Leistungsbewertung rekonfigurierbarer Verbindungsnetze für Multiprozessorsysteme

Günter Hommel und Dietmar Tutsch
Technische Universität Berlin
Institut für Technische Informatik und Mikroelektronik
DietmarT@cs.tu-berlin.de
http://pdv.cs.tu-berlin.de

Überblick

1. Einleitung
2. Multistage Interconnection Networks
3. Bidirectional Multistage Interconnection Networks
4. Rekonfigurierung
5. Modellierung
6. Zusammenfassung

Einleitung

- Multiprozessorsysteme erlangen immer wieder neue Einsatzfelder
- z.B. SunRay-System benötigt (idealerweise) Multiprozessorsystem zur Bedienung von SunRay-Terminals

- aber auch klassische Anwendungen
  z.B. Hochleistungsrechnen im wissenschaftlichen Bereich

Einleitung (Forts.)

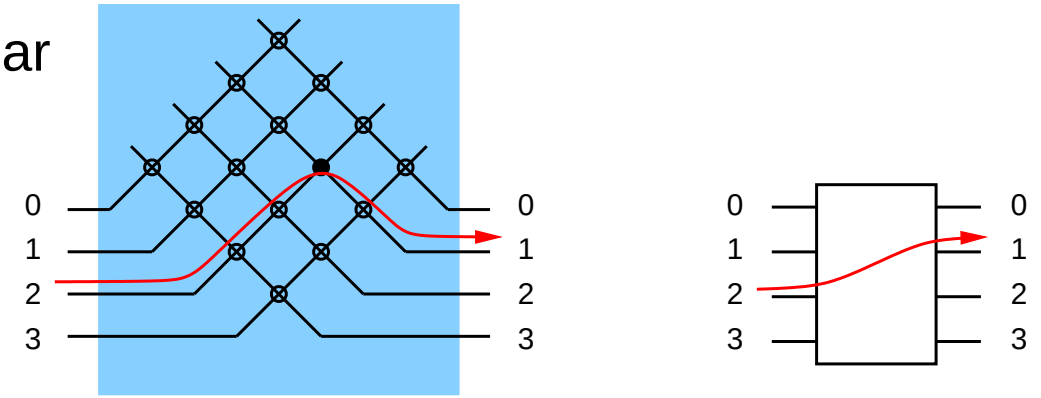
- Multiprozessorarchitekturen Gegenstand der Forschung
- insbesondere: deren Kommunikationsnetze
- zahlreiche Kommunikationsnetze dienen als Verbindungsnetz zwischen den Prozessoren
  - Bus
  - Ring
  - Crossbar
  - Multistage Interconnection Network

Multistage Interconnection Network (MIN)

- dtsch.: mehrstufiges Verbindungsnetz
- mehrstufige Koppelanordnungen (Fernmeldetechnik)
- theoretische Untersuchungen durch Beneš in 60er Jahren
- verbindet hohe Zahl von Kommunikationsteilnehmern, z.B. hohe Zahl von Prozessoren
- in Stufen angeordnete Crossbars mit spezieller Zwischenleitungsführung

Multistage Interconnection Network (Forts.)

- Crossbar
  - Kreuzschienenverteiler, Koppelvielfach
  - jeder Eingang kann mit jedem Ausgang verbunden werden
  - Verbindung läuft über *einen* Schalter
  - 4×4-Crossbar

