

Rechnerstrukturen im SS 2009

Verbindungsstrukturen

Oliver Mattes

mattes@ira.uka.de



Universität Karlsruhe (TH)

Forschungsuniversität • gegründet 1825

Institut für Technische Informatik
Lehrstuhl für Rechnerarchitektur

25. Juni 2009

Verbindungsstrukturen:

- Charakterisierung von Verbindungsnetzen
- Statische Verbindungsnetze
 - Aufgabe 1
- Dynamische Verbindungsnetze
 - Aufgabe 2

Verständnisfragen:

- Aufgabe 3

Vergleich von Parallelrechnern:

- Aufgabe 4

Beurteilungskriterien:

- **Verbindungsgrad**: Anzahl der Kanten von einem Knoten zu anderen Knoten
- **Durchmesser**: maximale Distanz zwischen zwei Knoten
- **Blockierung** → blockierungsfrei?
- **Erweiterbarkeit** um zusätzliche Knoten
- **Skalierbarkeit** des Verbindungsnetzes: Vergrößerung möglich ohne die wesentlichen Eigenschaften zu verlieren
- **Ausfalltoleranz** (Redundanz)
 - minimale Bisektionsbreite
 - Diskonnektivität
- **Bandbreite**
- **Latenz**
- Komplexität der Pfadberechnung / Wegfindung
- ...

Minimale Bisektionsbreite:

Schneidet man einen Graphen in zwei gleich große in sich zusammenhängende Teile und betrachtet die Menge der Kanten, die diesen Schnitt kreuzen, so bezeichnet man die Kardinalität der kleinsten Kantenmenge – über alle möglichen Schnitte – als minimale Bisektionsbreite.

Verbindungsstrukturen: Charakterisierung

Minimale Bisektionsbreite:

Schneidet man einen Graphen in zwei gleich große in sich zusammenhängende Teile und betrachtet die Menge der Kanten, die diesen Schnitt kreuzen, so bezeichnet man die Kardinalität der kleinsten Kantenmenge – über alle möglichen Schnitte – als minimale Bisektionsbreite.

Diskonnektivität:

Diskonnektivität = #Knoten / min. Bisektionsbreite

Kosteneffektivität:

Kosteneffektivität = Verbindungsgrad * max(Durchmesser, Diskonnektivität)

Übertragungsbandbreite / Durchsatz:

Die maximale Übertragungsleistung des Verbindungsnetztes oder einzelner Verbindungen. Meist theoretisch errechnet.

Bisektionsbandbreite:

Maximale Datenmenge, die das Netzwerk über die Bisektionslinie, die das Netzwerk in zwei Hälften teilt, pro Sekunde transportieren kann.

Ausfalltoleranz (Redundanz):

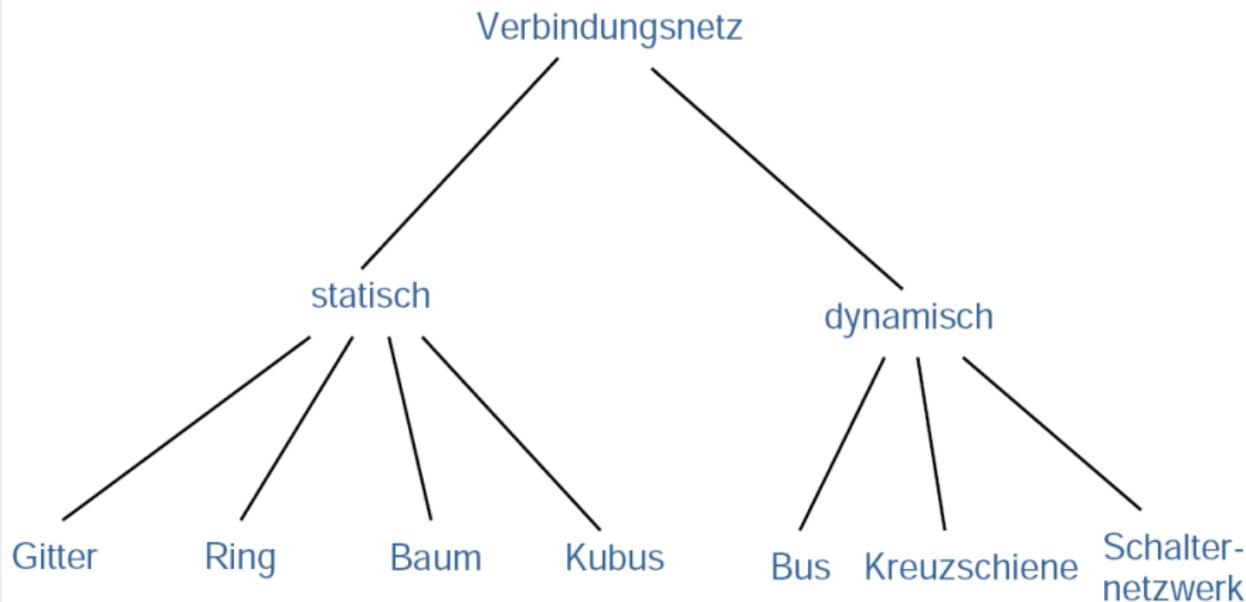
Kriterien:

- minimale Bisektionsbreite
- Diskonnektivität
- Wurzelknoten
- Flaschenhals
- Ausweichverbindungen
- ...

