Are Neural Networks Effective in Physics Data Analysis?

IPC Study Trip Presentation

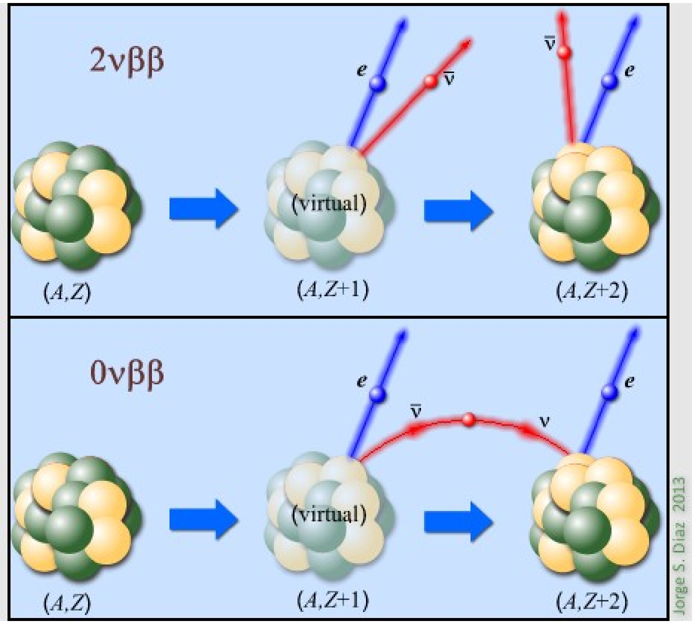
B. Temuge, CANDLES group
March 16, 2018

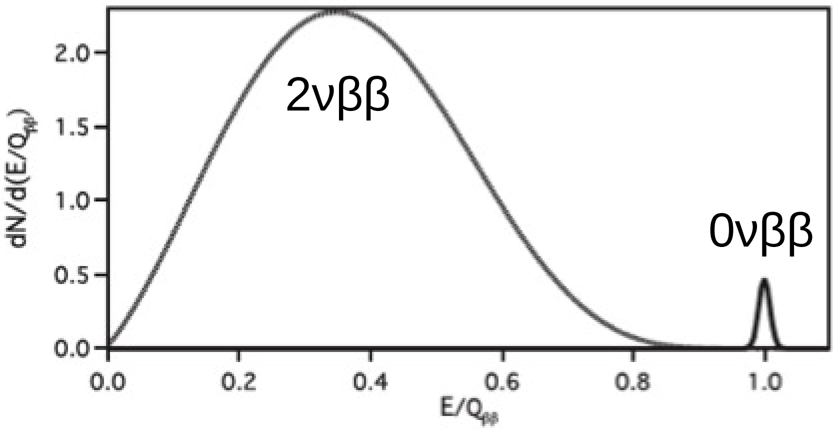
Contents

- Introduction to 0νββ decay
- Introduction to CANDLES experiment
- Background reduction of 208Tl – current study
- Preparation of Training examples
- *Pattern fitting method for background reduction - skipped*
- Introduction to neural nets
- Application of neural nets
- Discussion
- Conclusion

Introduction to 0νββ decay

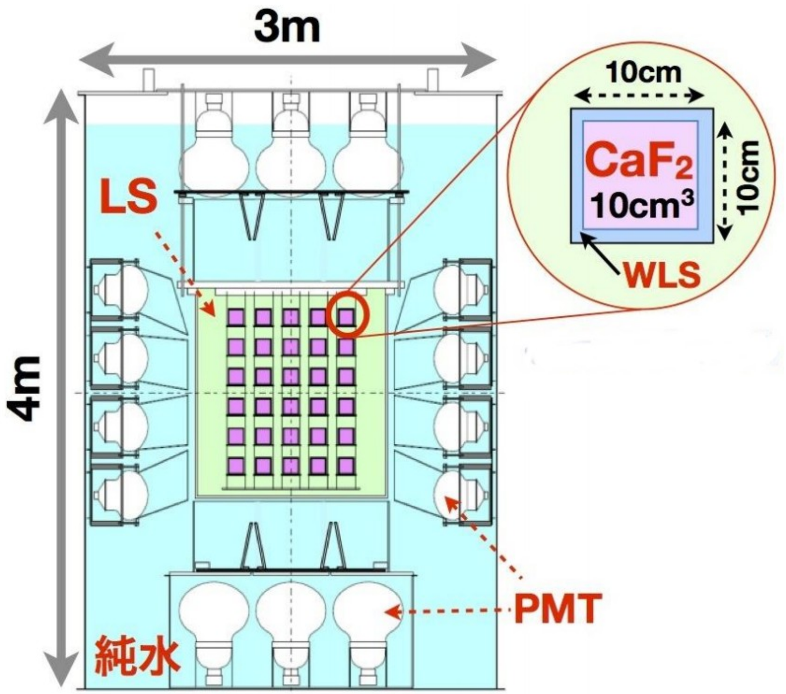
- A 2νββ decay occurs when a neutron decay from Z to (Z+1) nuclei is not energetically allowed but Z to (Z+2) is allowed. There are not many nuclei with this property. The 2νββ decay has been observed before.
- If the neutrino is a Majorana type particle, then the neutrino and anti-neutrino are the same particle and they can undergo pairwise annihilation. This process would be in violation of lepton number conservation. The remaining ββ values are expected to be observed with the full Q value.





