

Artificial Intelligence in High Energy Physics I

First, how did we see the Higgs Boson ?

David Rousseau

IJCLab-Orsay

rousseau@ijclab.in2p3.fr

Rencontres d'été de physique

de l'infiniment grand à l'infiniment petit

juillet 2023



La Matière: les FERMIONS

Modèle Standard
des particules

Quarks

u up	c charm	t top
d down	s strange	b bottom

Leptons

e electron	μ muon	τ tau
ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino

Les Forces: les BOSONS

?
La masse!

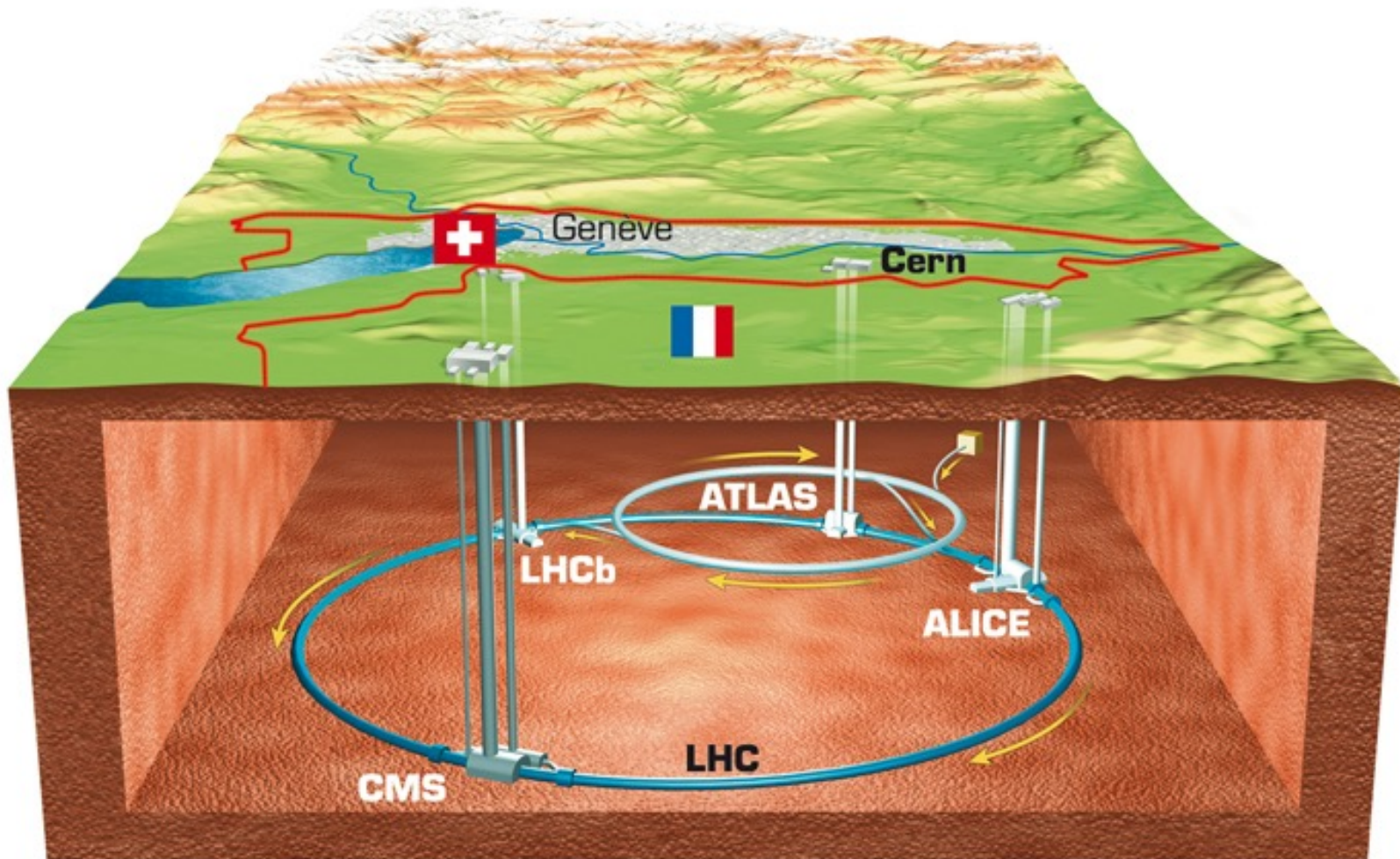
H
Higgs boson

Z Z boson	γ photon
W W boson	g gluon

+Antiparticules

VOIR les particules ?

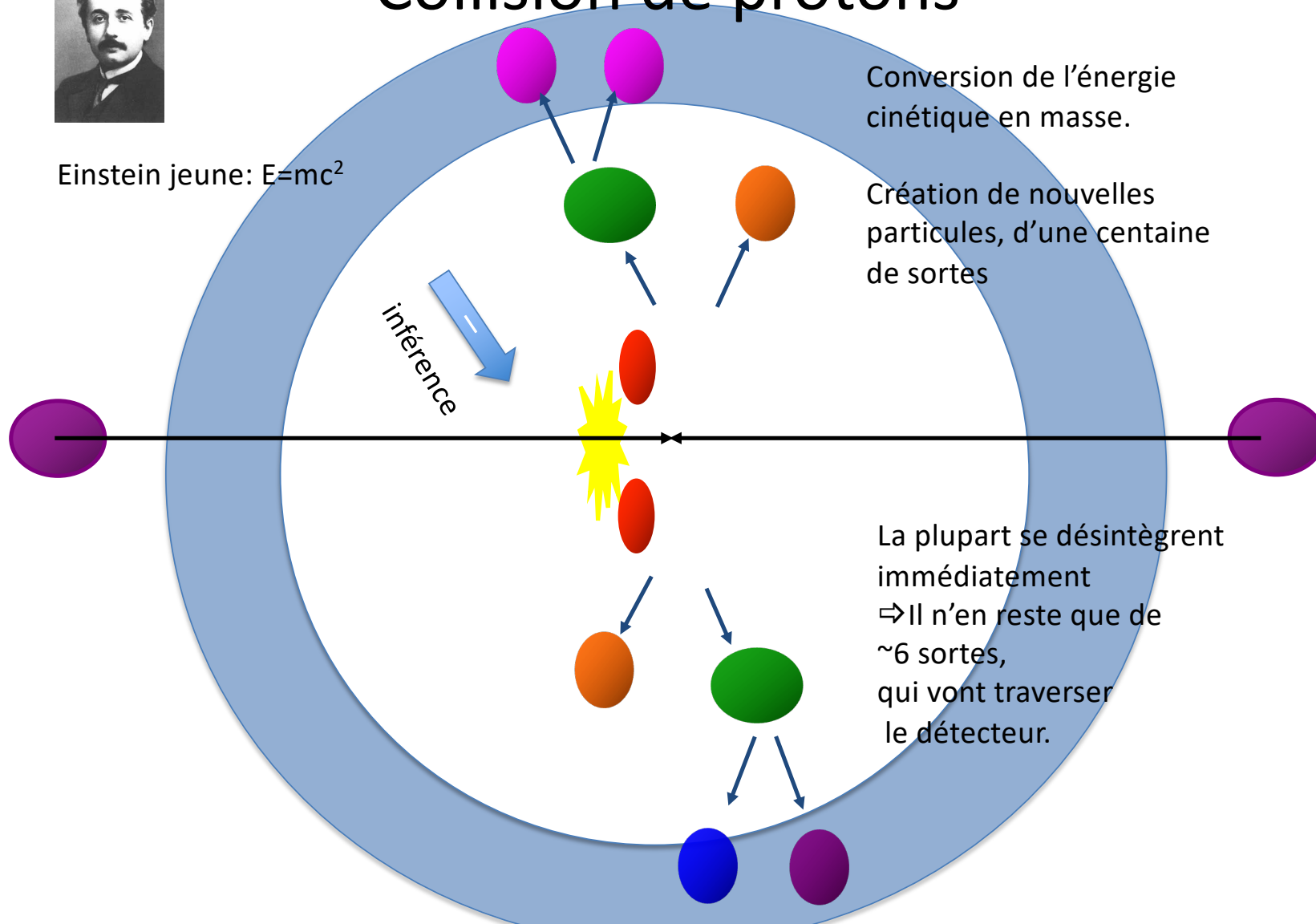
Le LHC





Einstein jeune: $E=mc^2$

Collision de protons



Conversion de l'énergie cinétique en masse.

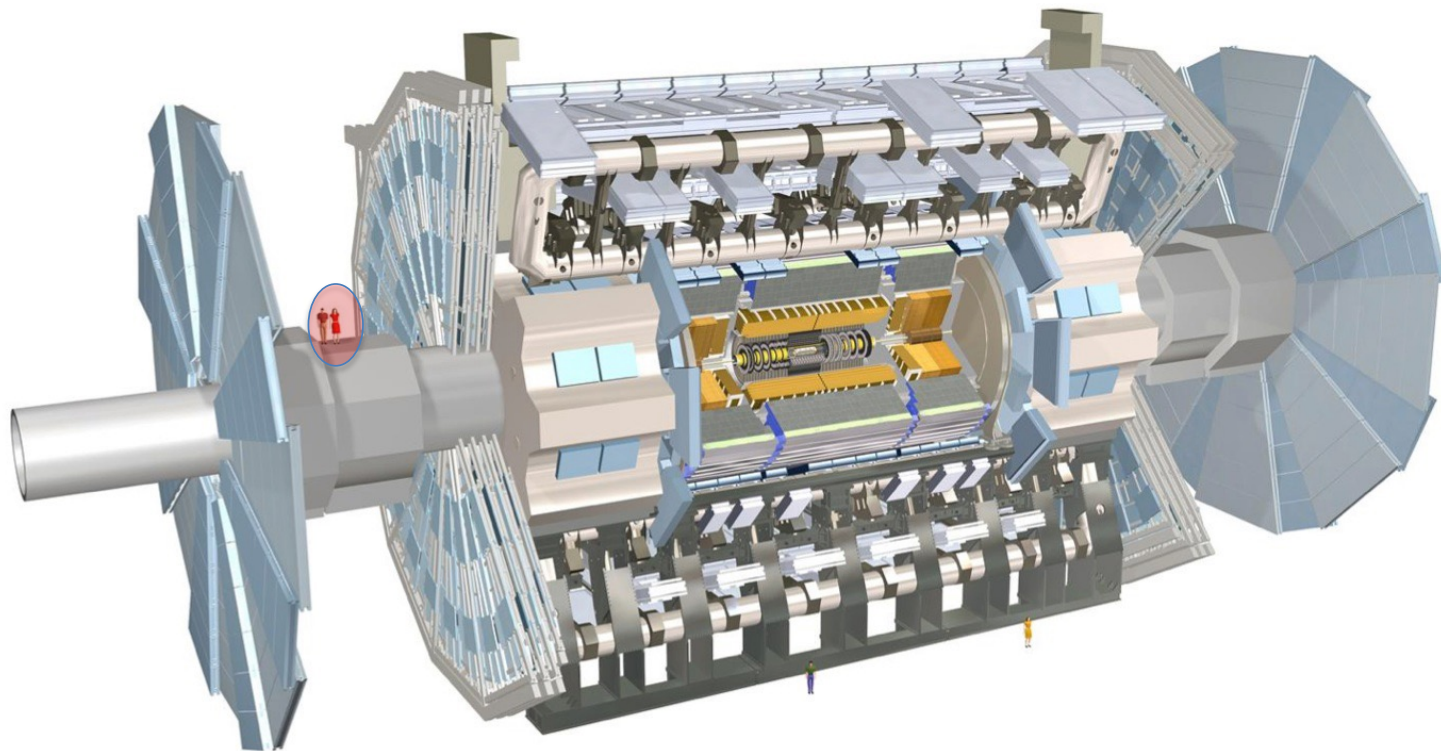
Création de nouvelles particules, d'une centaine de sortes

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

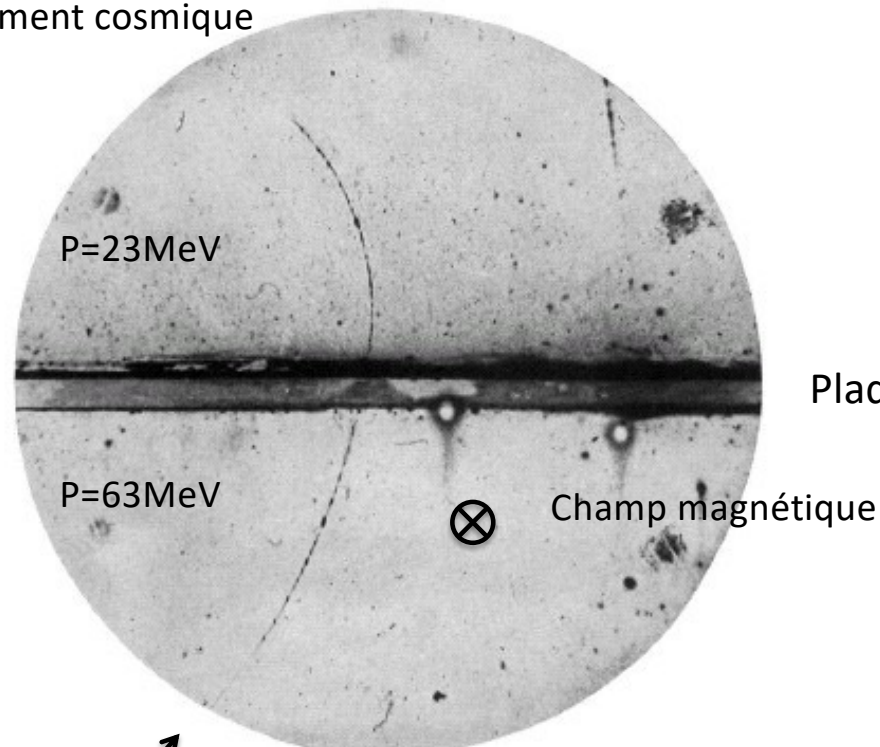


FIG. 1. A 63 million volt positron ($H_p = 2.1 \times 10^6$ gauss-cm) passing through a 6 mm lead plate and emerging as a 23 million volt positron ($H_p = 7.5 \times 10^4$ gauss-cm). The length of this latter path is at least ten times greater than the possible length of a proton path of this curvature.

