

6D-3

SSS-MIN 用 LSI チップの実装*

周 洛、天野英晴†

慶応義塾大学理工学部‡

1 はじめに

SSS(Simple Serial Synchronized)-MIN[1]は、スイッチングエレメントの構造が単純で、シリアル転送によりピンネックが解消されているため、1チップ内に多数のエレメントを実装することが可能である。また、従来のMINでは実装が困難であった Message Combining 機能を簡単に組み込むことができる。ここでは、試作が終了した16入出力(32エレメント)を持つSSS-MIN用LSIについて報告する。

2 チップの構成

試作したチップは、16入出力を持つbanyan網で、直列に接続することにより、TBSF(Tandem Banyan Switching Fabrics)網を構成することができる。1.0 μ mのCMOS sea-of-gatesを用いており、安定動作のため、クロックはやや低めの50MHzに設定されている。このことによる動作速度の遅れを補うため部分的にビットパラレル化(アドレスは3bit、データは2bit)が行われ、各入力当たり最大250Mbit/secの転送容量を実現している。

2.1 エレメントの構成と基本動作

図1に各スイッチングエレメントの構成を示す。エレメントはアドレス、応答信号(ACK/NAK)、データ(双方向)の3つのブロックから構成されており、各部には専用の信号線とマルチプレクサが用意されている。先に述べたように試作チップではアドレスは3bit、データは2bitとなっている。

SSS-MINにおいて、スイッチの状態はアドレスパケット内のルーティングタグにより決定される。2つの入力パケットが同一の出口に向かった場合、衝突が発生し、希望の方向に進めなかったパケットの conflict bit がセットされる。以降、この bit がセットされたパケットは他のパケットの進行を妨害しない。アドレスパケットの先頭がMINの出口(メモリ側)に達した時、すべてのエレメントの状態が決定し、入出力間に論理的なパスが設定される。このパスをトレースと呼ぶ。応答信号、入出力データはすべてこのトレースに沿って送られる。このため、図1のようにSSS-MINのエレメントはアドレスパケット転送路に対してのみコントローラが用意され、他の部のマルチプレクサは全てこの信号により設定される。このことにより、複雑なハードウェアを要するコントローラが1つで済み、ハードウェア量が大幅に簡単化されている。応答信号はアドレスパケットの先頭がMINの出口(メモリ側)に達した時に、即座にプロセッサに戻される。一

方、データパケットの転送は次のフレームのアドレスパケットの転送とオーバーラップされる。この動作をパイプライン化サーキットスイッチングと呼ぶ。

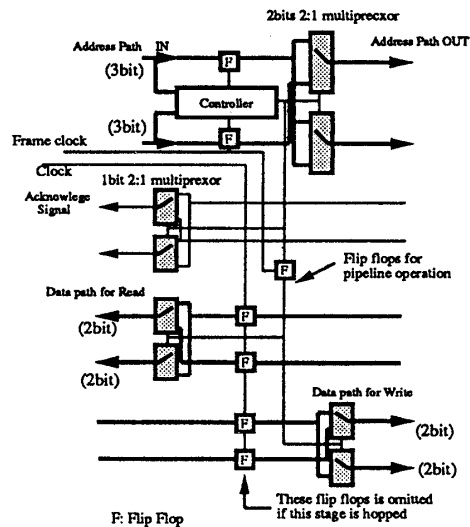


図1. スイッチングエレメントの構成

2.2 トップビットルーティング

大多数のMINではスイッチの状態を決めるヘッダには行き先の番号が直接用いられ、各ステージでは自分の段数に相当する bit をチェックして状態を決める。SSS-MINでは各エレメントでの遅延を減らすため、各段で状態決定するのに用いた bit をけずっていき、常に先頭の bit のみで状態を決定していく。試作チップでは性能の向上のため、アドレス転送線路は3bit化されている。このうち1bitは行き先の番号、残りの2bitは conflict bit, priority bit に用いられ、1クロックでエレメントの状態を決定する。

2.3 ステージの飛び越し

アドレスパケットのヘッダにより一度トレースが設定されてしまえば、アドレスパケットの残りも、データパケットはすでに設定された経路をたどっていけば良い。このため、各ステージで記憶を行う必要はなく、試作チップでは4ステージ(つまりチップ全体)を飛び越して転送を行う。このことにより、網通過のための遅延時間を大幅に低減できる。

*An implementation of an LSI chip for SSS-MIN

†Luo ZHOU, Hideharu AMANO

‡Keio University

2.4 Message Combining 機能

Message Combining 機能は同一宛先のパケットをエレメント内で結合し、1つのパケットとして扱うことにより、Hot stop に対するアクセス集中に伴う Tree saturation を回避する方法である。この機能の効果はシミュレーション等で確認されているが、従来型の MIN でこの機能を実現するためには6倍から32倍のハードウェアを要する[2]ため、今までに実現された例はない。