CANDLES Experiment

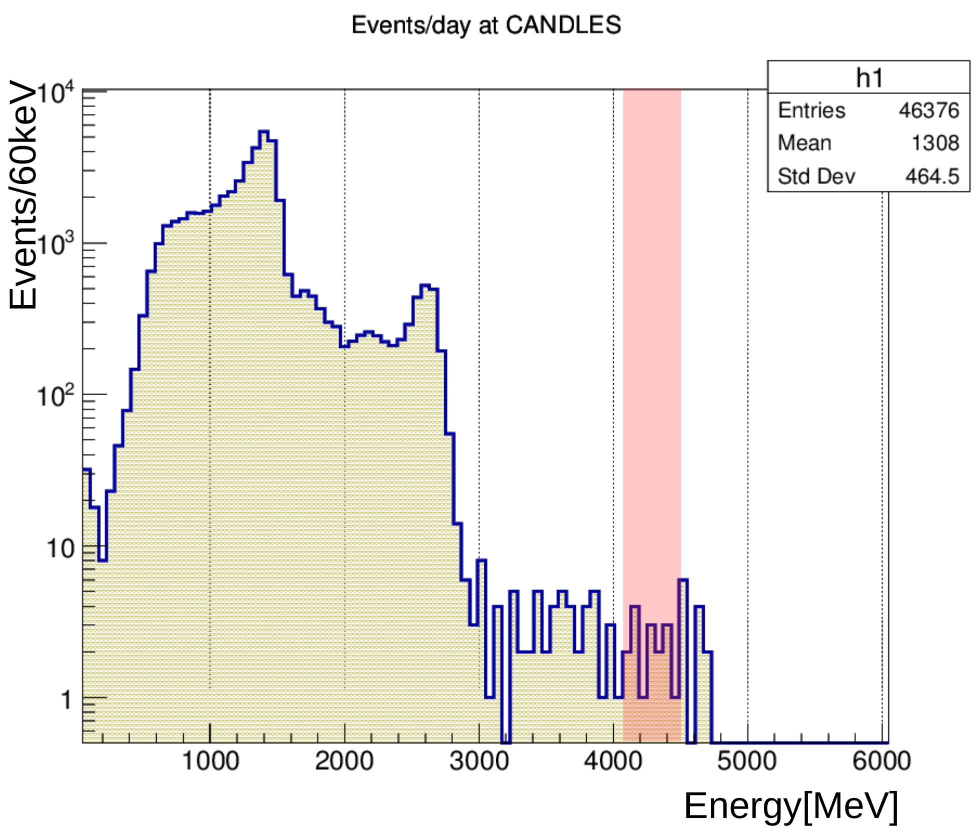
- CANDLES experiment is looking for the 0νββ using 96 CaF_2 scintillating crystals that act as source and detector. The crystals are scintillating and contain the 0νββ candidate 48Ca isotope.
- 48Ca is chosen because it has the highest Q value among candidate nuclei. This means detection is easier due to fewer number of low energy background.
- Detector is surrounded inside by 62 PMTs that detect the signal.

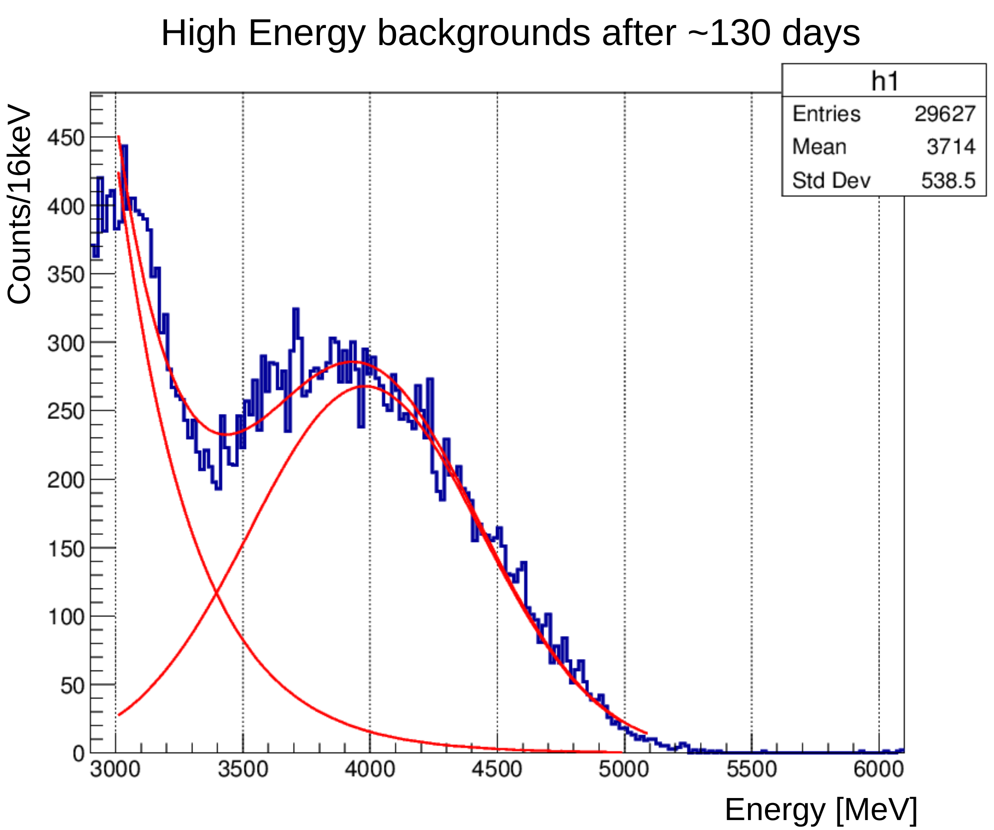


Nuclei	$Q_{\beta\beta}$ value [keV]	Natural Abundance [%]	2νββ half-life [year]	Experiment
^{48}Ca	4274	0.187	0.44 x E(20)	CANDLES
^{76}Ge	2039	7.8	15 x E(20)	GERDA, MAJORANA
^{82}Se	2996	9.2	0.92 x E(20)	SuperNEMO, LUCIFER
^{100}Mo	3035	9.6	0.07 x E(20)	NEMO-3, MOON, AMoRe
^{116}Cd	2809	7.6	0.29 x E(20)	COBRA
^{130}Te	2530	34.5	9.1 x E(20)	CUORE, SNO+
^{136}Xe	2458	8.9	21 x E(20)	KamLAND-Zen, EXO, NEXT
^{150}Nd	3367	5.6	0.08 x E(20)	SNO+, DCBA/MTD

Backgrounds at CANDLES

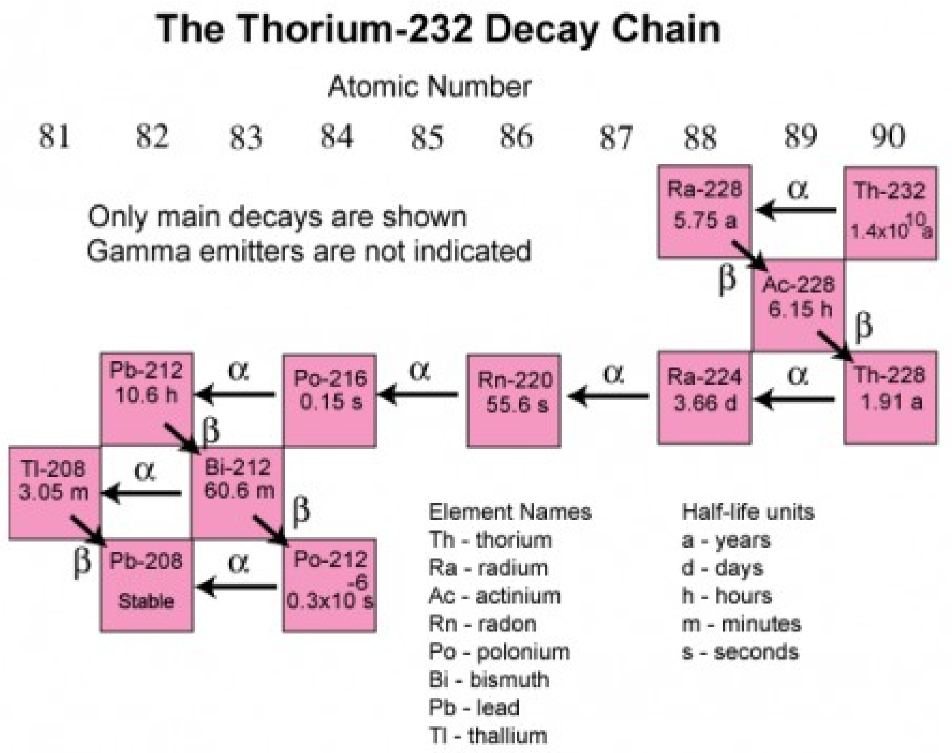
- The number of background events in the $Q_{\beta\beta}$ value region is not very many.
- Many of the backgrounds in the $Q_{\beta\beta}$ value region comes from 212Bi natural background.