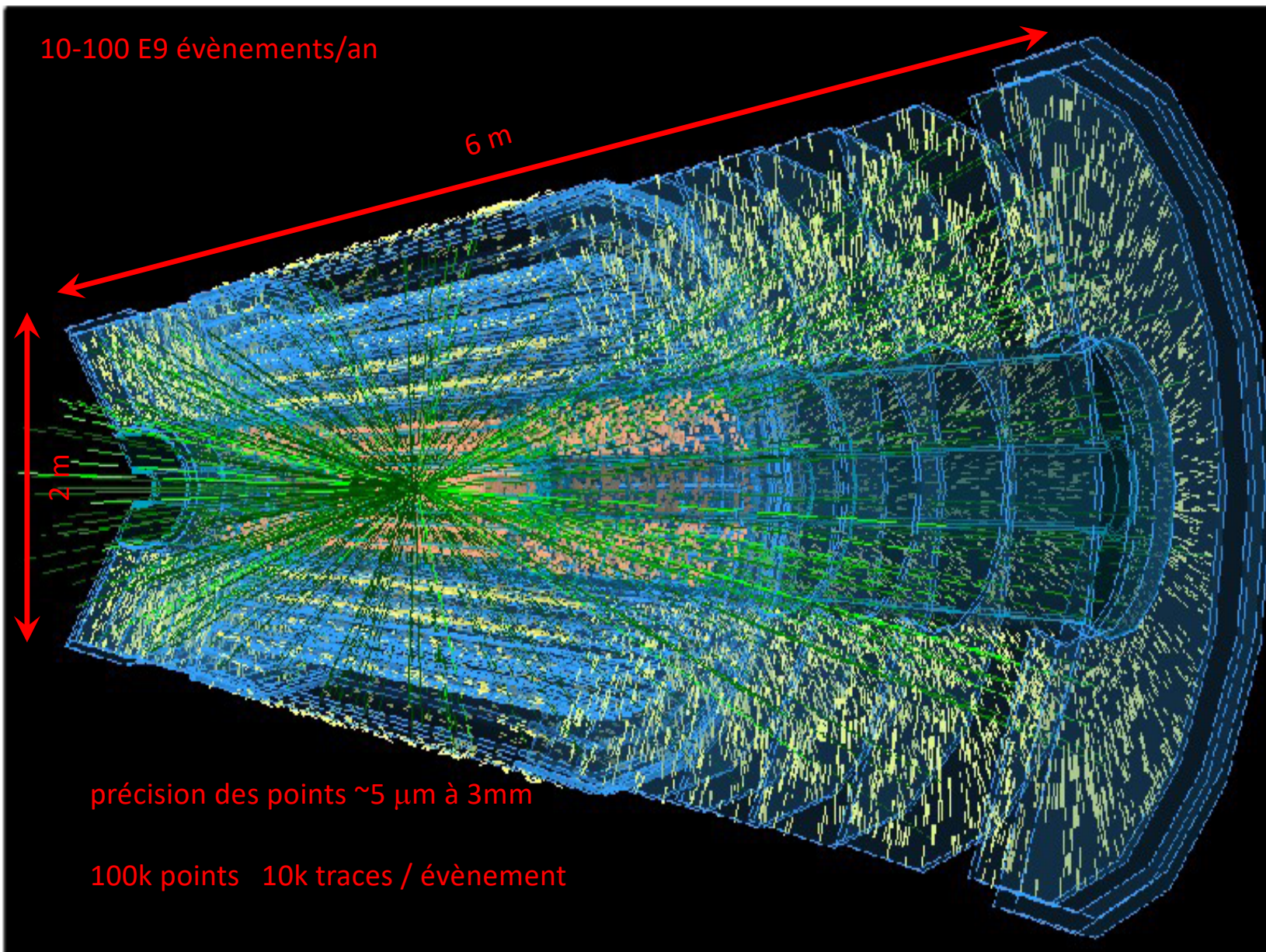
10-100 E9 évènements/an

6 m

2 m

précision des points $\sim 5 \mu\text{m}$ à 3mm

100k points 10k traces / évènement



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%

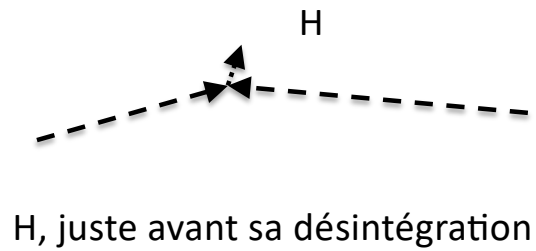
$$E=mc^2$$



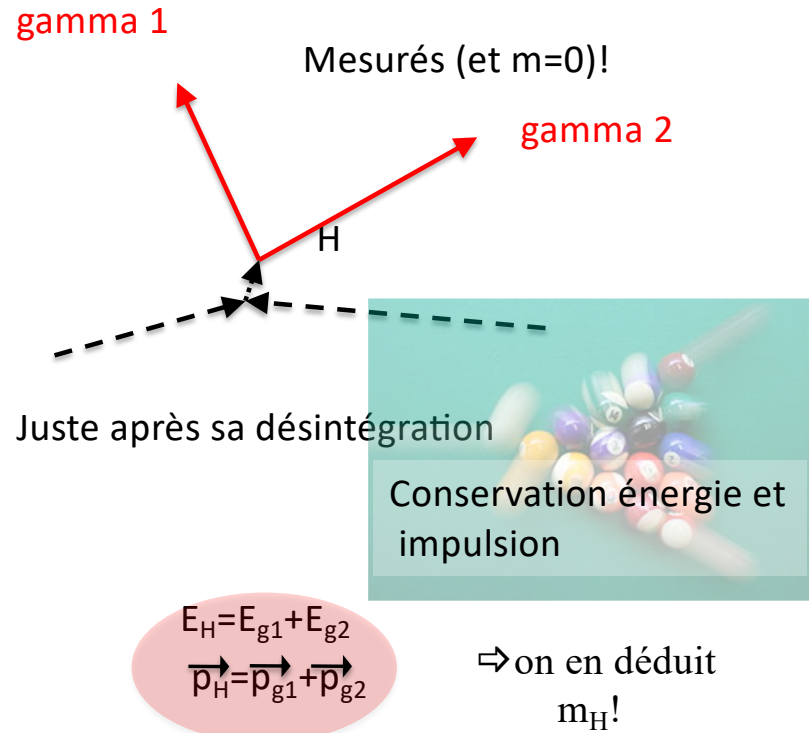
Einstein en 1905

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:

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



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



Finalemment...

10^{14} collisions



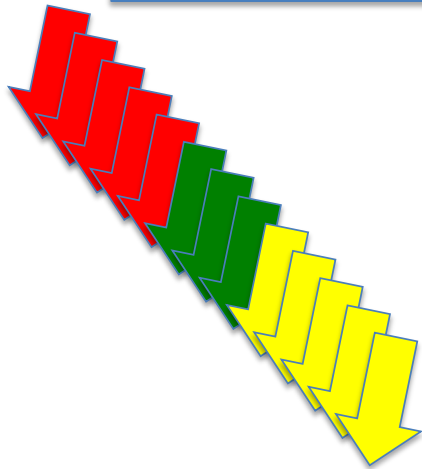
Tri rapide et grossier

10^9 événements sur disque

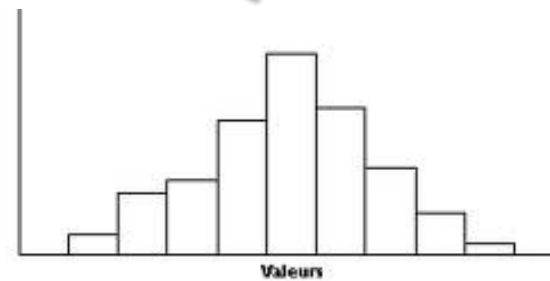


Tri précis

10^5 événements à 2 gamma

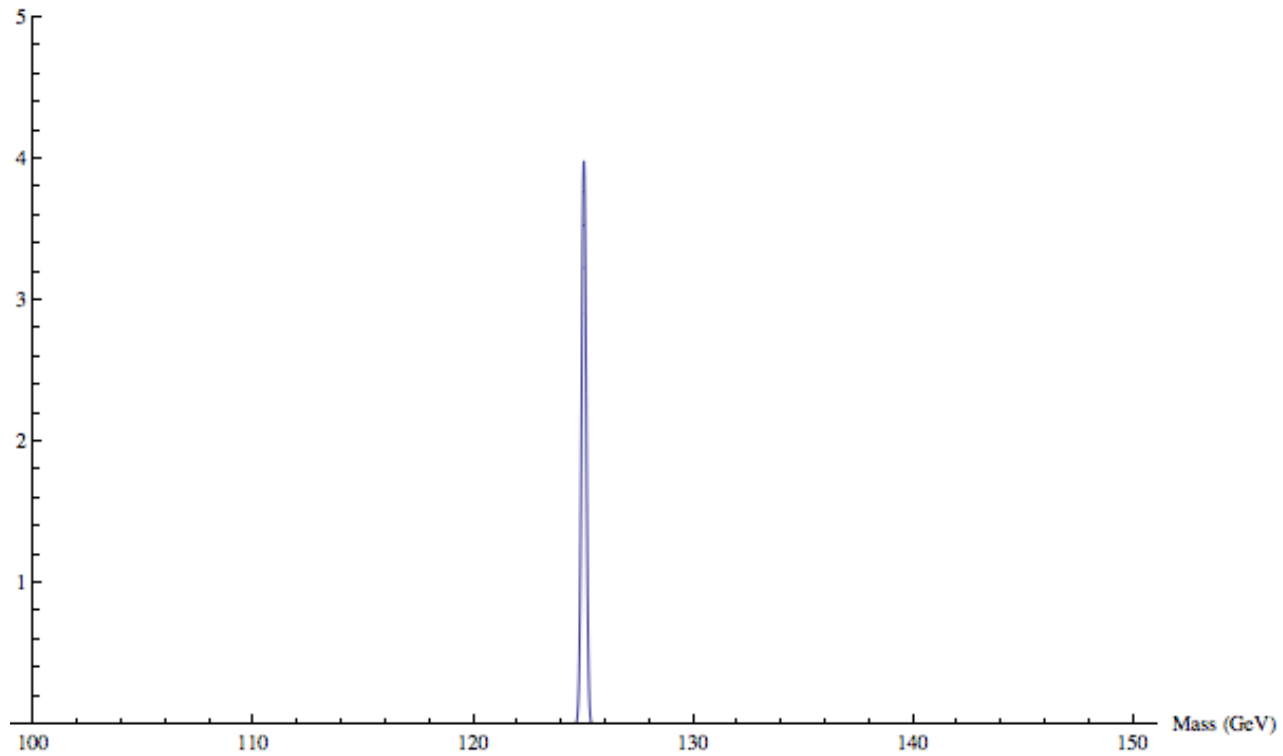


Calcul de la masse
→ histogramme



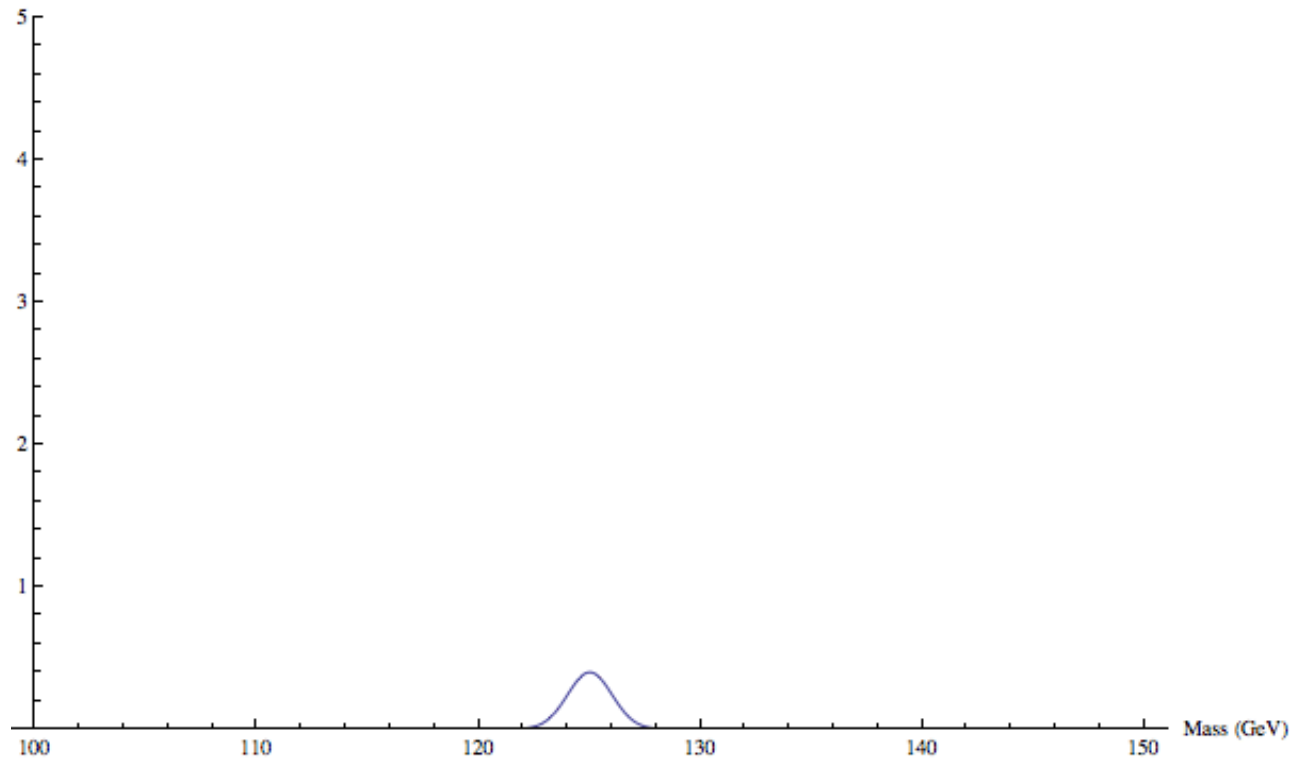
Effet de la précision du détecteur

Cliquer pour animer



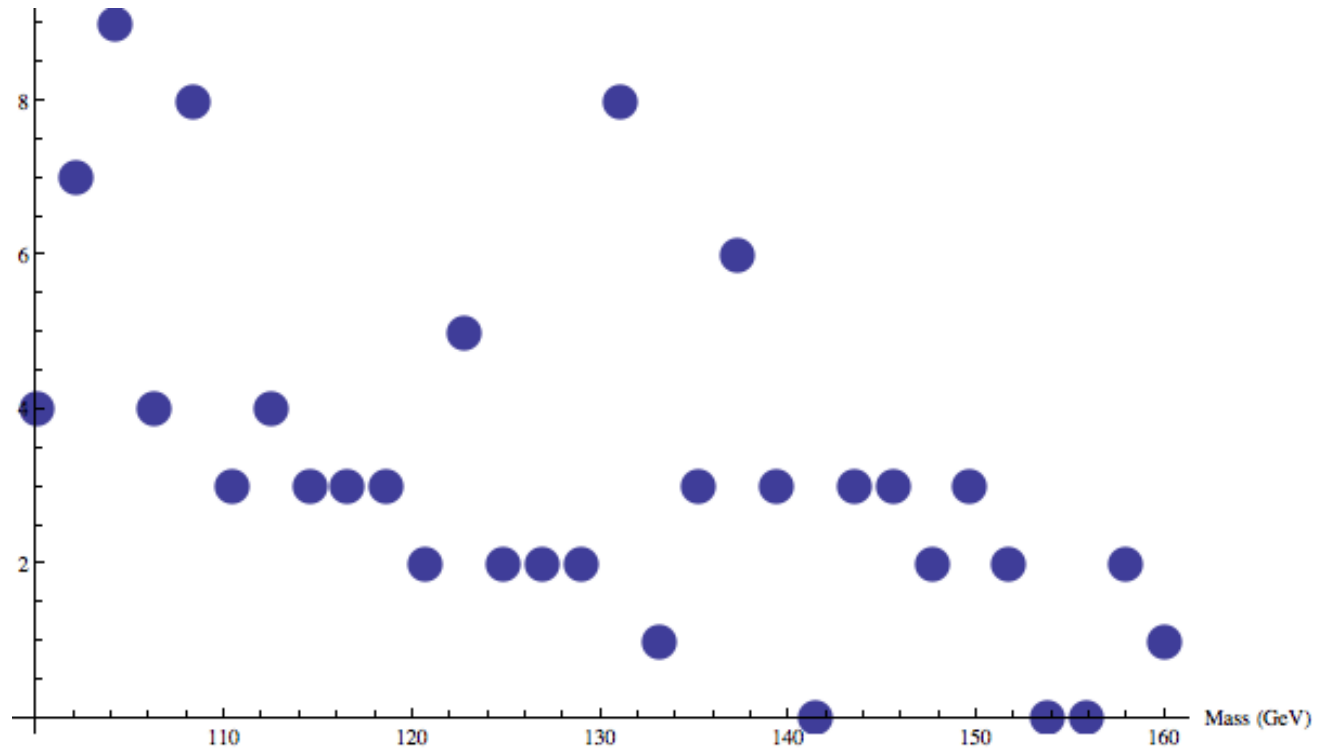
Effet du bruit de fond

Cliquer pour animer

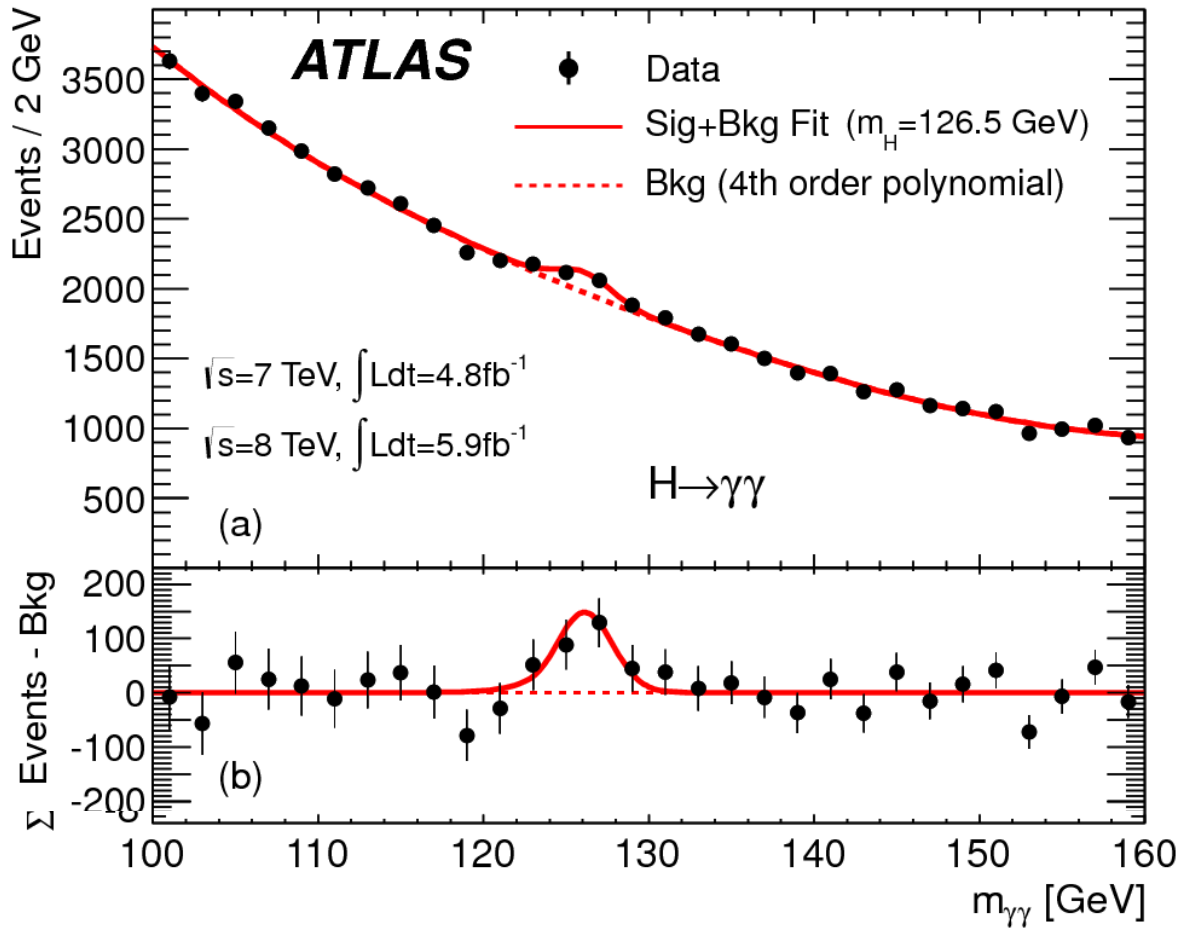


Effet de la statistique

Cliquer pour animer



Et maintenant « en vrai » (Juillet 2012)



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.



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. **PAGES 3-5**

Les derniers feux des pharaons
Au musée Jacquemart-André, à Paris, une exposition passionnante s'attaque sur la période sombre de l'Égypte ancienne, ses rois oubliés.

Suicides chez France Télécom: l'ancien patron mis en examen
Olivier Lombard, qui dirigeait l'opérateur téléphonique lors de la vague de suicides ayant touché l'entreprise en 2009, est visé par une enquête de la justice pour harcèlement moral.

A nos lecteurs
En raison d'un mouvement de grève dans les imprimeries, certains de nos journaux n'ont pu être distribués que sous forme numérique. Désolé pour ces quelques jours de retard.



MOST POPULAR U.S. Edition ▼

The New York Times

Wednesday, July 4, 2012 Last Update: 4:00 AM ET

DIGITAL SUBSCRIPTION: 4 WEEKS FOR 99¢

Search

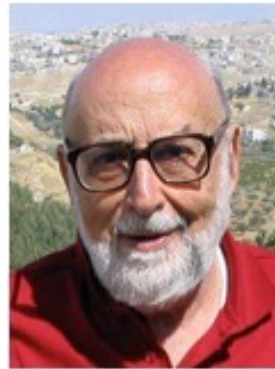
OPINION »
EDITORIAL
Too Quiet, A Health Care
The Obama ca
forcefully cou
Republican m
the reform law

MARKETS »
Britain Ge

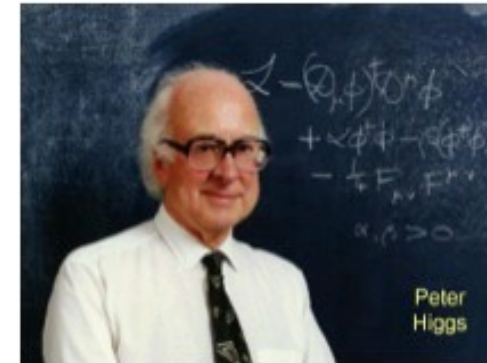




Robert Brout 1928-2011



François Englert 1932-



Peter Higgs 1929-

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



Champ de Higgs

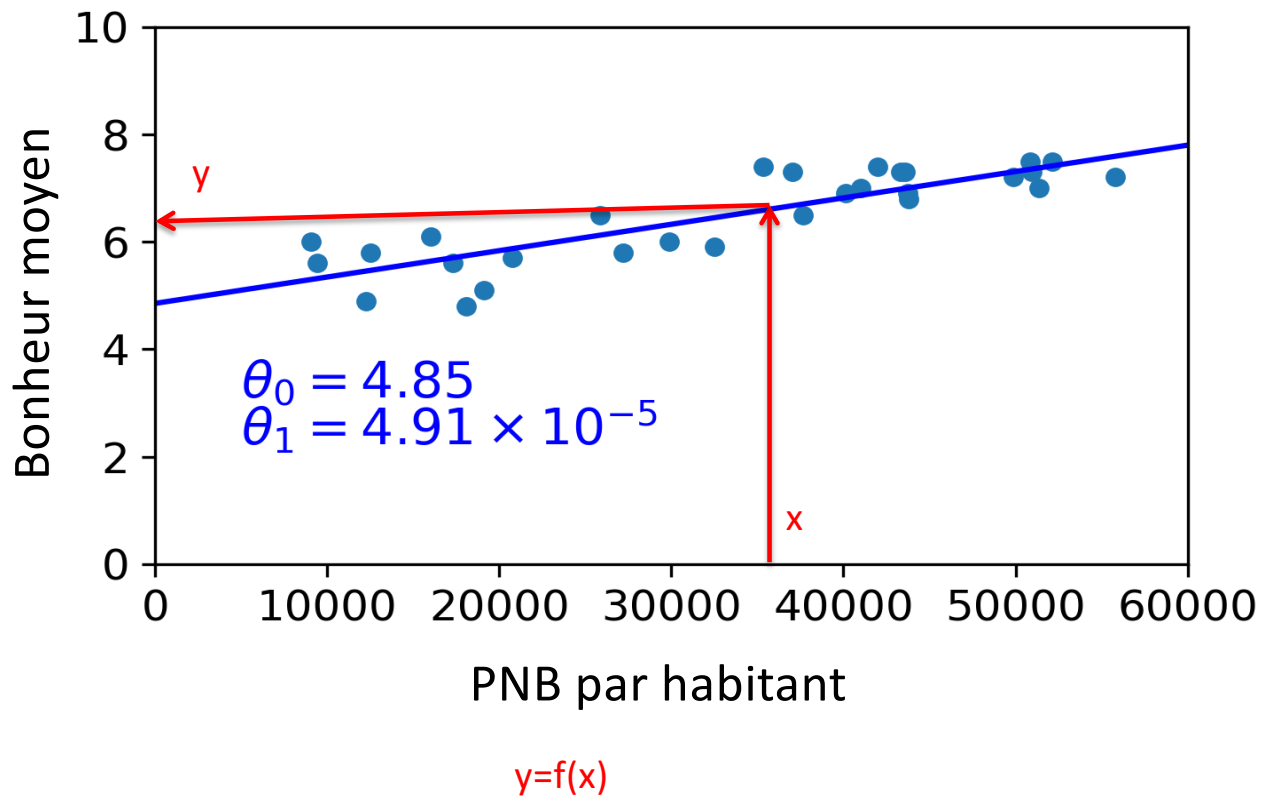
Boson de Higgs

« 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 »

Intelligence Artificielle

Régression Linéaire

Boskovic, Legendre, Laplace, Gauss



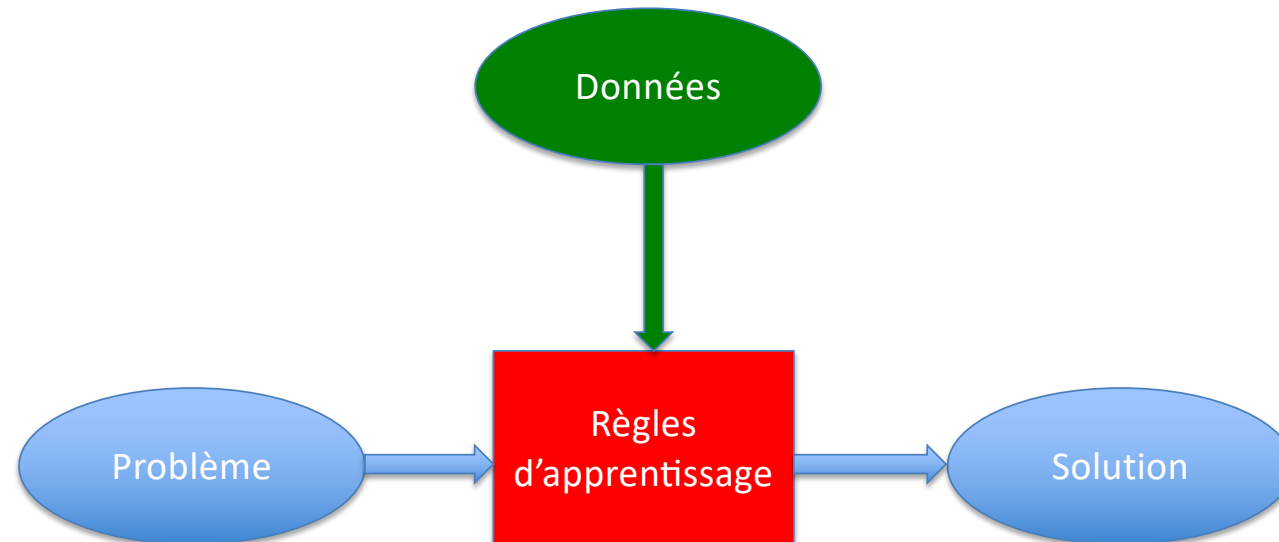
Etant donné x , on veut y \rightarrow comment construire f ?

