Pipelining and Parallel Processing

Lan-Da Van (范倫達), *Ph. D.*
Department of Computer Science
National Chiao Tung University
Taiwan, R.O.C.

Spring, 2007

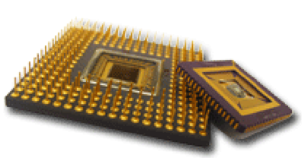
ldvan@cs.nctu.edu.tw
http://www.cs.nctu.edu.tw/~ldvan/

Outlines

- *Introduction*
- Pipelining of FIR Digital Filter
- Parallel Processing
- Pipelining and Parallel Processing for Low Power
- Conclusions

Introduction

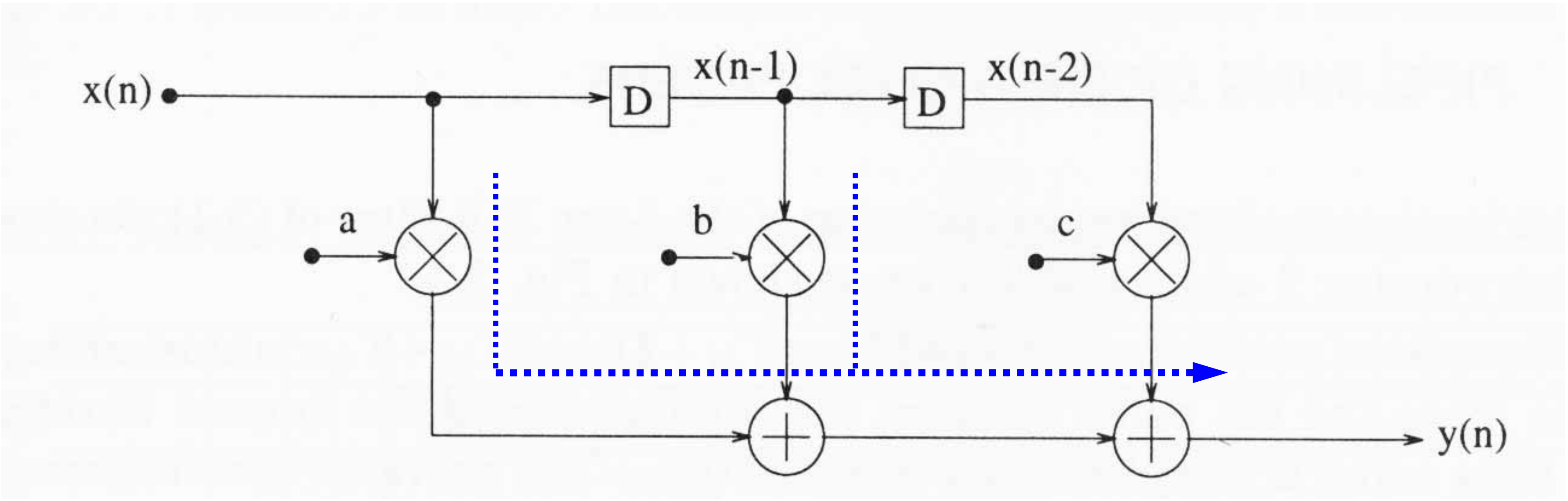
- Pipelining
  - Reduce the critical path
  - Increase the clock speed or sample speed
  - Reduce power consumption
- Parallel processing
  - Not reduce the critical path
  - Not increase clock speed, but increase sample speed
  - Reduce power consumption



A 3-tap FIR Filter

- Direct-form structure

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



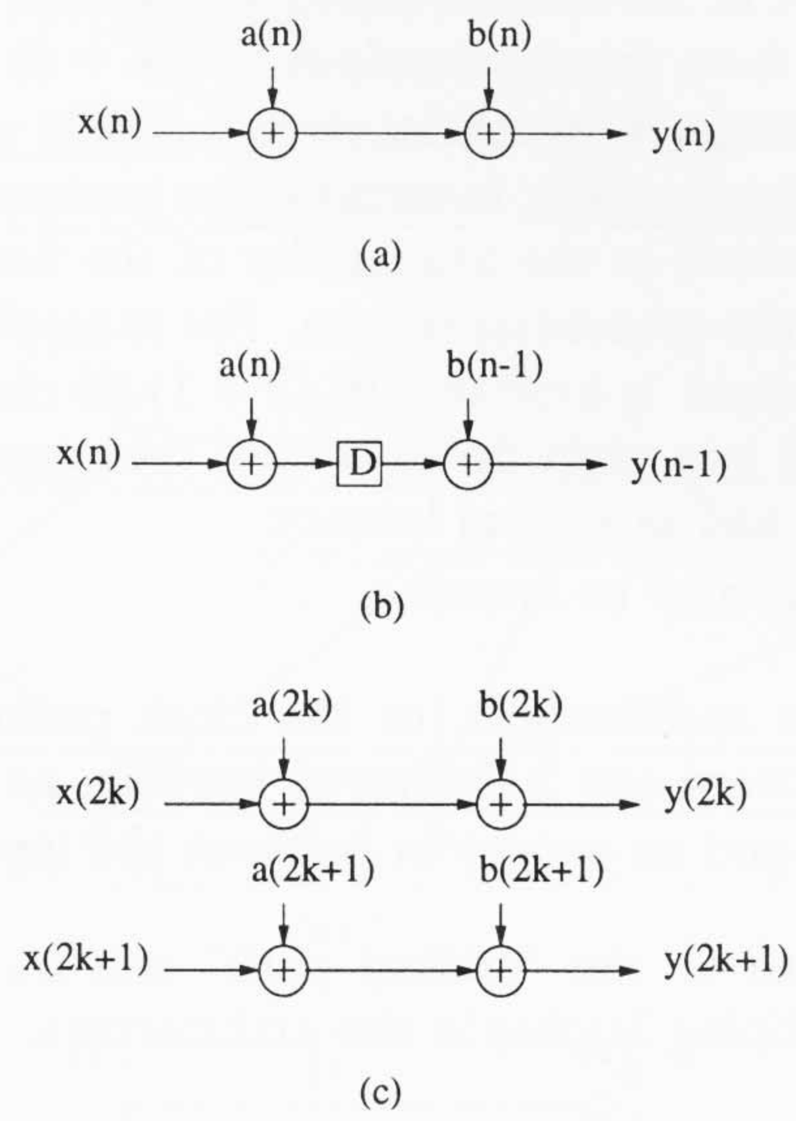
$$T_{sample} \geq T_M + 2T_A \qquad f_{sample} \leq \frac{1}{T_M + 2T_A}$$

Outlines

- Introduction
- *Pipelining of FIR Digital Filter*
- Parallel Processing
- Pipelining and Parallel Processing for Low Power
- Conclusions

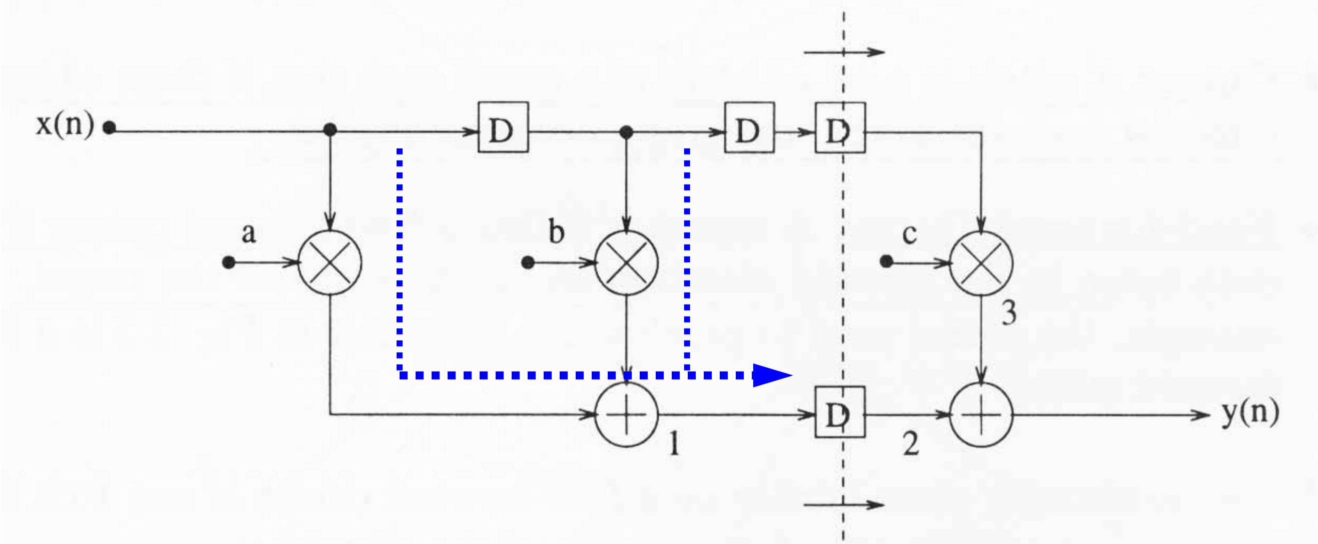
Pipelining and Parallel Concept

- Pipelining
  - Introduce pipelining latches along the datapath
- Parallel processing
  - Duplicate the hardware



