PUBLIC SUBMISSION

Received: May 29, 2025
Tracking No. mba-6ndp-czul
Comments Due: May 28, 2025 **Submission Type:** API

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0327
Comment on FR Doc # 2025-07332

Submitter Information

Name: Yizhou Sun

General Comment

See attached file(s): AI for Chip Design

Attachments

NSF AI for EDA Workshop Report (Concise Version)

Report for NSF Workshop on AI for Chip Design

Jason Cong (UCLA) and Yizhou Sun (UCLA)

Vancouver

Dec. 10th 2024

Background. The advanced integrated circuits (ICs), also called computer chips, consist of over 100 billion transistors in sub-5nm fabrication technologies, are arguably the single most complex man-made object created in the human history. Their design and manufacture are enabled by a set of sophisticated Electronic Design Automation (EDA) tools, which support the automation of design creation, synthesis, simulation, test, and verification processes so that the engineers can keep up with the exponential growth of design complexity following the Moore's Law. As the government is bringing the chip manufacturing capability back to USA, the lack of chip design talents is a serious concern. The data from the 2023 US Bureau of Labor Statistics shows the significant gap between the number of software developers (approximately 1,795,300) and hardware designers (78,100) highlights a labor crisis in hardware design. Moreover, beyond the general-purpose processors, in the past few years domain-specific accelerators (DSAs), such as Google's Tensor Processing Unit (TPU), have shown to offer significant performance and energy efficiency over general-purpose CPUs, which suggests that more chip designs are needed. A wide range of deep learning accelerators have been introduced in both industry and academia. However, it is well-known that the DSAs in field-programmable gate-arrays (FPGAs) or application-specific integrated circuits (ASICs) are hard to design and require deep hardware knowledge to achieve high performance.

Executive Summary

This report summarizes the key discussions and recommendations from the NSF Workshop on AI for Electronic Design Automation (EDA), held alongside the NeurIPS 2024 conference in Vancouver, Canada. The workshop brought together leading researchers and practitioners from both machine learning (ML) and Chip Design communities to address the urgent need for AI-driven automation in chip design. The co-organizers include: Jason Cong (UCLA), Sergio Guadarrama (Google), Stefanie Jegelka (MIT & TU Munich), David Z. Pan (UT Austin), Yizhou Sun (UCLA).

The central challenge in hardware design is the increasing complexity of modern chips. Domain-specific accelerators (DSAs) offer substantial performance and energy efficiency gains, but their design demands deep hardware expertise, leading to long turnaround times and hindering rapid innovation. AI techniques (such as Large Language Models),

in combination with traditional EDA tools, have the potential to offer a transformative solution to democratize hardware design, making it domain-specific, cheaper, and far more effective.

The workshop explored four key themes:

1. **AI for Physical Synthesis and Design for Manufacturing (DFM):** This theme addressed the challenges in translating a chip design into a manufacturable product, including multi-objective optimization, complex constraints, and long turnaround times. Current AI efforts leverage Reinforcement Learning (RL) for placement, surrogate models for analysis, LLMs for rule creation, and generative AI for design acceleration. Key showstoppers include generalization issues across designs and nodes, scalability limitations, data quality and availability, and the significant computational resources required for AI models. Opportunities lie in fostering cross-community collaboration, developing open-source frameworks and datasets, and exploring hybrid AI approaches that combine deep learning with traditional methods. Recommendations to NSF include supporting foundational AI research for EDA, encouraging open-source initiatives, promoting scalable compute infrastructure, and investing in workforce development.

