Discerning WIMPs from Neutrons in XENON1T

Alexander Shaw

Whitman College

August 2, 2017
Dark Matter - A way we know it exists [1]
WIMP basics [9]

- **Weakly Interacting** - WIMPs interact solely via the weak force and gravity
- **Massive** - with a range of 1 GeV - 100 TeV (5 order of magnitude!)
- **Particles** - KE of WIMP \(< 100 \text{ keV}\) – Elastic scattering with normal matter nuclei would produce energies in 1 - 100 keV!
- How might we see these particles as directly as possible?
XENON1T [2]
XENON detector concept [5]
Consider the Neutron

It, like DM, is neutrally charged
It will make energy depositions within the WIMP 1 - 100 keV range
It is a major background
The signal for a single scattering event will look identical to a WIMP’s!
The Important Difference:

The cross section is much larger! If we have a TPC large enough...

Multiple Scatter

Single Scatter
The decision process

- ER / NR - Ratio of S1 to S2, fitted with calibration data.
- MS / SS - Look at the features of the S2s, fit with calibration data?
The important Features of S2

Main S2

Largest other S2 Possibilities

True Scatter Single Electron Multi-electron Pileup

Main S2

Largest other S2 Possibilities

Width 50% area

Main S2

Largest other S2 Possibilities

Area

Main S2

Largest other S2 Possibilities

Time Delay from Main S2
Understanding signals of an event

That is very messy! The features of the peaks are used to sort through and classify a multiple scatter and a single scatter.
How are decisions made with these features?

- How much physics is behind it?
- After physics and calibration, just deciding “by eye”? 
- Humans and “by eye”... seems nonphysical
- Why not use the ultimate nonphysical decision maker - a machine learner

My project: establish a better method of seeing single scatters through machine learning
Current Single Scatter Cut

How well can we define this line? Have no pure multiple scatter source to fit with.
How can the line classify these events?

It really can’t.
What is Machine Learning?

For my purposes, it is a purely numerical method of classifying data.

Basic method:

1. Find data with known classifications
2. Train an algorithm with this data
3. Use the algorithm to predict unclassified data
Using the Random Forest [8]

Random Forest Classifier: A randomly grown collection of decision trees. Will I go out today?

A decision tree grows from training data. It makes decisions like these based on maximizing information gain in the training process.
Random Forest: Many decision trees [6]
A Single Scatter S2 cut with Machine Learning

The most important thing a learner needs is purely classified data. I want to train on Monte Carlo simulation data.

\[ \text{MonteCarlo} \rightarrow \text{WaveformSimulation} \rightarrow \text{Analysis} \]

1. Monte Carlo outputs “truth information” : Info of the event fed into the detector (Don’t have this with real data)
2. Truth information \rightarrow MS or SS
3. Feed event to the Waveform Simulator, get simulation data just like real
4. But, you know if it is a MS or SS

Need purely classified data, and accurate data
Need to match Sim data with real data
Need to match Sim data with real data
So, it matches well enough

I can train the Random Forest, now!
We understand which features are important to discern a MS from a SS.
But, does the learner?
Optimizing the Random Forest

**Figure:** Optimizing on training sample size

**Figure:** Optimizing on depth of decision trees
Classifying with trained and optimized learner

**Figure:** Able to pull out data previously classified as single scatter

Votes of decision trees $\rightarrow$ probability. But, the probability is nonphysical.

**Figure:** The simple Single Scatter cut
Defining the probability cutoff
Defining the probability cutoff

Want to optimize our rejection and acceptance in a certain region of events.
Compare RFC and the simple SS cut

Better rejection! Could be used to better reject background
In short - What I’ve done

- Created a unique way to discern single scatters from multiple scatters
- Created framework for classifying simulation data
- Created framework for applying machine learning to classify data

This can be used for many different applications! And expanded upon...
Current Status [4]
Acknowledgements

Thank you to NSF and the REU Program at Nevis Labs for the amazing opportunity, especially Prof. Karagiorgi and Prof. Parsons. Thank you to Prof. Aprile and the entire XENON Dark Matter Project for teaching me so much, especially Fei Gao, Patrick de Perio, Tianyu Zhu, Joey Howlett, and Qing Lin.
ER and NR [4]

- Rn - an electron source - produces Electronic Recoil (ER)
- AmBe - a neutron source - produces Nuclear Recoil (NR)
- Define regions in S2 and S1 plot as ER or NR
- How do we know if the dark matter search signal is a neutron or a WIMP?

Figure: Dark Matter Search data and ER/NR fits

https://3c1703fe8d.site.internapcdn.net/newman/gfx/news/hires/2017/1-xenon1tthemoe.jpg


https://d2wh20haedxe3f.cloudfront.net/sites/default/files/random_forest_diagram_complete.png

https://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm.