AI 4 physics

David Rousseau

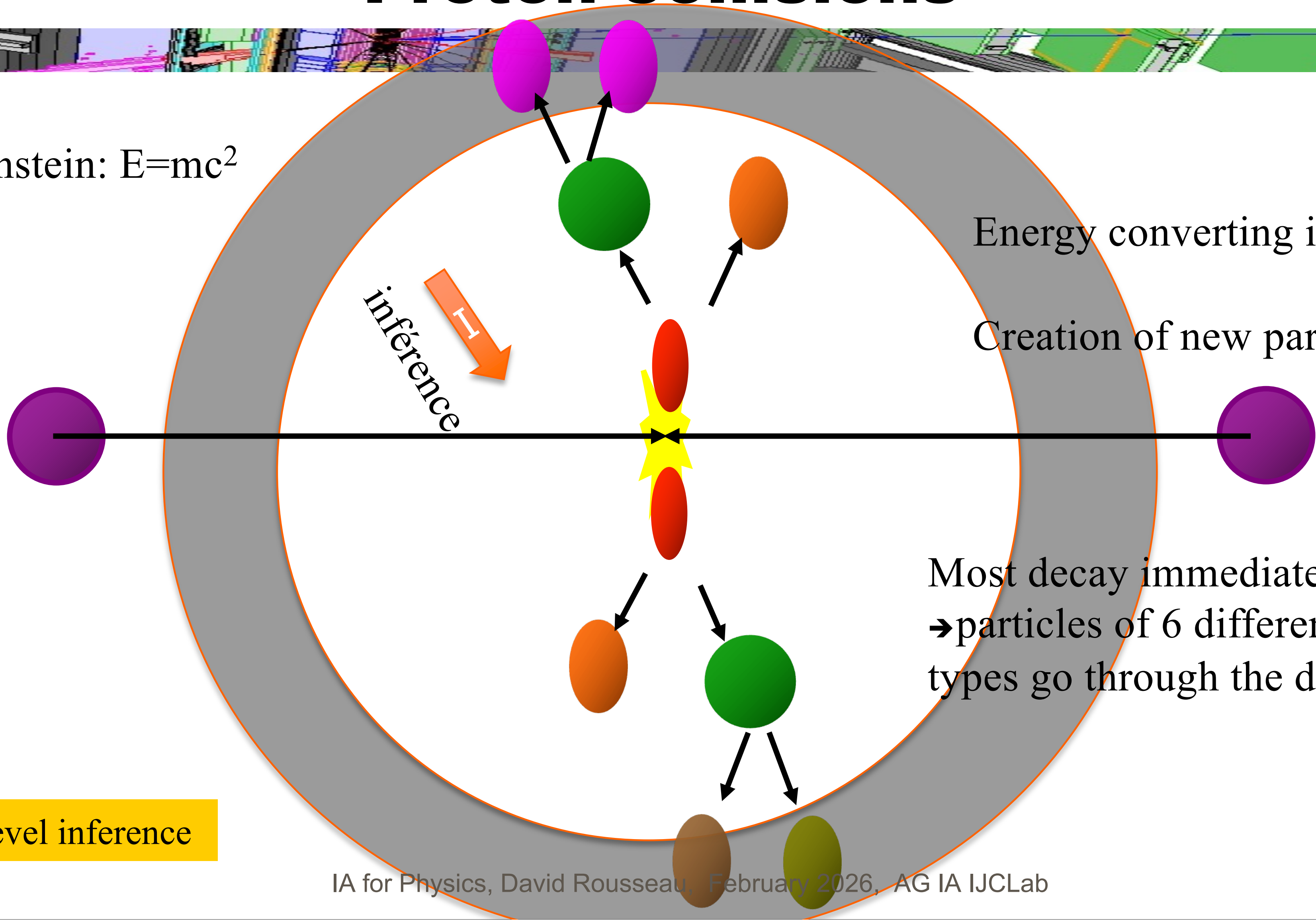
IJCLab-Orsay

david.rousseau@in2p3.fr

AG IA, février 2026

Proton collisions

Young Einstein: $E=mc^2$



Energy converting into mass

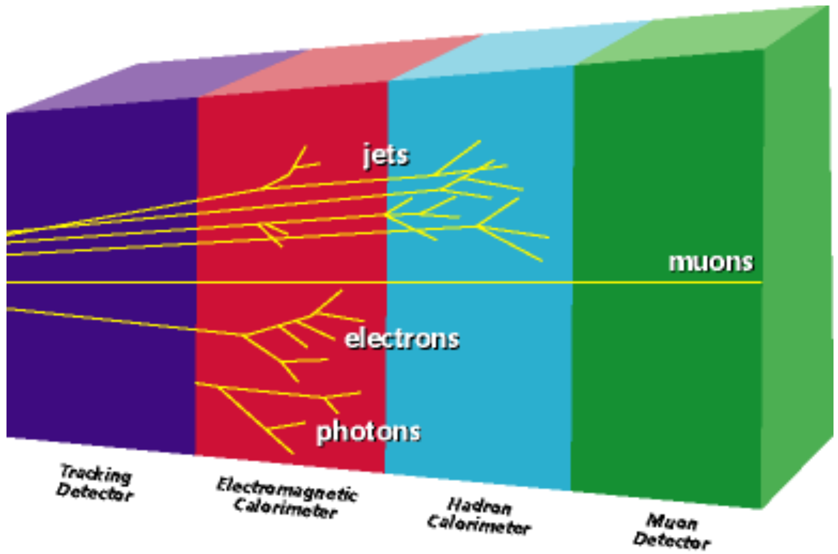
Creation of new particles

Most decay immediately
➔particles of 6 different
types go through the detector

Event level inference

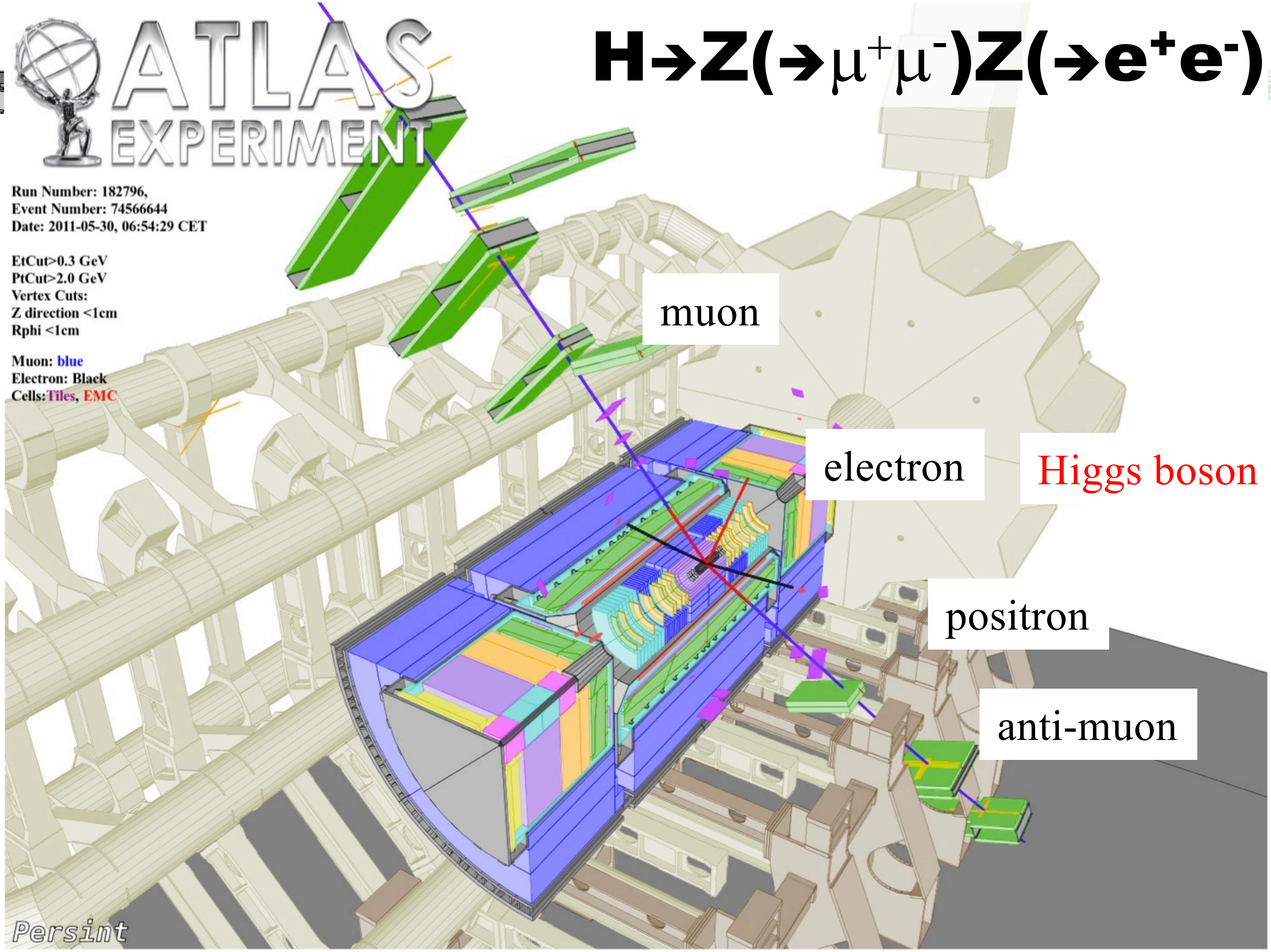
Particle-level inference

- Particles identified (pattern)
- And measured : 3D direction and energy, origin



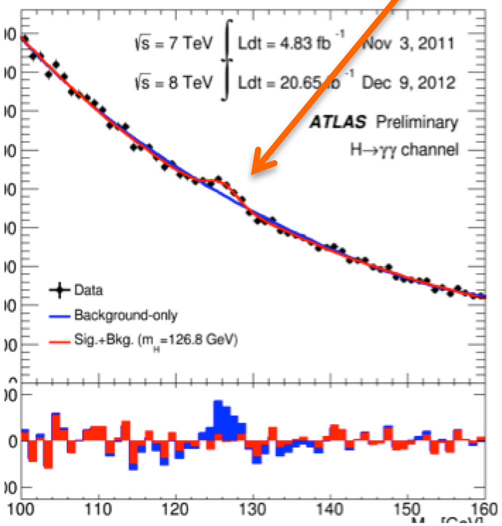
Particle level inference

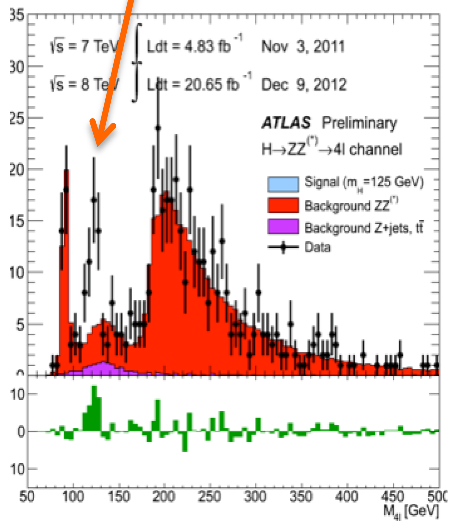
$H \rightarrow Z(\rightarrow \mu^+\mu^-)Z(\rightarrow e^+e^-)$

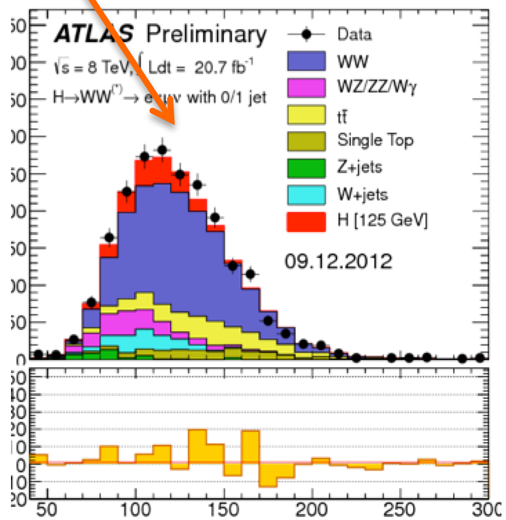


To « see » the Higgs boson

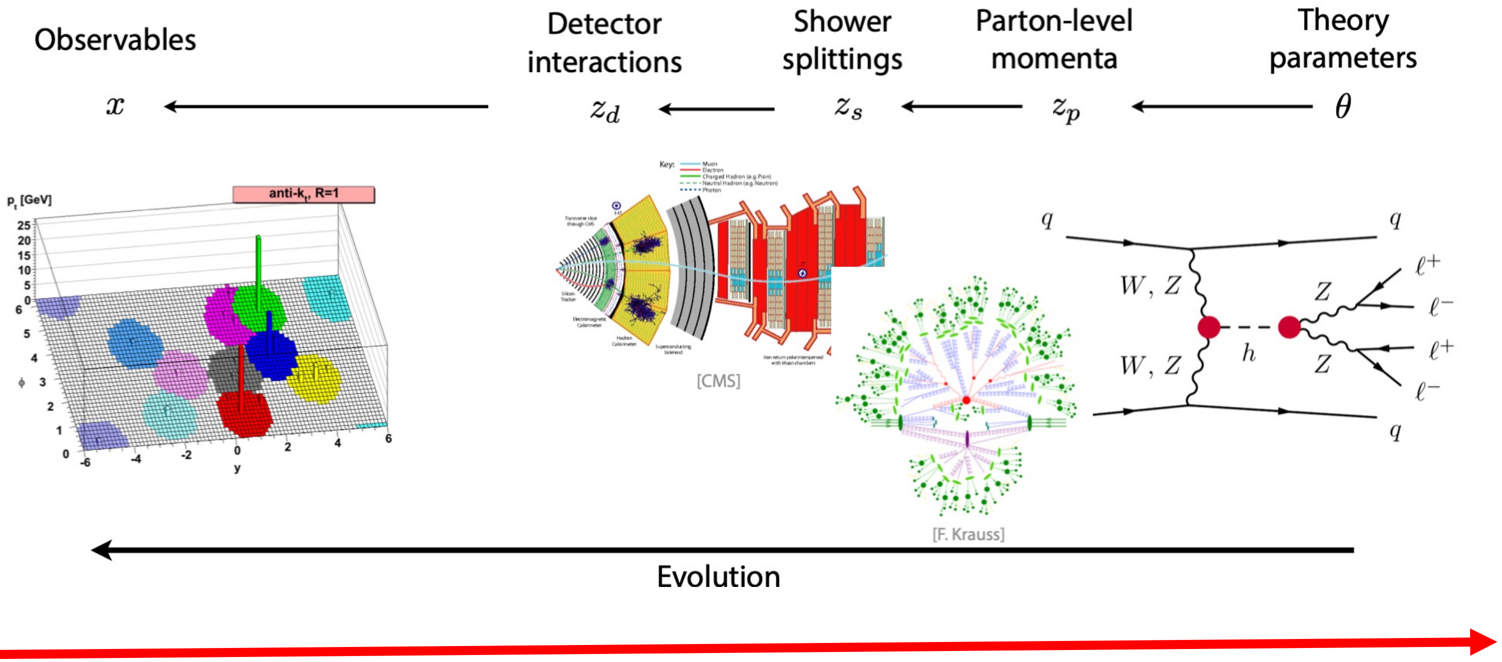
Experiment level inference





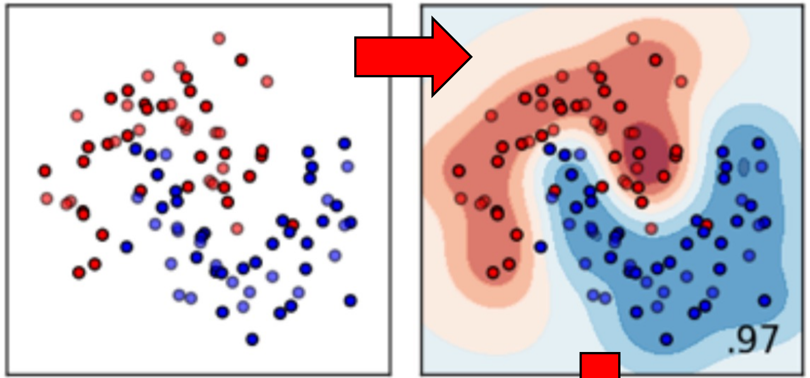


Modélisation et Inference

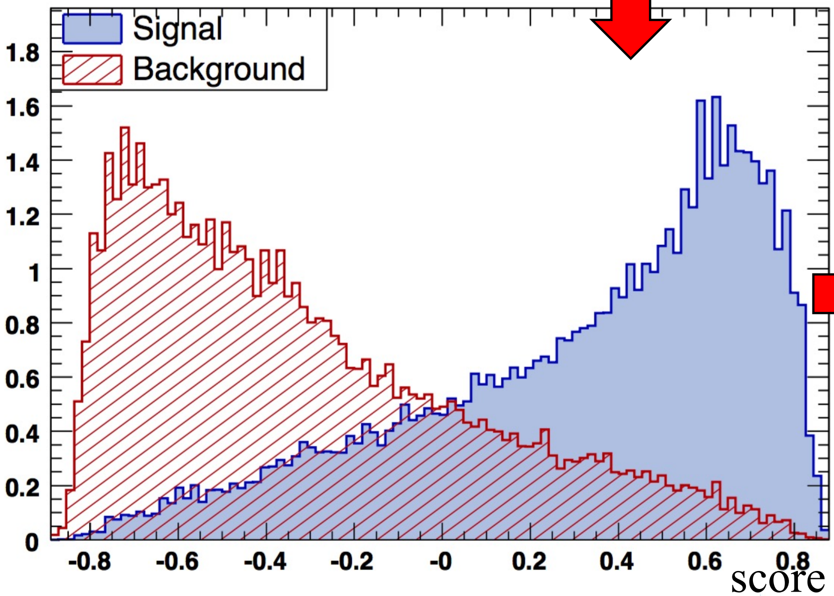


Inférence

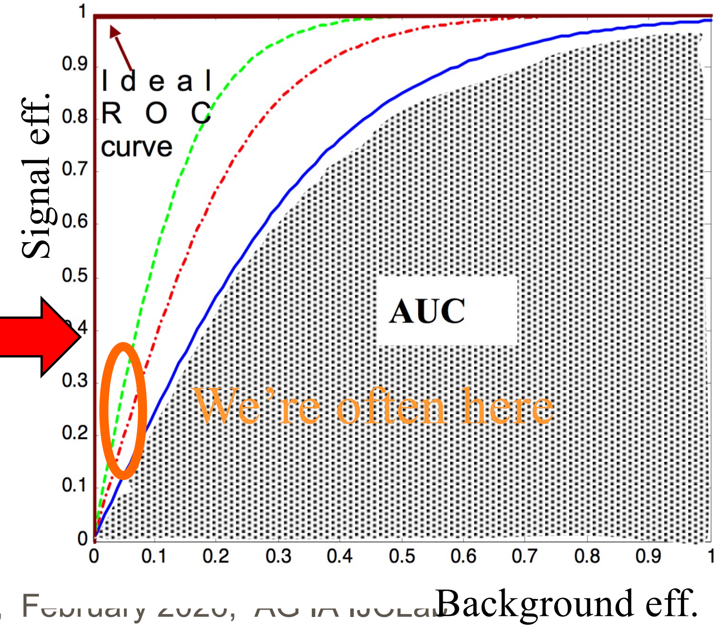
Classifier basics



Train on Signal and Background Monte-Carlo
➔learn the separation between S and B distribution
Optimize a score from most background
like to most signal like.



