

シングルチップマルチプロセッサにおける JPEG エンコーディングのマルチグレイン 並列処理

小 高 剛[†] 内 田 貴 之[†],
木 村 啓 二[†] 笠 原 博 徳[†]

近年の JPEG, MPEG などを用いたマルチメディアコンテンツの増加にともない, これらマルチメディアアプリケーションを効率良く処理できる低コスト, 低消費電力かつ高性能なプロセッサの開発が望まれている. 特に, 複数のプロセッサコアを搭載したシングルチップマルチプロセッサは命令レベル以外の並列性も自然に引き出すことができ集積度向上に対しスケーラブルな性能向上が得られるアーキテクチャとして注目されている. 本論文では, JPEG エンコーディングのシングルチップマルチプロセッサ用マルチグレイン 並列処理手法を提案するとともに, その性能評価を行う. 評価の結果, シンプルなシングルイシュープロセッサを 4 基搭載した OSCAR 型シングルチップマルチプロセッサアーキテクチャでは逐次実行に対して約 3.59 倍の性能向上が得られスケーラブルな性能向上が得られることが確かめられた.

JPEG Encoding Using Multigrain Parallel Processing on a Single Chip Multiprocessor

TAKESHI KODAKA,[†] TAKAYUKI UCHIDA,[†] KEIJI KIMURA[†]
and HIRONORI KASAHARA[†]

With the recent increase of multimedia contents using JPEG and MPEG, low cost, low power consumption and high performance processors for multimedia application have been expected. Particularly, single chip multiprocessor architectures having simple processor cores that will attain scalability and cost performance are attracting much attention to develop such processors. Single chip multiprocessor architectures allow us to exploit coarse grain task level and loop level parallelism in addition to the instruction level parallelism, so parallel processing technology is indispensable to get us scalable performance improvement. This paper describes a multigrain parallel processing scheme for the JPEG encoding for a single chip multiprocessor and its performance. The evaluation shows an OSCAR type single chip multiprocessor having four single-issue simple processor cores gave us 3.59 times speed-up.

1. はじめに

最近のマルチメディアコンテンツの増加に従い JPEG, MPEG などのメディア系アプリケーションを効率良く処理できる低コストかつ低消費電力の高性能プロセッサの開発が望まれている. このようなニーズに対応するために CPU ベンダからは Intel の PentiumIII などに搭載されている SSE¹⁾ や, 日立の SH4²⁾, 富士通の FR500³⁾ などのようにマルチメディア用命令セットを追加し処理能力を向上させる試みが行われ

ている. これらのプロセッサは, スーパスカラアクトクチャや VLIW アーキテクチャによる命令レベル並列性の利用や SIMD 命令によるマルチメディア処理で多用される内積演算やベクトル変換演算など同一命令が連続する処理の高速化により処理能力の向上をねらっている. しかし, これらのような命令レベル並列性を用いるアーキテクチャでは命令レベル細粒度並列性の限界により投入されたトランジスタ資源に見合うスケーラブルな性能向上が困難になってきている.

このような状況から, プロセッサコアを複数搭載したシングルチップマルチプロセッサアーキテクチャは集積トランジスタ数の増加に対しスケーラブルな性能向上が得られるアーキテクチャとして注目を集めている. また, シングルチップマルチプロセッサアーキテクチャでは, 命令レベル並列性に加え, ループレベル,

[†] 早稲田大学電気電子情報工学科
Department of Electrical, Electronics and Computer
Engineering, Waseda University
現在, 三菱電機株式会社
Presently with Mitsubishi Electric Corporation

サブルーチンレベルなどより多くの並列性を利用することも可能である。ただし、これら複数レベルの並列性を利用するためにはアプリケーションからの並列性抽出が重要となる。

シングルチップマルチプロセッサを用いたメディア系アプリケーションの高速化として、NECのMP98⁴⁾では、1チップ上に4CPUを搭載しマルチスレッド処理により高速化を行っており、Stanford大学では、2次キャッシュ共有型のシングルチップマルチプロセッサHydra⁵⁾上でのメディアアプリケーションの高速化が提案されている⁶⁾。

本論文では、離散コサイン変換や量子化、エントロピー符号化といったマルチメディア処理で多用されるアルゴリズムを利用した画像圧縮処理であるJPEGエンコーディングにおけるマルチグレイン並列処理手法を提案し、提案手法を適用したJPEGエンコーディングのOSCAR型シングルチップマルチプロセッサ上での評価結果を述べる。

以降、2章でマルチグレイン並列処理、3章で提案するJPEGエンコーディングのマルチグレイン並列処理手法、4章で本論文の評価で用いるOSCAR型シングルチップマルチプロセッサアーキテクチャ、5章でマルチグレイン並列処理を適用したJPEGエンコーディングのOSCAR型シングルチップマルチプロセッサ上で行った性能評価について述べる。

2. マルチグレイン並列処理

ここでは、OSCAR型シングルチップマルチプロセッサで扱うマルチグレイン並列処理技術について述べる。マルチグレイン並列処理とは、ループやサブルーチンなどの粗粒度タスク間の並列処理を利用する粗粒度タスク並列処理(マクロデータフロー処理⁷⁾)、ループイタレーションレベルの並列処理である中粒度並列処理、基本ブロック内部のステートメントレベルの並列性を利用する近細粒度並列処理⁸⁾を階層的に組み合わせてプログラム全域にわたる並列処理を行う手法である。

2.1 粗粒度タスク並列処理⁷⁾(マクロデータフロー処理)

