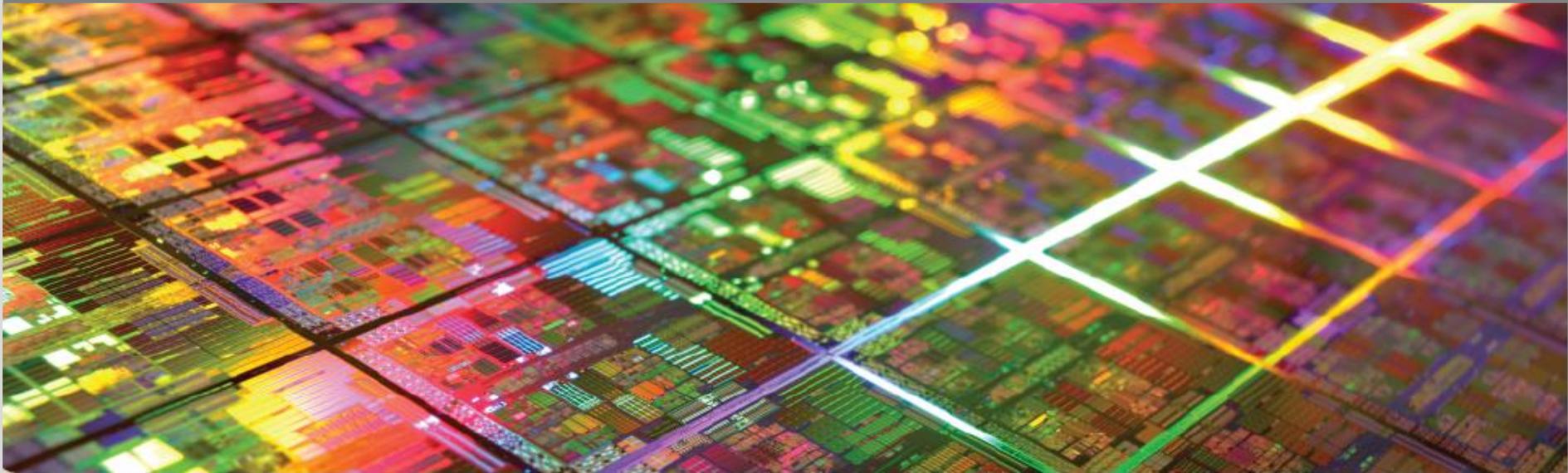


# Rechnerstrukturen

Vorlesung im Sommersemester 2016

Prof. Dr. Wolfgang Karl

Institut für Technische Informatik (ITEC), Lehrstuhl für Rechnerarchitektur und Parallelverarbeitung



# Vorlesung Rechnerstrukturen

## Kapitel 3: Multiprozessoren – Parallelismus auf Prozess-/ Blockebene

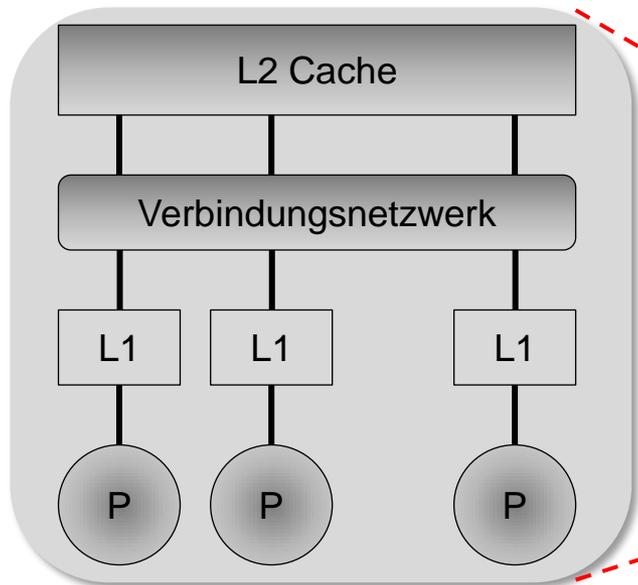
- 3.1 Motivation
- 3.2 Allgemeine Grundlagen
- 3.3 Parallele Programmierung
- 3.4 Quantitative Maßzahlen
- 3.5 Verbindungsnetzwerke

# Verbindungsstrukturen

## Verbindungsnetzwerke in einem Multiprozessorsystem

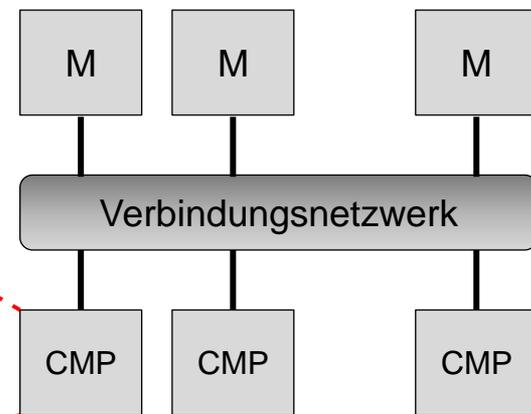
### ■ Beispiel

CMP mit gemeinsamem L2 Cache



Verbindungsnetzwerk auf dem Chip

Multiprozessor

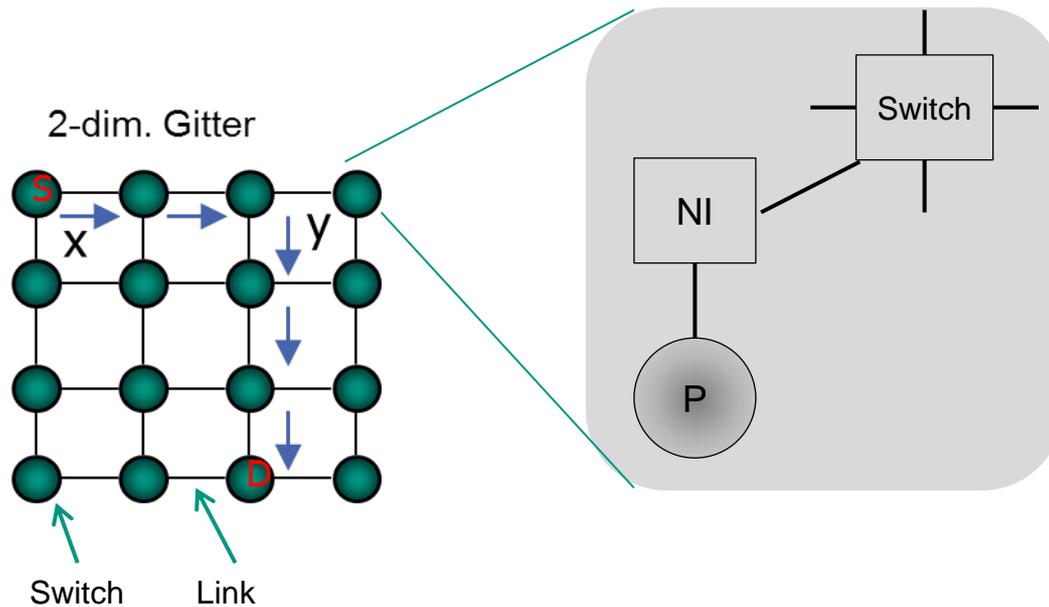


Verbindungsnetzwerk im System

# Verbindungsstrukturen

## Verbindungsnetzwerke in einem Multiprozessorsystem

### ■ Beispiel



# Verbindungsstrukturen

## Verbindungsnetzwerke in Multiprozessoren

- Ermöglichen die Kommunikation und Kooperation zwischen den Verarbeitungselementen (Knoten)
  - Zuverlässiger Austausch von Informationen
- Einsatz eines Verbindungsnetzwerks
  - Chip-Multiprozessor (CMP)
    - Network on Chip (NoC)
  - Multiprozessor mit verteiltem Speicher (nachrichtenorientierter Multiprozessor)
    - Verbinden physikalisch jeden Knoten für das Versenden von Nachrichten
    - Direkte Send/Receive-Kommunikation zwischen den Knoten
  - Multiprozessor mit gemeinsamem Speicher
    - Ermöglicht den Zugriff aller Knoten auf den gemeinsamen Speicher
    - Kommunikation durch Lesen und Schreiben auf gemeinsame Daten

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Latenz

#### ■ Übertragungszeit einer Nachricht $T_{msg}$

- die Zeit, die für das Verschicken einer Nachricht von einer bestimmten Länge zwischen zwei Prozessoren benötigt wird

#### ■ Die Übertragungszeit setzt sich zusammen aus:

##### ■ der **Startzeit** $t_s$ (Message Startup Time):

- Die Zeit, die benötigt wird, um die Kommunikation zu initiieren

##### ■ **Transferzeit** $t_w$ pro übertragenem Datenwort:

- hängt von der physikalischen Bandbreite des Kommunikationsmediums ab.

##### ■ Voraussetzung:

- Verbindungsnetz ist konfliktfrei, da sonst die Übertragungszeit nicht fest berechnet werden kann

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

- **Latenz** (Übertragungszeit einer Nachricht, latency)
  - **Software-Overhead**

Erzeugerprozess:

`send(proci, processi, @sbuffer, num_bytes)`

**Sender**

Systemaufruf  
Prüfe Schutzbed.  
DMA Init.

