

Monday, August 12, 2024

Today's top trending papers in Computer Science

646,966 papers ranked by PageRank*. +314 new papers added in the last 9 hours. [Read more.](#)

Filter and sort 

Chip-Chat: Challenges and Opportunities in Conversational Hardware Design

PageRank: 37,343 Growth: +275% Citations: 59

Blocklove, Jason | Garg, Siddharth | Karri, Ramesh | Pearce, Hammond

May 22, 2023 – This paper discusses the challenges and opportunities in using artificial intelligence (AI) and conversational language models for hardware design. The authors present a case study where a...

TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings

PageRank: 30,291 Growth: +177% Citations: 97

Jouppe, Norman P. | Kurian, George | Li, Sheng | Ma, Peter | Nagarajan, Rahul | Nai, Lifeng | Patil, Nishant | Subramanian,...

Apr 3, 2023 – TPU v4 is a new supercomputer developed by Google for machine learning models. It utilizes optical circuit switches to improve performance and efficiency, and includes SparseCores that...

Retrospective: Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors

PageRank: 51,233 Growth: +166% Citations: 46

Mutlu, Onur

Jun 28, 2023 – The ISCA 2014 paper introduced the RowHammer vulnerability in DRAM chips, demonstrating that it is possible to induce bitflips in real systems by repeatedly accessing a DRAM row...

Sustainable AI: Environmental Implications, Challenges and Opportunities

PageRank: 10,267 Growth: +156% Citations: 145

Wu, Carole-Jean | Raghavendra, Ramya | Gupta, Udit | Acun, Bilge | Ardalani, Newsha | Maeng, Kiwan | Chang, Gloria |...

Oct 30, 2021 – This paper examines the environmental impact of AI and proposes ways to reduce its carbon footprint through hardware-software design and optimization. It also highlights the challenges an...

ChatEDA: A Large Language Model Powered Autonomous Agent for EDA

PageRank: 69,773 Growth: +141% Citations: 37

He, Zhuolun | Wu, Haoyuan | Zhang, Xinyun | Yao, Xufeng | Zheng, Su | Zheng, Haisheng | Yu, Bei

Aug 20, 2023 – This research paper introduces ChatEDA, an autonomous agent for Electronic Design Automation (EDA) that utilizes a large language model called AutoMage. ChatEDA streamlines the design...

RTLLM: An Open-Source Benchmark for Design RTL Generation with Large Language Model

PageRank: 72,108 Growth: +133% Citations: 36

Lu, Yao | Liu, Shang | Zhang, Qijun | Xie, Zhiyao

Aug 10, 2023 – This paper introduces an open-source benchmark called RTLLM for generating design RTL using natural language instructions. It also proposes a prompt engineering technique called self-...

Chasing Carbon: The Elusive Environmental Footprint of Computing

PageRank: **11,455** Growth: **+112%** Citations: **59**

Gupta, Udit | Kim, Young Geun | Lee, Sylvia | Tse, Jordan | Lee, Hsien-Hsin S. | Wei, Gu-Yeon | Brooks, David | Wu, Carole-...

Oct 28, 2020 – This paper highlights the environmental impact of computing and quantifies the carbon emissions associated with computer systems. It reveals that while operational energy consumption is...

ChipGPT: How far are we from natural language hardware design

PageRank: **73,726** Growth: **+109%** Citations: **28**

Chang, Kaiyan | Wang, Ying | Ren, Haimeng | Wang, Mengdi | Liang, Shengwen | Han, Yinhe | Li, Huawei | Li, Xiaowei

May 23, 2023 – This work explores the potential of using large language models (LLMs) like ChatGPT to assist hardware engineers in generating hardware logic designs from natural language specifications. Th...

Gemmini: Enabling Systematic Deep-Learning Architecture Evaluation via Full-Stack Integration

PageRank: **34,202** Growth: **+99%** Citations: **35**

Genc, Hasan | Kim, Seah | Amid, Alon | Haj-Ali, Ameer | Iyer, Vighnesh | Prakash, Pranav | Zhao, Jerry | Grubb, Daniel | Liew...

Nov 22, 2019 – Gemmini is an open-source DNN accelerator generator that considers the cross-stack, system-level effects in real-world environments, allowing for the evaluation of deep-learning architectur...

