



CMS Experiment at the LHC, CERN

Data recorded: 2011-Jun-25 06:34:20.986785 GMT (08:34:20 CEST)

Run / Event: 167675 / 876658967

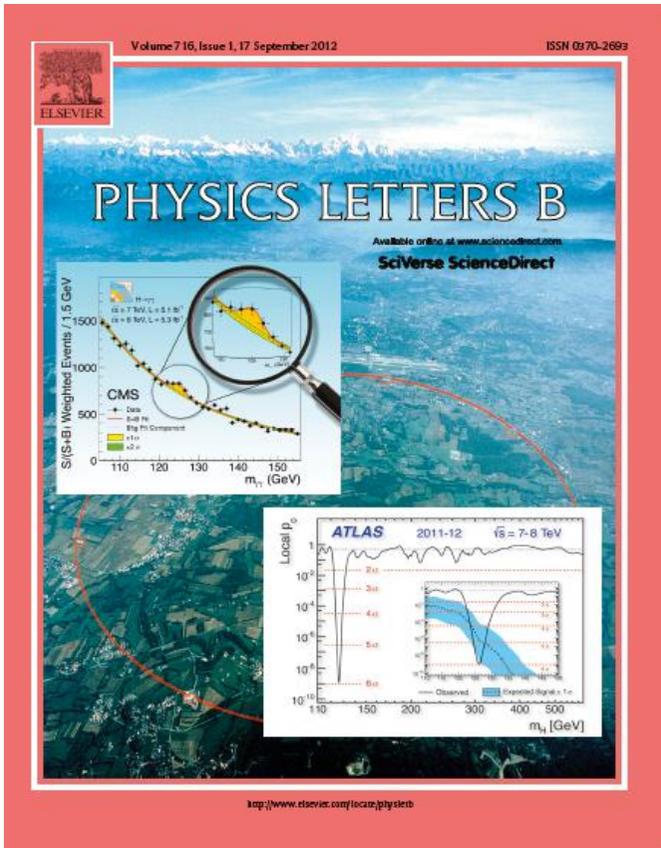
Découverte du boson Higgs : Rôle du computing/Feedback utilisateur

30 mai 2013,

Rencontre LCG-France, LLR/Palaiseau

Christophe Ochando
Laboratoire Leprince Ringuet
CNRS/IN2P3/Ecole Polytechnique





ATLAS Collaboration,

“Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”,
Phys.Lett.B **716** (2012) 1.

CMS Collaboration,

“Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”,
Phys.Lett.B **716** (2012) 30.

Soumis le 31 Juillet 2012 (Publiés le 17 Septembre 2012)

> 1000 citations chacun

(dans les 400 papiers les plus cités du domaine)

Remerciements

Prix Spécial du « Fundamental Physics Prize Foundation » (2012)



Prix de la Société Européenne de Physique en physique des hautes énergies (2013)

The High Energy and Particle Physics Prize 2013

ATLAS collaboration (CERN, CH)

CMS collaboration (CERN, CH)

For the discovery of a Higgs boson as predicted by the Brout-Englert-Higgs mechanism

Michel Della Negra (CERN, now at Imperial College London, UK)

Peter Jenni (CERN, now at Freiburg University, DE)

Tejinder Virdee (Imperial College London, UK)

For their pioneering and outstanding leadership rôles in the making of the ATLAS and CMS experiments.

In July 2012, the ATLAS and CMS collaborations announced the discovery of a new heavy particle at a mass around 125 GeV. Its properties were strikingly similar to those of a Higgs boson, a long-sought particle expected from the mechanism for electroweak symmetry breaking that was introduced almost 50 years ago by Robert Brout, François Englert and Peter Higgs. Experimental confirmation of the Higgs boson presented monumental challenges because of its relatively large mass and small production rates in cleanly detectable modes. The observation has required the creation of experiments of unprecedented capability and complexity, designed to discern the signatures that correspond to the Higgs boson. Their creation has required the use, and in many cases the development, of cutting-edge technologies. In addition, the gigantesque structures were supplemented with appropriate software and computing systems that enabled the analysis of the vast amounts of data that had to be collected. This work has required the

Remerciements

Prix Spécial du « Fundamental Physics Prize Foundation » (2012)



Prix de la Société Européenne de Physique en physique des hautes énergies (2013)

The High Energy and Particle Physics Prize 2013

ATLAS collaboration (CERN, CH)

CMS collaboration (CERN, CH)

Un grand merci à tous les techniciens, ingénieurs & physiciens du software & computing sans qui rien de tout ceci n'aurait été possible !!!

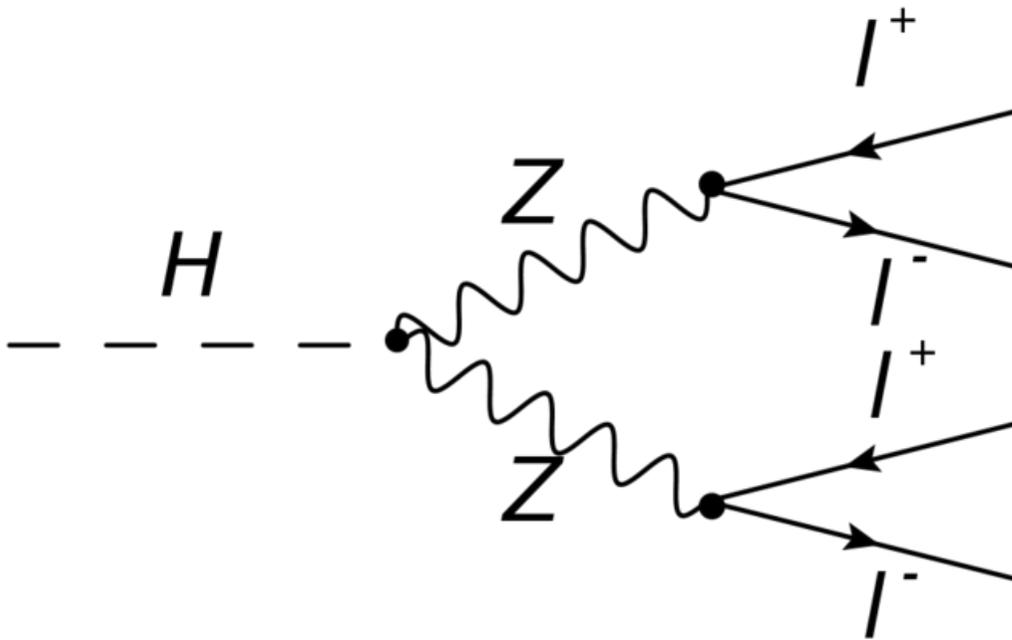
mechanism

ATLAS and CMS

In July 2012, the ATLAS and CMS collaborations announced the discovery of a new heavy particle at a mass around 125 GeV. Its properties were strikingly similar to those of a Higgs boson, a long-sought particle expected from the mechanism for electroweak symmetry breaking that was introduced almost 50 years ago by Robert Brout, François Englert and Peter Higgs. Experimental confirmation of the Higgs boson presented monumental challenges because of its relatively large mass and small production rates in cleanly detectable modes. The observation has required the creation of experiments of unprecedented capability and complexity, designed to discern the signatures that correspond to the Higgs boson. Their creation has required the use, and in many cases the development, of cutting-edge technologies. In addition, the gigantesque structures were supplemented with appropriate software and computing systems that enabled the analysis of the vast amounts of data that had to be collected. This work has required the

« Analyse de données CMS: Retour utilisateur »

- Exemple: $H \rightarrow ZZ \rightarrow 4$ leptons : un des canaux de découverte
 - Contexte (CMS, Période Avril-Juillet 2012,...)
 - Schéma d'analyse
 - Rôle/utilisation de la grille dans une analyse de CMS (+ feedback)

