

Calibration de l'étiquetage des jets issus de quarks b dans ATLAS

Sébastien CORRÉARD

sous la direction d'Alexandre ROZANOV

- Introduction : le LHC et ATLAS
- L'étiquetage des jets b
- Calibration de l'étiquetage
- Le canal $tt H \rightarrow bb$

Motivations du LHC : le boson de Higgs²

Le modèle standard a été testé avec une grande précision (10^{-3} - 10^{-4}) par les accélérateurs récents (LEP, Tevatron...)

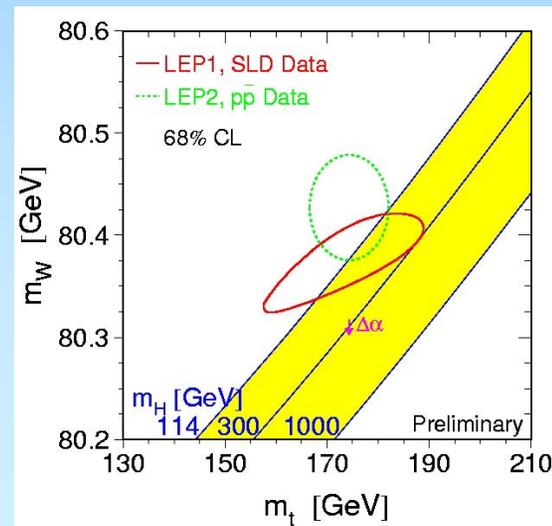
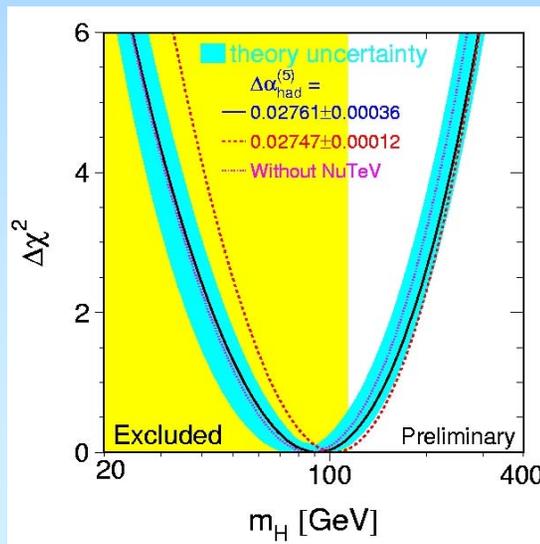
Masse des particules ?

→ Mécanisme de brisure spontanée de la symétrie électrofaible

→ Recherche du boson de Higgs : la pièce manquante

$m_H > 114.4 \text{ GeV}/c^2$ (LEP) et $m_H < 1 \text{ TeV}/c^2$ (théorie)

$m_H = 114^{+69}_{-45} \text{ GeV}/c^2$ (ajustement électrofaible)



Motivations du LHC : nouvelle physique³

absence de boson de Higgs → nouvelle physique à l'échelle du TeV

Le modèle standard, théorie effective à basse énergie d'une théorie plus fondamentale ?

→ Super symétrie, Théorie Grand Unifiée, ...

Questions ouvertes :

- Pourquoi 3 familles de fermions ?
- Famille additionnelle de quarks et leptons ?
- Gravité ?
- Déconfinement quark/gluon ?
- ...

Le LHC : Large Hadron Collider

Accélérateur en construction, installé dans l'ancien tunnel du LEP.
Mise en service prévue en 2007

Interactions p-p à 14 TeV (également Pb-Pb à 1.15 PeV)
Taux de croisement des faisceaux : 40 MHz

Luminosité :

2007-2009 : $L \sim 10^{33} \text{ cm}^{-2}\text{s}^{-1}$, soit $\int L dt \sim 10 \text{ fb}^{-1}$ par an

2009-20XX : $L \sim 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, soit $\int L dt \sim 100 \text{ fb}^{-1}$ par an



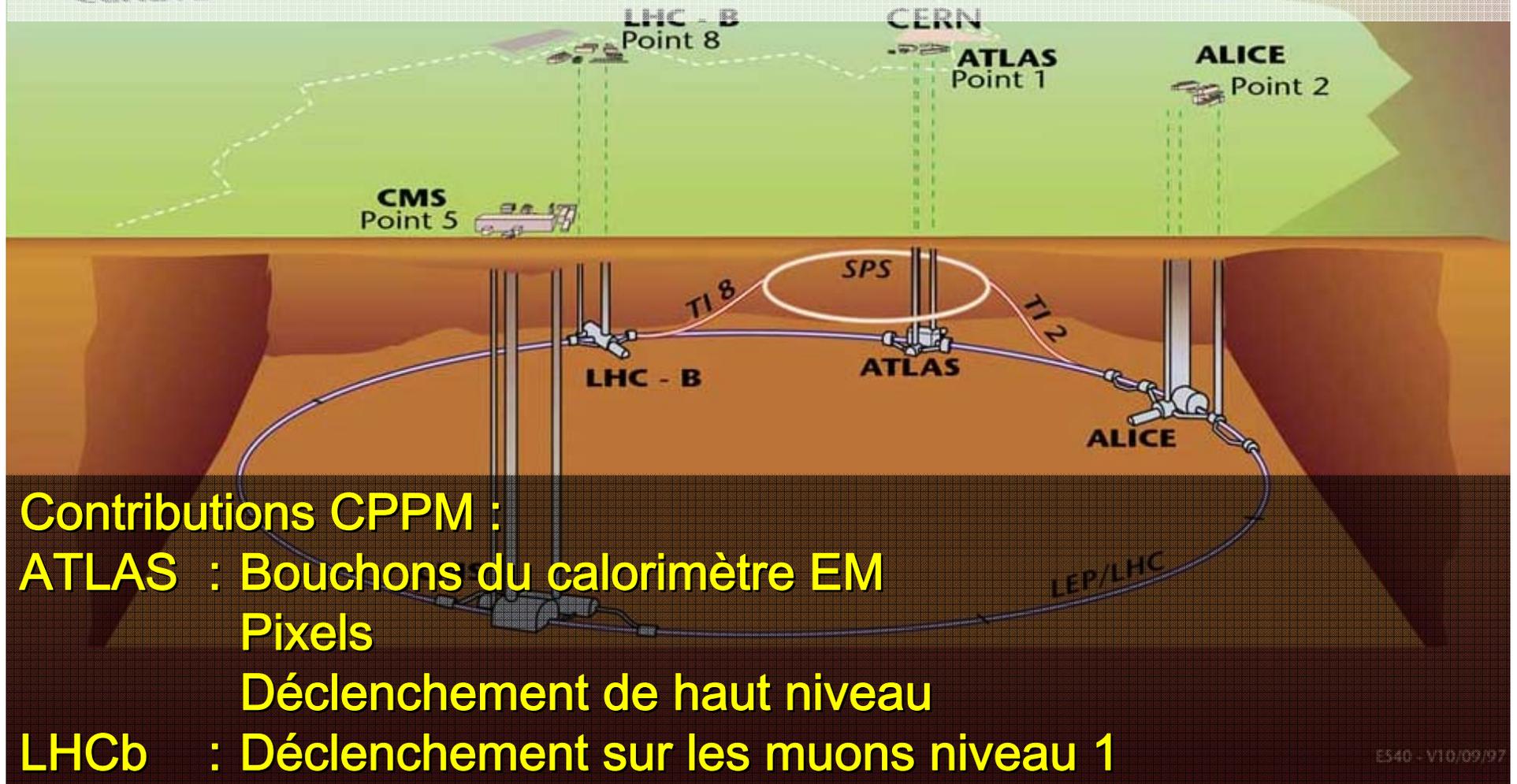
conçu pour découvrir le boson de Higgs $< 1 \text{ TeV}$
en cas d'absence, nouvelle physique $\sim \text{TeV}$

Les expériences auprès du LHC

ATLAS et CMS : généralistes

LHCb : violation CP et physique du b

ALICE : ions lourds



Contributions CPPM :

ATLAS : Bouchons du calorimètre EM
Pixels

Déclenchement de haut niveau

LHCb : Déclenchement sur les muons niveau 1

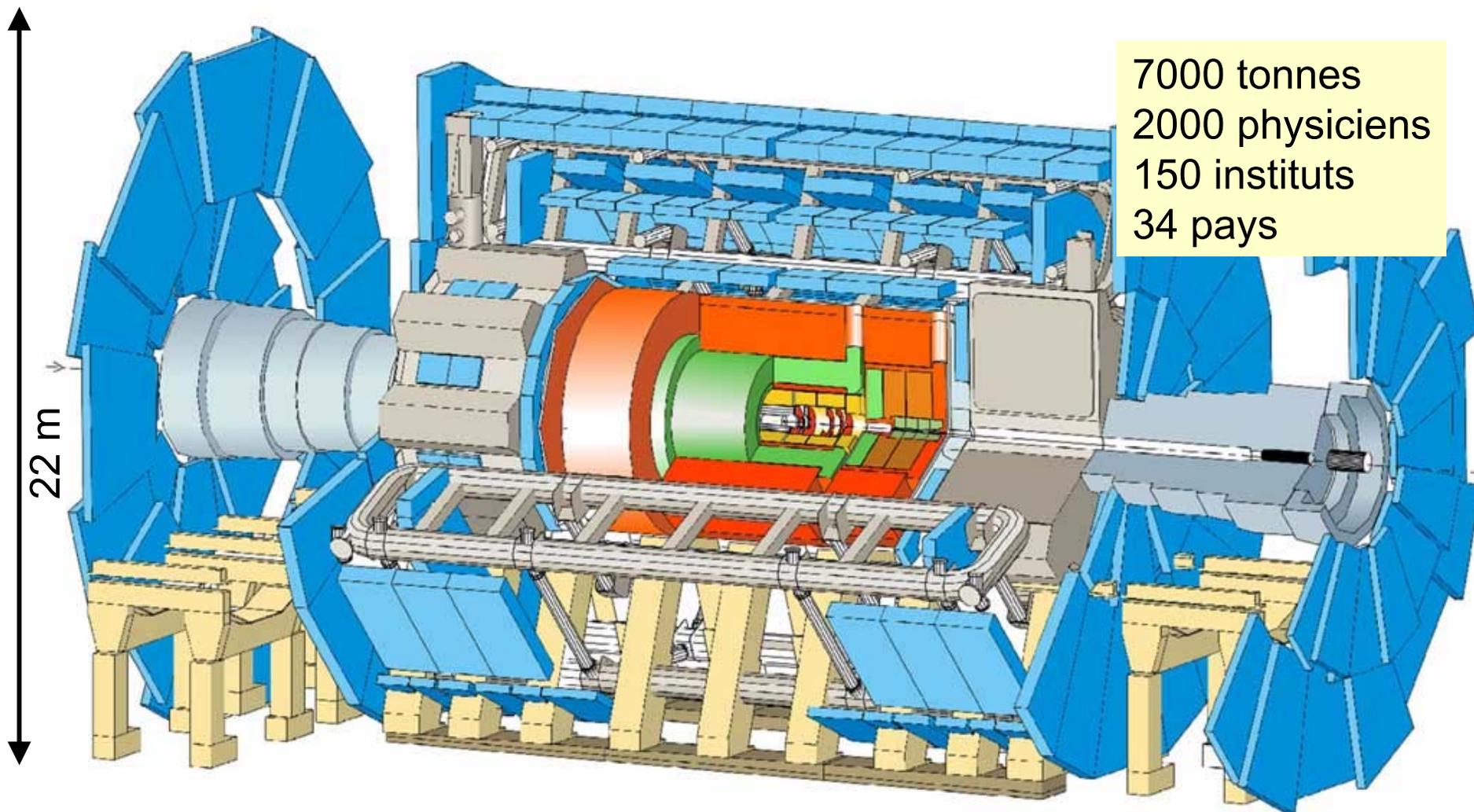
ATLAS

6

40 m

22 m

7000 tonnes
2000 physiciens
150 instituts
34 pays



ATLAS

7

Trajectographe interne

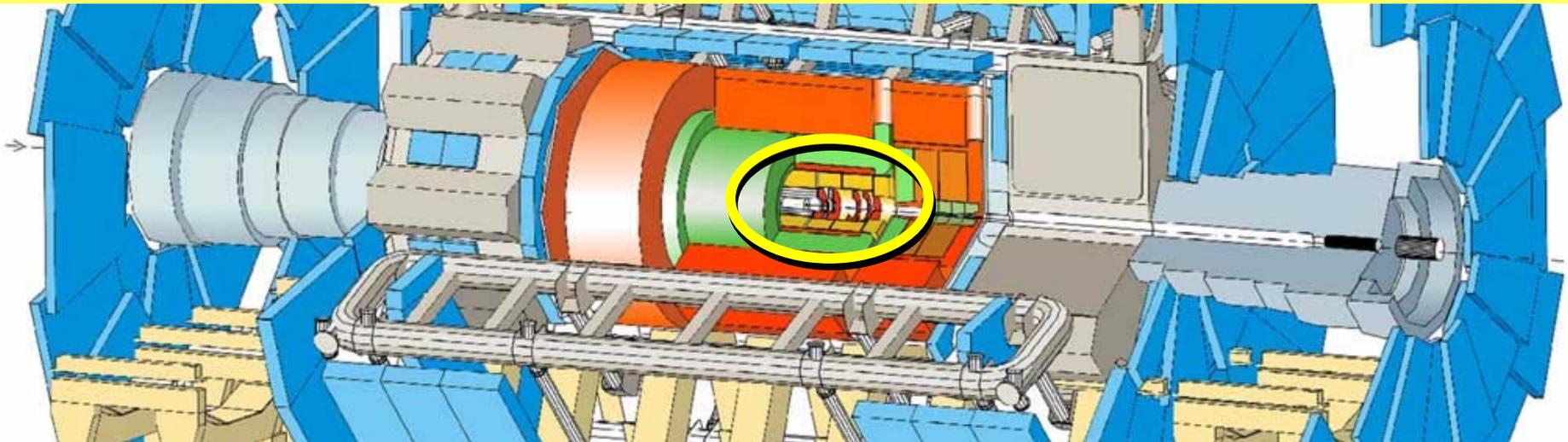
Pixels : détecteur à silicium

SCT : micropistes silicium

TRT : trajectographe à pailles avec
fonction radiation de transition

} dans un aimant solénoïde

tt $H \rightarrow bb$



- trajectoires des particules chargées
- paramètres des traces : paramètre d'impact, impulsion, ...
- reconstruction des vertex
- séparation électrons/pions (radiation de transition)

ATLAS

Calorimètre

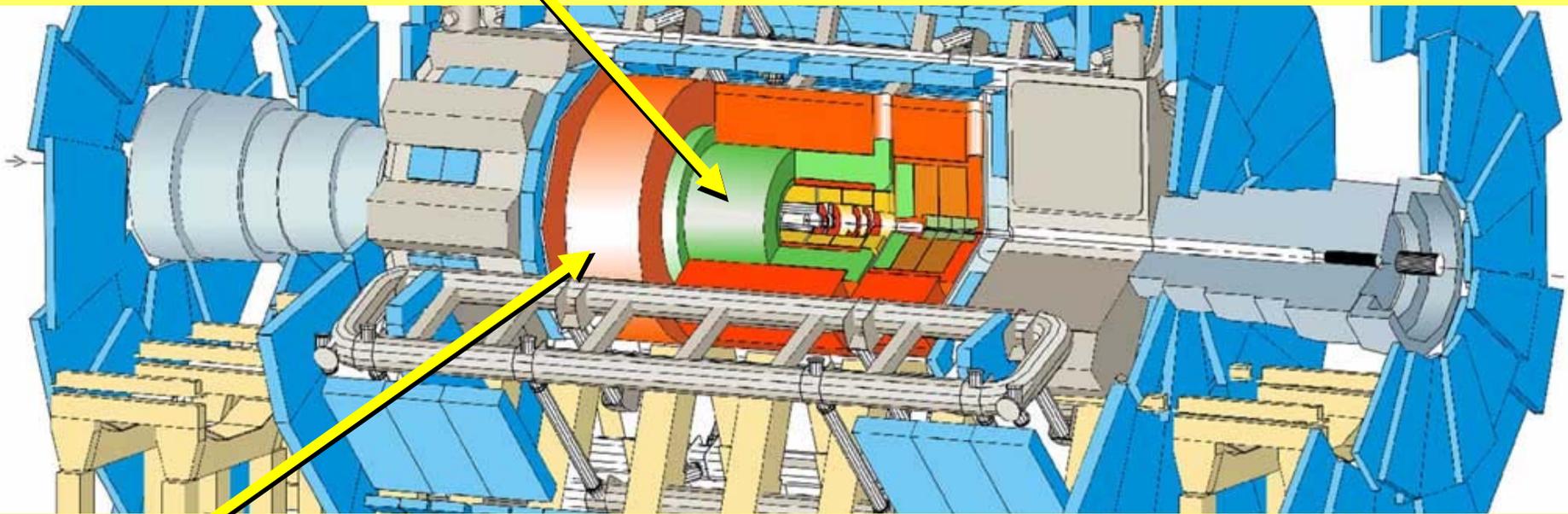
Électromagnétique : calorimètre à échantillonnage Pb/Ar, dans un cryostat

