

研究課題別評価

1 研究課題名: 命令列の多重入出力構造を利用した演算高速化

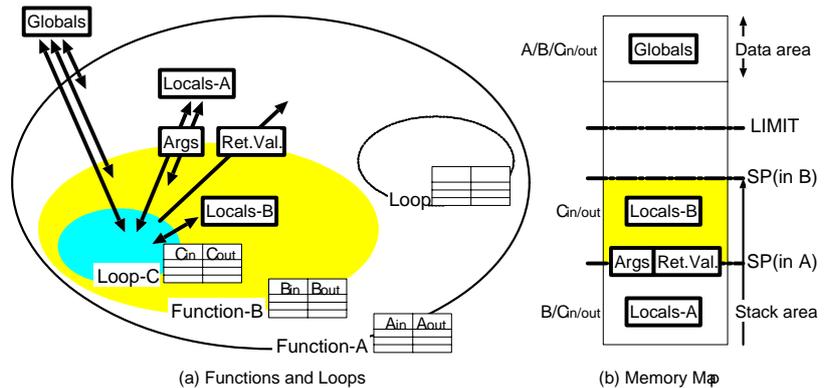
2 研究者氏名: 中島 康彦

3 研究の狙い:

一般ユーザにとって、プロセッサの性能向上は当然、かつ、バイナリ非互換は論外である。この厳しい制約条件の下で、演算および主記憶アクセスをいかに減らすかについて、様々な可能性を追求し、解決方法を提示する。関連研究には、値再利用と投機的マルチスレッド実行がある。しかし、1) 既存オブジェクトを高速化できない; 2) より大規模な再利用区間を認識できない; 3) 入力の変動変化に対応できない; 4) 主プロセッサの実行機構に投機的要素を残している; 5) 高速化機構を意識したプログラミングができない; など多くの問題点がある。これに対し、本研究は以下をねらう。

【ABI(Application Binary Interface)を利用した再利用区間の大規模化】 命令レベルではなく、関数やループを再利用区間とし、またABI

を利用することにより、再利用区間における大量のレジスタ/主記憶参照から、局所的なレジスタ/主記憶参照を排除する。この結果、専用命令を追加することなく、より大規模な再利用区間を確保するとともに、再利用の可

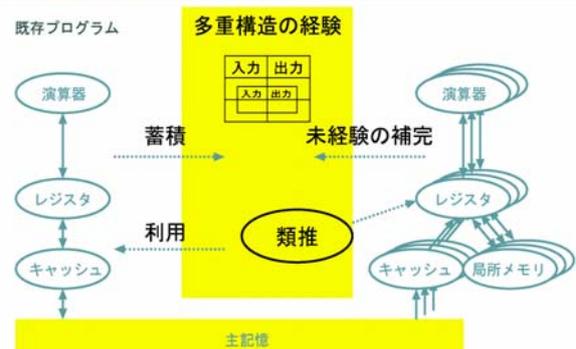


否検査に必要なレジスタ/主記憶参照を必要最小限(関数やループが意図する本質的な入出力)に抑える。

【多重再利用機構による入れ子構造への対応】 関数やループが入れ子になった構造において、各レベルに局所的ではないレジスタ/主記憶参照を上位レベルの入出力としても多重に登録することにより、各レベルの本質的な入出力を把握し、入れ子構造を一度だけ実行した場合においても各レベルの再利用を可能にする。大規模な再利用が不可能であっても、より小規模な再利用により補完できる。

【主プロセッサから投機的要素を排除することによる見通しのよい事前実行】 主プロセッサ(MP)とは別に設けた副プロセッサ(SP)に、再利用区間(関数やループ)の事前実行を行わせることにより、主プロセッサがはじめて遭遇する演算パターンについても、実行結果の検証を行うことなく、命令実行を完全に省略する。主プロセッサの命令実行機構に投機的要素が全く含まれない点が、既存研究とは大きく異なる。

多重構造を「経験と類推」に写像



【既存のプログラミングスタイルから意識可能な高速ハードウェア機構】 従来の命令レベル再利用機構では、粒度が小さすぎて、プログラマがその存在を意識したり明示的に利用することが困難である。これに対し、関数やループの挙動は、プログラミング時にある程度意識することができる。従来ハッシュなどを利用して検索処理を関数呼出しの形に記述することにより、関数実行時における再利用表の高速検索を利用した直接的な高速化を図ることができる。従来のプログラミングの枠組みの一部がハードウェアにより高速処理されることは、従来のベクトル機構を意識したプログラミングと同様に、新たな刺激になると考える。