マクロデータフロー処理では、ソースとなるプログラムを疑似代入文ブロック(BPA)、繰返しブロック(RB)、サブルーチンブロック(SB)の3種類の粗粒度タスク(マクロタスク(MT))に分割する。ここで、BPAは基本的には通常の基本ブロックであるが、並列性抽出のために単一の基本ブロックを複数に分割したり、逆に複数の基本ブロックを融合して1つのBPAを生成する。MT生成後、コンパイラはBPA、RB、

SBなどのMT間のコントロールフローとデータ依存を解析しそれらを表したマクロフローグラフ(MFG)を生成する。さらにMFGからMT間の並列性を最早実行可能条件解析により引き出し、その結果をマクロタスクグラフ(MTG)として表現する。その後、コンパイラはMTG上のMTをプロセッサあるいは複数のプロセッサエレメント(PE)を1つのグループとしたプロセッサグループ(PG)に割り当てる。なお、このグループはプログラム中の並列性に応じソフトウェア的に形成される仮想的なものでハードウェア的なグループ化とは異なる。

2.2 中粒度並列処理(ループ並列処理)

PGに割り当てられたMTがDoall可能なRBである場合、このRBはPG内のプロセッサエレメント(PE)上で、イタレーションレベル並列実行される。

2.3 近細粒度並列処理⁸⁾

PGに割り当てられたMTが、BPAや中粒度並列処理あるいはループボディ部に粗粒度並列処理を適用できないRBである場合、それらはステートメントレベルのタスクに分割され、PG内のPEにより並列処理される。

近細粒度並列処理においては、基本的にBPA内のステートメント、もしくはIF-THEN-ELSEなどで囲まれた複数ステートメントから構成される疑似代入文を1つの近細粒度タスクとして定義する。その後、近細粒度タスク間のデータ依存を解析してタスクグラフを作成し、このタスクグラフ上のタスクを、データ転送・同期オーバーヘッドを考慮して実行時間を最小化できるように各PEにスタティックにスケジューリングする。スケジューリング後、PEに割り当てられたタスクに対応する命令列を順番に並べデータ転送命令や同期命令を必要な箇所に挿入し各PEごとに異なるマシンコードを生成する。

3. JPEGエンコーディングアルゴリズムのマルチグレイン並列処理手法

ここでは、JPEGエンコーディングアルゴリズムのマルチグレイン並列処理手法について述べる。本論文で扱うJPEGエンコーディングアルゴリズムは、MediaBench⁹⁾に収録されているJPEGベンチマークプログラムである“jpeg-v6a”をベースとする。まず、JPEGエンコーディングにおける各処理について簡単に述べた後、マルチグレイン並列性を利用する手法を述べる。

3.1 JPEGエンコーディングアルゴリズム^{10),11)}

JPEGエンコーディングは以下の6つの処理から

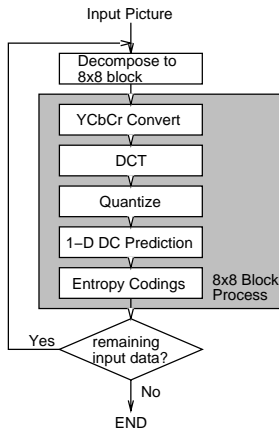


図 1 JPEG エンコーディング実行ブロック図
Fig. 1 Execution flow of JPEG encoding.

なる。

8 × 8-pixel ブロック分割

入力画像データを JPEG 基本処理単位である 8 × 8-pixel ブロックに分割する。

YCbCr 変換 (YCbCr)

分割された 8 × 8-pixel ブロックのデータを JPEG のデータフォーマットである YCbCr フォーマットへ変換する。

2次元離散コサイン変換処理 (DCT)

YCbCr フォーマットへ変換された 8 × 8 ブロックデータに対し 2次元離散コサイン変換により画像信号を空間周波数へ変換する。なお、変換後の 2次元配列 S_{ij} は、 S_{00} から S_{77} までの 64 要素存在し、 S_{00} を直流 (DC) 係数、その他の 63 要素は交流 (AC) 係数と呼ばれる。ここで、 S_{ij} は i または j が大きいほど高い周波数成分に相当する。

一様量子化 (Quantize)

量子化テーブルを用い DCT 係数に対し 1 次元一様量子化を行う。

量子化 DC 係数 1 次元予測

処理を行っている 8 × 8 ブロックの量子化 DC 係数と 1 つ前の 8 × 8 ブロックの量子化 DC 係数を減算し 1 次元 DC 予測値を計算する

エントロピー符号化

1 次元 DC 予測値および量子化 AC 係数を可変長符号を用いてエントロピー符号化を行う。

上記 6 つの処理を入力画像データが終わるまで繰り返して行われる。図 1 に JPEG エンコーディング実行ブロック図を示す。

3.2 JPEG エンコーディングのマルチグレイン並列処理

ここでは、JPEG エンコーディングのマルチグレイン並列処理手法について述べる。

3.2.1 粗粒度並列性の抽出

3.1 節で述べた基本的な JPEG エンコーディングアルゴリズムでは、量子化 DC 係数の 1 次元予測 (1-D DC Prediction) で 1 つ前の 8 × 8 ブロックの量子化 DC 係数を用いるため、8 × 8 ブロック間でのデータ依存が存在する。しかし、その他の YCbCr 変換、DCT およびエントロピー符号化処理では 8 × 8 ブロック内のデータのみ扱うため 8 × 8 ブロック間のデータ依存は存在しない。これらのデータ依存を考慮し、並列処理を効果的に行うために 1-D DC Prediction 処理および、1-D DC Prediction の処理結果の 1 次元 DC 予測値を用いる 1 次元 DC 予測値のエントロピー符号化処理 (DC EC) を図 2 (a) のように YCbCr, DCT, Quantize および量子化 AC 係数のエントロピー符号化処理 (AC EC) の後ろに移動する。そして、8 × 8 ブロック処理を JPEG エンコーディングの基本処理単位として、YCbCr, DCT, Quantize および AC EC を 1 つのマクロタスク (図 2 (a) MT_a) として定義し、同様に 1-D DC Prediction および DC EC も 1 つのマクロタスク (図 2 (a) MT_b) として定義する。

MT_a , MT_b の各 PG への割当ては、各 PG で処理される MT_a および MT_b の数が等しくなるように割り当てる。その際、データ転送の最小化を考慮し同じ 8 × 8 ブロックを同一 PG 内で処理できるように MT_a , MT_b のスケジューリングを行いスタティックに各 MT を割り当てる。たとえば、16 × 16 pixel の入力データを 2 つの PG で実行する場合の各 MT の割当ては、図 2 (b) に示したようになる。ここで、 MT_{ai} および MT_{bi} は、入力データを 8 × 8 ブロックに分割したときシーケンシャルスキャン順での i 番目のブロックである block i を処理する MT に相当する。このときの実行イメージは、まず、入力データは block1 から block4 の 4 つの 8 × 8 ブロックに分割され、PG1 では block1 と block2 を処理する MT_{a1} および MT_{a2} が実行され、同様に PG2 では block3 と block4 を処理する MT_{a3} および MT_{a4} が実行される。このとき、各 MT_a 間では依存がないためこれらの MT_a は全 PG で並列に処理される。すべての MT_a が実行された後、 MT_b 実行前に、PG2 で処理されている block3 の 1-D DC Prediction で必要とされる block2 の量子化 DC 係数を PG1 から PG2 へ転送する。データ転送後、 MT_b は全 PG で並列に処理される。

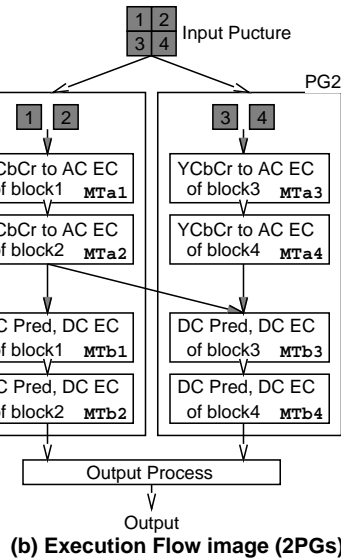
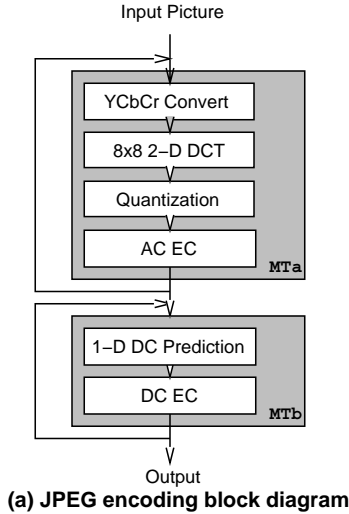


図 2 並列化 JPEG エンコーディング
Fig. 2 Parallel execution flow of JPEG encoding.

この粗粒度タスク並列性の抽出は一般最適化により抽出可能である。まず、図 3 (a) のようにループにより表現された JPEG エンコーディング処理をループアンローリングにより展開する (図 3 (b)) 。ループアンローリング後、1 つの 8×8 ブロックを処理する YCbCr, DCT, Quantize, AC EC, 1-D DC Prediction, DC EC をそれぞれ粗粒度タスクとして定義し、粗粒度タスク間並列性抽出を行う。図 3 (c) に、ループアンローリング後の粗粒度タスク間並列性抽出結果を表したマクロタスクグラフを示す。各ノードは粗粒度タスクを示しノード内の番号は図 3 (b) の各処理に対応する。また、ノードから出る実線はデータ依存を表す。この

```

YCbCr of block1 (1, 1)
DCT of block1 (1, 2)
Quantize of block1 (1, 3)
1-D DC Prediction of block1 (1, 4)
DC EntropyCoding of block1 (1, 5)
AC EntropyCoding of block1 (1, 6)
do i=1, 4
YCbCr of 8x8block
DCT of 8x8block
Quantize of 8x8 block
1-D DC Prediction of 8x8block
DC EntropyCoding of 8x8block
AC EntropyCoding of 8x8block
enddo
YCbCr of block2 (2, 1)
DCT of block2 (2, 2)
Quantize of block2 (2, 3)
1-D DC Prediction of block2 (2, 4)
DC EntropyCoding of block2 (2, 5)
AC EntropyCoding of block2 (2, 6)
YCbCr of block3 (3, 1)
DCT of block3 (3, 2)
Quantize of block3 (3, 3)
1-D DC Prediction of block3 (3, 4)
DC EntropyCoding of block3 (3, 5)
AC EntropyCoding of block3 (3, 6)
YCbCr of block4 (4, 1)
DCT of block4 (4, 2)
Quantize of block4 (4, 3)
1-D DC Prediction of block4 (4, 4)
DC EntropyCoding of block4 (4, 5)
AC EntropyCoding of block4 (4, 6)

```

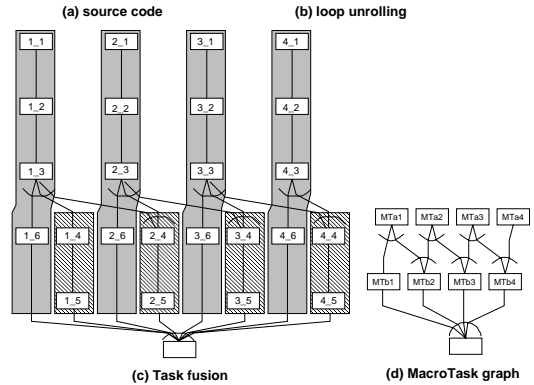


図 3 粗粒度並列性抽出

Fig. 3 Exploiting coarse grain parallelism.

