

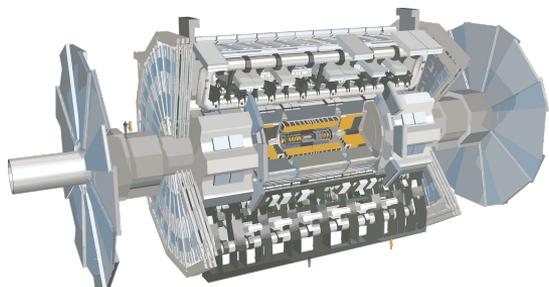
Analyse de données

Nicolas **Morange**, *IJCLab*



Présentation du cours

Détecteur
Cours de P. Schune



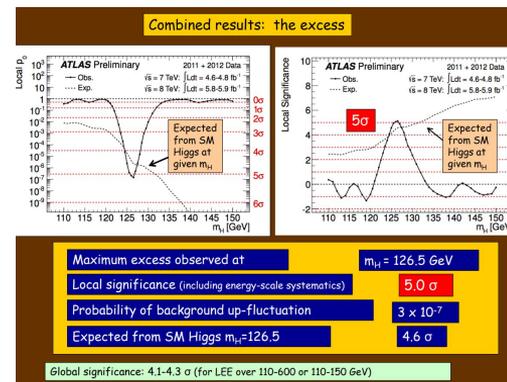
Analyse de données



Des données brutes
du détecteur à une
publication:

1. Acquisition des données
2. Principes de la reconstruction
3. Calcul
4. Simulation
5. Performance de la reconstruction
6. Mesure d'observables
7. Recherche d'une nouvelle particule

Résultats de physique des particules
Cours de E. Petit



L'essentiel des exemples de ce cours portent sur l'expérience ATLAS au LHC

- On parle mieux de ce qu'on connaît bien
- Et je sais où trouver les illustrations :-)

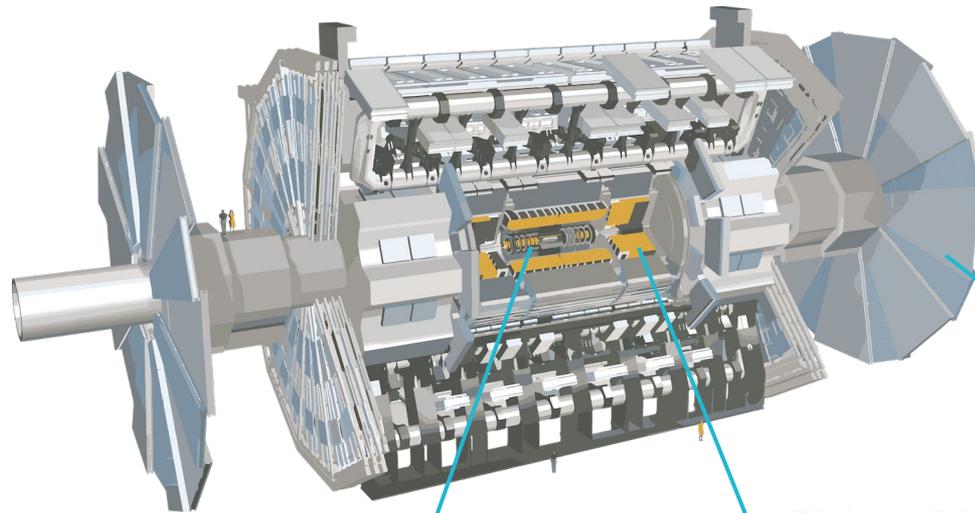
Les concepts abordés se transposent à la grande majorité des expériences de physique des particules

- À des **échelles** différentes

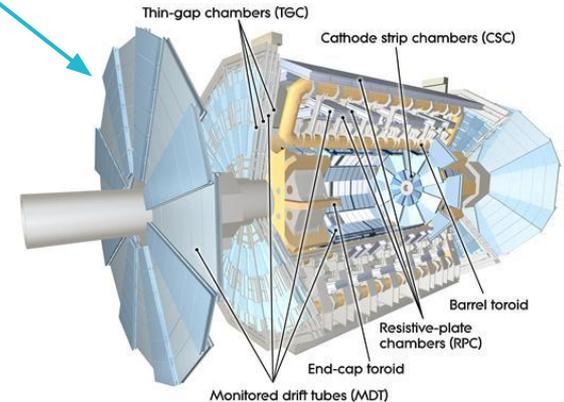
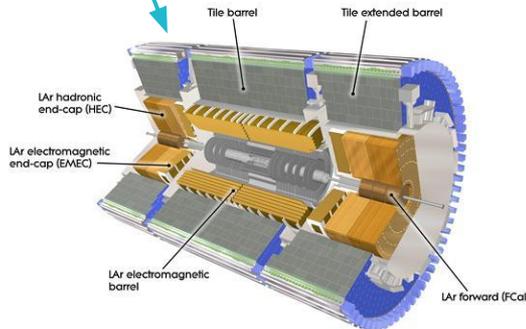
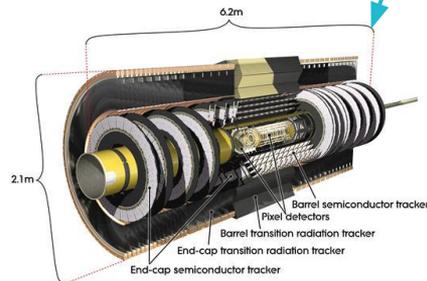
- Collisionneur p-p ou Pb-Pb
 - Mise en service en 2009
 - Programme jusqu'en >2035
- Énergie des collisions:
 - 7 TeV (2010 – 2011)
 - 8 TeV (2012)
 - 13 TeV (2015 – 2018)
 - 13.6 TeV (2022 –)
- Fréquence des collisions:
 - Croisements de faisceaux : ~40 MHz
 - Environ 40 collisions à chaque croisement de faisceau (empilement)
- 4 expériences principales
 - ATLAS, ALICE, CMS, LHCb



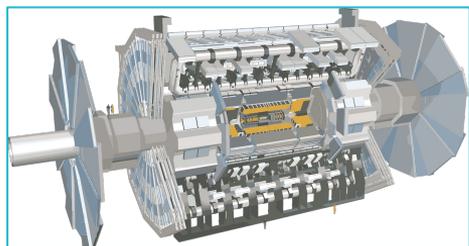
Le détecteur ATLAS



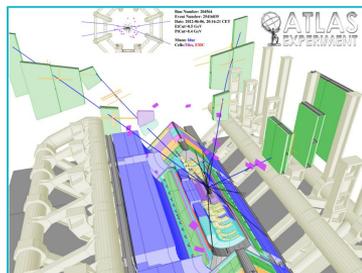
$L = 45\text{m}$, $D = 25\text{m}$
 $m = 7000\text{ T}$
3000 km de câbles
~100 millions de canaux



Cheminement des données



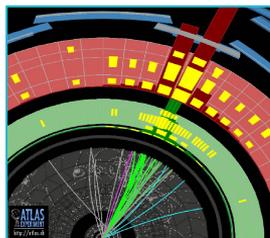
Détecteur



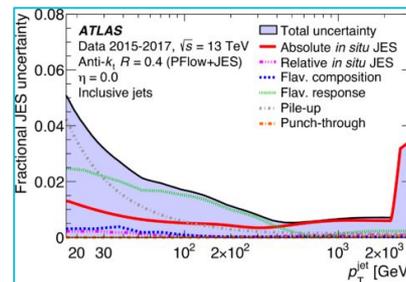
Simulation



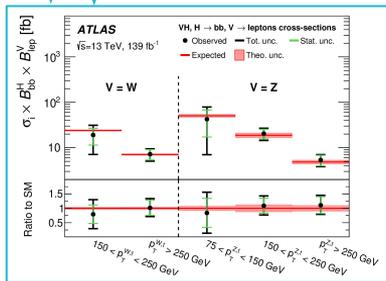
Trigger / Acquisition



Calibration
Qualité des données
Reconstruction



Performance de la reconstruction



Analyse de données

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector

ATLAS Collaboration*

Abstract: Measurement of the Standard Model Higgs boson decaying into a $b\bar{b}$ pair and produced in association with a W or Z boson decaying into leptons, using proton-proton collision data collected between 2015 and 2018 by the ATLAS detector, are presented. The measurements are combinations produced by the Large Hadron Collider at a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 139 fb⁻¹. The production of a Higgs boson in association with a W or Z boson is established with observed (expected) significance of 4.0 (3.1) and 3.5 (3.5) standard deviations, respectively. Cross-sections of associated production of a Higgs boson decaying into bottom quark pairs with an electron, muon, W or Z , decaying into leptons are measured as a function of the gauge boson transverse momentum in kinematic fiducial volumes. The cross-section measurements are all consistent with the Standard Model expectations, and the total uncertainties vary from 30% in the high gauge boson transverse momentum regions to 80% in the low regions. Limits are subsequently set on the parameters of effective Lagrangian operators in modulation of the WZ and ZZ processes as well as the Higgs boson decay into $b\bar{b}$.

1 Introduction

1.1 Motivation

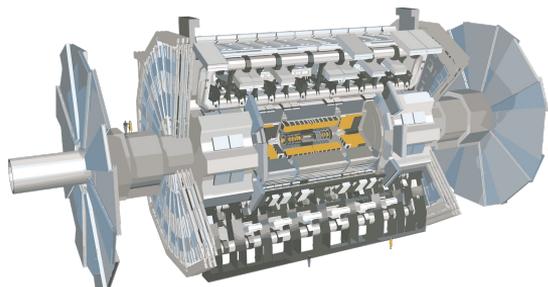
1.2 The Higgs boson [1–6] was discovered in 2012 by the ATLAS and CMS Collaborations [7, 8] with a mass of approximately 125 GeV from the analysis of proton-proton collisions produced by the Large Hadron Collider (LHC) [9]. Since then, the analysis of data collected at center-of-mass energies of 7 TeV, 8 TeV and 13 TeV in Run 1 and 2 of the LHC

7.1 Experimental uncertainties 11
7.2 Background uncertainties 11
7.3 Signal uncertainties 15
8 Statistical analysis 15
9 Results 16
9.1 Signal strength measurements 16
9.2 Cross-section measurements 18
10 Constraints on effective interactions 19
11 Conclusions 22
References 23

Publication

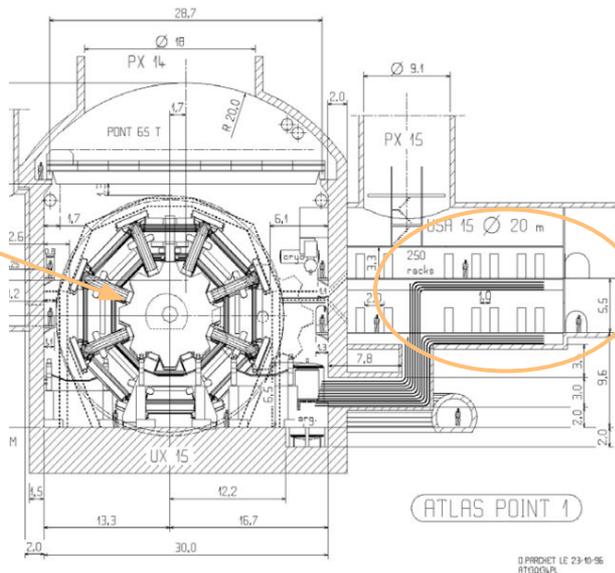
Acquisition des données

Acquisition des données



Détecteur

- Trajectographe: 100 millions de canaux
- Calorimètres: 200000 canaux avec précision 10-bit
- Détecteurs de muons: 1.5 millions de canaux



Caverne ATLAS



Système d'acquisition

1 collision:
~1.5 Mo

Flux de données:
60 To/s
Impossible de tout enregistrer !

0x00000015	0x20000e3f	536874559	lv1 trigger info[0]
0x00000016	0x100000c0	268435648	lv1 trigger info[1]
0x00000017	0x8000043f	2147484735	lv1 trigger info[2]
0x00000018	0x00021007	135175	lv1 trigger info[3]
0x00000019	0x00000e10	3600	lv1 trigger info[4]
0x0000001a	0x00080000	524288	lv1 trigger info[5]
0x0000001b	0x02c00400	46138368	lv1 trigger info[6]
0x0000001c	0x00020001	131073	lv1 trigger info[7]
0x0000001d	0x00000816	2070	lv1 trigger info[8]
0x0000001e	0x100000c0	268435648	lv1 trigger info[9]
0x0000001f	0x80000018	2147483672	lv1 trigger info[10]
0x00000020	0x00021001	135169	lv1 trigger info[11]
0x00000021	0x00000e10	3600	lv1 trigger info[12]
0x00000022	0x00000000	0	lv1 trigger info[13]
0x00000023	0x02c00400	46138368	lv1 trigger info[14]
0x00000024	0x00020000	131072	lv1 trigger info[15]
0x00000025	0x00000010	16	lv1 trigger info[16]
0x00000026	0x00000000	0	lv1 trigger info[17]
0x00000027	0x00000008	8	lv1 trigger info[18]
0x00000028	0x00000000	0	lv1 trigger info[19]
0x00000029	0x00000810	2064	lv1 trigger info[20]
0x0000002a	0x00000000	0	lv1 trigger info[21]
0x0000002b	0x00000400	1024	lv1 trigger info[22]
0x0000002c	0x00000000	0	lv1 trigger info[23]

Taux des processus physiques

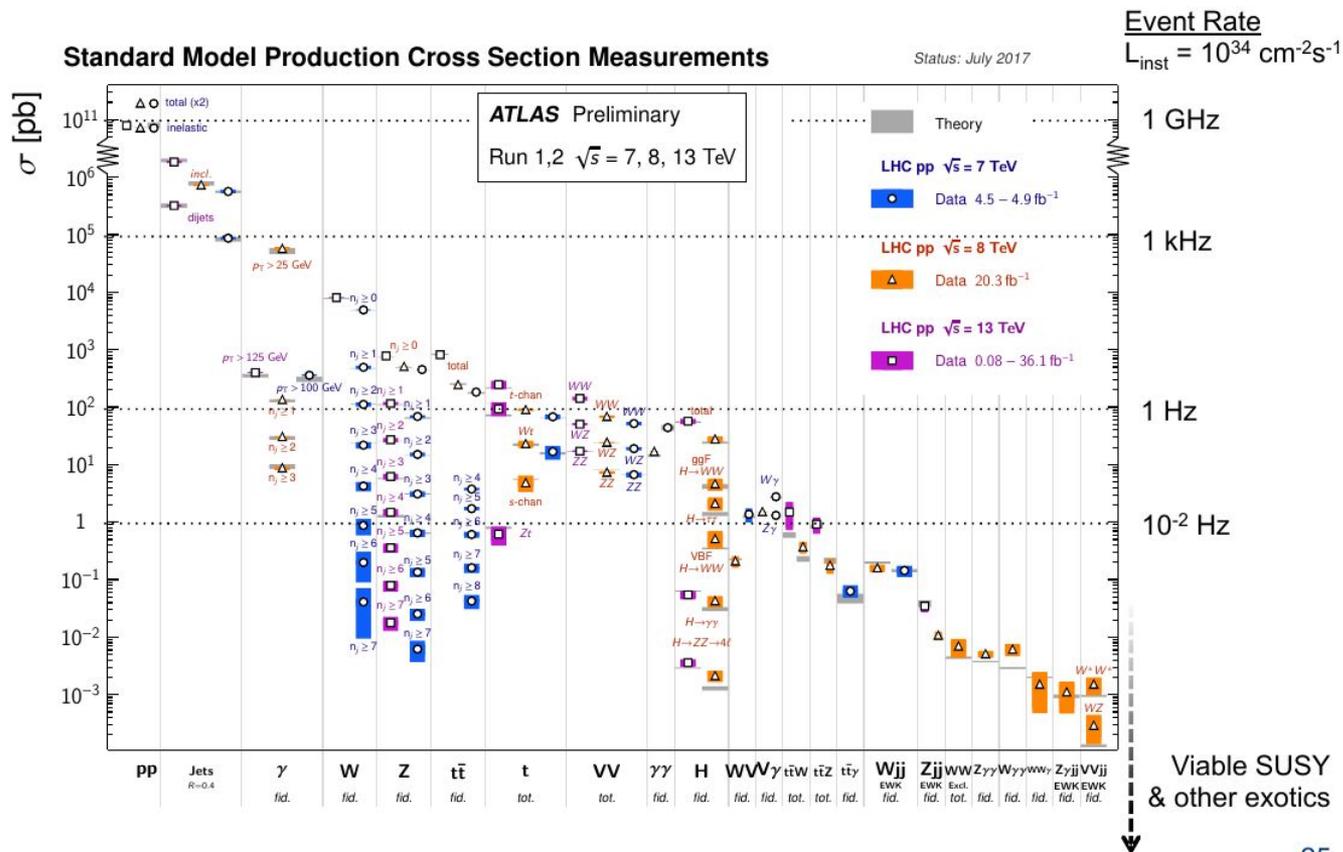
Pour un processus donné:

$$N = \sigma \cdot \mathcal{L}$$

section efficace luminosité

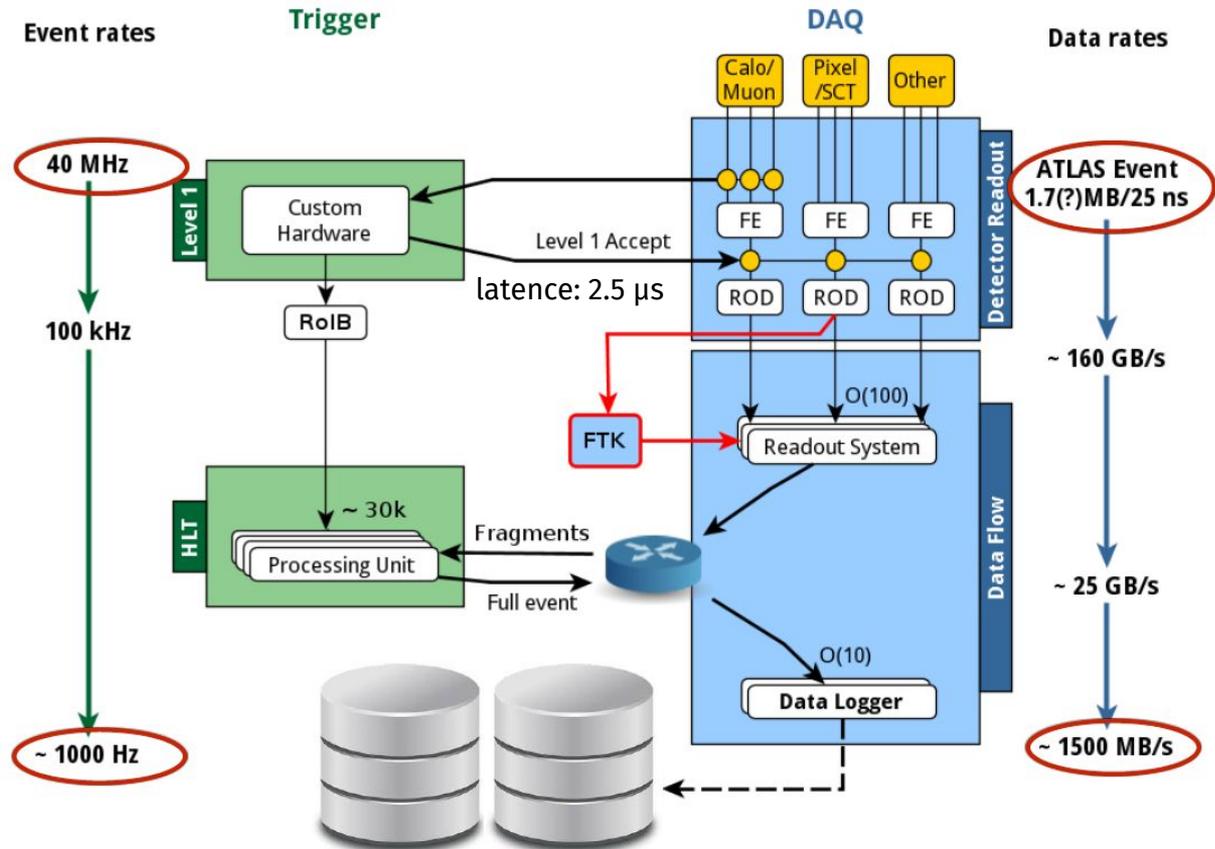
Processus les plus intéressants aussi les plus rares

=> possibilité d'un **tri en temps réel** pour ne garder que les collisions les plus intéressantes

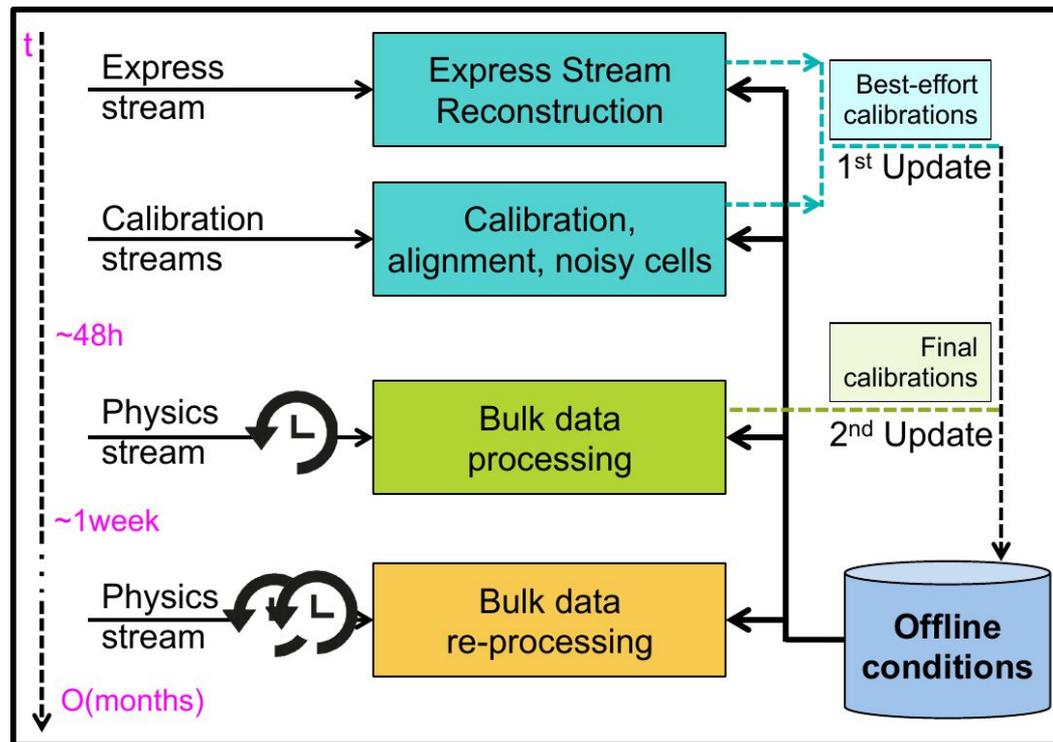


Système de déclenchement et acquisition

- **Système de déclenchement**
reconstruction simplifiée en temps réel
- **Objectif**
sélection sur la base de **critères simples**:
"électron $E > 25$ GeV" ||
"muon $p_T > 23$ GeV" ||
"électron $E > 7$ GeV && muon $p_T > 7$ GeV" ||
...

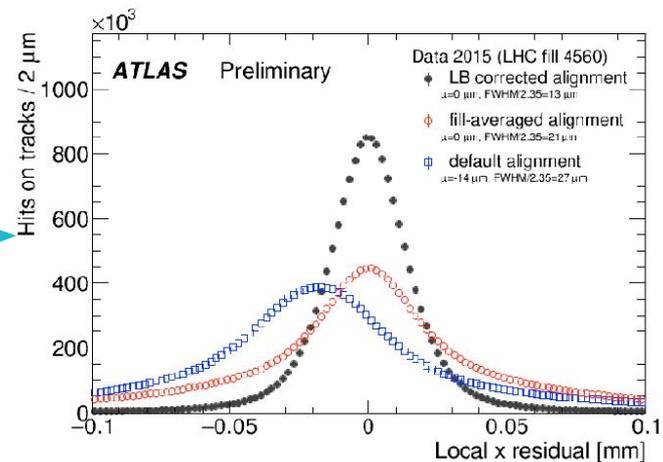
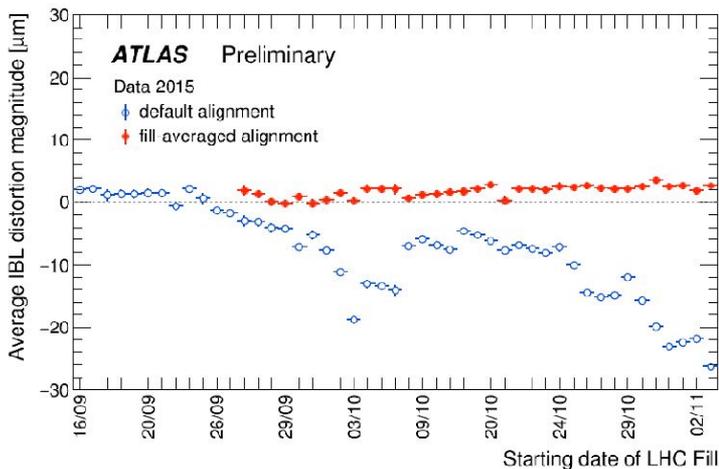


Calibration des données



- La reconstruction des données ne peut pas être refaite souvent
- Utilisation de jeux de données spécifiques pour **calibrer** les données avant reconstruction
- Nécessaire pour atteindre la performance désirée du détecteur

Calibration des données: alignement



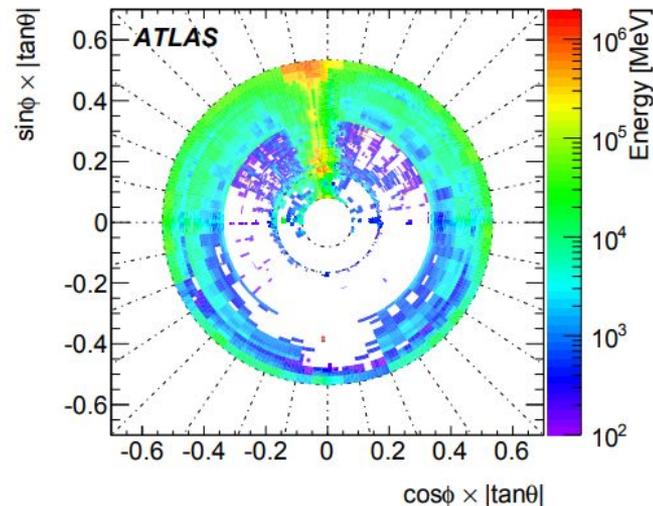
- En particulier: **alignement des détecteurs** de traces
- Echauffement en cours de prise de données
=> dilatation
=> torsions et déformations
- Prise en compte nécessaire pour une bonne résolution sur les traces

- Objectif: données de **bonne qualité** pour les analyses de physique
 - On ne veut pas “découvrir” une nouvelle particule à cause d’un bruit dans des capteurs !
- **Masquage** des données problématiques
 - Données corrompues
 - Bruits anormaux
 - Événements transitoires
- >95 % de bonnes données

ATLAS pp Run-2: July 2015 – October 2018

Inner Tracker			Calorimeters		Muon Spectrometer				Magnets	
Pixel	SCT	TRT	LAr	Tile	MDT	RPC	CSC	TGC	Solenoid	Toroid
99.5	99.9	99.7	99.6	99.7	99.8	99.6	100	100	99.8	98.8

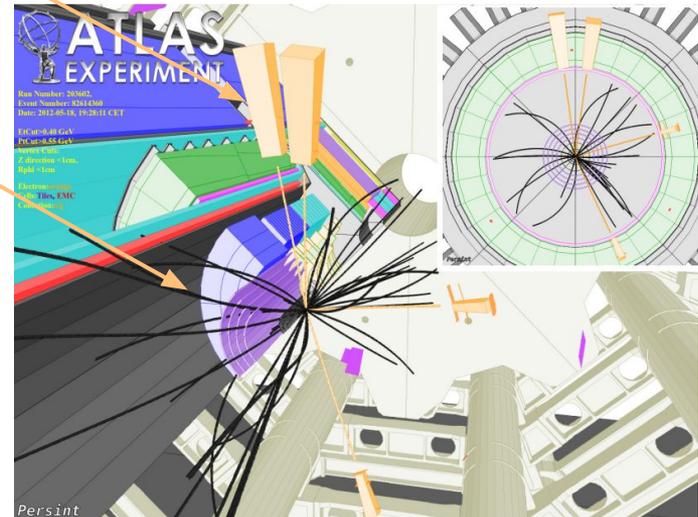
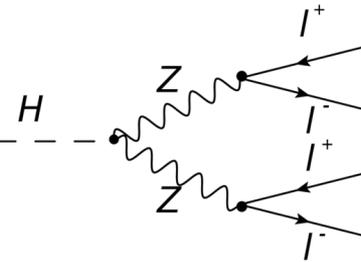
Good for physics: 95.6% (139 fb⁻¹)



Reconstruction

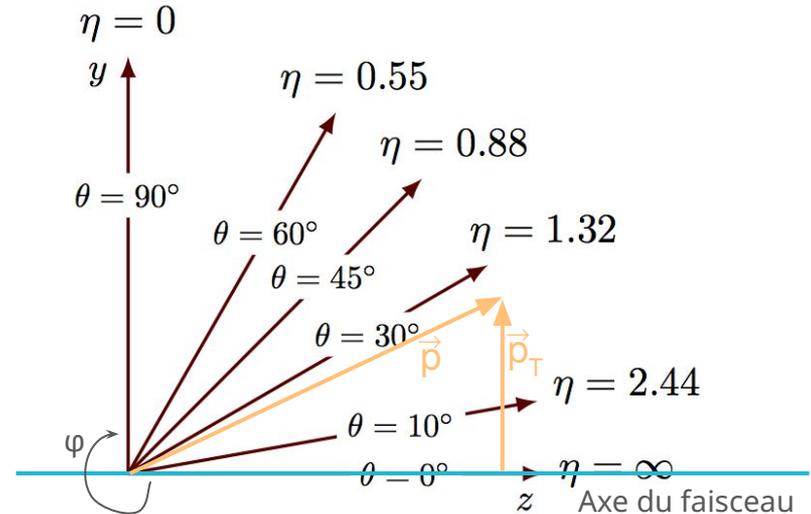
Reconstruction ?