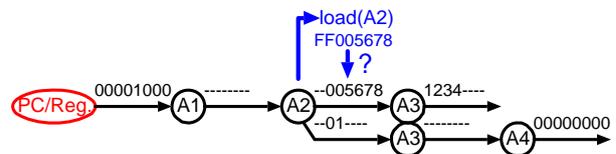
4 研究成果:

4.1 命令区間の入出力を動的に把握する仕組みを考案(特開 2004-258905、355397)

- 命令区間において、WRITE前にREADされるレジスタ/主記憶アドレスは入力
- WRITE後のREADは入力ではない
- レジスタ/主記憶アドレスへのWRITEは出力
- 以上を把握するには全てのWRITEを記録する必要があるが、局所変数領域については入出力とはなり得ないため記録から除外できる

4.2 複数組の入力系列を効率よく格納する仕組みを考案(特開 2005-92354)

- マスクが値固有のためハッシュ不可
- マスクと値を木構造の枝に格納
- 節には次に参照するアドレスを格納



4.3 入力列が一致するパスを効率よく探す仕組みを考案(特願 2004-176140、266056)

- 枝をマスク付きCAM(各図の左側)、枝に続く節をRAM(各図の右側)に格納
- 内容が変更されない主記憶アドレスは比較を省略できる
- 具体的には、木構造に分岐がなく、後続節の内容が全て登録時のまま(ストアされていない)であれば、後続節および枝の検索が不要(左図)
- 登録後に内容が変更された場合、該当アドレスに対応する節に枝を追加して、その部分だけを比較(右図)
- 木構造を分割して複数の枝を同時に検索することによる高速化も可能



PC/Reg.		Cache			
		Adr.	UP	Alt.	DN
00					
01	FF	E	A1	FF	
02					
03	01	E	A2	01	
04					
05	--005678	E	A3	03	
06					
07	1234----	E		05	

Select Output

PC/Reg.		Cache			
		Adr.	UP	Alt.	DN
00					
01	FF	E	A1	FF	A2 03
02					
03	01	C	A2	01	
04	--01---	X	A3	03	
05	--005678	E	A3	03	
06					
07	1234----	E		05	

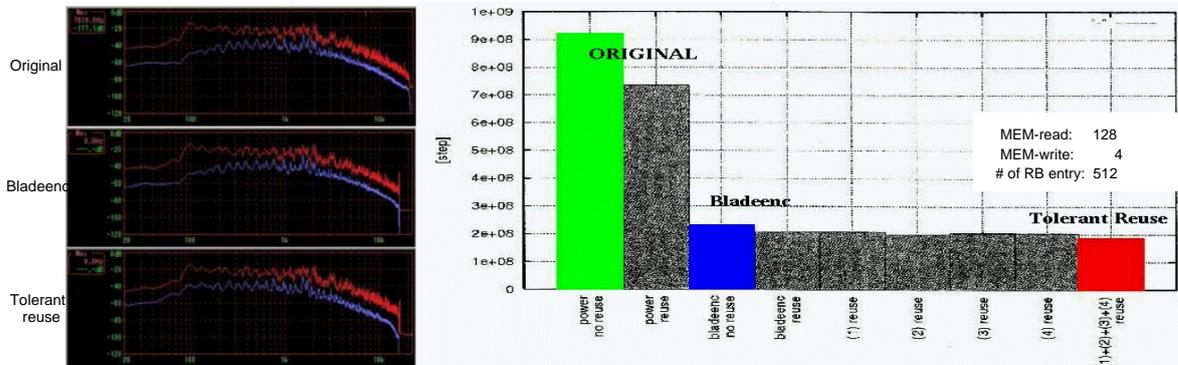
Select Output

4. 4 計算を曖昧化しても精度が落ちにくいマルチメディア処理への応用(ソース書換え)

