

# AI Lunch

## Problem of the Quarter

Jacob Murphy

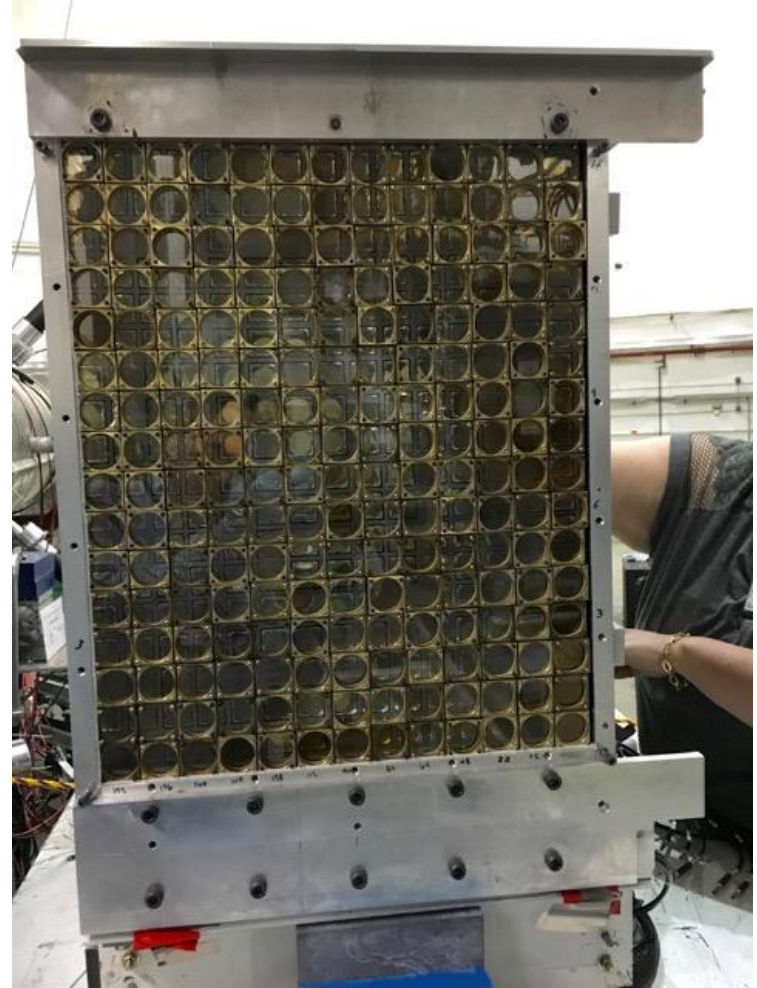
Ohio University

[jm443918@ohio.edu](mailto:jm443918@ohio.edu)