AUC : Area Under the (ROC) Curve



Données Tabulaires

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa

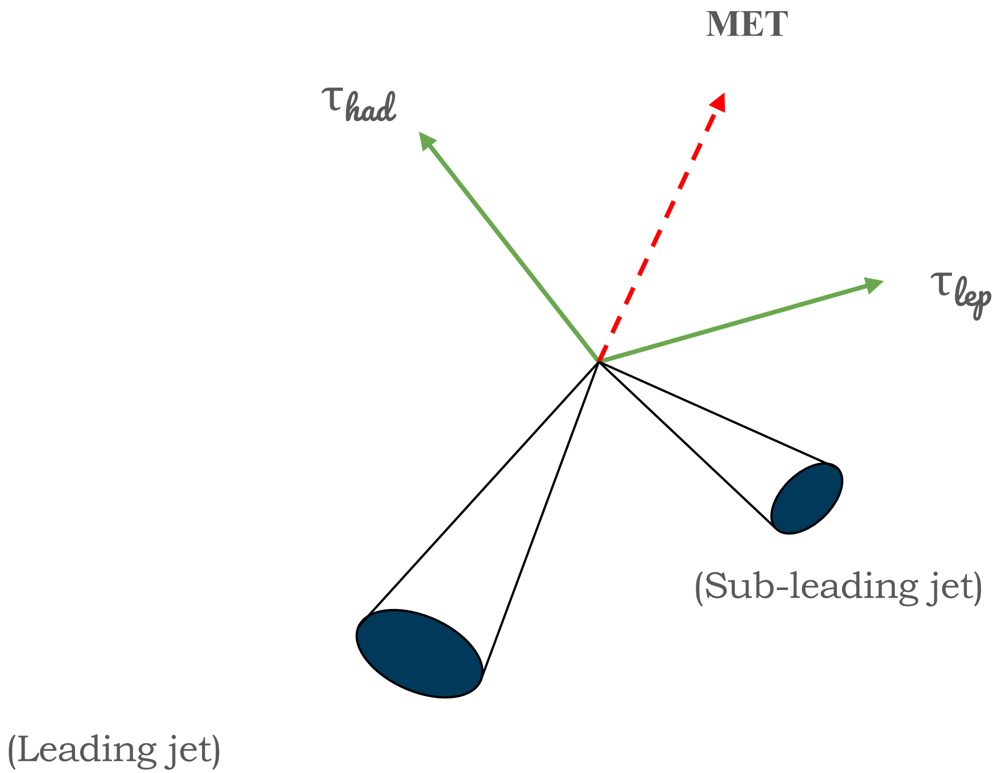
```
df = pd.read_csv('assets/train.csv')
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 2117
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 1759
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450

	dijet_invmass	dijet_deltaeta	jet_pt_0	jet_pt_1	eta_zepp_ZZ	min_dR_jZ
39	502.338409	3.694098	140.582153	43.095196	1.529987	1.799823
76	166.194427	0.426846	171.107452	81.588737	0.663560	1.020612
97	269.543396	2.568801	81.123795	64.938507	0.404464	0.050431
107	130.786301	0.119691	171.627014	31.095165	1.329497	0.539539
129	139.976868	2.145803	51.312862	37.323059	3.293238	0.423458

- Qq dizaines de variables (=colonnes)
- Arbre de Décision Boosté (BDT) pour apprendre une ou plusieurs colonnes a partir des autres
- ➔fonctionne très bien

Four-momentum final state : H →tautau



Four Momentum final state

ASCII csv file, with mixture of Higgs to tautau signal and corresponding background, from official GEANT4 ATLAS simulation

Weight and signal/background (for training dataset only)

weight (fully normalised)

label : « s » or « b »

Event features

- DER_mass_MMC
- DER_mass_transverse_met_lep
- DER_mass_vis
- DER_pt_h
- DER_deltaeta_jet_jet
- DER_mass_jet_jet
- DER_prodeta_jet_jet

VBF signature

- DER_deltar_tau_lep
- DER_pt_tot
- DER_sum_pt
- DER_pt_ratio_lep_tau
- DER_met_phi_centrality
- DER_lep_eta_centrality

Primitive 3-vectors allowing to compute the conf note variables (mass neglected),

16 independent variables:

- PRI_tau_pt
- PRI_tau_eta
- PRI_tau_phi
- PRI_lep_pt
- PRI_lep_eta
- PRI_lep_phi
- PRI_met
- PRI_met_phi
- PRI_met_sumet
- PRI_jet_num (0,1,2,3, capped at 3)
- PRI_jet_leading_pt
- PRI_jet_leading_eta
- PRI_jet_leading_phi
- PRI_jet_subleading_pt
- PRI_jet_subleading_eta
- PRI_jet_subleading_phi

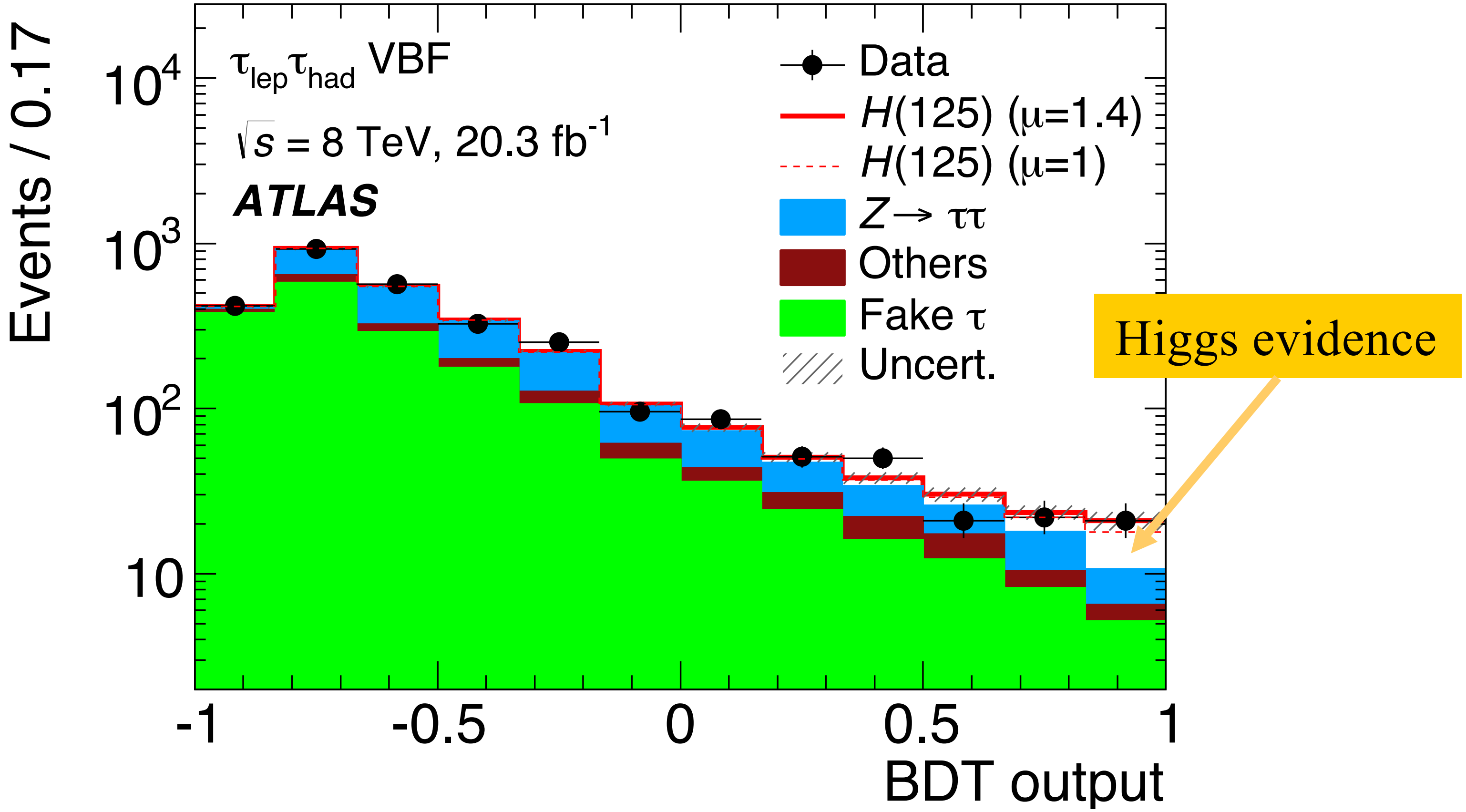
VBF signature

- PRI_jet_all_pt

Classifier

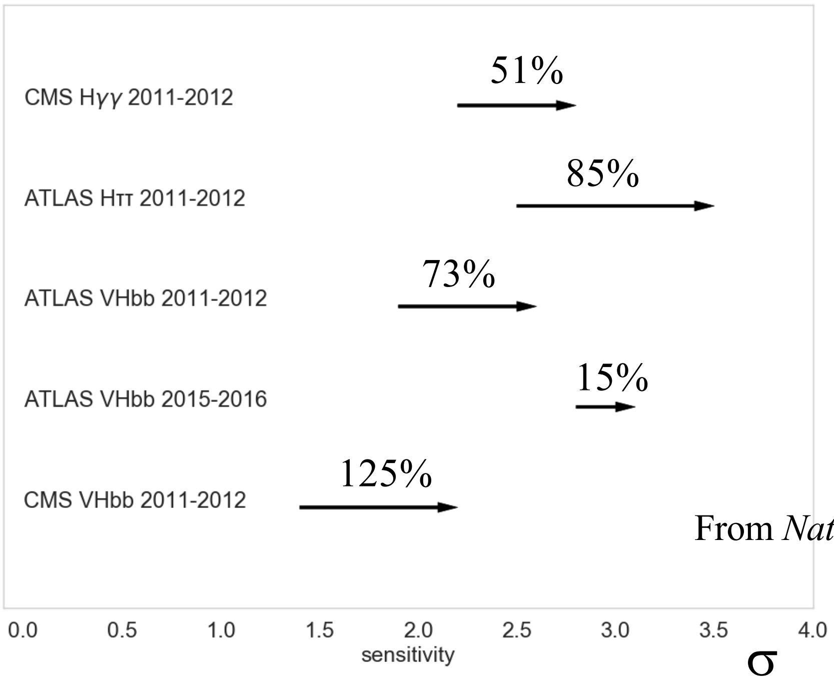
JHEP 04, 117 (2015) 1501.04943

BDT using ~dozen of high level variables



ML on Higgs Physics

- At LHC, Machine Learning used almost since first data taking (2010) for reconstruction and analysis
- For example, impact on Higgs boson sensitivity at LHC:



From *Nature* **560**, 41–48 (2018) |

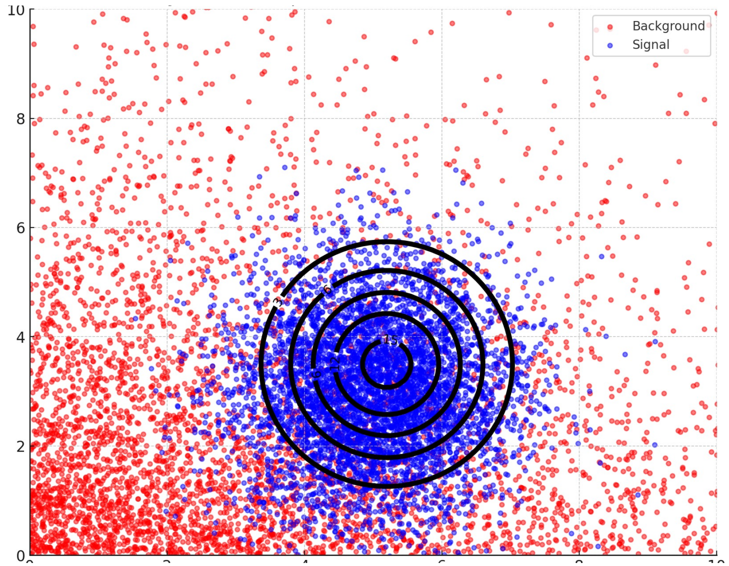
➔~50% gain on LHC running

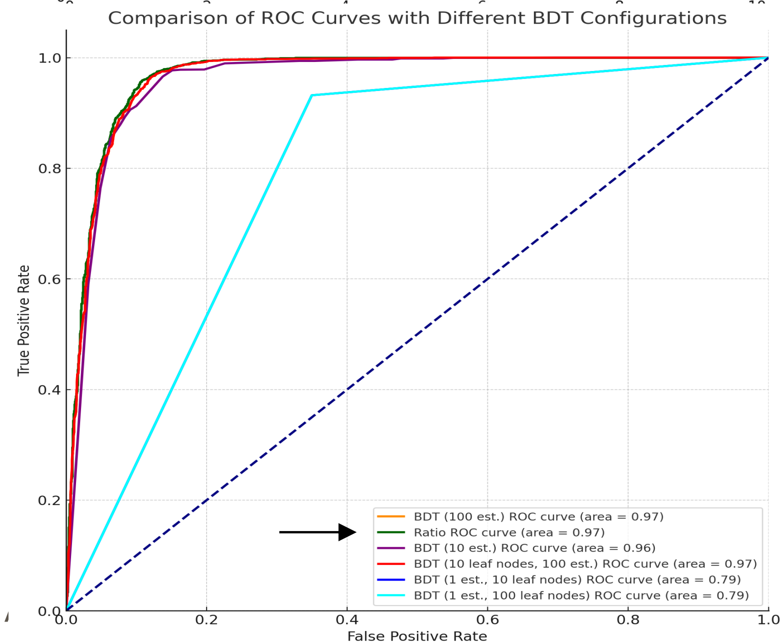
No miracle

- ML (nor Artificial Intelligence) does not do any miracles
- For selecting Signal vs Background and underlying distributions are known, nothing beats density ratio! (often called "Bayesian limit"):
  - $pdf_S(x)/pdf_B(x)$
- OK but quite often L_S L_B are unknown
