DEEP LEARNING FOR HIGGS BOSON IDENTIFICATION AND SEARCHES FOR NEW PHYSICS AT THE LARGE HADRON COLLIDER

Allocation: Illinois/645.929 Knh
PI: Mark Neubauer[1]
Collaborators: Robert Gardner[2], Dewen Zhong[1]

[1]University of Illinois at Urbana–Champaign
[2]University of Chicago

EXECUTIVE SUMMARY

The Large Hadron Collider (LHC) is the world's most powerful particle accelerator, designed to study the fundamental nature of matter and the forces that govern its interactions by colliding beams of protons at the highest-available energies. The research team is using Blue Waters to process, simulate, and analyze high-energy proton–proton collision data produced by the ATLAS experiment at the LHC and to improve researchers' sensitivity to new phenomena by developing novel approaches to identifying Higgs bosons produced with high momentum at the LHC by using machine learning techniques.

RESEARCH CHALLENGE

The goal of particle physics is to understand the universe at its most fundamental level, including the constituents of matter, their interactions, and the nature of space and time itself. This quest is one of the most ambitious and enduring of human endeavors.

The Standard Model (SM) of particle physics describes all known fundamental particles and their interactions, including the Higgs boson, which was discovered at the LHC in 2012 with significant contributions by the Illinois (Neubauer) Group. The discovery led to François Englert and Peter W. Higgs receiving the 2013 Nobel Prize in Physics. The SM has withstood the last 40 years of experimental scrutiny, with important exceptions being neutrino mass, dark matter, and dark energy. Recent developments in particle physics and cosmology raise the exciting prospect that we are on the threshold of a major step forward in our understanding.

It is an enormous challenge to process, analyze, and share the 15 petabytes of data generated by the LHC experiments each year with thousands of physicists around the world. To translate the observed data into insights about fundamental physics, the important quantum mechanical processes and the detector's responses to them need to be simulated to a high level of detail and with a high degree of accuracy.

A key thrust of this project is to use the Higgs boson to search for new physics in novel ways enabled by the Blue Waters supercomputer. The enormous energy available in proton–proton collisions at the LHC leads to the production of particles with very high velocity relative to the ATLAS detector (lab frame). Even massive particles like the Higgs boson can have a large momentum and, therefore, large Lorentz factor in the lab frame. When these "boosted" particles decay, their decay products are highly collimated (aka parallel rays) and not easily distinguished in the detector instrumentation (*e.g.*, by calorimeters). This limits the sensitivity of searches for new physics such as $X \rightarrow hS$, where X is a new massive (~TeV/c^2) particle, h is the Higgs boson, and S is a scalar particle that could be the h or a new particle yet to be discovered (Fig. 1).

The research team has integrated Blue Waters into their production processing environment to simulate and analyze massive amounts of LHC data. Blue Waters resources are made available to the ATLAS computing fabric using a system called ATLAS Connect, which is a set of computing services designed to augment existing tools and resources used by the U.S. ATLAS physics community. Docker images are run via Shifter to create an environment on Blue Waters' nodes that is compatible with the ATLAS job payload.

Through the work of Dewen Zhong using Blue Waters and with collaborators at Indiana University and the University of Göttingen, the research team developed a novel four-prong jet tagger for boosted $WW \rightarrow 4q$ identification based on machine learning techniques, including feed-forward deep, 2D convolutional, and recursive neural networks. Among the best performing approaches is a deep neural network (DNN) that has been the primary study of Zhong.

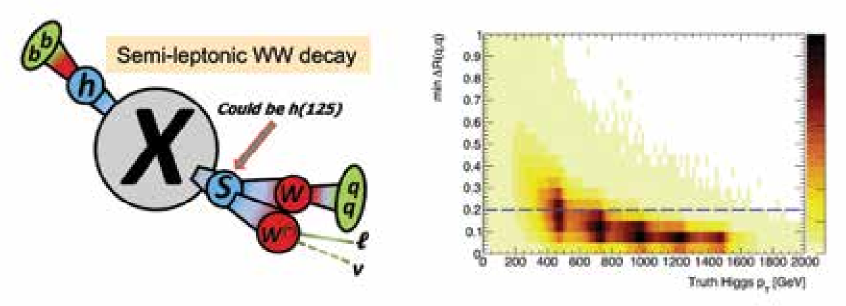

Figure 1: (Left) Decay of a heavy hypothetical particle X to a Higgs boson and a scalar particle that decays to a W boson pair. The daughter particles from this decay are highly collimated in the lab frame and therefore difficult to reconstruct using traditional methods. (Right) Minimum distance between final-state quarks versus Higgs boson transverse momentum. The dashed line shows the typical jet clustering radius, indicating that the majority of decay products are in close proximity to each other.

