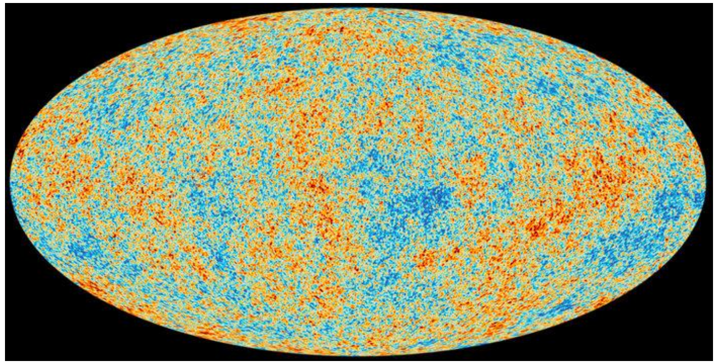
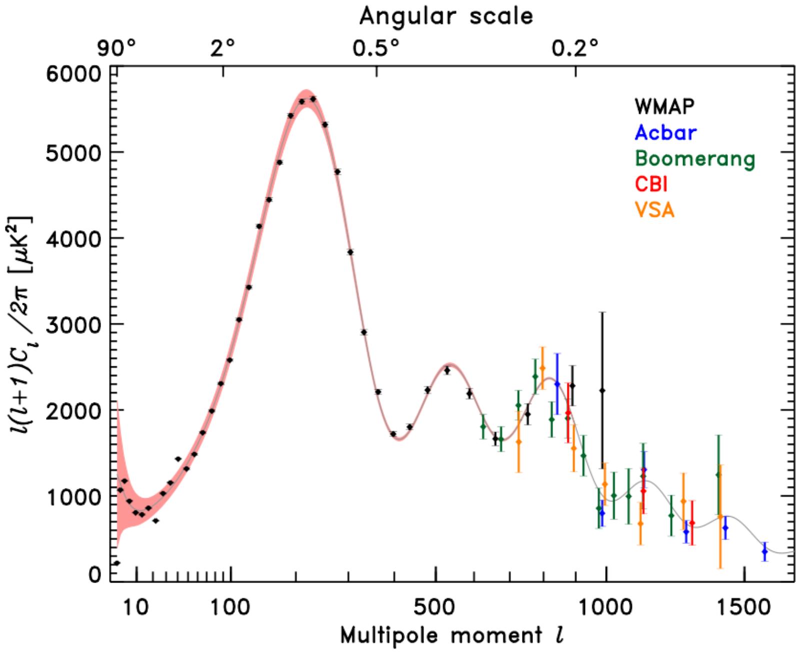
SEMI SUPERVISED LEARNING
FOR PARTICLE IDENTIFICATION IN PICO-60

Gevy Cao on behalf of Brendon Matusch
For the PICO Collaboration
ACAT 2019
Saas Fee, Switzerland
March 14, 2019

