Cyberrio : An AI-generated RISC-V CPU

Guohua Yin, Xinze Wang, Yifei Zhu, Dawei Feng, Lei Ren, Zhangxi Tan

RIOS Lab, Tsinghua University

1 Abstract

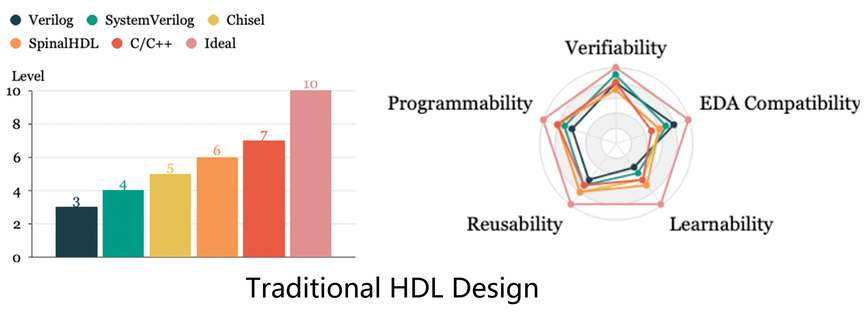
CyberRio is the first RISC-V CPU design with assists from the GPT-4 large language Model. Combining extensive usages of AI and OpenEDA in our agile hardware design flow, we have taped out CyberRio with the SkyWater 130nm technology as one of the winning designs in the Efabless AI-generated design contest. We share the overall design experience of CyberRio and shows an astonishing 1000% design productivity improvement over a traditional design flow.

2 Introduction

Due to challenges in CMOS technology scaling and the rise of open-source movements, we're observing a boom in domain-specific architectures, especially with the RISC-V open ISA[1]. However, this growth has made design more complex, emphasizing the need for agile hardware development methods[2]. Meanwhile, advancements in Machine Learning (ML) and Artificial Intelligence (AI) show potential to refine traditional ASIC design flows[3], though their full integration remains in the early stages.

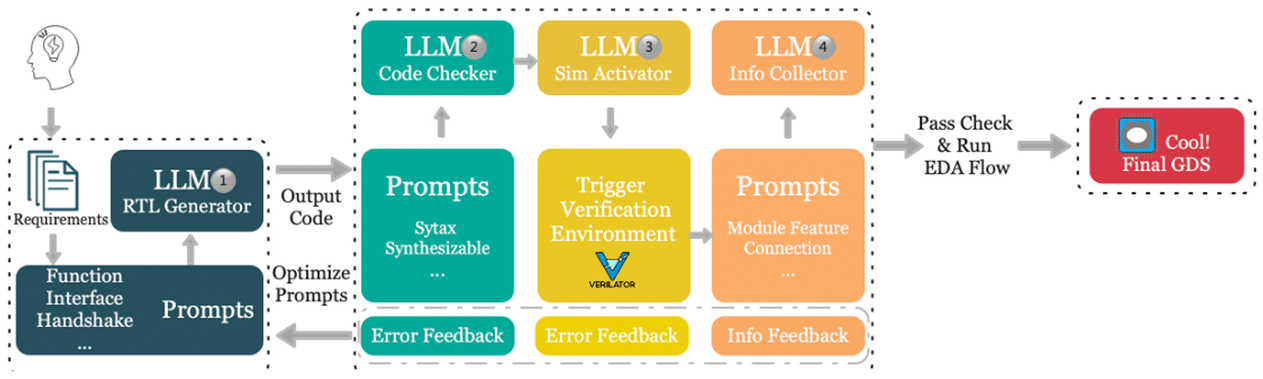
In this context, our study introduces CyberRio, the pioneering RISC-V CPU design crafted using GPT-4 LLM. The design, finalized with the OpenLane flow[5], won accolades in Efabless's open-source silicon design contest[6]. Our experiences, shared on Github (CyberRio: https://github.com/hello-eternity/Cyberrio), highlight LLM's potential in modern microprocessor design and hint at future AI-powered agile design methodologies.

