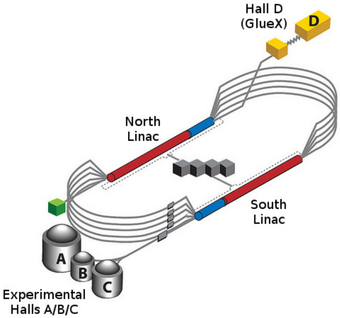
Machine Learning and Precision Analysis at GlueX

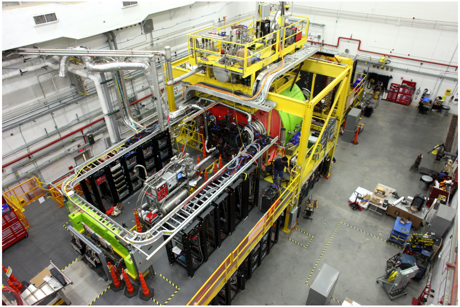
Daniel Lersch, Sean Dobbs

Florida State University

12.10.2018

GlueX at Thomas Jefferson National Laboratory





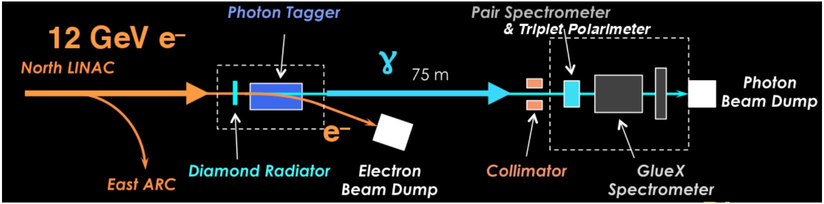
Experimental Hall D:

- Over 125 scientists from:
  - 28 Institutions
  - 10 Countries
- Experiments with polarized photon beam

Physics Program at GlueX:

- Study properties of strong force
  (Binds quarks into protons, protons-neutrons into nuclei)
- Search for new particles or new particle states
  $\Rightarrow$ Baryon- /Meson - Spectroscopy
- Test fundamental symmetries in physics:
  $\Rightarrow$ Rare decay modes of the $\eta^{(\prime)}$-meson

Photo-Production Data



- Photon beam with energies $\in [3\,\mathrm{GeV}, 12\,\mathrm{GeV}]$
- Do not produce just one particle, but a whole bunch of them

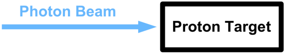
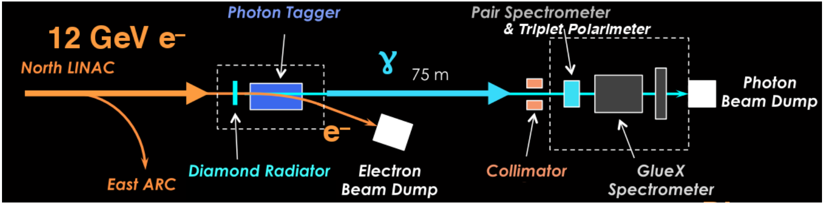


Photo-Production Data



- Photon beam with energies $\in [3\,\mathrm{GeV}, 12\,\mathrm{GeV}]$
- Do not produce just one particle, but a whole bunch of them
- Some production mechanisms are more dominant
- Final states with similar topology, but different particles:
  - $\eta \rightarrow \pi^+\pi^-\gamma \leftrightarrow \eta \rightarrow e^+e^-\gamma$
  - $\Phi \rightarrow K^+K^- \leftrightarrow \rho \rightarrow \pi^+\pi^-$
  - $\rho \rightarrow \pi^+\pi^-$ + fake photon $\leftrightarrow \eta \rightarrow \pi^+\pi^-\gamma$
- In general: Pions are dominating

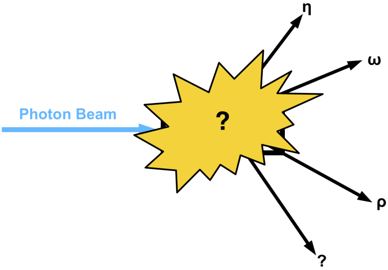
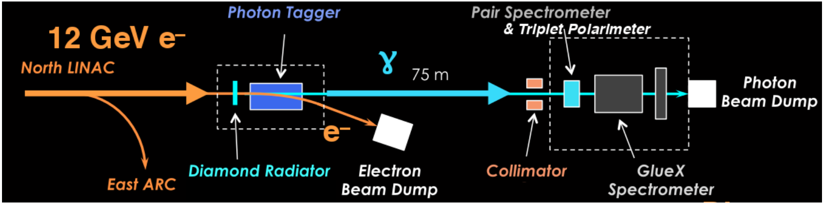
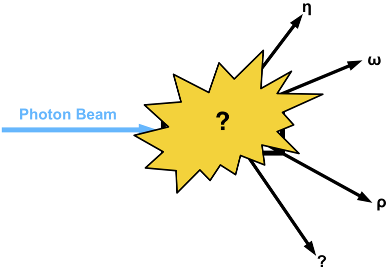