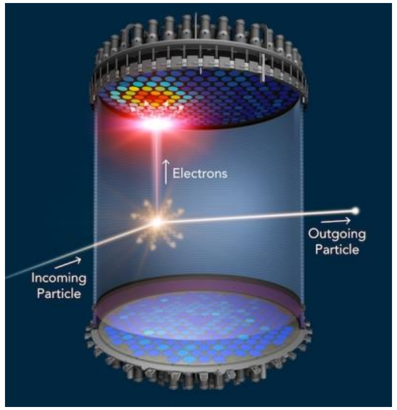
DARK MATTER EVIDENCE

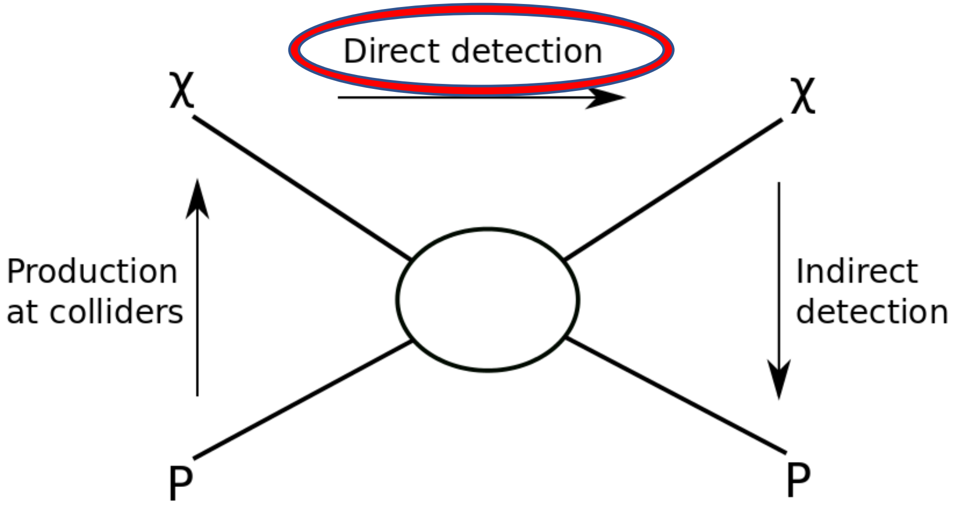




- Baryon Acoustic Oscillation in the early universe
  - Universe expands smoothly but not perfectly. Dark matter in over-dense regions collapse gravitationally. Radiation pressure repels the collapse.
  - Damping in the power spectrum -> high photon diffusion -> indicate high density of dark matter
  - Model fitting verifies the hypothetical density of dark matter

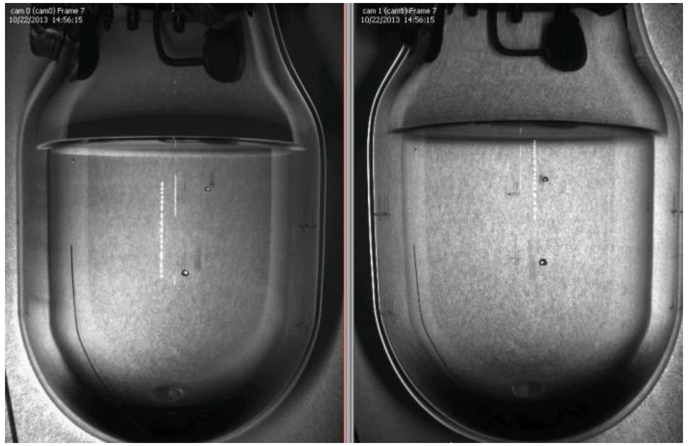
DARK MATTER DETECTION



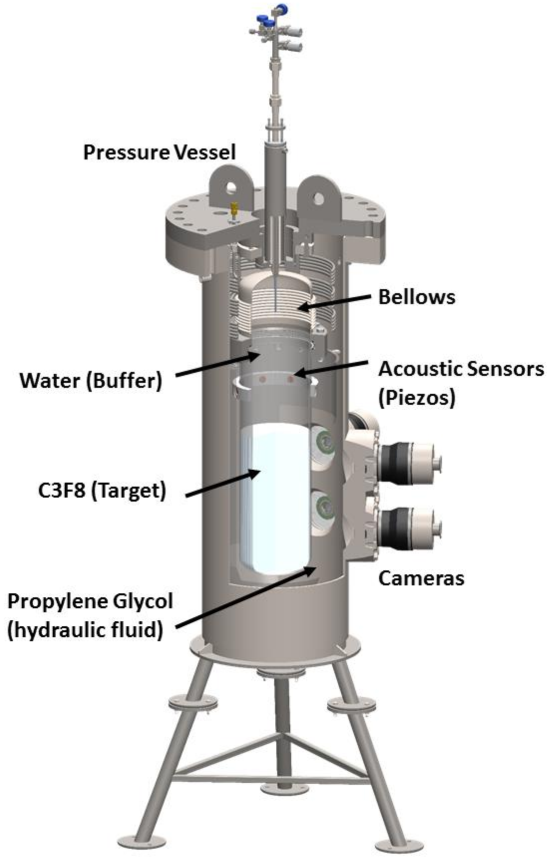


- Direct Detection of Dark Matter:
  - Gaseous detectors
    - TPC (directional sensitivity)
  - Solid crystal detectors
    - Cryogenic
  - Liquid detectors
    - Superheat