2. **AI for High-Level and Logic-Level Synthesis (HLS and LLS):** This theme focused on automating the translation of high-level behavioral descriptions into gate-level specifications. The core problem is the manual, error-prone, and time-consuming nature of traditional synthesis, exacerbated by the labor disparity in hardware design. AI can automate design decisions, predict performance outcomes, and enhance design space exploration. Existing efforts include HLS coupled with deep learning optimization (HLS+DL), using Graph Neural Networks (GNNs) and LLMs for performance prediction and pragma insertion, and more ambitious natural language to RTL (NL2RTL) generation. Showstoppers include the difficulty in adapting AI models to toolchain changes and new designs (domain shift), and the critical need for high-quality, labeled training data. Significant opportunities exist for the AI community to tackle novel problems in multi-modality fusion, domain/task transfer, and optimization, while the EDA community can leverage AI for accelerated pipelines, synthetic data generation, and black-box optimization. Recommendations to NSF include encouraging large-scale benchmarks for HLS and LLS, investing in research on foundation models, hybrid AI techniques, and agent-based methods tailored to HLS/LLS workflows, and supporting workforce development to bridge the gap between software and hardware designers.

3. **AI Toolbox for Optimization and Design:** This theme focused on challenges raised to AI from chip design. Existing AI techniques rely heavily on well-maintained offline data, a clearly defined objective function, and an offline evaluation. Key recommendations for NSF include the following. Generalization capability and uncertainty estimation become even more critical in chip design. Reinforcement learning and black-box optimization can lead to more effective and efficient search, which is beyond the traditional prediction tasks in general ML. LLM agents could serve as a very powerful tool by automatically putting long pipelines together and getting end-to-end solutions from data collection, tool calling, to verification.

4. **AI for Test and Verification:** This theme discussed frontier AI approaches in test and verification of chips and software systems, from pre-silicon logic bugs and post-silicon electrical faults to manufacturing defects, reliability failures, and hardware/software security vulnerabilities. Testing and verification are critical to ensuring the correctness of hardware systems. Current efforts learn semantic representations of hardware designs for verification tasks; leverage generative AI models for analysis, bug detection, providing fixing hints, and developing end-to-end development systems. Opportunities lie in the combination of AI and non-AI techniques, hybrid AI-plus-formal verification flows, shared large-scale open benchmarks, and collaboration between EDA, hardware security, and software engineering communities. Recommendations to NSF include funding foundational research on AI for test and verification and seeding large-scale open benchmark datasets.

The workshop had 17 invited speakers and panelists covering four themes and attracted 130+ attendees, on a wide variety of methods and problems in this area. The speakers and panelists include various experts from both academia and industry: Deming Chen (UIUC), Jeff Dean (Google), Vijay Ganesh (Georgia Tech), Aditya Grover (UCLA), Farinaz Koushanfar (UCSD), Koen Lampaert (Broadcom), Yingyan (Celine) Lin (Georgia Tech), Yong Liu (Cadence Design Systems), Igor Markov (Synopsys), Subhasish Mitra (Stanford), Bryan Perozzi (Google), Pranay Prakash (Synopsys), Ruchir Puri (IBM), Mark Ren (Nvidia), Shobha Vasudevan (Google), Yusu Wang (UCSD), and Lingming Zhang (UIUC). The workshop website is: https://ai4eda-workshop.github.io.

Overall, the workshop highlighted the immense potential of AI to revolutionize chip design. Realizing this potential requires a concerted effort to overcome challenges related to data availability, generalization, scalability, and integration. Key recommendations for the NSF and the broader research community include:

- **Fostering cross-disciplinary collaboration** through shared datasets, benchmarks, open-source tools, and joint research programs.
- **Investing in foundational AI research** tailored for EDA, with a focus on explainability, scalability, and robustness.
- **Developing robust data infrastructures** to efficiently extract and manage EDA data, alongside establishing diverse and large-scale benchmark datasets.
- **Promoting scalable compute infrastructure** and encouraging the development of hybrid AI approaches that combine the strengths of AI with traditional EDA methodologies.
- **Investing in workforce development** to train a new generation of professionals at the intersection of AI and EDA.

By addressing these critical areas, the workshop aims to accelerate the adoption of AI in chip design, ultimately democratizing hardware innovation and enabling the creation of next-generation energy-efficient computing systems.