DNN Design Phase 1

The design has two stages. In the first stage, four multiply-accumulate (MAC) units are used to compute the output of the first layer. Each MAC unit multiplies the input values x0, x1, x2, and x3 by the corresponding weight values w04, w14, w24, and w34 (for node_4), w05, w15, w25, and w35 (for node_5), w06, w16, w26, and w36 (for node_6), and w07, w17, w27, and w37 (for node_7), respectively. The output values of the MAC units (y4, y5, y6, and y7) are then passed through a rectified linear unit (ReLu) to generate four output values (z4, z5, z6, and z7) for the second layer.

In the second stage, two MAC units (out_node_0 and out_node_1) are used to compute the final output of the neural network. Each MAC unit multiplies the output values of the ReLu unit (z4, z5, z6, and z7) by the corresponding weight values w48, w58, w68, and w78 (for out_node_0), and w49, w59, w69, and w79 (for out_node_1), respectively. The output values of the MAC units (out0 and out1) represent the final output of the neural network.

There are also some internal signals in the design, such as in_ready1_q, in_ready2_q, and in_ready3_q, which are used to synchronize the input ready signal with the clock.

GNN Design Phase 2

Graph Neural Network (GNN) accelerator is an accelerator, which performs graph convolution operations on a graph dataset. The GNN has 4 nodes, each with 4 input features and 2 output features. It takes in the input features of all the nodes, the weights of the GNN, and the clock signal. It outputs the 2 output features for each of the 4 nodes, and a signal indicating whether each node's output is ready.

The GNN operation is implemented with a post-ReLU unit. The input features are first aggregated into a 7-bit representation, then passed through the post-ReLU unit to produce 12-bit hidden layer features, which are then passed through 4 independent DNNs to produce the final output features.

We include input and output ports for all the node's input and output features, the weights, the clock signal, and control signals indicating when the input is ready and when each node's output is ready. The internal signals are also defined to represent the aggregated and hidden layer features of each node, as well as the output features after passing through the DNNs.

Area(mm2)	Max Frequency (Mhz)	Min Latency (ns)	Power (mW)	Energy (pJ)	EDAP (pJ ns-mm2)
0.002519	748.50	2.67	4.91	13.1	0.086