# Expansion of the Capabilities of the MMS Ground Loop System

ETHAN DAVIS
DR. MATTHEW ARGALL
UNIVERSITY OF NEW HAMPSHIRE
PHYS 795/799

# Project Background and Motivation

- MMS was launched in 2015, can only downlink 4% of high-resolution data.
- Scientist-in-the-Loop (SITL) manually searches low resolution data to label possible events.
- SITL is good but unsustainable long term.
- Argall et al. 2020[1] proposed using an ensemble of machine learning models to form the Ground Loop System (GLS), which will ultimately automate event detection.
- Two parts: work on speeding up labeler and model creation/evaluation.

# Labeler

- Originally created by Vyacheslav Olshevsky[2] during a joint project by KTH and IRFU in Sweden.
- At the start, the program was mostly complete but not working.
- Most of the work was spent on replacing the SciKit[3] Primary Component Analysis (PCA) with SciKit-CUDA , which interfaces with NVIDIA's CUDA toolset through PyCuda[4] .
- The CUDA toolset is published by NVIDIA and allows certain operations to be run on an NVIDIA GPU rather than a CPU.

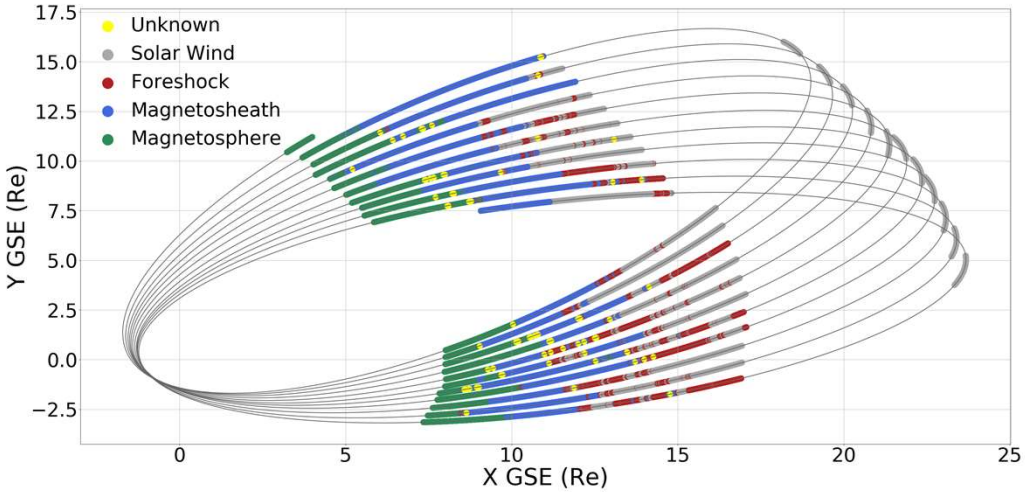


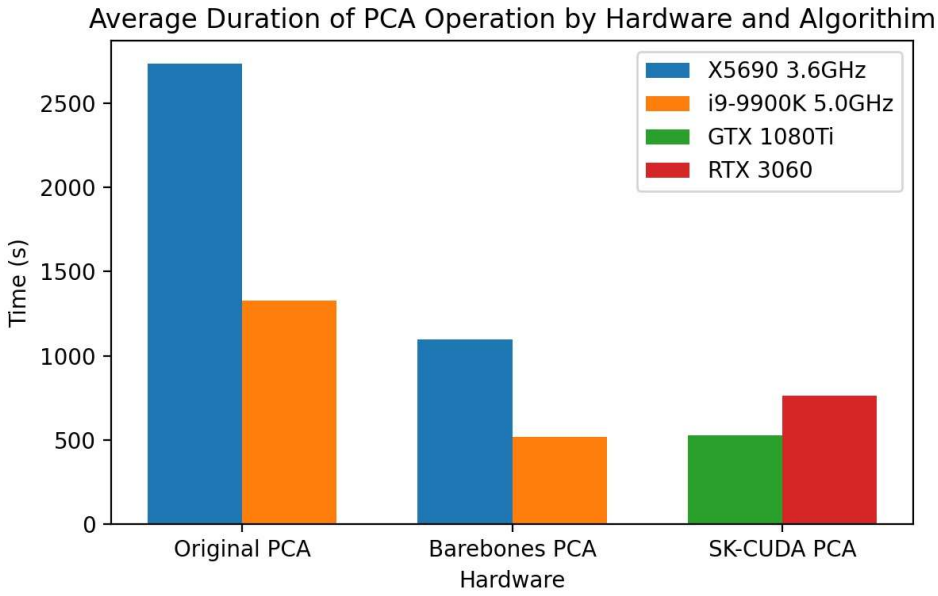

Image credit: Olshevsky, 2020

# Why a GPU?

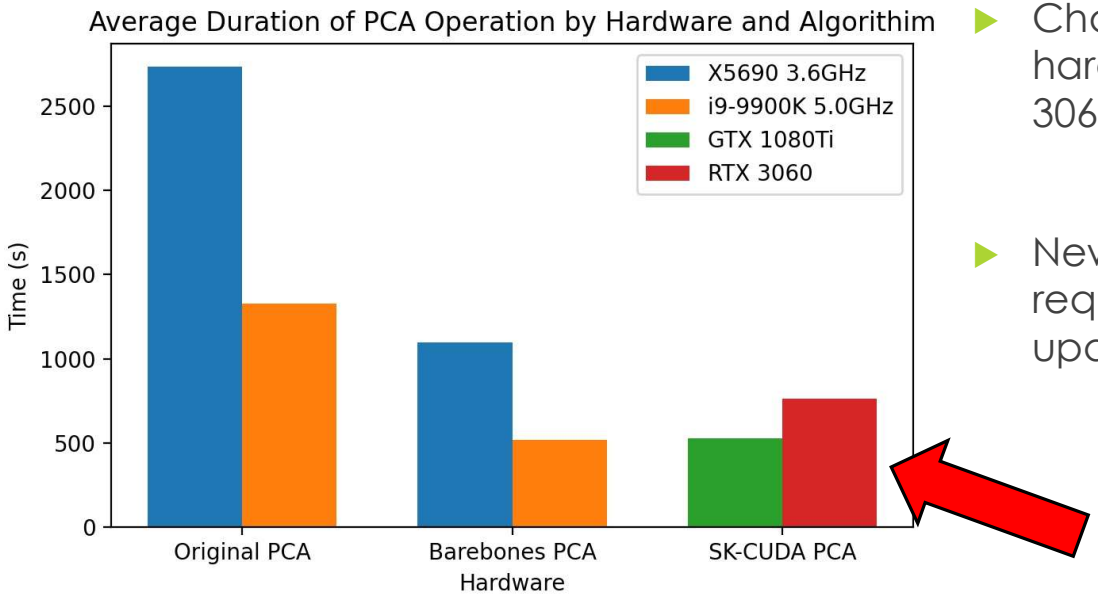
- Reason #1: Structure
- Example: my desktop's CPU has 8 cores at max 5.0GHz each, 16 threads, a 16MB cache. The GPU has 3584 cores at max 1837MHz each (not including Tensor cores), and 12GB of memory it can access at 15Gb/s.

- Reason #2: Data Transfer
- Because the CPU has a small cache, it must constantly move data from RAM to threads. The GPU can store an entire month of data in its memory and distribute it to the cards cores efficiently.