- mesurer l'énergie et la direction des gerbes électromagnétiques
- séparation γ/π^0

$$\frac{\sigma_E}{E} = \frac{0.1}{\sqrt{E}} \oplus \frac{0.3}{E} \oplus 0.01$$

$H \rightarrow ZZ^{(*)} \rightarrow 4e$

$H \rightarrow \gamma\gamma$



Hadronique : calorimètre acier/scintillateur

- déterminer l'impulsion transverse manquante
- mesurer la contribution hadronique de l'énergie des jets

$$\frac{\sigma_E}{E} = \frac{0.5}{\sqrt{E}} \oplus 0.03$$

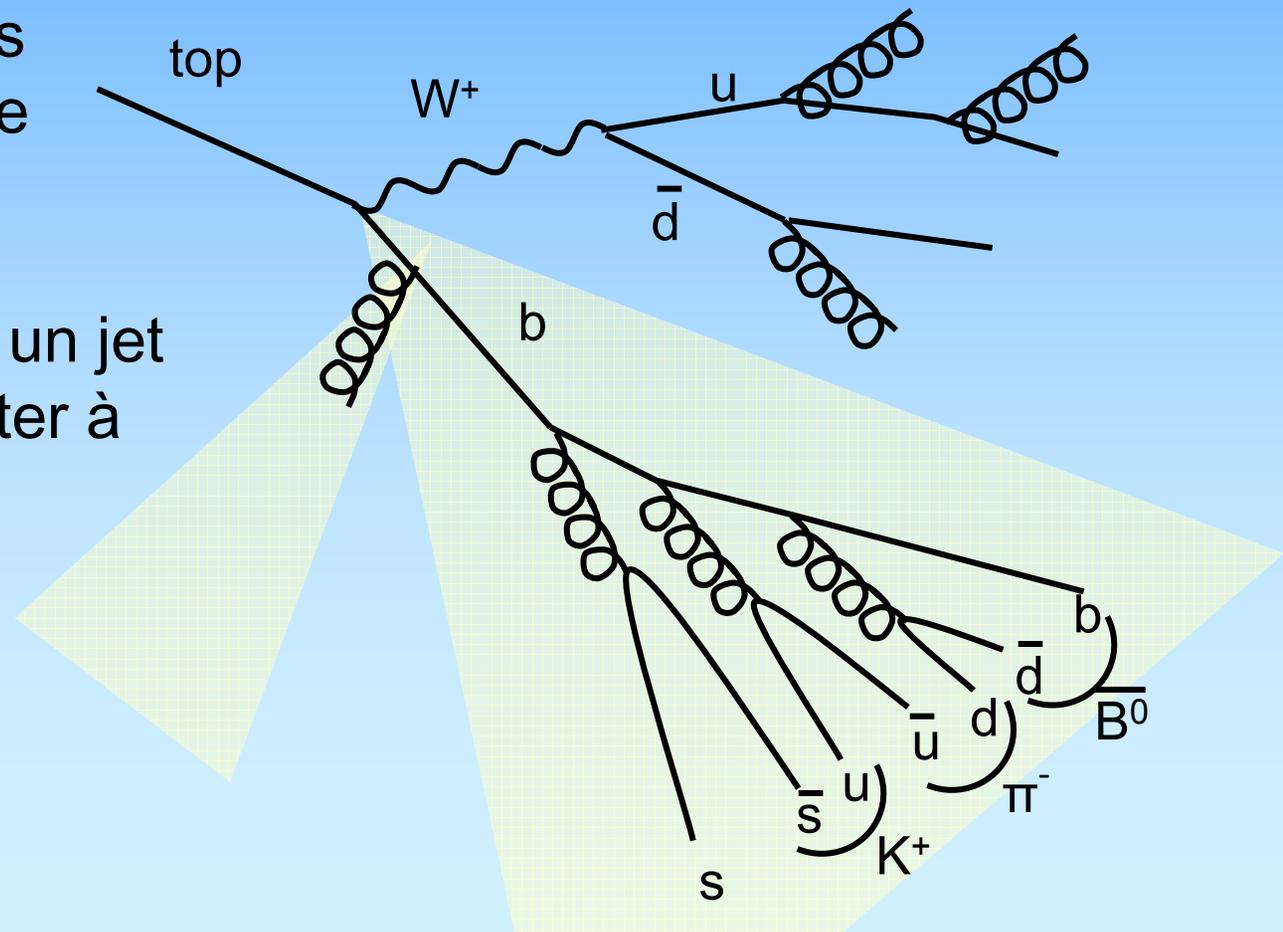
ATLAS

Calorimètre

Permet de déterminer l'énergie et la direction des jets

Le jet contient des informations sur le parton initial

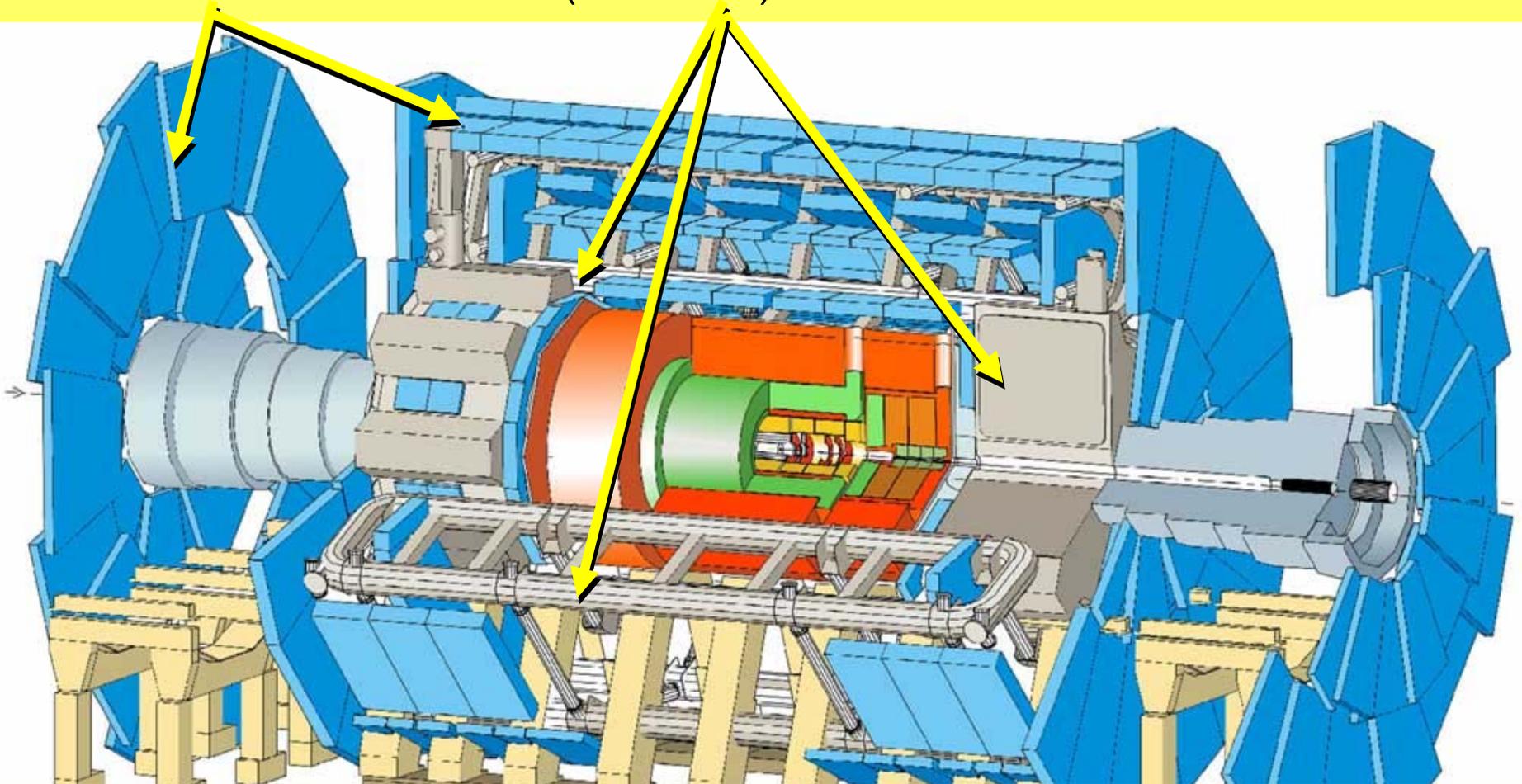
Bien reconstruire un jet permet de remonter à cette information



ATLAS

Spectromètre à muons

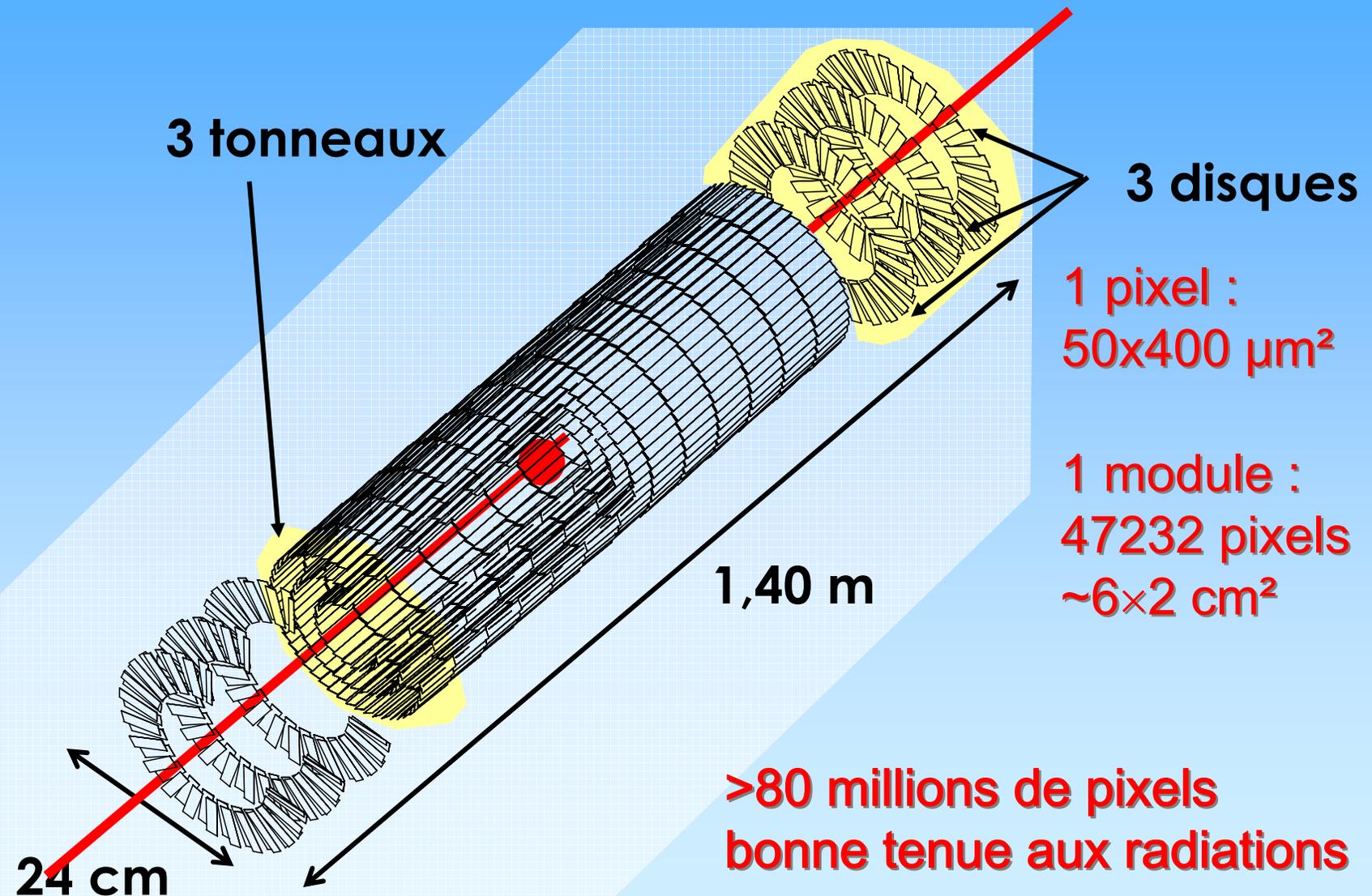
Chambres à dérive et aimants (toroïdaux)



