1B－1

SSS型MINアーキテクチャを用いたマルチプロセッサSNAILの実装*

寺田 純　笹原 正司　安川 英樹　塙 敏博　天野 英晴†

慶應義塾大学 理工学部‡

1 はじめに

多段結合網 (Multistage Interconnection Network: MIN) は中規模 (数十から数百プロセッサ) の並列計算機におけるプロセッサとメモリ間の接続に多く用いられる結合方式である. しかし、従来型のMINは, ビットパラレル転送であるため,LSI実装時にピンネックとなり, 高密度実装が困難である. また, パケット衝突時の格納バッファ制御やスイッチングエレメント相互のフロー制御等の複雑な動作を伴うため, 高速化も困難である.

これらの問題点の解決のため, 我々は, 単純なエレメントから構成され, 高い通過率を持つ結合網により, ビットシリアルに同期して入力されたパケットを交換する方式,SSS(Simple Serial Synchronized) 型MINを提案しており [4][5], このSSS型MINのプロトタイプチップを実装した, 16プロセッサからなるマルチプロセッサプロトタイプSNAIL(SSS Network Architecture ImpLementation) を実装中で, 既に12プロセッサの構成で稼働している. 今回はこのSNAILのハードウェア構成について説明する.

2 SNAILの構成

2.1 プロトタイプの概観

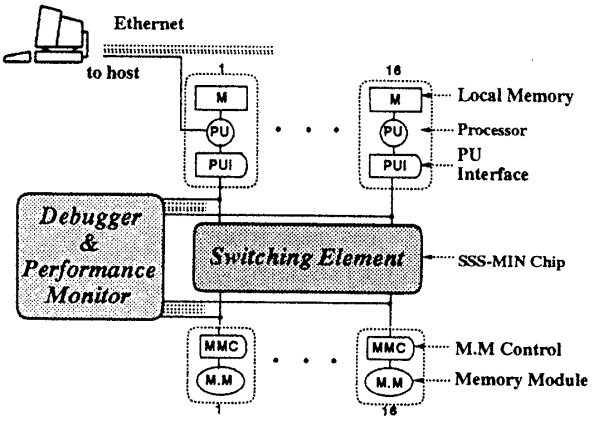

図 1: SNAILの構成

SNAILの構成は, 典型的なMIN結合型のマルチプロセッサ [1][2] であり, 図1に示すように, プロセッサ (MC68040) とローカルメモリからなる16個のPUがSSS型MINを介してメモリモジュール16個と接続されている. プロセッサのうち4個に1個はEthernetインタフェースを持ち, ホストのワークステーションと結合されている. モニタ／デバッグ用ハードウェアは, 別に簡単な16bit CPU(MC68000) を持ち, MINの両側を監視しながら, 本体とは独立に動作する. 全システムは, 単一のシステムクロック (50MHz) と, フレームクロック (50/16=3.125MHz) を共有し, 同期動作する. さらに,SNAILは絶対性能よりも柔軟性を重視した設計になっており,SSS型MIN本体以外はFPGA(Xilinx XC3090:トグル動作 125MHz) で構成されている. 従って, 構成をある程度変更して実験を行なうことが可能である.

2.2 共有メモリのアクセス

プロセッサの共有メモリへのアクセスは, パケットの形でフレーム信号に同期してMINに対して入力される. 動作の高速化のため, 書き込み及びPrefetchにはパケットバッファが用意されており, バッファに空きがあればノンブロッキングでアクセスできるようになっている.

共有メモリは,4ワード (128byte) 単位に16モジュールにインタリーブされており, 以下に示すアクセスが可能である.

- 読み出し: プロセッサは読み出しを完了するまで, 待ち状態になる. Combine可能である.
- Prefetch: プロセッサはパケットバッファ(2パケット分) が空いていれば, 次の命令に進むことができる. データが必要になった時, プリフェッチバッファ (2ワード) に対し読み出しを行なう. この時データがまだ到着していなければ, 到着するまでプロセッサは待ち状態になる.Combine可能である.
- 書き込み: プロセッサはパケットバッファ(2ワード) が空いていれば, 次の命令に進むことができる.
- ブロック転送: プロセッサのオンチップキャッシュの1ブロック (4ワード) 分を高速に読み込む. ローカルメモリと共有メモリ間のデータのブロック転送に有効である.
- 同期操作: Test&SetとFetch&Decが可能である.Test&SetはSSS型MIN内でCombineされるが,Fetch&Decはメモリインタフェース内でのみ行なわれ, Combineされない.

いずれの場合も, アクセスを発生したフレームをiとすると, i+1のフレーム信号に同期してアドレスパケットの転送が行なわれ, プロセッサからメモリモジュールへの経路 (SSS型MINでは**トレース**と呼ぶ) が設定される. データの転送は, このトレースに従ってi+2フレームで行なわれ, 32bitデータの場合, その前半で読み出し可能となる. パケットバッファにパケットが入っている場合, これらの操作は全てフレーム単位でパイプライン的に実行される (パイプライン化サーキットスイッチン

*An Implementation of a Multiprocessor SNAIL based on the SSS-Network Architecture

†Jun TERADA, Masashi SASAHARA, Hideki YASUKAWA, Toshihiro HANAWA, Hideharu AMANO

‡Keio University

グ). MIN 内で衝突が発生すると,転送が成功するフレームまでアクセスが引き延ばされる.

