超広帯域光ネットワークにおける分散型マルチプロトコルルータ

林　秀樹*　岩田　誠*　孔　令杰　米田　進**　寺田　浩詔***

***高知工科大学**
〒782-8502 高知県香美郡土佐山田町宮ノ口 185
Tel:0887-53-1020, FAX:0887-57-2220
036013c@gs.kochi-tech.ac.jp,iwata@info.kochi-tech.ac.jp,terada@info.kochi-tech.ac.jp
****日本テレコム株式会社　情報通信研究所**
〒104-0032 東京都中央区八丁堀 2-9-1
Tel:03-5540-8493,Fax:03-5540-8485
{kong,yone}@japan-telecom.co.jp

あらまし　本稿は，飛躍的な進展を続けている光通信技術を背景に，真に自由な通信空間を提供可能なパケットルータ処理装置の基本的な構成法を提案している．本構成法では，超広帯域光ネットワークの構成要素となり，かつ，多様なプロトコルの複雑な処理をソフトウェアにより柔軟に行うことを目的に，光技術を補完する電気的なルータ処理装置として，多重処理性能に優れたデータ駆動型プロセッサを適用している．本提案方式の初期的性能特性評価の結果，その実現可能性が確認され，現行の IPv4/v6 パケット，および，ATM セルに対するヘッダ処理を同時に多重処理しても，全く性能が劣化せずに，優れた緩衝特性を有することが確認された．

キーワード　動的データ駆動方式，マルチプロトコル，ルータ

A Distributed Multiple-protocol Router for All-Optical Networks

Hideki Hayashi*　Makoto Iwata*　LingJie Kong**　Susumu Yoneda**　Hiroaki Terada*
*Kochi University of Technology
185 Miyanokuchi, Tosayamada-cho, Kami-gun, Kochi, 782-8502 Japan
Tel:+81-887-53-1020, Fax:+81-887-57-2220
036013c@gs.kochi-tech.ac.jp,iwata@info.kochi-tech.ac.jp,terada@info.kochi-tech.ac.jp
**Information & Communication Lab., Japan Telecom Co., LTD.
2-9-1 Hatchobori, Chuou-ku, Tokyo, 104-0032 Japan
Tel:+81-3-5540-8493, Fax:+81-5540-8485
{kong,yone}@japan-telecom.co.jp

Abstract　We propose a novel packet router network architecture that would freely utilize a communication space and fully take advantage of the rapidly progressed photonic technologies. This paper applies the data-driven technology that would show a prominent performance for multiple protocol processing at a packet router. Since this technology flexibly and simultaneously operates multiple protocols controlled by software, the speed of optical technologies and transmitters can be realized. The feasibility of simultaneous multiple protocol processing in terms of IP v4, v6, and ATM is demonstrated, and even when those protocols are multiplexed together, there was no degradation of their performance.

Keywords　dynamic data-driven scheme, multi-protocol, router

1. はじめに

ファイバ内に非常に多数の物理的空間あるいは仮想自由空間が見え，あらゆる形式の情報を疎通できる柔らかなネットワークの構築は，近年の光通信技術の進展に伴い，大きな課題になりつつある[1]．このようなネットワークの構築のため，超高速光伝送リンクを多数収容して，多様なサービスや新しいプロトコルを QoS 要求に応じて柔軟に処理できる，超高速ソフトウェアスイッチの実現が望まれている．また，近年の伝送システムの大容量化に伴い，遅延時間がネットワーク性能に及ぼす影響は相対的に益々大きくなっている．このため，要求されたプロトコルのパケットをそのまま疎通させ，ネットワーク内あるいはノード内で生起する様々な遅延時間の影響を局所化できる分散型動作モードを徹底的に追究した，プログラマブルなルータの実現が必須である．

