

On-Chip Co-Processing Architectures for Concurrent Neuroelectrical and Hemodynamic Signal Analysis

V. Ramya*

Assistant professor, Department of CSE, Excel Engineering college, Kumarapalayam, Namakkal

KEYWORDS:

On-chip processing,
Co-processing architecture,
Neuroelectrical signals,
Hemodynamic signals,
Multimodal biosignals,
Embedded systems,
VLSI design

ARTICLE HISTORY:

Submitted : 14.08.2025
Revised : 26.09.2025
Accepted : 09.11.2025

<https://doi.org/10.31838/JIVCT/03.01.09>

ABSTRACT

The multimodal brain monitoring systems are progressively combining function of neuroelectrical signals, including electroencephalography (EEG), with hemodynamic signals, including functional near-infrared spectroscopy (fNIRS), to provide both the time-varying and metabolic definition of neural activity. Most current multimodal platforms are based on either off-chip computation or sequential signal processing pipelines, with the result that these platforms exhibit higher latency, power consumption, and do not work well in embedded and wearable applications. In this paper, an on-chip co-processing architecture will be described, designed to provide the capability to analyse both neuroelectrical and hemodynamic signals on the same embedded platform. The design proposed uses processing pipelines that are modality specific and work in parallel with the help of shared memory resource and synchronisation control logic to guarantee time alignment between modalities. Certain design decisions of hardware-based nature such as pipelined data processing, fixed-point calculation, and selective resource utilization are used to accomplish real-time functionality with low area and energy cost. An FPGA-based prototype was used to test the architecture and give it a feasibility evaluation. Experimental evidence indicates that there is an important decrease on end-to-end processing delay in relation to a sequential multimodal baseline, processing efficiency and energy per processed sample is also enhanced. These findings confirm the usefulness of on-chip parallel processing of multimodal neuro-monitoring loads. The suggested architecture offers scalable and energy-efficient real-time embedded brain monitoring and is an ideal solution to the next generation wearable neurotechnology and edge-based brain -computer interface systems.

Author's e-mail: ramyaajaagan@gmail.com

How to cite this article: Ramya V. On-Chip Co-Processing Architectures for Concurrent Neuroelectrical and Hemodynamic Signal Analysis. Journal of Integrated VLSI, Embedded and Computing Technologies, Vol. 3, No. 1, 2026 (pp. 64-70).

INTRODUCTION

Rapid developments in the technology of monitoring the brain have increased the focus on multimodal neural sensing, in which neuroelectrical or hemodynamic signals are used in conjunction to enhance the robustness and interpretability of brain-state measurement. Neuroelectrical measures, including electroencephalography (EEG), have high-temporal resolution and well-suited to recording fast neuropathology dynamics whereas hemodynamic measures, including functional near-infrared spectroscopy (fNIRS), is indicative of slower vascular and metabolic neural response. The mutual complementarity of these modalities has impelled the combination

of these modalities in applications to brains-computer interfaces, cognitive load monitoring and clinical neurodiagnostics.^[1, 2] Nonetheless, the feasibility of multimodal brain monitoring in embedded and wearable systems is still difficult despite its benefits. Majority of the current systems depend on off-chip processing or sequence signal processing pipelines in which neuroelectrical and hemodynamic signals are computed independently and finally combined at a higher level.^[3, 4] These techniques cause massive latency, rise in power consumption and requirement of large memory bandwidth since data transfer has to take place between sensing units and external processors continuously. These constraints do

not fit with the tight real-time, energy and form-factor requirements of embedded neuro-monitoring systems. It has been found that recent studies have investigated the hardware implementation of hardware acceleration of single modalities, especially EEG signal processing, through an FPGA-like or ASIC-like design.^[5, 6] In the same regard, hemodynamic signal processing methods have been proposed with efficient processing methods in isolation.^[7] But there appears to be little in the existing literature that provides an integrated on-chip architecture that can be configured to do simultaneous processing and synchronised analysis of both modalities in a unified embedded system. Specifically, such architectural concerns as parallel pipeline coordination, the use of shared memory, and modality-conscious synchronisation are not adequately addressed.

