

Démarrage du LHC et préparation des premières analyses « top » dans l'expérience CMS

Jérémy Andrea
Brown University



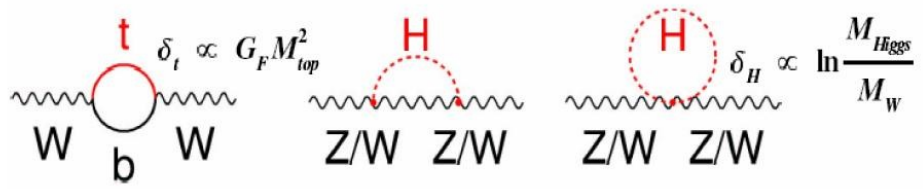
Introduction

- Premières données de collisions fournies par le LHC fin 2009, phase de compréhension des détecteurs et premiers résultats.
- Préparation des mesures de physique et de la mesure de la section efficace $t\bar{t}$ avec les premières données (20-100 pb⁻¹) issues de collisions de « haute » énergie (≥ 7 TeV).
- Les sélections d'événements doivent être robustes et leurs efficacités de sélection doivent être estimées à partir des données :
 - Faible statistique d'événement,
 - Accord données-Monte Carlo éventuellement imparfait.
- Dans ce cadre, une attention particulière a été portée à:
 - L'estimation des bruits de fond à partir des données,
 - Une première estimation des performances du détecteur (ici du trajectographe) dans les données cosmiques et de collisions à 900 GeV.

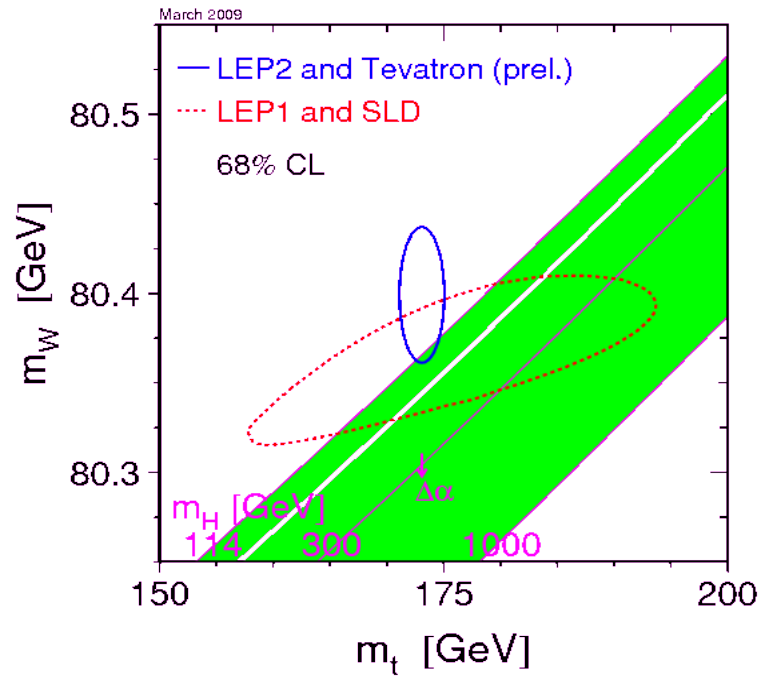
Physique du top au LHC



- **Masse du Top** ($173.1 \pm 0.6(\text{stat}) \pm 1.1(\text{sys}) \text{ GeV}/c^2$):
- Le Top et le boson de Higgs interviennent dans les corrections radiatives des propagateurs des bosons W et Z,
- Mesurer précisément la masse du top (et du W) permet de déterminer des limites sur la masse du boson de Higgs.

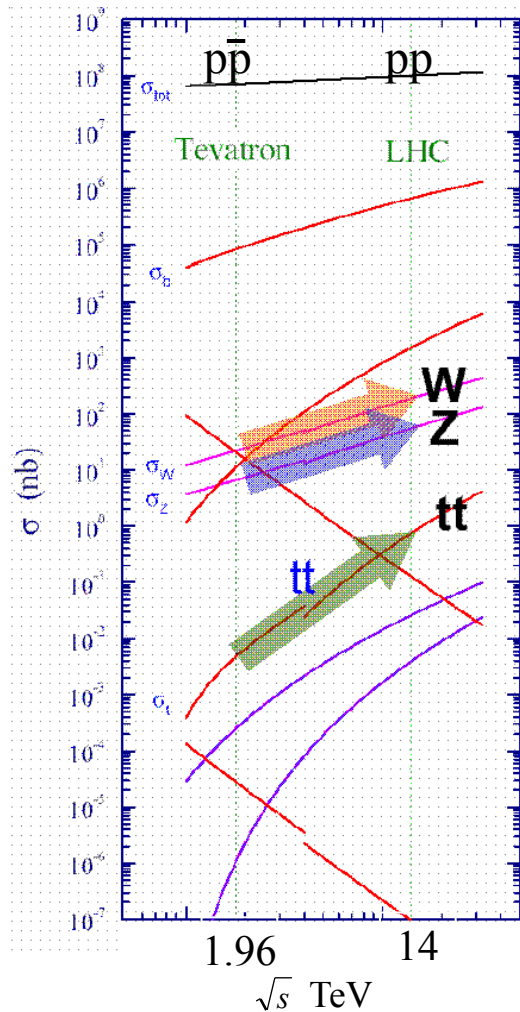


- **Mesure de la section efficace:**
 - Large augmentation au LHC,
 - Mesure de précision,
 - Sonde de la nouvelle physique.
- **Top** = un bruit de fond principaux de nombreux canaux de **nouvelle physique**.



Higgs chargé, résonances de grande masse, di-leptons de même charges, couplages anormaux...

Du Tevatron au LHC



	Tevatron pb	LHC (14 TeV) pb	LHC (10 TeV) pb	LHC (7 TeV) pb
Ttbar pairs	7.60	908	414	187
Single top s-channel	0.88	11	5	4
Single top t-channel	1.98	246	130	64
Single top Wt channel	0.15	66	29	11

$\sigma_{\text{top}} \neq \sigma_{\text{antitop}}$
 $\sim 10^9$ événements/10fb⁻¹ **W** (200 par seconde)
 $\sim 10^8$ événements/10fb⁻¹ **Z** (50 par seconde)
 $\sim 10^7$ événements/10fb⁻¹ **tt** (1 par seconde)

**Une large quantité d'événements du Modèle Standard
 va être collectée au LHC
 (échantillons de contrôle et de calibration)**

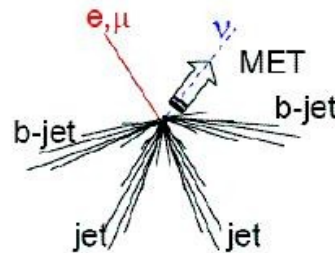
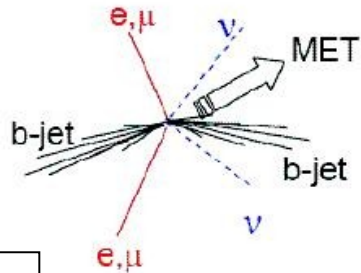
Le prix à payer : une large quantité d'événements QCD $\approx 10^9$ pb



Canaux de désintégration de paires de top

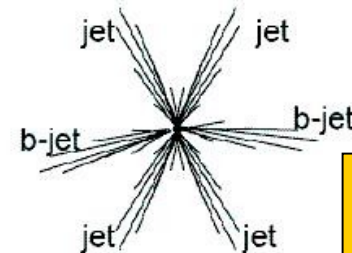
Canal di-leptonique
Faible statistique
Faible bruit de fond

top, W+jets, Z+jets, dibosons



Canal Semi-leptonique
Grande stat.
Bruit de fond modéré

W(bb)+jets, QCD multijet



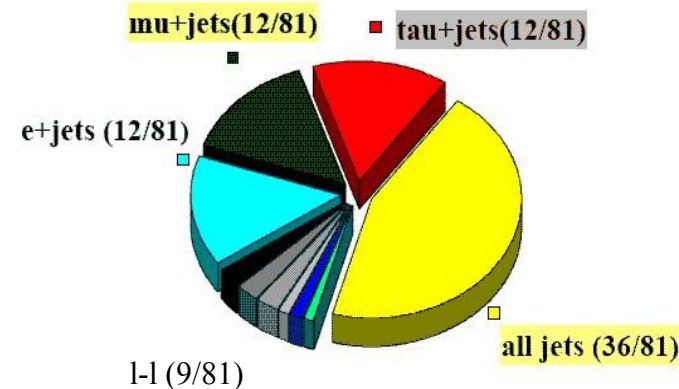
$|V_{tb}| \approx 1$

Canal full-hadronique
Grand stat.
Fort bruit de fond

QCD multijet

Bruits de fond instrumentaux: faux leptons(isolés)/MET

- $T\bar{T}$ important pour la compréhension globale des détecteurs : la majorité des objets reconstruits peut être utilisée!
- **Reconstruction des Jets,**
- **Reconstruction et identification des leptons,**
- **E_T manquante,**
- **Identification des jets b.**
- Peut être utilisé pour la calibration en énergie des jets ou les mesures d'efficacité de b-tagging.



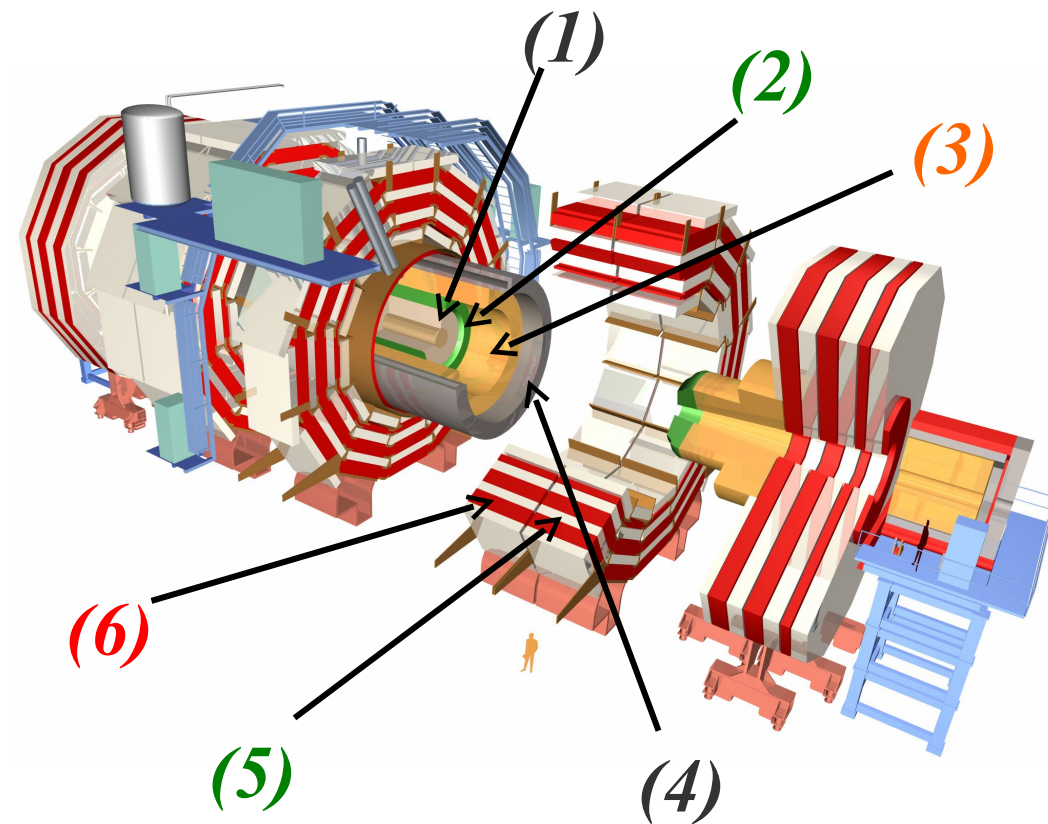
Le LHC



- Collisionneur proton-proton (ion-ion) situé au CERN à Genève dans l'ancienne cavité du LEP (27 km de circonférence).
- Dans son mode de fonctionnement nominal, des faisceaux de protons seront accélérés jusqu'à une énergie de 7 TeV chacun et mis en collision.
- De grandes luminosités ($10^{34} \text{ cm}^{-2}\text{s}^{-1}$) seront atteintes.
- LHC inauguré le 10 septembre 2008.
- Premières collisions observées en Septembre 2009 (collisions à 900 GeV et 2.36 TeV dans le centre de masse)
- Plan pour 2010-11 : collisions de 7 TeV dans le centre de masse, environ 1 fb^{-1} de données collectées.
- >2011, arrêt de « longue durée » puis collisions à 14 TeV.



Le détecteur CMS



- De forme cylindrique centré sur l'axe du faisceau. Dimensions : diamètre de 15 m, longueur de 22 mètres (12500 tonnes).
- Composé d'une partie centrale et de deux bouchons.
- (1) Trajectographe au silicium.
- (2) Calorimètre électromagnétique (ECAL) au tungstate de plomb.
- (3) Calorimètre hadronique (HCAL) : plaques de cuivre et scintillateurs plastiques.
- (4) Solénoïde supraconducteur : 4 T.
- (5) Chambres à muons.
- (6) Structure de soutien et de retour des boucles de champ.



Premiers événements vus par CMS

1. La physique du top au LHC,
2. **Premiers événements vus par CMS,**
3. Mesure de la section efficace $t\bar{t}$ bar,
4. Conclusion



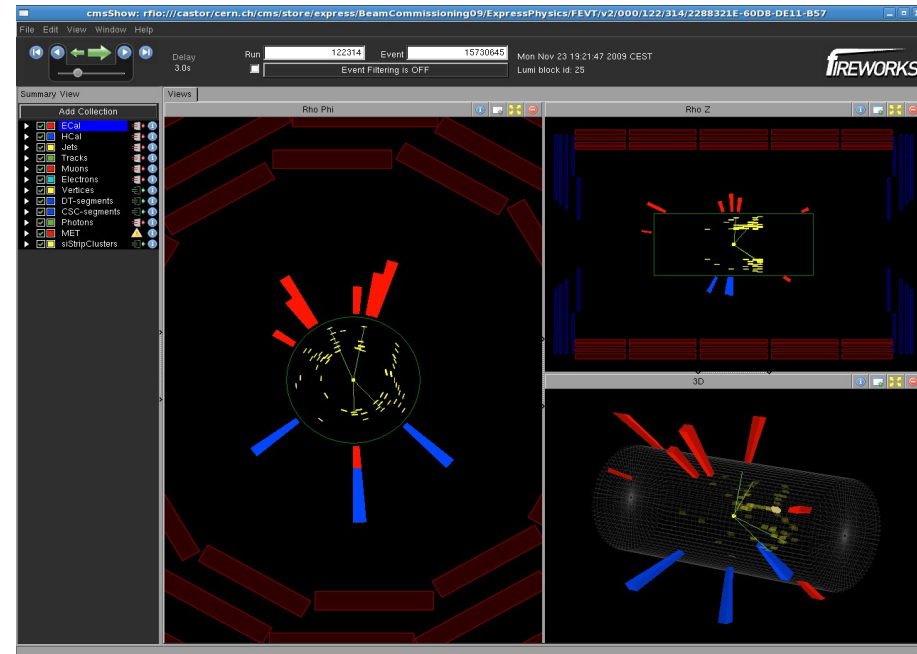
Démarrage du LHC vu par CMS (1)

Le jour J...