CHARM: Composing Heterogeneous Accelerators for Matrix Multiply on Versal ACAP Architecture

PageRank: **68,952** Growth: **+88%** Citations: **13**

Zhuang, Jinming | Lau, Jason | Ye, Hanchen | Yang, Zhuoping | Du, Yubo | Lo, Jack | Denolf, Kristof | Neuendorffer, Stephe...

Jan 5, 2023 – The CHARM framework is proposed to address the challenge of efficiently utilizing computation resources in heterogeneous architectures for deep learning applications with multiple matri...

GPT4AIGChip: Towards Next-Generation AI Accelerator Design Automation via Large Language Models

PageRank: **103,167** Growth: **+85%** Citations: **24**

Fu, Yonggan | Zhang, Yongan | Yu, Zhongzhi | Li, Sixu | Ye, Zhifan | Li, Chaojian | Wan, Cheng | Lin, Yingyan

Sep 19, 2023 – This paper explores the use of large language models (LLMs) to automate the design of AI accelerators, aiming to democratize the process and make it more accessible to non-experts. The author...

RTLCoder: Outperforming GPT-3.5 in Design RTL Generation with Our Open-Source Dataset and Lightweight Solution

PageRank: **113,549** Growth: **+79%** Citations: **24**

Liu, Shang | Fang, Wenji | Lu, Yao | Zhang, Qijun | Zhang, Hongce | Xie, Zhiyao

Dec 13, 2023 – This study introduces a new open-source language model for generating RTL code, which outperforms commercial models like GPT-3.5. The model is efficient, with a small parameter count and t...

Splitwise: Efficient generative LLM inference using phase splitting

PageRank: **106,096** Growth: **+79%** Citations: **37**

Patel, Pratyush | Choukse, Esha | Zhang, Chaojie | Shah, Aashaka | Goiri, Íñigo | Maleki, Saeed | Bianchini, Ricardo

Nov 30, 2023 – The paper discusses the challenge of efficient inference in large language models (LLMs) and proposes a technique called Splitwise, which splits the two main phases of LLM inference onto...

Assessing requirements to scale to practical quantum advantage

PageRank: **62,441** Growth: **+75%** Citations: **16**

Beverland, Michael E. | Murali, Prakash | Troyer, Matthias | Svore, Krysta M. | Hoefler, Torsten | Kliuchnikov, Vadym | Low,...

Nov 14, 2022 – This article discusses the challenge of scaling quantum computers to achieve practical quantum advantage and proposes a framework for estimating the resources required for large-scale...

Quantitative Information Flow for Hardware: Advancing the Attack Landscape

PageRank: **81,228** Growth: **+74%** Citations: **4**

Reimann, Lennart M. | Erdönmez, Sarp | Sisejkovic, Dominik | Leupers, Rainer

Nov 30, 2022 – This research paper discusses the limitations of current security analysis in Electronic Design Automation (EDA) tools and proposes a novel quantitative analysis approach called 2D-QModel t...

RTLFixer: Automatically Fixing RTL Syntax Errors with Large Language Models

PageRank: **121,320** Growth: **+70%** Citations: **23**

Tsai, Yun-Da | Liu, Mingjie | Ren, Haoxing

Nov 28, 2023 – The paper introduces RTLFixer, a framework that uses Large Language Models (LLMs) to automatically fix syntax errors in Verilog code. The framework demonstrates high proficiency in resolving...

hls4ml: An Open-Source Codesign Workflow to Empower Scientific Low-Power Machine Learning Devices

PageRank: **46,741** Growth: **+64%** Citations: **35**

Fahim, Farah | Hawks, Benjamin | Herwig, Christian | Hirschauer, James | Jindariani, Sergo | Tran, Nhan | Carloni, Luca P. | ...

Mar 9, 2021 – The hls4ml project has developed an open-source software-hardware codesign workflow called hls4ml, which allows machine learning algorithms to be implemented on FPGA and ASIC...

An Electro-Photonic System for Accelerating Deep Neural Networks

PageRank: **66,747** Growth: **+63%** Citations: **19**

Demirkiran, Cansu | Eris, Furkan | Wang, Gongyu | Elmhurst, Jonathan | Moore, Nick | Harris, Nicholas C. | Basumallik, Ayo...