DMA nach NI

Verbraucherprozess:

`receive(@rbuffer, max_bytes)`

**Empfänger**

DMA vom Netzwerk in den Puffer  
BS Interrupt und Dekodierung der Nachricht  
BS kopiert Systempuffer in Userpuffer  
Reschedule Benutzerprozess  
Lesen der Nachricht

Zeit



NI: Netzwerkschnittstelle  
DMA: Direktspeicherzugriff

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

- **Latenz** (Übertragungszeit einer Nachricht, latency)
  - **Kanalverzögerung** (channel delay)
    - Dauer für die Belegung eines Kommunikationskanals durch eine Nachricht
    - Kanal:
      - Physikalische Verbindung zwischen Schalterelementen oder Knoten mit einem Puffer zum Halten der Daten während ihrer Übertragung
  - Verbindung (link)
    - Menge von Leitungen

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

- **Latenz** (Übertragungszeit einer Nachricht, latency)
  - **Schaltverzögerung, Routing-Verzögerung** (switching delay, routing delay)
    - Zeit, einen Weg zwischen zwei Knoten aufzubauen
    - Pfadberechnung oder Wegefindung (Routing)
      - die Art, wie der Weg einer Nachricht vom Sende- zum Zielknoten berechnet wird
      - Zu einer Verbindungsstruktur kann es mehrere Wegefindungsalgorithmen geben
      - einfache Implementierung in Verbindungselementen mit Hilfe eines schnellen Hardware-Algorithmus

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

- **Latenz** (Übertragungszeit einer Nachricht, latency)
  - **Blockierungszeit** (contention time)
    - Wird verursacht, wenn zu einem Zeitpunkt mehr als eine Nachricht auf eine Netzwerkressource zugreifen
  - **Blockierung** (contention)
    - Ein Verbindungsnetzwerk heißt blockierungsfrei, falls jede gewünschte Verbindung zwischen Prozessoren oder zwischen Prozessoren und Speichern unabhängig von schon bestehenden Verbindungen hergestellt werden kann

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

- **Durchsatz oder Übertragungsbandbreite** (bandwidth)
  - Maximale Übertragungsleistung des Verbindungsnetzwerkes oder einzelner Verbindungen, meist in Megabits pro Sekunde (MBit/s) oder Megabytes pro Sekunde (MB/s)
  
- **Bisektionsbandbreite** (bisection bandwidth)
  - Maximale Anzahl von Megabytes pro Sekunde, die das Netzwerk über die Bisektionslinie, die das Netzwerk in zwei gleiche Hälften teilt, transportieren kann

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

- **Diameter oder Durchmesser  $r$  (diameter):**
  - maximale Distanz für die Kommunikation zweier Prozessoren, also die Anzahl der Verbindungen, die durchlaufen werden müssen. Man spricht auch von der maximalen Pfadlänge zwischen zwei Knoten.
  
- **Verbindungsgrad eines Knotens  $P$  (node degree, connectivity)**
  - ist definiert als die Anzahl der direkten Verbindungen, die von einem Knoten zu anderen Knoten bestehen.
  
- **Mittlere Distanz  $d_a$  (average distance) zwischen zwei Knoten**
  - Anzahl der Links auf dem kürzesten Pfad zwischen zwei Knoten
  - $P/d_a$  ist die maximale Anzahl neuer Nachrichten, die von jedem Knoten in einem Zyklus in das Netzwerk eingebracht werden können

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Komplexität oder Kosten:

- Kosten für die Implementierung einer Hardware
- Aufwand für das Verbindungsnetz gemessen in der Anzahl und der Art der Schaltelemente und Verbindungsleitungen.

### ■ Erweiterbarkeit:

- Multiprozessoren können begrenzt, stufenlos oder nur durch Verdopplung der Anzahl der Prozessoren erweiterbar sein.

### ■ Skalierbarkeit:

- Fähigkeit, die wesentlichen Eigenschaften des Verbindungsnetzes auch bei beliebiger Erhöhung der Knotenzahl beizubehalten.

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

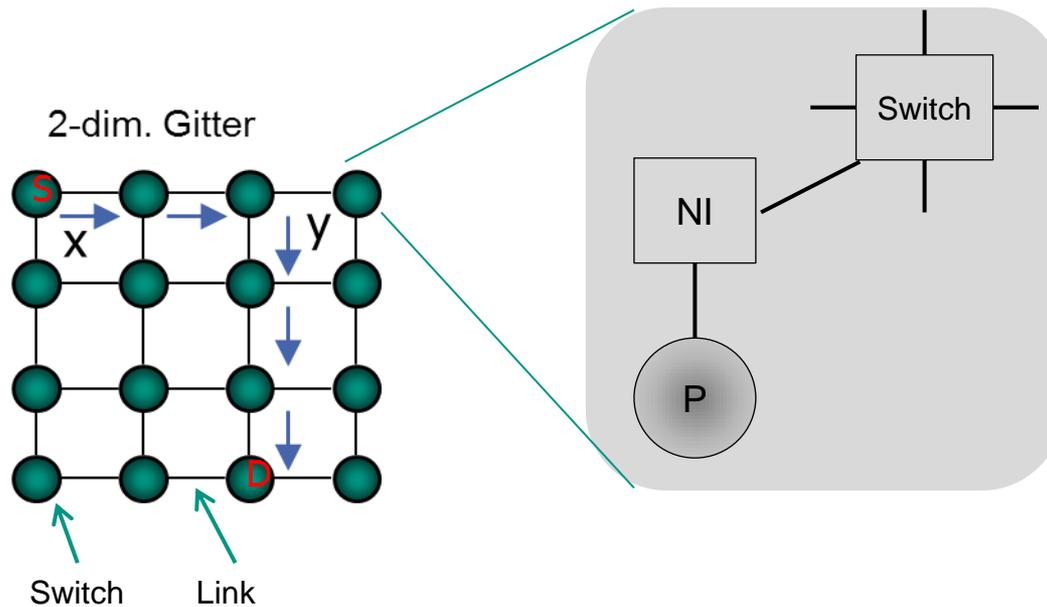
### ■ Ausfalltoleranz oder Redundanz :

- Verbindungen zwischen Knoten sind selbst dann noch zu schalten, wenn einzelne Elemente des Netzes (Schaltelemente, Leitungen) ausfallen.
- Ein fehlertolerantes Netz muss also zwischen jedem Paar von Knoten mindestens einen zweiten, redundanten Weg bereitstellen.  
Die Eigenschaft eines Systems, bei Ausfall einzelner Komponenten unter deren Umgehung funktionstüchtig zu bleiben, wenn auch mit verminderter Leistung, wird als Graceful degradation bezeichnet.

# Verbindungsstrukturen

## Verbindungsnetzwerke in einem Multiprozessorsystem

- Beispiel: 4 x 4 Gitter



# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Verbindungsnetzwerk (IN)

- Verbindet eine Anzahl von Knoten miteinander, so dass Informationen von einem Quellknoten (S) zu einem Zielknoten (D) verschickt werden können
- **Knoten:**
  - Beispiele: Cache-Modul, Speichermodul, Rechnerknoten, ...
  - Ist über eine Netzwerkschnittstelle (NI) mit dem Verbindungsnetzwerk verbunden
- **Schaltelement (Switch):**
  - Hat eine Menge von Ein- und Ausgängen
  - Setzt eine Verbindung zwischen einem Eingang und einem Ausgang auf, um eine Information zu übertragen, solange die Verbindung besteht
  - $n \times n$  Schaltelement:  $n$  Eingänge- und  $n$  Ausgänge mit Grad  $n$ , wobei die Ein- und Ausgänge zum Knoten nicht mitgezählt werden
  - Zur Vermeidung von Blockierung an einem Ausgang können Puffer bereitgestellt werden.

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Verbindungsnetzwerk

#### ■ Link:

- Verbindet einen Ausgang eines Schaltelements oder einer Netzwerkschnittstelle mit dem Eingang eines anderen Schaltelements oder Netzwerkschnittstelle
- Besteht aus Leitungen (auch aus optischen Materialien), über die digitale Informationen transportiert werden
  - Synchroner Übertragung: Links und Schaltelemente haben dieselbe Taktquelle, dessen Zykluszeit vom langsamsten Element bestimmt wird
  - Asynchroner Übertragung: die Komponenten haben unterschiedlichen Takt und die Synchronisation erfolgt über Handshake
- Die Breite ist die Anzahl der Bits, die parallel in einem Taktzyklus übertragen werden können
- Bandbreite:  $w/t$  Bits pro Zeiteinheit, mit Breite  $w$  und Taktzykluszeit  $t$

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Verbindungsnetzwerk

#### ■ Nachricht (message):

- Die Informationseinheit, die von der Quelle zum Ziel über das Verbindungsnetzwerk gesendet wird

#### ■ Länge der Nachricht:

- Variiert von einem Wort bis zu einer beliebigen Anzahl von Wörtern

#### ■ Arten der Nachrichtenübertragung

##### ■ Unicast:

- ein Knoten schickt eine Anforderungsnachricht an einen Zielknoten

##### ■ Multicast:

- Ein Knoten schickt eine Anforderungsnachricht an mehrere Knoten

##### ■ Broadcast:

- Ein Knoten schickt eine Anforderungsnachricht an alle Knoten
- Anforderungsnachrichten erwarten in vielen Fällen eine Antwortnachricht

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Verbindungsnetzwerk

#### ■ Nachricht (message):

##### ■ Pakete fester Länge

- Auftrennen einer Nachricht in eine Folge von Paketen fester Länge
- Auftrennen in Pakete erfolgt auf höheren Protokollebenen
- Erfolgt in Netzwerkschnittstelle mit Hilfe einer Kombination von Hardware- und Software-Mechanismen
- Pakete werden in einen Puffer der Netzwerkschnittstelle gespeichert und anschließend in das Verbindungsnetzwerk geschickt

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Verbindungsnetzwerk

#### ■ Aufbau eines Pakets:



#### ■ Header

- Enthält die Routing Information zum Bestimmen des Wegs vom Quell- zum Zielknoten über eine Menge von Schaltelementen
- Anforderungs- und Antworttyp (kann auch im **Trailer** stehen)

#### ■ Errorcode

- Pakete können verloren gehen aufgrund von Fehlern
- Fehlerbehandlung auf Protokollebene des IN (ECC für die Fehlererkennung oder Fehlerkorrektur von Übertragungsfehlern)

#### ■ Payload

- Zu übertragende Daten, irrelevant für die Verbindung

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Art des Datentransfers: Switching strategy

#### ■ Durchschalte- oder Leitungsvermittlung (circuit switching):

- Eigenschaft eines Netzes eine direkte Verbindung zwischen zwei oder mehreren Knoten eines Netzes zu schalten.

Die physikalische Verbindung bleibt für die gesamte Dauer der Informationsübertragung bestehen.

