Quiz Questions for Module 8

1. For a tiled 1D convolution, if the output tile width is 250 elements and mask width is 7 elements, what is the input tile width?

    a. 250
    b. 254
    c. 256
    d. 7

    Answer: (C)
    Explanation: The radius of the mask is 3. The input tile needs to cover all input elements needed to calculate the output tile. We need 3 additional elements on each side. The input tile size is 250+3+3 = 256.

2. In Question 1, if we assume that the input tile does not involve any ghost element, what would be the ratio of global memory reduction for generating the output tile by loading the input tile into the shared memory?

    a. 250*7/256
    b. 256*7/250
    c. 7
    d. 250

    Answer: (A)
    Explanation: The total number of global memory accesses to the input array is 7*250: 7 input elements are accessed for each output element and 250 output elements are generated.

3. In a tiled 2D convolution with 12x12 output tiles and 5x5 mask, how many warps in each thread block will have control divergence?

    a. 16
    b. 6
    c. 4
    d. 2

    Answer: (B)
    There are 16X16=256 threads in each block. Four out of every 16 threads will be inactive during the calculation phase. Out of the 8 warps, top 6 warps will have divergence. The bottom two warps will be entirely inactive, thus no divergence.

4. For a tiled 3D convolution, assume that we load an entire input tile, including the halo elements into the shared memory when calculating an output tile. Further assume that the tiles are internal and thus do not involve any ghost elements. If the mask is a cube with 5 elements on each side what is the trend of the average number of times each input element will be accessed from the shared memory during the calculation an output tile as a function of the input tile width?

    a. Increases with the width of the tile size with a limit of 25
    b. Decreases with the width of tile size with a limit of 25
    c. Increases with the width of the tile size with a limit of 125
    d. Decreases with the width of the tile size with a limit of 125

    Answer: (C)
    The average number of times an input tile element is accessed from the shared memory is (o_tile_width3*mask_widt3)/(o_tile_width+mask_width-1)3. For a given mask_width, the value increases as the o_tile_width increases. When o_tile_width becomes much larger than mask_width, the mask_width term can be dropped from the denominator. This makes the expression mask_width3.

5. For a tiled 2D convolution, if each output tile is a square with 12 elements on each side and the mask is a square with 5 elements on each side, how many elements are in each input tile?

    a. 12*12 = 144
    b. 5*5 = 25
    c. (12+2) * (12+2) = 194
    d. (12+4) * (12+4) = 256

    Answer: (D)
    Explanation: As shown in Lecture 8.4, the number of elements in an input tile is (o_tile_width+mask_width-1)*(o_tile_widt+mask_width-1), where o_tile_width is the width the output tiles.

6. For a tiled 2D convolution, assume that we load an entire input tile, including the halo elements into the shared memory when calculating an output tile. Further assume that the tiles are internal and thus do not involve any ghost elements. If each output tile is a square with 12 elements on each side and the mask is a square with 5 elements on each side, which of the following best approximate the average number of times each input element will be accessed from the shared memory during the calculation of an output tile?

    a. 256
    b. 37
    c. 14
    d. 4.9

    Answer: (C)
    Explanation: As shown in Lecture 8.4, Slide, the answer is o_tile_width2*mask_width2/ (o_tile_width+mask_width-1)2 = 122 * 52 / (12+5-1)2 = 14.1

7. For a tiled 3D convolution, assume that we load an entire input tile, including the halo elements into the shared memory when calculating an output tile. Further assume that the tiles are internal and thus do not involve any ghost elements. If the mask is a cube with 5 elements on each side and due to the limited size of the shared memory, each output tile is a cube with 8 elements on each side, what is the average number of times each input element will be accessed from the shared memory during the calculation an output tile?

    a. 256
    b. 37
    c. 14
    d. 4.9

    Answer: (B)
    Explanation: As can be generalized from Lecture 8.4, the answer is (tile_width^3*mask_width^3)/((tile_width+mask_width-1)^3) = ((8^3)*(5^3))/((8+5-1)^3)= 37