  - + x is n-dimensional
- ML starts to be interesting when there is no proper model

Toy model : gauss signal on exponential background

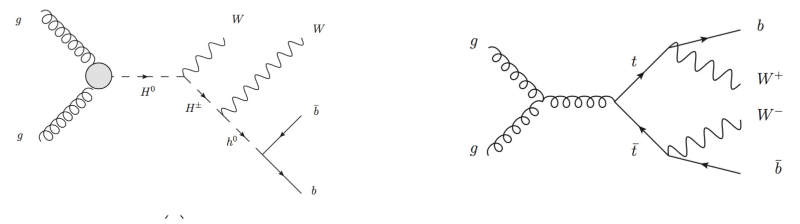
Contour lines of density ratio



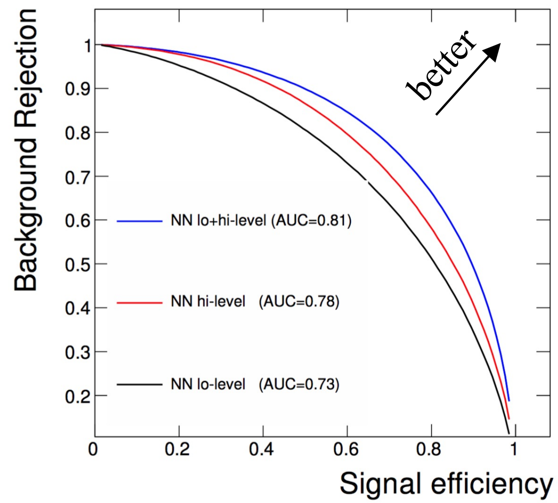


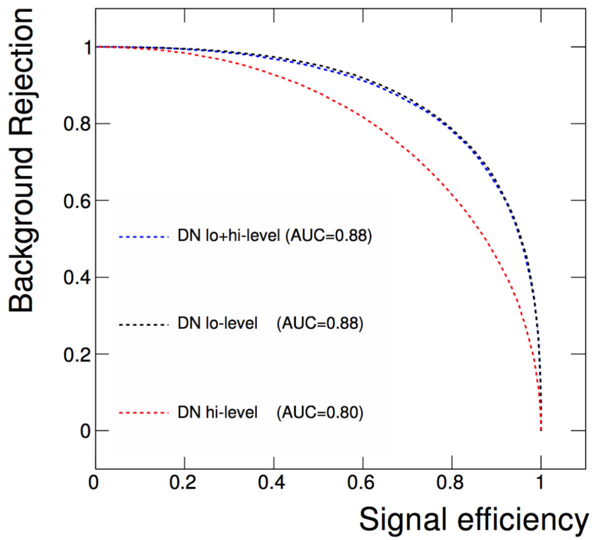
Deep learning for analysis

1402.4735 Baldi, Sadowski, Whiteson



- MSSM at LHC : $H^0 \rightarrow WWbb$ vs $tt \rightarrow WWbb$
- Low level variables:
  - 4-momentum vector
- High level variables:
  - Pairwise invariant masses
- Deep NN outperforms NN, and does not need high level variables
- DNN learns the physics ?
- Gros problème : 10 Millions d'événements pour entrainer, 100 x ce qui est disponible d'habitude



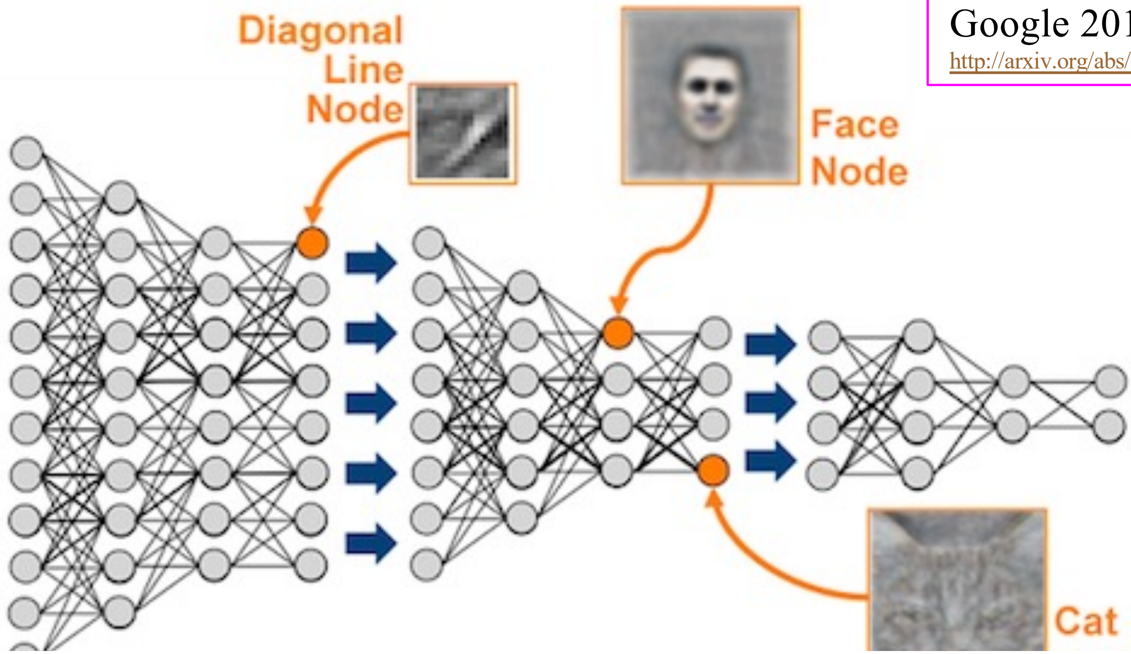


Neural Network sur des images

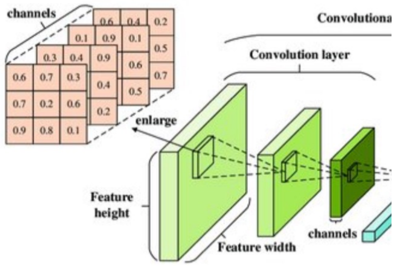
Réseau convolutionnel

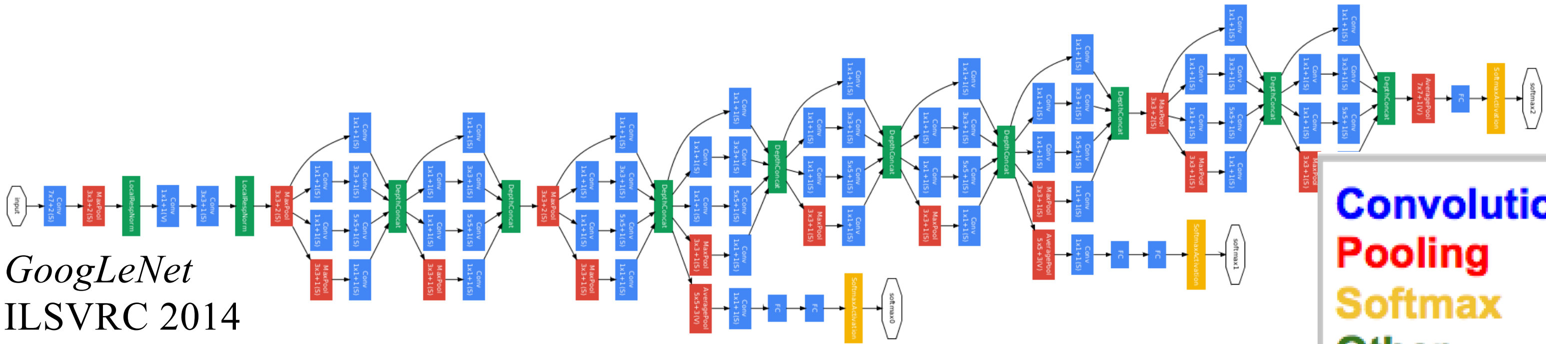
Convolutional Neural Network

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



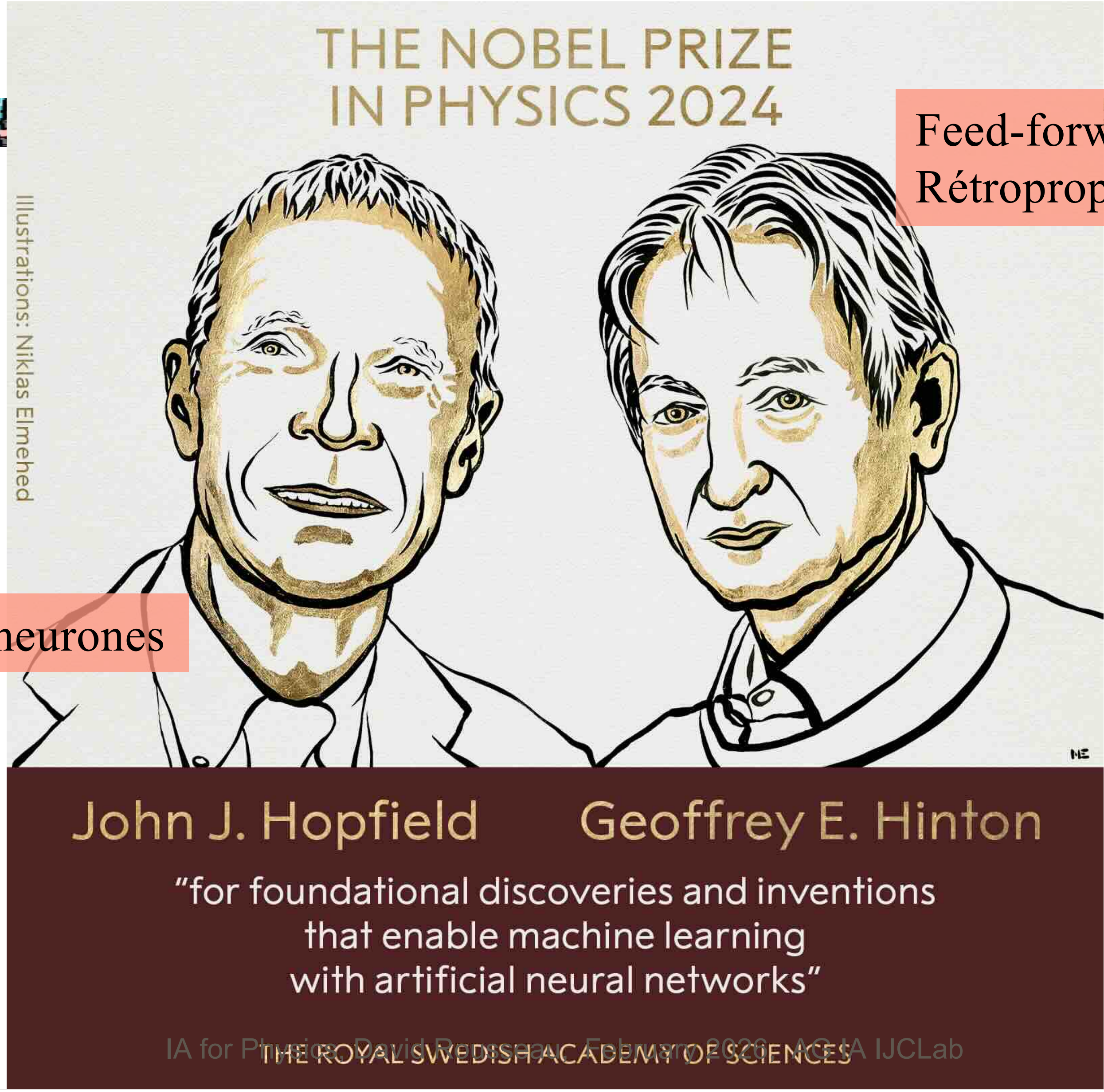
Dog
Cat
…





GoogLeNet
ILSVRC 2014
Winner

Convolution
Pooling
Softmax
Other



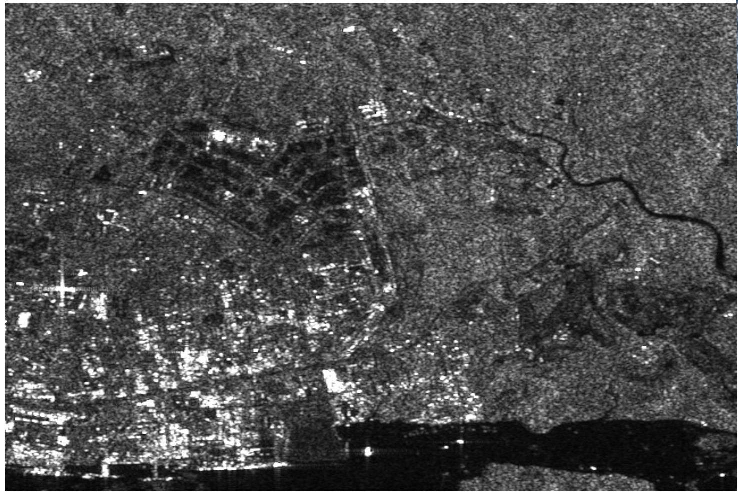
Feed-forward NN
Rétropropagation du gradient

Réseau de neurones

Press release


A.M.
TURING
AWARD
2018
acm
YOSHUA BENGIO,
GEOFFREY E. HINTON
AND YANN LECUN
Feed-forward NN
Rétropropagation du gradient
NN Convolutionnel
NN Convolutionnel
For conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing

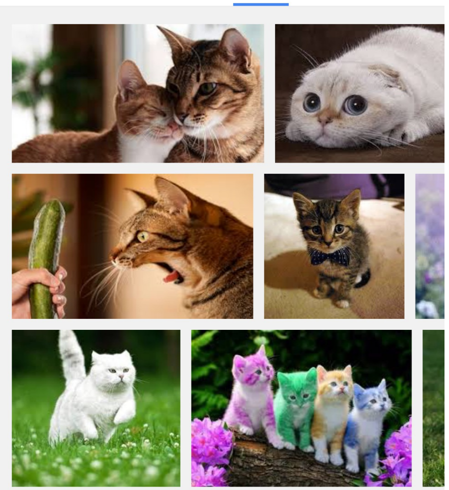
Applications innombrables NN Convolutionnels



Jet Images

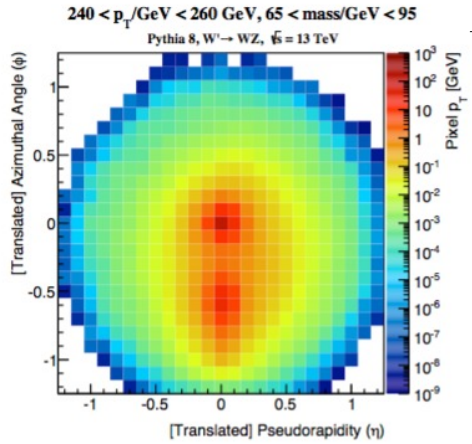
arXiv 1511.05190 de Oliveira, Kagan, Mackey, Nachman, Schwartzman

- Distinguish boosted W jets from QCD
- Particle level simulation
- Réseau convolutionnel

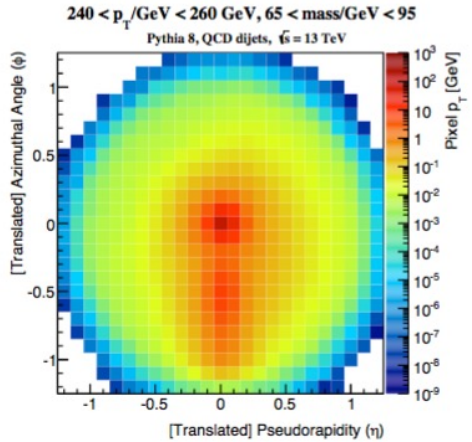


Images moyennes

Boosted W→qq jet



QCD

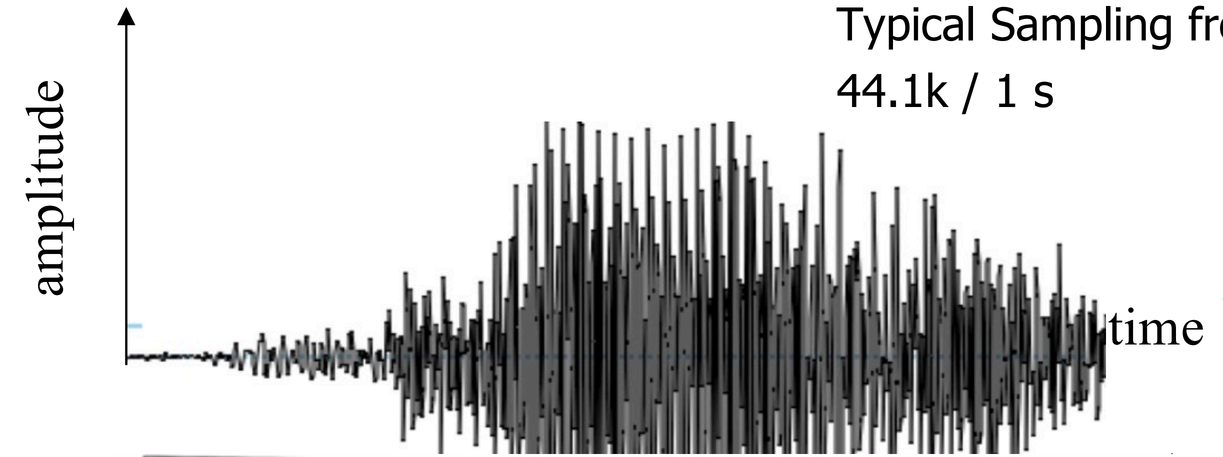


Pas de suite,
approche trop naïve

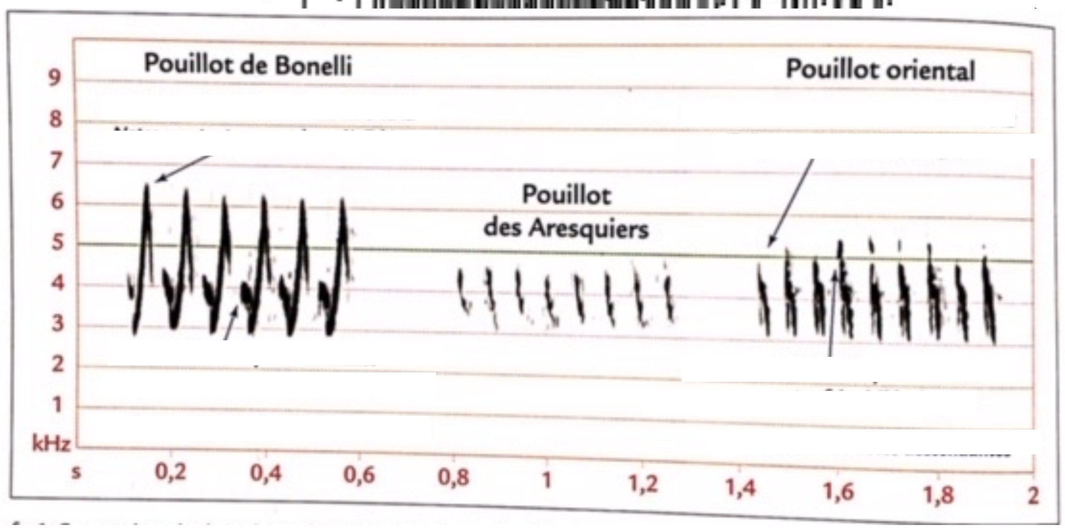
Les données scientifiques ne sont souvent pas des images

Inputs

Typical Sampling frequency 44.1kHz
44.1k / 1 s



General comment : preparing the data



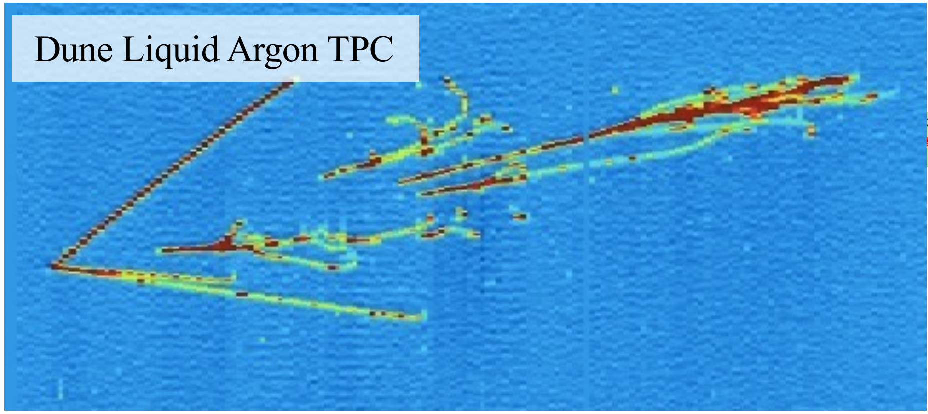
Time-frequency diagram

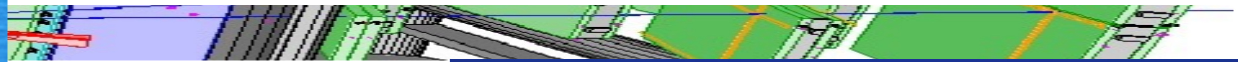
Stanislas Wroza

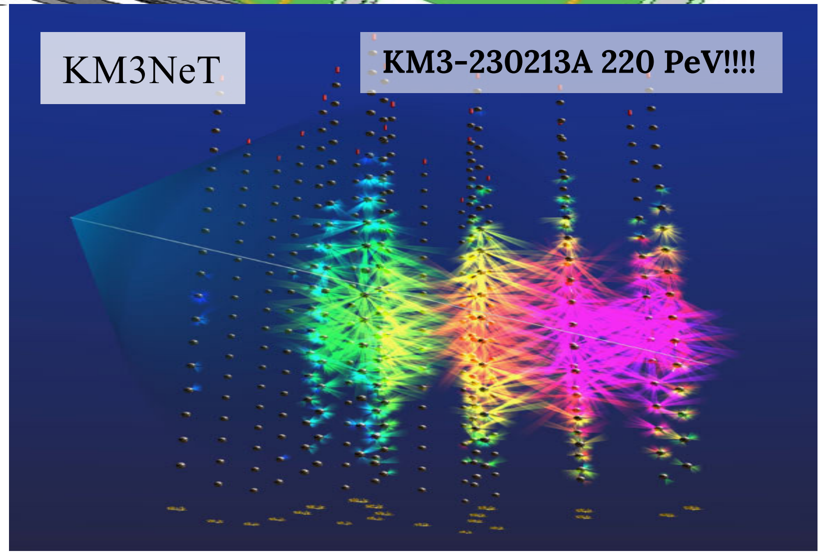
➔ images

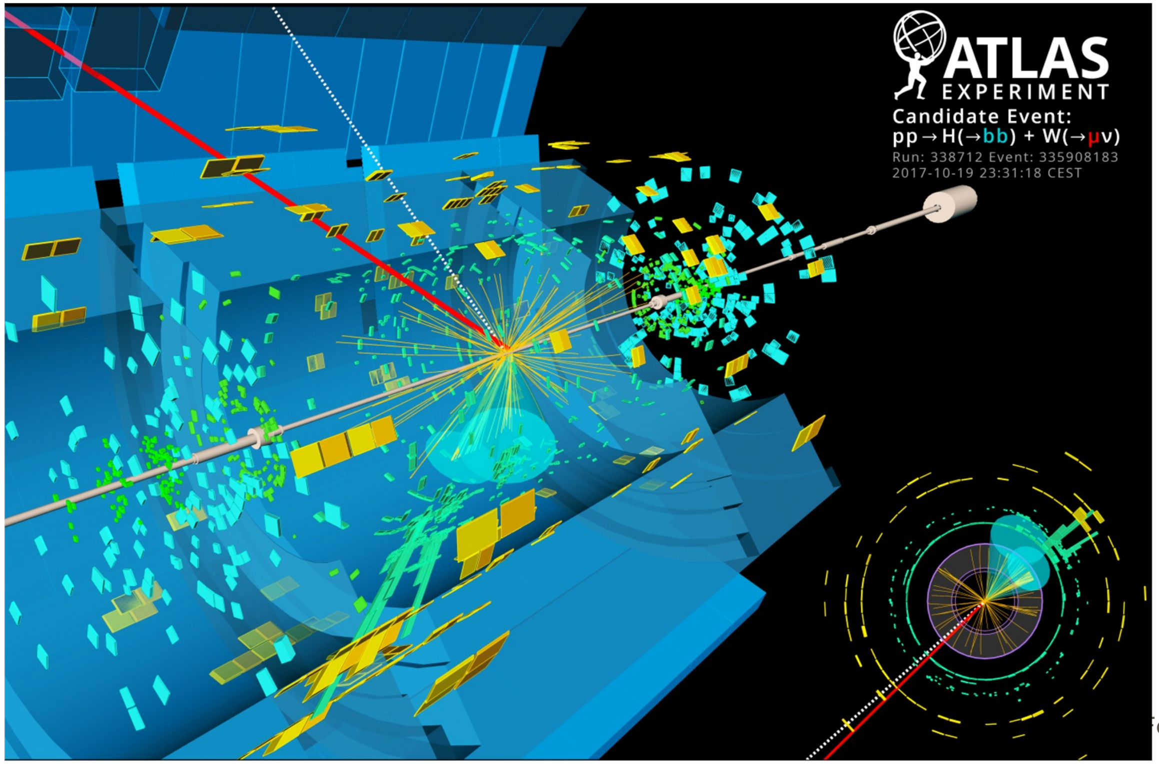
https://environnement.ijclab.in2p3.fr/2024/06/17/sorties-nature-bio-diversite

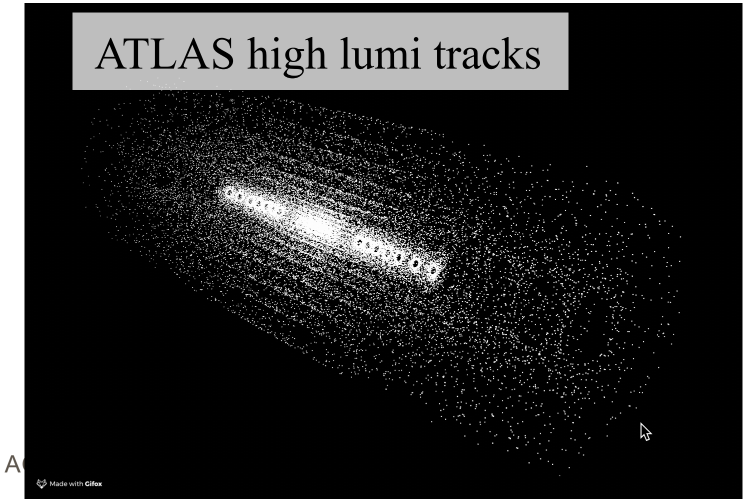
Des images, pas les données





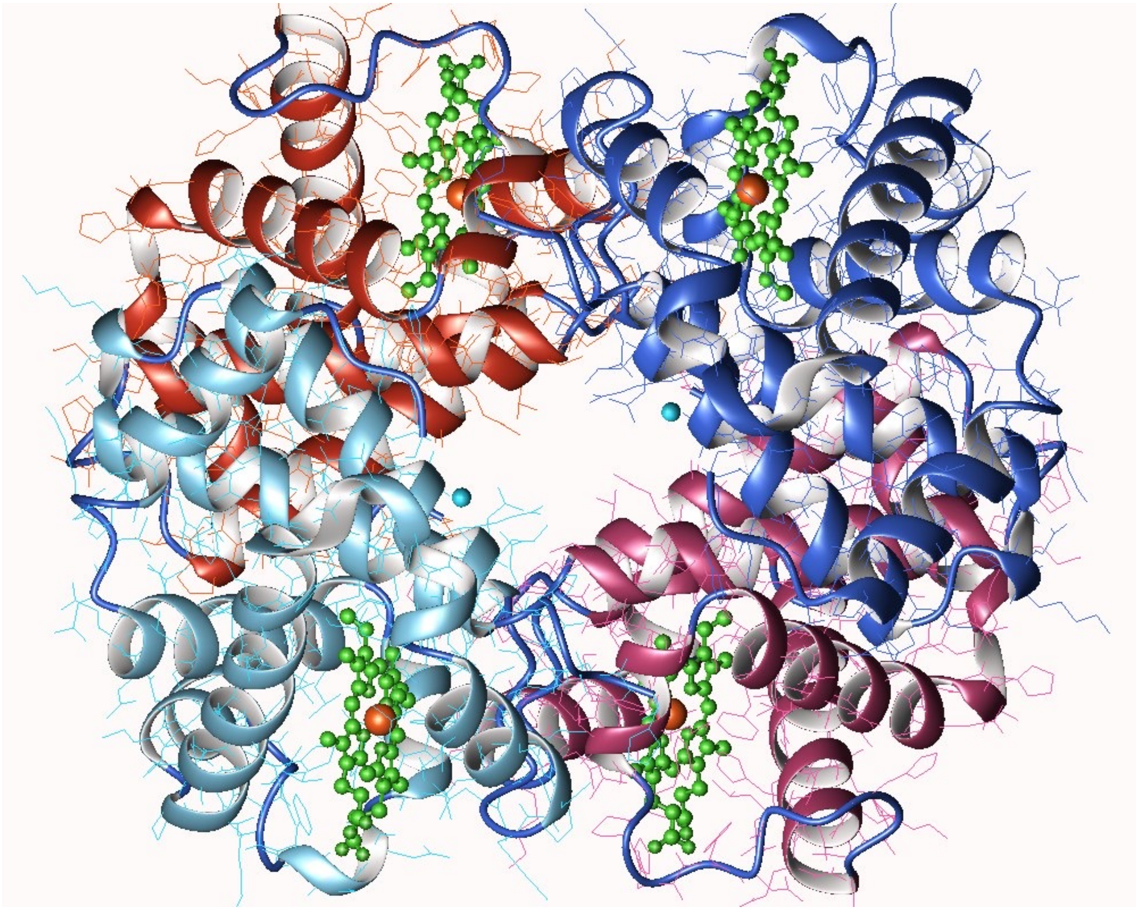






Une image, pas les données

Structure de l'hémoglobine

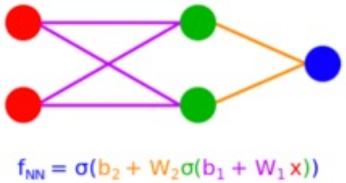


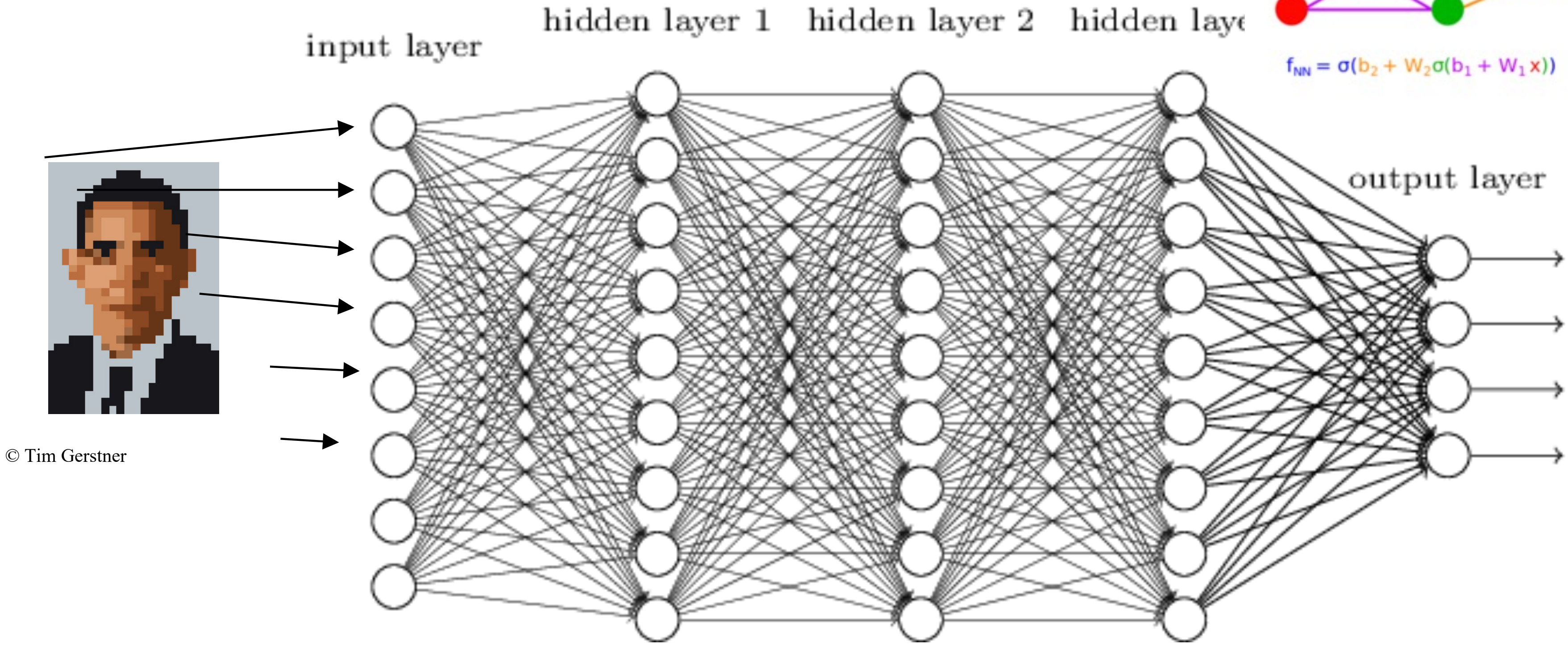
