Seminar at INFN-Genoa 2022

MACHINE LEARNING FOR HADRON SPECTROSCOPY

Sunset in Madrid

César Fernández Ramírez

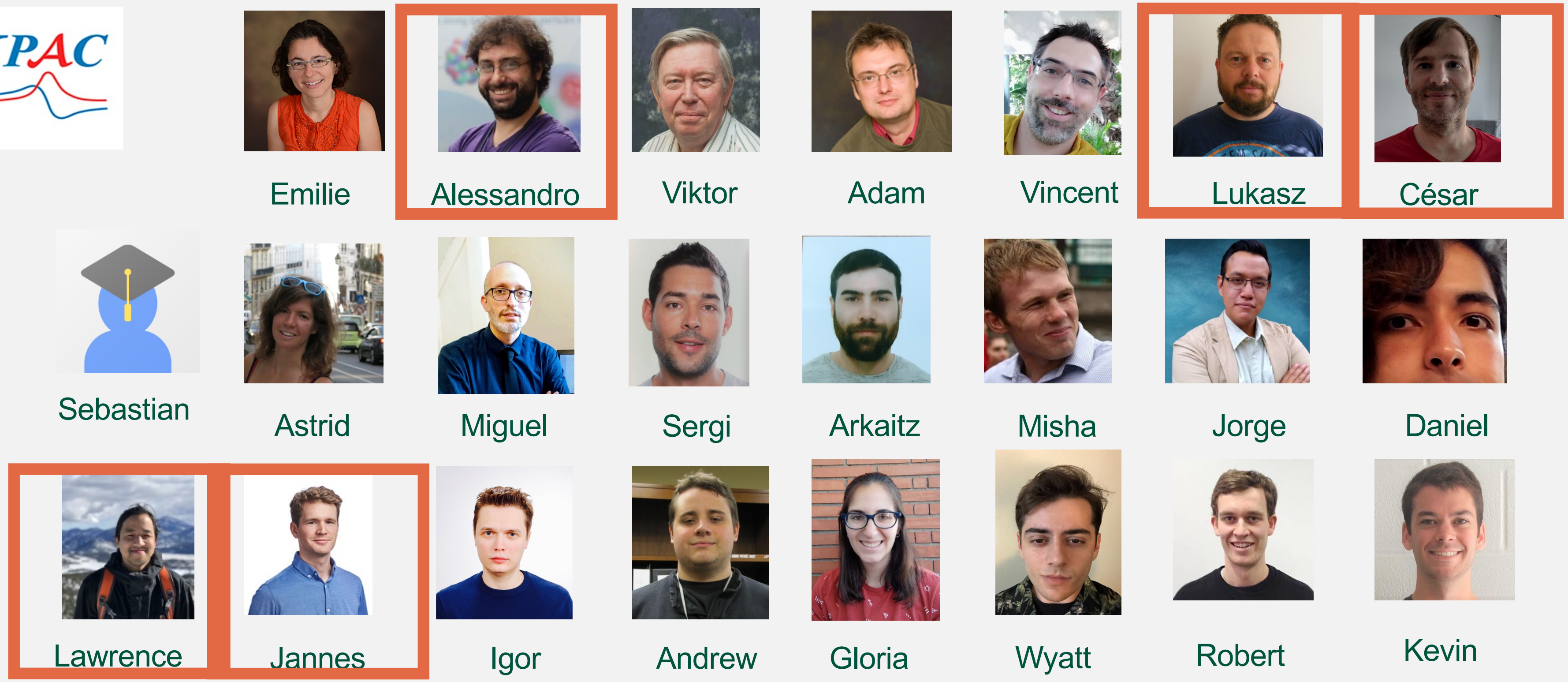
#SOMOS2030 GOBIERNO DE ESPAÑA MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL CONACYT dgapa PAPIIT - PAPIME

JPAC Instituto de Ciencias Nucleares UNAM UNED

Joint Physics Analysis Center

Institutions:

Nationalities:



Outlook

Motivation

- *QCD exotics, huge amount of data coming from experiments*

Linesahape analysis

- *Standard approach, $P_c(4312)$ example*

Neural networks

- *ML/AI, classifiers*

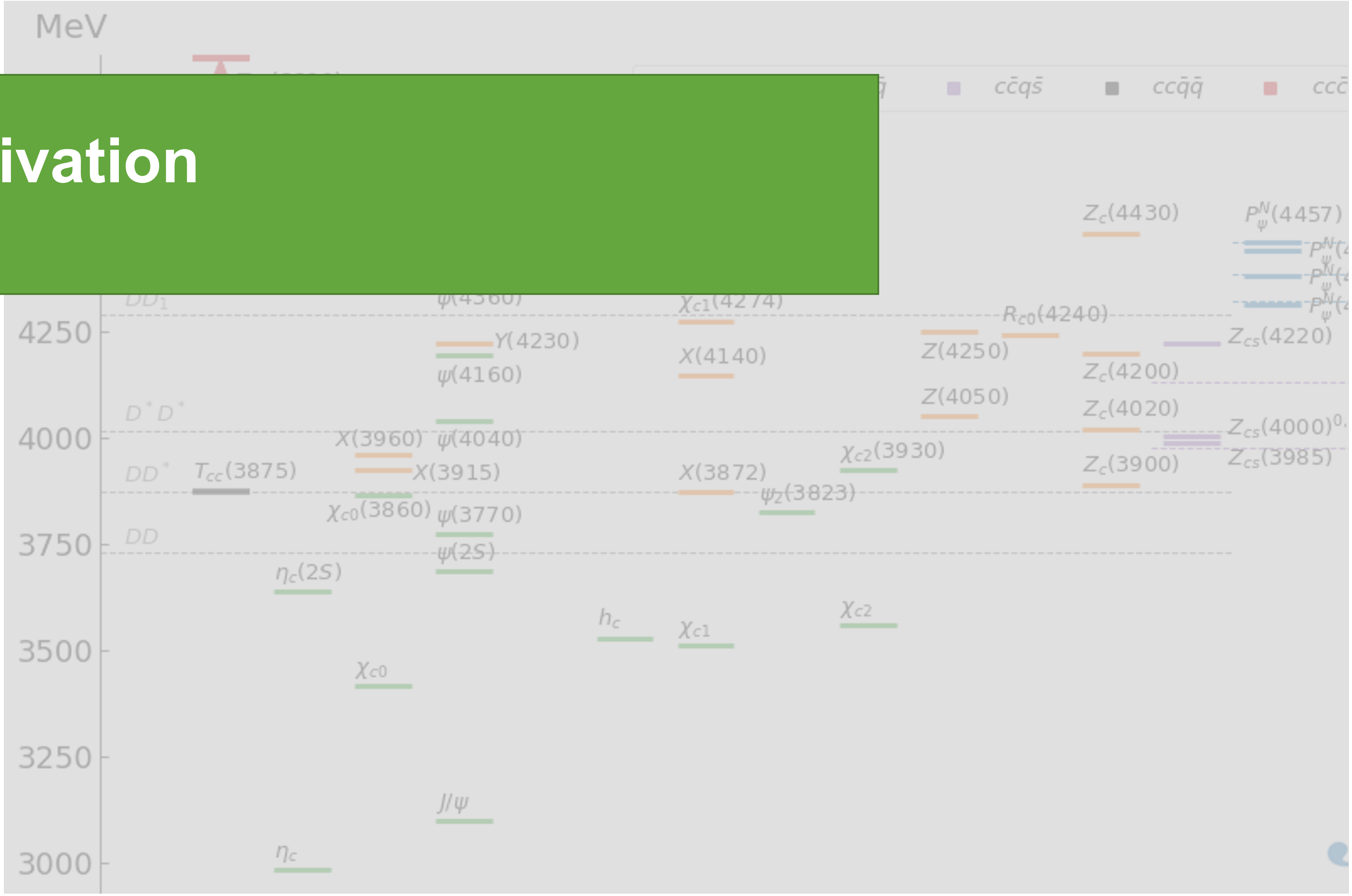
Neural networks applied to hadron spectroscopy

- *Benchmark example: $P_c(4312)$*

Takeaways

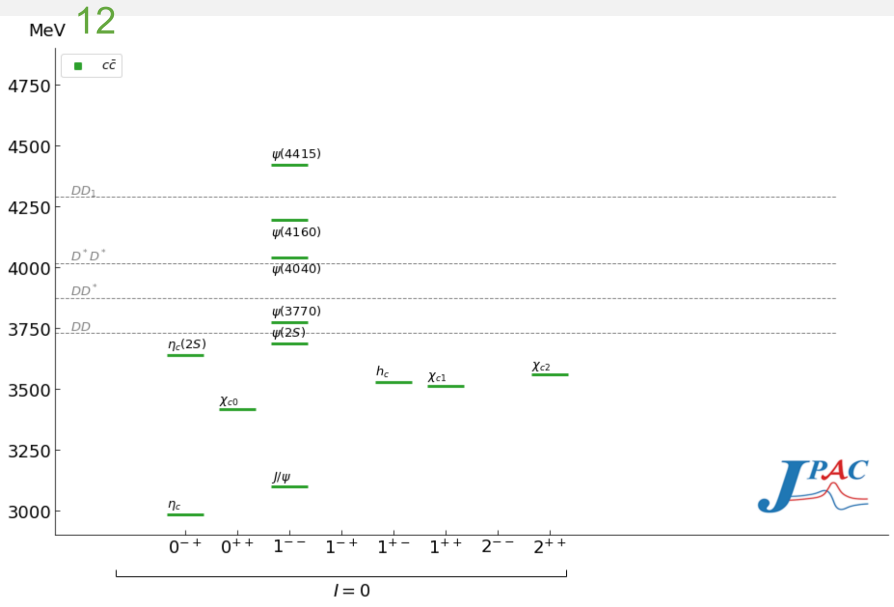
- *Explainability*

Motivation

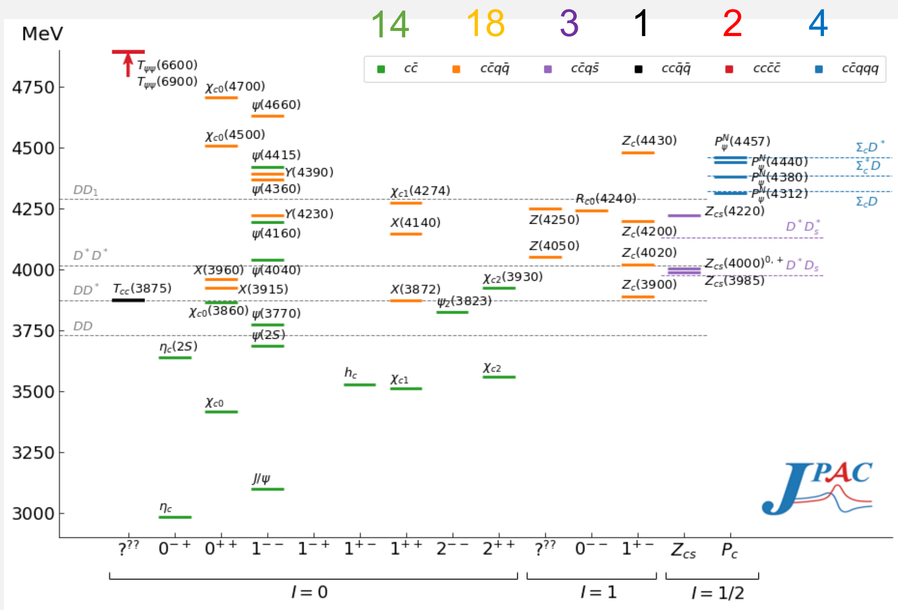


Charmonia(-like)

