Artificial Intelligence in High Energy Physics I

First, how did we see the Higgs Boson ?

David Rousseau

IJCLab-Orsay

rousseau@lal.in2p3.fr

Rencontres d'été de physique

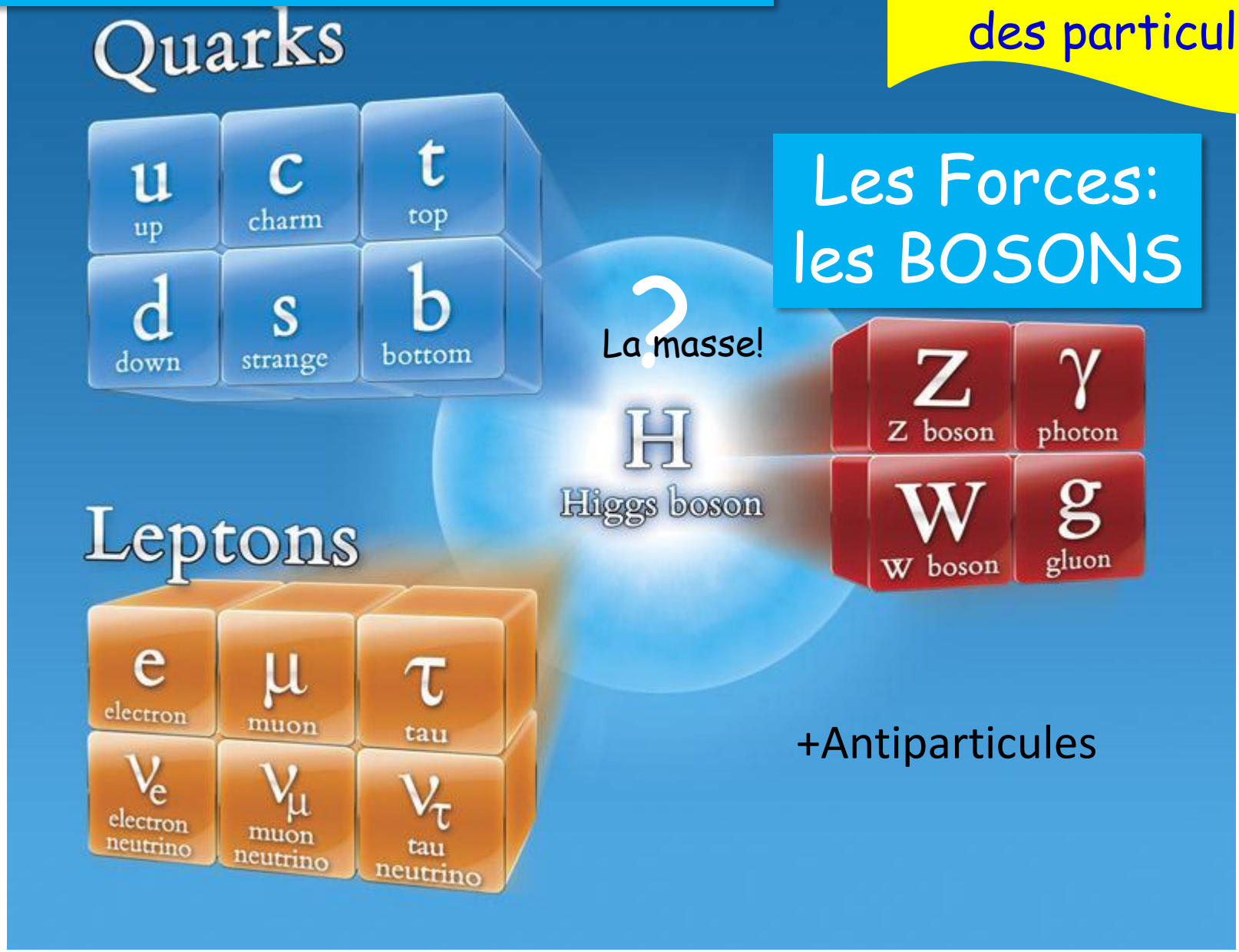
de l'infiniment grand à l'infiniment petit

Juillet 2022

Voir le boson de Higgs ?

La Matière: les FERMIONS

Modèle Standard des particules

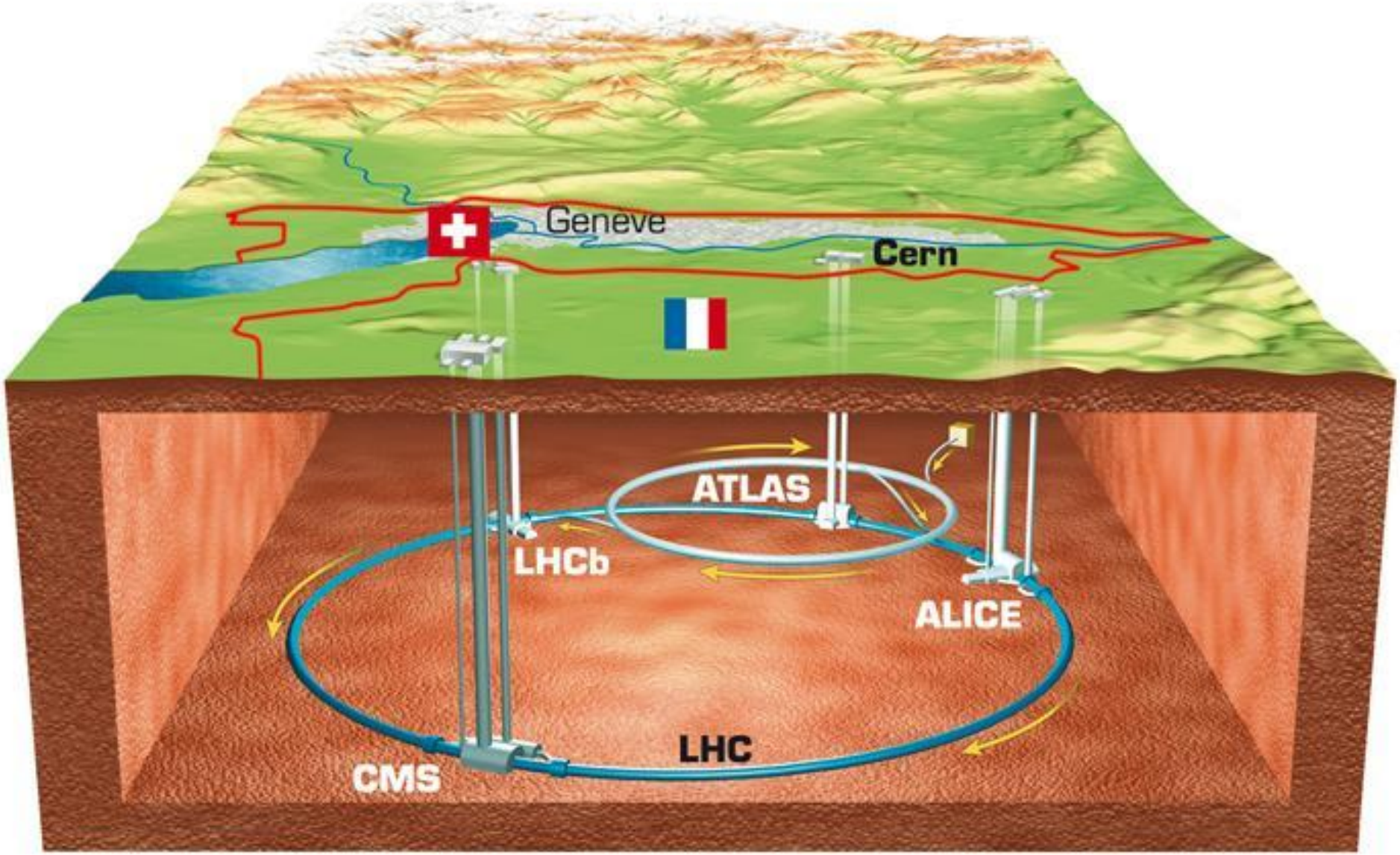


Les Forces: les BOSONS

+Antiparticules

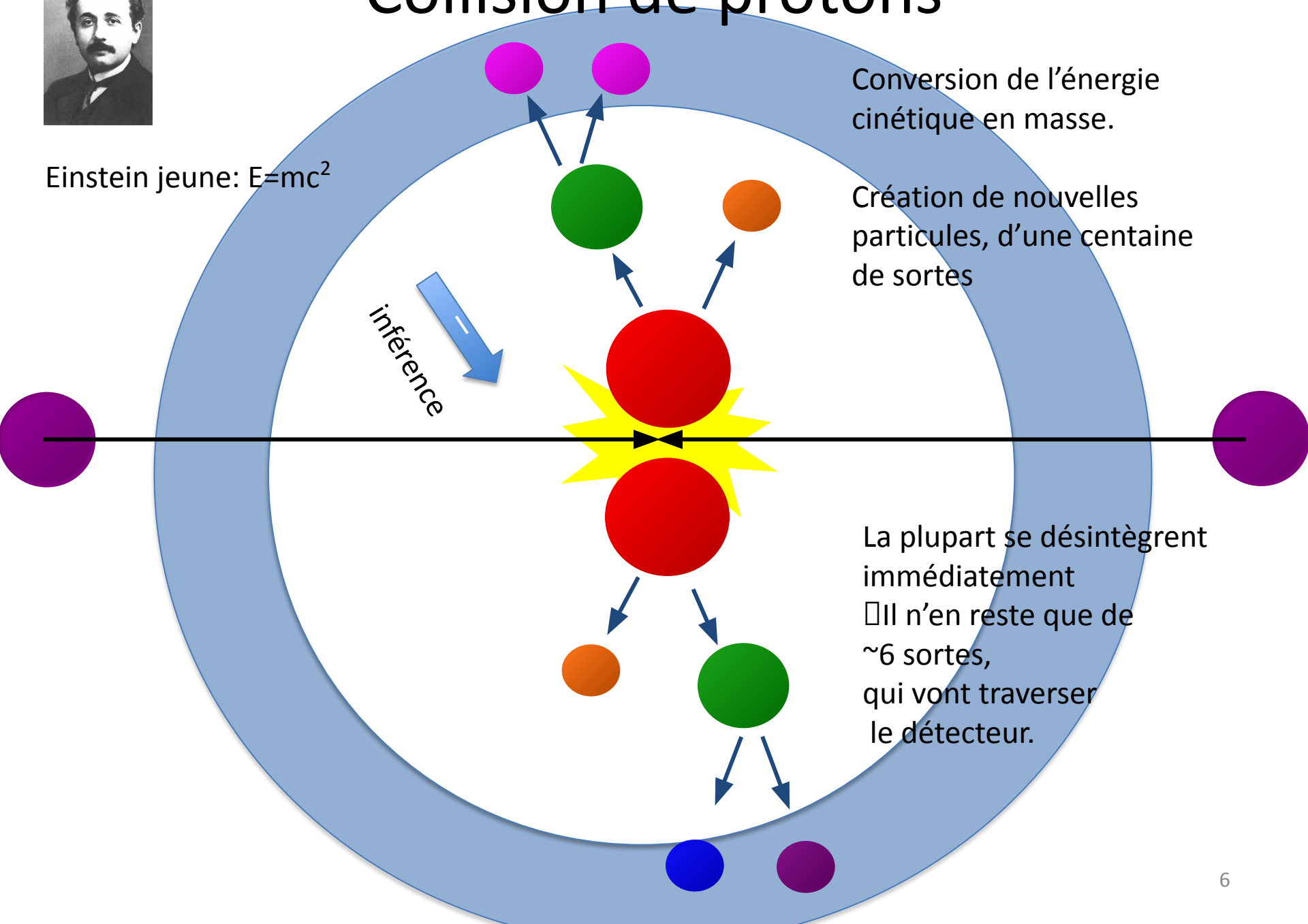
VOIR les particules ?

Le LHC



Collision de protons

Einstein jeune: $E=mc^2$



Conversion de l'énergie cinétique en masse.

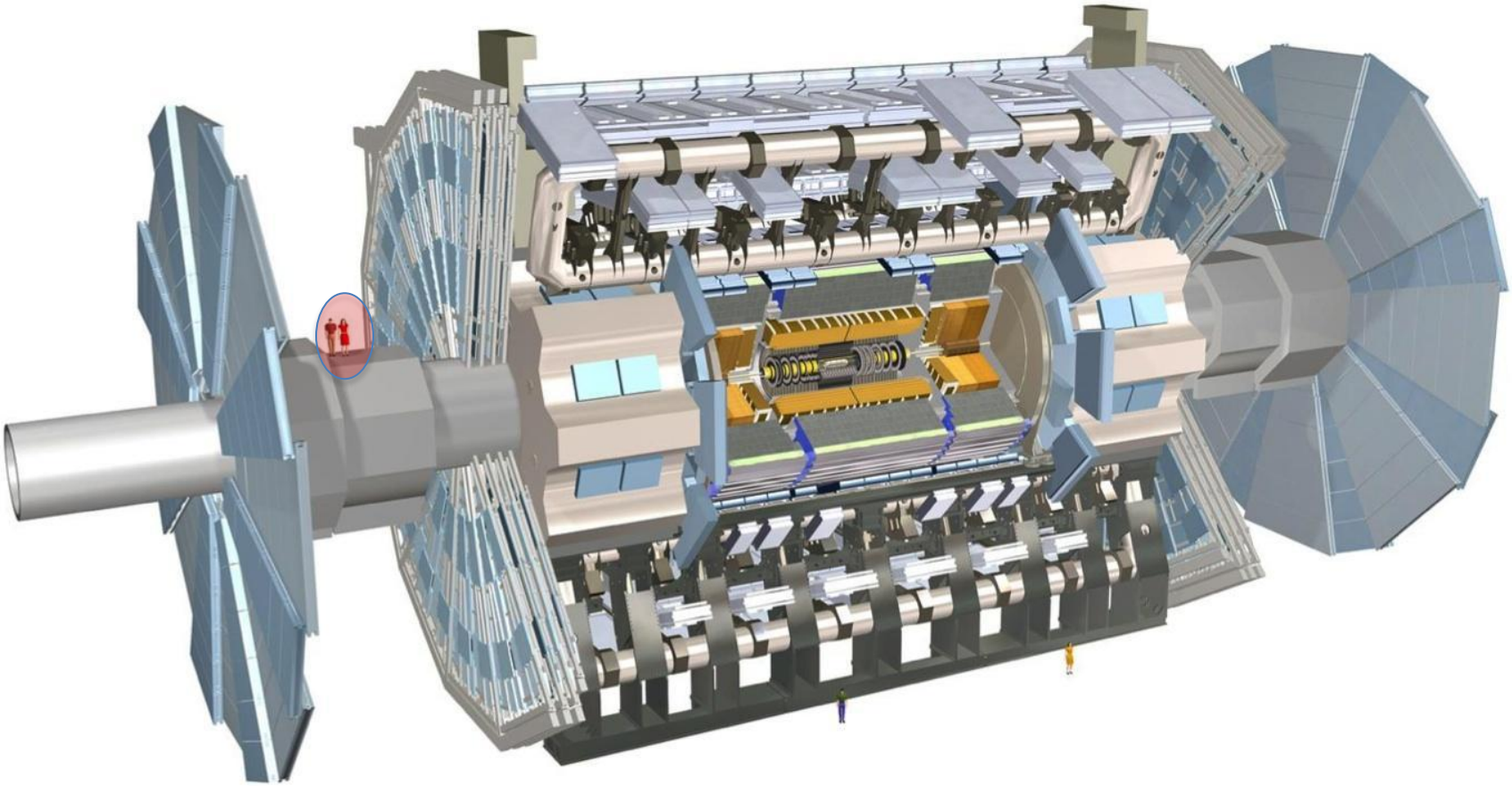
Création de nouvelles particules, d'une centaine de sortes

La plupart se désintègrent immédiatement
Il n'en reste que de ~6 sortes,
qui vont traverser le détecteur.

Le détecteur Atlas

Diamètre: 25m
Longueur: 46m
Poids: 7000 tonnes

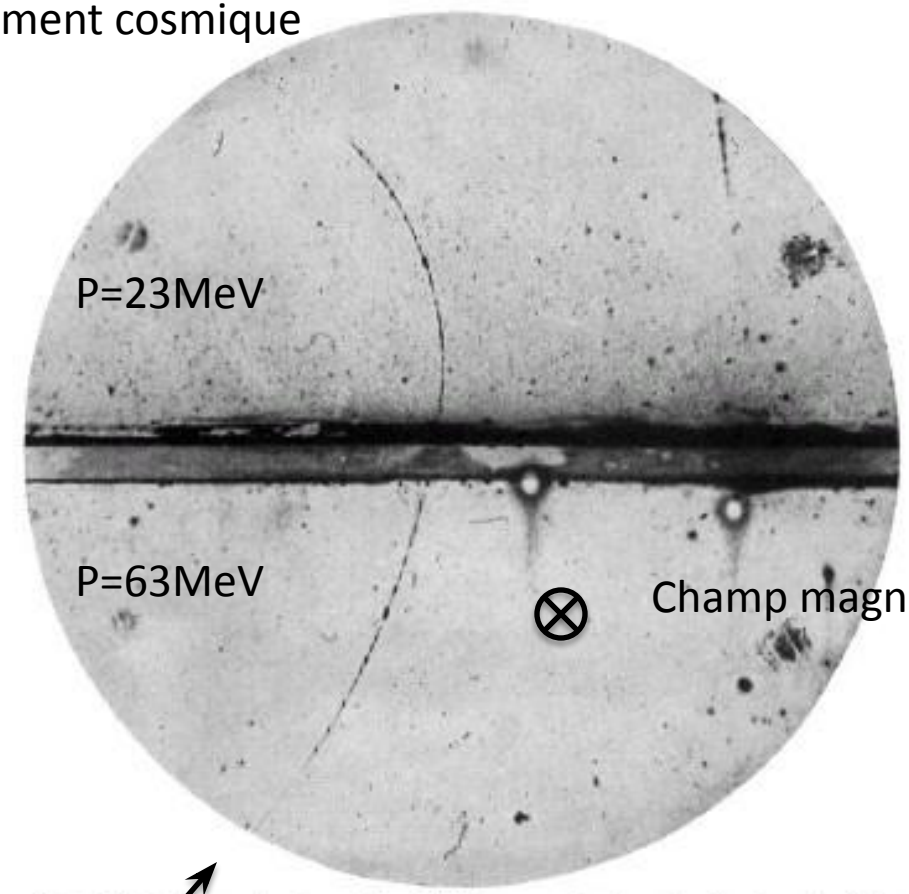
3000 km de câbles
100 millions de canaux



Découverte du positron (Anderson 1932), l'anti-électron postulé par Dirac (mais Anderson l'ignorait)
(en joignant les équations de la mécanique quantique et de la relativité restreinte, Dirac a vu apparaître comme solutions des électrons d'énergie négative)