PICO EXPERIMENT

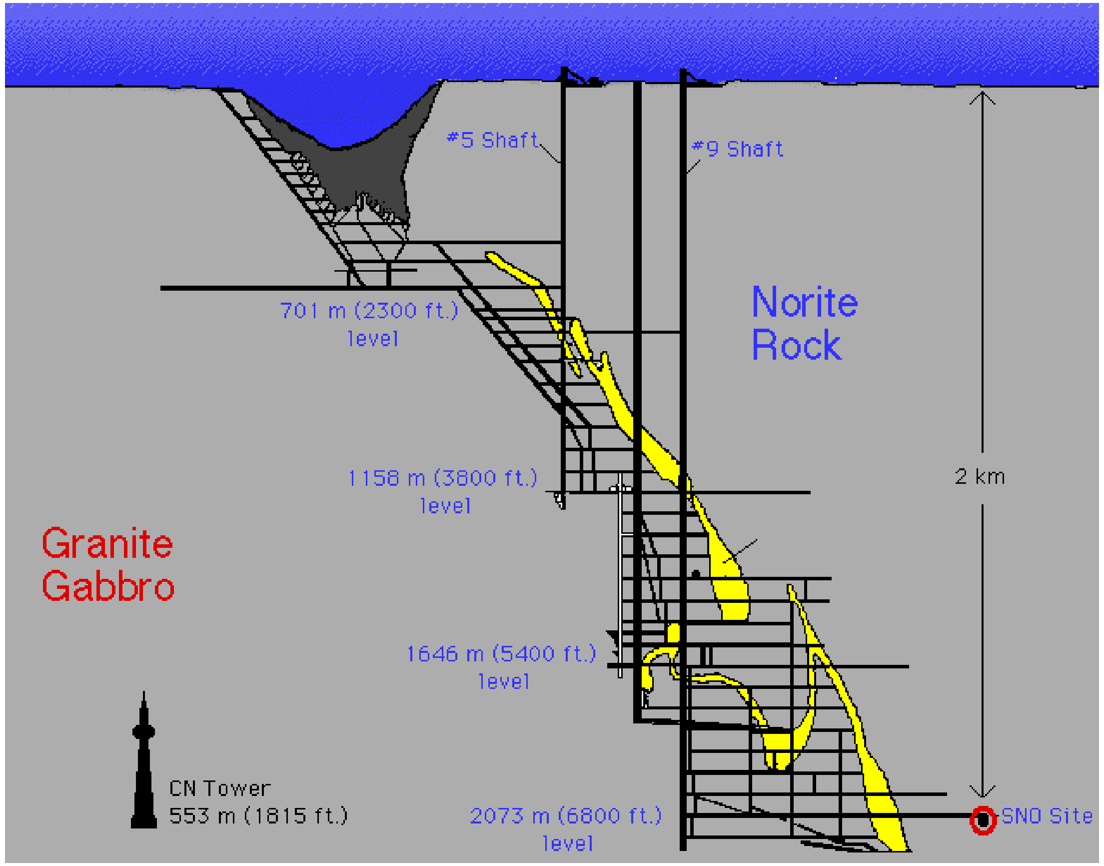


- Superheated bubble chamber technology
  - Slowly heat up the target liquid (C_3F_8) above its boiling point while applying pressure.
  - Slowly lower the pressure to reach superheat -> metastable state.
  - Energy deposition in the form of nuclear recoil in the detector causes the state of fluid to jump over the Gibbs thermodynamic potential -> nucleation of a bubble.

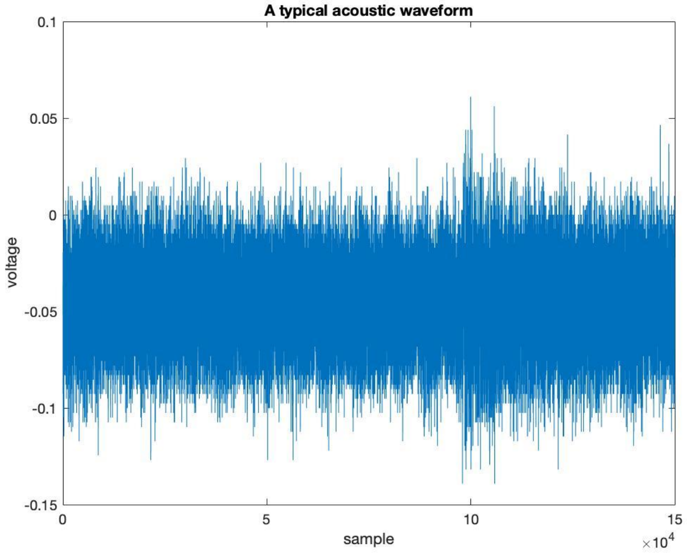


BACKGROUND

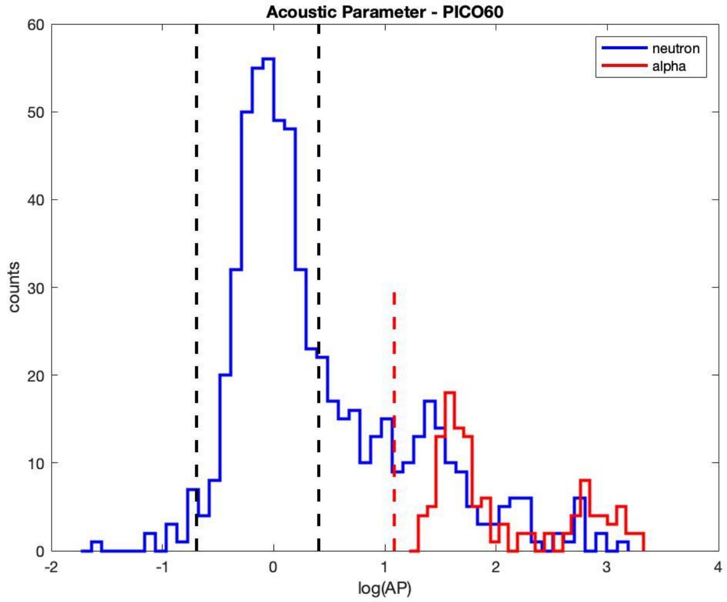
- Cosmic muons/neutrons
  - Shielded by 2km of rock
  - More neutron shielding provided by 1.5m of water
- Gamma radiation
  - Excellent gamma rejection (10^{-12})
  - New electron recoil model in progress
  - Scintillating bubble chamber
- **Alpha recoils**
  - **Acoustic discrimination**



