

# ベクトルプロセッサ用キャッシュメモリにおける MSHR の性能評価

佐藤義永<sup>1</sup>, 撫佐昭裕<sup>1, 2</sup>, 江川隆輔<sup>1</sup>, 滝沢寛之<sup>1</sup>, 岡部公起<sup>1</sup>, 小林広明<sup>1</sup>

1 東北大学, 2 日本電気株式会社

{yoshiei, musa}@sc.isc.tohoku.ac.jp, {egawa, tacky, okabe, koba}@isc.tohoku.ac.jp

**概要:** ベクトルコンピュータにおける演算性能(Gflop/s)あたりのメモリバンド幅(GB/s)が低下しており, その影響で実行効率が低下する恐れがある. そこで我々はベクトルプロセッサ向けのキャッシュメモリ機構(ベクトルキャッシュ)を提案している. 本報告では, 姫野ベンチマークによる性能評価を通じ, Miss Status Handling Register (MSHR)をベクトルキャッシュ内部に備えることで, キャッシュミスが起きた場合の冗長なベクトルロード時間を削減し, 実行効率が向上することを明らかにする. MSHRの有効性を検証した結果, ベクトルキャッシュがない場合と比較して最大で94%の実行効率の向上を達成できることを示した.

## 1 はじめに

スーパーコンピュータによるシミュレーションは理論と実験に並ぶ第3の手法として先端科学分野で重要な役割を担っており, 高精度・大規模なシミュレーションが求められることから, より高性能なスーパーコンピュータが必要とされている. しかし近年, CPUに実装可能な入出力ピン数の物理的限界から, メモリバンド幅の向上が今まで以上に困難となっている. その結果, 演算性能(Gflop/s)あたりのメモリバンド幅(GB/s)が減少し, 実アプリケーションの実行効率が低下する恐れがある[1]. NEC製ベクトルスーパーコンピュータであるSXシリーズにおいても, SX-7で4B/flop, SX-9では2.5B/flopと演算性能あたりのメモリバンド幅が下降傾向にある. そこで我々は, 高い実効性能の実現を目的として, ベクトルプロセッサ用キャッシュメモリ機構(以下ベクトルキャッシュ)を提案している[2].

本報告では, ベクトルキャッシュの効果をより高めるため, キャッシュミス時のロード時間削減を目的としてMiss Status Handling Register (MSHR)を導入し, その性能評価を行う.

## 2 MSHR 付きベクトルキャッシュ機構

本研究では, 演算性能向上に比べて相対的に性能低下が起きているメモリバンド幅をカバーすると共にメモリアクセスレイテンシを削減するために, ベクトルプロセッサ向けキャッシュについて検討する. 本ベクトルキャッシュには, キャッシュミス時のアドレス情報管理のために, MSHR[3]を用意し, ノンブロッキングキャッシュを実現する. 図1にベクトルキャッシュとMSHRを加えたベクトルプロセッサを示す. MSHRはキャッシュミスを起こしたベクトルロードのメモリアドレスを記録しておくレジスタである. 後続するベ

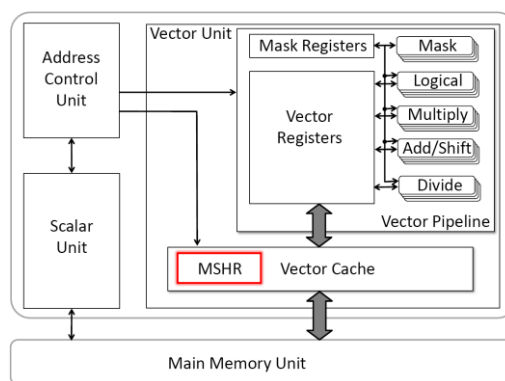


図1. ベクトルキャッシュとMSHRを加えたベクトルプロセッサ

クトルロードはメインメモリへアクセスを行う前にMSHRを確認し, 共通するメモリアドレスが存在する場合にはメモリへのアクセスは行われず, データの到着を待つ. 結果として, キャッシュミス時の冗長するメモリアクセスを削減でき, メモリロード時間の増加を抑制する効果が期待できる.

## 3 性能評価

MSHRを加えたベクトルキャッシュをNEC SX-7に導入した場合の性能評価を行う. 本評価ではキャッシュのラインサイズ8バイトの2-wayセットアソシアティブ方式を採用し, キャッシュ容量は16KBから2MBまで変化させ評価する. また, キャッシュ-レジスタ間を4B/flop, メモリ-キャッシュ間を4B/flop, 2B/flop, 1B/flopと変化させて比較評価する. 評価にはポアソン方程式解法をヤコビの反復法で解く際の主要ループで構成された姫野ベンチマークのカーネルを用いる.