Par Deposition authors: Fermi, G., Perutz, M.F.; visualization author: User:Astrojan — https://www.rcsb.org/structure/3hhb, https://commons.wikimedia.org/w/index.php?curid=48714519

Architectures spécialisées

simplest Neural Networks

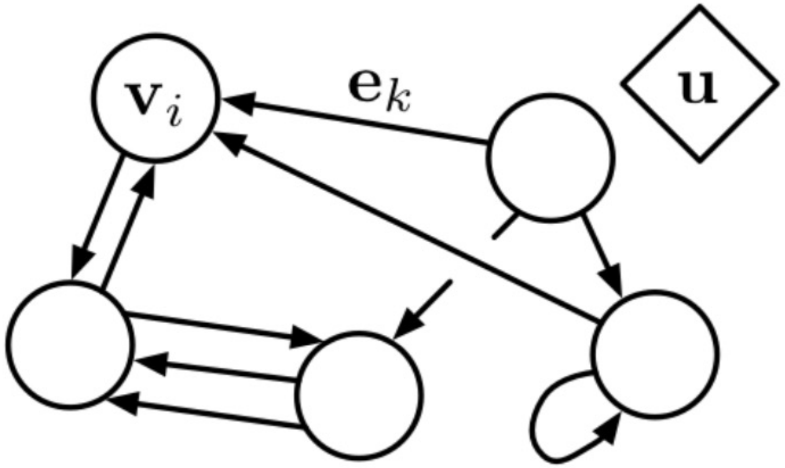
- Simple dense NN : a structure-less vector or matrix is transformed from layer to layer (Convolution NN : use neighbouring information)



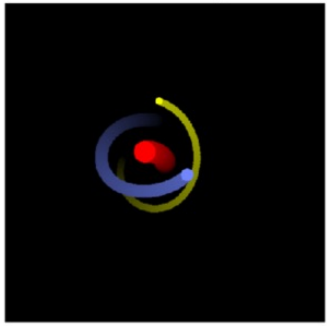


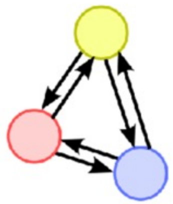
Graph Neural Networks (GNN)

- Now some structure:
  - o v_i : nodes
  - o e_k : edges
  - o u : global



n-body





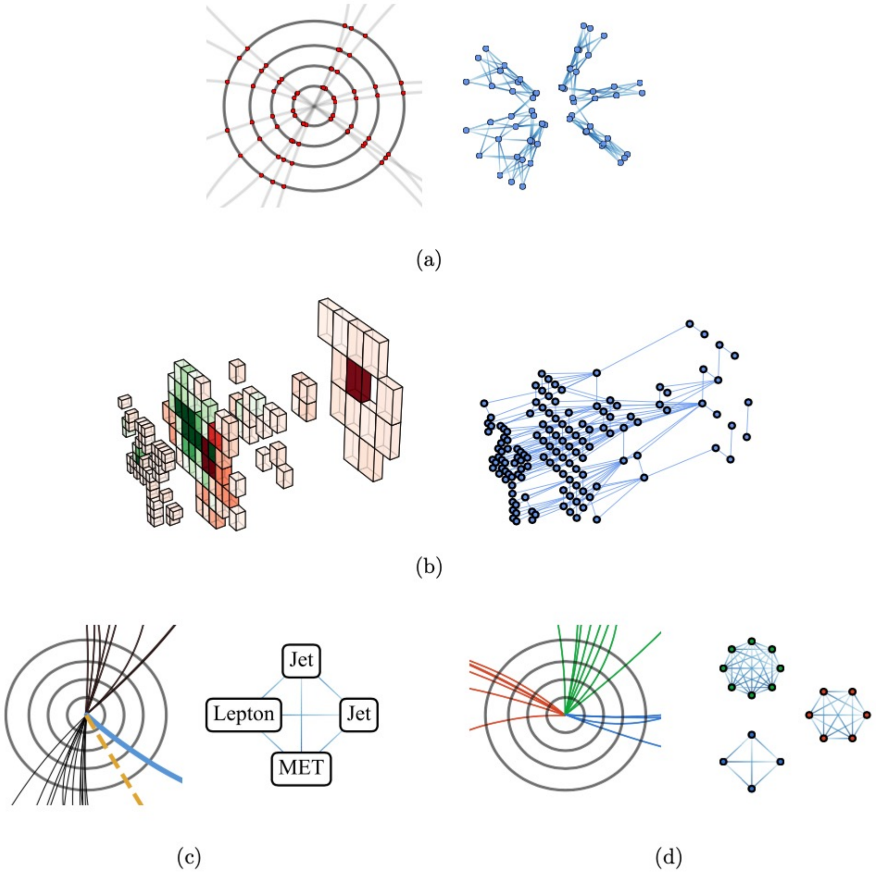
Nodes: bodies

Edges: gravitational forces

Global : potential energy

Graph to HEP data

from 2007.13681



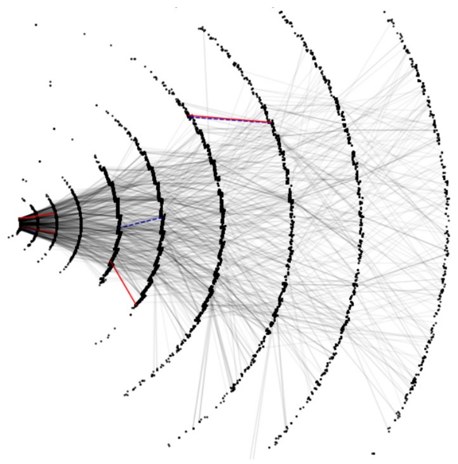
(a) (b) (c) (d)

- Construction du graphe en fonction du problème

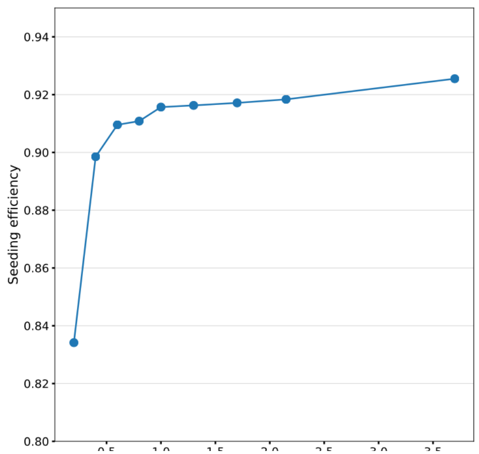
- Nouvelle évolution Transformers : corrélations entre les objets➔construction du graphe

Track Seeding with GNN

[2007.00149](2007.00149)



- Build edges between neighbour
- Then GNN trained to classify double and triplet
- High efficiency reached with subsecond computing time (also very parallelisabled)
- ➔can be used as a filtering stage before traditional Kalman filter



Intermède

Aparté on ML in Physics history

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

NEURAL NETWORKS AND CELLULAR AUTOMATA IN EXPERIMENTAL HIGH ENERGY PHYSICS

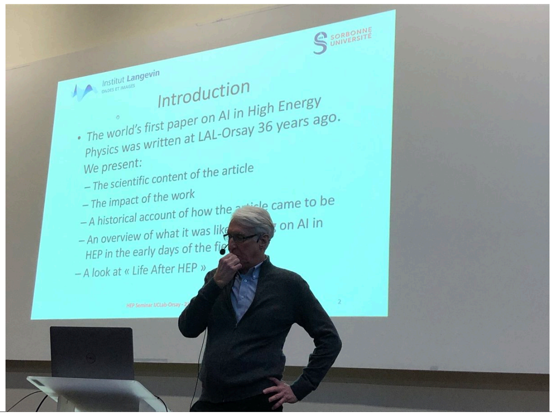
B. DENBY

Laboratoire de l'Accélérateur Linéaire, Orsay, France

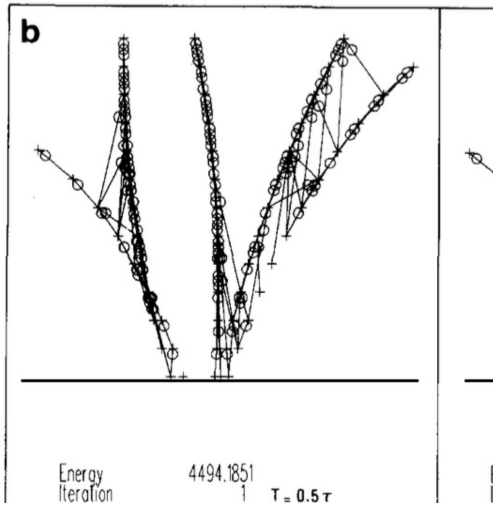
Received 20 September 1987; in revised form 28 December 1987

Bruce Denby

- 1987 Very first known paper on Neural Net in High Energy Physics

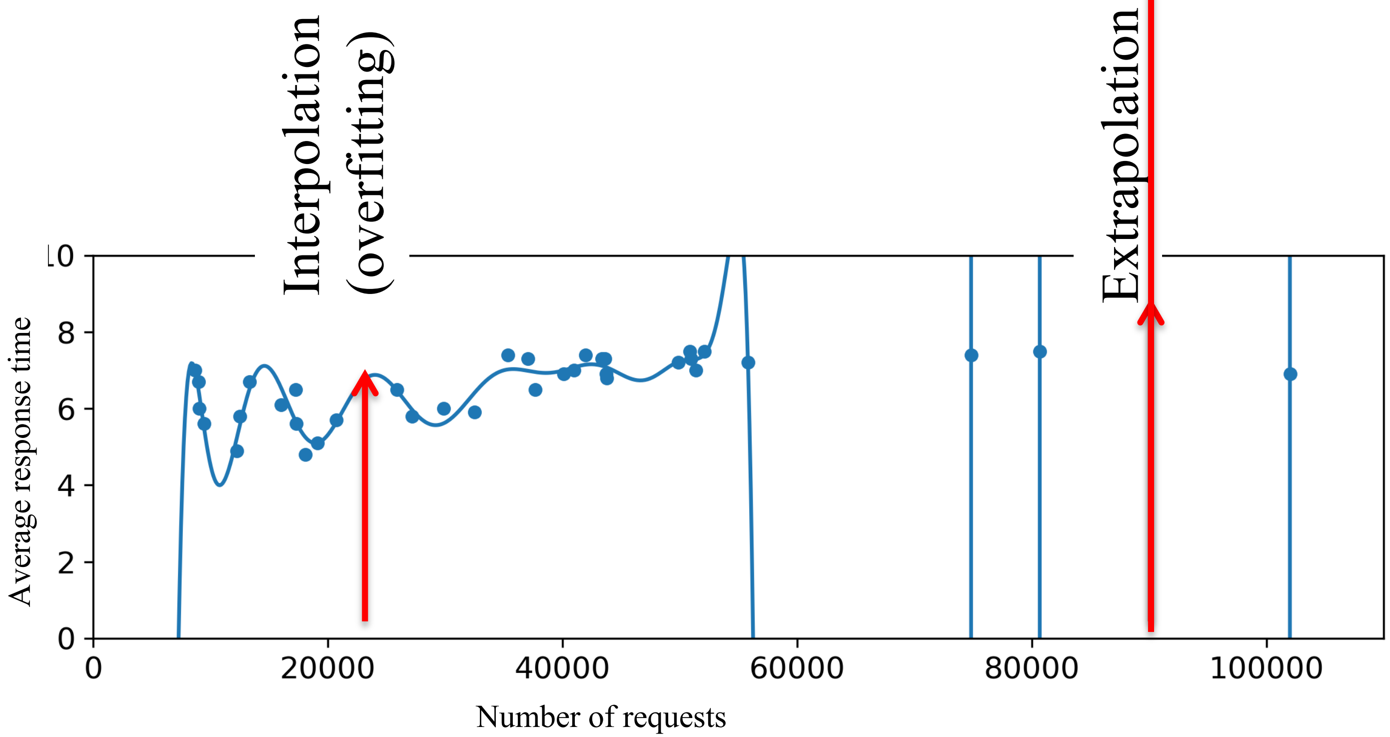


B. Denby seminar
at ex-LAL in 2023



NN interpolation

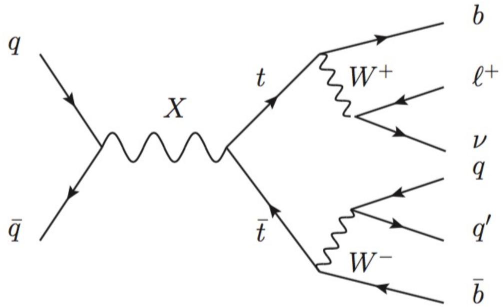
Interpolation vs Extrapolation

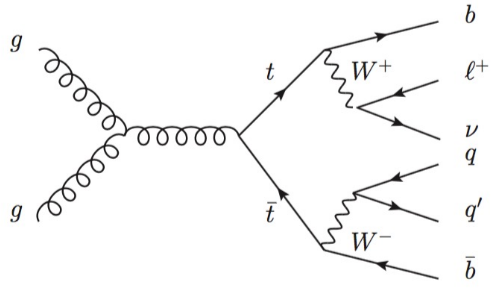


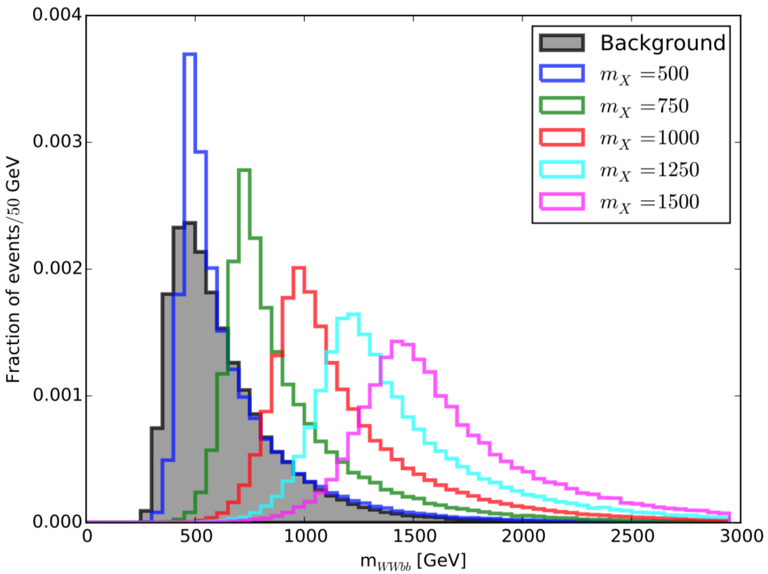
Interpolation is easier than extrapolation

Parameterised learning

1601.07913 Baldi, Cranmer, Faucett, Sadowksi, Whiteson

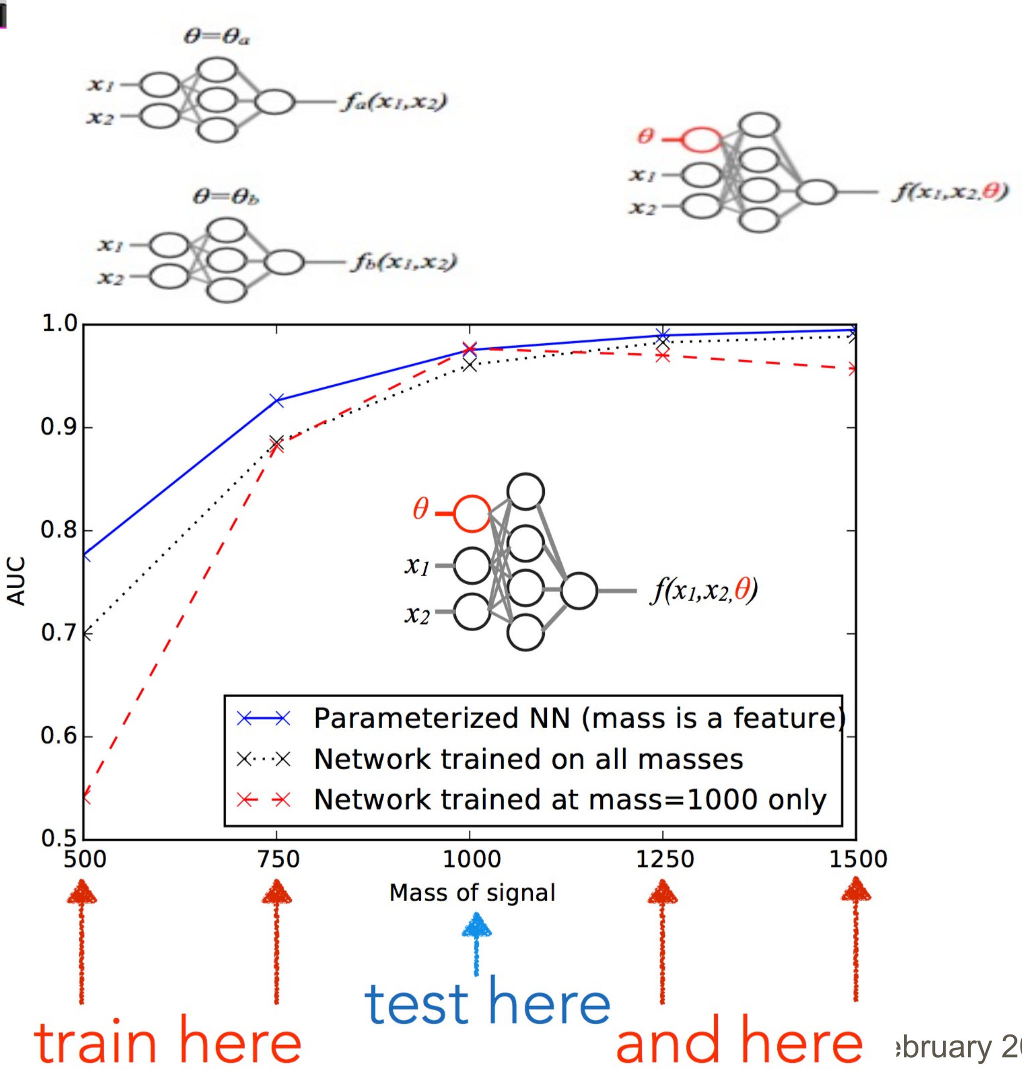






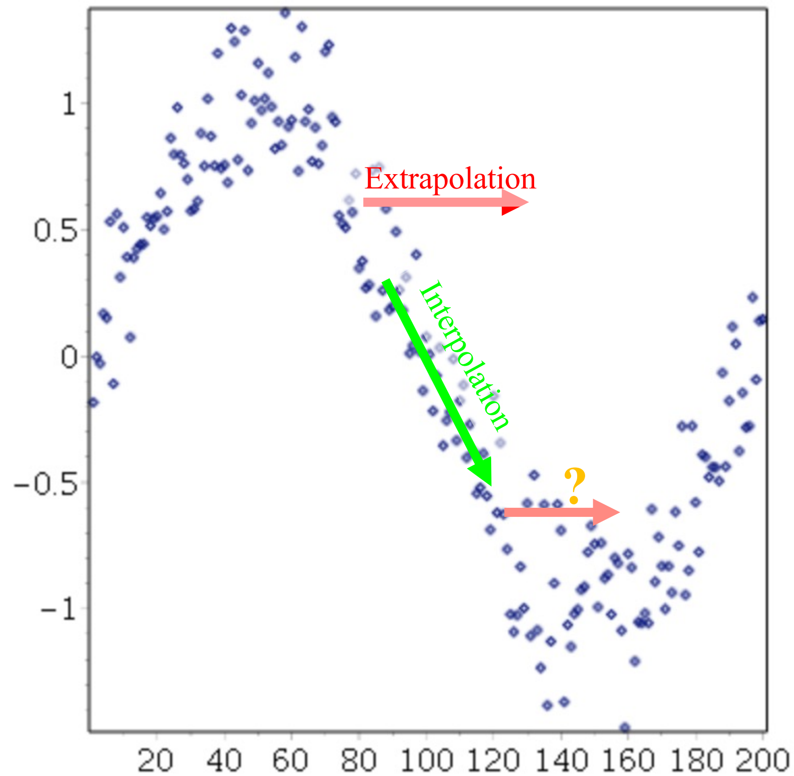
- Typical case: looking for a particle of unknown mass
- E.g. here tt decay

Parameterised learning (2)



- Train on 28 features plus true mass
- Parameterised NN as good as single mass training
- ➔clean interpolation
- (mass just an example)
- Used by CMS bblνl ν search
  https://arxiv.org/pdf/1708.04188.pdf
- Nowadays commonly used in searches

Interpolation vs Extrapolation



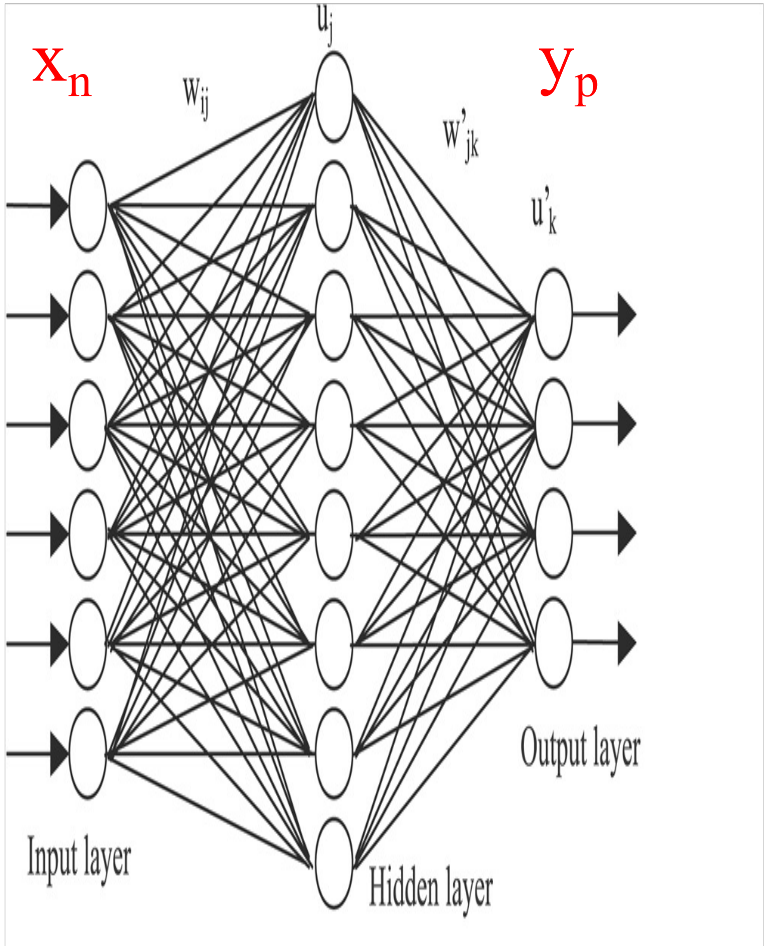
Interpolation/Extrapolation already ill-defined in 2D, what about large dimensions ?