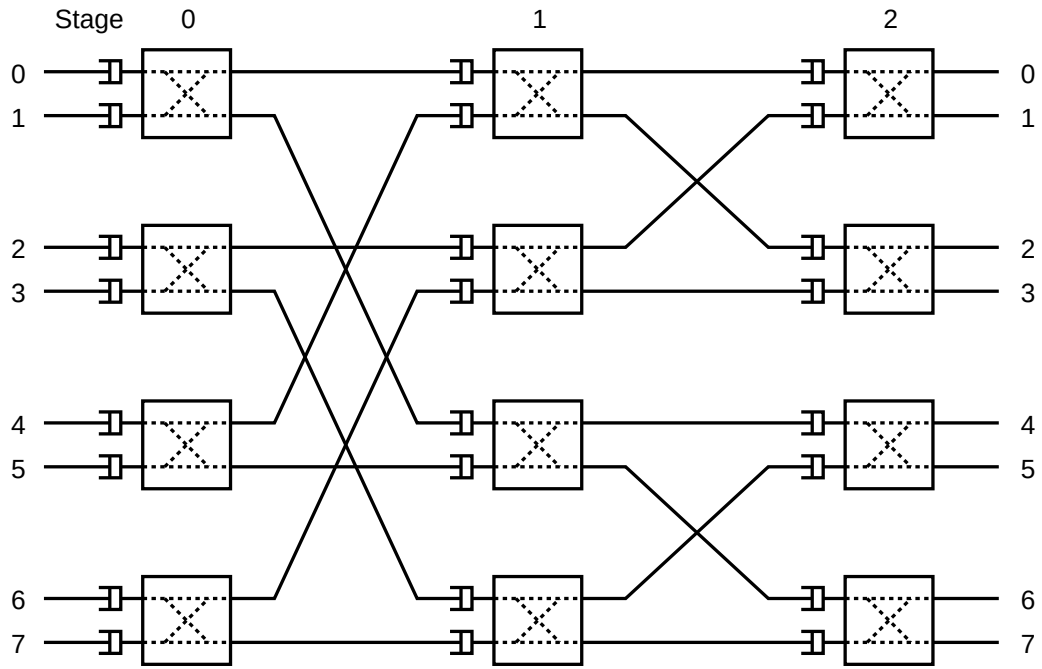


  - $N \times N$-Crossbar: N^2 Kreuzungspunkte
  - skalierbar, aber quadratisches Ansteigen der Komplexität
    → Multistage Interconnection Networks

Multistage Interconnection Networks (Forts.)

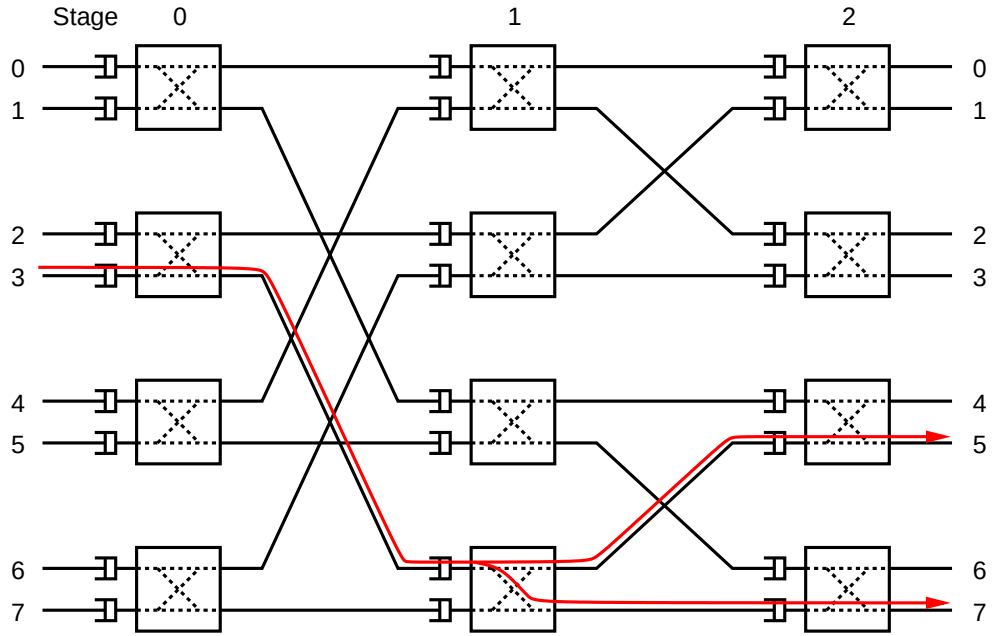
Parameter eines Multistage Interconnection Networks:

- $c \times c$ Crossbars (Switching Elements, SE)
- Netzgröße $N \times N$
- Vermittlungsart
- Pufferungsart
- Puffergröße
- Verlust/Backpressure
- Taktung
- spez. Eigenschaften (Banyan, Delta, etc.)