はじめに, 問題サイズを変化させた時の実行効率を評価する. 用いるキャッシュメモリの容量は2MBである. 実験の目的は, キャッシュによるメモリバンド幅低下の影響を軽減する効果と短ベク

トル時のメモリアクセスレイテンシ削減の検証である。結果を図2に示す。図2より、問題サイズが大きくなるにつれて実行効率が向上していくことが分かる。問題サイズが大きくなると演算時間が増加しメモリロード時間が隠ぺいされ、実行時間に対する演算時間の割合が増加するためである。しかし、B/flop が減少するとメモリロード時間が増加し、隠ぺいできなくなるために実行効率が大きく低下する。これに対し、本研究で提案するベクトルキャッシュは、このような状況においてもメモリロード時間を削減し、実行効率の向上に貢献する。また、SSやSにおいて4B/flop時でもキャッシュを利用することでより高い実行効率が得られる。SSやSにおいてはベクトル長が64, 128と短く、メモリレイテンシが表面化するが、キャッシュを導入することでレイテンシが削減され、実行効率が向上する。

次に、キャッシュ容量毎の実行効率とMSHRの効果の検証を行う。結果を図3に示す。横軸はキャッシュ容量、縦軸は実行効率である。MSHRを用いない場合には、キャッシュ容量が256KBより小さくなるにつれ、2B/flop以下では実行効率が低下していく。これはキャッシュ容量の不足によるキャッシュミス率の増加によるものである。そこでMSHRを導入した場合に注目すると、全ての場合において実行効率が向上する。特にキャッシュミスが増える256KB以下の場合において高い効果を示し、MSHRを用いることで最大20%の性能向上が得られることが分かる。図4にMSHRによるメモリロード時間の削減率を示す。1B/flopの場合では最大で20%のメモリロード時間を削減でき、その結果として2MBのベクトルキャッシュ

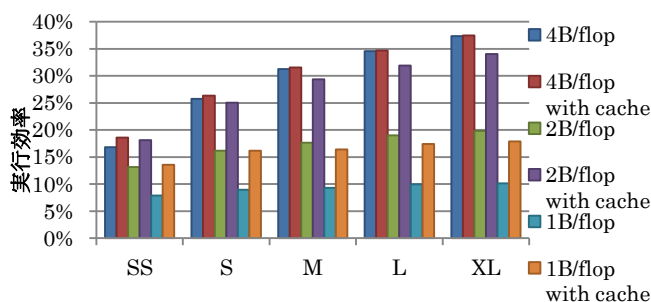


図2. 問題サイズ毎の実行効率

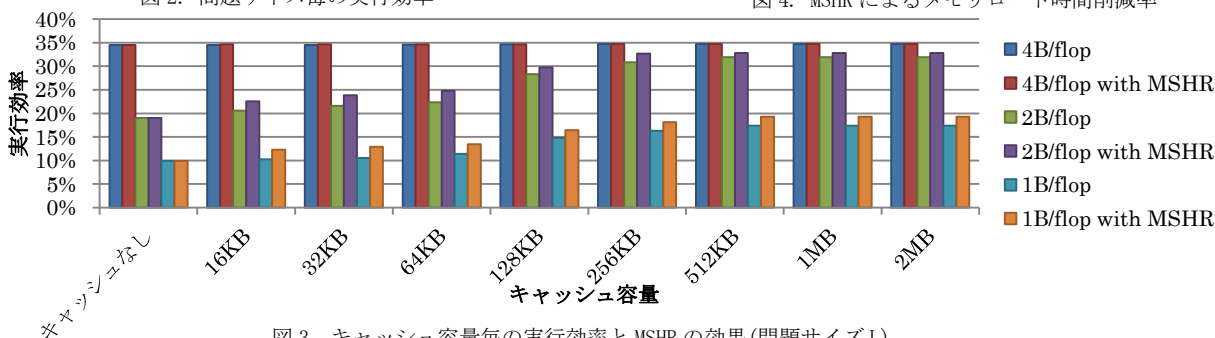


図3. キャッシュ容量毎の実行効率とMSRRの効果(問題サイズL)

にMSHRを導入することで、キャッシュメモリを用いない場合と比較して94%の実行効率の向上を得られることが示された。

#### 4 まとめ

本報告では、姫野ベンチマークを用いてベクトルキャッシュの性能を評価した。評価結果より、全てのサイズにおいてベクトルキャッシュがB/flopの減少に伴う実行効率低下の改善に有効であることを示した。さらに、キャッシュミスが頻出する大きな問題サイズの場合ではMSHRのメモリロード時間削減の効果が大きく、実行効率の向上が得られた。また、小さい問題サイズの場合ではキャッシュの高メモリバンド幅よりも低レイテンシ効果が影響し、実行効率の向上が得られる。

#### 参考文献

- [1] Kobayashi, H., et al, "Implication of Memory Performance in Vector-Parallel and Scalar Parallel HEC," High Performance Computing on Vector Systems 2006, Springer-Verlag, pages 21-50, 2006.
- [2] Kobayashi, H., et al, "The Potential of On-Chip Memory Systems for Future Vector Architectures," High Performance Computing on Vector Systems 2007, Springer-Verlag, pages 247-264, 2007.
- [3] D. Kroft., "Lockup-Free Instruction Fetch/Prefetch Cache Organization," International Symposium on Computer Architecture Proceedings of the 8th annual symposium on Computer Architecture, pages 81-87, 1981.

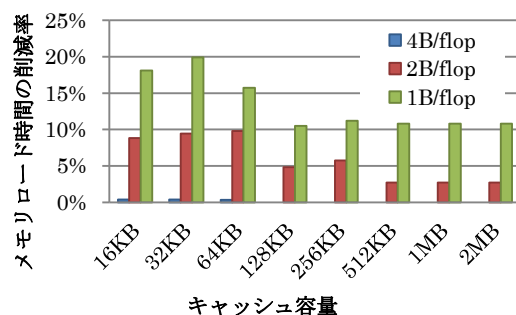


図4. MSRRによるメモリロード時間削減率