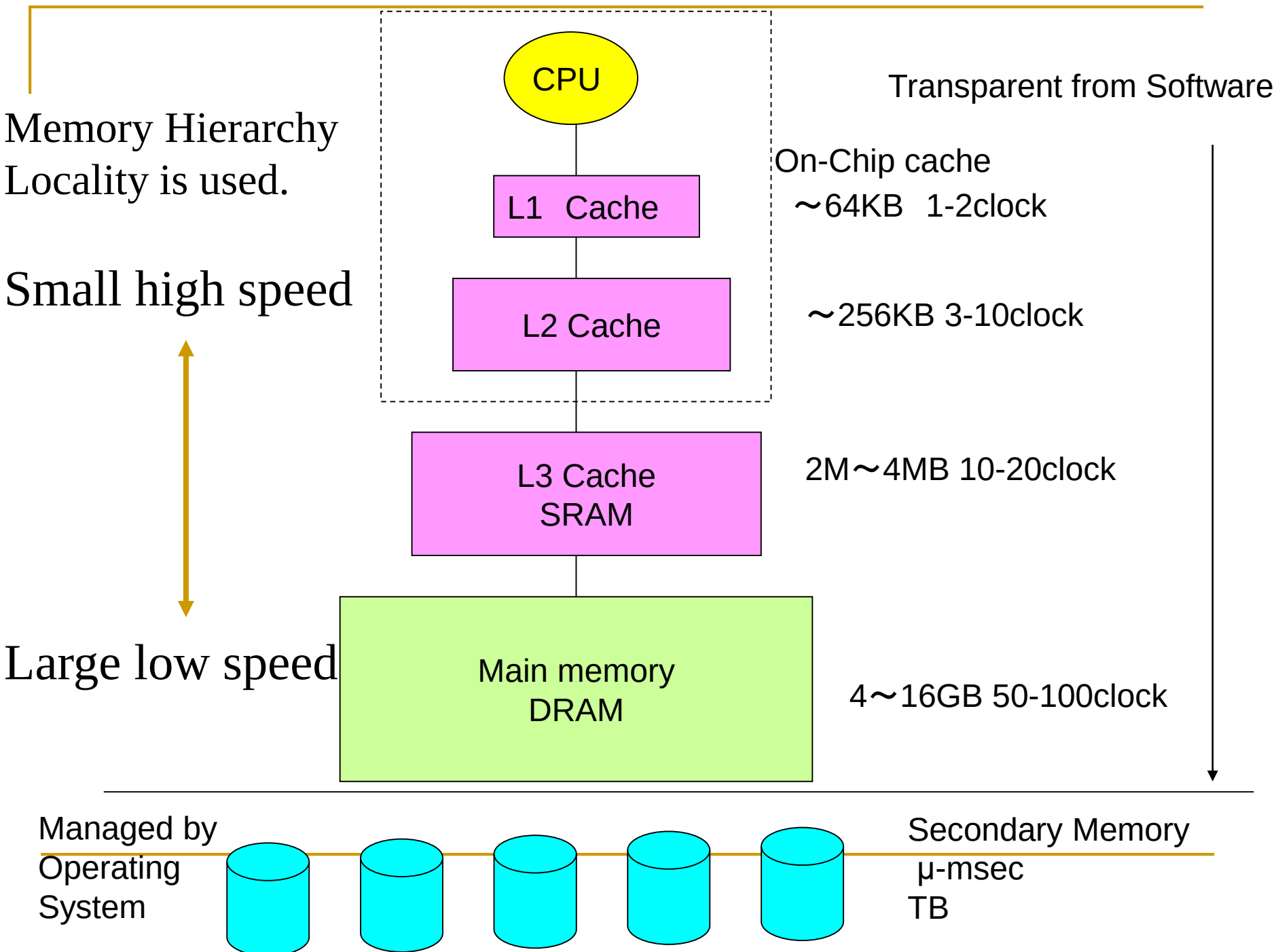
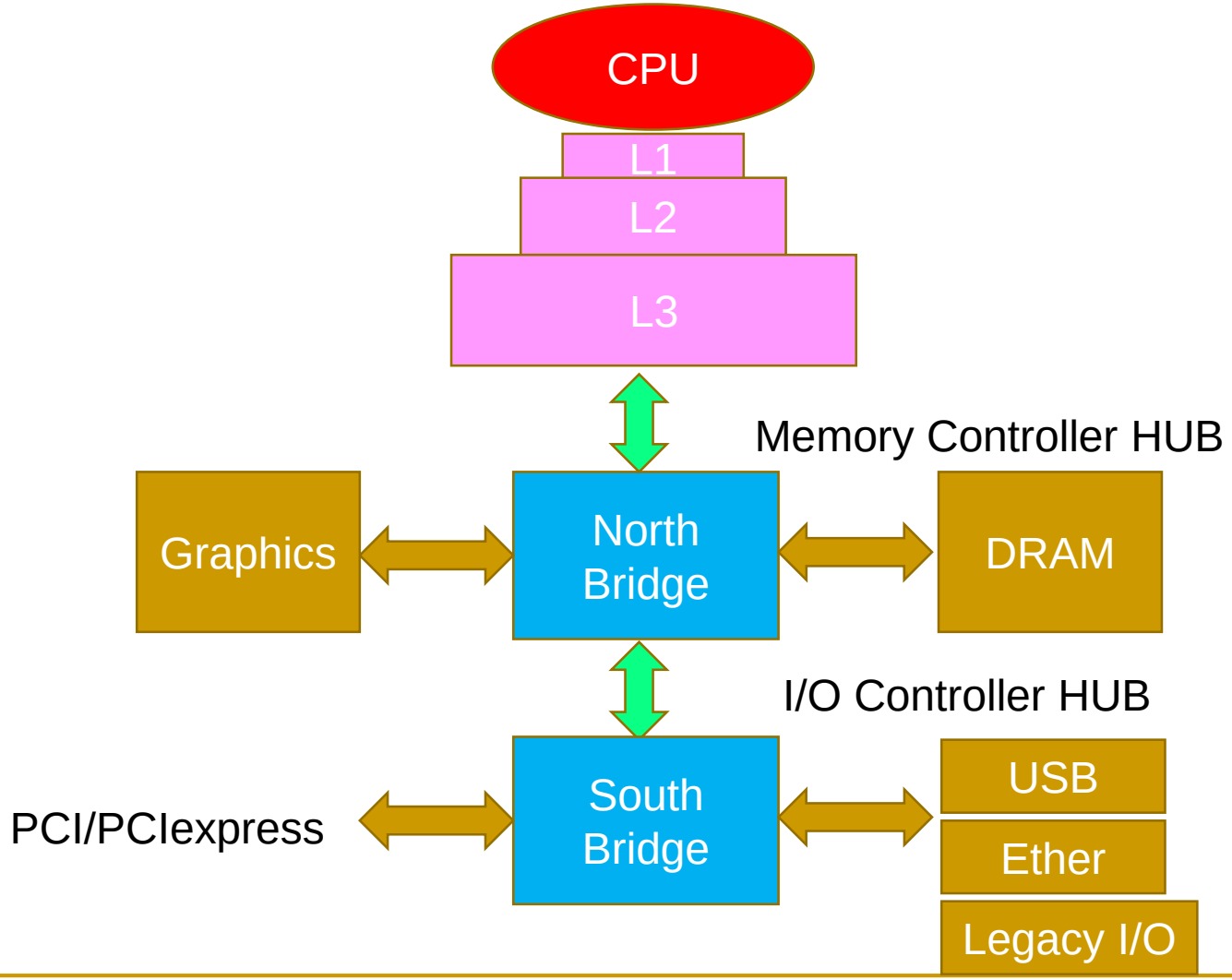
Bus / Crossbar Switch

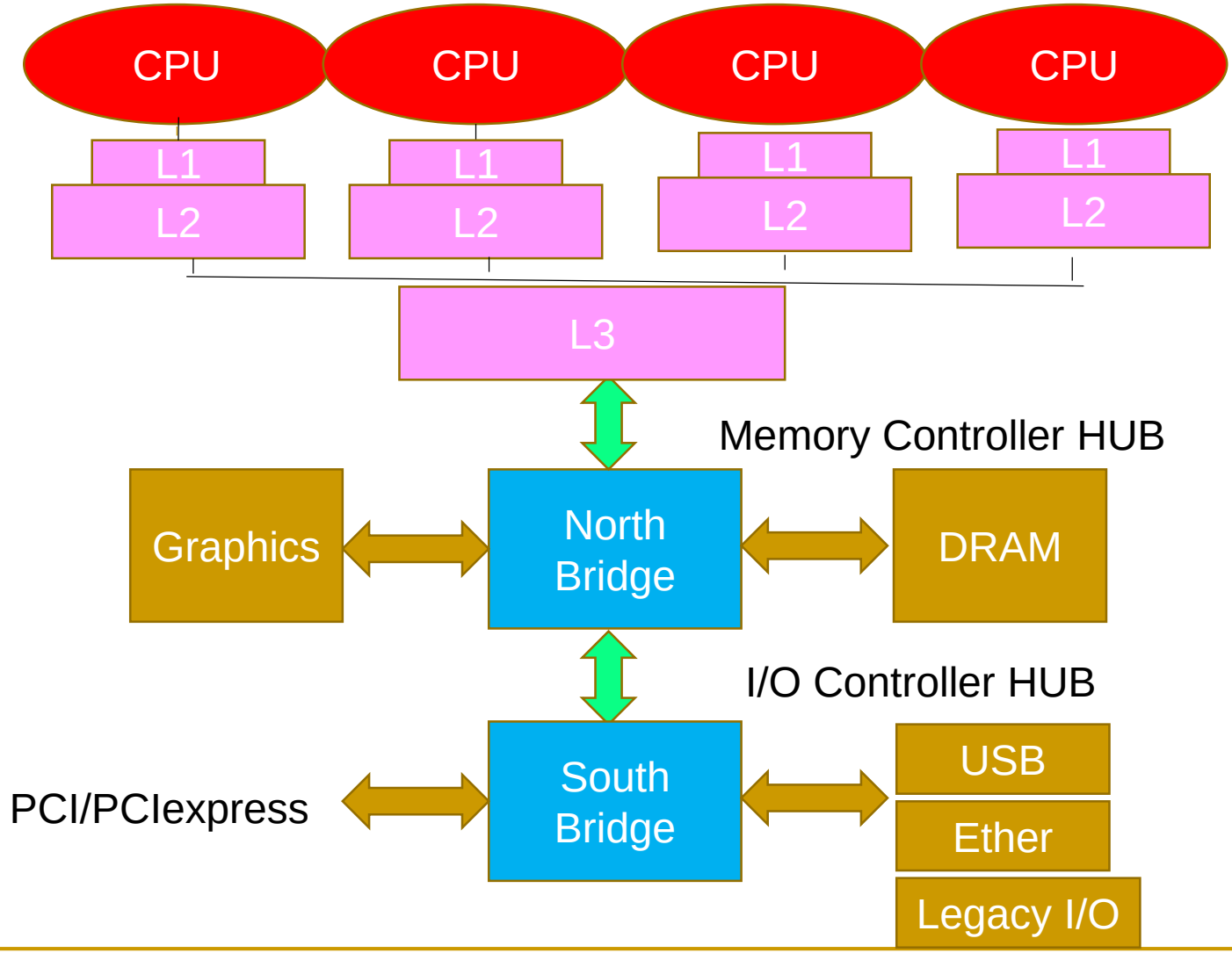
AMANO, Hideharu
hunga@am. ics. keio. ac. jp


Memory Hierarchy
Locality is used.
Small high speed
Large low speed
CPU
L1 Cache
L2 Cache
L3 Cache
SRAM
Main memory
DRAM
Transparent from Software
On-Chip cache
～64KB 1-2clock
～256KB 3-10clock
2M～4MB 10-20clock
4～16GB 50-100clock
Managed by
Operating
System
Secondary Memory
μ-msec
TB

