

Parallel EEG processing on GPULab, a GPU-enabled container-based distributed system in the SLICES EU research infrastructure

May 2023

Zeyu Wang, Iffah Syafiqah binti Suhaili, Zoltan Juhasz

University of Pannonia Veszprem, Hungary

Content outline

1. Background

- 1.1 About the GPULab
- 1.2 About EEG processing
- 1.3 HPC requirements for EEG processing

2. The usage of GPULab

- 2.1 Hardware infrastructure
- 2.2 Access modes
- 2.3 Our workflow on GPULab

3. Multi-GPU access test

- 3.1 Unified Memory
- 3.2 Unified Virtual Address
- 3.3 Peer-to-peer access

4. Future works



1.1 About the SLICES

Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies (SLICES)

SLICES-RI (Research Infrastructure)
SLICES-DS (Design Study)
SLICES-SC (Starting Community)
SLICES-PP (Preparatory Phase)

SLICES-SC (Starting Community)

- 1. Consiglio Nazionale Delle Ricerche (IT)
- 2. SZTAKI (HU)

3. Imec (BE)

- 4. University of Thessaly (GR)
- 5. Sorbonne Universite (FR)
- 6. French Institute for Research in Computer Science (FR)
- 7. Eurecom (FR)
- 8. University of Oulu (FI)
- 9. Poznan Supercomputing and Networking Center (PL)
- 10. IMDEA Networks Institute (ES)
- 11. Cosmote (GR)
- 12. Technische Universitat Munchen (DE)



imec iLab.t testbed

Virtual Wall



Perform large networking and cloud experiments

Portable testbed



A testbed which can be deployed in the wild

W-iLab.t testbeds



Wireless Testbed and Officelab

CityLab testbed



City of Things smart cities FIRE testbed



HomeLab



HomeLab test environment

GPULab architecture



GPULab is a distributed system for running jobs in GPU-enabled Docker-containers.



1.2 About EEG processing

- The brain is the most complex system known to human
- EEG contains physiological and pathological information of brain activity.





Time

- Time-frequency analysis
- Connectivity analysis
- Source localization analysis
- Signal decomposition analysis
- Resting state and task-related analysis

....

1.3 HPC requirements for EEG processing

High computational workload factors:

- High temporal resolution
 - 1024Hz, 2048Hz
- High electrode density
 - 128, 256 channels
- Long record
 - Minutes to half hour
- Large number of subject
 - Start from 20 subjects usually
- MATLAB based processing script
 - Low execution efficiency
- Complex processing algorithms
 - ICA EMD

Compromise solutions:

- Down sampling
 - 1024Hz, 2048Hz → 256Hz, 128Hz
- Reduce spatial resolution
 - 128, 256 → 32, 16 channels
- Segmentation
 - Minutes \rightarrow epoch in seconds
- Limited number of subjects
 - 20 \rightarrow less than 10
- C based processing program
 - Low \rightarrow high efficiency
- Algorithm optimization
 - Time/space complexity: high \rightarrow low

2.1 Hardware infrastructure of GPULab

- Several GPU clusters with different types of GPU cards
- We take Cluster 6 as an example:
 - NVIDIA HGX-2 with 16 Tesla V100 GPUs, 96 2.7Ghz vCPU cores with 1.5TB RAM
 - NVLink switch connection fabric





JupyterHub based interactive mode

÷	- C ふ む https://jupyter	rhub.ilabt.imec.be/u	user/rcueoost@ilabt.imec.be/lab/tree/project_ghent
\bigcirc	File Edit View Run Kernel Git	Tabs Settings	Help
	+ 🗈 🛨 C 🚸		Ircueoost@3a19266c457f: /pr × +
0	Filter files by name	Q	(base) rcueoost@3a19266c457f:/project_ghent\$ nvidia-smi Fri May 12 12:11:51 2023
	/ project_ghent / Storage space		++ NVIDIA-SMI 515.86.01 Driver Version: 515.86.01 CUDA Version: 11.7
٩	Name dataset	Last Modified 7 months ago	GPU Name Persistence-M Bus-Id Disp.A Volatile Uncorr. ECC Fan Temp Perf Pwr:Usage/Cap Memory-Usage GPU-Util Compute M.
≣	 eegProcessingPipeline eigenValueDecomposition 	3 months ago a month ago	MIG M. ===================================
*	 GPU_Pipeline_test headFiles 	4 months ago 3 months ago	++
	 testProgramForMEMD testProgramForNvshmem testProgramForP2P 	3 months ago 24 days ago 3 months ago	+
	 benchmark benchmark.cu 	3 months ago 3 months ago	=====================================
	 benchmarkFFTandForward benchmarkFFTandForward.cu 	3 months ago 3 months ago	(base) rcueoost@3a19266c457f:/project_ghent\$
	benchmarkMultiGPU benchmarkMultiGPU.cu	3 months ago 3 months ago	
	 ceemdanReducedEpochs_dev.cu CUDA_CEEMDAN eeqSampleDataCH1_244032 bin 	2 months ago 2 months ago 2 months ago	Interactive shell
	multi_test.cu	4 months ago	

CLI batch mode



2.3 Our workflow with GPULab

Local development environment



GPULab clusters

3 Multi-GPU access test

Case 1

Case 2 & 3



Naïve pattern

Unified Memory & Unified Virtual Address

3 Multi-GPU access test



Case 4

- Input signal is processed by 2 GPUs
- Different lengths with same SR (2048 Hz)
- Kernel: multiplication and addition
- Hardware: HGX-2
- 50 repetitions

	512s	1024s
Case1	4.28	69.99
Case2	7.47	43.14
Case3	0.93	6.41
Case4	0.79	5.79

Execution time in microseconds

Peer-to-Peer access pattern

4 Future works

- Several single-GPU EEG processing algorithms have bee developed
 - Empirical Mode Decomposition (EMD) and its several variants
 - Independent Component Analysis (ICA)
 - Source localization
 - •
- Single-GPU algorithm optimization
- Multi-GPU-based EEG processing algorithm development
- Extend the EEG pre-processing pipeline to multi-GPU system
 - filtering, mean removal, averaging, artifacts removal, statistical analysis
- Research on the scalability of multi-GPU signal processing algorithm