



Acquisition de données

Exemple de l'expérience LHCb et prospective

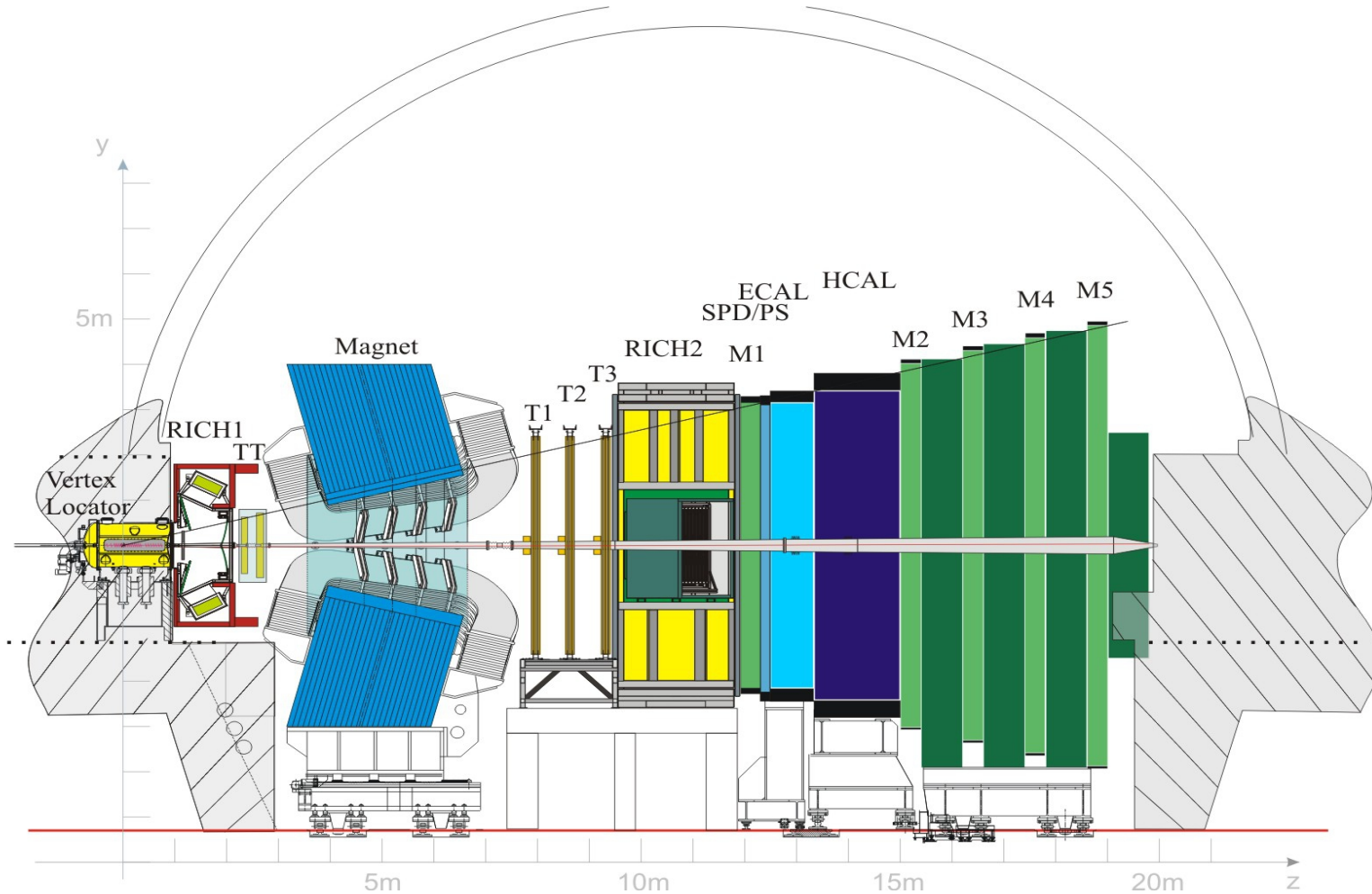


J.P. Cachemiche
Centre de Physique des Particules de Marseille

Plan

- **Présentation du trigger LHCb**
- **Démarche de réalisation du trigger à muon**
- **Techniques de base pour le parallélisme**
- **Evolution de l'architecture d'acquisition LHCb pour 2019**

Le détecteur LHCb



Etude des asymétries matière/anti-matière dans la physique du méson B

Fonctions du système

- Lecture d'environ 1 million de canaux
- Production de 100 000 paires $b\bar{b}$ par secondes
- Filtrage des événements non intéressants
- Acquisition
- Identification
- Stockage (quelques kHz)

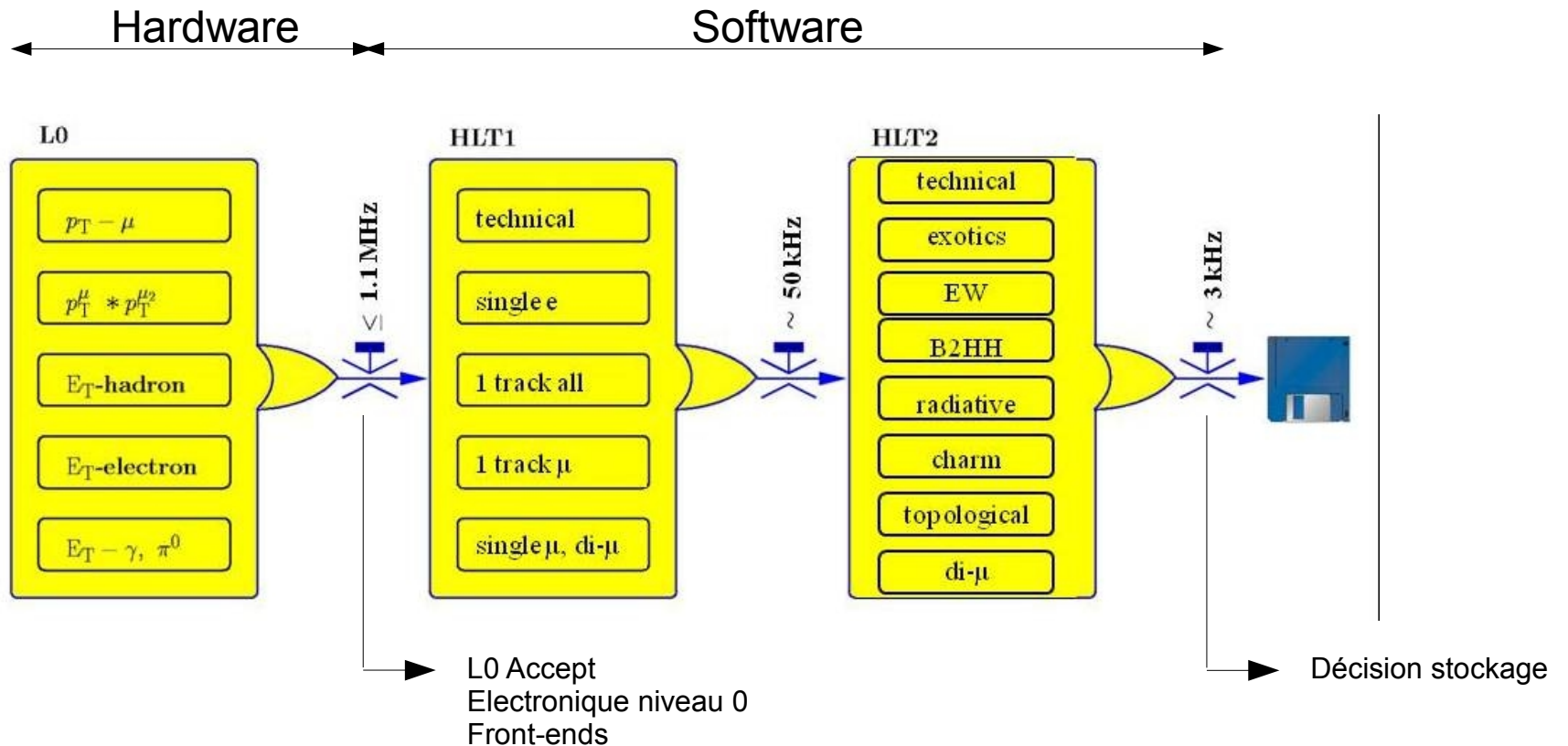
Decay Modes	Visible Br. fraction	Offline Reconstr.
$B_d^0 \rightarrow \pi^+ \pi^- + \text{tag}$	0.7×10^{-5}	6.9 k
$B_d^0 \rightarrow K^+ \pi^-$	1.5×10^{-5}	33 k
$B_d^0 \rightarrow \rho^+ \pi^- + \text{tag}$	1.8×10^{-5}	551
$B_d^0 \rightarrow J/\psi K_S + \text{tag}$	3.6×10^{-5}	56 k
$B_d^0 \rightarrow \bar{D}^0 K^{*0}$	3.3×10^{-7}	337
$B_d^0 \rightarrow K^{*0} \gamma$	3.2×10^{-5}	26 k
$B_s^0 \rightarrow D_s^- \pi^+ + \text{tag}$	1.2×10^{-4}	35 k
$B_s^0 \rightarrow D_s^- K^+ + \text{tag}$	8.1×10^{-6}	2.1 k
$B_s^0 \rightarrow J/\psi \phi + \text{tag}$	5.4×10^{-5}	44 k

Expected numbers of events reconstructed offline in one year (10' s of data taking) with an average luminosity of $2 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, for some channels.

Trigger

Filtrage des événements

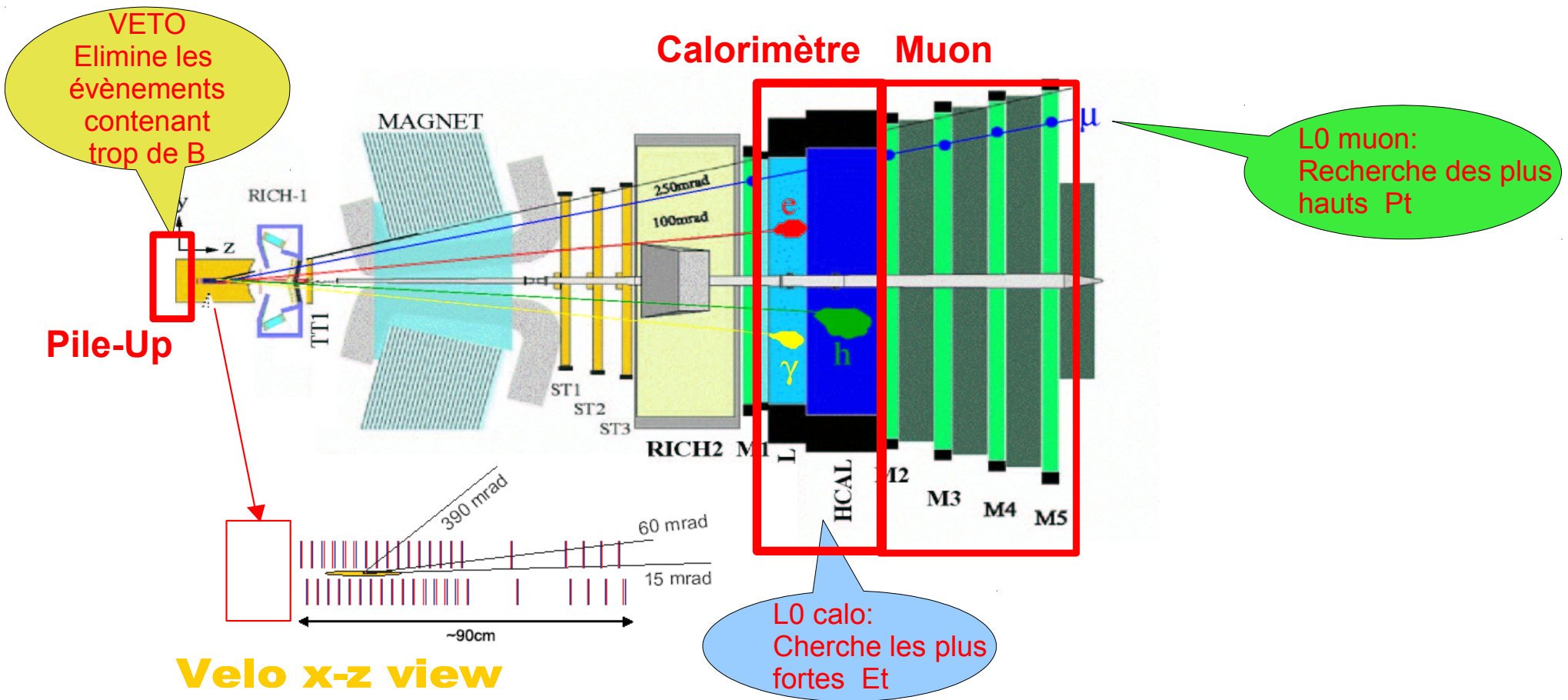
3 étapes de réduction :



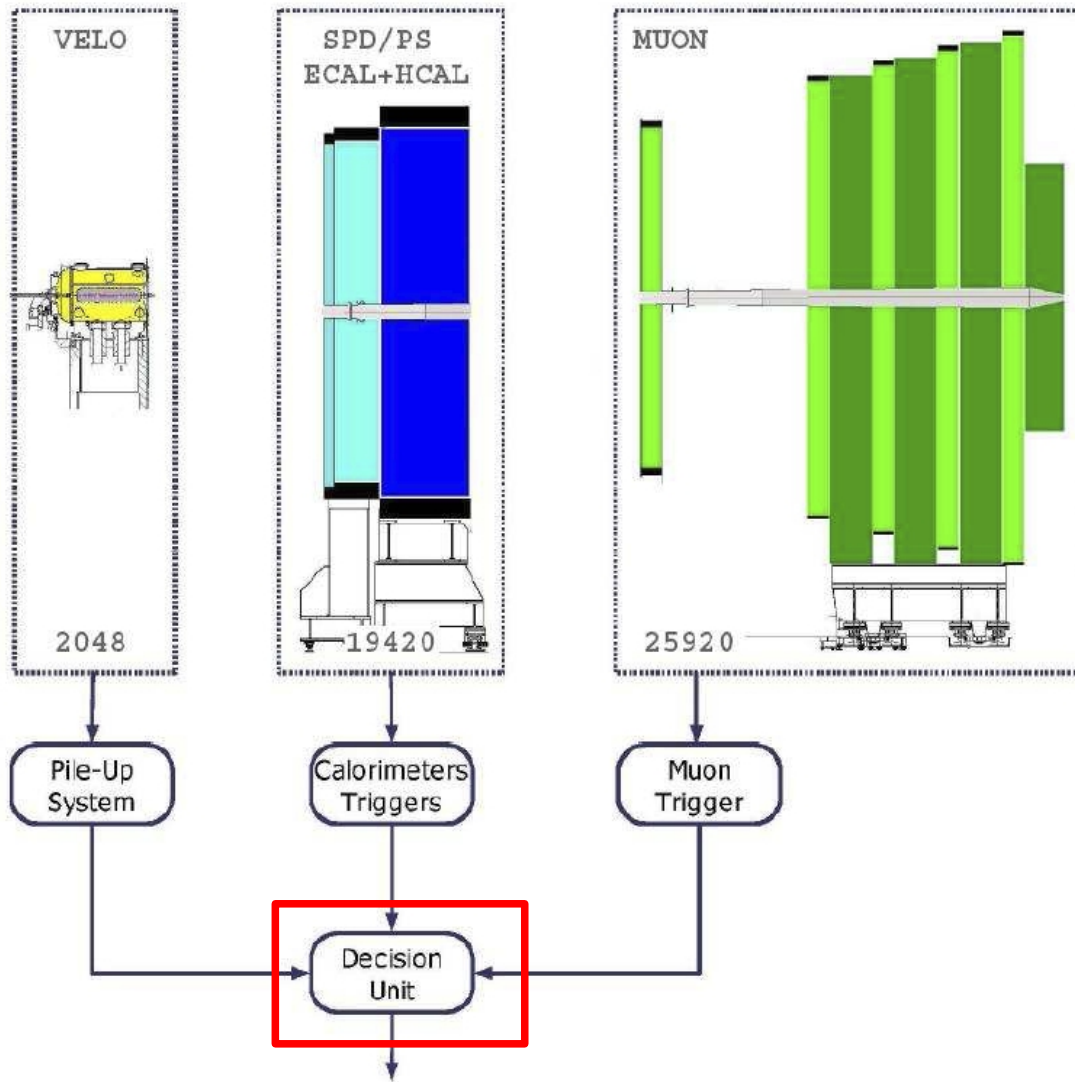
Sous-détecteurs participant au trigger de niveau 0

3 détecteurs

- Calorimètres, muons et pile-up veto

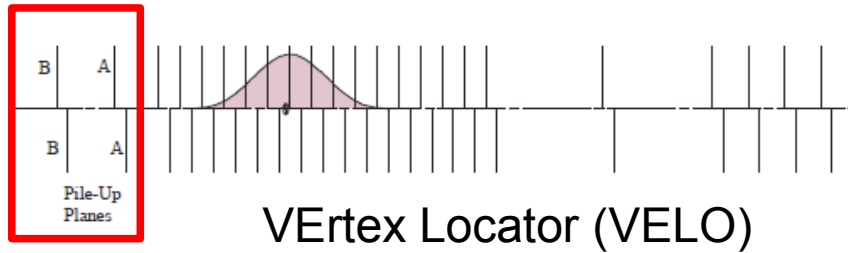


Unité de décision



Trigger final réalisé par l'unité de décision

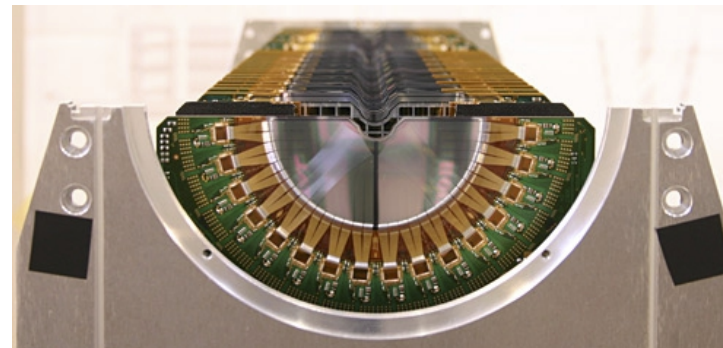
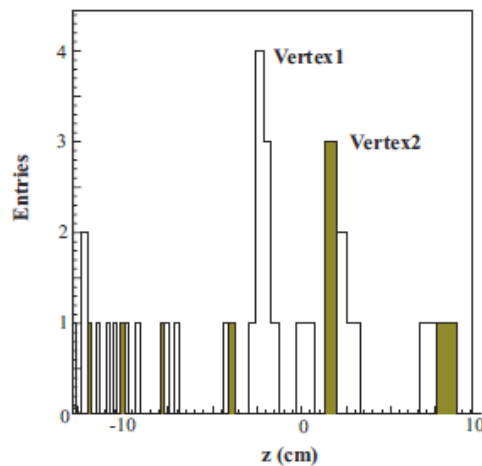
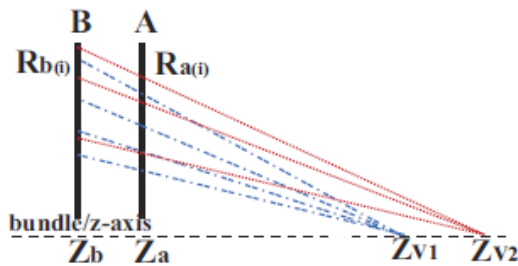
Pile-up veto



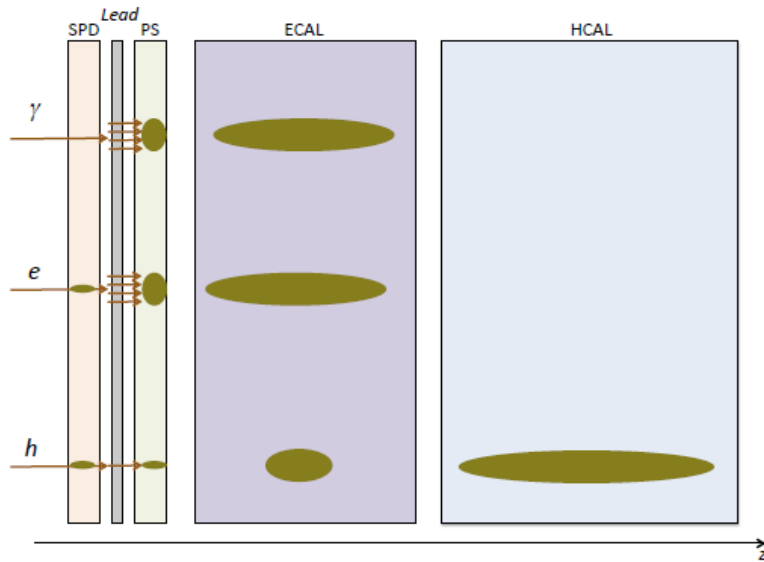
Détection des croisements contenant trop d'interactions

→ Événements trop difficiles à analyser

- Détection de tous les vertex déterminés par les hits des plans A et B
- Elimination des hits correspondant aux 2 vertex de plus grande énergie
- S'il reste un ou plusieurs vertex, élimination de l'événement (VETO)