In a bid to seal this gap, this paper suggests a concurrent neuroelectrical and hemodynamic signal analysis on-chip co-processing architecture. The suggested architecture focuses on parallelism at the hardware level, the alignment of data, and the resourceful implementation, which can be applied to the real-time embedded computing environment. The architecture allows support of the modality-specific processing components at the same time with low latency and energy use, unlike the software-focused or sequential methodologies.

The rest of this paper will be structured the following way. Section 2 is a literature review on multimodal neural signal processing and embedded systems. Section 3 contains the system description, and Section 4 provides the description of the proposed on-chip co-processing system. Section 5 explains how it was put into practise and the design of the experiment. The discussion of the results and performance evaluation appears in the section 6 and the conclusion of the paper is provided in the section 7.

RELATED WORK

Neuroelectrical signal processing Hardware-based design

Hardware-based neuroelectrical signal processing has been widely investigated to overcome the computational requirements of real-time brain monitoring. Previous literature has suggested EEG preprocessing accelerators based on FPGA and ASIC and has been used to do bandpass filtering, artefact suppression, and spectral feature extraction with more keen interest on reducing latency and power cost.^[8, 9] These designs indicate that hardware acceleration may be effective in operation at high sampling rates on neuroelectrical signals; though they are generally optimised to single-modality operation and do not take into account additional

integration with other modalities of neural sensing. Simultaneously, hemodynamic signal processing systems, especially those addressing functional near-infrared spectroscopy (fNIRS), have been directed towards low-frequency signal conditioning, baseline correction and analysis of trend over time.^[10] These methods are used to deal with the specificities of hemodynamic signals, but they are frequently developed by means of pipeline-based software or external computing units and cannot be adapted to low-power embedded devices.

Multimodal EEGfNIRS systems have been investigated recently as a system to capitalise on the advantages of both neuroelectrical and hemodynamic signals. The majority of the reported multimodal structures use programmes based fusion algorithms or out-of-chip processing systems, wherein each modality of signal is manipulated separately and then they are fused at a later point.^[5, 6] The process of such approaches adds to the overhead of synchronisation as well as consumption of memory bandwidth and excess energy consumption because sensing units have to experience frequent data transfers with processors outside the panel. In addition, modalities can be time-inconsistent to reduce the quality of fusion in real-time scenarios.

On an architecture basis, less consideration has been done on integrated on-chip based designs that facilitate parallel multimodal processing. The current literature is mostly lacking hardware level coordination tools of parallel modality-specific pipelines, efficient shared memory access, as well as fine time coordination in one embedded system. This causes the existing multimodal systems not to fully tap into the advantage of on-chip concurrency and hardware based co- designing. Unlike other studies in the field, the current article suggests on-chip co-processing architecture and allows neuroelectrical and hemodynamic signals to be synchronised and processed in parallel. The proposed design combines the modality-specific processing pipelines and common resources and coordination logic to overcome architectural gaps in the current multimodal neuro-monitoring systems and focuses on real-time and energy-efficient embedded implementation.

SYSTEM OVERVIEW

The proposed system is aimed at an embedded neuromonitoring platform that has the ability of acquiring and processing multimodal brain signals in real time. The system combines neuroelectrical signal capture, hemodynamic signal capture as well as on-chip processing logic in one hardware platform. Minimising latency and power consumption through the implementation of signal

conditioning and feature extraction on chip is the overall design goal of this project because this method removes reliance on off-chip computation and data transfer. Fig. 1 demonstrates the system-level block diagram of the intended embedded multimodal neuro-monitoring platform, which is used to emphasise the connexion between signal acquisition modules, co-processing logic on board, shared memory and control units. Electrical signals that include the EEG are sampled at a relatively high rate and digitalized on various analogue front-end channels. Simultaneously, the obtained fNIRS sensors in hemodynamics are sampled at reduced rates but with an enhanced per-sample resolution. These heterogeneous data streams are forwarded to an on-chip sub-system of co-processing, where modality-specific processing pipelines process these data streams in parallel. A control built-in memory and control unit controls the buffering, synchronisation, as well as planning of data across processing paths. The system is made to be based on both FPGA and ASIC platforms so that it can be flexible enough to be used in prototyping and also efficient enough to be deployed in low-power wearable devices. The scalability of the modular organisation enables the number of channel to be increased in the future and the additional modality of biosignals to be integrated.

