

C-010

仮想出力キューイング方式に基づいた低遅延ルータのFPGA実装・評価

グエン・チュオン・ソン†

Truong Son Nguyen

上原子 正利†

Masatoshi Kamiharako

小柳 滋†

Shigeru Oyanagi

1 はじめに

従来, System-on-Chip (SoC) の IP コア間の通信方式はバスが用いられているが, 近年 IP コアが大量に増えると高性能, 低遅延, スケーラビリティ等を提供できる NoC (ネットワークオンチップ) を幅広く使用する傾向になっている. NoC の主な構成要素であるルータは全体の NoC のパフォーマンスに極めて影響する. 本研究では, NoC の低遅延化を目指した仮想出力キューイング方式に基づいたルータの設計方針を提案する. 提案したルータを, FPGA 上に実装し, 回路規模と動作周波数の面から検討・評価を行う.

2 ルータのアーキテクチャと低遅延化

ルータの基本アーキテクチャを, 図1に示す[1]. ここで, 入力はいくつかの VC (Virtual Channel) というバッファがある. 到着したパケットは, どの出力へ進むかを「Routing Computation」モジュールによって確定する. また, 次ルータでどの VC に到着するかを「VC Allocator」モジュールから決める. そして, 「Switch Allocator」モジュールは全ての入力と出力との接続を調停する. 一般に, ルータの処理プロセスは, flit (flow control digit) レベルで4段階のパイプラインが流れる. それらは図2に示した様に RC (Routing Computation), VA (Virtual-channel Allocation), SA (Switch Allocation) と ST (Switch Traversal) である. 各ステージに1クロックとすると, ルータのレーテンシは4クロックかかり, クリティカルパスとなる. 高性能な NoC を設計するためにこのクリティカルパスを最小することが課題となっている.

2001年に Li-S. Peh により提案されたルータは, Speculation 方式を用いて VA と SA を並列に処理し, パイプラインステージ数を4段から3段に減らした[2]. Speculation が失敗すると, 4段のパイプラインに戻る. また, 2007年に G. Michelogiannakis により提案された手法は次ルータの RC を実行し, Speculation 方式を用いて RC, VA, SA を並列に処理すると4段階のパイプラインを2段パイプラインに改良できた. [3]. さらに, H. Matsutani が2009年に予測ルータを提案した. 予測が当たれば RC, VA, SA を省略でき, 70%に達すると

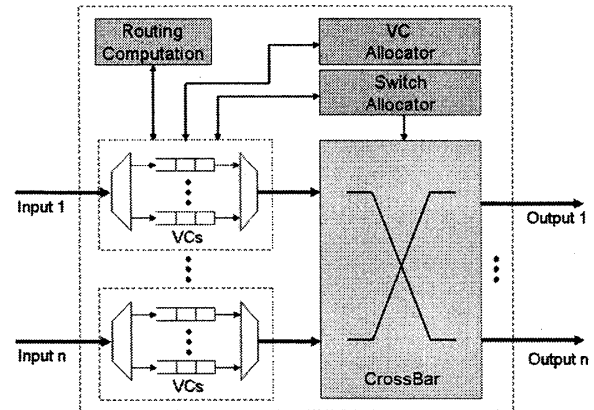


図1: ルータの基本アーキテクチャ

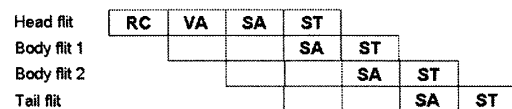


図2: ルータの処理プロセス

パイプラインステージ数は1.6まで削減できた[4]. しかしながら, 手法[2]は, 軽負荷に対して有効であるが, 重負荷には Speculation の失敗が多くて役に立たない. そして, 後者の手法[3]は, 1段のステージが複数になりすぎる. [4]の場合は, 1個のバッファに1つの「Predictor」モジュールが必要であり, 制御ロジック等がオーバーヘッドになっている.

