Interactions physiciens-informaticiens

Quelques exemples du point de vue d'un physicien

Avant propos

Dans les transparents suivants, je présente des exemples d'interactions physiciens-informaticiens et leur évolution.

Comme tous les membres du laboratoire, je bénéficie du support du groupe info, mais je n'en parlerai pas par la suite

Je suis membre de CMS, donc c'est un point de vue partiel qui est présenté dans la suite (je ne fais pas d'online par exemple)

En tant que tel, je classe les activités que nous avons avec le groupe informatique en plusieurs catégories

Les "services" rendus à la collaboration avec échéances définies par CMS

La R&D avec des échéances définies par le groupe

Services rendus à la collaboration

Au LLR, ils peuvent être de différentes nature:

- Gestion du Tier2
- Production d'histogrammes de validation
- Écriture de code (simulation de trigger, validation, acquisition, etc)

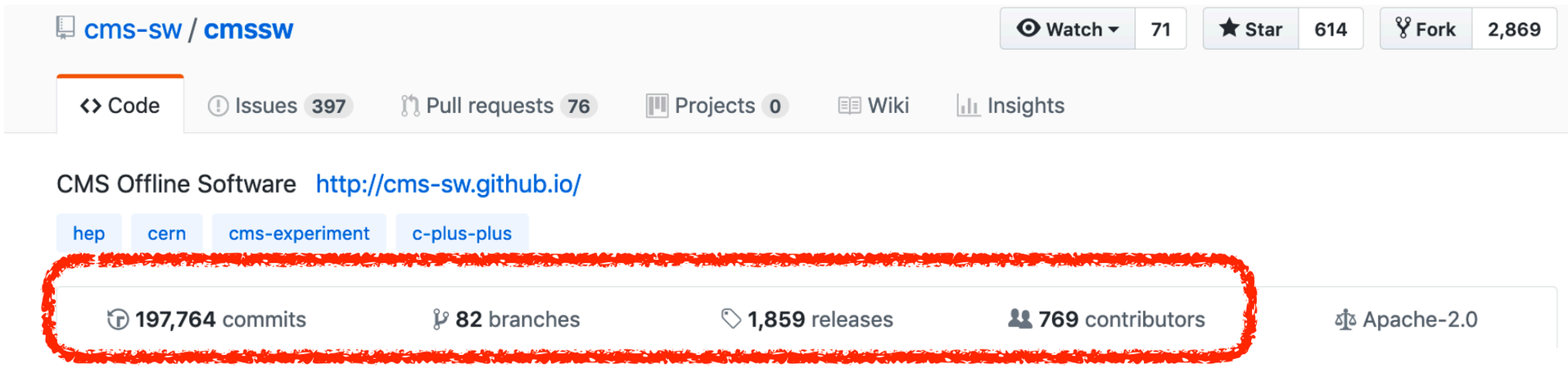
Ces travaux contribuent aux EPR du groupe

Ce sont des tâches soumises à des agendas très contraints et nécessitent un grand professionnalisme

Du point de vue de CMS, c'est un point important que l'on puisse afficher l'implication d'ingénieur(s) dans un projet

Écriture de code pour CMS

En 2013, CMS est passé de CVS à git (github.com) pour la gestion du code



Depuis, les standards de qualité ont régulièrement augmenté

- thread-safe
- règles de codage très strictes
- tests approfondis (partiellement automatisés)
- passage progressif au C+11
- configuration en python de plus en plus sophistiquée (support simultané des configurations Run 1, Run 2, upgrade)
- cycles de développements très contraints
- performances

Écriture de code pour CMS

Le code qui est intégré à CMSSW doit être irréprochable

Conséquences:

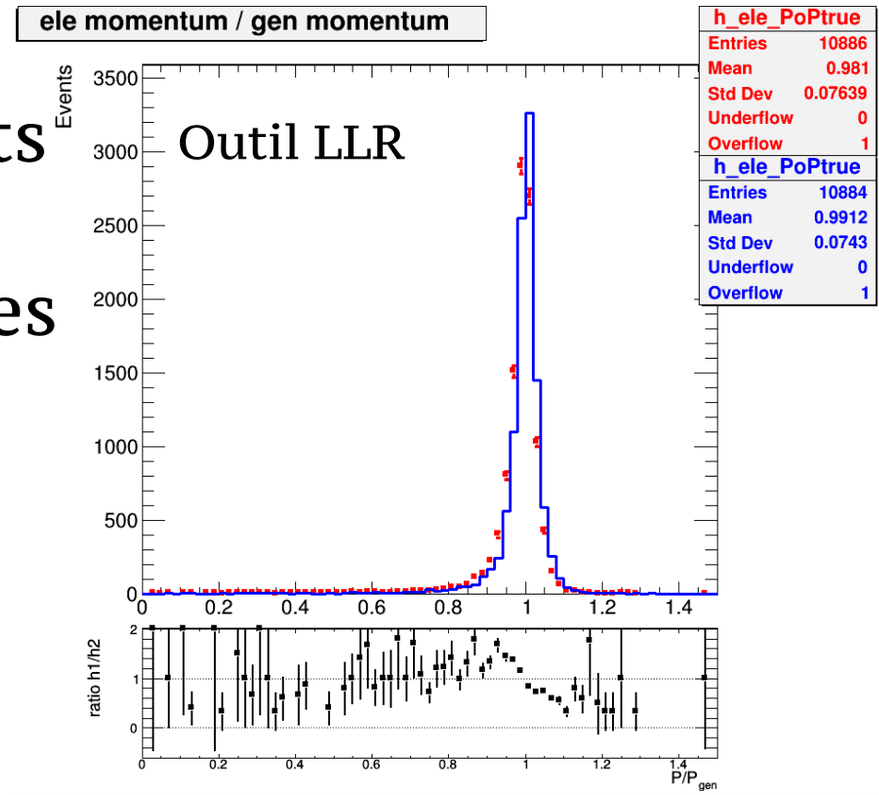
- il est de plus en plus difficile pour un physicien de contribuer au code. A fortiori pour un étudiant en début de thèse…
- courbe d'apprentissage très raide
  (à titre personnel, j'ai renoncé à contribuer à CMSSW pendant ~2 ans du fait du passage au git et à l'intransigeance des release managers)
- Le niveau d'expertise nécessaire est tel que le codage occasionnel par un physicien est très inefficace

- L'expérience des informaticiens est cruciale pour l'écriture d'un code de qualité. La définition et le suivi par les physiciens sont indispensables.
  ⇒ forte synergie

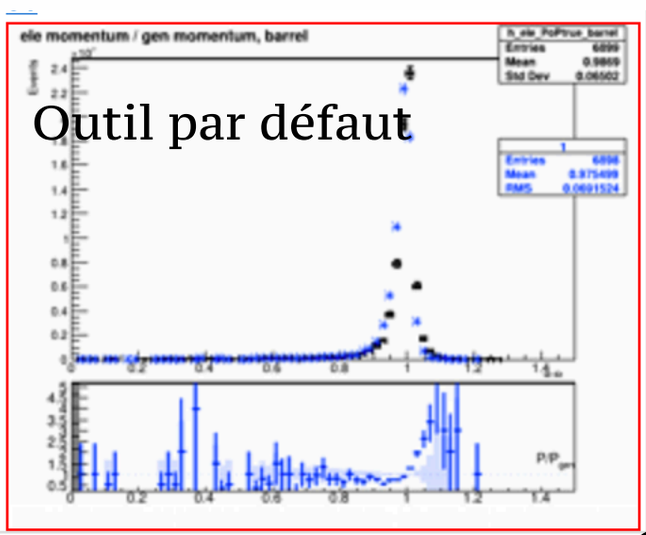
Outil de validation de la reconstruction des électrons

Pour chaque nouvelle release du programme de reconstruction, les performances de physique sont validées.

- les fichiers contenant les histos sont produits par CMS
- A. Chiron produit les pages web contenant les comparaisons
  - ★ L'outil par défaut de CMS peu pratique pour les physiciens, et peu flexible
- inspection & rapport par un physicien



Le code qui produit les histogrammes pour les électrons est de la responsabilité du LLR.
Il évolue régulièrement (multi-threading, nouveaux histos,...) en fonction des demandes et suggestions des physiciens du groupe et de CMS.
→axe actuel: détection de défaut par ML



Tier 2 & Tier 3

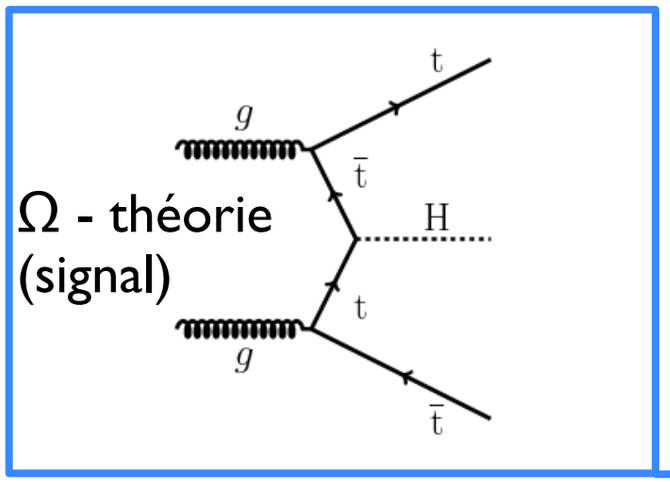
En complément des informations données par Andrea, il faut souligner que l'impact des analyses du groupe doit largement à la présence du Tier2 et du Tier 3

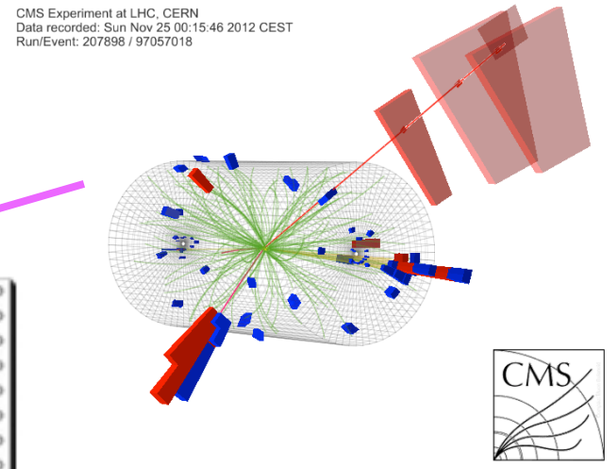
- Accès rapide et efficace aux données
- Fiabilité
- Flexibilité
- Maîtrise du temps

Impact majeur lors de la découverte et l'étude du boson de Higgs car le Tier 2 accueille officiellement les datasets correspondants.

Analyse ttH avec élément de matrice sur GPU

Des variables discriminantes de haut niveau sont construites à partir d'intégrales multi-dimensionnelles






```
Event 97057018
==============
muon pT : 32 GeV
τ pT : 44 GeV
jet 1 : 80 GeV
jet 2 : 36 GeV
...
```

Matrix Element

simplified formula

$$P(\mathbf{x}|\Omega) = \frac{1}{\sigma_\Omega} \iiint dx_1 dx_2 d\mathbf{y} \mathcal{P}_s(x_1, x_2) |\mathcal{M}_\Omega(x_1, x_2, \mathbf{y})|^2 W(\mathbf{x}, \mathbf{y})$$

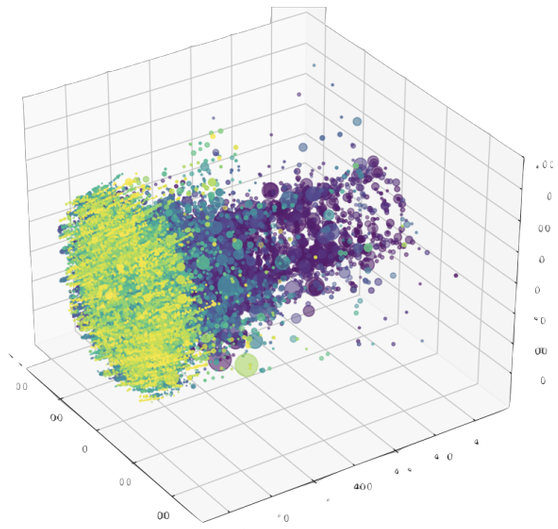
Réponse du détecteur

G. Grasseau impliqué dès le début du projet pour que le code se prête à la parallélisation.
Mise au point au labo (MPI → GPU) pour les 2 première thèses, production au CC pour la thèse en cours.

