Variable latency tracking @ 30 MHz

Connection
- PWGs attaché.e
- Computing attaché.e
- Simulation attaché.e
- Online attaché.e
- Upgrade 2 attaché.e

Coordination
- PL & Deputy/ies
- IB chair (ex-officio)
- Work package coordinators
- WP1 Data Structures
- WP3 Selections
- WP5 QA
- WP2 Reconstruction
- WP4 Align & Calib
- WP6 Accelerators

Institutional Board
- IB chair
- PL & Deputy/ies (ex-officio)
- Institute representatives

Implementation
- WP deliverable responsibles
- Piquets
- Release shifters
- Voluntary developers and PWG line authors

V. V. Gligorov, CNRS/LPNHE

IN2P3 Calcul et Données Town Hall, 17.10.2019

European Research Council
Established by the European Commission

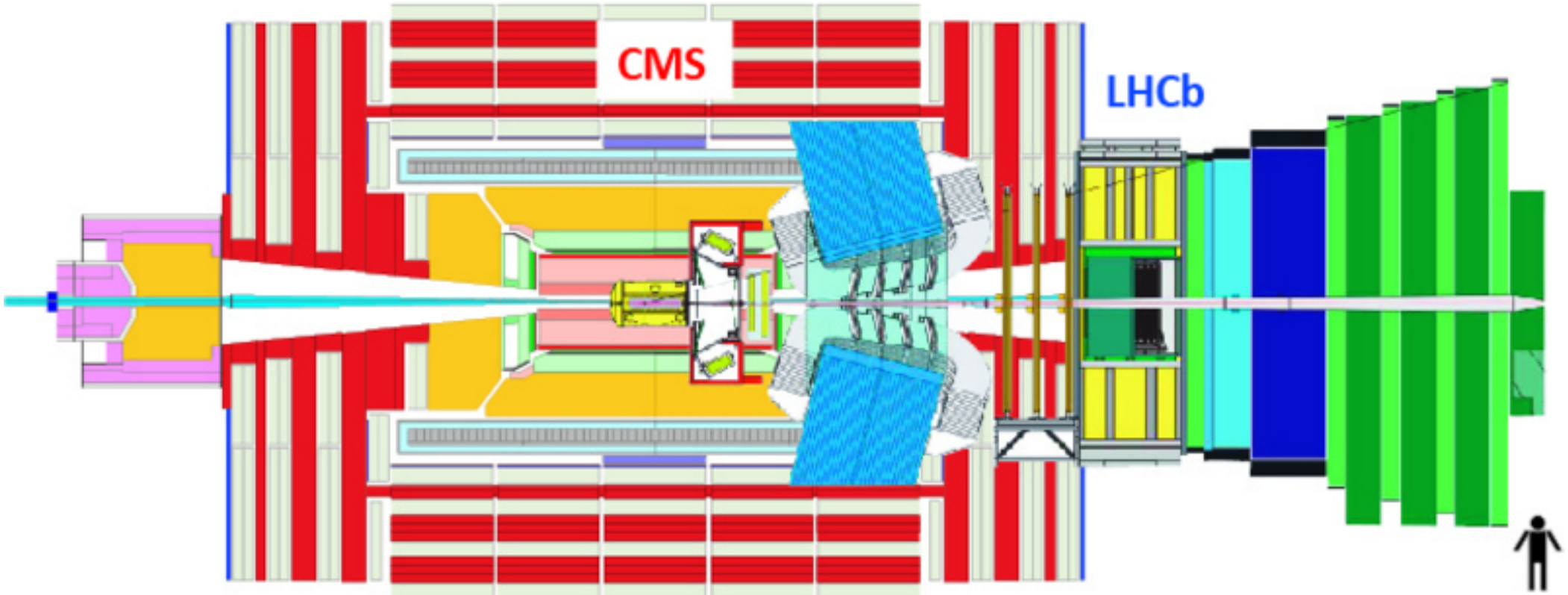
Objectifs de cette presentation

1. Motiver pourquoi il est intéressant d'effectuer une trajectographie à variable latency en temps réel, ce qui, au LHC, signifie à 30 MHz

2. Décrivez les défis liés à la fourniture d'un tel suivi pour l'expérience LHCb en faisant référence à deux architectures spécifiques: x86 et GPU.

3. Donne mes pensées personnelles sur ce que nous avons appris au cours de ce processus de développement à LHCb, et des pensées sur ce que cela va devenir dans le futur

Le détecteur LHCb



Spectrometre “forward" optimisé pour la physique des saveurs lourds

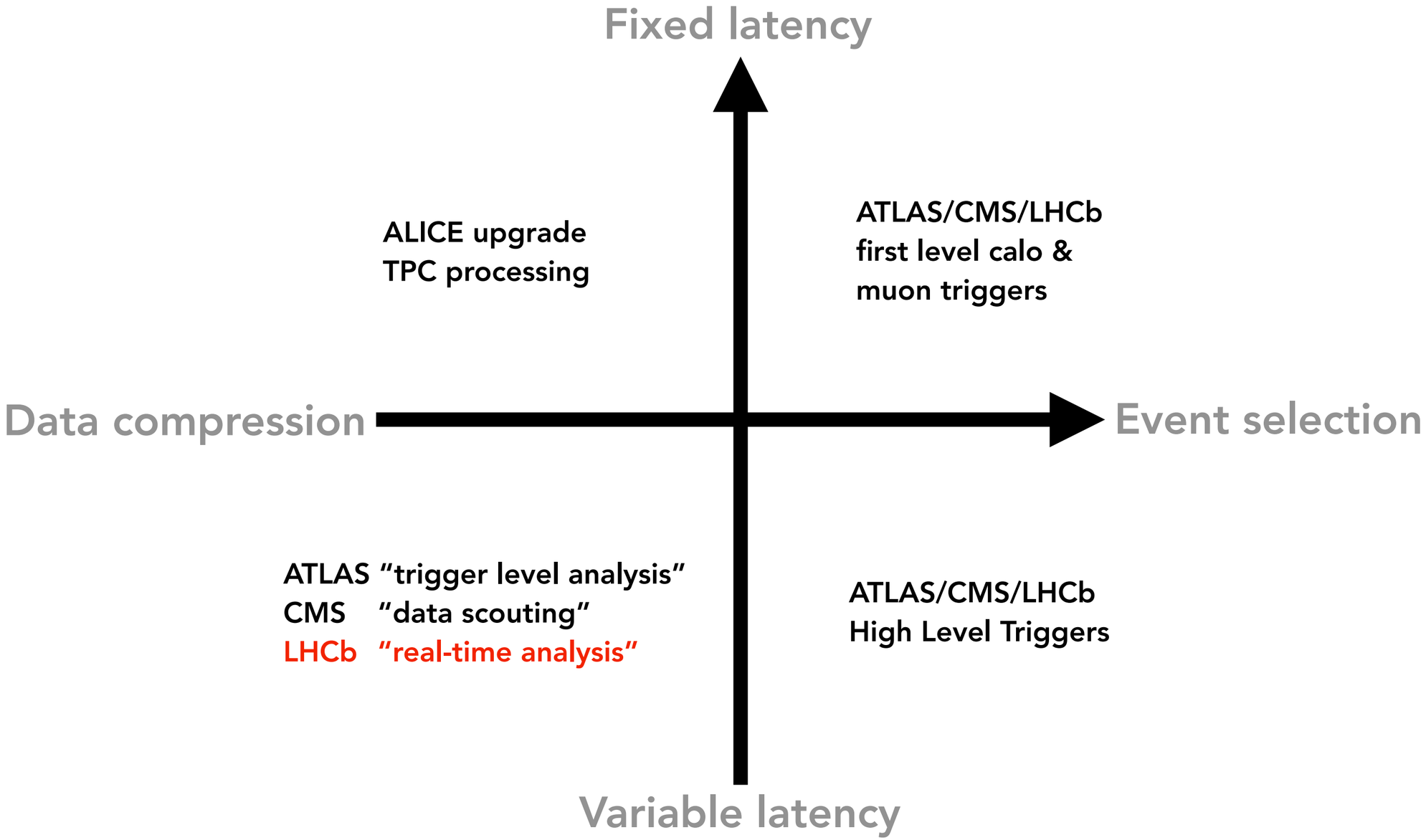
Pourquoi une trajectographie @30 MHz?

Pourquoi variable latency?

LHC processing au temps reel circa 2018

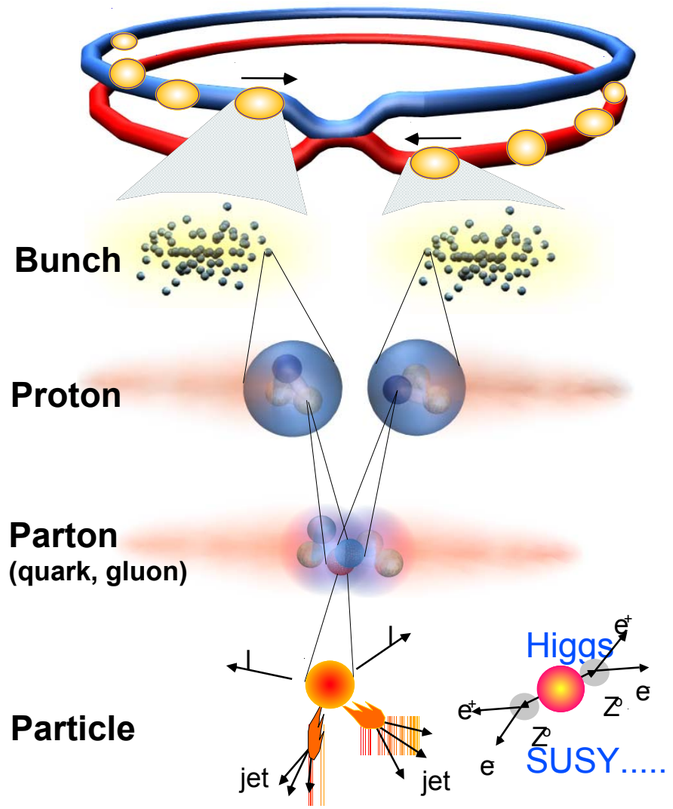
	LHCb		CMS/ATLAS
Volume des données au detecteur	~30 Eb/ année	Taux des données mondiales 2015 ~640 Eb/ année	~1 Zb/ année
Volume des données pour l'analyse	~30 Pb/ année		~40 Pb/ année

Quelles genres du processing en temps reel existent-il?



Processing traditionel en temps reel – "triggering"

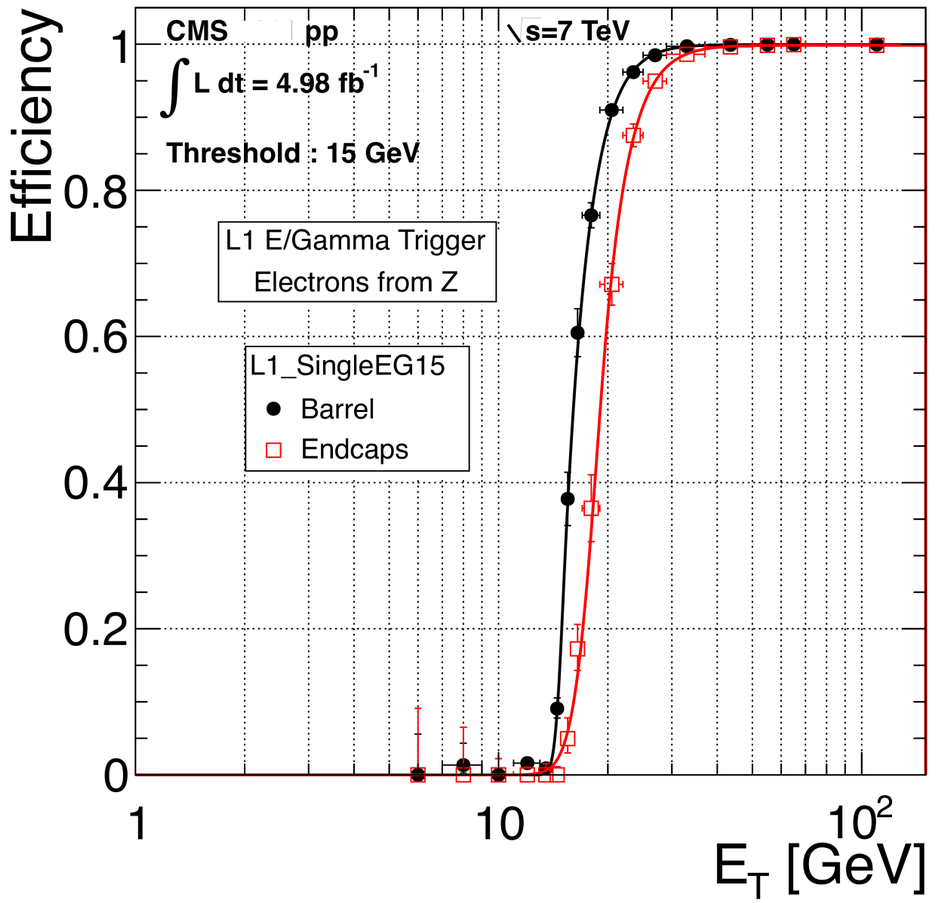
Collisions at the LHC: summary



Proton - Proton	2804 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	$10^{34} cm^{-2} s^{-1}$
Crossing rate	40 MHz
Collision rate ≈	10^{7}-10^{9}

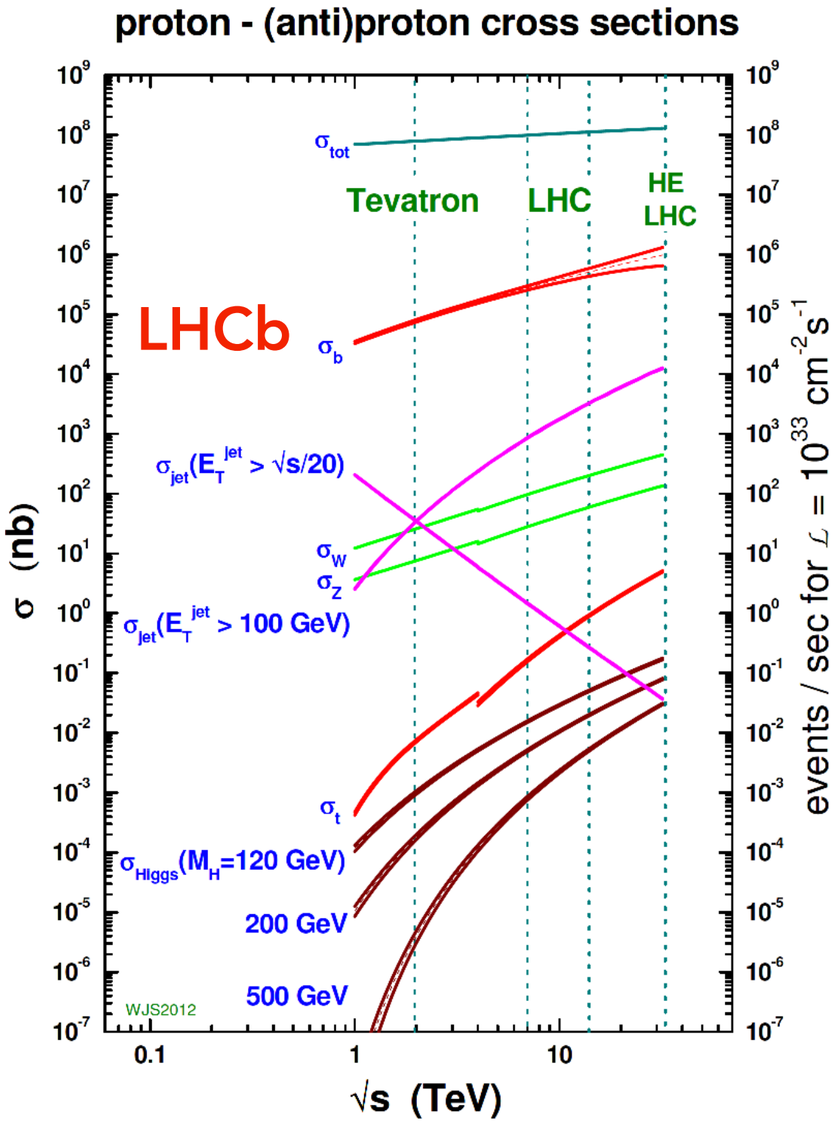
New physics rate ≈ .00001 Hz

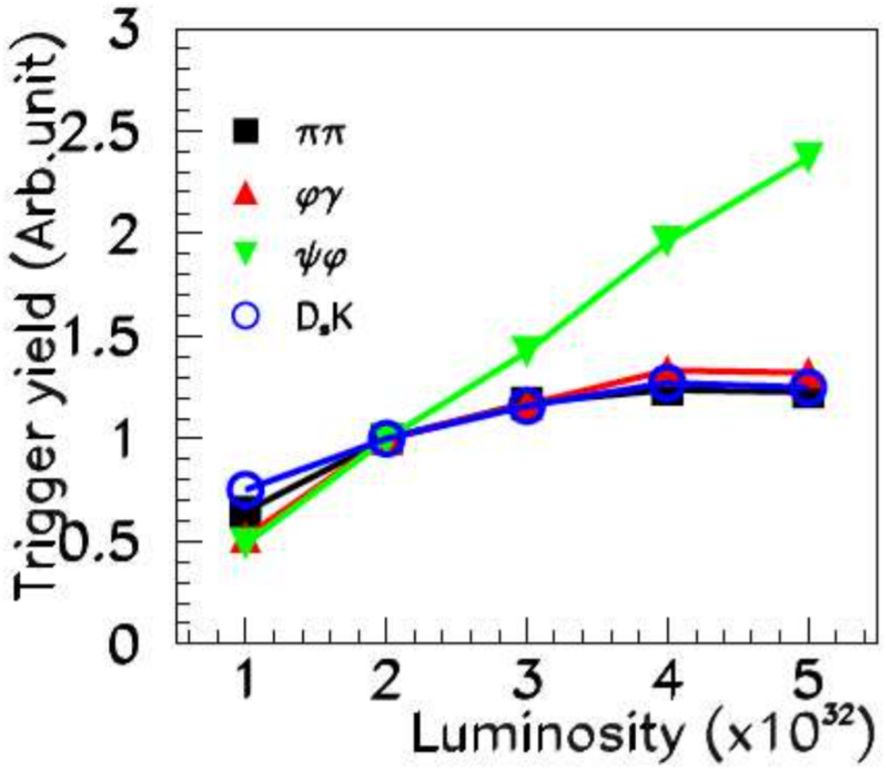
Event selection:
1 in 10,000,000,000,000



