Direct energy estimation for waveforms produced by the EXO-200 detector

Michael Jewell, Alexander Rider

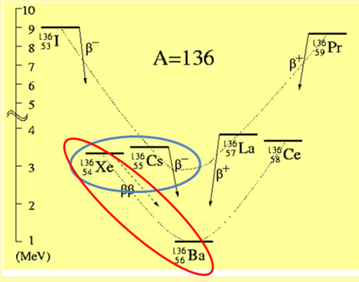
Department of Physics, Stanford University

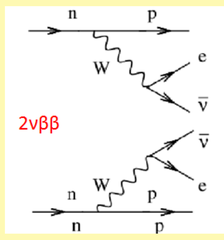
Abstract

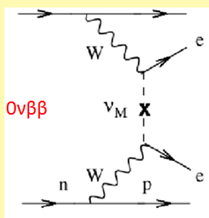
We report on a new analysis technique for estimating the amount of charge associated with waveforms produced by a simulation of the EXO-200 detector. We use a set of filters to compress the number of parameters associated with each waveform from 2048 to 10. The filter templates are generated from clustering raw waveforms. A boosted regression tree is then used to estiamte the mapping from this reduced parameter space to the energy associated with each waveform.

The EXO-200 detector

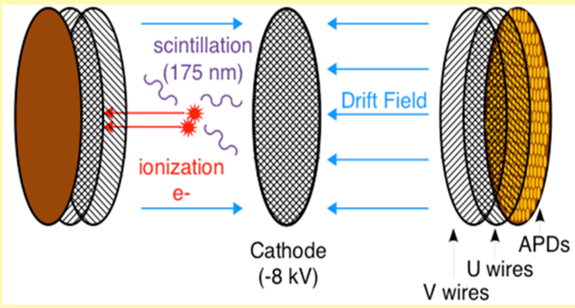
The EXO-200 detector is a time projection chamber (TPC) searching for neutrinoless double beta decay [1]. If neutrinoless double beta decay is observed it would indicate that the neutrino is its own anitparticle [2]. This would be clear evidence of letpon number violation, the Majorana nature of the nuetrino, and would indicate the neutrino mass.





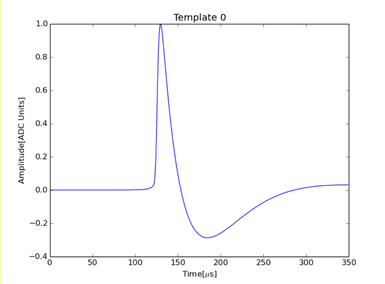


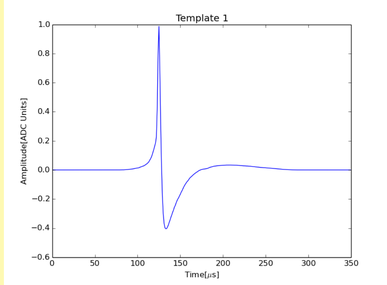
Events deposit energy in the liquid xenon through both scintillation light and free ionization charge. Scintillation is detected on either side by APDs(avalanche photodiodes). Charge deposited in the detector is drifted to either end where it is collected by the U-wires. Drifted charge produces signal on electrical channels by inducing current while drifting and collecting onto channels. The shape of the signals from collected charge is different from the shape of siganls due to induced current. The goal of this analysis is to develop a technique for determing the amount of charge collected on a channel that is robust against contamination from induced signals.



Pulse shape templates

The first step in our analysis is to determine a set of pulses that are representitive of as many pulses produced by EXO-200 as possible. To do so, we employ k-means clustering on the raw waveforms in the training set. The two most significant templates are displayed below. To process the data we used 10 such templates.





Optimal filters

The cluster centers from k-means clustering were used as templates for optimal filters. The cluster centers were othonormalized using the Gram-Schmidt process. Optimal filtering is equavalent to LMS fitting to the templates in the frequency domain. For each template, s, and each waveform, we minimize

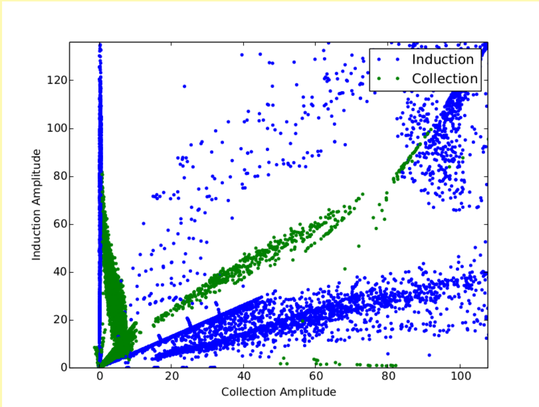
$$\chi^2 = \int df \frac{|\tilde{v}(f) - A\tilde{s}(f)|^2}{J(f)}$$

References

[1] J B. Albert et al. Phys. Rev. C, 89, 015502 (2014)
[2] M Auger et al. JINST 7 P05010 (2012)
[3] J.B. Albert et al. arXiv:1402.6956(2014)
[4] F. Pedregosa, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825-2830, 2011

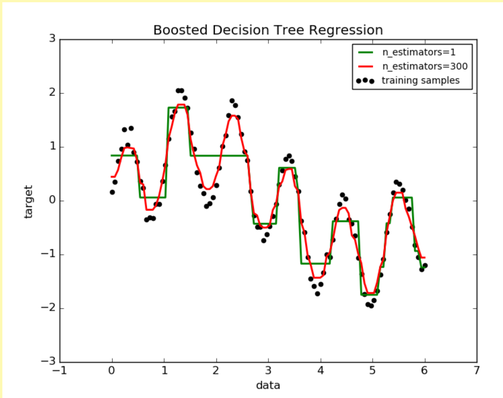
Template amplitude space

The figure below shows the training data projected into the space of the first two filter amplitudes. We train a boosted decision tree in the full 10 dimensional space of all of the filter amplitudes for each waveform to predict the amount of collected charge associated with each waveform.



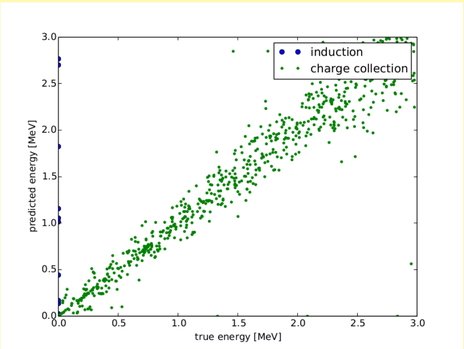
Boosted regression tree

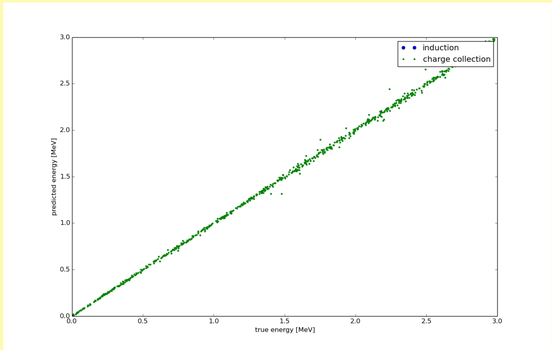
A regression tree fits a function to a series of threshold functions. We use regression trees with 20 threshold functions. We then use AdaBoost to boost together 300 regression trees. The figure below shows the improvement gained by boosting 300 regression trees together to estimate a function over using a single regression tree. The figure below is from [4]



Energy prediction

A boosted regression tree (BRT) is used to estimate the mapping from the 10 dimensional space of filter amplitudes to the energy associated with each waveform. A boosted regression tre estimates a function by boosting together threshold functions. The plot on the left shows the result of training the BRT with just the first two filter amplitudes. With just the first two filter amplitudes, 7 induction wavefroms were erronously given energies and the average error in energy estimation of collection waveforms was 10.2%. Once we trained in the full 10 dimensional space, no induction waveforms were given energies and the average error was reduced to 0.1%.





Outlook: comprison to conventional analysis.

This algorithm for estimating the amount of energy associated with each waveform outperfoms the traditional analysis for the very simple simulated data set used here. Future work will see if this can be extended to more complicated events with noise and multiple charge depositions.

Acknowledgments: The work was supported by the NSF and DOE