Vol.2011-ARC-193 No.17
2011/1/21

優先度を考慮したオンチップルータ VIX の設計及び実装

向後 卓磨[†1]　山崎 信行[†1]

メニーコア CMP の性能は，各コアを接続する NoC の性能に大きく影響される．特に NoC におけるパケットの転送遅延はアプリケーションの実行時間に大きく影響する．今後，メニーコア CMP 上で異なるアプリケーションを同時に実行する要求が高まることが予想されるため，システム全体の性能向上を目的に各アプリケーションのパケットの転送遅延の保証を行ったり，差をつけたりする優先度制御が重要となってくる．しかし NoC を構成するオンチップルータに単純に優先度比較論理を追加すると，ハードウェア量が劇的に増大する問題と高優先度を与えられたパケットが低優先度のものにブロックされることが頻発する問題が生じる．本論文では，これらの問題を解決する優先度制御を考慮したオンチップルータのアーキテクチャを提案する．実システムに近い条件でネットワークの評価を行ったところ，提案方式により平均転送遅延，ジッタ及び最悪転送遅延の削減が実現された．

Design and Implementation of a VIX Priority-Aware On-Chip Router

Takuma Kogo[†1] and Nobuyuki Yamasaki[†1]

Performance of many-core CMPs is significantly influenced by performance of NoC which interconnects each cores for on-chip communication. Specially, latency of packets in NoC remarkablly effect performance of applications executing on a many-core CMP. Because it is projected that executing different applications concurrently on a many-core CMP becomes highly required in the future, it will be important that a priority control which guarantees or differenciates latency of packets. If priority is naively introduced to on-chip router composing NoC, it is arouses two problems: drastic increase in router area and increase in latency of higher priority packets which is caused by lower priority packets frequently blocking higher priority packets. In this paper, in order to solve these problems we propose an on-chip router architecture for priority-aware resource allocation. We evaluated network performance under a condition near a real system. As a result, average lantency, jitter and worst-case latency are reduced by our proposal.

1. はじめに

半導体技術により数十億ものトランジスタが集積可能となることから，1 つのチップに複数のコアを搭載する Chip-Multiprocessors（CMPs）がプロセッサ性能を向上させるアーキテクチャとして有望である．既に多数のコアを搭載するメニーコア CMP[10)–12),20)] が実現されており，今後も CMP のコア数は増加し続けると多くの研究者および開発者が予想している．メニーコア CMP ではコア同士の通信のためにスケーラビリティ，転送遅延，バンド幅，ハードウェア量，消費電力などを考慮して，Networks-on-Chip（NoC）[3)] が用いられる．メニーコア CMP における NoC の性能はシステム全体に大きな影響を与える．特に NoC のパケットの転送遅延はアプリケーションの実行時間に大きく影響する．パケットの転送遅延は以下の式で与えられる．

$$Delay = H \times D_{router} + D_{contention} + D_{serialization}$$

H は平均ホップ数，D_{router} はルータ遅延，$D_{contention}$ は衝突遅延，$D_{serialize}$ はシリアル化遅延である．コア数の増加に伴い第 1 項の転送遅延の影響が大きくなるため，D_{router}（ルータのパイプラインステージ数）を削減するルータアーキテクチャが数々提案されてきた[14),16),17),19)]．また H を削減する効率の良いトポロジ[1),5),13)] も提案されており．これらの先行研究により第 1 項の遅延の影響は最小限に抑えられつつある．第 3 項はパケット長（フリット数）で決まり固定である．さらなる性能向上のためには第 2 項の $D_{contention}$ を小さくする必要がある．

システムレベルで考えた場合，全てのパケットの $D_{contetion}$ を削減する必要はなく転送遅延がクリティカルなパケットを優先することが十分有効であることがわかっている[2),6),7),9),15)]．クリティカルなパケットの転送遅延を制御するには，パケットに優先度を付加して，ネットワークを構成する各ルータでクロスバや仮想チャネルなどの資源を優先度に基づいて割り当てれば良い．しかし従来のオンチップルータに単純に優先度比較論理を追加した場合，ルータのハードウェア量が大幅に増大すると同時に，高優先度パケットが低優先度パケットにブロックされて高優先度パケットの転送遅延が増大してしまう優先度逆転問題が発生しやすいという問題がある．そこで本論文は，優先度制御を考慮したオンチップルータ VIX を提案

