

Artificial Intelligence in High Energy Physics II

David Rousseau

IJCLab-Orsay

[**rousseau@ijclab.in2p3.fr**](mailto:rousseau@ijclab.in2p3.fr)

*Rencontres d'été de physique
de l'infiniment grand à l'infiniment petit
juillet 2023*



Apprentissage profond

'Godfathers of AI' honored with Turing Award, the Nobel Prize of computing

3

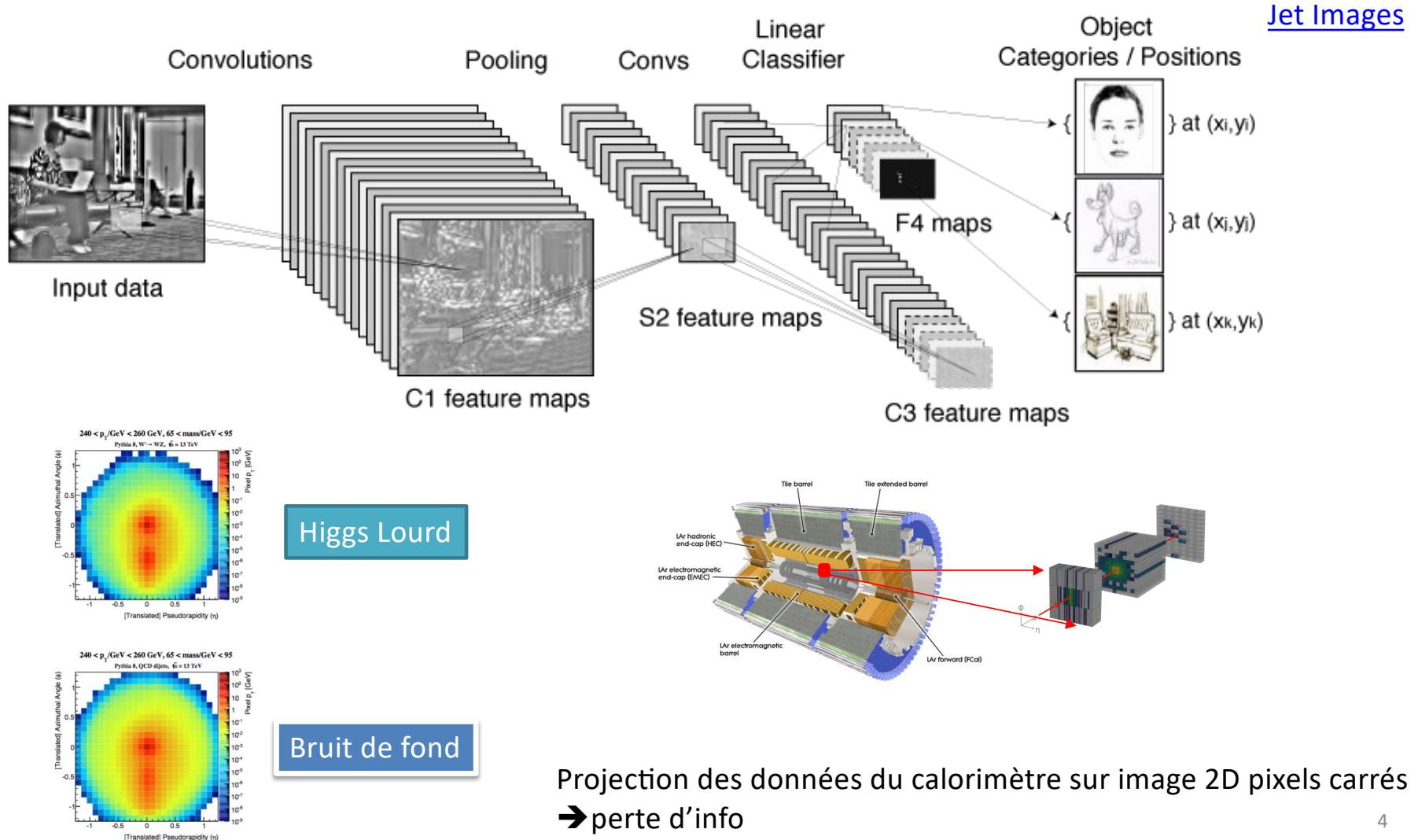
Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI

By James Vincent | Mar 27, 2019, 6:02am EDT

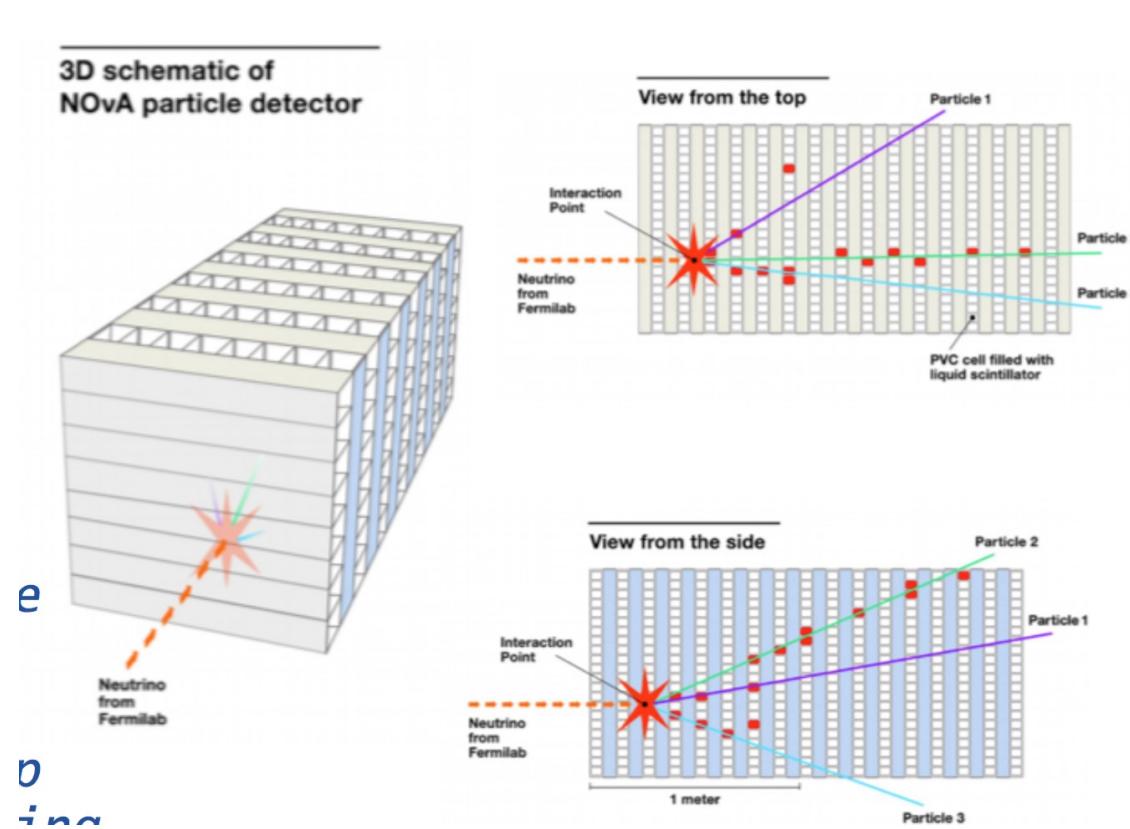
2019



Réseau de Neurones Convolutifs (CNN)



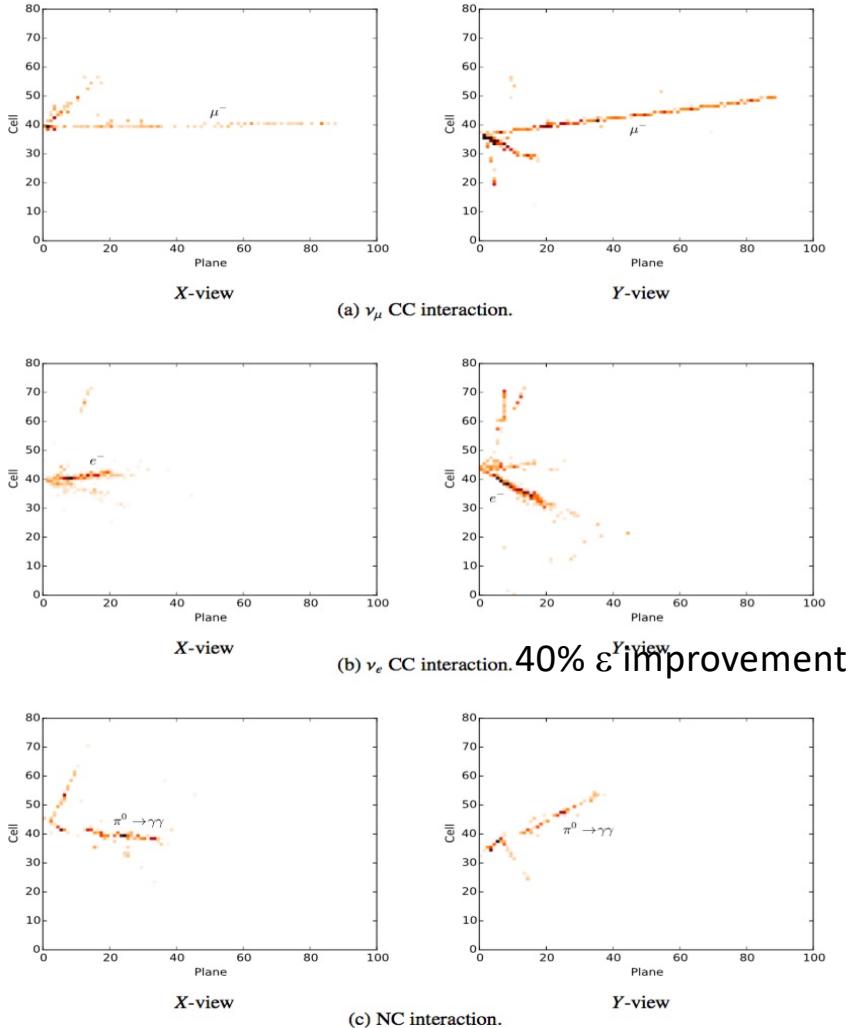
Un succès du Deep Learning : NOVA



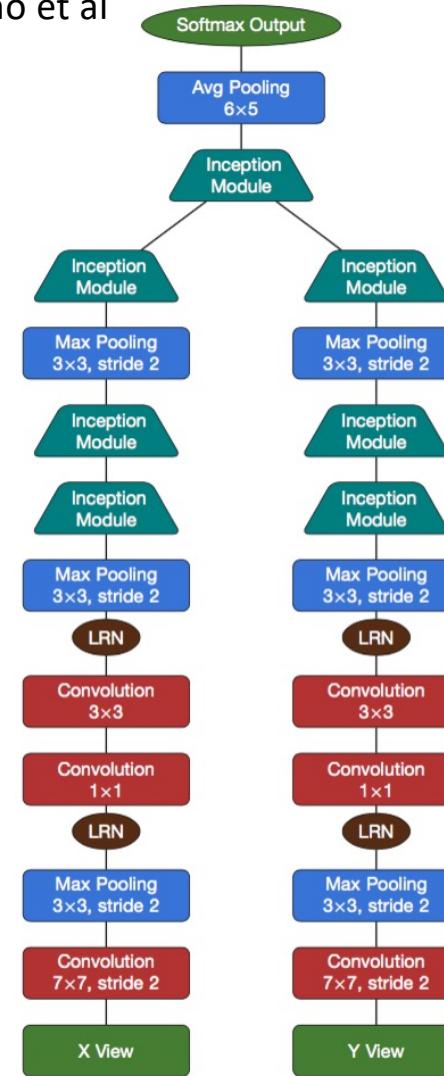


NOVA (2)

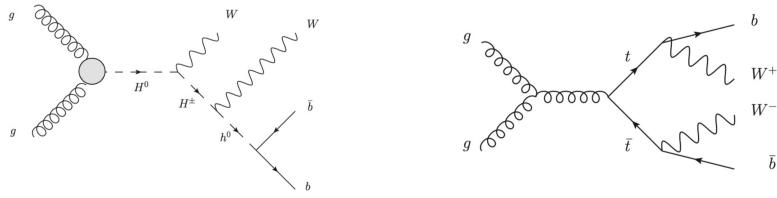
[arXiv 1604.01444](https://arxiv.org/abs/1604.01444) Aurisano et al



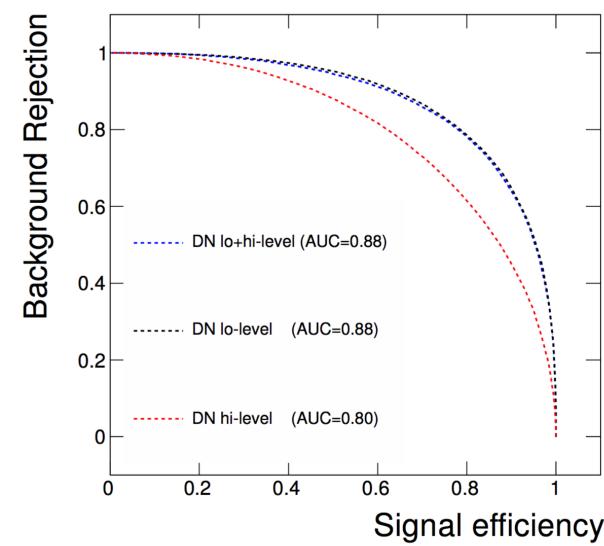
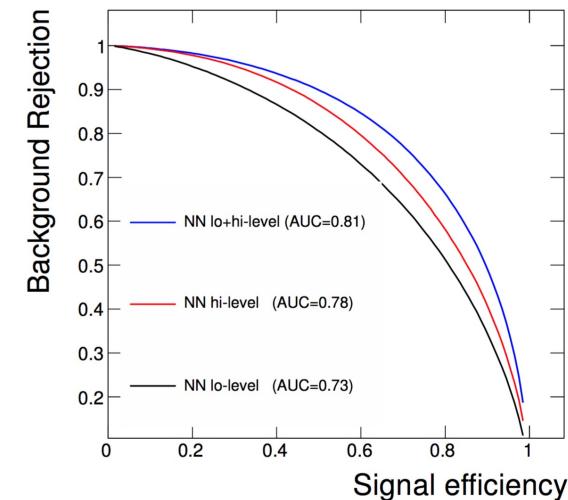
Neutrino interaction classification
Using Convolutional Neural Network (GoogLeNet)
Actually used in physics results 1703.03328 and 1706.04592



