Seminarium Wydziału Fizyki i Informatyki Stosowanej AGH
2 czerwca, 2023

Uczenie maszynowe w analizie pentakwarku $P_c(4312)$

Łukasz Bibrzycki
Katedra Informatyki Stosowanej i Fizyki Komputerowej
WFiIS AGH

Joint Physics Analysis Center (JPAC)


Mikhail Mikasenko
TU Munich


Cesar Fernández-Ramírez
National Autonomous U. of Mexico


Daniel Winney
South China Normal U.


Victor Mokeev
Jefferson Lab


Sergi Gonzalez-Solis
Los Alamos National Lab.


Alessandro Pilloni
U. Messina


Łukasz Bibrzycki
AGH Kraków


Astrid Hiller Blin
U. of Heidelberg


Vincent Mathieu
U. Barcelona


Adam Szczepaniak
Indiana U.


Miguel Albaladejo
CSIC-Valencia


Arkaitz Rodas
College of William and Mary


Robert Perry
U. Barcelona


Jorge A. Silva-Castro
National Autonomous U. of Mexico


Lawrence Ng
Florida State U.

Outline

- Motivation
- Physical model
- ML model
- Feature refinement
- Model predictions and explanation
- Outlook and open questions

Applicability of the QCD

- QCD is a relativistic field theory with local SU(3) gauge symmetry with matter quanta called quarks and gauge quanta called gluons

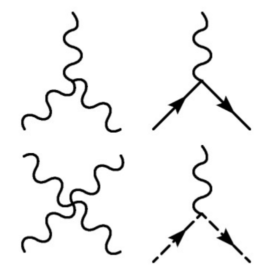
$$\mathcal{L}_{\rm QCD} = \overline{\psi}_i \left(i (\gamma^\mu D_\mu)_{ij} - m\, \delta_{ij} \right) \psi_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a , \quad \text{where}$$

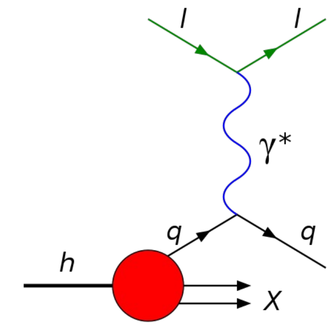
$$G^a_{\mu\nu} = \partial_\mu \mathcal{A}^a_\nu - \partial_\nu \mathcal{A}^a_\mu + g f^{abc} \mathcal{A}^b_\mu \mathcal{A}^c_\nu , \quad \text{and}$$

$$(D_\mu)_{ij} = \partial_\mu \delta_{ij} - i g \, (T_a)_{ij} \, \mathcal{A}^a_\mu$$

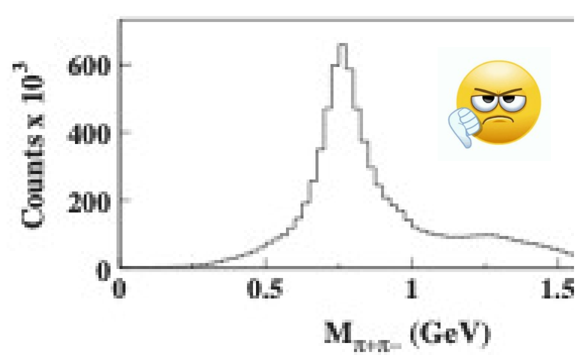
- No free params (apart from coupling and quark masses), a lot of symmetries – very decent theory

- Unfortunately, perturbative approach is limited to the high energy region



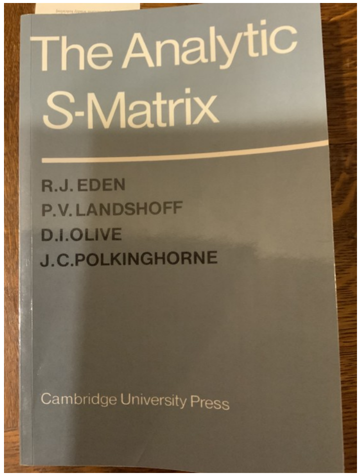


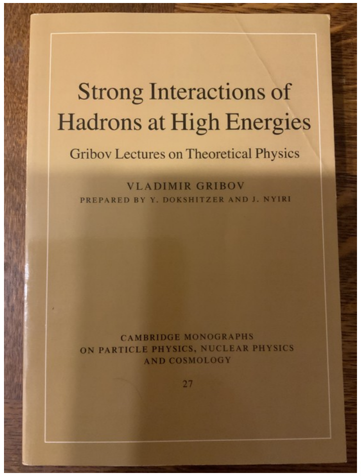
- Thus not applicable for hadron spectroscopy and description of hadronic processes



QCD substitutes at hadronic scale ~1 GeV

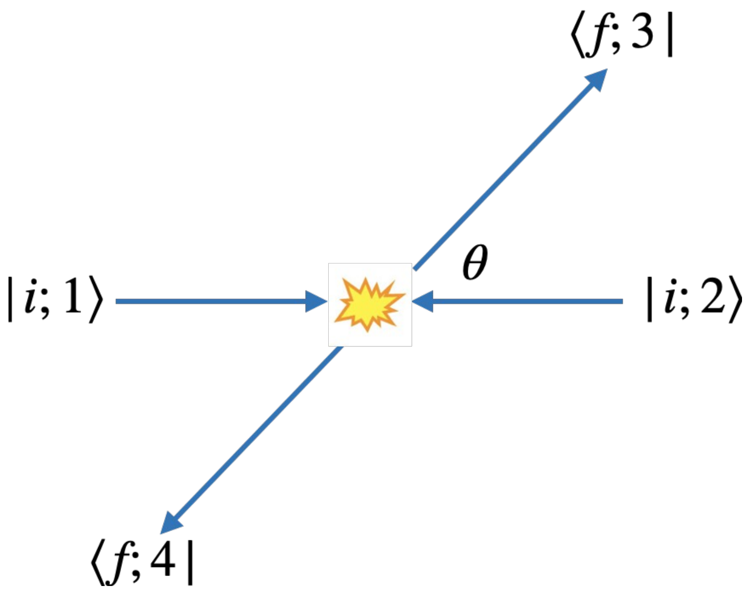
- Quark model – it is NOT QCD, non-relativistic, no gauge/no dynamical gluons, application for scattering processes is problematic but… works !
- Lattice QCD – computing the QCD observables ie. masses, widths (but also scattering parameters), the only first principles based approach to the QCD, very promising, very actively studied but … you can't fit it !
- Analytic scattering theory





Quantum scattering

- Wheeler (1937), Heisenberg (1943)



$$|i\rangle = |i;1\rangle|i;2\rangle \qquad |f\rangle = |f;3\rangle|f;4\rangle$$

$$S_{fi} = \langle f|S|i\rangle$$

- S-matrix is a measure of the overlap between the initial and final state – the larger the overlap, the stronger the interaction
- Laws of physics are encoded as relations between S-matrix elements

- Expressing the S-matrix in terms of Lorenz invariant variables, like Mandelstam invariants, $s=(p_1+p_2)^2$, $t=(p_1-p_3)^2$ ensures that it is manifestly Lorenz invariant
- Usually we are interested in situations where scattering really happened, rather than particles missed each other without interaction

$$S_{fi} = \delta_{fi} + iT_{fi};$$

$$T_{fi} \sim f(k,\theta) - \text{scattering amplitude}$$

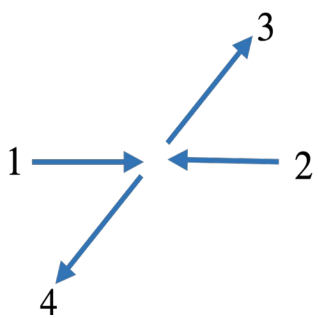
$$\frac{d\sigma}{d\Omega} = |f(k,\theta)|^2$$

