

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 9 日現在

機関番号：35309

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26730032

研究課題名(和文)再構成型HPCシステムにおけるデバイス間的高速通信基盤の開発

研究課題名(英文)Development of a high speed communication platform for the FPGA-based high performance computing systems

研究代表者

近藤 真史 (Kondo, Masafumi)

川崎医療福祉大学・医療技術学部・講師

研究者番号：90590133

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：本研究では、複数のFPGAを用いた高性能演算基盤を前提に、FPGA間を接続する非同期バスとして、環状のバスを動的に接続する分散型リングセグメントバス(RSB)を提案した。特にその制御を担うアービタについては、FPGA間のバス遅延に基づいて最短経路を構築することで優れた拡張性を実現している。これに基づいた分散型RSBシステムを非同期式順序回路として設計し、4つのFPGA(Xilinx KC705)に対する回路実装を通じて所望の動作を確認した。さらに、実装時の回路遅延を用いてモンテカルロ法に基づいた性能評価を行い、FPGA数とバス遅延が速度性能に与える影響を明らかにした。

研究成果の概要(英文)：Recently, a Field-Programmable Gate Array (FPGA) has been noted as the hardware accelerators for high-performance computing. The distributed ring segmented bus (RSB) which connects an annular bus dynamically was proposed for the asynchronous bus between FPGA. In order to improve extensibility, the arbiters which control the divided buses connect the shortest path based on the wire delay. The distributed RSB system was designed as asynchronous circuits, and its operation was confirmed by implementing to four FPGA (Xilinx KC705). In addition, the effects of the FPGAs and wire delay on performance was clarified through the Monte Carlo simulation using implemented wire delay.

研究分野：計算機工学

キーワード：GALSシステム 非同期式回路 アービタ リングバス FPGA

1. 研究開始当初の背景

近年、スーパーコンピュータ等の高性能計算機システム(High Performance Computing : HPC)では、複数の FPGA を用いてアプリケーション毎に専用回路を動的に構築する FPGA-based HPC (FbHPC)が注目されている。この種のシステムでは、複数の FPGA が相互にデータ通信を行う必要があるが、各 FPGA は個別のクロックパルスに同期して動作している。したがって、FbHPC はその構造上、FPGA 間のデータ通信は非同期的に、FPGA 内のデータ通信は同期的に行う、いわゆる大域非同期局所同期(Globally Asynchronous Locally Synchronous : GALS)システムとして構成される。

GALS システムの速度性能は、同期通信に比して非同期通信に係るオーバーヘッドが大きく、これを担う通信トポロジがシステム全体の速度性能に強く影響する。この種の通信トポロジはバス型とネットワーク型に大別されるが、FbHPC のようなオフチップシステムにおいては、FPGA 間の配線遅延のためルータ間のハンドシェイク処理に係るオーバーヘッドが大きく、速度性能が著しく劣化する。したがって、高性能な FbHPC を実現する上では、バス型に基づいてその転送効率の改善を図ることが重要となる。

一方、研究代表者は、GALS システムに対する新たな非同期バスの構成法として、リングセグメントバス(RSB)を提案している。RSB は、環状のバスを複数のセグメントに分割した上で、データ転送に関連するセグメントのみを動的に接続する形態であり、優れた並列転送を実現できる。しかし、バスの制御を担うアービタが全てのセグメントの制御を集中的に行っていることから、クロックドメイン(CD, FbHPC では FPGA に対応する)数の増加に伴って処理時間が増大する上に拡張性に欠けるという、実用化に向けた課題が残されていた。

2. 研究の目的

本研究課題では、FbHPC に対する非同期バスの構成法として RSB を採用し、より高性能な FbHPC 基盤の実現を目指すものである。まず、拡張性の改善を図るため、(1)集中型アービタの機能をそれぞれの FPGA に分散化し、非同期バスの制御機能を備えた FPGA モジュールを構成する。次に、FPGA 内の同期バスにも RSB を適用し、(2)階層的な RSB 構造を採ることで更なる速度性能の改善を図る。そして、(3)複数の FPGA に対する RSB の実装を通じてその性能を明らかにする。

3. 研究の方法

本研究では、FPGA 間の非同期バス、FPGA 内部の同期バスと順に RSB を適用するため、各バスの制御を担うアービタの機能が重要となる。以下、各アービタの機能を予め簡単に説明する。

ローカルアービタ(LArb) : 同期バスの使用権に関する競合を処理する。また、競合に勝った構成要素が異なる CD との通信を要求している場合、GArb に非同期バスの使用権を要求する。