CURRENT P.I.D IN PICO



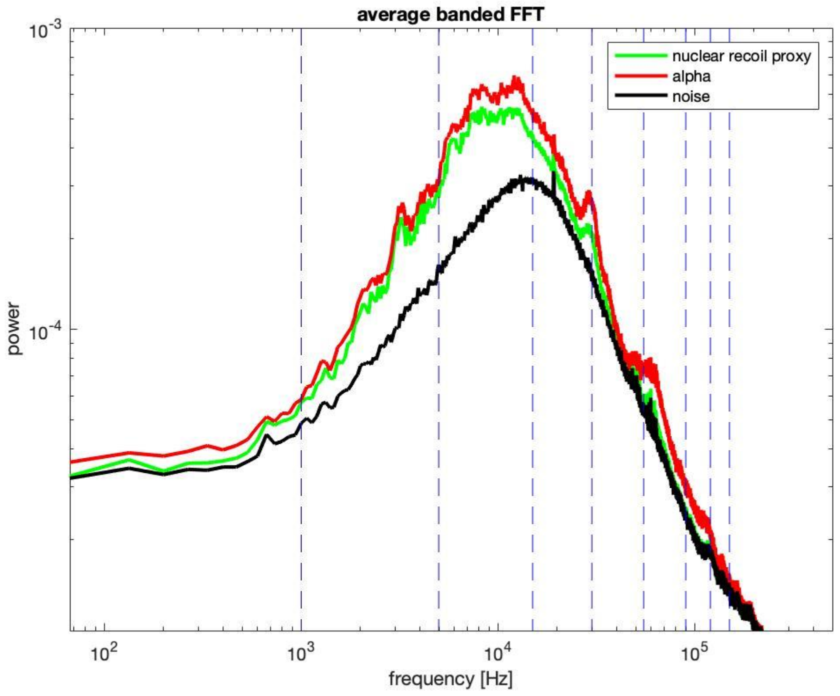
- Amplitude of the acoustic wave is measured for each bubble in time domain
- Fourier transform -> filter -> Acoustic Parameter (AP) calculated in most distinguishable frequency bands.
- Nuclear recoil events normalized to AP = 1.
- Cut set to 2 σ from mean.
- **Problem**: technique tuning required for a new detector and after large calibration data sets; mid-AP events
- Can machine learning help?



MACHINE LEARNING IN PICO

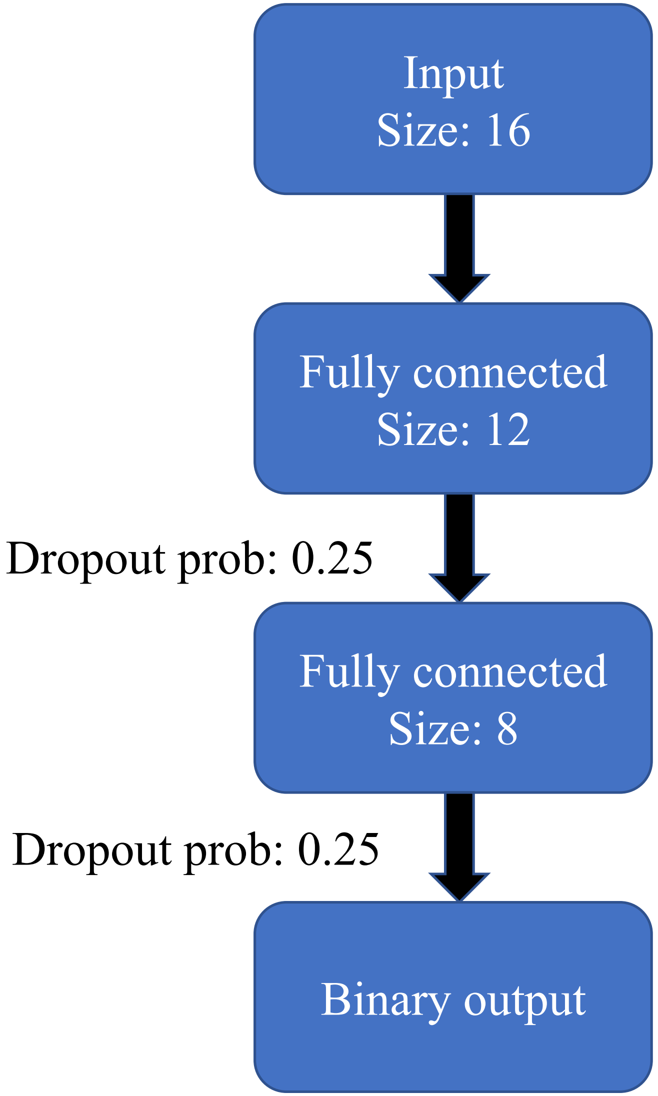
- Goal: Discriminate alpha recoils from nuclear recoils
- Performance assessed against AP
- Input choices: raw waveform, integrated FFT, full FFT, images, position corrected FFT.