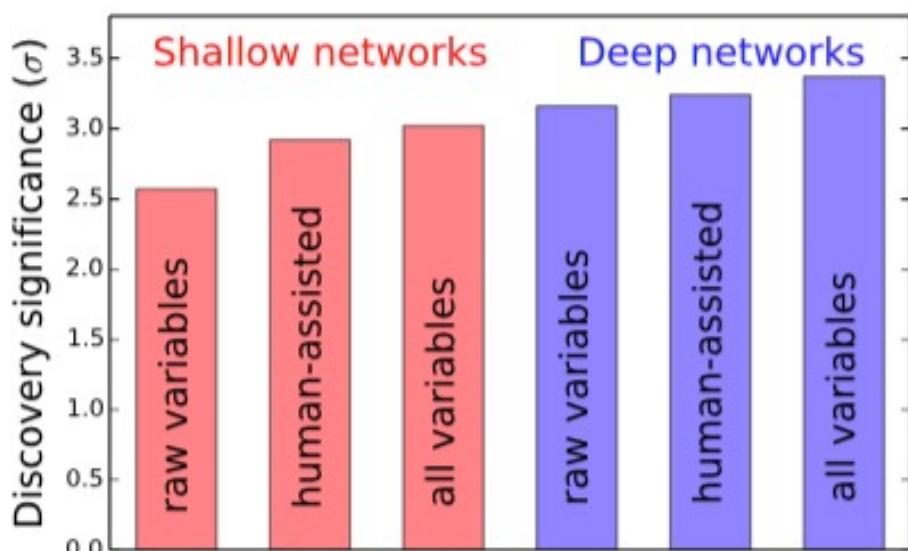
[1402.4735](#) Baldi, Sadowski, Whiteson



- Recherche d'un Higgs rare : $H^0 \rightarrow WWbb$ vs $tt \rightarrow WWbb$
- Simulation simplifiée d'événements LHC
- Variables de bas niveau
 - Quadri-vecteur particules
- Variables de haut niveau
 - Masses invariantes, angles etc...
- \rightarrow DNN (Deep Neural Network) marche mieux que réseau simple
- \rightarrow DNN n'a pas besoin de variables de haut niveau
- \rightarrow le DNN apprend la physique ????



- Analyse H tautau : $H \rightarrow \tau\tau$ vs $Z \rightarrow \tau\tau$
 - Bas niveau
 - Haut niveau



- ❑ DNN marche mieux que réseau simple
- ❑ Mais cette fois il a besoin de variables de haut niveau !
- ❑ Ca se complique...
- ❑ Pour l'instant, pas de résultat de physique utilisant DNN sauf ...

Tracking challenge

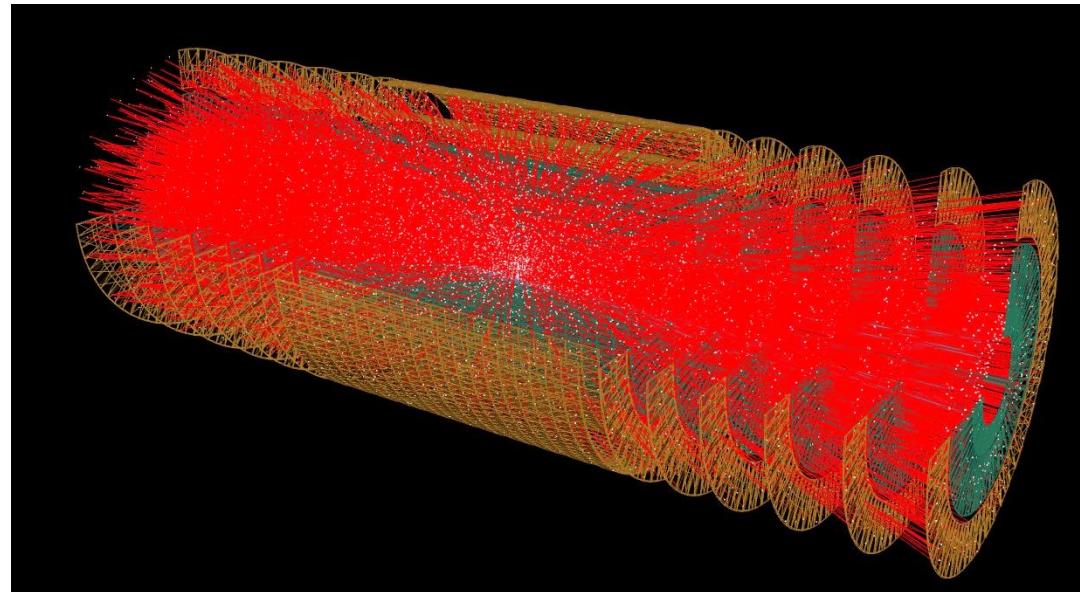
[**https://sites.google.com/site/trackmlparticle/**](https://sites.google.com/site/trackmlparticle/)

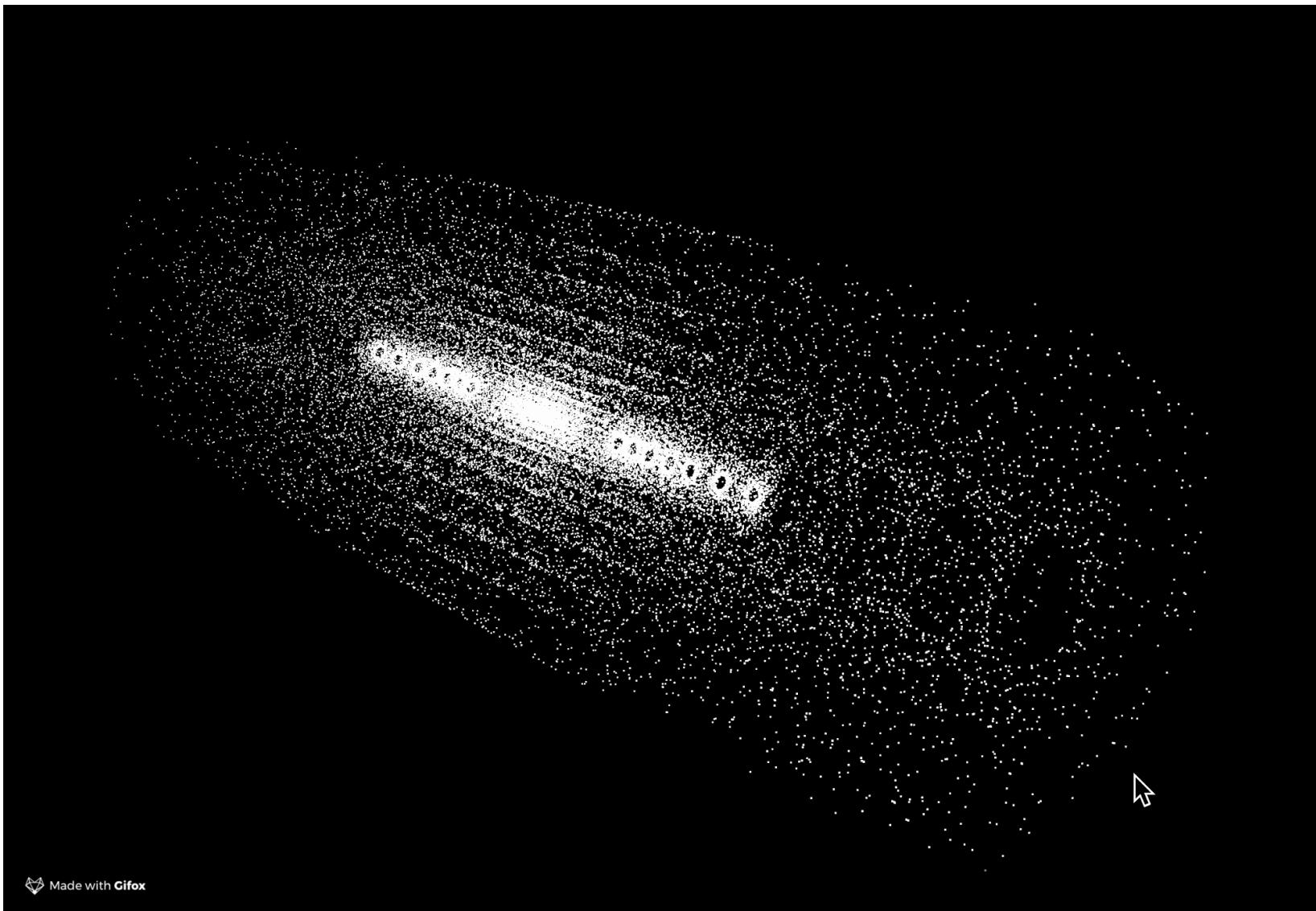
[**https://www.kaggle.com/c/trackml-particle-identification**](https://www.kaggle.com/c/trackml-particle-identification)

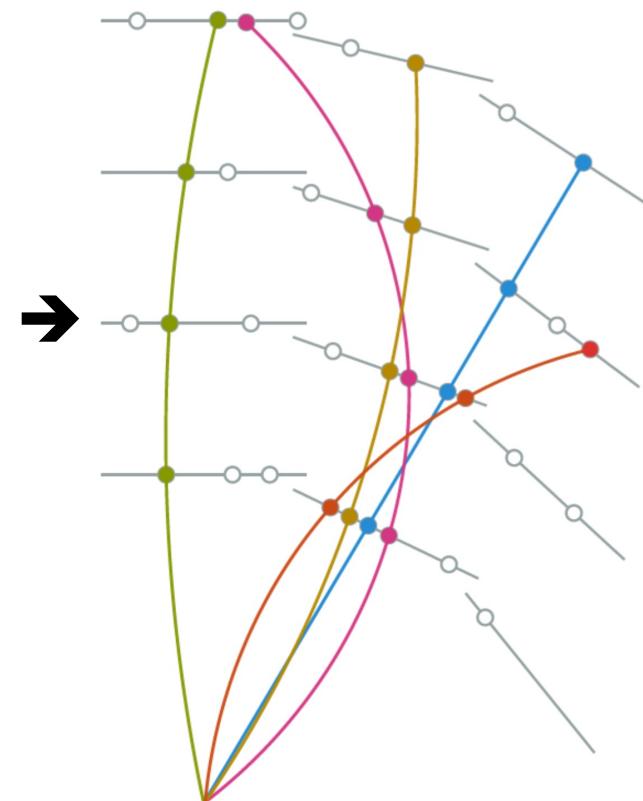
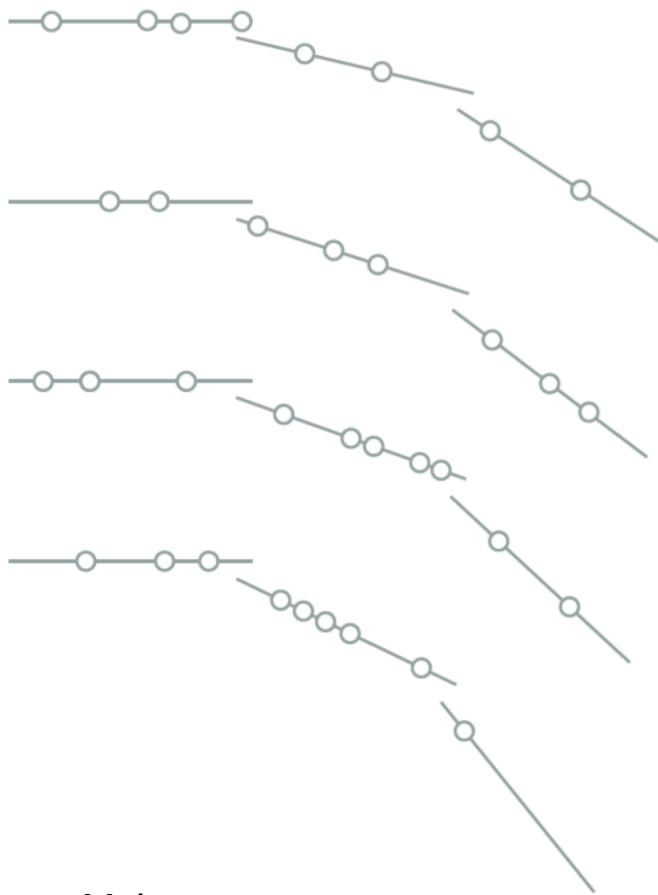
[**https://twitter.com/trackmllhc**](https://twitter.com/trackmllhc)

L'enjeu

- LHC en 2029
- Augmentation du nombre de collisions de proton
- →superposition d'événements ~200 (contre ~50 maintenant)
- →événements très complexes
- En particulier pour la trajectographie : relier les points
- Algorithmes actuels ~10 fois trop lents







