

## 消費電力を 50%削減する動的電圧/周波数制御型 H.264/AVC HDTV デコーダアーキテクチャ

川上 健太郎<sup>†</sup> 竹村 淳<sup>‡</sup> 黒田 光彦<sup>†</sup> 川口 博<sup>¶</sup> 吉本雅彦<sup>¶</sup>

<sup>†</sup> 神戸大学大学院自然科学研究科 〒657-8501 兵庫県神戸市灘区六甲台町 1-1

<sup>‡</sup> 金沢大学大学院自然科学研究科 〒920-1192 石川県金沢市角間町

<sup>¶</sup> 神戸大学工学部情報知能工学科 〒657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: <sup>†</sup> <sup>‡</sup> <sup>¶</sup> {kawakami@cs28., takemura@cs28., mitsuhiro@cs28., kawapy@, yosimoto@}.cs.kobe-u.ac.jp

あらまし 専用 LSI 回路に対して動的電圧スケールリングを適用可能にするエラスティックパイプラインアーキテクチャを提案する。提案アーキテクチャの適用例として H.264/AVC HDTV デコーダへの適用を検討した。提案アーキテクチャは H.264 HDTV デコードの処理サイクル数を最大 48%削減することができる。削減されたサイクルを利用して DVS を適用した場合、90nm プロセスにおいて H.264/AVC HDTV デコーダの消費電力をクロックゲーティング手法と比較して最大 50%削減できることが見積もられた。

キーワード H.264/AVC, DVS (dynamic voltage scaling), エラスティックパイプライン, 低消費電力

## Dynamic Voltage Scaling in an Elastic Pipeline and Its Application to an H.264/AVC HDTV Video Decoder LSI

Kentaro KAWAKAMI<sup>†</sup> Jun TAKEMURA<sup>‡</sup> Mitsuhiro KURODA  
Hiroshi KAWAGUCHI and Masahiko YOSHIMOTO<sup>¶</sup>

<sup>†</sup> Graduate School of Science and Technology, Kobe University 1-1 Rokkodai-Cho, Nada-ku, Kobe, 657-8501 Japan

<sup>‡</sup> Graduate School of Natural Science & Technology, Kanazawa University

Kakuma-machi, Kanazawa, Ishikawa, 920-1192 Japan

<sup>¶</sup> Computer and Systems Department, Kobe University 1-1 Rokkodai-Cho, Nada-ku, Kobe, 657-8501 Japan

E-mail: <sup>†</sup> <sup>‡</sup> <sup>¶</sup> {kawakami@cs28., takemura@cs28., mitsuhiro@cs28., kawapy@, yosimoto@}.cs.kobe-u.ac.jp

**Abstract** We propose an elastic pipeline that can apply dynamic voltage scaling (DVS) to hardwired logic circuits. In order to demonstrate its feasibility, a hardwired H.264/AVC HDTV decoder is designed as a real-time application. The designed decoder reduces 48% of execution cycles for H.264 HDTV decoding. In case of DVS is applied with this reduced cycles, the proposed decoder achieves a power reduction of 50% in a 90-nm process technology, compared to the conventional clock-gating scheme.

**Keyword** H.264/AVC, DVS (dynamic voltage scaling), elastic pipeline, low power

### 1. まえがき

DVS は汎用プロセッサにおいて、高いピーク性能を維持しながら平均消費電力を削減することを可能とする手法である [1]。CMOS デジタル回路の動作周波数  $f$  は式(1)[2]で表され、電源電圧  $V_{dd}$  が高いほど動作周波数は高くなる。

$$f = k \frac{(V_{dd} - V_{th})^\alpha}{V_{dd}} \quad (1)$$

ただし、 $\alpha$  は短チャネル長 MOSFET の velocity

saturation index であり、本原稿の検討で用いた 90nm プロセスでは 1.6、 $V_{th}$  は MOSFET の閾値電圧である。一方、LSI の消費電力を  $P$  とおくと、 $P$  は式(2)で表され、電源電圧の 2 乗に比例する。

$$P = a \cdot f \cdot V_{dd}^2 \quad (2)$$

図 1 は LSI の動作周波数と消費電力の関係を表している。DVS では LSI を最大動作周波数で動かさないとときは LSI の電源電圧を下げるため、消費電力を大きく削減することができる。DVS では高いピーク性能が必要とされるときは最大の動作周波数と電源電圧で LSI を動作させることで対応し、平均的な性能が必要など

<sup>†</sup> 現在、株式会社 ルネサス テクノロジ。

きには動作周波数と電源電圧を最大値から下げること  
で消費電力を低減する。

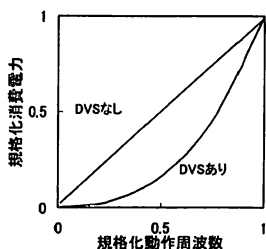


図1 動作周波数と消費電力の関係

式(1)および(2)の関係式は、CMOS デジタル回路一般  
に対して成立するため、DVS による消費電力削減は汎  
用プロセッサに限らず、特定の処理に専用化して設計  
される LSI に対しても有効である。しかし、特定の処  
理に専用化して設計される LSI は処理のスループット  
向上のためにパイプラインアーキテクチャを採用する  
ことが多いが、従来のパイプラインアーキテクチャで  
は DVS の適用は困難になる。従来のパイプラインア  
ーキテクチャでは、パイプライン処理の開始はあらかじ  
め決められた固定のサイクルに同期したサイクルとす  
るため、処理に必要なサイクル数は何回パイプライン  
処理を実行するか、言い換えると、何個のデータを  
パイプラインで処理するかによってのみ決定され、各デ  
ータが処理にどれだけの演算負荷を必要とするかには  
依存しなかった。したがって、汎用プロセッサ上での  
ソフトウェア処理の場合のような必要な処理負荷に依  
存した処理サイクル数の変動が発生しないために、動  
作周波数を下げることができず、DVS を適用するこ  
とができなかった。

本稿では、パイプラインアーキテクチャを採用した  
専用 LSI に対して DVS を適用可能にするエラスティッ  
クパイプラインアーキテクチャを提案する。エラステ  
ィックパイプラインアーキテクチャでは、入力される  
データが処理に要求する演算負荷に応じてパイプライン  
処理に必要なサイクル数を削減することができ、削  
減されたサイクルによって得られる時間余裕を用いて  
動作周波数を削減することができるようになるため  
DVS が適用可能になる。本稿ではエラスティックパイ  
プラインアーキテクチャを H.264/AVC HDTV デコーダ  
に適用し、消費電力削減効果を検討する。

## 2. エラスティックパイプラインアーキテクチャ の構成

### 2.1. 従来のパイプラインアーキテクチャ

図 2 に従来のパイプラインアーキテクチャのタイミ  
ング図を示す。Worst-case execution cycles (WCEC) は 1  
回のパイプライン処理に必要なサイクル数の最大値を  
表している。図中のハッチングされた領域は各ステー

ジが処理を実行中であるサイクルを、それ以外の領域  
はアイドル中であるサイクルを示している。従来のパ  
イプラインアーキテクチャでは、各パイプライン処理  
の実行は WCEC に同期したサイクルで開始されるため、  
斜線の領域で示されたすべての演算器がアイドル状態  
であるサイクルが存在する。

