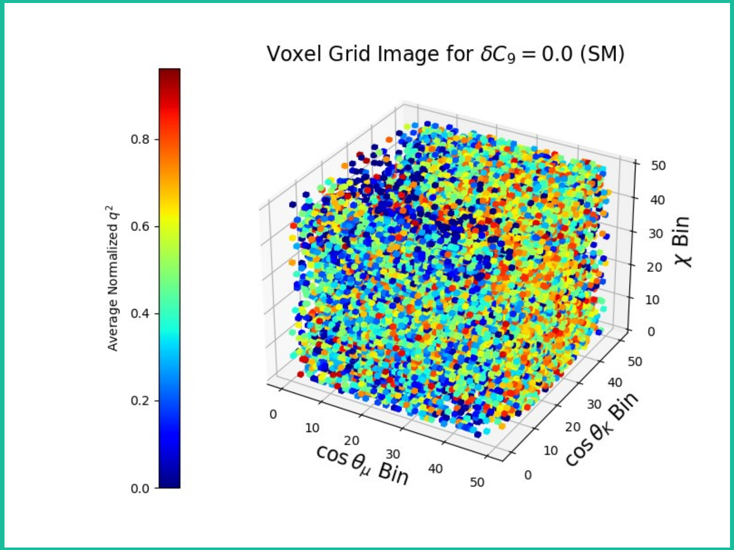


BROWN

Machine Learning for New Physics in $B \rightarrow K^{*0}\mu^+\mu^-$ Decays

Shawn Dubey (shawn_dubey@brown.edu)
Brown University Center for the Fundamental Physics of the Universe

UH Manoa Physics Journal Club Presentation

01/09/2024

Introduction and Motivation

Introduction

BROWN

- In recent years, several experiments seem to have seen hints of lepton flavor universality violation in certain B meson decay modes
  - BaBar (USA)
  - Belle (Japan)
    - The Belle II experiment is now online
  - LHCb (Switzerland)

Introduction

$$R_{K^*} = \frac{\mathcal{B}(B \to K^* \mu^+ \mu^-)}{\mathcal{B}(B \to K^* e^+ e^-)}$$

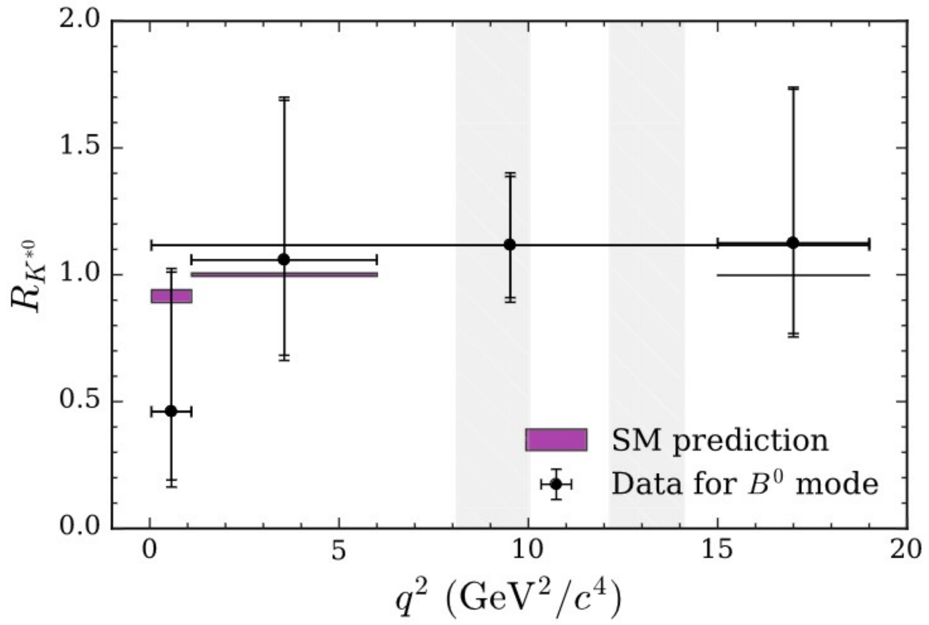
Introduction

$$R_{K^*} = \frac{\mathcal{B}(B \to K^* \mu^+ \mu^-)}{\mathcal{B}(B \to K^* e^+ e^-)}$$

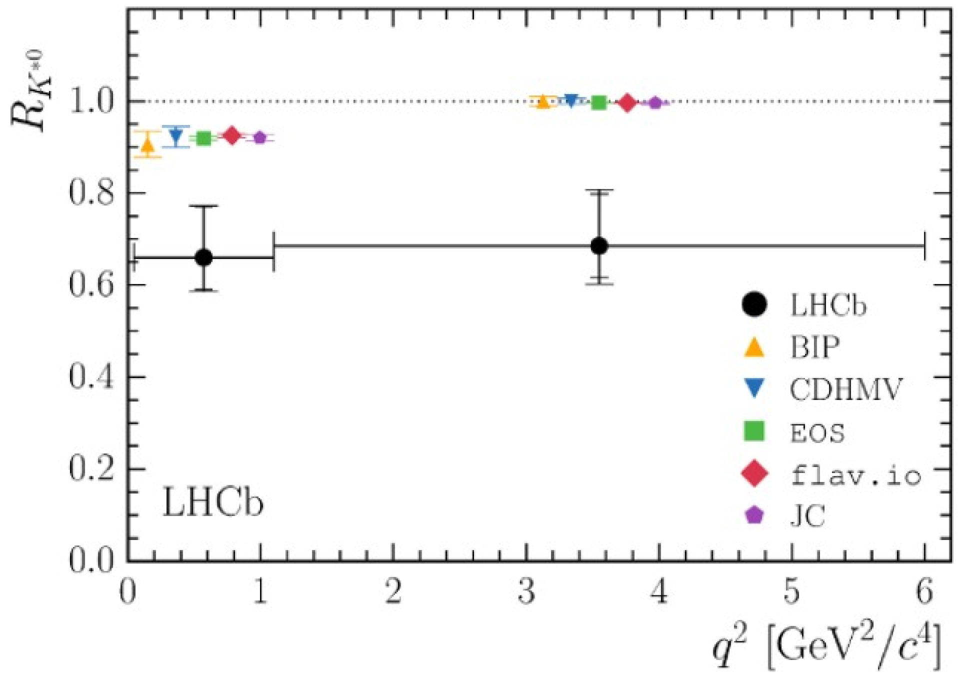
Hints at:
Belle: Phys. Rev. Lett. 126, 161801 (2021)
LHCb: JHEP 08 (2017) 055
BaBar: Phys. Rev. D 86, 032012 (2012)

Introduction

Phys. Rev. Lett. 126, 161801 (2021)

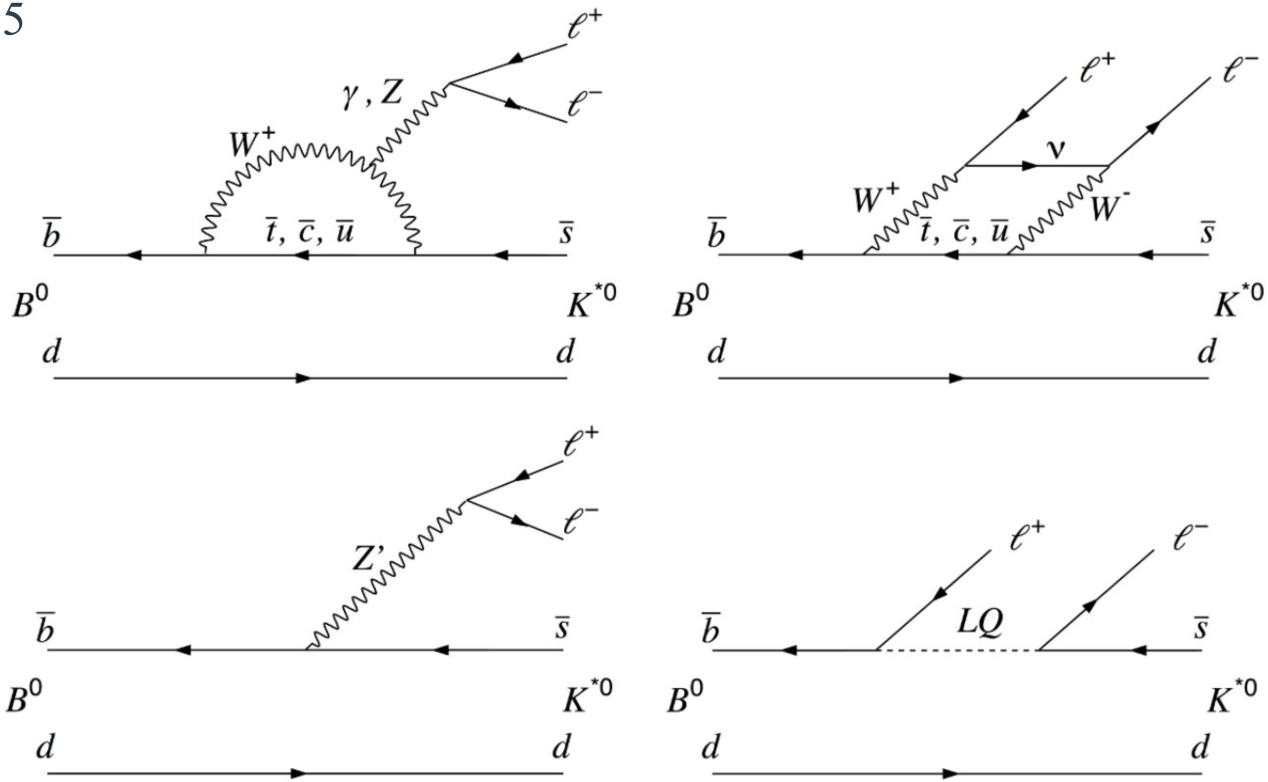


JHEP 08 (2017) 055



Introduction

JHEP 08 (2017) 055



Introduction

$$R_{K^*} = \frac{\mathcal{B}(B \to K^* \mu^+ \mu^-)}{\mathcal{B}(B \to K^* e^+ e^-)}$$

LHCb: arXiv:2212.09152 (2022)
Belle: Phys. Rev. Lett. 126, 161801 (2021)
LHCb: JHEP 08 (2017) 055
BaBar: Phys. Rev. D 86, 032012 (2012)

Introduction

$$R_{K^*} = \frac{\mathcal{B}(B \to K^* \mu^+ \mu^-)}{\mathcal{B}(B \to K^* e^+ e^-)}$$

LHCb: arXiv:2212.09152 (2022)
Belle: Phys. Rev. Lett. 126, 161801 (2021)
LHCb: JHEP 08 (2017) 055
BaBar: Phys. Rev. D 86, 032012 (2012)

Introduction

BROWN

$$R_{K^*} = \frac{\mathcal{B}(B \to K^* \mu^+ \mu^-)}{\mathcal{B}(B \to K^* e^+ e^-)}$$

LHCb: arXiv:2212.09152 (2022) Results currently consistent with SM
Belle: Phys. Rev. Lett. 126, 161801 (2021)
LHCb: JHEP 08 (2017) 055
BaBar: Phys. Rev. D 86, 032012 (2012)

Introduction

BROWN

$$\text{low-}q^2 \begin{cases} R_K &= 0.994 \,^{+0.090}_{-0.082}\,(\text{stat})\,^{+0.029}_{-0.027}\,(\text{syst}), \\ R_{K^*} &= 0.927 \,^{+0.093}_{-0.087}\,(\text{stat})\,^{+0.036}_{-0.035}\,(\text{syst}), \end{cases}$$

$$\text{central-}q^2 \begin{cases} R_K &= 0.949 \,^{+0.042}_{-0.041}\,(\text{stat})\,^{+0.022}_{-0.022}\,(\text{syst}), \\ R_{K^*} &= 1.027 \,^{+0.072}_{-0.068}\,(\text{stat})\,^{+0.027}_{-0.026}\,(\text{syst}). \end{cases}$$

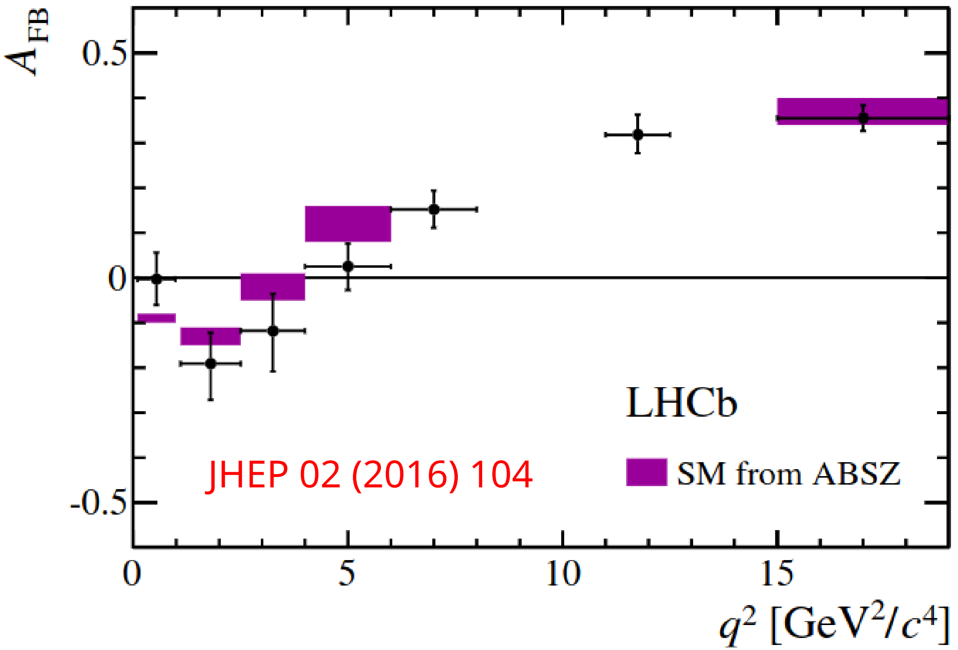
arXiv:2212.09152 (2022)

Introduction

- Measuring these ratios seems less profitable for NP signals (at least for this mode).
- However, all is not lost.
- We can try to look for NP signals in the angular observables obtained from the decay.

Introduction

- **For example, at LHCb**



$$q^2 \equiv M^2(\mu^+\mu^-)$$

$$B \to K^{*0}\mu^+\mu^-$$

$$A_{FB}(q^2) = \frac{\left[\left(\int_0^1 - \int_{-1}^0\right) \mathrm{d}\cos\theta_\ell\right] \mathrm{d}(\Gamma - \bar{\Gamma})}{\int_{-1}^1 \mathrm{d}\cos\theta_\ell \, \mathrm{d}(\Gamma + \bar{\Gamma})}$$

Unaccounted for hadronic effects or NP?