2003

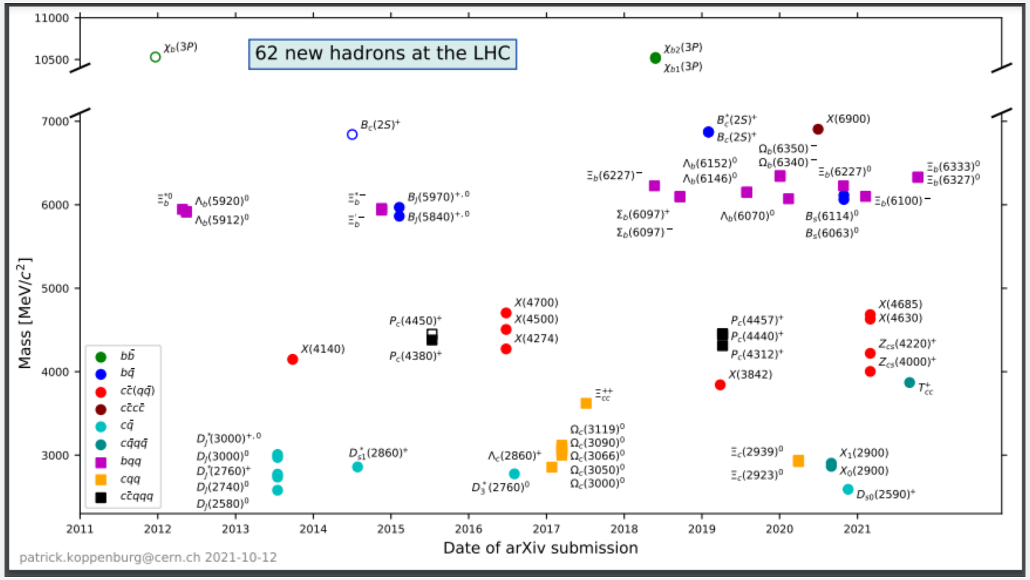


2022

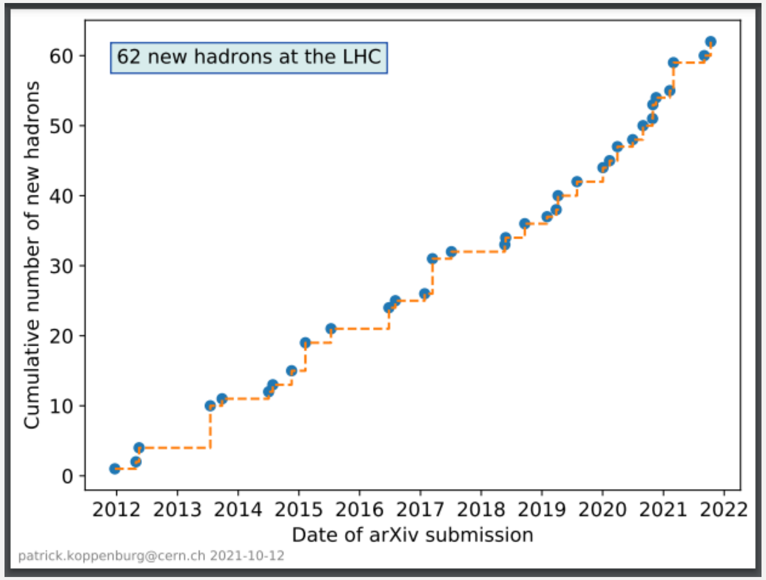


and $P_{\psi s}^{\Lambda}(4338), T_{c\bar{s}0}^{a}(2900)^{0,++}$

Just LHC

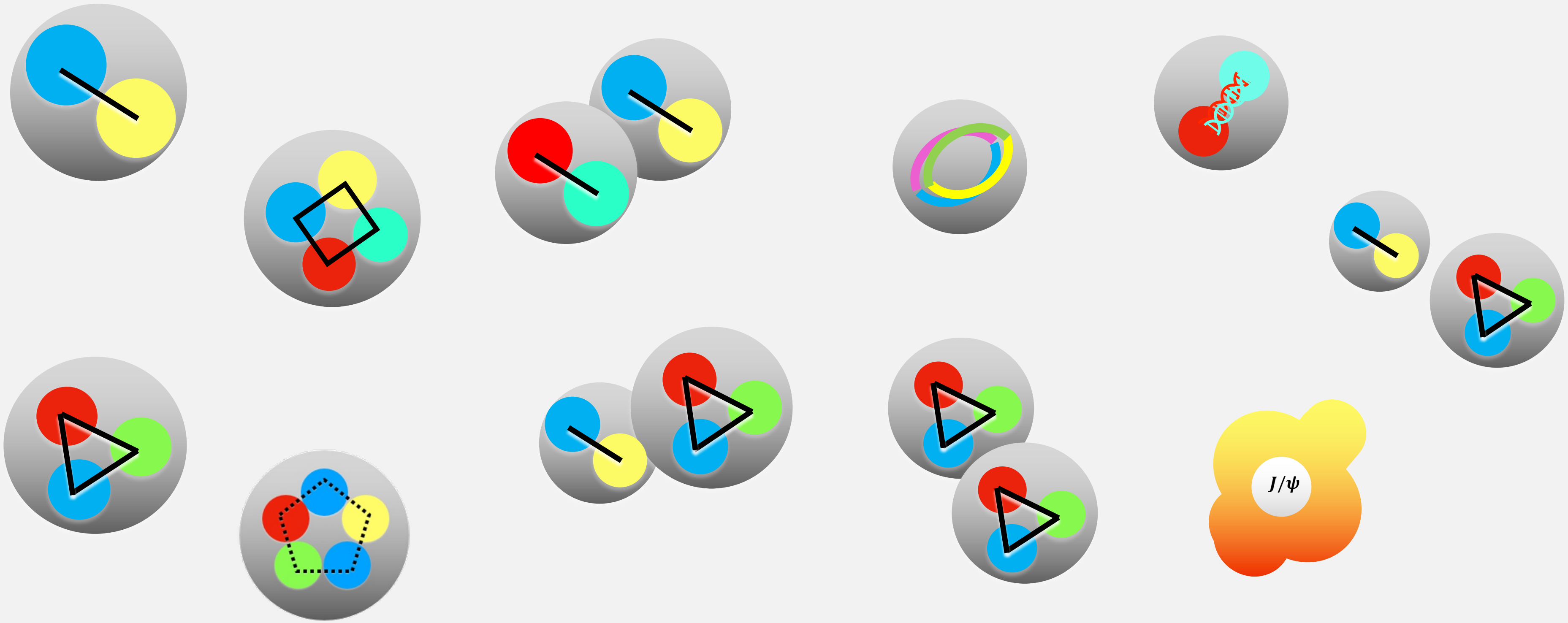


62 new hadrons at the LHC
Mass [MeV/c²]
Date of arXiv submission
patrick.koppenburg@cern.ch 2021-10-12



62 new hadrons at the LHC
Cumulative number of new hadrons
Date of arXiv submission
patrick.koppenburg@cern.ch 2021-10-12

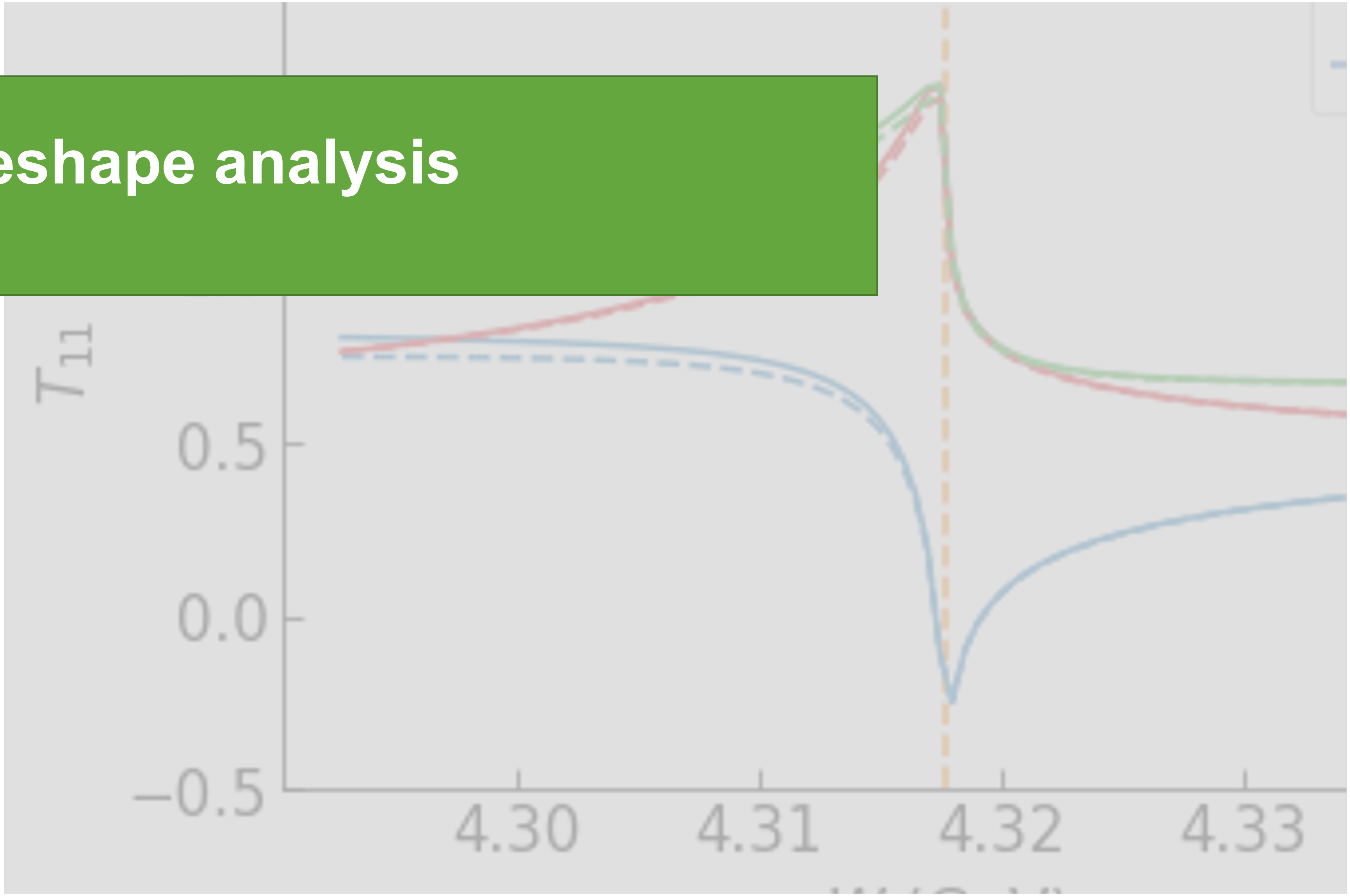
Understand resonances



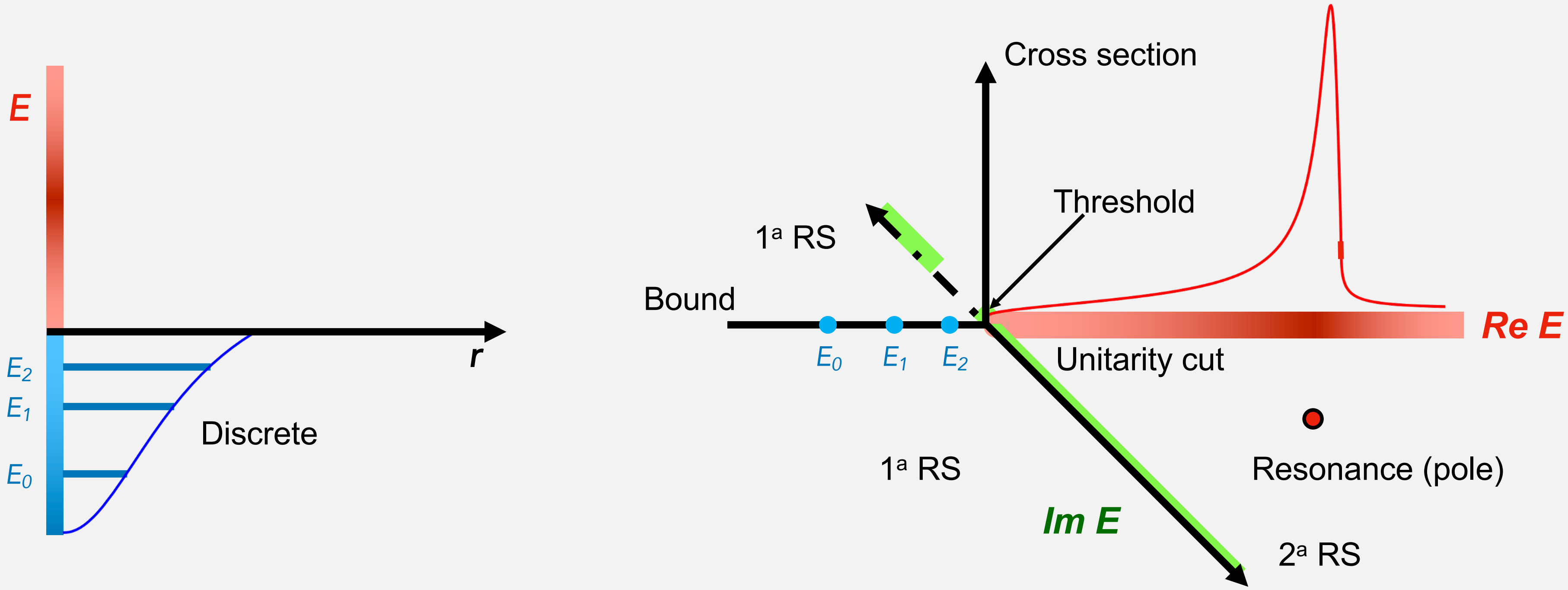

panda
NICA
NUCLOTRON BASED ION COLLIDER FACILITY
ВЭПП 2000
BESIII
mami
MAINZER MIKROTRON
J-PARC
Belle II
Jefferson Lab
clas
GlueX
CERN
COMPASS
LHCb
ATLAS
EXPERIMENT
CMS
ALICE

Theory is trying to catch-up with experiments

Lineshape analysis



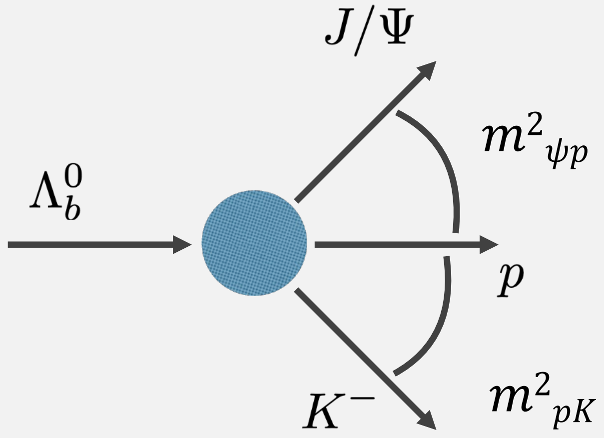
Resonances (aka poles in unphysical Riemann sheets)