†1 慶應義塾大学大学院理工学研究科開放環境科学専攻
Department of Computer Science, Graduate School of Science and Technology, Keio University

Vol.2011-ARC-193 No.17
2011/1/21

し，これらの問題を解決する．我々はこれまでも優先度制御を考慮したオンチップルータの研究を行ってきたが[21]，本論文は以下の点で異なる．

- 実システムを想定して，メッセージクラスをサポートするための拡張を行う．
- 実システムに近いトラフィックパターンと優先度割当方式を用いてネットワーク性能を評価する．

本論文の構成は次の通りである．第 2 章で優先度付き NoC のサーベイと従来のルータアーキテクチャに関する考察を述べ，第 3 章で提案する VIX ルータの設計，実装及び拡張について述べる．第 4 章では VIX ルータの評価を行う．第 5 章で本論文をまとめる．

2. 背　　景

本章では，はじめに優先度付き NoC のサーベイについて述べた後，従来のオンチップルータに単純に優先度を導入した場合の問題について述べる．

2.1 優先度付き NoC のサーベイ

NoC に優先度を導入する目的は以下のように分類できる．

- ネットワークの特性に基づいた資源割当制御[2]
- アプリケーションの特性に基づいた資源割当制御[6),7]
- 転送遅延を保証する Quality-of-Services (QoSs) 制御[9),15]

1 つ目は，純粋にネットワークレベルでの転送遅延削減を目的としている．パケット長の短いパケットに高い優先度を与え，パケット長の長いパケットには低い優先度を与える方式が提案されている[2]．

2 つ目は，ネットワークレベルの転送遅延削減が目的ではなく，アプリケーションの実行時間削減が目的である．通信遅延がクリティカルなアプリケーションのパケットに高い優先度を与える方式[6] や，アプリケーションの実行時間に大きな影響を与えるパケットに高い優先度を与える方式[7] が提案されている．

3 つ目は，転送遅延削減でもアプリケーションの実行時間削減でもなく，通信のリアルタイム性が目的である．これはチップ内通信にリアルタイム性が必要なシステムに限らない．通信パターンに対して特定のパケットの転送遅延が何十倍以上のオーダーで増大するようではシステムを構成する上で支障となりかねない．この通信時間予測性の低下はコア数の増加に伴うため，今後は QoS 制御の重要性が増していくと考えられる[15),18]．

2.2 従来のオンチップルータ

優先度の導入によりオンチップルータのハードウェア量は劇的に増大する．仮想チャネルアロケータとクロスバアロケータを構成する各アービタに優先度比較論理が追加されるが，図 1 に示すように仮想チャネルアロケータは多数の大きなアービタで構成されている[19] ため，優先度比較論理の追加によってハードウェア量が大幅に増大してしまう．

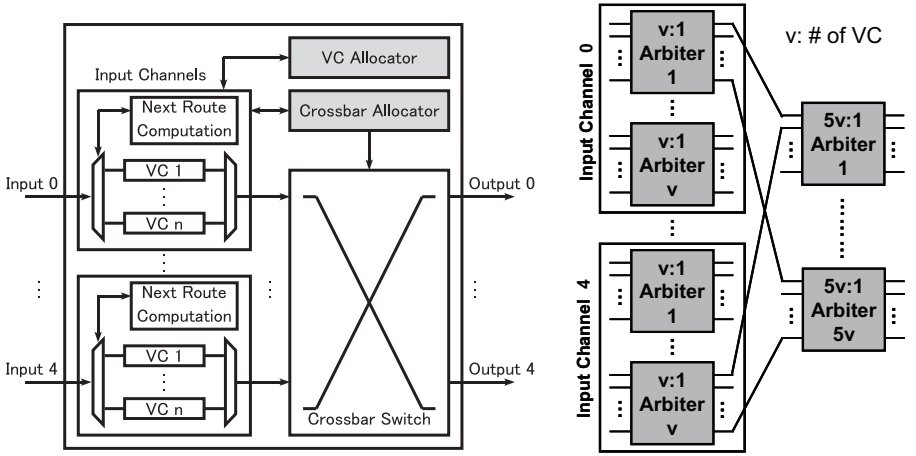


図 1　従来のオンチップルータのアーキテクチャと仮想チャネルアロケータ

また，従来のオンチップルータでは仮想チャネル割当とクロスバ割当が独立して実行されるため，クロスバが割り当てられることのない低優先度パケットが仮想チャネルを獲得してしまう．低優先度パケットが獲得した仮想チャネルは高優先度パケットがルータから抜けるまで使用されることがないので，ルータの実質的な仮想チャネル数が低下する．仮想チャネルが全て占有されると，高優先度パケットですら仮想チャネルが空くまで待つ必要がある．これが優先度逆転問題であり，高優先度パケットの転送遅延を増大させる原因である．