AGH

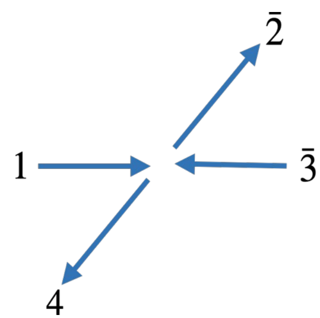
JPAC

Properties of the S-matrix

- Unitarity: $$S S^\dagger = 1 \Leftrightarrow \hat{T} - \hat{T}^\dagger = i \hat{T} \hat{T}^\dagger$$
- Crossing symmetry: Processes



and



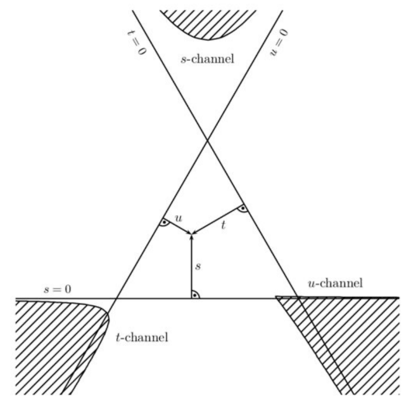


Fig. From PhD thesis of Stefan Lanz

are described by the same amplitude (up to analytical continuation from the physical region of one to the other)

- Analytical properties of the amplitude bear dynamical interpretation:
  - Poles – bound states or unstable “particles” produced in the intermediate state
  - Branch cuts – thresholds for production of particles (unitary cut) or particles exchanged in the crossed channels (dynamical cut)

Partial wave expansion (spinless case)

$$f_{ji}(s,t) = \sum_{l=0}^{\infty}(2l+1) f^l_{ji}(s) P_l(\cos\theta(s,t))$$

- Unitarity relation for partial wave amplitudes

$$f^l_{ji} - f^{l*}_{ij} = 2i \sum_n p_n f^l_{jn} f^{l*}_{in} \quad \text{or} \quad \hat{f}^l - \hat{f}^{l\dagger} = 2i \hat{f}^l \hat{p} \hat{f}^{l\dagger}$$

- This can be solved in the general form:

$$\hat{f}^{-1} = \hat{M}(s) - i\hat{p}$$

- It was shown by *Frazer, Hendry, Phys. Rev. 134 (1964)* that $\hat{M}$ is symmetric and free from unitarity cuts, so can be Taylor expanded in *s*.

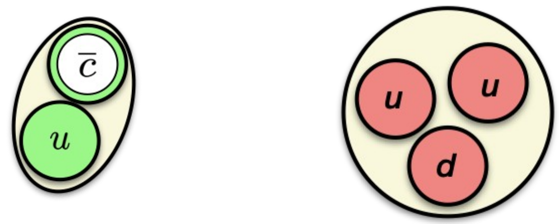
$$\hat{M} \approx \hat{m} - \hat{c}s$$

Scattering length

Effective range

Understanding resonances

- For decades much more resonances were predicted than observed

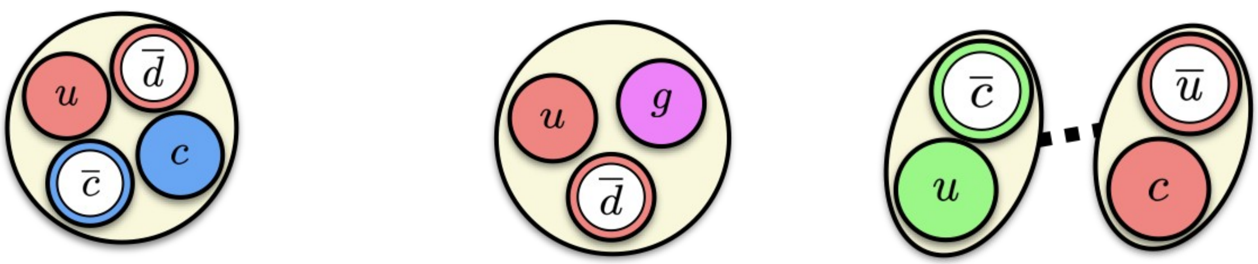


$$|\text{meson}\rangle = \alpha_1|q\bar{q}\rangle + \alpha_2|q\bar{q}q\bar{q}\rangle + \alpha_3|q\bar{q}g\rangle + \alpha_4|gg\rangle + \ldots$$

$$|\text{baryon}\rangle = \alpha_1|qqq\rangle + \alpha_2|q\bar{q}qqq\rangle + \alpha_3|qqqg\rangle$$

$$\sum_{i=0}^{\infty}|\alpha_i|^2 = 1$$

- With the observation of exotic charmonium X(3872) in the $D^0\bar{D}^{0*}$ channel by Belle in 2003 the avalanche of observations of exotics started



Motivation

Plethora of potentially multiquark states observed in last decade

PHYSICAL REVIEW LETTERS **122,** 222001 (2019)

Editors' Suggestion | Featured in Physics

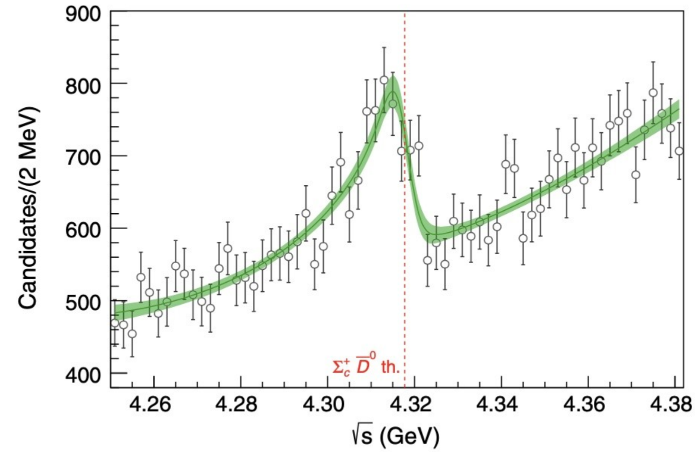
Observation of a Narrow Pentaquark State, $P_c(4312)^+$, and of the Two-Peak Structure of the $P_c(4450)^+$

R. Aaij *et al.**
(LHCb Collaboration)

(Received 6 April 2019; published 5 June 2019)

A narrow pentaquark state, $P_c(4312)^+$, decaying to $J/\psi p$, is discovered with a statistical significance of 7.3σ in a data sample of $\Lambda_b^0 \to J/\psi p K^-$ decays, which is an order of magnitude larger than that previously analyzed by the LHCb Collaboration. The $P_c(4450)^+$ pentaquark structure formerly reported by LHCb is

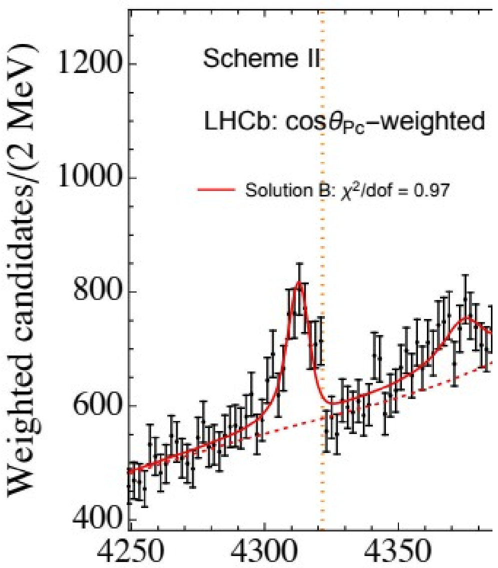
Possible interpretation as $duuc\bar{c}$ pentaquark



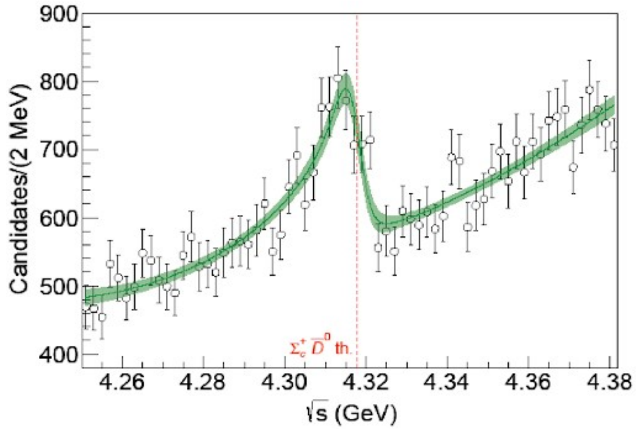
Intensity in the $P_c(4312)$ neighbourhood and the JPAC fit *C. Fernandez-Ramirez Phys.Rev.Lett. 123 (2019) 9, 092001*