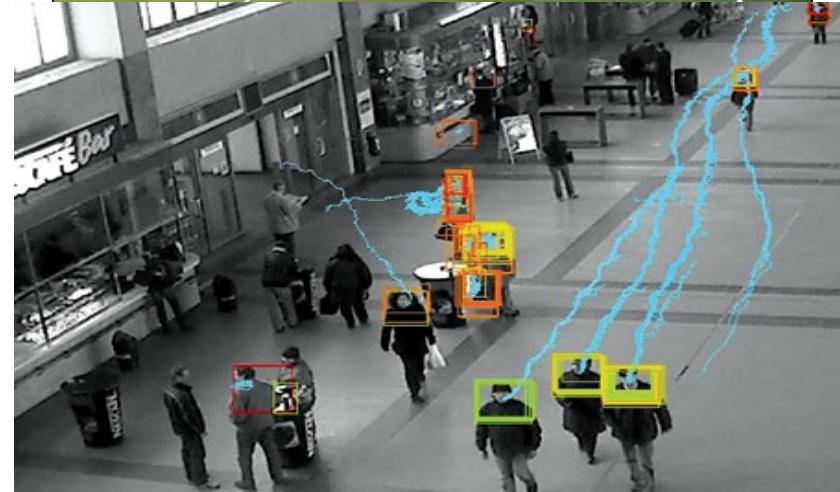
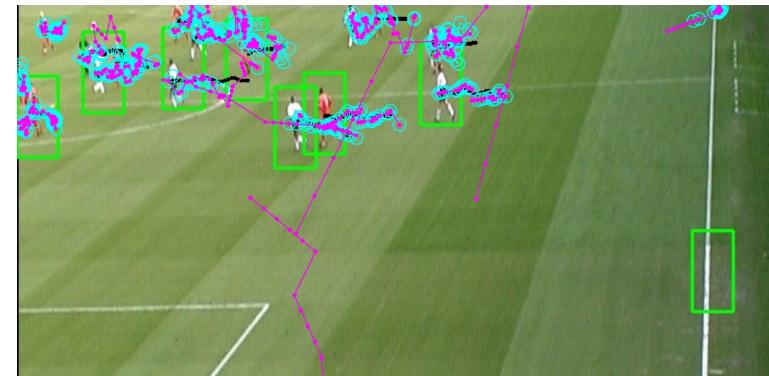
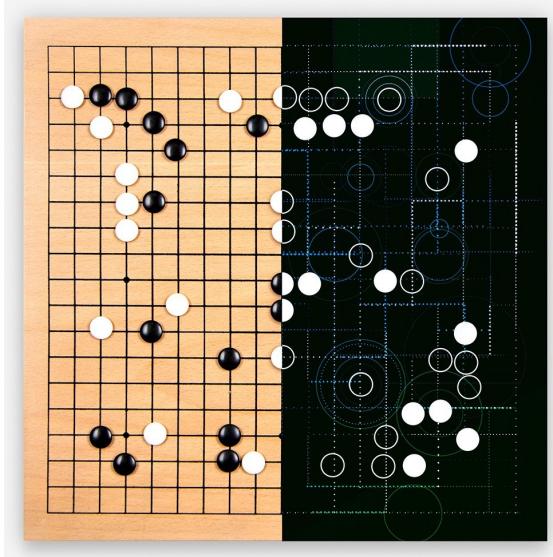
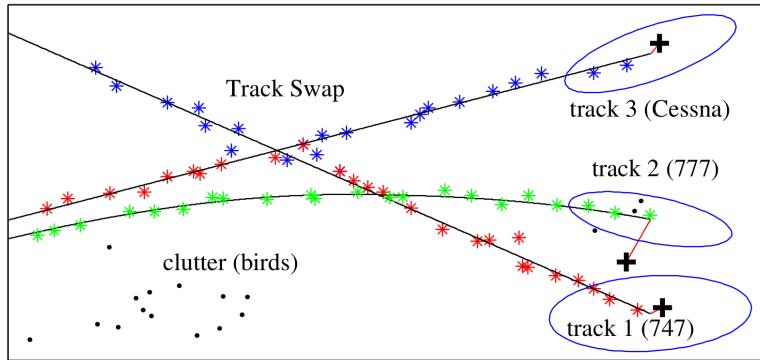
Mais :

- 3 dimensions
- 10'000 traces x 10 points

Pourquoi c'est difficile ?

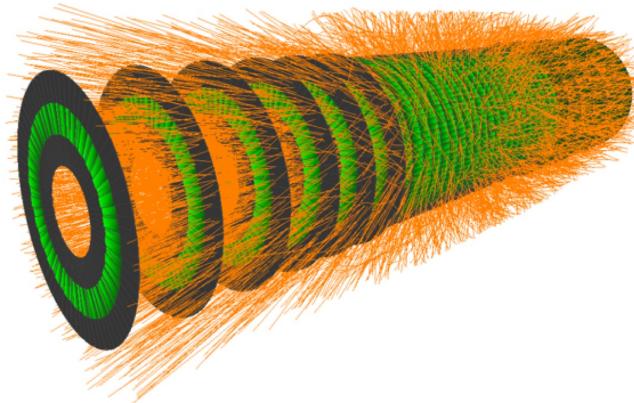
- 100'000 points à grouper en 10'000 traces de 10 points
 - $\rightarrow \sim 10^{450'000}$ combinaisons
 - \Rightarrow la “force brutale” n'a (vraiment) aucune chance
- Precision des points : $\sim 50\mu\text{m}$ sur un volume $\sim 40\text{ m}^3$
 - $\rightarrow 3 \cdot 10^{14}$ voxels!
 - projection 2D $\rightarrow 2 \cdot 10^9$ pixels !
 - \Rightarrow algorithme de traitement d'image n'ont (vraiment) aucune chance
- Pas un problème classique

Trajectographie (Tracking)



Tracking Machine Learning challenge

<https://twitter.com/trackmllhc>



Idée : mettre les points (x, y, z) sur le web

demander aux participants de relier les points

Objectif : obtenir de nouveaux algorithmes

- Accuracy phase sur Kaggle, uniquement la précision : mai-aout 2018
- Throughput phase (compétition officielle NeurIPS 2018) précision + vitesse : octobre 2018-mars 2019

Conclusion : Machine Learning aided reconstruction

Visualisation spin-off

- Visit at CERN Tobias Isenberg visualisation scientist at LRI-Orsay with PhD student Xiyaow Wang
- Will use TrackML dataset to experiment with visualisation/interaction with Microsoft' Hololens



Aparté historique.

Le tout premier article Intelligence Artificielle appliqué
à la physique des particules!

Computer Physics Communications 49 (1988) 429–448
North-Holland, Amsterdam

NEURAL NETWORKS AND CELLULAR AUTOMATA IN EXPERIMENTAL HIGH ENERGY PHYSICS

B. DENBY

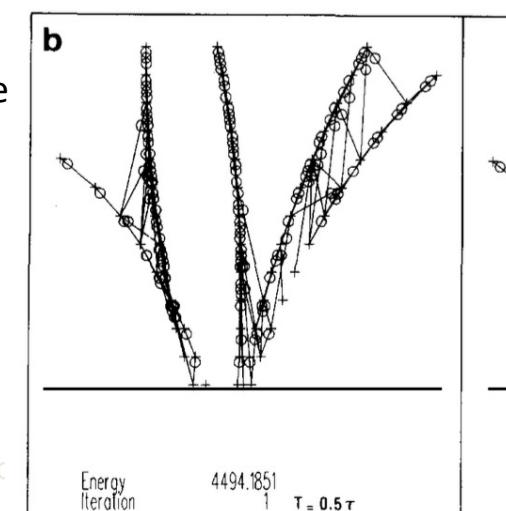
Laboratoire de l'Accélérateur Linéaire, Orsay, France

Received 20 September 1987; in revised form 28 December 1987

- ❑ Bruce Denby post doc au LAL-Orsay sur Delphi (expérience sur le LEP au CERN)
- ❑ Puis Post-doc à Fermilab sur Tevatron
- ❑ Quitte HEP en 1993 (annulation de SSC)
- ❑ Actuellement Professeur U Sorbonne, Institut Langevin Ondes et Images): “silent speech” (analysis of ultrasonic video of the tongue) [slides](#) [video](#)

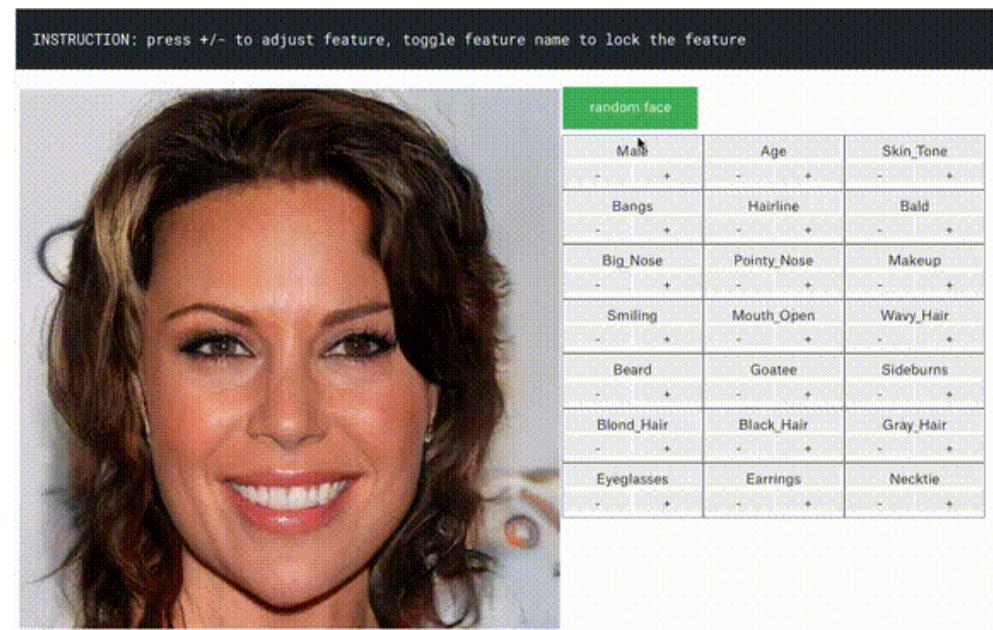


Bruce Denby

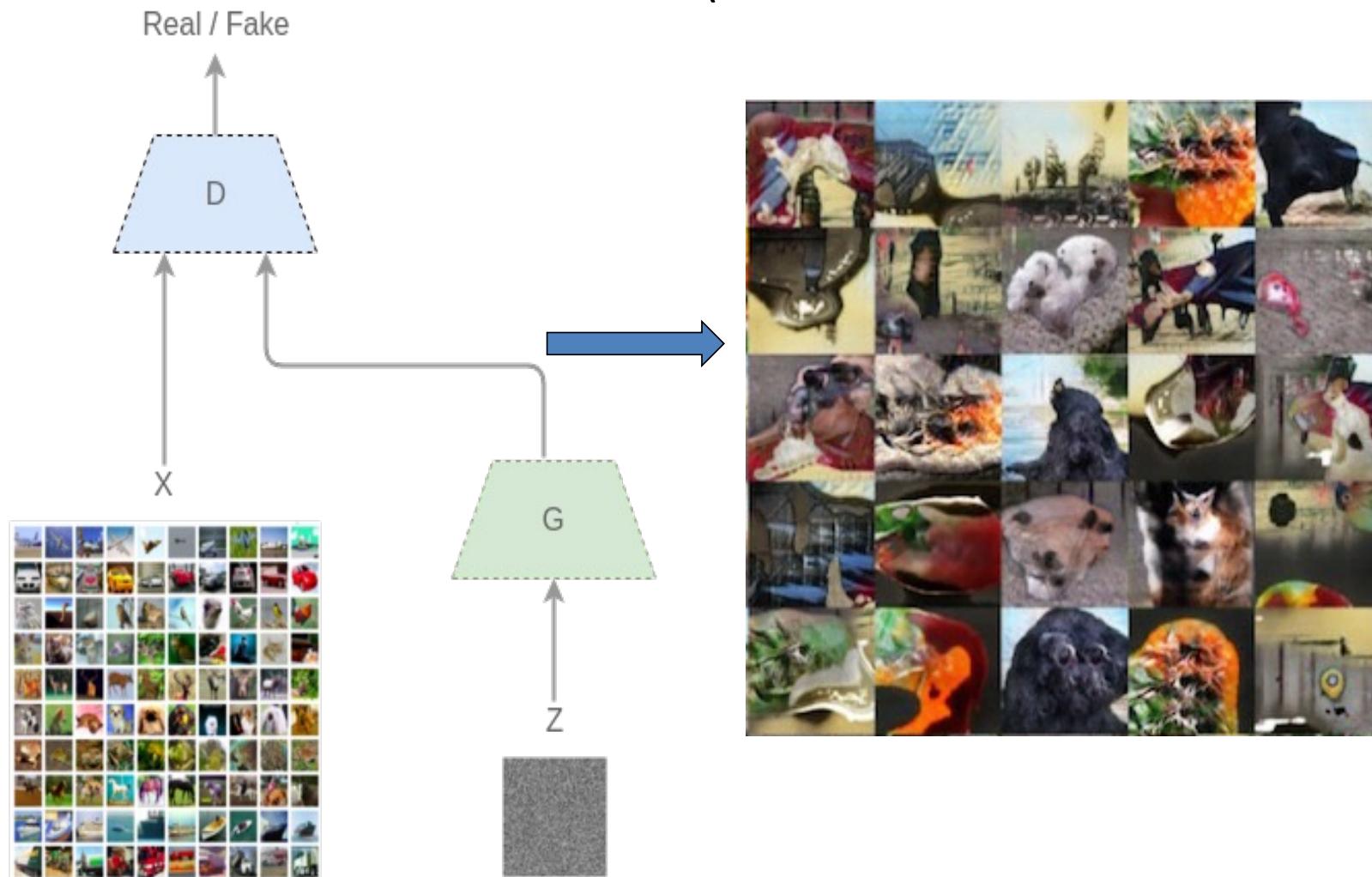


Generative Adversarial Network

Generative model



Réseau Génératif Adversaire (Generative Adversarial Network GAN)



Condition GAN

Text to image

this small bird has a pink breast and crown, and black primaries and secondaries.



this magnificent fellow is almost all black with a red crest, and white cheek patch.