Thorium C Background

- Thorium C is a historic name given to 212Bi decay. It can decay to 208Tl through α decay or to 212Po though β decay.
- Because half-life of 212Po is very short, 212Bi → 212Po → 208Pb decay is recorded as a single event by CANDLES
- 208Tl → 208Pb decay has a Q value of 5MeV. But, because the neutrino carries away some energy, detected energy in CANDLES is very close to $Q_{\beta\beta}$ of 48Ca.



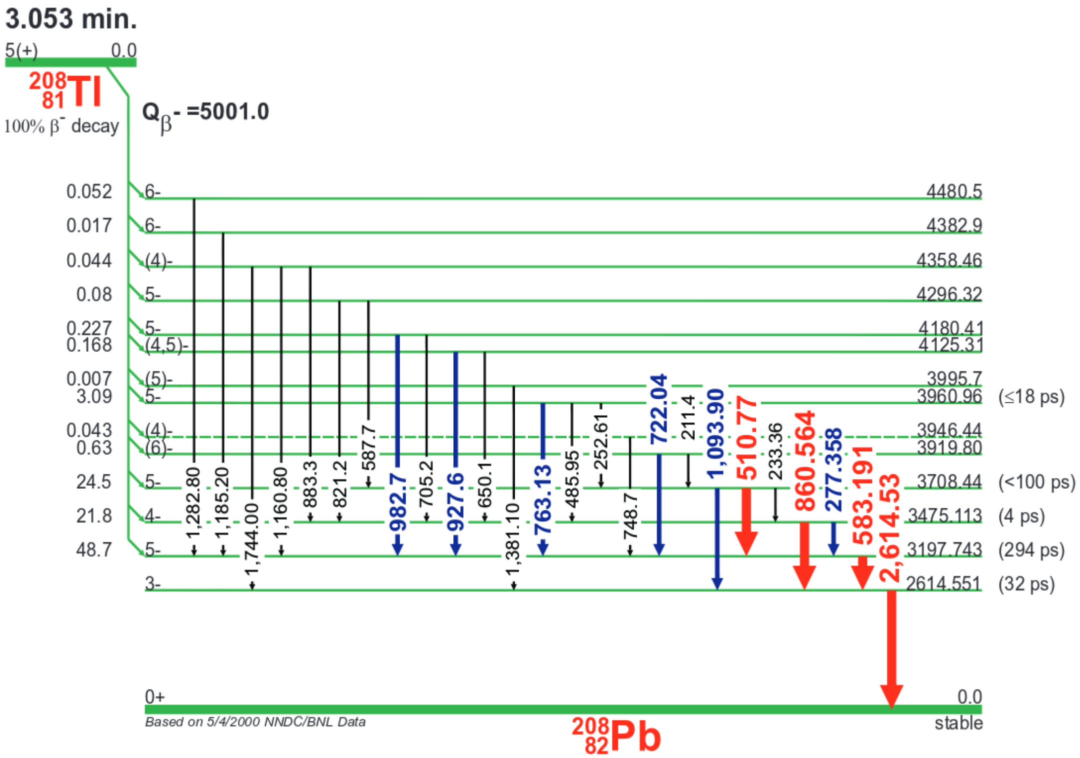
$$({}^{212}Bi) \xrightarrow{\beta} ({}^{212}Po) \xrightarrow{\alpha} ({}^{208}Pb) : \ 5MeV$$

$$({}^{208}Tl) \xrightarrow{\beta+\gamma} ({}^{208}Pb) : 5MeV$$

^{208}Tl Background

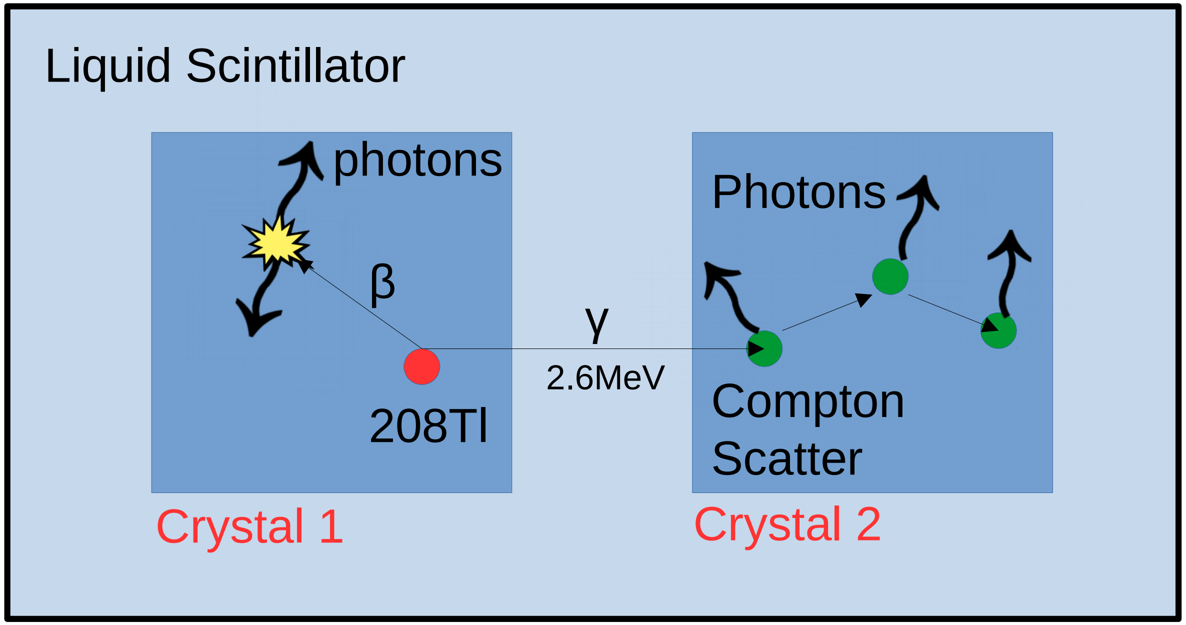
- ^{208}Tl is one of the major backgrounds in CANDLES.
- 208Tl rejection is based on the delayed coincidence of ^{212}Bi α event with a high energy event in the same crystal.
- My current research is focused on reduction of background 208Tl events from CANDLES data.

