演算加速機構を持つオンチップメモリプロセッサの検討

高橋睦史[†1] 佐藤三久[†1] 高橋大介[†1]
朴泰祐[†1] 宇川彰[†1] 中村宏[†2,†1]
青木秀貴[†3] 澤本英雄[†4] 助川直伸[†3]

本稿では電力性能の向上に有効であるオンチップメモリプロセッサアーキテクチャSCIMA に，演算あたりのハードウェアおよび電力コストに有利な演算加速機構を導入することとし，その電力性能を評価する．演算加速機構としてベクトル型および SIMD 型の 2 種の方式を提案し，シミュレーションにより評価を行った結果，おおむねベクトル型が優位であったが，同コア数かつ同 FMA 数であればベンチマークの特性により各方式の特性に優劣が左右された．今後は電力モデルの詳細な検討や多様なアプリケーションを用いた評価を行う必要がある．

1. はじめに

オンチップメモリプロセッサはプロセッサコアと同一チップ上にメモリを持つプロセッサである．チップ外の主記憶にアクセスするには相対的に大きな電力が必要となるが，オンチップメモリプロセッサではオンチップメモリ・主記憶間のデータ転送をソフトウェアで効率的に制御することにより，電力性能の改善することができる．

他方，電力性能改善のアプローチとして，構成が単純な演算器を多数追加することで単位時間当たりの演算量を増やし，消費電力の増加を最小限に抑えつつ性能を向上させる方法が考えられる．そのためには効率的に演算器を利用するための機構が必要となる．

そこで本稿では多数の演算器からなる演算加速機構を搭載したオンチップメモリプロセッサの構成について検討を行い，シミュレータにより評価を行う．

2. 演算加速機構を持つオンチップメモリプロセッサ

本稿では演算加速機構を付加したオンチップメモリプロセッサについての検討と評価を行う．

このプロセッサは単層のオンチップメモリと演算加速機構を持つ．演算加速機構は既存のプロセッサコア（スカラコア）に付加される形で実装されるものし，このスカラコアと演算加速機構で 1 つのコアを構成する．このコアをプロセッサチップ内に複数配置し，L2 キャッシュおよびオンチップメモリを共有する．なお，メモリ階層は演算加速機構とオンチップメモリ，およびオンチップメモリと主記憶を直接接続する方法を想定する．検討するプロセッサの基本構成を**図 1** に示す．

以降，本プロセッサの特徴であるオンチップメモリと，検討する 2 種類の演算加速機構について詳細を述べる．

2.1 オンチップメモリ

オンチップメモリを用いたプロセッサアーキテクチャは数多く提案されている．我々は SCIMA と呼ぶアーキテクチャを提案している[1),2)]．我々はこの SCIMA を基本的なオンチップメモリアーキテクチャとして考える．

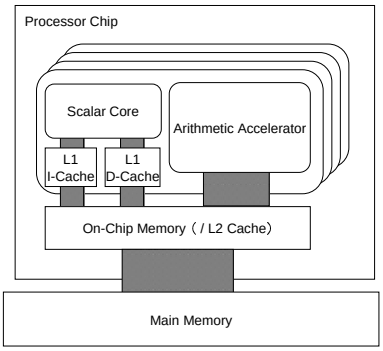

図 1 演算加速機構を持つオンチップメモリプロセッサの基本構成（4 コア時）

SCIMA は高性能コンピューティング向けのオンチップメモリプロセッサアーキテクチャであり，キャッシュと同じメモリ階層に実装したオンチップメモリをメモリウェイごとに，データ転送のソフトウェア制御可能なメモリか従来型のキャッシュとして使い分けることが可能である．

オンチップメモリを利用することで，キャッシュプリフェッチと同様，データ転送のタイミングを明示的に指定できる．また，必要なデータを明示的に指定できることから，従来のキャッシュで生じていた意図しないキャッシュミスや固定されたキャッシュラインサイズでのデータ転送で発生する不要なデータ転送を抑制する．また，事前にデータをオンチップメモリへ転送することで，各コア間でアクセスするアドレスがコンフリクトしなければ，演算加速機構とオンチップメモリ間のデータ転送のアクセスレイテンシが一定とあることを保証でき，プログラムの最適化が容易となる．

