**ESE 461 Final Project Proposal**

Alan Coe

Christopher Ricciardi

Washington University in St. Louis

**Implementation**

We will use a single pipelined MAC between each layer for each of the 10 inputs. In total there should be 20 pipelined MACs. However, because going from the hidden layer to the output layer will have substantial delays due to its dependency on the previous computation, it is may be possible to reuse MACs in the second layer. For both layers we plan on broadcasting one value of the weight function at a time to each of the inputs, thus simplifying the top-level FSM.

For the first layer, we will have 200 neurons per input that require 784 MAC computations each. Once the MAC completes its operations on 784 values (one column) of the weight function, it will pass the value to the activation function and repeat with the next column. Rather than have a 200 x 16-bit array of values, we plan to “encode” the MAC output into a single binary bit. To do this we are simply going to compare the MAC output to ~4 (the approximate x-value of the flat portion of the sigmoid graph) and give it a corresponding binary value.

After completing enough computations, we will repeat this process with the second weight and activation functions. This will give us our final output.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Task** | 26-Nov | 29-Nov | 2-Dec | 6-Dec | 10-Dec | 13-Dec |
|  |  |  |  |  |  |  |
| **Proposal** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **Architecture** |  | X | X |  |  |  |
|  |  |  |  |  |  |  |
| **Control (FSM)** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **Presentation** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **Final Report** |  |  |  |  |  |  |

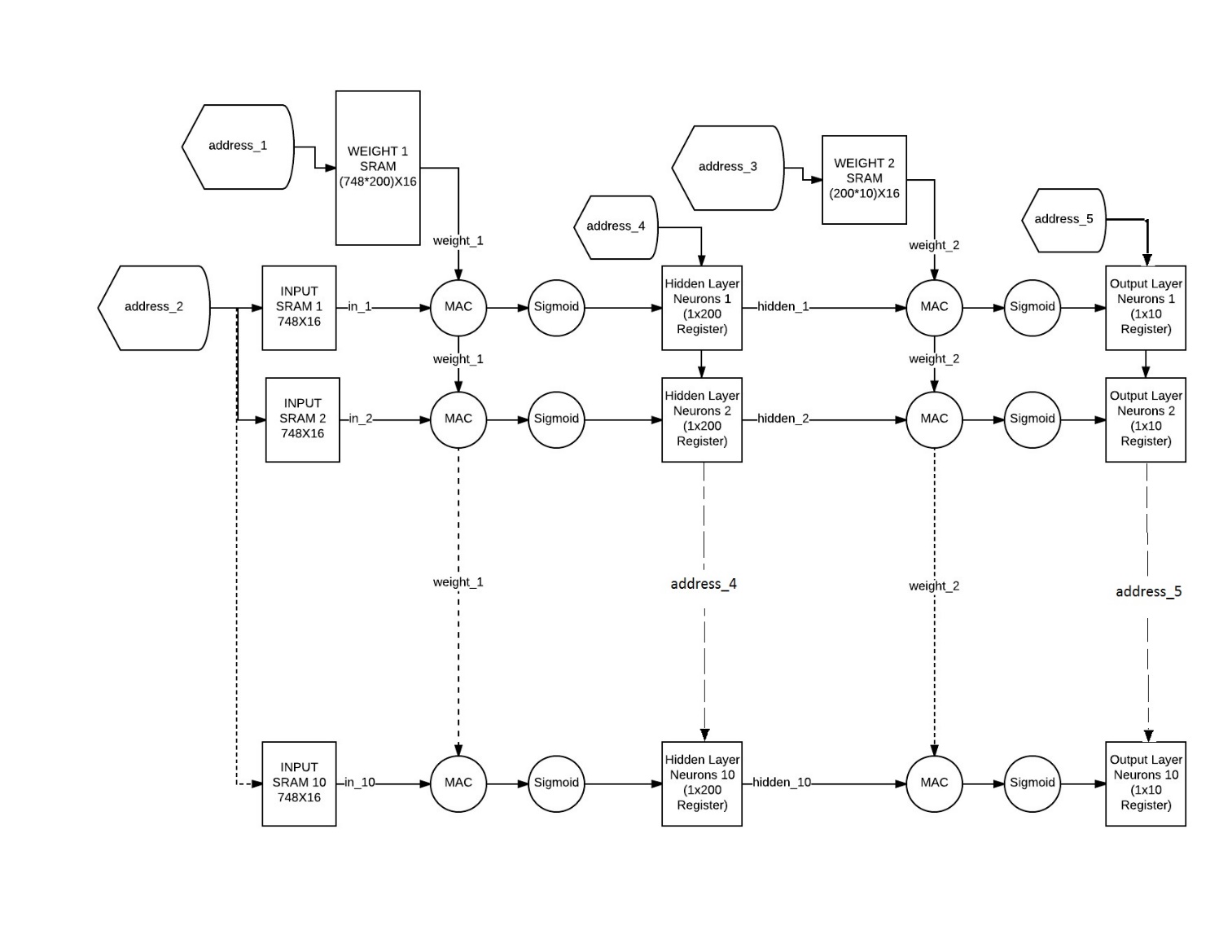


Figure - MLP Accelerator Data Path