- Paket wird ohne Unterbrechung vom Sender zum Empfänger übertragen
- Paket muss keine Routing-Information mitführen, da der Weg aufgebaut wird, bevor das Paket versendet wird
- Vermeidet Routing-Overhead in jedem Schaltelement
- Alle Netzwerkressourcen auf dem Kommunikationspfad sind nicht für andere Pakete verfügbar, bis das gesamte Paket am Ziel angekommen ist

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Art des Datentransfers: Switching strategy

#### ■ Paketvermittlung (packet switching):

- Datenpakete fester Länge oder Nachrichten variabler Länge werden entsprechend einem Wegefindungsalgorithmus (routing) vom Absender zum Empfänger geschickt
- Nachrichten mit Adresse und Daten werden durch das Netzwerk verschickt
- Adresse wird in jedem Knoten gelesen und die Nachricht wird zum nächsten Knoten weitergeleitet, bis die Nachricht das Ziel erreicht
- Die Ressourcen eines Schaltelements werden nur solange belegt wie sie benötigt werden
- Es kann mehrere Paketübertragungen geben, wobei deren Wege gemeinsame Link- oder Schaltelement-Ressourcen haben können
  - Falls ein Schaltelement belegt ist und der Weg eines anderen Pakets führt zu diesem Schaltelement, tritt ein Konflikt auf
  - Flusskontrolle (flow control)

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Wegefindung: Routing Algorithmus

- Verfahren zur Wegebestimmung
  - Hängt von der Topologie ab
  - Beispiel: 4 x 4 Gitter
    - Das Paket wird vom Sender zunächst in x-Richtung (horizontal) und dann in y-Richtung (vertikal) übertragen

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Latenz- und Bandbreitenmodelle

- Bewertung der Entwurfsalternativen für Verbindungsnetzwerke bezüglich der Kommunikationsleistung

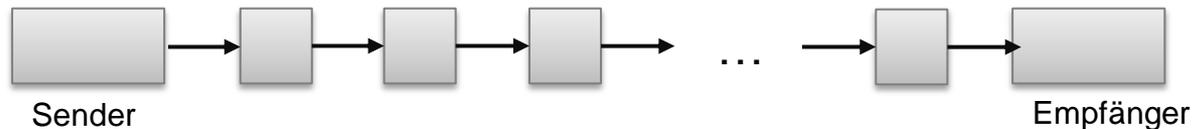
### ■ End-to-end packet latency model

- Betrachtet die Übertragung eines Pakets vom Sendeknoten zu einem Empfängerknoten
- Annahme: das Paket ist bereit zur Übertragung in einem Puffer an der Quelle
- **End-to-end latency:**
  - Zeit, die benötigt wird von diesem Zeitpunkt an bis das gesamte Paket über das Verbindungsnetzwerk übertragen worden ist und in einem Puffer am Empfängerknoten abgelegt ist
  - $N_P$ : Anzahl der Bits im Payload des Pakets
  - $N_E$ : Anzahl der Bits für Verwaltungsinformationen im Header, Errorcode und Trailer des Pakets
- $N_P + N_E$ : Anzahl der zu übertragenden Bits

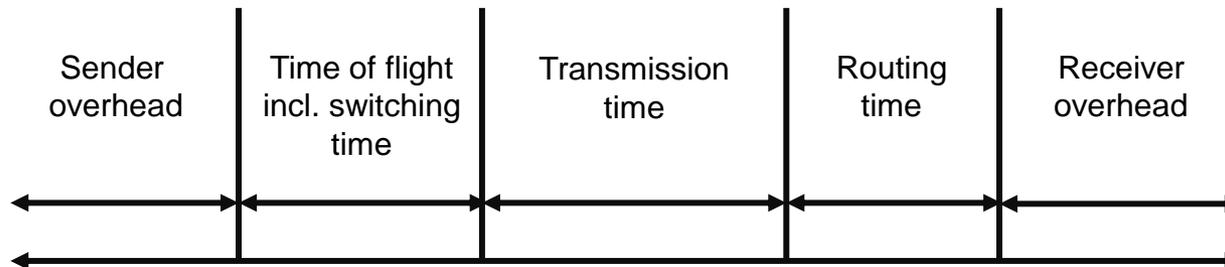
# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

- End-to-end packet latency Modell
  - Übertragungspipeline



- End-to-end latency:



*End-to-end packet latency = Sender OV + Time of flight + Transmission time + Routing time + Receiver OV*

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

- **End-to-end packet latency model**
  - **Sender overhead:**
    - Zusammenstellen des Pakets (Header, Errorcode, Trailer) und Ablegen in Sendepuffer der NI
  - **Time of flight**
    - Zeit, um ein Bit von der Quelle zum Ziel zu senden, wenn der Weg festgelegt ist (ohne Konflikte)
  - **Transmission time (Übertragungszeit)**
    - Zusätzliche Zeit, die benötigt wird, alle Bits eines Pakets zu übertragen, nachdem das erste Bit beim Empfänger angekommen ist
    - Hängt von der Linkbandbreite ab
    - Phit (physical transfer unit): Informationseinheit, die in einem Zyklus auf einem Link übertragen wird
    - Für ein Paket:
      - $(N_P + N_E) / (N_{Ph} \times f)$
      - mit  $N_{Ph}$ : Phit-Größe

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

- **End-to-end packet latency model**
  - **Routing time:**
    - Zeit, um den Weg aufzusetzen, bevor ein Teil des Pakets übertragen werden kann
  - **Switching time**
    - Hängt von der Switching-Strategie ab
  - **Receiver overhead**
    - Ablegen der Verwaltungsinformation und Weiterleiten des Pakets aus dem Empfangspuffer

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Latenz- und Bandbreitenmodelle

- Bewertung der Entwurfsalternativen für Verbindungsnetzwerke bezüglich der Kommunikationsleistung

### ■ Effektive Bandbreite

- effektive Bandbreite = 
$$\frac{\text{Paketgröße}}{\max(\text{Sender OH}, \text{Empfänger OH}, \text{Übertragungszeit})}$$

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Switching strategy

- Bestimmt, wie ein Weg in einem Verbindungsnetzwerk aufgebaut und ein Paket vom der Quelle zum Ziel übertragen wird
- Modellannahmen:
  - Modellierung des Pfads zwischen Quelle und Ziel als Übertragungspipeline
  - Ein Paket hat auf dem Weg von der Quelle zum Ziel  $L$  Schaltelemente und ein Paket umfasst  $N$  Phits

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Switching strategy

#### ■ Modellierung Circuit switching

- Aufbau des Weges zwischen Quelle und Ziel, anschließend wird das Paket übertragen (Pipelining)
- **Routing time**
  - ist die Zeit, die notwendig ist, ein einzelnes Paket von der Quelle zum Ziel zu senden und wieder zurück, um die Quelle zu informieren, dass der Weg aufgebaut ist
  - Die Routing- Entscheidung in einem Schaltelement benötigt  $R$  Netzwerkzyklen
  - $routing\ time = L \times R + time\ of\ flight = L \times R + L = L(R + 1)$
- **End-to-end packet latency**
  - $end - to - end\ packet\ latency =$   
 sender OH + time of flight + transmission time + routing time +  
 receiver overhead = sender OH +  $L$  +  $N$  +  $L(R + 1)$  + receiver OH =  
 sender OH +  $L(R + 2)$  +  $N$  + receiver OH

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Paketvermittlung (packet switching):

- Datenpakete fester Länge oder Nachrichten variabler Länge werden entsprechend einem Wegefindungsalgorithmus (routing) vom Absender zum Empfänger geschickt
- Der Weg wird festgelegt, so wie das Paket von der Quelle zum Ziel fortschreitet
- Die Wegeentscheidung wird in jedem Schaltelement getroffen, sobald das Paket es erreicht
- Die Ressourcen eines Schaltelements werden nur solange belegt wie sie benötigt werden

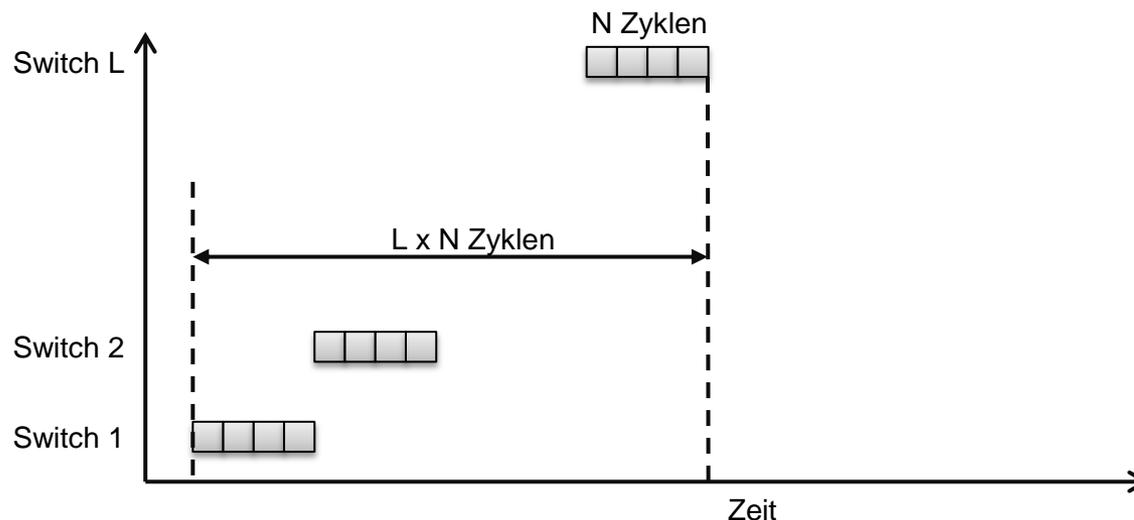
# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Paketvermittlung (packet switching):

#### ■ Store-and-forward-Modus

- Jeder Knoten enthält einen Puffer zum Aufnehmen der vollständigen Nachricht
- Für ein Paket wird ein Pfad bestimmt, wenn es ein Schaltelement erreicht
- Alle Teile eines Pakets müssen bei einem Knoten angekommen sein, bevor ein Teil des Pakets weitergeleitet wird



# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Paketvermittlung (packet switching):

#### ■ Modellierung Store-and-forward

- Time of flight:
  - umfasst die Zeit für die Übertragung eines Bits von der Quelle zum Ziel unter der Annahme, dass es keine Routing OH gibt
  - Hängt von der Paketgröße ab:
    - Jedes Paket verbringt N Zyklen in einem Schaltelement
    - Es dauert  $L \times N$  Zyklen, bis das erste Bit am Ziel angekommen ist
- Übertragungszeit: N Zyklen
- Routing Time:  $L \times R$  Zyklen (mit R Routing OH in jedem Schaltelement)
  
- **End-to-end packet latency:**
- *end – to – end packet latency = sender OH +  $L \times N$  + N +  $L \times R$  + receiver OH = sender OH +  $N(L + 1)$  +  $L \times R$  + receiver OH*

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Paketvermittlung (packet switching):