- Ce qu'on veut connaître: **processus microscopique**
- Ce qu'on mesure: hits / dépôts d'énergie des particules **de l'état final** dans le détecteur
- **Reconstruction:**
 - Identification
 - Mesure de la quadri-impulsion
 - Pour toutes les particules **qui arrivent dans le détecteur**: nos "objets".
- Étape objets → processus microscopique: **"analyse de physique"**



- Collisionneur p-p: grandeurs dans le plan transverse
 - Impulsion transverse $p_T = p \sin \theta$
- Pseudo-rapidité
 - $\eta = -\ln(\tan(\theta/2))$
 - Processus physiques en η plutôt qu'en θ
- Relativité restreinte
 - Conservation de la quadri-impulsion
 - Masse invariante d'une collection de particules:

$$m = \sqrt{(\sum_i E_i^2 - \|\sum_i \vec{p}_i\|^2)}$$



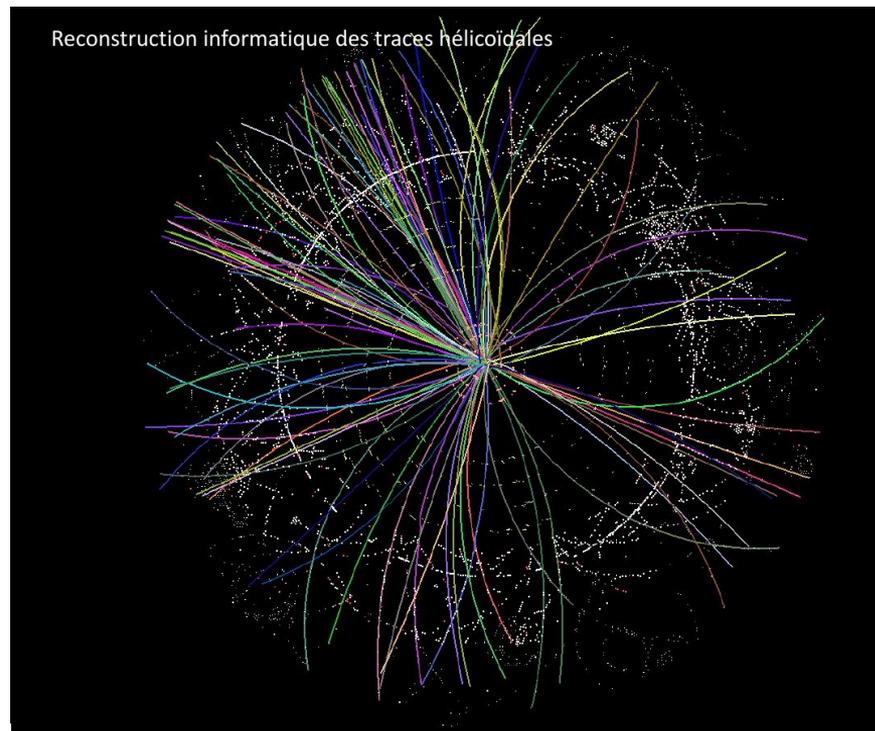
Reconstruction des traces chargées

- **Trajectomètres**

- Points de passages des particules chargées
- Détecteurs silicium ou gazeux

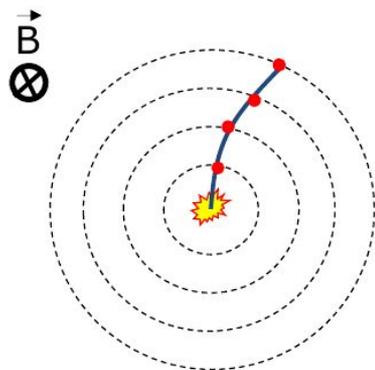
- **Reconstruction: jouer aux points à relier...**

- En tenant compte du champ B
- Des incertitudes de mesure
- Des incertitudes d'assignation des hits à une trace donnée
- De la confusion des hits
- De la diffusion multiple
- Du bruit
- Des inefficacités

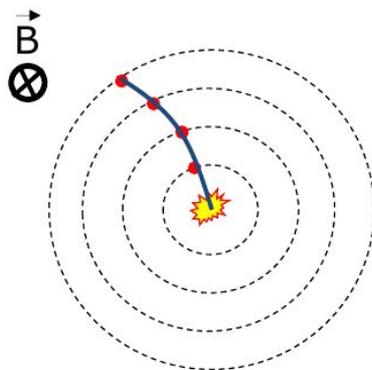


Champ magnétique dans la reconstruction des traces

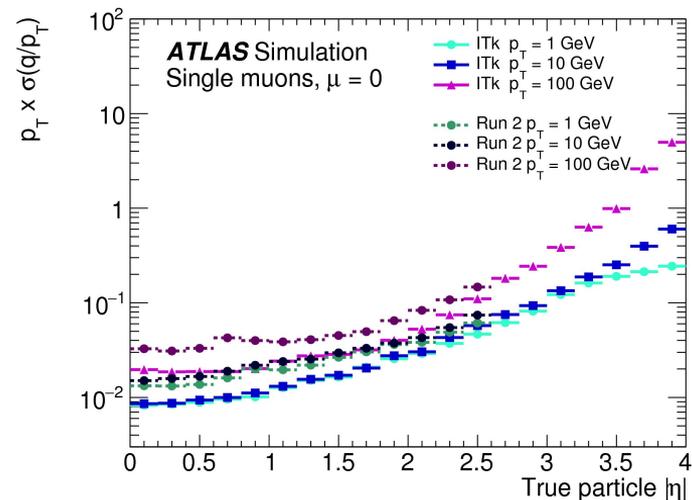
- Champ magnétique de solénoïde
 - Courbure dans le plan transverse
 - $\rho = qRB$
 - La courbure donne **la charge et l'impulsion** !
 - La mesure de R doit être la plus précise possible
 - Meilleure précision à petite impulsion (R grand)



Électron

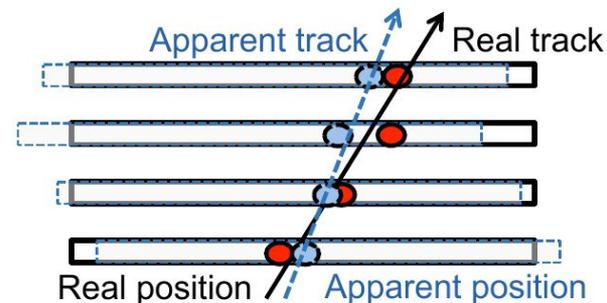
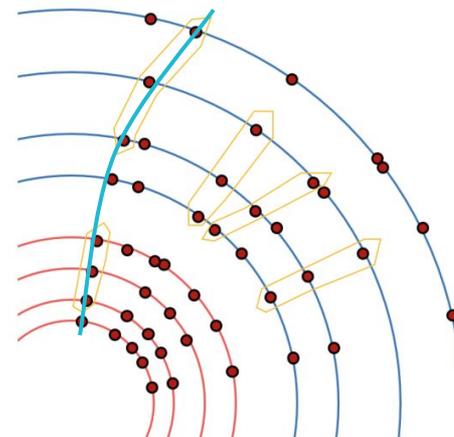


Positron



Algorithmes de reconstruction

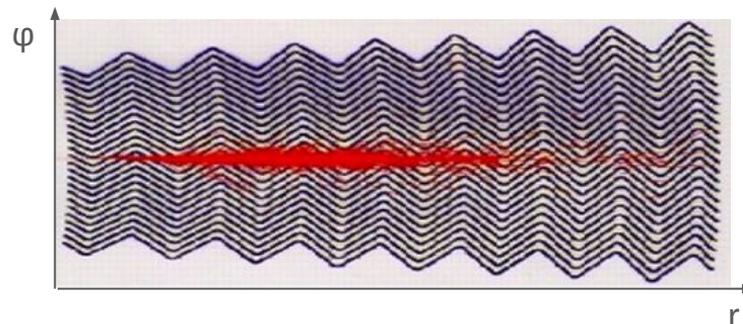
- **Algorithmes de reconstruction**
 - Difficulté: combinatoire
 - Domine le temps total de reconstruction
 - Localisation de graines
 - Ajustement: filtre de Kalman
 - Détermine paramètres des traces et incertitudes
 - Reconstruction des vertex
- **Incertitudes**
 - Incertitudes de mesures, méconnaissance du champ B, désalignements...
 - Se propagent à la mesure de l'impulsion et des paramètres d'impact



Reconstruction des amas d'énergie

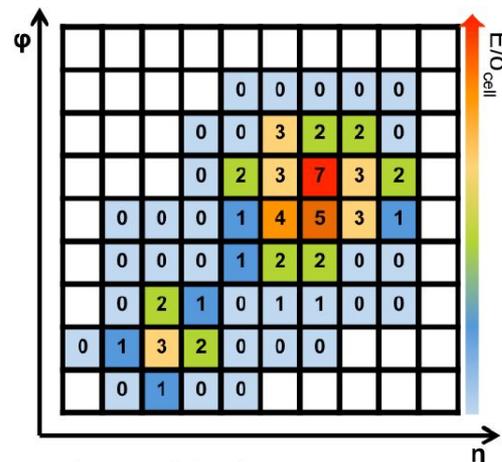
- **Calorimètres**

- Arrêtent les particules, qui y déposent leur énergie en **gerbes**
- Gerbes électromagnétiques / hadroniques
- Cellules: segmentation longitudinale / transverse



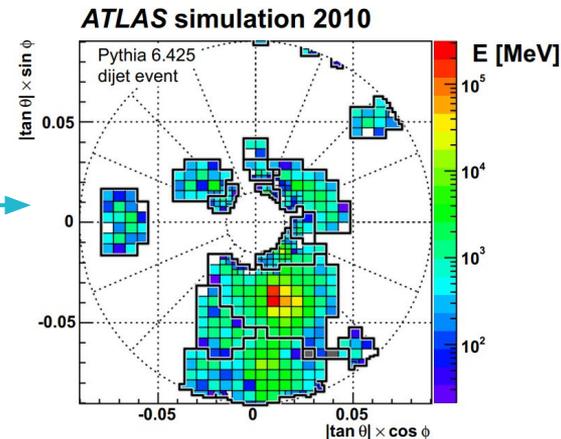
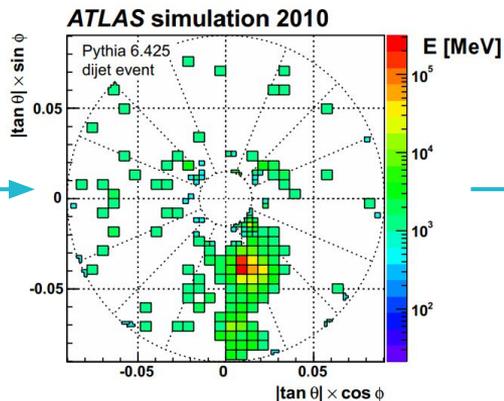
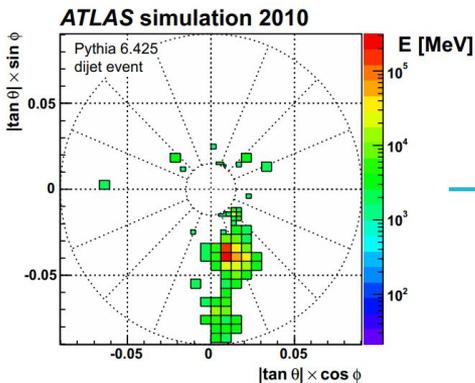
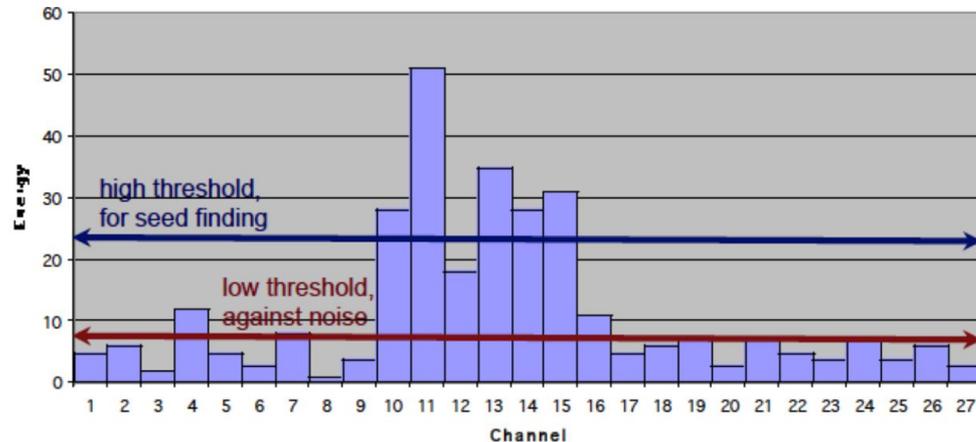
- **Agrégation ("clustering")**

- Regrouper les cellules appartenant à une même gerbe
- La somme de leurs énergies donne une **mesure de l'énergie** de la particule initiale
- Difficultés: bruit du calorimètre, confusion entre particules proches



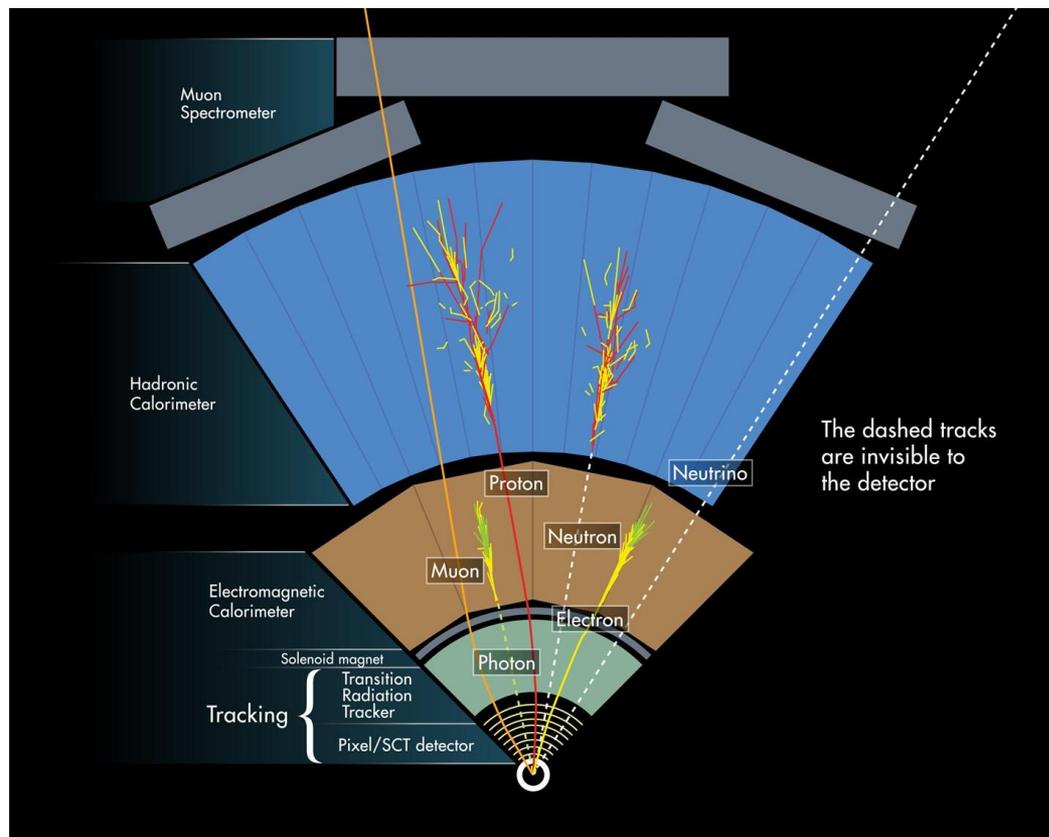
Ambiguïtés du clustering

- Confusion entre amas proches
 - 1 ou 2 amas sur l'exemple ?
 - Seuil haut pour trouver des graines
- Problème du bruit
 - Nécessité d'un seuil bas pour ne pas polluer la mesure d'énergie

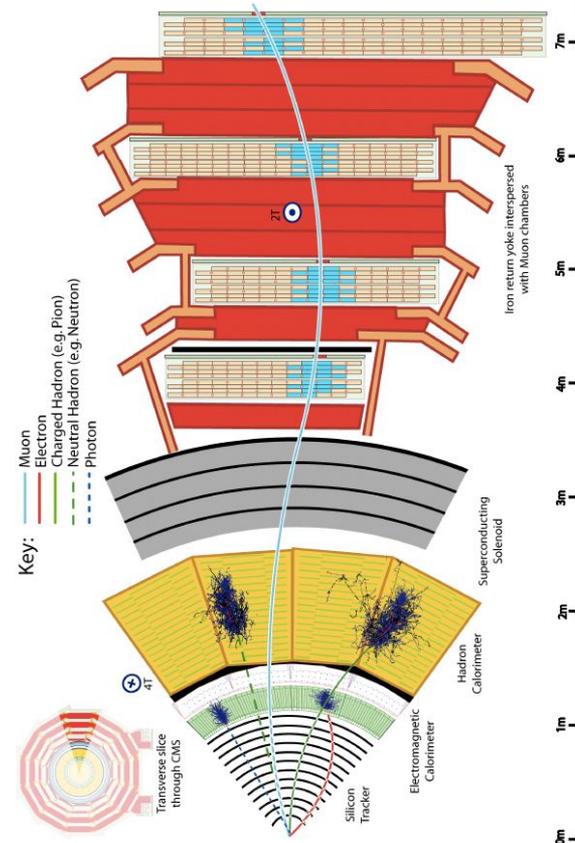


Reconstruction des particules dans ATLAS

- Cours de détecteurs
 - Comment chaque type de particule interagit dans la matière
- **Reconstruction**: processus inverse !
 - Détermination du **type de particule** et **mesure de sa quadri-impulsion** à partir des informations des traces et des amas

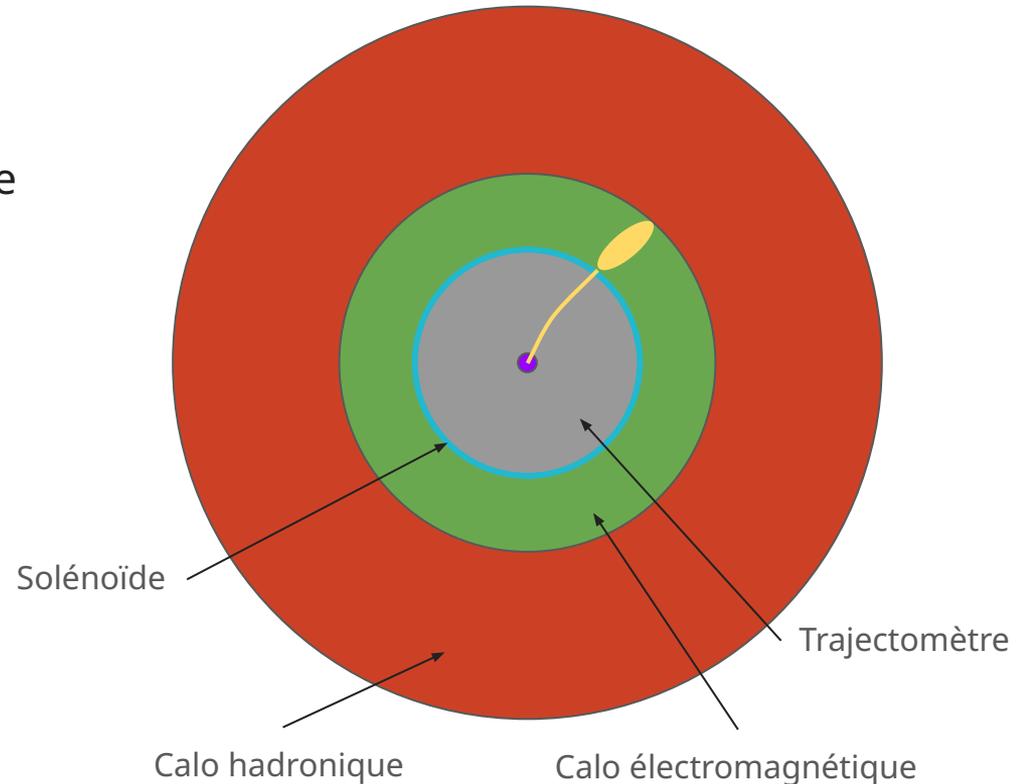
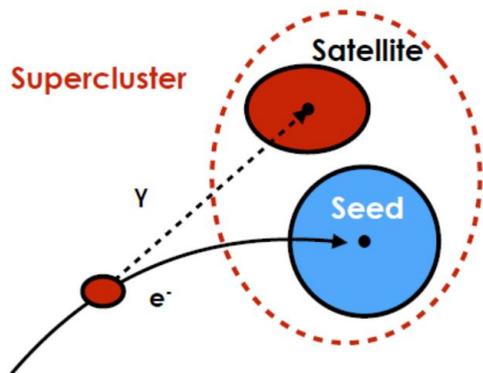


- Le plus facile !
 - $\tau_\mu = 2.2 \mu\text{s}$ donc $c\tau_\mu = 650\text{m}$: muons stables
 - **Traces dans le détecteur interne et dans les chambres à muons**
 - Faible dépôt d'énergie dans les calorimètres
 - Excellente résolution en énergie

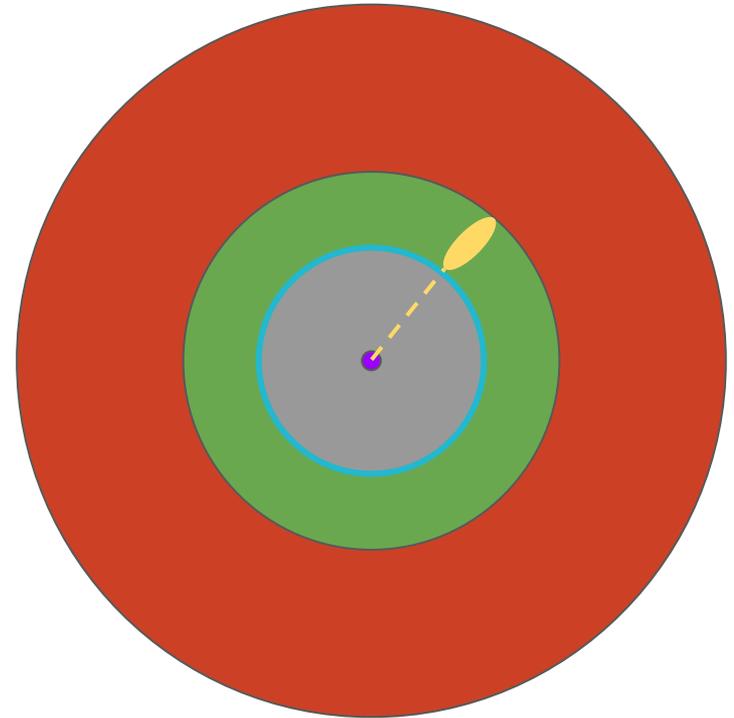
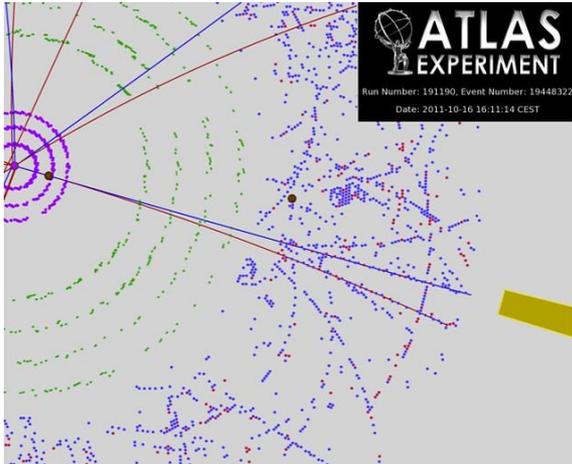


Electrons

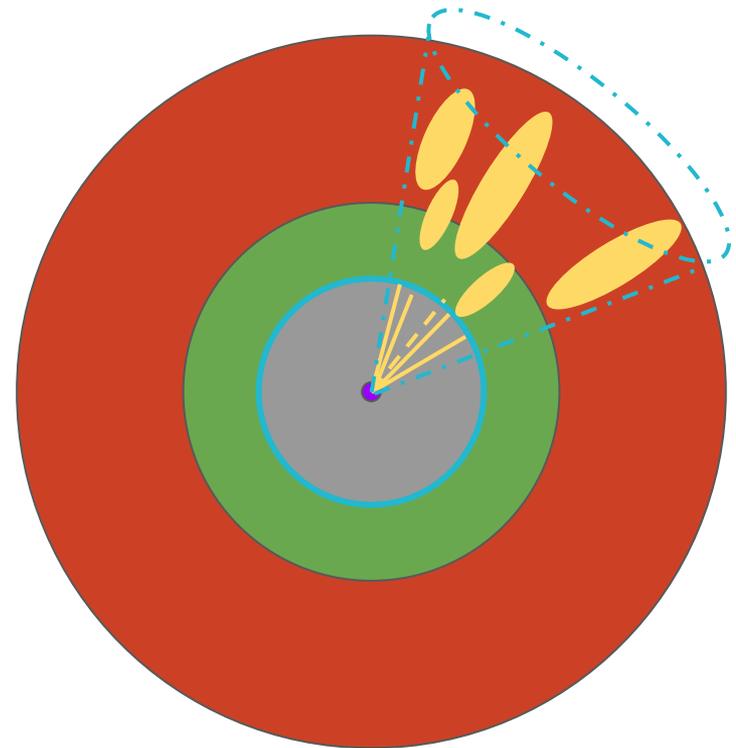
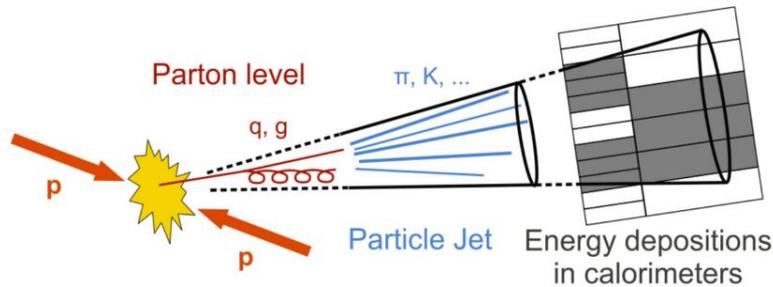
- Simple en théorie
 - Trace chargée + amas électromagnétique
- Un peu plus compliqué en pratique
 - Bremsstrahlung !
 - Reconstruction de trace adaptée
 - Dépôt d'énergie à côté de l'amas principal: "supercluster"



- **Un amas électromagnétique**
 - Pas de trace chargée associée
 - Très bonne résolution
- **Sauf...**
 - Conversions $\gamma \rightarrow e^+e^-$
 - Deux traces venant d'un vertex décalé



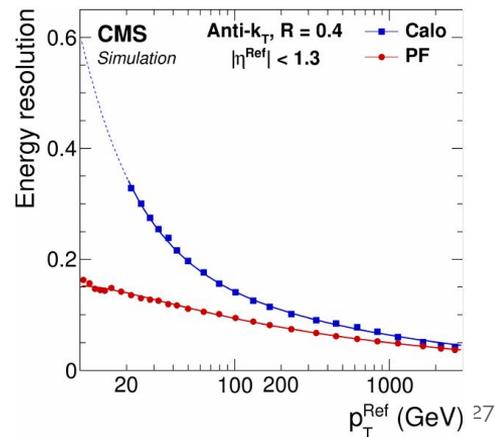
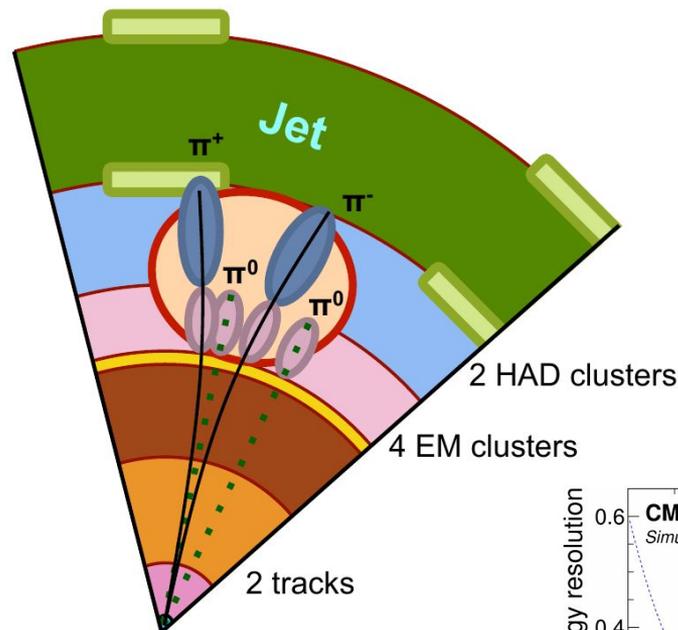
- **Jet:** manifestation d'un quark ou gluon de haute énergie
 - Multiples photons et hadrons collimatés, chargés et neutres
 - Conservation de la quadri-impulsion



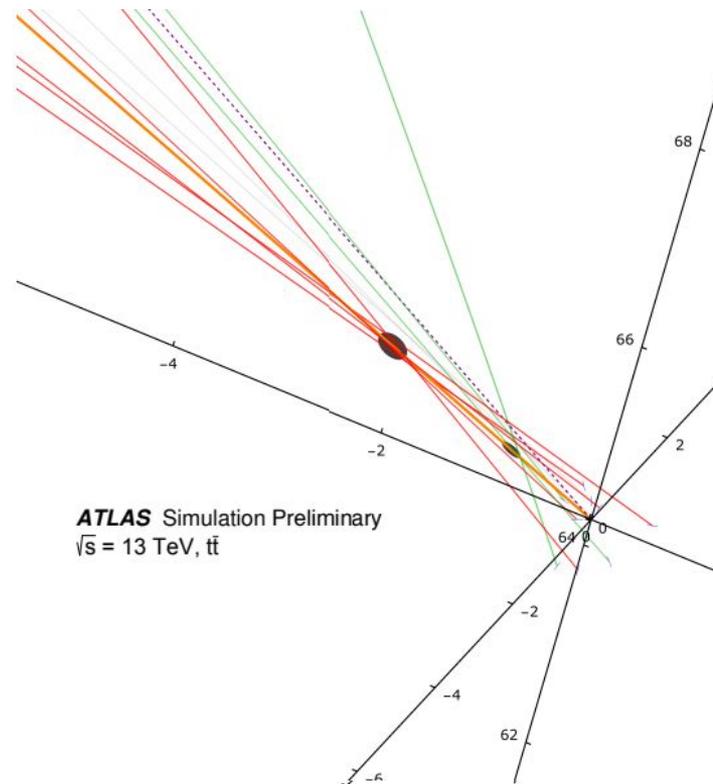
- Dans le détecteur
 - **Multiples traces et amas**
 - Rassemblés dans un cône
 - Algorithmes de formation de jets

L'algorithme de "Particle Flow"

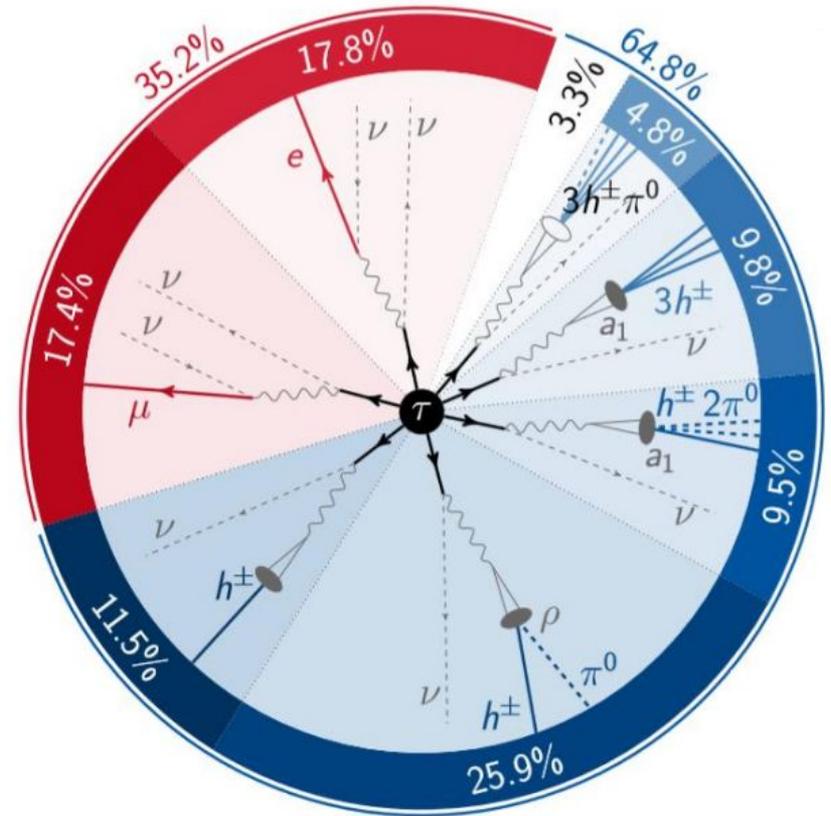
- Mesure d'énergie des hadrons dans le calorimètre pas très précise à basse énergie
- Pour les hadrons chargés, possibilité d'utiliser la mesure de la trace
 - Puis soustraction de l'énergie de l'amas
 - Difficulté: confusion entre amas au sein d'un jet
- Très efficace dans CMS
 - Moins nécessaire dans ATLAS



- **Temps de vie des hadrons b: ~ 1 ps**
 - $\gamma_{ct} \sim 2$ mm - 3 cm
 - Désintégration en hadron c et autres particules
 - Hadron c vole également avant de se désintégrer
- **Dans le détecteur**
 - Un jet reconstruit comme les autres
 - Mais dans le trajectomètre:
 - **Vertex secondaires**
 - **Paramètre d'impact des traces**

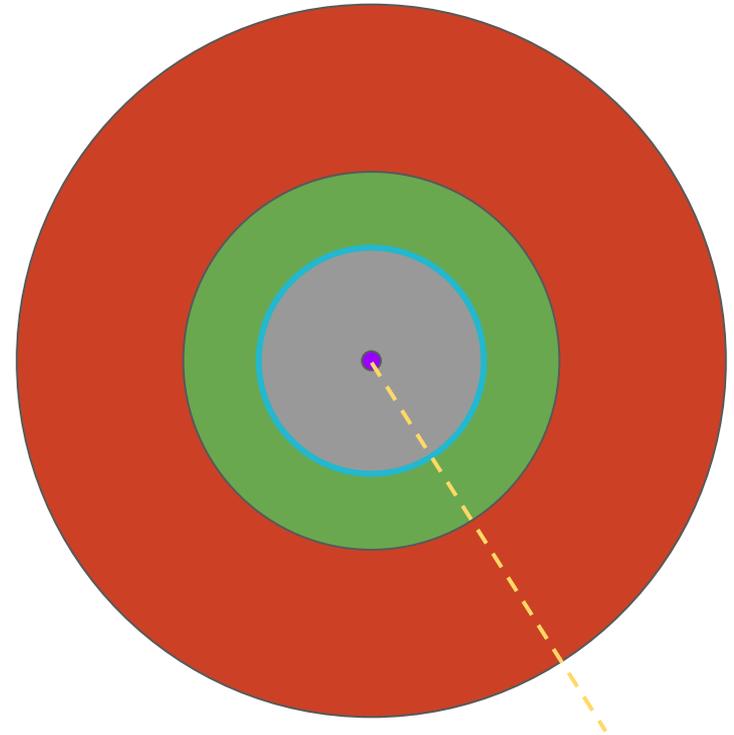


- **Temps de vie: $c\tau_\tau = 87 \mu\text{m}$**
 - Désintégration dans le tube à vide
- **Nombreux modes de désintégration**
 - Modes leptoniques: électron ou muon de plus basse énergie + neutrino
 - Modes hadroniques: jet étroit, faible multiplicité de hadrons (1 ou 3)

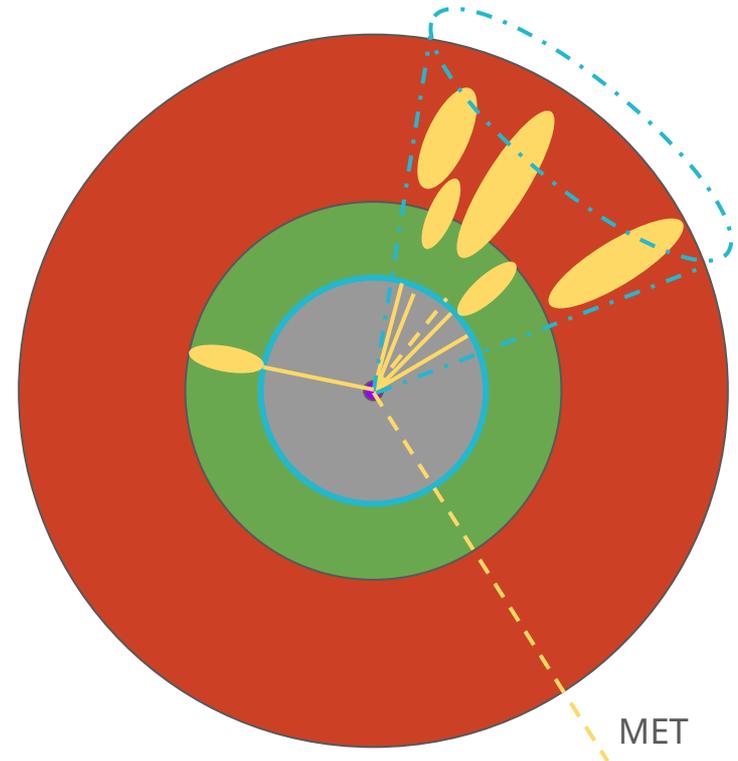


Neutrinos

- Les neutrinos n'interagissent pas dans le détecteur

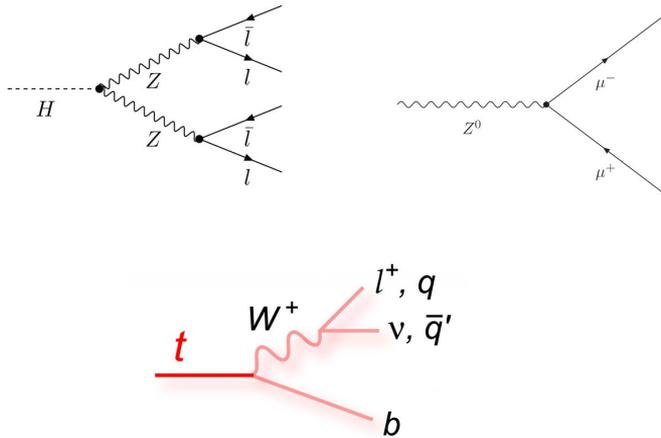


- Les neutrinos n'interagissent pas dans le détecteur
- Mais: conservation du mouvement dans le plan transverse
 - $\sum_i \vec{p}_{T,i} = 0 = \sum_{\text{visible}} \vec{p}_T + \sum_{\text{invisible}} \vec{p}_T$
 - On définit l'**énergie transverse manquante**:
$$\text{MET} = - \sum_{\text{visible}} \vec{p}_T$$
 - Mesure la somme des particules indétectables de l'événement
 - Neutrinos
 - **Possibles nouvelles particules**

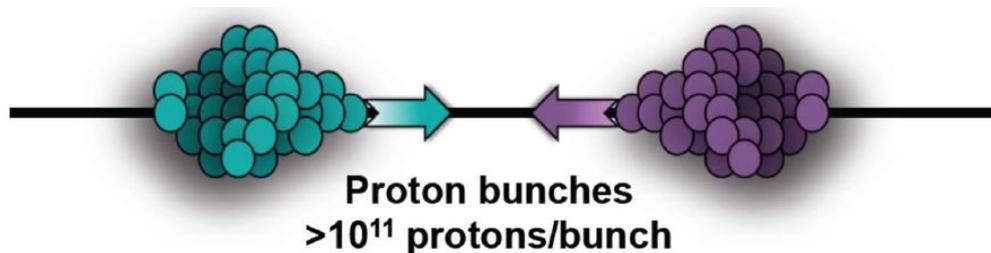


Et les autres ?

