

仮想チャネル数と動作周波数を考慮した適応ルータの性能評価

堀 田 真 貴[†] 吉 永 努^{††}
大 津 金 光[†] 馬 場 敬 信[†]

並列計算機ネットワークの性能向上のためには、採用するルーティングアルゴリズムを検討する必要がある。適応ルーティングを行うこと、また多くの仮想チャネル(VC)を使用することはルーティングの自由度を高め、通信性能を向上させるなどの利点がある。しかし、ルーティングアルゴリズムが複雑になり必要なハードウェア量が増えるため、ルータの動作周波数が低下する。本稿では、2次元トラスネットワークにおいて、このルーティング自由度と動作周波数のトレードオフを考慮し、いくつかのアルゴリズムでのVC数と通信性能の関係を考察する。実験の結果、物理チャネルあたり4本のVCを持つルータが良い性能バランスを示した。また、適応ルータは動作周波数において非適応ルータに劣る傾向があるものの、非ユニフォーム通信に対して高バンド幅、低レイテンシとなる。特にVC数が4本のRecover-xルータは、通信パターンにかかわらず堅牢な性能を示すことが判明した。

Performance Evaluation of Adaptive Routers Based on the Number of Virtual Channels and Operating Frequencies

MAKI HORITA,[†] TSUTOMU YOSHINAGA,^{††} KANEMITSU OOTSU[†]
and TAKANOBU BABA[†]

In order to improve the communication performance of the parallel computer network, we should evaluate the various routing algorithms. Adaptive routing or virtual channels (VCs) can improve communication performance by increasing routing flexibility. However, the operating frequencies of the router become degraded, since the adaptive routing and the VCs require a complex and huge amount of hardware resources. Therefore, it is important to consider the trade-off between the routing flexibility and the operating frequency. We clarify this trade-off by evaluating the communication performance in 2D tori network for typical routers, taking into account the operating frequency of the routing circuits. Our experimental results show that the routers with four VCs per physical channel attain a good trade-off between routing flexibility and operating frequency. Adaptive routers show higher performance than non-adaptive routers due to their higher routing flexibility, especially in the case of a non-uniform communication pattern. The Recover-x router with four VCs per physical channel shows robust performance both in uniform and non-uniform traffic.

1. はじめに

並列計算機においては、プロセッサ間通信の効率がシステム全体の性能に大きく影響する。この通信性能を決定する要因の1つとしてルーティングアルゴリズムがあげられる。ルーティングアルゴリズムによる性能向上の試みとして、種々の適応ルーティング手法が提案されており、その効果を示す研究成果が報告されて

いる^{4),6),8)}。一般に、適応ルーティングではデッドロック防止のために仮想チャネル(VC)を利用する。また、非適応ルーティングにおいても、近年の集積回路技術の発達により物理チャネルあたり4本程度のVCを実装するものが多くなっている⁹⁾。適応ルーティングを行うことや多くのVCを実装することには、共通したプラス面とマイナス面が存在する。すなわち、ルーティング自由度を増大させて通信性能の向上をもたらす半面、経路選択ロジックの複雑化とハードウェア量の増加がチップの動作周波数の低下を招くことが指摘されている³⁾。したがって、並列計算機ネットワークの性能を高めるためには、ルーティング自由度とチップの動作周波数のトレードオフを十分考慮する必要がある。この関係を明らかにするため、ハードウェア記述言語(HDL)

[†] 宇都宮大学工学部情報工学科

Department of Information Science, Faculty of Engineering, Utsunomiya University

^{††} 電気通信大学情報システム学研究所

Graduate School of Information Systems, University of Electro-Communications

を用いて3種類のルーティングアルゴリズムに基づくルータを設計し、評価する。本稿では非適応ルーティングである Dimension-order, デッドロック回避型の適応ルーティングを行う *-channel²⁾, そしてデッドロック回復適応ルーティングを行う Recover-x¹¹⁾を比較した評価結果を示す。Recover-x は、我々が提案する適応ルーティングアルゴリズムである。同様なデッドロック回避機構を持つルータとして代表的なものに DISHA¹⁾がある。Recover-x は DISHA よりもデッドロック回復のオーバーヘッドが小さく、効率の良いデッドロック回復を行う方式である。

まず、論理合成を行うことにより、各アルゴリズムにおける物理チャネルあたりの VC 数を変化させた場合のハードウェア量と動作周波数を評価する。次に、100 ノードのネットワーク性能をシミュレータで評価することにより、各ルーティングアルゴリズム間における VC 数と通信性能の関係を明らかにする。通信性能は、すべてのルータが同一の動作周波数で動作する場合と、それぞれのルータが HDL 設計の論理合成結果から得られた最大動作周波数で動作する場合について示す。実験の結果、各ルーティングアルゴリズムにおいて、物理チャネルあたり4本の VC を持つルータが良い性能バランスを示した。ルーティングアルゴリズムについて分析すると、適応ルータは動作周波数で非適応ルータに劣るものの、ルーティング自由度の高さによって、特に非ユニフォーム通信パターンに対する通信性能が優れることが分かった。

以降、2章では本稿で用いたルーティングアルゴリズム、3章ではルータの設計仕様を示す。4章では、各ルータの論理合成結果を示し、動作周波数と回路規模について考察する。5章では、RTLシミュレーションの結果から各ルータの性能を比較・評価する。最後に6章で、本稿のまとめと今後の予定について述べる。

2. ルーティング

2.1 基本仕様

本研究では、基本的なネットワークである2次元トラスを対象にする。また、フロー制御には、比較的小容量なバッファで実現できるワームホール方式を用いる。また、すべてのルータが最短経路ルーティングを行う。