- There is a close relation between QCD spectrum and the analytic structure of amplitudes (production thresholds → branch points, resonances/bound states → poles)
- Currently this relationship is impossible to derive from first principles of QCD (top down approach)
- One can utilize the general properties of amplitudes, like unitarity, analyticity or crossing symmetry, but then model parameters must be derived from data – bottom up approach

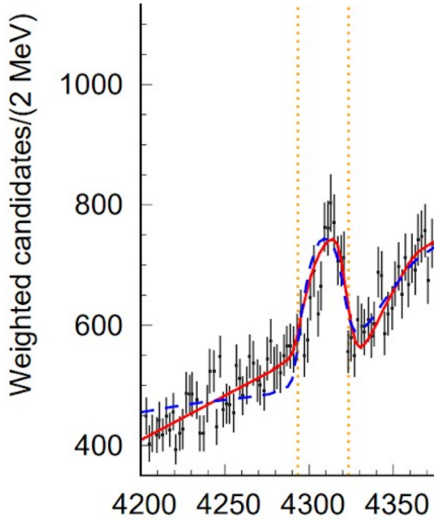
Discrepant interpretations of the $P_c(4312)$ nature



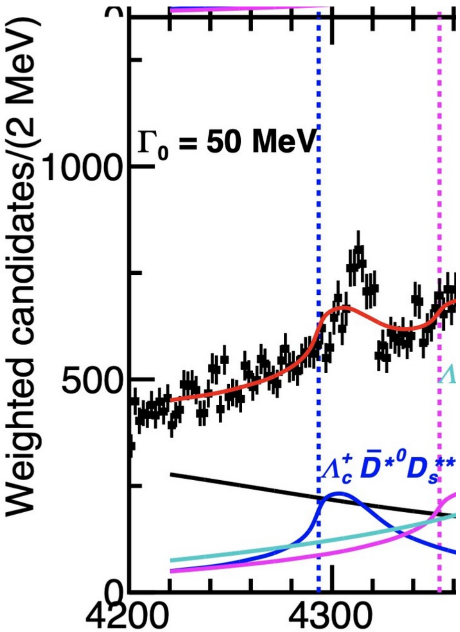
Molecule
Du et al., 2102.07159



Virtual
C. F-R et al. (JPAC), Phys. Rev. Lett. 123, 092001 (2019)



Double-triangle (w. complex coupl. in the Lagrangian)
Nakamura, Phys. Rev. D 103, 111503 (2021)



Single triangle (ruled out)
LHCb, Phys. Rev. Lett. 122, 222001 (2019)

Interpretation of amplitude poles – Morgan-Pennington criterion

- A resonance is a pole in the closest unphysical Riemann sheet (2-nd – in one channel case)
- With increasing channel number one resonance can have poles on several Riemann sheets
- In two channel case (4 sheets) the emergence of additional pole on the 3-rd sheet close to the physical region indicates the quark model – like nature of the state
- Pole on the 4-th sheet indicates the virtual state

AGH

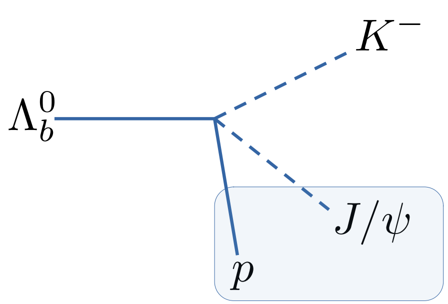
JPAC

We want to use ANN to:

- Go beyond the standard χ^2 fitting
- Specific questions:
  - Can we train a neural network to analyze a line shape and get as a result the probability of each possible dynamical explanation ?
  - If possible, what other information can we gain by using machine learning techniques?
- First attempts to use Deep neural networks as model classifiers for hadron spectroscopy:

Sombillo et al., 2003.10770, 2104.141782, 2105.04898

Physics model



- $P_c(4312)$ seen as a maximum in the pJ/ψ energy spectrum
- $P_c(4312)$ has a well defined spin and appears in single partial wave
- Background contributes to all other waves
- $\Sigma^+_c \bar{D}^0$ channel opens at 4.318 GeV -coupled channel problem

- Intensity

$$\frac{dN}{d\sqrt{s}} = \rho(s)\left[|P_1(s)T_{11}(s)|^2 + B(s)\right]$$

where

$$\rho(s) = pqm_{\Lambda_b}$$ phase space

$$p = \lambda^{\frac{1}{2}}(s, m^2_{\Lambda_b}, m^2_K)/2m_{\Lambda_b}, \;\; q = \lambda^{\frac{1}{2}}(s, m^2_p, m^2_\psi)/2\sqrt{s}$$

$$P_1(s) = p_0 + p_1 s$$ production term

$$B(s) = b_0 + b_1 s$$ background term

Physics model

- Coupled channel amplitudes

$$T_{ij}^{-1} = M_{ij} - ik_i\delta_{ij} \quad \text{where} \quad k_i = \sqrt{s - s_i}$$

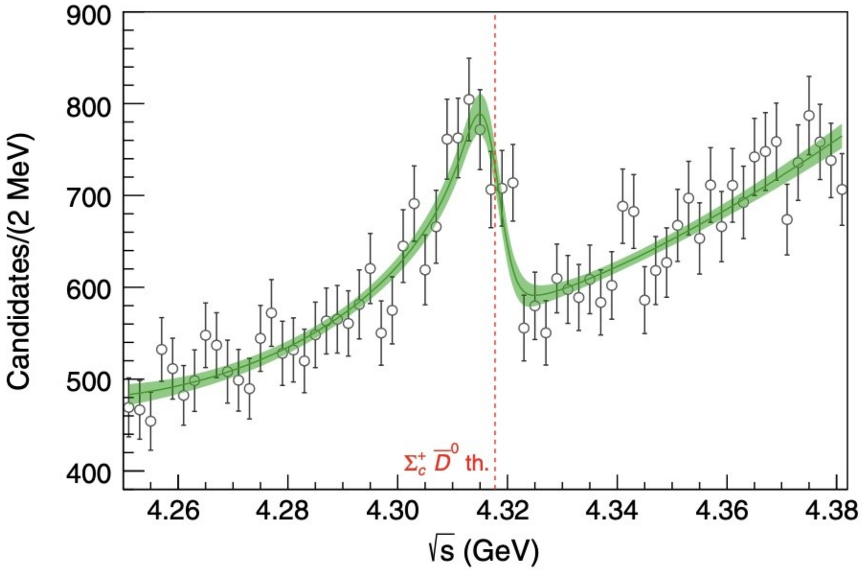
$$s_1 = (m_p + m_{J/\psi})^2 \quad \text{and} \quad s_2 = (m_{\Sigma_c^+} + m_{\bar{D}^0})^2$$

- Unitarity implies that M_{ij} is free from singularities near thresholds s_1 and s_2 so that it can be Taylor expanded *Frazer, Hendry Phys. Rev. 134 (1964)*

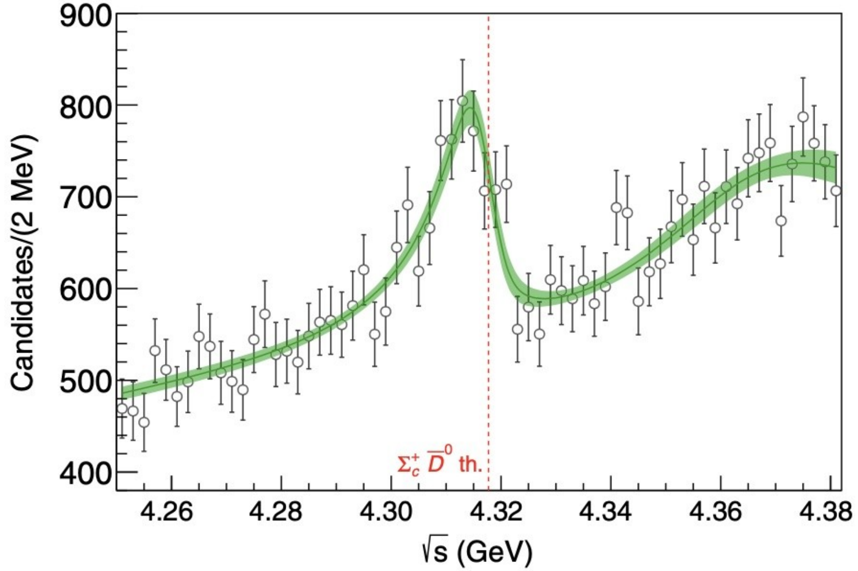
$$M_{ij}(s) = m_{ij} - c_{ij}s$$

- In principle the off-diagonal term of the amplitude $P_2(s)T_{21}$ could be included but we are interested in the analytical structure ("denominator") – so it's effect can be absorbed to the background and production terms.