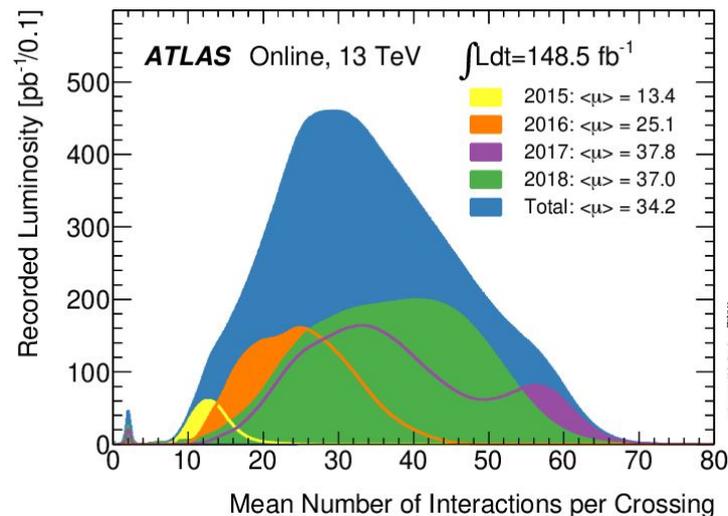
- Quark top, bosons W, Z, H
 - Temps de vie extrêmement bref
 - Se **désintègrent** en d'autres particules
 - Doivent être **reconstruites** à partir de leurs désintégrations

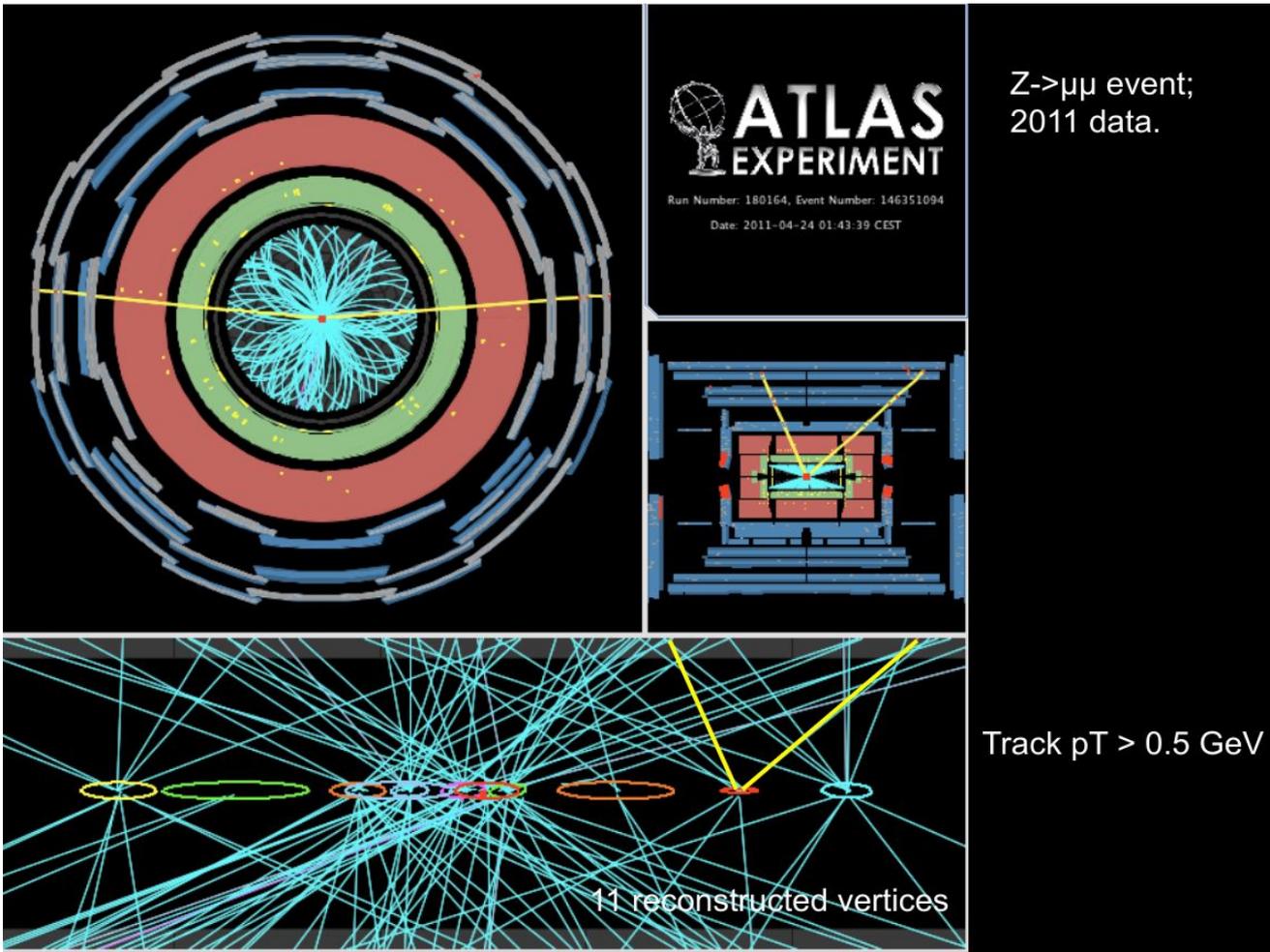


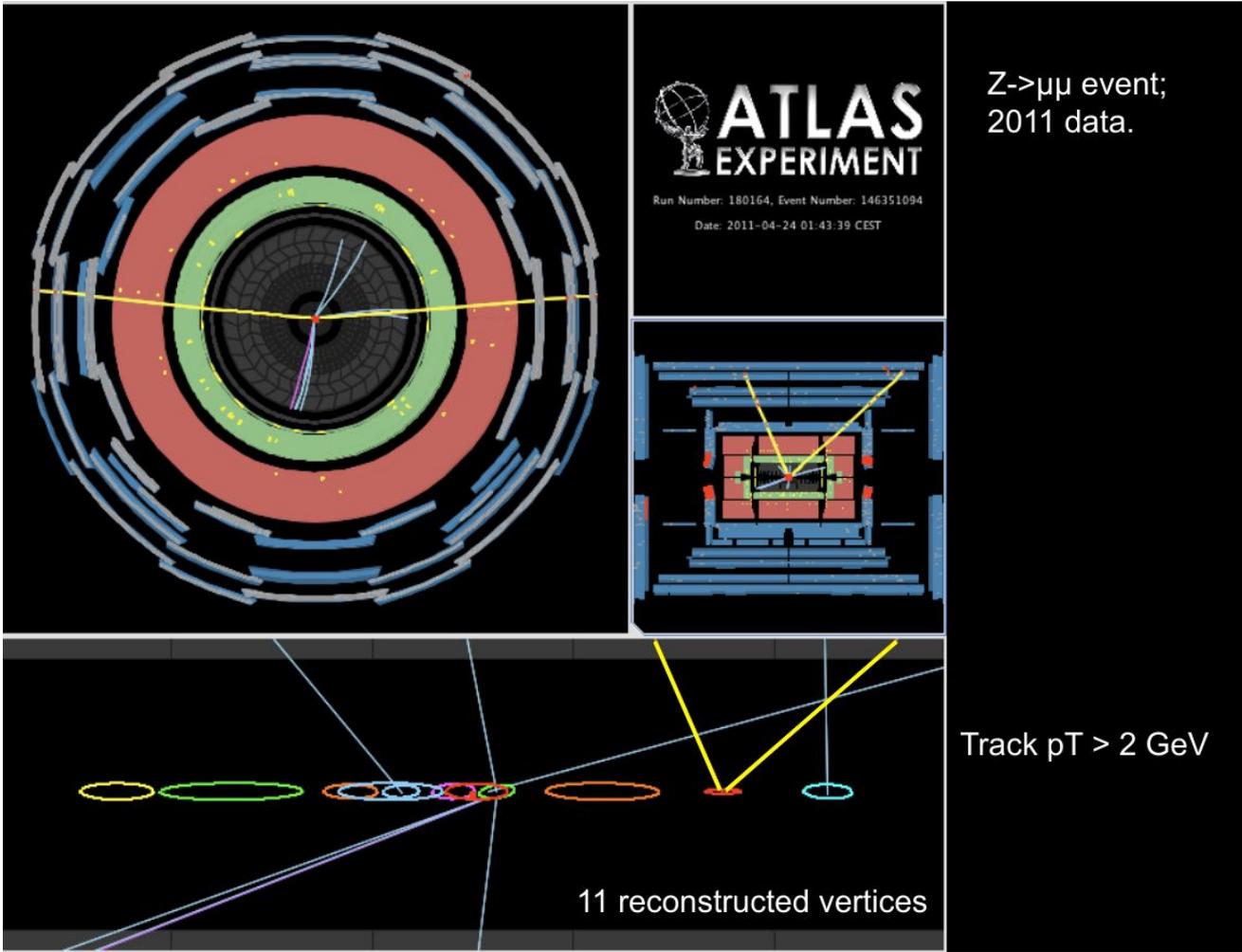
	masse → $\approx 2.3 \text{ MeV}/c^2$	$\approx 1.275 \text{ GeV}/c^2$	$\approx 173.07 \text{ GeV}/c^2$	0	$\approx 126 \text{ GeV}/c^2$
charge →	2/3	2/3	2/3	0	0
spin →	1/2	1/2	1/2	1	0
	u up	c charm	t top	g gluon	H boson de Higgs
QUARKS	$\approx 4.8 \text{ MeV}/c^2$	$\approx 95 \text{ MeV}/c^2$	$\approx 4.18 \text{ GeV}/c^2$	0	
	-1/3	-1/3	-1/3	0	
	1/2	1/2	1/2	1	
	d down	s strange	b bottom	γ photon	
	$0.511 \text{ MeV}/c^2$	$105.7 \text{ MeV}/c^2$	$1.777 \text{ GeV}/c^2$	$91.2 \text{ GeV}/c^2$	
	-1	-1	-1	0	
	1/2	1/2	1/2	1	
	e électron	μ muon	τ tau	Z⁰ boson Z ⁰	
LEPTONS	$< 2.2 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 15.5 \text{ MeV}/c^2$	$80.4 \text{ GeV}/c^2$	
	0	0	0	± 1	
	1/2	1/2	1/2	1	
	ν_e neutrino électronique	ν_μ neutrino muonique	ν_τ neutrino tauique	W[±] boson W [±]	
					BOSONS DE JAUGE

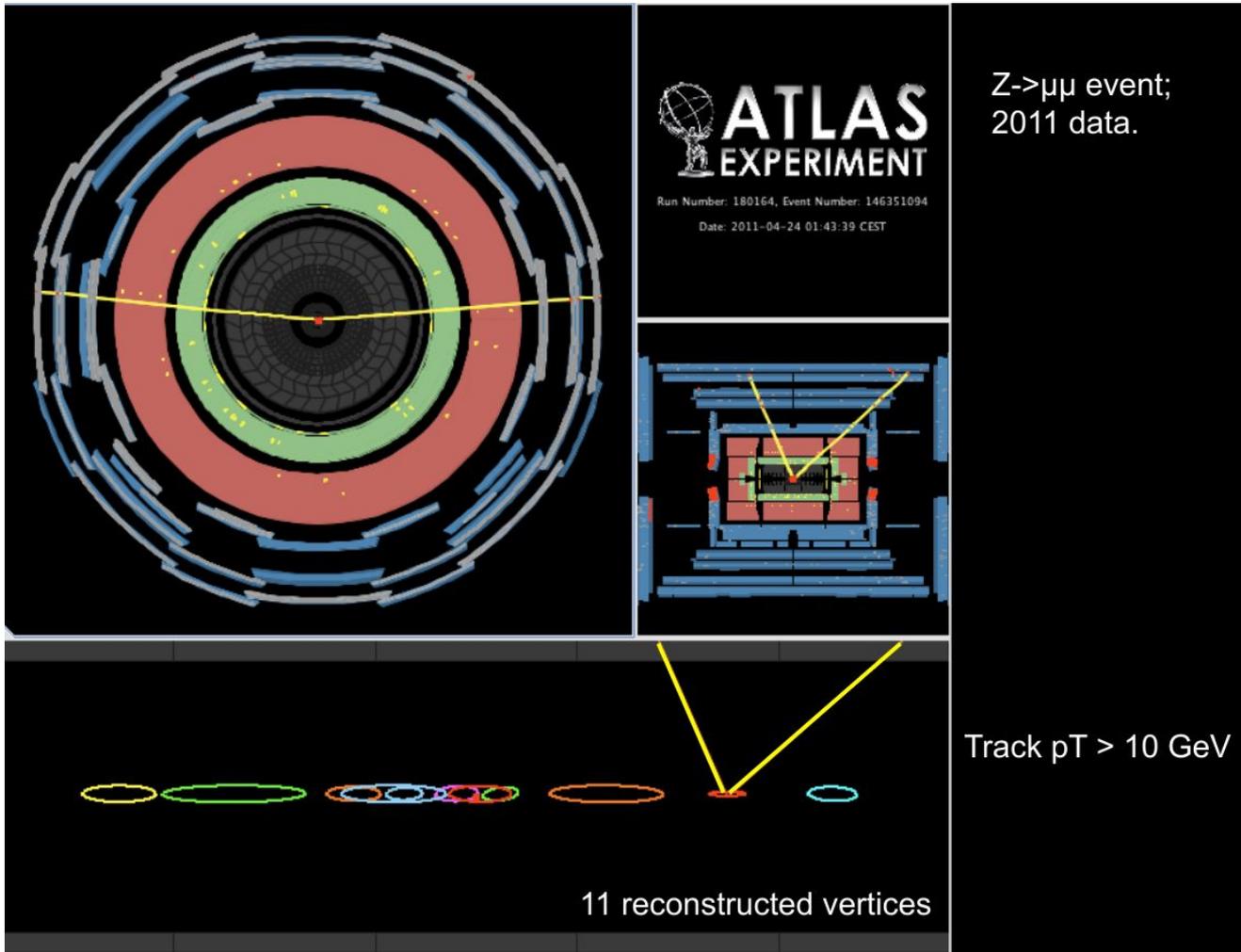


- Jusqu'à 60 collisions par croisement de faisceau
- Au plus une collision "intéressante"
 - Les autres sont de **l'empilement**
 - Interactions de basse énergie, particules de faible impulsion



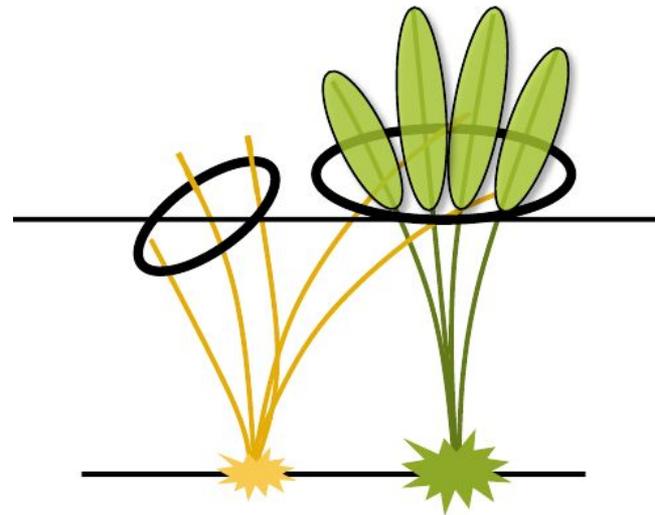






Conséquences de l'empilement

- Complexité combinatoire et confusion pour la reconstruction des traces
 - Problème principalement pour le temps CPU
- Confusion dans la reconstruction des amas dans les calorimètres
- Conséquences **principalement pour les jets**
 - Jets additionnels
 - Bruit additionnel dans la mesure d'énergie
 - Techniques spécifiques de calibration pour limiter ses effets



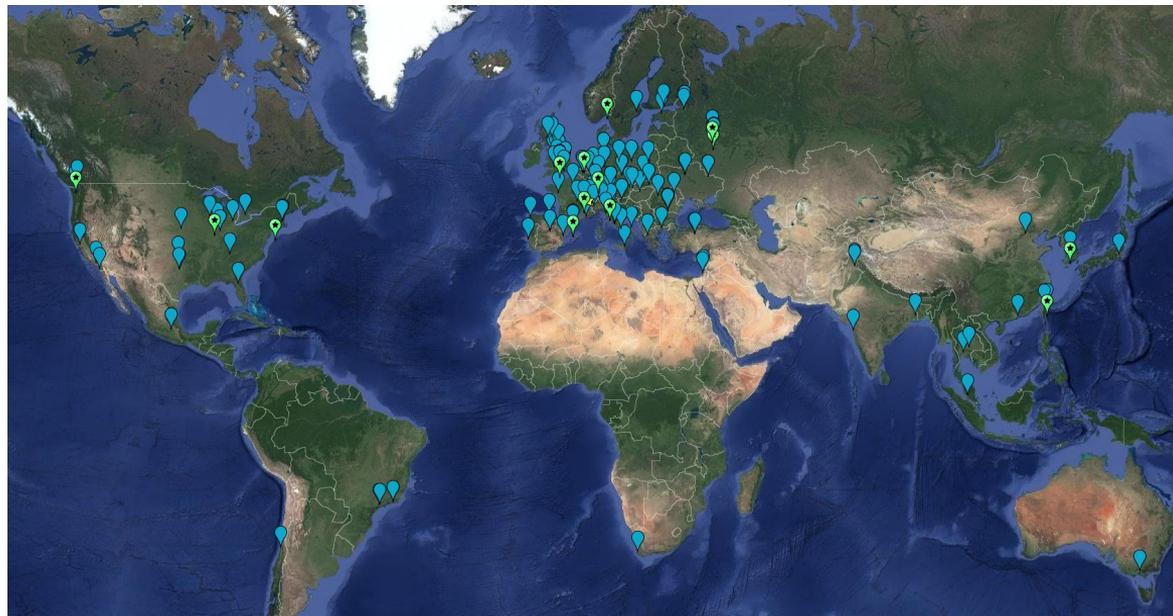
La grille de calcul

La grille de calcul

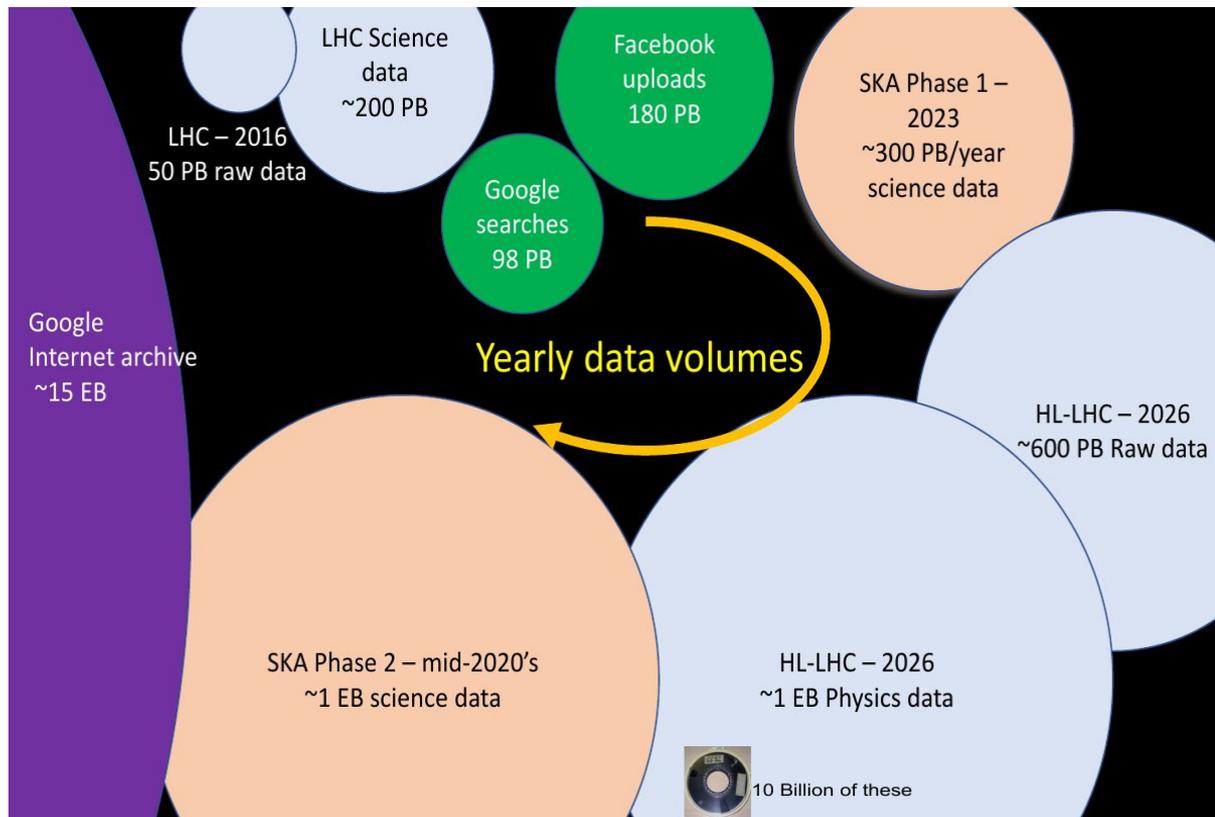
- Le CERN seul n'a pas la **puissance de calcul et le stockage** pour gérer l'ensemble des données du LHC
- Worldwide LHC Computing Grid (WLCG): **Fédération mondiale** de centres de calcul
 - Créé en 2002
 - Distribution, stockage, traitement et analyse des données
 - 42 pays, 170 sites
 - >8000 utilisateurs, venant des 4 expériences
 - 2 millions de jobs / jour, sur 750000 coeurs et >1000 Po de stockage
- **Avantages de la grille**
 - Données copiées à plusieurs endroits
 - Support 24/7 avec des sites sur tous les continents
 - Système souple et adaptable en fonction des incidents dans les différents centres

Les sites de la grille

- **Tier-0**
 - Le CC du CERN
 - Reconstruction des données
- **Tier-1**
 - Quelques (14) gros centres de calcul, i.e CC-IN2P3
 - Simulations, campagnes de re-reconstruction
 - Stockage des données
- **Tier-2**
 - Nombreux sites dans les universités (~140)
 - Analyse de données, simulations
- **Supercalculateurs**
 - Utilisation additionnelle ad-hoc
 - Simulations



D'énormes quantités de données



- Les infrastructures de transfert doivent suivre les besoins de stockage
 - Transferts à 60 Go/s maintenus pendant des jours

La grille en pratique



● Systèmes centraux

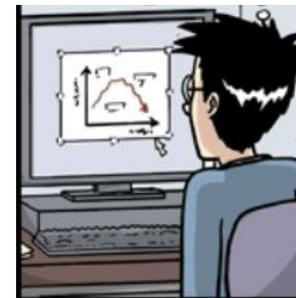
- Catalogues des données
- Gestion des queues de jobs
- Bases de données diverses

● Utilisateurs

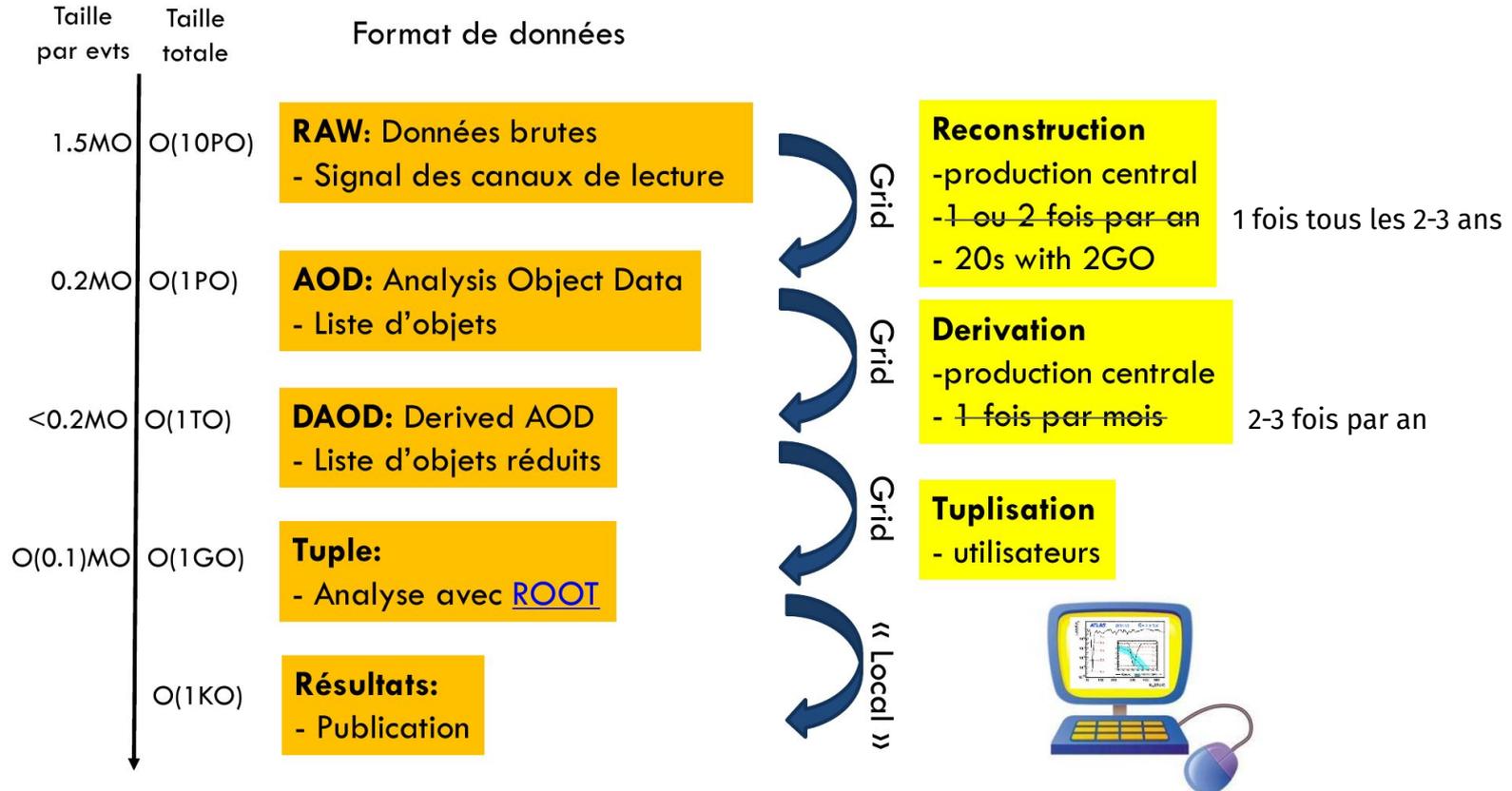
- Prototypage des programmes sur laptop ou serveur interactif
- Envoi du programme 'sur la grille' pour tourner sur l'ensemble des données
 - **Le programme va là où les données sont**
- Téléchargement des résultats

● Sites

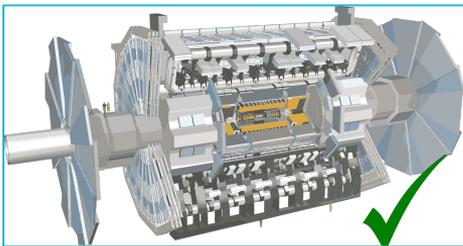
- Matériel hétérogène
- Système d'exploitation et bibliothèques uniformes
- Accès aux logiciels de physique des particules et des expériences
- Accès aux bases de données centrales



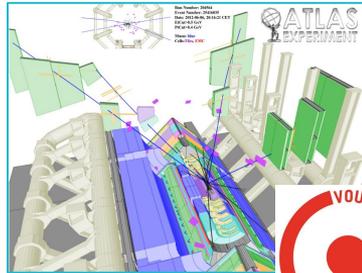
Flux d'analyse des données



Rappel: Cheminement des données



Détecteur



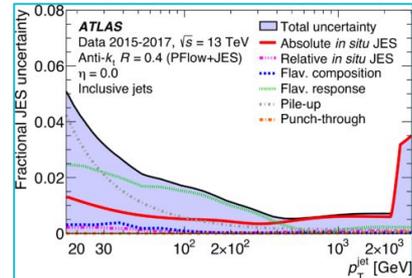
Simulation



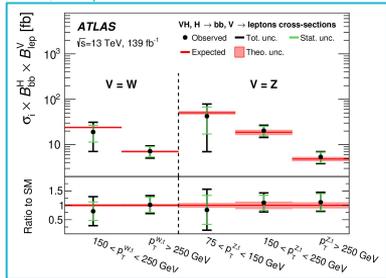
Trigger / Acquisition



Calibration
Qualité des données
Reconstruction



Performance de la reconstruction



Analyse de données

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Measurements of W and ZH production in the $H \rightarrow b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector

ATLAS Collaboration*

Abstract

1 Introduction

2.1 Experimental uncertainties

2.2 Background uncertainties

2.3 Signal uncertainties

3 Statistical analysis

4 Results

5.1 Signal strength measurement

5.2 Cross-section validation

6 Cross-section measurements

7 Constraints on effective interactions

8 Conclusions

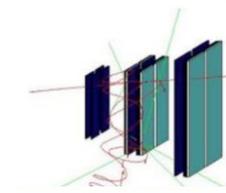
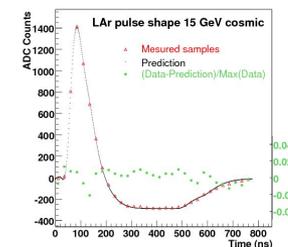
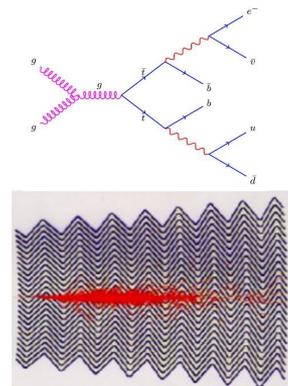
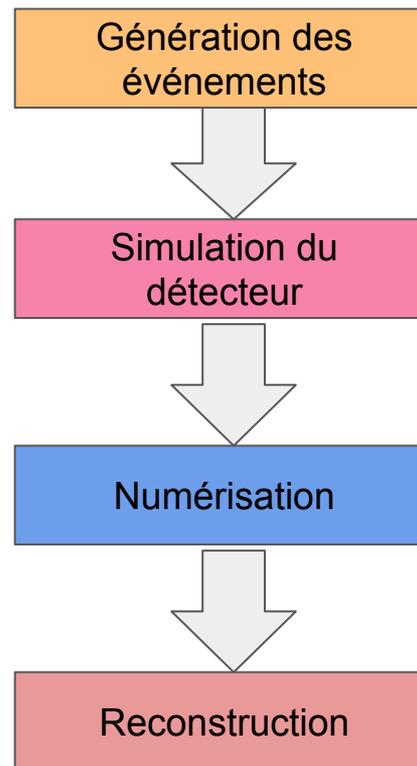
References

Publication

Simulation

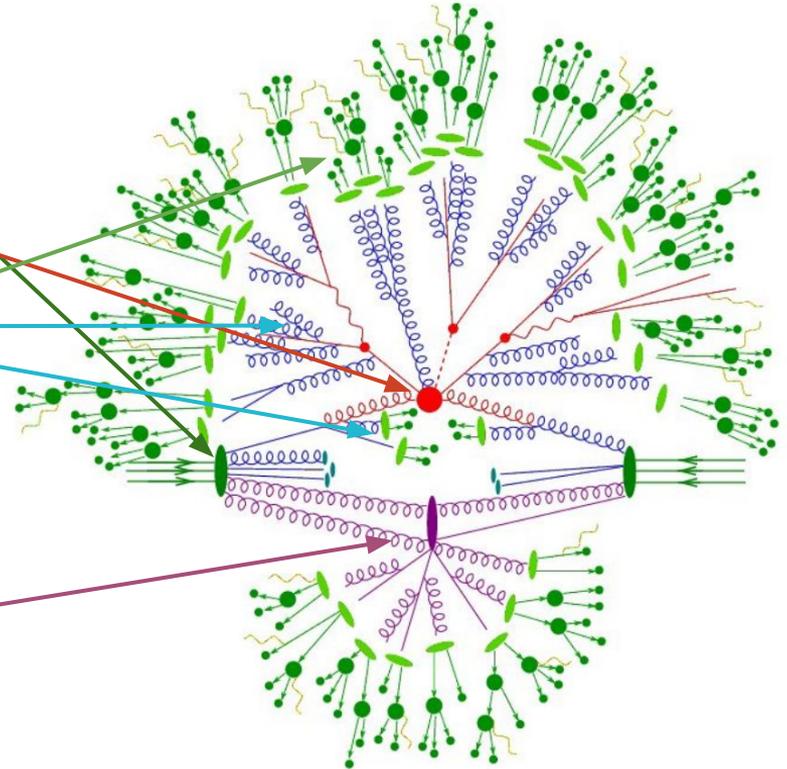
Intérêt de la simulation

- Lors du design du détecteur
 - Permet de **prédire et optimiser** ses performances
- Lors de la préparation de l'analyse de données
 - Prédiction et optimisation de la mesure / recherche
- Lors de la réalisation de l'analyse
 - **Confrontation permanente** des données et de la simulation pour bien comprendre les phénomènes en jeu



Génération des événements

- Une collision: un processus complexe
 - Quelle partie du proton interagit ?
 - Processus "dur" (diagrammes de Feynman)
 - Radiations
 - Formation des gerbes de partons
 - Hadronization
 - reconnexion de couleur
 - désintégrations des hadrons
 - Événement sous-jacent



Les “générateurs Monte Carlo”

- Nombreux programmes disponibles
 - Certains spécialisés sur certains aspects
 - Ou sur certains processus
 - De précision et de complexité croissante
 - Temps de génération d'un événement: de qq secondes à ~1 heure



Monte Carlo ?

