



TECHNISCHE
UNIVERSITÄT
DRESDEN

Fakultät Informatik, Professur für VLSI-Entwurfssysteme, Diagnostik und Architektur

Energieeffizienz und Performance von Networks-on-Chip

Marco Zulkowski
Marco.Zulkowski@mailbox.tu-dresden.de

Dresden, 11.01.2012



DRESDEN
concept
Exzellenz aus
Wissenschaft
und Kultur

Übersicht

Einleitung

NoC Architekturen im Vergleich

- **Untersuchte Architekturen**
- **Energieeffizienz**
 - **Leistungsaufnahme**
 - **Energieaufnahme je Paket**
- **Performance**
 - **Paketlatenz - gleichmäßig verteilter Traffic**
 - **Paketlatenz - lokalisierter Traffic**
 - **Paketlatenz - Streaming Traffic**

Auswertung

Einleitung

- Performance und Energiebedarf moderner SoC zunehmend durch Kommunikation bestimmt
- Sehr großer Anspruch an Skalierbarkeit (vgl. Shared Bus)
- Möglichkeiten für Einsatz von QoS steigen
 - Leichter umzusetzen
 - Verbesserte Fehlertoleranz
- Flit – Flow Control Unit
 - Elementares Datenpaket der Flusskontrolle

NoC Architekturen im Vergleich

Untersuchte Architekturen

- Vier Architekturen die sehr großes Spektrum abdecken
 - Circuit Switched Router
 - Wormhole Router
 - Virtual Channel Router mit QoS
 - Spekulativer Virtual Channel Router

NoC Architekturen im Vergleich

Circuit Switched Router

- Einfache statisch geschedulete Daten-Pfade
- Statische Allokation von Ressourcen
- Packed-Switched Netzwerk zum Initialisieren
- Flit : 64 bit

NoC Architekturen im Vergleich

Wormhole Router

- Dynamische Allokation
 - Header-Flit bestimmt Route für folgende Flits
- Aufwand der Allokation aber gering
- Flit : 74 bit

NoC Architekturen im Vergleich

Virtual Channel Router mit QoS (GuarVC)

- Wormhole Router mit Virtual Channels
- Source-Routing
- Zusätzlicher Header-Flit mit Routeninformation
- QoS
 - Guaranteed Throughput (GT)
 - Best Effort (BE)
- Flit : 68 bit

NoC Architekturen im Vergleich

Spekulativer Virtual Channel Router (SpecVC)

- Spekulation, dass wartende Pakete VC allokiieren
 - Switch Allokation wird vorzeitig ausgeführt
- Verringerte Latenzen
- Große Anzahl an Allokationslogik für gute Ressourcenteilung
- Flit : 82 bit

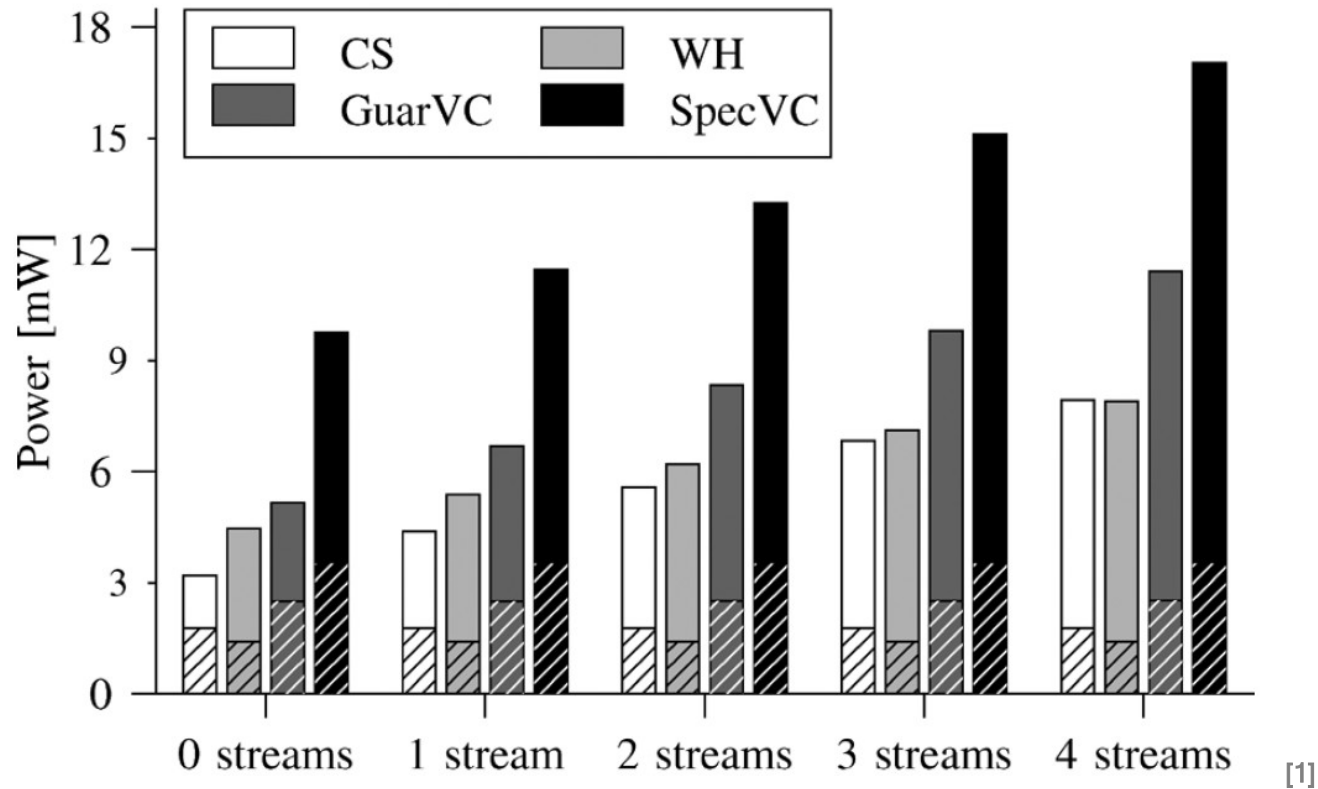
Energieeffizienz

Leistungsaufnahme

- Datenstrom durch einzelnen Router
- Ein Datenstrom je Port (N, S, O, W) zum gegenüberliegenden Port
- Paketgröße 256 Bit Nutzdaten
- Datenrate beträgt 30% der maximalen Linkbandbreite