動的データ駆動方式は，これらの要件に合う優れた特徴を有している．すなわち，複数のデータ組の処理を完全に独立して実行できる自律分散型の処理原理で動作し，特別な文脈切替機構やスケジューリング機構なしに，複数または単一のプログラムを同時に多重処理が可能である[2]．このため，異種プロトコル，あるいは，速度の異なる複数のパケットストリームをそのまま受理して，オーバヘッドなくソフトウェアで並列に処理することが可能である．

本稿では，この動的データ駆動型処理方式の導入によって，様々な形式のパケットをソフトウェアで処理でき，かつ，スケーラブルに性能向上が可能な分散型マルチプロトコルルータの基本構成を提示すると共に，そのデータ駆動型実現法を述べる．また，本実現法が多重処理に関して高い過負荷耐力を示すことを定量的に明らかにする．

2. 分散型マルチプロトコルルータ

本稿で述べる分散型マルチプロトコルルータは全光ネットワークにおける高機能ノードとして構想され，光では実現困難な複雑な処理のみを，電気的に柔軟に実現することを目指している．本章では，このルータの基本構想とその核となる汎用パケット処理エンジンの要件を示す．

2.1 基本構想

分散型マルチプロトコルルータは，全光ネットワークによる仮想自由空間的な通信を実現するために，光のままでの情報疎通も可能とし，かつ，光のまま通信を行うには困難な複雑な機能（サービス/プロトコル）処理のみを電気的に行う構成を採る．具体的には，図１に示すように，単一または複数の光伝送路を収容する多数のルーティングプロセッサ RP(Routing Processor)が光スイッチ OSU(Optical Switch Unit)により相互接続する形態を採り，例えば波長分割多重方式等の論理的メッシュスイッチ機能を用いて, RP 間の全光パケットスイッチが実現される．

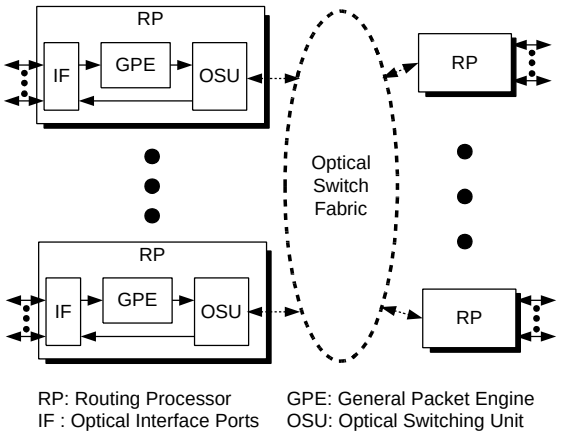


図 1 分散型マルチプロトコルルータの基本構成

汎用パケット処理エンジン GPE(Generalized Packet Engine)は，光伝送路を収容するポート群 IF および OSU を介在して電気変換されたパケットの複雑なヘッダ処理等を柔軟に実現する．

2.2 パケット処理エンジン

今後，ネットワークで扱われるサービスやプロトコルは益々多様になる．このため，GPE には，これらの多種多様なプログラムの多重処理が要求されるため，超高速プログラマブルプロセッサの導入が必須である．

動的データ駆動処理方式は，優れた多重処理性能と，パイプライン処理との高い親和性のため，メモリアクセスやプロセッサ間通信に起因する遅延時間の影響を受けず，システム全体の処理率を最大に維持できる特徴がある．さらに，擬似一致(psude-coincidence)回路による自己タイミング型パイプライン機構[5]の導入により，省電力・超並列処理性能に加えて，負荷変動に対する緩衝能力をシステム全体に分散して付与でき，一時的な過負荷に対してもオーバフローせずに処理を継続できる耐性を備えている. この方式を活用して既に，映像信号処理向けではあるが, VLSI マルチプロセッサ DDMP (Data-Driven Multimedia Processor)が実

用化され，消費電力あたりの処理性能が既存のノイマン型 DSP チップに比べて格段に高いことも実証されている[3],[4].