Nobel 1936

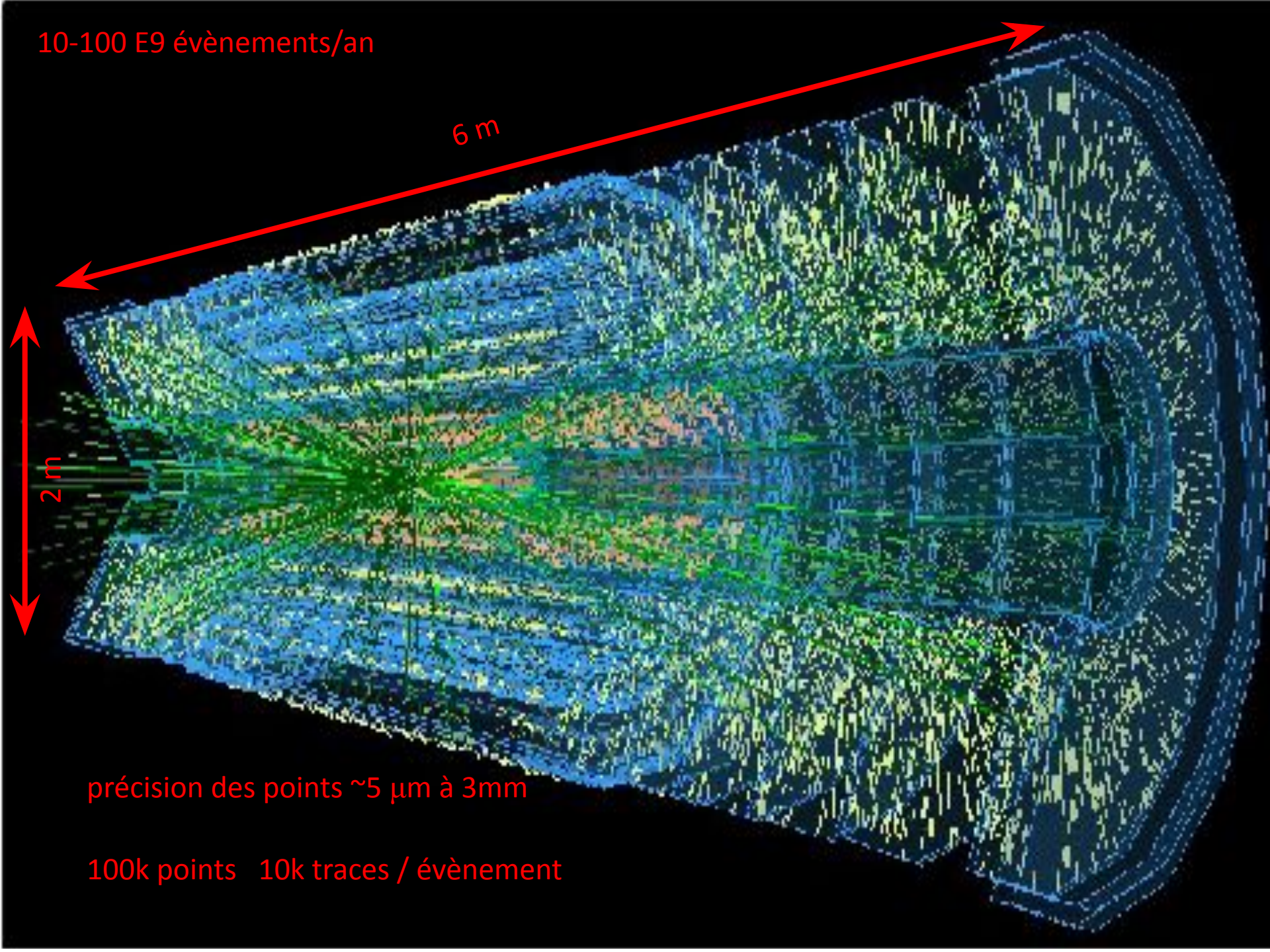
Photo dans une chambre à brouillard (« de Wilson »), soumise au rayonnement cosmique



Plaque de plomb

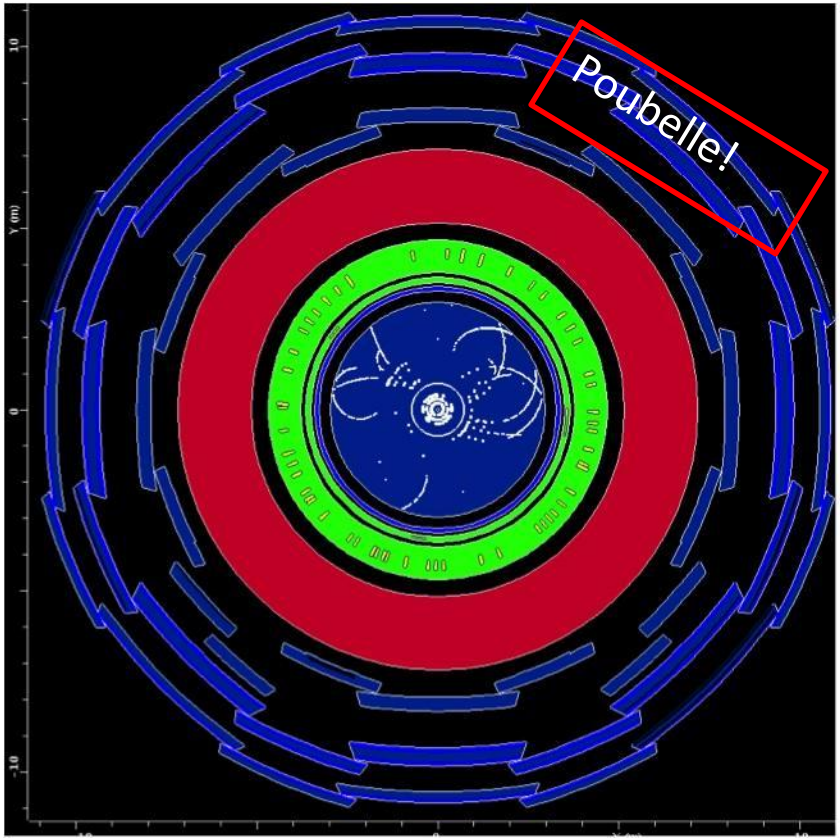
Champ magnétique

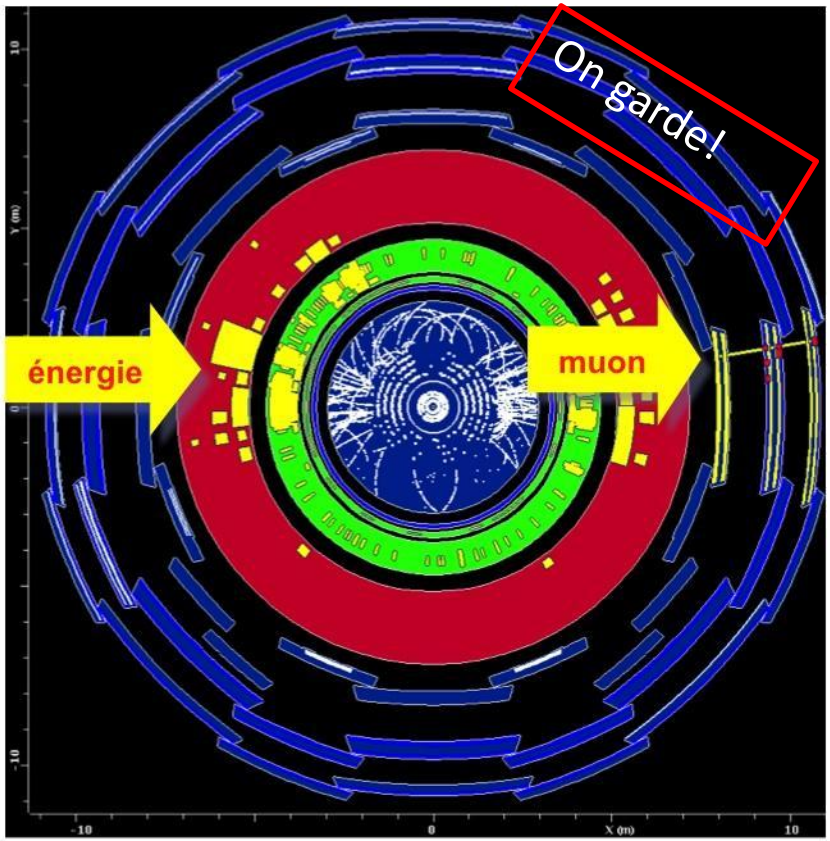
FIG. 1. A 63 million volt positron ($H\rho = 2.1 \times 10^4$ gauss-cm) passing through a 6 mm lead plate and emerging as a 23 million volt positron ($H\rho = 7.5 \times 10^3$ gauss-cm). The length of this latter path is at least two times greater than the possible length of a proton path of this curvature.


10-100 E9 évènements/an
6 m
2 m
précision des points ~5 µm à 3mm
100k points 10k traces / évènement

Sélection des événements en temps réel

- 20 millions de collision de paquets par seconde
- 400 événements sélectionnés (1/50.000) au vol
- échantillonnage en cascade, décision en 1µs-1s sur des *signatures*





VOIR le boson de Higgs

Avant de le voir, on savait tout sur le boson de Higgs, sauf sa masse

Particule très instable (10^{-22}s), se désintègrant immédiatement en paire d'autres particules, de façon imprévisible (sauf en moyenne)

Probabilités de désintégration prédites pour une masse de 125 GeV

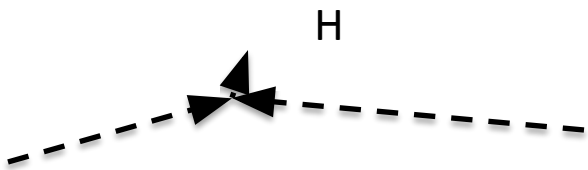
$H \rightarrow bb$	58%
$H \rightarrow WW^*$	21%
$H \rightarrow \tau^+\tau^-$	6.4%
$H \rightarrow ZZ^*$	2.7%
$H \rightarrow \gamma\gamma$	0.2%

$E=mc^2$

Einstein en 1905

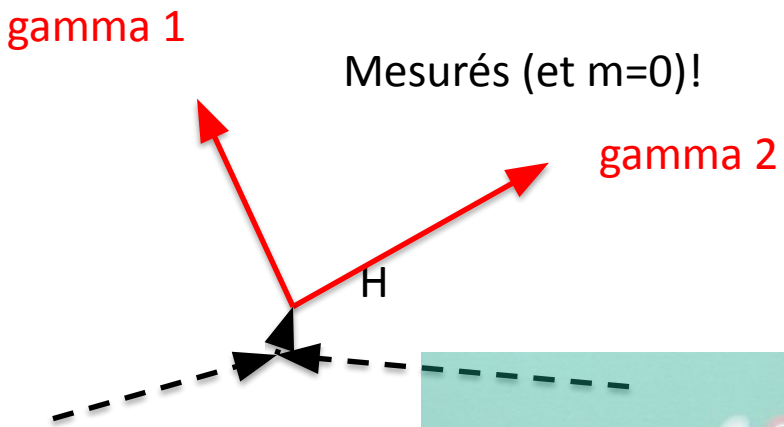
En fait, la formule complète est $E^2=p^2c^2+m^2c^4$
p est l'impulsion, mv en mécanique classique
En choisissant bien les unités, on se débarrasse de c:

$$E^2=p^2+m^2$$



H, juste avant sa désintégration

$m_H^2=E_H^2-p_H^2$



Juste après sa désintégration

Conservation énergie et impulsion

$E_H=E_{g1}+E_{g2}$

$\vec{p}_H=\vec{p}_{g1}+\vec{p}_{g2}$

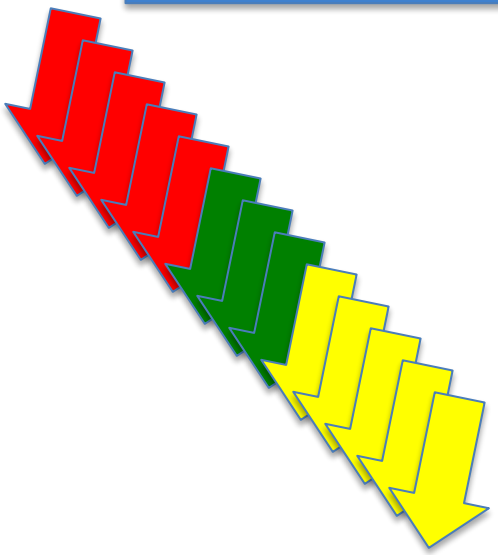
on en déduit m_H!

