

Homework 1

Due Wednesday, September 14, 2022, 11:59pm

Late submission: 4% penalty per day; cannot be late by more than 6 days

Notes: (1) Please submit your solution electronically on CMS in a single PDF file named *hw1.pdf*; (2) Please put your name and NetID on the top of the first page; (3) Scanned handwritten copies are acceptable as long as your solution is legible.

Q1. In each of the following situations, indicate whether $f = O(g)$, or $f = \Omega(g)$, or both.

- (a) $f(n) = n - 100$; $g(n) = n - 200$
- (b) $f(n) = 100n + \log n$; $g(n) = n + (\log n)^2$
- (c) $f(n) = n!$; $g(n) = 2^n$

Q2. Please answer the following questions about arbitrary precision data types.

(a) In the following code segment, what is the minimum bitwidth of e that preserves the full precision of the expression? Please justify your answer.

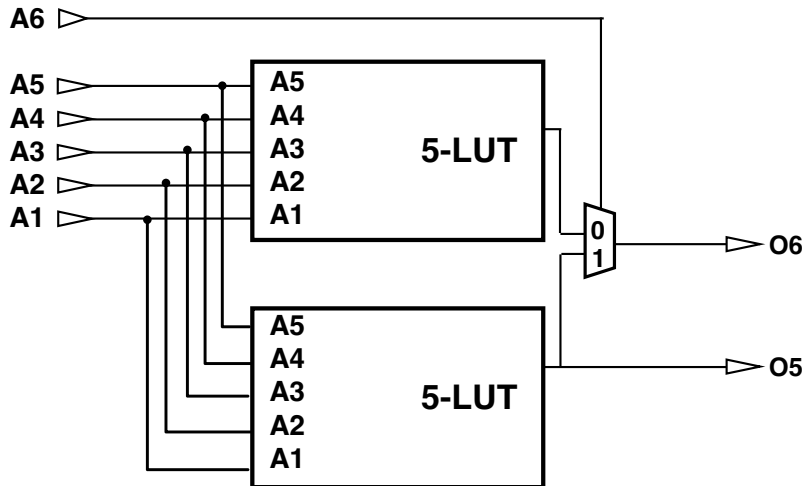
```
ap_uint<8> a, b; // 8-bit unsigned integer
ap_int<9> c, d; // 9-bit signed integer
ap_int<?> e = a*b + c*d; // How to set the bitwidth for e?
```

(b) In the following code segment, what is the minimum bitwidths of h that preserves the full precision of the expression? Please justify your answer.

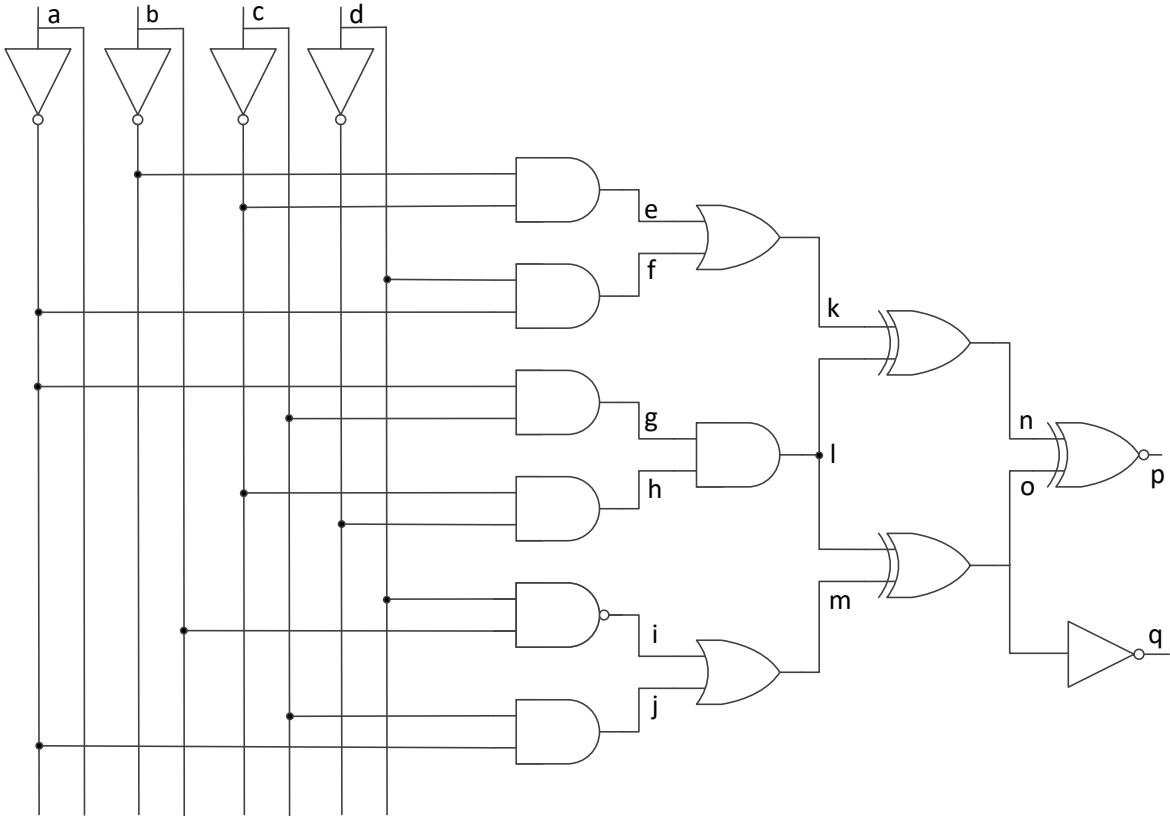
```
ap_ufixed<5,2> f; // 5-bit unsigned fixed point with 2 integer bits
ap_fixed<5,4> g; // 5-bit signed fixed point with 4 integer bits
ap_fixed<?,?> h = f*g; // How to set the bitwidth for h?
```

Q3. Consider functions $f = ab+bc$ and $g = ac$. Draw the reduced and ordered BDD corresponding to $f \oplus g$. Please use variable order (a, b, c) .

Q4. Can you use Boolean algebra to show that in the following diagram, output **O6** can realize any 6-input 1-output Boolean function $F(A1, A2, A3, A4, A5, A6)$? In other words, a 6-input LUT can be “fractured” into two 5-input LUTs and a 2:1 multiplexer.



Q5. Implement the circuit shown below using a **minimum** number of 3-input lookup tables (LUTs). Please do NOT simplify the gate-level circuit before mapping it to LUTs. Indicate your answer by creating a table similar to Table 1. Please add a new row per LUT, and use the signal names from the circuit to label the inputs (i.e., the associated cut) and output of each LUT. Mark an unused LUT input with a “-”.



LUT #	Input 1	Input 2	Input 3	Output
1	b	c	-	e

Table 1: Example: using a 3-input LUT to cover the cone from b, c to e.

Q6. Consider a matrix-vector multiplication (MV), where A is a square matrix, vectors x and y are N -dimensional, and all data elements are 32-bit floating-point (FP) numbers, i.e., 4 bytes each. For an accelerator with an on-chip memory of size $4N$ bytes and a few 32-bit registers, what would be the **maximum operational intensity (OI)** when running the given MV workload? Please explain your answer.

Here we assume all input data (i.e., A and x) are initially stored in an off-chip DRAM. Please also consider the memory accesses for storing output y off-chip. We count FP Mult and Add as two operations.

$$y = Ax = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N \\ \vdots \\ a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{NN}x_N \end{bmatrix}$$