- détecter les muons
- mesurer leur impulsion

$$H \rightarrow ZZ^{(*)} \rightarrow 4\mu$$

Le détecteur à pixels

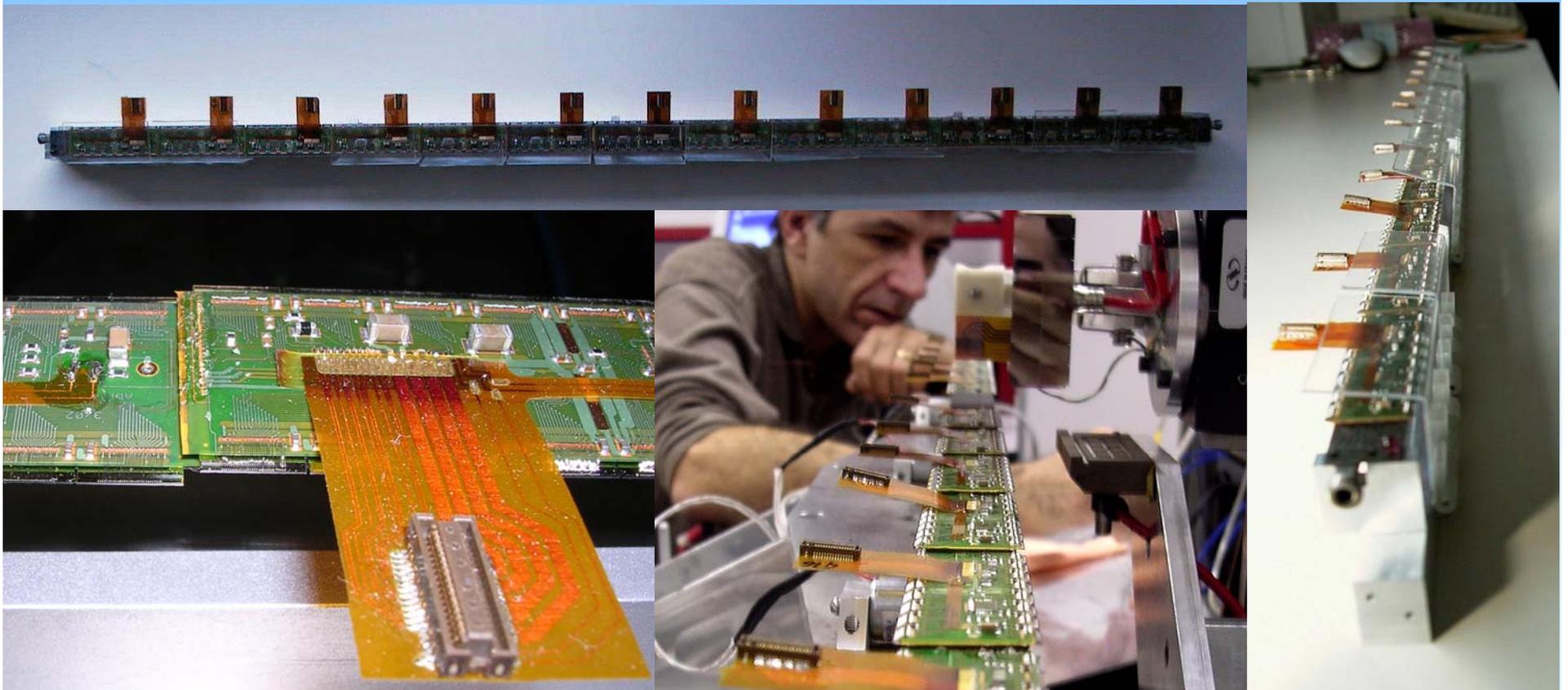


⇒ Reconstitue les vertex et les trajectoires des particules chargées

Le détecteur à pixels

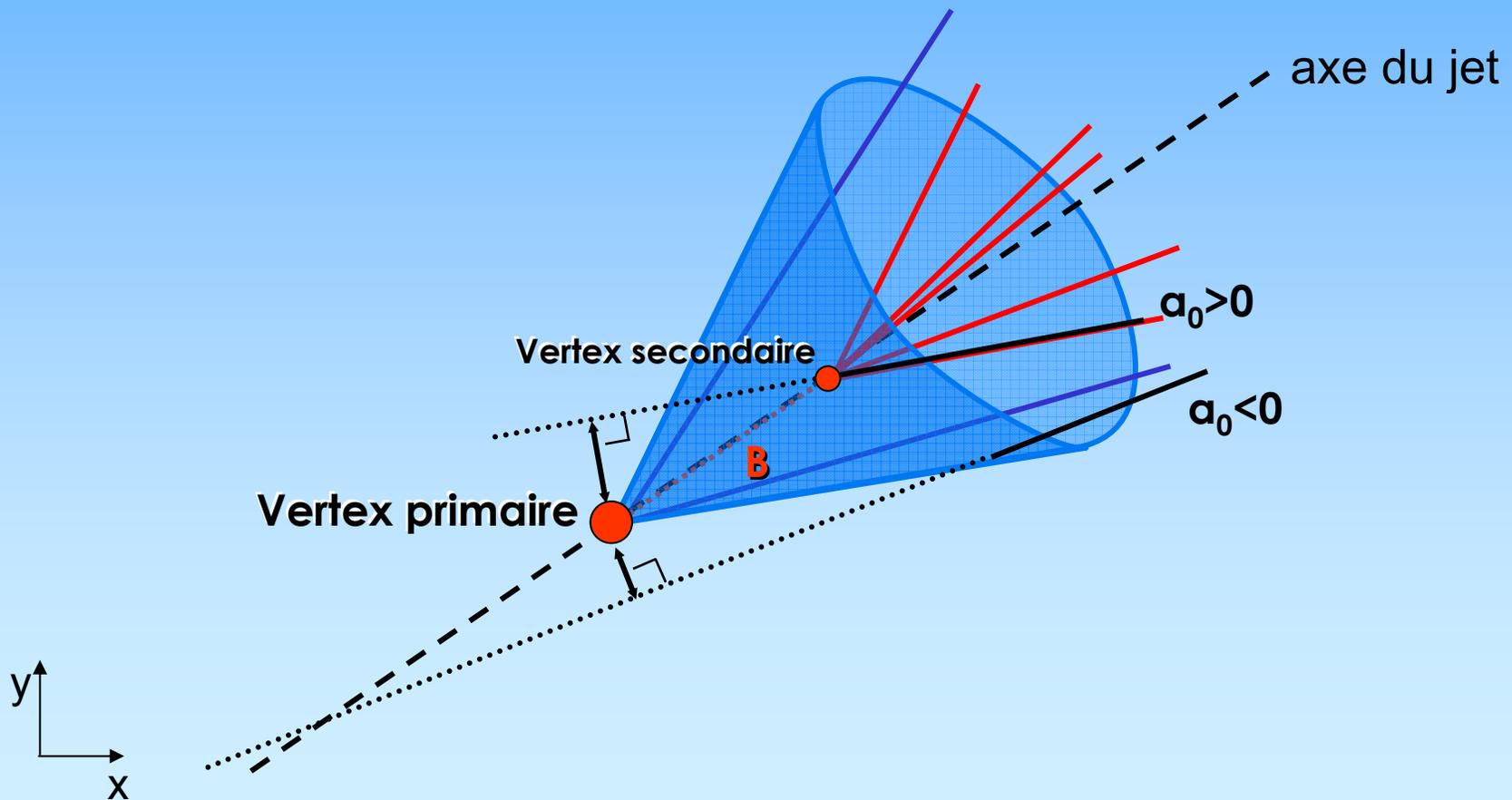
Implication du CPPM :

- chargé de construire 1/3 du tonneau (avec Allemagne et Italie)
- 3 échelles déjà produites, dont 1 avec l'électronique finale
- production lancée



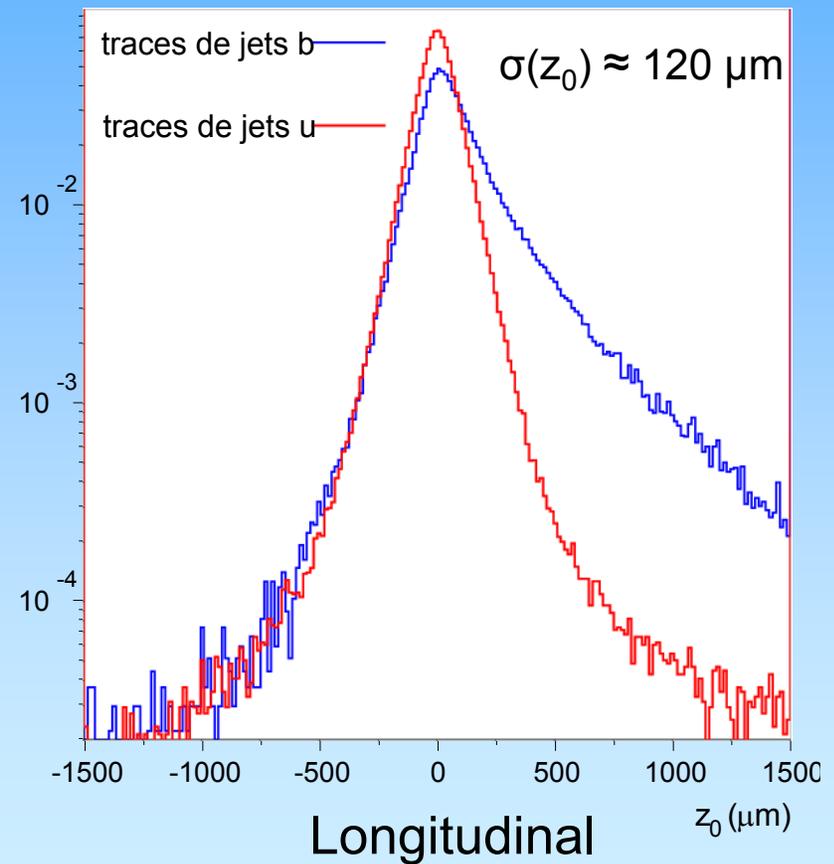
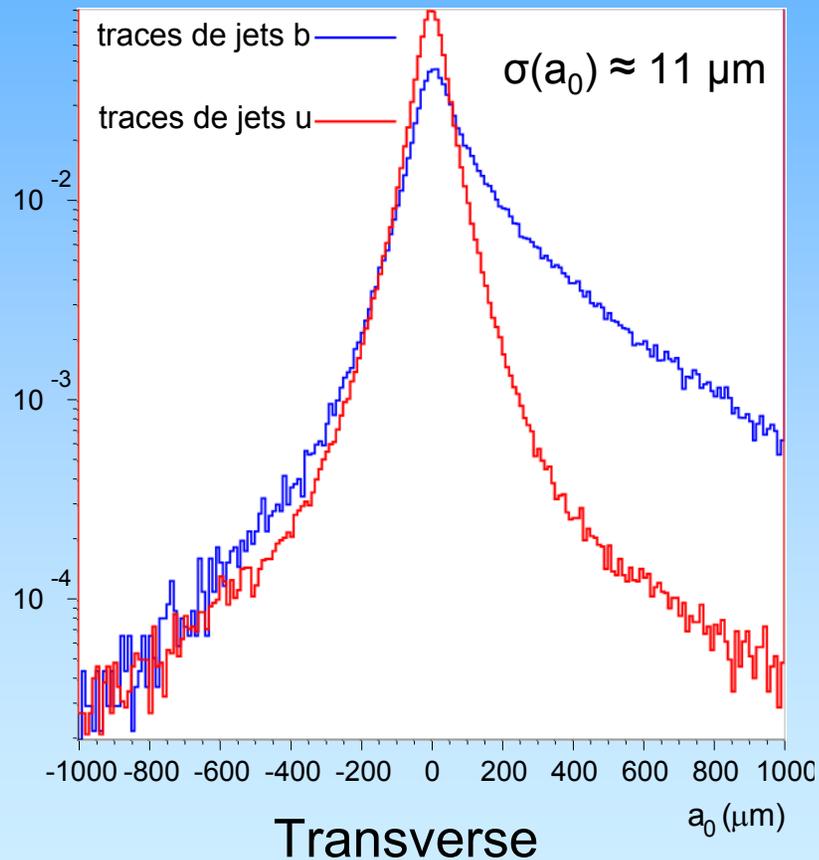
Paramètre d'Impact

- Définition du paramètre d'impact signé



Paramètre d'Impact

- Distribution du paramètre d'impact



Méthodes d'étiquetage des jets b ¹⁵

- **Durée de vie des hadrons beaux :**

$\tau \sim 1.5$ ps, si $E=50$ GeV, distance de vol ~ 5 mm

- Paramètre d'impact 2D et 3D
- Vertex Secondaire

- **Leptons mous ($p_T \sim 10$ GeV/c) non isolés :**

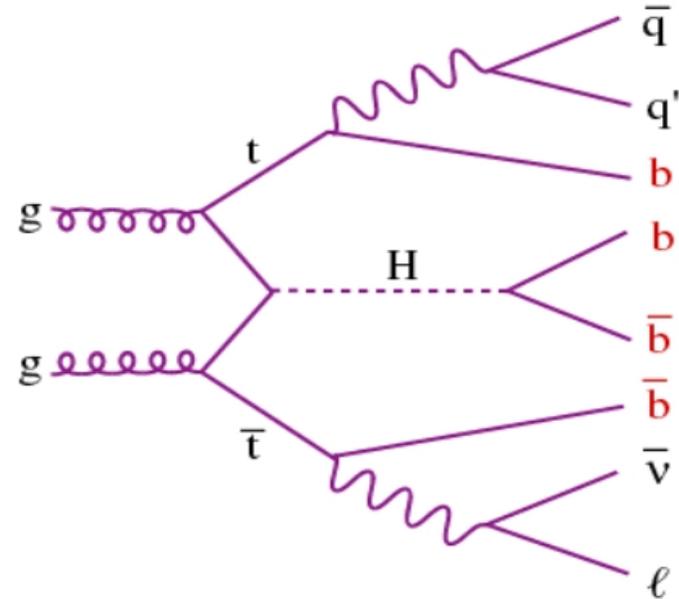
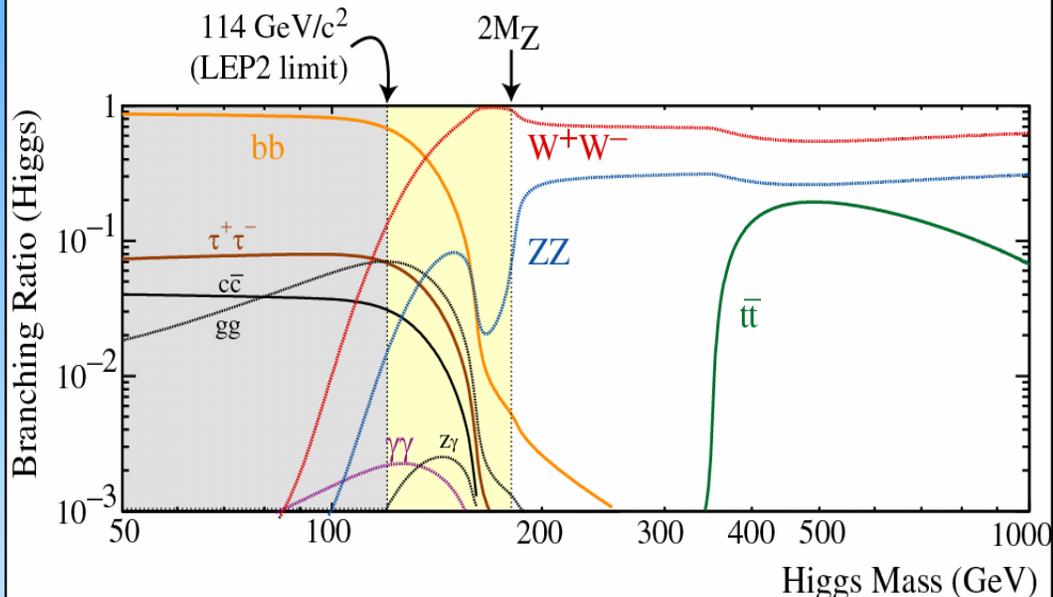
- électrons/muons issus de $B \rightarrow D \rightarrow \ell$

Canaux sensibles à l'étiquetage

- tt → $\sigma(tt)=833\text{pb}$, physique du top : m_{top} , Γ_{top} , corrélation de spin...
 $m_{\text{top}} \rightarrow$ contraintes sur m_H
- tt H→bb → recherche d'un boson de Higgs léger
 $114 \text{ GeV}/c^2 < m_H < 140 \text{ GeV}/c^2$
- W H →bb → recherche du boson de Higgs
- SUSY → recherche du \tilde{t} et du \tilde{b} , du h→bb MSSM
 $(m_h < 130 \text{ GeV}/c^2)$

le canal $t\bar{t} H \rightarrow bb$: signal

Rapport d'embranchement du boson de Higgs



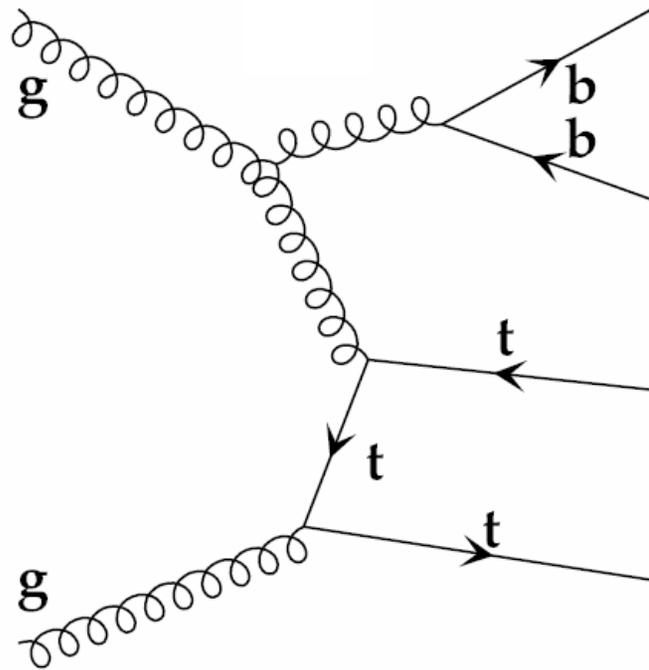
Un top $\rightarrow \ell\nu b$ (déclenchement)

Un top $\rightarrow jjb$

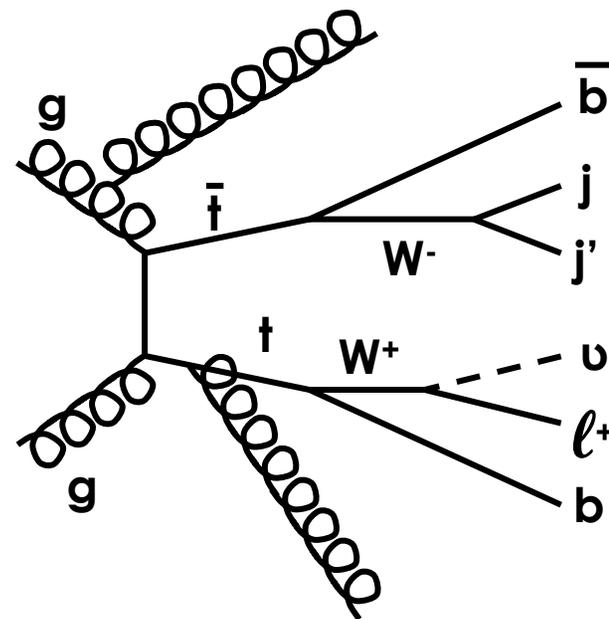
État final complexe : 6 jets, 1 ℓ , E_t^{miss} + jets ISR/FSR
 Requiert l'étiquetage de **4 jets b** $\rightarrow (\epsilon_b)^4$

le canal $tt H \rightarrow bb$: bruit de fond

| processus | $tt(H \rightarrow bb)$ (120 GeV/c ²) | ttbb | ttjj |
|-----------------------------|-----------------------------------------------------|------|------|
| σ_{incl} (pb) | 0.36 | 9.5 | 57 |



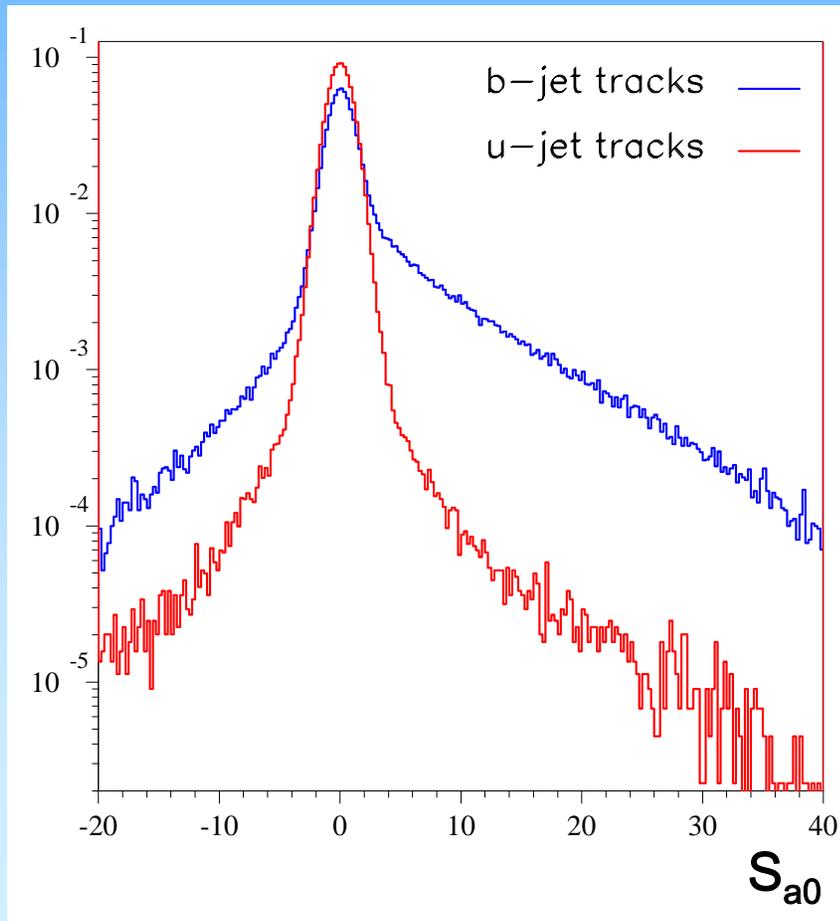
Bruit de fond irréductible ttbb



Bruit de fond réductible ttjj

Étiquetage des jets b

- **Distribution de signification** $S(a_0) = \frac{a_0}{\sigma(a_0)}$ des traces
(pour donner plus de poids aux traces les mieux mesurées)



