Taeyoung Kim

Ph.D in Computer Science (System/Architecture)
Fremont, CA, 94536

Phone: (650) 690-0092
Email: taeyoung_work@icloud.com
Webpage: https://tkimva.github.io

Key Contributions

Led development of scalable local LLM frameworks, enabling performance/power evaluation framework for CPUs and Platforms.

Merged academic research and industry expertise to deliver innovative solutions in system/platform optimization and modeling.

Extensive hands-on experience with simulation development/embedded system design/firmware design, bridging the gap between hardware and software.

Summary of Experience

AI Software Architect for local LLM Applications & Cloud Services (4+ years in industry)

Developed a local LLM client application and its accompanying cloud-based model service for Intel CPUs and Platforms (Intel AI-Assistant Builder released 02/2025).

Developed an LLM benchmarking framework to improve silicon performance and manage thermal limitations for Intel CPUs and Platforms (Internally released 03/2024).

Developed on-device CPU overclocking techniques for silicon performance optimization (Intel AI-Assistant Overclocking released Q3/2023).

Software Engineer for Presilicon System Modeling and Optimization (2017-2021 | 4 years in industry)

Developed ML-based solutions for thermal simulations for CPU die, benefiting 100+ Intel Thermal Engineers in analyzing advanced hardware systems.

Developed an ML-based system-level architectural modeling tool to enable EQ optimization at scale for platform I/O.

Developed a high-speed I/O simulation (channel performance simulatior) product utilized by 400+ Intel Analog/Signal Integrity Engineers.

VLSI/Computer Architecture and Embedded System Design (2005-2017 | 12 years in academia)

Developed ML-based workload partitioning algorithms for energy, reliability and performance for multicore processors.

Developed resource-based GPU workloads balancing algorithms for energy, reliability and performance.

Developed a system-level reliability optimization framework for datacenters and dark silicon many-core platforms.

Developed HW/SW co-design and optimization frameworks to accelerate system performance in FPGA.

Developed embedded systems, sensor data collection and performance tuning for various IOT platforms.

Software Engineer for Webservices (2001-2005 | 4 years in industry)

Developed and designed backend web services, ensuring seamless functionality and scalability.

Created a dictionary-based searching system for web services, improving search efficiency and usability.

Managed domain name infra services integrated with web service platforms.

Professional Experience

Staff AI Software Architect,
Client Computing Group–Ecosystem Development Division–AI Solution Group,
Intel Corporation, Santa Clara, CA, USA Jun.2021–Present

Senior Software Engineer,
Design Engineering Group–Cores & Client CAD Division–System Modeling Group,
Intel Corporation, Hillsboro, OR, USA Jun.2017–Jun.2021

Software Intern Graduate,
Platform Engineering Group–Design Technology Solutions–System Modeling Group,
Intel Corporation, Hillsboro, OR, USA Jun-Sep, 2016

Research Assistant,
ML-based System/Design Optimization,
Department of Computer Science and Engineering, University of California, Riverside, CA, USA, Sep.2013–Jun.2017

Research Staff,
Design Tool Optimization Development, Department of Electrical and Computer Engineering, University of Virginia,
Charlottesville, VA, USA, Jun.2012–Feb.2013

Research Assistant,
Embedded/Firmware Systems Design for MCUs, Department of Electrical and Computer Engineering, University of
Virginia, Charlottesville, VA, USA, Jan.2010–May.2012

College Instructor
Software Engineering and Computer Security Classes,
Kyungmin College, Uijeongbu, Korea, Aug.2007–May.2008

Research Assistant,
Embedded/Firmware Systems Design for MCUs, Department of Electronics and Computer Engineering, Korea University, Seoul, Korea, Mar.2005–Feb.2008