#### ■ Cut-through switching-Modus

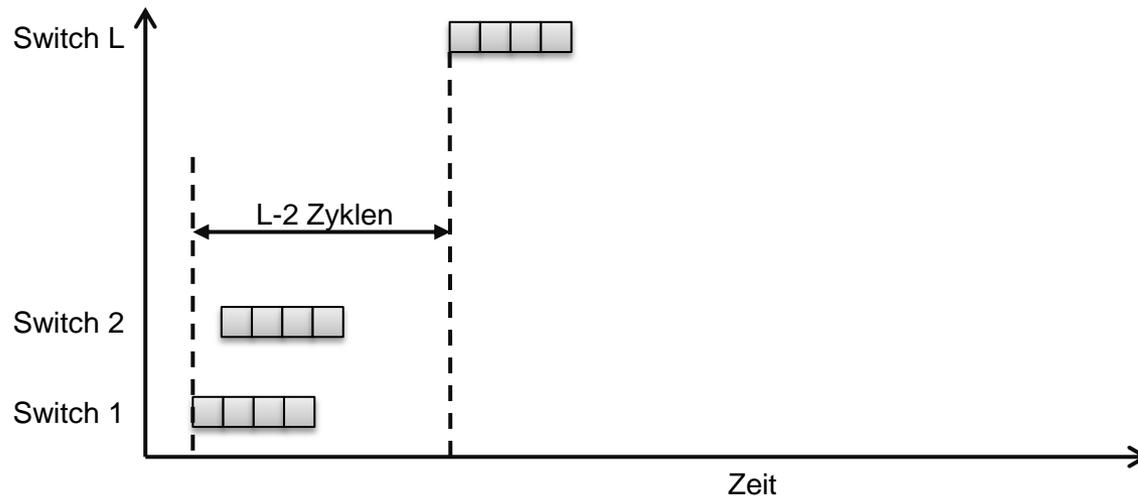
- Der Kopfteil der Nachricht enthält die Empfängeradresse und wird in jedem Schaltelement bei der Ankunft des Pakets dekodiert, so dass der einzuschlagende Weg bestimmt werden kann, was zu einem Routing OH von R Zyklen führt
- Solange keine Blockierung auftritt, wird das Paket durch die Schaltelemente gemäß der Pipeline-Verarbeitung von der Quelle zum Ziel übertragen
- Kopf-Information wird festgehalten bis letztes Phit angekommen ist.
- Nachdem das gesamte Paket von einem Schaltelement weitergeleitet worden ist, wird es freigegeben
- Bei einer Blockierung wird das gesamte Paket aufgehalten
- Flow control unit (Flit) ist der Teil eines Pakets, der bei einer Blockierung aufgehalten wird, bei cut-through ist das gesamte Paket ein Flit

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Paketvermittlung (packet switching):

#### ■ Modellierung Cut-through



- *end – to – end packet latency = sender OH + L + N + L × R + receiver OH = sender OH + L × (R + 1) + N + receiver OH*

# Verbindungsstrukturen

## Charakterisierung von Verbindungsnetzwerken

### ■ Art des Datentransfers:

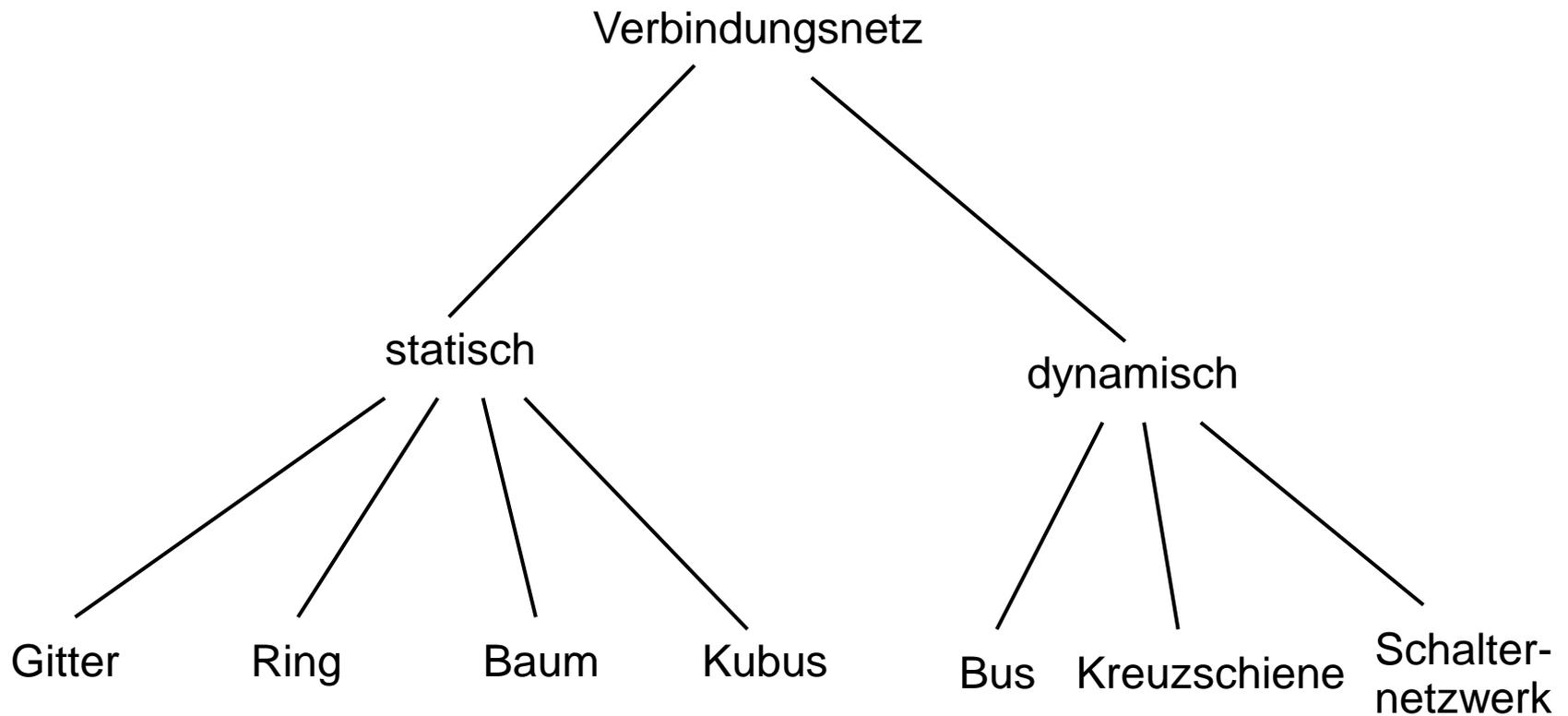
#### ■ Paketvermittlung (packet switching):

#### ■ Übertragungsmodi: **Wormhole-routing-Modus:**

- solange keine Übertragungskanäle blockiert sind, mit dem Cut-through-Modus identisch.
- Falls der Kopfteil der Nachricht auf einen Kanal trifft, der gerade belegt ist, wird er abgeblockt. Alle nachfolgenden Übertragungseinheiten der Nachricht verharren dann ebenfalls an ihrer augenblicklichen Position, bis die Blockierung aufgehoben ist. Durch das Verharren werden die Puffer nachfolgender Kanäle auch für weitere Nachrichten blockiert.
- Es werden nur die Phits festgehalten, die die Wegeinformation enthalten, weshalb der Flit keiner als ein Paket ist

# Verbindungsstrukturen

## Topologie: Klassifizierung



# Verbindungsstrukturen

## Topologie:

### ■ Statische Verbindungsnetze

- Nach Aufbau des Verbindungsnetzes bleiben die Verbindungen fest
- Gute Leistung für Probleme mit vorhersagbaren Kommunikationsmustern zwischen benachbarten Knoten

# Verbindungsstrukturen

## Topologie:

### ■ Statische Verbindungsnetze

#### ■ Vollständige Verbindung

- Jeder Knoten ist mit jedem anderen Knoten verbunden
- Höchste Leistungsfähigkeit
  - Arbeitet für alle Anwendungen mit allen Arten von Kommunikationsmustern effizient
- Aber: nicht praktikabel in Parallelrechnern
  - Netzwerkkosten steigen quadratisch mit der Anzahl der Prozessoren

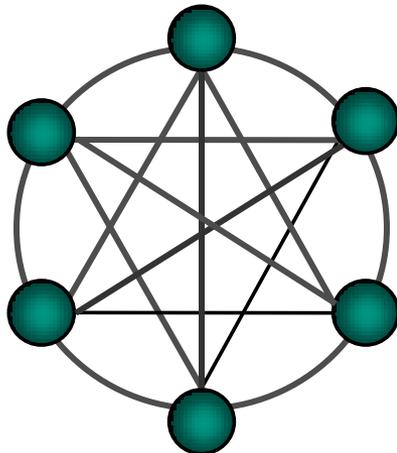
# Verbindungsstrukturen

## Topologie:

### ■ Statische Verbindungsnetze

### ■ Vollständige Verbindung

- Jeder Knoten ist mit jedem anderen Knoten verbunden
- Höchste Leistungsfähigkeit
  - Arbeitet für alle Anwendungen mit allen Arten von Kommunikationsmustern effizient
- Aber: nicht praktikabel in Parallelrechnern
  - Netzwerkkosten steigen quadratisch mit der Anzahl der Prozessoren



# Verbindungsstrukturen

## Topologie:

### ■ Statische Verbindungsnetze

### ■ Gitterstrukturen

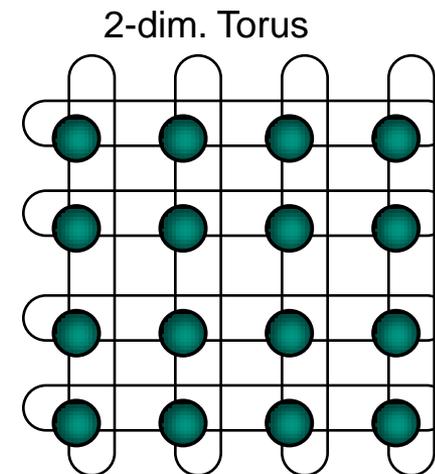
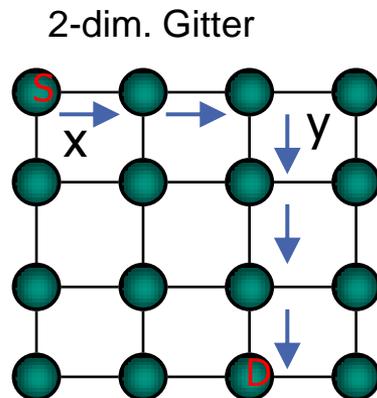
#### ■ 1-dimensionales Gitter (**lineares Feld, Kette**)