粗粒度タスク間並列性抽出結果より粗粒度タスクを網掛け部分の粗粒度タスクを 1 つの粗粒度タスクとして融合することにより図 3 (d) のようなマクロタスクグラフが得られ、本章で提案した粗粒度並列処理が可能となる。

3.2.2 近細粒度並列性の抽出

ここでは、 MT_a , MT_b 内の近細粒度タスク定義およびスケジューリングについて述べる。

YCbCr 変換の変換式は、入力信号により違うが今回の評価では一般的に用いられる RGB 信号からの変換を使用する。RGB から YCbCr への変換は

$$\begin{aligned}
 Y_{ij} &= R_{ij} * 0.29 + G_{ij} * 0.58 + B_{ij} * 0.11 \\
 Cb_{ij} &= -R_{ij} * 0.16 - G_{ij} * 0.33 + B_{ij} * 0.50 \\
 Cr_{ij} &= R_{ij} * 0.50 - G_{ij} * 0.41 - B_{ij} * 0.08 \\
 (0 \leq i \leq 7, 0 \leq j \leq 7)
 \end{aligned}$$

で表される。jpeg-v6a では、この式をループによる繰返しにより各要素の演算を行っているが、近細粒度並列性を高めるため今回の評価ではループの展開を行う。各ステートメントを近細粒度タスクとして定義すると式よりそれぞれの近細粒度タスク間にはデータ依存が存在しない。そのため並列処理可能であることが分かる。

2次元離散コサイン変換 (DCT) は、1次元DCTを行方向へ1回、列方向へ1回処理することにより

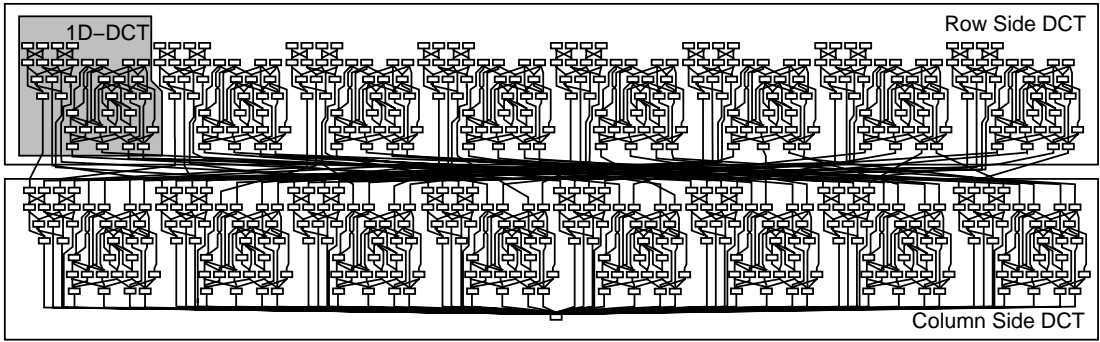


図 4 2次元 DCT タスクグラフ
Fig. 4 Task graph of 2-D DCT.

実現される。jpeg-v6a では、この処理を 1 次元処理をループによる繰返し実行を行うことにより実現しているが、近細粒度並列性を高めるため今回の評価ではループを展開した。ステートメントを近細粒度タスクとして定義し、そのときの 2 次元離散コサイン変換のタスクグラフを図 4 に示す。各ノードは近細粒度タスクを示しノードより出る実線はデータ依存を表している。図のように、ループを展開することによって行方向 1 次元 DCT 処理中では各行方向の処理間ではデータ依存が存在しないため並列処理可能なタスクが多数存在していることが分かる。列方向 1 次元 DCT 処理中に関しても同様である。

一様量子化処理は、以下の式で示される。

$$Sq_{ij} = \text{round} \left(\frac{S_{ij}}{q_{ij}} \right) \quad (0 \leq i \leq 7, 0 \leq j \leq 7)$$

この場合も、YCbCr 変換同様、ループを展開し各ステートメントを近細粒度タスクとして定義することでタスク間でのデータ依存は存在せず並列実行可能であることが分かる。

エントロピー符号化処理は、1 つの量子化 DCT 係数を符号化する処理は図 5(b) のように IF-THEN-ELSE により記述された処理になる。データ転送のオーバーヘッドや同期オーバーヘッドなどを考慮し、この IF-THEN-ELSE 文で囲まれた複数のステートメントを疑似代入ステートメントとして扱い近細粒度タスクとして定義する。エントロピー符号化処理は、非 0 係数と非 0 係数が出現するまでの間の 0 の連続値を用いて符号化を行うため、図 5(b) のように直前までの 0 の連続値によるデータ依存が存在し並列性の抽出が難しい。そこで、今回の評価では図 5(c) のようにエントロピー符号化処理を仮分割し並列性を抽出する。この手法は、ある量子化 DCT 係数のエントロピー符号化処理のところで仮分割を行うことにより、分割され