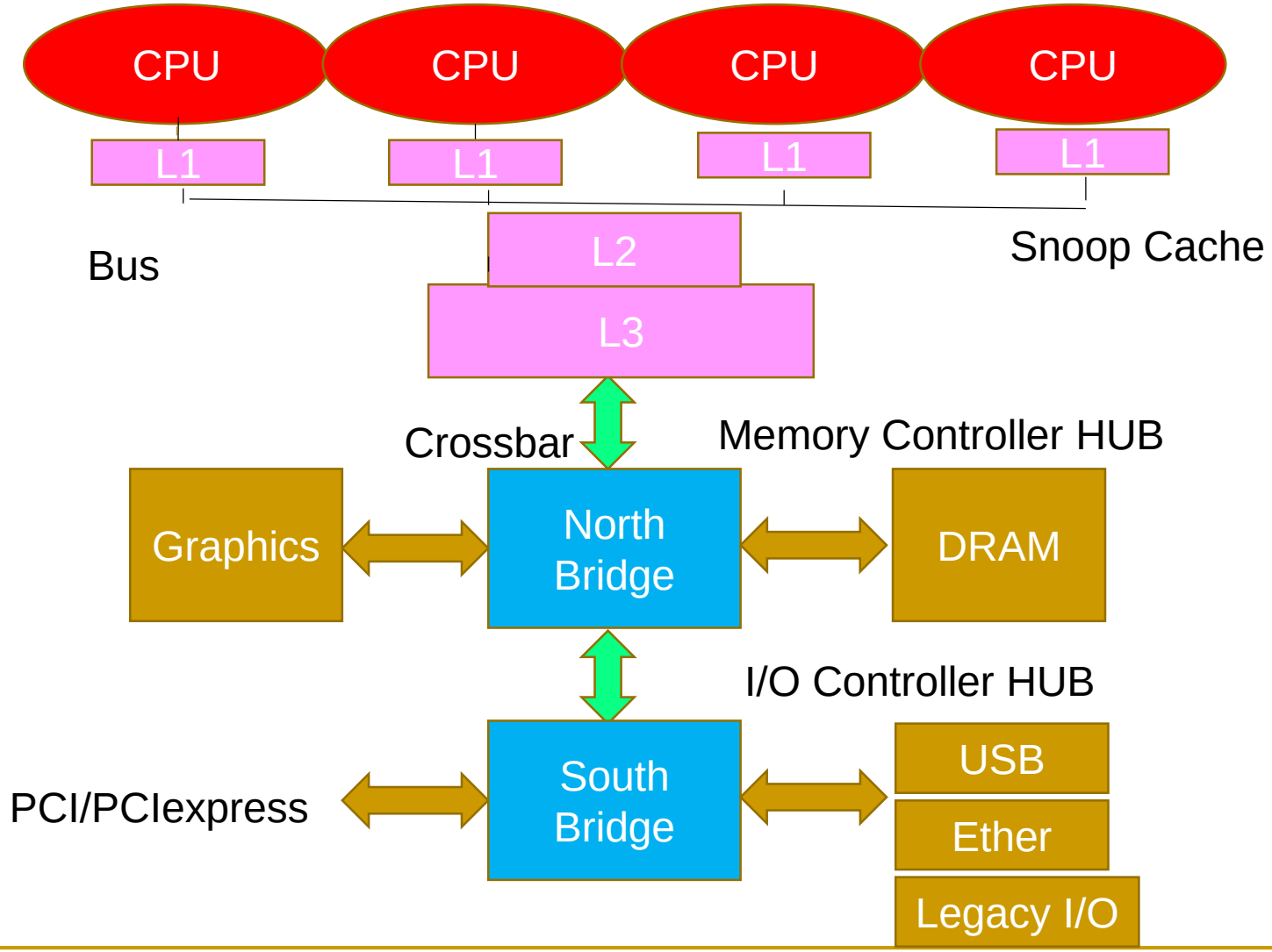
Uni-processor structure



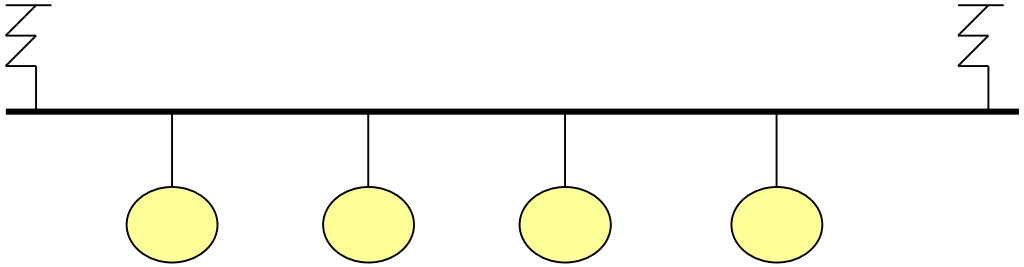
Consistency Problem



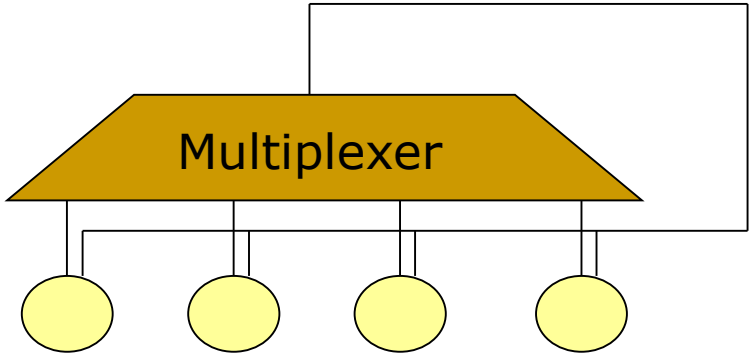
Snoop Cache is provided



Implementation of buses



Passive Bus:
Board level
implementation



Active Bus:
Chip level
implementation

A single module sends data to all other modules

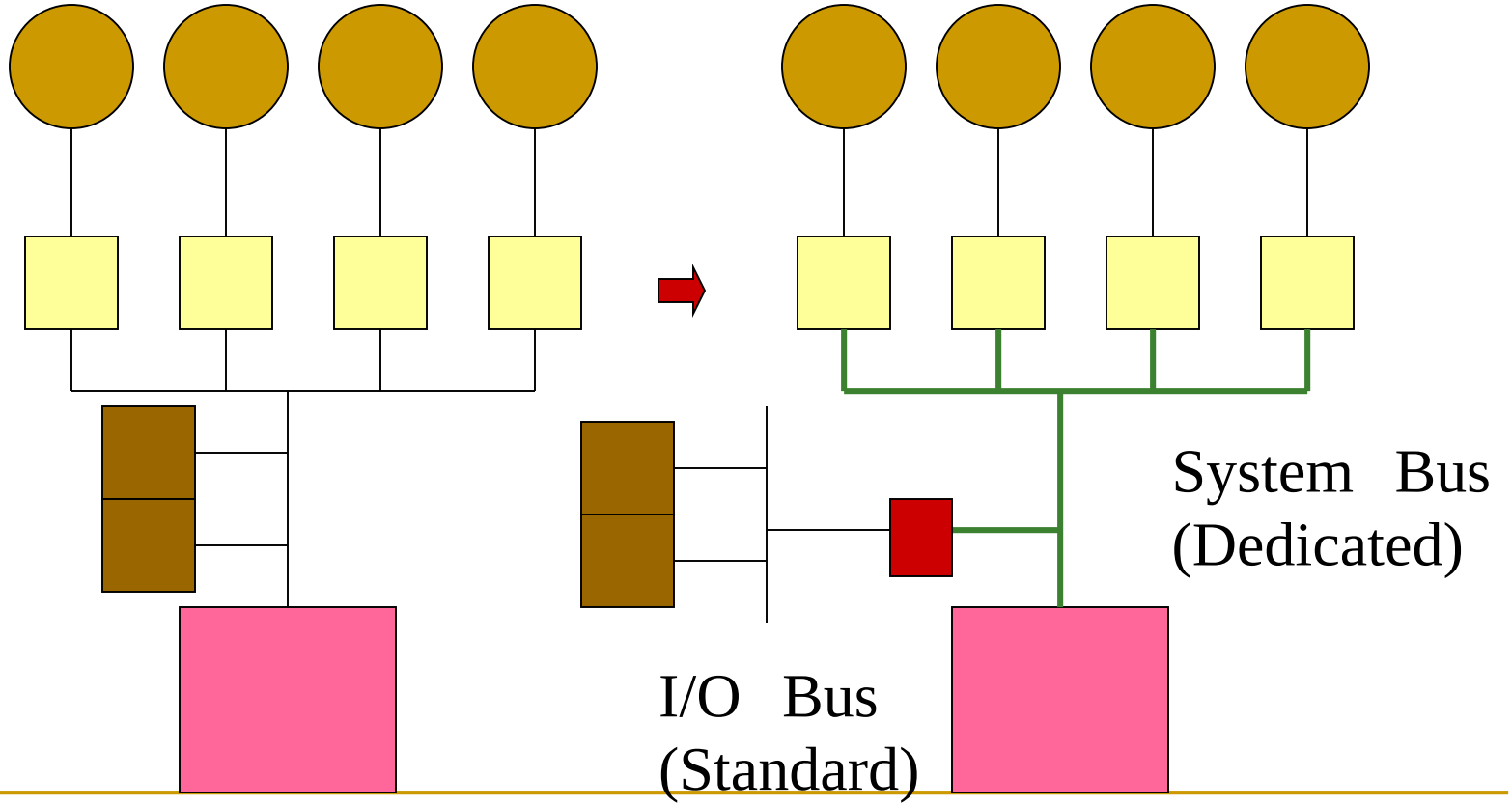
Requirements

- High Performance
  - Bandwidth（Throughput）
  - Latency
- Flexibility（Universality)
  - The number of modules
  - Clock frequency
  - Electrical characteristics

➡ Dedicated Bus

➡ Standard Bus

System bus vs. I/O bus



Synchronous vs. Asynchronous

- Synchronous bus
  - Data is sent synchronized with a clock
    - Easy to handshake, block (continuous) data transfer
    - Module numbers/types are limited
  - PCI、Mbus、PCIx、PCI express, On chip buses
  - Performance centric
- Asynchronous bus
  - Data is sent without a system clock
    - Variable modules can be connected
  - VME、Futurebus+

Recently, asynchronous buses are not commonly used

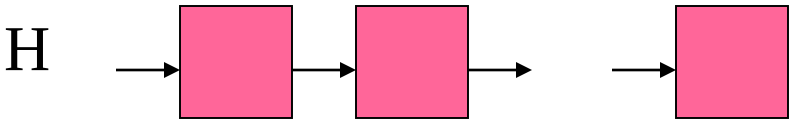
Terms around bus

- Transaction: A continuous data transfer of address and data
- Arbitration : An operation for taking a right to control the bus
- Bus Master: a module which had a right of controlling the bus through the arbitration
- Bus Slave : modules except the bus master

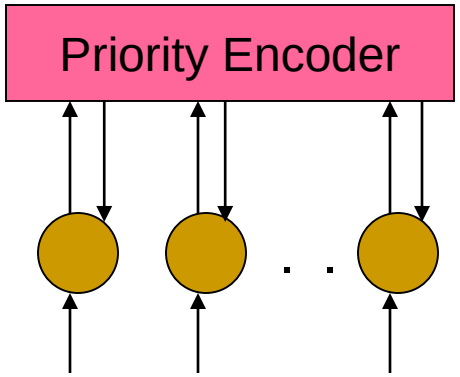
A sequence of data transfer with the bus

- Get the mastership with the arbitration → Arbiter hardware
- Bus Transaction
  - Address transfer → Handshake
  - Data transfer (repeated if necessary)
  - End of transaction
- Release the mastership

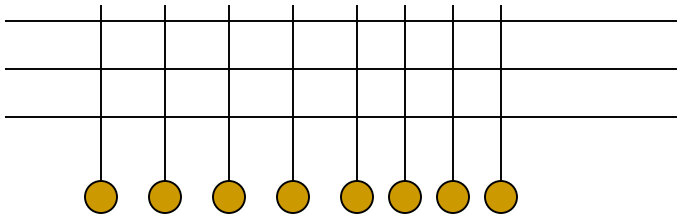
Arbiter



Daisy Chain



Centralized



Distributed bus

Distributed

Centralized arbiter is used inside the chip

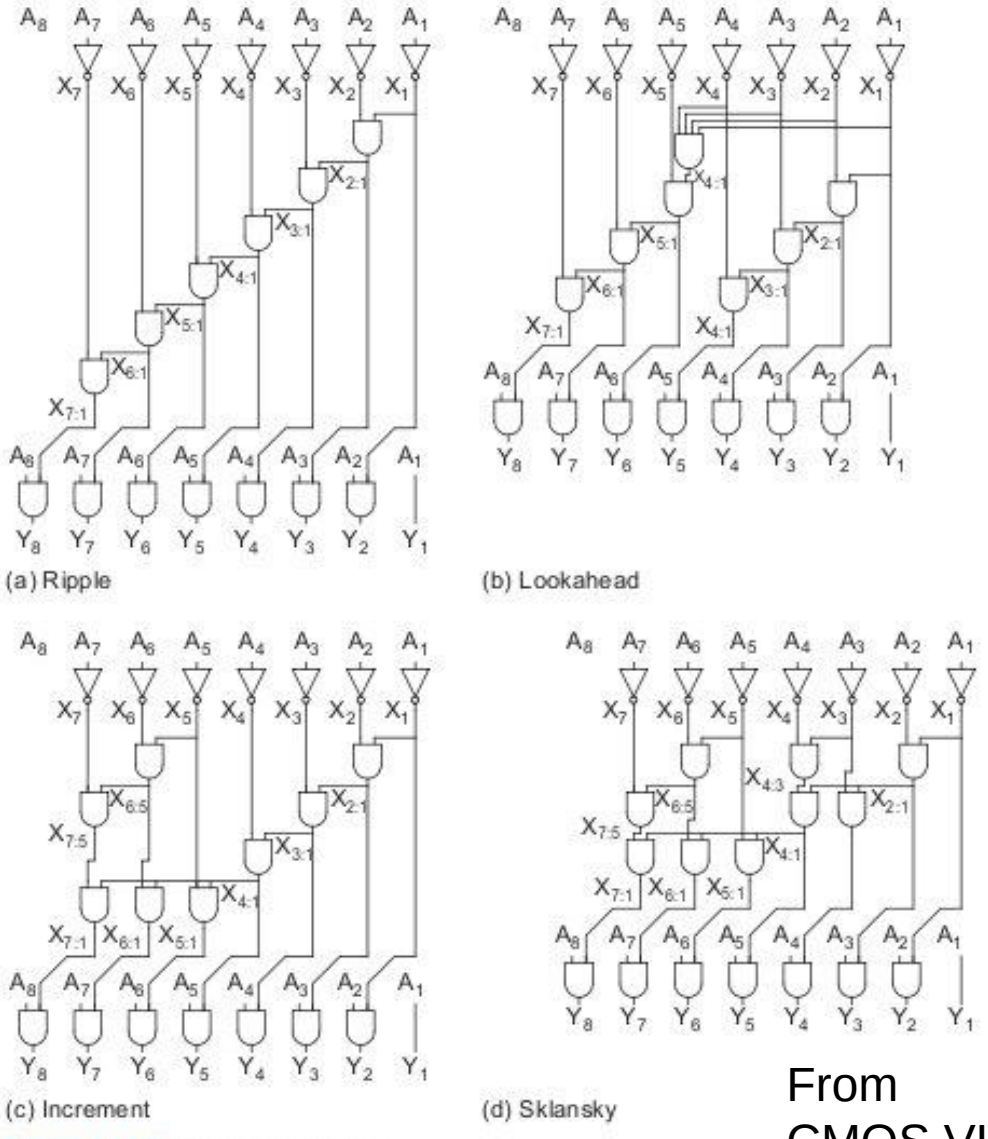
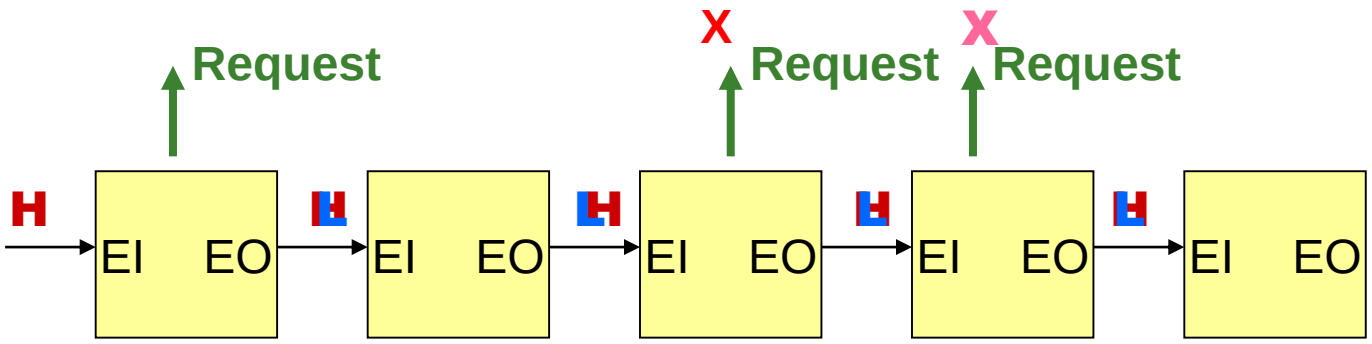


