Projects with UC Irvine Department of Physics & Astronomy (working in the lab of Dr. Daniel Whiteson):

Working with Dr. Whiteson's group in the UC Irvine Department of Physics and Astronomy, I developed methods for analyzing data from high-energy particle collisions. Results were greater than 90% accuracy for high momentum collisions and < 2% error for heavy particle reconstruction.

1. Heavy Particle Reconstruction (June 2022 – Present):

I developed methods of analyzing properties of particles that decay too quickly to be observed by detectors. These methods will be used to analyze the properties of a hypothetical new fundamental particle.

I simulated particle collisions, developed methods, and reconstructed heavy intermediate particles. By reverse-engineering the decays, I was able to reconstruct masses with < 2% error.

MadGraph was used to simulate collisions, Pythia8 for hadronization, Delphes for detection, and ROOT for analysis. Reconstruction programs were written in Python and C++ (using functions from the PyROOT and ROOT libraries).

Some examples of decays I analyzed are listed below.

Z^0 Boson Decay:

A proton – proton collision was simulated, resulting in the production of two Z^0 bosons, each of which decayed into a muon – anti-muon pair. After successfully reconstructing leptonic decays, Z^0 bosons were reconstructed from jets with each boson decaying into two jets (Fig. 1).

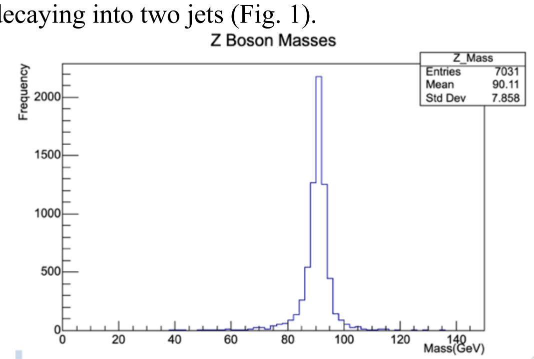

Fig. 1: Results for Z^0 jet decay reconstruction. Average reconstructed mass is 90.11 GeV, true mass is 91.18 GeV. Histogram generated using ROOT.

Top Quark Decay:

A proton – proton collision was simulated, resulting in the production of a top – anti-top pair. Each top quark decayed into a W boson and a bottom quark (Fig. 2), resulting in the production of up to 16 jets.

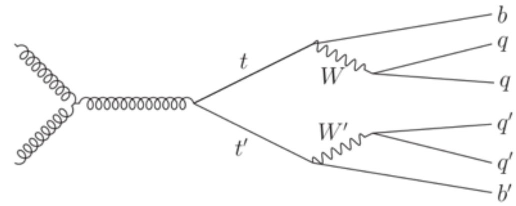

Fig 2: Feynman Diagram for $t - \bar{t}$ decay.

An algorithm I devised, and later incorporated the χ^2 method from arXiv:2010.09206 (Permutationless Many-Jet Event Reconstruction with Symmetry Preserving Attention Networks), reconstructed both top quarks (Fig. 3).

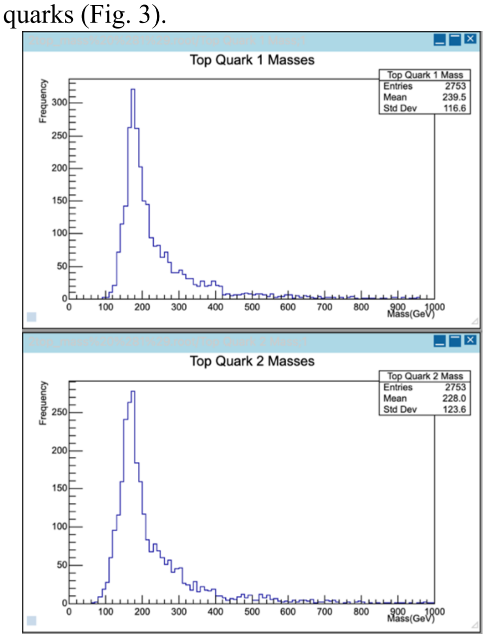

Fig. 3: Results for $t - \bar{t}$ decay reconstruction. Median reconstructed mass is approximately 175 GeV, true mass is 172.76 GeV. Histogram generated using ROOT.

2. High Momentum Collision Analysis (February 2022 – June 2022):

Often in High-Energy Physics, large amounts of data are available for collisions with low transverse momentum (p_t), but not for those with high p_t. As a result, it is difficult to analyze data from collisions with high p_t because there is not enough data to train a machine learning model effectively. I trained neural networks on low p_t data and analyzed their ability to extrapolate to high p_t data depending on the information they were given.

We found that parameterized networks perform significantly better than unparameterized with approximately 90% accuracy and with an ROC AUC that was approximately 0.1 greater than their unparameterized counterparts (Fig. 7).

