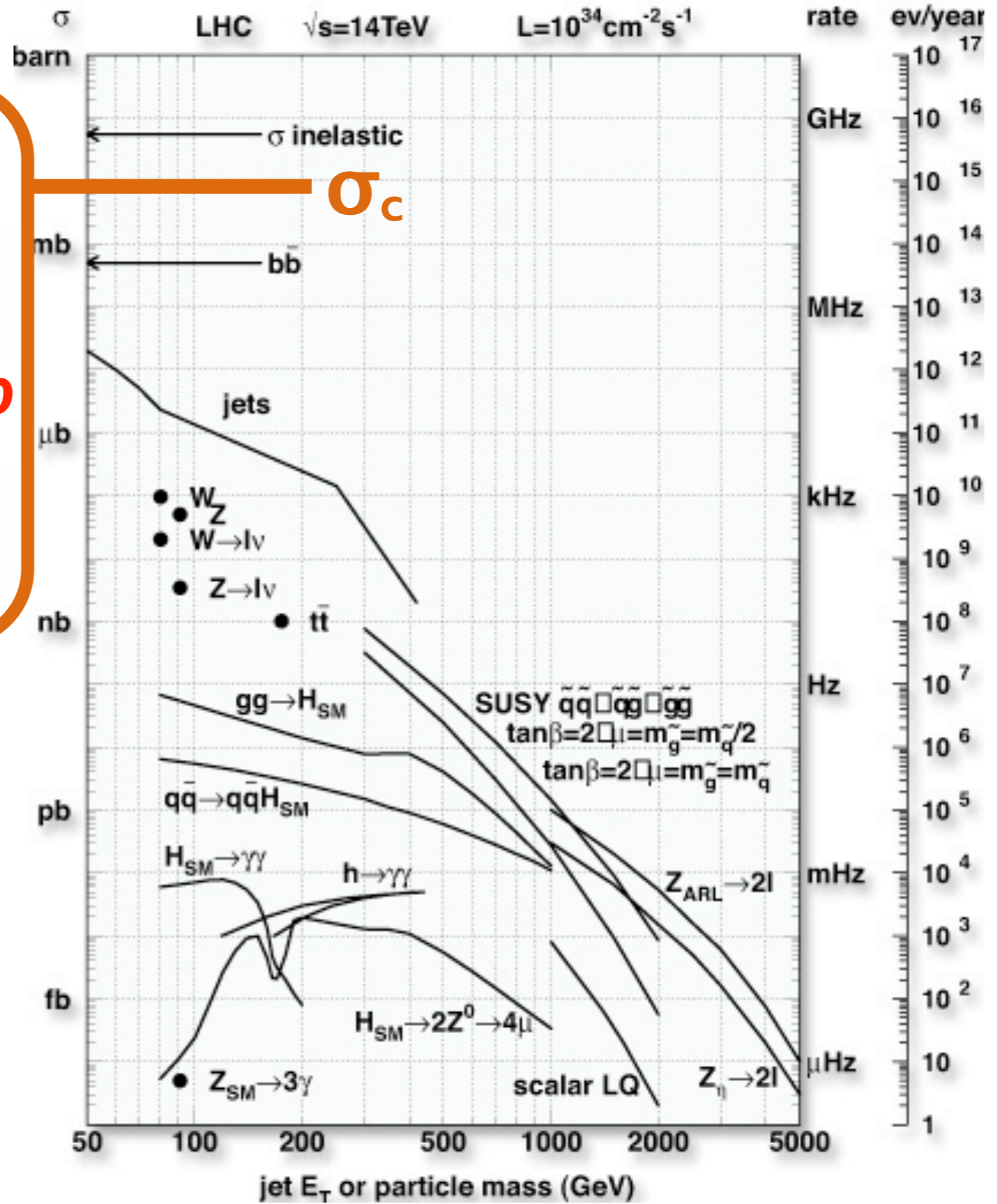
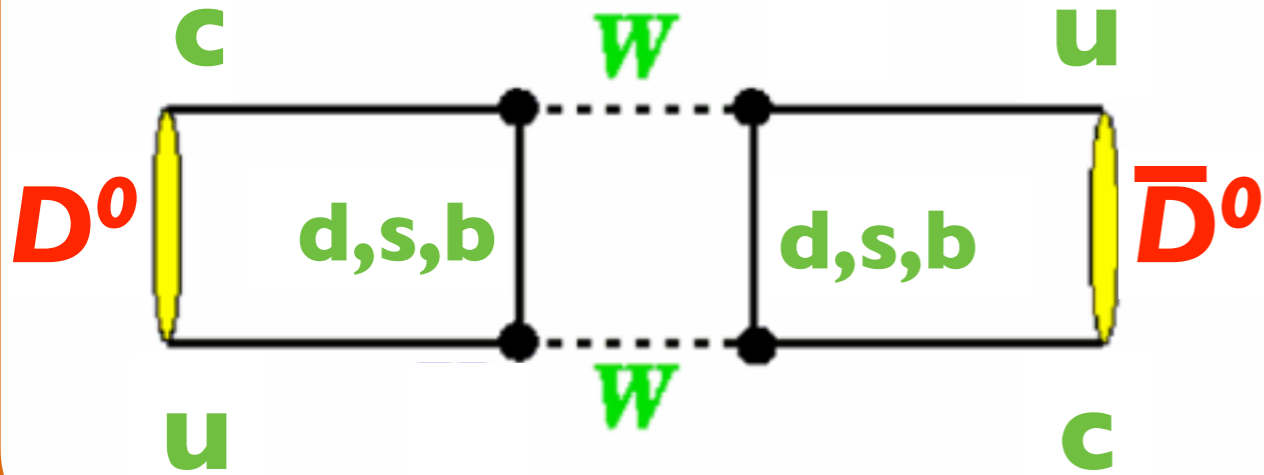
MLHEP in trigger





# Charm physics

And it is interesting!



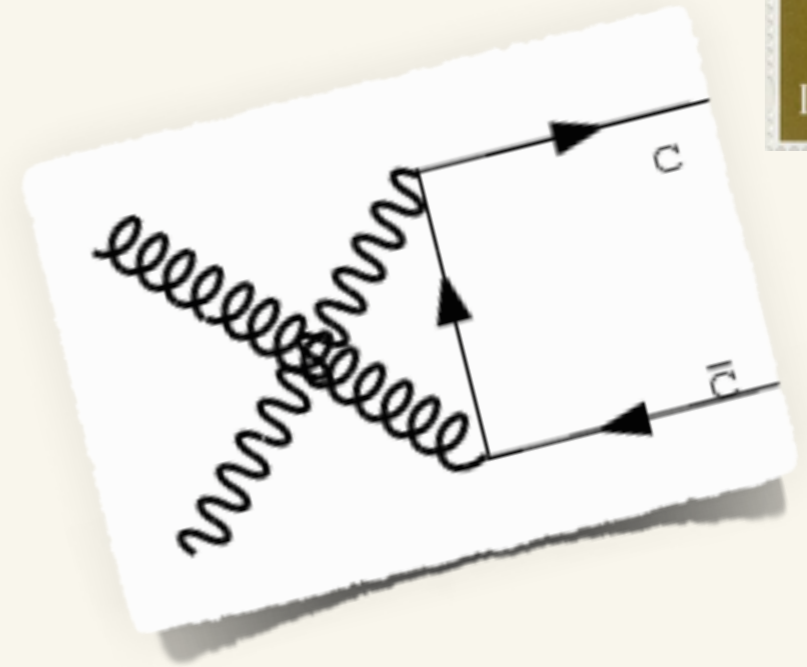
# How much charm?

**LHCb**

$\sigma_c$  around 2 mb

Lumi:  $4 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$

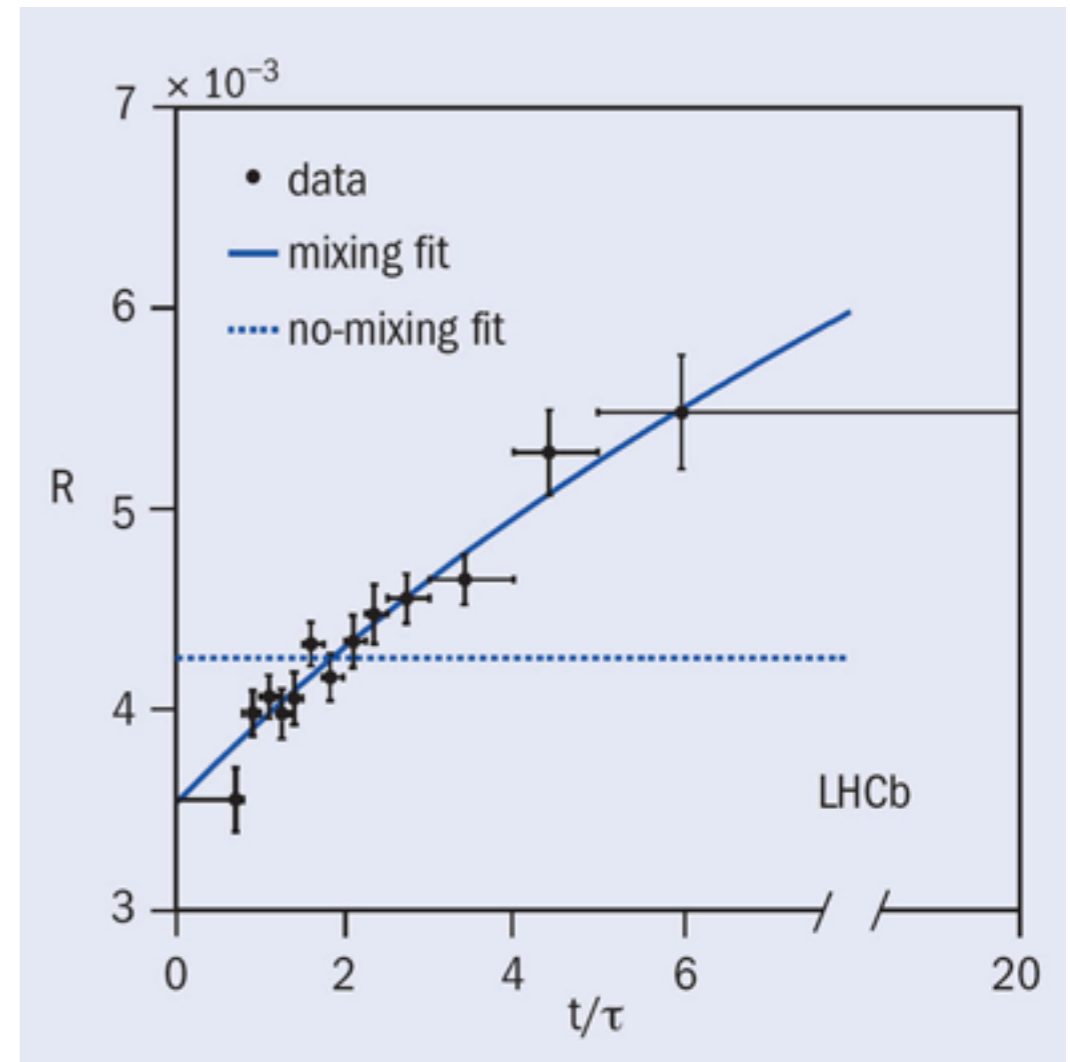
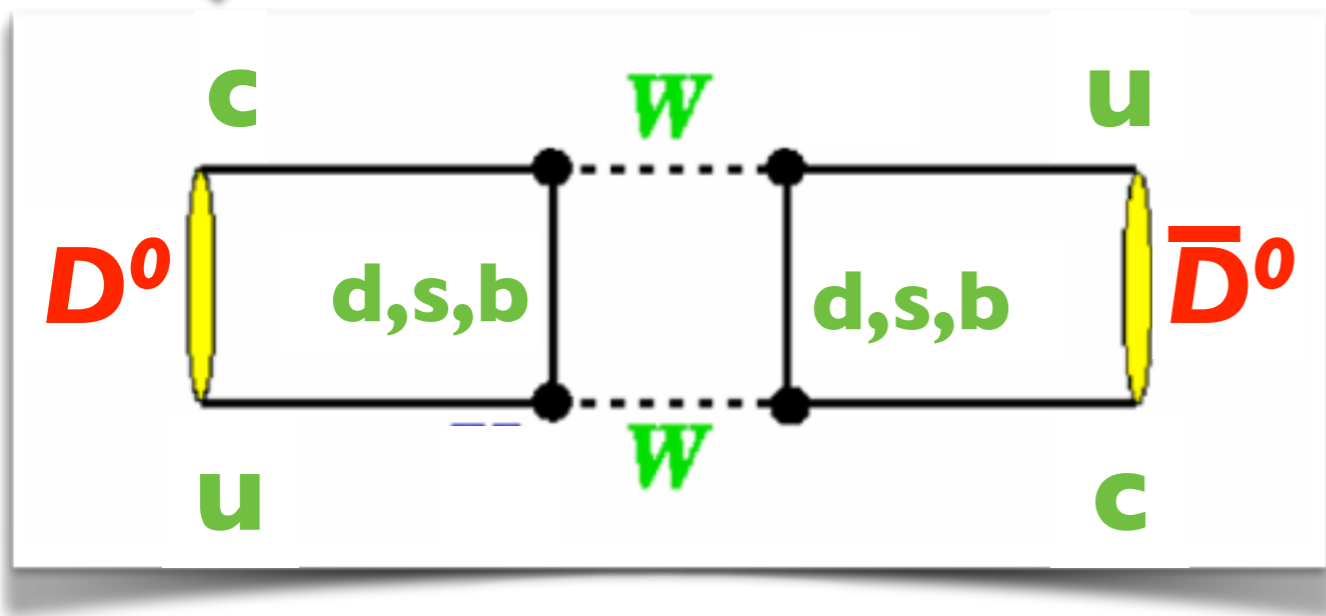
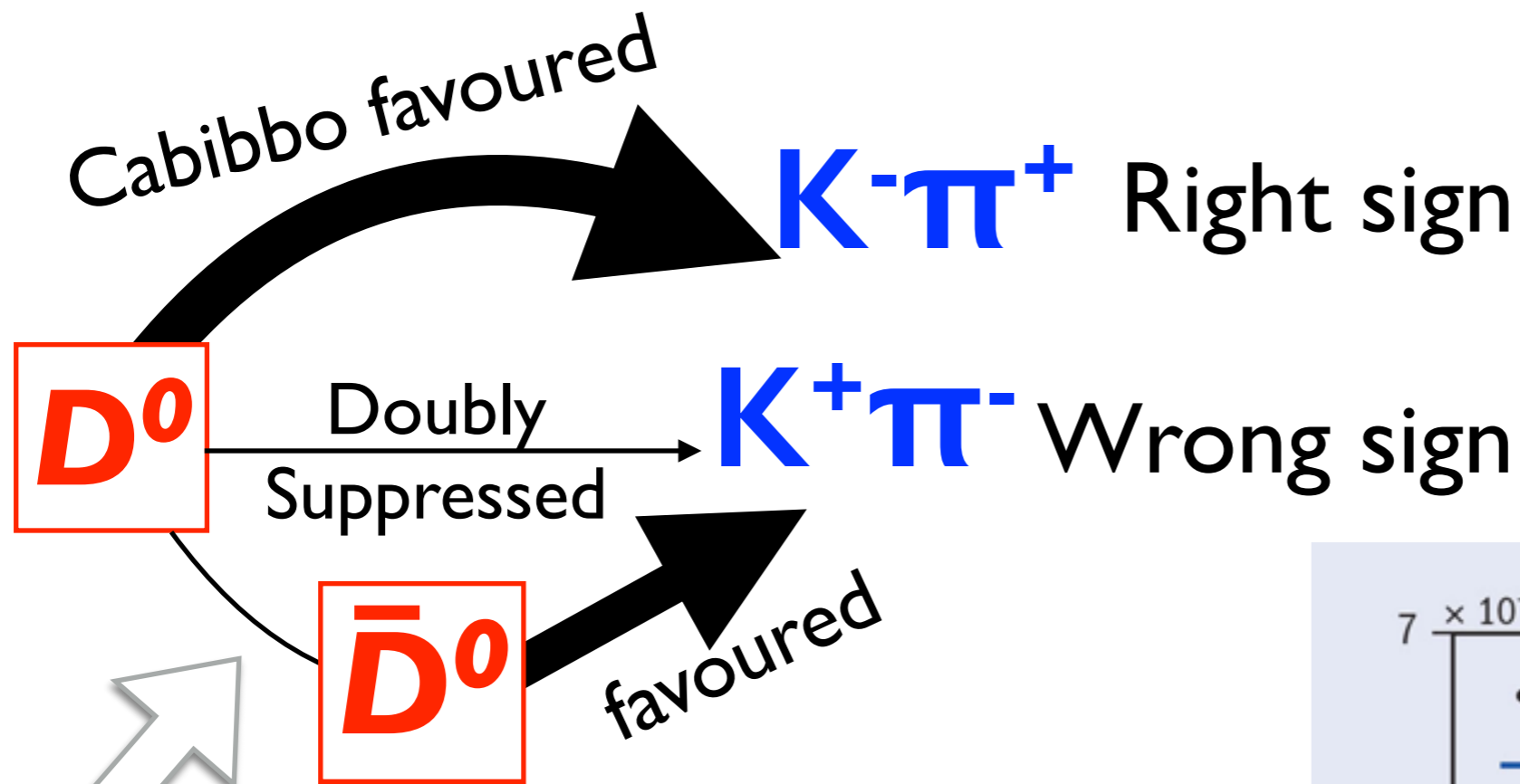
**Around 1 MHz of charm!**



Not all charm is interesting though...

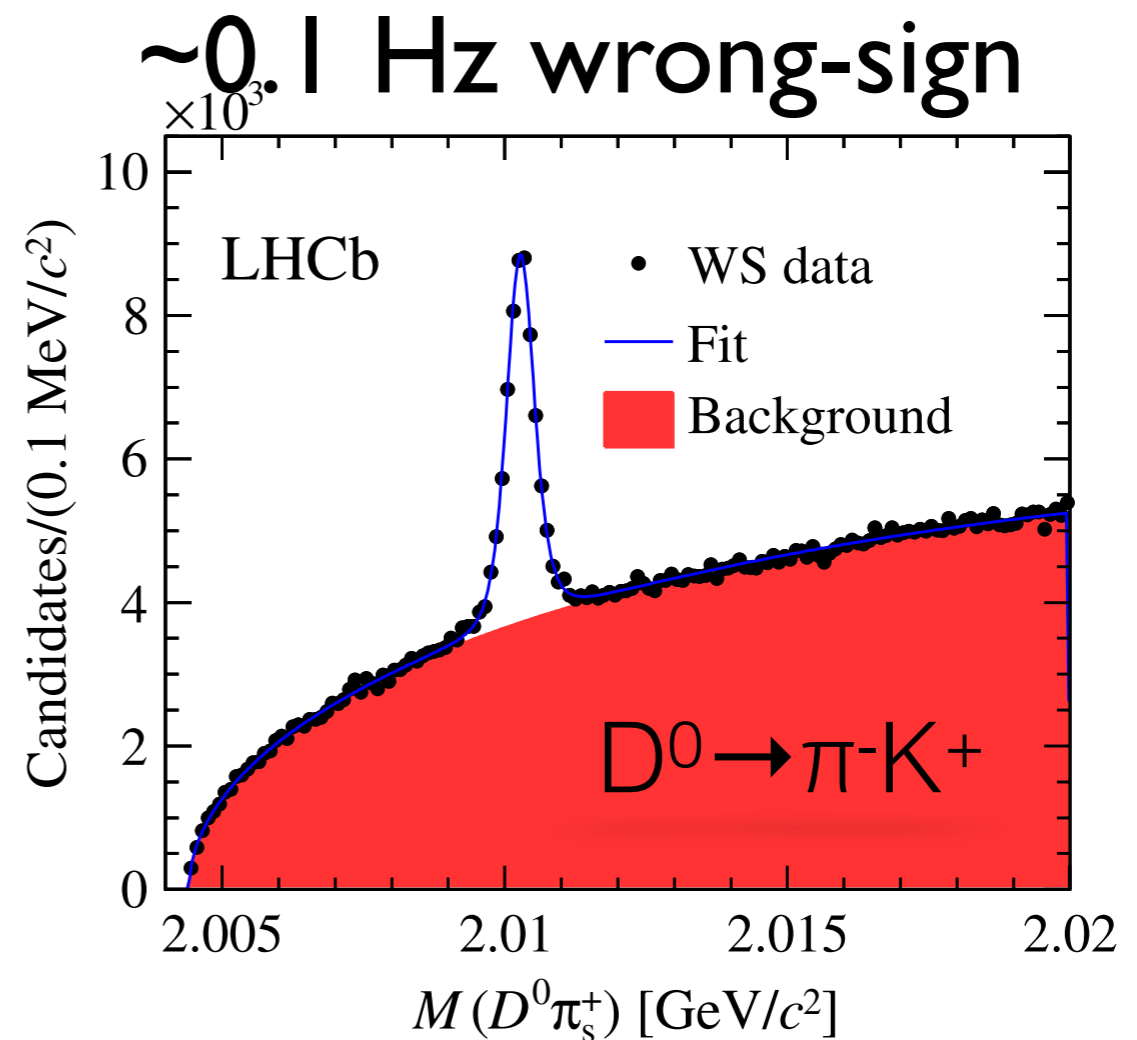
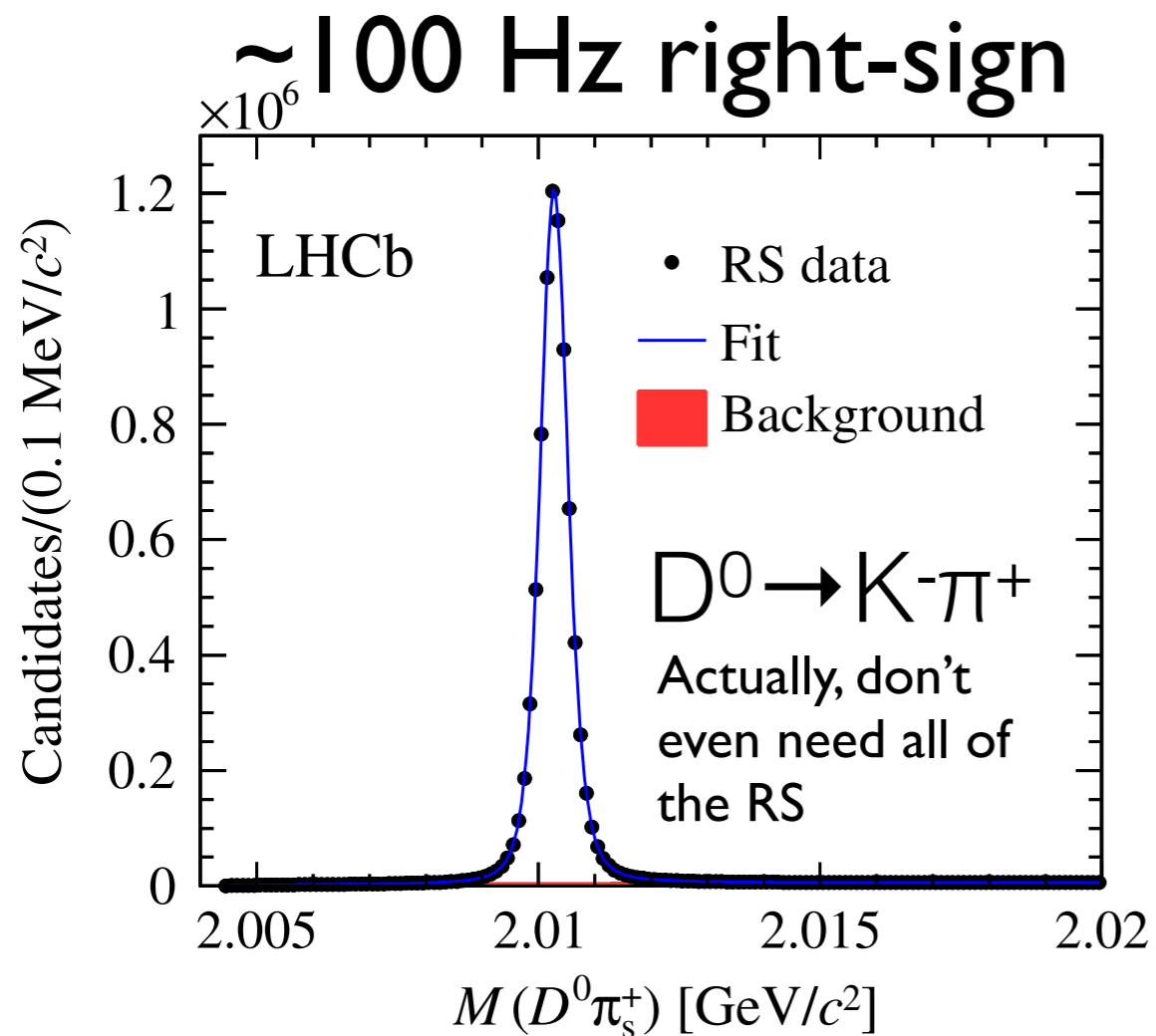


# Example of neutral D meson mixing



# Exclusive approach

Out of our 1 MHz of charm, we needed...



In reality LHCb has ~100 exclusive charm lines for different physics interests with 2-4 kHz



# High $p_T$ physics



# Charm physics



# Low level triggers



# Another bottleneck

- Ideally we would wait until we have all of the best possible information to hand, before deciding which events to keep.
- Offline event reconstruction: 1 second with CPU core
- Would need 40 million processes



**few billion €**  
 $(\text{€}1\text{k}/\text{PC}) * (40\text{e}6 \text{ procs}) / (40 \text{ cores})$

# And imagine moving all that data

- ATLAS/CMS @ 40 MHz means 40 TB/s
- And you need to assemble complete events to send to a single PC.

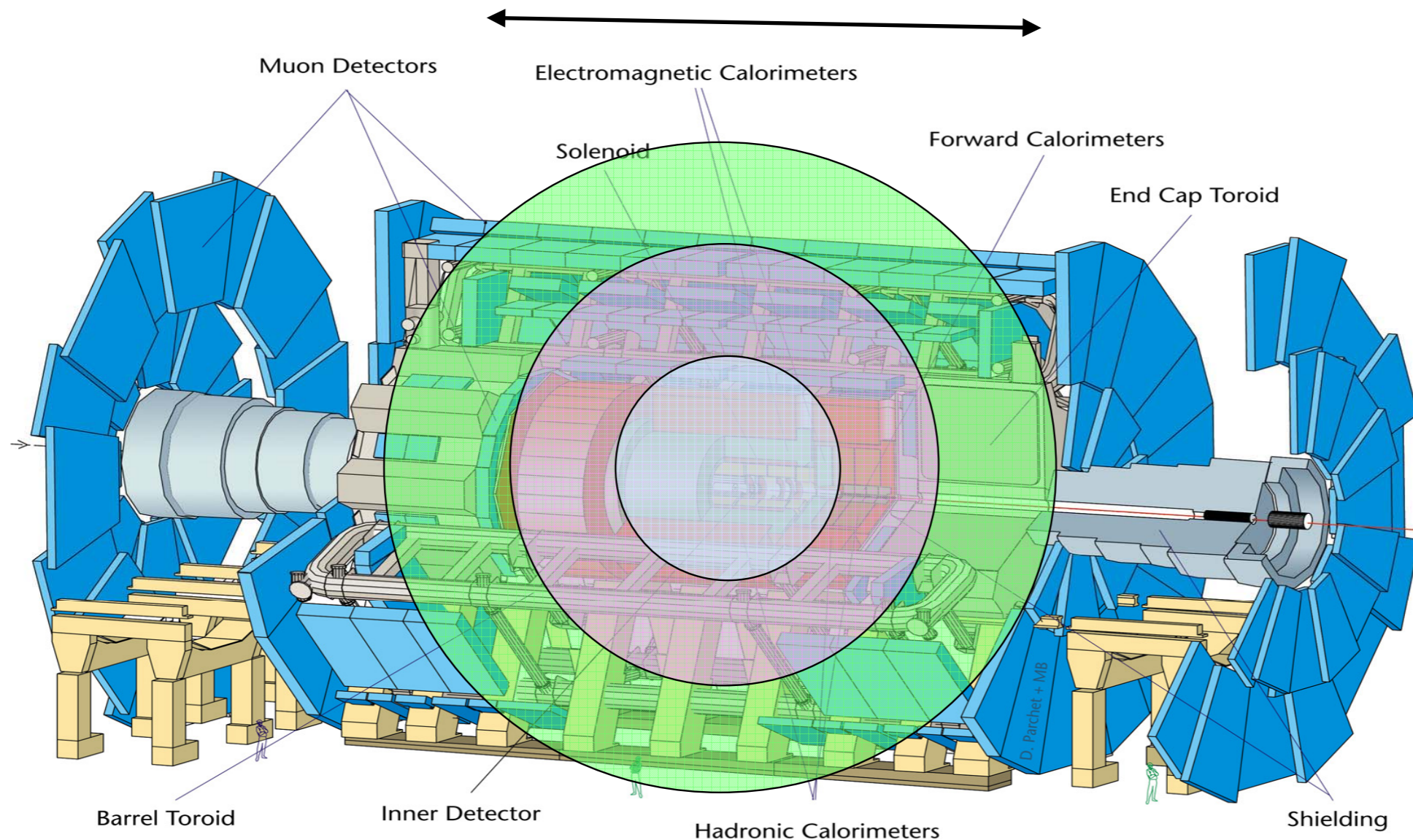
**Forget it!**

...until the LHCb (which has relatively small events) upgrade (2019).



