Verteilte Systeme
(Betriebssysteme II)

Kapitel 2: Parallelrechnerarchitekturen

Prof. Dr. Wolfgang Küchlin

Dipl.-Inform., Dr. sc. techn. (ETH)

Arbeitsbereich Symbolisches Rechnen
Wilhelm-Schickard-Institut für Informatik
Fakultät für Informations- und Kognitionswissenschaften

Universität Tübingen

Steinbeis Transferzentrum
Objekt- und Internet-Technologien (OIT)

Wolfgang.Kuechlin@uni-tuebingen.de
http://www-sr.informatik.uni-tuebingen.de

Geschichte

Quelle: www-5.ibm.com/es/press/fotos/mainframe/

- Mainframe 1964: IBM 360

Geschichte

- Mainframe 1965: IBM 360 Model 20 in Böblingen

Quelle: www-5.ibm.com/es/press/fotos/mainframe/

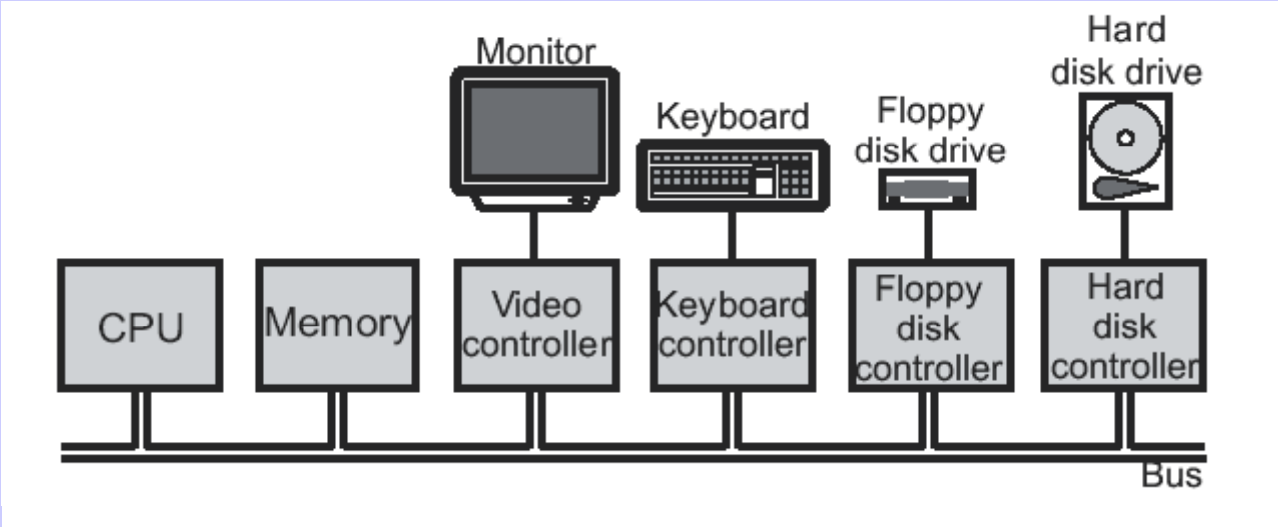
Geschichte: erster IBM-PC (1981)

- Systemeinheit (9,5 kg)
  - 16 bit 8088 Proc. / 8 bit Bus
  - 64 kB RAM (max 640kB)
  - extern. Kassette
  - 5,25“ Diskette 360kB
  - BASIC in ROM 40kB
  - DOS
- Monitor (7,9 kg)
  - 25 Zeilen x 80 Zeichen
  - 720x350 Pixel, grün
- Erweiterungseinheit (12 kg)
  - 10 MB Festplatte

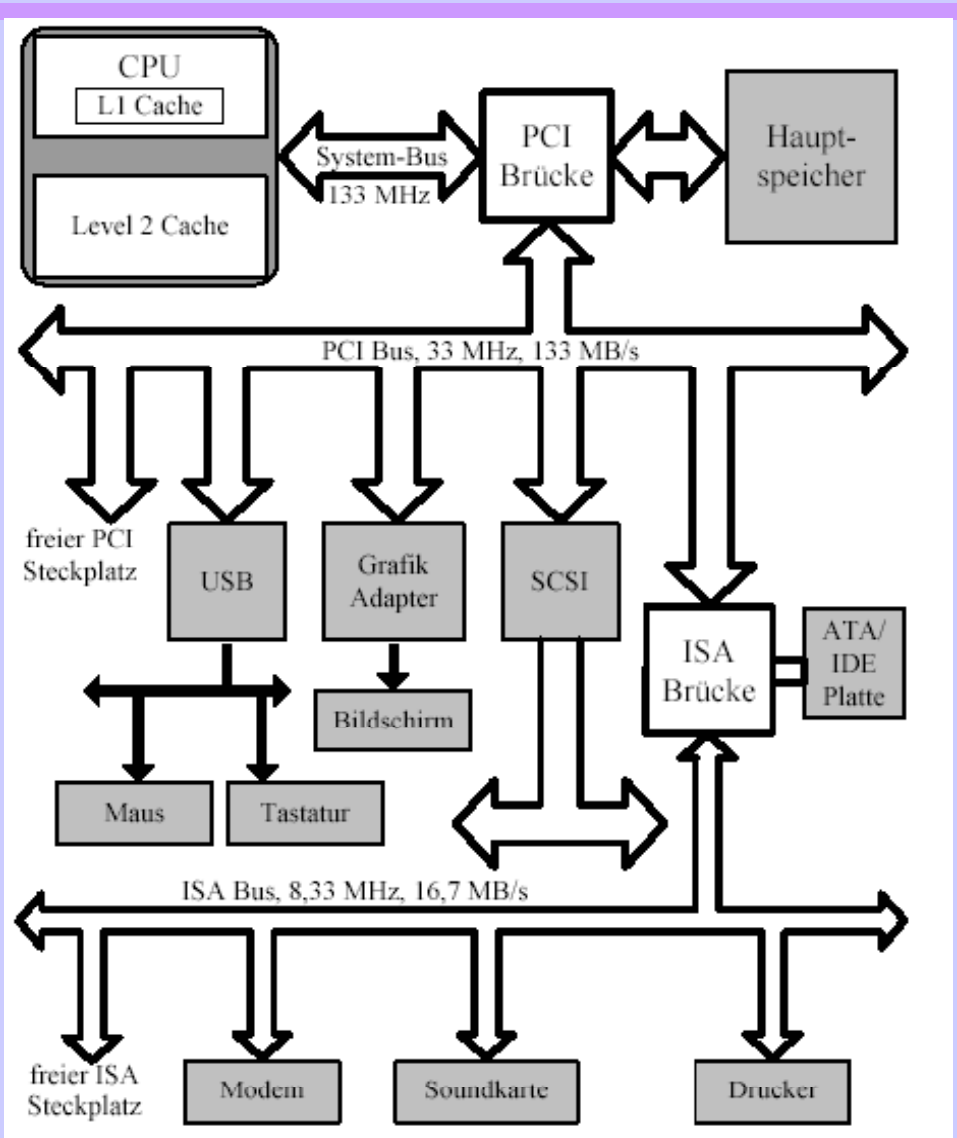
Quelle: http://www-5.ibm.com/es/press/fotos/20aniversario/index.html

Klassische Sequentielle Rechner

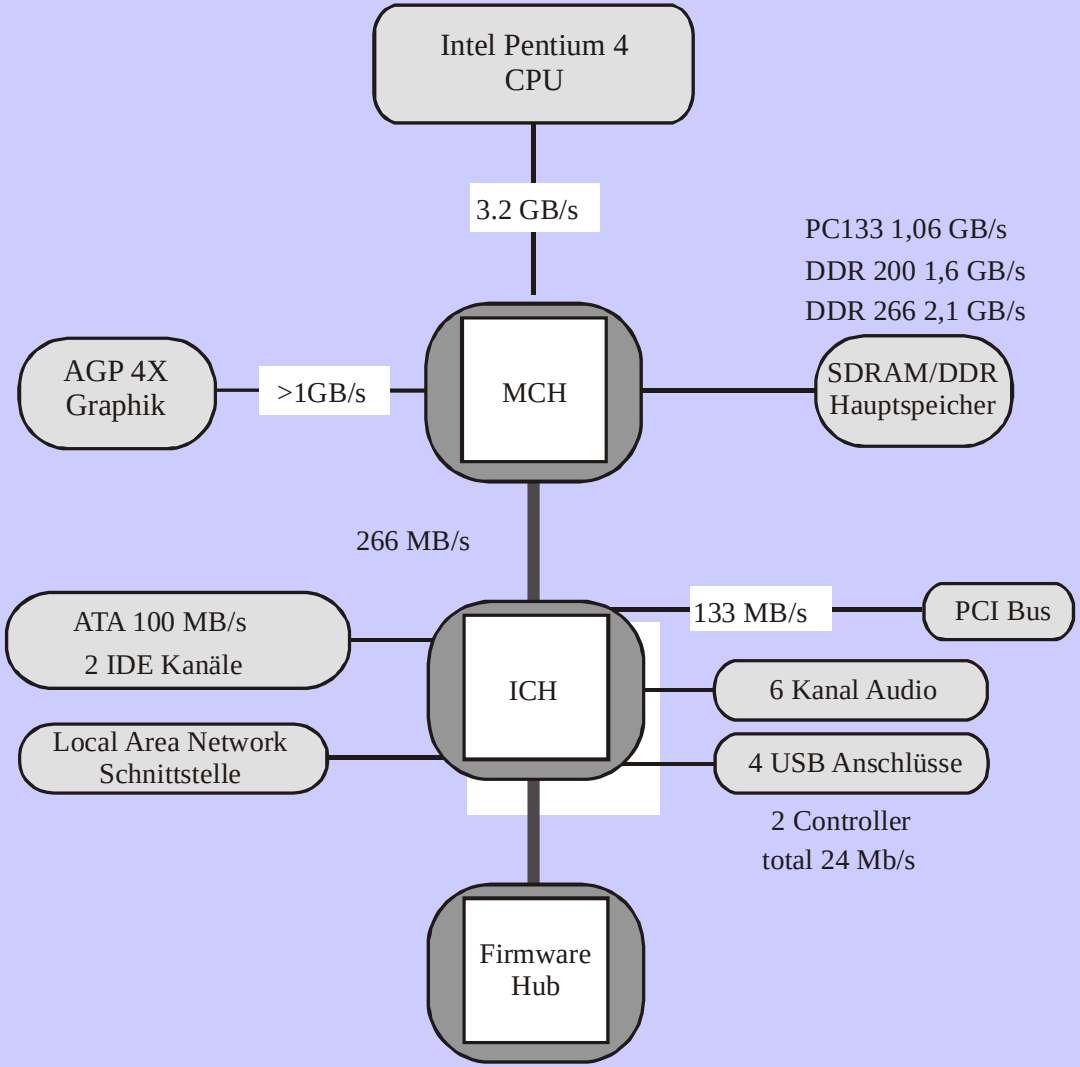
- Architektur eines einfachen Computersystems mit Bus