Configuration	Max Accuracy	Precision	Recall	CWSD
$DeepConv(\omega)$	95%	97%	97%	0.52
$FourierMLP(PosCor(\beta_8))$	96%	98%	97%	0.42
$FourierMLP(\beta_8)$	98%	100%	98%	0.29
$FourierMLP(\beta_{50,001})$	95%	95%	98%	0.46
$ImageConv(\iota)$	68%	68%	100%	0.97



NETWORK STRUCTURE

- Input: 8 position-corrected, banded fourier power for each of the two working piezos
- Cost function: simple mse
- Work is done in python using tensorflow



SEMI-SUPERVISED LEARNING

Gravitational Differentiation

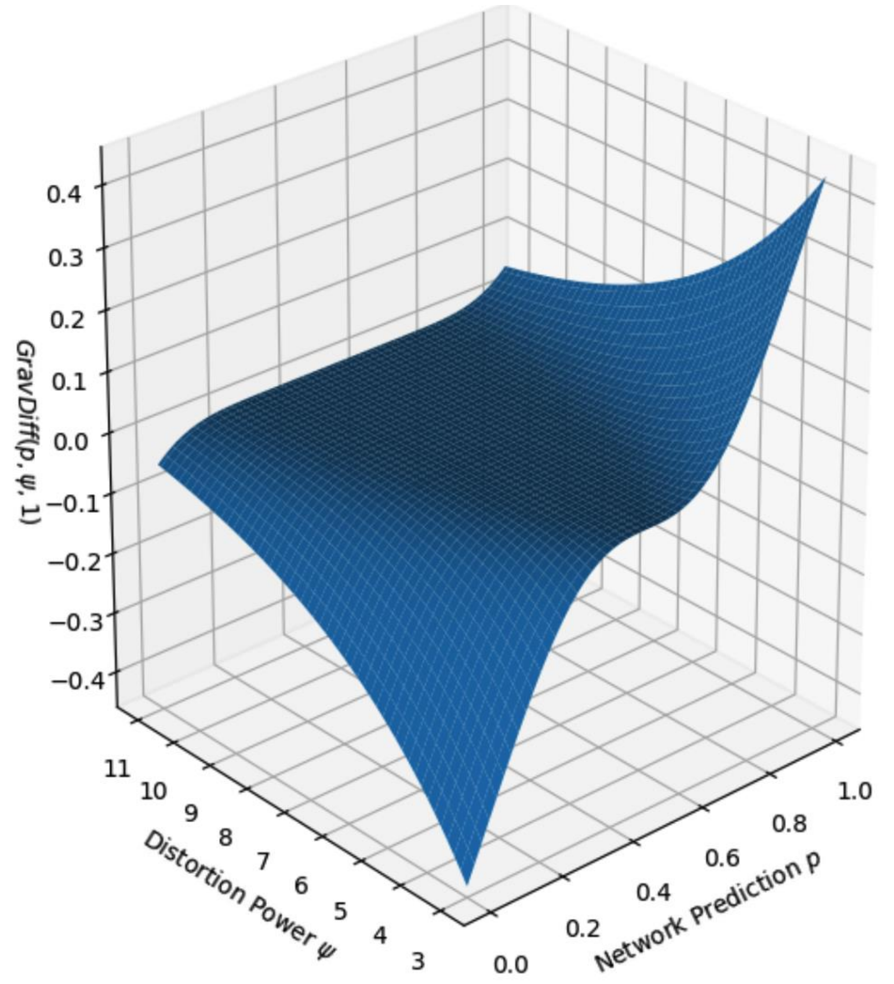
- Set aside some labeled (GT 0 or 1) training data
- All unlabeled training data have a ground truth value of 0.5
- As training proceeds, the ground truth value is updated such that:

$$GT = p + GravDiff$$

$$GravDiff = g * \mathrm{sgn}(p) * abs\left(\tanh\left(2(p - 0.5)\right)\right)^{\varphi}$$

- Produces large gradients for predictions close to 0 or 1, and flattens out low-confidence predictions.

