University of BRISTOL

DEPARTMENT OF PHYSICS

FINAL YEAR PROJECT

NAME:	Toby Twigger
DEGREE COURSE:	MSci Physics
PROJECT TITLE:	Isolating quantum correlated charm systems for new constraints on symmetry violation
YEAR OF SUBMISSION:	2019
SUPERVISOR:	Paras Naik
NUMBER OF WORDS:	7582

Declaration

All data generation through RapidSim was completed by the project supervisor, Dr. Paras Naik. Through this, the focus on $D^0\bar{D}^0$ mesons was also suggested. Initially, code written by Jenny Ogden was adapted; this was later rewritten and bears very little resemblance to the initial code. All data reported on was generated ourselves, as was the analysis of the results.

My lab partner, Tom Williams, focused predominantly on the generation of plots and further analysis of predictions made by `Keras`, as well as the classification algorithm used in the final code. My focus was on creating a high-level API with which to interact with `Keras` to allow for a network to be adapted significantly easier, as well as implementing the grid search algorithm for the network layer configuration and attempting to upgrade this to a custom evolutionary algorithm to allow for much faster configuration setting.

Contents

Acknowledgements

Many thanks to our project supervisor, Dr. Paras Naik, for all his guidance and help throughout this project, and for the generation of simulated data. Thanks also to Dr. Jonas Rademacker for his support and help, and colleagues of Paras who offered us guidance as to network classification methods.

Most of all, thanks to my lab partner Tom Williams for his dedication and insight into the direction of the project and high level of commitment to ensuring the project ran as smoothly as possible.

New constraints on symmetry violation via quantum-correlated charm systems

Toby Twigger

August 7, 2025

Abstract

The use of a multilayer perceptron neural network for the classification of CP eigenstates of $D^0\bar{D}^0$ mesons generated through the decay of $X_{c1}(3872)$ is evaluated. The network is found to be able to classify $X_{c1}(3872)$ decays producing a π^0 as a secondary decay product with high precision. The separation of γ-producing $X_{c1}(3872)$ decays from a simulated background signal is found to result in a significantly lower degree of separation.

1 Introduction

In most branches of physics, theories are introduced and adapted as experimental data confirms or invalidates physical consequences of the theory. For example, Newtonian gravity was widely accepted when published in 1687, but it failed to describe accurately the motion of Mercury. In 1915, Einstein's Theory of General Relativity superseded Newtonian gravity, and was able to predict the correct orbital motion [1].

The evolution of particle physics has followed a similar path, with it roots said to be in the musings of Greek philosophers Leucippus and his better known student Democitius [2]. As technology adapts and improves, it has been possible to delve further into the inner workings of particles and atoms, and construct a wider picture to account for new experimental results.

However, the full theory still remains hidden. Despite many groundbreaking theoretical proposals, and many experimental confirmations, the Standard Model (SM) as it stands is still incomplete. Some of the problems, amongst many, are that: it cannot be reconciled with general relativity, does not explain or predict the existence of dark matter and dark energy, and is unable to describe the abundance of matter when compared to antimatter [3]. The discrepancy in the ratio of observed matter to antimatter is one of the most puzzling questions posed to theoretical and experimental physicists. There is no obvious reason for this symmetry violation, thus it holds heavy implications for the completeness of the SM. It is therefore highly likely there is undiscovered physics which would illuminate and hopefully solve this observed but unexpected violation of CP symmetry.

To narrow the search for CP violating physics, it is essential to constrain measured parameters which dictate the magnitude of the violating effects. One such fundamental constant used for a quantifying measurement of CP violation within the SM is named γ, and is the least precisely measured CP violating phase [4]. By imposing further constraints on this constant γ, it is hoped that indicators to new physics will become clear, and guide the evolution of the SM towards a more accurate and complete theory.

CP violation has been observed and measured before, but only very recently has it been seen in the charm system [5]. The CPT theorem states that the combined symmetries of Charge (C), Parity (P) and Time (T) should be constant [6]. Both CP and T violation have been independently measured in the B system through utilisation of B-meson mixing [7, 8]. However, through similar techniques in the charm system, specifically the quantum-correlated charm system $D^0\bar{D}^0$, the constraints imposed on the SM phase γ will hopefully be further increased.

Historically, these $D^0\bar{D}^0$ systems have been produced through e^+e^- collisions at the $\psi(3770)$ threshold [9, 10]. However, such methods exist to create large numbers of quantum correlated $D^0\bar{D}^0$ systems through decays of the exotic meson $X_{c1}(3872)$ [11]. Machine learning can be employed to aid identification of these decays and isolate $X_{c1}(3872)$ produced $D^0\bar{D}^0$ systems for measurement to lead to precise constraints on the SM phase γ.

2 Theory

2.1 CPT Violation in the SM

Through attempting to find the SM mechanism by which the observed degree of asymmetry between matter and antimatter is generated, Andrei Sakharov published three conditions to which a system must abide in order to produce a majority of matter particles over antimatter particles [12, 13]. The second of these conditions states that a system must violate CP-symmetry.

CP violating sources have been theorised and measured in the SM, but the combination of sources does not account for the amount of CP violation required to describe the observed matter and antimatter quantities [14]. One of these sources, the CKM matrix, governs the strength of flavour changing weak interactions [15]. The Wolfenstein parameterisation of the CKM matrix (Equation 1) is a commonly used parameterisation, correct to $\mathcal{O}(\lambda^3)$, around the small parameter $\lambda \approx 0.22$, and requires four parameters: λ, ρ, η and A [16, 17].

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1-\frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} \quad (1)$$

CP violation requires $\eta \neq 0$; the constraints on ρ and η are commonly represented in the complex plane, where they form the apex of the unitary triangle (figure 1).

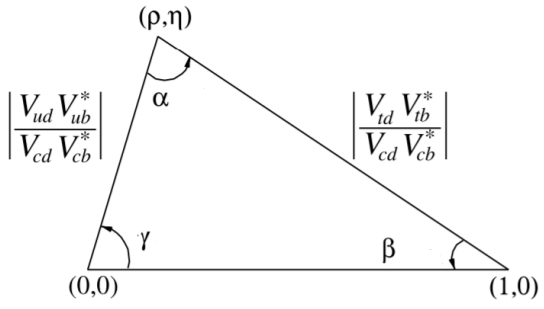

Figure 1: Standard unitary triangle representation of the CKM matrix elements in the complex plane.

Current measurements of the angle γ place the value at $\gamma = 74.0^{+5.0}_{-5.8}$ [18]. These measurements come from tree-level decays through B meson decays. Specifically, the interference between the favoured $b \to cW$ and the suppressed $b \to uW$ transitional amplitudes are exploited [19].

Commonly, the decays $B^- \to D^0 K^-$ and $B^- \to \bar{D}^0 K^-$ are used for the interference calculations. Charged mesons are used since they are easier to identify and are not involved in B-mixing. The Feynman diagrams for these decays are shown in figure 2 [20].

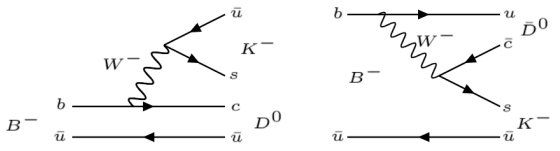

Figure 2: A Feynman diagram of the decay of a B^- meson to a D^0 or $\bar{D}^0$.

The only contribution to CP violation, and in turn γ, comes from the V_{ub} element of the CKM matrix. Assuming the D^0 and $\bar{D}^0$ decay to a common final state f, γ may be extracted from the amplitudes defined in equation 2 [20]. The four strong phases present in the decay paths may be reduced to two relative phases, $\delta_B = \delta_u - \delta_c$ and $\delta_D = \delta_{\bar{F}} - \delta_f$. δ_B may be deduced through the yield of B^+ and B^-, and quantum correlated charm mesons may be used to measure δ_D [21].

$$\begin{aligned} A(B^- \to D^0 K^-) &= A_c e^{i\delta_c} \\ A(B^- \to \bar{D}^0 K^-) &= A_u e^{i(\delta_u - \gamma)} \\ A(D^0 \to f) &= A_f e^{i\delta_f} \\ A(\bar{D}^0 \to f) &= A_{\bar{f}} e^{i\delta_{\bar{f}}} \end{aligned} \tag{2}$$

The benefit of this technique is that only tree-level decays are utilised in the interference experiments [22]. These processes are very unlikely to be affected by new physics, such as heavier particles and quantum loops. They can therefore set a SM benchmark against which any further measurements of γ may be compared.

This fundamental benefit of using tree-level bottom decays to measure γ is also the reason for requiring further experimentation outside the bottom system. A new technique more open to new physics is required in order to develop further the SM and hopefully to reveal signs of physics beyond the SM through a difference in branching fractions or unexpected measurements of CP violation.

Another motivation for a new technique is measurements of T violation. CPT is assumed to represent an exact symmetry, meaning that if CP violation is measured in a system, an equal violation of T should also be present. Until 2012, no such violation had been experimentally verified in the B system. However the BaBar collaboration published a paper in which they directly observed T symmetry violation in the B system [8]. They used the method proposed by Bernabéu, Martínez-Vidal and Villanueva-Pérez' [23], which utilises the property of the $\Upsilon(4S)$ decay to produce ERP-entangled neutral B mesons. These B mesons are produced in an entangled state $|i\rangle$ as seen in equation 3.

$$\begin{aligned} |i\rangle &= \frac{1}{\sqrt{2}}\big[B^0(t_1)\bar{B}^0(t_2) - \bar{B}^0(t_1)B^0(t_2)\big] \\ &= \frac{1}{\sqrt{2}}\big[B_+(t_1)B_-(t_2) - B_-(t_1)B_+(t_2)\big] \end{aligned} \tag{3}$$

Here, the B meson states may be written in terms of flavour eigenstates and CP eigenstates. Observations of the transitions between these states allows for a measurement of T violation, and when performed BaBar found a significant level of T violation.