Physics model – final version



Scattering length approximation



Effective range approximation

See *C. Fernandez-Ramirez Phys.Rev.Lett. 123 (2019) 9, 092001*

Finally we use the scattering length approximated amplitude as the basis for ML model

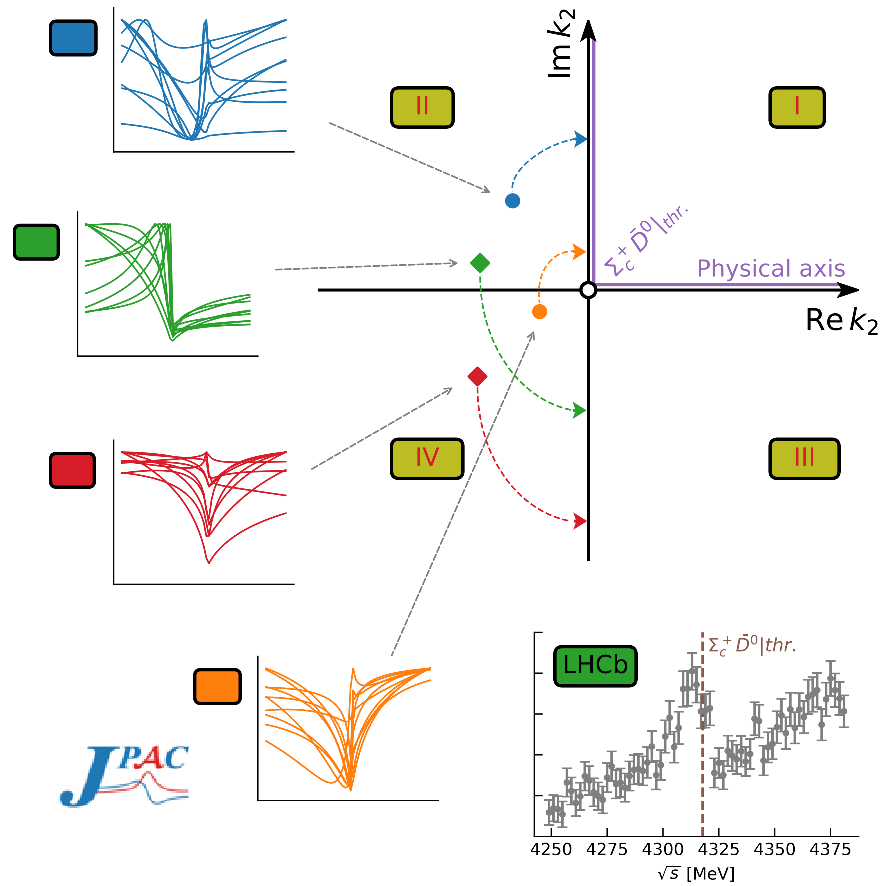
$$T_{11} = \frac{m_{22} - ik_2}{(m_{11} - ik_1)(m_{22} - ik_2) - m_{12}^2}$$

7 model parameters in total: m_{11}, m_{22}, m_{12}, p_0, p_1, b_0, b_1.

AGH

ML model – general idea

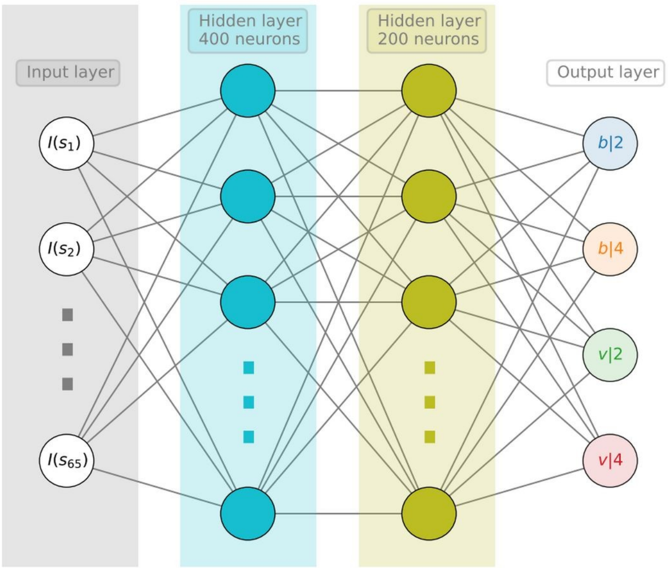
- From the physical model we produce:
  - Sample intensities (computed in 65 energy bins) – produced with randomly chosen parameter samples – **examples**
  - For each parameter sample we are able to compute the **target class** – one of the four: b|2, b|4, v|2, v|4
  - Symbolically:



$$K : \{[I_1, ..., I_{65}](m_{11}, m_{22}, m_{12}, p_0, p_1, b_0, b_1)\} \to \{b|2, b|4, v|2, v|4\}$$

ML model – MLP

Layer	Shape in	Shape out
Input		(B, 65)
Dense	(B, 65)	(B, 400)
Dropout(p=0.2)	(B, 400)	(B, 400)
ReLU	(B, 400)	(B, 400)
Dense	(B, 400)	(B, 200)
Dropout(p=0.5)	(B, 200)	(B, 200)
ReLU	(B, 200)	(B, 200)
Dense	(B, 200)	(B, 4)
Softmax	(B, 4)	(B, 4)



Training dataset preparation:

1. Parameters were uniformly sampled from the following ranges: b_0 = [0 ; 700], b_1 = [-40 ; 40], p_0 = [0 ; 600], p_1 = [-35 ; 35], M_{22} = [-0.4 ; 0.4], M_{11} = [-4 ; 4], M_{12}^2 = [0 ; 1.4]
2. The signal was smeared by convolving with experimental LHCb resolution:

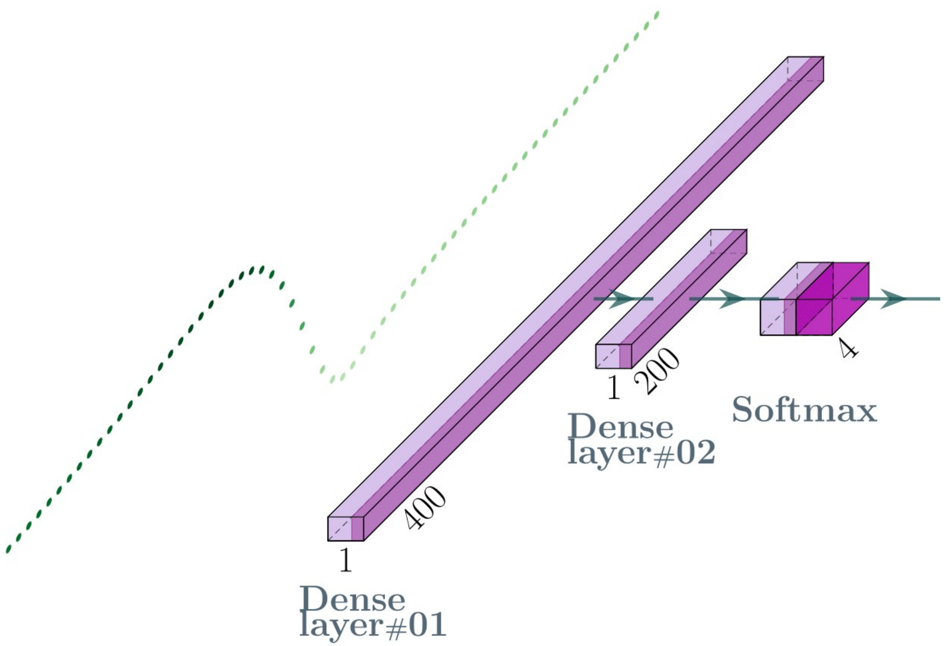
$$I(s) = \int_{m_\psi+m_p}^{m_{\Lambda_b}-m_K} I(s')_{\text{theo}} \exp\left[-\frac{(\sqrt{s}-\sqrt{s'})^2}{2R^2(s)}\right] d\sqrt{s'} \Bigg/ \int_{m_\psi+m_p}^{m_{\Lambda_b}-m_K} \exp\left[-\frac{(\sqrt{s}-\sqrt{s'})^2}{2R^2(s)}\right] d\sqrt{s'}\,,$$