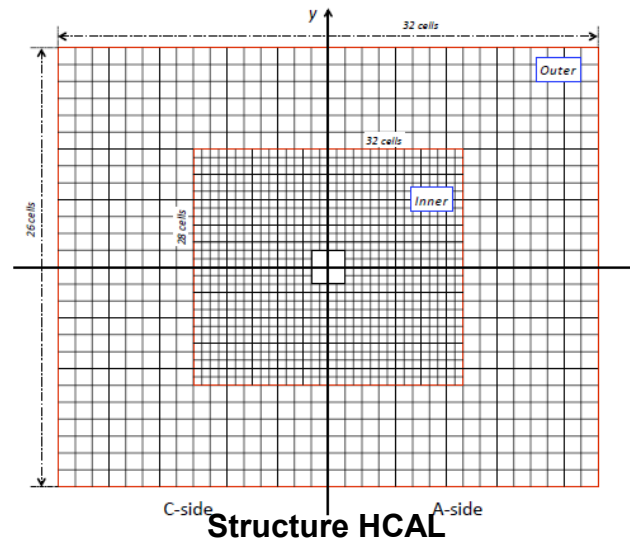
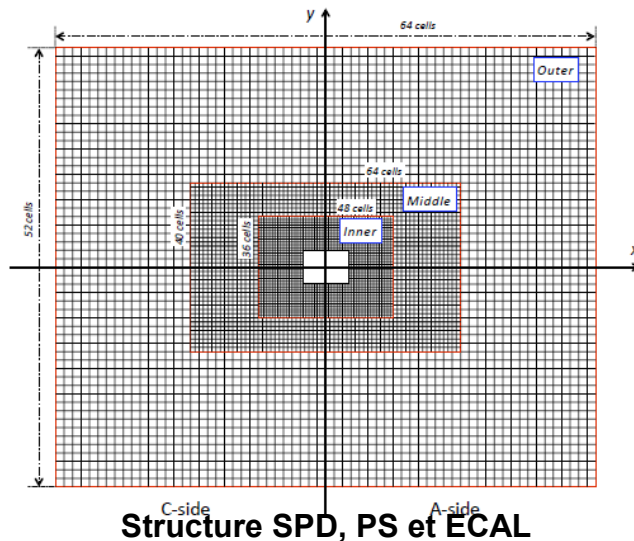
Calorimètres



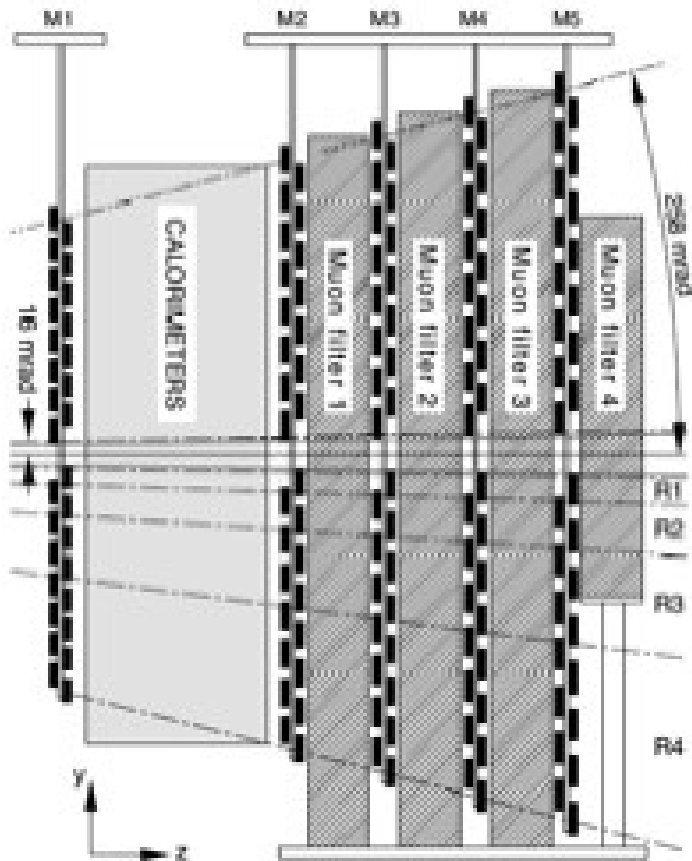
Détection des particules avec une E_T élevée

Plusieurs sous-systèmes :

- **SPD** (Scintillator Pad detector)
 - Identifie les particules chargées et sépare les électrons des photons
- **PreShower** (détecteur de pied de gerbes)
 - Identifie les électrons et photons
- **Calorimètre Electro-magnétique**
 - Mesure l'énergie des électrons et photons
- **Calorimètre Hadronique**
 - Mesure l'énergie des hadrons



Trigger à muon

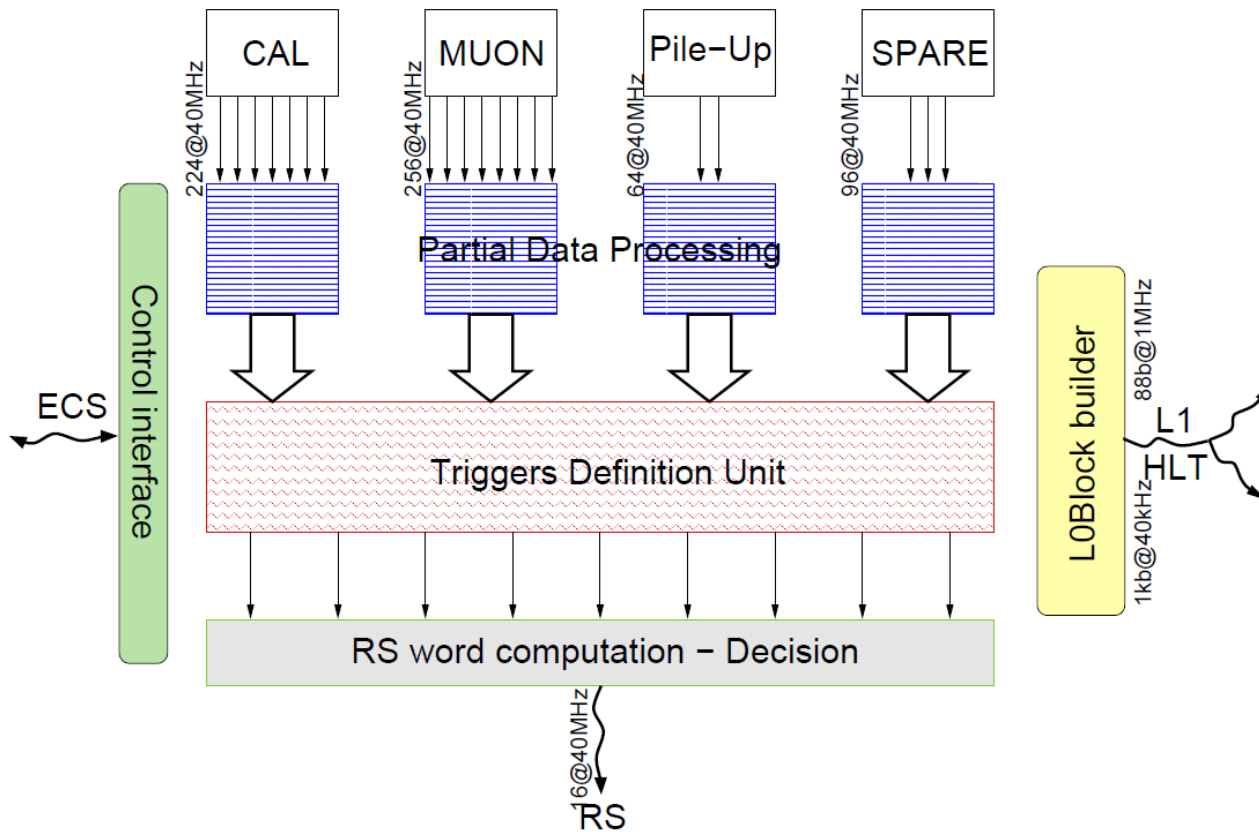


Détection des muons avec une impulsion transverse (P_t) élevée

- 1400 GEM et MWPC répartis sur 5 plans
- 120000 canaux
- 435 m²
- 2.5 millions de cables

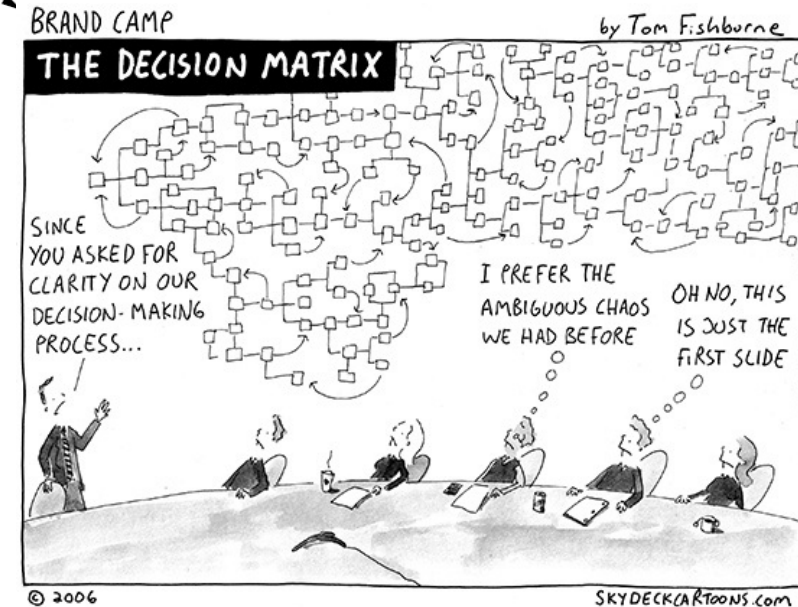


Unité de décision

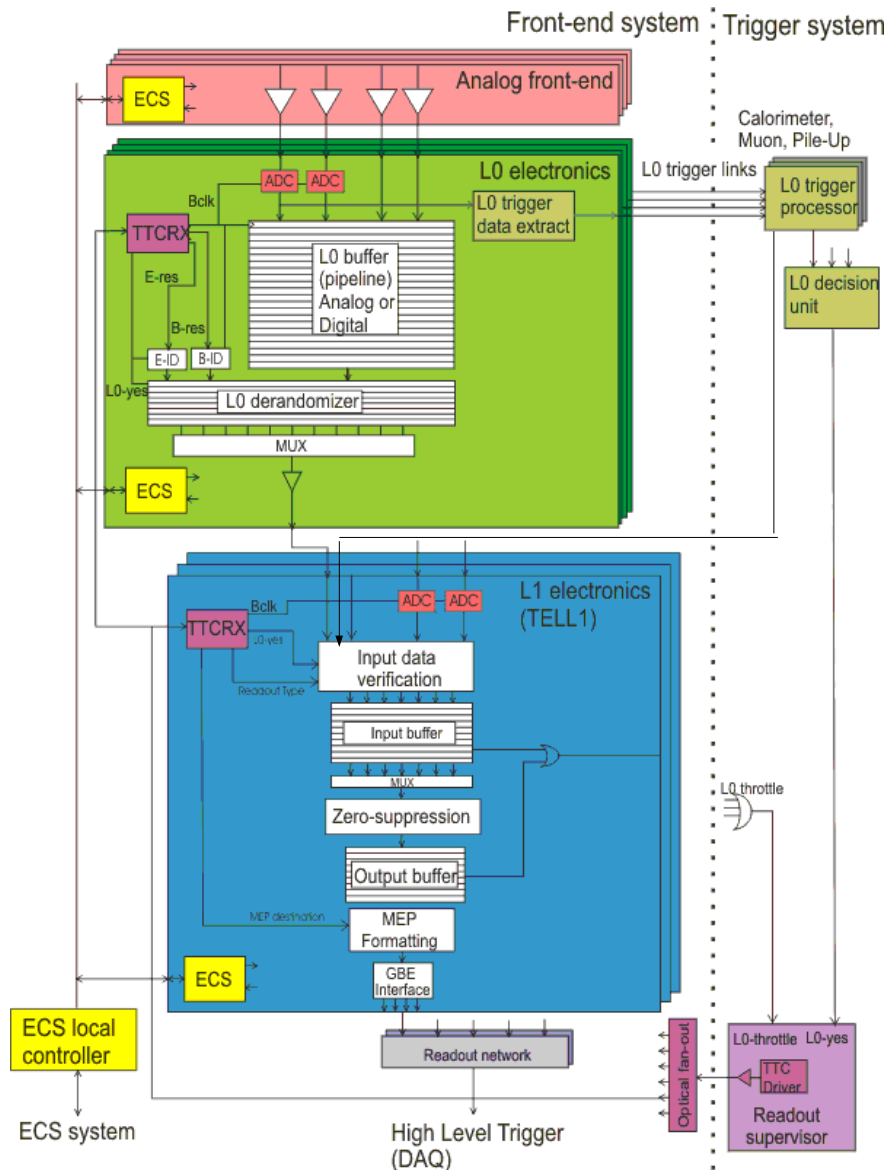


Opérations

- Mise en temps
- Algorithme trigger global
- Envoi décisions au TFC supervisor
- Envoi décision au readout ainsi qu'au HLT



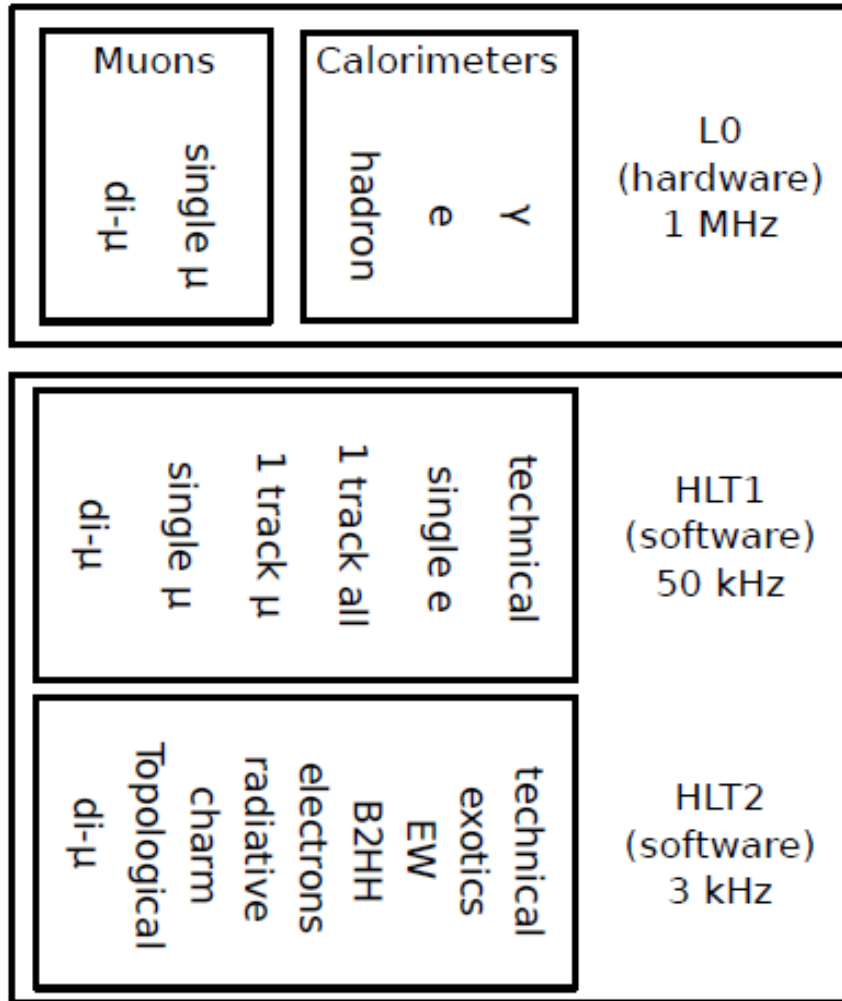
Carte de readout TELL1



Fonctionnalités

- Interface optique ou analogique avec cartes Front-Ends
- **Traitements sur mesure**
- **Compression des données**
- **Bufferisation**
- Regroupements et formation de MEPs (Multi Event Packets)
- Envoi des événements vers les fermes

HLT



HLT1

- Confirmation des résultats du trigger hardware en associant les traces du calorimètre et du muon avec celles du VELO et du Tracker
 - Reçoit les données du VELO et reconstruit les vertex primaires
 - Utilise les données du tracker pour mesurer le P et le P_T des traces correspondantes
 - Élimination par seuillage
- Réduit le débit de 1 MHz à ~50 kHz

HLT2

- Analyse et identification de la totalité des événement
- Débit sortant 3 à 5kHz

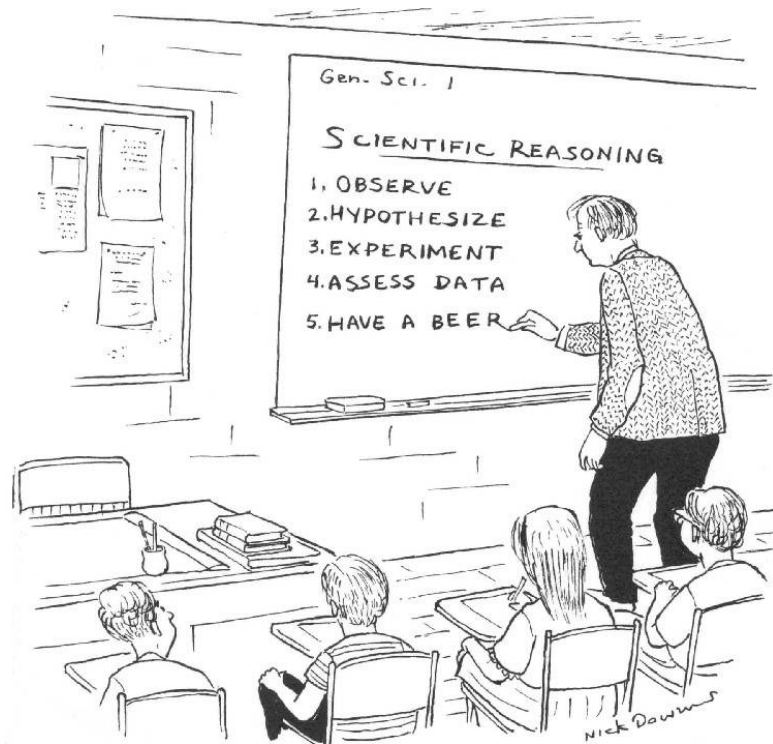
HLT



Dimensionnement HLT

- 1820 CPUs 24 ou 32 cœurs

	CPU type	Moore/s	number of physical cores	number of logical cores	total RAM	local harddisk space	number of units
DELL C6100	Intel x5650	650	12	24	24	2 TB	520
Action Solar 820 S4	Intel x5650	630	12	24	24	2 TB	400
ASUS RS720QA-E6-RS12	AMD 6272	680	32	32	32	2 TB	100
Intel S2600KP	Intel E5-2630v3	980	16	32	32	4 TB	800

