[HW6_prob1] VGGNet_Hardware_Mapping_Tiling (10 pts)

- The VGGNet 3rd layer (assuming layers are indexed from 0)
- Weight stationary with an array size = 16 X 16
- Input and output channels are both 64
- Needs to be tiled into 16 pieces
- Goal is to format input, psum, and output in the following form:
- Input (a_tile) format: [ic_tile, oc_tile, ic index (row #), nij (time step)]
- psum (psum) format: [ic_tile, oc_tile, oc index (col #), nij (time step), kij]
- Output (out) format: [oc index, o_nij]

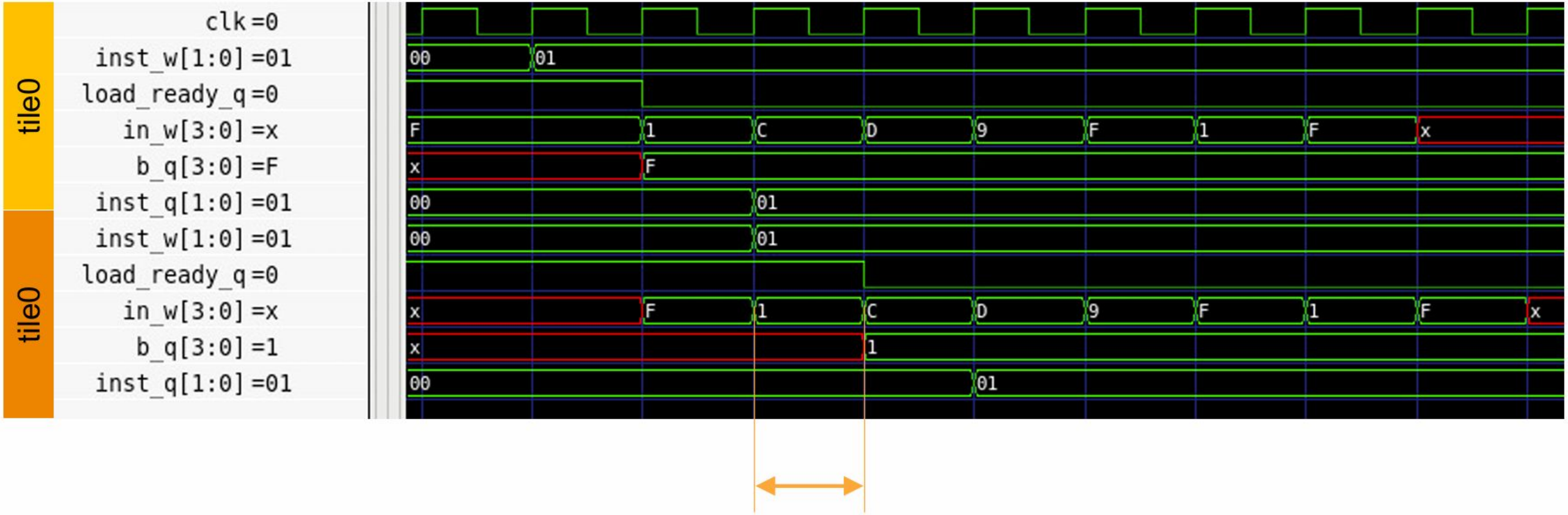
Abbreviation - ic: input channel, oc: output channel, o_nij: nij index for output

Refer “[Code14]_VGGNet_Hardware_Mapping.ipynb" to understand the indices better with an example.

[HW6_prob2-1] mac_tile Design (no need to simulate compilation error should be zero) (10 pts)

- Download from github: hardware/w5/hw1_2
- Equip ports: in_w, out_e, in_n, out_s, inst_w (input), inst_e (output), clk, reset (synchronous)
- Add latches: inst_q[1:0], a_q (activation), b_q (weight), c_q (psum), load_ready_q
- a_q is connected to out_e, and inst_e is connected to inst_q
- When reset ==1, inst_q[1:0] becomes all 0, and load_ready_q becomes 1
- Accept your inst_w[1] (execution) always into inst_q[1] latch.
- When either inst_w[0] or inst_w[1], accept the new in_w into a_q latch.
- When inst_w[0] (kernel load) == 1 and also load_ready_q ==1, accept the new weight in the b_q latch via in_w port. At the same time, load_ready_q becomes 0.
- When load_ready_q ==0, latch inst_w[0] into inst_q[0], which is connected to inst_e

Waveform for 2 PEs in a Row



Bubble cycle

[HW6_prob2-2] mac_row & mac_array (no need to simulate compilation error should be zero) (10 pts)

- Use for loop to build mac_row based on mac_tile from prob1
- PE row should provide a “valid[col-1:0]” vector to let ofifo know
- Inst_w[1:0] is also propagated from col0 to col7
- valid for the column is inst_e[1] for the column
- Look at “temp” wire use-case

- Similarly, use for loop to build mac_array based on mac_row
- PE row should provide a “valid[col-1:0]” vector to let ofifo know
- Here, only the last row’s valid signals are used.