Selection des événements rares et avec des critères simples et locales (par example un cluster energetique) au fixed latency

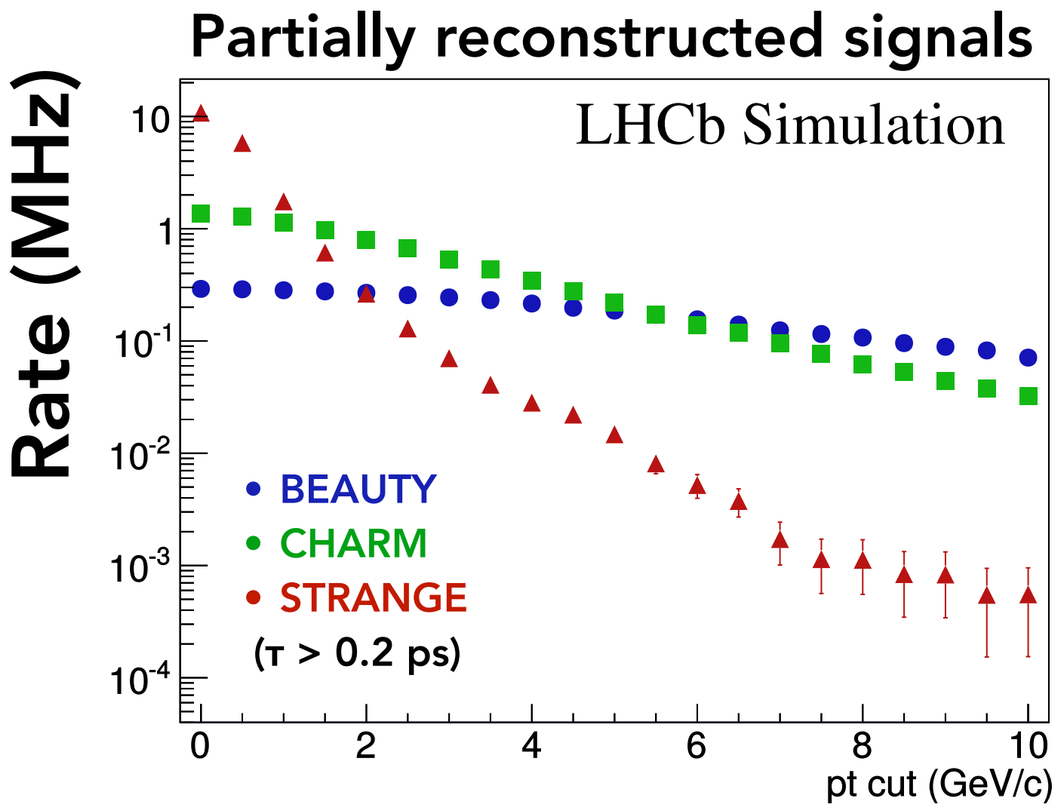
Pourquoi ca s'applique pas au LHCb?





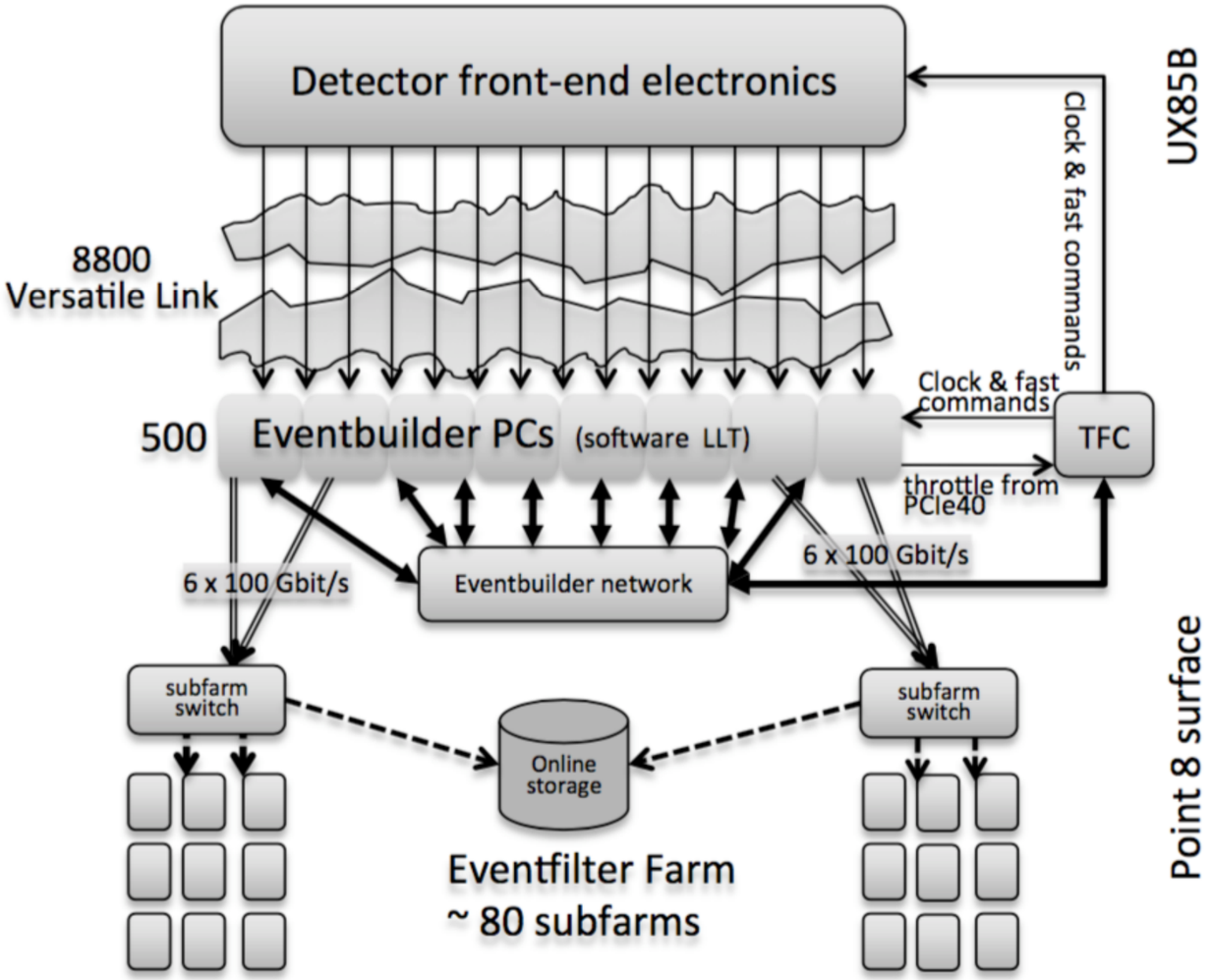
Bruit de QCD trop important après une certaine luminosité!

Taux des signals @ LHCb upgrade au 2021



Des MHz des signals reconstructibles!
Aucune possibilité de discrimination sans une trajectographie efficace au 30 MHz

Le design du DAQ LHCb pour l'upgrade



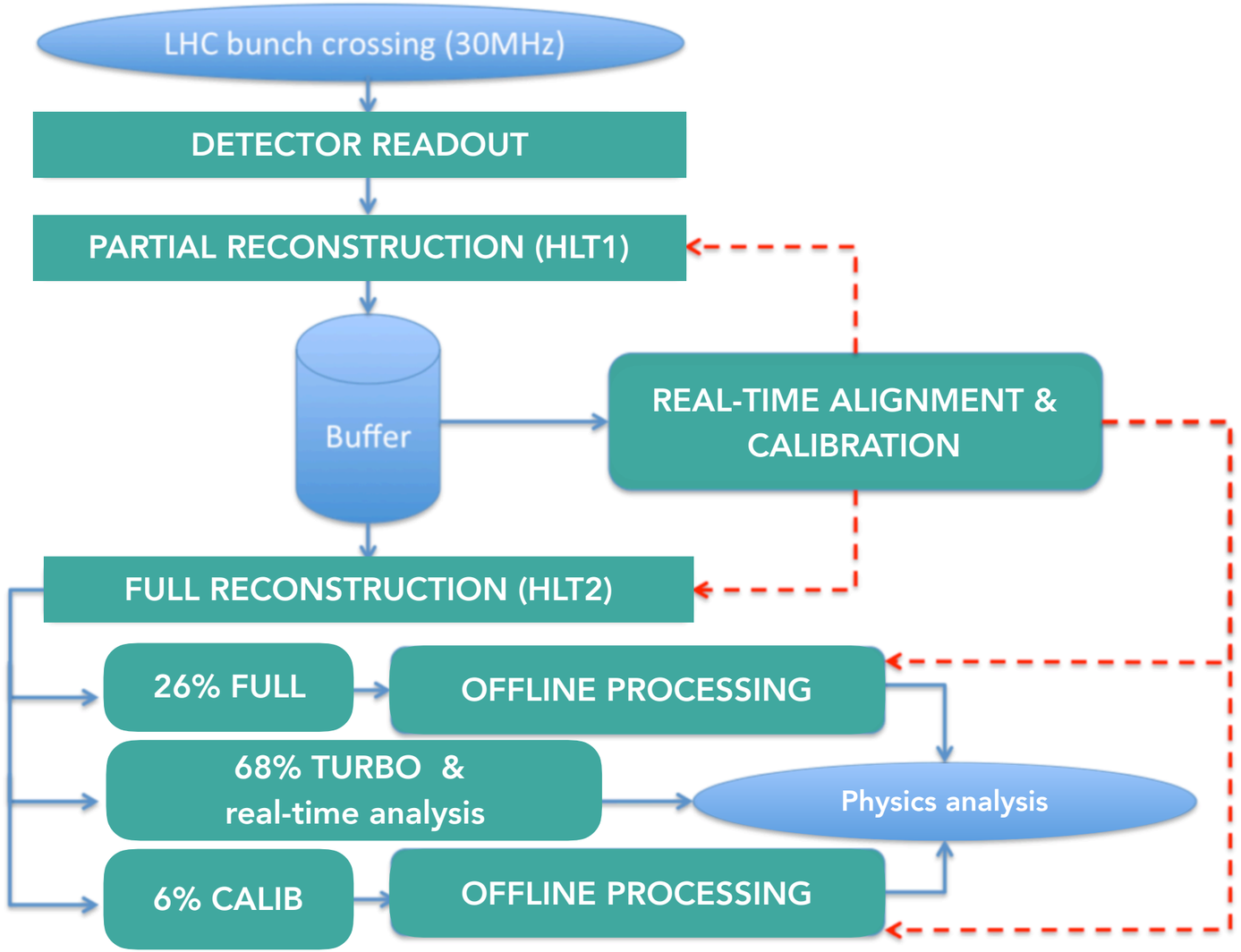
40 Tbit/s des événements complets qui sont envoyé dans un data centre (flexible archi, baseline CPU)

