

## FPGA化のためのk-ary n-cube型相互結合網用のルータの設計

佐野雅彦 Andrew C. Flavell 杉原泰次郎  
岡田智士 栗栖一洋 高橋義造

徳島大学工学部知能情報工学科

770 徳島市南常三島町2-1

TEL 0886-56-9445 , FAX 0886-56-7489

E-mail: { sano , flavell , sugi , okada , chris , taka }@is.tokushima-u.ac.jp

あらまし

我々は、MIMD超並列計算機向きのルータのFPGA化を前提として、k-ary n-cube型ネットワークの実装方法について検討し、試みとして16-ary2-cubeの2次元トーラスネットワークルータをXilinx FPGA XC4013PQ240に実装した。k-ary n-cube型ネットワークルータのアーキテクチャには我々の研究室で提案されたTokkyūルータを採用している。このルータは特定方向のパケットをバイパスするPacket Express Wayにより低遅延のルーティングが可能なアーキテクチャを持つ。

本報告では、FPGAへ実装方法について検討し、その実装の結果について述べるものである。

キーワード 相互結合網, k-ary n-cubeネットワーク, Tokkyūルータ, FPGA

### Design of the k-ary n-cube interconnection network router for considering FPGA implementaion

Masahiko Sano, Andrew C. Flavell, Taijiro Sugihara,  
Satoshi Okada, Kazuhiro Chrisu, Yoshizo Takahashi

Department Information Science and Intelligent Systems  
Faculty of Engineering, Tokushima University, Tokushima 770 Japan

TEL 0886-56-9445 , FAX 0886-56-7489

E-mail: { sano , flavell , sugi , okada , chris , taka }@is.tokushima-u.ac.jp

Abstract

Aiming the implimentation on FPGA, we have designed a router for k-ary n-cube interconnection network. As a result, a router for 16x16 at bidimensional torus network, Tokkyū router, was implimented on Xilinx FPGA XC4013PQ240. In this router, low routing latency is achieved by using the method of the "packet expressway" for packets to bypass specific direction. This paper discusses the FPGA implementation and presents experimental results.

key words

Interconnection network, k-ary n-cube network, Tokkyū router, FPGA

## 1. はじめに

並列計算機の総合的性能は計算機アーキテクチャと実行されるソフトのアルゴリズムに依し、これまでに多くの方式の提案とその成果が報告されている。特に数値計算、データベース、シミュレーション等の分野での成果は目覚ましい。そして並列処理の応用範囲の拡大により、今後、ますます多様な並列処理アルゴリズムが考案されると推測される。一方ハードウェアにおいてはその実現コストが高く、考案されたアルゴリズムに対するネットワーク特性の評価をネットワーク方式の変更により評価・比較することがコスト的に難しい。しかし近年のFPGAなどの書き換え可能デバイスの高性能化により、現在では10万ゲート相当のFPGAが開発されている。これによって従来では数個～数十個必要であった回路が1～数個に集積可能となり、FPGAによるルータの実現が容易になりつつある。特にSRAM形式のFPGAは不揮発性のFPGAと比較して書き換え可能回数が大幅に多く、頻繁に内部論理を変更する用途に向いている[1,2,3]。

FPGAなどのインサーキットプログラム可能(ICを装着したままプログラム可能) デバイスをルータに用いた場合、次の特徴が挙げられる。

(1) ゲートアレイやカスタムVLSIを用いたルータの実現では、少数個の作成はコストが高く、頻繁にネットワーク構成を変化させて評価を行う場合には再プログラム可能なデバイスを使用の方が低コストで、1個からの作成でき、加えて書き込みのコストも低いため、低コストでネットワークの試作・評価ができる。

(2) 内部回路の再設計により、異なるルーティングアルゴリズムやバッファ数などのパラメータを変化させた評価が可能になり、アプリケーションに対するネットワーク構成の影響を評価できる。