Modèles substitutifs

Quand l'IA ne fait pas mieux mais beaucoup plus vite

Théorème d'universalité

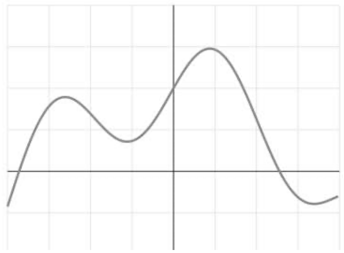
- Théorème mathématique 1991
  https://en.wikipedia.org/wiki/Universal_approximation_theorem
- Toute fonction continue $R^n \rightarrow R^p$
- … peut être approximée suffisamment bien
- … avec un NN à une couche cachée avec un nombre suffisant de neurones
- (Ne dit rien sur la construction)
- ➔on peut entrainer un NN à émuler un calculateur existant complexe
- ➔gains en vitesse x10… x100000
- (mais précision… mais excursion…)



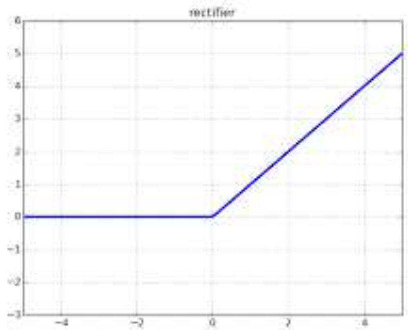
Universal Theorem at work

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

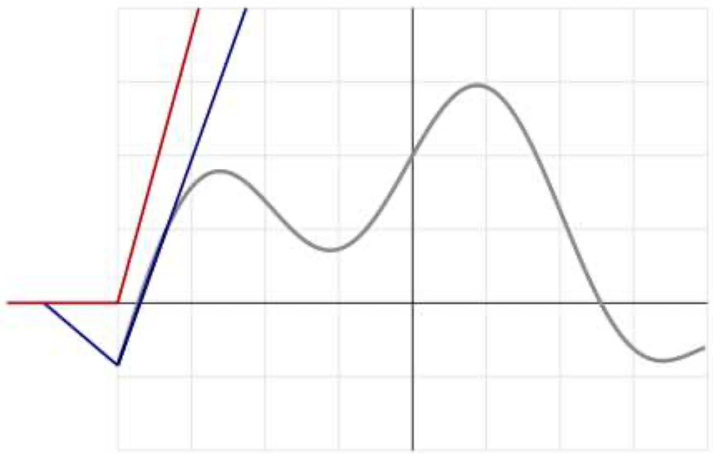


Universal approximation

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

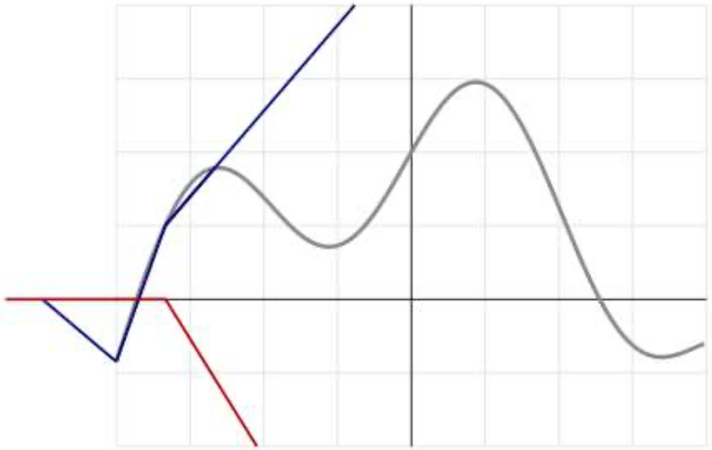
Universal approximation

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



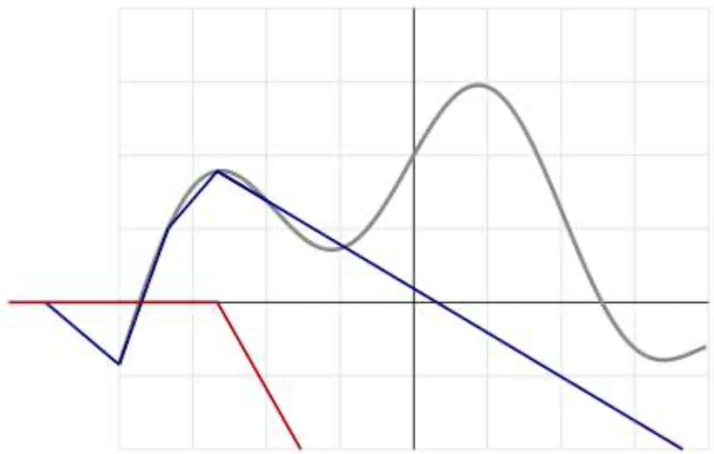
Universal approximation

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



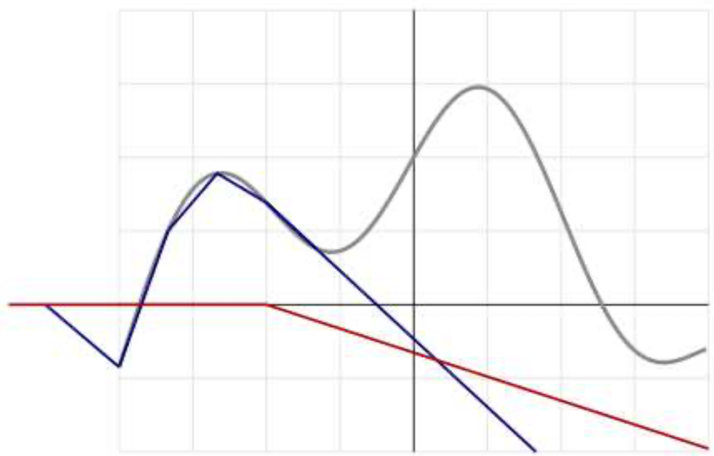
Universal approximation

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



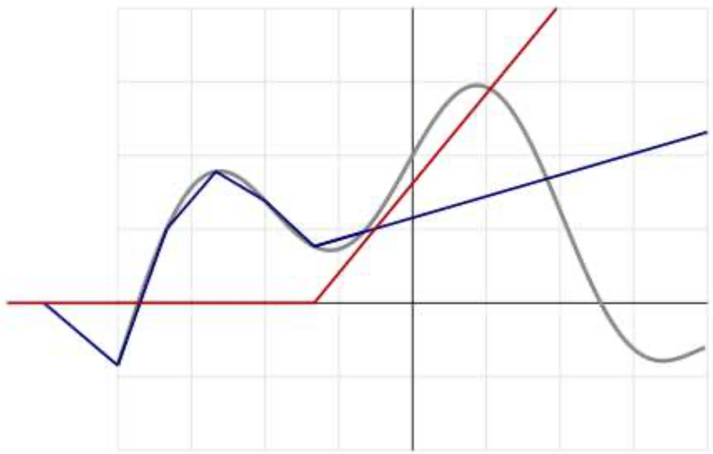
Universal approximation

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



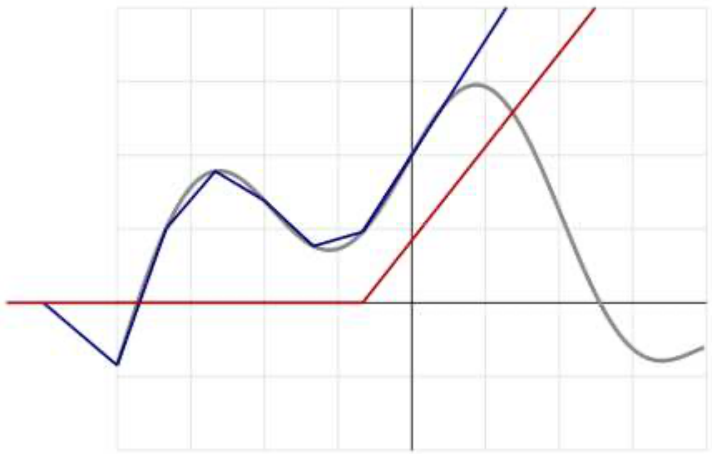
Universal approximation

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



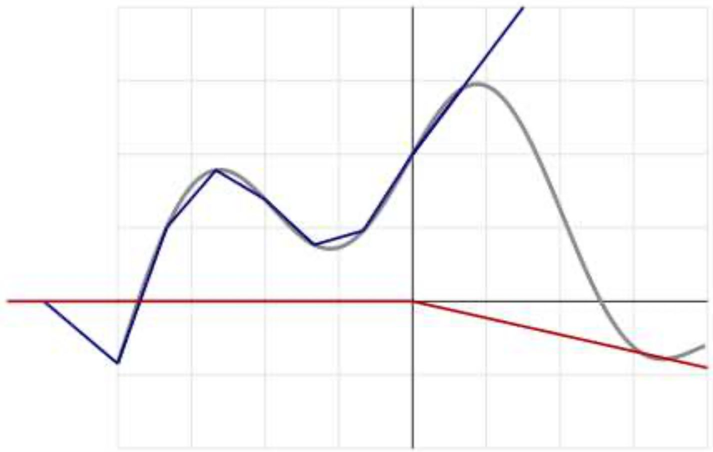
Universal approximation

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



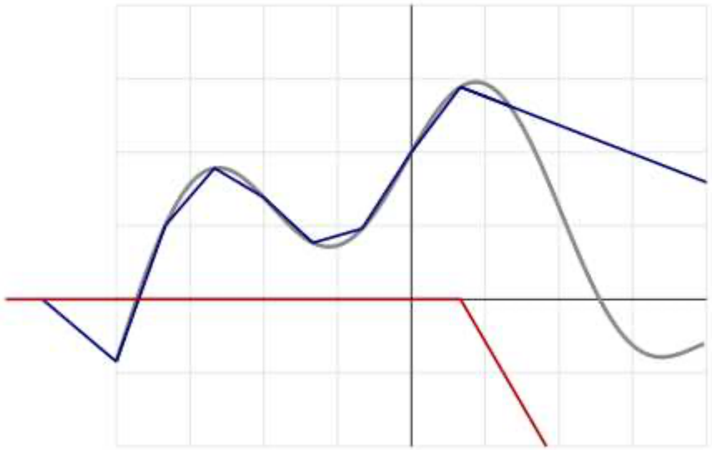
Universal approximation

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



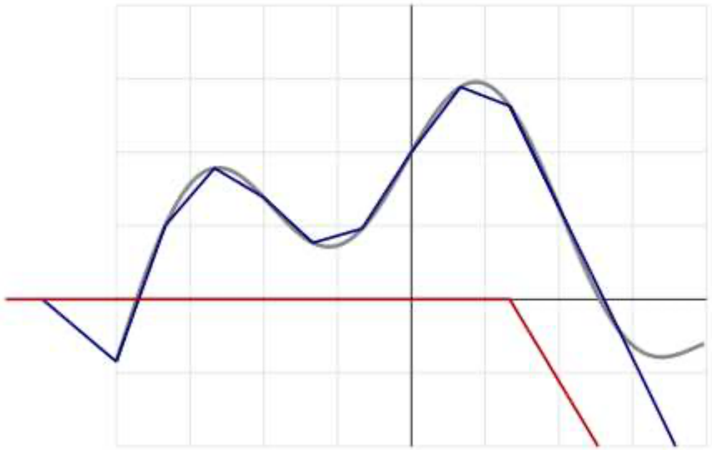
Universal approximation

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



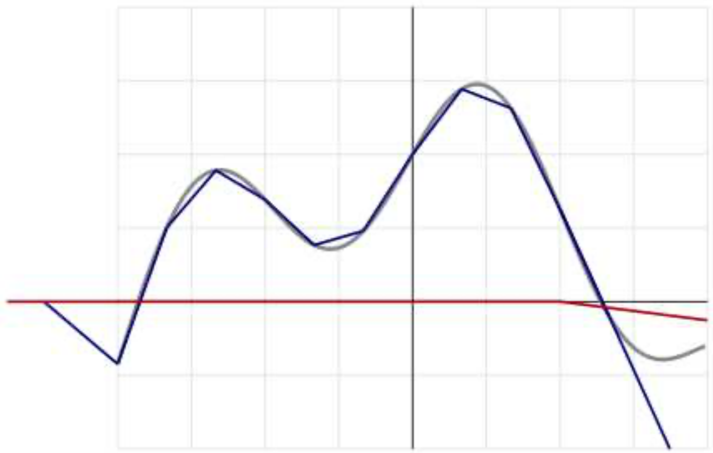
Universal approximation

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



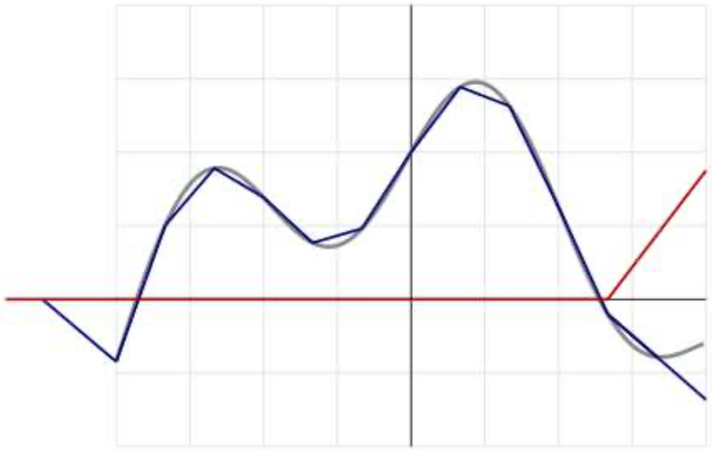
Universal approximation

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



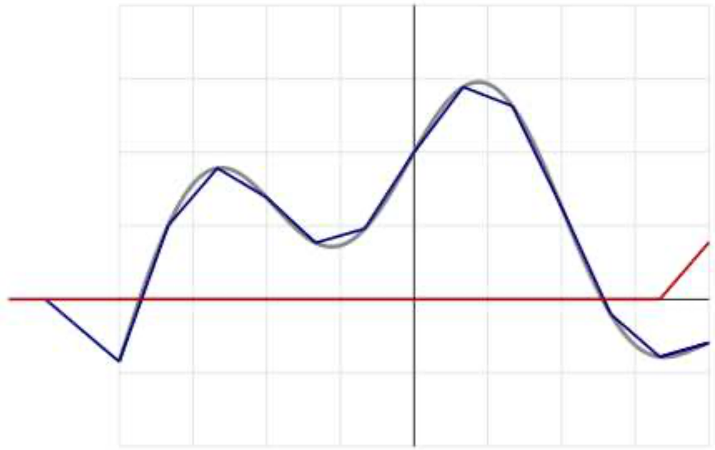
Universal approximation

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



Universal approximation

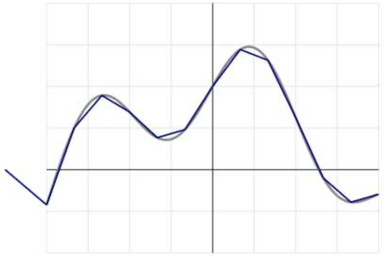
We can approximate any $f \in \mathscr{C}([a, b], \mathbb{R})$ with a linear combination of translated/scaled ReLU functions

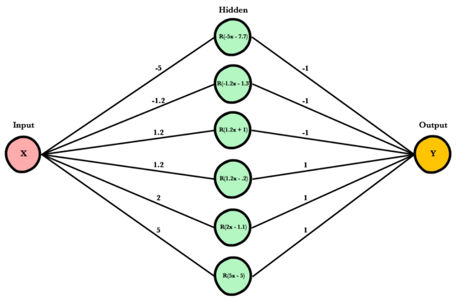


Universal Theorem at work

Universal approximation

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

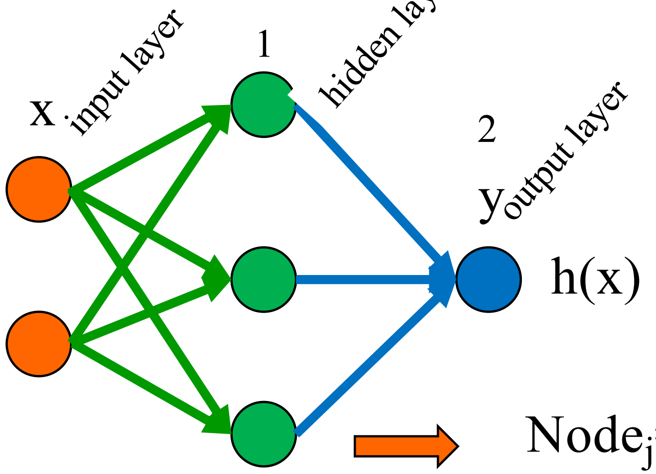




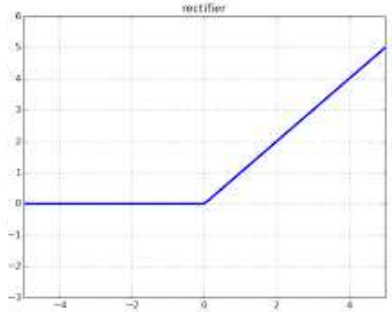
$$y = \sum_i \mathrm{Relu}(\mathrm{a_i} \times \mathrm{x} + \mathrm{b_i})$$