Architektur Intel Pentium II



Architektur Intel Pentium 4

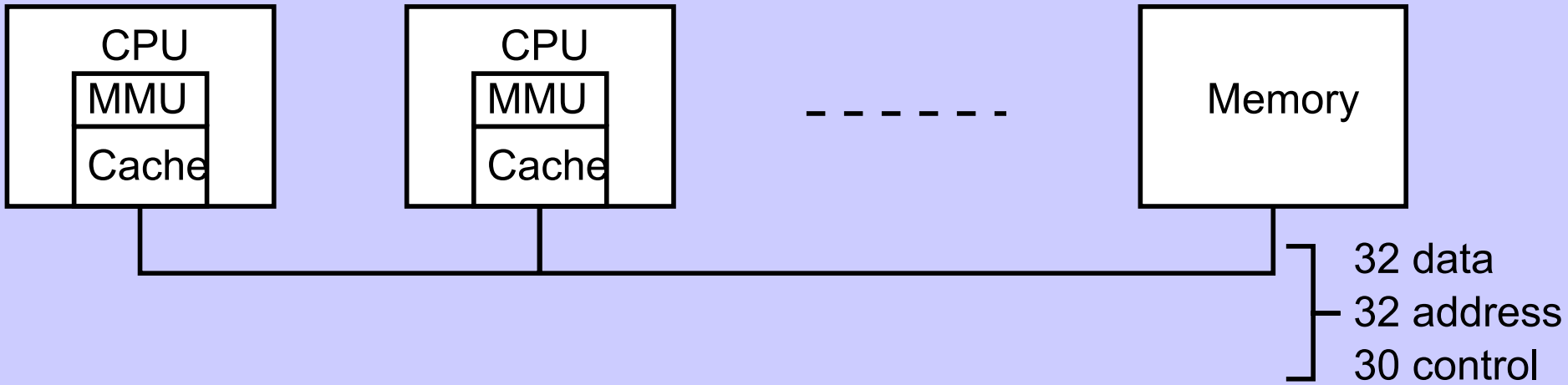


Parallele und verteilte Rechner

- Multiprozessoren
  - Mehrere Prozessoren (Prozessor-Chips) in einem Computer
  - teilen sich Hauptspeicher und Geräte
- Multicomputer
  - Mehrere zu einem Gesamtsystem verbundene Computer
  - Jeder Computer hat eigenen Speicher und Geräte
  - Gewisse Geräte können gemeinsam sein (LAN-Karten, …)
- Multicore Chips (mehrkernige Prozessoren)
  - Jeder Prozessor-Chip hat mehrere Prozessoren
  - Schwächer: ein Prozessor, mehrere Registersätze für multi-threading