2.2 ルーティングアルゴリズム

今回検討するルーティングアルゴリズムは、次に述べる3種類である。

2.2.1 Dimension-order

Dimension-order は、2次元の各次元を X, Y で表

すと、X 方向のルーティングが完了後 Y 方向のルーティングを行う非適応ルーティングである。ハードウェア構成が簡単だが、非適応チャネルのみの設計であるため経路選択の自由度はない。

2.2.2 *-channel

*-channel は、各ネットワークポートに適応チャネルとデッドロック回避用の非適応チャネルを持つ。メッセージは、まず各ポートの適応チャネルを用いて適応ルーティングを行う。適応ルーティングを行っているメッセージがブロックされると、非適応チャネルに回避するが、適応チャネルが使用可能なノードでは再び適応チャネルに戻り、適応ルーティングを行う。このように、メッセージは適応/非適応チャネル間を自由に行き来するため VC 切替え数は多くなる。しかし、全ポートに非適応チャネルを装備するためデッドロック回避チャネルのバンド幅は大きい。この *-channel と類似したアルゴリズムとして Duato の提案がある⁵⁾。

2.2.3 Recover-x

Recover-x は、あらかじめ用意した退避パスへ移動するメッセージを、Y 次元方向のルーティングが完了しており、あとは X 次元方向に進むのみのメッセージに限定することにより効率的なデッドロック回復を行う適応ルーティングである。退避メッセージの限定は、デッドロックサイクルを形成するメッセージ群から一部のメッセージを取り除くことによりデッドロックを解消できる、という考えに基づく。

このことにより、*-channel と比べて必要な非適応チャネル数が少なくなり、適応チャネルとして使用可能な VC を多く確保できる。さらに、非適応チャネルから適応チャネルにメッセージが移動しないため、ルーティングロジックが簡単になる。また、DISHA はデッドロック回復に全ポート共有のデッドロックバッファを使用するため、ハンドシェイクによる排他制御などが必要になり、調停が複雑になる。一方、Recover-x ではデッドロック回復に使用する非適応チャネルは複数のネットワークポートに共有されない。そのため、デッドロック回復時の調停が容易である。

3. 基本構成

3.1 ハードウェア構成

図1にルータのハードウェア構成を示す。どのルータも、図1(a)に示すように4つのネットワークポート(±X, ±Y Port)とPEインタフェース(I/F)を持つ。図中の実線は非適応ルータに必要な結線を示し、適応ルータではさらに破線で示す結線が必要となる。

図1(b)はポートの内部構成である。各ポートは、

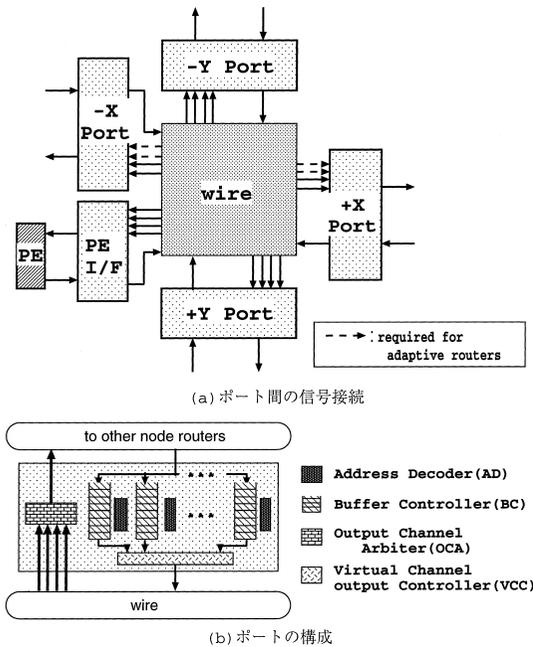


図1 ルータの構成

Fig. 1 Hardware organization of the routers.

必要な VC 数と同数の BC (Buffer Controller), および AD (Address Decoder) を持つ . また , BC は , フロー制御のための 8 フリット分の FIFO を持つ . 以下にルータの動作を簡単に説明する .

- (1) ネットワークからメッセージが入ってくると , 指定された VC 内の BC がそのメッセージを受信し , AD が受信メッセージのアドレスをデコードする .
- (2) AD は出力候補のポートに出力要求を行い , 許可を待つ . 同時にデッドロック回復ルータである Recover-x ではデッドロック検出のためにブロック時間のカウンタを開始する .
- (3) 出力要求を受けたポートの OCA (Output Channel Arbiter) は , 物理チャネルの使用状況と隣接ノードのバッファ状態を確認し , 各 VC ごとに他のポートからの出力要求と調停する . そして , 出力 VC を調停し , 選択したポートへ出力許可を返す .
- (4) OCA から出力許可を受けた AD は出力ポートを選択し , BC にメッセージ出力を許可する . その際 , VCC (Virtual Channel output Controller) において VC の出力制御を行う .

本ルータのハンドシェイクはフリット単位で行われ , 異なる VC に対してはメッセージはカットスルーが可能である . アービトレーションは固定とし , VC0 優

表 1 各 VC 構成におけるポートごとの VC 数

Table 1 The number of VCs for each VC configuration.

	X	Y	PE I/F	計
VC3 本構成	3	3	2	14
VC4 本構成	4	4	2	18
VC5 本構成	5	5	2	22

$$\text{計} = 2 \cdot X + 2 \cdot Y + \text{PE I/F}$$

先 , 適応ルーティングならばさらに適応チャネルが優先である . VCC においてもメッセージ出力に同様な優先順があり , 逐次的に処理される . また , メッセージのヘッダのみを格納するバッファを備えておらず , メッセージ本体も含めてすべて同一のバッファに格納される .

