

HDL 設計に基づく Recover-x 適応ルータの評価

3H-9

堀田 真貴 林 匡哉 吉永 努 大津 金光 馬場 敬信
宇都宮大学工学部

1 はじめに

並列計算機では、PE(Processing Element)がネットワークを通して互いにメッセージを送受信することにより処理を行なう。ルーティングアルゴリズムは、ネットワークの通信性能に影響を与える一因であり、我々はその実装コストと性能について検討している [3]。

本稿では、ハードウェア記述言語 Verilog-HDL を用いて我々が提案する Recover-x^[4] 完全適応ルータと他の代表的なルータを設計し、各ルータを比較する。そして、その論理合成、シミュレーション結果に基づいて、Recover-x の有効性を示す。

2 ルータの設計仕様

本研究では、ネットワークトポロジーとして2次元トーラスを採用する。また、フロー制御方式としてワームホール方式 (1-flit=1-phit) を採用する。

2.1 ルーティングアルゴリズム

以下に述べる4種類のルーティングアルゴリズムについて比較検討する。また、メッセージの転送は最短経路で行なうものとする。

- (1) Dimension-order: 定められた次元順 (X-Y 順) にメッセージを転送する。必要なバーチャルチャネル (VC) 数は2/ポートであり、ハードウェア構成が簡単だが、ルーティングの自由度がない。(非適応)
- (2) *-channel^[2]: 一般に、各ネットワークポートに1本の適応 VC と2本の非適応 VC をもつことで、エスケープチャネルのバンド幅を大きく確保することができるが、最小構成が大きい。(適応)
- (3) DISHA-seq^[1]: 全ポート共有のデッドロックバッファ(DB)と、各ポートに1本の適応 VC のみで構成できる。ルーティングの自由度は高いが、デッドロック発生率が上がると性能が低下する。(適応)
- (4) Recover-x^[4]: X次元でのみデッドロック回復することで、並列デッドロック回復が可能である。また、各ポートのVC数は、X次元が3本、Y次元が2本で構成できる(適応)

2.2 ハードウェア構成

図1(a)にルータの基本構成を示す。どのルータも、4つのネットワークポート (North, East, West, South Port) と PE インタフェース (PE I/F) を持つ。また、DISHA-seq ではさらに DB を追加する。各ポートは wire で結線する。図中に実線で示した矢印は、Dimension-order に必要な結線であり、適応ルータでは破線の結線を追加する。DISHA-seq ではさらに DB との結線が必要となる。

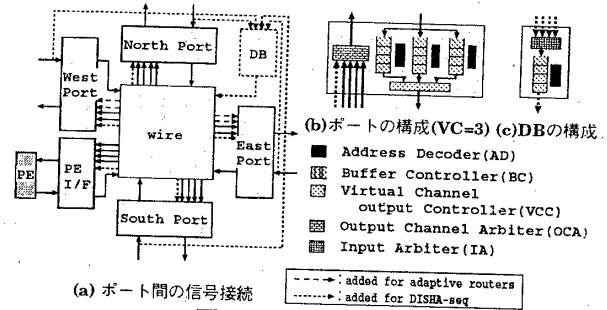


図1: ルータの構成

VC数が3の場合のポートの構成を図1(b)に、DBの構成を図1(c)に示す。各ブロックの機能は以下のとおりである。

Buffer Controller (BC): ネットワークからの入力メッセージを受信し、そのバッファリングを制御する。

Address Decoder (AD): メッセージのアドレスをデコードし、出力候補のポートに出力要求を行ない、出力ポートを選択する。デッドロック回復ルータではメッセージのブロック時間をカウントしてデッドロックの検出を行なう。

Output Channel Arbiter (OCA): 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基にメッセージの出力要求を調停する。

Virtual Channel output Controller (VCC):

OCAからの出力許可によりVCの出力を制御する。

Input Arbiter (IA): DBにおいて、BCの使用状況とトークン獲得の有無によってDBへの入力要求を調停する。

公平な比較をするため、各ルータの全VC数を統一する。東西南北ポートのVC数は3にする。また、PE I/FのVC数はDimension-order、*-channel、Recover-xでは2であり、DISHA-seqではDBがあるため1である。PE I/FのVC数が少ないとネットワーク混雑時にメッセージの出力を抑制し、デッドロック発生率が下がるので、DISHA-seqにとってこれは不利にはならない。

3 論理合成

これらのルータをHDLで記述し、その論理合成結果を表1に示す。なお、論理合成は、クロック制約条件を除き、どのルータも次のような同一条件のもとで行った。シンセサイザ: Synopsys HDL Compiler Version 1998.02
ライブラリ: LSI Logic 0.6 μ m Array-Based Gate Array
バッファ容量: 32-bit \times 4 words / VC

3.1 動作速度による比較

動作速度は、各ルーティングアルゴリズムにおいて必要なVC切り替えの組合せ数とOCAがマルチプレックスするリンク数に依存する。よって、経路選択ロジック、調停回路の複雑な適応ルータの最大クロック速度が非適応ルータよりも低くなる。

表 1: ルータの論理合成結果

Router	最大クロック速度 (MHz)	総面積 (Kgetes)
Dimension-order	71.4	57.2
*-channel	65.3	63.3
DISHA-seq	64.9	68.4
Recover-x	66.7	67.1

また、適応ルータのみで比較すると、*-channel よりも必要な VC 切り替え数が少なく、DISHA-seq よりもマルチプレックス数が少ない Recover-x のクロック速度がもっとも高くなる。

