Sleep Spindles as a Driver of Low Latency, Low Power ML in HLS4ML & TinyML

Neural Data – Sleep Spindles

Sleep Spindles Introduction
- Rare low-frequency brain signals
- Primarily occur during sleep or rest
- Are believed to contribute to learning
- Lack of mechanistic understanding
- Our goal: Design and build a system that can help neuroscientists to understand the mechanism behind the theory

The Proposed System

The brain signals will be analyzed by a deep learning model, which will be pushed through the HLS4ML.

Methods (HLS4ML & TinyML)

Baseline Deep Learning Model

Latent Factor Analysis via Dynamical Systems (LFADs)
- RNN variational autoencoder (VAE) in tf.keras API
- Input: Neural spiking data
- Output: Firing Rates & LFADs Latent Factors

Modified LFADs Architecture

By removing the gaussian sampling layer, LFADs are converted to an autoencoder, which is easier to be pushed through HLS4ML flow.

Implementing the gaussian sampling layer in HLS4ML is through HLS4ML flow.

The negative poisson log-likelihood is the evaluation metric of the LFADs. Minimized negative poisson log-likelihood indicates an optimal performance.

For the same testing dataset, the numerical value of the negative poisson log-likelihood from the modified LFADs matches to the original LFADs, which indicates that removing the gaussian sampling from LFADs is acceptable.

Future Work

MRAE contains multiple LFADs-like models, which are separated to different sections to process neural data in different frequency. Deploy MRAE onto FPGA.

Acknowledgements & References

Thanks to everyone in the HLS4ML community, Hardware team and the Neural teams.

This work is supported by National Science Foundation under OAC-2117997