LHCb upgrade dataflow

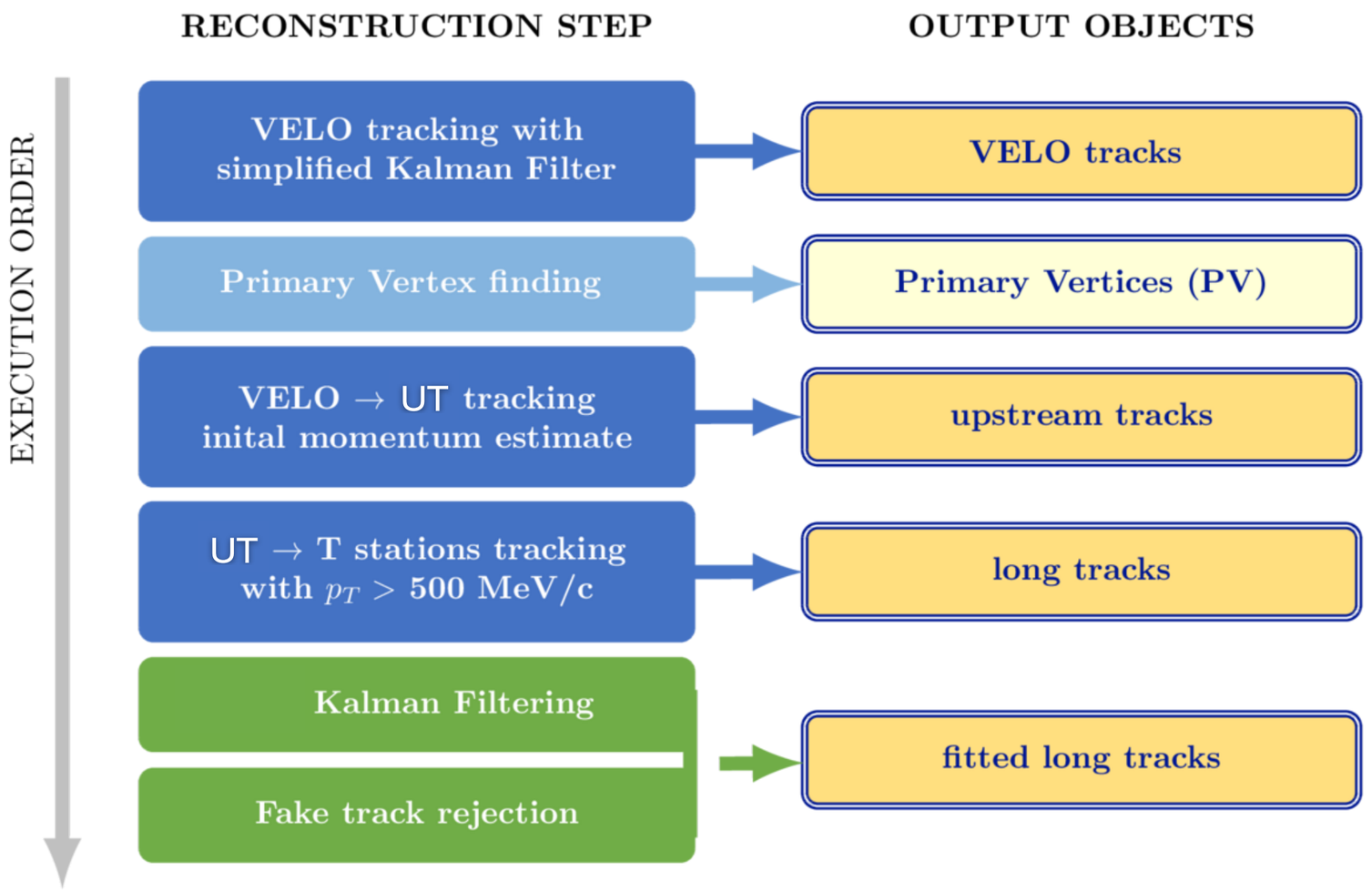
40 Tb/s
30 MHz non-empty pp

1-2 Tb/s
0.5 - 1.5 MHz

80 Gb/s

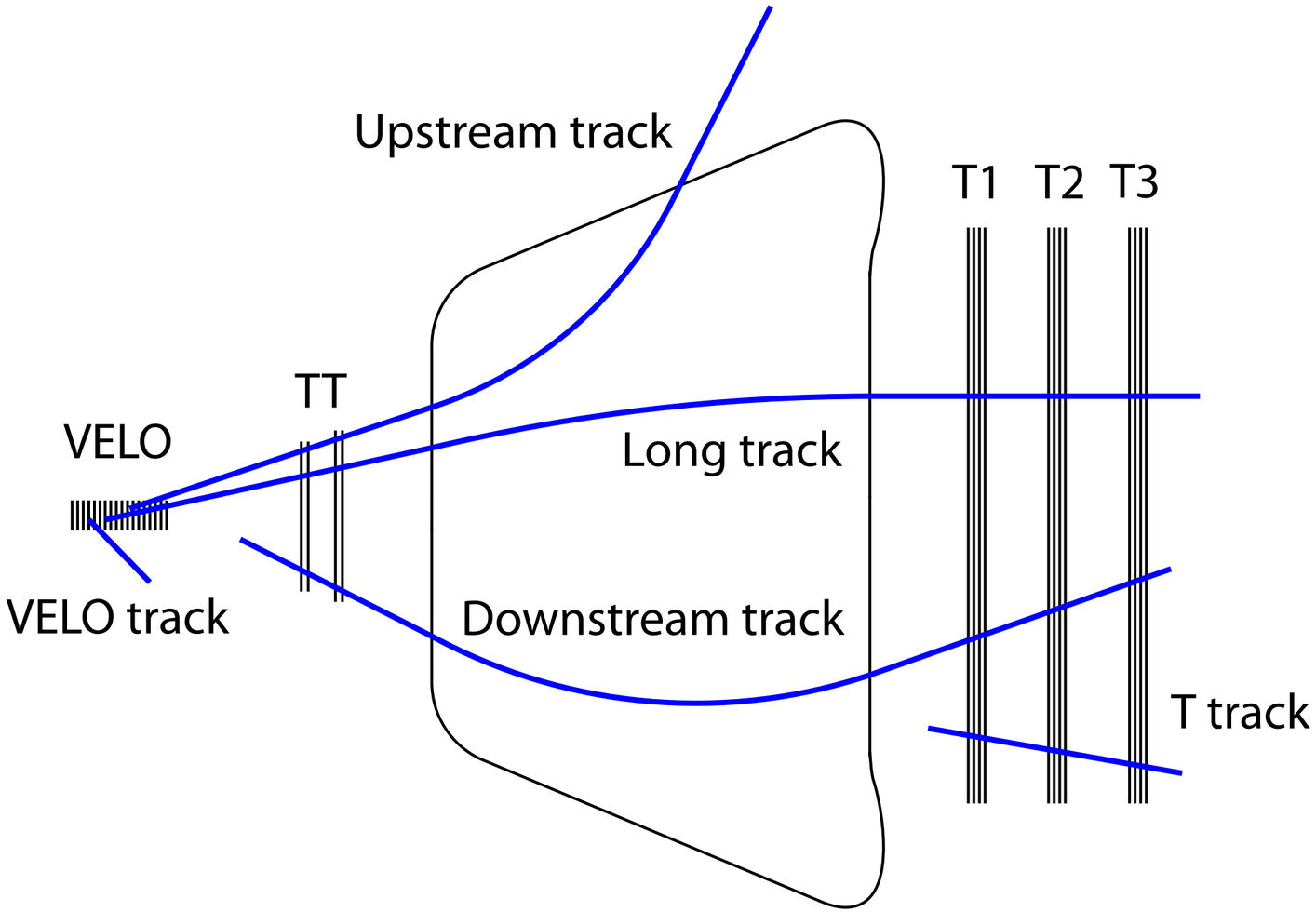


Contenu du premiere étape de processing (HLT1)



Trajectographie des traces au dela du 500 MeV P_T @ 30 MHz!

Mais pourquoi variable latency?



Étant donné la géométrie du LHCb la trajectographie nécessite par conséquent de réunir des données non locales provenant de plusieurs sous-détecteurs.

Vous pouvez créer un trajectographie à latence fixe avec des FPGAs, mais vous devrez néanmoins construire la plus grande partie de la lecture du détecteur. Vous pouvez également tout lire à l'avance et travailler à latence variable.

Ce n'est pas un argument de ne pas utiliser des FPGA, il faut simplement d'abord créer des événements, puis les traiter de la manière la plus rentable.

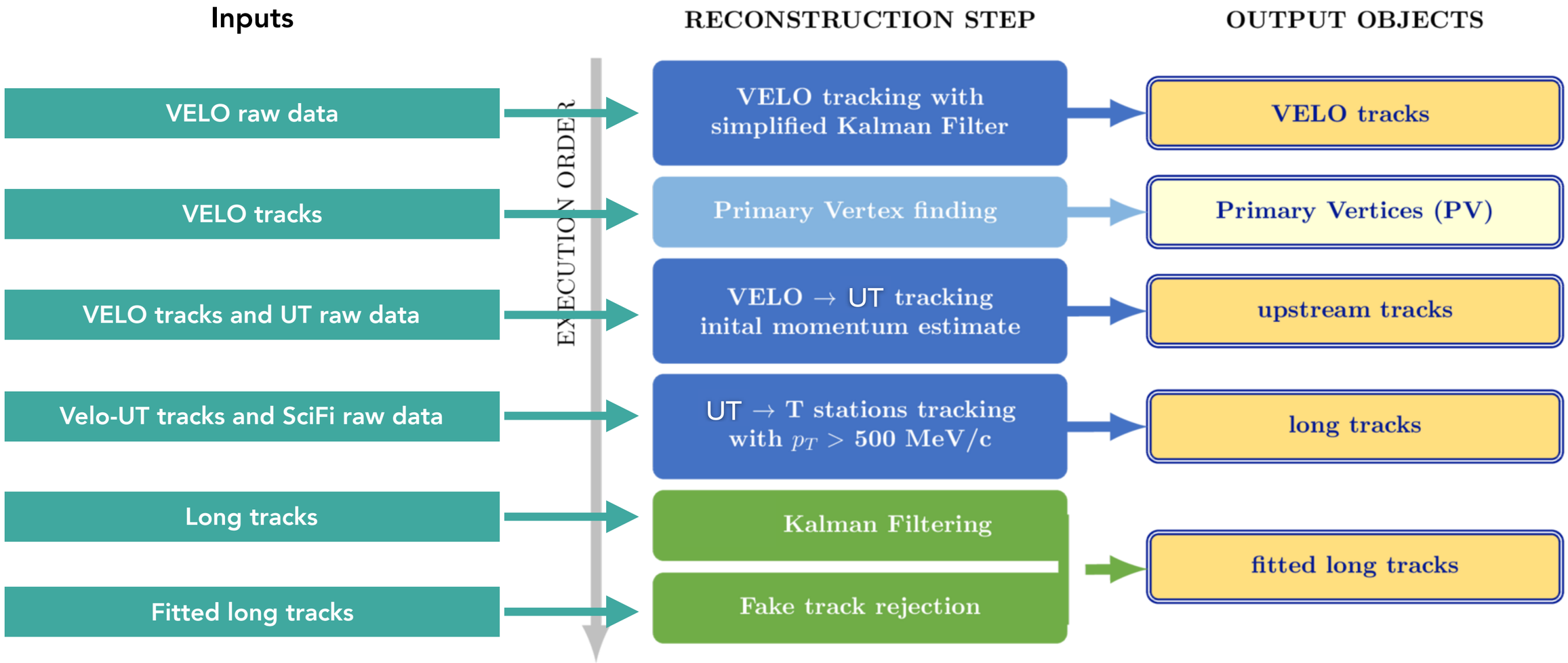
Comparaison au ATLAS/CMS HL-LHC

CMS detector	LHC Run-2	HL-LHC Phase-2	
Peak ⟨PU⟩	60	140	200
L1 accept rate (maximum)	100 kHz	500 kHz	750 kHz
Event Size	2.0 MB [a]	5.7 MB [b]	7.4 MB
Event Network throughput	1.6 Tb/s	23 Tb/s	44 Tb/s
Event Network buffer (60 seconds)	12 TB	171 TB	333 TB
HLT accept rate	1 kHz	5 kHz	7.5 kHz
HLT computing power [c]	0.5 MHS06	4.5 MHS06	9.2 MHS06
Storage throughput	2.5 GB/s	31 GB/s	61 GB/s
Storage capacity needed (1 day)	0.2 PB	2.7 PB	5.3 PB

Meme taux des données mais 6 ans avant et avec un budget beaucoup plus faible...

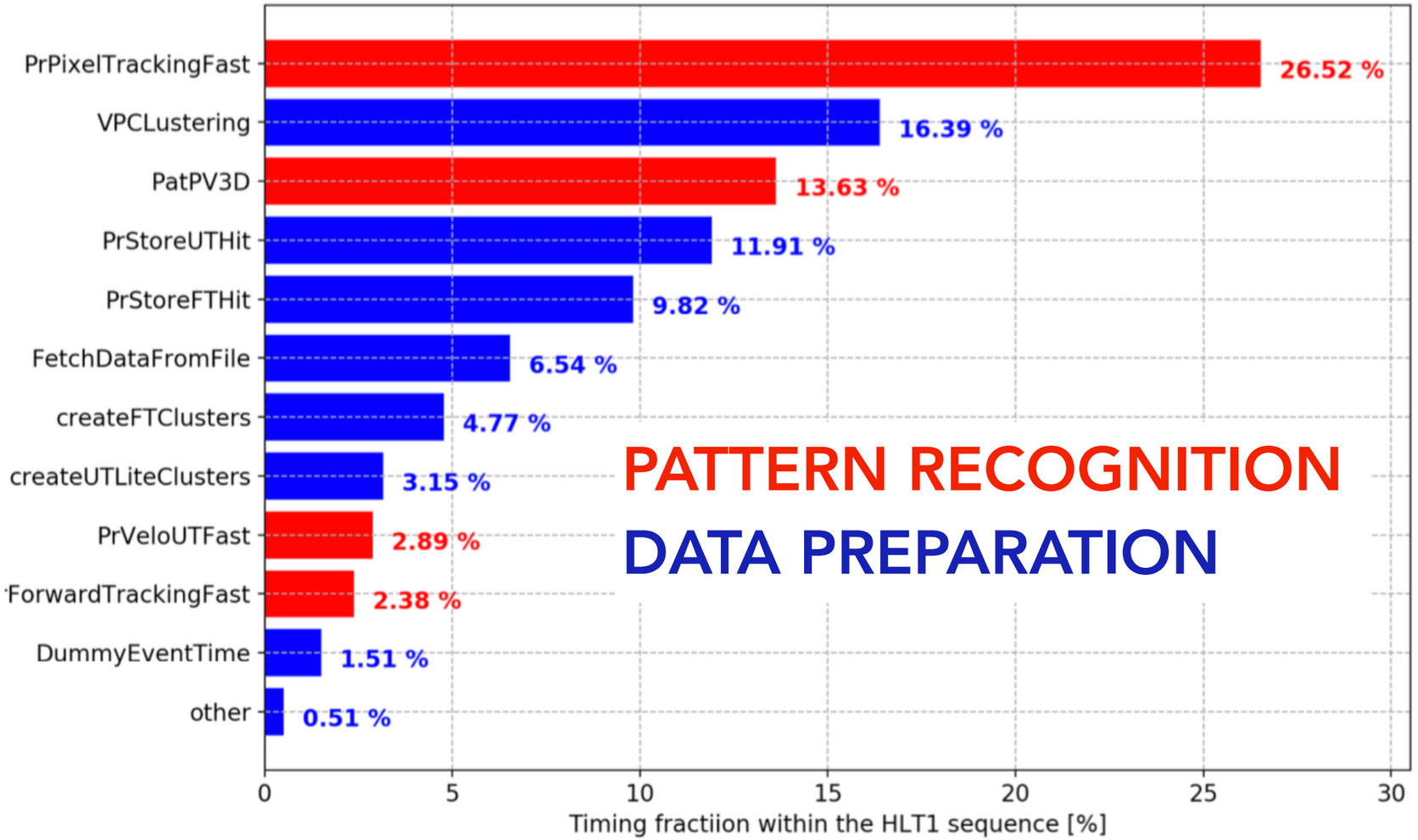
Challenges et solutions

Regardons la sequence en detail



Il faut d'abord recevoir les données, les transformer dans les coordonnées globales, et puis faire le pattern recognition...

On a commence a environ 3 MHz au 2018...

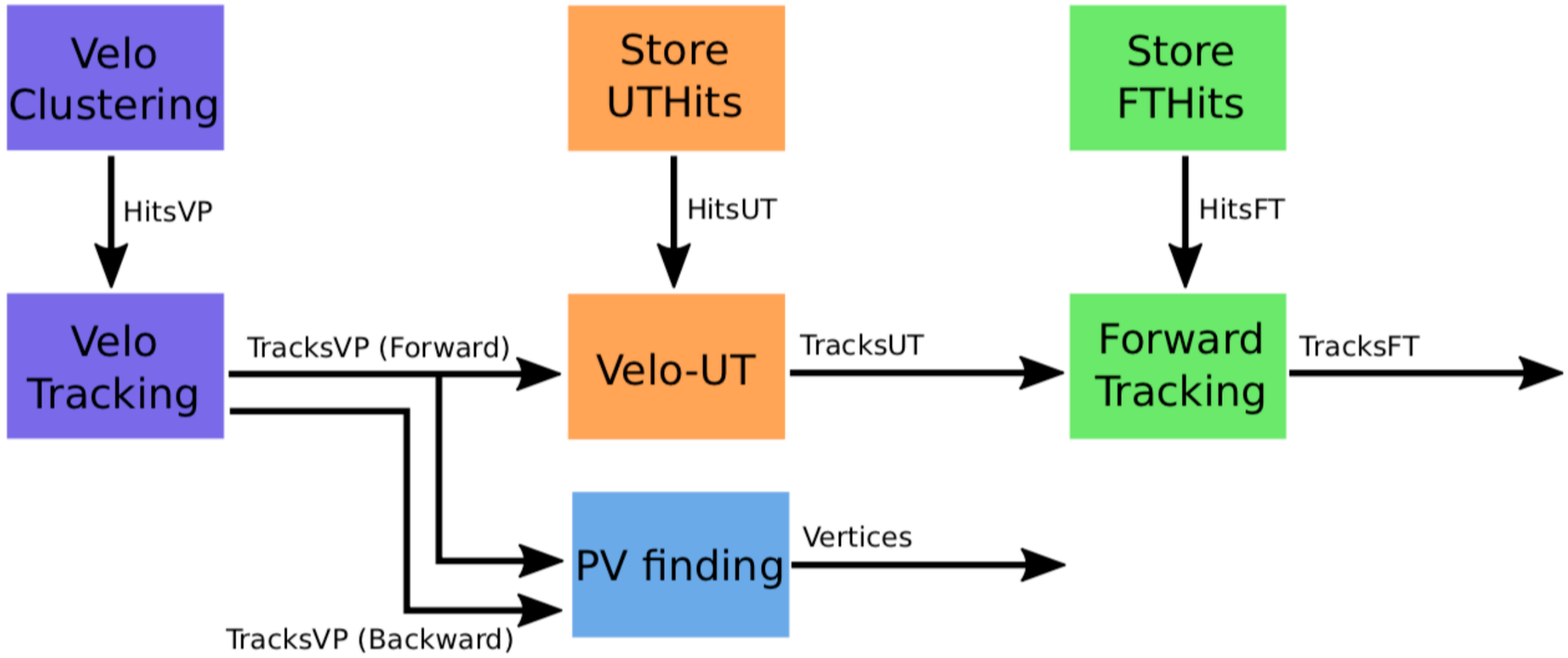


La preparation des données est aussi important que la trajectographie!