このような特性は，マルチスレッドアーキテクチャを基礎にした従来型ネットワークプロセッサ[6],[7]にはない優れた特性である．このことから，GPE の実現法に自己タイミング型のデータ駆動型マルチプロセッサを導入した.

動的データ駆動方式では，データ駆動パケットに付加したタグ情報により，各パケットに関連する文脈を識別しながら処理を行うため，多重処理時でも，オーバヘッド損失なく，本来の処理性能を発揮できる．この特性を活用するために，GPE では次のような固有のタグ情報を用いている．

行き先ノード情報(Dest. Node)：各プロトコル種別に対応したプログラム情報を識別する情報.

コンテキスト識別子 CID：或るプロトコル *i* が対象とする或るパケット P*i*，すなわち，同一プログラム内での文脈を識別する情報.

世代識別子 GID：P*i* 中の断片データワード，すなわち，同一文脈内でのデータ駆動パケット系列の順序を識別する情報.

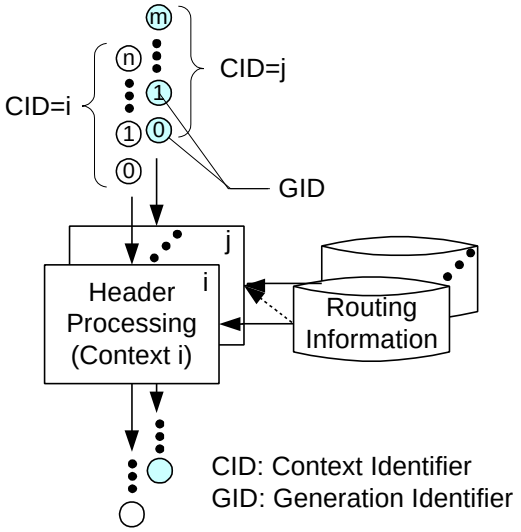


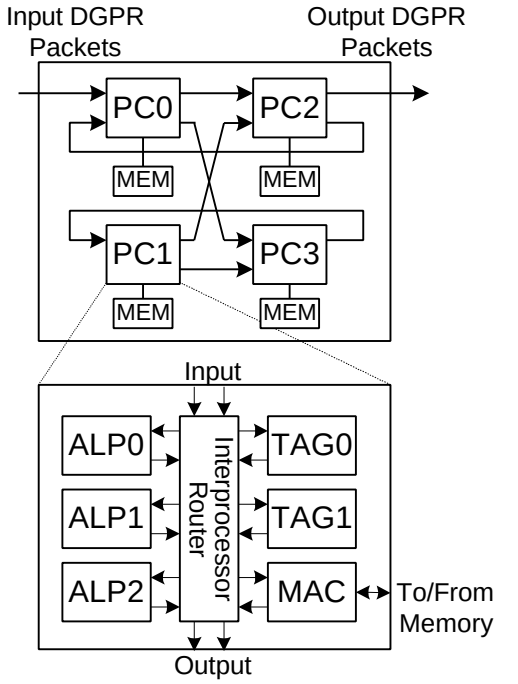
図 2 データ駆動並列多重処理

GPE では，これらの識別子をデータおよび演算コードに付加することによって，図２に示す多重処理動作を可能にしている.

2.3 データ駆動型 GPE の実現法

自己タイミング型データ駆動マルチプロセッサによる GPE アーキテクチャの実現法を図３に示す．本実現法では，数種の専用化した命令セットを有するナノプロセッサ nPE (nano-Processing Element) を階層的に相互接続した，負荷・機能分散型マルチプロセッサ構成を採っている．この nPE はデータ駆動型プログラムを解釈実行するプロセッサ・コアであり，整数/論理演算専用の ALP-nPE，データ駆動方式固有のタグ処理専用の TAG-nPE，テーブル参照/更新専用の MAC-nPE の 3 種類が用意されている.