Pipelining FIR Filter

- Critical path
  - $2T_A+T_M \dashrightarrow T_A+T_M$



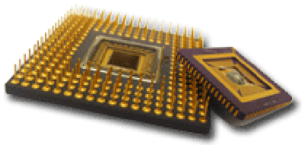
Clock	Input	Node 1	Node 2	Node 3	Output
0	$x(0)$	$ax(0)+bx(-1)$	$-$	$-$	$-$
1	$x(1)$	$ax(1)+bx(0)$	$ax(0)+bx(-1)$	$cx(-2)$	$y(0)$
2	$x(2)$	$ax(2)+bx(1)$	$ax(1)+bx(0)$	$cx(-1)$	$y(1)$
3	$x(3)$	$ax(3)+bx(2)$	$ax(2)+bx(1)$	$cx(0)$	$y(2)$

Pipelining (1/2)

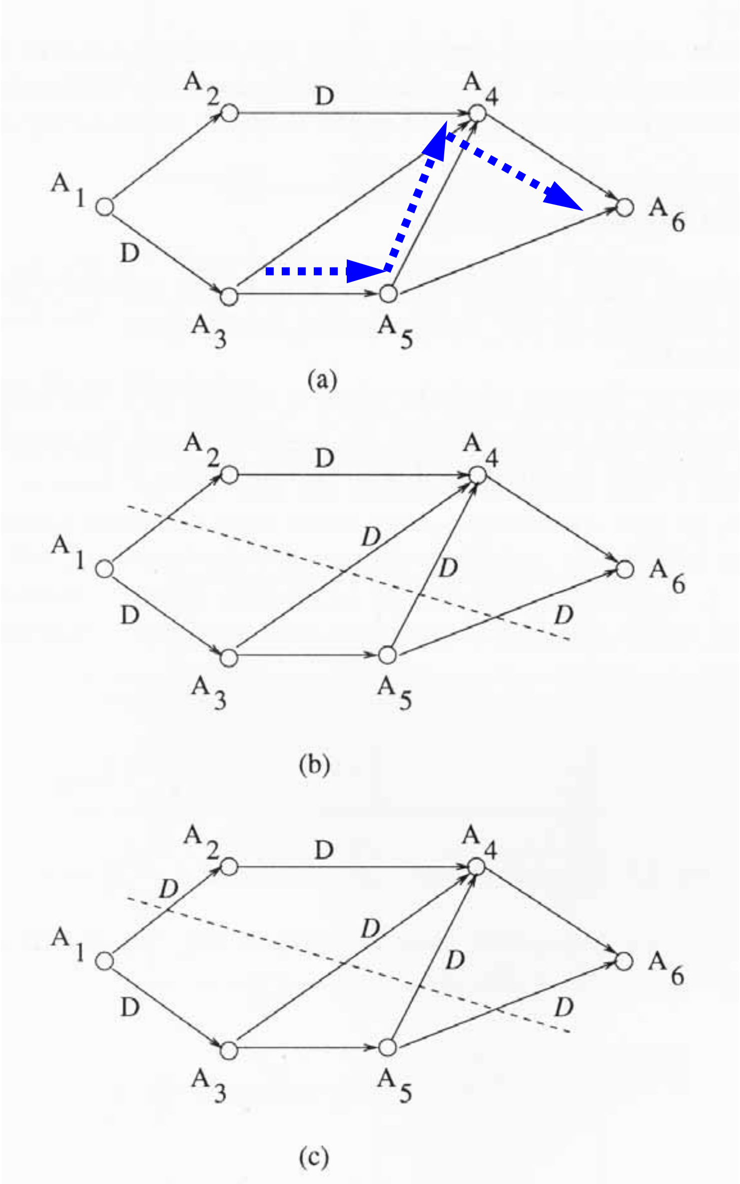
- Drawbacks
  - Increase number of delay elements (registers/latches) in the critical path
  - Increase latency
- Clock period limitation: critical path may be between
  - An input and a latch
  - A latch and an output
  - 2 Latches
  - An input and an output
- Pipelining latches can only be placed across any *feed-forward cutset* of the graph

Pipelining (2/2)

- Cutset: A cutset is a set of edges of a graph such that if these edges are removed from the graph, the graph becomes disjoint.
- Feed-forward cutset: A cutset is called a feed-forward cutset if the data move in the forward direction on all the edges of the cutset.



Example 3.2.1



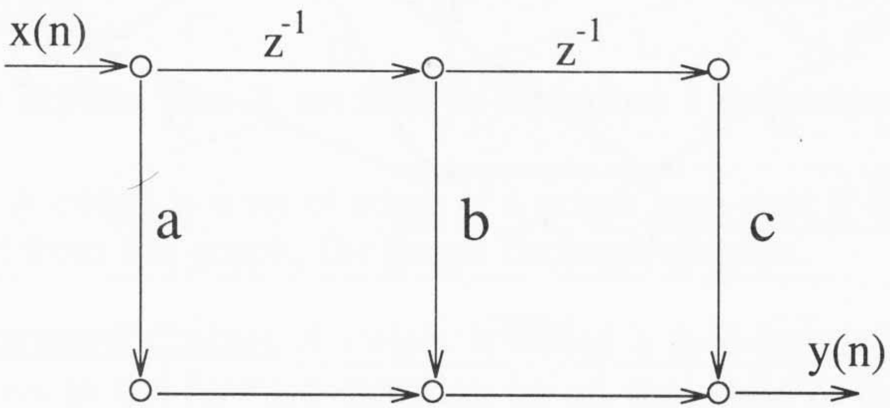
4 u.t.

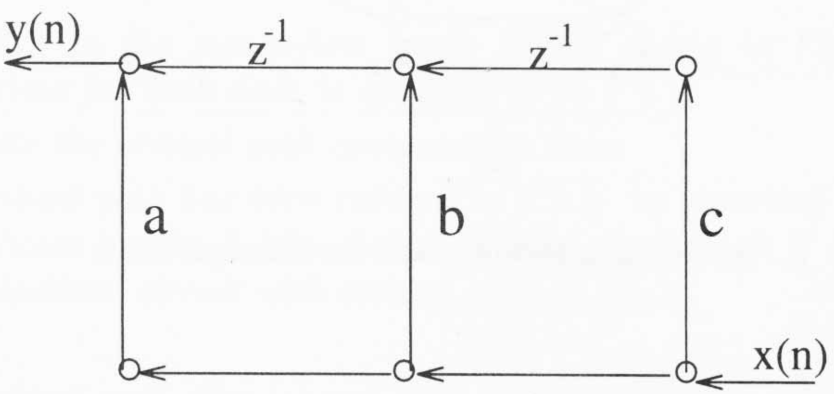
Error!

2 u.t.

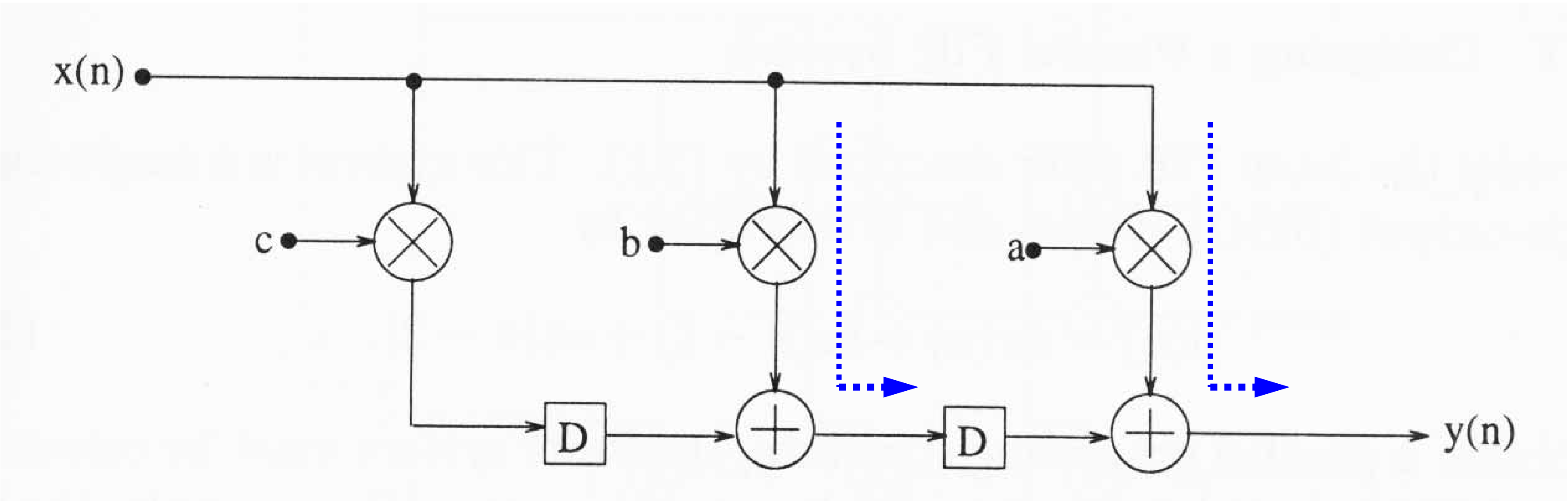
Transposition Theorem

- Reversing the direction of all edges in a given SFG and interchanging the input and output ports preserve the functionality of the system.