3.2 仮想チャネル構成

各ルーティングアルゴリズムに対して , 表 1 に示すように各ネットワークポート内に 3 , 4 , 5 本の VC を持つルータを設計した (それぞれ VC3 本構成 , VC4 本構成 , VC5 本構成と呼ぶ) . いずれの場合も , PE I/F 内には 2 本の VC を装備する . ここでは , 図 2 , 3 , 4 に示す VC 構成図を用いて , 各ルーティングアルゴリズムにおける VC 構成とその VC 切替えについて説明する . また , 3.2.2 項において VC 数の違いによる VC 切替えについて説明する . VC 構成図において , node#A と node#B , node#B と node#C はそれぞれ隣接ノードを表し , node#A はメッセージの送信元 , node#C は node#A から距離 2 のノードとする . なお , 我々の設計は入力バッファ方式を使用しているため , 図中の X と Y は入力ポートを表す .

3.2.1 ルーティングアルゴリズムにおける比較

図 2 に , Dimension-order を X , Y ポートそれぞれ 3 本の VC で設計した場合の VC 構成とメッセージがとりうる経路を示す . Dimension-order は非適応ルーティングであるため , すべて非適応チャネルで構成する . メッセージがラップアラウンド・チャネルを使用するか否かによって用いる VC を静的に定めることでトラスサイクルによるデッドロックを防ぐ .

図 3 は , *-channel の VC3 本構成における VC の適応/非適応チャネルの別とメッセージがとりうる経路である . *-channel はすべてのネットワークポートを 1 本の適応チャネルと 2 本の非適応チャネルで構成する . チャネルの使用状況に応じて適応チャネル/非適応チャネル間を自由に切り替えるため , VC 切替え数も多く , 調停ロジックが複雑になる .

図 4 の Recover-x では , X 方向の VC のうち 1 本は適応チャネルとして , 2 本はデッドロック回復のための非適応チャネルとして使用する . Y 方向はすべ

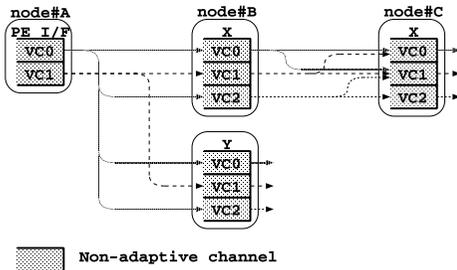


図2 Dimension-order の VC 切替え

Fig. 2 VC assignment for a Dimension-order router.

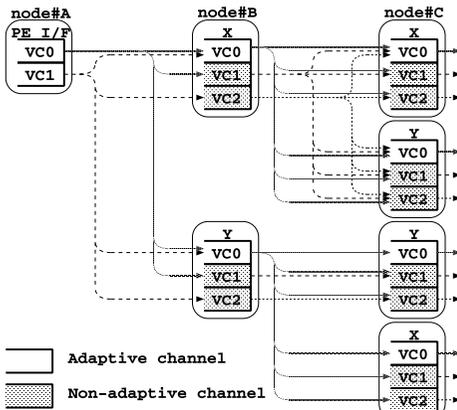


図3 *-channel の VC 切替え

Fig. 3 VC assignment for a *-channel router.

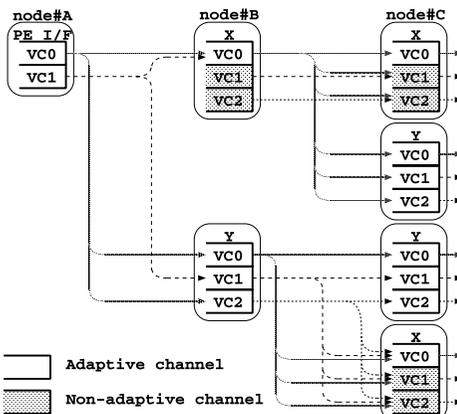


図4 Recover-x の VC 切替え

Fig. 4 VC assignment for a Recover-x router.

て適応チャネルとして使用可能である。したがって、*-channel より使用可能な適応チャネル数が多くなり、今回比較するアルゴリズムの中で最もルーティング自由度の高い構成である。非適応チャネルを使用するデッドロックからの退避メッセージが制限されるので、

VC 切替え数が少ない。

3.2.2 VC 数の違いによる比較

ここでは、VC を追加したときの VC 構成について述べる。VC 増加のイメージは文献 7) を参照されたい。

Dimension-order では、追加する VC は非適応チャネルのみである。したがって、ルーティング自由度の増加率は小さいが、動作周波数の低下を小さく抑えることができる。*-channel や Recover-x では、追加する VC は適応チャネル、非適応チャネルのどちらとしても実装できる。一般に適応チャネルを追加すると、非適応チャネルを追加するよりも AD や OCA が複雑になり、動作周波数の低下を招く。しかし、ルーティングの自由度が増加し、バンド幅が向上する。また我々の予備評価により適応チャネルの効果を確認している。したがって、適応ルーティングアルゴリズムではルーティングの自由度を優先し、適応チャネルを追加する。VC を追加すると VC 切替え数は増加するが、VC3 本から 4 本時よりも 4 本から 5 本時の方が増加する VC 切替え数は多い。また、*-channel は Recover-x より増加する VC 切替え数が多い。

3.3 メッセージの構造

メッセージの 1 フリットは物理チャネル幅と等しいものとし、32 ビットで構成する。各メッセージの 1 フリット目はヘッダとする。ヘッダには、2 次元の各次元ごとに、宛先ノードアドレス (絶対アドレス)、方向ビット、VC 番号の情報を保持する。メッセージサイズは 1 フリット (ヘッダのみ) 以上の不変長任意サイズを設定可能である。これらの詳細については文献 10) を参考されたい。

4. 論理合成

4.1 論理合成条件

3 章までに説明したルータを Verilog-HDL で設計し、以下の同一条件で論理合成した。

シンセサイザ: Synopsys HDL Compiler version 1999.05

回路の動作条件: 民生用最悪条件

マッピング最適化: Medium effort

ターゲットライブラリ: LSI Logic 0.6 μ m Array-Based Gate Array

各ルータの配線負荷はセル面積により自動選択される。すべてのルータは、クロックの立ち上りエッジ駆動である。なお合成において、すべてのサブ回路の境界最適化を行った。また、Verilog-HDL ソースプログラムでは、シンセサイザへのディレクティブを適宜指定し、クリティカルパスが短くなるよう配慮した。我々

表2 論理合成結果
Table 2 Synthesis results.