各 nPE は相互結合網(Interprocessor Router)で接続されているため，データ駆動パケットの行き先ノードが同一 nPE でも異なる nPE であっても同一コストでプログラムを実行でき，スケーラブルな性能向上が図れる．つまり，特別なマルチプロセッサ・スケジューリング機構を必要としない，既存のマルチスレッドアーキテクチャには見られない特性を持っている.



PC i : Processor Cluster
ALP i : Arithmetic and Logical Processing nanoprocessor
TAG i : TAG manipulation nanoprocessor
MAC : Memory Access Control nanoprocessor

図 3 データ駆動型 GPE の例

一方，VLSI チップ上に搭載する各 nPE については，対象とするプロトコル処理に応じて最適な台数で構成することが望ましい．したがって本研究ではまず，IPv4，IPv6，ATM の 3 種類のプロトコル処理をそれぞれ実行したときに駆動されるプリミティブ演算の動的な分布を事前評価した．その結果，整数/論理演算，タグ処理，テーブル参照

/更新に関する演算の比率が，約 3:2:1 になることが確認されたため，図３のマルチプロセッサ構成例では，3 種類の nPE の数をそれぞれ 3，2，1 としたクラスタを単位として拡張している．

3. 性能特性評価

本章では，GPE 実現のための評価目的として，
i) 動的データ駆動方式によってスケジューリングなく多重処理が可能なこと，および，
ii) 自己タイミング型パイプライン機構の自己緩衝能力によって過負荷時の耐性を付与可能なこと，
をシミュレーション評価した結果を示し，本実現法の有効性を明らかにする．

3.1 評価手法

提案実現法の多重処理特性を評価するために，まず，GPE への入力パケットの到着率が一定であるとして，単一リンク単一プロトコルの処理性能を測定し，この性能を評価基準値とした．この基準に基づき同一/異種プロトコル多重処理時の緩衝能力を評価した．性能評価対象は，IPv4，IPv6，ATM のパケットヘッダ処理（通常時の平均パケット処理性能評価のため，オプション処理等は含まない）である．次に，入力パケットの到着率に変動がある場合の GPE の自己緩衝能力を評価した．

これらの測定には，動的データ駆動方式による多重処理動作と，自己タイミング型パイプラインが有している「データ流量の変動に対する緩衝能力」を含めた，総合的な性能特性評価が可能であり，マルチプロセッサ構成の負荷分散効果の測定できる構成(図３)をとるシミュレータを用いた．

3.2 入力パケット到着率一定の状況での評価

(a) 単一リンクの評価

本評価では，多重処理性能の特性評価の基準値となる単一リンク･単一プロトコル処理について，IPv4/v6，ならびに，ATM の最大処理性能を測定した．その結果，信号処理用命令セットしか持たないデータ駆動型マルチプロセッサチップ[3]でも，それぞれ 7.9，6.8，19.0 MPPS(Packets/sec.)の性能(注 1)を達成することが判った．SONET リンク速度に換算するとそれぞれ，3.0，3.7，8.1 Gbps となる(注 2)．この結果から，命令セットをパケット処理に合わせて最適化すれば，倍以上の性能を達成できることが期待される．

(b) 同一プロトコル多重処理

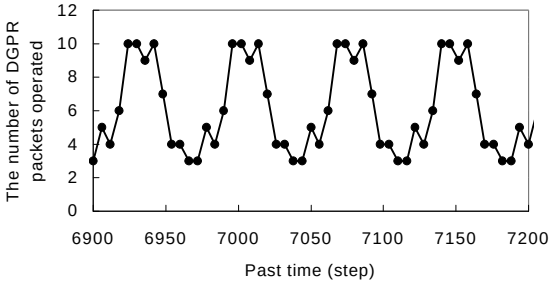


図 4　nPE の負荷変動例

一般に，データ駆動型プロセッサ内の負荷は静的なプログラム構造ならびに入力トラフィックの到着率に依存した周期で変動する（図４）．

複数リンクから到着する同一プロトコルのパケットストリームを多重する場合，図のような負荷変動パターンをスケジューリングせずに重畳して多重処理が可能である．ただし，負荷変動パターンの位相差により，相補的に負荷が平滑化される状況と，共振的に倍増した変動を自己緩衝作用で吸収して実行する状況が起こる．前者はスループットが向上し，後者はスループットをやや落としながら処理を継続すると考えられる．