Finalement…

10^{14} collisions

Tri rapide et grossier

10^{9} événements sur disque

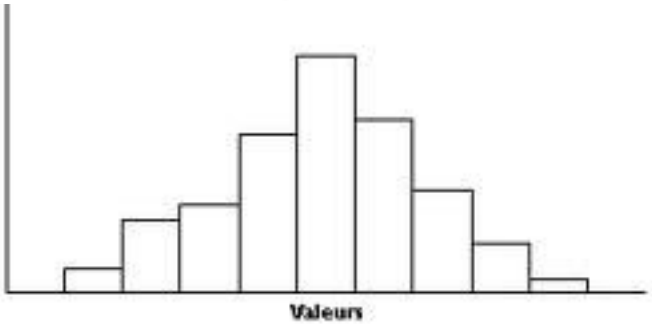


Tri précis

10^{5} événements à 2 gamma

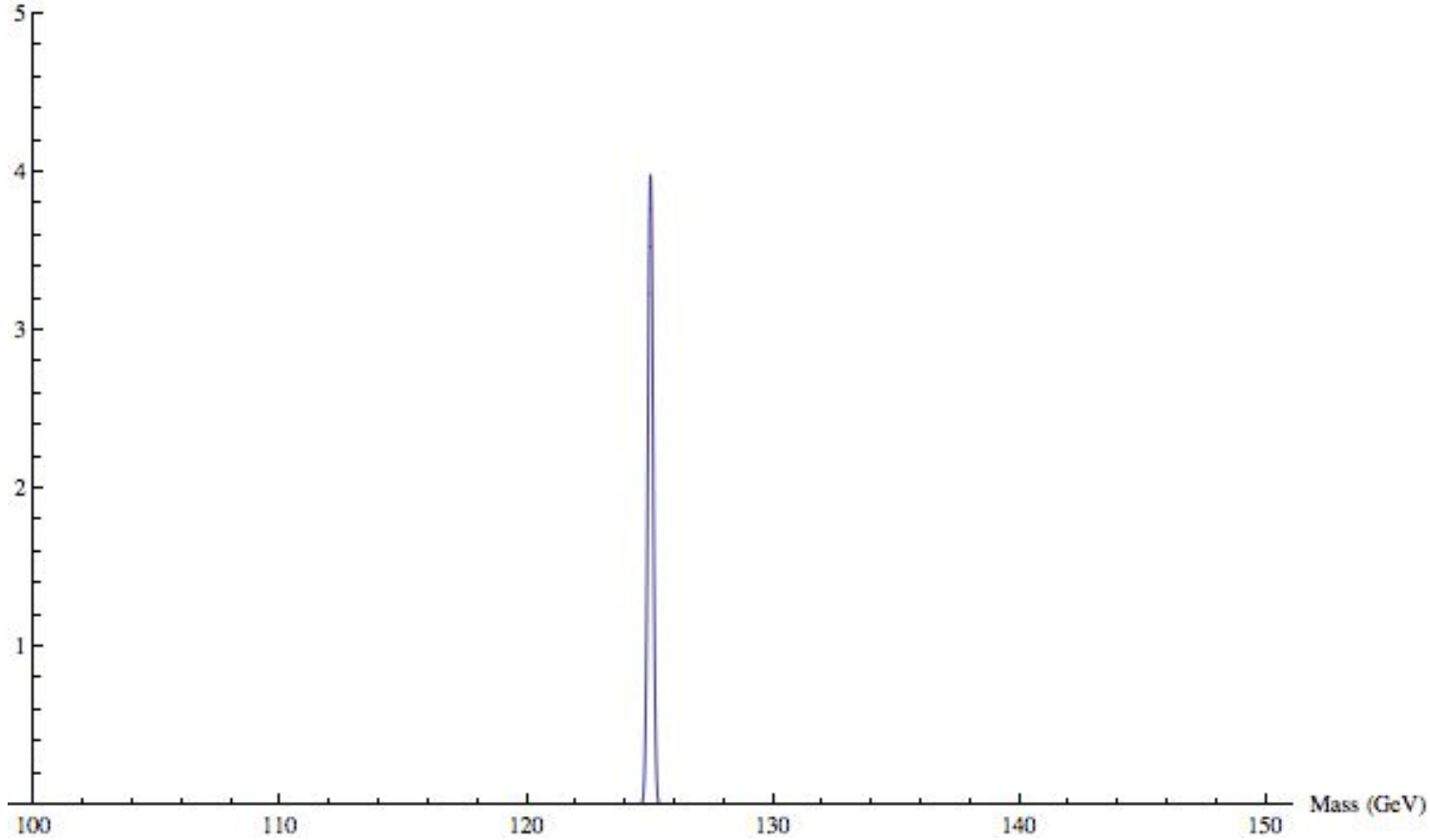
Calcul de la masse

histogramme



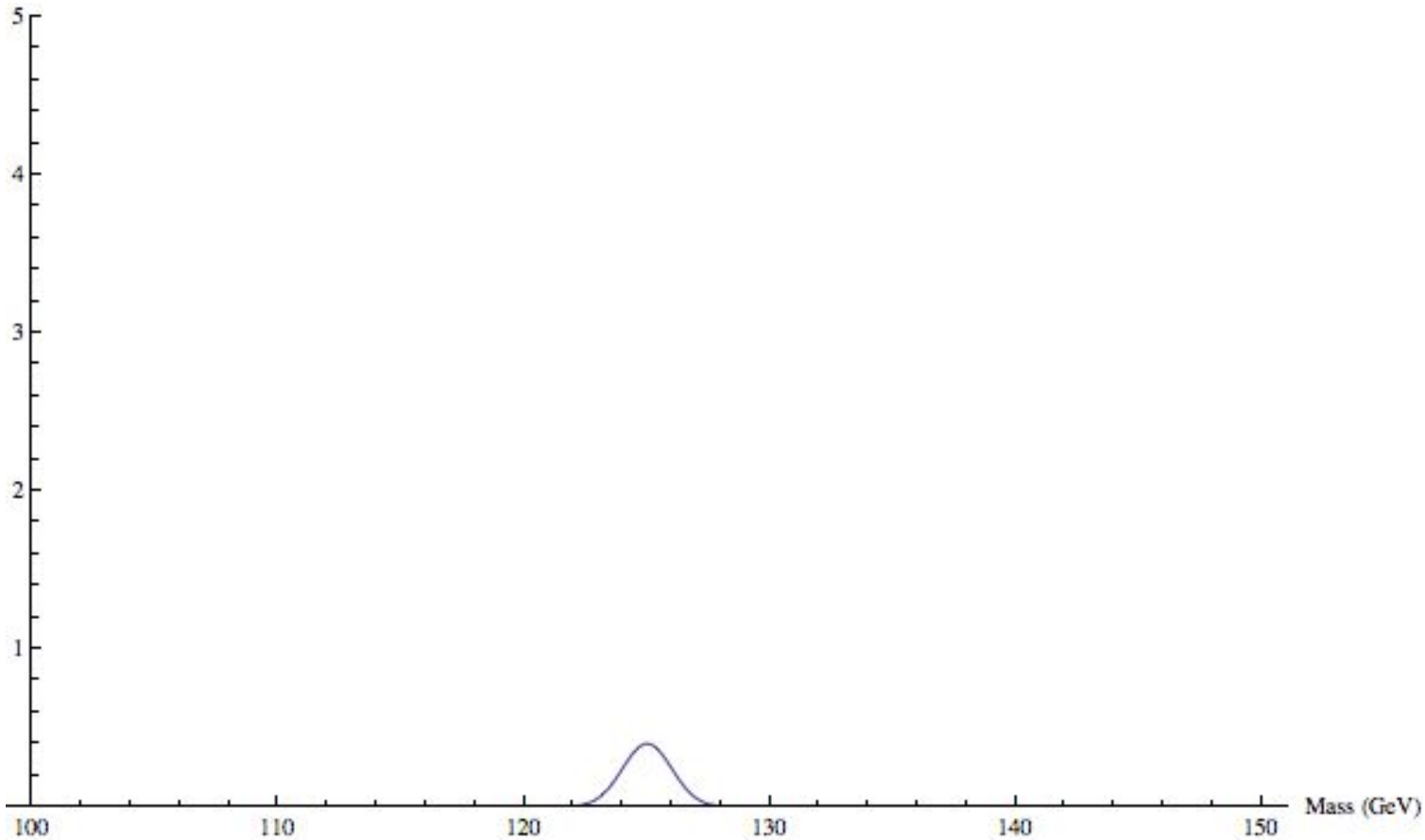
Effet de la précision du détecteur

Cliquer pour animer



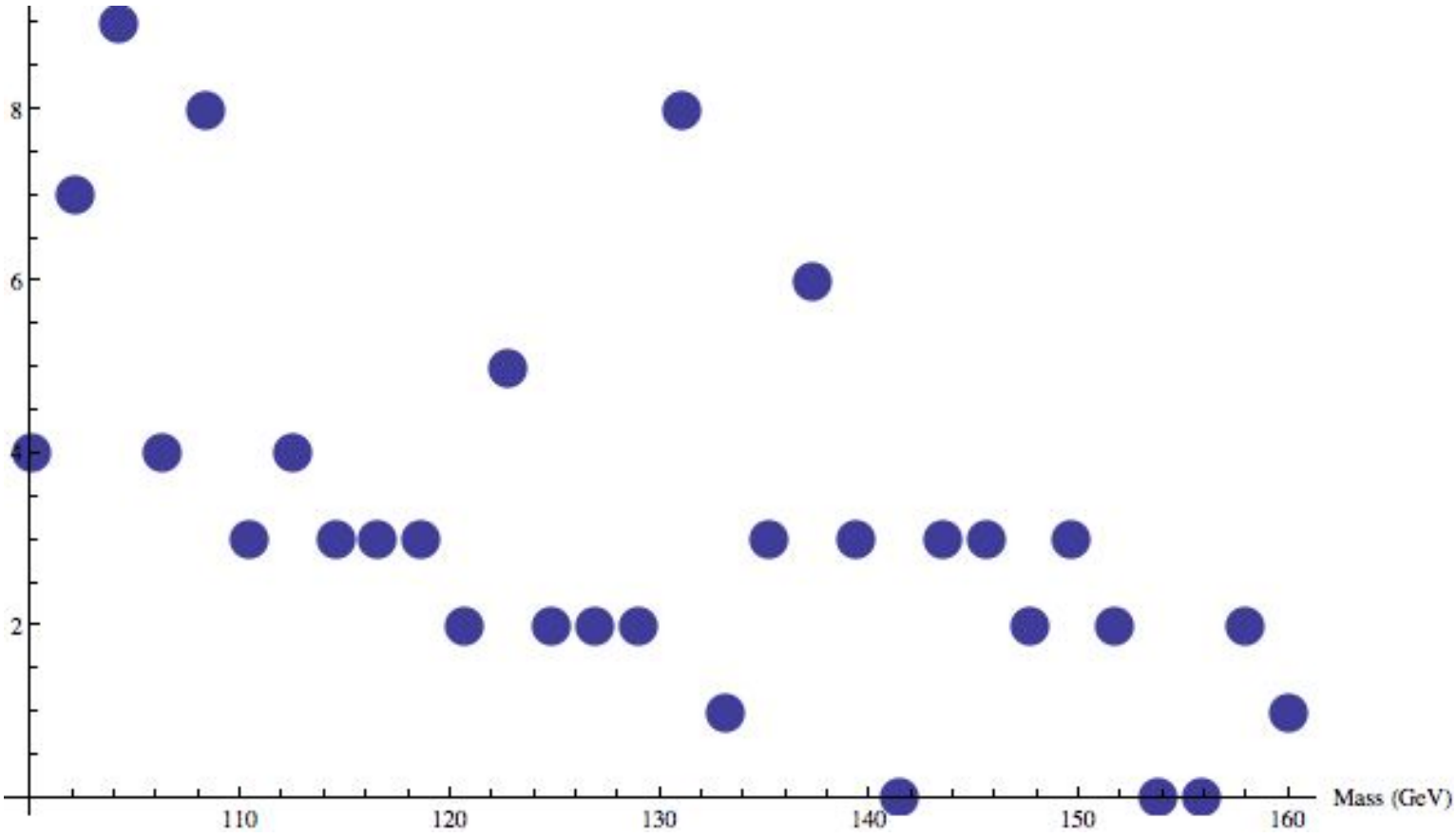
Effet du bruit de fond

Cliquer pour animer

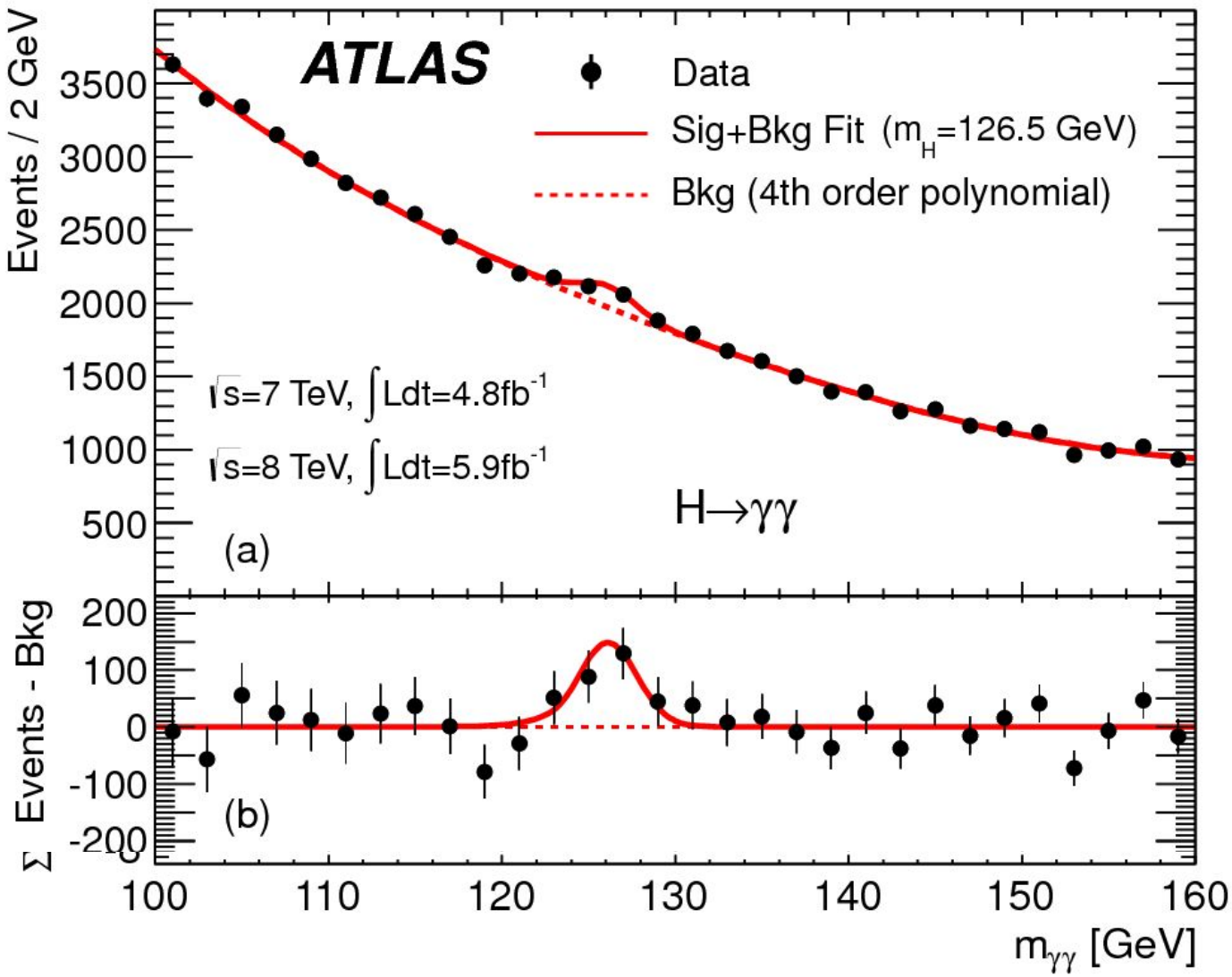


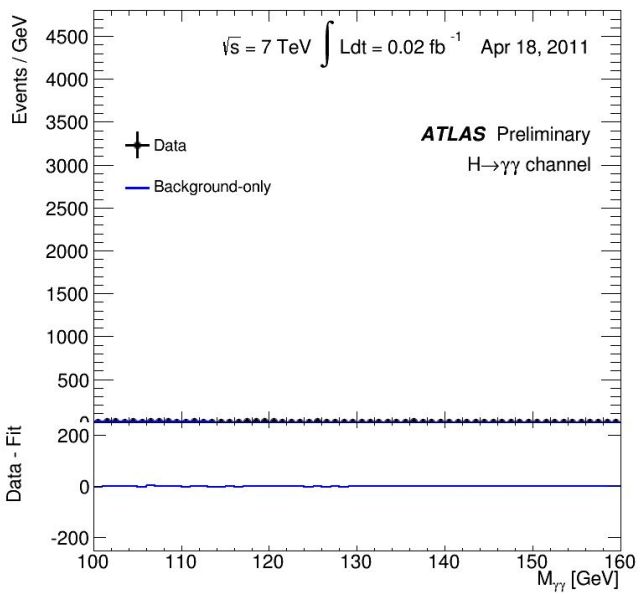
Effet de la statistique

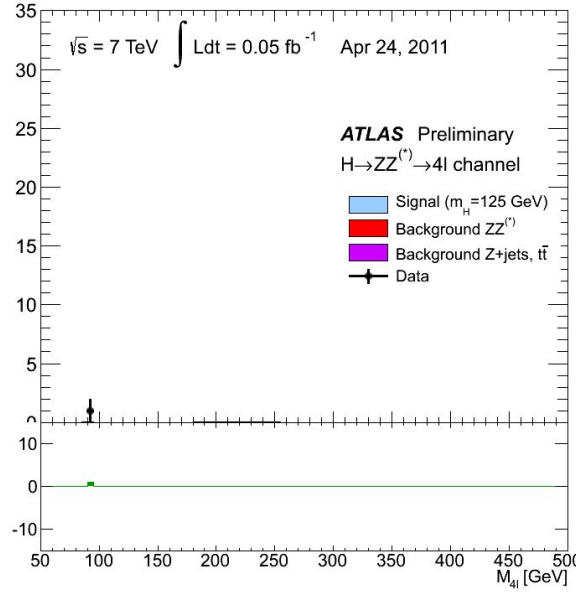
Cliquer pour animer

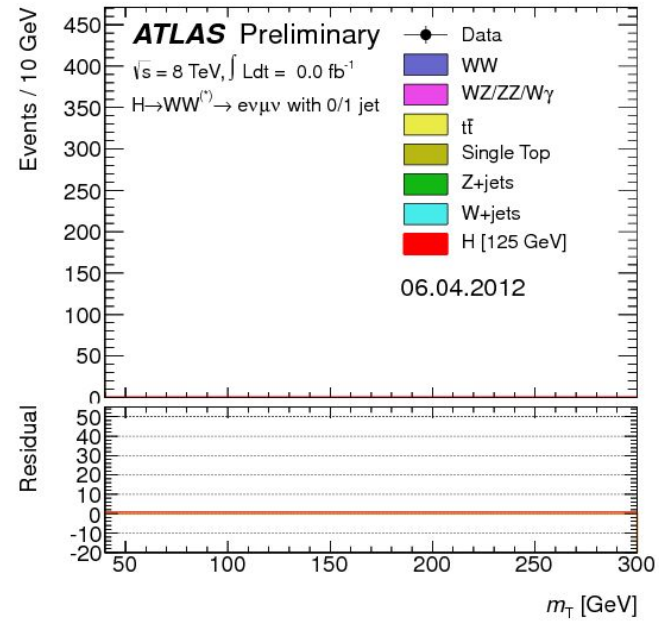


Et maintenant « en vrai » (Juillet 2012)




√s = 7 TeV ∫ Ldt = 0.02 fb⁻¹ Apr 18, 2011
ATLAS Preliminary
H→γγ channel
Data
Background-only
Events / GeV
Data - Fit
M_γγ [GeV]


√s = 7 TeV ∫ Ldt = 0.05 fb⁻¹ Apr 24, 2011
ATLAS Preliminary
H→ZZ(*)→4l channel
Signal (m_H=125 GeV)
Background ZZ(*)
Background Z+jets, tt̄
Data
M_4l [GeV]


ATLAS Preliminary
√s = 8 TeV, ∫ Ldt = 0.0 fb⁻¹
H→WW(*)→ eνμν with 0/1 jet
Data
WW
WZ/ZZ/Wγ
tt̄
Single Top
Z+jets
W+jets
H [125 GeV]
06.04.2012
Events / 10 GeV
Residual
m_T [GeV]

Séminaire du 4 juillet 2012 au CERN

rapt crowd watches as physicists Fabiola Gianotti (standing, left), Rolf Dieter-Heuer (right) and Joe Incandela (far right) unveil evidence for the Higgs boson.

theguardian
Diamond set to come out fighting as he faces MPs
Pakistan ends supply route dispute

Libération
Physique des particules
La masse est dite
Le Cern a réussi à mettre en évidence le boson de Higgs qui résout une énigme fondamentale et ouvre une nouvelle étape scientifique.
Les derniers feux des pharaons
Suicides chez France Télécom: l'ancien patron mis en examen
A nos lecteurs

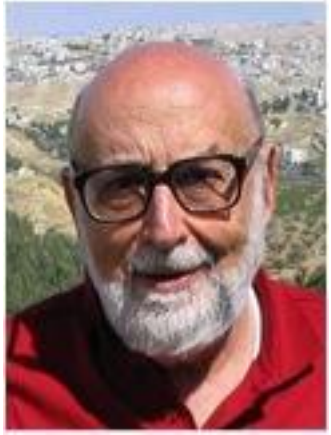
Le Monde
Science : la matière dévoilée

EL PAÍS
Hallada "la más sólida evidencia" de la existencia del bosón de Higgs

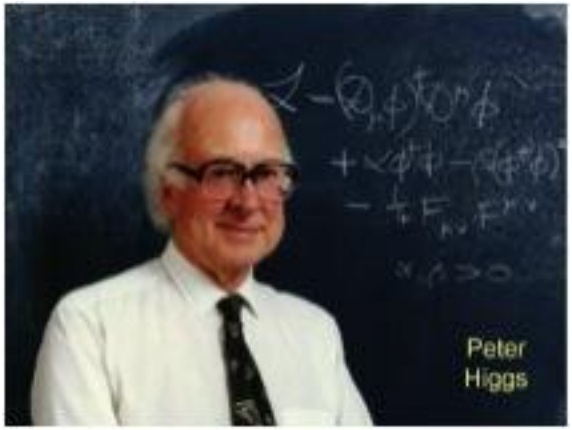
OST POPULAR
U.S. Edition
The New York Times
Wednesday, July 4, 2012
Last Update: 4:00 AM ET
MES DIGITAL SUBSCRIPTION: 4 WEEKS FOR 99
Search
OPINION »
EDITORIAL
Too Quiet, A
Health Care
MARKETS »
Britain

LE FIGARO
Le Cern a mis la main sur le boson de Higgs

Robert Brout 1928-2011



François Englert 1932-



Peter Higgs 1929-

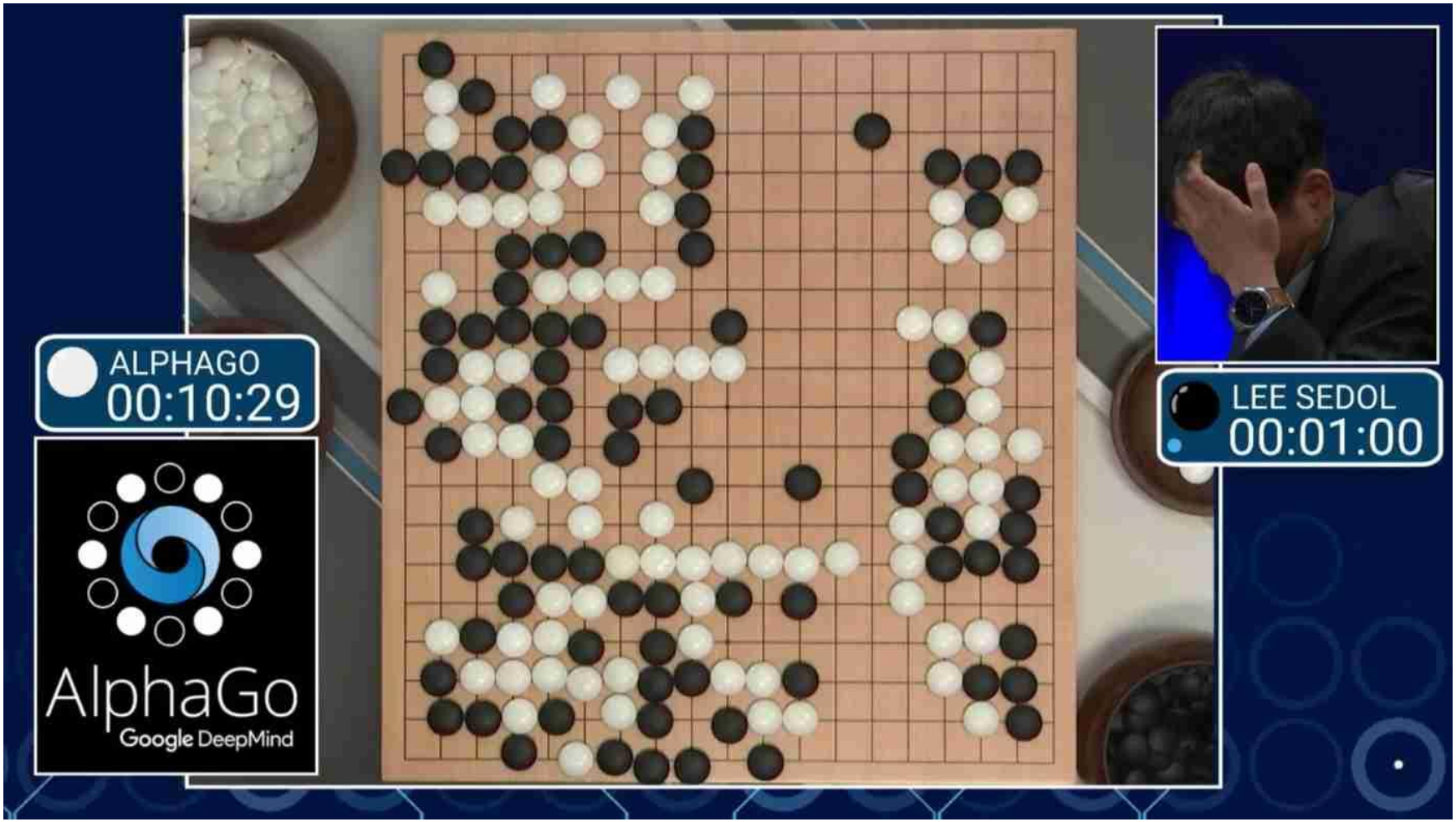
Également : G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble,

Champ de Higgs

« pour la découverte théorique d'un mécanisme qui contribue à notre compréhension de l'origine de la masse des particules subatomiques, qui a récemment été confirmé par la découverte de la particule fondamentale prédite, par les expériences ATLAS et CMS au grand collisionneur de hadrons (LHC) du CERN »

