

局所的な混雑を予測する オンチップアダプティブルータの設計と実装

谷口将一^{†1} 向後卓磨^{†2} 山崎信行^{†2}

Networks-on-Chip (NoC) はメニーコアプロセッサのスケラブルな通信網として提案されているが、これは実装面積の制約により単純なルーティングアルゴリズムを用いることが多い。単純なルーティングを行うとトラフィックの偏り及びパケットの衝突が起こり、ネットワークの性能が低下する。

本研究では、局所的な混雑を予測するオンチップアダプティブルータを提案する。これは過去の混雑情報と未来の予測された混雑情報を受信し、これらの情報と自身のローカルの情報を合成して Minimal Adaptive Routing を行う。実装したルータは、2.6%の面積と物理チャネルあたり 3 本の配線のオーバヘッドのみで、各トラフィックにおけるスループットを平均 17.2%向上させた。

Designing and Implementation of On-chip Adaptive Router with Predictor for Regional Congestion

MASAKAZU TANIGUCHI,^{†1} TAKUMA KOGO^{†2}
and NOBUYUKI YAMASAKI^{†2}

Networks-on-Chip (NoC) has been proposed as a scalable fabric for communication on many core processors. However, as on-chip routers have constraint in implementation area, they generally use simple routing algorithms. Simple routing algorithms lower the network performance by imbalanced traffic and conflict of packets.

In this paper, we propose On-chip Adaptive Router with Predictor for Regional Congestion. It receives past congestion information and future predicted congestion information. It synthesizes these information and its local congestion information and use Minimal Adaptive Routing. It improved the average throughput for each traffic by 17.2% with the overhead of only 2.6% area and 3 wires for each physical channel.

1. はじめに

近年、プロセッサのメニーコア化により Networks-on-Chip (NoC)¹⁾ が注目を集めているが、NoC はチップ上に実装するため、ハードウェアコストが低くなるよう設計する必要がある。そのため、オンチップルータの実装面積には制約があり、TILE64²⁾ は DOR(2.1 節参照) を採用するなど、単純なルーティングを行う事が多い。単純なルーティングはトラフィックの偏り及びパケットの衝突を引き起こし、性能の低下を招く。これを回避するため、オンチップルータ向けの Adaptive Routing の研究がなされているが、Adaptive Routing の性能とそれを行うための実装面積にはトレードオフの関係がある。本研究ではハードウェアコストを抑えながら混雑を回避するオンチップアダプティブルータを提案する。このルータは、各出力ポートの混雑状況を監視及び予測し、その情報を周辺ルータへ伝播する。さらに、パケットを送信するという情報をパケットフローの下流に先送りすることで、混雑箇所の迂回を行うだけでなく、混雑をそもそも起こりにくくすることができる。このルータは混雑状況の監視及び予測に単純な演算しか用いないため、低ハードウェアコストで実装することができる。本研究では提案したルータを HDL で実装し、論理合成及びネットワークシミュレーションを行った。本研究の構成は、次の通りである。まず、第 2 章で既存のルーティングアルゴリズムとそのトレードオフについて述べる。第 3 章で本研究で提案するルータの設計について述べ、第 4 章でその実装の詳細について述べる。第 5 章で設計及び実装したルータの評価及びその考察について述べ、第 6 章で本研究をまとめる。

2. 既存のルーティング手法

本研究の目的は、ハードウェアコストを抑えながら効率的に Adaptive Routing を行うことである。本章では、既存の各ルーティング手法とそれらの性能及びハードウェアコストについて述べる。

^{†1} 慶應義塾大学理工学部情報工学科

Department of Information and Computer Science, Faculty of Science and Technology, Keio University

^{†2} 慶應義塾大学大学院理工学研究科開放環境科学専攻

Department of Computer Science, Graduate School of Science and Technology, Keio University

2.1 次元順ルーティング (Dimension Order Routing : DOR)³⁾

予め各次元に順序を決めておき、出力可能なポートが複数ある場合、その順序が最も先である次元のポートに出力を行う。例えば 2D-Mesh では、初めに宛先ノードと X 座標が等しくなるまで X 軸方向にルーティングを行い、その後で Y 軸方向へとルーティングを行う。この手法は初めからデッドロックフリーであり、ハードウェアコストも低いため NoC において多用される。