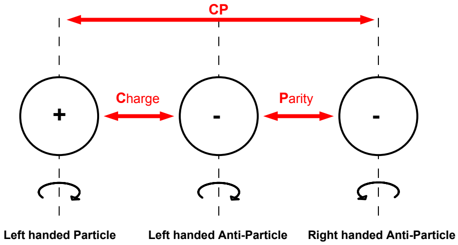
Photo-Production Data

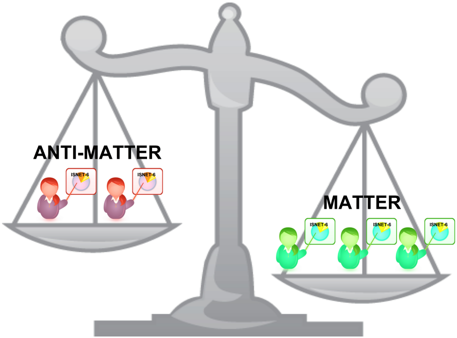


- Photon beam with energies $\in [3\,\mathrm{GeV}, 12\,\mathrm{GeV}]$
- Do not produce just one particle, but a whole bunch of them
- Some production mechanisms are more dominant
- Need reliable algorithms/methods to:
  - i) Reconstruct the measured data properly **(Kalmann-Filter, Clustering,...)**
  - ii) Identify particle final states correctly **(Kinematic Fitting,...)**
  - iii) Figure out what is going on in the yellow area **(Partial Wave Analysis,...)**



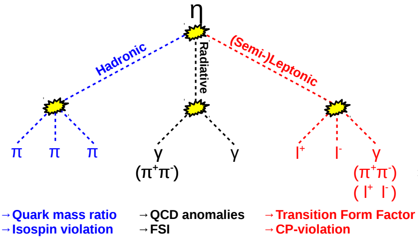
Symmetries and CP-Violation





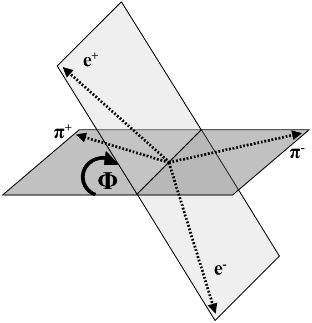
- Symmetries in physics: **C**harge-, **P**arity- and **T**ime- conjugation
- CP-Violation is one (of three necessary) condition(s) required to cause an imbalance between matter and anti-matter
  (A. Sakharov)
- Candidates to study CP-Violation: K^0-, B^0- and $\eta^{(\prime)}$-Decays

The Anomalous Decay $\eta^{(\prime)} \to \pi^+\pi^-e^+e^-$



- $\eta^{(\prime)}$-Mesons are allrounder for interesting physics studies
- Look at decay: $\boldsymbol{\eta^{(\prime)}} \to \pi^+\pi^-e^+e^-$ to study CP-violation:

$\Rightarrow$ Asymmetry A_Φ of angle Φ between $\pi^+\pi^-$-e^+e^--decay planes



- Upper limit predicted by theory: $A_\Phi \sim 1\%$

  D. Gao. *Mod. Phys. Lett.*, A17:1583-1588,(2002)
- Current experimental results:
  $A_\Phi = (-0.6 \pm 2.5_{stat} \pm 1.8_{sys}) \cdot 10^{-2}$

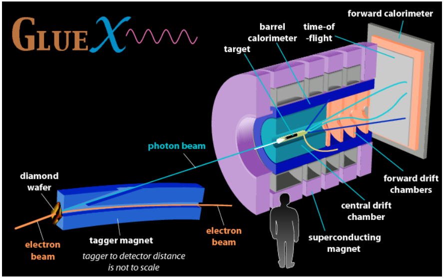
  KLOE coll. *Phys. Lett.*, B675:283-288,(2009)

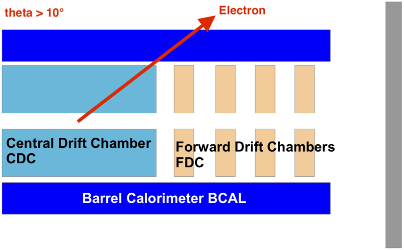
$\Rightarrow$ Particle Identification is crucial for precise/sensitive measurement

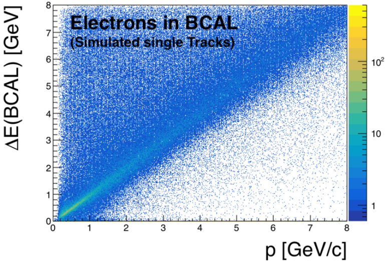
$\Rightarrow$ **Utilize machine learning for classification between $\pi^\pm$ and $e^\pm$**

The GlueX-Detector and Particle Identification

- Reconstruction of charged particles:
  - Magnetic field + Drift Chamber
  - Energy deposits in calorimeters
  - Different detector sub-parts used depending on θ-Angle of particle
- Goal(s):
  i) Reproduce detector response for each particle species
  ii) Use detection pattern for classification

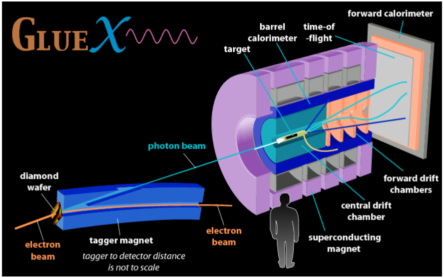


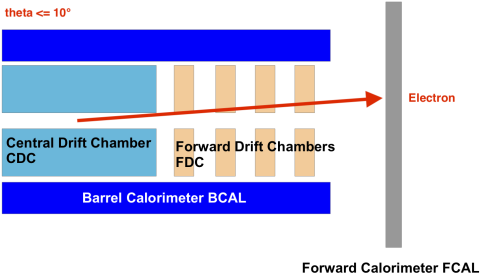


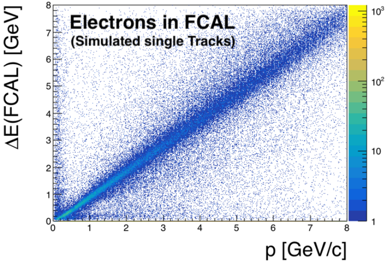


The GlueX-Detector and Particle Identification

- Reconstruction of charged particles:
  - Magnetic field + Drift Chamber
  - Energy deposits in calorimeters
  - Different detector sub-parts used depending on θ-Angle of particle
- Goal(s):
  - i) Reproduce detector response for each particle species
  - ii) Use detection pattern for classification