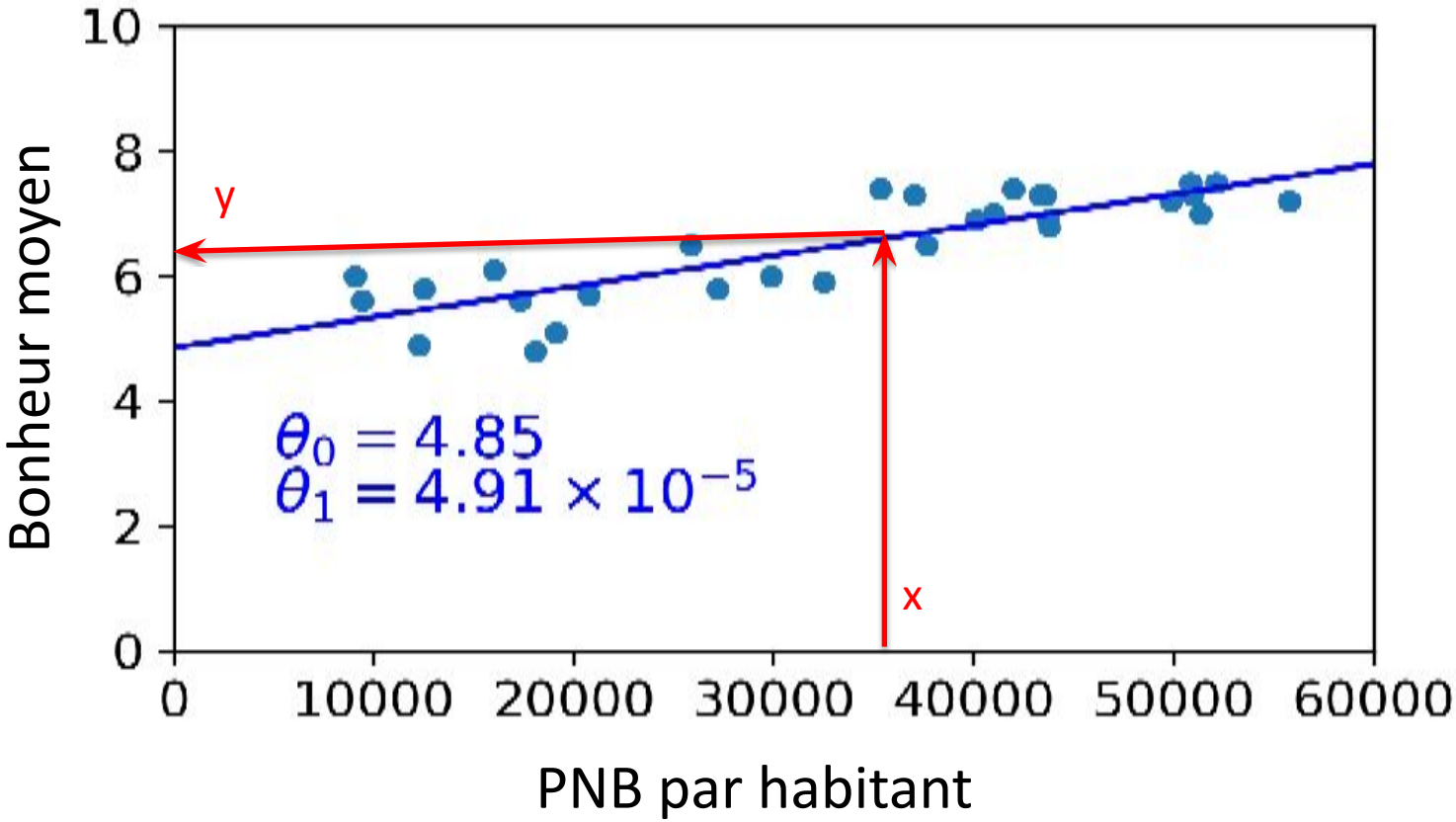
Boson de Higgs

Intelligence Artificielle


ALPHAGO
00:10:29
AlphaGo
Google DeepMind
LEE SEDOL
00:01:00

Régression Linéaire

Boskovic, Legendre, Laplace, Gauss



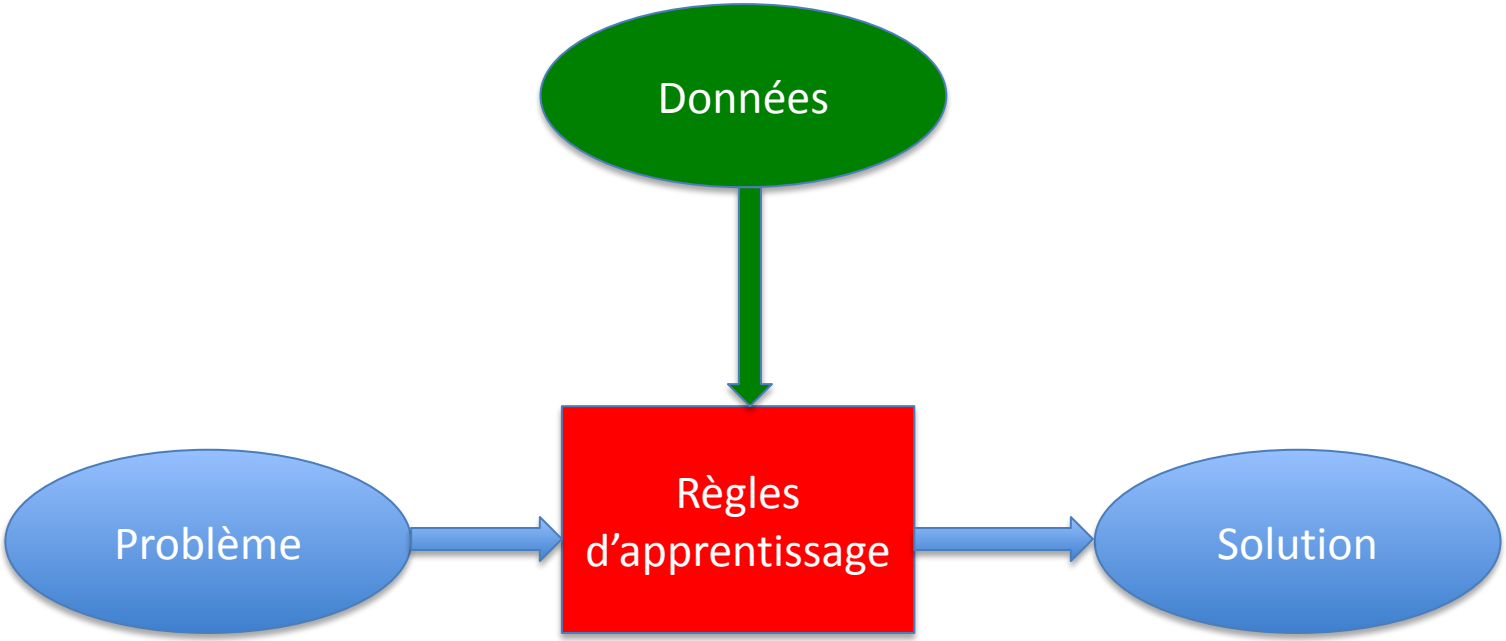
y=f(x)

Etant donné x, on veut y comment construire f ?

x	f	y
Ecriture manuscrite		texte
Image		chien ou chat ?
Photo		maman ou mamie ?
« Comment ça va ? »		كيف حالك ؟
Parole		texte
Compte facebook		publicité ciblée
Position des pièces		prochain coup
Camera + capteurs+GPS		action sur volant

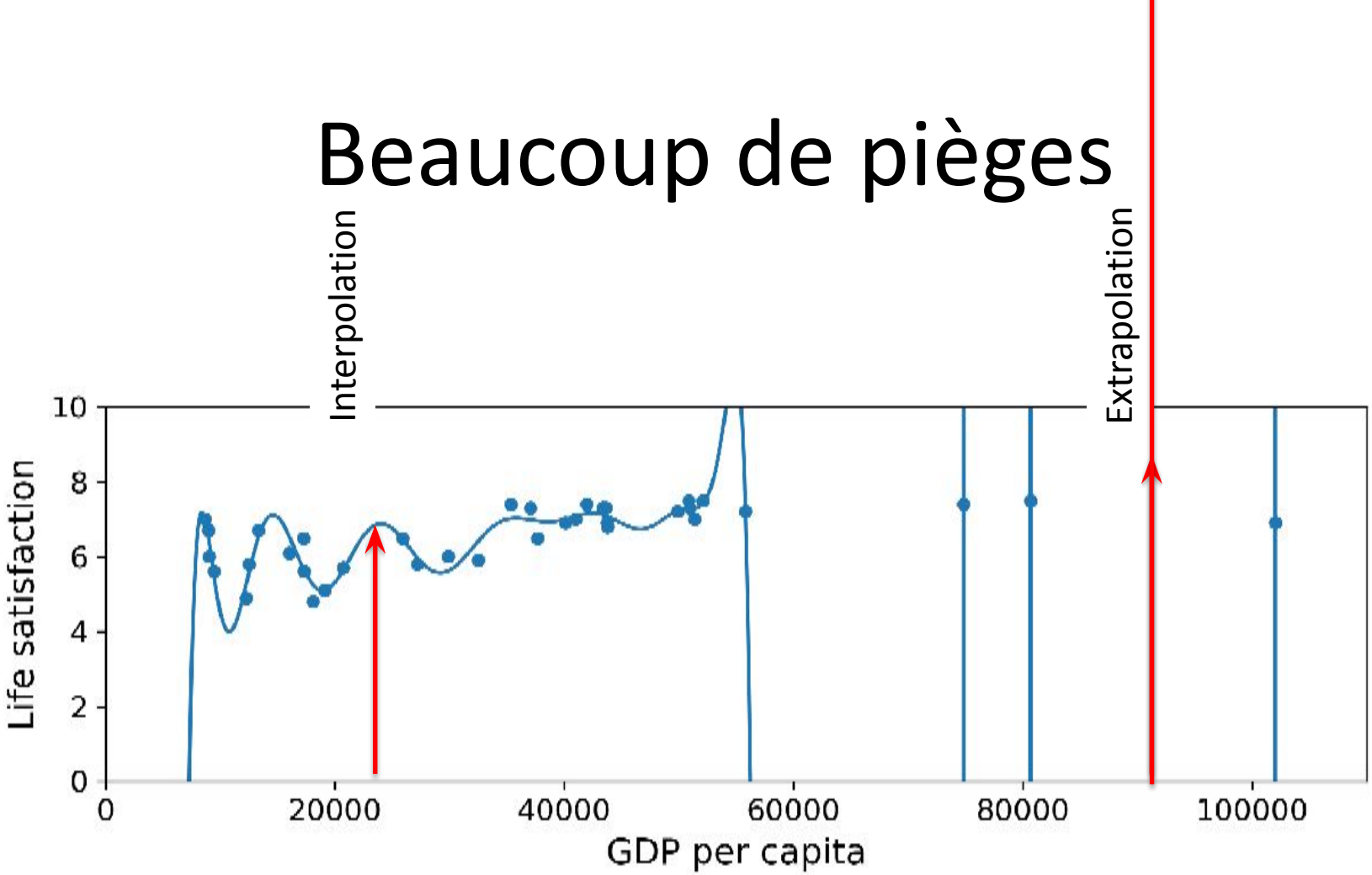
Informatique traditionnelle

Apprentissage Automatique

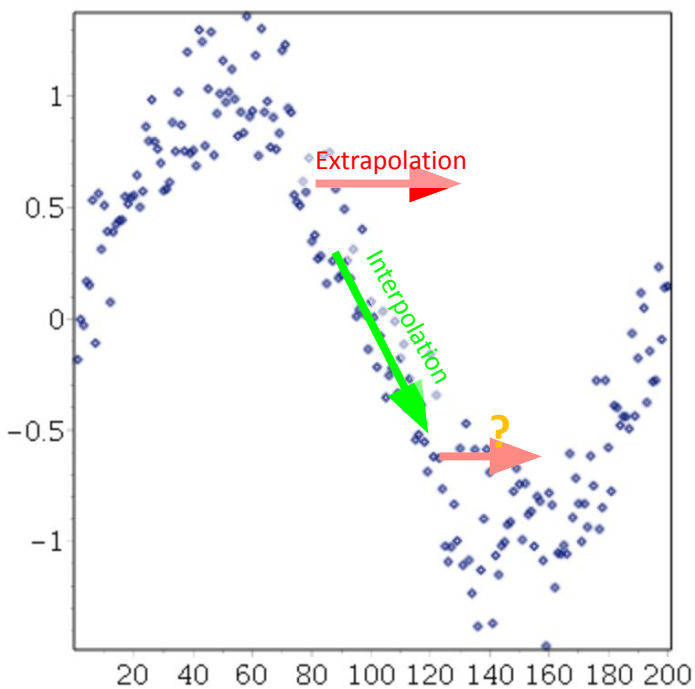


Apprendre = optimiser les paramètres internes de l'algorithme : n=2 - millions

Beaucoup de pièges



Interpolation vs Extrapolation

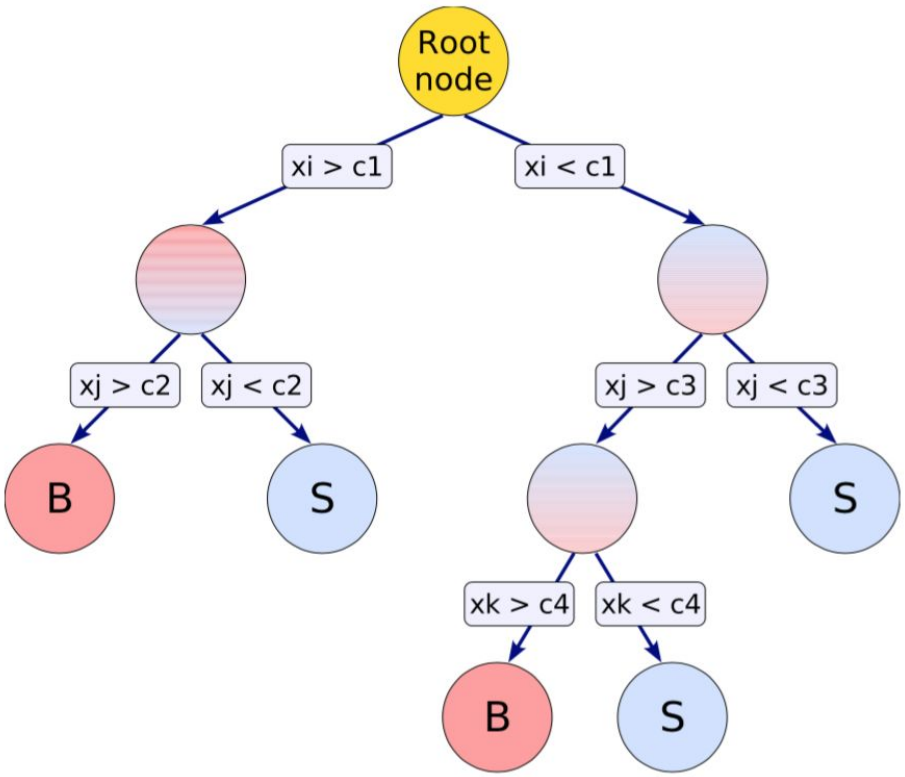


Interpolation/Extrapolation already ill-defined in 2D, what about large dimensions ?

Comment ça marche ?

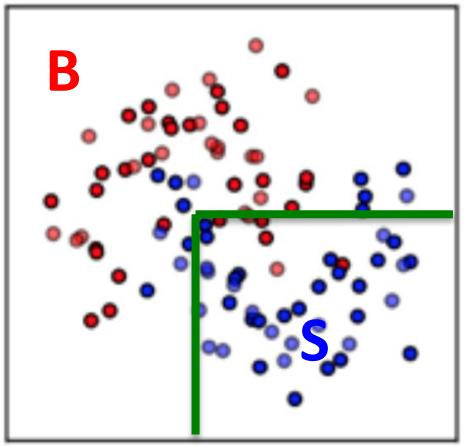
Arbre de Décision Boosté

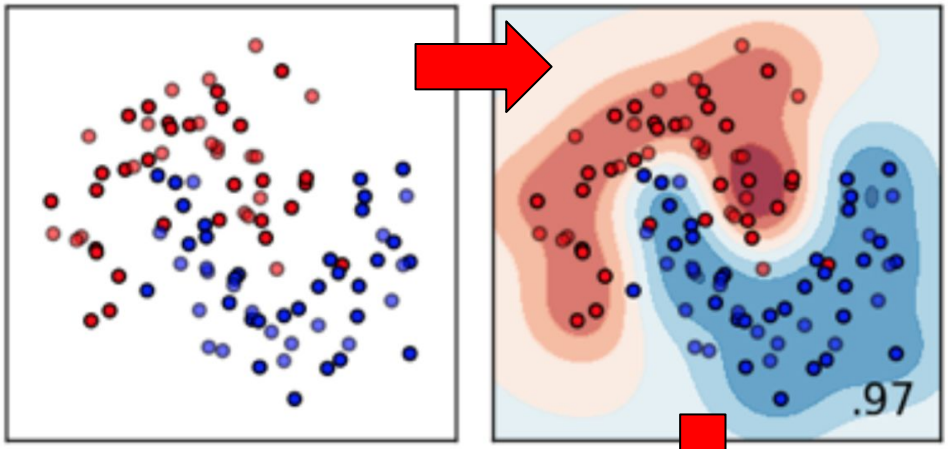
(Boosted Decision Tree)

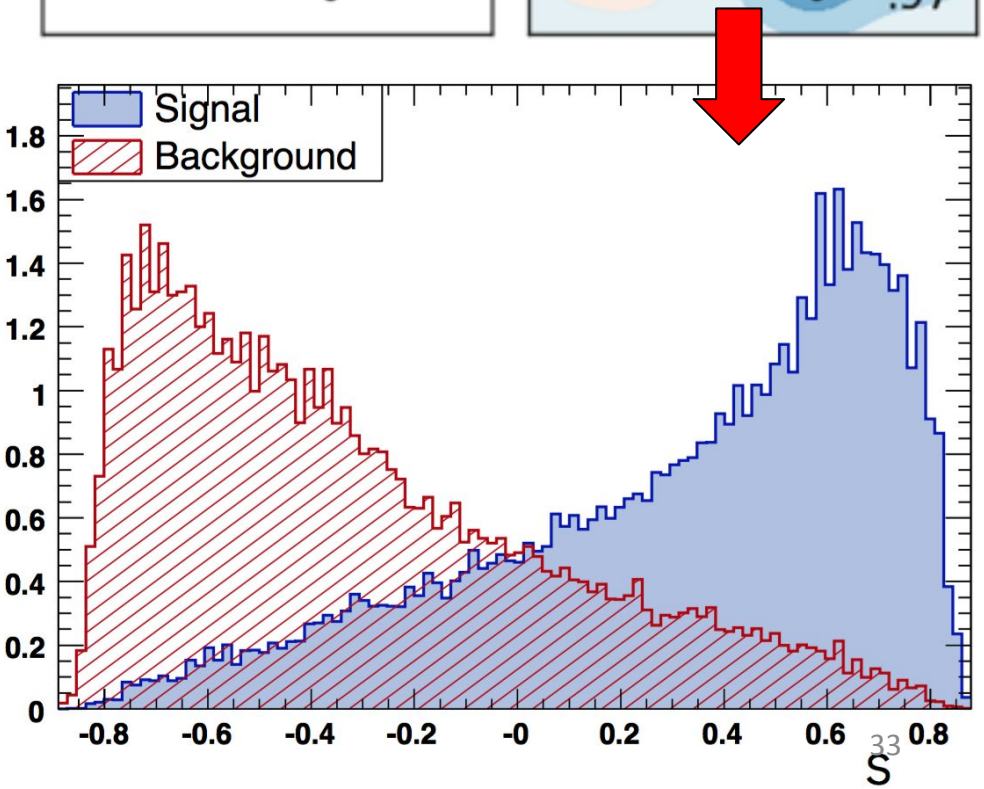


- Single tree (CART) <1980
- AdaBoost 1997 : rerun increasing the weight of misclassified entries
  Boosted Decision Trees (Gradient BDT XGBoost, random forest…)