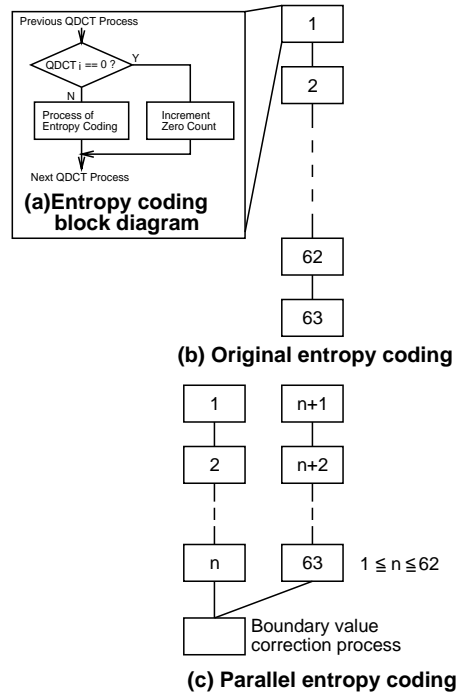


図 5 エントロピー符号化処理の近細粒度タスク
Fig. 5 Near fine grain tasks of entropy coding.

た処理はそれぞれデータ依存がなくなり並列実行可能とするものである。ただし、分割境界では 0 の連続値が分離されてしまうため処理の最後に境界部分の値を補正する処理を行う必要がある。

仮分割の際、分割数は近細粒度並列処理を行うプロセッサエレメント数で分割し負荷のバランスを考慮して分割境界より前の一連の処理と後の一連の処理とで演算コストが均等になるように分割を行う。しかし、量子化 DCT 係数が 0 か非 0 かにより処理内容が異なるためタスク演算コストの推定が難しい。また、JPEG エンコーディングに用いられる入力画像は一般に自然画像を用いるため量子化 DCT 係数は DCT による空

間周波数変換により低周波数成分に非 0 係数が偏り高周波成分は 0 になる傾向があるため、複数の入力画像においてプロファイルを取り、その結果に基づき各タスクの演算コストを見積もった。

近細粒度タスクの定義後、近細粒度タスクはスケジューリングにより PE へ割り当てられる。このとき、データ転送オーバーヘッドを考慮し実行時間を最小化するヒューリスティックスケジューリングアルゴリズムである CP/DT/MISF 法、CP/ETF/MISF 法、ETF/CP 法および DT/CP 法の 4 手法を同時に適用し最良のスケジュールを選びスタティックに近細粒度タスクを割り当てる。

4. OSCAR 型シングルチップマルチプロセッサアーキテクチャ⁸⁾

ここでは、OSCAR 型シングルチップマルチプロセッサ (SCM) アーキテクチャおよびそのプロセッサコアアーキテクチャについて述べる。

4.1 メモリアーキテクチャ

OSCAR 型 SCM のネットワークおよびメモリアーキテクチャは、図 6 のように CPU、データ転送ユニット (DTU)、ローカルプログラムメモリ (LPM)、ローカルデータメモリ (LDM)、および分散共有メモリ (DSM) を持つプロセッサエレメント (PE) を相互接続網 (バス結合、クロスバ結合など) で接続し 1 チップ上に搭載したアーキテクチャである。今回の評価では、データ転送を CPU の処理とオーバーラップして行える DTU についてはオーバーラップデータ転送スケジューリングアルゴリズムの開発が終わっていないため利用していない。また、PE 間相互結合網は 3 本バスを利用している。

LPM は各々の CPU で実行するプログラムを格納し、1 クロックでアクセスできるものとする。同様に、LDM は PE 固有のデータを保持するために使用し、その容量は 1 チップあたりの総容量を 1 M バイトとする。たとえば、1 チップに 2 PE 搭載する SCM では、各 PE あたり 512 K バイトとして評価を行っている。また、LDM へのアクセスレイテンシは 1 クロックとする。DSM は、自 PE と他 PE の双方から同時にアクセス可能なマルチポートメモリであり、近細粒度タスク間のデータ転送などに使用する。DSM の容量は 1 PE あたり 16 K バイトとし、自 PE からのアクセスレイテンシは 1 クロック、他 PE からのアクセスレイテンシは 4 クロックとする。さらに、チップ外部には集中共有メモリ (CSM) が接続され、各 PE で共有されるデータを格納する。この CSM のアクセス

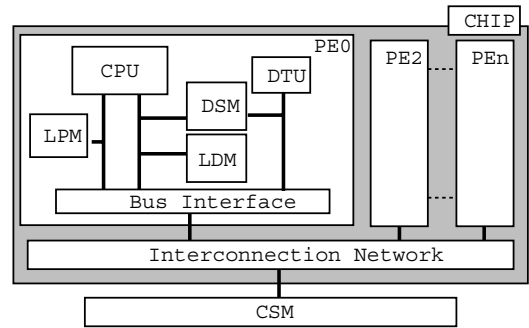


図 6 OSCAR 型 SCM アーキテクチャ
Fig. 6 OSCAR type SCM architecture.

表 1 プロセッサコア仕様
Table 1 Specifications of processor cores.

