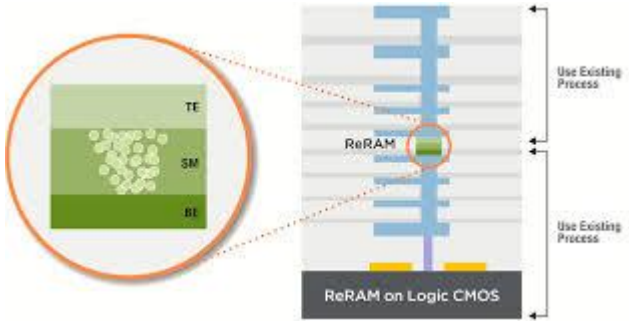
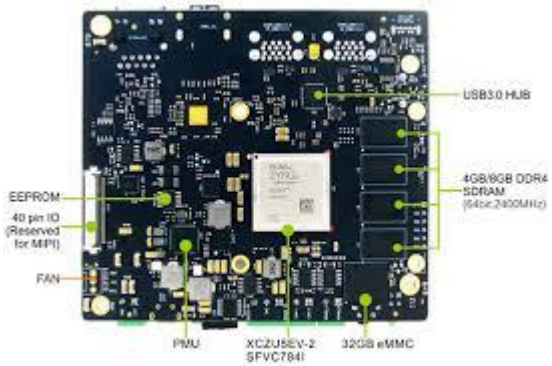
Product Research

By: Noah Mack, Sam Burns, Travis Jakl, and Olivia Price

Project Overview

Matrix-vector multiplication(MVM) is the most common operation in most machine learning applications. Performing large numbers of MVM operations takes a lot of time and energy in a typical CPU. This project aims to utilize Resistive Random Access Memory (ReRAM) to perform in-memory MVM computations, which can be dispatched by a CPU, acting as an accelerator for machine learning applications.





Problem Statement

This project aims to develop and evaluate a versatile ReRAM-based chip using the Skywater 130nm process, seeking to enhance matrix vector multiplication performance while addressing the trade-offs of different ReRAM configurations.

Related Products

NVIDIA Converged Accelerators

- It is designed to speed up data heavy workloads by integrating GPU and DPU units onto a single card which also saves PCIE bandwidth

Google’s Tensor Processing Unit

- Designed by Google, runs on CISC instructions, more I/O operations per joule.

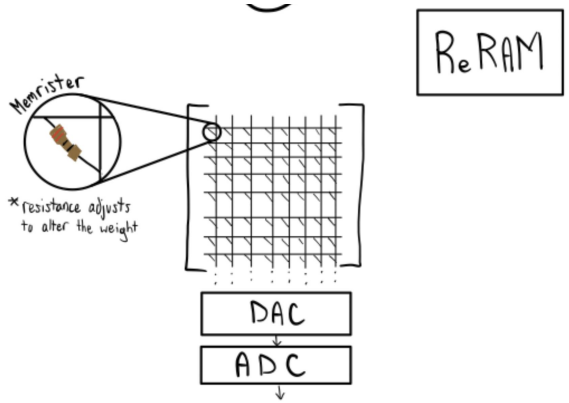
HAILO-8 AI Accelerator

- Provides a dataflow-centric design, minimizing data flow. Offers 26 Tera operations Per Second with only 2.5 W power consumption

There are many chips similar to ours, but not quite what we are doing. Our chip is a learning tool that will help students underneath us and above us understand chip layout. ReRAM is also a new upcoming technology so not much has been done with it.

Market Gap

As of now there is nothing like this on the market. We are trying to create a tool that does matrix multiplication in the memory block. This tool is a part of research projects through our professor and many senior design teams. If this chip is designed successfully, then students will be able to learn about the three chips architecture and how they were implemented. They will also let students see up and coming memory storage devices that will hopefully help to bridge the gap between the future and now.

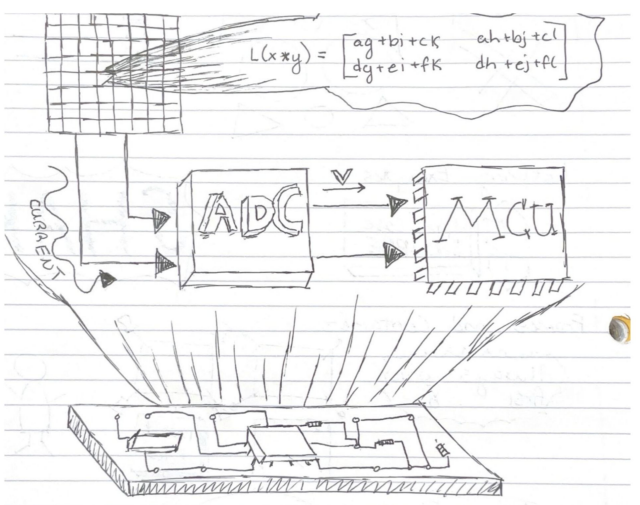


Ideas Generated by Product Research

Our Product Research consisted of looking at teams before us

Components we need to design and create:

- ADC
- Transimpedance amplifier
- Microcontroller
- ReRAM crossbar cell
- Feedback network
- Peripherals



With these components, hopefully we can get our product to work and successfully get it through pre-check and post layout.

Conclusion

In conclusion if this product succeeds we will fill in many gaps. These gaps include:

- Being able to make calculations in memory, instead of having it go to the ALU and back to memory.
- We will bridge a gap between students knowledge and new memory technologies.
- Research students will be able to use this chip as a building block and be able to innovate upon it.
- This design document will also be open to the public furthering research in the field and helping the field evolve.