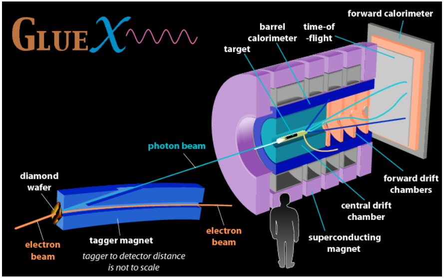


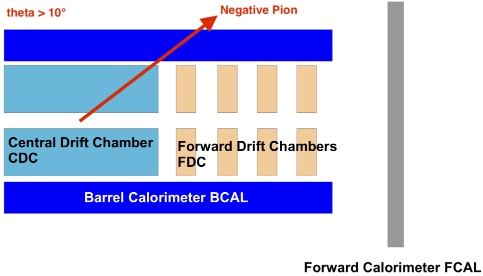


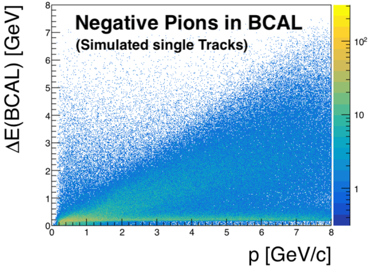


The GlueX-Detector and Particle Identification

- Reconstruction of charged particles:
  - Magnetic field + Drift Chamber
  - Energy deposits in calorimeters
  - Different detector sub-parts used depending on θ-Angle of particle
- Goal(s):
  i) Reproduce detector response for each particle species
  ii) Use detection pattern for classification

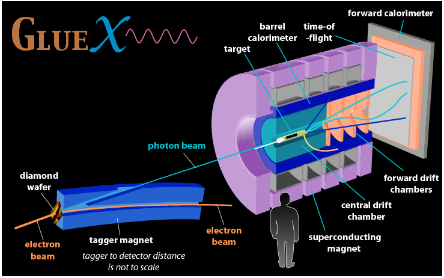


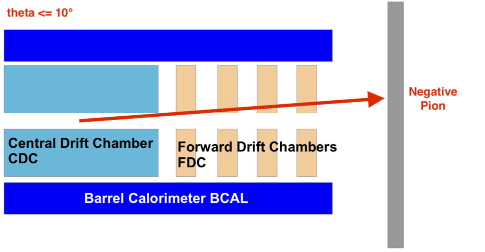


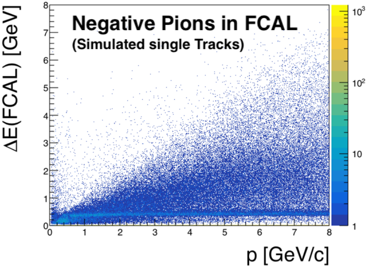


The GlueX-Detector and Particle Identification

- Reconstruction of charged particles:
  - Magnetic field + Drift Chamber
  - Energy deposits in calorimeters
  - Different detector sub-parts used depending on θ-Angle of particle
- Goal(s):
  - i) Reproduce detector response for each particle species
  - ii) Use detection pattern for classification





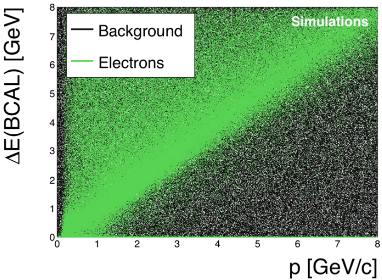


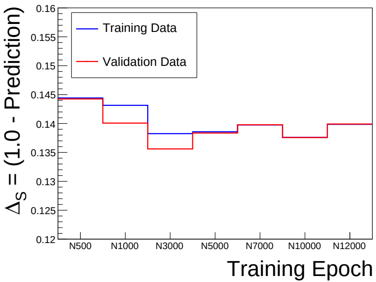
Data Set(s) and Training

- Information used for classification:

Sub-Detector	Momentum p	Angle θ	Energy Deposit ΔE
CDC	x	x	x
BCAL	-	-	x
FDC	x	x	x
FCAL	-	-	x

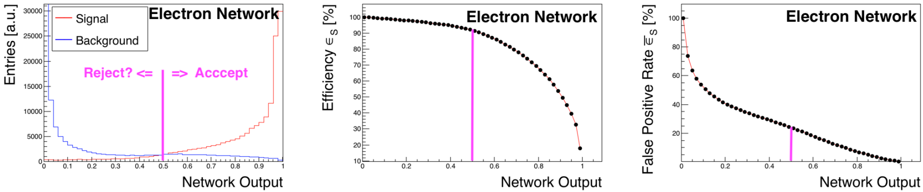
- "Classical" Approach: Train a classifier with electrons as signal and pions as background $\Rightarrow$ Not done here
- Trained neural network with simulated single particle tracks (signal) + random flat detector response (background) for: e^+, e^-, π^+ and π^-
$\Rightarrow$**One neural network per particle and per charge**

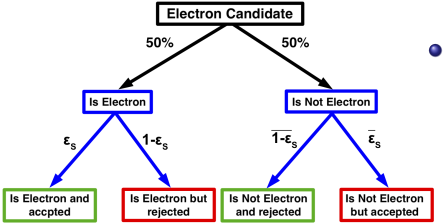




Using the Classifier-Output *

* Inspired by PhD-Thesis from Daniel Coderre (2012)