$N$  個のデータを  $M$  段で構成されるパイプラインで処  
理される場合に必要なサイクル数を  $EC_{conv}$  とすると、  
 $EC_{conv}$  は式(3)で求まる。ここで、 $W_{M,N} (\leq WCEC)$  は  $N$   
番目の入力データが  $M$  段目のパイプラインステージで  
の処理に必要なサイクル数である。 $N$  が充分大きい場  
合、 $EC_{conv} \approx N \times WCEC$  となり、処理に必要なサイ  
クル数は入力データの個数  $N$  のみに依存し、処理負荷  
には依存しない。したがって動作周波数を下げて DVS  
を適用する余地がないことになる。

$$EC_{conv} = (M + N - 2) \times WCEC + W_{M,N} \quad (3)$$

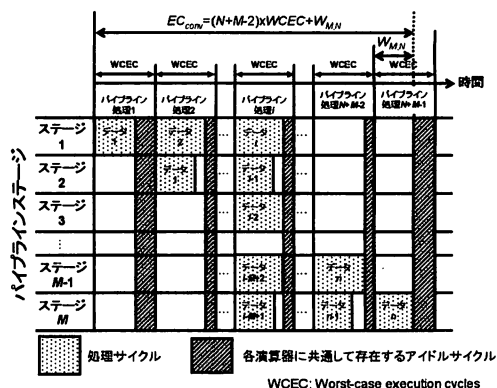


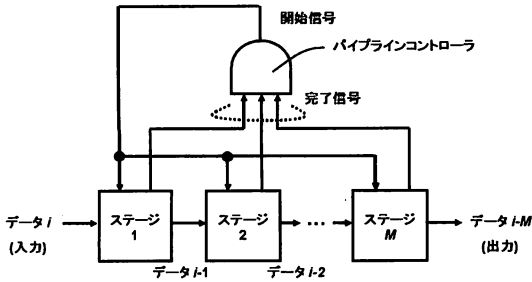
図 2 従来のパイプラインアーキテクチャの  
タイミング図

### 2.2. エラスティックパイプラインアーキテクチャ

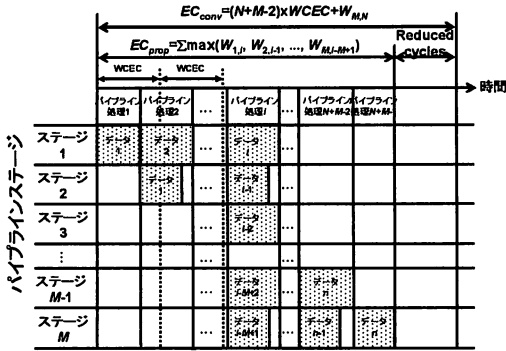
図 3(a)および(b)エラスティックパイプラインア  
ーキテクチャの概念図とタイミング図を示す。エラステ  
ィックパイプラインアーキテクチャでは各ステージは自  
身の処理が完了した後、処理完了信号をパイプライン  
コントローラへ送信する。パイプラインコントローラ  
はすべてのステージから処理完了信号を受け取った後、  
すべてのステージに対して処理開始信号を送信する。  
すべてのステージは処理開始信号の受信を待つて、一  
斉に次のパイプライン処理を開始する。このように、  
提案パイプラインでのパイプライン処理開始サイクル  
は、従来パイプラインのような WCEC に同期したサイ  
クルではなく、すべてのステージにおいて一つ前のパ  
イプライン処理が完了次第開始される。図 3(b)に示す  
ように、すべてのステージがアイドルであるサイ  
クルを省略することで、1 回のパイプライン処理に必要  
なサイクルを WCEC から削減している。

提案パイプラインアーキテクチャで  $N$  個の入力データを  $M$  段のステージで構成されるパイプラインで処理するために必要なサイクル数を  $EC_{prop}$  とすると、 $EC_{prop}$  は式(4)で求められる。ここで、 $W_{p,q}$  は  $q$  番目のデータが  $p$  番目のステージでの処理に必要なサイクル数を表し、 $q < 1$  または  $q > N$  の場合、 $W_{p,q} = 0$  である。

$$EC_{prop} = \sum_{i=1}^{M+N-1} \max(W_{1,i}, W_{2,i-1}, \dots, W_{M,i-M+1}) \quad (4)$$



(a)