Interpretation der Daten:

- Vergleich mit anderen Netzen
- Beachtung von Teilnetzen (z.B. Ring-Würfel-Netzwerk)
- ...

Klassifizierung von Verbindungsnetzen:



Statische Verbindungsstrukturen

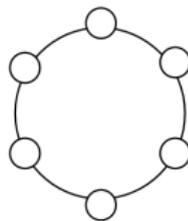
- In statischen Netzen existieren fest installierte Verbindungen zwischen Paaren von Netzknoten
- Steuerung des Verbindungsaufbaus ist Teil der Knoten

Dynamische Verbindungsstrukturen

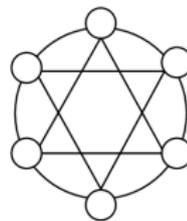
- Dynamische Netze enthalten eine Komponente „Schaltnetz“, an die alle Knoten über Ein- und Ausgänge angeschlossen sind.
- Direkte, fest installierte Verbindungen zwischen den Knoten existieren nicht.
- Alle notwendigen Steuerungsfunktionen sind im Schaltnetz konzentriert

Statische Verbindungsstrukturen

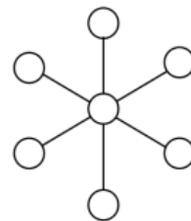
- Kette
- Ring
- Stern
- Baum (event. mit Fat-Tree)
- Chordaler Ring
- Gitter



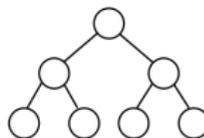
Ring



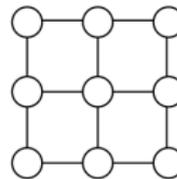
Chordaler Ring



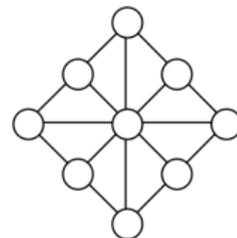
Stern



Baum



Gitter mit vier
Nachbarknoten

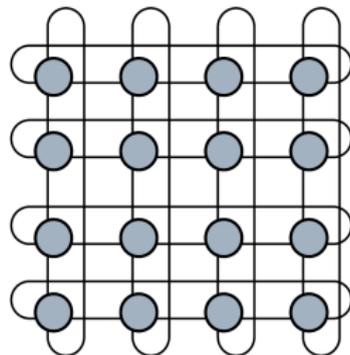


Gitter mit acht
Nachbarknoten

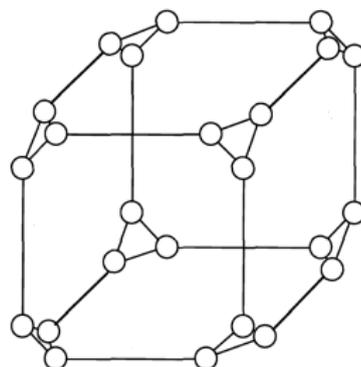
Statische Verbindungsstrukturen

- Torus
- Pyramide
- Würfel
- n-dimensionaler Hyperwürfel
- Ring-Würfel-Netzwerk
Cube-Connected-Cycle (CCC)

2-dim. Torus

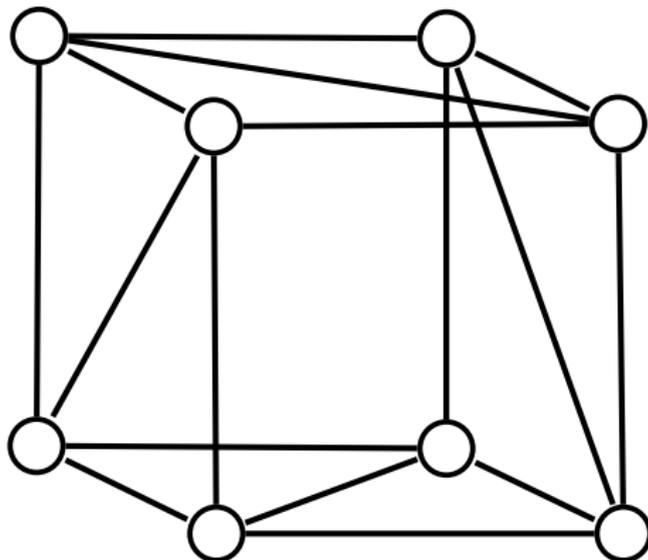


CCC



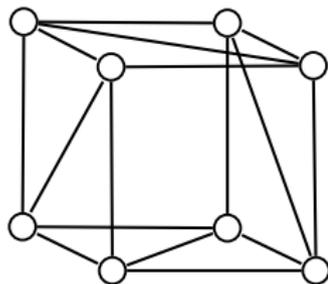
Aufgabe 1 – Statische Verbindungsstrukturen

