Trending Papers   ☰

Friday, June 21, 2024

Today's top trending papers in Computer Science

628,929 papers ranked by PageRank*. **+955** new papers added in the last 3 days. <u>Read more</u>.

Filter and sort ⌄

Sustainable AI: Environmental Implications, Challenges and Opportunities

PageRank: **11,681**   Growth: **+220%**   Citations: **128**

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...

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

PageRank: **36,819**   Growth: **+218%**   Citations: **80**

Jouppi, 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...

Chip-Chat: Challenges and Opportunities in Conversational Hardware Design

PageRank: **46,897**   Growth: **+205%**   Citations: **45**

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...

Chasing Carbon: The Elusive Environmental Footprint of Computing

PageRank: **12,043**   Growth: **+205%**   Citations: **57**

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...

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

PageRank: **61,781**   Growth: **+126%**   Citations: **37**

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....

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

PageRank: **33,425**   Growth: **+111%**   Citations: **34**

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...

Assessing requirements to scale to practical quantum advantage

PageRank: **67,956**   Growth: **+99%**   Citations: **12**

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...

ChatEDA: A Large Language Model Powered Autonomous Agent for EDA

PageRank: **94,410**   Growth: **+94%**   Citations: **26**

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...

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

PageRank: **46,685**   Growth: **+89%**   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...

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

PageRank: **98,845**   Growth: **+85%**   Citations: **24**

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-...

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

PageRank: **69,107**   Growth: **+83%**   Citations: **12**

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...

ChipGPT: How far are we from natural language hardware design

PageRank: **93,825**   Growth: **+76%**   Citations: **20**

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...

An Electro-Photonic System for Accelerating Deep Neural Networks

PageRank: **64,958**   Growth: **+73%**   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...

Accelerating Sparse Deep Neural Networks

PageRank: **37,125**   Growth: **+61%**   Citations: **113**

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 significant speedups in sparse deep neural networks. It introduces the design and behavior of Sparse...

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

PageRank: **94,445**   Growth: **+59%**   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...

Quantitative Information Flow for Hardware: Advancing the Attack Landscape

PageRank: **88,649**   Growth: **+59%**   Citations: **3**

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...

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

PageRank: **105,729**   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...

The Dirty Secret of SSDs: Embodied Carbon

PageRank: **127,563**   Growth: **+56%**   Citations: **6**

Tannu, Swamit | Nair, Prashant J.

Jul 8, 2022 – This paper explores the environmental impact of Solid-State Drives (SSDs) and proposes strategies to improve their sustainability compared to traditional Hard Disk Drives (HDDs). The...

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

PageRank: **50,541**   Growth: **+56%**   Citations: **32**

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...

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

PageRank: **97,063**   Growth: **+56%**   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...

AI and ML Accelerator Survey and Trends

PageRank: **108,850**   Growth: **+55%**   Citations: **13**

Reuther, Albert | Michaleas, Peter | Jones, Michael | Gadepally, Vijay | Samsi, Siddharth | Kepner, Jeremy

Oct 8, 2022 – This paper provides an updated survey of AI accelerators and processors, summarizing their peak performance and power consumption. It also analyzes trends based on release dates and...

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

PageRank: **147,177**   Growth: **+53%**   Citations: **16**

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...

RTGPU: Real-Time GPU Scheduling of Hard Deadline Parallel Tasks with Fine-Grain Utilization

PageRank: **65,713**   Growth: **+53%**   Citations: **3**

Zou, An | Li, Jing | Gill, Christopher D. | Zhang, Xuan

Jan 25, 2021 – The paper proposes RTGPU, a real-time scheduling algorithm for multiple GPU applications that meets hard deadlines. It takes into account CPU and memory copy segments, as well a...

HEROv2: Full-Stack Open-Source Research Platform for Heterogeneous Computing

PageRank: **71,868**   Growth: **+53%**   Citations: **6**

Kurth, Andreas | Forsberg, Björn | Benini, Luca

Jan 11, 2022 – The HEROv2 research platform is an FPGA-based system that allows for accurate and fast exploration of heterogeneous computers, combining general-purpose processors with domain-specific...

Hardware Approximate Techniques for Deep Neural Network Accelerators: A Survey

PageRank: **69,082**   Growth: **+51%**   Citations: **23**

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....

Eliminating the Barriers: Demystifying Wi-Fi Baseband Design and Introducing the PicoScenes Wi-Fi Sensing Platform

PageRank: **70,563**   Growth: **+51%**   Citations: **11**

Jiang, Zhiping | Luan, Tom H. | Ren, Xincheng | Lv, Dongtao | Hao, Han | Wang, Jing | Zhao, Kun | Xi, Wei | Xu, Yueshen | Li,...