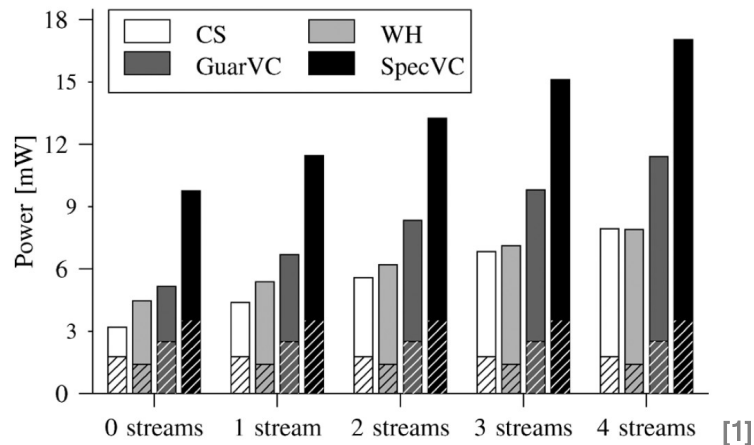
Leistungsaufnahme

Router



Leistungsaufnahme

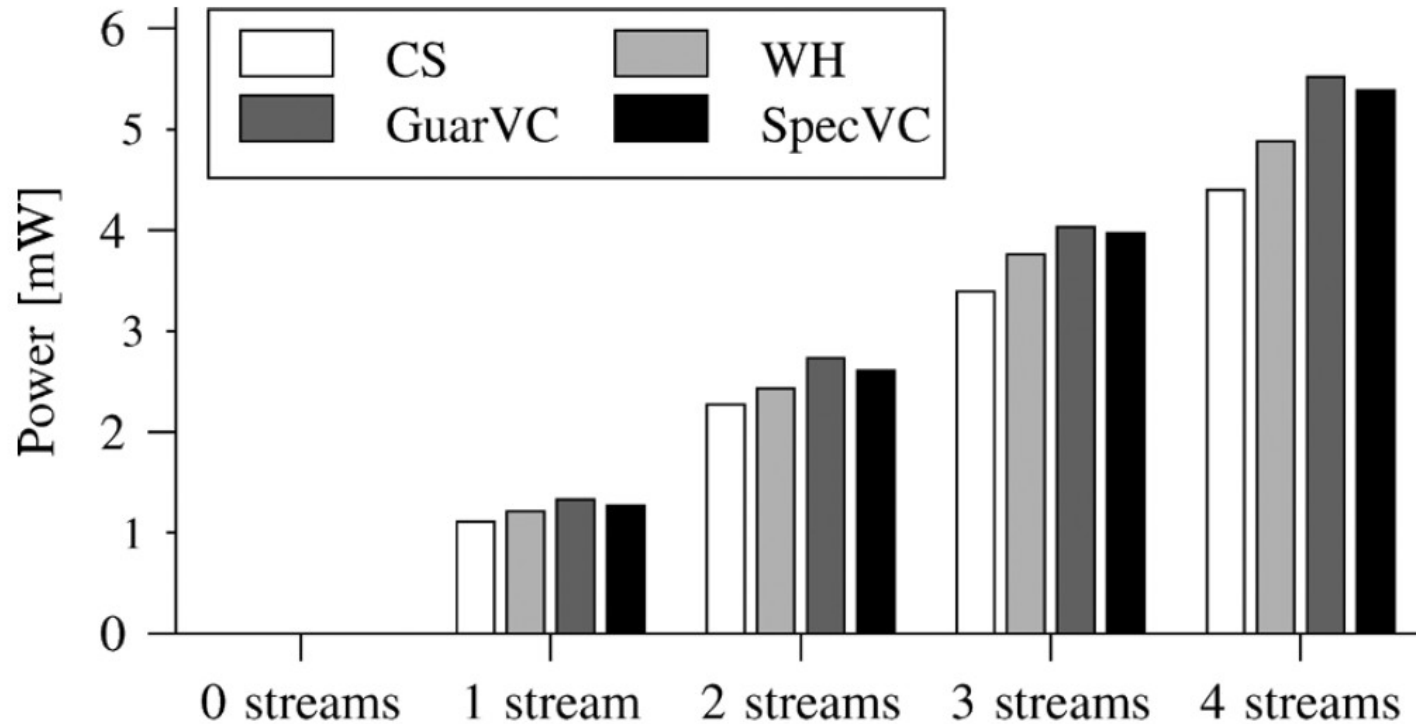
Router



- Hohe Leistungsaufnahme ohne Belastung
 - Aktivität im Taktbaum
 - FIFOs
- Keine Techniken zur Reduktion der Verlustleistung (schraffierter Bereich)

Leistungsaufnahme

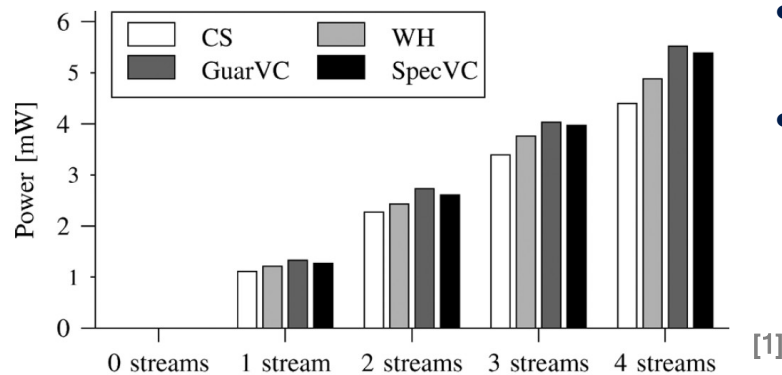
Link



[1]