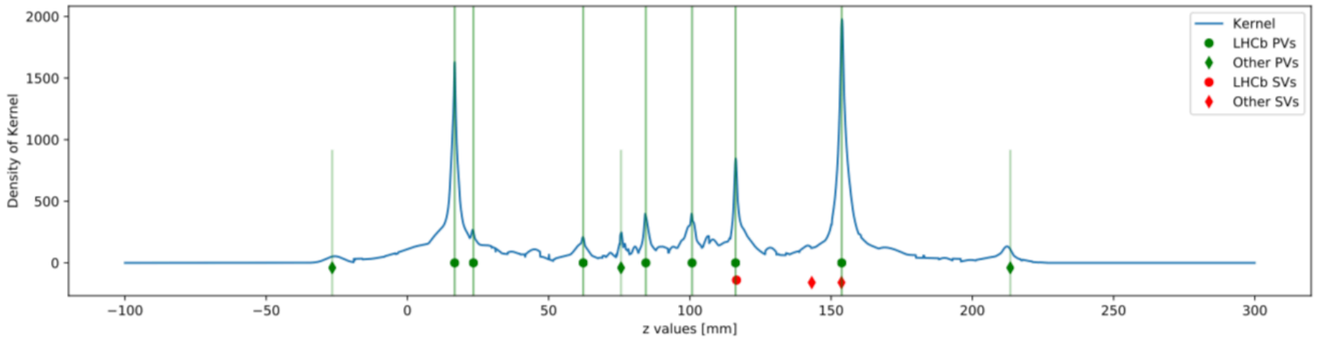
Comment améliorer?

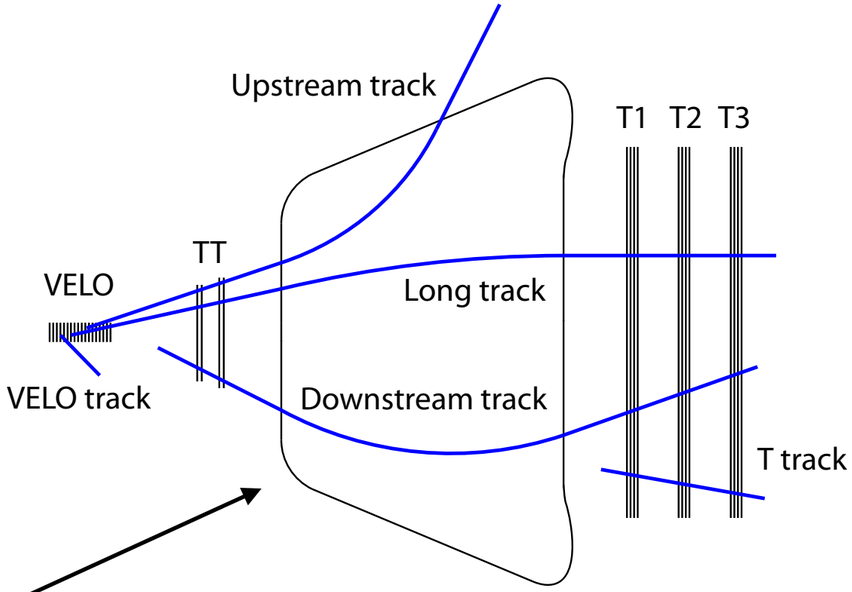
1. Faites ce que vous pouvez sur les FPGAs! Produisez les données dans le format le plus utile possible, effectuez le clustering dans le readout si vous le pouvez.
2. Écrivez des structures de données orientées sur le débit, qui ne contiennent que le minimum absolu requis par le pattern recognition. “Plain old data”
3. Travaillez autant que possible avec les structures SOA pour permettre la vectorisation.
4. Minimisez la copie des informations en décomposant de grandes structuresen morceaux plus petits - par exemple séparer les track states et track hits.

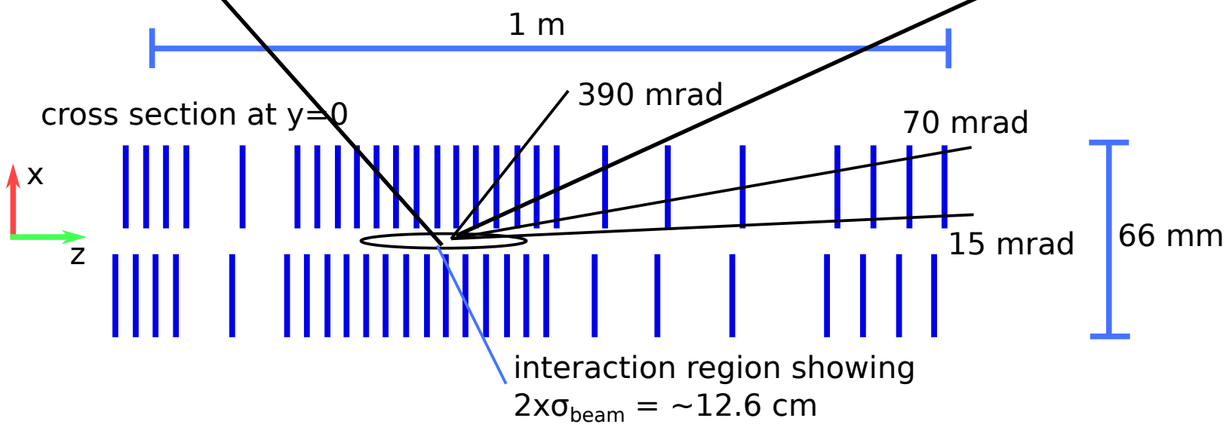
Regardons la sequence en concret...



Petit example illustratif – pourquoi decomposer le VELO?

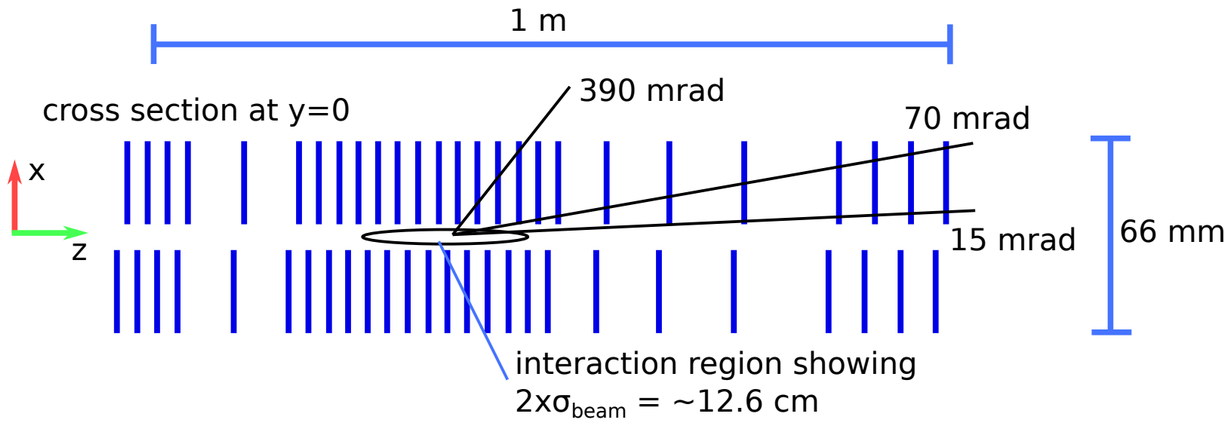


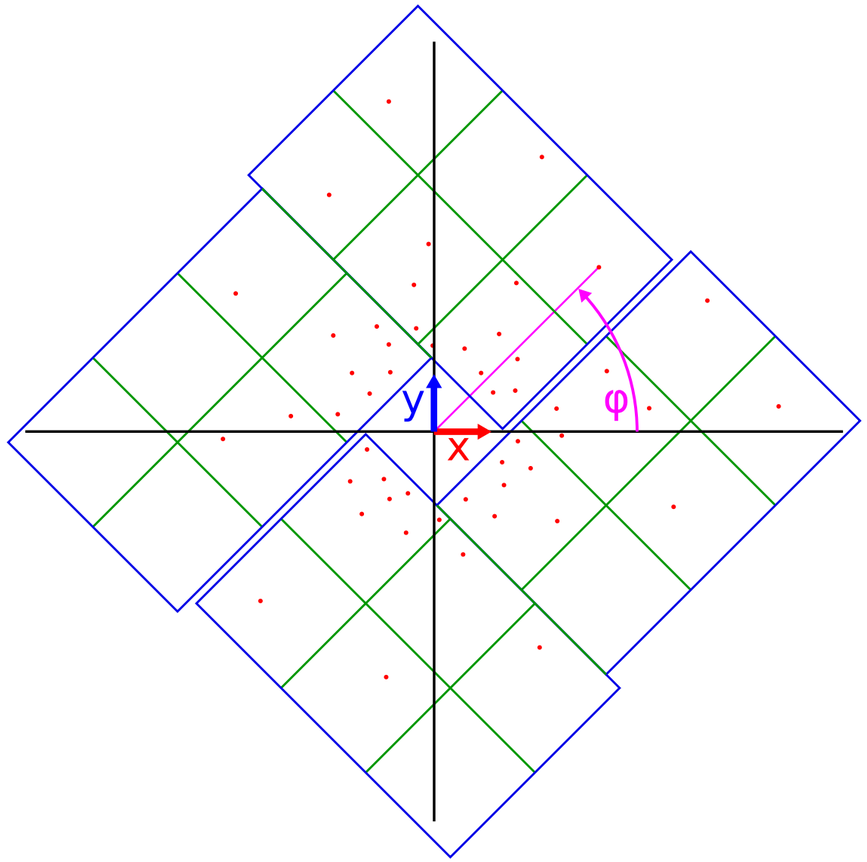




Accentuer le “memory locality" des objets liées

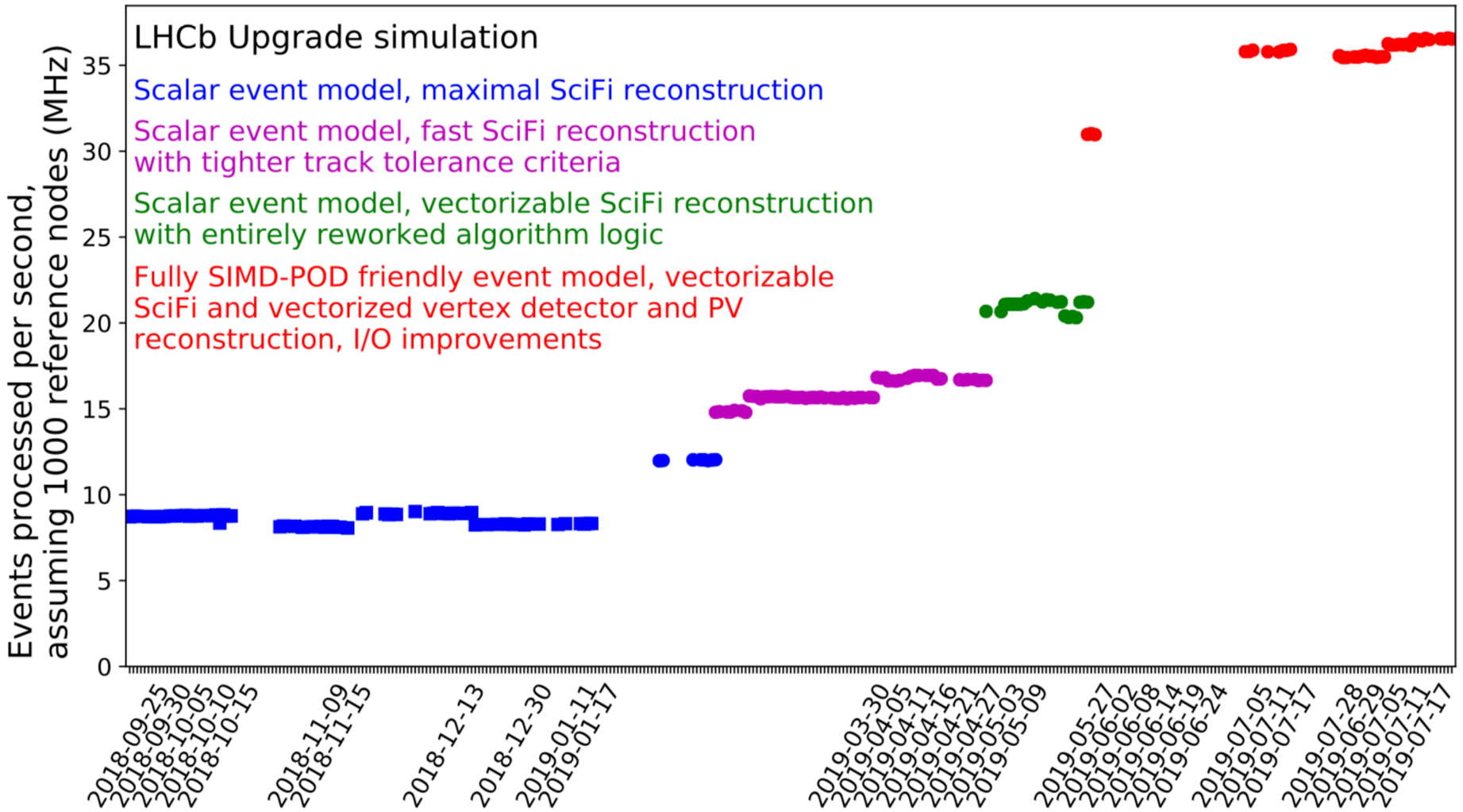
Comment utiliser la géométrie du détecteur?





Les traces traversent lignes du constant φ
Cherchers les N nearest neighbours est plus effectif que chercher les hits dans un "search window" d'une taille fixe en φ

Et nous sommes la!



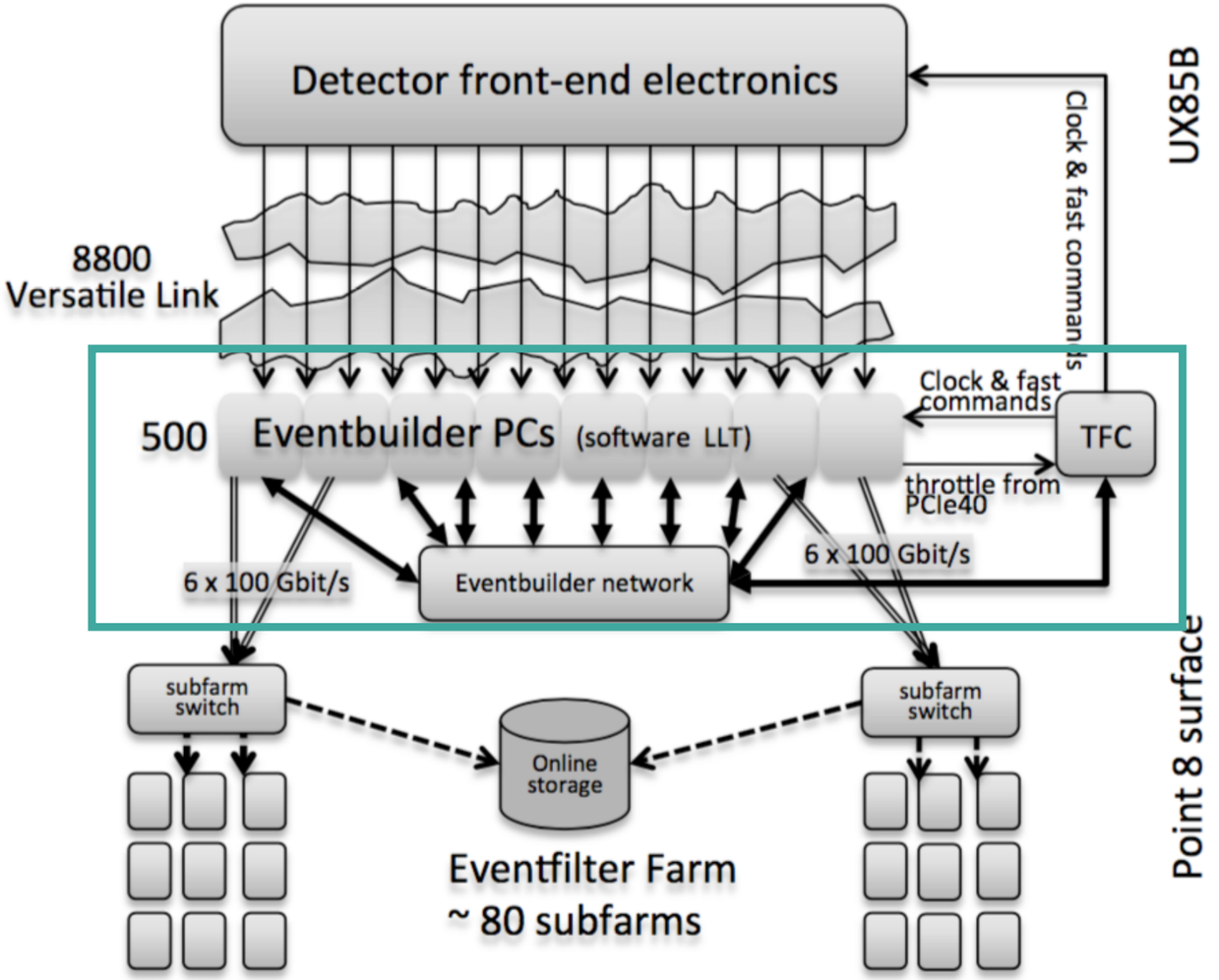
Trigger du premiere niveau au CPUs est faisable! La simplification des données et la vectorisation ont un gain tres complementaire

Et nous avons aussi développée un HLT1 sur les GPUs!