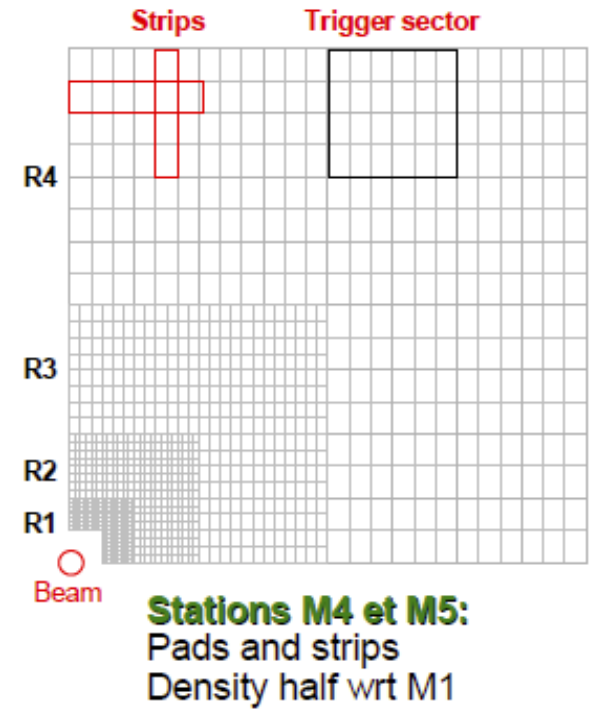
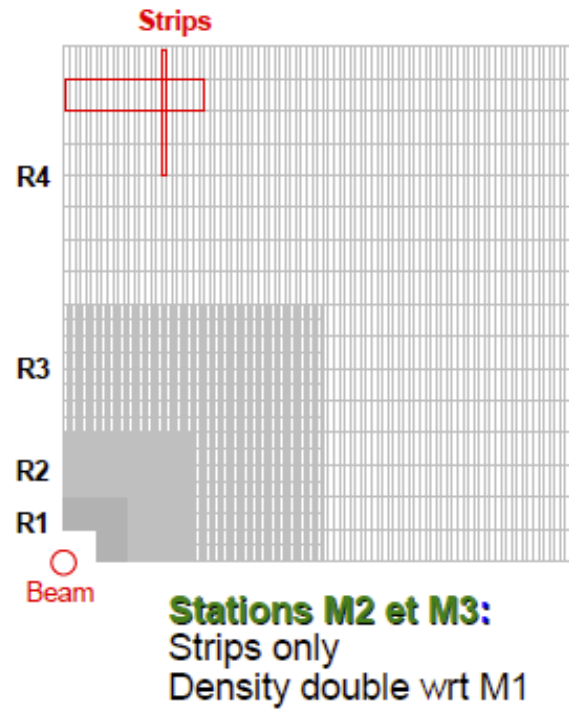
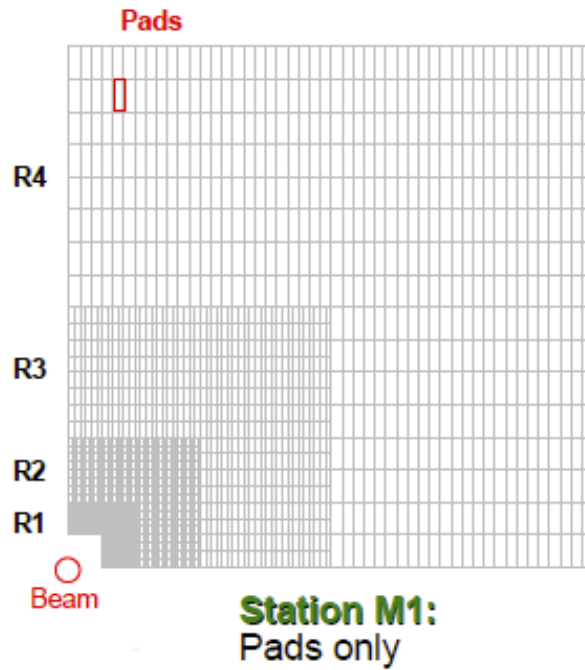
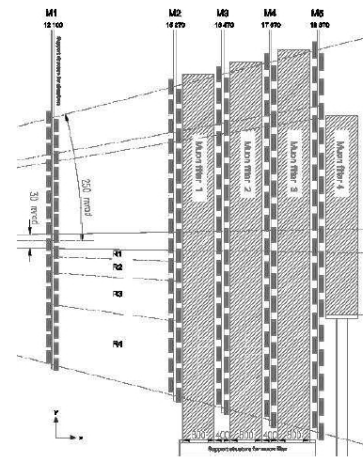


Démarche de réalisation du trigger à muon

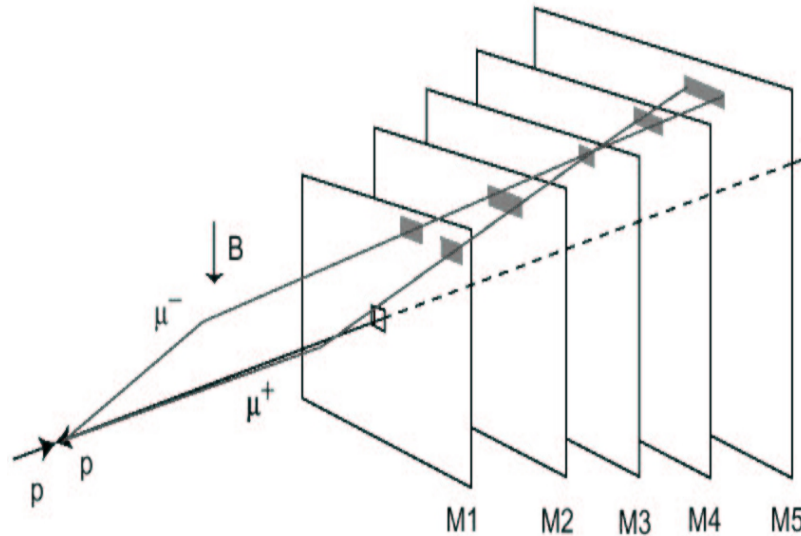
Chambres à muons

5 stations, 4 régions par station :

- Chambres à fils
 - Détection de muons à P_T élevé



Recherche des candidats



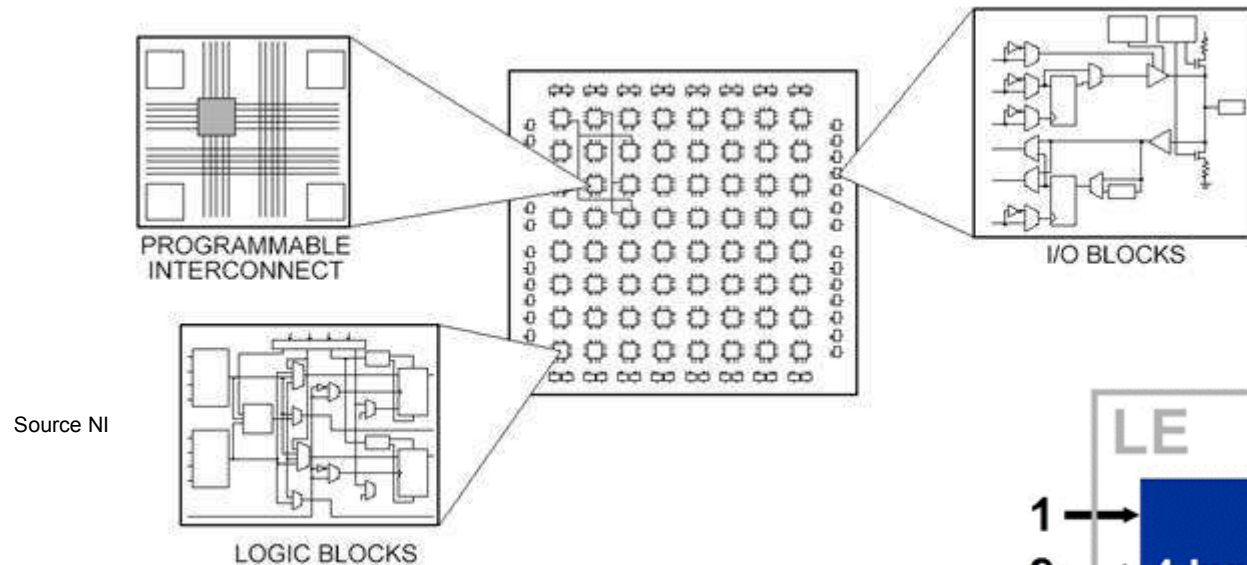
Principe de l'algorithme:

- 1- Trouver un pad touché en M3
- 2- Définir un axe de recherche centré sur le PAD
- 3- Ouvrir 2 cones le long de cet axe
- 4- Sélectionner une trace si un pad est touché dans le couloir dans les plans M5 et M4 et M2
- 5- Le point de passage en M1 est extrapolé en suivant la droite partant de M3 et passant par le pad touché de M2
- 6- Recherche d'un hit dans la zone extrapolée
- 7- Ce point dans M1 donne l'angle de la trace par rapport au faisceau donc P_T (impulsion transverse)

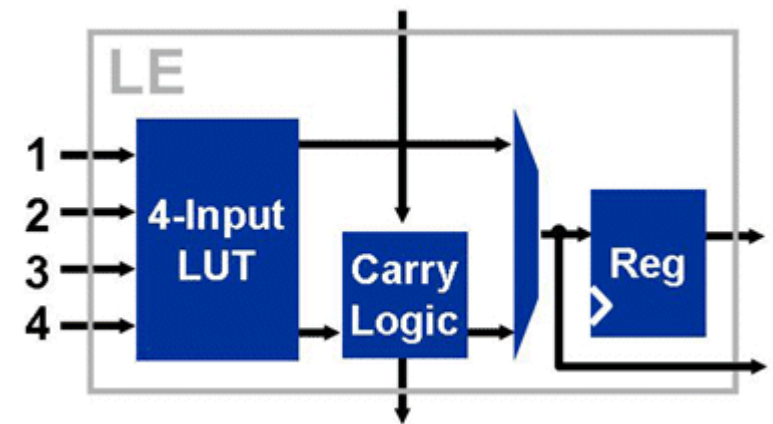
Unité de traitement : un FPGA

Field Programmable Gate Array

- **Matrice de cellules logiques** interconnectables de façon programmable



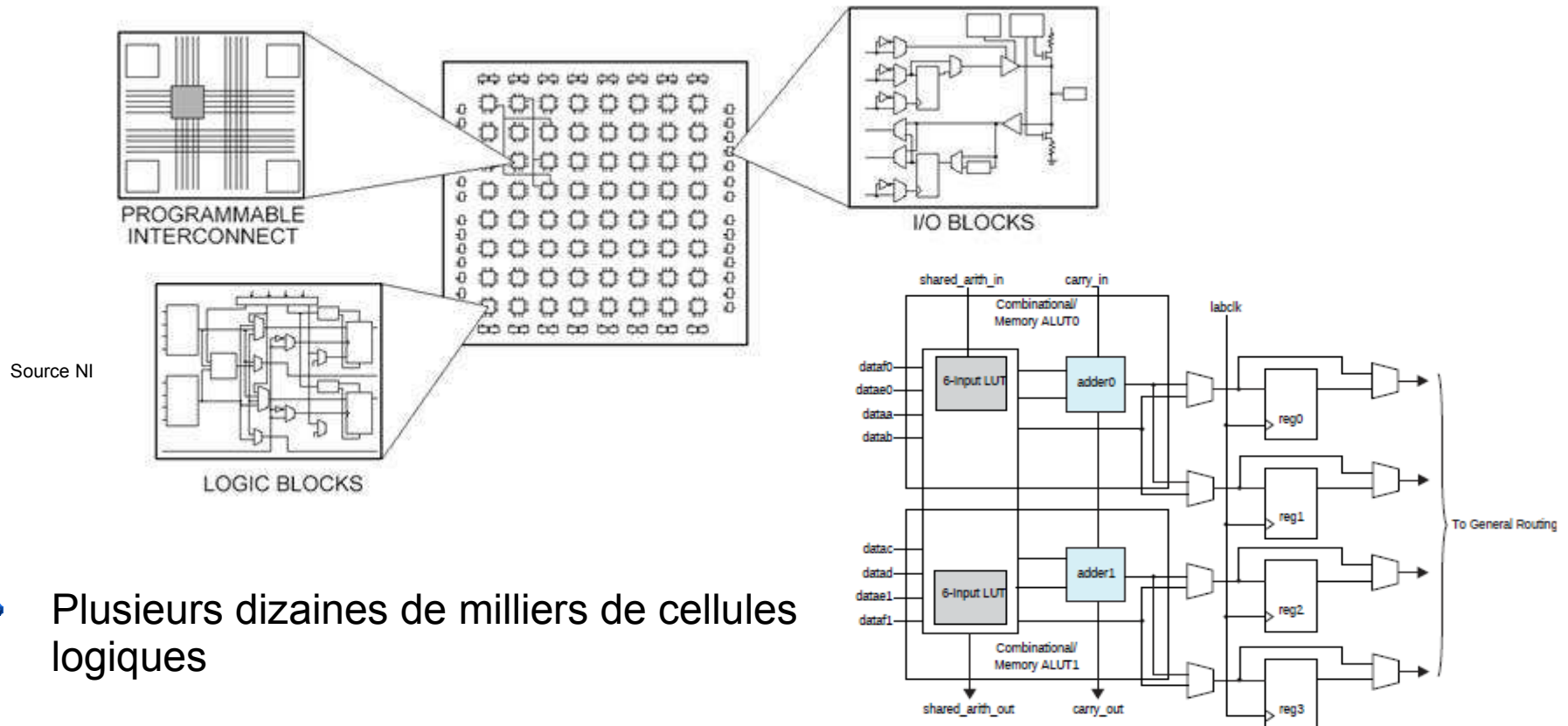
- Plusieurs dizaines de milliers de cellules logiques



Unité de traitement : un FPGA

Field Programmable Gate Array

- **Matrice de cellules logiques** interconnectables de façon programmable



- Plusieurs dizaines de milliers de cellules logiques

Unité de traitement : un FPGA

Entrées sorties programmables

- LVDS, CML, HSCL, CMOS, SSTL,
- Avec des fonctions de filtrage : préaccentuation, égalisation

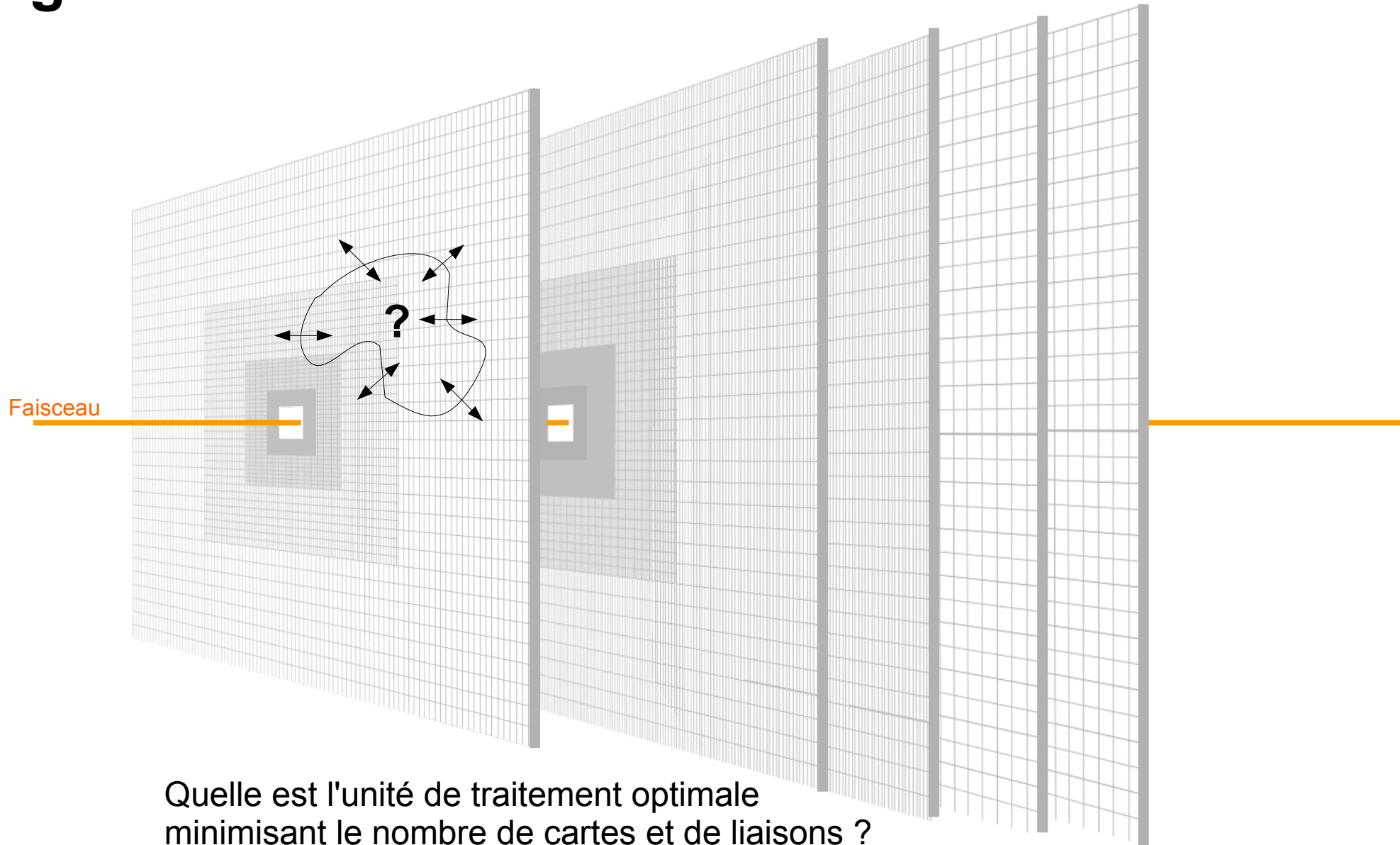
Contient également des structures câblées

- Mémoires
- PLLs
- Cellules DSP
- Sérialiseurs/désérialiseurs multigigabits
- Hardware IP blocks
 - interfaces mémoires : DDR3, DDR, QDR, ...
 - Interfaces protocoles de communication : PCIe, GbE, Interlaken, ...
- Hardware CPU : ARM
- Convertisseurs Analogiques Digitaux

Peut contenir des fonctions autrefois dédiées aux instruments de mesure

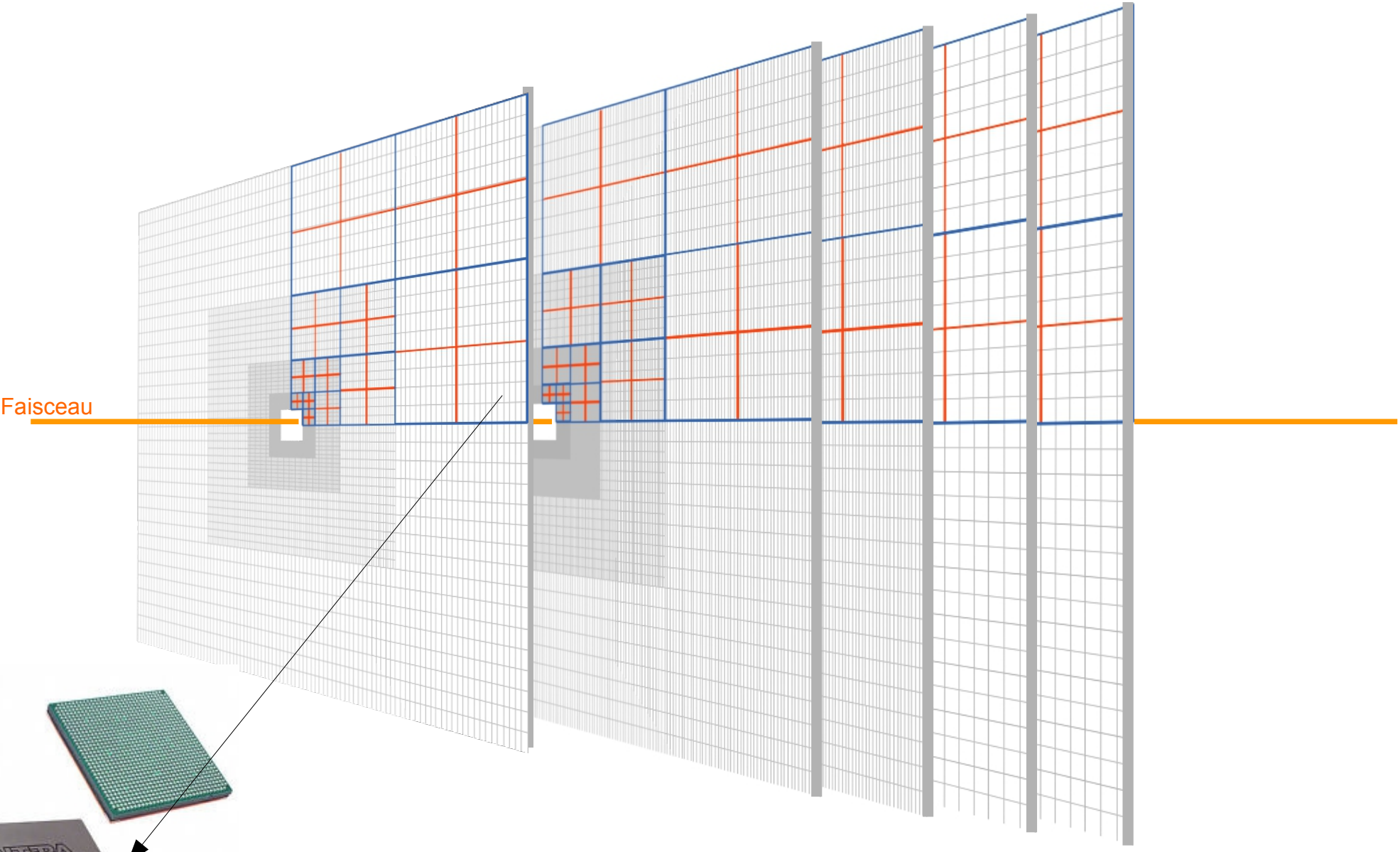
- Analyse logique,
- Serial Data Analyser

Segmentation du traitement



Quelle est l'unité de traitement optimale
minimisant le nombre de cartes et de liaisons ?

