

SPARCアーキテクチャに基づく
“スーパー scaler マイクロプロセッサ” FLARE” の性能評価

岡本 光正^{*1} 的場 司^{*1} 山上 宣彦^{*2} 山田 朝彦^{*3} 須藤 英彦^{*1}

*1 (株) 東芝 青梅工場 ASIC技術部

*2 (株) 東芝 情報処理・機器技術研究所 開発第8部

*3 (株) 東芝 情報処理・機器技術研究所 開発第7部

〒198 東京都青梅市末広町二丁目九番地

あらまし

SPARCアーキテクチャ(V7)仕様の2命令スーパー scaler マイクロプロセッサ“FLARE”を開発した。FLAREは命令、データに分離された4KB毎の1次キャッシュメモリを内蔵している。また、SPARCレファレンスMMU仕様のメモリ管理機構も内蔵している。さらに、浮動小数点演算チップ、128KBの2次キャッシュメモリをそれぞれ専用のインターフェイスで接続することによりシステム性能を向上させることができる。FLAREはHS1.0 μ CMOSスタンダードセルの半導体技術を用いて開発された。本稿ではFLAREの仕様、開発手法、2命令スーパー scaler アーキテクチャの評価及びスーパー scaler を意識したコード最適化について述べる。

和文キーワード SPARC、RISC、スーパー scaler、マイクロプロセッサ、コード最適化

Performance of the Super scaler Microprocessor "FLARE"
based on SPARC architecture

^{*1}Kousei Okamoto, ^{*1}Tsukasa Matoba, ^{*2}Nobuhiko Yamagami,
^{*3}Asahiko Yamada & ^{*1}Hidehiko Sudou

*1 ASIC engineering Dept, OME works, TOSHIBA

*2 Engineering workstation development Dept, Information systems Lab, TOSHIBA

*3 Advanced computer architecture Dept, Information systems Lab, TOSHIBA

2-9 suehiro-cho, ome-shi, Tokyo, JAPAN 198

Abstract

We have developed two instructions super-scaler microprocessor "FLARE" based on SPARC architecture(V7). FLARE has the internal cache memories which are separated 4KB instruction cache and 4KB data cache. Also, FLARE has the memory management mechanism specified with SPARC reference MMU. Moreover, FLARE can be coupled with Floating unit and the secondary 128KB cache memory directly in order to enhance the system performance. FLARE employed HS1.0 μ -CMOS standard cell semiconductor technology. This paper describes the FLARE outline, development method, evaluation of two instruction super-scaler architecture and the code optimizer.

英文 key words SPARC, RISC, super scaler, microprocessor, code optimizer

1. はじめに

近年、高性能RISCマイクロプロセッサの出現により、ワークステーションは急激に性能を向上させている。RISCプロセッサは動作周波数を上げることにより性能向上を図ってきたが、高速信号の取扱いや発熱などの実装上の問題、あるいは半導体設計の困難さ（開発期間の長期化）などの問題が発生してきた。

FLAREはSPARCアーキテクチャV7に基づき、スーパースケラ（2命令並列実行）アーキテクチャを採用することにより高速化を目指すとともに、スタンダードセル方式で実現することによって開発期間の短縮を図った。

本稿ではFLAREの仕様、開発フロー、スーパースケラアーキテクチャの評価及びコード最適化について報告する。

2. FLAREアーキテクチャ

2.1. FLARE概略仕様

FLAREはSPARCアーキテクチャV7の命令セットをサポートする。またチップ上にSPARCレファレンスMMU仕様のMMUと命令、データ各々に分離された4KB毎の1次キャッシュメモリを持つ。MMU内には命令用に32エントリーのTLB、データ用に16エントリーTLBがある。