LHCb-ANA-20XX-YYY
May 31, 2019

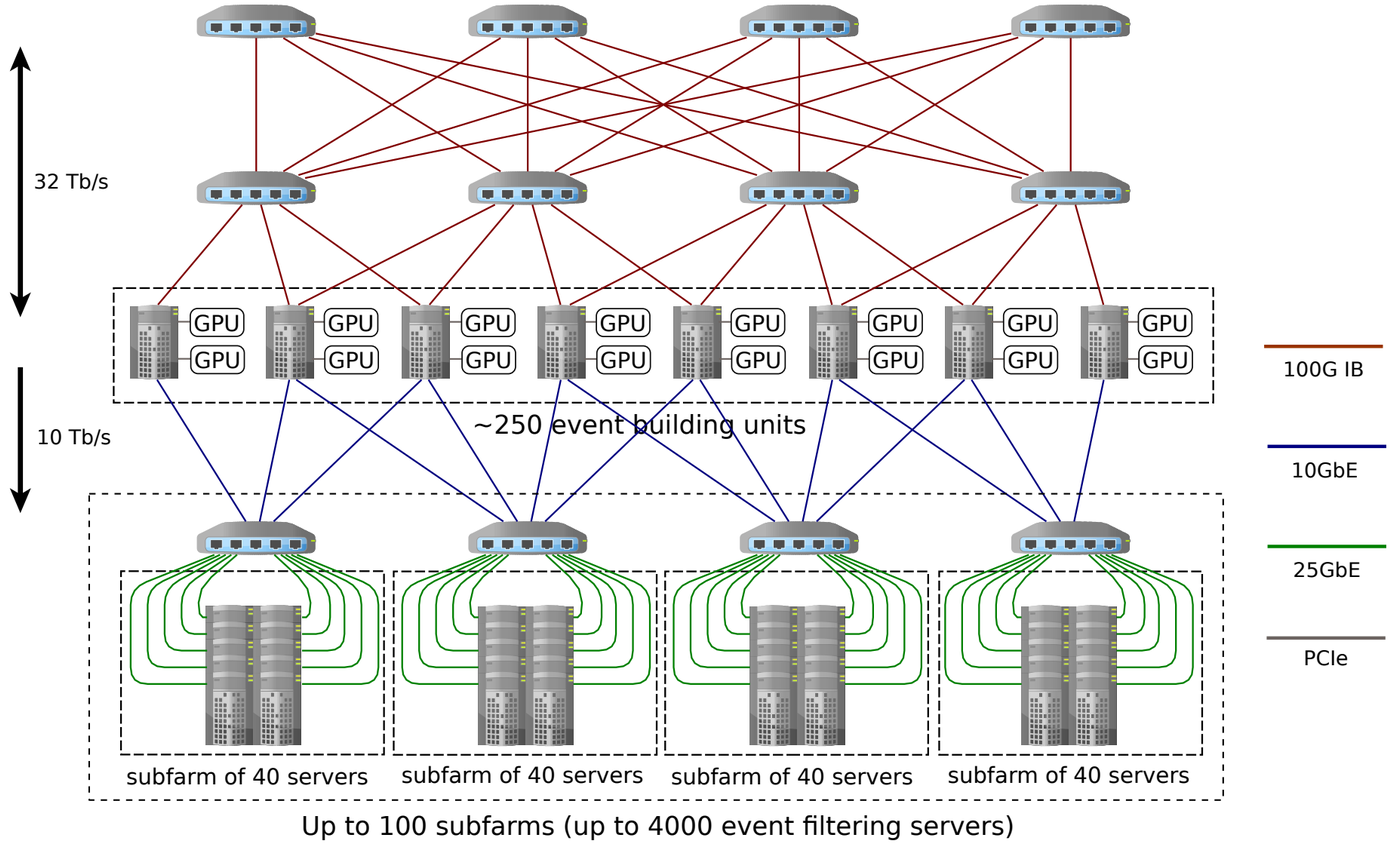
Proposal for an HLT1 implementation on GPUs for the LHCb experiment

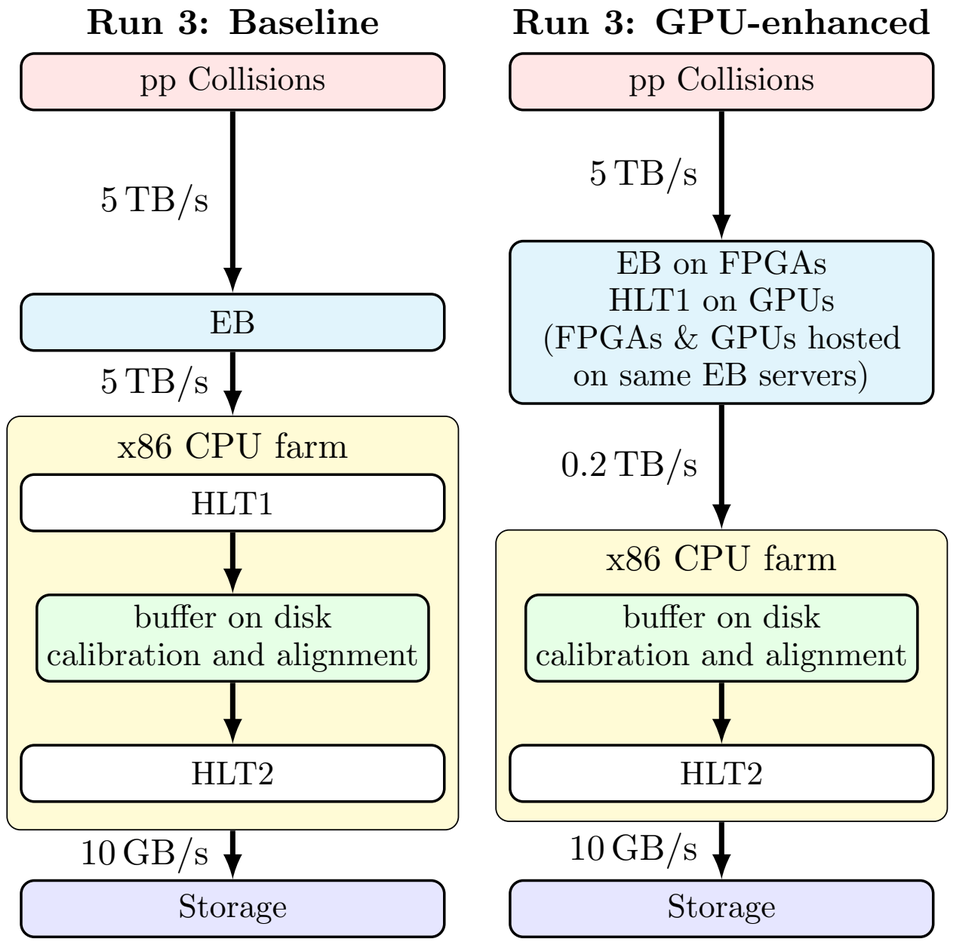
R. Aaij[1], J. Albrecht[2], M. Belous[a,3], T. Boettcher[4], A. Brea Rodríguez[5], D. vom Bruch[6], D. H. Cámpora Pérez[b,7], A. Casais Vidal[5], P. Fernandez Declara[c,7], L. Funke[2], V. V. Gligorov[6], B. Jashal[9], N. Kazeev[a,3], D. Martínez Santos[5], F. Pisani[d,e,7], D. Pliushchenko[f,3], S. Popov[a,3], M. Rangel[10], F. Reiss[6], C. Sánchez Mayordomo[9], R. Schwemmer[7], M. Sokoloff[11], A. Ustyuzhanin[a,3], X. Vilasís-Cardona[8], M. Williams[4]



Exploite la flexibilité de notre DAQ Run 3 en implémentant HLT1 directement dans les serveurs recevant les données du détecteur. Jugée viable par un review externe, une analyse coûts-avantages complète est en cours pour déterminer si nous utiliserons cela lors Run 3.

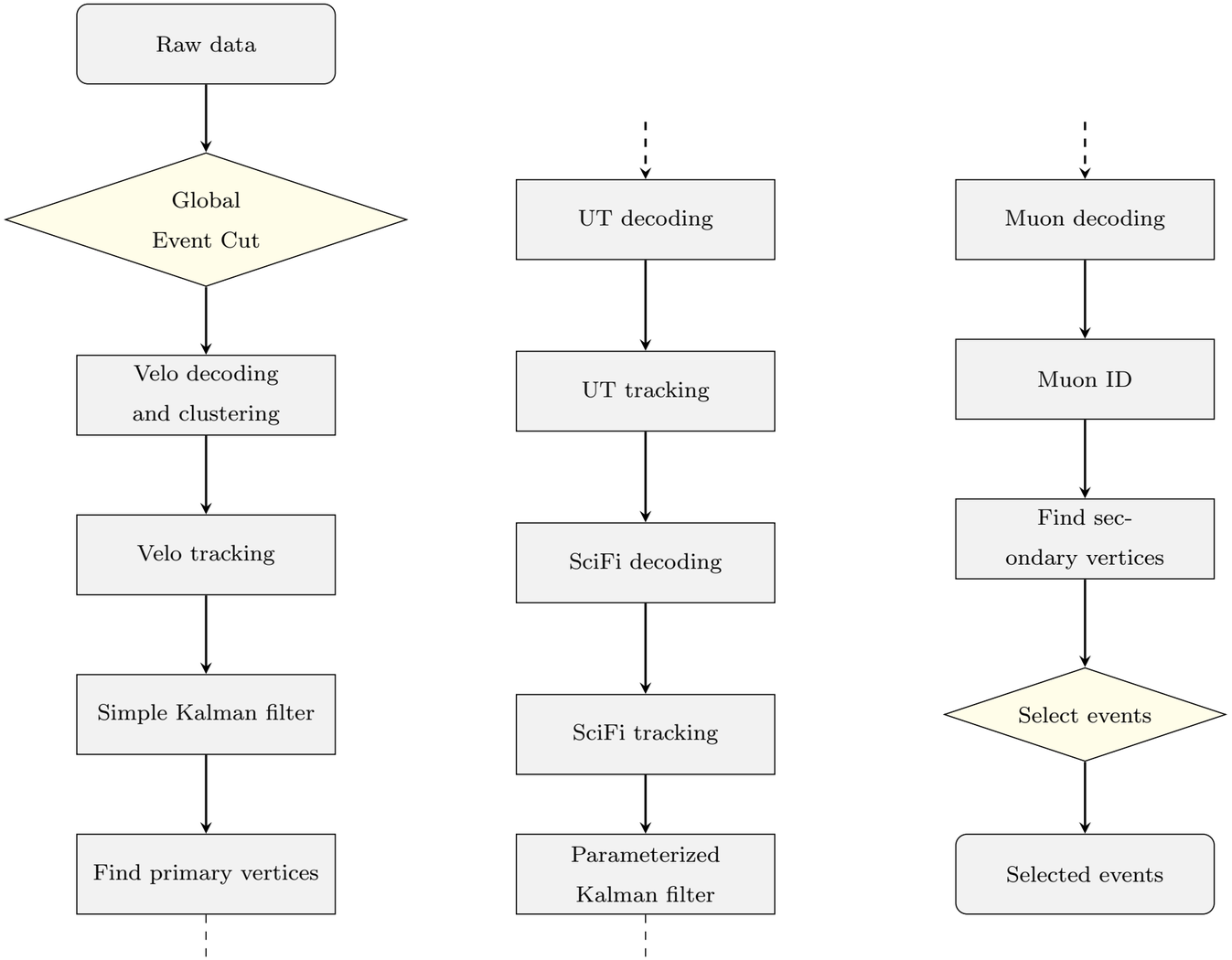
Architecture d'un trigger GPU @ 30 MHz





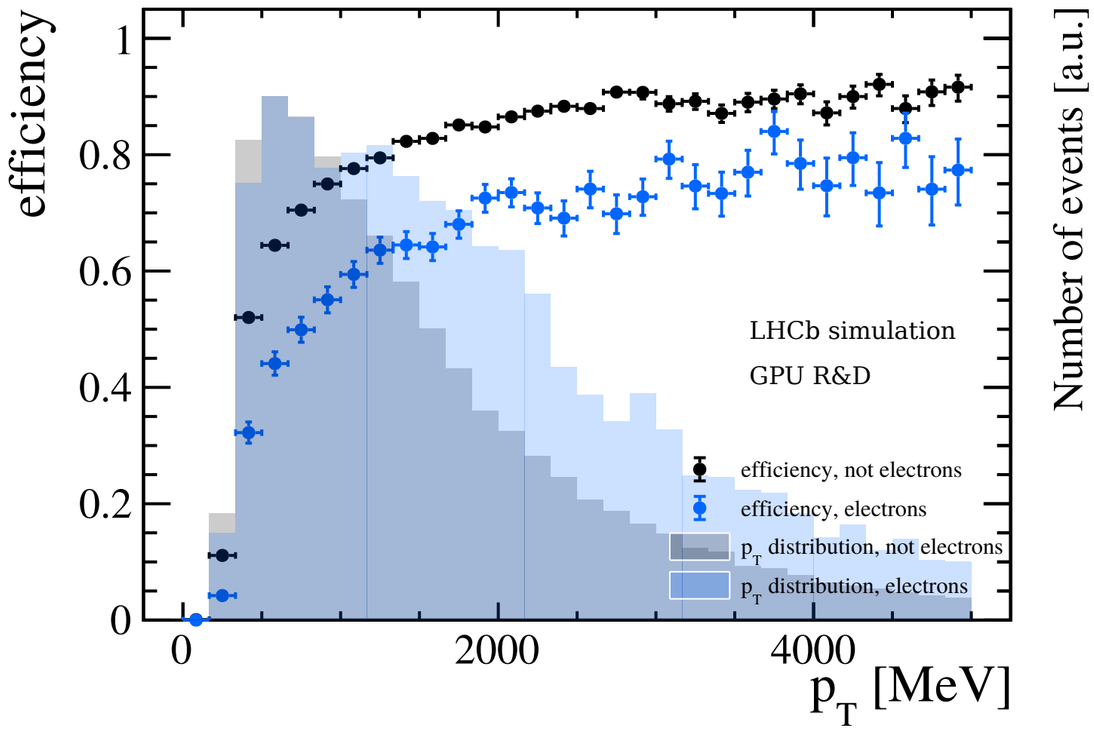
Exploiter les emplacements vides sur les serveurs - opportuniste mais efficace
Chaque GPU consomme 6 GB/s - les premiers tests d'intégration donnent de bons résultats pour le I/O, testes plus severes sont en cours de finalisation

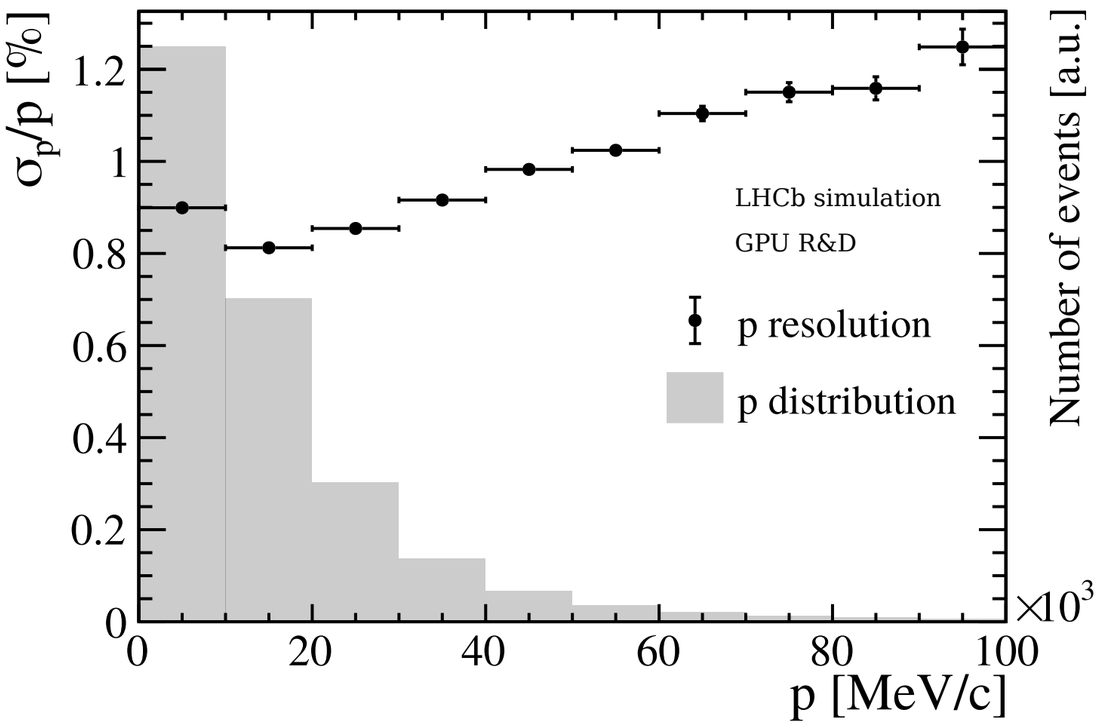
La reconstruction GPU est composée de...