2.2 空き仮想チャンネル数最大出力ポート選択機構⁴⁾

利用可能な仮想チャンネルの数を各出力ポートごとに集計し、最も多かった出力ポートを選択する。ネットワークの状態を考慮しているため、比較的空いているチャンネルを選択することができ、性能が向上する。また、ルータ間の配線を追加する必要は無く、ハードウェアコストは低い。但し、各ルータはローカルの情報のみ使用しているため、その性能向上には限界がある。

2.3 Regional Congestion Awareness (RCA)⁵⁾

Adaptive Routing では混雑情報の精度がネットワークの性能に影響するが、RCA は、混雑情報を伝播し集約することで、ネットワークの広域の混雑情報を利用可能にしている。各ルータは他のルータからの混雑情報を受信し、その情報と自身の持っている情報を合成して更に他のルータに伝播する。この情報伝達をパケットフローの上流方向に再帰的に行うことで広域の混雑情報が利用可能となる。広域の混雑情報を利用可能であるため、様々なトラフィックにおいて性能が向上する。但し、混雑情報の伝播及び集約には多数の加算器が必要であったり、物理チャンネルあたり 8 本または 16 本のルータ間のワイヤの増加が必要など、ハードウェアコストは高い。

3. 設 計

前章で述べたように、ルーティングアルゴリズムの性能とその実装面積にはトレードオフの関係があり、ローカルな情報のみを使用するとハードウェアコストは小さいが性能は向上しにくい。一方で、広域の情報を使用すると性能は向上するがハードウェアコストが高くなる。また、遠方の情報を利用するとその情報のフレッシュネスが問題となる。そこで、本研究では局所的な混雑を予測するオンチップアダプティブルータ (On-chip Adaptive Router with Predictor for Regional Congestion : PRC) を提案する。PRC は中程度の距離の混雑状況の情報を使用して混雑を回避し、さらに予測を用いてフレッシュネスの向上を目指す。PRC は RC (Route Computation), VSA (Virtual Channel Allocation と Speculative

Switch Allocation) 及び SA (Switch Traversal) の 3 段パイプライン構成である投機ルータ⁶⁾ をベースラインとしており、軽量ルータを想定している。また、Minimal Adaptive Routing を行うことを前提としており、デッドロック回避の手法には West-First⁷⁾ を用いる。PRC の設計は複雑であるため、例を示しながら順を追って設計を説明する。ここで座標は最北西を原点 (0, 0) とし、座標 (x, y) に位置するルータを $R(x, y)$ と表記する。また、 $R(x_1, y_1)$ と $R(x_2, y_2)$ の間に位置するチャンネルを $C(R(x_1, y_1), R(x_2, y_2))$ と表記する。また、混雑しているチャンネルを破線で、そうでないチャンネルを実線で示す。

3.1 伝播情報の生成と伝達

PRC は 2 ホップ先までの混雑情報を使用する。1 ホップ先までの情報は各ルータが初めからローカルに持っているため、実際に情報を伝播するのは 1 ホップで良い。各ルータは東西南北それぞれの出力ポートの混雑情報をそれぞれ 1 ビットで表現し、4 ビットのベクトルを生成する。この 4 ビットのベクトルを伝播情報と定義する。各ルータは伝播情報のうち 3 ビットを隣接ルータに送信する。PRC は Minimal Adaptive Routing を行うため、パケットが元来たチャンネルを逆戻りすることはなく、送信する情報はそのチャンネルを除いた 3 つの出力ポートに対応する 3 ビットで良い。伝播情報を受信したルータは、伝播情報と自身が初めからローカルに持っていた情報を合成し、最も混雑度の低い経路にルーティングを行う。図 1 はその例を示したもので $R(1, 1)$ は $R(1, 2)$ に西、北、東それぞれの出力ポートの混雑情報を送信している。 $R(1, 2)$ は $C(R(1, 1), R(2, 1))$ が混雑していることを知ることができるため、この混雑を回避して東側の $R(2, 2)$ へパケットを送信することができる。

3.2 先送り情報及び予測付き伝播情報の生成

図 2 はパケットの衝突の例を示したものである。PRC は各チャンネルの情報をそれぞれ別々に送信しているため、 $R(1, 2)$ は $C(R(1, 1), R(2, 1))$ が混雑していないことを知っている。そのため、 $C(R(1, 1), R(2, 1))$ を有効に使うことができるのだが、逆にこのことにより $R(0, 1)$ が送信したパケットと $R(1, 2)$ が送信したパケットの衝突が起こっている。実アプリケーションにおいてはバーストトラフィックが生じることが多く、このように複数のパケットが送信されることが多いため、このようなパケットの衝突は多発し性能の低下をもたらす。このような衝突は、従来手法のようにパケットフローの上流にのみ情報を伝播するだけでは回避することはできない。なぜならば、パケットフローの上流にのみ情報を伝播するだけでは、実際に混雑が生じてからしかそれを回避することはできないためである。