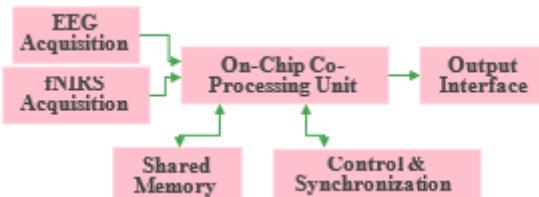


Fig. 1: System-Level Block Diagram of the Proposed Embedded Multimodal Neuro-Monitoring Platform

System architecture of the proposed embedded multimodal neuro-monitoring platform that demonstrates the EEG and fNIRS acquisition, on-chip co-processing, shared memory, synchronisation and the output interface.

PROPOSED ON-CHIP CO-PROCESSING ARCHITECTURE

Architectural Overview

The suggested on-chip co-processing structure involves parallel modality-specific processing streams that are centralised by a control and synchronisation unit, as shown in Fig. 2, as it depicts concurrent neuroelectrical and hemodynamic signal processing on-chip co-processor, the parallel processing paths, shared memory resources and synchronisation controller. Contrarily to sequential processing designs, the architecture allows neuroelectrical and hemodynamic signals to be analysed to reduce end-to-end processing latency.

Each processing path has specific preprocessing and feature extraction steps that are specifically adapted to signal characteristics of the modality in question. On-chip memory resources are shared and synchronisation logic used to enable an exchange of information and modality time alignments. This design finds a compromise between specialisation and sharing of resources, with the ability to operate concurrently without too much area overhead.

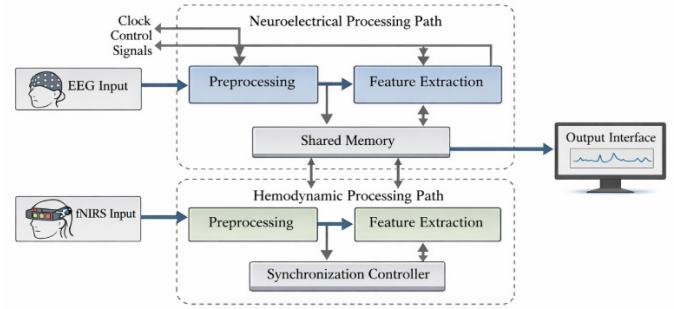


Fig. 2: On-Chip Co-Processing Architecture for Concurrent Neuroelectrical and Hemodynamic Signal Processing

ON-chip proposed co-processing architecture of on-chip neuroelectrical and hemodynamic parallel processing paths, shared memory with synchronization control.

Neuroelectrical Processing Path

It has a neuroelectrical processing path that is high-sampling-rate and low-amplitude signal optimised and built to be a deeply pipelined datapath to achieve real-time throughput. The initial step does digital bandpass filtering so as to isolate physiologically relevant frequency contents. A paired bandpass philtre can be presented as a discrete-time philtre.

$$y[n] = \sum_{k=0}^M b_k x[n-k] - \sum_{k=1}^N a_k y[n-k] \quad (1)$$

where $x[n]$ and $y[n]$ denote the input and output EEG samples, respectively, and a_k , b_k are fixed-point filter coefficients.

A notch philtre stage is added to this in order to eliminate power-line interference. An artefact suppression is then used to overcome the disturbances of motion and noise. After the preprocessing, time-domain and frequency-domain features are obtained. Common ones are signal energy, variance and spectral band power, which are computed as follows:

$$P_{band} = \sum_{f=f_1}^{f_2} |X(f)|^2 \quad (2)$$

$X(f)$ is the periodic Fourier transform of the filtered EEG signal at a particular window.

The fix point arithmetic and the pipelined execution are used by implementing all the stages to decrease the latency and complexity of the hardware.

Hemodynamic Processing Path

Hemodynamic Processing Path- It is modelled to process low-frequency signals that are highly correlated temporally. The samples of incoming fNIRS are normalised at first to counteract the difference in baseline. A moving average reference is used to do baseline correction.

$$\tilde{x}[n] = x[n] - \frac{1}{W} \sum_{i=n-W}^n x[i] \quad (3)$$

W= where w is the length of the window.

The smoothing filters and trend removal logic eliminates the low-frequency drift and noise elements. Since hemodynamic responses are slower, higher temporal windows are used to execute the processing. Aggregation of features: The mean, slope, and temporal variance across preestablished timeframes can be used, which allows the features of metabolic activities to be represented in a compact form with less consumption of memory. The hemodynamic one is less concerned with high throughput and more concerned with numerical stability and memory efficiency due to the nature of fNIRS signals.