3 Background

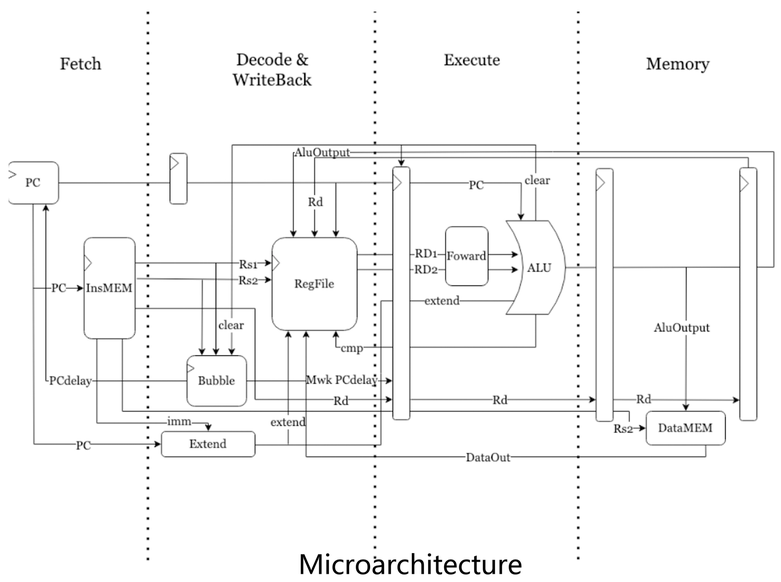


Traditional HDL Design

4 Frontend Flow

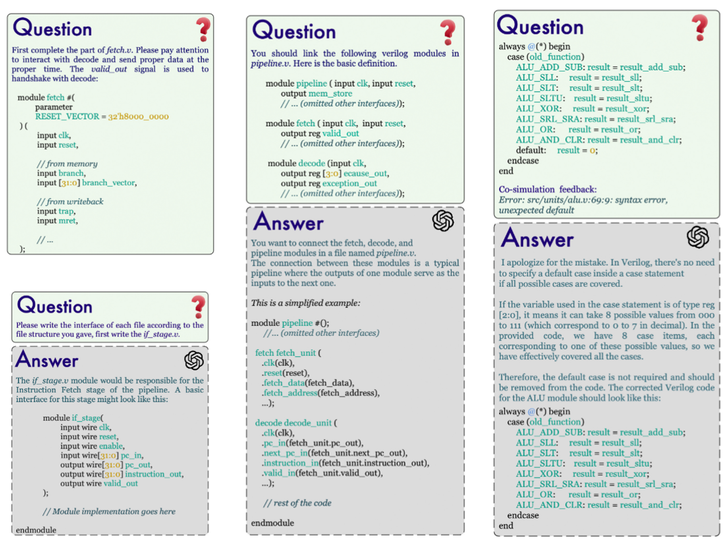


Overall FLow



Microarchitecture

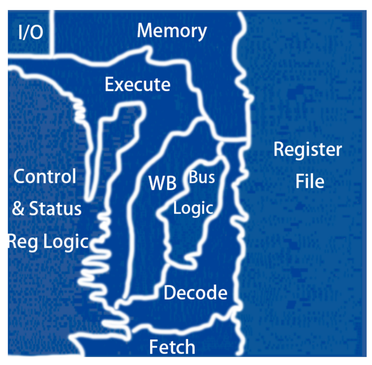
The collaborative development flow shown in Figure Overall Flow mimics how humans approach complex projects. In this flow, we integrate different LLMs at each stage of the development process. Each LLM plays a specialized role, focused on the implementation of its corresponding subtask. We aim to achieve an efficient design process by collaborating and dividing the labor across these specialized LLMs. To avoid issues stemming from long-term context, each LLM operates independently in a separate chat window of GPT-4. This approach proves a few major advantages: 1) no conversational loops, 2) generated content can be self-checked, and 3) enables increased extensibility for the flow.



Example of generate

5 Backend

EDA Flow

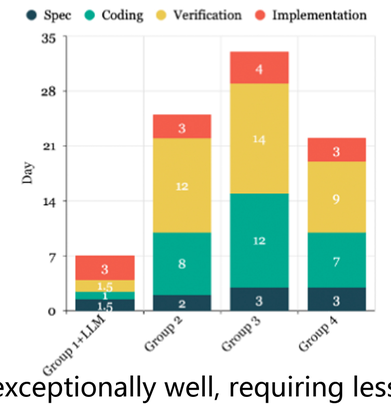


GDS of Cyberrio

Feature	Parameters
Gate Count	40,000
Pipeline	5
Frequency	100MHz
Area	0.12mm2

6 Evaluation

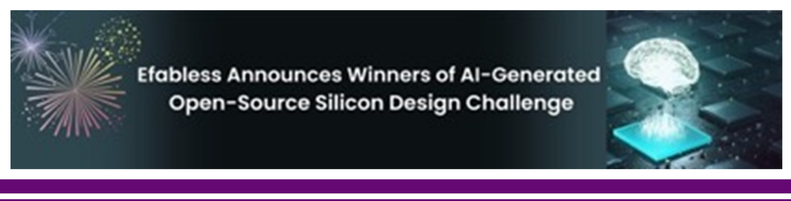
Role		LLM
Design Specification	Microarchitecture	D
	File Structure	A
	Functional Modules	A
Coding	Fetch	A
	Decode	D
	Execute	B
	Memory	C
	WriteBack	A
	Modules Connection	D
	Environment Setup	D
Verification	Syntax Check Lint	B
	Top Simulation	A



- Category A: The LLM performed exceptionally well, requiring less than 10% human modification.
- Category B: The LLM's performance was good, necessitating 10-25% human modification.
- Category C: The LLM's performance was moderate, requiring 25-40% human modification.
- Category D: The LLM's performance was unsatisfactory, with more than 40% human modification needed.

7 Current Status

Second-place prize in the Efabless AI Generated Design Contest
https://efabless.com/second-place-winner



8 Reference

[1] Waterman, Andrew; Lee, Yunsup; Avizienis, Rimas; Cook, Henry; Patterson, David; Asanovic, Krst. "The RISC-V Instruction Set." In Proc. 25th IEEE Hot Chips Symposium, 2013, pp. 1-1. doi:10.1109/HOTCHIPS.2013.7478332.

[2] Lee, Yunsup et al. "An Agile Approach to Building RISC-V Microprocessors." IEEE Micro., vol. 36, no. 2, pp. 8-20, 2016.

[3] Huang, Guyue et al. "Machine Learning for Electronic Design Automation: A Survey." ACM Trans. Des. Autom. Electron. Syst., vol. 26, no. 5, pp. 1-46, 2021. doi:10.1145/3451179.

[4] Sanderson, Katharine. "GPT-4 Is Here: What Scientists Think." Nature., vol. 615, no. 7954, pp. 773-773, 2023.

[5] Shalan, Mohamed; Edwards, Tim. "Building OpenLANE: A 130nm Openroad-Based Tapeout-Proven Flow." In Proc. 39th International Conference on Computer-Aided Design, 2020, pp. 1-6. doi:10.1145/3400302.3415735.