Multistage Interconnection Networks (Forts.)

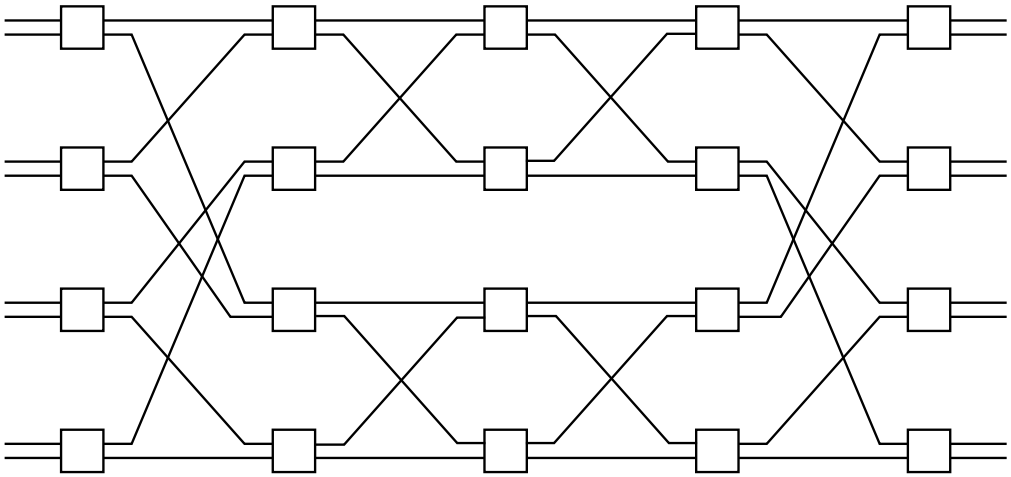
- Multicast (Mehrfachsendung) durch interne Vervielfachung (CRWR)



- weniger Pakete in den ersten Netzstufen
- Anzahl der Pakete wächst von der ersten zur letzten Stufe an

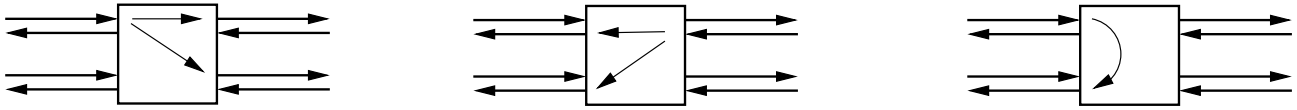
Multistage Interconnection Networks (Forts.)

- MIN mit zusätzlichen Stufen
  - fehlertolerant
  - nichtblockierend
- Beispiel: Beneš-Netz
  - Leitungsvermittlung: umdirigierend nichtblockierend
  - Paketvermittlung: nichtblockierend



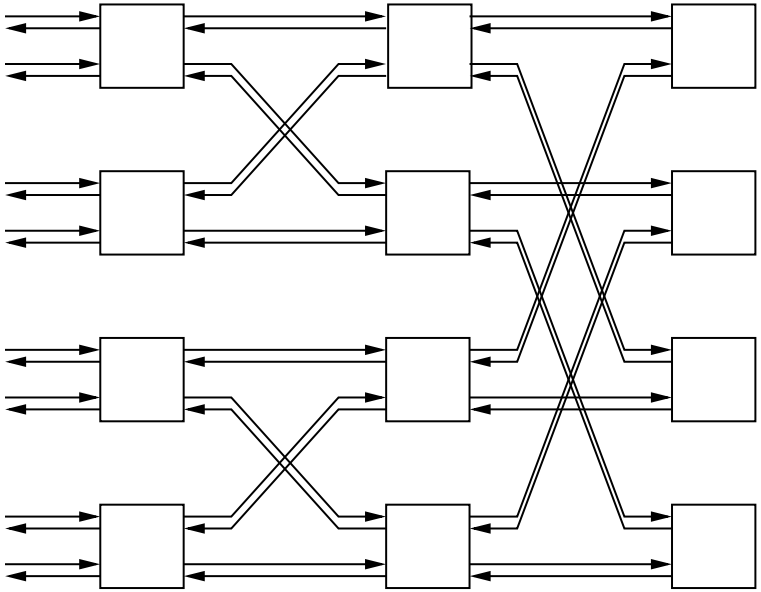
Bidirectional Multistage Interconnection Networks