	Simple	4-issue
パイプライン段数	9	
同時命令発行数	1	4
IEU	1	2
FPU	1	2
LSU	1	1
命令発行タイプ	in-order	

レイテンシは 20 クロックとする。

OSCAR 型 SCM では、これら 4 種類のメモリに対し最適なデータ配置を行うことにより効率の良い並列処理を行うことができる。

4.2 プロセッサコアアーキテクチャ

各 PE が持つ CPU は、SPARC V9 規格に準拠したプロセッサである Sun Microsystems 社の UltraSPARC II¹²⁾ のパイプライン構成をベースとし、バリア同期機構など用の特殊レジスタや特殊レジスタを操作するための命令を付加したプロセッサである。

今回の評価で用いる OSCAR 型 SCM のプロセッサコアは、整数演算ユニット (IEU) を 1 本、ロードストアユニット (LSU) を 1 本、浮動小数点ユニット (FPU) を 1 本持つシングルイシューのシンプルな構成とした (表 1: Simple)。

また、スーパスカラとの比較のため 4 イシュー in-order 発行スーパスカラプロセッサについても評価を行う。このプロセッサは、IEU を 2 本、LSU を 1 本、FPU を 2 本持つ 4 イシュー in-order 発行構成で、動的スケジューリング用命令バッファエントリ数は 12 である (表 1: 4-issue)。なお、メモリアーキテクチャは OSCAR 型 SCM と同様である。

4.3 トランジスタ数の概算

ここでは、シングルイシュープロセッサを搭載した OSCAR 型 SCM の回路規模を既存のプロセッサを基に概算する。

表 2 プロセッサコアのトランジスタ数推定
Table 2 Number of transistors for processor cores.

チップ	microSPARC	UltraSPARC
総トランジスタ(万)	80	520
命令キャッシュ(KByte)	4	16
データキャッシュ(KByte)	2	16
ライン(Byte)	32(I)/16(D)	32
キャッシュのトランジスタ(万)	35	244
コアのトランジスタ(万)	45	276

表 3 各アーキテクチャのトランジスタ数推定
Table 3 Number of transistors for each architectures.

アーキテクチャ	SCM (4PE)	4-issue
コアのトランジスタ(万)	180	276
DSMのトランジスタ(万)	104	0
総トランジスタ(万)	284	276

プロセッサコアのようなランダムロジックが必要とするトランジスタ数は、論理回路の最適化の度合いや高速化のために費やす回路などにより大きく変化するが、ここではプロセッサコアに要するトランジスタ数を以下の式¹³⁾で概算値を求め推定する。ただし、ここでは簡単のためにI/Oドライバに関しては扱わない。

プロセッサコアのトランジスタ数

= チップの総トランジスタ数

- キャッシュのトランジスタ数

- I/Oドライバのトランジスタ数

キャッシュのトランジスタ数の概算は、

キャッシュのトランジスタ数

= data array + tag array + LRU array

+ data MUX + tag comparators

により算出する。ここで各arrayは、メモリアレイを示しSRAMセルであり1セル(=1bit)6トランジスタで構成される。データマルチプレクサ(data MUX)のトランジスタ数は $associativity \times (2 + associativity)$ 、tag comparatorsのトランジスタ数は $associativity \times (tag + status) \times 4$ で求められる¹³⁾。

参照するプロセッサとしては、シングルイシュープロセッサコアの推定にmicroSPARC¹⁴⁾を用い、また、今回比較に用いる4イシューin-orderスーパスカラプロセッサコアの推定にはUltraSPARC¹⁵⁾を用いる。各プロセッサの仕様、および算出したトランジスタ数を表2に示す。

以上のプロセッサコアの推定値をもとにプロセッサコアとDSMのトランジスタ数を合計し、SCMのトランジスタ数として表3に示す。LDMは各アーキテクチャで同容量としているのでここでは扱わない。また、バスはトライステートバッファのみ、調停回路はBus Interfaceのみで構成されるとし簡単のためにこ

こでのトランジスタ数見積りでは考慮しない。表より、シングルイシュープロセッサを4基搭載したOSCAR型SCMアーキテクチャと4イシューin-orderスーパスカラアーキテクチャは、ほぼ同程度のトランジスタ数であると推定できる。

5. 性能評価

ここでは、JPEGエンコーディングプログラムにマルチグレイン並列処理を適用しOSCAR型メモリアーキテクチャシングルチップマルチプロセッサ(SCM)上で評価した結果について述べる。なお、性能評価にはクロックレベルシミュレータを用いた。

評価に用いるプログラムは、プログラムソースレベルで3章で提案したマルチグレイン並列性抽出のためのループアンローリング、粗粒度タスク融合、エンタロピーコーディングのプログラムリストラクチャリングを適用し、タスク粒度を3章で述べた形に成形したプログラムソースをOSCARマルチグレイン自動並列化コンパイラを用いてマルチグレイン並列性の抽出およびタスクスケジューリングを行いSCM用バイナリを作成した。なお、コンパイラにより生成されたタスクスケジューリング結果は本論文3章で提案した結果と同一である。

入力画像は、MediaBenchで用いられる入力画像(testimg.ppm:自然画像)をシミュレーション時間短縮のために32×32pixelに縮小した画像とし、CSM上に配置した。シングルチップマルチプロセッサでのデータ配置は、プログラム開始後、CSM上に置いている入力画像データを各PE内のLDMへコピーし、エンコード処理は各PE内のLDMで行い、データ転送はDSMを用いたPE間データ転送により行われる。処理完了後、処理結果はCSMへ格納する。4イ

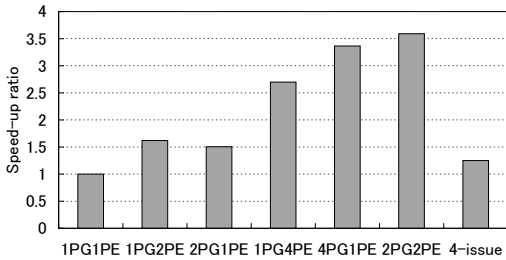


図 7 JPEG エンコーディング評価結果

Fig. 7 Evaluation result.