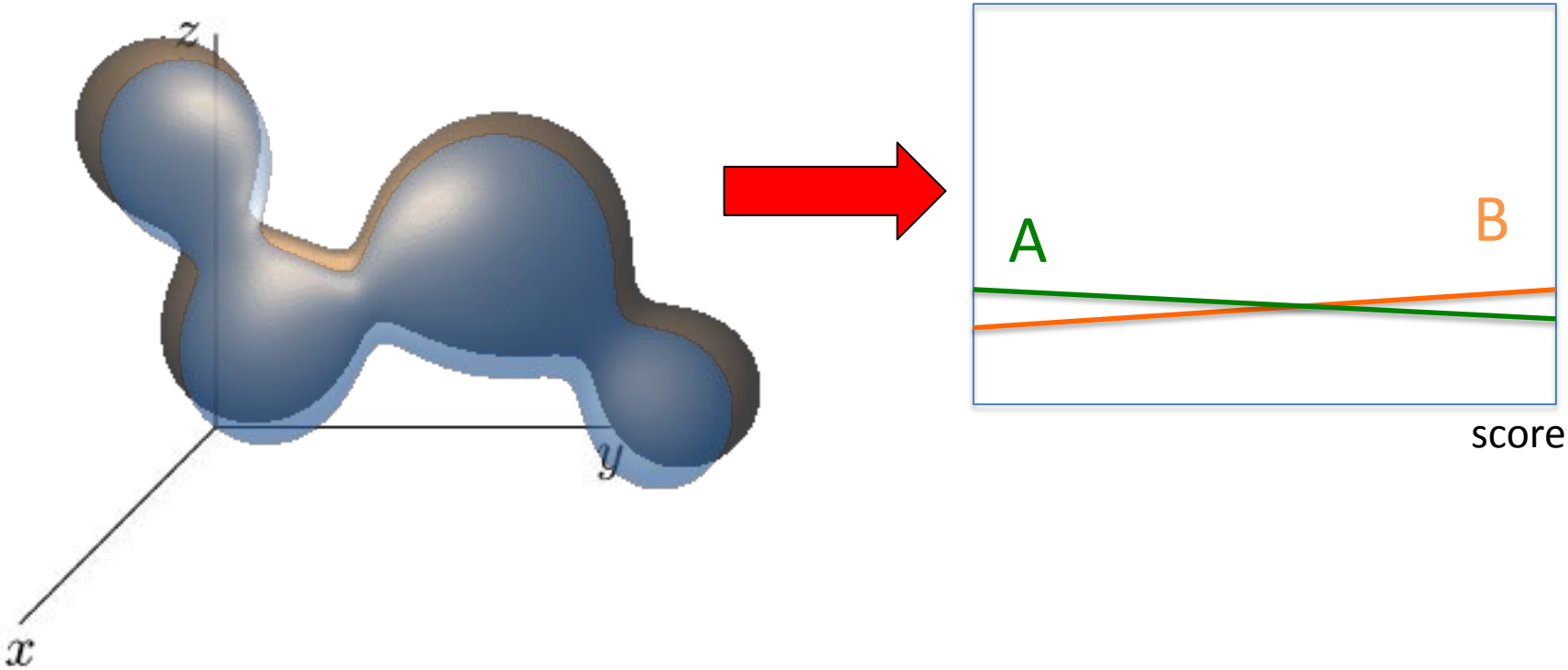
Classifier


B
S


.97


Signal
Background
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.8
-0.6
-0.4
-0.2
-0
0.2
0.4
0.6
0.8
S

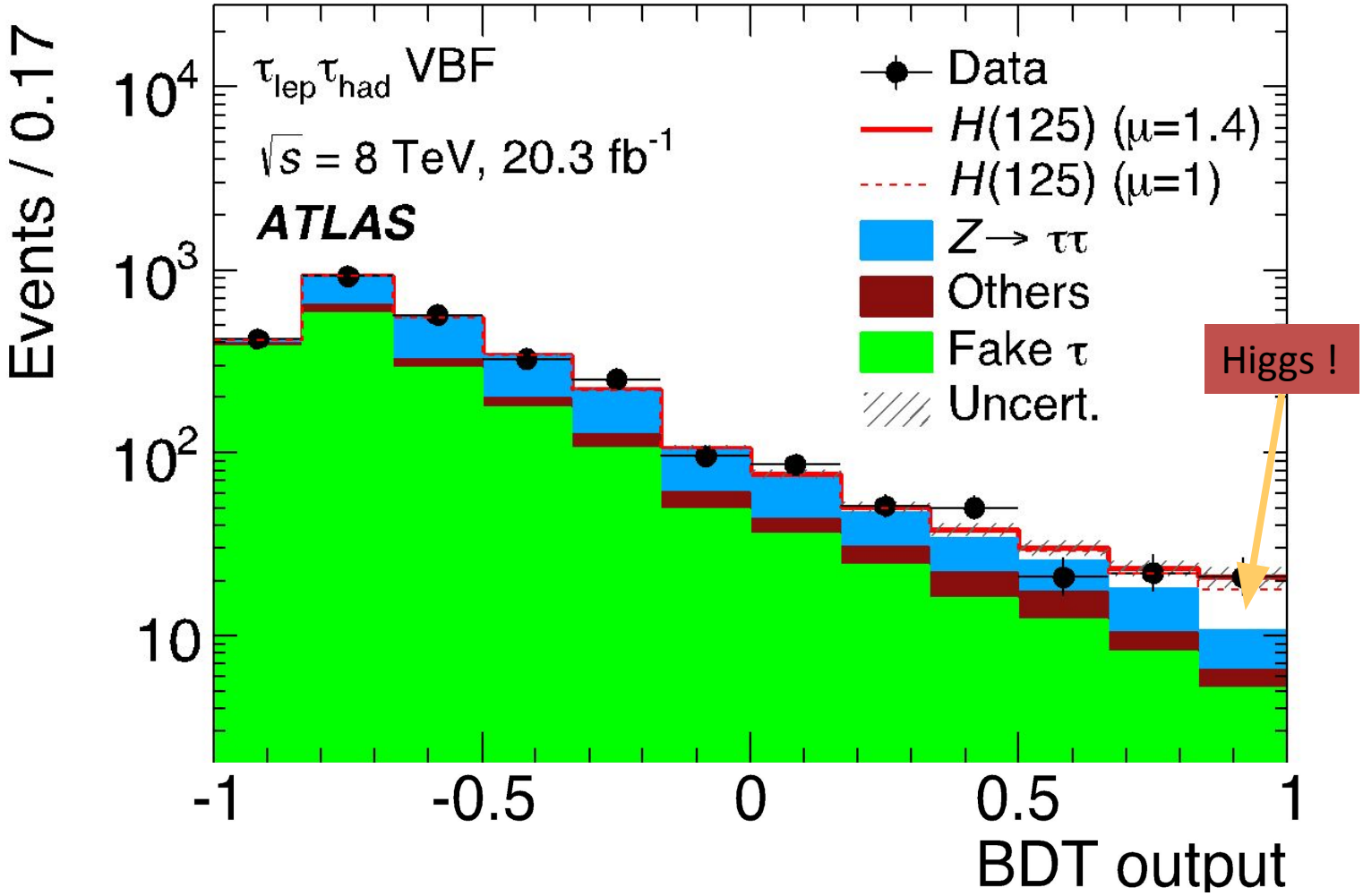
Qu’est ce qu’un classifieur fait?

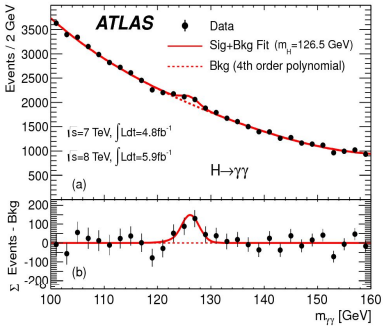


- Le classifieur “projette” les deux “blobs” multidimensionnels en maximisant leur difference

Application H→$\tau^+\tau^-$

BDT sur ~10 variables : masses invariantes, angles, etc…

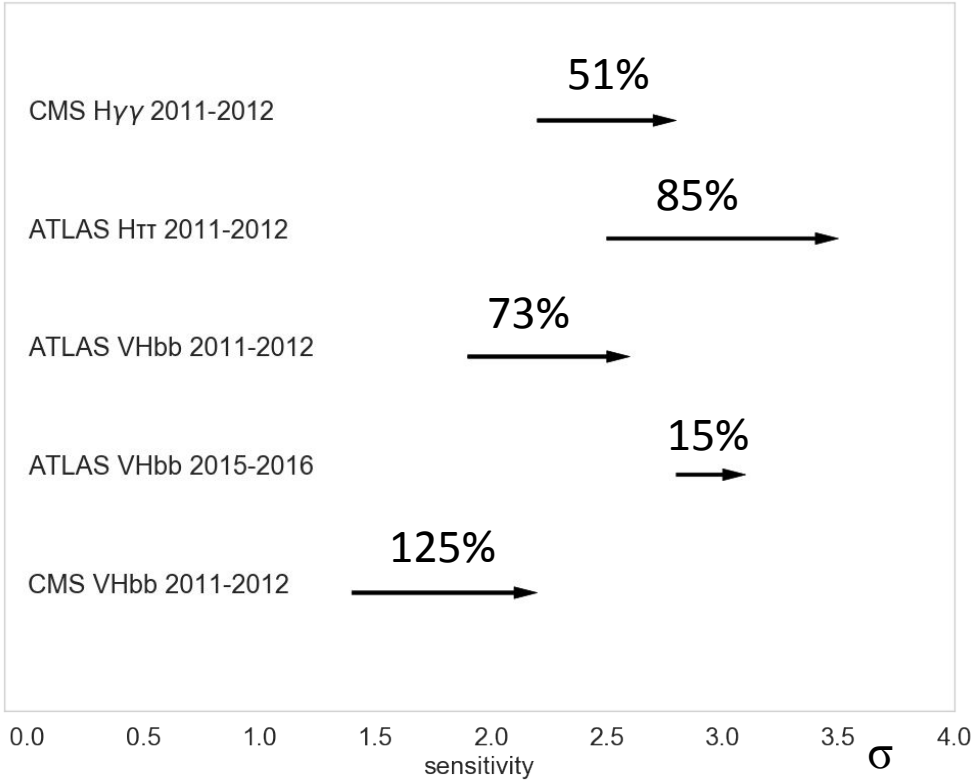




JHEP 04, 117 (2015) 1501.04943

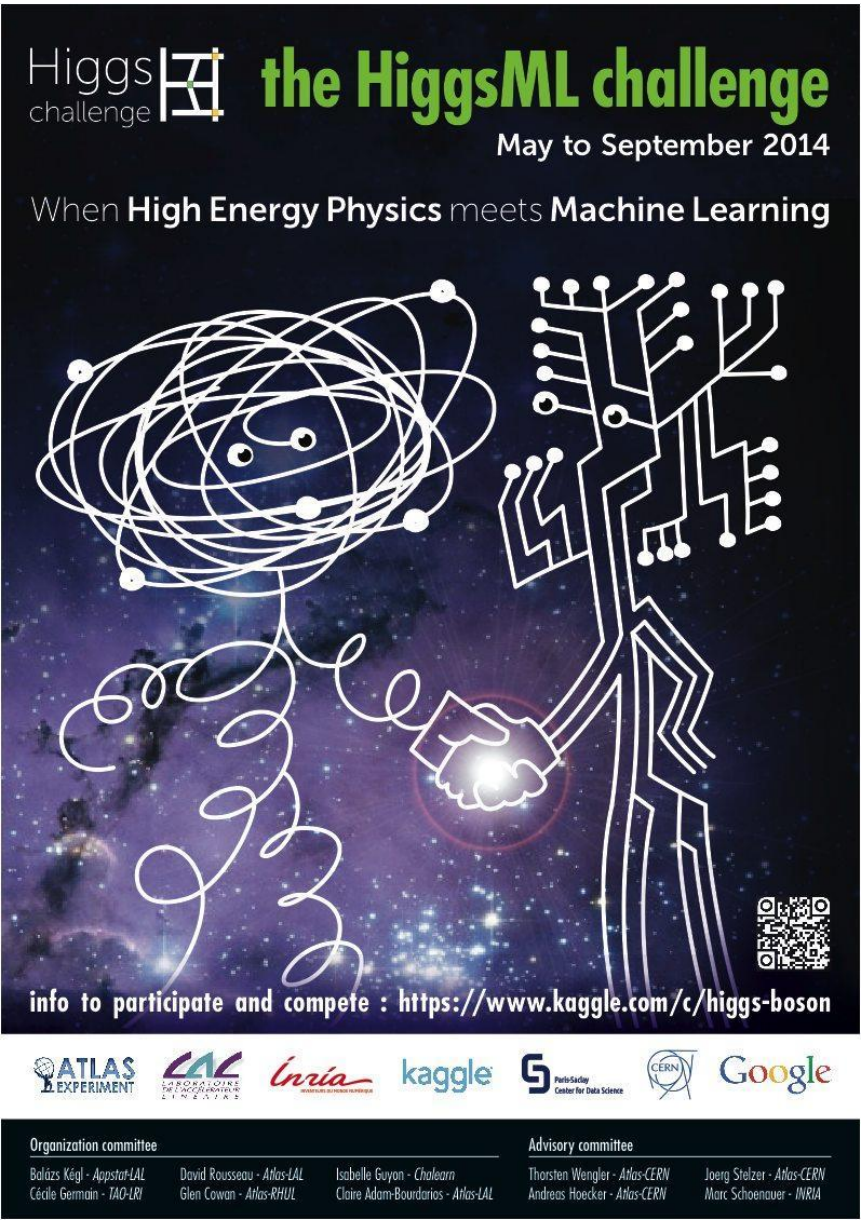
- En général, jusqu'en ~2015, nous n'avons utilisé que les Arbres de Décision Boosté sur une dizaine de variables
- Impact sur la sensibilité de découverte du boson de Higgs:

D'après *Nature* **560**, 41–48 (2018)



équivalent à ~50% de données en plus
(le LHC a coûté 4 milliards d'euros, budget du cern 1 milliard CHF par an)

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

Completed • $13,000 • 1,785 teams

Higgs Boson Machine Learning Challenge

Mon 12 May 2014 – Mon 15 Sep 2014 (34 days ago)

Dashboard

Private Leaderboard - Higgs Boson Machine Learning Challenge

This competition has completed. This leaderboard reflects the final standings.

See someone using multiple accounts?
Let us know.

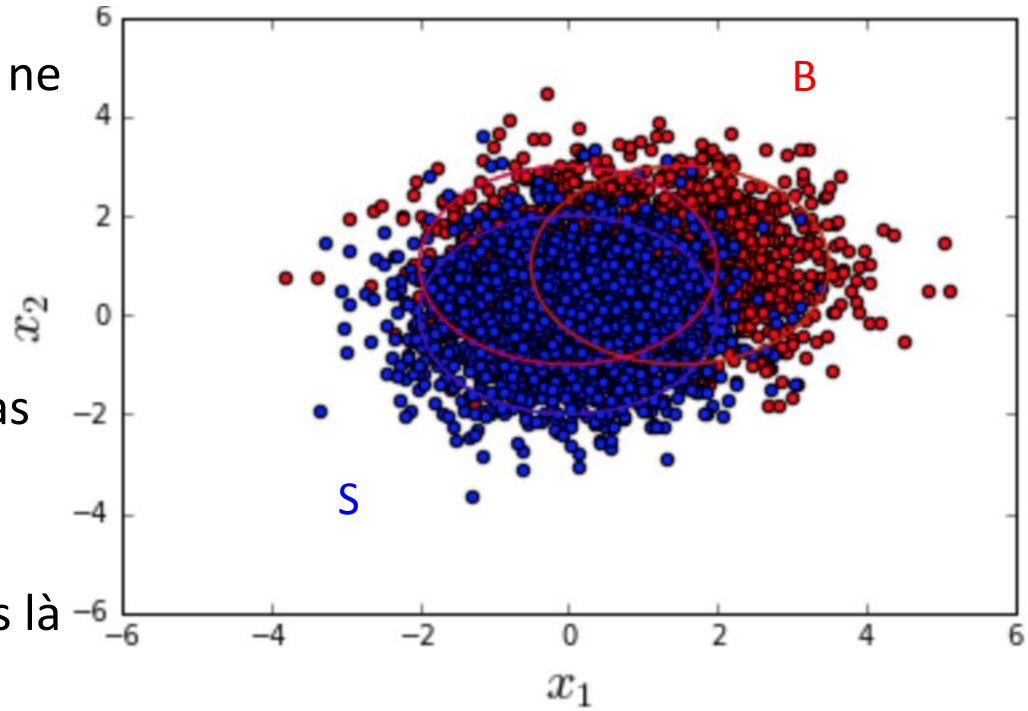
Ecart standard

#	Δ1w	Team Name ‡ model uploaded * in the money		Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑4	Gábor Melis ‡ *	7000$	3.80581	110	Sun, 14 Sep 2014 09:10:04 (-0h)
2	↓1	Tim Salimans ‡ *	4000$	3.78913	57	Mon, 15 Sep 2014 23:49:02 (-40.6d)
3	—	nhlx5haze ‡ *	2000$	3.78682	254	Mon, 15 Sep 2014 16:50:01 (-76.3d)
4	↑55	ChoKo Team		3.77526	216	Mon, 15 Sep 2014 15:21:36 (-42.1h)
5	↑23	cheng chen		3.77384	21	Mon, 15 Sep 2014 23:29:29 (-0h)
6	↓2	quantify		3.77086	8	Mon, 15 Sep 2014 16:12:48 (-7.3h)
7	↑73	Stanislav Semenov & Co (HSE Yandex)		3.76211	68	Mon, 15 Sep 2014 20:19:03
8	↓1	Luboš Motl's team		3.76050	589	Mon, 15 Sep 2014 08:38:49 (-1.6h)
9	↓1	Roberto-UCIIIM		3.75864	292	Mon, 15 Sep 2014 23:44:42 (-44d)
10	↑5	Davut & Josef		3.75838	161	Mon, 15 Sep 2014 23:24:32 (-4.5d)

