20 November 1997

Public domain Crossbarkommunikationssystem

Die GMD-FIRST hat für das Gebiet der konfigurierbaren, massiv-parallelen Rechner ein Hochleistungskommunikationssystem entwickelt, das auch in anderen Bereichen mit hohen Kommunikationsanforderungen einsetzbar ist. Das gesamte System und seine Entwicklungsunterlagen stellen wir jetzt als public domain Hardware zur Verfügung. Dieses Kommunikationssystem kann sehr einfach bei verschiedenen Sorten digitaler Kommunikation verwendet werden. Es erlaubt, die Zahl der Anschlüsse über einen sehr großen Bereich zu wählen, von 16 als kleinster Wert bis zu mehreren Tausenden [http://www.first.gmd.de/~sergio/public.html].

Die Grundbausteine dieses Kommunikationssystems sind ein 16x16- byte paralleler Crossbar mit einer Übertragungsbandbreite von 2.24 Gbyte/sec und ein universeller Link Interface Device. Beide wurden als ASIC implementiert.

An jeden Anschluß kann ein Knotenrechner, ein I/O-Device oder ein Interface zu einem Standard Bus angeschlossen werden. Die Link-Anschlüsse sind bidirektional und vollduplex. Das Netzwerk hat einen integrierten Router. Kleine Systeme (mit 16 oder weniger Anschlüssen) haben eine Routing-Zeit von weniger als einer viertel Mikrosekunde. Große Systeme mit mehreren Tausend Knoten haben eine Routing-Zeit von 2 bis 3 Mikrosekunden. Im Netz können viele Verbindungen gleichzeitig existieren, ohne sich gegenseitig zu stören oder Bandbreite zu stehlen.

Das Kommunikationssytem wurde bei GMD-FIRST bereits in meheren Projekten erfolgreich eingesetzt. Es fand als erster Einsatz in das MANNA-System. MANNA (1992) ist ein sehr flexibles, leistungsfähiges und skalierbares MIMD-Testbett für massiv parallele Applikationen. Es ist ein verteiltes Rechnersystem mit verteiltem Speicher. Seine extrem hohe Flexibilität erlaubt dem Benutzer, viele verschiedene Kommunikations- und Programmiermodelle zu testen, ohne dabei auf Leistung verzichten zu müssen. Durch das Kommunikationssystem konnte einen linearen Speed up erreicht werden.

Das neue Projekt Power-MANNA [http://www.first.gmd.de/PowerMANNA] (1997) basiert ebenfals auf demselben Kommunikatonssystem wie MANNA, seine Knotenrechner basieren auf einem der schnellsten 64-Bit Prozessoren (PowerPC 620). Ein Knotenrechner besteht aus zwei PowerPC 620 Prozessoren, die mit einem innovativen Speicher-Interface gekoppelt sind. Die Knotenrechner implementieren ein Pipelined Speicher-Interface mit einer Übertragungsrate von 640 MBytes/s. Dieses Interface wurde auch als ASIC implementiert.

Um verschiedene Einheiten problemlos miteinander verbinden zu können, wurde ein universeller Verbindungskanal (“LINK”) geschaffen und für diesen ein “LINK-Protokoll” definiert. Dieses Protokoll wird nicht nur benutzt, um die verschiedenen Einheiten mit dem Netzwerk zu verbinden, sondern auch, um netzwerkintern die verschiedenen Netzkomponenten zu koppeln. Auf diese Weise ist es möglich, jede Einheit unmittelbar mit jeder anderen zu verbinden.

Crossbar

Der Grundbaustein des Verbindungsnetzwerkes ist der byte parallele16x16 Crossbar.

Jara, Montenegro & Mousa; GMD-FIRST, Rudower Chaussee 5;
D-12489 Berlin-Adlershof
Tel +49/30/6392-1878, email: sergio@first.gmd.de

Er wude als ASIC realisiert, unter Einhaltung des defenierten LINK-Protokolls. Die Steuerung der Kanäle ist detzentral, alle Kanäle können simultan activ sein, so eine Übertragungsbandbreite von 2.24 GByte/sec erreichbar ist. Die FIFOs in den Eingangskanälen sind 32 Bytes (32x9-Bit) tief und soren für eine “weiche” Datenflußkontrolle.