2.2 CPT Violation in the Charm System

Although CP and T violation measurements have been successful in the bottom system, very few experiments have been able to detect the same in the charm system. In fact, the first paper to confirm an observation of CP violation within the charm system was published in March 2019 by CERN [5].

To measure the CP violation, CERN used the decays $D^0 \to K^- K^+$ and $D^0 \to \pi^- \pi^+$ present in a pp collision. The flavour of the charm meson was inferred using the decay products from $D^*(2010) \to D^0 \pi^+$ and the D^0 producing decay $\bar{B} \to D^0 \mu^- \bar{\nu}_\mu X$, more specifically the charge of the pion and the muon.

The CP violation measured for the K producing and π producing decay was $[-18.2 \pm 3.2(stat.) \pm 0.9(syst.)] \times 10^{-4}$ and $[-9 \pm 8(stat.) \pm 5(syst.)] \times 10^{-4}$ respectively, thus confirming CP violation presence in the charm system.

2.3 Production of Correlated Charm Systems

To test CP violation in the charm system in a wider range of decay branches, the generation of a $D^0\bar{D}^0$ system in an eigenstate of C is required. Quantum correlated $D^0\bar{D}^0$ mesons are usually generated from e^+e^- collisions at the $\psi(3770)$ threshold. However, to ensure the $D^0\bar{D}^0$ system is in a CP-even state, which makes measurements of T violation significantly easier, a virtual photon must be emitted during the decay [9]. This significantly reduces the branching fraction.

Fortunately, following the confirmation of the fixed quantum numbers $J^{PC} = 1^{++}$ of the exotic meson $X_{c1}(3872)$ [24], this may now be used as a new source of quantum correlated $D^0\bar{D}^0$ systems generated in an eigenstate of C.

The decay of an $X_{c1}(3872)$ follows the decay pattern in equation 4, where m and n are integers [9].

$$X_{c1}(3872) \to D^0\bar{D}^0 + m\gamma + n\pi^0 \tag{4}$$

The CP even and CP odd eigenstates of the $D^0\bar{D}^0$ system are shown in equations 5 and 6 respectively.

$$\frac{\big|D^0\bar{D}^0\big\rangle + \big|\bar{D}^0 D^0\big\rangle}{\sqrt{2}} \tag{5}$$

$$\frac{\left|D^0\bar{D}^0\right\rangle - \left|\bar{D}^0 D^0\right\rangle}{\sqrt{2}} \tag{6}$$

Another benefit of using the $X_{c1}(3872)$ decay over the $\psi(3770)$ decay is that the exact configuration of the $D^0\bar{D}^0$ pair, when produced with an $X_{c1}(3872)$, depends on the other particles produced in the decay. To ensure a pure CP eigenstate is generated, the condition $m = 0, n = 1$ or $m = 1, n = 0$ must be met [25]. By monitoring the secondary decay product, one can tag the $D^0\bar{D}^0$ system as having a CP eigenvalue of $+1$ or -1.

However, LHCb was designed to detect charged particles with high precision. This results in neutral particles being very hard to reconstruct accurately. By analysing reconstructable information extracted from the $X_{c1}(3872)$ decay through the utilisation of a neural network, it may be possible to determine the decay product produced, and so the configuration of a $D^0\bar{D}^0$ pair, without having to reconstruct these neutral particles.

2.4 Introduction to Neural Networks

A mutlilayered perceptron neural network (figure 3) takes its inspiration from the human brain. It is made of a set of interconnected nodes, and allows for complex relationships between input variables to be analysed in such a way as to produce an activation of an output node corresponding to a particular classification [26]. This allows for a set of data to be passed to the network, and through training allows the network to understand which input variable values signify a particular output.

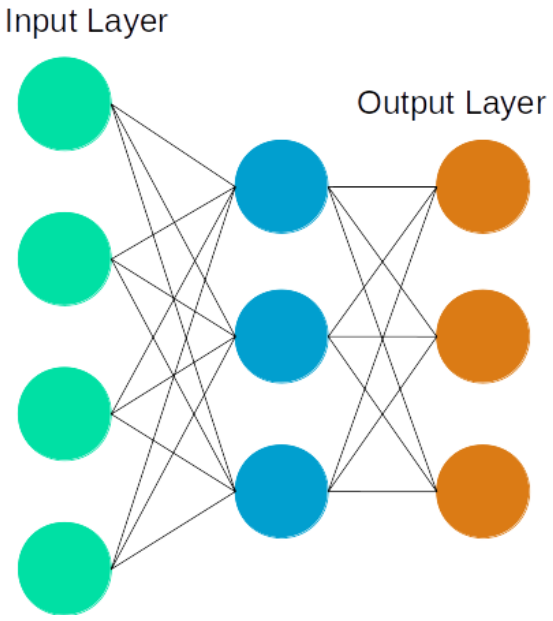

Figure 3: A representation of a multilayer perceptron neural network. Each circle corresponds to a single node, and each line is a connection between two nodes.

The input layer contains as many nodes as input variables. Each variable is passed to a single node, which performs a calculation on the data and outputs a value to the next layer. The calculation is dictated by the activation function, whose shape is altered through training by adjusting the bias and weighting of the node.

This is mathematically governed by the matrix in equations 7, 8 and 9, where m is the number of nodes in the first layer and n the number of nodes in the second layer. The input data passed to the first layer is represented by $\boldsymbol{a}^{(0)}$. Each of the nodes is also assigned a bias, given by $\boldsymbol{b}$. Through the connections between each node in the first and second layer, given weightings defined in $\boldsymbol{W}$, the network may populate each layer with a new calculated dataset given by $\boldsymbol{a}^{(l)}$, where l is the layer.

$$\boldsymbol{a}^{(0)} = \begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ \vdots \\ a_{m^{(0)}} \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} \tag{7}$$

$$\boldsymbol{W} = \begin{bmatrix} w_{0,0} & w_{0,1} & \dots & w_{0,m} \\ w_{1,0} & w_{1,1} & \dots & w_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,0} & w_{n,1} & \dots & w_{n,m} \end{bmatrix} \tag{8}$$

$$\boldsymbol{a}^{(l)} = \sigma(\boldsymbol{W}\boldsymbol{a}^{(l-1)} + \boldsymbol{b}) \tag{9}$$

Training the network can occur through backwards propagation of errors. Every weight is randomly generated and assigned when the network is created. The first set of training data is then inputted into the network and the output observed. Since training data is labelled, a comparison between the predicted output and the true result can be made through a loss function, which is then passed back through the network and weights adjusted to minimise this loss function. This process is done many times, with the hope that after training, previously unseen data may pass into the network and be correctly classified.

By inputting reconstructed data from LCHb or simulated events, it is hoped that a neural network will be able to identify correctly if an $X_{c1}(3872)$ decay produced a π^0 or a γ without providing the network with any data measured directly from these decay products. By training the network on a simulated sample set in which the $D^0\bar{D}^0$ configuration is given, the weights should be adjusted accordingly and be able to discover relationships between input variables which are too complex for a human to find.

3 Experimental Details

3.1 Simulation of $X_{c1}(3872)$ decays

To generate data that the neural network can use to train, `RapidSim` was used [27]. This is an open-source software package which simplifies the simulation of heavy-quark hadron decays, and is built on the `ROOT` framework [28]. To ensure the focus of the network development could fall on utilising tools to increase the

accuracy of the network and not on simulating accurate data, simple non-resonant direct prompt decays of the $X_{c1}(3872)$ were created.

Each event generated a set of variables which were saved into a `.root` file. Using `Python` with the `ROOT` framework, variables which were to be used for the neural network could then be extracted and saved as a `Numpy` array for processing with the high level package `Keras` [29] using `TensorFlow` [30].

Each `.root` file contained data relevant to a single $D^0\bar{D}^0$ configuration. These were merged to create the `Numpy` array, but since there was an initial separation it was trivial to label each event as having produced a π^0, γ or being background.

Of course, the aim of the simulation was to recreate accurate data. In any detector, there is an issue of background and detector resolution. Detector resolution was at first simulated by adding a 1% gaussian variation to the input data, but later was included directly within the simulation.

The background was generated through `RapidSim`, and made use of $D^0\bar{D}^0$ pairs produced through a prompt decay. However, a sharp cut-off was implemented at approximately $3.8GeV/c^2$ (figure 4). This was done since the $X_{c1}(3872)$ produced $D^0\bar{D}^0$ mesons fall into a narrow mass range, whereas the background $D^0\bar{D}^0$ mass was much more varied. It was thought that this initial separation in invariant mass would render identification of heavier background $D^0\bar{D}^0$ mesons trivial to the network assuming the invariant mass was an input variable, and thus a generation of heavy background mesons would not make the simulation any more realistic.

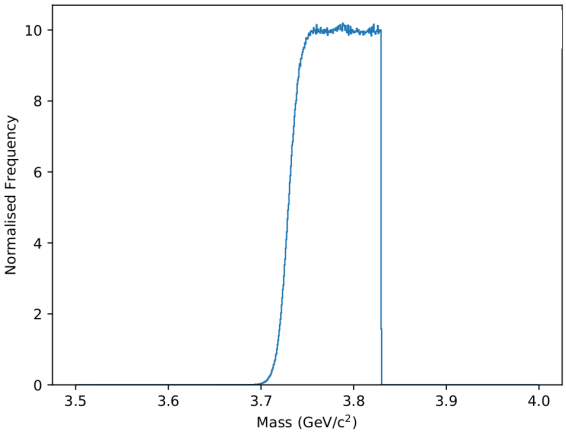

Figure 4: The invariant mass of $D^0\bar{D}^0$ mesons produced by background simulations. A non-physical cut can clearly be seen at the high mass range.