Dalitz plot

Data from

LHCb, 1904.03947



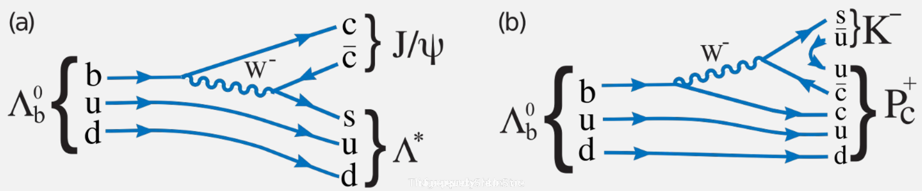
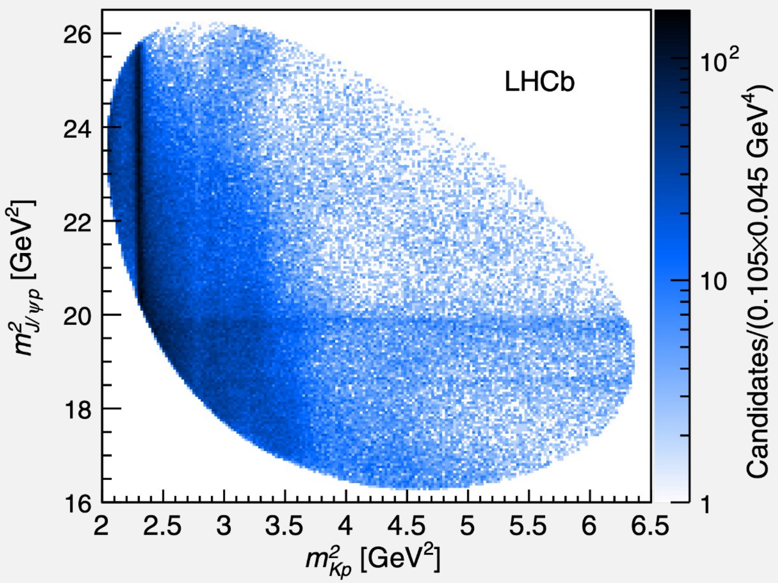


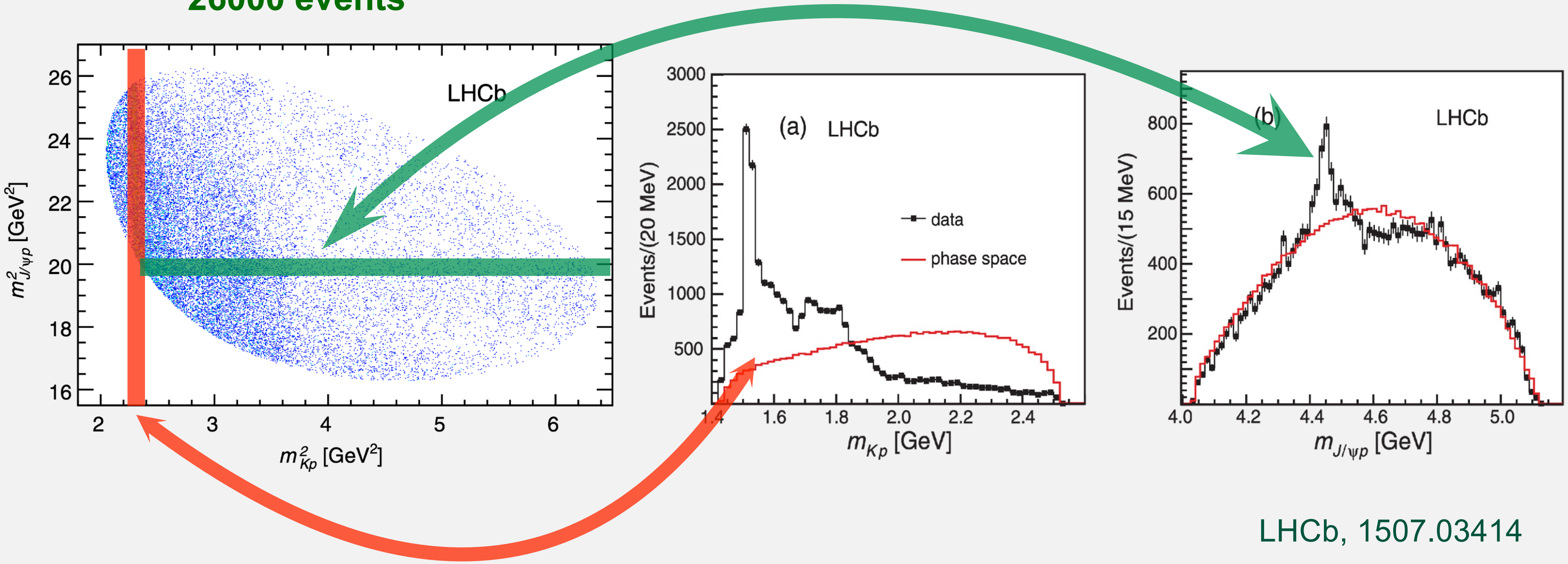
Fig from: LHCb, 1507.03414



246000 events

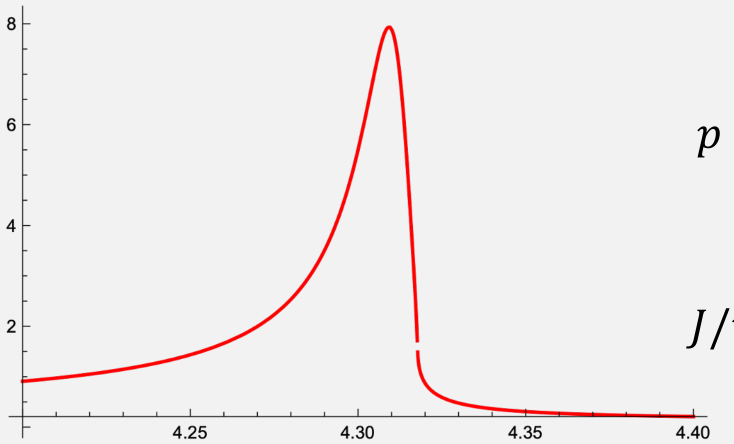
Decay

26000 events



LHCb, 1507.03414

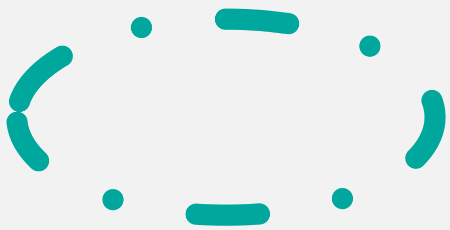
Pentaquark



$$J/\psi p \rightarrow J/\psi p$$

$p = |uud\rangle$

$J/\psi = |c\bar{c}\rangle$



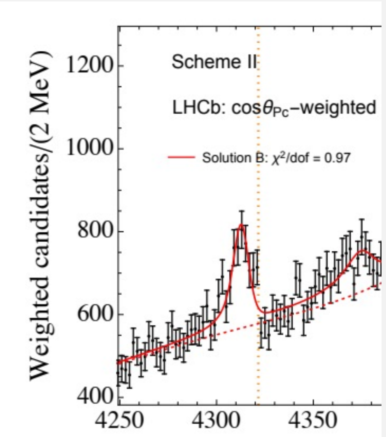
$$\tau = \frac{1}{\Gamma} \sim 10^{-24} s$$

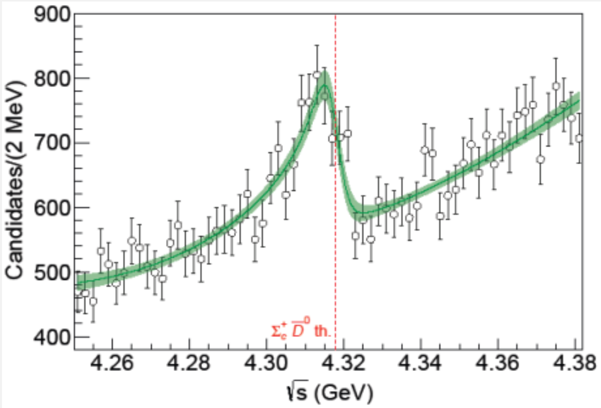
Standard approach to lineshape analysis

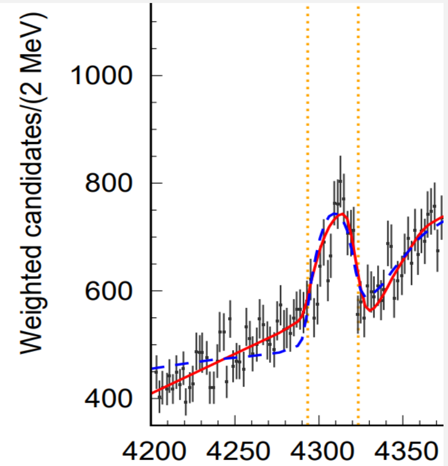
- Build a model for the amplitude and assume it is true
- Fit data using χ^2
- Extract model parameters and get pole positions and compute uncertainties
- Asses the probability that those data were generated by your model
- If everything is fine, you can claim that the interpretation embedded in the model is a possible explanation of the data
- You can do this with different models with different underlying dynamics
- Compare models? Compare dynamics?

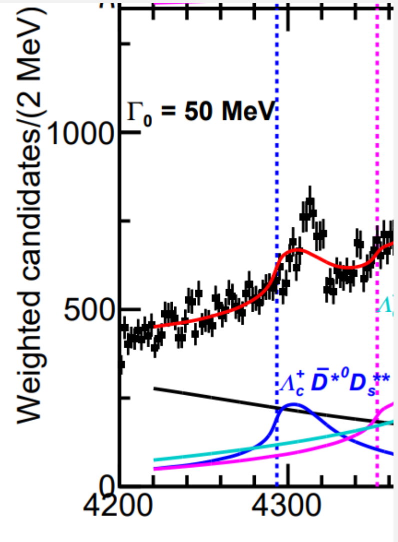
Example $P_c(4312)$

Data from









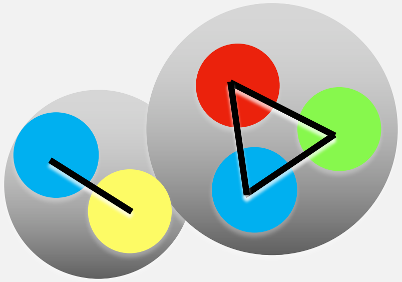
Molecule
Du et al., 2102.07159

Virtual
CFR et al. (JPAC), 1904.10021

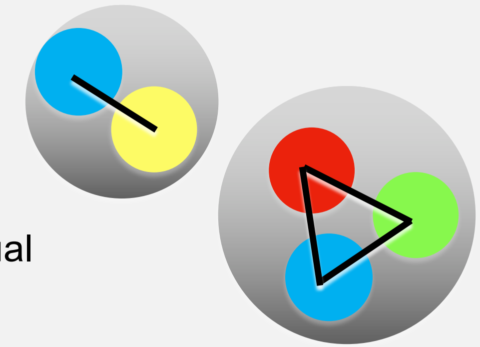
Double-triangle (w. complex coupl. in the Lagrangian)
Nakamura, 2103.06817

Single triangle (ruled out)
LHCb, 1904.03947

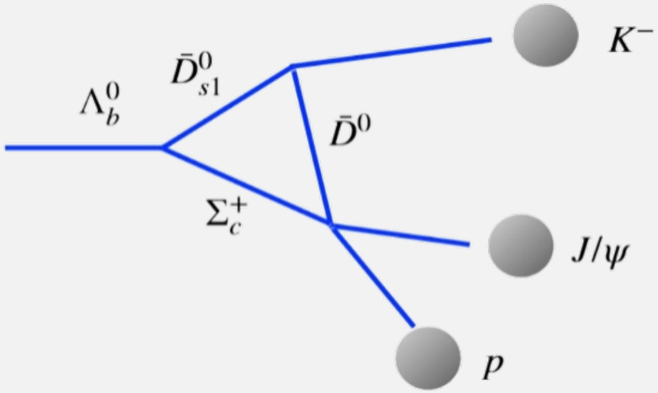
Internal dynamics



Molecule



Virtual

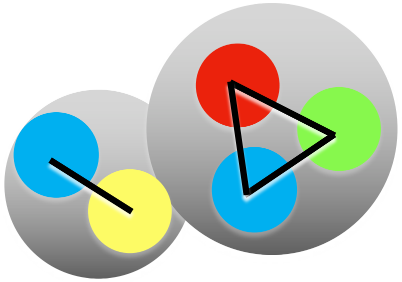


Single triangle
(double triangle=same idea but more complicated)

Threshold generated (residual interaction)

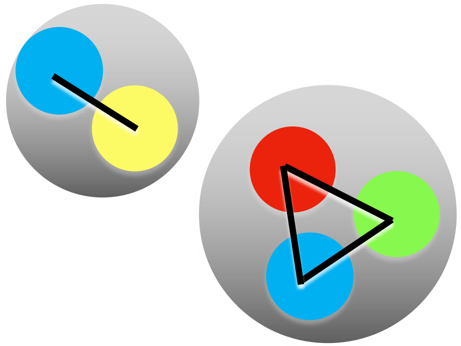
Molecular state

- Generated through the opening of a new channel
- Residual part of the strong interaction bounds the system
- Finite radius