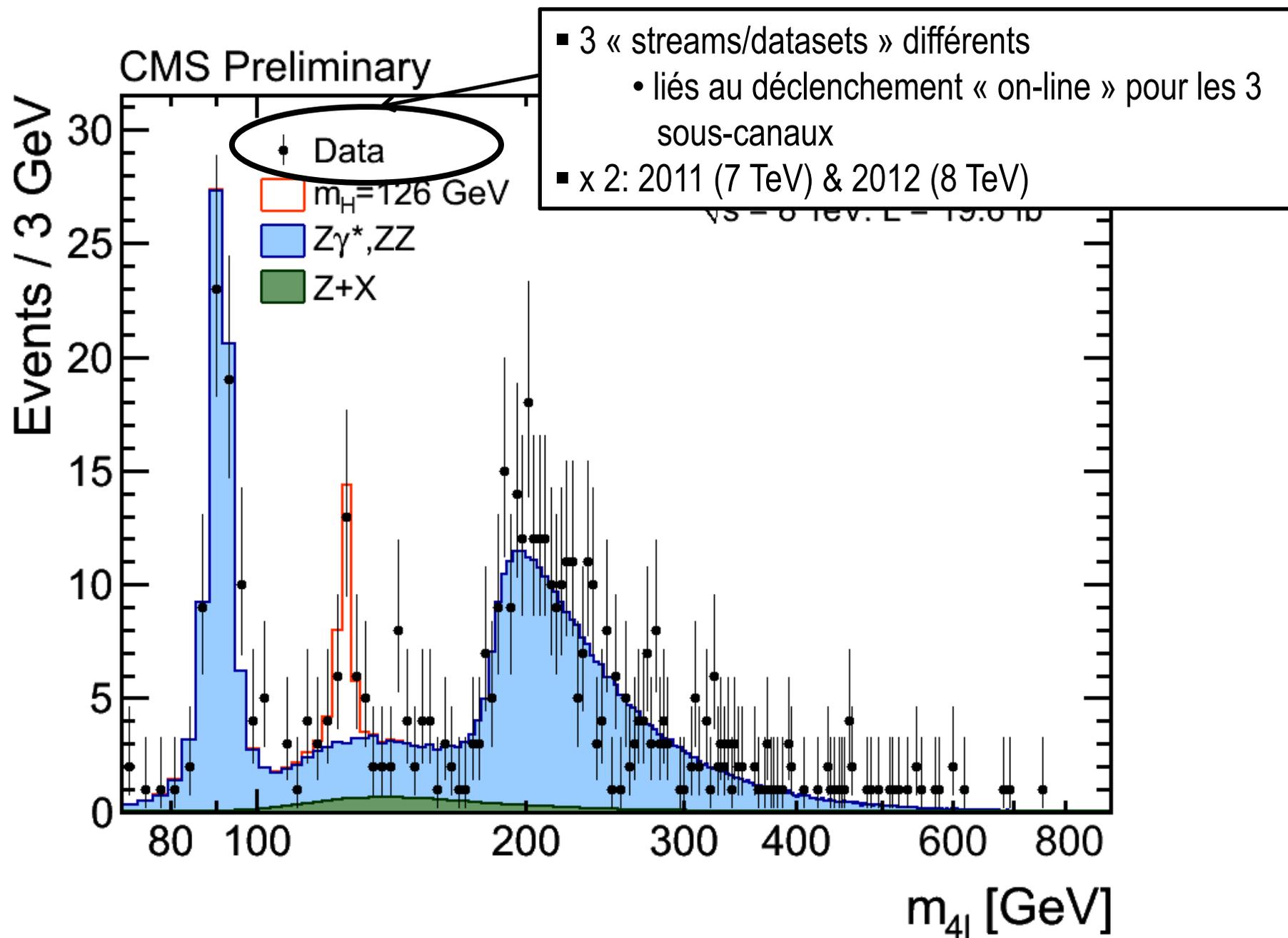


- l = électron ou muons,
Très bien mesurés dans les détecteurs

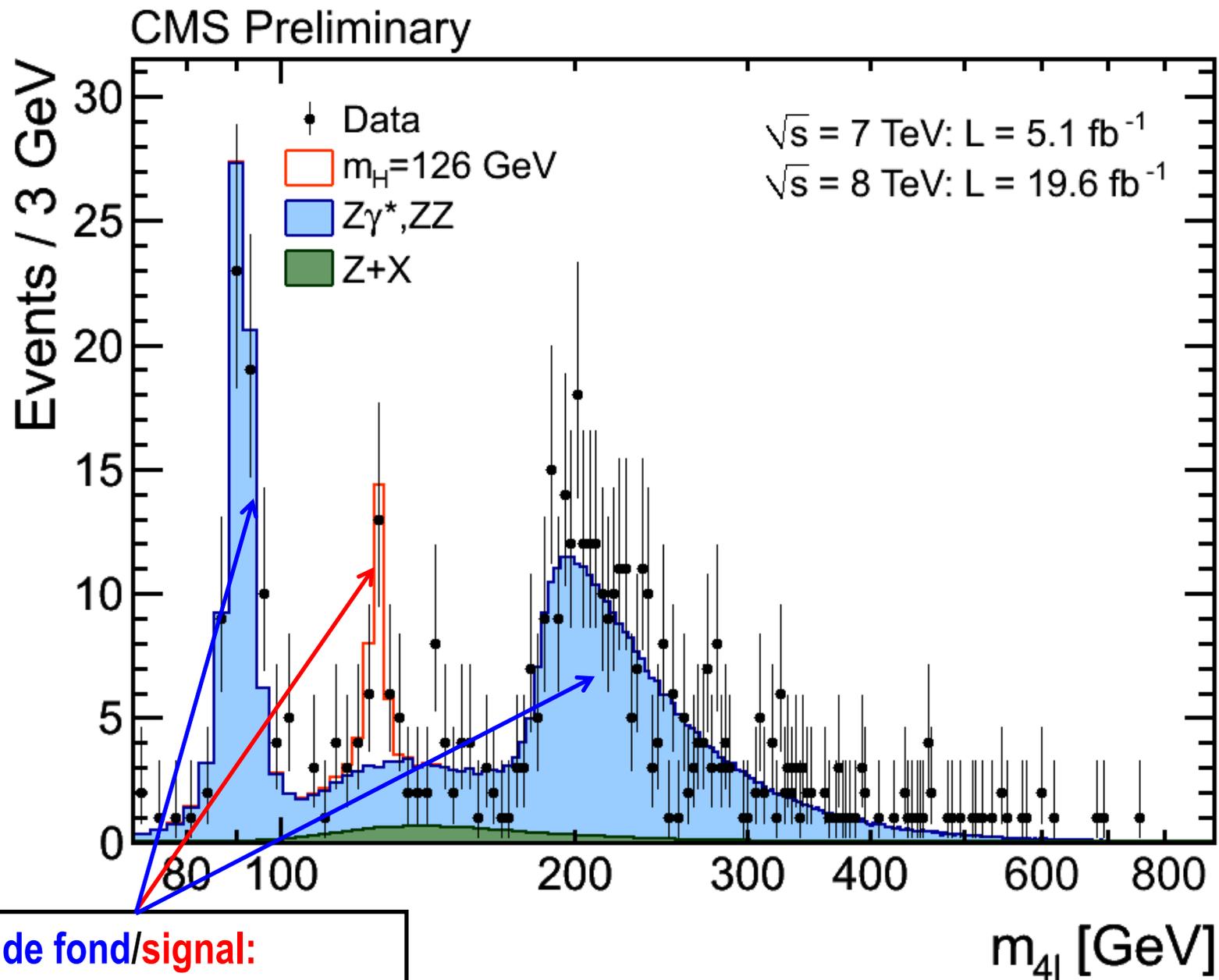
- 3 sous-canaux:
 - 4 électrons ($4e$),
 - 4 muons (4μ),
 - 2 électrons, 2 muons ($2e2\mu$)

- Processus (très) rare...
... mais très bon S/B.

Analyse: $H \rightarrow ZZ \rightarrow 4$ leptons (HZZ4l)

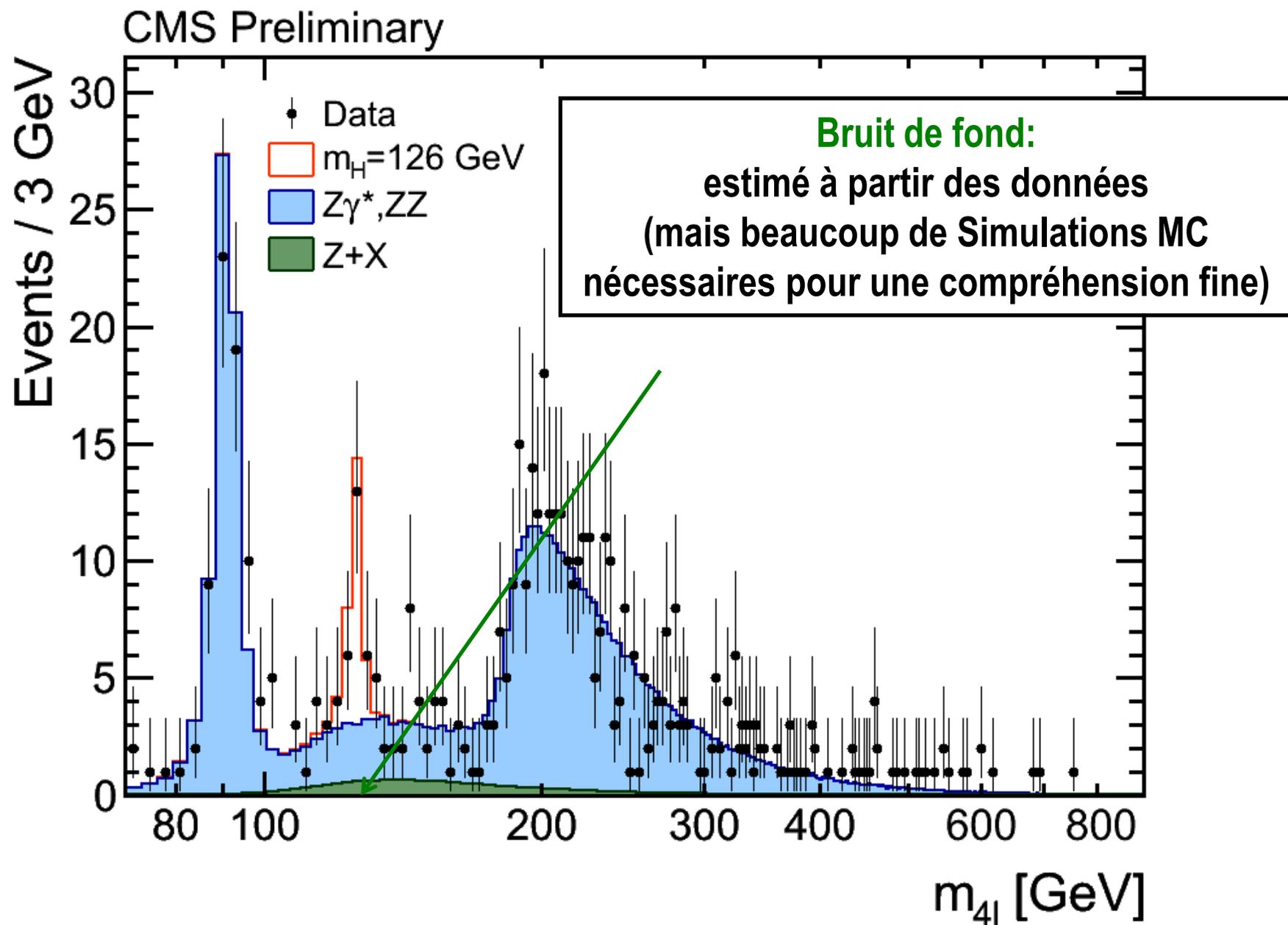


Analyse: $H \rightarrow ZZ \rightarrow 4$ leptons (HZZ4l)