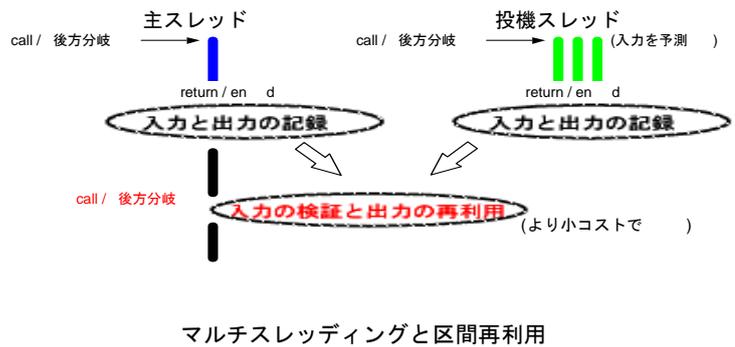
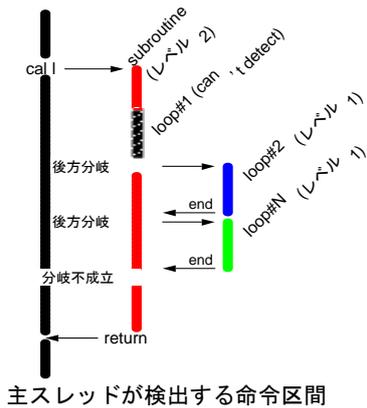
- 差分計算の精度を落とした曖昧再利用により、ステレオ画像処理を2倍以上高速化