(b)

図3 提案するエラスティックパイプラインアーキテクチャ、(a)概念図、(b)タイミング図

### 3. H.264/AVC デコード処理における処理サイクル数変動の要因

図4にH.264/AVCデコード[3]処理のブロック図を示す。各処理に必要なサイクル数は3.1から3.5に示す理由により変動する。

#### 3.1. エントロピーデコード処理

エントロピーデコードは入力されたビットストリームを復号化し、後段の処理に必要なMBタイプ、イントラ予測モード、動きベクトル、コーデッドブロックパターン、有効係数などの種々のシンタックスエレメントを生成する。1つのMBに含まれるシンタックスエレメントの数はMBごとに異なる。情報量の多いMBであれば100以上のシンタックスエレメントを含

み、情報量の少ないスキップマクロブロックであれば1つしか含まない。1つのMBのエントロピーデコード処理に必要なサイクル数は、対象MBがいくつのシンタックスエレメントを含むか、それぞれのシンタックスエレメントのデコードに何ビット分のビットストリームを必要とするかなどによって変動するため、MBごとに処理サイクル数が変動する。

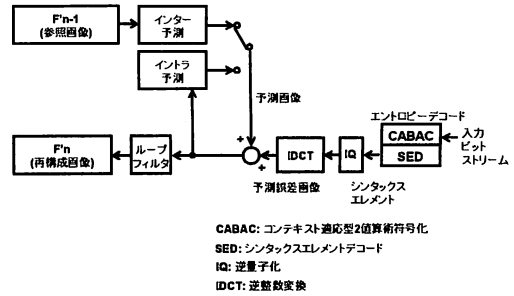


図4 H.264 デコード処理のブロック図

#### 3.2. IQ/IDCT 処理

エントロピーデコーダによって生成された有効係数は逆量子化(IQ)処理、逆整数変換(IDCT)処理を経て予測誤差画像に変換される。IQ/IDCT処理は $4 \times 4$ 画素で構成されるブロック単位に行列の演算で定義されており、有効係数は $4 \times 4$ 行列の形に並べ替えられ演算される。 $4 \times 4$ 行列が0行列である場合、IQ/IDCT処理後の行列も0行列となるのが分かっているため、演算を省略することができる。1つのMBは24個の $4 \times 4$ 行列を含むが、1つのMBのIQ/IDCT処理に必要なサイクル数は、対象MBが何個の0でない行列を含むかに依存する。

#### 3.3. イントラ予測処理

MBがイントラMBとして符号化されている場合、すでに復号化処理を完了している隣接するMBの画素値を用いて、対象MBのイントラ予測画像が生成される。H.264では13種類のイントラ予測モードが規定され、イントラMBにはそれぞれ1つの予測モードが適用される。予測画素生成に必要な演算量は予測モード毎に異なる。例えばIntra\_4x4\_Horizontal予測モードやIntra\_4x4\_Vertical予測モードの場合、予測画素は隣接するMBの画素値をコピーするだけでよいが、Intra\_4x4\_DC予測モードの場合、8タップのフィルタ処理によって計算しなければならない。したがって、イントラ予測処理に必要なサイクル数は対象MBのイントラ予測モードに依存する。

#### 3.4. インター予測処理

MBがインターMBとして符号化されている場合、動きベクトルと参照ピクチャからインター予測画像が生成される。H.264では動きベクトルの精度として整数

画素精度，1/2 画素精度，1/4 画素精度が規定されている。対象 MB の動きベクトルが整数画素精度である場合，参照ピクチャから動きベクトルで示された座標の画像が切り出され，その画像が予測画像とされる。動きベクトルが 1/2，1/4 画素精度の場合，動きベクトルで示された座標の画像にそれぞれ 6 タップ，8 タップフィルタ処理を行って予測画像を生成する。したがって，インター予測処理に必要なサイクル数は対象 MB の動きベクトルの精度に依存する。