FLAREは2レベルの階層キャッシュメモリシステムを採用しており、外部に接続される128KBのキャッシュメモリチップを直接制御する。

浮動小数点演算は外部に接続されたFPUにより処理される。FPUは4段の演算パイプラインで構成され、このパイプラインはFLARE内で制御される。

また、FLAREはシステムに用いられるEWS標準バス（Mbus）を直接チップから出力する。

FLAREはHS1.0μCMOSスタンダードセル方式のセミカスタム半導体技術を用いており、33MHzの動作周波数を実現した。

表1にFLARE概略仕様、図1に概略ブロック図を示す。

表1. FLARE概略仕様

命令セット	SPARC (V7) 互換
パイプライン段数	4段
同時実行命令数	2命令
内蔵キャッシュメモリ	命令 (4KB) + データ (4KB)
内蔵MMU	SPARCレファレンスMMU仕様
FPU	外付け (WTL4064)
2次キャッシュメモリ	T9490E × 2 (128KB)
チップバス	Mbus
使用半導体技術	HS1.0μCMOS スタンダードセル
動作周波数	33MHz

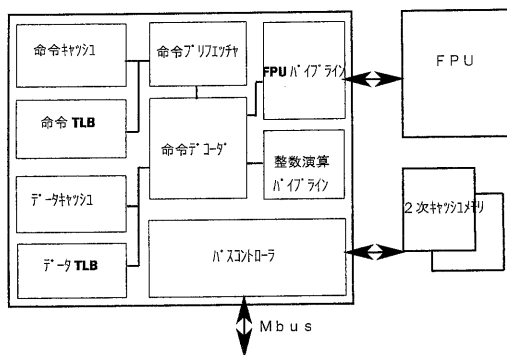


図1. FLARE概略ブロック図

2.2. パイプラインアーキテクチャ

FLAREは命令フェッチ（IF）、命令デコード（DE）、実行（EX）、データ書き込み（WR）の4段のパイプラインで構成される。また、各ステージでは2命令の処理が同時に実行される。

IFステージでは命令フェッチアドレスを用いて命令キャッシュ、命令TLBが検索される。命令キャッシュにヒットした場合は命令バッファに2つの命令が読み込まれる。命令キャッシュに命令が無い（ミス）場合は命令キャッシュの更新シーケンスに入り、更新が終了するまでIFステージの状態が続く。

DEステージでは2つの命令が同時にデコードされる。2つの命令が同時実行可能な組み合わせであり、かつレジスタの競合が発生しない場合に2つの命令が同時にEXステージに移される。同時実行条件を満足しない場合は先行命令のみEXステージに移される。また、同時にレジスタソースオペランドが読み出される。Load/Store命令のオペランドアドレスもこのステージで算出される。分岐命令がデコードされた場合はDEステージで分岐アドレスが算出され、IFステージに入る。

EXステージでは命令で指定される演算が行われる。また、Load命令ではデータキャッシュ、データTLBの検索が行われる。

WRステージでは演算結果、Load命令オペランドがレジスタファイルに格納される。

図2にパイプラインの各ステージの機能、図3に分岐命令実行時のパイプライン動作を示す。

	IF	DE	EX	WR
命令キャッシュ検索	命令TLB検索	命令デコード	整数演算実行	レジスタ書き込み
		分岐アドレス算出	データキャッシュ検索	
		オペランド算出	データTLB検索	

図2. パイプラインステージの機能

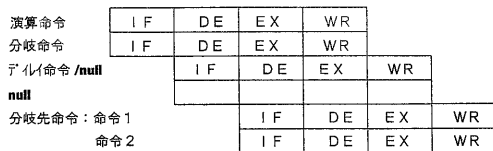


図3. 分岐命令実行時のパイプライン動作

浮動小数点演算命令はDEステージ以降に専用パイプラインで処理される。DEステージでデコードされた浮動小数点演算命令は、次に浮動小数点レジスタ競合のチェック及び演算制御信号ステージ(FDE)に入る。その後、浮動小数点レジスタ読みだしステージ(RR)、演算ステージ1(C1)、演算ステージ2(C2)、演算結果書き込みステージ(WR)の4段のFPU内のパイプラインで処理される。