Gegeben sei ein Verbindungsnetzwerk mit der nachfolgend dargestellten Topologie:



Aufgabe 1 – Statische Verbindungsstrukturen

- a) Bestimmen Sie den Verbindungsgrad, den Diameter, die minimale Bisektionsbreite, die Diskonnektivität und die Kosteneffektivität.



Verbindungsgrad:	4
Durchmesser:	2
min. Bisektionsbreite:	6
Diskonnektivität:	$8/6 = 4/3 = 1,33$
Kosteneffektivität:	$4 * \max(2, 4/3) = 8$

Aufgabe 1 – Statische Verbindungsstrukturen

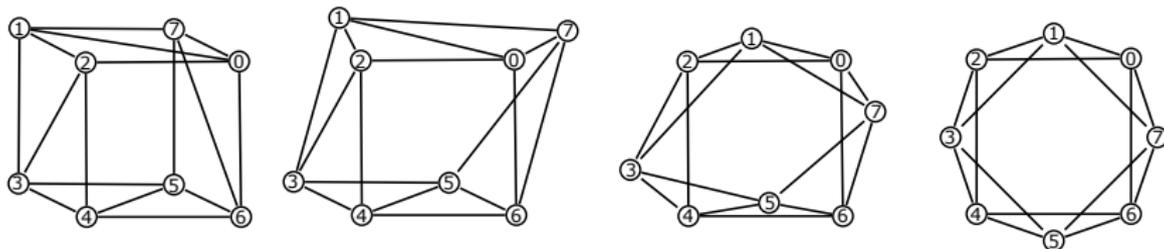
- b) **Um welche Art eines Verbindungsnetzwerkes handelt es sich in diesem Fall?**

Chordaler Ring mit Knotengrad 4

Aufgabe 1 – Statische Verbindungsstrukturen

- b) Um welche Art eines Verbindungsnetzwerkes handelt es sich in diesem Fall?

Chordaler Ring mit Knotengrad 4



Aufgabe 1 – Statische Verbindungsstrukturen

- c) **Liegt Redundanz vor? Wenn ja, wieviele Verbindungsleitungen können ausfallen bevor eine Verbindung zwischen zwei beliebigen Knoten nicht mehr geschalten werden kann?**
- Es liegt Redundanz vor.
 - Da der Verbindungsgrad jedes Knotens 4 ist und bidirektionale Leitungen verwendet werden, können bis zu drei Leitungen ausfallen und dennoch jeder Knoten von einem anderen erreicht werden.
 - Allerdings kann beim Ausfall einer Kante der Durchmesser steigen, das heißt es könnten längere Wege notwendig sein.

Aufgabe 1 – Statische Verbindungsstrukturen

- d) **Vergleichen Sie diese Netzwerktopologie mit den Topologien (unidirektionaler) Ring, 2D-Gitter, (binärer) Baum und Hyperkubus in den Punkten Verbindungsgrad, Durchmesser, minimale Bisektionsbreite, Diskonnektivität und Kosteneffektivität.**

$N = \#$ Knoten

	Aufgabe 2	Ring	2D-Gitter	Baum	Hyperkubus
Verbindungsgrad	4	2	$2 - 4$	$1 - 3$	$\log_2 N$
Durchmesser	$\lfloor \sqrt{N} \rfloor$	$\lfloor N/2 \rfloor$	$2(\sqrt{N} - 1)$	$2(\lceil \log_2 N \rceil - 1)$	$\log_2 N$
min. Bisektionsbreite	6	2	\sqrt{N}	1	$N/2$
Diskonnektivität	$N/6$	$N/2$	\sqrt{N}	N	2
Kosteneffektivität	$4\lfloor \sqrt{N} \rfloor$	N	$8(\sqrt{N} - 1)$	$3N$	$(\log_2 N)^2$

Statische Verbindungsstrukturen

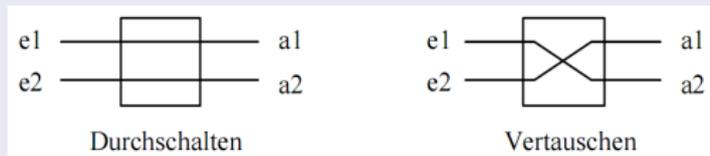
- In statischen Netzen existieren fest installierte Verbindungen zwischen Paaren von Netzknoten
- Steuerung des Verbindungsaufbaus ist Teil der Knoten