3. VIX ルータアーキテクチャ

本章では前章で挙げた問題を解決するオンチップルータアーキテクチャVIX を提案する．VIX の設計及び実装について述べた後，拡張について説明する．

3.1 VIX ルータの設計及び実装

仮想チャネルアロケータの多大なハードウェア量と低優先度パケットによる仮想チャネルの浪費による高優先度パケットの転送遅延増大を同時に解決するために，仮想チャネルアロケータの構成を変更することが望ましい．優先度付き NoC では優先度を考慮しない NoC

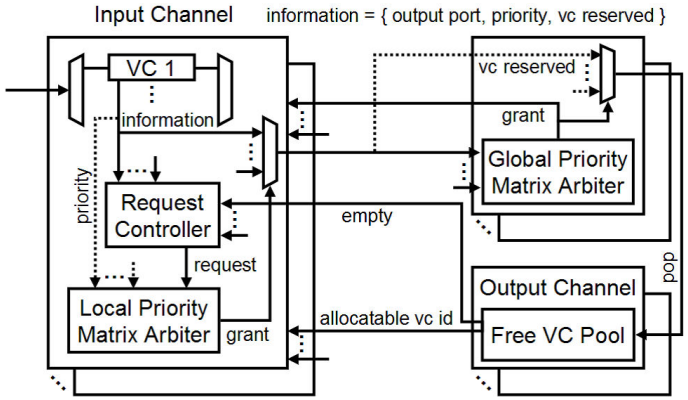


図 2 VIX ルータのアーキテクチャ

と比較すると，**クロスバ割当の機会は高優先度パケットに限られるため，低優先度パケットに仮想チャネルを割り当てたとしてもスループットは向上せず，反対に割り当てることによって仮想チャネルを浪費する可能性がある．**この洞察に基づき我々は仮想チャネル割当方式を変更することによってハードウェア量と高優先度パケットの転送遅延の問題を同時に解決する．

本論文では仮想チャネル割当とクロスバ割当のパイプラインステージを統合する VIX (Virtual channel allocation integrated with crossbar allocation) ルータを提案する．VIX ルータでは，同じパイプラインステージでクロスバ割当，仮想チャネル割当の順で逐次的に行う．仮想チャネル獲得の有無に関わらず全てのパケットがクロスバ割当のリクエストを出し，グラントを得たパケットは次のパイプラインステージで転送を開始する．このときパケットが仮想チャネルを獲得していなければクロスバ割当のグラントをトリガに仮想チャネルを獲得する．VIX ルータは各出力チャネルに対して 1 サイクルで最大 1 本の仮想チャネルしか割り当てないため，仮想チャネル割当論理を大幅に削減することができる．また VIX ルータではクロスバが割り当てられる高優先度のパケットのみが仮想チャネルを獲得するので，仮想チャネル利用効率が向上し，優先度逆転問題の発生確率を低下させることができる．

図 2 に VIX ルータのアーキテクチャを示す．VIX ルータでは仮想チャネルアロケータを除外し，代わりに同等の機能となる**仮想チャネルプール**を各出力チャネルに 1 つずつ追加している．仮想チャネルプールは割当可能な仮想チャネルを管理し，そのうち 1 本の仮想チャネルの識別子を出力する．また各入力チャネルに**リクエストコントローラ**を 1 ずつ追加している．リクエストコントローラはクロスバ割当リクエストのうち不要なリクエストをフィルタリングする．リクエストコントローラは仮想チャネルを獲得しているリクエストはそのまま通過させ，仮想チャネルを獲得していないリクエストは割当可能な仮想チャネルが存在すれば通過させ，存在しなければ無効にする．仮想チャネルプールの更新は当該リクエストが仮想チャネルを獲得したときに行われる．VIX ルータは大幅なハードウェア量を削減しつつ，クリティカルパスも削減する．実装の詳細と面積及びクリティカルパスの評価は文献[21] に記している．