図4に浮動小数点演算のパイプラインを示す。

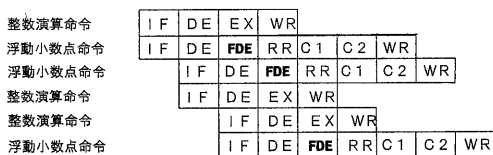


図4. 浮動小数点演算パイプライン

FLAREはパイプラインの実行効率を上げるため、ステージ間のデータバイパスを持っている。しかしながら、以下の場合にはDEステージからEXステージへの命令の発行が禁止される。

- ・Load/Store命令のアドレスレジスタがEXステージの命令のディステーションレジスタと一致する場合。

- ・分岐命令のインデックスレジスタがEXステージの命令のディステーションレジスタと一致する場合。
- ・IN,OUT,LOCALレジスタを使用する命令であり、EXステージでSAVEあるいはRESTORE命令が実行されている場合。
- ・Load命令であり、EXステージでStore命令を実行している場合。
- ・浮動小数点分岐命令であり、RWステージで浮動小数点比較命令が実行されている場合。

2. 3. 並列実行命令の組み合わせ

FLAREは整数演算命令・分岐命令・Load/Store命令・浮動小数点演算命令の内の任意の2命令を同時に実行する。

表2に詳細な2命令の組み合わせを示す。表中のシンボルは以下の通り。

- LOAD : LD,LDUB,LDUH,LDSB,LDSH
 STORE : ST,STB,STH,STD,LDD
 ALU : SETH,ADD,AND,OR,XOR,SUB,ANDN,ORN,XORN,ADDcc,ANDcc,ORcc,XORcc,SUBcc,ANDNcc,ORNcc,XORcc,ADDXcc,SLL,SRL,SRA,RDY,RDPSR,RDWIN,RDTBR
 FLDST : LDF,LDDF,LDFSR,STF,STDF,STFSR
 FPop : FPop1,FPop2
 SINGLE : TADDcc,TSUBcc,TADDccTV,TSUBccTV,WRY,WRPSR,WRWIM,WRTBR,Ticc,IFLUSH,SAVE,RESTORE,MULScc

表2. 同時実行可能命令の組み合わせ

次命令	Load	Store	ALU	FLDST	FPop	Bicc, CALL	JMPL, RETT	FBfcc	SINGLE
先行命令									
Load	不可	不可	可	不可	可	可	可	可	不可
Store	不可	不可	不可	不可	可	可	不可	可	不可
ALU	可	可	不可	可	可	可	可	可	不可
FLDST	不可	不可	可	不可	不可	可	可	可	不可
FPop	可	可	可	不可	不可	可	可	不可	不可
Bicc,CALL	可	可	可	可	可	不可	不可	不可	不可
JMPL,RETT	可	不可	可	可	可	不可	不可	不可	不可
FBfcc	可	可	可	可	可	不可	不可	不可	不可
SINGLE	不可	不可	不可	不可	不可	不可	不可	不可	不可

3. 開発手法

3. 1. 開発フロー

FLAREの開発ではC言語によるパイプラインモデル(FSIM)と機能記述言語(H2DL)によるハードウェア機能モデル及びテストプログラムを同時に作成した。FSIMはSPARCアーキテクチャモデルあるいはSPARCワークステーションを用いてテストプログラムを実行させることにより、その記述の正当性を検証した。その後、FSIMとハードウェア機能モデルとの動作を比較することにより、ハードウェア機能モデルの正当性を検証した。

ハードウェア機能モデルの正当性が確認されると論理合成、一部タイミングの厳しい論理回路は人手により論理回路を作成した。作成された論理回路はハードウェア機能モデルの機能シミュレーション結果から自動的に作成される論理シミュレーションテストデータを用いて論理シミュレーションを行った。