Dynamische Verbindungsstrukturen

- Dynamische Netze enthalten eine Komponente „Schaltnetz“, an die alle Knoten über Ein- und Ausgänge angeschlossen sind.
- Direkte, fest installierte Verbindungen zwischen den Knoten existieren nicht.
- Alle notwendigen Steuerungsfunktionen sind im Schaltnetz konzentriert

- **Bus**, Mehrfachbus
- **Kreuzschienenverteiler** (Crossbar Switch)
Alle angeschlossenen Prozessoren und Speicher können paarweise disjunkt gleichzeitig und blockierungsfrei miteinander kommunizieren.

- **Bus**, Mehrfachbus
- **Kreuzschienenverteiler** (Crossbar Switch)
Alle angeschlossenen Prozessoren und Speicher können paarweise disjunkt gleichzeitig und blockierungsfrei miteinander kommunizieren.
- **Schalternetzwerke** aus Zweierschaltern

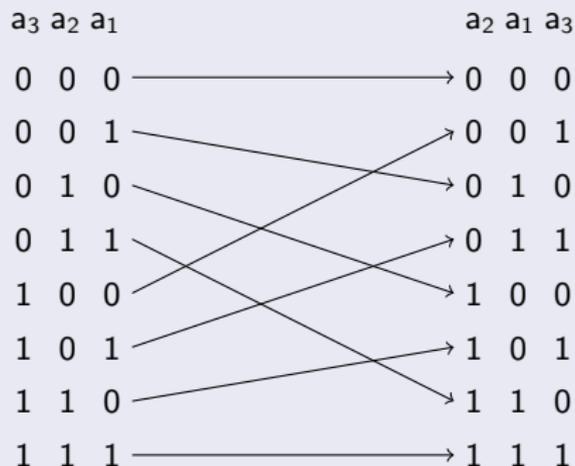


- **Permutationsnetz**
 - Einstufige, mehrstufige Permutationsnetze
 - Reguläre Permutationsnetzwerke:
 p Eingänge, p Ausgänge, k Stufen mit je $p/2$ Zwischenschaltern
 - Irreguläre Permutationsnetzwerke:
weisen Lücken auf

Mischpermutation

Kreisverschiebung der Adressbits:

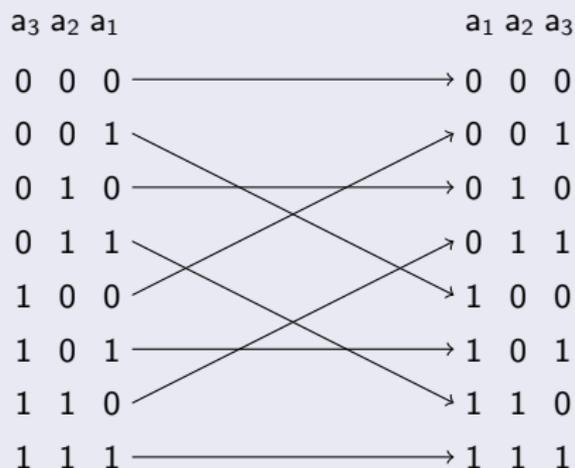
$$U(a_n, a_{n-1}, \dots, a_2, a_1) = (a_{n-1}, \dots, a_2, a_1, a_n)$$



Kreuzpermutation

Vertauschen des hochwertigsten mit dem niedrigwertigsten Adressbit:

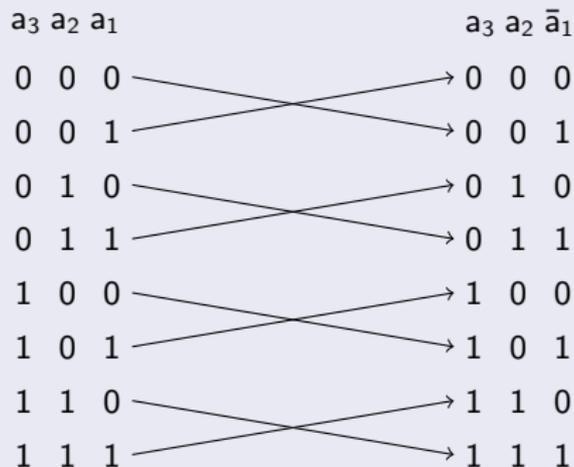
$$U(a_n, a_{n-1}, \dots, a_2, a_1) = (a_1, a_{n-1}, \dots, a_3, a_2, a_n)$$