3.2 VIX ルータの拡張

NoC を構成する上でメッセージクラス[⋆1]のサポートは重要である．メッセージクラスサポート（仮想チャネル予約）の主な目的を以下に挙げる．

- 適応型ルーティングにおけるデッドロック回避
- プロトコルデッドロック[4] 回避
- 最高優先度パケットなどの特殊なパケットのプリエンプションを保証

オリジナルの VIX ルータはメッセージクラスをサポートしていない．そこでメッセージクラスサポートのための拡張について本節で述べる．オリジナルの VIX ルータは各出力チャネルにある仮想チャネルプールから割当可能な仮想チャネル識別子を 1 つだけ出力しているが，メッセージクラス 1 つ毎に割当可能な仮想チャネルを管理し，それぞれから識別子を 1 つ出力するように変更する．またリクエストのタプル（information）にメッセージクラス（mc）を追加し，リクエストコントローラのリクエスト無効化の論理は当該出力チャネルかつ当該メッセージクラスの割当可能な仮想チャネルの存在をチェックするように変更する．仮想チャネルの獲得は当該出力チャネルかつ当該メッセージクラスの仮想チャネル識別子を受ける取るように変更する．仮想チャネルプールの更新はどのメッセージクラスに対してポップするか決められるように仮想チャネル獲得信号（vc resereved）と同様にリクエストのタプルの 1 つして要求するメッセージクラス（mc）を伝搬させ，仮想チャネルプールが受け取ることができるように変更する．

⋆1 メッセージクラスとはある仮想チャネルを専用に予約するパケットの集合のことである．言い換えればメッセージクラスのサポートはあるパケットの集合に対して仮想チャネルを専用に予約することである．

Vol.2011-ARC-193 No.17
2011/1/21

4. 評　　価

Verilog HDL を用いて提案方式及び従来方式のオンチップルータを実装し，Cadence NC-Verilog を用いてネットワークの評価を行った．評価環境について述べた後，評価結果を示す．

4.1 評 価 環 境

表 1 に評価のパラメータを示す．評価対象のルータは提案方式の VIX と従来方式の SPC である．SPC は先読みルーティング[8] と投機[19] を組み合わせることでパイプラインを 2 段とした．VIX は先読みルーティングを用いることでパイプラインを 2 段とし，SPC と同じにした．トラフィックに関するパラメータは実システムを想定して以下の通りとした．

- **パケット長**：1 フリットのショートパケット (制御パケット) と 5 フリットのロングパケット（データパケット）を同確率で生成する．
- **リクエスト・アクノレッジ制約**：リクエストパケットがディスティネーションノードに到着するとソースノード宛てにアクノレッジパケットが生成される．またアクノレッジパケットのパケット長はリクエストパケットがショートならばロング，ロングならばショートと対応付けている．プロトコルデットロックを回避するために 4 本の仮想チャネルの内 1 本をリクエスト用に，1 本をアクノレッジ用に予約してある．
- **優先度**：ショートパケットに高い優先度を与える方式と[2] とバッチ方式[⋆1][6),7)] を組み合わせる．優先度の上位 3 ビットをバッチ方式に，下位 1 ビットをパケット長方式に使用するものとした．
- **トラフィック**：空間的局所性を考慮した Local トラフィックを定義する．Local トラフィックはディスティネーションが 75%の確率で隣接ノードの中からランダムに，25%の確率で全ノードからランダムに決まる．

4.2 シミュレーション結果

図 3 は Uniform トラフィックにおける各方式の平均転送遅延，ジッタと最悪転送遅延を示している．ジッタは転送遅延の標準偏差である[4]．各グラフでは，ショートパケットとロングパケットそれぞれの測定値を示しており，平均転送遅延に関しては全パケットの測定値も示している．図 3(a) より各パケットの転送遅延は飽和状態前の負荷 0.16〜0.24 の時に削減されいることがわかる．これは VIX のクロスバ割当に連動した仮想チャネル割当により仮想チャネルの利用効率が向上し，高優先度パケットが低優先度パケットにブロックされにくくなったためである．このことは図 3(b) と図 3(c) により，安定した転送が実現されていることからも確認できる．