Lissage des distributions



densités de probabilité
b(S) et u(S)

poids d'un jet :

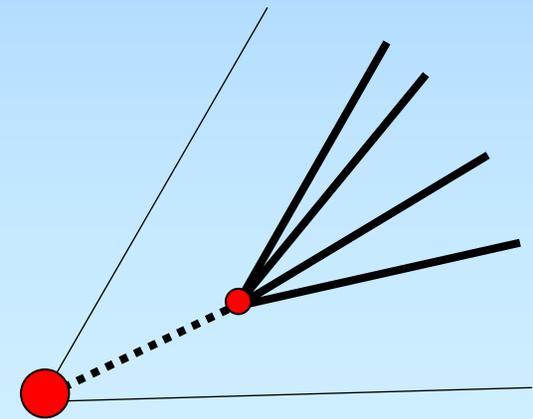
$$W_{2D} = \sum_{\text{traces}} \ln \left(\frac{b(S)}{u(S)} \right)$$

Vertex Secondaire

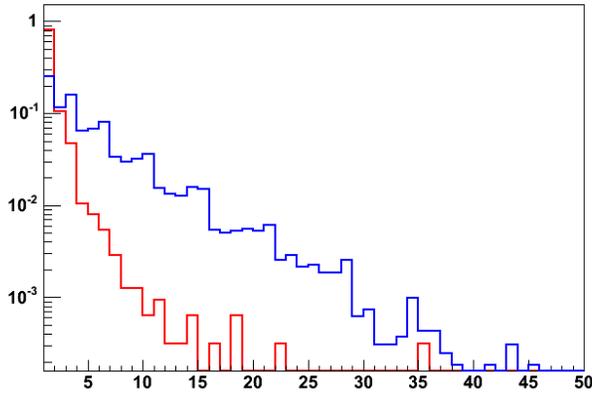
- Reconstruction explicite du vertex secondaire (SVX)
⇒ 3 variables supplémentaires :
 - Masse reconstruite du vertex secondaire : M
 - Rapport $E_{\text{SVX}}/E_{\text{jet}}$: F
 - Nombre de vertex secondaires à 2 traces : N

On ajuste la distribution de ces 3 paramètres par les fonctions $u(M,F,N)$ et $b(M,F,N)$

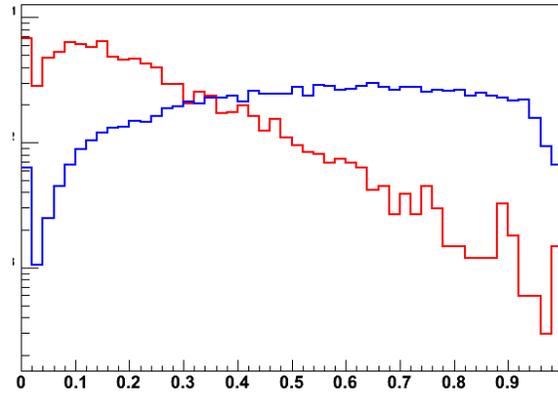
$$\rightarrow w_{\text{SVX}} = \ln \left(\frac{b(M,F,N)}{u(M,F,N)} \right)$$



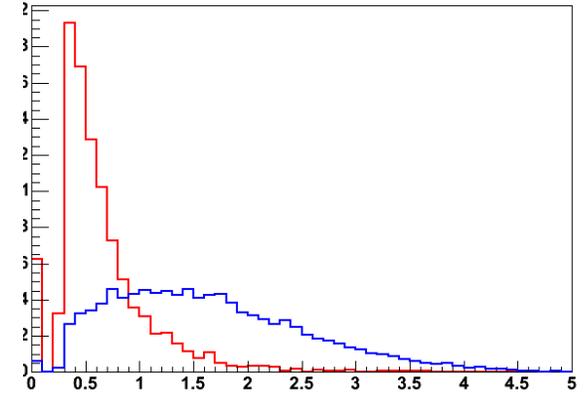
Méthode combinée



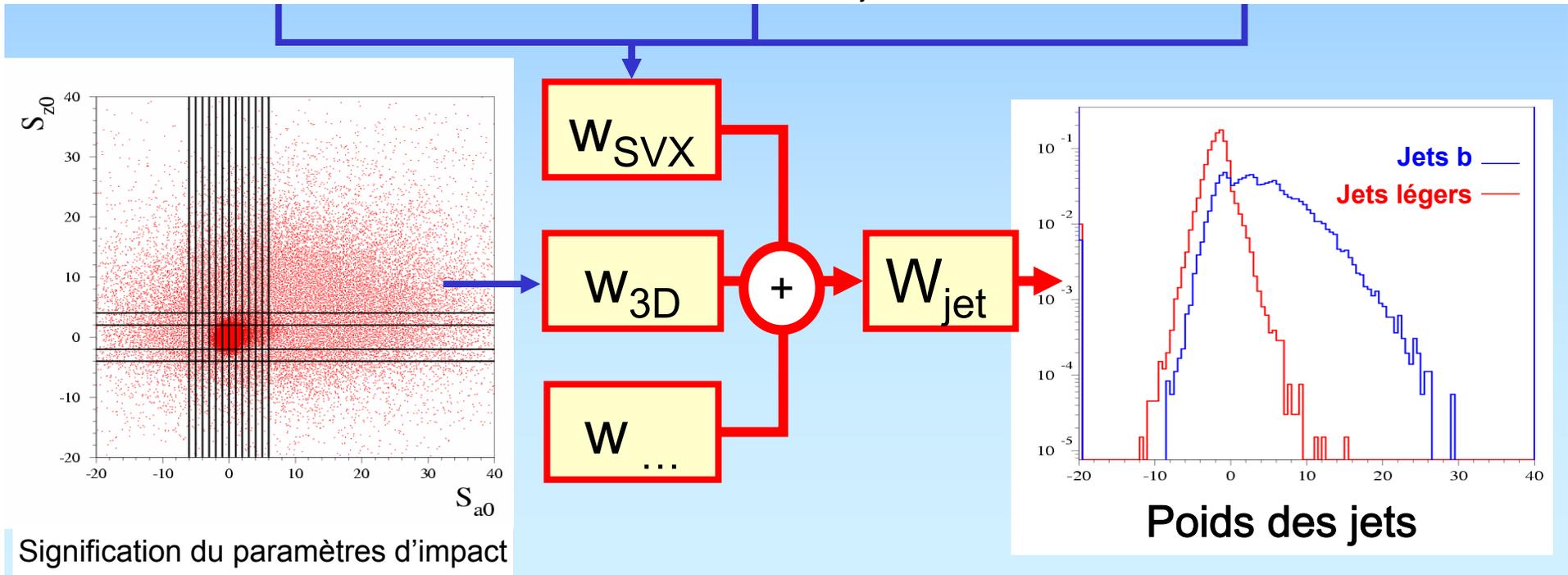
Nb de vertex à 2 traces



Rapport E_{SVX}/E_{jet}

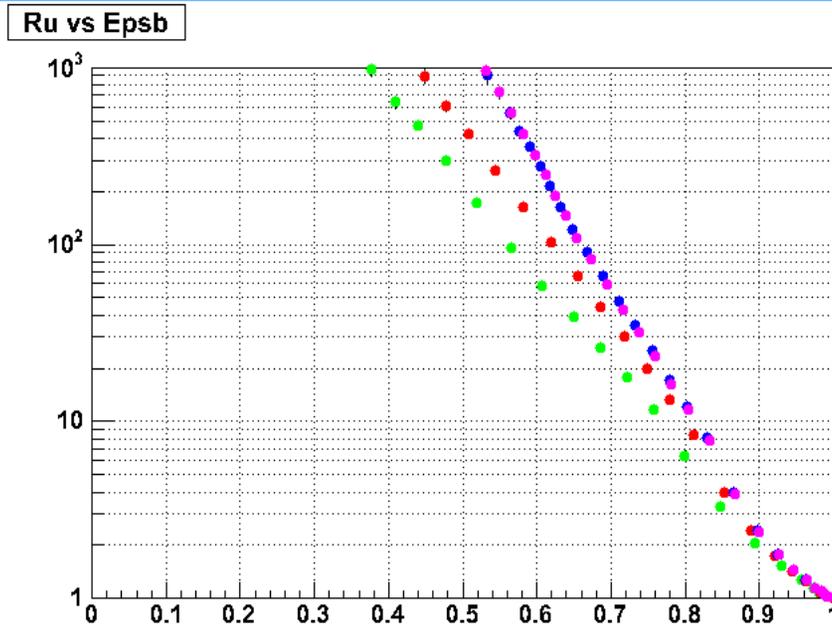


Masse du SVX



Définition des jets b

- Poids du jet → étiquetage de la saveur du jet



Efficacité de sélection :

$$\varepsilon_b = \frac{N_{b \text{ sélectionnés}}}{N_{b \text{ total}}} = 60\%$$

Facteur de rejet des jets légers :
(u, d, s ou g)

$$R_{\text{léger}} = \frac{N_{\text{léger total}}}{N_{\text{léger sélectionnés}}} \approx 150$$

Effets réalistes

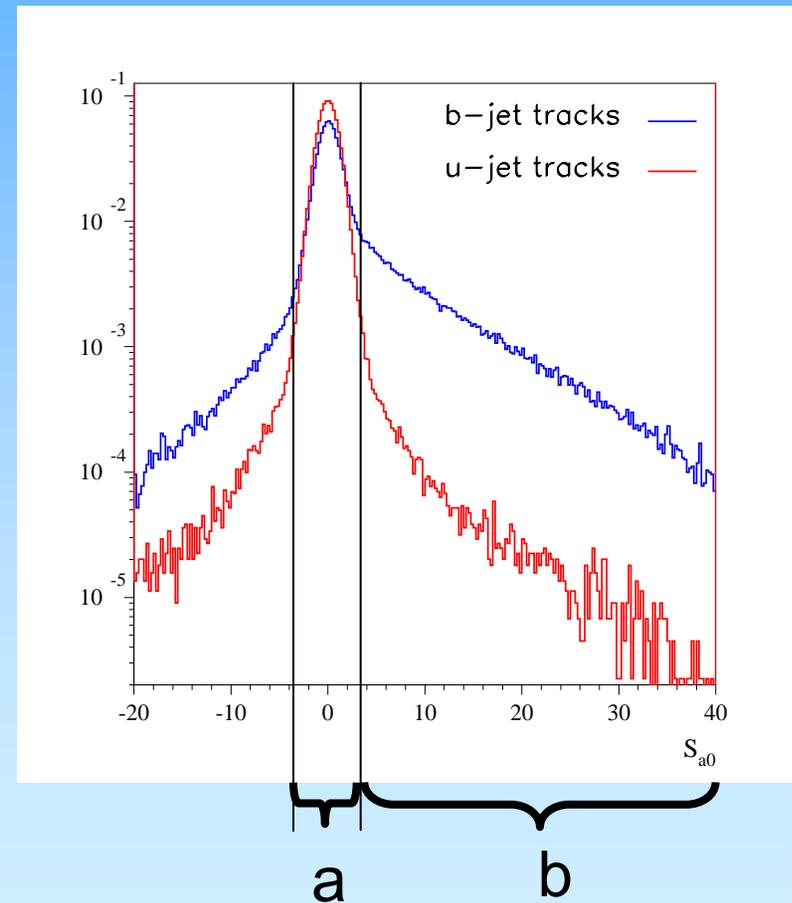
L'algorithme d'étiquetage tient compte des effets « réalistes » suivants :

- Inefficacité des pixels : 3%
- Inefficacité des modules et chips : 1 et 2%
- détecteur initial (2 couches de pixels)
- bruit d'empilement des événements de biais min.
- descriptions détaillées de la matière et de l'électronique
- effets d'alignement

Calibration de l'étiquetage des jets b ²⁴

Calibration = construction des fonctions de densité de probabilité

1. **Calibration sur simulations MC**
description du détecteur
2. **Calibration sur les données**
exemples :
 - a. impact de la résolution
 - b. propriétés physiques des jets b

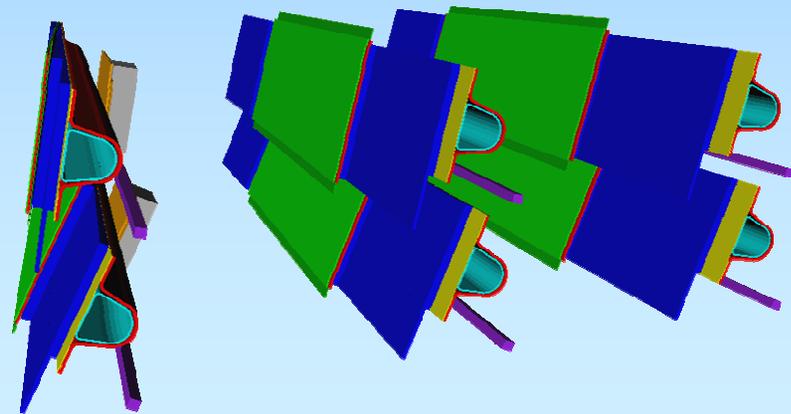
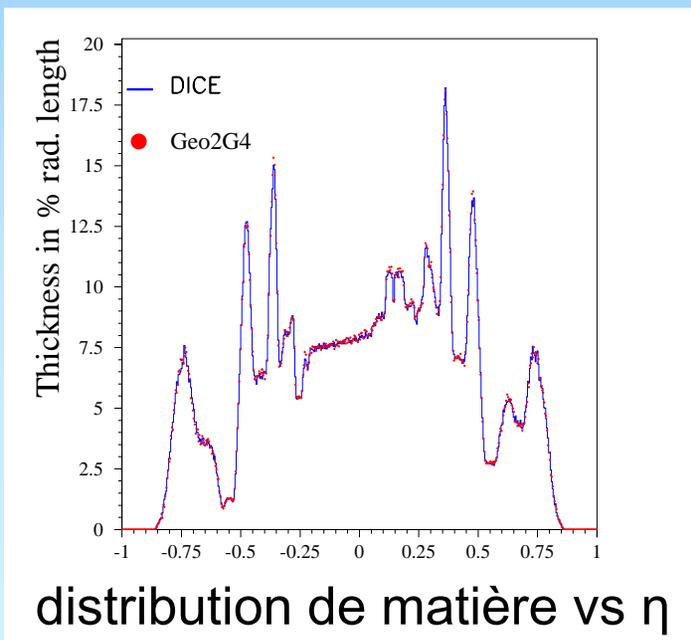


Calibration sur simulations MC

Une simulation aussi proche que possible de la réalité est nécessaire pour construire de bonnes courbes de calibration