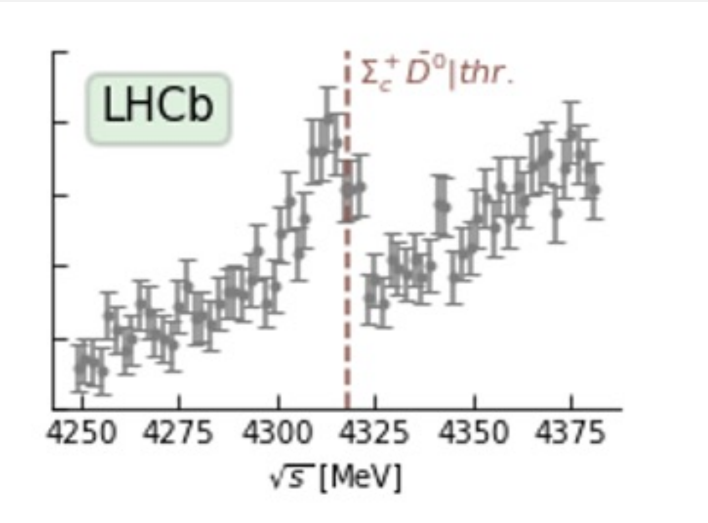


Virtual state

- Generated through the opening of a new channel
- Residual part of the strong interaction generates the signal
- Infinite radius



$J/\psi p$ projection data



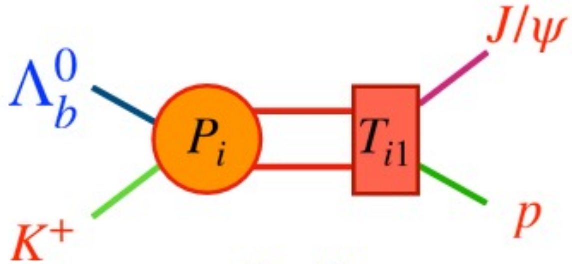
Data from LHCb, 1904.03947

- We focus on $\Sigma_c^+ \overline{D}^0$ threshold region
- One partial wave contributes to the $P_c(4312)$
- The threshold is responsible for the dynamics
- Other singularities are irrelevant

Near-threshold model (two channels)

$$\frac{dN}{d\sqrt{s}} = \rho(s)\left[\,|F(s)|^2 + B(s)\right]$$

$$F(s) =$$



$1 : J/\psi p$

$2 : \Sigma_c^+ \bar{D}^0$

$$F(s) = P_1(s)T_{11}(s) \qquad \left(T^{-1}\right)_{ij} = M_{ij} - ik_i\delta_{ij}$$

Inverse of the scattering length

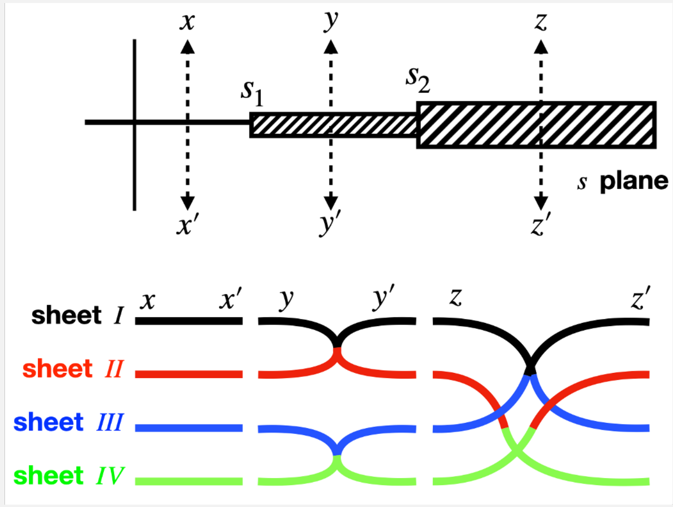
$$M_{ij}(s) = \underset{\downarrow}{m_{ij}}$$

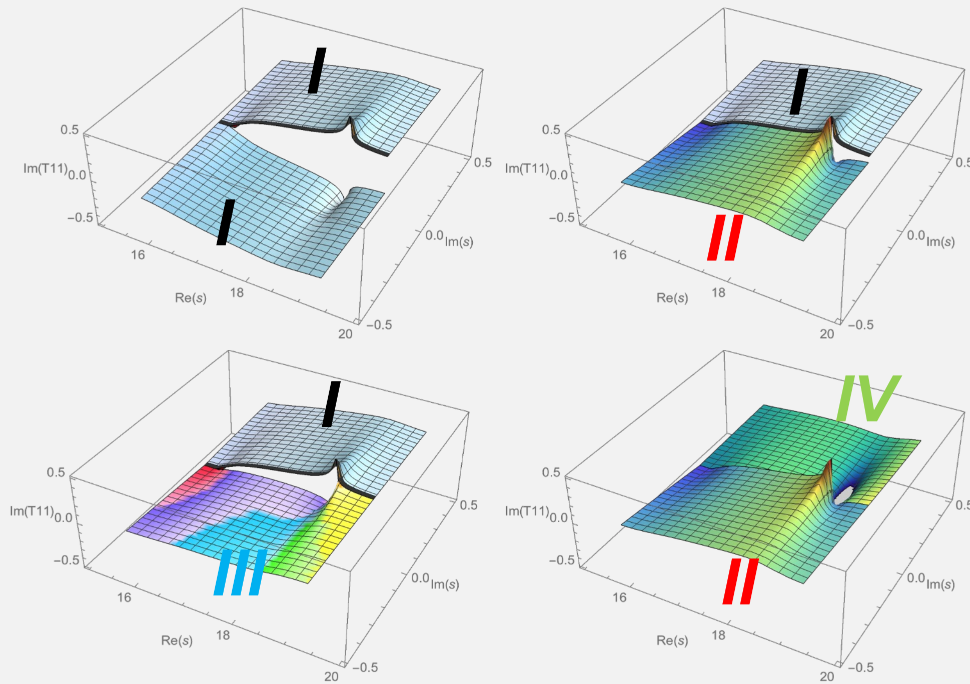
Matrix elements M_{ij} are singularity free and can be Taylor expanded

Frazer, Hendry, PR134 (1964) B1307

Riemann sheets structure (two channels)

$$\begin{bmatrix} J/\psi p \to J/\psi p & J/\psi p \to \Sigma_c \bar{D}^0 \\ \Sigma_c \bar{D}^0 \to J/\psi p & \Sigma_c \bar{D}^0 \to \Sigma_c \bar{D}^0 \end{bmatrix}$$

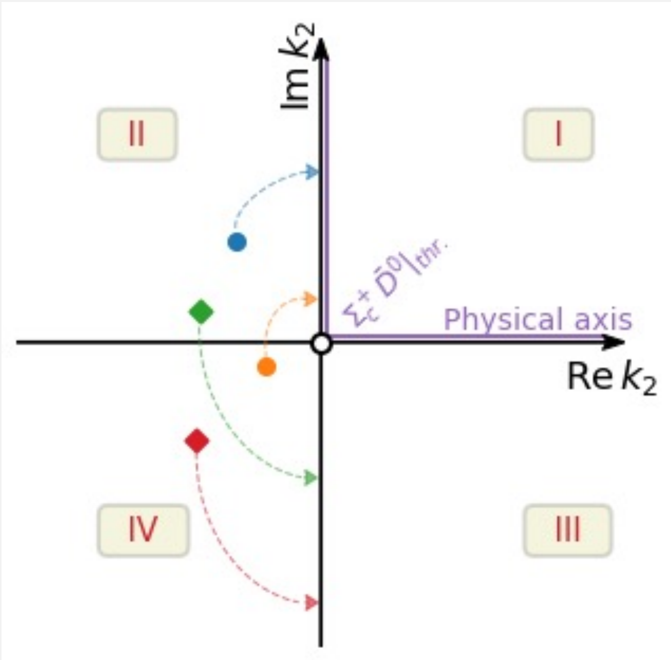




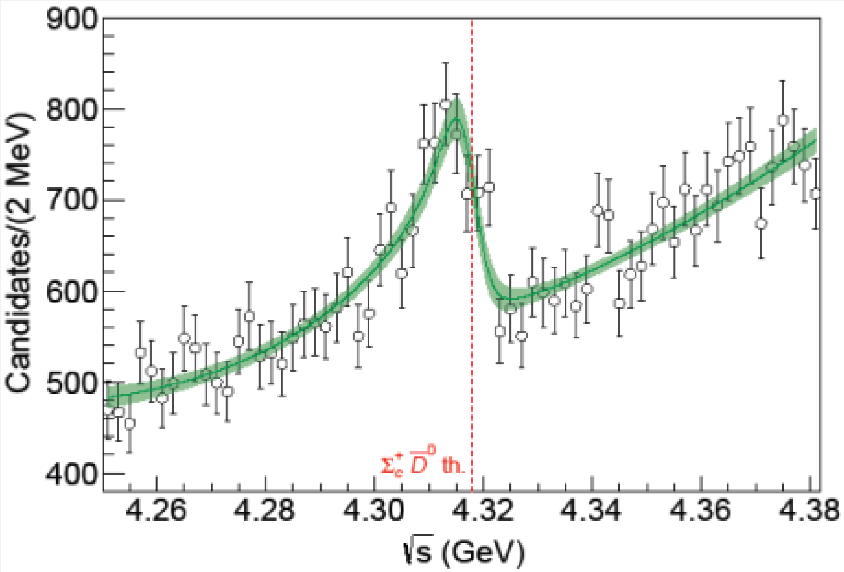
Virtual and bound states

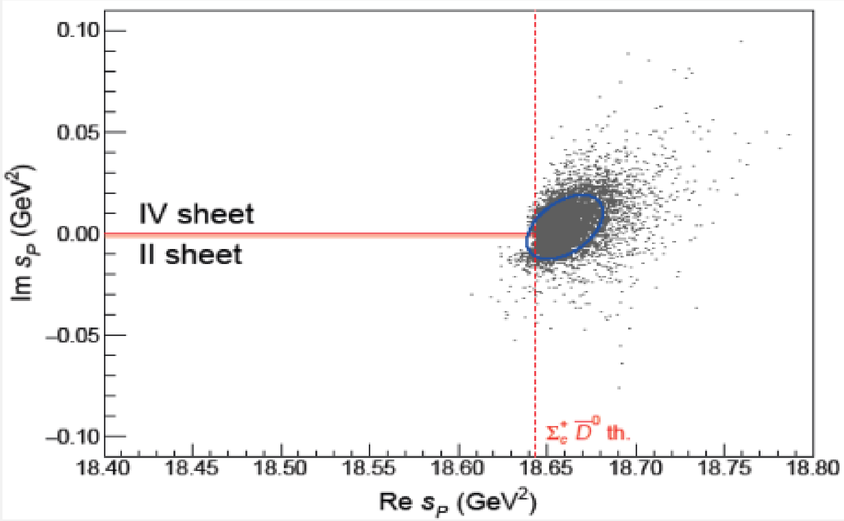
Under the the scattering length approximation the physical interpretation is given by the sign of the m_{22} parameter. Four options:

- Bound state on IV RS: b|4
- Virtual state on IV RS: v|4
- Bound state on II RS: b|2
- Virtual state on II RS: v|2



Amplitude analysis result

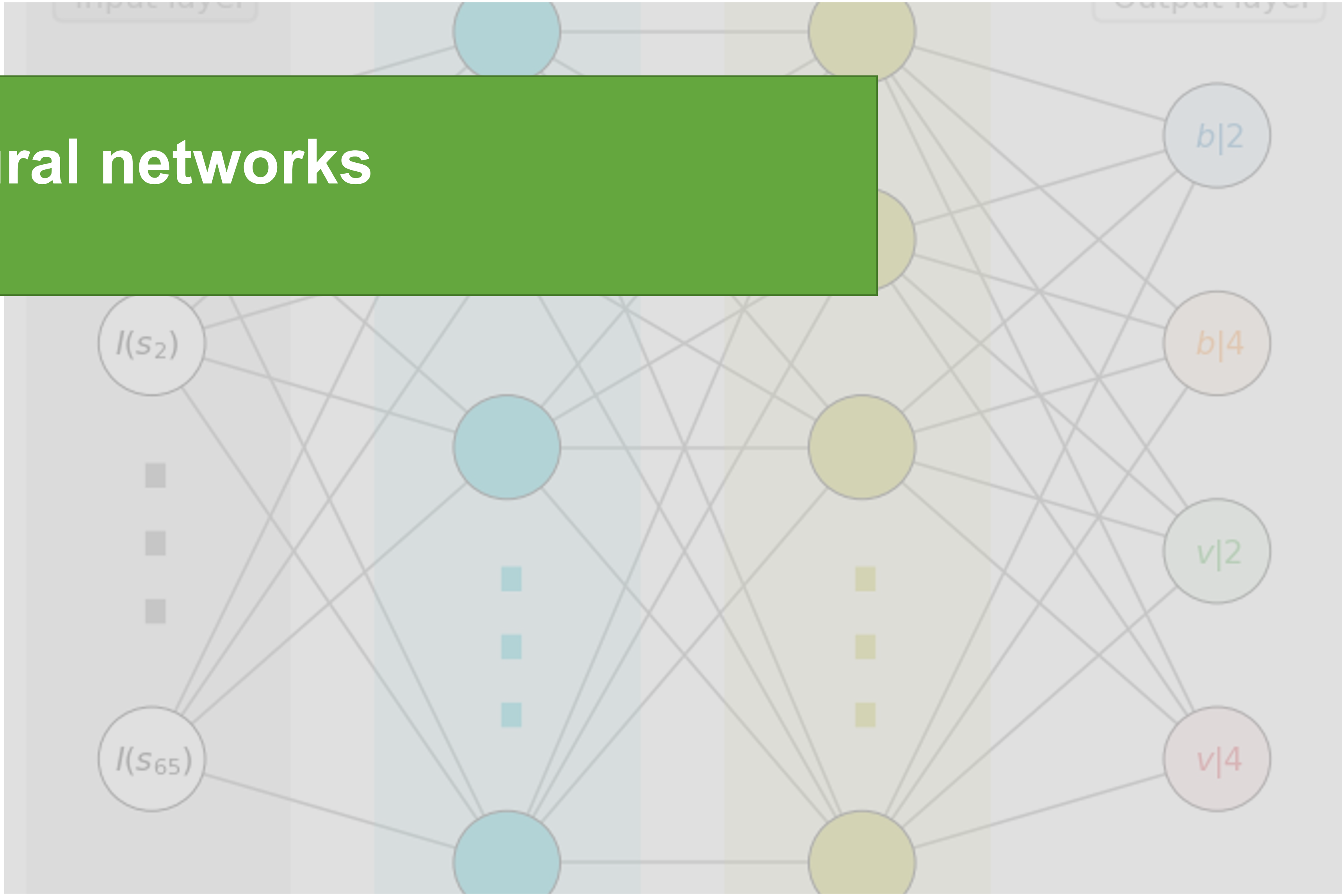




Interpretation obtained: Virtual state on IV RS (v|4)

