

AiNalog: Machine Learning Powered by Analog Compute

Jordan de Souza, Mingze Jin, Patrick Li, Yihong Liu, Anujan Mathisekaran, Max Zhou
Consultant: Professor Peter Levine, PhD, PEng

Background and Objective

- Machine learning operations, such as matrix multiplication, are costly in resource-constrained environments when performed on digital processors
- Industry almost exclusively uses digital processors
- Potential for more AI on the edge (processed directly on-device) such as wearables, smart sensors, and IoT
- Objective:** Design a custom, low power, mixed-signal 16-bit matrix multiplying SoC suited for running ML models on the edge

Theory of Operation

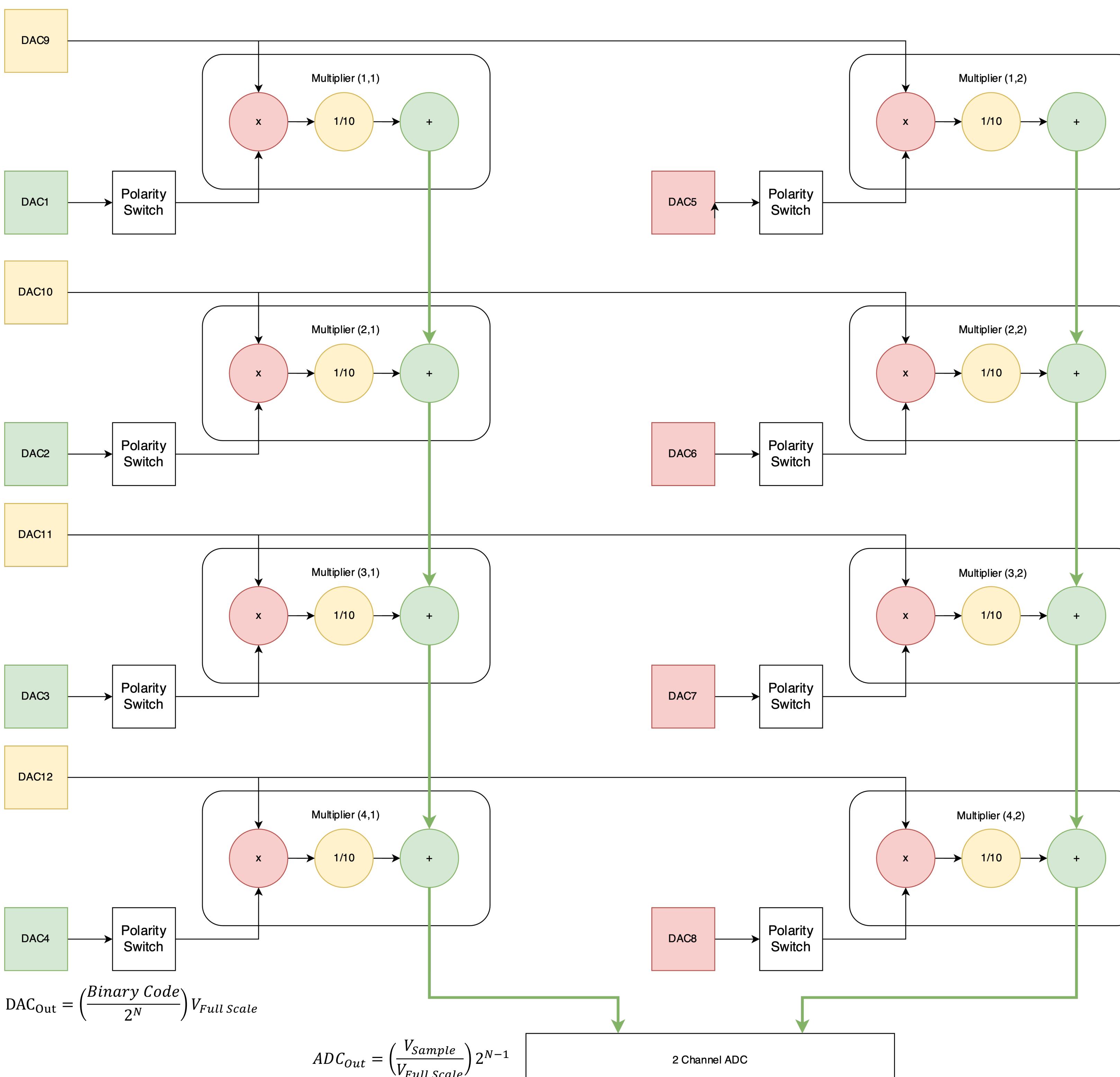


Figure 1. Detailed diagram of the analog multiplier array. AiNalog's input stage consists of 12 digital-to-analog converters (DACs). DACs 1 through 8 have polarity switches which enable negative inputs to be processed.