グローバルアービ(GArb) : 非同期バスの使用権に関する競合を処理する。セグメントの使用状況に基づいて、通信経路を構築可能な要求を適宜承認し、セグメント間を接続するバスコネクタを制御して通信経路を構築する。

(1) 分散型グローバルアービタの設計

グローバルアービタの機能を各 FPGA に分散し、RSB に基づいた FbHPC 基盤における拡張性の向上を図る。

競合処理機能の分散 : RSB の構造に着目し、環状に接続されたセル上を使用権(トークン)が循環する形態であるリングアービタを採用する。具体的には、各 FPGA にセル(RArb)を配置し、環状バス上に使用権を循環させることで競合処理機能を分散する。特に RSB では多資源(セグメント)に対する競合処理が必要となるが、これについては、トークンを得たセルが転送経路を構築・解放できるか否かに関わらず、これに係る一連の処理が完了した後にトークンを伝搬する。これにより、セル上を常にトークンが循環するため、構築可能な要求が生起している限り、複数の転送経路が適宜構築されることとなる。

最短経路構築機能の分散 : 既存の集中型 GArb では、現在のセグメントの使用状況をレジスタにより集中管理しており、これに基づいて最短経路の構築判定やバスコネクタの制御を行っている。しかし、アービタの分散化に伴って他のセグメントの使用状況を管理できない上に、経路上のバスコネクタを直接制御することができない。このアービタの分散化に伴う構造的な問題については、環状の非同期バスを利用してセグメントの使用状態の確認処理と構築処理を伝搬的に行うことで解決する(図 1)。まず、転送元の RArb は時計周りと反時計周りの両経路に対してバスの使用要求信号を送り、これを受けた転送先の RArb は先に要求が到達した経路に対してのみ使用承認信号を返す。そして、承認信号

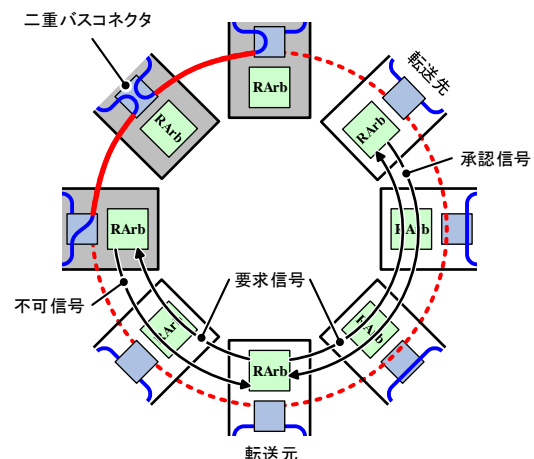


図 1 分散型 GArb の概要

がセグメントを接続しながら転送元 RArb に返送されることにより最短経路が構築される。また、使用不可信号が返送された場合は、その経路の一部が既に使用中であるものとして処理する。

(2) RSB の階層化

CD 内の同期バスに対して RSB を適用することにより、階層的な RSB 構造を構築する。
バスコネクタの二重化: 同期バスにバスコネクタを挿入し、同期ブロック内に RSB 構造を構築する。特に非同期バスとの接続を担うバスコネクタについては、非同期バスと接続可能な同期バスが 2 経路に拡張されるため、バスコネクタを二重に設けてそれぞれの経路に割り当てる(図 1 参照)。

ローカルアービタの機能拡張: 環状となった同期バスの競合処理および制御機能については、同様の機能を備えた既存の集中型 GArb を同期式回路として設計することにより実現する。ここで、GArb に対する非同期バスの要求機能については、要求を発する前に上述の二重バスコネクタに至るまでの同期バスを獲得しておく必要がある。これについては、二重バスコネクタを通信先 CD に見立てて競合処理を行い、それに至るまでの同期バスを獲得した上で、非同期バスの使用要求を発するように機能を拡張する。

グローバルアービタの機能拡張: 非同期バスに接続可能な同期バスが 2 経路存在するため、異なる二つの構成要素が同じ転送経路を要求するとデッドロックが生じる可能性がある。階層化に伴うこの問題については、同一転送先を要求している構成要素に対する競合処理を別途行った上で、それに勝った要求のみを非同期バスの使用権に関する競合処理に参加させることで対応する。