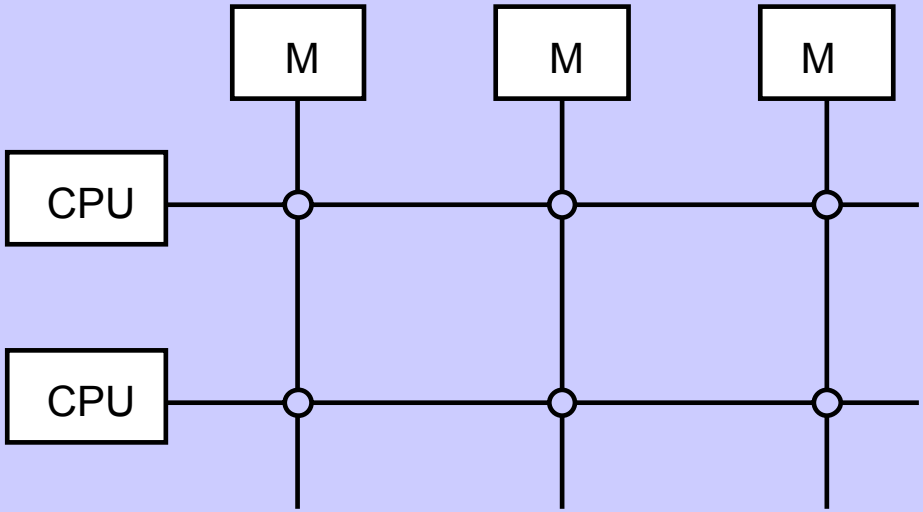
Parallele und verteilte Rechner

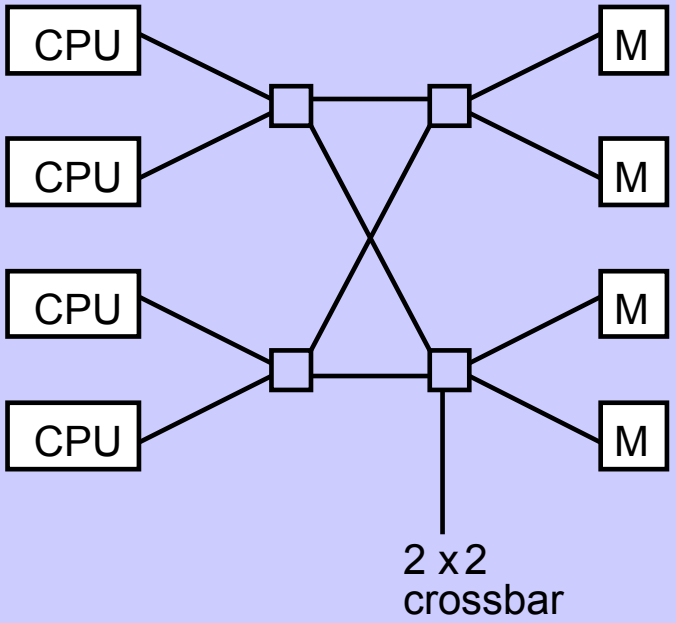
- Bus basierte Multiprozessoren



Parallele und verteilte Rechner

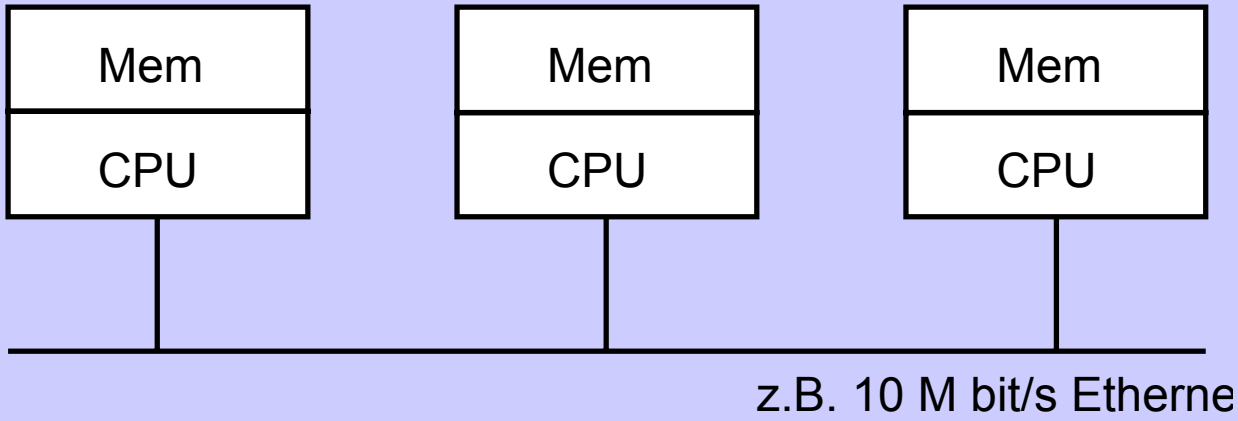
- Schaltnetzbasierte Multiprozessoren





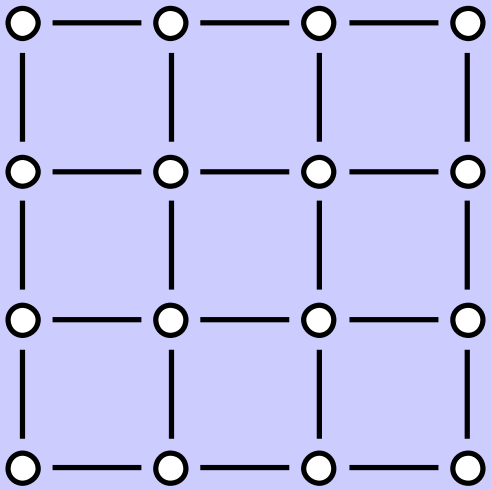
Parallele und verteilte Rechner

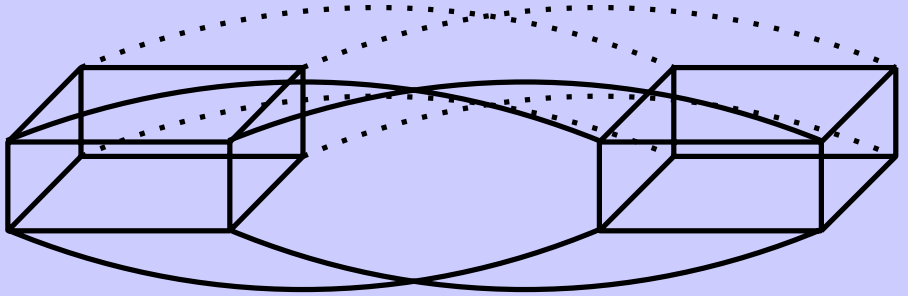
- Bus-basierte Multicomputer



Parallele und verteilte Rechner

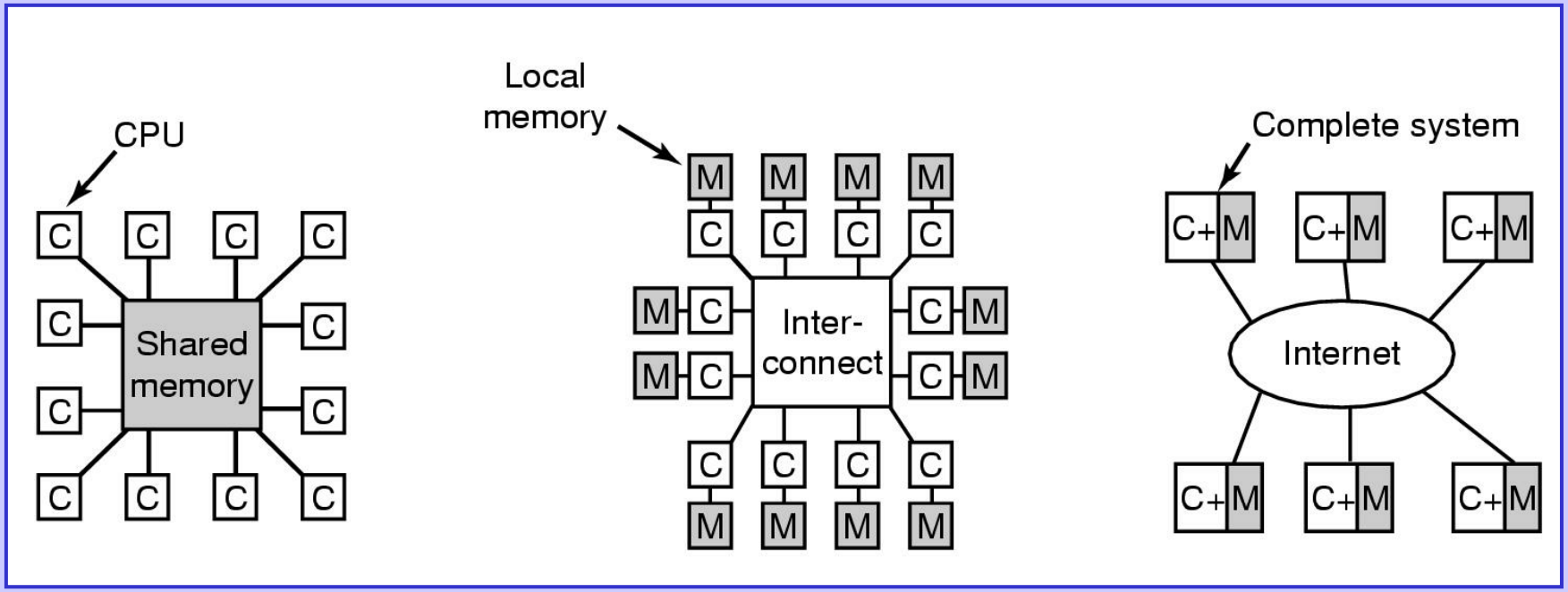
- Schaltnetzverbundene Multicomputer
  - Gitter
  - Hypercube





Multiprozessorsysteme

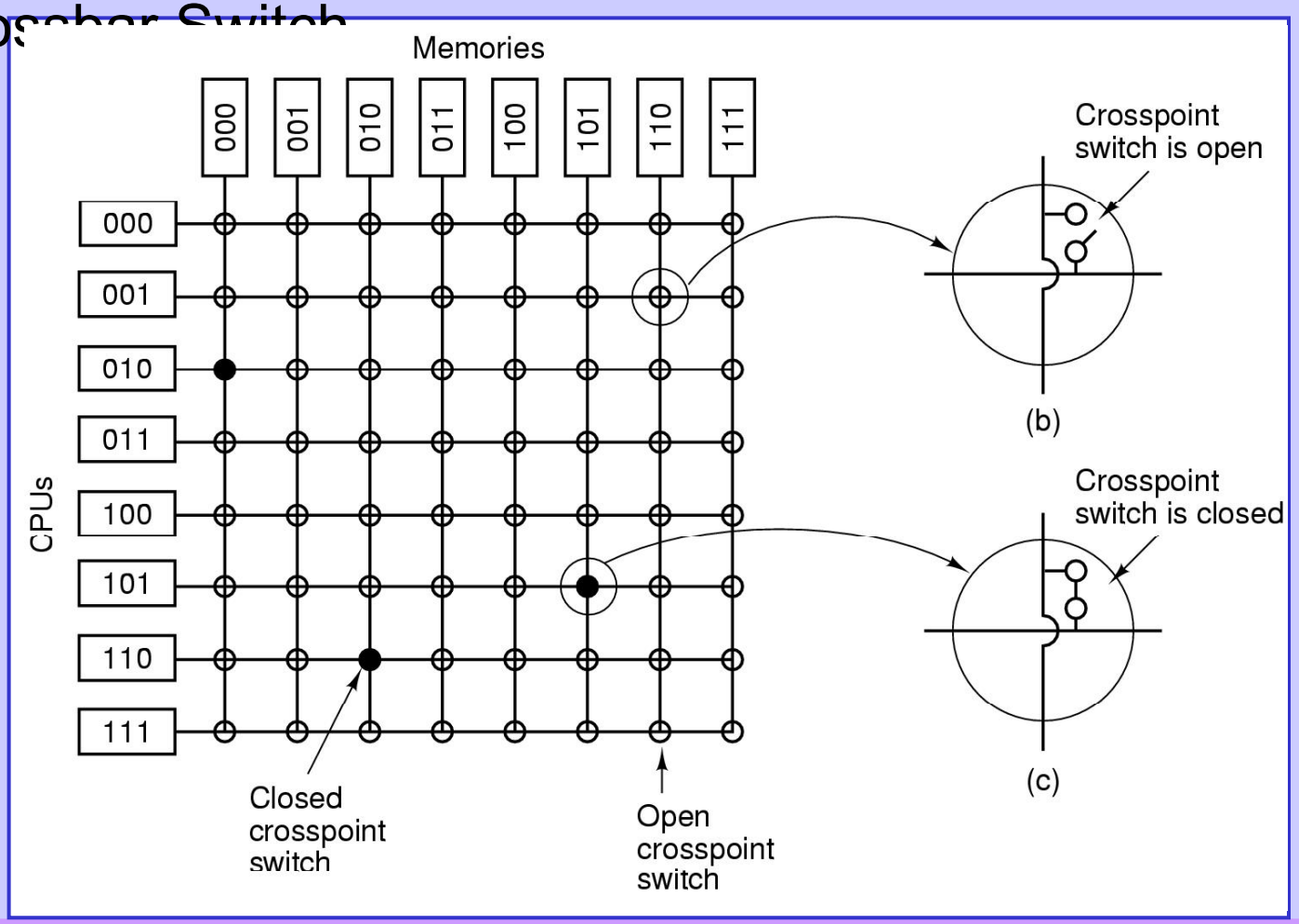
- Definition: Multiprozessorsystem
  - Ein System in dem zwei oder mehr Prozessoren sich den Zugriff auf gemeinsamen Hauptspeicher teilen.



Quelle: „Operating Systems“, Tanenbaum, Abb.8-?

Multiprozessorsysteme – Hardware

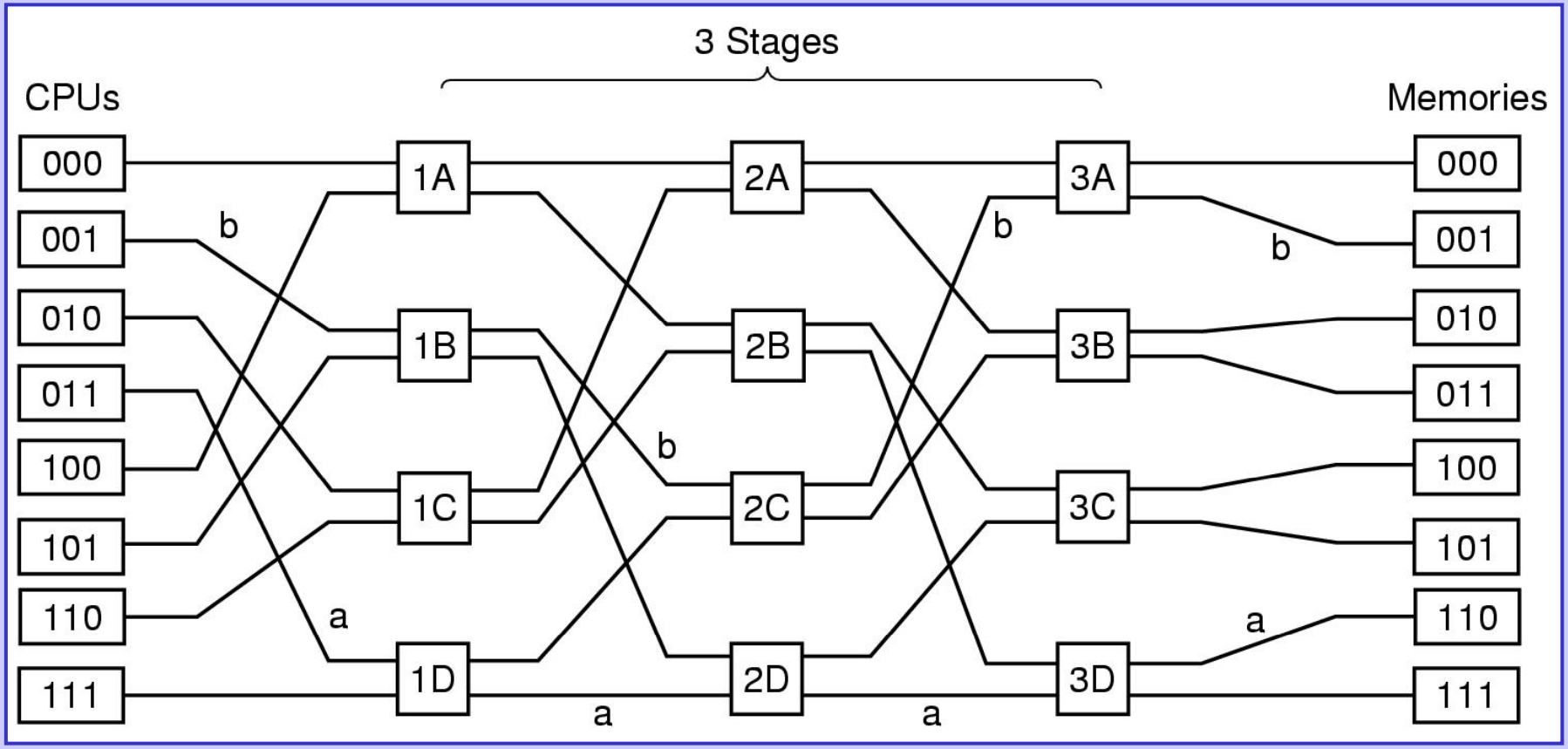
- UMA (uniform memory access time) Multiprozessoren mit Crossbar Switch