Our parameterized networks used p_t (Fig. 4) to assist in prediction. Two features (f_1 and f_2) were generated as a function of p_t and a variable θ randomly drawn from a normal distribution (Eqs. 1, 2, & 3).

$$f_1 = p_t^{0.2} + l \cdot p_t \cdot (\sin(\theta) - 0.5) + 50 \quad (1)$$
$$f_2 = p_t^{1.1} + l \cdot p_t \cdot (\cos(\theta) - 0.5) + 5 \quad (2)$$
$$\theta \sim N(\mu = 0, \sigma = \tfrac{\pi}{3}) \quad (3)$$

Three signals were generated: signal source (src), signal target (tar), and a background signal. Each data point was composed of a p_t, f_1, and f_2 component, with signal source having low p_t and signal target having high p_t (Figs. 4 & 5).

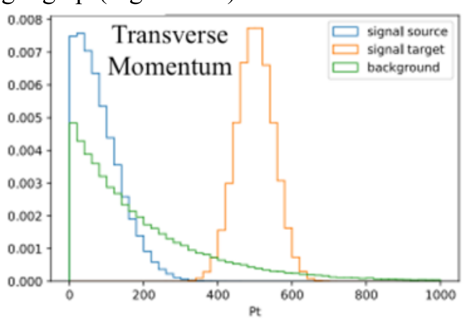

Fig. 4: Unweighted p_t distributions.

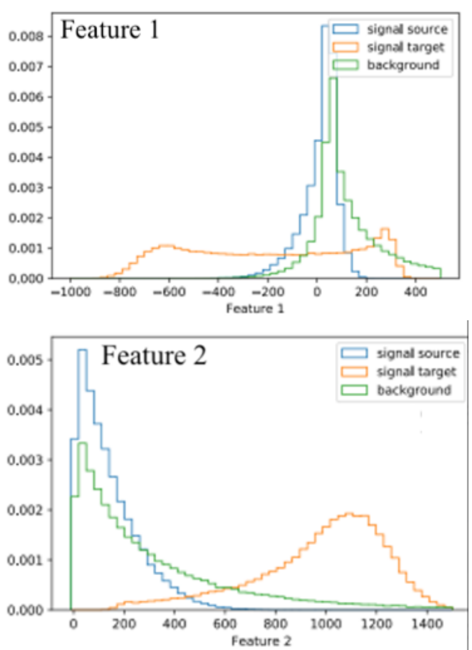

Fig. 5: Unweighted f_1, and f_2 distributions.

To mask the discriminating power of p_t during training, sample weights were assigned to each data point, reweighting signal source to match background (Fig. 6).

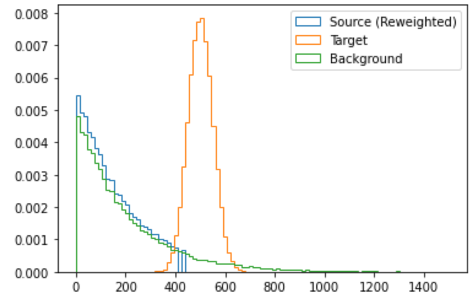

Fig. 6: Reweighted p_t distributions.

For deep neural network development, Python, Tensorflow, and Keras were used. 4 networks were constructed. 2 were provided only f_1 and f_2 as inputs (unparameterized), the remaining 2 were provided p_t, f_1, and f_2 (parameterized). Each network had a single output node (sigmoid output function) indicating the probability that the data point was background. Building on the work outlined in arXiv:1601.07913 (Parameterized Machine Learning for High-Energy Physics), each network was trained for signal-background classification between signal source and background and evaluated on classification ability between signal target and background. Results were analyzed using the Area Under the Curve (AUC) of Receiver Operating Characteristic (ROC) Curves (Fig. 7).

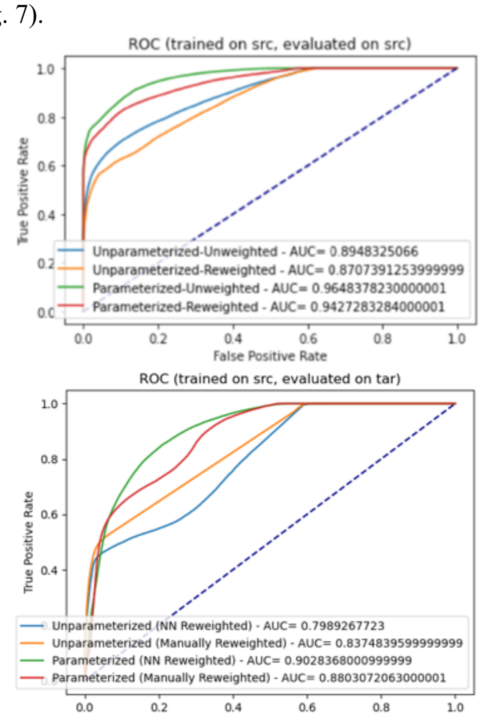

Fig. 7: ROC Curves for models. Note that the closer the Area Under the ROC Curve (AUC) is to 1, the better the model