しかし、SSS-MIN ではシリアル転送の特性を生かして簡単に実現できる。図2に読みだし同士の結合の実現例を示す。各エレメントはアドレスパケット通過時に入力された2つのパケットの全アドレスを比較する。そして、両方のアドレスが完全に等しければ、データ転送路のマルチプレクサを図2に示すような Backward Broadcasting モード (B) にセットし、全体として Tree 状のトレースを形成する。読み出されたデータはこのトレースにしたがってマルチキャストされ、同時に各プロセッサに到着する。この方法では、比較とマルチキャストがビットシリアルに行われるため、付加回路が簡単で、高速性を害することもない。

試作チップでは Test&Set の結合も実現されている。図2同様にアドレス通過時に各エレメントは Tree 状のトレースが形成されるが、データの転送時片方のデータ転送路からは強制的に H レベル (all-1) がプロセッサ側に戻される。このため、ただひとつのプロセッサだけが、メモリからのデータを受けとり、残りのプロセッサは all-1 のデータを受けとる。読み出し直後、メモリコントローラによってメモリの値は all-1 とされるので、全体として Test&Set 操作の結合が実現される。

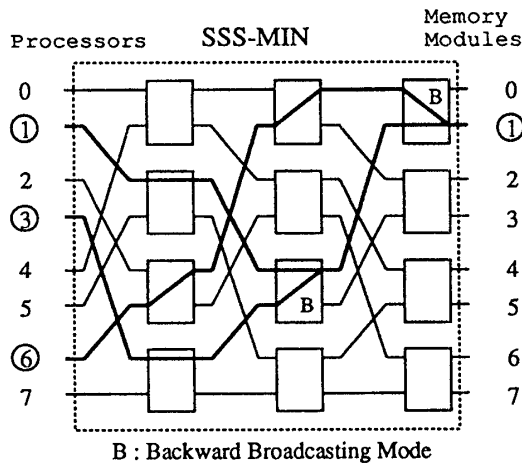


図2. Message Combining の実現

3 試作チップの仕様

今回実装した試作チップは、16入出力、エレメント数32、クロック50MHz、アドレス3bit、データ2bit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。信号ピン数は180、使用セル数は9838、セル利用率49%であった。sea-of-gatesの各セルの配置、配線はCADにより自動的に行われたため、図

3に示すように各セルがチップ全体にちらばったレイアウトになっている。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)で収まっている。エレメント1つは約200セルで実現されている。図4にセル利用の内訳を示す。これによると Message Combining に要するハードウェアは全体の20%に過ぎないことがわかる。このチップは直列に並べることにより、TBSFのbanyan網数の増加により通過率を改善することができ、bit方向に並べることによりビットパラレル化を実現できる。現在、このチップを4個(2個並列、2個直列)用いたMINにより16プロセッサを結合した1ボードマルチプロセッサを設計中である。この構成では、フレーム時間220nsec、メモリ読みだし時間480nsecを達成できる。

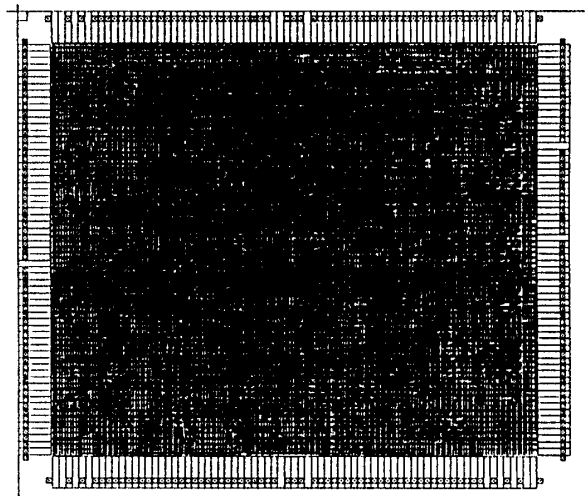


図3. 試作チップのレイアウト

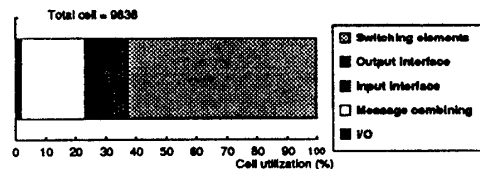


図4. セル利用の内訳

謝辞

チップ実装にご協力いただいた川崎製鉄株式会社 LSI 事業部に感謝します。

参考文献

- [1] 天野、Gaye, "SSS スイッチングアーキテクチャに基づく並列計算機," 信学報 CPSY91-11, Jul. 1991.
- [2] G.F. Phister, and V.A. Norton, "Hot Spot Contention and Combining in Multistage Interconnection Networks," IEEE Trans. on Comput. vol. c-34, No.10, Oct. 1985.