Die detzentralen Routers des Crossbars sorgen für die automatische logische Verbindungen zwischen den Eingangs- und Ausgangskanälen des Crossbars. Im kollisionsfreien Fall benötigt ein Verbindungsaufbau zwischen 0.2 und 0.25 Mikrosekunden. Eine logische Verbindung wird durch ein *route*-Kommando aufgefordert, gefolgt von vielen Kommandos und Daten und wird durch *close*-Kommando geschlossen. Nachrichten im Netzwerk werden in wormhole routing style transferiertt.

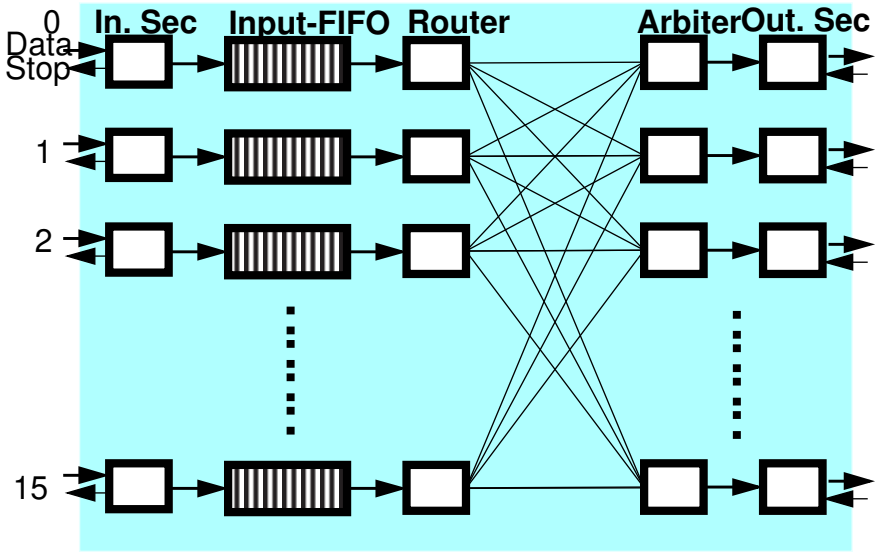

Abbildung 1: 16x16 Crossbar

Link Interface

Um einen Anschluß einer 64-Bit Einheit (z.B. Knotenrechner) zu ermöglichen, wurde ein LINK-Interface-Chip entwickelt. Außer der Protokoll-Konvertierung und dem Datentransfer führt das LINK-Interface zur Erhöhung der Zuverlässigkeit der Nachrichten auch eine CRC-Checksummen-Generierung und Prüfung durch. Das Link-Interface Device hat einen Sende- und einen Empfangskanal (Abbildung), die voneinander unabhängig und simultan arbeiten können. Jede Richtung hat einen FIFO-Puffer von 32 64-Bit-Wörter (256 Bytes). Die FIFOs werden zur Entkpplung der beiden Seiten des LINK-Interface (CPU und Crossbar) wegen unterschiedlicher Übertragungsbandbreite benutzt. Der Sende-FIFO kann außerdem kleine Nachrichten (bis zu 256 Bytes) zwischenspeichern, so daß der Sendeprozeß nicht auf den Verbindungsaufbau zum Empfänger warten muß. Die Sende- und Empfangsrichtung kann alternierend benutzt werden, d.h alternierend ein Wort senden und ein Wort empfangen.

Jara, Montenegro & Mousa; GMD-FIRST, Rudower Chaussee 5;
D-12489 Berlin-Adlershof
Tel +49/30/6392-1878, email: sergio@first.gmd.de

Die Übertragungsbandbreite auf der 64 Bit-Seite beträgt 400 MBytes/s, wobei sie auf der 8-Bit -Seite 60 Mbyte/sec je Richtung beträgt.

