

Low Latency On-Chip Networks through Compression and Multicasting

その他のタイトル	圧縮とマルチキャストを用いた低遅延オンチップネットワーク
学位授与年月日	2014-03-24
URL	http://doi.org/10.15083/00006809

審査の結果の要旨

氏 名 和 远

本論文は「**Low Latency On-Chip Networks through Compression and Multicasting**」と題し、6つの章から構成されている。半導体技術の微細化に伴い、同一VLSI上に多数のプロセッサコアを搭載するメニーコアアーキテクチャが広く使われており、搭載されるプロセッサコアは今後さらに増加すると予想されている。その場合、プロセッサコア間を接続するネットワークの性能、特にプロセッサコア間のデータ転送に要する遅延時間(latency)の短縮が重要になる。本論文では、チップ上のネットワークを構成するスイッチに焦点を当て、スイッチにおけるデータ圧縮で転送量を削減する手法、スイッチにおいて最も時間を要する転送先の判断と転送路の確保を高速化する手法として、転送先を予測する手法、ならびに、転送先を予測せずに転送を開始する手法を提案し、これらの提案手法により、プロセッサコア間の転送時間を短縮し、VLSI全体の性能を向上できることを明らかにしている。

第1章「**Introduction**」では、本論文の背景と目的、および本論文の構成を述べている。メニーコアアーキテクチャで想定されるチップ上のネットワークの重要性、プロセッサコア間の latency に影響を与える要因、その要因の中で、本論文が扱うオンチップネットワークのスイッチが与える影響の分析と、本論文が解決しようとする問題の定義を明らかにしている。

第2章「**Background: Low Latency Techniques for On-chip Networks**」では、研究の背景として、VLSI上のオンチップネットワークに関して、これまでになされたスイッチ構成手法、およびネットワークの latency の短縮手法に関する研究をサーベイしている。

第3章「**Latency Reduction through Traffic Compression**」では、データの圧縮に着目し、圧縮による得失利害の整理、およびメリットを最大化するための手法を提案している。圧縮の利点はデータ量が減ることによる遅延時間の短縮と、ネットワークの混雑緩和による遅延時間の短縮であり、欠点はデータの圧縮・復元のオーバーヘッドによる遅延時間増加である。そこで、この欠点を顕在化させずに利点を活用する方法として、オーバーヘッドの少ない選択的圧縮を

することを提案している。具体的には、圧縮できない場合は圧縮しない、転送路のバンド幅が狭いところに限って圧縮をする、この2つを同時に用いる、という3つの手法を提案し、その有効性をシミュレーションによる評価で提示している。

第4章「**Latency Reduction through In-router Multicasting: Predict-more Router**」では、ルータ内のスイッチにおいて最も時間を要する転送先の判断と転送路の確保を高速化する手法として、転送先を複数予測し、予測された転送先全てに同時に転送する方式を提案している。これまでも転送先を予測する手法は多く研究されていたが、いずれも、転送先は1カ所しか選択できないという仮定のもと、どれだけ転送先を正確に予測するか、を主眼にしていた。本論文は、複数個所に転送しても、ルータ内に当該データがある期間は、間違っただけの転送先へのデータを容易に無効化できるということに着目し、その無効化機構を実現する構成手法、ならびに、転送先を複数予測し同時転送することで、ネットワークの **latency** を短縮し性能を向上できることを明らかにしている。

第5章「**Latency Reduction through In-router Multicasting: McRouter**」では、第4章の手法をさらに推し進め、転送先を予測することなく、ルータ内のスイッチにおいて全ての転送先を候補としてデータを転送し、正しい転送先が判明してから間違っただけの転送先のデータをルータ内で無効化する手法を提案している。この手法は、ルータ内部で不要なデータ転送を行うため、トラフィックが多いときはルータ内でデータ転送の衝突が頻発し性能が低下する可能性がある。提案する手法は実際のプログラムを実行した場合にはトラフィックは極めて低いため衝突による性能低下はほとんどない、という着眼に基づくものである。評価では、実際のプログラムに適用した場合に提案手法により性能が向上することを明らかにしている。

第6章「**Conclusion**」では、以上の成果を要約している。

以上を要するに、本論文はメニーコアアーキテクチャの性能向上を目的とし、プロセッサコア間ネットワークのデータ転送に要する遅延時間(**latency**)を短縮する手法を提案しその有効性を明らかにしたもので、非常に意義がある研究であり、その成果は工学的に貢献するところが大きいと考えられる。

よって本論文は博士（工学）の学位請求論文として合格と認められる。