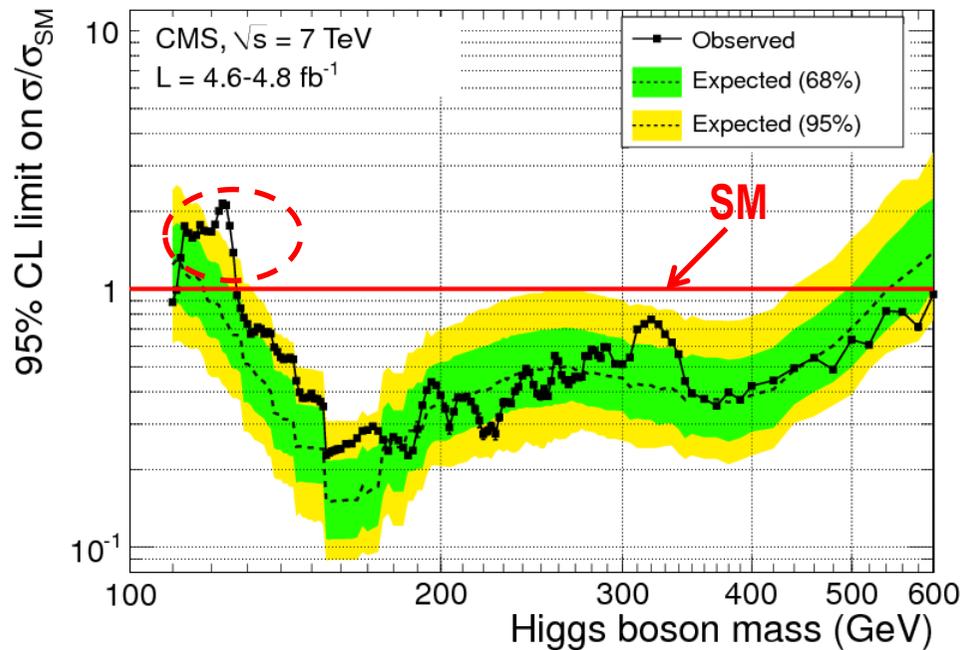


Bruit de fond/signal:
estimé à partir de Simulations (MC)

Analyse: $H \rightarrow ZZ \rightarrow 4 \text{ leptons}$ (HZZ4l)



Contexte



- Fin 2011/Début 2012: Ensemble des données à 7 TeV analysées.
- Standard Model (SM) Higgs exclu à 95% de confiance partout...
... sauf autour de 125 GeV (**petit excès**)
- Si c'est un signal, **découverte possible en doublant la statistique** (workshop Chamonix)

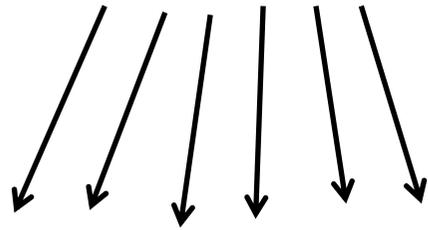
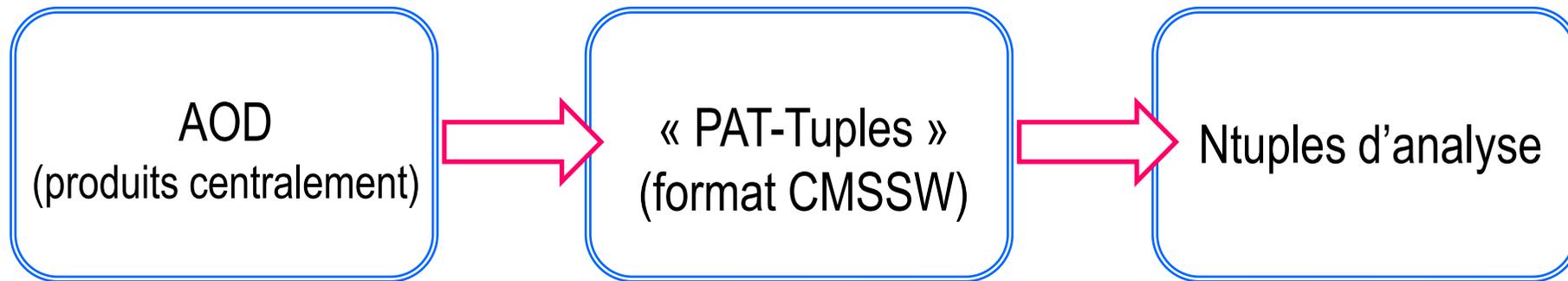
➤ Conférence ICHEP début Juillet 2012.

- **30 Mars 2012:** Reprise des opérations au LHC:
 - **But: délivrer au moins autant de collisions qu'en 2011...**
... mais en quelques mois !! (à temps pour ICHEP)
 - Energie dans le centre de masse : 7 → 8 TeV
 - Empilement (PileUp) : ~x2/x3 wrt 2011.
- **HZZ4l:** « Very High Priority Analysis » de CMS (ie, « must no fail »)

Schéma d'analyse

➤ Caveat:

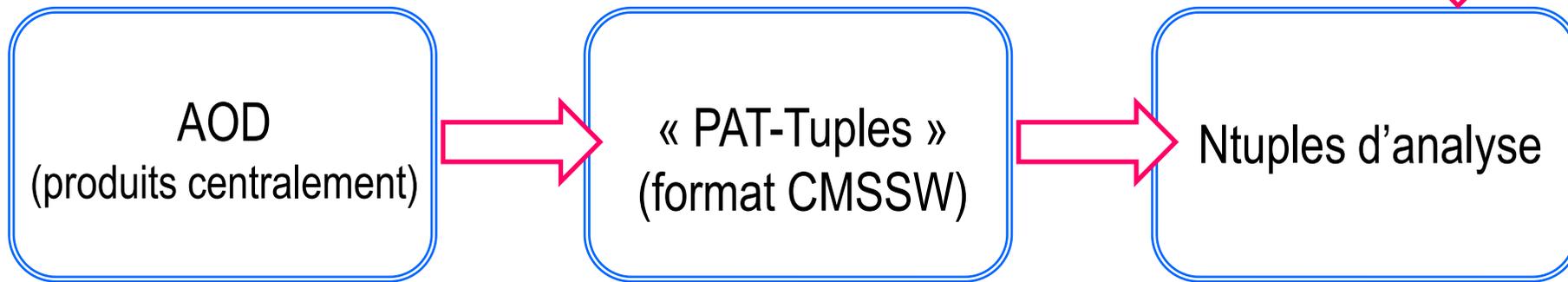
- Schéma « standard » de l'analyse de base (réduction/estimation du bruit de fond, etc...)
- Cloner X fois pour les autres tâches essentielles: calibrations, validation/correction de la simulation, ...



Distribués aux T2's

- Réduction d'un facteur ~20
 - informations non-utiles enlevées,
 - ajout de « user variable »
 - possible « skimming » pour les gros samples
 - Créés avec GRID
 - Stockés localement (Tier3)
 - ~10-15 jours pour avoir données/MC complets
- Réduction d'un facteur ~7000
 - coupures d'analyses appliquées
 - conserve seulement quelques variables utiles
 - Créés avec Batch System / Tier 3
 - Stockés sur disque / laptop

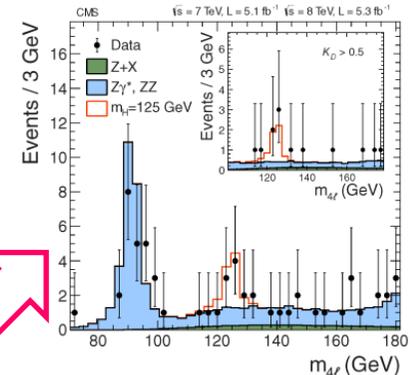
Schéma d'analyse



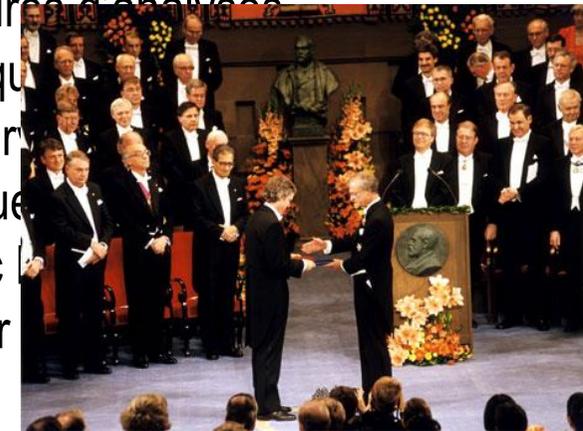
↓ ↓ ↓ ↓ ↓
Distribués aux T2's

- Réduction d'un facteur ~20
 - informations non-utiles enlevées,
 - ajout de « user variable »
 - possible « skimming » pour les gros samples
- Créés avec GRID
- Stockés localement (Tier3)
- ~10-15 jours pour avoir données/MC complets

- Réduction d'un facteur ~7000
 - coupures d'analyse appliquées
 - conservation de quelques événements
- Créés avec
- Stockés sur



Découverte !