the flower has petals that are bright pinkish purple with white stigma



this white and yellow flower have thin white petals and a round yellow stamen



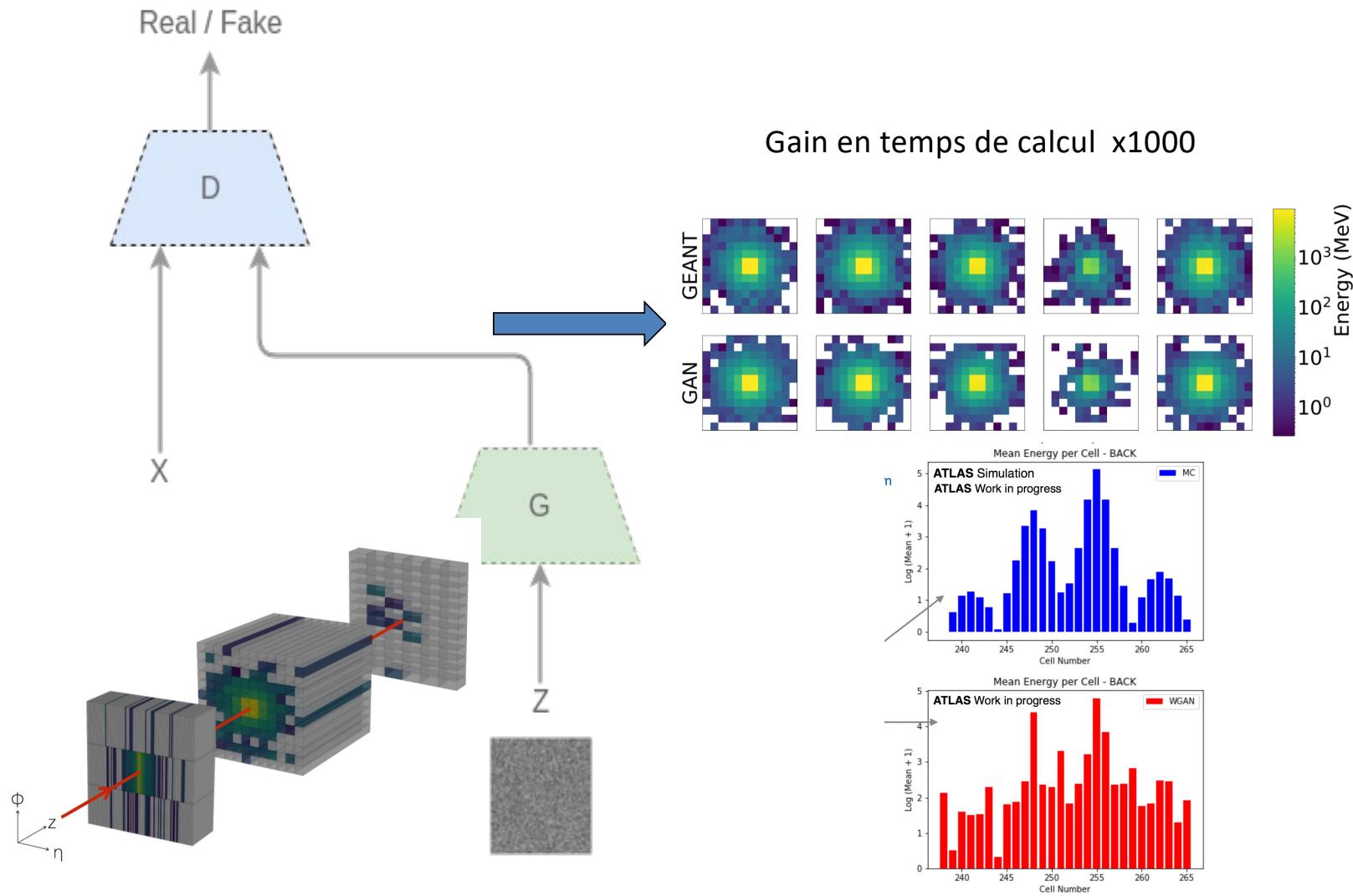
Text to image

Dall.e

2022 « Large Foundation Model »



GAN pour la simulation de particules



La méthode scientifique

Publication scientifique

[Phys.Rev.Lett. 114 \(2015\)191803](#)

Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments

(ATLAS Collaboration)[†]

(CMS Collaboration)[‡]

(Received 25 March 2015; published 14 May 2015)

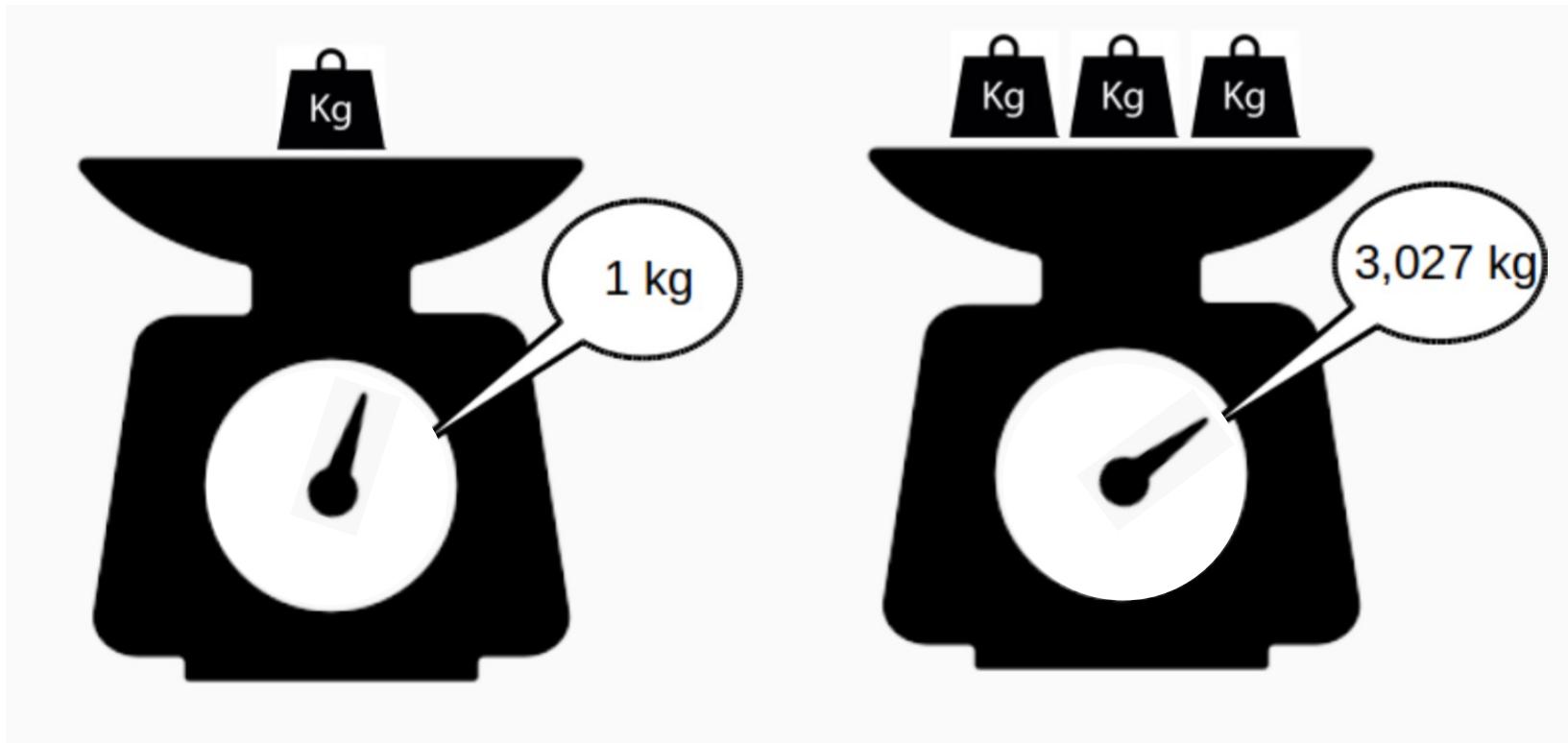
A measurement of the Higgs boson mass is presented based on the combined data samples of the ATLAS and CMS experiments at the CERN LHC in the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4\ell$ decay channels. The results are obtained from a simultaneous fit to the reconstructed invariant mass peaks in the two channels and for the two experiments. The measured masses from the individual channels and the two experiments are found to be consistent among themselves. The combined measured mass of the Higgs boson is $m_H = 125.09 \pm 0.21 \text{ (stat)} \pm 0.11 \text{ (syst)} \text{ GeV}$.

Incertitudes systématiques :

tout ce dont on n'est pas sur...

Et si ? Et si ? ... Les inconnues connues et inconnues jusqu'à ce que nos collègues soient convaincus

Biais expérimental



Pres de 500 termes de systématiques

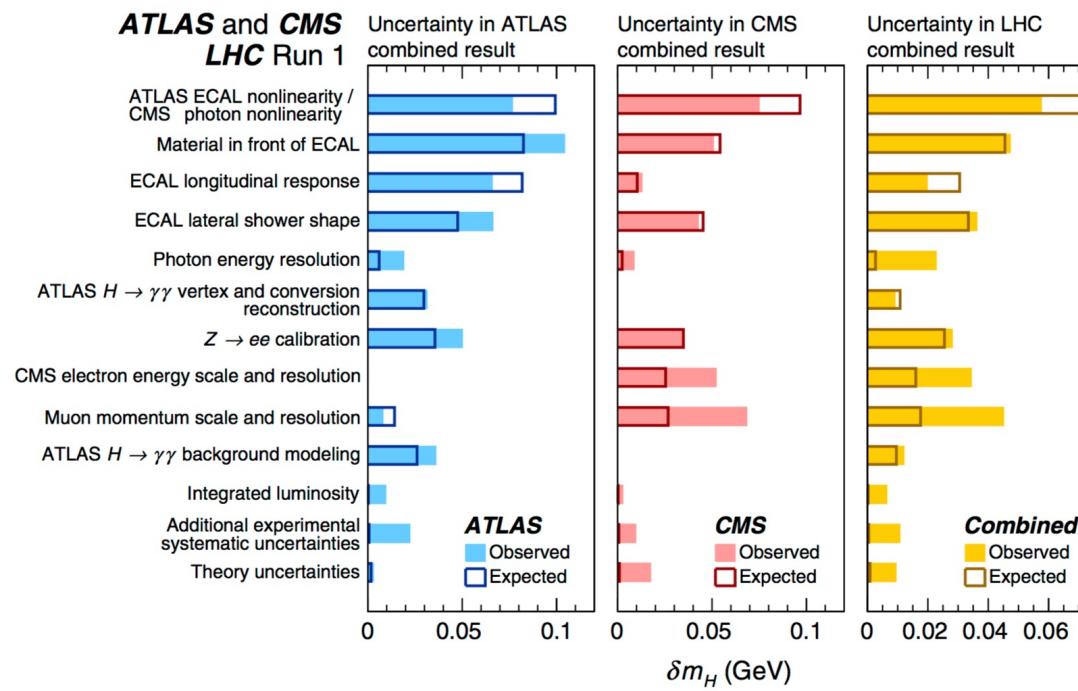


FIG. 3 (color online). The impacts δm_H (see text) of the nuisance parameter groups in Table I on the ATLAS (left), CMS (center), and combined (right) mass measurement uncertainty. The observed (expected) results are shown by the solid (empty) bars.

>>>89-90: Where is this measurement documented in this note?

*That has now been finished and published, and references will be updated in the notes
JHEP 09 (2014) 079 arXiv:1407.5532

>>>116: Doesn't "3 GeV" contradict your previous statement that we cover a pT range "from 1 GeV..."
**No, those muons with pT~1 GeV and p~3 GeV are at high eta. But this is all irrelevant, as we only use muons with pT>4 GeV.

>>>116: I'm not sure what you mean when you say that they "require" the combined reconstruction. I would just say "Muons used in this analysis are reconstructed using a statistical combination of an MS track and ID track."

**Only the muons that have a successful statistical combination of an MS track and ID track are used in this analysis, the text is updated.

>>>128: what is fully efficient for signal candidates? It's not clear what the word "which" refers to.
**This related to the vertex quality criteria applied to the fit.

>>>142: Please mention the total number of di-muon candidates here.

**This is added to the text, 7.8M for 2011, and ~65M for 2012.

>>>171: could you provide more information about how the psi is assigned to a primary vertex? I'm surprised by your statement that "few" events contain multiple vertices; I thought pileup was a significant issue at 7 TeV.

** Not really. The only relevance the primary vertex has is to measure Lxy, which is measured in the transverse direction only, and hence is not changing much from one collision vertex to another. However, the determination of the primary vertex position depends on whether the two muon tracks were used in its fit or not, hence we need to know which vertex was it. But at 7 TeV there was not much ambiguity, the vertex which jpsi came from was almost always the main primary one. Studies from the 2011 Jpsi Phi analysis -- where vertex choice may have been an issue -- showed that there was no impact in the few cases of an incorrect choice of vertex.

>>>Fig 1: Do you understand the eta dependence of this plot? What is the z-axis -- number of events per bin? The bins are extremely small -- what size bins did you choose?

** Yes. This is the scatter plot of dimuon candidates. The x-axis is the absolute rapidity "y" of the dimuon candidate, with the structure roughly reflecting the (smeared) geometry of the muon chambers, with dips near/around the cracks and edges.

The z-axis is just the candidate yield, the bin-width in $|y|$ is $5e-3$ and in pT is 320 MeV

>>>228: The phi* definition is unclear to me, in particular what the "psi production" is. Do you mean the psi momentum vector?