Effectivement les memes elements que le x86 baseline, mais avec des algorithmes beaucoup plus parallèles.

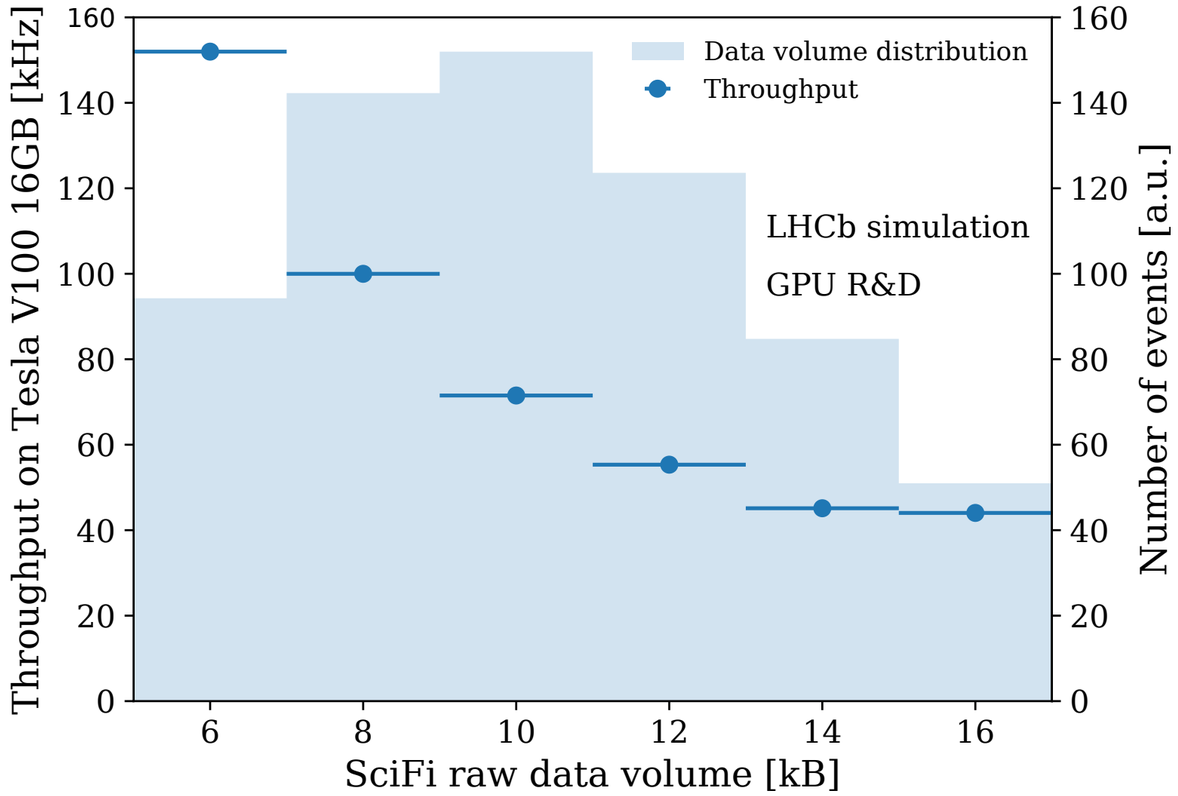
Physics performance

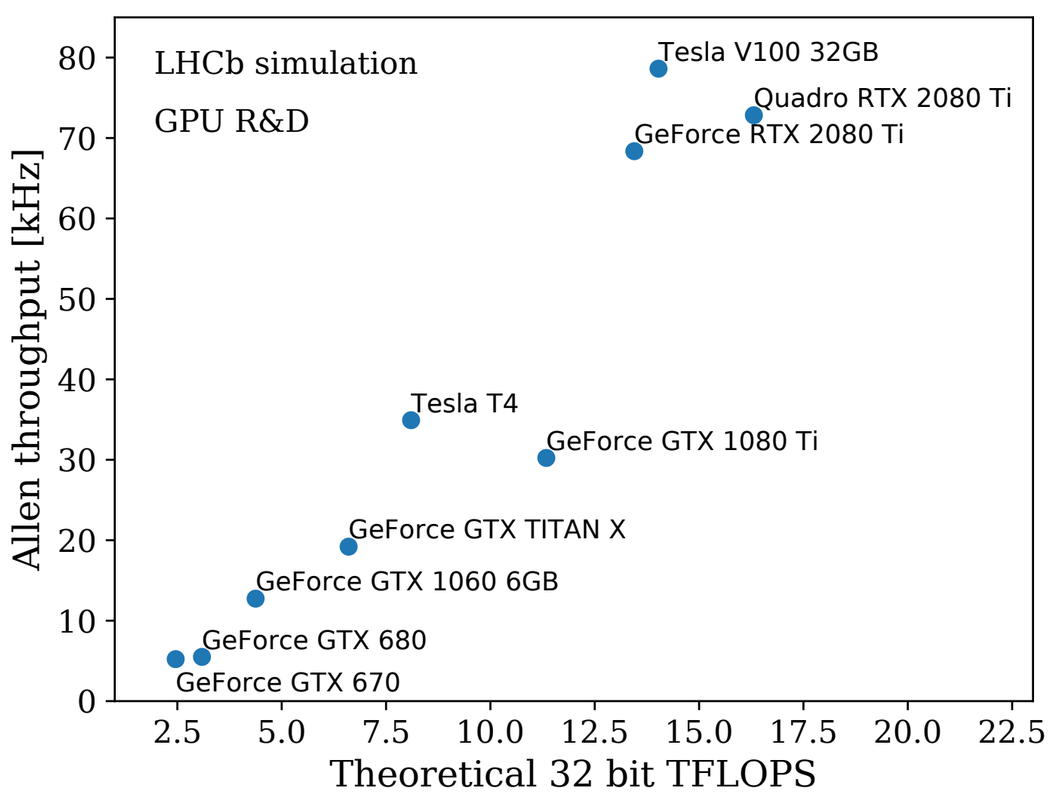




Meme qualité que le baseline

Evolution des performances GPU





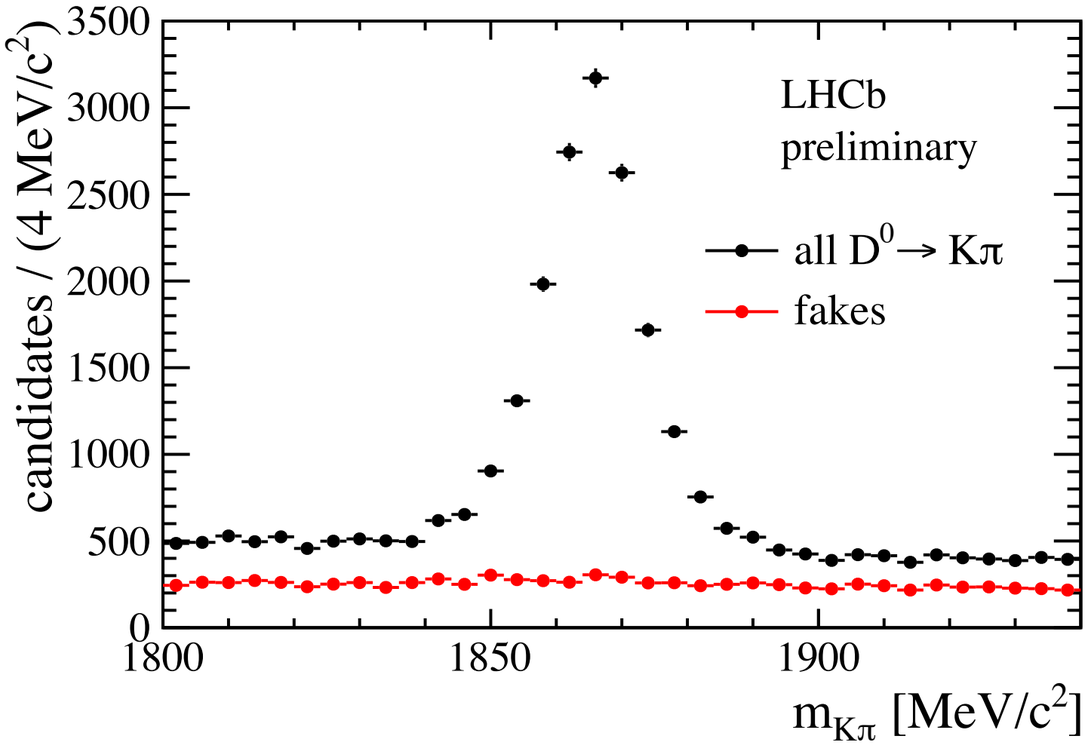
Evolution linéaire avec le detector occupancy! Relation quasi linéaire entre les TFLOPS théoriques et la performance.

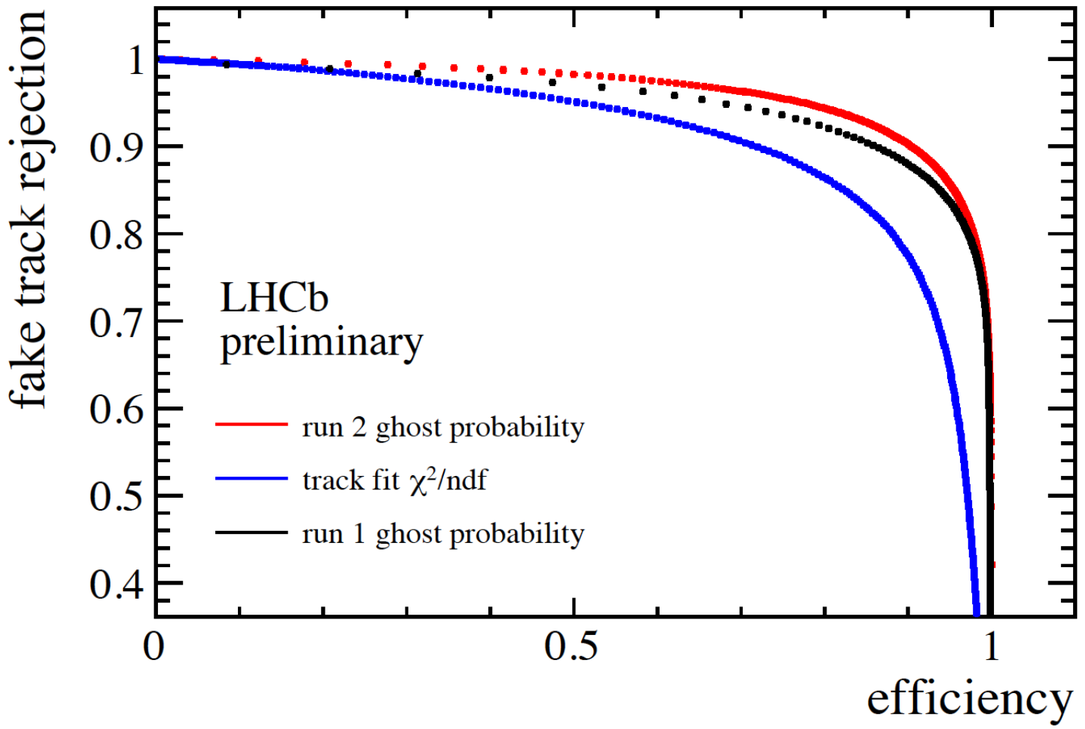
Utilisation du ML

LHCb reconstruction en temps reel et machine learning

Dans LHCb on utilisent du machine learning dans le système temps reel depuis 2011

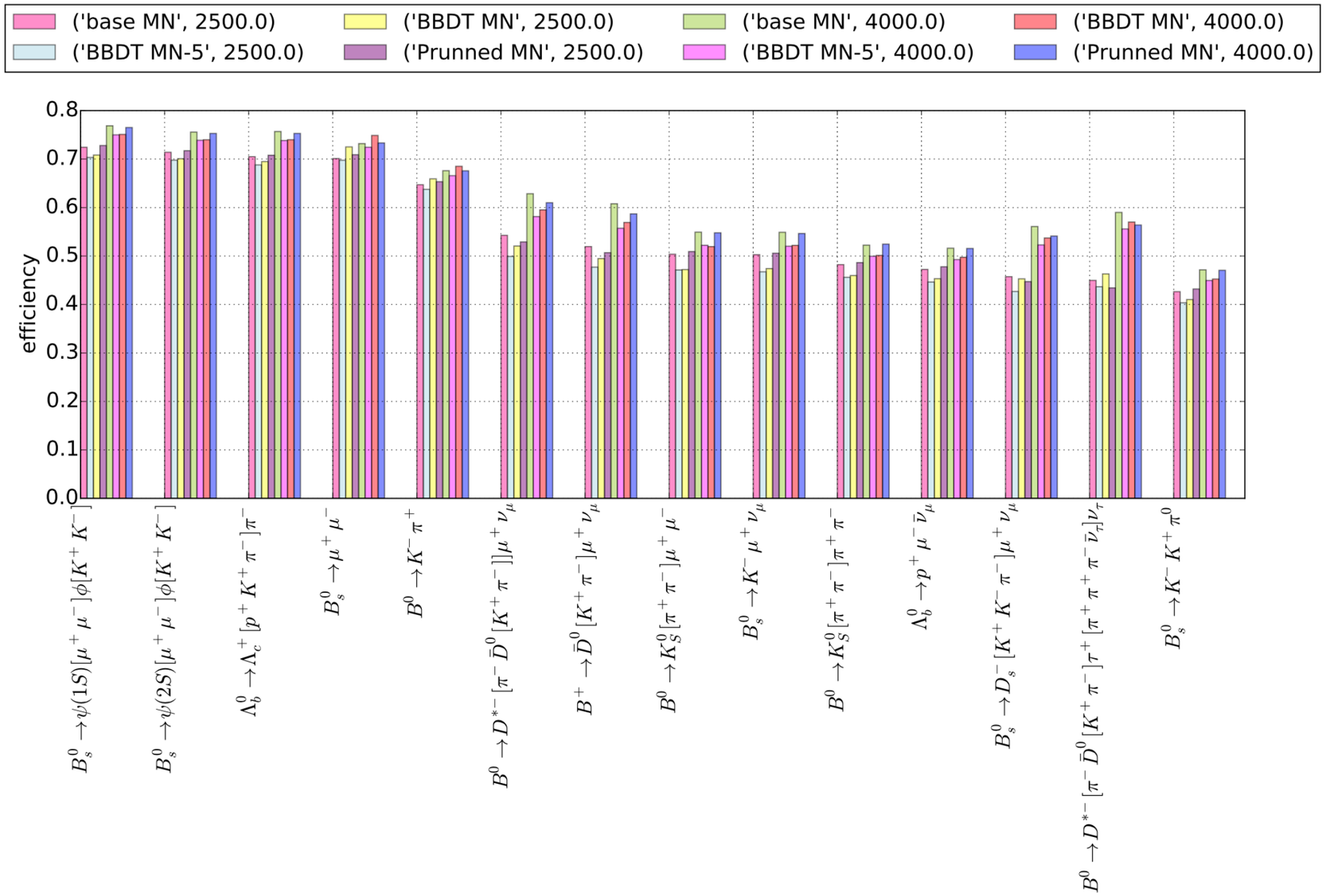
Très différentes architectures correspondent aux différentes problèmes: BDT, réseaux neurones, “deep” et “shallow”, customise pour une evaluation rapide. Architecture flexible nous permets une énorme liberté de choix meme en temps reel!