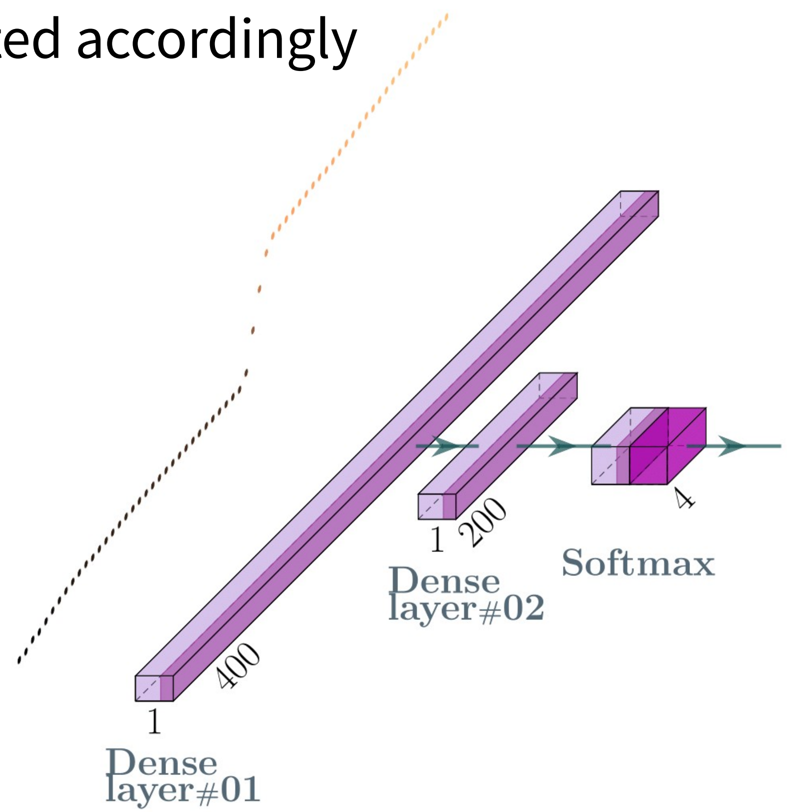
$$R(s) = 2.71 - 6.56\times 10^{-6-1} \times \left(\sqrt{s} - 4567\right)^2$$

3. To account for experimental encertainty the 5% gaussian noise was added

Caveats (on using MLPs)

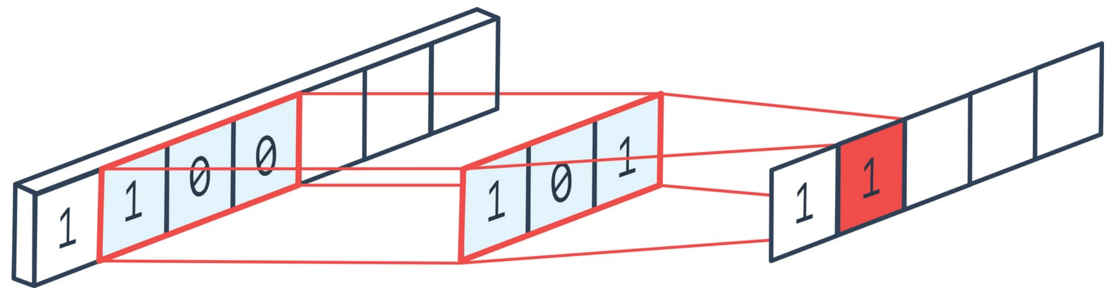
- Even though we want to recognize ordered sets (series) of data, the MLP rather recognizes just sets
- One can permute the data arbitrarily and get basically the same classification quality
- Provided the prediction dataset is permuted accordingly

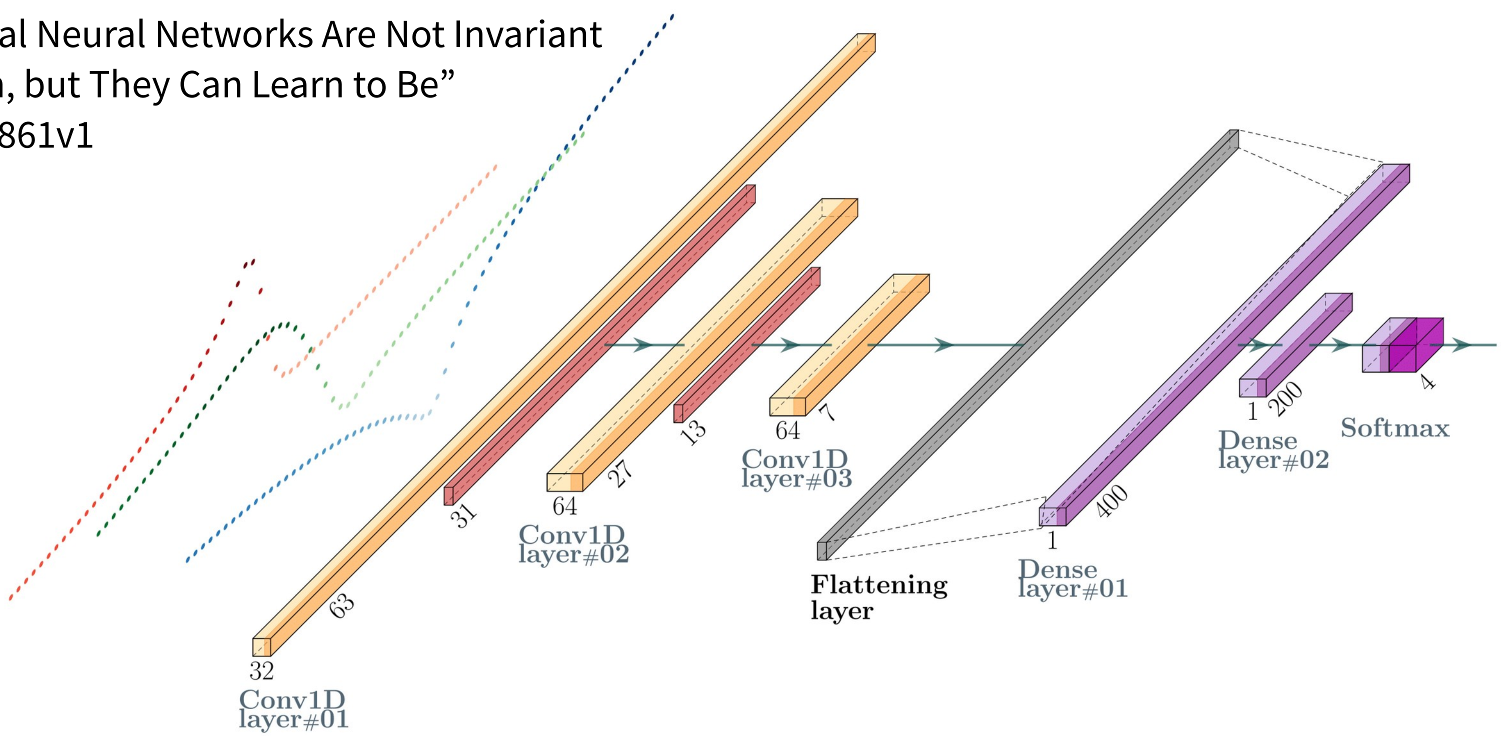




CNN as an alternative

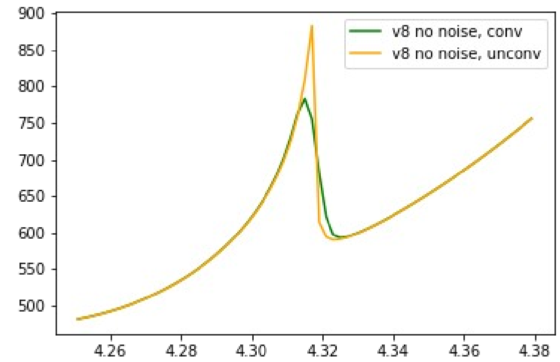
- Convolution neural network is able to detect local patterns
- Unfortunately it does it in fixed location (it's not translationally invariant)
- There are some (partial) remedies however: V. Biscione, J. S. Bowers, "Convolutional Neural Networks Are Not Invariant to Translation, but They Can Learn to Be" arXiv:2110.05861v1