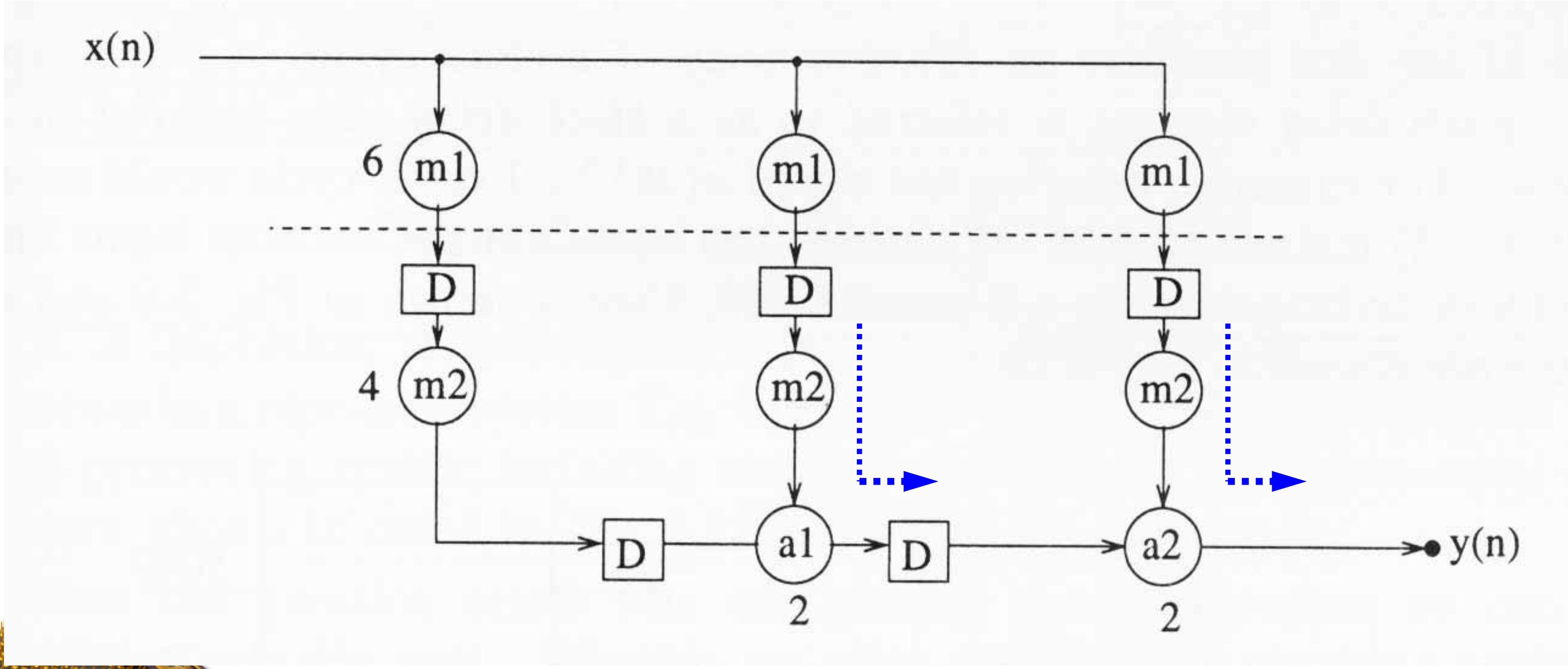
Data-Broadcast Structure



Critical path is reduced to (T_M+T_A).

Fine-Gain Pipelining

- Let T_M=10 u.t., T_A=2 u.t., and the desired clock period=6 u.t.
- Break the MULTIPLIER into 2 smaller units with processing time of 6 and 4 units.

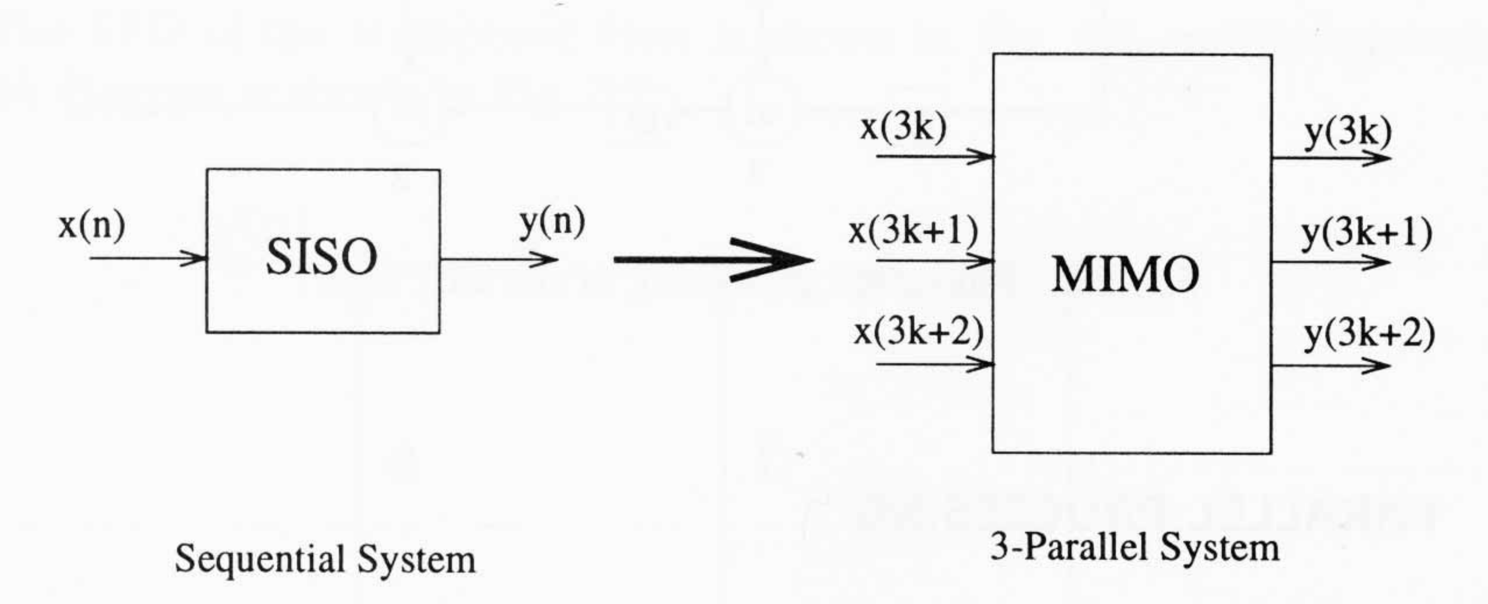


Outlines

- Introduction
- Pipelining of FIR Digital Filter
- *Parallel Processing*
- Pipelining and Parallel Processing for Low Power
- Conclusions

Parallel Processing

- Parallel processing and pipelining are dual
- If a computation can be pipelined, it can also be processed in parallel.
- Convert a single-input single-output (SISO) system to multiple-input multiple-output (MIMO) system via parallelism



Parallel Processing of 3-Tap FIR Filter (1/2)

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

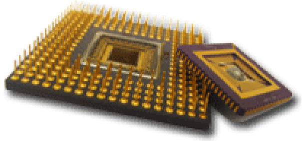
$$y(3k) = ax(3k) + bx(3k-1) + cx(3k-2)$$

