

Tae Jun Ham

CONTACT	Building 944, 7th floor	Voice: (+82)-10-5314-0529	
INFORMATION	1-Gwanak-ro, Gwanak-gu Seoul 08826, South Korea	E-mail: ham.taejun@gmail.com Website: https://taejunham.github.io	
CURRENT POSITION	Seoul National University , Seoul, Republic of Korea <i>Postdoctoral Researcher</i>		July. 2018 - Present
	Research Area: Software-hardware co-design for emerging applications such as big data analytics and machine learning. Supervisor: Professor Jae W. Lee This position also fulfills the mandatory military service duty required for all Korean men.		
EDUCATION	Princeton University , Princeton, NJ USA <i>School of Engineering and Applied Science</i> Ph.D in Electrical Engineering M.A in Electrical Engineering Dissertation: Data Access Optimization in Accelerator-oriented Heterogeneous Architecture through Decoupling and Memory Hierarchy Specialization Advisor: Professor Margaret Martonosi and Professor Juan Luis Aragon		Sep. 2012 - Jun. 2018
	Duke University , Durham, North Carolina USA <i>Pratt School of Engineering</i> B.S.E in Electrical and Computer Engineering (GPA : 3.95/4.00) <i>Summa Cum Laude, with Distinction in Electrical and Computer Engineering</i>		Aug. 2009 - Dec. 2011
PROFESSIONAL EXPERIENCE	Microsoft Research , Cambridge, UK <i>Graduate Research Intern</i> Research on an efficient secure memory design with near-data computation		May - Aug, 2016
	Intel Labs — Parallel Computing Lab, Santa Clara, USA <i>Graduate Technical Intern</i> Research on a custom hardware accelerator for graph analytics applications.		May - Nov, 2015
	AMD Research , Austin, USA <i>Co-op Engineer</i> Research on an efficient use of high-perf energy-efficient heterogeneous system consists of large, low memory bandwidth processors and small, high memory bandwidth processors.		Jun - Aug, 2013
	Samsung Advanced Institute of Technology , Yongin, Republic of Korea <i>Research Intern</i> Research on a GPU branch divergence problem.		Jun - Aug, 2012
	Duke University — BCL Research Group, Durham, USA <i>Research Assistant</i> Research on a heterogeneous memory system.		Jan - May, 2012
HONORS AND AWARDS	<ul style="list-style-type: none">• IEEE MICRO Top Picks (2021)• IEEE MICRO Top Picks Honorable Mention (2021)• ISPASS Best Paper Award Nominee (2020)• MICRO-49 Best Paper Award (2016)• IEEE MICRO Top Picks Honorable Mention(2016)• Facebook Graduate Fellowship Finalist (2016-2017), Facebook, Inc.• Gordon Y.S. Wu Fellowship (2012-2017), Princeton University• Samsung Scholarship (2012-2017), Scholarship that supports up to \$50,000 per year• Summa Cum Laude, Duke University		

PUBLICATIONS **[ISCA '21] ELSA: Hardware-Software Co-design for Efficient, Lightweight Self-Attention Mechanism in Neural Networks**

Tae Jun Ham^{*}, Yejin Lee^{*}, Seong Hoon Seo, Soosung Kim, Hyunji Choi, Sung Jun Jung, Jae W. Lee

The 47th ACM/IEEE International Symposium on Computer Architecture (ISCA), 2021 [To Appear]

Acceptance Rate : 76/406 \approx 19%

^{*}Two authors contributed equally.

[ISCA '21] BOSS: Bandwidth-Optimized Search Accelerator for Storage-Class Memory

Jun Heo, Seungyul Lee, Sunhong Min, Yeonhong Park, Sung Jun Jung, **Tae Jun Ham**, Jae W. Lee

The 47th ACM/IEEE International Symposium on Computer Architecture (ISCA), 2021 [To Appear]

Acceptance Rate : 76/406 \approx 19%

[ASPLOS '21] MERCI: Efficient Embedding Reduction on Commodity Hardware via Sub-Query Memoization

Yejin Lee, Seong Hoon Seo, Hyunji Choi, Hyoung Wook Sul, Soosung Kim, Jae W. Lee, **Tae Jun Ham (Corresponding)**

The 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2021 [To Appear]

Acceptance Rate : 75/398 \approx 19%

[FAST '21] FlashNeuron: SSD Enables Large-Batch Training of Very Deep Neural Networks