# Time of flight

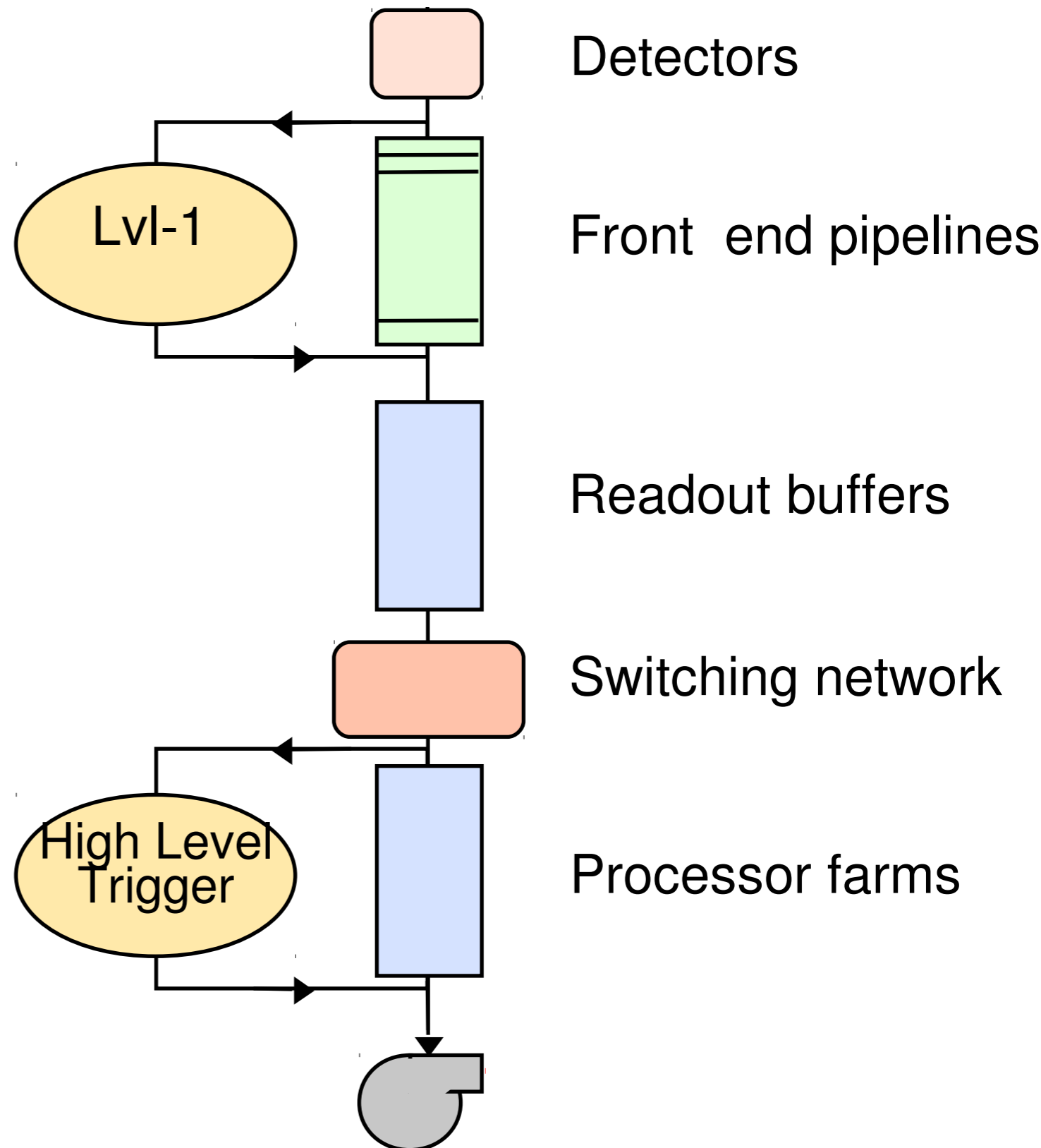
~25 meters



In 25ns, light travels 7.5m

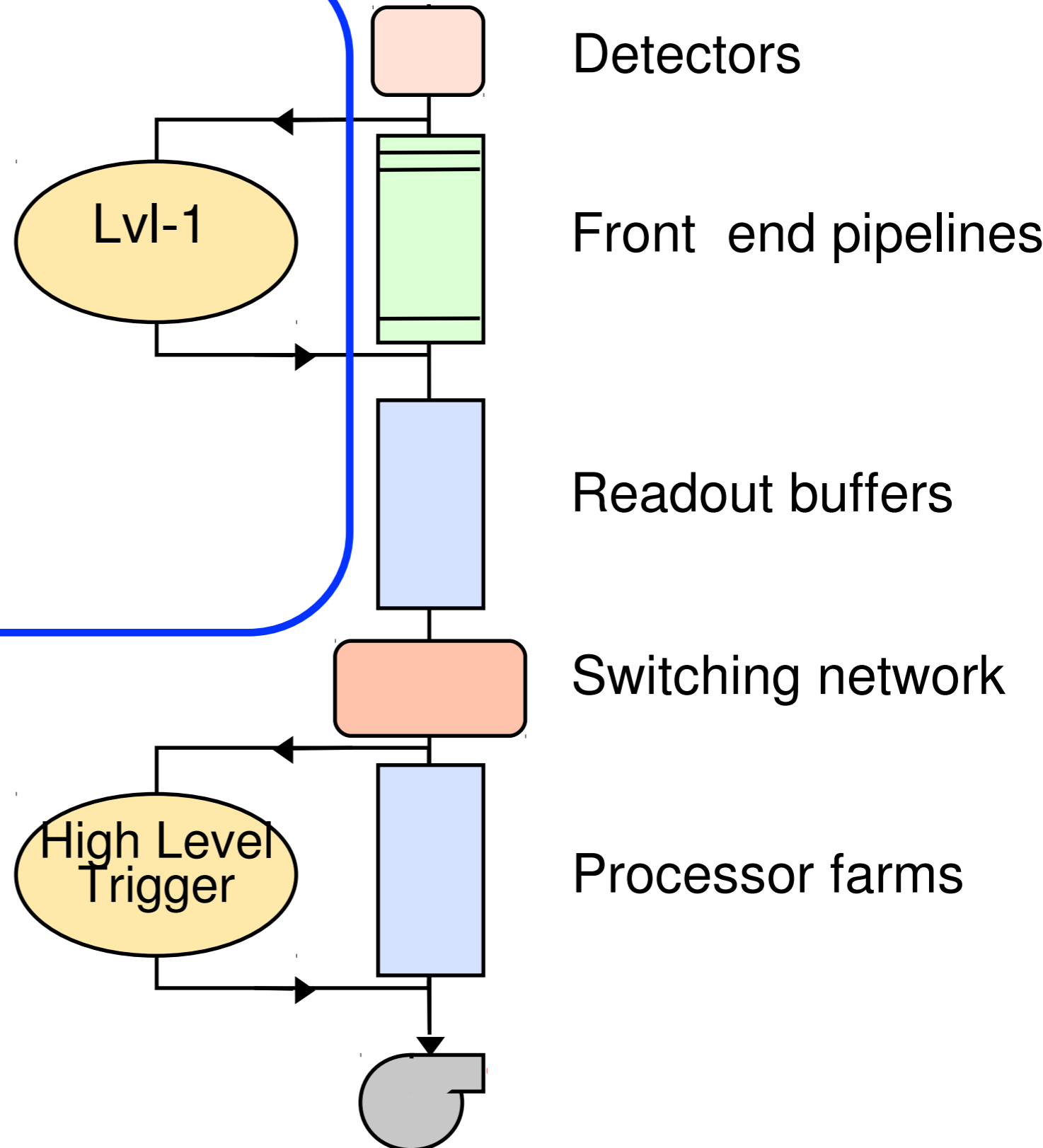
Already means that we need to buffer many bunch crossings while we wait for a decision.

# Real trigger and DAQ



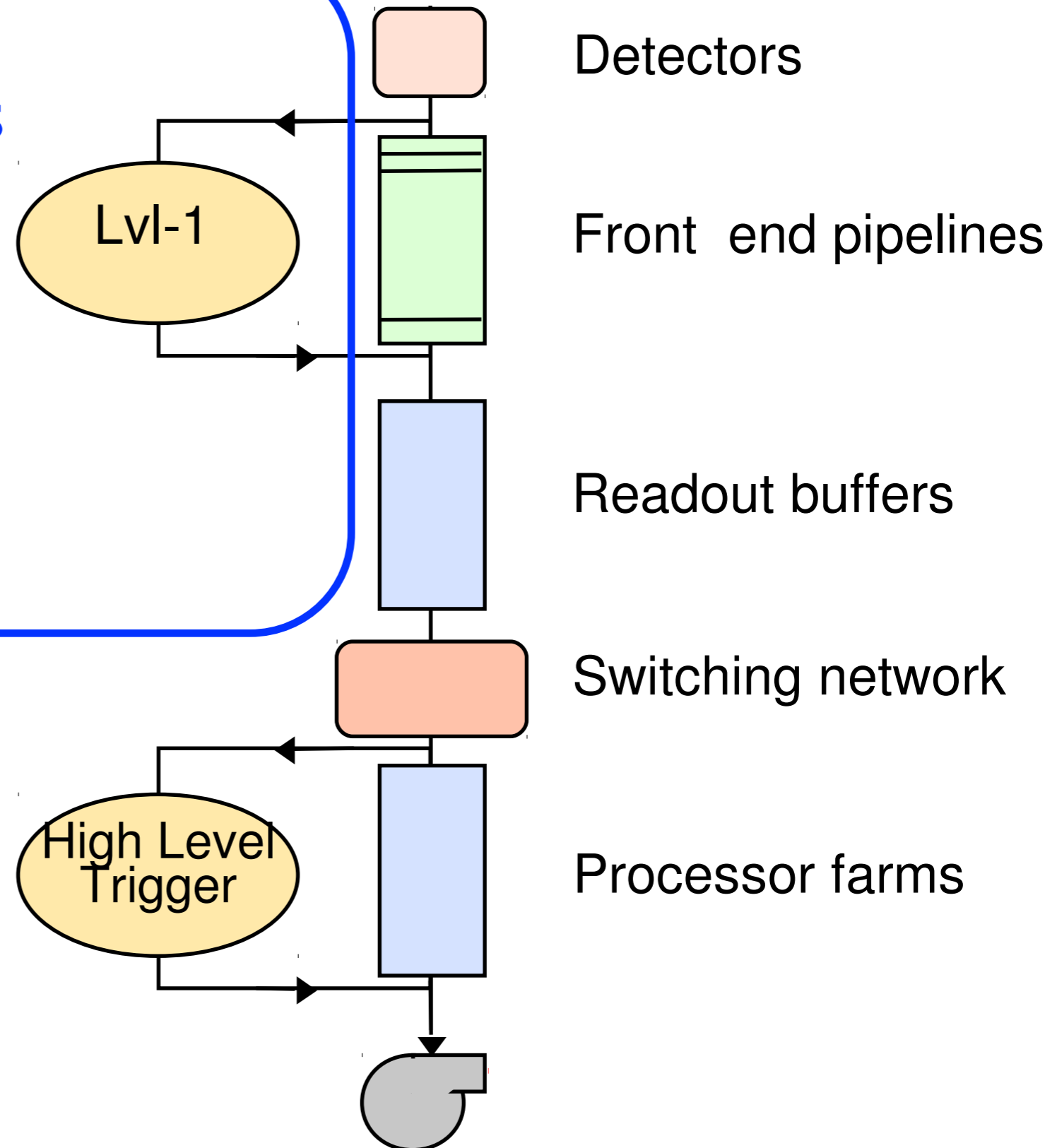
# Real trigger and DAQ

- Low level trigger to reduce the event rate to 0.1 - 1 MHz



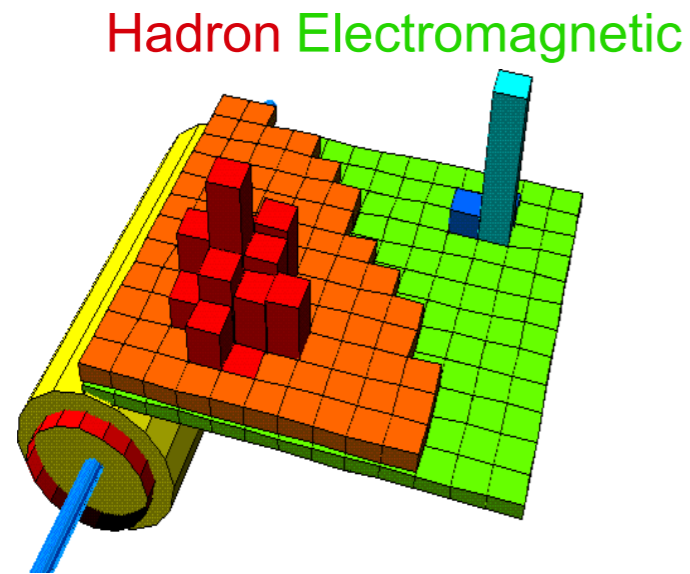
# Real trigger and DAQ

- New event every 25 ns
- FE pipelines can store around 100 crossings
- LLT needs to decide with a few  $\mu\text{s}$ !

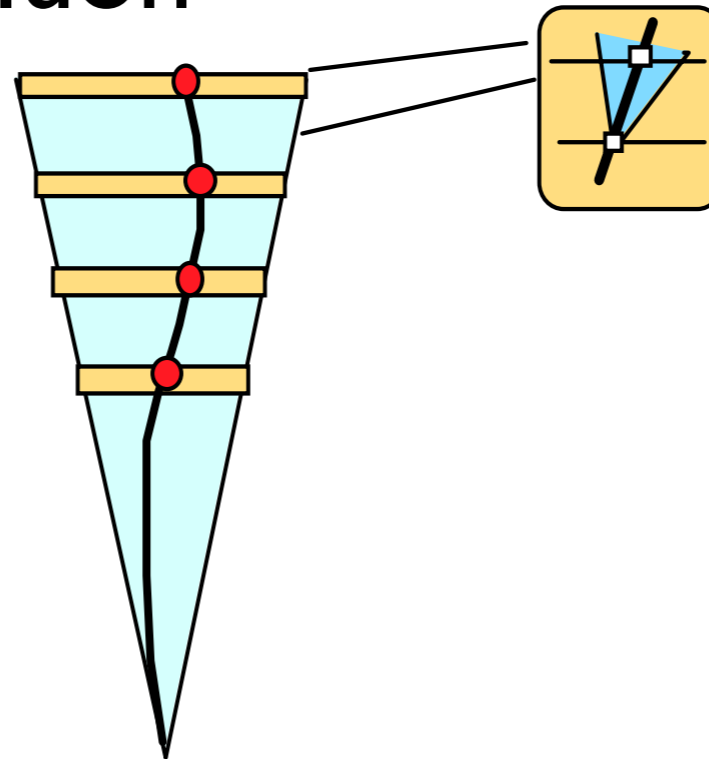


# Look at the easy stuff first

CALO



Muon

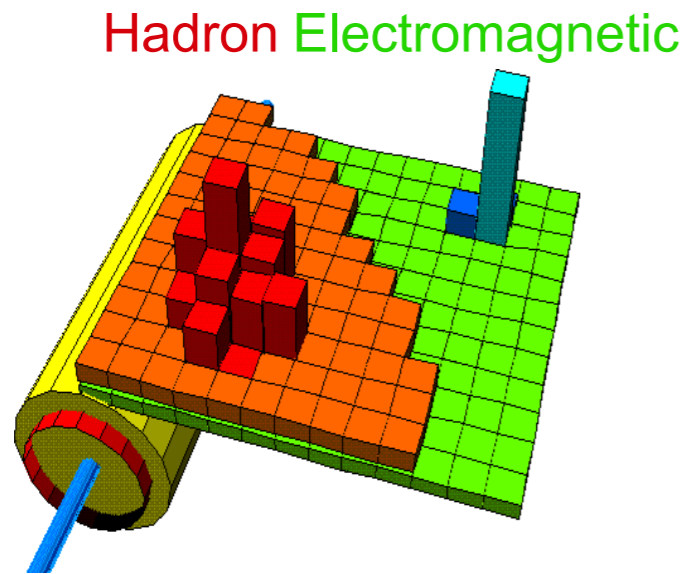


- Small amount of data
- Simple algorithms
- Local

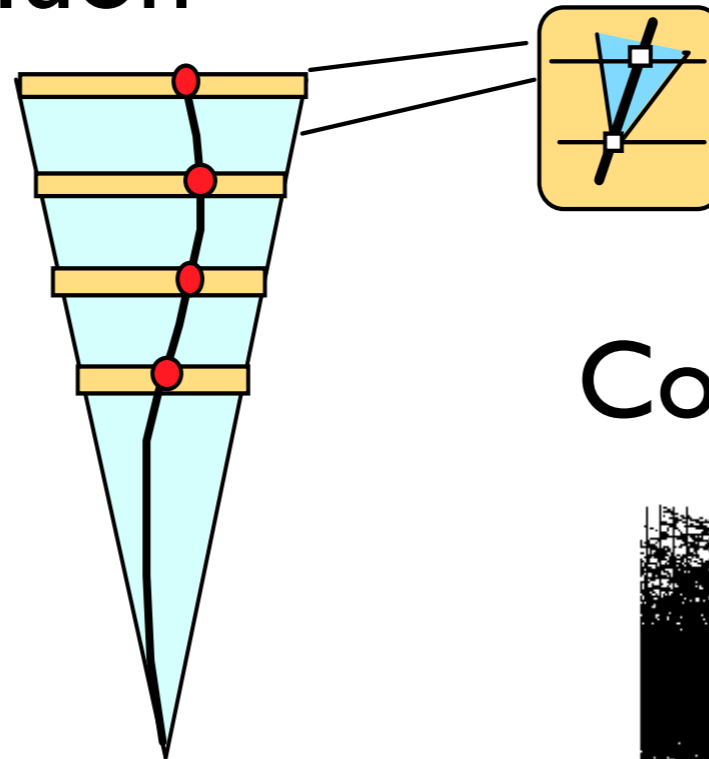


# Look at the easy stuff first

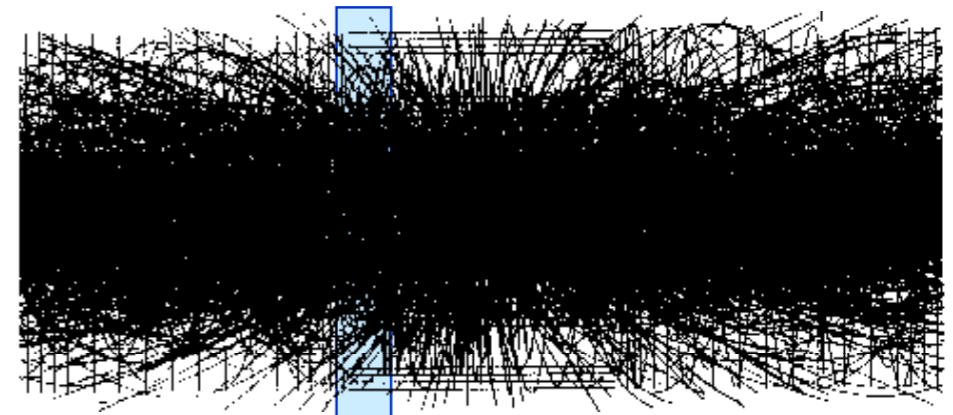
## CALO



## Muon

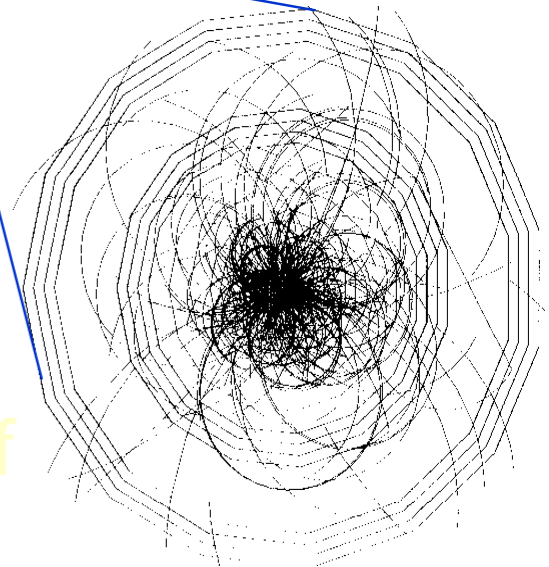


Compare with tracker!



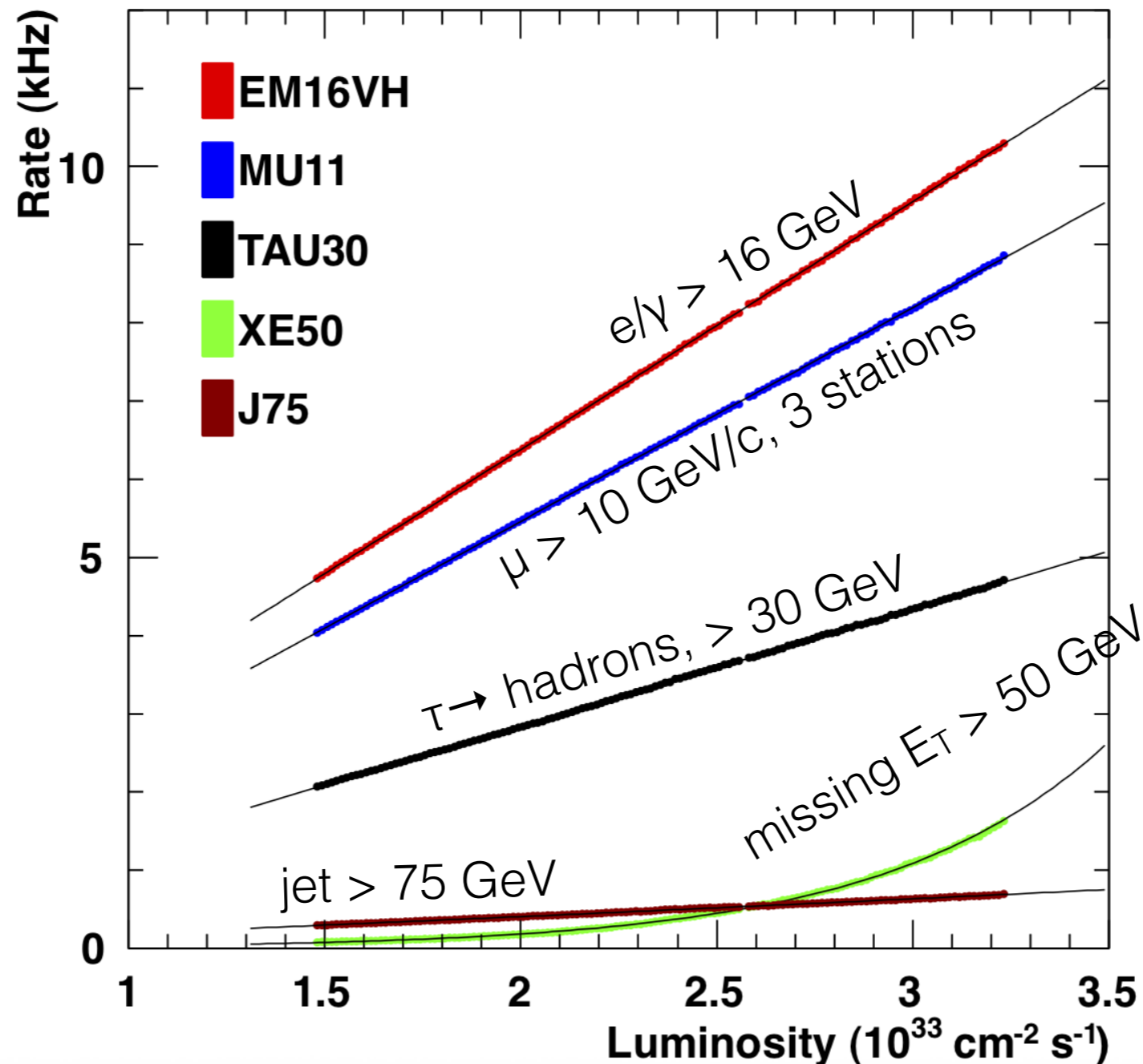
- Small amount of data
- Simple algorithms
- Local

- Complex algorithms
- Huge amounts of data



# Inclusive high $p_T$ muon and calo triggers

ATLAS Trigger Operations (Oct. 22, 2011)

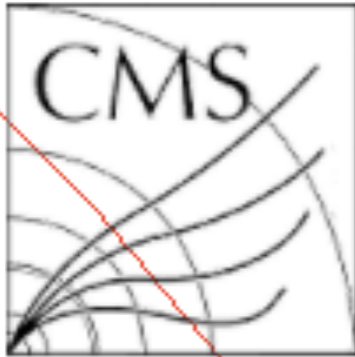


ATLAS and CMS can be efficient on the high  $p_T$  physics with  $\sim 100$  kHz LI output rate

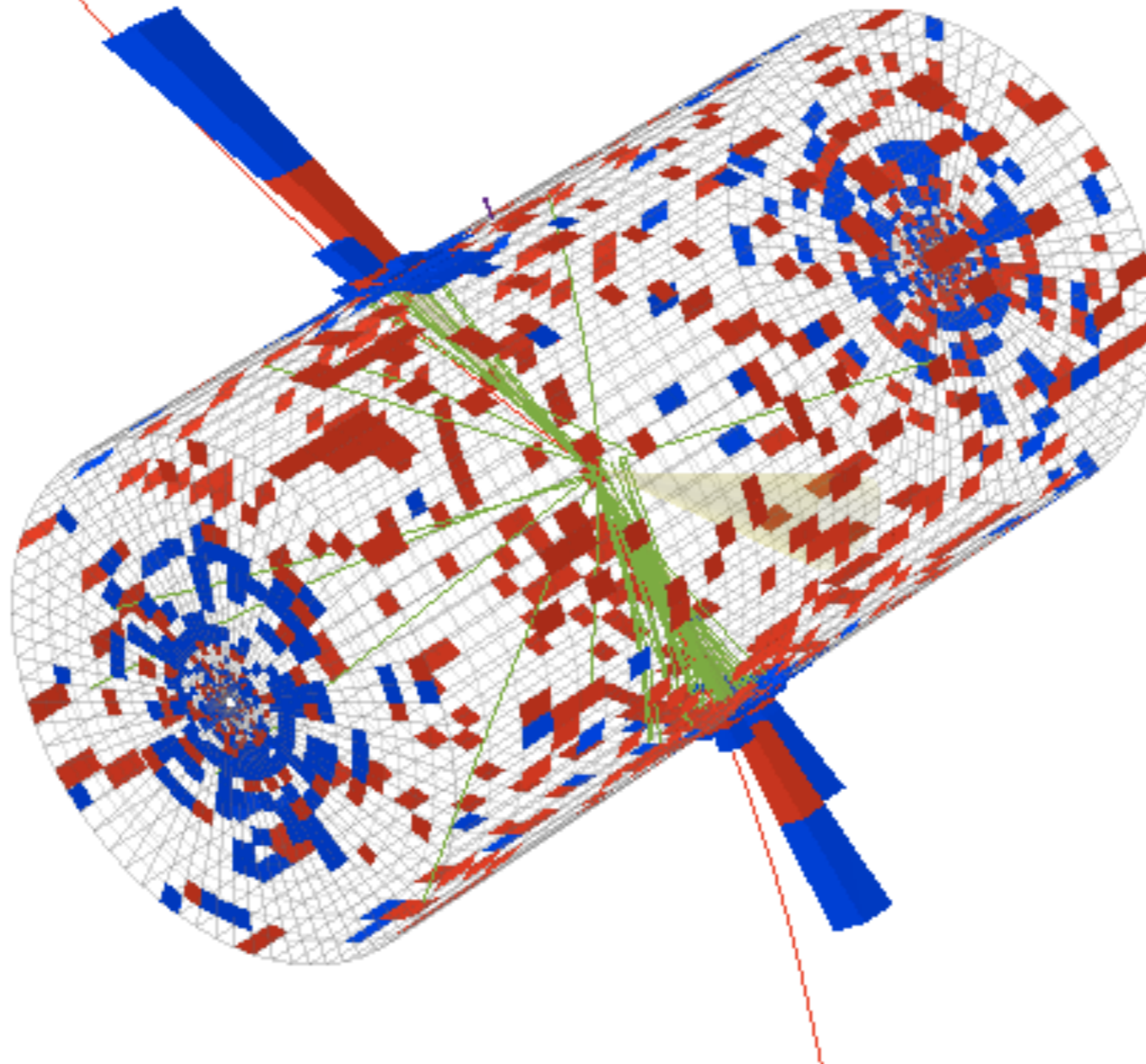
# Low level triggers

- Efficient when asked to find a clear feature that discriminates between signal and background.
- That can be identified with simple information

# For example



CMS Experiment at LHC, CERN  
Data recorded: Fri Oct 5 12:29:33 2012 CEST  
Run/Event: 204541 / 52508234  
Lumi section: 32