- bidirektionale Leitungen → bidirektionales MIN
  - Verwendung von Turnaround-Crossbars mit 3 Modi



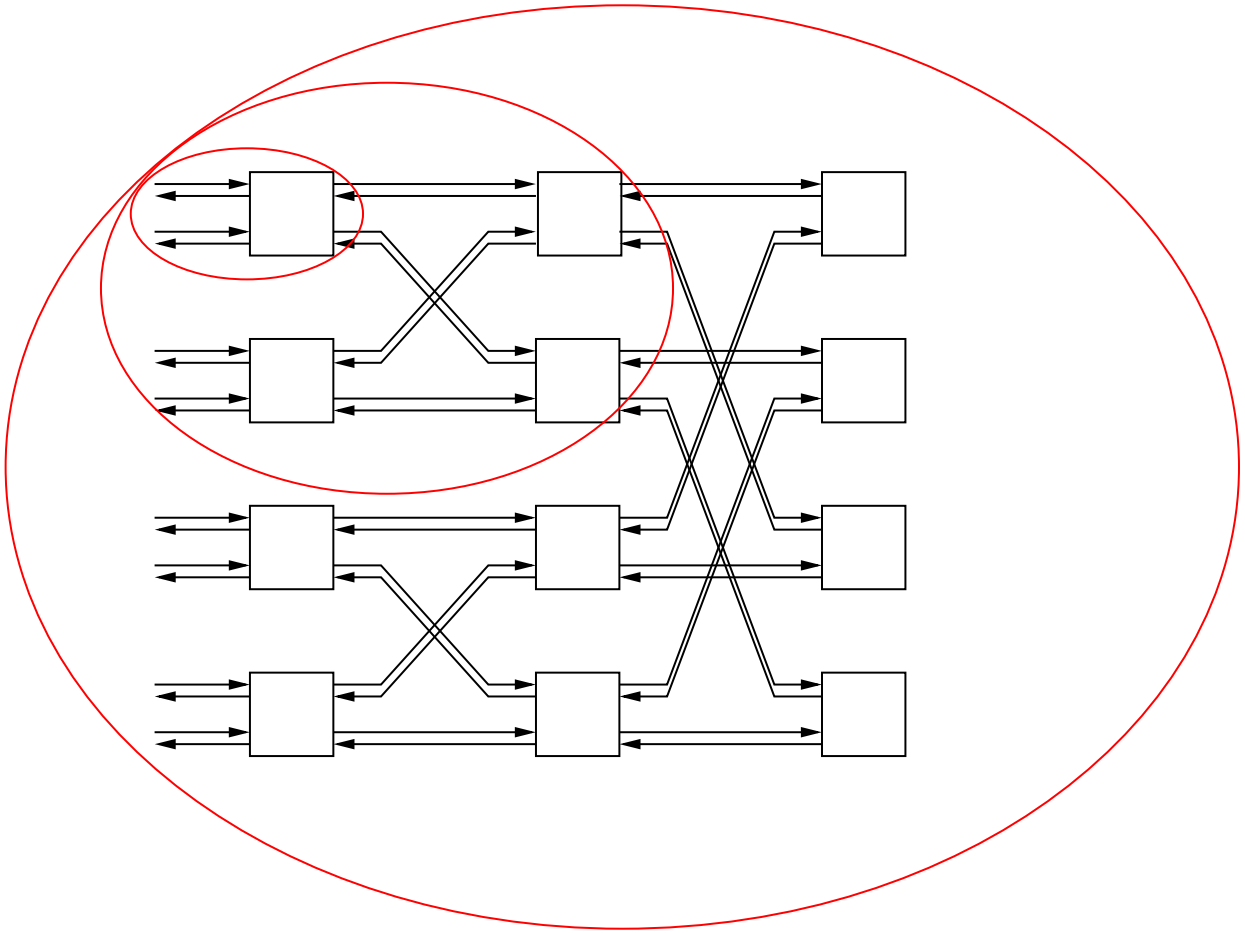
vorwärts rückwärts umkehr

  - Turnaround-MIN



Bidirectional Multistage Interconnection Networks (Forts.)

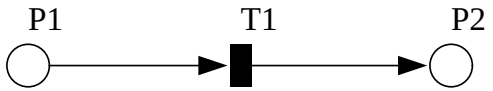
- Lokalität der Nachrichtenübertragungen



Rekonfigurierung

- „Verschieben“ der Leitungen bei neuer Aufgabe, so dass Lokalität entsteht → Rekonfigurierung des MIN
- offene Fragen
  - Zeitpunkt der Rekonfigurierung (alte Pakete!)
  - transientes Verhalten während der Rekonfigurierung
  - Leistungsgewinn durch Rekonfigurierung
- Beurteilung durch Leistungsbewertung