FIGURE 11.96 Priority encoder trees

Centralized Arbiter = Priority Encoder Tree

From
CMOS VLSI Design
by Weste and Harris

Daisy Chain

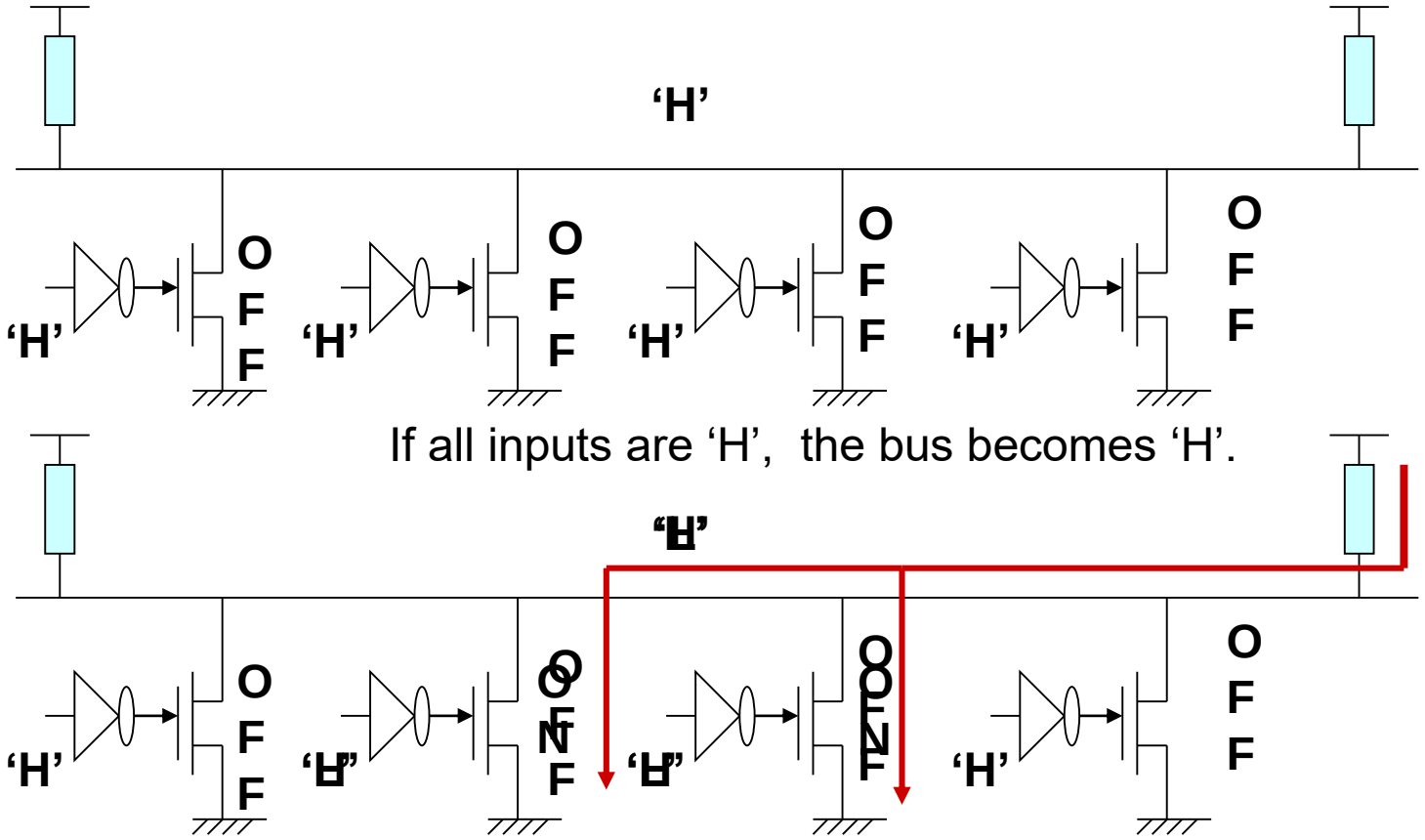


If no request EI→EO
The request can be issued only if EI is H level
When the request is issued, EO becomes L level

Right side module has a low priority

Left side module has a high priority

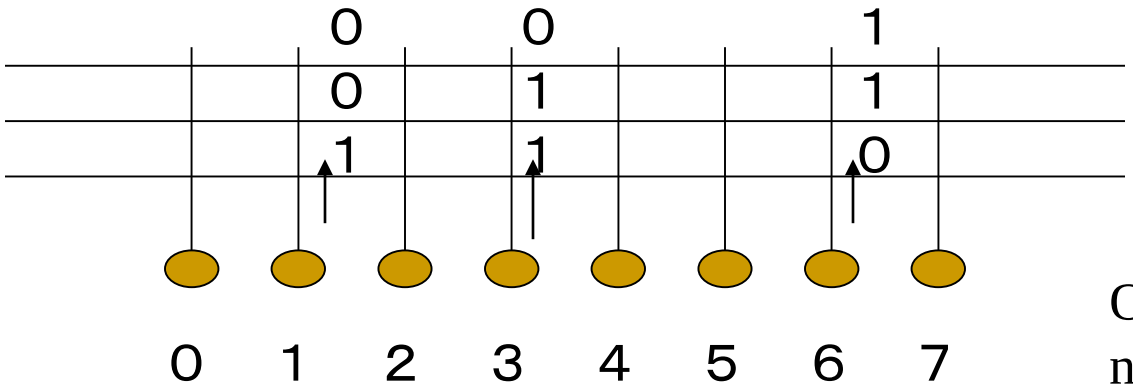
Open Drain bus



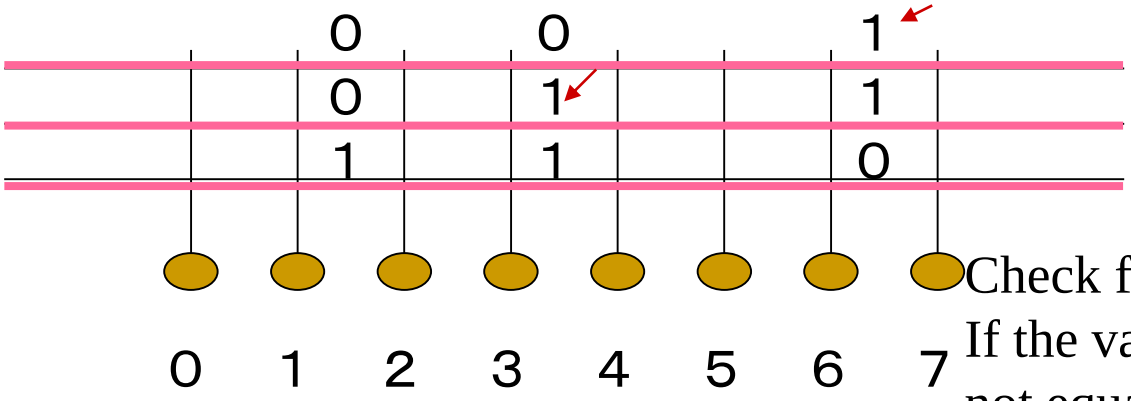
➡ Wired-OR(AND Tie)

Distributed bus arbiter

Open Drain:
0 overtakes 1



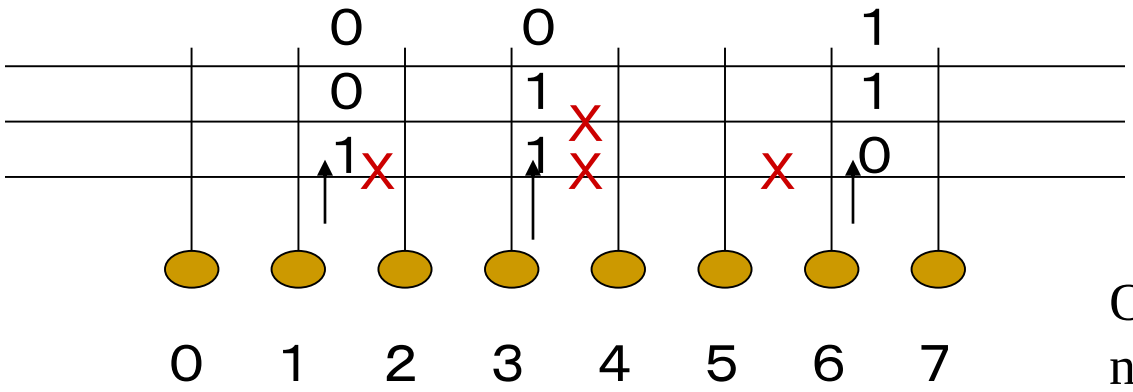
Output its own number



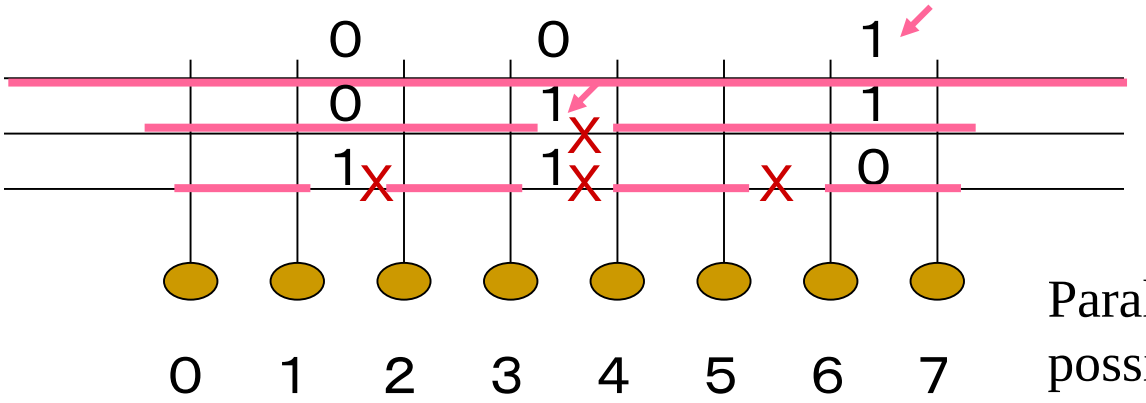
Check from the upper line. If the value on the line is not equal to its output number, then stop the output.

Modified method (Keio's patent)

Set cut-points on the bus



Output its own number



Parallel check is possible

Starvation Problem

- If the priority of the arbiter is fixed, a weak module cannot use the bus continuously.
- Central arbiter

→ Round robin priority scheduling

- Distributed arbiter

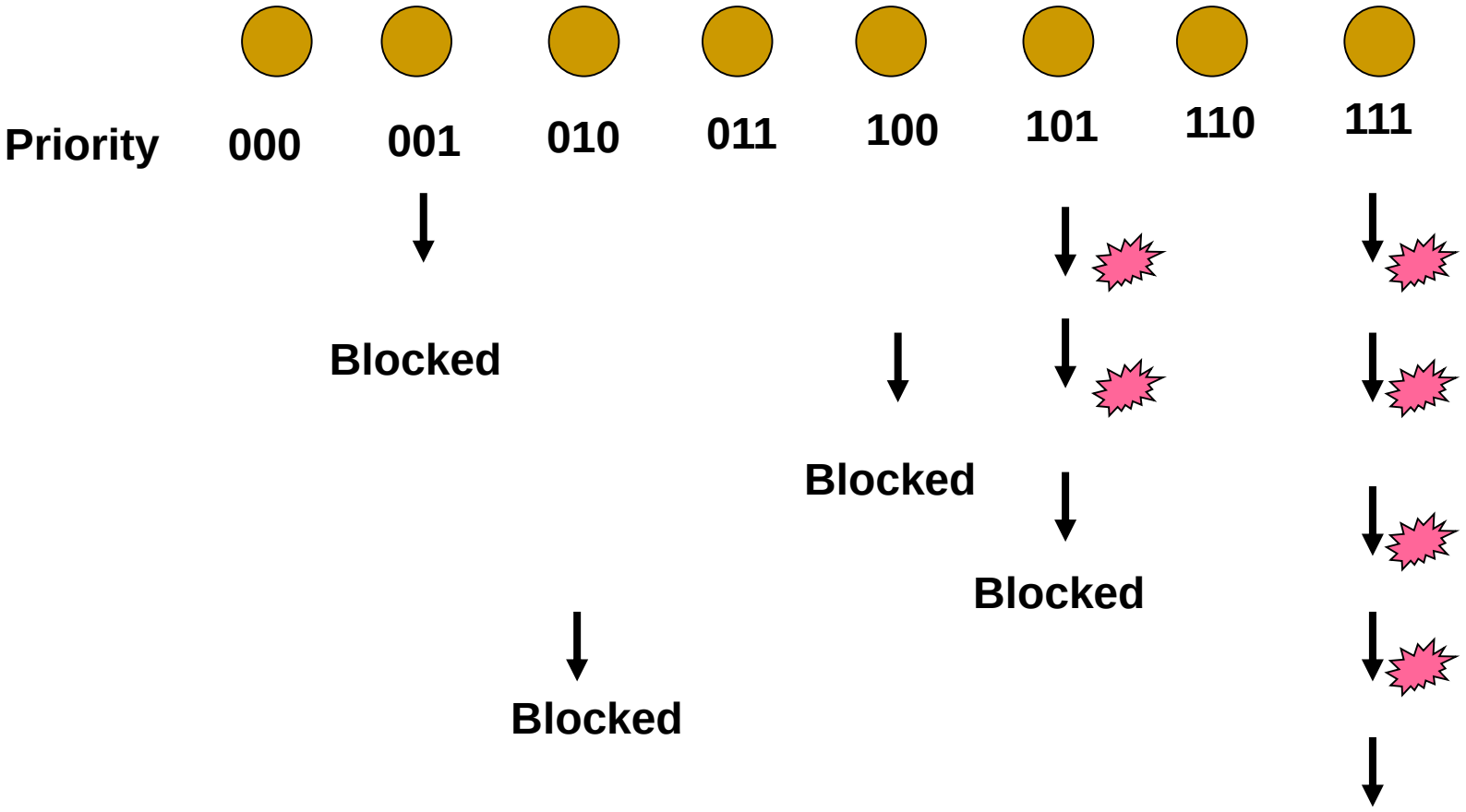
→ The next request cannot be issued until all requesting modules satisfy their requests.

Round Robin

Priority	000	001	010	011	100	101	110	111
	001	010	011	100	101	110	111	000
	010	011	100	101	110	111	000	001
	011	100	101	110	111	000	001	010

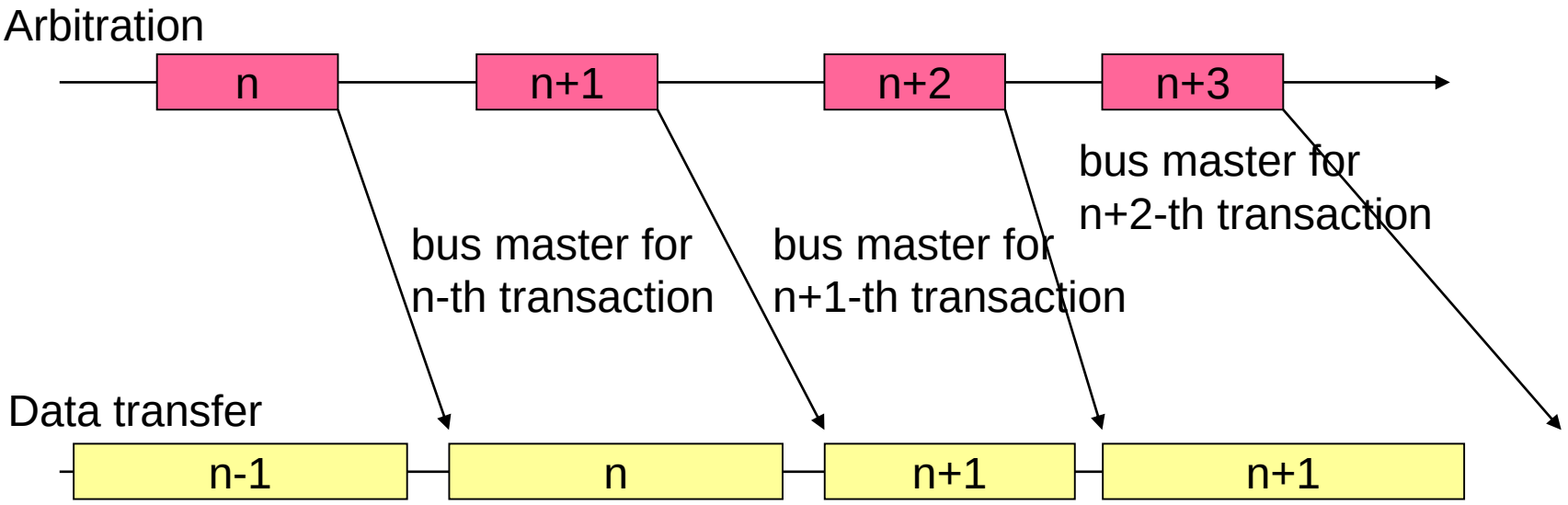
Practical Starvation Avoidance

Assume that 0 is the strongest.