Sep 2, 2021 – This paper presents ADEPT, an electro-photonic accelerator that combines photonic computing with electronic components to improve the performance of deep neural networks (DNNs). Th...

ReckOn: A 28nm Sub-mm² Task-Agnostic Spiking Recurrent Neural Network Processor Enabling On-Chip Learning over Second-Long Timescales

PageRank: **86,348** Growth: **+61%** Citations: **17**

Frenkel, Charlotte | Indiveri, Giacomo

Aug 20, 2022 – The paper introduces a new processor that enables on-chip learning for autonomous edge devices, allowing them to adapt to different variables in real-world scenarios. The processor...

Hardware Approximate Techniques for Deep Neural Network Accelerators: A Survey

PageRank: **60,555** Growth: **+60%** Citations: **25**

Armeniakos, Giorgos | Zervakis, Georgios | Soudris, Dimitrios | Henkel, Jörg

Mar 16, 2022 – This survey article explores the use of hardware approximation techniques in deep neural network accelerators to improve energy efficiency and address the computational complexity of DNNs....

Memory-Aware Denial-of-Service Attacks on Shared Cache in Multicore Real-Time Systems

PageRank: **108,852** Growth: **+58%** Citations: **4**

Bechtel, Michael | Yun, Heechul

May 21, 2020 – This paper explores the impact of memory performance on denial-of-service attacks on shared cache in multicore real-time systems. The authors introduce new cache DoS attacks that can...

Cheshire: A Lightweight, Linux-Capable RISC-V Host Platform for Domain-Specific Accelerator Plug-In

PageRank: **99,847** Growth: **+58%** Citations: **6**

Ottaviano, Alessandro | Benz, Thomas | Scheffler, Paul | Benini, Luca

May 8, 2023 – Cheshire is a lightweight and modular host platform designed for domain-specific accelerators in IoT and TinyML applications. It features a low-pin-count DRAM interface, configurable...

TheHuzz: Instruction Fuzzing of Processors Using Golden-Reference Models for Finding Software-Exploitable Vulnerabilities

PageRank: **47,517** Growth: **+58%** Citations: **18**

Tyagi, Aakash | Crump, Addison | Sadeghi, Ahmad-Reza | Persyn, Garrett | Rajendran, Jeyavijayan | Jauernig, Patrick |...

Jan 24, 2022 – The paper presents a novel hardware fuzzer called TheHuzz, which addresses limitations of existing hardware fuzzing techniques and improves the state of the art. TheHuzz successfully detects...

ScaleHLS: A New Scalable High-Level Synthesis Framework on Multi-Level Intermediate Representation

PageRank: **95,504** Growth: **+57%** Citations: **17**

Ye, Hanchen | Hao, Cong | Cheng, Jianyi | Jeong, Hyunmin | Huang, Jack | Neuendorffer, Stephen | Chen, Deming

Jul 24, 2021 – The paper introduces ScaleHLS, a new scalable and customizable high-level synthesis (HLS) framework that utilizes a multi-level compiler infrastructure called MLIR. ScaleHLS represents HLS...

HAMMER: boosting fidelity of noisy Quantum circuits by exploiting Hamming behavior of erroneous outcomes

PageRank: **99,906** Growth: **+57%** Citations: **10**

Tannu, Swamit | Das, Poulami | Ayanzadeh, Ramin | Qureshi, Moinuddin

Aug 19, 2022 – The paper introduces a post-processing technique called HAMMER that leverages the Hamming behavior of erroneous outcomes in quantum computers to improve the fidelity of the output...

Accelerating Sparse Deep Neural Networks

PageRank: **35,442** Growth: **+56%** Citations: **117**

Mishra, Asit | Latorre, Jorge Albericio | Pool, Jeff | Stosic, Darko | Stosic, Dusan | Venkatesh, Ganesh | Yu, Chong |...

Apr 16, 2021 – The paper discusses the challenges of maintaining model accuracy and achieving

Login or create account

Thanks for reading Trending Papers! Create your free account or log in to continue reading.

Email address

Continue