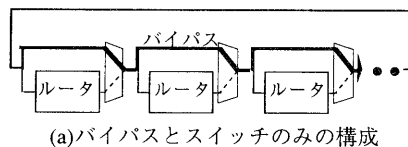
(3) SRAM型の揮発性FPGAを用いれば、アプリケーションに対して効率の高いルータアーキテクチャを実行時に選択できるようになり、アプリケーションに特化したアーキテクチャを用意できる。将来、デバイスの部分再プログラムが可能になれば、複数の特性の異なるアプリケーションの実行に対して動的に変換可能なネットワークが実現できる[3]。

(4) ゲートアレイやカスタムVLSIと比較して利用できるゲート数が少ないが、現在ではゲート数の多いFPGAも開発されつつあり、動作速度の問題も改善されつつある。

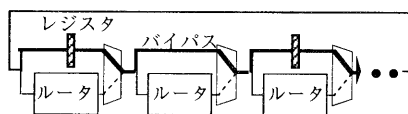
## 2. k-ary n-cube ネットワーク

k-ary n-cube ネットワークにはメッシュ、トーラス、ハイバキューブ等のトポロジが含まれており、多くの並列計算機に用いられている。一方我々の研究室でもk-ary n-cube ネットワークルータとしてTokkyū ルータ

[4,5]が提案された。Tokkyū ルータは、バイパスとスイッチからなるエクスプレッスウエイと呼ばれるバイパス機構を持ち、双方向リング結合のk-ary n-cube ネットワークを想定している。Tokkyū ルータでは、図1(a)に示すように単にスイッチを用いたバイパス機構ではなく、リングの配線経路長とスイッチの遅延からデータクロック周波数の上限の低下を防ぐために、図1(b)に示すようにレジスタを挿入することでネットワークをシフトレジスタ的に動作させる[5]。図2にn=2の場合のルータの構成を示す。



(a) バイパスとスイッチのみの構成



(b) レジスタを挿入した構成

(2つ毎にレジスタを挿入した例)

図1 エクスプレッスウエイの構成

Tokkyū ルータではカットスルー方式によるパケット転送を採用し、進行方向が変化しないパケットをエクスプレッスウエイを用いてバイパスする事により遅延の増加を抑える。エクスプレッスウエイが使用できない場合はアウトプットキューに空きがない場合には、ランダムルーティングが行われる[4,6]。

Tokkyū ルータは図2に示すように各チャネル毎にm ウエイのキューがある。このウエイ数mと各キューの容量をパケットサイズの1倍としてシミュレーションを行った。図3(a)は通信負荷に対するミスルーティング率、図3(b)はスループットに対する遅延、図3(c)は負荷に対するスループットのグラフで、16-ary 2-cube ネットワークに於いて様な通信下のものである。これら結果から、少なくともm=2, l=2であれば十分な性能を維持できると判断した[4]。

## 3. 実装

本研究ではTokkyū ルータをXilinx社のXC4000シリーズ[7]への実装問題を取り上げた。XC4000シリーズは図4に示す構造を持ち、周囲の入出力ブロック(IOB)、アレイ状に配置された機能ブロック(CLB)、配線リソースから構成されており、ルータの実装に対して以下の利点がある。

- ・内蔵RAM CLBの一部をRAMとして使用できる。パケット交換型のルータでは内部に複数のバッファを必

要とするので、このRAMの使用により回路を効率化ができる。

・バスリソース CLB当たり2つの割合で3ステートバッファがあり、水平方向に並ぶバスリソースを用いたバスが構成できる。ルータにはクロスバスイッチやデータセクタが不可欠であるが、バスを用いることにより、機能ブロックを殆ど使用せずに実現できる。