Unfortunately, this was later found to be untrue; some prompt produced signal decays, especially for γ-producing signal, had $m_{D^0\bar{D}^0}$ greater than the applied cut-off (figure 5) so this cut off was eventually extended to cover the whole range of the prompt produced $D^0\bar{D}^0$ masses (figure 6).

Another non-trivial simulation deficiency applicable to the background data was a lack of final-state radiation (FSR). FSR occurs when gluon or photon radiation after the initial decay significantly alters the event topology. Since the $X_{c1}(3872)$ decays did include FSR, it was found that the signal tended to have a longer tail below the peak $D^0\bar{D}^0$ mass, whereas the background contained a much sharper cutoff, as shown in figure 7. It was thought that this could have an effect on the identification in the network, since the network could learn that background had a shorter tail than signal and so more accurately separate the signal and background. The FSR effect was eventually removed from the signal to ensure that simulation effects would not affect the physics (figure 8).

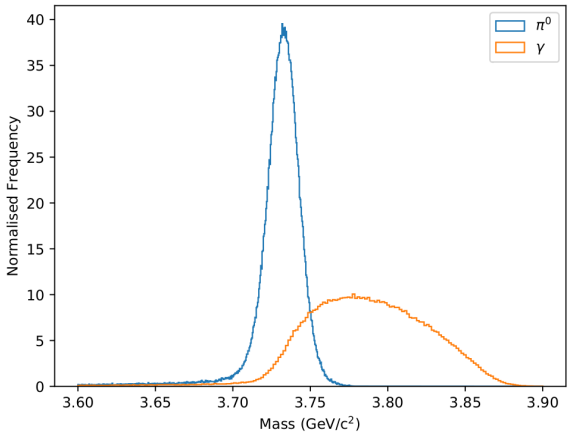

Figure 5: The invariant mass of $D^0\bar{D}^0$ mesons produced by π^0-producing and γ-producing decays of $X_{c1}(3872)$.

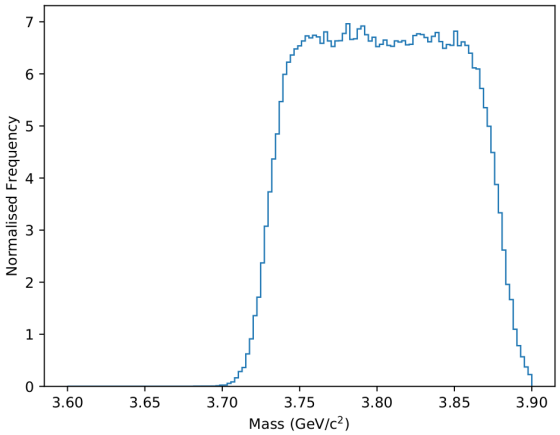

Figure 6: The invariant mass of $D^0\bar{D}^0$ mesons produced by background simulations. The non-physical cut has been removed and so the mass distribution spans the entire relevant range of $D^0\bar{D}^0$ mesons.

3.2 Evolution of the Neural Network

There were multiple stages involved in optimising the neural network to provide maximal separation in decay classification. By altering each configuration parameter of the network, a quantatitive effect of the parameter on the network accuracy could be established.

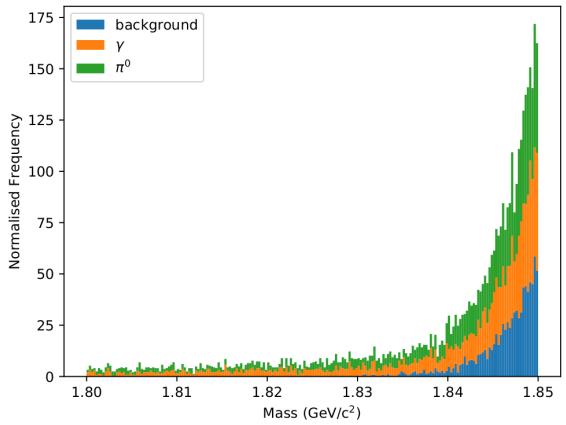

Figure 7: The invariant mass of the D^0 meson for a limited mass range for each decay classification. The tail caused by FSR for the π^0 and γ-producing decays can be clearly seen to extend past that of the background.

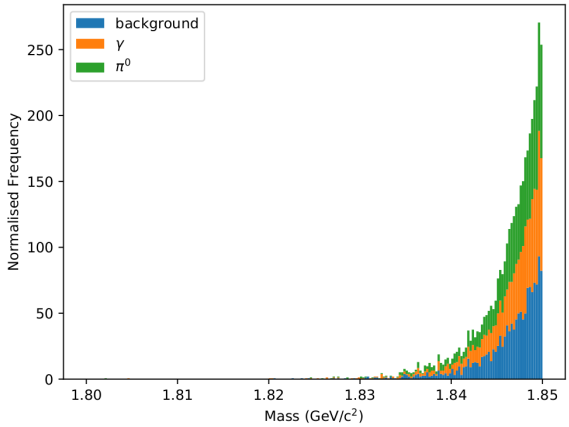

Figure 8: The invariant mass of the D^0 meson for a limited mass range for each decay classification. The lack of an FSR tail for all three decay types is clear.

3.2.1 Input Variables

One of the most fundamental configuration parameters of a neural network is the inputs. Neural networks tend to achieve higher accuracies when presented with separable data. Each input variable provides the network with an additional dimension of separability. For example, suppose a set of events were indistinguishable when comparing the invariant mass of the produced $D^0\bar{D}^0$ mesons. Given another parameter, such as the transverse momentum of the $D^0\bar{D}^0$, which shows a difference in behaviour related to the CP eigenvalue of the $D^0\bar{D}^0$, the network may use this second dimension to make a clean cut between the datasets and so accurately classify the decay paths.

Figure 9 demonstrates this graphically. Using a single dimension, the data would be very hard to separate. The introduction of an additional parameter with a higher degree of separability allows for a much higher classification accuracy.

Since there were 39 variables available in the simulation data towards the end of the project, it was hoped that a number of these would have a high degree of separation. By inputting all available variables,

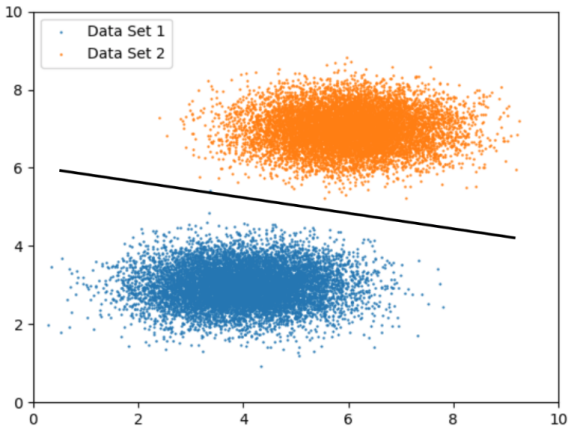

Figure 9: A representation of data separability through additional dimensions. Only by adding an additional dimension can a clear cut be made, since the data distributions overlap in one dimension.

the weight associated with each one may be analysed through a summation of the $\boldsymbol{W}$ matrix from equation 8 corresponsing to the particular input value. Through this method, a set of highly separable variables may be found without requiring an excessive load on the network.

3.2.2 Network Configuration

The network configuration itself also affected the results. The network used was configurable in three main aspects: the number of layers and nodes, the batch size and the number of epochs.

The number of layers and nodes predominantly affects the depth of the relationships the network is sensitive to. In order to construct a simple method to refer to specific network layer configurations, the notation used lists out the size of each layer with a colon delimiter. Therefore, a network with an input layer with eight nodes, two hidden layers with four and two nodes respectively, and and one output node will be denoted as a 8:4:2:1 network.

The batch size specifies how much data the network should be given access to each training round, whereas the number of epochs alters the number of times the full dataset is passed to the network. A training round is a single forwards calculation and backwards propagation of errors, and so by adjusting the two parameters an optimal configuration may be found to ensure the network is able to train quickly without introducing errors in the gradient estimation.

3.2.3 Random Seed

When a network is created, as previously specified, the weights and biases assigned to nodes and connections between nodes are random. This initial random set makes use of the `TensorFlow` and `Numpy` random number generators, which in turn rely on a random seed. By manually setting the random seed, the reproducibility of the training of the network may be ensured.

Through investigating the effect on the random seed, an algorithm was designed to pick a selection of possible seeds from a pool and test each one, therefore allowing for a constant random seed for all model training, but one that will not adversely affect the accuracy of the data.

3.2.4 Dropout and Normalisation

Two additional techniques were used alongside the neural network in an attempt to further increase the accuracy of the predictions and the speed of training. These were dropout and normalisation.

Dropout randomly allows for individual input variables to be ignored for a particular training set, which prevents overfitting of the network. Through the training process, nodes can develop a co-dependency which reduces the contributions from individual nodes. Through the utilisation of dropout, no single node or set of nodes can become too important, and thus the network is able to train on a variety of variable relationships as opposed to becoming entirely dependent on a single relationship.

Normalisation allows a network to train significantly faster, and reduces the risk of the network getting stuck in a local minima. Given a multilayered network, the input variables statistical distribution will change over several iterations as each layer performs the standard weight and bias calculations. By applying batch normalisation, the mean of data passed to layers will always be 0, and hence the network may train on the same statistical distribution of data each iteration. Furthermore, batch normalisation handles the vanishing gradient problem. Given an activation function which is constrained by horizontal asymtotes, a non-normalised mean will require significant training to set the bias of the node activation function close to the mean. By normalising the mean initially to within the range the activation function is most sensitive to, the response of training to a small shift in the input value is much more significant.