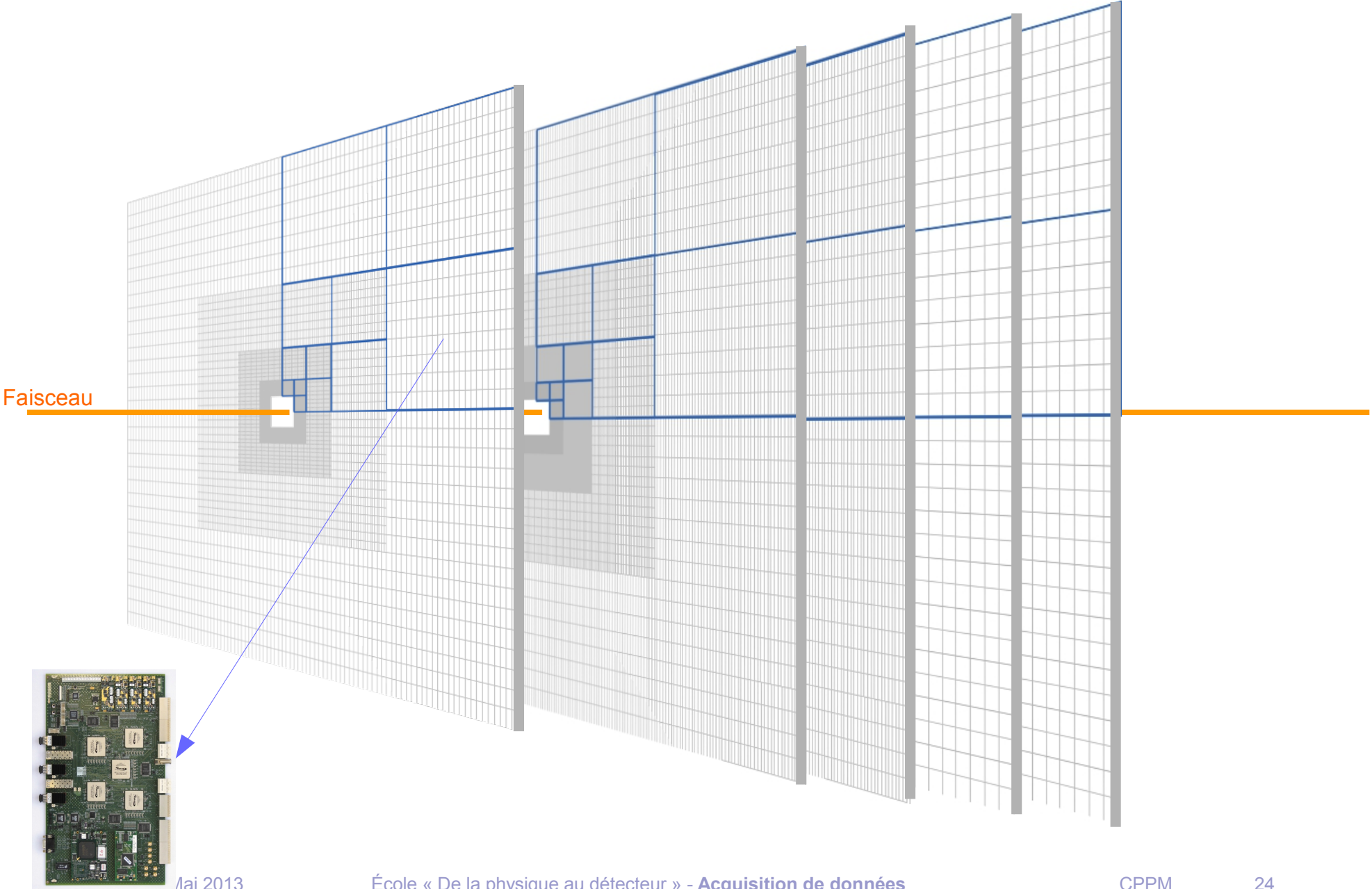
Traitement FPGAs



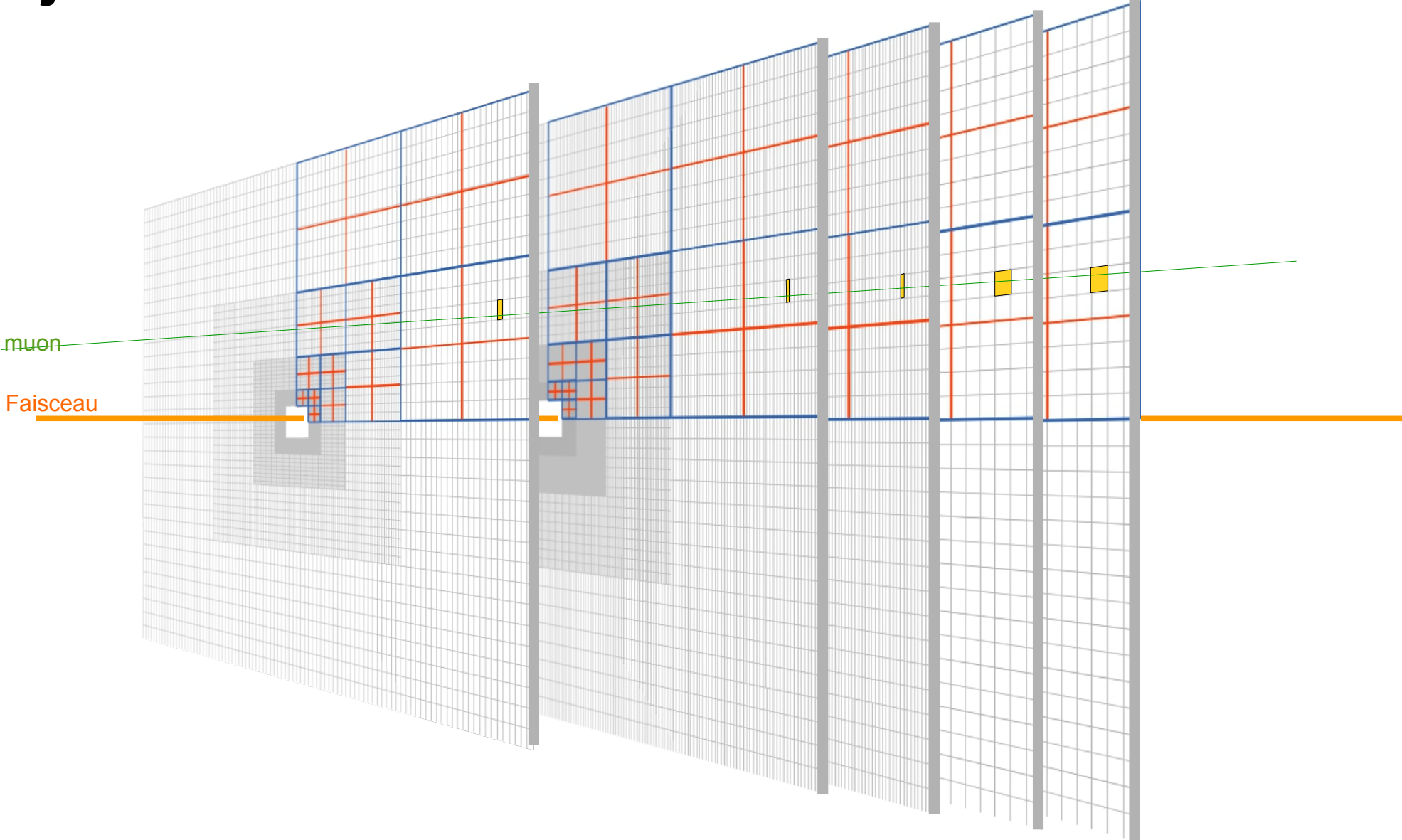
Faisceau

FPGA (Field Programmable Gate Array)

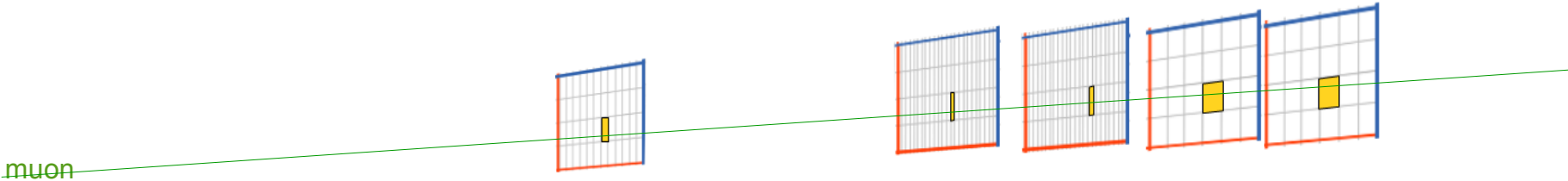
Traitement cartes



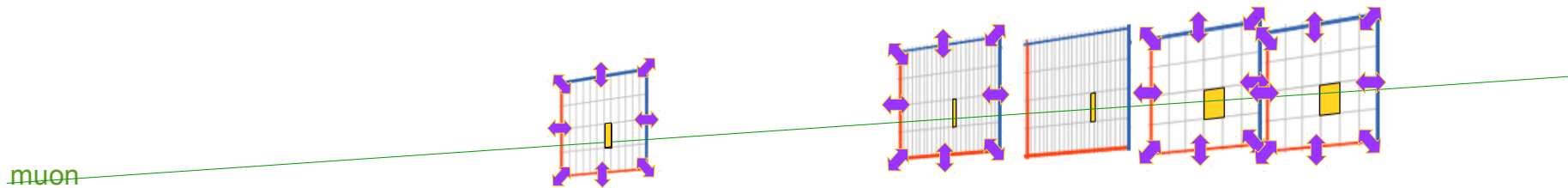
Trajectoire muon



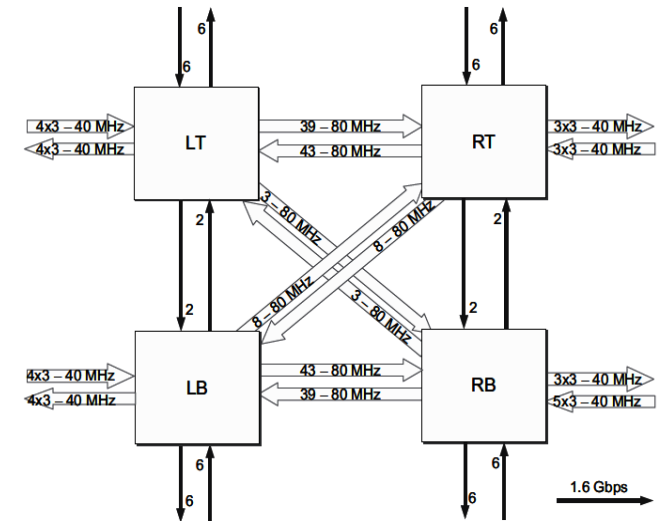
Ce que voit un FPGA



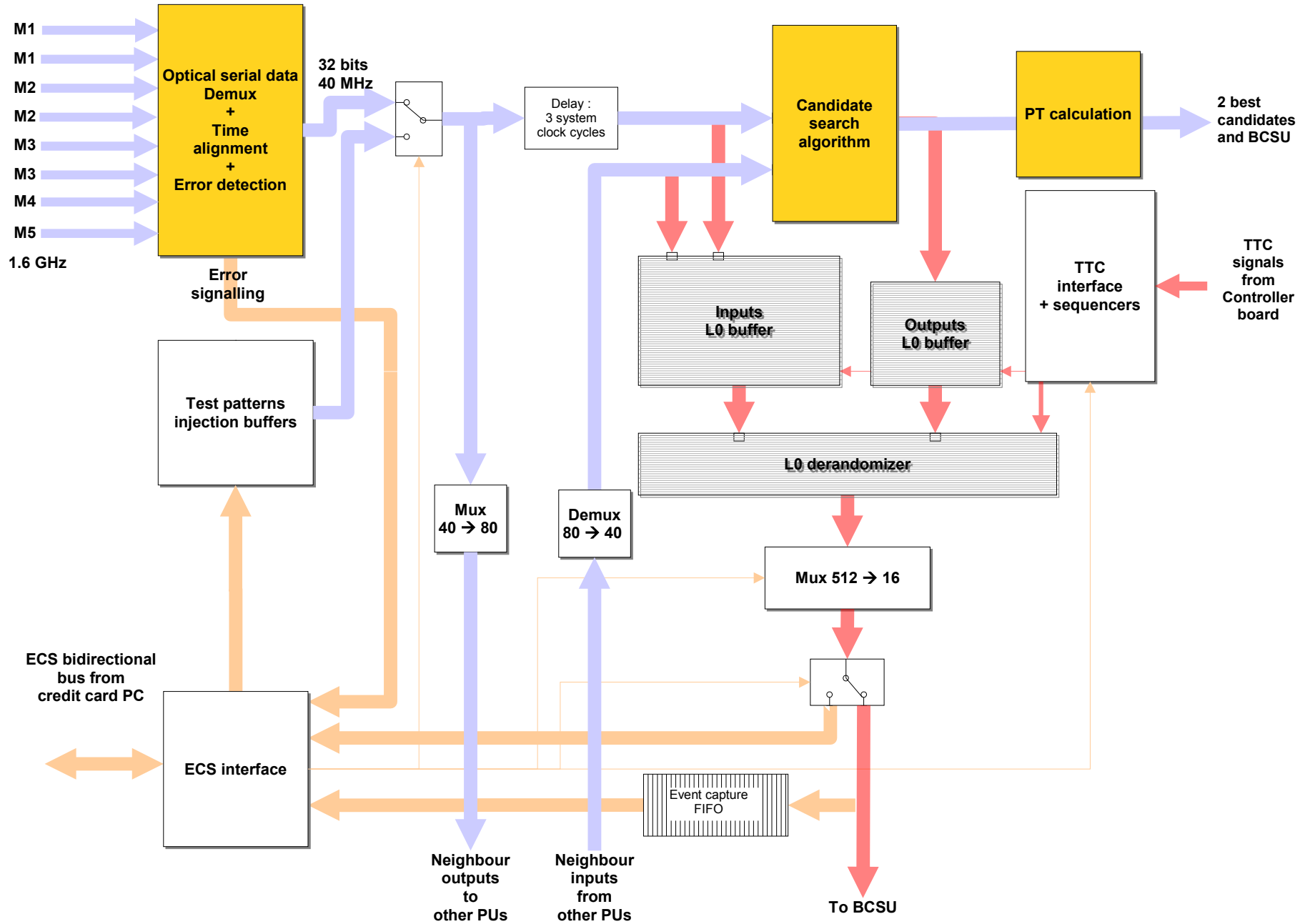
Echanges de données



**Nombreuses communications
pour traiter les détections aux limites**

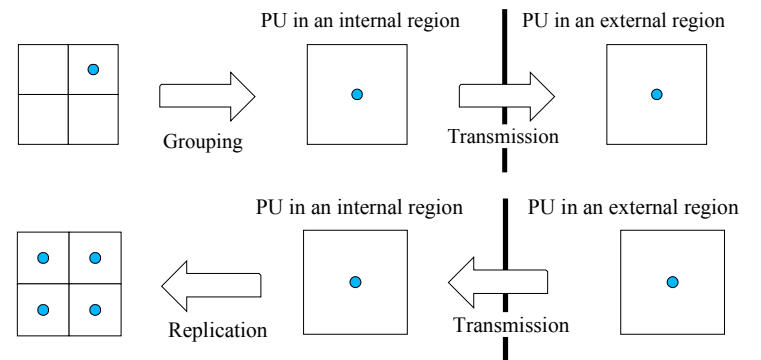
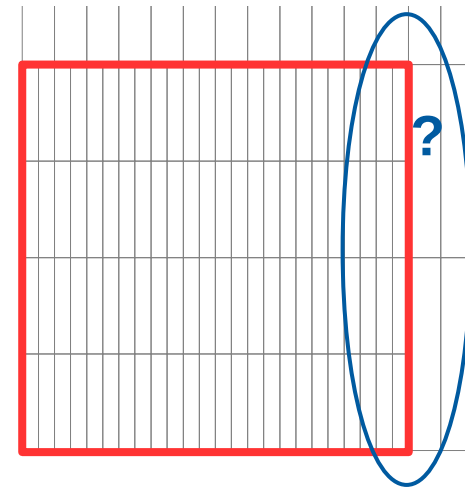
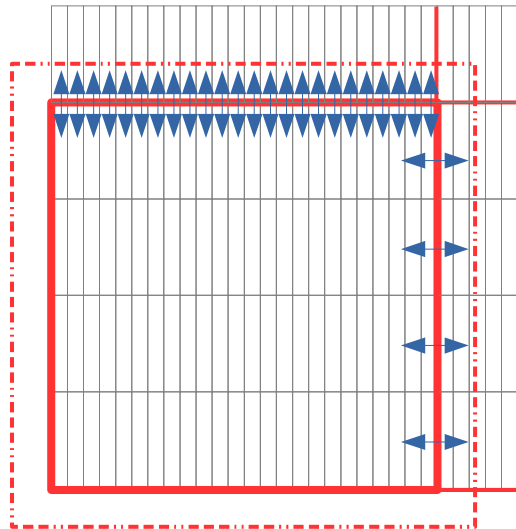


Traitement de données



Homogénéisation de l'espace de travail

Échanges de données

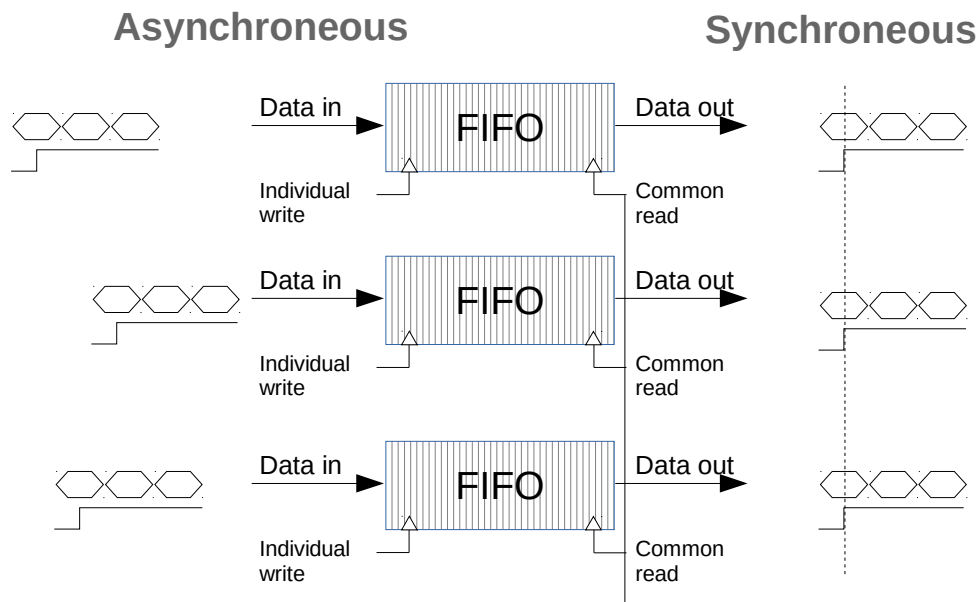


Mise en temps

Toutes les données arrivent décalées

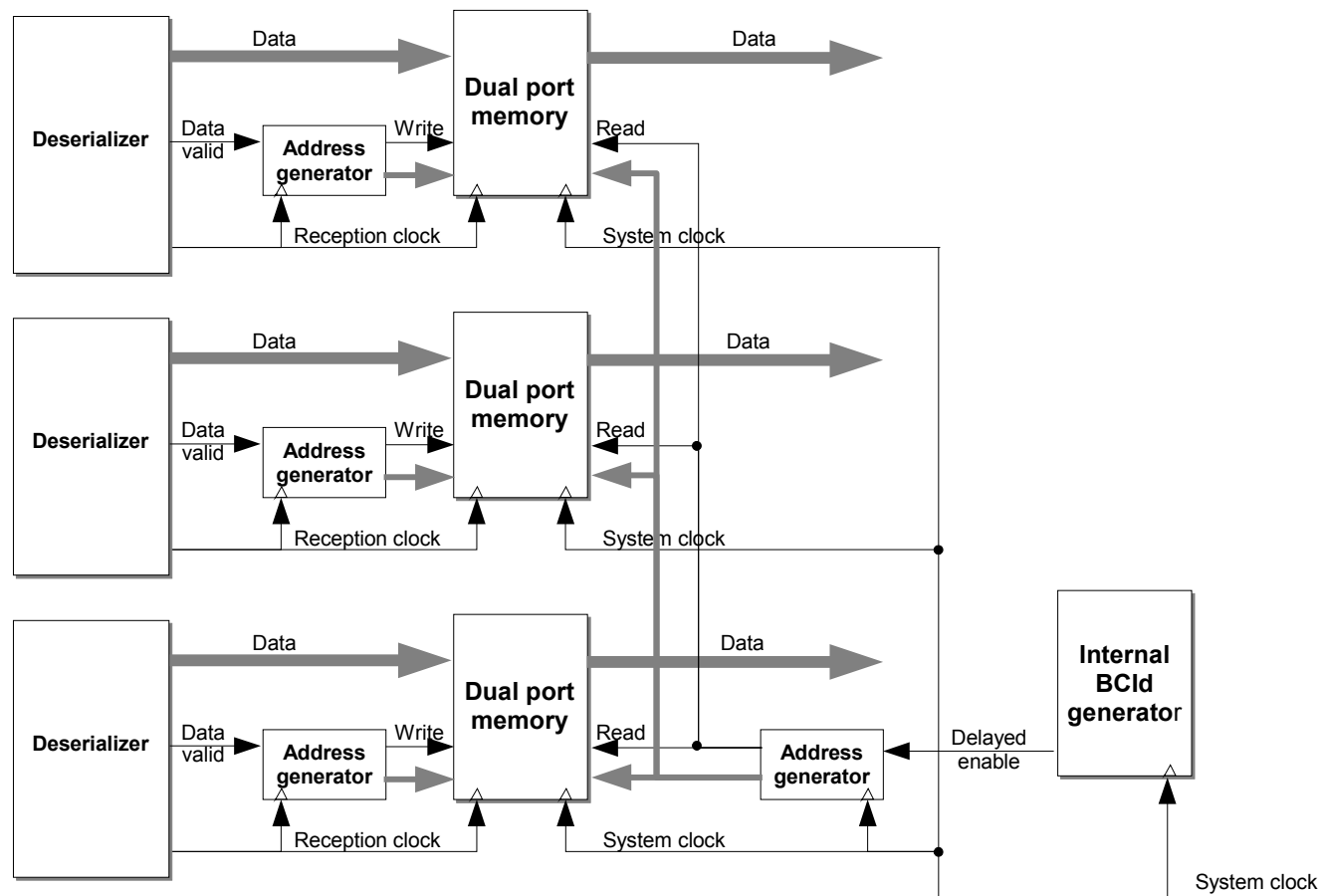
- Chemins de longueur différente
- Dérives thermiques
- Données de voisinage