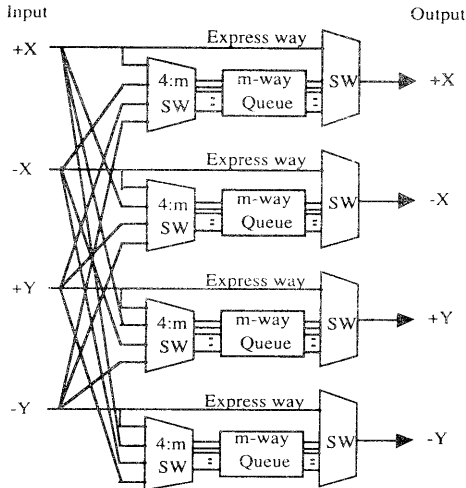


図2 Tokkyūルータの構成 (n=2の場合)

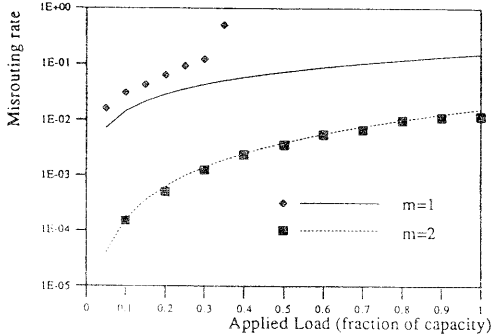


図3 (a) 通信負荷に対するミスルーティング率

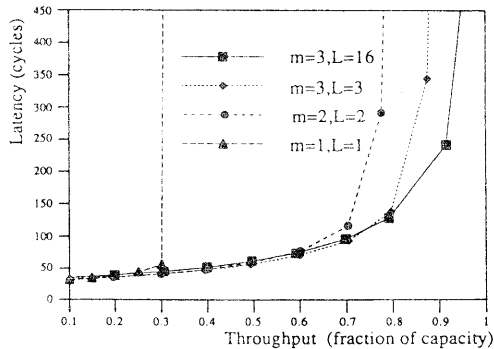


図3 (b) スループットに対する遅延

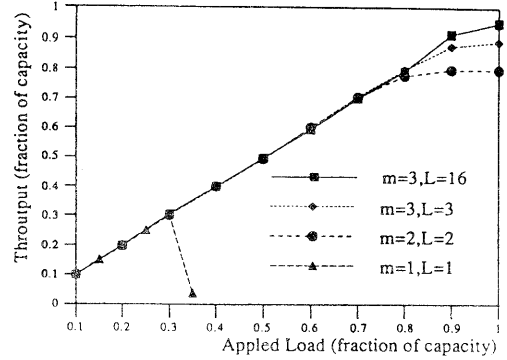


図3 (c) 通信負荷に対するスループット  
図3 シミュレーション結果 (16-ary 2-cube)

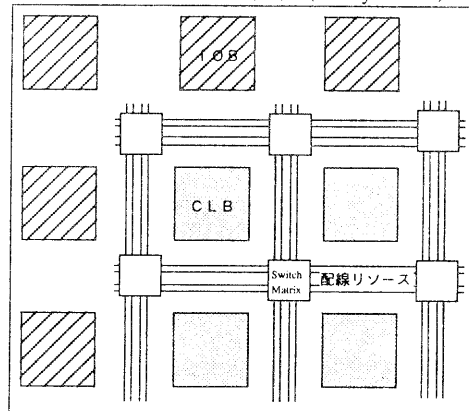


図4 Xilinx FPGA内部構造の概要

### 3.1 方針

前述の特徴を考慮して、本研究では以下の実装方針を採った。

(1) 使用するデバイスはXC4013PQ240

13,000ゲート相当のFPGAであり、現時点で我々の研究室で開発可能な最大のものを選択した。表1にこのFPGAの仕様を示す[7]。

表1 XC4013PQ240の仕様

項目	数値	
ゲート数	13,000 ゲート相当	
I/Oピン数	192	
CLB数	576	
フリップフロップ数	1152	
バスリソース	ロングライン	96
	3ステートバッファ	1248
RAM容量 (ビット)	18,432	