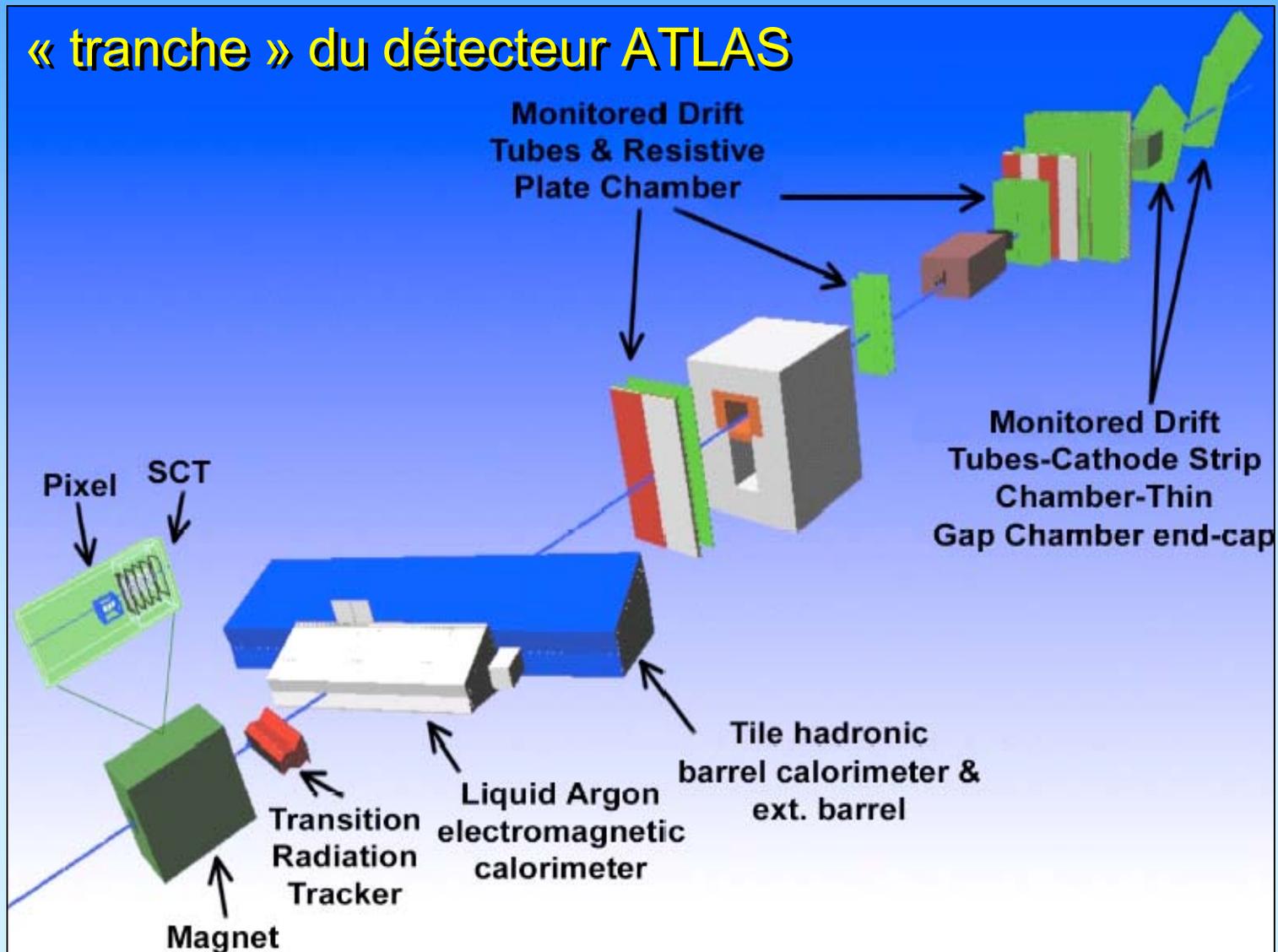
- ⇒ Bon ajustement du MC grâce aux tests en faisceau
- ⇒ Bonne description du détecteur

Développement de la description du détecteur :



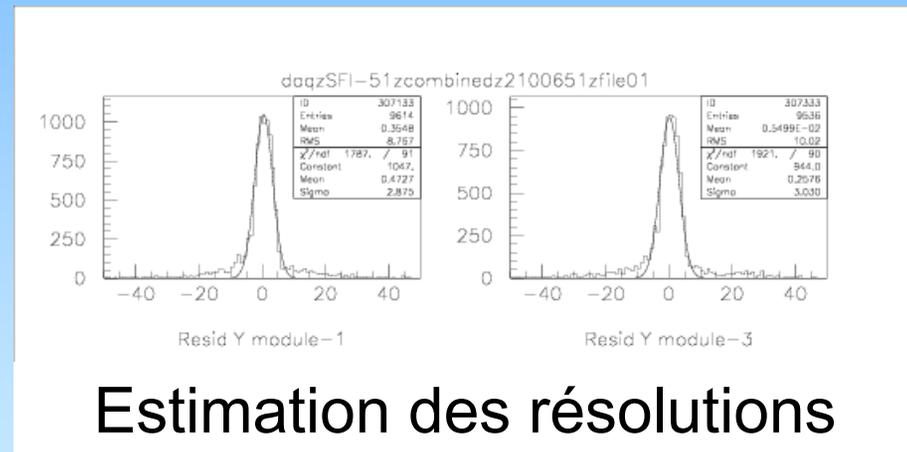
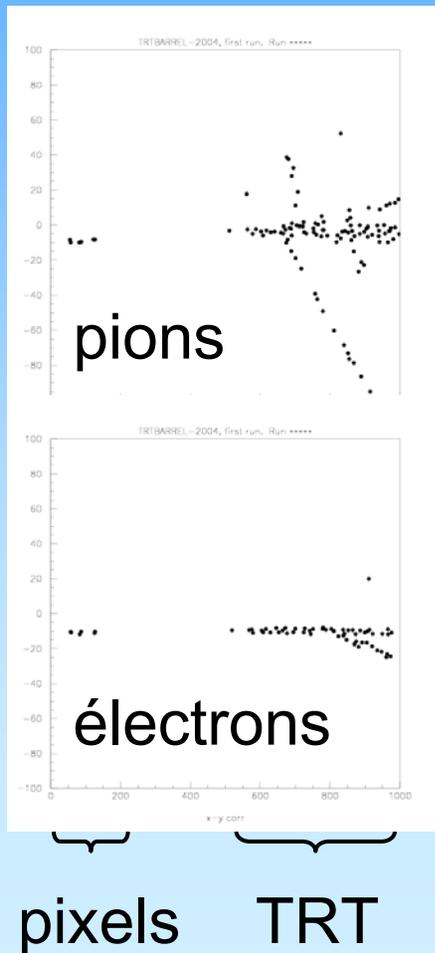
Validation description : tests faisceau

Été-automne 2004 : tests combinés en faisceau



Validation description : tests faisceau

Résultats des tests combinés en faisceau



Vérification/ajustement MC

Étiquetage : impact de la résolution

Exemple de dégradation des résolutions sur les traces :

$$\sigma_{a_0} \rightarrow \sigma'_{a_0} = \sqrt{\sigma_{a_0}^2 + (30 \mu\text{m})^2}$$

événements ttH et tt

$$\sigma_{z_0} \rightarrow \sigma'_{z_0} = \sqrt{\sigma_{z_0}^2 + (150 \mu\text{m})^2}$$

Impact de la dégradation de la résolution sur les traces :

| | | |
|-------------------------------------------------------|----------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| résolution sur les traces | σ_{a_0} original σ_{z_0} original | $\sigma'_{a_0} = \sqrt{\sigma_{a_0}^2 + (30 \mu\text{m})^2}$ $\sigma'_{z_0} = \sqrt{\sigma_{z_0}^2 + (150 \mu\text{m})^2}$ |
| performances d'étiquetage pour $\varepsilon_b = 60\%$ | $R_u = 352 \pm 20$ | $R_u = 168 \pm 6$ |

-50 %

Calibration : impact de la résolution

Impact de la calibration sur les performances d'étiquetage :

| | |
|-----------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| <p>résolution sur les traces</p> | $\sigma'_{a_0} = \sqrt{\sigma_{a_0}^2 + (30 \mu\text{m})^2}$ $\sigma'_{z_0} = \sqrt{\sigma_{z_0}^2 + (150 \mu\text{m})^2}$ |
| <p>calibration ne correspondant pas à la réalité (résolutions originales)</p> | $R_u = 143 \pm 4$ |
| <p>calibration correspondant à la réalité (résolutions dégradées)</p> | $R_u = 168 \pm 6$ |

$\epsilon_b = 60\%$

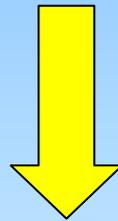


➡ Important d'avoir de bonnes résolution, mais il faut aussi les connaître

Impact de la résolution

Résumé :

- avoir une bonne résolution sur les traces est la priorité
- à résolution donnée, une bonne calibration améliore les performances d'étiquetage

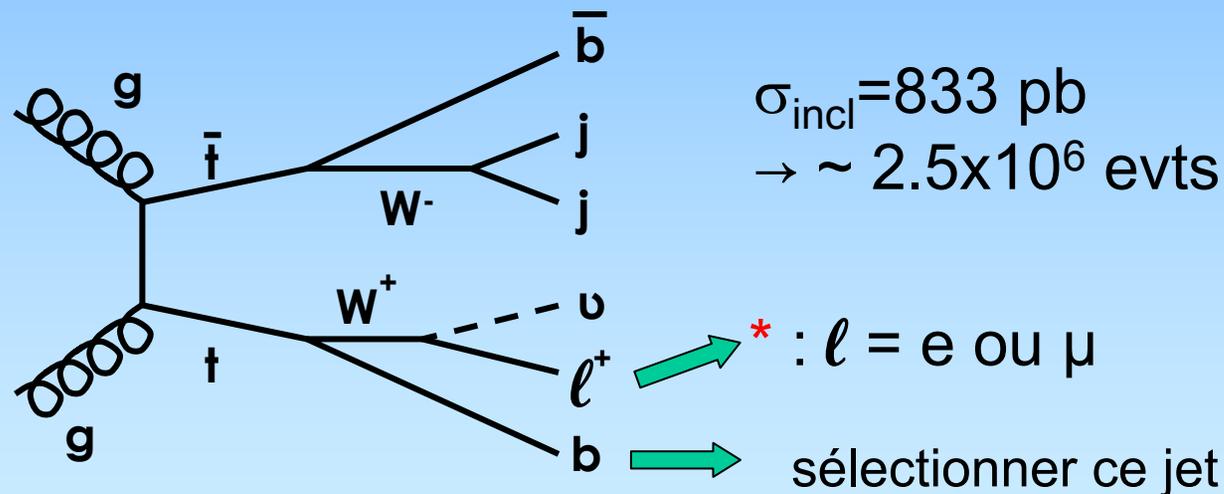


On pourrait extraire les résolutions des données réelles, et produire des courbes de calibration adaptées à ces résolutions.

Calibration : propriétés physiques des jets b

Canal tt : bruit de fond important pour l'étude ttH, mais peut être mis à profit pour sélectionner des jets b dans les données

Canal semileptonique* : pour déclenchement, reconstruction plus facile, et peu de fond

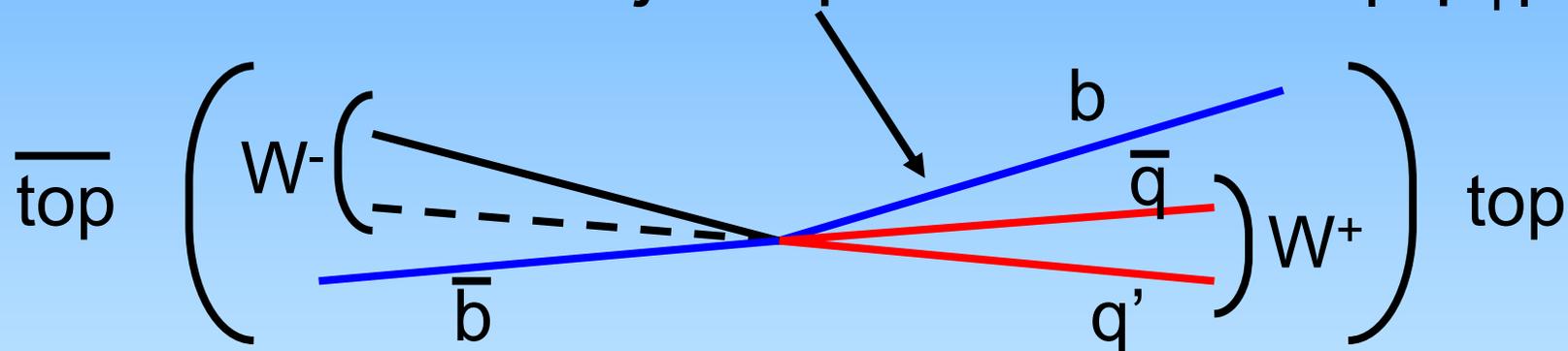


Deux quarks b dans l'état final. Idée : en étiqueter un, et sélectionner l'autre à l'aide d'une reconstruction cinématique.

Calibration : propriétés physiques des jets b

Reconstruction complète de l'événement :

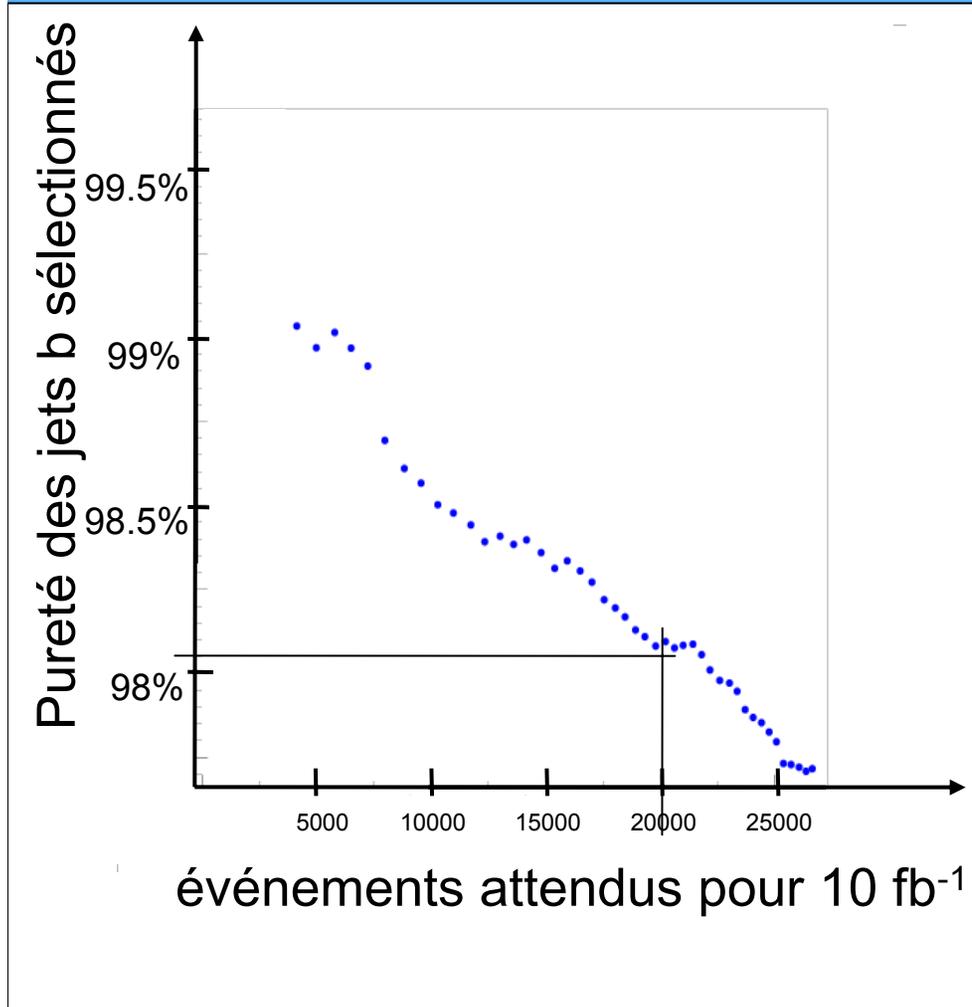
- Recherche de 3 jets qui maximisent $|\Sigma \vec{p}_T|$



- Un jet étiqueté b parmi ces 3 jets
- Fit contraint \rightarrow coupure sur χ^2

On sait quels sont les jets légers et quels sont les jets b
 \Rightarrow sélection d'un échantillon de jets b

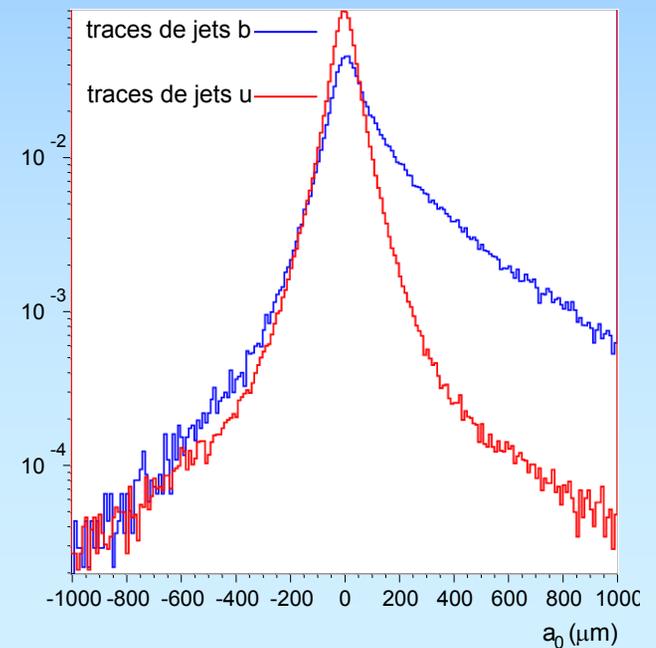
Calibration : propriétés physiques des jets b