これらの利点により性能が向上し，加えて主記憶へのアクセス削減による電力性能の向上も見込まれる．

2.2 演算加速機構

本稿では SIMD 型とベクトル型の 2 種類の演算加速機構について検討を行う．

SIMD 型演算加速機構は 4 個の倍精度浮動小数点数積和演算器（Fused Multiply-Add, 以降 FMA）を持ち，128 個の SIMD レジスタを持つ．1SIMD レジスタあたり 64bit レジスタを 4 要素を持ち，演算時には 1 サイクルで 4 要素がそれぞれ対応する演算器へと送られる．また，ロード・ストア命令用にインデックスレジスタを導入し，スカラコアでのアドレス演算のオーバヘッド軽減を図る．

ベクトル型演算加速機構は 16 個の FMA を持ち，32 個

†1 筑波大学 計算科学研究センター
{takahasi,msato,daisuke,taisuke,ukawa}@ccs.tsukuba.ac.jp
†2 東京大学 先端科学技術研究センター
nakamura@hal.rcast.u-tokyo.ac.jp
†3 （株）日立製作所 中央研究所
{hidetaka.aoki.rt, naonobu.sukegawa.eb}@hitachi.com
†4 （株）日立製作所 エンタープライズサーバ事業部
hideo.sawamoto.ad@hitachi.com

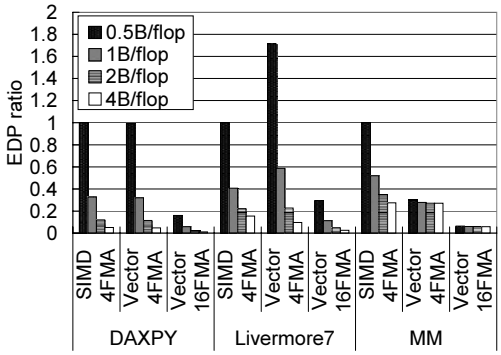

図 2 加速機構・オンチップメモリ間バンド幅による電力性能（4 コア）

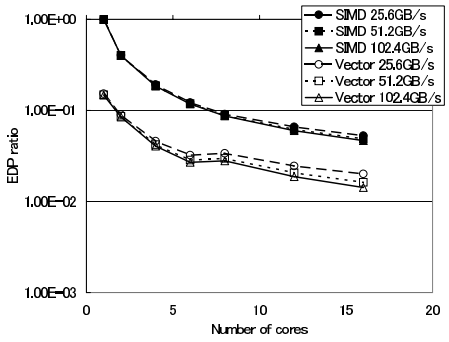

図 3 メモリバンド幅別 スケーラビリティ評価結果（CPU 2GHz 時）

のベクトルレジスタを持つ．1 ベクトルレジスタは通常のベクトルプロセッサよりも少ない 128 要素 (64bit) で構成される．1 サイクルに 1 ベクトルレジスタあたり FMA 数分のデータが演算器に送り込まれ，これが複数サイクル繰り返される．チェイニング機構を持つものとし，後続のベクトル演算のオペランドとして用いられる場合には，バイパスして後続の演算に供給される．なお，ベクトル型はアドレス演算のオーバヘッドが少ないと考えられるためインデックスレジスタは導入しない．

3. 性能評価

3.1 評価環境

本稿では想定するプロセッサアーキテクチャの性能をシミュレーションにて評価した．スカラコアとして SH-4A プロセッサを想定し，GDB5.0 付属の SH エミュレータにクロックシミュレーション機能，オンチップメモリ機能および演算加速機構を駆動するためのベクトル型および SIMD 型の命令セットを拡張することでシミュレータを開発した．

電力シミュレーションについては，11 に分割した各機能ブロックについてあらかじめ電力単価を算出しておき，シミュレーションにより得られる機能ブロックの駆動率を掛け合わせることで求める．なお，本稿では主記憶自体の消費電力とリーク電力は評価に含めない．