- FPGA sends ML model weights and matrix multiplicands to 12-bit DACs via SPI
- Polarity switch set for positive or negative multiplication
- Each stage of analog multiplier performs operation and sets carry out for following stages
- Results are accumulated via a 16-bit ADC and results are read out via SPI to FPGA

System Diagrams

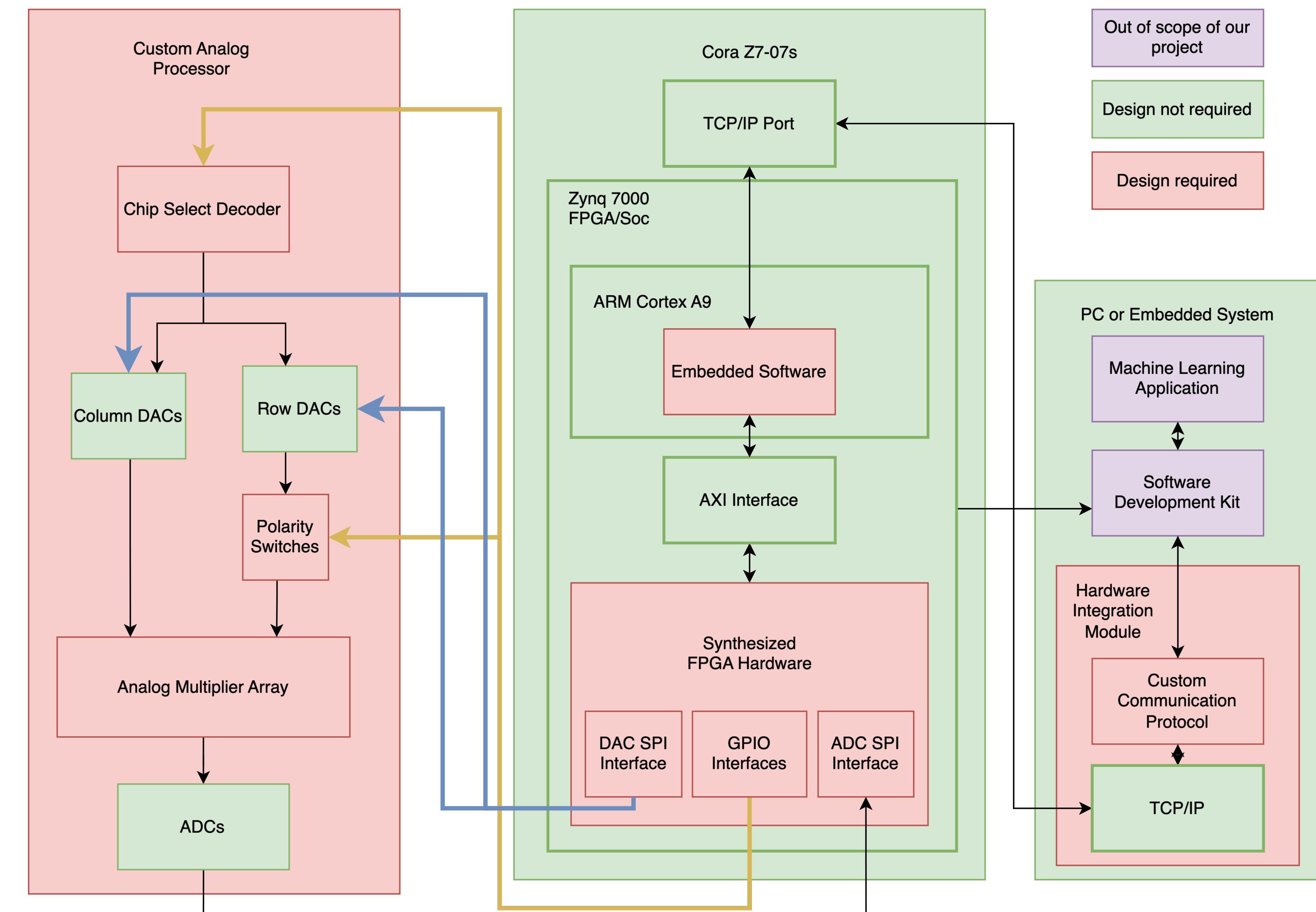


Figure 2. Block diagram of AiNalog system. Line colours are used for visibility and have no other meaning.