同一メモリアドレスに対し, 読み出し／書き込み, Test&Set, Fetch&Dec, Prefetch の 4 つのアドレス空間が割り付けらており, 上記の操作はアドレスによって識別される. ブロック転送は共有メモリに対し, オンチップデータキャッシュを ON にし, その領域に対して読み出しを行なうことにより起動される.

SNAIL では, フレーム周期は CPU のクロック周期の 8 倍になり, 最大周波数動作時の共有メモリのアクセス時間の最小値は表 1に示すようになる.

表 1: 共有メモリのアクセス時間

フレーム時間	320nsec
32bit data 読み出し/同期	380nsec
1 ラインブロック転送	800nsec

2.3 各部の構成

各部の構成は以下の通りである.

2.3.1 プロセッサ部

MC68040(25MHz) を CPU として用い, ローカルメモリは SRAM で 512Kbyte, ウエイトなしにアクセス可能である. プロセッサの周辺回路と SSS 型 MIN に対するインタフェースは Xilinx 社の FPGA (XC3090) 2 個で構成されている. このインタフェースはプロセッサ–MIN 間のパラレル／シリアル変換, 入力パケットバッファ2,Prefetch バッファ2 ワードを含む.

2.3.2 ネットワーク部

SSS 型 MIN は MIN 内にバッファを持たないため, 衝突による性能低下が大きい. このため, われわれは, 同一宛先への複数パケットの通過を許す MIN に着目し, オメガネットワークを直列に複数接続した TBSF(Tandem Banyan Switching Fabrics), 三次元的に構成する PBSF(Piled Banyan Switching Fabrics)[5] について検討を行なってきた.

SNAIL では, 比較的構成の簡単な TBSF 用のチップを実装し, これを利用した. このチップは,16 入出力, エレメント数 32, クロック 50MHz で動作する[1]. アドレス 3bit, データ 2bit (双方向) パラレルで各入力当たりの最大転送容量 250Mbit/sec, 全体で 500Mbyte/sec である. 詳細な仕様を表 2に示す. 16 入出力,32 エレメントが実装されているにもかかわらず, 全体として 9838 セル (簡単なゲートならば 1 ゲート 1 セル, フリップフロップ等は 4 セルに当たる) で収まっており、1 エレメントが約 200 セルで実現されている. また、セル利用の内訳を分析するとメッセージ結合に要するハードウェアは全体の 20%に過ぎない. 実装されたプロトタイプは複数 Omega 網での実装は行なっていない代わりに 4 チップをパラレルに接続してバンド幅を広げている.

[1]このチップは川崎製鉄株式会社の協力のもとに作成された.

表 2: SSS-TBSF チップの仕様

Max bandwidth	$250Mbits/sec \times 16$ (50MHz)
Bit width	3bit(address)2bit(data)/channel
テクノロジ	$1.0\mu m$ CMOS sea-of-gates
セル使用数	9838 (利用率:49%)

2.3.3 共有メモリ部

共有メモリは各モジュールは 4Mbyte の DRAM で構成され, 全体で 64Mbyte であり, Xilinx 社の FPGA1 個を用いたメモリインタフェースによって制御される. 前述のようにキャッシュブロックに合わせた 4 ワード (16byte) 単位にインタリーブされるが, この構成は FPGA のコンフィグレーションの変更により, 任意のサイズに変更することができる. メモリコントローラは,MIN とメモリ間のパラレル／シリアル変換を行ない, 加算器により Fetch&Dec を実現すると共に,DRAM の制御, リフレッシュ等も行なう. リフレッシュは通常は使用されないフレームを利用して行なわれ, メモリの連続利用のためにこれが不可能な場合は, 全体の動きを 1 フレーム中断して行なう. このため, リフレッシュにより, プロセッサの動作のタイミングが狂うことはない.

3 現状

現在 SNAIL は 16 プロセッサシステムが構築されており,12 プロセッサが稼働中で, ネットワークのパフォーマンス, システム全体のパフォーマンス等の評価が行なわれている.

参考文献

[1] Gottlieb, A. Grishman, R. Kruskal, C.P. Mcauliffe, K.P. Rudolf, L. and Snir, M. : *The NYU Ultracomputer - Designing an MIMD Shared Memory Parallel Computer*, IEEE Trans.on Comput. vol. c-32, No.2, (1983).

[2] Phister, G.F. et al.: *The IBM Research Parallel Processor Prototype (RP3): Introduction and architecture*, Proc. of 1985 Int. Conf. Parallel Processing, (1985).

[3] Konicek, J. et al,: *The Organization of the Cedar System*, Proc. of 1991 Int. Conf. Parallel Processing, (1991).

[4] Amano, H. Zhou, L. Gaye, K.:*SSS (Simple Serial Synchoronized)-MIN: A novel multi stage interconnection architecture for multiprocessors*, Proc. of IFIP Congress 92, Sept. pp. 571-577, (1992).

[5] 天野, 周, 藤川: SSS (Simple Serial Synchronized) 型マルチステージネットワーク, 情報処理学会論文誌 Vol.34, No.5, pp. 1134-1143, (1993).