## (2) 16-ary 2-cubeのトラスネットワーク

Tokyoルータの場合次元数  $n$  に対して、少なくとも入出力にそれぞれ  $2n$  個のポートとノードプロセッサとの通信のための  $2$  個の合計  $4n+2$  個のポートが必要である。仮にポート当たりのピン数を  $10$  本（データ幅  $8$  ビット、制御線  $1$  ビット）とすれば、 $n=2$  の場合  $90$  本、 $n=3$  の場合、 $126$  本となり、I/Oピン数  $190$  本程度使用できるFPGAであれば十分である。しかし、今回の実装では回路規模の増加を抑えるために  $n=2$  とした。

(3) 8ビットデータ幅 (1)より  $n=2$  の場合、 $16$  ビット幅とすると、I/Oピンは十分であるが、後述するバスリソースの問題を検討した結果、本研究では  $8$  ビットのデータ幅とした。

(4) バスリソース 図3のシミュレーション結果よりアウトプットキューのウエイ数  $m$  は  $2$  以上が望ましいが、バスリソースを用いた  $4:m$  スイッチを構成するには、表1に示すバスリソース数のロングラインが  $96$  本以下でなければならない。データ幅を  $16$  ビットとすれば  $128$  のリソースが必要であるため、データ幅は  $8$  ビットとなる。この結果、ウエイ数は最大で  $m=3$  となるが、シミュレーション結果から回路増加に見合う性能改善が得られないと判断し、 $m=2$  とした。

(5) パケットサイズ FPGAのRAMの最小構成は  $16$  ワード  $\times$   $1$  ビットなので、これに合わせて  $16$  ワードとした。パケットの先頭の  $1$  ワードが宛先を示し、現在位置からの相対距離を  $X$ 、 $Y$  方向それぞれ  $4$  ビットで表す。残りの  $15$  ワードはデータである。ヘッダの  $X$ 、 $Y$  方向の相対距離が共に  $0$  となったパケットを受けたルータはノードプロセッサにパケットを出力する。

(6) キュー RAMベースのキューとしFPGAの内蔵RAMを用いて効率化を図った。但しデュアルポート化により入出力の同時処理を実現し、パケットの遅延を極力抑えるようにした。

アウトプットキューの容量は、シミュレーションの結果から、パケットサイズの  $2$  倍以上が望ましいが、今回の実装ではパケットサイズと同容量とした（今後増加させる予定）。

(7) グローバルアービタ 上記の方針よりグローバルアービタは  $4:2$  のデータ選択制御回路が  $4$  系統必要となるが、 $4$  系統の並列回路を用いた場合、FPGAに収まらない可能性があるため今回は逐次処理するものとした（最短  $2$  クロック、最悪  $4$  クロック）。これにより入力パケットの調停に最悪  $12$  クロック必要となるため、入力側に  $16$  ワードの容量のキューを設け、遅延を吸収することとした。（トラフィック密度が低い場合、最悪値で  $6$  クロック）

### 3.2 実装

以上の方針に基づいたルータのブロック構成を図4に示す。以下では各ブロックの詳細について述べる。

・エクスプレスウエイ (EW) 入力から出力へのパイパス経路となるので、FPGA内部の横方向に延びる配線リソースを活用し、配線混雑を避けるレイアウトを行った。

・インプットコントローラ (IC) 入力パケットのヘッダ解析後、エクスプレスウエイが使用可能か判定し、可能な場合はエクスプレスにデータを送り、不可能な場合はキューにデータを蓄えたと共にグローバルアービタにリクエストを出す。入力パケットのワード数計測とキューのアドレスカウンタは共通化し、回路の効率化を行った。

・キュー (Q) (6)で述べたように、内蔵RAMの時分割使用によるデュアルポート化により、読み込みと書き込みの同時処理を行った。また、 $2$  ウエイのアウトプットキューでは双方が同時に出力しないので、アドレスカウンタを共用化して全体で約  $20$  CLBの削減を図った。