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

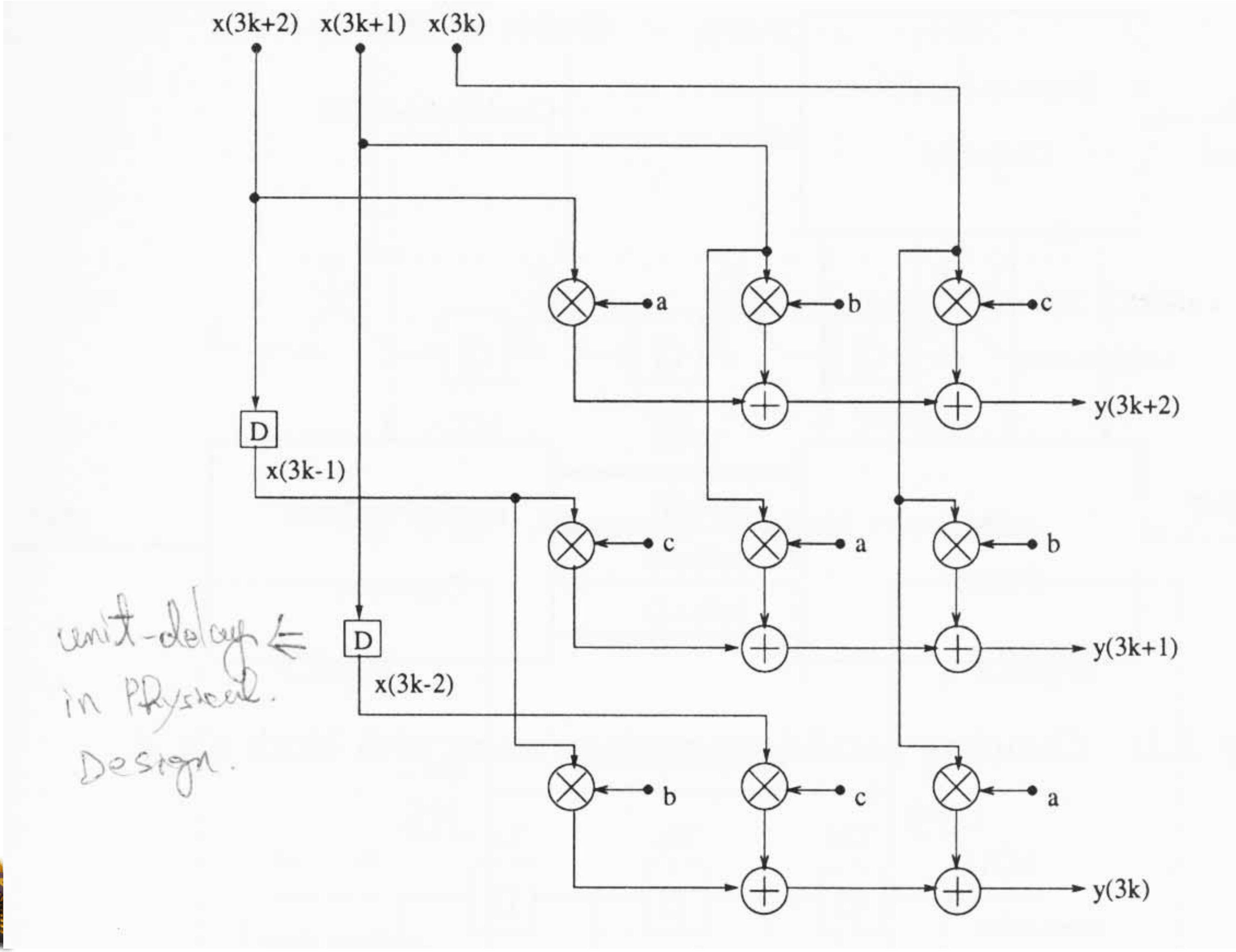
$$y(3k+2) = ax(3k+2) + bx(3k+1) + cx(3k)$$

$$T_{iter} = T_{sample}$$

$$= \frac{1}{L} T_{clk} \geq \frac{1}{3}(T_M + 2T_A)$$

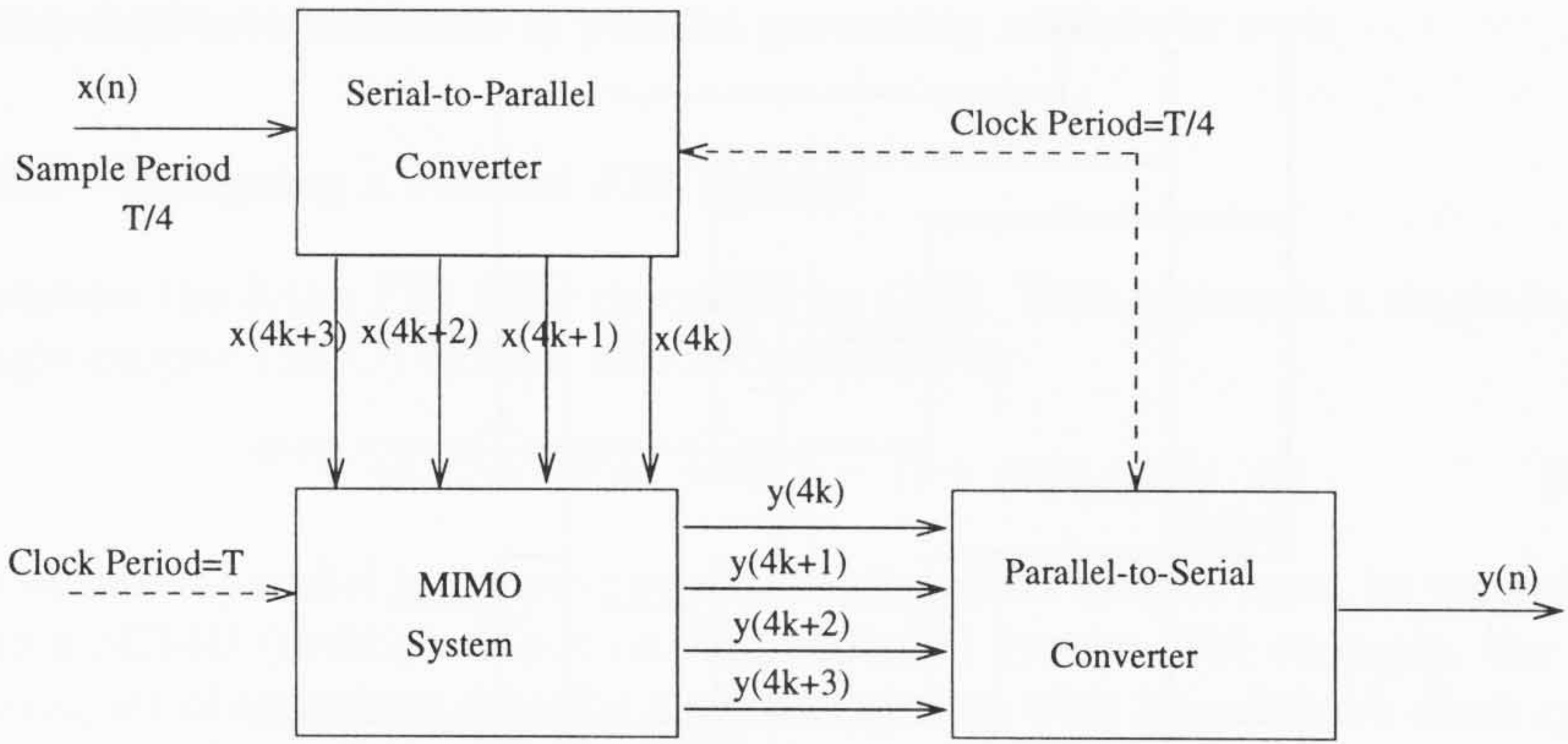


Parallel Processing of 3-Tap FIR Filter (2/2)