August 2020

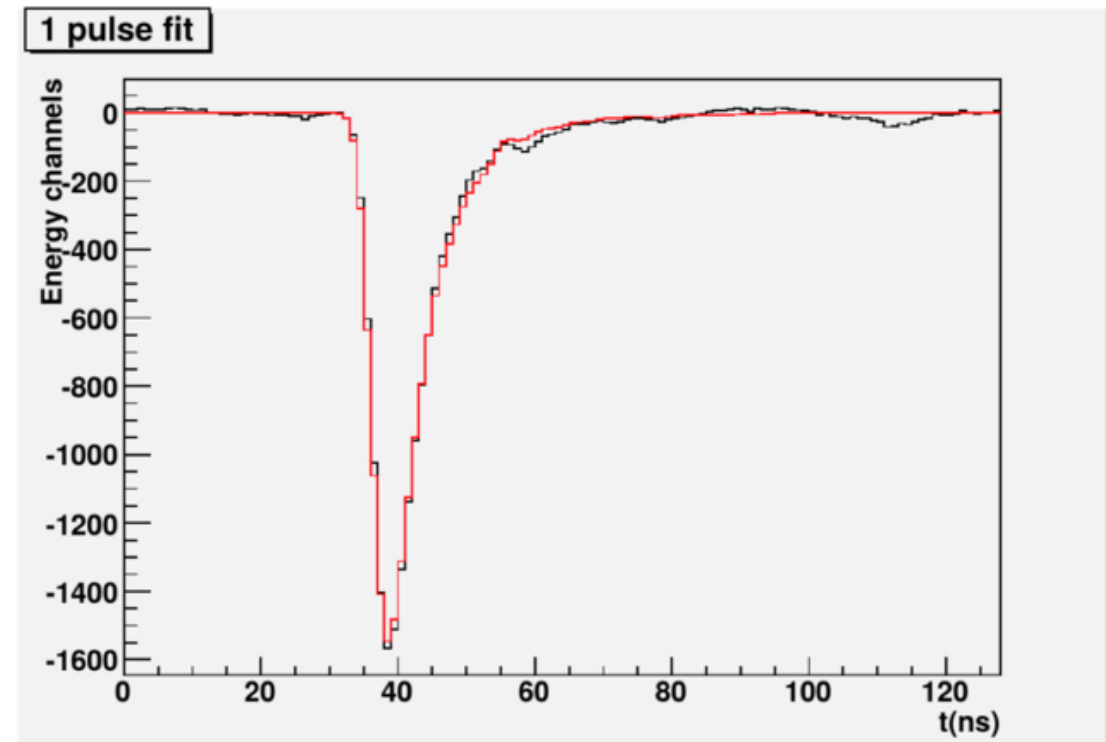
# DVCS in Hall A Calorimeter

- A detector consisting of 208 blocks struck by high-energy photons
- Photon Events saved in 128-ns chunks, voltage value 1 per ns
- Interested in arrival-time (to better than 1ns) and amplitude of photon pulses (to within 2%)



# 128-ns Event Signal

- One-pulse event, with an amplitude of  $\sim 1600$  and arrival time of  $\sim 39$  ns
- Signal values recorded at time 0 ns to time 127 ns
- Most events with a pulse are One-pulse, though most recorded events have no pulse at all (flatline or noise)