Prix Nobel, etc...

Schéma d'analyse... un peu plus compliqué dans la réalité...

Données arrivent en continu... (à répéter plusieurs fois)

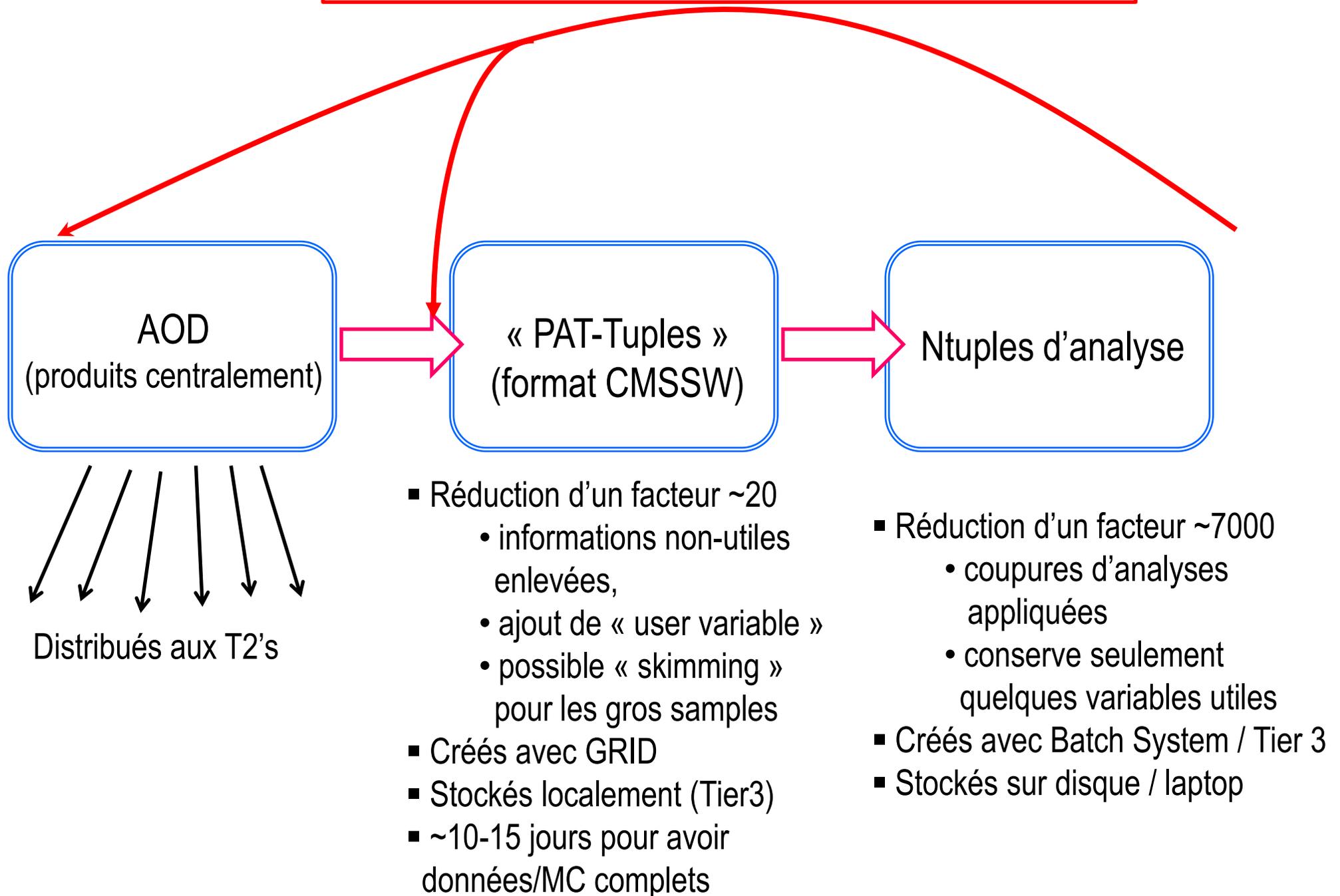


Schéma d'analyse... un peu plus compliqué dans la réalité...

Données arrivent en continu... (à répéter plusieurs fois)

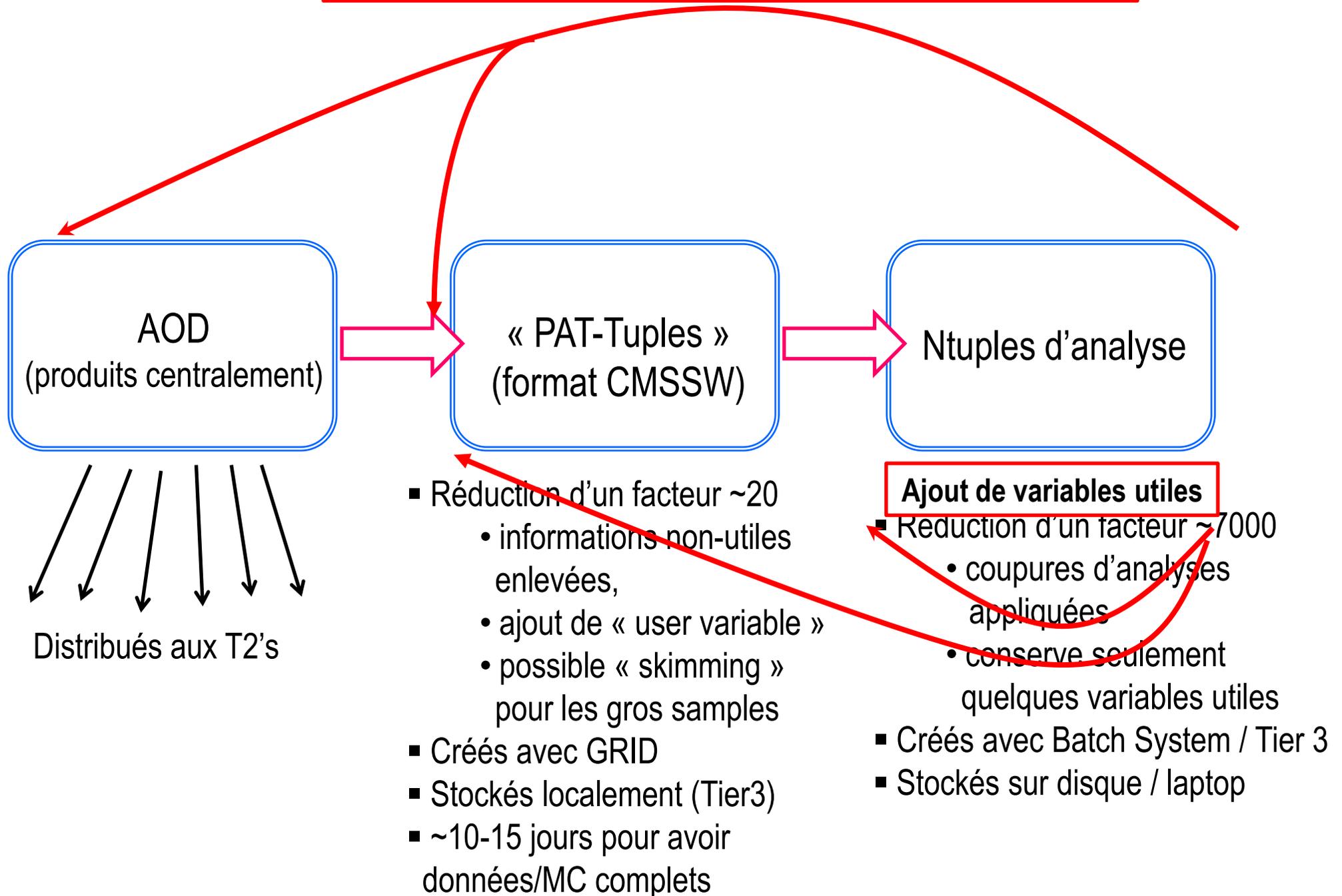


Schéma d'analyse... un peu plus compliqué dans la réalité...