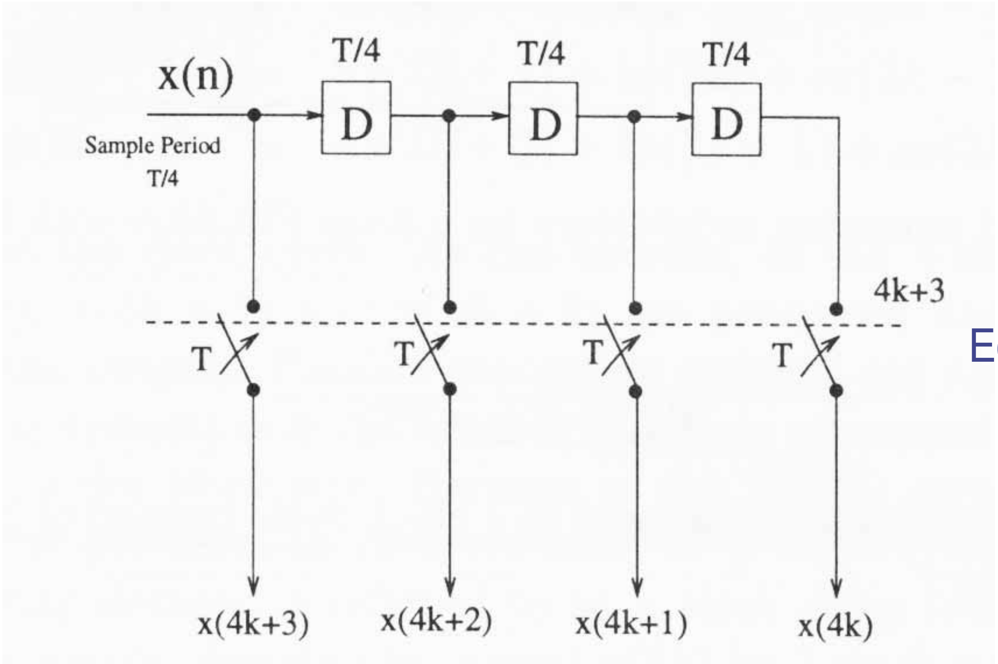


Complete Parallel Processing System

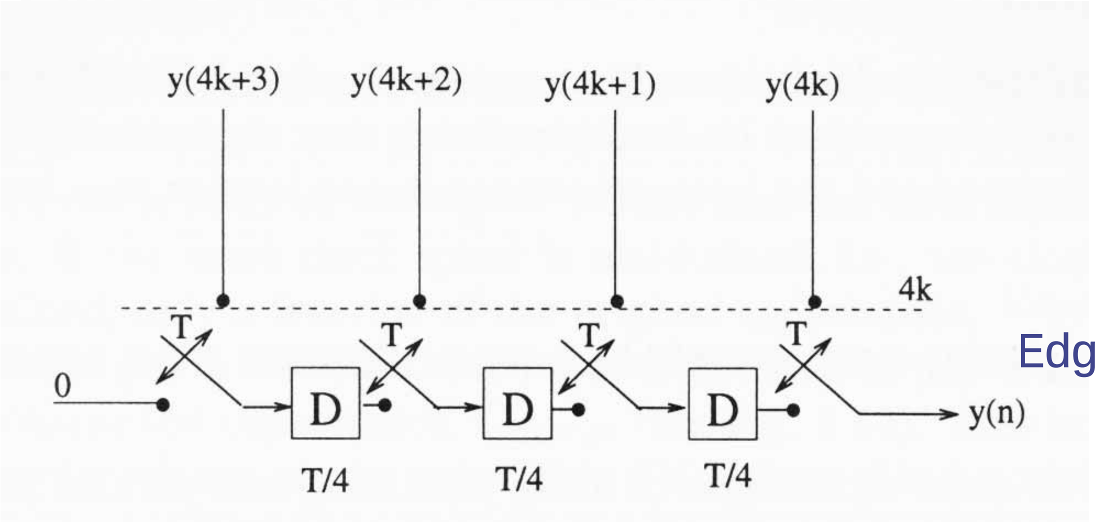
- Critical path has remained unchanged.
- But the iteration period is reduced.



S/P and P/S Converter



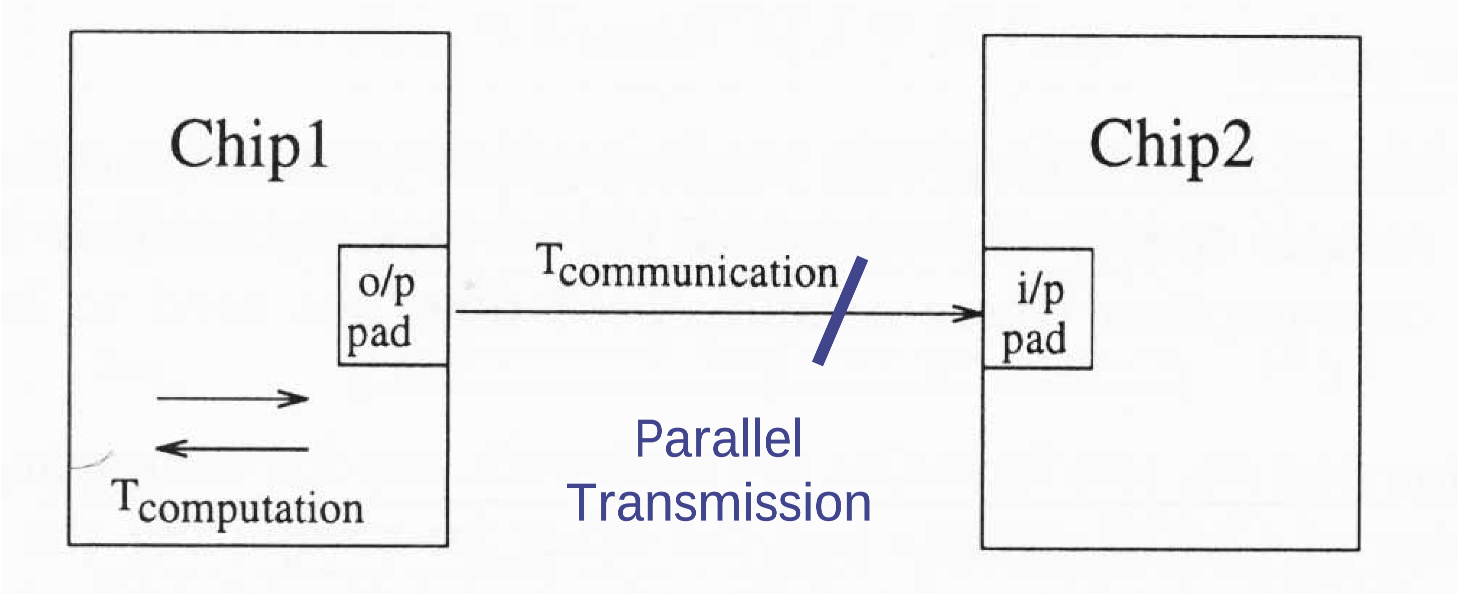
Edge Trigger!



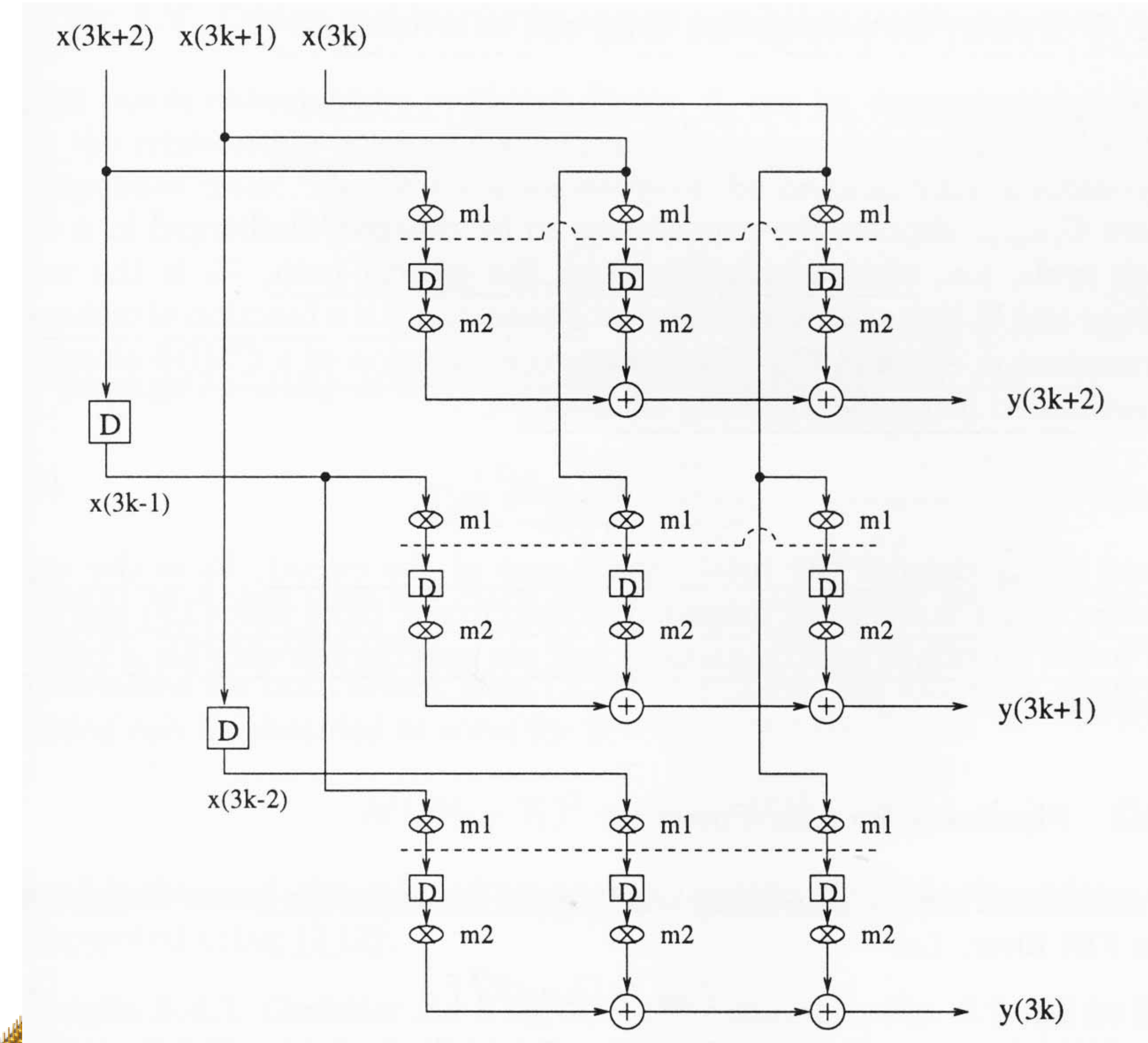
Edge Trigger!