VCs/port	Dimension-order			*-channel			Recover-x		
	3	4	5	3	4	5	3	4	5
最大動作周波数 (MHz)	161.2	156.2	147.0	120.4	114.9	107.5	142.8	133.3	117.6
セル面積 (K gates)	70.9	90.2	109.1	72.2	96.5	120.9	75.6	94.5	118.2
配線領域 (K gates)	40.0	51.6	63.0	43.1	60.2	78.7	43.1	58.2	76.1
総面積 (K gates)	110.9	141.8	172.1	115.3	156.7	199.6	118.7	152.7	194.3

の設計においては、バッファに用いるメモリは比較的小さなものであるため、フリップフロップによって構成した。そのため、クリティカルパスへの影響は、メモリアクセスよりもルーティングロジックの複雑さの方が大きい。

4.2 論理合成結果

表2に、各ルータの論理合成結果を示す。表中の最大動作周波数とは、論理合成結果がタイミング条件を満たす中で最も高速であった場合の値(小数第2位切り捨て)である。また、各面積は最大動作周波数時のゲート数(小数第2位切り上げ)を表す。

4.2.1 動作周波数による比較

同数のVC構成で比較した場合、Recover-xは*-channelよりVC切替えの組合せ数が少ないので、最大動作周波数が高くなる。また、Dimension-orderは、ハードウェア構成が簡単なので最も動作周波数が高くなる。

次に、各ルータにおいて、VC数の違いによる影響を考察する。VC数を増加させるとともに、動作周波数は低くなる。これは、経路選択ロジックの複雑化と配線遅延の増加による。ロジックの複雑化は、主にVC切替えの組合せ数の増加による調停回路の複雑化に依存する。また、配線遅延の増加は面積増加に起因する。どのルーティングアルゴリズムにおいても、VC3本から4本にした場合よりもVC4本から5本にした場合の方が動作周波数の低下率がわずかに大きい。これは、ロジックの複雑化の影響が大きいためである。論理合成結果から得られる回路のクリティカルパスは、それぞれのルーティングアルゴリズムにおけるVC数によって異なるが、VC3本ではアドレスのデコード部分、VC5本ではOCAの調停部分がクリティカルパスになるという傾向がある。これより、VC3本のときには経路選択ロジックが動作周波数に与える影響の方が大きい、VC数が増えるとルーティングの自由度が増加することによるOCAの複雑化の影響の方が大きくなる。

4.2.2 回路規模による比較

同数のVC構成で比較した場合、Recover-x、

表3 Recover-xの各ブロック面積(単位:K gates)

Table 3 Each block area for Recover-x routers (K gates).

	VC3本構成		VC4本構成		VC5本構成	
	±X	±Y	±X	±Y	±X	±Y
AD	0.43	0.82	0.67	1.03	0.92	1.32
BC	13.11	13.44	17.57	18.29	21.29	21.37
OCA	2.15	1.01	3.17	2.00	4.81	2.86
VCC	0.21	0.22	0.40	0.36	0.44	0.47
計	15.89	15.47	21.80	21.68	27.44	26.01

-channelの適応ルータに比べて、非適応ルータであるDimension-orderの面積は小さい。これは、図1に示したとおり、非適応ルータのポート間結線が少なく、経路選択ロジックも簡単なためである。また、Recover-xと-channelを比較すると、VC3本構成では前者の面積が大きい。これは、今回設計したRecover-xがデッドロック検出回路を持っていることと面積よりも動作周波数を優先して論理合成したことにより、*-channelよりも高速な回路に合成されたためである。しかしVC4本/5本構成になると、カウンタなどで構成されるデッドロック検出回路よりもVC切替え数の増加によるロジック複雑化の影響が大きくなる。そのため、*-channelの面積の方が大きくなる。

次にVC数の影響について考察する。VC数を増すと、バッファ容量、ポート間結線の増加とロジック複雑化の影響で面積が増加する。Dimension-orderは同程度の割合で面積が増加するが、適応ルータではVC数が増すにつれて面積増加率が上がる。これは、適応ルーティングでは特にVC追加によるポート間結線の増加やロジック複雑化の影響が大きいからである。また、表3にRecover-xのポートを構成する各ブロックが占める面積を示す。これから分かるようにバッファを含むBCの面積が支配的である。適応チャンネルを追加することは非適応チャンネルを追加するよりもADやOCAを複雑にする。したがって、VC数を追加するとXポートのADとYポートのOCAの面積増加率が高い。また、ADとBCは非適応チャンネルを持つXポートの面積がYポートよりも小さいが、OCAは出力するVC切替えの組合せ数の関係でXポートの面積が大きいことが分かる。

5. シミュレーション

5.1 シミュレーション条件

ルータのシミュレーション条件を以下に示す。

シミュレータ： Cadence Verilog-XL

ネットワークサイズ： 10×10 (= 100 ノード)

通信パターン：

- Hot-spot 通信—すべてのノードが 100 個のメッセージを連続して送信するが、そのうちの 25%は、ノードアドレス $(4, j)$, $0 \leq j \leq 9$ の 10 個のノードに集中させる。残りのメッセージは自分以外の任意のノードに等確率で送信する。
- Random 通信—各ノードが 100 個のメッセージをランダムな宛先に連続して送信する。

また、本稿で評価するルータは各ノードが近距離で密に結合した並列計算機システムのもを前提とし、そのノード間データ転送遅延は各ルータにおける動作周波数の 1 クロック以内と仮定する。設計したルータは、メッセージヘッダが 1 つのルータを通過するために 3 クロックを要するので、ノード間遅延を含めて 1 ホップ時間は 4 クロックである。宛先 PE に到着したメッセージは、随時 PE に取り込まれるものとする。また、PE I/F は送信と受信を並列処理することが可能である。