All Blocked modules are released

Overlap between the arbitration and data transfer

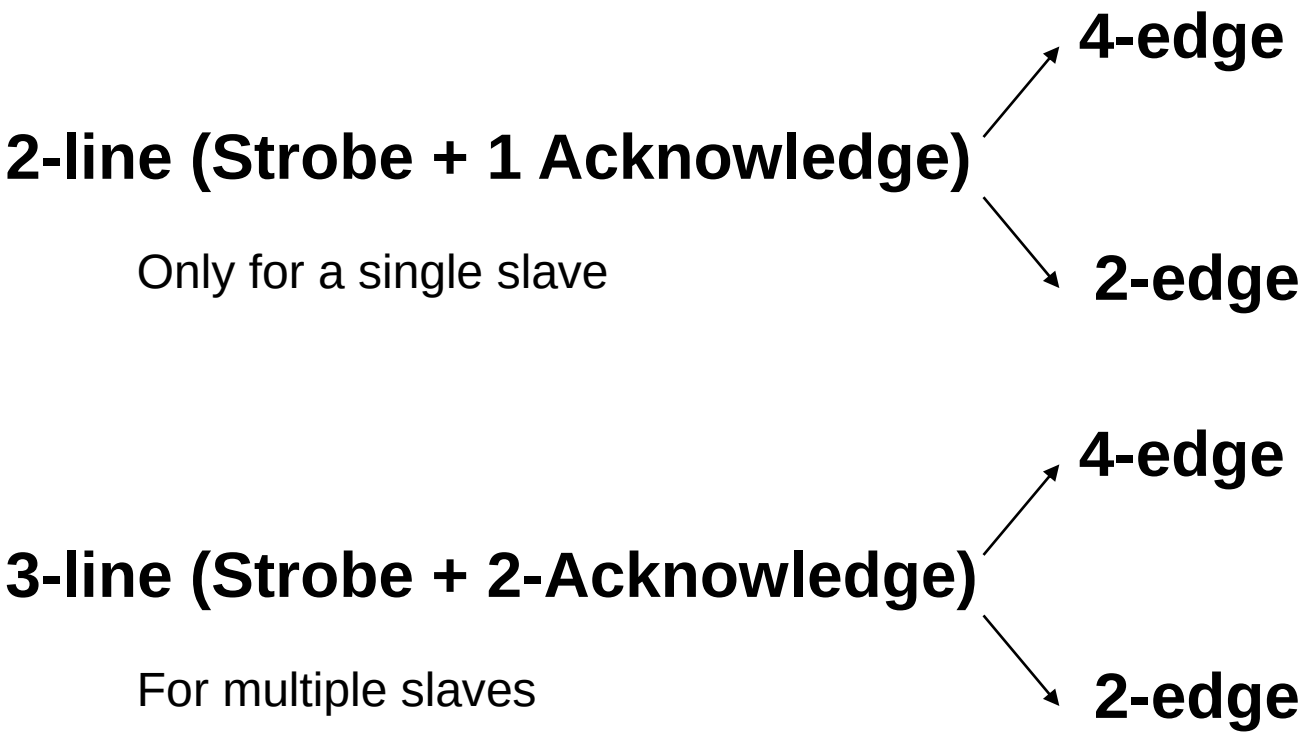


So, the arbitration time is not critical in most cases.

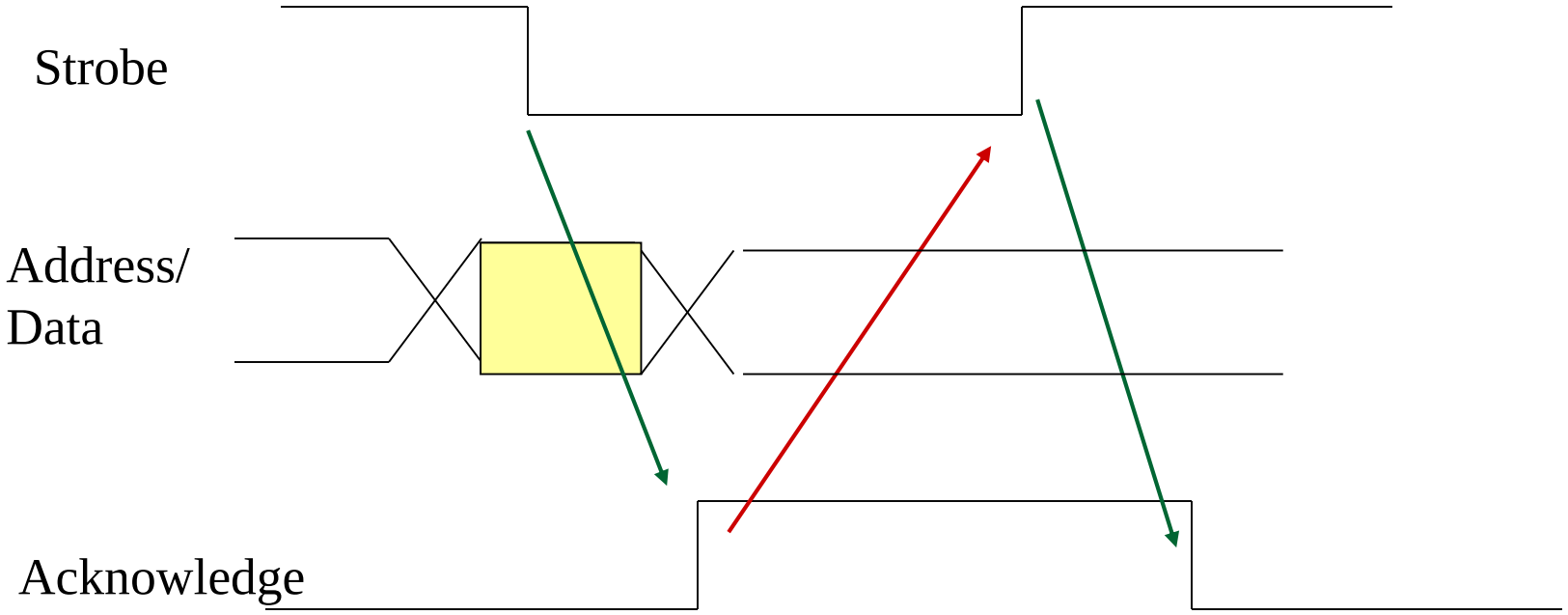
glossary-1

- Arbiter 調停回路
- Arbitration　調停操作、バスマスタを選ぶ
- Bus master　バスマスタ、バスの利用権を管理するモジュール
- Bus slave　バススレーブ、バスの利用権を持たないモジュール（マスタからスレーブに常にデータを転送するわけではないので注意！）
- Centralized 集中型　⇔　Distributed 分散型
- Daisy Chain　Arbiterの一方法で、ヒナゲシの花輪から来ている
- Transaction　バス上でデータを転送するための一連の操作
- Open drain　オープンドレイン、バスの作り方の一つで、出力トランジスタをオープンにして抵抗につなぐ。全てがOFFのときのみHレベルになり、どれか一つでもONになるとLレベルになる。この操作をワイヤードORと呼ぶ。
- Starvation　飢餓状態、バスの利用権を獲得できない状態が長期間続くこと
- Round-robin　ラウンドロビン、優先順位をArbitration毎に隣りのモジュールに移動していく方法