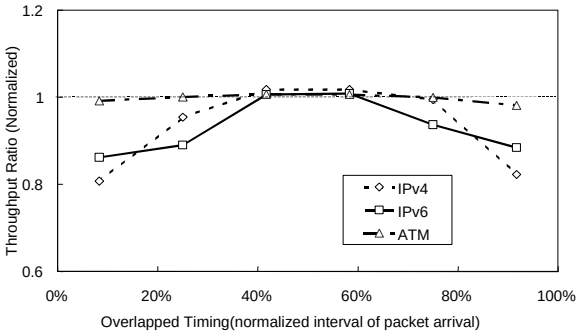


図 5　2 リンク多重時の入力タイミング処理性能特性（同一プロトコル）

図５は，各リンクから入力してくるパケットの到着タイミング（位相）の変化が，処理性能に与える影響を，(a)の測定結果を基準にして，示した結果である．この結果では，先の仮説通りスループットが約 80%～102%の範囲で変動している．100%を超える変動は，本提案方式により，スケジューリングオーバーヘッドなく，多重実行されていることを示している．ATM セルヘッダ処理の場合にはプログラムの並列処理性の静的な変動が少

(注 1) 自己タイミング型パイプラインの最大スループット値については，0.18μm の LSI 製造プロセスを前提として，240M [GPE パケット/秒]として換算．

(注 2) IP パケットの平均ペイロード長については，2000bit[8]として換算(ただし，ppp オーバヘッドを 8Bytes とする)

ないため変動幅が小さく，IPv4 や IPv6 の場合には逆に変動幅が大きくなる．

この測定結果は，統計多重効果の観点で最悪条件である 2 リンク多重処理時でも，本 GPE は自律的な負荷平滑化により，オーバフローせずに約 80%の処理率を維持できることを示している．このような過負荷状況では，パケット入力から出力までの遅延時間が若干増加するが，データ駆動プロセッサ内にパケット流量の変動を吸収する可変長 FIFO バッファを分散配置する方法[9] や，負荷分散用の nPE の追加する方法等により，nPE 内の負荷の平滑化を進め，処理率ならびに遅延時間を改善できる．

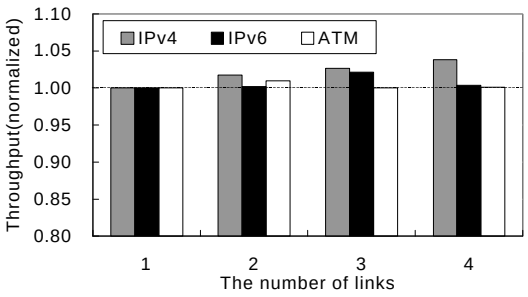


図 6 複数リンク多重時の処理性能特性（同一プロトコル）

次に，入力リンク数を増加させ，各プロトコル毎にスループットがどのように変化するか測定した．この場合の測定は，複数リンクを多重化した場合の統計多重効果を観測するために，位相差が均等な場合を対象とした．その結果を図６に示す．

この図では，入力リンク数を２から４まで変化させ，スループット向上率を各プロトコル毎に比較している．IPv4/v6，ATM いずれの場合も高い過負荷耐性を示している．2 リンクの場合と同様に複数リンクの場合でも統計多重効果があり，GPE の処理能力は入力リンク数の増加の影響を受けないことが分かる．