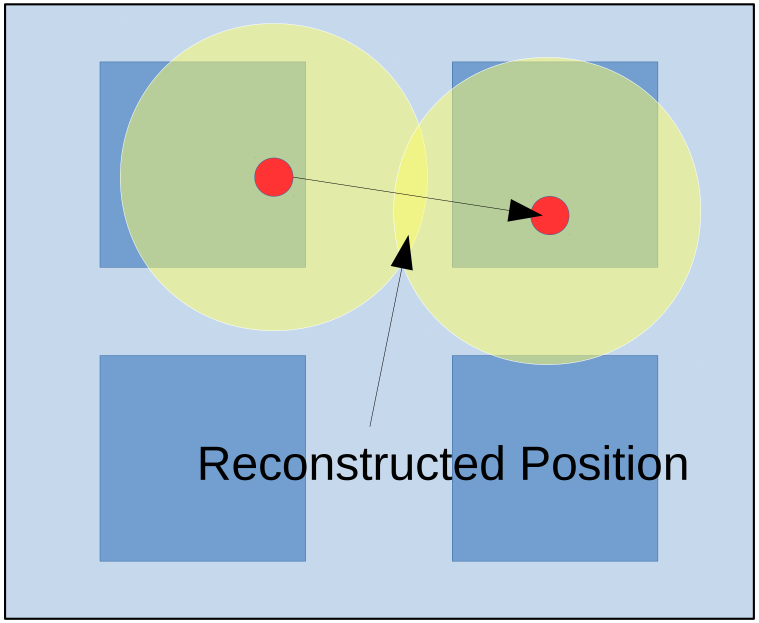
$$^{212}Bi \xrightarrow[\alpha]{} {}^{208}Tl \xrightarrow[\beta+\gamma]{} {}^{208}Pb$$



Multi-Crystal 208Tl Event

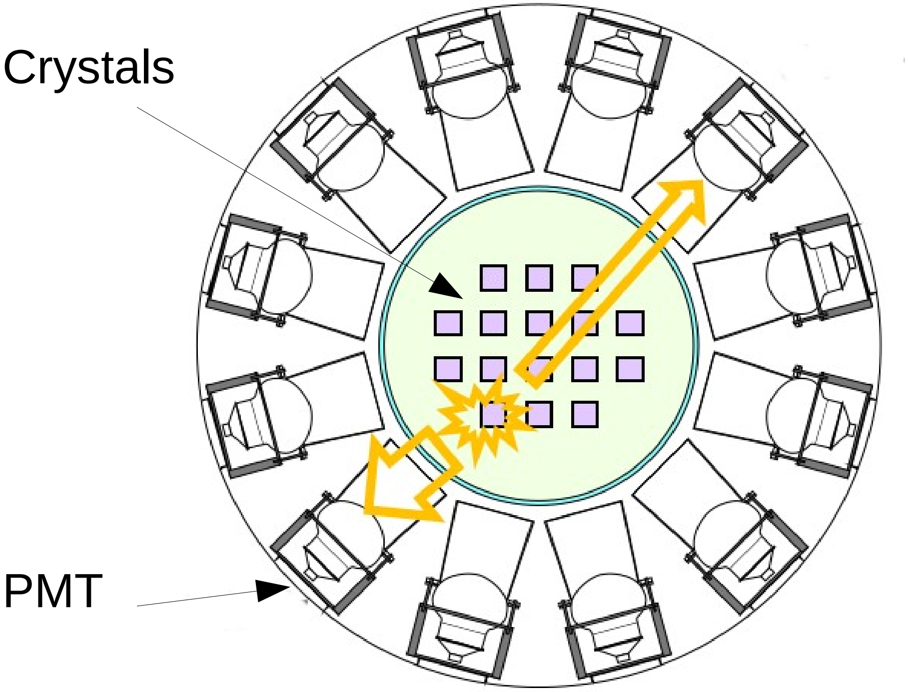
- ^{208}Tl β and γ decay could become a multi-crystal event. The current detection method is applied on the same crystal and may not detect multi-crystal events. We want to further reduce ^{208}Tl through multi-crystal event detection.
- Therefore, need to develop a method for detecting multi-crystal events.
  - Pattern Fitting Method
  - Neural Networks