3.3 Method

Throughout training and testing, the same method was used. The relevant data was extracted from the `.root` files using `Python`, before being passed into a Neural Network built with `TensorFlow` and `Keras`. The network was allowed to train, before previously unseen prediction data was passed in to evaluate the accuracy of the network.

3.4 Network Evaluation Methods

A fundamental tool for analysing a network's performance is being able to determine how good the network was at separating the data. The simplest evaluation method is simply outputting the accuracy at the end of a prediction. The accuracy is a measure of how many correct predictions the network made. However this metric should not be solely relied upon as it is heavily dependent on the inputs. Given an uneven distribution of classification data, for example a larger number of background $D^0\bar{D}^0$ mesons than $D^0\bar{D}^0$ mesons produced alongside a γ, the network is simply able to classify everything as background and achieve a higher accuracy than guessing. This would imply the network is able to separate the input data, when in reality it is just classifying everything as background.

To avoid this problem, there are a multitude of additional metrics one can use to analyse the performance of a neural network. By observing the confidence of each output node individually, easily separable decays may be inferred since there will be a greater difference in the confidence output. Furthermore, areas of weakness for the network may be identified, such as areas in which the network is equally confident in two results.

To build up a greater understanding of the ability of the network to predict data accurately, a Receiver Operating Characteristic (ROC) curve or Precision Recall (PR) curve may be used.

ROC curves can be used to evaluate a network accuracy when there is an equal number of events in each classification class. They plot the false positive rate, or the number of events classified as a specific class which are not from that class, against the true positive rate, or the number of events classified in the correct class compared to the whole set of events. The area under the curve (AUC) is a quantifiable representation of the network accuracy. However, ROC curves tend to present an optimistic view of the network accuracy given an imbalance in the event classification distribution [31].

PR curves handle this imbalance by plotting the precision against the recall instead. The precision is a ratio of the number of true positives over the sum of true and false positives, and thus represents a sample size independent measure of how many predicted classes are actually classified as that class. The recall is the ratio of true positives over the sum of true positives and false negatives, and instead represents the proportion of positives that the network was able to classify.

Through a combination of these methods, the accuracy of the network can be evaluated at each stage of its evolution to allow for conclusions to be drawn.

3.5 Network Classification Algorithm

Once a network has a given set of confidences from output nodes, an algorithm must be employed to allow the network to decide which classification to label the event as. A simple method of doing this is simply to allow the network to classify the event as whatever it was most confident about. However, given a set of confidences with similar magnitudes, this will not necessarily be the most accurate classification algorithm.

Instead, the precision and recall from a PR curve may be used to set a threshold above which a network is allowed to classify an event. The F-measure defines a harmonic mean which can be used to find an optimum balance between the precision and recall of a training set. Generally, an F_1 measure is used, which treats

recall and precision as equals and returns a measure of the accuracy of the network based on the average of these two measures.

In this case, precision is thought to be significantly more important than recall, since a small but precise sample is preferred to a larger but less accurate sample. Therefore, the $F_{0.5}$ measure was used, which favours precision. Equation 10 defines F_β for any $\beta \in \mathbb{R}_{>0}$.

$$F_\beta = (1+\beta^2)\frac{precision \cdot recall}{\beta^2 \cdot precision + recall} \tag{10}$$

4 Network Evolution and Evaluation

4.1 Random Seed

Since the initial network configuration is entirely dependent on the random seed chosen for `Numpy` and `TensorFlow`, which must be manually set to allow for reproducibility of results, the effect this seed has on the final accuracy of the network must be evaluated. Training was completed using the 4-momenta of the $D^0\bar{D}^0$ system from π^0-producing and γ-producing decays, and the accuracy measured for a range of random seeds, as seen in figure 10.

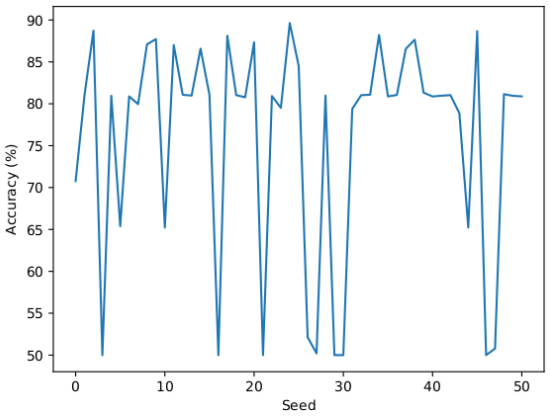

Figure 10: The effect of the initial random seed on the final accuracy of a MLP Neural Network. A large dependence can be seen.

Clearly, the initial random seed chosen can seriously affect the final accuracy of a network. Accuracies tend to fall either at the 50% level or between 80% and 90% aside from a few seeds. Since the random seed defines the initial configuration of the network, it also defines the local topology of the loss function, which the network attempts to minimise through training. Therefore, some seeds will place the network a distance from the true mimima, allowing the network to fall into a local minima and stop training.

Although a more complex network should dissipate some of this effect, since additional dimensions will lead to a more complex topology, it should be noted that most local minima are the same depth as global minima. Therefore, by simply allowing a selection of seeds to be tested and the one resulting in the highest accuracy selected before a training process starts will allow for optimum accuracy to be achieved.

Configuration	Accuracy (%)
8 : 8 : 7 : 1	81.03
8 : 8 : 7 : 4 : 1	80.94
8 : 8 : 7 : 4 : 4 : 1	80.96
8 : 8 : 6 : 4 : 1	80.91
8 : 8 : 6 : 4 : 4 : 1	90.73
8 : 8 : 6 : 4 : 2 : 1	86.02
8 : 8 : 6 : 5 : 4 : 1	80.88
8 : 8 : 6 : 3 : 4 : 1	65.39
8 : 8 : 6 : 3 : 3 : 1	65.35
8 : 8 : 5 : 4 : 1	80.84
8 : 8 : 5 : 4 : 4 : 1	89.52
8 : 8 : 4 : 4 : 1	81.05
8 : 8 : 4 : 4 : 4 : 1	80.96
8 : 8 : 4 : 4 : 4 : 4 : 1	87.76

Table 1: The accuracy of the network when trained with different network sizes. Values with very high accuracy are highlighted.

4.2 Network Configuration

Since the network configuration may be trivially adapted to suit a particular problem, an analysis of the size and shape of the network was carried out with the intent of maximising the accuracy of the network. The output and input layer shapes are decided by the information in the data, and so only the hidden layers need to be adjusted.

A grid search algorithm was written to enable a simple analysis of a set of hidden layer configurations. Table 1 was generated detailing the configurations attempted and the resulting accuracy for each one.

The most accurate network configuration was the 8 : 8 : 6 : 4 : 4 : 1 network. As with the random seed, minor modifications of the network configuration are able to affect the total accuracy significantly, and no clear pattern is observed. This implies that the grid search, as with the random seed, should be run before each training to determine the best configuration for the network given a particular set of input data.

In general, however, deeper networks tended to give higher accuracies. This implies a level of complexity in the data, since a deeper network is able to find more abstracted relationships between input data.

Further configuration testing focused on the batch size and the number of epochs. Through a similar testing method of training the network on a set of different configurations and recording the final accuracy, conclusions may be drawn about the dependence of the network on batch size and number of epochs.

Smaller batch sizes tended to converge quicker, but the network will take longer to train to the same accuracies. This is as expected, since given a smaller batch size, the algorithm which defines the backwards propagation of errors will be executed more times per epoch. This means that the gradient will make multiple

smaller updates as opposed to a single large gradient update when presented with a larger batch size, thus allowing for a mimima to be reached faster. However, since the network never trains on the complete training dataset, it may never settle at the true minima and instead just oscillate around the minima, thus there may be a sacrifice of accuracy for quicker convergence. Furthermore, smaller batch sizes take longer to run since they require the computational power to execute multiple backwards propagation of errors when compared to a gradient descent method with a single batch.

Large numbers of epochs are also considered to be preferable when taking into account accuracy. The longer a network trains for, the closer it can get to a true minima. However, an increase in the number of epochs leads to an increase in the time to train, and additional epochs often provide a smaller increase in accuracy since the gradient of the loss function reduces as a network approaches the minima.

To take into account and optimise all network configuration parameters, a more advanced algorithm could be written using an evolution technique. By starting with a set of different network configurations, including different hidden layer counts and sizes, different batch sizes and different numbers of epochs, and testing each one, the most accurate configuration may be taken and evolved into a set of child network configurations, each with only minor modifications to the network. This process may continue until an accuracy threshold is met. The advantages of this algorithm over the simple grid search implemented is that networks with fundamentally lower initial accuracy are instantly discarded, allowing for a much quicker optimisation of a network.

4.3 Input Variables

In designing a network to separate different decay products, arguably one of the most important parameters is the variables exposed to the network. By manipulating data from a reconstructed decay into different forms, or selecting different variables to input to the network, a variety of complex relationships between different variables may be indirectly deduced in the hidden layers of the network.

4.3.1 Invariant Mass

Initially the invariant mass of the $D^0\bar{D}^0$ pair, $m_{D^0\bar{D}^0}$, was calculated from the 4-momenta of each meson and inputted through a single node. The invariant mass is calculated using equation 11.

$$\begin{aligned} m^2_{D^0\bar{D}^0} &= E^2_{D^0\bar{D}^0} - p^2_{D^0\bar{D}^0} \\ E_{D^0\bar{D}^0} &= E_{D^0} + E_{\bar{D}^0} \\ \boldsymbol{p}_{D^0\bar{D}^0} &= \boldsymbol{p}_{D^0} + \boldsymbol{p}_{\bar{D}^0} \end{aligned} \tag{11}$$

Figure 11 demonstrates the invariant mass distribution for the set of events. There is a small overlap present, but some of the signal events should be classified correctly by implementing a cut in the overlap.