- Les processus à l'œuvre sont régis par la mécanique quantique
 - **Intrinsèquement aléatoires !**
 - À toutes les étapes
 - La théorie ou les modèles phénoménologiques donnent des **distributions de probabilité**
- Génération Monte-Carlo
 - **Méthode numérique pour approcher une valeur numérique en utilisant des procédés aléatoires**
 - Permet:
 - Tirages de nombres aléatoires suivant des distributions arbitraires
 - Intégration de ces distributions pour calculer des sections efficaces
 - Dans des espaces de grande dimension

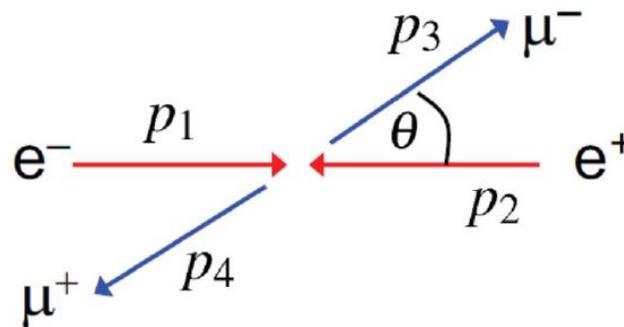
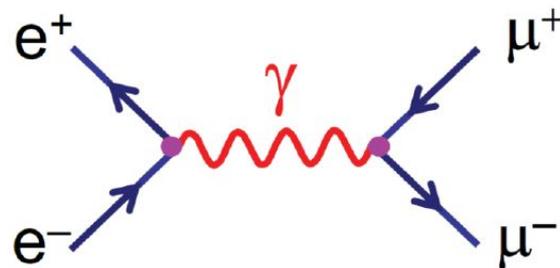
Exemple: le processus $e^+e^- \rightarrow \gamma \rightarrow \mu^+\mu^-$

- Processus simple de QED
 - État final décrit par 2 variables: θ et ϕ
- La théorie permet de calculer la section efficace différentielle:

$$\frac{d\sigma}{d\cos\theta d\phi} = \frac{\alpha_{\text{em}}^2}{4s} (1 + \cos^2\theta)$$

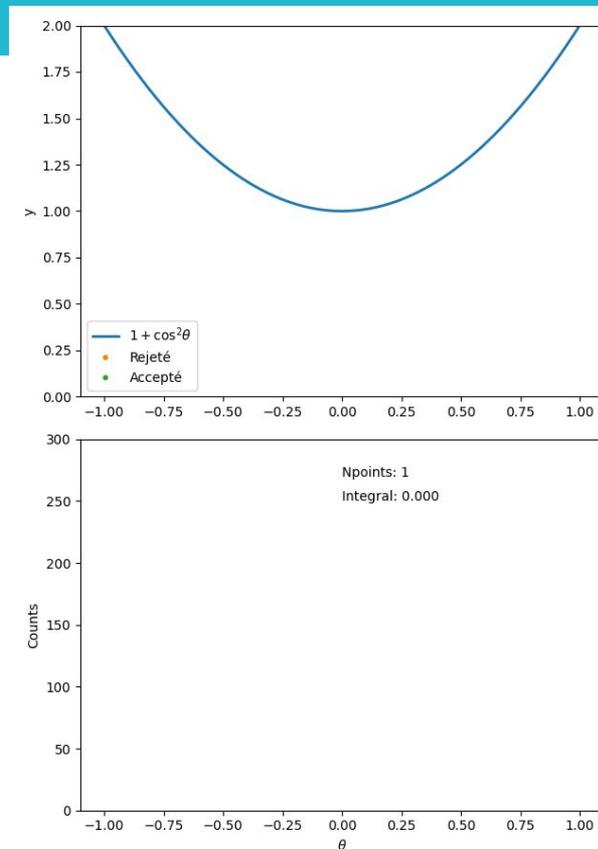
- Section efficace totale:

$$\sigma = \frac{\alpha_{\text{em}}^2}{4s} \int_{\Omega} (1 + \cos^2\theta) d\cos\theta d\phi = \frac{4\pi\alpha_{\text{em}}^2}{3s}$$



Génération d'événements $e^+e^- \rightarrow \gamma \rightarrow \mu^+\mu^-$

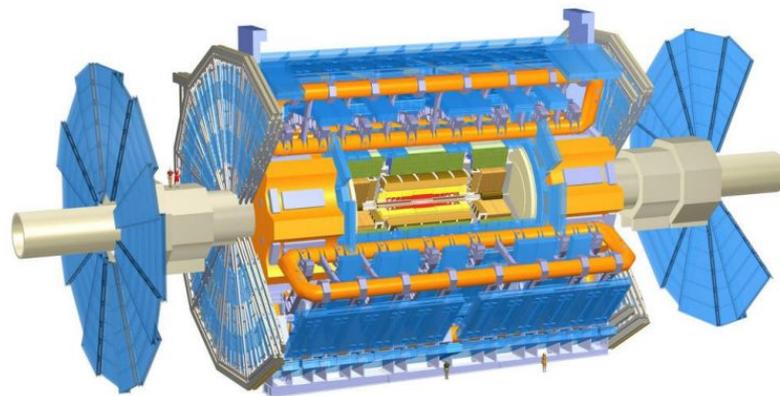
- Distribution uniforme en ϕ
 - Facile: aléatoire primitif de l'ordinateur
- Distribution en $1+\cos^2\theta$
 - Plus difficile: à générer à partir de distributions uniformes
 - Algorithme "hit and miss"
 - On tire $\cos \theta$ entre -1 et 1 et y entre 0 et 2
 - 10000 tirages
 - On accepte le tirage si $y < 1+\cos^2\theta$
 - **La distribution du tirage suit $1+\cos^2\theta$**
 - Le rapport du nombre de tirages acceptés sur le total donne l'intégrale de $1+\cos^2\theta$
 - $0.6684 \sim \frac{2}{3}$



$$\frac{d\sigma}{d\cos\theta d\phi} = \frac{\alpha_{em}^2}{4s} (1 + \cos^2\theta)$$

Simulation du détecteur

- Les particules générées doivent passer dans une simulation du détecteur
 - Phase de **transport**
- Il faut:
 - Représentation virtuelle du détecteur
 - Géométrie, matériaux
 - Propriétés des particules
 - Simulation du passage des particules dans la matière
 - Modélisations des interactions, voir cours de P. Schune
- Encore des processus aléatoires
 - Utilisation de méthodes de Monte Carlo



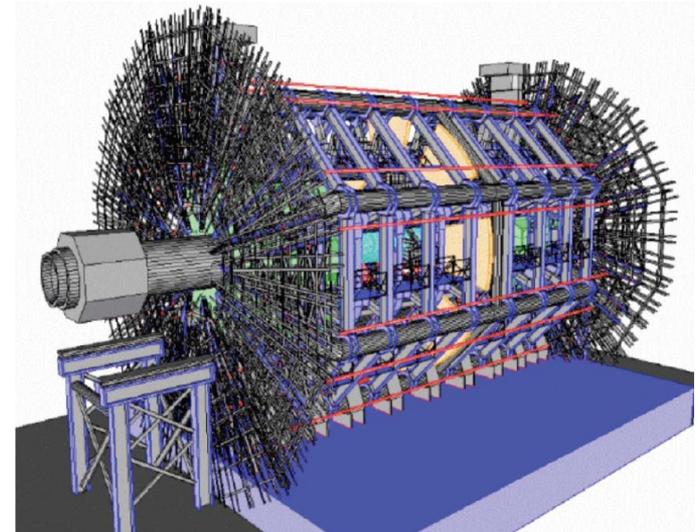
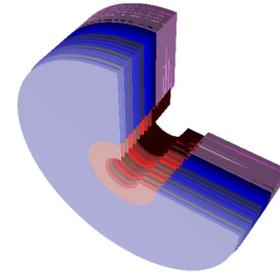
GEANT 4: GEometry ANd Tracking

- Framework initié par le CERN en 1994
 - Aujourd'hui: ~100 collaborateurs internationaux
 - Entièrement ouvert et gratuit
- Programmation orientée objet (C++)
- Permet la **description des détecteurs** et la **simulation du passage des particules**
- Utilisé en physique des particules, physique nucléaire, physique des accélérateurs, physique médicale



Modélisation du détecteur

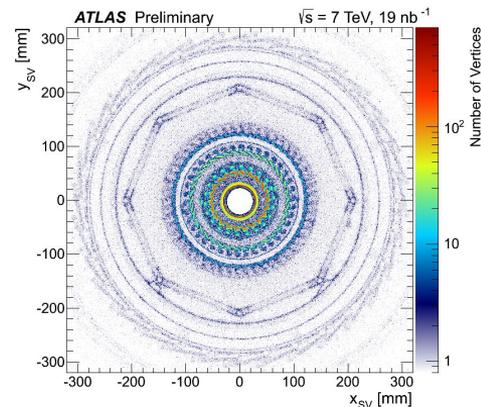
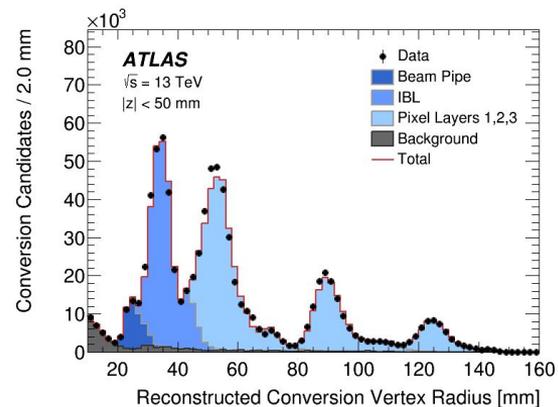
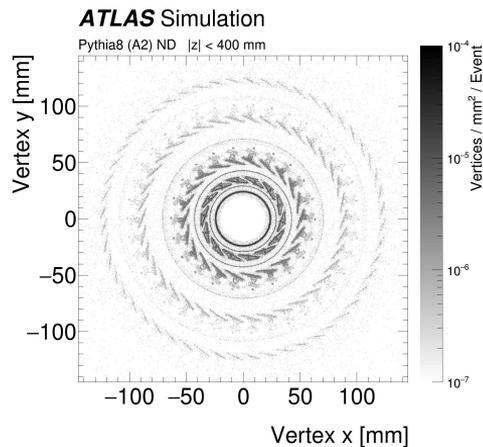
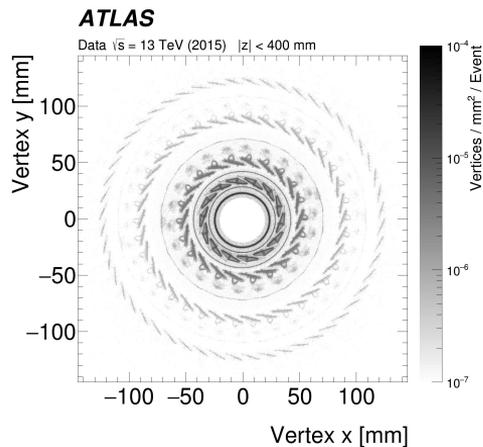
- Géométrie créée en combinant des formes simples
 - boîtes, cylindres, trapèzes...
 - possibilités d'extruder, de couper, de combiner...
 - similaire à d'autres logiciels de CAO
- Nombreux matériaux prédéfinis
- ATLAS: 5 millions de volumes !
 - description fidèle nécessaire: détecteurs, mais aussi supports et services
 - ex: performance des calorimètres dépendante de la quantité de matière dans le trajectomètre



Vérification de la simulation

Les données permettent de vérifier certaines parties

- Photons convertis
- Interactions hadroniques



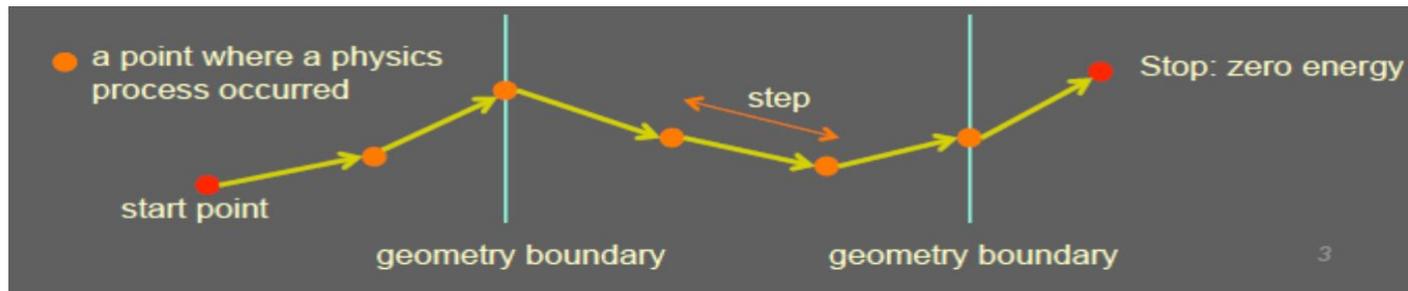
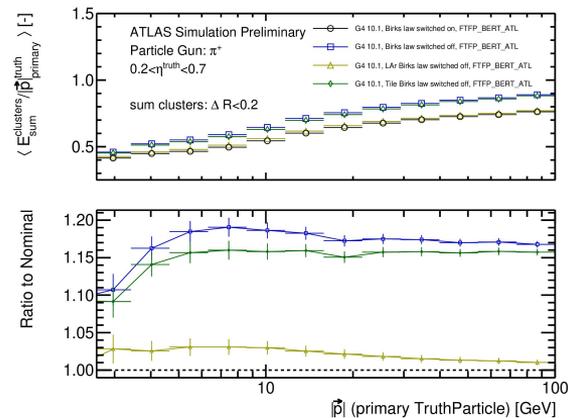
Interactions particules-matière

- **Simulation pas à pas**

- A chaque pas, on choisit ce qui se produit en fonction des probabilités d'interaction
- Dépend du type de particule et de la matière traversée
- Importance de la bonne description du détecteur
- Simulation des gerbes peut être très coûteuse en CPU

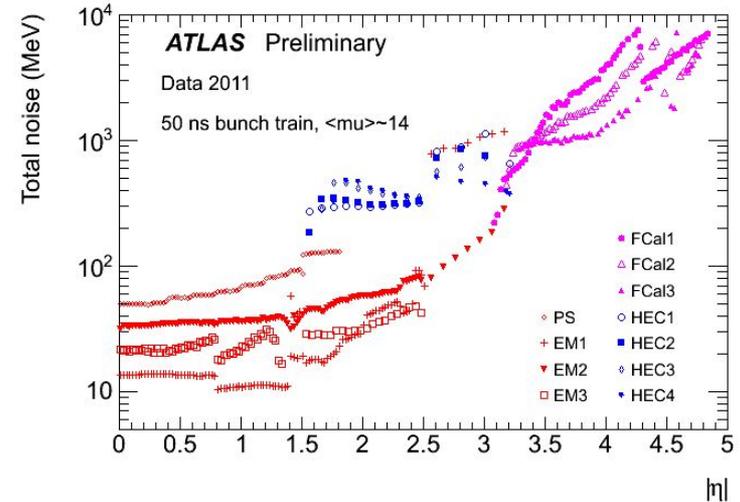
- **Interactions particules-matière**

- Modèles phénoménologiques pour les différents types d'interaction
- Choix de 'physics lists' et nombreux réglages pour trouver la meilleure description pour chaque application



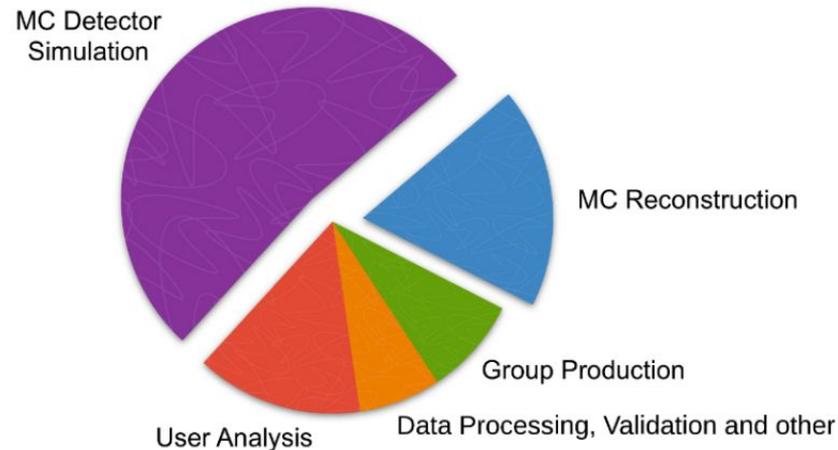
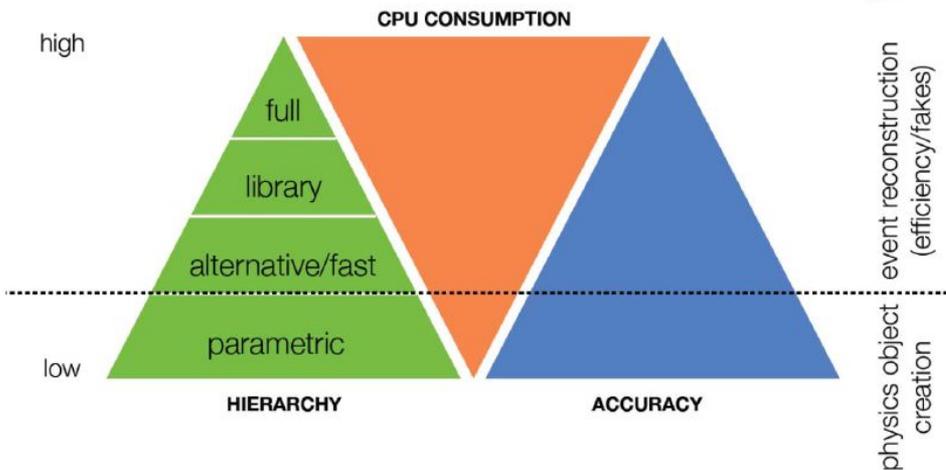
Numérisation des données

- Énergie déposée dans les volumes de détection collectée pour fabriquer des fausses données **au même format que les données de collision**
 - Par exemple on regroupe l'énergie déposée dans les volumes actifs des cellules des calorimètres
 - On applique la réponse attendue de la voie d'électronique (seuil de détection...)
 - On ajoute le bruit de l'électronique
- On prend également l'empilement en compte à cette étape
 - À partir de jeux de données d'empilement préalablement conçus
- Les données simulées peuvent alors passer par les **mêmes algorithmes de reconstruction** que les vraies données



Vers la simulation rapide

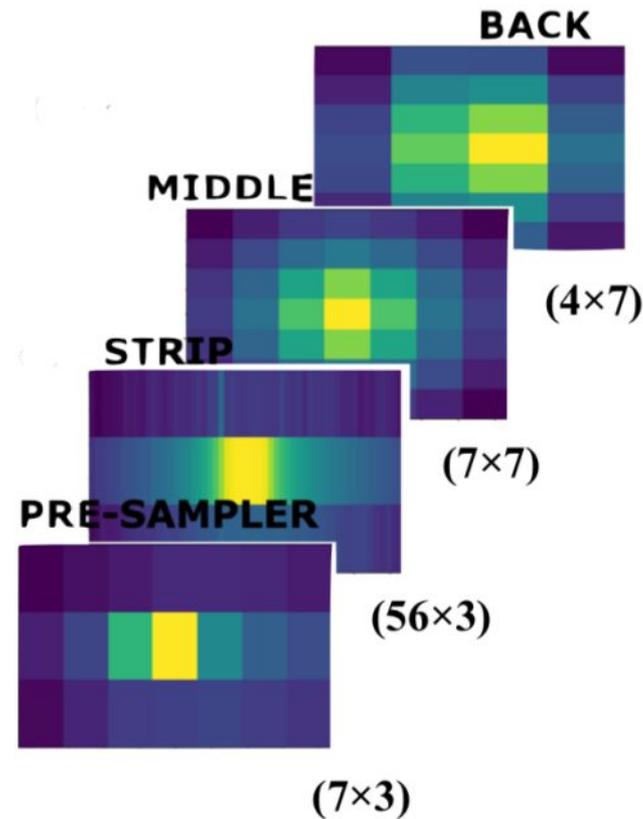
- Simulation complète du détecteur: un gouffre à CPU et à RAM
 - ~10min / événement
 - La moitié des ressources de la grille !
 - Mais c'est ce qu'on a de plus précis



- Alternatives plus rapides
 - Paramétrisation des résultats de la simulation complète
 - Beaucoup plus rapide, mais moindre précision (corrélations)
 - Peut-on avoir le meilleur des deux mondes ?

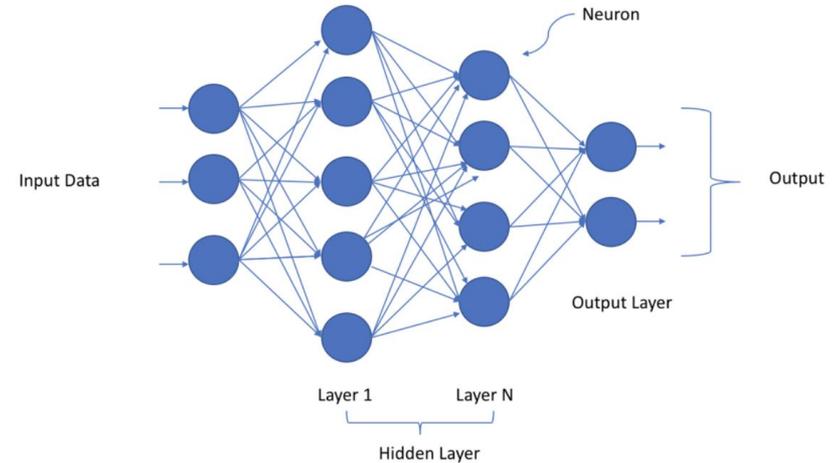
Simulations rapides des gerbes

- Les gerbes dans le calorimètre
 - Formes caractéristiques de gerbes dans chaque couche
 - Corrélées entre les couches
 - Dépendant du point d'impact de la particule dans le calorimètre
 - La somme des énergies doit correspondre à celle de la particule initiale
 - Variabilité intrinsèque entre les gerbes (processus aléatoires)
- Paramétrisations
 - Beaucoup de cas différents à traiter
 - Difficile de tenir compte de toutes les corrélations
- Solution: **laisser l'ordinateur faire le boulot !**



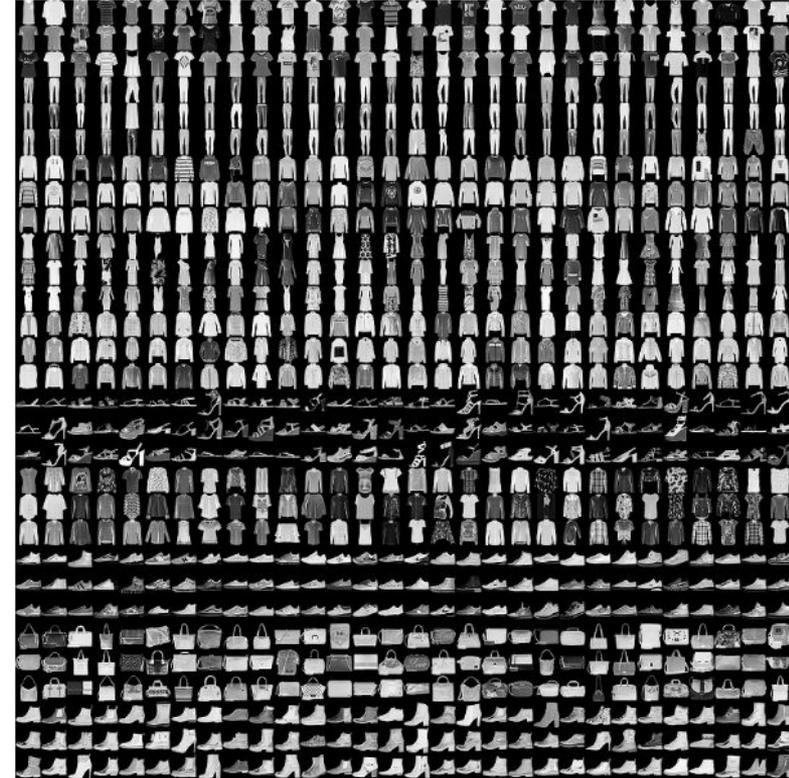
L'apprentissage automatique à la rescousse

- Les **réseaux de neurones**
 - **Neurone**: fonction simple à N entrées et M sorties, dépend de paramètres additionnels
 - **Réseau**: collection de neurones reliés entre eux
 - **Architecture du réseau**: choix des types de neurones, nombre de couches, relations entre eux
 - **Entraînement**: optimisation des paramètres additionnels sur la base d'un jeu de données, afin de minimiser une métrique astucieusement choisie "loss function"
 - **Inférence**: Une fois entraîné, le réseau peut être utilisé pour la tâche qu'il a optimisée à partir de nouvelles entrées



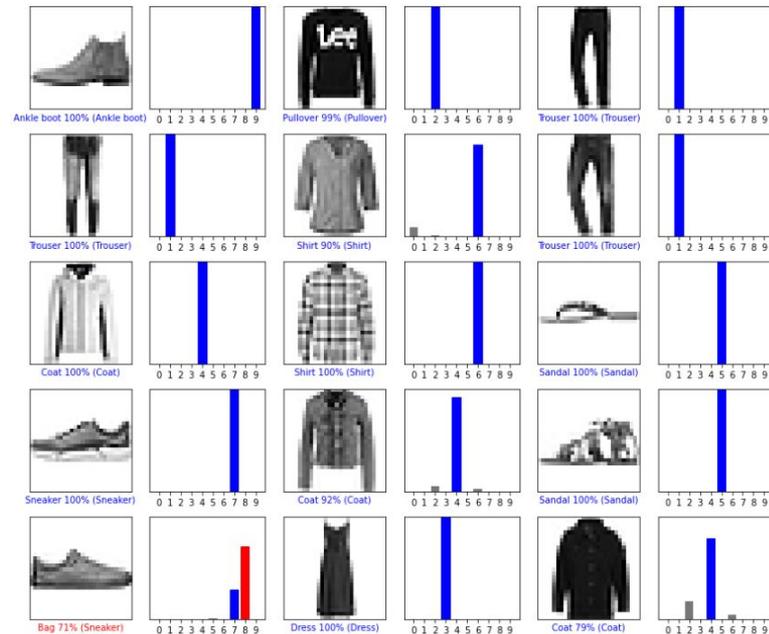
Apprentissage automatique: exemple

- **Classification d'images de vêtements**
 - Variables d'entrée: les 28x28 pixels d'une image en niveaux de gris
 - Sortie: probabilité d'être dans une des classes (pull, robe, chaussure...)
- **Entraînement**
 - Images dont la catégorie est connue
 - Loss function:
 - petite valeur si on choisit la bonne catégorie
 - grande valeur si on se trompe



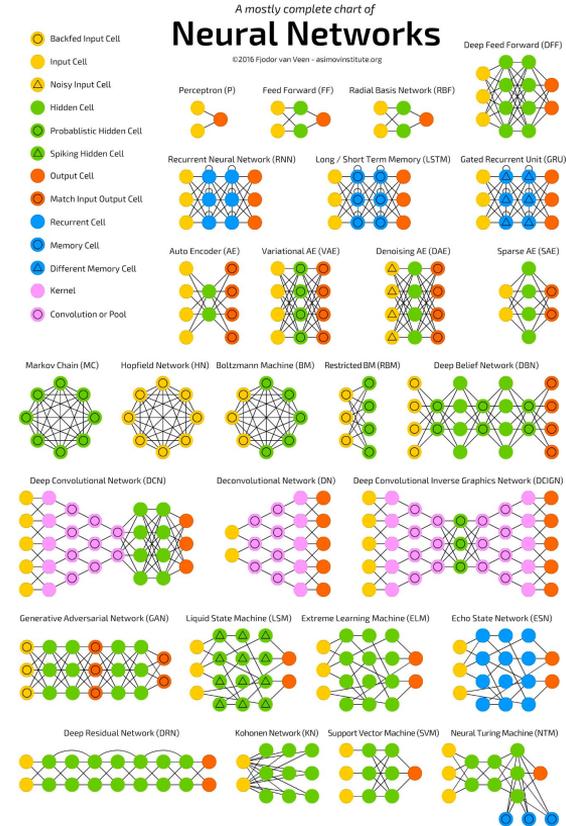
Apprentissage automatique: exemple

- Classification d'images de vêtements
 - Variables d'entrée: les 28x28 pixels d'une image en niveaux de gris
 - Sortie: probabilité d'être dans une des classes (pull, robe, chaussure...)
- Entraînement
 - Images dont la catégorie est connue
 - Loss function:
 - petite valeur si on choisit la bonne catégorie
 - grande valeur si on se trompe
- Application à des images inconnues
 - Permet de **prédire** leur catégorie