M= 4319.7±1.6 MeV	Γ= -0.8±2.4 MeV

Neural networks



NN ⊃ ML ⊃ AI

Machine Learning provides methods to learn from data to perform tasks

Widely used nowadays thanks to:

- *Improvement in algorithms and hardware*

Examples:

- *Neural networks, Random forest, Genetic algorithms, …*

Widely used in physics

- *Fitters, universal interpolators, classifiers*

Explainability

- *NNs are usually black boxes, so statistical methods have been developed to study them*

ML/NNs in hadron physics

Experiments

- *Particle ID, A(I)DAPT*

Universal interpolators

- *NNPDF*

Regression

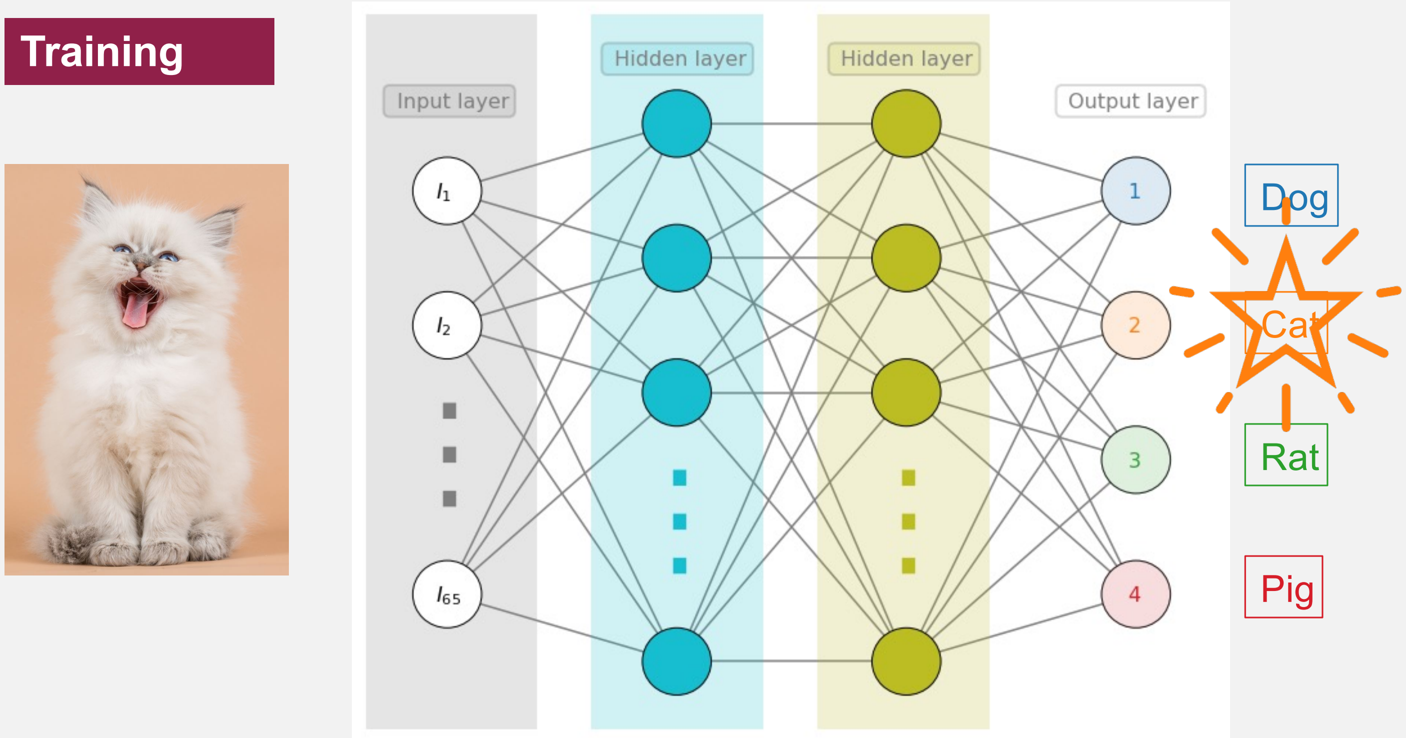
- *Lattice QCD*

Classifiers

- *Hadron spectroscopy (this talk)*

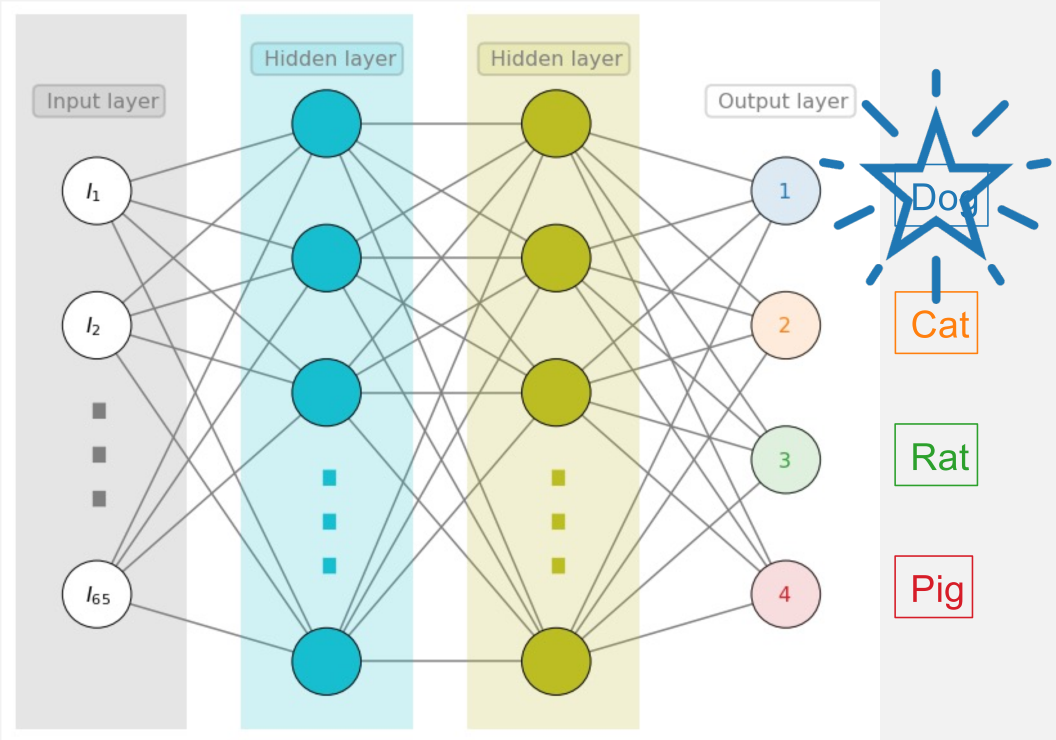
Neural networks as classifiers

Training



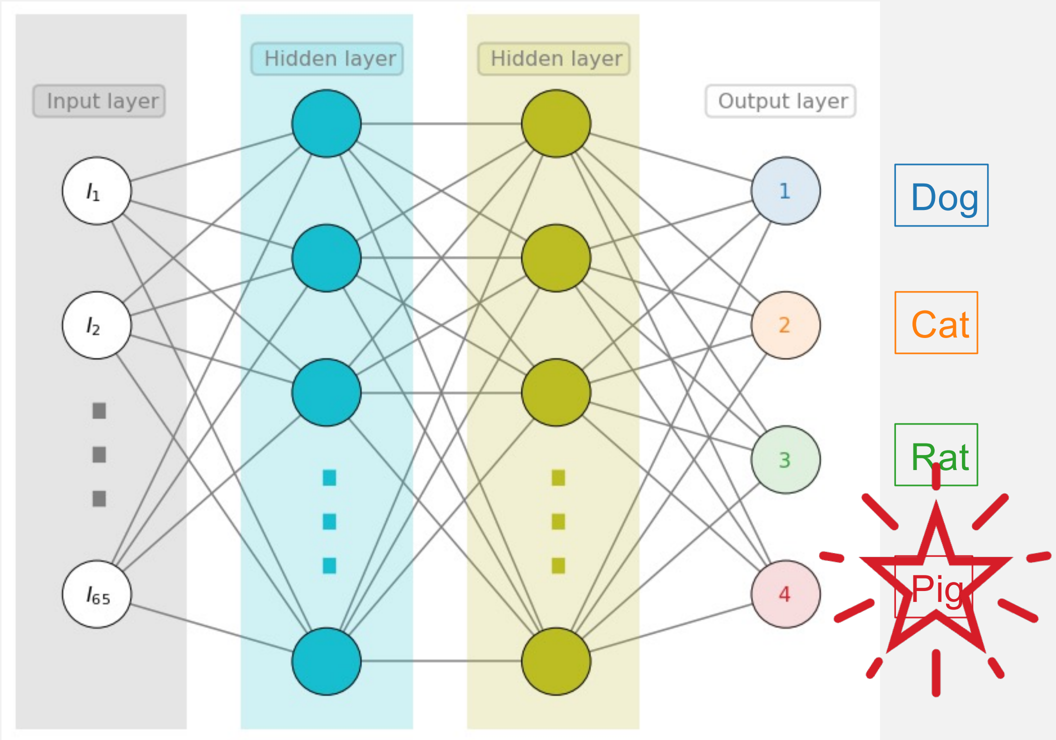