Recover-x において、デッドロックが生じてから回復を始めるまでのメッセージのブロック時間は、どの VC 構成でも 4 クロックとした。これは、カウント時間を変更して評価した種々の実験により、どのメッセージサイズでも特にバンド幅に落ち込むところがなく高バンド幅を示した値である。

5.2 シミュレーション結果

シミュレーション結果として、各ルータの上記通信パターンに対するネットワーク全体のバンド幅とメッセージの平均レイテンシを示す。各ルータの動作周波数は、同一の 100 MHz で動作する場合と、表 2 で示した最大動作周波数で動作する場合を示す。なお、ネットワークの定常状態における評価を行うため、ウォームアップとしてシミュレーション開始から 2000 番目までの到着メッセージを除いて、それに続く 5000 メッセージについてバンド幅とレイテンシを評価した。

5.3 Hot-spot 通信

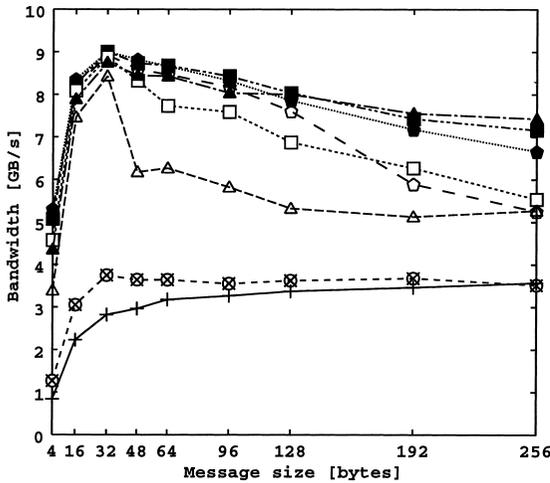
5.3.1 バンド幅

図 5 (a) に、全ルータが 100 MHz で動作する場合の Hot-spot 通信に対するバンド幅を示す。バンド幅は、メッセージサイズを 4 ~ 256 バイトに変化させて測定した。このグラフから、Recover-x、*-channel、

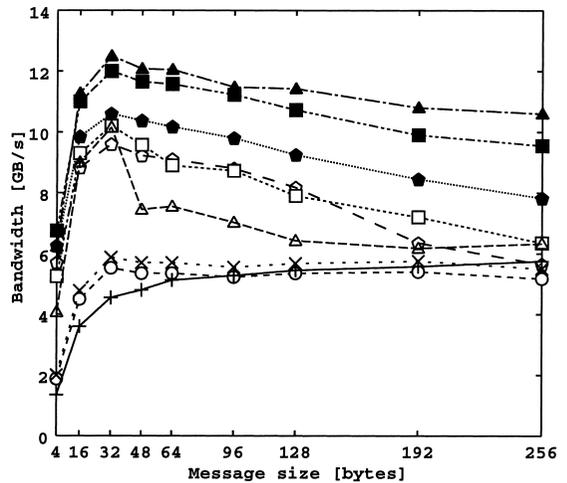
Dimension-order の順で高バンド幅であることが分かる。これは、Hot-spot 通信がネットワーク負荷の一例でない非ユニフォーム通信パターンであるため、アルゴリズム的にルーティング自由度の高い順となっていることを表す。特に適応ルータにおいて、バッファサイズまでの小さなメッセージサイズの場合、2 つのバッファを占めることなく個々に代替経路を選択してルーティングするのでチャンネル利用率が良い。しかしバッファサイズ以上のメッセージサイズになると、メッセージは 2 つ以上のルータのチャンネルを使用するので、代替経路を見つけにくくなり適応ルーティングによる自由度増加の効果が減少する。したがって、バッファサイズと等しい 32 バイトにおいてピーク性能となる。適応ルータはメッセージサイズが大きくなるにつれてバンド幅低下が大きいが、経路が 1 通りしかない Dimension-order に比べてルーティング効率が良いため、さらに大きなメッセージサイズになっても結果の逆転はないと考えられる。また、*-channel においてメッセージサイズがバッファサイズを超えると、適応ルーティングを行っているメッセージがブロックされることが多くなり非適応チャンネルを使用する確率が多くなる。したがって、適応ルーティングによる効果が小さくなりバンド幅が急激に低下する。VC 追加の効果は、*-channel と Dimension-order の VC3 本構成と 4 本構成で確認できるが、それ以外については顕著に見られない。Recover-x では VC3 本、4 本、5 本すべての構成において同程度のバンド幅を示す。すなわち、与えられた通信パターンに対してネットワークが十分なルーティング自由度を達成した段階でバンド幅は飽和するといえる。場合によっては、適応チャンネルの調停時に発生する待ち時間、VCC におけるメッセージ出力待ち時間、さらに VC 切替え時に発生するパルスサイクルなどがオーバーヘッドとして現れることもある。これは波形出力で実行結果を詳細に検討した結果観測された。そのため、Recover-x や *-channel の VC5 本構成が VC3 本、4 本構成に比べて大きく低下する。

図 5 (b) は、各ルータが最大動作周波数で動作する場合のバンド幅を示す。このグラフから、Dimension-order の VC4 本構成は 256 バイトで増加した自由度の効果が小さくなり 3 本構成と同程度のバンド幅になるが、その他のメッセージサイズでは高バンド幅である。また、*-channel では、64 バイトまでの小さなメッセージサイズに対して VC4 本構成で十分なルーティング自由度を有するため、動作周波数の影響が大きくなる。64 ~ 128 バイトまでのサイズでは追加した