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

Informatique traditionnelle

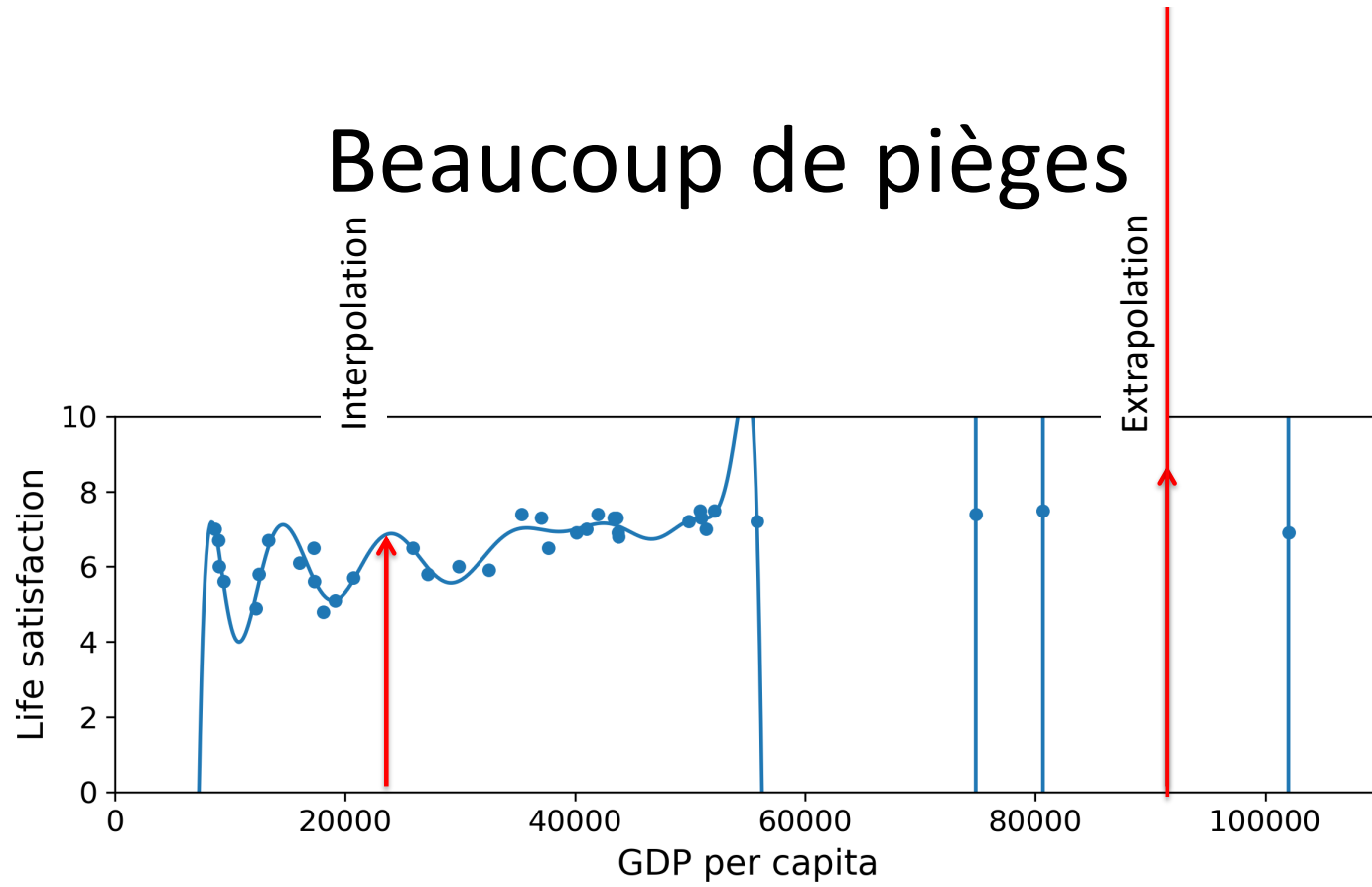


Apprentissage Automatique

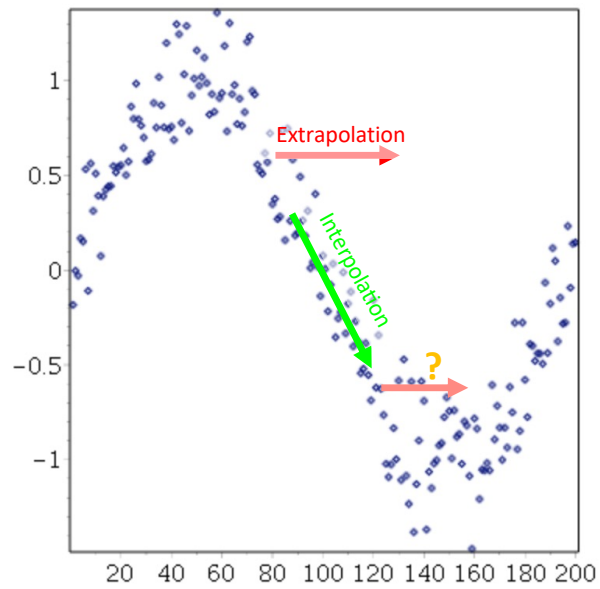


Apprendre = optimiser les paramètres internes de l'algorithme : $n=2$ à 10^{12}

Beaucoup de pièges



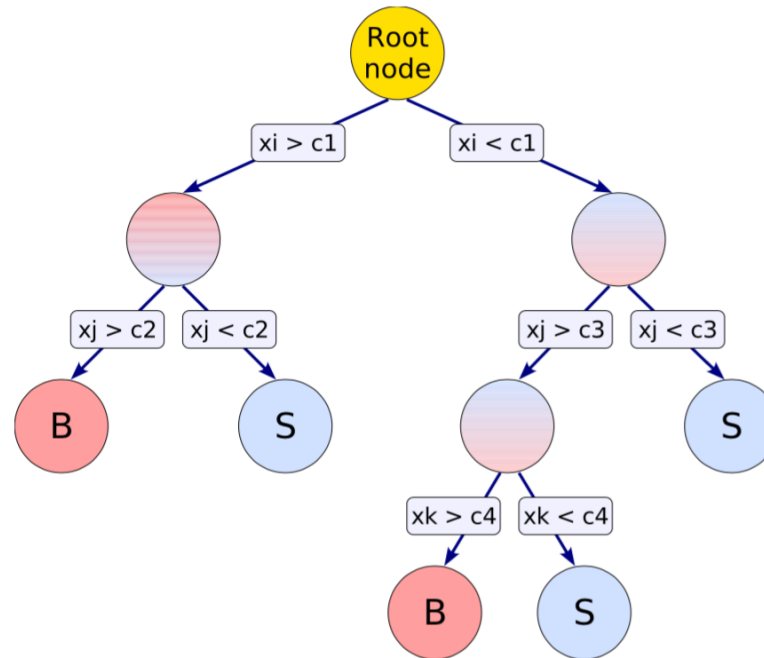
Interpolation vs Extrapolation



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

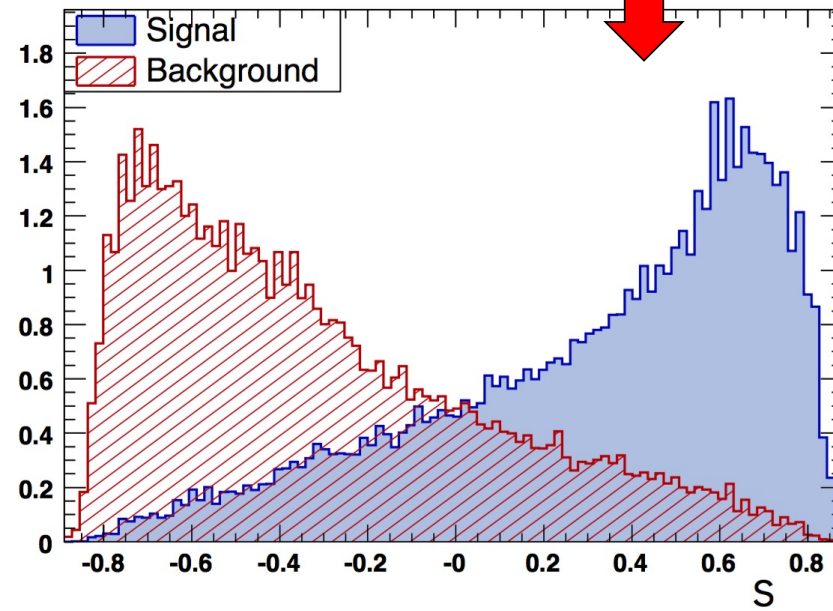
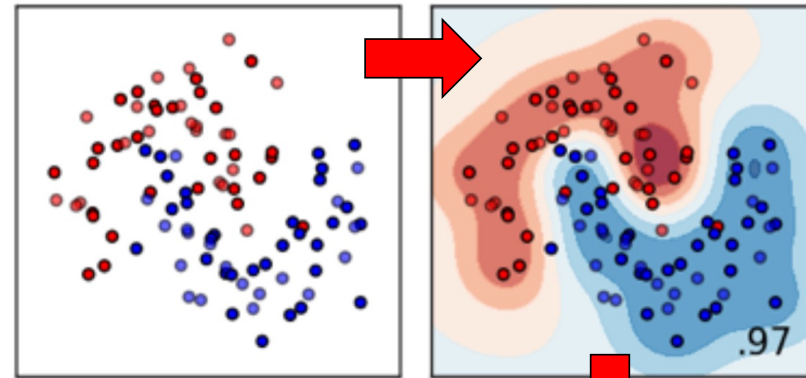
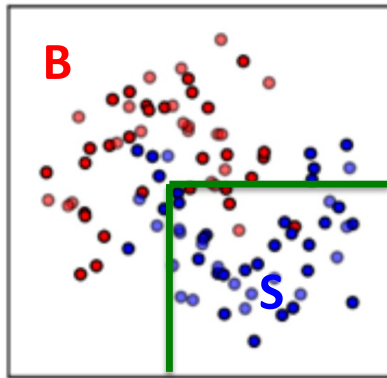
Comment ça marche ?

Arbre de Décision Boosté (Boosted Decision Tree)

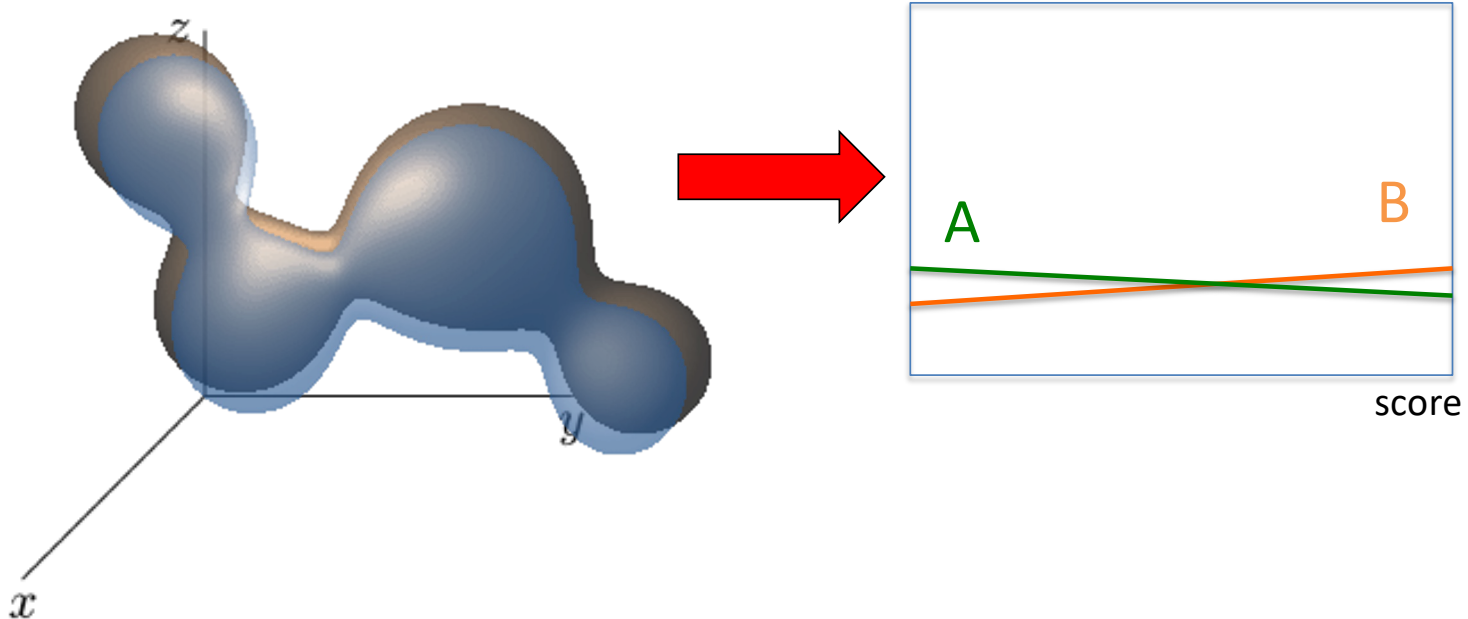


- Single tree (CART) <1980
- AdaBoost 1997 : rerun increasing the weight of misclassified entries
→ Boosted Decision Trees (**Gradient BDT XGBoost**, random forest...)

Classifier



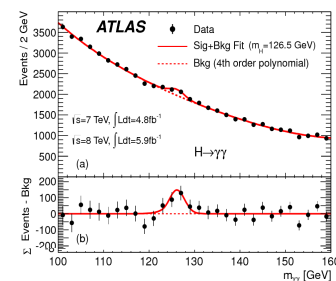
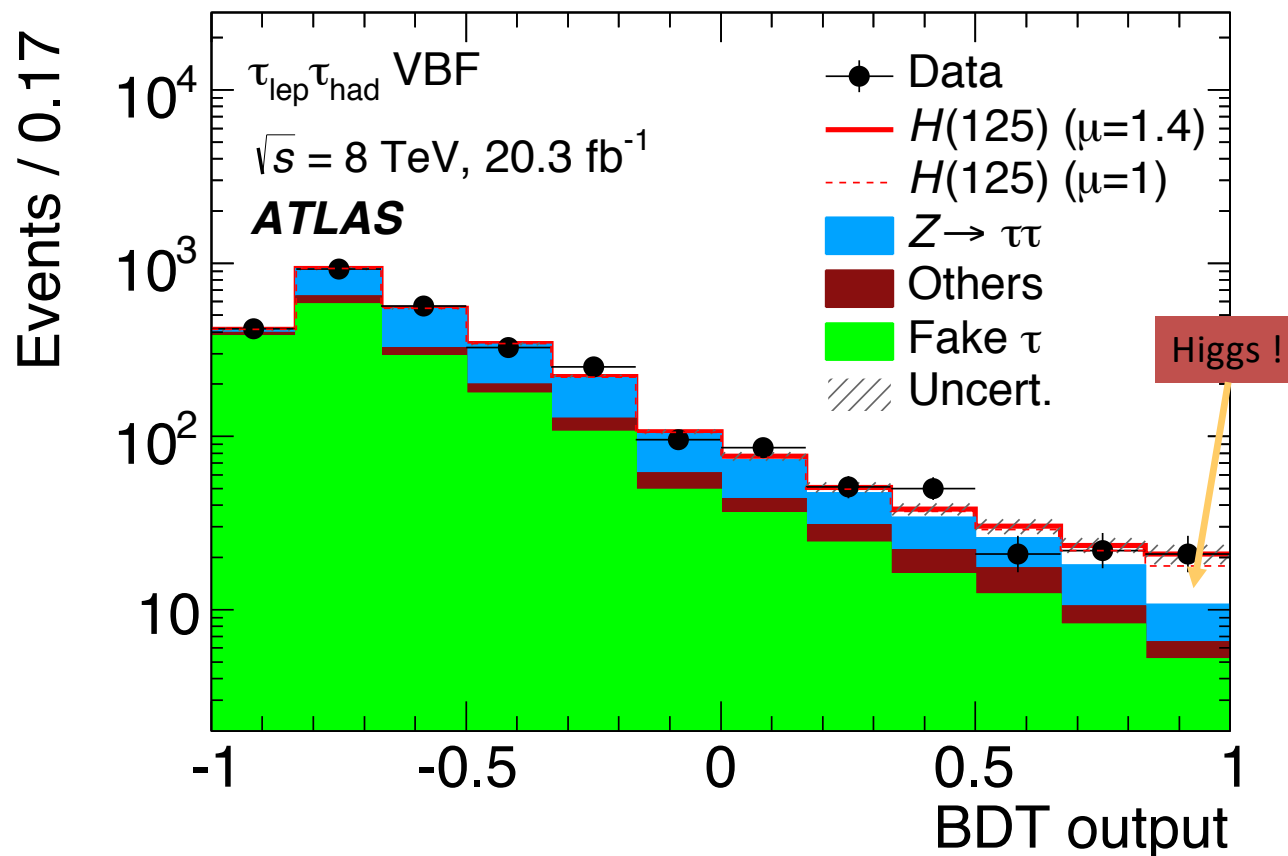
Qu'est ce qu'un classifieur fait?



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

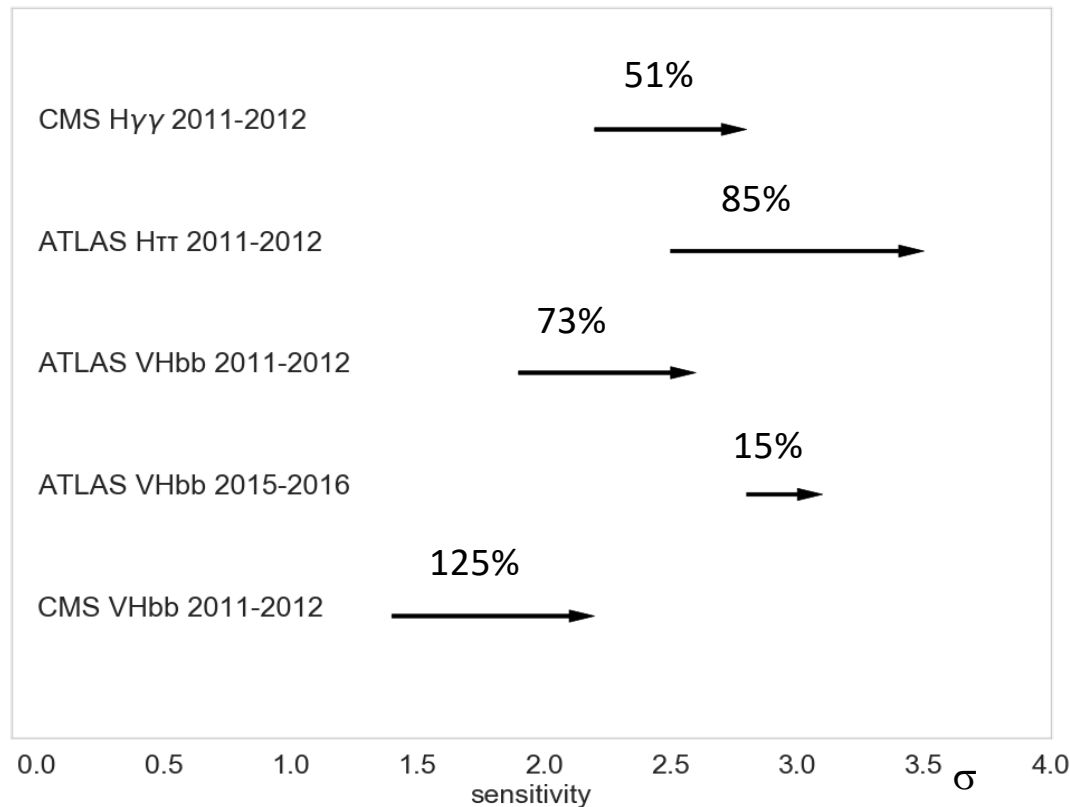
Application $H \rightarrow \tau^+ \tau^-$

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



- 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:

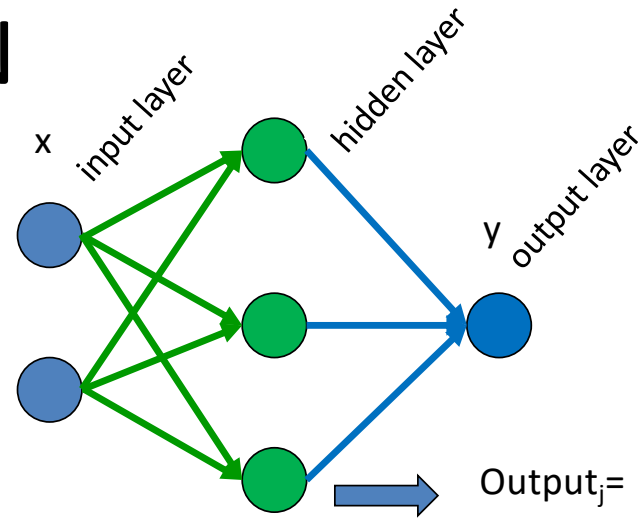
D'après *Nature* **560**, 41–48 (2018)



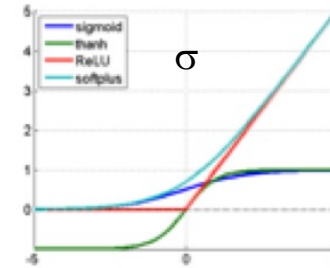
→ équivalent à ~50% de données en plus
 (le LHC a coûté 4 milliards d'euros, budget du cern 1 milliard CHF par an)

Neural Networks

Simplest NN



activation function



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

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

Beware: superscript are layer indices!