図5に開発フローを示す。また、表3にモデル記述量及びテストデータ量の設計データ量をを示す。

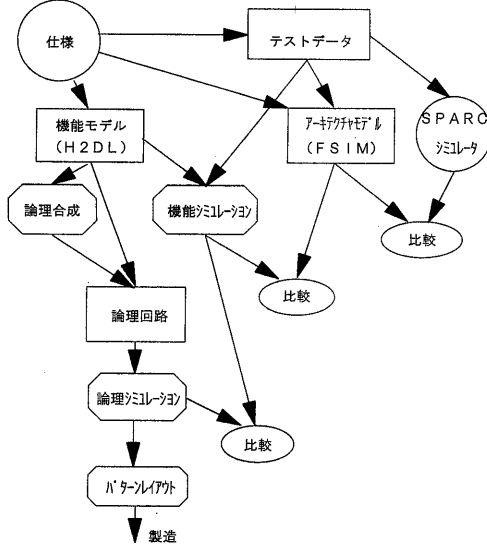


図5. FLARE開発フロー

表3. 設計データ量

設計データ	データ量
テストデータ	約 1,000,000step
FSIM (C言語)	19,000行
機能記述 (H2DL)	19,000行
論理記述 (TDL)	24,000行

3. 2. LSIレイアウト

LSIのレイアウトはチップサイズ、動作周波数に影響が大きい。FLAREはキャッシュメモリ、レジスタファイルといった大きなメガセル持ち、また加算器、ALU等のマクロセルも持っている。

これらのマクロセル間を効率よく配線するためにグルー回路は50のブロックに分けて階層設計した。

またレイアウトにあたっては、TLB、整数演算データパス、浮動小数点パイプラインレジスタは繰り返しの多い規則的なパターンのため、人手により標準セルを用いてレイアウトした。

最終的にはメガセル、マクロセル、人手レイアウトブロックを配置し、その間に3つのブロックに分けた自動レイアウトブロックを作成し、さらにその間のブロック間自動レイアウトを行った。

FLARE内のレイアウトブロックを表4に示す。

表4. FLAREレイアウトブロック

使用セル	RAM, 多ポートRAM, ALU, adder, シフト
人手レイアウトブロック	CAM, IUデータパス, FPUパイプライン
自動レイアウトブロック	5K, 10K, 15Kブロック
総トランジスタ数	約1Mトランジスタ

4. アーキテクチャ評価

FSIMを用いて、dhrystoneベンチマークテストプログラムの実行解析を行い、FLAREアーキテクチャの評価を行った。

表5に1クロックあたりの命令実行数を示す。2命令同時実行を行ったサイクルは21%である。表6に実行命令の種類を示す。表6から解るように、dhrystoneでは整数演算の頻度が高く、2つの整数演算を同時に実行できないFLAREのアーキテクチャでは2命令の同時実行頻度が少なくなっている。

また、表7にnullサイクル発生要因を示す。レジスタ更新待ちの頻度が最も高く、52%である。これはSPARC標準コンパイラがFLAREのアーキテクチャを考慮していないために発生している。また、分岐後の命令バッファへの命令読み込みに時間がかかるため命令を発行できない場合が35%ある。分岐によるnullサイクルの発生を減少させるには分岐予測機構が有効であると思われる。

表5. 1サイクルあたりの実行命令数頻度 (dhrystone)