3 提案方式

本研究では, 仮想出力キューイング方式 (VOQ) [5] に基づいたルータアーキテクチャを提案する. VOQ では RC と VA が不要になり, パイプラインステージ数を2段のみに減少できる. 提案したルータのアーキテクチャを図3に示す. ここでは, 出力用の VC を動的に割り当てる代わりに, 各出力に対して1個の VC が固定的に割り当てられる. そうすると, 到着した flit では進行方向の出力と次ルータでの到着 VC を明確に確定できる. そのため, パイプラインは図4に示したように SA と ST のみである. さらに, SA と ST を並列に処理し, パイプラインステージ数を1段のみに削減することも可能である (図4). [2] の VA, SA の並列化と

† 立命館大学情報システム学科

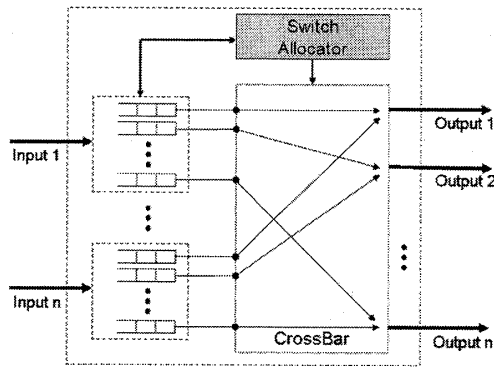


図3: 提案したルータのアーキテクチャ

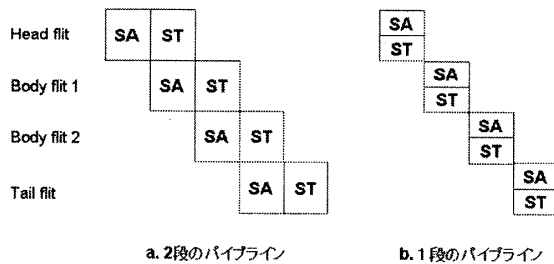


図4: 提案したルータの処理プロセス

[3][4]のRC, VA, SAの並列化に比べて容易に実装できると考えられる。

今回設計したルータは、隣接ルータとつながるためのポートN(北), S(南), E(東), W(西)とIPコアに接続するポートL(ローカル)の5つの入出力をもつ。1つのポートには4個のFIFOがある。そのデータのバンド幅は16ビットであり、FIFOの深さは4である。トポロジはハードウェア開発やシミュレーションにおいては幅広い範囲に応用でき、スケーラビリティ等の特徴を持っている2Dメッシュを利用した。また、2Dメッシュにおけるデッドロックを防止するためXYルーティングというプロトコルを選んだ。

4 評価と考察

今回の実装では、システムデザイン用のXilinx社ISE 10.1iを利用した。回路設計にはVerilog HDLを用い、またシミュレーションにModelSim XE 6.3cを利用した。ターゲットデバイスはXilinx社製のFPGA XC5VFX70Tである。実装結果を表1に示す。その結果、従来アーキテクチャと比べるとパイプラインステージ数は4段から2段に減少することができた。ルータは回路規模の面で277スライスを占め、最大周波数は230 MHzである。さらに、これらの2段を並列に処理し、1段のみに削減することも可能である。その回路規模は、2段

表1: 提案したルータの性能

本デザイン	2段のパイプライン	1段のパイプライン
スライス	277	269
Flip-Flop	347	275
LUT	1070	1037
BlockRAM	10	10
動作周波数	230 MHz	148 MHz
レーテンシ	2-cycle	1-cycle

のパイプラインの場合と比較するとほぼ同じであるが、動作周波数は35.65%下がる。

5 まとめ

本稿では、仮想出力キューイング方式に基づいてルータの処理プロセスのパイプラインステージ数を削減し、ルータを低遅延化できる手法を提案した。ルータを、FPGA上に実装し、回路規模と動作周波数の面から検討・評価した。その結果、従来アーキテクチャと比べるとパイプラインステージ数は4段から1段までに減少することができた。ルータは回路規模の面で277スライスを占め、最大周波数は230 MHzである。現在、従来アーキテクチャとの定量的な比較を行っている。

今後の課題としては、単一のルータだけではなく全体のNoCの実装・性能評価も上げられる。

参考文献

- [1] W. J. Dally and B. Towles. Principles and Practices of Interconnection Networks, 2004.
- [2] Li-Shiuan Peh and W. J. Dally. A Delay Model and Speculative Architecture for Pipelined Routers, 2001. Proc. 7th Int. Symp. on HPCA.
- [3] G. Micheliogiannakis et al. Approaching Ideal NoC Latency with Pre-Configured Routers, 2007. Proc. Int. Symp. on NoC.
- [4] H. Matsutani et al. Prediction Router: Yet Another Low Latency On-Chip Router Architecture, 2009. Proc. Int. Symp. on HPCA.
- [5] N. McKeown et al. Achieving 100% Throughput in an Input-Queued Switch, 1999. IEEE Trans. on Comms., vol.47, no.8, pp.1260-1267.