### 3.5. ループフィルタ処理

予測画像と予測誤差画像の加算によって生成された再構成画像は，ループフィルタ処理が施されて最終的な復号画像となる。ループフィルタ処理では MB と MB の境界の画素や，MB に含まれるブロックとブロックの境界の画素に対してスムージング処理を行う。1 つの MB は 48 個のブロック境界を含むが，それぞれの境界に対してループフィルタ処理を行うか行わないかは，MB タイプや境界の画素値に応じて適応的に判定される。ループフィルタ処理に必要なサイクル数は，対象 MB の 48 個のブロック境界のうち，何個の境界に対してスムージング処理されるかに依存する。

## 4. 提案パイプラインアーキテクチャによる消費電力削減効果の見積もり

3 節で述べたように，H.264/AVC デコード処理において必要な各要素処理は入力データの特徴によって処理に必要なサイクル数が変動するため，エラスティックパイプラインアーキテクチャを適用することでパイプライン処理に必要なサイクル数を削減できると考えられる。本節では H.264/AVC HDTV デコーダ LSI に適用した場合のサイクル数の削減効果と，DVS を適用した場合の消費電力削減効果を見積もった。

### 4.1. 提案デコーダアーキテクチャ

提案する H.264/AVC HDTV デコーダのアーキテクチャのブロック図を図 5 に示す。HDTV 解像度対応のデコーダの場合，デコードピクチャバッファとして 96Mbit のメモリが必要であるが，これを外部 DRAM 上に確保する。

処理モジュール間にはバッファ RAM を配置している。ダブルバッファリングのためにバッファ RAM は 2 バンク構成としている。例えば IQ/IDCT 処理モジュールが予測誤差画像をバンク 0 に書き込み動作をしているとき，同時に予測誤差加算器はバンク 1 から読み出し動作を行う。イントラ予測処理モジュールから読み出される RAM はアクセスがビジーでないので，この RAM のみ 1 バンク構成としている。

エラスティックパイプラインで構成されたデコーダコアに DVS を適用する場合でも，DRAM インターフェース(I/F)は一定の動作周波数と電源電圧で動作さ

せるのが望ましい。なぜなら，DRAM I/F はプレ/ポストビデオ処理などのデコード処理以外の処理を実行する他のコアと共有するリソースであるため，デコーダコアに合わせて DRAM I/F の電源電圧や動作周波数を変えた場合，他のコアでタイミング上の問題やインターフェースの問題が発生することが懸念されるためである。

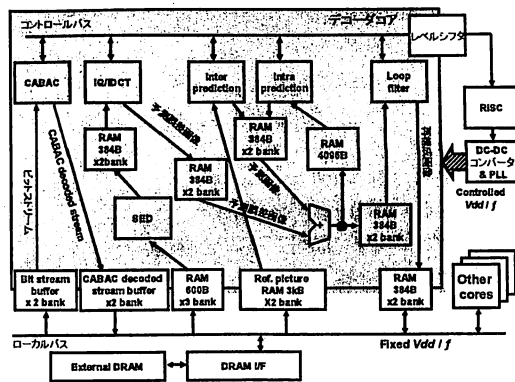


図 5 提案する H.264 デコーダアーキテクチャのブロック図

せる場合，デコーダコアの動作周波数と電源電圧が低い値に制御されているときにミスマッチが起こる。このミスマッチを解消するためにローカルバスとデコーダの間には図 6 に示すマルチバンク構成の SRAM を配置する。デコーダコアからアクセスされるバンクは，デコーダコアと同じクロックと電源電圧が供給される。DRAM I/F とアクセスするバンクは DRAM I/F と同じクロックと電源電圧が供給される。例えばインター予測処理モジュールがバンク 0 の SRAM に格納されている参照画像を読み出すとき，バンク 0 のクロックと電源電圧はデコーダコアと同じクロックと電源電圧に制御される。このときバンク 1 には次の MB のインター予測処理に必要な参照画像が DRAM I/F を介して書き込まれるが，このときバンク 1 のクロックと電源電圧は DRAM I/F のクロックと電源電圧に制御される。インター予測処理モジュールが次の MB の処理を行うときはバンク 0 とバンク 1 の役割が入れ替わり，バンク 1 はデコーダコアと同じクロックと電源電圧に，バンク 0 は DRAM I/F と同じクロックと電源電圧に制御される。