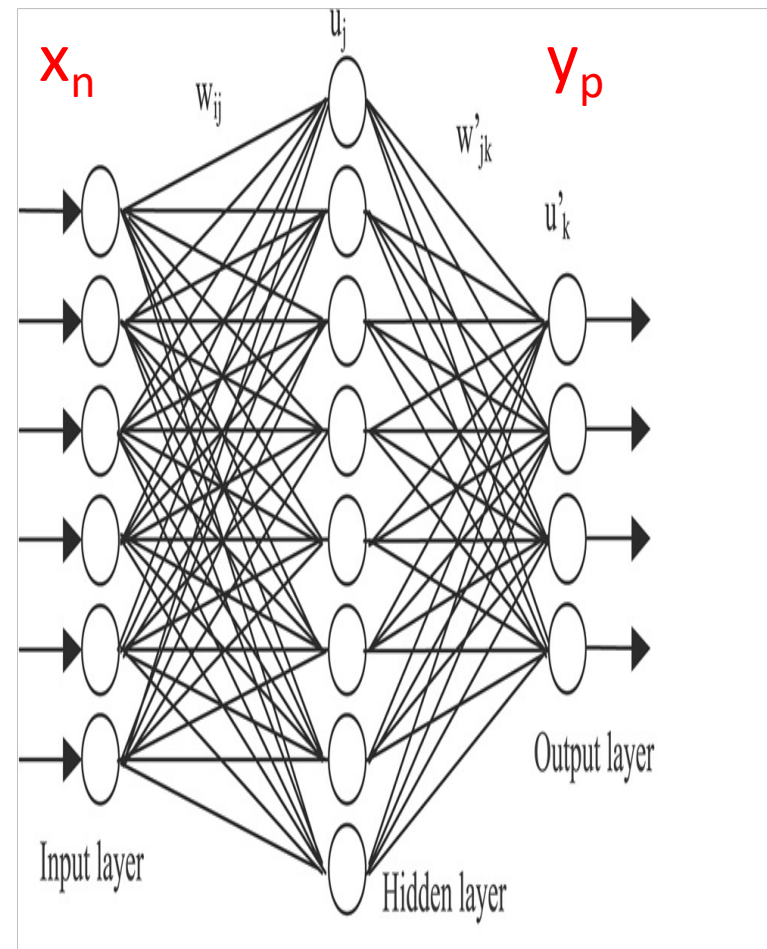
Now with dimensions

$$h(x_{(2)}) = \sigma(b_{(1)}^2 + W_{(1,3)}^2 \sigma(b_{(3)}^1 + W_{(3,2)}^1 x_{(2)}))$$

Universal Approximation theorem

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

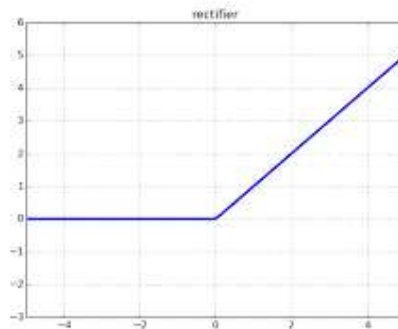
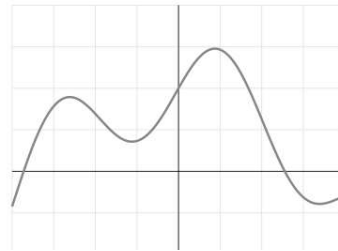
[Addendum ResNet 1 neuron sufficient depth](#)