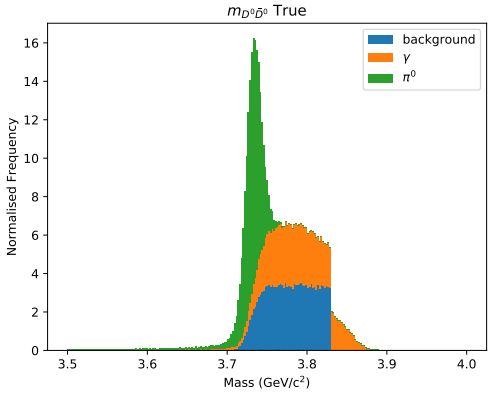

Figure 11: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair produced through π^0-producing, γ-producing and background decays.

For this reason, a sigmoid activation function is used as this can effectively simulate a cut. The background spans most of the mass distribution, so it is expected that this will be very difficult to separate, especially from the γ-producing decay.

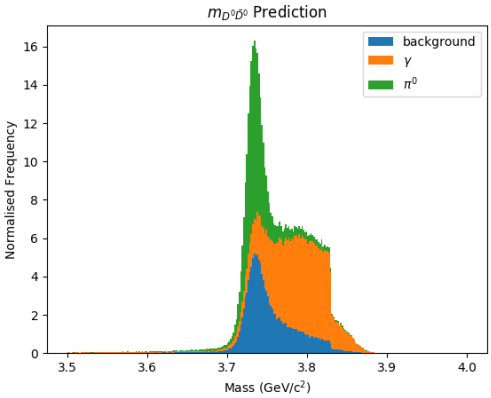

Figure 12: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair as classified by `Keras` as having been produced through π^0-producing, γ-producing and background decays.

From the predicted mass distributions in figure 12, it appears as though there have been a large number of misclassifications. The low mass tail of the π^0-producing decay is almost fully misclassified as a mixture of π^0-producing and γ-producing decays and background signal. This can be explained by considering that, since the network only has access to the invariant mass, and the $D^0\bar{D}^0$ meson mass from γ-producing decays and background signal span almost the entire mass range (albeit to a smaller degree at the low mass range), just through a mass cut very little classification should be possible.

The high mass tail of γ-producing decays above the mass cut implemented in the background is almost perfectly classified. This is unsurprising since the only $X_{c1}(3872)$ decay which is able produce a $D^0\bar{D}^0$ mass above this threshold will also produce a γ. The net-

work has simply made a mass cut and classified anything above this mass as a γ-producing decay.

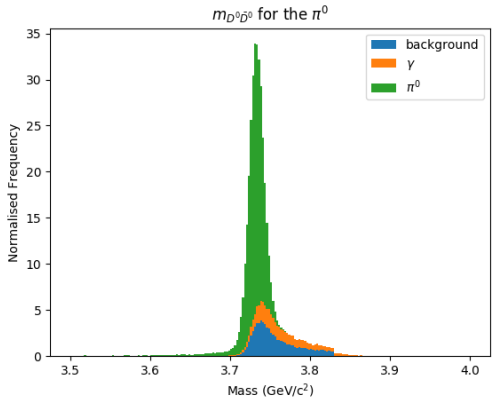

Figure 13: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair classified as having been produced through π^0-producing decays. The data has been grouped by the true classification of the event.

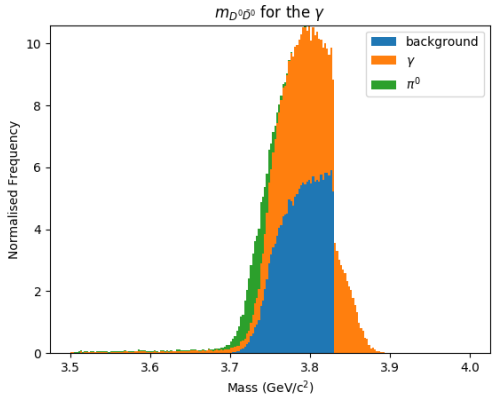

Figure 14: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair classified as having been produced through γ-producing decays. The data has been grouped by the true classification of the event.

The mass distribution of π^0-producing decays in figure 13 asserts that the separation of π^0-producing decays is reasonably difficult using only invariant mass. A significant number of misclassifications can be seen, and surprisingly these span the whole mass range although are less dominant at higher masses. The peak in the misclassifications corresponds to the π^0-producing decay mass, which implies that a soft mass cut has been made, in which the network is not simply classifying anything below a certain mass as a π^0-producing decay. This is because a significant proportion of the low mass signal is made of a mixture of decay types, and hence a hard mass cut will lead to a large proportion of misclassfications.

Misclassifications are also prevalent in figures 14 and 15. Gamma classifications appear to be very difficult to make, since a significant proportion of the classified γ-producing decays are actually background. Similarly, γ-producing decays are seen in the background classifications, thus implying the two look very similar.

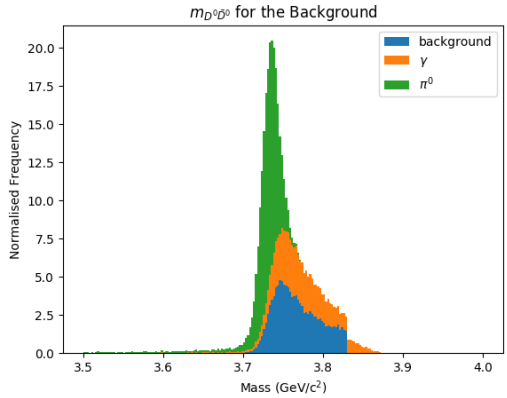

Figure 15: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair classified as having been produced through background simulation. The data has been grouped by the true classification of the event.

A surprising result is that of the misclassifications of π^0-producing decays as background. It is likely that there are a number of π^0-producing decays which have an invariant mass closer to that of the average background as opposed to within the standard range for the π^0-producing decay, thus leading the network to classify them as background.

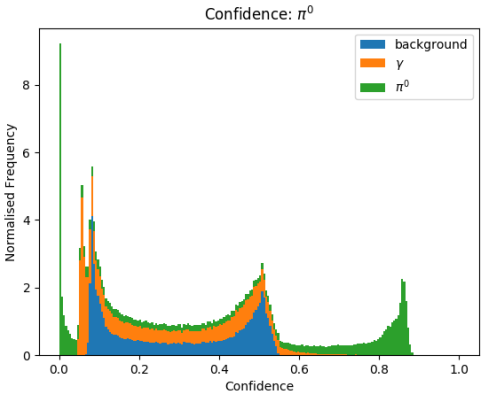

Figure 16: The confidence of the `Keras` output nodes when presented with data from the $D^0\bar{D}^0$ mesons produced by π^0-producing decays. The data has been grouped by the output node, since each node refers to a different classification class.

This theory is supported by the confidence of the π^0-producing decay classifications in figure 16. A large output confidence range demonstrates there is a multitude of events which look very similar, as the network is unable to distinguish between them with any confidence. There is a high confidence peak of π^0-producing decays which comes from the low mass tail of the π^0-producing decays, and is coupled with low confidence classifications of γ-producing decays and background. However, there is a peak in the π^0-producing decay confidence at 0, coupled with a confidence peak in both γ-producing decays and background, hence implying that a set of π^0-producing decay events have masses closer to that of the γ-producing decay $D^0\bar{D}^0$ or back-

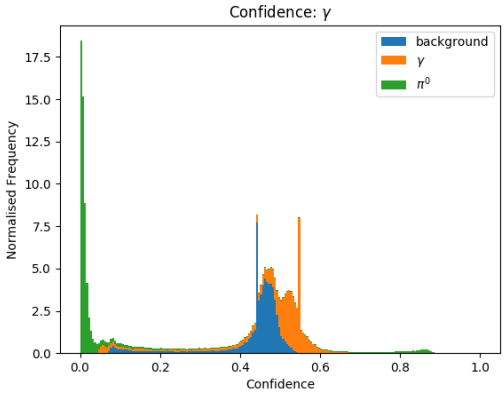

Figure 17: The confidence of the `Keras` output nodes when presented with data from the $D^0\bar{D}^0$ mesons produced by γ-producing decays. The data has been grouped by the output node, since each node refers to a different classification class.

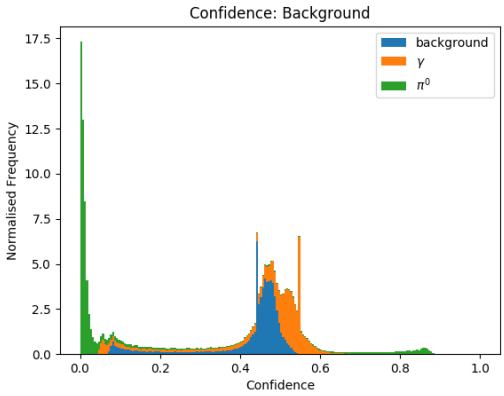

Figure 18: The confidence of the `Keras` output nodes when presented with data from the $D^0\bar{D}^0$ mesons produced in simulated background data. The data has been grouped by the output node, since each node refers to a different classification class.

ground, which in turn look identical to one another. It is thus extremely likely that there is a small set of π^0-producing decay events in which the $D^0\bar{D}^0$ mass is significantly higher than would be expected for a π^0-producing decay. These high mass $D^0\bar{D}^0$ mesons will be referred to as the extreme mass π^0-producing decays.

Very similar results are seen in figures 17 and 18, in which a low confidence π^0-producing decay peak is seen and coupled with two peaks in background and γ-producing decays, which must look very similar since the network is unable to draw confident conclusions. These π^0-producing classifications are considered to be the average mass decay products, since the network is able to classify them to a certain degree. Both confidences also contain a very high confidence peak corresponding to π^0-producing deacys, which is due to the extreme mass π^0-producing decays.