Le “Deep learning”

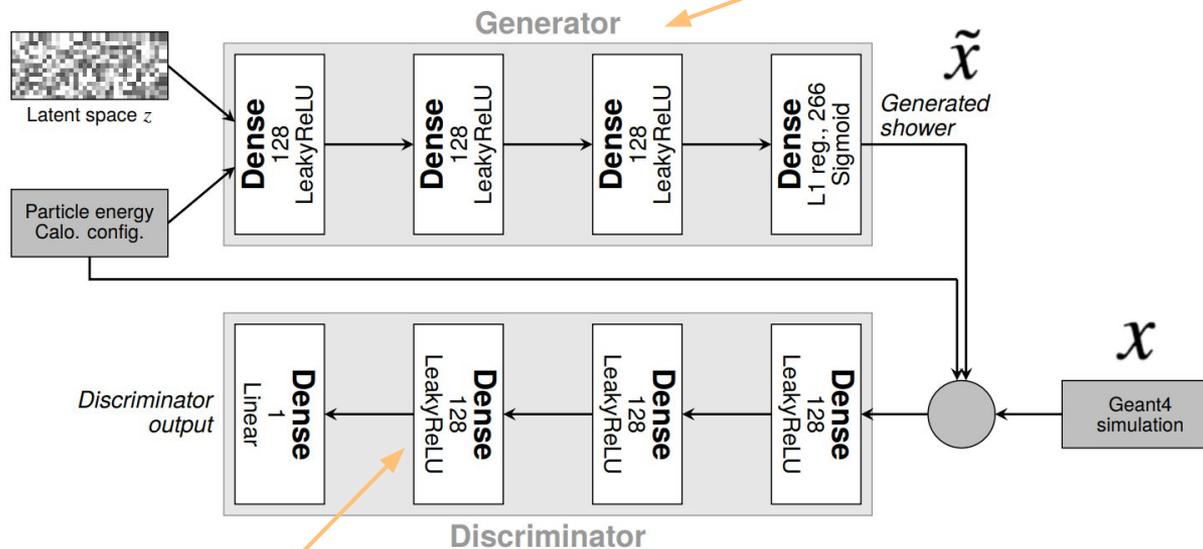
- Avancée majeure de ces 10 dernières années
- Capacité à entraîner de très grands réseaux avec de **nombreuses entrées** et **beaucoup de paramètres**
 - Meilleurs algorithmes d’optimisation “descente de gradient”
 - Capacités hardware, programmation GPU
- Explosion des possibilités
 - Beaucoup de créativité dans les architectures
 - De plus en plus d’applications en physique des hautes énergies



Réseaux générateurs adversariaux (GAN)

Deux réseaux de neurones jouent l'un contre l'autre

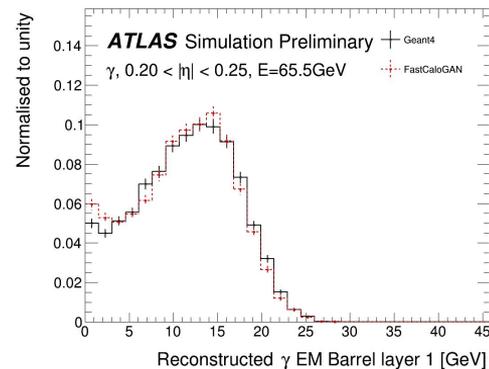
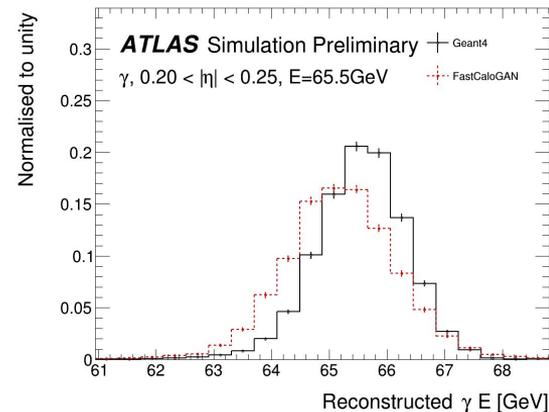
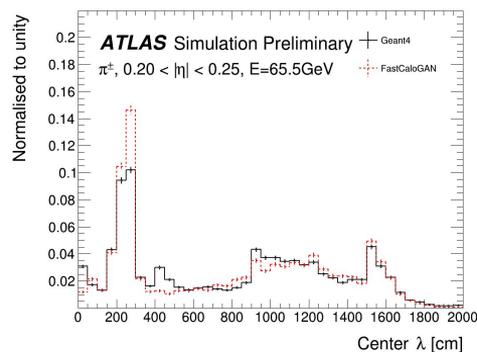
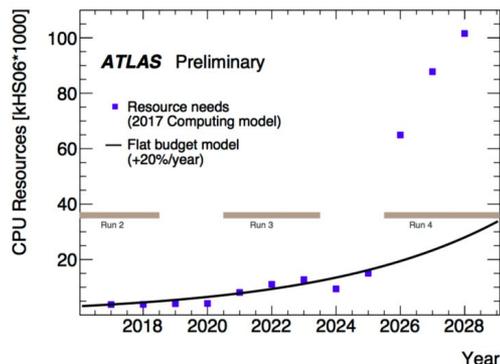
Génère des gerbes simulées



Essaie de séparer les gerbes
générées de gerbes produites
par GEANT

Est-ce que ça marche ?

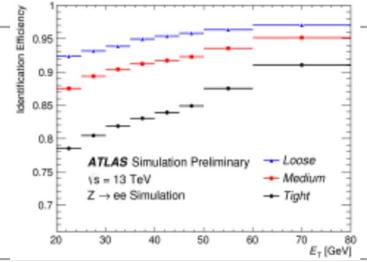
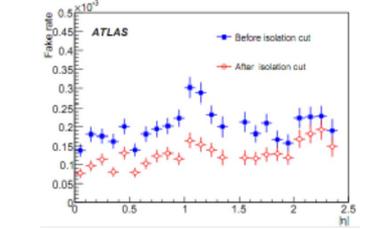
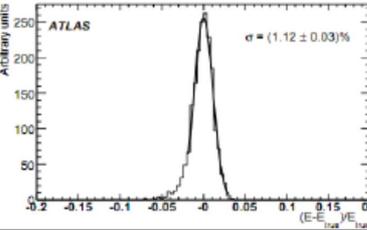
- Pas encore tout à fait
 - Performance comparable aux paramétrisations
 - ~100 fois plus rapide que GEANT (10 fois moins que la paramétrisation)
 - Mais progrès très rapides
- De grands espoirs pour le HL-LHC
 - La quantité de simulations requise impose des gains de CPU significatifs



Performance de la reconstruction

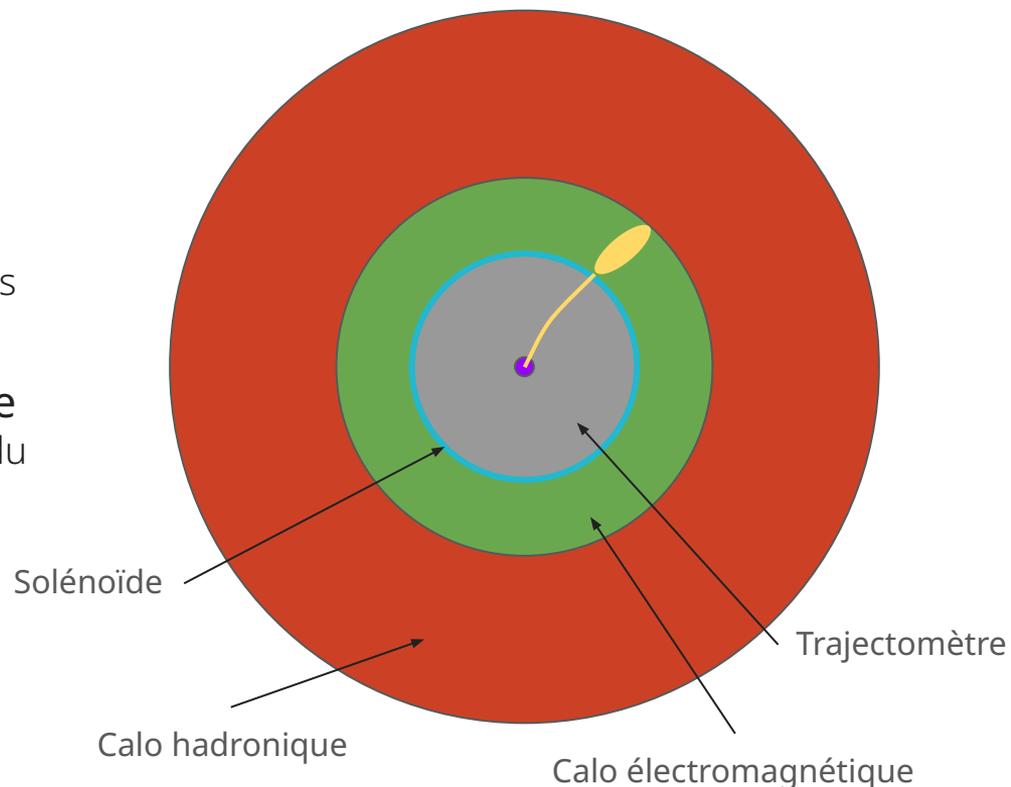
Sur deux exemples choisis

Objectifs de la reconstruction: facteurs de mérite

	Définition	Exemple		Objectif
Efficacité	$\frac{\# \text{ objets reconstruits}}{\# \text{ objets total}}$	Efficacité des électrons: $\frac{\# \text{ électrons reconstruits}}{\# \text{ électrons produits}}$		Grande
Taux de faux positifs	$\frac{\# \text{ objets B reconstruits comme A}}{\# \text{ objets B}}$	Jets mal reconstruits comme électrons: $\frac{\# \text{ jets pris pour des électrons}}{\# \text{ jets}}$		Bas
Résolution	Précision de la mesure (énergie, position...)	Résolution en énergie d'un électron: $\frac{E(\text{mes}) - E(\text{vrai})}{E(\text{vrai})}$		Fine

Exemple 1: mesure de l'énergie des électrons

- Rappel: reconstruction des électrons
 - Trace chargée + amas calorimétrique
- Reconstruction du quadrivecteur
 - Direction donnée par la trace
 - Énergie donnée par l'énergie de l'amas
 - **Pas assez précis !**
- Nécessité de l'étalonnage de l'énergie
 - Pertes dans les matériaux en amont du calorimètre
 - Fuites d'énergie hors de l'amas
 - latérales et longitudinales
 - Inhomogénéités en θ et ϕ

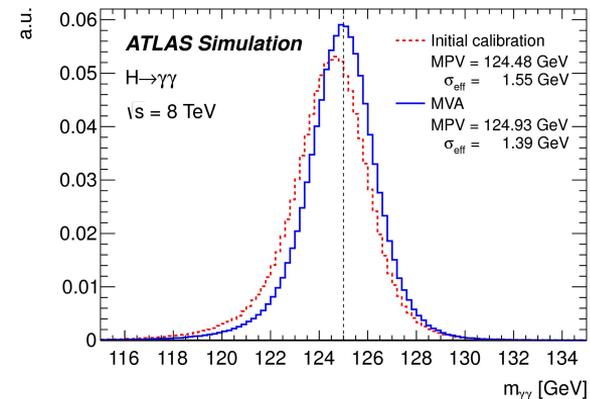
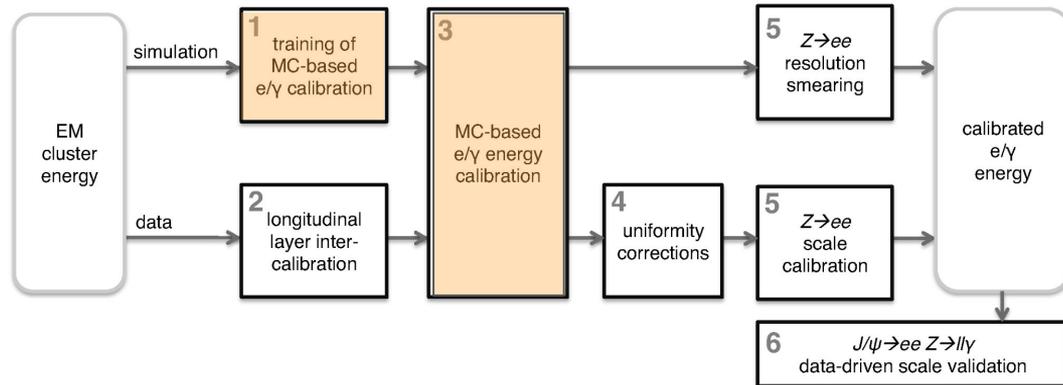


Procédure d'étalonnage

- Correction de l'énergie obtenue par la simulation
 - Régression multivariée
 - Apprentissage automatique à partir de l'énergie mesurée de l'amas, de sa position, et de caractéristiques de sa forme
 - But: améliorer la résolution en énergie

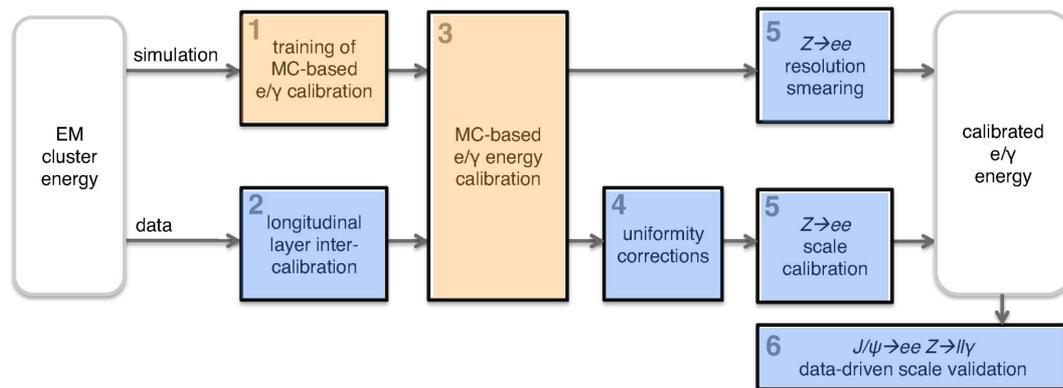
- Résultats

- Corrections de 5 à 15%
- Amélioration de la résolution de $\sim 10\%$



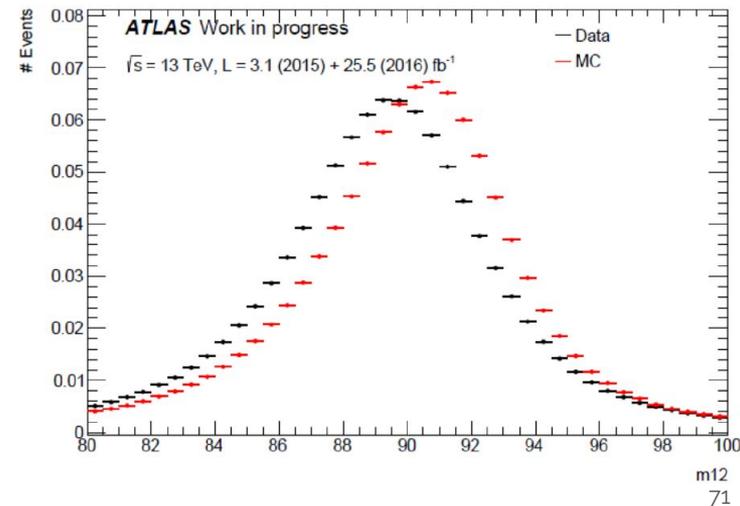
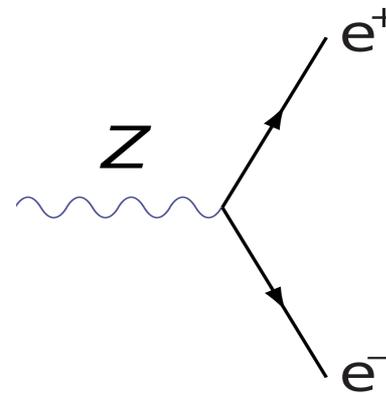
Procédure d'étalonnage

- Correction de l'énergie obtenue par la simulation
- Étalonage final à partir des données
 - Puisque la simulation ne représente pas le détecteur parfaitement
 - Nombreuses étapes



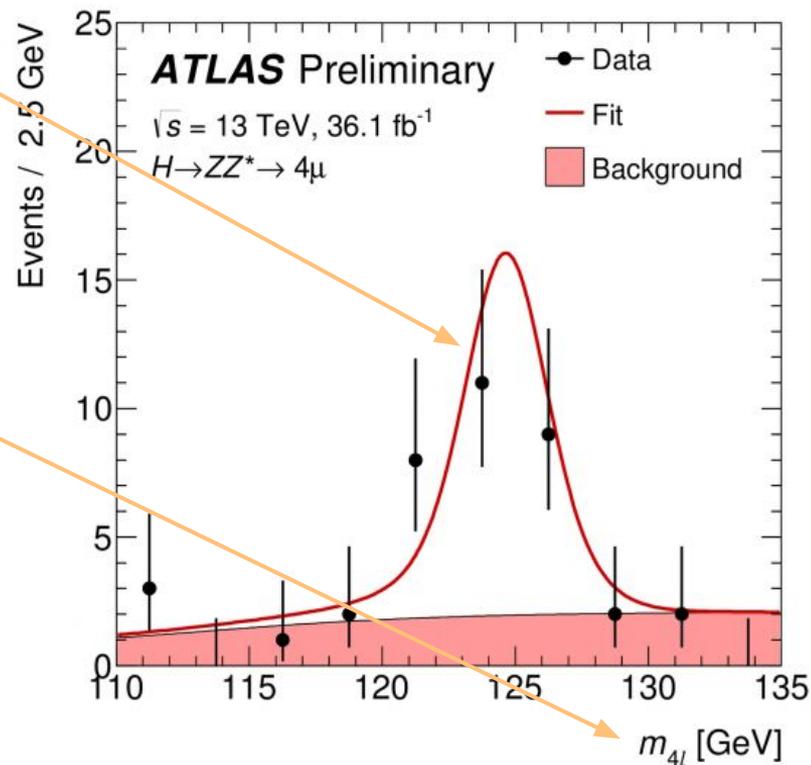
Étalonnage avec le boson Z

- Masse du boson Z connue très précisément depuis les expériences au LEP:
 - $m_Z = 91.1876(21)$ GeV
- Masse invariante à partir d'une paire e^+e^-
 - $m_{ee} = \sqrt{(2 E_1 E_2 (1 - \cos \theta_{12}))}$
 - mesure de θ_{12} bien plus précise que celles de E_1 et E_2
- Comparaison données et simulation
 - Position du maximum donne l'échelle d'énergie
 - Largeur du pic donne la résolution
 - Facteurs correctifs extraits par un **maximum de vraisemblance**

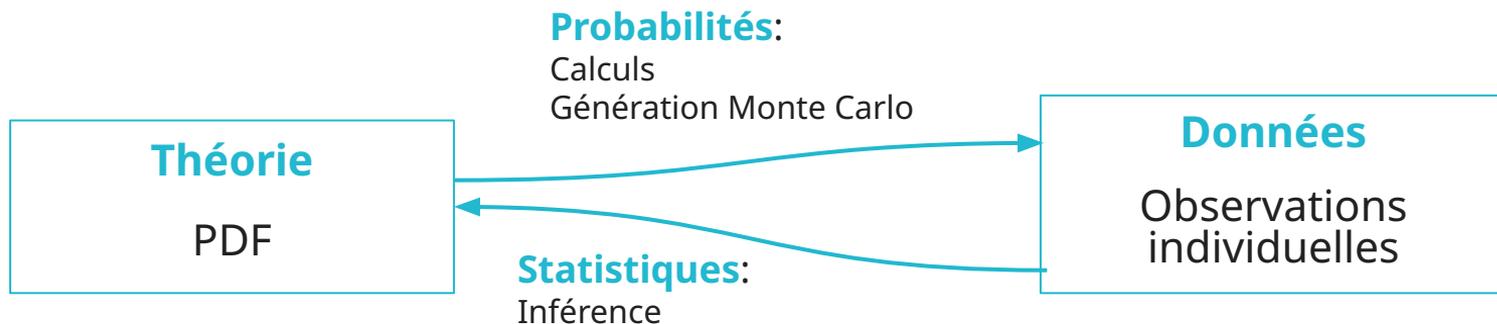


Détour: l'inférence statistique

- La théorie / les simulations: **fonctions de densité de probabilité** (PDF) pour un processus donné
 - Dépendent de **paramètres** qu'on veut mesurer (masse du boson de Higgs...)
- La distribution est celle d'une **variable aléatoire**



L'inférence statistique



- Objectifs des analyses:
 - **Estimation de paramètres** des PDFs à partir des mesures
 - **Test d'hypothèses**: est-ce que la théorie est compatible avec les observations ?

Fonction de vraisemblance

- Pour réaliser des inférences, il faut pouvoir **quantifier l'accord** entre les données et la théorie
 - On veut obtenir des **estimateurs non biaisés et efficaces** (faible variance) des paramètres
 - Plusieurs méthodes: chi2, fonction de vraisemblance
- **Fonction de vraisemblance**
 - Maximum de vraisemblance: pour des mesures données, calculer θ les qui maximisent \mathcal{L}
 - Ces $\hat{\theta}$ sont des estimateurs de la valeur réelle
 - La méthode donne aussi les incertitudes sur les $\hat{\theta}$

$$\mathcal{L}(x_1, \dots, x_n \mid \theta_1, \dots, \theta_k) = \prod_i f(x_i \mid \theta_1, \dots, \theta_k)$$

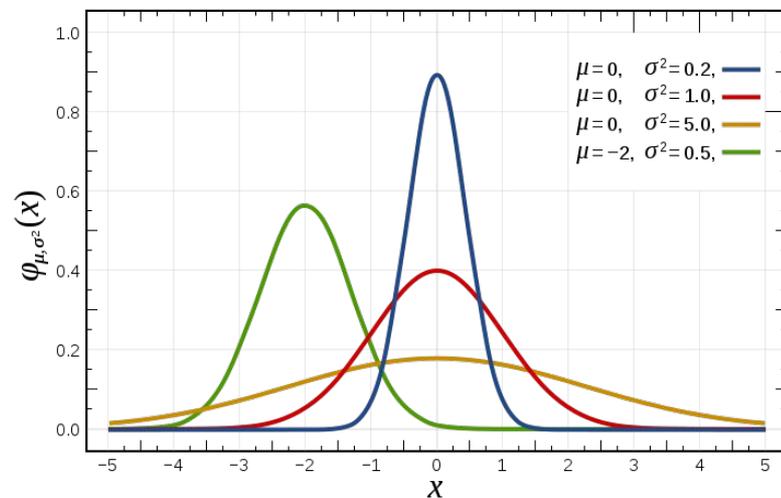
Mesures des observables

Paramètres

PDF

Exemple: estimation des paramètres d'une gaussienne

- Processus régi par une distribution gaussienne, dont on ne connaît pas la moyenne ni l'écart-type
 - Exemple: Distribution de masse invariante du boson $Z \rightarrow e^+e^-$
 - **Paramètres:** $\theta \mapsto \mu$ valeur centrale, σ écart-type
 - **Observables:** x_i
 - **PDF:** $f \mapsto G(x | \mu, \sigma) = 1/\sqrt{2\pi\sigma^2} \exp(- (x-\mu)^2 / 2\sigma^2)$



Estimation des paramètres

- On veut maximiser:

$$\mathcal{L}(x_i|\mu, \sigma) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

- En pratique, on préfère minimiser:

- Cela donne:

$$NLL = -\log \mathcal{L}(x_i|\mu, \sigma) = \frac{N}{2} \log(2\pi\sigma^2) + \sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$\frac{\partial -\log \mathcal{L}(x_i|\mu, \sigma)}{\partial \mu} = 0$$

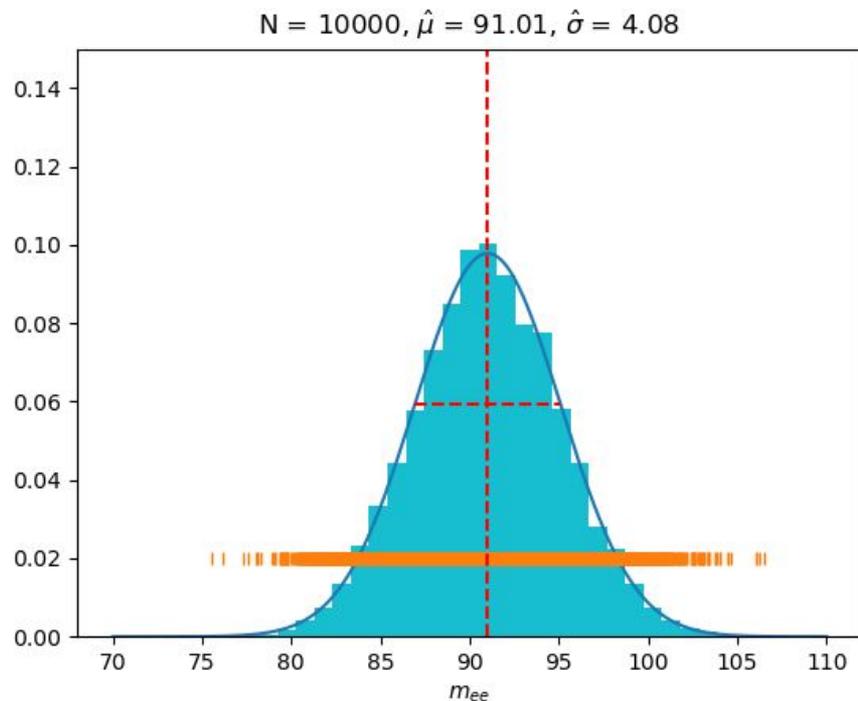
$$\frac{\partial -\log \mathcal{L}(x_i|\mu, \sigma)}{\partial \sigma} = 0$$

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2}$$

Moyenne arithmétique !

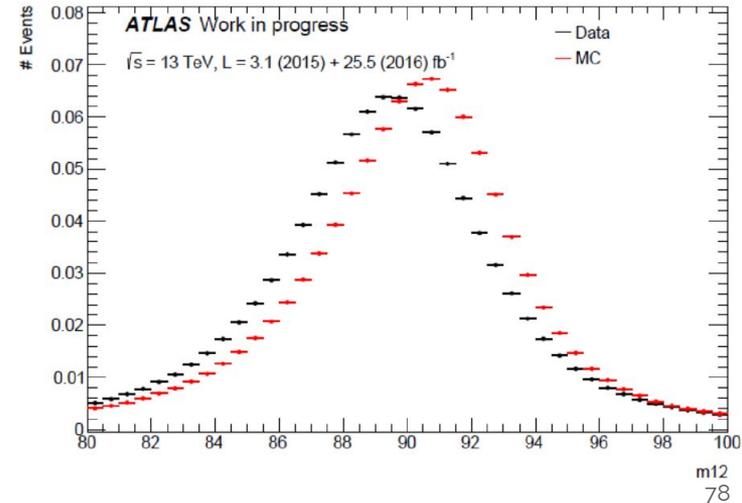
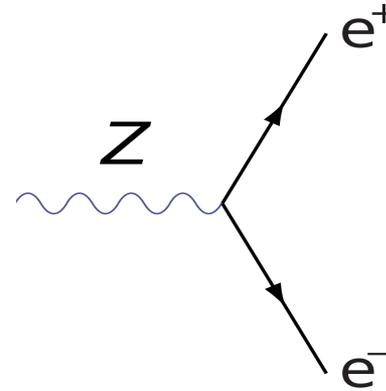
Exemple en action



- Quand on ajoute des données
 - $\hat{\mu}$ converge vers $\mu=91$
 - $\hat{\sigma}$ converge vers $\sigma=4$
 - L'incertitude sur l'estimation diminue également
- Maximisation de la fonction de vraisemblance
 - Calcul analytique sur cet exemple
 - **Méthodes numériques** le plus souvent: programme **Minuit**

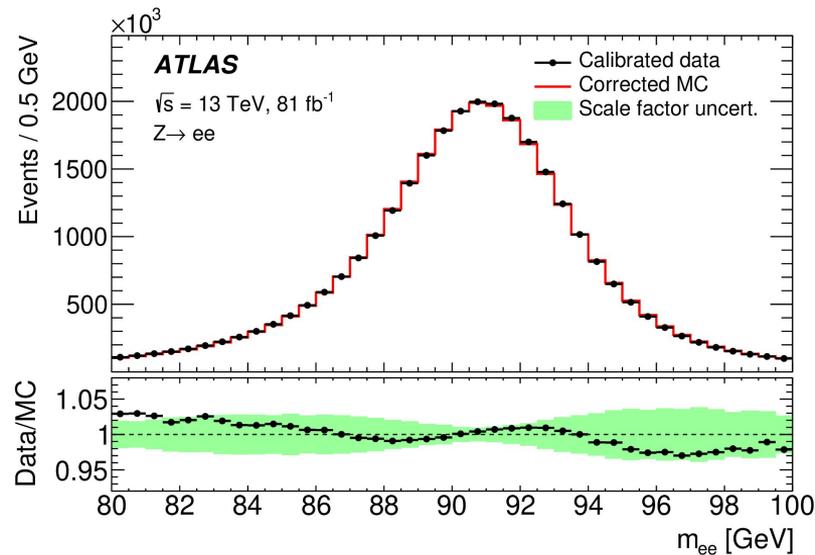
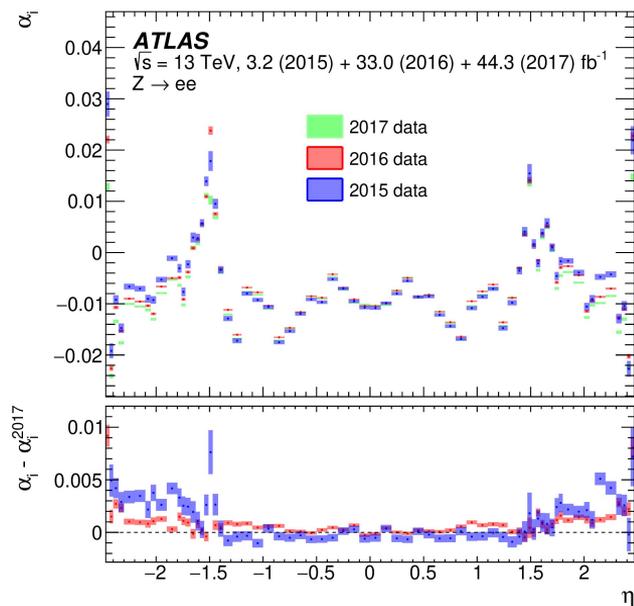
Reprise: Étalonnage avec le boson Z

- Masse du boson Z connue très précisément depuis les expériences au LEP:
 - $m_Z = 91.1876(21)$ GeV
- Masse invariante à partir d'une paire e^+e^-
 - $m_{ee} = \sqrt{(2 E_1 E_2 (1 - \cos \theta_{12}))}$
 - mesure de θ_{12} bien plus précise que celles de E_1 et E_2
- Comparaison données et simulation
 - Position du maximum donne l'échelle d'énergie
 - Largeur du pic donne la résolution
 - Facteurs correctifs extraits par un **maximum de vraisemblance**