・スイッチ (SW) (4)で述べたようにバスリソースを用いて  $4:2$  のスイッチを構成した。これによりCLBのみのスイッチ構成の場合と比較して  $56$  CLB（全CLBの  $10\%$ ）の削減を図った。

・グローバルアービタ (GA) (7)で述べたように、各インプットコントローラを順に調べ、キューにパケットの一部でも格納されていればヘッダにより要求されるアウトプットキューの空きを調べ、割り当てを行う。空きがなければ他の空いているチャンネルをランダムに選択して割り当てる。

・アウトプット・コントローラ (OC)

OCではエクスプレスウエイを最優先に、以下、アウトプットキュー、ノードプロセッサからのパケット注入の優先順位で出力選択をする。ただし、そのチャンネルのアウトプットキューが全て塞がっている場合は注入を行わない。これは、注入中に別のルータからパケットが入ってくる可能性があるためである。また、OCではパケットのヘッダを出力方向毎に更新、 $X$ 、 $Y$  のそれぞれの方向のアドレスを同時に更新することはないので、 $4$  ビットアダーを用いて更新し、ヘッダ更新時には  $+1$  または  $-1$  を、データの場合は  $0$  をヘッダに加算することで、データバスを増やさずに実装した。

以上の実装により、エクスプレスウエイ経由のルーティングの遅延が  $3$  クロック。キュー経由では最小  $10$  クロック、トラフィックが少ない場合は平均  $14$  クロック、最悪  $22$  クロックの遅延となった。またパケット同士では、エクスプレスウエイの連続使用時には最小  $2$  クロックの間隔が必要となり、エクスプレスウエイ、キュー、パケット注入の任意のパケット同士間では  $4$  クロックとなった。図5に全体のブロック図を示す。

#### 4. 実装結果・評価

今回用いたFPGAの開発システムはXilinx社のPCベースのXact Step Ver. 6を使用し、全体の80%を回路図で作成、グローバルアービタ、アウトプットコントローラの一部をNTTデータ社製HDL、パルテノンを用いて記述した。また、FPGA内部の配線リソースを有効に活用するためにIOBの80%、CLBの約1/3を手動で配置し、残りを自動配置させた。論理最適化後のリソースの使用率は表2に示す結果となり、実装の目安である必要CLB数は全体の62%となった。表3に各部のCLBの使用数の概算値を示す。図6に配置の結果、図7に配線結果をそれぞれ示す。

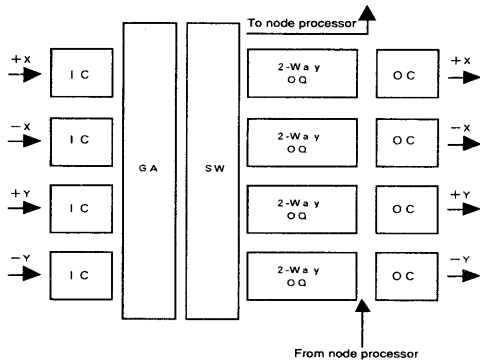


図5 全体のブロック図

表2 最適化後の使用率(\*1:配置後のCLB使用率)

項目	最適化後(割合)	
I/Oピン数	101(53%)	
CLB数	359(62%), 532(92%) *1	
フリップフロップ数	337(29%)	
バスリソース	ロングライン	64(20%)
	3ステートバッファ	256(20%)
RAM容量(ビット)	1,536(8%)	

表3 各ブロックのCLB使用数の概算値

各ブロック	CLB使用数の概算値
インプットコントローラ	22 × 4
グローバルアービタ	32 × 1
アウトプットキュー(2-way)	28 × 4
アウトプットコントローラ	21 × 4
スイッチ	24 × 1
概算値による合計	340