$$\mathbb{R} \to \mathbb{R} \text{ generalised to } \mathbb{R}^n \to \mathbb{R}^p$$

Simple NN as a math function



activation function (non linear!)



σ

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

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

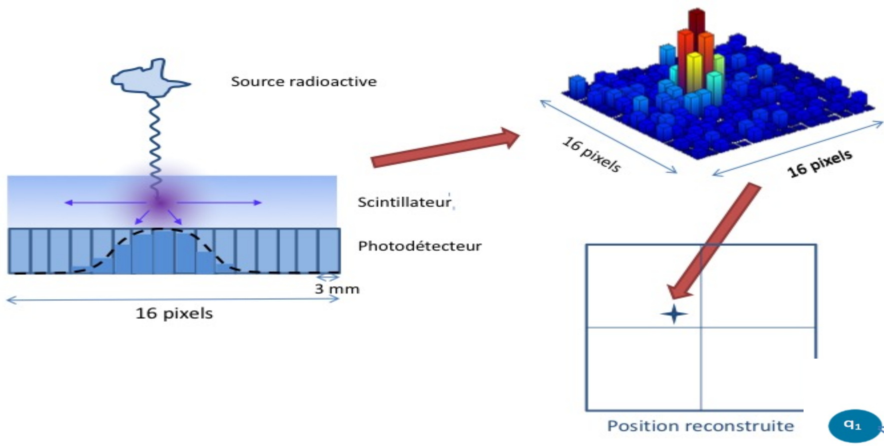
Beware: superscript are layer indices!

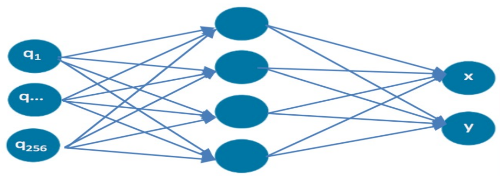
Now with dimensions

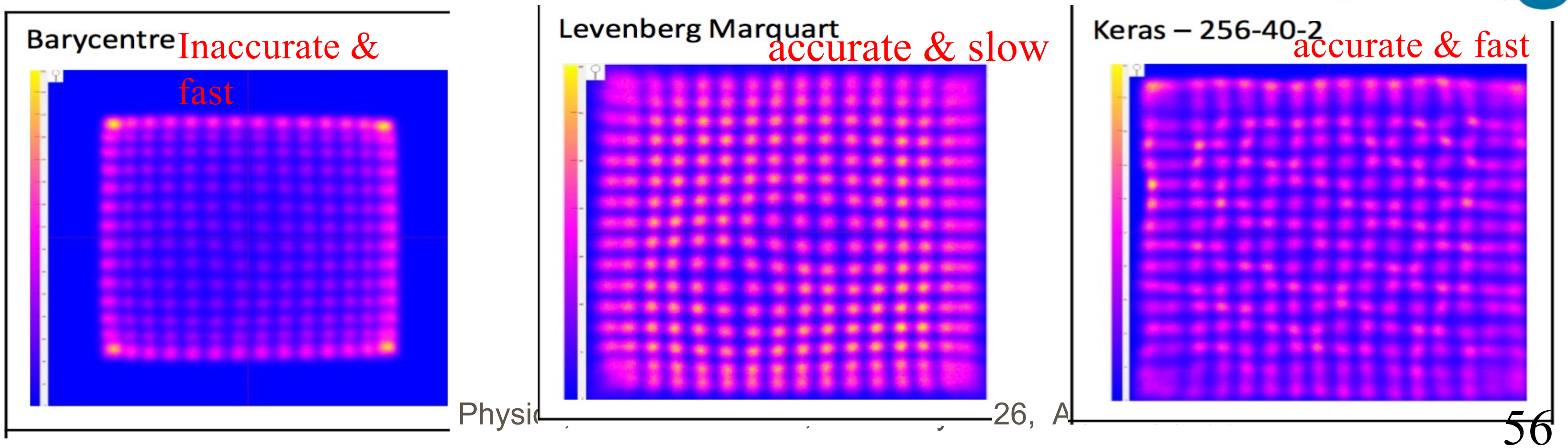
$$h(x_{(2)}) = \sigma(b^2_{(1)} + W^2_{(1,3)} \sigma(b^1_{(3)} + W^1_{(3,2)} x_{(2)}))$$

Exemple : β,γ camera pour chirurgie

Françoise Bouvet
IJCLab







Inaccurate & fast

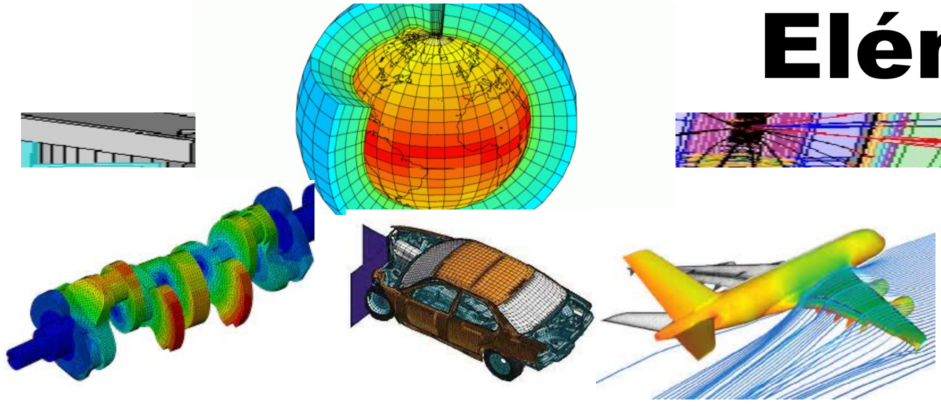
accurate & slow

accurate & fast

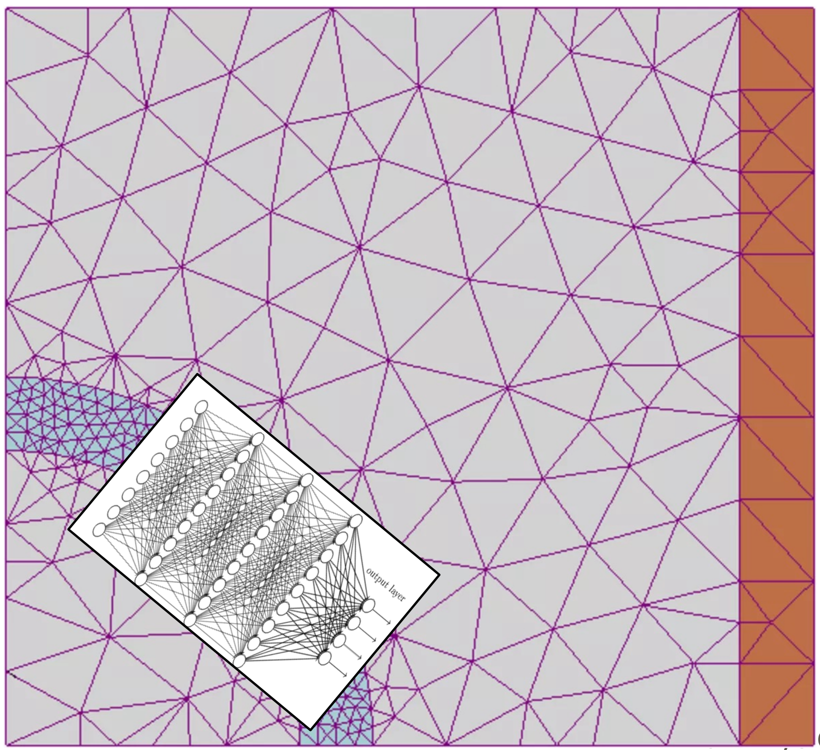
56

Eléments finis

Eléments finis

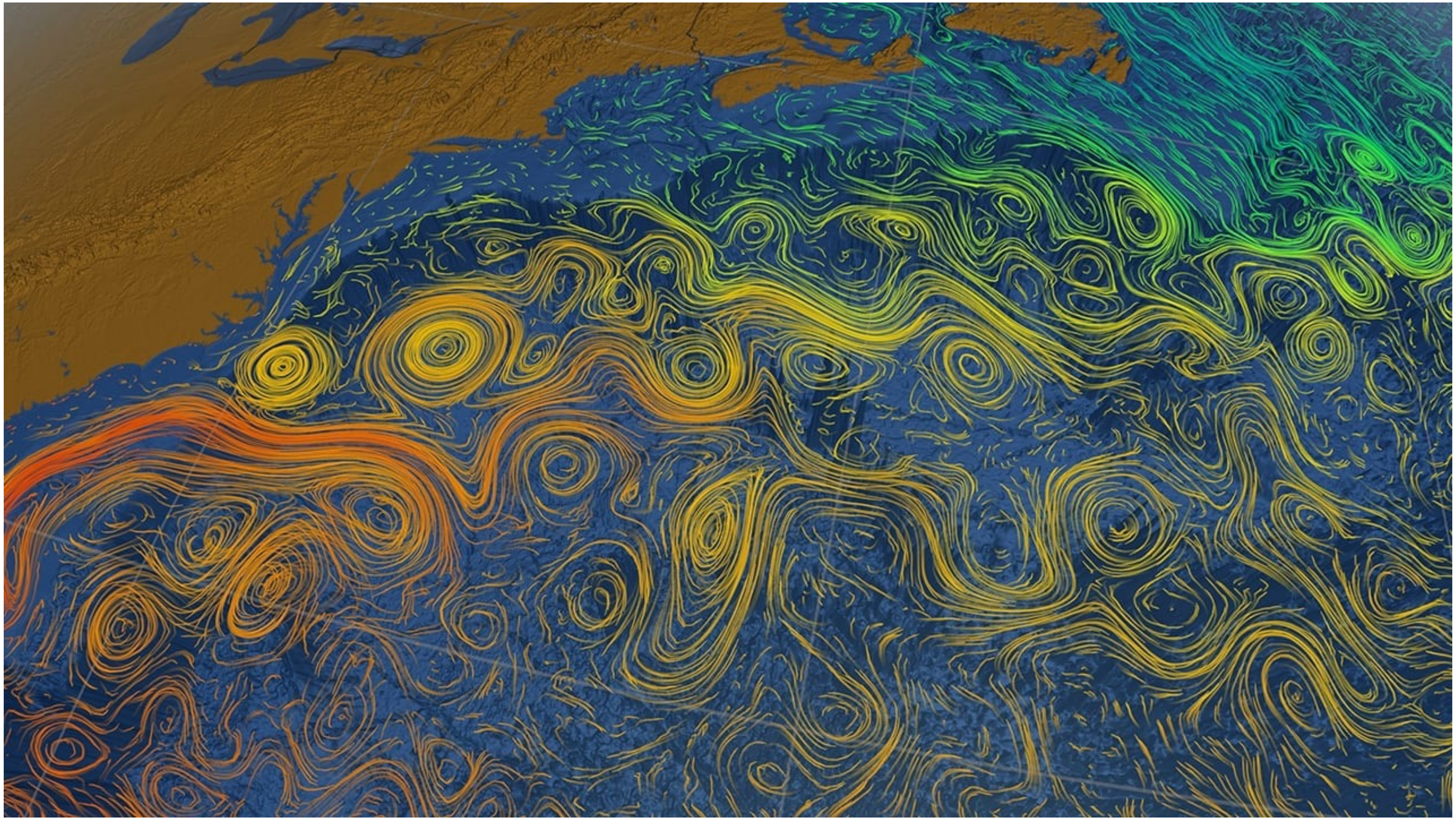


- Technique fondamentale
- Précision limitée par taille de maille
- maille / 2 ➔ ressources x 16 (mémoire, temps de calcul)
- Au lieu de réduire la maille, entrainer un réseau de neurone à reproduire un modèle à petite maille

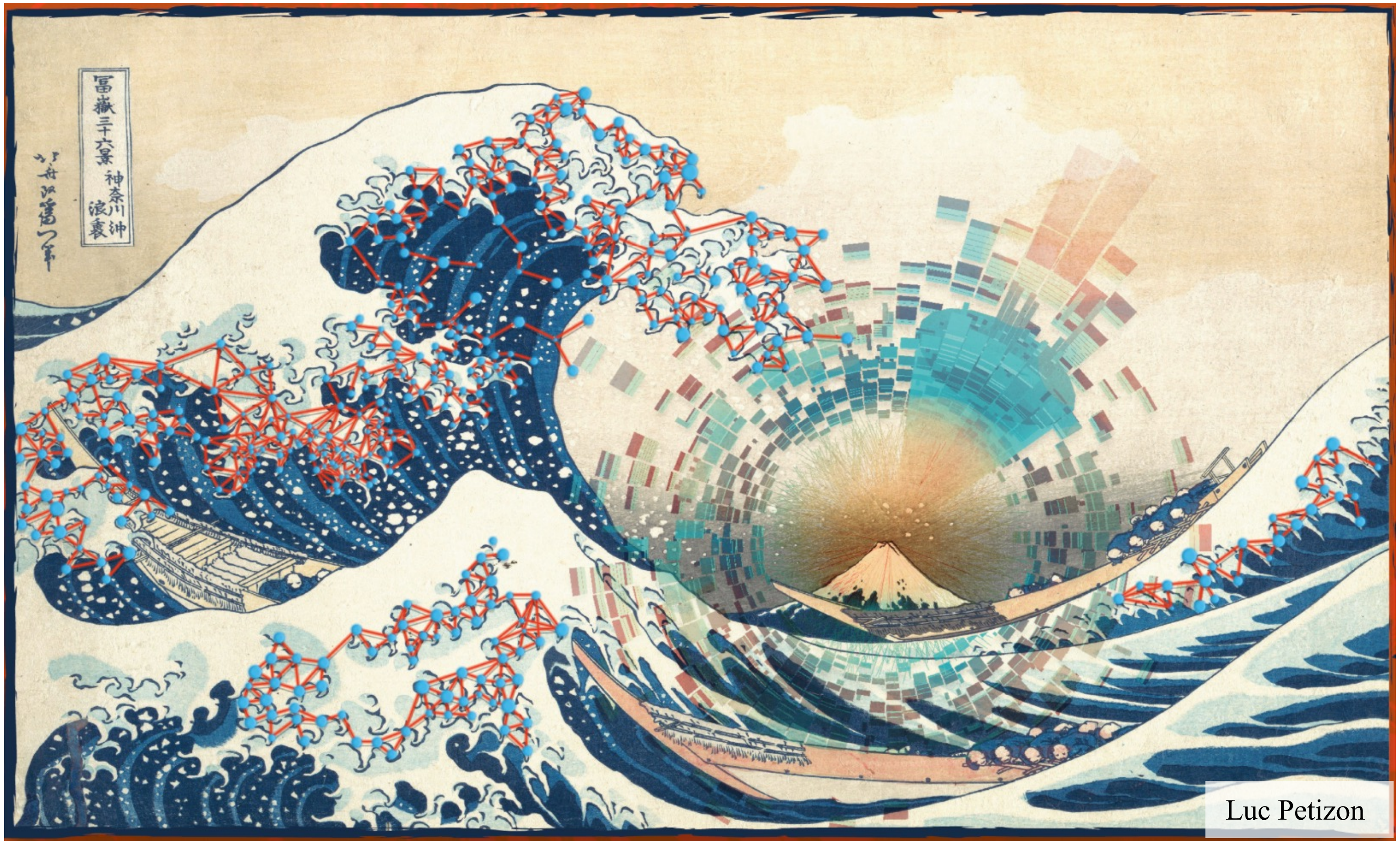


Application océanographie

NASA



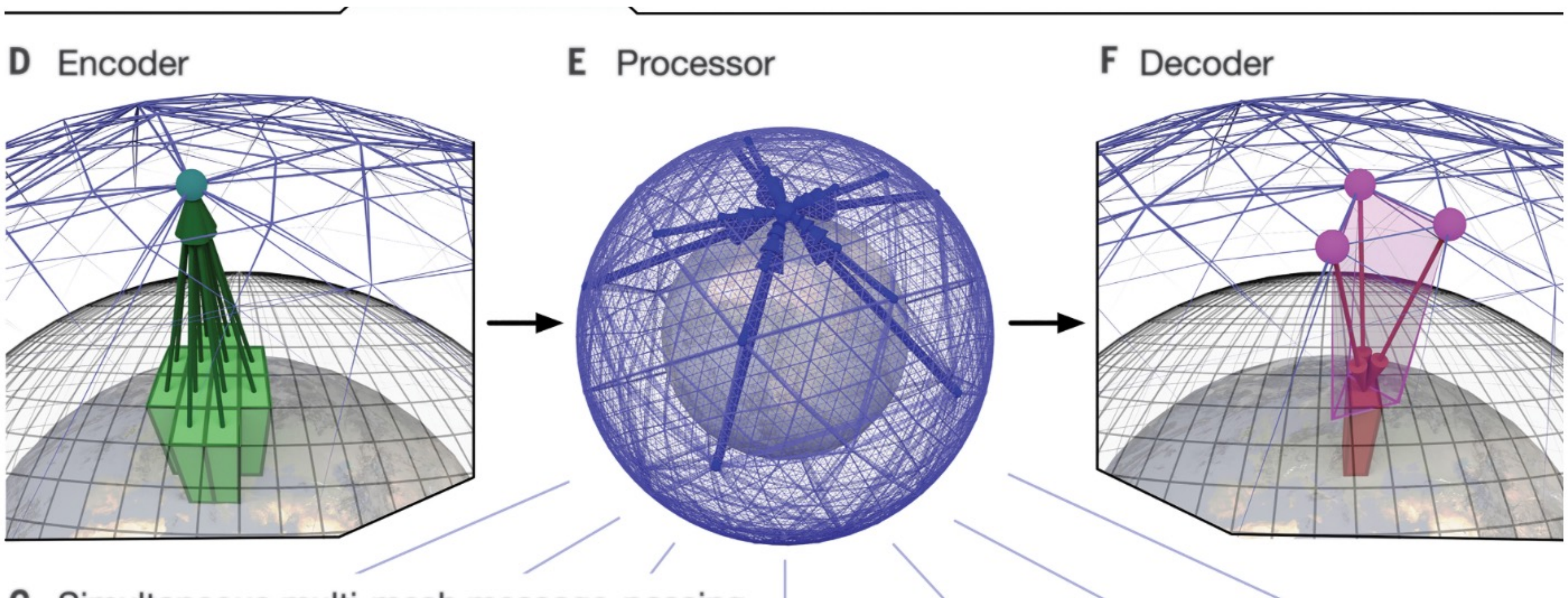
Modèliser l'inconnu



Si un phénomène n'est pas calculable, entrainer un réseau de neurones sur des données expérimentales pour l'émuler

Ex: étude vague scélérate Hokusai

Modèle Météo GraphCast Deepmind

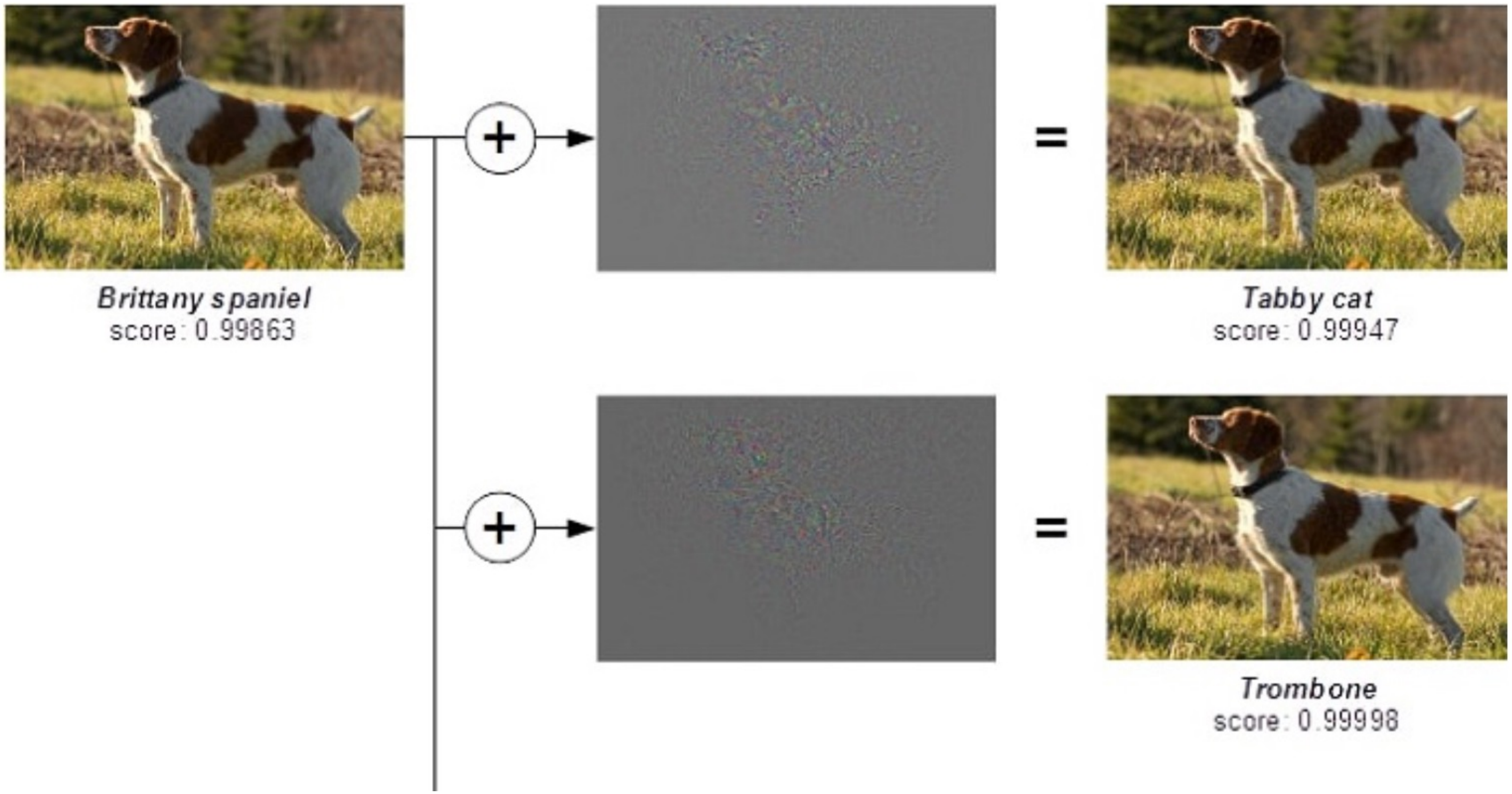


Prédiction à 10 jours en qq minutes, plus fiables que modèles traditionnels
(mais pas d' »assimilation »)

Intermezzo on Adversarial examples

Adversarial examples

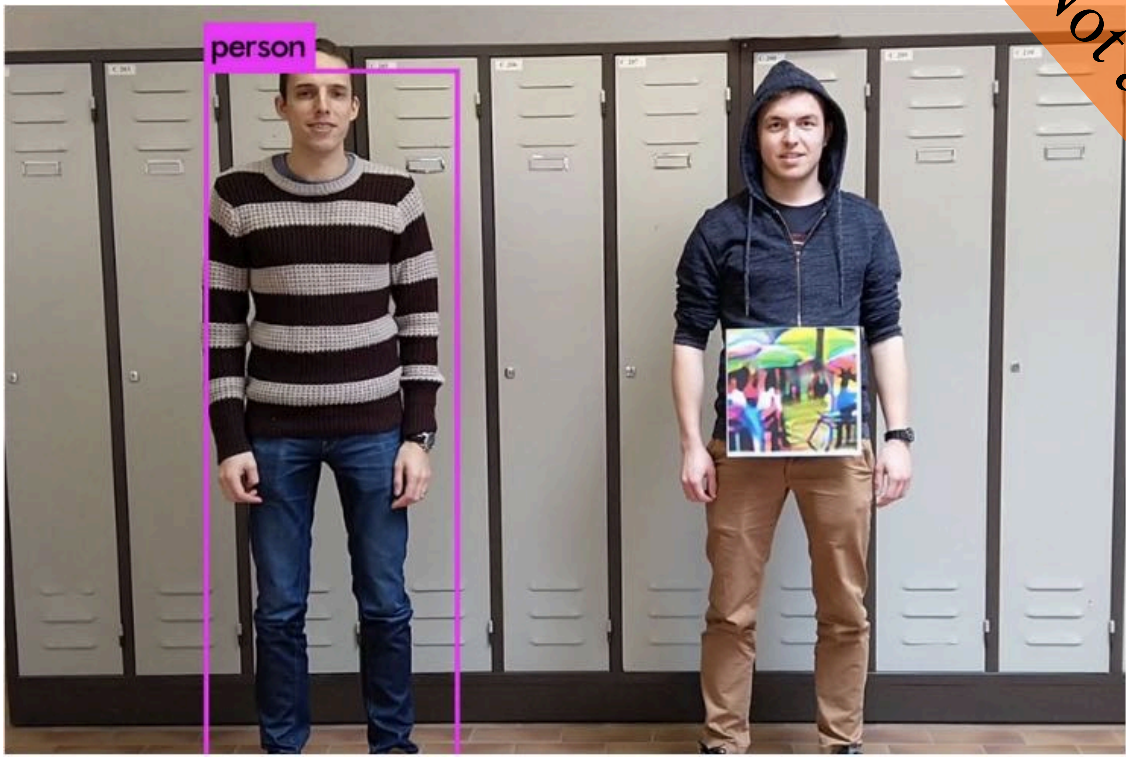
- Subtle non random alteration of an image fooling a classifier



Adversarial examples (2)

the verge 2019/04/23 (see also video)

- Extraneous object
- ➔more worrying, for HEP it would be e.g. a glitch in the data which is not simulated



Not a problem for physics

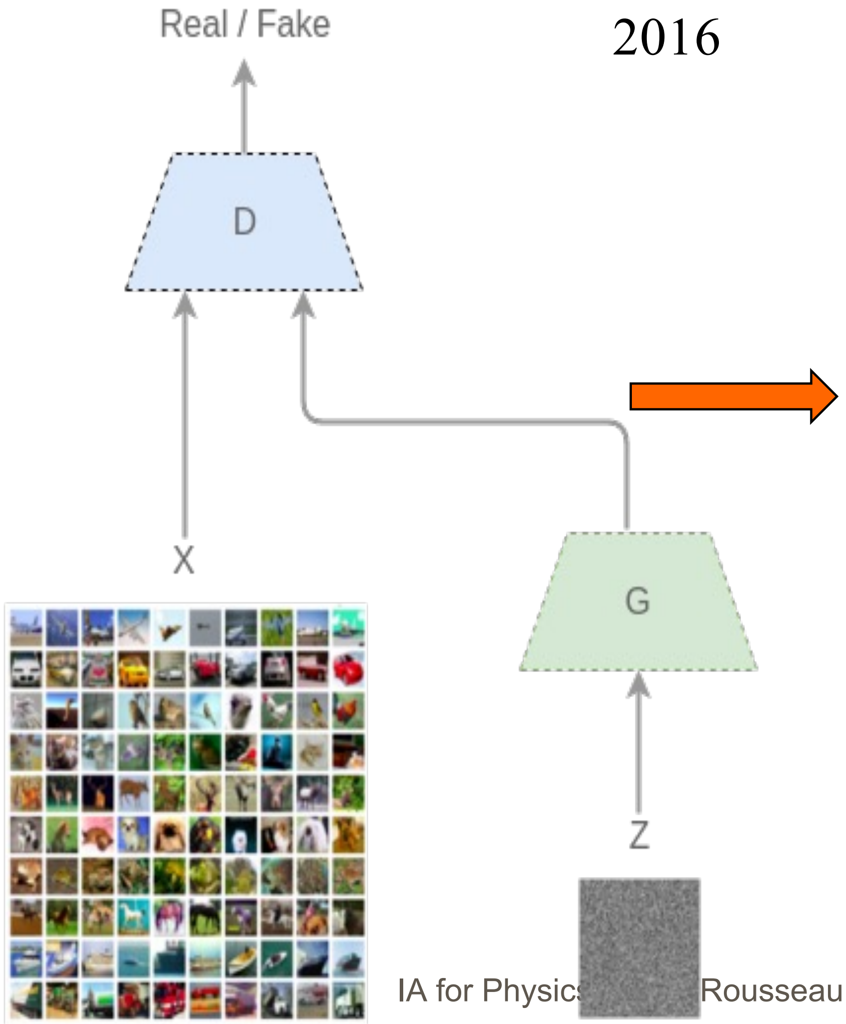
Adversarial examples

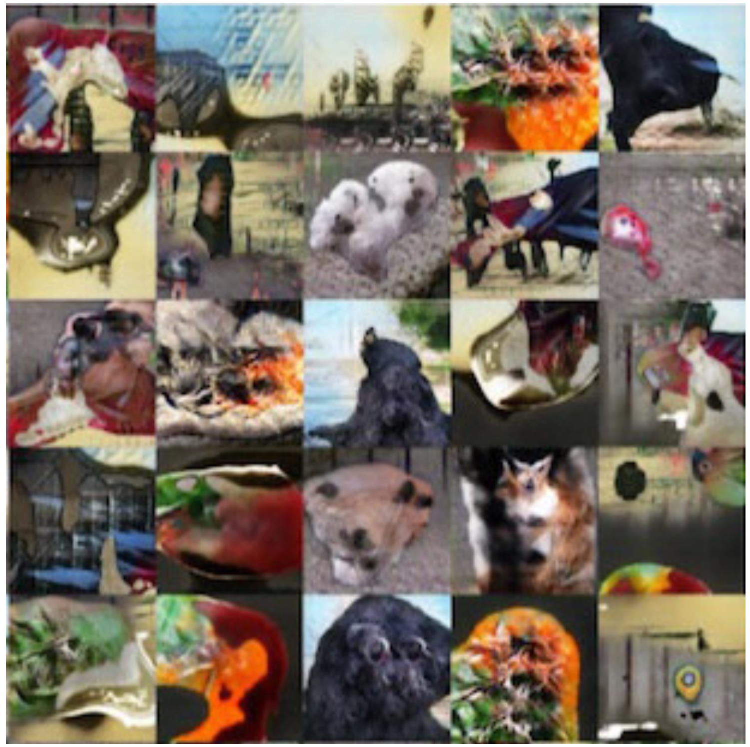
- Like an optical illusion for a NN
- Tuned for a specific known NN ➔ can still fool other NN (share a common behavior)
- Dangerous for physics if we rely more and more on NN ?
- Definition : « inputs that are specifically designed to cause the target model to produce erroneous outputs »
- we could not think of a case a detector or physics glitch would fulfill the definition above

Modèles génératifs

Generative Adversarial Network

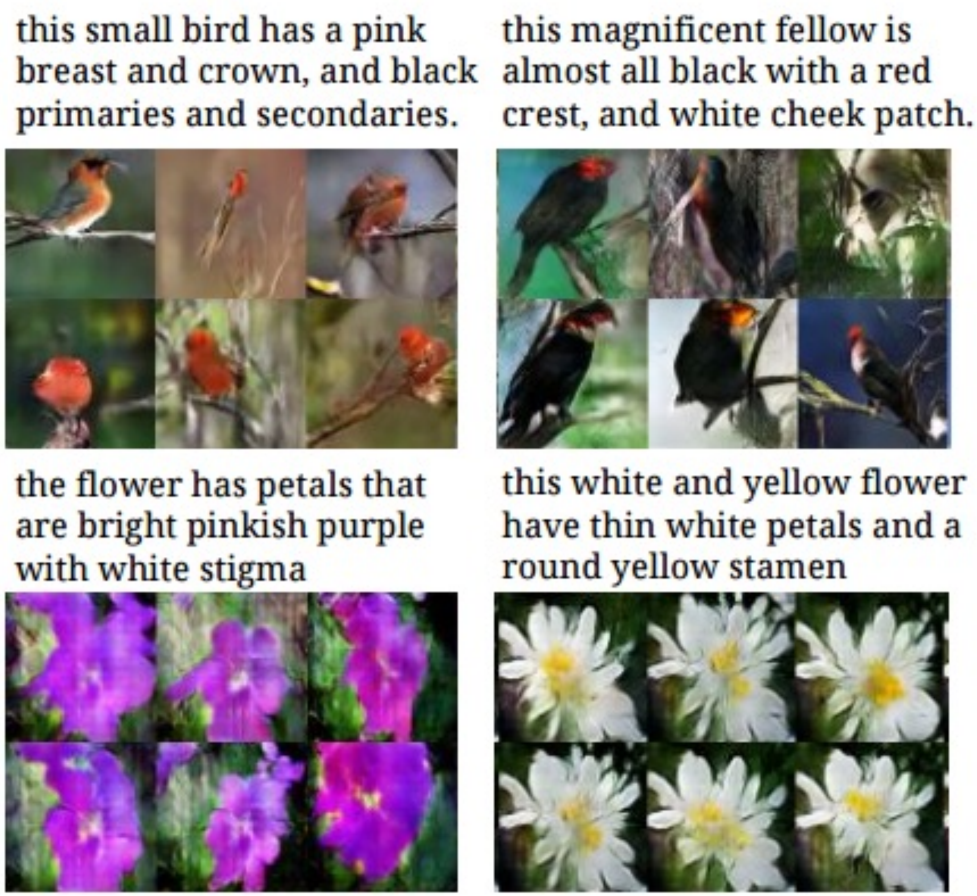
2016



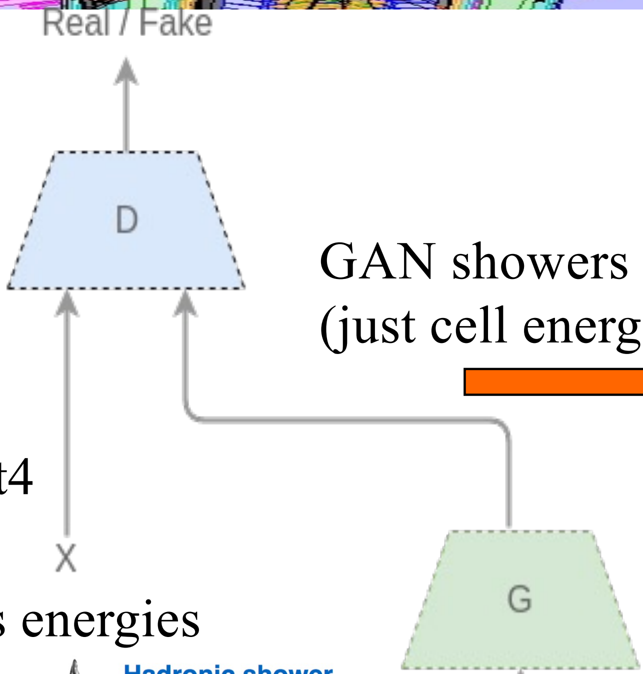


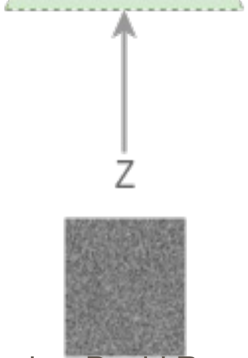
Condition GAN

Text to image



GAN for simulation

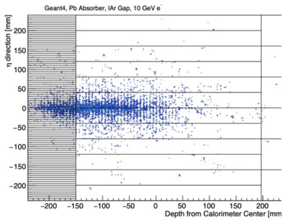