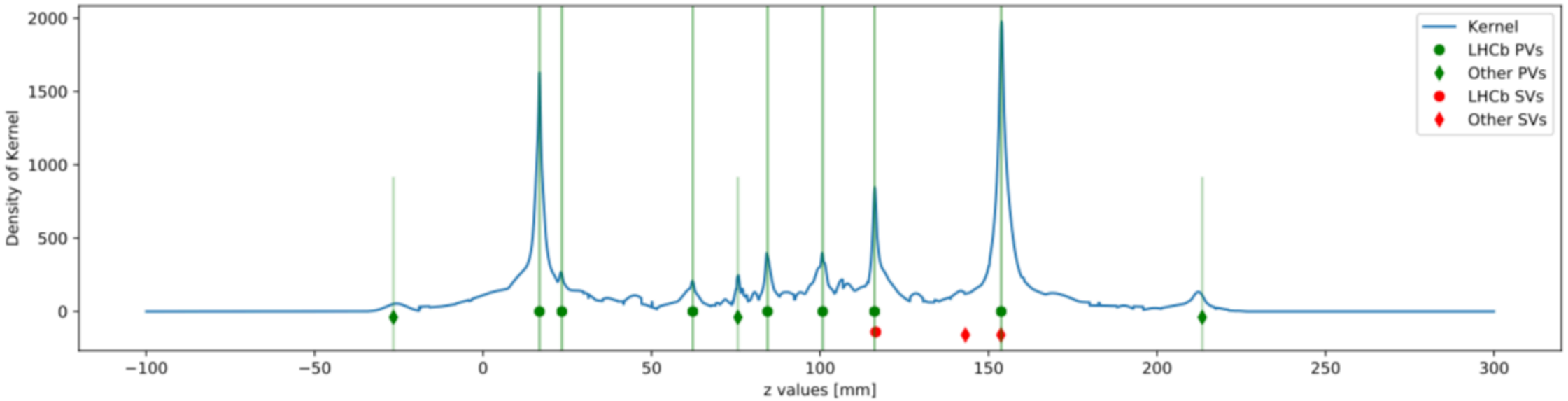
De Cian et al. LHCb-PUB-2017-011

ML pour classification au temps reel

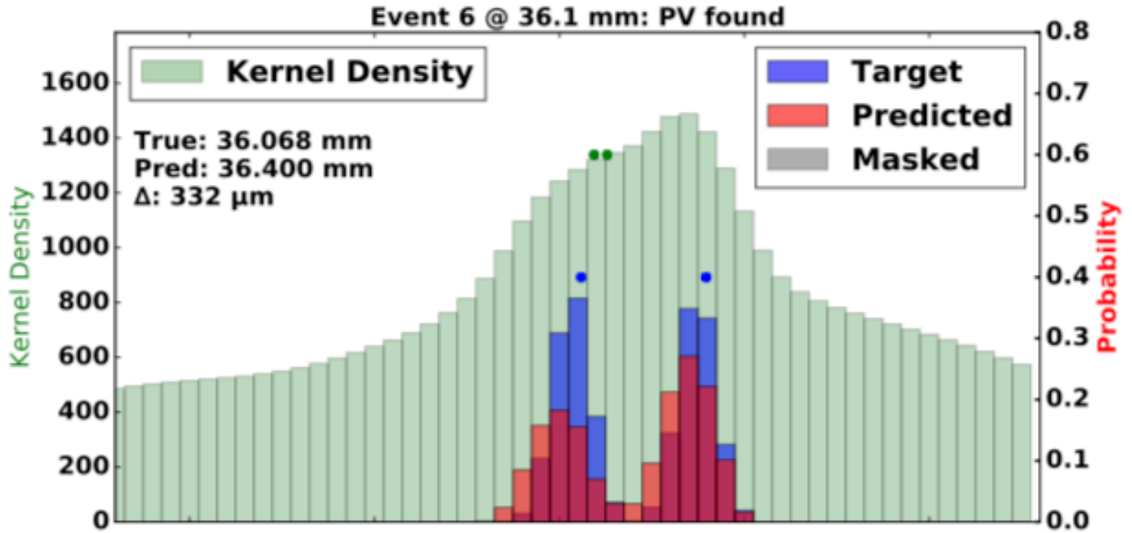


Likhomanenko et al. http://cds.cern.ch/record/2019813?ln=en

Example de réseau de neurones fait pour le upgrade



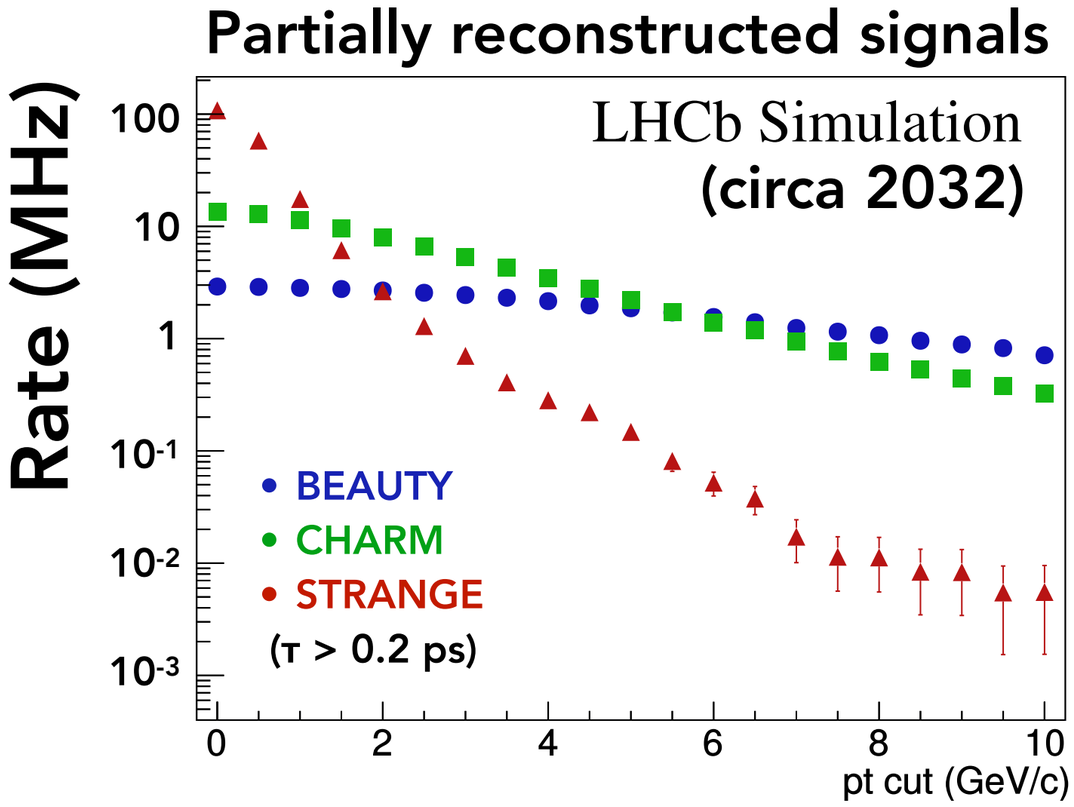
Trouver les collisions *pp* a partir des traces
4 convolutional layers
Performances encourageants mais sera-t-il plus rapide que l'algorithme classique? Il faut voir.

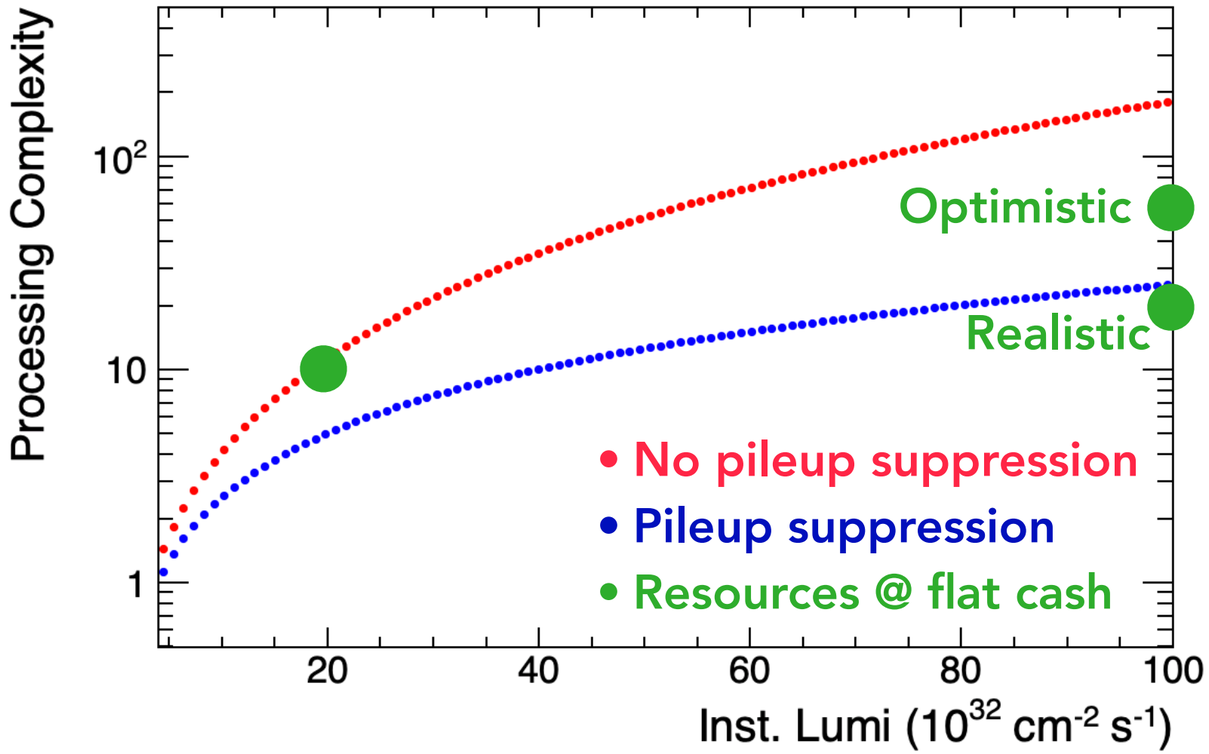


Slides and Paper available

En regardent vers l’avenir

Vers une deuxieme upgrade du LHCb a 10^{34} du luminosite?





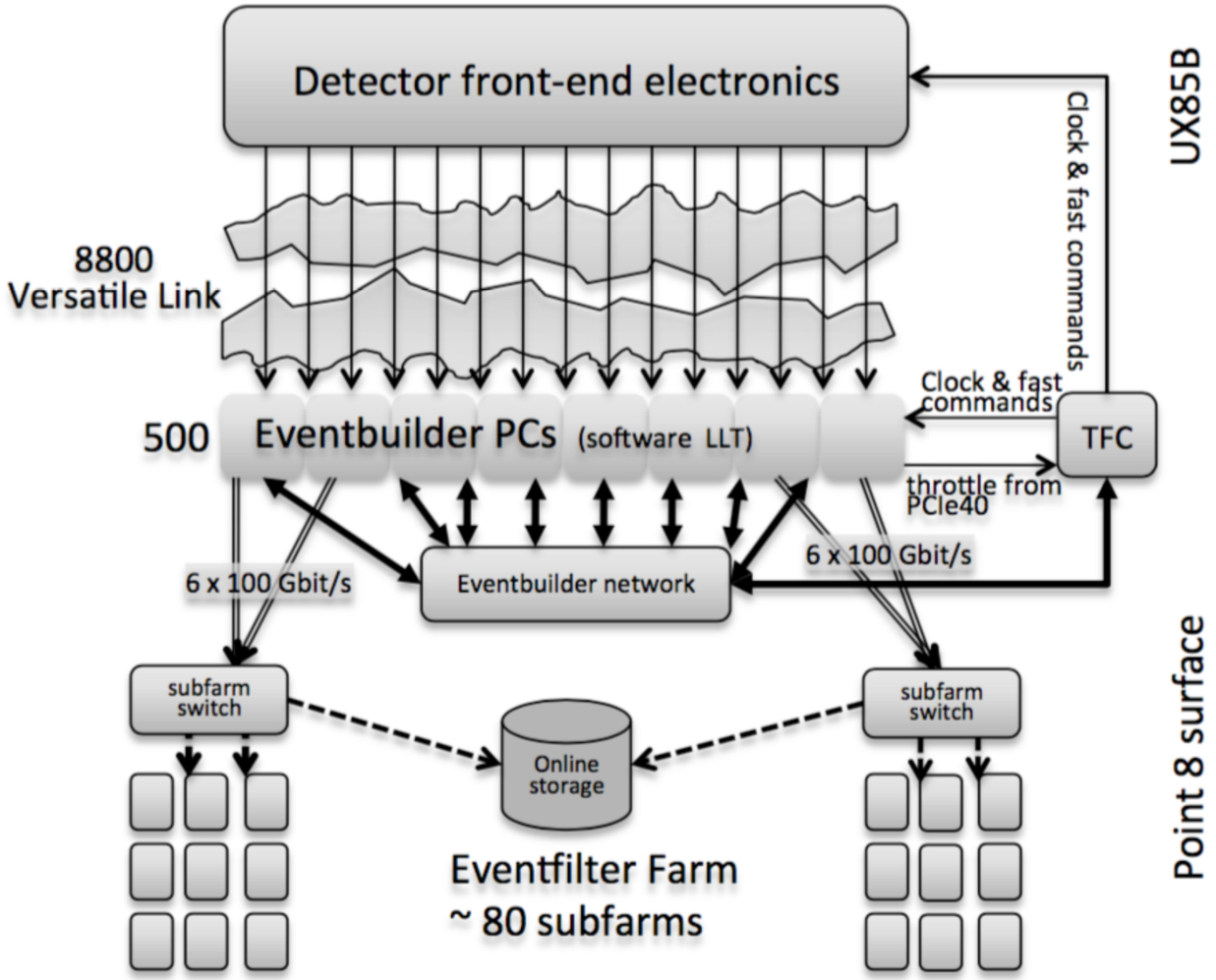
Faut il faire la reconstruction complete du détecteur a 30 MHz?

Maintenir la flexibilité de notre processing sera crucial

GBT link: 4.8 Gb/s Upgrade I
Evolution au 10 Gb/s pour HL-LHC mais il faut donc aller encore 5 fois plus loin pour pas exploser le cout

Event-building: passer au 200 Gb/s dans les années prochaines et suivre l'evolution technologique après ca

Centre de calcul: quelle technologie sera la meilleure? Hybride ou non? Proche de l'experience ou loin?



Il faut developer et maintenir des competences techniques au travers des archis hybrides si on veut être prêt pour l'avenir, et nous sommes parti dans le bon sens.