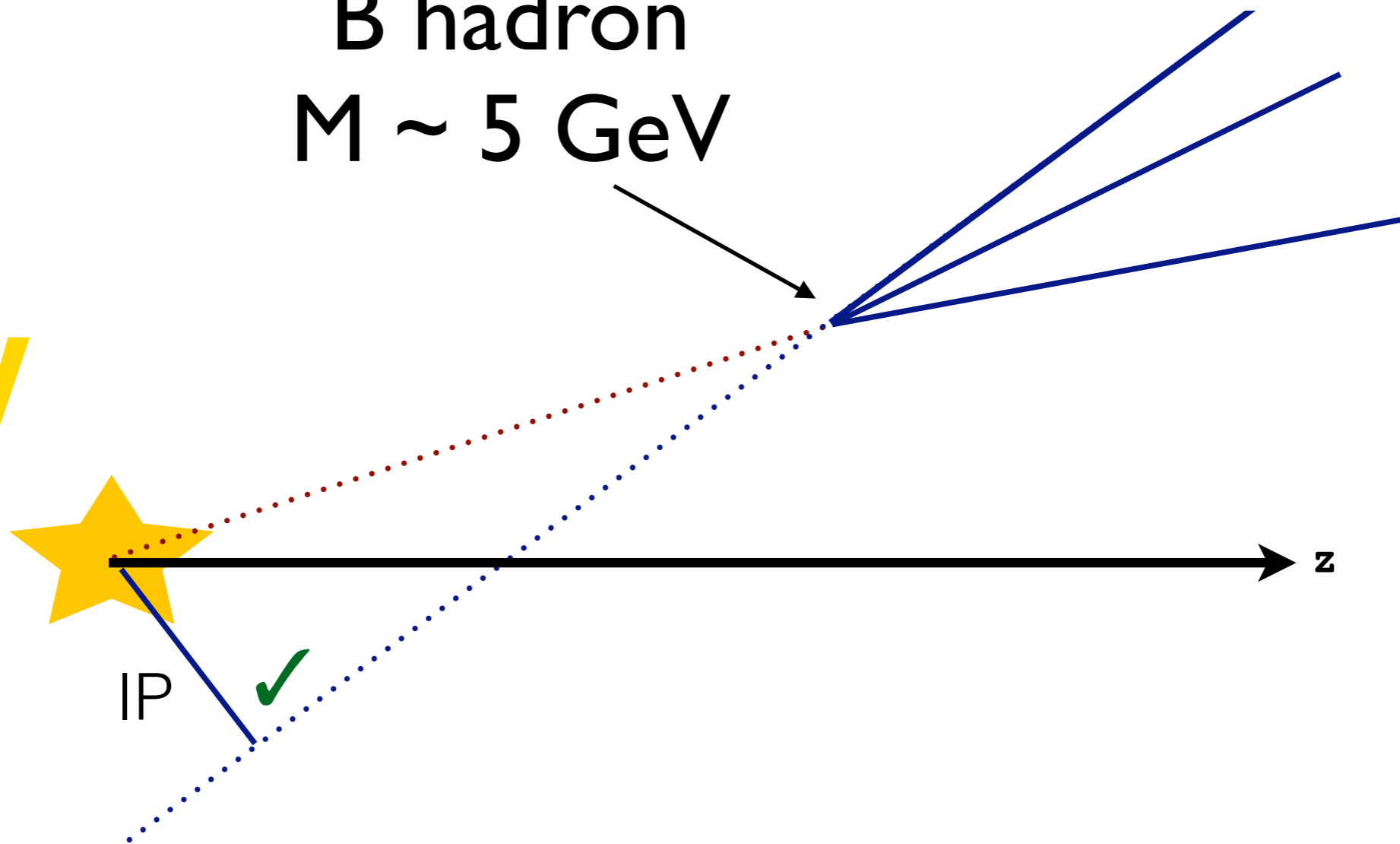


# LHCb case a little harder

Typical child,  
 $p_T \sim 1 \text{ GeV}$

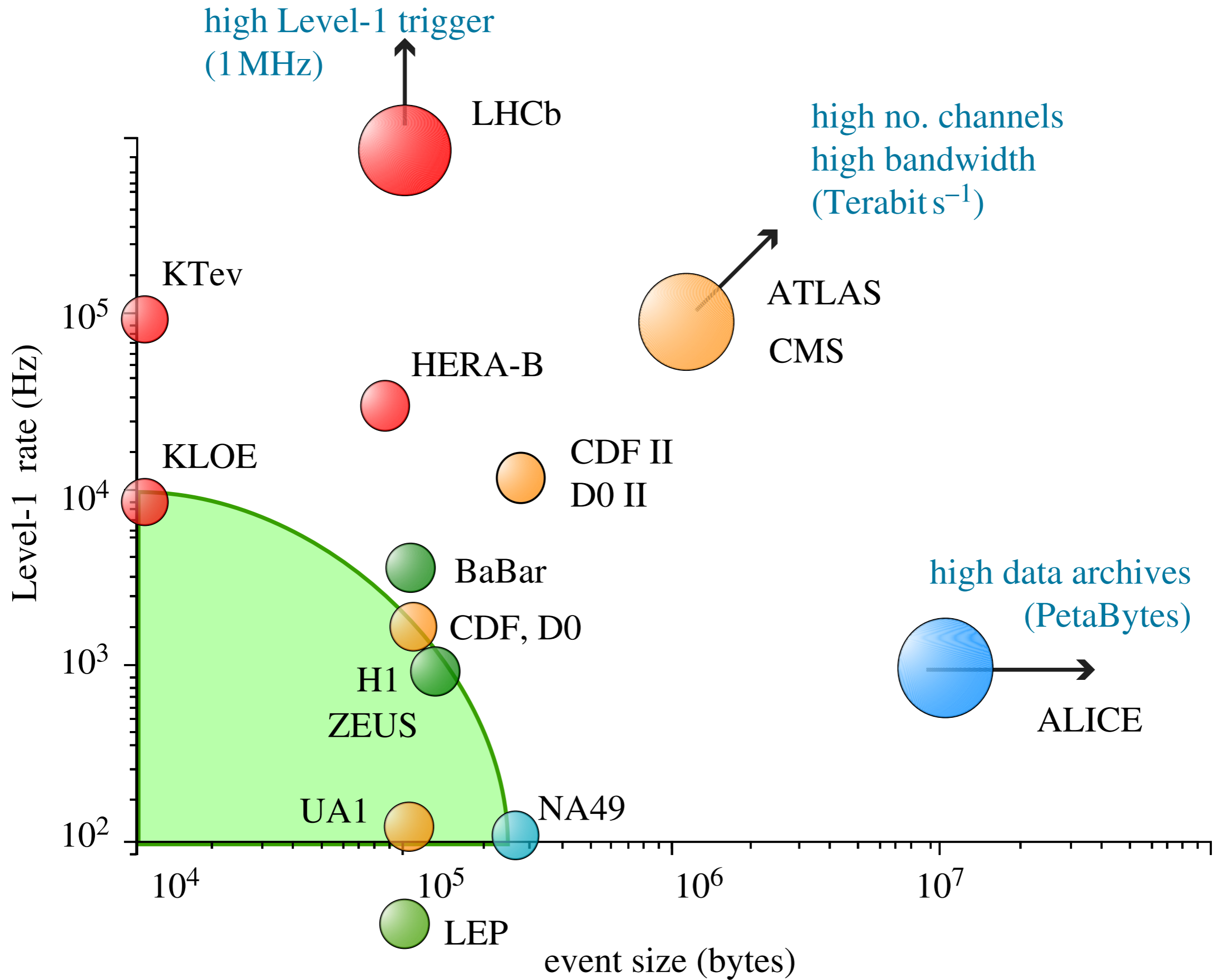
B hadron  
 $M \sim 5 \text{ GeV}$

PV

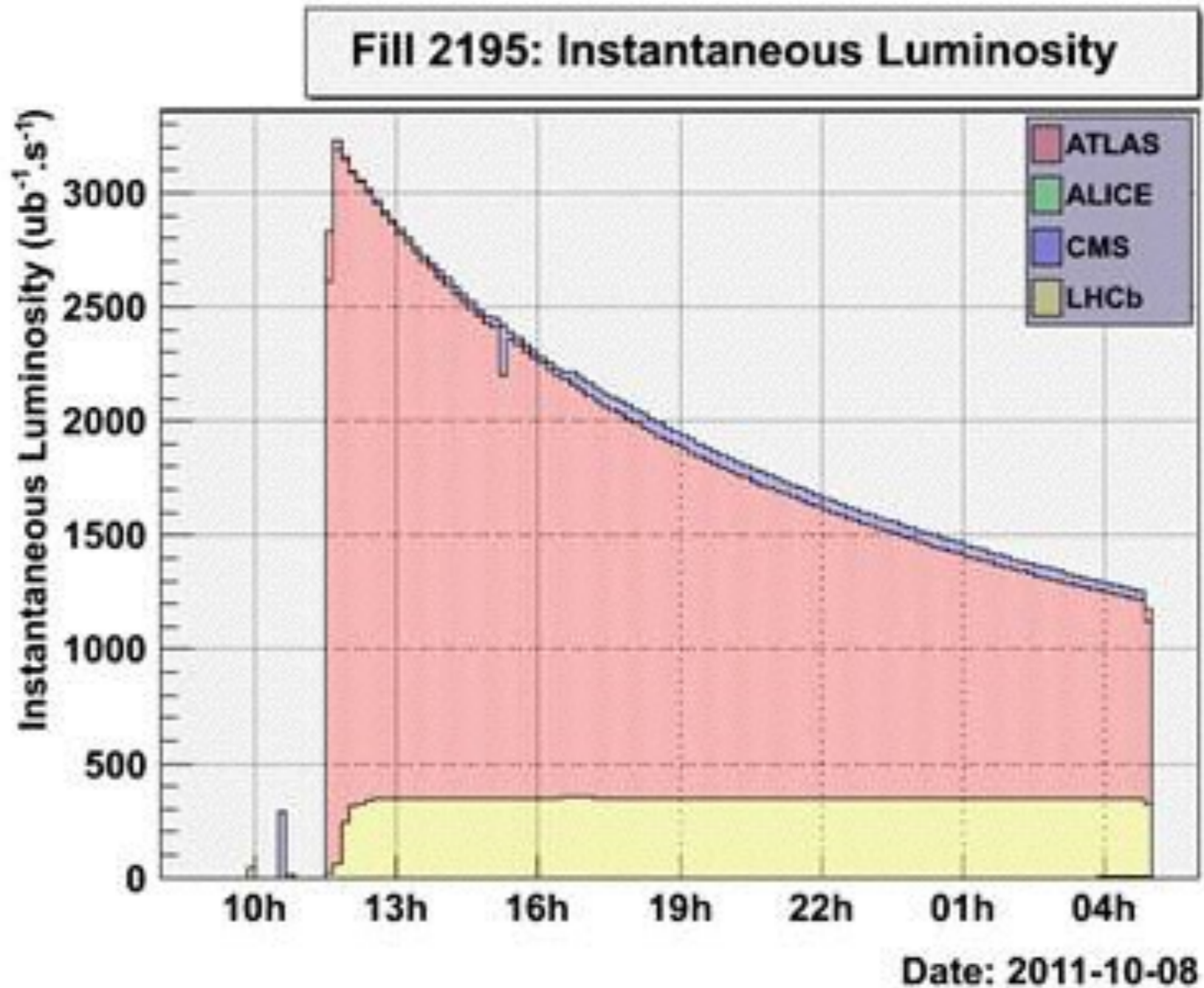


Can't reduce the rate as efficiently with  
low-level hardware trigger



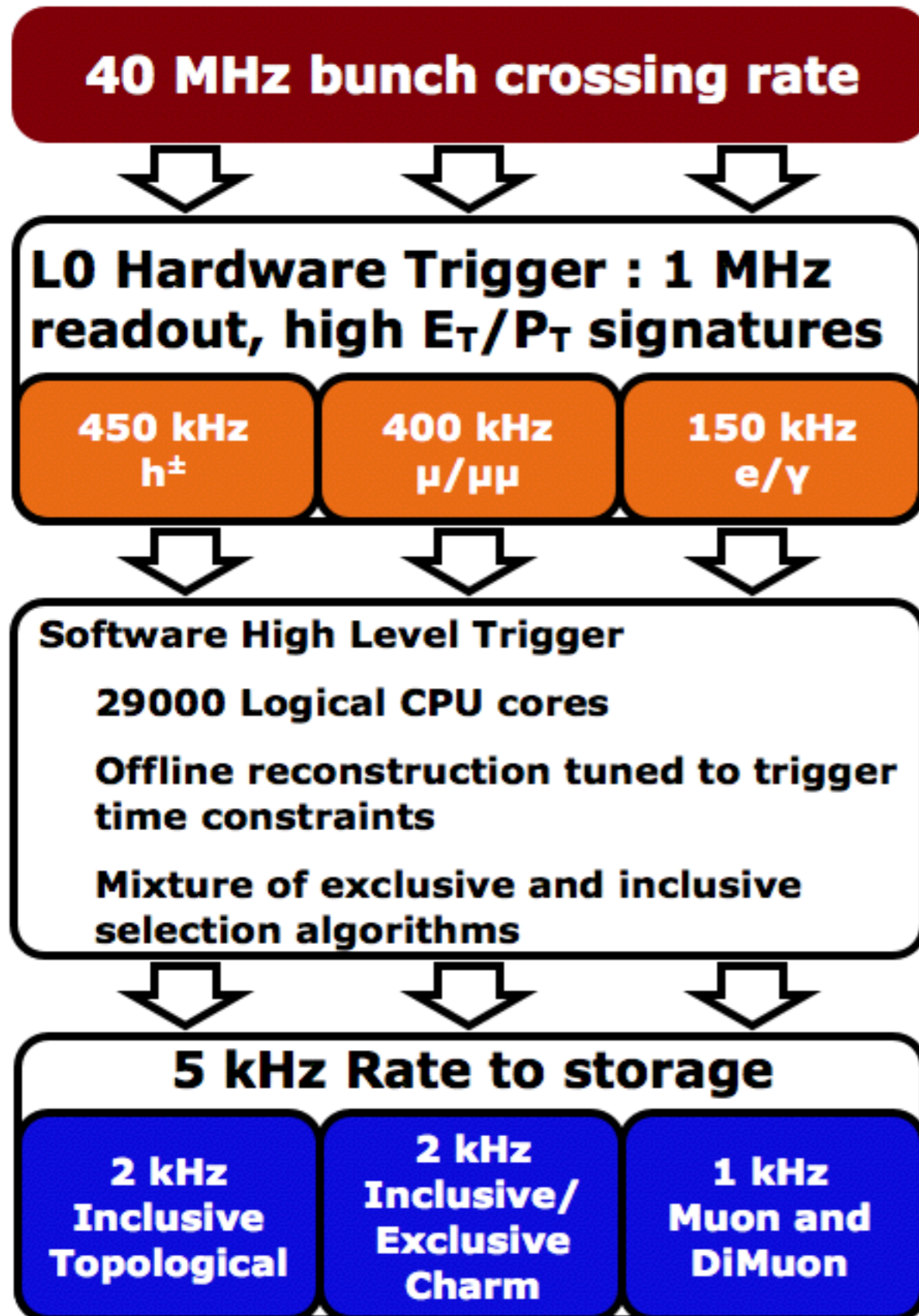


# Lumi levelling



LHCb levelled at  $4 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$

# LHCb Run-1 trigger



For 50:50 split of the 1 MHz between hadrons and muons:

- Muon  $p_T > 1.7$  GeV
- Hadron  $E_T > 3.5$  GeV

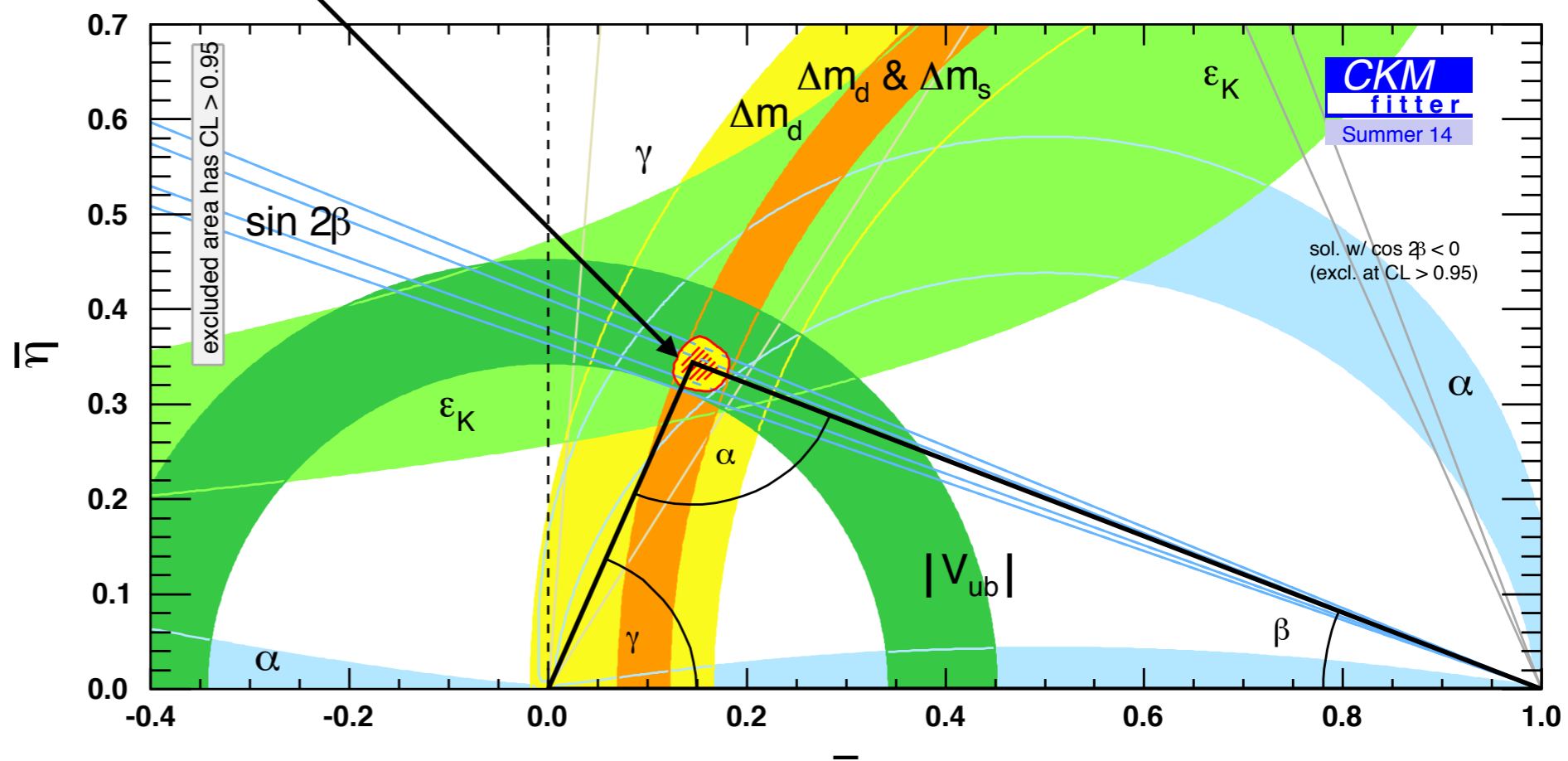
(Run-1)

# We need hadronic B for physics



$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}.$$

Unitarity





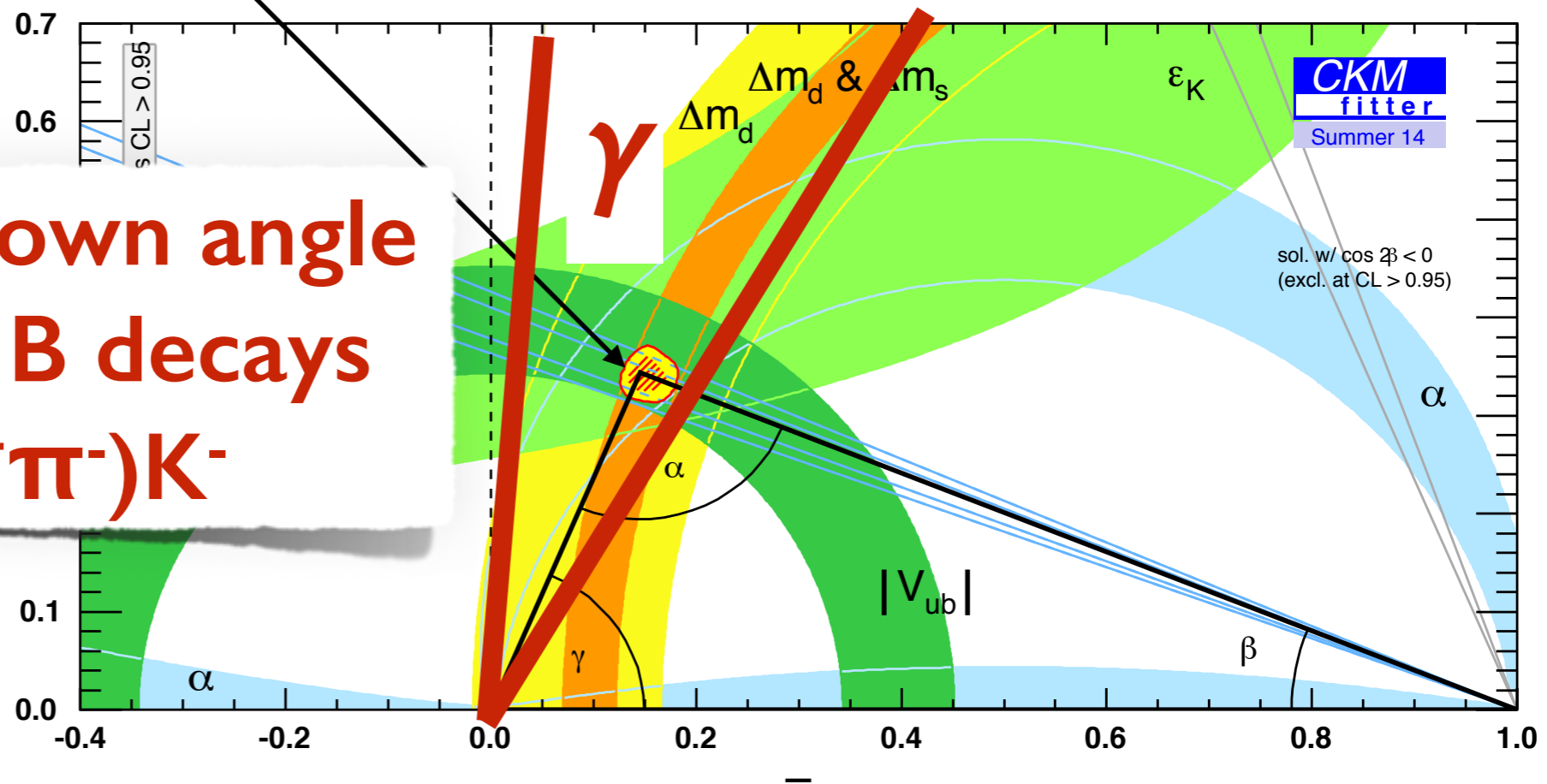
# We need hadronic B for physics



$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}.$$

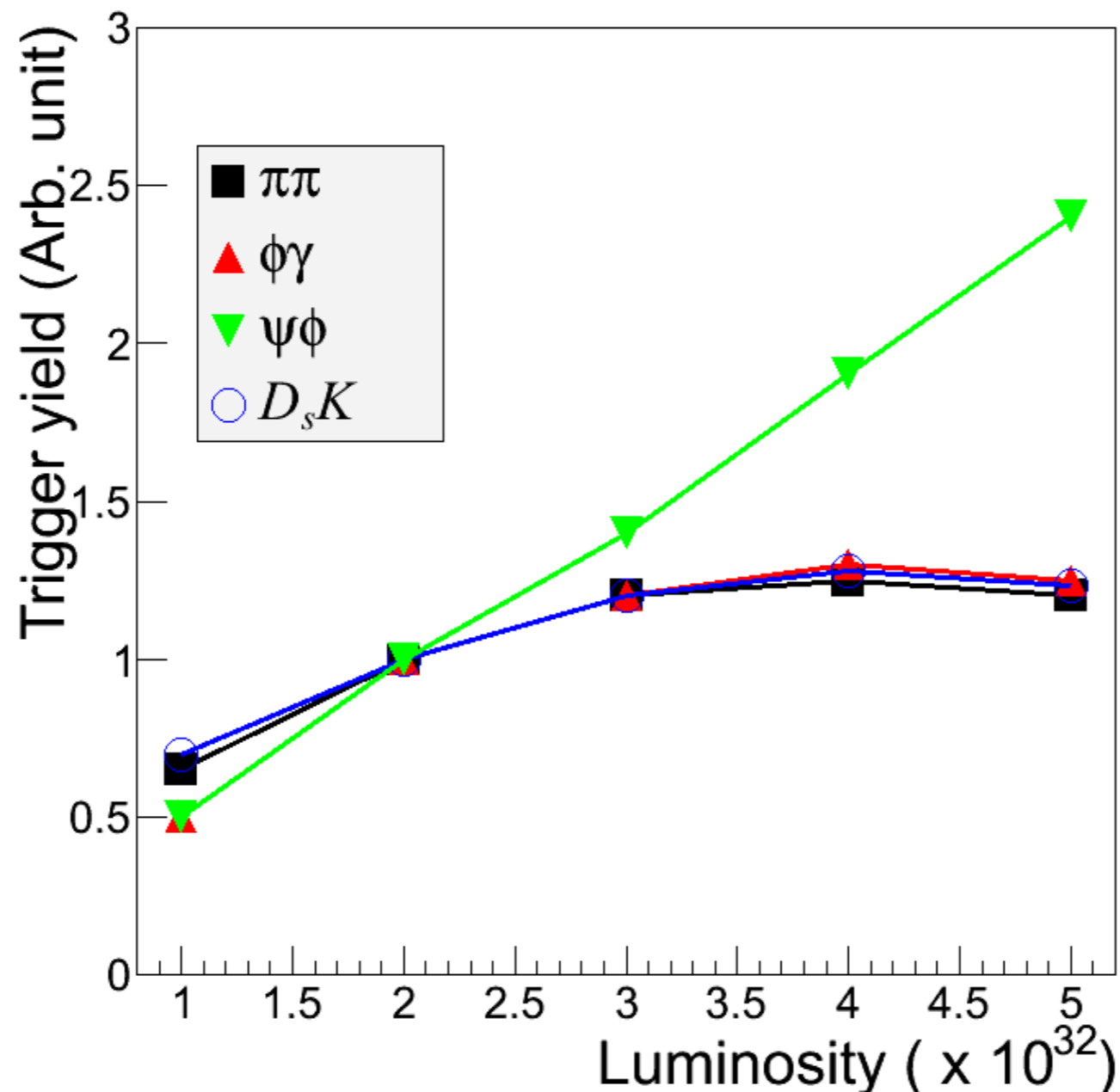
Unitarity

E.g. least well known angle  
needs hadronic B decays  
like  $D^0(\rightarrow K^+\pi^-)K^-$





# The LHCb upgrade (2019)



In Run-III (2020), LHCb will have a full software trigger at 30 MHz

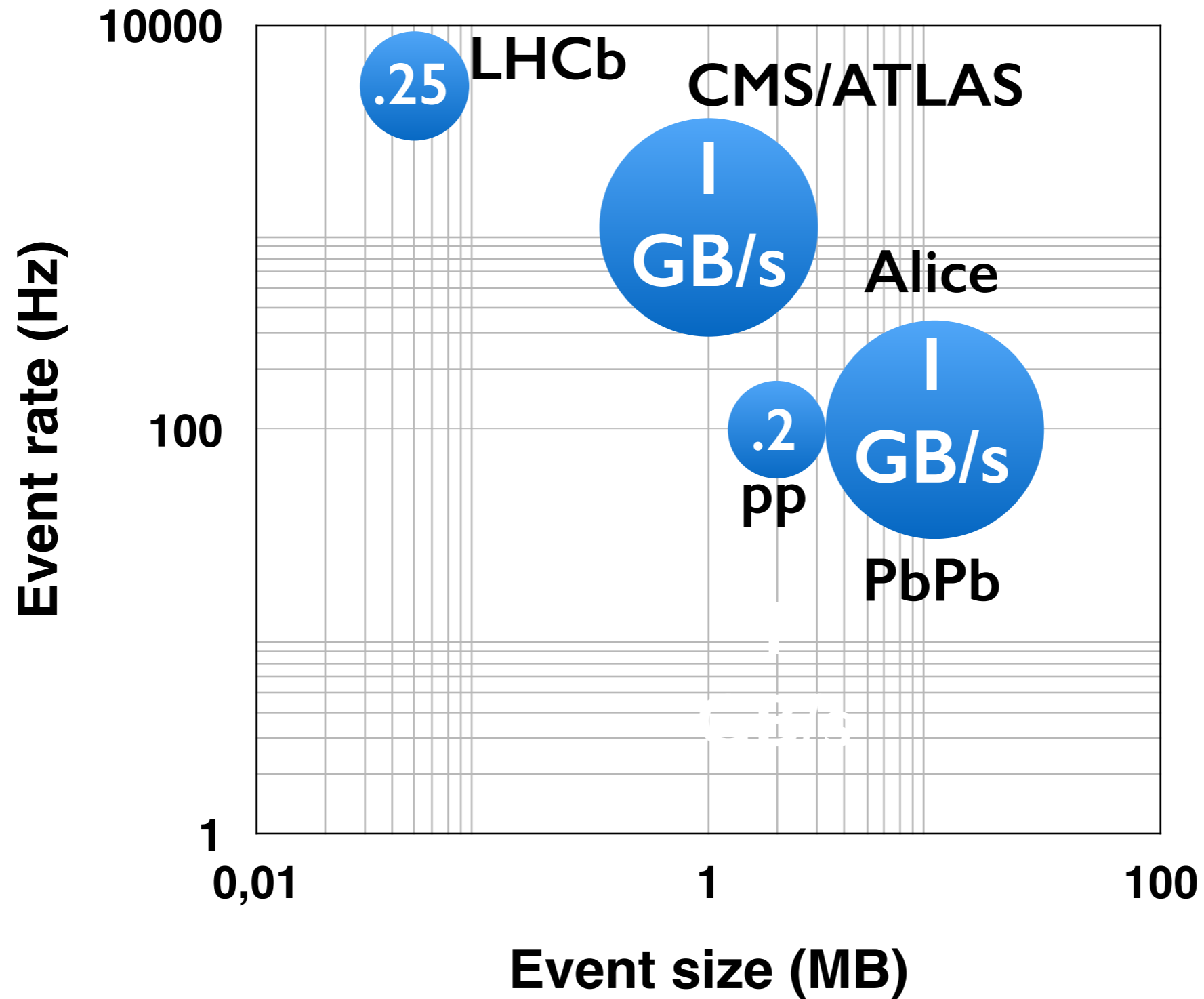
# High level triggers

# How much data can we write out?



Limited by tape and disk storage,  
and ability to reconstruct and process the data

# Run-I numbers



# The High Level Triggers

Thanks to the LLT and event builder, we can now perform full event reconstruction with off-the-shelf PCs

Run-I specifications:

<b>ATLAS</b>	<b>17k CPU cores*</b>
<b>CMS</b>	<b>13k CPU cores</b>
<b>LHCB</b>	<b>29k CPU cores</b>

Use the same event reconstruction software as used offline, but tuned/configured differently.