Synchronisation



Mise en temps

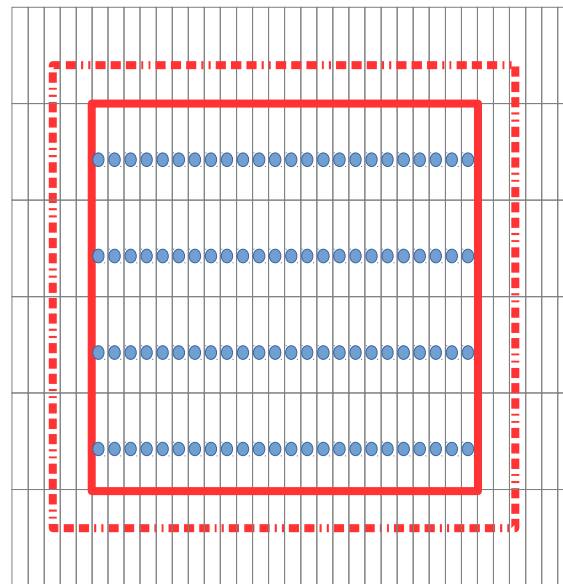
Implémentation effective



Traitement

Parallelisme massif

- Un résultat à donner toutes les 25 ns
- Pas le temps de chercher séquentiellement
 - L'algorithme de recherche est effectué sur l'ensemble des cellules simultanément



48 algorithmes

M3 seed

Traitement

Structure pipeline

- Le temps de recherche est supérieur à l'écart entre deux collisions (25 ns)
- Multiplication des unités de traitement trop coûteuse
 - On adopte une structure pipeline



	F1	F2	F3	F4	F5	F6	F7	F8
T0	Dn	Dn-1	Dn-2	Dn-3	Dn-4	Dn-5	Dn-6	Dn-7
T0 + 25	Dn+1	Dn	Dn-1	Dn-2	Dn-3	Dn-4	Dn-4	Dn-6
T0 + 50	Dn+2	Dn+1	Dn	Dn-1	Dn-2	Dn-3	Dn-4	Dn-5
T0 + 75	Dn+3	Dn+2	Dn+1	Dn	Dn-1	Dn-2	Dn-3	Dn-4
T0 + 100	Dn+4	Dn+3	Dn+2	Dn+1	Dn	Dn-1	Dn-2	Dn-3
T0 + 125	Dn+5	Dn+4	Dn+3	Dn+2	Dn+1	Dn	Dn-1	Dn-2
T0 + 150	Dn+6	Dn+5	Dn+4	Dn+3	Dn+2	Dn+1	Dn	Dn-1
T0 + 175	Dn+7	Dn+6	Dn+5	Dn+4	Dn+3	Dn+2	Dn+1	Dn

Operation	Estimated time [ns]	Estimated number of clock periods
Time of flight to M5	63	
FE board processing	70	
Transmission to IB and ODE (15 m)	105	13
IB processing	40	
ODE processing	30	
Transmission to processing (100 m)	600	24
Muon processing	1200	48
Transmission to L0 decision Unit	50	2
L0 Decision Unit processing	525	21
L0 Decision Unit distribution	800	32
Contingency	500	20
Total	3983	160

- Profondeur du pipeline muon trigger LHCb : 48 coups d'horloge

Calcul du P_T

Formule simple

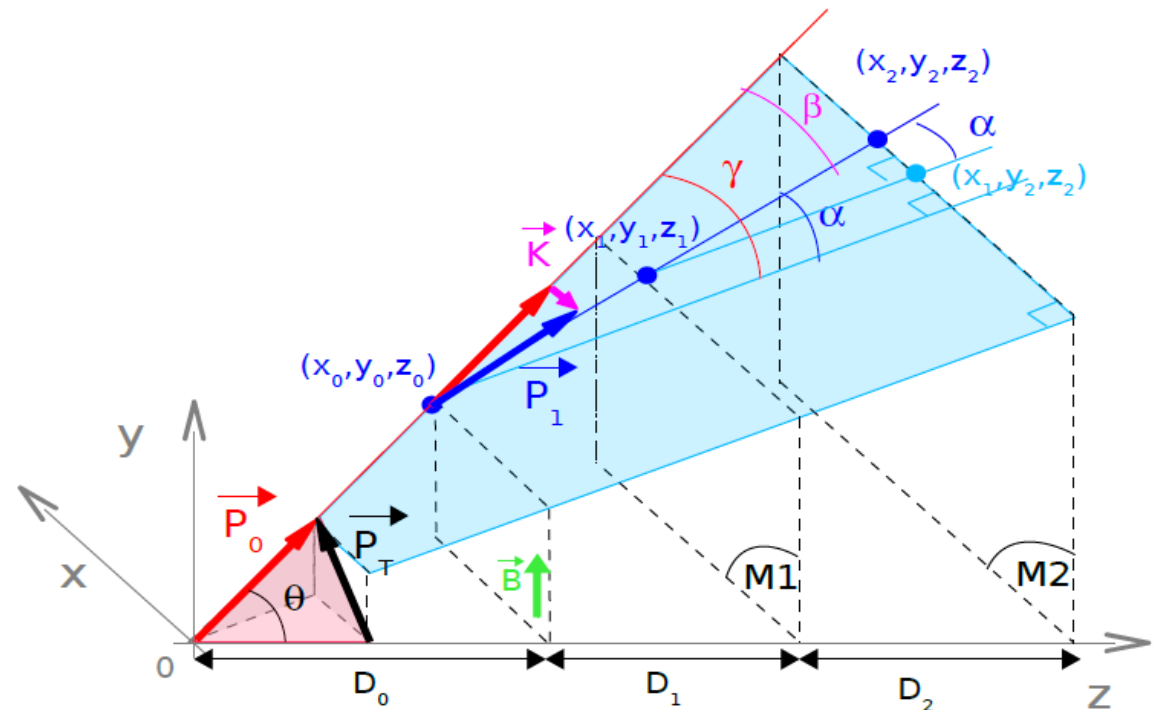
$$P_T = P_0 \sin(\theta)$$

Calcul du P_T

Formule simple

$$P_T = P_0 \sin(\theta)$$

sauf que



- P_0 is obtained from β , the deflection angle and \vec{K} :

$$P_0 = \frac{K}{2\sin(\frac{\beta}{2})} = \frac{K}{2\sin(\frac{\gamma-\alpha}{2})}$$

where $\tan(\gamma) = \frac{x_0}{\sqrt{D_0^2 + y_0^2}}$ and $\sin(\alpha) = \frac{x_2 - x_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + D_2^2}}$

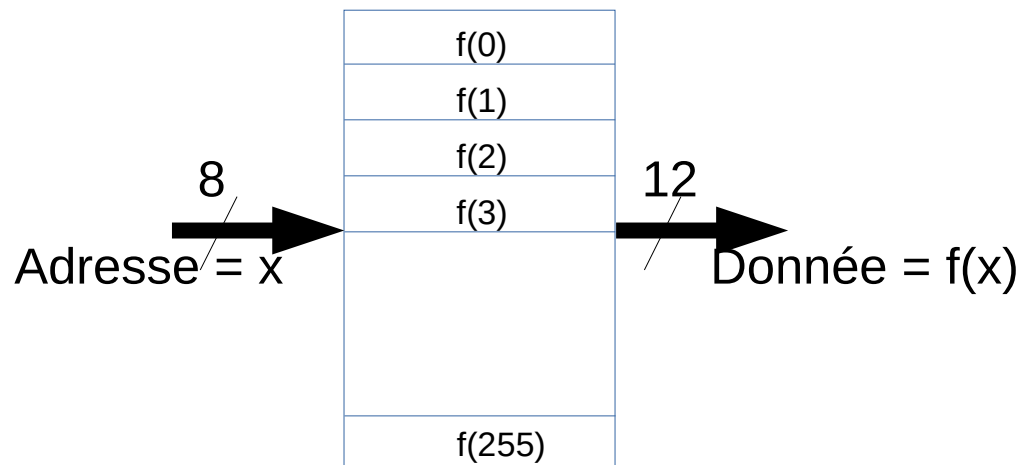
- $\sin(\theta)$ is given by:

$$\sin(\theta) = \frac{R_0}{\sqrt{(R_0^2 + D_0^2)}} = \frac{\sqrt{x_0^2 + y_0^2}}{\sqrt{x_0^2 + y_0^2 + D_0^2}}$$

Calcul du P_T

Utilisation de LUT

- Permet de calculer n'importe quelle fonction de type $f(x)$ même complexe en un coup d'horloge



- Faisable tant que range (x) reste faible

Quelques chiffres

Trigger muon LHCb

- Avec 240 FPGAs interconnectés, ceci permet de réaliser **740 milliards** d'algorithmes de recherche par seconde.

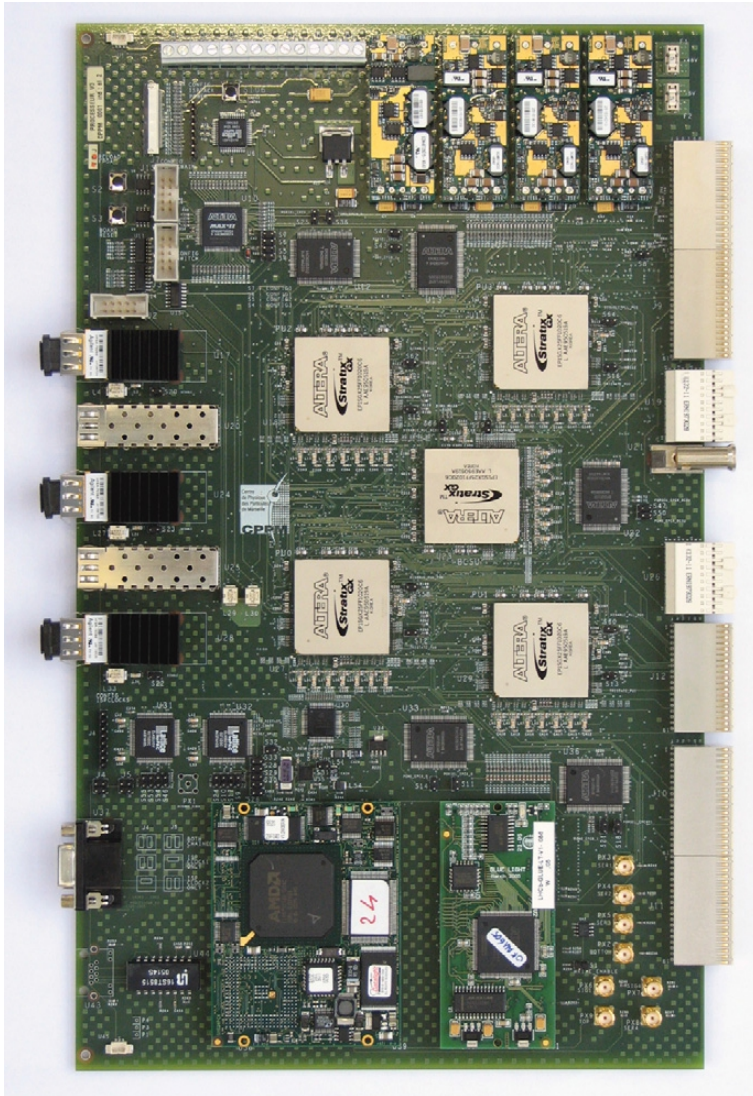


Testabilité

Primordial de comprendre les anomalies quand elles surviennent

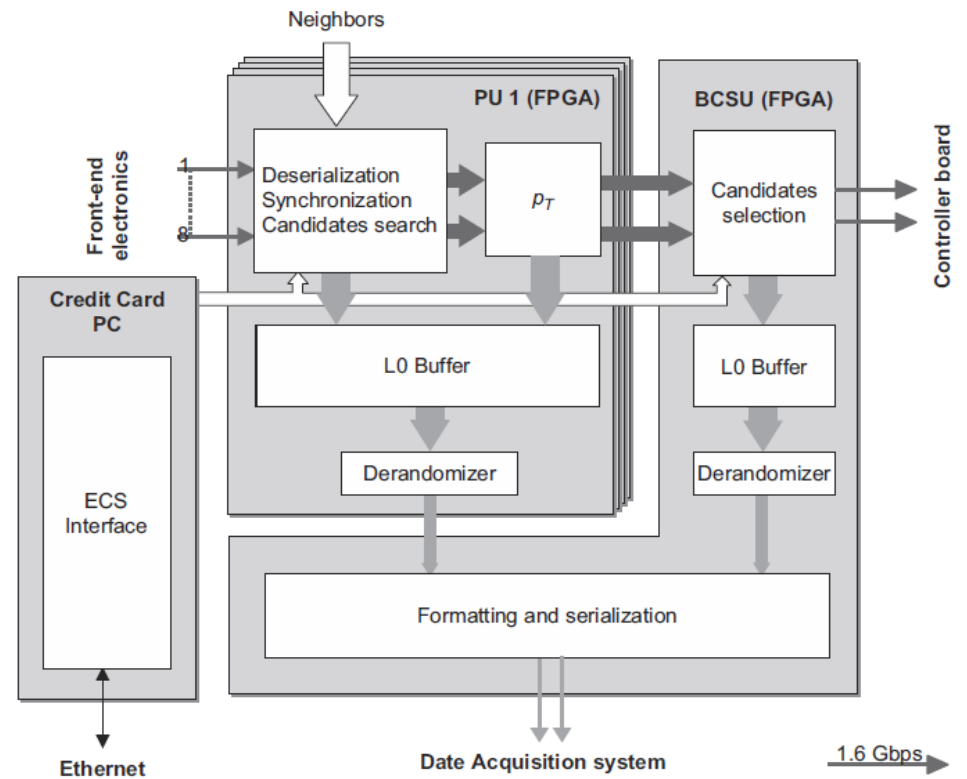
- L'algorithme ne représente pas plus de 50 % de l'occupation du FPGA
- Le reste est occupé par des fonctions de test et de monitoring
 - Injection de données simulées
 - Event capture
 - Relecture à différents endroits de la chaîne de traitement.

Carte trigger

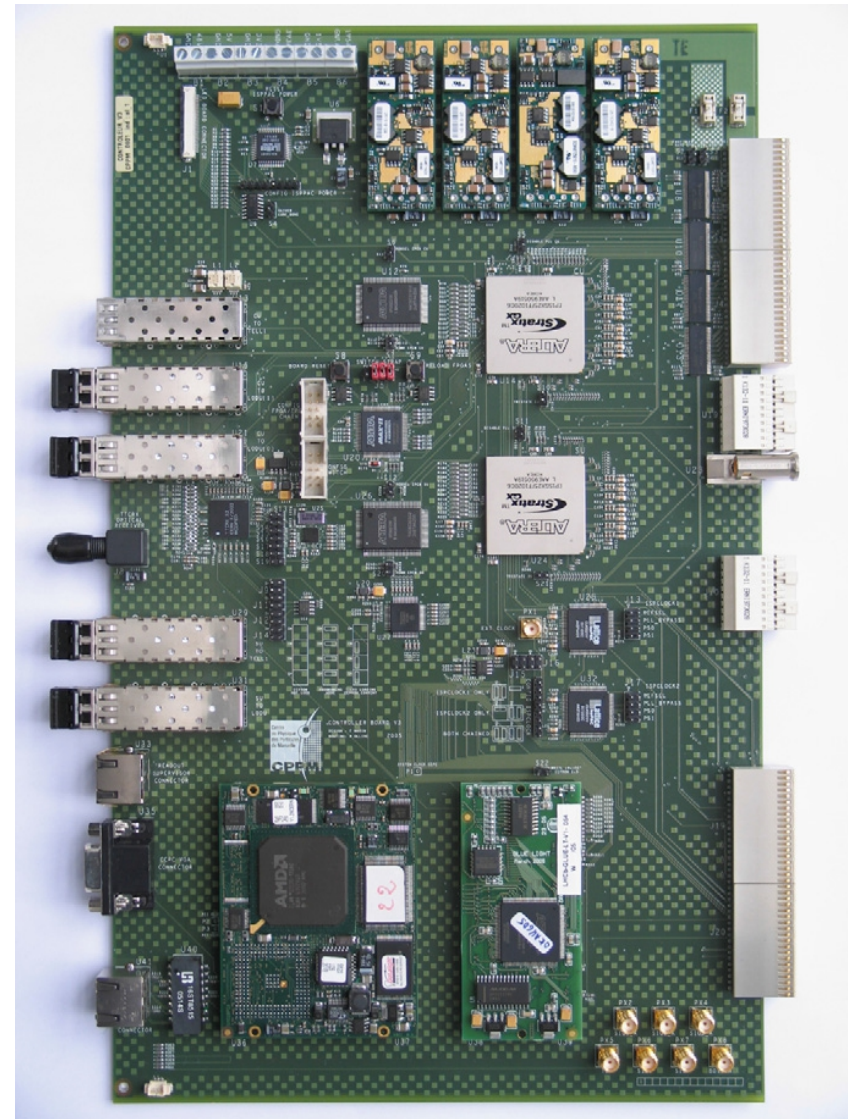
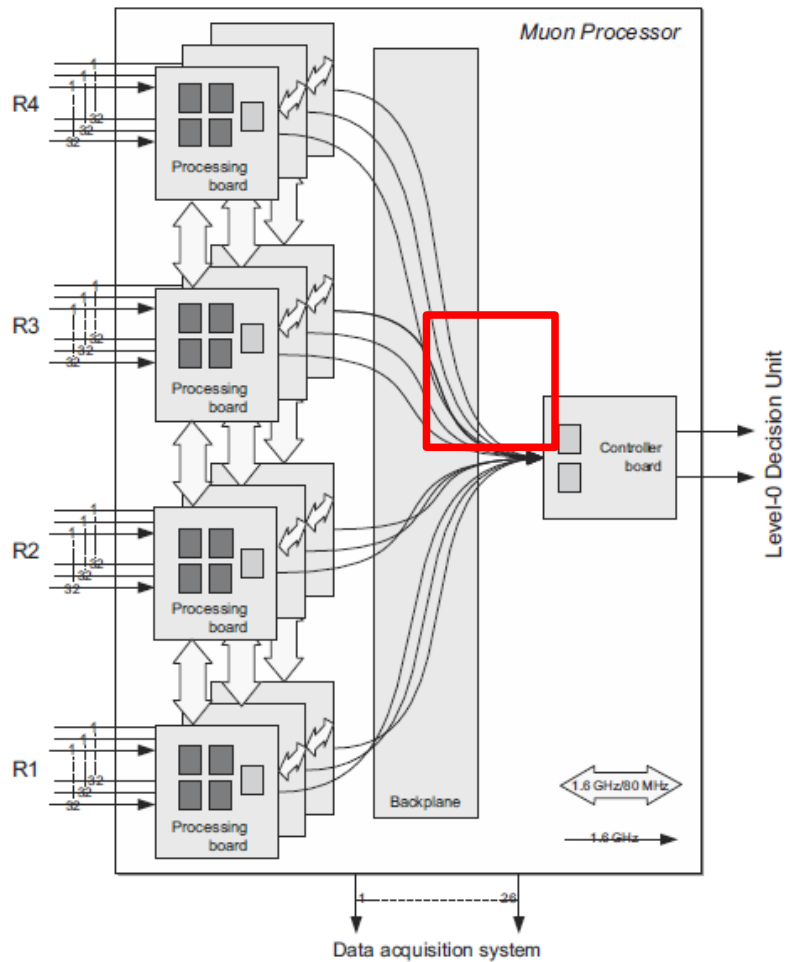