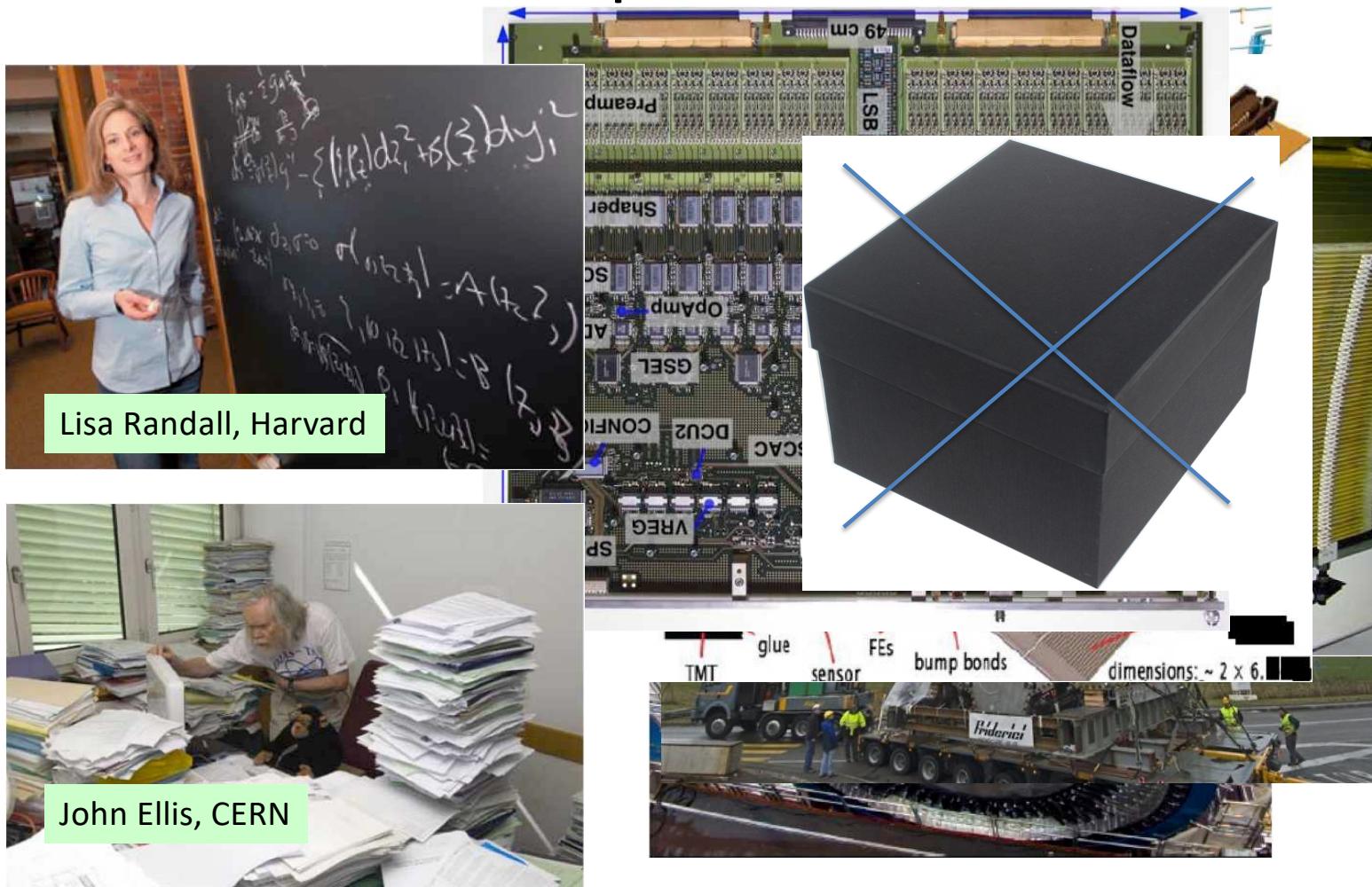
**phi* is the angle between the psi production plane (defined by psi momentum and colliding proton

ue

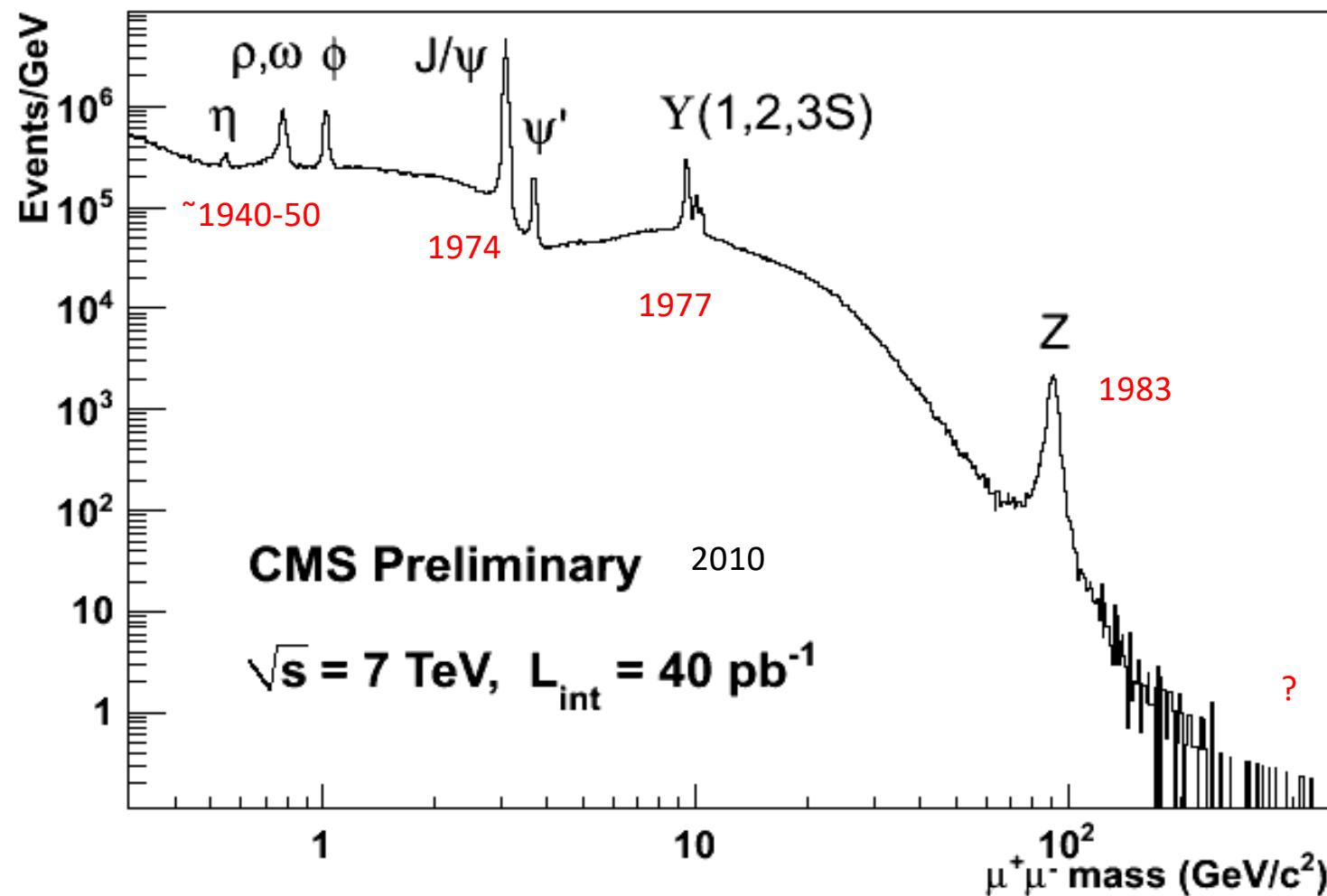


Près de 200 pages et plusieurs semaines de questions/réponses avant l'approbation de la publication !

Trust but verify : de la théorie à l'expérience



Ré-observation de particules connues

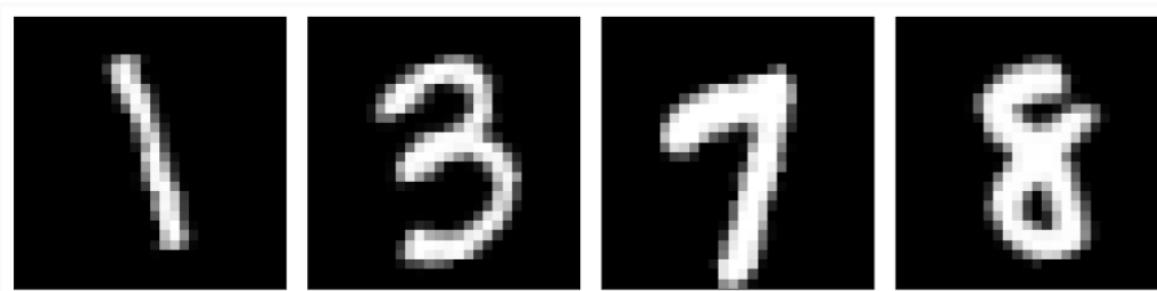
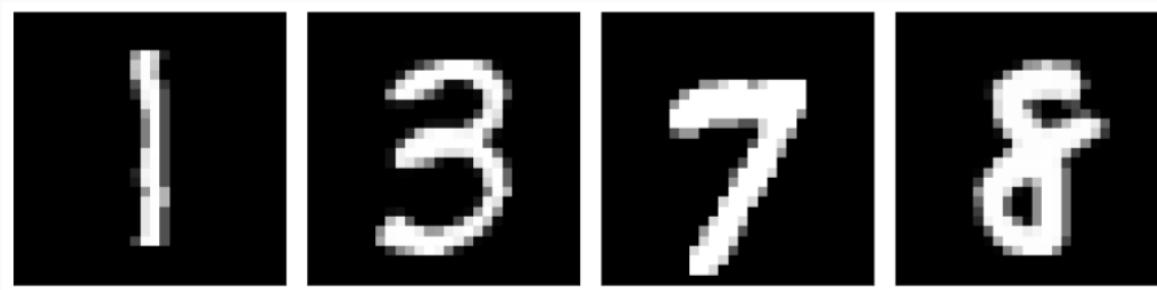


ML et erreur systématique

- Nos papiers expérimentaux se concluent souvent par
 - $\text{measure} = m \pm \sigma(\text{stat}) \pm \sigma(\text{syst})$
 - $\sigma(\text{syst})$ erreur systématique : inconnues connues, inconnues inconnues
 - On cherche à minimiser la somme quadratique:
$$\sigma(\text{stat}) \pm \sigma(\text{syst})$$
 - Technique ML standard minimise $\sigma(\text{stat})$
- Comment indiquer au ML de minimiser $\sigma(\text{stat}) \pm \sigma(\text{syst})$?