Données arrivent en continu... (à répéter plusieurs fois)

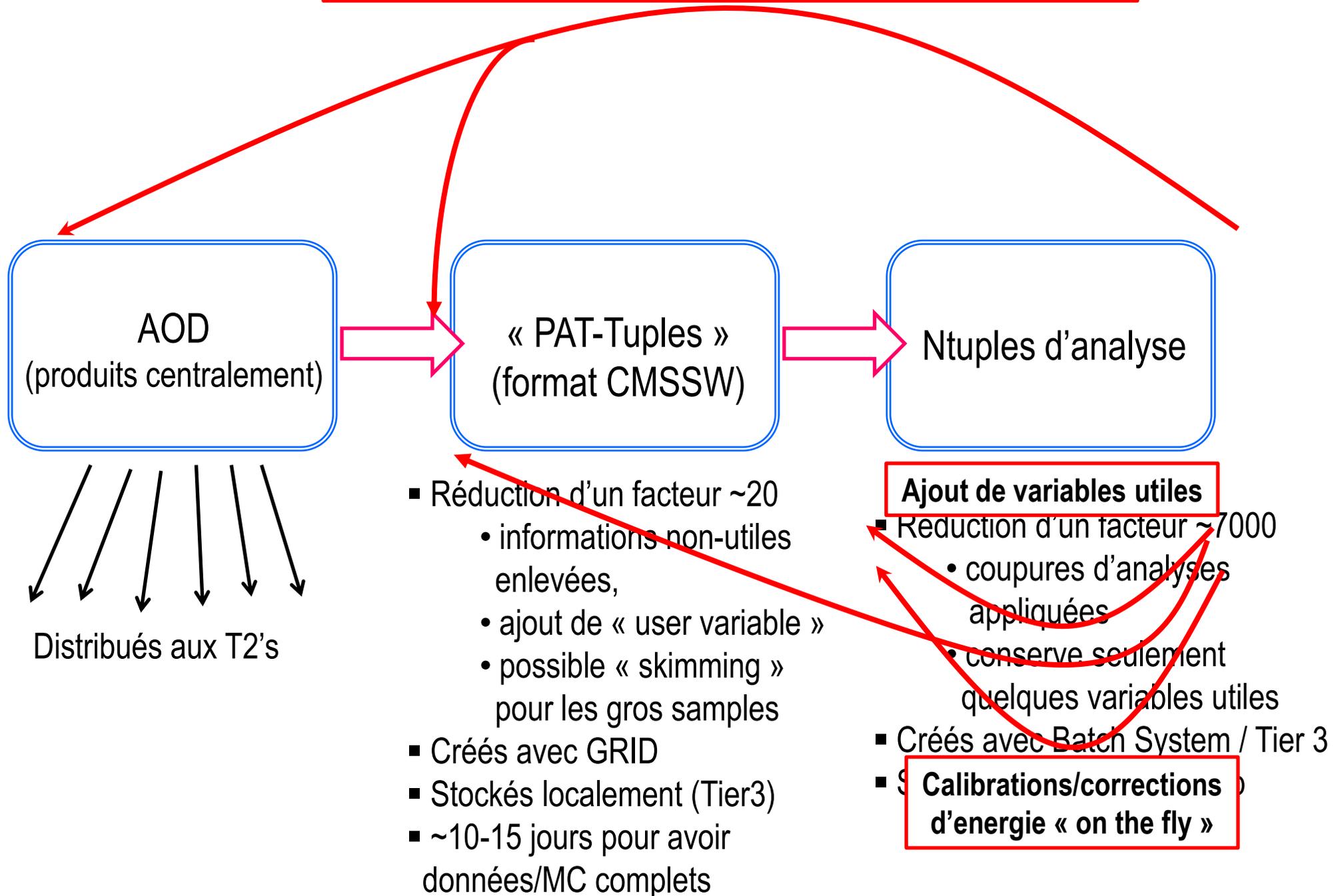


Schéma d'analyse... un peu plus compliqué dans la réalité...

Données arrivent en continu... (à répéter plusieurs fois)

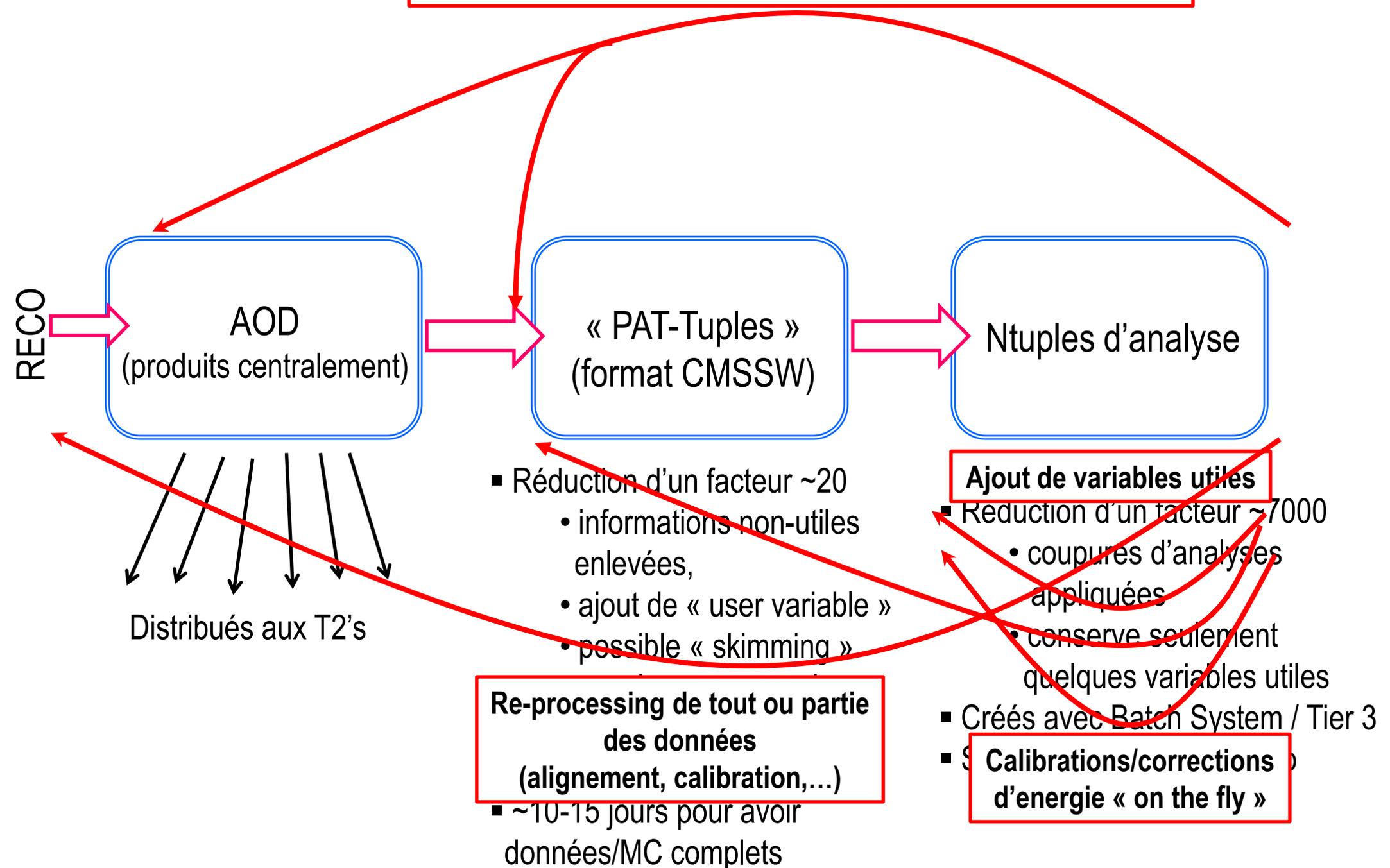


Schéma d'analyse... un peu plus compliqué dans la réalité...

Données arrivent en continu... (à répéter plusieurs fois)

Simulations MC arrivent en continu...
(à répéter plusieurs fois)

RECO

AOD
(produits centralement)

« PAT-Tuples »
(format CMSSW)

Ntuples d'analyse

Distribués aux T2's

- Réduction d'un facteur ~20
 - informations non-utiles enlevées,
 - ajout de « user variable »
 - possible « skimming »

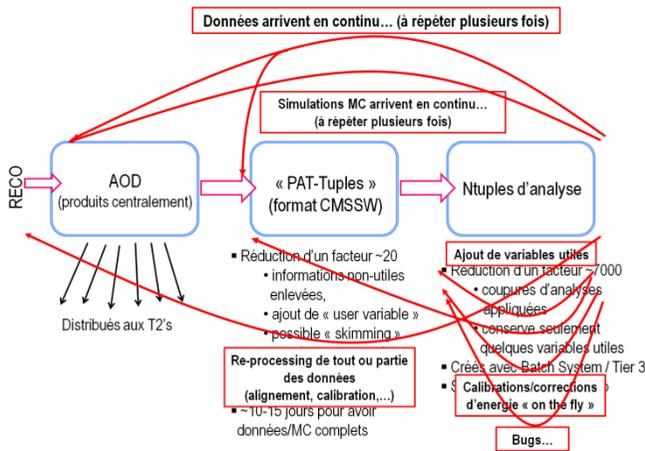
Ajout de variables utiles

- Réduction d'un facteur ~7000
 - coupures d'analyses appliquées
 - conserve seulement quelques variables utiles
- Créés avec Batch System / Tier 3
- Calibrations/corrections d'énergie « on the fly »

- Re-processing de tout ou partie des données
(alignement, calibration,...)
- ~10-15 jours pour avoir données/MC complets

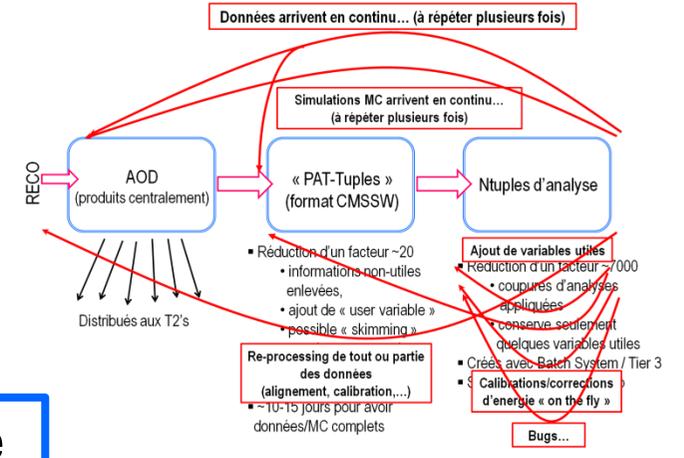
Bugs...