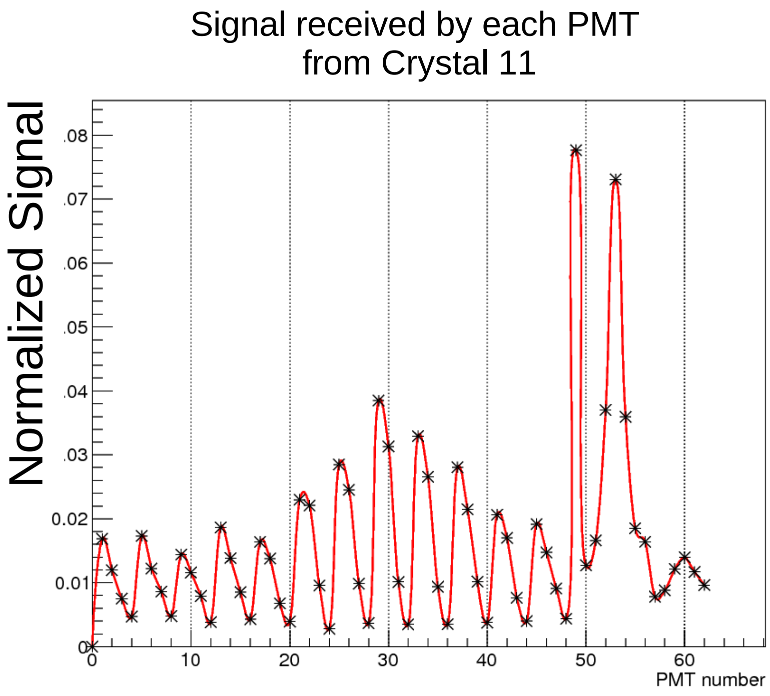




Making Training Examples

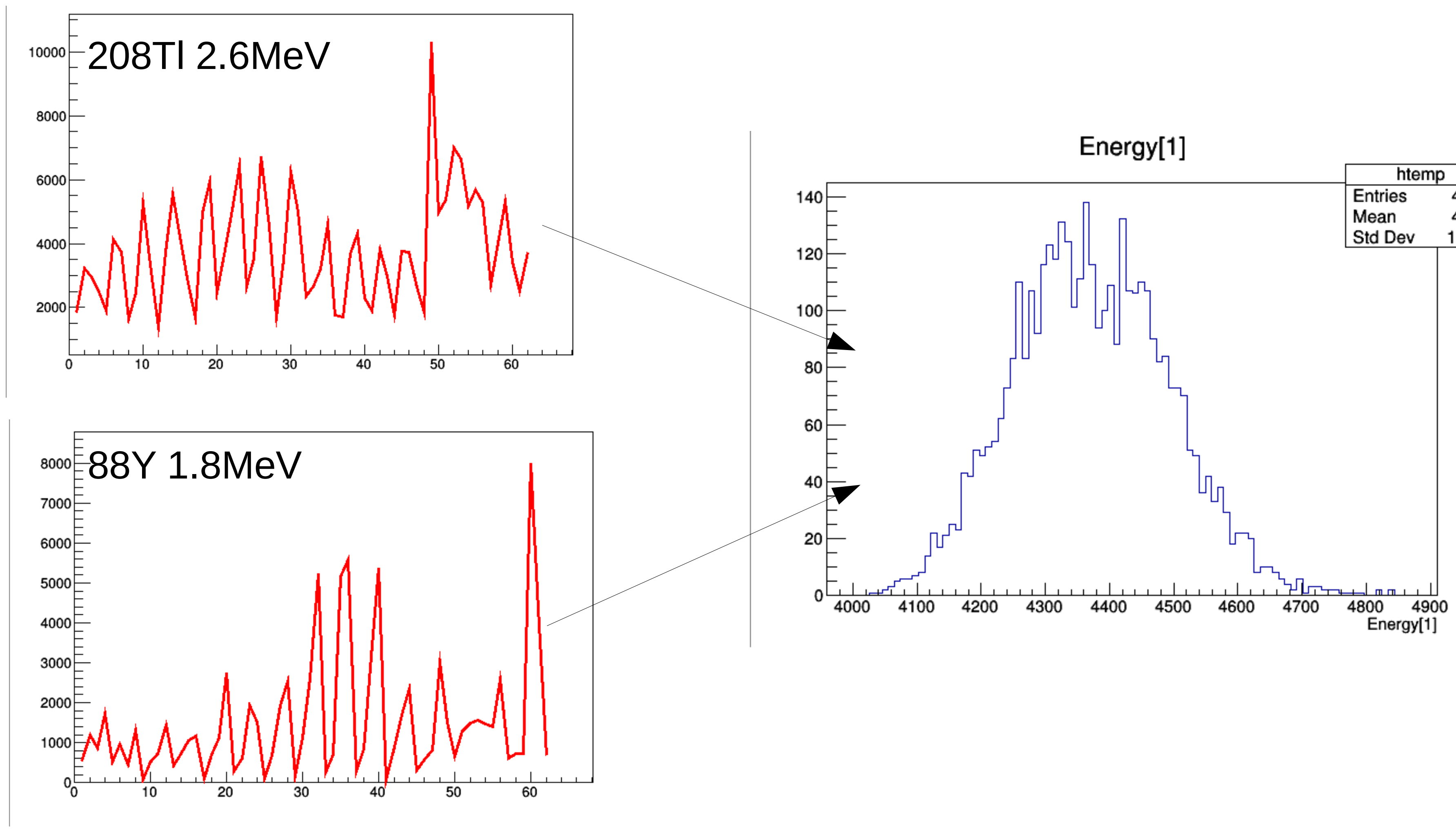
- Detection of multi-crystal events is hard because we don’t have real examples of such events. To study them, we have to make fake multi-crystal events.
- Each PMT receives a different ratio of photons from each crystal. By adding number of photons from on crystal to another crystal’s signal, we can make multi-crystal events.





