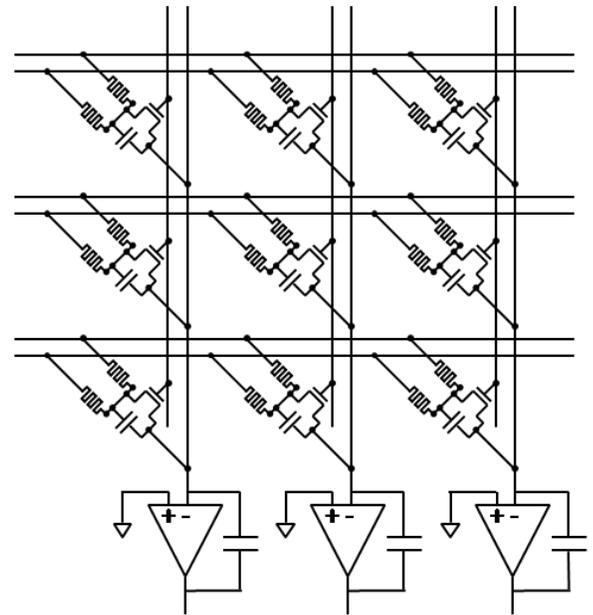


ANALOG MATRIX VECTOR MULTIPLICATION

Patent Pending
 SD# 15180
 Technology Readiness Level: 3
 Concept Demonstrated Analytically or Experimentally

Reduces the area and energy consumption of neural network accelerators using nonvolatile memory components (NVM)

Analog matrix vector multiplication (MVM) using nonvolatile memories is limited by the accuracy of the memory elements and their analog variability. Variations in “on” or “off” state current, or the low or high resistance state, directly translate to variability in value represented by the memory, thus limiting the accuracy of analog matrix multiplication. Capacitive based analog sums can be extremely accurate as the relative capacitance across elements in a local analog block is very well controlled; however, they currently require large SRAM based memory elements to charge the capacitors and therefore require extra power to store the non-volatile state. Using a nonvolatile memory to store the state and a capacitor for the analog sum realizes advantages for both systems.



Hybrid nonvolatile and capacitive analog MVM array

Sandia researchers have developed a system that uses nonvolatile memory elements (NVM) to charge capacitors arranged in an array to perform matrix vector multiplications (MVM). This development can be utilized by memory manufacturers to reduce both the area and energy consumption of accelerators, significantly improving the efficiency of binary accelerators. This allows for analog matrix multiplication to be performed with the accuracy of capacitive elements, while benefiting from the energy and area advantages of new non-volatile memories. The system can be altered to include multibit weights and multibit inputs. The system can ultimately be used to improve the efficiency of neural network accelerators.

TECHNICAL BENEFITS

- Reduces total accelerator area
- Reduces accelerator energy consumption
- Efficiency gains in memory

INDUSTRIES & APPLICATIONS

- Neural Network Accelerometers
- RAM
- Memory