

# 実世界情報システムプロジェクト ~VR 研究グループ~

## 高性能低消費電力プロセッサの研究

南谷崇 中村宏

情報理工学系研究科システム情報学専攻 (先端科学技術研究センター)

### 1 はじめに

人間がリアルタイムで現実の世界とのインタラクションをとることが可能な VR 空間を構築するためには、高品質なマルチメディア情報端末を廉価に提供することが重要となる。しかしながら、マルチメディア情報端末に対する性能向上の要求は、デバイスの微細化による素子の速度向上を凌いでおり、従来のデジタルシステムの VLSI アーキテクチャを踏襲しては、その要求を満たすことができなくなる、という問題が指摘されている。この問題は具体的には以下のように説明される。大容量のデータを扱うマルチメディア処理においては、システム全体の処理能力は演算処理部の能力だけではなくメモリシステム側のデータ供給能力によっても決定される。半導体技術の進展により、素子の集積度と演算処理の速度は今後も向上すると期待されているが、DRAM, SRAM 等の記憶素子の速度は大きくは向上しないと予測され、また VLSI を実装するパッケージのピン数も物理的な制約によりそれほど増加しないため、プロセッサとメモリ間の転送スループットを向上させるのは困難な状況である。しかも、携帯容易さが重要な情報端末であるため、消費電力を極力抑えてこの問題を解決しないといけない。

そこで、本研究では素子集積度の向上を活用すべく、同一 VLSI 上にプロセッサコアと大容量のメモリを搭載する新しい VLSI アーキテクチャ SCIMA (Software Controllable Integrated Memory Architecture) を提案している。大容量メモリは、従来型のキャッシュだけでなく、新たにソフ

トウェアから制御可能なメモリとしても用いる。この可制御メモリは、アドレス指定可能としソフトウェアによるデータ配置と入れ替えを制御可能とする点が従来のキャッシュとは本質的に異なる。また、可制御メモリはキャッシュと違い、動作がソフトウェアによって陽に指定されるため、性能が予測可能でリアルタイム性の維持が容易である、という特徴がある。

本年度は、SCIMA アーキテクチャの検討を、性能と消費電力の両面から定量的に行った。

### 2 SCIMA

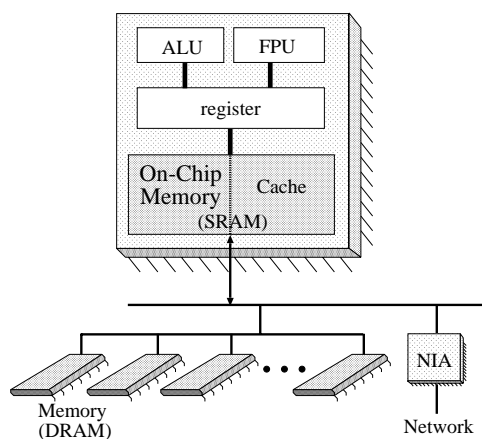


図 1: SCIMA の構成

SCIMA の構成を図 1 に示す。SCIMA はプロセッサ上にキャッシュだけでなく、アドレス指定可能な SCM (Software Controlled Memory) を搭

載する。従来のキャッシュでは、データのアドレスロケーションやリプレースメントがハードウェアで暗黙的に制御されるのに対し、SCM はそれらの制御をソフトウェアで明示的に行う。

**拡張命令** SCM へのデータ転送を制御するため、page-load/page-store と呼ぶ主記憶・SCM 間のデータ転送命令を追加する。本命令によるデータ転送は大きな粒度で行ない、オフチップメモリーレーテンシの影響を抑えることを狙う。

レジスタ・SCM 間のデータ転送は従来の load/store 命令により行う。前節で述べたアドレスマッピング機構により、load/store の対象アドレスが SCM 領域か否かを判定し、SCM 領域であれば SCM へのアクセスを、そうでない場合は通常のキャッシュアクセスを行なう。

**キャッシュ・SCM 統合機構** キャッシュと SCM に割り振られる容量をアプリケーションの性質に応じて変えるべく、総容量一定のもと SCM とキャッシュの容量比を実行時に再構成できる機構を提案しており [1]、具体的には、n-way 連想キャッシュの連続する一部の way を SCM に割り当てる。

### 3 SCIMA における低消費電力化

**メモリトラフィック削減による低消費電力化** キャッシュはハードウェア制御により決められたアルゴリズムでそれらの制御が行なわれるため、個々のプログラムに最適なデータアドレスロケーション、リプレースメントを行わせることは難しい。例えば、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キャッシュブロッキング (タイリング) を適用する場合、キャッシュでは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) により、オフチップメモリトラフィックが増加してしまう。

それに対し、SCIMA は SCM・主記憶間のデータ転送を page-load/page-store 命令により明示的に行うことで、SCM のデータアドレスロケーション・リ

