GA1: the Generated Accelerator 1 Designed for the efabless SoC Ecosystem

Guohua Yin
SIGS
Email: ygh158@yeah.net

Xinze Wang
SIGS
Email: send2alphatu@gmail.com

***Abstract*—The advent of generative AI models, such as ChatGPT, Gemini, and Claude, has revolutionized various domains, including hardware design. This proposal presents GA1 (Generated Accelerator 1), an AI-generated hardware accelerator specifically tailored for Keyword Spotting (KWS) applications within the efabless SoC ecosystem. By leveraging the power of large language models (LLMs) and the OpenLane framework, we aim to demonstrate the feasibility and potential of AI-generated hardware designs.**

I. Introduction

Keyword Spotting (KWS) is a critical technology in voice-enabled devices, enabling always-on voice recognition and triggering more complex speech recognition systems. However, designing efficient and low-power hardware accelerators for KWS remains a challenging task, often requiring extensive domain expertise and manual effort. This proposal explores the application of generative AI models to automate and accelerate the design process, potentially unlocking new levels of efficiency and innovation.

II. Proposed Design Methodology

Building upon the experiences and methodologies of CyberRio [1], one pioneering AI-generated hardware design in 1st efabless AI-generated design contest, we propose an end-to-end LLM-driven design flow for GA1. This approach leverages the vast knowledge and reasoning capabilities of LLMs to generate hardware specifications, architecture exploration, and optimized Register-Transfer Level (RTL) code.

Specification Generation: We will prompt LLMs to generate high-level specifications for the GA1 accelerator, considering performance, power, and area constraints specific to the efabless SoC ecosystem.

Architecture Exploration: LLMs will be utilized to explore various architectural options, evaluating trade-offs between efficiency, flexibility, and hardware resource utilization.

RTL Generation: Based on the selected architecture, LLMs will generate optimized RTL code for the GA1 accelerator, leveraging their knowledge of hardware design best practices and optimization techniques.

Verification: The generated RTL code will be rigorously verified using open-source verification tools, ensuring functional correctness and adherence to design specifications.

Physical Implementation: The OpenLane framework will be employed for the physical implementation of GA1, enabling automated synthesis, placement, and routing stages.

Pre-check: Comprehensive validation and testing procedures will be performed, including pre-silicon simulations, formal verification, and post-silicon validation on efabless shuttle runs.

III. Current Progress and Expected Deliverables

At the time of this proposal, we have successfully generated high-level specifications and partial RTL code for the GA1 accelerator using LLMs. Our ongoing efforts are focused on architecture exploration, optimization, and the generation of complete RTL code. We anticipate delivering a fully functional and validated GA1 accelerator design before the competition deadline, along with a comprehensive report detailing our methodology, results, and insights.

IV. Code Repository

The GA1 project, including all generated code, prompts, and documentation, will be open-sourced and made available on GitHub (https://github.com/0616ygh/GA1) by the end of April.

V. Conclusion

GA1 represents a pioneering effort in leveraging generative AI models for hardware design within the efabless SoC ecosystem. By combining the power of LLMs with traditional hardware design methodologies, we aim to demonstrate the potential of AI-generated hardware accelerators for KWS applications. This project not only contributes to the advancement of AI-driven hardware design but also serves as a foundation for future exploration and innovation in this emerging field.

References

[1] CyberRio. https://github.com/hello-eternity/Cyberrio