Software Engineering Manager,
eStop, Inc., Seoul, Korea, May.2004–Feb.2005

Software Engineering Manager,
Netpia, Inc., Seoul, Korea, Nov.2001–Apr.2004

Software Engineer,
eStop, Inc., Seoul, Korea, Aug.1999–Apr.2000

Education

Ph.D. Computer Science, University of California, Riverside, CA, USA
Advisor: Dr. Sheldon X.-D. Tan GPA: 3.87/4.0
Sep.2013–June 16th 2017

M.S. Electrical Engineering, University of Virginia, Charlottesville, VA, USA
Aug.2009–May 20th 2012

M.Eng. Electronics and Computer Engineering, Korea University, Seoul, Korea
Mar.2005–Feb 24th 2007

B.S. Electronics Engineering, Konkuk University, Seoul, Korea
Mar.1997–Feb 22nd 2005

Honors and Awards

[2025/Q2] Intel AI Assistant Builder Project Release V1.2, Divisional Recognition Award (DRA), PC Ecosystem Division, Intel Corporation

[2025/Q2] Building the 1st SOC thermal-aware floorplanner,Divisional Recognition Award (DRA), Intelligent Systems Research, Intel Corporation

[2025/Q1] Intel AI Assistant Builder Project Release V1.1, Divisional Recognition Award (DRA), Client Ecosystem Division, Intel Corporation

[2024/Q4] Intel AI Assistant Builder Project Release V1.0, Divisional Recognition Award (DRA), Client Ecosystem Division, Intel Corporation

[2023/Q4] AI-assist Overclocking Product Release, Divisional Recognition Award (DRA), Client Ecosystem Division, Intel Corporation

[2023/Q4] AI-assist Overclocking Product Release, Divisional Recognition Award (DRA), Platform Software Engineering Division, Intel Corporation

[2016] Dissertation Year Program (DYP) Fellowship Award, University of California, Riverside

[2016] Finalist at ACM student Research Competition (SRC), ICCAD

[2014, 2015, 2016] Travel Grant Award at ACM Student Research Competition (SRC), ICCAD

[2016] Travel Grant Award at Young Faculty Workshop, Design Automation Conference (DAC)

[2015] Best Poster Research Award at ACM PhD Forum, Design Automation Conference (DAC)

[2015] Travel Grant Award at ACM PhD Forum, Design Automation Conference (DAC)

[2014] Richard Newton Fellowship Award, Design Automation Conference (DAC)

[2013-2015] Dean's Distinguished Fellowship Award, University of California, Riverside

[2013] In Recognition of Exceptional Presentation (2nd place), KSEA, Virginia Regional Conference

[2007] Outstanding Academic Performance Award, Korea University

Professional Activities for Engineering Community

[2023-Present], Forum Chair, Intel Data Science Center of Excellence

[2022-Present], Forum Chair, Intel Weekly AI Forum

[2017-Present], Associate Editor, AI/ML/System Integration, the VLSI Journal, ELSEVIER

[2023-Present], Technical Program Committee (TPC), AI/ML Track, Design, Automation and Test in Europe Conference (DATE)

[2022-Present], Technical Program Committee (TPC), AI/ML Track, Design Automation Conference (DAC)

[2021-Present], Technical Program Committee (TPC), AI/ML Track, ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)

[2021-Present], Program Committee, Young Excellence (WYE) Program at IEEE Solid-State Circuits Society (ISSCC), IEEE

[2022], AI Track Session Chair, Asia and South Pacific Design Automation Conference (ASP-DAC)

[2020-2022], Standard Committee, IEEE P1924.1 Standard Working Group, Energy Efficient Comm Hardware, IEEE

[2017-2018], Program Committee, ACM Student Research Competition at International Conference On Computer Aided Design (ICCAD)

[2019-Present], Reviewer, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)

[2018-Present], Reviewer, Microelectronics Journal, ELSEVIER

[2018-Present], Reviewer, Microelectronics Reliability Journal, ELSEVIER

[2018-Present], Reviewer, ACM Journal on Emerging Technologies in Computing Systems (JETC)