図 4 は図 3 と同様に Local トラフィックにおける各方式の測定値を示したグラフである．Uniform トラフィック同様に低い負荷から高い負荷まで転送遅延を削減しつつ，公平で安定した転送が実現できていることがわかる．また VIX はスループットを向上させていることから従来方式と比較してバーストトラフィックに対しても転送遅延を低く抑えることが可能だと考えられる．

表 1　評価パラメータ

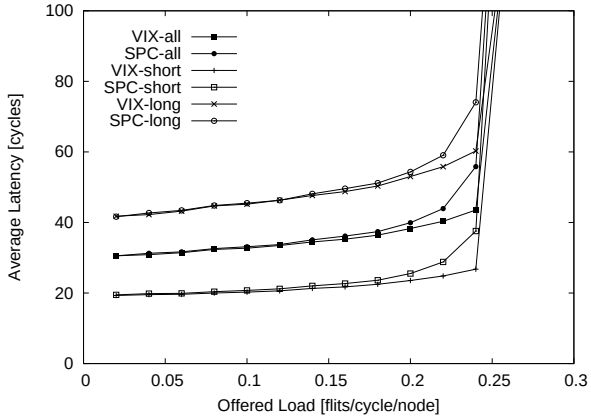
Topology	8-ary 2-mesh
Traffic	Uniform, Local
Routing	Demesion-order
Switching	Wormhole
Allocation	Least recent serve + priority
Link delay	1 cycle
Router pipeline	2-stages
# of VCs	4 /port
VC depth	4 flits/VC
Packet size	1 or 5 flits (50:50 chance)
Flit width	128 bits
Priority	4 bits
# of Message Classes	2 (VC2 reserved for MC0, VC3 reserved for MC1)
Simulation	Warmup: 2,000 cycles, Sample: 40,000 cycles

5. まとめと今後の課題

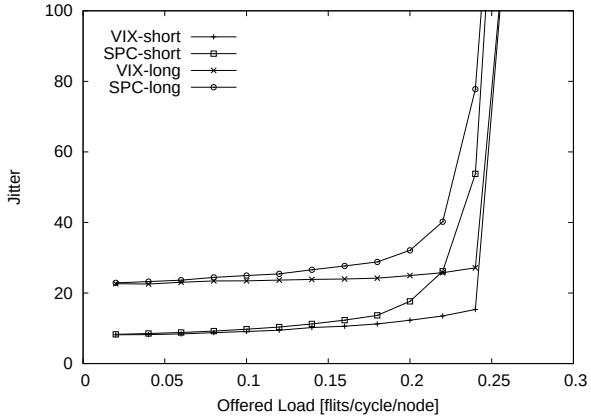
本論文では，優先度制御を考慮したオンチップルータのアーキテクチャVIX とメッセージクラスサポートのための拡張について提案し，実システムに近い条件で評価を行った．仮想チャネル利用効率の向上により，ジッタと最悪転送遅延を削減し，公平で安定した転送を実現した結果，平均転送遅延とスループットが改善した．

本論文では 1 つの優先度割当方式について評価を行ったが，様々な優先度割当方式について評価を取る必要があると考えている．またアプリケーションのトレースを用いた評価により実アプリケーションにどの程度の改善効果があるのか検証したい．

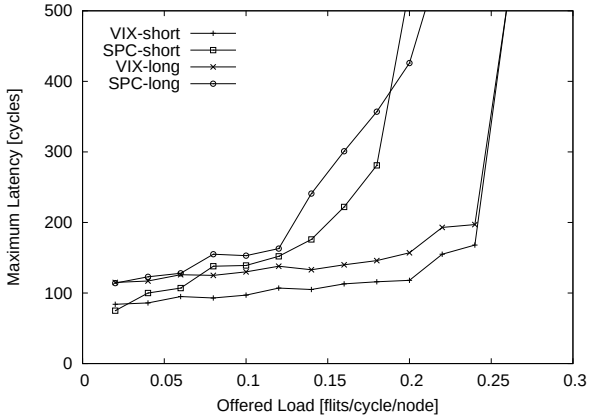
⋆1 バッチ方式は優先度をエイジングする方法であり一定サイクル毎にパケットの優先度を 1 上げる方法である．バッチ方式は低優先度パケットのスタベーションの発生を防ぐだけでなく，パケットの公平性も向上させる．詳細は文献を参照されたい．評価では，バッチインターバルを 64 サイクル，バッチ数を 16 としている．