Carte générique

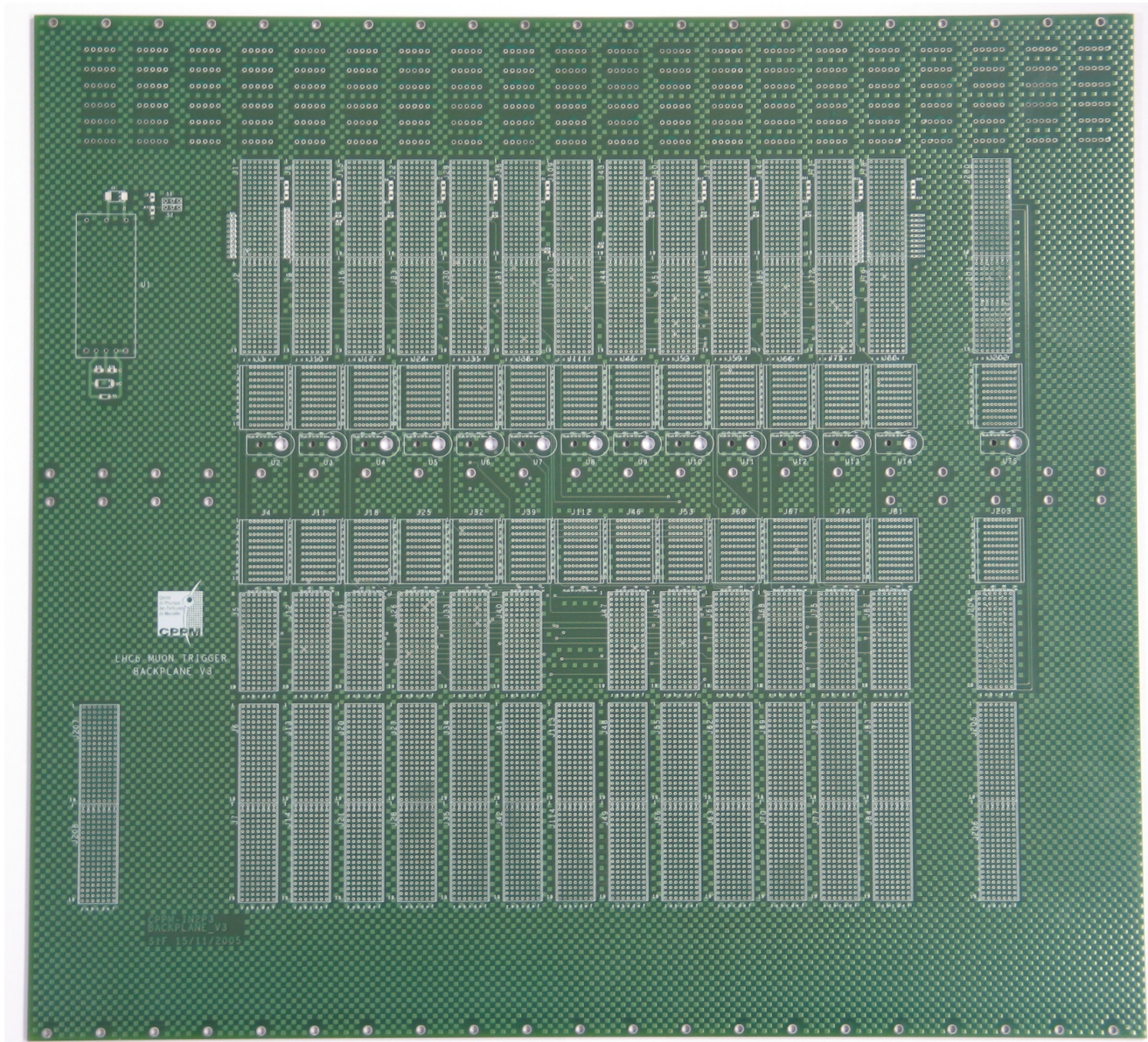
- Mais 48 configurations de FPGA différentes



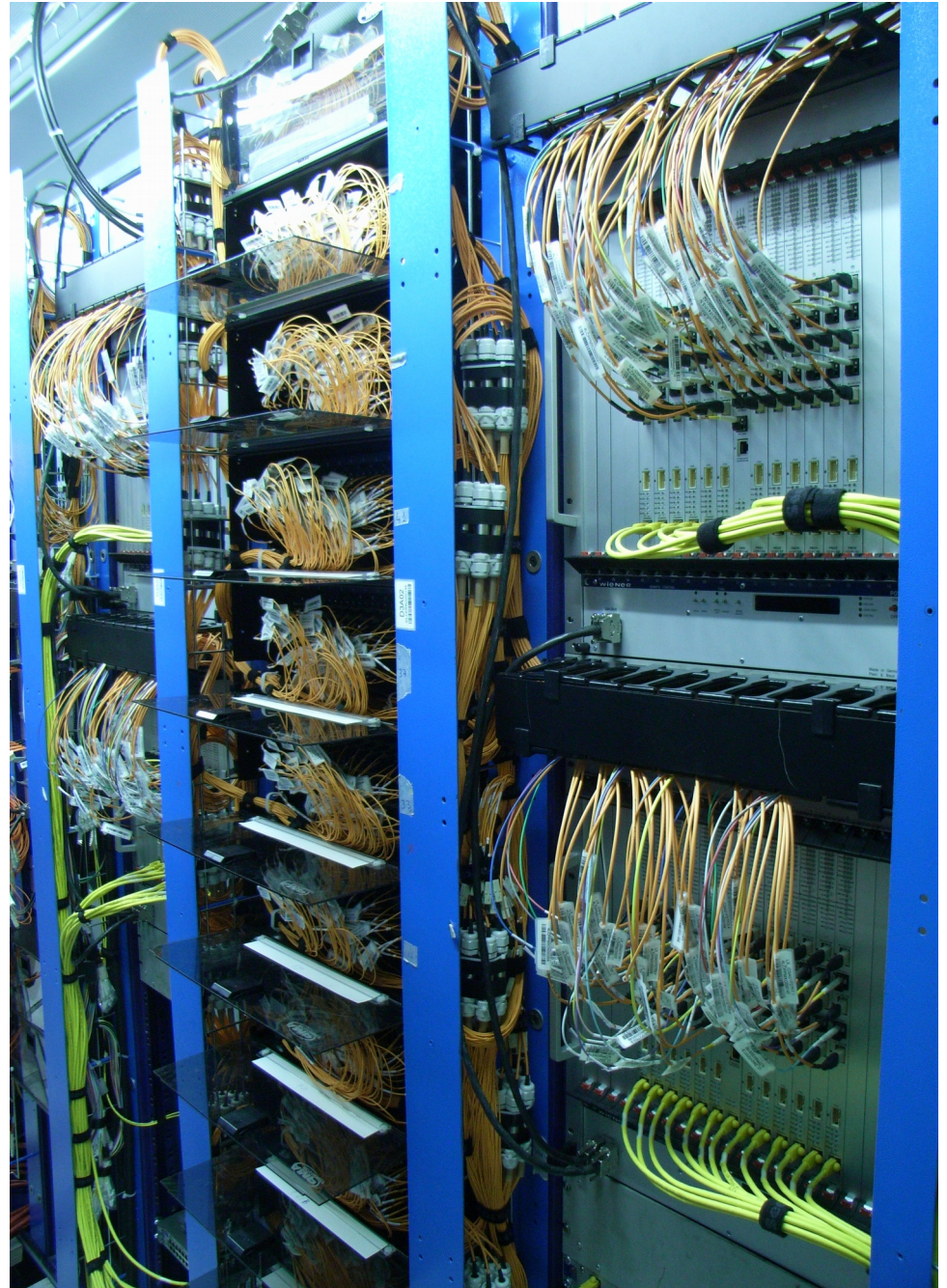
Carte de contrôle



Custom backplane



Le trigger à muons



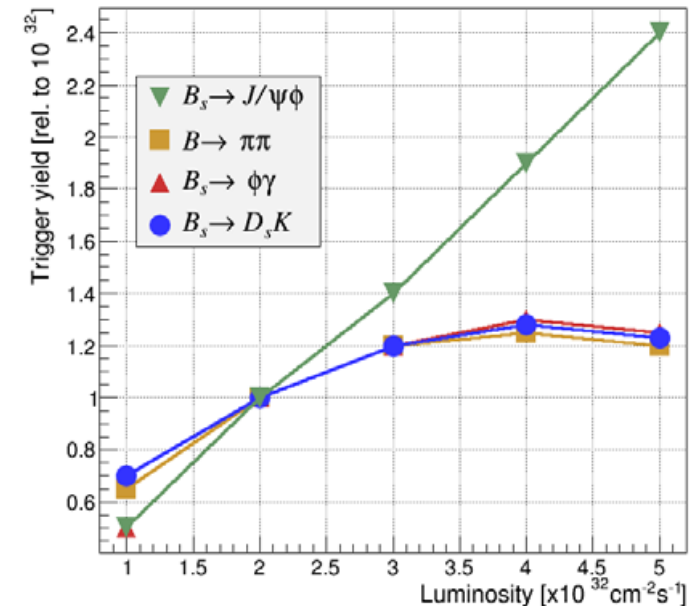
Evolution du détecteur : l'upgrade

LHCb Upgrade

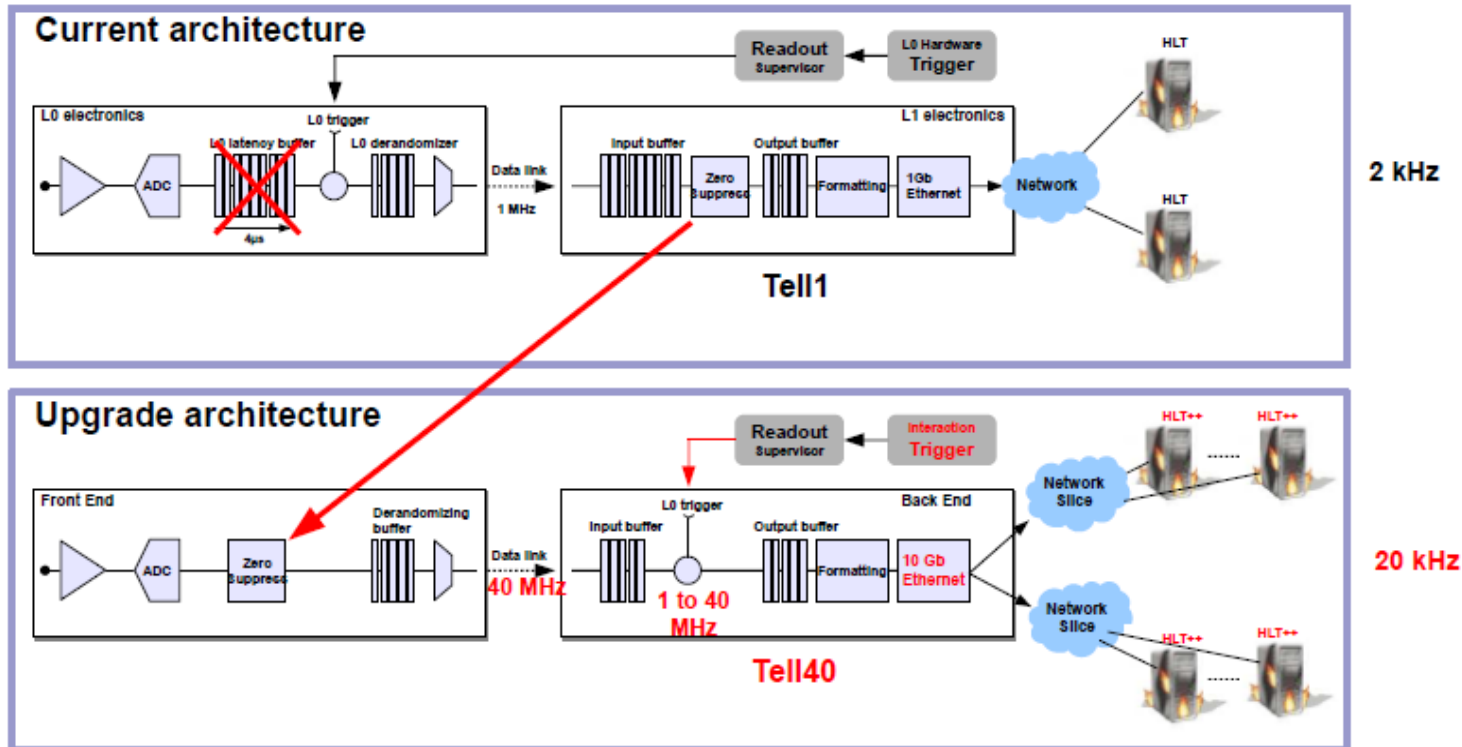
Motivation

- Luminosité maximale sous 5 ans : 5 fb^{-1}
- Au rythme actuel la précision statistiques des mesures varie très lentement
- En augmentant la luminosité de 2×10^{32} à $10^{33} \text{ cm}^{-2}\text{s}^{-1}$
 - Parvenir à une luminosité cumulée supérieure à 50 fb^{-1}

- Saturation du trigger sur les canaux hadroniques



Upgrade : triggerless readout

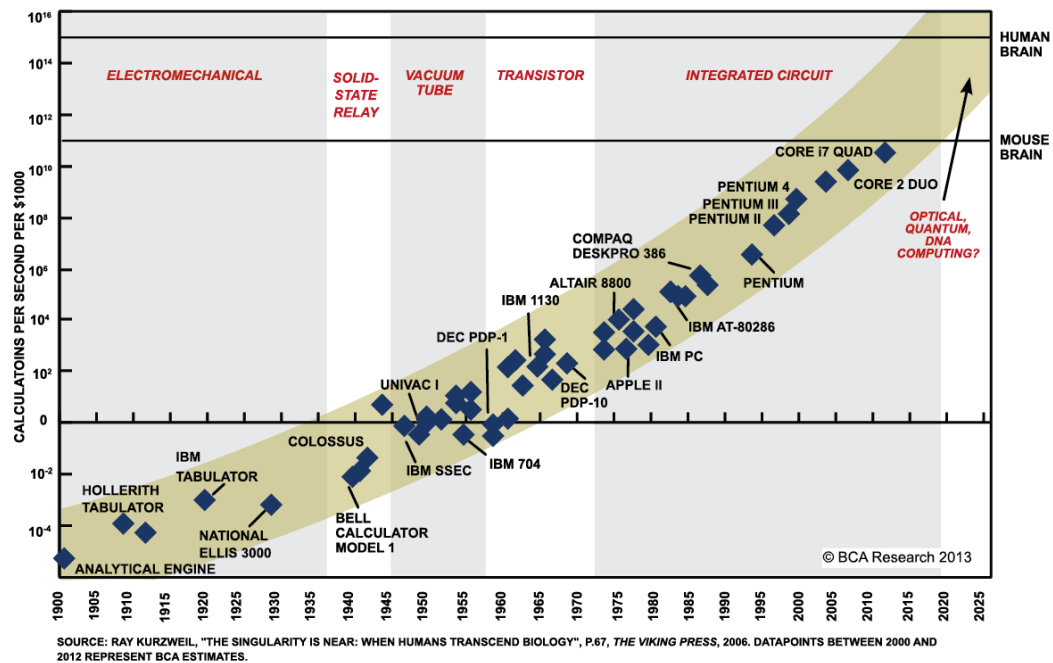


Migration vers une architecture sans trigger

- Fonction trigger réalisée dans la ferme
 - ➔ Relecture à 40 MHz au lieu de 1 MHz
- Compression dans les front-ends pour diminuer le nombre de liens optiques
- Liens à 10 Gbits/s vers les fermes

Loi de Moore

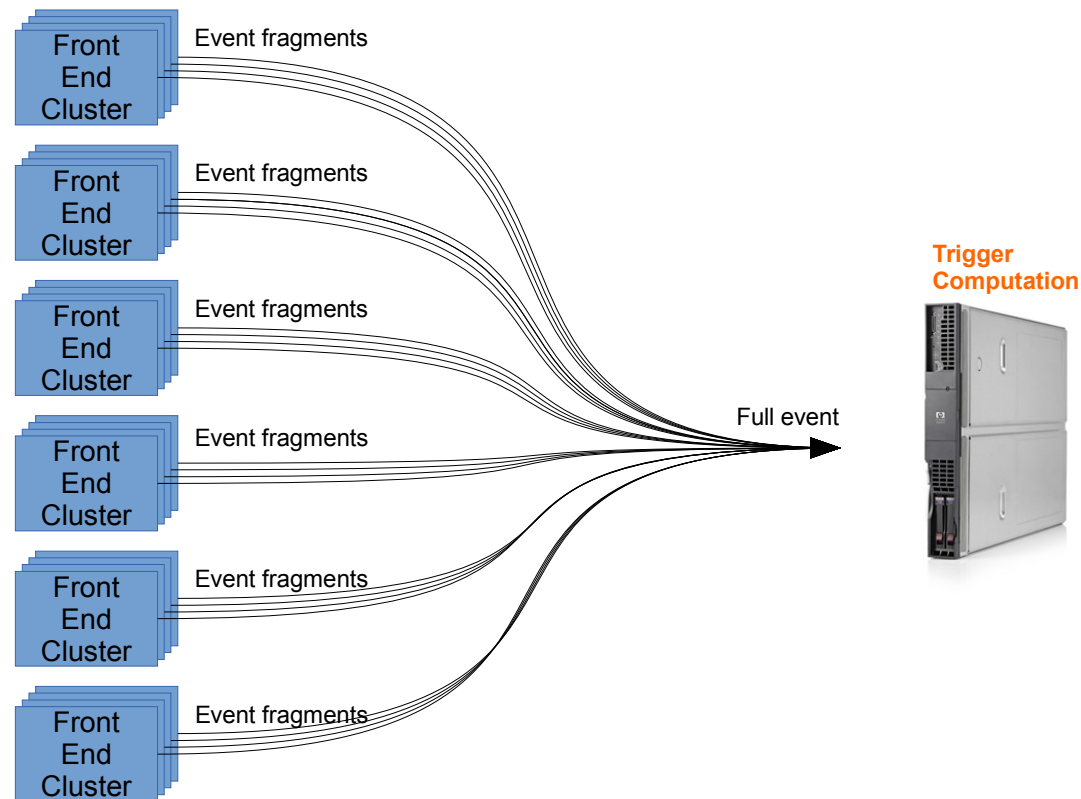
Si la ferme peut traiter les événements à 1 Mhz en 2008



Elle doit pouvoir traiter 40 Mhz entre 2018 et 2020

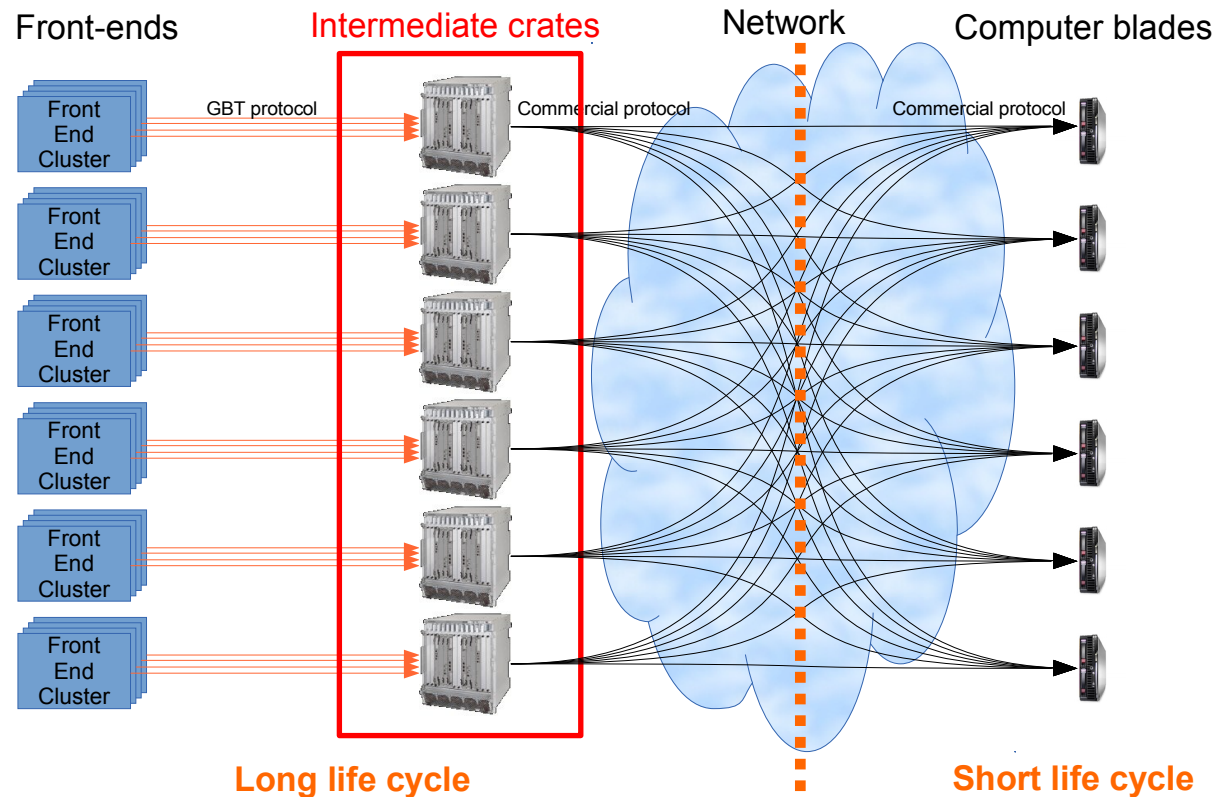
Principe du « triggerless readout »