Universal Theorem at work

Universal approximation

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



$\text{relu}(x) = x$ if $x > 0$ & 0 otherwise

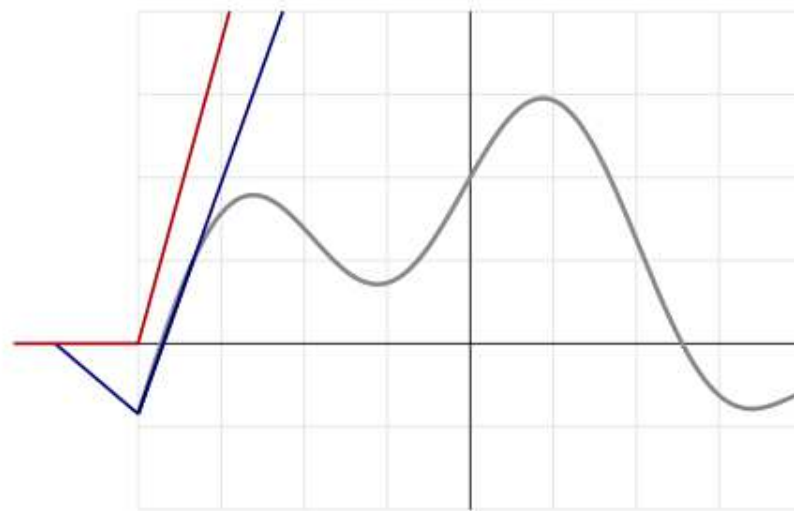
Universal approximation

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



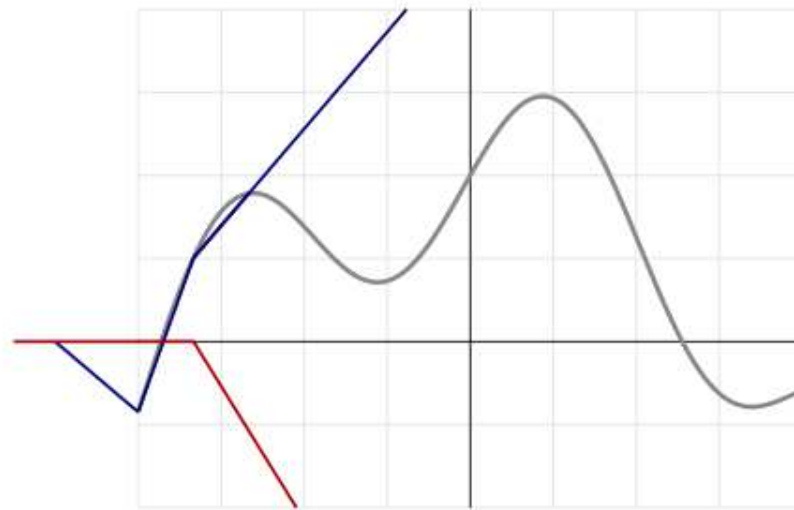
Universal approximation

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



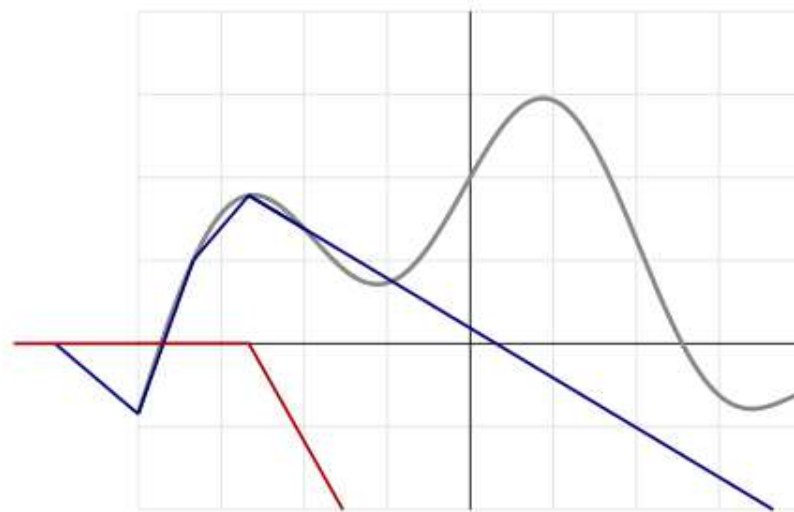
Universal approximation

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



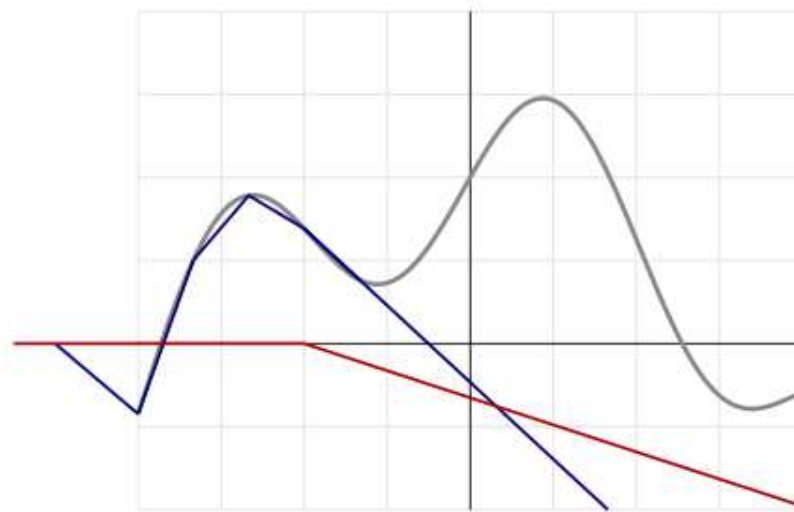
Universal approximation

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



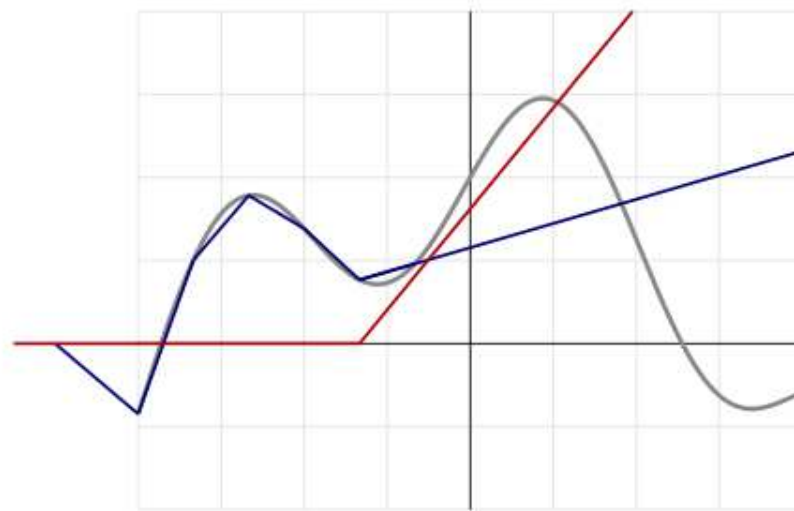
Universal approximation

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



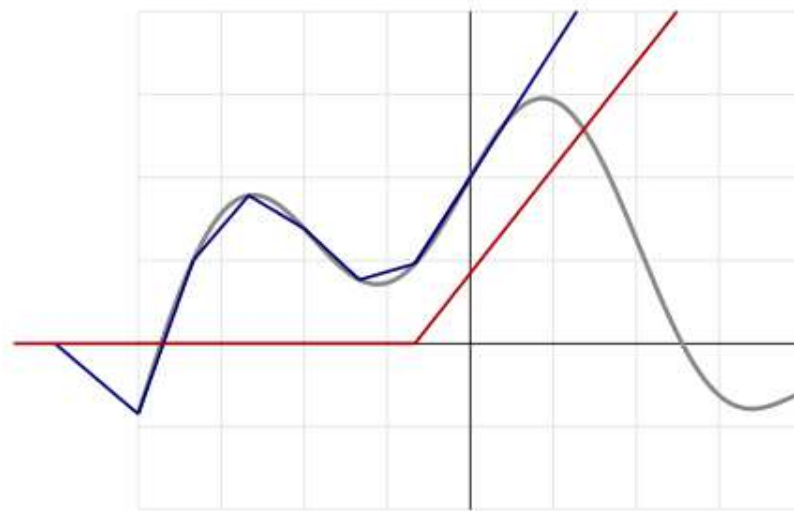
Universal approximation

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



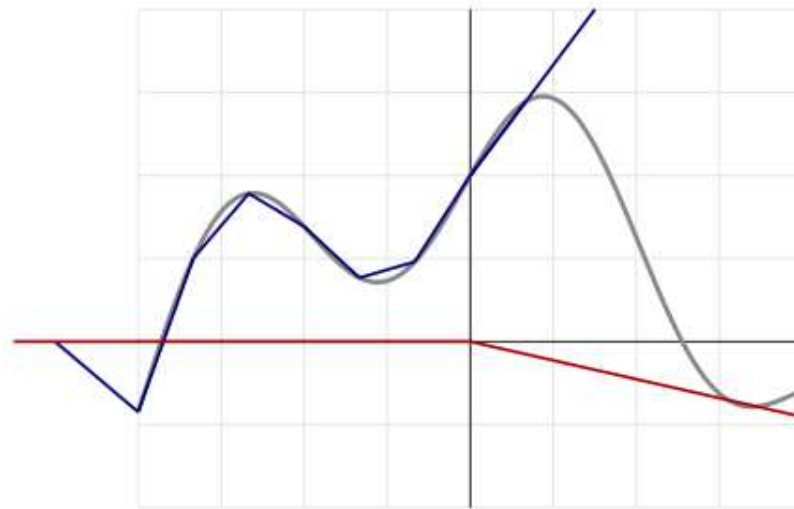
Universal approximation

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



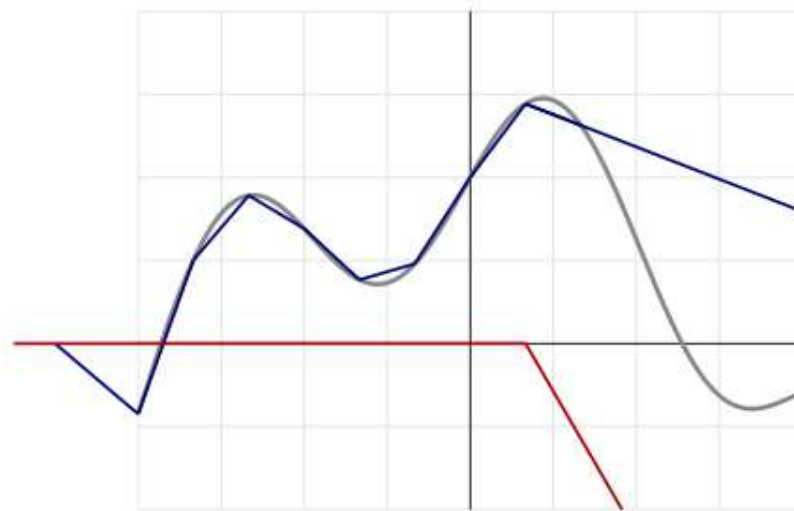
Universal approximation

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



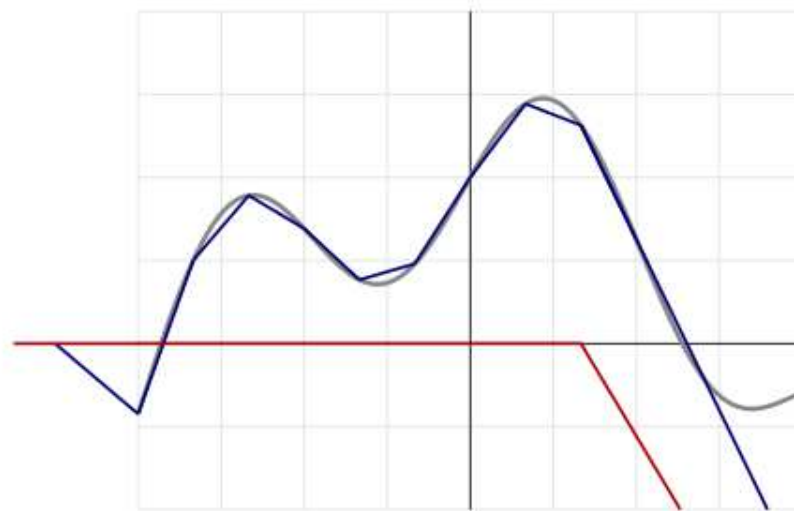
Universal approximation

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



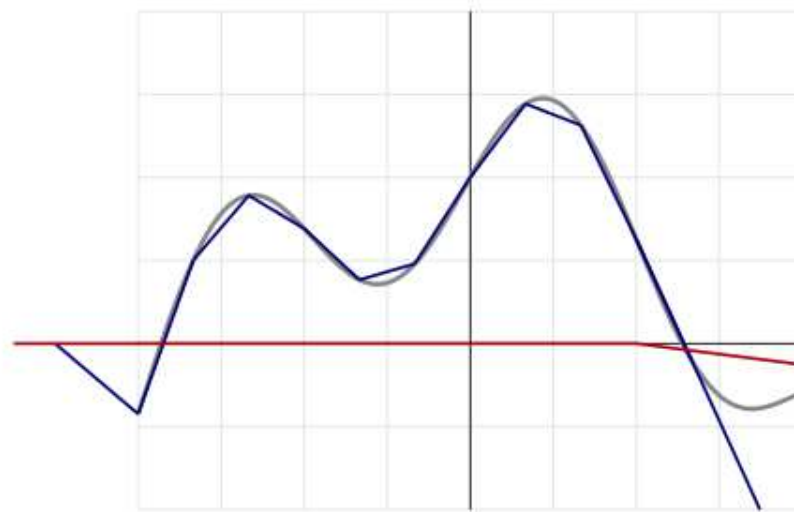
Universal approximation

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



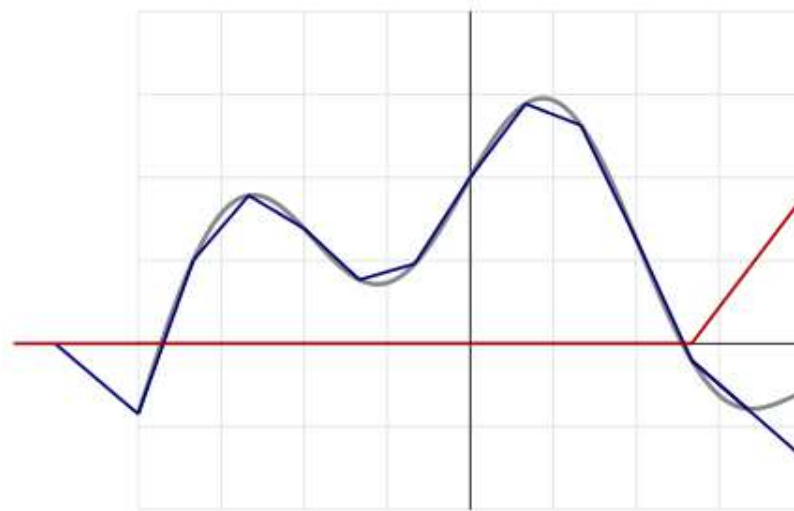
Universal approximation

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



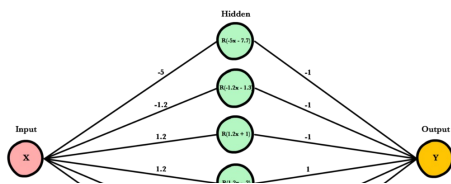
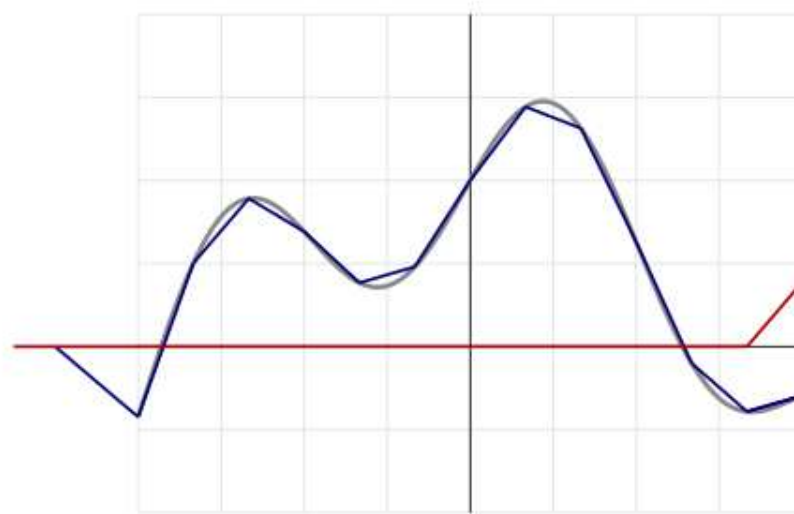
Universal approximation

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



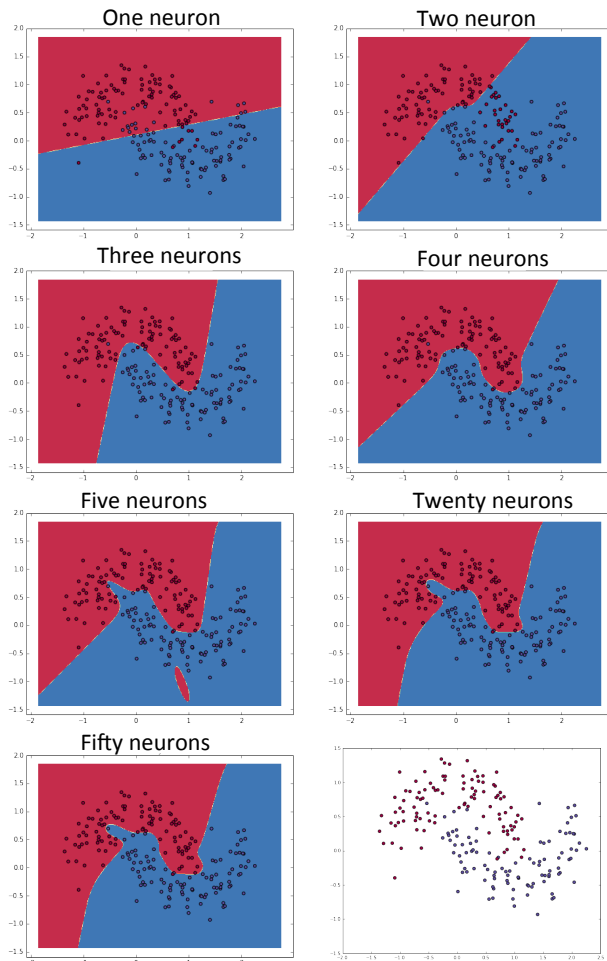
Universal approximation

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



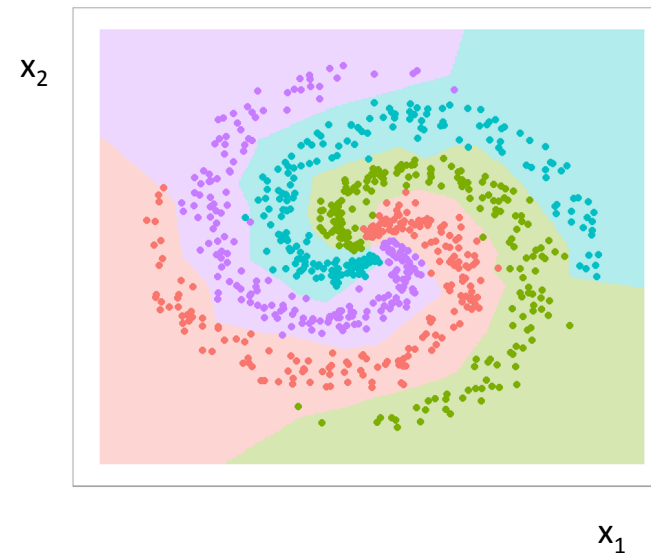
$$y = \sum_i \text{Relu}(a_i \times x + b_i)$$

NN at work



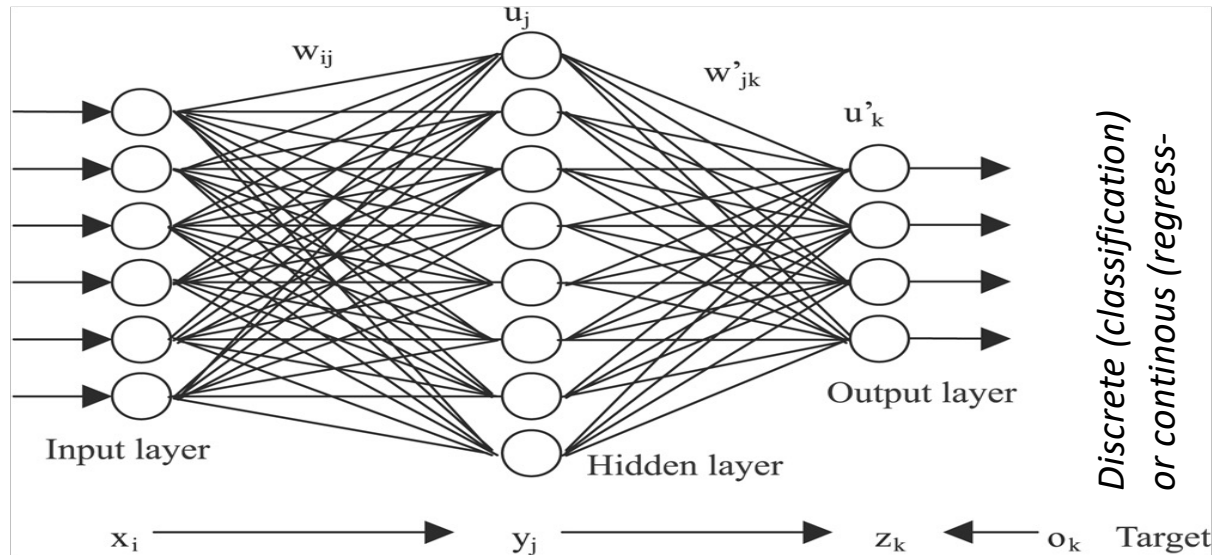
<http://www.wildml.com/2015/09/implementing-a-neural-network-from-scratch/>

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



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

Neural Net in a nutshell

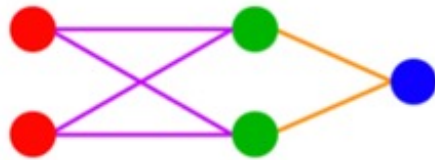


- Neural Net ~1950!
- But many many new tricks for learning, in particular if many layers (also ReLU instead of sigmoid activation)
- “Deep Neural Net” hundreds layers
- Computing power (DNN training can take days even on GPU)

Optimisation

Note : <https://www.college-de-france.fr/site/stephane-mallat/course-2019-03-20-09h30.htm>