Leistungsaufnahme

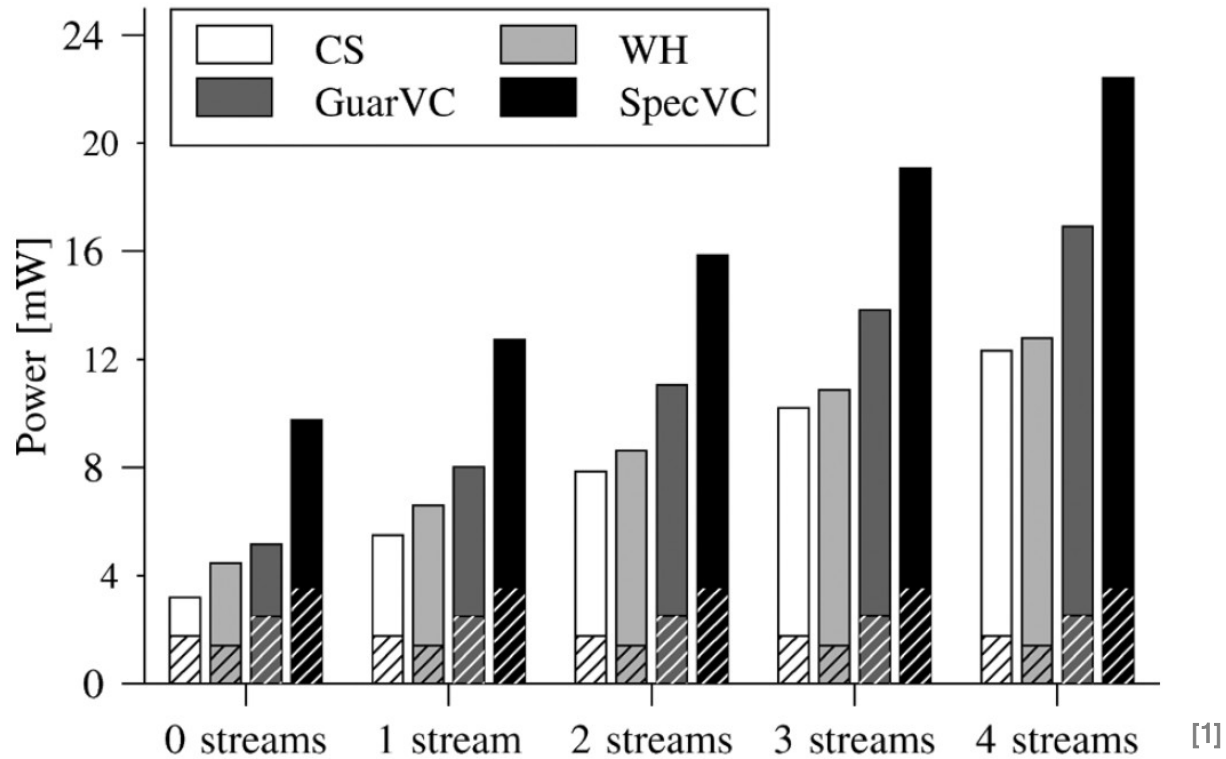
Link



- Höchste Aufnahme bei GuarVC
 - Zusätzliches Header-Flit pro Paket
- Niedrigste Aufnahme bei CS
 - Nutzt kleinste Flit-Größe

Leistungsaufnahme

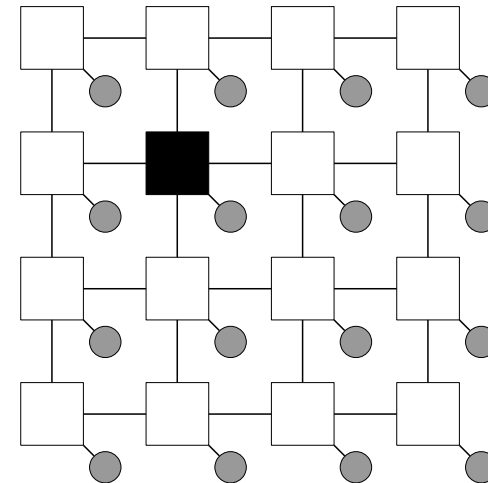
Gesamtleistungsaufnahme



Energieeffizienz

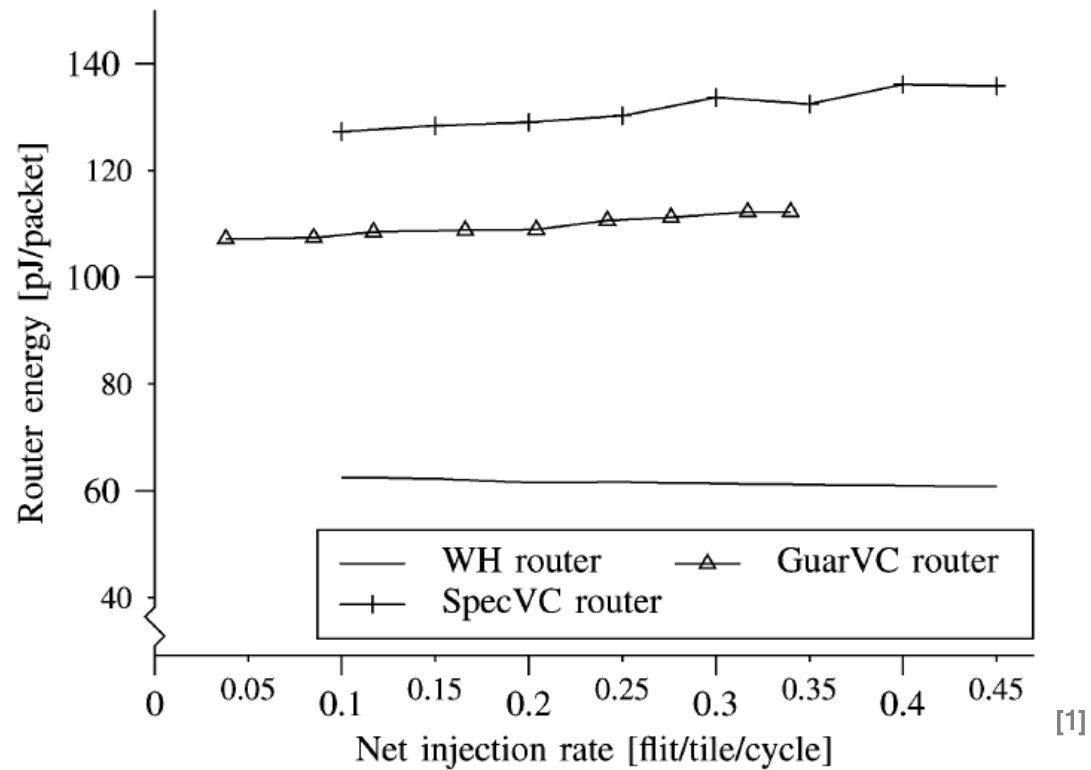
Energieaufnahme je Paket

- 4x4 Mesh
- Injektion von Paketen an jedem Router
 - Zufälliges Ziel (ohne sich selbst)
 - Variierende Injektionsrate
- Paketgröße 256 Bit
- GuarVC mit BE Paketen
- Messung an einem Router
 - Position (2,2)