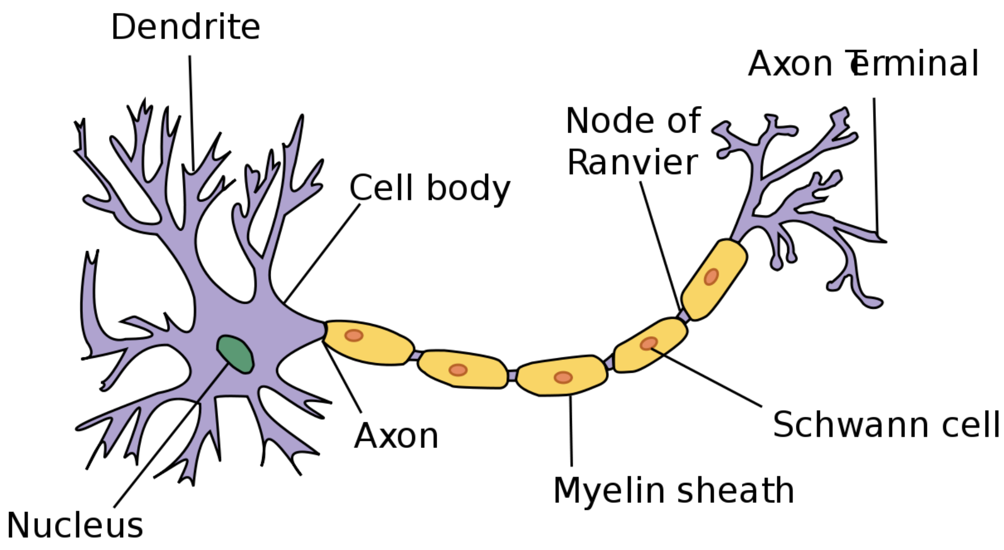
Making Training Examples

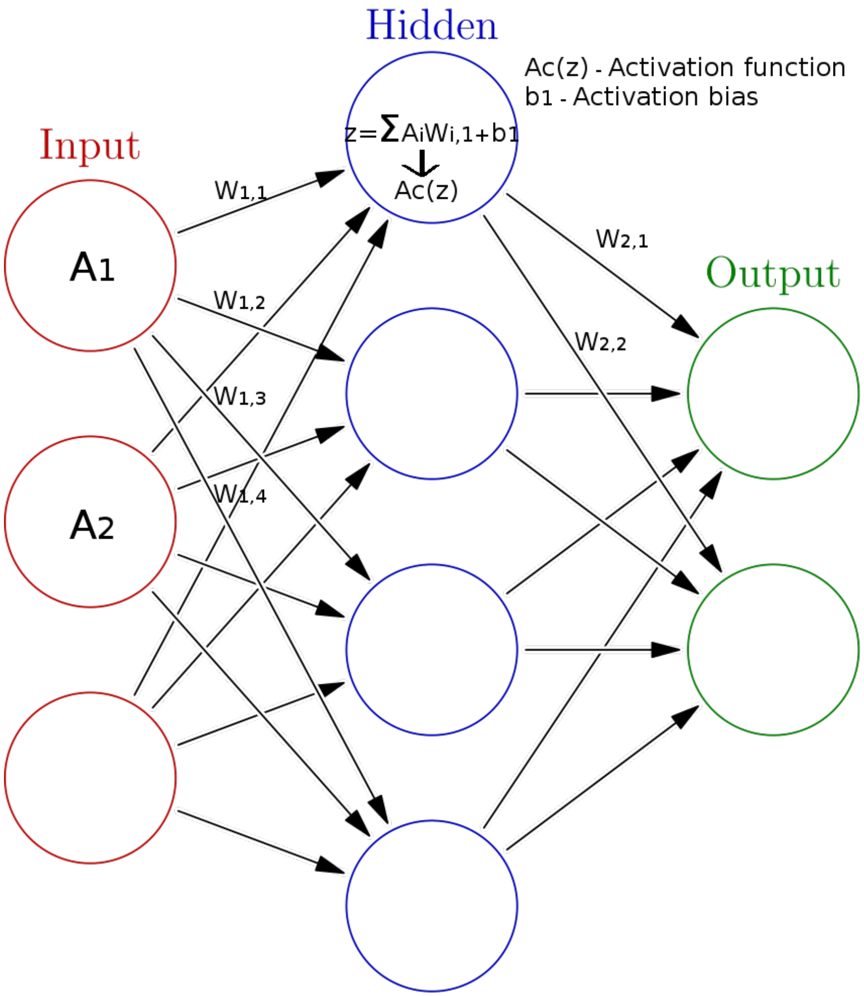
- Multi-crystal events are made by adding 2.6MeV gamma events to 1.8MeV gamma events. The output energy spectrum is shown below.
- 2000 Multi-crystal and 2000 single-crystal events were made.



Introduction to Neural Nets

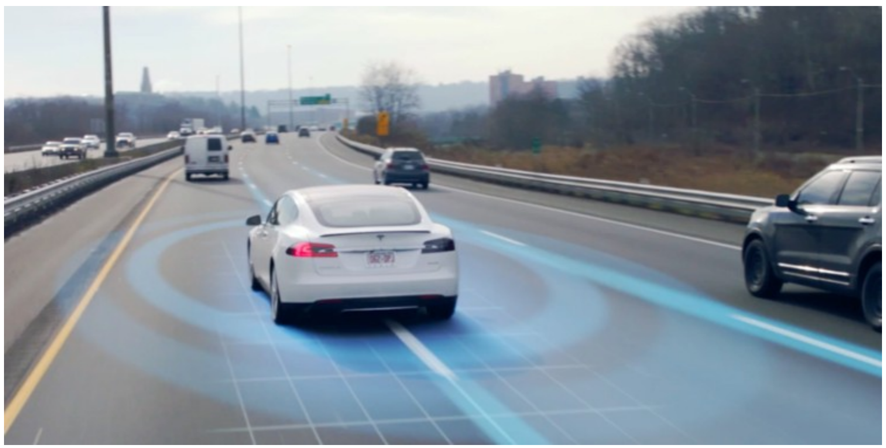
- An attempt to mimic human brain. Human brain has ~10^10 neurons
- Each node is activated when enough input signal is accumulated





Examples of Neural Nets

- Neural nets research is developing in an extremely fast rate. They are used in many fields such as computer science, engineering, economics etc. They are able to do things that humans can't do.



Self driving cars

Self landing rockets

Colorado National Park, 1941 | Textile Mill, June 1937 | Berry Field, June 1909 | Hamilton, 1936

Coloring pictures

Real time translation

Training Neural Nets

- In supervised training, input vector and desired output vector is known: $(\overline{x}_i, \overline{y}_i)$, where i = 1,…, N, are number of training samples
- An overall error function is defined on all the training examples:

$$E(W, x, y) = \frac{1}{2N} \sum_{i}^{N} \|y(W, \overline{x_i}) - \overline{y}_i\|^2$$

$y(W,\overline{x}_i)$ – output from the neural network

W – weights and bias matrix

- The neural net is trained by minimizing this function for all training examples with respect to the weights. The weights are modified by gradient descent:

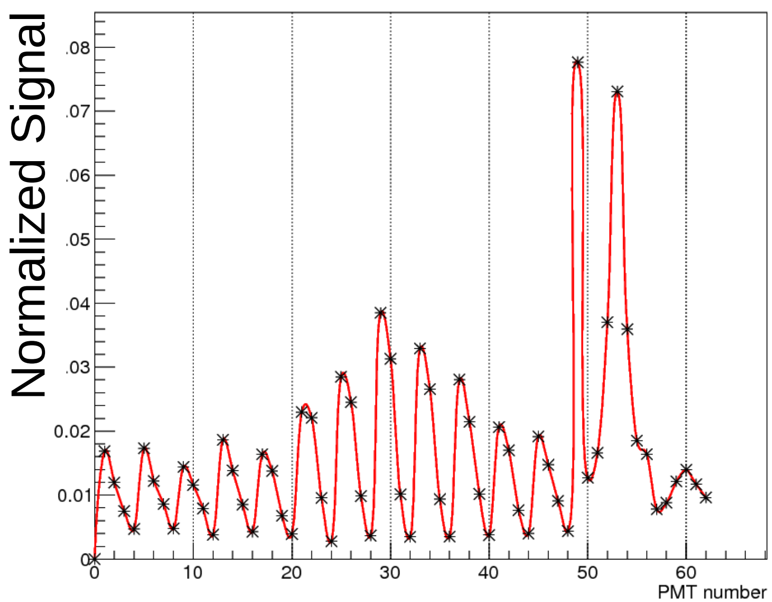
$$w_{i+1} = w_i - \alpha \frac{\partial E}{\partial w_i}$$

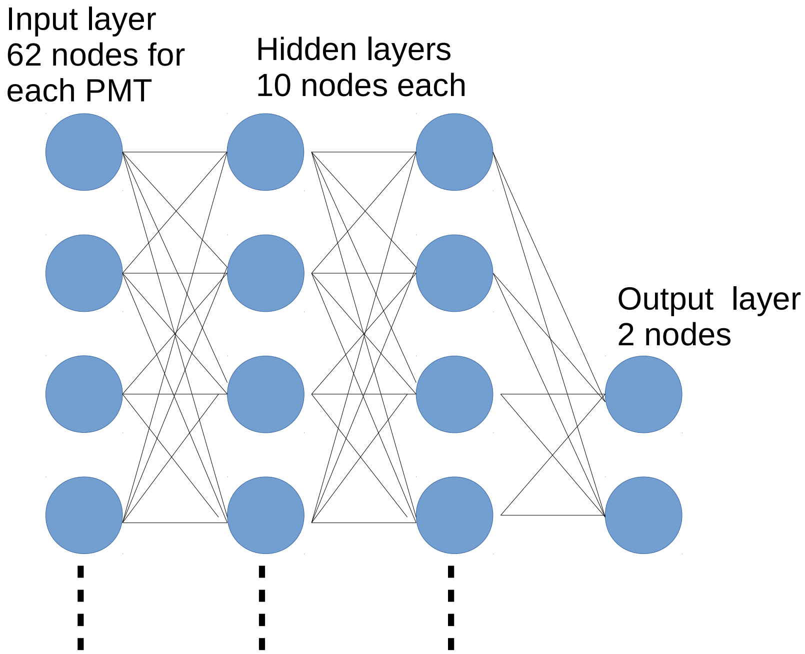
- Mathematically, neural nets are simple. The gradient of E can be written down exactly. But minimizing analytically is hard because there are too many simultaneous equations to solve, for each w_i in each layer.

