

IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS

CALL for PAPERS

Generative AI Compute: Algorithms, Architectures, and Applications to CAS

Guest editors

Chuan Zhang, Southeast University, China (chzhang@seu.edu.cn)

Naigang Wang, IBM T.J. Watson Research Center, USA (nwang@us.ibm.com)

Jongsun Park, Korea University, Korea (jongsun@korea.ac.kr)

Li Zhang, University of Leeds, UK (l.x.zhang@leeds.ac.uk)

Scope and purpose

Generative Artificial Intelligence (GenAI) is one of the most heated and emerging topics. Both academia and industry are actively embracing GenAI for its astounding capabilities and wide range of applications. Towards its full implementation and ultimate prosperity, the assistance of the circuits and systems (CAS) community plays a pivotal role. However, the increasing scale of GenAI models poses unprecedented challenges to their practical deployment, creating a huge gap between GenAI and the field of circuits and systems. The challenges faced by contemporary GenAI implementations are multifaceted:

1. *Resource limitations*: Circuits and systems, which are limited by fixed computational resources, find it challenging to keep pace with the relentless expansion of GenAI models.
2. *Architectural rigidity*: The one-size-fits-all endeavor of existing hardware architectures fails to cater to the diverse, personalized requirements that different GenAI applications necessitate.
3. *Energy constraints*: The escalating computational demands of GenAI, growing virtually at an exponential rate, cannot be met sustainably by energy-constrained circuits and systems.

Addressing these challenges calls for joint efforts from algorithm, implementation, and design perspectives. First, algorithm optimizations towards efficient GenAI deployment is essential. Researchers are actively exploring complexity reduction techniques to streamline generative models without significantly compromising their performance. Though recent algorithm research has made progress on pruning and quantization, such downsized GenAI models remain resource-intensive. Thus, there is an urgent need for hardware-aware GenAI algorithms, while keeping the superior performance. Second, efficient circuits and systems for GenAI are imperatively needed. Innovative hardware and architectures for GenAI are continuously being proposed, aiming to strike a balance among scalability, flexibility, and efficiency. Companies in industry are making strides, but there is a continuous need for specialized GenAI accelerators and energy-efficient computing paradigms for GenAI. Third, GenAI for accelerating circuits and systems design is of great need and promising. GenAI also has the potential to enhance electronic design automation tools, emulate circuits, optimize simulations, and accelerate verification. However, challenges remain in ensuring reliability, efficiency, and trust.

The objective of this special issue is to solicit and present the latest research findings in the field of circuits and systems

for GenAI, and GenAI-aided design for circuits and systems.

Topics of interest

To give a comprehensive introduction to this field, we plan to solicit papers presenting the latest developments in algorithms, implementations, and applications related to GenAI, as well as GenAI-aided design for circuits and systems. Within the broad scope, we prioritize the following areas of interest:

1. *Algorithm optimizations for efficient GenAI deployment.* This special issue is interested in algorithm optimizations for efficient GenAI deployment, particularly optimizing algorithms with respect to hardware compatibility, and presenting low-complexity, hardware-friendly models for emerging applications, e.g., image generation, language translation, and genomic research.
2. *Circuits and Systems for efficient GenAI implementation.* This special issue invites papers on designing circuits and systems for efficient GenAI implementation, including architecture design for GenAI, circuit design for ASIC/FPGA, and embedded software implementation on multi-core CPU/GPU/DSP. This special issue also welcomes papers that address implementation-related issues, such as architecture security, performance evaluation and comparison, and scalability.
3. *GenAI for accelerating circuits and systems design.* This special issue also highlights how AI can transform the design methodology of circuits and systems. This can be applied to design automation, efficient verification and other related cases. Examples may include large language models (LLMs) and other GenAI for ASIC/FPGA circuits.

Given the multidisciplinary flavor of this special issue, we welcome papers with topics in any of the following areas and beyond:

- Computational complexity reduction of GenAI
- Resource efficient frameworks for GenAI workloads
- Hardware-friendly optimizations of GenAI
- Software/hardware co-design of GenAI
- Innovative key-value caching strategies of GenAI
- Ethical and security considerations in the deployment of GenAI
- Efficient architectures of GenAI
- Hardware accelerator design for GenAI
- Resource efficient implementations of GenAI
- Energy efficient implementations of GenAI
- Memory optimizations for GenAI implementations
- In-memory computing architectures for GenAI
- Implementation of GenAI in various fields
- Frameworks for secure and scalable implementation of GenAI
- GenAI and circuits and systems in emerging applications
- GenAI-aided very large-scale integrated circuits design
- LLM-aided circuits and systems design
- Neural networks for customized circuits and systems

Submission procedure

Prospective authors are invited to submit their papers following the instructions provided on the IEEE JETCAS website: <https://iee-cas.org/publication/JETCAS/manuscript-submission-guide>. The submitted manuscripts should not have been previously published, nor should they be currently under consideration for publication elsewhere.

The IEEE JETCAS submission site: <https://iee.atyponrex.com/journal/jetcas>

Important dates

- Manuscript submissions due: Dec. 02, 2024
- First round of reviews completed: Jan. 20, 2025
- Revised manuscripts due: Mar. 03, 2025
- Second round of reviews completed: Mar. 31, 2025
- Final manuscripts due: Apr. 14, 2025

Request for information

Corresponding Guest Editor: Chuan Zhang (chzhang@seu.edu.cn)