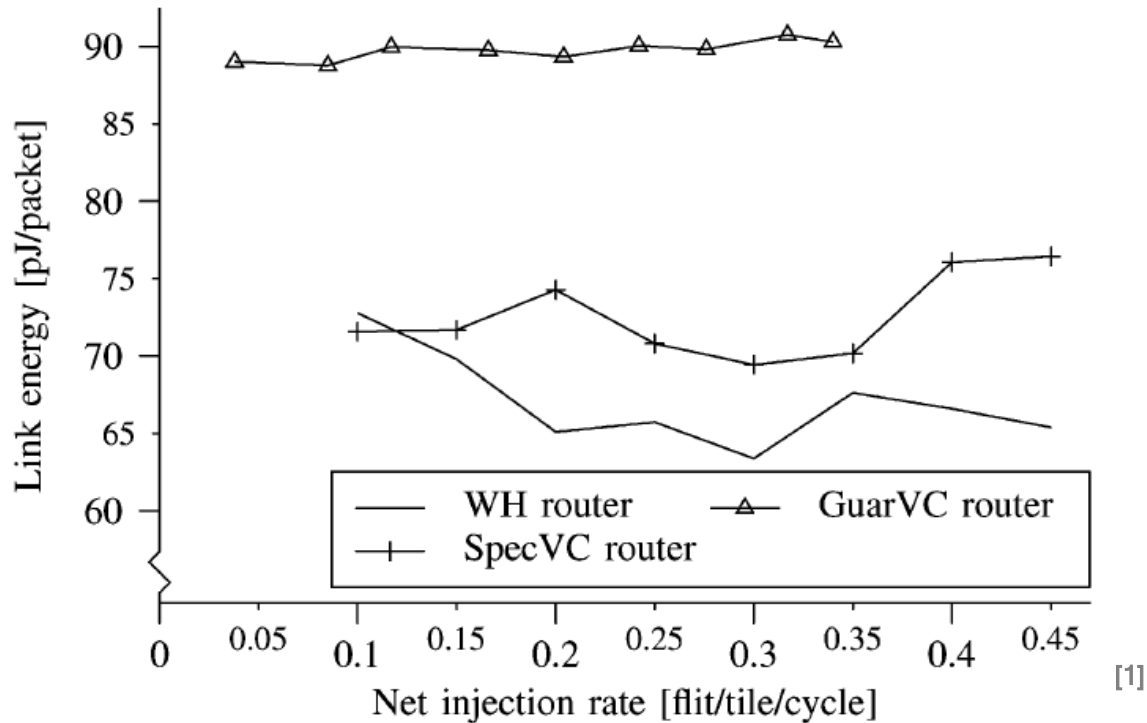
Energieaufnahme

Router



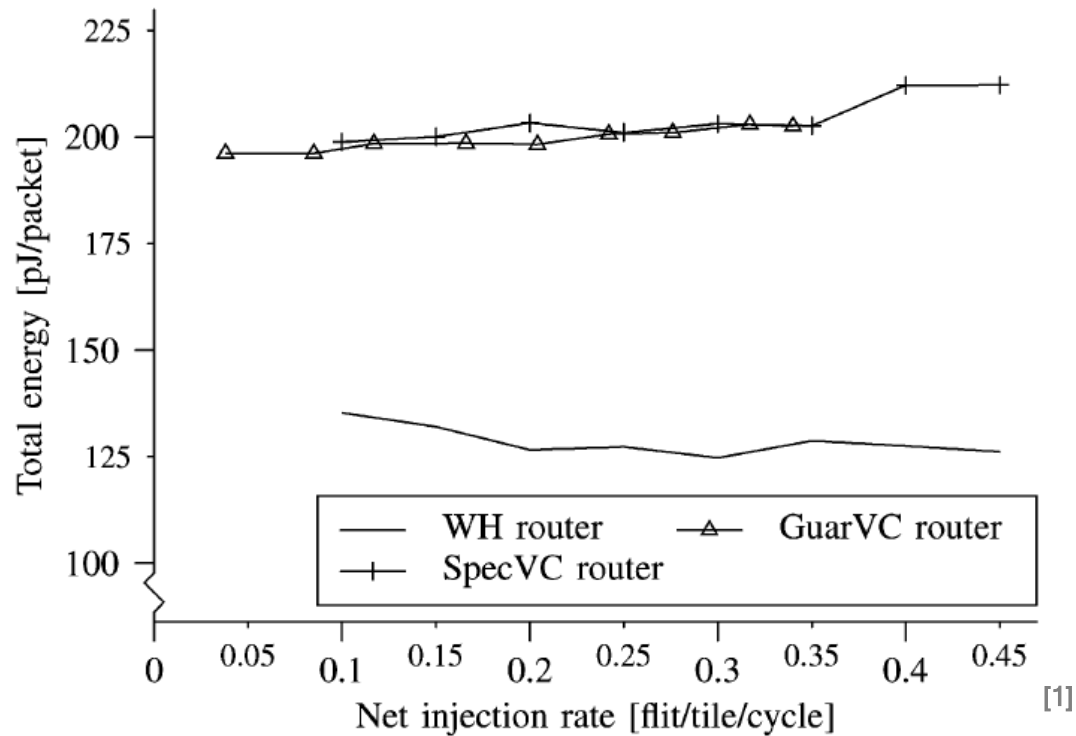
Energieaufnahme

Link



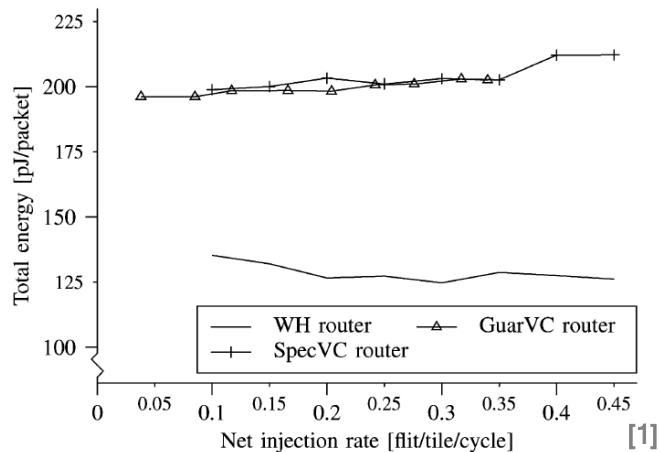
Energieaufnahme

Gesamtenergieaufnahme



Energieaufnahme

Gesamtenergieaufnahme

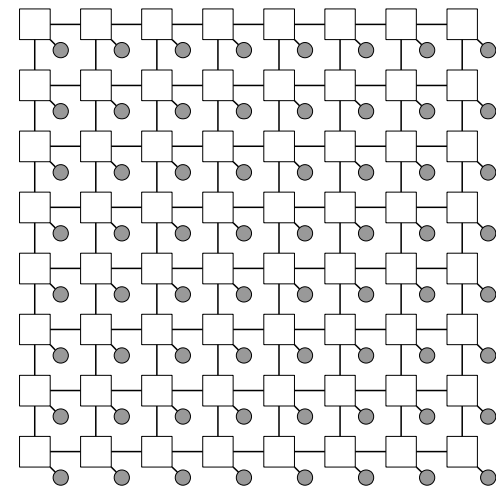


- Einfluss des zunehmenden Traffics sehr gering
 - Effektives Clock-Gating
- Anstieg bei VC Router
 - Mehr Aktivität bei Allokation und Flow-Control durch höhere Belegung der Ressourcen