(3) 複数の FPGA を用いた RSB の実装

RSB の FPGA 実装にあたって、実装対象となるデバイスの選定を行う。

分散型 RSB において FPGA 間で交わす信号を整理すると、バスの制御を担う RArb 間で非同期的に交わされる制御信号と、転送経路の構築後に FPGA 内の構成要素間で交わされるデータ信号に大別される。前者の制御信号については、バスの構築頻度を勘案すると高速な伝送速度は冗長となる。よって本研究では、制御信号に対して汎用 IO を用いたシングルエンド伝送を前提とする。後者のデータ信号については、大容量のデータ転送を想定し、SMA ケーブルを用いた高速な差動伝送により実現する方針を採る。以上の異なる伝送方式を実現するため、本研究では Xilinx FMC 拡張カード XM105 (汎用 IO×40)と XM104 (SMA×8)により、高性能汎用 FPGA ボード(KC705)の IO を拡張し、これを単位として 4 つの FPGA からなる分散型 RSB の実装環境を構築する。

4. 研究成果

(1) 分散型グローバルアービタの設計

提案した分散型 GArb の動作を信号遷移グラフ(STG)で表現し、それを入力として非同期式回路合成ツール Petrify を用いて論理合成を行った。そして、これにより得られた回路を統合し、XilinxISE 9.2 を用いて設計を行った。Xilinx Spartan3E (XC3S1200E-FT256)を対象に、CD 数 8 の場合におけるアービタのシミュレーションを行った結果、複数の転送経路の構築から解放に至るまで所望の動作を全て確認することができた。

続いて、設計した分散型 GArb と既存の集中型アービタに対して FPGA における使用資源量の比較を行った結果を表 1 に示す。同表より、分散型アービタは CD 数の増加に伴って所要面積が増加するものの、その増加割合は集中型のそれに比して緩やかである。特に、CD 数 16 においては Slice を約 1/2、LUT を約 1/3 と大幅に面積を低減できることを確認した。

表 1 アービタの資源量

	CD 数	Slice	FF	LUT
集中型	4	14	16	55
	8	53	32	167
	16	261	64	612
分散型	4	33	20	65
	8	61	40	121
	16	112	80	223

(2) RSB の階層化

階層化を施した RSB を VHDL により機能レベルで設計し、シミュレーションによりその動作確認を行った。設計環境および対象デバイスは(1)と同様である。対象とした RSB の構成は、CD 数 4、各 CD に内蔵される構成要素数 4 である。また、転送先については、階層化特有の並列転送(図 2)を想定し、 $C_{12} \rightarrow C_{13}$ (Trans1: 同期転送)、 $C_{11} \rightarrow C_{41}$ (Trans2: 非同期転送)、 $C_{14} \rightarrow C_{21}$ (Trans3: 非同期転送)とする。なお、 C_{ij} は i 番目の CD に内蔵される j 番目の構成要素を表す。以上の動作仕様におけるシミュレーション結果(図 3)より、 $4 \mu s \sim 5 \mu s$ にかけて上述した 3 つのデータ転送が並列的に実行されていることが確認された。

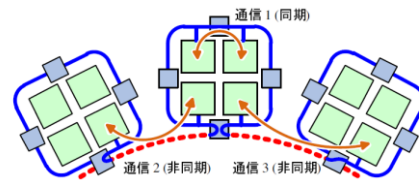


図 2 階層化に伴う並列転送の拡張

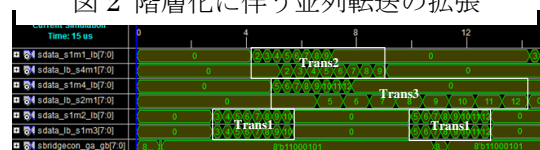


図 3 シミュレーション結果

(3) 複数の FPGA を用いた RSB の実装

構築した RSB の実装環境(図 4)に対して分散型 GArb の実装を行った。なお、これを構成する RArb は非同期式回路であるため、その動作を保証する遅延仮定が破綻することのないように、主要な信号全てに配置・配線制約を施している。Xilinx ChipScope を用いて RArb の内部信号を直接観測した結果、(1)と同様に分散型 RSB における所望の動作全てが実際の回路基板上でも確認された。

CD 数が増加した場合における性能を明らかにするため、上述した実装環境における RArb およびバスの遅延値を実測・抽出し、モンテカルロシミュレーションにより性能評価を行った。CD 数 4, 8, 16 におけるスループットの評価結果をそれぞれ図 5 (バス遅延 4.45ns)および図 6 (バス遅延 20.50ns)に示す。なお、横軸の ReqE は指数分布に従う要求の発生頻度である。この結果より、CD 数の増加、即ち経路構築時に経由するバス数の増加に伴って、その遅延値による性能劣化が顕著になることが示唆された。