Tous les fragments d'événements doivent être routés vers un seul CPU



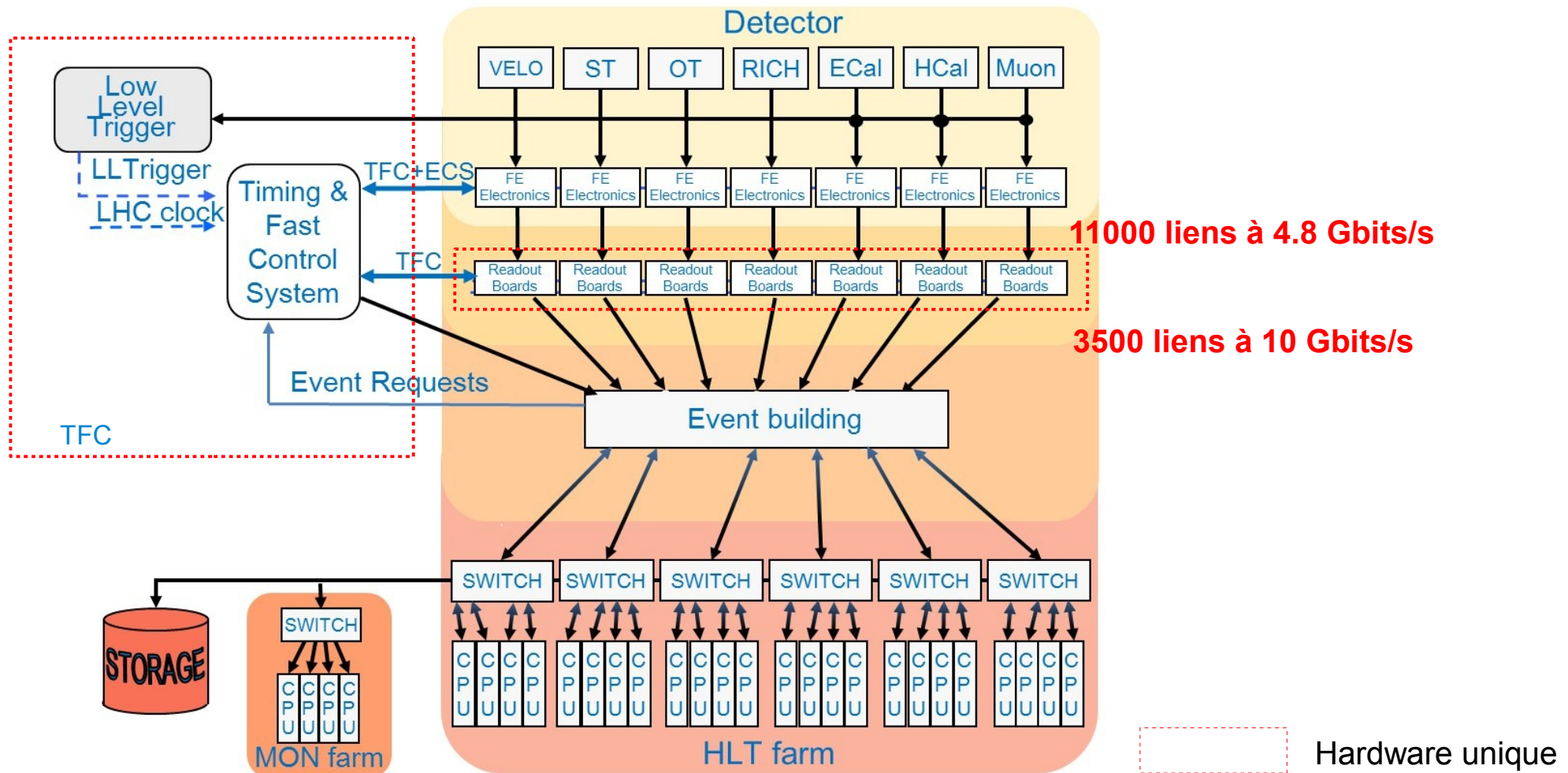
Choix d'architecture initiale du readout LHCb

Systeme éprouvé distinguant le back end des fermes à courte durée de vie



Architecture

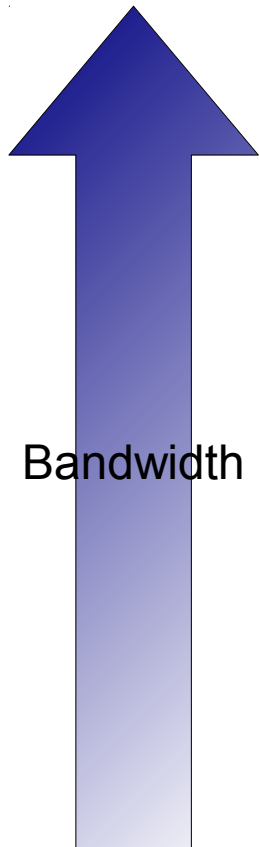
Une carte de readout commune et reconfigurable



Standard mécano-électrique

VME vieillissant

→ Besoin d'un nouveau standard

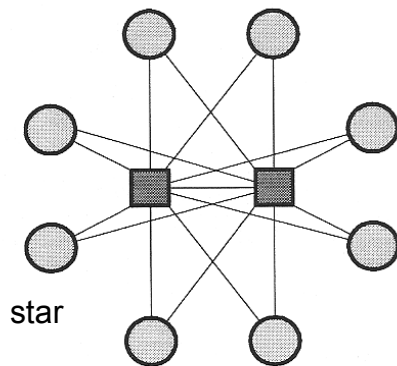


Standard	Bandwidth in Mbytes/s
ATCA 40Gb	1 820 000
ATCA 10Gb	455 000
VPX (VITA46)	112 500
VXS (VITA 41)	20 000
SHB Express	17 500
Compact PCIe/PSB	5 000
PCI 64 x 33 Mbits/s	533
VME 320	320
VME64x	160
PCI 32 x 32 Mbits/s	133
VME64	80
VME32	40
VME16	20

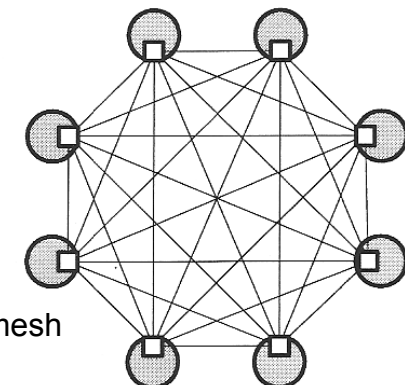
Plus de backplane custom

Utilisation du standard ATCA

- Nombreux avantages :
 - Bien adapté aux composants récents
 - Plus de place pour les radiateurs
 - Alimentation jusqu'à 3kW/crate
 - Refroidissement adapté
 - Backplane standard
 - Topologie basée sur des liens sériels
 - Mezzanines normalisées
 - Coûts similaires au VME
 - Redondance
 - Système normalisé de surveillance de l'état du système (IPMI)

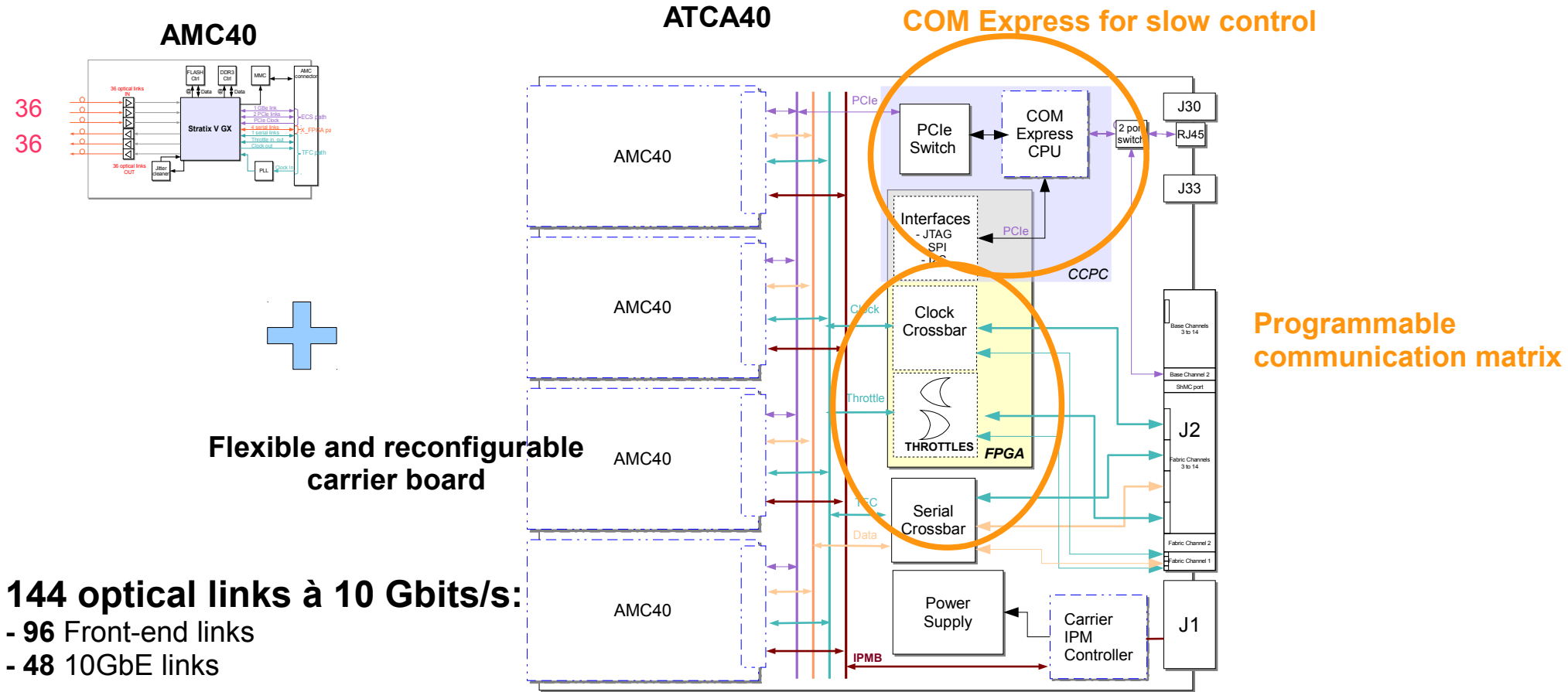


Topologie dual star



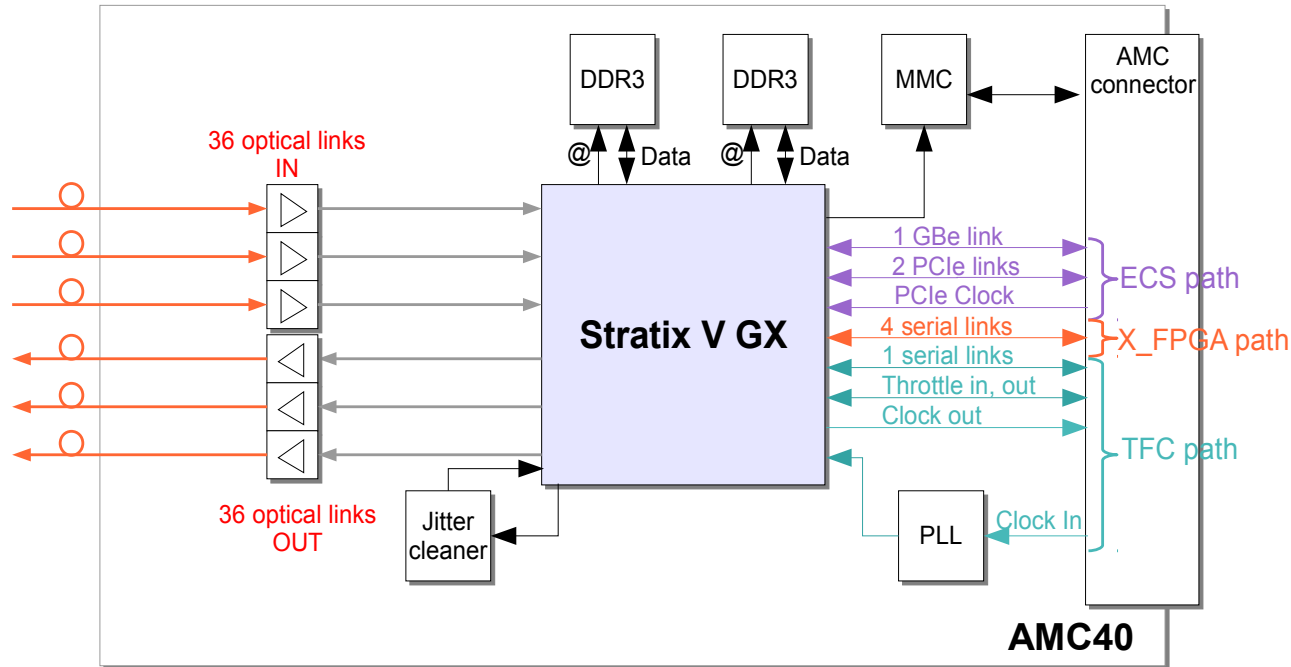
Topologie full mesh

Carte de readout générique



Obtention des fonctions readout, slow control, Timing and Fast Control ou Low Level Trigger interface par simple reprogrammation des FPGA et des chemins des crossbars

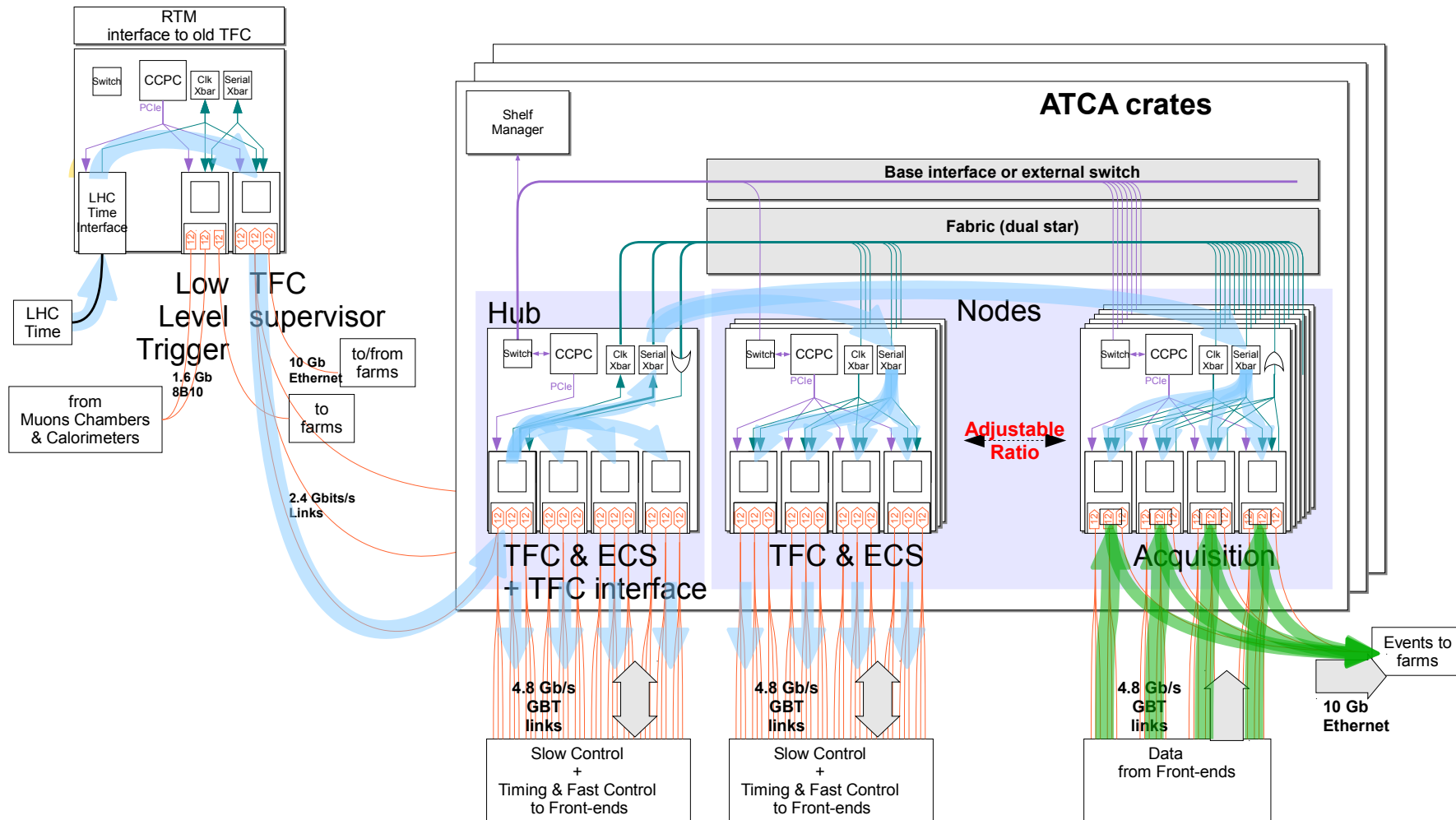
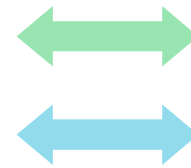
Carte mezzanine optique générique



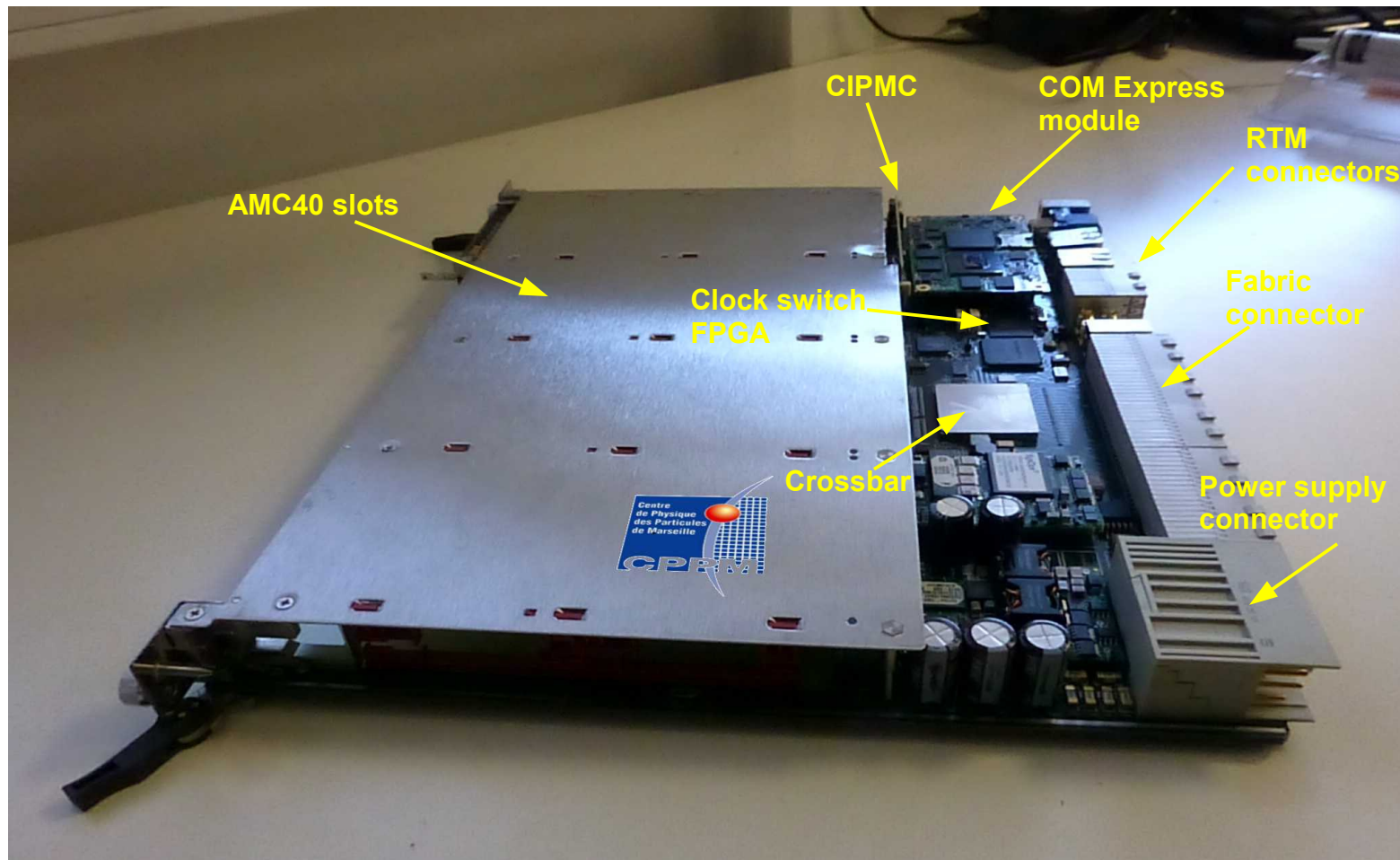
36 bidirectional optical links at up to 10 Gbits/s

622 kLE FPGA Stratix V GX: 5SGXEA7N2F45C3N

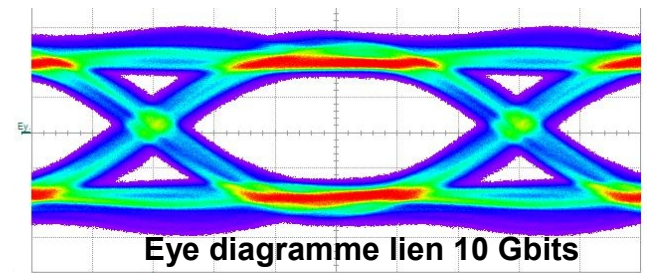
Chemins : Acquisition Timing and Fast Trigger