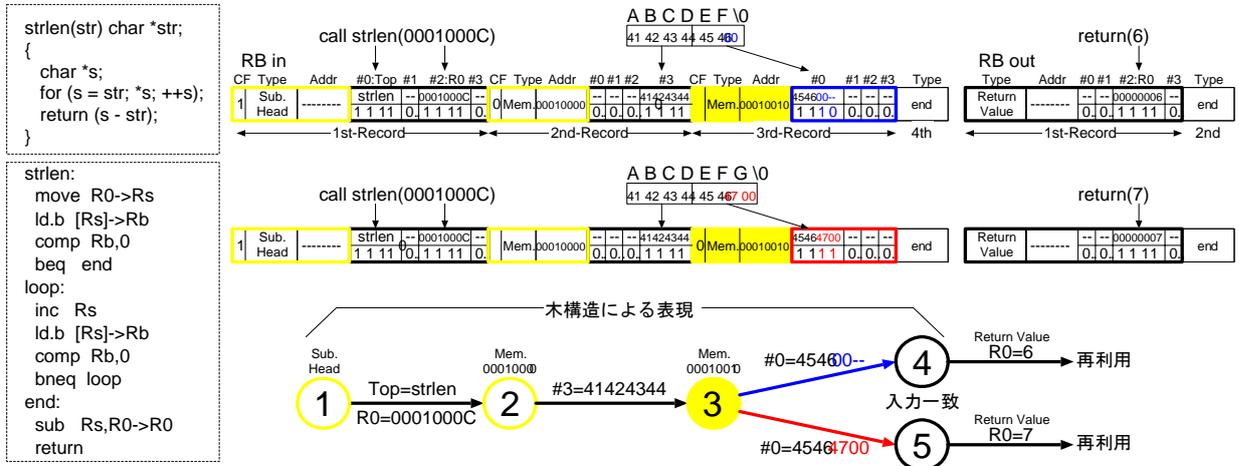
- 関数計算の精度を落とした曖昧再利用により、MP3圧縮処理を4倍以上高速化



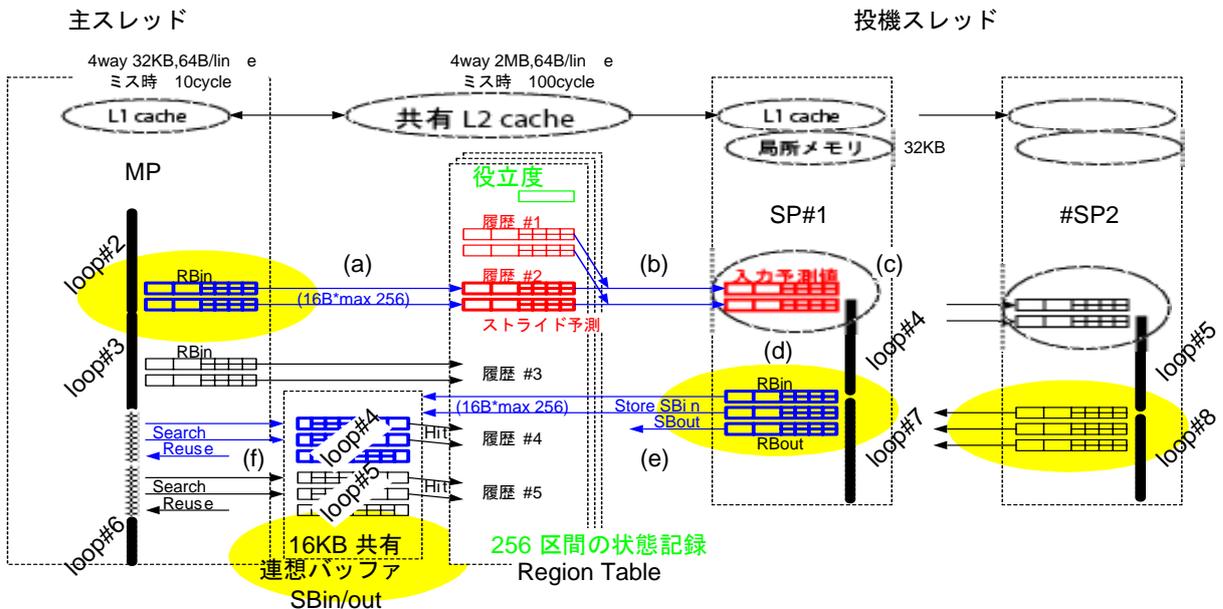
4. 5 再利用/事前実行の仕組みを既存プログラムに適用



各スレッドごとに命令区間の入出力をリスト(16バイト×64レコード×6レベル)に記録し、区間終了時に木構造として共有連想バッファ(16Kバイト)に集積するモデル



- 「入力履歴⇒予測⇒投機実行⇒記録⇒主スレッドが再利用」の連携による高速化



4. 6 シミュレーションによる実行モデルの評価

- 大容量CAMによるアクセラレータ(写真)を開発して本機構が正しく動作することを確認するとともに、シミュレーション速度を60倍高速化

- | | spec95 | spec2k | |
|-----------------------------|---------|---------|-----|
| ● 1スレッド最大/平均 | 28/ 3% | 52/ 6% | 高速化 |
| ● 4スレッド最大/平均 | 158/28% | 150/20% | 高速化 |
| ● 多重区間の投機実行が有効 | 20% | 7% | 改善 |
| ● 比較削減機構が有効 | 11% | 13% | 改善 |
| ● コスト評価機構が有効 | 8% | 9% | 改善 |
| ● 数百命令規模の命令区間が高速化に寄与する | | | |
| ● 二次キャッシュミスの低減効果を兼ね備えることを確認 | | | |



4.7 SP間待ち合わせ機構によるSPパイプライン機構と連想バッファのブロック化

- SP間のデータ依存関係を参照アドレスの履歴/値の履歴/ストア回数に基づいてハードウェアが調停する機構を考案(特願 2004-324348、347124)
- 連想バッファを分割して各ブロックの高速化と省電力を図る機構を考案(特願 2005-234806)
- ブロック分割しても再利用効率の低下が軽微であることを確認