Build an OR of 100-1000 different “lines”

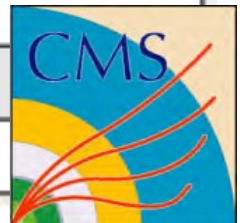
\*split between L2 and L3



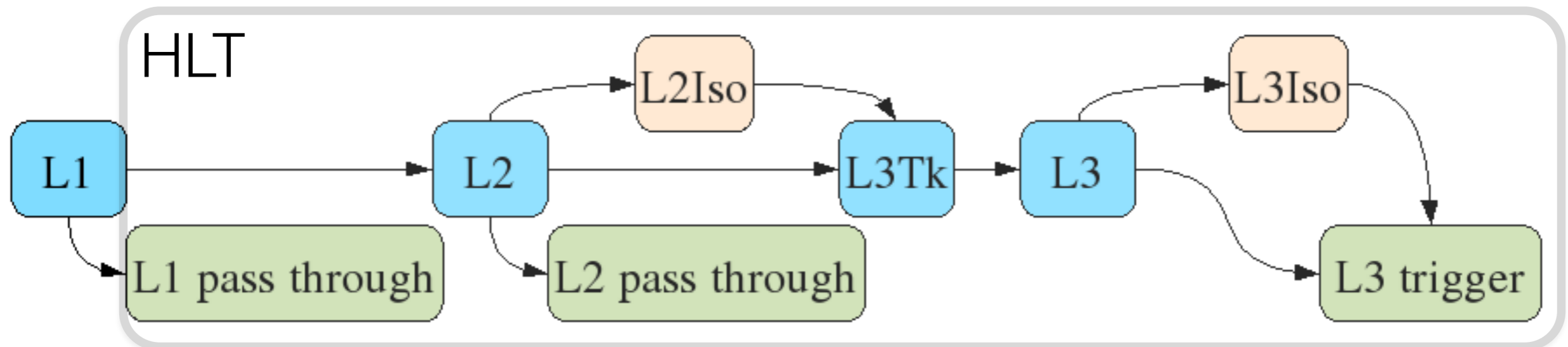
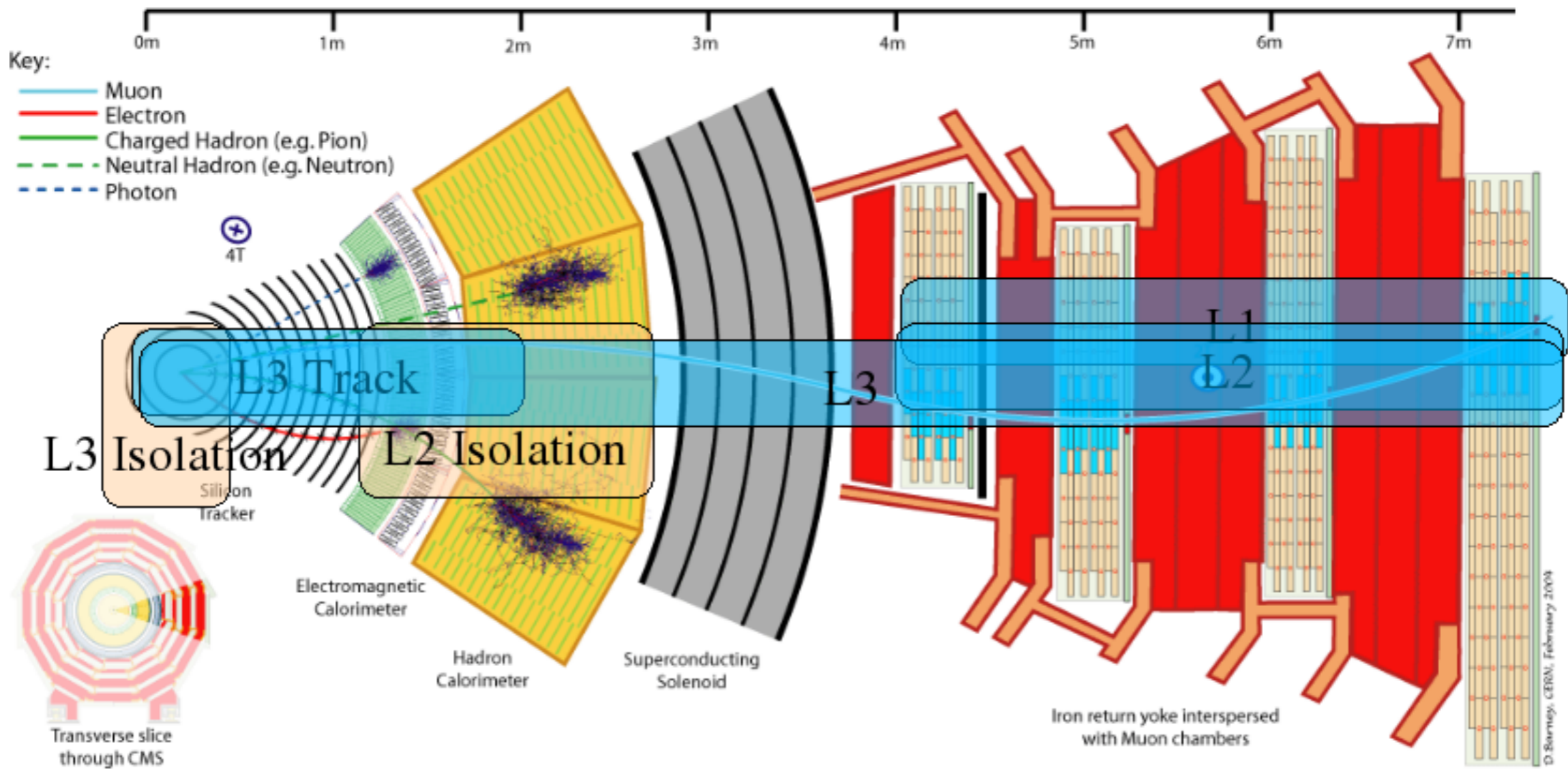
# ATLAS and CMS

“Menus” of mostly inclusive high  $p_T$  physics lines

(Unprescaled) Object	Trigger Threshold (GeV)	Rate (Hz)	Physics
Single Muon	40	21	Searches
Single Isolated muon	24	43	Standard Model
Double muon	(17, 8) [13, 8 for parked data]	20 [30]	Standard Model / Higgs
Single Electron	80	8	Searches
Single Isolated Electron	27	59	Standard Model
Double Electron	(17, 8)	8	Standard Model / Higgs
Single Photon	150	5	Searches
Double Photon	(36, 22)	7	Higgs
Muon + Ele x-trigger	(17, 8), (5, 5, 8), (8, 8, 8)	3	Standard Model / Higgs
Single PFJet	320	9	Standard Model
QuadJet	80 [50 for parked data]	8[100]	Standard Model / Searches
Six Jet	(6 x 45), (4 x 60, 2 x 20)	3	Searches
MET	120	4	Searches
HT	750	6	Searches

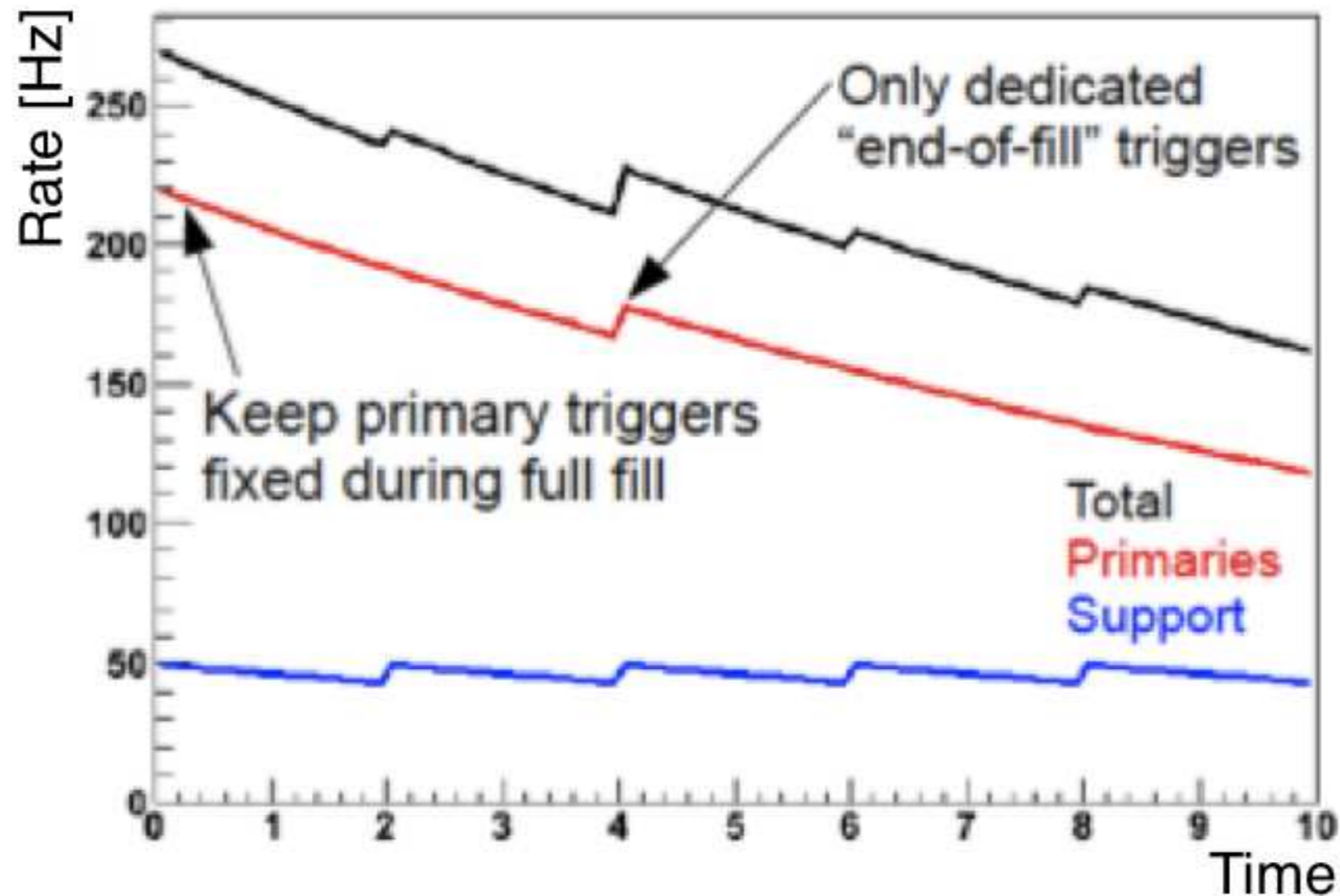


# CMS muon example



# Changing conditions

A fill of the LHC machine lasts  $\sim 10$  hours.  
Exponential decay of the luminosity.



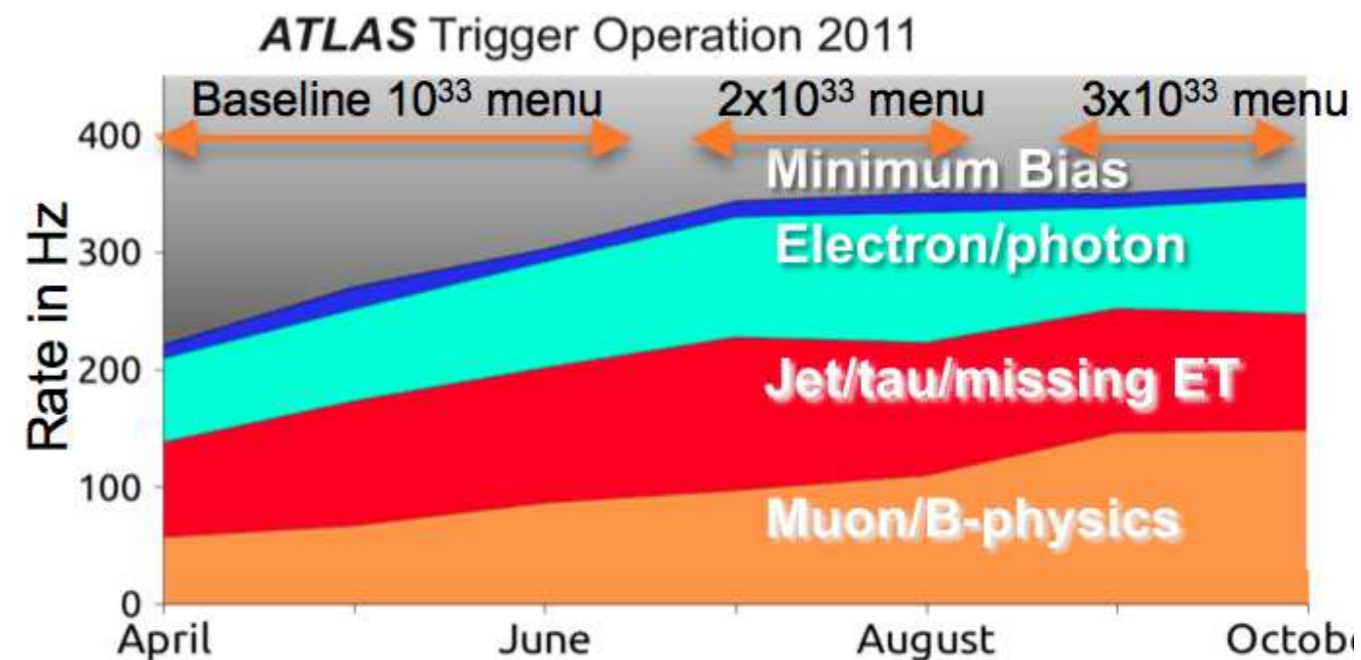
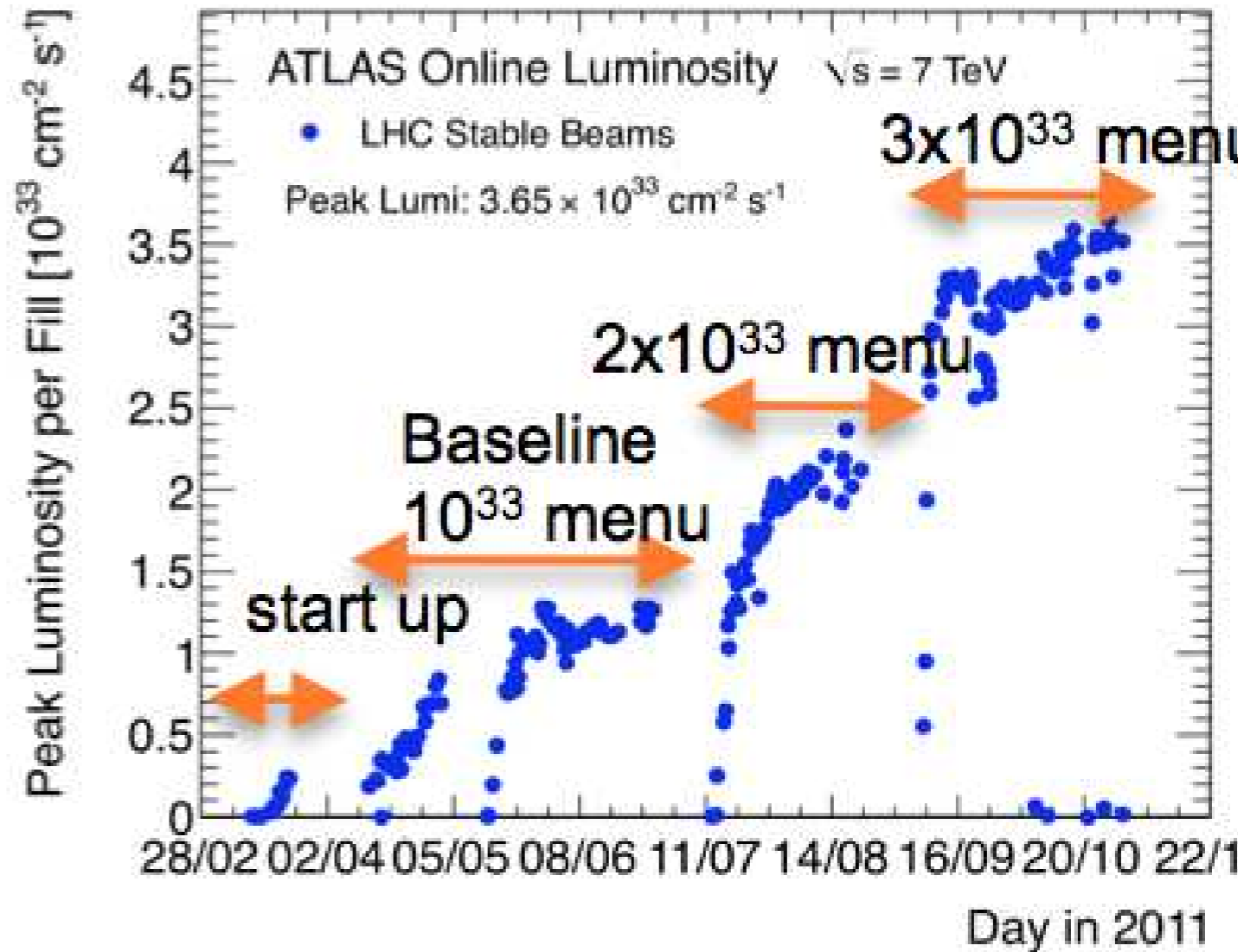


# Changing conditions

Output bandwidth is a finite resource.

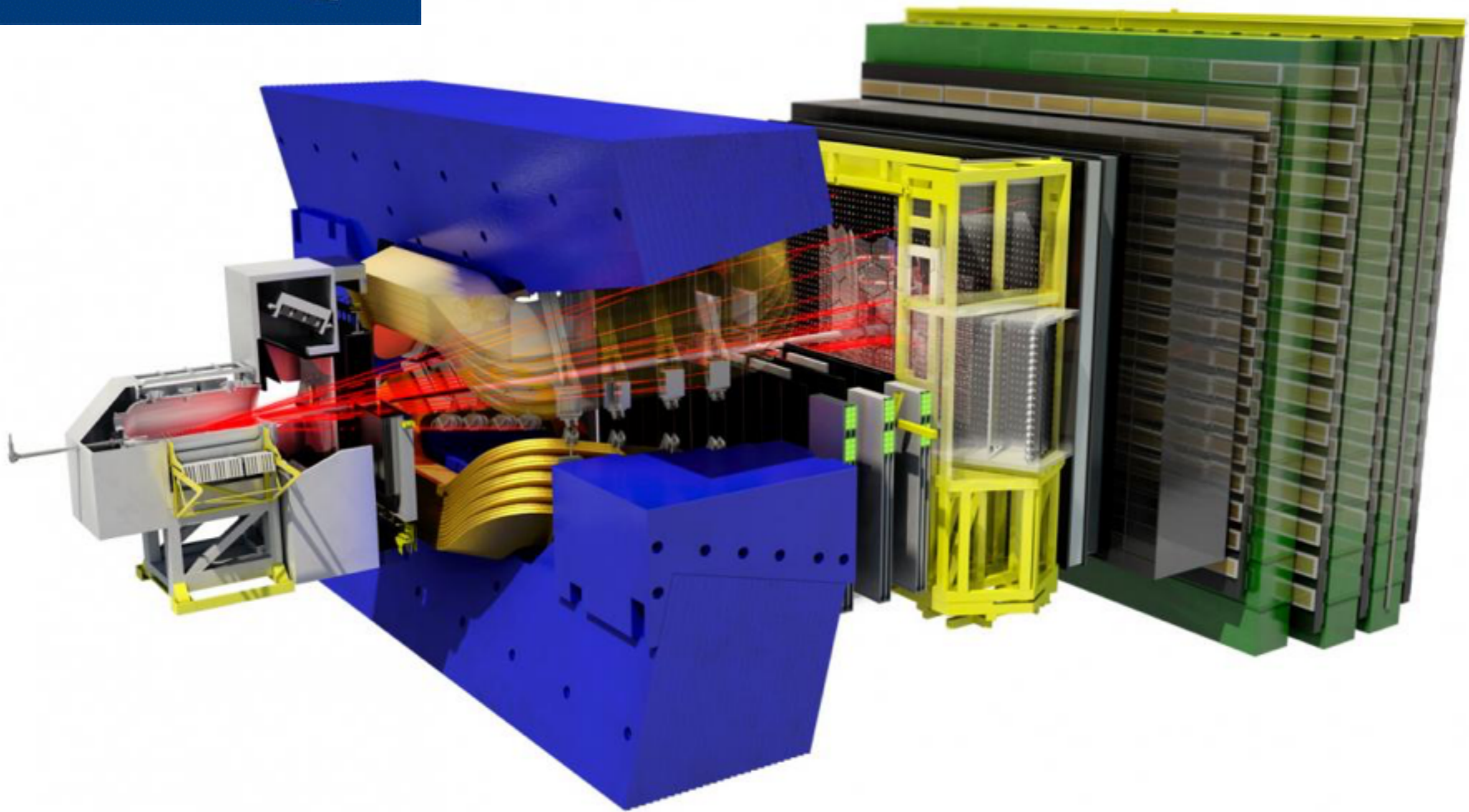
- The machine evolves
- And the physics priorities evolve too.

Requires plenty of interaction with analysis, and careful monitoring



# MLHEP and triggers





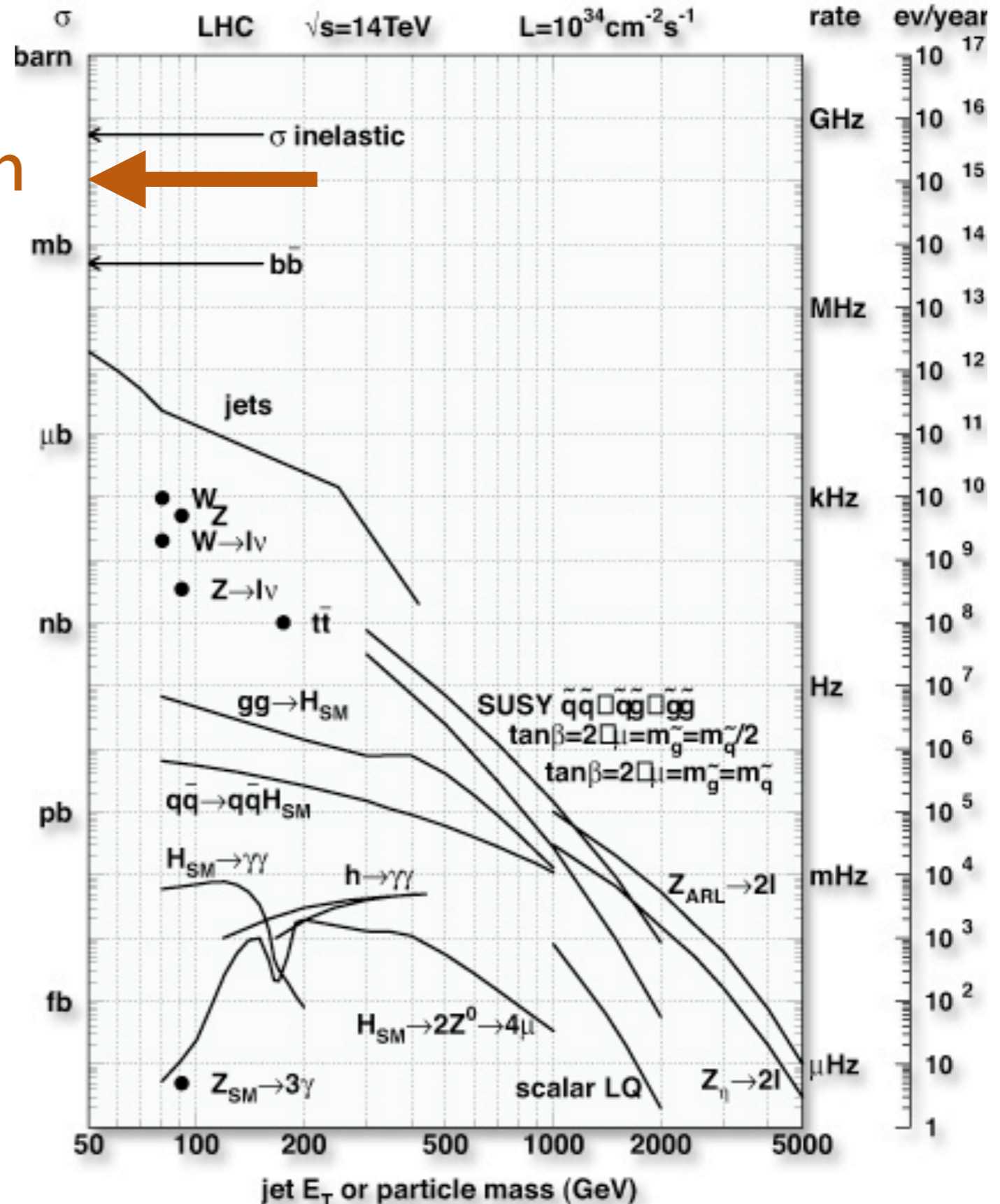
HLT needs to reduce the 1 MHz from L0 to  $\sim 10$  kHz

~1 MHz in LHCb  
(Run-I/II)

Charm

Even if we could inclusively select pure charm, we couldn't take all of it.

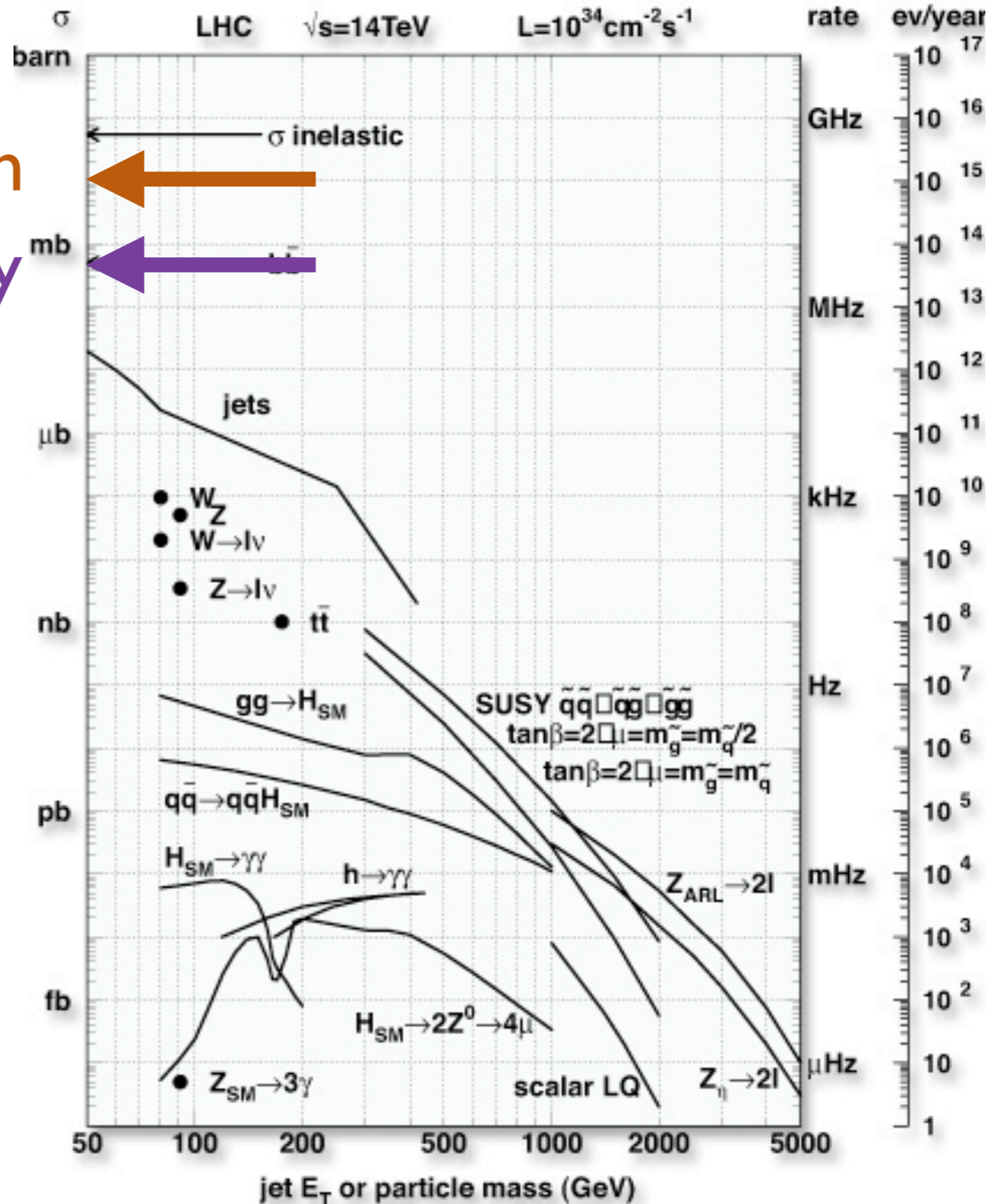
Need exclusive selections of the decay modes that we are interested in.





Still too much signal  
(50-100 kHz\*)

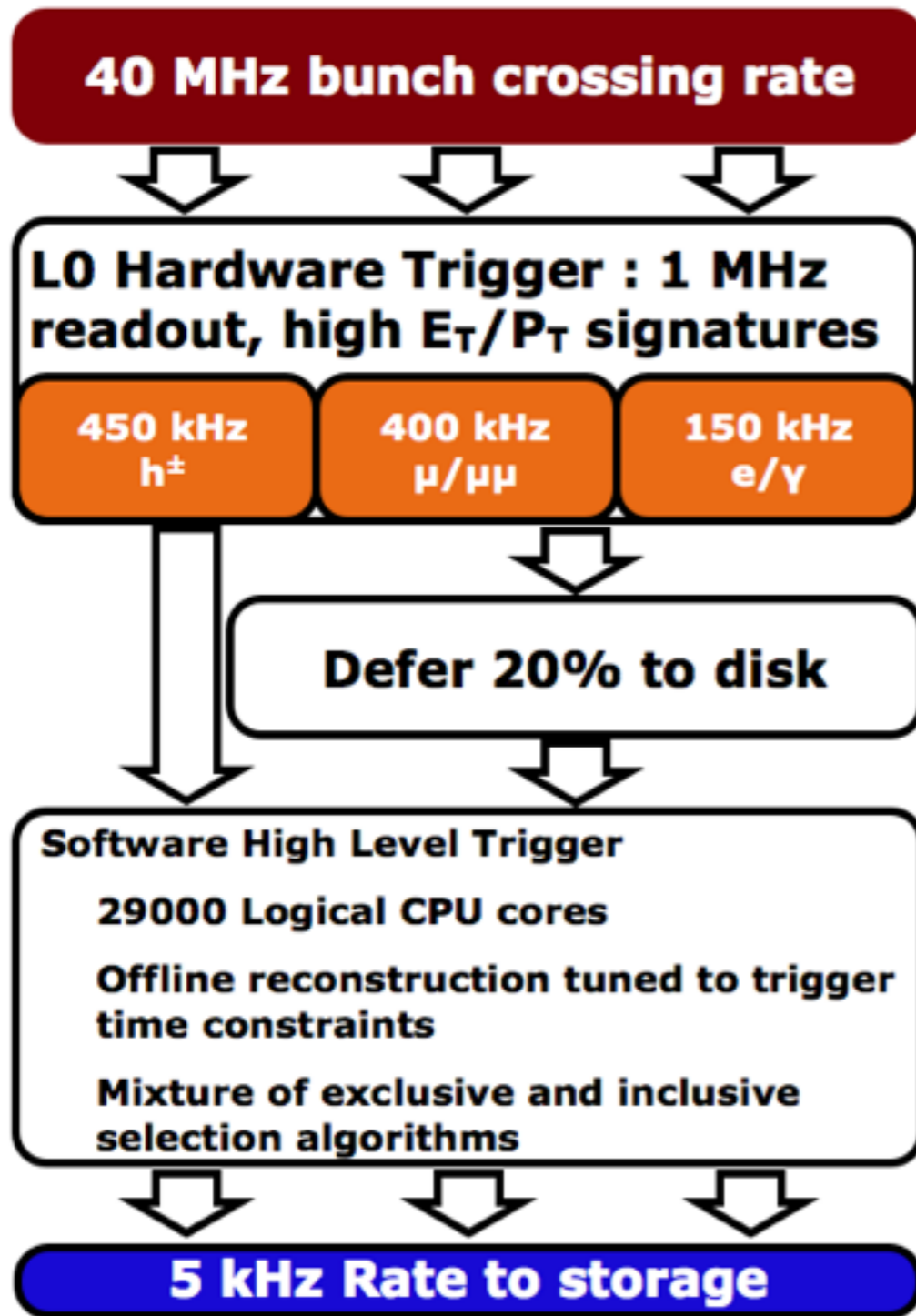
Charm  
Beauty



\*In Run-I/II, roughly speaking



# Run-1

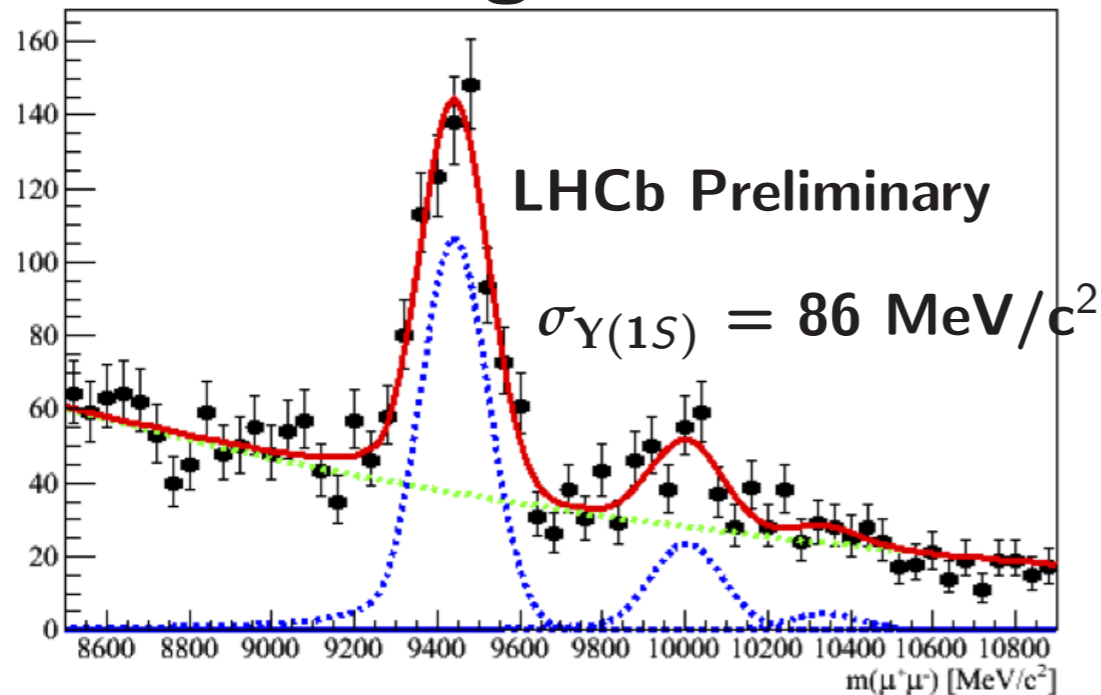




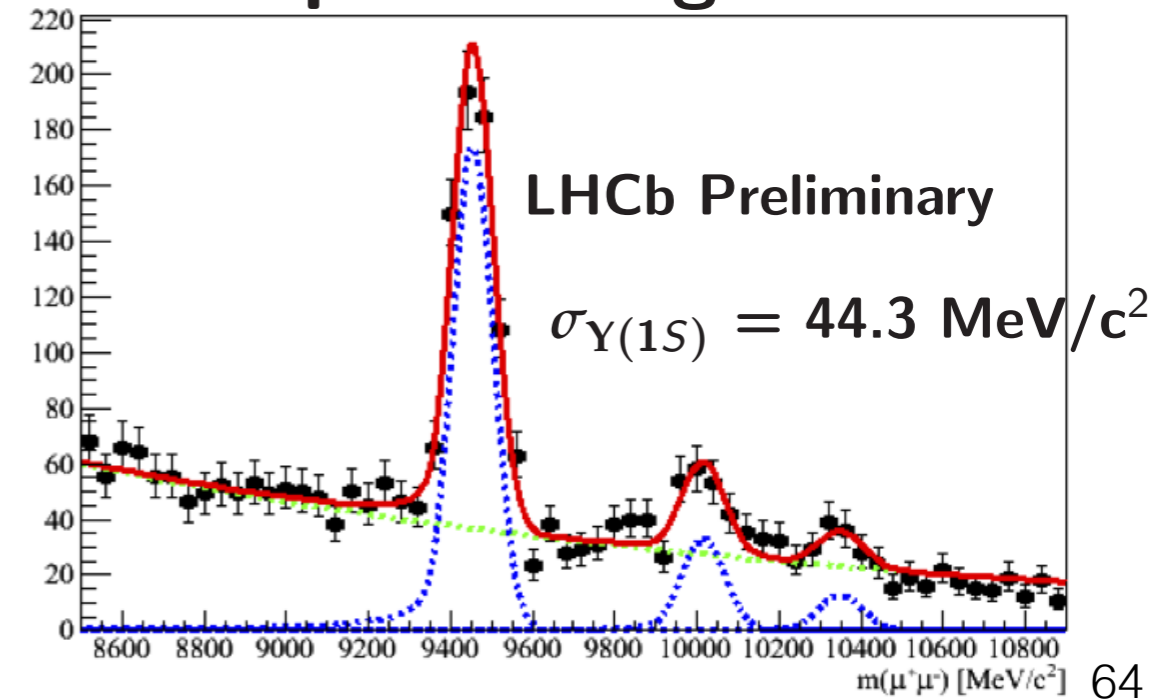
# Enabling real-time calibrations

In **Run-I**, the final alignments could arrive months after the data was taken — to be used in the final offline processing.