Quelle:
„Operating Systems“, Tanenbaum, Abb.8-?

Multiprozessorsysteme – Hardware

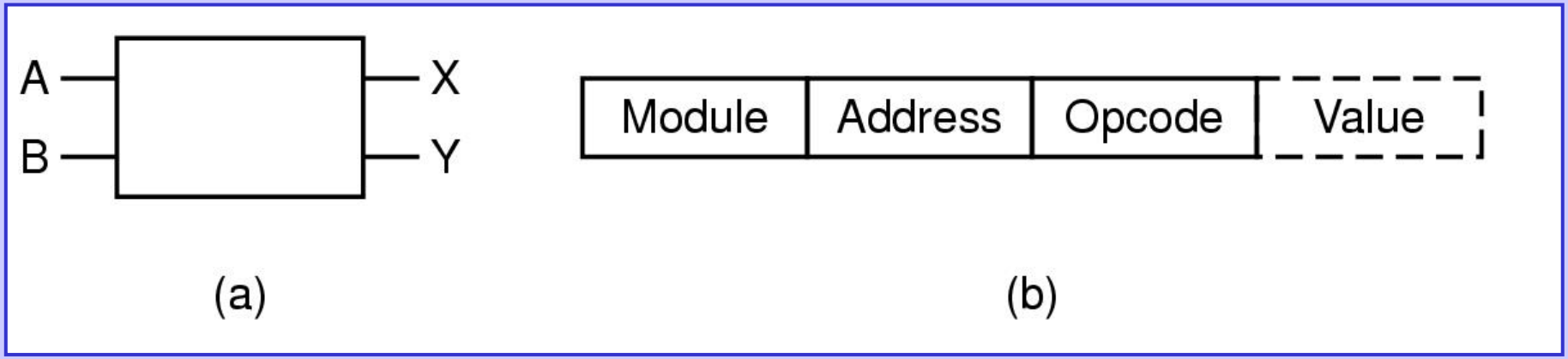
- Netzwerk mit Omega Switching



Quelle: „Operating Systems“, Tanenbaum, Abb.8-?

Multiprozessorsysteme – Hardware

- UMA-Multiprozessoren mit Multistage Switch Netzwerken können aus 2×2 Switches gebaut werden.



- Speicherzugriffe durch Messages, die geroutet werden
- Module = Adresse des Speichermoduls, Opcode = {R | W}
  - Switch in Stufe i interpretiert i-tes Bit (von links)
  - i=0 → oberer Ausgang X; i=1 → unterer Ausgang Y

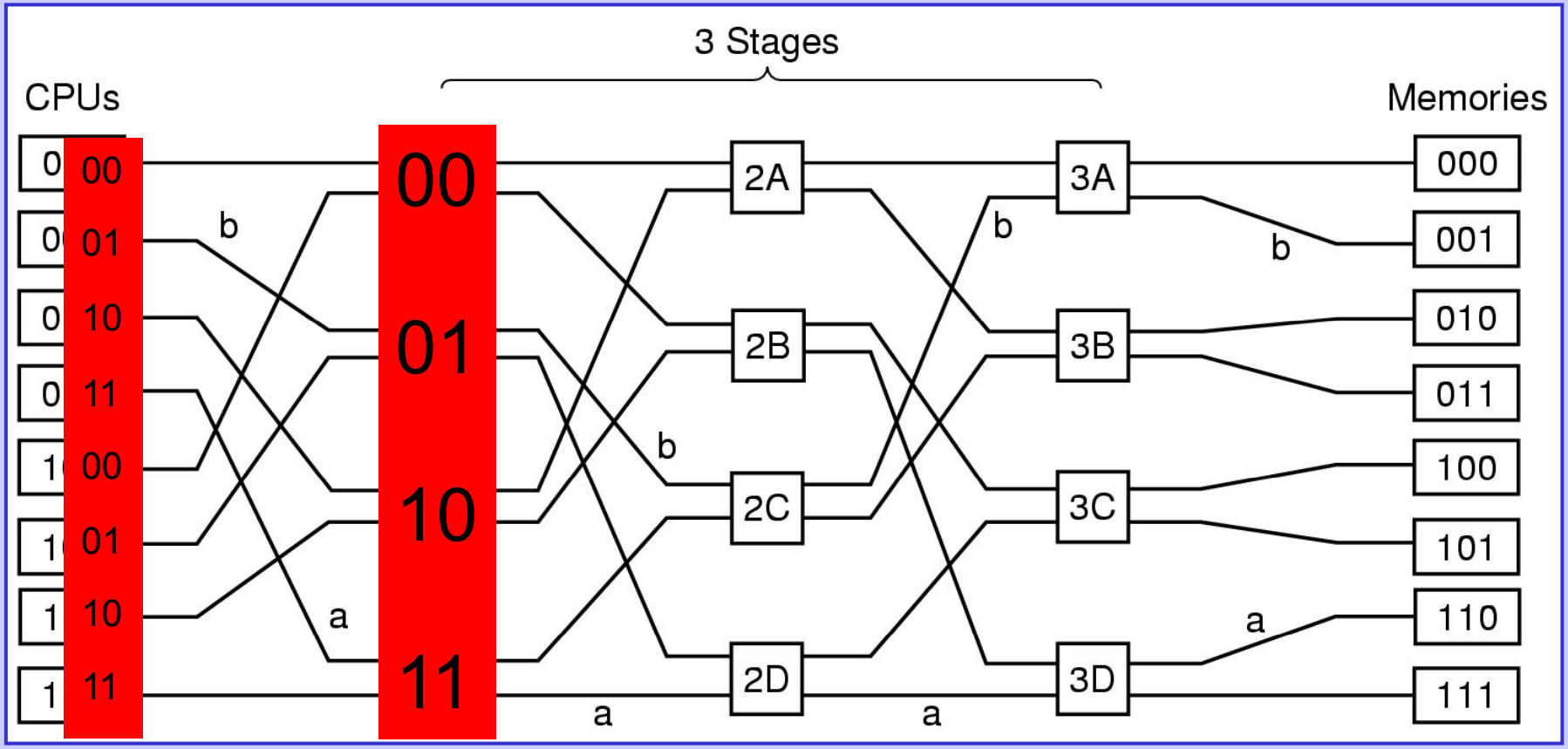
Quelle: „Operating Systems“, Tanenbaum, Abb.8-?

Multiprozessorsysteme – Hardware

- Bauprinzip von Netzwerken mit Omega Switching
  - Bei n CPU und n Memory-Einheiten log(n)*(n/2) Switches
  - (vgl. n*n Switches im crossbar)
  - Verschaltung analog zum Kartenmischen
  - Im Beispiel:
    - Message von jeder CPU a.b.c zu jedem Speicher m.n.o
    - Switch Stufe 1 Nummer x.y routet alle messages a.x.y.m.n.o
    - Switch Stufe 2 Nummer x.y routet alle messages a.b.x.y.n.o
    - Switch Stufe 3 Nummer x.y routet alle messages a.b.c.x.y.o

