

論文内容の要旨

論文題目 Software Controlled On-Chip Memory for High-Performance and
Low-Power Processor
(ソフトウェア可制御オンチップメモリを用いた高性能・低消費電力
プロセッサに関する研究)

氏名 近藤 正章

近年、プロセッサと主記憶の性能格差の問題が深刻化している。この問題に対処するために従来からキャッシュメモリが用いられているが、データセットの大きなハイパフォーマンスコンピューティング (HPC) 分野やマルチメディア系のアプリケーションなど、データ集約的 (data intensive) なアプリケーションではキャッシュが有効に機能しないことが多い。キャッシュ容量に比べデータセットが大きく、またデータの時間的局所性がほとんどないため、キャッシュミスによる主記憶へのアクセスが頻発し性能が大きく低下する。

キャッシュブロッキング (タイリング) は、ソフトウェア的な手法でデータの再利用性を向上させることで、データへのアクセスをキャッシュという上位階層の記憶に閉じるさせる手法であり、プロセッサチップ上に実装可能なキャッシュ容量がこれからも増大するという技術動向を考えると今後も有望な手法である。しかし、従来のキャッシュでは、データのアドレスとリプレースメントがハードウェアで制御されるため、どのデータをキャッシュに載せるか指定できない (配列 A は載せたいが B は載せたくない等)、ラインコンフリクトによる同一配列のブロック内データの干渉 (self-interference) や異なる配列間のデータの干渉 (cross-interference) を防げない、等の問題がある。self-interference の問題に対しては、適切なブロックサイズを選ぶアルゴリズムや、配列サイズを変更しコンフリクトを防ぐ手法などが提案されているが、いずれもキャッシュ構成に応じてプログラムを変更する必要がある、また cross-interference の問題は依然残る。

また、消費電力の面においてもキャッシュには問題点がある。近年では、キャッシュミス削減のため大容量で連想度の高いキャッシュが用いられるようになってきているが、このような状況のもと、チップ内の電力消費に対して、キャッシュにおける電力消費の占める割合が大きくなってきている。

従来の連想キャッシュでは、アクセスすべきデータは高々1つのウェイにしか存在しないにも関わらず、キャッシュアクセス時間の増加を防ぐために、タグの検索と同時に全てのウェイを並列にアクセスする。このため、該当するデータが存在しないウェイでは無駄に電力が消費されている。近年では、マルチメディア分野のアプリケーションではもちろんのこと、HPC 分野においても、計算機システムの消費電力、あるいは消費エネルギー削減の要求が高まっている。これは、大規模な並列計算機を構成する上で、放熱や設置面積などの面から消費電力を可能な限り小さく抑えることが重要なためである。今後、消費電力が全体の性能を制限する要素の1つになると考えるため、低消費電力化・低消費エネルギー化は今日の計算機システムにとって非常に重要な課題である。

上記のキャッシュの問題点に対処し、データ集約的なアプリケーションにおいて高性能かつ低消費電力なプロセッサの実現を目的として、本論文では従来のキャッシュに加えて、アドレス指定可能な主記憶の一部を SRAM として載せる (以降では、そのメモリを SCM: Software Controlled

on-chip Memory と呼ぶ)新しいアーキテクチャ, SCIMA (Software Controlled Integrated Memory Architecture)を提案する. また, キャッシュと SCM でハードウェアを共用し, 総容量一定のもと SCM とキャッシュ容量比を実行時に再構成できる機構も提案する.

SCM と主記憶との間のデータ転送は, ISA (命令セットアーキテクチャ)上に定義された命令により行われる. 従って, ユーザ(コンパイラ)による明示的なデータのアロケーション, 及びリプレースメントが可能となる. そのため, ラインコンフリクトによる性能低下は生じない. また, プロセッサチップと主記憶間のデータ転送を, より大きな粒度で行えることも SCIMA の利点である. メモリアクセス時のレーテンシの影響を抑え高い実効スループットを得るためには, データ転送の粒度を大きくし, 転送回数を減らすことが効果的である. しかし, キャッシュではラインサイズを大きくすると, 無駄な転送の増加やキャッシュ内のエントリ数が減少することによるコンフリクトミスの増加といった問題が生じる. SCIMA では SCM とオフチップメモリの間の転送をソフトウェアで制御するため, 必要なデータのみを大粒度で転送することができる.

SCIMA ではまた, SCM アクセスの際の消費電力を従来のキャッシュアクセスに比べ削減できる. SCM では, キャッシュとは異なりアクセスすべき位置がアドレスから一意に決定されるため, アクセスすべきデータが存在するウェイのみを選択的にアクセスすることができる. 従って, タグの検索と同時に全てのウェイを並列にアクセスする従来の連想キャッシュとは異なり, 無駄に電力が消費されることがない. さらに, SCIMA ではソフトウェア制御の SCM を用いることで, メモリトラフィックを最小限に抑えることが可能となる. この, オフチップメモリトラフィックの削減は, 計算機システムとしての低消費電力化を考える上で効果が大きい.

しかし, SCM を利用し高性能・低消費電力を得るためには, SCM 経由でアクセスするデータの選択とデータ転送のサイズやタイミングを指定しなければならない. それらの戦略は性能最適化・消費電力最適化のためには特に重要である. そのため, 本論文では SCM を用いるための最適化の指針, およびアルゴリズムも提案する. 最適化においては, 再利用性やデータアクセスの特徴に関して同じ配列内のデータは同じ性質を持つことが多いため, 配列を単位として考え, 利用される配列の特徴を, 再利用性とアクセスの連続性の観点から分類する. その分類に基づき各配列に対する最適化手法を示す.

本最適化手法を, NAS Parallel Benchmarks の CG および FT カーネルと, 実アプリケーションである QCD (Quantum ChromoDynamics) 計算に適用し, 本論文では SCIMA の性能と消費エネルギーに関して, 従来のキャッシュと比較評価を行なう. 評価は, サイクルレベルのシミュレーションにより行なった.

性能における評価結果より, SCIMA では従来のキャッシュのみを用いるアーキテクチャに比べ 1.5~3.0 倍高速であることがわかった. これは, SCIMA は SCM を用いることで, 利用性の有効活用, 大粒度データ転送による高実効スループット, そしてブロックストライドデータ転送機能による効率的なデータ転送とチップ内記憶領域の利用が実現でき, 高性能化が達成できるためである.

消費エネルギーにおける評価結果では, SCM アクセス時の選択的ウェイアクセス, およびオフチップメモリトラフィック削減により, 従来のキャッシュアーキテクチャに比べメモリシステムにおける低消費エネルギーを 5%~50%も削減できることがわかった.

これらの評価結果から, SCIMA では SCM を用いることにより, キャッシュアーキテクチャに比べ, 高性能かつ低消費電力を実現できるアーキテクチャであると結論付けることができる.