## First alignment



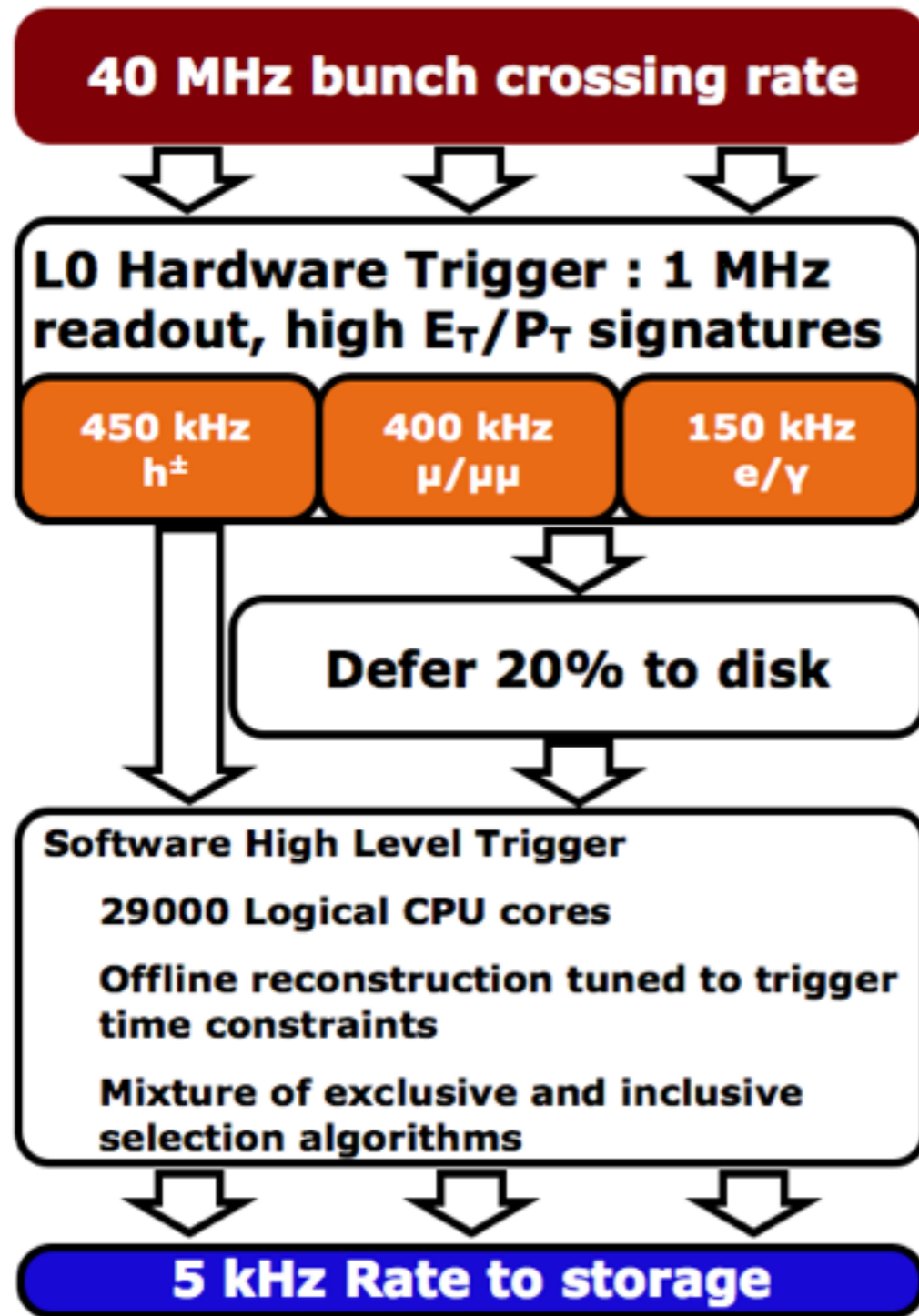
## Improved alignment



**Run-II:** alignment and calibration become part of the online workflow.

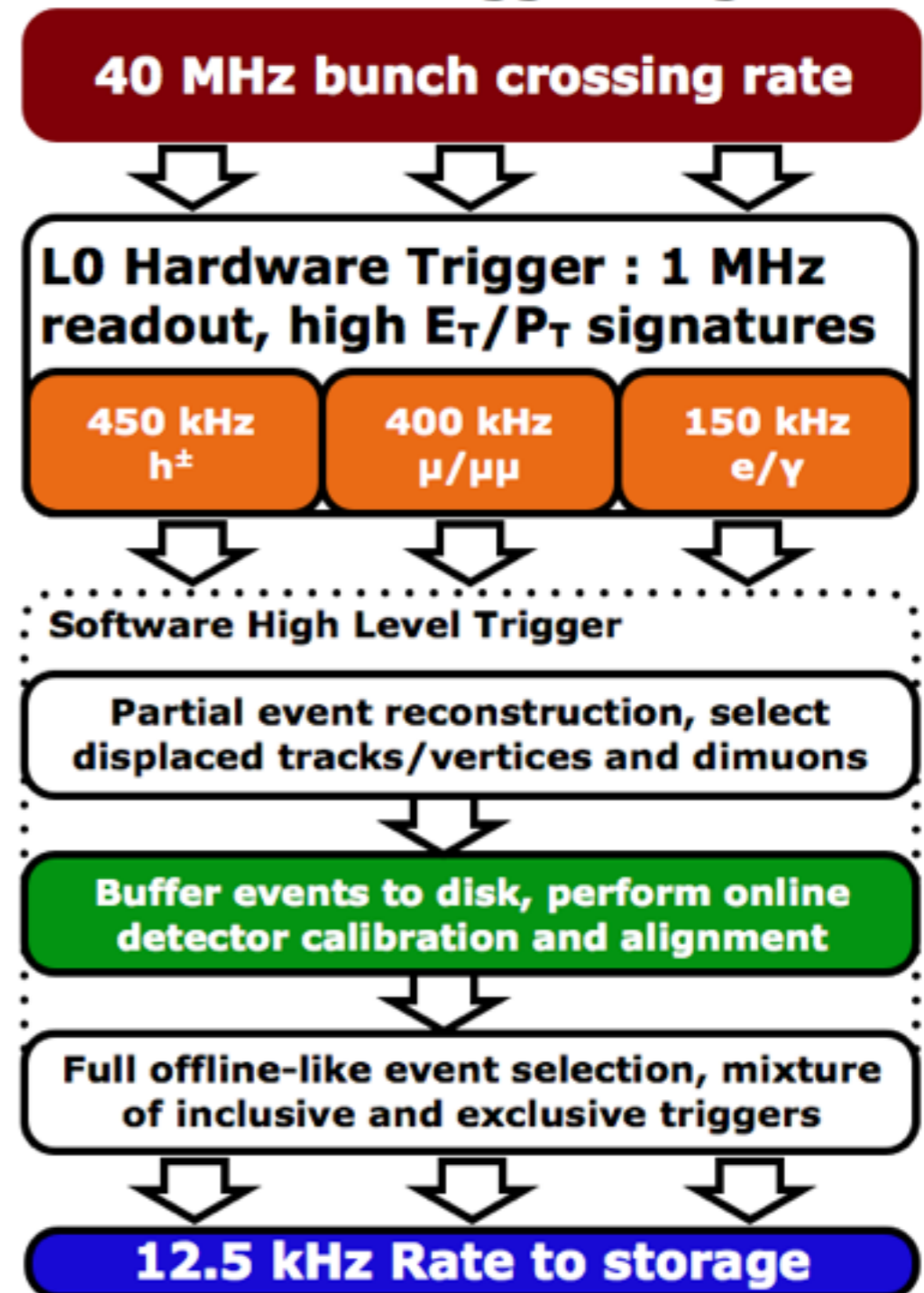
- Better information = better real-time selections
- And, we can do physics analysis on the HLT output!

# Run-I



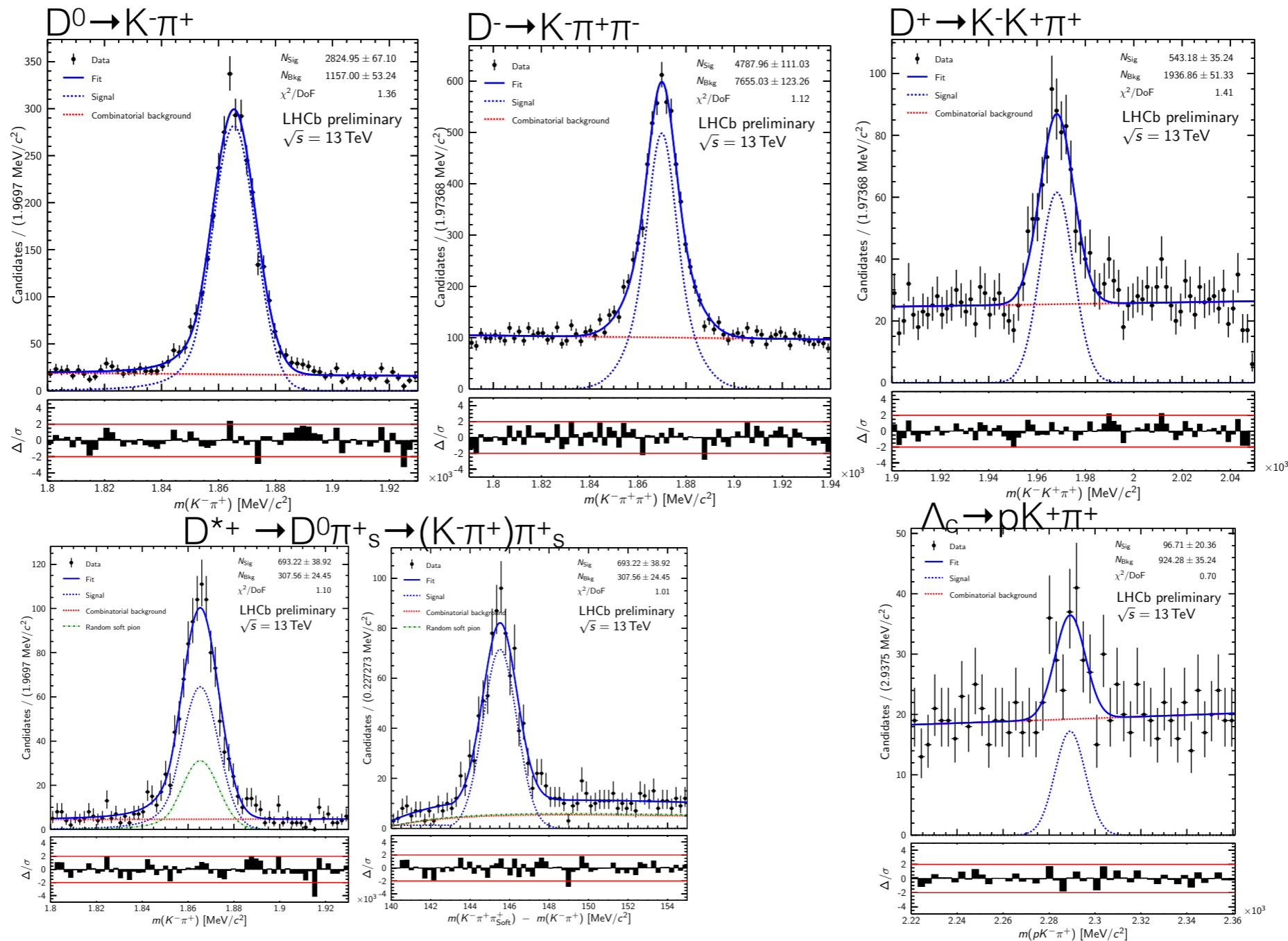
# Run-II

## LHCb 2015 Trigger Diagram



# The LHCb “Turbo” stream

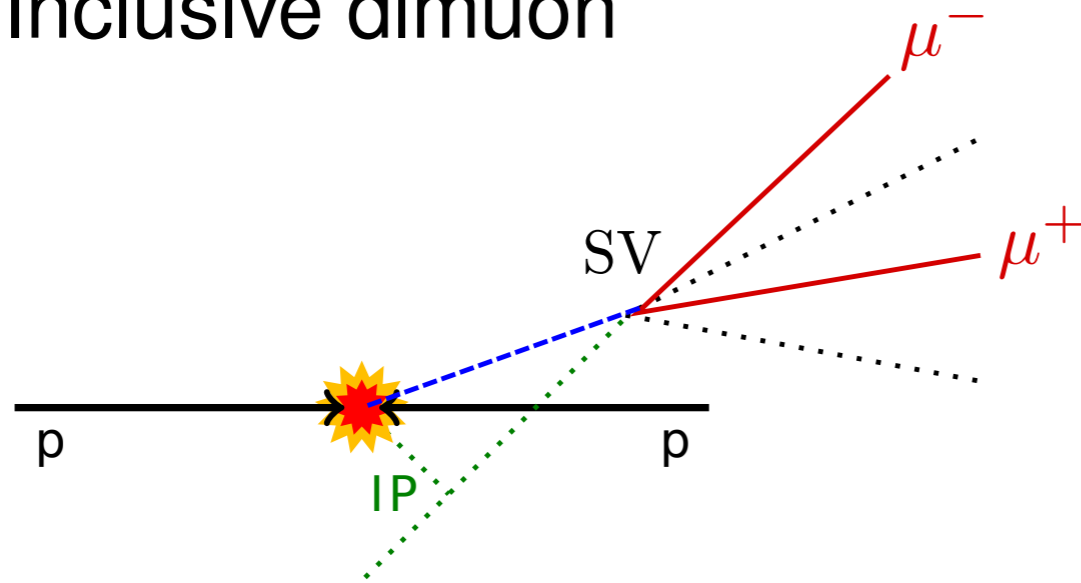
- Bypass offline reconstruction (70 kB/event → 5 kB/event)
- Perfect for the huge rate of charm exclusives.



# Allocating the 5-10 kHz bandwidth?

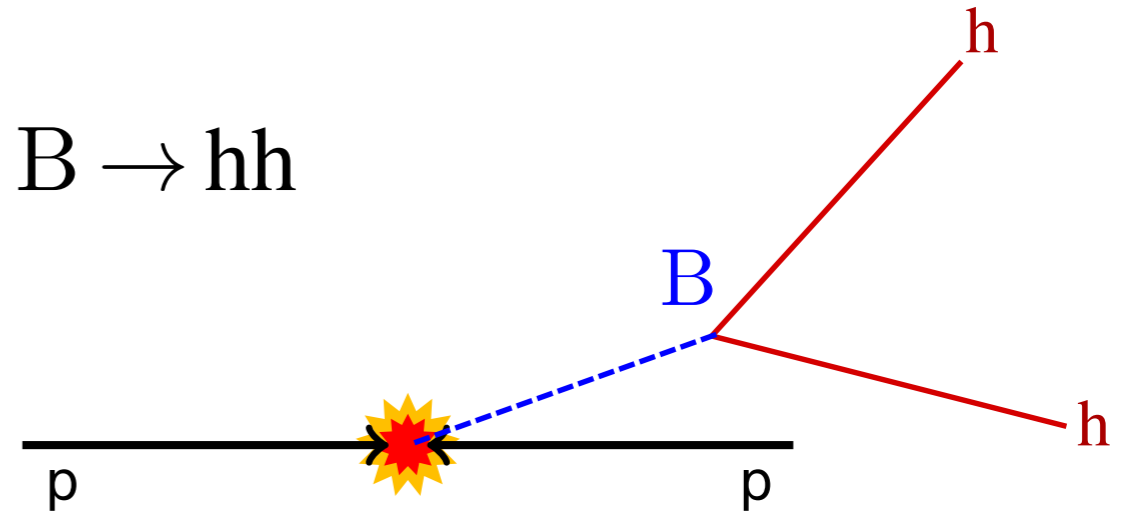
## Inclusives

Inclusive dimuon

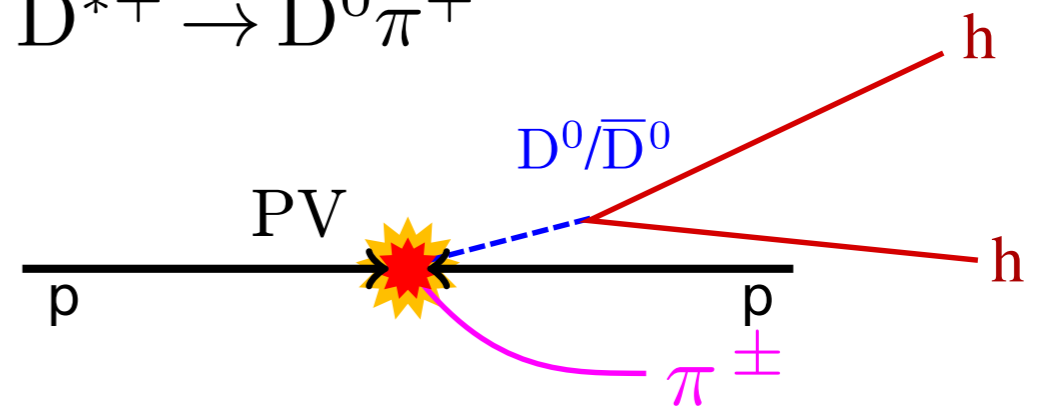


## Exclusives

$B \rightarrow hh$



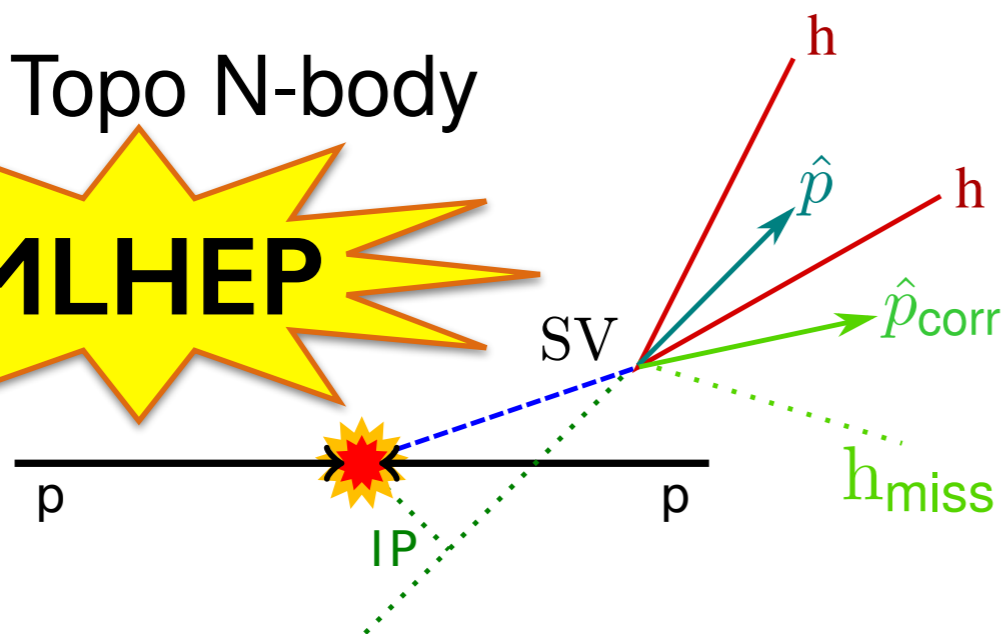
$D^{*+} \rightarrow D^0 \pi^+$



etc.. etc...

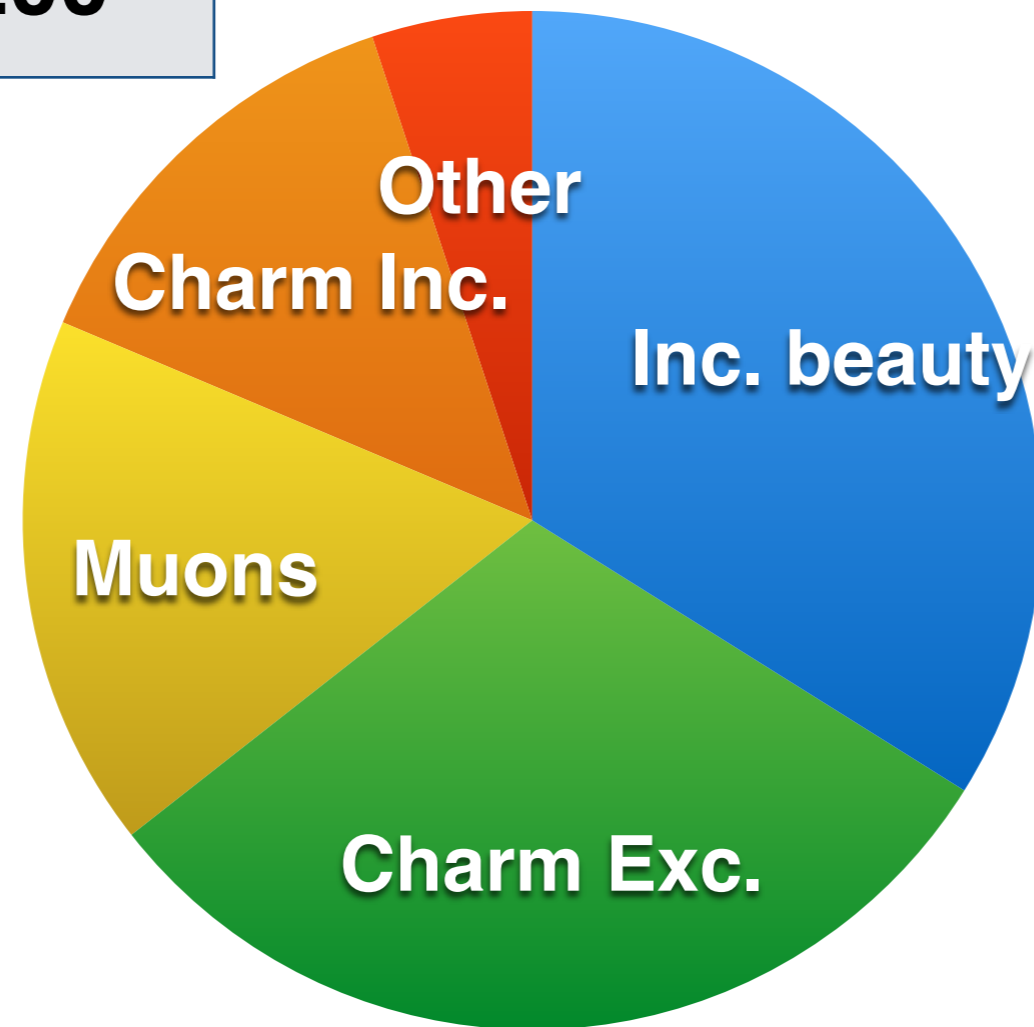
Topo N-body

**MLHEP**



# Bandwidth division in Run-I (2012)

<b>Rate</b>	<b>5 kHz</b>
<b>Lines</b>	<b>~200</b>



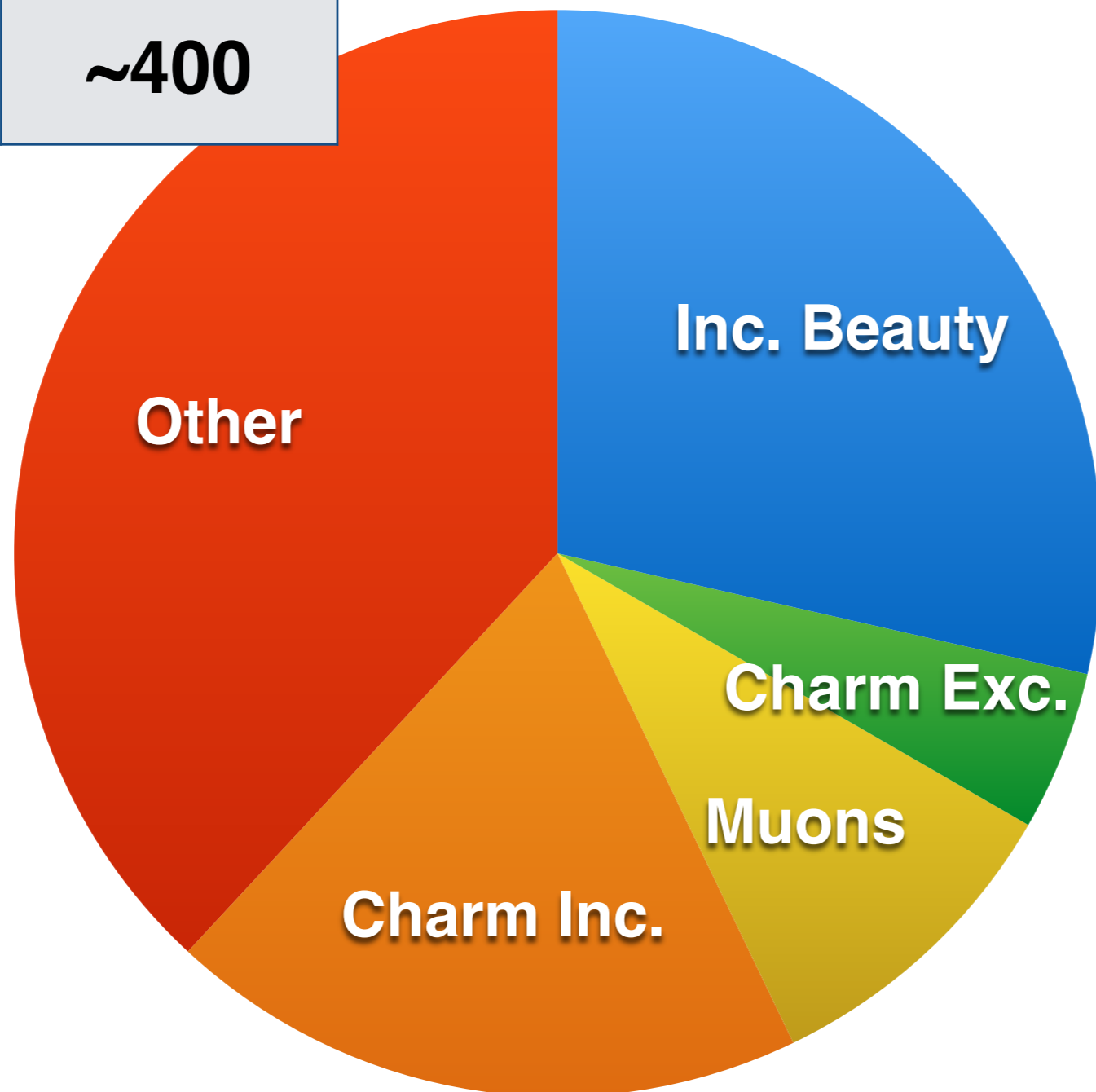


# Bandwidth division in Run-II

<b>Rate</b>	<b>12.5 kHz</b>
<b>Lines</b>	<b>~400</b>

10 kHz  
“Full” stream

+2.5 kHz  
“Turbo” stream







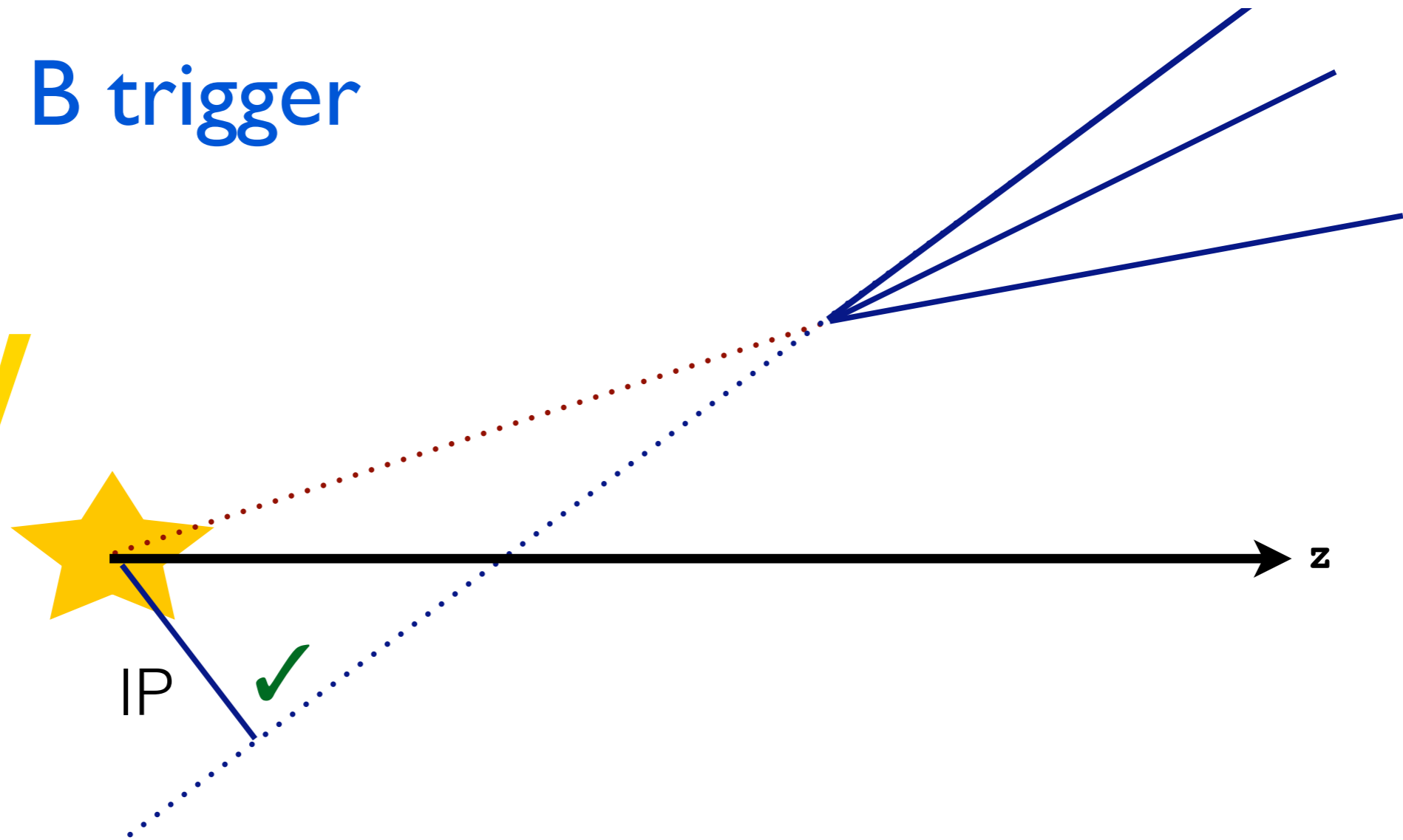
Managing 100s of exclusive lines

(Inclusives preferred)



# Inclusive B trigger

PV



## The goal (Run-I)

*A small set of lines which select any “reasonable”\**

*B hadron decay with  $> 80\%$  efficiency, and with an output rate of 2 kHz.*

*And fast:*

*$\sim 50k$  processes @ 1 MHz:  $\sim 50$  ms to make a decision*

\*All charged daughters reconstructible  
and  $p_T(B) > 2$  GeV and  $\tau(B) > 0.2$  ps.