Neural networks as classifiers

Training



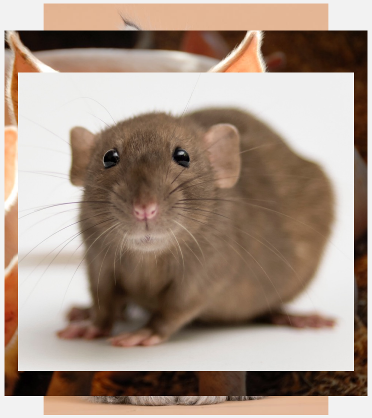
Neural networks as classifiers

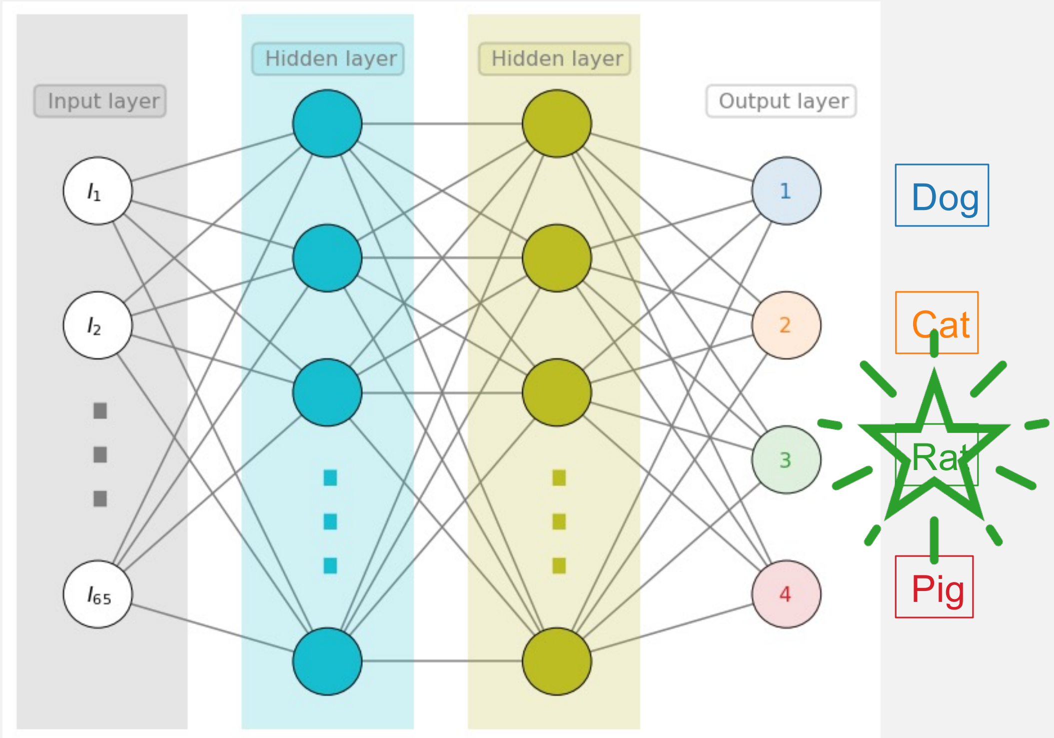
Training



Neural networks as classifiers

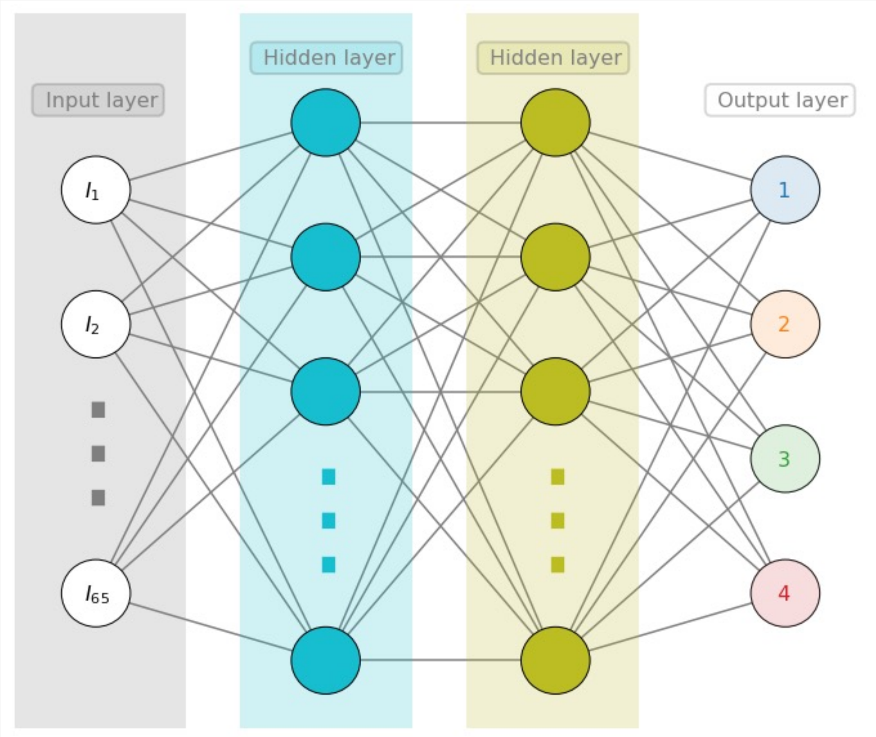
Training





Neural networks as classifiers

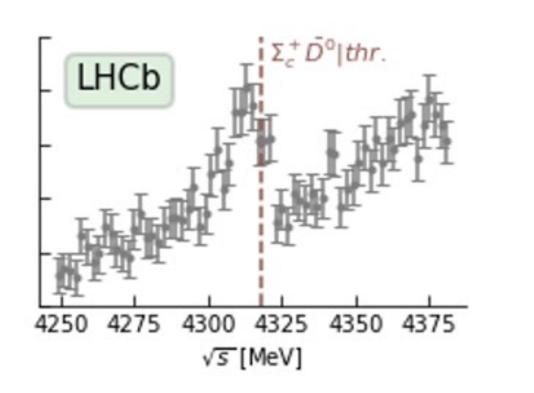
New picture

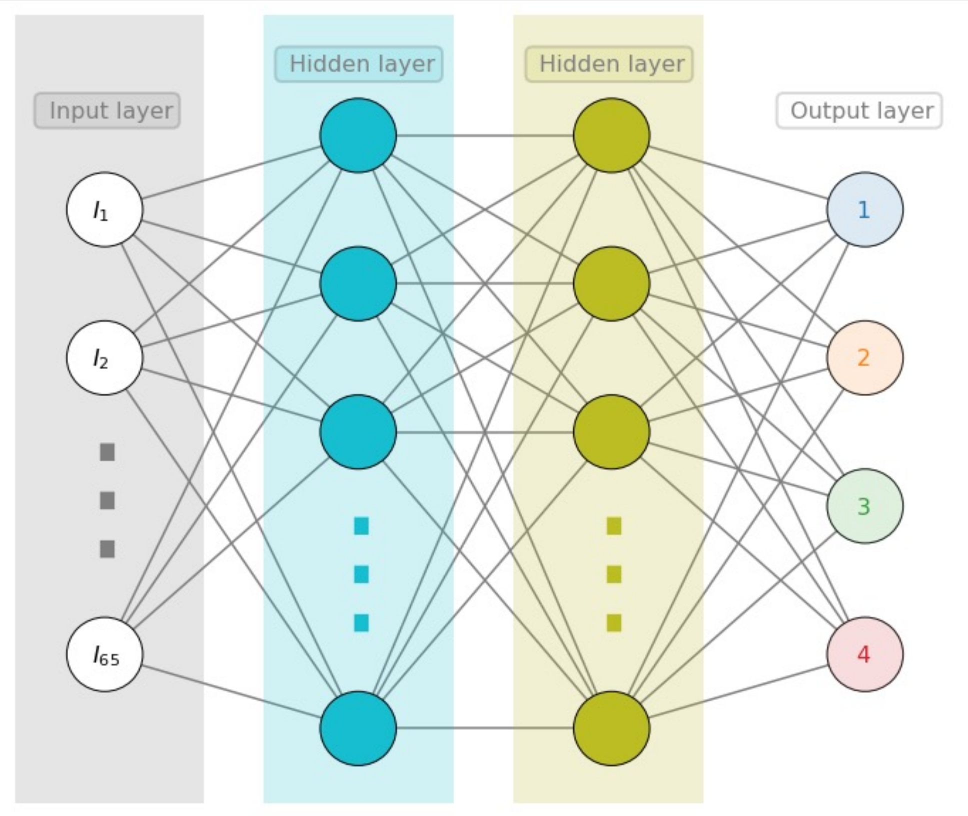


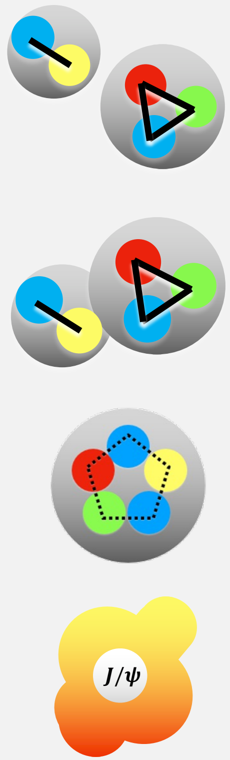
Dog	3%
Cat	96.8%
Rat	0.1%
Pig	0.1%

Disclaimer: Made up percentages

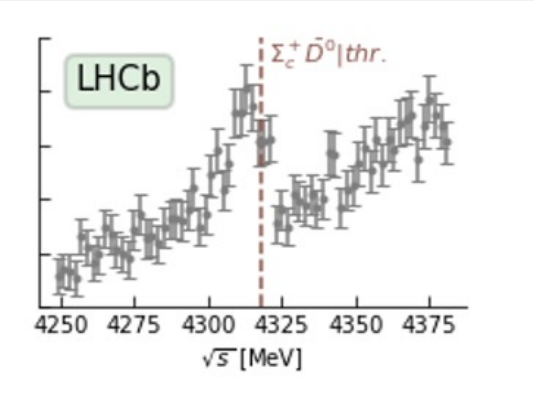
Neural networks as classifiers / lineshape analyzers

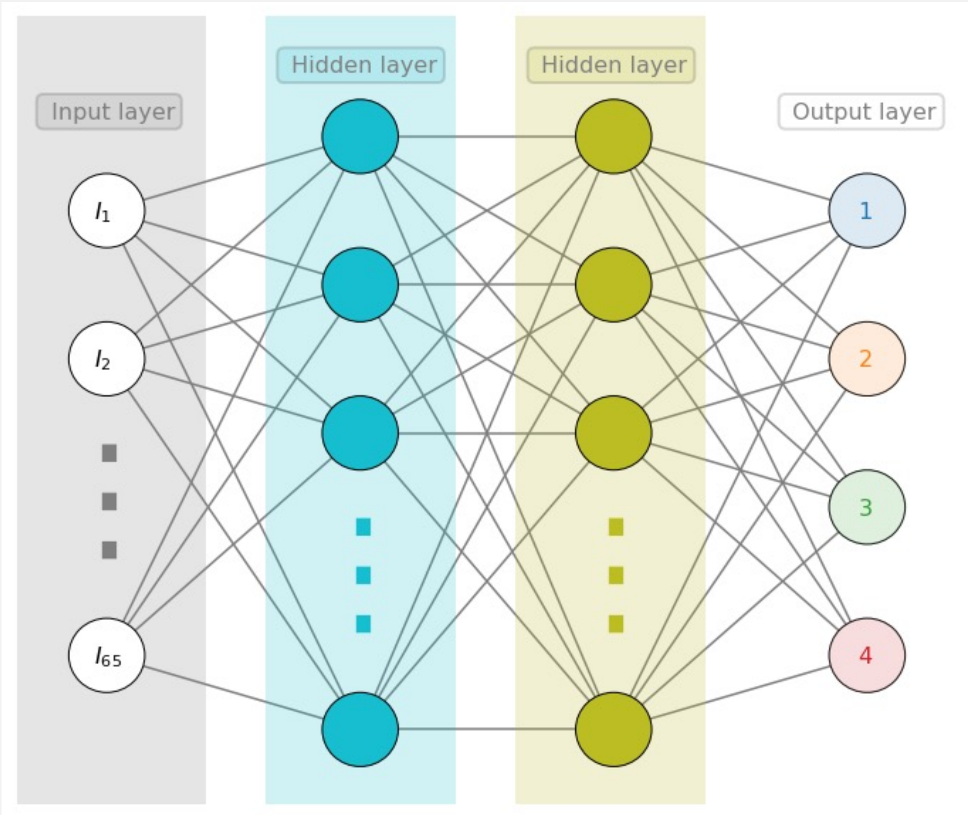






Neural networks as classifiers / model discriminator





- *Model 1*
- *Model 2*
- *Model 3*
- *Model 4*

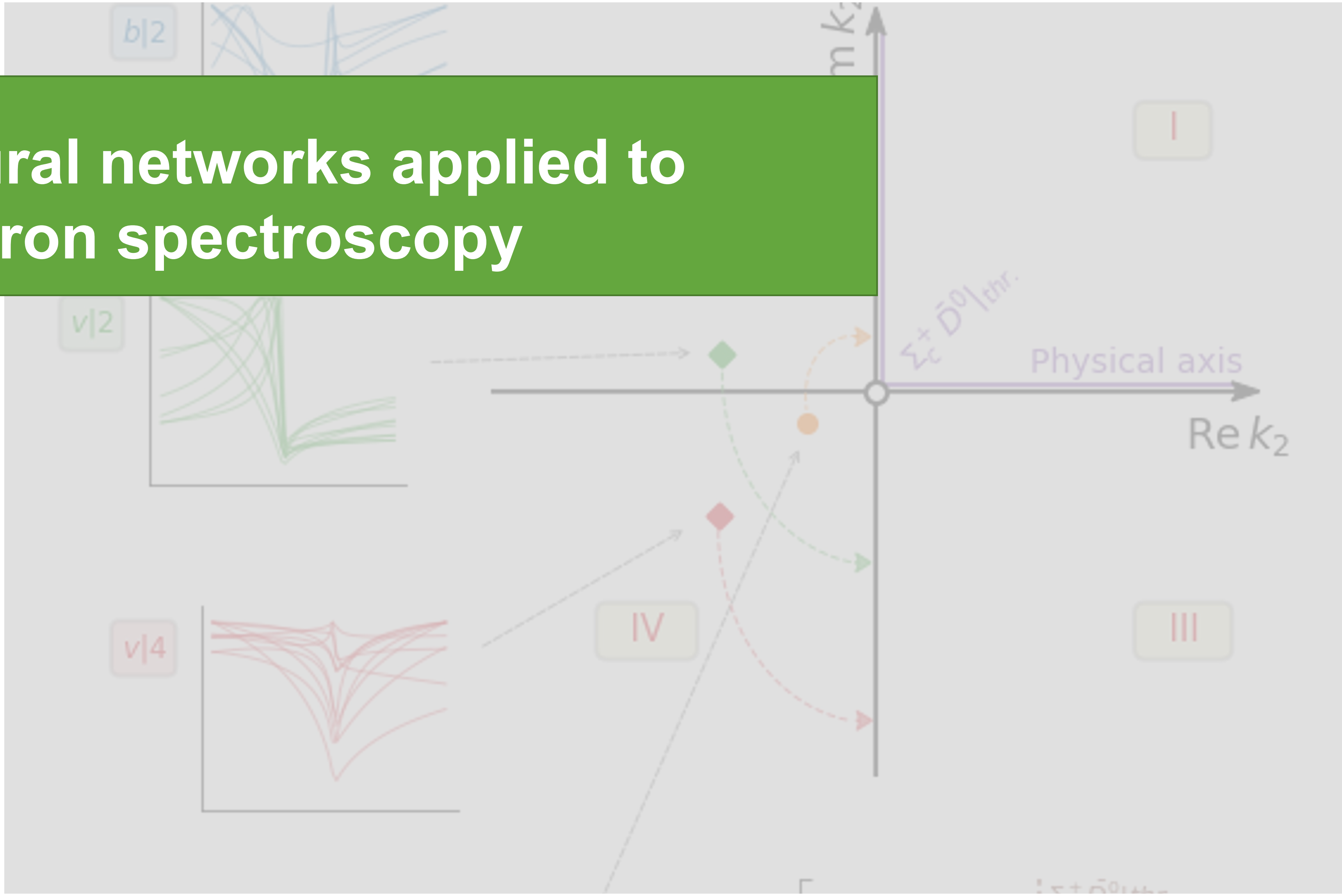
Warning! Be aware of bias!

