

## Tae Jun Ham

---

CONTACT	Building 944, 7th floor	<b>Voice:</b> (+82)-10-5314-0529
INFORMATION	1-Gwanak-ro, Gwanak-gu Seoul 08826, South Korea	<b>E-mail:</b> ham.taejun@gmail.com <b>Website:</b> <a href="https://taejunham.github.io">https://taejunham.github.io</a>
CURRENT POSITION	<b>Seoul National University</b> , Seoul, Republic of Korea <i>Postdoctoral Researcher</i>	<b>July. 2018 - Present</b>
	<b>Research Area:</b> Software-hardware co-design for emerging applications such as big data analytics and machine learning. <b>Supervisor:</b> Professor Jae W. Lee This position also fulfills the <b>mandatory military service duty</b> required for all Korean men.	
EDUCATION	<b>Princeton University</b> , Princeton, NJ USA <i>School of Engineering and Applied Science</i> <b>Ph.D</b> in Electrical Engineering <b>M.A</b> in Electrical Engineering <b>Dissertation:</b> Data Access Optimization in Accelerator-oriented Heterogeneous Architecture through Decoupling and Memory Hierarchy Specialization <b>Advisor:</b> Professor Margaret Martonosi and Professor Juan Luis Aragon	<b>Sep. 2012 - Jun. 2018</b>
	<b>Duke University</b> , Durham, North Carolina USA <i>Pratt School of Engineering</i> <b>B.S.E</b> in Electrical and Computer Engineering (GPA : 3.95/4.00) <i>Summa Cum Laude, with Distinction in Electrical and Computer Engineering</i>	<b>Aug. 2009 - Dec. 2011</b>
PUBLICATIONS	<b>[ASPLOS '21] MERCI: Efficient Embedding Reduction on Commodity Hardware via Sub-Query Memoization</b> Yejin Lee, Seong Hoon Seo, Hyunji Choi, Hyoung Wook Sul, Soosung Kim, Jae W. Lee, <b>Tae Jun Ham (Corresponding Author)</b> The 26th <i>ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)</i> , 2021 [To Appear] Acceptance Rate : 75/398 $\approx$ 19%  <b>[HPCA '21] Layerweaver: Maximizing Resource Utilization of Neural Processing Units via Layer-Wise Scheduling</b> Young H. Oh, Seonghak Kim, Yunho Jin, Sam Son, Jonghyun Bae, Jongsung Lee, Yeonhong Park, Dong Uk Kim, <b>Tae Jun Ham</b> , Jae W. Lee The 27th <i>IEEE International Symposium on High-Performance Computer Architecture (HPCA)</i> , 2021 [To Appear] Acceptance Rate : 63/258 $\approx$ 24%  <b>[ICCAD '20] Unlocking Wordline-level Parallelism for Fast Inference on RRAM-based DNN Accelerator</b> Yeonhong Park, Seung Yul Lee, Hoon Shin, Jun Heo, <b>Tae Jun Ham</b> , Jae W. Lee The 39th <i>IEEE/ACM International Conference on Computer-Aided Design (ICCAD)</i> , 2020 Acceptance Rate : 127/470 $\approx$ 27%  <b>[MICRO '20] Graphene: Strong yet Lightweight Row Hammer Protection</b> Yeonhong Park, Woosuk Kwon, Eojin Lee, <b>Tae Jun Ham</b> , Jung Ho Ahn, Jae W. Lee The 53rd <i>IEEE/ACM International Symposium on Microarchitecture (MICRO)</i> , 2020 Acceptance Rate : 82/424 $\approx$ 19%  <b>[ISCA '20] Genesis: A Hardware Acceleration Framework for Genomic Data Analysis</b> <b>Tae Jun Ham</b> , 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 <i>ACM/IEEE International Symposium on Computer Architecture (ISCA)</i> , 2020 Acceptance Rate : 77/428 $\approx$ 18%	

**[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] A<sup>3</sup>: 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%

WORK EXPERIENCE	<b>Microsoft Research</b> , Cambridge, UK <i>Graduate Research Intern</i> Research on an efficient secure memory design with near-data computation	<b>May - Aug, 2016</b>
	<b>Intel Labs</b> — Parallel Computing Lab, Santa Clara, USA <i>Graduate Technical Intern</i> Research on a custom hardware accelerator for graph analytics applications.	<b>May - Nov, 2015</b>
	<b>AMD Research</b> , 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.	<b>Jun - Aug, 2013</b>
	<b>Samsung Advanced Institute of Technology</b> , Yongin, Republic of Korea <i>Research Intern</i> Research on a GPU branch divergence problem.	<b>Jun - Aug, 2012</b>
	<b>Duke University</b> — BCL Research Group, Durham, USA <i>Research Assistant</i> Research on a heterogeneous memory system.	<b>Jan - May, 2012</b>

HONORS AND AWARDS	<ul style="list-style-type: none"><li>• <b>ISPASS Best Paper Award Nominee (2020)</b></li><li>• <b>MICRO-49 Best Paper Award (2016)</b></li><li>• <b>IEEE MICRO Top Picks Honorable Mention (2016)</b></li><li>• <b>Facebook Graduate Fellowship Finalist (2016-2017)</b>, Facebook, Inc.</li><li>• <b>Gordon Y.S. Wu Fellowship (2012-2017)</b>, Princeton University</li><li>• <b>Samsung Scholarship (2012-2017)</b>, Scholarship that supports up to \$50,000 per year</li><li>• <b>Summa Cum Laude</b>, Duke University</li></ul>
----------------------	--

COMPUTER SKILLS	<ul style="list-style-type: none"><li>• <b>Languages</b> : C/C++, CUDA, Python, Chisel, Verilog, Matlab, R</li><li>• <b>Applications/Frameworks</b> : PyTorch, Numpy, SciPy, Pandas, Intel Pin, LLVM, Cadence C-to-Silicon, <math>\LaTeX</math></li></ul>
--------------------	---

PATENTS	<b>Instruction, Circuits, and Logic for Graph Analytics Acceleration (US20170286122A1; WO/2017/172173)</b> with Lisa Wu, Nadathur Satish and Narayanan Sundaram
	<b>Method For Accelerating Candidate Selection based on Similarity and Accelerator for Performing Candidate Selection (Pending - Application No. 16/270.054)</b> with Jae W. Lee, Deog-Kyoon Jeong, Seonghak Kim, Sung Jun Jung, and Minsoo Lim
	<b>Hardware Accelerator Performing Search using Inverted Index Structure and Search System Including the Hardware Accelerator (Pending)</b> with Jae W. Lee, Jun Heo, Jaeyeon Won, and Yejin Lee
	<b>Scheduler, Method for Operating the Same and Neural Network Accelerator System Including the Same (Pending)</b> with Jae W. Lee, Young H. Oh, and Seonghak Kim
	<b>Hardware-based Demand Paging Technique (Pending)</b> with Jinkyu Jeong, Jae W. Lee, Gyusun Lee, and Wenjing Jin

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

#### 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
- **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

#### 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

#### 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)