  - Messung

  - Modellierung

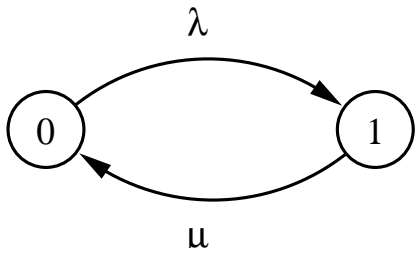


Modellierung

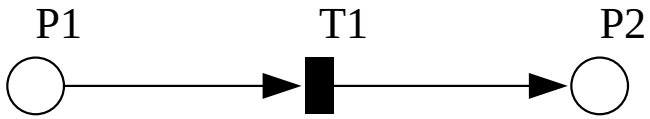
- Modellierung
  - Simulation
  - mathematische Analyse
- Simulation
  - Softwaremodell (keine Hardware)
    → schnell änderbar
  - stochastische Einflüsse:
    * Zuverlässigkeit der Ergebnisse
    * lange Simulationslaufzeiten
    * Konfidenzintervall, relativer Fehler

Modellierung (Forts.)

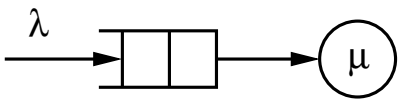
- mathematische Analyse
  - Markow-Ketten (CTMC, DTMC)



  - zeitbehaftete Petri-Netze

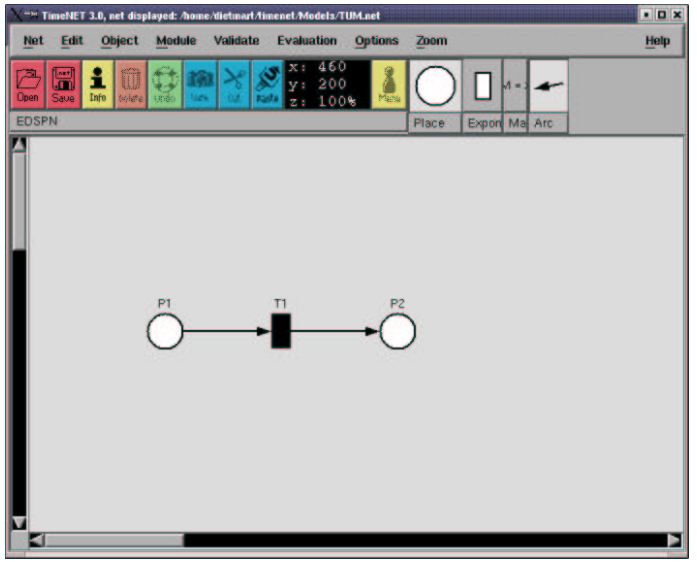


  - Warteschlangensysteme



Modellierung (Forts.)