Performance

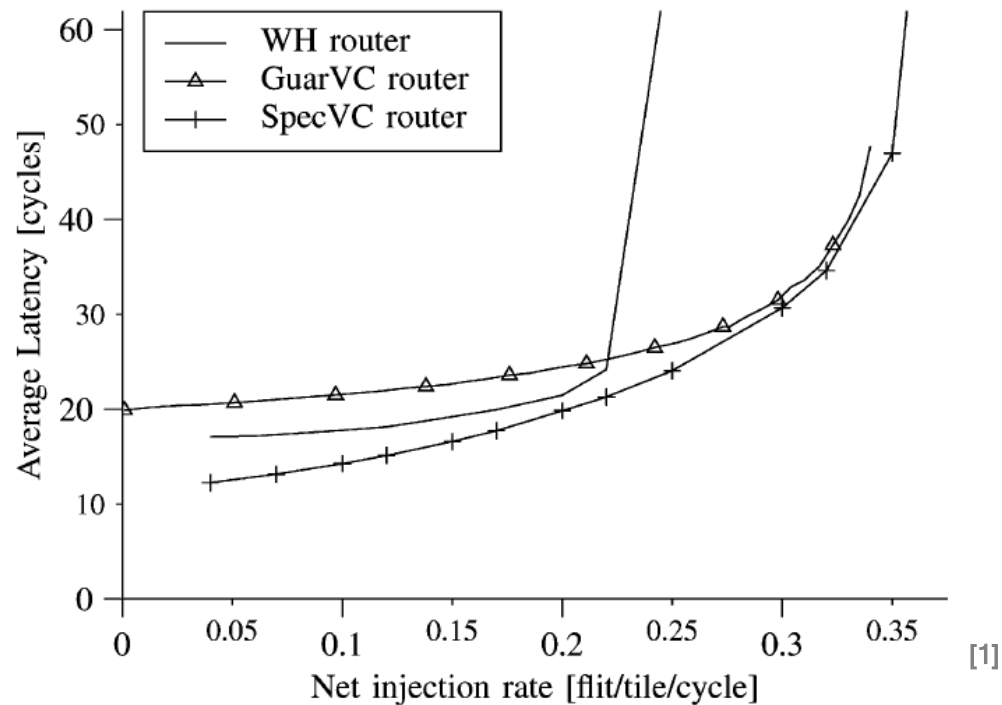
Paketlatenz

- 8x8 Mesh
- Injektion von Paketen an jedem Router
 - Zufälliges Ziel (ohne sich selbst)
 - Variierende Injektionsrate
- 4 Flit lange Pakete
- Messung der Zeit zwischen Quell- und Ziel-Router



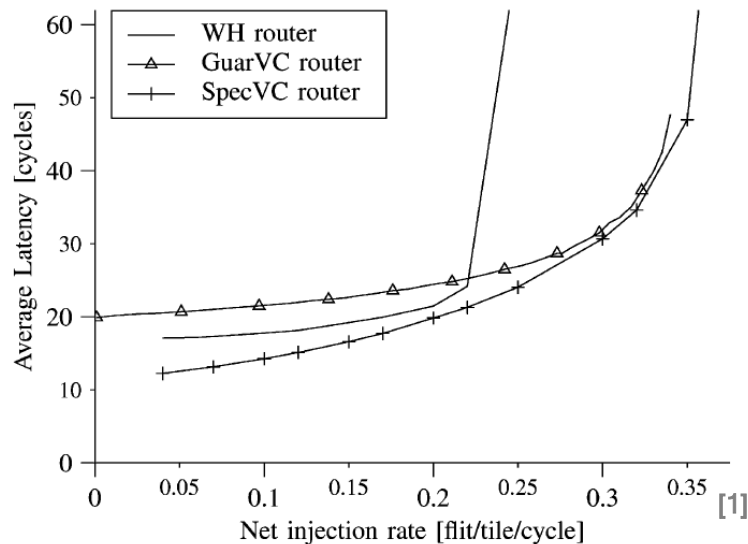
Paketlatenz

Gleichmäßig verteilter Traffic



Paketlatenz

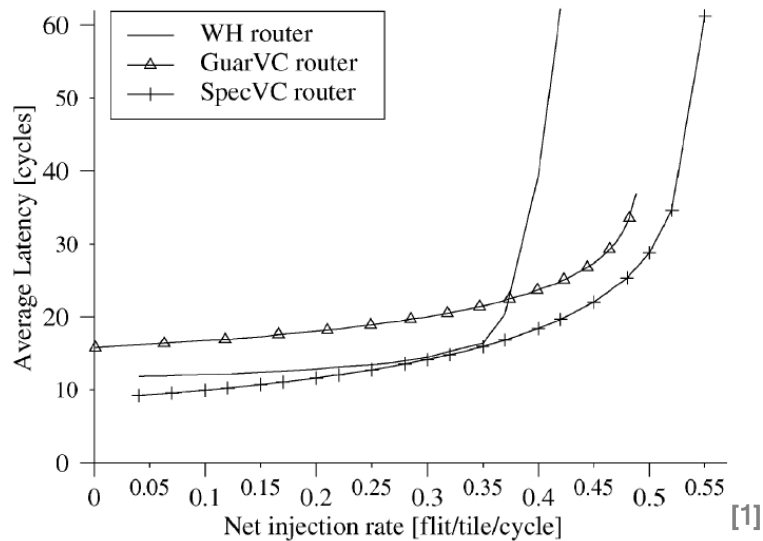
Gleichmäßig verteilter Traffic



- GuarVC mit höchstem Delay
 - Einfache, statische VC alloktion
 - VCs für gesamten Pfad an Quelle allokiert
- Sättigung bei VC Router später

