

Sleep Spindle as a Driver of Low Latency, Low Power ML in HLS4ML

Friday 7 October 2022 12:50 (3 minutes)

A specific type of Electroencephalography (EEG) signal, sleep spindle, is believed to contribute to neuronal plasticity and memory consolidation. In this project, we proposed a system that is based on ultra-low latency and power FPGA to detect and interact with the sleep spindles to help neuroscientists to understand the mechanism behind the theory. The proposed system will have a programmed FPGA and a headstage. The headstage will record the subject's brain signals and the FPGA will be connected with the headstage and process those signals to detect and interact with the sleep spindles.

This system is currently under development. We are working on implementing the HLS4ML to support the baseline deep learning model for this system. The baseline model is named Latent Factor Analysis via Dynamical Systems (LFADs). LFADs is an RNN variational autoencoder for analyzing spiking neural data. LFADs follows the encoder-decoder structure, and the main component for the LFADs' encoder is a bidirectional GRU, and the decoder is a unidirectional GRU. A dense layer is followed by the decoder to reduce the data to a set of low-dimensional temporal factors, which is LFADs latent. Another dense layer is followed by the latent information to create the log firing rate. The LFADs latent and the log firing rate are two important outputs from LFADs.

We faced several challenges during implementation. First, LFADs is a custom model, and this type of model is not supported by HLS4ML. We recreated the LFADs based on the Keras functional API and copied the weights and bias to test and ensure the recreated model has the same model architecture as the original one. Besides, the gaussian sampling and the bidirectional GRU layer are also not supported in HLS4ML. We removed the gaussian sampling from LFADs and found the performance does not decrease much based on the same dataset. Thus, we decided to have HLS4ML support this non-variational LFADs first and add the gaussian sampling back later. For the current stage, we are implementing the bidirectional GRU layer for HLS4ML and hope to finish it soon.

Author: LIU, Xiaohan

Presenter: LIU, Xiaohan

Session Classification: Lightening talks