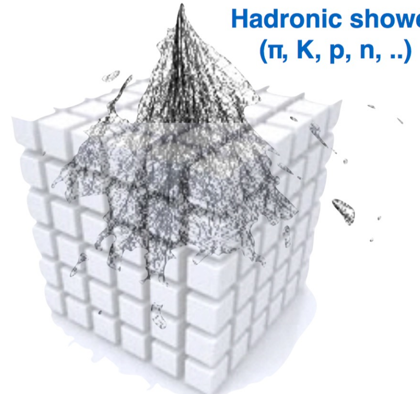


GAN showers
(just cell energies)

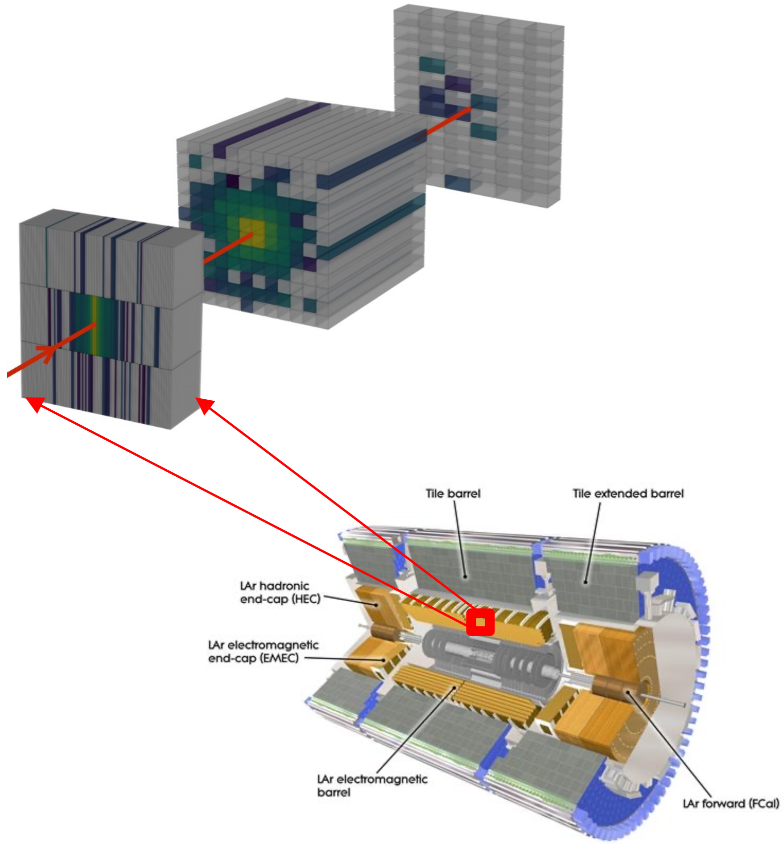
Geant4

Cells energies



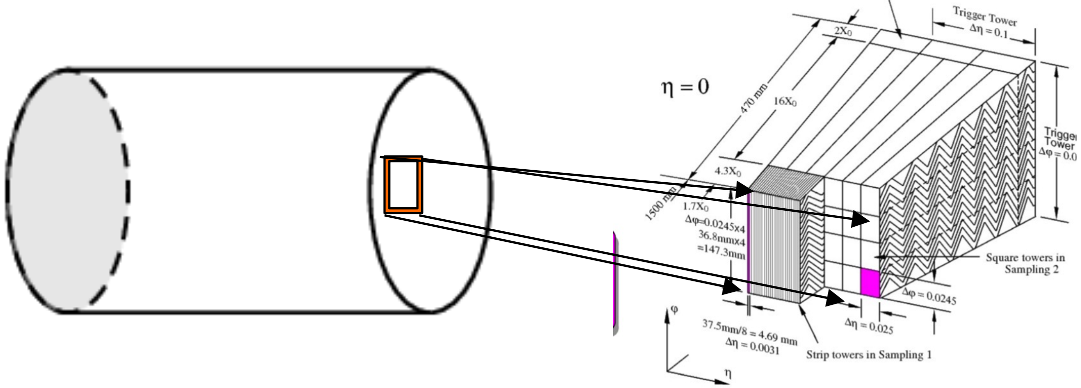


Hadronic shower
(π, K, p, n, ..)



ATLAS calo simulation

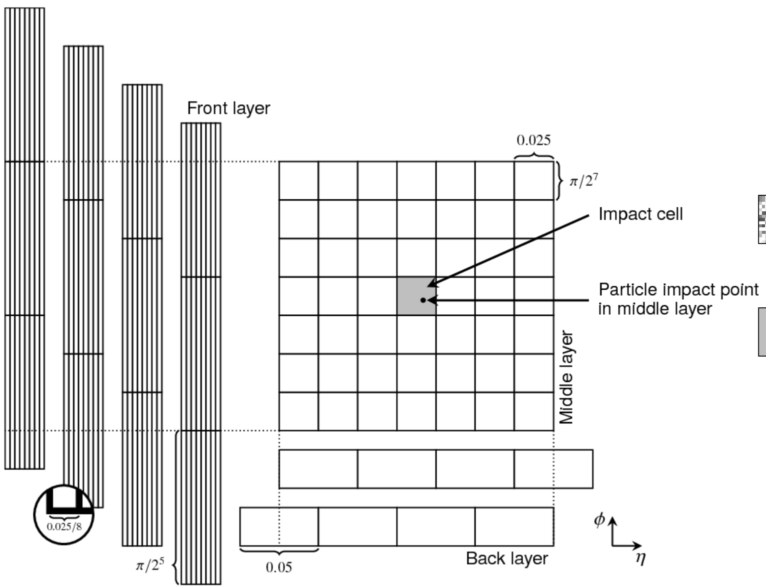
ATLAS Collaboration, [arXiv:1712.10321]

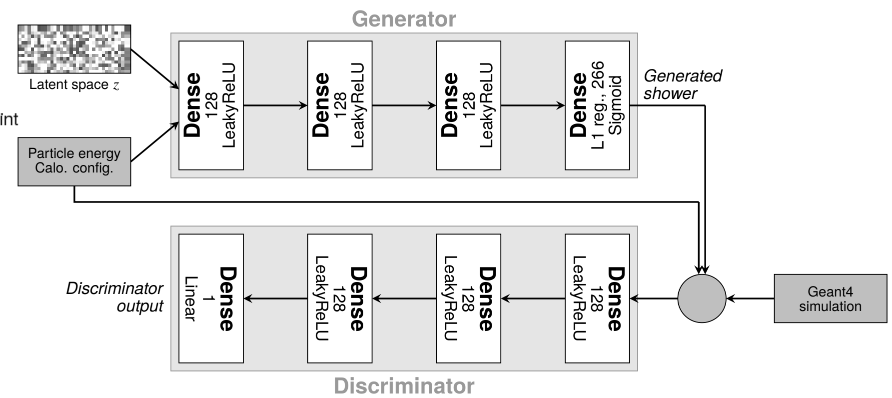


With major contribution from Gilles Louppe U Liège

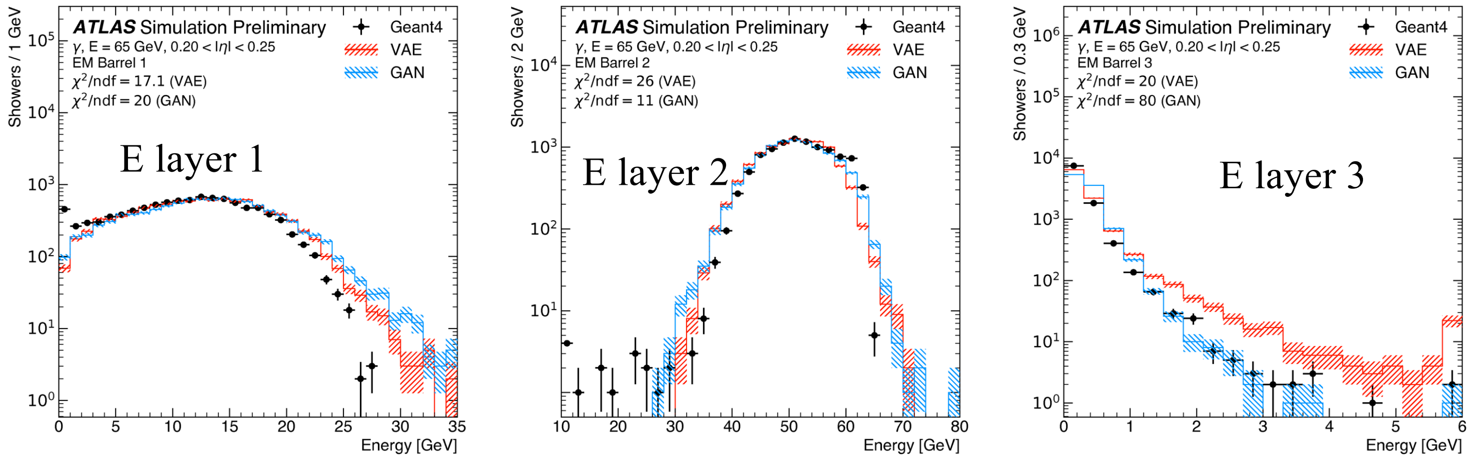
+ η, ϕ translation

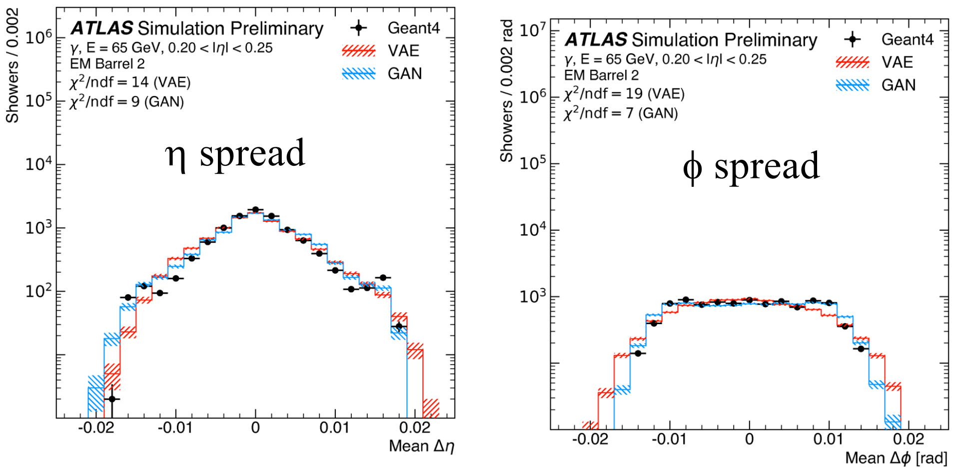
177000 cells ➔266 cells





Results

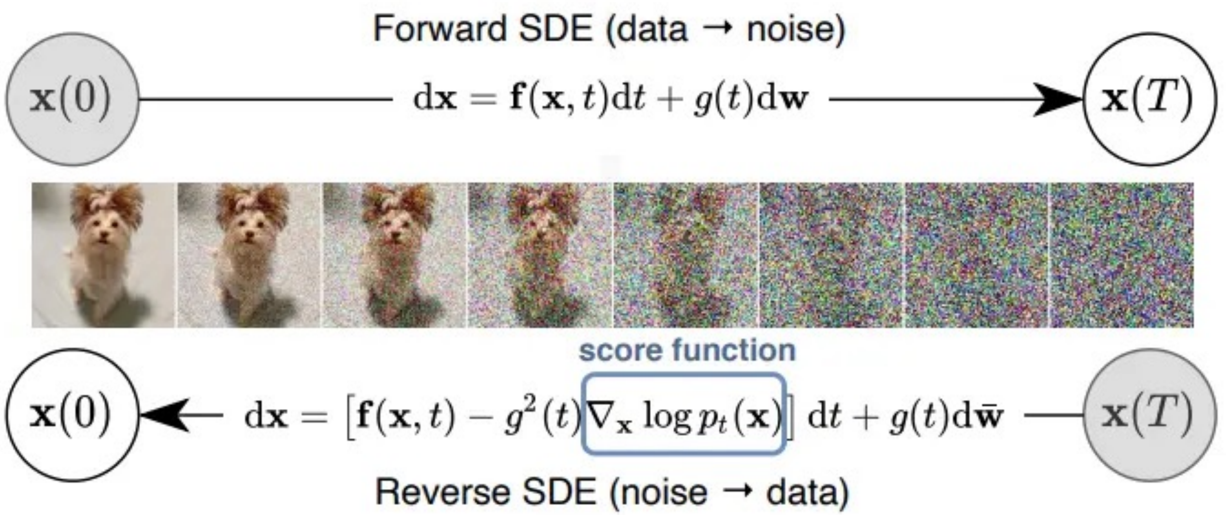




Works reasonably well,
but not good enough
Training very tricky

So much progress since 2016

See nice introduction #Dall-e Diffusion



an espresso machine that makes coffee from human souls, artstation

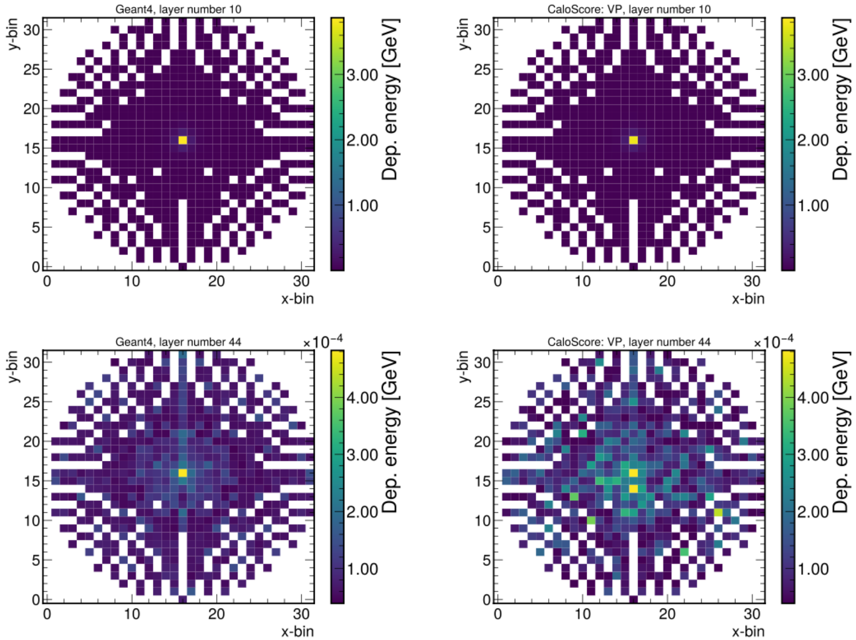
panda mad scientist mixing sparkling chemicals, artstation

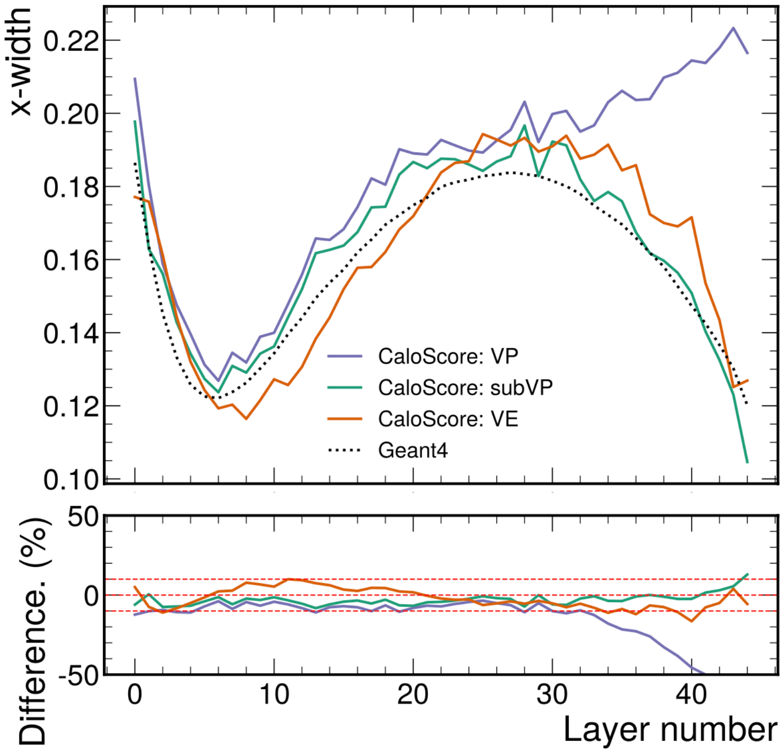
GAN becoming obsolete, can we do something (serious) with diffusion model ?

Simulation with diffusion model

Mikuli, Nachman

- On CaloChallenge dataset
- ➔seems to work, claim it is easier to train
- However inference (=generation) slower

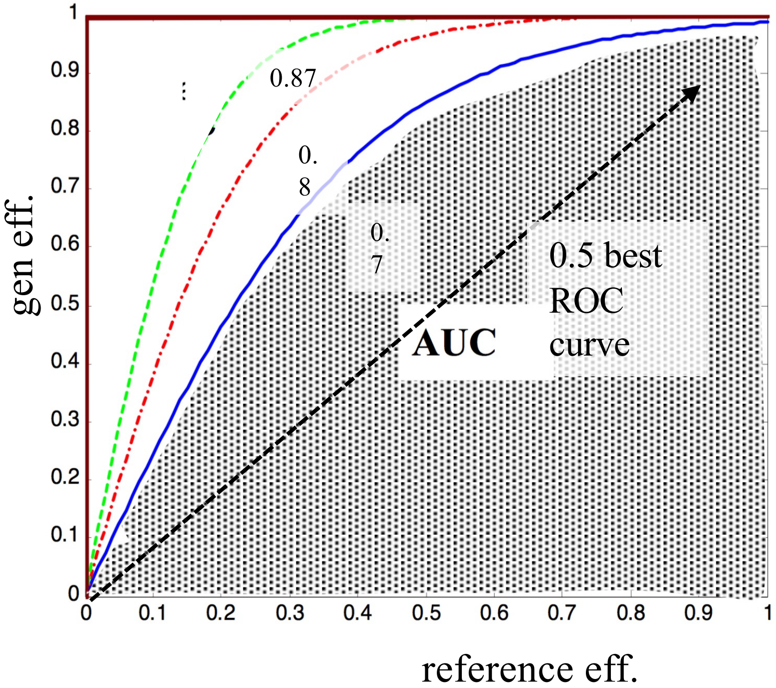




How to evaluate a generative model

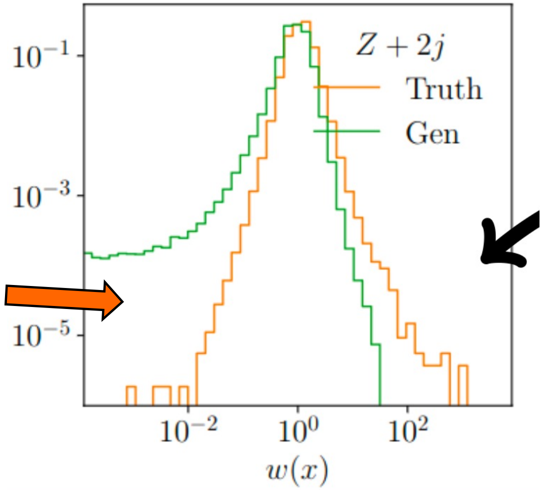
- For text or image generation, subjective human evaluation « by eye »
- For science we should be more quantitative:
- train a classifier to distinguish the generative model from reference=truth :

AUC : Area Under the (ROC) Curve



$$w = \frac{p_{gen}}{p_{truth}}$$ with calibrated classifier

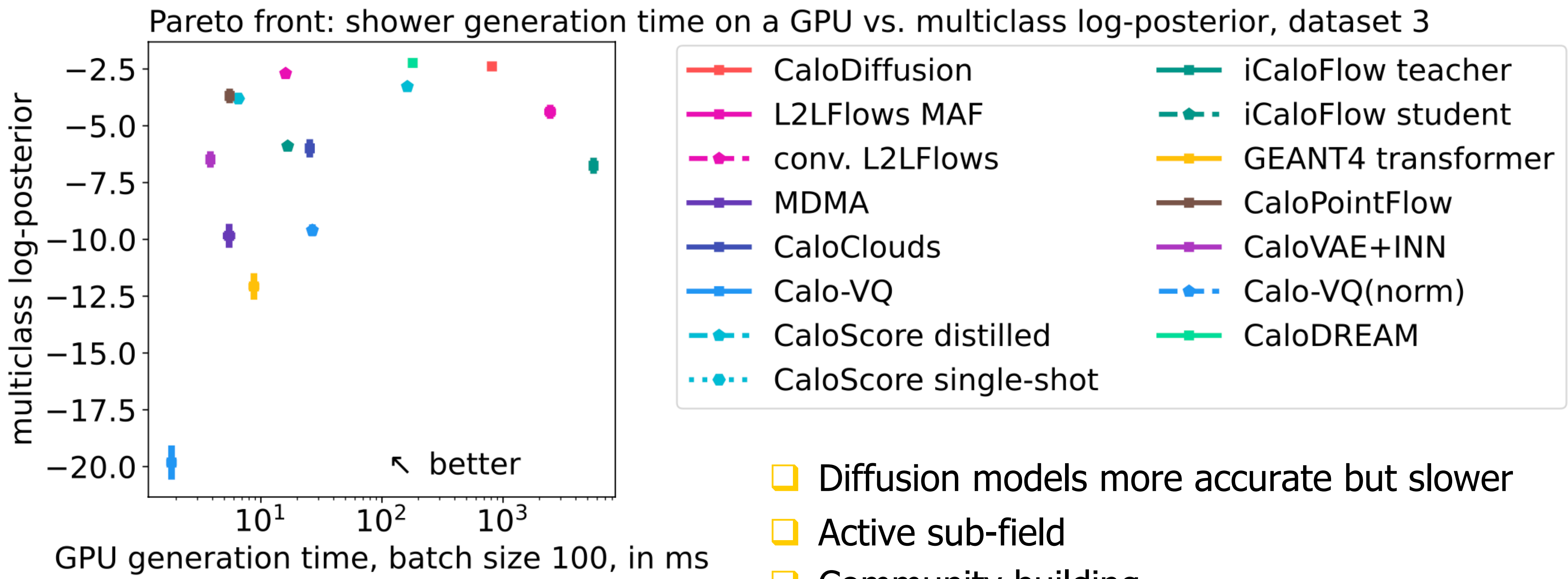
mis-modelled feature



missed feature

CaloChallenge results

C. Krause et al, CaloChallenge 2022 final paper arXiv:2410.21611

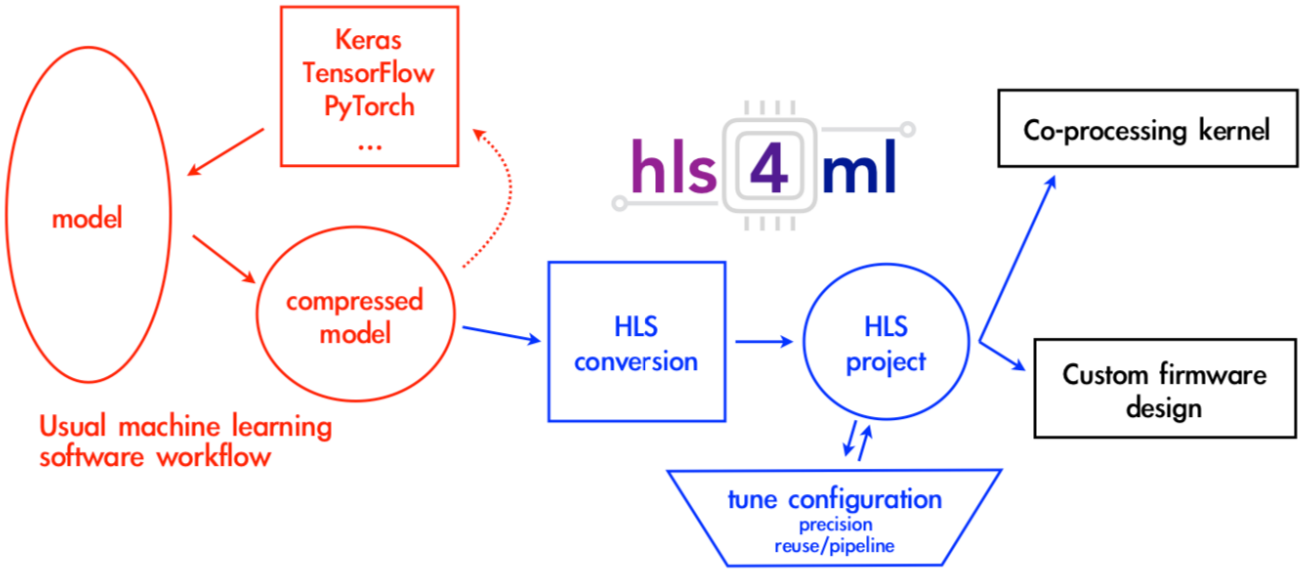


- Diffusion models more accurate but slower
- Active sub-field
- Community building

Fast Machine Learning

ML for DAQ/trigger

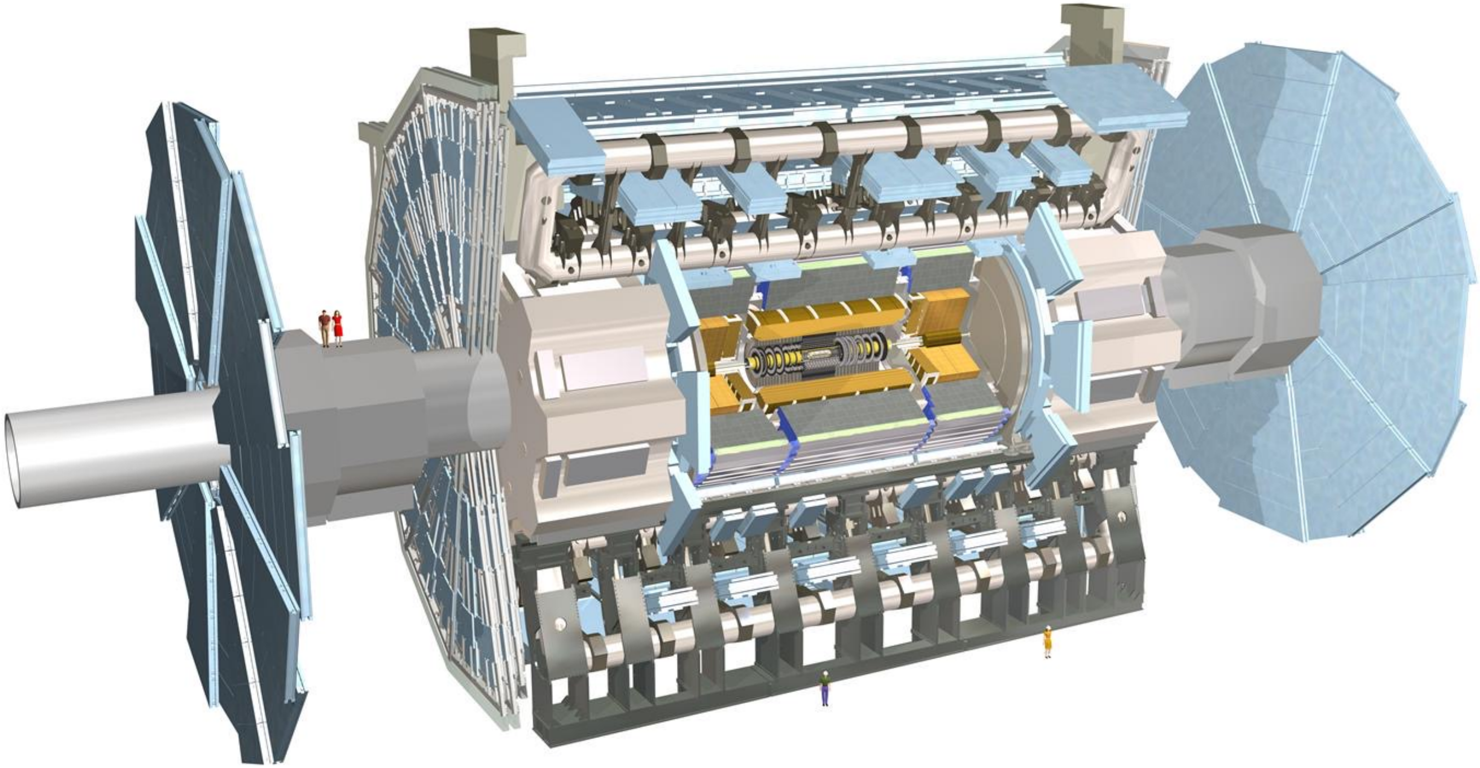
- Growing use of ML in trigger/readout
- Classifiers (BDT or NN) maybe long to train but evaluation is very fast
- Can be implemented on GPU or FPGA
  - GBDT on FPGA
  - NN on FPGA : hls4ML (arXiv:1804.06913)
- Manpower efficient, few experts can code directly for GPU or FPGA. However experts are still better (see CPPM work for ATLAS LArg arXiv:2510.11469)



Experiment design

ATLAS experiment

Currenty taking data at the LHC, designed in 1992



- Currently : we are developping AI techniques to analyse the data of current/near future experiments
- Future : use AI to design new experiments

Topological Design Optimisation

- Whole field of industry, boosted by 3D-printing capabilities, and by AI