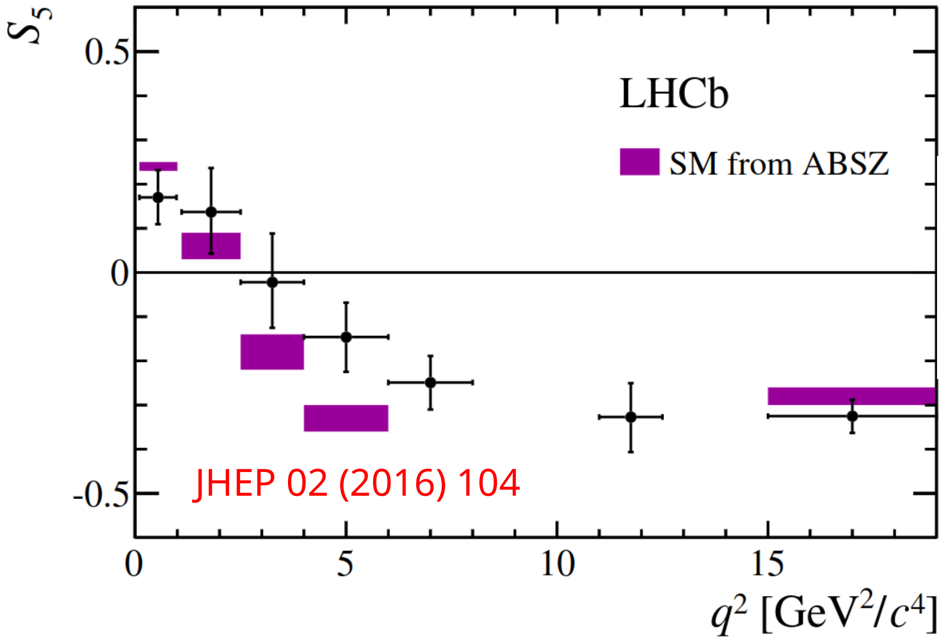
Introduction

BROWN

- **For example, at LHCb**

$$q^2 \equiv M^2(\mu^+\mu^-)$$

$$B \to K^{*0}\mu^+\mu^-$$



JHEP 02 (2016) 104

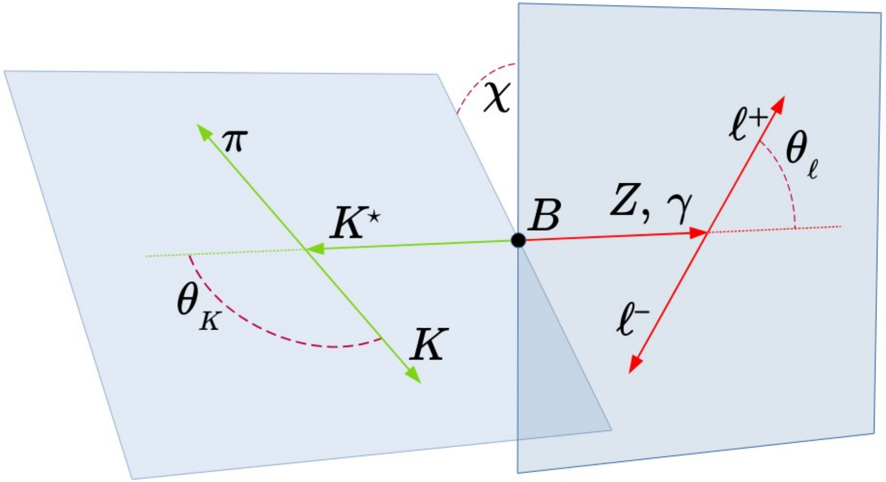
$$S_5(q^2) = \frac{4}{3}\frac{\left[\int_{-\pi/2}^{\pi/2} - \int_{\pi/2}^{3\pi/2}\right]\mathrm{d}\chi\left[\int_0^1 - \int_{-1}^0\right]\mathrm{d}\cos\theta_K \int_{-1}^1 \mathrm{d}\cos\theta_\ell \;\; \mathrm{d}(\Gamma - \bar{\Gamma})}{\int_0^{2\pi}\mathrm{d}\chi\int_{-1}^1 \mathrm{d}\cos\theta_K\int_{-1}^1 \mathrm{d}\cos\theta_\ell \;\mathrm{d}(\Gamma + \bar{\Gamma})}$$

Unaccounted for hadronic effects or NP?

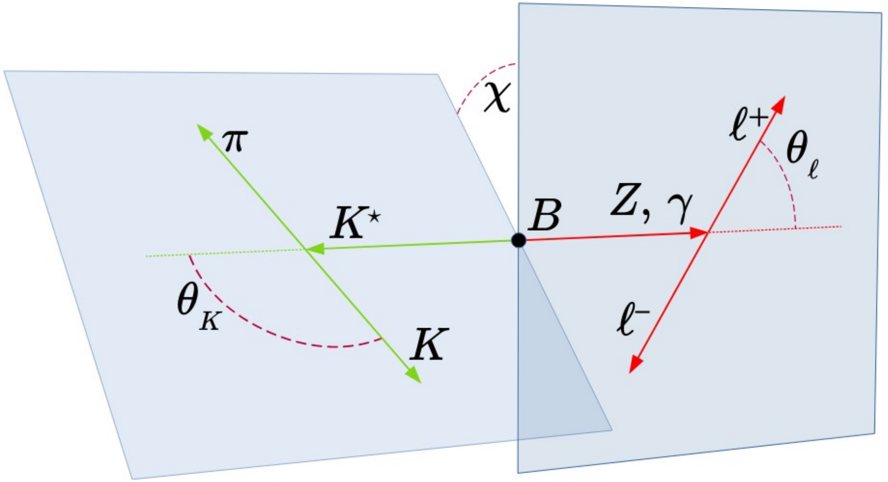
Introduction

- To conduct studies to find out, it would be ideal if we had some NP models/Monte Carlo (MC) generators.
- But to generate what?

Introduction



Introduction



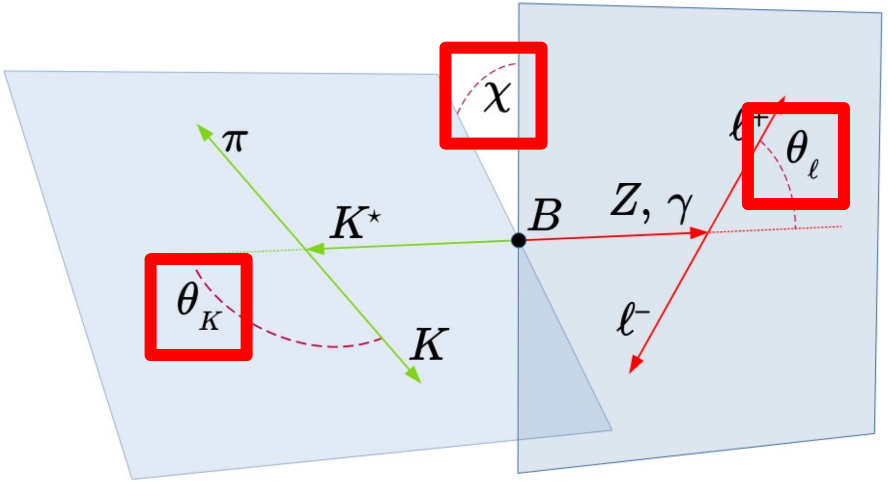
Effective Field Theory

$$\mathcal{M} = \frac{G_F \alpha}{\sqrt{2}\pi} V_{tb} V_{ts}^* \Big\{ \Big[\langle K^* | \bar{s} \gamma^\mu (C_9^{\text{eff}} P_L + C_9' P_R) b | \bar{B} \rangle - \frac{2 m_b}{q^2} \langle K^* | \bar{s} i \sigma^{\mu\nu} q_\nu (C_7^{\text{eff}} P_R + C_7' P_L) b | \bar{B} \rangle \Big] (\bar{\ell} \gamma_\mu \ell) + \langle K^* | \bar{s} \gamma^\mu (C_{10} P_L + C_{10}' P_R) b | \bar{B} \rangle (\bar{\ell} \gamma_\mu \gamma_5 \ell) \Big\},$$

$$\mathcal{O}_S^{(\prime)} = (\bar{s} P_{R(L)} b)(\bar{\mu}\mu) \quad \text{and} \quad \mathcal{O}_P^{(\prime)} = (\bar{s} P_{R(L)} b)(\bar{\mu}\gamma_5\mu)$$

Primes are RH terms.

Introduction



Effective Field Theory

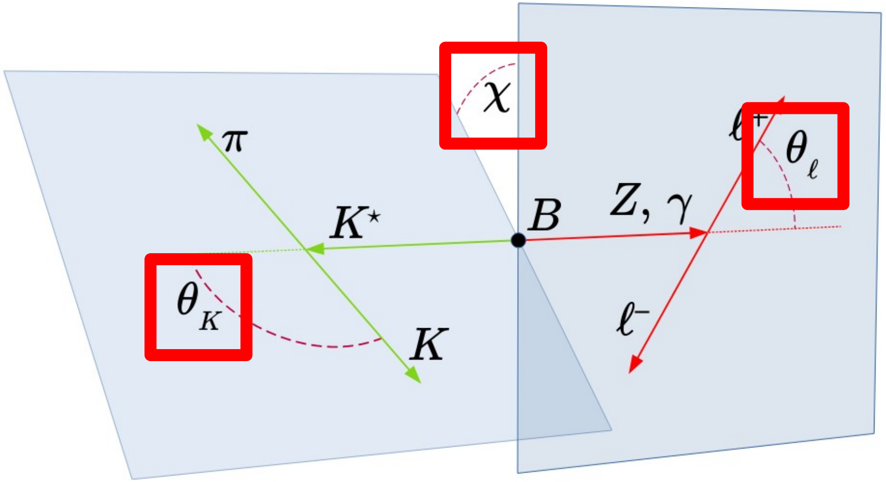
$$\begin{aligned}\mathcal{M} &= \frac{G_F\alpha}{\sqrt{2}\pi} V_{tb}V_{ts}^* \Big\{ \Big[\langle K^*|\bar{s}\gamma^\mu (C_9^{\text{eff}} P_L + C_9' P_R) b|\bar{B}\rangle \\ &- \frac{2m_b}{q^2} \langle K^*|\bar{s} i\sigma^{\mu\nu} q_\nu (C_7^{\text{eff}} P_R + C_7' P_L) b|\bar{B}\rangle \Big] (\bar{\ell}\gamma_\mu \ell) \\ &+ \langle K^*|\bar{s}\gamma^\mu (C_{10} P_L + C_{10}' P_R) b|\bar{B}\rangle (\bar{\ell}\gamma_\mu\gamma_5 \ell) \Big\},\end{aligned}$$

$$\mathcal{O}_S^{(\prime)} = (\bar{s}P_{R(L)}b)(\bar{\mu}\mu) \quad \text{and} \quad \mathcal{O}_P^{(\prime)} = (\bar{s}P_{R(L)}b)(\bar{\mu}\gamma_5\mu)$$

Primes are RH terms.

Introduction

BROWN



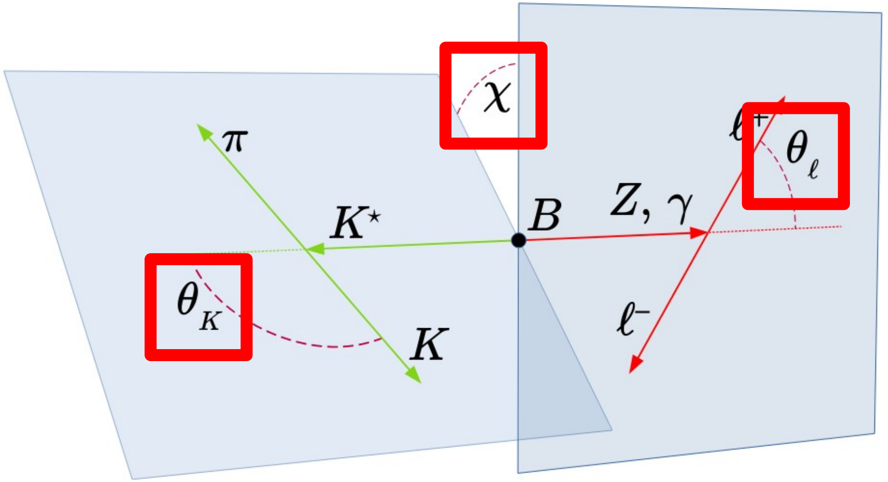
Effective Field Theory

$$\mathcal{M} = \frac{G_F \alpha}{\sqrt{2}\pi} V_{tb} V_{ts}^* \Big\{ \Big[\langle K^* | \bar{s}\gamma^\mu (C_9^{\text{eff}} P_L + C_9' P_R) b | \bar{B} \rangle$$
$$- \frac{2 m_b}{q^2} \langle K^* | \bar{s} i \sigma^{\mu\nu} q_\nu (C_7^{\text{eff}} P_R + C_7' P_L) b | \bar{B} \rangle \Big] (\bar{\ell}\gamma_\mu \ell)$$
$$+ \langle K^* | \bar{s}\gamma^\mu (C_{10} P_L + C_{10}' P_R) b | \bar{B} \rangle (\bar{\ell}\gamma_\mu \gamma_5 \ell) \Big\},$$

$$\mathcal{O}_S^{(\prime)} = (\bar{s} P_{R(L)} b)(\bar{\mu}\mu) \quad \text{and} \quad \mathcal{O}_P^{(\prime)} = (\bar{s} P_{R(L)} b)(\bar{\mu}\gamma_5 \mu)$$

Primes are RH terms.

Introduction



Effective Field Theory

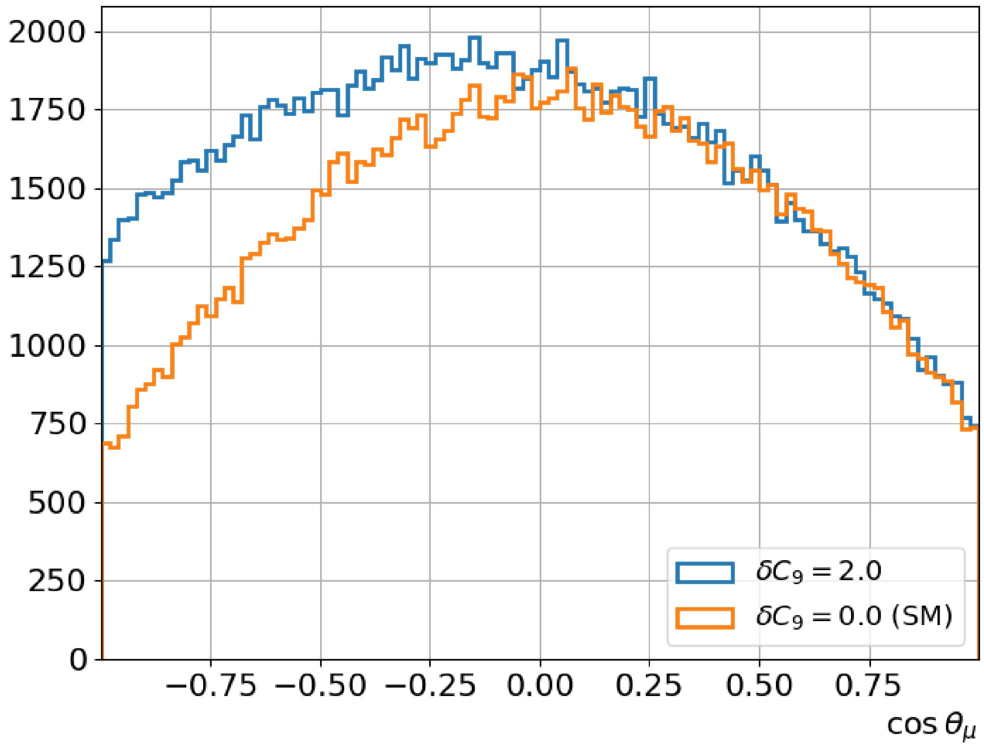
$$\mathcal{M} = \frac{G_F \alpha}{\sqrt{2}\pi} V_{tb} V_{ts}^* \Big\{ \Big[\langle K^* | \bar{s}\gamma^\mu (C_9^{\text{eff}} P_L + C_9' P_R) b | \bar{B} \rangle - \frac{2m_b}{q^2} \langle K^* | \bar{s} i \sigma^{\mu\nu} q_\nu (C_7^{\text{eff}} P_R + C_7' P_L) b | \bar{B} \rangle \Big] (\bar{\ell}\gamma_\mu \ell) + \langle K^* | \bar{s}\gamma^\mu (C_{10} P_L + C_{10}' P_R) b | \bar{B} \rangle (\bar{\ell}\gamma_\mu \gamma_5 \ell) \Big\},$$

$$\mathcal{O}_S^{(\prime)} = (\bar{s} P_{R(L)} b)(\bar{\mu}\mu) \quad \text{and} \quad \mathcal{O}_P^{(\prime)} = (\bar{s} P_{R(L)} b)(\bar{\mu}\gamma_5 \mu)$$

Examine only one B-flavor so the asymmetries are not washed out

Primes are RH terms.