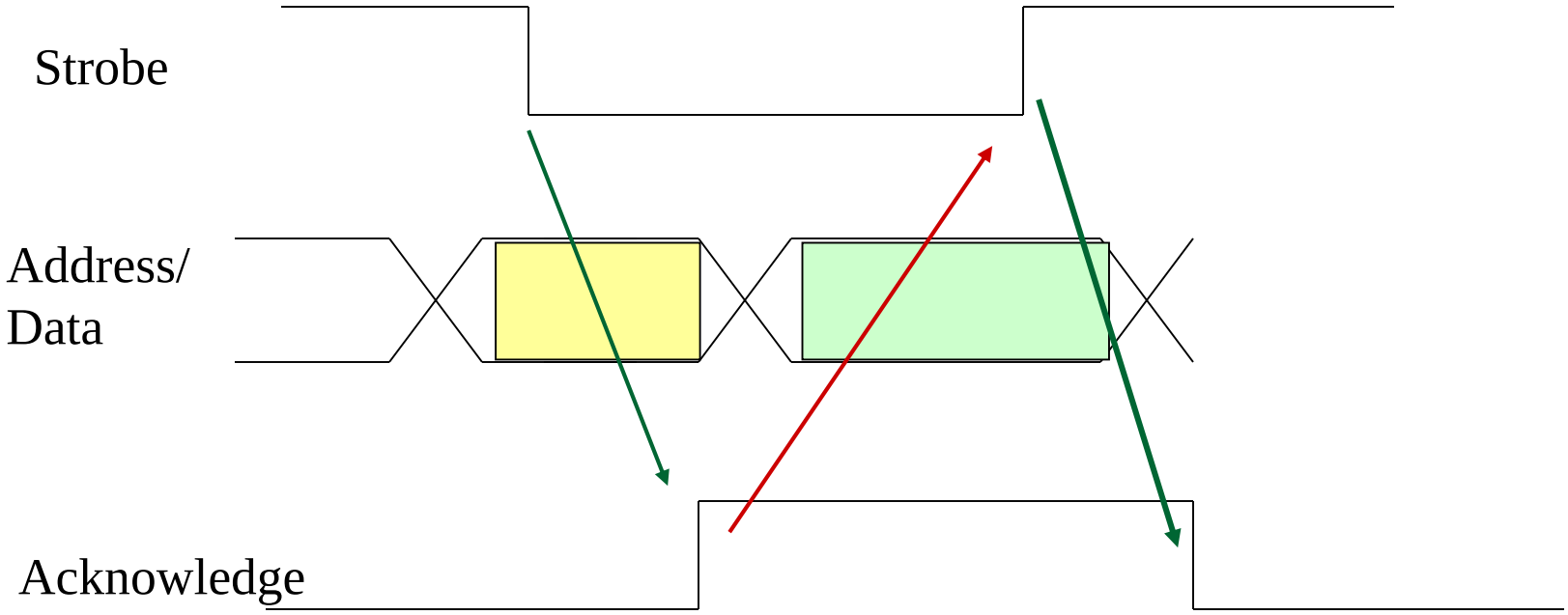
Handshake for data transfer



2-line 4-edge handshake

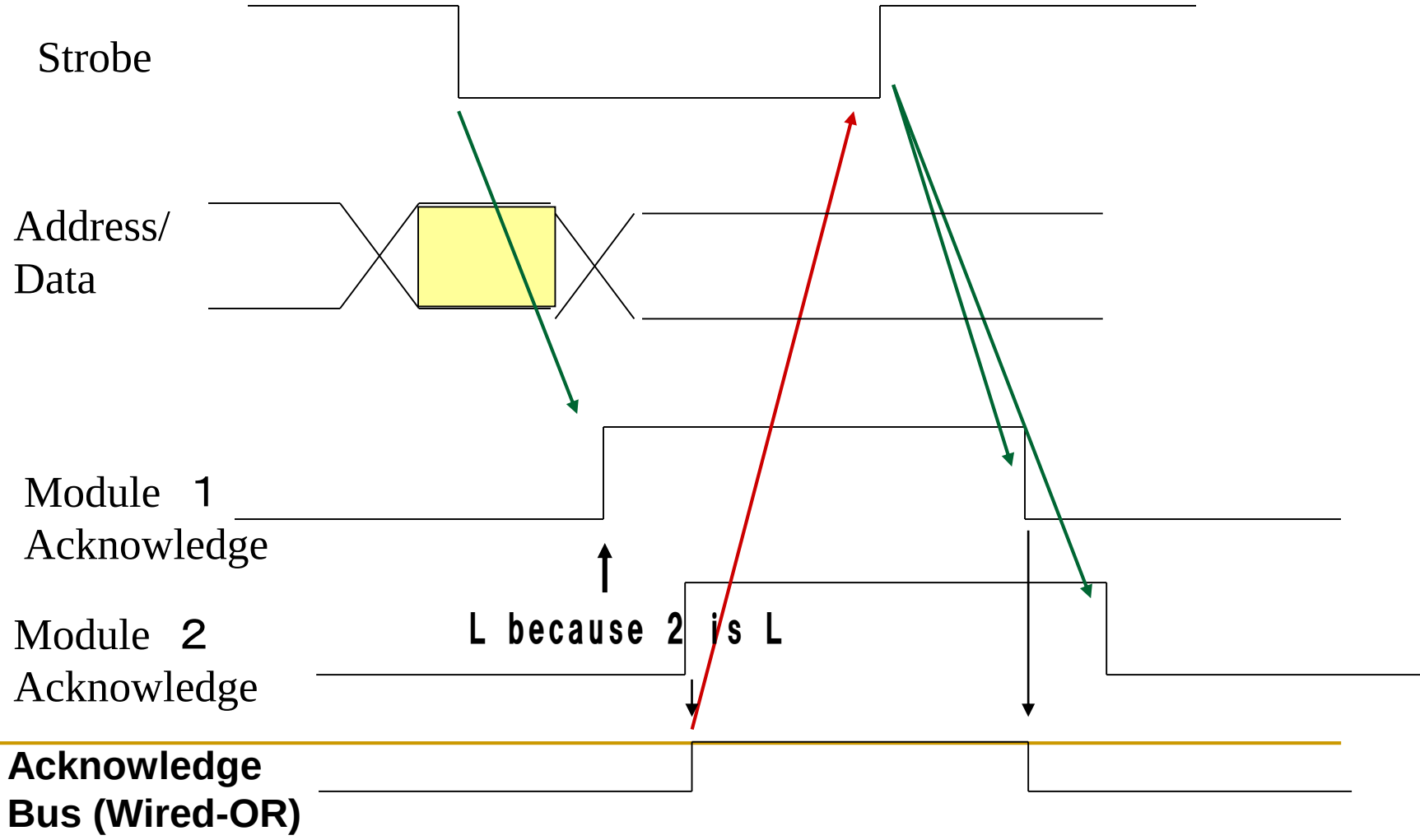


2-line 2-edge handshake



Data is transferred with both edges of the strobe

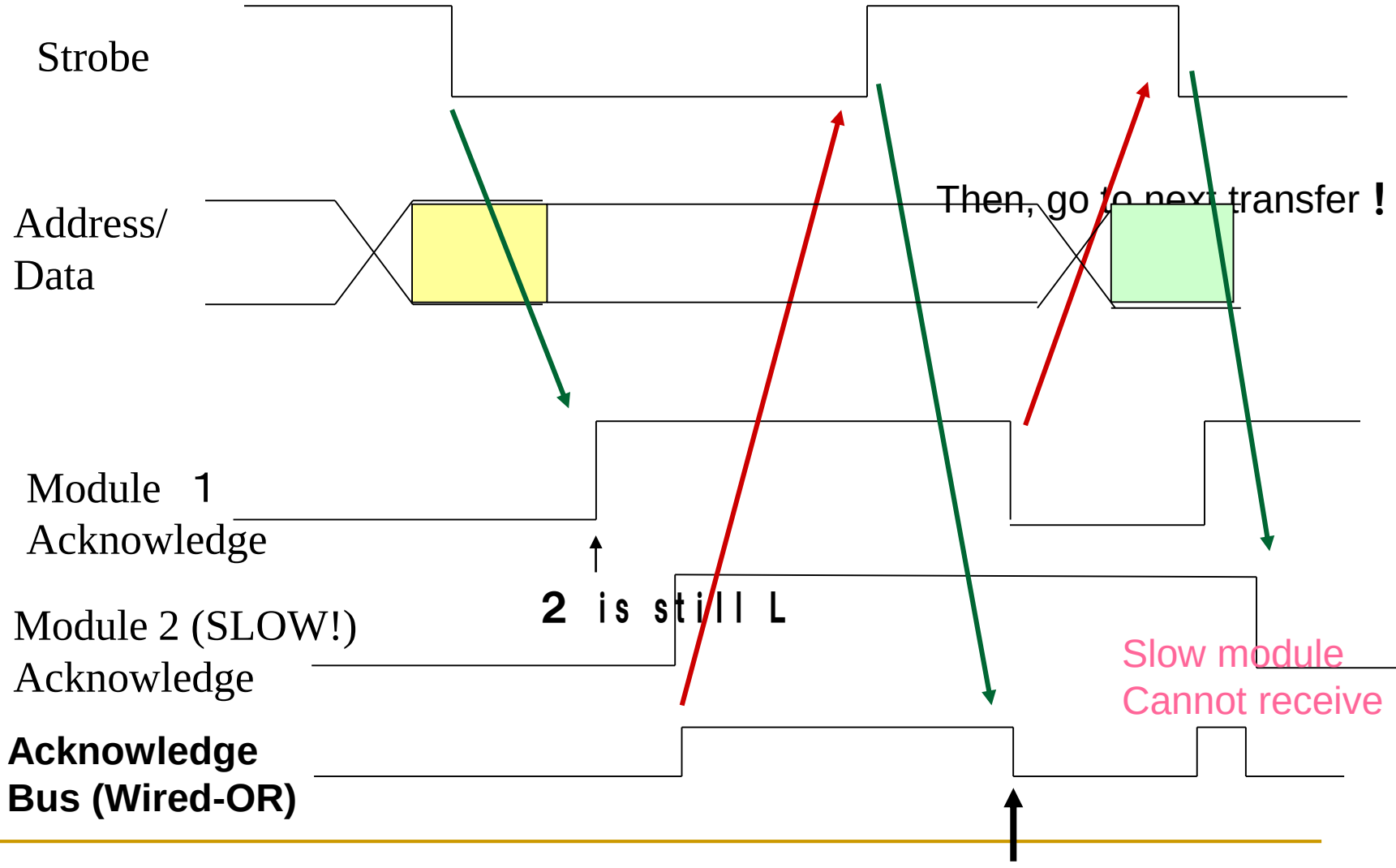
In the case of multiple slaves



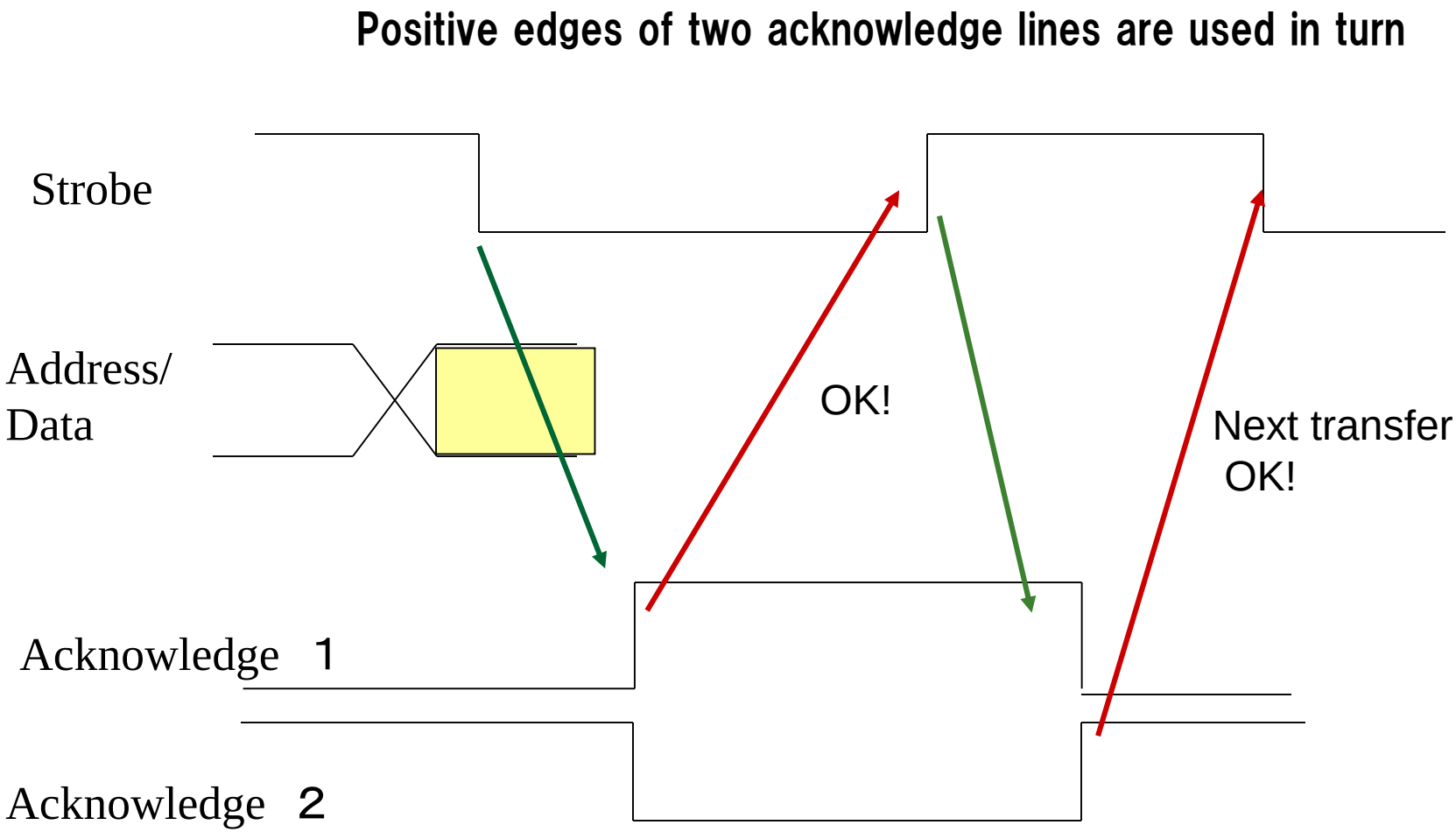
Quiz

- 3-line handshake (1 for strove and 2 for acknowledge) is used for multiple slaves.
- Why 2-line handshake cannot manage multiple slaves?

2-line cannot manage multiple slaves

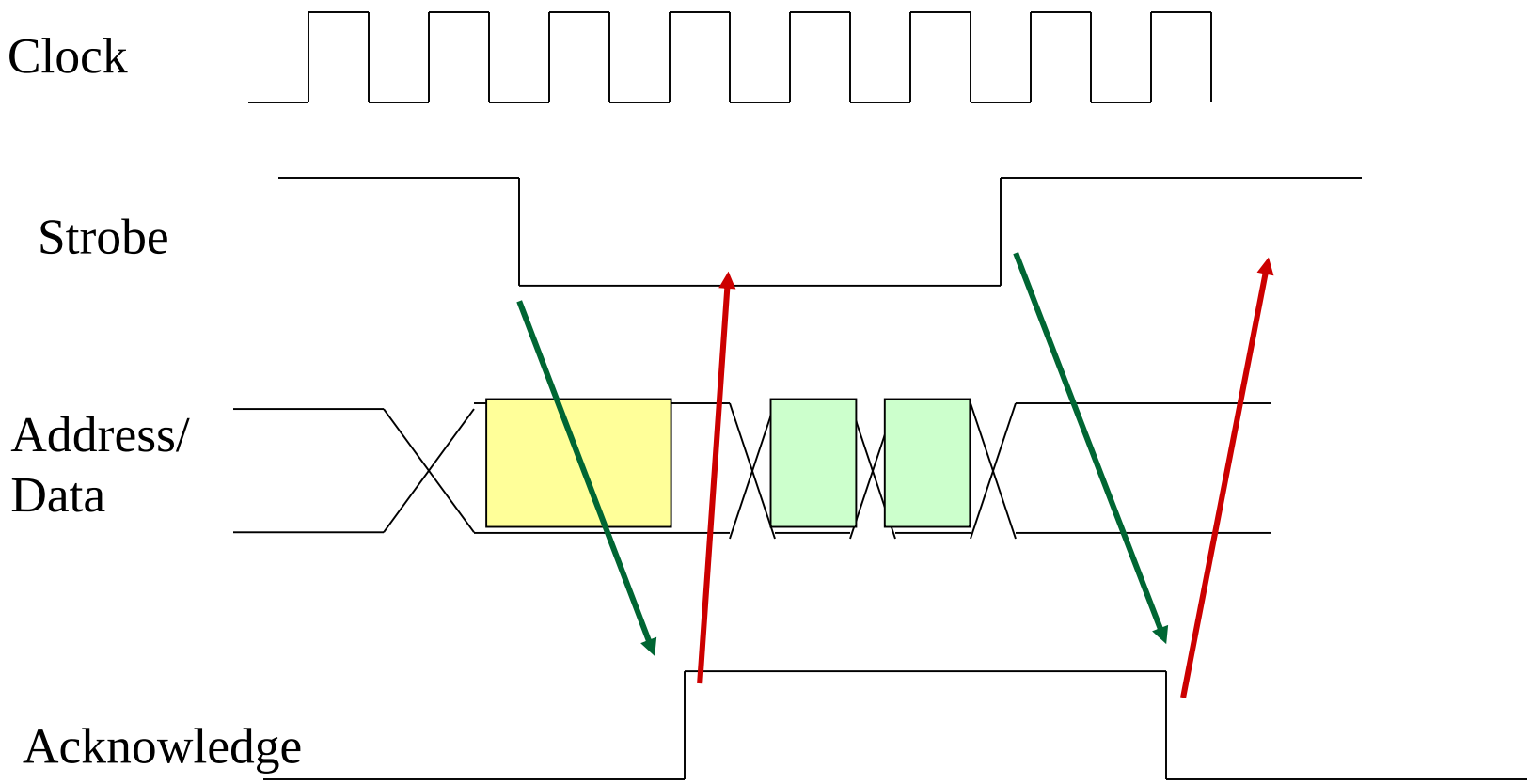


3-line handshake



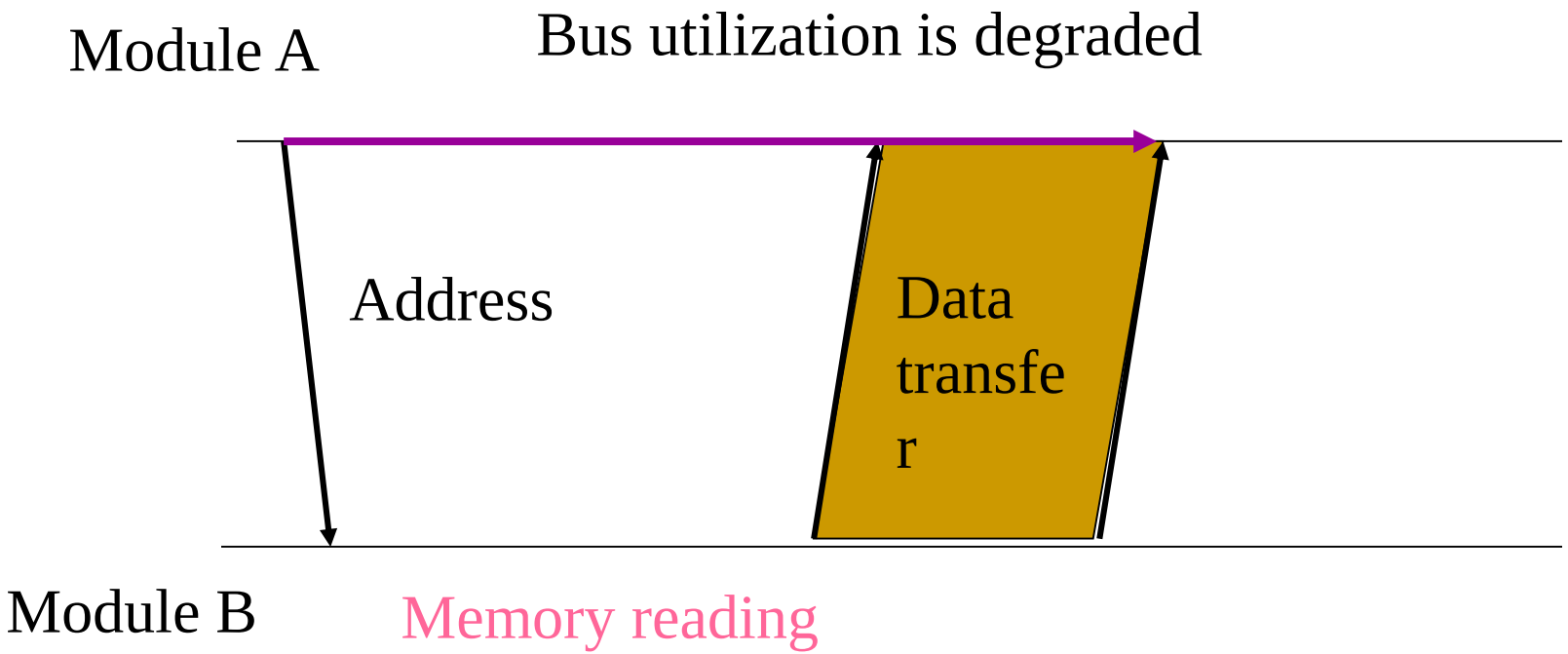
3-line 2-edge handshake is also possible