(c) マルチプロトコル多重処理

図７は，3 種類のプロトコルを，組み合わせを変えて同時に多重実行した場合の最大スループットを測定した結果である．縦軸は，(a)の測定結果を基準に正規化した性能向上率である．ただし，多重化される各プロトコルの影響（処理負荷）が均等になるように，各プロトコルパケットを混在させて測定した結果である．

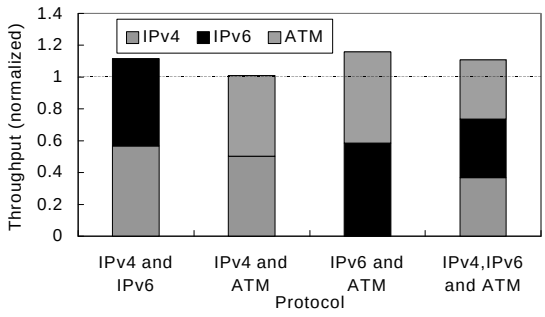


図 7 パケット処理性能特性（異種プロトコル）

この結果では，いずれの場合も各プロトコルのスループット向上率の合計が全て 1.0 以上となっている．むしろ，単独プロトコルの処理時よりも異種プロトコルの多重処理時の方が総合的なスループットが向上する傾向が見られ，最大で約 15%性能が向上している．これは，以下のような理由によるものと考えられる．

同一プロトコルの多重処理の場合には，同一の負荷変動パターンを持つ負荷が重畳するため，全体の負荷変動を平滑化させる作用が小さい．一方，異種プロトコルの多重実行の場合には，各プロトコル毎に負荷の変動パターンは異なり，重畳される負荷変動パターンの相関が低いため，これらが相互に打ち消し合い，全体の負荷変動を平滑化する方向に作用する．結果として，GPE の自己緩衝能力によって許容できる過負荷領域が広がるため，より性能が向上する．

3.3 入力パケット到着率変動時の評価

前節の単一リンクの評価において，プロセッサクラスタ内の nPE の負荷変動状況を観測すると，例えば，IPv6 パケット処理を実行中の TAG-nPE 内負荷の時間変動は図８のようになっている．

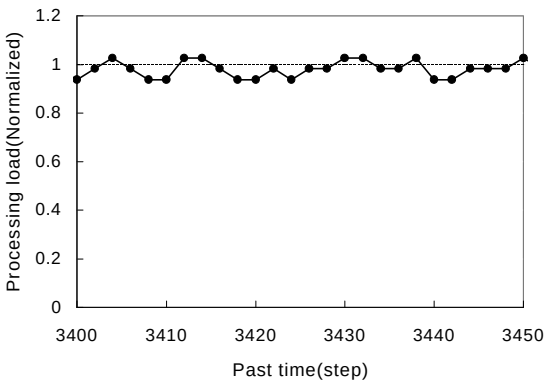


図 8　GPE の緩衝特性

横軸は時間経過, 縦軸は nPE の処理負荷である．ただし，処理負荷は，単位時間にパイプラインを通過して処理されるパケット数を，nPE の平均最大処理性能（図中点線）で正規化した値である．このグラフから明らかなように，GPE では，最大処理能力を一時的に超えても，自己タイミングパイプラインに内在する緩衝能力により過負荷を吸収して，継続的に処理可能なことを示している．

この結果に基づき，GPE にどの程度の自己緩衝能力があるのか，入力パケット間隔をランダムに変動させて IPv4 の場合を評価した．

図９は，平均入力パケット間隔に対して，入力ストリームが 60%～30%，および 10%変動した場合のスループットの変化を，(a)の測定結果で正規化したものである．ただし，入力間隔のゆらぎ幅をランダムに変動させて試行し，そのうち 95%以上疎通した場合について，プロットしている．この図から，入力パケットの到着率が一定の場合に比べて，到着間隔の 60%程度までは揺らいでも耐性があることが判る．