Schéma d'analyse... un peu plus compliqué dans la réalité...

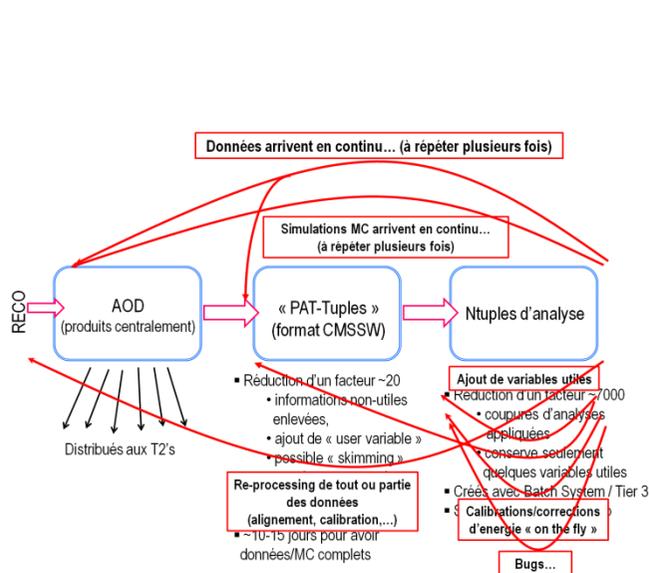


T2_FR_GRIF(*)
(Partie T3: LLR)

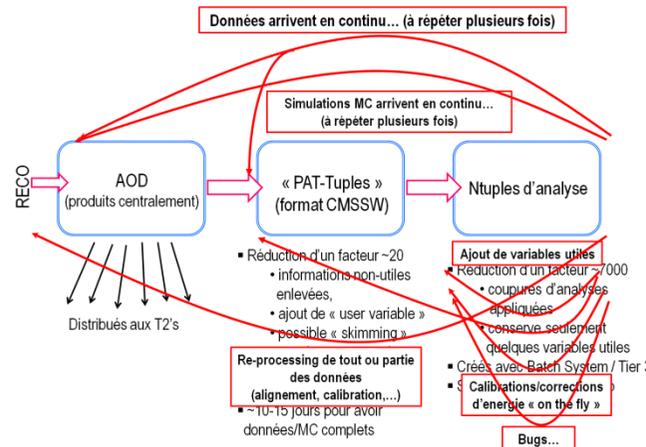
■ ≥ 5 groupes **indépendants** d'analyse
 ■ ... qui doivent être parfaitement synchronisés !



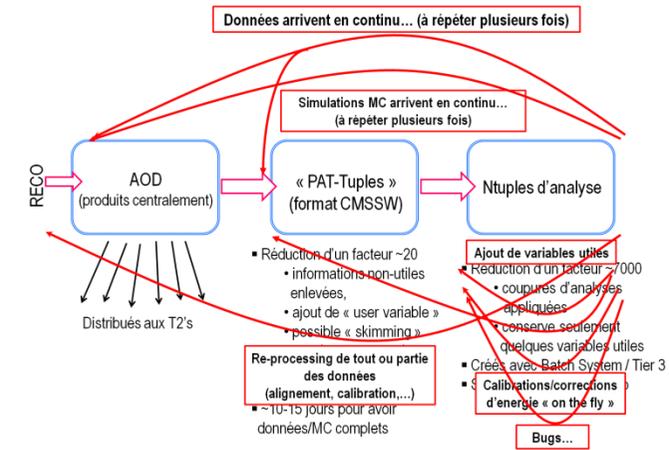
T2_US_UF (*)
(Partie T3: University of Florida)



T2_US_UCSD, T2_US_Caltech (*)
(Partie T3: UCSD, Caltech)

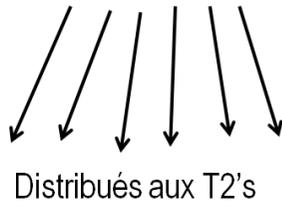
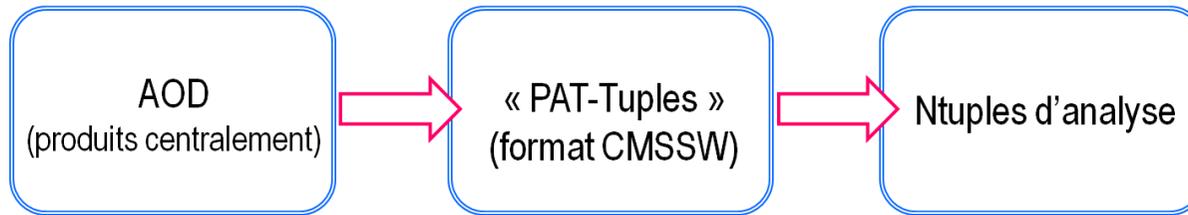


T2_US_MIT (*)
(Partie T3: MIT)



Other groups

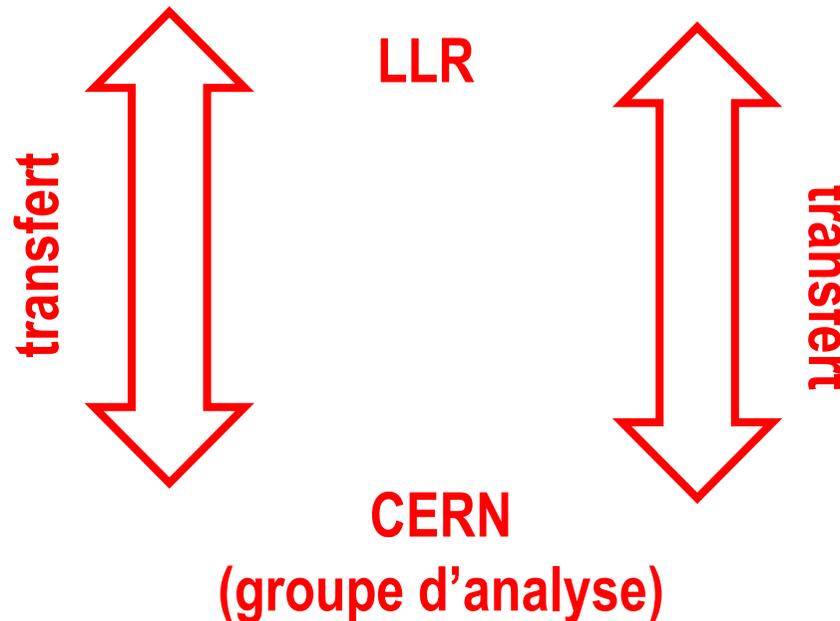
Collaborations



T2_FR_GRIF(*)
(Partie T3: LLR)