3.2 回路規模による比較

非適応ルータと適応ルータで比較すると、総面積に約 6~11Kgate の差がある。これは、ポート間結線が増加すると共に、マルチプレクサ等のスイッチが大きくなることと、経路選択ロジックが複雑化するためである。

適応ルータ間では、*-channel よりも他の 2 つの総面積が大きい。これは DISHA-seq と Recover-x にはデッドロック検出/回復のための回路があるためである。ただし、これらの回路は Recover-x においては省略可能である。なお、この適応ルータ間の数 K ゲートの面積差は、近年の技術では大きな制約とはならない。

4 HDL シミュレーション

ここでは、これまで述べてきたルータの RTL シミュレーション結果を示す。なお、ネットワークサイズは 5×5 の 25 ノードとし、シミュレータには Cadence 社の Verilog-XL を使用した。

各ルータの動作速度は、表 1 に示した最大クロック速度と仮定する。また、デッドロック回復ルータがエスケープチャネルを使用するまでのメッセージのブロック時間は、DISHA-seq では 32、Recover-x では 4 クロックとする。これは、種々の実験において最良の結果を示した値である。

メッセージ長を 16 バイト固定とし、各メッセージのヘッダフリットが PE を出てから、最終フリットが宛先の PE に到着するまでの時間を計測し、全メッセージの平均をとった。図 2 に負荷が均一でない (Non-Uniform) 通信、図 3 に負荷が均一な (Uniform) 通信の平均レイテンシを示す。なお、横軸はメッセージの生起確率を表す。また、その他のシミュレーション条件については文献 [3] と同様である。

ネットワーク負荷が軽いとき、Uniform 通信では高速動作する Dimension-order のレイテンシが短くなるが、適応ルーティングが有効な Non-Uniform 通信では性能差がほとんどない。ネットワーク負荷が重くなると、デッドロック発生率が上がるため、全ポート共有の DB を 1 つしかもたない DISHA-seq のレイテンシは長くなる。さらに、DISHA-seq ではメッセージのブロック時間が Recover-x よりも長い。Recover-x は、並列デッドロック回復が可能であり、メッセージのブロック時間が短い。そのため、ネットワーク負荷が重くなってもあまり性能低下がみられない。*-channel は、Recover-x よりもルーティングの自由度が低い。そのため、Recover-x に比べて性能が劣る。また、Non-Uniform 通信の方が、適応ルータである Recover-x と非適応ルータの性能差は大きい。

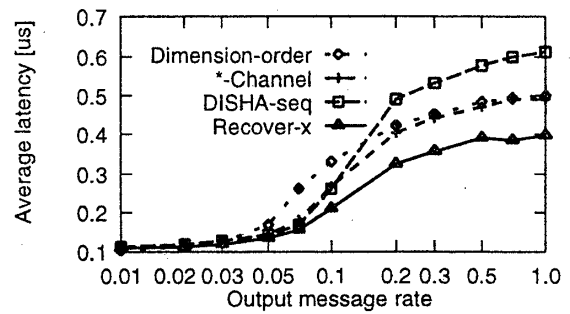


図 2: Non-Uniform 通信

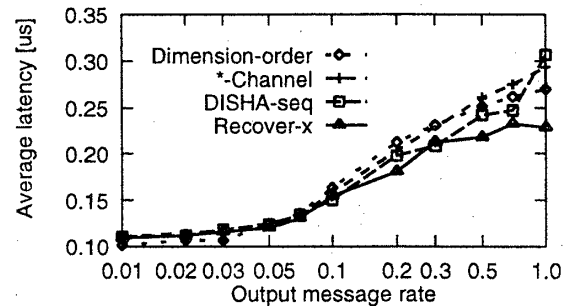


図 3: Uniform 通信

5 おわりに

本稿では、2次元トラスにおける Recover-x のコストと性能について評価した。これより、Recover-x は、従来のルータに比べて構成が簡単であり、高速動作が可能であること、また、高負荷なネットワークでの通信が低レイテンシであることがわかった。

今後は、大規模ネットワーク、並びに、3次元ネットワークでの評価を行なう予定である。

謝辞

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頂いた CAD ツールを使用しており、深く感謝する。

本研究は、一部文部省科学研究費 基盤研究 (C) 課題番号 09680324、基盤研究 (B) 課題番号 10558039、奨励研究 (A) 課題番号 09780237 の援助による。

参考文献

- [1] Anjan K.V., T.M. Pinkston and J. Duato: "Generalized Theory for Deadlock-Free Adaptive Wormhole Routing and its Application to *Disha* Concurrent", *Proc. IPPS*, pp.815-821(1996).
- [2] L. Gravano and G.D. Pifarré, P.E. Berman, and J.L.C. Sanz: "Adaptive Deadlock and Livelock-Free Routing with all Minimal Paths in Torus Networks", *IEEE Trans. on Parallel and Distributed System*, vol.5, No.12, pp.1233-1251(1994).
- [3] 林匡哉, 堀田真貴, 大津金光, 吉永努, 馬場敬信: "HDL 設計に基づく並列計算機ルータの評価", 情報処理学会研報, vol.98, No.70, pp.79-84(1998).
- [4] 林匡哉, 堀田真貴, 吉永努, 大津金光, 馬場敬信: "Recover-x: 2次元トラス用並列デッドロックリカバリ方式の提案", 情報処理学会第 58 回全国大会論文集, 3H-08(1999).