- Given geometrical, stiffness and feasability constraints, how to optimise cost ?
- ➔future experiments will be designed with AI, cf MODE collaboration

Incertitudes

Phys.Rev.Lett. 114 (2015)191803

Peut-être la mesure la plus complexe

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

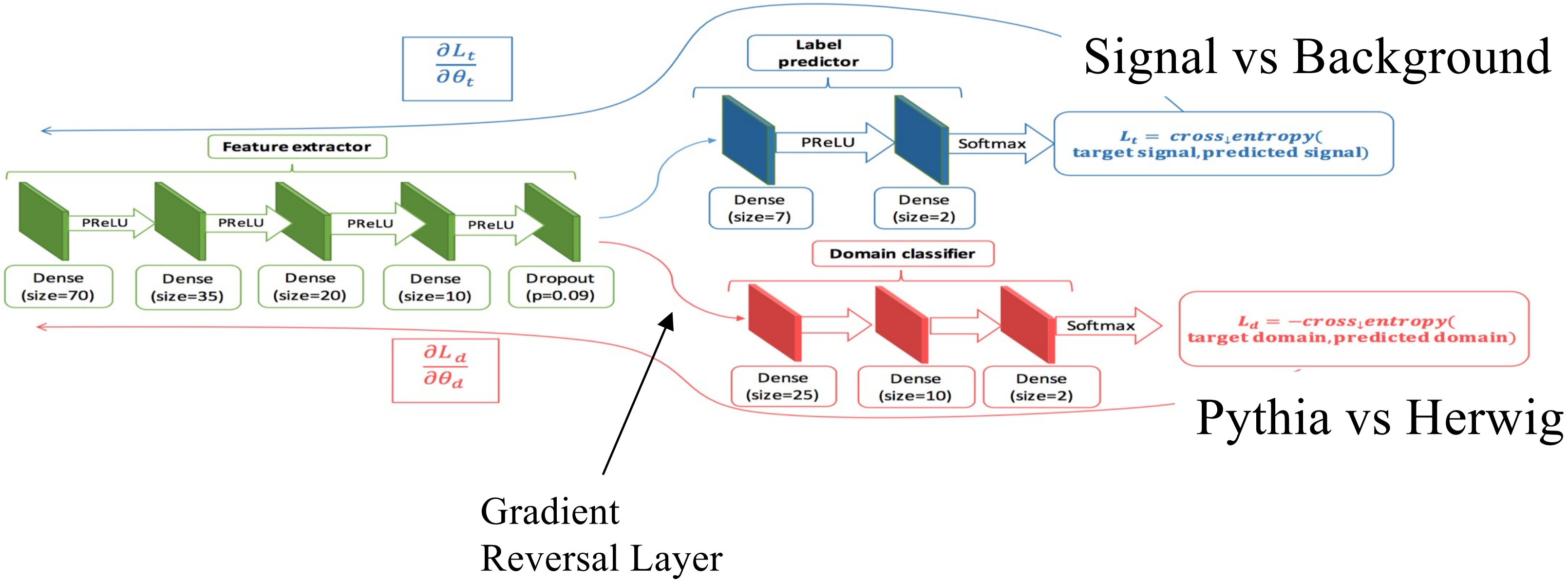
(ATLAS Collaboration)[†]
(CMS Collaboration)[‡]

(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 $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ 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 $m_H = 125.09 \pm 0.21\,(\mathrm{stat}) \pm 0.11\,(\mathrm{syst})$ GeV.

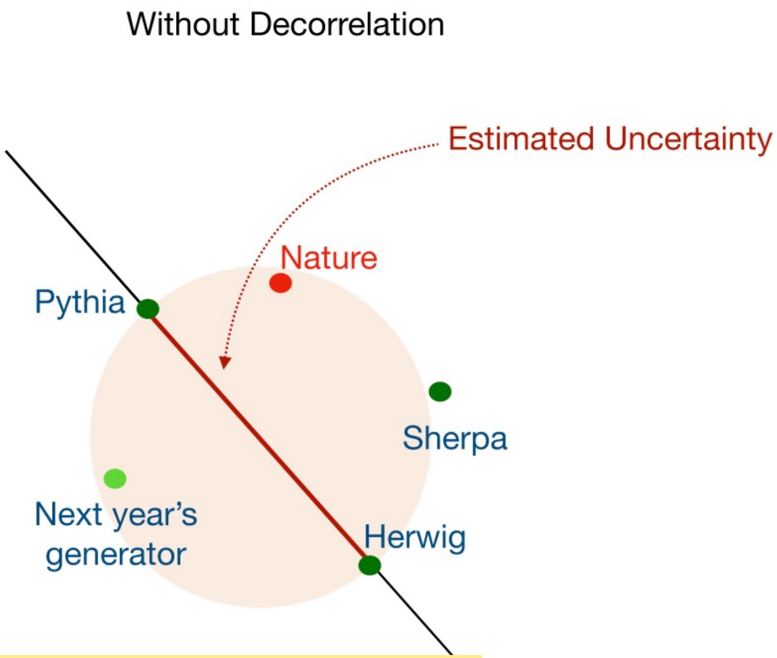
Réd. d’incertitude par entrainement adversaire

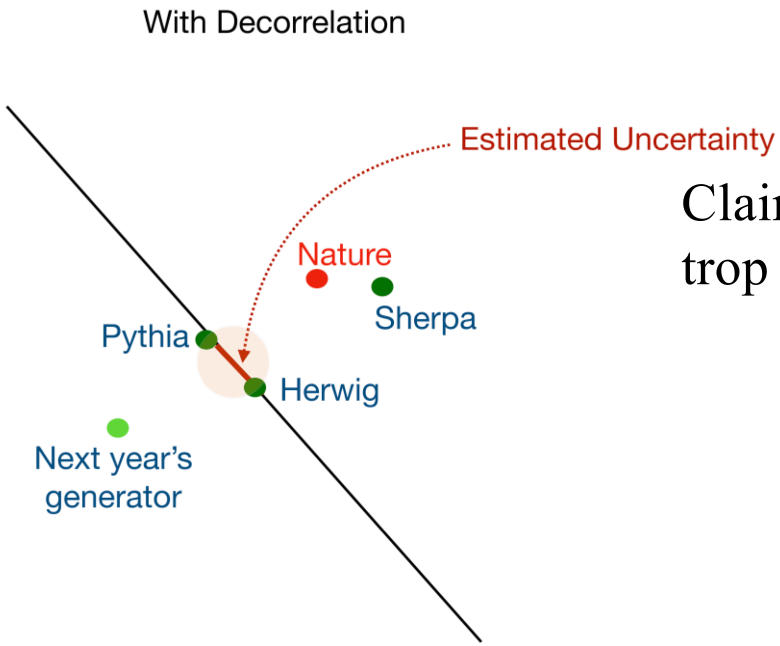
Inspired from 1505.07818 Ganin et al :



Les idées simples ne marchent pas

Ghosh & Nachman EPJC **82** 46 (2022)





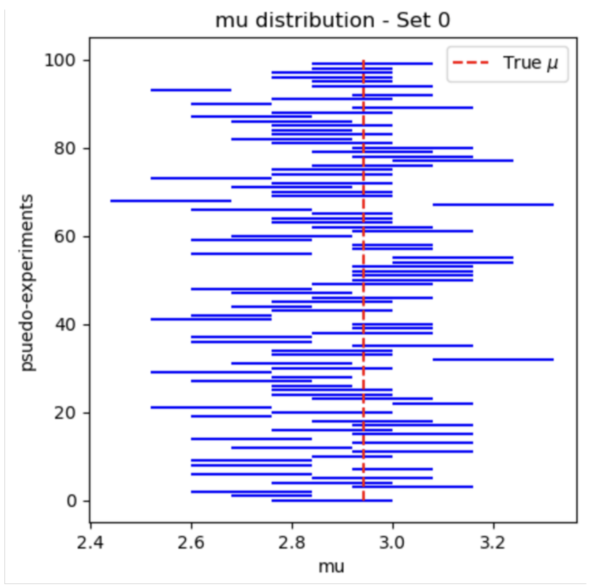
Clairement trop petite !

Contraintes par 70 ans de données de Physique des particules

Loi de Goodhart "Quand un indicateur devient un objectif, ce n'est plus un bon indicateur."

Fair Universe

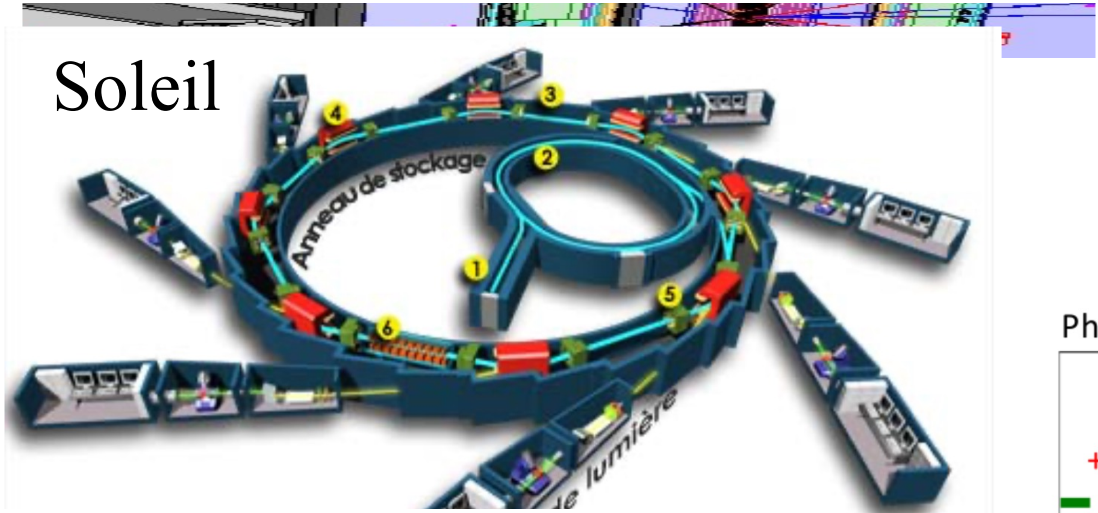
- Une competition pour un modèle évaluant un Intervalle de Confiance
- https://fair-universe.lbl.gov

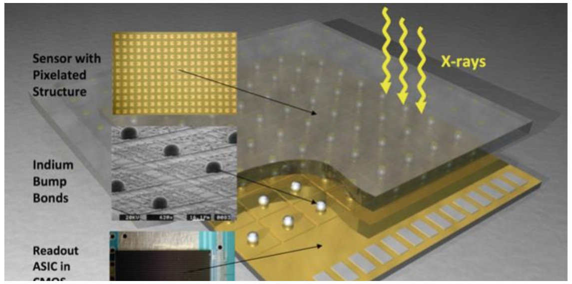


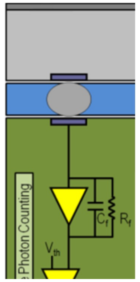
Task:				Fact Sheet Answers	Higgs Uncertainty Challenge			
#	Participant	Entries	Date of last entry	Method Name	Quantile Score	Interval	Coverage	Detailed Results
1	ragansu	30	2024-01-22	Histogram_10	**1.45**	0.226	0.57	
2	ragansu	30	2024-01-22	One_bin NLL	**1.07**	0.333	0.57	
3	laurensslu	20	2023-12-01	cheat7	**0.68**	0.504	0.63	
4	laurensslu	20	2023-12-01	cheat7	**0.61**	0.544	0.68	
5	laurensslu	20	2023-12-01	cheat4	**0.31**	0.732	0.61	
6	laurensslu	20	2023-12-01	cheat4	**0.16**	0.852	0.71	
7	laurensslu	20	2023-12-01	Cheat2	**-0.44**	1.55	0.62	
8	laurensslu	20	2023-12-01	Cheat2	**-0.74**	1.375	0.55	
9	ragansu	30	2024-01-22	tes_finder	**-0.95**	1.124	0.54	
10	laurensslu	20	2023-12-01	Cheat2	**-1.59**	1.325	0.53	
11	Ihsan Ullah	4	2024-01-18	Sascha sys aware 8	**-2.69**	0.329	0.47	
12	Rafał Masełek	10	2023-12-01	1binNLL	**-2.9**	1.233	0.5	
13	ihsanchalearn	16	2023-12-18	1 bin NLL	**-2.9**	1.233	0.5	
14	Rafał Masełek	10	2023-12-01	1binNLL	**-2.9**	1.233	0.5	
15	ihsanchalearn	16	2023-12-18	Sascha sys aware 8	**-3.01**	0.33	0.46	

IA peut être très simple

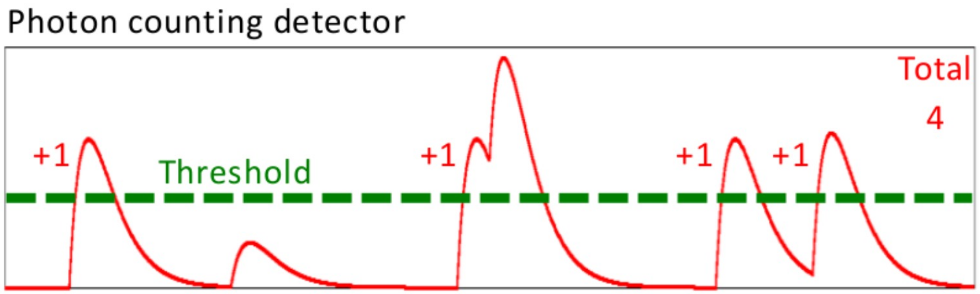
Impact IA »simple »

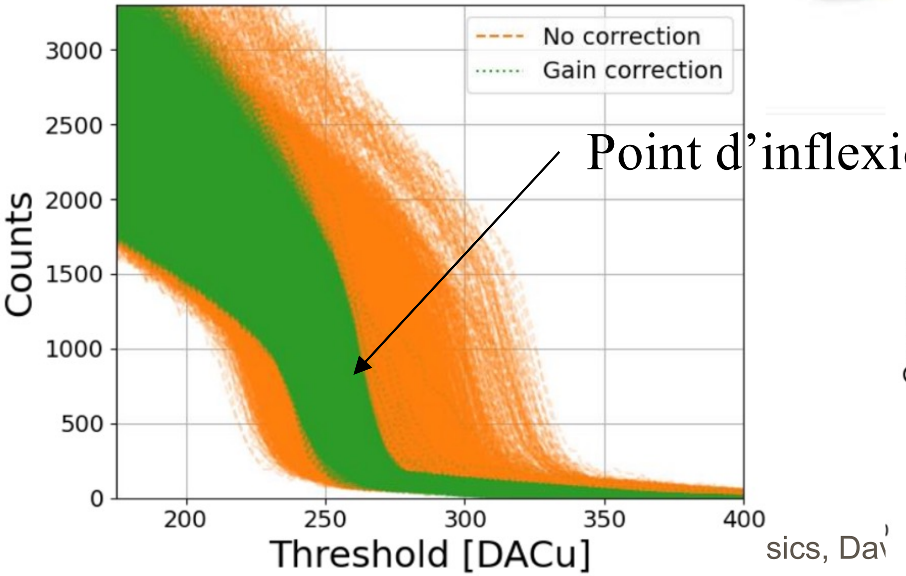


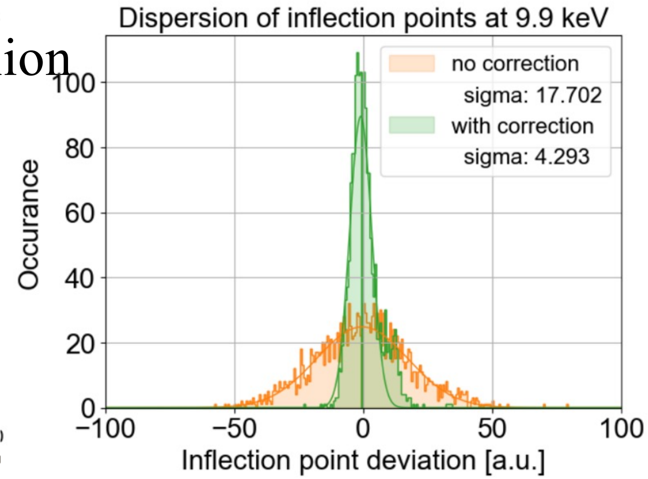




[François Caud (Dataia), Marie Andrä, Martin Chauvin, Arkadiusz Dawiec (SOLEIL)]()







- Calibration « classique » :1 journée
- Random forest : 1/2h