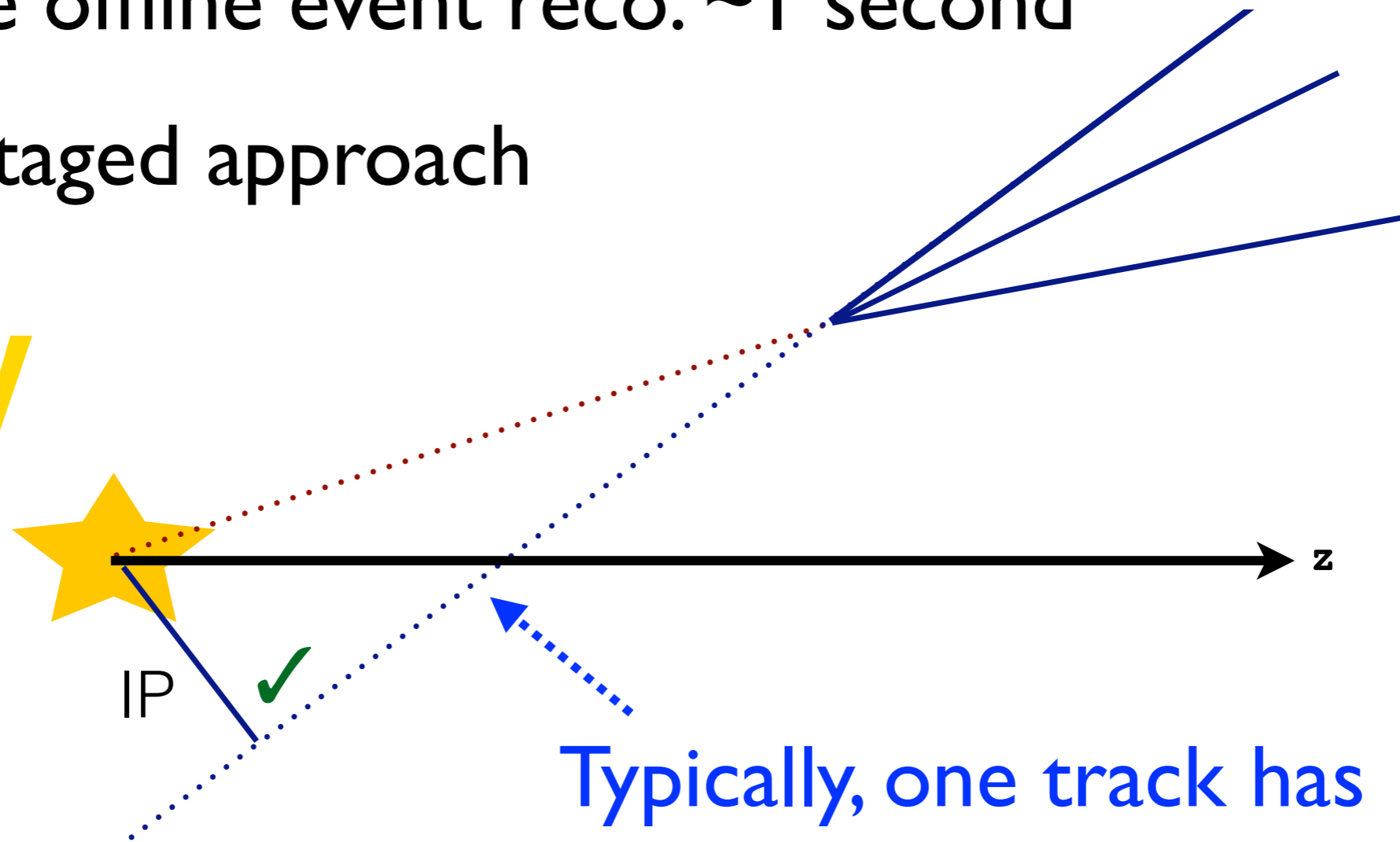
# Staged HLT

CPU budget: 50(25) ms in Run-II(I)

Complete offline event reco: ~1 second

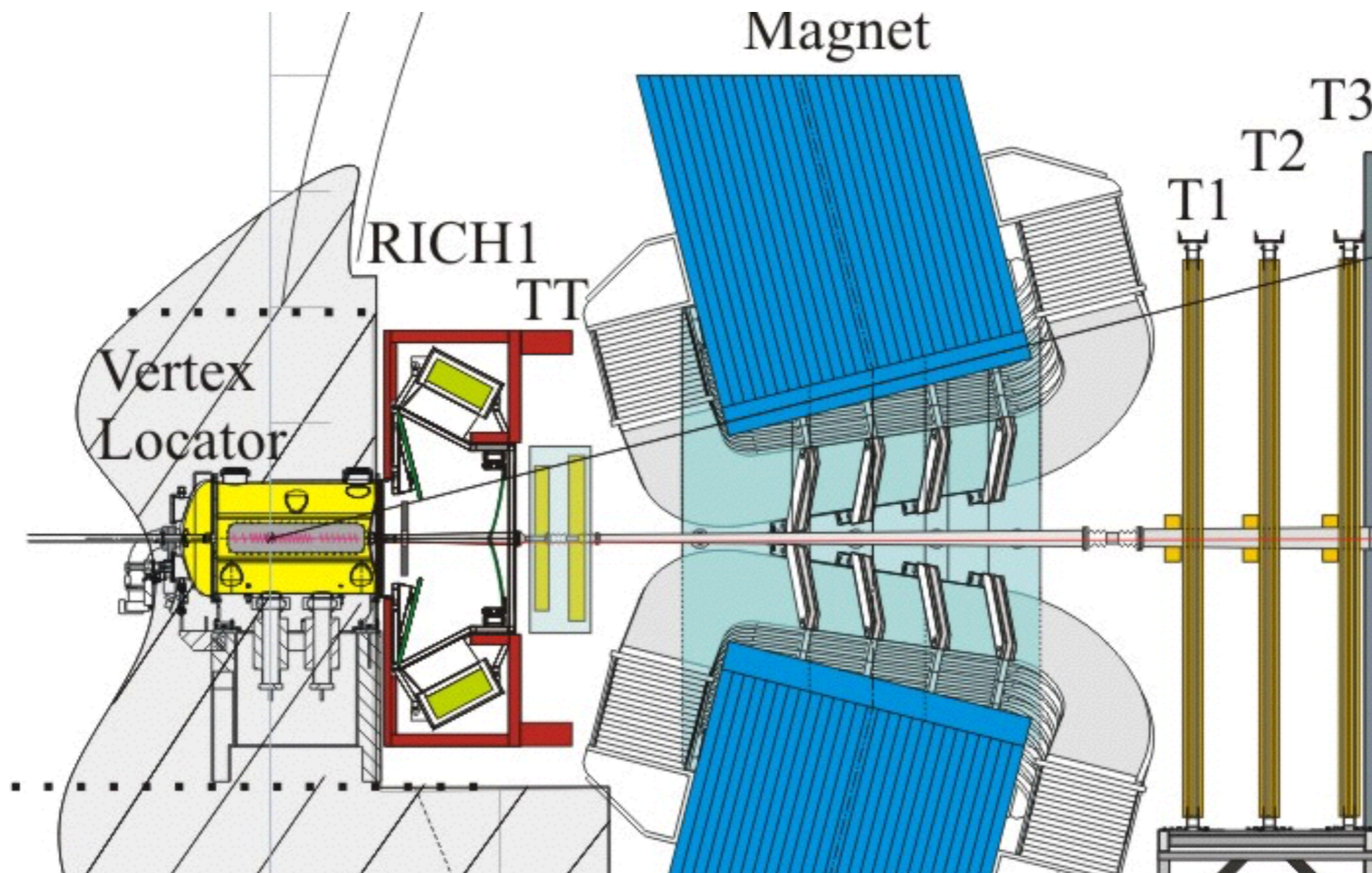
Need a staged approach

PV



Typically, one track has large  $p_T$  and large IP

# Start with tracking



**20-200 ms**

depending primarily on how low you want to go in  $p_T$   
and how well you want to fit the tracks



# Run-1



1 MHz

Tracking  
 $p_T > 1250$  MeV  
Incl. selection

HLT1  
15 ms/ev

## HLT1 inclusive beauty (and charm)

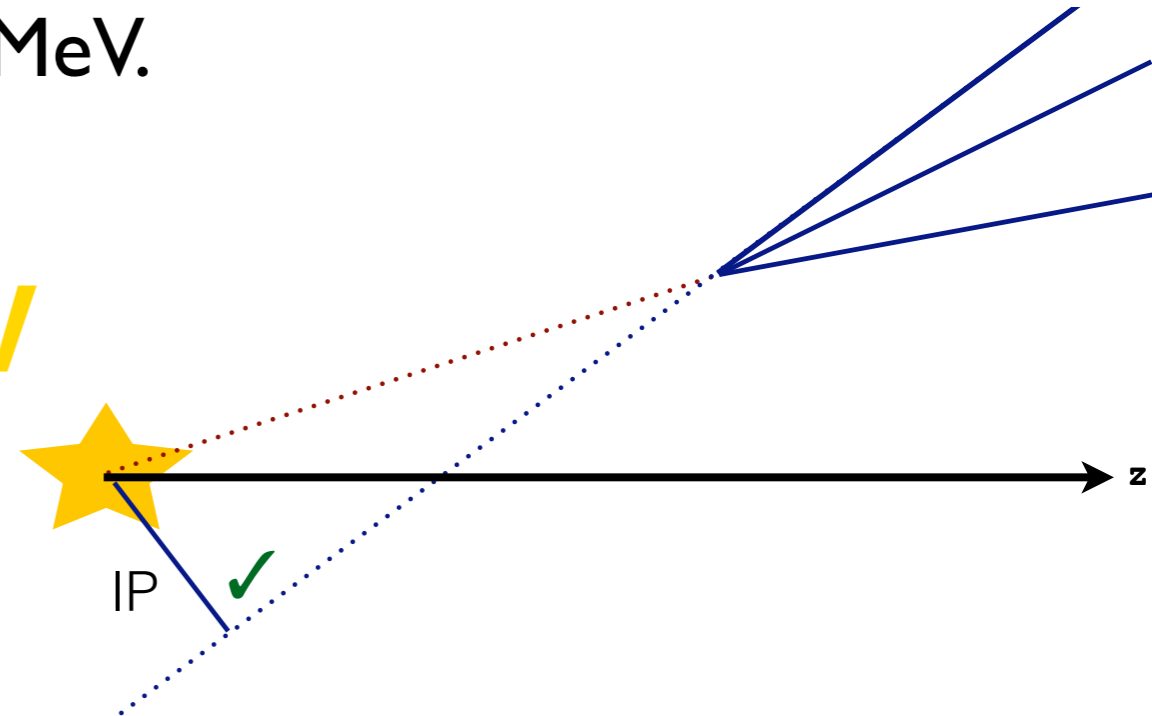
- Single track line with  $p_T > 1700$  MeV and IP  $> 100$   $\mu\text{m}$
- Little to be gained from looking for other tracks with  $p_T > 1250$  MeV.

70 kHz

Tracking  
 $p_T > 300$  MeV  
Incl. + Excl.

HLT2  
200 ms/ev

PV



5 kHz



# Run-II

1 MHz

Tracking  
 $p_T > 500$  MeV  
Incl. + excl.



HLT1  
35 ms/ev

HLT1 inclusive beauty (and charm)

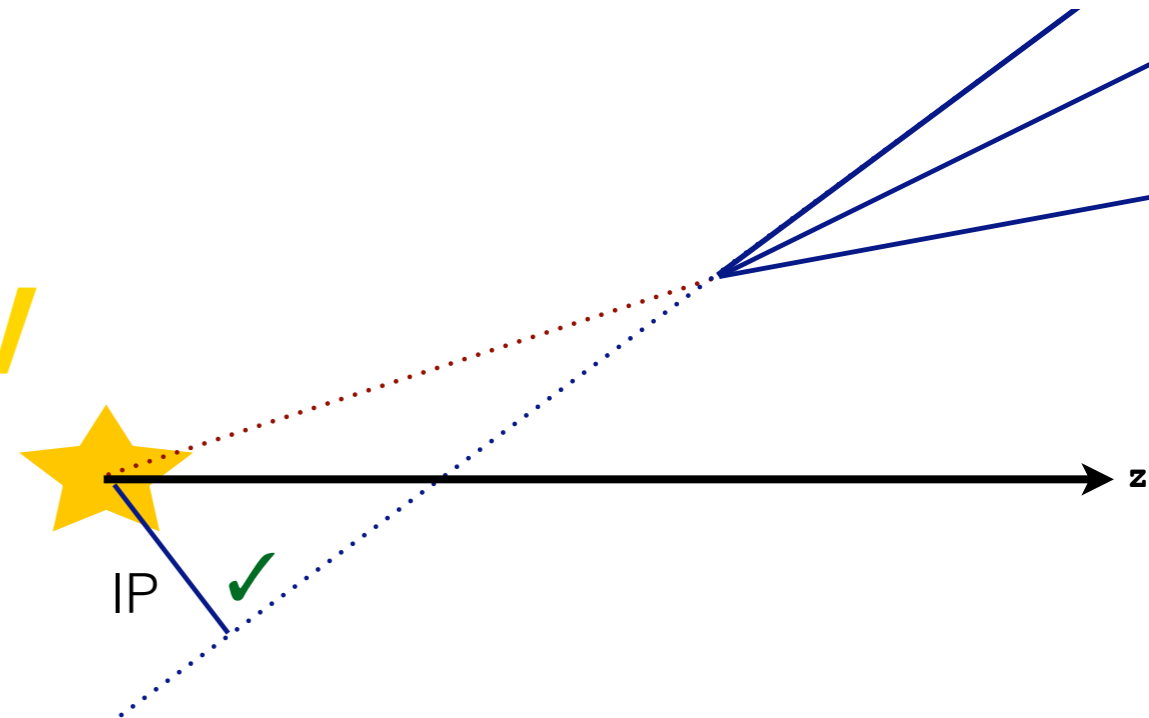
- At **500** MeV, now we start to see more of the beauty daughters
- Start to gain with 2-track triggers

150 kHz

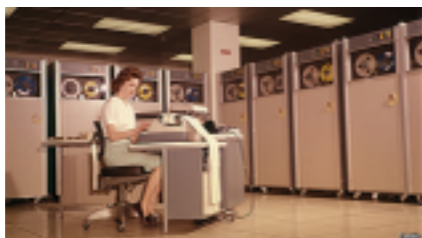
Tracking  
 $p_T > 70$  MeV  
Incl. + Excl.

HLT2  
500 ms/ev

PV



12.5 kHz



# Run-II inclusive 1- and 2-track triggers



## 1 Track:

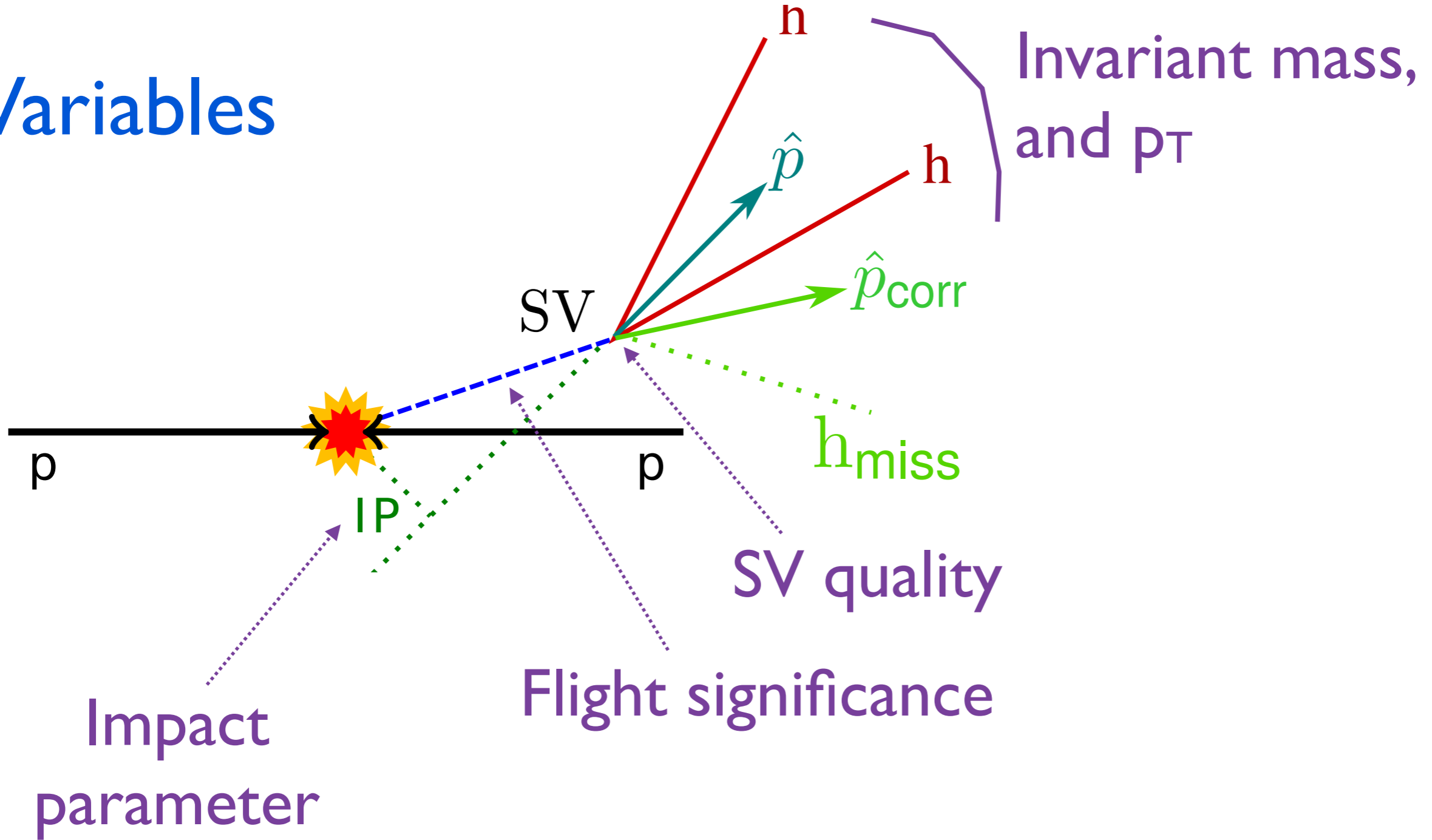
- ❖ track  $p_t > 500$  MeV; IP  $\chi^2 > 4$ ;  
track  $\chi^2/\text{ndof} < 3$ ;
- ❖ BDT using  $p_t$ , IP  $\chi^2$ ;
- ❖ Adding cut-based TrackMuon requiring  $p_t > 600$  MeV, IP  $\chi^2 > 6$  adds about 20 kHz to the total.

## 2-body SV:

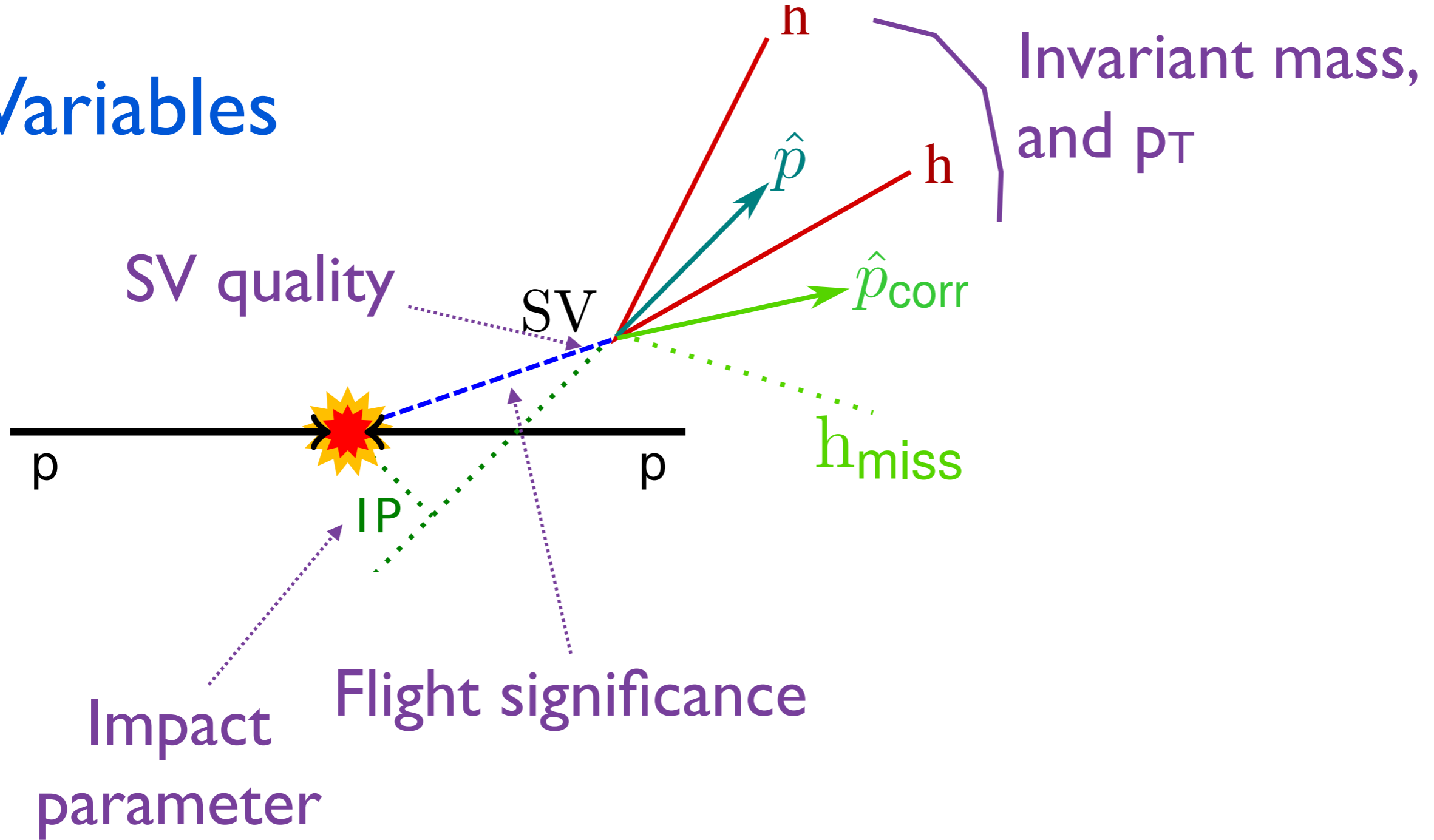
- ❖ tracks with  $p_t > 500$  MeV; IP  $\chi^2 > 4$ , track  $\chi^2/\text{ndof} < 2.5$ ;
- ❖ SV  $p_t > 2$  GeV; vertex  $\chi^2 < 10$ ;  
 $1 < \text{MCOR} < 10$  GeV;  $2 < \eta < 5$   
(PV to SV).
- ❖ BDT using  $p_t$ , MCOR, vertex  $\chi^2$ , min track  $p_t$ , FD  $\chi^2$ ,  
 $N(\text{tracks with IP } \chi^2 < 16)$ .

The HLT2 Topo lines are very similar, except that we have 3- and 4-body versions too.