### 4.2. エラスティックパイプラインによる削減サイクル数

エラスティックパイプラインによって削減されるサイクル数を見積もった。Celoxica Handel-C [4]を用いて各処理モジュールを設計し，ビットレートを 10Mbps に設定して H.264/AVC 参照ソフトウェア JM9.6[5]を用いて生成したビットストリームをデコード処理するた

めに必要なサイクル数をシミュレーションした。

図9に1フレームの処理に必要なサイクル数を示す。横軸はフレーム番号、縦軸はすべてのパイプライン処理がWCECサイクル必要な場合を1として正規化したサイクル数を示す。1フレームの処理に必要なサイクル数は平均してWCECの場合の50%程度で済むことが分かる。すなわち、実際の復号化処理では動作周波数を50%まで下げても所定の時間内に復号化処理を完了できることを表している。

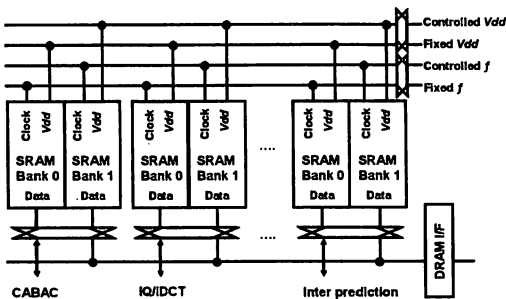


図6 デコーダコアとローカルバス間のインターフェース用SRAM

### 4.3. 動作周波数制御アルゴリズム

DVS環境下で処理のリアルタイム性を保証しながら低い動作周波数での動作時間を設定するアルゴリズムとしてフィードバック型の動作周波数制御アルゴリズムが提案されている[6]。フィードバック型制御アルゴリズムでは1フレームをスロットと呼ぶ処理単位に分割し、動作周波数はスロット単位に設定される。1スロットは決められた数のMBから構成される。

図10にフィードバック型制御アルゴリズムによって3番目のスロットの動作周波数が決定される方法を示す。エラスティックパイプラインにより1番目と2番目のスロットの処理に必要なサイクル数は、すべてのパイプライン処理がWCECサイクルかかる場合と比較して削減される。したがって、削減されたサイクル数に相当する時間 $\Delta H$ だけ3番目のスロットの処理開始時刻が前倒しされる。スロット3の処理には、もともとスロット3の処理に割り当てられていた時間 $T_{slot}$ に加えて $\Delta H$ の時間を費やすことができる。電源電圧と動作周波数を切り替える場合、切り替えに必要な待ち時間 $T_{id}$ が必要になるが、スロット3に含まれるすべてのMBのパイプライン処理に必要なサイクル数がWCECである場合を考慮しても、 $f/2$ まで動作周波数を下げてもスロット3が本来処理を完了しなければならない時刻 $3T_{slot}$ の前にスロット3の処理完了を保證できることが分かる。 $f/4$ まで動作周波数を下げた場合、時刻 $3T_{slot}$ までに処理完了は保證できない。この場合、フィードバック型制御アルゴリズムはスロット3の動