In general, figure 19 suggests the network is able to learn and classify at a rate better than guessing us-

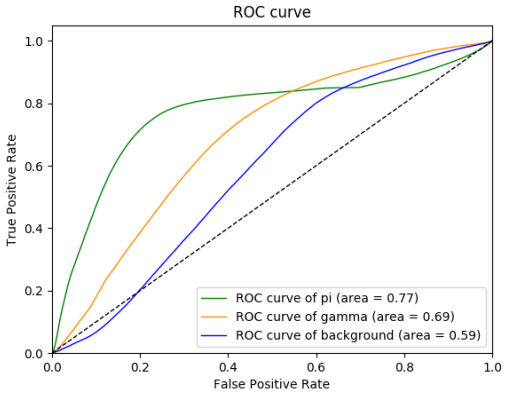

Figure 19: The ROC curve for a neural network having been trained on only the invariant mass of the $D^0\bar{D}^0$ pair.

ing just invariant mass. Due to the relatively small number of π^0-producing decays with an extreme mass, there is a small increase in the gradient at high false positive rate. This is because when the network allows for a high false positive rate, the few extreme mass π^0-producing decays are then classified as π^0-producing decays.

Additionally, inputting the 4-momenta alongside the mass had no effect on the overall accuracy of the network. This is due to the network's ability to discover relationships itself, and since the invariant mass is a manipulation of the 4-momenta, there are no benefits to implementing both. It is hoped that, by adding additional variables unrelated to the invariant mass, the π^0-producing decays will be separated with much higher accuracy since the network will have additional dimension in which to identify the extreme mass π^0-producing decays.

4.3.2 Additional input variables

Through the reconstructed decays, both the D^0 and $\bar{D}^0$ mesons expose thirteen different variables that may be utilised in the network. These are

- Energy E
- Transverse Energy E_t
- Flight Distance d
- Mass m
- Mass Squared m^2
- Transverse Mass m_t
- Momentum p
- x-momentum p_x
- y-momentum p_y
- z-momentum p_z
- Pseudo-rapidity

- Azimuthal angle

- Rapidity

Since charged particles can often be reconstructed at LHCb, additional variables are exposed from additional charged decay products. The impact parameter and minimum impact parameter are collected from the final decay products of the system, namely K^+, K^-, π^+ and π^-. Finally, the angle between the K^- and $\bar{D}^0$ in the D^0 rest frame, the angle between the K^+ and D^0 in the $\bar{D}^0$ rest frame, the angle between the D^0 and $\bar{D}^0$ in the D^0 and $\bar{D}^0$ rest frame and the $D^0\bar{D}^0$ invariant mass were also inputted.

Figure 20 shows the predicted mass distribution as opposed to the the true mass distribution in figure 11.

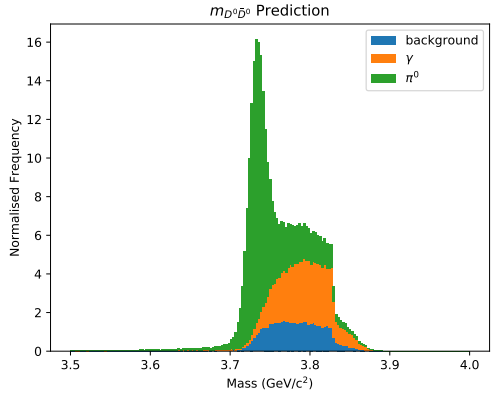

Figure 20: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair as classified by `Keras` as having been produced through π^0-producing, γ-producing and background decays.

The most noticeable difference between the predicted mass distributions in figures 12 and 20 is that the π^0-producing decay appears to be classified to a much greater accuracy at the peak. It also appears as though the network is misclassifying background and γ-producing decays as π^0-producing decays, since π^0-producing decays are predicted over a much larger mass range than they should be. The addition of new variables other than the invariant mass has clearly allowed the network to better separate π^0-producing decays at low masses.

Through creating a mass distribution plot for a specific set of classified events, the reason for the misclassifications in γ-producing decays and background should be derivable. These are seen in figures 21, 22 and 23.

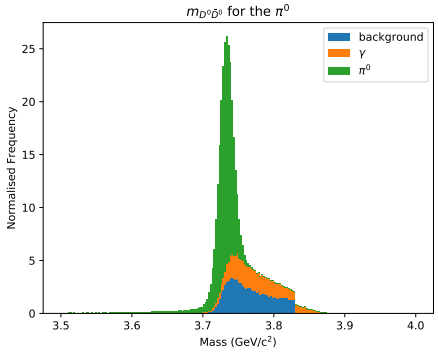

Figure 21: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair classified as having been produced through π^0-producing decays. The data has been grouped by the true classification of the event.

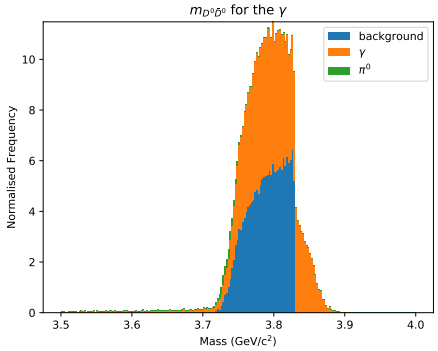

Figure 22: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair classified as having been produced through γ-producing decays. The data has been grouped by the true classification of the event.

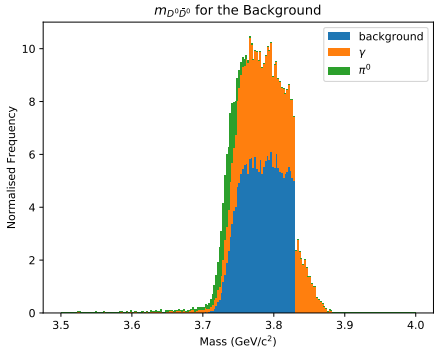

Figure 23: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair classified as having been produced through background simulation. The data has been grouped by the true classification of the event.

Figure 21 clearly shows that the majority of π^0-producing classified decays are from π^0-producing decays as expected. This reiterates that the network is highly dependent on the invariant mass for the separation of π^0-producing decays from the γ-producing decays and background, although the additional variables have enabled the network to refine this classifi-

cation to a much higher degree. The peak of misclassifications is still observed, which again implies that the invariant mass is still extremely important; since most π^0-producing decays are found at the low mass, the network is overpredicting the number of low mass π^0-producing decays.

Figures 22 and 23 show that the further separation of γ-producing decays and background is significantly more difficult, despite the additional variables. It appears as though the network is unable to distinguish between γ-producing decays and background signal, since both classifications consist of an approximately equal distribution of true γ-producing decay classifications and background. The background cut is clearly visible in the classifications, and so must be taken into consideration when evaluating the network accuracy.

Fortunately, the peak of π^0-producing decays in the background classification appears to have been reduced. This is again due to the additional dimensions of separation available to the network with the introduction of the additional variables; in fact this shows that almost every π^0-producing decay can now be classified as such. The extreme mass π^0-producing decays are no longer a source of inaccuracy for the network, which is a large improvement on only using the invariant mass.

Further analysis is possible through the output node confidence for each classification class. These are shown in figures 24, 25 and 26.

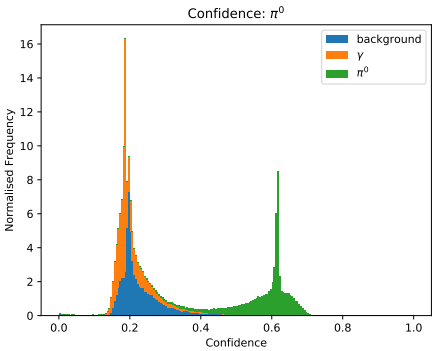

Figure 24: The confidence of the `Keras` output nodes when presented with data from the $D^0\bar{D}^0$ mesons produced by π^0-producing decays. The data has been grouped by the output node, since each node refers to a different classification class.

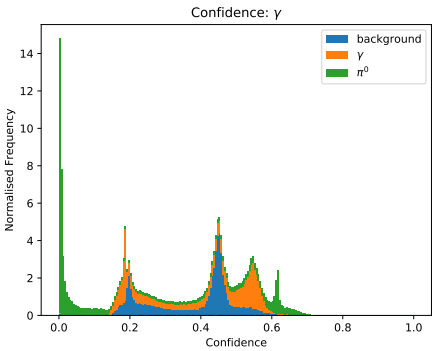

Figure 25: The confidence of the `Keras` output nodes when presented with data from the $D^0\bar{D}^0$ mesons produced by γ-producing decays. The data has been grouped by the output node, since each node refers to a different classification class.

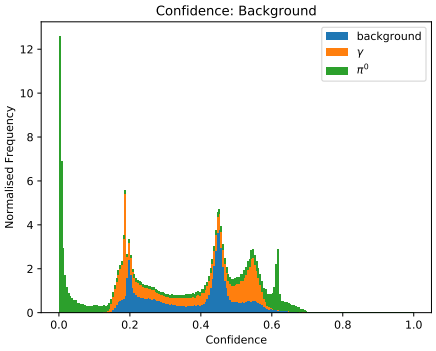

Figure 26: The confidence of the `Keras` output nodes when presented with data from the $D^0\bar{D}^0$ mesons produced in simulated background data. The data has been grouped by the output node, since each node refers to a different classification class.

The π^0-producing decay confidence is as expected, and has significantly improved. The network is able to classify a π-producing decay with a large separation in confidence from an incorrect classification. The two peaks in confidence tail into one another between the peaks, thus leave space for improvement since a sharper cut in classification would be better. This is a huge improvement from figure 16, and shows that the network is almost entirely able to separate π^0-producing decays from the γ-producing decays and the simulated background with confidence. The π^0-producing decays which were classified as π^0-producing decays with 0 confidence have been classified confidently as π^0-producing decays now, hence this confirms that the extreme mass π^0-producing decays were able to be classified based on additional variables.