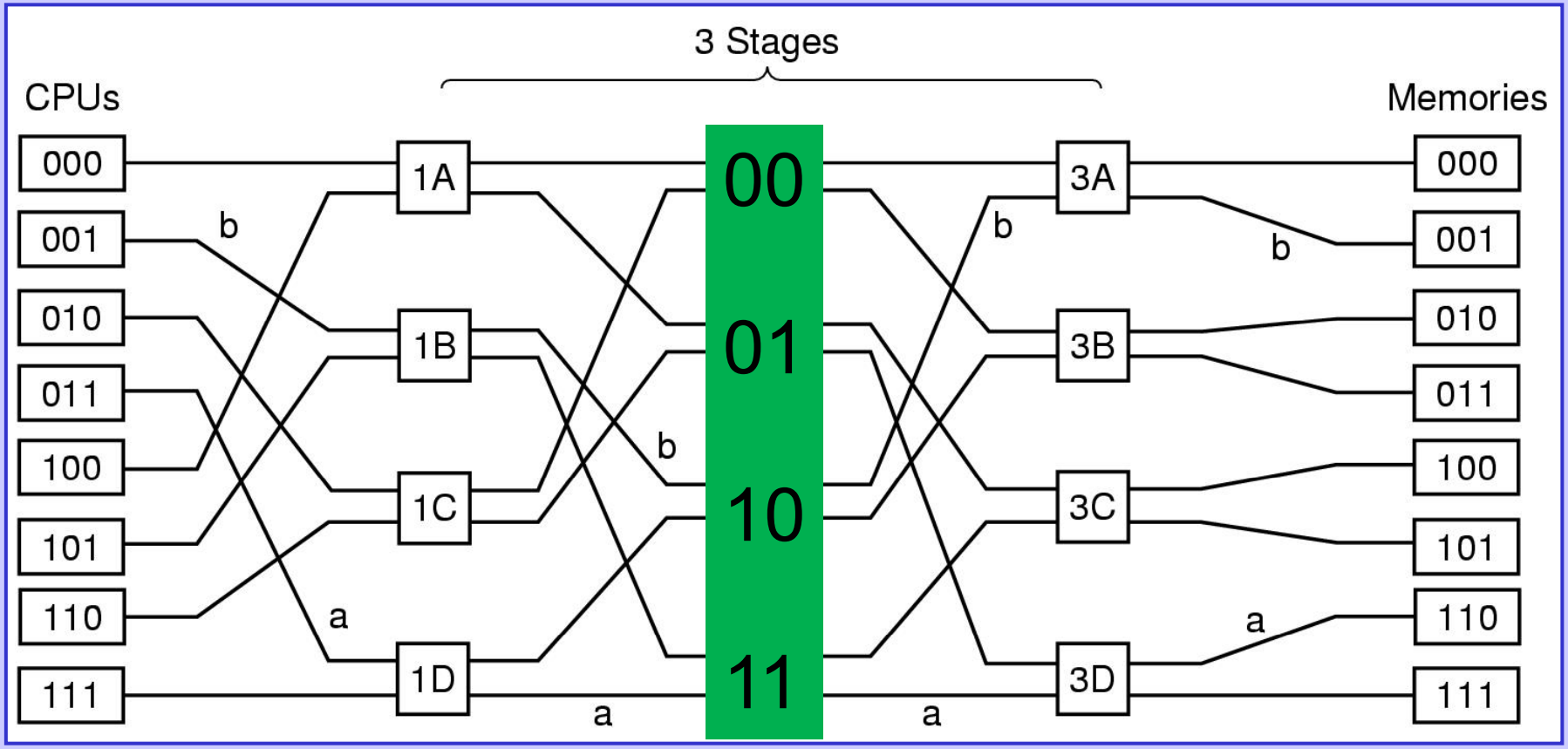
Multiprozessorsysteme – Hardware

- 1. Stufe: Switch $x_1.y_1$ routet alle Pakete a.$x_1.y_1$.r.s.t



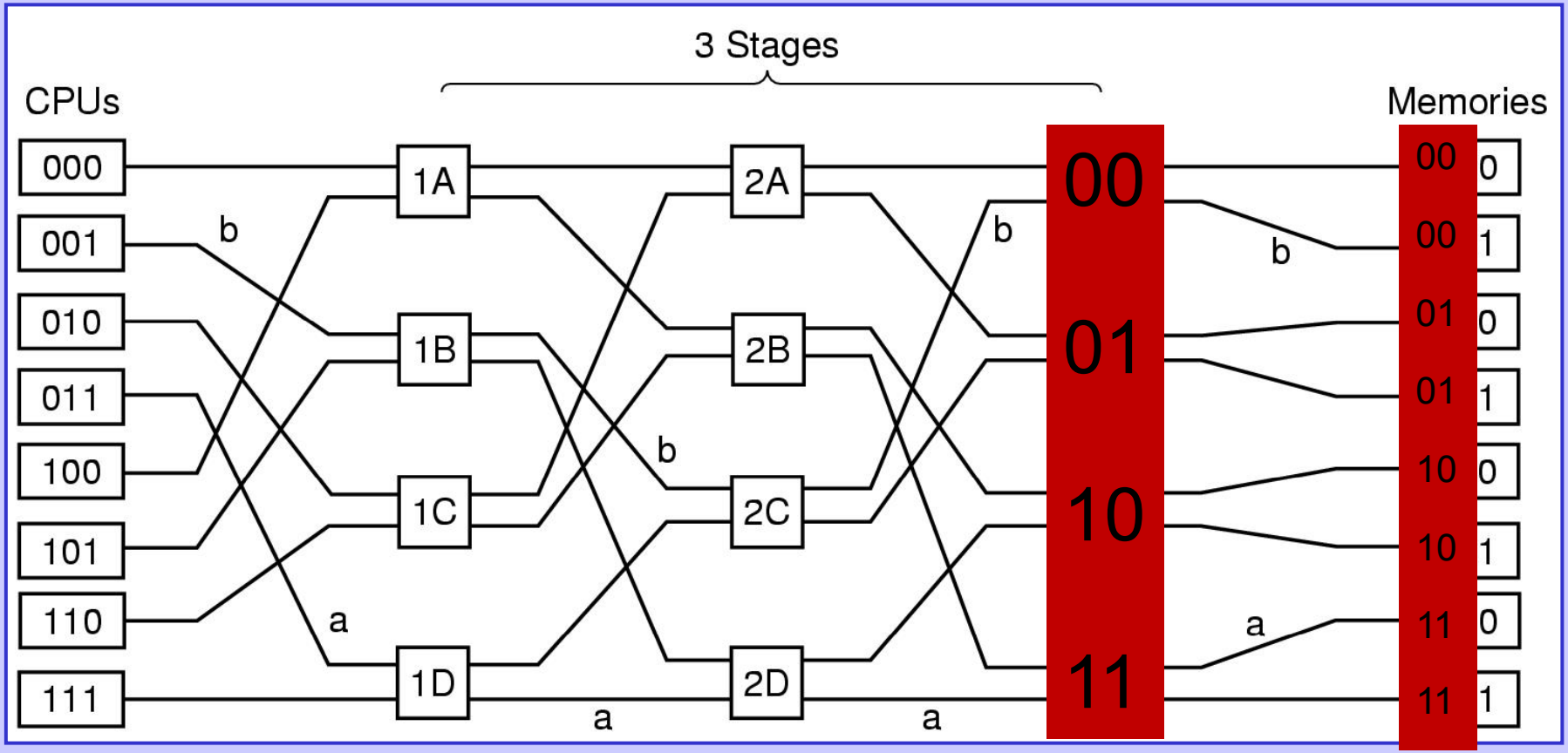
Multiprozessorsysteme – Hardware

- 2. Stufe: Switch $x_2.y_2$ routet Pakete a.b.$x_2.y_2$.s.t



Multiprozessorsysteme – Hardware

- 3. Stufe: Switch $x_3.y_3$ routet Pakete a.b.c.$x_3.y_3$.o



Multiprozessorsysteme – Hardware

- Begründung des Bauprinzips (am Beispiel 8x8)
  - Switch Stufe 1 Nummer $x_1.y_1$ routet alle messages a.$x_1.y_1$.r.s.t
    - Es gibt je 2 CPUs mit Adresse a.x.y.
    - Zu zeigen: von switch $x_1.y_1$ sind alle Speichermodule r.s.t erreichbar.
    - Genügt zu zeigen: Es gibt 2 switches der Stufe 2, die über die beiden Ausgänge 0 und 1 von $x_1.y_1$ erreichbar sind, und von denen aus die Module 0.s.t und 1.s.t erreichbar sind. (Dies sind $x_2.y_2$ = $y_1.0$ und $y_1.1$).
  - Switch Stufe 2 Nummer $x_2.y_2$ routet alle messages a.b.$x_2.y_2$.s.t
    - Zu zeigen: von switch $x_2.y_2$ sind alle Speichermodule y_2.s.t erreichbar.
    - Genügt zu zeigen: Es gibt 2 switches der Stufe 3, die über die beiden Ausgänge 0 und 1 von $x_2.y_2$ erreichbar sind, und von denen aus die Module y_2.0.t und y_2.1.t erreichbar sind. (Dies sind $x_3.y_3$ = $y_2.0$ und $y_2.1$).
  - Switch Stufe 3 Nummer $x_3.y_3$ routet alle messages a.b.c.$x_3.y_3$.t
    - Von switch $x_3.y_3$ sind die Speichermodule $x_3.y_3$.t erreichbar.