- Le 20 novembre 2009, le LHC a vu ses faisceaux circuler.
- Le 23 Novembre, les **premières collisions à 900 GeV** (c.d.m.) sont enregistrées dans CMS !

- Le 14 décembre, environ 15 événements sont enregistrés pour des collisions de **2.36 TeV** (c.d.m.), record battu !
- En 2009, environ 500 événements “biais minimum” ont été enregistrés par CMS (**>80%** des périodes de faisceau, **CMS complètement allumé**), $\approx 16 \mu\text{b}$.
- Prochains transparents : **rapide vue d'ensemble des résultats de CMS**.

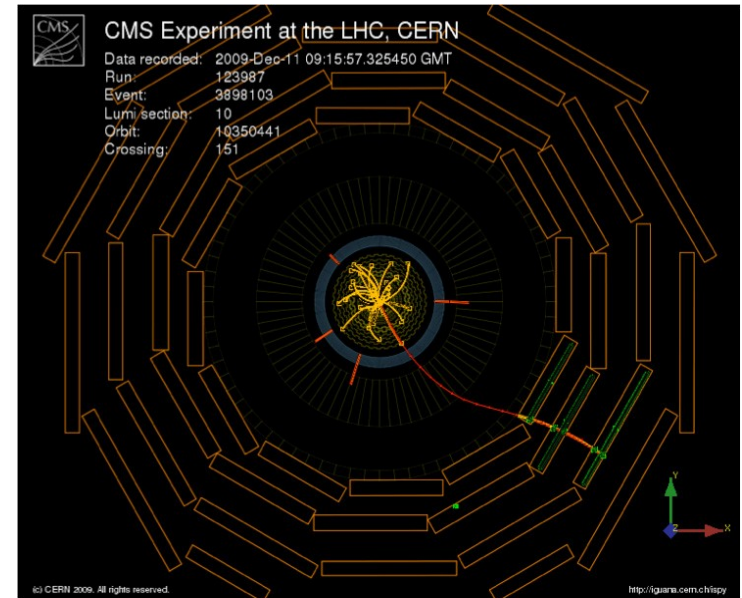
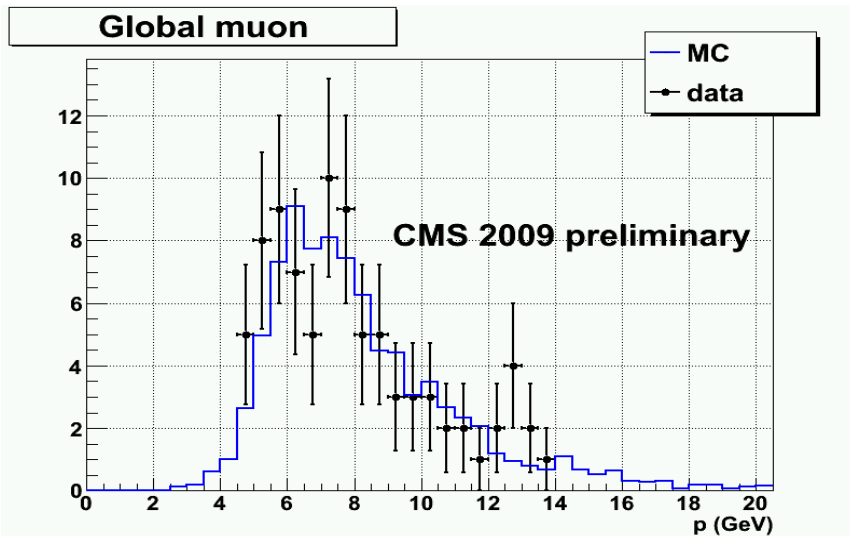


Démarrage du LHC vu par CMS (2)

Reconstruction des muons



- Reconstruction des premiers muons issus de collisions.
- “event display” d'un candidat muon reconstruit en combinant l'information du trajectographe et des chambres à muon.



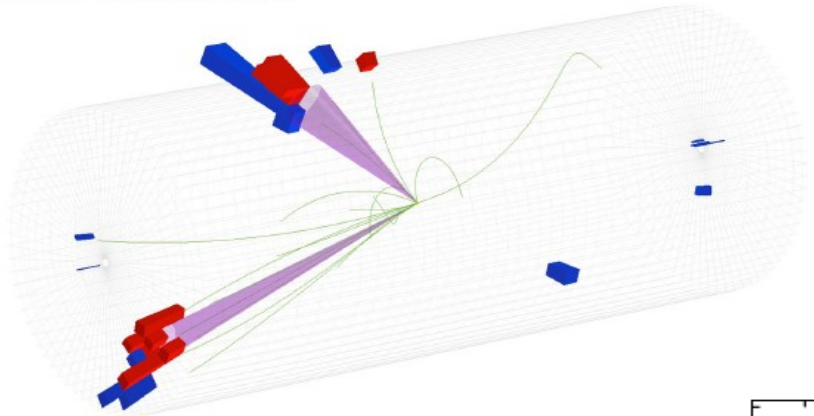
- Comparaison entre muons reconstruits dans les données et dans le Monte-Carlo.
- Compte tenu des larges erreurs statistiques, accord raisonnable.



Démarrage du LHC vu par CMS (3) reconstruction des jets

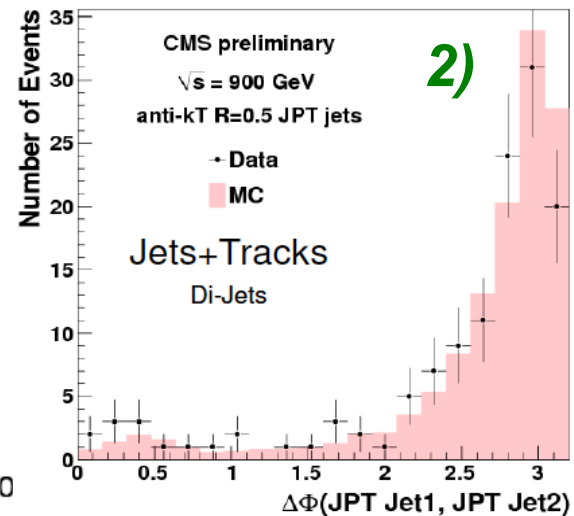
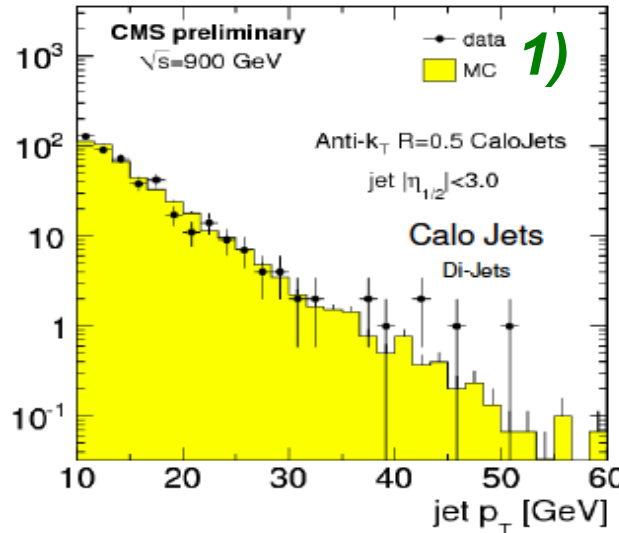


CMS Experiment at the LHC, CERN
Date Recorded: 2009-12-06 07:18 GMT
Run/Event: 123596 / 6732761
Candidate Dijet Collision Event

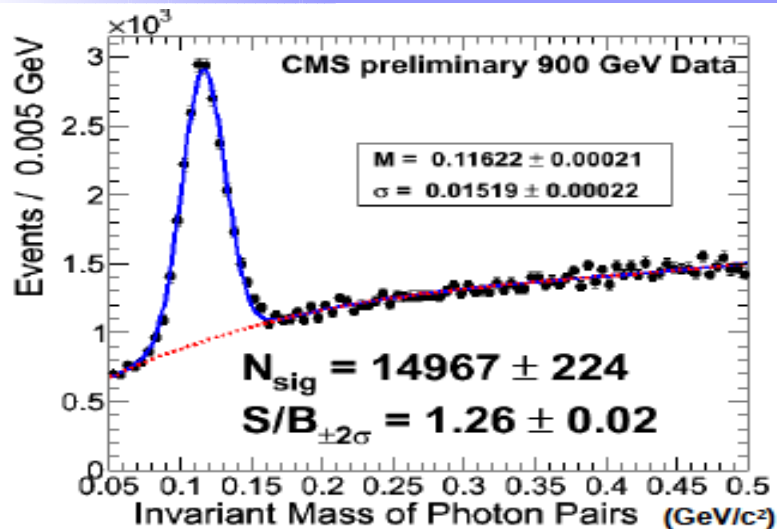


- “event display” d'un **candidat d'événement di-jet** (900 GeV).
- Jets **dos-à-dos** dans le plan transverse, de **p_T comparable** (≈ 25 GeV/c) et des fractions d'énergie hadronique $\approx 50\%$.

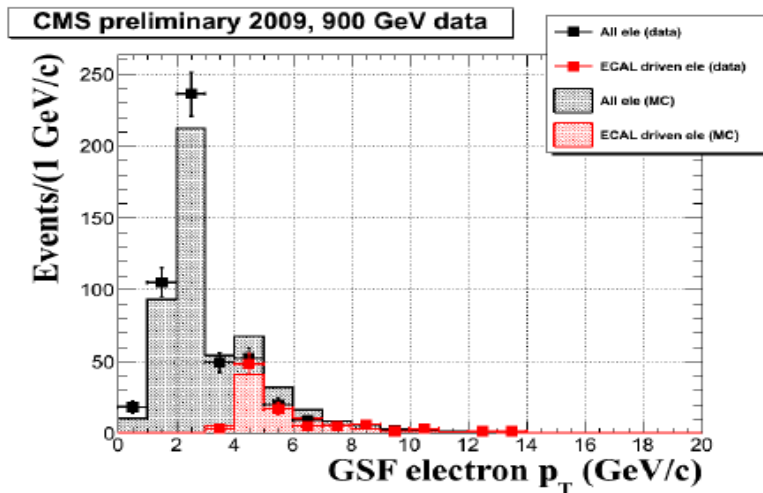
- Comparaisons données-MC,
 - du spectre en p_T des jets **1)**
 - De la distribution $\Delta\Phi$ di-jet (jets corrigés utilisant l'information du trajectographe) **2)**



Démarrage du LHC vu par CMS (4) calorimétrie électromagnétique



- Distribution non corrigée de la masse **invariante di-photons**.



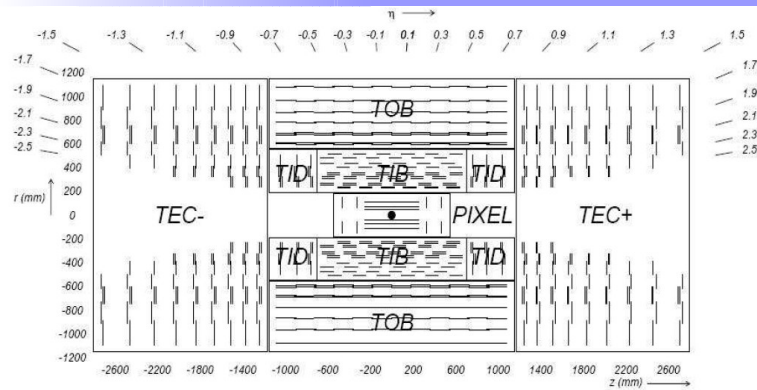
- Comparaison données-MC du **spectre en p_T des électrons** reconstruits.
- Bon accord



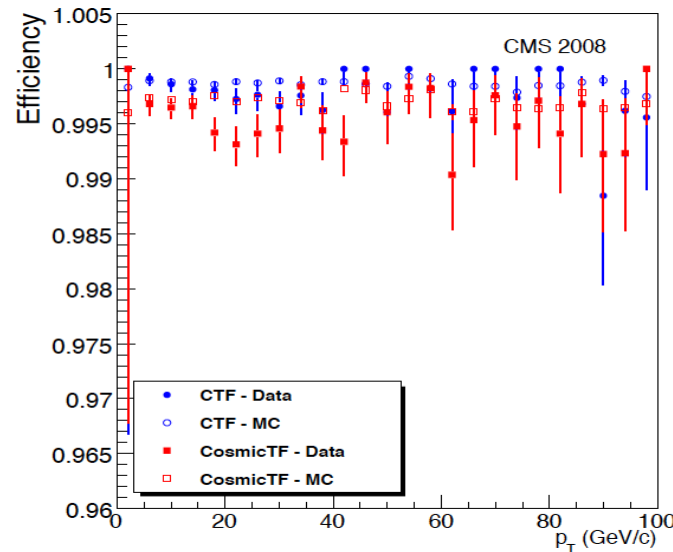
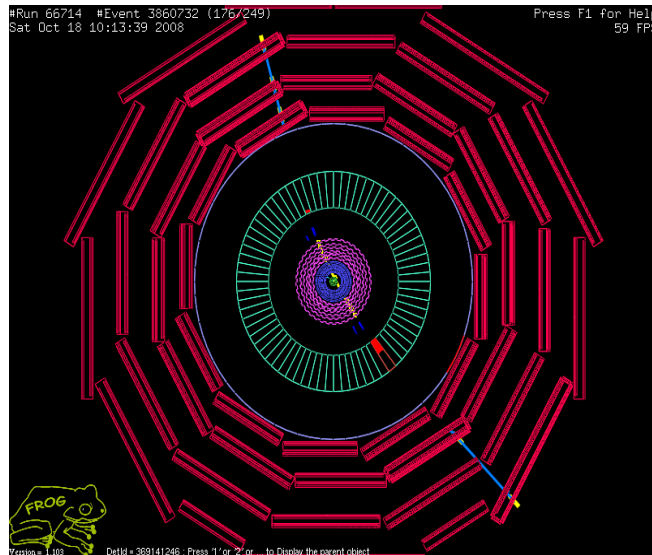
Reconstruction des traces dans les données cosmiques



* = mes contributions



- Première validation de la reconstruction des traces a été effectuée à partir de données « cosmiques ».
- Etude des efficacités de reconstruction des traces à partir des données cosmiques sans ou avec champ magnétique.
- Comparaison p_T des traces et des muons : test du champ magnétique.