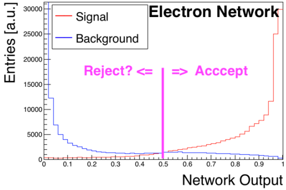


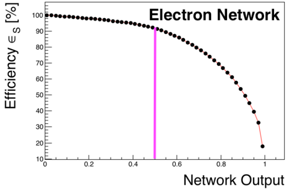
- Instead of network output, use ROC (i.e. efficiency, false identification rate) for classification
- Calculate two probabilities:

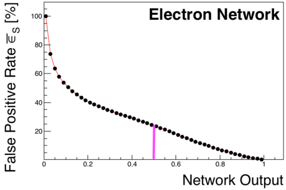
1. $$P_e = \frac{0.5 \times \epsilon_S}{0.5 \times \epsilon_S + 0.5 \times \bar{\epsilon}_S}$$
2. $$P_{\bar{e}} = \frac{0.5 \times (1 - \bar{\epsilon}_S)}{0.5 \times (1 - \bar{\epsilon}_S) + 0.5 \times (1 - \epsilon_S)}$$

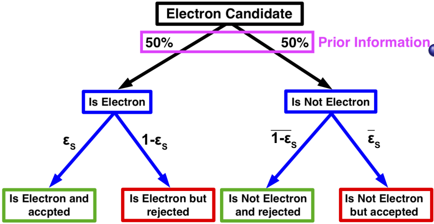
Using the Classifier-Output *

* Inspired by PhD-Thesis from Daniel Coderre (2012)







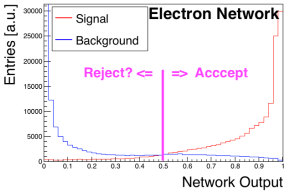


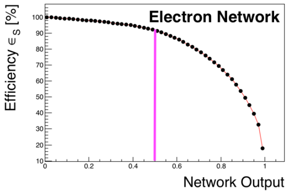
- Instead of network output, use ROC (i.e. efficiency, false identification rate) for classification
- Calculate two probabilities:

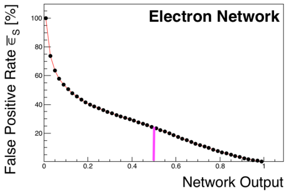
1. $$P_e = \frac{\boxed{0.5 \times \epsilon_S}}{\boxed{0.5 \times \epsilon_S} + \boxed{0.5 \times \bar{\epsilon}_S}}$$
2. $$P_{\bar{e}} = \frac{\boxed{0.5 \times (1 - \bar{\epsilon}_S)}}{\boxed{0.5 \times (1 - \bar{\epsilon}_S)} + \boxed{0.5 \times (1 - \epsilon_S)}}$$

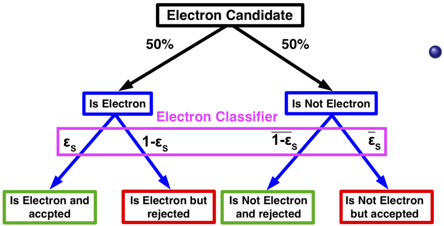
Using the Classifier-Output *

* Inspired by PhD-Thesis from Daniel Coderre (2012)







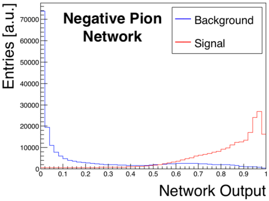


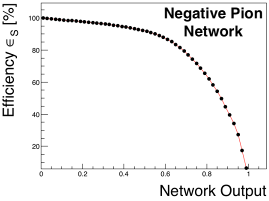
- Instead of network output, use ROC (i.e. efficiency, false identification rate) for classification
- Calculate two probabilities:

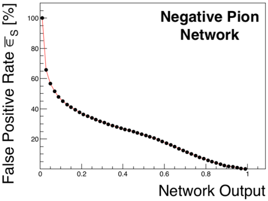
1. $P_e = \dfrac{\boxed{0.5 \times \epsilon_S}}{\boxed{0.5 \times \epsilon_S} + \boxed{0.5 \times \bar{\epsilon}_S}}$
2. $P_{\bar{e}} = \dfrac{\boxed{0.5 \times (1 - \bar{\epsilon}_S)}}{\boxed{0.5 \times (1 - \bar{\epsilon}_S)} + \boxed{0.5 \times (1 - \epsilon_S)}}$

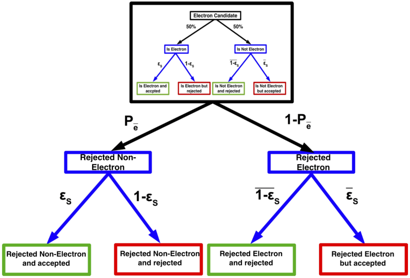
Using the Classifier-Output *

* Inspired by PhD-Thesis from Daniel Coderre (2012)





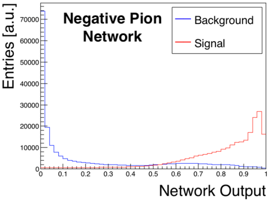


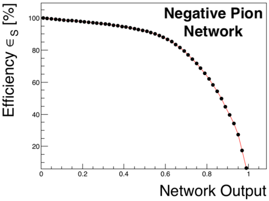


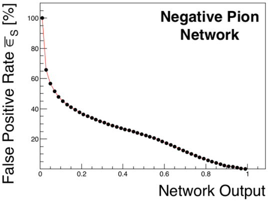
- Instead of network output, use ROC (i.e. efficiency, false identification rate) for classification
- Calculate two probabilities:
  1. $P_{\bar{e}} = f(\text{Electron Network})$
  2. $P_\pi = \dfrac{\boxed{P_{\bar{e}} \times \epsilon_S}}{\boxed{P_{\bar{e}} \times \epsilon_S} + \boxed{(1 - P_{\bar{e}}) \times \bar{\epsilon}_S}}$
- $P_{\bar{e}}$ serves as prior probability for the pion-hypothesis