991 TMVA (CERN standard) 3.2

Pas de miracle

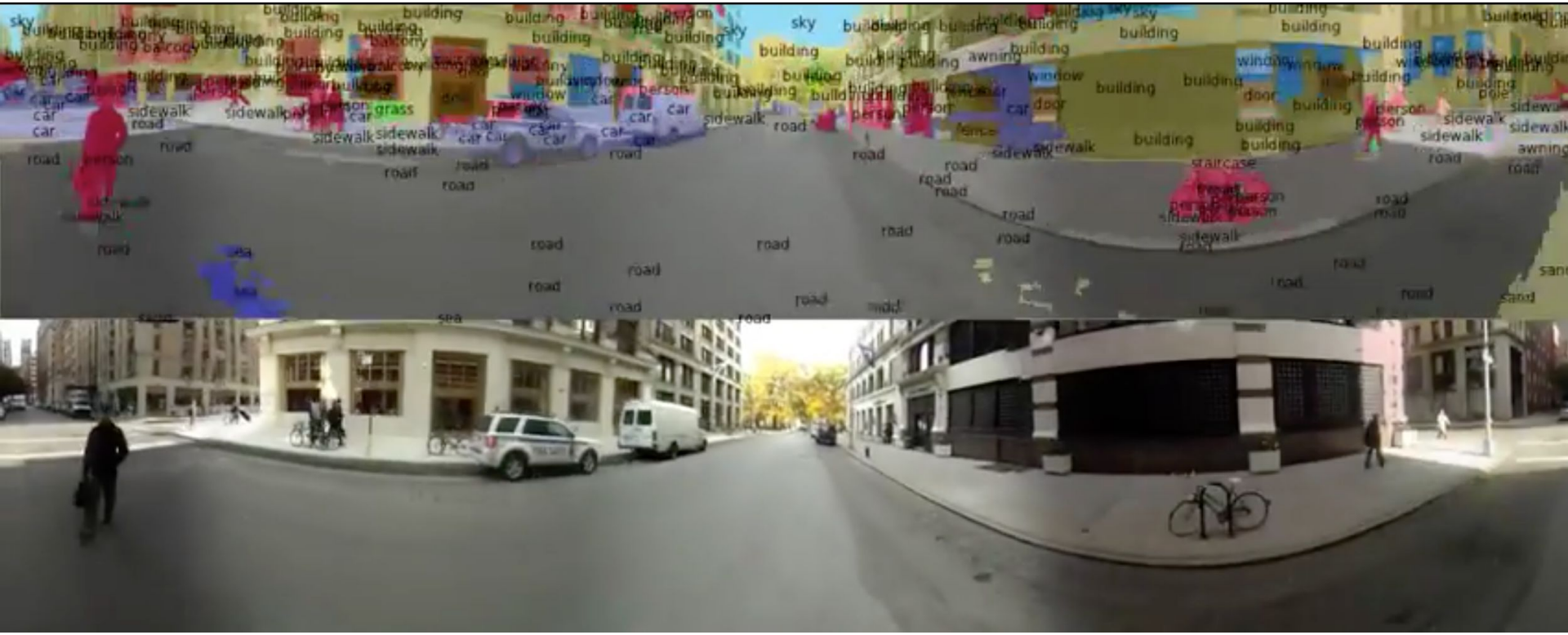
- Si les fonctions de densité de probabilité (pdf) sont connues, on ne peut faire mieux que le rapport de vraisemblance (théorème de Neyman-Pearson) :
  - $L_S(x)/L_B(x)$
- ❑ OK mais en général L_S L_B ne sont pas connus (même pas leur forme)
  - ❑ + x de grande dimension
- ML devient intéressant dans ce cas là



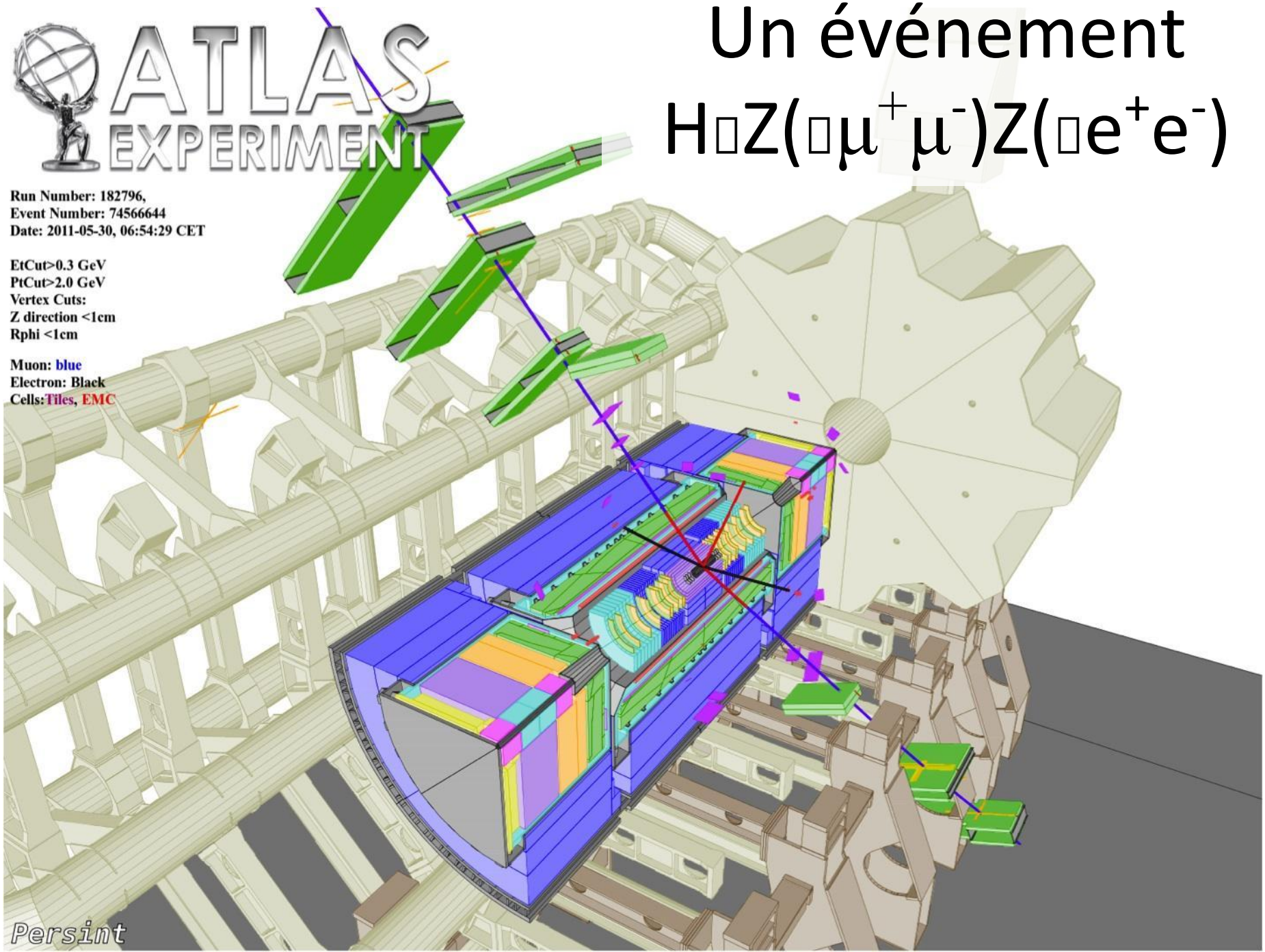
- approche mixte

Identification

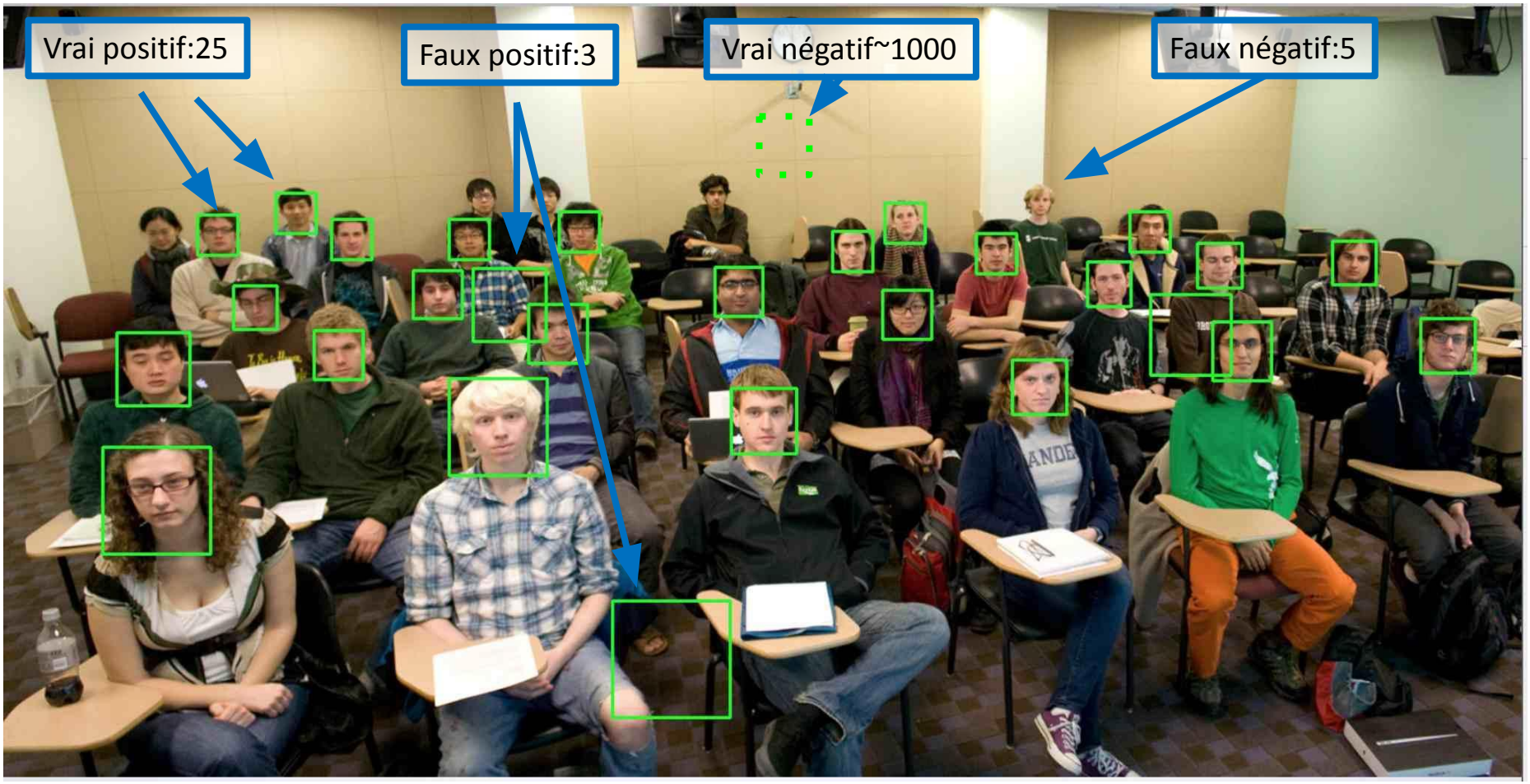
Typical Deep Learning application



Un événement $H \to Z(\to \mu^+\mu^-)Z(\to e^+e^-)$



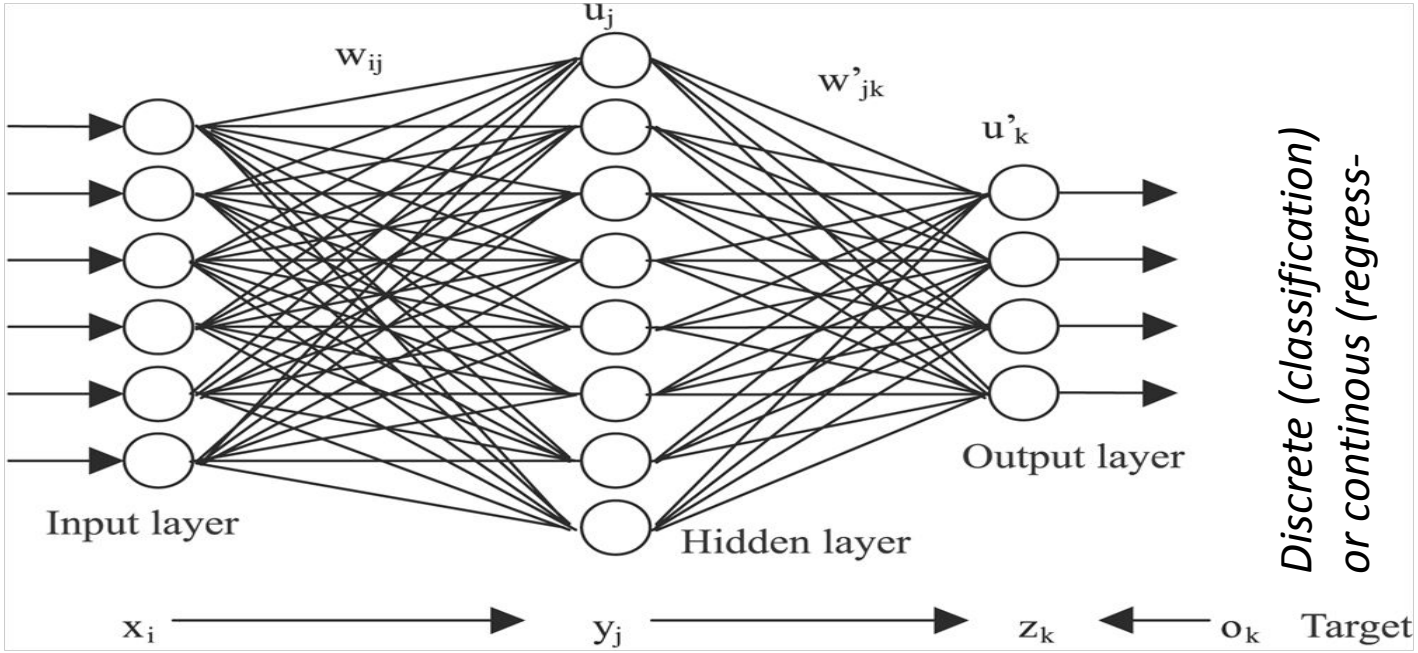
Real-time face detection



Efficacité
Pureté
Vitesse

Neural Networks

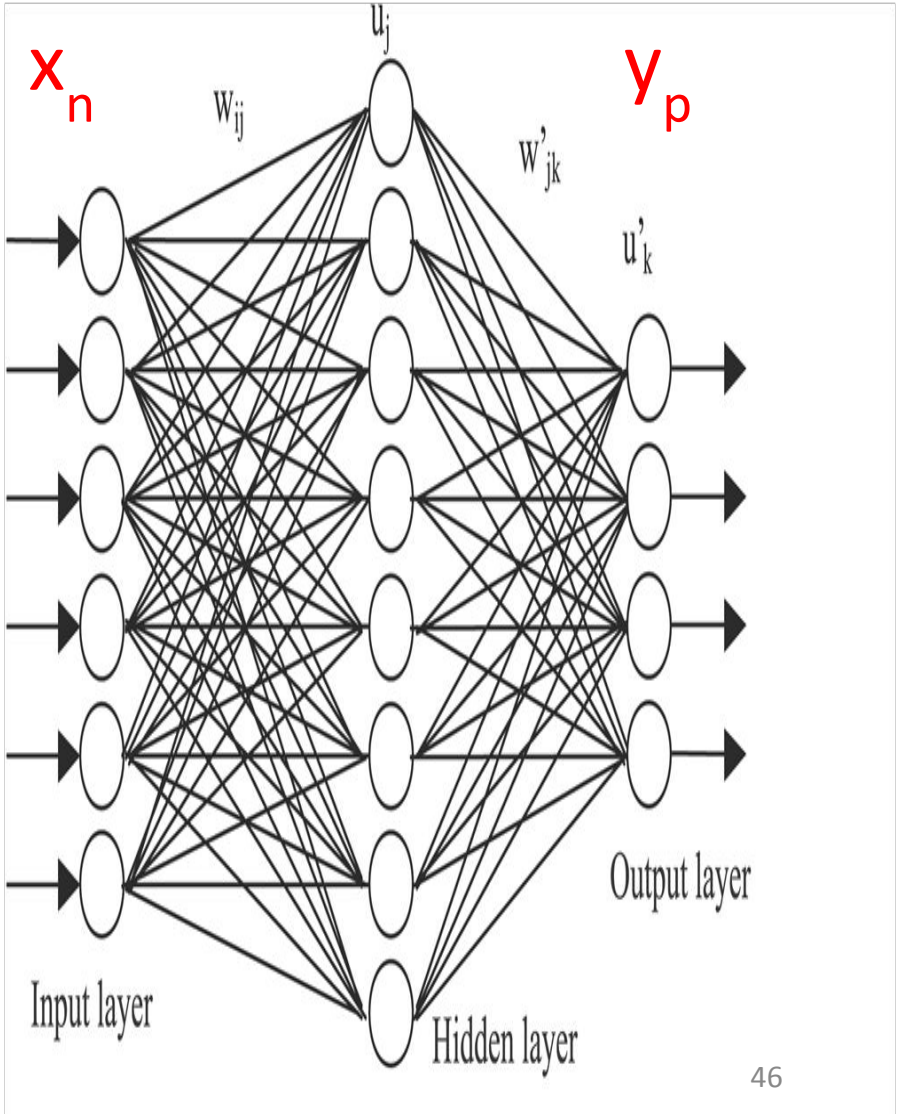
Neural Net in a nutshell



- Neural Net ~1950!
- But many many new tricks for learning, in particular if many layers (also ReLU instead of sigmoïd activation)
- “Deep Neural Net” hundreds layers
- Computing power (DNN training can take days even on GPU)

Universal Approximation theorem

- Mathematical theorem 1991
  https://en.wikipedia.org/wiki/Universal_approximation_theorem
- Any continuous, bounded function $R^n \square R^p$
- … can be approximately sufficiently well (better than a given ε)
- … with a sufficiently large single hidden layer neural net
- But how to build it ?