Why Parallel Processing ?

- Parallel leads to duplicating many copies of hardware, and the cost increases! Why use?
- Answer lies in the fact that the fundamental limit to pipelining is at I/O bottlenecks, referred to as *Communication Bound,* composed of I/O pad delay and the wire delay.



Combined Fine-Grain Pipelining and Parallel Processing



$$T_{iter} = T_{sample}$$

$$= \frac{1}{LM} T_{clk} = \frac{1}{6}(T_M + 2T_A)$$

Outlines

- Introduction
- Pipelining of FIR Digital Filter
- Parallel Processing
- *Pipelining and Parallel Processing for Low Power*
- Conclusions

Underlying Low Power Concept

- Propagation delay

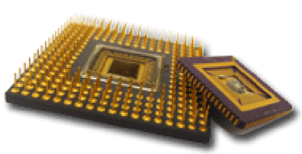
$$T_{pd} = \frac{C_{charge} V_0}{k(V_0 - V_t)^2}$$

- Power consumption

$$P = C_{total} V_0^2 f$$

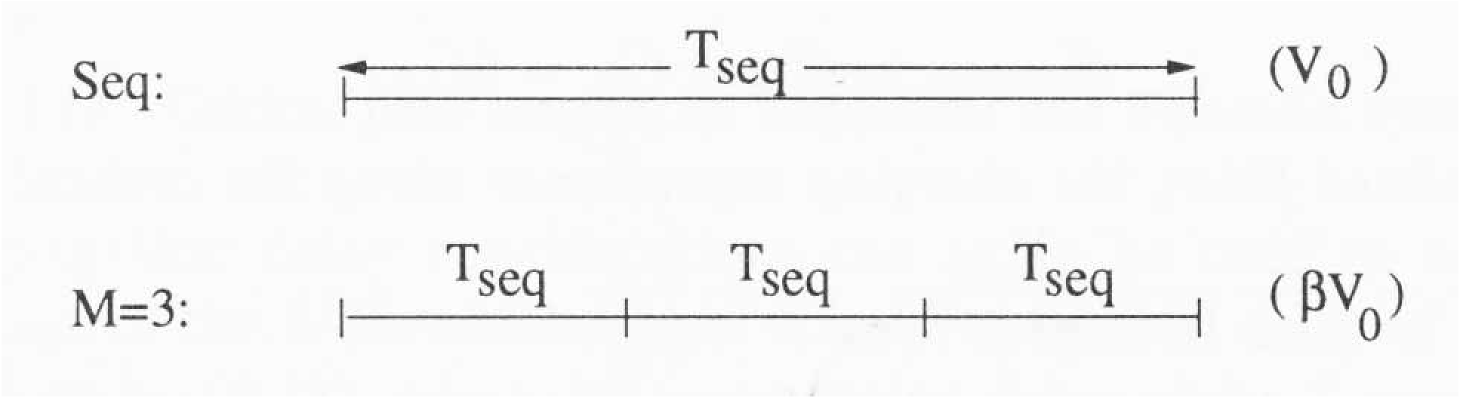
- Sequential filter

$$P_{seq} = C_{total} V_0^2 f, \qquad T_{seq} = \frac{C_{charge} V_0}{k(V_0 - V_t)^2}, \qquad f = \frac{1}{T_{seq}}$$



Pipelining for Low-Power (1/2)

- M-level pipelined system
- Critical path-->1/M, capacitance to be charged in a single clock cycle-->1/M
- If the clock frequency is maintained, the power supply can be reduced to βV_0 ($0<\beta<1$)



Pipelining for Low-Power (2/2)

- Power consumption

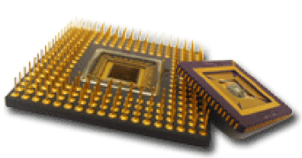
$$P_{pip} = C_{total}\beta^2 V_0^2 f = \beta^2 P_{seq}$$

- Propagation delay

$$T_{seq} = \frac{C_{charge} V_0}{k(V_0 - V_t)^2}, \quad T_{pip} = \frac{\frac{C_{charge}}{M}\beta V_0}{k(\beta V_0 - V_t)^2}$$

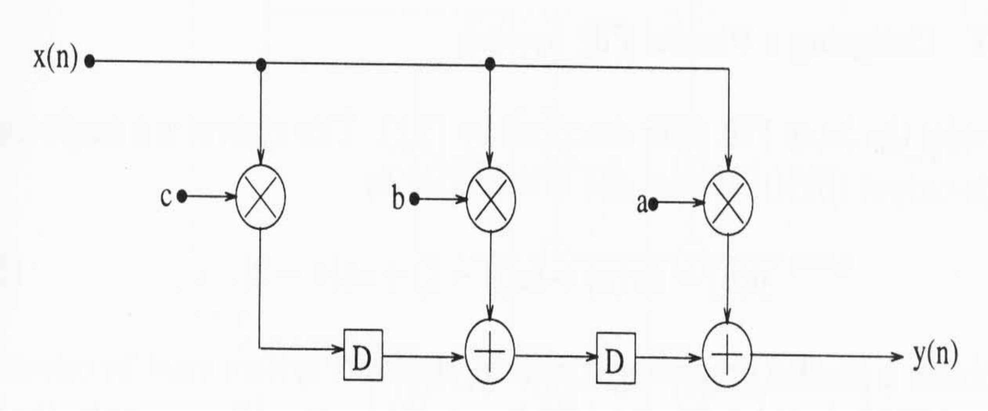
- Let $T_{seq}=T_{pip}$

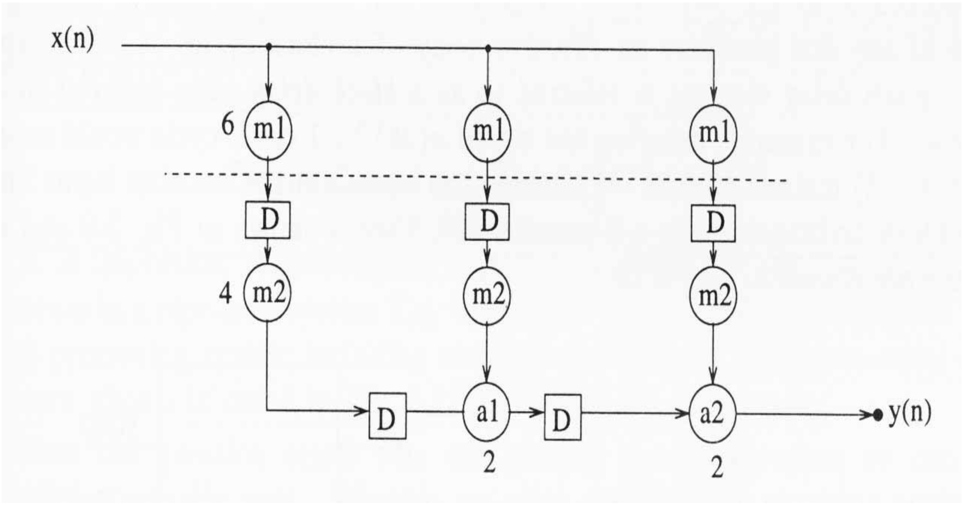
$$M(\beta V_0 - V_t)^2 = \beta(V_0 - V_t)^2 \implies \text{get } \beta$$



Example 3.4.1 (1/2)

- Consider an original 3-tap FIR filter and its fine-grain pipeline version shown in the following figures. Assume T_M=10 ut, T_A=2 ut, V_t=0.6V, V_o=5V, and C_M=5C_A. In fine-grain pipeline filter, the multiplier is broken into 2 parts, m1 and m2 with computation time of 6 u.t. and 4 u.t. respectively, with capacitance 3 times and 2 times that of an adder, respectively. (a) What is the supply voltage of the pipelined filter if the clock period remains unchanged? (b) What is the power consumption of the pipelined filter as a percentage of the original filter?