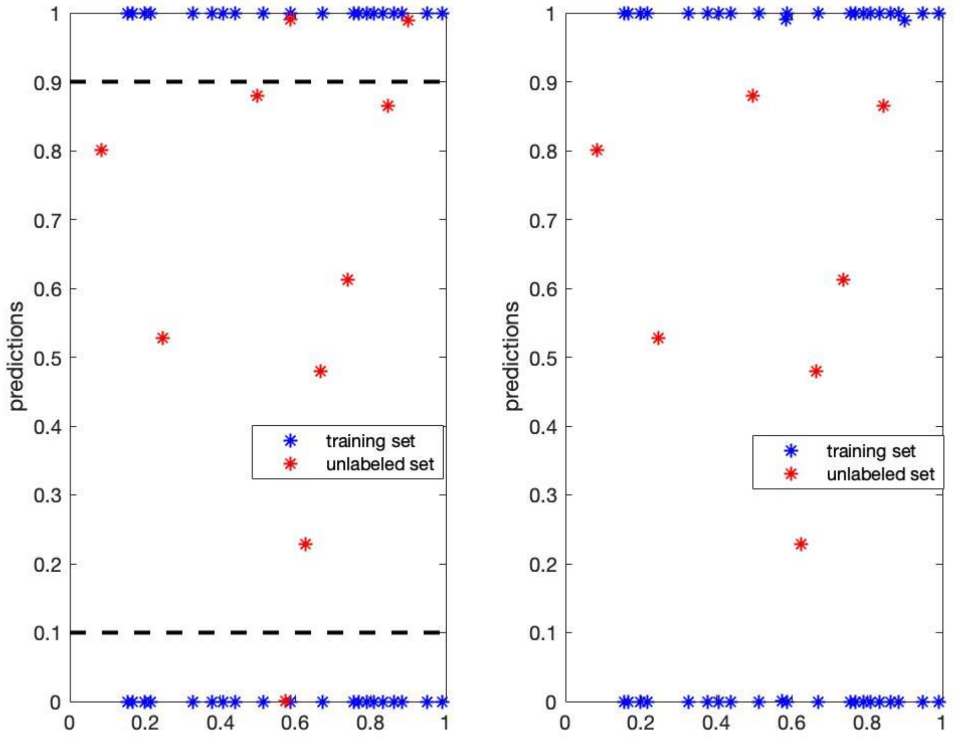
$$\frac{dE}{dw_i} = \sum_{i=0}^{m} (GT_i - p_i) * f(x_i)$$



SEMI-SUPERVISED LEARNING

Iterative Clustering

- Train a set of labeled data for N epochs
- Produce predictions on unlabeled data
- If the prediction is within a distance σ from 0 or 1, add the data sample to labeled training data

