6.004 Recitation 19

L19 – Introduction to Pipelining

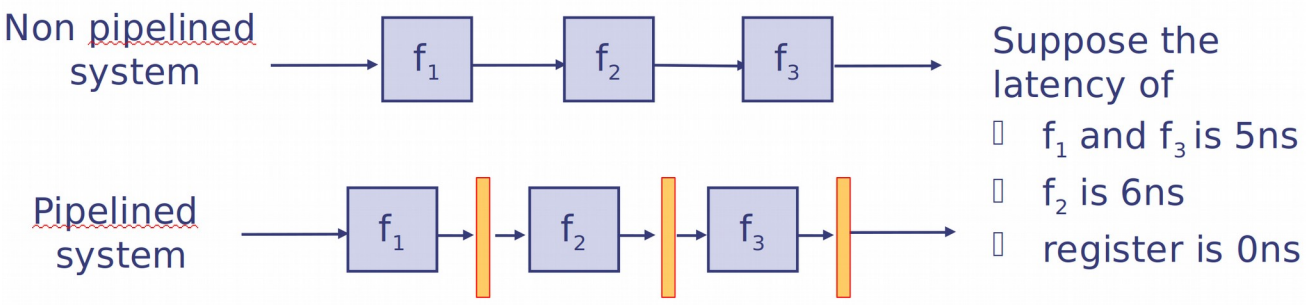
Key concepts review

Latency: The total amount of time needed for a system to produce an output following being given an input.

Throughput: The rate at which items are processed, typically expressed in terms of 1 item / time period.

Pipelining: Transform a system into one that can process multiple inputs simultaneously via the introduction of *stages*, where the system is partitioned into segments that run concurrently, with intermediate information that travels between stages being stored in registers or queues between cycles.

Important idea: Pipelining trades improved Clock Period and Throughput for worse Latency. Consider the example given in lecture:



	Non pipelined	Pipelined
Clock period (ns)	5+6+5=16	6
Throughput (1/ns)	1/16	1/6
Latency (ns)	16	3*6 = 18

Pipelining Methodology

1) Draw a line crossing all outputs

2) Draw a line to partition the graph into two parts;

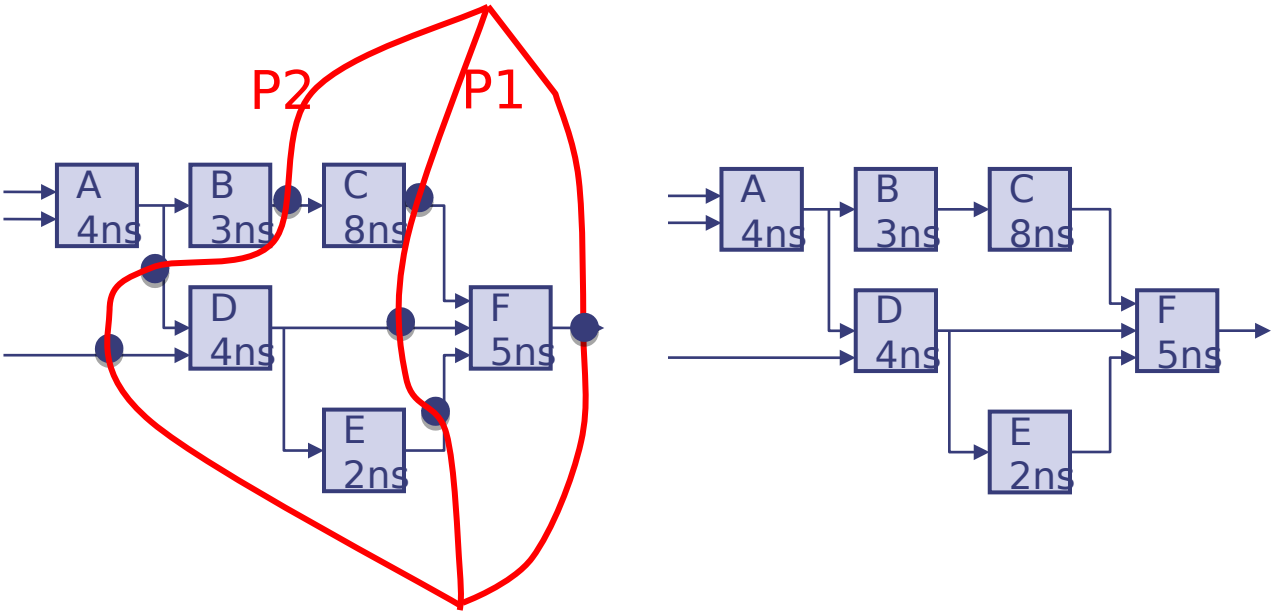
- all the arrows must cross the partition line in the same direction
- add a pipeline register at every point where the partition line crosses an edge