Synchronous bus is suitable for block transfer

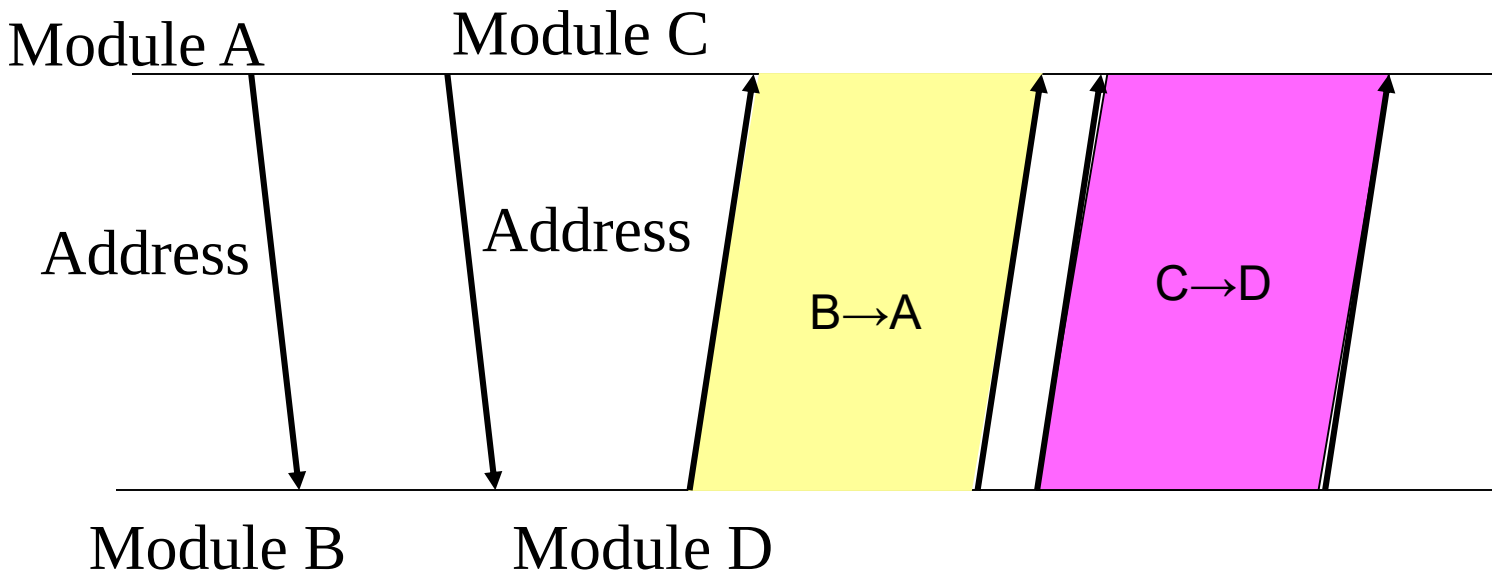


The start/end handshake is the same, but block transfer is possible synchronized with a clock

Non-Split Transaction



Split Transaction



Split transaction of A→B

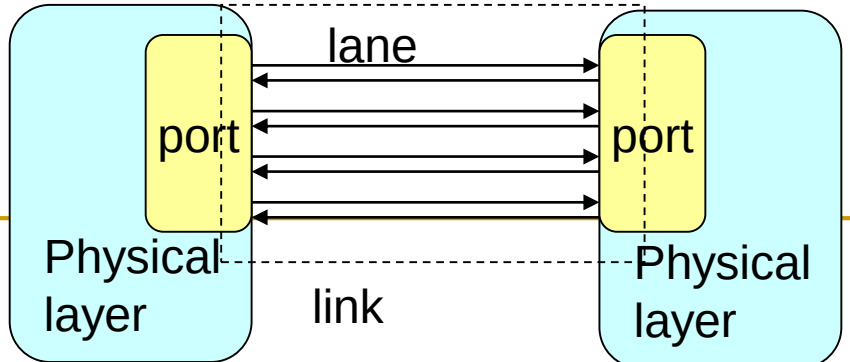
Transaction C→D is executed

Advanced I/O Buses

- PCI bus was widely used, but it could not cope with recent computer system.
  - 32bit/33MHz, 64bit/66MHz
- New standard I/O bus
  - PCI-X
    - 64bit/133MHz DDR/QDR
  - PCI Express
    - Point-to-point serial data transfer
    - 1 lane:2.5Gbps
    - x2, x4, x8
  - Now, PCI Express is used instead of PCI bus.

PCI Express

- Consisting of serial one-to-one bidirectional connection wires called lanes.
- Each lane supports 2.5Gbps/5Gbps (Physical Speed)
- Multiple lanes can be used as a link(x4, x8, x16 and x32).
- The data is transferred in a packet called TLP (Transaction Layer Packet).
- Interconnection network rather than the bus, but the protocol of traditional PCI bus is supported.

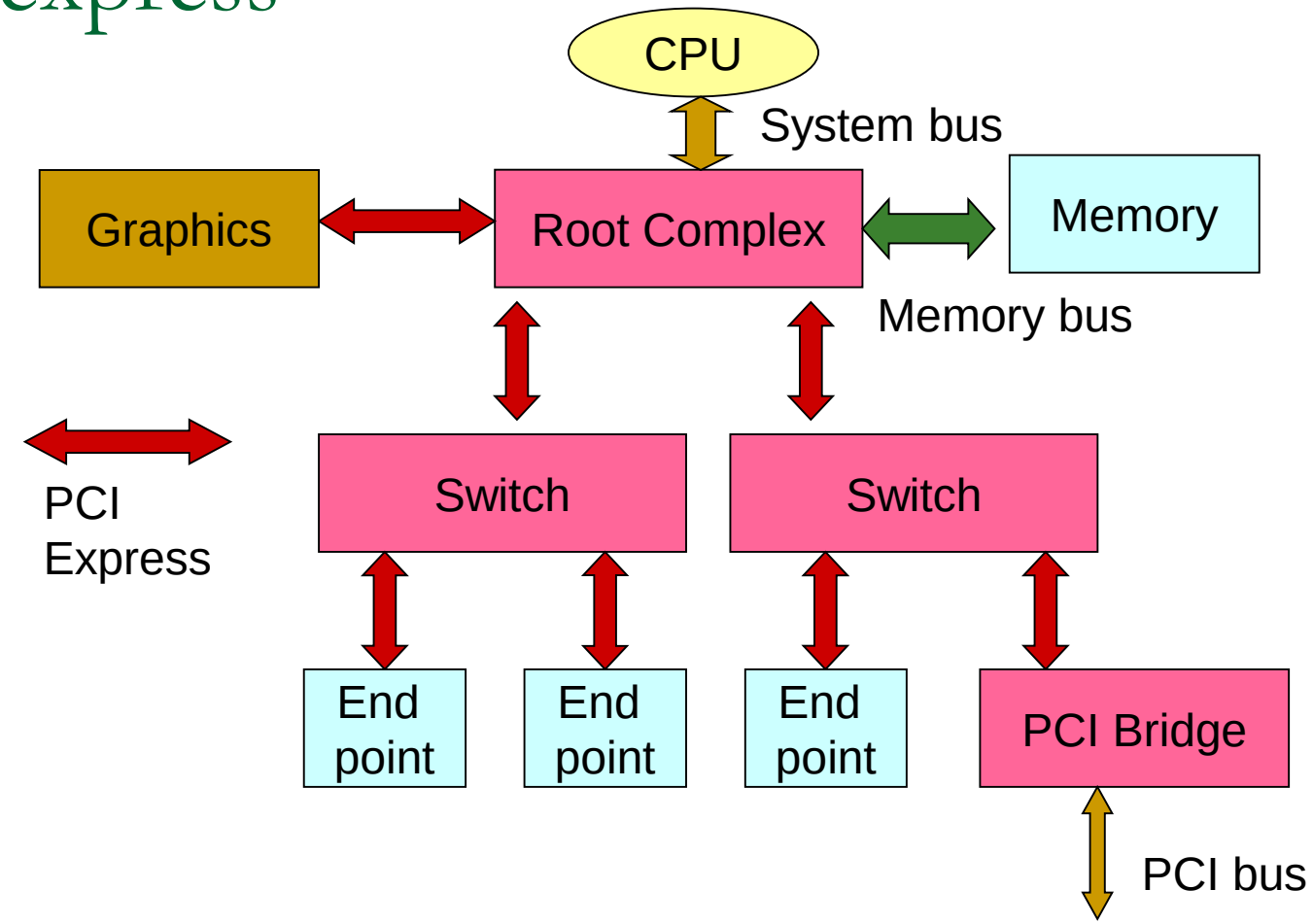


PCIe standard

	Gen1	Gen2	Gen3
Physical speed（Gbps）	2.5	5	8
Bandwidth (GB/sec)	0.25	0.5	1.0
x8 bandwidth (GB/sec)	2.0	4.0	7.9
Encoding	8b/10b	8b/10b	128b/130b

Physical speed is x1.6, but almost twice practical performance is realized by changing the encoding method.

An example of bus system using PCI express

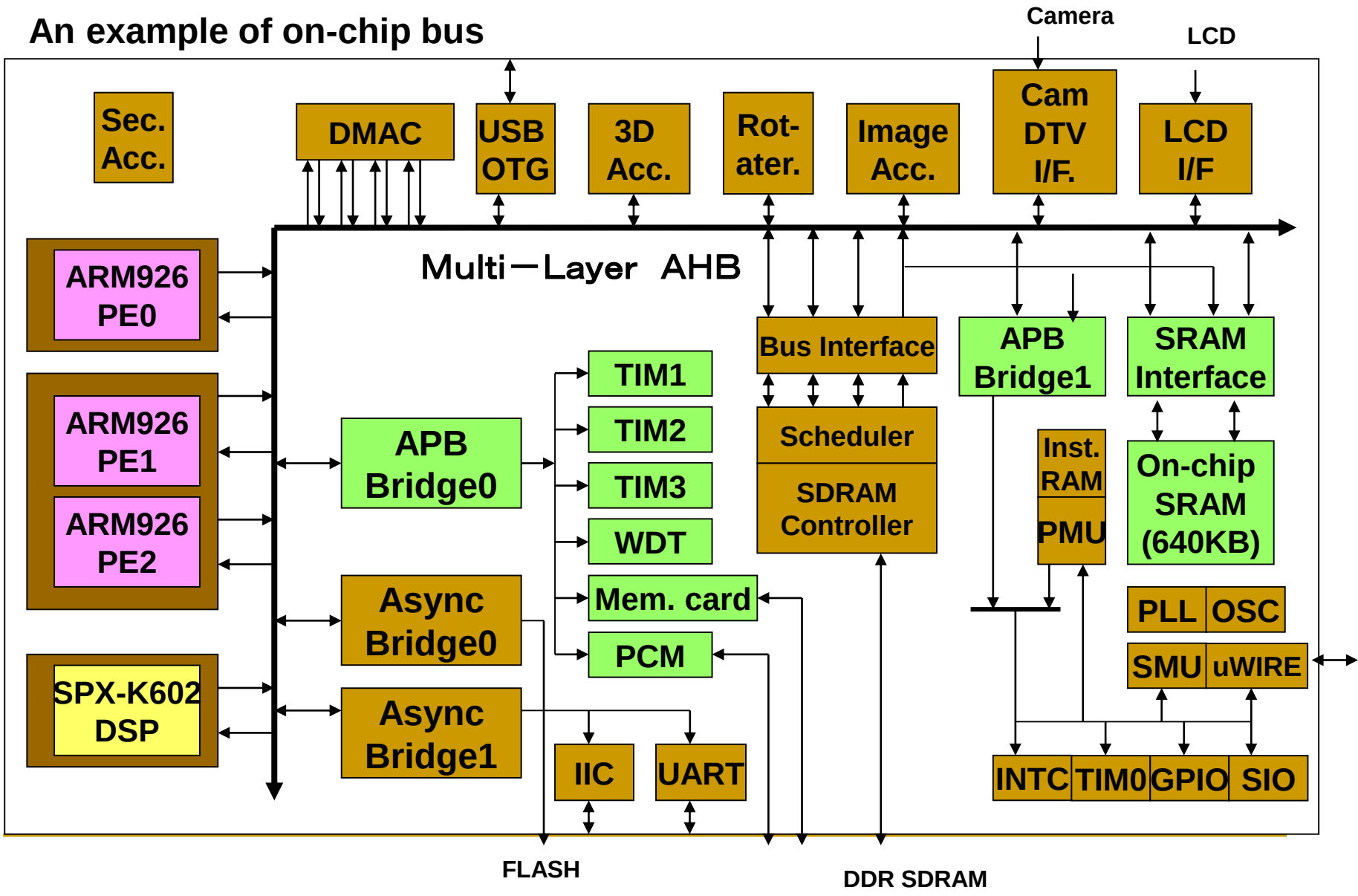


On-chip bus

- For on-chip implementation, various types of IP (Intellectual Property) must be connected.
- Standard bus is required.
  - AMBA (Advanced Microcontroller Bus Architecture): a bus for ARM cores.
  - CoreConnect: a bus for PowerPC cores.
  - Wrapper based buses
    - IPs are wrapped in the standard interface.
- For further performance improvement, NoCs (Network on Chips) are introduced.

→ Introduced in the later part of this lecture

NEC MP211


An example of on-chip bus
Camera
LCD
Sec. Acc.
DMAC
USB OTG
3D Acc.
Rot-ater.
Image Acc.
Cam DTV I/F.
LCD I/F
Multi－Layer AHB
ARM926 PE0
ARM926 PE1
ARM926 PE2
SPX-K602 DSP
APB Bridge0
Async Bridge0
Async Bridge1
TIM1
TIM2
TIM3
WDT
Mem. card
PCM
IIC
UART
Bus Interface
Scheduler
SDRAM Controller
APB Bridge1
SRAM Interface
Inst. RAM
PMU
On-chip SRAM (640KB)
PLL
OSC
SMU
uWIRE
INTC
TIM0
GPIO
SIO
FLASH
DDR SDRAM

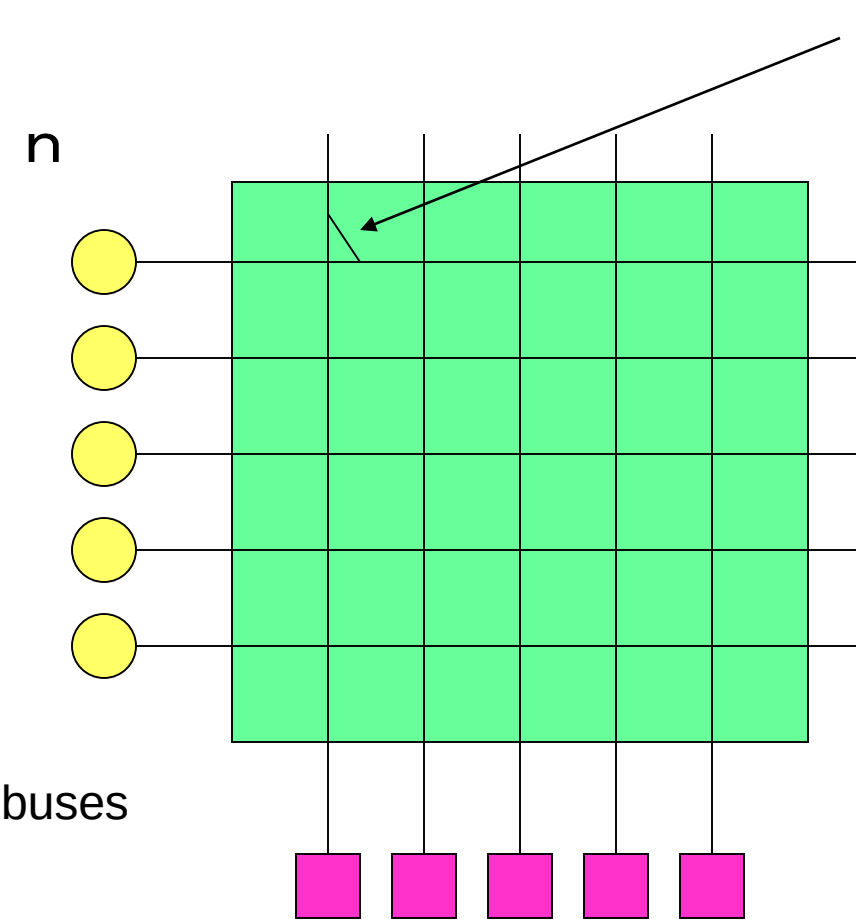
Summary of Bus

- Classic bus with passive wires has been changed to active bus with a kind of switches
- High Speed Bus
  - Synchronous bus with Split Transaction
  - Using active devices
  - It becomes somehow like a packet transfer with switching hub.

