CSE4210
Architecture and Hardware for DSP

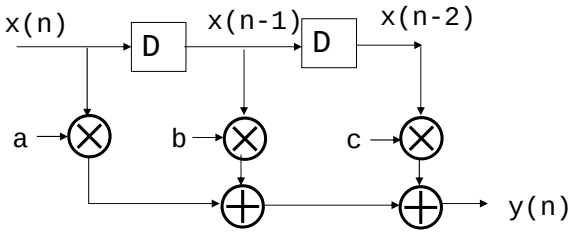
Chapter 3

Pipelining and Parallel Processing

Pipelining -- Introduction

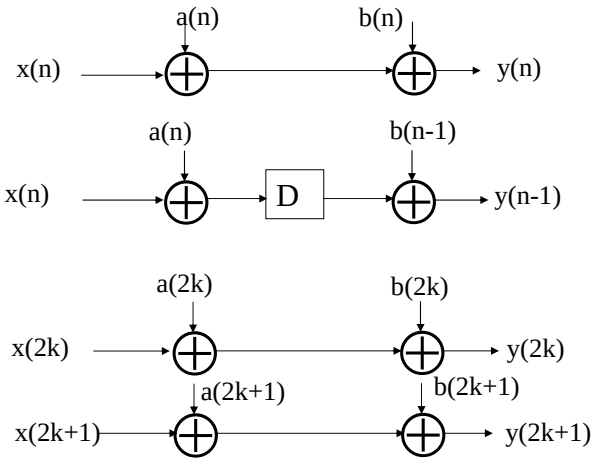
- Pipelining can be used to reduce the the critical path.
- That can lead to either increasing the clock speed, or decreasing the power consumption
- Multiprocessing can be also used to increase speed or reduce power.

Pipelining



The critical path here is two additions and one multiplication

$$T_s = 2T_{add} + T_{mul} \quad , \quad f_s = 1/(2T_{add} + T_{mul})$$



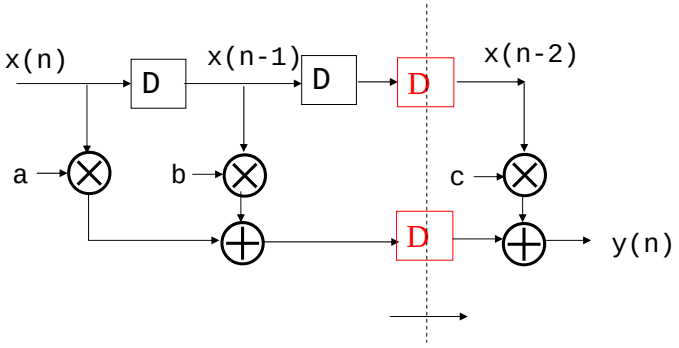
Pipelining

Parallel processing

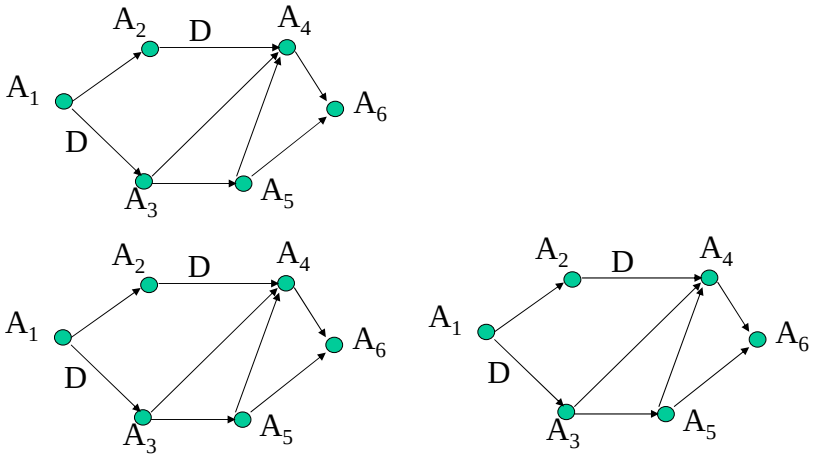
Pipelining

- Cutset: is a set of edges of a graph if removed, the graph becomes partitioned
- Feed forward cutset: a cutset where the data moves in the forward direction on all the edges in the cutset
- We can place latches on a feed-forward cutset without affecting the functionality of the graph.