- Verbindet  $N$  Knoten mit  $(N-1)$  Verbindungen
- Endknoten haben den Grad 1, Zwischenknoten den Grad 2 und sind mit benachbarten Knoten verbunden
- Diameter  $r$  ist  $N-1$
- Disjunkte Bereiche des linearen Netzwerkes können gleichzeitig genutzt werden, aber es sind mehrere Schritte notwendig, um eine Nachricht zwischen zwei nicht benachbarte Knoten zu verschicken



# Verbindungsstrukturen

- Topologie:
- Statische Verbindungsnetze
- Gitterstrukturen
  - k-dimensionales Gitter mit N Knoten



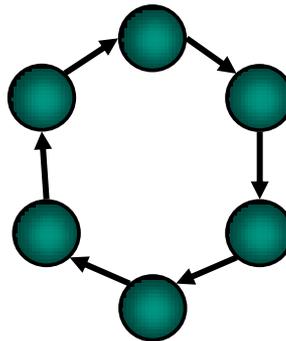
# Verbindungsstrukturen

## Topologie:

### ■ Statische Verbindungsnetze

### ■ Ring

- Erhält man, wenn man die Endknoten eines linearen Feldes miteinander verbindet
- Unidirektionaler Ring mit N Knoten
  - Nachrichten werden in einer Richtung vom Quellknoten zum Zielknoten verschickt
  - Diameter  $r$  ist  $N-1$
  - Bei Ausfall einer Verbindung bricht die Kommunikation zusammen



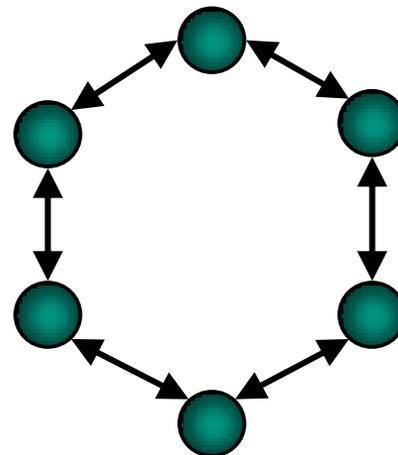
# Verbindungsstrukturen

## Topologie:

### ■ Statische Verbindungsnetze

### ■ Bidirektionaler Ring mit N Knoten

- symmetrisches Netzwerk
- Der längste Pfad, den eine Nachricht nehmen muss, ist nicht länger als  $N/2$
- Bei Ausfall einer Verbindung bricht die Kommunikation noch nicht zusammen, während zwei Ausfälle von Verbindungen das Netzwerk in zwei disjunkte Teilnetzwerke aufteilen



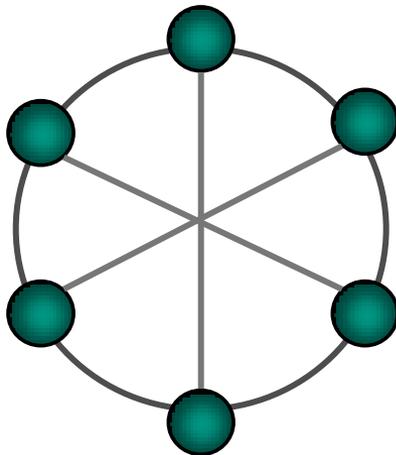
# Verbindungsstrukturen

## Topologie:

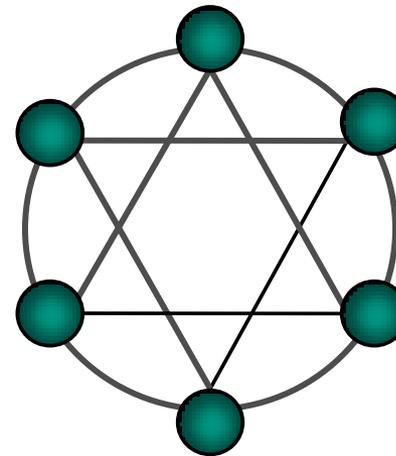
- **Statische Verbindungsnetze**

- **Chordaler Ring**

- Hinzufügen redundanter Verbindungen
  - erhöht Fehlertoleranzeigenschaft des Verbindungsnetzwerkes
  - Höherer Knotengrad und kleinerer Diameter gegenüber Ring



Chordaler Ring mit Knotengrad 3



Chordaler Ring mit Knotengrad 4

# Verbindungsstrukturen

## Topologie:

### ■ Statische Verbindungsnetze

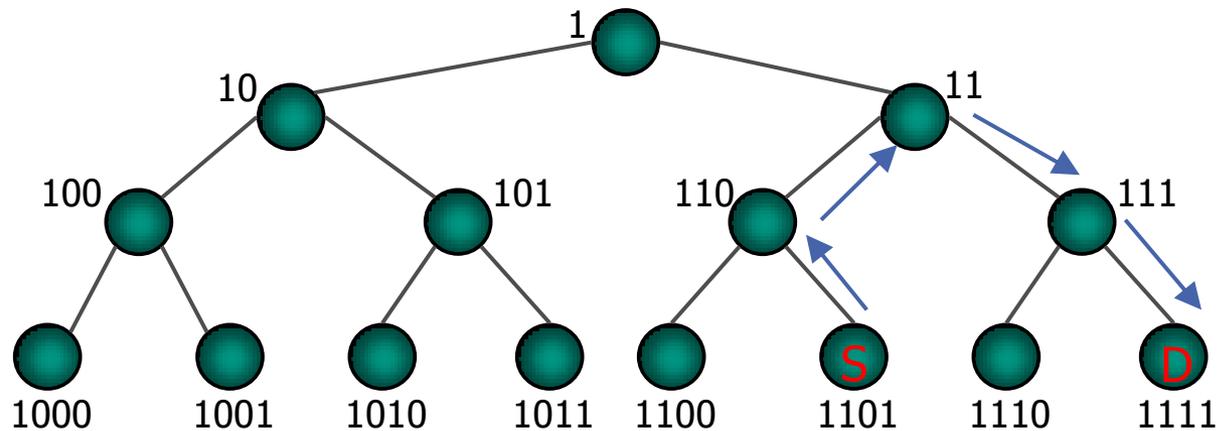
### ■ Baumstrukturen

- Binärer Baum mit  $m$ -Ebenen:
- Auf Ebene  $m$ :  $N=2^m-1$  Knoten
- Diameter:  $2(m-1)$
- Adressierung der Knoten:
  - Die Knotennummer auf Ebene  $m$  besteht aus  $m$  Bits
  - Der Wurzelknoten hat die Nummer 1
  - Die Nummer des linken Kindknotens erhält man durch Hinzufügen einer 0 an die niederwertige Stelle der Adresse des Elternknoten
  - Die Nummer des rechten Kindknotens erhält man durch Hinzufügen einer 1 an die niederwertige Stelle der Adresse des Elternknoten

# Verbindungsstrukturen

## Topologie:

- Statische Verbindungsnetze
- Baumstrukturen: Routing
  - Finde gemeinsamen Elternknoten P von S und D
  - Gehe von S nach P und von P nach D



# Verbindungsstrukturen

## Topologie:

- Statische Verbindungsnetze
- Baumstrukturen: Routing

- Die Binärdarstellung der Adresse eines Quellknotens  $S$  auf Ebene  $i$  sei  $S_i S_{i-1} \dots S_1$  und die der Adresse des Zielknotens  $D$  auf Ebene  $j$  sei  $D_j D_{j-1} \dots D_1$
- Finde die gemeinsamen höchstwertigen Bits von  $S$  und  $D$ , so dass die Adresse des Elternknotens  $P$  gleich  $D_j D_{j-1} \dots D_x = S_i S_{i-1} \dots S_{(i-j+x)}$  ist
- Steige von  $S$  ( $i-j+x$ ) Ebenen auf nach  $P$
- for  $k=x-1$  step 1 until 0
  - {steige nach links ab, falls  $D_x=0$
  - steige nach rechts ab, falls  $D_x=1$ }

# Verbindungsstrukturen

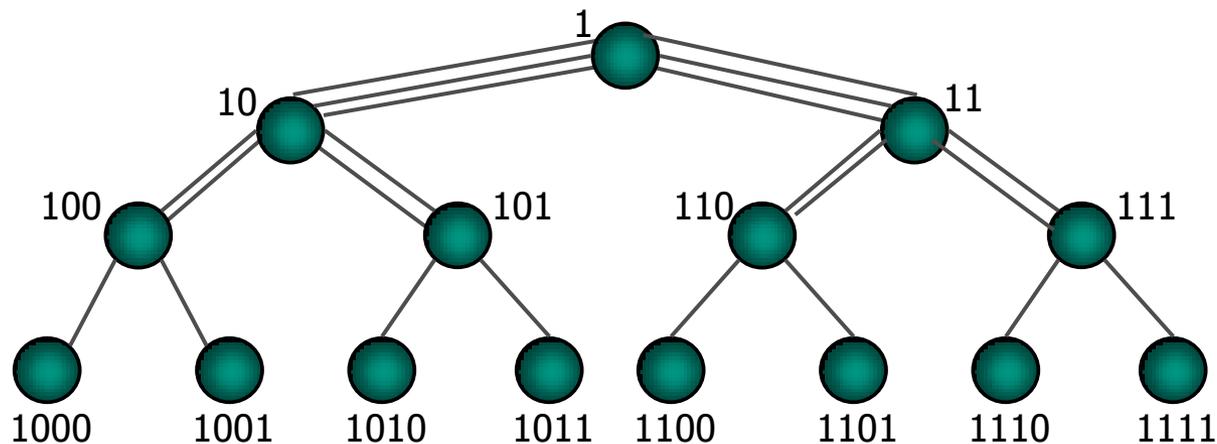
## Topologie:

- Statische Verbindungsnetze

- Baumstrukturen

- Fat Tree:

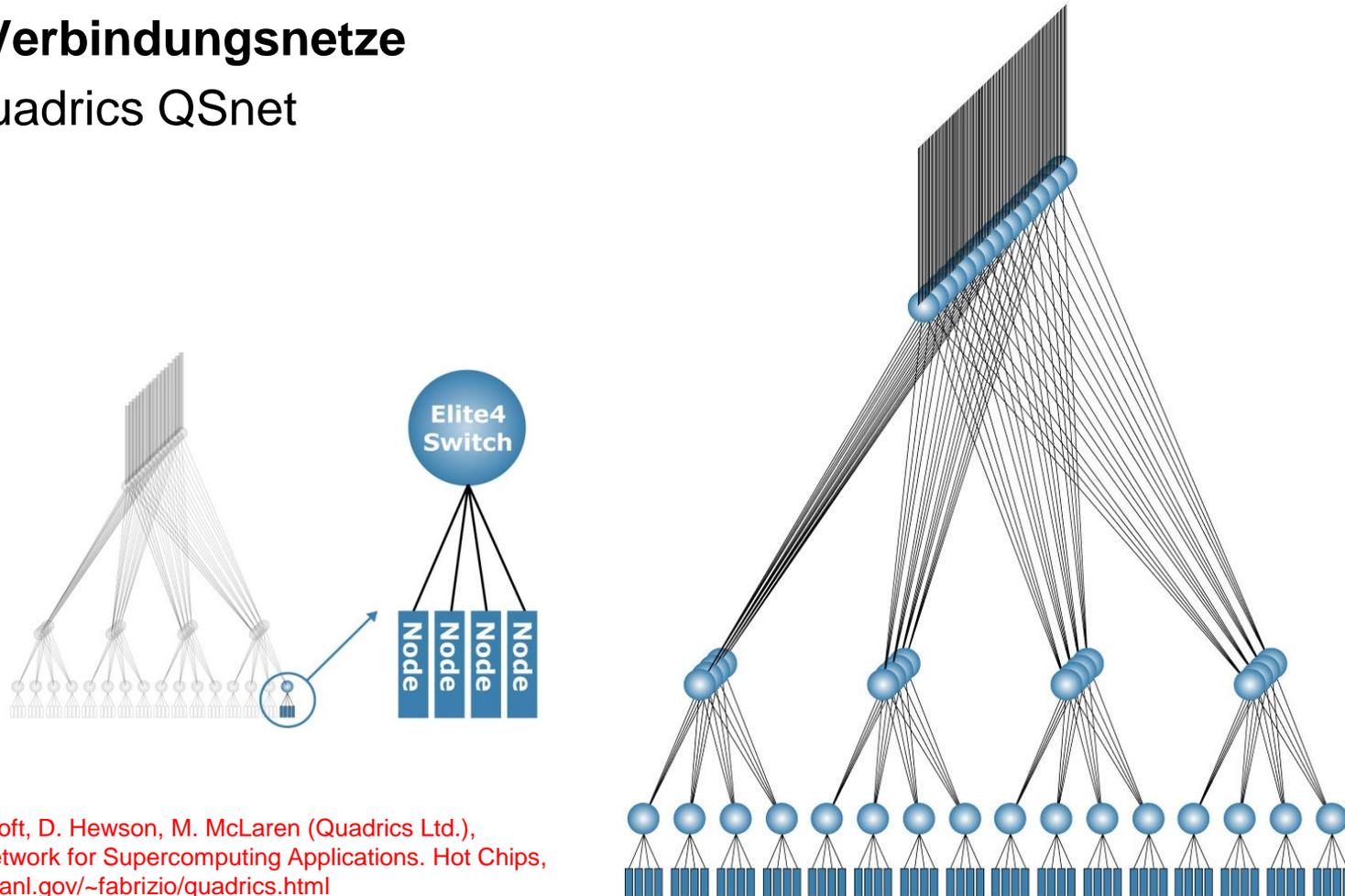
- Lösung des Blockierungsproblems in Richtung Wurzel
- Kommunikationskanäle werden größer, je näher man sich der Wurzel nähert



# Verbindungsstrukturen

## Topologie:

- Statische Verbindungsnetze
- Beispiel: Quadrics QSnet



Quelle: D. Addison, J. Beecroft, D. Hewson, M. McLaren (Quadrics Ltd.),  
Fabrizio Petrini (LANL): A network for Supercomputing Applications. Hot Chips,  
August 2003, <http://www.c3.lanl.gov/~fabrizio/quadrics.html>

# Verbindungsstrukturen

## Topologie:

- **Statische Verbindungsnetze**
- Beispiel: Quadrics QSnet

Elan 4 network interface card:



Elite 4 Switch Component:



QsNet II Switch:

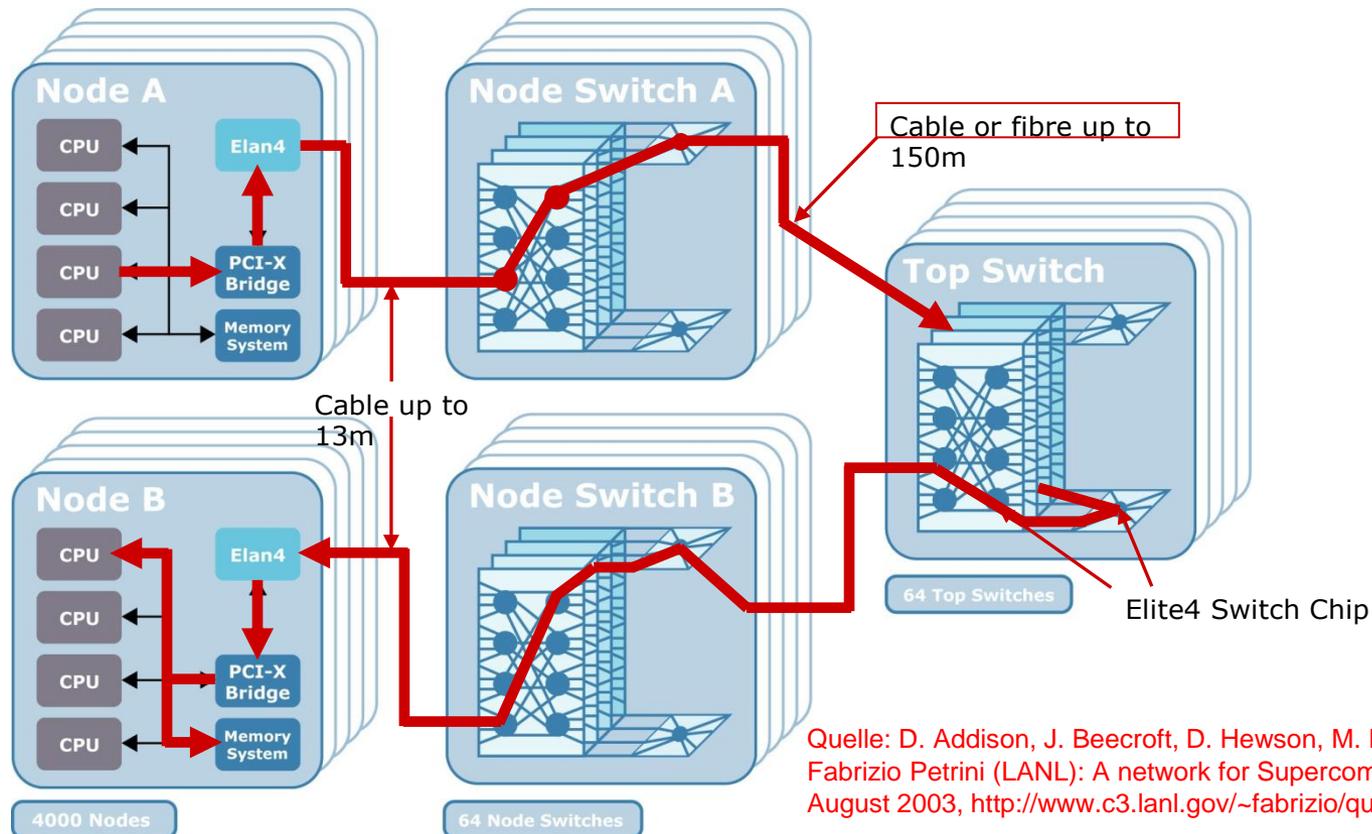


Quelle: D. Addison, J. Beecroft, D. Hewson, M. McLaren (Quadrics Ltd.),  
Fabrizio Petrini (LANL): A network for Supercomputing Applications. Hot Chips,  
August 2003, <http://www.c3.lanl.gov/~fabrizio/quadrics.html>