Reconstruction des courbes de calibration



Échantillon de jets b



Le canal $t\bar{t} H \rightarrow b\bar{b}$

- Continuation de l'analyse précédente
- Hypothèse : $m_H = 120 \text{ GeV}/c^2$
- Deux types d'analyse : coupures et likelihood (LH)
- Résultats sur événements de simulation rapide
- « vrai » étiquetage sur événements de simulation détaillée

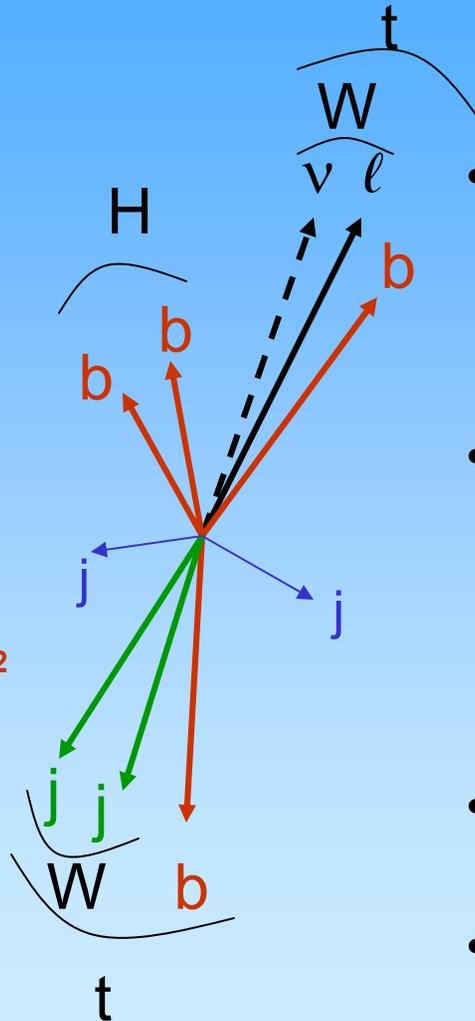
Deux analyses

Méthode des coupures

- Déterminer p_v^z avec la contrainte m_W (échec dans 25% des cas)
- Trouver la combinaison de 2 jets légers, 4 jets b qui minimise

$$\Delta = (m_{\ell vb} - m_{top})^2 + (m_{jjb} - m_{top})^2$$

- Garder les événements où m_{jjb} , $m_{\ell vb}$ consistant avec m_t , et m_{bb} consistant avec m_H

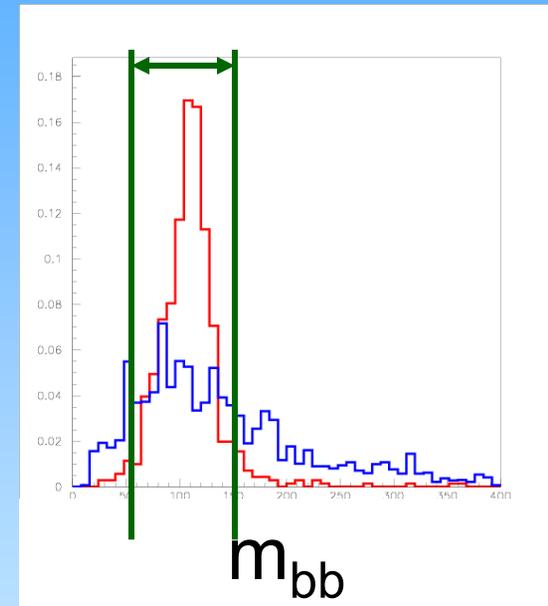
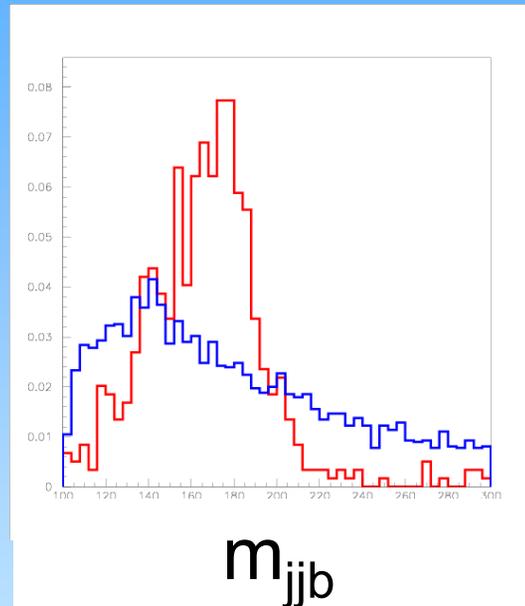
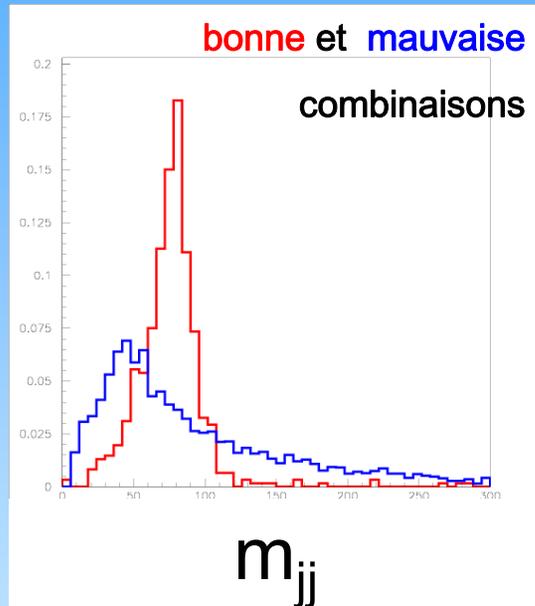


Méthode Likelihood

- Déterminer p_v^z avec la contrainte m_W (si échec, approximation $p_v^z = p_\ell^z$)
- Utilise un **likelihood** à 6 variables pour trouver la meilleure combinaison
- Garder les événements où m_{bb} consistant avec m_H
- Utilise **likelihood** à 8 variables pour séparer encore plus le signal du fond

Analyse en coupures

Événements Atfast (simulation rapide),
étiquetage aléatoire $\varepsilon_b=60\%$, $R_u=100$, $R_c=10$



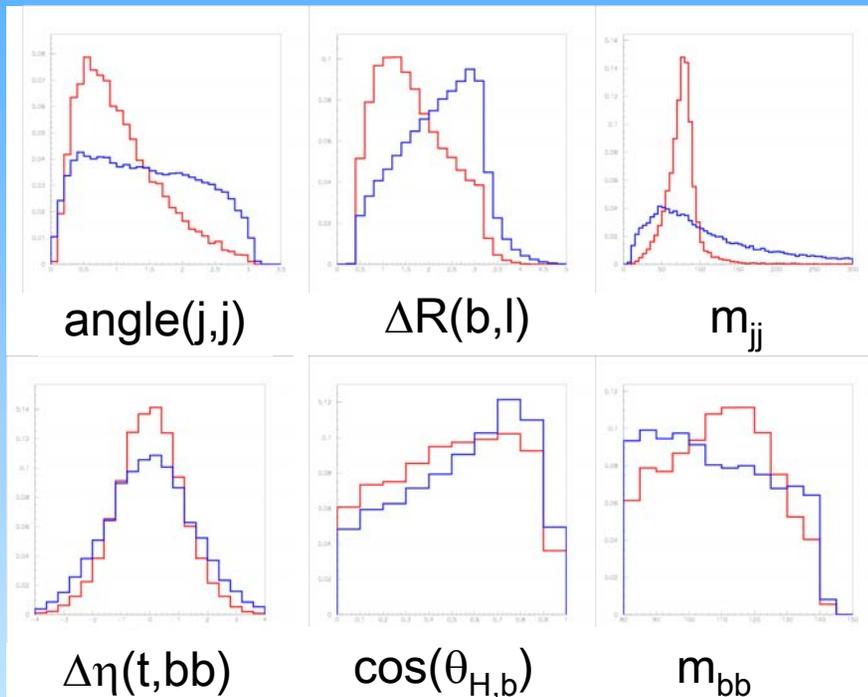
Événements attendus pour 30 fb^{-1} :

$$S/\sqrt{B_{\text{red}}} = N(\text{ttH}) / \sqrt{N(\text{ttjj})}$$

| Analyse en coupures | ttH(120) | ttbb | ttjj | ΣB_{kg} | S/\sqrt{B} | $S/\sqrt{B_{\text{red}}}$ |
|---------------------|----------|------|------|------------------------|--------------|---------------------------|
| Nb evt | 27 | 164 | 46 | 180 | 2.0 | 3.9 |

Analyse likelihood (LH)

Événements Atfast (simulation rapide),
étiquetage aléatoire $\varepsilon_b=60\%$, $R_u=100$, $R_c=10$



LH combinatoire :

Distinguer les bonnes
combinaisons des mauvaises
dans le signal

LH de signal :

Distinguer les événements ttH
du bruit de fond

| Analyse LH | ttH(120) | ttbb | ttjj | ΣB_{kg} | S/\sqrt{B} | S/\sqrt{B}_{red} |
|---------------|----------|------|------|-----------------|--------------|--------------------|
| Nb evt | 43 | 167 | 58 | 226 | 2.8 | 5.6 |

paramétrisation de l'étiquetage

Paramétrisation dépendant du p_T , basée sur :

- des événements tt/ttH
- l'algorithme 3D
- un détecteur à 2 couches de pixels
- des inefficacités modules/chips de 1/2 %

→ Plus réaliste que l'étiquetage « théorique » $\varepsilon_b=60\%$, $R_u=100$, $R_c=10$

Conclusions de l'analyse précédente :

| Analyse LH | ttH(120) | ttbb | ttjj | ΣB_{kg} | S/\sqrt{B} | S/\sqrt{B}_{red} |
|---------------|----------|------|------|-----------------|--------------|--------------------|
| Nb evt | 49 | 222 | 136 | 358 | 2.6 | 4.2 |

⇒ Dégradation en termes de S/\sqrt{B} : ~10%

étiquetage basé sur les traces

Étape suivante: événements issus de simulation détaillée et étiquetage des jets b basé sur les traces reconstruites

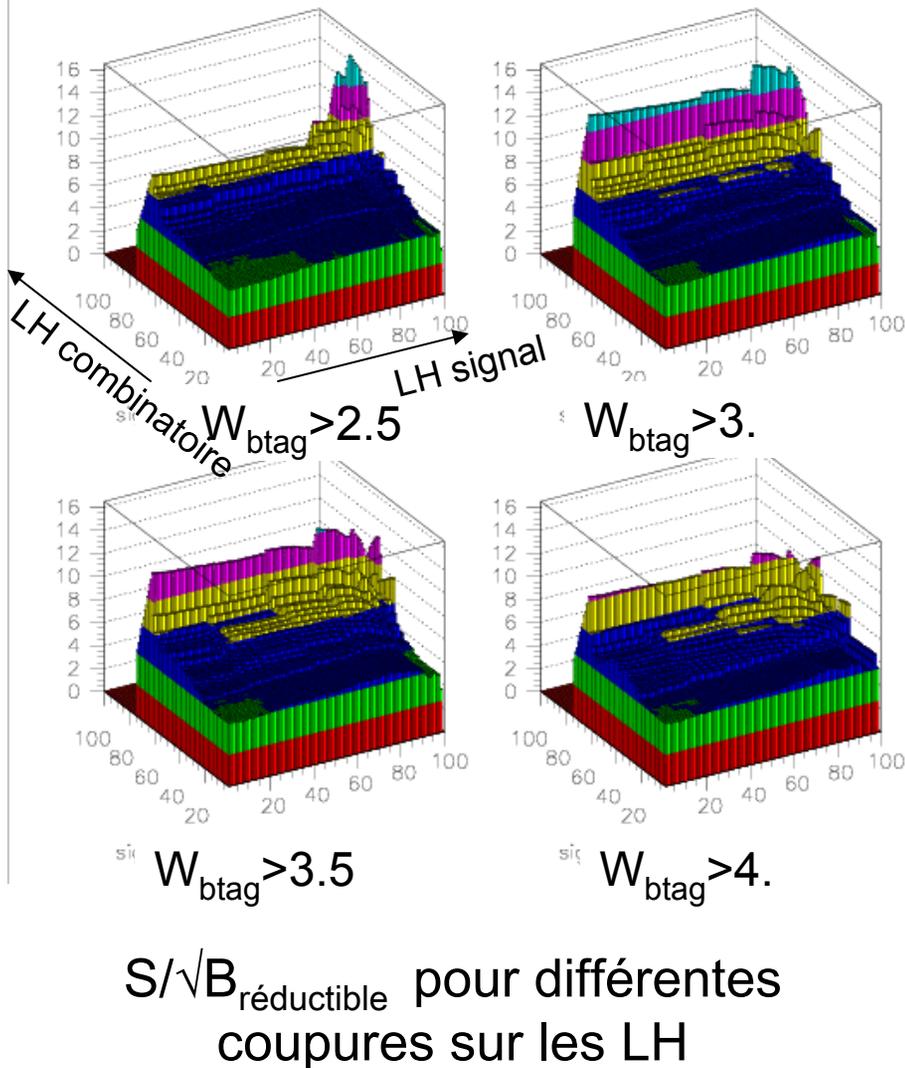
- plus réaliste : basé sur les paramètres des traces
- plus efficace : utilise les derniers algorithmes

données :

| Canal | ttH(120) | ttj | ttbb |
|------------|----------|---------|--------|
| événements | 20,000 | 100,000 | 20,000 |

petite stat.

Optimisation des coupures sur LH⁴⁰



Optimisation des 3 likelihoods :

Likelihood combinatoire

Likelihood de signal

Likelihood d'étiquetage des b

Pour le moment, l'optimisation a porté sur $S/\sqrt{B}_{\text{réductible}}$

Les pics correspondent à des coupures LH pour lesquelles un seul événement du bruit de fond réductible survit.

"vrai" étiquetage des jets b sur événements de simulation détaillée 41

"vrai" étiquetage $w_{btag} > 3$:

$$\varepsilon_b = 65\%$$

$$R_{\text{léger}} = 60$$

$$R_c = 6.5$$

étiquetage aléatoire fixe :

$$\varepsilon_b = 60\%$$

$$R_{\text{léger}} = 100$$

$$R_c = 10$$

| Analyse | ttH(120) | ttbb | ttjj | ΣB_{kg} | S/\sqrt{B} | S/\sqrt{B}_{red} |
|------------------------------|----------|------|------|-----------------|--------------|--------------------|
| analyse précédente, coupures | 27 | 164 | 46 | 180 | 2.0 | 3.9 |
| cette analyse, coupures | 41 | 114 | 77 | 191 | 3.0 | 4.7 |

⇒ même sur une analyse rudimentaire, apport important

“vrai” étiquetage des jets b sur événements de simulation détaillée

"vrai" étiquetage $w_{btag} > 3$:

$$\varepsilon_b = 65\%$$

$$R_{\text{léger}} = 60$$

$$R_c = 6.5$$

étiquetage paramétrisé :

$$\varepsilon_b = 60\%$$

$$R_{\text{léger}} = 40$$

$$R_c = 6$$

| Analyse | ttH(120) | ttbb | ttjj | ΣB_{kg} | S/\sqrt{B} | S/\sqrt{B}_{red} |
|-----------------------|----------|------|------|-----------------|--------------|--------------------|
| analyse précédente LH | 49 | 222 | 136 | 358 | 2.6 | 4.2 |
| cette analyse LH | 71 | 185 | 19 | 204 | 5.0 | 16.0 |