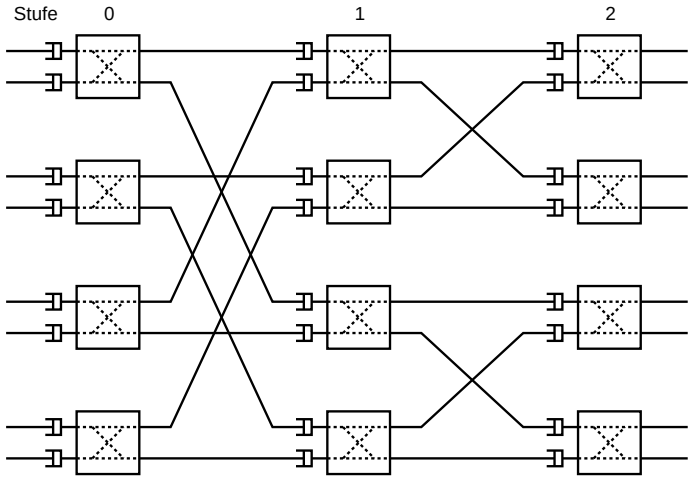
– viele Werkzeuge vorhanden

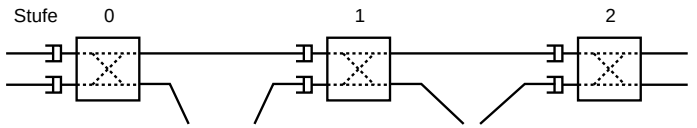


– dennoch häufig „Handarbeit“
  * komplexe Gleichungserstellung
  * hohe Entwicklungszeit
  * Lösung: automatische Generierung der Gleichungen über Regeln

Modellierung (Forts.)

- bisher: Multistage Interconnection Network
- Modellierung als zeitdiskrete Markow-Kette
- Annahmen zur Reduzierung des Zustandsraums
  - gleichförmiger Verkehr
  - Unabhängigkeit der Pakete
- Puffer: zwei Zustandsräume
  - vorderstes Paket
  - Anzahl der Pakete





Modellierung (Forts.)

- Hunderte von Zustandsübergängen
- Übergänge nach gleichem Muster innerhalb bestimmter Zustandsgruppen
  Beispiel:
  - Übergang (Paketanzahl): 4 Pakete ⟶ 3 Pakete
  - gleich zu: 3 Pakete ⟶ 2 Pakete
- Idee:
  Entwicklung von Regeln, wie Gleichungen aufzustellen sind
  ⟶ automatische Gleichungsgenerierung
- leichte Änderung der Regeln bei Variation des Netzes
- ungleichförmiger Verkehr und bidirektionale MIN:
  erhöhte Anzahl der Regeln, aber Struktur der Regeln bleibt

Modellierung (Forts.)

Vergleich verschiedener Modellierungsmethoden

Modell	Entwicklung	Auswertung	Genauigkeit	Speicher
Petri-Netz	100 Personenstunden	> 2 Wochen	3,0%	8,6 MByte
iteratives Petri-Netz	200 Personenstunden	20 Stunden	4,7%	5,5 MByte
C++/Akaroa	400 Personenstunden	4 Stunden	—	7,1 MByte
Analyse	1500 Personenstunden	< 1 Sekunde	3,1%	0,9 MByte
generierte Analyse	400 Personenstunden	< 1 Sekunde	3,1%	0,9 MByte

Zusammenfassung und Ausblick

- Multistage Interconnection Network als Kommunikationsnetz eines Multiprozessorsystems oder Multi-FPGA-Systems
- Rekonfigurierung je nach benötigter Lokalität
- Arbeitsschritte der ersten Phase
  - mathematisches Modell eines bidirektionalen MIN
  - ungleichförmiger und transienter Verkehr
  - Validierung durch Simulation
- spätere Phase
  - Konzepte zur Verkehrsvorhersage
  - Hardware-Realisierung