Introduction



Introduction

- Develop an MC model for **EvtGen** (B-physics event generator)
- Now available and under review in Phys. Rev. D [arXiv:2203.06827v4]
  - Wilson Coefficients (WCs) encode short distance/high energy information
  - MC generator is tunable in terms of NP parameters, δC_i, the deviations of the WCs from their SM values $\delta C_i = C_i^{\mathrm{NP}} - C_i^{\mathrm{SM}}$

Machine Learning Motivation

- Standard HEP procedure to extract WC information is fitting
  - Fit after a projection
  - High-dim fit
- Standard HEP usage of ML is classification
  - Classify events: signal vs background
  - Model training and classification is done on event-by-event basis (particle decay event)

Machine Learning Motivation

- Instead, we use machine learning for regression, *not* classification
  - Extract a continuous value (δC_i) rather than a class (SM vs NP or signal vs background)
- Use MC to create images to train a regression model.

Machine Learning Motivation

- Advantages of NN regression over 4D maximum likelihood fitting
  - Straightforward to include background events into images, nothing special needs to be done
  - With proper model training and optimization, should have competitive sensitivity

The Neural Network

The Neural Network

BROWN

- Our goal is to map the angular observables and q^2 into images that can be understood by a neural network.
- For images, the natural choice of model is the convolutional neural network (CNN)
  - In this case, after studying other models, we find the Residual Network (ResNet) variation of the CNN to be most useful
  - We will use a dense layer to perform *regression* and extract δC_9 values.
  - ResNets first developed in 2015 by K. He et al [arXiv:1512.03385]

Intro to Convolutional Neural Networks

- A Convolutional neural network (CNN) is a neural network model that is built from convolutional layers
  - Convolutional layers (CLs) are designed such that the neurons in the first layer are not connected to all the pixels/voxels in the input image, only a subset (the receptive field).
  - The neurons in the second CL are in turn only connected to a subset of neurons in the first CL.
  - And so on.

Intro to Convolutional Neural Networks

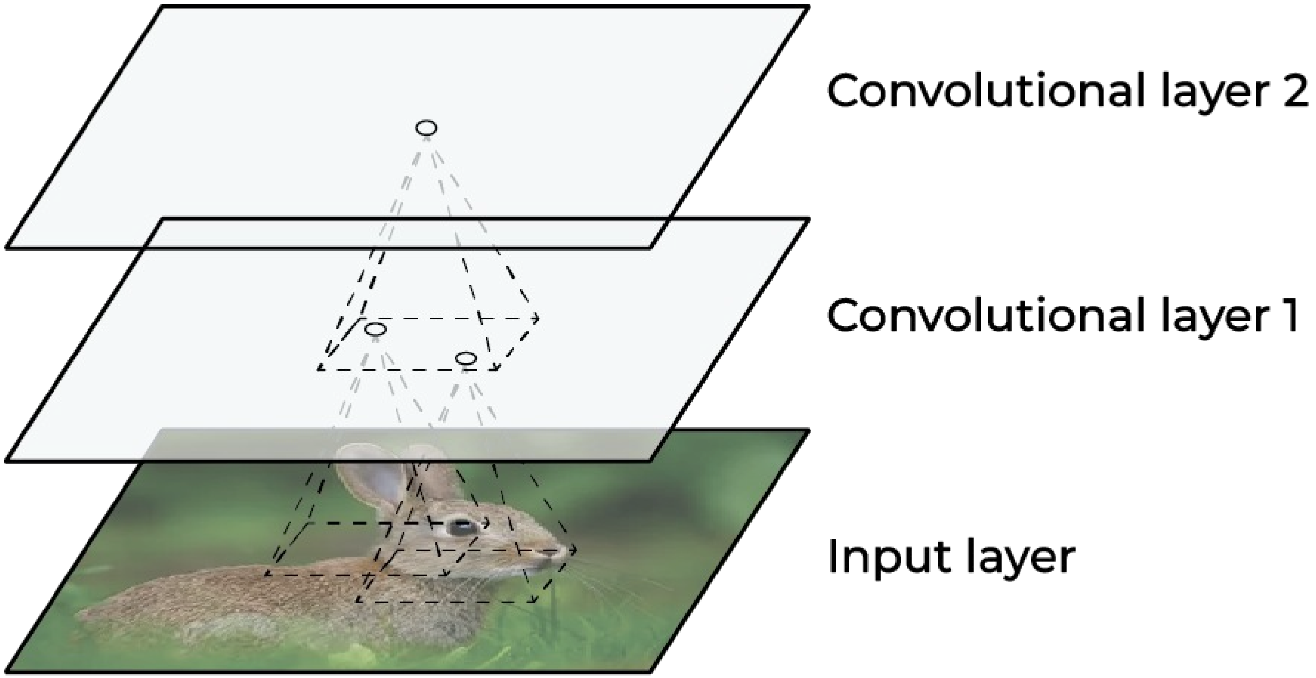


Image based on book by Aurélien Géron (ISBN-10: 1098125975), modified by M. Wan

Intro to Convolutional Neural Networks

BROWN

- Neuron weights are given by filters, which look for features in the image
- A CL can have multiple filters
- When CLs are stacked, earlier layers can processes smaller features and propagate them up the ladder so the layers together reconstruct high-level features

Intro to Convolutional Neural Networks

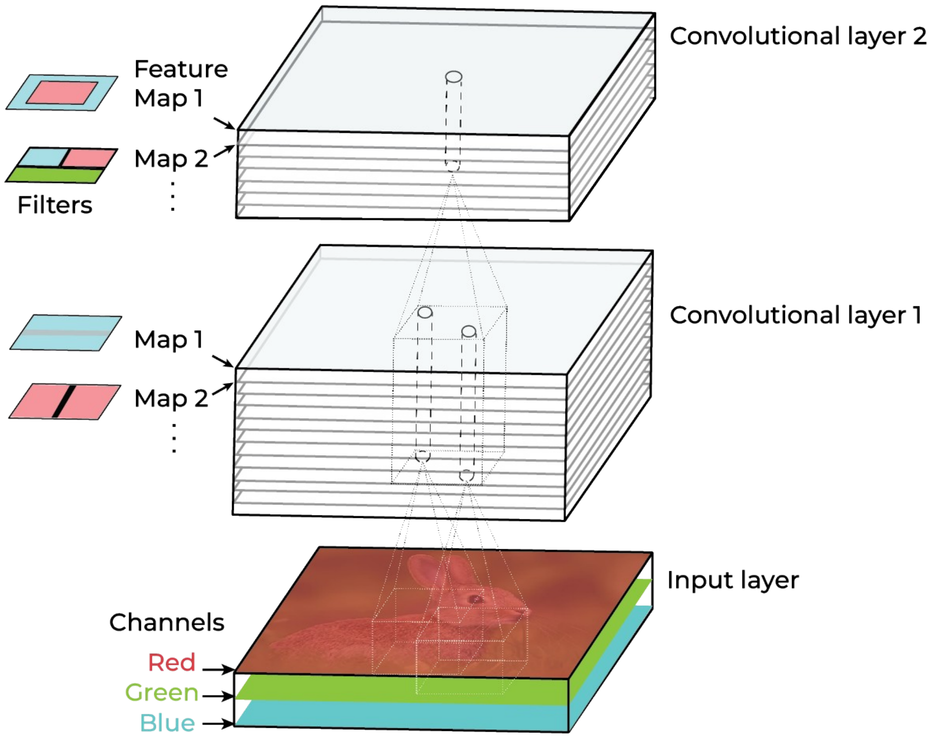


Image based on book by Aurélien Géron (ISBN-10: 1098125975), modified by M. Wan

Intro to Convolutional Neural Networks

- A CNN can have other layers, such as a fully-connected layer at the output to perform classification or regression tasks
  - We will use a fully-connected layer to perform regression and extract δC_9 values.

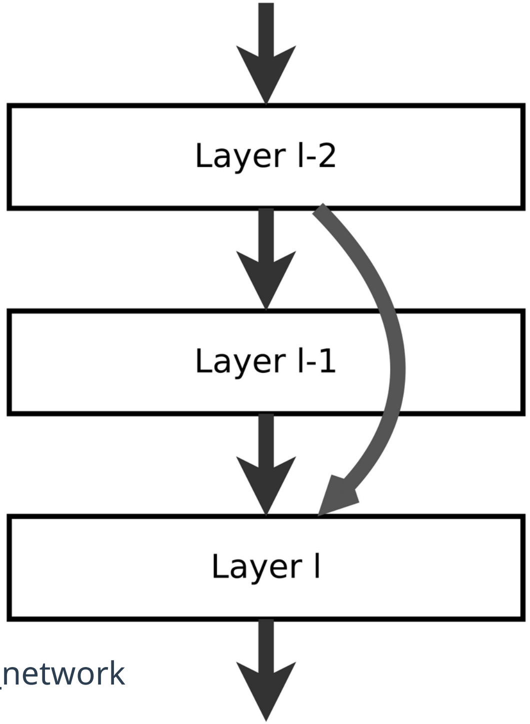
Residual Networks

Intro to Residual Networks

- Residual (neural) networks, or "ResNets", are a type of CNN
- Models the residual of the underlying function
- First developed in 2015 by K. He et al [arXiv:1512.03385]
- Developed to address training issues with very deep neural networks, e.g. vanishing gradients

Intro to Residual Networks

- **Models a residual function by using a “shortcut” or “skip” connection.**



https://en.wikipedia.org/wiki/Residual_neural_network

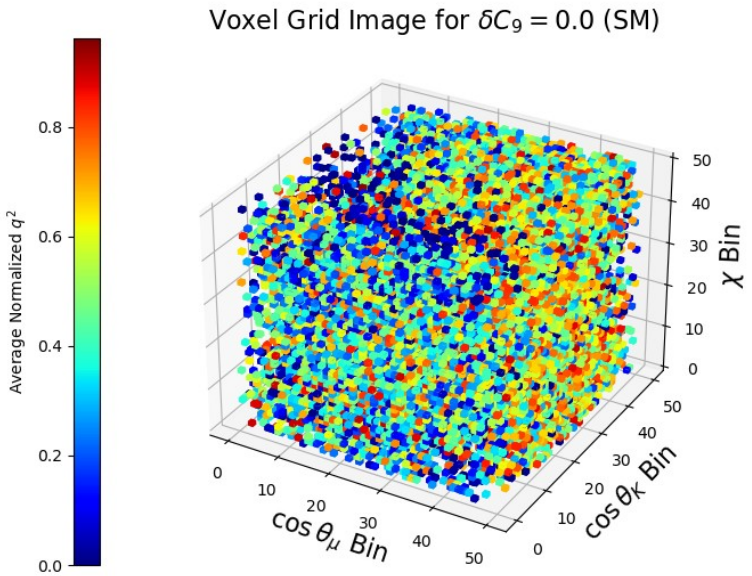
Creating the Images

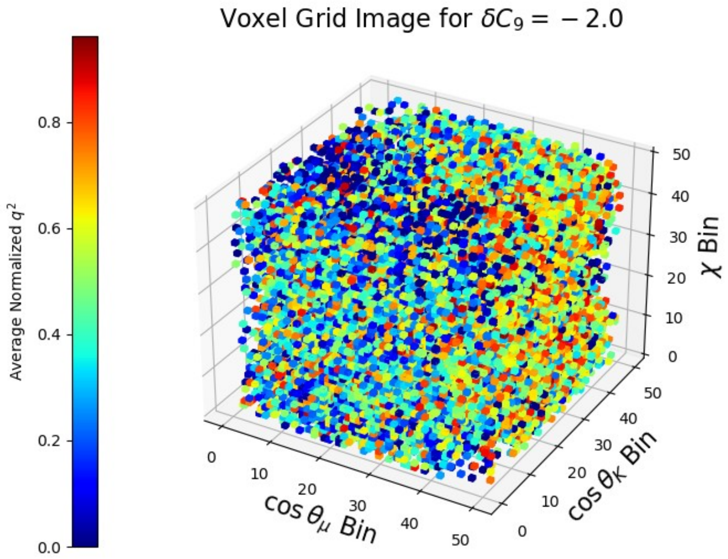
Creating the Images

BROWN

- To create images for training,
  1) Generate $1\text{x}10^6$ MC events for each of 22 different δC_9 values chosen in [-2, 0]
  2) Populate each image with ~250/ab equivalent events
     1) 5x Belle II expected integrated luminosity
  3) Each event is one voxel in the image
- More concretely...
  - We bin the average q^2 in bins of the decay angles to create the voxel grid image
- Proof-of-concept
  - No detector simulation or backgrounds, yet

Images


Voxel Grid Image for $\delta C_9 = 0.0$ (SM)
Average Normalized q^2
0.0
0.2
0.4
0.6
0.8
$\cos\theta_\mu$ Bin
$\cos\theta_K$ Bin
χ Bin
0
10
20
30
40
50


Voxel Grid Image for $\delta C_9 = -2.0$
Average Normalized q^2
0.0
0.2
0.4
0.6
0.8
$\cos\theta_\mu$ Bin
$\cos\theta_K$ Bin
χ Bin
0
10
20
30
40
50

ResNet Design

ResNet Design

We implement a ResNet for regression to <u>extract δC_9 values</u> directly from decay information.

ResNet Design

Use the Keras ML API (https://keras.io/) and TensorFlow (https://www.tensorflow.org/)

- Implement a 34-layer ResNet based on arXiv:1512.03385
- Loss: MAE $\frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$
- Use *stochastic gradient descent* for optimization
- ReLU activation
- Two dense layers after the last convolutional block, with one 50% dropout layer in between
  - First dense layer has 1000 neurons with ReLU activation
  - Second dense layer has one neuron with a linear activation function that performs regression
    - This gives the δC_9 predictions

ResNet Training

ResNet Training

BROWN

- The set of training images is split, with 20% reserved for validation
- If no improvement seen in val loss, reduce LR
- Implement early stopping to mitigate overfitting
- Train on the GPU nodes of the University of Hawaii's MANA cluster
  - Requires us to implement the tensorflow GPU libraries

ResNet Training

- UH’s HPC cluster
- MANA~”divine power”*

https://datascience.hawaii.edu/about-mana/

*wehewehe.org

ResNet Training

- 357 nodes
- 63.19 TB RAM
- 120 GPUs
- > 1 PB disk space

https://datascience.hawaii.edu/about-mana/

ResNet Training

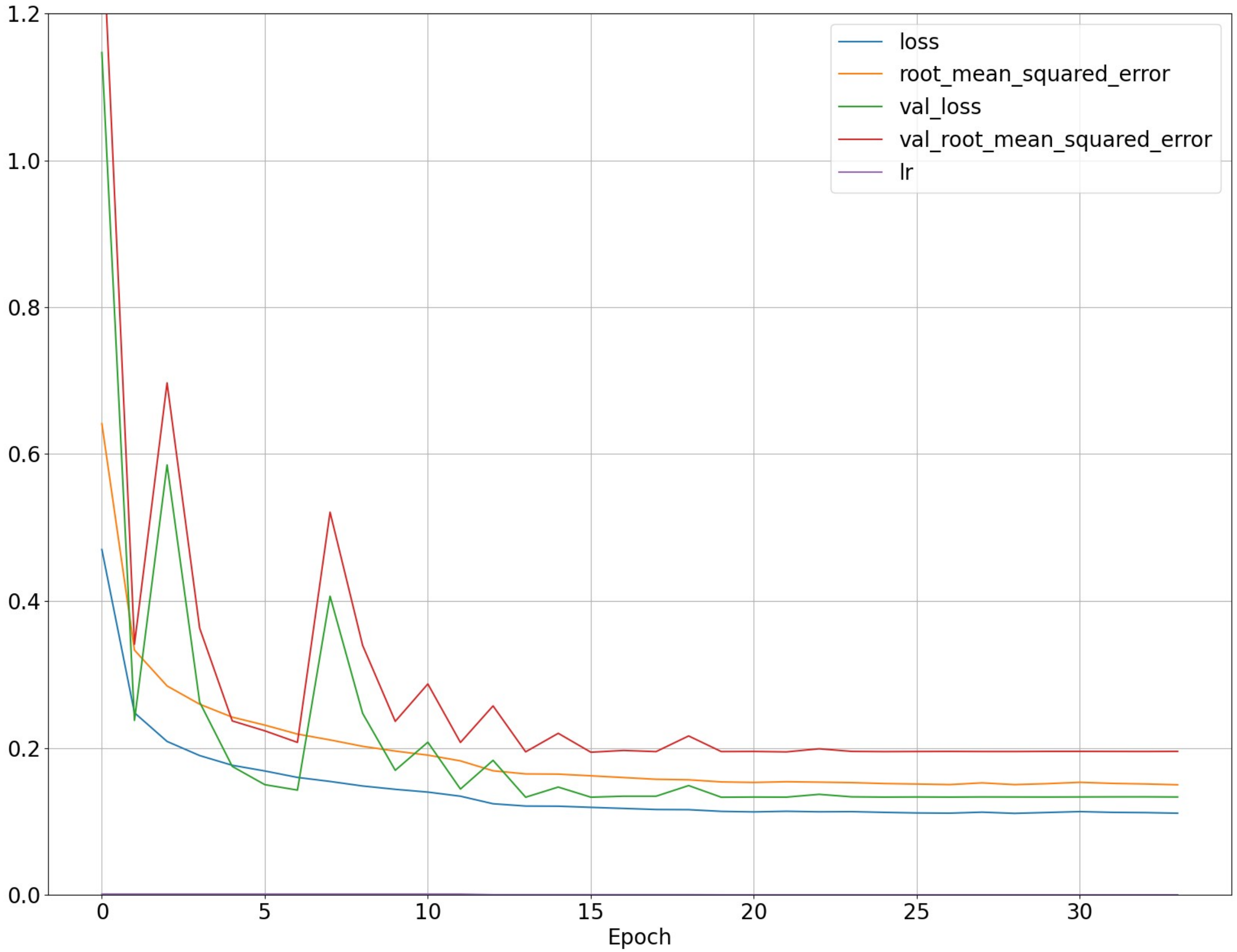
- 357 nodes
- 63.19 TB RAM
- 120 GPUs
- > 1 PB disk space