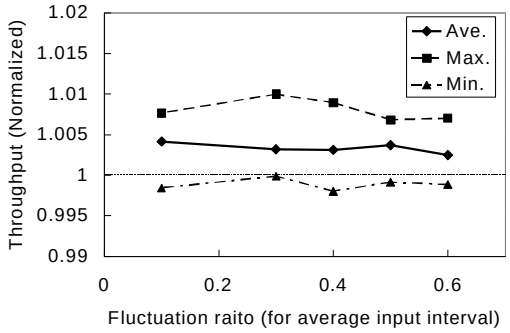


図 9　GPE の入力変動耐性

以上の結果から，入力パケットの到着率がある範囲内で変動しても GPE の自己緩衝能力によって吸収可能なため，全体のスループットに影響を与えず高い過負荷耐性を示すことが確認された．

4. 結論

本稿では，柔軟な機能変更を可能にする新しいプログラマブル・ハードウェアによる分散型マルチプロトコルパケットルータの構想，ならびにその動的データ駆動型方式による高速実現法を提案した．さらに，データ駆動型マルチプロセッサチップによる汎用パケット処理エンジン GPE の性能特性を実験的に評価した結果，複数リンクからの異種プロトコルトラフィックをオーバヘッドなしに多重処理可能なことが確認された．これらの結果から，命令セットやアーキテクチャをさらに最適化すれば，過負荷耐性に優れた柔軟な超高速 GPE チップを実現できる見通しが得られた．

近年，ネットワーク処理を効率的に行うことを目的とした，ネットワークプロセッサが注目を浴び各所で開発が行なわれている[10]．本提案方式のように，異種プロトコルソフトウェアの多重処理を特別なスケジューリング機構なしに自己緩衝能力によって柔軟に実現する手法は他では見られない特徴であり，将来の高機能ルータの構成要素の核になりうると考えている．

謝辞 本研究の機会を与えていただいた日本テレコム（株）情報通信研究所 弓削副所長に深謝します．また，本研究に際し有益なご議論を頂いたシャープ（株）およびパシフィックソフトウェア開発（株）の関係各位に感謝します．

参考文献

[1] H. Terada, “Impact of Photonic Technology on the Future Communication,” IEICE Trans. on Comm., vol. E77-B, no. 2, pp. 96-99, Feb. 1994.

[2] Arvind, and R.Iannucci, “A Critique of Multi-Processing von Neuman Style, in Proc. 10th Int. Symp. Computer Architecture, pp.426-436, 1983

[3] H.Terada, S.Miyata, and M.Iwata, “DDMP’s: Self-Timed Super-Pipelined Data-Driven Multimedia Processors,” Proc. IEEE, vol.87, no.2, pp.282-296, Feb. 1999.

[4] 岩田, 宮田, 寺田. 自己タイミングスーパパイプライン型データ駆動プロセッサ, 信学論(D-Ⅰ), vol.J81-D-Ⅰ, no.2 pp.62-69, Feb. 1998.

[5] M. Iwata, H. Terada, et al., “Flow-thru Processing Concept and its Application to Softcomputing,” Computers & Electrical Eng., pp.3-15, Mar. 1998.

[6] Intel IXP1200. http://developer.intel.com/design/network/products/npfamily/ixp1200.htm

[7] IBM PowerNP. http://www-3.ibm.com/chips/techlib/techlib.nsf/products/IBM_PowerNP_NP4GS3

[8] C.Partridge, et al., “A 50-Gb/s IP Router,” IEEE/ACM Trans. on Networking, vol.6, no.3, pp.237-248, June 1998.

[9] 上方, 岩田, 滝根, 寺田, “分散キューバッファを持つデータ駆動型プロセッサ Qv-x の性能評価,” 電学論(C), vol.116-C, no.11, pp.1295-1300, 1996.

[10] L.Gwennap, B.Wheeler, “A Guide to Network Processors,” MicroDesign Resources, 2000.