作周波数として $f/2$ を選択する。このように、フィードバック型制御アルゴリズムはスロット毎に動作周波数を決定する。

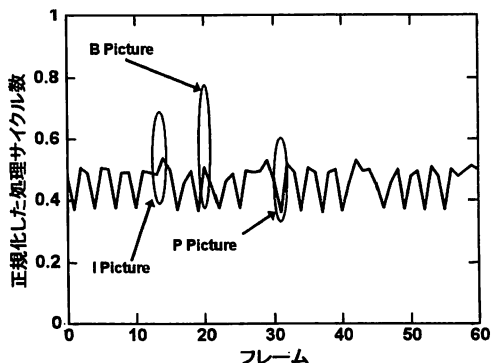


図9 デコーダコアとローカルバス間のインターフェース用SRAM

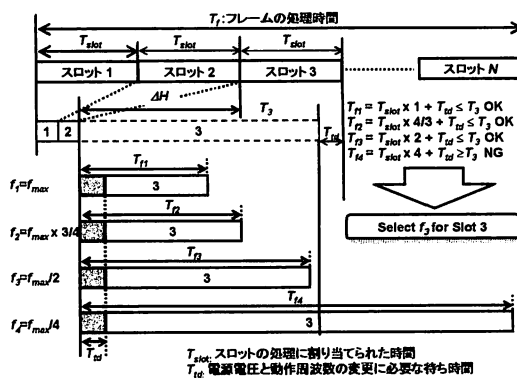


図10 フィードバック型制御アルゴリズム

スロット数は小さすぎても大きすぎても十分な低消費電力効果を得ることができない。スロット数が小さすぎる場合、適切な時刻で動作周波数を下げる機会を得られず、高い動作周波数のまま処理が進んでしまい低電力化されない。スロット数が大きすぎる場合、頻りに動作周波数の上げ下げが発生するが、電源電圧と動作周波数の切り替えには有限の待ち時間が必要となるため[1]、これらの頻繁な変更により多くの時間が必要となる。エラスティックパイプラインによって得られた処理時間の余裕は、この待ち時間によって消費されてしまうため、処理時間の余裕が減り、結果的に低い動作周波数での動作時間が少なくなってしまう低電力化が妨げられる。

### 4.4. 消費電力削減効果の見積もり

エラスティックパイプラインアーキテクチャにDVSを適用した場合の消費電力削減効果を見積もった。各演算器の動作周波数、消費電力は、①32ビットALUの電源電圧、動作周波数、消費電力の関係をSPICE

シミュレーションによって求め、各演算器と 32 ビット ALU のクリティカルパスのゲート段数比とゲート数比から動作周波数と消費電力を計算した。SPICE のモデルファイルとして 90nm ジェネリックプロセスを用いた。

図 11 に消費電力削減効果のスロット数依存性を示す。電源電圧と動作周波数の組み合わせは  $f_{max}=108\text{MHz}@1.0\text{V}$ ,  $f_{max}/2=54\text{MHz}@0.7\text{V}$  の 2 種類を用意し、電源電圧と動作周波数の切り替えには 50us[1] の待ち時間を仮定した。消費電力削減効果は従来のパイプラインアーキテクチャでクロックゲーティングした場合の消費電力で規格化した提案アーキテクチャの消費電力を表している。スロット数が小さすぎても大きすぎても消費電力は削減されず、1 フレームを 60 スロットとしたときに消費電力削減効果が極値を持つことが分かる。