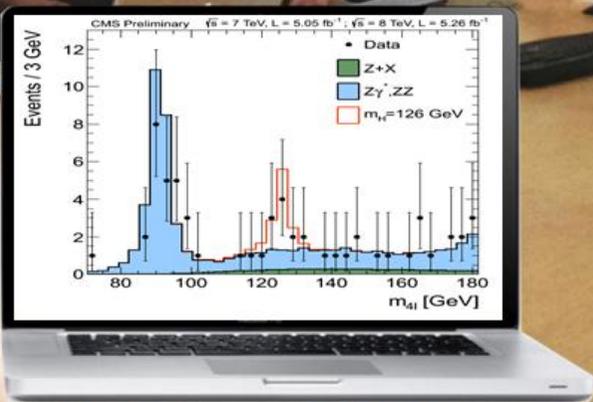
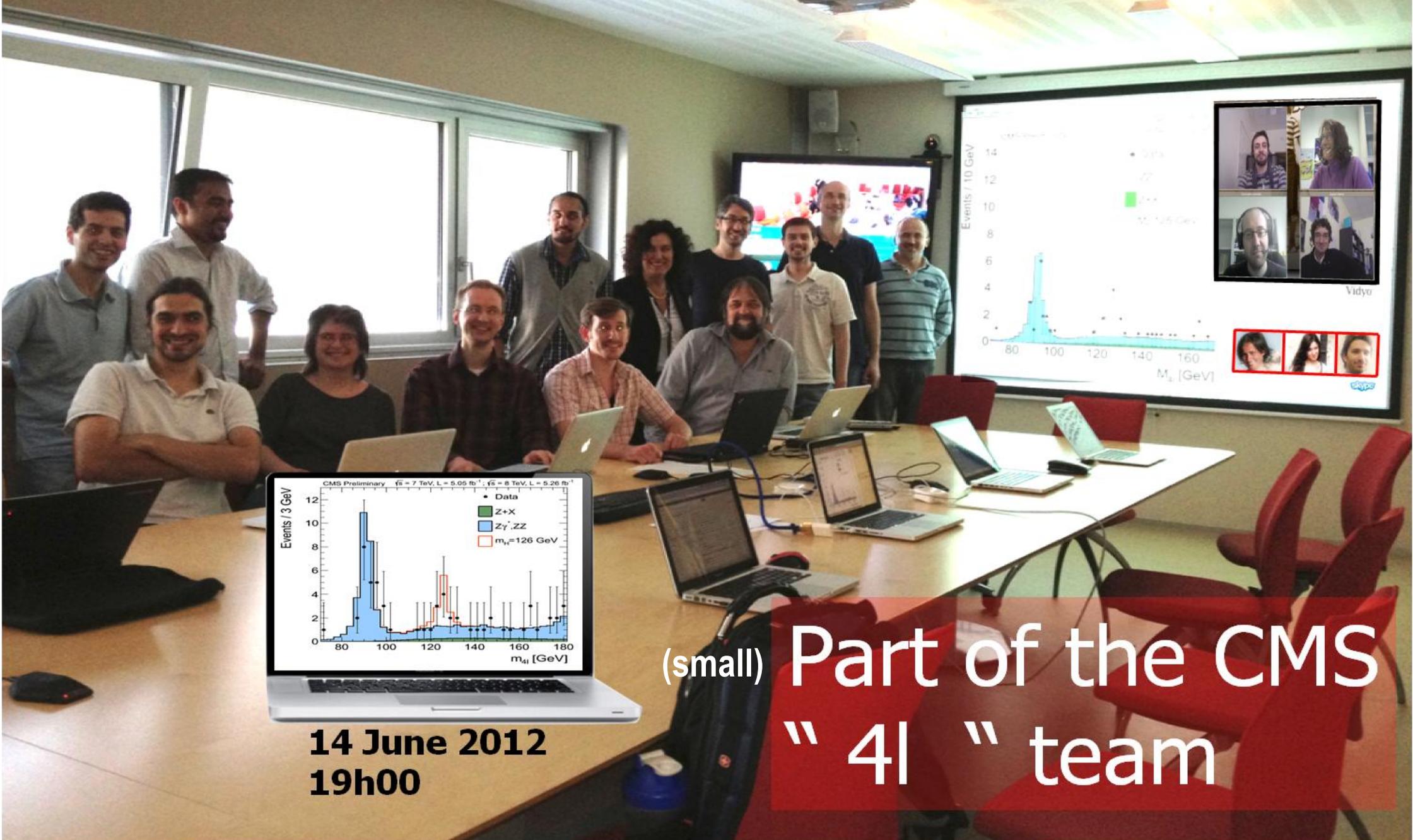
- Réduction d'un facteur ~20
 - informations non-utiles enlevées,
 - ajout de « user variable »
 - possible « skimming » pour les gros samples
 - Créés avec GRID
 - Stockés localement (Tier3)
 - ~10-15 jours pour avoir données/MC complets
- Réduction d'un facteur ~7000
 - coupures d'analyses appliquées
 - conserve seulement quelques variables utiles
 - Créés avec Batch System / Tier 3
 - Stockés sur disque / laptop

+ Collaboration étroite
(échange de code/data) avec :
CERN, JHU (US), Split, Torino



- Partage de production
- Réplication (transfert via FTS)
- CRAB adapté pour utilisation en soumetteur de job local

Analysis groups meet separately around mid-June 2012 CMS unblinds the critical masse range ...



14 June 2012
19h00

(small) Part of the CMS
"41" team

Quelques chiffres...

Données (*):

- (O)10 MHz: Rate LHC, (O)300 Hz en sortie de High Level Trigger
- >170 000 000 d'évènements analysés (AOD)
- 14 datasets
- ~40 TB (AOD)
- 2 re-processing (partiels): fin mai, mi-juin...
- Dernières données analysées: ~25 juin.

- 172 évènements sélectionnés
- ~8 évènements de signal attendus (mH=126 GeV)

Simulations (*):

- ~200 000 000 d'évènements analysés (AOD)
- ~240 datasets
- ~50 TB (AOD)

(*): données/simulations 2011 (7 TeV) et 2012 (8 TeV): CMSSW releases différentes...

Feedback « GRID T2 ». Principaux problèmes:

➤ **Scheduling:**

- Glite, remoteGlidein, condor,...
- Pas toujours évident de comprendre la distribution des jobs aux différents T2... (ne semble pas forcément la plus optimale/évidente)

➤ « **Baby-sitting** » des jobs: prend beaucoup (trop) de temps lorsque de nombreuses erreurs sont rencontrées.

➤ « **stage-out** »: notamment avec certains sites US...

➤ **Informations sur les sites** pas assez transparentes/accessibles facilement:

- disponibilité, qui/comment contacter en cas de problème,

Tous ces petits soucis deviennent **cruciaux** dans un **cycle/période de découverte** comme décrit précédemment (pression considérable, compétition interne, timing ultra serré, ...)

« Bottle Neck » / Feedback / Problèmes

Feedback « GRID T2 ». Principaux problèmes:

- **Scheduling:**
 - Glite, remoteGlidein, condor,...
 - Pas toujours évident de comprendre la distribution des jobs aux différents T2...
(ne semble pas forcément la plus optimale/évidente)

- « **Baby-sitting** » des jobs: prend beaucoup (trop) de temps lorsque de nombreuses erreurs sont rencontrées.

- « **stage-out** »: notamment avec certains sites US...

- **Informations sur les sites** pas assez transparentes/accessibles facilement:
 - disponibilité, qui/comment contacter en cas de problème,

- Il est apparu assez vite que pour rester dans le tempo et travailler efficacement, une **condition nécessaire (mais pas suffisante...)** était:
 - **Avoir un T2(+T3) « ami » proche, stable, efficace, bien (inter-)connecté,**
avec contact efficace & réactif (ie, c'est le cas de **T2_FR_GRIF**)
afin de lancer/re-lancer tout ou partie des jobs lors des différents cycles
(même chose pour les groupes US...)

Conclusion

- **La découverte du boson de Higgs au LHC n'aurait pas été possible sans la Grille & le travail des techniciens, ingénieurs et physiciens pour qu'elle fonctionne.**

- **Conditions d'analyse extrêmes au LHC pour une découverte :**
 - Données & simulation arrivent en continu
 - timing très serré
 - pression considérable
 - compétition interne
 - nombreux cycles de (re-)processing

- **Pose des contraintes fortes sur le computing**
 - Quelques difficultés:
 - « stage-out », manque d'informations/difficulté à contacter admins pour les sites distants,
 - « baby-sitting » trop prenant,...

 - **condition nécessaire (mais pas suffisante...) pour suivre la cadence:**
 - **Avoir un T2(+T3) « ami » proche, stable, efficace, bien (inter-)connecté,**
avec contact efficace & réactif,
afin de lancer/re-lancer tout ou partie des jobs lors des différents cycles

BACK UP SLIDES

Le boson de Higgs : principaux “canaux” de recherche

H → quark b, anti-quark b

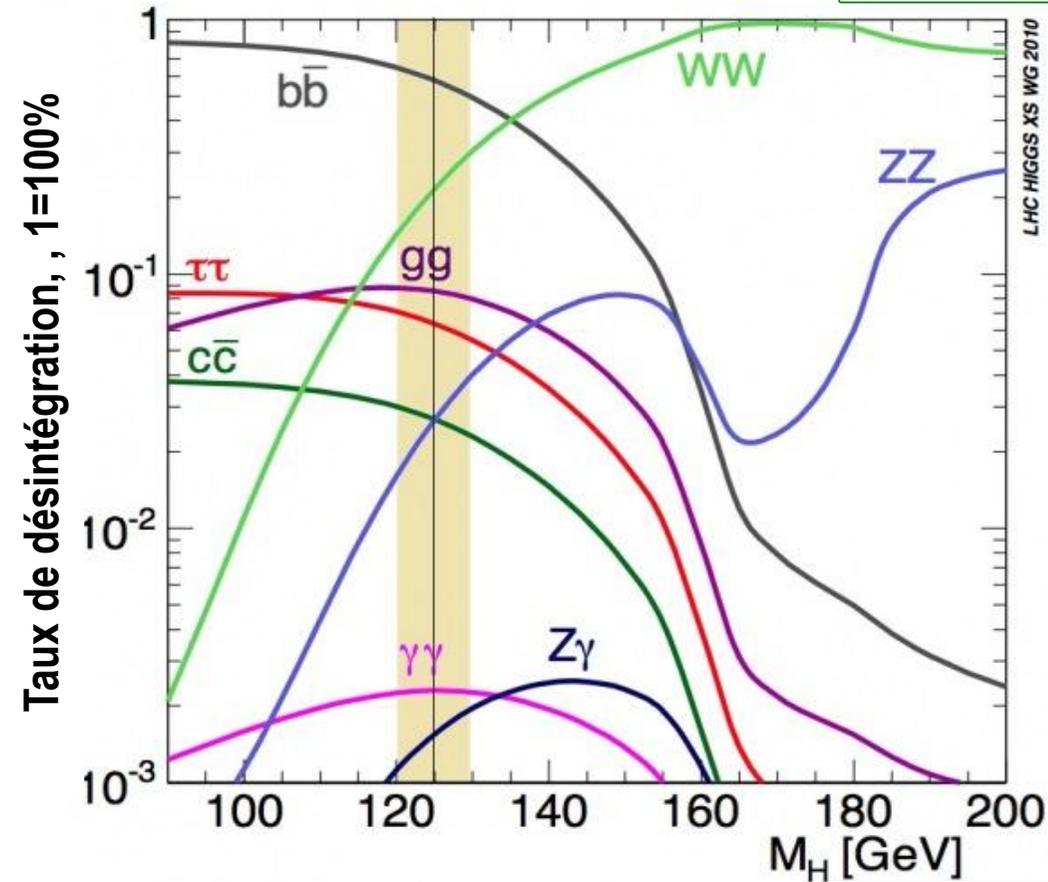
Important à “basse” masse...
... mais très difficile à détecter