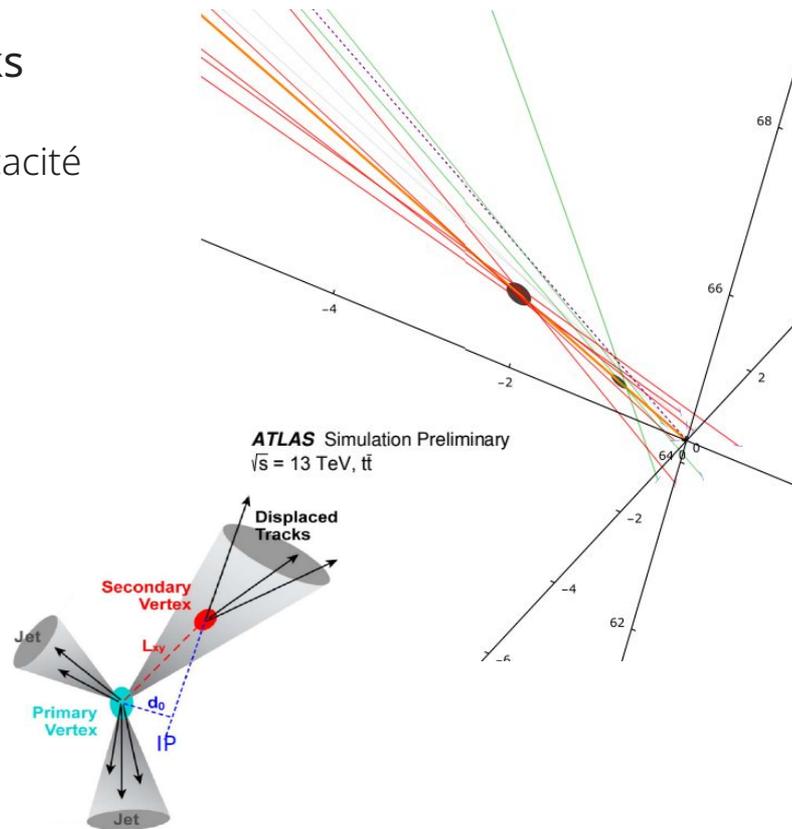
Étalonnage avec la masse du boson Z

- Corrections de la simulation: 1-2%
- Incertitudes associées: <0.3%



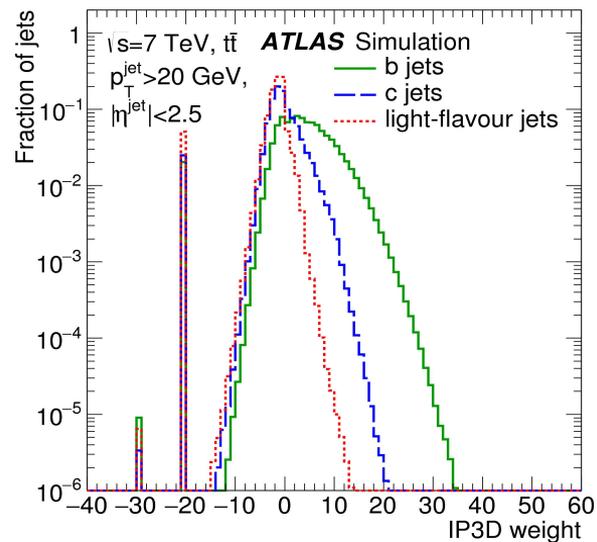
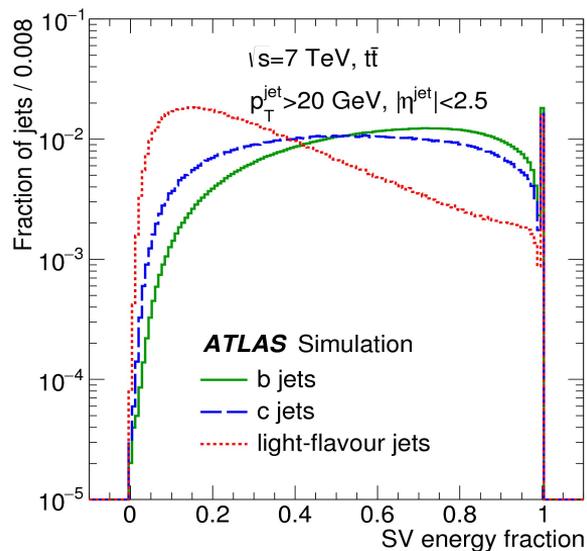
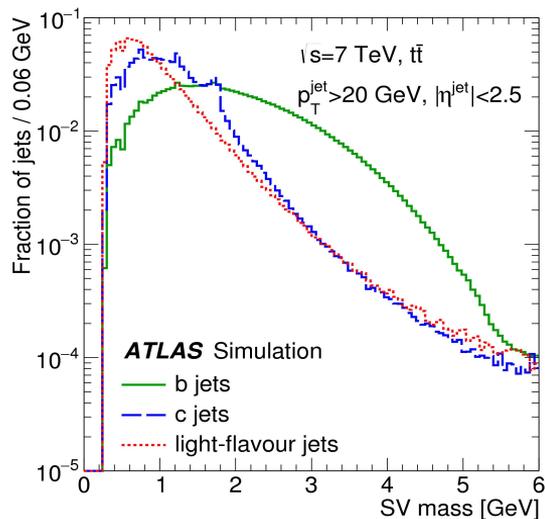
Exemple 2: Identification des jets de quarks b

- But: identification des jets b vs jets de quarks légers et de gluons
 - Plus petit taux de faux positifs possibles à efficacité donnée
 - C'est un **problème de classification**
 - Vertex secondaires de désintégration
 - Traces avec paramètres d'impact
- Problème à beaucoup de degrés de liberté
 - Nombre de traces chargées variable
 - Corrélations entre les traces (chaînes de désintégration)
 - Cas typique où l'apprentissage automatique permet de gagner beaucoup



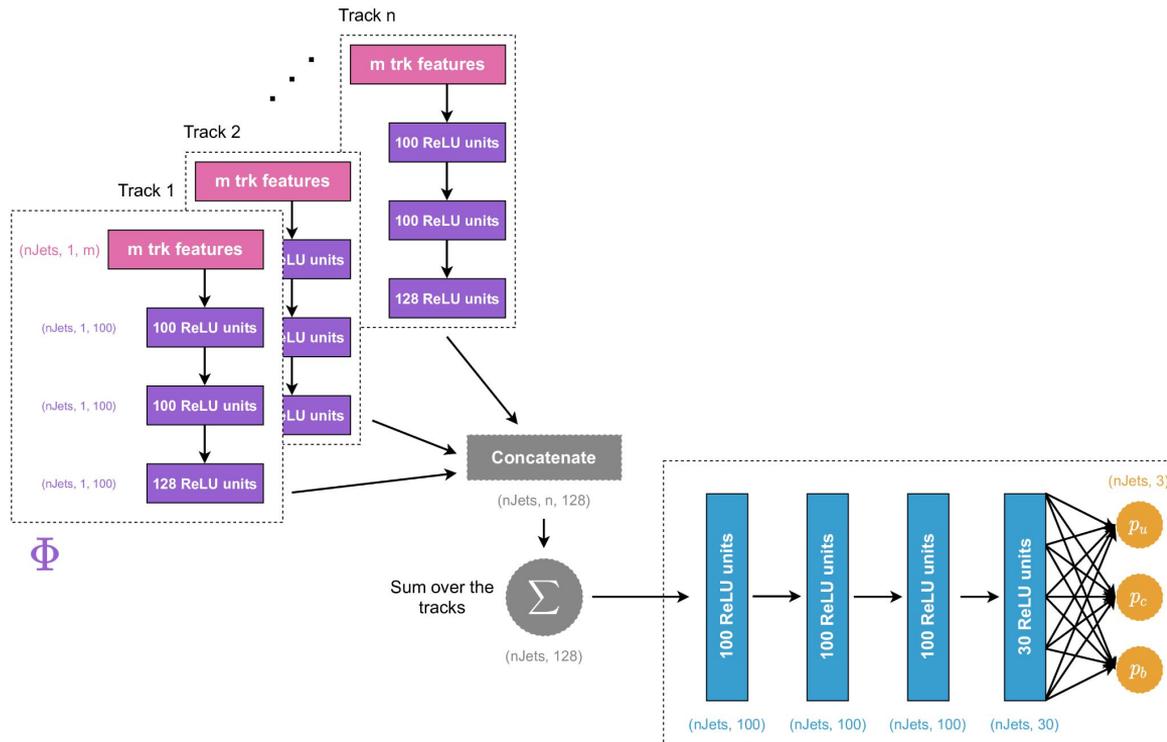
Variables discriminantes

- Création de variables discriminantes
 - Bon potentiel de séparation b-jets / jets de quarks légers
 - Possibilité de les combiner pour augmenter l'efficacité



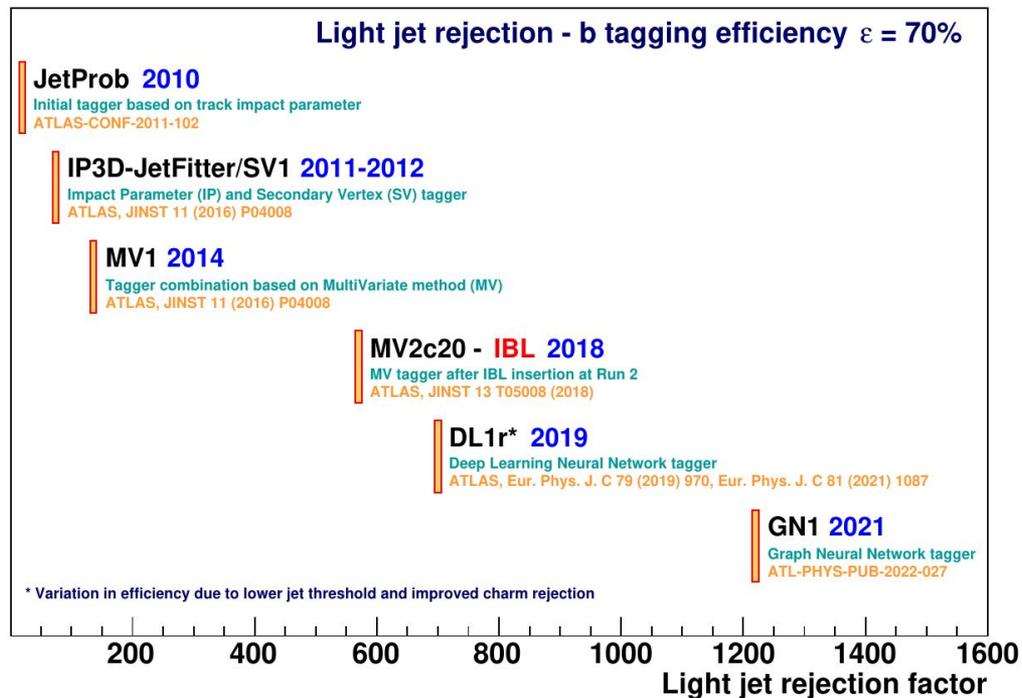
L'apprentissage automatique à la rescousse

- Exploite un grand potentiel caché dans les **corrélations entre variables**
 - Plus subtil qu'avec des variables discriminantes classiques
 - Peut reconstruire des chaînes de désintégration complètes
 - Un réseau de neurones bien construit exploite les traces au maximum
- De nos jours **employé pour de très nombreuses tâches de reconstruction** (electron ID, tau ID...)
 - Utilisation de l'état de l'art des réseaux de neurones: deep sets, GNN, transformers, etc...



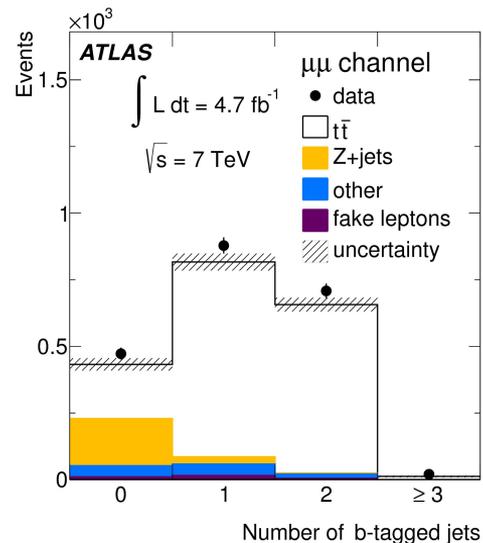
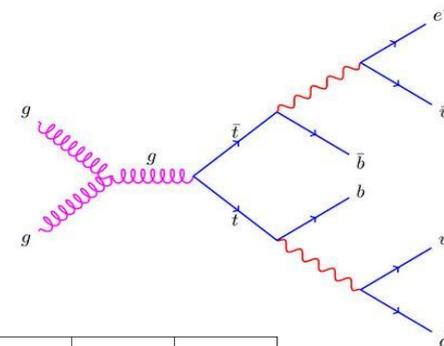
Identification des jets b: évolution

- L'un des gains en performance les plus spectaculaires en 10 ans
 - Réjection $\times 50$ à 70% d'efficacité entre 2010 et l'état de l'art

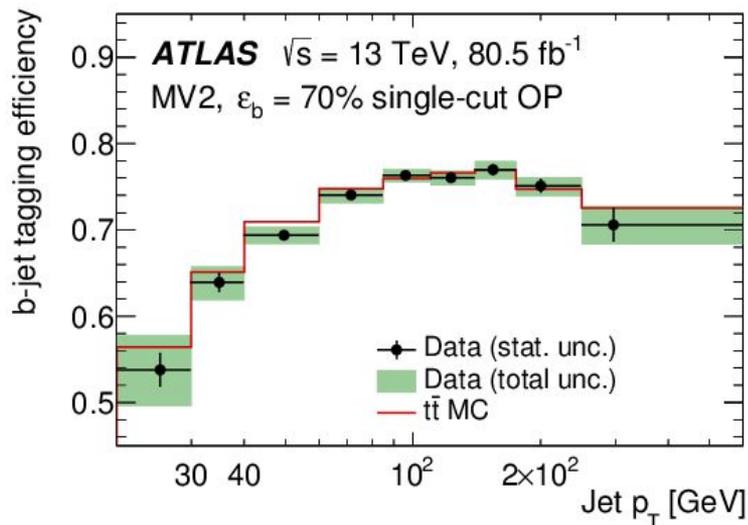


Mesure de l'efficacité d'identification

- Les algorithmes sont **optimisés à partir de la simulation**
- Mesure de la performance réelle faite **à partir des données**
 - Avec incertitude associée
- Nécessité d'un échantillon 'pur' en jets b
 - Événements avec une paire top anti-top
 - Sélection: 2 électrons ou muons, énergie transverse manquante, et 1 jet identifié b
 - On s'attend alors à trouver un autre jet dans l'événement, et à ce que celui-ci soit un jet b
- On mesure alors l'efficacité d'identification du jet additionnel
 - "Quelques" complications sont passées sous silence

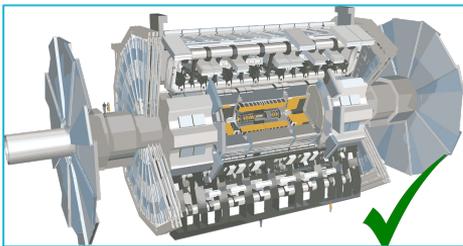


- Différence de quelques % entre simulation et données
 - Avec incertitude de quelques % également
- On corrige la simulation pour la réalisation des analyses de physique
- Plus long terme: **trouver les sources de désaccord** pour réduire l'écart
 - Meilleure simulation du détecteur
 - Meilleure génération des jets b et des événements $t\bar{t}$
- Efficacités de sélection des faux positifs également mesurées

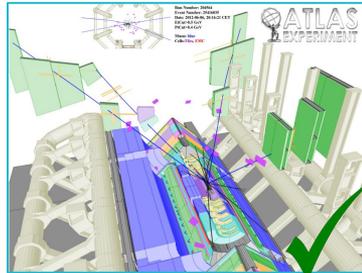


Analyses de physique

Rappel: Cheminement des données



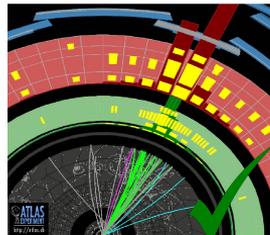
Détecteur



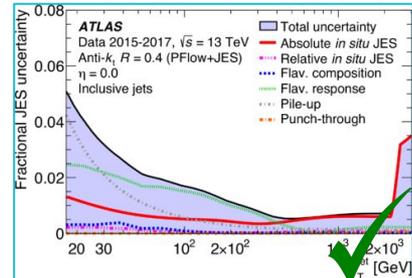
Simulation



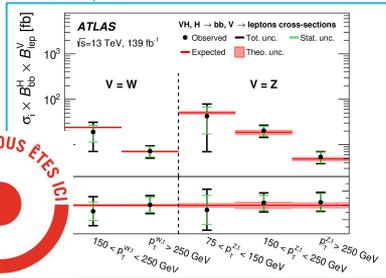
Trigger / Acquisition



Calibration
Qualité des données
Reconstruction



Performance de la reconstruction



Analyse de données

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector

ATLAS Collaboration*

Abstract: Measurements of the Standard Model Higgs boson decaying into a $b\bar{b}$ pair and produced in association with a W or Z boson decaying into leptons, using proton-proton collision data collected between 2015 and 2018 by the ATLAS detector, are presented. The measurements use collision events produced by the Large Hadron Collider at a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 36.1 fb $^{-1}$. The production of a Higgs boson in association with a W or Z boson is established with observed statistical significance of 4.0 (3.1) and 3.5 (3.5) standard deviations, respectively. Cross-sections of associated production of a Higgs boson decaying into bottom quark pairs with an electron, muon, W or Z , decaying into leptons are measured as a function of the gauge boson transverse momentum in kinematic fiducial volumes. The cross-section measurements are all consistent with the Standard Model expectations, and the total uncertainties vary from 30% in the high gauge boson transverse momentum regions to 80% in the low regions. Limits are subsequently set on the parameters of effective Lagrangian operators in modifications of the WZ and ZZ propagators as well as the Higgs boson decay into $b\bar{b}$.

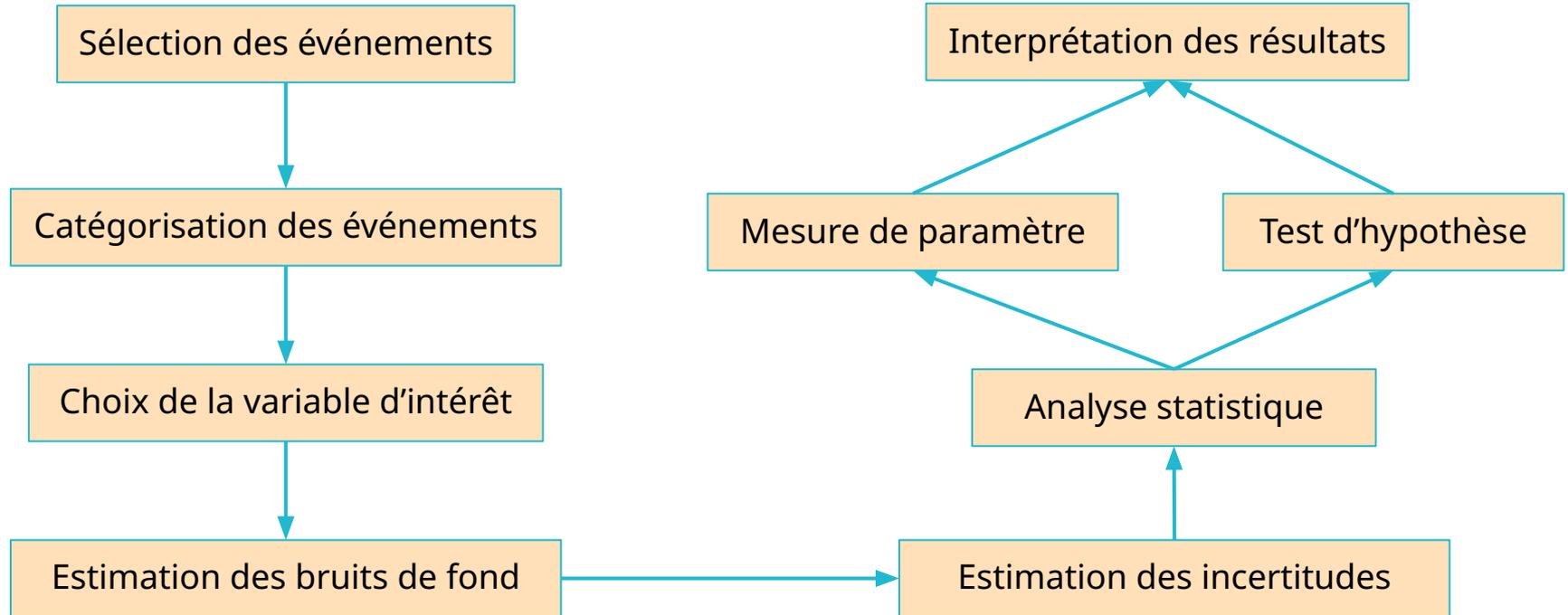
1 Introduction

The Higgs boson [1–6] was discovered in 2012 by the ATLAS and CMS Collaborations [7, 8] with a mass of approximately 125 GeV from the analysis of proton-proton collisions produced by the Large Hadron Collider (LHC) [9]. Since then, the analysis of data collected at center-of-mass energies of 7 TeV, 8 TeV and 13 TeV in Run 1 and of the LHC

Publication

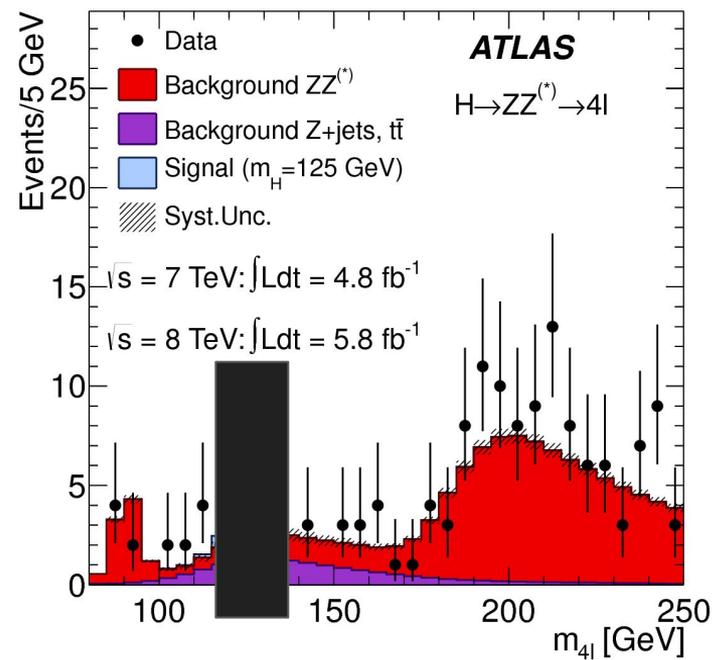
Étapes d'une analyse de physique

Processus itératif



Les analyses en aveugle

- Risques de biais
 - La réalisation d'une analyse est un processus itératif
 - Risque de **biaiser** l'analyse en fonction des résultats pour "améliorer" celui-ci
 - Même **inconsciemment** !
- Analyse en aveugle
 - Analyse optimisée sur **simulations**
 - Données sensibles **gardées secrètes** tant que l'optimisation n'est pas terminée et que l'analyse n'est pas validée
 - On ne dévoile le résultat qu'à la fin !



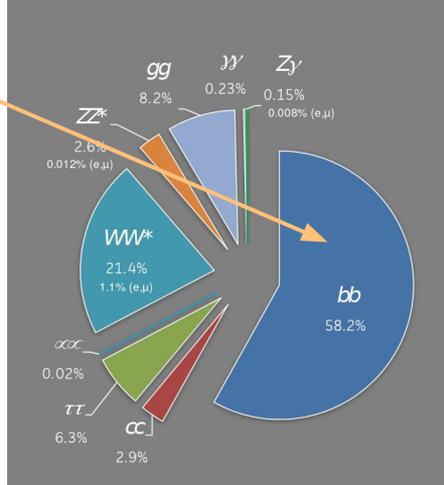
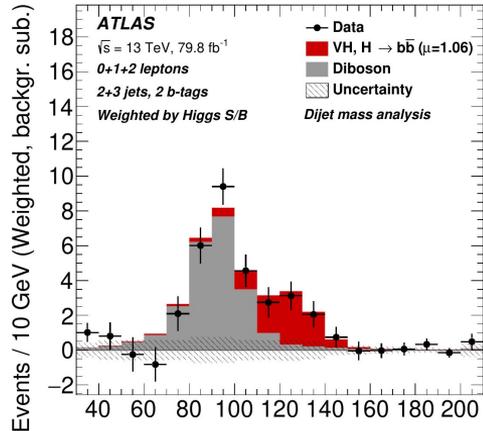
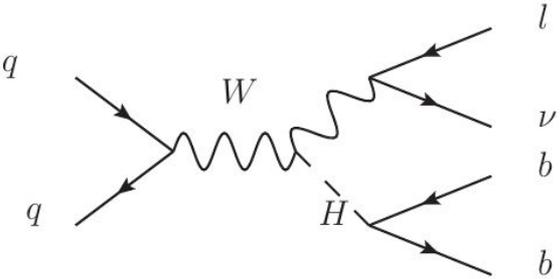
- Framework et bibliothèques pour l'analyse de données, créés par le CERN en 1994
 - Gratuit et libre
 - Nombreux contributeurs internationaux
- Programmation objet en C++, interpréteur C++, intégration python
- Toutes les briques pour nos applications
 - Stockage et lecture de données (ntuples)
 - Analyse des données
 - Histogrammes et visualisation des données
 - Analyse statistique, fits
 - Manipulation de quadri-vecteurs
 - Algèbre linéaire
 - ...



Mesure d'un processus

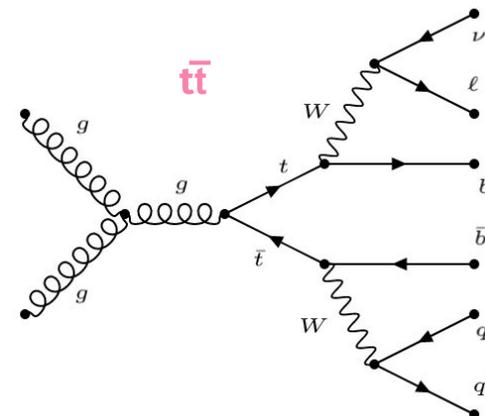
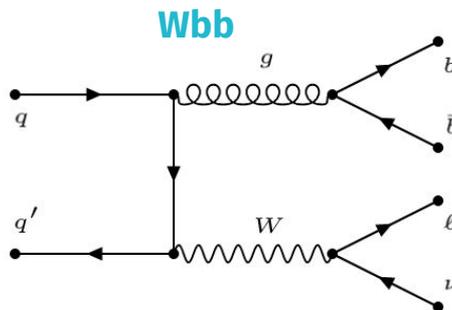
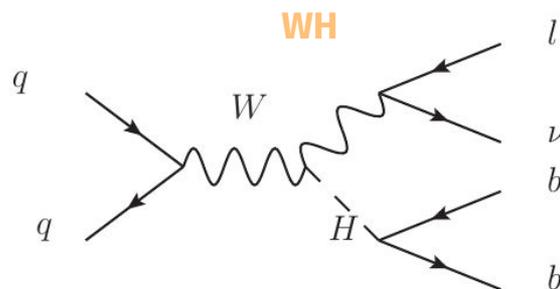
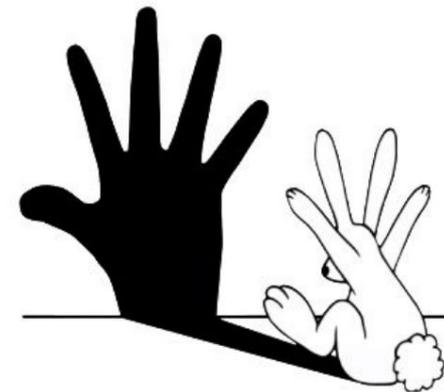
Exemple d'analyse: VH, H→bb

- Exemple de ce qu'on peut faire 12 ans après la découverte du boson de Higgs
 - Programme de recherche: mesure de l'ensemble des propriétés du Higgs
- Analyse beaucoup plus difficile que H→γγ et H→4ℓ
 - Mais permet d'étudier la désintégration H→bb
 - Observation en 2018



Signal et bruits de fond

- **Bruits de fond** physiques: autres processus microscopiques qui donnent le **même état final reconstruit** que le signal étudié:
 - **Signal étudié**: le processus qu'on veut mesurer
 - **Bruit de fond irréductible**: même état final en terme de particules
 - **Bruit de fond réductible**: contenu en particules différent
 - Peut être diminué par une optimisation de la reconstruction



Taux des processus physiques

Pour un processus donné:

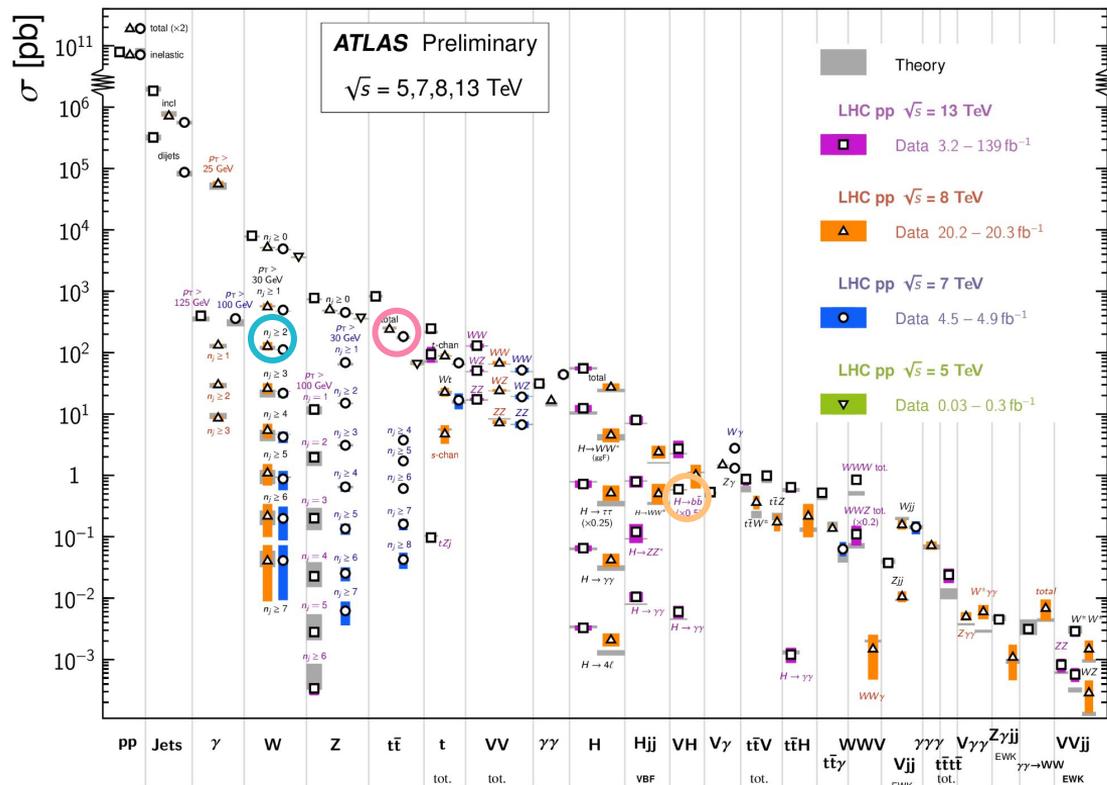
$$N = \sigma \cdot \mathcal{L}$$

section efficace luminosité

Les bruits de fond ont parfois une section efficace **bien supérieure** à celle du signal étudié

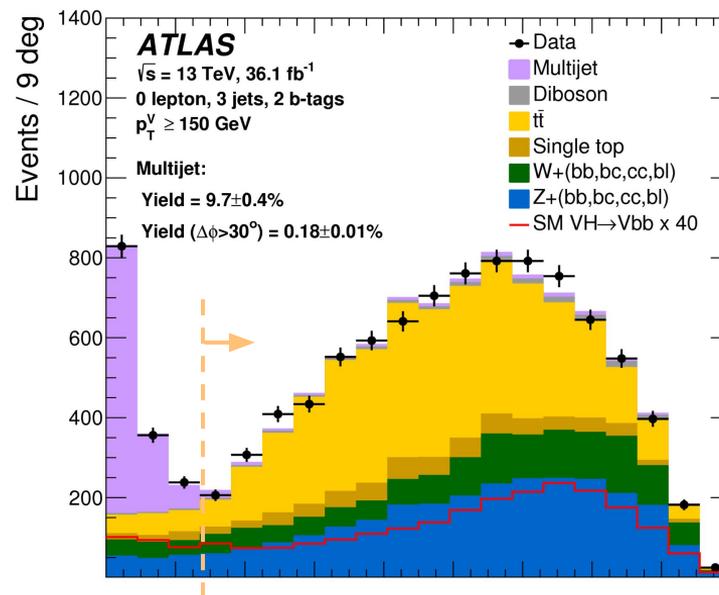
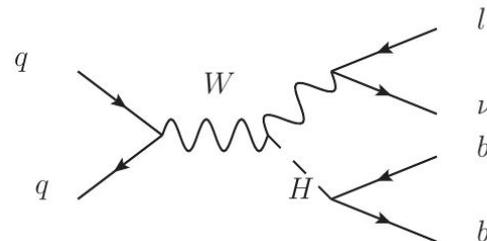
Standard Model Production Cross Section Measurements