Multiprozessorsysteme – Hardware

NUMA Multiprocessor Characteristics

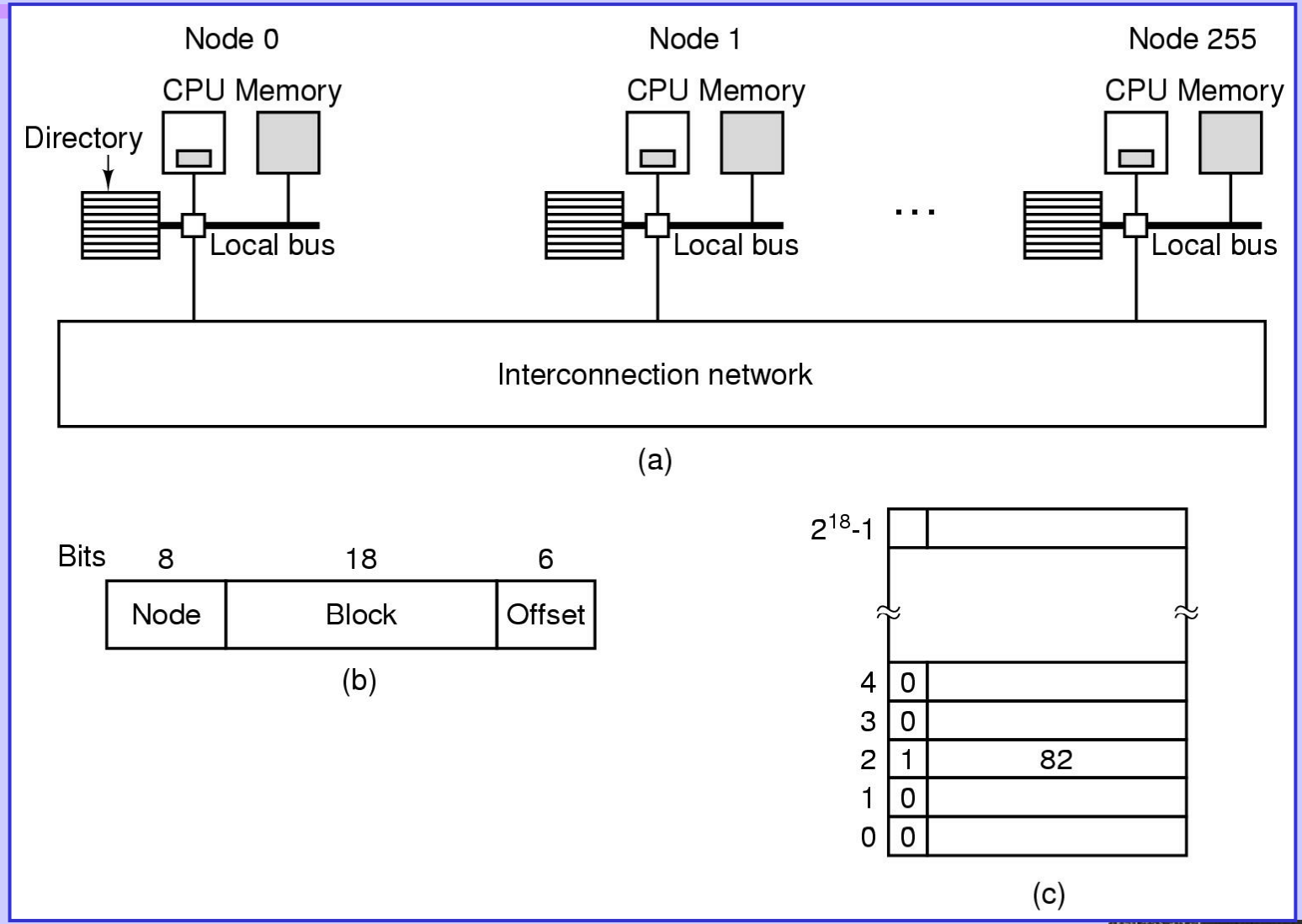
(non-uniform memory access time)

- Single address space visible to all CPUs
- Access to remote memory via commands
  - LOAD
  - STORE
- Access to remote memory slower than to local

Multiprozessorsysteme – Hardware

CC-NUMA

- cache coherent NUMA
- Directory-basierter Multiprozessor mit 256 Knoten
- Felder einer 32-Bit Adresse (Block = Cache Line)
- Verzeichnis im Knoten 36: Zeile 2 ist im Knoten 82 gecachet



Quelle: „Operating Systems“, Tanenbaum, Abb.8-?

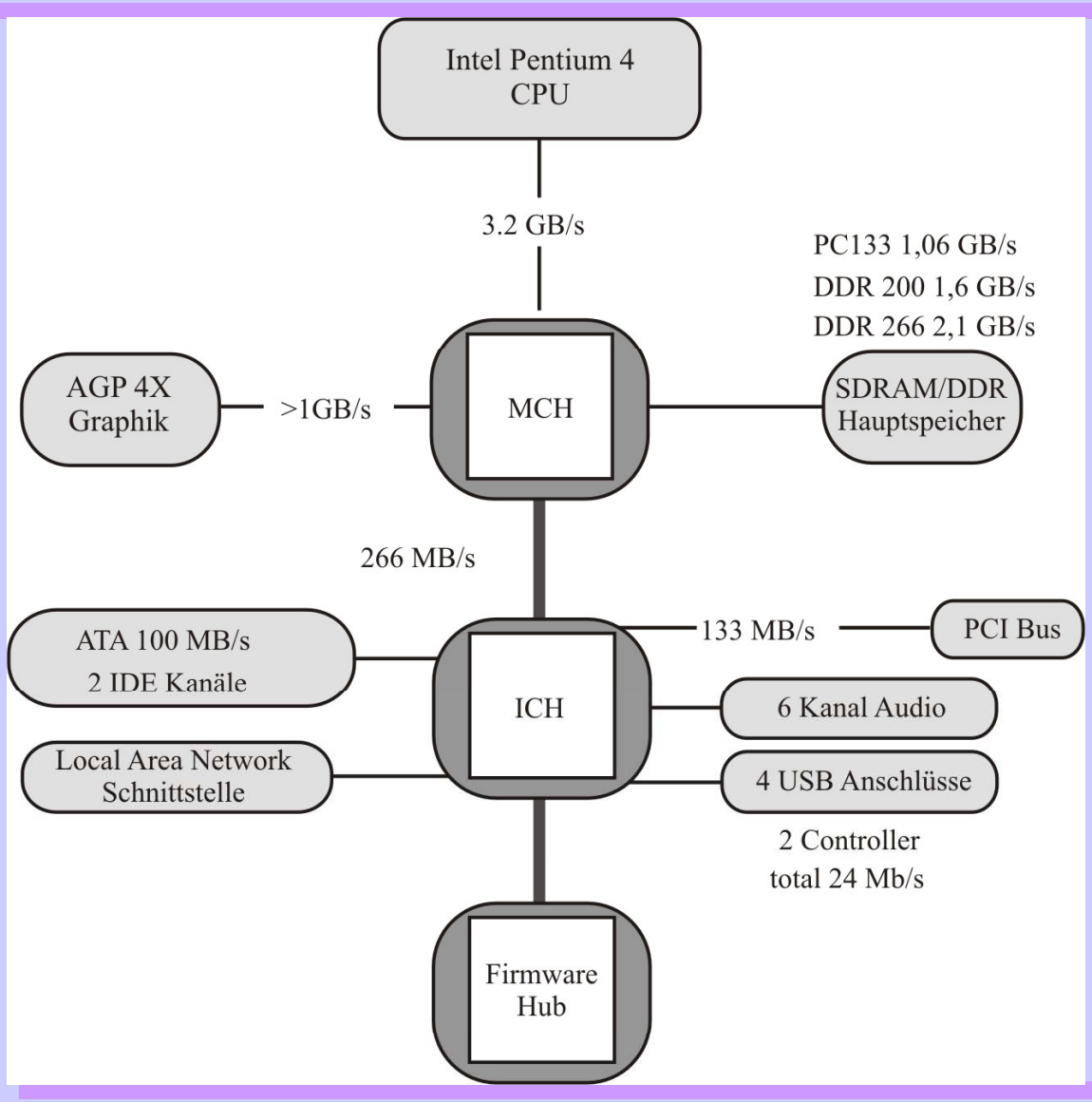
Multiprozessorsysteme – Hardware

- Verzeichnisbasierter Multiprozessor mit 256 (= 2^8) Knoten
- 16 MB RAM in jedem Knoten (= 2^{24} Byte)
- Speicher insgesamt: 2^{32} Bytes
  - d.h. 2^{26} Cache Lines zu 2^6 Bytes
- 32-bit Speicheradressen aufgeteilt in drei Felder
  - 8bit Knotenadresse
  - 18bit Cacheblock-Adresse
  - 6bit (Byte-)Offset innerhalb des Cacheblocks
- Cacheblock-Verzeichnis in jedem Knoten
  - 2^{18} Einträge: 1bit Cache-Flag + 8bit Knotenadresse

Überblick

1. Single Processor
2. UMA Multiprocessor
3. NUMA Multiprocessor
4. NORMA Multicomputer
5. SUN T1 (Niagara) 8 way Multicore