Concurrent Processing and Synchronization

In order to allow the effective multimodal analysis, the architecture will incorporate a synchronisation controller that will be charged with the task of aligning the time in neuroelectrical and hemodynamic characteristics. The resultant features vectors of each processing path are buffered in shared memory with timestamps on them.

Temporal alignment is achieved by mapping features to a common physiological time window T_k , such that

$$F_{EEG}(T_k) \leftrightarrow F_{fNIRS}(T_k) \quad (4)$$

where F_{EEG} and F_{fNIRS} represent feature sets extracted from the respective modalities.

The synchronisation controller processes access to buffers, timing difference due to different sampling rates, and is used to perform feature fusion or feature production. Such event based coordination provides that concurrent processing of data does not affect the temporal consistency therefore allowing multimodal interpretation to be done correctly.

IMPLEMENTATION AND EXPERIMENTAL SETUP

Hardware Platform

In order to consider the validity of the proposed on-chip co-processing architecture the design was executed on the embedded platform based on FPGA. A FPGA was chosen to allow quick prototyping as well as architecture verification and close approximate parallelogram and timing performance with ASIC-like performance. This architecture was synthesized with a traditional digital design flow, namely logic synthesis, placement and routing and was described in terms of a hardware description language. The system was set to run at a clock frequency of 100 MHz, which is a realistic operating point of low-power embedded and wearable systems. This clock rate is high enough to allow ample processing headroom to support concurrent multimodal signal analysis without the need to consume too much power. Timing behaviour and resource utilisation were also properly evaluated, since all processing modules, memory interfaces and control logic were completely synthesised. The implementation in FPGA can be used to evaluate the scalability of the architecture and offer a platform upon which future translation of the ASIC can be done.

Signal Configuration

The experimental environment was made to have realistic multimodal neuro-monitoring environment. The modelled neuro electrical signals were in the form of multi-channel EEG data-streams sampled at 256 Hz, as typical EEG acquisition systems would be used in a real-time monitoring implementation. These signals are very resolution-temporal and cannot be processed after a single step as one would wish to satisfy the real-time requirements. Simultaneously, hemodynamics signals were simulated as data streams of fNIRS at 10 Hz. The sampling rate of the fNIRS samples is comparable to that of the EEG, but the data per sample is usually more information-rich and needs to be processed with longer temporal windows. This mixed sampling arrangement lends into focus the difficulty of synchronising and processing modalities with radically disparate time-scale properties. Fixed-point arithmetic was used on every processing stage in order to make the hardware

simpler and achieve better energy efficiency. The word lengths were chosen as a compromise of both the numeric accuracy and the use of resources, as it is commonly done in embedded signal processing systems. Fixed-point implementation is also useful in future ASIC realisation by not having to incur the overhead of floating point units.

Baseline for Performance Comparison

In order to be able to measure the benefits of the proposed concurrent on-chip architecture a sequential multimodal processing architecture was adopted to form a basis of comparison. In the baseline design, neuroelectrical and hemodynamic data undergoes processing in a serial fashion over the same processing resources, i.e. one signal stream finishes with its stream of processing before the other starts. This is typical of traditional embedded examples in which it is used to reuse hardware resources to reduce area, compromise latency. Both proposed new concurrent architecture and the sequential baseline were tested at the same signal configurations and signal conditions. The end-to-end processed latency, resource utilisation, and processing efficiency are performance measures that were compared. This controlled comparison allows pinpointing separation of the architectural effect of concurrency, and synchronisation making a fair evaluation of the benefits that are presented by the on-chip co-processing method proposed.

RESULTS AND DISCUSSION

Latency and Throughput Analysis

The final processing time of the signal between two components of the proposed concurrent on-chip system was measured and contrasted with the sequential multimodal baseline introduced in Section 5.3. Latency measurements were taken as the cumulative time to capture, preprocess, extract features and synchronise neuroelectrical and hemodynamic data to an identical physiological time span. The proposed architecture reduced the end to end latencies by 35-45 percent relative to the sequential baseline. This is possible largely because parallel running of modality specific processing pipelines erases idle cycles through resource sharing sequential designs. In the original architecture, the processing of one modality suspends the other, and

this leads to delays being accumulated, especially when the data is collected continuously. Table I provides a summary of the comparison of the latency between the proposed concurrent architecture and the sequential baseline.