実行命令数 0	20%
実行命令数 1	59%
実行命令数 2	21%

表6. 実行命令種類の出現頻度

整数演算命令	57%
分岐命令	20%
Load/Store命令	20%
単独実行命令	3%

表7. nullサイクル発生要因頻度

レジスタ更新待ち	52%
分岐後の命令待ち	35%
Store後のLoad命令	7%
ウィンドウ切り換え待ち	6%

5. FLARE用オプティマイザ

5. 1. オプティマイザ概要

スーパースカラ・プロセッサFLAREのCPU内並列化による効果を引き出すための言語処理系としては、アセンブル・テキスト変換プログラムを作成した。アセンブル・テキスト変換プログラムは、SPARC用のコンパイラが生成したアセンブル・テキストを読み込み、並列実行度を上げるような命令の並べ替え（命令スケジューリング）を行ない、アセンブル・テキストを出力する。実行に当たっては、アセンブル・テキスト変換プログラムの出力をSPARC用のアセンブラによりオブジェクト・コードに変換して実行させる。

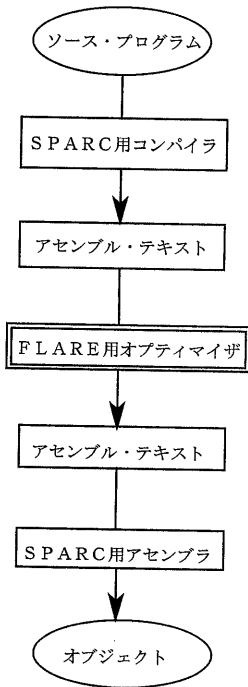


図6. オプティマイザの位置づけ

アセンブル・テキスト変換プログラムは、アセンブラ・テキストを読み込んで内部形式に変換し、その内部形式を基本ブロックに分割し、基本ブロック内の各命令間の依存関係を求めてDAG (Directed Acyclic Graph) を作成し、その依存関係をもとに命令スケジューリングを行ない、内部形式をアセンブル・テキストに変換する。

アセンブル・テキスト変換プログラムの役割は、命令スケジューリングにより並列実行命令を増加させ実行効率を向上させることであるが、その際パイプラインのインタロックを考慮することが不可欠である。すなわち、本変換プログラムの中心である命令スケジューラは、並列実行を考慮したパイプライン・スケジューラである。

5. 2. 命令スケジューラ

以下、命令スケジューラについて説明する。なお、命令スケジューリングの本質的な部分は基本ブロック内のスケジューリングにあると考え、今回は命令の並べ替えを基本ブロック内に限定した。

命令スケジューラは、一般に知られているパイプライン・スケジューラの技法に基づいて作成した。第1のフェイズで、予め決めた各々の命令固有の重みを基にスケジューリングの対象となるプログラムの命令依存DAGの各命令に重みを付ける。第2のフェイズで、命令に付けられた重みとプロセッサのハードウェア・リソースの状況に従い命令のスケジューリングを行なう。

5. 2. 1. 重みの計算

第1のフェイズで計算される重みは、以下のように求める。

各命令固有の重みは、先行命令とインタロックを起こさずに実行された時に先行命令の実行に隠れない分のクロック数とする。これはほとんどの命令に対しては1である。DAGの各エッジには、エッジをはさむ2命令が継続して実行された場合のインタロックのサイクル数の重みを付ける。

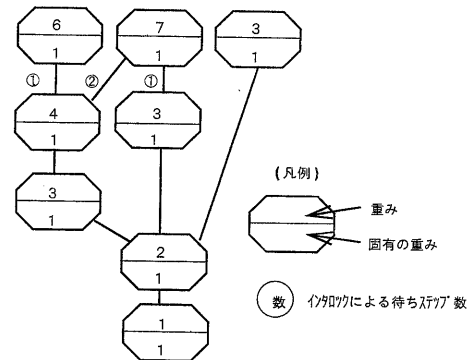


図7. Directed Acyclic Graph

DAG上の命令の重みは帰納的に求める。DAGの終端の命令の重みは命令固有の重みとする。終端命令でない命令の重みは、その命令に依存する命令それぞれについて、依存する命令の重みとはさまれるエッジの重みの和の最大値とする。すなわち、命令の重みは、その命令から終端に至る経路に沿って継続実行したときのクロック数の最大をほぼ反映するものである。

