Question 1
You are doing a dot-matrix calculation on large arrays
\[ A \cdot B = (a_0*b_0, a_1*b_1, \ldots, a_n*b_n) \]
You have the following design elements:
- An SRAM with a 5 ns read cycle. All the entries for A and B are stored in ONE SRAM.
- A parallel single cycle multiplier that takes 22 ns and takes 1 sq.um. of area
- A parallel serial multiplier that takes 8 cycles of 3 ns (24 ns total) to perform a multiply but takes 0.3 sq.um. of area
- For flip-flops, assume tsu+tck-Q = 1 ns, and the area is (effectively) 0

What is the overall design approach and (approximate) clock(s) choices that are likely to maximize the throughput with minimum area?