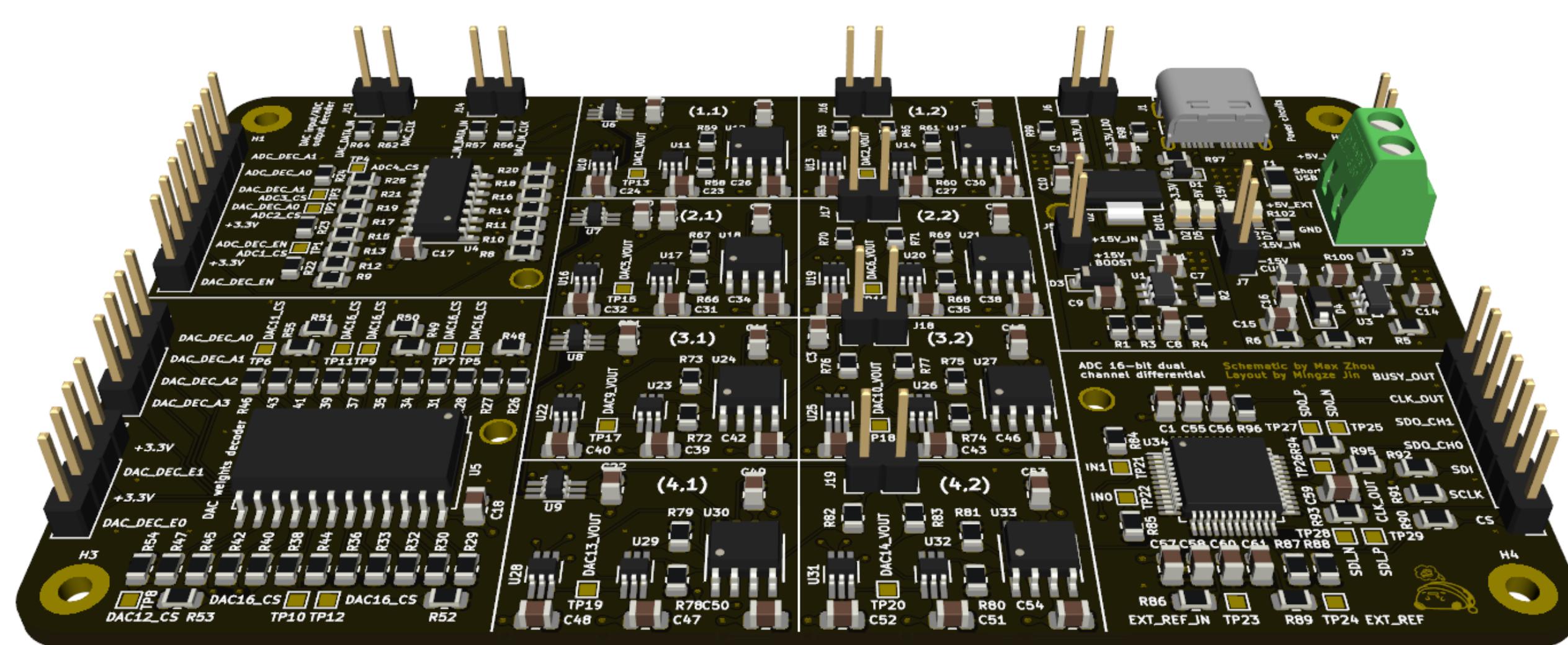


Figure 3. Render of the AiNalog PCB with all components mounted. While currently a 4-layer PCB, AiNalog is envisioned to be manufactured in the form of a compact integrated circuit in the future.