Single Processor

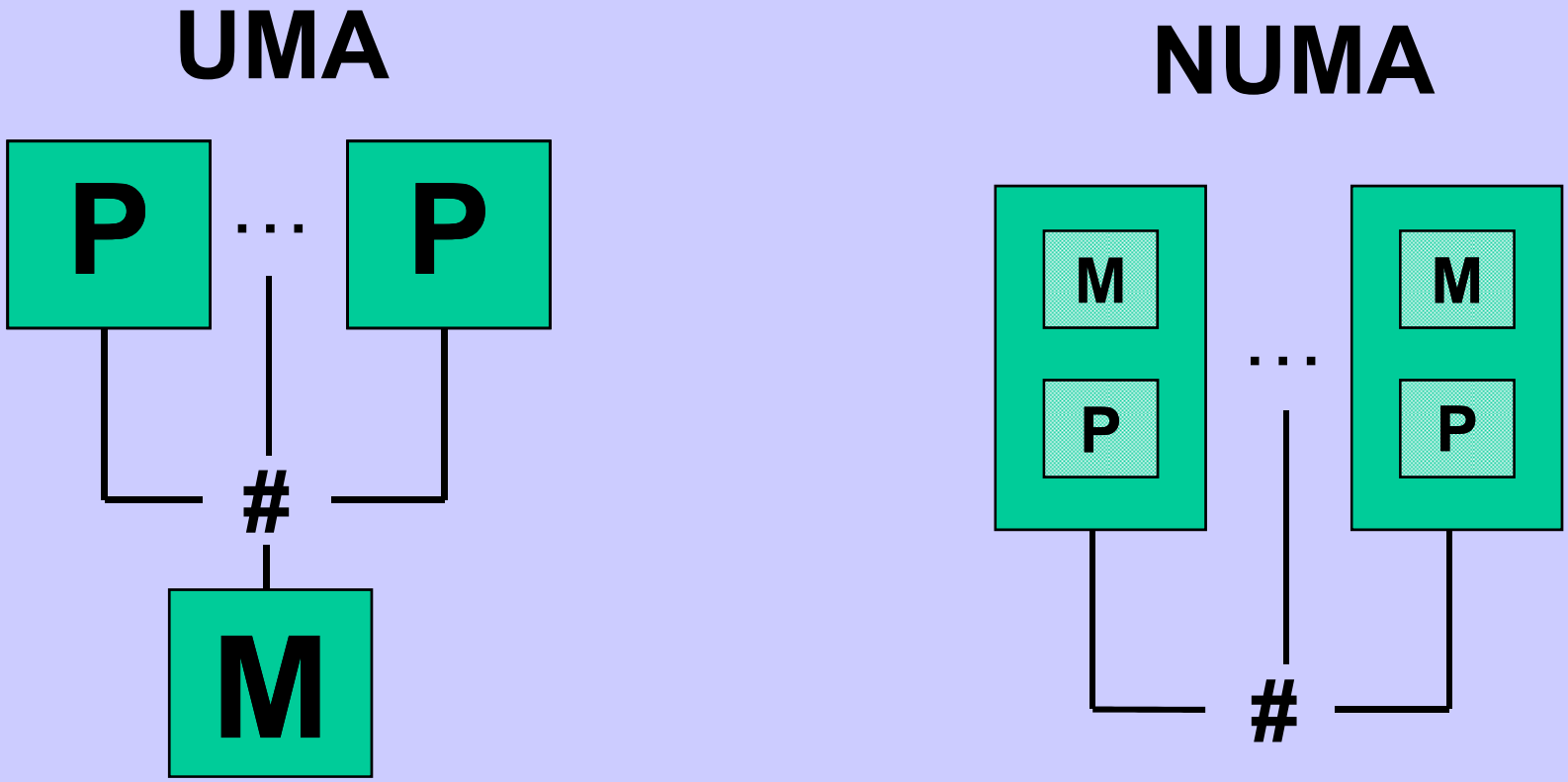


Single

P

M

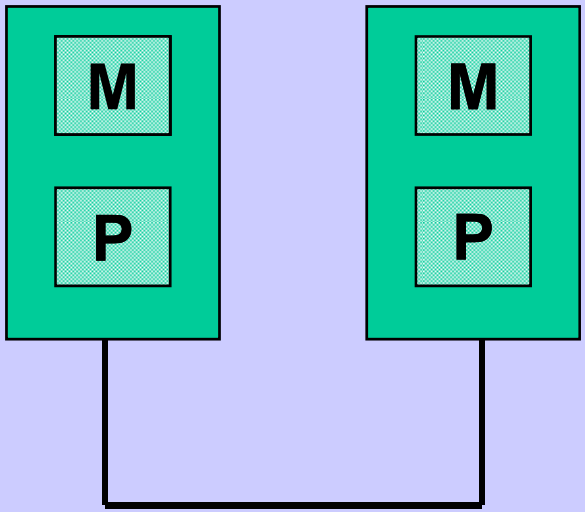
UMA / NUMA



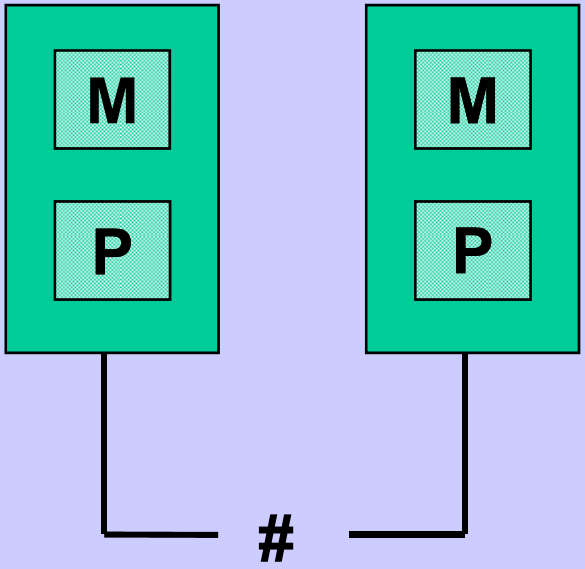
UMA – Uniform Memory Access
NUMA – Non Uniform Memory Access