# Variables

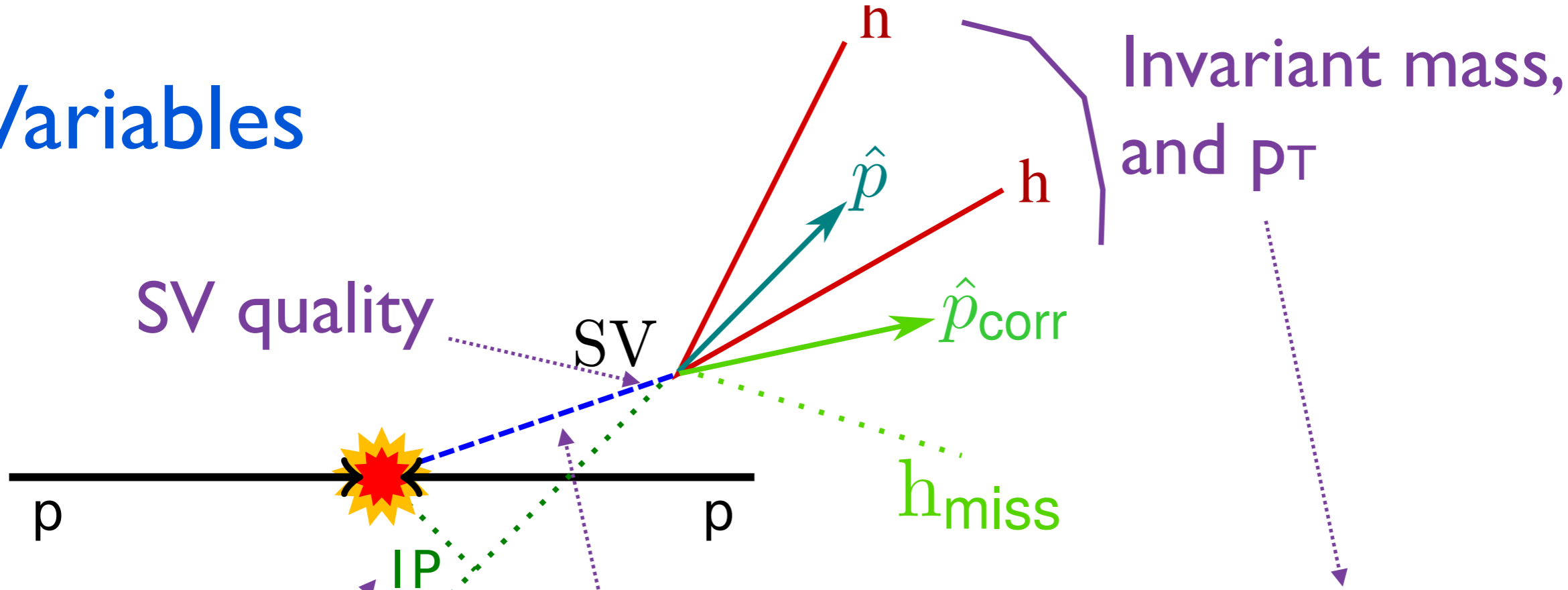


# Variables



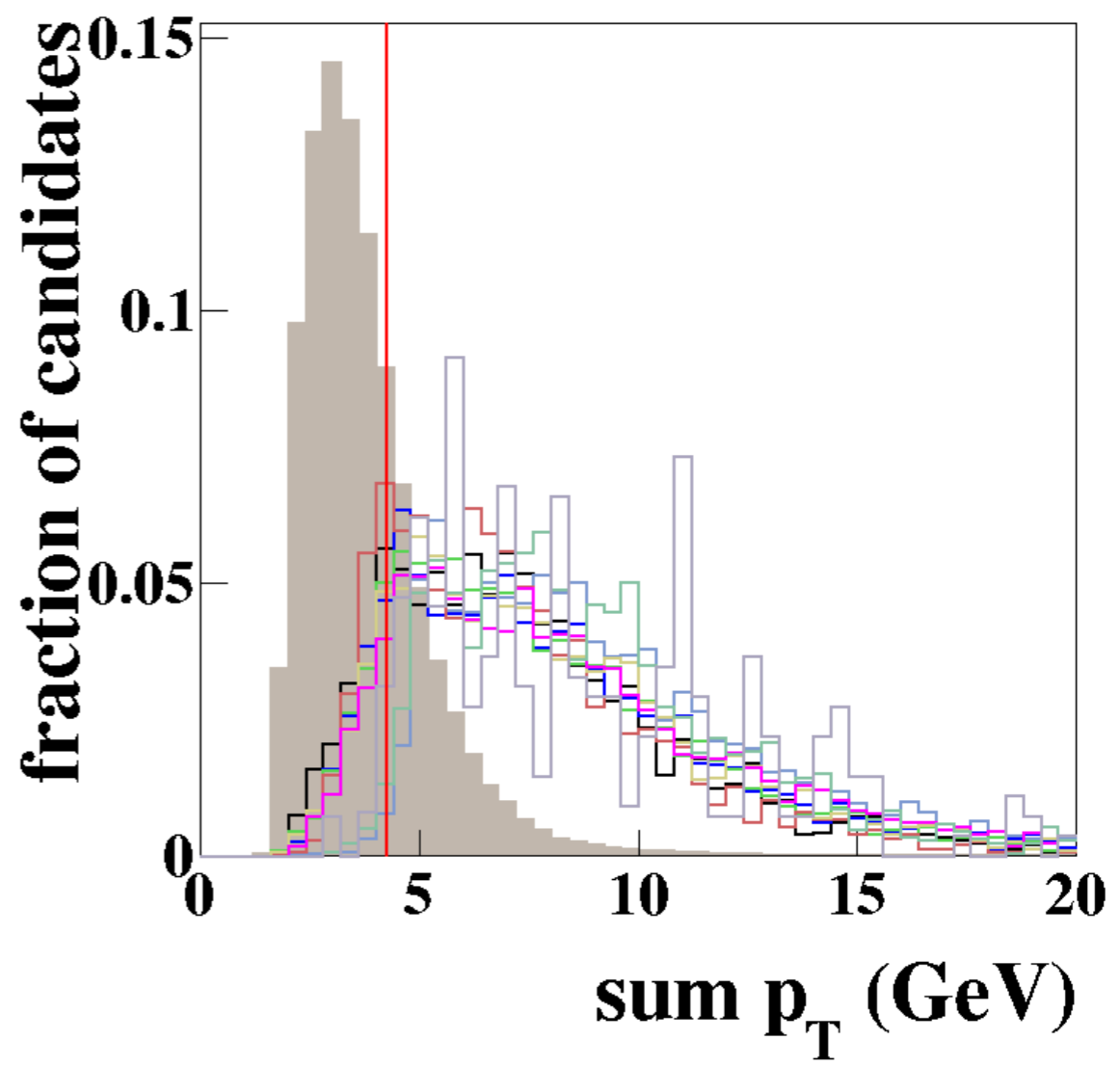


# Variables

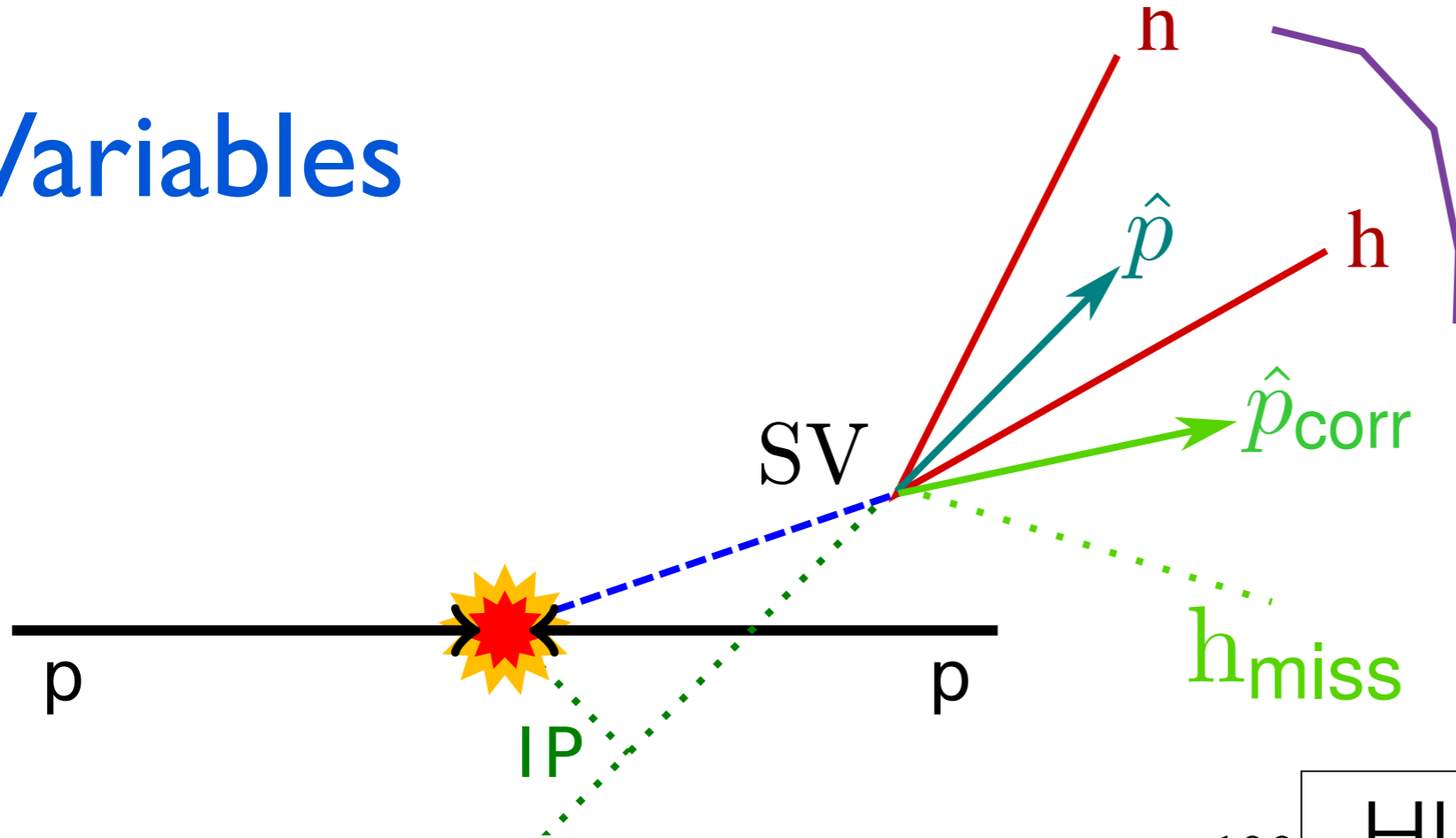


Impact parameter

Flight significance

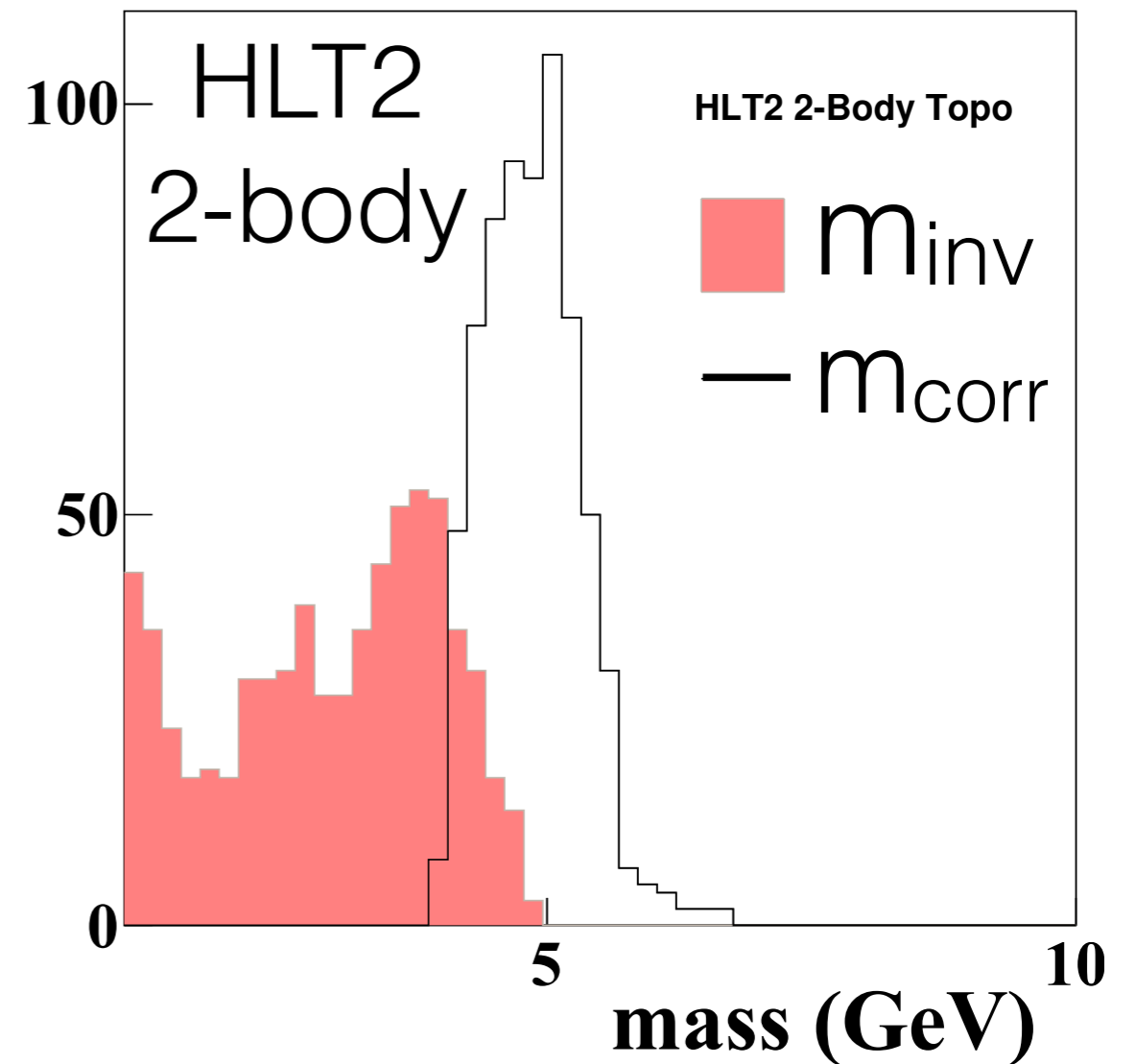


# Variables



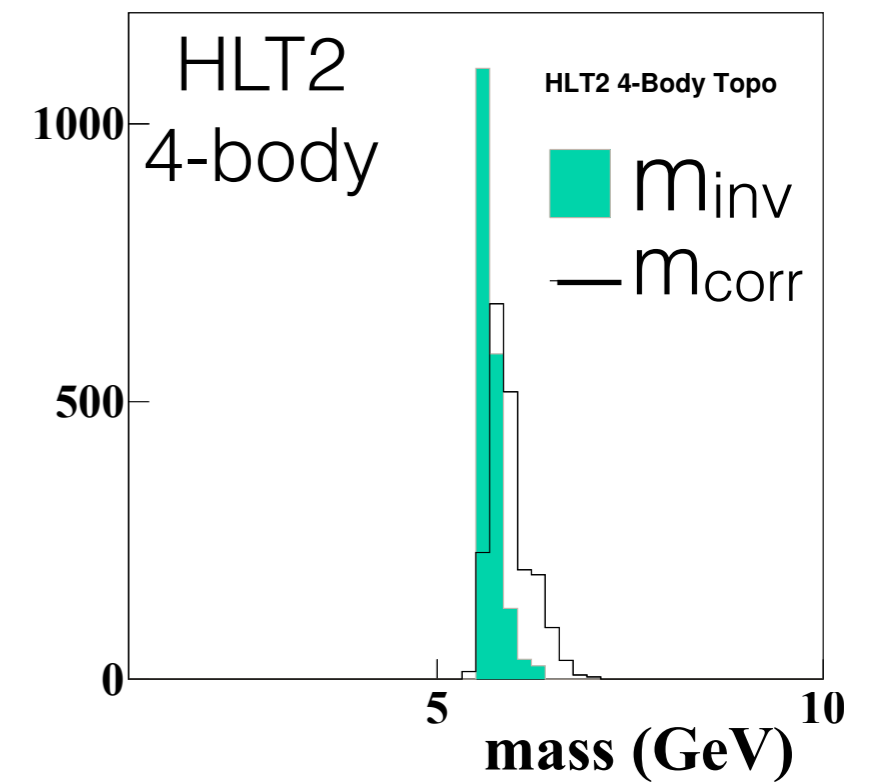
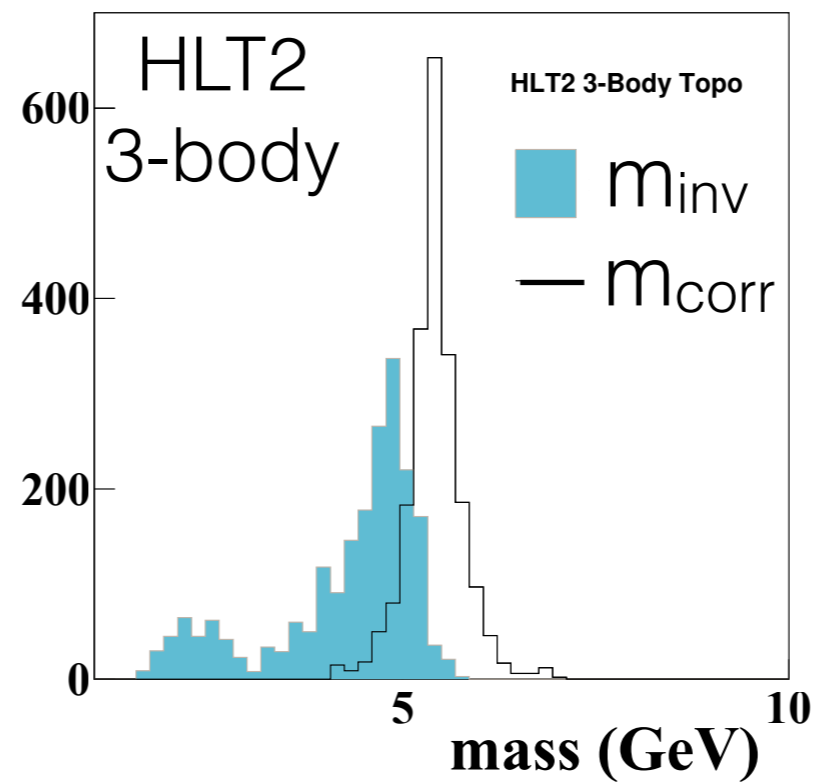
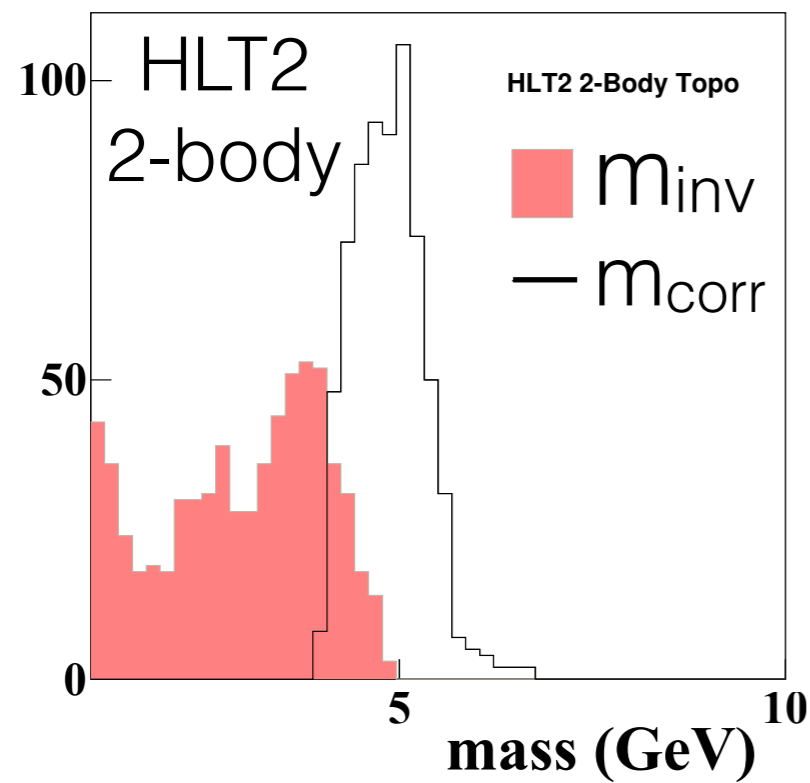
## The “corrected mass”

$$m_{\text{corr}} \equiv \sqrt{m_{\text{inv}}^2 + |P_{T\text{miss}}|^2 + |P_{T\text{miss}}|}$$

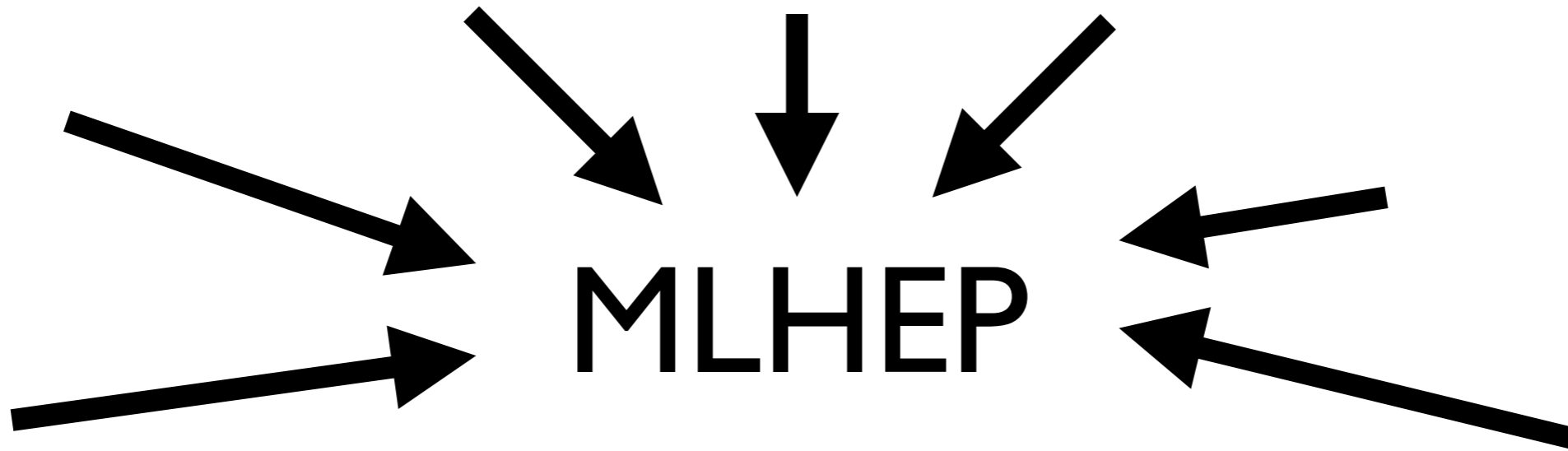
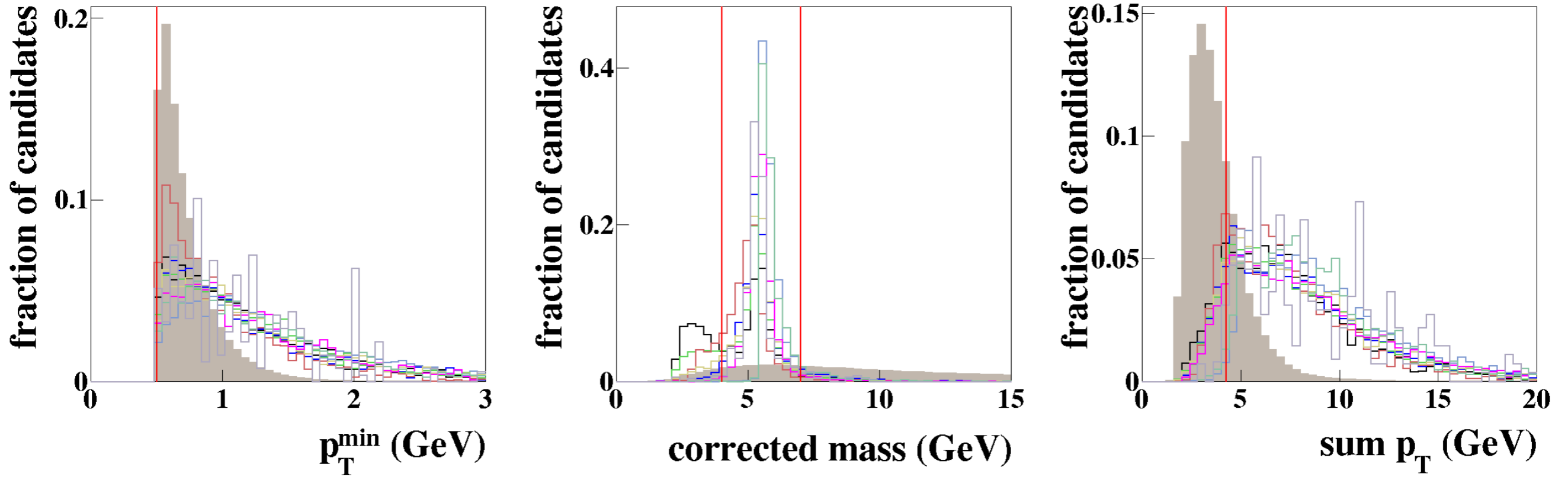


# 2-, 3-, and 4-body versions

Same 3-body B decay as seen by the 2-, 3- and 4-body Topos



3-body candidates: BKGD = shaded grey,  $B \rightarrow$  anything = various colors



# Discussing the selection cuts, pre MVA



Cut harder on  
 $p$

The  $m$   
should be 7.9 not  
8.0 GeV!



# The Topo with MLHEP



Signal  
efficiency

Single BDT  
threshold



Output rate

# Alternative view: MVA in my HLT!



MVA first introduced into the inclusive B triggers in 2011, when we were forced to reduce the background rate by a factor of 3 given the fast ramp up of the luminosity compared to 2010. Some in LHCb had concerns though.

# Concerns

- **Efficiency:** needs to be stable and well understood.
- **Generality:** Want to learn generic b decays not just the limited set that we can realistically train on
- **Speed:** Decision in a few ms.

One of the most commonly used classifiers in HEP is the BDT

# Concerns

- 👍 ● **Efficiency:** needs to be stable and well understood.
- 👍 ● **Generality:** Want to learn generic b decays not just the limited set that we can realistically train on
- 👍 ● **Speed:** Decision in a few ms.

One of the most commonly used classifiers in HEP is the BDT

**Solution:** Control what the tree can learn by discretising the variables as we choose.







# Bonzai Boosted Decision Tree

**Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree.**

---

**V. V. Gligorov<sup>1</sup> and M. Williams<sup>2,a</sup>**

<sup>1</sup>*Organisation Européenne pour la Recherche Nucléaire, Geneva, Switzerland*

<sup>2</sup>*Imperial College London, London SW7 2AZ, UK*

<sup>a</sup>*Current address: Massachusetts Institute of Technology, Cambridge, MA, USA*

**ABSTRACT:** High-level triggering is a vital component in many modern particle physics experiments. This paper describes a modification to the standard boosted decision tree (BDT) classifier, the so-called *bonsai* BDT, that has the following important properties: it is more efficient than traditional cut-based approaches; it is robust against detector instabilities, and it is very fast. Thus, it is fit-for-purpose for the online running conditions faced by any large-scale data acquisition system.

# Discretisation of the variables

Table 4: Allowed split points in the bonsai boosted decision tree.

variable	cuts(2,3,4-body)	allowed splits
$\sum  p_T $	$> 3,4,4$ GeV	3.5,4,4.5,5,6,7,8,9,10,15,20 (GeV)
mass	$< 7$ GeV	2.5,4.75 (GeV)
DOCA	$< 0.2$ mm	0.05,0.1,0.15 (mm)
$IP\chi^2$		20
corrected mass		2,3,4,5,6,7,8,9,10,15 (GeV)
$p_T^{\min}$	$> 0.5$ GeV	0.6,0.7,0.8,0.9,1,1.25,1.5,1.75,2,2.5,3,4,5,10 (GeV)
$FD\chi^2$	$> 100$	2,3,4,5,6,7,8,9,10,25,50,100 $\times 100$

# Training (run-1)

parent	daughters
$B^\pm$	$K\pi\pi, D_{[K\pi]}\pi, D_{[hhhh]}K, D_{[K_S\pi\pi]}K, D_{[K\pi\pi]}K\pi$
$B^0$	$K_{[K\pi]}^*\mu\mu, K_{[K\pi]}^*ee, D_{[K\pi\pi]}\pi, K\pi, D_{[K\pi]}K\pi, D_{[D(K\pi)\pi]}^*\mu\nu, D_{[K\pi\pi]}K\pi\pi$
$B_s$	$D_{s[KK\pi]}\pi, D_{s[KK\pi]}K\pi\pi, K_{[K\pi]}^*K_{[K\pi]}^*$
$\Lambda_b$	$\Lambda_{c[pK\pi]}\pi, \Lambda_{c[pK\pi]}K\pi\pi$

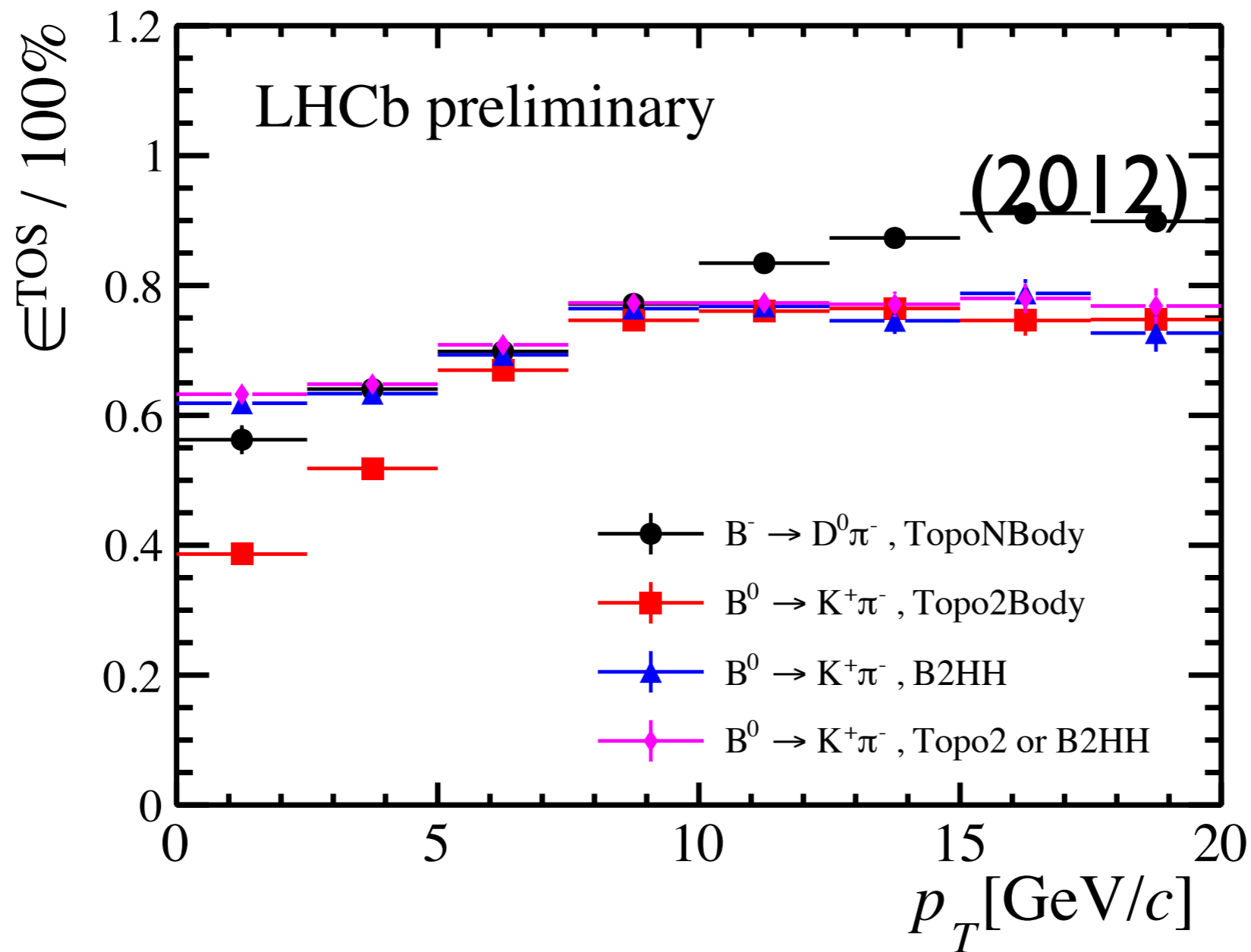
# Performance

Stage	Rate	Beauty fraction
L0	1 MHz	1%
Hlt1	70 kHz	9%
Hlt2Topo	2 kHz	100%

3 lines (2-,3-,4- body) which give 2 kHz of ~100% beauty  
(plus looser muon and electron versions)

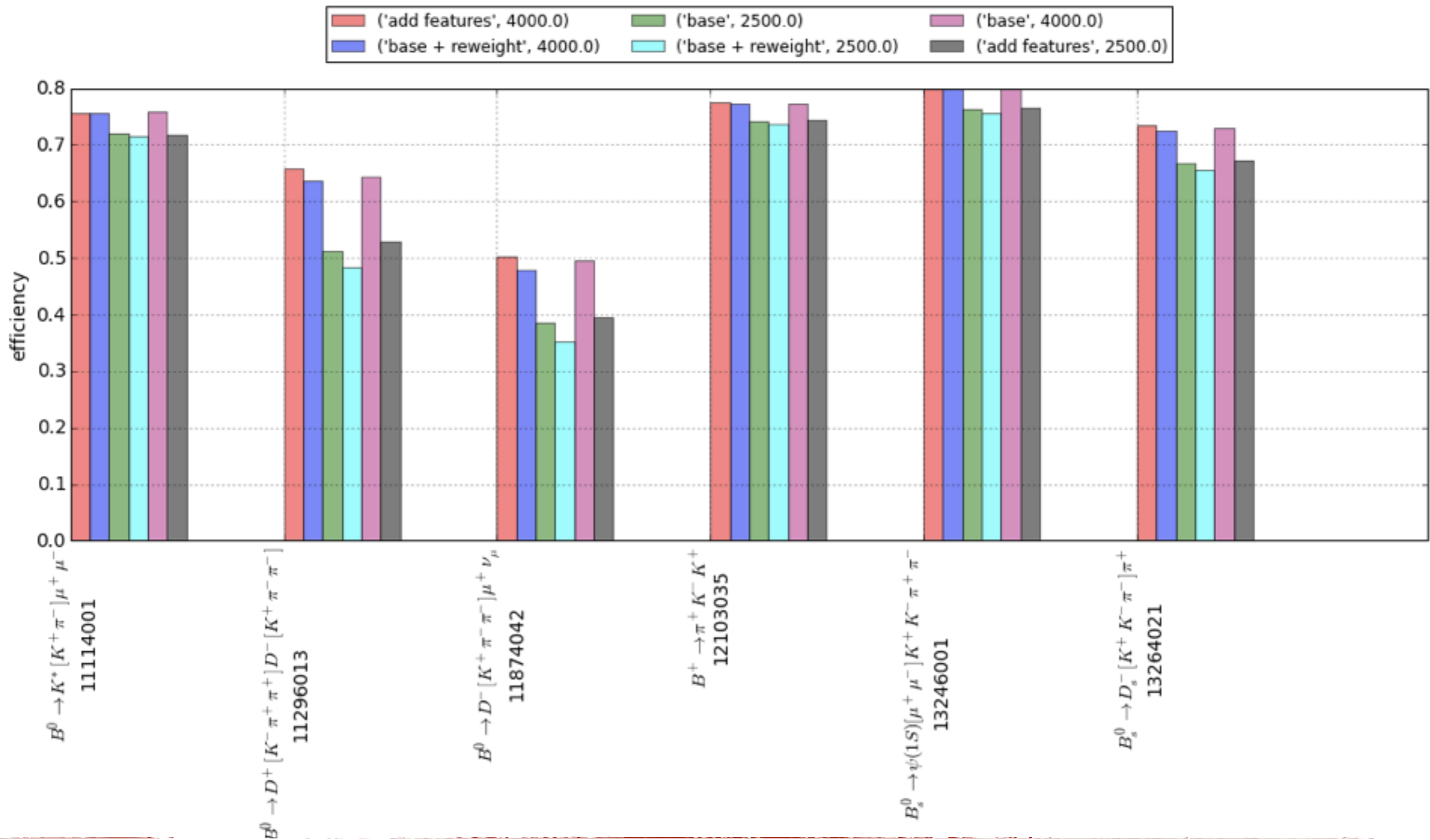


# Performance



60-80% efficient on various fully hadronic B decays  
output rate: 2 kHz

# Classifier optimisation for Run-II



# Performance improvements for Run-II

Denominator is now just “reconstructible” and  $p_T > 2 \text{ GeV}$ ,  $\tau > 0.2 \text{ ps}$