Dimension-order		*-channel	Recover-x
VC3	+	VC3	▲
VC4	x	VC4	■
VC5	o	VC5	●

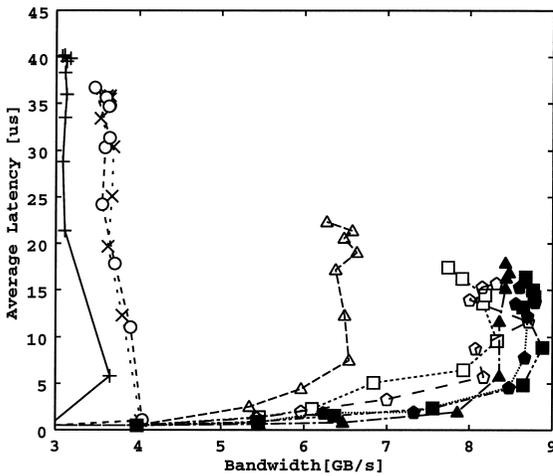


(a) 100MHz

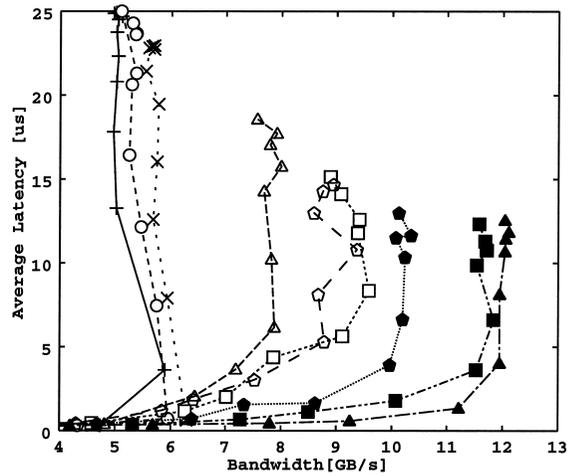


(b) 最大動作周波数

図5 Hot-spot 通信のバンド幅
Fig.5 Bandwidth for the Hot-spot traffic.



(a) 100MHz



(b) 最大動作周波数

図6 Hot-spot 通信の平均レイテンシ
Fig.6 Latency for the Hot-spot traffic.

適応チャネルによる自由度増加の効果がやや大きく、VC5本構成が良好な値を示す。しかし、128バイト以上になるとVC5本構成のバンド幅低下が大きくなったため、再びVC4本構成が高バンド幅になる。このようにVC4本構成は、一部のメッセージサイズにおいてVC5本構成に逆転されるものの、それらに対してもほぼ同程度のバンド幅を示すことから堅牢なバンド幅特性を示すといえる。Recover-xでは、VC3本構成

成ですでにHot-spot通信に対して十分なルーティング自由度を提供するため、VC追加による動作周波数低下の悪影響が生じている。同様な理由で、どのアルゴリズムにおいてもVC5本構成は4本構成にやや劣る結果となった。

5.3.2 平均レイテンシ

図6にHot-spot通信におけるメッセージの平均レイテンシを示す。平均レイテンシは、メッセージサイ

ズが 64 バイトのときの、メッセージが PE 内で生成されてから最終フリットが宛先ノードの PE に到着するまでの時間を表す。グラフの横軸は、メッセージの送信間隔を変えることによってネットワークに与える負荷をバンド幅で表している。

いずれのルータも、あるネットワーク負荷に達するとレイテンシが上昇する。このレイテンシが上昇するときのネットワーク負荷はネットワークが飽和しているところであり、これを飽和容量と呼ぶ。図 6(a) から、ネットワーク飽和容量はルーティング自由度の高さを反映しているといえる。すなわち、適応チャネル数の多い *channel や Recover-x の飽和容量が大きく低レイテンシである。この通信パターンにおける Recover-x では VC 数が 3 本のときに十分な自由度を達成しているため、レイテンシにおいても差がほとんどない。また、*channel ではネットワークポートに適応チャネルが 1 本しかない VC3 本構成は自由度が低く、他の構成に大きく劣る。図 6(b) を見ると、各ルータの最大動作周波数を反映して、Recover-x と *channel の飽和点の差が広がり、Dimension-order と *channel の差が縮まっていることが分かる。ただし、Dimension-order が *channel を逆転するには至っていない。Recover-x では動作周波数低下の影響が大きく、VC5 本構成が劣る。Dimension-order や *channel は VC3 本の自由度が低いいため、動作周波数を考慮しても VC4 本、5 本構成に近づくのみで逆転には至らなかった。これらのアルゴリズムでは VC4 本構成が飽和容量が大きく、低レイテンシである。

なお、他のメッセージサイズにおいても評価したが、一般にそのサイズにおいて高バンド幅なルータのネットワーク飽和容量が大きく、低レイテンシである傾向が見られる。すなわち、図 5 においてメッセージサイズによってルータ間のバンド幅の高低に逆転がある場合にはレイテンシ特性の優劣もメッセージサイズによって変化する。しかし、概して Recover-x、*channel、Dimension-order の順、Recover-x 間では VC3 本、4 本、5 本の順にレイテンシ特性が優れている傾向は変わらない。

5.4 Random 通信

5.4.1 バンド幅

図 7 に Random 通信におけるバンド幅を示す。Random 通信はユニフォーム通信パターンであるため、ネットワークが一様に混雑すると代替経路を探そうとする適応ルーティングの効果は現れにくくなる。したがって、図 7(a) の 100 MHz 動作において、Hot-spot 通信と同様な理由でチャネル利用効率の良い 32 バイト