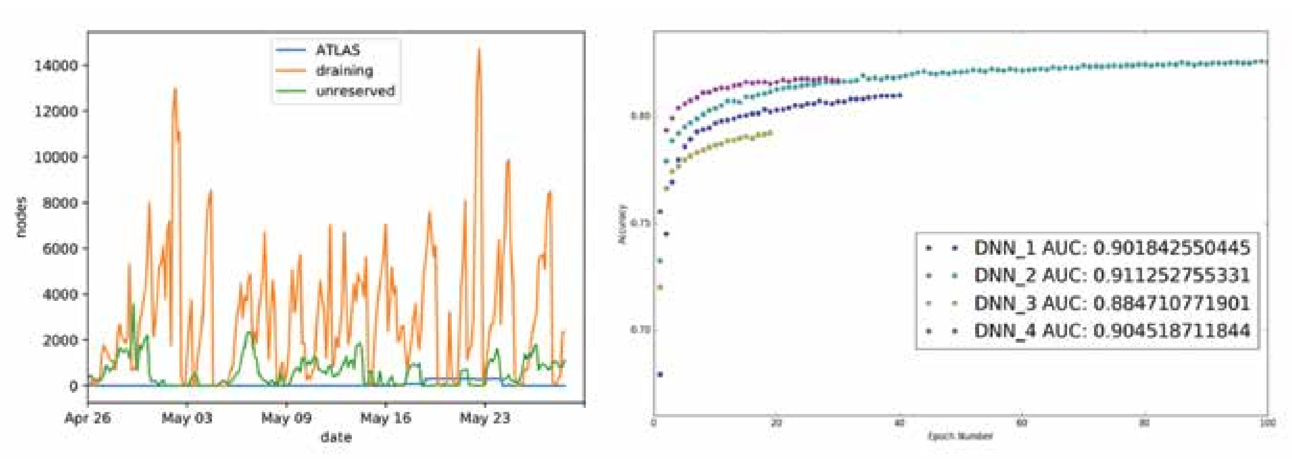

Figure 2: (Left) This shows a period of time during which 35 million ATLAS events were processed using 300 Blue Waters nodes. Utilization during this period averaged 81%, which is typical for Blue Waters. (Right) Higgs boson identification accuracy versus training epoch for the boosted Higgs boson tagger developed on Blue Waters.

The research team is working to extend this approach to the single-lepton channel, which brings additional challenges, and to develop it as a general WW tagger for ATLAS that can be applied to searches involving boosted dibosons and not limited to the semileptonic Higgs+scalar searches described here. The team is using Blue Waters to implement this approach as a jet-mass decorrelated tagger to avoid learning the mass and training DNNs using both calorimeter and tracking information.

RESULTS & IMPACT

Fig. 2a shows a particular one-month period in 2018 in which 35,000 Blue Waters cores were utilized to process 35 million collision events. The top panel of this figure shows that this approach is cost-effective, boosting cluster utilization, and has no adverse effect on other high-performance computing workloads. The job output was made available to the rest of the ATLAS collaboration for use in analysis of the LHC data to improve SM measurements and to search for new physics beyond the SM. Fig. 2b shows the Higgs boson identification accuracy as a function of the number of training epochs for a variety of DNN configurations and hyperparameter settings. The team is also using Hyperas, a convenience wrapper using Hyperopt with Keras models, on Blue Waters to automate the scanning of hyperparameters in a variety of machine and deep learning approaches to improve the Higgs boson identification over backgrounds. The techniques show promise in addressing the challenges of boosted Higgs boson identification and improving the sensitivity of new physics searches at the LHC.

WHY BLUE WATERS

Blue Waters, as a large CPU and GPU resource with high data-throughput capabilities, greatly facilitated this research. The strong support for containers allowed the research team to deploy their science application on Blue Waters' nodes. Also, Blue Waters provided a means for a highly parallelized and automated scanning of free parameters in the team's machine learning configurations and, therefore, rapid optimization of the researchers' boosted Higgs boson identifier.

PUBLICATIONS & DATA SETS

M. Belkin *et al.*, "Container solutions for HPC systems: A case study of using shifter on Blue Waters," in *Practice and Experience on Advanced Research Computing 2018, Proc. PEARC '18*, Pittsburgh, PA, U.S.A, July 22–26, 2018, pp. 43.1–43.8.

E. A. Huerta, R. Haas, S. Jha, M. Neubauer, and D. S. Katz, "Supporting high-performance and high-throughput computing for experimental science," *Comput. Softw. Big Sci.*, vol. 3, no. 5, 2019.