# Verbindungsstrukturen

## Topologie:

- Statische Verbindungsnetze
- Beispiel: Quadrics QSnet

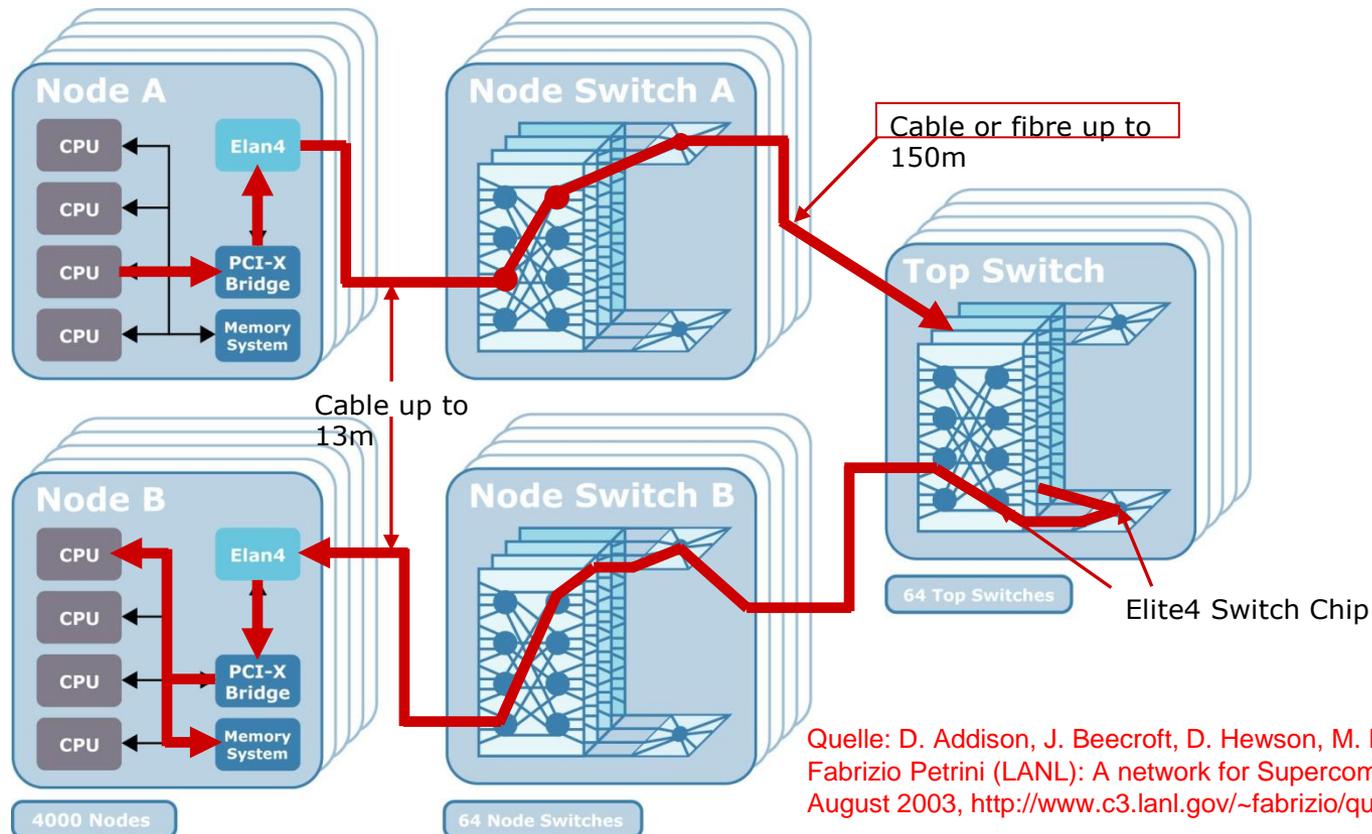


Quelle: D. Addison, J. Beecroft, D. Hewson, M. McLaren (Quadrics Ltd.), Fabrizio Petrini (LANL): A network for Supercomputing Applications. Hot Chips, August 2003, <http://www.c3.lanl.gov/~fabrizio/quadrics.html>

# Verbindungsstrukturen

## Topologie:

- Statische Verbindungsnetze
- Beispiel: Quadrics QSnet



Quelle: D. Addison, J. Beecroft, D. Hewson, M. McLaren (Quadrics Ltd.), Fabrizio Petrini (LANL): A network for Supercomputing Applications. Hot Chips, August 2003, <http://www.c3.lanl.gov/~fabrizio/quadrics.html>

# Verbindungsstrukturen

## Topologie:

### ■ Statische Verbindungsnetze

### ■ K-ärer n-Kubus (Cubes, Würfel)

- Allgemeine Form eines Kubus-Verbindungsnetzwerkes
- Ringe, Gitter, oder Hyperkubi sind topologisch isomorph zu einer Familie von K-ären n-Kubus Netzwerken
  - n ist die Dimension
  - Radius K ist die Anzahl der Knoten, die einen Zyklus in einer Dimension bilden
- Enthält  $N=K^n$  Knoten
- Die Knoten werden über eine n-stellige k-äre Zahl der Form  $a_0, a_1, \dots, a_{n-1}$  adressiert
  - Jede Stelle  $0 \leq a_i < K$  stellt die Position des Knotens in der entsprechenden i-ten Dimension dar, mit  $0 \leq i \leq n-1$
  - Ein Nachbarknoten in der i-ten Dimension zu einem Knoten mit Adresse  $a_0, a_1, \dots, a_{n-1}$  kann erreicht werden mit  $a_0, a_1, \dots, a_{(i \pm 1)} \bmod k \dots a_{n-1}$ .
- Knotengrad ist  $2n$  und der Diameter ist

# Verbindungsstrukturen

## Topologie:

### ■ Statische Verbindungsnetze

### ■ Hyperkubus (Hypercubes)

#### ■ Verallgemeinerter Würfel:

- die  $N = 2^n$  Prozessoren sind Ecken eines  $n$ -dimensionalen Würfels, wobei die Verbindungen dann die Kanten des Würfels darstellen.

#### ■ Komplexität ist $(N \cdot \log_2 N) / 2$ .

#### ■ Diameter beträgt $\log_2 N$ .

#### ■ Lange Zeit häufigste Verbindungsstruktur bei den nachrichtengekoppelten Multiprozessoren, aber:

#### ■ Skalierbarkeit:

- Jede Erweiterung benötigt mindestens die Verdopplung der Prozessorenanzahl.
- Aus räumlichen Anordnungsgründen begrenzt.

# Verbindungsstrukturen

## Topologie:

- **Statische Verbindungsnetze**

- **Hyperkubus**

- **e-Cube Routing**

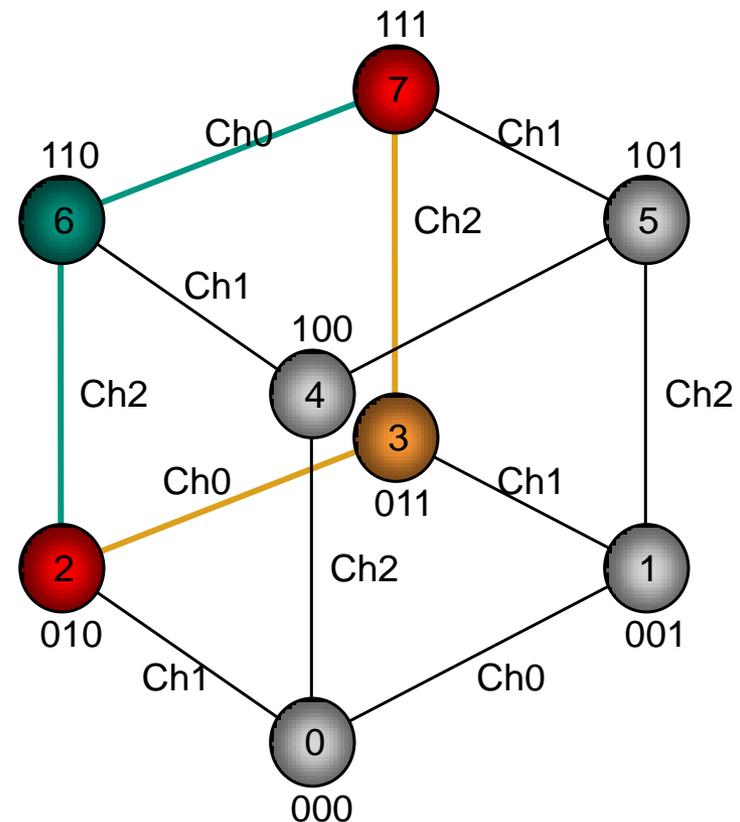
- Die Knotennummern werden als Binärzahlen geschrieben, dadurch unterscheiden sich benachbarte Knoten in genau einer Stelle, die zudem die Richtung der Verbindung angeben kann (Hamming Distanz)
    - Eine einfache Wegewahl:  
die Bits in Start- und Zieladresse werden mittels einer XOR-Verbindung verknüpft und das Resultat bestimmt die möglichen Wege.

# Verbindungsstrukturen

## Topologie:

- Statische Verbindungsnetze
- Hyperkubus
  - Beispiel:

- $A = (010)$  und  $B = (111)$
- $W = A \text{ XOR } B = 101$
- $(010) \rightarrow (011) \rightarrow (111)$ ,
- $(010) \rightarrow (110) \rightarrow (111)$



# Verbindungsstrukturen

## Topologie:

### ■ Dynamische Verbindungsnetzwerke:

- Geeignet für Anwendungen mit variablen und nicht regulären Kommunikationsmustern

### ■ Bus:

- Wird von den am Bus angeschlossenen Prozessoren gemeinsam benützt
- Ein Datentransport zu einem Zeitpunkt
- Nachricht von einer Quelle zu jedem Ziel in einem Schritt
- Busbandbreite =  $w * f$ 
  - $w$ : Anzahl der Datenleitungen (Busbreite)
  - $F$ : Frequenz
  - Bestimmt maximale Anzahl der Prozessoren, d. h. die Bandbreite muss mit dem Produkt der Anzahl der Prozessoren und ihrer Geschwindigkeit abgestimmt werden

# Verbindungsstrukturen

## Topologie:

### ■ Dynamische Verbindungsnetzwerke:

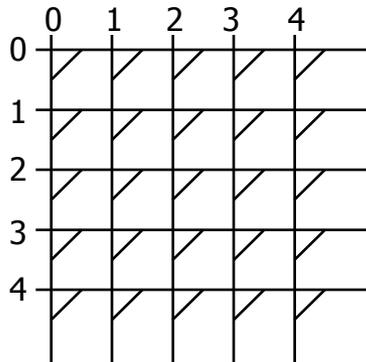
### ■ Bus:

- Reduzierung des Busverkehrs
  - Verwendung von Cache-Speichern mit Cache-Kohärenz-Protokollen
- Verwendung von sog. Split-Phase Busprotokollen
  - Das Protokoll gibt den Bus nach der Übertragung einer Speichereferenzanforderung wieder frei
  - Wenn der Speicher bereit ist, das Datum zu liefern, fordert dieser den Bus an und schickt die Daten als Antwort
  - Ermöglicht, dass andere Prozessoren in der Zwischenzeit den Bus anfordern können, vorausgesetzt, dass ein verschränkter Speicher vorliegt oder Pipelining möglich ist

# Verbindungsstrukturen

## Topologie:

- **Dynamische Verbindungsnetzwerke:**
- **Kreuzschienenverteiler (Crossbar)**
  - Vollständig vernetztes Verbindungswerk mit allen möglichen Permutationen der N Einheiten, die über das Netzwerk verbunden werden



# Verbindungsstrukturen

## Topologie:

### ■ Dynamische Verbindungsnetzwerke:

### ■ Kreuzschienenverteiler (Crossbar)

- Hardware-Einrichtung, die so geschaltet werden kann, dass in einer Menge von Prozessoren alle möglichen disjunkten Paare von Prozessoren gleichzeitig und blockierungsfrei miteinander kommunizieren können.
  - In Abhängigkeit vom Zustand der Schaltelemente im Kreuzschienenverteiler können dann je zwei beliebige Elemente aus den verschiedenen Mengen miteinander kommunizieren.
  - Alle  $N!$  Permutationen sind möglich
  - An den Kreuzungspunkten sitzen Schaltelemente: hoher Hardware-Aufwand
  - Kosten:  $N^2$  Schaltelemente (bei  $N$  Knoten pro Dimension)
  - Ein Schaltelement entspricht einem Paar von Quelle und Ziel, so dass die Darstellung einer Permutation als eine Liste solcher Paare direkt zu der korrekten Schaltung der Schalterelemente führt.