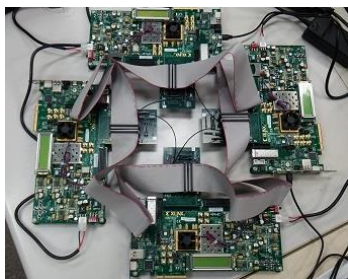


図 4 実装環境

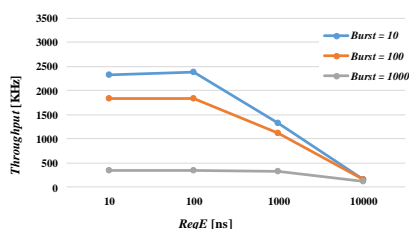


図 5 スループット (バス遅延 4.45ns)

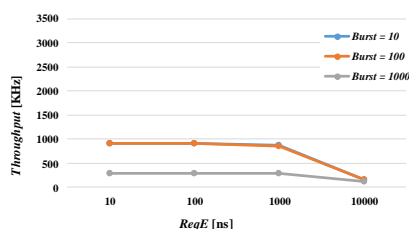


図 6 スループット (バス遅延 20.05ns)

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1 件)

[1]近藤真史, 横川智教, 佐藤洋一郎, 有本和民, “GALS システムの最適設計化を目的とした性能評価ツール,” コンピュータソフトウェア, Vol.32, No.4, pp.115-130 (Dec.2015) 査読有

[学会発表] (計 9 件)

[1]中野涼太, 近藤真史, 横川智教, 佐藤洋一郎, 有本和民, “複数の FPGA に対する分散型非同期式アービタの実装と評価,” LSI とシステムのワークショップ 2016, 2016 年 5 月 16 日~17 日, 東京大学 (東京都・目黒区)

[2]中野涼太, 荒木英友, 近藤真史, 横川智教, 佐藤洋一郎, 有本和民, “複数の FPGA に対する分散型非同期式アービタの実装と評価,” 2016 年電子情報通信学会総合大会, ISS-SP-151, 2016 年 3 月 15 日~18 日, 九州大学 (福岡県・福岡市)

[3]森光勇太, 横川智教, 近藤真史, 宮崎仁, 佐藤洋一郎, 有本和民, “Verilog-HDL による大規模ハードウェア設計の検証支援ツールの開発,” 電子情報通信学会研究技術報告 VLD, Vol.115, No.465, pp.1-6, 2016 年 2 月 29 日~3 月 2 日, 沖縄青年会館 (沖縄県・那覇市)

[4] 中野涼太, 近藤真史, 横川智教, 佐藤洋一郎, 有本和民, “分散型アービタに基づいたリングセグメントバスシステムの性能評価,” 第 66 回電気・情報関連学会中国支部連合大会, pp.378-379, 2015 年 10 月 22 日, 山口大学 (山口県・宇部市)

[5]森光勇太, 横川智教, 近藤真史, 宮崎仁, 佐藤洋一郎, 有本和民, “大規模ハードウェア IP への記号モデル検査の適用事例,” 電子情報通信学会研究技術報告 VLD, Vol.115, No.87, pp.31-36, 2015 年 6 月 17 日~18 日, 小樽商科大学 (北海道・小樽市)

[6]小田桐由樹, 近藤真史, 横川智教, 佐藤洋一郎, 有本和民, “リングセグメント型 GALS システムを対象とした分散型アービタの設計,” 2015 年電子情報通信学会総合大会, ISS-SP-174, 2015 年 3 月 10 日~13 日, 立命館大学 (滋賀県・草津市)

[7]赤利昌紀, 小田桐由樹, 近藤真史, 佐藤洋一郎, 横川智教, 有本和民, “分散制御型リングセグメントバスシステムの構成法,” 第 65 回電気・情報関連学会中国支部連合大会, pp.378-379, 2014 年 10 月 25 日, 福山大学 (広島県・福山市)

[8]近藤真史, 横川智教, 佐藤洋一郎, 有本和民, “リングセグメント型 GALS システムの階層化,” 電子情報通信学会研究技術報告 VLD, Vol.114, No.231, pp.19-24, 2014 年 10 月 2 日~3 日, 東北大学 (宮城県・仙台市)

[9]小田桐由樹, 赤利昌紀, 近藤真史, 横川智教, 佐藤洋一郎, 有本和民, “リングセグメント型 GALS システムを対象とした分散型非同期式アービタ,” 電子情報通信学会研究技術報告 VLD, Vol.114, No.123, pp.237-242, 2014 年 7 月 9 日~11 日, 北海道大学 (北海道・札幌市)

6. 研究組織

(1)研究代表者

近藤 真史 (KONDO, Masafumi)

川崎医療福祉大学・医療技術学部・講師

研究者番号: 90590133