# Reviewer’s Report

On the PhD dissertation of
Zeyu Wang
entitled
“GPU-Accelerated Signal Decomposition for Efficient EEG Processing: Methods and Applications “

Dr. István Zoltán Reguly
Associate Professor
Pázmány Péter Catholic University
Faculty of Information Technology and Bionics

Introduction and Context

The dissertation developed by Zeyu Wang addresses the computationally demanding challenges posed by processing Electroencephalography (EEG) signals, focusing specifically on the performance and efficiency of signal decomposition methods using Graphics Processing Units (GPUs). This field of research is inherently multi-disciplinary and has a direct and significant impact on biomedical signal processing and computational neuroscience. The topic is highly relevant and timely, given the increasing computational demand presented by high-frequency multi-channel EEG recording devices (with up to 512 channels), as well as the size of publicly available datasets. The applications are wide ranging, from the growing interest in brain-computer interfaces and neurofeedback, through real-time monitoring applications, to understanding diseases from large amounts of historical data.
The dissertation studies signal decomposition methods on non-stationary, nonlinear signals, the most computationally expensive step in the signal processing pipeline. Appropriately, GPUs are used to implement these algorithms, which offer high throughput for these kinds of data-parallel tasks. The field of GPU processing itself is also a very current and dynamic one, particularly with the introduction of Tensor Cores and extensions to the CUDA programming model, which are utilized and evaluated by the candidate.

Summary of Contributions

The dissertation tackles the fundamental algorithmic and performance bottlenecks associated with applying advanced signal decomposition techniques to EEG data on GPUs. The candidate structures his contributions around three thesis points

1. The development and optimization of the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), allowing effective handling of non-stationary signals, and addressing the computational costs.
2. Addressing the performance bottlenecks in processing multi-channel EEG data with Multivariate Empirical Mode Decomposition (MEMD), allowing for the decomposition of multi-channel EEG data, preserving cross-channel correlations.
3. A novel approach to utilizing Tensor Cores to speed up the Independent Component Analysis (ICA) algorithm, specifically targeting efficient artifact removal in EEG processing.

Overall Assessment of the Dissertation

- Structure and Organization: the dissertation has a logical and coherent structure. It begins with the introduction of GPU parallelism and EEG (Chapter 1), then continues with the algorithmic background of the studied signal decomposition approaches (Chapter 2). The following three chapters describe the development and evaluation of each of the core contributions (Chapters 3-5). The thesis concludes with a summary (Chapter 6), then spells out the thesis points in Chapter 7.

- Clarity and Language: The dissertation is written in clear and precise English, with the appropriate use of technical terminology and abbreviations, that was easy to follow for someone working in a related field. The explanations of complex algorithms and implementations is particularly well and commendably done, with multiple very well-designed diagrams to help understanding.
- Presentation: The dissertation includes numerous well-designed figures that give a high-level overview of the algorithms and data structures in questions. These are clear and easy to read. The figures are largely consistent in style across the chapters, and can be read well.
- Length and scope: The dissertation is approx. 85 pages, which is appropriate for a thesis in this field. The scope is well-defined and sufficiently focused. The depth of technical detail is appropriate to support the author's conclusions. It is interesting to see how these innovations fit into EEG pipelines, and what novel types of analyses they might enable - even if these are not directly the results of the candidate.
- References: The bibliography is extensive, with over 190 entries, and covers relevant foundational work, specific algorithms, GPUs, and applications. The breadth of references underlines the multi-disciplinary nature of the research. The sources are reasonably up to date (mainly with respect to the original publication date of the individual contributions).