そこで、図 3 に示すように、パケットを送信するという情報をパケットフローの下流に先送りすることを提案する。 $R(0, 1)$ は $R(0, 2)$ からパケットを受信し、そのパケットが東

側にルーティングされることが分かると、 $R(1, 1)$ にパケットを送信するという情報 1 ビットを先送りする。この情報を先送り情報と定義する。先送り情報が送信されるのは、実際にパケットが送信されるより最低でも 2 クロック早いタイミングとなる。PRC はこのタイミングを更に早めるため、予測を用いて RC ステージが終わる前に先送り情報を送信する。予測アルゴリズムについては 3.4 節で述べる。

先送り情報を受信したルータはそのパケットがどちらに送信されるのかを予測し、その予測情報と自身がローカルに持っている伝播情報を合成して未来の混雑情報を生成する。この混雑情報を予測付き伝播情報と定義する。予測付き伝播情報は伝播情報同様、各ビットが各出力ポートの混雑状況に対応する 4 ビットのベクトル情報である。先は説明のため各ルータは伝播情報を送信すると述べたが、実際に送信するのはこの予測付き伝播情報である。予測付き伝播情報を受信したルータは 2 ホップ先までの未来の混雑情報を知ることができる。

図 3 の例では、 $R(1, 1)$ は $R(0, 1)$ からの先送り情報を受け取り、そのパケットが東側の $R(2, 1)$ に送信されるだろうと予測している。そのため、 $R(1, 1)$ が $R(1, 2)$ に送信する予測付き伝播情報には、 $C(R(1, 1), R(2, 1))$ が混雑するだろうという情報が含まれ、これを知った $R(1, 2)$ は混雑を回避するため、 $R(1, 3)$ から受信したパケットを $R(2, 2)$ に送信することができる。PRC はこのようにしてパケットの衝突を回避することができる。

3.3 ハードウェアコストの削減

PRC は 2 ホップ先までの混雑情報を考慮してルーティングを行うが、2 ホップ先までの経路は最大 4 経路あり、更に宛先の座標によって使用可能な経路の数は変化する。4 経路のうちどの経路が使用可能なかを判断し、さらにその中から最も混雑度合いの低い経路を選択するためにはハードウェアコストがかかり、動作周波数を下げる要因にもなる。そこで本研究では、宛先が同じ軸上に無ければどの座標であっても有効な、右折または左折する経路のみの混雑情報を考慮することにし、他の 2 つのルートは無視して混雑情報の比較を行う。つまり、図 3 においては、 $R(1, 2)$ から $R(3, 0)$ にルーティングする際、 $C(R(1, 0), R(1, 1))$ を使用する経路と、 $C(R(2, 2), R(3, 2))$ を使用する経路の混雑情報は無視し、 $R(2, 1)$ を通る経路の混雑状況の比較のみを行う。これにより、宛先が同じ軸上に無いとき、比較すべき経路は常に 2 経路となり、ハードウェアを簡素化できる。更に予測付き伝播情報に関しては、直線方向の情報を伝える必要がなくなるため、実際に送信するビット数を 3 ビットから 2 ビットに削減できる。