https://slurm.schedmd.com/

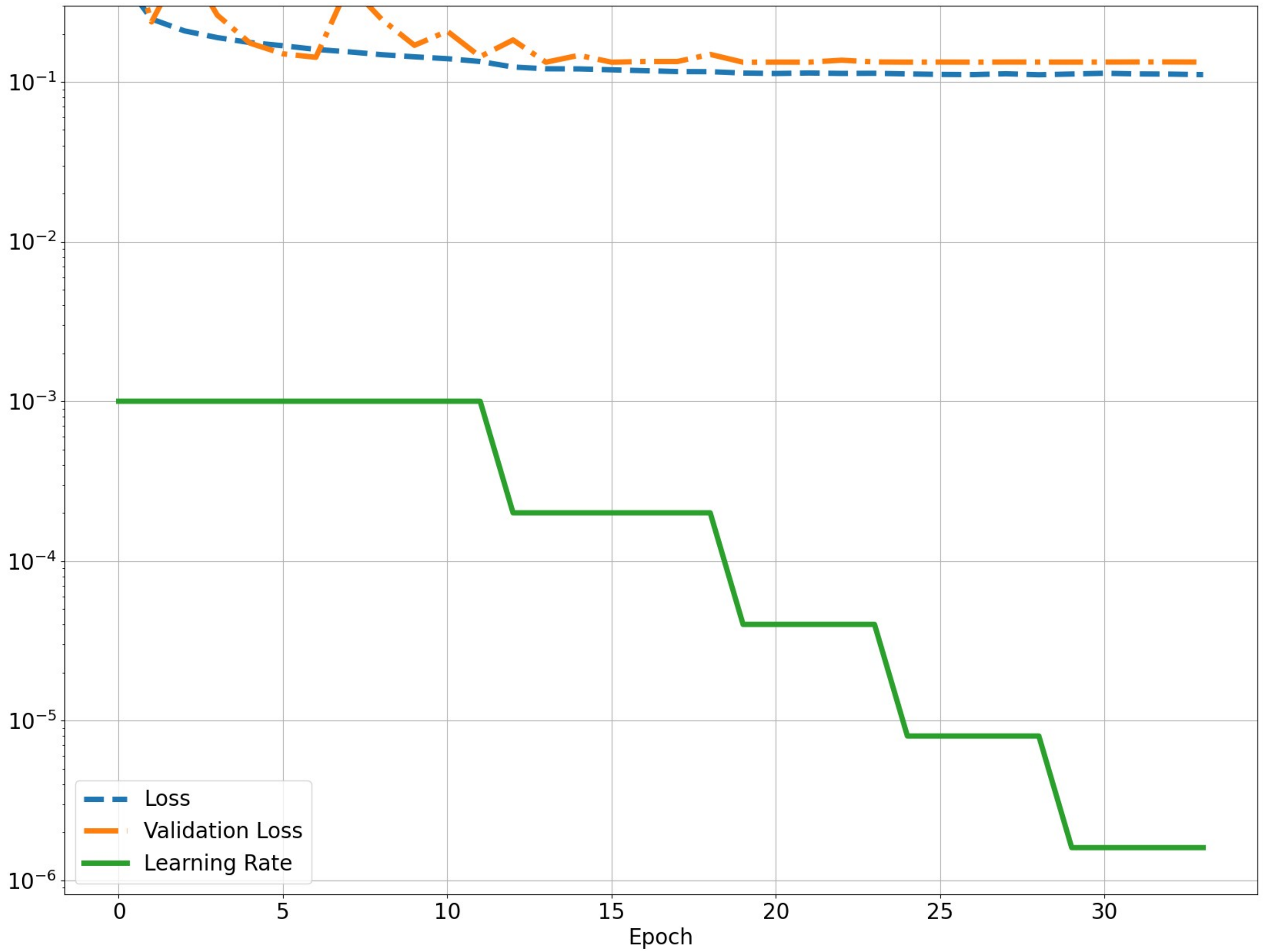
https://datascience.hawaii.edu/about-mana/

ResNet Training

- We have the model.
- We have the images.
- So, let’s train.



1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Epoch
loss
root_mean_squared_error
val_loss
val_root_mean_squared_error
lr



10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
0
5
10
15
20
25
30
Epoch
Loss
Validation Loss
Learning Rate

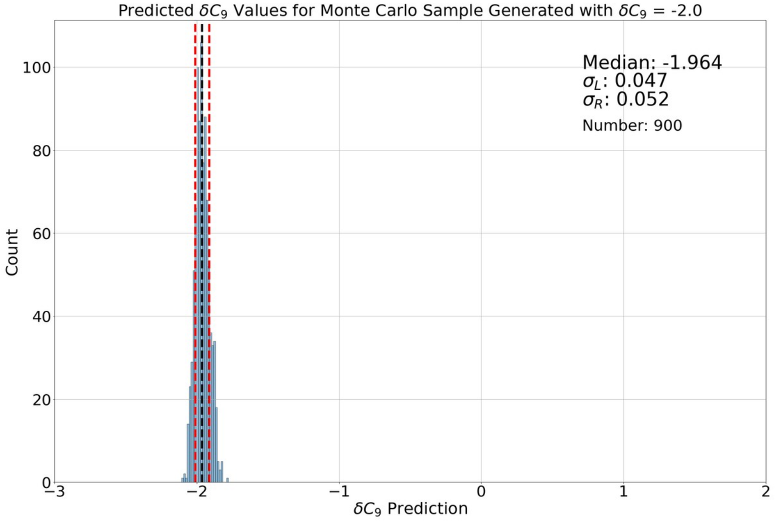
Testing the Trained Model

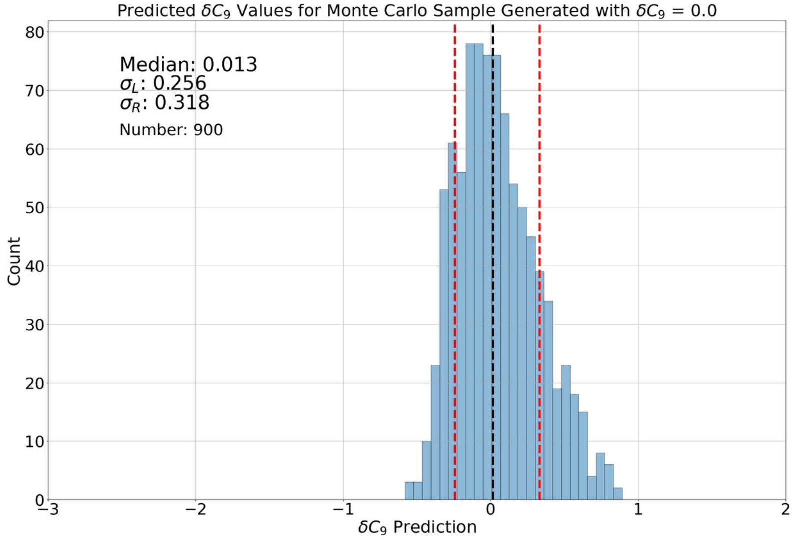
Testing the Trained Model

- How do we test model?
- Have the trained ResNet extract δC_9 values from unseen test images
  - 22 sets of images (for 22 different δC_9 values), as well as for values in between, i.e. images generated for values the NN has not been trained on
  - 900 images in each test set
- Fit resulting δC_9 distributions and produce linearity plot
  - Output δC_9 vs Input δC_9 (Predicted vs Actual)

Testing the Trained Model

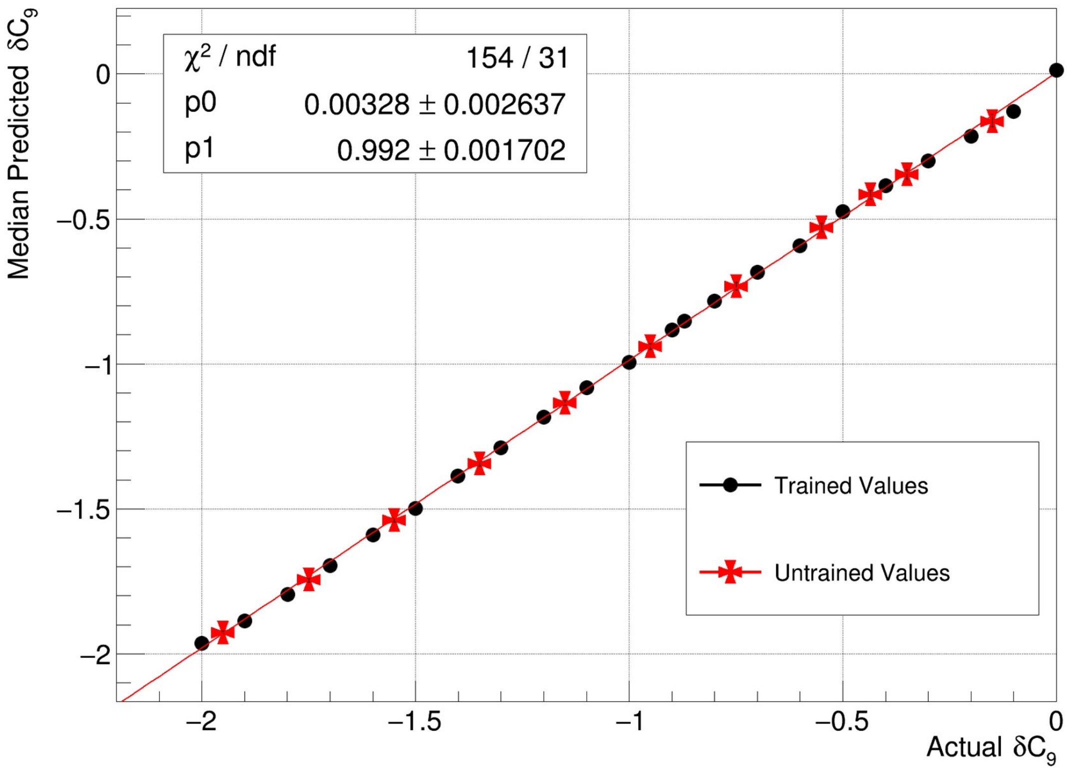
We are able to perform regression to obtain the correct δC_9 values in image ensemble tests.