程度の小さなメッセージのバンド幅が高い。Hot-spot 通信と比較すると、一般にユニフォーム通信特性の良い Dimension-order のバンド幅が相対的に高くなっている。同じ理由で、非適応チャネルで次元順ルーティングを行う *channel が高バンド幅を示している。また、メッセージサイズが大きくなるほど VC 数の違いが小さい傾向があるのは、適応ルーティングによる自由度増加や VC 数追加による自由度増加の効果が少なくなるからである。ただし、動作周波数を考慮した図 7(b) を見ると、*channel よりも動作周波数の高い Dimension-order や Recover-x の方が高バンド幅となった。またいずれのルーティングアルゴリズムにおいても、ルーティング自由度と動作周波数のバランスから VC4 本構成のバンド幅特性が良いことが分かる。

5.4.2 平均レイテンシ

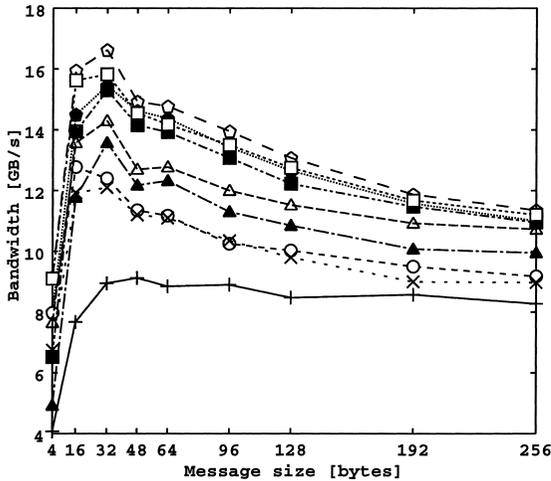
図 8 に Random 通信におけるメッセージの平均レイテンシを示す。図 8(a) の 100 MHz 動作では、デッドロック回復までの時間をカウントする必要のない *channel がどの VC 構成においても最も飽和容量が大きく、低レイテンシである。ただし、*channel と Recover-x の飽和容量の差は小さい。また、Dimension-order は図 6(a) に比べて適応ルータに接近しているが、依然として適応ルータよりも飽和容量が小さく、レイテンシも高い。Random 通信では VC 追加によって増加した自由度が大きく貢献し、VC 数が多いルータほど低レイテンシとなる。

図 8(b) では、各ルータの性能が接近しており、最大動作周波数の関係で Dimension-order と *channel では飽和容量に逆転も見られる。Recover-x は、どの VC 構成においても最も飽和容量が大きく、低レイテンシとなっている。また、どのアルゴリズムにおいても同一周波数の場合に低レイテンシを示した VC5 本構成は動作周波数が低いので VC4 本構成、Recover-x においては 3 本構成にも逆転される。このように、ルーティング自由度の増加によるメリットと動作周波数の低下によるデメリットから、VC4 本構成のレイテンシ特性が良いといえる。

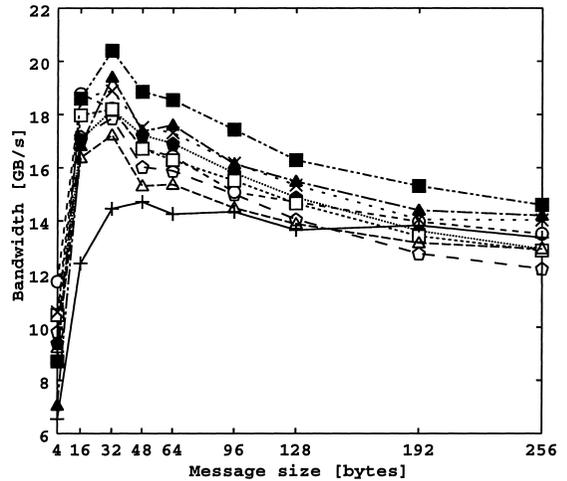
6. おわりに

本稿では、Dimension-order、*channel、Recover-x の 3 種類のルーティングアルゴリズムにおいて、VC 数を変えた場合のハードウェアコストと通信性能を定量的に評価した。ルータチップの動作周波数とルーティング自由度のトレードオフを考慮した結果、物理チャネルあたり VC3 本から 4 本への増加についてはルー

Dimension-order		*-channel	Recover-x
VC3	+	△	▲
VC4	x	□	■
VC5	o	○	●

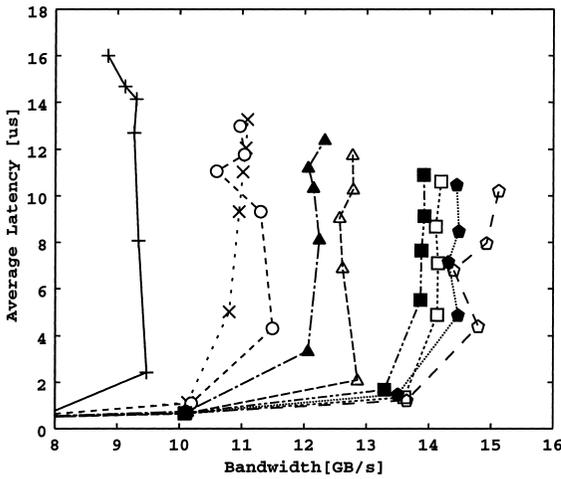


(a) 100MHz

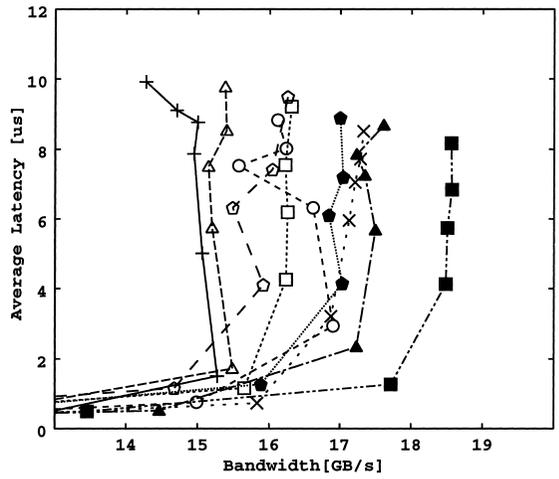


(b) 最大動作周波数

図7 Random通信のバンド幅
Fig.7 Bandwidth for the Random traffic.