3.4 出力ポートの予測

出力ポートの予測は各入力チャンネルごとに行い、2 回連続で同じ方向にルーティングを

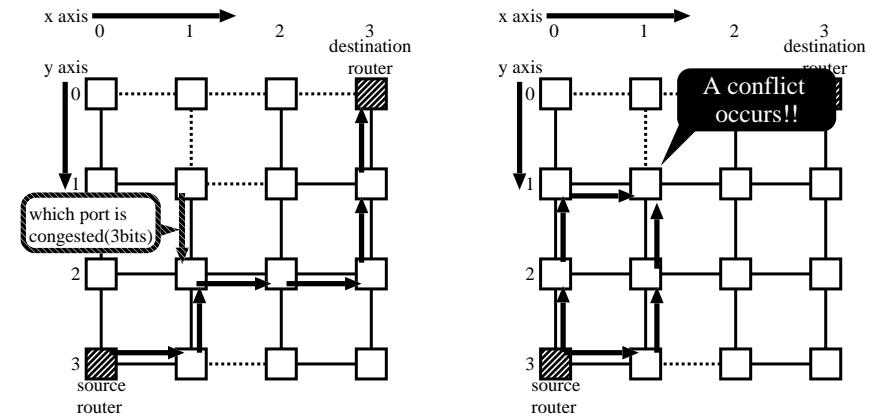


図 1 伝播情報

図 2 パケットの衝突

Fig. 1 transmitted information

Fig. 2 packet conflict

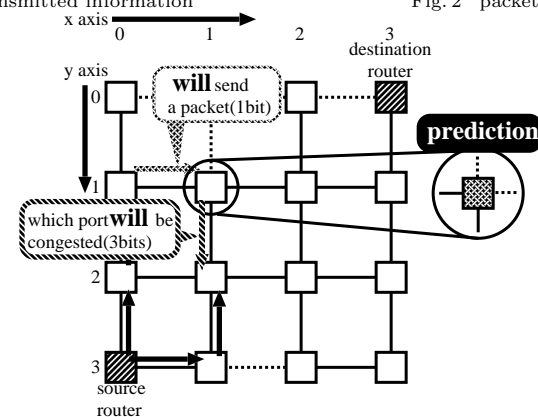


図 3 先送り情報及び予測付き伝播情報

Fig. 3 ahead information and predicted transmitted information

行った場合、次回以降その方向にルーティングするであろうと予測する。これはプロセッサにおける 2 ビット予測器と同じ発想で、一度異なる方向にルーティングされたとしても、その次からは元の方向にルーティングする可能性が高いであろうという思想に基づく。宛先の同じパケットが大量に流れているところに、宛先の異なる 1 パケットが混入した場合、1 ビット予測器^{*1} では予測ミスが 2 回起こってしまうが、2 ビット予測器では 1 回で済む。また、パーストラフィックが流れ始める際、3 ビット予測器では初めに数回予測ミスが起こるが、2 ビット予測器では 1 回のみで済む。また、プロセッサにおける予測器では、予測の方向は分岐するか否かの 2 方向だけだったが、NoC では東西南北とローカルの 5 方向あり、3 ビット予測器では予測の方向が切り替わりにくくなってしまうため、本研究では 2 ビット予測器を採用した。また、ハードウェアコストを下げるため、先送り情報の生成及び予測付き伝播情報の生成には、同じ予測アルゴリズムを用い、同一の予測器を共有して使用する。

4. 実装

図 4 に PRC のアーキテクチャを示す。点線で囲った部分が本研究で追加したハードウェアである。

Route Predictor 各入力仮想チャネルの状態とその出力ポートを監視し、3.4 節で述べた通りの予測を行う。

Local Congestion Transmitter 先送り情報を生成するためのハードウェアであり、各入力仮想チャネルの状態とそれらがパケットを出力するポートを監視し、東西南北それぞれのポートの混雑状況 1 ビット × 4 を出力する。パケットを保持している全仮想チャネルを調べ、それらがパケットを出力するポートに対応するビットを 1 にする。但し、RC ステージの仮想入力チャネルについては予測情報を用いる。この 4 ビットの情報は、それぞれ 1 ビットずつ隣接ルータに送信される同時に、ローカルの Regional Congestion Transmitter にも送信される。この理由は次節で述べる。

Regional Congestion Transmitter 隣接ルータの Local Congestion Transmitter からの各ポート 1 ビットの先送り情報とローカルの伝播情報から、予測付き伝播情報を生成する。先送り情報に関しては、各入力チャネルについて 1 が立っていれば、Route Predictor を用いてその出力ポートを予測し、対応するビットを 1 にする。このように

して得られた 4 ビットのベクトルとローカルの伝播情報 4 ビットとの論理和を取ったものが予測付き伝播情報であり、これが本ハードウェアの出力である。この出力は隣接ルータへ送信されると共に、Regional Congestion Mixer にも送信される。この理由は次節で述べる。ここで考察したいのが、伝播情報も先送り情報も各出力ポートに対応する 4 ビットの混雑情報であるということである。その差は予測情報の有無だが、伝播情報に予測情報が加わっても問題はなく、両者は同一の情報として扱うことができる。そこで、これらの情報はどちらも Local Congestion Transmitter の出力を用いることにしハードウェアコストを削減する。つまり、Local Congestion Transmitter の出力はその使用のされ方により呼び方が変わるだけであり、ローカルで使用されれば伝播情報、隣接ルータに送信されれば先送り情報と呼ばれる。