- Efficacité proche de 100%.
- arXiv:0911.4996v1, JINST.

Méthode en cours d'étude dans les données de collisions

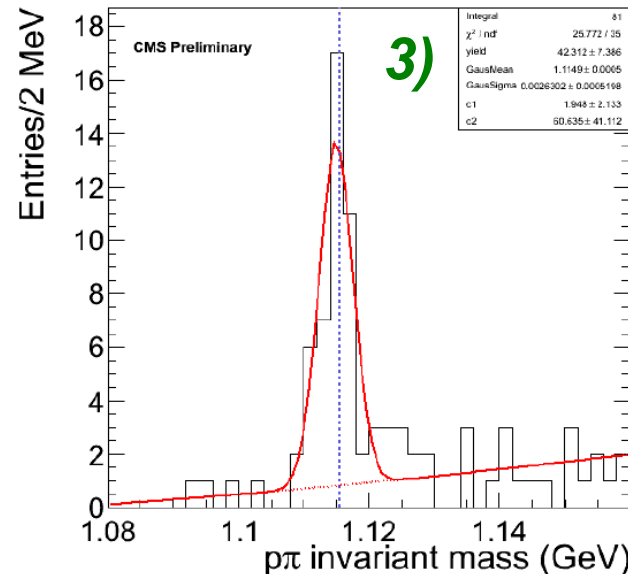
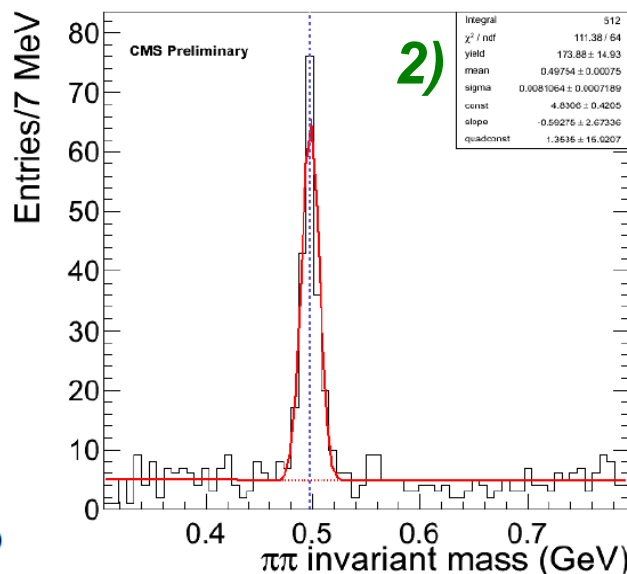
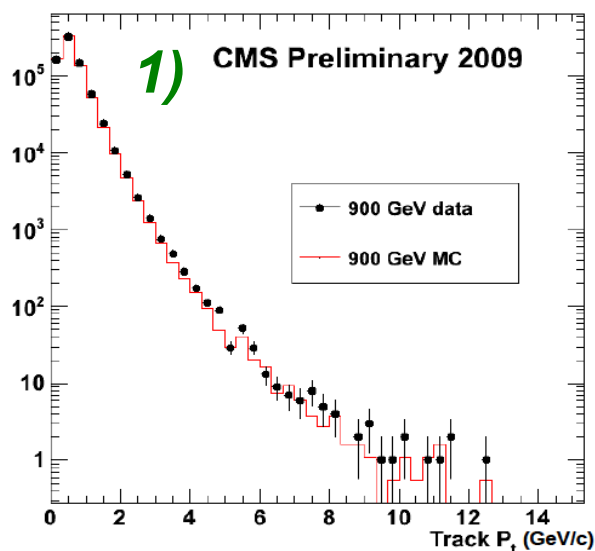
Reconstruction des traces dans les données de collisions



- La reconstruction des traces a été testée avec succès sur des données de collisions, ici à 900 GeV:
 - 1) comparaison MC-données du spectre en pT des traces reconstruites
 - 2) et 3), masses invariantes de paires de traces « proches » avec des hypothèses $\pi\pi$ et $\pi\text{-p+vertex}$ fit.

$M=497.9 \text{ MeV}/c^2$
 $\sigma=7.7 \text{ MeV}/c^2$

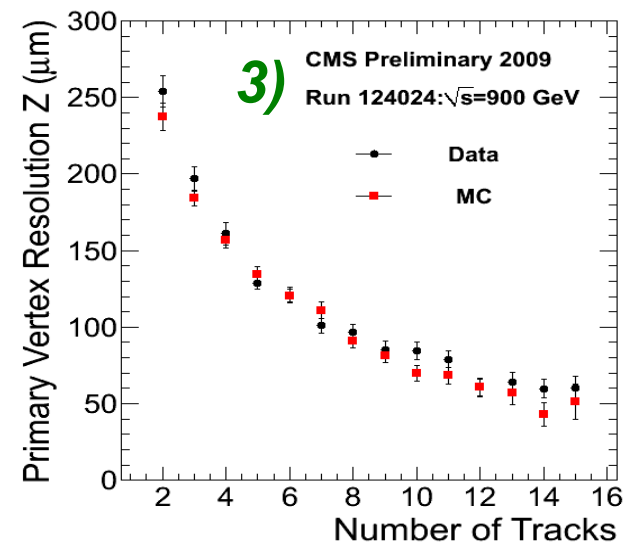
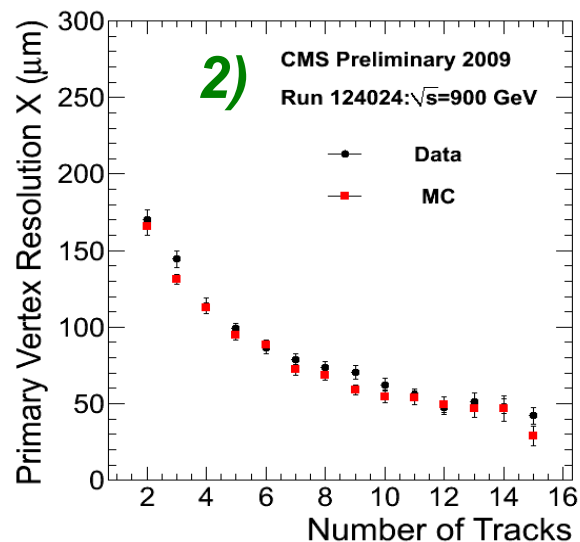
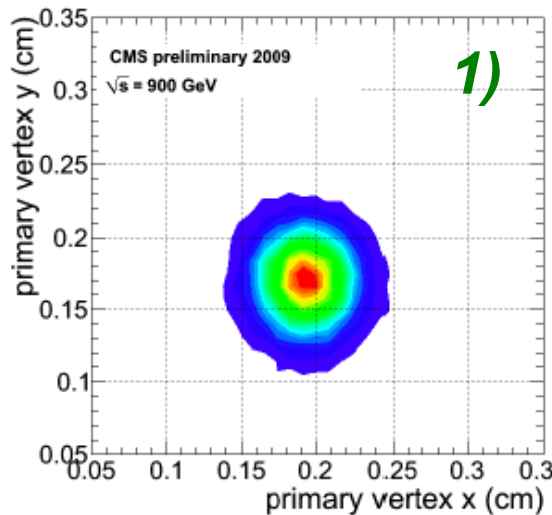
$M=1.116 \text{ GeV}/c^2$
 $\sigma=3.1 \text{ MeV}/c^2$



Reconstruction des traces dans les données de collisions (3)



- Reconstruction des **vertex primaires** à partir des traces :
 - 1) **positions** des vertex primaires reconstruits dans le plan transverse
 - 2) et 3) **résolutions** en x et z en fonction du nombre de traces associées aux vertex
 - Résolution obtenue
 - en **séparant** aléatoirement, pour chaque événement, la **collection de traces en 2** puis en reconstruisant **2 vertex** différents à partir de chaque collection de trace,
 - Puis à partir du σ de l'**ajustement d'une gaussienne** de la distribution **position(vertex1)-position(vertex2)**



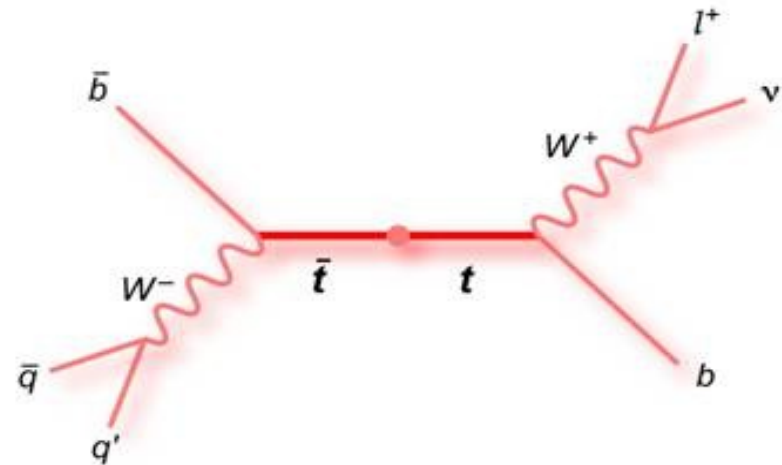
Section efficace $t\bar{t}b\bar{b}$



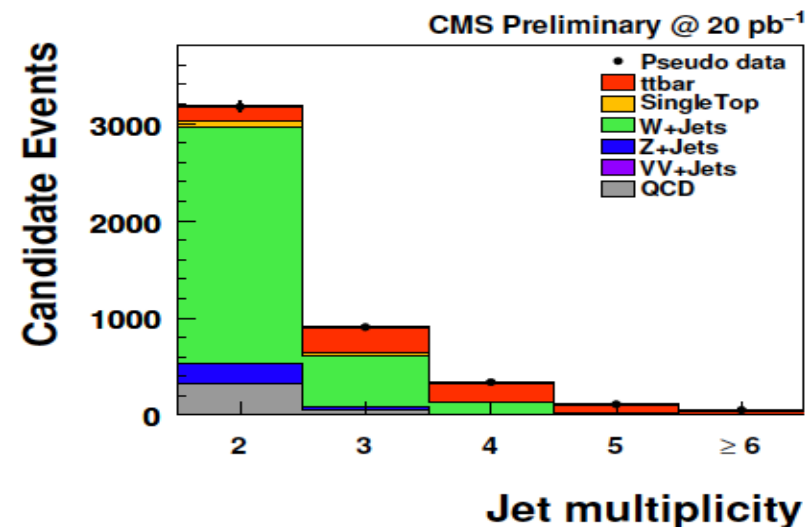
1. La physique du top au LHC,
2. Premiers événements vus par CMS,
3. **Mesure de la section efficace $t\bar{t}b\bar{b}$,**
4. Conclusion



Canal lepton + jets



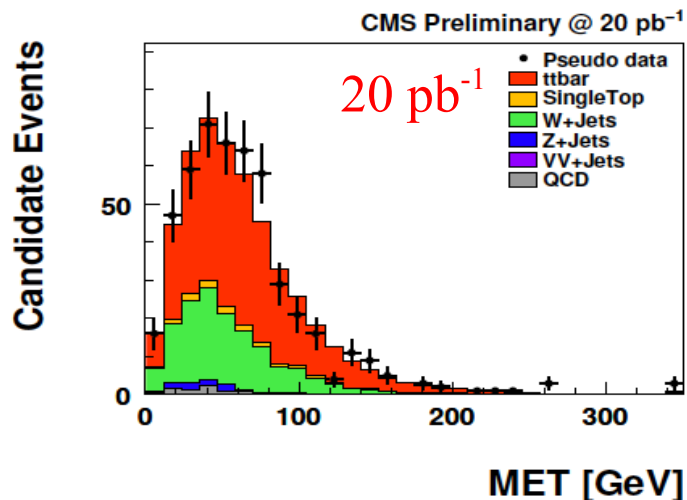
- Canal **lepton+jets**, sensible aux bruits de fond instrumentaux.
- Topologie des événements:
 1. **1 lepton isolé** de “grand p_T ”,
 2. **2 jets** légers ,
 3. E_T manquante (**neutrino**) ,
 4. 2 jets b.



- Pour les premières collisions, la sélection des événements demande uniquement **un lepton bien isolé** avec $p_T > 20$ GeV/c et **4 jets** ayant un p_T corrigé > 30 GeV + **veto** sur la présence de **leptons isolés** supplémentaires.
- Analyse présentée : **canal muon+jets** (moins de bnf instrumentaux) pour $\sqrt{s} = 10\text{TeV}$ et 20 pb⁻¹.

Canal lepton +jets (2)

20 pb ⁻¹	$t\bar{t}$ +jets s.l. μ	$t\bar{t}$ +jets other	Single top			W+jets	Z+jets	VV+jets	QCD
			s-Ch.	t-Ch.	tW				
AllEvents	1,220	7,060	32	832	580	912,000	76,240	236	2,546,279
Trigger	978	1,418	10	260	147	168,633	20,952	100	2,032,021
≥ 1 tight μ	620	345	5	140	69	110,509	15,296	73	7,200
< 2 tight μ	620	309	5	140	66	110,509	9,300	62	7,200
no tight e	620	264	5	140	62	110,508	9,292	53	7,200
veto on loose μ	618	228	5	140	60	110,503	5,492	44	7,192
veto no loose e	616	183	5	140	56	110,469	5,415	34	7,188
≥ 1 jet	614	180	4	125	55	16,998	1,325	18	2,701
≥ 2 jets	593	158	3	63	47	3,076	256	5	387
≥ 3 jets	489	99	1	18	27	651	51	1	60
≥ 4 jets	277	43	0	5	9	140	10	0	7



- Rapport signal sur bruit après la sélection $\approx 1,3$.
- Echantillon **moins pur** que pour les canaux di-leptoniques.
- Bruit de fond principal = **W+jets**.
- Erreur statistique sur le nombre de $t\bar{t}$ sélectionnés $\approx 6\%$.