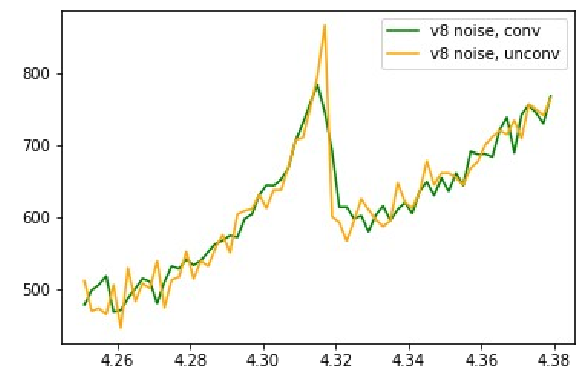




ML model - training

- Input examples (effect of energy smearing and noise):





- Computing target classes:
  - $m_{22}>0$ – bound state, $m_{22}<0$ – virtual state
  - To localize the poles on Riemann sheets we need to find zeros of the amplitude denominator in the momentum space:

$$p_0 + p_1\, q + p_2\, q^2 + p_3\, q^3 + q^4 = 0$$

with

$$p_0 = (s_1 - s_2)\, m_{22}^2 - \left(m_{12}^2 - m_{11} m_{22}\right)^2$$

$$p_1 = 2\,(s_1 - s_2)\, m_{22} + 2 m_{11}\left(m_{12}^2 - m_{11} m_{22}\right)$$

$$p_2 = -m_{11}^2 + m_{22}^2 + s_1 - s_2$$

$$p_3 = 2 m_{22}$$

Then poles appear on sheets defined with (η_1,η_2) pairs:

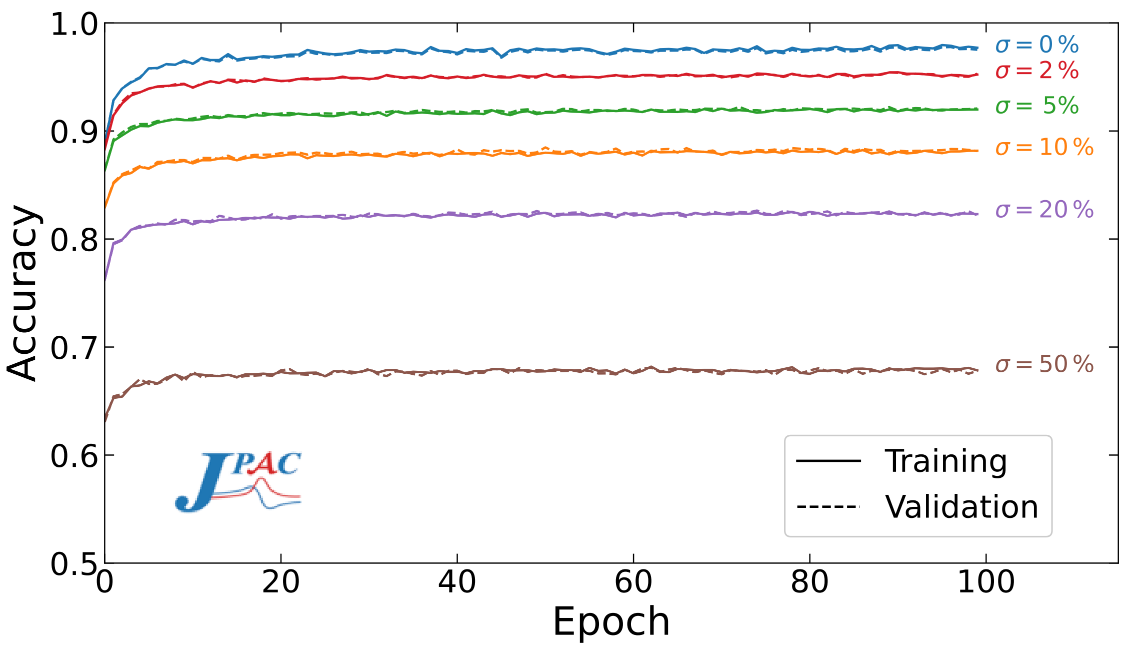
(-,+) - II sheet

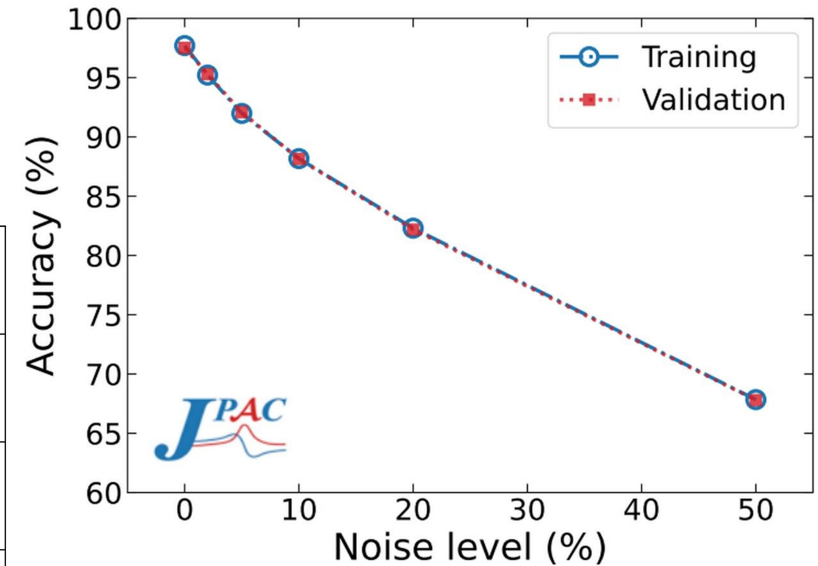
(+,-) - IV sheet

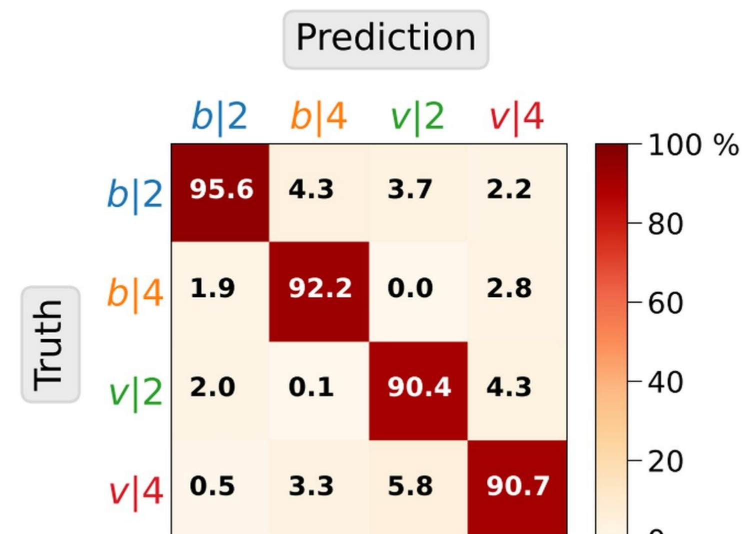
$$\eta_1 = \text{Sign Re}\left(\frac{m_{12}^2}{m_{22}+q} - m_{11}\right) \quad \eta_2 = \text{Sign Re} q$$

