ECE 564 ASCI Design Project Report

Convolutional Neural Network - 12 x 12 Matrix Dot Product

Abstract

A Convolutional Neural Network is designed to train a system for machine learning purpose by performing a series of matrix operations include multiplication and addition. The matrix and its respective “filter” vector will be read from different locations in two separate SRAM modules. The computed dot product results shall be written back to another separate SRAM containing just the output. The approach was to do things in a parallel processing-like design where multiple values were read from memory and computed in parallel for speed while still keeping sizing reasonable.

Introduction

A Convolutional Neural Network is designed to train a system for machine learning purpose by performing a series of matrix operations include multiplication and addition. The matrix and its respective “filter” vector will be read from different locations in two separate SRAM modules. The computed dot product results shall be written back to another separate SRAM containing just the output. The approach was to do things in a parallel processing-like design where multiple values were read from memory and computed in parallel for speed while still keeping sizing reasonable. In the Micro-Architecture part of the report the high level architecture drawing with interfaces will be presented and the datapath shall be elaborated on.