(a) 100MHz



(b) 最大動作周波数

図8 Random通信の平均レイテンシ
Fig.8 Latency for the Random traffic.

ティング自由度増加のメリットが確認できた。しかし、4本から5本へ増加すると、ルーチチップの動作周波数の低下によりルーティング自由度増加のメリットが相殺されてしまうことを示した。また、ルーティングアルゴリズムについて分析すると、適応ルーチは動作周波数で非適応ルーチに劣るものの、ルーティング自由度の高さによって特に非ユニフォーム通信パターンに対する通信性能が優れることが分かった。チップ

の動作周波数を考慮したネットワークの通信性能評価では、Recover-xの物理チャンネルあたりVC4本の構成はユニフォーム/非ユニフォーム通信において他のルーチに比べて堅牢な結果を示し、高バンド幅、低レイテンシを達成した。

今後の課題として、クラスタ結合を想定してルーチ間のデータ伝送遅延を媒介変数とした評価、また実アプリケーションに対する通信性能の評価があげられる。

謝辞 本研究において貴重なご意見をいただきました筑波大学の山口喜教氏、宇都宮大学馬場研究室の諸氏に深く感謝いたします。本研究の一部は東京大学大規模集積システム設計教育研究センターより提供していただいたCADツールを使用しています。

本研究は、一部文部省科学研究費基盤研究(B)課題番号10558039, 奨励研究(A)課題番号11780190, 実吉奨学会の援助による。

参考文献

- 1) Anjan, K.V. and Pinkston, T.M.: An Efficient, Fully Adaptive Deadlock Recovery Scheme: DISHA, *Proc. 22nd ISCA*, pp.201-210 (1995).
- 2) Berman, P.E., Gravano, L., Pifarré, G.D. and Sanz, J.L.C.: Adaptive Deadlock and Livelock Free Routing with all Minimal Paths in Torus Networks, *Proc. SPAA* (1992).
- 3) Chien, A.A.: A Cost and Speed Model for k-ary n-Cube Wormhole Routers, *IEEE Trans. Parallel and Distributed Systems*, Vol.9, No.2, pp.150-162 (1998).
- 4) Dai, D. and Panda, D.K.: How Much Does Network Contention Affect Distributed Shared Memory Performance?, *Proc. ICPP'97*, pp.454-461 (1997).
- 5) Duato, J.: A New Theory of Deadlock-Free Adaptive Routing in Wormhole Network, *IEEE Trans. Parallel and Distributed Systems*, Vol.4, No.12, pp.1320-1331 (1993).
- 6) Flich, J., Malumbres, M.P., López, P. and Duato, J.: Performance Evaluation of Networks of Workstations with Hardware Shared Memory Model Using Execution-Driven Simulation, *Proc. ICPP'99* (1999).
- 7) 堀田真貴, 林 匡哉, 中村さゆり, 吉永 努, 大津金光, 馬場敬信: 適応ルータにおける最適な仮想チャネル数に関する考察, 並列処理シンポジウム JSP2000 論文集, pp.189-196 (2000).
- 8) 上樂明也, 舟橋 啓, 鯉淵道紘, 若林正樹, 天野英晴: 命令レベルシミュレーションによる adaptive routing の評価, *Proc. HOKKE2000*, 情報処理学会研究報告 2000-ARC-137-9, pp.47-52 (2000).
- 9) Vaidya, A.S., Sivasubramaniam, A. and Das, C.R.: LAPSES: A Recipe for High Performance Adaptive Router Design, *Proc. HPCA-5'99* (1999).
- 10) 吉永 努, 林 匡哉, 堀田真貴, 山口喜教, 大津金光, 馬場敬信: 適応ルータの出力チャネル選択における優先次元指定の効果, 情報処理学会論文誌, Vol.40, No.5, pp.1958-1967 (1999).
- 11) 吉永 努, 林 匡哉, 堀田真貴, 中村さゆり, 大

津金光, 馬場敬信: Recover-x ルーティング, 情報処理学会論文誌, Vol.41, No.5, pp.1360-1369 (2000).

(平成 12 年 8 月 31 日受付)

(平成 13 年 3 月 9 日採録)



堀田 真貴 (学生会員)

1999年宇都宮大学工学部情報工学科卒業。現在同大学大学院博士前期課程在学中。ハードウェア設計, 特に, 並列計算機アーキテクチャに興味を持つ。



吉永 努 (正会員)

1986年宇都宮大学工学部情報工学科卒業。1988年同大学大学院修士課程修了。同年より宇都宮大学工学部助手。1997年から翌年にかけて電子技術総合研究所・客員研究員。2000年8月より電気通信大学大学院情報システム学研究科助教授。博士(工学)。並列計算機アーキテクチャ, リコンフィギュラブル・コンピューティング等に興味を持つ。電子情報通信学会, IEEE 各会員。



大津 金光 (正会員)

1993年東京大学理学部情報科学科卒業。1995年同大学大学院修士課程修了。1997年同大学院博士課程退学, 同年より宇都宮大学工学部助手となり現在に至る。理学修士。高性能計算機システム, 特にマイクロプロセッサアーキテクチャに興味を持つ。



馬場 敬信 (正会員)

1970年京都大学工学部数理工学科卒業。1975年同大学大学院博士課程単位取得退学。同年より電気通信大学助手, 講師を経て, 現在宇都宮大学工学部教授。工学博士。1982年より1年間メリーランド大学客員教授。計算機アーキテクチャ, 並列処理等の研究に従事。電子情報通信学会, IEEE 各会員。1992年情報処理学会 Best Author 賞。著書「Microprogrammable Parallel Computer」(MIT Press)、「コンピュータアーキテクチャ」(オーム社)等。