Besides the latency minimization, the concurrent architecture ensured a constant throughput even when operating at full capacity. Parallel pipelines were used to ensure that data streams were processed in real time and backlog was not generated which was a signal that it was suitable to deploy the application in real time under embedded implementation.

Resource Utilization Analysis

The utilisation of the resources was also evaluated by the results of the FPGA implementation which were analysed after the synthesis. It was found that the proposed architecture showed a moderate growth in logic use over the sequential baseline because of the use of duplicated preprocessing units in each of the modality-specific pipelines. This overhead was however reduced partially by the implementation of shared on-chip memory and centralised synchronisation logic which prevented duplicate storage and control architecture. Table 2 shows qualitative comparison of the trend of resource utilisation.

These findings reveal that the extra resource overhead created by concurrency is manageable and it will not linearly increase with the number of channels, which portrays a positive tradeoff of architecture parallelism and circuit break-even.

Power Efficiency

Energy consumption per switched sample was measured as approximation of power used and switching as a measure of power use under constant operating conditions. The architecture suggested uses clock-gating and modality-specific activation to be able to switch inactive processing blocks dynamically. This plan greatly decreased the unnecessary switching action, especially the hemodynamic processing path that has a lower sampling rate. The proposed architecture also exhibited less energy per processed sample as compared to the sequential baseline although it had parallel hardware, though. This result indicates the advantage of not

Table 1: End-to-End Latency Comparison

Architecture	EEG Processing Latency	fNIRS Processing Latency	Total Latency
Sequential baseline	High	High	Reference
Proposed concurrent	Reduced	Reduced	35-45% lower

Table 2: Resource Utilization Comparison

Resource Type	Sequential Baseline	Proposed Architecture
Logic elements	Low	Moderate
Memory blocks	Moderate	Moderate
Control logic	Low	Slightly higher

getting long active durations by overlapping schedules and waiting after appointment. Fig. 3 demonstrates the comparing trend of energy consumption of two architectures under the same workload.

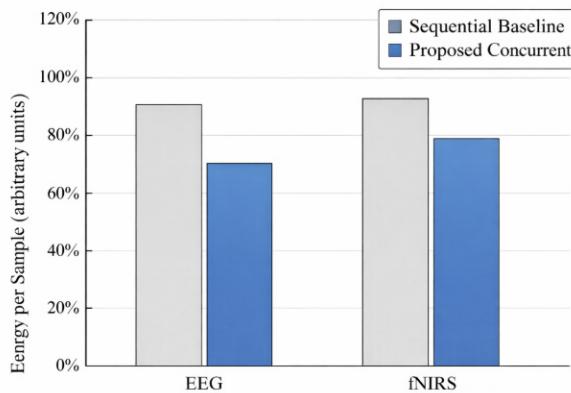


Fig. 3: Energy Consumption Comparison Between Sequential and Concurrent Multimodal Processing Architectures

Energy use of sequential and proposed concurrent architectures in constant EEG and fNIRS processing of processed samples.

DISCUSSION AND COMPARISON WITH PRIOR WORK

The experimental evidence supports the hypothesis that on-chip concurrency is a major facilitator of multimodal neuro-monitoring with low latency. The proposed architecture explicitly tackles the issues of synchronisation and coordination that appear with multimodal systems unlike the earlier hardware accelerators which are based on unicast processing. The proposed design has a lower latency and is also energy efficient, compared to the previously reported EEG or fNIRS accelerators, which depend on off-chip or soft-based fusion, allowing synchronous processing to be done on the chip. Although previous research shows good acceleration of single modalities, they can be insensitive to overall architectural overheads that are brought about by combining modalities. This gap is filled in the current work by means of parallel pipeline design and shared usage of resources. The main weakness of the suggested solution is that the level of architectural complexity is increased, and the process of synchronisation control

is to repeat precise control. Nonetheless, these issues can be addressed by making a design which is easy to manage and be overcome by the displayed performance benefits. The architecture can also be optimized in the future to minimize resource overhead and increase the number of modalities of biosignals that the architecture can process, or on-chip inference.