Carte ATCA40



Carte AMC40

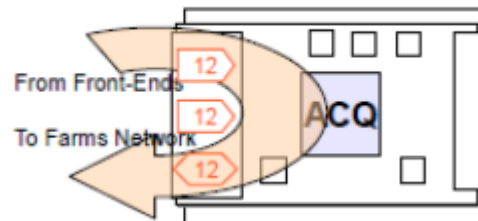


AMC40
1 Stratix V GX
36 optical inputs and
36 optical outputs at up to 10 Gbits/s
Slow control through PCIe

Élément dimensionnant : capacité d'entrées sorties

Les données sont compressées au niveau des front-ends

- Aucune compression supplémentaire n'est possible
 - Bande passante entrante = bande passante sortante

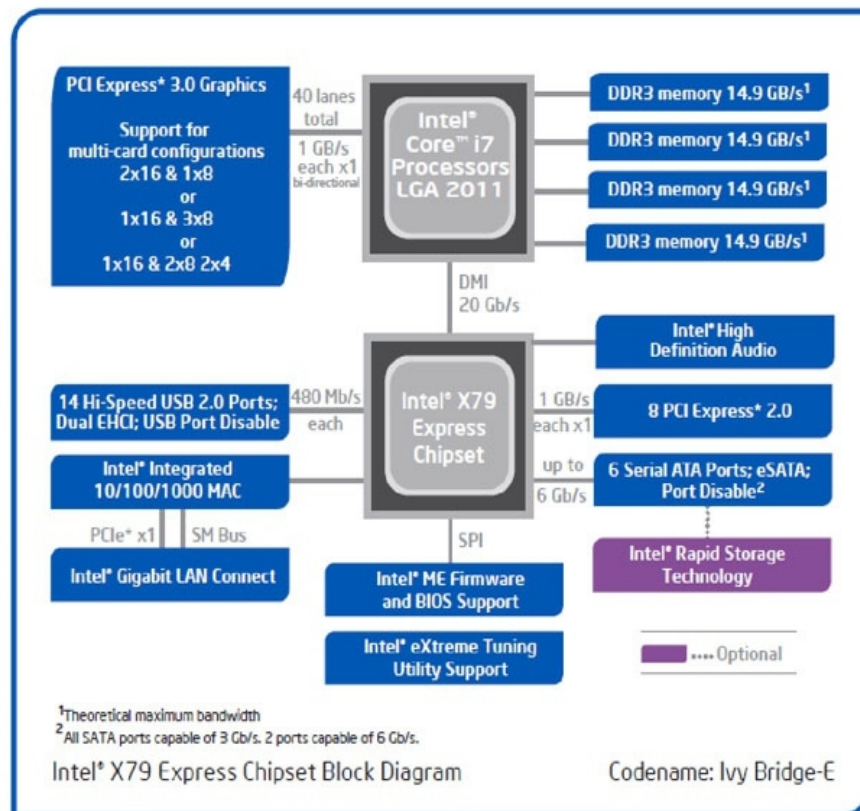


- Protocole GBT du CERN : 4.2 Gbits/line
 - Répartition optimale : 24 liens en entrées ~**100 Gbits**
10 liens 10GbE Ethernet en sortie ~**100 Gbits**

Amélioration architecture interne des CPUs

Libération de la bande passante des CPUs

- 40 canaux PCIe GEN3 à 8 Gbits/s
- Accès à la mémoire de passe plus par le processeur

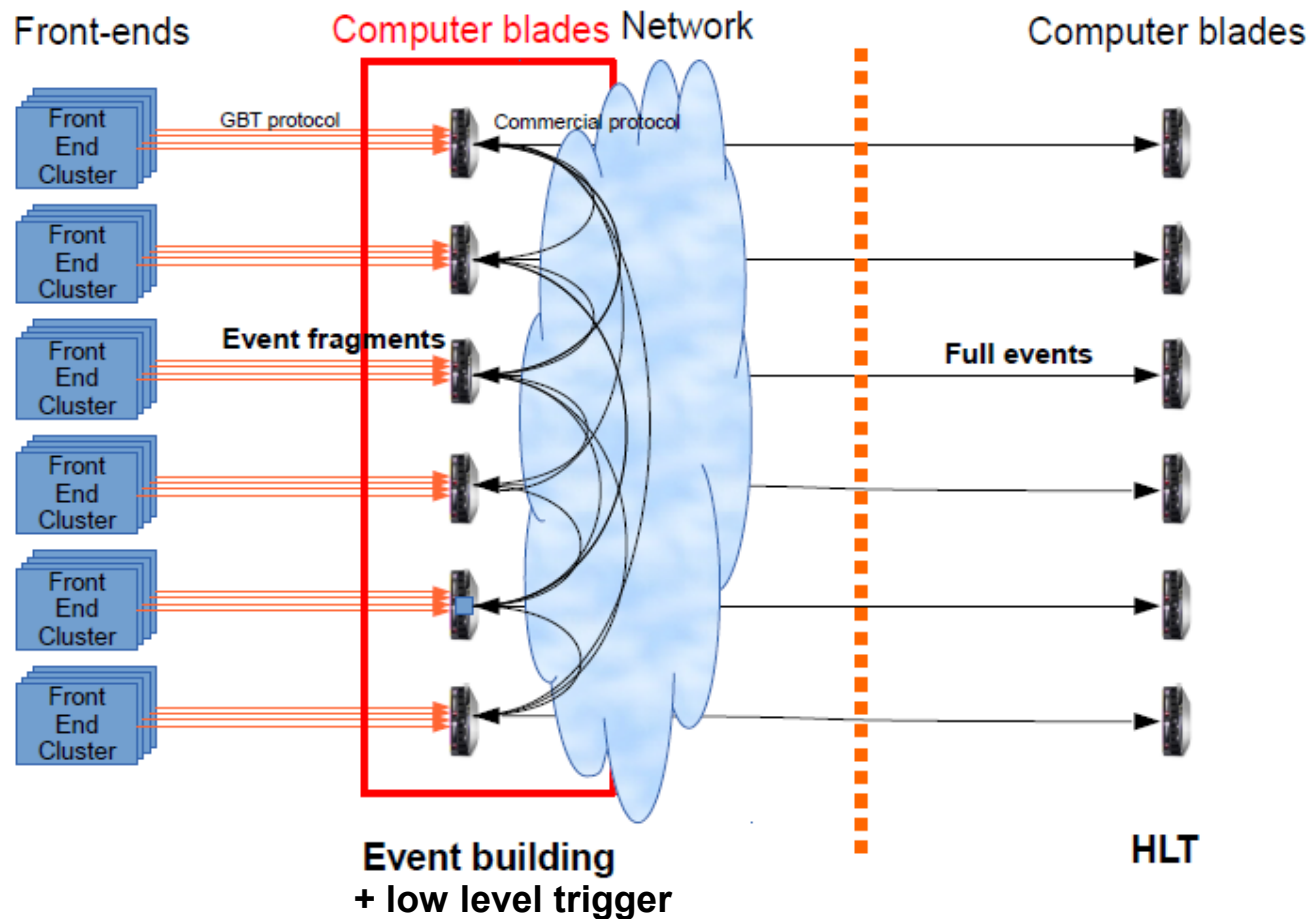


→ Donne la capacité au CPU de prendre en charge l'Event Building complet

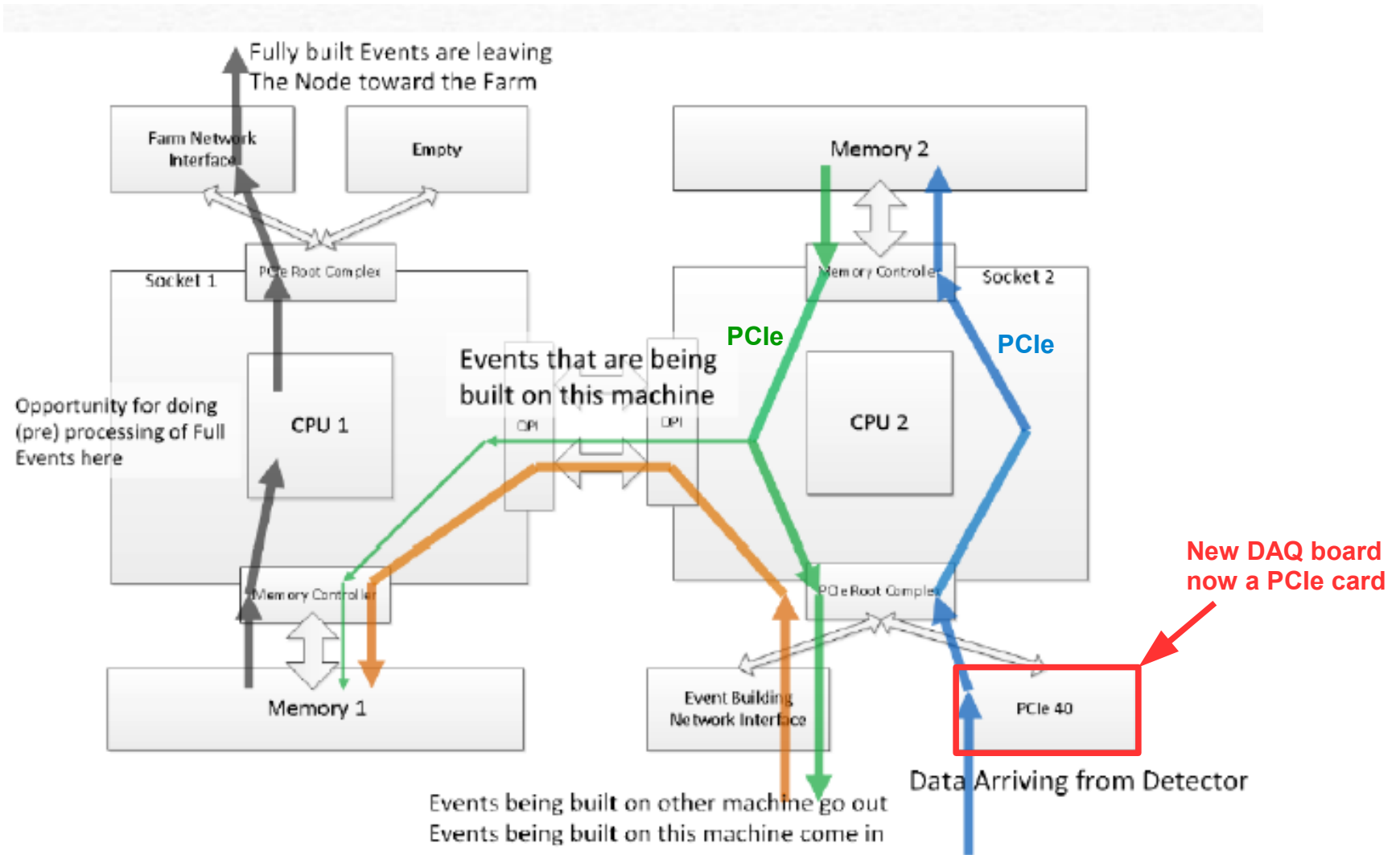
Ivy Bridge architecture

Nouveau schéma de readout

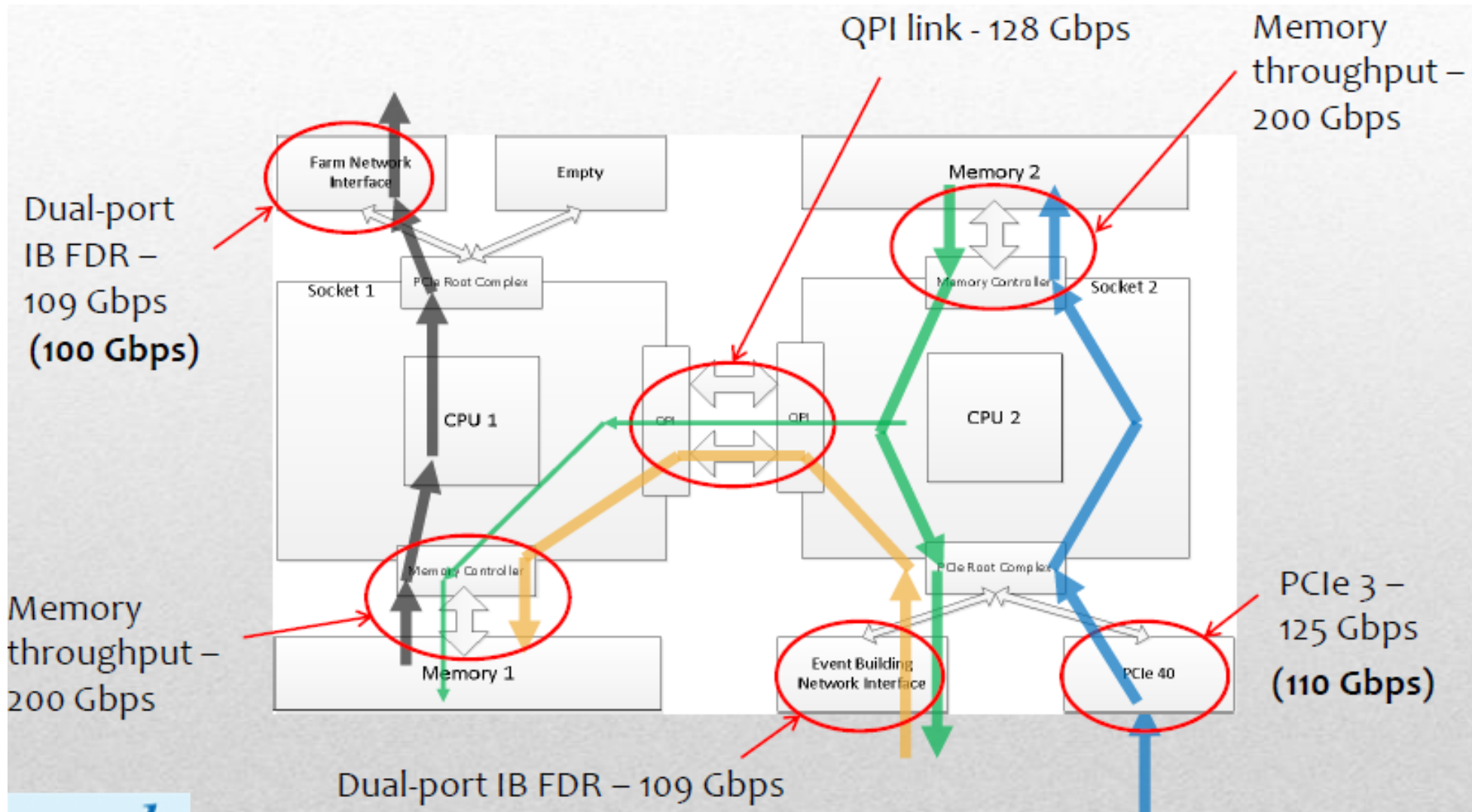
Déplacement des FPGAs back-ends dans les fermes de calcul



Data path



Bandwidth



Avantages et inconvénients

😊 Coûts

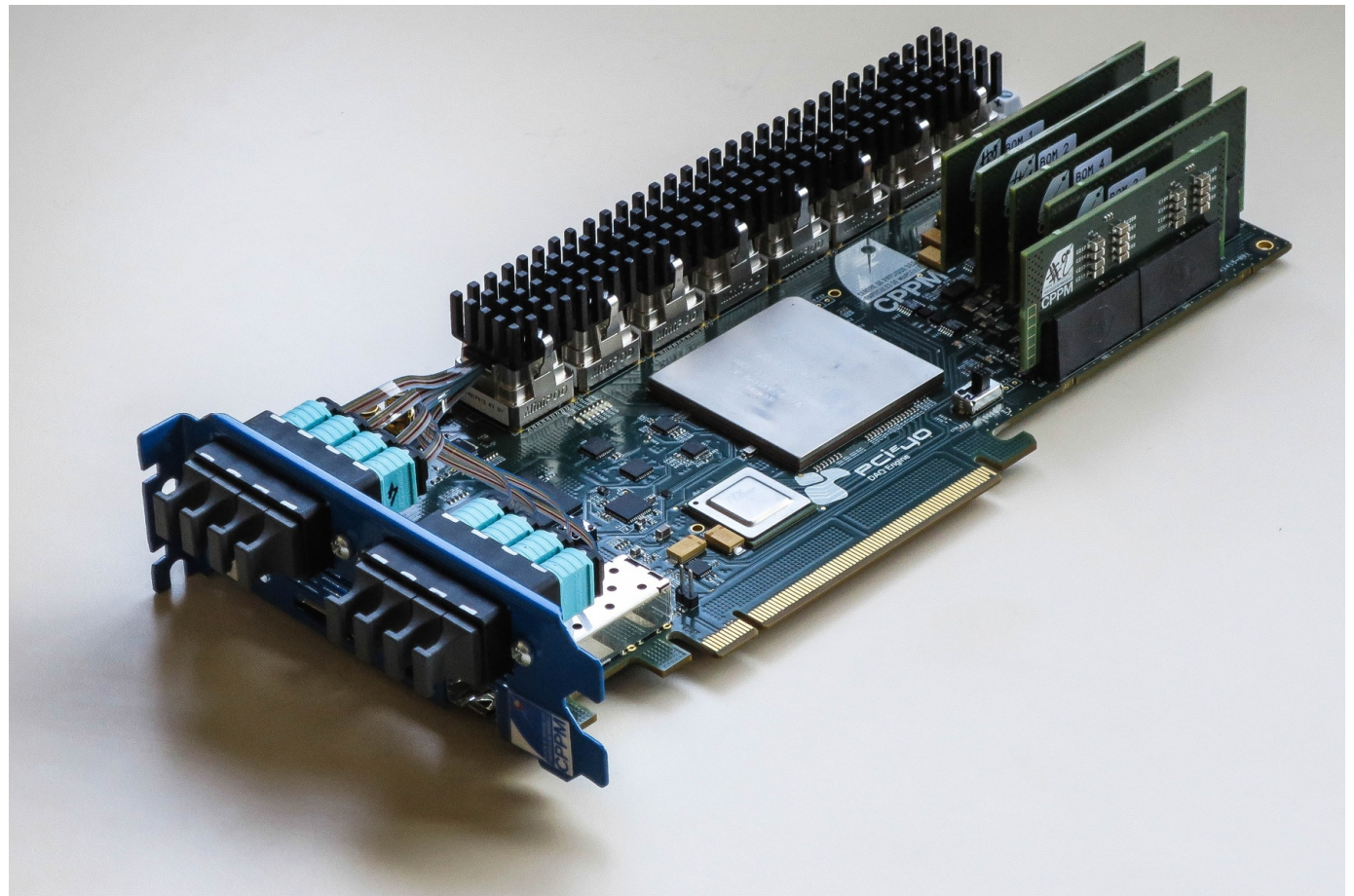
- Plus de crates intermédiaires
- Moins de liens optiques
- Beaucoup de mémoire dans le CPU → carte d'acquisition plus simple ou switches moins chers
- Possibilité de faire tourner le HLT dans les CPUs d'event building
 - Plus de 80 % de la puissance inoccupée

😞 Durée de vie du système

- Vie moyenne d'un PC = ~4 ans (jusqu'à 8 selon statistiques du CERN)
 - Que se passe-t-il si les slots cuivre PCIe disparaissent du PC ?
Risque de devoir redessiner et produire la carte PCIe40
 - Mitigation :
 - Prochaine génération encore compatible
 - Acheter des CPUs de réserve pour remplacer ceux qui tombent en panne
 - Utilisation de cartes mères avec slots mixtes

PCIE40

- Arria10 - Technologie 20nM - 1980 pins
- 1.15 millions de logic cells
- 72 liens 10 Gbits/s
- Bande passante : Optique 480 Gbits en entrée, 480 bits en sortie
PCIe 100 Gbits en entrée et en sortie



Schisme architectural ?

Choix effectués pour les systèmes de readout du CERN

	ALICE	LHCb	CMS	ATLAS
Hardware trigger	No	No	Yes	Yes
Software trigger input rate	50 kHz Pb-Pb 200 kHz p-Pb	30 MHz	500/750 kHz for PU 140/200	0.4 MHz
Baseline processing architecture	CPU/GPU/FPGA/ Cloud&Grid	CPU farm (+coprocessors)	CPU farm (+coprocessors)	CPU farm (+coprocessors)
Software trigger output rate	50 kHz Pb-Pb 200 kHz p-Pb	20-100 kHz	5-7.5 kHz	5-10 kHz

Conclusion

Tendances

- Migration de plus en plus de fonctions vers les fermes de calcul
 - Loi de Moore : le temps travaille pour les gens du online

	Event-size [kB]	Rate [kHz]	Bandwidth [Gb/s]	Year [CE]
ALICE	20000	50	8000	2019
ATLAS	4000	200	6400	2022
CMS	2000	200	3200	2022
LHCb	100	40000	32000	2019

Future DAQ in the LHC

Niko Neufeld, CERN

- Standards :
 - adoption progressive du standard xTCA par les expériences
 - ATCA pour ATLAS, μ TCA pour CMS
 - mais aussi du PCIe
 - LHCb, Alice
 - Coexistence probable des deux types de solutions