3) To increase the number of pipeline stages, further partition any partition using step 2

4) For the fastest possible pipeline make the slowest box be the determinant of the clock speed, i.e., the bottleneck

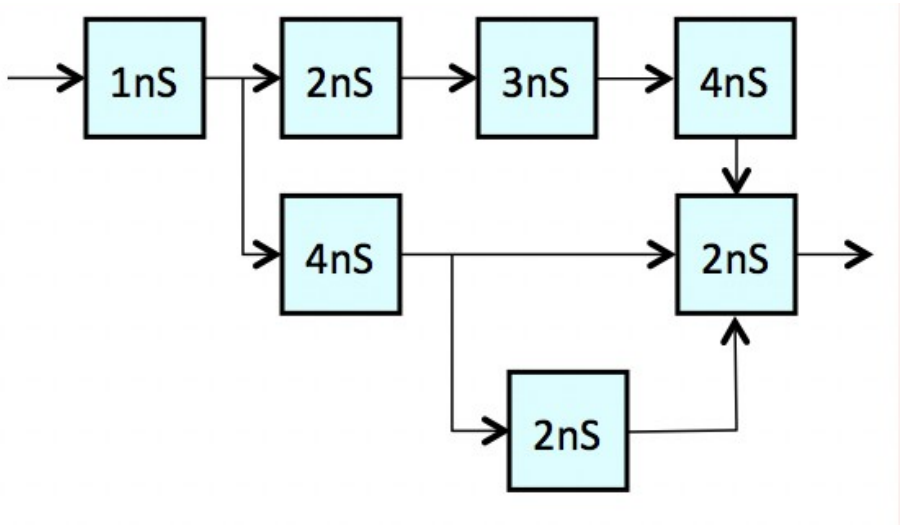
Warmup

Consider this pipelined system shown in lecture. Draw another three stage pipelining of the same system that does not increase the clock period.



Problem 1

Consider the following simple combinational circuit. Lem E. Tweakit is attempting to improve its performance by finding the optimal way to pipeline it. To this end, he wants to know what the optimal pipelining partitions are for each of the following stage counts, and has asked you to help him figure them out.



What are the optimal throughputs and the corresponding latencies for the following pipeline stage counts? What is the minimum number of registers required to achieve the optimum? Remember all outputs also require a register!

2 stages:

Tput: **Latency:** **Min Registers:**

3 stages:

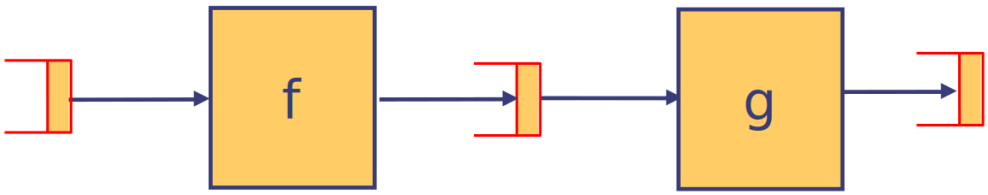
Tput: **Latency:** **Min Registers:**

4 stages:

Tput: **Latency:** **Min Registers:**

Problem 2

Consider the following system shown in lecture, where module **g** takes 1 cycle to run, while module **f** takes either 1 or 2 cycles to run, depending on the input.



Such a system has characteristics that are much more difficult to describe, as they are dependent on the set of inputs being given. Let **A** represent the portion of inputs that cause **f** to run in 1 cycle, and **B** represent the portion that cause it to run in 2 cycles (in other words, **B** = 1 - **A**).

A) Give expressions in terms of **A** and **B** for the average latency and throughput of the system.

B) What should the values of **A** and **B** be to achieve an average latency of 2**.3** cycles? What about an average throughput of **1/1.75** cycles?