1. Record Nr. UNISA996691662603316

Autore	Li Chao
Titolo	Advanced Parallel Processing Technologies : 16th International Symposium, APPT 2025, Athens, Greece, July 13-16, 2025, Proceedings / / edited by Chao Li, Xuehai Qian, Dimitris Gizopoulos, Boris Grot
Pubbl/distr/stampa	Singapore : , : Springer Nature Singapore : , : Imprint : Springer, , 2026
ISBN	9789819510214 9789819510207
Edizione	[1st ed. 2026.]
Descrizione fisica	1 online resource (598 pages)
Collana	Lecture Notes in Computer Science, , 1611-3349 ; ; 16062
Disciplina	004.35
Soggetti	Software engineering Operating systems (Computers) Computer systems Computers, Special purpose Artificial intelligence Software Engineering Operating Systems Computer System Implementation Special Purpose and Application-Based Systems Artificial Intelligence
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	-- Best Paper Candidates -- DACO: Unlocking Latent Dataflow Opportunities in Edge-side SIMT Accelerators -- ATLAS: Efficient Dynamic GNN System through Abstraction-Driven Incremental Execution -- Segmentation-Aware Optimization of Collective for Waferscale Chips -- Area-Efficient Automated Logic Design with Monte-Carlo Tree Search -- Chip and Accelerators -- NFMap: Node Fusion Optimization for Efficient CGRA Mapping with Reinforcement Learning -- A Unified Synthesis Framework for Dataflow Accelerators through Multi-Level Software and Hardware Intermediate Representations -- Defect-aware Task Scheduling and Mapping for Redundancy-Enhanced Spatial Accelerators -- Irregular Sparsity-

Enabled Search-In-Memory Engine for Accelerating Spiking Neural Networks -- Memory and Storage -- QRAMsim: Efficiently Simulating, Analyzing, and Optimizing Large-scale Quantum Random Access Memory -- CeDMA: Enhancing Memory Efficiency of Heterogeneous Accelerator Systems Through Central DMA Controlling -- PAMM: Adaptive Memory Management for CXL-/UB-Based Heterogeneous Memory Pooling Systems -- STAMP: Accelerating Second-order DNN Training Via ReRAM-based Processing-in-Memory Architecture -- Cloud and Networking -- Cochain: Architectural Support Mechanism for Blockchain-based Task Scheduling -- DyQNet: Optimizing Dynamic Entanglement Routing with Online Request in Quantum Network -- Veyth: Adaptive Container Placement for Optimizing Cross-Server Network Traffic of Microservice Applications -- Design for LLM and ML/AI -- Unifying Two Operators with One PIM: Leveraging Hybrid Bonding for Efficient LLM Inference -- AsymServe: Demystifying and Optimizing LLM Serving Efficiency on CPU Acceleration Units -- SparseTem: Boosting the Efficiency of CNN-Based Video Encoders by Exploiting Temporal Continuity -- TokenSim: Enabling Hardware and Software Exploration for Large Language Model Inference Systems -- Big Data and Graph Processing -- Achieving Efficient Temporal Graph Transformation on the GPU -- GASgraph: A GPU-accelerated Streaming Graph Processing System based on SubHPMAs -- Accelerating Large-Scale Out-of-GPU-Core GNN Training with Two-Level Historical Caching -- Understand Data Preprocessing for Effective End-to-End Training of DNN -- Secure and Dependable System -- TwinStore: Secure Key-Value Stores Made Faster with Hybrid Trusted/Untrusted Storage -- The Future of Fully Homomorphic Encryption: from a Storage I/O Perspective -- LASM: A Lightweight and General TEE Secure Monitor Framework -- Identifying Potential Anomalous Operations in Graph Neural Network Training -- APPT Posters -- DraEC: A Decentralized Routing Algorithm in Erasure-Coded Deduplication System -- Spatial-Aware Orchestration of LLM Attention on Waferscale Chips -- ACLP: Towards More Accurate Loop Prediction for High-Performance Processors -- DSL-SGD: Distributed Local Stochastic Gradient Descent with Delayed Synchronization -- Exploiting Large Language Models for Software-Defined Solid-State Drives Design -- Comber: QoS-aware and Efficient Deployment for Co-located Microservices and Best-Effort Tasks in Disaggregated Datacenters -- NISA-DV: Verification Framework for Neuromorphic Processors with Customized ISA -- Lembda: Optimizing LLM Inference on Embedded Platforms via CPU/FPGA Co-Processing -- QDLoRA: Enhanced LoRA Fine-Tuning on Quantized LLMs via Integrated Low-Rank Decomposition.

Sommario/riassunto

This book constitutes the refereed proceedings of the 16th International Symposium on Advanced Parallel Processing Technologies, APPT 2025, held in Athens, Greece, during July 13–16, 2025. The 17 full papers and 10 short papers included in this book were carefully reviewed and selected from 74 submissions. They were organized in topical sections as follows: Chip and Accelerators, Memory and Storage, Cloud and Networking, Design for LLM and ML/AI, Big Data and Graph Processing, and Secure and Dependable System.