Tauschpermutation

Negation des niedrigwertigsten Adressbit:

$$U(a_n, a_{n-1}, \dots, a_2, a_1) = (a_n, a_{n-1}, \dots, a_2, \bar{a}_1)$$



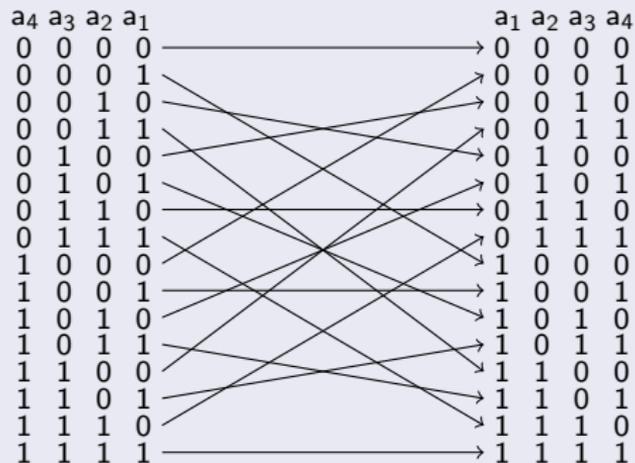
Umkehrpermutation

Spiegelung aller Adreßbits um die Mitte der Adressbitfolge:

$$U(a_n, a_{n-1}, \dots, a_2, a_1) = (a_1, a_2, \dots, a_{n-1}, a_n)$$

Für $n = 2$ und $n = 3$ ergibt sich dasselbe Grundmuster wie bei der Kreuzpermutation.

Umkehrpermutation

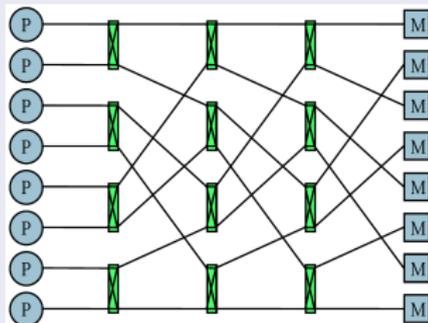


Mehrstufige Permutationsnetzwerke:

- jeweils aus einem bestimmten Grundmuster aufgebaut
- oft mit einer der eben vorgestellten Permutationen

Beispiele:

- Omega-Netzwerk
 - Mischpermutation

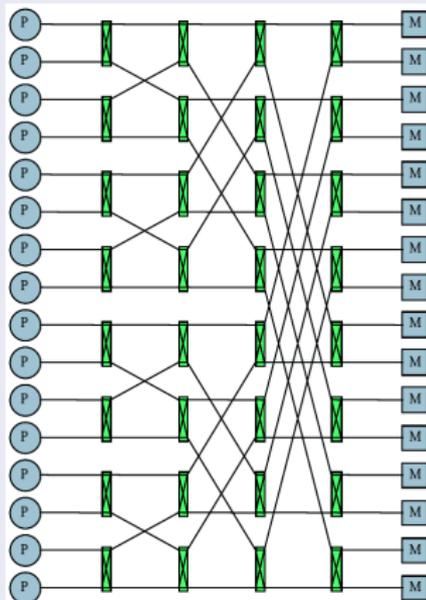


Mehrstufige Permutationsnetzwerke:

- jeweils aus einem bestimmten Grundmuster aufgebaut
- oft mit einer der eben vorgestellten Permutationen

Beispiele:

- Omega-Netzwerk
 - Mischpermutation
- Switching-Banyan-Netzwerk
 - Kreuzpermutation



Mehrstufige Permutationsnetzwerke:

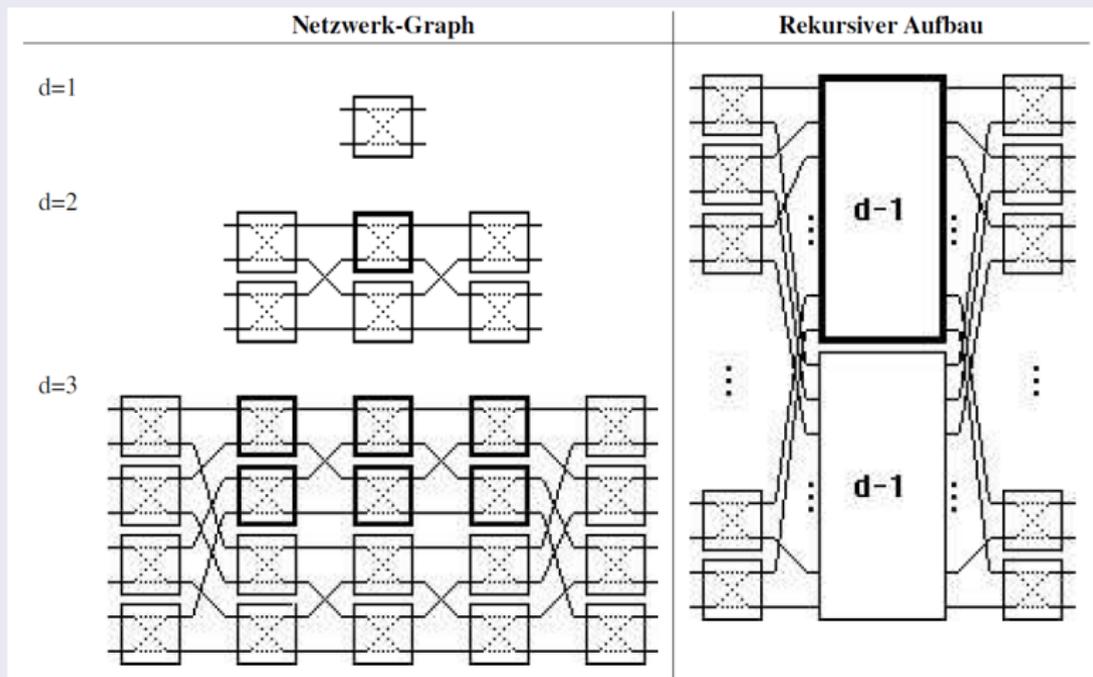
- jeweils aus einem bestimmten Grundmuster aufgebaut
- oft mit einer der eben vorgestellten Permutationen

Beispiele:

- Omega-Netzwerk
 - Mischpermutation
- Switching-Banyan-Netzwerk
 - Kreuzpermutation
- Benes-Netzwerk
 - rekursiver Aufbau

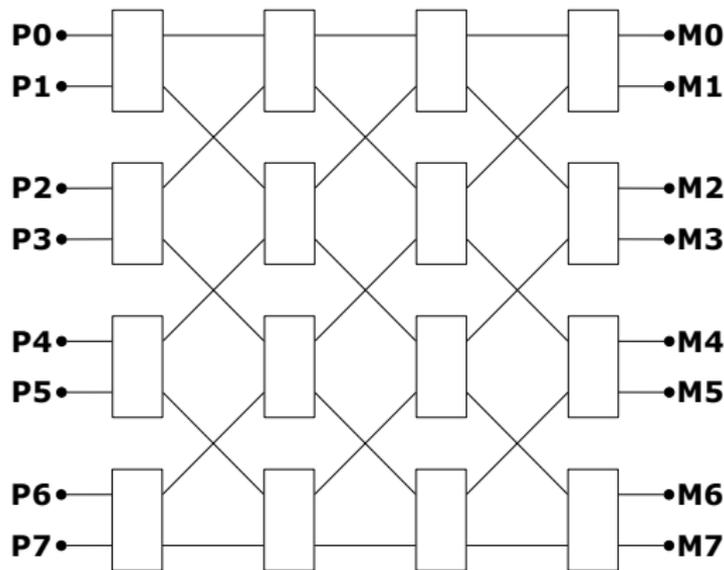
Benes-Netzwerk

- rekursiver Aufbau



Aufgabe 2 – Dynamische Verbindungsstrukturen

Gegeben sei ein dynamisches Verbindungsnetzwerk, das 8 Prozessoren (P0 – P7) mit 8 Speichern (M0 – M7) wie folgt verbindet:



Aufgabe 2 – Dynamische Verbindungsstrukturen

- a) **Kann zwischen jedem Prozessor- und Speicherpaar eine Verbindung hergestellt werden?**

Ja!

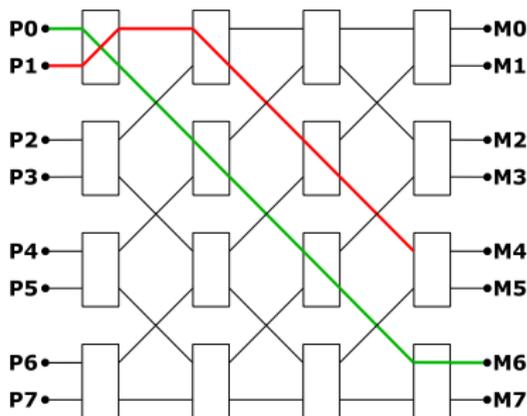
Aufgabe 2 – Dynamische Verbindungsstrukturen

- a) Kann zwischen jedem Prozessor- und Speicherpaar eine Verbindung hergestellt werden?

Ja!

- c) Was ist die minimale Verbindungszahl ab der eine Blockierung auftritt? Geben Sie ein Beispiel an.

Schon bei zwei Verbindungen kann eine Blockierung auftreten: z.B. bei $P_0 \rightarrow M_6$ und $P_1 \rightarrow M_7$



Aufgabe 2 – Dynamische Verbindungsstrukturen

- a) **Kann zwischen jedem Prozessor- und Speicherpaar eine Verbindung hergestellt werden?**

Ja!