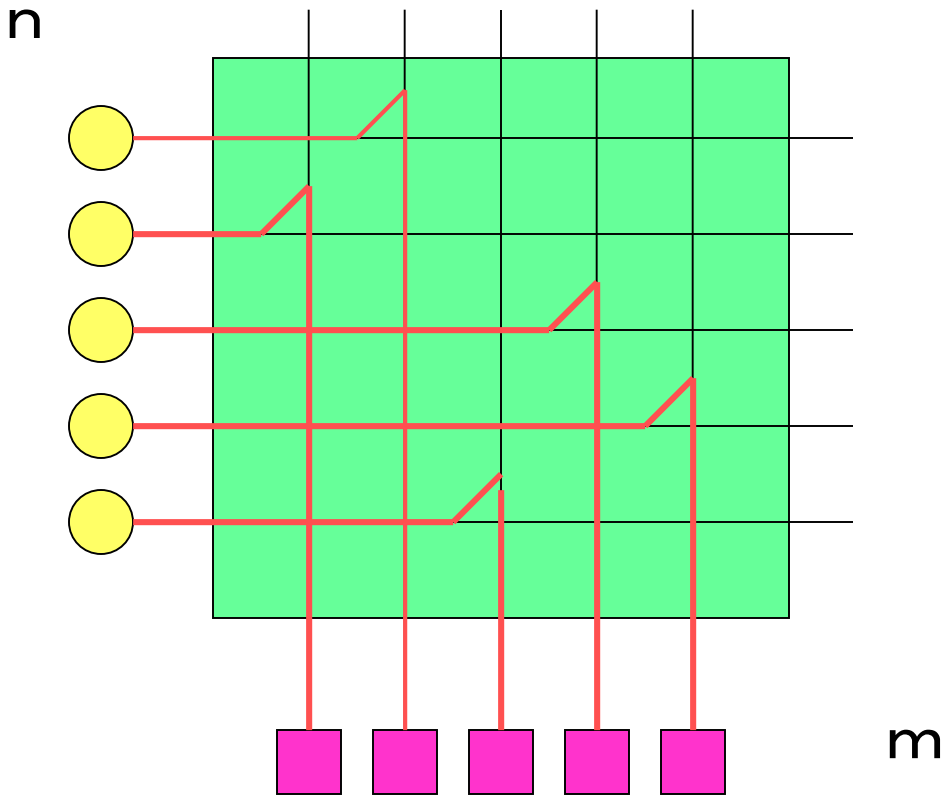
glossary 2

- Handshake 握手のことだがここでは正しく転送するための信号のやりとりを指す
- Synchronous 同期式⇔Asynchronous　非同期式
- Strobe　転送を起動を知らせる信号線
- Acknowledge　Strobeに対する応答用の信号線
- Edge　信号線の変化
- Split transaction　バス転送を中断して途中に他の転送を挟むことを可能にする方法

Crossbar switch

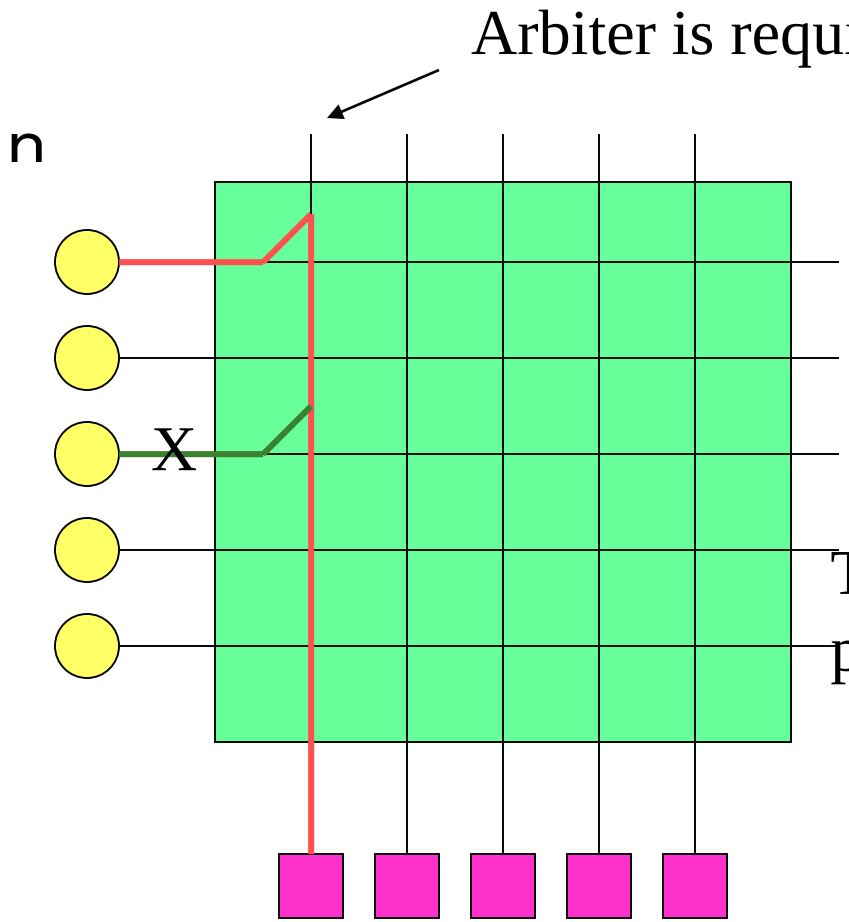


Non-blocking property



For different destination, conflict free

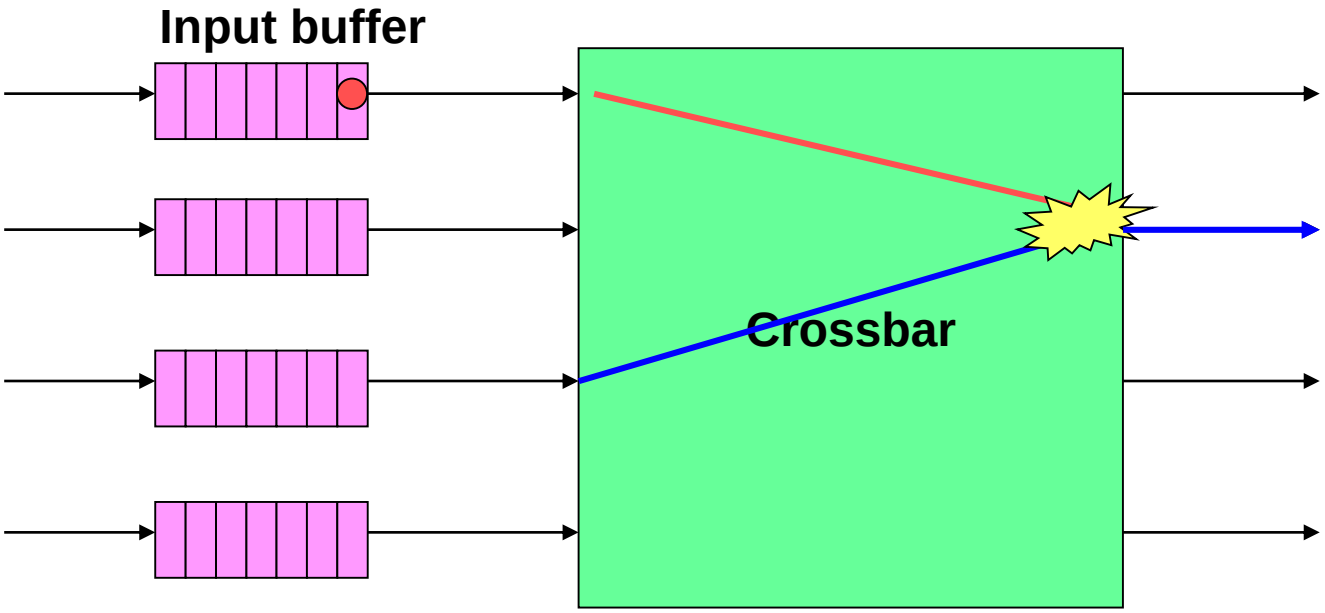
Head Of Line (HOL) conflict



The buffer is required

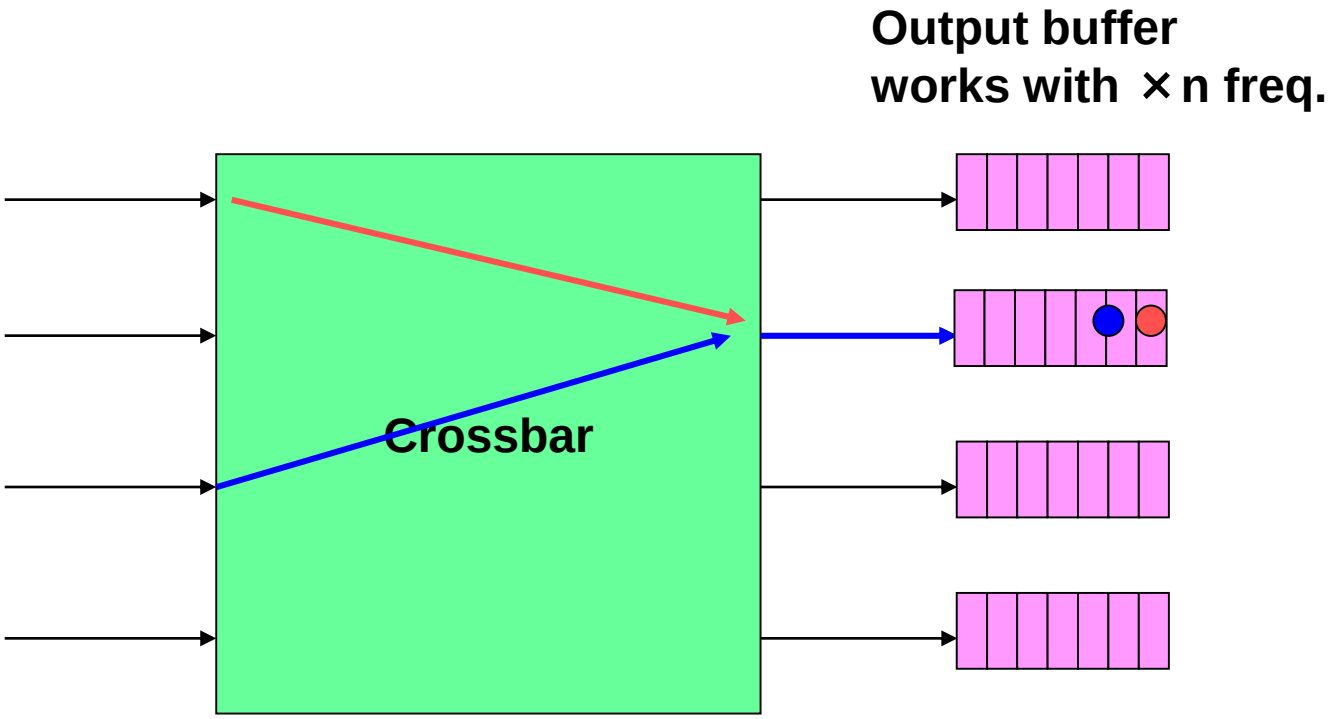
The number of cross point is not dominant.

Input buffer switch



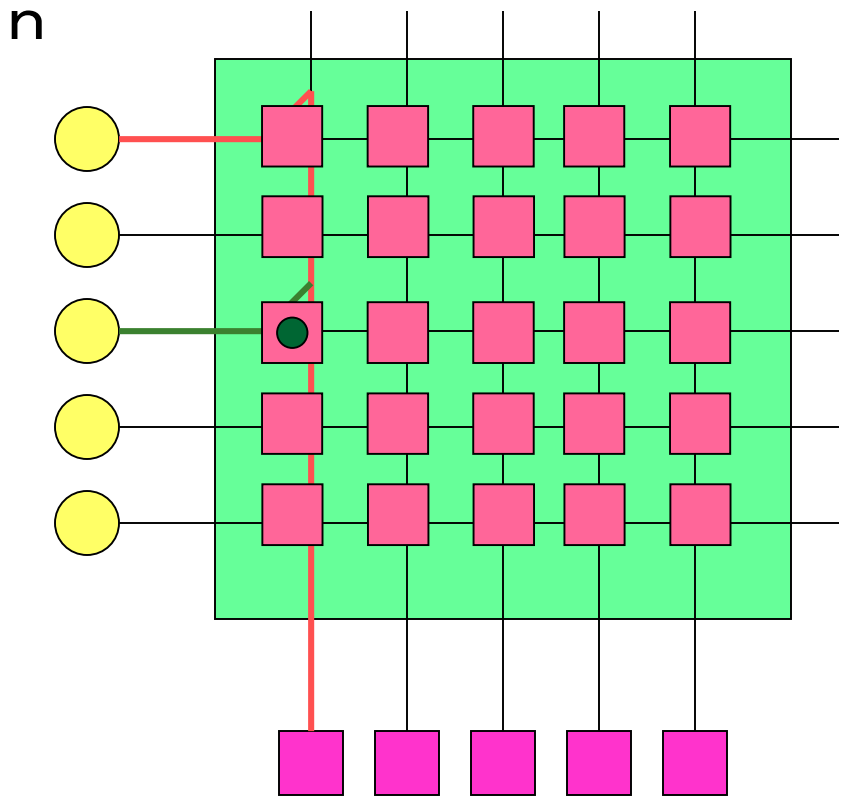
One of conflicting packets is selected.
Others are stored Into the input buffer

Output buffer switch



Crossbar must work with ×n frequency of input/output rate.
No HOL problem.
Used in switches in WAN, but for parallel machines it is difficult.

