

## Faculty of Engineering

### Summer Research Program 2022-2023

Project Title: Information storage in synthetic DNA

Supervisor(s): Prof. Emanuele Viterbo

Department: Electrical and Computer Systems Engineering

Email: emanuele.viterbo@monash.edu

Website profile of project supervisor: <https://www.monash.edu/engineering/emanueleviterbo>

---

#### Objective

Design codebooks for encoding DNA that offer reliable and efficient decoding using Machine Learning techniques.

#### Project Details

DNA can be used as a storage medium with exceptional data density. Improving the reading (DNA sequencing) and writing (DNA synthesis) operations is the current challenge in this application. More recently, significant progress has been made in sequencing technology, where the underlying DNA strand may be identified by measuring a small current signal. These signals are highly stochastic in nature, and in order to improve reliability of identification, signal processing and machine learning can be utilized.

You will conduct a preliminary review of DNA sequencing, and then, during the first phase of the project, using a simulator for the current signals, you will design and train a neural network to identify DNA sequences. Then, you will create a codebook for DNA data storage by selecting sequences such that the probability of the neural network misidentifying one as another is minimized. You will further train the network with this set to improve the accuracy of the identification process.

#### Prerequisites

Experience in employing machine-learning techniques

Experience with programming in Python and/or Matlab

Familiarity with the Linux environment or the Docker platform in Windows

A strong grasp of the fundamental concepts of probability and signal processing

#### Additional Information

Applicants may be invited to an online discussion with the supervisor.