Edge AI-optimized ASIC Prototype for Sleep Stage Detection

Tristan Robitaille (1006343397)

1 Background

As reported by Chaput *et al.* [1], insomnia impacts around 24% of Canadians adults. Sleep stage detection (SSD), followed by neuromodulation, has been recently found by Yoon [2] to be a promising treatment against insomnia. However, state-of-art SSD techniques such as polysomnography are cumbersome to use (at least 19 sensors are required, as explained by Levin and Chauvel [3]), require clinical supervision and do not offer neuromodulation. Thus, there is a need to develop a self-contained brain-machine interface (BMI) device for SSD and neuromodulation. To maximize treatment potential, the device should be as small and portable as possible.

2 Objectives

To maintain a reasonable scope, this thesis project focusses on prototyping the compute software and hardware for sleep stage detection. Multiple authors [4]–[6] have published high-accuracy results using a deep learning approach to SSD, and have done so with significantly fewer sensors than polysomnography. However, these AI models run on standard computers as software frameworks and are thus unsuitable for a lightweight integrated solution. Google sells small custom AI-accelerators (such as the Coral Edge TPU) that could run these AI models, but they still consume too much power (2W, [7]) and do not readily integrate with custom neuromodulation hardware. Thus, the objective of this thesis is to demonstrate the superiority of a hardware implementation of an AI model compared to state-of-the-art SSD methods. The proposed solution should match or exceed the accuracy of traditional polysomnography and published models in the literature while consuming less power than the currently available commercial hardware solutions.

3 Methodology

To achieve the aforementioned objectives, a three-step plan is proposed. First, a prototype Python model will be developed, trained, evaluated and optimized until its accuracy surpasses existing solutions and size cannot be reduced further. Then, this model will be translated into C. Finally, the model and associated system architecture will be designed, implemented on FPGA, benchmarked and continuously optimized for low power consumption and area. Most of the effort of this project will be spent on architectural optimizations. This high- to low-level approach will provide gradually increasing exposure to the computations involved and opportunities for design improvements while still providing relatively rapid debugging and optimization cycles. This final model will be benchmarked against published research and an equivalent model running on Google Coral.

References

[1] Jean-Philippe Chaput, Jessica Yau, Deepa P. Rao, et al. "Prevalence of insomnia for Canadians aged 6 to 79". In: *Health Reports* 29.12 (2018).

[2] Ho-Kyoung Yoon. "Neuromodulation for Insomnia Management". In: *Sleep Medicine and Psychophysiology* 28.1 (2021), pp. 2–5.

[3] Jessica Vensel Rundo and Ralph Downey. "Chapter 25 - Polysomnography". In: *Clinical Neurophysiology: Basis and Technical Aspects*. Ed. by Kerry H. Levin and Patrick Chauvel. Vol. 160. Handbook of Clinical Neurology. Elsevier, 2019, pp. 381–392. DOI: https://doi.org/10.1016/B978-0-444-64032-1.00025-4. URL: https://www.sciencedirect.com/science/article/pii/B9780444640321000254.

[4] Micheal Dutt, Surender Redhu, Morten Goodwin, et al. "SleepXAI: An explainable deep learning approach for multi-class sleep stage identification". In: *Applied Intelligence* 53.13 (2023), pp. 16830–16843.

[5] Mingyu Fu, Yitian Wang, Zixin Chen, et al. "Deep learning in automatic sleep staging with a single channel electroencephalography". In: *Frontiers in Physiology* 12 (2021), p. 628502.

[6] Emadeldeen Eldele, Zhenghua Chen, Chengyu Liu, et al. "An attention-based deep learning approach for sleep stage classification with single-channel EEG". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), pp. 809–818.

[7] *USB Accelerator datasheet*. Version 1.4. Coral. 2019. URL: https://coral.ai/docs/accelerator/datasheet/.