Regional Congestion Mixer 本ハードウェアはルータの持っている 3 種類の情報を合成し、4 ポート × 2 経路 = 計 8 経路の混雑度合いをそれぞれ数値化する。各情報の数値化は以下のように行う。

先送り情報 先送り情報はパケットフローの上流から下流への情報であり、隣接ルータの Local Congestion Transmitter から受信しているが、各入力チャネルごとに予測を用いて出力ポートを判断する必要がある。その判断はローカルの Regional Congestion Transmitter が既に行っているため、これを受信するだけで良い。各出力ポートの混雑度合いを 1 ビットで得ることができるため、その 0 または 1 の値を、その出力ポートを使用する経路の混雑度合いとする。

予測付き伝播情報 隣接ルータの Regional Congestion Transmitter からの入力チャネルあたり 2 ビットの情報であり、2 経路 × 4 出力ポートの計 8 経路の混雑情報である。パケットフローの下流から上流への情報であり、それぞれの混雑情報をそのまま 0 または 1 とする。

ローカル情報 初めからローカルに持っている情報であり、各出力ポートを通る 2 経路の混雑度合いをそれぞれ (仮想チャネル数 - 空き仮想チャネル数) とする。

上記の混雑度合いを各経路につき、それぞれ合計したものを各経路の混雑度合いとして出力する。Routing Unit 内の Least Congested Route Selector はこの情報を受け取り、使用可能な経路のうち最も混雑度合いの低い経路への出力ポートを選択する。

5. 評価及び考察

Cadence 社の NC-Verilog を用いて、Verilog-HDL で実装した PRC のシミュレーション

*1 実際には 1 ビット予測器ではないが、ここでは 1 ビット予測器のように 1 回のルーティングで予測方向が変わる予測器を 1 ビット予測器と呼ぶ。後述の 2 ビット予測器及び 3 ビット予測器についても同様である。

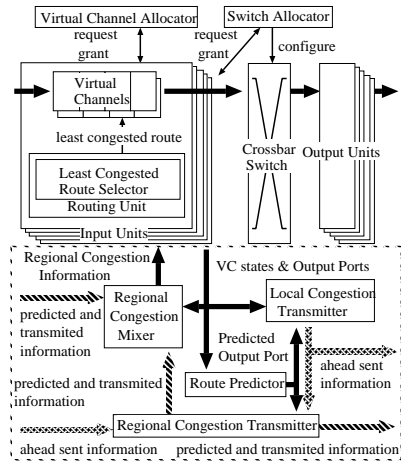


図 4 PRC のアーキテクチャ

Fig. 4 architecture of PRC
及び評価を行った．比較対象には次元順ルーティングを行うルータ (DOR) , 及びローカルの情報のみを用いて空き仮想チャネル数が最大の出力ポートを選択するルータ (Local) を用いた．いずれも投機 SA を行う 3 段パイプラインのルータである．評価で用いたパラメータは表 1 に示す通りである．さらにこのパラメータを以下に示すように変更しながら評価を行った．

- 基本性能：全てのパラメータを表 1 の通りに固定
- ネットワークサイズ：ネットワークサイズを 8×8 2D-Mesh に変更
- バースト性：常に一定の確率でパケットを挿入する非バーストトラフィックを使用

シミュレーションのサイクル数はネットワークサイズが 4×4 の時は 100000 サイクル, 8×8 の時は 40000 サイクル行った．そのうち初めの一割の期間をウォームアップ期間, 終わりの一割をドレインの期間とした．ヘッダフリットがネットワークの入り口のキューに入ってから, テイルフリットが宛先ノードに到着するまでの時間をレイテンシとした．

5.1 基本性能

図 5(a), 図 5(b), 図 5(c) に表 1 に示す通りのパラメータを用いたシミュレーションの結果を示す．Transpose³⁾ は, 負荷の偏ったトラフィックであるため, PRC 及び Local の双方がレイテンシ, スループット共に向上した．Uniform Random は初めから負荷が分散されており, Adaptive Routing を評価する上では悲観的な評価となるトラフィックである．そ