プレースメントをユーザから制御可能で、必要なデータのみを必要なタイミングで転送することができるため、SCIMA ではキャッシュに比べ以下の点でトラフィックが削減できる。

- データの再利用性を最大限に活用できる
  - コンフリクトミスが生じない
  - 再利用性のあるデータが再利用性のないデータにより追い出されることがない
- 転送サイズが可変であるため、ストライド転送となる場合に、必要ないデータの転送が抑制できる

プロセッサ・主記憶間のデータ転送では、負荷容量の高い外部のメモリバスや I/O パッドを駆動するため、消費される電力が多い。キャッシュに比べオフチップメモリトラフィックの削減が期待できる SCIMA では、低消費電力化を図ることができると考えられる。

**選択的ウェイクアクセスによる低消費電力化** 従来の n ウェイク連想キャッシュでは、キャッシュアクセスが発生すると、アクセスされたアドレスの Index 部をデコードすることで、該当するキャッシュ内のセットを決定する。次に、該当セットのタグ、およびデータアレイの全てのウェイクを並列にアクセスし、それぞれのデータを読み出す。そして、読み出されたタグとアクセスされたアドレスの Tag 部分を比較し、一致するものがあればキャッシュヒットとなり、該当ウェイクのデータが選択される。このように、選択すべきデータがタグアレイの内容に依存するため、高速化のためには全てのウェイクを並列に読み出す必要がある。

一方、SCIMA においては、SCM へのアクセス時には、アクセスすべきセットは、アドレスの Index 部のみで決定される。従って、選択すべきデータのウェイクをキャッシュアクセス時より早い段階で決定することができ、SCM アクセス時には 1 つのウェイクのみを選択的にアクセスすることができることとなり、消費電力の削減につながる。さらに、選択すべきデータがタグの内容に依存し

表 1: キャッシュ/SCM アクセス回数

	Cache	SCIMA
	#access (% WP hit)	#access (% SCM access)
CG	1686219 (69%)	1731899 (96%)
FT	65400698 (84%)	50253470 (92%)
QCD	8620112 (89%)	8806997 (55%)

ないため、タグアレイへのアクセスを抑制することによる消費電力の削減も期待される。

## 4 評価

評価におけるベンチマークプログラムとして、NAS Parallel Benchmarks の中から CG, FT の 2 つのカーネル、および実アプリケーションである QCD(量子色力学) 計算を用いる。

図 2 に、各ベンチマークプログラムにおける、キャッシュ/SCM アクセスにおける消費エネルギー (*Echip*) と、バスにおける消費エネルギー (*Ebus*) の合計を示す。従来のキャッシュアーキテクチャとして “Cache” と “CacheWP” の 2 つを採用する。前者は通常のキャッシュであるが、後者は MRU アルゴリズムによりアクセスするウェイを予測し、そのウェイだけをまずアクセスするものである。また、SCIMA アーキテクチャとして “SCIMAall” と “SCIMAsel” を採用する。前者は SCM アクセス時にも全ウェイを並列にアクセスするものであり、後者は選択的にウェイをアクセスするものである。SCIMA では、アドレスからアクセスすべきウェイを一意に決定できるため、常に選択的ウェイアクセスが可能であるが、比較のために “SCIMAall” を評価した。

また、表 1 に、32B のキャッシュラインにおける両モデルのキャッシュ/SCM アクセス回数を示す。表 1 には、Cache モデルにでのウェイ予測の際のヒット率 (% WP hit) と、SCIMA モデルのキャッシュ/SCM アクセス回数中に SCM アクセスが占める割合 (% SCM access) も示している。

まず、選択的ウェイアクセスを行わない “Cache” と “SCIMAall” を比較する。FT を除き、キャッシュ/SCM アクセス回数はほとんど変わら

ないため(表 1 参照)、チップ内の消費エネルギー (*Echip*) に差はない。しかし、SCIMA ではバスの消費エネルギーが削減されているため、合計の消費エネルギーは SCIMA の方が少なくなっている。

次に、全ウェイを並列にアクセスせずに、データの存在するウェイをのみ選択的にアクセスすることで、キャッシュ/SCM アクセス時の消費エネルギーを削減した “CacheWP” と “SCIMAsel” について比較する。

CacheWP では、ウェイ予測が成功した場合、今回仮定した連想度 4 のキャッシュではタグアレイアクセスの消費エネルギーは変わらないものの、データアレイアクセスの消費エネルギーが 4 分の 1 になる。今回評価したベンチマークプログラムでは、ウェイ予測のヒット率が高いため(表 1 参照)、チップ内の消費エネルギーが 60% ~ 72% ほど削減されている。