Buffers at cross-point

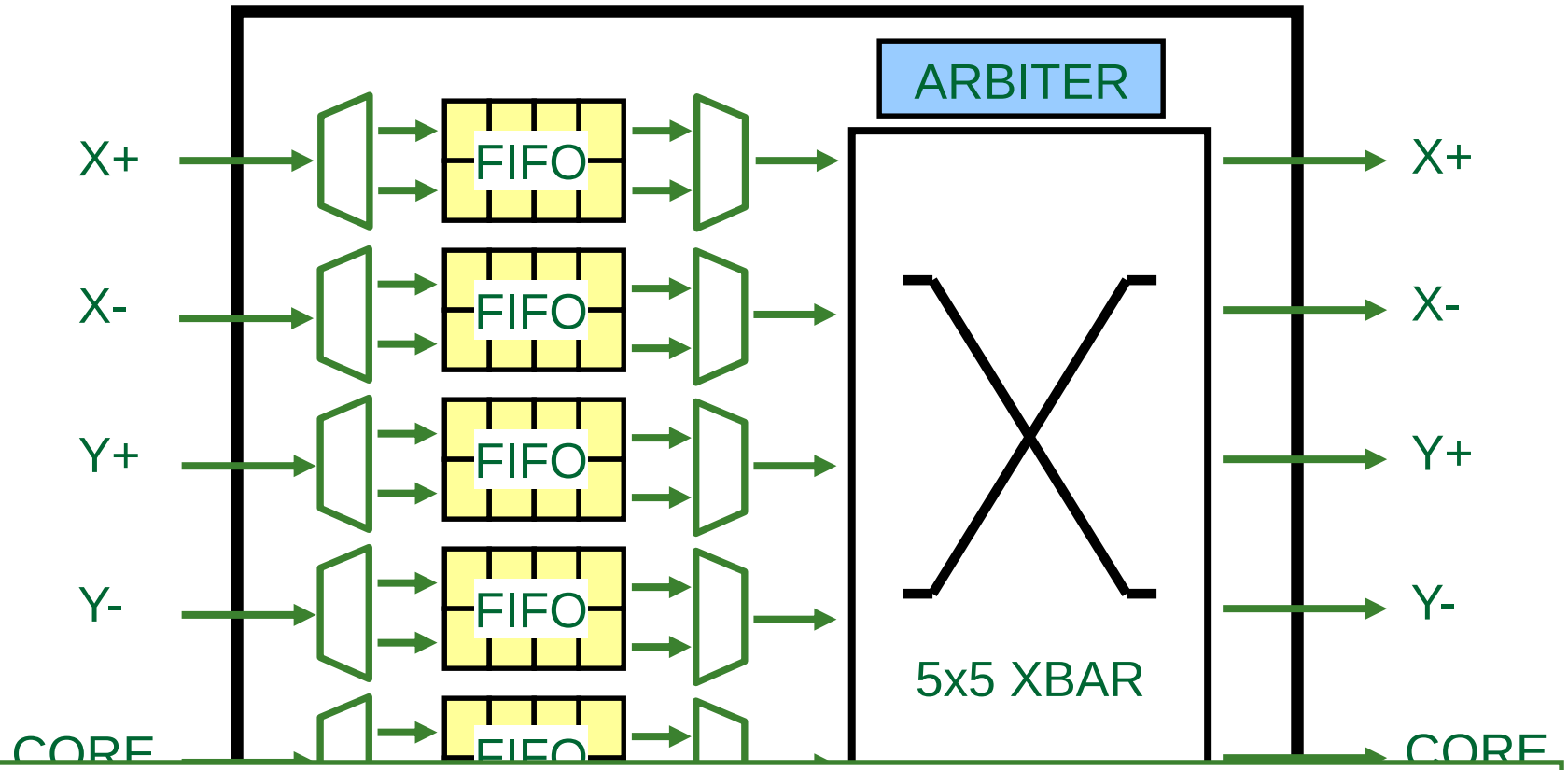


The buffer is provided at each cross-point. High performance but the total amount of buffer becomes large.

An example of a modern router

- WH router with two virtual channels

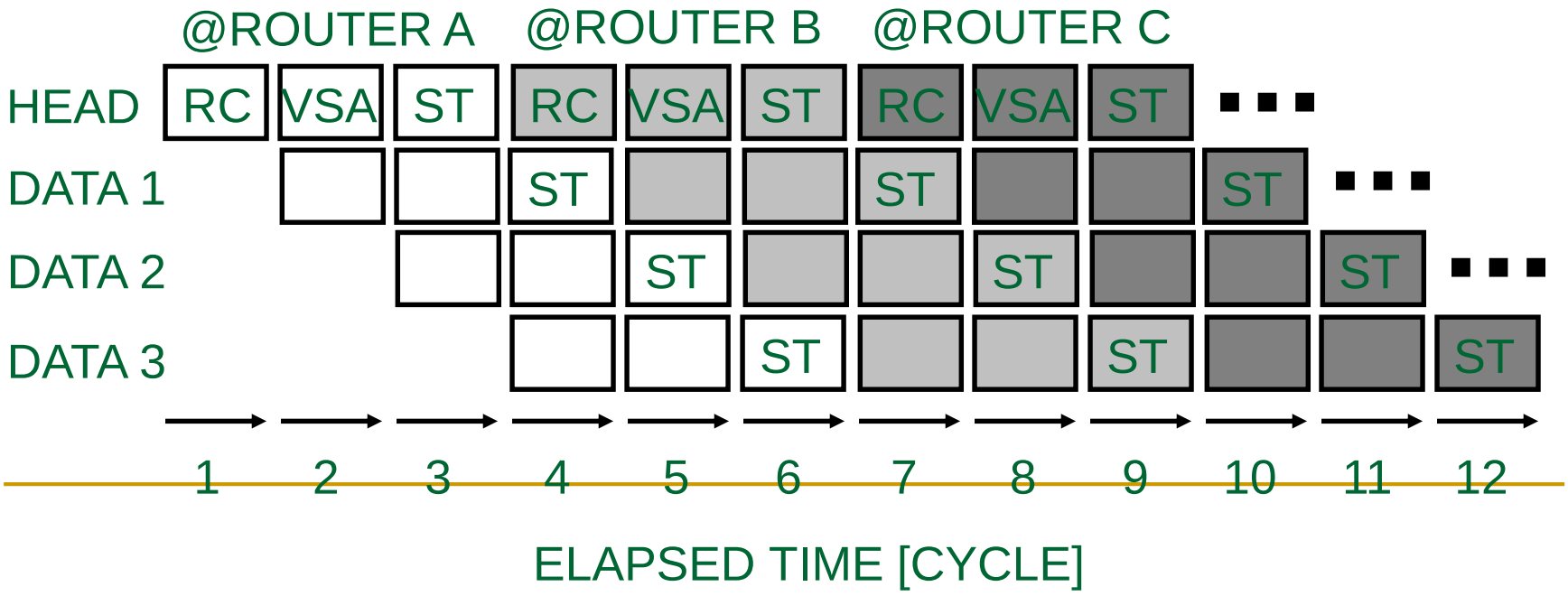
(Introduced later in this lecture)



If the bitwidth is 64bits, it uses 30～40 [kgates] FIFO occupies 60%

Pipelined operation

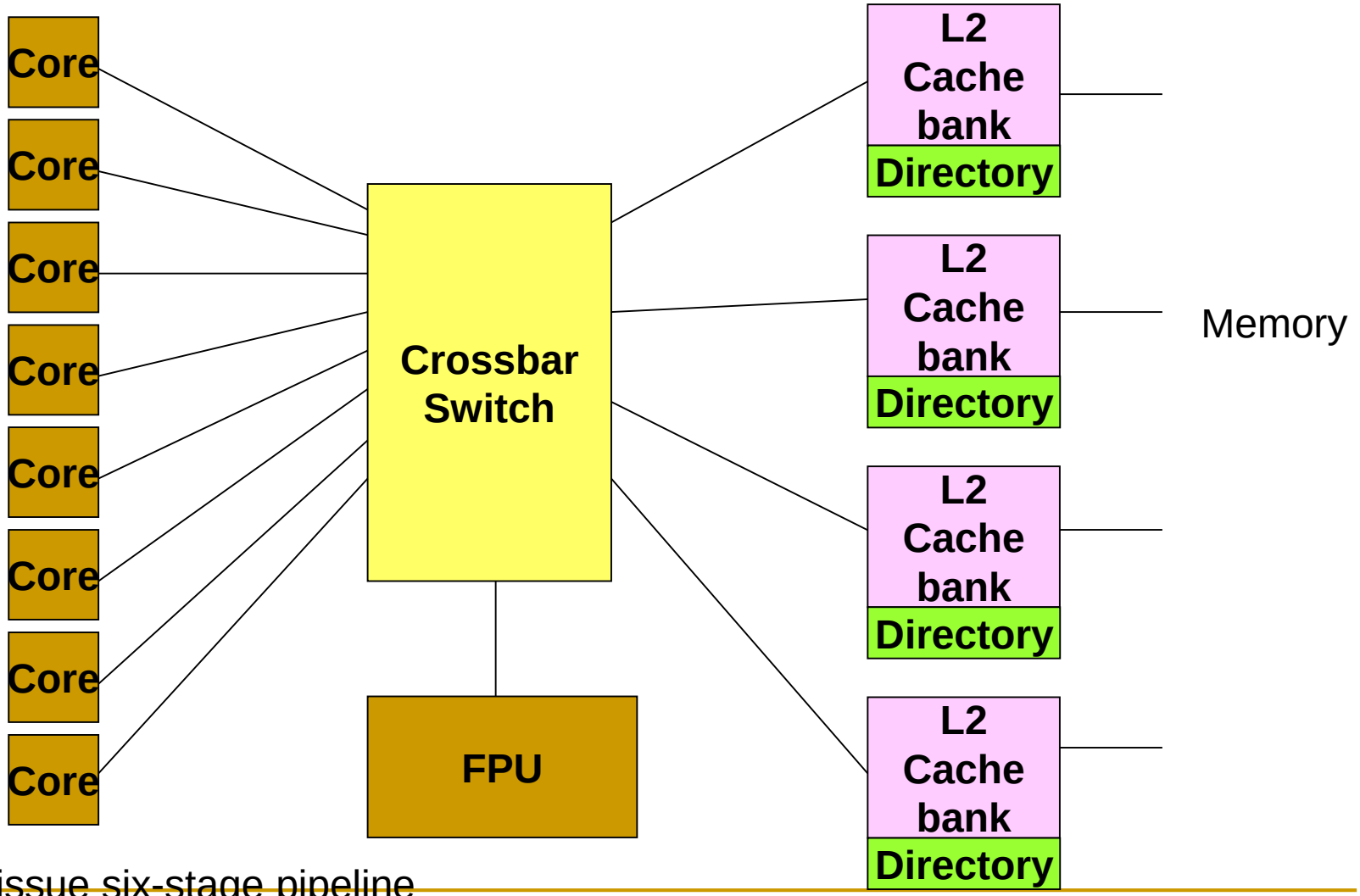
- It takes three clocks to pass through the switch
  - RC (Routing Computation)
  - VSA (Virtual Channel / Switch Allocation)
  - ST (Switch Traversal)



Merit/demerit of Crossbars

- Non-blocking property
- Simple structure/Control
- The hardware for cross-points usually do not limit the system (Fallacy of crossbars)
- Extension is difficult by the pin-limitation of LSIs
  - If pins can be used, a large crossbar can be constructed → Earth simulator

SUN T1

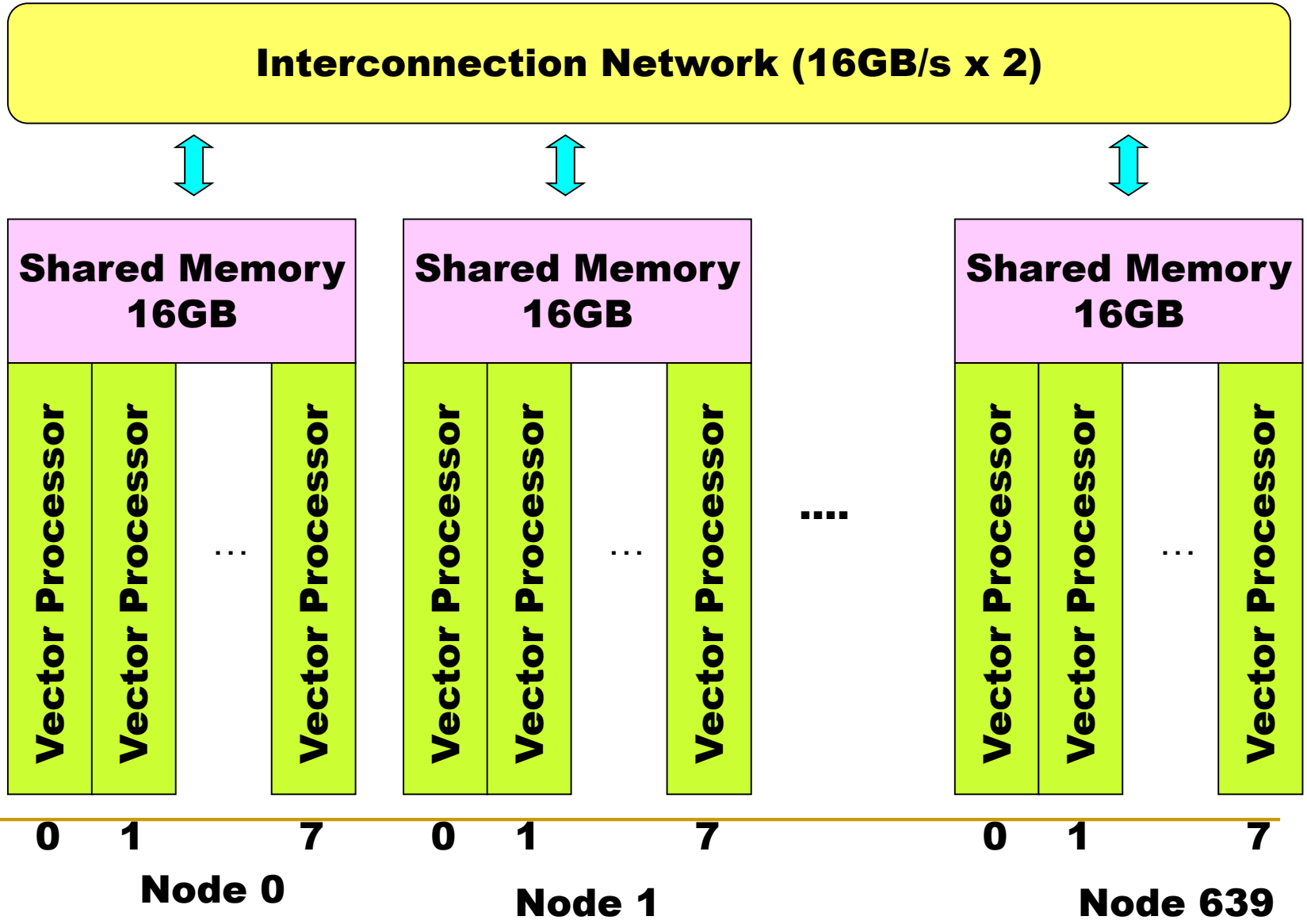


Single issue six-stage pipeline RISC with 16KB Instruction cache/ 8KB Data cache for L1

Total 3MB, 64byte Interleaved

The earth simulator

Peak performance
40TFLOPS



glossary 3

- Crossbar switch: クロスバスイッチ、ここでは主としてスイッチ本体を指すが、バッファも入れて考える場合もある
- Router:パケットを転送するためのハードウェア全体を指す
- WH, Virtual Channel:この授業のもっとあとで紹介するのでここでは深く追求しないでよい
- Non-blocking, blocking:出力ポートが重ならなければ、衝突が起きないのがノンブロッキング、出力ポートが重ならなくてもスイッチ内部で衝突するのがブロッキング
- HOL conflict:出線競合、出力ポートが重なることで起きる衝突

Homework 3

- Your computer uses PCIe gen2 x 8.

  1. How much maximum bandwidth can be used ?

  2. You want to improve the bandwidth.

  2-1. When you use PCIe gen2 x 16, how much maximum bandwidth can be used?

  2-2. You changed the bus to PCIe gen3 x 8, how much maximum bandwidth can be used?

  Just a simple calculation. You will spend only about 3 minutes.