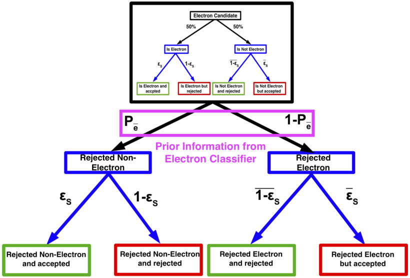
Using the Classifier-Output *

* Inspired by PhD-Thesis from Daniel Coderre (2012)





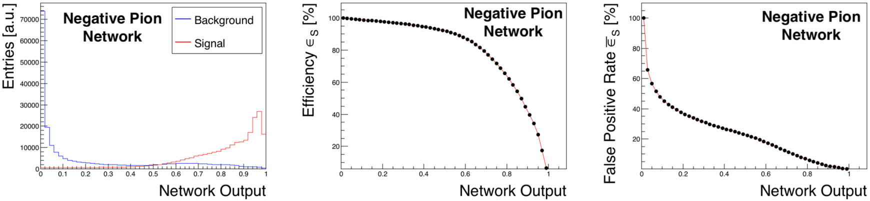


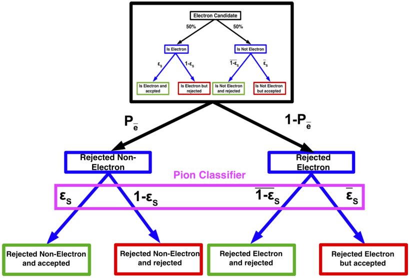


- Instead of network output, use ROC (i.e. efficiency, false identification rate) for classification
- Calculate two probabilities:
  1. $P_{\bar{e}} = f(\text{Electron Network})$
  2. $P_\pi = \dfrac{\boxed{P_{\bar{e}} \times \epsilon_S}}{\boxed{P_{\bar{e}} \times \epsilon_S} + \boxed{(1 - P_{\bar{e}}) \times \bar{\epsilon}_S}}$
- $P_{\bar{e}}$ serves as prior probability for the pion-hypothesis

Using the Classifier-Output *

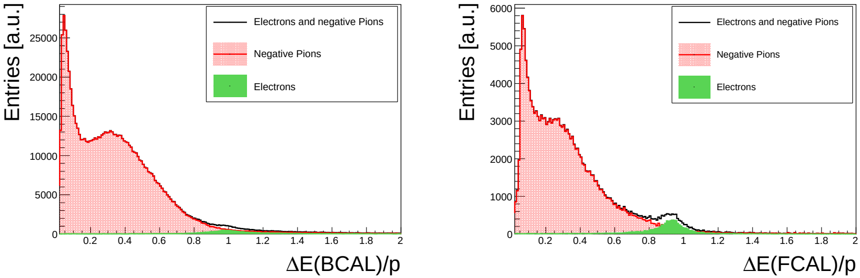
* Inspired by PhD-Thesis from Daniel Coderre (2012)





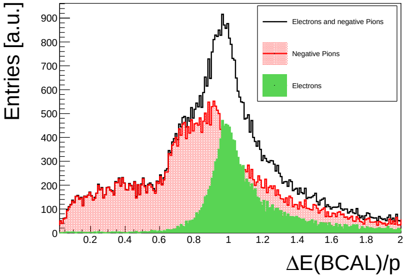
- Instead of network output, use ROC (i.e. efficiency, false identification rate) for classification
- Calculate two probabilities:
  1. $P_{\bar{e}} = f(\text{Electron Network})$
  2. $P_\pi = \dfrac{\boxed{P_{\bar{e}} \times \epsilon_S}}{\boxed{P_{\bar{e}} \times \epsilon_S} + \boxed{(1 - P_{\bar{e}}) \times \bar{\epsilon}_S}}$
- $P_{\bar{e}}$ serves as prior probability for the pion-hypothesis

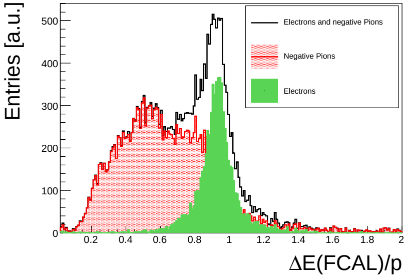
Application and Validation on Simulated single Particles



- Look at data set with simulated single particle tracks and: $N(\pi^{\pm}) \approx 20N(e^{\pm})$
- Top figures: No particle identification applied
- Signature of electrons: $\Delta E/p \approx 1$

Application and Validation on Simulated single Particles



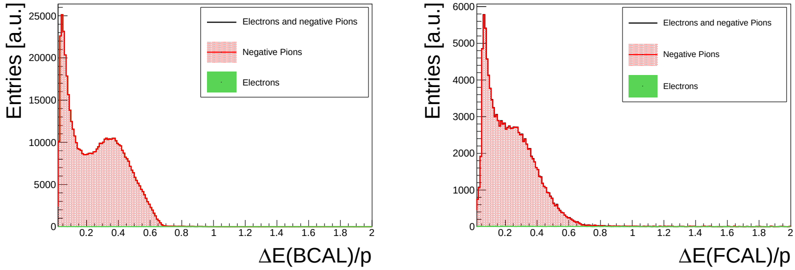


- Look at data set with simulated single particle tracks and: $N(\pi^{\pm}) \approx 20N(e^{\pm})$
- Top figures: Particle identification applied with: $P_e > P_\pi$