Et pour finir, petit point de réflexion

LHCb 2032

>1000 Eb/year

ATLAS+CMS 2027

260 Eb/year

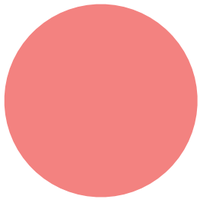
Square Kilometre Array (2030s)

~30000 Eb/year

Sequence genome of all humans on Earth

8000 Eb

Global internet dataflow 2021



2800 Eb/year

Backup

LHCb analysis methodology and role of calibration samples

Trigger Efficiency

Tag-and-probe calibration method exists & widely used

Tracking efficiency

Tag-and-probe

Existing	Developing
μ	e,π,K,p

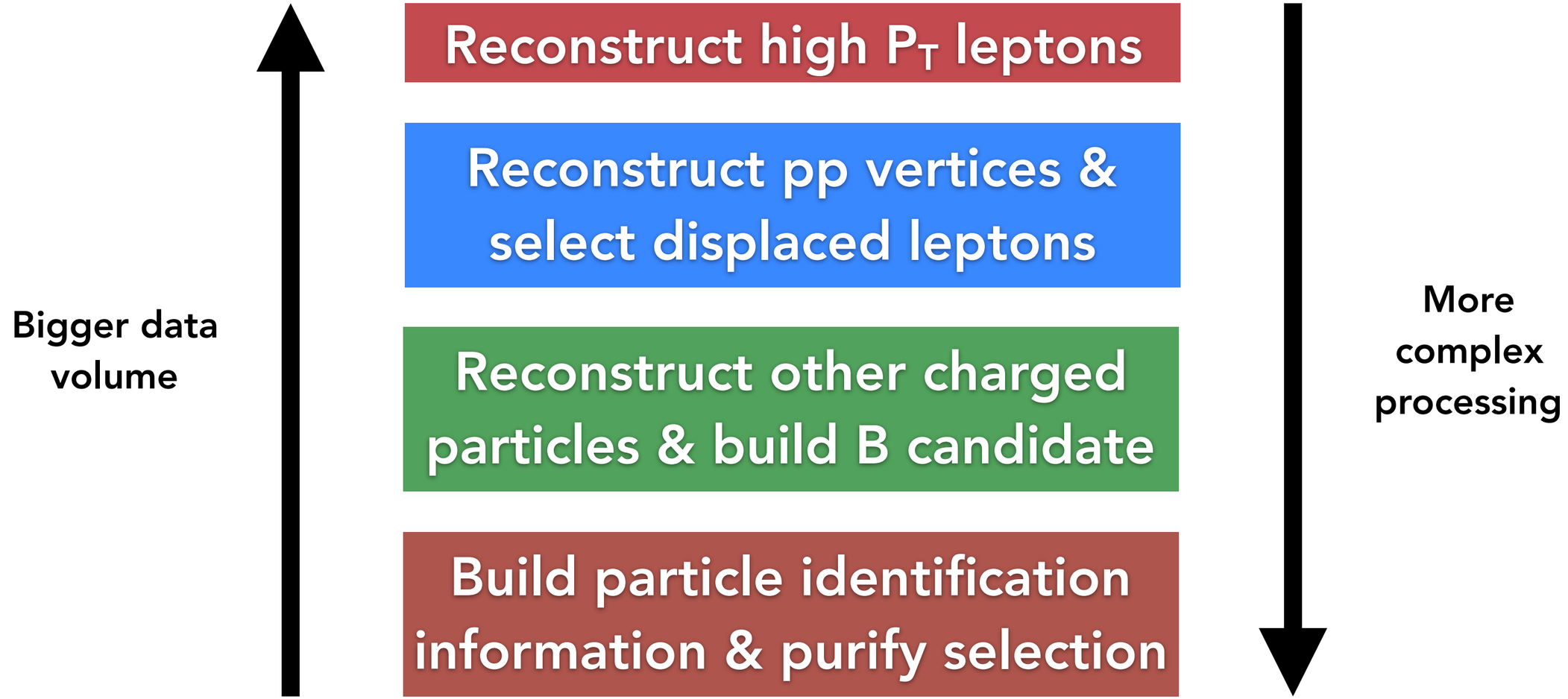
Particle identification

Tag-and-probe

Tag-and-probe calibrations exist for all charged particle species and for π^0/γ, with new sources added over time to improve coverage

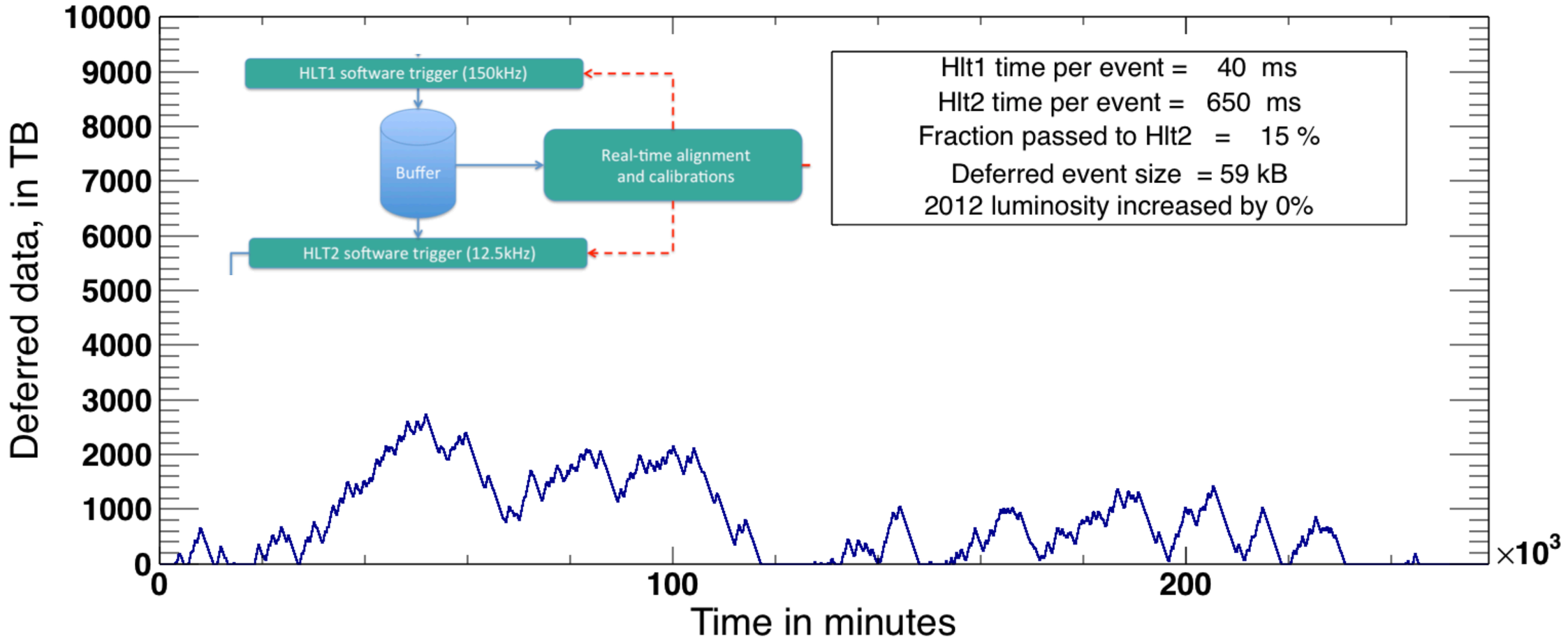
Data driven efficiency calibration key to precision physics

What is a cascade buffer?



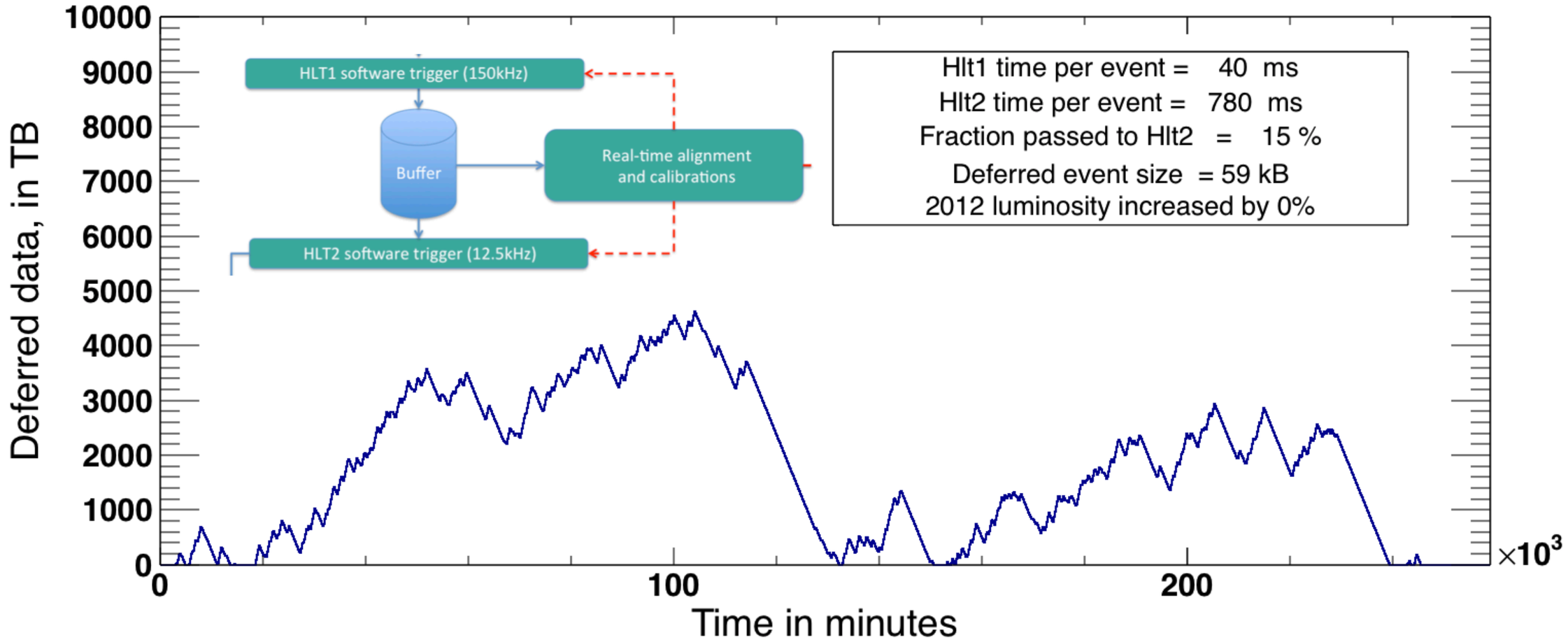
A staged data reduction using increasingly complex algorithms

Optimization of the Run 2 LHCb cascade buffer



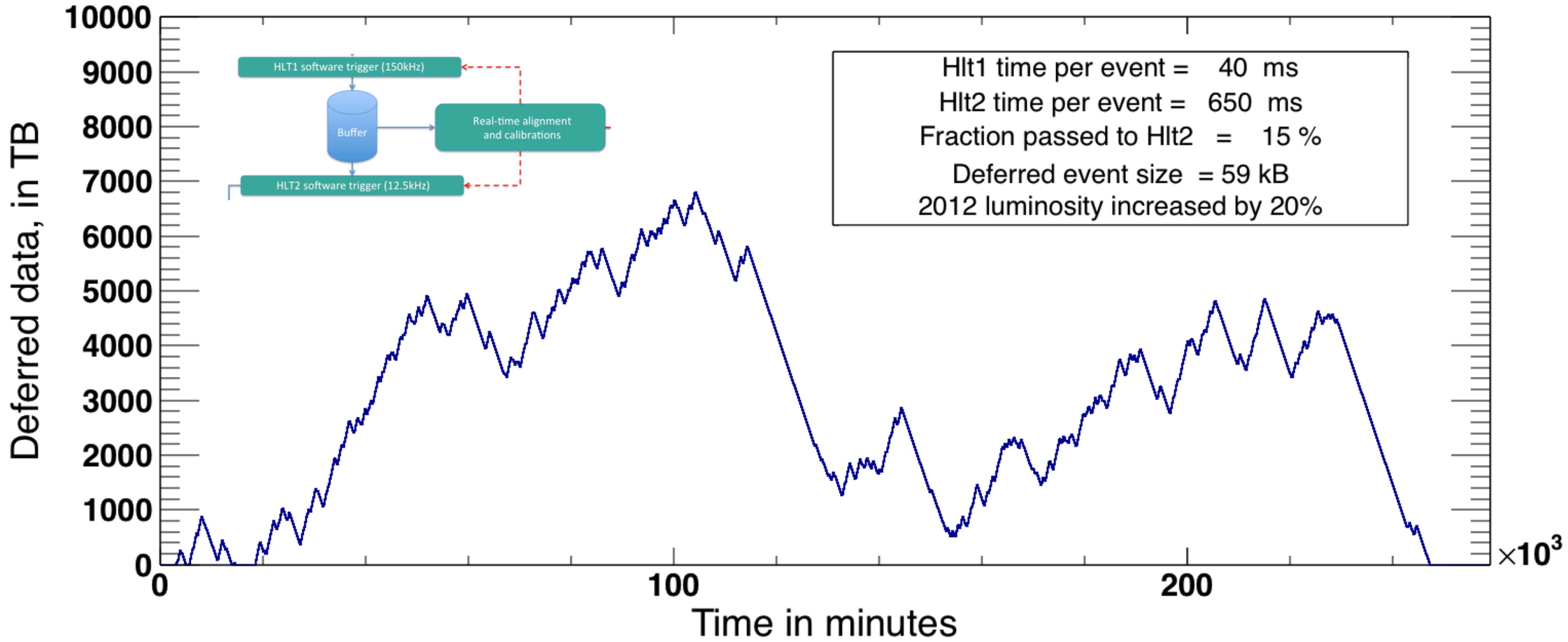
Use Run I LHC fill structure to simulate disk buffer usage

Optimization of the Run 2 LHCb cascade buffer



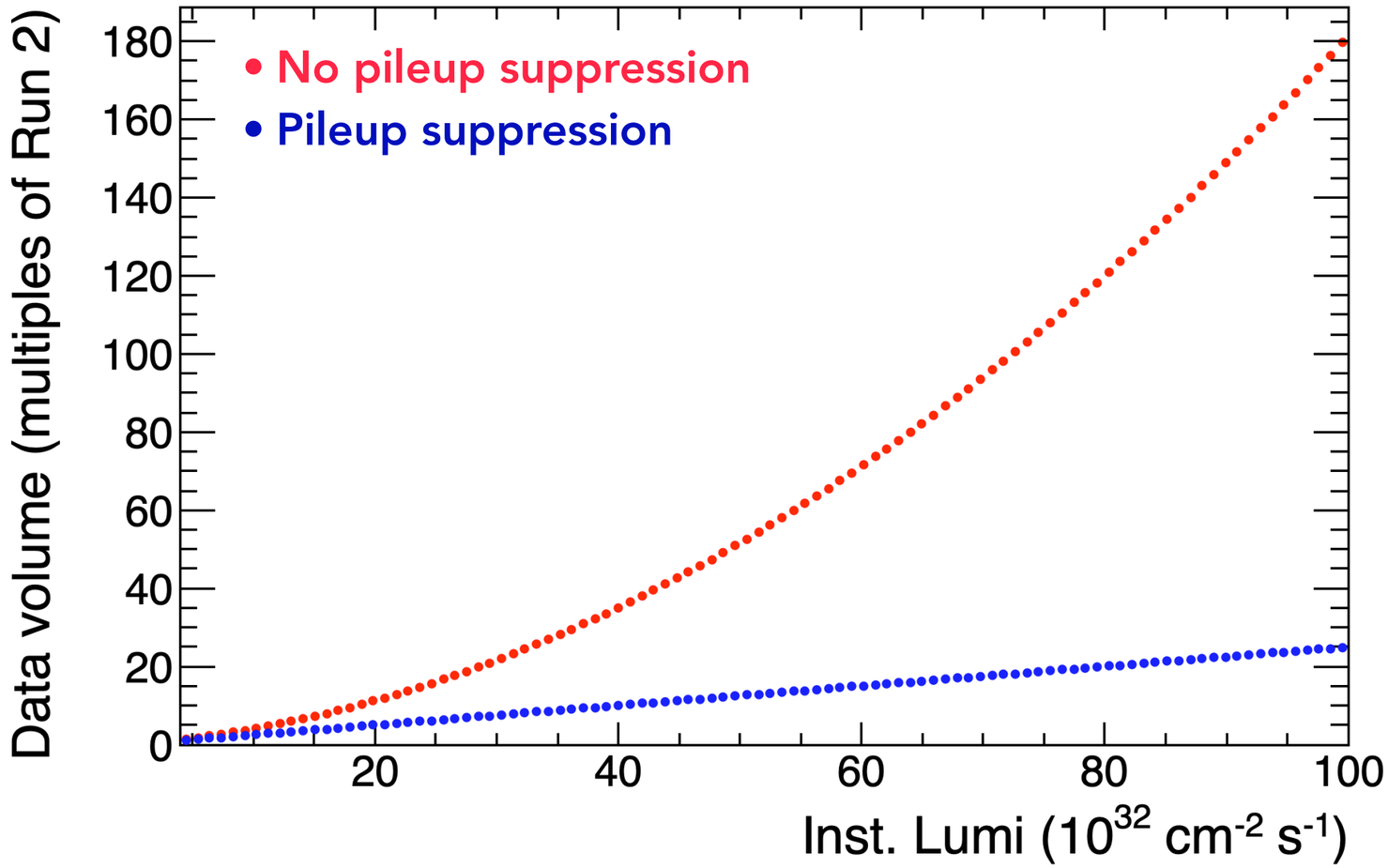
Use simulation to ensure robustness if timing estimates wrong

Optimization of the Run 2 LHCb cascade buffer



Use simulation to ensure robustness if LHC overperformed

And what about data volumes?



Data volume increases quadratically even with 0 background.
Select pp collisions, not bunch crossings, in real time!