Base: 2-waySS SPARC

1K-MW: 連想バッファ(16KB) 3台の SP 間待ち合わせ機能付き

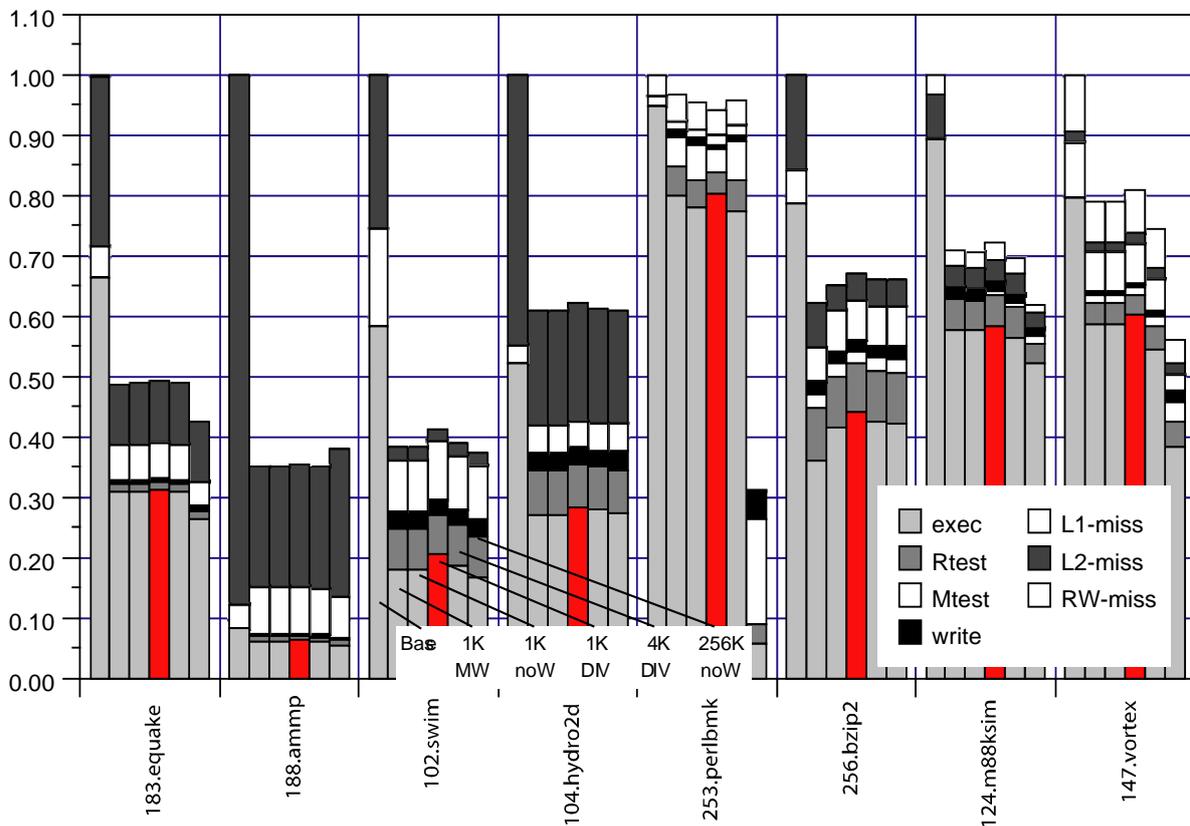
1K-noW: 連想バッファ(16KB) 待ち合わせ無しの従来方式【参考】

1K-DIV: 連想バッファ(16KB) 64 行×16 ブロックに分割(赤色部分)