次に速度性能についてのシミュレーションの結果について述べる。シミュレーションは配置配線後のLCAファイル(この後、ビットファイルに変換されてデバ

イスにダウンロード可能になる)を元にして行った。その結果、XC4013PQ240-6(CLB遅延6ns)で14MHz、-3で18MHzの動作が確認された。CLB遅延が半分となったが動作速度は余り向上しなかった。この理由は次のように推測される。すなわち、CLB遅延は半分となったが、配線リソース等の他の部分で半分に短縮されていない部分によりフリップフロップ、およびRMAのセットアップタイム、ホールドタイムが不足し、データのラッチが正確に実行されなくなるためである。このことから、遅延を考慮したレジスタ配置の工夫が必要であると思われる。

#### 5. おわりに

本報告では、我々の研究室で提案されたk-ary n-cube ネットワークルータであるTokkyūルータをXilinx社のFPGA4013PQ240に16-ary 2-cubeトラスネットワークルータとして実装し、配置配線後のシミュレーションにより、CLB遅延6nsのFPGAで14MHz、3nsのもので18MHzの動作が確認された。また実装に当たり、FPGAのもつRAMやバスリソースを活用のほか、各ブロックにおける実装の方針について検討した。実装の結果、回路規模を全体として約2割の削減が可能となった。

今回の実装では13,000ゲート相当のFPGAを使用した。配置前のCLB使用数で62%であった。FPGAでは回路の種類にも依存するが配置を注意して程度行えば80%程度の使用率でも実装可能なため、余裕が残されていると言える。一方、現在出荷されているFPGAは100,000ゲート相当(不揮発性)のものが利用可能になっており、SRAMタイプのFPGAの利用可能ゲート数も今後増加することを考慮すれば、高機能なルータの実装が可能になる。今後の課題は、バッファサイズの増加、ブロードキャストの追加、動作速度の改善などである。

最後に、パルテノンを使用させていただいたNTTデータ社の方々に感謝致します。

#### 参考文献

- [1]富田, 澄川, 菅沼, 平野: 汎用エンジンRM-IIの構成, 情報処理学会論文誌, Vol. 35, No. 4, pp. 639-645 (1994).
- [2]浅, 天野: データ駆動型制御機構付MPLDを用いた並列マシンWASMIIの仮想化, 情報処理学会論文誌, Vol. 35, No. 4, pp. 646-657(1994).
- [3]末吉, 杷野, 有田: 書換え可能なLSIによる可変構造型相互結合網の実現法, 情報処理学会論文誌, Vol 33, No.3, pp. 260-269(1992).
- [4]Flavell, A. C. and Takahashi, Y.: Tokkyū: A High-Performance, Randomizing, Adaptive Message Router with Packet Expressway, IEICE Trans. on Information and Systems, Vol. E78-D, No. 10, pp. 1248-1260(1995).
- [5]Flavell, A. C. and Takahashi, Y.: A Hybrid Time/Space Communications Paradigm for k-ary n-cubes, Proc. of the International Conf. on Parallel Processing 1994, Vol. I, pp. I38-I41 (1994).
- [6]Konstantinidou, S. and Snyder, L.: Chaos router: Architecture and Performance, SIGARCH, Vol. 19, No. 1, pp. 212-221 (1991).
- [7] The XC4000 Data Book, Xilinx, Inc. (1994).