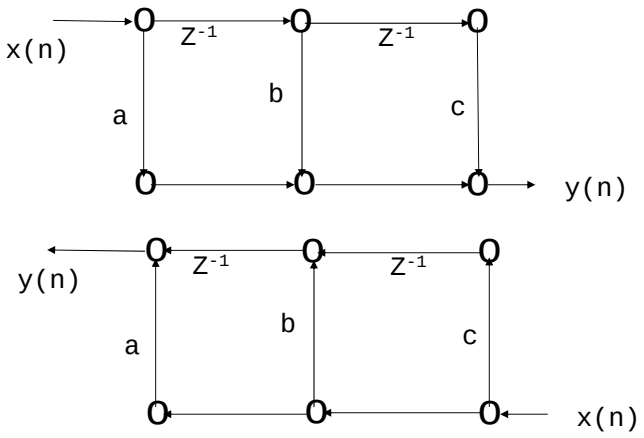
Pipelining



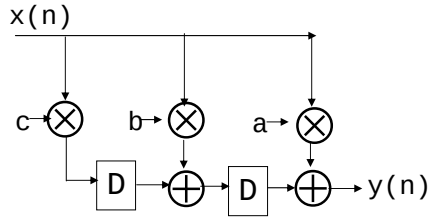
Example



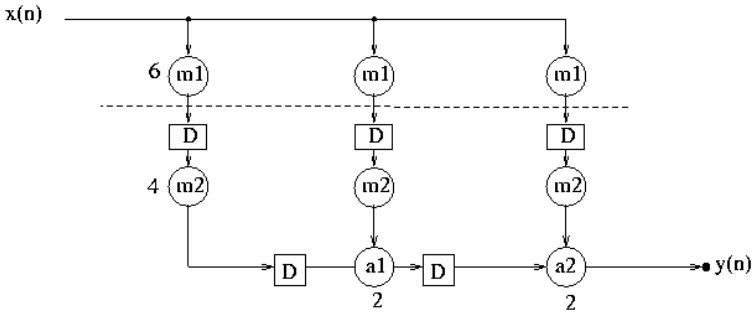
Data Broadcast



Pipelining



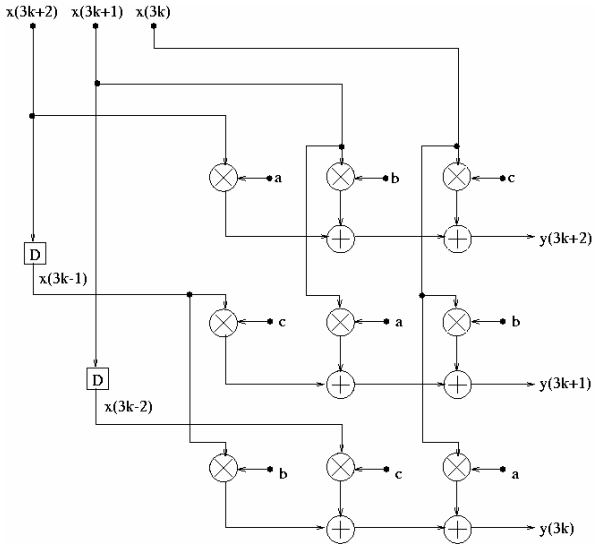
Fine grain Pipelining



Parallel Processing

$$y(n) = ax(n) + bx(n-1) + cx(n-2)$$

$$y(3k) = ax(3k) + bx(3k-1) + cx(3k-2)$$
$$y(3k+1) = ax(3k+1) + bx(3k) + cx(3k-1)$$
$$y(3k+2) = ax(3k+2) + bx(3k+1) + cx(3k)$$



Low Power

$$P = C_{total} V_o^2 f$$

$$T_{pd} = \frac{C_{\text{charge}} V_o}{k(V_o - V_t)^2}$$

Simple approximation for CMOS

C_{total} is the total capacitance of the circuit, Vo is the supply voltage. C_{charge} is the capacitance to be charged/discharged in a single clock cycle.

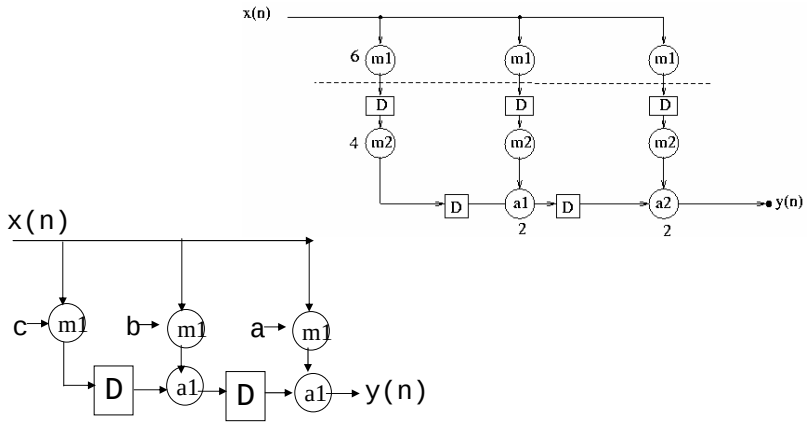
Pipelining and parallel processing could be used to minimize power or execution time.

Low Power

- What happens in case of M –pipelining?
- Critical path is reduced by M, so is C_{charge}
- If we keep the same f, then we have more time to charge C_{charge}, which means we can reduce the supply voltage

$$T_{seq} = \frac{C_{\text{charge}} V_o}{k(V_o - V_t)^2} = T_{pip} = \frac{\frac{C_{\text{charge}}}{M} \beta V_o}{k(\beta V_o - V_t)^2}, \beta = ?$$

Low Power –Example



Parallel processing

- What happen for L parallel System?
- Total capacitance increased by L
- Same performance, increase T by L
- More time to charge C_{charge}, can decrease V

$$LT_{seq} = L\frac{C_{\text{charge}}V_o}{k(V_o - V_t)^2} = T_{par} = \frac{C_{charge}\ \beta V_o}{k(\beta V_o - V_t)^2}, \beta = ?$$

Parallel Processing Example