[2017-Present], Reviewer, ACM Transactions on Embedded Computing Systems (TECS)

[2016-Present], Reviewer, ACM Transactions on Design Automation of Electronic Systems (TODAES)

[2015-Present], Reviewer, IEEE Transactions On Very Large Scale Integration (VLSI) Systems (TVLSI)

Patents

Optimized Group-Wise Quantized Matmul on Neural Network Accelerator, Patent pending, 2025

Optimized Rotary Positional Embedding on Neural Network Accelerator, Patent pending, 2025

System and Methods for Personalized GenAI Assistants on AI PC, Patent pending, 2025

Methods and Apparatus for Document Parsing for Retrieval-Augmented Generation, Patent pending, 2025

System and Methods for Seamless End-to-End Enterprise GenAI Solution, Patent pending, 2025

Systems, Methods, and Appratus for Autotuning of Retrieval Augmented Generation Parameters, Patent pending, 2025

Methods and apparatus to perform cloud-based artificial intelligence overclocking, US20230418622A1, Dec. 28, 2023

Methods and apparatus to adapt memory channel usage on a per-user basis, US20220188016A1, Jun. 16, 2022

Specific Detail of Intel Work

Staff AI Software Architect, AI Solution Group (2021-Present)

* AI-Assistant Builder Development and Productization and LLM Benchmark Development for Intel Silicon:

- Developed LLM Model quality evaluation and reporting pipelines for Local LLM Models.

- Developed/productize LLM model recommendation engine API for Intel Silicon in AWS Cloud.

- Developed/productize LLM model deployment pipeline development for Intel AI Assistant Builder App.

- Lead security development lifecycle (SDL) for cloud service architect for Local LLM Model Service.

- Developed and deployed AI-PC-specific LLM workloads/pipelines development for power and thermal analysis for Intel Silicon.

- Developed comparative study pipeline of local IPEX-LLM/OpenVINO model performance.

* AI-Assist Overclocking Solution Development and Productization for Intel Silicon:

- Lead developer a telemetry client app for CPU characterization to build the Overclocking AI model.

- Implemented application-level watchdog timer and auto-resume features for AI-Assist overclocking proof of concept software.

- Led a weekly working group for AI overclocking projects for two years.

- Conducted user journey studies for overclocking users.

- Enhanced software probing methodologies for minimum voltage measurement.

- Conducted a comparative study of CPU stress benchmarks for overclocking.

- Provided extensive training and support for large data collection in the lab, including test plans and training documents.

- Successfully productized and released overclocking features in the Intel Core 14th Generation Processor.

- Filed two patents related to AI-based overclocking and cloud-based AI overclocking.

* Software Tool Development for Intel Silicon Performance and Thermal Modeling:

- Developed cache cost modeling and power-performance database for L1/L2/L3/L4 cache sizing optimization.

- Led large-scale data collection for cache size impact studies and developed a practical design tool for cache sizing.

- Designed a backend database for CPU performance monitoring using MongoDB.

- Owned and managed an in-house thermal simulation solution, using grid and superposition methods for workload analysis.

- Managed a ticketing system for 50 engineering customers using Jira.

* Dynamic DRAM Tuning Study for Personalized n-User:

- Analyzed per-user memory usage from CPU telemetry big data in AWS Redshift.

- Developed an automated experiment framework for SOC and DRAM power measurement.

- Conducted benchmark studies for DRAM power and performance characterization per n-user requirements.

- Approved patent on adaptive memory channel usage on a per-user basis (US20220188016A1).

Senior Software Engineer, EDA Team (2017-2021)

* Pre-silicon Signal Integrity Simulation Tool Owner for Platform I/O and Chip-to-Chip I/O:

- Developed and supported in-house channel simulation tools for UPI and PCIe Gen4/Gen5/Gen6.

- Developed a new AMI parser for IBIS-AMI 7.0 compatibility.

- Correlated external industry solutions with in-house channel simulation methodologies.

- Productized FFT-based jitter decomposition algorithm development.