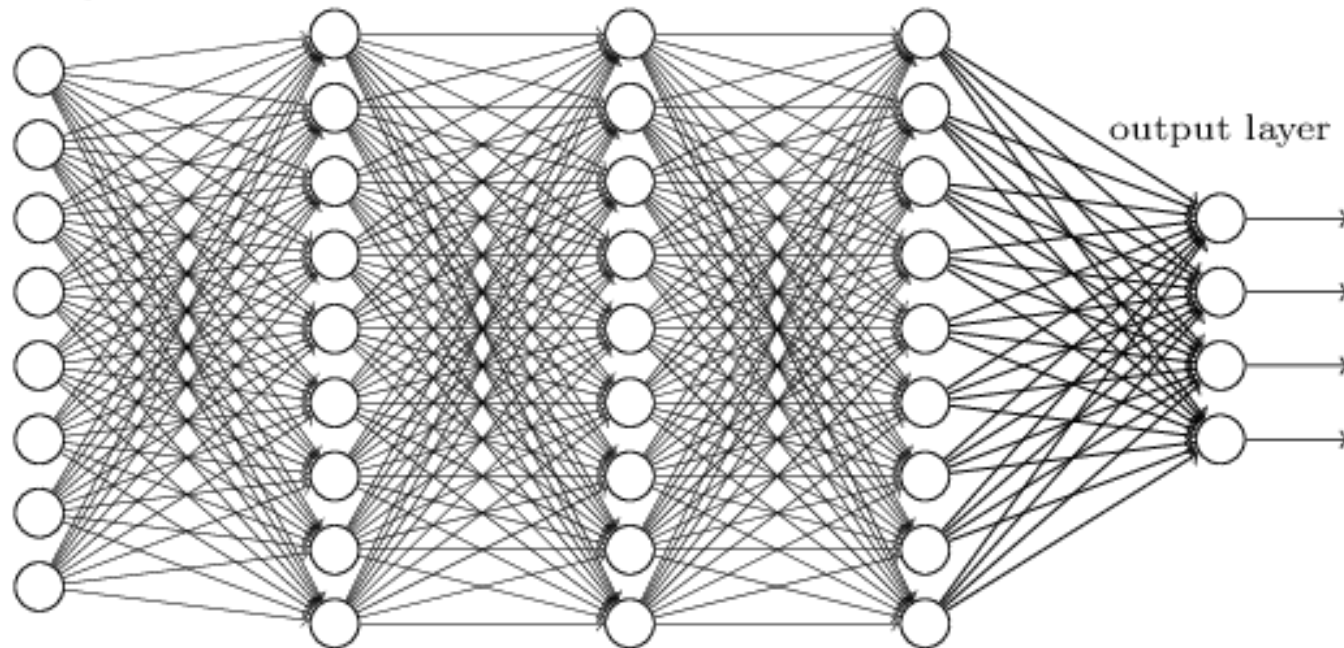
Neural Network Optimization



$$f_{NN} = \sigma(b_2 + W_2 \sigma(b_1 + W_1 x))$$

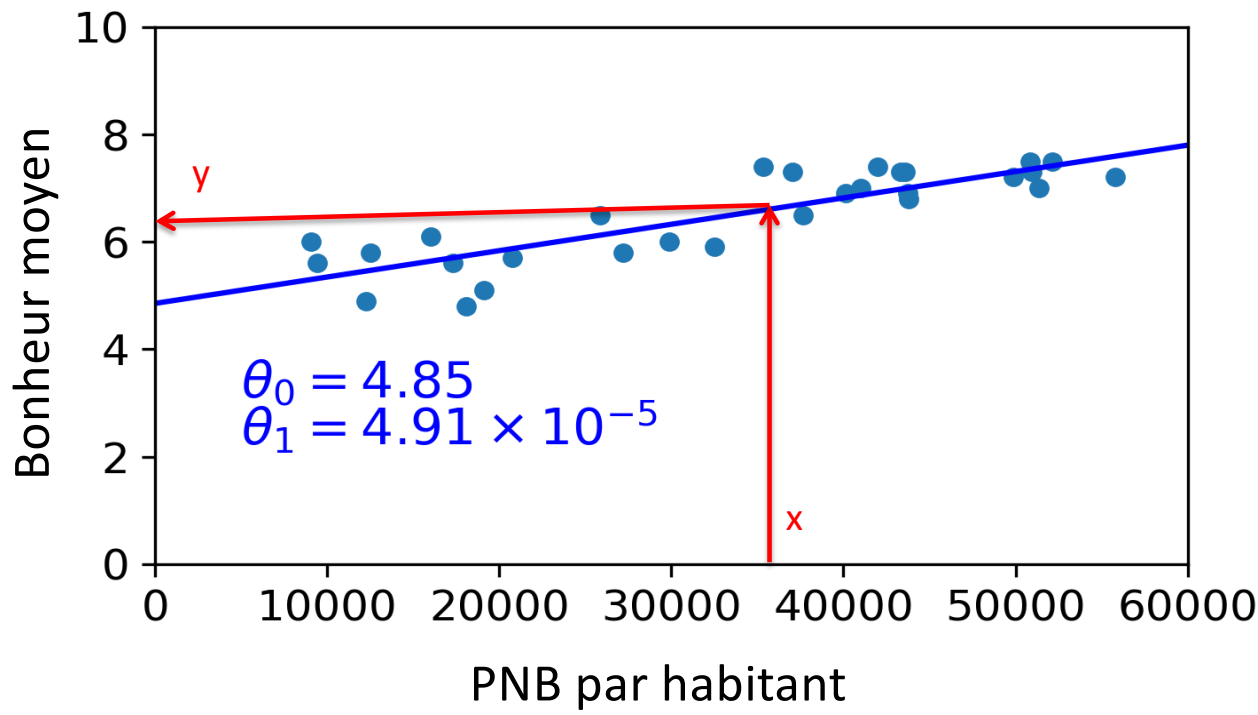
input layer

layer 1 hidden layer 2 hidden layer 3



Régression Linéaire

Boskovic, Legendre, Laplace, Gauss



$y=f(x)$

Neural Network loss function

- Neural Network Model: $h(\mathbf{x}) = \mathbf{w}^T \sigma(\mathbf{U}\mathbf{x})$

- **Classification:** Cross-entropy loss function

$$p_i = p(y_i = 1 | \mathbf{x}_i) = \sigma(h(\mathbf{x}_i))$$

$$L(\mathbf{w}, \mathbf{U}) = - \sum_i y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)$$

- **Regression:** Square error loss function

$$L(\mathbf{w}, \mathbf{U}) = \frac{1}{2} \sum_i (y_i - h(\mathbf{x}_i))^2$$

- Minimize loss with respect to weights \mathbf{w} , \mathbf{U}

BB

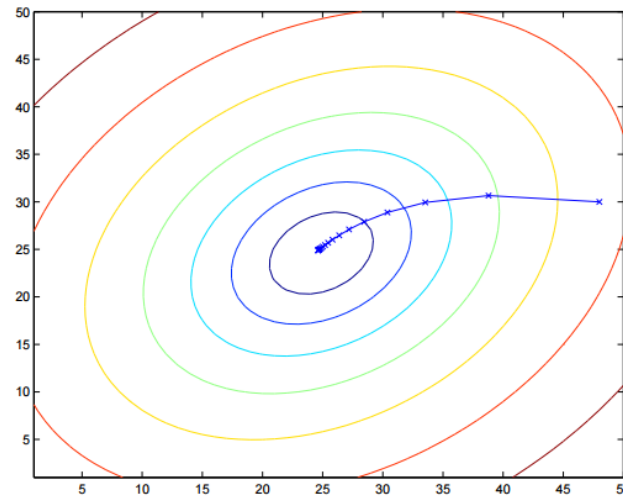
Gradient descent

- Minimize loss by repeated gradient steps

- Compute gradient w.r.t. parameters: $\frac{\partial L(\mathbf{w})}{\partial \mathbf{w}}$

- Update parameters: $\mathbf{w}' \leftarrow \mathbf{w} - \eta \frac{\partial L(\mathbf{w})}{\partial \mathbf{w}}$

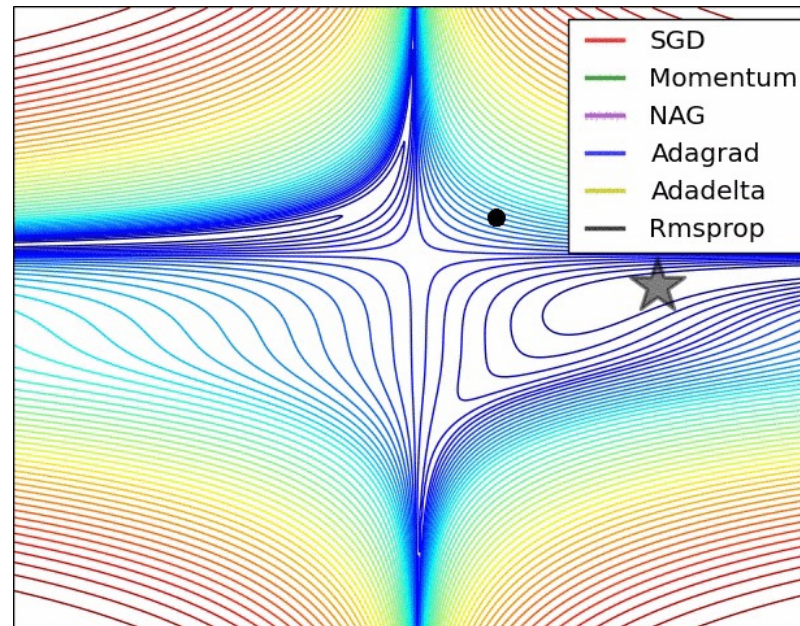
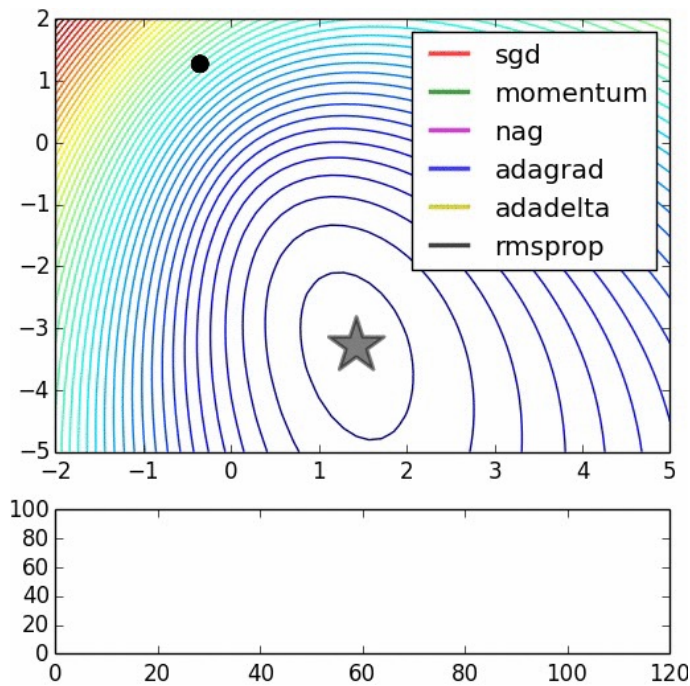
Computing Hessian not practical!



Optimisation

- Up to 10^{12} of parameters to optimise....
- Wealth of newish algorithms in particular Stochastic Gradient Descent (SGD) and more

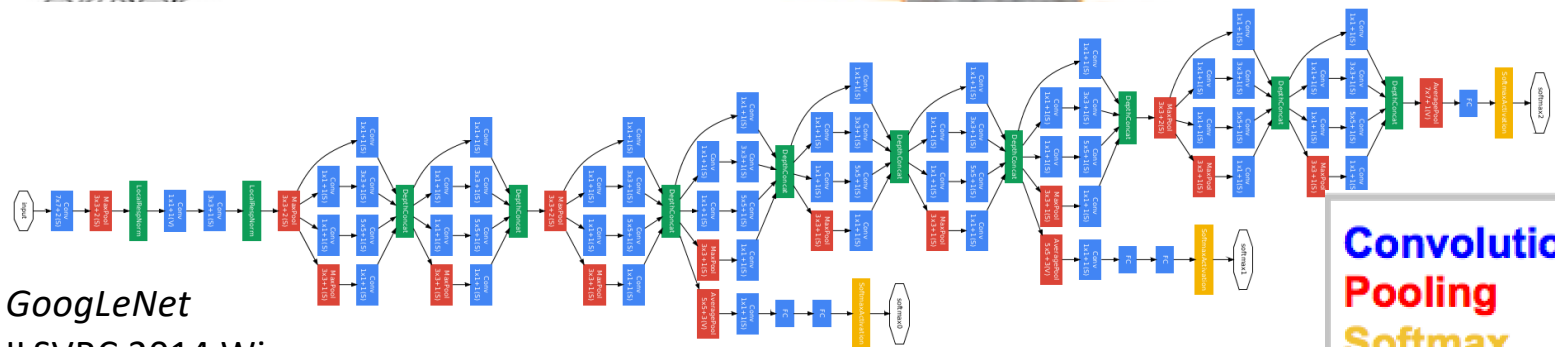
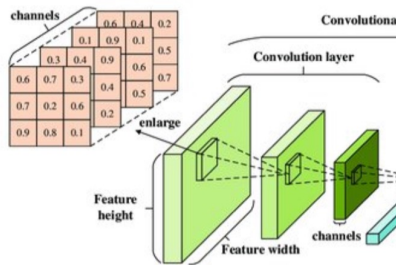
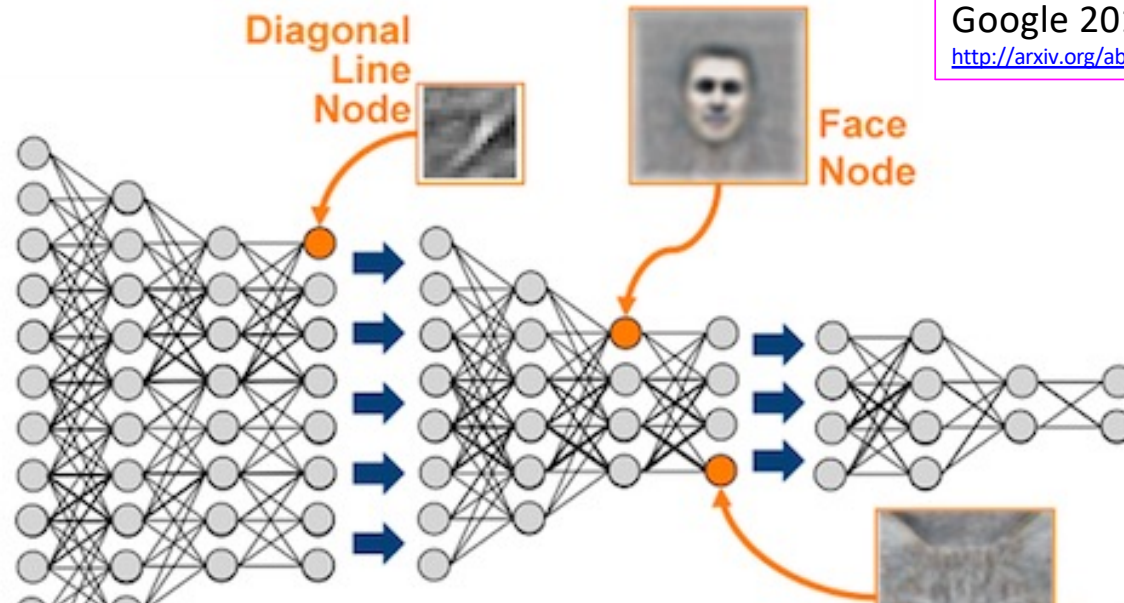
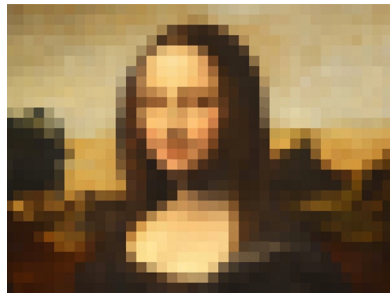
[Alec Radford](#)



Architectures spécialisées

Convolutional Neural Network

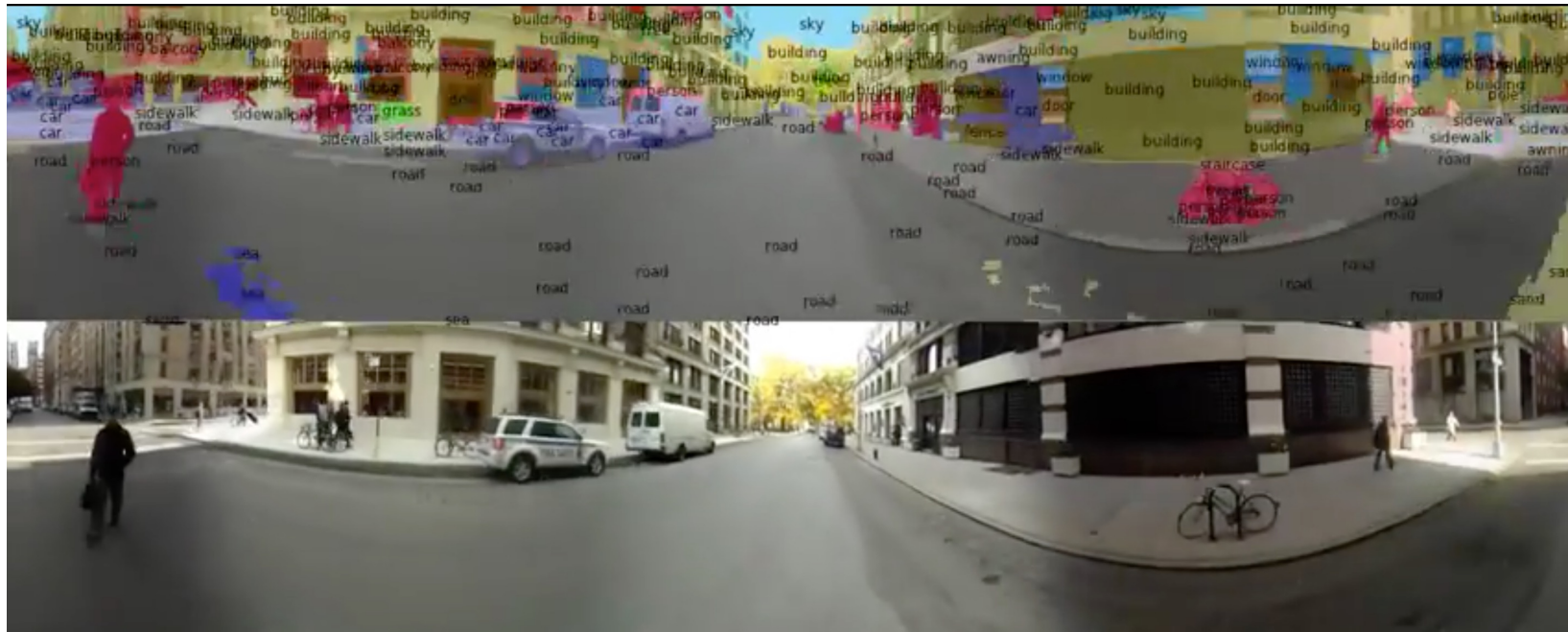
Google 2012
<http://arxiv.org/abs/1112.6209>