H → boson W+, boson W-

Fort taux de désintégration...
...mais détection peu précise

H → lepton τ , anti-lepton τ

Important à “basse” masse...
... mais très difficile à détecter



H → 2 photons

Très rare... mais très précis à
basse masse !

H → 2 boson Z^0 → 4 leptons:

$Z \rightarrow e+e-$, $Z \rightarrow e+e-$
ou $Z \rightarrow \mu+\mu-$, $Z \rightarrow \mu+\mu-$
ou $Z \rightarrow e+e-$, $Z \rightarrow \mu+\mu-$
“Canal en or”

Le Modèle Standard : une simplicité toute relative ...

$$-\frac{1}{2}\partial_\nu g_\mu^a \partial_\nu g_\mu^a - g_s f^{abc} \partial_\mu g_\nu^a g_\mu^b g_\nu^c - \frac{1}{4}g_s^2 f^{abc} f^{ade} g_\mu^b g_\nu^c g_\mu^d g_\nu^e + \frac{1}{2}ig_s^2 (\bar{q}_i^\sigma \gamma^\mu q_j^\sigma) g_\mu^a + \bar{G}^a \partial^2 G^a + g_s f^{abc} \partial_\mu \bar{G}^a G^b g_\mu^c -$$

$$\partial_\nu W_\mu^+ \partial_\nu W_\mu^- - M^2 W_\mu^+ W_\mu^- - \frac{1}{2}\partial_\nu Z_\mu^0 \partial_\nu Z_\mu^0 - \frac{1}{2}M^2 Z_\mu^0 Z_\mu^0 - \frac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - \frac{1}{2}\partial_\mu H \partial_\mu H - \frac{1}{2}m_H^2 H^2 -$$

$$\partial_\mu \phi^+ \partial_\mu \phi^- +$$

$$\frac{2M^2}{g^2}$$

$$W_\nu^- \partial_\nu$$

$$W_\nu^- \partial_\nu W$$

$$g^2 s_w^2 ($$

$$g \alpha [H^3$$

$$2(\phi^0)^2 H$$

$$\frac{1}{2}g[W_\mu^+ ($$

$$W_\mu^- \phi^+)$$

$$\frac{1}{4}g^2 W_\mu^+ W$$

$$W_\mu^- \phi^+)$$

$$W_\mu^- \phi^+)$$

$$\bar{d}_j^\lambda (\gamma^\mu + \tau$$

$$1 - \gamma^5)$$

$$(\bar{u}_j^\lambda \gamma^\mu (1$$

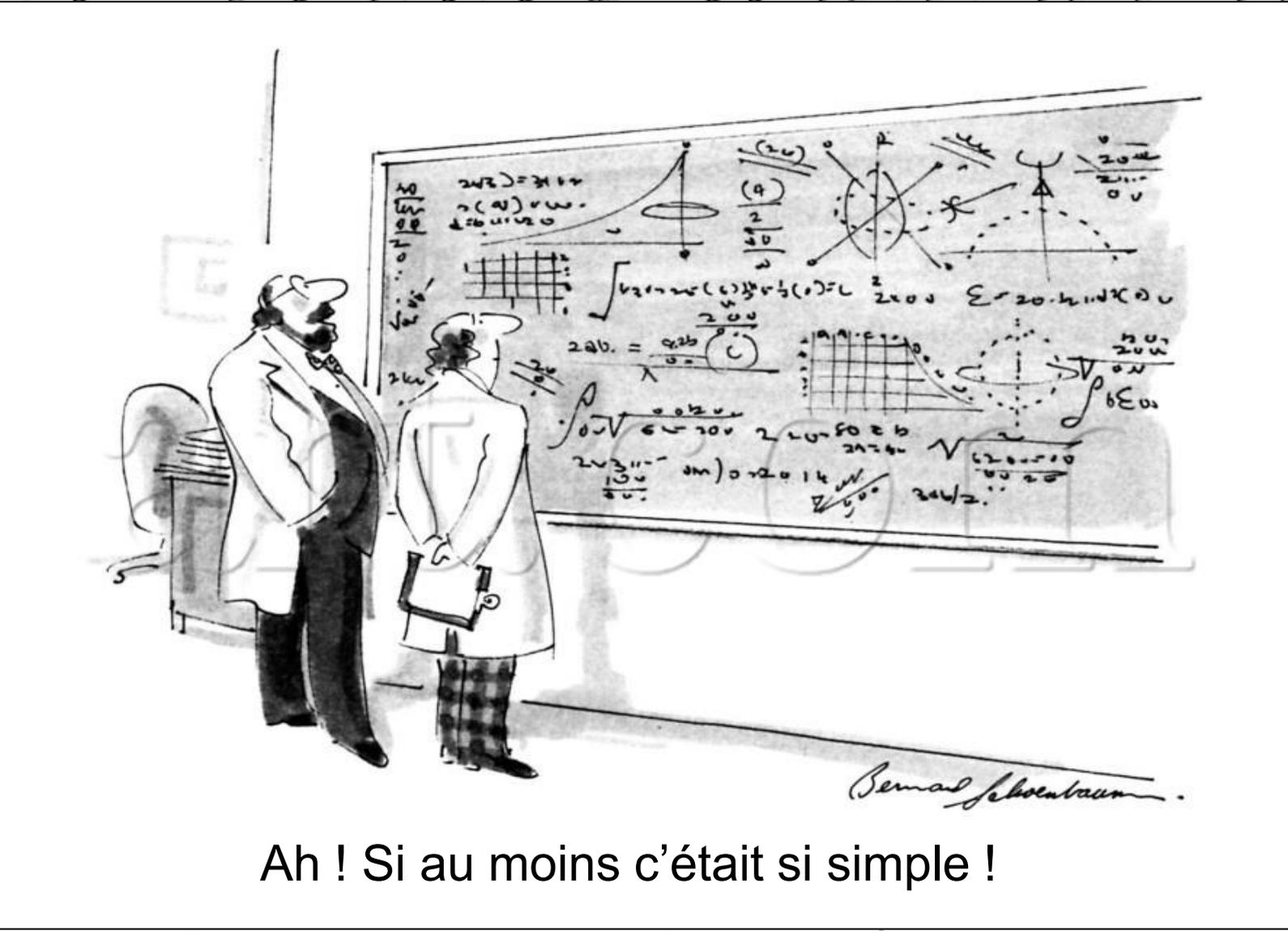
$$\gamma^5) e^\lambda)$$

$$m_u^\lambda (\bar{u}_j^\lambda ($$

$$\frac{g m_\lambda^1}{2 M}$$

$$\frac{M^2}{c_w^2}) X^0 +$$

$$\partial_\mu \bar{X}^0 X$$



Ah ! Si au moins c'était si simple !

$$\phi^-)] +$$

$$-$$

$$W_\mu^- -$$

$$W_\mu^-) +$$

$$[-] -$$

$$\phi^- +$$

$$\partial_\mu \phi^0)] +$$

$$W_\mu^+ \phi^- -$$

$$\partial_\mu \phi^+)$$

$$W_\mu^+ \phi^- +$$

$$\mu^+ \phi^- -$$

$$m_u^\lambda) u_j^\lambda -$$

$$\gamma^\mu (4s_w^2 -$$

$$) e^\lambda) +$$

$$(\bar{v}^\lambda (1 -$$

$$) d_j^\lambda) +$$

$$i_j^\lambda u_j^\lambda) -$$

$$\bar{X}^0 (\partial^2 -$$

$$\bar{X}^- X^0 -$$

$$+ X^+ -$$

Standard Model Lagrangian Density

$$\partial_\mu \bar{X}^- X^-) - \frac{1}{2}gM[\bar{X}^+ X^+ H + \bar{X}^- X^- H + \frac{1}{2}\bar{X}^0 X^0 H] + \frac{1-2c_w^2}{2c_w} igM[\bar{X}^+ X^0 \phi^+ - \bar{X}^- X^0 \phi^-] +$$

$$\frac{1}{2c_w} igM[\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + igMs_w[\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \frac{1}{2}igM[\bar{X}^+ X^+ \phi^0 - \bar{X}^- X^- \phi^0]$$