Example 3.4.1 (2/2)

- Solution:

$$\text{Original}: C_{charge} = C_M + C_A = 6C_A$$

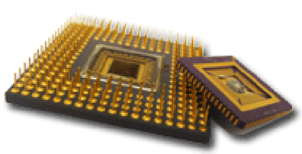
$$\text{Fine-Grain}: C_{charge} = C_{m1} = C_{m2} + C_A = 3C_A$$

$$2(5\beta - 0.6)^2 = \beta(5 - 0.6)^2$$

$$\Rightarrow \beta = 0.6033 \text{ or } 0.0239 \text{ (infeasible)}$$

$$V_{pip} = 3.0165\text{V}$$

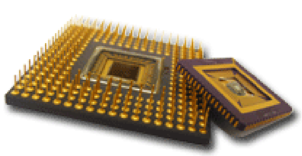
$$Ratio = \beta^2 = 36.4\%$$



Comparison

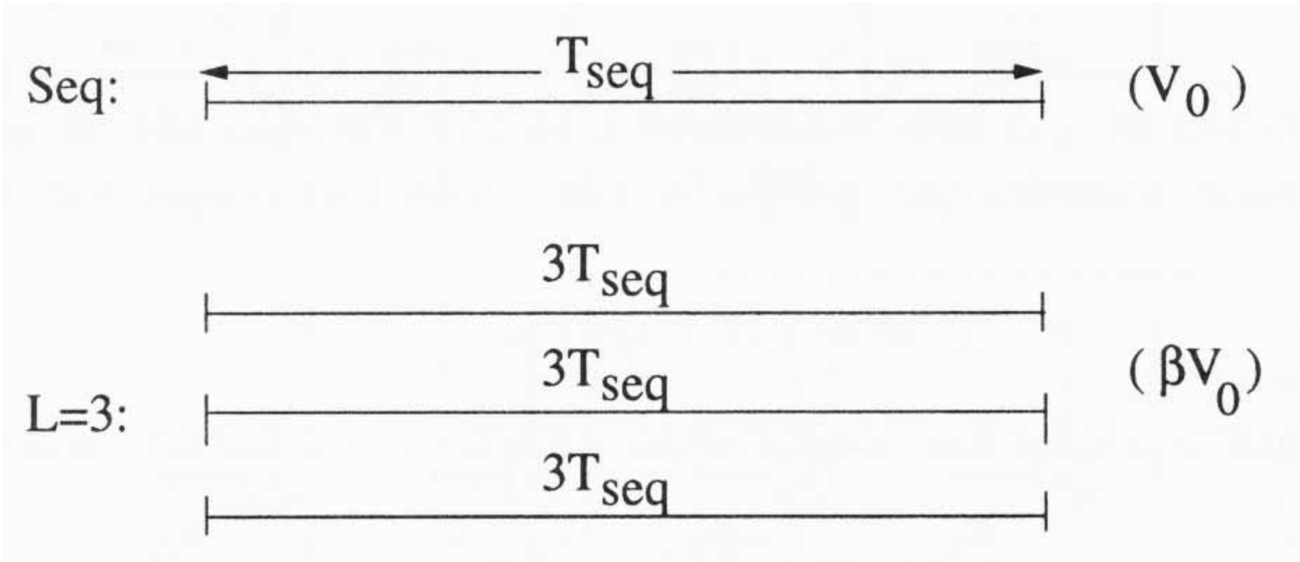
System	Sequential FIR (Original)	Pipelined FIR (Without reducing Vo)	Pipelined FIR (With reducing Vo)
Power (Ref)	Ref	2Ref	0.364Ref
Clock Period (u.t.)	12 ut	6 ut	12 ut
Sample Period (u.t.)	12 ut	6 ut	12 ut

Thinking Again!



Parallel Processing for Low-Power

- L-parallel system
- Maintain the same sample rate, clock period is increased to LT_{seq}
- This means that C_{charge} is charged in LT_{seq}, and the power supply can be reduced to βV_0



Parallel Processing for Low-Power

- Power consumption

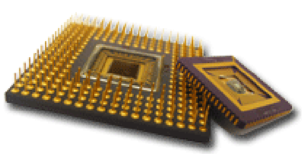
$$P_{par} = (LC_{total})(\beta V_0)^2 \frac{f}{L} = \beta^2 P_{seq}$$

- Propagation delay

$$T_{seq} = \frac{C_{charge} V_0}{k(V_0 - V_t)^2}, \qquad LT_{seq} = \frac{C_{charge} \beta V_0}{k(\beta V_0 - V_t)^2}$$

- $LT_{seq} = T_{pap}$

$$L(\beta V_0 - V_t)^2 = \beta (V_0 - V_t)^2 \text{ ==> get } \beta$$



Example 3.4.2 (1/2)

- Consider a 4-tap FIR filter shown in Fig. 3.18(a) and its 2-parallel version in 3.18(b). The two architectures are operated at the sample period 9 u.t. Assume T_M=8, T_A=1, V_t=0.45V, V_o=3.3V, C_M=8C_A (a) What is the supply voltage of the 2-parallel filter? (b) What is the power consumption of the 2-parallel filter as a percentage of the original filter?
- Solution:

$$\text{Original}: \ C_{charge} = C_M + C_A$$

$$2-\text{Parallel}: \ C_{charge} = C_M + 2C_A = 10C_A$$

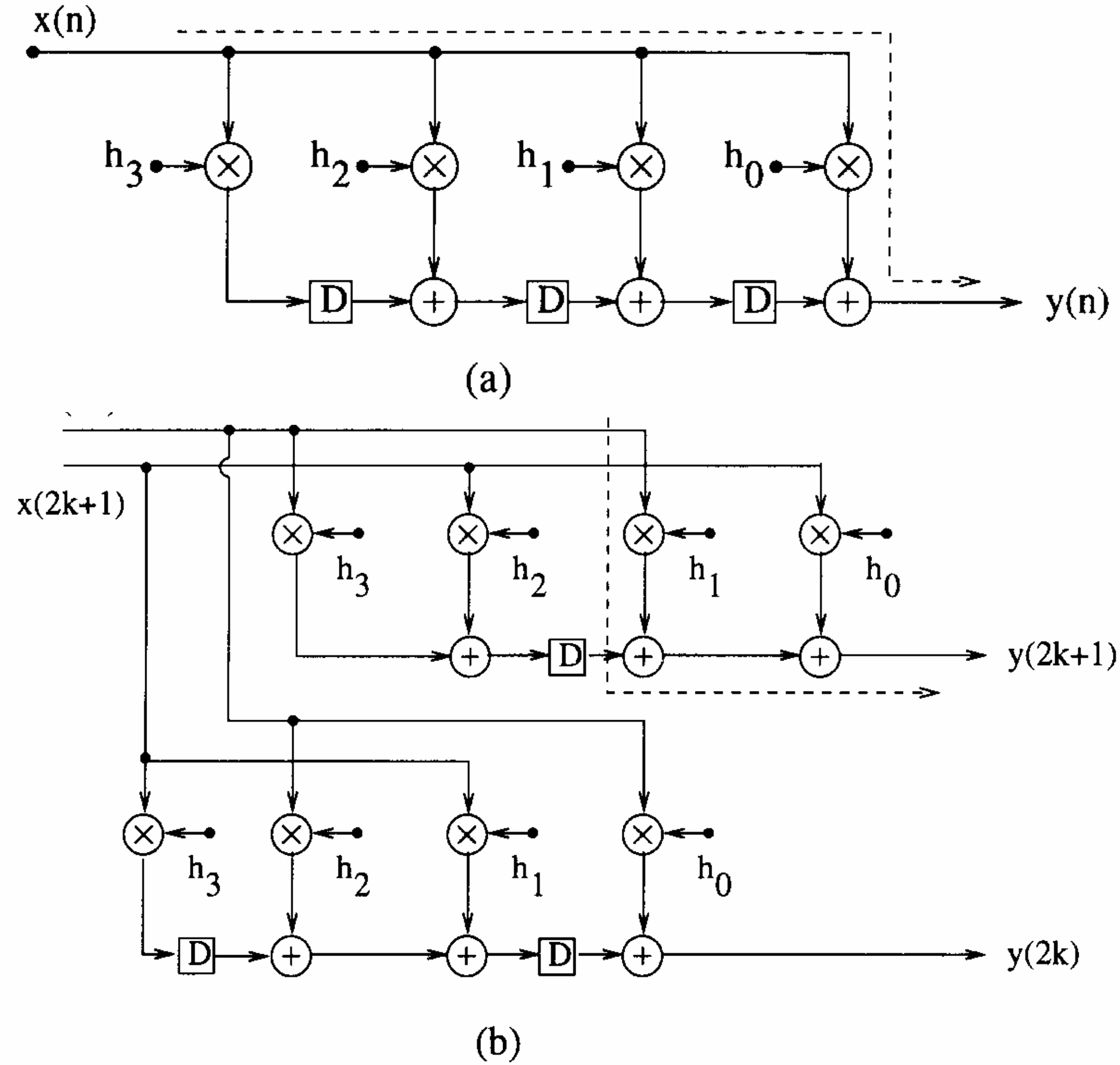
$$9(3.3\beta - 0.45)^2 = 5\beta(3.3 - 0.6)^2$$