シュースーパスカラプロセッサの場合も同様に、プログラム開始後 LDM へ入力画像をコピーし、LDM を用いてエンコード処理を行い、処理結果を CSM へ格納している。

OSCAR 型 SCM 上での評価結果を逐次実行時間に対する速度向上率として図 7 に示す。図中横軸の $mPGnPE$ は n 個のプロセッサエレメント (PE) をグループとした m 個のプロセッサグループ (PG) による処理を表し、SCM 中の総 PE 数は $m \times n$ であることを示す。この $mPGnPE$ という SCM 構成は、コンパイラから見たタスク割当て単位の仮想的な構成であり、PG 間では粗粒度タスク並列処理、PG 内 PE 間では近細粒度並列処理を行うことを前提とし、ハードウェアとしては同一である。次に図中の 4-issue は、4 イシュースーパーカプラプロセッサによる速度向上率である。ただし、ここでの 4 イシュースーパーカプラプロセッサの評価に関しては、OSCAR コンパイラがスーパーカプラプロセッサの性能を最大限に引き出すための最適化を行っていないため参考値として示す。

まず、4 イシュースーパーカプラプロセッサを用いた場合、シングルイシュープロセッサの逐次実行時間に対して約 1.25 倍の速度向上率を得た。ほぼ同等のトランジスタ数であるシングルイシュープロセッサコアを 4 基使用した SCM での 2PG2PE 実行は、逐次実行時間に対して約 3.59 倍の速度向上率が得られ、4 イシュースーパーカプラプロセッサに対しても約 2.87 倍の性能向上が得られたことが分かる。これは、JPEG エンコーディング処理は積和演算を多用するため乗算などのマルチサイクル命令によるパイプラインストールが発生し 4 イシュースーパーカプラプロセッサでは十分な性能が出せないがシングルイシュープロセッサコアを複数搭載した SCM ではプロセッサ間負荷分散により効果的な並列処理が行えるためである。

次に、OSCAR 型 SCM アーキテクチャでの粗粒度並列処理単体性能と近細粒度並列処理単体性能、およ

び粗粒度並列処理と近細粒度並列処理を階層的に組み合わせたマルチグレイン並列処理での性能を比較する。図 7 に示すように、総 PE 数が 2 のときでは、逐次実行時に対し粗粒度並列処理を用いた 2PG1PE では約 1.50 倍、近細粒度並列処理を用いた 1PG2PE では約 1.62 倍の速度向上率が得られた。同様に総 PE 数が 4 のとき、 8×8 pixel ブロック間の粗粒度並列性のみを用いた 4PG1PE では約 3.36 倍、 8×8 pixel ブロック内の近細粒度並列性のみを用いた 1PG4PE では約 2.70 倍、PG 間で 8×8 ブロック間の並列性を用いた粗粒度並列処理を行い PG 内 PE で 8×8 ブロック内の並列性を用いた近細粒度並列処理を行うマルチグレイン並列処理を行う 2PG2PE では約 3.59 倍の速度向上率が得られた。

以上より使用プロセッサ数が同じときには、階層的なマルチグレイン並列処理を使用した場合の方が高い実行効率を与えることが確かめられた。この理由を以下に述べる。まず、近細粒度並列処理のみを用いた場合は、使用プロセッサ数の増加により PE 間データ転送オーバーヘッドおよび同期オーバーヘッドが増加したため処理性能の低下が見られた。粗粒度並列処理のみの場合は、エントロピー符号化処理では、処理を行う量子化 DCT 係数の値が 0 か非 0 かにより実行時間が変化するので、粗粒度タスク MT_{ai} は入力データによりタスク実行時間が異なるため粗粒度並列処理のみを利用した場合、各タスク処理時間のばらつきが生じ各プロセッサに負荷のアンバランスが生じた (図 8 (a))。しかし、粗粒度並列処理と近細粒度並列処理を階層的に用いたマルチグレイン並列処理では、近細粒度並列処理により 1 つの MT が複数のプロセッサ上で処理されるため負荷バランスが向上し、より高い実行効率を得られたのである (図 8 (b))。

6. ま と め

本論文では、JPEG エンコーディング処理における、従来より行われていた命令レベル並列性の利用とは異なるアプローチの 8×8 ブロック間の粗粒度並列性と 8×8 ブロック内の近細粒度並列性を階層的に用いるシングルチップマルチプロセッサ用マルチグレイン並列処理手法を提案しその性能評価を行った。その結果シングルイシュープロセッサコアを 4 基搭載した OSCAR 型 SCM 上では、逐次実行時間に対し約 3.59 倍の速度向上率が得られた。以上よりシンプルなプロセッサコアを集積した OSCAR 型 SCM 上での JPEG エンコーディングのマルチグレイン並列処理は集積度向上に対しスケーラブルな性能向上を可能とするとい

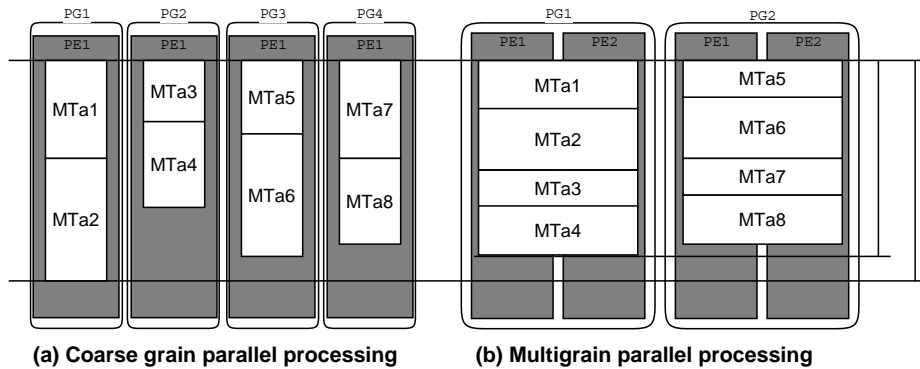


図 8 JPEG エンコーディング実行時間イメージ

Fig. 8 Execution time of JPEG encoding.