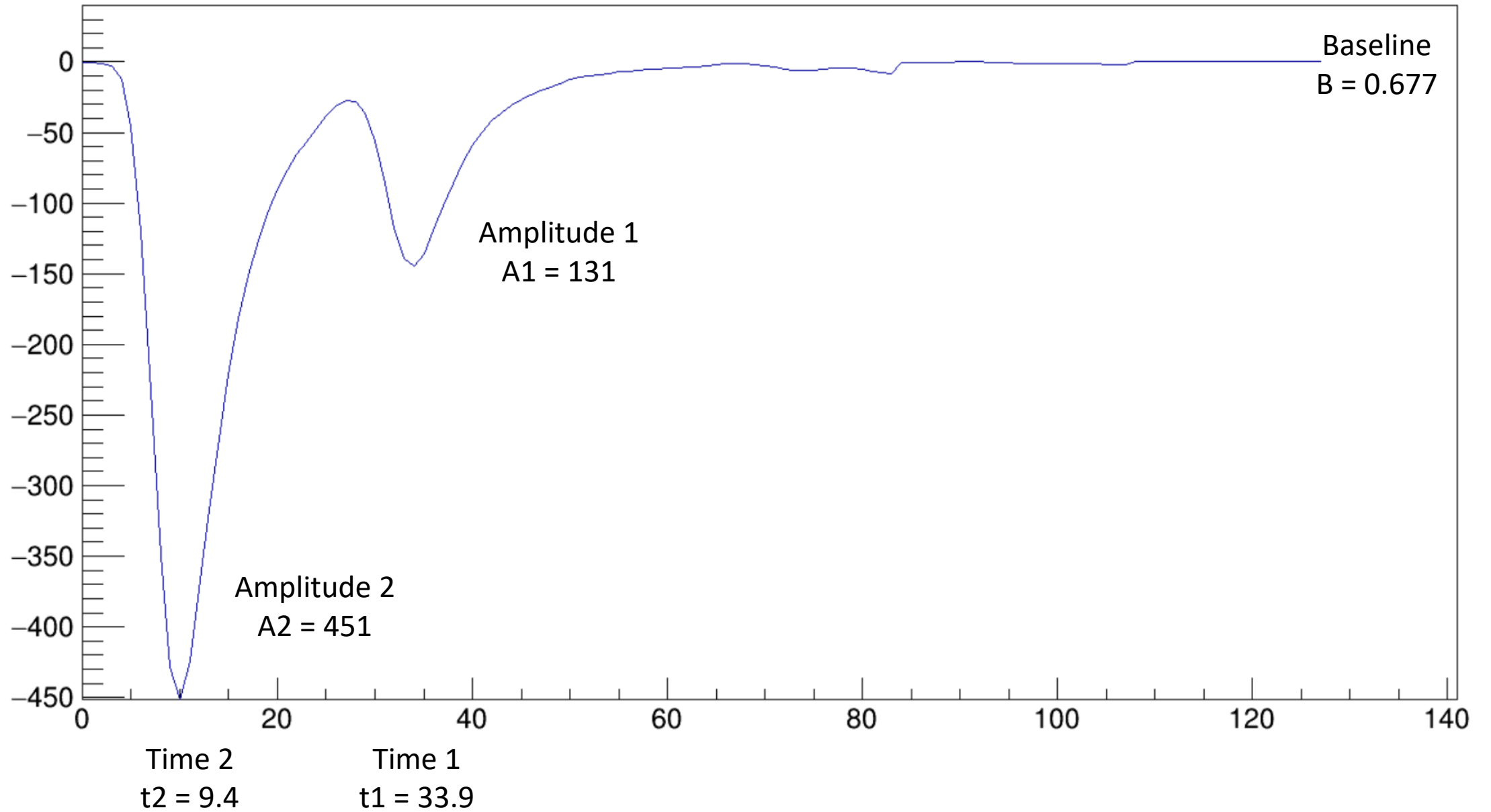
# Goal

- Create a model that inputs a 128-signal array and outputs
  - 2 amplitude values,  $A_1$  &  $A_2$
  - 2 arrival time values,  $t_1$  &  $t_2$
- Model should first determine number of pulses: 0, 1, or 2
  - For no pulses, both amplitudes should output as 0 while both times should output as -1
  - For one pulse, EITHER  $A_1$  and  $t_1$  OR  $A_2$  and  $t_2$  should output as 0 and -1
- Model should then find amplitude and arrival time of any pulses

# Training Data Set

- Participants are given a training set of data in the form of a CSV file with over 190k events
- Each row is an event with 133 values. The first 5 are:
  - Baseline value ( see next slide; this is to potentially help with your training but should not be an output in the test)
  - Amplitude 1 value
  - Amplitude 2 value
  - Time 1 value
  - Time 2 value
- The remaining 128 are the signal arrays, starting from time 0 to 127

# Example of a 2-Pulse Event



# Event Types in Training Set

- No Pulse
  - All amplitude values are 0, all time values -1
  - Most common
- One Pulse
  - One amplitude (A1) recorded with arrival time (t1)
    - $A2 = 0, t2 = -1$
- Two Pulse
  - Two amplitudes (A1 and A2) recorded with arrival times (t1 and t2)
  - Least common, pose potential issue with overlapping pulses

# Test Format

- The Test Data Set will be made available **Wednesday, November 4<sup>th</sup> at noon**
  - The structure will be similar to that of the training data, but with only the 128 signals
- Submissions will be due 48 hours later on **Friday, November 6<sup>th</sup> at noon**
  - Submission must include all code used for training and testing
- Results from test should be submitted as a CSV file
  - Each row should correspond to the events in the Test Set
  - There should be 4 columns: A1, A2, t1, t2
  - The Baseline value should not be included and does not need to be found



# Judging and Scoring

- Each submitted variable will be compared to the true value, taking the difference squared
  - Labels of pulse 1 vs 2 are arbitrary, so all submitted pulses will be compared to both true pulse value sets; the smaller difference will be used for judgement
- Number of pulses will also be compared to true value
  - This number does not need to be submitted separately
  - Number of pulses determined by non-zero amplitudes with times not -1
- Differences will be summed for each variable, points being given to lowest sums
  - 5 separate sums (A1, A2, t1, t2, pulse number)
  - Lowest sum in each category gets 3 points, then 2, then 1
- Winner has most points in the end, ties decided by score for pulse number

# File Locations and Turn-in

- Training data, along with this presentation and the accompanying write-up, are available at:
  - [https://halldweb.jlab.org/talks/ML\\_lunch/Aug2020/](https://halldweb.jlab.org/talks/ML_lunch/Aug2020/)
  - Testing Data will also be available here on November 4<sup>th</sup>
- Alternatively, you can request these documents directly from Jacob Murphy (me) at [jm443918@ohio.edu](mailto:jm443918@ohio.edu)
- Submissions (Due November 6<sup>th</sup>!) should also be sent to [jm443918@ohio.edu](mailto:jm443918@ohio.edu)