Status: July 2021



Sélection des événements

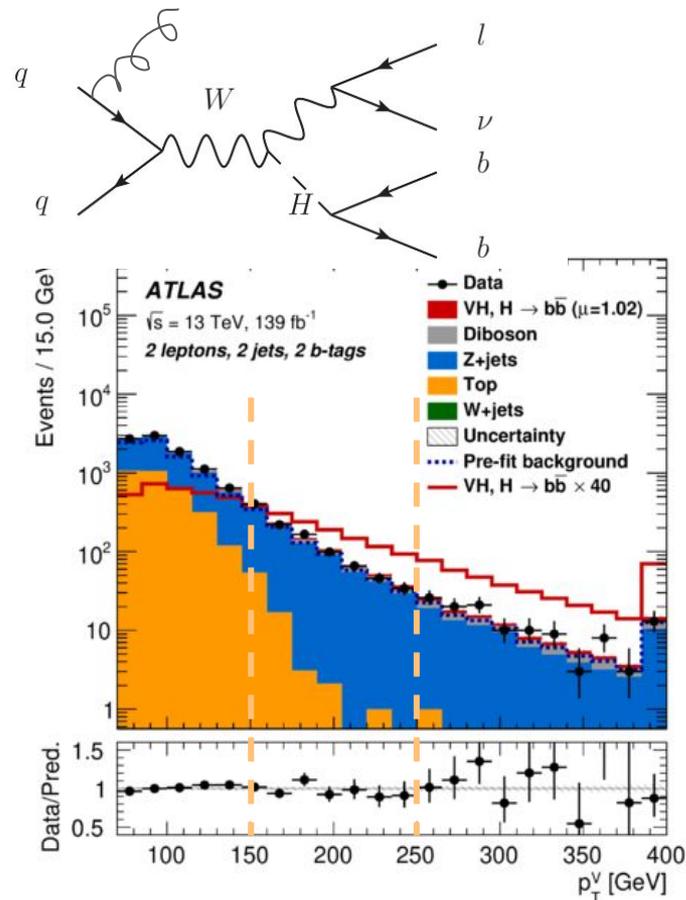
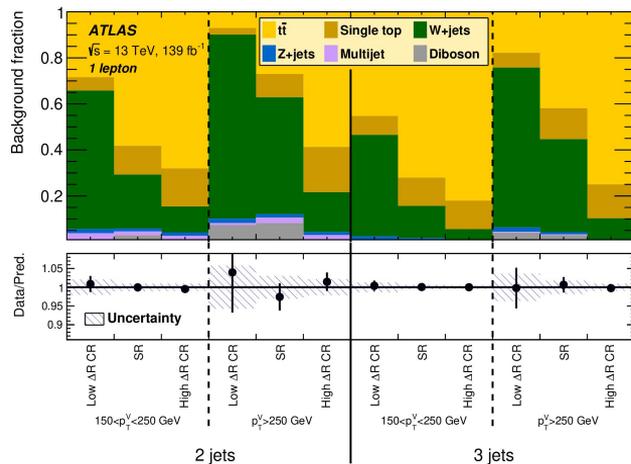
- Optimisation des données à étudier
 - Efficacité sur le signal: ↗
 - Quantité de bruit de fond: ↘
- Sélections ?
 - Lignes de déclenchement
 - Nombre et type des objets
 - 1 électron ou muon, énergie transverse manquante, 2 jets b
 - Énergie des objets
 - Jets $p_T > 20$ GeV, ...
 - Position des objets
 - Lepton $|\eta| < 2.5$
 - Cinématique plus avancée
 - $m_T^W > 30$ GeV
- Selon les analyses, le rapport S/B va de O(1%) voire O(0.1%) à >100



Catégorisation des événements

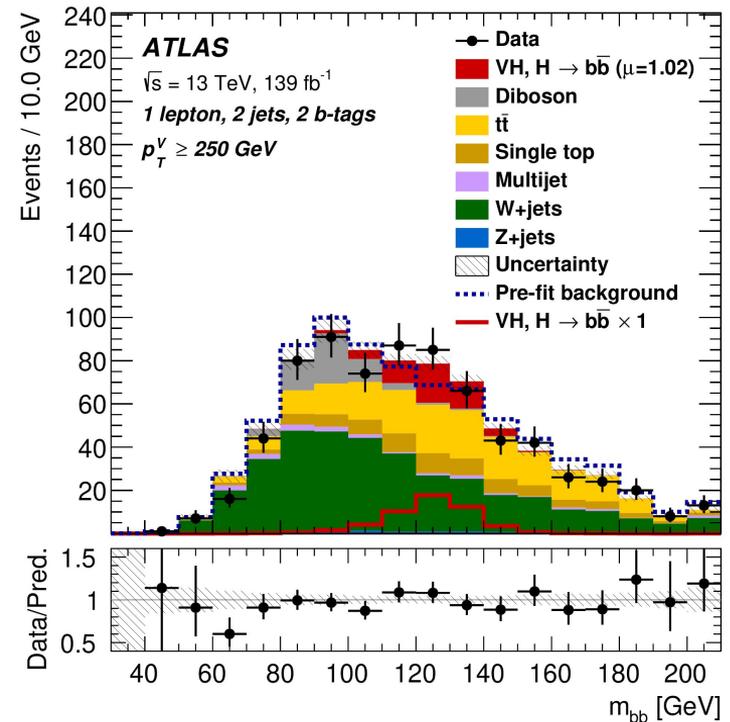
Plusieurs raisons pour **classer les événements** en plusieurs catégories

- Événements de topologies différentes
- Sections efficaces différentielles
- Séparation **régions de signal / régions de contrôle**



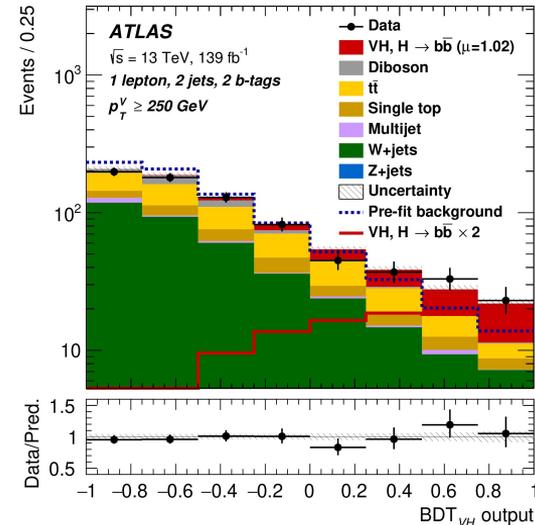
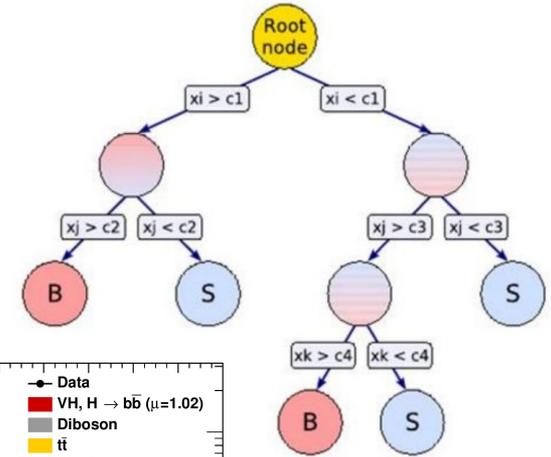
Variable d'analyse et méthodes multivariées

- Mesure d'un signal: ajustement de la distribution d'une variable par le signal et les bruits de fond
 - Maximum de vraisemblance
 - Incertitude réduite si S/B est grand
- Choix de la variable: meilleure séparation S / B
 - Masse invariante...
- Nouvelles variables construites pour maximiser la séparation
 - Méthodes multivariées pour utiliser toute la cinématique



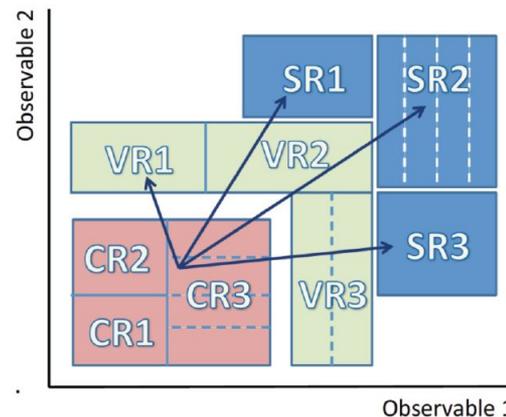
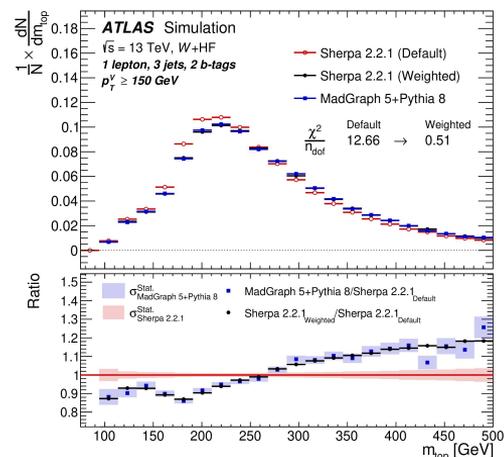
Les arbres de décision boostés

- Analyse de physique: **$O(10)$** variables cinématiques intéressantes
 - Deep Neural Nets non nécessaires
 - Arbres de décision boostés
- Arbres de décision boostés (BDT)
 - Existent depuis ~40 ans
 - Collection de séries de sélections
 - Classification: maximise la séparation S / B
 - Entraînement / optimisation plus simple que les réseaux de neurones
- Gain **~30% performance** par rapport à une variable cinématique



Estimation des bruits de fond

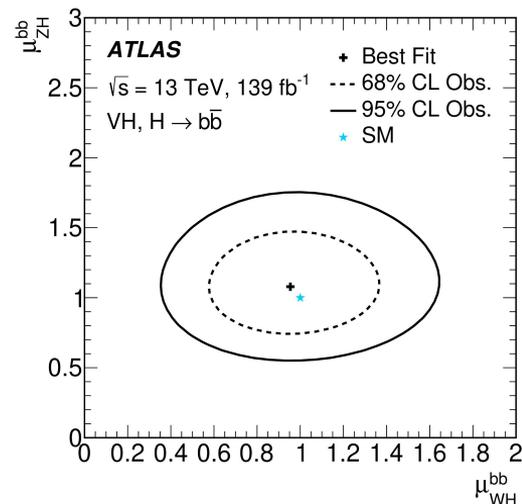
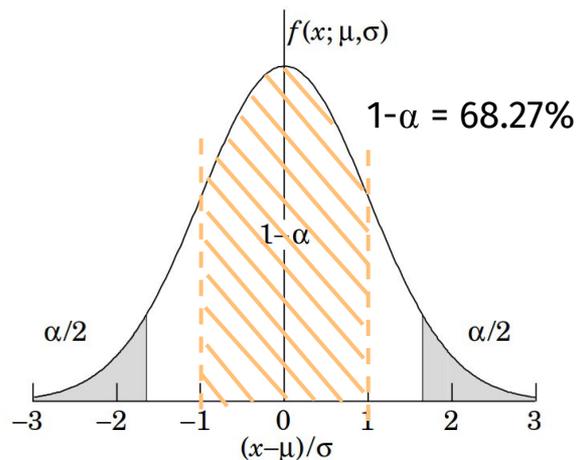
- Mesure d'un signal: $S \sim N - B$
 - Il faut estimer B **précisément** pour diminuer l'erreur sur S
- Estimations à partir des simulations
 - Donnent une bonne indication
 - Difficile d'estimer l'incertitude associée (comparaison de différents générateurs)
 - Incertitudes souvent grandes
- Les données: le meilleur Monte-Carlo
 - Nombreuses techniques utilisant les données ou données+simulation
 - Reposent sur des **régions / échantillons de contrôle**
 - Incertitudes liées à l'extrapolation entre régions de contrôle et région de signal
 - **Réduction des incertitudes** / simulations seules



Détour: qu'est-ce qu'une barre d'erreur ?

- Buts d'une mesure

- Publier la **valeur** d'une observable (section efficace, masse...)
- Avec une **incertitude**
- Matrices de corrélation si plusieurs mesures simultanées



- Barre d'erreur

- **Intervalle de confiance** à 68.3%
- Définition (fréquentiste):
 - Si on répète la mesure N fois, alors dans 68.3% des cas la vraie valeur de l'observable sera contenue dans l'intervalle
- Correspond à **1 σ** sur une gaussienne

L'incertitude statistique

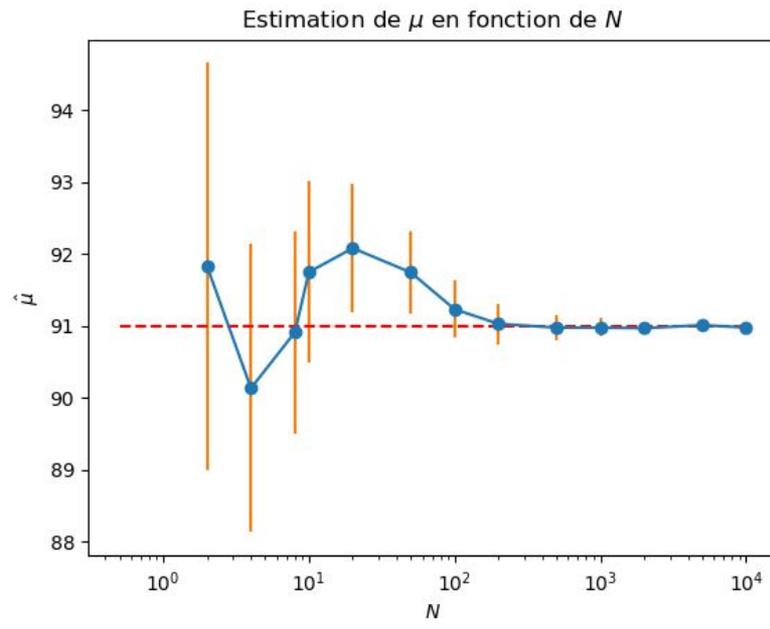
- Estimation de la valeur centrale d'une gaussienne
- On a vu l'estimateur de la valeur centrale:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

- Incertitude associée:

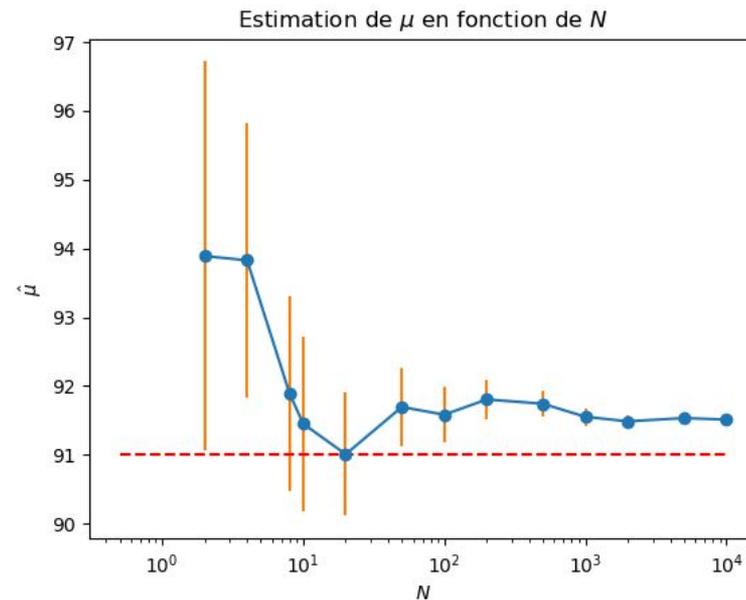
$$\sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{N}}$$

Incertainde statistique: décroît comme la **racine carrée** de la quantité de données



Les incertitudes systématiques

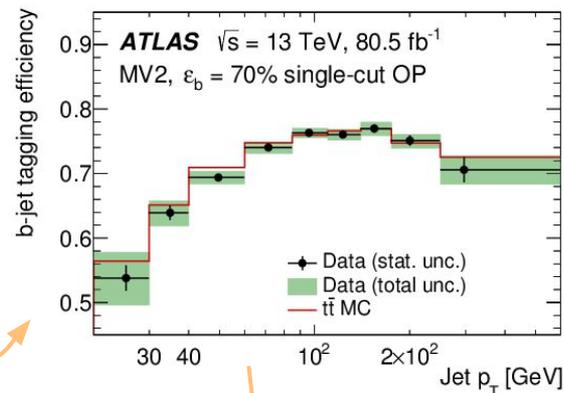
- Les ingrédients de notre mesure peuvent comporter des **biais**
- Ces biais vont se propager jusqu'au résultat
 - Exemple: mauvaise calibration de l'énergie des électrons de 0.5%
 - Histogramme des mesures décalé
 - Estimation de valeur centrale de la gaussienne déplacée de ~ 0.5 GeV
- Ils ne se **réduisent pas** quand on ajoute des données
- Incertitudes systématiques:
 - Lister toutes les **sources indépendantes** de biais
 - Estimer la taille possible de ces biais
 - Ajouter leur effet sur la mesure à la barre d'erreur



Incertitudes systématiques expérimentales

Nombreuses sources:

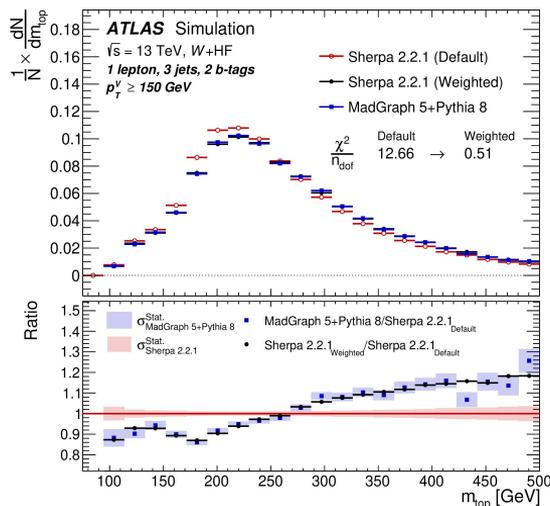
- Calibration des objets:
 - Calibration en énergie, linéarité
 - Connaissance de la résolution
 - Efficacité
 - Subtilité: Incertitudes de calibration diminuent avec l'accumulation des données
- Luminosité délivrée
 - $N = \sigma \cdot \mathcal{L}$ donc $\sigma = N/\mathcal{L}$



Source of uncertainty	VH	σ_μ WH	ZH
Total	0.177	0.260	0.240
Experimental uncertainties			
Jets	0.043	0.050	0.057
E_T^{miss}	0.015	0.045	0.013
Leptons	0.004	0.015	0.005
b-tagging	b-jets	0.045	0.025
	c-jets	0.035	0.068
	light-flavour jets	0.009	0.004
Pile-up	0.003	0.002	0.007
Luminosity	0.016	0.016	0.016

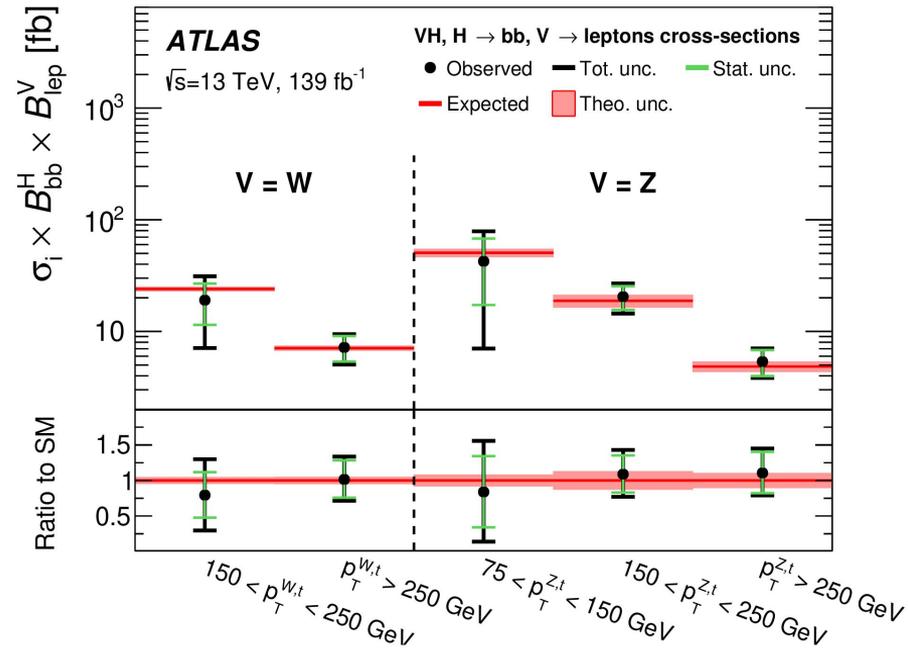
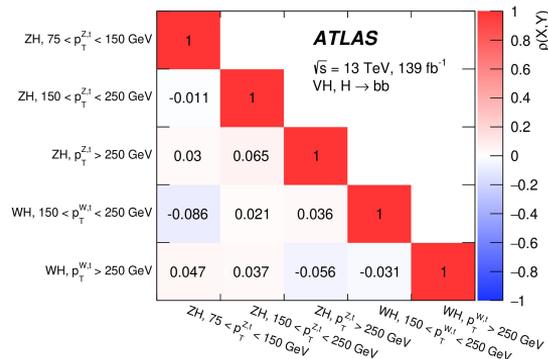
Incertitudes systématiques sur les bruits de fond

- À estimer pour chaque type de bruit de fond
 - Comparaisons de prédictions théoriques
 - Contraintes par les régions de contrôle
 - Effet propagé sur la mesure
- Incertitudes sur l'acceptance du signal
- Taille des jeux de données simulés

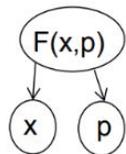


Source of uncertainty	σ_μ		
	VH	WH	ZH
Total	0.177	0.260	0.240
Theoretical and modelling uncertainties			
Signal	0.072	0.060	0.107
$Z + \text{jets}$	0.032	0.013	0.059
$W + \text{jets}$	0.040	0.079	0.009
$t\bar{t}$	0.021	0.046	0.029
Single top quark	0.019	0.048	0.015
Diboson	0.033	0.033	0.039
Multi-jet	0.005	0.017	0.005
MC statistical	0.031	0.055	0.038

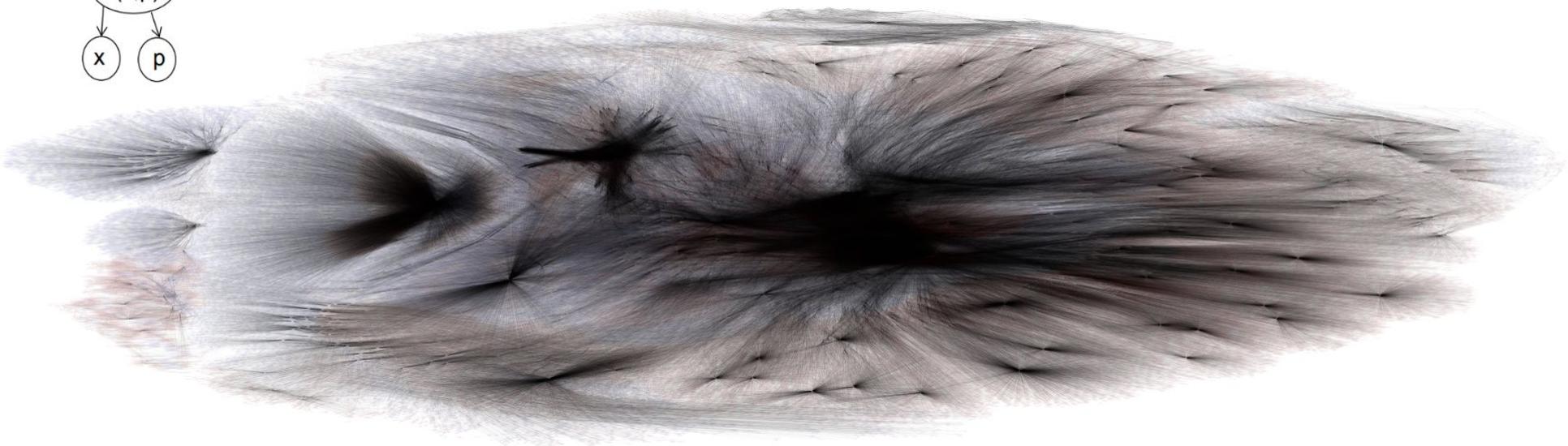
- Extraction des résultats par maximum de vraisemblance
 - Incluant les incertitudes systématiques
- Section efficace différentielle des processus WH et ZH
 - Avec matrice de corrélation
 - Incertitudes statistiques ~ systématiques
 - En bon accord avec le modèle standard



Analyse statistique: un cas extrême



Fonction de vraisemblance des mesures combinées du boson de Higgs

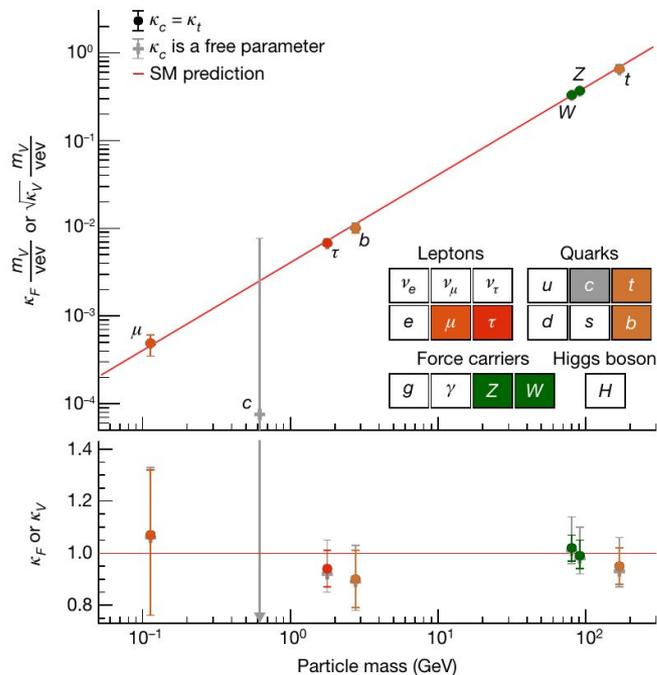


- **Combinaison des résultats sur le boson de Higgs**
 - Mesure de 36 sections efficaces simultanées
 - $O(5000)$ points de mesure
 - 2600 systématiques

Interprétation des résultats

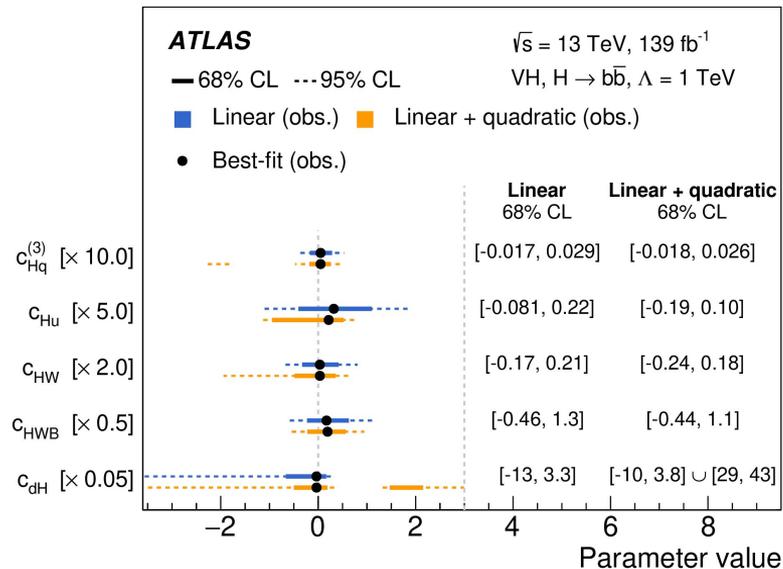
- **Comparaison avec la théorie**

- Estimation des paramètres du Modèle Standard



- **Contraintes sur de la nouvelle physique**

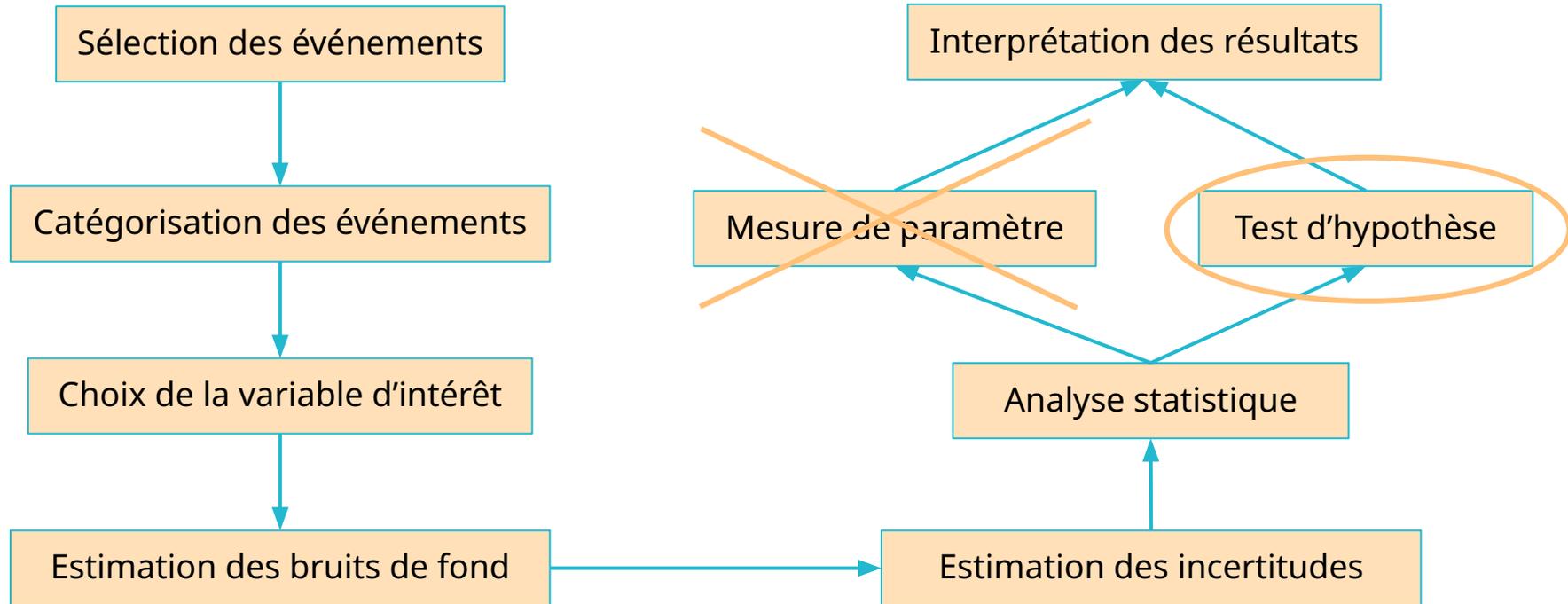
- Interprétation dans différents modèles



Recherche d'une nouvelle particule

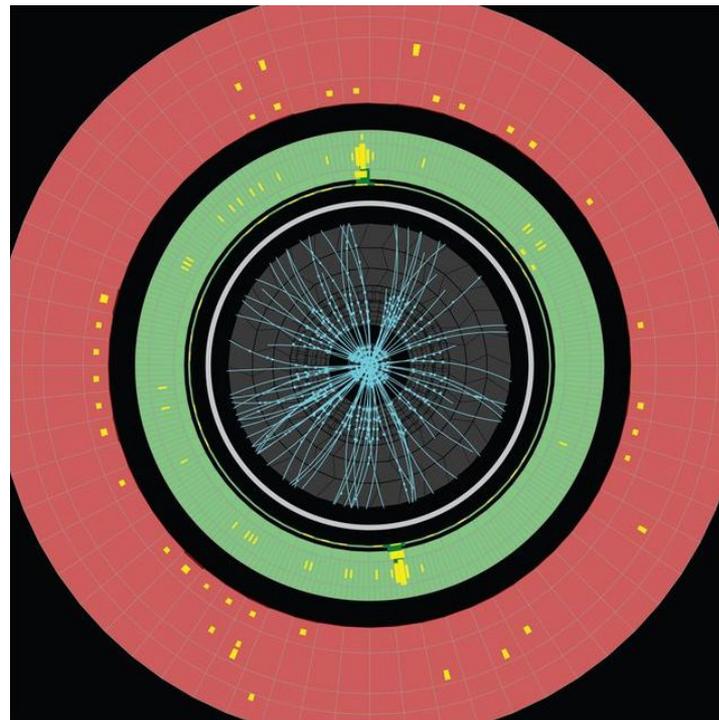
Étapes de recherche d'une nouvelle particule

Presque pareil qu'une mesure



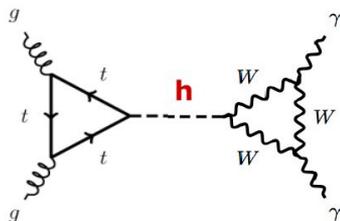
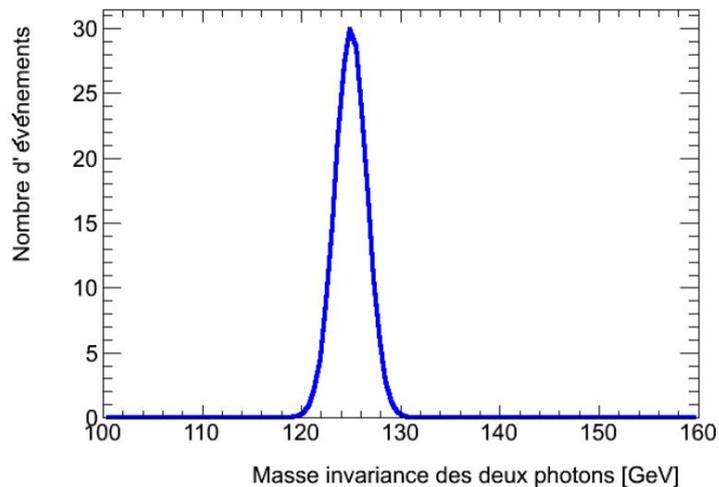
Exemple: observation du Higgs en $H \rightarrow \gamma\gamma$

- Un des canaux “en or” pour la recherche du boson de Higgs
- Caractéristiques:
 - État final simple: 2 photons
 - Leur masse invariante doit piquer à m_H
 - Bruit de fond important
 - Mais excellente résolution (1.5 GeV)
- Retour en 2012: on ne sait pas
 - Si le boson de Higgs existe
 - Et si jamais il existe, on ne connaît pas sa masse
- Recherche sur une large plage

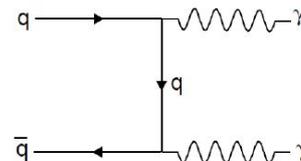
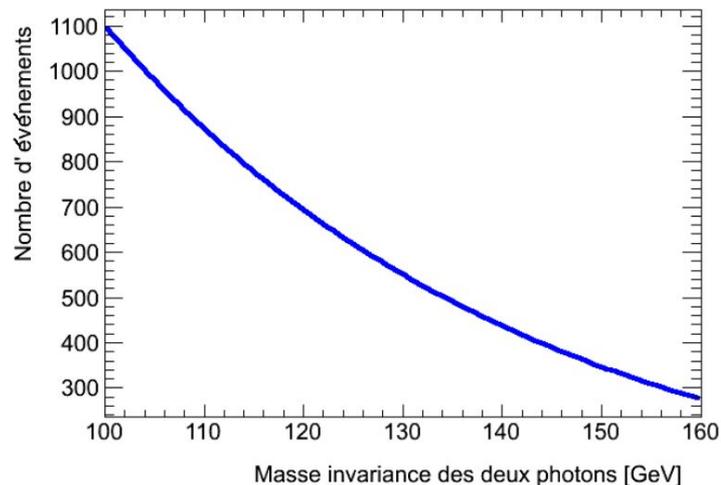


Signal et bruits de fonds

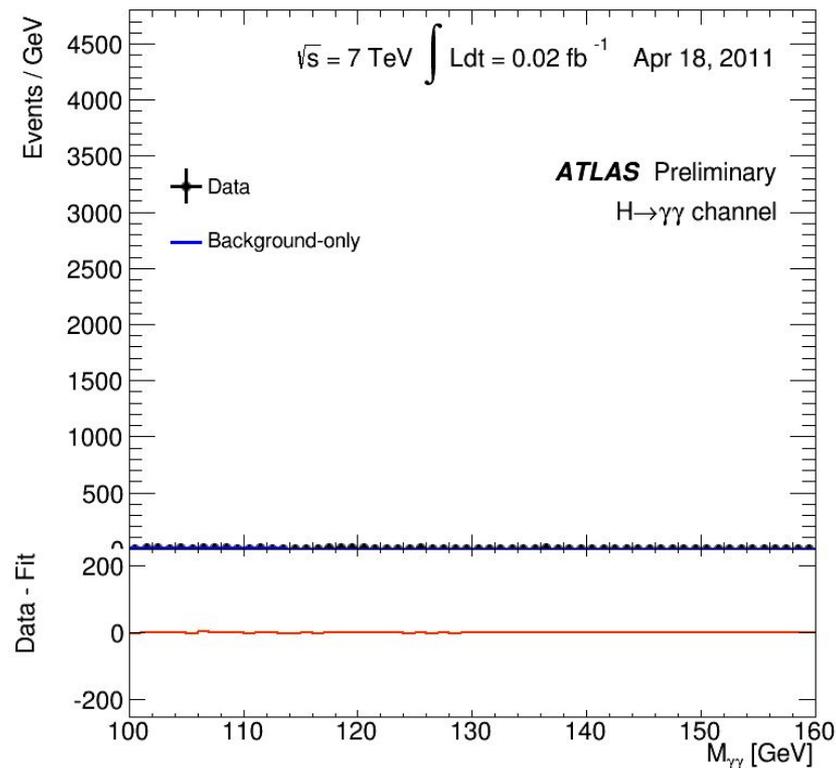
Higgs ($m_H = 125$ GeV)



Bruit de fond



- L'analyse est prête: on prend des données et on attend
- On "voit" une bosse à 125 GeV:
 - Une découverte ou non ?