Runtime Analysis

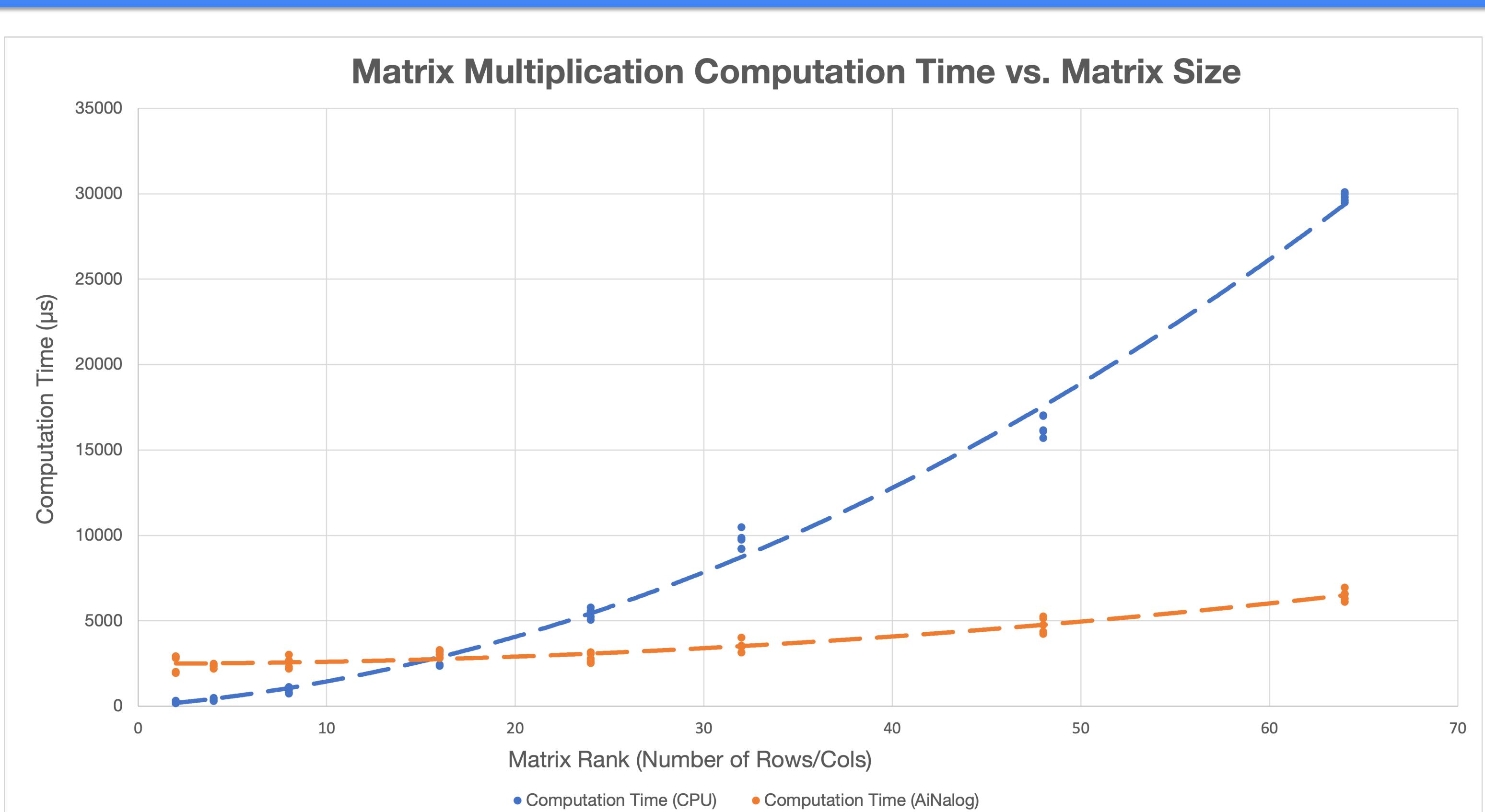


Figure 4. Plot of measured computation time (in microseconds) for multiplication of two N x N matrices using a CPU (Intel® Core™ i5-8259U) and AiNalog. Five trials conducted per matrix size, per processor.

- Computation time on digital processors grows drastically for larger matrices compared to AiNalog