Examples

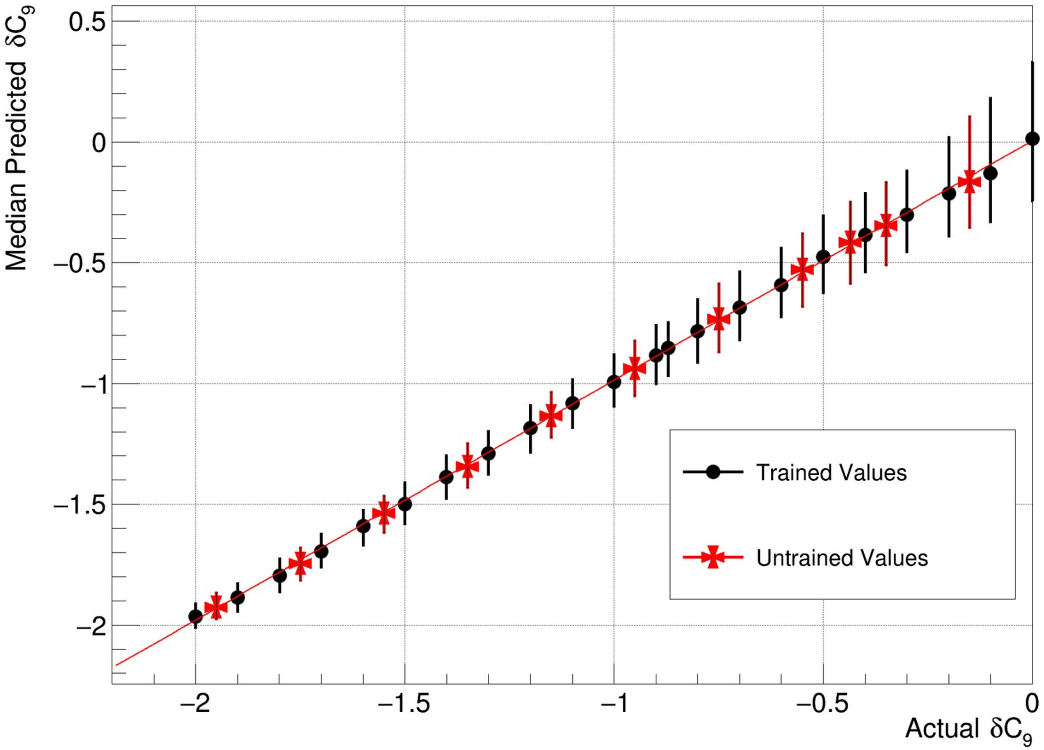
ResNet Linearity Test from MC Ensembles



Higher uncertainty toward SM values

Likely mitigated with more training data

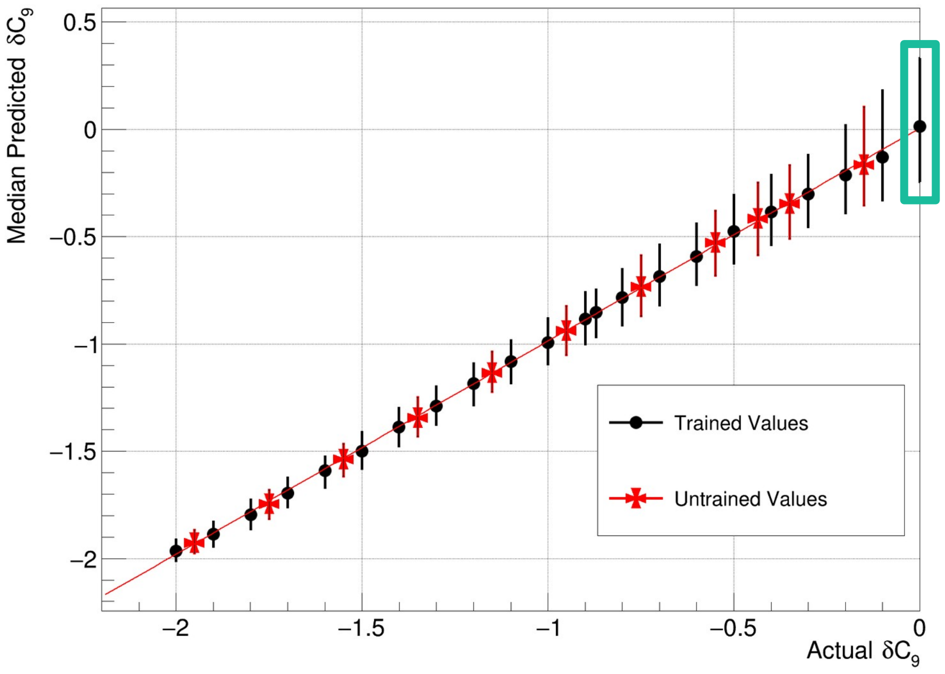
ResNet Linearity Test from MC Ensembles

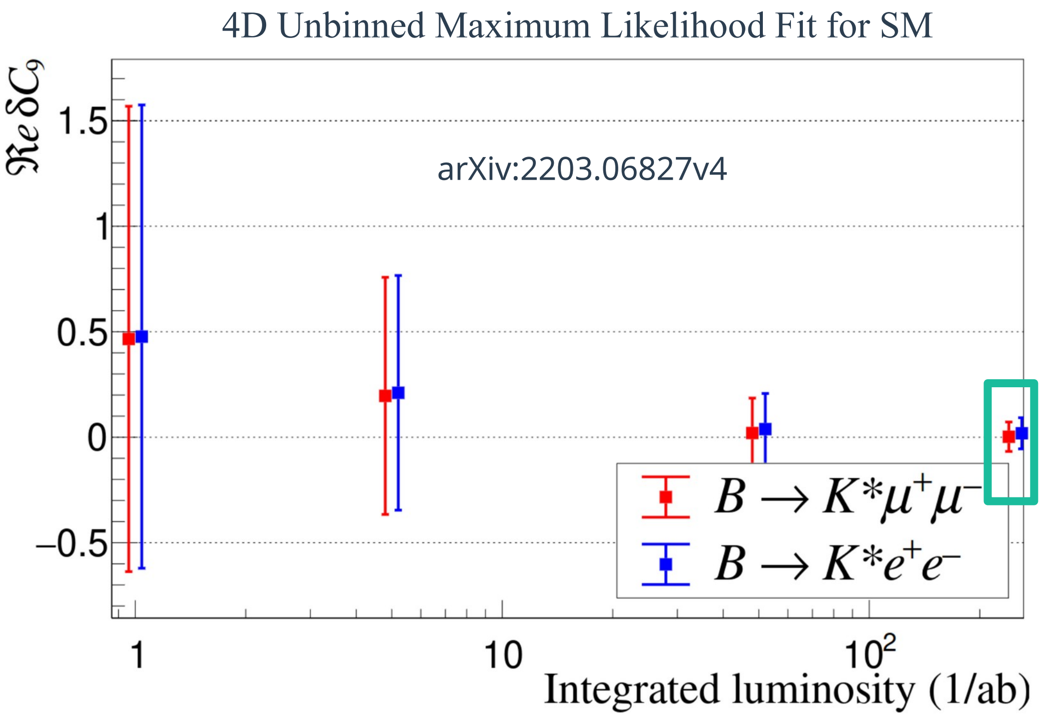


Higher uncertainty toward SM values

Likely mitigated with more training data

ResNet Comparison to Maximum Likelihood Fitting





Summary and Conclusion

Summary and Conclusion 1/3

- **What is our motivation?**
  - Ratios of BFs $B \rightarrow K^{*0} l^+ l^-$ seem to be consistent with SM
  - However, there may be anomalous behavior in angular asymmetries
- **What have we done?**
  - Implemented a NP MC generator for $B \rightarrow K^{*0} l^+ l^-$
  - Generated high statistics MC samples
  - Each MC sample has a different NP parameterization
  - Used the decay information to produce 3D images that contain the decay angular info
  - Trained a NN (ResNet) with these images

Summary and Conclusion 2/3

BROWN

- **Proof-of-concept ResNet CNN regression model is able to learn correlations between images and NP values, i.e. between decay angle distributions and NP values**
  - Can directly extract δC_9 values
  - Higher uncertainty for values closer to SM
  - Currently generating more training images to improve error bars (compared to 4D max. like. fitting)
- **To do: add background; add Geant4 detector simulation and Belle II reconstruction**
  - These require intense computing resources
    - Assuming the Belle reconstruction efficiency (~25%), require at least 128×10^6 signal MC events for this integrated luminosity
    - Dominant background is K* with incorrect lepton combinations
      - Full background simulations will be expensive for high statistics studies so to save computational power, simulate only K* with two leptons
    - But NN does not know about statistics; need to generate images for [1, 5, 50, 100]/ab dataset sizes

Summary and Conclusion 3/3

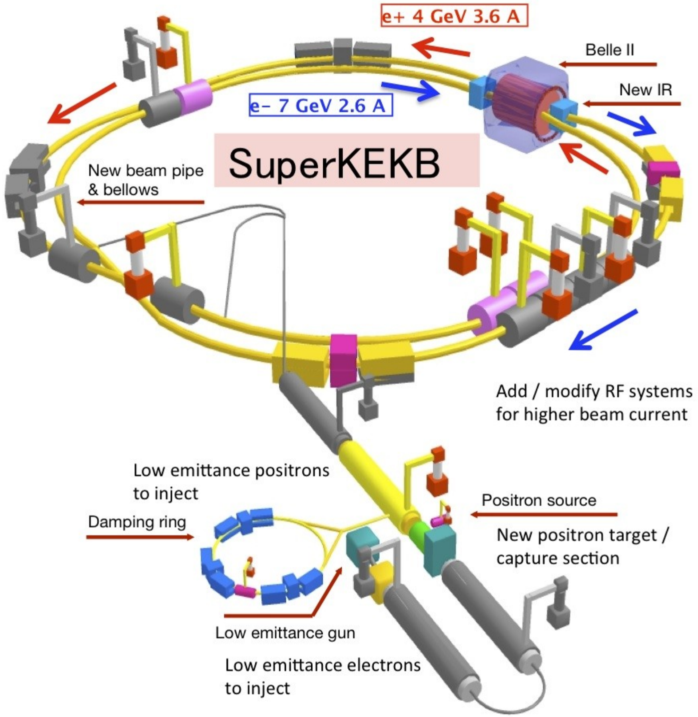
- **Similar project planned for $\overline{B} \rightarrow D^{*+} l^- \overline{\nu}$**
  - Phys. Rev. D 107, 015011 (2023)

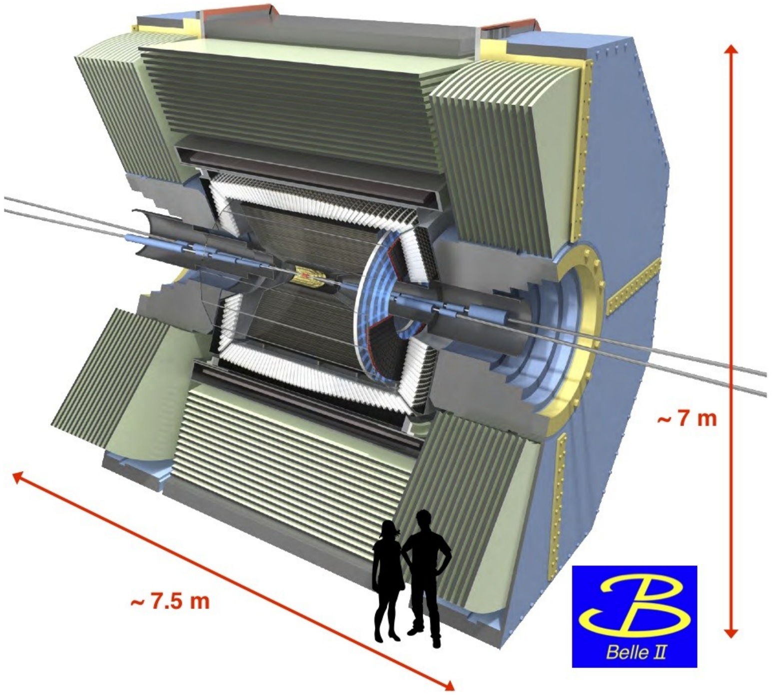
Acknowledgments

- Done in collaboration with T.E. Browder (U. Hawaii Manoa), S. Kohani (U. Hawaii Manoa), R. Mandal (IIT Gandhinagar), S. Sandilya (IIT Hyderabad), A. Sibidanov (U. Hawaii Manoa), R. Sinha (IMSc, U. Hawaii Manoa), and S.E. Vahsen (U. Hawaii Manoa)
- Thanks to Hongyang Gao (ISU CS) for suggesting computer vision and Chunhui Chen (ISU Physics)
- Thanks to Peter Sadowski (UHM ICS) and Jeff Schueler (UNM) for helpful ML technical discussions and advice