(a) 平均転送遅延



(b) ジッタ



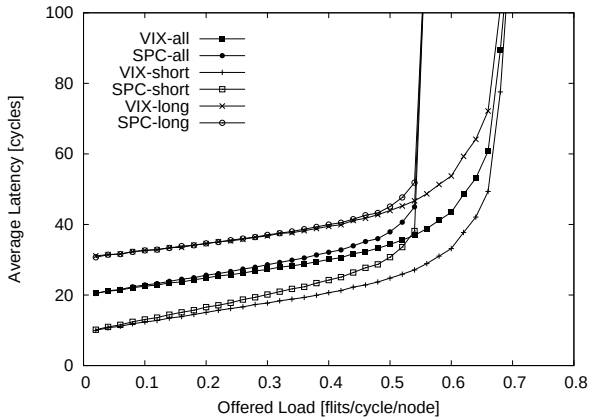
(c) 最悪転送遅延

図 3 Uniform トラフィックにおける各評価

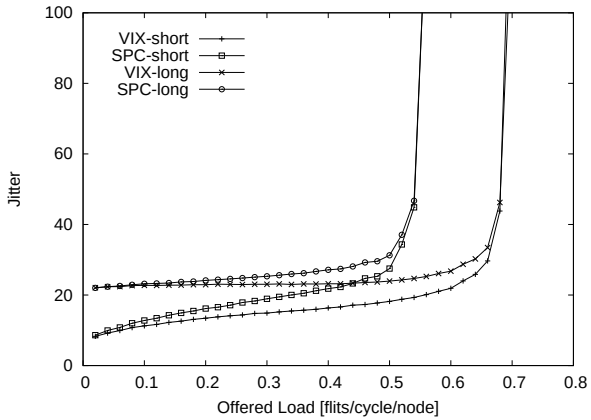
謝辞 本研究は科学技術振興機構 CREST の支援によるものであることを記し，謝意を表す.

参 考 文 献

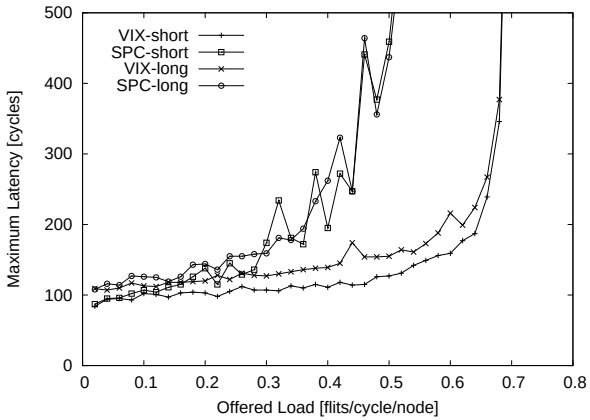
1) Balfour, J. and Dally, W.: Design and Tradeoffs for Tiled CMP On-Chip Networks, *Proceedings of the International Conference on Supercomputing (ICS'06)*, pp.187–198 (2006).
2) Bolotin, E., Guz, Z., Cidon, I., Ginosar, R. and Kolodny, A.: The Power of Priority: NoC Based Distributed Cache Coherency, *Proceedings of the 1st ACM/IEEE International Symposium on Networks-on-Chip (NOCS'07)*, pp.117–126 (2007).
3) Dally, W.J. and Towles, B.: Route Packets, Not Wires: On-Chip Interconnection Networks, *Proceedings of the Design Automation Conference (DAC'01)*, pp.684–689 (2001).
4) Dally, W.J. and Towles, B.: *Principles and Practices of Interconnection Networks*, Morgan Kaufmann (2004).
5) Das, R., Eachempati, S., Mishra, A. K., Narayanan, V. and Das, C. R.: Design and Evaluation of a Hierarchical On-Chip Interconnect for Next-Generation CMPs, *Proceedings of the Symposium on High-Performance Computer Architecture (HPCA'09)*, pp.175–186 (2009).
6) Das, R., Mutlu, O., Moscibroda, T. and Das, C.R.: Application-Aware Prioritization Mechanisms for On-Chip Networks, *Proceedings of the International Symposium on Microarchitecture (MICRO'09)*, pp.280–290 (2009).
7) Das, R., Mutlu, O., Moscibroda, T. and Das, C.R.: Aergia: Exploiting Packet Latency Slack in On-Chip Networks, *Proceedings of the International Symposium on Computer Architecture (ISCA'10)*, pp.106–116 (2010).
8) Galles, M.: Scalable pipelined interconnect for distributed endpoint routing: The SGI SPIDER chip, *Proceedings of the International Symposium on High-Performance Interconnects (HOTI'96)*, pp.141–146 (1996).
9) Grot, B., Keckler, S.W. and Mutlu, O.: Preemptive Virtual Clock: A Flexible, Efficient, and Cost-effective QoS Scheme for Networks-on-Chip, *Proceedings of the International Symposium on Microarchitecture (MICRO'09)*, pp.89–100 (2009).
10) Hoskote, Y., Vangal, S., Singh, A., , Borkar, N. and Borkar, S.: A 5-GHz Mesh Interconnect for a Teraflops Processor, *IEEE Micro*, pp.51–61 (2007).
11) Howard, J., Dighe, S., Hoskote, Y., Vangal, S., Finan, D., Ruhl, G., Jenkins, D., Wilson, H., Borkar, N., Schrom, G., Pailet, F., Jain, S., Jacob, T., Yada, S., Marella, S., Salihundam, P., Erranguntla, V., Konow, M., Riepen, M., Droege, G., Lindemann, J., Gries, M., Apel, T., Henriss, K., amd S.Steibl, T. L.-L., Borkar, S., De, V., Wijngaart, R. V.D. and Mattson, T.: A 48-Core IA-32 Message-Passing Processor



