KAITLYN CHEN

kaichen@g.hmc.edu ⋄ github.com/kaitlynchen1 ⋄ kaitlynchen1.github.io

EDUCATION

Harvey Mudd College *Expected B.S. Computer Science/Physics 2026*
Relevant Courses: Data Structures/Program Development (C++), Data Science, Probability, and Statistics (Python), Principles of Computer Science (Java, DrRacket), Intro to Computer Science (Python), Discrete Math, Quantum Physics, Electromagnetic Theory & Optics, Mechanics & Wave Motion, Linear Algebra, Single & Multivariable Calculus

EXPERIENCE

1. **National Radio Astronomy Observatory Intern**, Green Bank Obs. *May 2024-present*
Skills: Open Source Software Developing, Documentation, GitHub, Data Analysis, Scientific Writing, Python, C++, Python Scripts, Unix

   Analyzing the effectiveness of various radio frequency interference (RFI) mitigation algorithms with Dr. Evan Smith. Creating an open source pip installable package with the analyzed RFI mitigation tools. With the package, work can be replicated and tested on more types of data sets with different parameters. Used matplotlib to have functions that can plot spectrograms, fine channelized spectra, and phase plots, allowing the user to easily visualize their data before and after RFI mitigation. Documenting package on Read the Docs.

2. **Carnegie Astrophysics Intern**, Carnegie Observatories *June 2023-present*
Skills: Data Science, Machine Learning, Markov Chain Monte Carlo (MCMC), Scientific Writing, GitHub, Python: matplotlib, pandas, numpy, emcee, Python Scripts, Unix

   Using machine learning to analyze yellow supergiant spectra from the Magellan Inamori Kyocera Echelle spectrograph with Dr. Trevor Dorn-Wallenstein. Established a Bayesian framework to regress fundamental parameters on our highly-dimensional dataset by using MCMC to fit models to our data. Enhanced technical skills in Python, data visualization with matplotlib, data manipulation with pandas and numpy, and Unix commands in Terminal. Worked with different fits and hdf5 files to load and save spectral data. Tabulated the results with pandas and created visualizations of the sample in Hertzsprung-Russell (HR) and spectroscopic HR diagrams. Calculated binomial confidence intervals with astropy to determine the significance of factors that played into the stars' evolutionary status and plotted them with matplotlib.

3. **Training Machine Learning Model: SPOCK**, Harvey Mudd *December-May 2023*
Skills: Machine Learning: scikit-learn, GitHub, Terminal, Ubuntu, SSH, conda, dask, Python: numpy, matplotlib

   Worked with Prof. Daniel Tamayo updating and training the machine learning (ML) model Stability of Planetary Orbital Configurations Klassifier (SPOCK) with scikit-learn. Pulled and pushed the repository of the ML model to make accessible to other users. Used the gradient-boosted decision tree algorithm, XGBoost. Took high-dimensional inputs and reduced it to 10 features. Debug original code to work with a newer version of python and changes in the dask library used for parallelization. Used GitHub to make the updates accessible and conda commands to have multiple python environments to work between the old and updated ML model.

4. **Inner Edge of Planetary Period Ratio Distribution**, Harvey Mudd *August 2023-present*
   Skills: Machine Learning, Statistics: Pearson-R, K-S p-tests, Kernel Density Estimations, Terminal, Linux, GitHub, LaTeX, Python: pandas, numpy, matplotlib

   Simulating data using the machine learning model Stability of Planetary Orbital Configurations Klassifier (SPOCK) to analyze compact period ratios with Prof. Daniel Tamayo. Use Linux commands to work with high performance workstations. Use SPOCK to simulate planetary systems and analyze the significance of eccentricity, mass, and dispersion, of system survivability. Use scipy.stats to calculate Kolmogrov-Smirnov p-values to relate the simulated data to the NASA Exoplanet Archive, which was sorted through with pandas. Used matplotlib to graph changes in p-value for varying inputs of the parameters and used numpy to work with arrays of information stored from each simulation. Understand how changes in kernel density estimations for SPOCK's input parameters (eccentricity, mass, number of planets) affect the statistical test results.

5. **Identifying GD-1 Stellar Stream**, American Astronomical Society *January 2024*
   Skills: SQL: queries, joins, filtering, ADQL: polygon, Python: astroquery, pandas, gala, astropy, File Formats: hdf, fits, csv

   Learned to query astronomical databases. Use Gaia and Pan-STARRS data to identify and visualize the GD-1 stellar stream. Conduct coordinate transformations from ICRS to SkyCoord to GD-1 to work with the data.

PUBLICATIONS

1. **A Spectroscopic Hunt for Post-Red Supergiants in the Large Magellenic Cloud: I. Preliminary Results** *March 2024*
   Kaitlyn M. Chen, Trevor Z. Dorn-Wallenstein
2. **Characterizing Early Compact Planetary Formation with a Correlation Between Instability and Planet Loss** *in prep*
   K. Chen, Oswaldo Cardenas, B. Bonifacio, N. Hall, D. Tamayo
3. **Quantifying Radio Frequency Interference Mitigation Algorithms on Green Bank Telescope Data** *in prep*
   Kaitlyn M. Chen, Evan T. Smith

PRESENTATIONS

Hunting for Post-Red Supergiants iPoster, Chambliss Award Finalist, American Astronomical Society (AAS) *January 9, 2024*
Compact Planetary Formation Poster, , AAS Division on Dynamical Astronomy *May 16, 2024*
Hunting for Post-Red Supergiants Talk, Carnegie Observatories *August 25, 2023*

OUTREACH

Spectroscopy Demonstration, Upward Bound *July 2023*
Lead a comprehensive spectroscopy demonstration for 40 low-income/first generation high school students in a pre-college program.

OPEN SOURCE

movel *June 2024*
Developed and maintain a pip installable package that can calculate the first three moments of velocity in accordance with the equations from Briquet and Aerts (2002).

nettingi *June 2024*
Developing code with algorithms to "net" radio frequency interference.