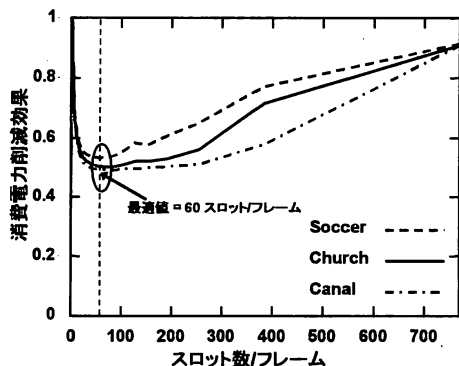


図 11 消費電力削減効果のスロット数依存性

図 12 にスロット数を 60 とした場合の 7 種類のシーケンスの消費電力削減効果を示す。動作周波数は  $f_{max}$ ,  $f_{max}/2$  の他に  $3f_{max}/4$ ,  $f_{max}/4$  が用意された場合について示している。フレームの処理に必要なサイクル数が前もって分かる場合、DVS 環境下で消費電力を最小とする動作周波数と動作時間の決定方法は一意に求まるが [7]、この場合の理論的な最小消費電力も合わせて示した。理論的な最小消費電力にどれだけ近づけるかは、動作周波数制御アルゴリズムのよしあしの目安となる。電源電圧と動作周波数を 4 組用意した場合、最も消費電力が削減されるのはシーケンス "Intersection" で 50% の削減が見積もられた。消費電力削減効果の平均値は動作周波数が 2 組と 4 組の場合でそれぞれ 45% と 49% と見積もられた。動作周波数は 2 組用意すれば、4 組用意した場合の理論的な最小電力の場合と比較して、最も消費電力削減効果が劣化する場合のシーケンス "Whale" で 7.8% と見積もられた。

## 5. まとめ

特定の処理に専用化されて設計されたパイプライン

アーキテクチャに DVS を適用可能とするエラスティックパイプラインアーキテクチャを提案した。エラスティックパイプラインアーキテクチャは、処理対象として入力されるデータの特徴に応じて処理に必要なサイクルが変動する場合に、処理に必要なサイクル数を削減することができる。削減された処理サイクル数により処理時間の余裕が発生するため、この時間余裕を利用して DVS による低消費電力化が可能になる。

H.264 HDTV デコード LSI を例にエラスティックパイプラインアーキテクチャに DVS を適用した場合の消費電力削減効果を見積もった。最も消費電力が削減されるシーケンスの場合、従来のパイプラインアーキテクチャと比較して 50% の低消費電力効果が見積もられた。

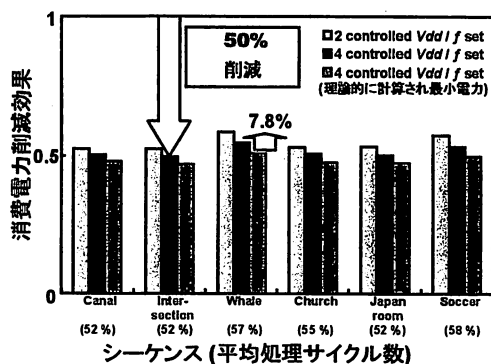


図 12 消費電力削減効果

## 文献

- [1] K. J. Nowka, G. D. Carpenter, E. W. MacDonald, H. C. Ngo, B. C. Brock, K. I. Ishii, T. Y. Nguyen, and J. L. Burns, "A 32-bit PowerPC system-on-a-chip with support for dynamic voltage scaling and dynamic frequency scaling," IEEE J. Solid-State Circuits, vol.37, no.11, pp.1441-1447, Nov. 2002.
- [2] T. Sakurai, and A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," IEEE J. Solid-State Circuits, vol.25, no.2, pp.584-594, Apr. 1990.
- [3] Joint Video Team (JVT) of ISO/IEC MPEG&ITU-T VCEG, "ISO/IEC 14496-10," May, 2003.
- [4] <http://www.celoxica.com/>
- [5] H.264/AVC reference software, <http://iphome.hhi.de/suehring/tml/>
- [6] H. Kawaguchi, Y. Shin, and T. Sakurai, "μITRON-LP: power-conscious real-time OS based on cooperative voltage scaling for multimedia applications," IEEE Trans. Multimedia, vol.7, no.1, pp.67-74, Feb. 2005.
- [7] K. Kawakami, M. Kanamori, Y. Morita, J. Takemura, M. Miyama, and M. Yoshimoto, "Power-minimum frequency/voltage cooperative management method for VLSI processor in leakage-dominant technology era," IEICE Trans. Fundamentals, Vol. E88-A, No. 12, pp.3290-3297 Dec. 2005.