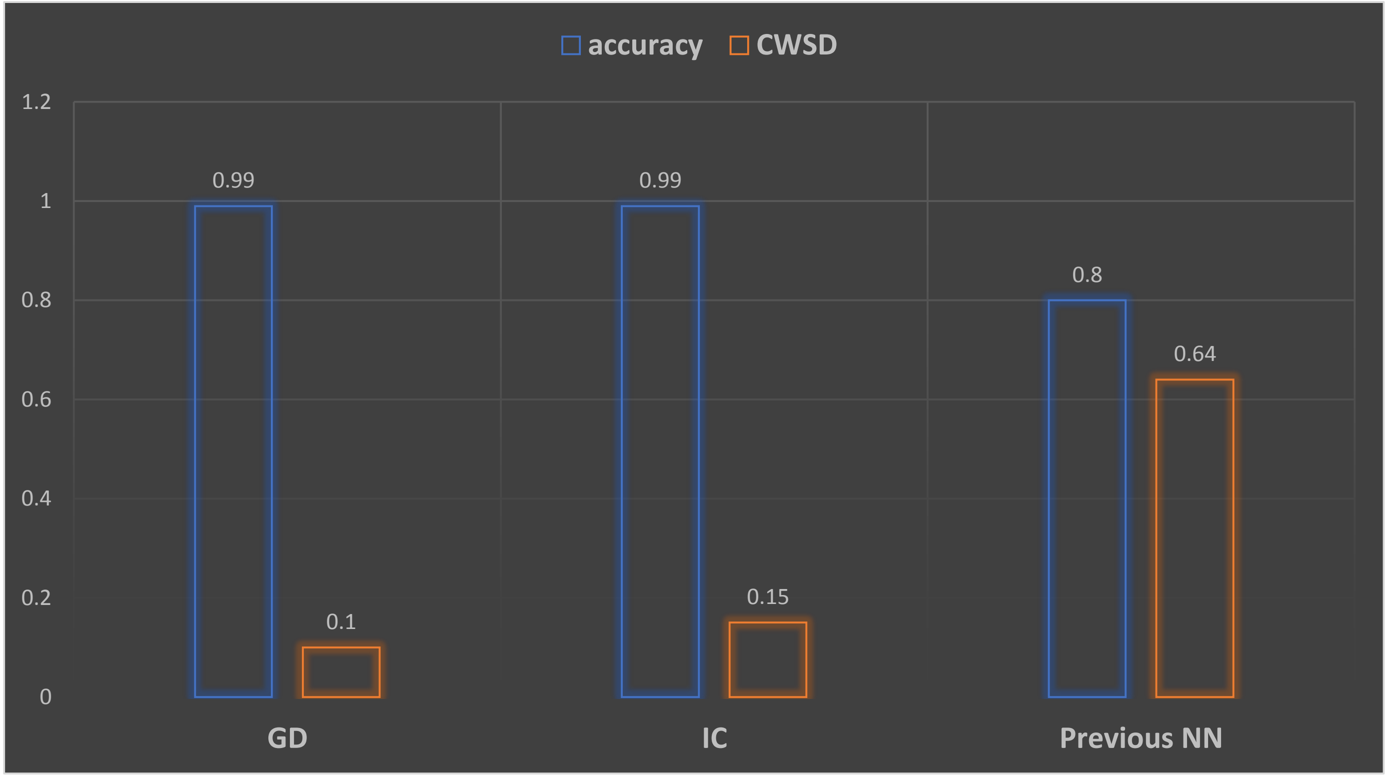


PERFORMANCE ASSESSMENT

Nuclear recoil dataset N, Alpha recoil dataset A, the class-wise standard deviation (CWSD) is defined as:

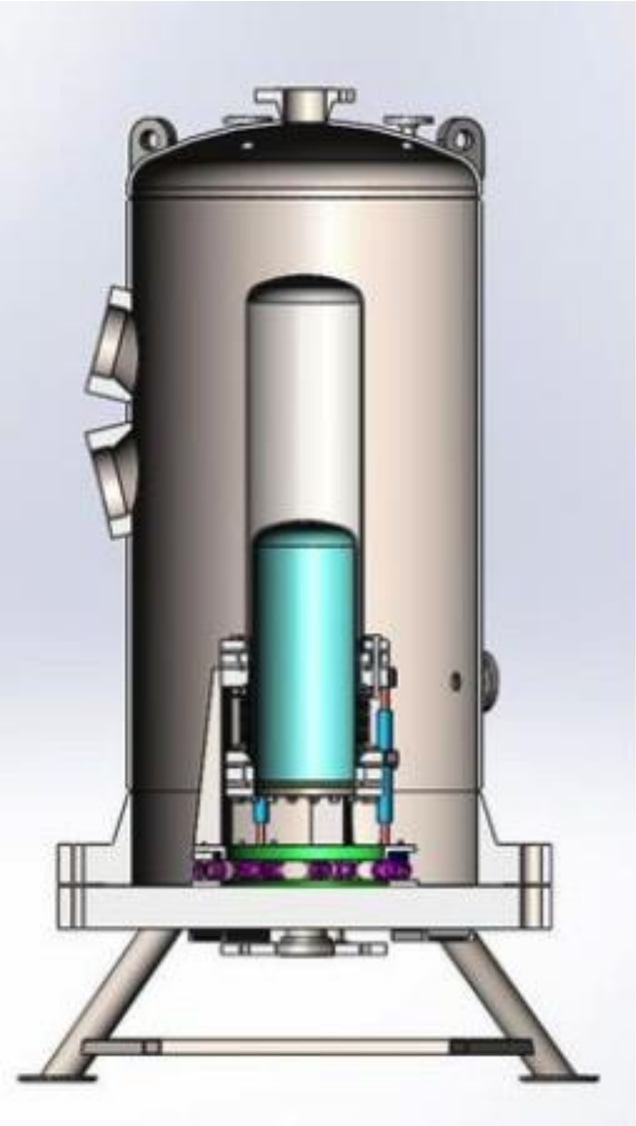
$$S = std([N, A])$$

$$CWSD = \frac{std\left(\frac{N}{S}\right) + std\left(\frac{A}{S}\right)}{2}$$



CONCLUSION AND OUTLOOK

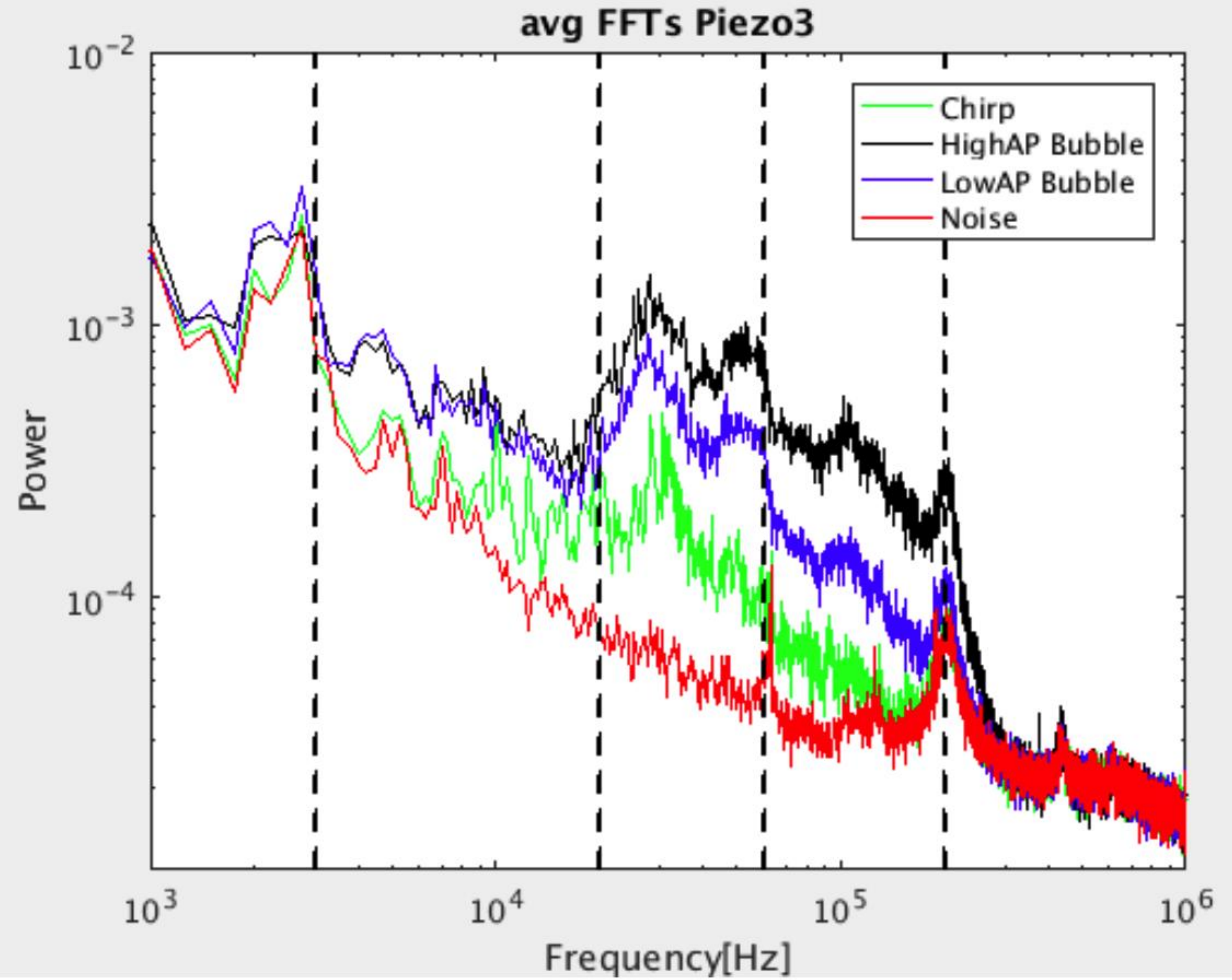
- ML can not only reproduce AP results, but can also provide a quick intermediate check on particle identification without the need to renormalize AP for each calibration set.
- PICO-40L will be operating in the next month or two
  - Preliminary P.I.D using ML to complement AP
  - Can be applied to a small set of calibration data
- Machine learning may help us refine acoustic cuts with a more definite classification of mid-AP events.



Results published on ArXiv 1811.11308

FROM THE AUTHOR

The author of this work, Brendon Matusch, welcomes any further discussions or opportunities.
He can be reached at: brendon-m@outlook.com
More of his work can be found at: https://github.com/bfmat



avg FFTs Piezo3
Chirp
HighAP Bubble
LowAP Bubble
Noise
Power
Frequency[Hz]