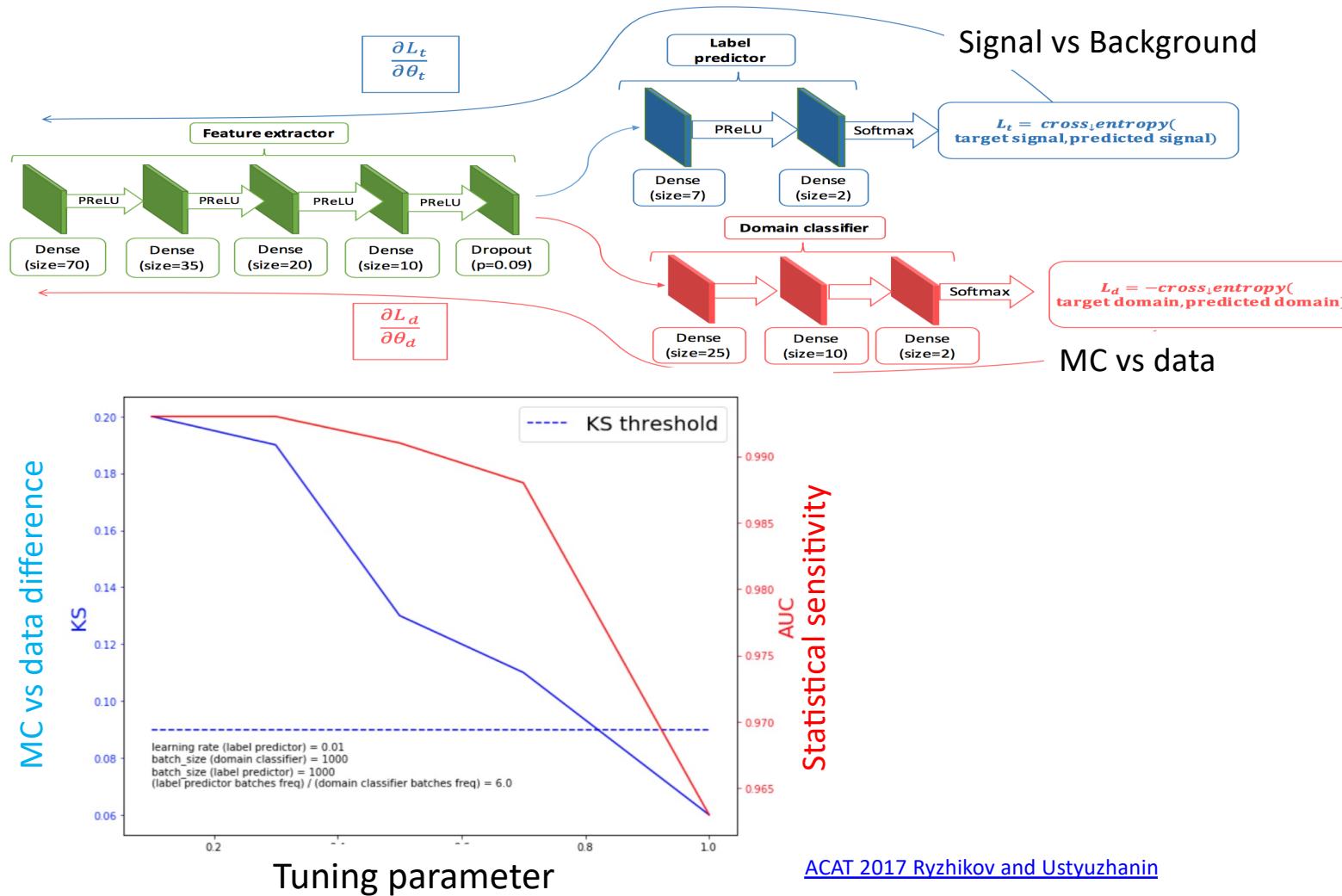
Effet systématique

Entrainement sur chiffres manuscrits. Effet de l'angle

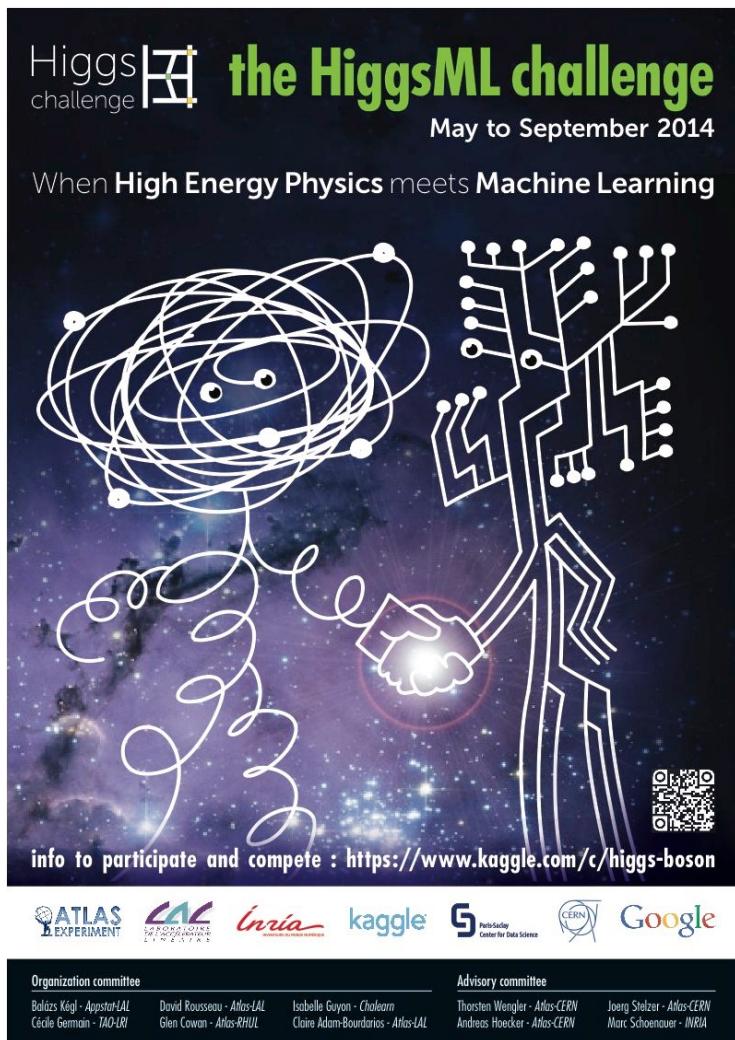


Entrainement adversaire

Inspired from 1505.07818 Ganin et al :



Higgs Machine Learning challenge

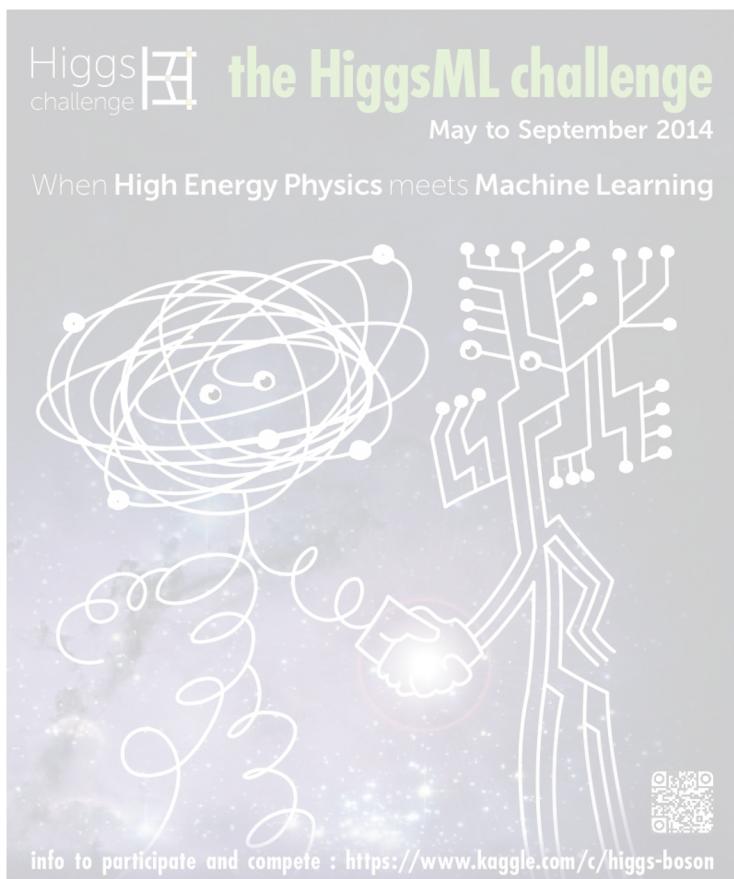


Idée : simplifier une analyse
Higgs, poster les données
simulées sur le web,
demander aux spécialistes
de l’Intelligence Artificielle
de l’améliorer

Organisé en 2014 sur la
plateforme [Kaggle](#)

Fair Universe 2024

- Collaboration U Berkeley U Washington U Paris-Saclay Chalearn
- Même principe que HiggsML mais en incluant les incertitudes systématiques



Higgs challenge **the HiggsML challenge**
May to September 2014
When High Energy Physics meets Machine Learning

info to participate and compete : <https://www.kaggle.com/c/higgs-boson>

ATLAS EXPERIMENT CERN Inria kaggle Paris-Saclay Center for Data Science Google

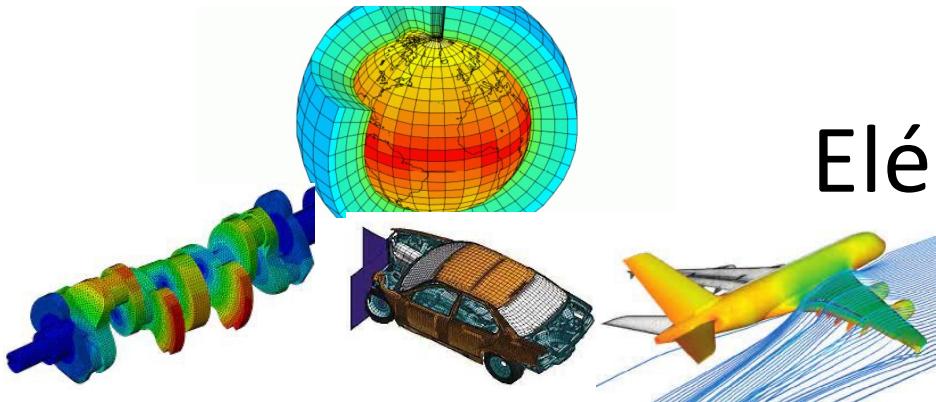
Organization committee

Balázs Kégl - <i>Appstat-LAL</i>	David Rousseau - <i>Atlas-LAL</i>	Isabelle Guyon - <i>Chalearn</i>
Cédile Germain - <i>TAO-LRI</i>	Glen Cowan - <i>Atlas-RHUL</i>	Clémie Adam-Boudard - <i>Atlas-LAL</i>

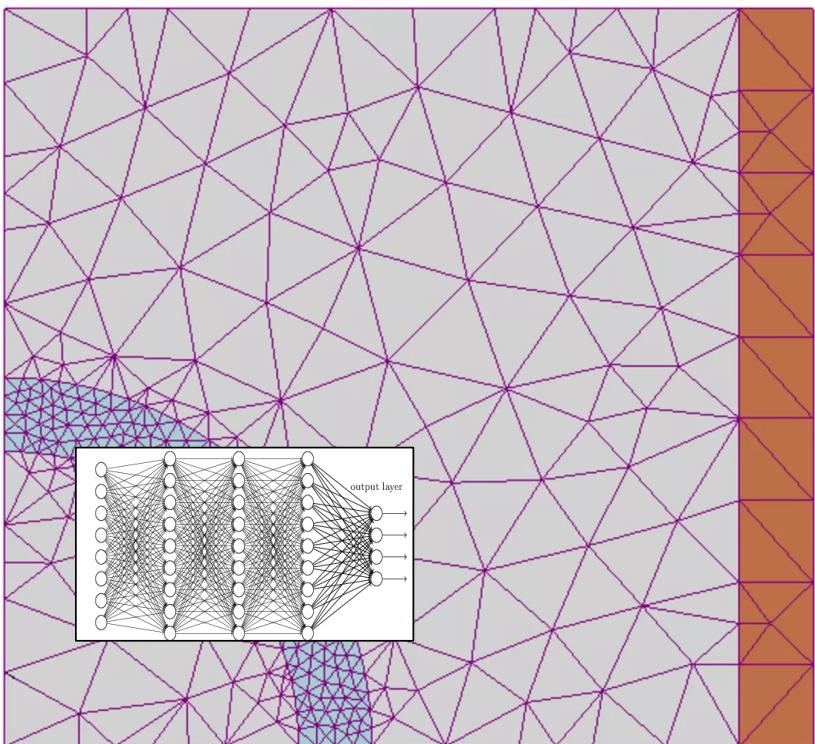
Advisory committee

Thorsten Wengler - <i>Atlas-CERN</i>	Joerg Stelzer - <i>Atlas-CERN</i>
Andreas Hoecker - <i>Atlas-CERN</i>	Marc Schoenauer - <i>INRIA</i>

Calcul multi-échelles



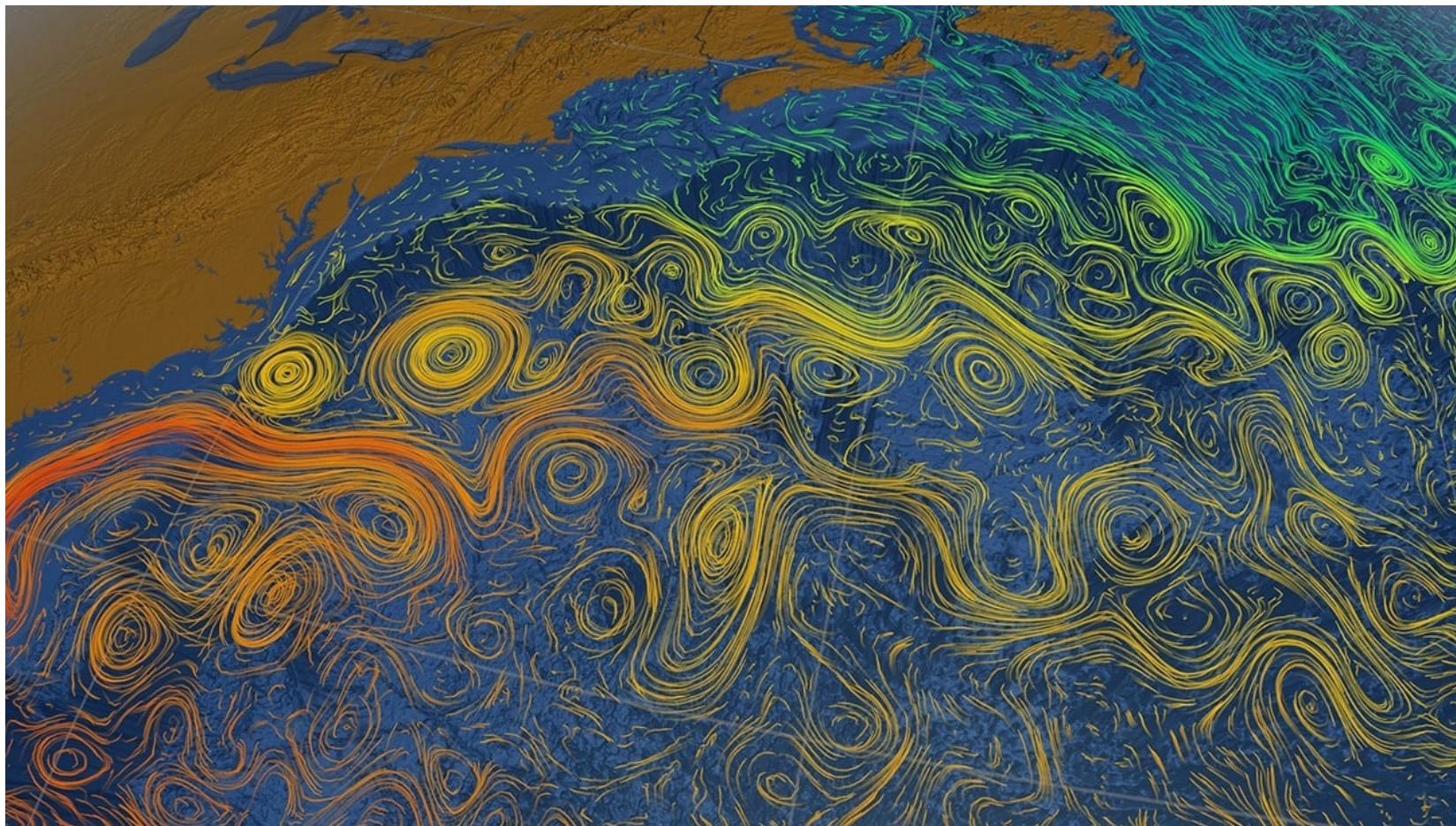
Eléments finis



- ❑ Technique ubiquitaire
- ❑ Précision limitée par taille de maille
- ❑ maille / 2 → ressources x 8
(mémoire, temps de calcul)
- ❑ Au lieu de réduire la maille,
entrainer un NN à reproduire un
modèle à petite maille

Application océanographie

[NASA](#)