You get out of the NN what you put in its training

If you train your NN to identify {🐶,🐱,🐷,🐹} ,don’t expect it to identify 🐼!

The NN doesn’t know what a 🐼 is

Neural networks applied to hadron spectroscopy



Can machine learning help us?

The question

- ***Can we train a neural network to analyze a lineshape and get as a result what is the probability of each posible dynamical explanation/model?***

First explorations of neural networks as classifiers for hadron spectroscopy

- *Sombillo et al., 2003.10770, 2104.141782, 2105.04898*

If posible…

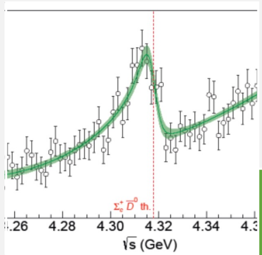
- *What other information can we gain by using machine learning techniques?*

Benchmark case

- *The $P_c(4312)$ lineshape: Ng et al. (JPAC)* 2110.13742

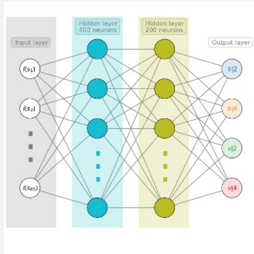
Still far away from answering this question but we are advancing

Dictionary



AMPLITUDE ANALYSIS

a. Datapoints (lineshape)
b. Model convoluted with experimental resolution
c. Experimental uncertainties
d. Physical interpretation
e. Objective (minimization) function



MACHINE LEARNING

a. Features
b. Training set with noise and convolution
c. Bootstrap
d. Classes
e. Optimize the NN (weights)

Building the training set

10^5 training curves

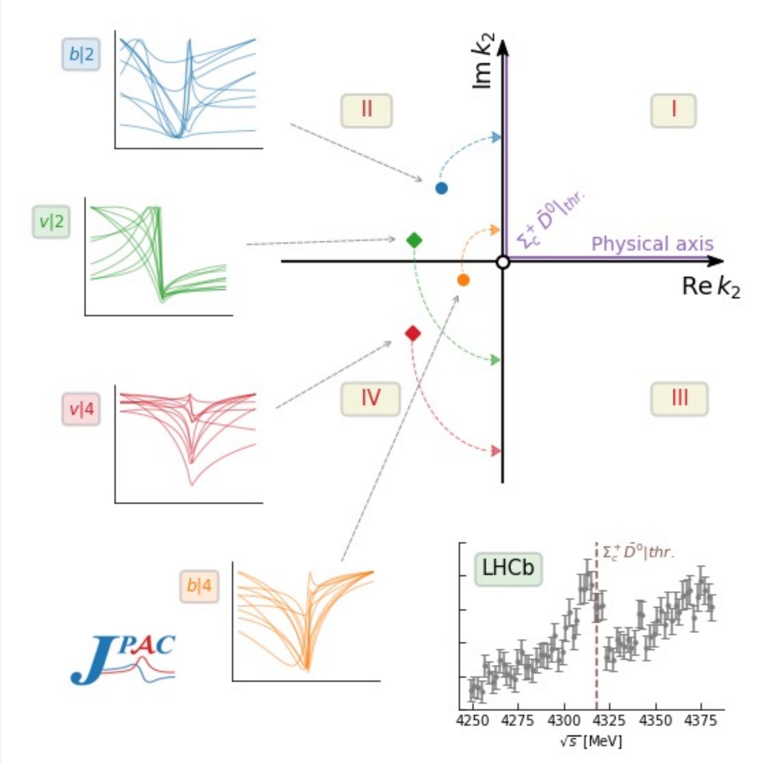
- *Generated by randomly setting parameter values in a wide range*
- *Curves are computed at the experimental energies*

Convolution

- *The lineshapes are convoluted with the experimental resolution*

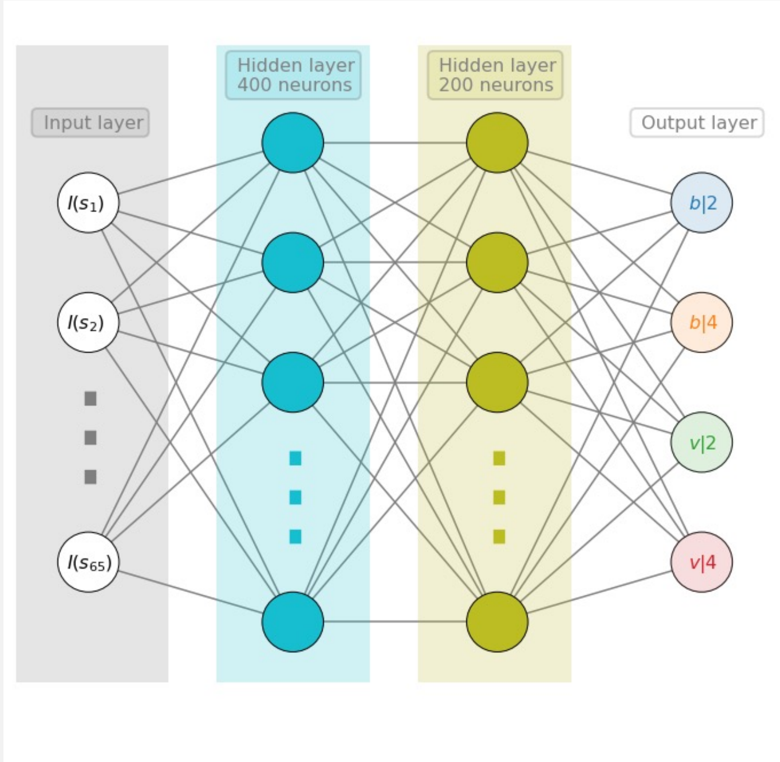
Gaussian noise

- *Included to mimic uncertainties*
- *Compare "blurry" to "blurry"*

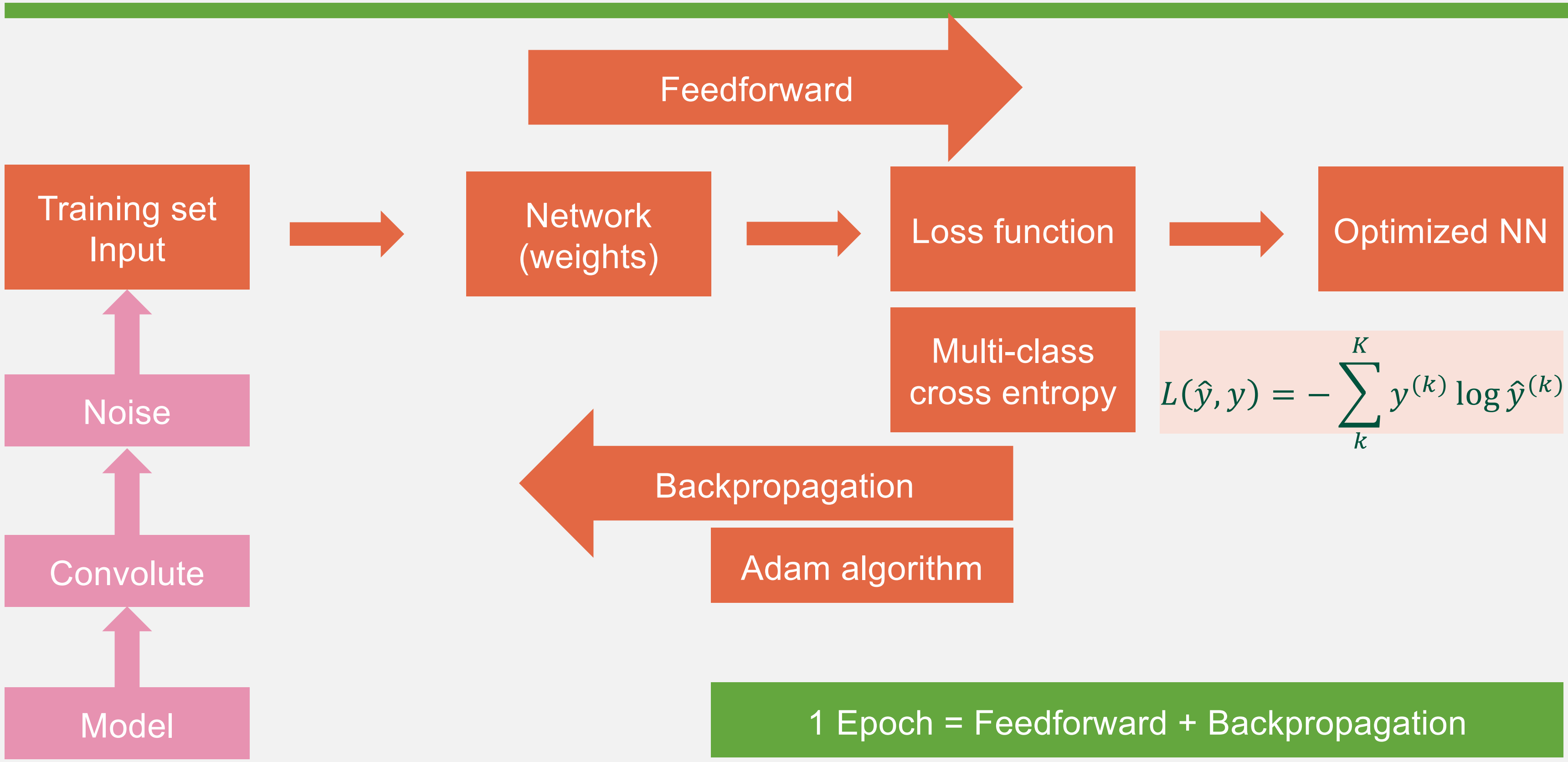


Neural network architecture

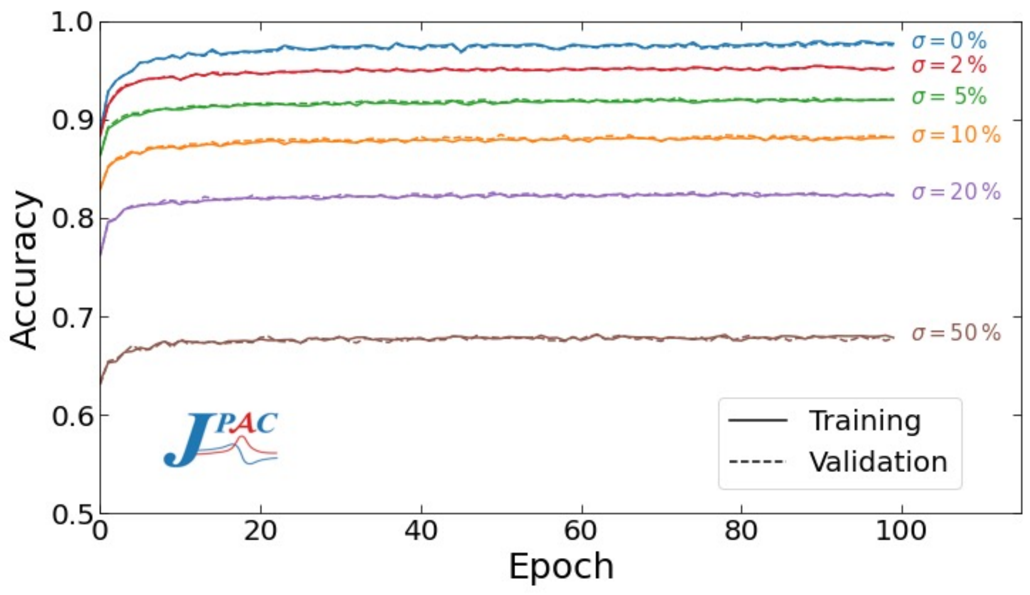
Input layer	(B,65)		
Dense	(B,65)	(B,400)	
Dropout	(B,400)	20%	(B,400)
ReLu	(B,400)	$f(x) = \max(0, x)$	(B,400)
Dense	(B,400)	(B,200)	
Dropout	(B,200)	50%	(B,200)
ReLu	(B,200)	$f(x) = \max(0, x)$	(B,200)
Dense	(B,200)	(B,4)	
Output layer	(B,4)	$p(x_i) = e^{x_i} / \sum e^{x_j}$	(B,4)

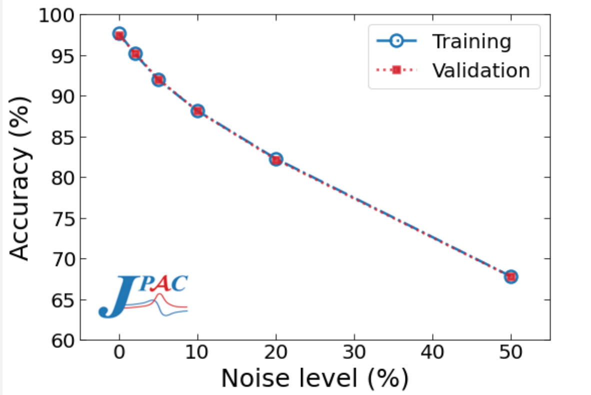


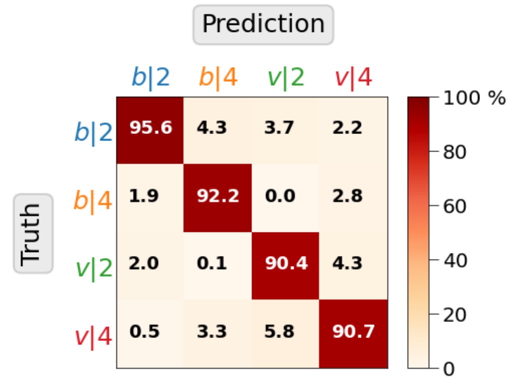
Optimizing (training) the neural network: fitting weights