Particle	Acceptance BCAL [%]	Acceptance FCAL [%]
Electron	75	83
Pion	5	12

$\Rightarrow$ $\sim 80\%$ **of Electrons accepted and** $\sim 90\%$ **of Pions rejected**

Application and Validation on Simulated single Particles



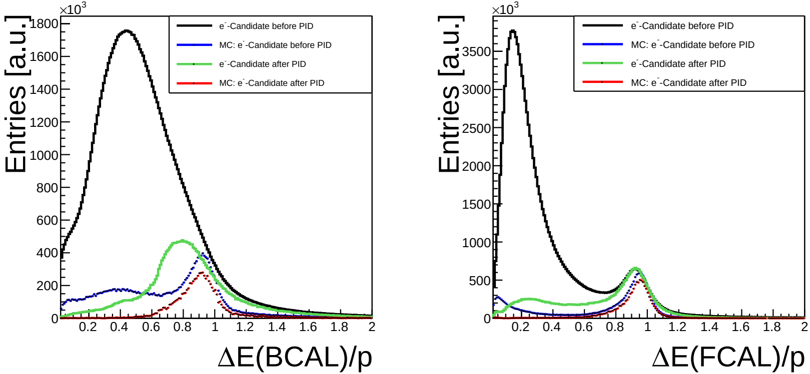
- Look at data set with simulated single particle tracks and: $N(\pi^{\pm}) \approx 20N(e^{\pm})$
- Top figures: Particle identification applied with: $P_{\pi} > P_{e}$

Particle	Acceptance BCAL [%]	Acceptance FCAL [%]
Electron	3	8
Pion	68	77

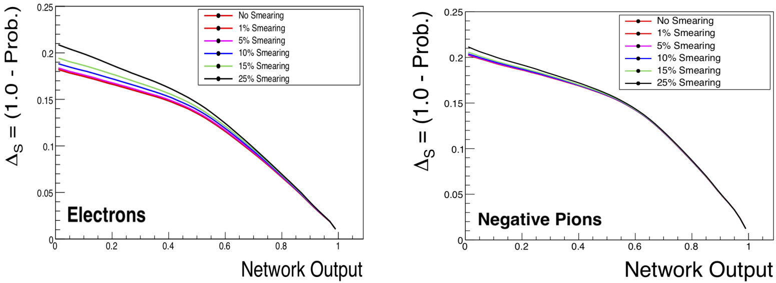
$\Rightarrow$ $\sim 90\%$ **of Electrons rejected and** $\sim 70 - 80\%$ **of Pions accepted**

Application and Validation in Analysis of $\eta' \to \pi^+\pi^-e^+e^-$ with GlueX-Data

- Apply method on GlueX data from run period 2017 (2018 to come)
- Significant reduction of pion-background, but still noticeable contribution
  $\Rightarrow$ Room for improvement
- Promising response for FCAL in measured and simulated data



First Checks on Reliability and Stability

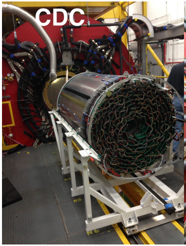


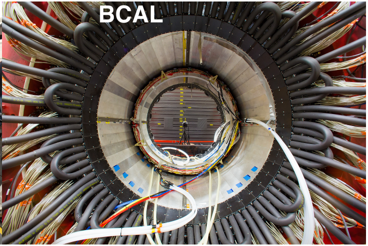
- Smeared variables in the test data set randomly with a gaussian function: variable $\mapsto$ variable $\times$ Gauss$(1, \delta)$
- Used relative smearing $\delta = 1\%$, 5%, 10%, 15% and 20%:

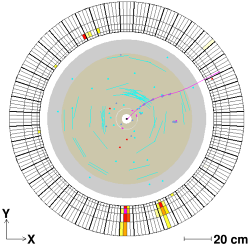
Particle	$> 5\%$ effect on Δ_S	Effect on Δ_S for $\delta = 25\%$
$e^{\pm}$	$\delta \gtrsim 15\%$	14%
$\pi^{\pm}$	$\delta \gtrsim 25\%$	8%

- Ongoing test: Apply method on "clean" channel: $\rho \to \pi^+\pi^-$ and compare response between data and simulations

Where to go from here (?)







- Used machine learning to reproduce detector response for electrons and pions:
  - i) Combine information from different detector sub-systems
  - ii) Take angular dependency into account
  - ⇒ Preliminary stage for particle track reconstruction (aka tracking)
- Particle Track Reconstruction: Go one step down in information hierarchy
  - Momentum $p \Leftrightarrow$ Helix defined by hits in Drift Chamber
  - Energy deposits ⇔ Group/Cluster of hits in Calorimeter

⇒ **Extend usage of machine learning towards track reconstruction**

Summary and Outlook

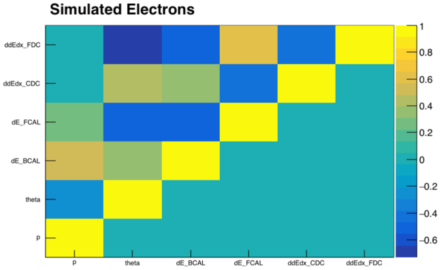
- **Application of machine learning based algorithms for particle identification**
  - ☑ Classification between electrons and pions with neural networks and boosted decision tree (latter one not shown today)
    $\Leftrightarrow$ developed/tested on decay $\eta^{(\prime)} \rightarrow \pi^+\pi^-e^+e^-$
  - ☑ First reliability and stability checks
  - ☐ Detailed comparison between measured and simulated data (ongoing)
  - ☐ Classification between kaons, pions and protons (ongoing)
    $\Leftrightarrow$ Include knowledge from $e^{\pm}/\pi^{\pm}$- classification
  - ☐ Include other algorithms for comparison (SVM, Likelihood,...)
  - ☑ Use neural networks to identify properly reconstructed photons and reject falsely reconstructed ones (not discussed today)
- **Explore possible applications for machine learning**
  - ☐ Reconstruction of particle tracks (tracking)
  - ☐ High level physics analysis
    (Final state selection, partial wave analysis,...)