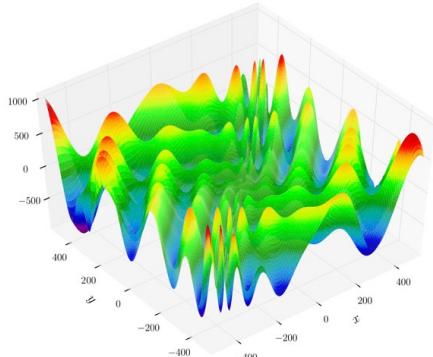
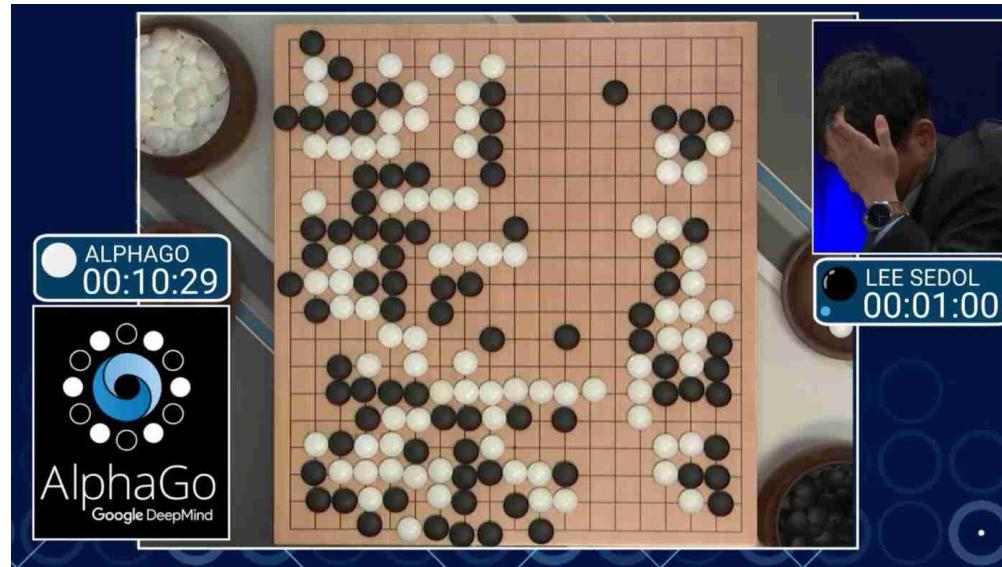
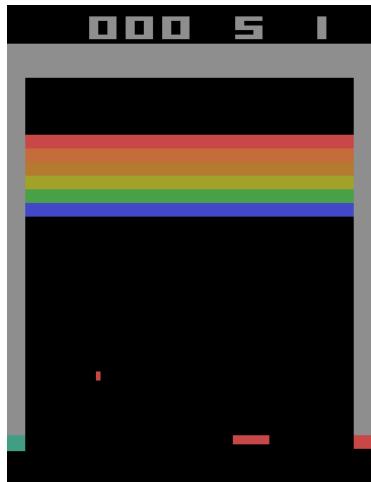


IA et Sciences, David Rousseau, Avril 2023, CSNUM

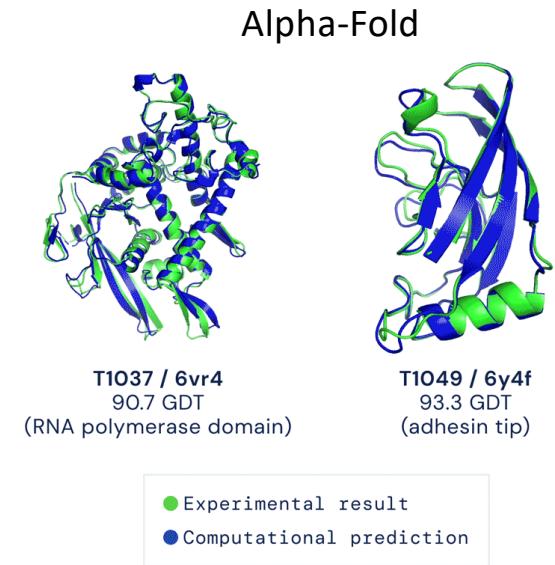
Apprentissage par renforcement (RL)

Ce n'est pas la destination qui compte mais le chemin

RL s'impose si:



- Les méthodes classiques renoncent
- Univers fini, mais combinatoire importante
- Objectif final clair mais pas d'objectif intermédiaire
- Réponse rapide aux essais

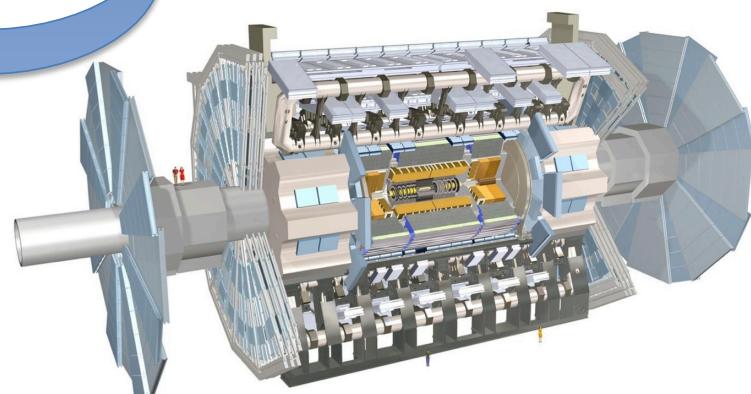


Application en physique des particules

- Conception expérience
- Prise de données
- Echec/succès ?



Voir collaboration [MODE](#)



➔ En pratique, pourrait être utilisable pour le tri rapide d'événements, facilement paramétrable et se prêtant bien à la virtualisation de l'expérience



CHAT GPT Vers une nouvelle façon de travailler?

▶ 3:44

Intellige
Faut-
la révolu
ChatGPT?

Capable de rédiger des textes impressionnantes de réalisme dans tous les domaines, ce robot conversé aux Etats-Unis rencontre un succès phénoménal.



ChatGPT, un stagiaire brillant mais brouillon

La folie ChatGPT Il a tout à tout, sait écrire des lettres

Le monde éducatif cherche la parade contre ChatGPT

Avec ses textes générés automatiquement, le robot américain est un redoutable outil de triche et de plagiat pour les élèves. Les enseignants tentent de riposter.

«C'est un nouveau monde. Adieu aux devoirs à la maison!» La provocation, signée

gnant. Depuis la mise en ligne de ce programme au succès planétaire fin novembre, néo-

mens. Interdiction pure et simple ou autorisation pour un usage raisonné... le moindre

28 | MEDIAS et PUBLICITE

ChatGPT : médias et artistes veulent faire valoir leurs droits d'auteur

Sur du contenu protégé. Alors que des procès guettent, une solution

TU NOUS VOIS TRAVAILLER
AVEC CHAT-GPT
JUSQU'À 64 ANS?!

DEMANDE-LUI



Pessin

Base de connaissance

- En général ChatGPT est plus complet que Bing (qui a accès au web) mais absences de sources souvent rédhibitoires (et base de connaissances gelée à septembre 2021)
- Très bon pour savoir livresque A ou B
- Nettement moins pour traiter A dans B (IA et science par exemple...), surtout si pointu et récent
- → potentiel très important si « fine tuning » pour une base de connaissance interne à l'expérience (articles publics et les discussions internes correspondantes, notes internes, liste emails, transparents de réunion, etc... depuis ~2000)

Prompt engineering



What does Scheimpflug mean ?



The "Scheimpflug principle" is a basic principle in optical engineering and photography. Named after the Austrian army Captain Theodor Scheimpflug, who applied it to aerial photography, it describes how the orientation of the plane of focus (the sharp region in a photo) is related to the orientation of the lens plane and the image plane.

The principle is typically used in photography and videography to control the field of view and the plane of focus, particularly in situations where a wide depth of field is required or when you want to have a specific part of the image in focus.

In more technical terms, the Scheimpflug principle states that when a lens is tilted relative to the image plane (as is possible with certain types of specialized camera lenses, such as tilt-shift lenses), the plane of focus (PoF) is also tilted relative to the image plane. Furthermore, the lens plane, the image plane, and the PoF intersect at a common line. This is known as the "hinge line". By adjusting the tilt of the lens, a photographer can move and rotate the PoF to achieve the desired focus effect.

It's important to note that the Scheimpflug principle isn't applicable to most ordinary photography because it requires specialized lenses or cameras that allow for movements to adjust the lens plane and/or the image plane. However, for specialized applications such as architectural or product photography, the Scheimpflug principle is very useful.



I am a physicist, quite familiar with scientific matter. But there are topics I'm not quite familiar with. I will ask you questions, please answer them in just one paragraph.



Of course, I'm here to assist you. Please go ahead and ask your questions!



what is Scheimpflug ?



The Scheimpflug principle, named after Austrian Captain Theodor Scheimpflug, is a geometric rule in optical engineering describing the orientation relationship between the lens plane, the image plane, and the plane of focus in an optical system such as a camera. According to this principle, when the lens plane is tilted relative to the image plane, the plane of focus also tilts, and all three planes intersect along a common line known as the "hinge line". This principle is often used in photography to control depth of field, particularly with tilt-shift lenses, enabling selective focus or extended depth in images.

Can Normalizing Flows be used at the LHC?

ChatGPT

Yes, normalizing flows can be used at the Large Hadron Collider (LHC) for various tasks, such as data analysis and simulation. Normalizing flows are a type of generative model that can be used to learn the probability distribution of high-dimensional data. In the context of particle physics, normalizing flows can be used to model the probability distribution of particle collision events. One application of normalizing flows in particle physics is in the simulation of particle collisions. Simulating particle collisions is essential for predicting the outcomes of experiments at the LHC. However, simulating particle collisions is a computationally expensive task that can require significant computing resources. Normalizing flows can be used to speed up the simulation of particle collisions by learning the probability distribution of the particles produced in a collision and generating new samples from that distribution. This can reduce the computational resources required to simulate particle collisions, which can be especially useful for rare events.

Normalizing flows can also be used for data analysis at the LHC. For example, they can be used to learn the probability distribution of signal events and background events in a particle physics experiment. This can be useful for identifying signal events, which are events that are of particular interest to researchers, and distinguishing them from background events, which are events that are produced by other processes and can obscure the signal events.

In summary, normalizing flows can be a useful tool for various tasks in particle physics, including simulation and data analysis at the LHC.

Bing

Normalizing Flows are generative models which produce tractable distributions where both sampling and density evaluation can be efficient and exact 1. They are used for distribution learning 2.

I found a research paper titled "Normalizing Flows for LHC Theory" which discusses the use of normalizing flows in LHC theory 3. However, I couldn't find any other information about the use of normalizing flows at LHC.

Is there anything else I can help you with?

1. <https://arxiv.org/abs/1908.09257>
2. <https://arxiv.org/abs/1505.05770>
- 3.

https://www.researchgate.net/publication/368556193_Normalizing_Flows_for_LHC_Theory

◀ best primary reference

- ❑ Bing has tried to make sense from the first few hits
- ❑ ChatGPT more thorough but absence of source a clear handicap
- ❑ ...however one should ask follow-up questions (A then B):

How can normalizing flows be used to learn probability distribution exactly ?
(...)
thanks, please provide as a full latex document

Let's suppose we have a dataset \mathcal{D} of N samples, each of which is a high-dimensional vector $\mathbf{x} \in \mathbb{R}^D$. The goal of a normalizing flow model is to learn a transformation $f_\theta : \mathcal{Z} \rightarrow \mathcal{X}$, parameterized by θ , that maps a simple base distribution, such as a Gaussian distribution, to the target distribution.

Specifically, we want to learn the transformation such that the probability density of the transformed samples $p_X(f_\theta(\mathbf{z}))$ matches the true probability density of the dataset $p_Z(\mathbf{z})$.

The transformation f_θ is constructed by composing a sequence of invertible mappings f_i :

$$f_\theta = f_K \circ f_{K-1} \circ \dots \circ f_2 \circ f_1$$

Each f_i is an invertible function that takes as input a vector \mathbf{u}_i and outputs a vector \mathbf{v}_i :

$$\mathbf{v}_i = f_i(\mathbf{u}_i)$$

The inverse function of f_i is denoted by g_i , such that $g_i(\mathbf{v}_i) = \mathbf{u}_i$. The log-likelihood of a sample \mathbf{x} can be computed using the change of variables formula:

$$\log p_X(\mathbf{x}) = \log p_Z(\mathbf{u}) - \sum_{i=1}^K \log \left| \det \frac{\partial f_i}{\partial \mathbf{u}_i} \right|$$

where $\mathbf{u} = g_K \circ g_{K-1} \circ \dots \circ g_2 \circ g_1(\mathbf{x})$ is the inverse transformation of \mathbf{x} , and $\frac{\partial f_i}{\partial \mathbf{u}_i}$ is the Jacobian matrix of the transformation f_i with respect to \mathbf{u}_i .

During training, the parameters θ are learned by maximizing the log-likelihood of the dataset \mathcal{D} with respect to θ :

$$\theta^* = \arg \max_{\theta} \sum_{\mathbf{x} \in \mathcal{D}} \log p_X(\mathbf{x}; \theta)$$

This optimization can be performed using gradient-based methods, such as stochastic gradient descent.

Lecture scientifique

- médiocre: »Quels sont les points importants de ce papier ? » → tous les points énoncés sont corrects mais:
 - Certains ne viennent pas du papier
 - Notion « d'important » est trop vague
- Par contre, dialogue peut être intéressant si question précise :
« pourquoi la technique A est utilisée » ?
 - → mais toujours 80/20 correct vs BS

Ecriture scientifique

- Très bon pour suggérer une structure/sous-structure d'un article, et donner des idées
- Nul pour texte complet
- Pas mal pour résumer un projet (80/20)
- De bonnes idées pour vulgarisation scientifique : je suis un lycéen, explique moi le boson de higgs



what would be a good analogy for the Higgs boson if I am 18 ?