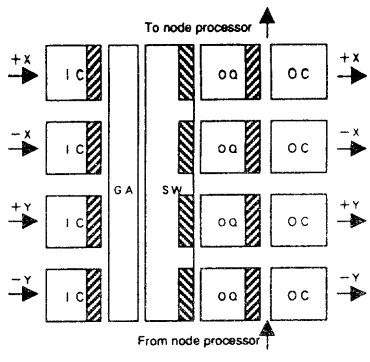


図 6 (b) 配置の概略図

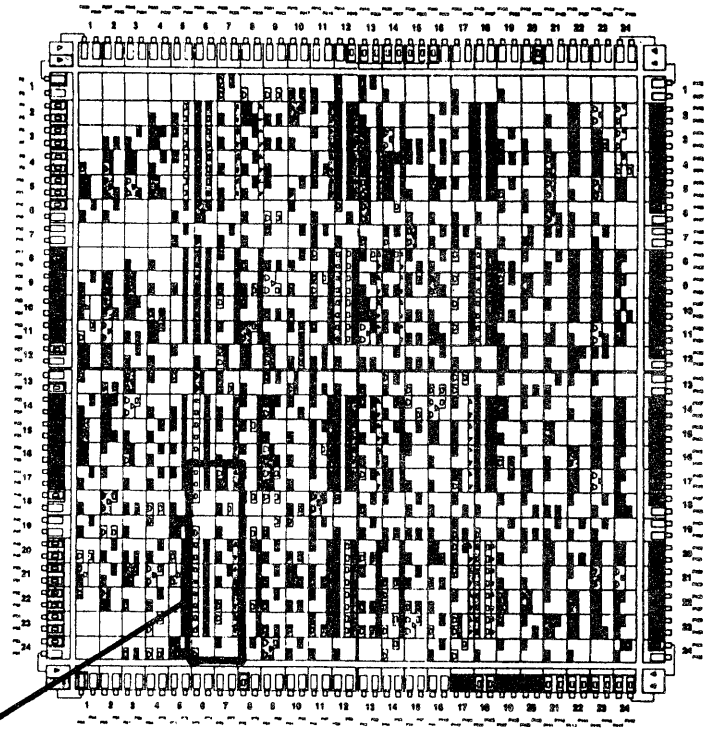


図 6 (a) 配置図

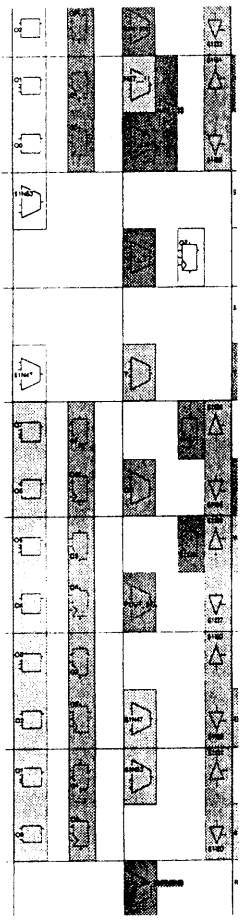


図 6 (c) 配置の部分拡大図

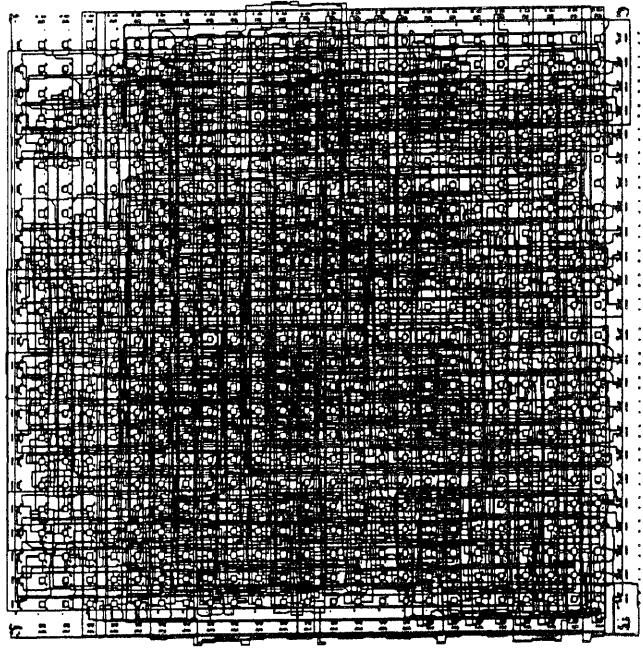


図 7 配線結果