(a) 平均転送遅延



(b) ジッタ



(c) 最悪転送遅延

図 4 Local トラフィックにおける各評価

with DVFS in 45nm CMOS, pp.108–109 (2010).

12) Kelm, J.H., Johnson, D.R., Johnson, M.R., Crago, N.C., Tuohy, W., Mahesri, A., Lumetta, S.S., Frank, M.I. and Patel, S.J.: Rigel: An Architecture and Scalable Programming Interface for a 1000-core Accelerator, *Proceedings of the International Symposium on Computer Architecture (ISCA'09)*, pp.140–151 (2009).

13) Kim, J., Balfour, J. and Dally, W.J.: Flattend Butterfly Topology for On-Chip Networks, *Proceedings of the International Symposium on Microarchitecture (MICRO'07)*, pp.172–182 (2007).

14) Kumar, A., Peh, L.-S., Kundu, P. and Jha, N.K.: Express Virtual Channels: Towards the Ideal Interconnection Fabric, *Proceedings of the International Symposium on Computer Architecture (ISCA'07)*, pp.150–261 (2007).

15) Lee, J.W., Ng, M.C. and Asanovic, K.: Globally-Synchronized Frames for Guaranteed Quality-of-Service in On-Chip Networks, *Proceedings of the International Symposium on Computer Architecture (ISCA'08)*, pp.89–100 (2008).

16) Matsutani, H., Koibuchi, M., Amano, H. and Yoshinaga, T.: Prediction Router: Yet Another Low Latency On-Chip Router Architecture, *Proceedings of the Symposium on High-Performance Computer Architecture (HPCA'09)*, pp.367–378 (2009).

17) Mullins, R., West, A. and Moore, S.: Low-latency virtual-channel routers for on-chip networks, *Proceedings of the International Symposium on Computer Architecture (ISCA'04)*, pp.188–197 (2004).

18) Owens, J.D., Dally, W.J., Ho, R., Jayasimha, D.J., Keckler, S.W. and Peh, L.-S.: Rereach challenges for on-chip interconnection networks, *IEEE Micro*, pp.96–108 (2007).

19) Peh, L.-S. and Dally, W. J.: A deley model and speculative architecture for pipelined routers, *Proceedings of the Symposium on High-Performance Computer Architecture (HPCA'01)*, pp.255–266 (2001).

20) Wentzlaff, D., Griffin, P., Hoffmann, H., Bao, L., Edwards, B., Ramey, C., Mattina, M., Miao, C.-C., III, J. F.B. and Agarwal, A.: On-Chip Interconnection Architecture of the Tile Processor, *IEEE Micro*, pp.15–31 (2007).

21) 向後卓磨，山崎信行：優先度付きオンチップネットワーク向けのルータアーキテクチャ，電子情報通信学会技術研究報告：デザインガイア，pp.13–18 (2010).