Clairement, un expert est indispensable tout au long du projet

- CHEP 2015, 2018 et plusieurs autres conférences

Upgrade [HGCAL]



Déclenchement et reconstruction dans HGCAL
- haute granularité → millions de canaux répartis sur des dizaines de couches donnent lieu à une combinatoire gigantesque
- multiplicité des particules incidentes

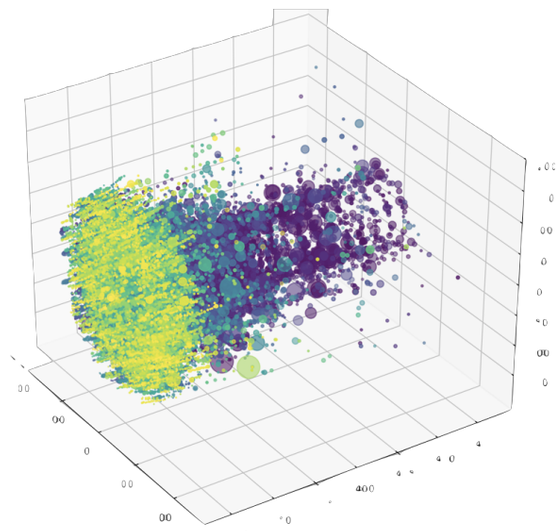
constituent un défi majeur à la fois pour le déclenchement
et pour la reconstruction

La maîtrise de la consommation CPU doit d'emblée être prise en compte
- nécessité de mise en commun des réflexions des physiciens et des ingénieurs
- terrain de jeu privilégié pour le parallèlisme (expertise)

Upgrade [HGCAL] et ML

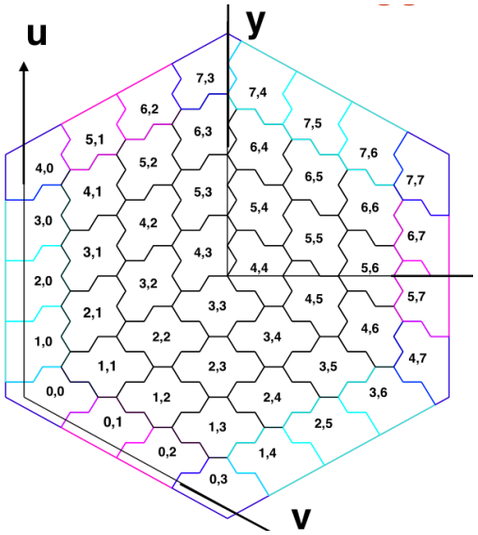
Déclenchement et reconstruction dans HGCAL
- haute granularité → millions de canaux
- multiplicité des particules incidentes

constituent à la fois un défi et un exemple parfait pour les techniques de Machine Learning.



Des approches similaires sont envisagées pour le déclenchement et la reconstruction.

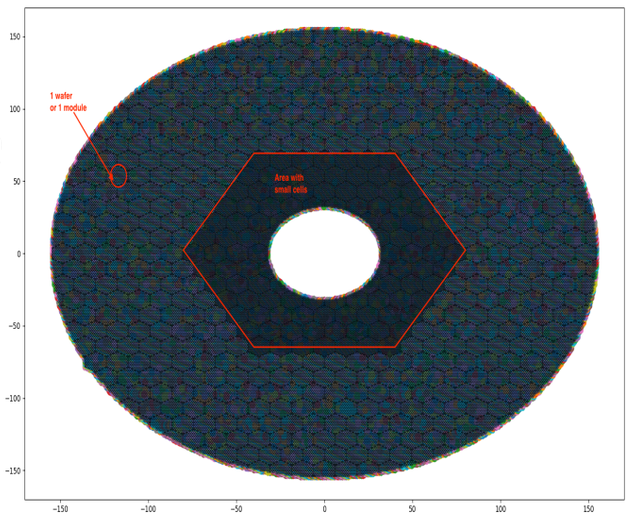
Cellules de déclenchement



Déclenchement: identifier les candidats électromagnétiques de haut pT

Hors-ligne: reconstruire les clusters

Géométrie “offline"



→Développement d’une plateforme d’entraînement commune pour le ML

Conclusion

La complexité croissante des techniques mises en oeuvre à tous les niveaux (déclenchement, reconstruction, analyse…) ainsi que les contraintes de plus en plus fortes font que les collaborations entre physiciens et informaticiens qui ont toujours existé et été fructueuses par le passé sont en train de devenir indispensables.

Les axes de développements présentés par Gilles sont pleinement cohérents avec cette évolution.