表 1 評価環境
Table 1 evaluation environment

トポロジ	2D-Mesh
ネットワークサイズ	4×4
仮想チャネル数	2
バッファサイズ	2 フリット
フリットサイズ	128 ビット
パケットサイズ	5 フリット
バースト性	平均 4 パケット連続
トラフィックパターン	Transpose, Uniform Random, Bit Complement

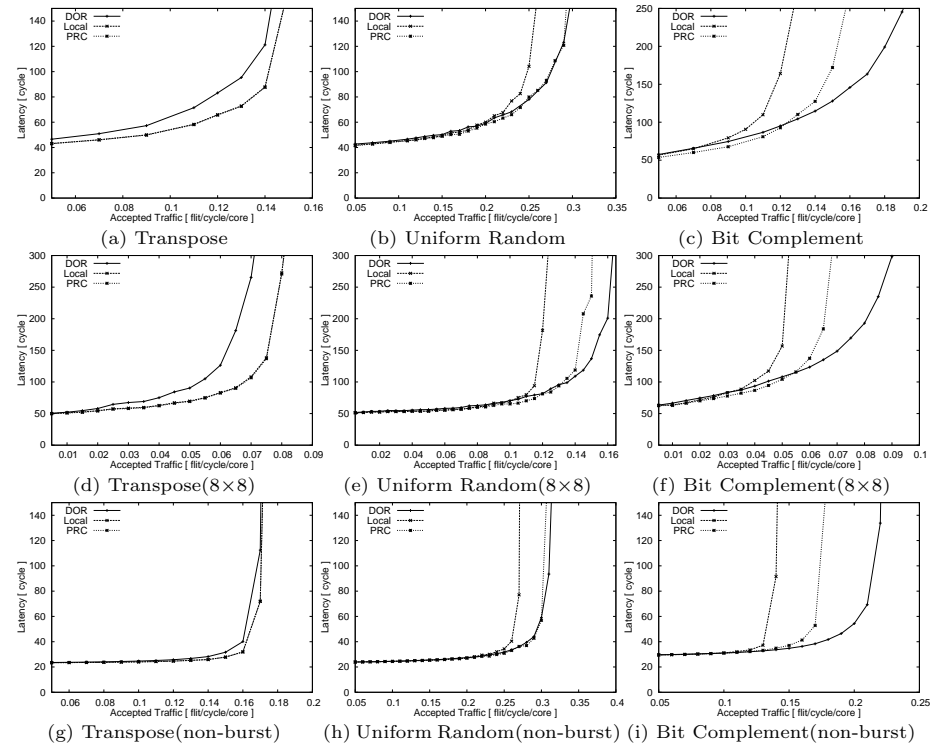


図 5 レイテンシ
Fig. 5 latency

のため, Local のスループットが低下している．本研究で設計した PRC のスループットは DOR と比べてほとんど低下せず, Local と比べて 14.0% 向上した．Bit Complement³⁾ は, DOR を行うと完全に負荷が分散するトラフィックパターンであり, Adaptive Routing の評価としては最も悲観的な最低ラインを示すトラフィックパターンである．そのため, DOR のスループットが最も高く, Local の性能は最も低下した．しかし, PRC のスループットは Local ほど低下することなく, Local と比べて 24.8% 向上した．

5.2 ネットワークサイズ 8×8

図 5(d), 図 5(e), 図 5(f) にネットワークサイズを 8×8 にした場合の各トラフィックのレイテンシを示す．DOR と比べた PRC のスループットの向上率は, Uniform Random 及

表 2 最大動作周波数
Table 2 max frequency

	最大動作周波数 [MHz]
DOR	388(± 0.0%)
Local	365(5.9%)
PRC	365(5.9%)

表 3 実装面積
Table 3 implementation area

	実装面積 [mm ²]	面積の平方根 [mm]
DOR	0.3551(± 0.0%)	0.596(± 0.0%)
Local	0.3573(+ 0.6%)	0.598(+ 0.3%)
PRC	0.3643(+ 2.6%)	0.604(+ 1.3%)

び Bit Complement についてそれぞれ、22.3%及び 31.9%となった。これはどちらもネットワークサイズが 4×4 の時に比べ大きな向上率となった。これは、ネットワークサイズが大きくなったことで、混雑を回避できる経路が増えたためであると考えられる。