- Developed comprehensive observability for channel equalization and waveform processing in simulations.

- Optimized receiver hardware flow for PCIe Gen4/Gen5.

- Created new PCIe Tx/Rx equalization optimization flows with IBIS-AMI models.

- Productized machine learning-based TX AMI equalization model with common SPICE models.

- Managed a ticketing system for approximately 400 engineering customers using Jira.

- Developed a big data-based Elasticsearch system for proactive monitoring of simulation tools.

- Published 4 internal conference papers on platform I/O simulation methodology (not publicly available).

Publications

Books

S. Tan, M. Tahoori, **T. Kim**, S. Wang, Z. Sun, and S. Kiamehr, Long-Term Reliability of Nanometer VLSI Systems: Modeling, Analysis and Optimization, 1st ed. Springer International Publishing, 2019. (Co-authored, 40% contribution)

Theses

T. Kim, *System-Level Electromigration-Induced Dynamic Reliability Management,* Ph.D. thesis University of California, Riverside, June, 2017

T. Kim, *Detection and Prevention of Forward Head Posture with Body Sensor Networks,* M.S. thesis University of Virginia, Charlottesville, May, 2012

T. Kim, *A Large Scale Indoor Localization System Based on Wireless Sensor Networks,* M.Eng. thesis, Korea University, Seoul, Feb, 2007

Journal Articles

[VLSIJ'18] **T. Kim**, S. X.-D. Tan, C. Cook, and Z. Sun, "Detection of Counterfeited ICs Via On-Chip Sensor and Post-Fabrication Authentication Policy", *Integration, the VLSI Journal*, vol. 63, pp. 31-40, Sep. 2018.

[MJ'18] **T. Kim**, Z. Liu, and S. X.-D. Tan, "Dynamic reliability management based on resource-based EM modeling for multi-core microprocessors," *Microelectronics Journal*, vol. 74, pp. 106-115, Apr. 2018.

[TVLSI'18] S. Wang, **T. Kim**, Z. Sun, S. X.-D. Tan, and M. B. Tahoori, "Recovery-Aware Proactive TSV Repair for Electromigration Lifetime Enhancement in 3-D ICs" *IEEE Transactions on Very Large Scale Integration Systems (TVLSI)*, vol. 26, no. 3, pp. 531–543, Mar. 2018.

[TVLSI'18] S. Peng, H. Zhou, **T. Kim**, H. Chen, S. X.-D. Tan, "Physics-based Compact TDDB Models for Low-k BEOL Copper Interconnects with Time-Varying Voltage Stressing," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems (TVLSI)*, vol. 26, no. 2, pp. 239–248, Feb. 2018

[VLSIJ'18'] S. X.-D. Tan, H. Amrouch, **T. Kim**, Z. Sun, C. Cook, and J. Henkel, "Recent Advances in EM and BTI induced Reliability Modeling, Analysis and Optimization," *Integration, the VLSI Journal*, vol. 60, pp. 132–152, Jan. 2018.

[TDMR17] H. Chen, S. X.-D. Tan, **T. Kim**, and J. Chen, "Analytical Modeling of Electromigration Failure for VLSI Interconnect Tree Considering Temperature and Segment Length Effects," *IEEE Transactions on Device and Materials Reliability (TDMR)*, vol. 17, no. 4, pp. 653-666, Dec. 2017.

[VLSIJ'17] X. Huang, V. Sukharev, **T. Kim**, and S. X.-D. Tan, "Dynamic electromigration modeling for transient stress evolution and recovery under time-dependent current and temperature stressing," *Integration, the VLSI Journal*, *Integration, the VLSI Journal*, vol. 58, pp. 518–527, Jun. 2017.

[TVLSI'17] **T. Kim**, Z. Sun, H. Chen, H. Wang, and S. X.-D. Tan, "Energy and Lifetime Optimizations for Dark Silicon Manycore Microprocessor Considering both Hard and Soft Errors", *IEEE Trans Very Large Scale Integration (VLSI) Systems (TVLSI)*, vol. 25, no. 9, pp. 2561-2574, Sep. 2017.