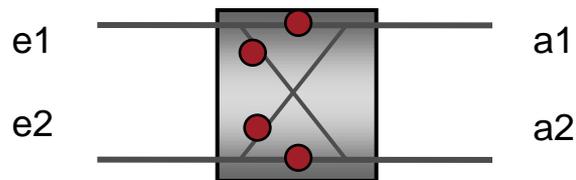
# Verbindungsstrukturen

## Topologie:

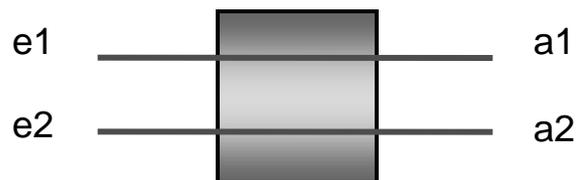
### ■ Dynamische Verbindungsnetzwerke:

### ■ Schalterelemente (2x2 Kreuzschienenverteiler)

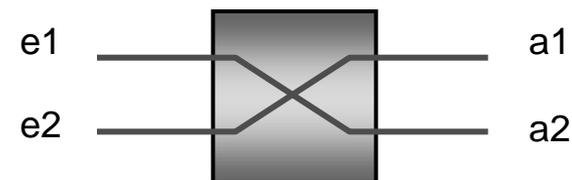
- bestehen aus Zweierschaltern mit zwei Eingängen und zwei Ausgängen, die entweder durchschalten oder die Ein- und Ausgänge überkreuzen können



● Kontakt, der geöffnet oder geschlossen werden kann



Durchschalten



Vertauschen

# Verbindungsstrukturen

## Topologie:

- **Dynamische Verbindungsnetzwerke:**
- **Mehrstufige Verbindungsnetzwerke (Schalernetzwerke, Permutationsnetzwerke)**
  - Kompromiss zwischen der niedrigeren Leistungsfähigkeit von Bussen und hohem Hardware-Aufwand von Kreuzschienenverteilern
  - Oft 2 x 2 Kreuzschienenverteiler (Schalterelement) als Grundelement

# Verbindungsstrukturen

## Topologie:

### ■ Dynamische Verbindungsnetzwerke:

### ■ Permutationsnetze

- $p$  Eingänge des Netzes können gleichzeitig auf  $p$  Ausgänge geschaltet werden und somit wird eine Permutation der Eingänge erzeugt.

### ■ Einstufige Permutationsnetze

- enthalten eine einzelne Spalte von Zweierschaltern,

### ■ Mehrstufige Permutationsnetze

- enthalten mehrere solcher Spalten
- Spalten: Stufen des Permutationsnetzwerkes

# Verbindungsstrukturen

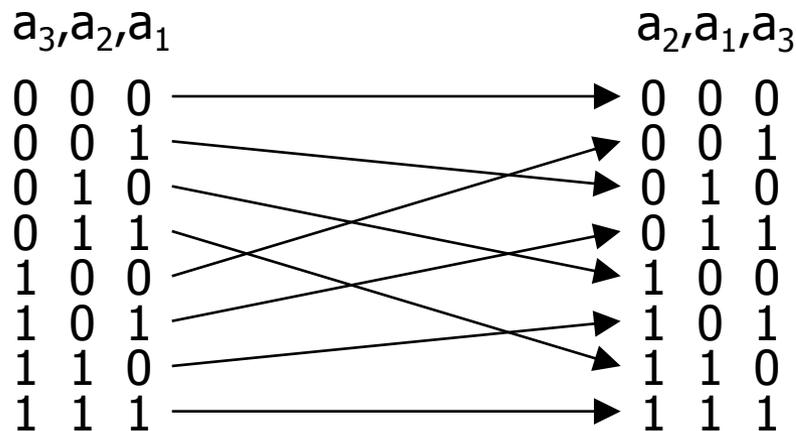
## Topologie:

- **Dynamische Verbindungsnetzwerke:**
- **Permutationsnetze**
- **reguläre Permutationsnetzwerke:**
  - $p$  Eingänge,  $p$  Ausgänge und  $k$  Stufen mit jeweils  $p/2$  Zweierschaltern, wobei die Zahl  $p$  normalerweise eine Zweierpotenz ist.
- **Irreguläre Permutationsnetzwerke**
  - weisen gegenüber der vollen regulären Struktur Lücken auf

# Verbindungsstrukturen

## Topologie:

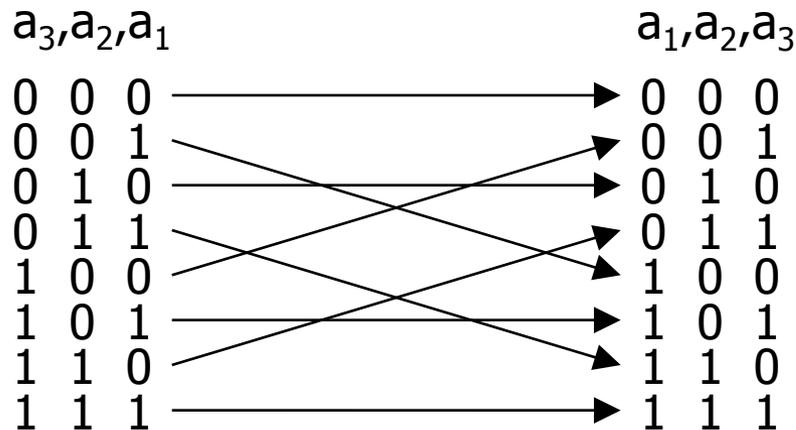
- Dynamische Verbindungsnetzwerke:
- Permutationen
  - Mischpermutation M (Perfect Shuffle):
    - $M(a_n, a_{n-1}, \dots, a_2, a_1) = (a_{n-1}, \dots, a_2, a_1, a_n)$



# Verbindungsstrukturen

## Topologie:

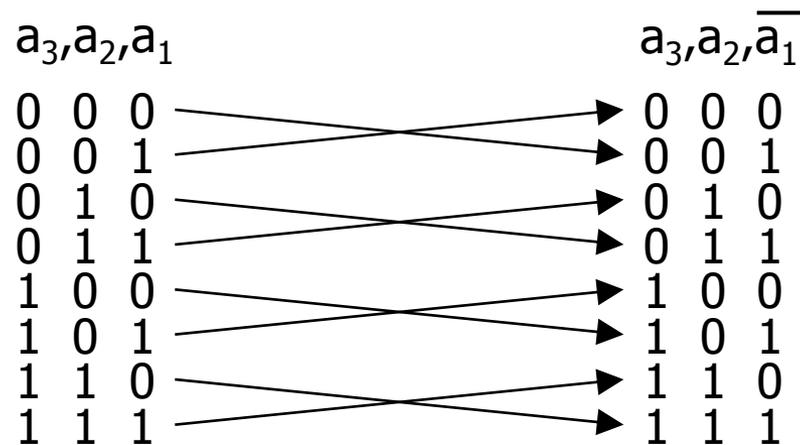
- Dynamische Verbindungsnetzwerke:
- Permutationen
  - Kreuzpermutation K (Butterfly):
    - $K(a_n, a_{n-1}, \dots, a_2, a_1) = (a_1, a_{n-1}, \dots, a_2, a_n)$



# Verbindungsstrukturen

## Topologie:

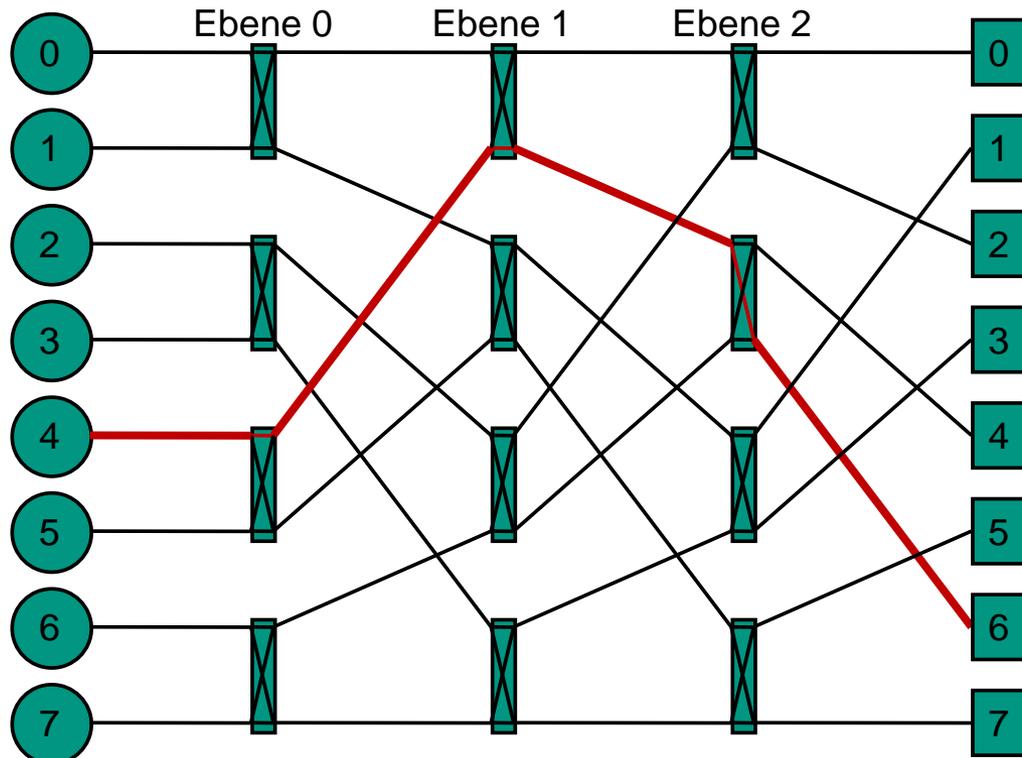
- Dynamische Verbindungsnetzwerke:
- Permutationen
  - Tauschpermutation T (Butterfly):
    - Negation des niedrigwertigen Bits
    - $T(a_n, a_{n-1}, \dots, a_2, a_1) = (a_n, a_{n-1}, \dots, a_2, \overline{a_1})$



# Verbindungsstrukturen

## Topologie:

- Dynamische Verbindungsnetzwerke:
- Speichergekoppeltes Omega-Netzwerk mit 8 Eingängen und 8 Ausgängen



Wegefindungsalgorithmus:

Zieladresse:  $[d_2, d_1, d_0]$

Verwende das  $i$ -te Bit der Zieladresse und bestimme die Richtung:

$d_i = 0$ : up

$d_i = 1$ : down

Zieladresse:  $[1, 1, 0]$

Ebene 0: up

Ebene 1: down

Ebene 2: down