Jonghyun Bae, Jongsung Lee, Yunho Jin, Sam Son, Shine Kim, Hakbeom Jang, **Tae Jun Ham**, Jae W. Lee

USENIX Conference on File and Storage Technologiess (FAST), 2021

Acceptance Rate : 28/130 \approx 21%

[FAST '21] Behemoth: A Flash-centric Training Accelerator for Extreme-scale DNNs

Shine Kim, Yunho Jin, Gina Sohn, Jonghyun Bae, **Tae Jun Ham**, Jae W. Lee

USENIX Conference on File and Storage Technologiess (FAST), 2021

Acceptance Rate : 28/130 \approx 21%

[HPCA '21] Layerweaver: Maximizing Resource Utilization of Neural Processing Units via Layer-Wise Scheduling

Young H. Oh, Seonghak Kim, Yunho Jin, Sam Son, Jonghyun Bae, Jongsung Lee, Yeonhong Park, Dong Uk Kim,

Tae Jun Ham, Jae W. Lee

The 27th IEEE International Symposium on High-Performance Computer Architecture (HPCA), 2021

Acceptance Rate : 63/258 \approx 24%

[ICCAD '20] Unlocking Wordline-level Parallelism for Fast Inference on RRAM-based DNN Accelerator

Yeonhong Park, Seung Yul Lee, Hoon Shin, Jun Heo, **Tae Jun Ham**, Jae W. Lee

The 39th IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2020

Acceptance Rate : 127/470 \approx 27%

[MICRO '20] Graphene: Strong yet Lightweight Row Hammer Protection

Yeonhong Park, Woosuk Kwon, Eojin Lee, **Tae Jun Ham**, Jung Ho Ahn, Jae W. Lee

The 53rd IEEE/ACM International Symposium on Microarchitecture (MICRO), 2020

Acceptance Rate : 82/424 \approx 19%

• IEEE Micro Top Picks Honorable Mention

[ISCA '20] Genesis: A Hardware Acceleration Framework for Genomic Data Analysis

Tae Jun Ham, David Bruns-Smith, Brendan Sweeney, Yejin Lee, Seong Hoon Seo, U Gyeong Song, Young H. Oh, Krste Asanovic, Jae W. Lee, Lisa Wu

The 47th ACM/IEEE International Symposium on Computer Architecture (ISCA), 2020

Acceptance Rate : 77/428 \approx 18%

• IEEE Micro Top Picks

[ISCA '20] A Case for Hardware-based Demand Paging

Gyusun Lee*, Wenjing Jin*, Wonsuk Song, Jeonghun Gong, Jonghyun Bae, **Tae Jun Ham**, Jae W. Lee, Jinkyu Jeong
The 47th *ACM/IEEE International Symposium on Computer Architecture (ISCA)*, 2020
Acceptance Rate : 77/428 \approx 18%

* Two authors contributed equally.

[ISCA '20] A Specialized Architecture for Object Serialization with Applications to Big Data Analytics

Jaeyoung Jang, Sung Jun Jung, Sunmin Jeong, Jun Heo, Hoon Shin, **Tae Jun Ham**, Jae W. Lee
The 47th *ACM/IEEE International Symposium on Computer Architecture (ISCA)*, 2020
Acceptance Rate : 77/428 \approx 18%

[ISPASS '20] MosaicSim: A Lightweight, Modular Simulator for Heterogeneous Systems

Opeoluwa Matthews, Aninda Manocha, Davide Giri, Marcelo Orenes-Vera, Esin Tureci, Tyler Sorensen, **Tae Jun Ham**, Juan Luis Aragon, Luca P. Carloni, Margaret Martonosi
IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2020
Acceptance Rate : 25/73 \approx 34%

• **Nominated for the Best Paper Award**

[ASPLOS '20] IIU: Specialized Architecture for Inverted Index Search

Jun Heo, Jaeyeon Won, Yejin Lee, Shivam Bharuka, Jaeyoung Jang, **Tae Jun Ham**, Jae W. Lee
The 25th *ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2020
Acceptance Rate : 86/476 \approx 18%

[HPCA '20] A³: Accelerating Neural Network Attention Mechanism with Approximation

Tae Jun Ham, Sung Jun Jung, Seonghak Kim, Young H. Oh, Yeonhong Park, Yoonho Song, Jung-Hun Park, Sanghee Lee, Kyoung Park, Jae W. Lee, Deog-Kyoon Jeong
The 26th *IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 2020
Acceptance Rate : 48/235 \approx 20%