Applying Neural Nets

- 2000 single-hit and 2000 multi-hit events were made. Used 1750 events from each as training sample and use remaining 250 as testing sample. 3500 training examples total and 500 testing examples total.
- The samples have the form: (x_i, y_i), where x_i is a 62 element vector representing signal for each PMT. The y_i vector is a 2 element vector that has only two possible values: $y_i = (1, 0)$ if single-crystal and $y_i = (0, 1)$ if multi-crystal event.
- Each weight is updated 1000 times per sample. 2.5 * 10^9 weight adjustments.
- Trained this way, the network has 100% accuracy to separate single and multi-hit events.

An example event





Discussion on Neural Nets

- Using traditional analysis techniques, the best accuracy I was able to achieve on the same training examples was 92.5%. But Neural nets were able to achieve 100% accuracy. But how? What about uncertainty?
- Does neural nets discover hidden features that we missed in analysis? Or does it only focus on features that are important and discard other features?
- Analysis of data is hard:
  - Collect data → Select features → Fit functions → Error analysis → X^2 minimization → classify/reduce backgrounds
- Neural Nets are easy:
  - Collect data → Feed data to Neural Net → Obtain better classifier/background reducer ?
- Issue #1: Neural nets does not concern itself with uncertainties. Any uncertainty in the input data disappears when input into the neural network. The output of the neural net is given a probabilistic interpretation even though there is no valid reason to do so. As a result, uncertainties cannot be assigned to neural nets outputs.
- Issue #2: Neural nets are trained on specific examples. There is no guarantee that the trained neural nets can be applied to new data. Because we don't know which features the neural nets use from training data, there is no guarantee that it can be applied to data collected in a similar but not exactly the same manner.