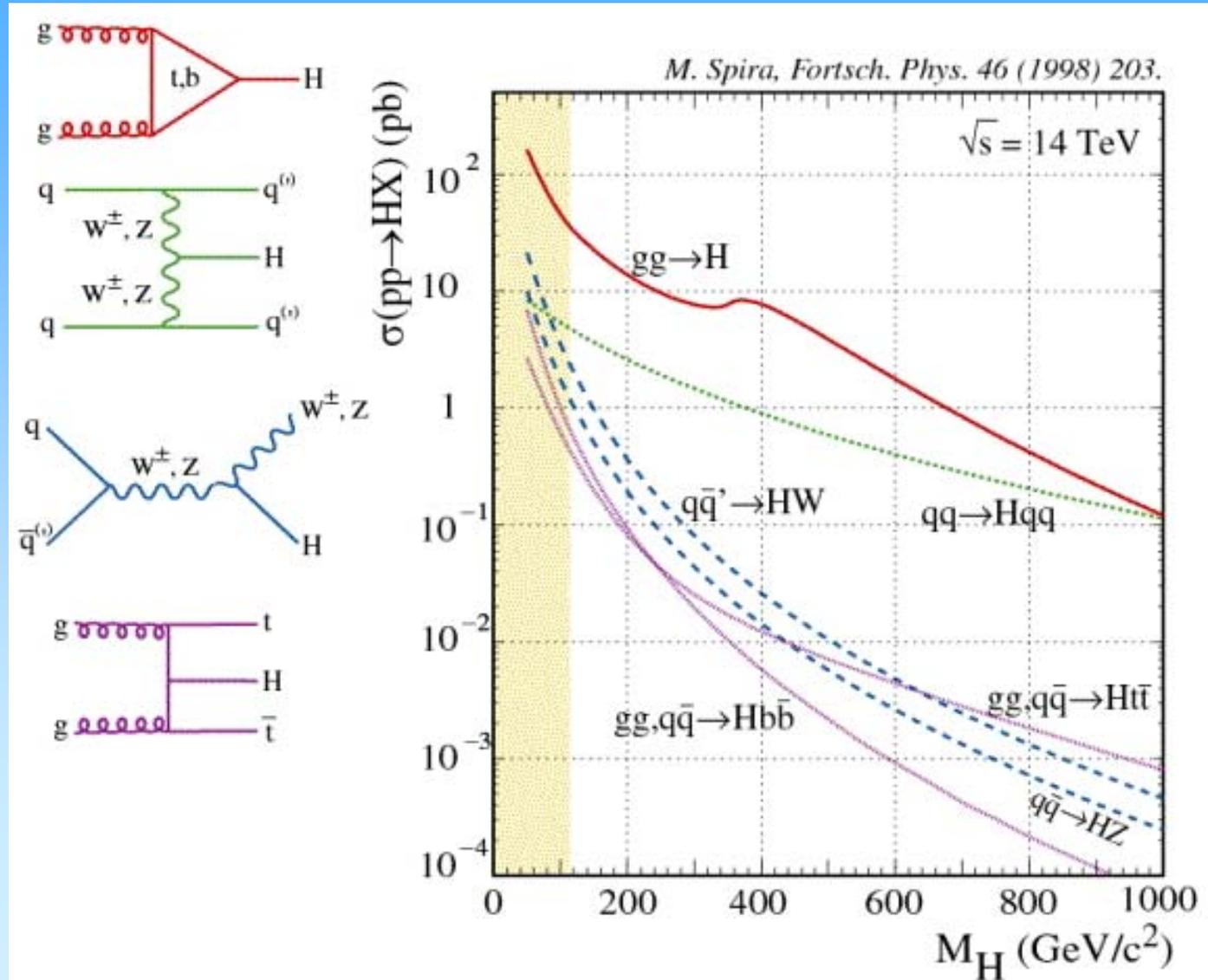
⇒ apport majeur des dernières méthodes d'étiquetage

Conclusions

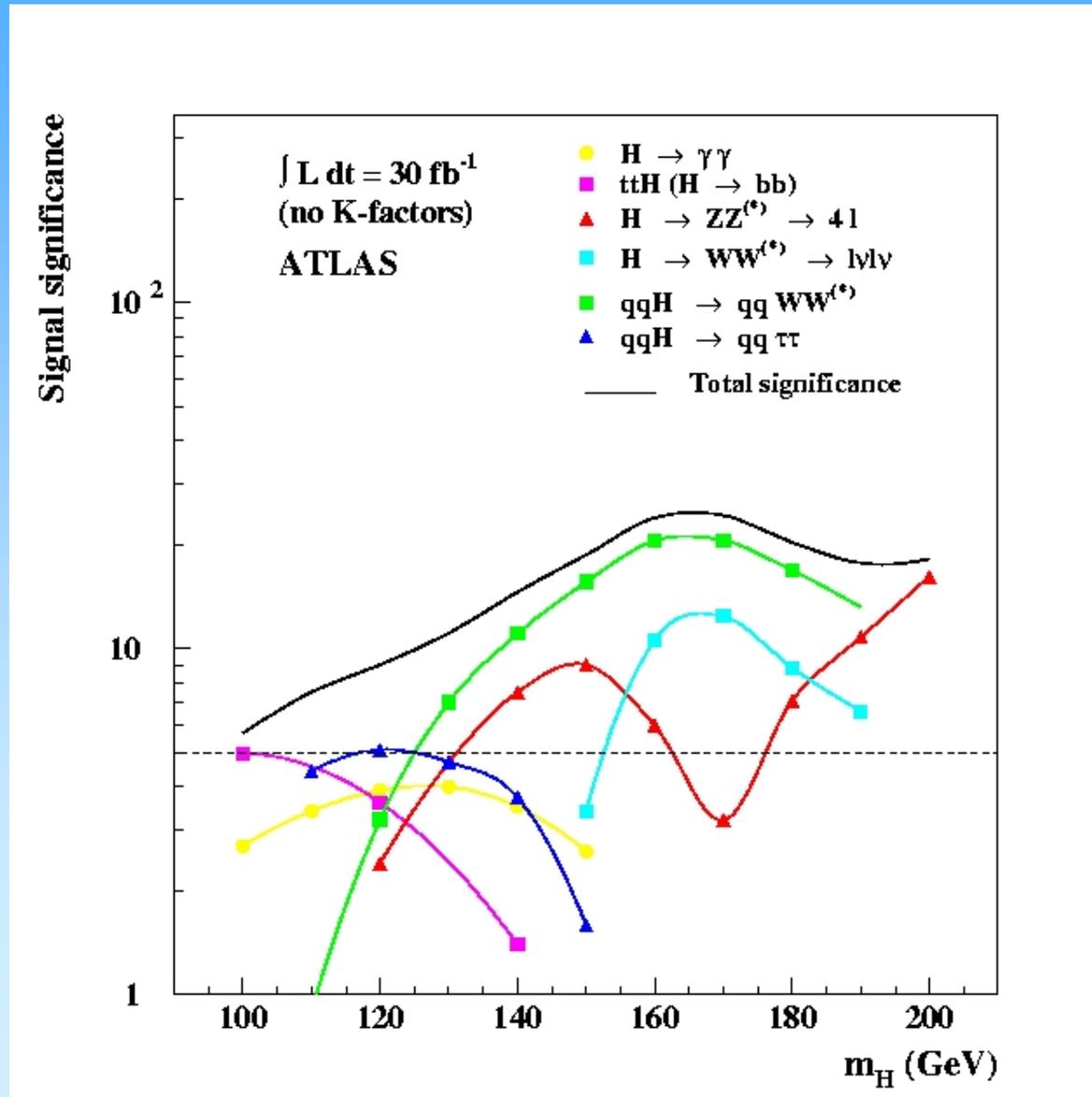
43

- L'étiquetage des jets b requiert de très bonnes performances du détecteur → tout faire pour optimiser celles-ci
- La calibration de l'étiquetage est possible (ajuster les résolutions, extraire les propriétés des jets b à partir des données)
- Les performances d'étiquetage des jets b sont déterminantes pour la physique (2D → 3D → SVX → futur...)
- L'amélioration de la découverte du boson de Higgs dans le canal $tt H \rightarrow bb$ a été démontrée → 5σ . Restent encore des améliorations potentielles à explorer

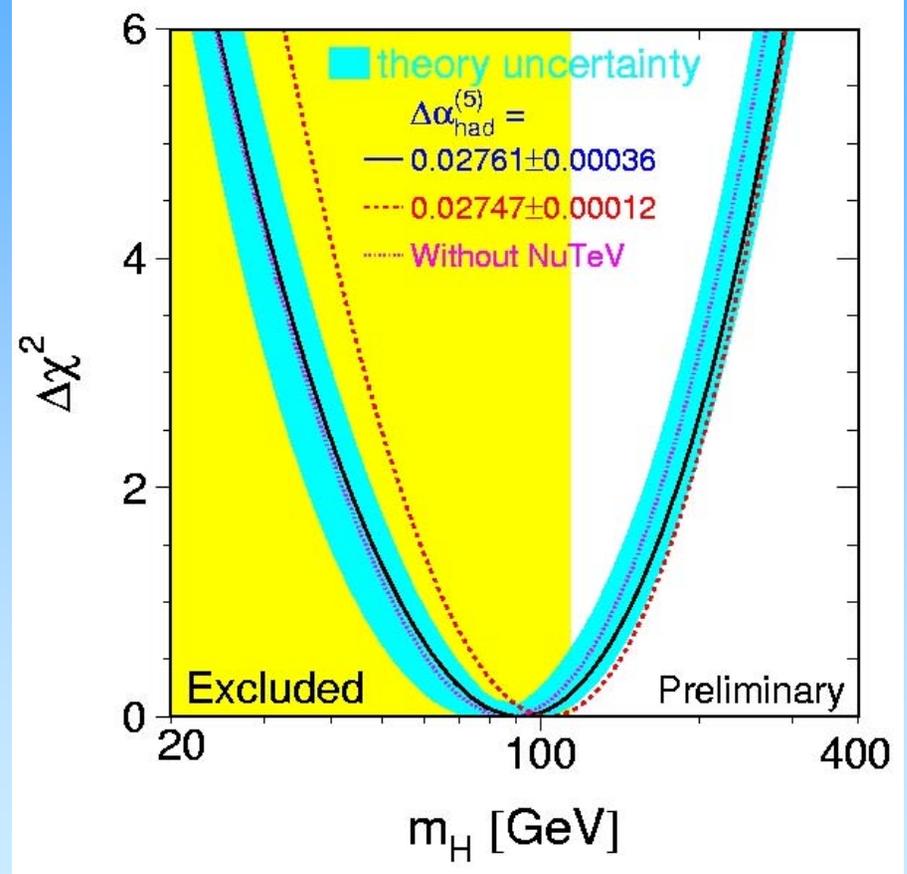
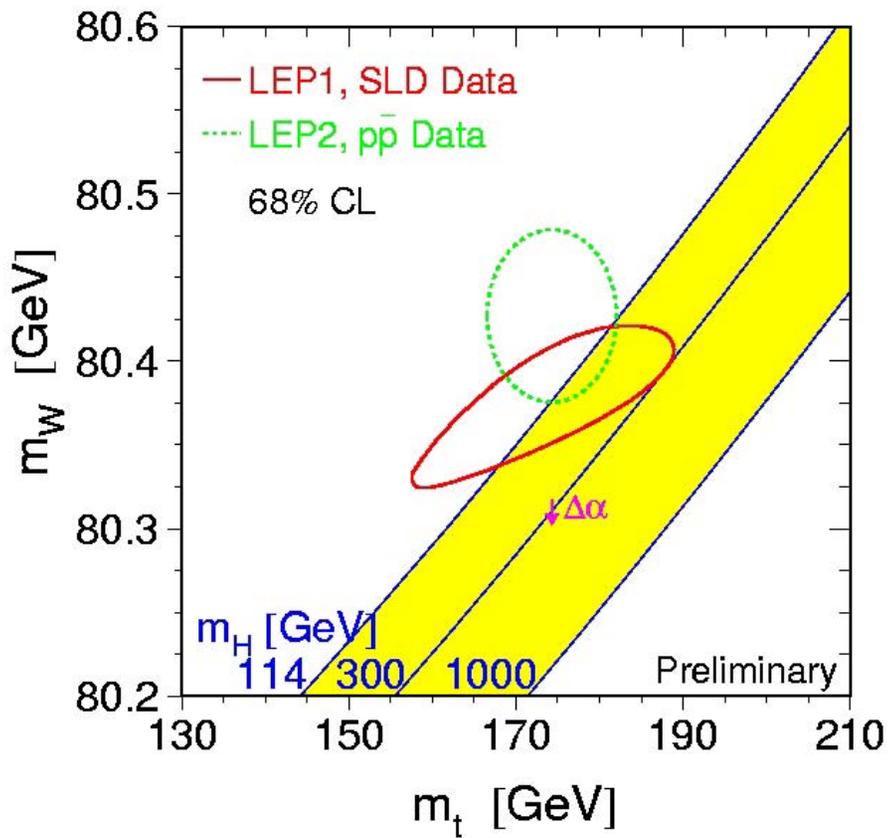
Le boson de Higgs dans ATLAS



Le boson de Higgs dans ATLAS



Ajustement à partir de mesures de précision électrofaibles



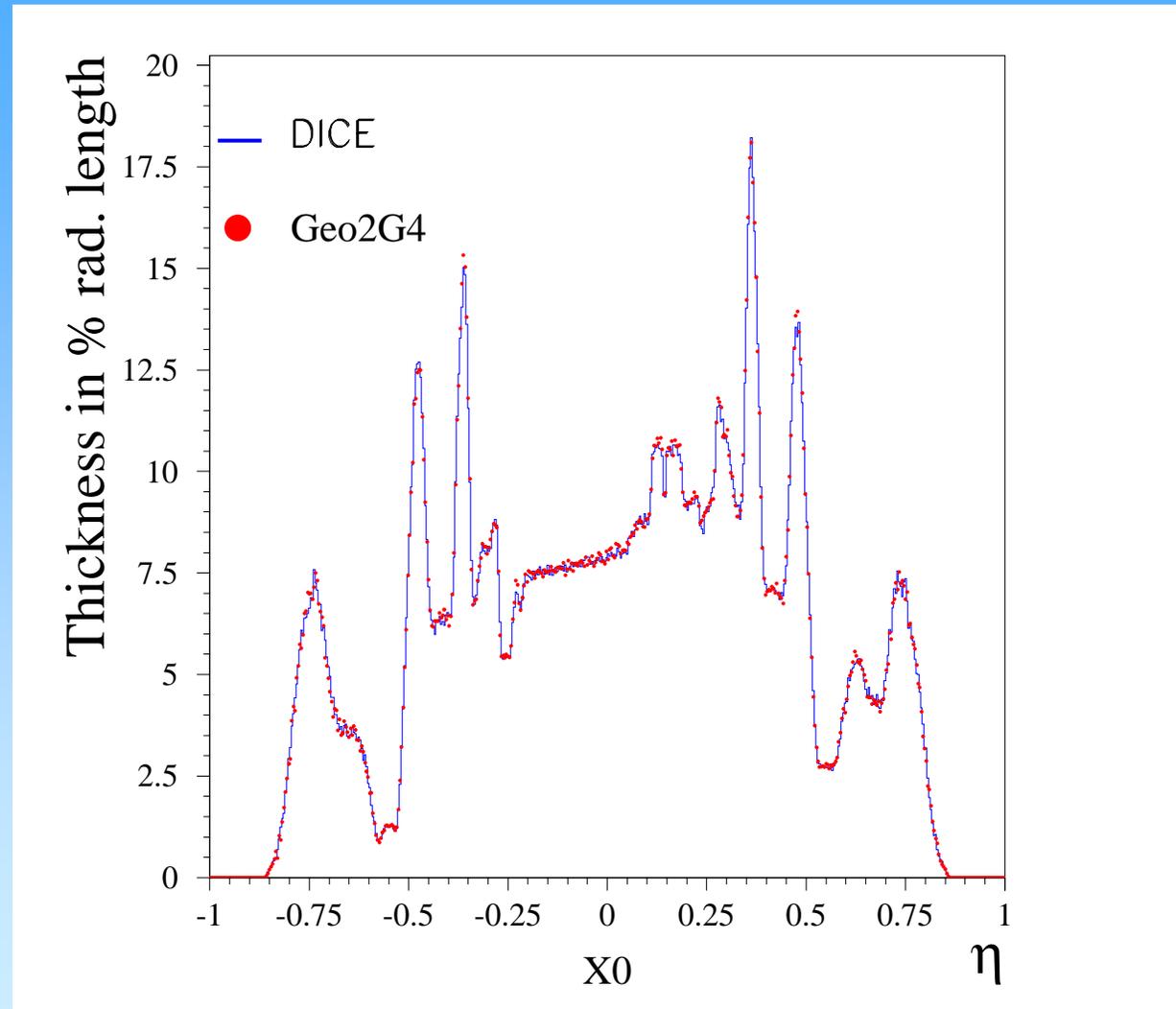
$m_H < 211 \text{ GeV}/c^2$ à 95 % C.L.

$$m_H = 114^{+69}_{-45} \text{ GeV}/c^2$$

propriétés physiques des jets b

Peuvent différer entre la réalité et le MC :

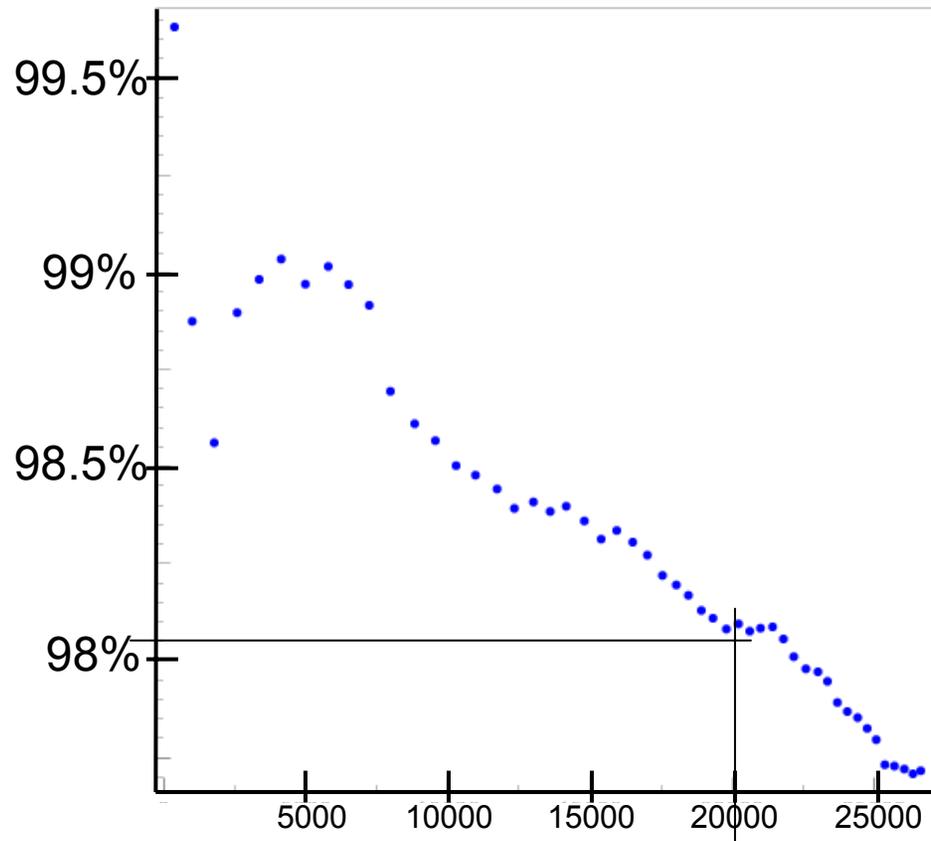
- Multiplicité des jets b
- Rapport multiplicité chargée/neutre : B^+/B^0
- Rapport baryons/mésons : Λ_B/B
- Branching de désintégration
- Corrélation des désintégrations due aux éléments de matrices