Estimation du bruit de fond W+jets

- * Estimation du bruit de **fond W+jets** (bruit de fond principal) => sections efficaces de production de W^+ et W^- sont différentes au LHC (collision p-p).
- L'**asymétrie de charge** résultante peut être définie comme $1/R_{\pm}$ avec :

$$R_{\pm}(W) = \frac{N_{W^+} + N_{W^-}}{N_{W^+} - N_{W^-}} = \frac{A_+ \sigma_{W^+} + A_- \sigma_{W^-}}{A_+ \sigma_{W^+} - A_- \sigma_{W^-}}$$

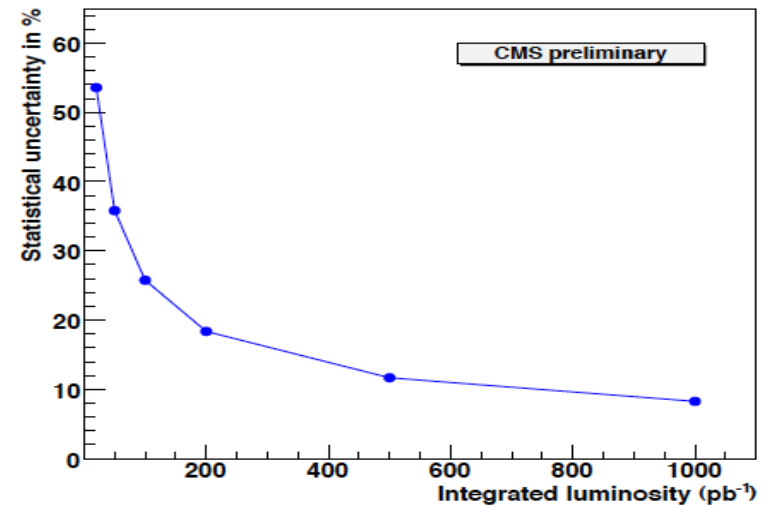
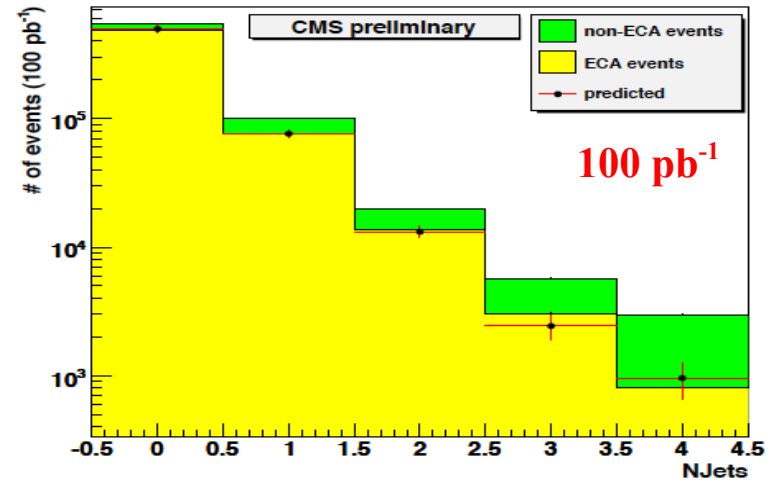
- La **contamination W+jets** peut être estimée en comptant le nombre d'événements sélectionnés avec un lepton (N_-) et un anti-lepton (N_+) :

$$(N_+ + N_-)_{data} = R_{\pm}(W) \times (N_+ - N_-)_{data}$$

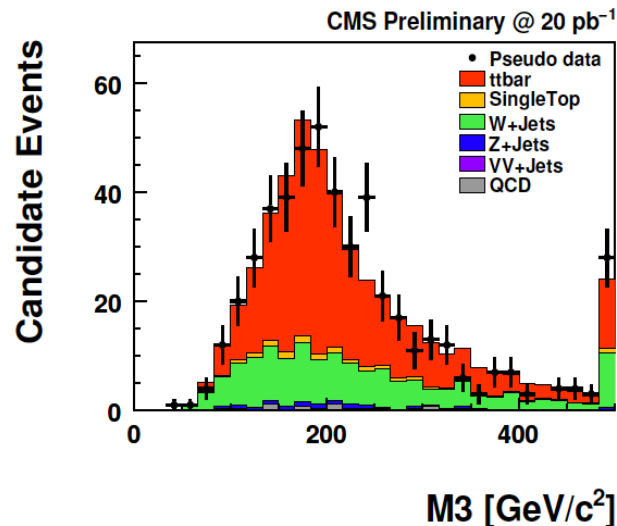
Estimation du bruit de fond W+jets (2)



- ✱ Asymétrie de charge W+jets estimée à partir du [Monte-Carlo](#), peut être mesurée dans les données à basse multiplicité de jet.
- Différentes sources d'[erreur systématique](#) ont été étudiées : JES, PDFs, contamination d'événements ayant des asymétries de charge autres que W+jets (ECA).
- [Sources principales d'erreur](#) : PDFs (8%), ECA (6%).
- [Erreur statistique importante](#) : méthode testée pour $100 \text{ pb}^{-1} \Rightarrow \approx 25\%$.



Canal lepton +jets : estimation de la section efficace



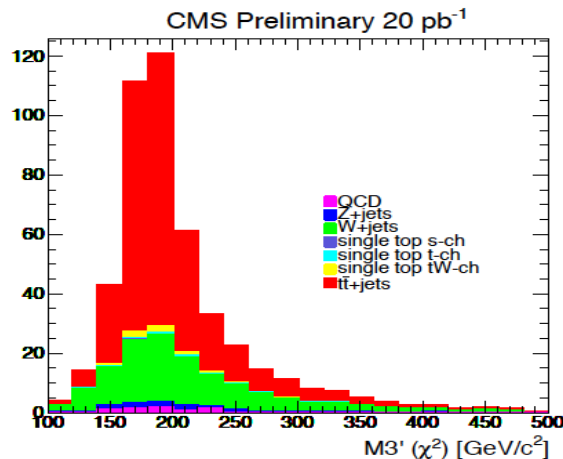
- Bruit de fond QCD estimé à partir des données (méthode ABCD)
- $N_{t\bar{t}}$ est estimé en ajustant les allures des distributions attendues pour le signal et les bruits de fond sur les données (vraisemblance binée)

$$\sigma(t\bar{t}) = \frac{N_{t\bar{t}}}{A \cdot \epsilon \cdot L}$$

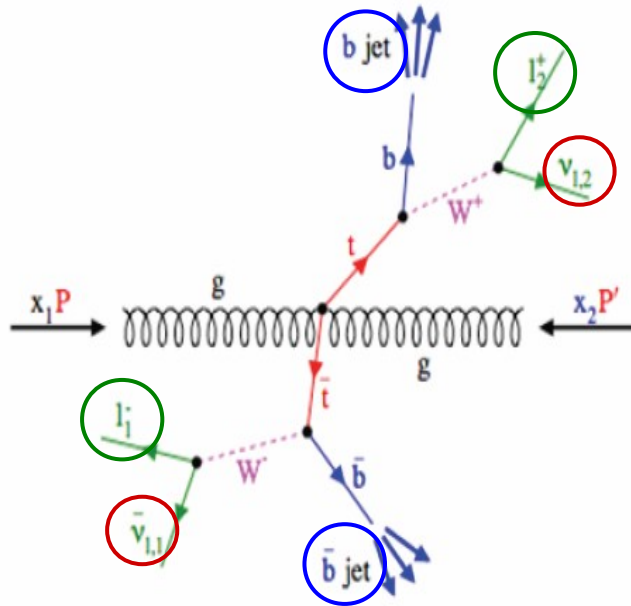
- Plusieurs variables étudiées:
 - M3 : sum p_T 3 jets le plus grand
 - M3'(χ^2):

$$\chi^2 = \frac{(m_{j_1 j_2} - m_W)^2}{\sigma_{jj}^2} + \frac{(m_{j_1 j_2 j_3} - m_t)^2}{\sigma_{jjj}^2} + \frac{(m_{\mu \nu j_4} - m_t)^2}{\sigma_{\mu \nu j}^2}$$

- Incertitude (20 pb⁻¹) 12-18% stat 20-25% syst



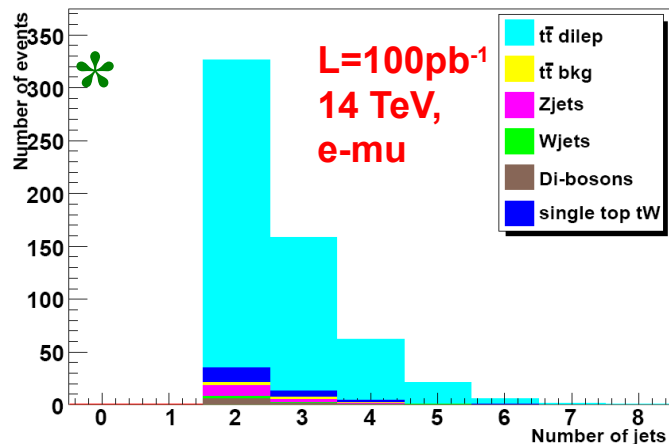
Canal di-lepton : sélection des événements



- **Canal dilepton** (e-e, μ - μ et e- μ), moins sensible aux bruits de fond instrumentaux (QCD multi-jets).
- Sélection des événements :
 - **2 leptons** isolés de “grand” p_T (>20 GeV/c) et de charges opposées \bigcirc ,
 - **2 jets** de “grand” p_T (>30 GeV/c) \bigcirc ,
 - **Veto sur la masse invariante di-lepton** $M_{inv} \neq M_Z$,
 - Large E_T manquante (**MET** $>20-50$ GeV) \bigcirc ,
 - Identification de 2 b-jets (point de fonctionnement “lâche”).
- **Estimation des erreurs systématiques**: liées aux bruits de fond et à l'efficacité de sélection.
- Résultats pour des collisions à **14 TeV**, analyse similaire à **10 TeV** donne des résultats compatibles.

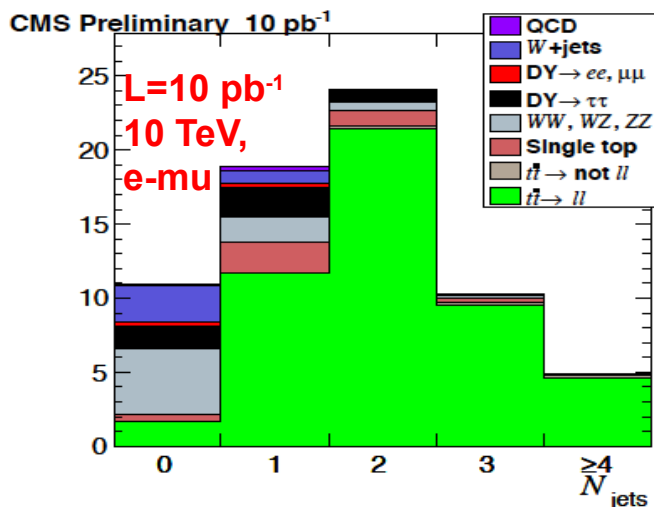


Sélection des événements (2)



sélection ($e - \mu$)	Signal	T T bkg	Z+jets	W+jets	WW	WZ	ZZ	tW	S/B
$\cancel{E}_T > 20$ GeV	544.57 ± 4.34	6.39 ± 0.53	13.89 ± 1.03	3.57 ± 0.63	6.54 ± 0.32	1.26 ± 0.13	0.31 ± 0.06	22.89 ± 0.72	9.93
1 jet b identifié	453.47 ± 4.05	4.86 ± 0.46	3.07 ± 0.46	0.58 ± 0.26	1.24 ± 0.14	0.19 ± 0.05	0.11 ± 0.04	17.07 ± 0.62	16.72
2 jets b identifiés	201.02 ± 2.84	1.59 ± 0.26	0.36 ± 0.16	- -	0.14 ± 0.05	0.01 ± 0.01	0.04 ± 0.02	3.57 ± 0.28	35.20

- **Bruits de fond principaux:** single top voie tW, Z+jets.
- A 14 TeV, lots d'événements purs ($S/B \approx 10$), **sans l'identification des jets b** . Après l'application des coupures en **b-tagging**, la plupart des événements de **bruit de fond** sont rejetés.
- A 10 TeV : S/B plus faible 7, 11 et 16 pour 0, 1 et 2 jets b .
- **Erreurs statistiques 14 TeV (100 pb⁻¹):** 4.3, 4.8 et 7.2% pour 0, 1 ou 2 jets b demandés.
- **Erreurs statistiques 10 TeV (100 pb⁻¹):** 5.3, 6.0 et 7.8% pour 0, 1 ou 2 jets b demandés.
- Canal le plus intéressant au démarrage: bonne statistique et bruits de fond plus faibles.



Estimation des bruits de fond: la méthode de la matrice



- Initialement développée par l'expérience DØ pour estimer le bruit de fond QCD dans le canal lepton+jets.
- Dans les canaux di-leptons: l'isolation des 2 leptons est variée, ce qui conduit à un système de 3 équations à 3 inconnues.
- 3 ensembles de coupures :
 - loose = pas de coupure d'isolation,
 - medium = au moins un lepton sélectionné qui passe les coupures d'isolation,
 - tight = au moins deux leptons sélectionnés qui passent les coupures d'isolation => sélection standard.
- Trois types d'événements sont définis:
 - de type signal (2 vrais leptons isolés, ttbar signal, Z, dibosons),
 - de type W+jets (1 vrai+un faux lepton isolé, W, ttbar lepton+jets),
 - de type QCD (2 faux leptons isolés, QCD, ttbar tout hadronique).
- Connaissant les probabilités qu'un événement loose passe les coupures medium et tight, estimation des bruits de fond.
- Les probabilités qu'un événement loose passe les coupures medium et tight seront déterminées à partir des données (ϵ_s et ϵ_{fake}).