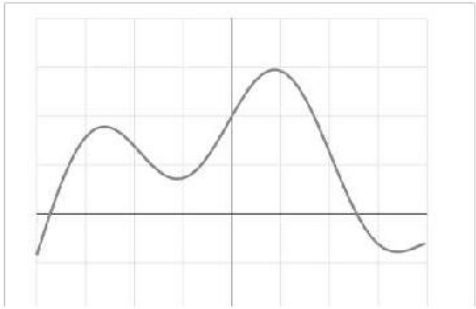


Addendum ResNet 1 neuron sufficient depth

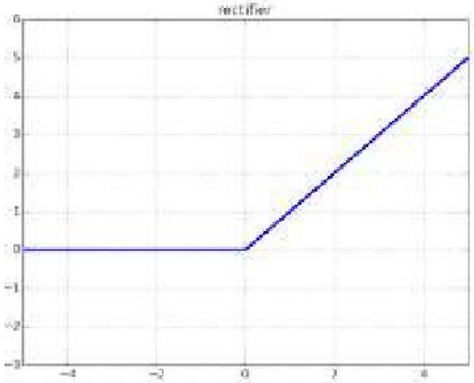
Universal Theorem at work

Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions

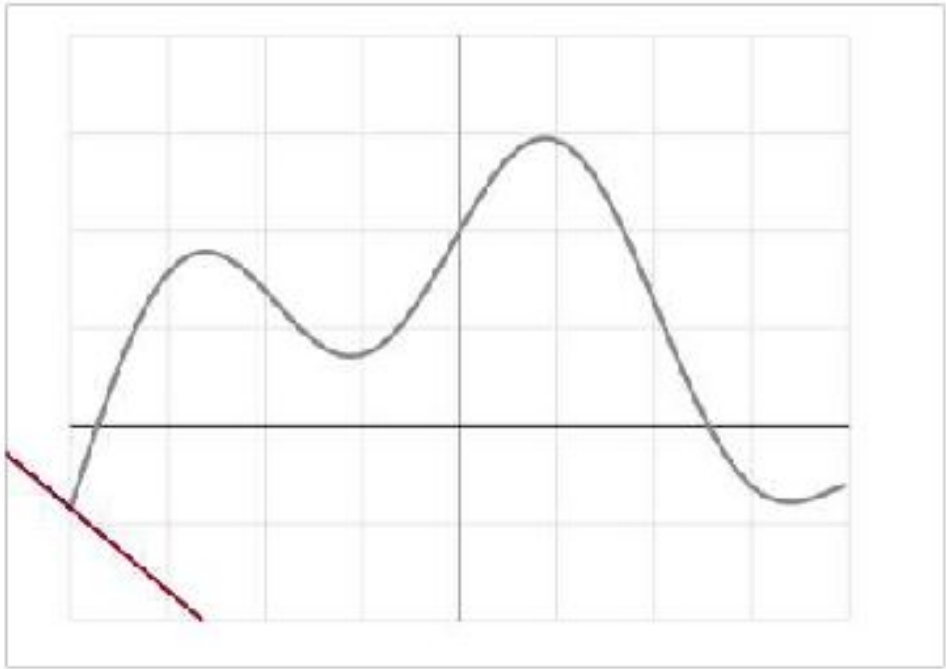


relu(x) = x if x>0 & 0 otherwise



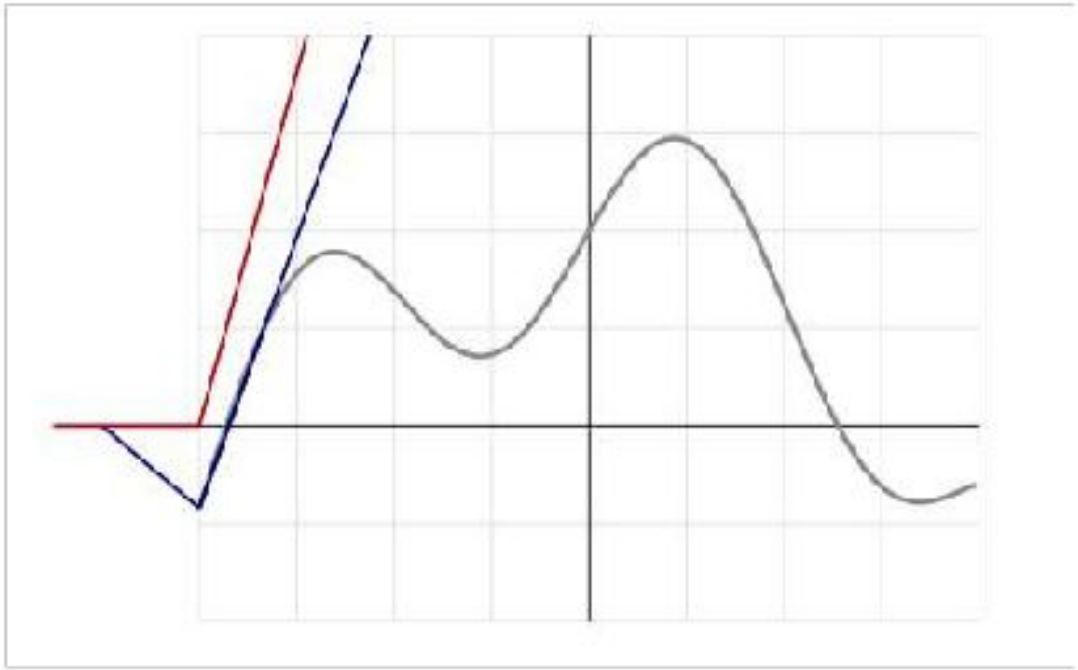
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



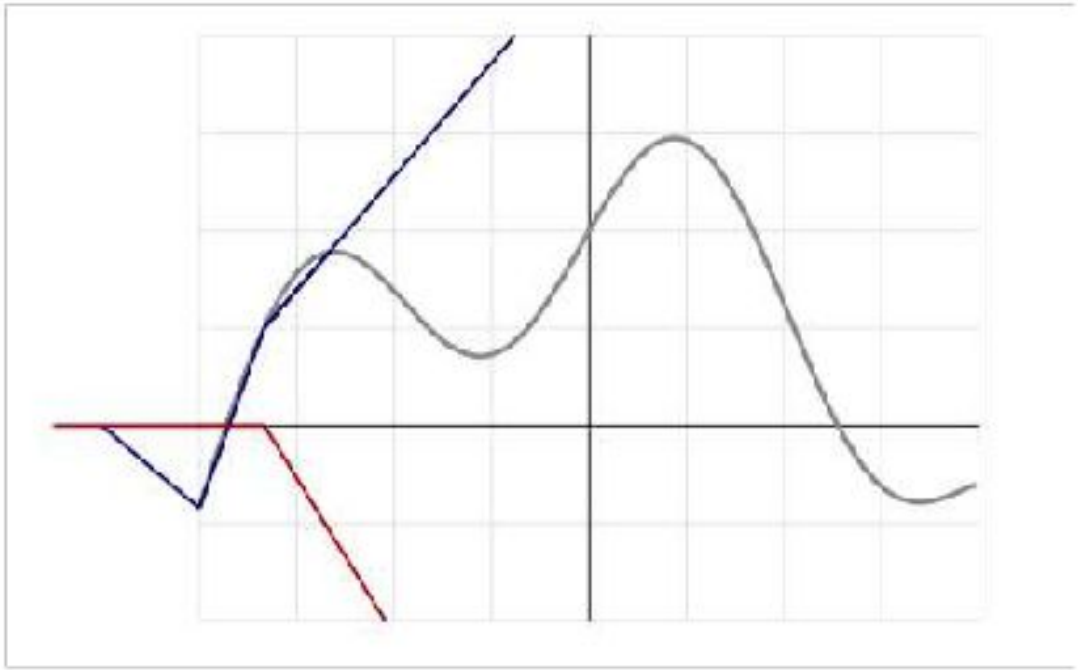
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



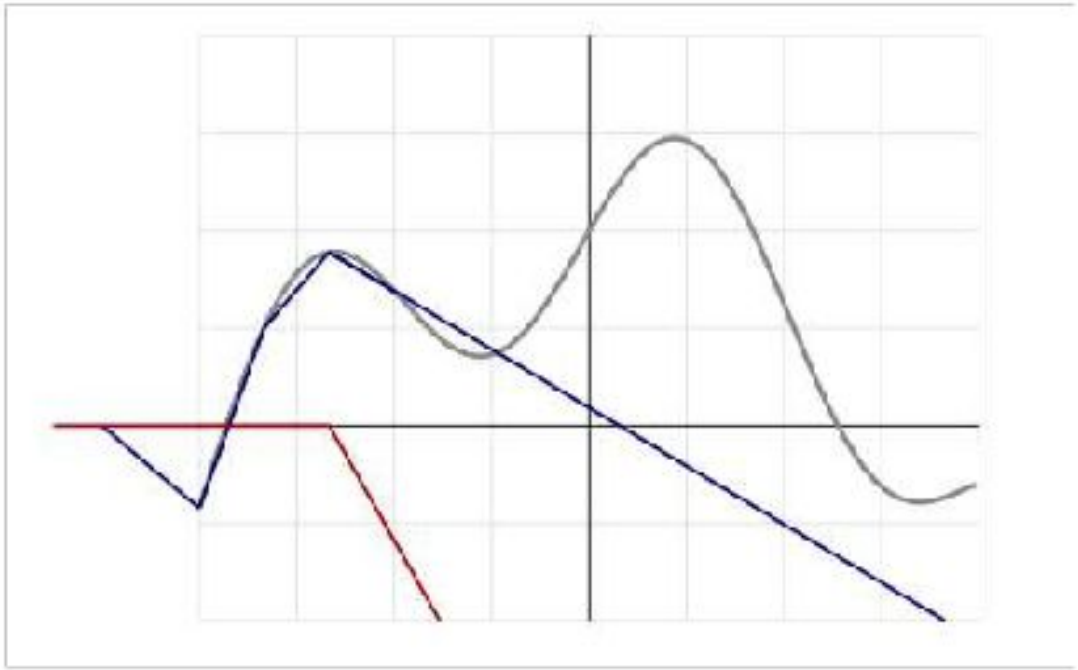
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



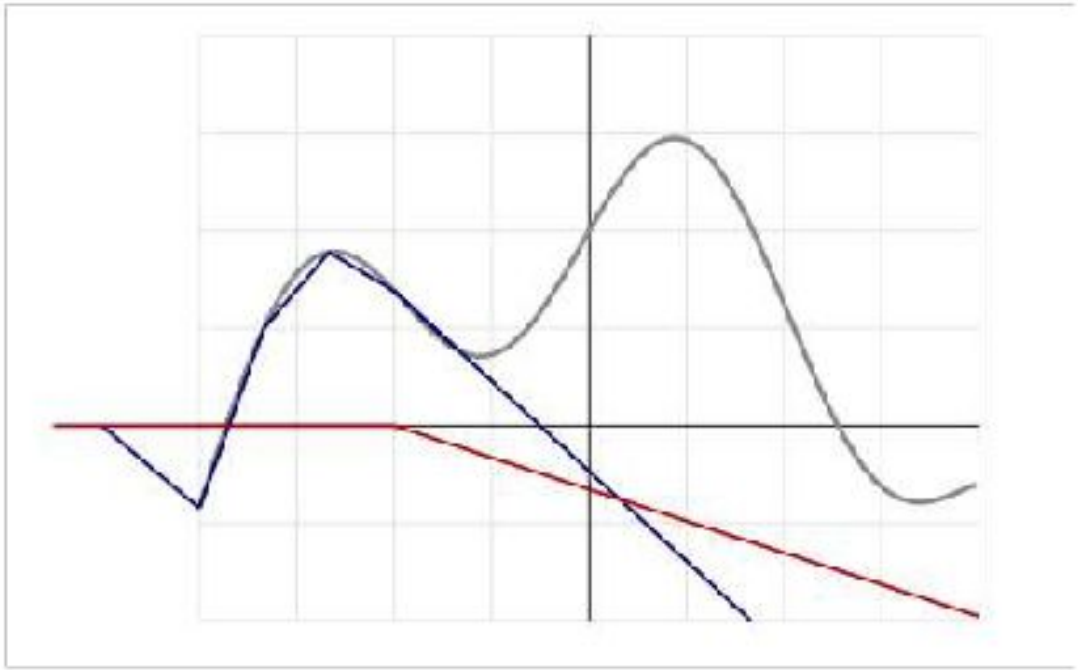
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



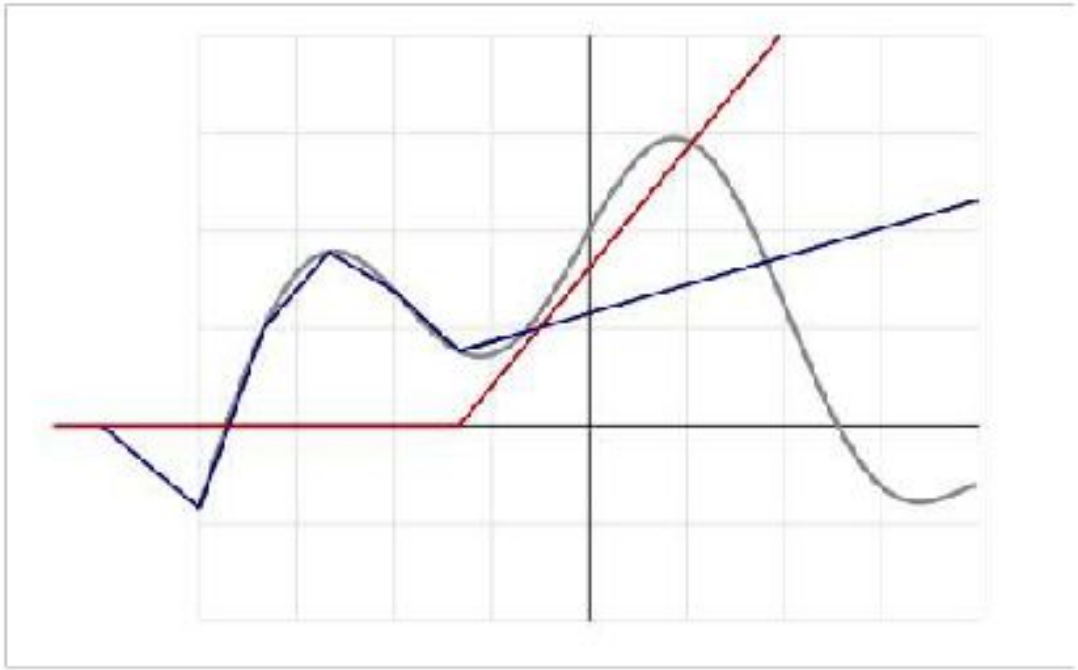
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



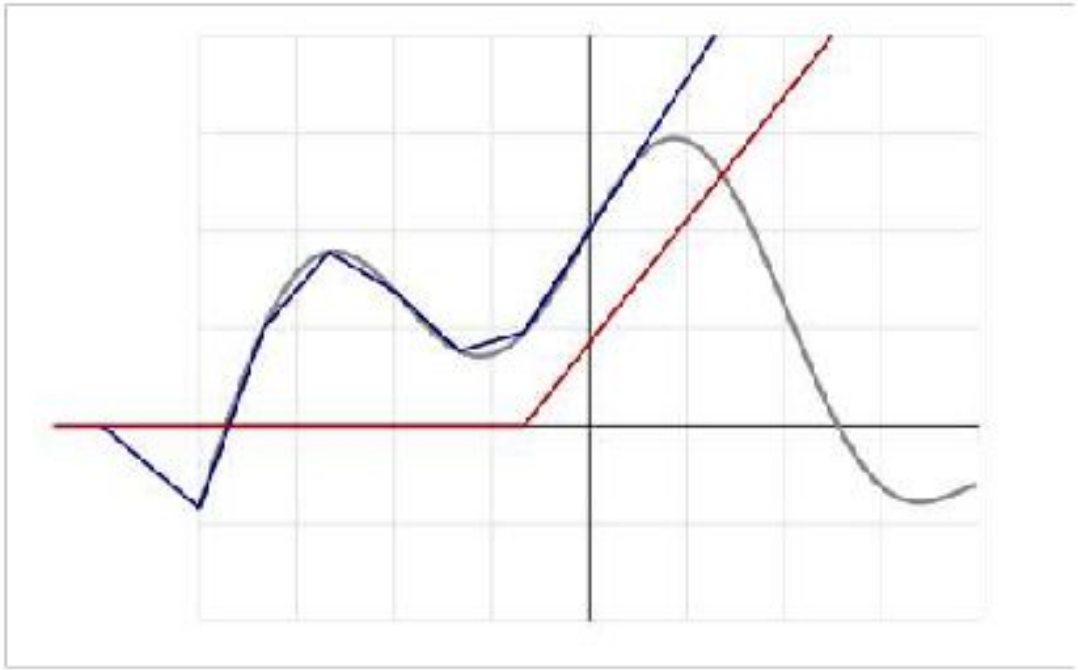
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



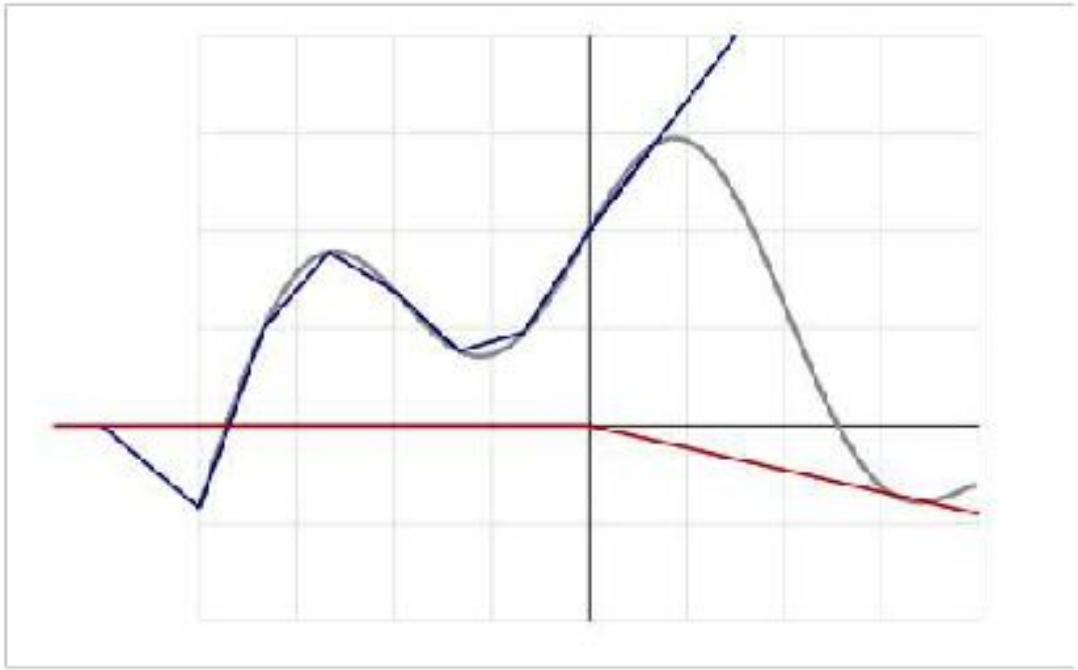
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



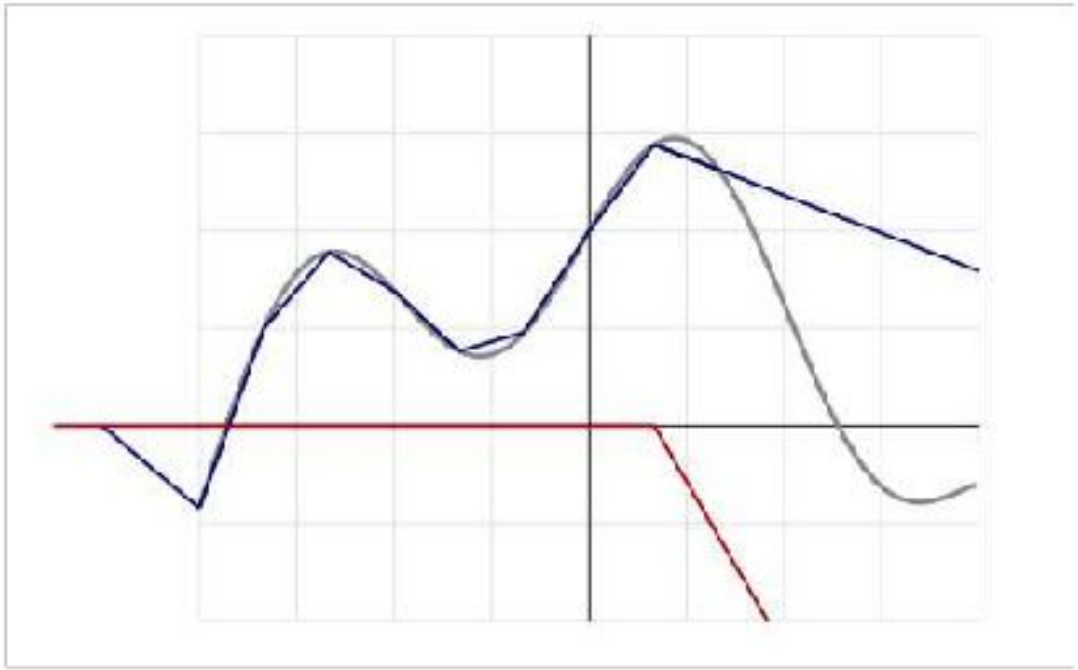
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



