

SPIN 2021 Exercise

<https://wiki.ncsa.illinois.edu/display/~rhaas/SPIN+2021+Exercise>

This exercise aims to provide a flavor of the expertise required for the SPIN project as well as giving you an idea of the work expected of you. If you choose to apply to work with me on this project then I will ask you to work through the exercise and send in your results before the interview with me. I provide a worked out solution so you are free to look at it to verify your own solution or to copy and past from it. However please keep in mind that the exercise mostly serves to give **you** an idea of the skills required to avoid signing up for more than you can handle.

Background

The Laser Interferometer Gravitational-Wave Observatory's (LIGO) detection of gravitational waves from merging black holes in September 2015 inaugurated a new era in astronomy and astrophysics, opening a window to observe the Universe through gravitational radiation. Occurring 100 years after Einstein's announcement of his theory of general relativity, the detection spurred world-wide interest in physics and science in general, making headline news around the world. The recent Nobel Prize awarded for this detection and the announcement of the detection of the double binary neutron star system by LIGO/Virgo underline the importance of these efforts and the interest that the wider society has in it.

Project details

This project is part of the ongoing effort in the [NCSA gravity group](#) to study gravitational waves produced by colliding compact objects like black holes and neutron stars. We use machine learning techniques to improve LIGO's detection capabilities to be able to detect gravitational waves more quickly (lower latency) and increase the set of waveforms that can be detected (for example detecting black holes on elliptic orbits as well as circular orbits).

Training these networks requires large numbers of training waveforms to be available to the network. We currently use LIGO's LIGO Algorithm Library (LAL) to produce these waveforms offline then read them in during the network training phase. For large parameter spaces this become infeasible since the training dataset increases in size to 100s of Terabytes. This project aims to produce training waveforms on the fly, interleaving a training epoch (on the GPU) with waveform production for the next epoch (on the CPU). You will use Python, LAL and TensorFlow to set up a pipeline that uses CPU cores to produce waveforms while using the produced waveforms to train neural networks. Students working on this project will have access to the HAL cluster at UIUC and Summit at Oak Ridge National lab.

Skills required:

- good understanding of Python
 - working knowledge of Python multiprocessing module
 - willing to learn new Python modules (h5py, numpy)
- experience using Linux and the command line interface
- some basic experience using Tensorflow

Before applying for this project please work through the exercise available on [SPIN 2021 Exercise](#) as this will be part of the interview with prospective candidates. I will not consider your application if I have not received the exercise **before** you apply.

Exercise

- install lalsuite in a Python virtualenv

```
mkdir lal
cd lal
virtualenv $PWD
source bin/activate
pip install lalsuite
```

- run the example code [exercise.py](#), which will produce several output files named `wv_NN.NNN_MM.MMM.dat`
- update `exercise.py` to use `multiprocessing.Pool` to generate waveforms in parallel, check that the produced output files are identical
- send the resulting script to rhaas@illinois.edu
- if you get stuck, you can peek at my [solution.py](#)