La méthode de la matrice (2)



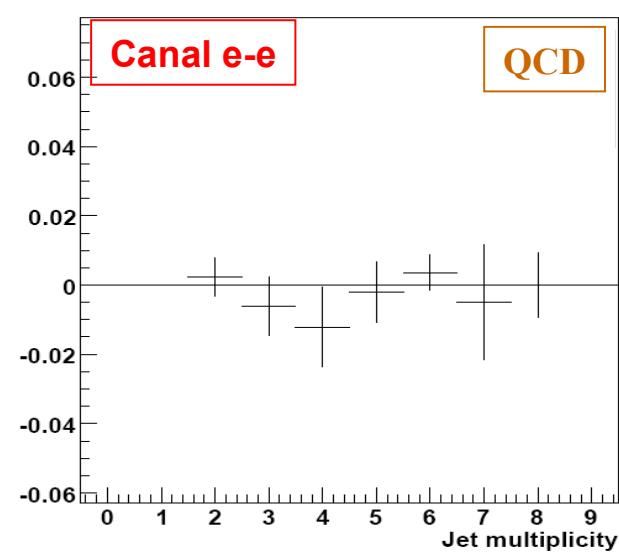
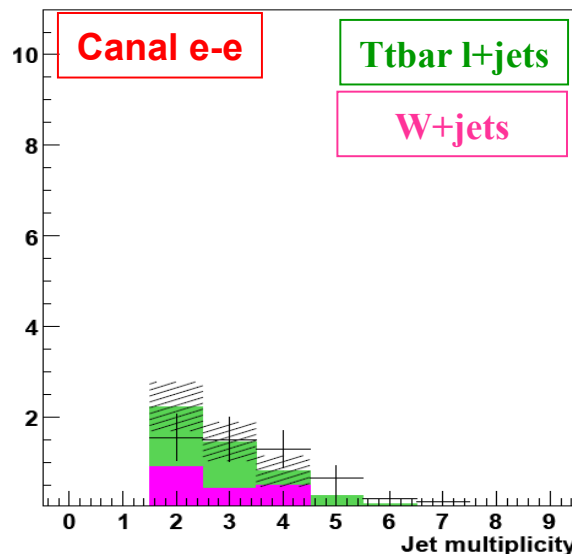
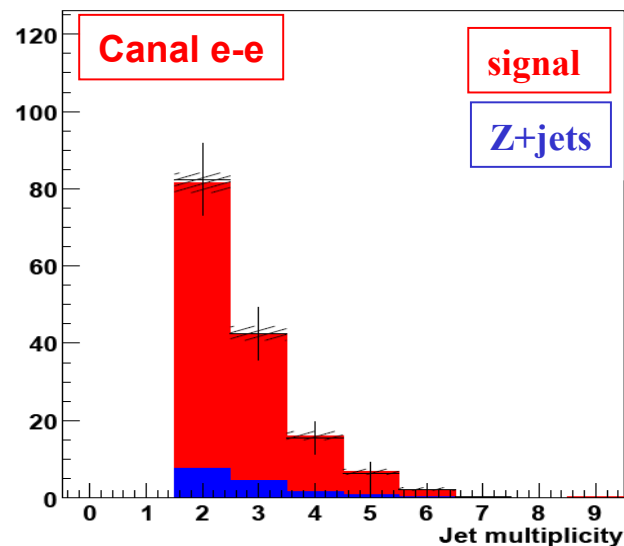
- Les erreurs sont estimées au moyen de **pseudo-expériences**.

- Les sources d'erreur sont :
 - statistique...
 - Erreurs sur ϵ_s et ϵ_{fake} , estimées à 5 et 10% respectivement.

	Type Signal	Type W +jets	Type QCD
Nbr. d'événements	149.1	5.3	-0.019
Erreur stat.	12.7	0.7	0.012
Erreur $\epsilon_s, \epsilon_{fake}$	0.9	0.9	0.013
Erreur totale	12.7	1.1	0.018
Nbr. d'événements (MC)	149.6 ± 3.5	4.9 ± 0.8	0

Erreur rel : 9%

Erreur rel : 21%



Précision de la mesure à 14 TeV



- **Précision** sur la mesure de la section efficace estimée pour les **erreurs systématiques** principales.

$$\sigma_{t\bar{t}} \times BR = \frac{N_{sel} - N_{bdf}}{\varepsilon_{t\bar{t}} \times \mathcal{L}}$$

e-μ	0 jet b	1 jet b	2 jet b
JES	7%	7%	7%
B-tagging	-	5%	15%
Frag. jets légers	3%	3%	3%
Frag. jets lourds	3%	3%	3%
Ev. sous jacent	2%	2%	2%
Masse du top	3%	3%	3%
ISR/FSR	17%	19%	19%

Probablement surestimé
(paramètres ajustés à partir de
données LEP et Tevatron!)

Résultats similaires à 10 TeV

Les incertitude sur les **bruits de fond instrumentaux** : méthode de la matrice (20%) + ajustement du pic de masse du Z (30%).

Les incertitudes sur les **bruits de fond physiques** (principalement tW) restent à être étudiées...

Pour rester conservatif, l'incertitude sur (tW) est prise à 50%.

Efficacité d'identification de jets b et taux d'identification estimé à partir des données*.

$$\begin{aligned} \frac{\Delta\sigma}{\sigma}(0 \text{ jet } b) &= 3.8 \% (\text{Stat.}) \oplus 19.6 \% (\text{Syst.}) \oplus 10.0 \% (\text{Lumi.}) \\ \frac{\Delta\sigma}{\sigma}(1 \text{ jet } b) &= 4.0 \% (\text{Stat.}) \oplus 21.7 \% (\text{Syst.}) \oplus 10.0 \% (\text{Lumi.}) \\ \frac{\Delta\sigma}{\sigma}(2 \text{ jets } b) &= 5.8 \% (\text{Stat.}) \oplus 30.2 \% (\text{Syst.}) \oplus 10.0 \% (\text{Lumi.}) \end{aligned}$$



Conclusion



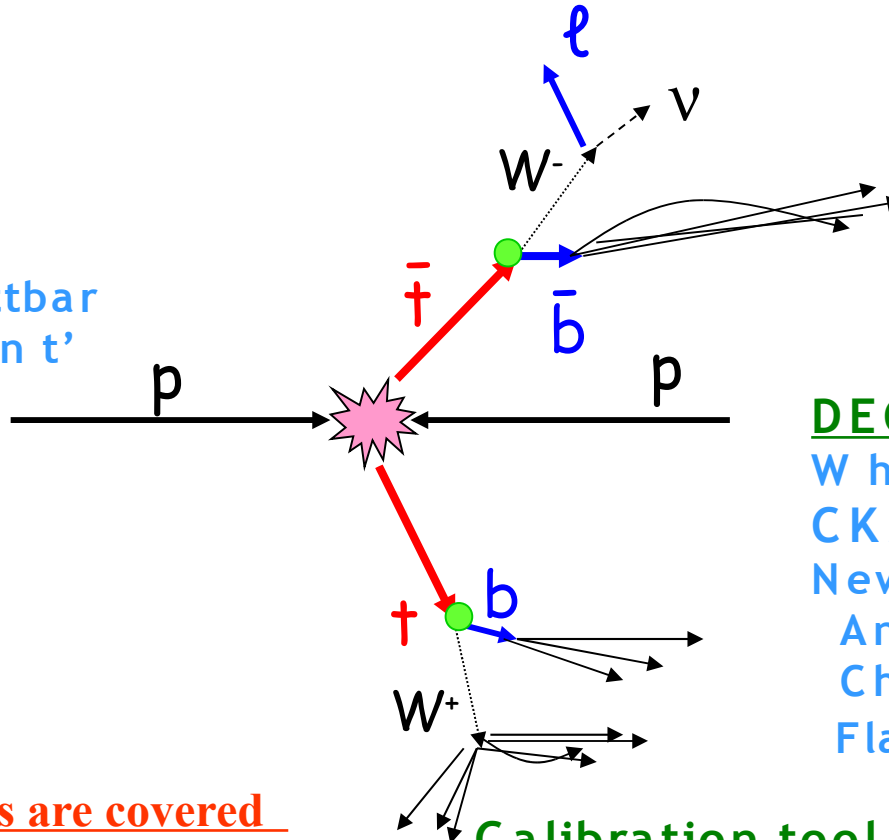
- Les analyses de physique liées au top sont **en préparation** depuis plusieurs années à l'aide d'événements Monte-Carlo.
- Des **stratégies d'analyses** ont été mises en place pour des **luminosités $\geq 20 \text{ pb}^{-1}$** et des énergie de 10 à 14 TeV (analyses à 7 TeV sont en préparation).
- Pour de plus faibles luminosités, **la compréhension du détecteur**, des **événements** et des **objets** reconstruits sera l'étape **primordiale** à toute analyse de physique.
- Cette étape importante a commencé avec la prise de données, mais ne pourra être validée qu'avec **des collisions de haute énergie** ($\geq 7 \text{ TeV}$ dans le c.d.m.).
- Une stratégie précise de **compréhension des données du point de vue du groupe top** est en train d'être mise en place : compréhension des chemins de déclenchement, validation des données et du MC, compréhension des toutes premières données etc...

Backups

Topics in top quark physics

PRODUCTION

Top anti-top
 Single top
 Spin-correlations
 New physics :
 Resonances $X \rightarrow t\bar{t}$
 Fourth generation t'
 Charged Higgs
 (Susy)



PROPERTIES

Mass
 Spin
 Charge
 Life-time and width

$$\tau = 0.4 \cdot 10^{-24} \text{ s}$$

DECAY

W helicity
 CKM matrix elements
 New physics:
 Anomalous couplings
 Charged Higgs
 Flavour physics (FCNC)

Calibration tool
 Jet Energy Scale
 B-tagging

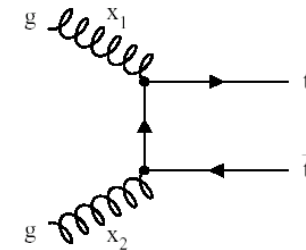
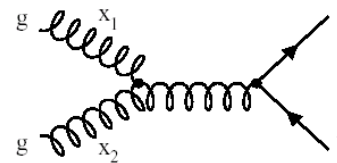
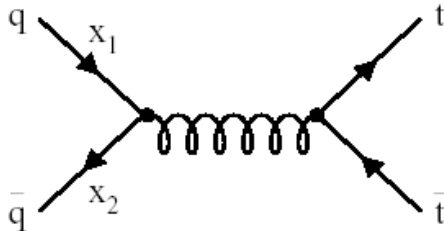
Most of these items are covered
 by CMS analysis

Production du quark top

10 paires $t\bar{t}$ par jour au Tevatron
 $q\bar{q} \rightarrow t\bar{t} : 85\%$

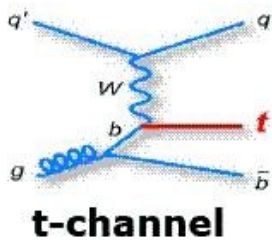


1 paire par seconde au LHC
 $g\bar{g} \rightarrow t\bar{t} : 87\%$

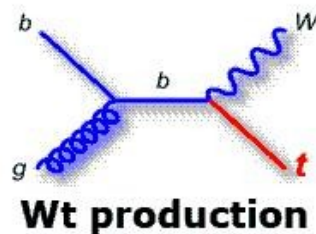


single-top au Tevatron
 Première mise en évidence (à 3.4σ)

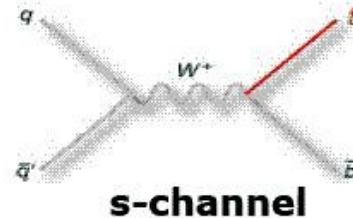
→ 30 single-tops par minute au LHC



NLO = 153 pb
 NLO = 90 pb



NLO = 32 pb
 NLO = 32 pb



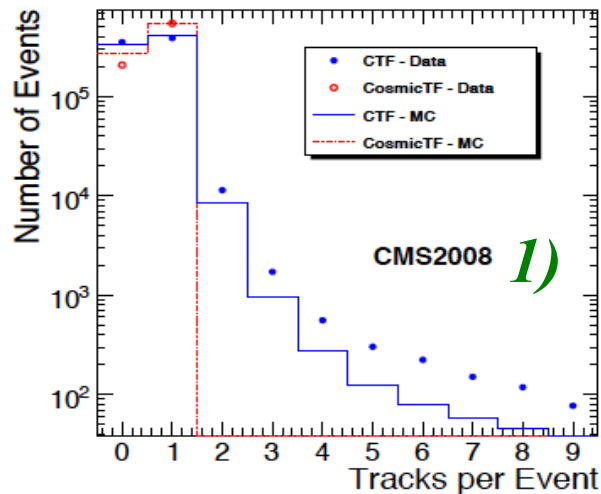
NLO = 6.6 pb
 NLO = 4.1 pb
 Top
 Anti-Top

σ_{top} & $\sigma_{\text{anti-top}}$ sont différents

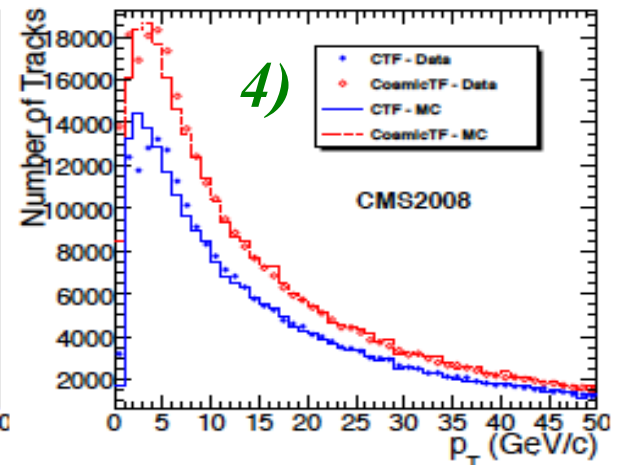
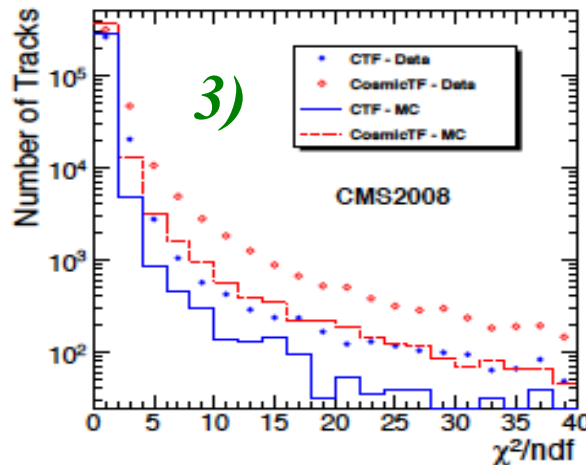
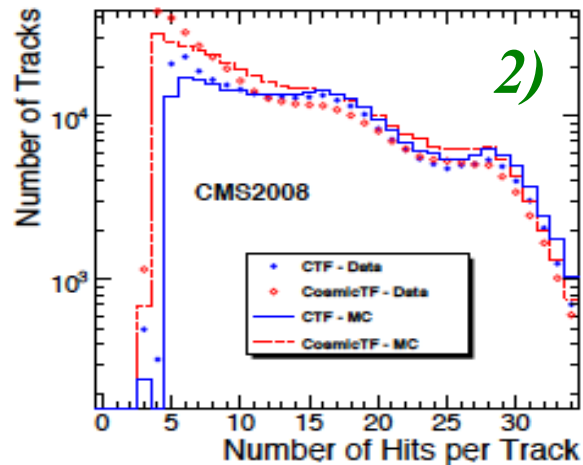
NLO(total) = 373 pb
 → ~3.7M événements/10fb⁻¹

14 TeV

Résultats des analyses cosmique sélection



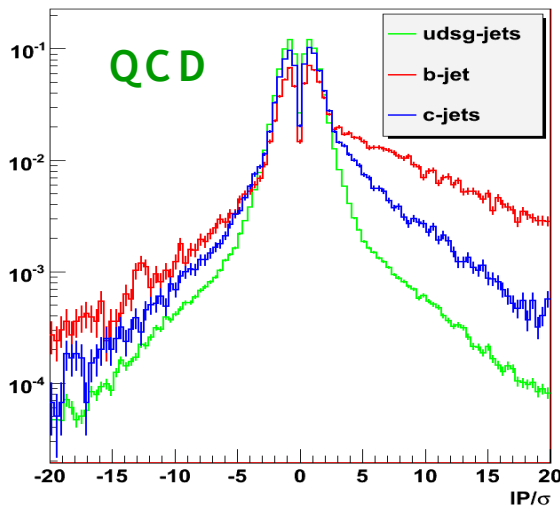
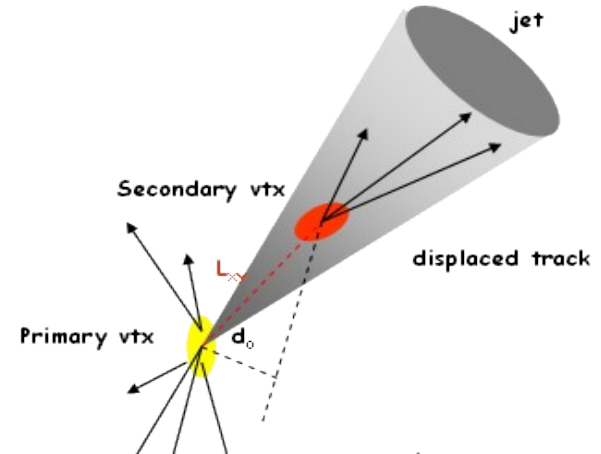
- 1) nombre de traces par événements,
- 2) nombre de coups par traces,
- 3) χ^2 normalisé,
- 4) impulsion transverse



Identification des jets b



- **Caractéristiques des hadrons beaux:**
 - ✱ – Grand temps de vie ($\approx \text{ps}$, $c\tau \approx 500 \mu\text{m}$),
 - Vertex secondaire,
 - Désintégrations leptoniques,
 - Masses plus élevées.
- **Le paramètre d'impact (IP)** des traces (par rapport à l'axe des jets).
- **Signe du IP** : signe du cosinus de l'angle d_0 -axe du jet.