[MICRO '19] Charon: Specialized Near-Memory Processing Architecture for Clearing Dead Objects in Memory

Jaeyoung Jang, Jun Heo, Yejin Lee, Jaeyeon Won, Seonghak Kim, Sung Jun Jung, Hakbeom Jang, **Tae Jun Ham**, Jae W. Lee
The 52nd *IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2019
Acceptance Rate : 79/343 \approx 23%

[IEEE Micro] SSDStreamer: Specializing I/O Stack for Large-Scale Machine Learning

Jonghyun Bae, Hakbeom Jang, Jeonghun Gong, Wenjing Jin, Shine Kim, Jaeyoung Jang, **Tae Jun Ham**, Jinkyu Jeong, Jae W. Lee
IEEE Micro, September 2019

[ATC '19] Asynchronous I/O Stack: A Low-latency Kernel I/O Stack for Ultra-Low Latency SSDs

Gyusun Lee, Seokha Shin, Wonsuk Song, **Tae Jun Ham**, Jae W. Lee, Jinkyu Jeong
USENIX Annual Technical Conference (ATC), 2019
Acceptance Rate : 71/356 \approx 20%

[ATC '19] Practical Erase Suspension for Modern Low-latency SSDs

Shine Kim, Jonghyun Bae, Hakbeom Jang, Wenjing Jin, Jeonghun Gong, Seungyeon Lee, **Tae Jun Ham**, Jae W. Lee
USENIX Annual Technical Conference (ATC), 2019
Acceptance Rate : 71/356 \approx 20%

[ACM TACO] Efficient Data Supply for Parallel Heterogeneous Architectures

Tae Jun Ham, Juan L Aragon, Margaret Martonosi
ACM Transactions on Architecture and Code Optimization (TACO), June 2019
• Presented on **HiPEAC** 2020 Conference

[ACM TACO] Decoupling Data Supply from Computation for Latency-Tolerant Communication in Heterogeneous Architectures

Tae Jun Ham, Juan L Aragon, Margaret Martonosi

ACM Transactions on Architecture and Code Optimization (TACO), June 2017

[MICRO '16] Graphicionado: A High-Performance and Energy-Efficient Accelerator for Graph Analytics

Tae Jun Ham, Lisa Wu, Narayanan Sundaram, Nadathur Satish, Margaret Martonosi

The 49th *IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2016

Acceptance Rate : $61/283 \approx 22\%$

- **MICRO-49 Best Paper Award**

[MICRO '15] DeSC: Decoupled Supply-Compute Communication Management for Heterogeneous Architectures

Tae Jun Ham, Juan L Aragon, Margaret Martonosi

The 48th *IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2015

Acceptance Rate : $61/283 \approx 22\%$

- **IEEE Micro's Top Picks from the Computer Architecture — Honorable Mention** (Top 23 Computer Architecture Papers of 2015)
- **Motivated \$5.8million DARPA-funded DECADES project** (<https://decades.cs.princeton.edu/>)

[HPCA '13] Disintegrated Control for Energy-Efficient and Heterogeneous Memory Systems

Tae Jun Ham, Bharath K. Chelepalli, Neng Xue, Benjamin C. Lee

The 19th *IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 2013

Acceptance Rate : $51/249 \approx 20\%$

COMPUTER
SKILLS

- **Languages** : C/C++, CUDA, Python, Chisel, Verilog, Matlab, R
- **Applications/Frameworks** : PyTorch, Numpy, SciPy, Pandas, Intel Pin, LLVM, Cadence C-to-Silicon, \LaTeX

PATENTS

**Instruction, Circuits, and Logic for Graph Analytics Acceleration
(US20170286122A1; WO/2017/172173)**

with Lisa Wu, Nadathur Satish and Narayanan Sundaram

Method For Accelerating Candidate Selection based on Similarity and Accelerator for Performing Candidate Selection (Pending - Application No. 16/270.054)

with Jae W. Lee, Deog-Kyoon Jeong, Seonghak Kim, Sung Jun Jung, and Minsoo Lim

Hardware Accelerator Performing Search using Inverted Index Structure and Search System Including the Hardware Accelerator (Pending - Application No. 17/118.085)

with Jae W. Lee, Jun Heo, Jaeyeon Won, and Yejin Lee

Hardware-based Demand Paging Technique (Pending - Application No. 16/950,370)