4K-DIV: 連想バッファ(64KB) 256 行×16 ブロックに分割

256K: 連想バッファ(4M) 分割無しの従来方式【参考】

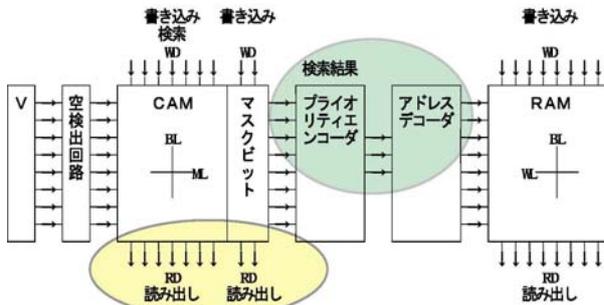
縦軸は、再利用なしの通常実行を1とした場合のサイクル数



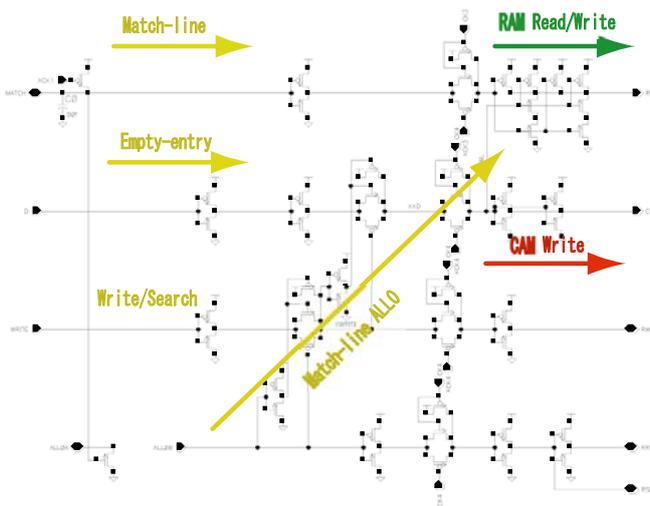
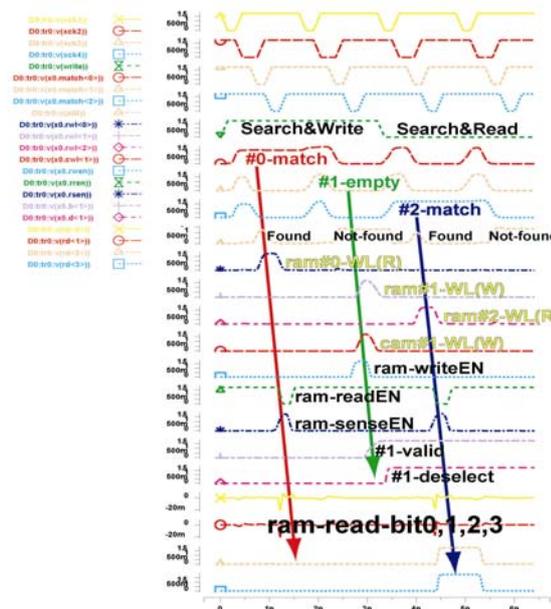
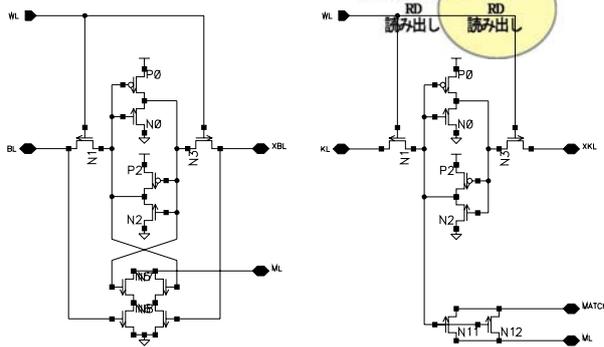
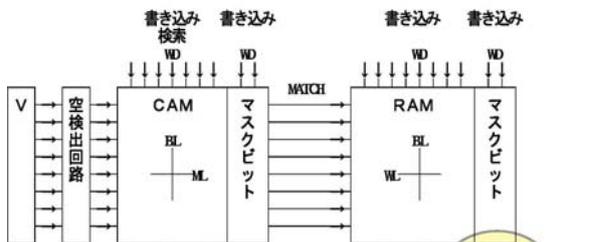
4. 8 連想バッファの高速化(特願 2005-234806)

- 汎用CAMでも再利用機構を実現できたが、もっと高速化できるはず
- 書き込む前に検索することで、検索の結果アサートされるマッチラインをたかだか1本に保障できるならば、プライオリティエンコーダやアドレスデコーダを省略できる
- 命令区間の入力登録に必要な「検索 & 書き込み」は、R/W用ビットラインと Comparand用ビットラインを共有することにより、高速化可能
- 命令区間の実行履歴の生成に必要な「検索 & 読み出し」は、CAMのマスクビット部分をRAM側にも持たせることで、高速化可能
- 検索結果に従って、直ちにCAM+RAM書き込みまたはRAM読み出しを行う機構を考案
- 共有連想バッファ 1K行(16KB)を 64行×16ブロックに分割した各々をトランジスタレベル設計(0.18um)しHSPICEでサイクルタイム1.6ns動作を確認

【汎用CAM】 検索と書き込みは別サイクル、検索と読み出しは別サイクル



【専用CAM】 検索と書き込みが同時サイクル、プライオリティエンコーダが不要
RAMのマスクビットを読み出すことにより、検索と読み出しを同時サイクルに可能



5 自己評価:

研究計画書に記載した当初の研究構想に対する達成状況を列挙する。

【関数およびループを対象とする区間再利用に必要な様々な機構および性能の評価】 スタックポインタを用いる多重区間再利用機構や、命令区間評価機構によるスレッド投機との連携を柱とする様々な関連機構を考案し、定量的な評価を行った。論文6件、口頭発表3件、特許9件を報告した。十分な成果をあげたと考えている。