$$=> \beta = 0.6589 \ \text{ or } \ 0.0282$$

$$\text{Vpar} = 2.1743\text{V}$$

$$\text{Ratio} = \beta^2 = 43.41\%$$

Example 3.4.2 (2/2)



Example 3.4.3 (1/2)

- A more efficient structure than the previous one is depicted in Fig. 3.18(c). (a) What is the supply voltage of the efficient 2-parallel filter? (b) What is the power consumption of the efficient 2-parallel filter as a percentage of the original filter?
- Solution:

$$\text{Original}: C_{charge} = C_M + C_A = 9C_A$$

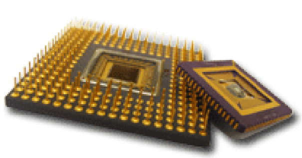
$$\text{New 2-Parallel}: C_{charge} = C_M + 4C_A = 12C_A$$

$$2 \times 9(3.3\beta - 0.45)^2 = 12\beta(3.3 - 0.6)^2$$

$$\beta = 0.745 \text{ or } 0.025 \text{ (infeasible)}$$

$$V_{pip} = 2.45857V$$

$$\text{Ratio} = \frac{P_{par}}{P_{seq}} = \frac{55C_A \cdot \beta^2 V_0{}^2 \cdot \frac{1}{2} f_s}{35C_A \cdot V_0{}^2 \cdot f_s} = 43.6\%$$



Example 3.4.3 (2/2)

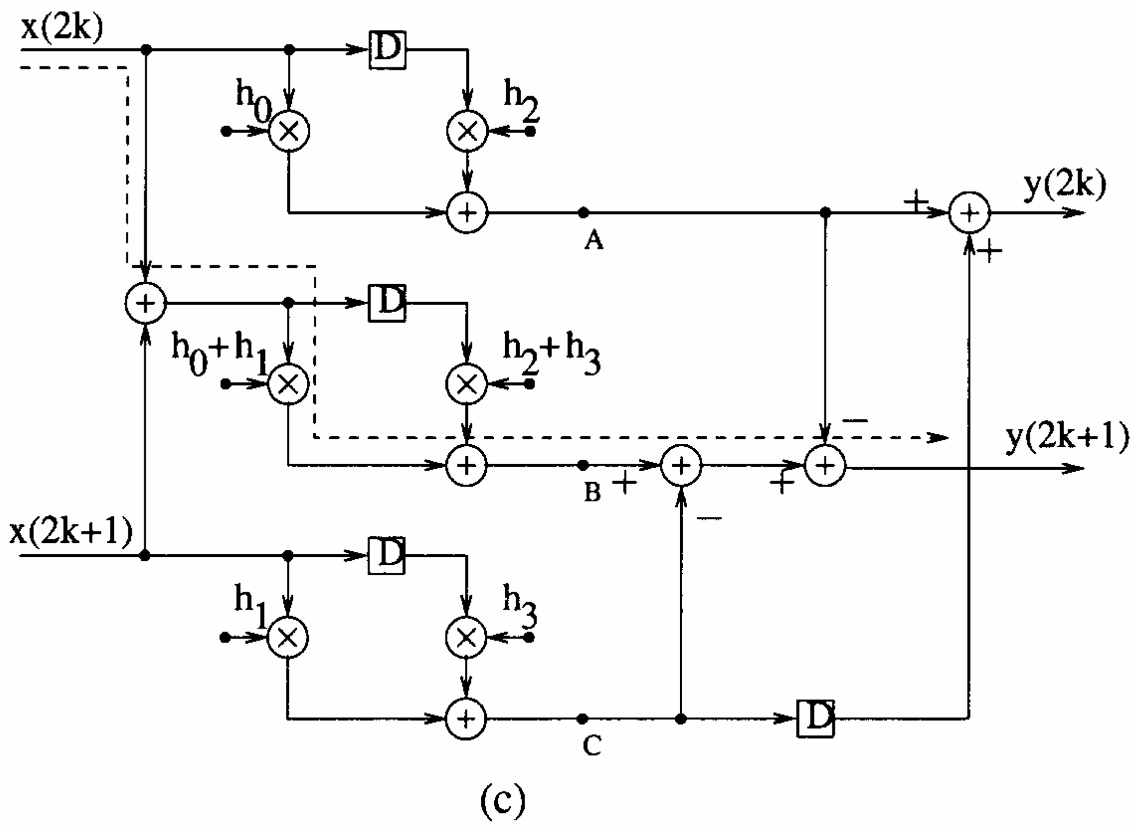


Fig. 3.18 (a) A 4-tap FIR filter. (b) A 2-parallel filter. (c) An area-efficient 2-parallel filter.

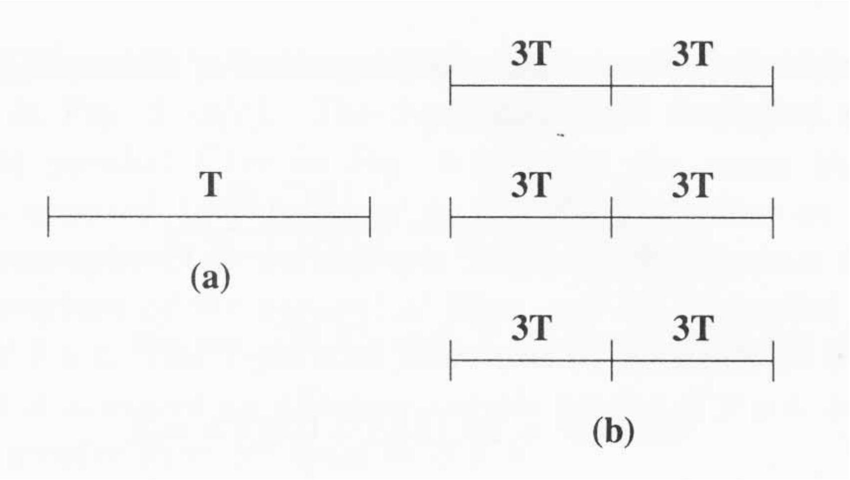
Combining Pipelining and Parallel Processing

- Parallel-pipelined structure

$$T_{seq} = \frac{C_{charge} V_0}{k(V_0 - V_t)^2}, \qquad T_{pp} = \frac{\frac{C_{charge}}{M} \beta V_0}{k(\beta V_0 - V_t)^2}$$

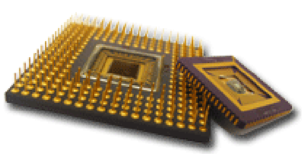
$$T_{pp} = L T_{\mathrm{seq}} \quad ==> \quad ML(\beta V_0 - V_t)^2 = \beta (V_0 - V_t)^2$$

- M=L=2, V_0=5V, V_t=0.6V-->β=0.4, β^2=0.16



Conclusions

- Methodologies of pipelining of 3-tap FIR filter
- Methodologies of parallel processing for 3-tap FIR filter
- Methodologies of using pipelining and parallel processing for low power demonstration.
- Pipelining and parallel processing of recursive digital filters using look-ahead techniques are addressed in Chapter 10.



Self-Test Exercises

- STE1: Problem 8 of Chap 3 in text book.
- STE2: Problem 9 of Chap 3 in text book.
- STE2: Problem 10 of Chap 3 in text book.