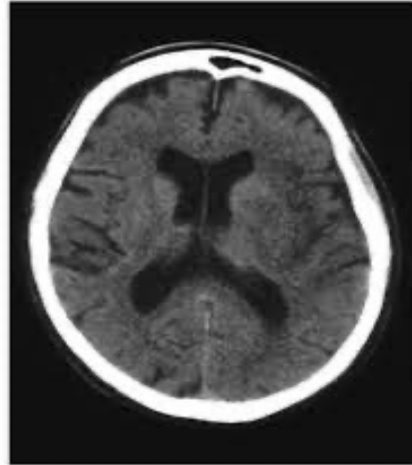
GoogLeNet
 ILSVRC 2014 Winner
 4M parameters

Convolution
Pooling
Softmax
Other

Typical Deep Learning application

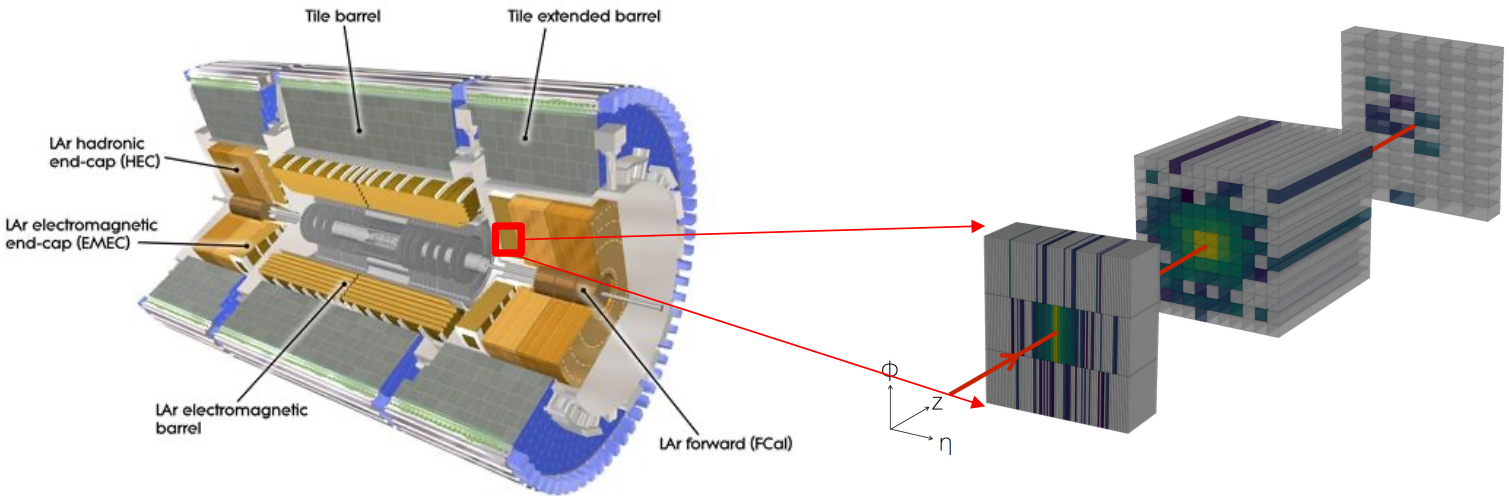


Applications innombrables



Les données scientifiques ne sont souvent
pas des images

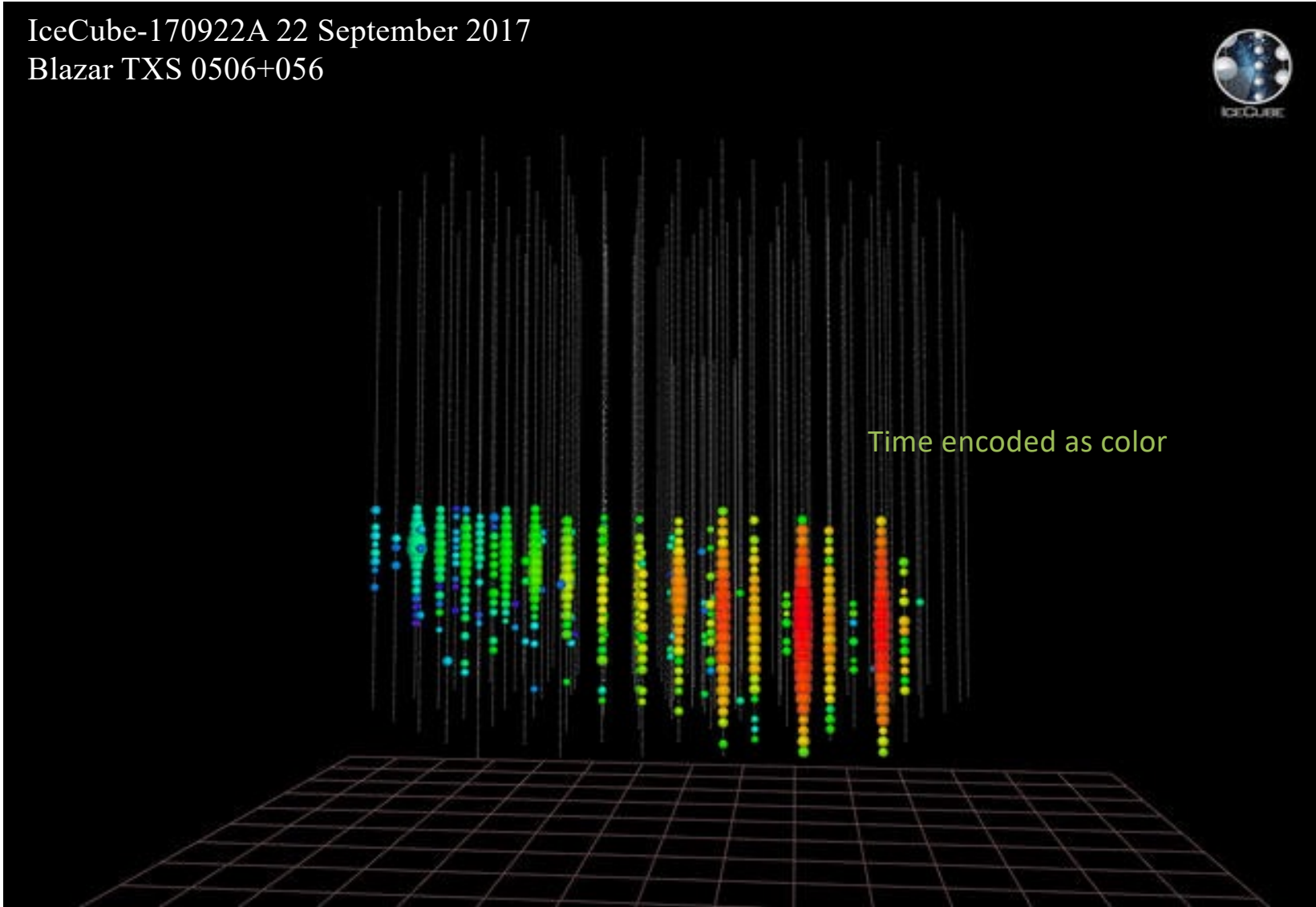
Les instruments dont des objets 3D complexes



IceCube-170922A 22 September 2017
Blazar TXS 0506+056

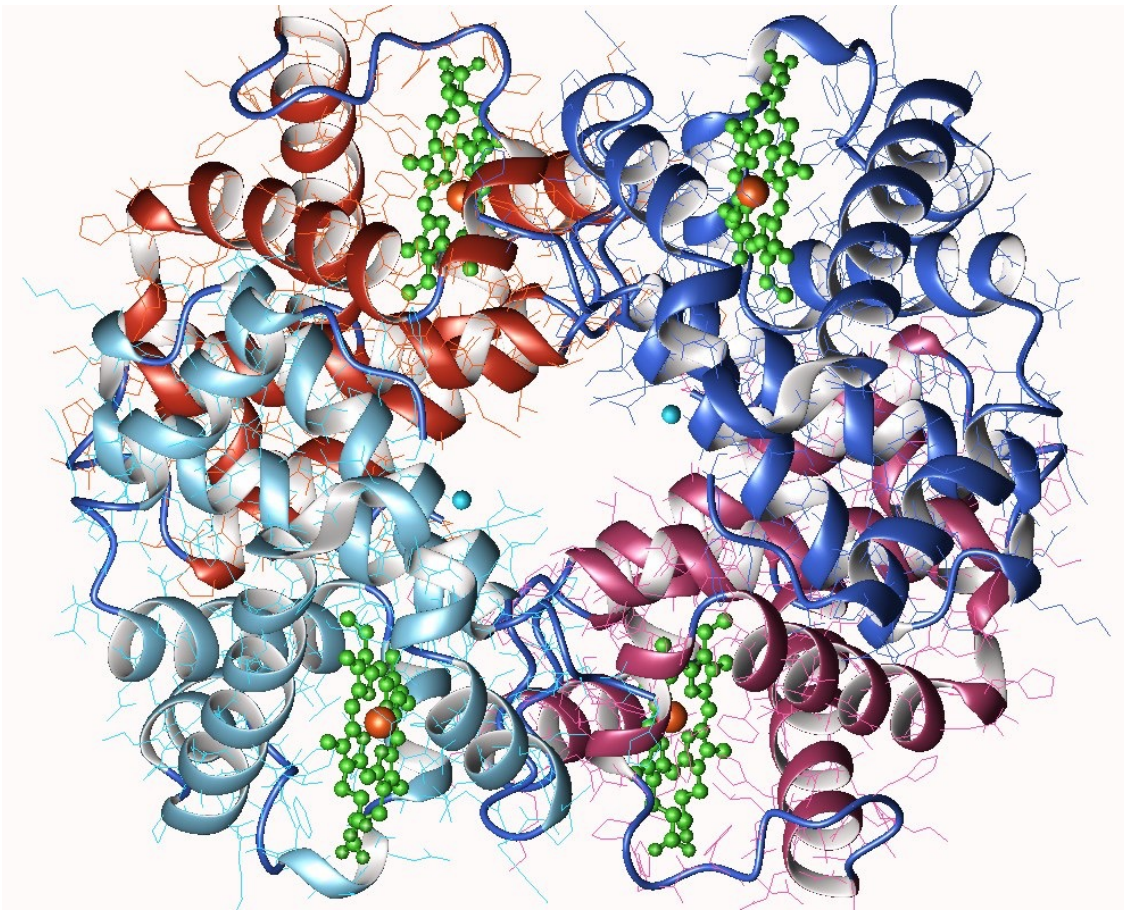


Time encoded as color



Une image, pas les données

Structure de
l'hémoglobine

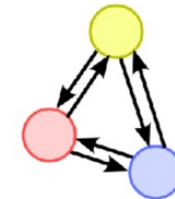
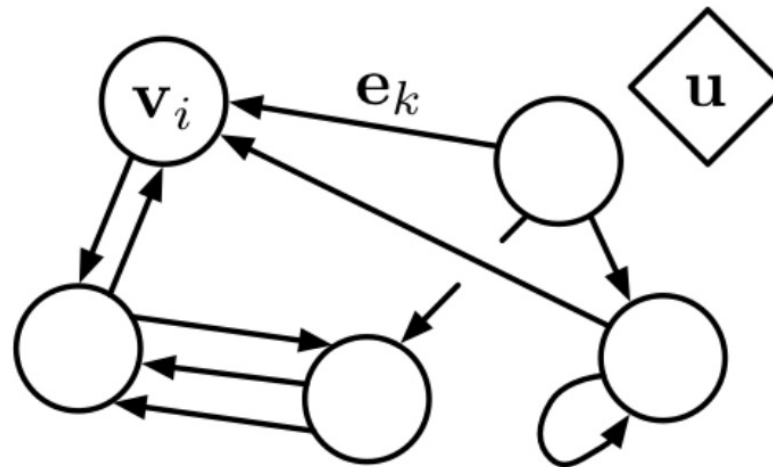
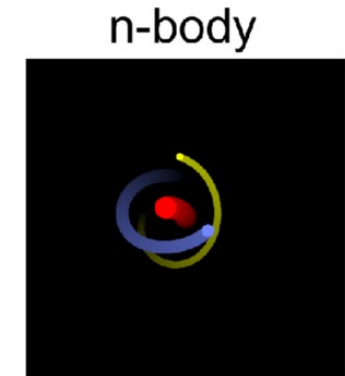
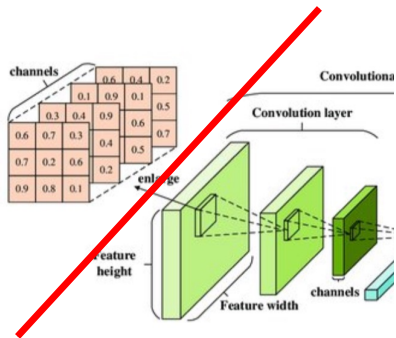


Par Deposition authors: Fermi, G., Perutz, M.F.;
visualization author: User:Astrojan —
<https://www.rcsb.org/structure/3hbb>,
CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=48714519>

Architectures spécialisées

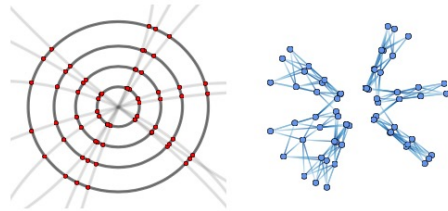
Graph Neural Network

- Structure définie
 - v_i : noeud
 - e_k : arête
 - u : global

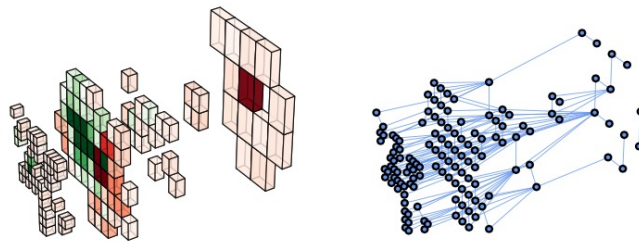


Nodes: bodies
Edges: gravitational forces
 Global : potential energy

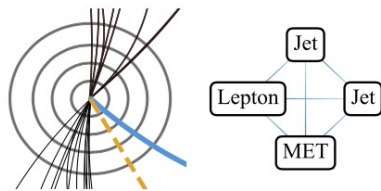
Graph on HEP data [from 2007.13681](#)



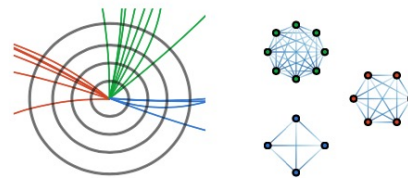
(a)



(b)



(c)



(d)