[TCAD'16] H. B. Chen, S. X.-D. Tan, X. Huang, **T. Kim** and V. Sukharev, "Analytical Modeling and Characterization of Electromigration Effects for Multibranch Interconnect Trees," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, vol. 35, no. 11, pp. 1811-1824, Nov. 2016.

[VLSIJ'16] X. Huang, V. Sukharev, J.-H. Choy, M. Chew, **T. Kim**, S. X.-D. Tan, "Electromigration assessment for power grid networks considering temperature and thermal stress effects," *Integration, the VLSI Journal*, vol. 55, pp. 307–315, Sep. 2016

[TODAES'16] Z. Yue, **T. Kim**, H. Shin, S. X.-D. Tan, X. Li, H. Chen and H. Wang, "Statistical Rare Event Analysis and Parameter Guidance by Elite Learning Sample Selection", *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 21, no. 4, p. 56:1–56:21, May 2016

[IJACCS'13] S. Chen, J. S. Brantley, **T. Kim**, S. A. Ridenour, and J. Lach, "Characterising and minimising sources of error in inertial body sensor networks," *Int. J. Auton. Adapt. Commun. Syst. (IJAACS)*, vol. 6, no. 3, pp. 253-271, May. 2013.

[WPC'12] W. Y. Lee, K. Hur, **T. Kim**, D. S. Eom, and J. O. Kim, "Large scale indoor localization system based on wireless sensor networks for ubiquitous computing," *Wireless Personal*, vol. 63, no. 1, pp. 241-260, Mar. 2012.

[TCE'08] D Eom, **T. Kim**, H. Jee, H. Lee and J. Han, "A Multi-Player Arcade Video Game Platform with A Wireless Tangible User Interface", *IEEE Transactions on Consumer Electronics*, , vol. 54, no. 4, pp. 1819–1824, Nov. 2008.

Conference Proceedings

[SMACD'19] Z. Sun, **T. Kim**, M. Chow, S. Peng, H. Zhou, H. Kim, D. Wong and S. X.-D. Tan, "Long-Term Reliability Management For Multitasking GPGPUs" *International Conference on Synthesis, Modeling Analysis and Simulation Methods and Applications to Circuit Design (SMACD2019)*, Lausanne, Switzerland, Jul. 2019.

[SMACAD'17] Y. Ye, **T. Kim**, S. X.-D. Tan, H. Chen and H. Wang, "Comprehensive Detection of Counterfeit ICs Via On-Chip Sensor and Post-Fabrication Authentication Policy," *International Conference on Synthesis, Modeling Analysis and Simulation Methods and Applications to Circuit Design (SMACD2017)*, Taormina, Italy, Jun. 2017.

[ICCAD'16] **T. Kim**, Z. Sun, C. Cook, J. Gaddipati, H. Wang, H. Chen, S. X.-D. Tan, "Dynamic Reliability Management for Near-Threshold Dark Silicon Processors", *Proc. IEEE/ACM International Conference on Computer-Aided Design (ICCAD2016)*, Austin, TX, Nov. 2016.

[ICCAD'16] Z. Sun, E. Demircan, M. D. Shroff, **T. Kim**, X. Huang and S. X.-D. Tan, "Voltage-Based Electromigration Immortality Check for General Multi-Branch Interconnects", *Proc. IEEE/ACM International Conference on Computer-Aided Design (ICCAD2016)*, Austin, TX, Nov. 2016.

[SMACD'16] C. Cook, Z. Sun, **T. Kim** and S. X.-D. Tan, "Finite Difference Method for Electromigration Analysis of Multi-Branch Interconnects," *International Conference on Synthesis, Modeling Analysis and Simulation Methods and Applications to Circuit Design (SMACD2016)*, Lisbon, Portugal, Jun. 2016.