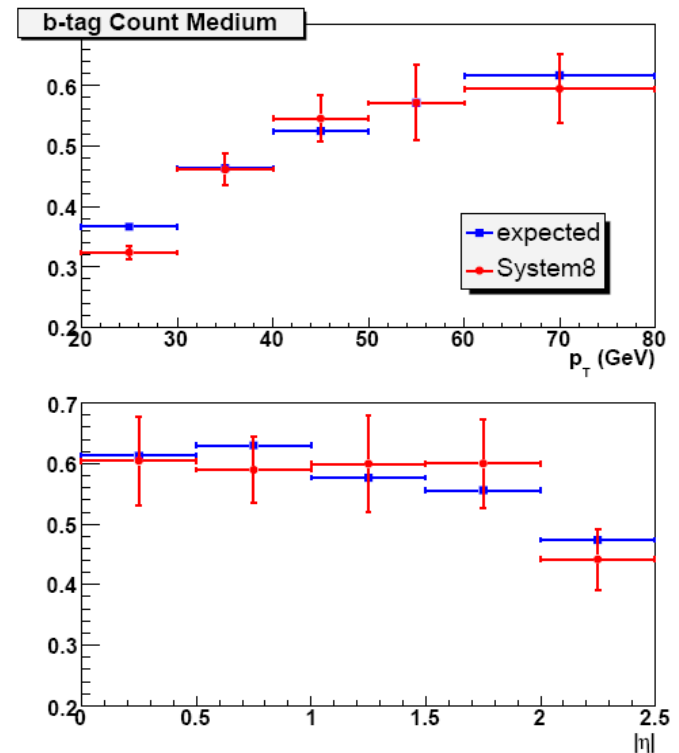
- **Effets de résolution** => signification du paramètre d'impact IP/σ .
- **Paramètre d'impact des traces:**
- **Pour les jets légers (udsg)** : symétriques autour de 0, effets de résolution. **Pour les jets b(c)** : asymétriques avec une plus grande fraction de traces à grand paramètre d'impact positif.
- Algorithme de comptage de traces : au moins 2(3) traces ayant une $IP/\sigma >$ à une coupure.
- Pour un taux de mauvaise Id de **1%**, efficacité \approx **65%**.



Mesure de l'efficacité d'identification



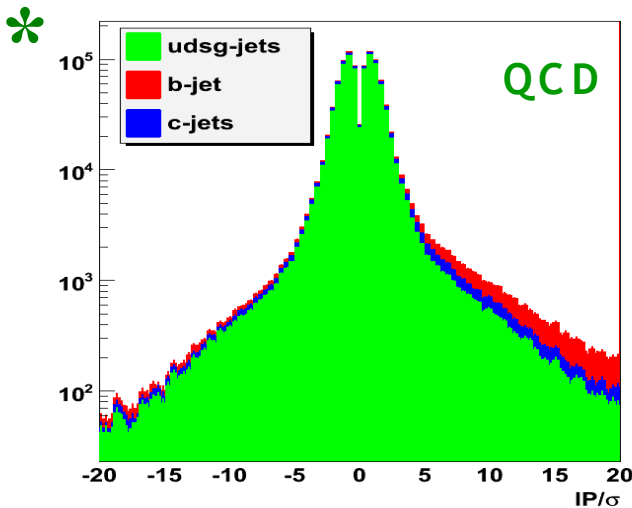
- ✱ **Méthode System8** (initialement développée par l'expérience D0).
- 3 critères d'identification faiblement corrélés** sont utilisés sur des événements contenant **des muons dans les jets**:
 - L'algorithme étudié,
 - Une coupure sur le p_T^{rel} du muon,
 - La présence d'un jet b additionnel.
- Un système de **8 équations à 8 inconnues** est construit. Sa résolution permet d'estimer **l'efficacité d'identification des jets b**.
- Des facteurs sont introduits pour prendre en compte les corrélations entre les différents critères d'identification. Ils sont estimés à partir du **Monte-Carlo**.
- Le système est résolu pour différents **bins en p_T et $|\eta|$** des jets.



Pour une $L=100\text{pb}^{-1}$ et un taux de mauvaise identification de 1%, l'erreur attendue est de 9%.

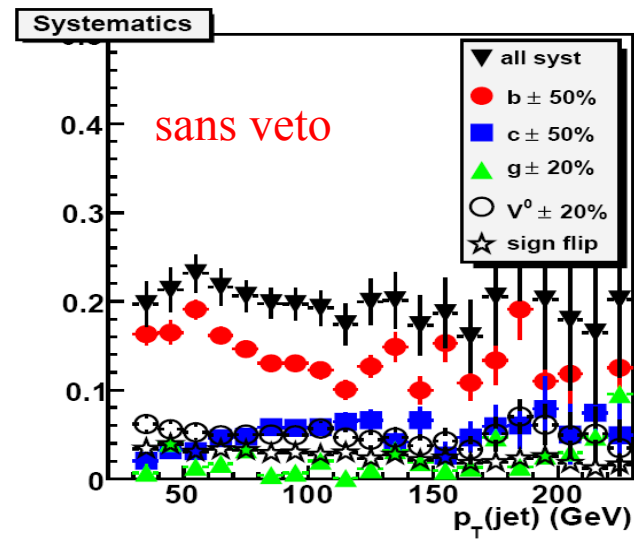


Taux de mauvaise identification



- Utilise la symétrie (au premier ordre) de la distribution de IP/σ pour les jets de quarks légers et de gluon.
- On peut définir un algorithme “négatif” qui utilise les traces avec $IP/\sigma < 0$ pour estimer $\epsilon^{IP/\sigma < 0}$ dans les données = estimation du taux de mauvaise identification.
- Utilisation de données ayant une faible fraction de jets b et c: QCD multi-jets.

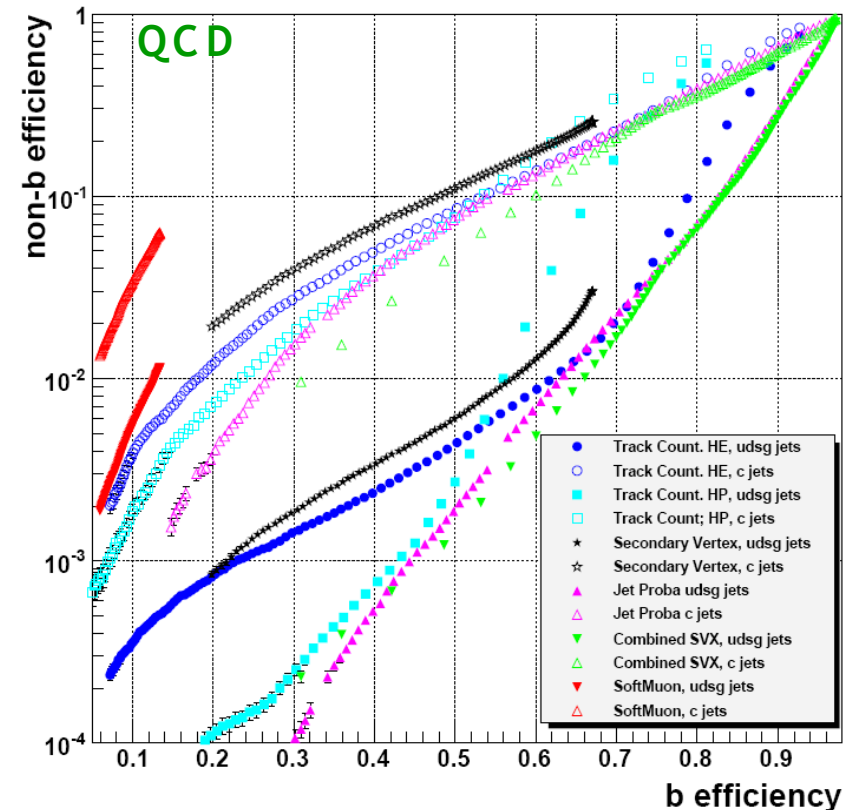
- Estimation doit être corrigée (facteur MC):
 - Asymétrie de la distribution IP/σ pour les jets légers ($V0$, $\gamma \rightarrow ee$, fausses traces), fraction de saveur lourde étiquetée négativement (source principale).
- Peut être réduite d'un facteur 4 par l'introduction d'un veto sur les jets (jet contenant des traces avec un grand IP/σ).
- Pour un taux de mauvaise identification de 1% et une $L=100\text{pb}^{-1}$, l'incertitude attendue est de $\approx 8\%$.



Identification des jets b



- Les autres algorithmes d'identification de jets b de CMS:
 - Reconstruction du **vertex secondaire** dans les jets (efficacité de reconstruction $\approx 65\%$). Le discriminant est la distance de vol.
 - Algorithme combiné** : IP/ σ , vertex secondaire, masse invariante, pseudo-rapidité par rapport à l'axe du jet et pt des traces associées au VS etc...
 - Soft lepton** : dans $\approx 40\%$ des cas, un jet b contient un électron ou un muon (désintégration faible des hadrons beaux et charmés) \Rightarrow discriminant possible IP/ σ , p_{T}^{rel} du lepton
 - Algorithme le plus performant** : Algorithme combiné. Il manque cependant de robustesse au démarrage car il est calibré à partir d'événements Monte-Carlo.
 - Les algorithmes **de comptage de traces** donnent de très bons résultats compte tenu de leur **simplicité**.



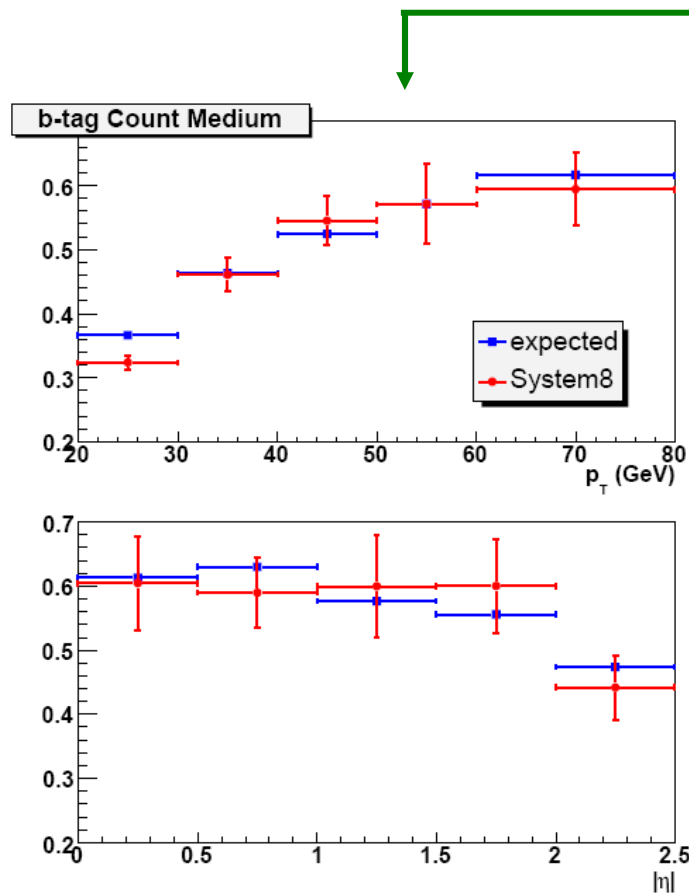
Mesure de l'efficacité d'identification



- Méthode System8 (initialement développée par l'expérience DØ).
- 3 critères d'identification faiblement corrélés sont utilisés sur des événements contenant des muons dans les jets:
 - L'algorithme étudié,
 - Une coupure sur le p_T^{rel} du muon,
 - La présence d'un jet b additionnel.
- Un système de 8 équations à 8 inconnues est construit. Sa résolution permet d'estimer l'efficacité d'identification des jets b.
- Des facteurs sont introduits pour prendre en compte les corrélations entre les différents critères d'identification. Ils sont estimés à partir du Monte-Carlo.
- Le système est résolu pour différents bins en p_T et η des jets.



Mesure de l'efficacité d'identification (3)



- Résultats pour l'algorithme du **comptage de traces avec N=2** pour le point de fonctionnement medium.
- Différentes sources d'incertitudes ont été étudiées:
 - Les erreurs **liées aux facteurs de corrélations**,
 - Les erreurs **liées à la valeur de la coupure en p_T^{rel}** ,
- Les **précisions attendues** pour les différents points de fonctionnement et pour différentes luminosités sont résumées (détecteur parfait!).