ML model – training results

Accuracy for various noise levels



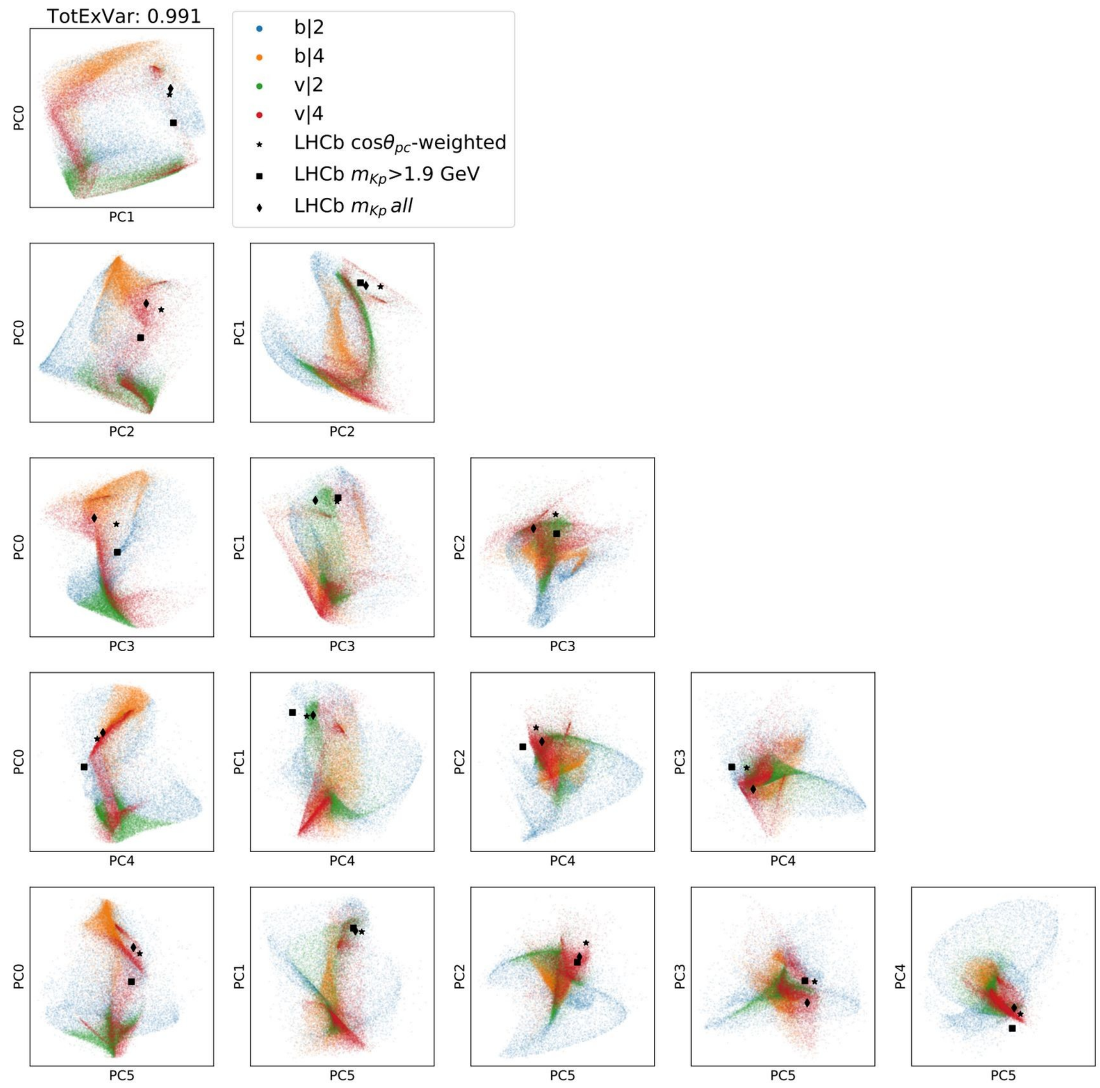




Confusion matrix for the 5% noise

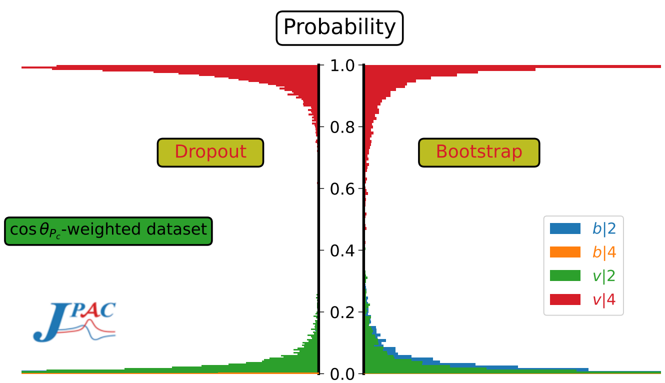
Feature refinement

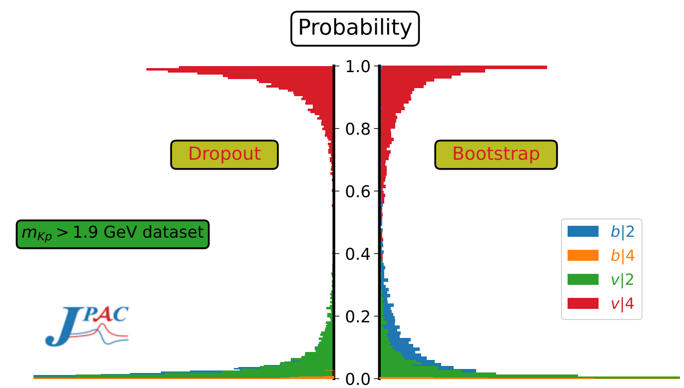
- Dimensionality reduction - Principal Component analysis
- Over 99% of the variance can be explained with just 6 features
- Experimental data projected onto principal components are well encompassed within the training dataset

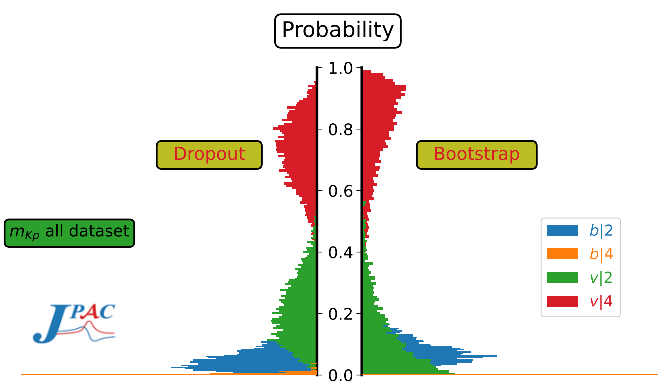


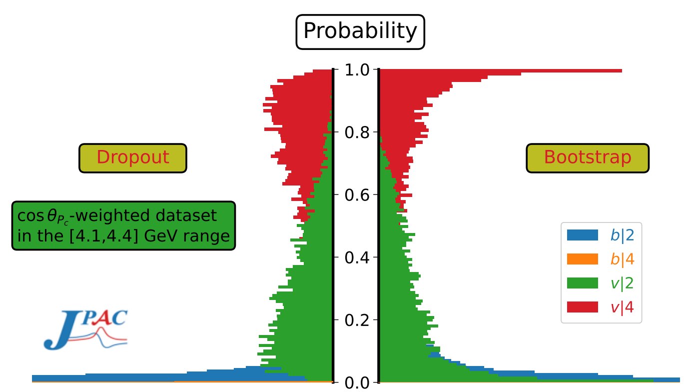
Model predictions – statistical analysis

- The distribution of the target classes was evaluated with
  - the bootstrap (10 000 pseudo-data based on experimental mean values and uncertainties) and
  - dropout (10 000 predictions from the ML model with a fraction of weights randomly dropped out)