Unfortunately, the same improvements are not seen in the γ-producing or background decay classifications. Interestingly, the most extreme confidences correspond to π^0-producing decays. Reassuringly most of these fall into the low confidence category, implying that the network distinguishes them from the γ-producing or background decays. However, there is a relatively smaller

peak at the highest confidence level, implying that a significant number of the γ-producing and background decays look like π^0-producing decays. This is likely to do with the overdependence on the invariant mass. Since the network will have made a mass cut by which is relies heavily on to classify π^0-producing decays, any γ-producing or background decays with a mass falling at the extremity of the cut will potentially be classified confidently as a π^0-producing decay event.

The double peak classification of background and γ-producing decays is very interesting. The peak at about 0.2 corresponds to the same peak in the π^0-producing decay confidence in figure 24, and so is likely to have been caused by the γ or background producing decays which the network perceives to be π^0-producing. By reducing the dependence on the mass, it is possible that both the incorrect high confidence π^0-producing decay classification and the peaks at 0.2 can be removed.

The second background and γ-producing classification peak at a confidence of approximately 0.45 is more realistic. Since both classifications are of similar confidence, this supports the idea that γ-producing and background decays are extremely difficult to classify. This is further supported by the fact that the background decay is less confidently classified than the γ-producing decays in both the background and γ-producing decays. This implies that the network is observing both particles to be almost identical, and thus uses the same relationships to classify both.

Since the ratio of background to signal was equal, a ROC curve can quantify the accuracy of the network. Figure 27 shows this, and supports the idea that π^0-producing decay classifications are much easier to separate than the γ-producing and background decays. It also indicates that γ-producing and background decay classifications are not completely random, thus the network must have some ability to separate some γ-producing decays from background decays. The mass cut in the background will significantly affect this, since it allows for a large proportion of the γ-producing decays to be perfectly separated due to non-physical reasons.

Since the background will be discarded, it is less important to classify this accurately relative to accurate classifications of the γ-producing decays. Therefore, it may be possible to require a lower false positive rate of around 0.3 for a γ-producing decay classification to result in a smaller but cleaner γ-producing decay signal if the network is still unable to separate these decays. However, since the ROC curve is very similar for γ-producing decays and background signal, and both have a positive gradient throughout the false positive range as opposed to the π^0-producing decay which levels off, the accuracy improvement would be limited.

Rank	Variable	Importance (10^{-1})
1	$m_{D^0\bar{D}^0}$	6.38
2	m_{D^0}	4.33
3	$m_{\bar{D}^0}$	4.21
4	$\theta_{D^0\bar{D}^0}$	3.31
5	β_{D^0}	2.48
6	$\beta_{\bar{D}^0}$	2.17
7	$E_{T\bar{D}^0}$	1.36
8	$m_{T\bar{D}^0}$	1.28
9	p_{TD^0}	1.28
10	$p_{z\bar{D}^0}$	1.27

Table 2: The top ten variables the network uses to make classifications.

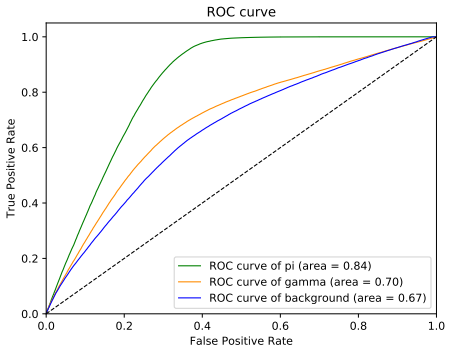

Figure 27: The ROC curve for a neural network having been trained on 39 variables extracted from the $D^0\bar{D}^0$ pair and additional decay products.

The relative importance of each variable as seen by the network was extracted in order to determine how dependent the network truly is on mass. Table 2 was generated, with the most important variables listed at the top.

Indeed, the top three most important variables in classification are the combined invariant mass $m_{D^0\bar{D}^0}$ and the individual masses m_{D_0} and $m_{D_0\bar{D}_0}$. By reducing the background cut, and reducing the effects of FSR, the dependence on the masses should be reduced.

The angle $\theta_{D^0\bar{D}^0}$ is also considered to be an important variable to the network. However, this variable should always be 180°, and so indicates a simulation effect as opposed to a physical effect.

4.4 Background Simulation

Since the sharp cut in the background sample has been shown to affect the final accuracy of the network, it is important to reduce this effect by removing the cut. Alongside this, the effects of FSR have been removed from the signals, since the generation of FSR for the background sample was considered to be significantly more difficult. Figure 6 demonstrates the new background, in which the cut has been removed, and figures 7 and 8 demonstrate the mass distribution of the frequency before and after the FSR removal. The true invariant mass distribution of the input data sets can be seen in figure 28.

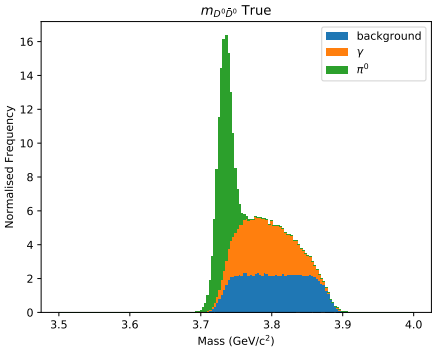

Figure 28: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair produced through π^0-producing, γ-producing and background decays. The removal of the background cut can be seen.

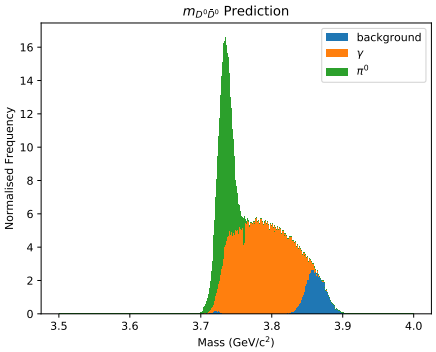

Figure 29: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair as classified by `Keras` as having been produced through π^0-producing, γ-producing and background decays.

Figure 29, the predicted mass distribution of the data set, seems to show again a large misclassification between γ-producing and background decays. The network appears to favour classifying decays as γ-producing. The π^0-producing decay separation appears to have been reasonably successful, but nearly everything else has been classified as γ-producing decay aside from at high mass. It is likely that this is due to the reduction in signal at these high masses; almost everything at the high mass extremity is background and hence the network is very confident an event with high mass can be classifed as background.

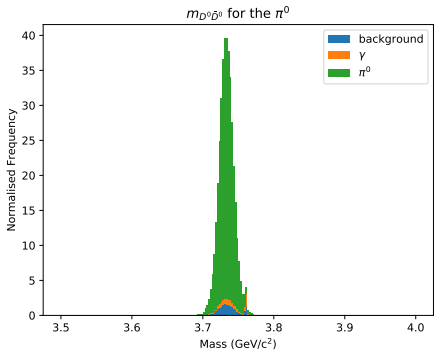

Figure 30: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair classified as having been produced through π^0-producing decays by `Keras`. The data has been grouped by the true classification of the event.

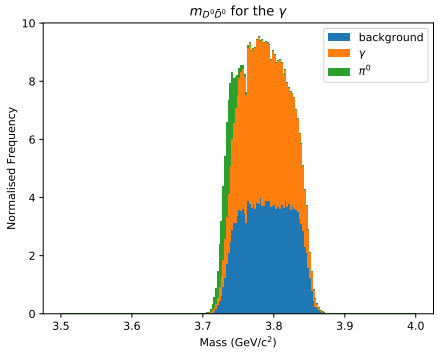

Figure 31: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair classified as having been produced through γ-producing decays. The data has been grouped by the true classification of the event.

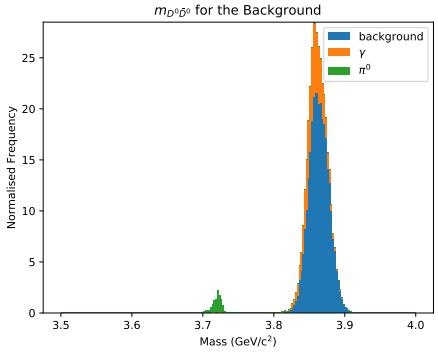

Figure 32: The invariant mass distribution of the $D^0\bar{D}^0$ meson pair classified as having been produced through background simulation. The data has been grouped by the true classification of the event.

Figure 30 supports the successful separation of π^0-producing decays from the γ-producing and background decays. Only a very small peak of misclassifications can be seen, which is again a large improvement when compared to figures 13 and 21. There is no longer a high mass misclassification; instead the range in which π^0-producing decays are classified is

very small. Similar accuracies are observed in the background classification in figure 32. Although a few misclassifications are observed, `Keras` was able to identify background with reasonable accuracy. However, figure 31 shows a different story for the γ-producing decay classifications. `Keras` has incorrectly classified a lot of background decays as γ-producing decays. This is likely since the background and γ-producing decay samples have been shown to be difficult to distinguish.

Since background and γ-producing decays look so similar to the network, despite the additional variables, the network classifies everything as a γ-producing decay unless there is a distinguishing feature separating the background and γ-producing decays or the background is outside the main $D^0\bar{D}^0$ mass range for γ-producing decays. This explains the small peak in background classifications in the predicted invariant mass distribution in figure 29 at high mass, and also the small π^0-producing decay peak in the background classifications in 32. The latter peak is likely to be a small subset of π^0-producing decays which vary in some way to the majority of the π^0-producing decays, thus letting the network believe these could be background.