5. 2. 2. 命令スケジューリング

命令スケジューリングが参照する情報は静的情報と動的情報に分類できる。静的情報は、スケジューリング以前にDAGのみにより決まる情報であり、DAG上の各命令については重みのほか直接依存する命令の数、エッジについてはインタロックの最大クロック数の情報がある。動的情報は、スケジューリングの過程で決まる情報

であり、既にスケジューリングされた命令により決まるハードウェア・リソースであるレジスタ、バス等の使用状況、命令の実行状態等の情報である。命令に関する静的情報としては、重みを直接依存する命令の数よりも重視する。

命令のスケジューリング方法は、基本的には

- ・後続命令とインタロックを起こす命令は後続のインタロックに応じて優先的に選択する
- ・浮動小数点演算については、特に後続命令とのインタロックを起こし易いので、特別に優先的に選択する
- ・その時点で最後に選択されている命令と並列実行可能な命令は優先的に選択する
- ・既に選択された命令によりインタロックが生じる命令は選択しないようにする

であり、パイプライン・スケジューラの一般的な技法に従っている。具体的には、上に挙げた4点それぞれに対して決めた定数を静的情報である重みに加減し、いわば動的な重みを求めることにより、命令の選択を行なう。

5. 3. 効果

本命令スケジューラによる効果を見るためにいくつかのベンチマーク・プログラムについてシミュレータ上で実行、測定した。

使用したベンチマーク・プログラムは、

`dhrystone`

`livemore(#2,#3,#4,#5,#6,#8,#13)`

`linpack(87s,87d)`

である。なお、シミュレータ上での実行を考慮して、プログラムを部分的に書き換えた。ループ回数については、`dhrystone` と `livemore` については10回、`linpack` については各ループを約5分の1に短縮した。

コンパイラとアセンブラは、当社ASシリーズ付属のものをそれぞれ使用した。コンパイラの最適化オプションについては、無し、レベル2、レベル3について測定を行なった。

以下の表における数値は、左からインタロック・サイクル数、並列実行サイクル数、及び実行サイクル数それぞれのスケジューリング版の非スケジューリング版に対する比率である。また、左端の数の0は最適化無し、2は最適化レベル2、3は最適化レベル3を表す。

表8. `dhrystone` 実行サイクル数比

<code>dhrystone</code>	インタロック	並列	実行
0	0. 8 1 3	1. 0 7 0	0. 9 5 2
2	0. 7 8 7	0. 9 8 3	0. 9 6 9
3	0. 8 1 3	0. 9 8 2	0. 9 7 4

表9. `livemore #2` 実行サイクル数比

<code>livemore#2</code>	インタロック	並列	実行
0	0. 8 8 2	3. 1 9 6	0. 7 9 9
2	1. 0 6 8	1. 4 6 1	0. 9 6 3
3	0. 9 0 5	1. 2 2 0	0. 9 5 7

表10. `livemore #3` 実行サイクル数比

<code>livemore#3</code>	インタロック	並列	実行
0	0. 8 7 6	1. 8 6 0	0. 8 7 9
2	0. 9 9 6	2. 3 2 1	0. 8 6 7
3	0. 6 6 5	1. 2 7 6	0. 8 9 6

表11. `livemore #4` 実行サイクル数比

<code>livemore#4</code>	インタロック	並列	実行
0	0. 7 3 1	1. 8 3 5	0. 8 5 0
2	0. 5 6 4	1. 2 3 3	0. 8 7 4
3	1. 1 1 8	1. 1 9 1	0. 9 8 5

表12. `livemore #5` 実行サイクル数比

<code>livemore#5</code>	インタロック	並列	実行
0	0. 7 3 5	2. 2 2 8	0. 8 7 3
2	0. 9 7 3	1. 3 7 1	0. 9 5 9
3	0. 9 6 7	1. 0 7 3	0. 9 8 4