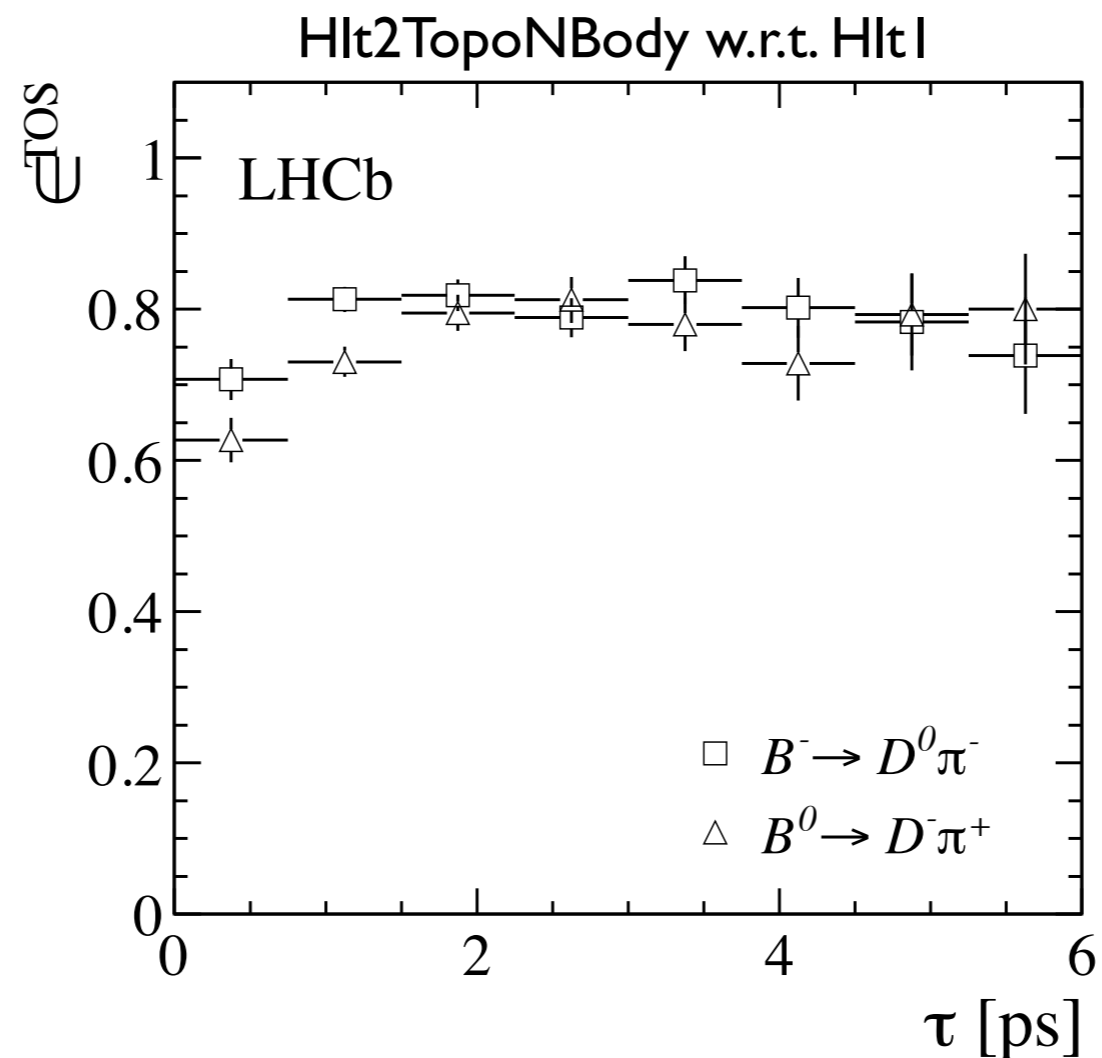
## LHCb unofficial

mode	HLT1	HLT2/HLT1		
		Run 1 (“as was”)	New @ 2.5 kHz	New @ 4 kHz
$B^0 \rightarrow K^*[K^+\pi^-]\mu^+\mu^-$	94	47	77	81
$B^+ \rightarrow \pi^+K^-K^+$	95	49	78	81
$B_s^0 \rightarrow D_s^-[K^+K^-\pi^-]\mu^+\nu_\mu$	95	36	41	53
$B_s^0 \rightarrow \psi(1S)[\mu^+\mu^-]K^+K^-\pi^+\pi^-$	97	48	78	82
$B_s^0 \rightarrow D_s^-[K^+K^-\pi^-]\pi^+$	96	48	70	73
$B^0 \rightarrow D^+[K^-\pi^+\pi^+]D^-[K^+\pi^-\pi^-]$	98	35	49	65

# If we want to be picky though....

Want to select b decays inclusively without biasing the kinematic and decay time distributions.

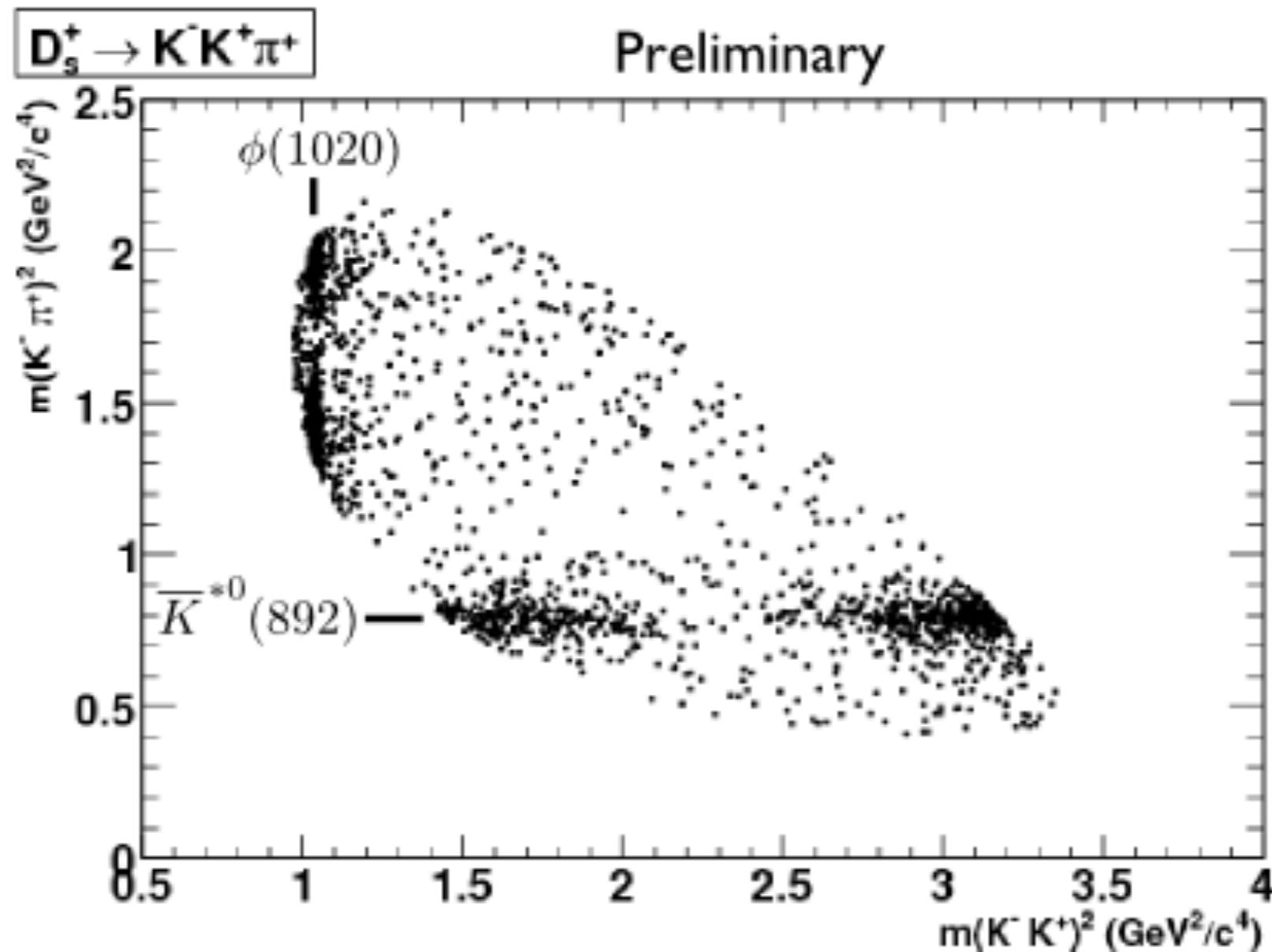
*Want to measure lifetimes of b hadrons, e.g., due to their indirect sensitivity to CPV parameters.*



# The Dalitz plot

- Useful way to visualise a decay 3-body decay

E.g.,  $D_s \rightarrow K^+ K^- \pi^-$



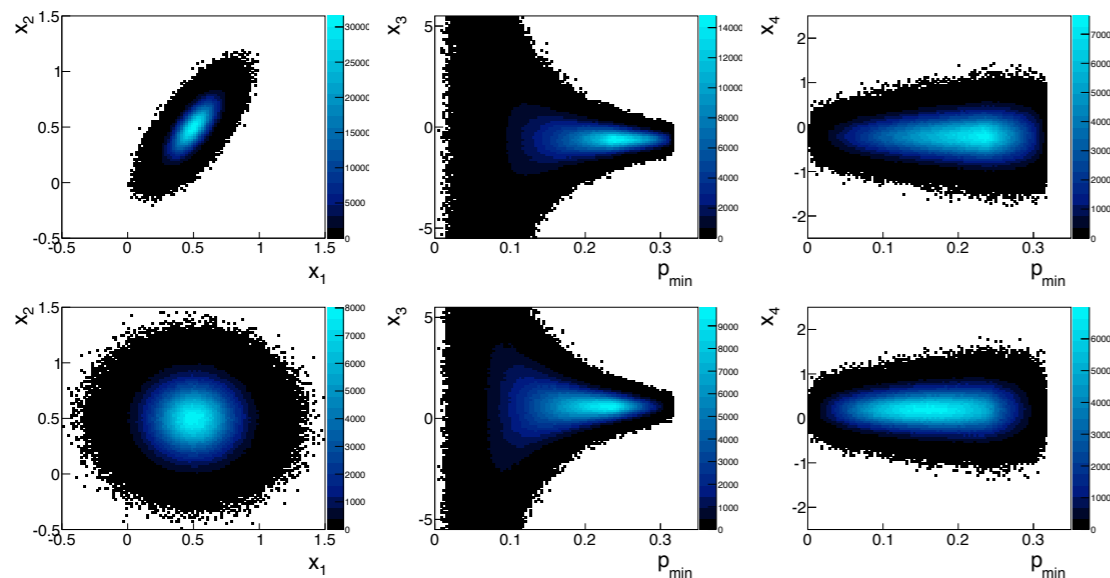
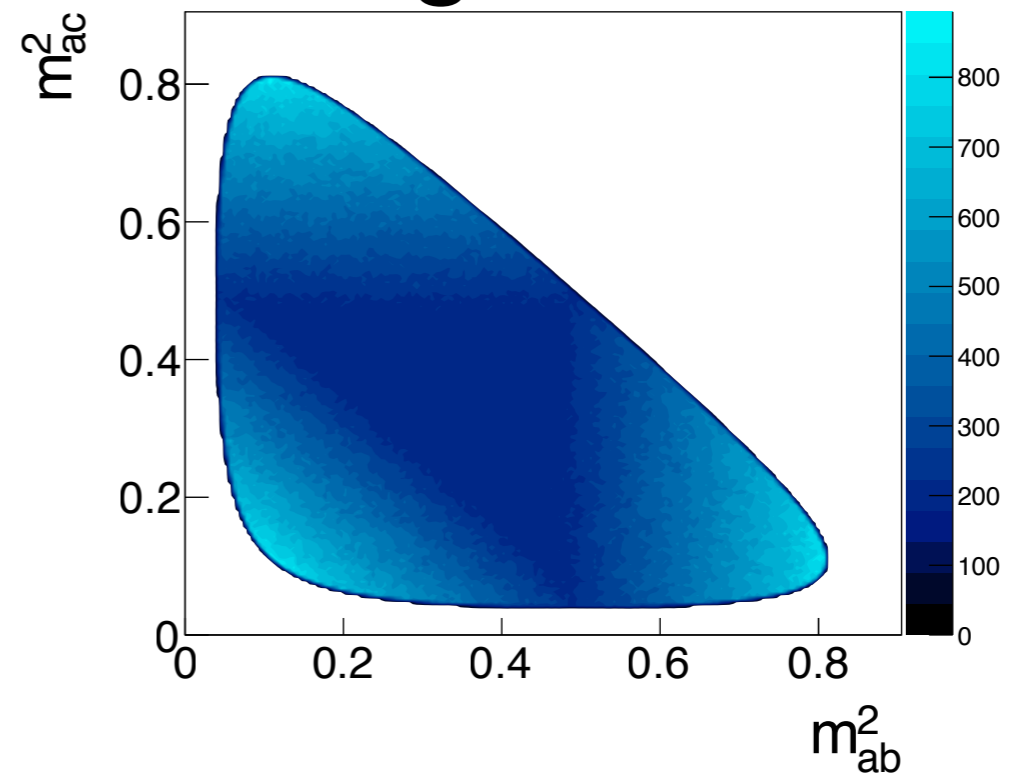
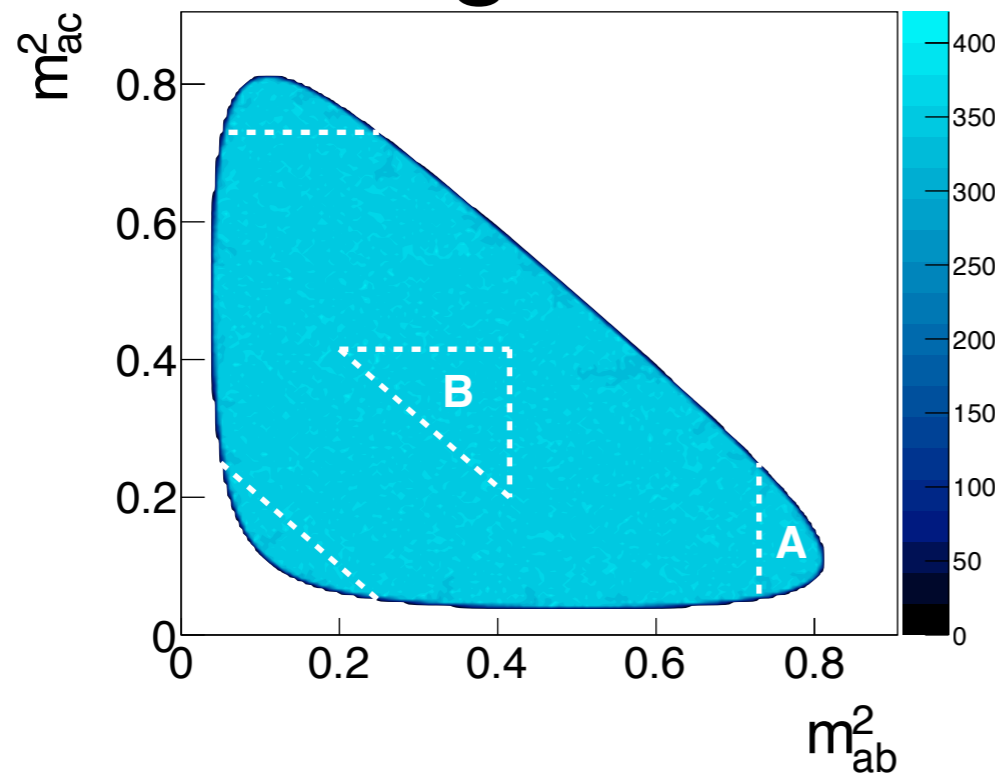


# The problem

Toy Dalitz example  $X \rightarrow abc$

Signal

Background



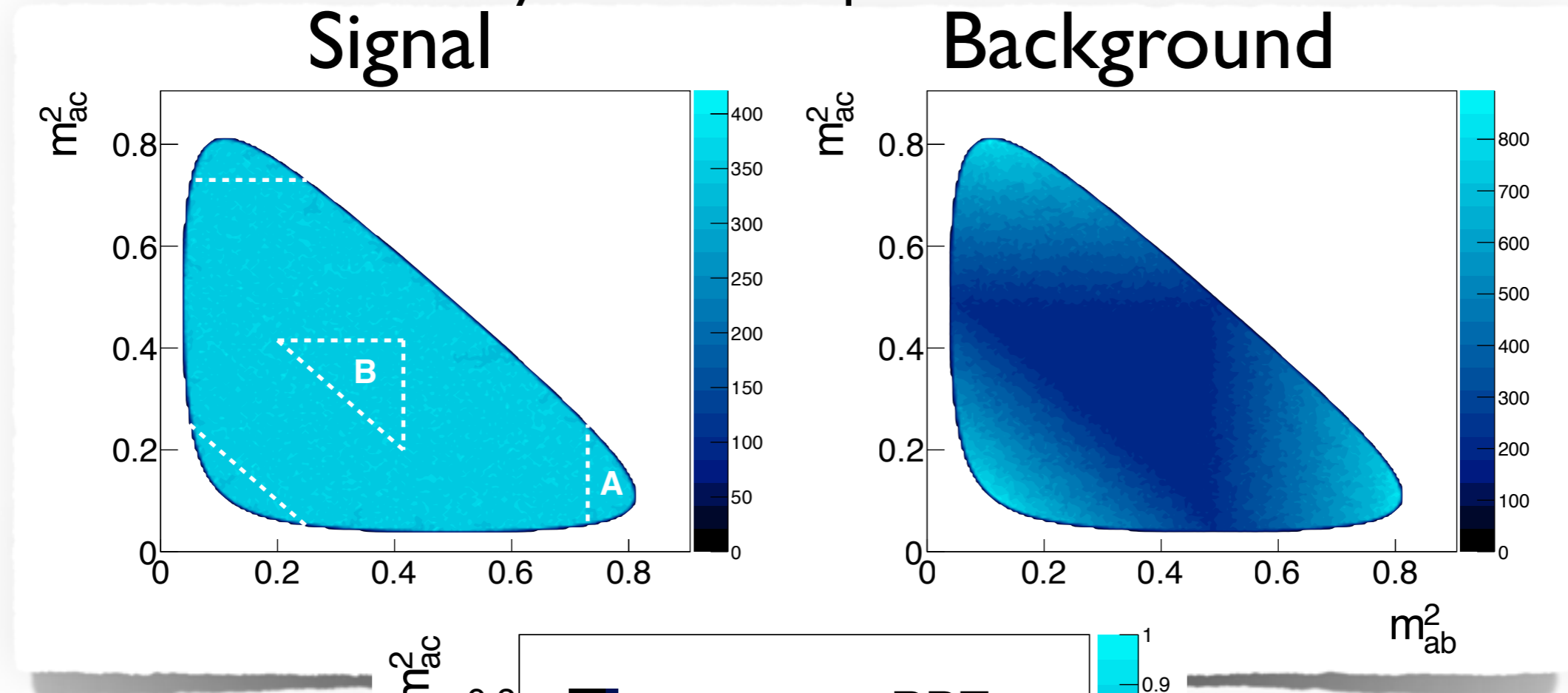
- 4 variables
- 2 correlated with Dalitz
- 2 uncorrelated



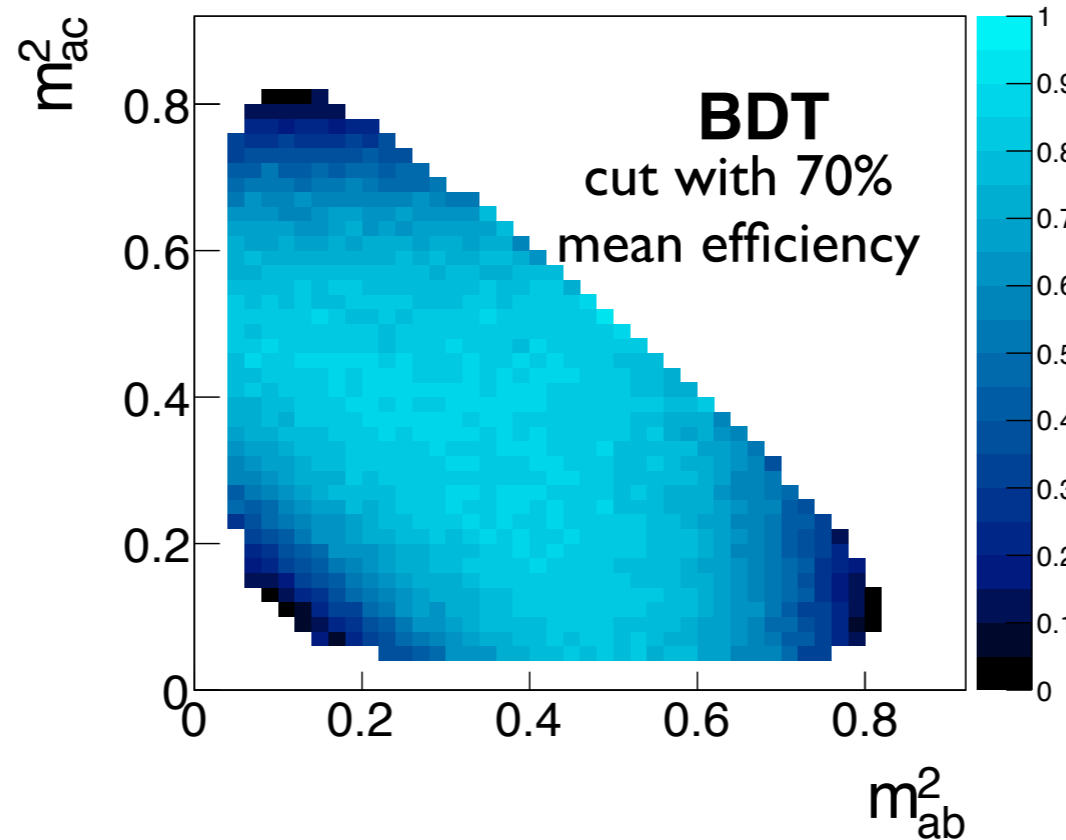
BDT

# The problem

Toy Dalitz example  $X \rightarrow abc$



Biased Dalitz distribution!



Efficiency



# The uBoost method

arXiv:1305.7248v2 [nucl-ex] 24 Dec 2013

## **uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers.**

---

**Justin Stevens and Mike Williams**

*Massachusetts Institute of Technology, Cambridge, MA, United States*

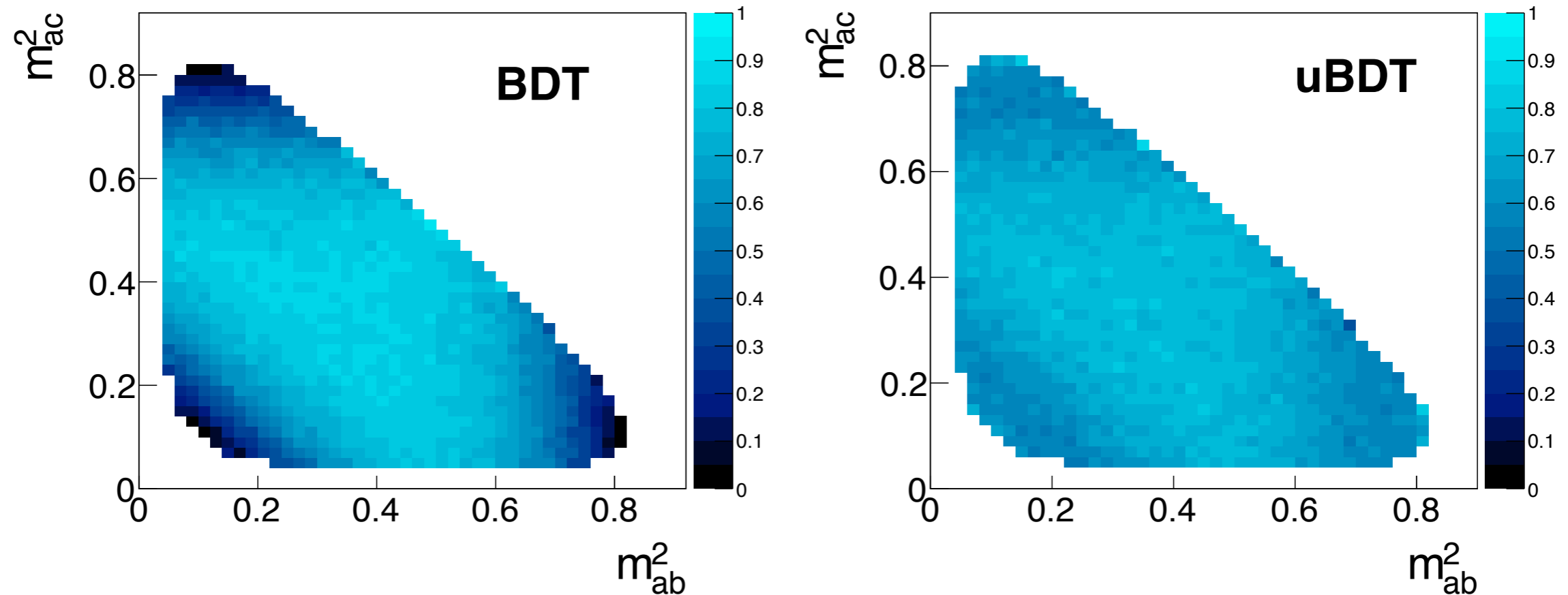
### ABSTRACT:

The use of multivariate classifiers, especially neural networks and decision trees, has become commonplace in particle physics. Typically, a series of classifiers is trained rather than just one to enhance the performance; this is known as boosting. This paper presents a novel method of boosting that produces a uniform selection efficiency in a selected multivariate space. Such a technique is well suited for amplitude analyses or other situations where optimizing a single integrated figure of merit is not what is desired.

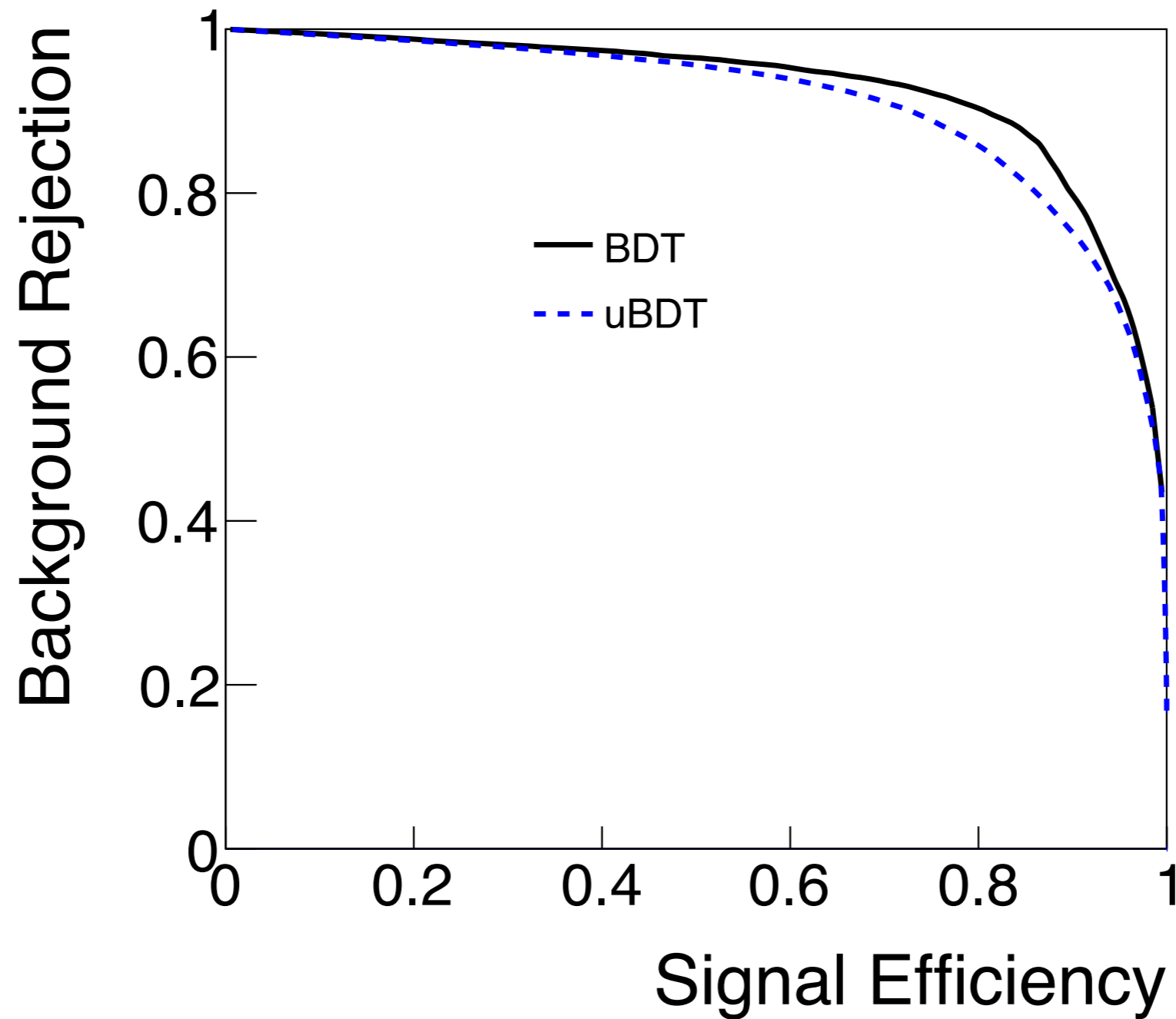
# The uBoost method

- Boosting: train many trees, and assign larger weights to events if they were mis-classified in previous rounds.
- uBoost: assign larger weights to events in which the local efficiency is smaller.

For the same 70% mean efficiency



# The uBoost method





## Closing thought

Surely this is only the beginning of MLHEP in triggers.

- Pattern recognition
- Detector and DQ monitoring
- Data popularity / trigger bandwidth division
- Softer multibody signatures in ATLAS/CMS
- More efficient exclusive lines in LHCb
- ....