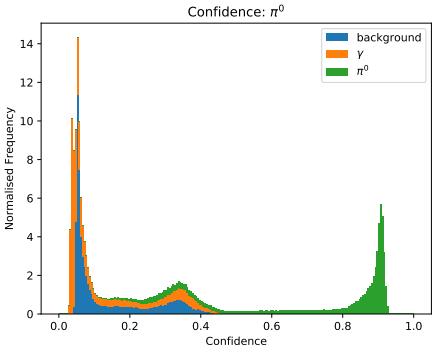

Figure 33: The confidence of the `Keras` output nodes when presented with data from the $D^0\bar{D}^0$ mesons produced by π^0-producing decays. The data has been grouped by the output node, since each node refers to a different classification class.

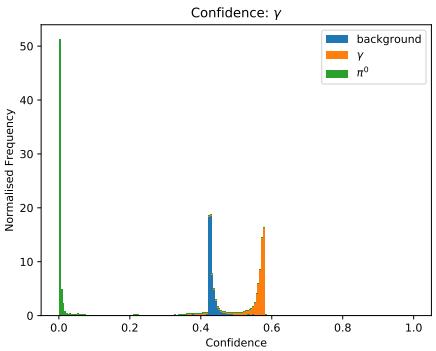

Figure 34: The confidence of the `Keras` output nodes when presented with data from the $D^0\bar{D}^0$ mesons produced by γ-producing decays. The data has been grouped by the output node, since each node refers to a different classification class.

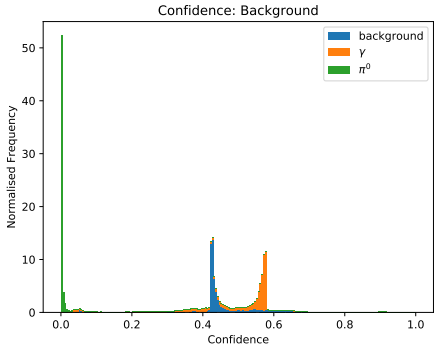

Figure 35: The confidence of the `Keras` output nodes when presented with data from the $D^0\bar{D}^0$ mesons produced in simulated background data. The data has been grouped by the output node, since each node refers to a different classification class.

The confidence plots show that the network has significantly improved. Figure 33 shows that the network is very confidently able to classify π^0-producing decays. The confidence plots of γ-producing and background decays in figures 34 and 35 respectively show the confidence outputs due to π^0-producing decay misclassifications has been entirely reduced. However, the confidence of a classification being a γ-producing or background decay are still very close, although separated very well. The confidence peaks are also in the same order; the network is always more confident a given event is a γ-producing decay as opposed to a background decay. This seems to suggest that it is impossible to distinguish between the background and γ-producing decay samples, since `Keras` responds identically to both types of decay. This also explains why the background classification sample is much cleaner than the γ-produciing decay classification sample, since a particle must have a distinguishing feature in order to be classified as background. In fact, the subset of background signal which `Keras` was able to classify accurately comes from the high confidence tail in figure 35 which extends beyond the γ-producing decay confidence peak.

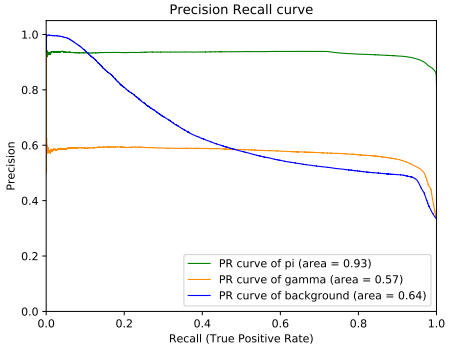

Figure 36: The PR curve for a neural network having been trained on 39 variables extracted from the $D^0\bar{D}^0$ pair and additional decay products. The cut in the background has been removed, and FSR is no longer simulated in the π^0 and γ producing samples.

The PR curve supports all conclusions already drawn from this data. When the network has low recall, so classifies very few background events as background, it is able to identify background decays with a very high accuracy. However at a higher recall this is no longer possible, thus indicating that the accuracy of the background is due to the network only classifying something as background when it is very confident.

The π-producing decays are very precise for all recalls, thus supports the conclusion that π^0-producing decays are very separable.

5 Conclusion

The network has been able to separate π^0-producing decays from the γ-producing decays and the background signal with a very high accuracy. This is due to the inclusion of the invariant mass, but also a few transverse variables which depend on the mass but are affected by boosts into different frames of reference. However, the γ-producing decays and the background are seen by the network to be almost identical, and hence a relatively low degree of separation can be achieved between these classes. This is thought to be due to the network dependence on invariant mass.

6 Further Work

Further work to improve the classification ability of the neural network when attempting to tag configurations of $D^0\bar{D}^0$ mesons generated in $X_{c1}(3872)$ decays should predominantly focus on reducing the network reliance on the invariant mass, since this is the predominant method of classification. Furthermore, by reducing the number of input variables by removing those which the network does not use, a smaller network may be achieved to allow training to occur significantly faster.

References

[1] Einstein, A. Explanation of the Perihelion Motion of Mercury from the General Theory of Relativity. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1915**, 831 (1915).

[2] Morrison, J. S. Ancilla to the Pre–Socratic Philosophers: A Complete Translation of the Fragments in Diels, Fragmente der Vorsokratiker. *The Journal of Hellenic Studies* **69**, 92 (2006).

[3] Marciano, W. J. The Standard Model and Beyond. *Annals of the New York Academy of Sciences* **461**, 367–383 (1986).

[4] LHCb collaboration *et al.* Measurement of CP observables in $B^{\pm} \to D^{(*)}K^{\pm}$ and $B^{\pm} \to D^{(*)}\pi^{\pm}$ decays (2017).

[5] LHCb collaboration *et al.* Observation of CP violation in charm decays (2019).

[6] Tureanu, A. CPT and Lorentz Invariance: Their relation and violation. In *Journal of Physics: Conference Series*, vol. 474, 12031 (2013).

[7] Oblakowska-Mucha, A. Neutral B-meson mixing and CP violation at lhcb. In *Journal of Physics: Conference Series*, vol. 770 (2016).

[8] Collaboration, T. B. Observation of time-reversal violation in the B^0 meson system. *Physical Review Letters* **109** (2012).

[9] Goldhaber, M. & Rosner, J. L. Mixing of neutral charmed mesons and tests for CP violation in their decays. *Physical Review D* **15**, 1254–1259 (1977).

[10] Dungel, W. *et al.* Measurement of the form factors of the decay $B^0 \to D^{*-}\ell^+\nu_\ell$. *Physical Review D* **82**, 112007 (2010).

[11] Braaten, E. & Kang, D. The $J/\psi\omega$ decay channel of the $X(3872)$ charm meson molecule (2013).

[12] Sakharov, A. D. Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe. *JETP Lett.* **5** (1967).

[13] Steigman, G. & Scherrer, R. J. Is the universal matter-antimatter asymmetry fine tuned? Tech. Rep. (2018).

[14] Moyotl, A. New sources of CP violation. In *Journal of Physics: Conference Series*, vol. 912, 12019 (2017).

[15] Charles, J. *et al.* CP violation and the ckm matrix: Assessing the impact of the asymmetric B factories. *European Physical Journal C* **41**, 1–131 (2005).

[16] Wolfenstein, L. Parametrization of the Kobayashi-Maskawa Matrix. *Phys.Rev.Lett.* **51**, 1945 (1983).

[17] Chau, L. L. & Keung, W. Y. Comments on the parametrization of the kobayashi-maskawa matrix. *Physical Review Letters* **53**, 1802–1805 (1984).

[18] Kenzie, M. W. & Whitehead, M. P. Update of the lhcb combination of the ckm angle γ (2018).

[19] Vos, K. K., Sevoir, M. & Perazzini, S. Progress and challenges in the study of direct CP violation and γ determinations: summary of ckm 2018 working group v (2019).

[20] Rama, M. Measurement of the ckm angle γ (ϕ_3) using $B \to DK$ decays. 003 (2009).

[21] Rama, M. Measurement of strong phases, D – $\bar{D}$ mixing, and cp violation using quantum correlation at charm threshold (2016).

[22] Brod, J. & Zupan, J. The ultimate theoretical error on γ from $B \to DK$ decays. *Journal of High Energy Physics* **2014** (2014).

[23] Bernabéu, J., Martínez-Vidal, F. & Villanueva-Pérez, P. Time reversal violation from the entangled B^0B^0 system. *Journal of High Energy Physics* **2012** (2012).

[24] Aaij, R. *et al.* Determination of the $X(3872)$ meson quantum numbers. *Physical Review Letters* **110**, 222001 (2013).

[25] Hanhart, C., Kalashnikova, Y. S., Kudryavtsev, A. E. & Nefediev, A. V. Reconciling the $X(3872)$ with the near-threshold enhancement in the $D^0\bar{D}^{*0}$ final state. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **76** (2007).

[26] Priddy, K. L. & Keller, P. E. *Artificial Neural Networks: An Introduction* (SPIE, 2009).

[27] Cowan, G. A., Craik, D. C. & Needham, M. D. Rapidsim: an application for the fast simulation of heavy-quark hadron decays. Tech. Rep. (2017). URL `https://arxiv.org/pdf/1612.07489.pdf`.

[28] Brun, R. & Rademakers, F. Root - an object oriented data analysis framework. URL `https://github.com/root-project/root`.

[29] Keras. Keras. URL `https://github.com/keras-team/keras`.

[30] Google. Tensorflow. URL `https://github.com/tensorflow/tensorflow`.

[31] Davis, J. & Goadrich, M. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning - ICML '06*, 233–240 (ACM Press, New York, New York, USA, 2006).

Project Report/Dissertation presented as part of, and in accordance with, the requirements for the Final Degree of MSci at the University of Bristol, Faculty of Science.

Author:	Toby Twigger
Title:	Isolating quantum correlated charm systems for new constraints on symmetry violation
Date of Submission:	2019

Signed: Toby Twigger
Full Name: Toby Twigger
Date: August 7, 2025