Backup

Belle II



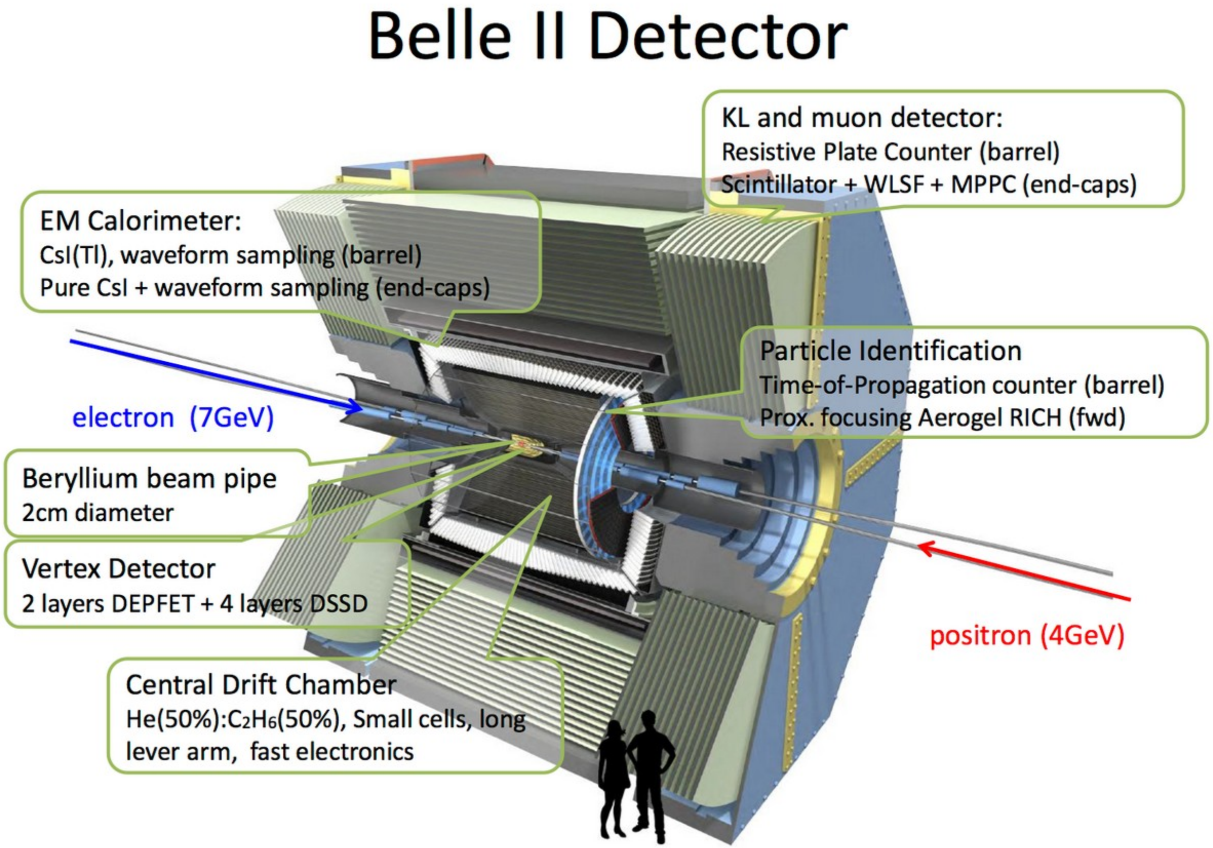


https://www.belle2.org/project/super_kekb_and_belle_ii

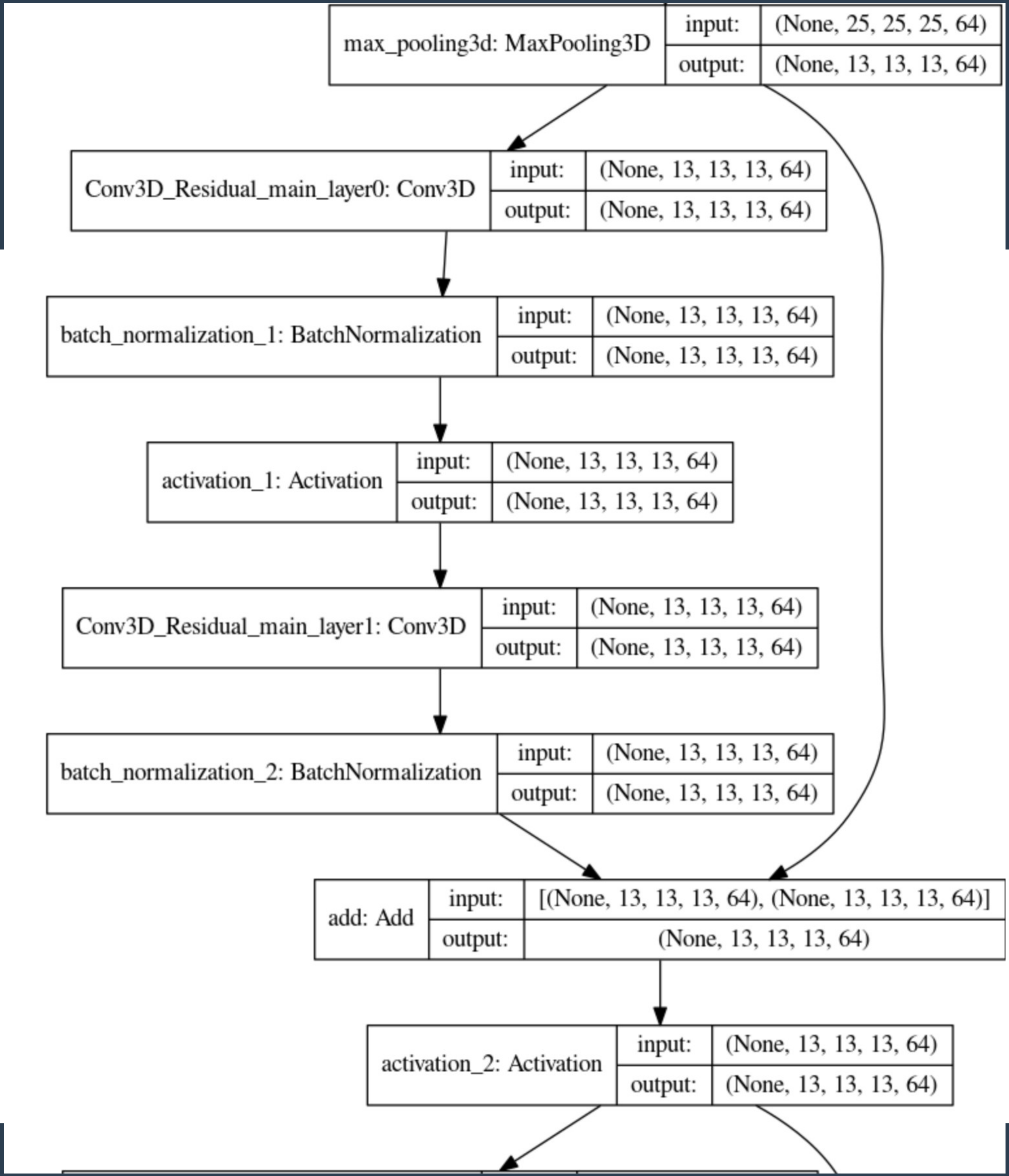
Belle II

https://www.nature.com/articles/nj0320

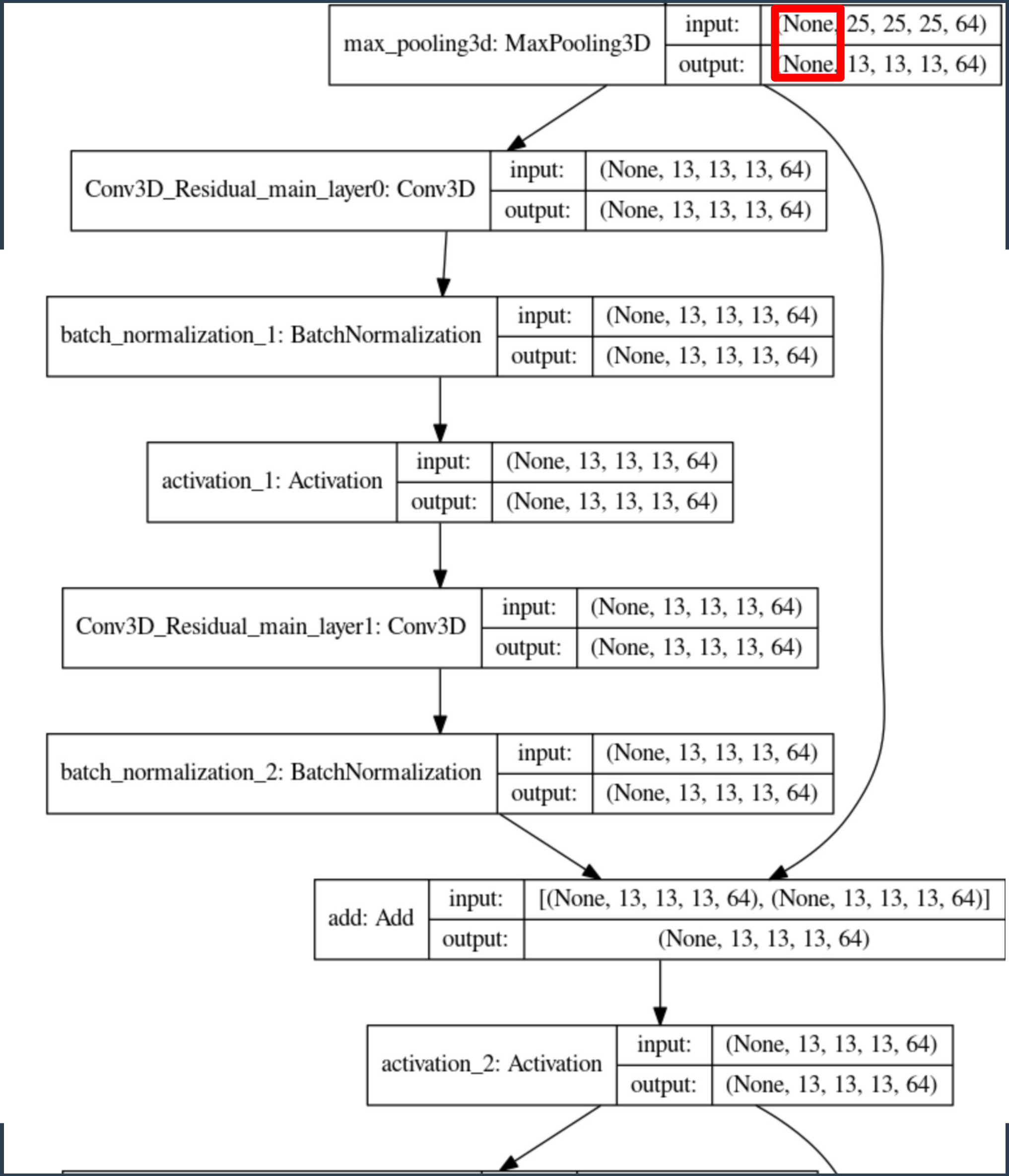
Belle II



Belle II Detector
KL and muon detector:
Resistive Plate Counter (barrel)
Scintillator + WLSF + MPPC (end-caps)
EM Calorimeter:
CsI(Tl), waveform sampling (barrel)
Pure CsI + waveform sampling (end-caps)
Particle Identification
Time-of-Propagation counter (barrel)
Prox. focusing Aerogel RICH (fwd)
electron (7GeV)
Beryllium beam pipe
2cm diameter
Vertex Detector
2 layers DEPFET + 4 layers DSSD
positron (4GeV)
Central Drift Chamber
He(50%):C2H6(50%), Small cells, long
lever arm, fast electronics