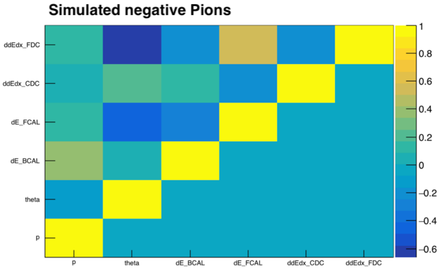
Content

Backup Stuff

- Contents
- Variable Correlations
- Training, Validation and Testing
- Current Model Selection
- Comparison to Decision Tree (GBT)
- The big Picture
- The Purity
- Combinatorics in Analysis of $\eta^{(\prime)} \rightarrow \pi^+\pi^- e^+ e^-$
- Details on $\eta^{(\prime)} \rightarrow \pi^+\pi^- e^+ e^-$

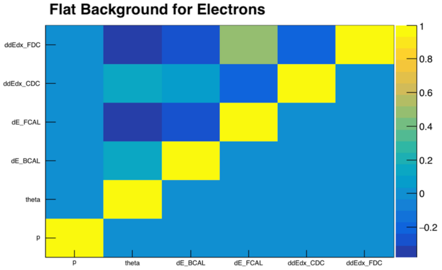
Backup: Variable Correlations

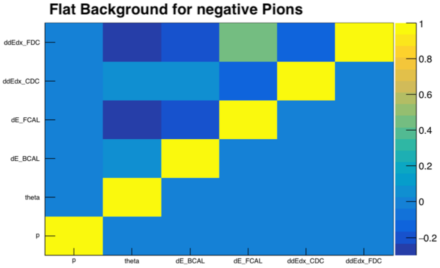




- Shown here is the Pearson Correlation Coefficient between the classification variables
- Correlation in flat background data due to detector geometry

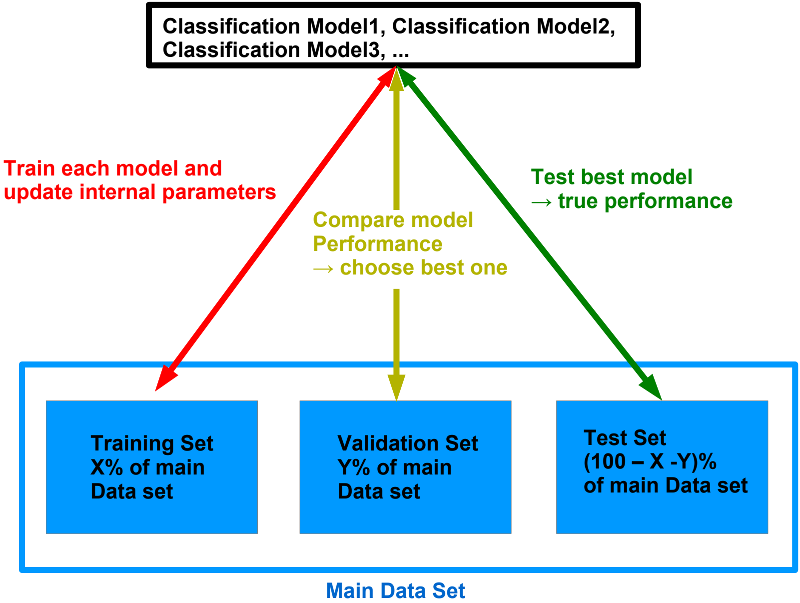
Backup: Variable Correlations



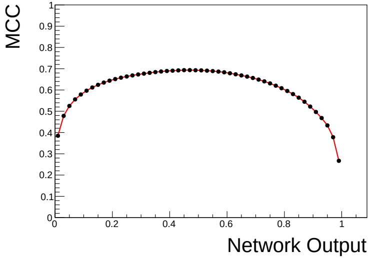


- Shown here is the Pearson Correlation Coefficient between the classification variables
- Correlation in flat background data due to detector geometry

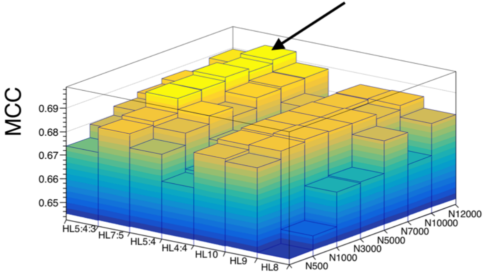
Backup: Training, Validation and Testing