Paketlatenz

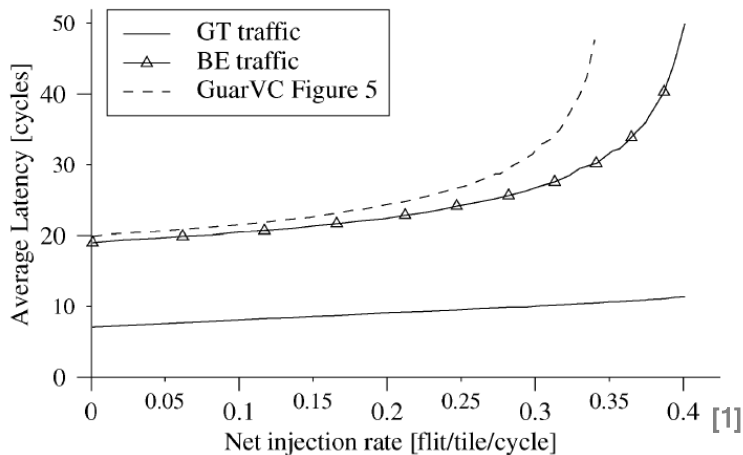
Lokalisierter Traffic



- Allgemein
 - Verringerte Latenz
 - Höherer Sättigungspunkt
- Statische VC Allokation des GuarVC mit negativem Einfluss auf Sättigung

Paketlatenz

Streaming Traffic



- Für ein Paket mit GT gilt:
 - Alle Ressourcen sind pre-allokiert
→ Serialisierungs-Delay und Hop-Distanz verursachen Latenz
- Keine Sättigung für Pakete mit GT
 - Da mind. 50% Linkbandbreite
- Sättigung für Pakete mit BE höher
 - Weniger Flits werden für längere Perioden geblockt

Auswertung

- Auswertung mittels „Energy – Latency – Product“

$$\text{IELP}_{\text{sum}} = \int_0^{r_{\text{sat}}} \frac{1}{E(r) \cdot L(r)} dr$$

Network	IELP_{sum} [×10⁷] (excl. links)		
	Uniform random	Local random	Streaming
WH	8.9 (19.1)	21.6 (45.5)	–
GuarVC – BE	6.8 (12.4)	12.8 (23.4)	8.5 (15.5)
– GT	–	–	29.7 (55.9)
SpecVC	9.6 (15.0)	19.4 (30.2)	–

[1]

Auswertung

- WH und SpecVC ungefähr gleichauf
- Zusätzlicher Energiebedarf des SpecVC durch erhöhte Leistung ausgeglichen
- Spezialisierung des GuarVC liefert hohe Effizienz für GT
 - Aber abgeschlagen bei BE
- SpecVC als favorisierte Wahl gegenüber WH
 - Höchste Leistungsfähigkeit
 - Keine Reduktion des Energieeffizienz

Quellen

[1] Banerjee et al.:

An energy and performance exploration of network-on-chip architectures, IEEE Trans. Very Large Scale Integr. Syst., 2009. Bandnummer: 17, Nummer: 3, Seiten: 319–329.

[2] Robert Mullins : The Design and Implementation of a Low-Latency On-Chip Network

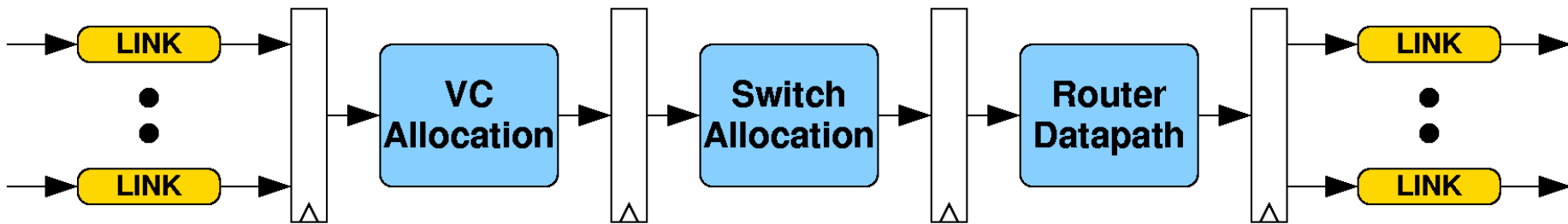
**Vielen Dank
für ihre Aufmerksamkeit!**



»Wissen schafft Brücken.«

Anhang

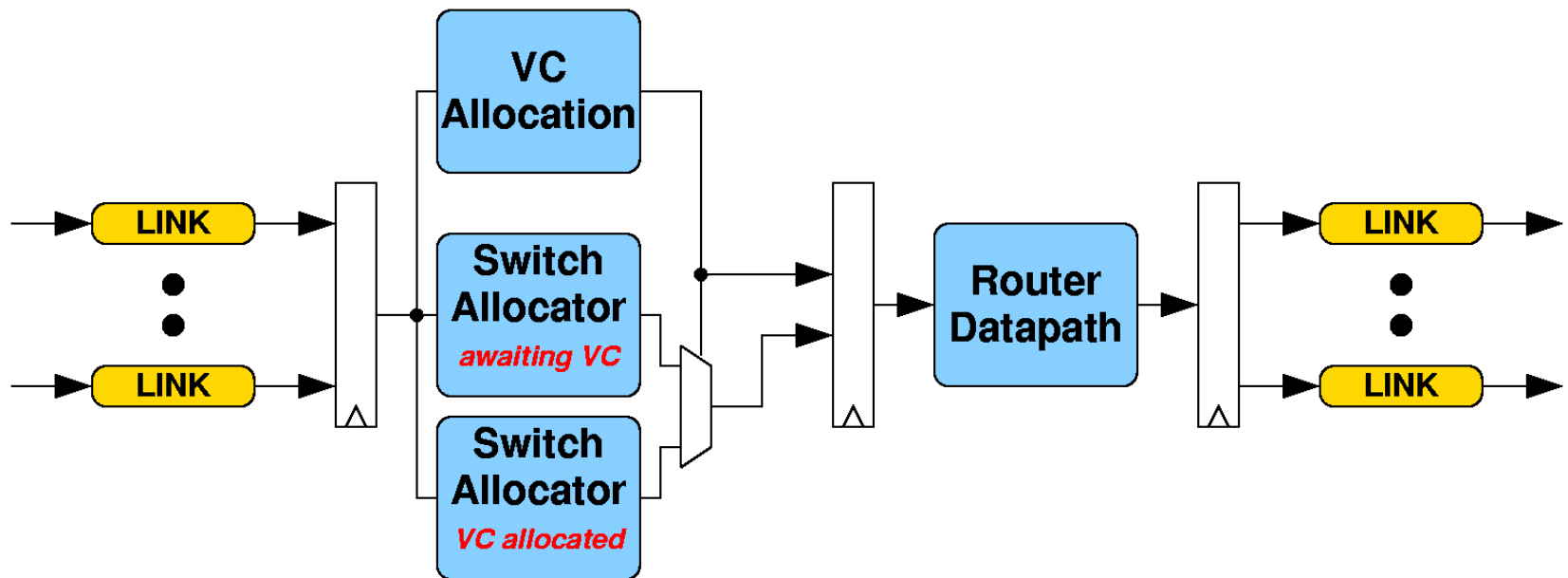
Synchrone Pipeline



[2]