Calibration sur tt : les coupures

Sélection des événements tt :

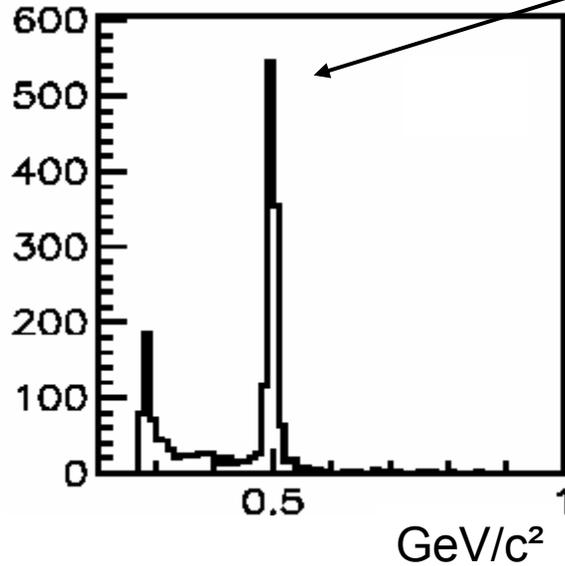
- 4 jets de $p_T > 15$ GeV/c, dont 2 avec $|\eta| < 2.5$ (candidats b)
- 1 lepton isolé de $p_T > 15$ GeV/c et $|\eta| < 2.5$
- $p_T^{\text{all}} > 300$ GeV : événements bien dos-à-dos
- Pour les 3 jets formant le top : $|\Sigma p_T| > 130$ GeV/c
- Fenêtres de masse avant fit contraint :
 - $m_W - 9 \text{ GeV}/c^2 < m_{W\text{rec}} < m_W + 9 \text{ GeV}/c^2$
 - $m_{\text{top}} - 25 \text{ GeV}/c^2 < m_{\text{THrec}} < m_{\text{top}} + 25 \text{ GeV}/c^2$
 - $m_{\text{top}} - 18 \text{ GeV}/c^2 < m_{\text{TLrec}} < m_{\text{top}} + 13 \text{ GeV}/c^2$



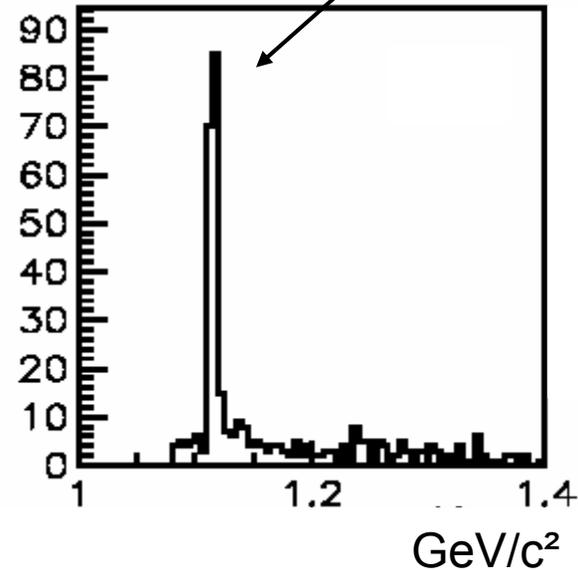
Pureté vs nombre de jets b
sélectionnés (pour 10 fb^{-1})

Amélioration des méthodes

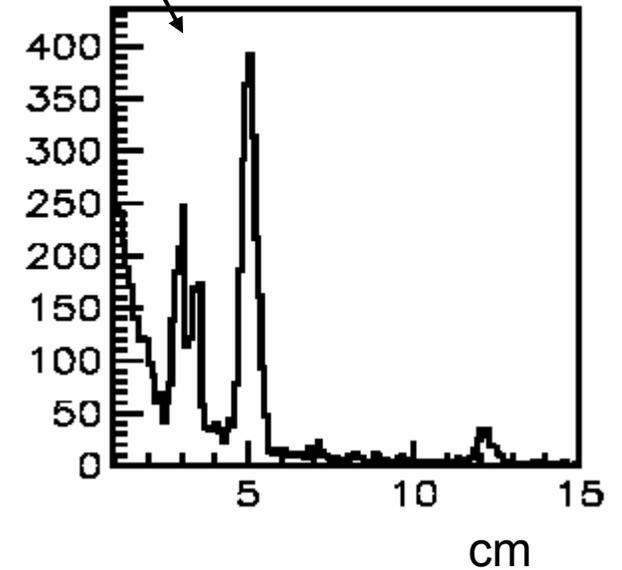
- **Suppressions des traces provenant d'un vertex :**
 - d'interaction avec la matière du détecteur
 - De désintégration de K ou Λ



Masse invariante $\pi\pi$



Masse invariante $p\pi$



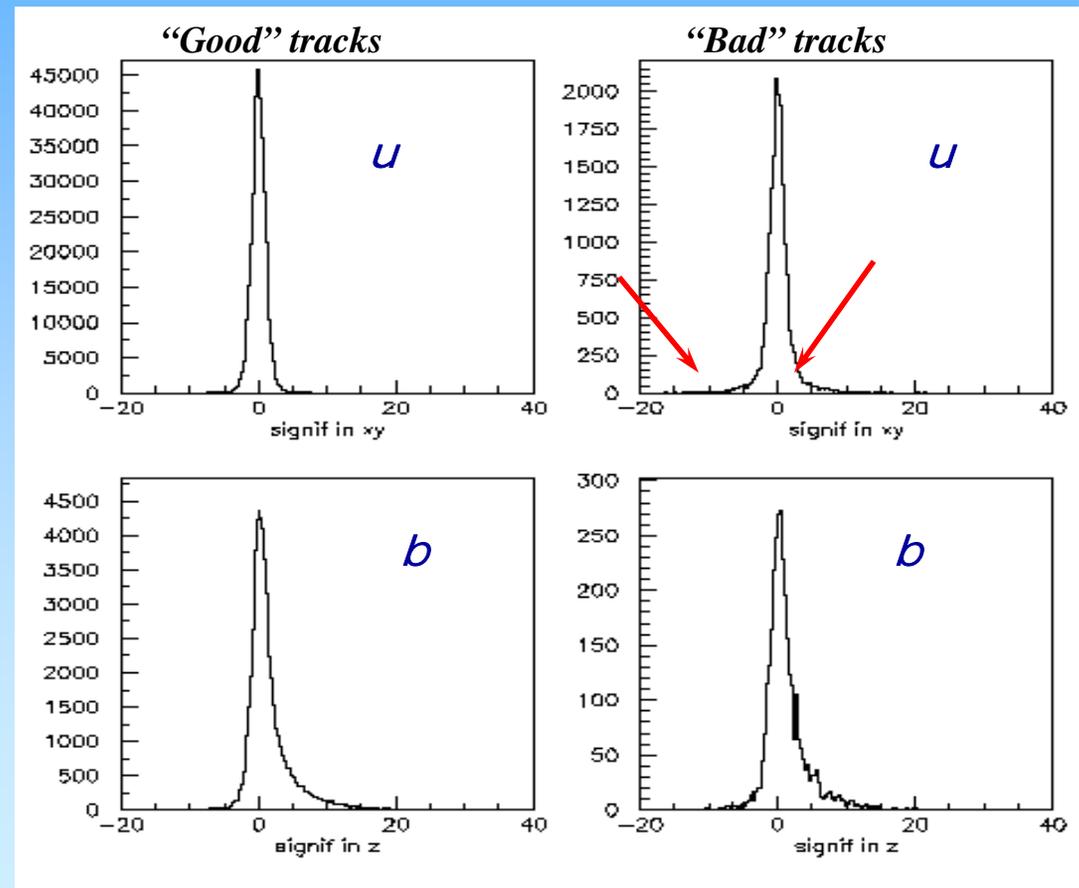
Distance $vtx\ 1^{aire} / vtx\ 2^{ndaire}$

Amélioration des méthodes

“mauvaises” traces

Bad tracks if one of conditions:

- one shared hit in b-layer
- one shared hit in pixels
- two shared hits in SCT
- one ambiguous hit in b-layer



Amélioration des méthodes “mauvaises” traces

ttH/ttjj events, mH=120 GeV, with pile-up, 2 pixel layers Initial layout, b-layer 400 μm ,
inefficiencias modules/chips 1-2 % , b-layer inef. 0.5-1.0 %

b- from ttH, u- from ttjj, ATLFast jets, reconstructed vertex , SV1 method

$$\Delta R_{jj} > 0.8 , \quad \Delta R_{jbquark} > 0.8$$

| | No bad tracks | With bad tracks | $R_{\text{with/no}}$ |
|-----------------------------|----------------|-----------------|----------------------|
| $R_u \quad \epsilon_b=50\%$ | 1635 \pm 169 | 1693 \pm 178 | 1.04 |
| $\epsilon_b=60\%$ | 386 \pm 19 | 418 \pm 22 | 1.08 |
| $\epsilon_b=70\%$ | 75 \pm 2 | 79 \pm 2 | 1.05 |

Leptons

- $B \rightarrow e^\pm X$: $\sim 20\%$

Identification e^\pm dans un jet :

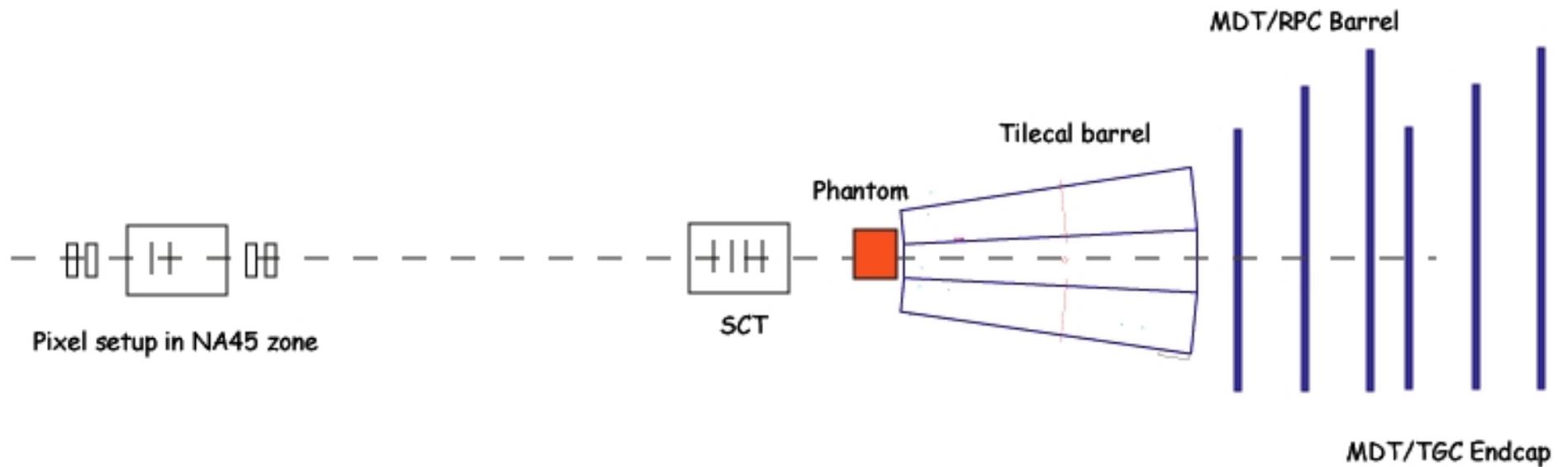
- Calorimètre électromagnétique
- Informations du TRT

- $B \rightarrow \mu^\pm X$: $\sim 20\%$

Identification μ^\pm dans un jet :

- Chambre à muons ⁺ Correspondance d'une trace dans le détecteur interne
- Calorimètre hadronique

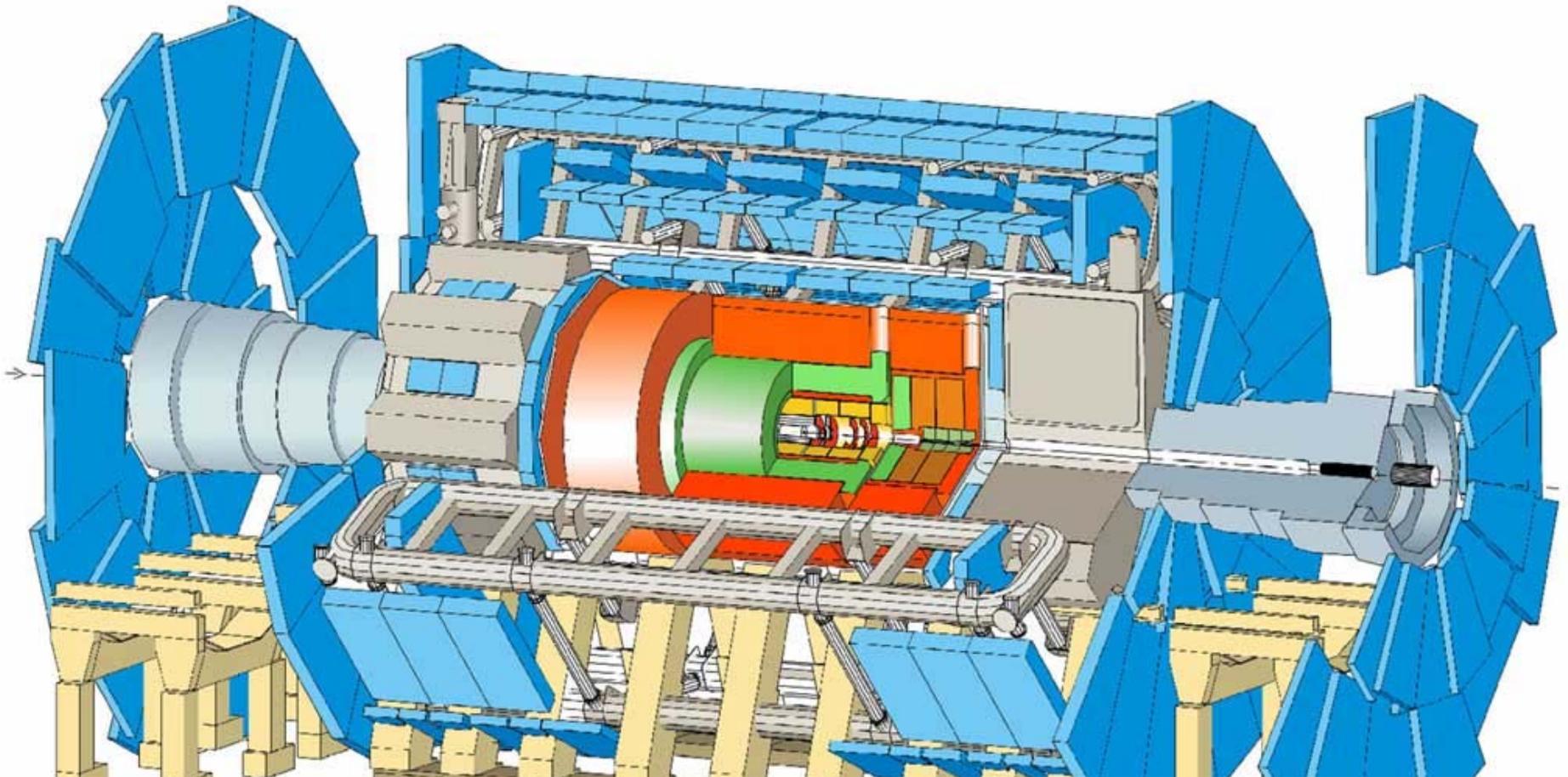
Test en faisceau combine



ATLAS

Systeme de déclenchement

Ferme de PC



- sélectionner les événements *a priori* intéressants
- 3 niveaux de déclenchement