- c) **Was ist die minimale Verbindungszahl ab der eine Blockierung auftritt? Geben Sie ein Beispiel an.**

Schon bei zwei Verbindungen kann eine Blockierung auftreten: z.B. bei $P_0 \rightarrow M_6$ und $P_1 \rightarrow M_7$

- d) **Ist das Netzwerk redundant? Begründen Sie Ihre Antwort.**

Nein. Auf der einen Seite gibt es für bestimmte Paare mehrere Möglichkeiten (vgl. $P_2 \rightarrow M_4$), aber ebenso gibt es Paare, bei denen schon der Ausfall einer Verbindung die Weiterleitung ausschließt (z.B. $P_0 \rightarrow M_6$).

Aufgabe 2 – Dynamische Verbindungsstrukturen

b) Kann jede Permutation generiert werden? Begründen Sie Ihre Antwort!

Nein. Beweis durch Widerspruch.

Annahme: jede Permutation kann generiert werden.

Gesucht: mindestens eine Permutation, für die die Annahme nicht gilt.

- Bei einer paarweisen Mischpermutation (Kreisverschiebung), hier also Verbindung von P0 und P1 mit M6 bzw. M7 gibt es nur einen möglichen Verbindungsweg, der gleichzeitig für beide Verbindungen benutzt werden müßte

⇒ Blockierung

Aufgabe 3 – Verständnisfragen

Kein Anspruch auf Vollständigkeit, es kann noch weitere Punkte geben!

a) **Welche Vorteile bieten Netzwerke auf Basis eines 3D-Torus?**

konstanter Verbindungsgrad, einfache Erweiterbarkeit, einfaches Routing, hohe Fehlertoleranz,...

Aufgabe 3 – Verständnisfragen

Kein Anspruch auf Vollständigkeit, es kann noch weitere Punkte geben!

a) **Welche Vorteile bieten Netzwerke auf Basis eines 3D-Torus?**

konstanter Verbindungsgrad, einfache Erweiterbarkeit, einfaches Routing, hohe Fehlertoleranz,...

b) **Welche Vorteile bieten Netzwerke auf Basis eines Fat-Trees?**

einfacher Aufbau, einfaches Routing, einfaches Adressierungsschema,...

c) Was sind die Vor-/Nachteile des Shared-Memory-Programmiermodells?

- Vorteile: Niedrige Latenzen, hohe Kommunikationsbandbreite, (implizite Kommunikation)
- Nachteile: Skalierbarkeit

c) Was sind die Vor-/Nachteile des Shared-Memory-Programmiermodells?

- Vorteile: Niedrige Latenzen, hohe Kommunikationsbandbreite, (implizite Kommunikation)
- Nachteile: Skalierbarkeit

d) Was sind die Vor-/Nachteile des Message-Passing-Programmiermodells?

- Vorteile: Skalierbarkeit
- Nachteile: im Vergleich mit Shared-Memory hohe Latenzen, komplexe Programmierung (explizite Kommunikation)

Aufgabe 4 – Vergleich von Parallelrechnern

a) Wieviel GFLOPs trägt jeder einzelne Prozessor zur theoretischen Spitzenleistung bei?

- BlueGene/L:
 $478200 \text{ GFLOPS} / 212992 \text{ Proz} = 2,25 \text{ GFLOPS/Proz}$
- HP HX6000:
 $1900 \text{ GFLOPS} / (101 * 2 + 10 * 8) \text{ Proz} = 6,74 \text{ GFLOPS/Proz}$
- Achtung, dies sind sehr theoretische und vereinfachte Werte!

Aufgabe 4 – Vergleich von Parallelrechnern

a) Wieviel GFLOPs trägt jeder einzelne Prozessor zur theoretischen Spitzenleistung bei?

- BlueGene/L:
 $478200 \text{ GFLOPS} / 212992 \text{ Proz} = 2,25 \text{ GFLOPS/Proz}$
- HP HX6000:
 $1900 \text{ GFLOPS} / (101 * 2 + 10 * 8) \text{ Proz} = 6,74 \text{ GFLOPS/Proz}$
- Achtung, dies sind sehr theoretische und vereinfachte Werte!

b) Was für ein Netzwerktyp/-struktur wird verwendet? (Topologie, Hersteller, statisches oder dynamisches Netz,...)

- BlueGene/L:
3-dimensionaler Torus, Eigenentwicklung von IBM, statisches Netz
- HP HX6000:
Fat-Tree (Baumstruktur), Quadrics QsNet II Interconnect, dynamisches Netz, Rechnerknoten sind nicht im Netzwerk auf verschiedenen Ebenen verteilt

c) Wie groß ist der Durchmesser, d.h. die längste Verbindung zwischen zwei Knoten?