# PCA Improvement Results



# PCA Improvement Results



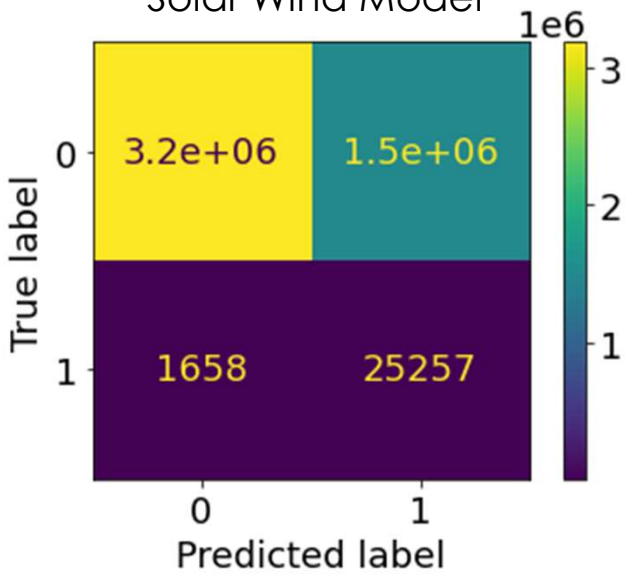

- Challenge: Incongruous hardware/software combination with RTX 3060
- New card architecture and CUDA requirements combined with non-updated packages result in slow down

# Event Detection Models

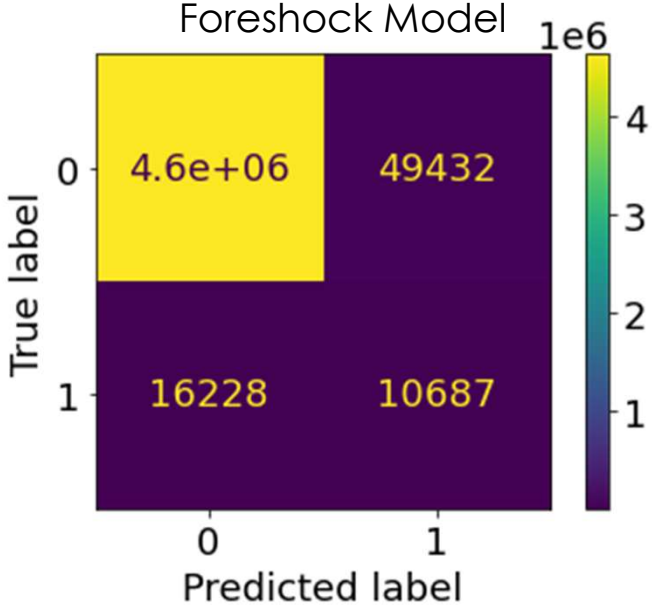
- Neural Networks built in TensorFlow Keras in Python.
- Trained on data pulled from MMS Science Center via PyMMS.[5]
- A model is created for each event, so individual models are effectively doing binary classification
- Models created for Solar Wind, Magnetosheath crossings, Foreshock crossings, and Bow Shock crossings.

# Event Detection Model Results

Solar Wind Model




Foreshock Model



# Challenges with Networks

- Very disproportionate classes. Events selected by SITL frequently made up <0.1% of all data points. To counter this, oversampling and undersampling were applied
- Networks can be sensitive and have many hyperparameters to tune
- Not all datasets or events are predisposed to similar model designs

# Future Work

- Improvement of quality of current models

# Works Cited


- [1] M.R. Argall, C.R. Small, S. Piatt, L. Breen, M. Petrik, K. Kokkonen, J. Barnum, K. Larsen, F.D. Wilder, M. Oka, W.R. Paterson, R.B. Torbert, R.E. Ergun, T. Phan, B.L. Giles, and J.L. Burch, Frontiers in Astronomy and Space Sciences **7**, (2020).
- [2] Olshevsky, V. (n.d.). *Bitbucket*. https://bitbucket.org/volshevsky/mmslearning/src/master/.
- [3] Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., … Varoquaux, G. (2013, September 1). API design for machine learning software: experiences from the scikit-learn project. arXiv.org. https://arxiv.org/abs/1309.0238.
- [4] Andreas Klöckner, Nicolas Pinto, Yunsup Lee, Bryan Catanzaro, Paul Ivanov, Ahmed Fasih, PyCUDA and PyOpenCL: A scripting-based approach to GPU run-time code generation, Parallel Computing, Volume 38, Issue 3, March 2012, Pages 157-174.