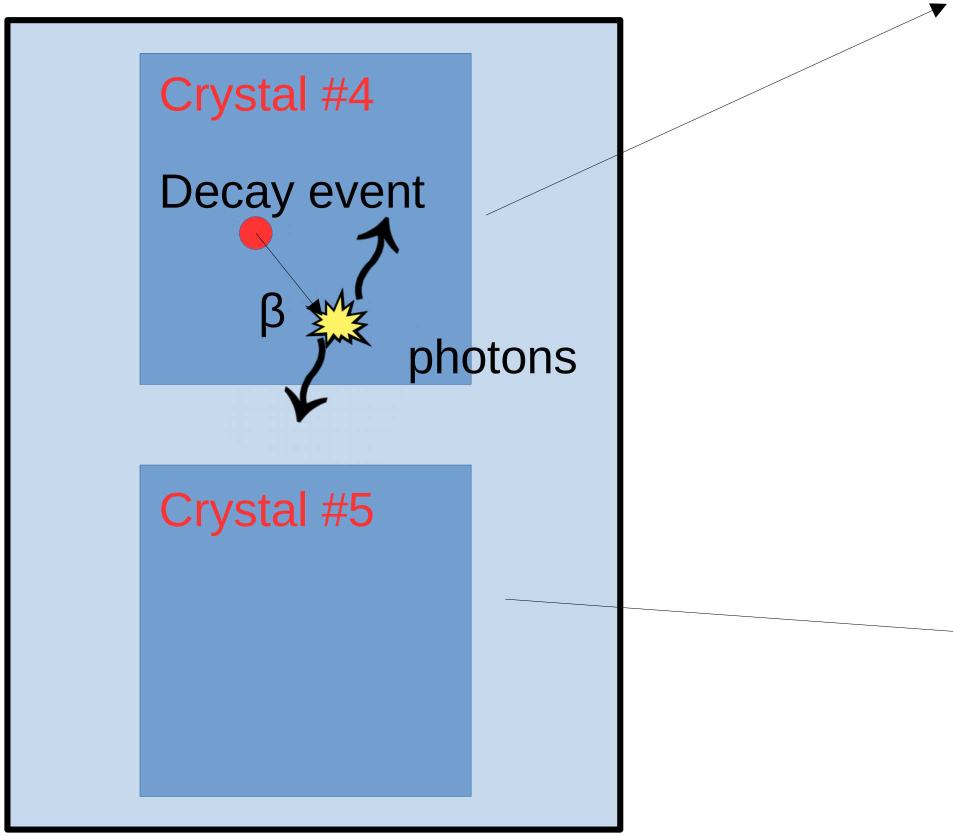
Conclusion

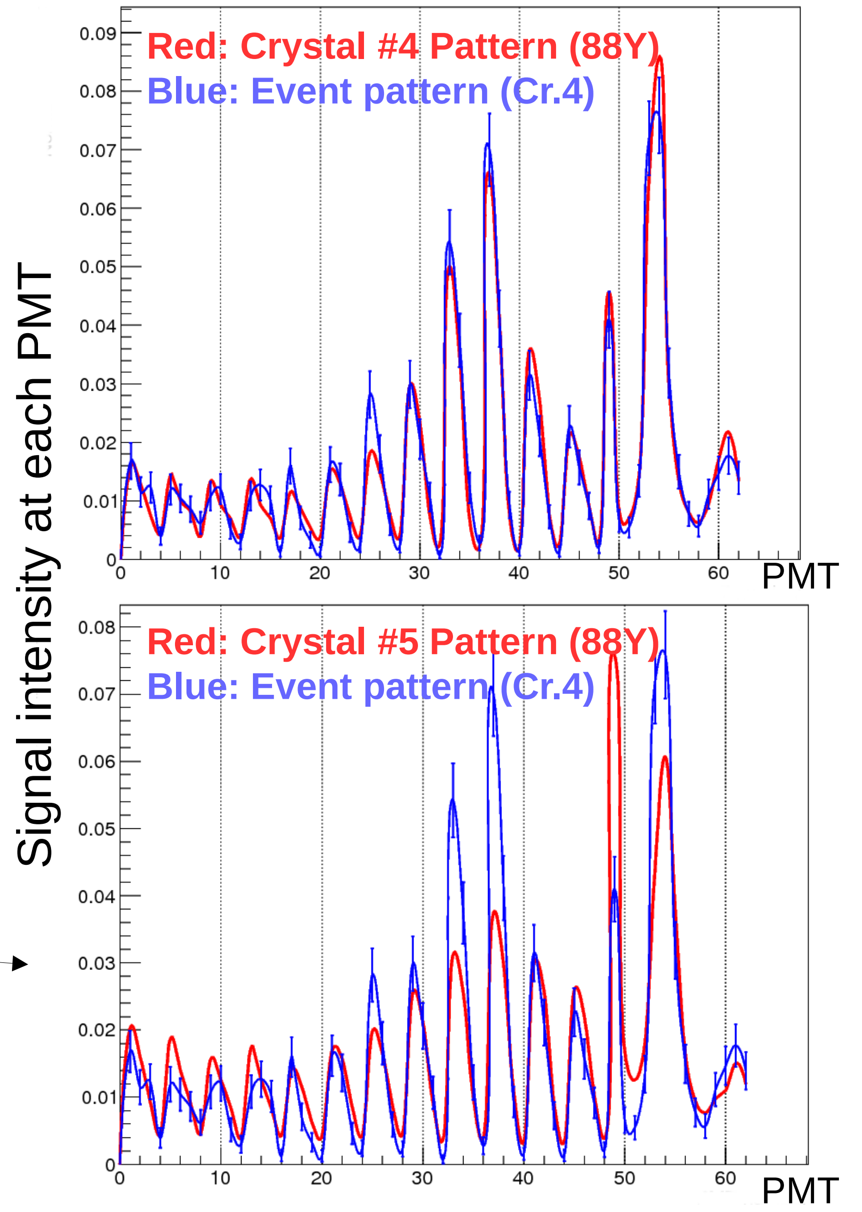
- CANDLES experiment is looking for the 0νββ decay using 48Ca isotope.
- A major source of background comes from the decay of 212Bi. The observed energy of the 212Bi is very close to the $Q_{\beta\beta}$ value of 48Ca. Therefore, significant effort is made to reduce this background.
- An application of a simple neural network was presented to classify single-crystal and multi-crystal 208Tl events. Neural networks provide 100% accuracy to classify while analysis techniques provide 92.5% accuracy on the same data set.
- Application of neural nets are easy but there are some drawbacks. For example, uncertainties in data are not propagated through neural nets. Also, underlying physics processes are not revealed through neural nets.
-

Backups

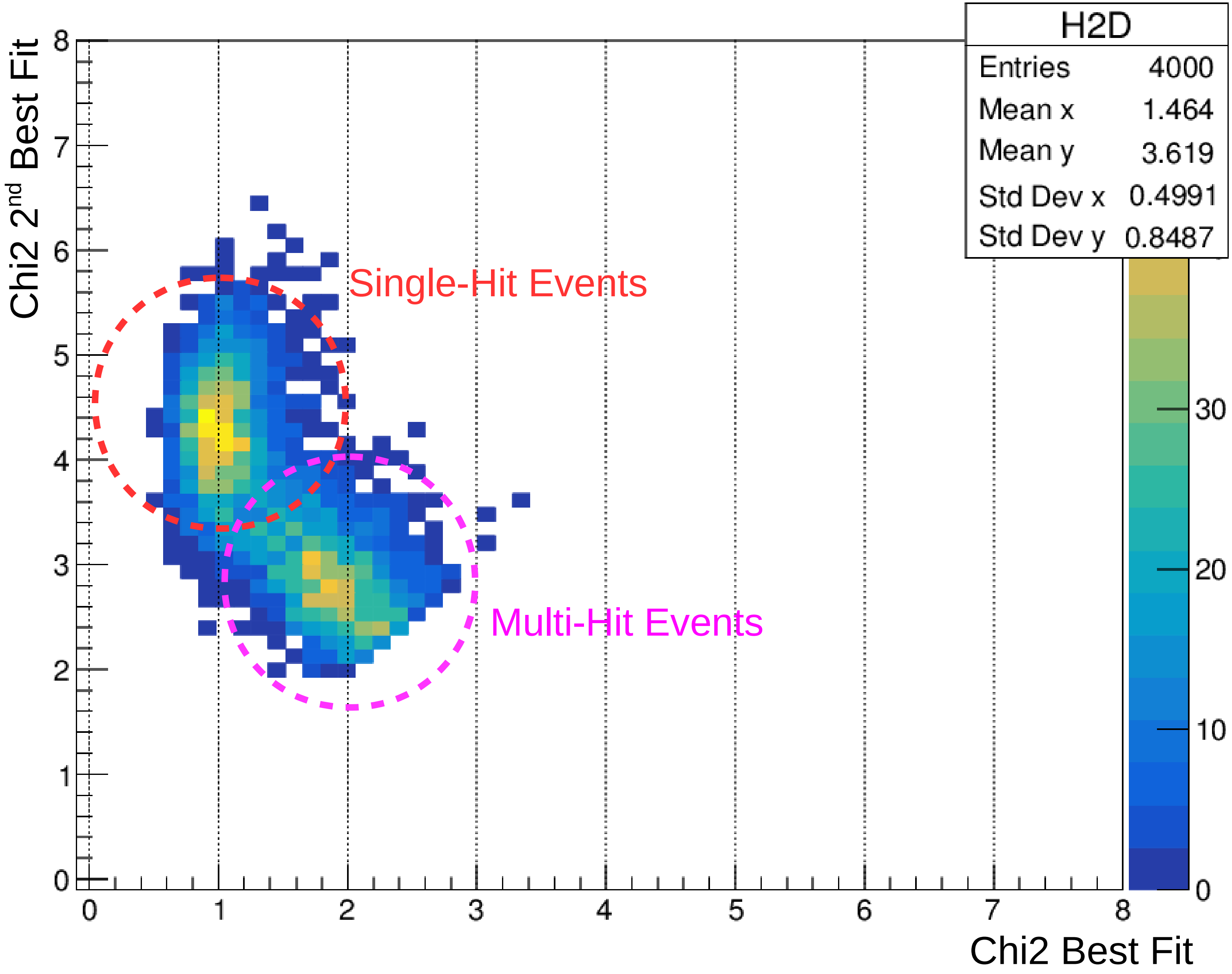
Pattern Fitting Method

For a single-crystal event, signal is received from only 1 crystal. In this example, beta decay in crystal #4. The event pattern matches crystal #4 pattern, obtained from 88Y. Not good match with crystal #5 pattern.





- Apply the hit patterns to fake events
  - Separation of multi-hit and single-hit events is possible



- Separation accuracy
  - Apply a cut based on the Chi2 difference
  - Accuracy is defined as $(N_{correct})/(N_{total})$
  - Best Accuracy : 92.4%