【本機構を前提としてプログラミングを行う場合の適用範囲と効果の評価】 画像圧縮、音声圧縮、ステレオ画像処理、音声認識などのアプリケーションを対象として、特に曖昧再利用に着目したプログラミングを追及し、メディア命令を用いる場合とあわせて評価を行った。論文5件、口頭発表4件を報告した。曖昧再利用に関する研究はまだ少なく、十分な成果をあげたと考えている。

【回路レベルの実現方法の検討および評価】 実用化の鍵となる、極めて高速な Search&Write および Search&Read 機能のための回路構成を考案し、トランジスタレベル設計を併用する性能評価を行った。口頭発表3件、特許1件を報告した。論文は現在投稿中である。

【CAM内蔵型FPGAを用いた試作と回路規模の評価】 CAM内蔵型FPGAは容量が極めて小さいため、汎用大容量CAMの導入を決定し、FPGA-CAM混載PCIボードの設計および評価を行った。インタフェース仕様書策定、ドライバ作成、動作検証、様々な障害回避手段の考案と適用を行い、安定動作にこぎつけた。実CAMによる論理確認ができたことは意義深いと考えている。口頭発表3件を報告した。

以上、当初の目標は十分に達成できたと考えている。今後は、SPARCアーキテクチャに依存しない一般的な手法として改良を重ね、組み込み用プロセッサなどに対しても適用可能な技術として育てていきたい。

6 研究総括の見解:

プロセッサと主記憶の速度比が拡大し、従来の命令レベル並列処理方式が限界に近づきつつある状況において、命令区間の再利用性および予測可能性を利用して、既存プログラムをさらに大幅に高速化する基本技術である。バイナリアノテーション(ロードモジュールの静的解析に基づく付加情報の埋め込み)や再コンパイルを必要とすることなく、既存プログラムをそのまま使うことができるため、早期の実用化が可能と思われる。独創性の高い研究であり、数多くの特許も出しており新規性、進歩性にも優れていると認められる。今後は本研究成果の実用化のために努力されることを期待している。

7 主な論文等:

論文

- (1) 木郁真, 池内康樹, 津邑公暁, 中島康彦, 中島浩: “再利用によるGAの高速化手法”, 情報処理学会論文誌:コンピューティングシステム, ACS12, pp.129-143, Nov. (2005).
- (2) 中島康彦ほか: “外部連想バッファを備えるSpMTモデルの分析”, 先進的計算基盤システムシンポジウムSACIS2005 論文集, pp.397-406, May. (2005).
- (3) 津邑公暁, 清水雄歩, 中島康彦ほか: “ステレオ画像処理を用いた曖昧再利用の評価”, 情報処理学会論文誌:コンピューティングシステム, ACS3, pp.246-256, Sep. (2003).

- (4) 中島康彦ほか: “動的命令解析に基づく多重再利用および並列事前実行”, 情報処理学会論文誌:コンピューティングシステム, ACS2, pp.1-16, Jul. (2003).
- (5) 中島康彦ほか: “関数値再利用および並列事前実行による高速化技術”, 情報処理学会論文誌:ハイパフォーマンスコンピューティングシステム, HPS5, pp.1-12, Sep. (2002).

特許

- 中島康彦:“データ処理装置”, 特願 2005-234806 (H17. 8. 12)
- 中島康彦:“データ処理装置, データ処理プログラム, およびデータ処理プログラムを記録した記録媒体”, 中華民国出願番号 094109781 (2005. 03. 29)
- 中島康彦:“データ処理装置”, 特願PCT/JP2005/005591 (H17. 3. 25)
- 中島康彦:“データ処理装置, データ処理プログラム, およびデータ処理プログラムを記録した記録媒体”, 特開 2004-258905 (H15. 9. 12)
- 中島康彦:“データ処理装置, データ処理プログラム, およびデータ処理プログラムを記録した記録媒体”, 特開 2005-92354 (H15. 2. 25)

受賞 なし

招待講演

- 中島康彦:“コンピュータ・アーキテクチャに関わったこれまでの20年, これからの20年”, 奈良先端科学技術大学院大学 特別講演会, Feb. (2006).