Points de fonctionnement	<i>loose</i>			<i>medium</i>			<i>tight</i>		
Luminosité (pb^{-1})	10	100	1000	10	100	1000	10	100	1000
Systématiques (%)									
β	5.8	5.8	2.9	6.3	6.3	3.2	5.7	5.7	2.9
α	0.4	0.4	0.2	0.4	0.4	0.2	0.4	0.4	0.2
κ_b	3.4	3.4	1.7	3.6	3.6	1.8	3.3	3.3	1.7
κ_{cl}	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.1
p_T^{rel}	2.8	2.8	2.8	2.9	2.9	2.9	3.0	3.0	3.0
statistique MC (%)	2.3	2.3	2.3	2.6	2.6	2.6	2.7	2.7	2.7
statistique data (%)	7.2	2.3	0.7	8.4	2.6	0.8	8.7	2.7	0.9
Erreur totale(%)	10.5	8.0	6.4	11.8	8.6	5.4	11.6	8.2	5.3



$$n = n_b + n_{cl}$$

$$p = p_b + p_{cl}$$

$$n^{\text{tag}} = \varepsilon_b^{\text{tag}} n_b + \varepsilon_{cl}^{\text{tag}} n_{cl}$$

$$p^{\text{tag}} = \beta \varepsilon_b^{\text{tag}} p_b + \alpha \varepsilon_{cl}^{\text{tag}} p_{cl}$$

$$n^{\text{mu}} = \varepsilon_b^{\text{mu}} n_b + \varepsilon_{cl}^{\text{mu}} n_{cl}$$

$$p^{\text{mu}} = \varepsilon_b^{\text{mu}} p_b + \varepsilon_{cl}^{\text{mu}} p_{cl}$$

$$n^{\text{tag,mu}} = \kappa_b \varepsilon_b^{\text{tag}} \varepsilon_b^{\text{mu}} n_b + \kappa_{cl} \varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{mu}} n_{cl}$$

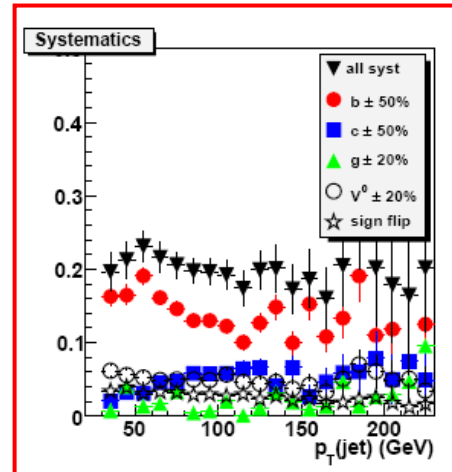
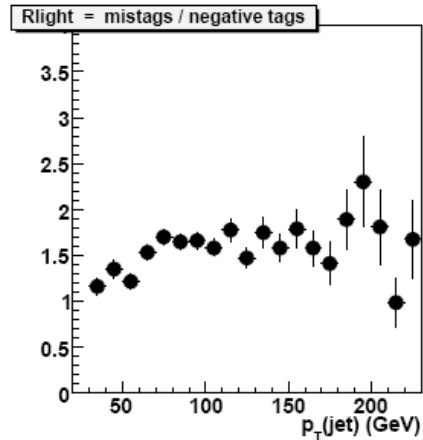
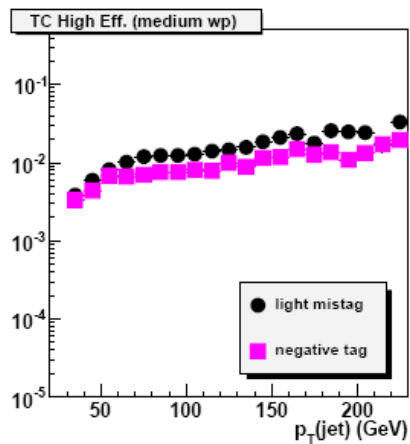
$$p^{\text{tag,mu}} = \kappa_b \beta \varepsilon_b^{\text{tag}} \varepsilon_b^{\text{mu}} p_b + \kappa_{cl} \alpha \varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{mu}} p_{cl} .$$

$$\kappa_b = \frac{\varepsilon_b^{\text{tag,mu}}}{\varepsilon_b^{\text{tag}} \varepsilon_b^{\text{mu}}}$$

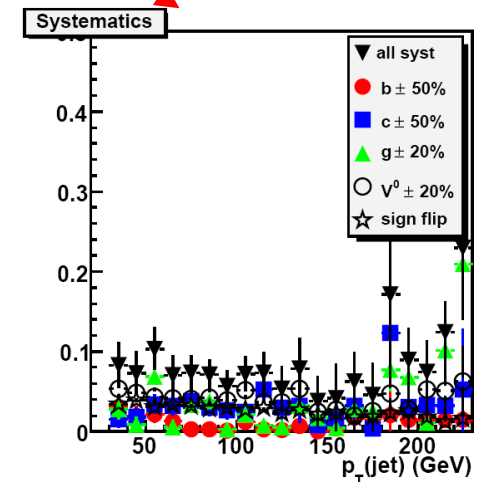
$$\kappa_{cl} = \frac{\varepsilon_{cl}^{\text{tag,mu}}}{\varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{mu}}}$$

$$\beta = \frac{\varepsilon_b^{\text{tag}} (\text{muon-jet} + \text{jet additionnel identifié})}{\varepsilon_b^{\text{tag}} (\text{muon-jet})} \quad \alpha = \frac{\varepsilon_{cl}^{\text{tag}} (\text{muon-jets} + \text{jet additionnel identifié})}{\varepsilon_{cl}^{\text{tag}} (\text{muon-jet})}$$

Taux de mauvaise identification



Avec Jet Veto



Comptage de traces
 $N=2$
Avec le point de
fonctionnement
Medium.



Taux de mauvaise identification



Point de fonctionnement	<i>Loose</i>			<i>Medium</i>			<i>Tight</i>		
Luminosité (pb ⁻¹)	10	100	1000	10	100	1000	10	100	1000
Systematiques (%)									
fraction de <i>b</i>	1.4	1.4	0.6	0.8	0.8	0.3	1.2	1.2	0.5
fraction de <i>c</i>	0.8	0.8	0.3	0.7	0.7	0.3	1.3	1.3	0.5
fraction de <i>g</i>	0.8	0.8	0.4	1.4	1.4	0.7	2.3	2.3	1.2
fraction de <i>V</i> ⁰	1.4	1.4	0.7	3.6	3.6	1.8	4.6	4.6	2.3
autres processus déplacé	1.4	1.4	0.7	3.6	3.6	1.8	4.6	4.6	2.3
Inversion de signe	0.7	0.3	0.2	4.5	1.9	1.4	24.0	10.2	7.6
statistique MC	0.1	0.1	0.1	0.4	0.4	0.4	1.2	1.2	1.2
statistique data	0.4	0.1	—	1.6	0.5	0.2	5.5	1.7	0.6
trigger	2.0	2.0	2.0	5.0	5.0	5.0	13.0	13.0	13.0
Total syst.	3.4	3.4	2.4	8.8	7.6	5.9	28.7	18.1	15.5

- Pour les algorithmes de comptage de traces, les incertitudes sur le taux de mistag ont été estimées pour différents points de fonctionnement et différentes luminosités (statistique).
- Des algorithmes d'étiquetage négatif ont été définis pour tous les algorithmes d'identification de jets *b*.
- L'estimation du taux de mistag est ainsi possible pour tous les algorithmes de *b*-tagging de CMS.



Variable d'isolation



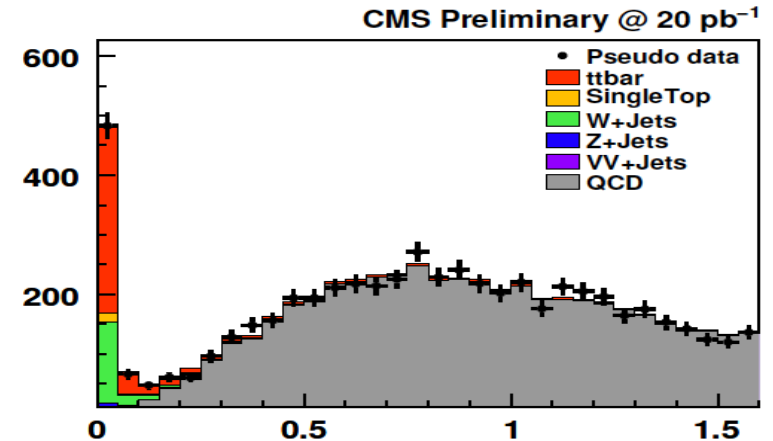
Lepton+jets :

- Isolation trajectographique (calorimétrique) = somme des p_T des traces (des tours calorimétriques) dans un cône $\Delta R=0.3$ autour du lepton – p_T du lepton.

$$I = \sum_{traces} p_T - p_T^{lepton}$$

- $Rellso = (iso\ calo + iso\ tracker) / p_T$.

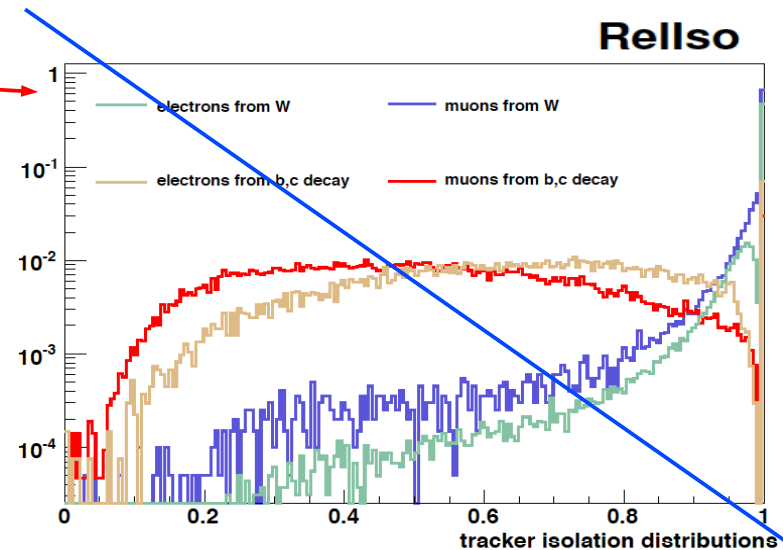
Candidate Events



Di-lepton :

- Isolation trajectographique (calorimétrique) = p_T du lepton / (p_T du lepton + somme des impulsions transverses des traces (des tours calorimétriques) dans un cône $\Delta R=0.3$ autour du lepton)

$$I = p_T / \sum_{traces} p_T$$



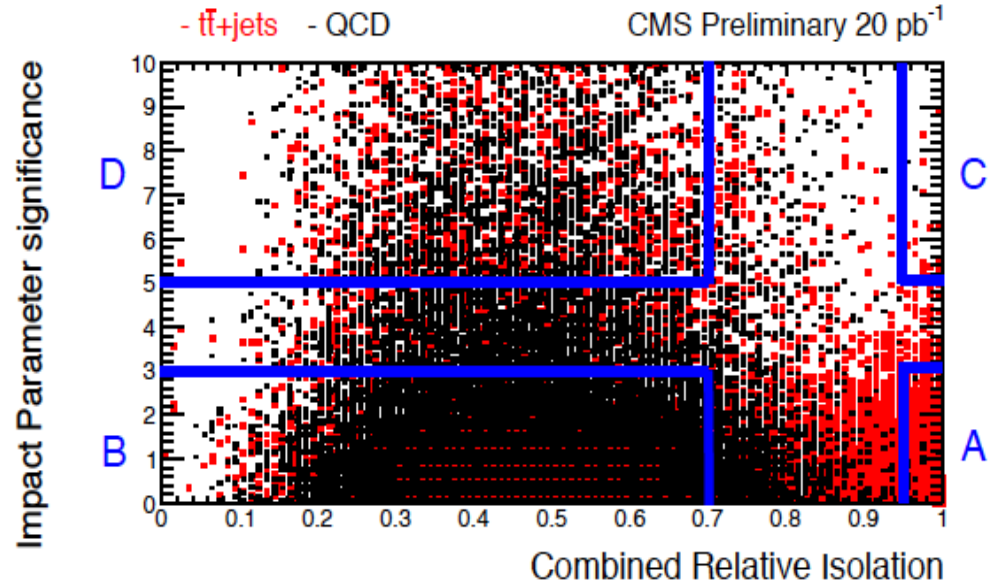
Estimation du bruit de fond QCD

- Estimation du bruit de fond QCD => la méthode ABCD.
- Utilise 2 variables faiblement corrélées ayant un bon pouvoir de discrimination entre signal et bruit de fond.
- Région A dominée par le signal, régions B, C et D par le bruit de fond.

Rapport $N_A/N_B = N_C/N_D$

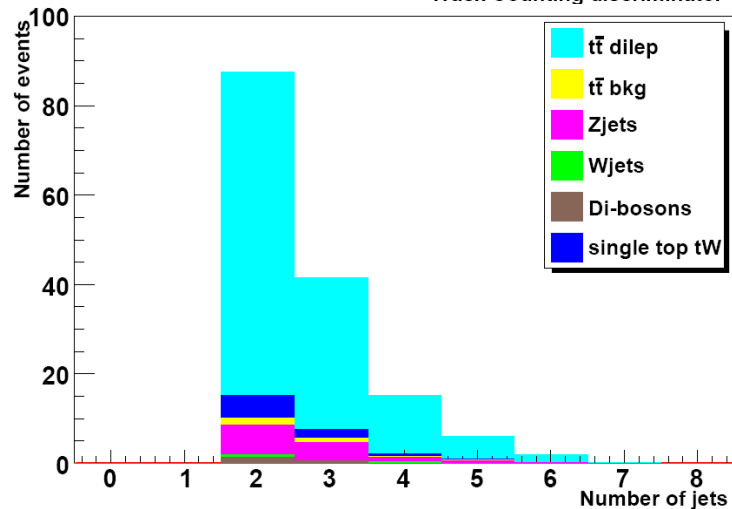
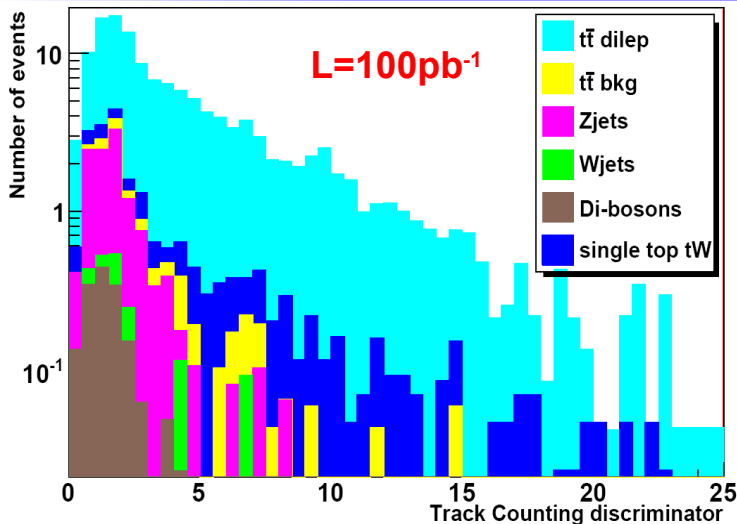
$$N_A = (N_B N_C) / N_D$$

- Méthode testée avec les variables : **significance du IP transverse** du muon vs **isolation** du muon.
- Erreur de la méthode $\approx 50\%$.