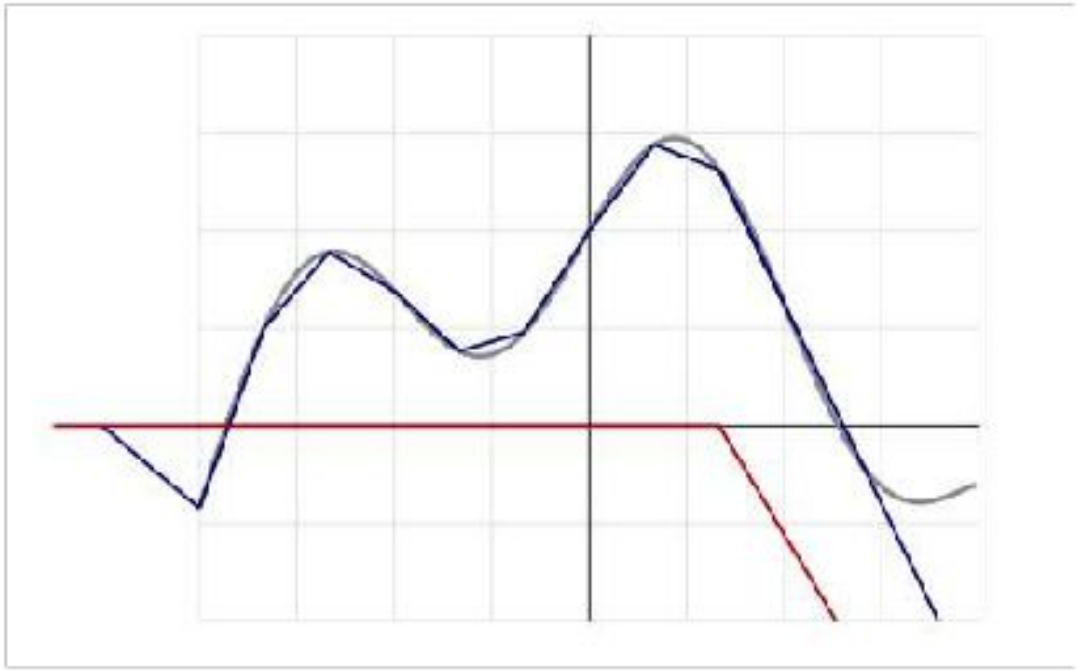
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



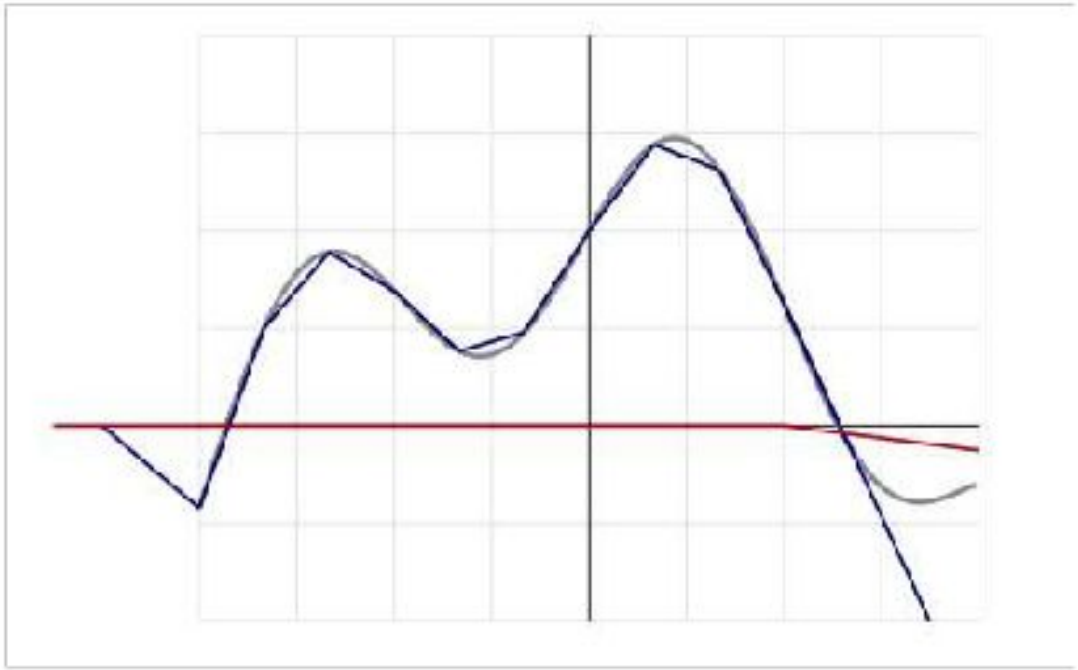
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



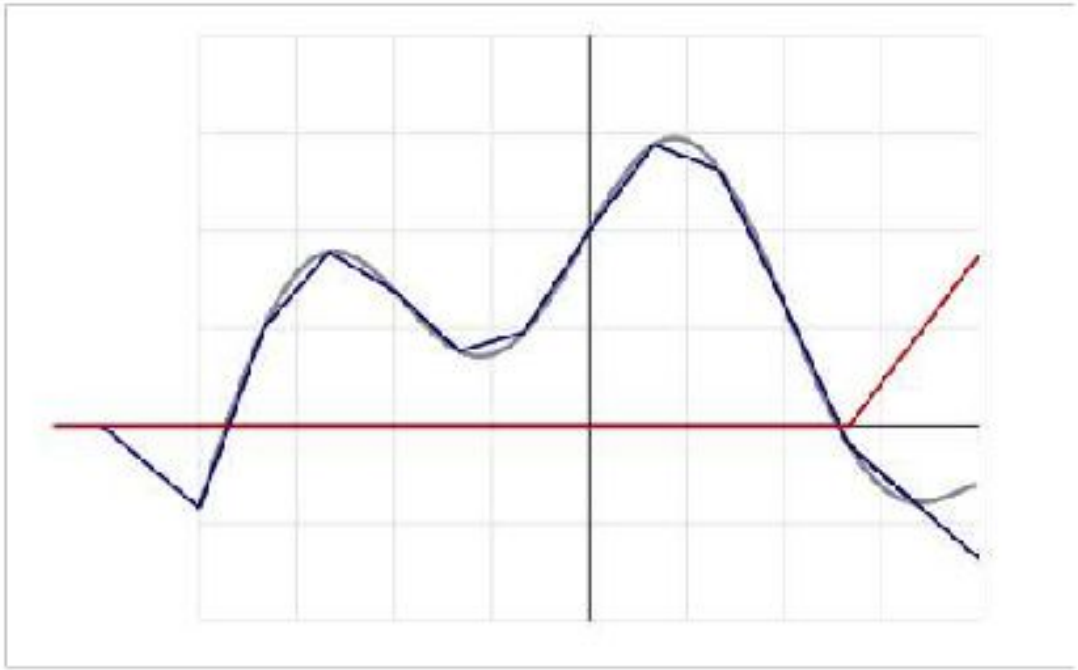
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



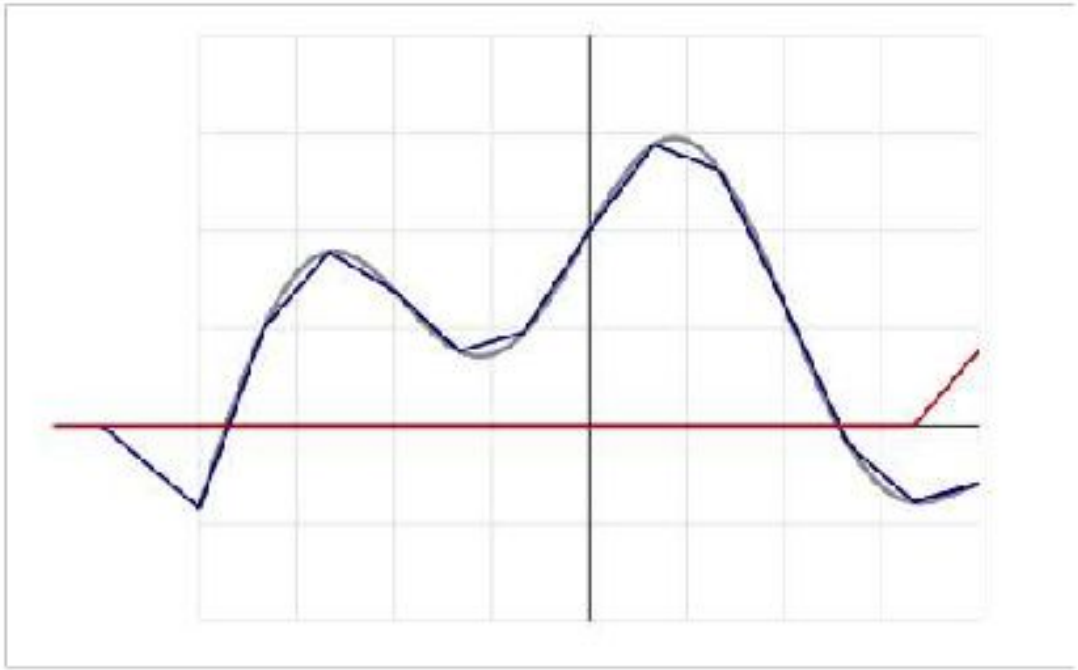
Universal approximation

We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions

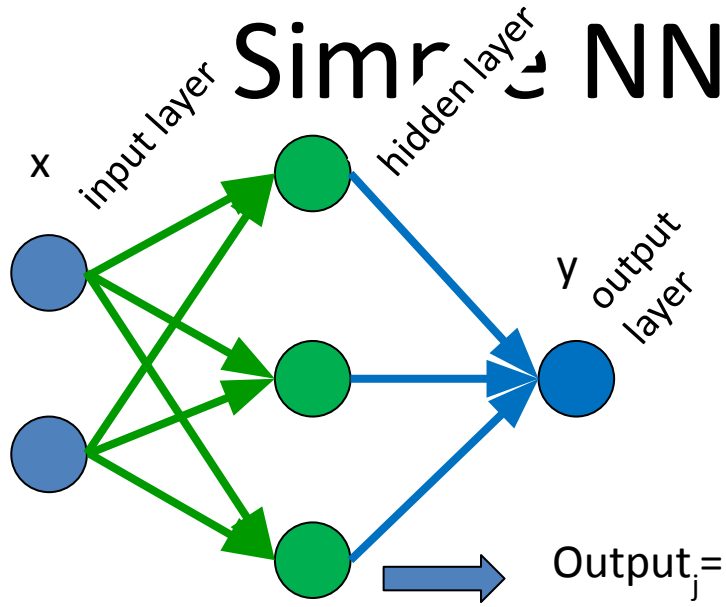


Universal approximation

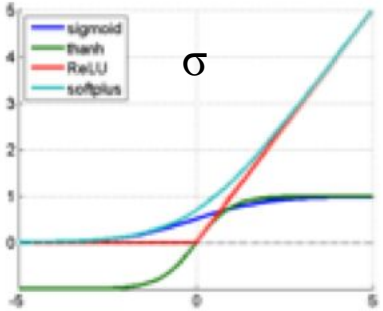
We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions



Simple NN



activation function



Output$_j$= $\sigma(b_j + \sum w_{ij} x_i)$

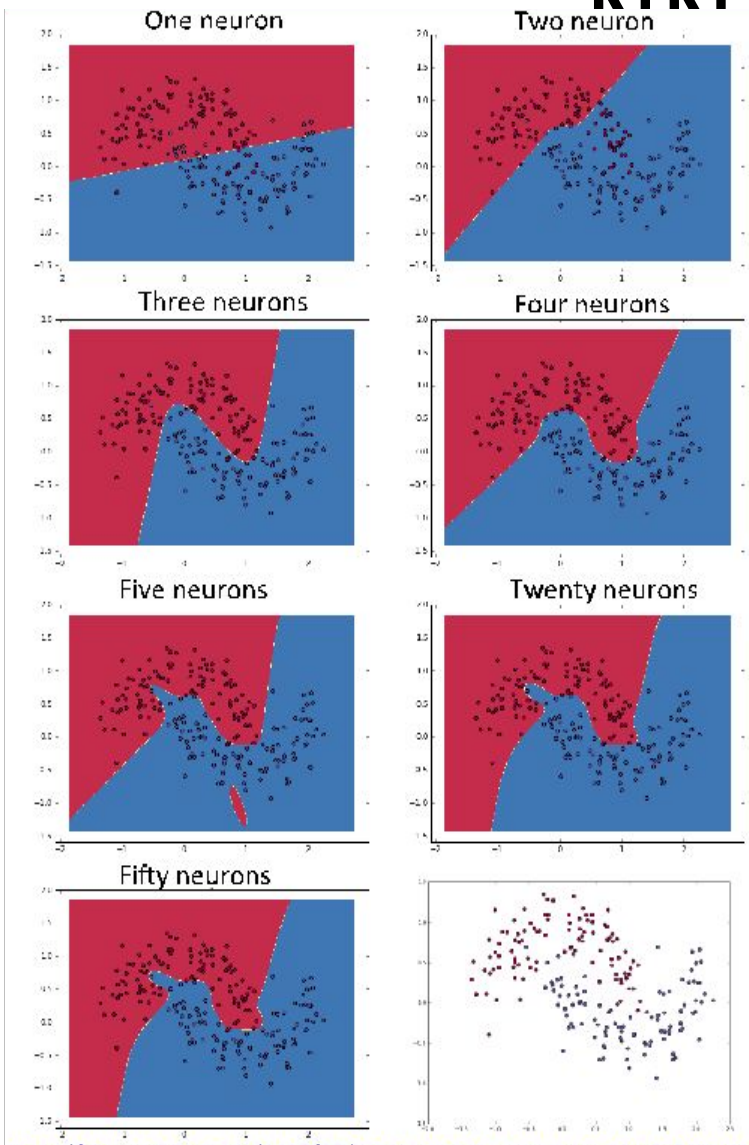
$$h(x) = \sigma(b^2 + W^2 \sigma(b^1 + W^1 x))$$

Beware: superscript are layer indices!

Now with dimensions

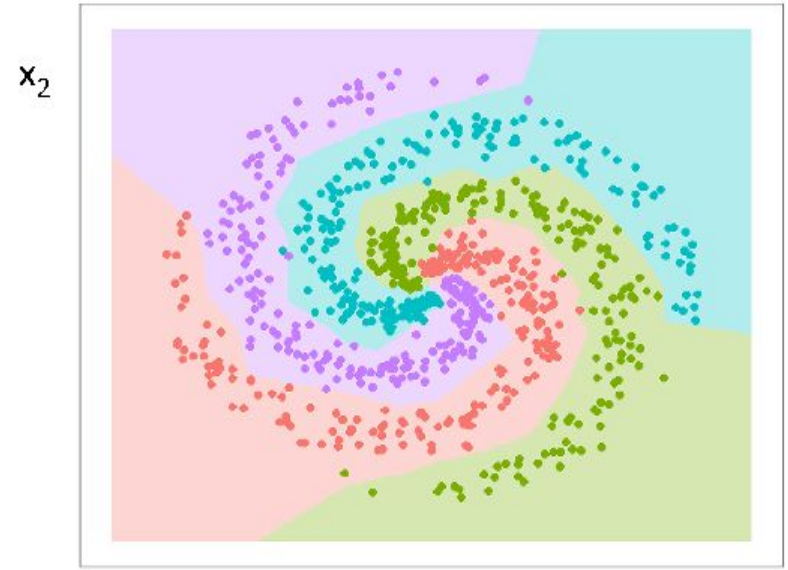
$$h(x_{(2)}) = \sigma(b^2_{(1)} + W^2_{(1,3)} \sigma(b^1_{(3)} + W^1_{(3,2)} x_{(2)}))$$

NN at work



http://www.wildml.com/2015/09/implementing-a-neural-network-from-scratch/

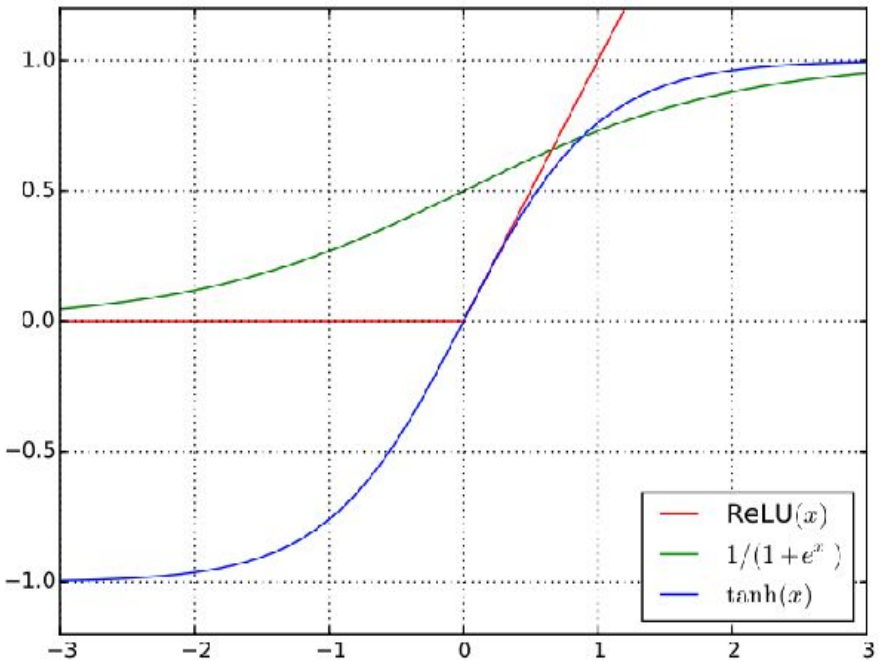
4-class classification
2-hidden layer NN
ReLU activations
L2 norm regularization



2-class classification
1-hidden layer NN
L2 norm regularization

http://junma5.weebly.com/data-blog/build-your-own-neural-network-classifier-in-r

Activation functions



~classic

- **Vanishing gradient problem**
  - Derivative of sigmoid:

$$\frac{\partial \sigma(x)}{\partial x} = \sigma(x)(1-\sigma(x))$$

  - Nearly 0 when x is far from 0!
  - Gradient descent difficult!

~NEW!

- **Rectified Linear Unit (ReLU)**
  - ReLU(x) = max{0, x}
  - Derivative is constant!

$$\frac{\partial ReLU(x)}{\partial x} = \begin{cases} 1 & when\ x > 0 \\ 0 & otherwise \end{cases}$$

  - ReLU gradient doesn't vanish

Aparté historique

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

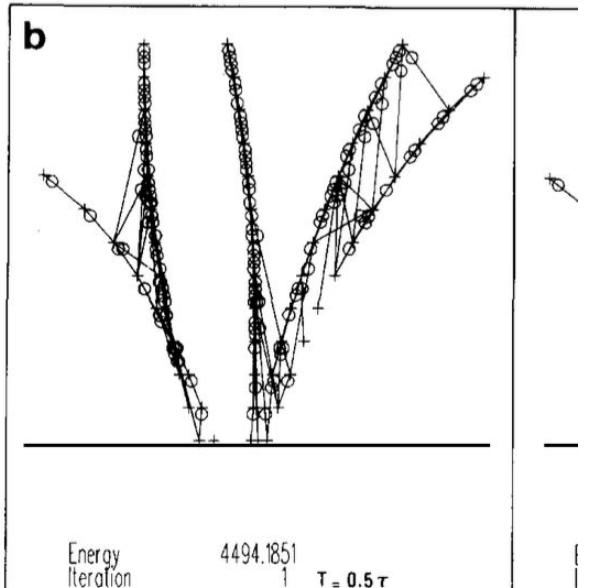
Computer Physics Communications 49 (1988) 429–448
North-Holland, Amsterdam

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

- 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