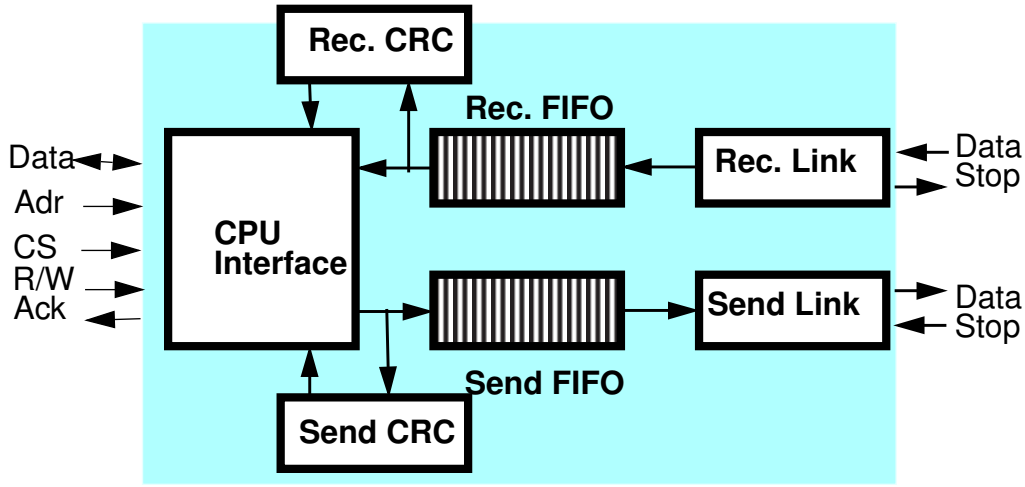

Abbildung 2: Link Interface

Topologie

Die effektive Leistung eines Parallelrechners hängt wesentlich von der Leistungsfähigkeit des Kommunikationssystems ab. Deshalb wurden in der Vergangenheit oft komplexe und teure Verbindungstopologien, wie z.B. der Hypercube, benutzt. Für massiv-parallele Systeme ist eine solche Topologie wirtschaflich kaum mehr realisierbar. Deshalb konzentrierten sich unsere Arbeiten auf die Entwicklung neuartiger Topologien, um gleich gute Verbindungseigenschaften wie beim Hypercube mit einfacheren Strukturen zu erreichen.

Um die Komplexität zu senken, wurden so viele Verbindungen wie möglich integriert. Die erste Integrationsstufe ist der 16x16 Ein-Chip-Crossbar, der 265 Verbindungen ersetzt da VLSI-Elemente billiger, kleiner, schneller und sicherer als Kabel sind. Kleine Systeme mit bis zu 16 Knoten lassen sich auf einfache Weise realisieren, indem man die Knoten über einen solchen Crossbar zu einem Cluster zusammenschaltet. Die zweite Verbindungssstufe kann durch eine backplane mit 2 oder 4 Crossbars und 20 oder 40 Anschlüsse realisiert werden.

Durch Zusammenschaltung mehrerer Crossbars (Hierarchie) können beliebig große Systeme realisiert werden.

Backplane

Eine Backplane aus zwei Crossbars mit 16 synchrone Anschlüsse und 16 asynchrone ECL-Strecken ist Verfügbar und wird auch als Public Domain Hardware angeboten (Siehe Abbildung 3).

Jara, Montenegro & Mousa; GMD-FIRST, Rudower Chaussee 5;
D-12489 Berlin-Adlershof
Tel +49/30/6392-1878, email: sergio@first.gmd.de

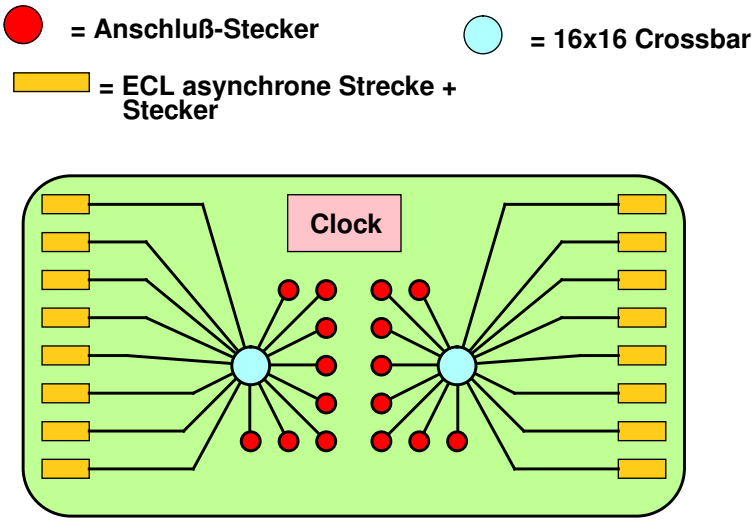

Abbildung 3: Backplane

Public Domain Hardware

Die benötigten Entwicklungsunterlagen sind bei der URL
http://www.first.gmd.de/~sergio/public_domain/
erhältlich. Für die Rechte über Chips und Platinen, sowie Schaltpläne und andere Entwicklungsunterlagen, bitte kontaktieren Sie uns:

Sergio Montenegro	Fihmi Mousa	Elizardo Jara
sergio@first.gmd.de	fihmi@first.gmd.de	jara@first.gmd.de
Tel 030/6392-1878	Tel 030/6392-1883	tel 030/6392-1884