Anhang

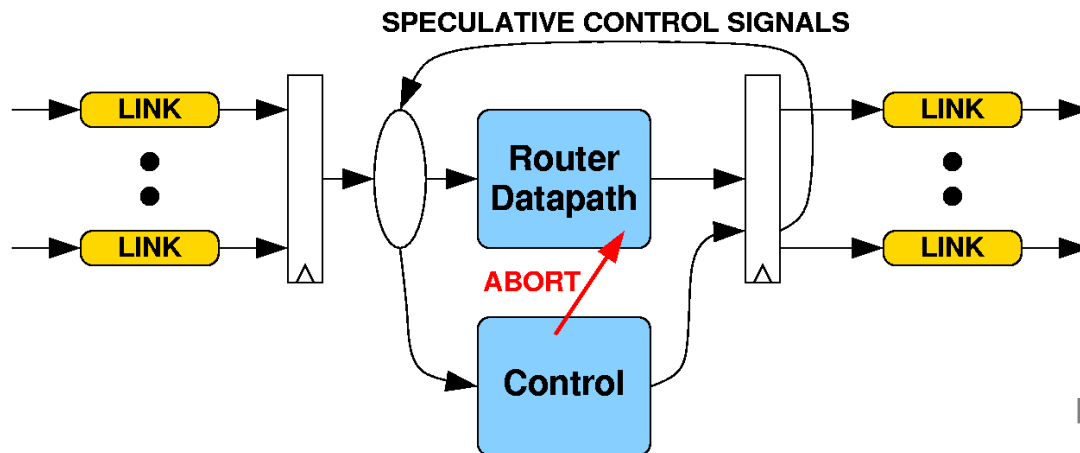
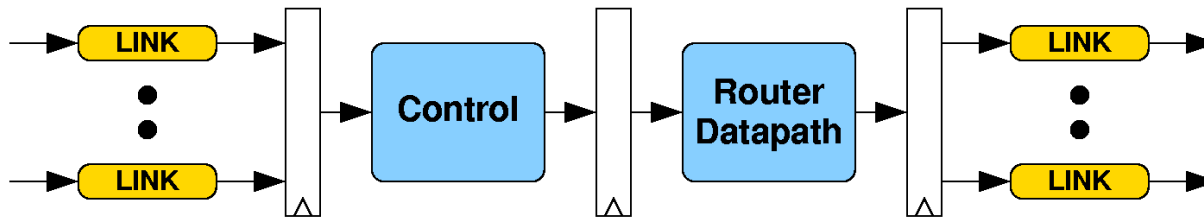
Speculative Router



[2]

Anhang

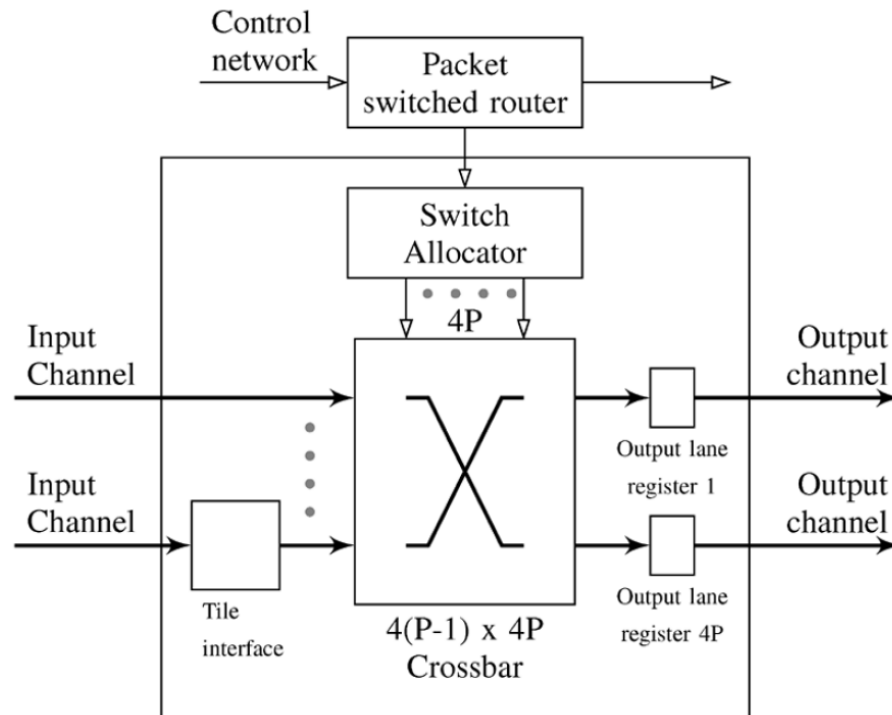
Single Cycle Speculative Router



[2]