Error Analysis

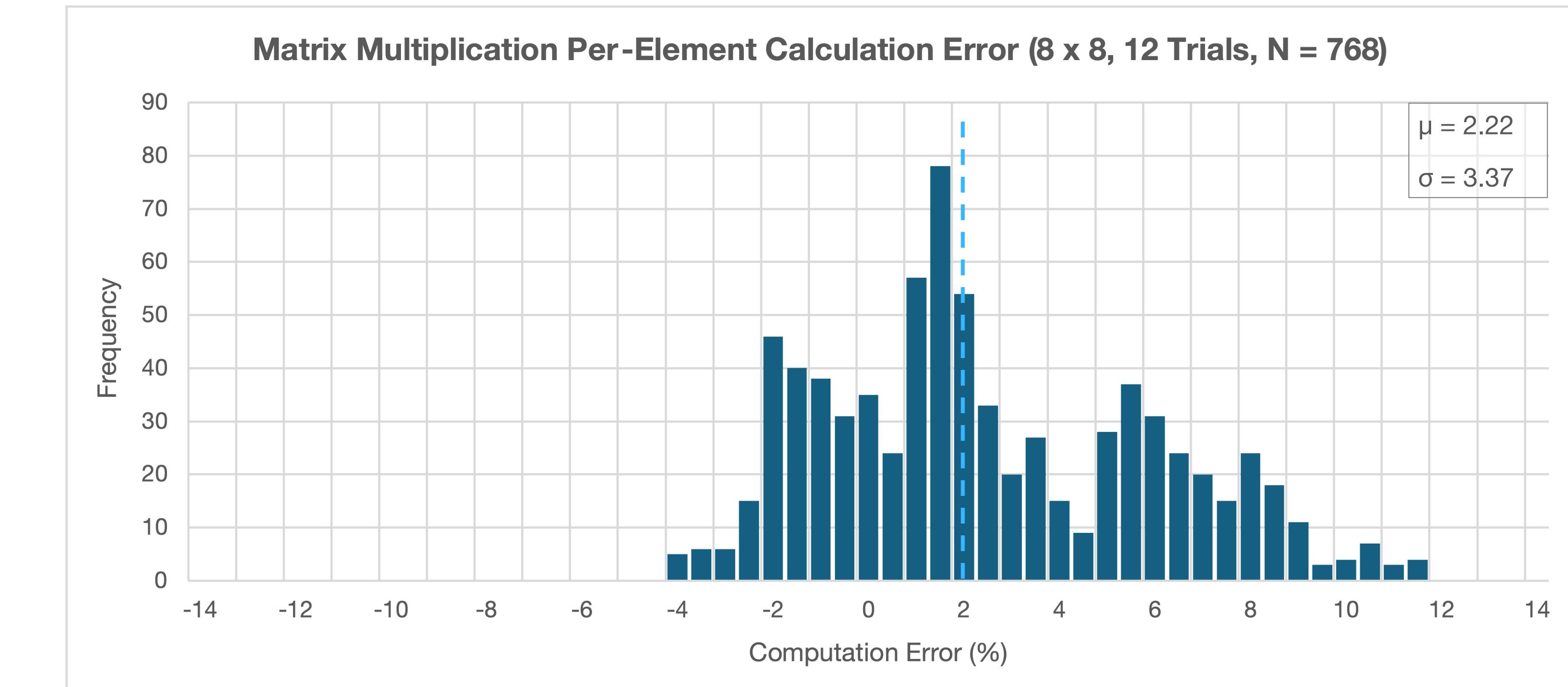


Figure 5. Histogram of observed per-element computation errors, as percentages, for multiplication of two 8 x 8 matrices using AiNalog. Errors calculated using Err = $\frac{\text{expected} - \text{actual}}{\text{actual}} \cdot 100\%$ for each element in result matrix.

- AiNalog can calculate matrix multiplications with an average error of 2.22%

Advantages Over Alternatives

- Matrix multiplication is expensive on digital systems (6 to 30 clock cycles), but nearly instantaneous in analog
- Low power consumption, consistent performance, and low latency

Discussion

- Power consumption is even lower than originally targeted: less than 3 W under all workloads
- Data transmission is the main bottleneck in the current design — combining the FPGA and analog multiplier into a single-board system would dramatically improve signal integrity and throughput
- While analog computations are somewhat noisy, neural networks have an inherent degree of noise resistance

Acknowledgements

- Thank you to Professor Peter Levine for his mentorship and support as consultant

References

- J. Hasler, "Analog Architecture Complexity Theory Empowering Ultra-Low Power Configurable Analog and Mixed Mode SoC Systems" [Jan. 2019]
- Analog Devices, "Low Cost Analog Multiplier", AD633 Datasheet, [Nov. 2002]
- Sedra, A.S. and Smith, K.C. "Microelectronic Circuits. 5th Edition", Oxford University Press [2004]
- Cora Z7 - Dilegent Reference. (n.d.). Dilegent.com. Retrieved February 13, 2024