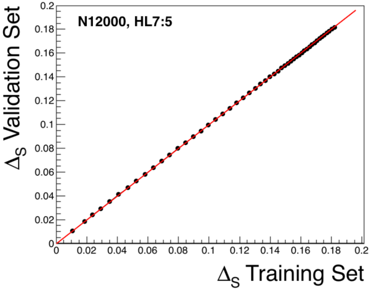


Backup: Current Model Selection

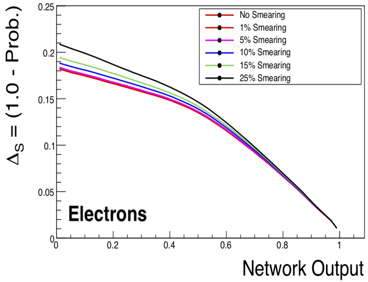


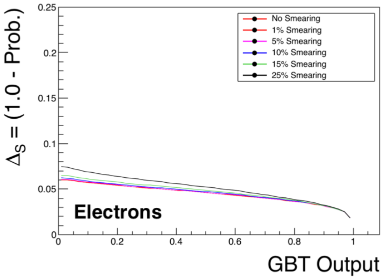
- Use Mathews Correlation Coefficient:
  $MCC \equiv \frac{\epsilon_S \times \epsilon_B - FPR \times FNR}{\sqrt{\frac{\epsilon_S}{P_S} \times \frac{\epsilon_B}{P_B}}} \in [-1, 1]$
- Currently (not best practice): Take model with largest MCC on validation data set
- Need to consider "costs": Number of parameters (e.g. training epochs, hidden layers,...)
- Bayesian Optimizer for machine learning algorithms: Spearmint





Backup: Comparison to Decision Tree (GBT)





- Smeared variables in the test data set randomly with a gaussian function: variable $\mapsto$ variable $\times$ Gauss$(1, \delta)$
- Used relative smearing $\delta = 1\%$, 5%, 10%, 15% and 20%:

Classifier for $e^{\pm}$	$> 5\%$ effect on Δ_S	Effect on Δ_S for $\delta = 25\%$
Network	$\delta \gtrsim 10\%$	14%
GBT	$\delta \gtrsim 10\%$	25%

- GBT is still under investigation: $e^{\pm}$-acceptance $\sim 8\% - 20\%$

Backup: The big Picture

INPUT

- Decisive Power
- Additional preparation needed ?(e.g. normalisation)
- How strong correlated?
- Use measured data or MC?
- Generality?
- Impact on classifier performance?

→ Know detector
→ Calibration
→ Match between data/MC

CLASSIFIER

- Which type?
- How to train?
- Implementaion?
- Handling?
- Influence on systematics?
- Handling of unknown data?
- Reliability?

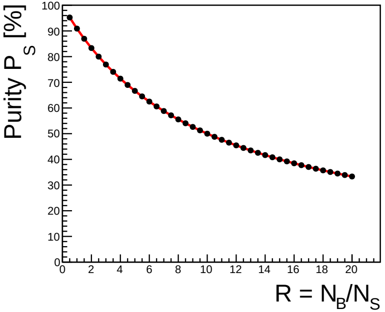
→ Training curve
→ ROC plot
→ Monitoring plots
→ Output variable
→ Use dedicated frameworks
→ Do not reinvent the wheel

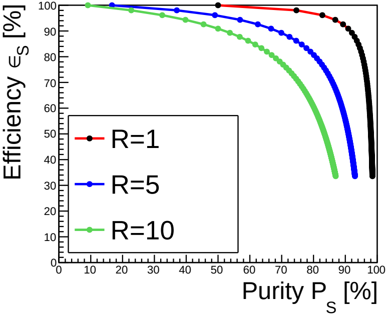
OUTPUT

- How used in further Analysis?
- Used at which analysis Stage?
- Assigned error?
- Trustworthy?
- Generality?

→ Systematic studies
→ Error handling

Backup: The Purity





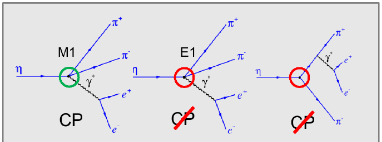
- Shown is the response of a neural network on a fake data set (not related to physics or anything else)
- Purity: $P_S = \left[1 + R \times \frac{\bar{\epsilon}_S}{\epsilon_S}\right]^{-1}$, with $R =$ ratio between number of background and signal events

Backup: Combinatorics in Analysis of $\eta^{(\prime)} \to \pi^+\pi^-e^+e^-$

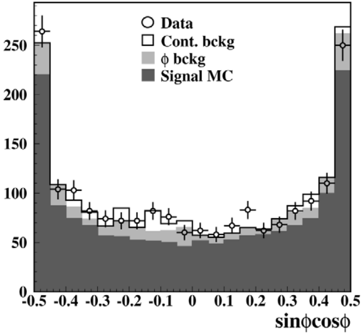
- To consider in data analysis: combinatorics
- Pass posterior probability form configuration i as prior probability to configuration i+1 (Adapted from PhD Thesis from Daniel Coderre (2012))
- Pick configuration with largest posterior probability

Configuration	Pos. Particle 1	Neg. Particle 1	Pos. Particle 2	Neg. Particle 2
1	π^+	π^-	e^+	e^-
2	π^+	e^-	e^+	π^-
3	e^+	e^-	π^+	π^-
4	e^+	π^-	π^+	e^-

Backup: Details on $\eta^{(\prime)} \to \pi^+\pi^- e^+e^-$



CP-allowed, and possible CP-forbidden amplitudes



- Underlying decay: $\eta^{(\prime)} \to \pi^+\pi^-\gamma$
- E_1-Transition of photon is CP-violating
  $\Leftrightarrow$ Need information about polarization of photon
  $\Leftrightarrow$ Experimental challenging
- Look at cases where: $\gamma^* \to e^+e^-$

- Determination of A_Φ via $\sin\Phi\cos\Phi$
- KLOE reconstructed: 1.6 k $\eta \to \pi^+\pi^- e^+e^-$ - events

KLOE coll. *Phys. Lett.*, B675:283-288,(2009)