Anhang

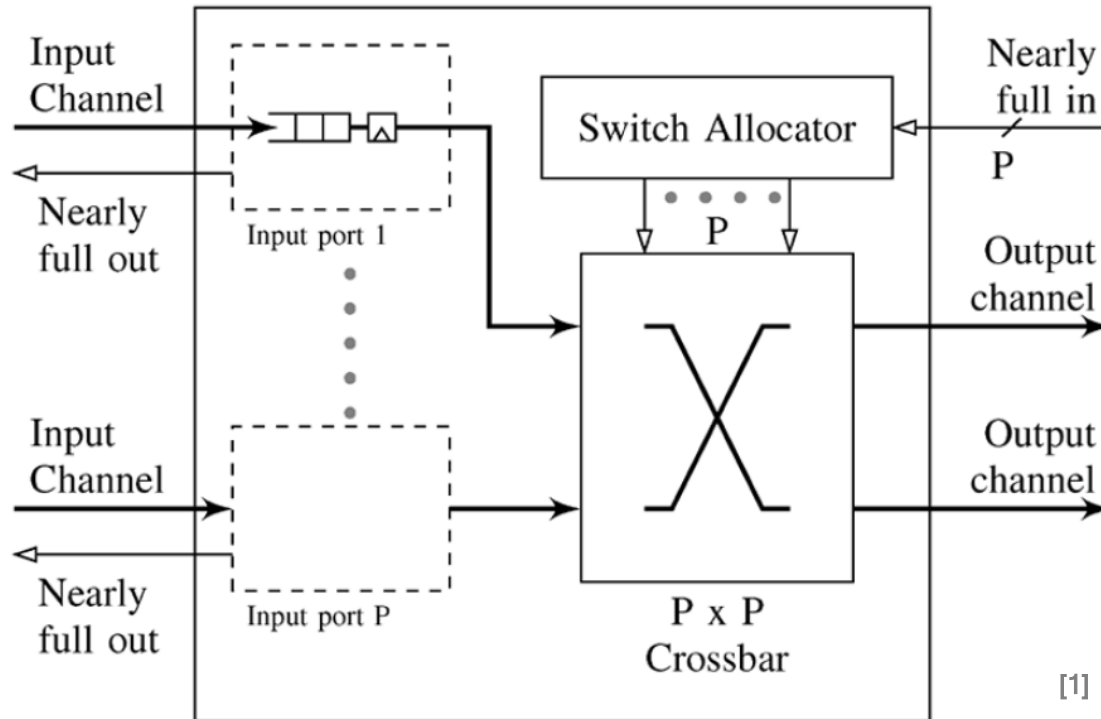
CS Router



[1]

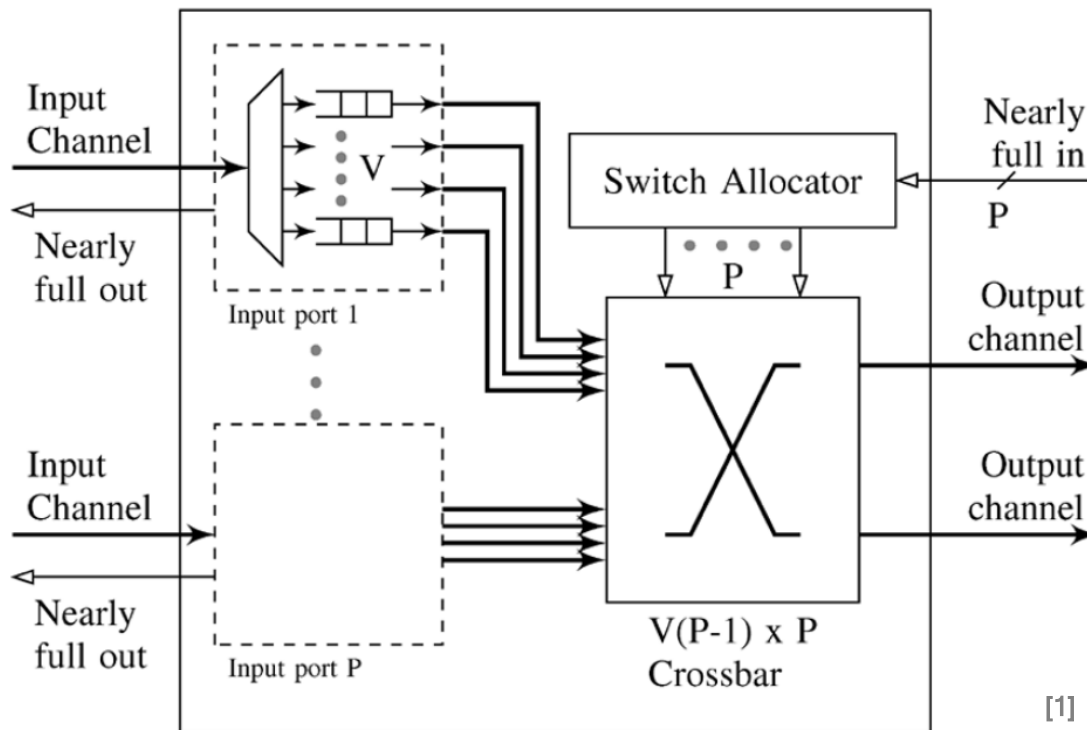
Anhang

WH Router



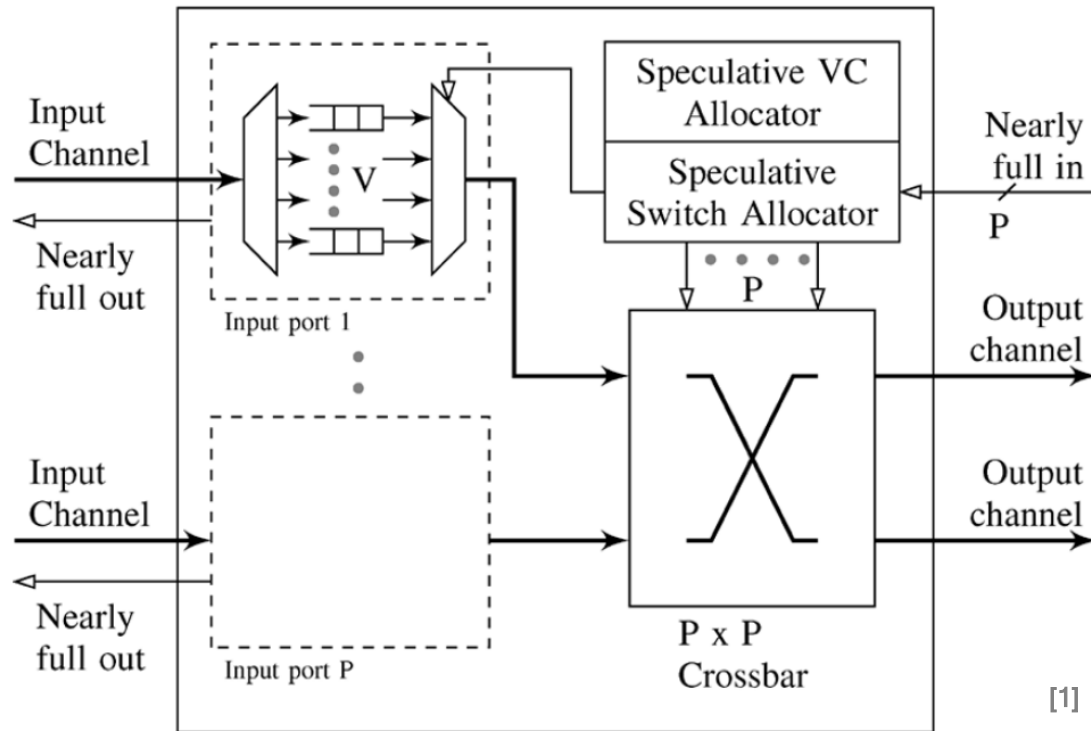
Anhang

GuarVC Router



Anhang

SpecVC Router



[1]