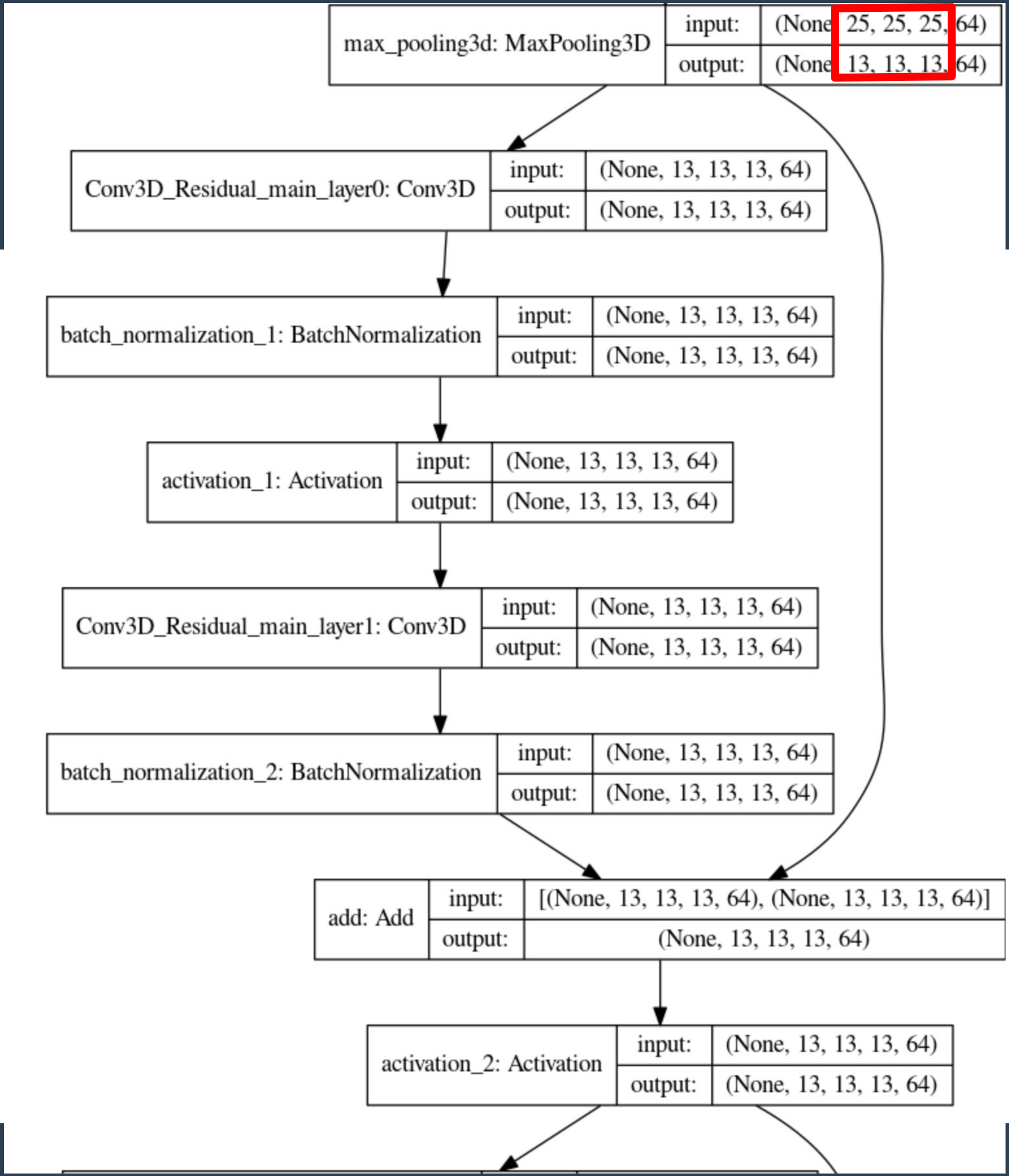


Based on arxiv:1512.03385

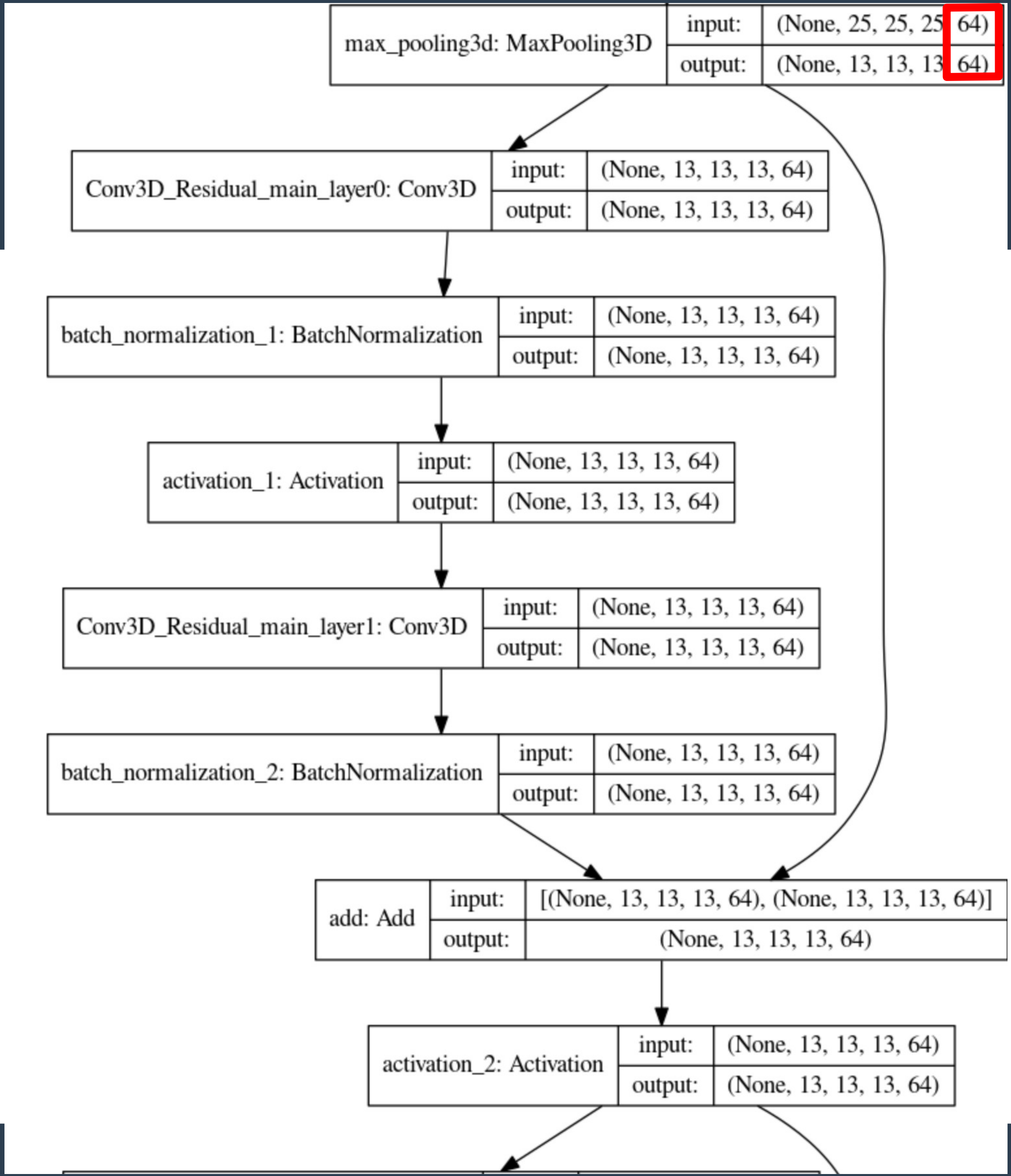


Specifies number of images; None = unspecified

Based on arxiv:1512.03385

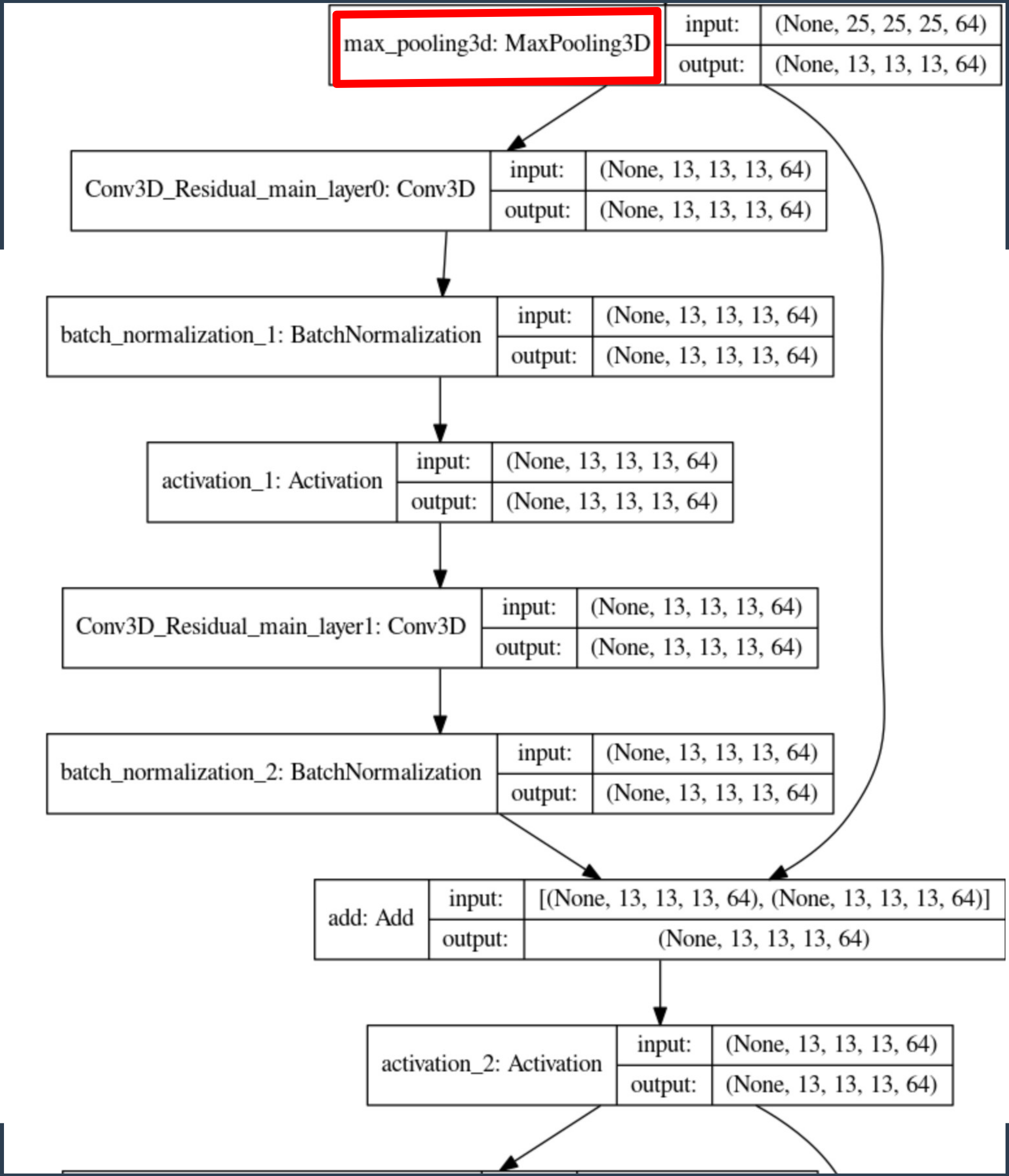


Downsampled image size



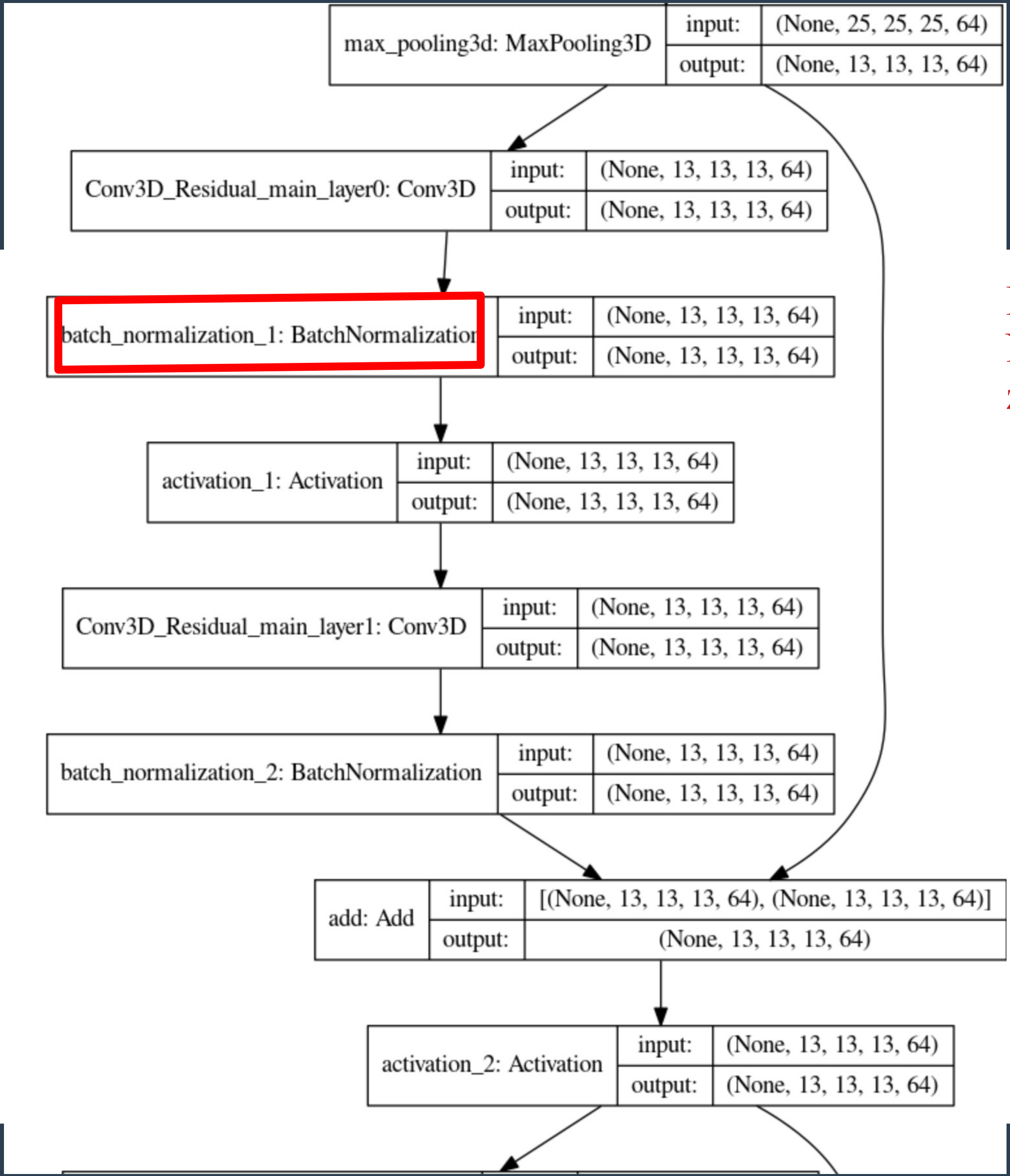
Number of filters

Based on arxiv:1512.03385



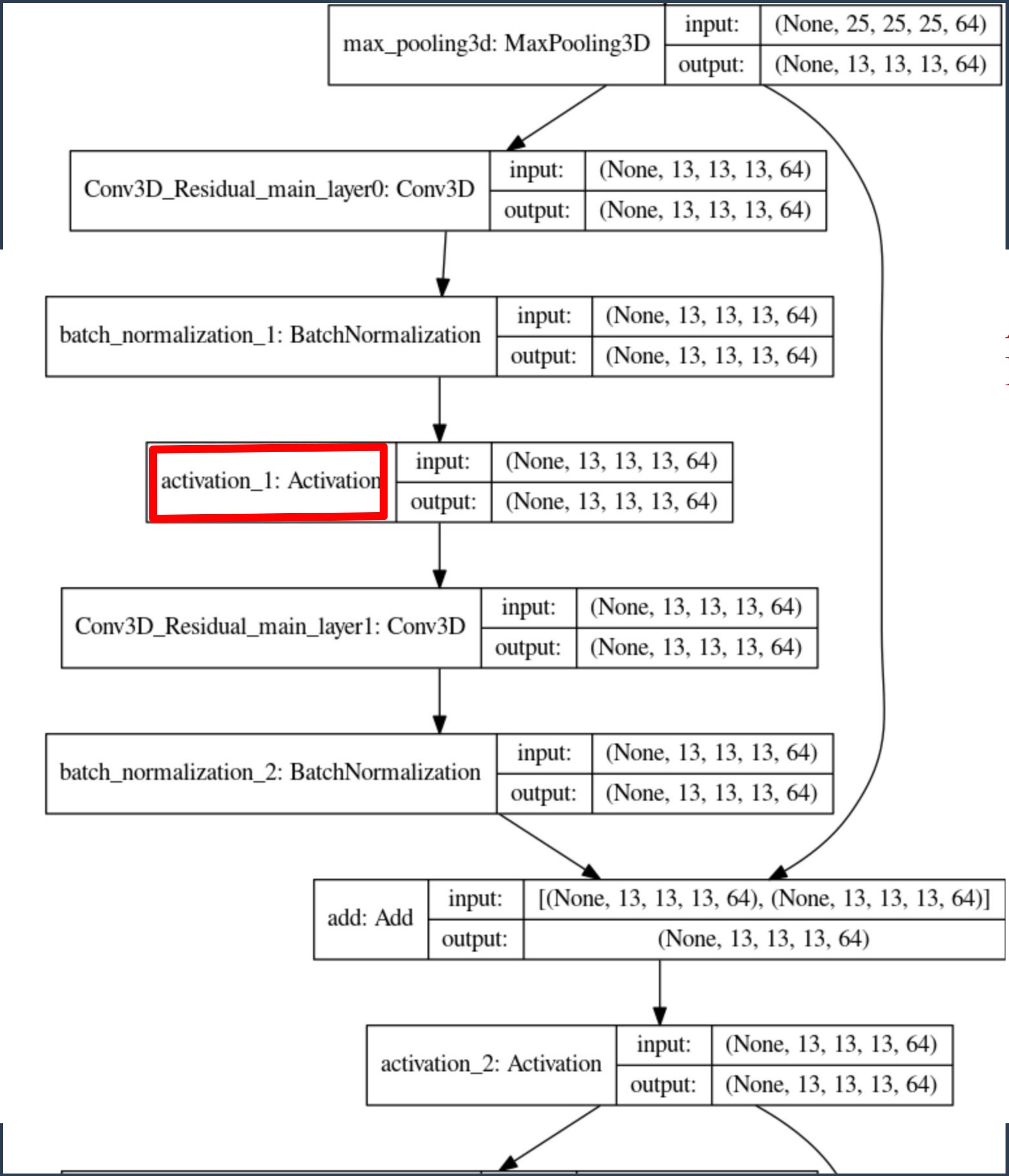
Downsampling layer

Based on arxiv:1512.03385



Batch Normalization layer; Normalizes feature to have zero mean and unit variance

Based on arxiv:1512.03385



Activation layer;
ReLU layer

Based on arxiv:1512.03385

ReLU

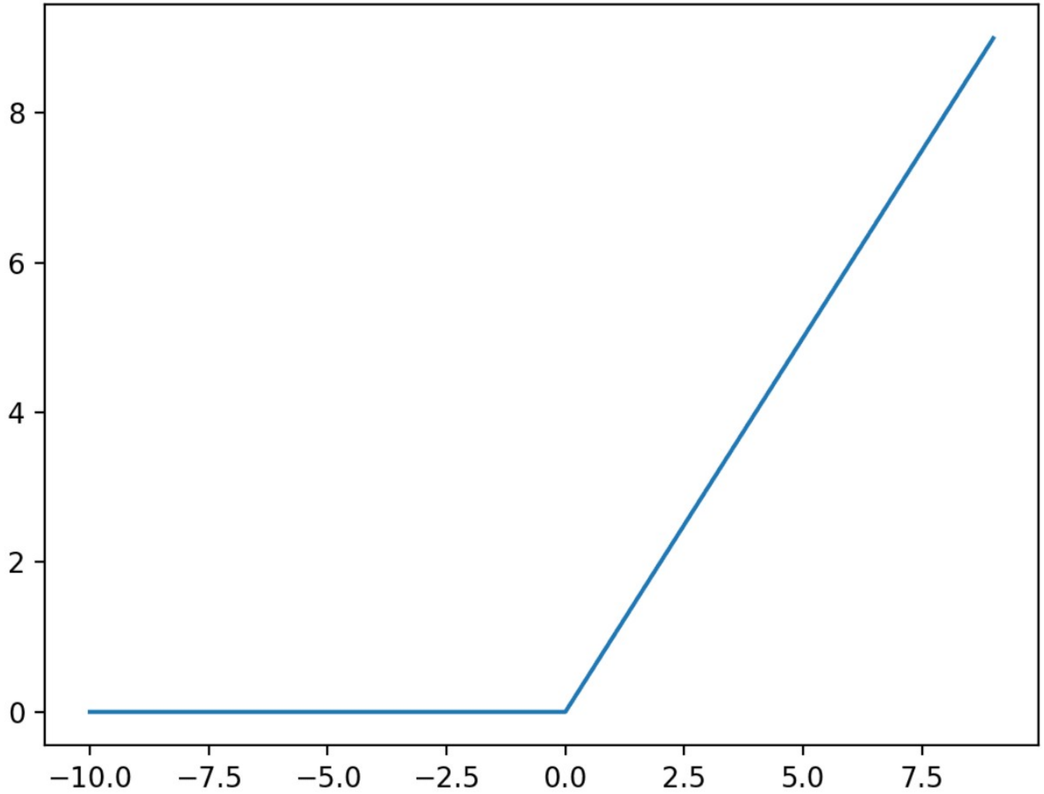
- **ReLU activation function**
  - rectified linear unit

$$f(x) = x^+ = \max(0, x) = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$f'(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x < 0. \end{cases}$$

https://en.wikipedia.org/wiki/Rectifier_(neural_networks)

ReLU



Backup – Previous Work

Fully Connected Neural Network

BROWN

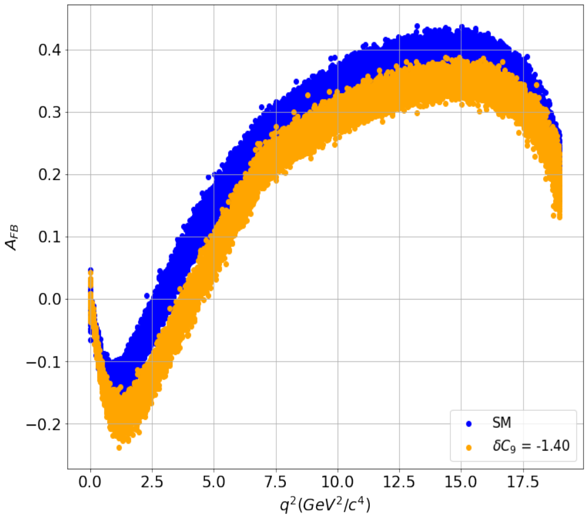
- We developed a fully connected neural network and used generator-level MC to train a classifier that distinguished between new physics and Standard Model scenarios
- We first try and train on angular asymmetries A_{FB} and S_5, and q^2
  - Loss is binary cross-entropy
  - Metric is accuracy
- Implement a likelihood-free inference method using binned template fitting to determine δC_9 values
  - Template histograms generated from output of NN classifier

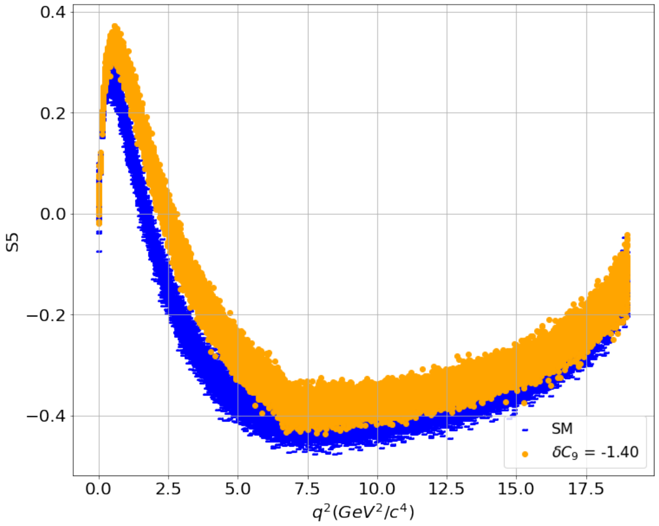
Fully Connected Neural Network

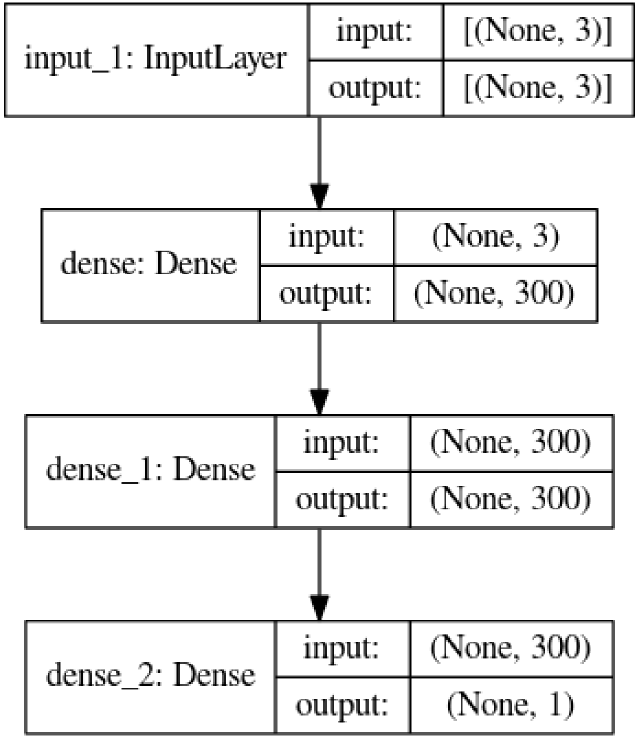
$$A_{\mathrm{FB}}(q^2) = \frac{\left[\left(\int_0^1 - \int_{-1}^0\right) \mathrm{d}\cos\theta_\ell\right] \mathrm{d}(\Gamma - \bar{\Gamma})}{\int_{-1}^1 \mathrm{d}\cos\theta_\ell \, \mathrm{d}(\Gamma + \bar{\Gamma})}$$

$$S_5(q^2) = \frac{4}{3} \frac{\left[\int_{-\pi/2}^{\pi/2} - \int_{\pi/2}^{3\pi/2}\right] \mathrm{d}\chi \left[\int_0^1 - \int_{-1}^0\right] \mathrm{d}\cos\theta_K \int_{-1}^1 \mathrm{d}\cos\theta_\ell \;\; \mathrm{d}(\Gamma - \bar{\Gamma})}{\int_0^{2\pi} \mathrm{d}\chi \int_{-1}^1 \mathrm{d}\cos\theta_K \int_{-1}^1 \mathrm{d}\cos\theta_\ell \; \mathrm{d}(\Gamma + \bar{\Gamma})}$$