- BlueGene/L:
 $\sqrt[3]{212992} \approx 60 \Rightarrow$ Durchmesser von 3D-Torus ≈ 180
Achtung, dies ist eine Schätzung ohne Berücksichtigung des wirklichen Netzwerkaufbaus!
- HP HX6000:
Aufsteigen im Baum bis zur Wurzel und zurück: 5

Aufgabe 4 – Vergleich von Parallelrechnern

c) Wie groß ist der Durchmesser, d.h. die längste Verbindung zwischen zwei Knoten?

- BlueGene/L:
 $\sqrt[3]{212992} \approx 60 \Rightarrow$ Durchmesser von 3D-Torus ≈ 180
Achtung, dies ist eine Schätzung ohne Berücksichtigung des wirklichen Netzwerkaufbaus!
- HP HX6000:
Aufsteigen im Baum bis zur Wurzel und zurück: 5

d) Vergleichen Sie Bandbreite, Latenz und Blockierungsfreiheit der beiden Netzwerke.

- BlueGene/L:
Netzwerk ist nicht blockierungsfrei, Bandbreitenengpässe können auftreten, die Latenz ist unterschiedlich je nach Verbindung
- HP HX6000:
Netzwerk ist blockierungsfrei, Bandbreite von mehr als 800 MB/s, geringe Latenz

e) Gibt es einen Flaschenhals?

- BlueGene/L:
Prinzipiell nein.
Je nach Wegwahlverfahren können aber Probleme auftreten.
- HP HX6000:
Nein, da ein „Dynamic Fat-Tree“ verwendet wird, bei dem jede Permutation geschaltet werden kann

e) Gibt es einen Flaschenhals?

- BlueGene/L:
Prinzipiell nein.
Je nach Wegwahlverfahren können aber Probleme auftreten.
- HP HX6000:
Nein, da ein „Dynamic Fat-Tree“ verwendet wird, bei dem jede Permutation geschaltet werden kann

f) Bewerten Sie die Skalierbarkeit und Erweiterbarkeit der beiden Netzwerkvarianten.

- BlueGene/L:
Sehr gut, das Netzwerk kann einfach um eine Ebene erweitert werden, prinzipiell unbeschränkt
- HP HX6000:
Sehr schlecht, erweiterbar um jeweils eine 2-er-Potenz, maximal 4096 angeschlossene Rechenknoten, d.h. maximal ≈ 40000 CPUs je nach Rechenknoten

Aufgabe 4 – Vergleich von Parallelrechnern

g) Nehmen Sie an, die Prozessorenzahl des HP XC6000 würde an die Größenordnung der Prozessorenzahl des BlueGene/L angepasst. Welches Problem hinsichtlich der Netzwerkkommunikation ergibt sich hierbei? Insbesondere welche Veränderungen am Netzwerk müssten durchgeführt werden, damit es die Anforderungen hinsichtlich Blockierungsfreiheit weiterhin erfüllt?

- Netzwerk besteht aus drei Schichten miteinander verknüpfter Switches
 - Ebene 1 - Switche haben genausoviel Verbindungen zur nächsten Ebene wie angebundene Rechenknoten
 - In jeder Ebene nimmt die Portzahl quadratisch zu
- ⇒ Netzwerkgröße limitiert durch die Größe (Portzahl) der Switche
- Grafik und Tabelle zum maximalen Ausbau von Quadrics QsNet II auf der Homepage <http://www.quadrics.com/>
- ⇒ Nach Erweiterung auf ~200000 Prozessoren müßten Switche mit mehr als 20000 Port verwendet werden.

Aufgabe 4 – Vergleich von Parallelrechnern

g) Beispiel: Infiniband-Switch mit 4096 Ports



h) Welche Vereinfachungen im Netzwerk könnten gemacht werden, um den Aufwand für Netzwerkhardware zu verringern und was wären die Auswirkungen hiervon?

- **Ausdünnung der Verbindungen in Richtung Wurzel des Baumes**
 - ⇒ Keine Blockierungsfreiheit mehr
 - ⇒ Durchsatz verringert sich
 - ⇒ Redundanz geringer

Anmeldung zur Klausur am 11.08.2009

- Die Anmeldung zur Klausur ist vom **29.06. bis 02.08.** möglich ist!
- ⇒ Online über das Studierendenportal
oder
- ⇒ Anmeldescheine in den orangen Kasten beim schwarzen Brett des Lehrstuhls für Rechnerarchitektur im **1. Stock in Gebäude 20.20**
(... wirklich im 1. Stock und nicht beim BIT8000...)

Spätere Anmeldungen sind nicht möglich!

- Rücktritt ist bis zum 09.08. elektronisch möglich

Fragen?