シミュレーションは，ベクトル型は FMA 数を 16 個と 4 個，SIMD 型は 4 個としてシミュレーションを行う（以降，“Vector-16FMA”，“Vector-4FMA” および “SIMD”）．

3.2 評価結果

まず各方式の基本性能を確認するために，すべてのデータがオンチップメモリにあると仮定して評価を行った．演算コアは 4 コアと仮定し，ベンチマークは 3 種類とし，DAXPY ループ，Livermore kernel 7 では N=200000，行列積演算（MM）では $N = 576 \times 576$ のデータサイズとした．結果を**図 2** に示す．本稿では指標としてエネルギー遅延積 (Energy Delay Product: 以降，EDP) を使用し，SIMD を 1 として各ベンチマークで正規化を行った．

DAXPY における SIMD と Vector-4FMA を比較すると，各バンド幅においてほぼ同じ EDP を示した．この時の性能は，バンド幅より求められる DAXPY の理想性能にほとんど等しい．また，Vector-16FMA については Vector-4FMA のほぼ 4 倍となる性能を示した．これらより基本的なベクトル処理においては演算加速機構によらず，高い性能を発揮できることが分かる．Livermore7 ではバンド幅が 2B/flop 以下の場合に SIMD の性能が Vector-4FMA の性能を上回った．これは SIMD 型はレジスタあたりの要素数が少ないためにループ間依存のあるデータを再利用でき，演算あたりのロード数を削減できるためである．これに対して行列積では，SIMD 型ではバンド幅に従って徐々に EDP が削減するのと比較して，ベクトル型では 4FMA・16FMA とも低いバンド幅から EDP が小さい．これは，SIMD 型のレジスタが 4×128 要素しかデータを保持できないのと比較して，ベクトル型は 128×32 要素と，レジスタブロッキングサイズを最大で 4 倍大きく取れるため，ベクトル型では低バンド幅でも高い電力性能を発揮できていると考えられる．

次に，オンチップメモリ・主記憶間データ転送を含めたシミュレーションで，コア数のスケーラビリティに関する評価を行った．行列サイズ $N = 1728 \times 1728$ の倍精度行列積演算において，DDR3 の実装を想定しオンチップメモリ・主記憶間のバンド幅を変化させて SIMD と Vector-16FMA についてシミュレーションを行った．演算加速機構とオンチップメモリ間のバンド幅は 4B/flop とした．電力性能比は SIMD 1 コア 25.6GB/s 時の性能を 1 として各 EDP を正規化した．結果を**図 3** に示す．

図 3 より，行列積については各方式において良好なスケーラビリティが得られることが分かった．またバンド幅を増加させることにより EDP は低下することが分かった．同 FMA 数で SIMD 型とベクトル型を比較した場合はベクトル型のほうが EDP が小さかった．これは前述のレジスタブロッキングサイズの差に加えて，FMA あたりのオーバヘッドはベクトル型のほうが少ないためだと考えられる．

4. まとめ

本稿ではオンチップメモリプロセッサへ多数の演算器を備えた演算加速機構を導入することを検討し，SIMD 型演算加速機構とベクトル型演算加速機構について性能評価を行った．シミュレーションの結果，おおむねベクトル型が優位であったが，同じコア・同 FMA 数であればベンチマークの特性により各方式の特性に優劣が左右された．

今後は実用に近い多様なアプリケーションでの評価が必要である．またリーク電流を含めた評価や，従来のキャッシュに演算加速機構を付加した評価との比較も行いたい．

謝辞 本研究の一部は文部科学省「次世代ＩＴ基盤構築のための研究開発」プロジェクト「低電力高速デバイス・回路技術・理論方式の研究開発」による．

参考文献

1) 中村宏ほか: ハイパフォーマンスコンピューティング向けアーキテクチャ*SCIMA*, 情報処理学会論文誌, Vol.41, No. SIG 5(HPS 1), pp. 15–27 (2000).
2) Kondo, M. et al.: Software-controlled on-chip memory for high-performance and low-power computing, *ACM SIGARCH Computer Architecture News*, Vol. 30, pp. 7–8 (2002).