Jae-Won **Chung**

PHONENUMBER | EMAILADDRESS | jaewonchung.me | jaywonchung | jae-won-chung-cs

Summary

I am a third year PhD candidate in CSE at the University of Michigan, working with Professor Mosharaf Chowdhury. I build efficient software systems for deep learning, with a recent focus on the efficient management of not only time, but also energy. I lead the ML Energy initiative.

Education

University of Michigan — *Ann Arbor, MI, USA*

PH.D. CANDIDATE IN COMPUTER SCIENCE AND ENGINEERING — *Sep 2021 - present*

University of Michigan — *Ann Arbor, MI, USA*

M.S. IN COMPUTER SCIENCE AND ENGINEERING — *Sep 2021 - Apr 2023*

Seoul National University — *Seoul, South Korea*

B.S. IN ELECTRICAL AND COMPUTER ENGINEERING — *Mar 2015 - Aug 2021*

- GPA: 4.04/4.3 (overall) 4.15/4.3 (major), Summa Cum Laude
- Period includes two years of military service.

Publications

- **Perseus: Removing Energy Bloat from Large Model Training**, Jae-Won Chung, Yile Gu, Insu Jang, Luoxi Meng, Nikhil Bansal, Mosharaf Chowdhury, Preprint, 2023
- **Chasing Low-Carbon Electricity for Practical and Sustainable DNN Training**, Zhenning Yang, Luoxi Meng, Jae-Won Chung, Mosharaf Chowdhury, **ICLR Workshop**: Tackling Climate Change with Machine Learning, 2023
- **Zeus: Understanding and Optimizing GPU Energy Consumption of DNN Training**, Jie You*, Jae-Won Chung*, Mosharaf Chowdhury, Symposium on Networked Systems Design and Implementation (**NSDI**), 2023 (Acceptance rate = 18.38%)
- **ShadowTutor: Distributed Partial Distillation for Mobile Video DNN Inference**, Jae-Won Chung, Jae-Yun Kim, Soo-Mook Moon, International Conference on Parallel Processing (**ICPP**), 2020 (Acceptance rate = 28.99%)

*** Equal contribution**

Research Experience

Energy-Efficient Systems for Machine Learning — *SymbioticLab, UMich*

ADVISOR: MOSHARAF CHOWDHURY — *Sep 2021 - Present*

- *Zeus*: Discovered the trade-off between DNN training time and energy. Designed a Multi-Armed Bandit solution for time-energy optimization.
- *Perseus*: A system for energy-efficient large model training. Cuts up to 30% energy without slowdown. Open-sourced as part of Zeus.
- *ML.ENERGY Leaderboard & Colosseum*: The first systematic benchmark and interactive comparison service for LLM energy consumption.

Software Systems for Machine Learning — *Software Platform Lab, SNU*

ADVISOR: BYUNG-GON CHUN — *Apr 2020 - Jun 2021*

- *Crane*: A GPU cluster manager for AutoML workloads. Built a Kubernetes backend that scaled to 288 GPUs. Contributed core features such as automatic bootstrapping on Docker Swarm and Kubernetes and log streaming through the EFK (Elasticsearch - Fluent Bit - Kibana) stack.

Online Model Specialization for Edge Video DNN Inference — *Virtual Machine and Optimization Lab, SNU*

ADVISOR: SOO-MOOK MOON — *Dec 2019 - Jun 2020*

- *ShadowTutor*: Knowledge distillation from the server to the edge device reduced network data transfer by 95% and increased throughput by 3x.

Few-Shot Learning with Meta-Learning — *Computer Vision Lab, SNU*

ADVISOR: KYOUNG MU LEE — *Jun 2019 - Dec 2019*

- Designed improved meta-initialization methods for Model-Agnostic Meta-Learning (MAML) with neural memory modules and convex programs.

Quantitative Susceptibility Mapping with Deep Learning — *Lab of Imaging Science and Technology, SNU*

ADVISOR: JONGHO LEE — *Jun 2019 - Aug 2019*

- Designed and implemented a full deep learning pipeline for QSM, a vision task for medical diagnostics with 3D MRI field data, including preprocessing (background removal, phase unwrapping, and patch slicing), augmentation (adding fake calcifications) and modeling (CAD-QSMNet).

Open Source Projects

- **BERT4Rec-VAE-Pytorch** (☆324 ⑂ 78), A PyTorch framework for recommendation model training, with abstract classes for **pluggable** model, dataset, and samplers. BERT4Rec and Netflix VAE models implemented.
- **Reason** (☆184 ⑂ 4), A shell for managing research papers, written in **Rust**. Supports importing papers from file and URL, attaching markdown notes, and creating an HTML book with notes. Uses `serde` to persist data in human-readable and cloud sync-friendly format.
- **Zeus** (☆117 ⑂ 16), A framework for deep learning **energy** measurement and optimization. Constantly **leading a team** of two to three student contributors. Integrates **best practices** such as full type-annotation, auto-generated code reference, Docker, Pytest, and examples.
- **Pegasus** (☆27 ⑂ 3), An SSH command runner with a focus on simplicity, written in **Rust**. Runs commands **asynchronously** using the `tokio` runtime and **streams** stdout and stderr back to the user. Battle-tested through multiple research projects and benchmarking.

Number of stars and forks are as of March 18th, 2024.

Honors & Awards

Nov 2022	**Carbon Hack '22 Second Best Solution**, Carbon-Aware DNN Training with Zeus, $25,000	*Green Software Foundation*
Jul 2021	**Kwanjeong Overseas Scholarship**, $100,000 over four years	*Kwanjeong Educational Foundation*
Mar 2019	**Kwanjeong Undergraduate Scholarship**, $20,000 over two years	*Kwanjeong Educational Foundation*

Grants & Funding

Jan 2024	**Research grant**, $20,000 for the development of the ML.ENERGY Initiative	*Salesforce*
Jan 2024	**Mozilla Technology Fund 2024**, $50,000 for the development of the Zeus project	*Mozilla*

Invited Talks

Oct 2023	**Energy-Efficient Software Systems for Machine Learning**	*Seoul National University*
Oct 2023	**Energy-Efficient Deep Learning with PyTorch and Zeus**	*PyTorch Conference*
Sep 2023	**Energy-Efficient Deep Learning with Zeus**	*Massachusetts Institute of Technology*

Service

- **Systems/Software Reading Group**, Paper reading group inside Michigan CSE, Organizer since Fall 2022

Teaching

- **Operating Systems (SNU, Spring 21)**, Lead TA, Managed Linux kernel hacking projects and led student team design reviews.
- **Computer Architecture (SNU, Fall 20)**, Peer tutor, Provided 30 hours of online lecture, **Best Tutor Award!**

Language Proficiency

Feb 2020	**TOEFL**, 120 (full credit)
Oct 2018	**TOEIC**, 900 (full credit)
Mar 2018	**GRE**, 167/170/4.5

Skills

Language	Python, Rust, CUDA, C++, Verilog, C, Bash
Framework	PyTorch, Pandas, Matplotlib, FastAPI
Methodology	Machine Learning, Deep Learning, Multi-Armed Bandit
Tool	Docker, Kubernetes, KubeFlow, LaTeX
English	TOEFL 120 (Perfect score), GRE 167/170/4.5