Detailed Critical Analysis

- Chapters 1-2 (Introduction and Background): The introduction effectively motivates the present work, highlighting the importance of EEG signal processing, and the associated challenges. It furthermore introduces GPU architectures, giving a brief by relevant summary of its development and key characteristics. Chapter 2 provides the theoretical background on frequency-based and spatial-based decomposition methods as they are applied to EEG signal processing. It is a clear and comprehensive overview. There is some context provided on how these signal decomposition methods fit into an end-to-end EEG processing pipeline to better stress just how much of a bottleneck these algorithms are.
- Chapter 3 (GPU ICEEMDAN): this chapter describes the first of three key contributions, the GPU implementation of the ICEEMDAN algorithm. The candidate presents a thorough description of the related work (highlighting key deficiencies), implementation details and performance results on various GPUs. The design choices for parallelization and implementation strategies are well thought-out. The performance bottlenecks are methodically explored and addressed, by e.g. incorporating CUDA graphs to mitigate launch latencies. The comparison in performance and numerical validity to MATLAB and libeemd is appropriate given the use cases, and demonstrated impressive speedups.
- Chapter 4 (GPU MEMD): this chapter addresses the challenge of efficiently processing multi-channel EEG data, while preserving inter-channel correlations. The design and implementation details are well considered and detailed, taking advantage of the parallelism in processing multiple direction vectors. The algorithms are particularly well explained with the help of block diagrams and figures. Once again performance is contrasted against MATLAB, but here prior implementations from literature are also evaluated – and significantly improved upon. There is detailed analysis across different channel counts, sample sizes and GPU architectures – demonstrating depth of the

study. To demonstrate numerical validation, a number of figures are presented and used to explain how well the results agree. Overall, all kernels developed perform very well, being reasonably close to the roofline. Once again, the tridiagonal solver appears to be a bottleneck.

- Chapter 5 (GPU ICA): this chapter presents the Tensor Core accelerated implementation of the Independent Component Analysis algorithm. The novelty of the approach is especially clear, with the recent introduction of this hardware. The explanation of how these cores work, how the WMMA API can be utilized and how the deep memory hierarchy is exploited for the implementation is very clear and well-illustrated. The results show significant speedups over Matlab (3-43x). Given the lower precision, the numerical accuracy is more thoroughly studied using a similarity index metric. In the results section, unexpectedly high channel counts are used – these could use some explanation. For Fig 5-12, it is somewhat surprising that the speedups are high at 2176 channels, but moderate at both fewer and more channels.
- Conclusions and Summary (Chapters 6-7): these chapters effectively summarize the work and results achieved by the candidate. The Summary explicitly and clearly spells out the thesis points.

Evaluation of Thesis Points

Thesis I: GPU implementation of ICEEMDAN – describing the massively parallel implementation of the ICEEMDAN algorithm significantly redacting computational time. The thesis point is clearly formulated, and it is supported by an appropriate literature review. The results were published at two conferences and a Q1 journal, underlining its novelty. I accept Thesis Point I as a new scientific result.
Thesis II: GPU implementation of MEMD – the candidate developed an efficient implementation of the MEMD algorithm, validated its numerical accuracy, and demonstrated its significant performance compared to the state of the art in Matlab, as well as other GPU codes. The results were published at a conference and a Q2 journal. I accept Thesis Point II as a new scientific result.
Thesis III: GPU implementation of ICA – describing a novel Tensor Core-enabled implementation of ICA, delivering unprecedented computational throughput. The candidate presents the details of how this new hardware can be exploited and describes the innovations behind mapping this algorithm to the GPU. The results of this work were presented at a conference. I accept Thesis Point III as a new scientific result.

Evaluation of Scientific Publications

The candidate has 10 publications, 7 directly associated the PhD theses. Notably, the candidate is the first author on two high-quality international journal publications and is a co-author on three further journal publications. These journals, as well as the conferences are appropriate and respected in this field. This publication record adequately supports the dissertation's claim to original research and demonstrate the candidate's ability to describe and disseminate the results of his work in peer-reviewed venues.

Questions

1. In multiple occasions, the performance of the tridiagonal solver appears to be a bottleneck, and as demonstrated it does not utilize the device well (far from the roofline). Is this an inherent limitation of the way tridiagonal equations can be solved in parallel, or how cuSPARSE is used in your case?

2. You refer to ICA-dependent pipelines as semi-automatic. Do you think there are sufficient samples available to train a machine learning model to perform artifact identification?

Conclusion

The dissertation of Zeyu Wang presents significant and original research on the GPU implementation of key EEG signal processing algorithms. The work has systematically studied the ICEEMDAN, MEMD, and ICA decompositions methods, their stat of the art, and gave novel implementations that produce numerically validated results at orders of magnitude higher performance, thereby opening the door to significantly more complex signal processing pipelines.
The candidate has demonstrated a strong ability to design and implement complex parallel algorithms, as well as to effectively communicate highly convoluted concepts and their realizations.
Based on this, I can confidently recommend the acceptance of these thesis points, and awarding the PhD degree.

Budapest, 2025. 06. 28

István Zoltán Reguly