Verteilt / NUMA

Verteilt

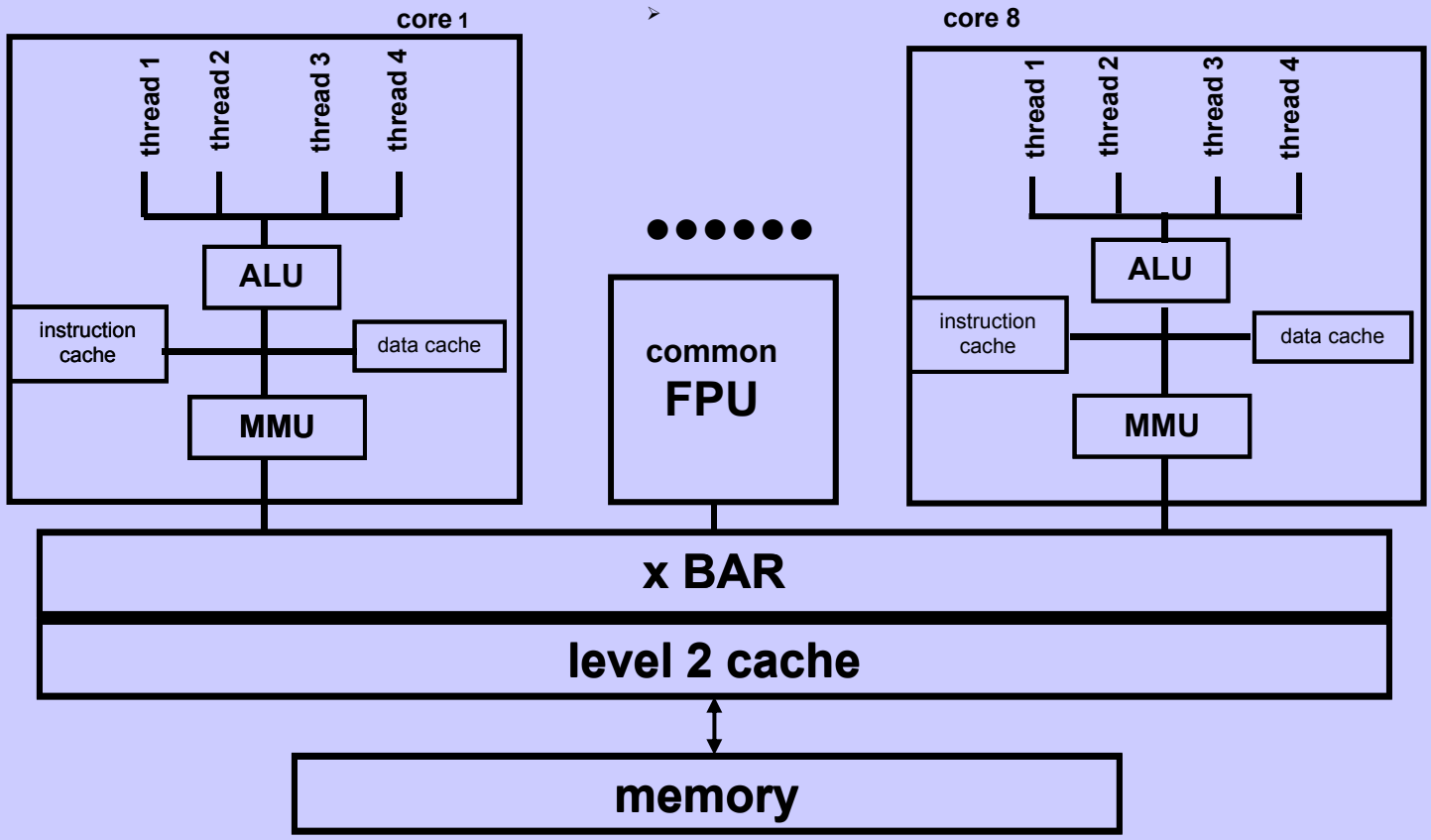


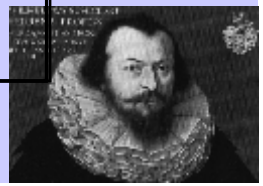
Numa



A Chip Multiprocessor

SUN Niagara 1 Multicore (T1)
8 Way Multicore x 4 Way Multi-Threaded

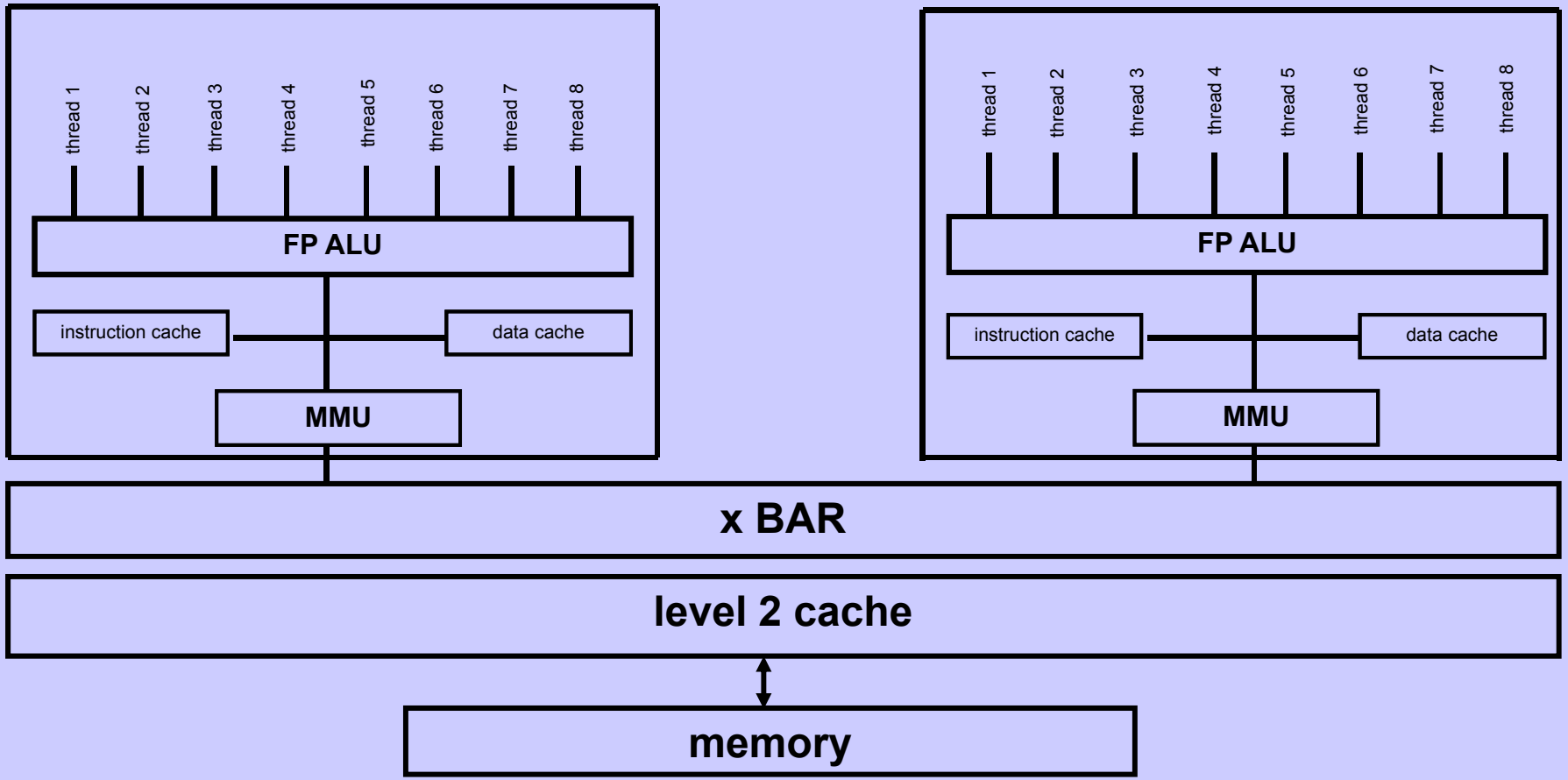




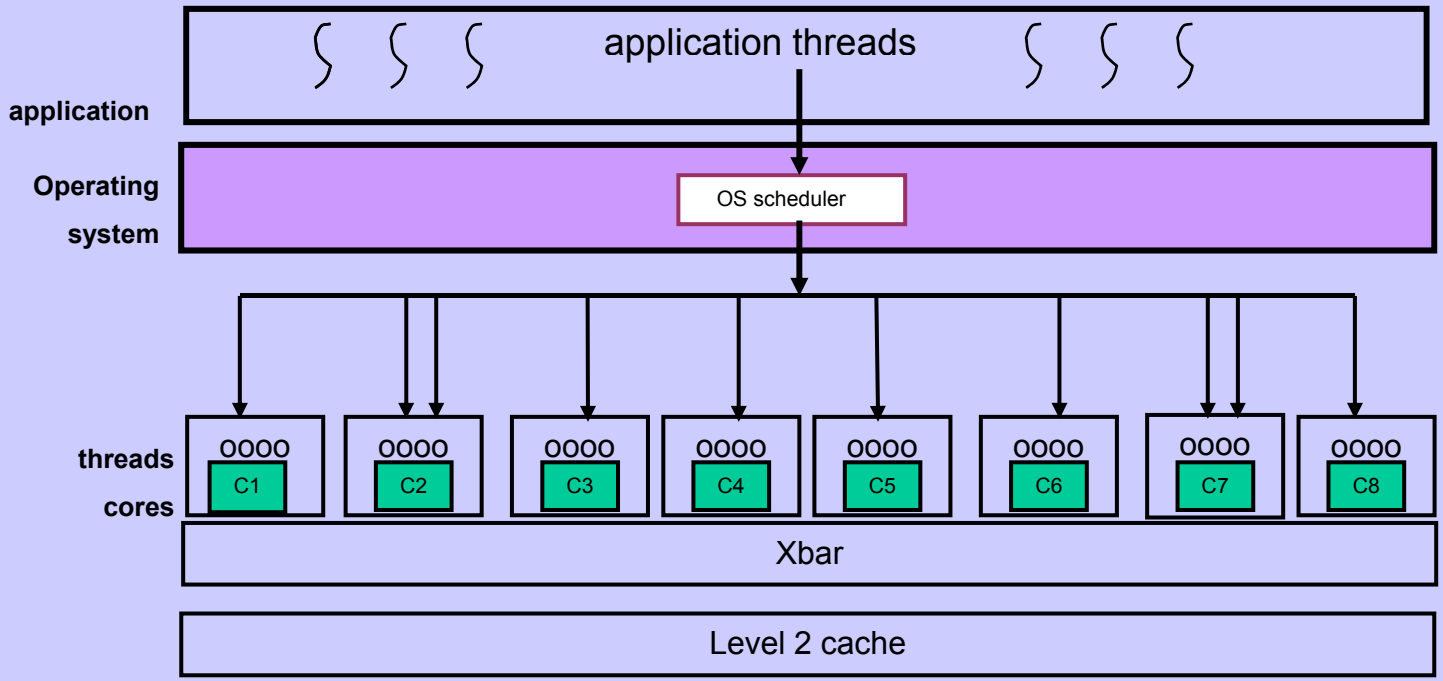
A Chip Multiprocessor

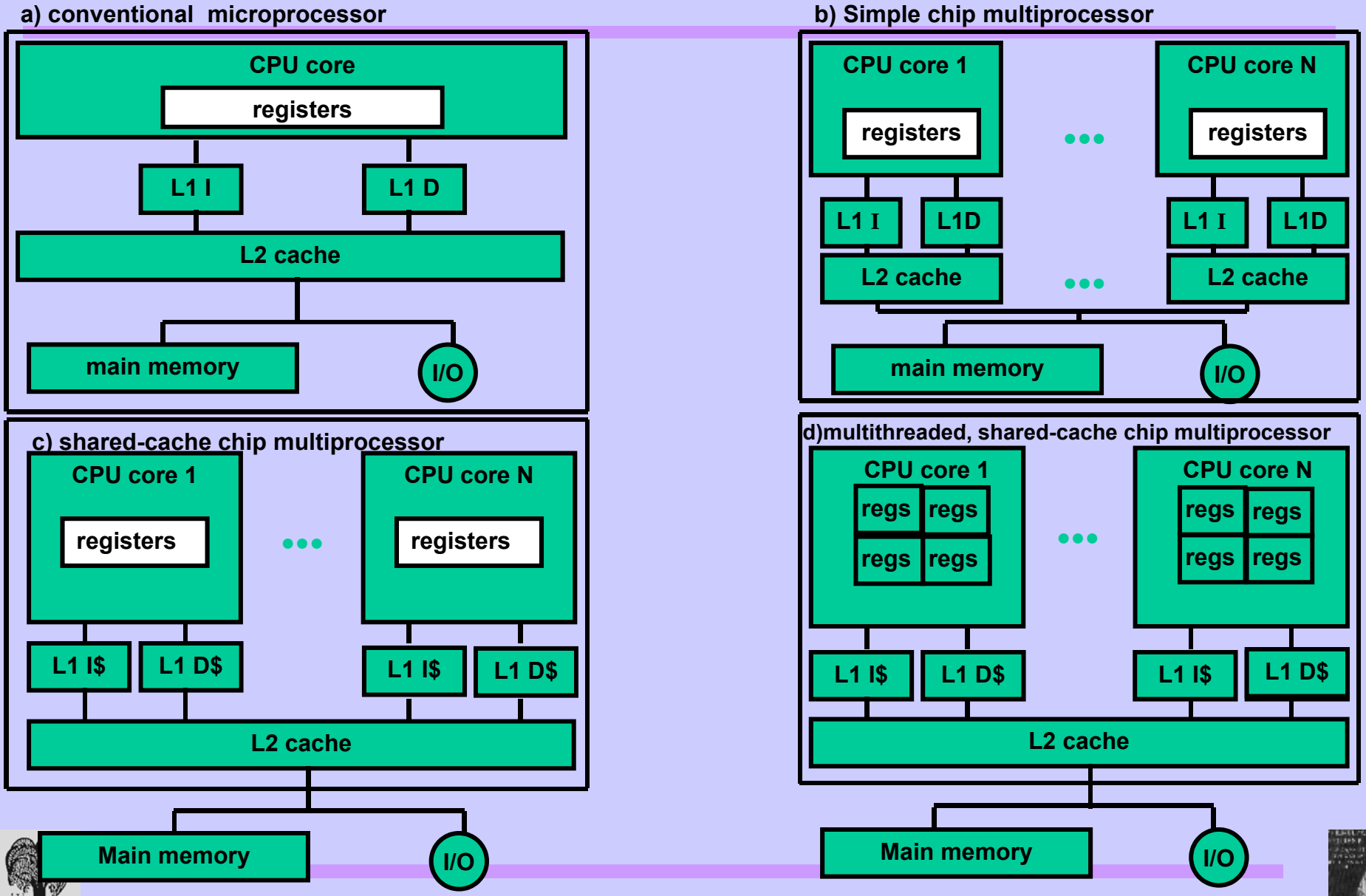
SUN Niagara 2 Multicore (T2)

16 Way Multicore x 8 Way Multi-Threaded



Software Threads Scheduling on CMP Cores/ Threads




a) conventional microprocessor
CPU core
registers
L1 I
L1 D
L2 cache
main memory
I/O
b) Simple chip multiprocessor
CPU core 1
registers
L1 I
L1D
L2 cache
CPU core N
registers
L1 I
L1D
L2 cache
main memory
I/O
c) shared-cache chip multiprocessor
CPU core 1
registers
L1 I$
L1 D$
CPU core N
registers
L1 I$
L1 D$
L2 cache
Main memory
I/O
d)multithreaded, shared-cache chip multiprocessor
CPU core 1
regs
regs
regs
regs
L1 I$
L1 D$
CPU core N
regs
regs
regs
regs
L1 I$
L1 D$
L2 cache
Main memory
I/O

IBM Power 5 Processor