表13. `livemore #6` 実行サイクル数比

<code>livemore#6</code>	インタロック	並列	実行
0	0. 7 8 9	2. 2 1 5	0. 8 3 5
2	1. 1 2 0	1. 2 8 1	0. 9 9 0
3	0. 9 1 2	1. 0 8 7	0. 9 6 1

表14. `livemore #8` 実行サイクル数比

<code>livemore#8</code>	インタロック	並列	実行
0	0. 9 1 5	1. 9 9 4	0. 8 1 2
2	0. 8 8 4	1. 6 1 7	0. 9 0 5
3	0. 9 3 0	1. 7 2 1	0. 9 1 4

表15. `livemore #13` 実行サイクル数比

<code>livemore#13</code>	インタロック	並列	実行
0	0. 7 5 3	1. 4 4 9	0. 8 6 9
2	0. 9 7 3	1. 4 9 2	0. 9 5 4
3	0. 9 7 4	1. 4 0 6	0. 9 5 9

表 1 6. linpack87s 実行サイクル数比

linpack87s	インタロック	並列	実行
0	0. 6 6 6	1. 3 3 2	0. 8 3 8
2	0. 8 9 0	1. 2 4 3	0. 9 3 2
3	0. 8 4 9	1. 0 6 0	0. 9 3 1

表 1 7. linpack87d 実行サイクル数比

linpack87d	インタロック	並列	実行
0	0. 6 5 7	1. 3 9 0	0. 8 3 8
2	0. 8 9 5	1. 2 3 9	0. 9 3 9
3	0. 8 9 5	1. 1 1 9	0. 9 4 9

インタロックの減少は 43.6%(livermore#4 最適化レベル 2) から -12.0%(livermore#6 最適化レベル 2), 並列実行の増加は 219.6%(livermore#2 最適化無し) から -1.8%(dhrystone 最適化レベル 3), 実行サイクル数の減少は 20.1%(livermore#2 最適化無し) から 1.0%(livermore#6 最適化レベル 2) となっている。

FLARE では整数命令どうしの並列実行が限られているため、浮動小数点演算を含まないプログラムである dhrystone での命令スケジューラの効果は小さい。一方、livermore 以下の浮動小数点命令を含むプログラムにおいては、実行サイクル数で比較すると、最適化無しの場合で 15% から 20% 程度、最適化レベル 2 では 5% から 15% 程度、最適化レベル 3 では 5% から 10% 程度の改善効果が現れている。

しかしながら、実行サイクル数で見れば命令スケジューラによる性能の劣化は生じてないが、いくつかの例においてインタロックの増加、並列実行の減少が見られる。並列実行を増加させたために生じるインタロック、インタロックを回避するために減少する並列実行の存在がある。インタロックの減少の要求と並列実行の増加の要求は、このように対立する場合がある。この 2 つの要求のより良いバランス関係を求めるための改良が必要である。

6. まとめ

SPARCアーキテクチャ (V7) に基づいた 2 命令スーパー スケラマイクロプロセッサ” FLARE” を HS1. 0 μ CMOS スタンダードセル方式で開発できた。しかしながら性能評価結果から、まだ以下の改良の余地が残されていることがわかった。

- ・実行性能向上にともなう命令の供給問題。

- 分岐予測による分岐実行時の命令の先取り。

- 命令キャッシュのヒット率の向上。

- ・並列実行度の向上。

- 整数演算の並列実行化。

- 3 命令以上の並列実行化。

今後ハードウェアにはこのような改善をほどこしていく。

また、RISC の高速化にはコンパイラの改善が必要不可欠である。今回はコード最適マイザにとどめたが、命令間のインタロックはコンパイラにさかのぼって改善して行かなければならない。

7. 謝辞

最後に、FLARE の開発にあたり、数々の半導体技術の支援と LSI レイアウトのために惜しみない努力をいただきました (株) 東芝半導体事業本部 IC センター産業用 LSI 開発部の技術者に感謝いたします。