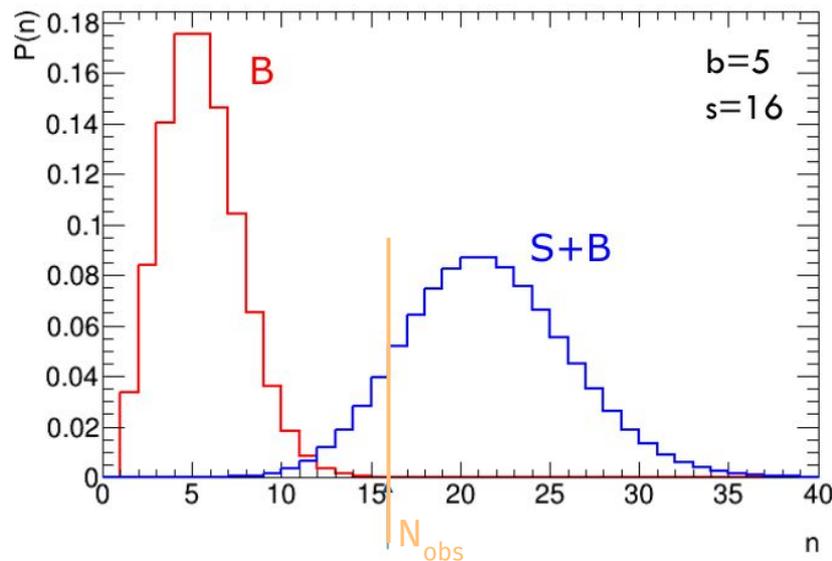


Test d'hypothèse

- **Hypothèse H_0** : Modèle standard **sans** boson de Higgs (B)
- **Hypothèse H_1** : Modèle standard **avec** boson de Higgs (S+B)
- **Test d'hypothèse**: conclure quelque chose sur H_0 ou H_1 à partir de la fonction de vraisemblance
 - La question doit être bien posée

- **Exemple: expérience de comptage**

- H_0 : Poisson $\lambda = 5$
- H_1 : Poisson $\lambda = 5+16 = 21$
- Observation: $N_{obs} = 16$
- Que peut-on conclure ?



Signification statistique d'un excès

- Prouver l'existence d'une nouvelle particule c'est démontrer que les données sont **incompatibles** avec l'hypothèse où elle **n'existe pas** (H_0)

- **Quantification: valeur p_0**

- $p_0 = P(n \geq N_{\text{obs}} | H_0)$

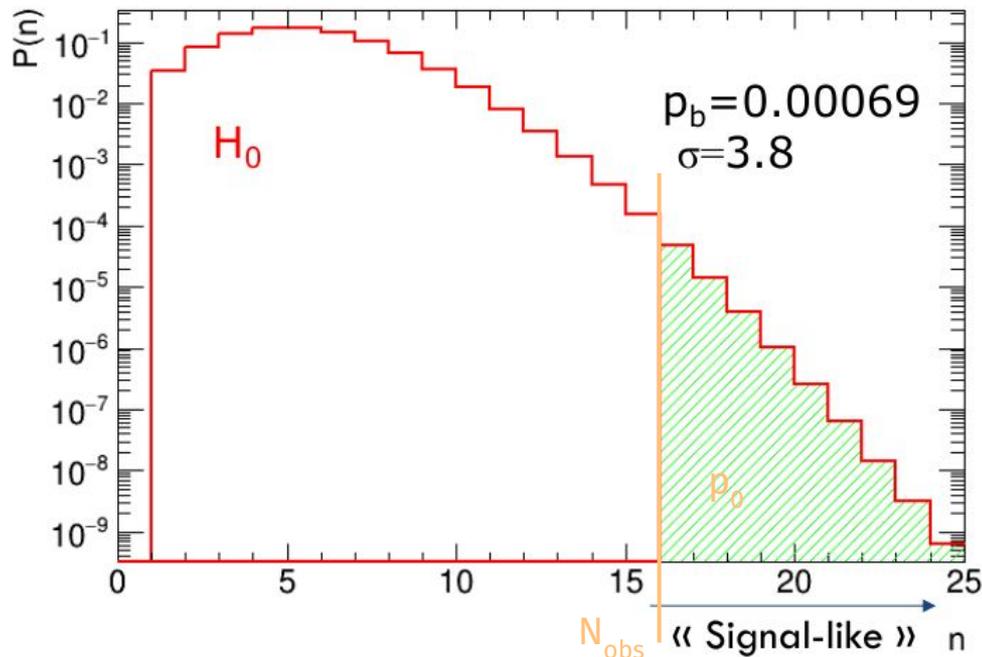
- **Convention: observation si $p_0 < 2.8 \times 10^{-7}$**

- Correspond à 5 déviations standard pour une distrib gaussienne: "5 σ "

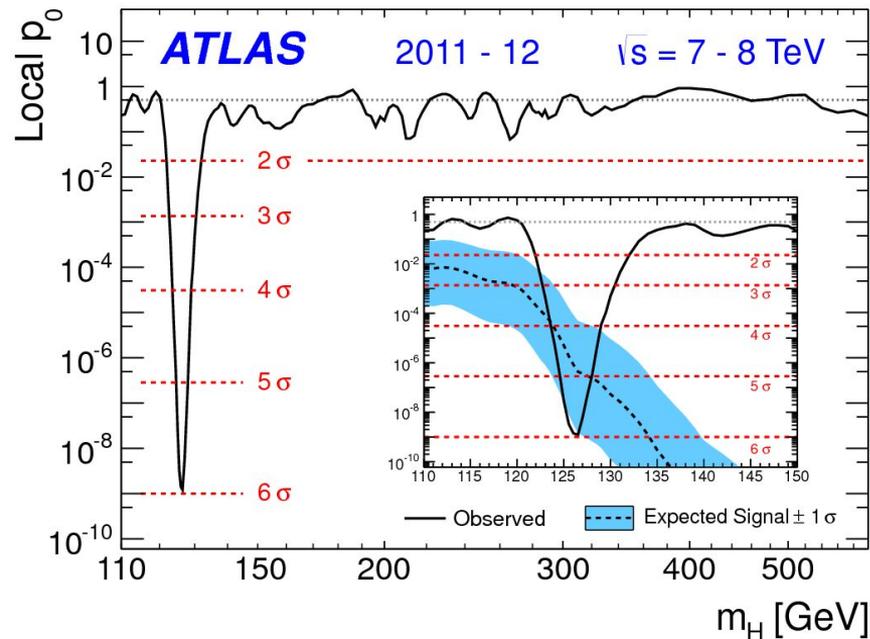
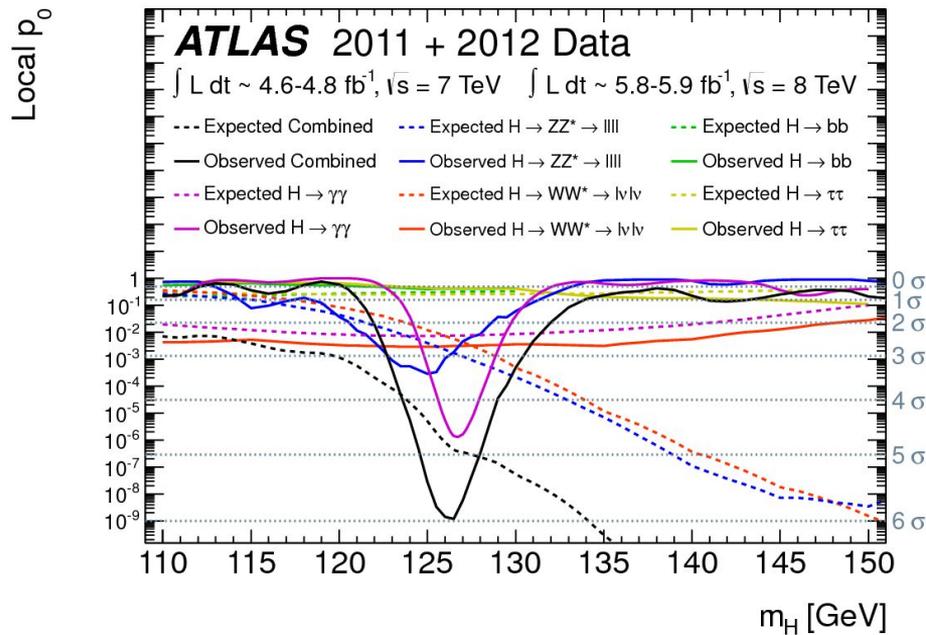
- **Signification statistique**

- Expérience de comptage

$$Z_0 = \sqrt{2 \left(N \ln \frac{N}{B} + B - N \right)} \sim \frac{N - B}{\sqrt{B}}$$

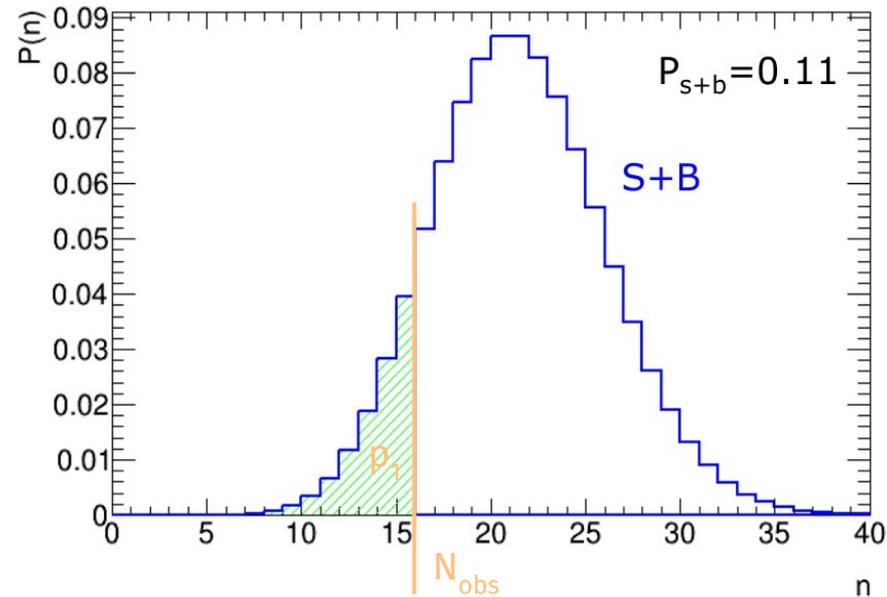


Observation du boson de Higgs



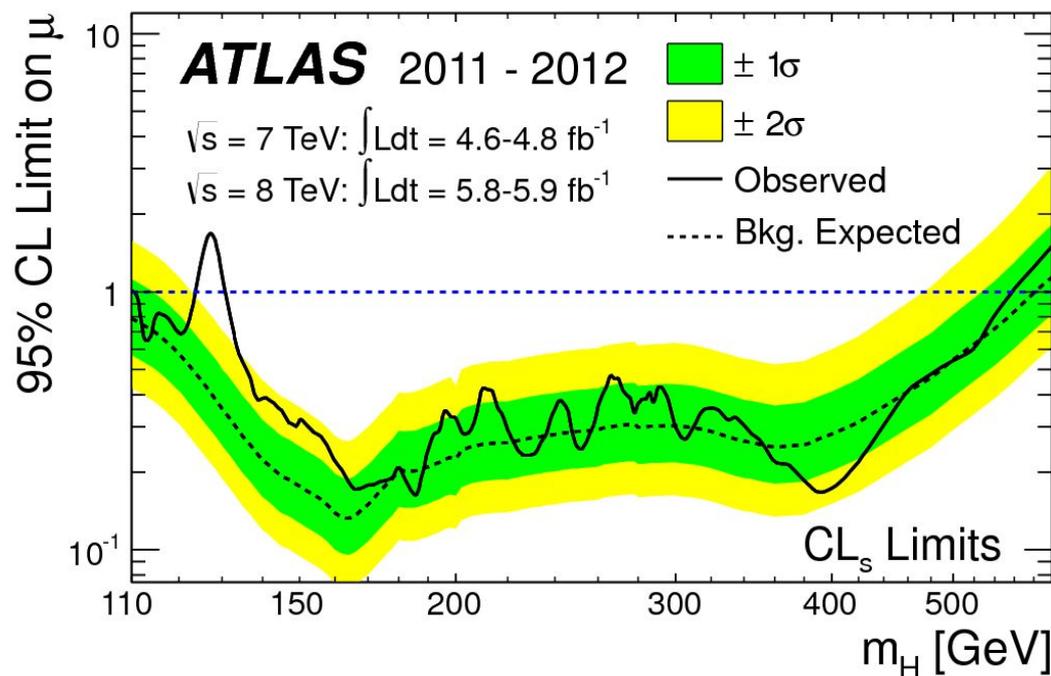
Pas d'excès: limites !

- Rejet d'une hypothèse de signal:
 - H_1 rejeté si $p_1 = P(n \leq N_{\text{obs}} \mid H_1) < 0.05$
- Limite à 95% de niveau de confiance:
 - On **scanne S** (\sim section efficace)
 - Jusqu'à trouver **$p_1 = 0.05$**
 - Cette valeur de S est la valeur de **limite à 95% de niveau de confiance**
 - Définition fréquentiste:
Valeur de S telle que si on répétait la mesure un grand nombre de fois, alors le nombre d'événements observés serait $< N_{\text{obs}}$ dans 5% des cas.



Limites sur le boson de Higgs

Section efficace du Higgs exclue pour chaque hypothèse de masse

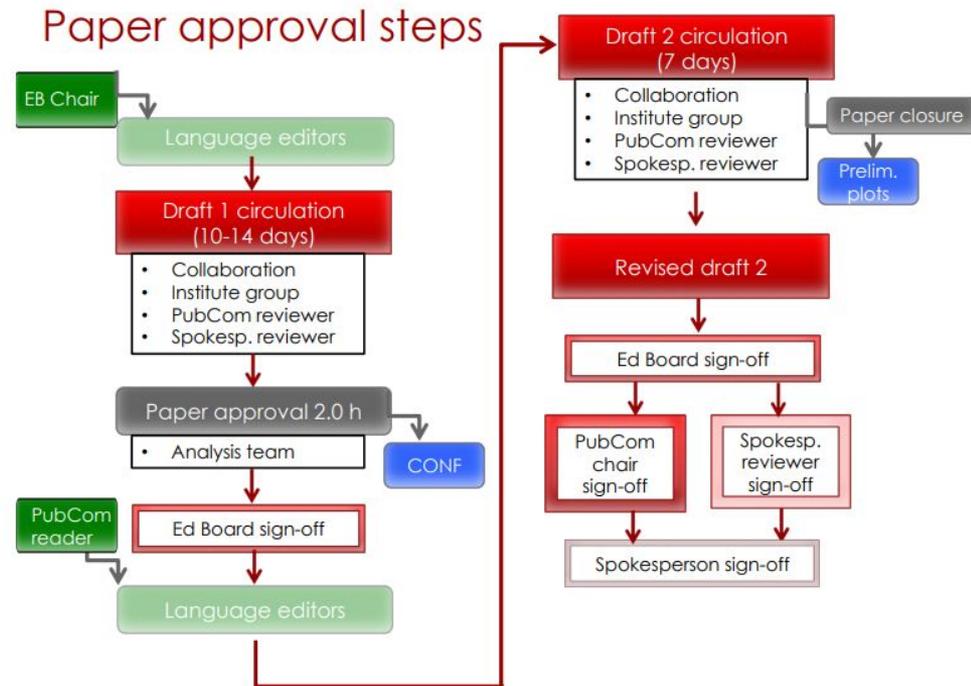


- En 2024:
 - Inutile pour le Higgs :-)
 - Employé dans toutes les analyses de recherche de nouvelle physique

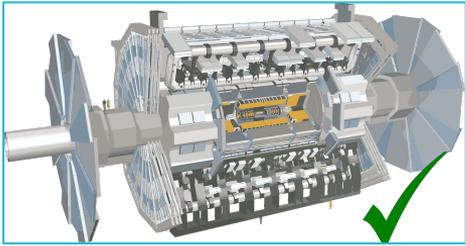
Publication des résultats

>3000 collaborateurs: procédures d'approbation par la collaboration

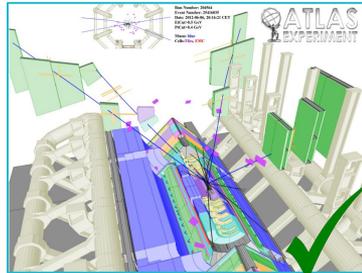
- Convaincre ses collègues:
 - Note interne: O(10) fois plus long que la publi
 - Unblinding approval, approval
- Rédiger la publication
 - Editorial board
 - Relecture et commentaires par la collaboration



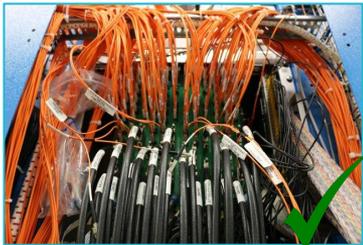
Conclusion



Détecteur



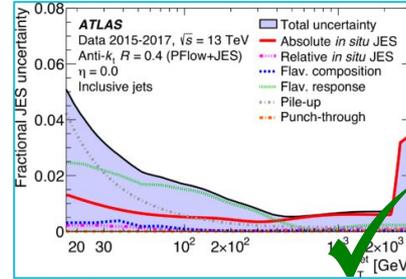
Simulation



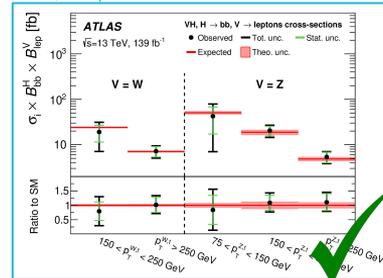
Trigger / Acquisition



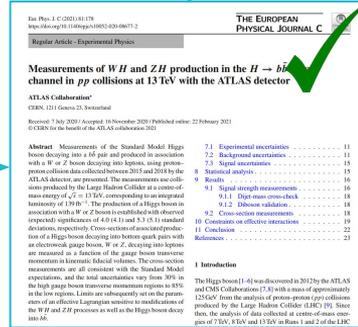
Calibration
Qualité des données
Reconstruction



Performance de la reconstruction



Analyse de données



Publication

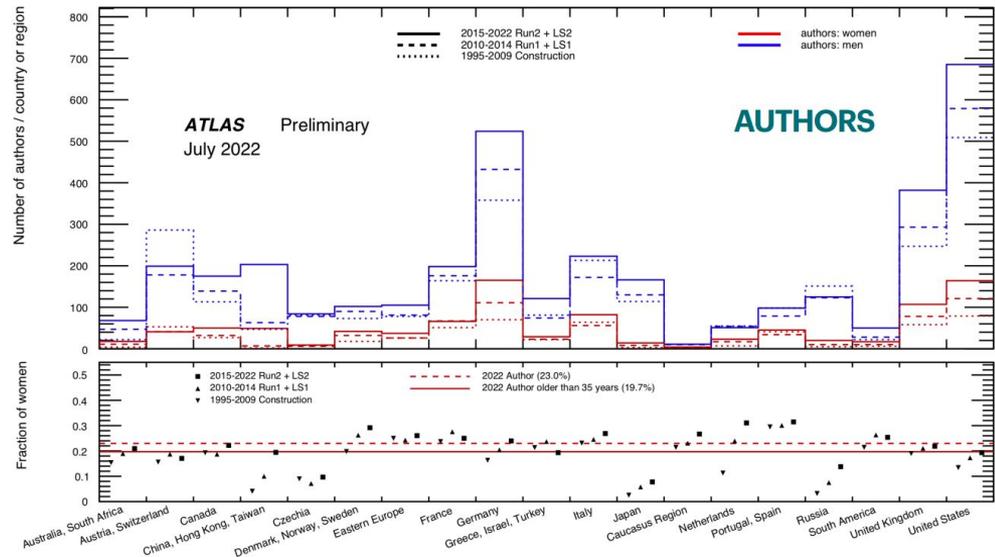
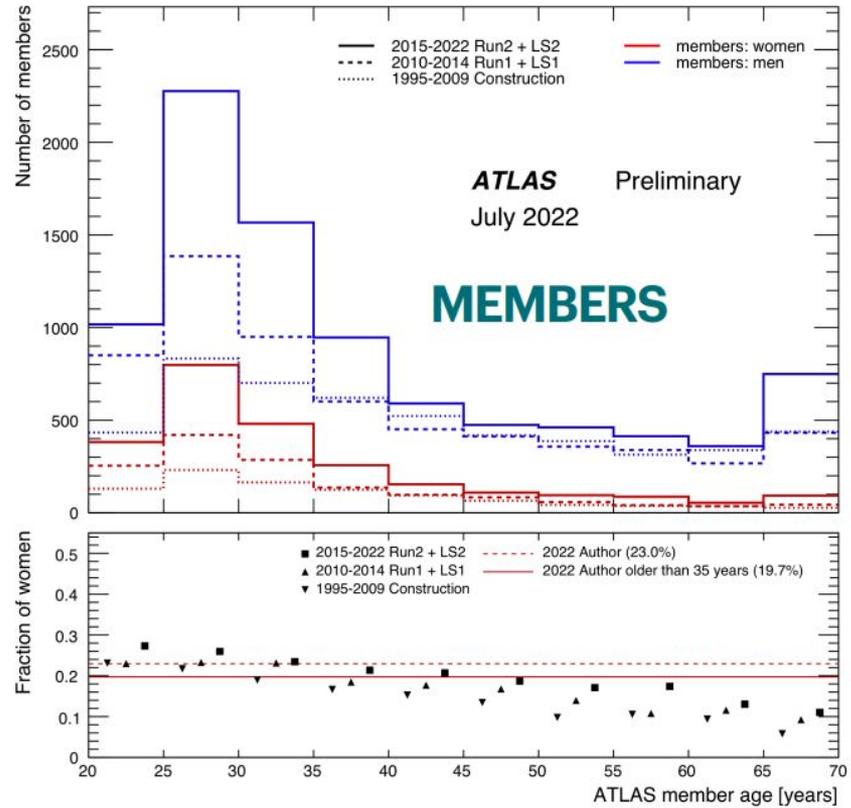
La Collaboration ATLAS

La collaboration ATLAS

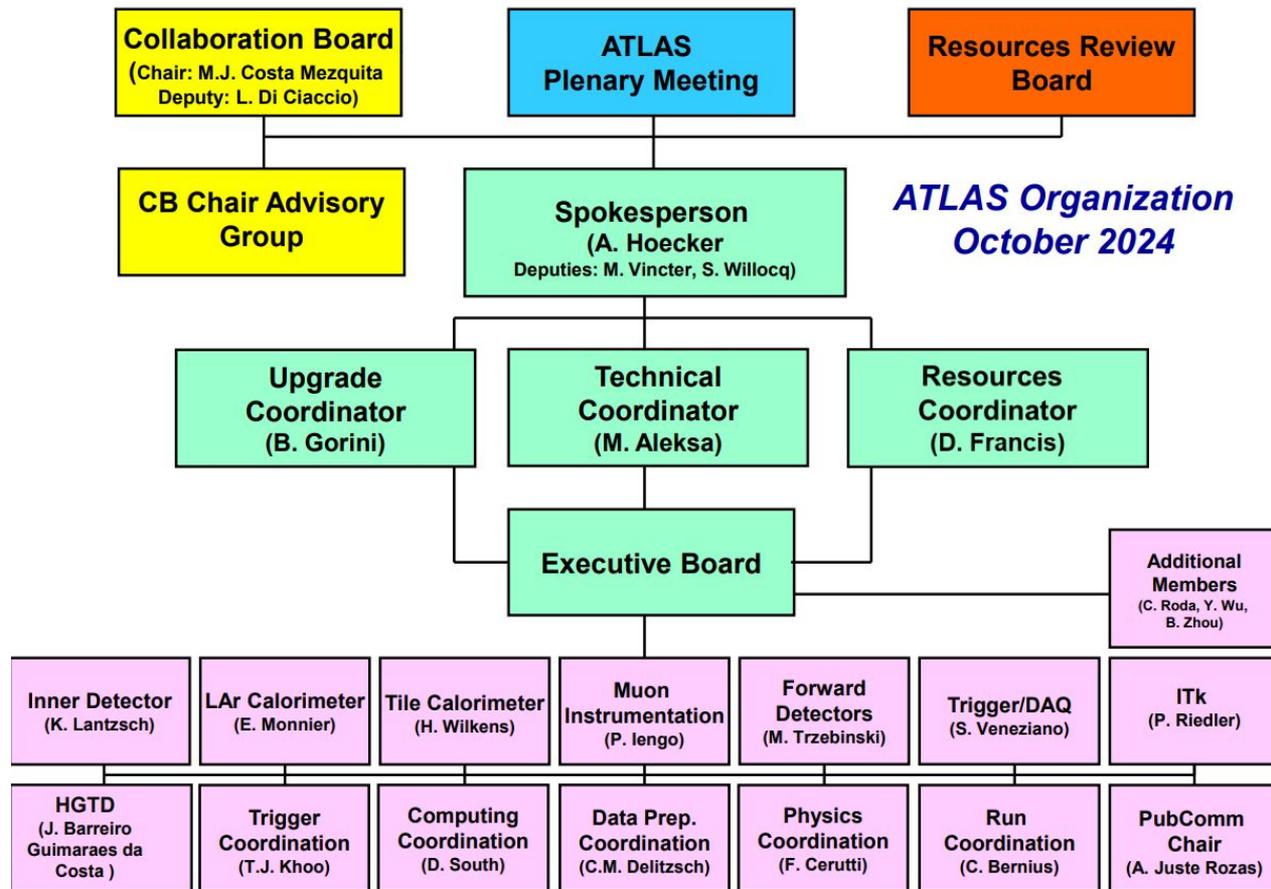
- 185 instituts dans 41 pays
- 6000 membres actifs
- ~2900 auteurs
- ~1200 doctorants



La collaboration ATLAS



Organisation



Organisation



Collaboration Board
(Chair: M.J. Costa Mezquita
Deputy: L. Di Ciaccio)

**ATLAS
Plenary Meeting**

**Resources Review
Board**

**CB Chair Advisory
Group**

Spokesperson
(A. Hoecker
Deputies: M. Vinciter, S. Willocq)

**ATLAS Organiza
October 2024**



**Upgrade
Coordinator**
(B. Gorini)

**Technical
Coordinator**
(M. Aleksa)

**Resources
Coordinator**
(D. Francis)



Executive Board

**Addit
Mem**
(C. Roda
B. Zl)



**orimeter
onnier)**

Tile Calorimeter
(H. Wilkens)

**Muon
Instrumentation**
(P. Iengo)

**Forward
Detectors**
(M. Trzebinski)

Trigger/DAQ
(S. Veneziano)

IT
(P. Ri)



**igger
ination
Khoo)**

**Computing
Coordination**
(D. South)

**Data Prep.
Coordination**
(C.M. Delitzsch)

**Physics
Coordination**
(F. Cerutti)

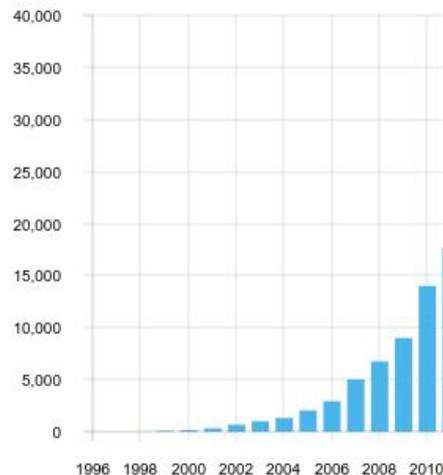
**Run
Coordination**
(C. Bernius)

**PubC
Ch**
(A. Juste)

Quelques réunions...

Number of events

The year is the one of the start date of t



472,579
TOTAL NUMBER OF EVENTS

Je vous rappelle
l'objet de cette réunion:



« Définir une méthodologie
pour lutter contre la réunionite »



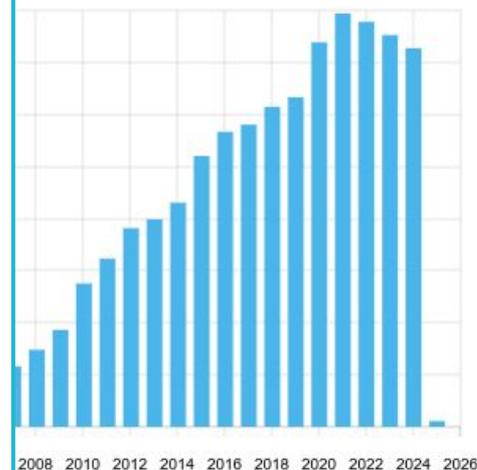
Bien sûr, cette question est tellement complexe
que nous ne la résoudrons pas en une seule réunion.



Xavier Gorce

ons

ate of the contribution.



1,479,670
NUMBER OF ATTACHMENTS

- **N. Makovec, cours d'analyse de données**
 - Ecole du détecteur à la Physique 2017
 - http://formation.in2p3.fr/PhysDet17/Support_PhysDet17.html
- **A. Sfyrla, P. Laycock, From raw data to physics**
 - Ecoles d'été du CERN 2018 et 2019
 - <https://indico.cern.ch/category/345/>
- **N. Berger, Introduction to statistical analysis**
 - Ecole d'été du CERN 2019
 - <https://indico.cern.ch/category/345/>
- **G. Cowan, Statistical data analysis**
 - Clarendon press, Oxford, 1998
- **PDG, Review, Statistics**
 - <https://pdg.lbl.gov/2021/web/viewer.html?file=%2F2021/reviews/rpp2020-rev-statistics.pdf>
- **Résultats et figures d'ATLAS**
 - <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/WebHome>

Animations