Training


1.0
0.9
0.8
0.7
0.6
0.5
Accuracy
0
20
40
60
80
100
Epoch
σ = 0 %
σ = 2 %
σ = 5%
σ = 10 %
σ = 20 %
σ = 50 %
Training
Validation
JPAC


100
95
90
85
80
75
70
65
60
Accuracy (%)
0
10
20
30
40
50
Noise level (%)
Training
Validation
JPAC


Prediction
b|2
b|4
v|2
v|4
Truth
b|2
b|4
v|2
v|4
95.6
4.3
3.7
2.2
1.9
92.2
0.0
2.8
2.0
0.1
90.4
4.3
0.5
3.3
5.8
90.7
100 %
80
60
40
20
0

Experimental uncertainties: bootstrap

Associate a distribution to each experimental datapoint

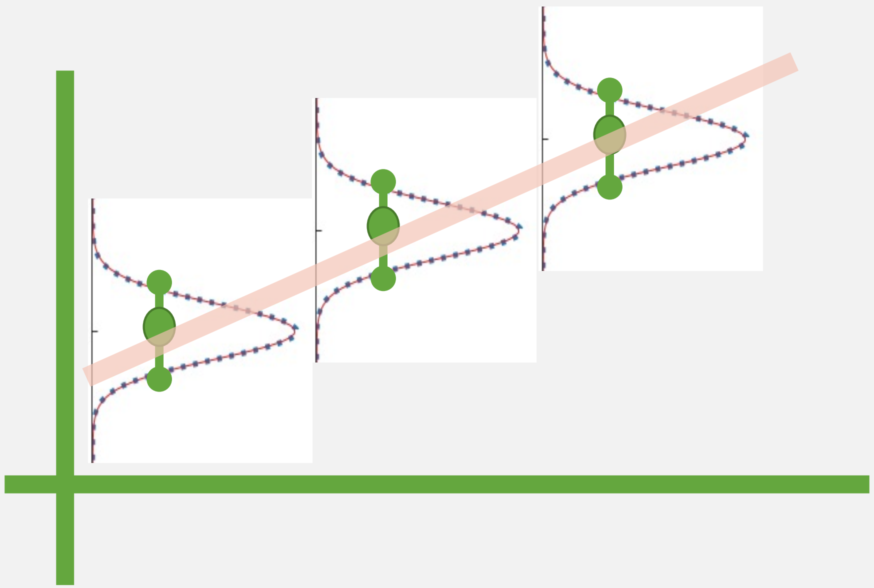
- *Typically a Gaussian with mean and sigma from experiment*

Monte Carlo

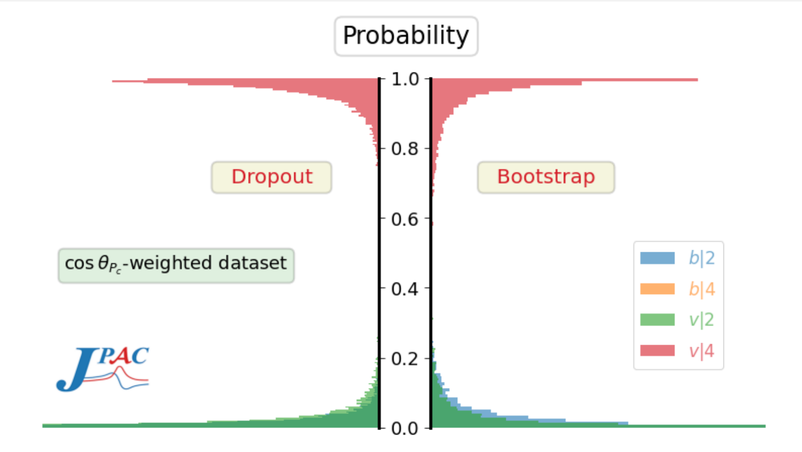
- *Generate pseudodata according to the chosen distribution*

Run statisitcs on the pseudadtasets

- *Compute distributions, mean, standard deviation, quantiles, …*



Applying the NN to LHCb data



We pass the data through the NN

- *Full, cut in m_{Kp} and weighted LHCb datasets through the NN to obtain an answer.*
- *Note that we pass them through the same NN*

Uncertainties

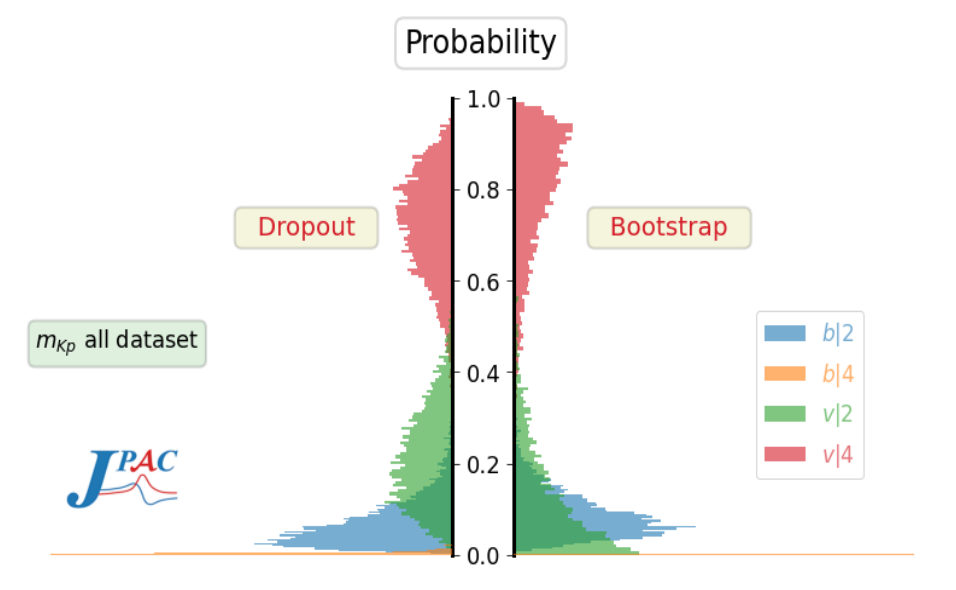
- *Bootstrap and dropout*

Obtain proability distributions

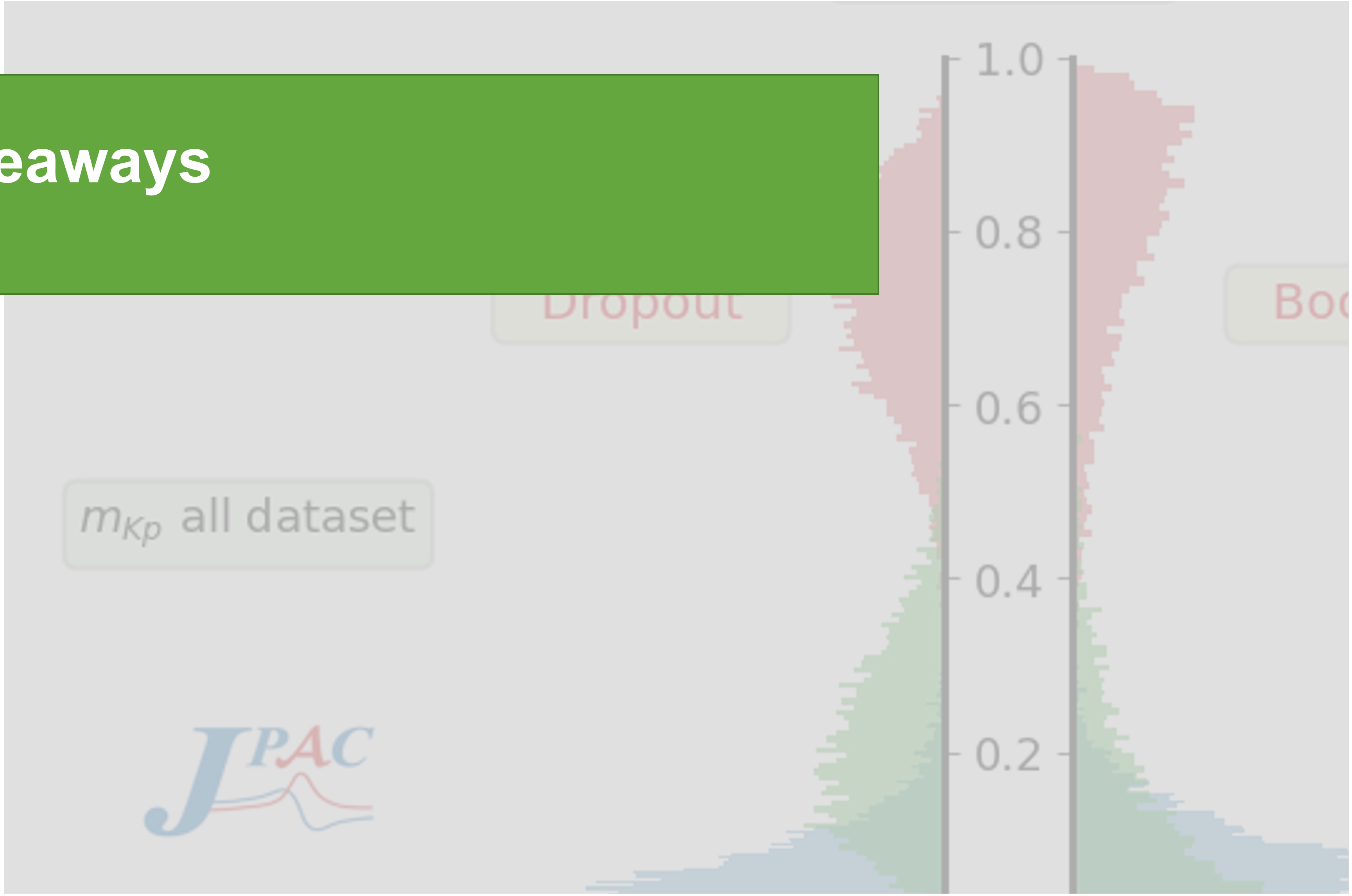
- *We (unsurprisingly) recover the same result as with the standard approach: $v|4$*

Three datasets analyzed with the same NN

	$b\|2$	$b\|4$	$v\|2$	$v\|4$
$\cos\theta_{P_c}$-weighted	0.6%	< 0.01%	1.1%	98.3%
$m_{Kp} > 1.9\,\mathrm{GeV}$	1.4%	< 0.1%	1.6%	97.0%
m_{Kp} all	5.4%	< 0.1%	21.0%	73.6%



Takeaways



What we get from the NN

The NN targets specific regions of the parameter space (which yield stable solutions) that might be difficult to reach during optimization or might require high-resolution data

Standard χ^2fit can be indeed unstable, and a small change in the input data can induce large changes in the parameter values and therefore in the physics interpretation

Rather than testing a single model hypothesis as a χ^2 fit would, the NN determines the probability of each of the classes of interest, given the experimental uncertainties. The latter is possible, since the NN learns the subtle classification boundary between the different classes

But, there's more…

Explainability

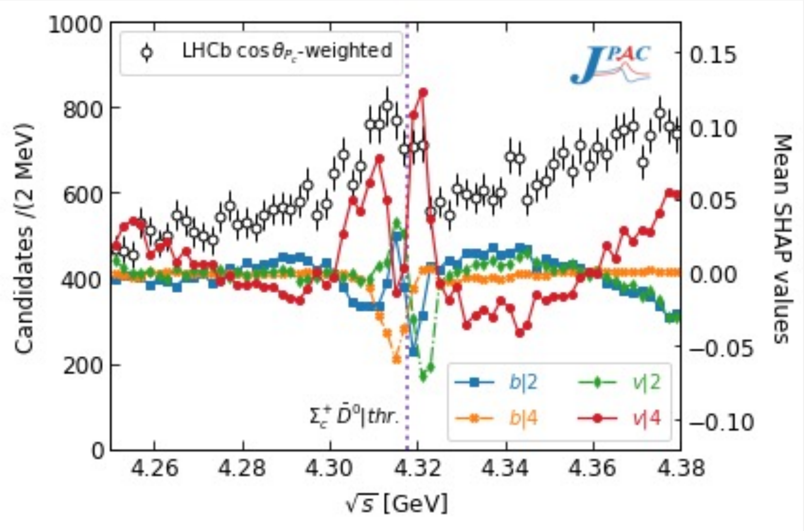
SHAP values

- *SHapley Additive exPlanations*

Inherited from game theory

Application

- *Allows to determine how a given feature in the input layer (in our case an experimental datapoint) impacts the decision made by the NN in the output layer (the classes)*



Takeaways

- Neural networks open new possibilities to answer the question on the underlying nature of a given resonance
- Possibility to gain physics insight on how the data impact the obtained interpretation
- NN does NOT substitute the standard approach. They are complementary (we still want the amplitude and the pole position)
- NN allows a true comparison among interpretations
- The BIG objective: Be able to train the DNN with every amplitude possibility we can devise with our twisted minds and throw the data to it, returning probabilities for each class
- Uncharted territory so we are taking baby steps: Ng et al. (JPAC), 2110.13742
- We are (hopefully) just in the begining…

Grazie! Domande?