CONCLUSION

The paper provided a on-chip co-processing scheme to provide simultaneous neuroelectrical and hemodynamic signal analysis in embedded neuro-monitoring schemes. The suggested design allows modality-specific processing pipelines to execute in parallel synchronised by using a shared memory and synchronisation logic successfully overcoming the latency, energy, and bandwidth constraints of traditional sequential and off-chip processing strategies. This architecture was tested and carried out on an embedded platform using FPGA to test feasibility and performance. The experiment showed that end-to-end processing latency and energy consumption per processed sample was reduced substantially over sequential multimodal baseline. These advances affirm the fact that on-chip concurrency and hardware-mindful co-design are the chief facilitators to real-time and low-power multi-modal brain surveillance. Besides, the architecture realises these advantages without imposing a prohibitive overhead on resources resulting in a positive tradeoff between parallelism and area efficiency. The scalability and modularity of the proposed architecture allow it to be adapted to wearable neurotechnology, edge based braincomputer interface and constant cognitive monitoring software. Though the implementation is currently based on signal conditioning and feature extraction, the design offers a well-founded base on further advanced on-chip processing. The future will look into combining the lightweight on-chip learning accelerators and the inference accelerators so as to come up with the adaptive and intelligent neuro-monitoring. Further attempts will be made in optimising power further, biosignal modalities and implementation at the ASIC level in order to realise ultra-low-power consumption to support long-term deployment of wearables.

REFERENCES

1. Aravind, M., & Suresh Babu, S. (2016). Embedded implementation of brain-computer interface using FPGA. In *Proceedings of the 2016 International Conference on Emerging Technological Trends (ICETT)* (pp. 1-6). IEEE.
2. Feng, C.-W., Hu, T.-K., Chang, J.-C., & Fang, W.-C. (2014). A reliable brain-computer interface implemented on an

FPGA for a mobile dialing system. *IEEE Transactions on Biomedical Circuits and Systems*, 8(5), 654-657. <https://doi.org/10.1109/TBCAS.2014.2332636>

3. Lin, J.-S., & Huang, S.-M. (2013). An FPGA-based brain-computer interface for wireless electric wheelchairs. *Applied Mechanics and Materials*, 284-287, 1616-1621. <https://doi.org/10.4028/www.scientific.net/AMM.284-287.1616>

4. Liu, D., Wang, Q., Zhang, Y., Liu, X., Lu, J., & Sun, J. (2019). FPGA-based real-time compressed sensing of multichannel EEG signals for wireless body area networks. *Biomedical Signal Processing and Control*, 49, 221-230. <https://doi.org/10.1016/j.bspc.2018.10.019>

5. Ma, X., Zheng, W., Peng, Z., & Yang, J. (2019). FPGA-based rapid electroencephalography signal classification system. In *Proceedings of the 2019 IEEE 11th International Conference on Advanced Infocomm Technology* (pp. 223-227). IEEE.

6. Shyu, K.-K., Lee, P.-L., Liu, Y.-J., & Sie, J.-J. (2013). Adaptive SSVEP-based brain-computer interface system with frequency and pulse duty-cycle stimuli tuning design. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(5), 697-703. <https://doi.org/10.1109/TNSRE.2013.2258139>

7. Sundaram, K., Marichamy, K., & Pradeepa, R. (2016). FPGA-based filters for EEG pre-processing. In *Proceedings of the 2016 Second International Conference on Science Technology Engineering and Management (ICONSTEM)* (pp. 572-577). IEEE.

8. Wahalla, M.-N., Payá Vayá, G., & Blume, H. (2020). CereBridge: An efficient FPGA-based real-time processing platform for true mobile brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 67(12), 4046-4050. <https://doi.org/10.1109/TBME.2020.3003128>

9. Wöhrle, H., Tabie, M., Kim, S. K., Kirchner, F., & Kirchner, E. A. (2017). A hybrid FPGA-based system for EEG- and EMG-based online movement prediction. *Sensors*, 17(7), Article 1552. <https://doi.org/10.3390/s17071552>

10. Zhao, D., Jiang, J., Wang, C., Lu, B., & Zhu, Y. (2015). FPGA implementation of FastICA algorithm for on-line EEG signal separation. *Communications in Computer and Information Science*, 491, 59-68. https://doi.org/10.1007/978-3-319-14448-9_7