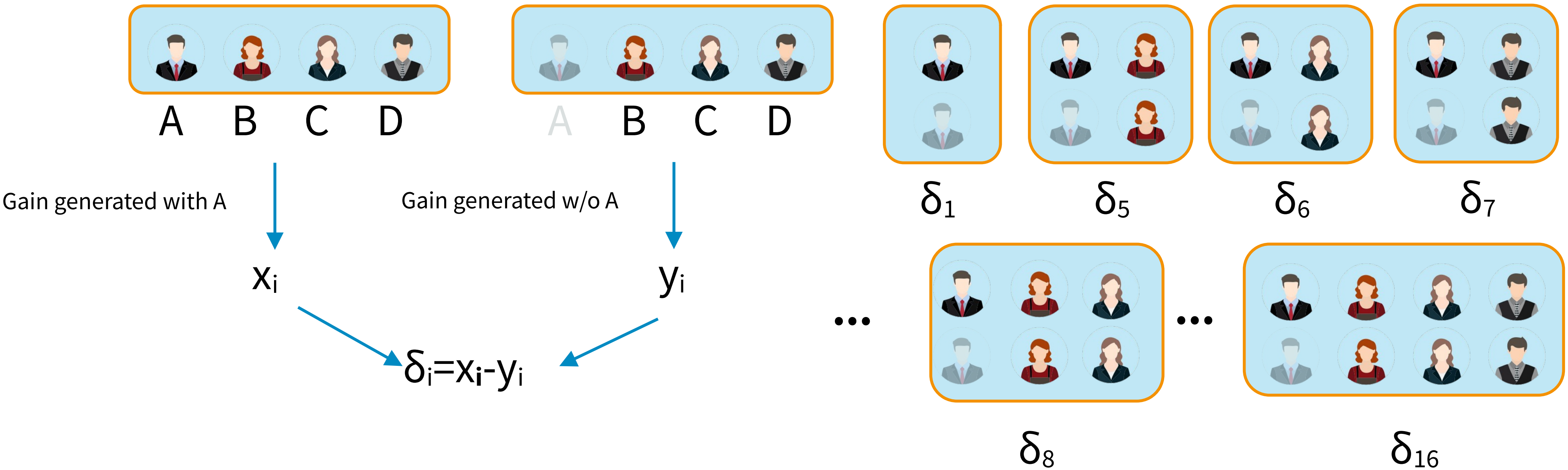




Model explanation with SHAP

- Shapley values and Shapley Additive Explanations

Shapley, Lloyd S. "Notes on the n-Person Game -- II: The Value of an n-Person Game" (1951)



Shapley value for member A: $$\phi_A = \frac{\delta_1 + \delta_2 + \cdots + \delta_{16}}{16}$$

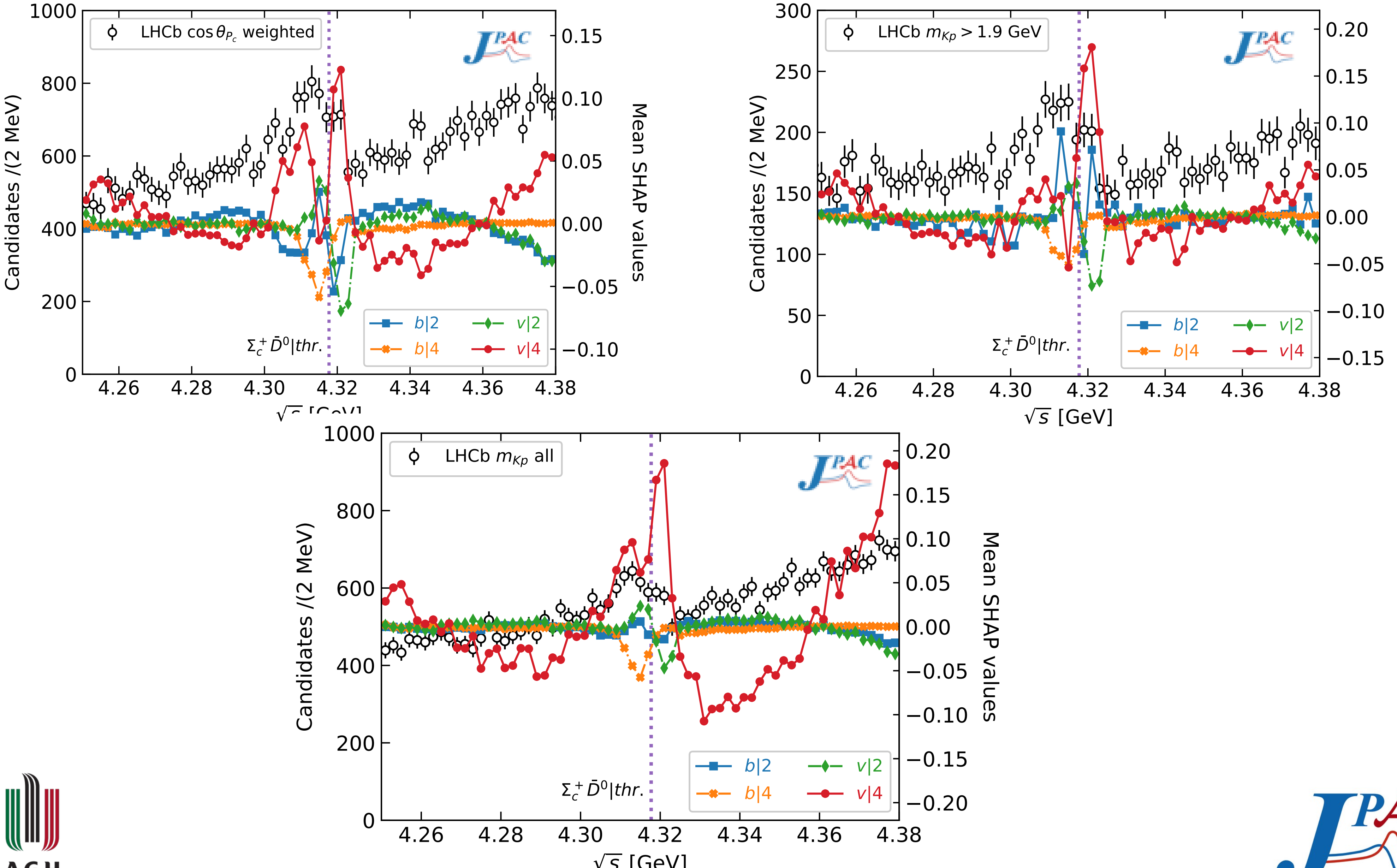
Model explanation with SHAP

- By making an association:
  - Member of a coalition → Feature
  - Game → Function that generates classification/regression result
  - Gain → Prediction
  - We define the Shapley values for features
- Caveats:
  - A number of possible coalitions grows like 2^N
  - Prohibitively expensive computationally (NP-hard)

Solution: Shapley additive explanations (Lundberg, Lee, arXiv:1705.07874v2, 2017)

AGH

Model explanation with SHAP

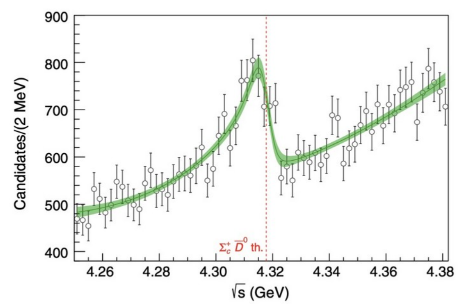


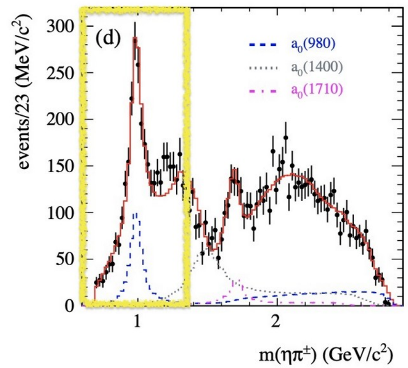
Summary

- Takeaways:
  - Standard χ^2 fit may be unstable, since small change in the input may result in large parameter fluctuations (change physics interpretation)
  - Rather than testing the single model hypothesis with χ^2, we obtained the probabilities of four competitive pole assignments for the $P_c(4312)$ state
  - The approach was model independent – meta model
  - By the analysis of the SHAP values we obtained an *ex post* justification of our scattering length approximation

Questions to be addressed

- Going beyond the limited generalization power - applying the method for larger class of resonances, described by the same physics, eg. $a_0/f_0(980)$ or other resonances located near thresholds





- Eg. we believe that these two resonances can be described by the same physics
  - MLPs and CNNs require inputs of the same size – re-binning required (but also kinematics and resonance parameters change: masses, widths, thresholds, phase spaces,...)
  - Alternatively we can use the length of the signal as part of the input information for RNNs
  - Difference between the models is not always as clear as above (different Riemann sheets) – need for model selection criteria.