5.3 非パーストラフィック

図 5(g), 図 5(h), 図 5(i) にトラフィックにパースト性を持たせないよう、一定の確率でパケットを挿入した場合のレイテンシを示す。DOR と比べて PRC のスループットの向上率は、Uniform Random 及び Bit Complement についてそれぞれ、13.8%及び 26.0%となった。本研究で提案した PRC は予測機構を使用しているため、非パーストラフィックでは予測的中率が低下し、性能が向上しにくい。しかし、パースト性を持たないトラフィックに関しても、性能は悪化することはない。全てのトラフィックにおいて Local と同じ又は向上する結果となった。

5.4 論理合成結果

Synopsys 社の Design Compiler を用いて TSMC 0.13 μm プロセスで論理合成を行った。表 2 に最大動作周波数を、表 3 に動作周波数を 333MHz をターゲットに論理合成した場合の実装面積を示す。但し括弧内は DOR との差を示したものである。最大動作周波数は DOR を行うルータに比べて 6%のみの低下のみとなり、Local と比べると同じ値となった。実装面積の増加は DOR 行うルータと比べて 2.6%、Local と比べて 2.0%のみである。これらは 2 つの予測情報の生成に同一の Route Predictor を共有して使用したことや、各種情報の生成及び合成に用いる演算を単純にしたことによる。また、ルータ間に追加した配線は、先送り情報のための 1 ビットと予測付き伝播情報のための 2 ビットの合計 3 ビットのみであり、配線のオーバーヘッドは小さい。

6. ま と め

本研究では、NoC 上でのロードバランシング及びパケットの衝突を回避するために、局所的な混雑を予測するオンチップアダプティブルータ (On-chip Adaptive Router with Predictor for Regional Congestion : PRC) を設計及び実装した。PRC は、混雑状況を周辺

のルータに伝え、その情報を用いて混雑箇所を回避したルーティングを行う。さらに各入力チャンネルごとに次回出力するであろう出力ポートの予測を行い、この情報をパケットフローの下流に送信することでパケット衝突の回避を行う。

各種パラメータを変えながらシミュレーションを行った結果、Adaptive Routing を行うと性能の上がりやすい Transpose において、全てのパラメータで Adaptive Routing を行わない DOR より良い性能となった。その一方で Adaptive Routing と相性の悪い Uniform Random 及び Bit Complement においても、全てのパラメータで性能をあまり低下させることなく、従来手法である Local と比べた場合、それぞれスループットを平均 17.9%、32.7%向上させた。

また、オーバーヘッドに関して、面積の増加は 2.6%、動作周波数の低下は 5.9%、ルータ間の配線の追加は物理チャンネルあたり 3 本のみで性能を向上させた。

謝辞 本研究は科学技術振興機構 CREST の支援によるものであることを記し、謝意を表す。また、本研究の一部は文部科学省グローバル COE プログラム「環境共生・安全システムデザインの先導拠点」に依るものであることを記し、謝意を表す。

参 考 文 献

- 1) Dally, W.J. and Towles, B.: Route Packets, Not Wires: On-Chip Interconnection Networks, *Proceedings of the Design Automation Conference*, pp.684–689 (2001).
- 2) Wentzlaff, D., Griffin, P., Hoffmann, H., Bao, L., Edwards, B., Ramey, C., Mattina, M., Miao, C.-C., III, J. F.B. and Agarwal, A.: On-Chip Interconnection Architecture of the Tile Processor, *IEEE Micro*, Vol.27, pp.15–31 (2007).
- 3) Dally, W.J. and Towles, B.: *Principles and Practices of Interconnection Networks*, Morgan Kaufmann (2004).
- 4) Dally, W. J. and Aoki, H.: Deadlock-Free Adaptive Routing in Multicomputer Networks Using Virtual Channels, *IEEE Transactions of Parallel and Distributed Systems*, Vol.4, No.4, pp.466–475 (1993).
- 5) Gratz, P., Grot, B. and Keckler, S. W.: Regional Congestion Awareness for Load Balance in Networks-on-Chip, *International Symposium on High-Performance Computer Architecture* (2008).
- 6) Peh, L.-S. and Dally, W. J.: A Delay Model and Speculative Architecture for Pipelined Routers, *International Symposium on High-Performance Computer Architecture* (2001).
- 7) Glass, C.J. and Ni, L.M.: The Turn Model for Adaptive Routing, *Proceedings of International Symposium on Computer Architecture*, pp.278–287 (1992).