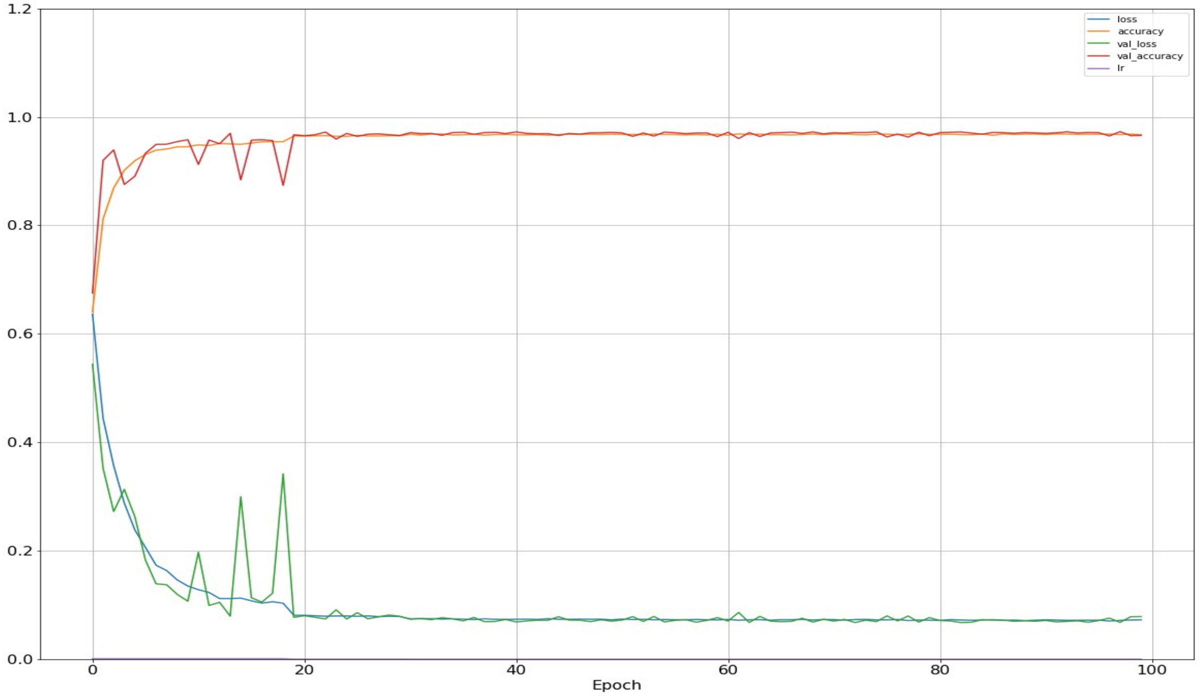
Fully Connected Neural Network



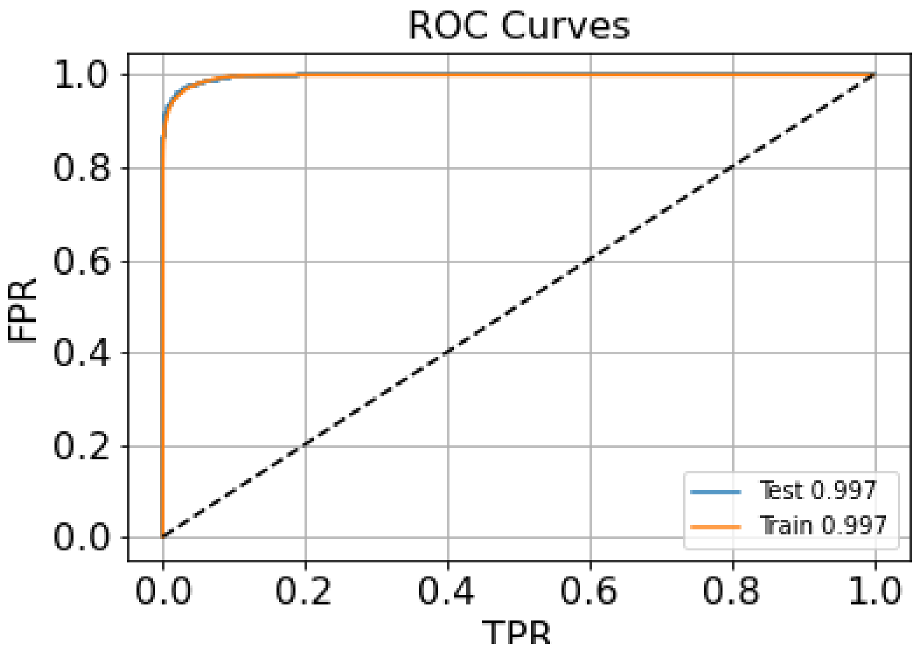


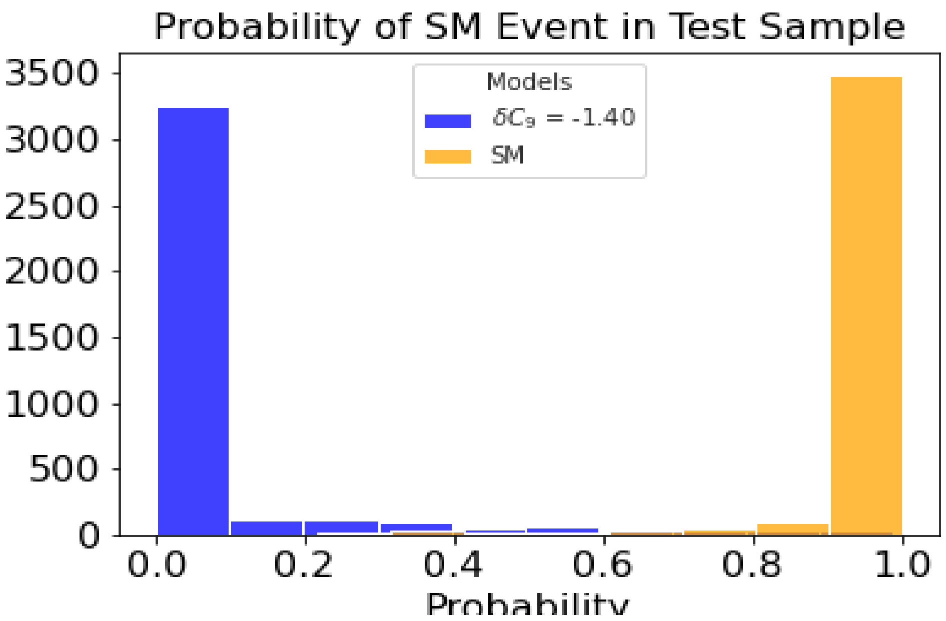


Fully Connected Neural Network Training

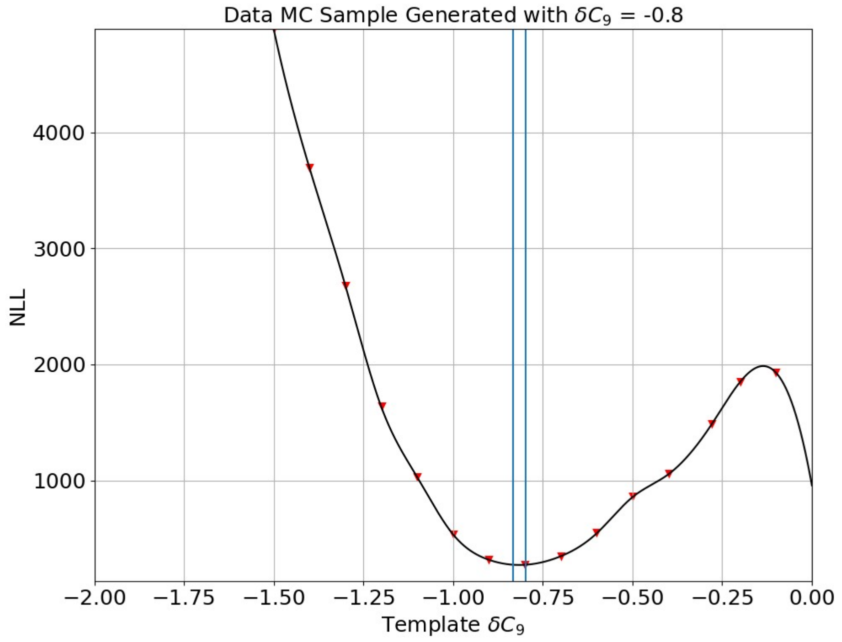


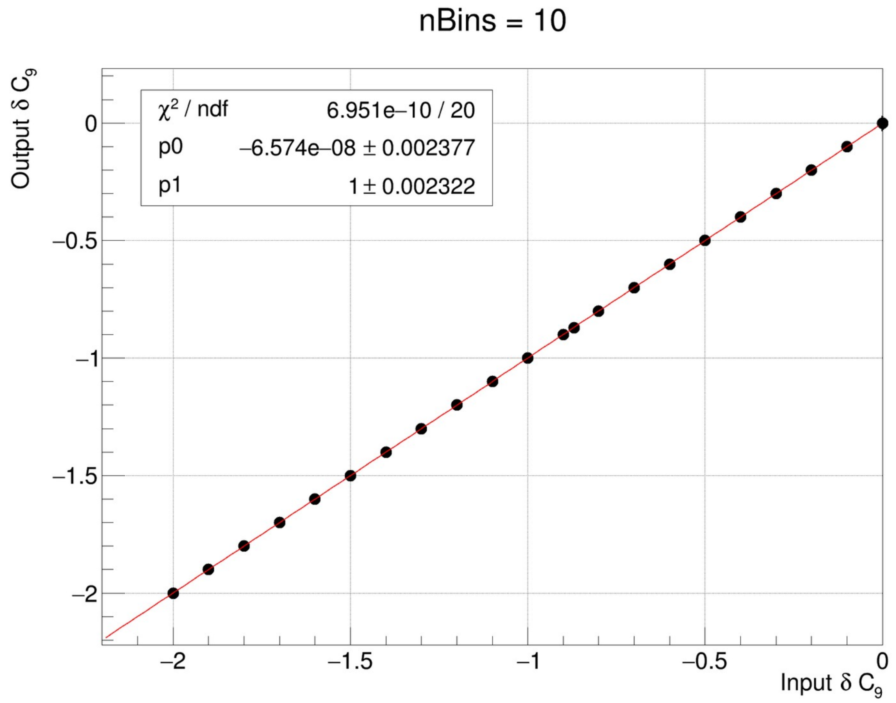
Fully Connected Neural Network Training





Fully Connected Neural Network Results



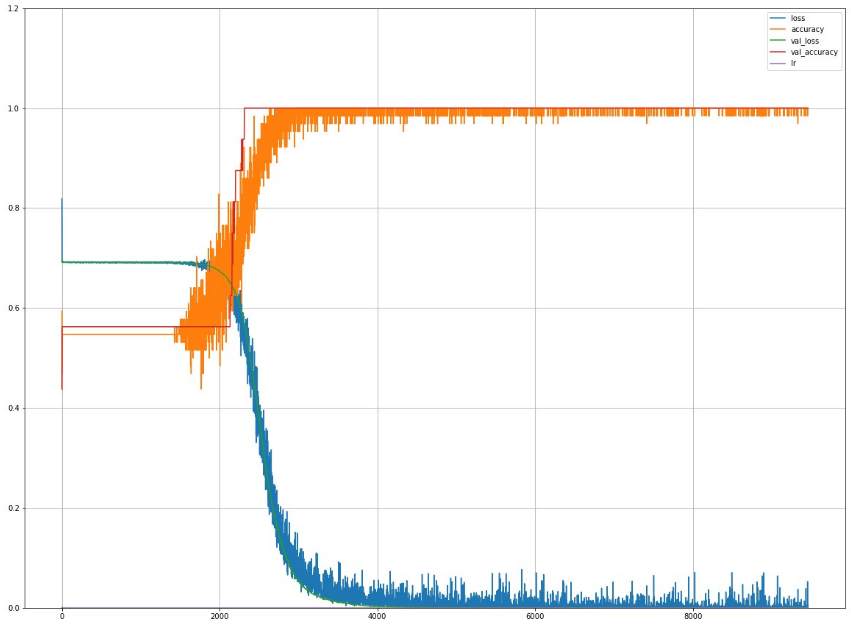


(Only one MC exp)

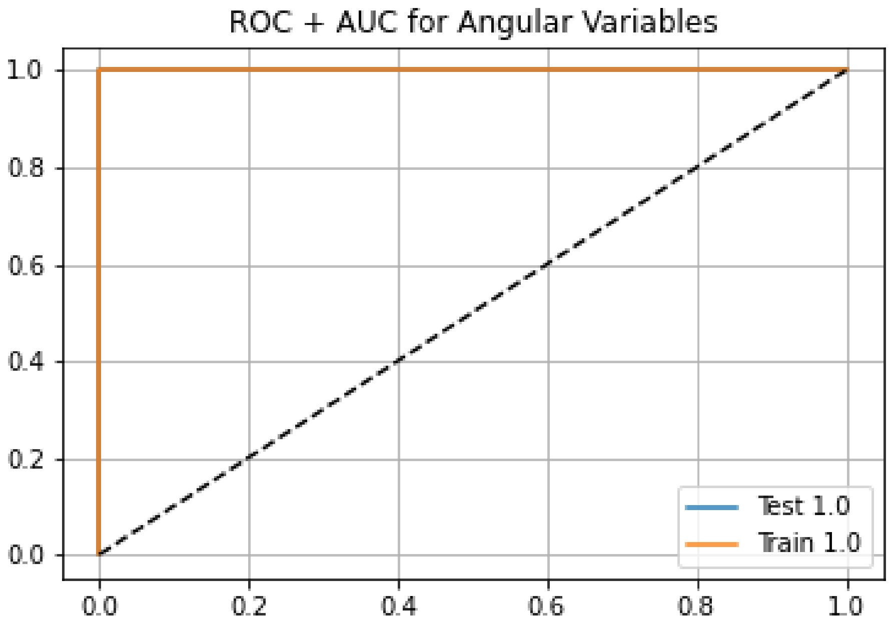
Fully Connected Neural Network Results

- **Fully connected NN using angular asymmetries seems to be a more difficult path**
  - Computationally intensive to generate many MC exps: extracting A_{FB} and S_5 requires a lot of resources and many MC samples
  - Not as easy to add background to angular asymmetries
  - Using angular asymmetries not ideal: don't directly use all information from the decay

Convolutional Neural Network Classifier Training



Convolutional Neural Network Classifier Training



Convolutional Neural Network Classifier

- Training seems to work very well but now what?
- Cannot use this method to analyze real data. Each data set will constitute only a single image.
  - Cannot reasonably fit one image with a template generated from MC (templates produced from multiple images given high statistics MC samples)
- What to do? Consider CNN regression.

Theory

BROWN

$$C_7^{\text{eff}} = -0.304,$$
$$C_9^{\text{eff}} = C_9 + Y(q^2) = 4.211 + Y(q^2)\,,$$
$$C_{10} = -4.103\,,$$

$$C_7^{\text{eff}} = C_7 - \frac{1}{3}C_3 - \frac{4}{9}C_4 - \frac{20}{3}C_5 - \frac{80}{9}C_6\,,$$
$$C_9^{\text{eff}} = C_9 + Y(q^2)\,,\ \text{with}$$
$$\begin{aligned} Y(q^2) = {} & h(q^2, m_c)\left(\frac{4}{3}C_1 + C_2 + 6C_3 + 60C_5\right) \\ & - \frac{1}{2}h(q^2, m_b)\left(7C_3 + \frac{4}{3}C_4 + 76C_5 + \frac{64}{3}C_6\right) \\ & - \frac{1}{2}h(q^2, 0)\left(C_3 + \frac{4}{3}C_4 + 16C_5 + \frac{64}{3}C_6\right) \\ & + \frac{4}{3}C_3 + \frac{64}{9}C_5 + \frac{64}{27}C_6\,. \end{aligned} \tag{A1}$$

$$\begin{aligned} h(q^2, m_q) = {} & -\frac{4}{9}\left(\ln\frac{m_q^2}{\mu^2} - \frac{2}{3} - z\right) \\ & - \frac{4}{9}(2+z)\sqrt{|z-1|} \times \begin{cases} \arctan\dfrac{1}{\sqrt{z-1}} & z > 1 \\ \ln\dfrac{1+\sqrt{1-z}}{\sqrt{z}} - \dfrac{i\pi}{2} & z \le 1 \end{cases} \end{aligned} \tag{A2}$$

$$h(q^2, 0) = \frac{8}{27} + \frac{4}{9}\left(\ln\frac{\mu^2}{q^2} + i\pi\right)$$