8 Systolic Arrays [30 points]

A systolic array consists of 3x4 Processing Elements (PEs), interconnected as shown in Figure 1. The inputs of the systolic array are labeled as H0, H1, H2 and V0,V1,V2,V3. Figure 2 shows the PE logic, which performs a multiply and accumulate operation (MAC), and it saves the result in an internal register (reg). Figure 2 also shows how each PE propagates its inputs. We make the following assumptions:

- The latency of each MAC is one cycle.
- The propagation of the values from i_0 to o_0, and from i_1 to o_1, takes one cycle.
- The initial value of all registers is zero.
- You can input a value more than once in the systolic array.

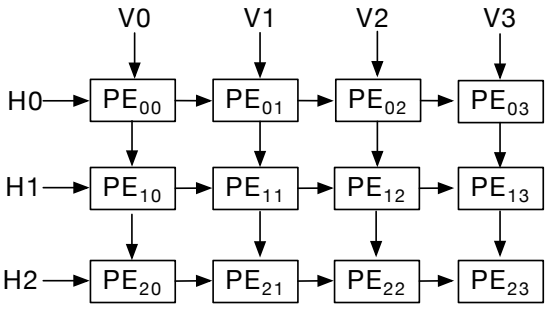

Figure 1: PE array

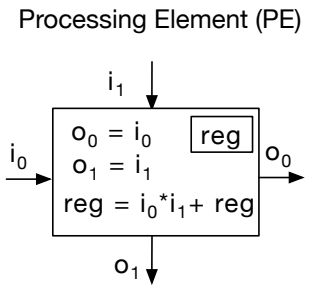

Figure 2: Processing Element (PE)

Your goal is to use this systolic array to perform the convolution of a 3x3 image (matrix I) with three 2x2 filters (matrices F, G, and H), to obtain three outputs (matrices O, U, and E):

$$\begin{matrix} I_{00} & I_{01} & I_{02} \\ I_{10} & I_{11} & I_{12} \\ I_{20} & I_{21} & I_{22} \end{matrix} \circledast \begin{matrix} F_{00} & F_{01} \\ F_{10} & F_{11} \end{matrix} = \begin{matrix} O_{00} & O_{01} \\ O_{10} & O_{11} \end{matrix}$$

$$\begin{matrix} I_{00} & I_{01} & I_{02} \\ I_{10} & I_{11} & I_{12} \\ I_{20} & I_{21} & I_{22} \end{matrix} \circledast \begin{matrix} G_{00} & G_{01} \\ G_{10} & G_{11} \end{matrix} = \begin{matrix} U_{00} & U_{01} \\ U_{10} & U_{11} \end{matrix}$$

$$\begin{matrix} I_{00} & I_{01} & I_{02} \\ I_{10} & I_{11} & I_{12} \\ I_{20} & I_{21} & I_{22} \end{matrix} \circledast \begin{matrix} H_{00} & H_{01} \\ H_{10} & H_{11} \end{matrix} = \begin{matrix} E_{00} & E_{01} \\ E_{10} & E_{11} \end{matrix}$$

As an example, the convolution of the matrix I with the filter F is computed as follows:

- $O_{00} = I_{00} * F_{00} + I_{01} * F_{01} + I_{10} * F_{10} + I_{11} * F_{11}$
- $O_{01} = I_{01} * F_{00} + I_{02} * F_{01} + I_{11} * F_{10} + I_{12} * F_{11}$
- $O_{10} = I_{10} * F_{00} + I_{11} * F_{01} + I_{20} * F_{10} + I_{21} * F_{11}$
- $O_{11} = I_{11} * F_{00} + I_{12} * F_{01} + I_{21} * F_{10} + I_{22} * F_{11}$

You should compute the three convolutions in the minimum possible amount of cycles. Fill the following table with:

1. The input values (matrices I, F, G, and H) in the correct input ports of the systolic array (the values can be repeated).
2. The output values and the corresponding PE where the outputs (matrices O, U, and E) are generated.

Fill the gaps only with relevant information.

cycle	**H0**	**H1**	**H2**	**V0**	**V1**	**V2**	**V3**	PE_{00}	PE_{01}	PE_{02}	PE_{03}	PE_{10}	PE_{11}	PE_{12}	PE_{13}	PE_{20}	PE_{21}	PE_{22}	PE_{23}
0	F_{00}			I_{00}															
1	F_{01}	G_{00}		I_{01}	I_{01}														
2	F_{10}	G_{01}	H_{00}	I_{10}	I_{02}	I_{10}													
3	F_{11}	G_{10}	H_{01}	I_{11}	I_{11}	I_{11}	I_{11}	O_{00}											
4		G_{11}	H_{10}		I_{12}	I_{20}	I_{12}		O_{01}			U_{00}							
5			H_{11}			I_{21}	I_{21}			O_{10}			U_{01}			E_{00}			
6							I_{22}				O_{11}			U_{10}			E_{01}		
7															U_{11}			E_{10}	
8																			E_{11}
9																			
10																			
11																			
12																			
13																			
14																			
15																			