The Higgs boson is a particle that helps give other particles mass, so here's an analogy that might help you understand it better:

Imagine a school hallway where students are moving around, but some of them are able to move faster than others. The faster students are like particles that have less mass, while the slower students are like particles with more mass. Now imagine that there's a teacher in the hallway, and every time a student passes by the teacher, the student gets slowed down a bit. This teacher is like the Higgs field, which particles interact with to acquire mass. The more a particle interacts with the Higgs field, the more massive it becomes.

And just like how some students might have more trouble getting past the teacher and thus end up moving slower, some particles have more trouble interacting with the Higgs field and thus end up being more massive.

Code

- Très bon pour écrire des bouts de code de complexité modérée
 - ➔ correct à 80% dans conditions réelles
 - Peut aussi expliquer code existant
- Peut aussi expliquer/debuguer du code ➔
- Moins bon pour du code spécifique (3 millions lignes de code C++/Python dans ATLAS) ➔ il faudrait un « fine tuning » dédié



I get an error when I try:

```
$ conda env create -f cpu_environment.yml python=3.9 -n exatrk-cpu  
Solving environment: failed
```

CondaValueError: Malformed version string '~': invalid character(s).



It looks like there might be an error in the `cpu_environment.yml` file that you are trying to use to create your conda environment. The error message you are seeing suggests that there might be an issue with the version number specified in the environment file.

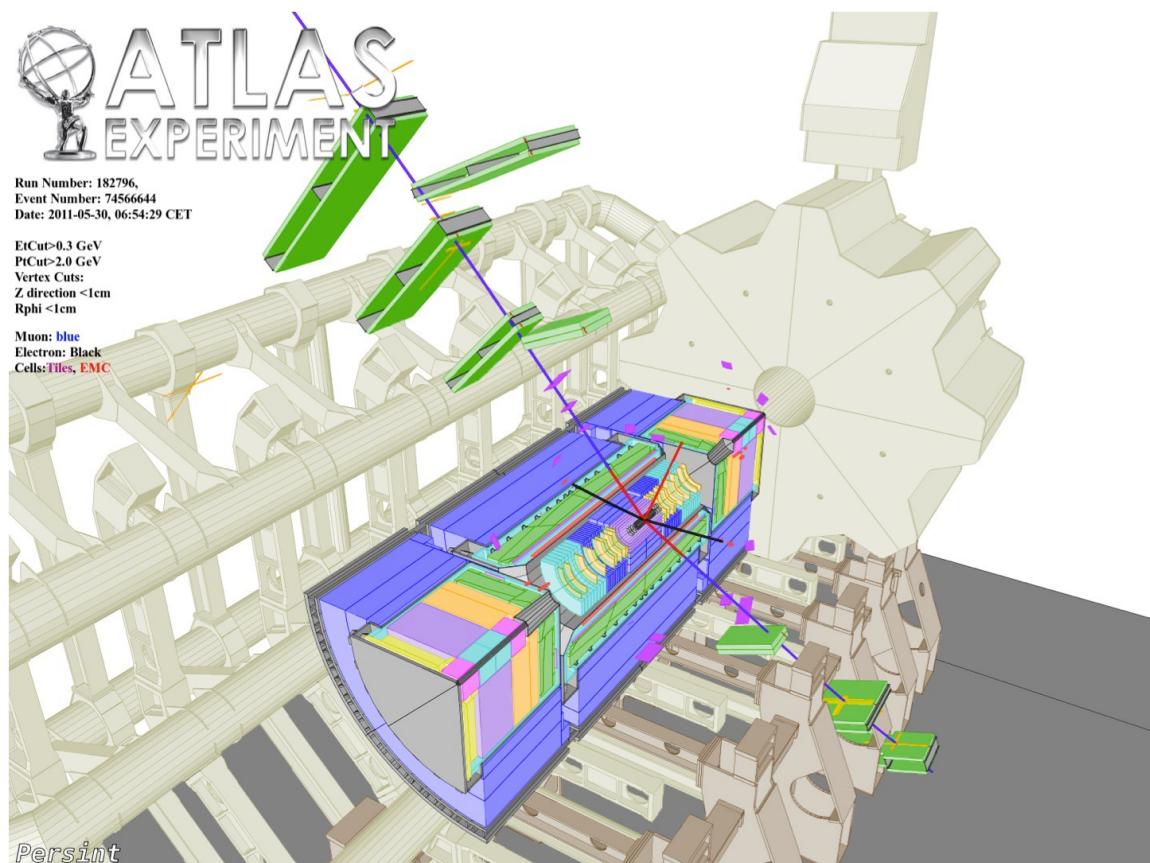
To resolve this issue, you could try removing the `python=3.9` argument from your command and see if that works. Here is an updated version of the command:

bash

Copy code

```
conda env create -f cpu_environment.yml -n exatrk-cpu
```

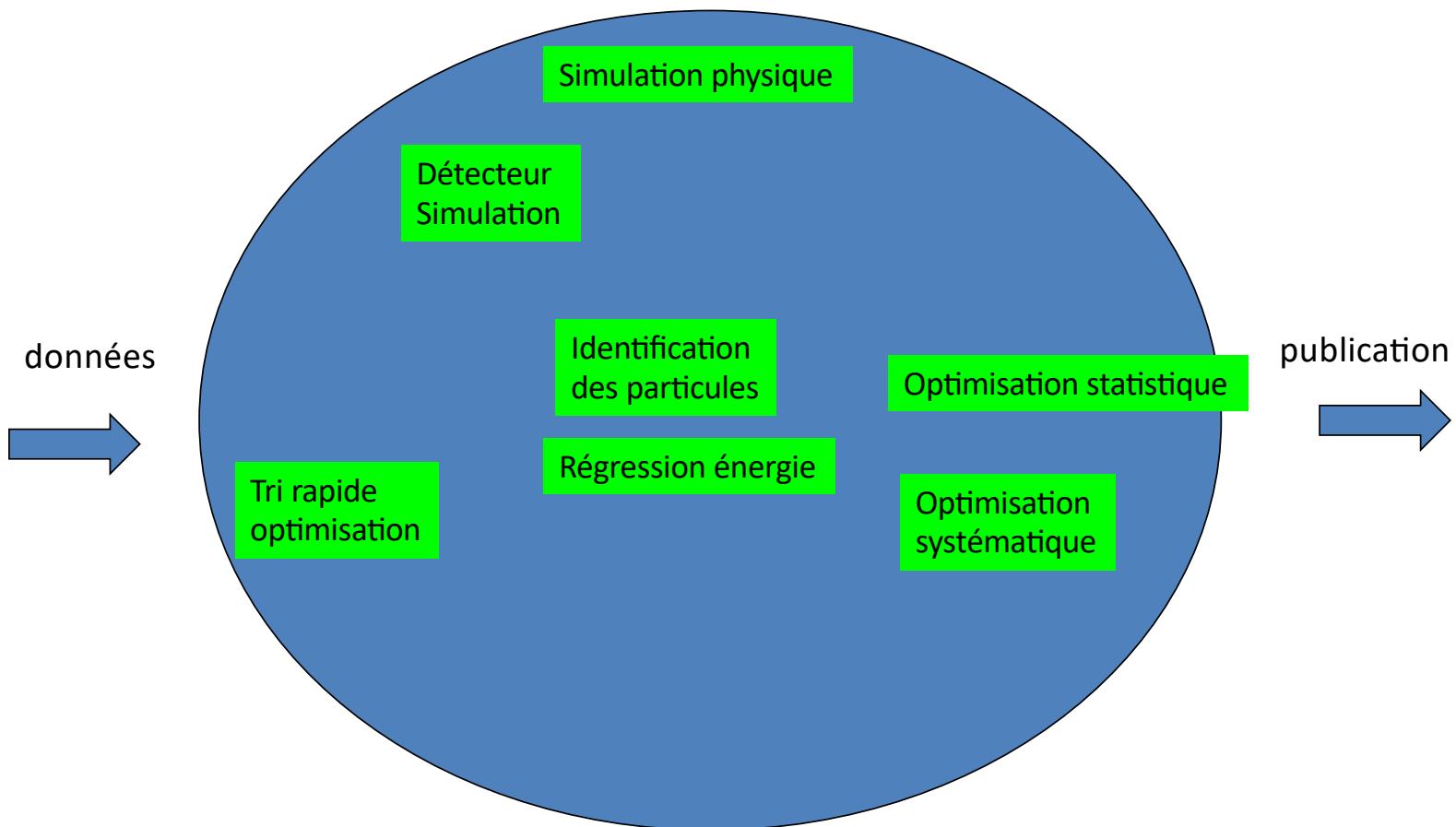
Rêve de physicien.ne



- Peut-on entraîner un grand modèle sur des dizaine de milliards d'événements simulés ?
- Qui « apprendrait la physique » ?

Finalement...

Terrain de jeu des algorithmes IA



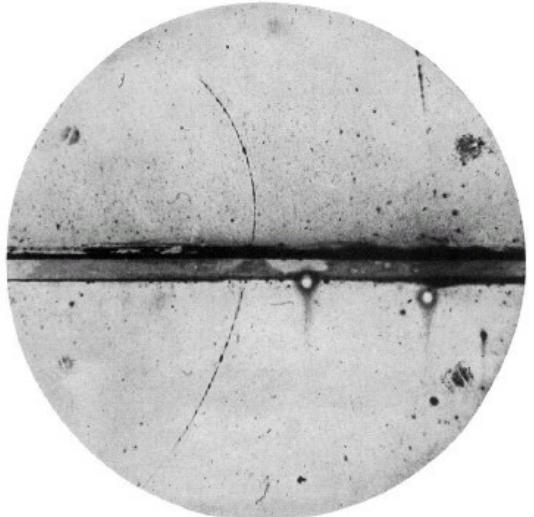
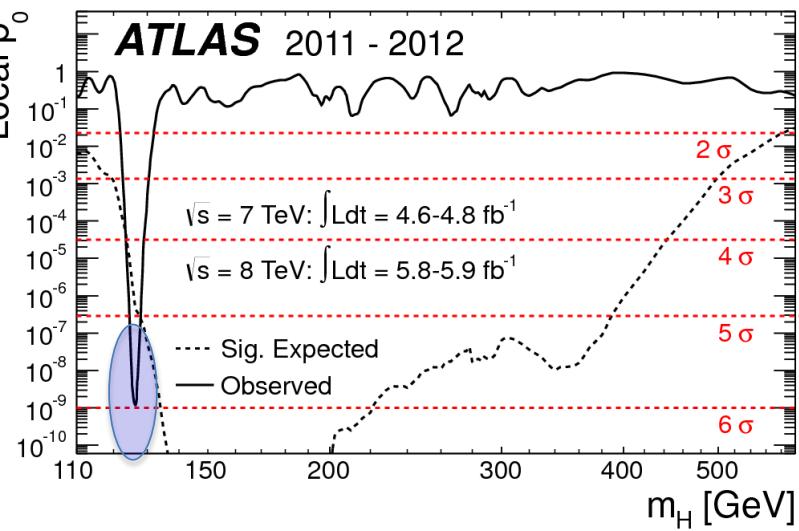
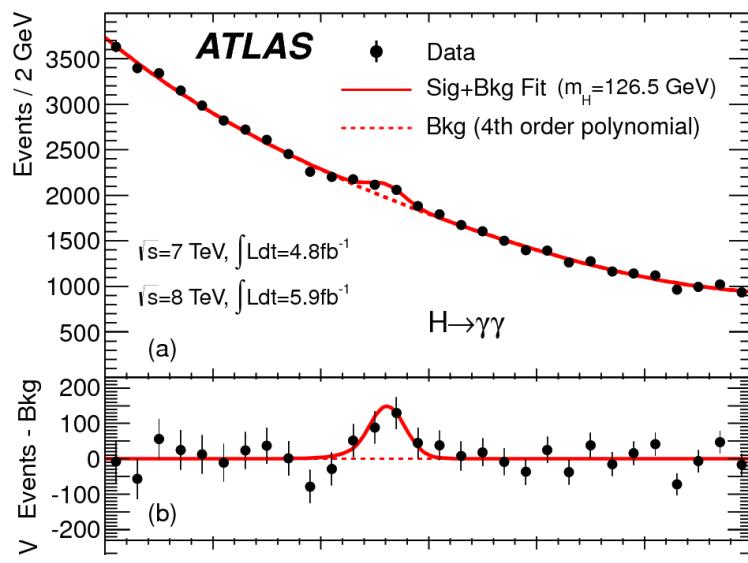
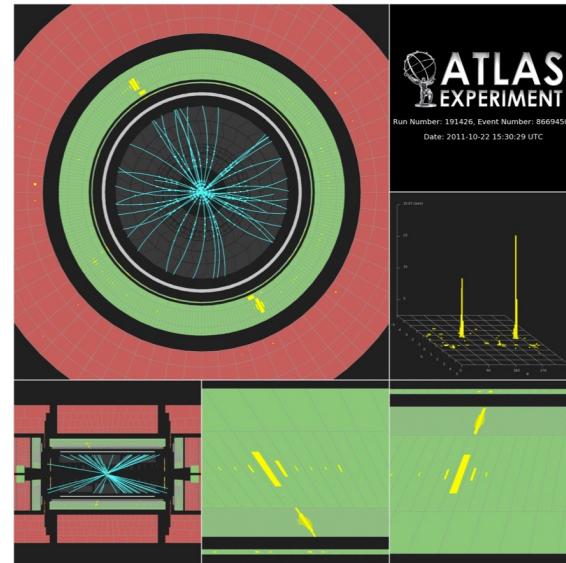


FIG. 1. A 63 million volt positron ($H_P = 2.1 \times 10^4$ gauss-cm) passing through a 6 mm lead plate and emerging as a 23 million volt positron ($H_P = 7.5 \times 10^4$ gauss-cm). The length of this latter path is at least ten times greater than the possible length of a proton path of this curvature.



Conclusion

- Intelligence Artificielle : beaucoup de promesses pour la physique des particules et la recherche scientifique
- Algorithmes très puissants
- Leur terrain de jeu s'agrandit
- Jusqu'où...
- ... ça sera l'enjeu des prochaines années!