うことが確認できた。

今後の課題として、MPEG-2 や JPEG2000 などのマルチメディアアプリケーションを含めたアプリケーションのマルチグレイン並列化手法の検討およびマルチグレイン並列処理をサポートするシングルチップマルチプロセッサアーキテクチャの開発があげられる。

謝辞 本研究の一部は、STARC「自動並列化コンパイラ協調型シングルチップマルチプロセッサの研究」および経済産業省/NEDO ミレニアムプロジェクト「アドバンス並列化コンパイラ」により行われた。本論文作成にあたり、有益なコメントをいただいた富士通高橋宏政氏、STARC 平田雅規氏、宮田操氏、東芝浅野滋徳氏、安川英樹氏、ソニー倉田隆弘氏、松下高山秀一氏に感謝いたします。

参考文献

- 1) Raman, S.K., Pentkovki, V. and Keshava, J.: Implementing Streaming SIMD Extensions on the Pentium III Processor, *IEEE MICRO*, Vol.20, No.4 (2000).
- 2) Arakawa, F., Nishi, O., Uchiyama, K. and Nakagawa, N.: SH4 RISC Multimedia Microprocessor, *IEEE MICRO*, Vol.18, No.2 (1998).
- 3) Suga, A. and Matsunami, K.: Introducing the FR500 Embedded Microprocessor, *IEEE MICRO*, Vol.20, No.4 (2000).
- 4) Edahiro, M., Matsushita, S., Yamashita, M. and Nishi, N.: A Single-Chip Multiprocessor for Smart Terminals, *IEEE MICRO*, Vol.20, No.4 (2000).
- 5) Hammond, L., Hubbert, B., Siu, M., Prabhu, M.K., Chen, M. and Olukotun, K.: The Stanford HYDRA CMP, *IEEE MICRO*, Vol.19, No.2 (1999).
- 6) Iwata, E. and Olukotun, K.: Exploiting Coarse-Grain Parallelism in the MPEG-2 Algorithm, CSL-TR-98-771, Stanford University Computer System Lab. (1998).
- 7) Kasahara, H., Obata, M. and Ishizaka, K.: Automatic Coarse Grain Task Parallel Processing on SMP using OpenMP, *Proc. 12th Workshop on Languages and Compilers for Parallel Computing* (2000).
- 8) 木村啓二, 加藤孝幸, 笠原博徳: 近細粒度並列処理用シングルチップマルチプロセッサにおけるプロセッサコアの評価, 情報処理学会論文誌, Vol.42, No.4 (2001).
- 9) Lee, C., Potkonjak, M. and Mangione-Smith, W.H.: MediaBench: A Tool for Evaluating and Synthesizing Multimedia and Communications Systems, *30th International Symposium on Microarchitecture (MICRO-30)* (1997).
- 10) Hamilton, E.: *JPEG File Interchange Format Version 1.02* (1992).
- 11) Aldus Developers Desk: *TIFFTM Revision 6.0* (1992).
- 12) Sun Microelectronics: *UltraSPARCTM User's Manual* (1997).
- 13) Geuskens, B. and Rose, K.: *MODELING MICROPROCESSOR PERFORMANCE* (1998).
- 14) Texas Instruments: *TMS390S10 Microprocessor* (2000).
- 15) Lev, L.A., et al.: A 64-b microprocessor with multimedia support, *IEEE Journal of SOLID-STATE CIRCUITS*, Vol.30, No.11 (1995).

(平成 14 年 1 月 25 日受付)

(平成 14 年 5 月 15 日採録)



小高 剛(学生会員)

昭和52年生。平成11年早稲田大学理工学部電気電子情報工学科卒業。同年(株)ユニシアジェックス入社。平成12年同社退社。同年早稲田大学大学院科目等履修生。平成13年同大学院修士課程入学。平成14年同大学院修士課程修了。同年同大学院博士課程進学。現在に至る。



木村 啓二(正会員)

昭和47年生。平成8年早稲田大学理工学部電気工学科卒業。平成13年同大学大学院理工学研究科電気工学専攻博士課程修了。博士(工学)。平成11年同大学同学部助手。平成14年同大学理工学総合研究センター客員講師。現在に至る。マルチグレイン並列処理用プロセッサアーキテクチャに関する研究に従事。



内田 貴之

昭和52年生。平成12年早稲田大学理工学部電気電子情報工学科卒業。平成14年同大学大学院修士課程修了。同年三菱電機(株)入社。現在に至る。



笠原 博徳(正会員)

昭和32年生。昭和55年早稲田大学理工学部電気工学科卒業。昭和60年同大学大学院博士課程修了。工学博士。昭和58年同大学同学部助手。昭和60年学術振興会特別研究員。昭和61年早稲田大学理工学部電気工学科専任講師。昭和63年同助教授。平成9年同大学電気電子情報工学科教授。現在に至る。平成元年~2年イリノイ大学 Center for Supercomputing Research & Development 客員研究員。昭和62年 IFAC World Congress 第1回 Young Author Prize。平成9年度情報処理学会坂井記念特別賞受賞。著書「並列処理技術」(コロナ社)。電子情報通信学会、IEEE等の会員。