また、SCIMAsel では、SCM アクセスの場合にデータアレイアクセスの消費エネルギーが約 4 分の 1 になるが、その他にタグアレイアクセスの消費エネルギーも削減できる。表 1 より、SCIMA ではロード/ストア命令の大部分が SCM へのアクセスであるため、チップ内の消費エネルギー削減率は 75% ~ 77% と CacheWP の場合に比べ、より高くなっている。

合計の消費エネルギーについて比較すると、SCIMAsel ではチップ内部、およびバスの消費エネルギーの両方が CacheWP に比べ削減されるため、両モデル間の消費エネルギーの差はさらに大きくなる。例えば FT では、約 50% もの消費エネルギーが、また最も低い CG の 128B ラインサイズの場合でさえ 5% の消費エネルギーが CacheWP に比べ削減されている。従って、SCIMA はメモリシステム全体として見た場合、キャッシュアーキテクチャに比べ低消費エネルギー化を達成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。

性能 次に、性能面での比較を行う。図 3 に Cache モデルと SCIMA モデルの実行時間を示す。両モデルとも、消費エネルギーは選択的ウェイアクセスを行なった場合を想定している。図中、“CPU

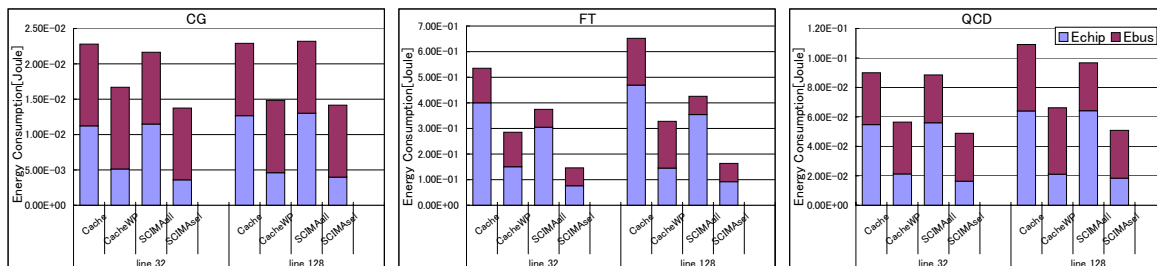


図 2: メモリシステム全体の消費エネルギー

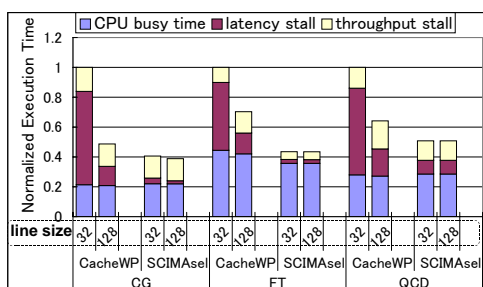


図 3: 実行時間

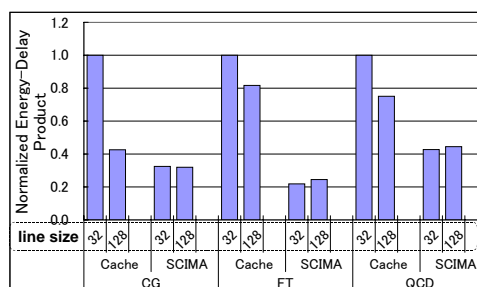


図 4: ED 積

“Busy Time” は CPU 処理に要する時間を表わし，“Latency Stall” と “Throughput Stall” は主記憶アクセスに要する時間であり，それぞれ主記憶アクセスのレイテンシに起因する処理時間と、主記憶のスループット不足に起因する処理時間を表わす。この図からわかるように，SCIMA は Cache に対して，主記憶アクセスに要する時間を短縮することで大きな性能向上を達成できることがわかる。

ED 積 性能と消費エネルギーを統合的に議論するため，図 4 に Cache モデルと SCIMA モデルの ED 積 (Energy-Delay Product) を示す。両モデルとも，消費エネルギーは選択的ウェイクアクセスを行なった場合を想定している。

図 4 では，SCIMA は Cache に比べ ED 積を大きく改善できることがわかる。この結果から，SCIMA はキャッシュアーキテクチャに対して，高性能かつ低消費エネルギーなプロセッサを実現す

るアーキテクチャであることがわかる。

## 5 まとめ

次世代マルチメディア情報端末を指向した，同一 VLSI 上にプロセッサコアと大容量のメモリを搭載する VLSI アーキテクチャを提案し，その有効性を性能と消費電力の観点から検討した。来年度以降は，今回扱っていないリーク電流に起因する消費電力の削減，および SCM を有効に活用可能なコンパイラの開発を引き続き実施する予定である。

## 参考文献

- [1] M.Kondo and H.Nakamura, “Reducing Memory System Energy by Software-Controlled On-Chip Memory”, IEICE Trans. on Electronics, Vol.E86-C, No.4, pp.580-588, 2003