[DAC'16] **T. Kim**, Z. Sun, C. Cook, H. Zhao, R. Li, D. Wong, S. X.-D. Tan, "Cross-layer modeling and optimization for electromigration induced reliability", *Proc. IEEE/ACM Design Automation Conference (DAC2016)*, Austin, TX, June, 2016.

[DAC'16] X. Huang, V. Sukharev, Z. Qi, **T. Kim**, H. Chen, S. X.-D. Tan, "Physics-Based Full-Chip TDDB Assessment for BEOL Interconnects", *Proc. IEEE/ACM Design Automation Conference (DAC2016)*, Austin, TX, June, 2016.

[DATE'16] **T. Kim**, X. Huang, H. Chen, V. Sukharev, S. X.-D. Tan, "Learning-based Dynamic Reliability Management for Dark Silicon Processor Considering EM Effects", *Proc. Design, Automation and Test in Europe (DATE2016)*, Dresden, Germany, March 2016.

[ASPDAC'16] X. Huang, V. Sukharev, **T. Kim**, H. Chen, S. X.-D. Tan, "Electromigration Recovery Modeling and Analysis under Time-Dependent Current and Temperature Stressing", *Proc. Asia South Pacific Design Automation Conference (ASP-DAC2016)*, Macao, China, Jan. 2016.

[3DTEST'15] **T. Kim**, X. Huang, V. Sukharev and S. X.-D. Tan, "Learning-Based Reliability Management for Dark Silicon Systems", *Sixth IEEE International Workshop on Testing Three-Dimensional Stacked Integrated Circuits (3D-TEST2015)*, Anaheim, CA, Oct, 2015.

[TECHCON'15] **T. Kim**, X. Huang, V. Sukharev, X. X.-D. Tan, "A Dynamic Reliability Management Framework for Dark Silicon", *TECHCON*, Austin, Sep, 2015.

[DAC'15] H. Chen, X. Huang, V. Sukharev, S. X.-D. Tan, **T. Kim**, " Interconnect reliability modeling and analysis for multi-branch interconnect trees," *Proc. IEEE/ACM Design Automation Conference (DAC2015)*, San Francisco, June, 2015.

[ICCAD'14] **T. Kim**, B. Zheng, H. Chen, Q. Zhu, V. Sukharev and S. X.-D. Tan, "Lifetime optimization for real-time embedded systems considering electromigration effects," *Proc. IEEE/ACM International Conference on Computer-Aided Design (ICCAD2014)*, San Jose, CA, Nov. 2014

[DAC'14] T. Wei, **T. Kim**, S. Park, Q. Zhu, S. X.-D. Tan, N. Chang, S. Ula, M. Maasoumy, "Battery management and application for energy-efficient buildings," *Proc. IEEE/ACM Design Automation Conference (DAC2014)*, San Francisco, June, 2014

[BSN'11] **T. Kim**, S. Chen, J. Lach, "Detecting and Preventing Forward Head Posture with Wireless Inertial Body Sensor Networks," *International Conference on Body Sensor Networks (BSN2011)*, PP. 125-126, May, 2011

[WH'10] A.T. Barth, B.C. Bennett, B. Boudaoud, J.S. Brantley, S. Chen, C.L. Cunningham, **T. Kim**, H.C. Powell, Jr., S.A. Ridenour, J. Lach, "Longitudinal High-Fidelity Gait Analysis with Wireless Inertial Body Sensors," *IEEE Wireless Health Conference*, 192-3, Oct. 2010.

[BODYNET'10] S. Chen, J. S. Brantley, **T. Kim**, and J. Lach, "Characterizing and Minimizing Synchronization and Calibration Errors in Inertial Body Sensor Networks," in *Proceedings of the Fifth International Conference on Body Area Networks (BodyNet2010)*, Corfu, Greece, pp. 138-144. Sep. 2010.

Last updated: May 30, 2025