Jets	N(QCD) Predicted	N_B	N_C	N_D	N(QCD) Estimated
2	327	86625	61	16240	325 ± 26
3	53	24216	10	5058	48 ± 9
≥ 4	7	5345	3	1148	12 ± 5

Sélection des événements (3)

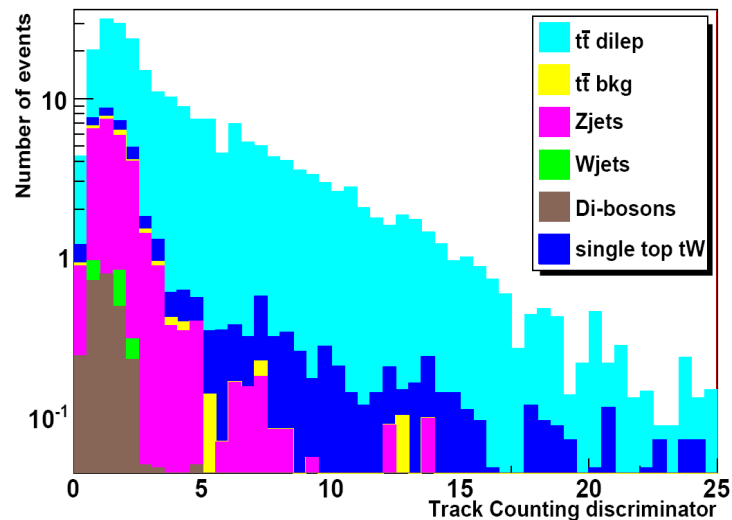


sélection ($e - e$)	Signal	TTbkg	Z+jets	W+jets	WW	WZ	ZZ	tW	S/B
$E_T > 50$ GeV	134.59 ± 2.30	2.69 ± 0.34	12.58 ± 0.98	0.85 ± 0.28	1.64 ± 0.16	0.25 ± 0.06	0.28 ± 0.06	8.05 ± 0.43	5.11
1 jet b identifié	112.15 ± 2.12	2.21 ± 0.31	3.33 ± 0.50	0.46 ± 0.20	0.37 ± 0.08	0.05 ± 0.03	0.09 ± 0.03	5.91 ± 0.37	9.03
2 jets b identifiés	49.59 ± 1.44	0.74 ± 0.18	0.83 ± 0.25	-	0.03 ± 0.02	-	0.01 ± 0.01	1.32 ± 0.17	16.92

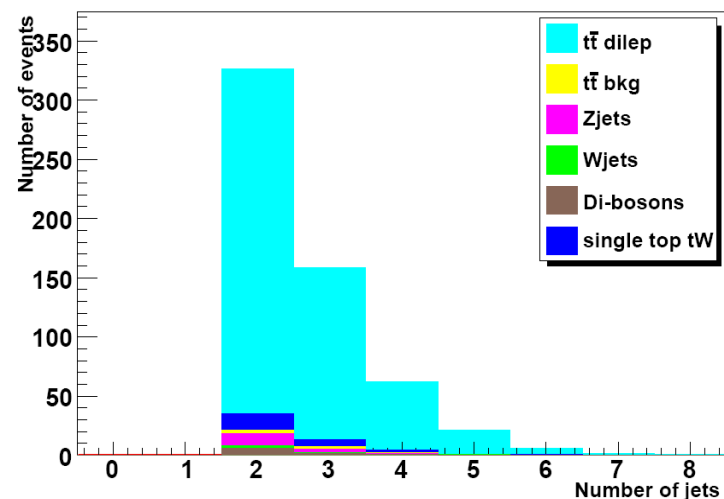
- Signal pur avec peu de bruit de fond (S/B entre 5 et 17).
- Avant b-tagging, bruit de fond principal Z+jets.
- Après b-tagging, bruit de fond principal single top.
- Erreurs statistiques 14 TeV: 9, 10 et 15% pour 0, 1 ou 2 jets b demandés.
- Erreurs statistiques 10 TeV: 10, 11 et 15% pour 0, 1 ou 2 jets b demandés.
- Les effets liés aux performances du détecteur peuvent être importants au démarrage: les bruits de fond « instrumentaux » doivent être estimés à partir des données



Sélection des événements



sélection ($\mu - \mu$)	Signal	TTbkg	Z+jets	W+jets	WW	WZ	ZZ	tW	S/B
$\cancel{E}_T > 50$ GeV	223.02 ± 2.84	1.92 ± 0.29	36.19 ± 1.69	0.94 ± 0.33	3.09 ± 0.22	0.49 ± 0.08	0.52 ± 0.08	12.23 ± 0.53	4.02
1 jet b identifié	188.09 ± 2.65	1.31 ± 0.24	8.96 ± 0.82	0.08 ± 0.08	0.49 ± 0.09	0.12 ± 0.04	0.12 ± 0.038	8.75 ± 0.45	9.48
2 jets b identifiés	87.74 ± 1.89	0.36 ± 0.13	1.57 ± 0.34	- -	0.06 ± 0.03	0.01 ± 0.01	0.03 ± 0.02	2.09 ± 0.22	21.30



La méthode de la matrice (2)



- N_s = signal + bruits de fond physiques (Z+jets, dibosons, tW),
- ✿ N_{W+jets} = W+jets + tt semi-leptonic (1 faux lepton isolé),
- N_{QCD} = QCD, tt full-hadronique (2 faux leptons isolés).
- On peut alors introduire les efficacités de « passer » des échantillons loose à medium et loose à tight: $\epsilon^{l \rightarrow t}$ et $\epsilon^{l \rightarrow m}$.

$$\begin{aligned} N^t &= \epsilon_S^{l \rightarrow t} N_S^l + \epsilon_W^{l \rightarrow t} N_W^l + \epsilon_{QCD}^{l \rightarrow t} N_{QCD}^l, \\ N^m &= \epsilon_S^{l \rightarrow m} N_S^l + \epsilon_W^{l \rightarrow m} N_W^l + \epsilon_{QCD}^{l \rightarrow m} N_{QCD}^l, \\ N^l &= N_S^l + N_W^l + N_{QCD}^l. \end{aligned}$$

Peut s'exprimer en fonction de l'efficacité d'isolation des leptons et du taux de faux leptons isolés.

- Si les efficacités $\epsilon^{l \rightarrow t}$ et $\epsilon^{l \rightarrow m}$ peuvent être mesurées avec une bonne précision, on a défini un système de 3 équations à 3 inconnues: N'_s , N'_W et N'_{QCD} .
- Après résolution du système, les nombres d'événements de type signal, de type W+jets et de type QCD sont:

$$\begin{aligned} N_S^t &= \epsilon_S^{l \rightarrow t} N_S^l, \\ N_W^t &= \epsilon_W^{l \rightarrow t} N_W^l, \\ N_{QCD}^t &= \epsilon_{QCD}^{l \rightarrow t} N_{QCD}^l \end{aligned}$$



La méthode de la matrice (3)



Erreurs systématiques:

- Erreur sur les nombres d'événements sélectionnés dans les lots loose, medium et tight.
- Erreur sur les efficacités ϵ_s et ϵ_{fake} . Hypothèse : ces valeurs sont connues à 5% et 10 % près.

Les erreurs sont estimées en variant ϵ_s et ϵ_{fake} de 5 et 10% et en variant les nombres d'événements dans les lots loose, medium et tight suivant une distribution de Poisson (méthode de pseudo-expériences)

	Type Signal	Type W+jets	Type QCD
Nbr. d'événements	149.1	5.3	-0.019
Erreur stat.	12.7	0.7	0.012
Erreur $\epsilon_s, \epsilon_{fake}$	0.9	0.9	0.013
Erreur totale	12.7	1.1	0.018
Nbr. d'événements (MC)	149.6 ± 3.5	4.9 ± 0.8	0

Erreur rel : 9%

Erreur rel : 21%

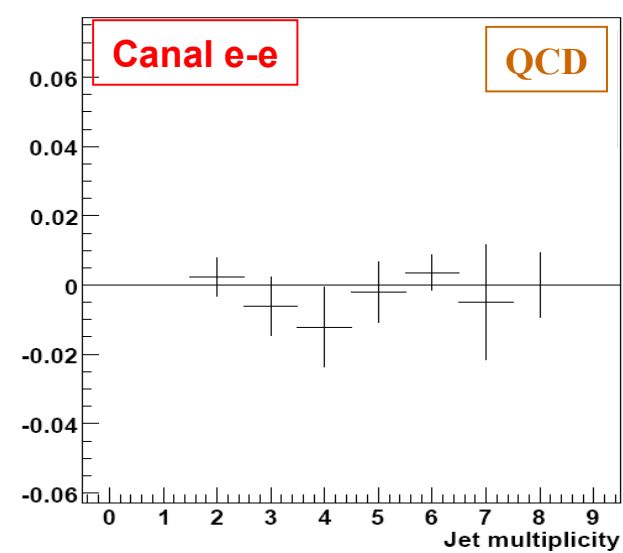
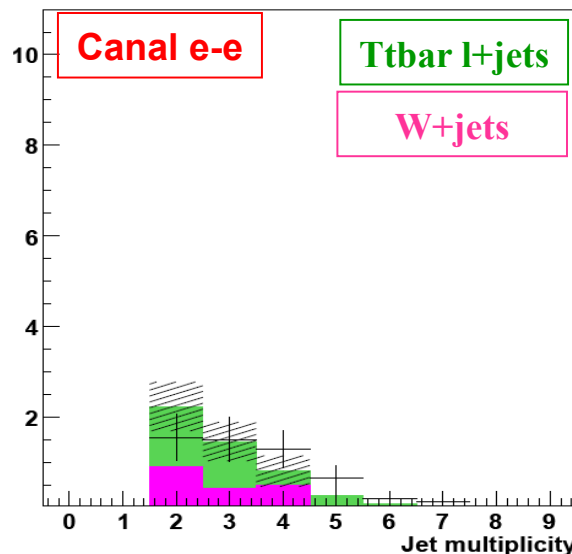
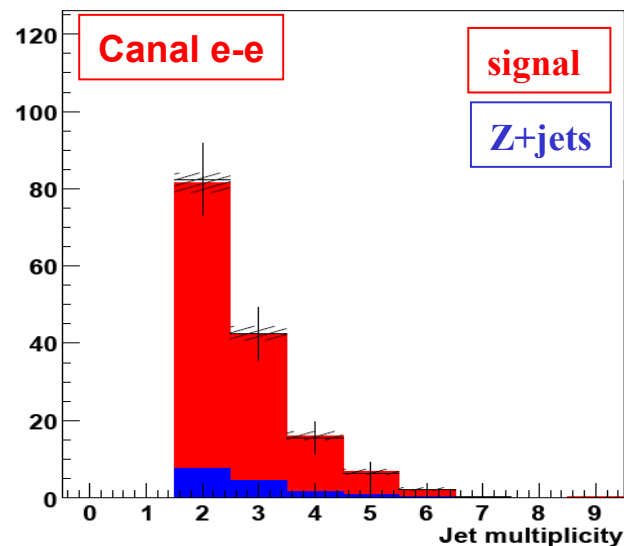
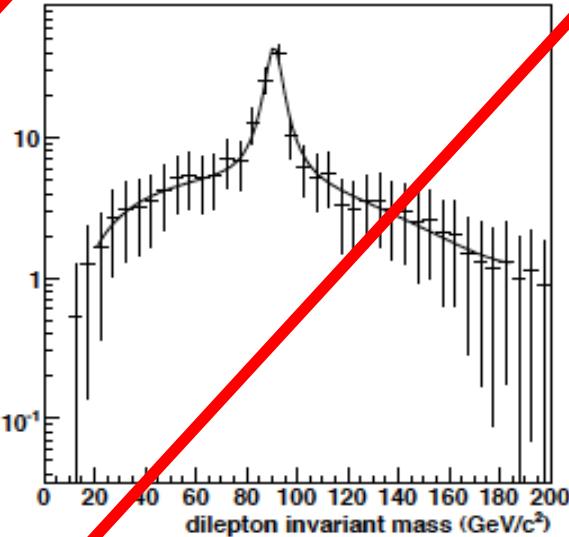
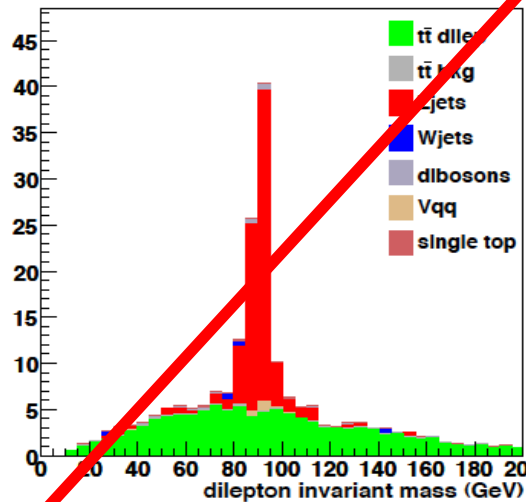


Table 5: Summary of statistical and systematic uncertainties in the template fits to the $\eta(\mu)$, M3 and M3' variables.

Source	Uncertainty [%]		
	Fit to $\eta(\mu)$	Fit to M3	Fit to M3'
Statistical Uncertainty (20 pb^{-1})	17.7	16.3	11.5
Jet Energy Scale	16.7	15.1	19
$t\bar{t}$ MC Generator	1.9	14.9	14
$t\bar{t}$ ISR/FSR	3.3	7.7	2
W+jets Factorization scale	4.4	4.7	4
W+jets Matching threshold	5.5	2.8	4
Single Top Shape	0.1	0.8	1
PDF Uncertainty	5.0	5.0	5.0
Total Systematic Error	19.2	23.8	25.0
Luminosity Error	10.0	10.0	10.0

dilepton Invariant mass for MET > 50 GeV



- Ajustement de la masse du Z avec Breit-Wigner+polynôme
- Les événements hors pique du Z sont comptés à basse MET.
- Rescalés au nombre d'événements sélectionnés à haute MET en utilisant les maximum de la BW et le maximum de la distribution a faible MET.

no b-tagging	Predicted	MC truth	statistical errors	scale factor	resolution	total
$e - e$	8.9	8.3	33.6%	13.2%	6.2%	36%
$\mu - \mu$	17.1	14.4	24.1%	4.9%	12.9%	28%