with Jinkyu Jeong, Jae W. Lee, Gyun Lee, and Wenjing Jin

Scheduler, Method for Operating the Same and Neural Network Accelerator System Including the Same (Pending)

with Jae W. Lee, Young H. Oh, and Seonghak Kim

Input-Aware Current Compensation for Reliable NVM Crossbar based In-Memory Computing (Pending)

with Jae W. Lee, Yeonhong Park, Seungyul Lee, Hoon Shin, and Jun Heo

**Method for Operating the same, and Electronic Device including the same
(for efficient quantized matrix multiplications) (Pending)**

with Jae W. Lee, Jaeyeon Won, and Seungwook Lee

RESEARCH I closely work with these students and supervise their work (along with their primary supervisor Jae W. Lee) through
 MENTORING frequent (often more than once a week) meetings with each group of students. I provide advices and help on research, technical implementation, and writing.

Past Mentees

- **Jaeyoung Jang**, Ph.D from Sungkyunkwan University Jul 2018 - Jan 2020
Now at Samsung Electronics
- **Jeonghun Gong**, M.S from Seoul National University Jan 2019 - Jan 2021
Now at Samsung Electronics

Graduate Students

- **Young H. Oh**, Ph.D Student at Sungkyunkwan University Jul 2018 - Present
- **Jun Heo**, Ph.D Student at Seoul National University Jul 2018 - Present
- **Jonghyun Bae**, Ph.D Student at Seoul National University Sep 2018 - Present
- **Shine Kim**, Ph.D Student at Seoul National University Sep 2018 - Present
- **Wenjing Jin**, Ph.D Student at Seoul National University Sep 2018 - Present
- **Sung Jun Jung**, M.S/Ph.D Student at Seoul National University Sep 2018 - Present
- **Yejin Lee**, M.S/Ph.D Student at Seoul National University Jan 2019 - Present
- **Yeonhong Park**, M.S/Ph.D Student at Seoul National University Jun 2019 - Present
- **Yunho Jin**, M.S/Ph.D Student at Seoul National University Jun 2019 - Present
- **Sam Son**, M.S/Ph.D Student at Seoul National University Jun 2019 - Present
- **Seung Yul Lee**, M.S/Ph.D Student at Seoul National University Jun 2019 - Present
- **Seong Hoon Seo**, M.S/Ph.D Student at Seoul National University Jun 2019 - Present
- **Soosung Kim**, M.S/Ph.D Student at Seoul National University Jan 2020 - Present
- **Hyunji Choi**, M.S/Ph.D Student at Seoul National University Jan 2020 - Present

Undergraduate Students

- **Jaeyeon Won**, Undergraduate at Seoul National University Jan 2019 - Aug 2019
Now a Ph.D student at MIT. May 2020 - Aug 2020
- **Woogyung Song**, Undergraduate at Seoul National University Jun 2019 - Aug 2019
- **Hyung Wook Sul**, Undergraduate at Seoul National University Jun 2020 - Aug 2020

PROFESSIONAL Paper Reviews

ACTIVITIES

- IEEE Transactions on Very Large Scale Integration Systems (TVLSI) - 2015
- IEEE Transactions on Mobile Computing (TMC) - 2016
- IEEE Transactions on Computer (TC) - 2017, 2018, 2020
- IEEE Computer Architecture Letters (CAL) - 2018
- IEEE Micro - 2019
- ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) - 2017
- ACM/IEEE International Symposium on Computer Architecture (ISCA) - 2020
- ACM Transactions on Architecture and Code Optimization (TACO) - 2014
- ACM Transactions on Parallel Computing (TOPC) - 2019
- Elsevier Future Generation Computer Systems (FGCS) - 2019

Invited Talks

- KAIST, POSTECH (Sep 2017)
DeSC: Decoupling Data Supply from Computation for Latency-Tolerant Communication in Heterogeneous Architectures
- DARPA HIVE PI Meeting (Oct 2017)
Graphicionado: A High-Performance and Energy-Efficient Accelerator for Graph Analytics
- HiPEAC 2020 (Jan 2020)
Efficient Data Supply for Parallel Heterogeneous Architectures

- Seoul National University AI Summer School (Aug 2020)
Accelerating Neural Network Attention Mechanism with HW/SW Codesign
- POSTECH Summer AI Seminar (Aug 2020)
Hardware/Software Co-design for Modern AI and Data Analytics Applications

Services

- External Review Committee Member, ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '19)
- Web Chair, ACM International Symposium on Code Generation and Optimization (CGO '21)