

## パーセプトロン分岐予測器を用いた 予測ミスする分岐命令の効率的分離

二ノ宮 康之<sup>†††</sup> 阿部 公輝<sup>††</sup>

分岐予測器の予測精度と構造の複雑さはトレードオフの関係にあり、制御ハザードをさらに削減したい場合はマルチパス実行との併用が望ましい。本稿ではパーセプトロン分岐予測器の累算値を利用した確信度付き分岐予測器を提案する。提案手法を用いることで、後続パスの同時実行を行う分岐予測を全体の4%とした場合、確信度カウンタを用いる従来法で発生する制御ハザードの24%を削減することができる。また、パーセプトロン分岐予測器と確信度カウンタの複合確信度付き分岐予測を使用することにより、同時実行を行う分岐命令を全体の10%とした場合、従来法で発生する制御ハザードの18%を削減することができる。

### Effective Separation of Miss-Predicted Branch Instructions Using Perceptron Branch Predictor

YASUYUKI NINOMIYA<sup>†††</sup> and KÔKI ABE<sup>††</sup>

There is a trade-off between prediction accuracy and hardware complexity of branch predictors. For reducing more control hazards, it is effective to execute multiple paths for a branch with little confidence on the prediction. In this paper, we propose a perceptron branch predictor which provides a confidence of the prediction using the accumulated value of the weights. The proposed method can reduce control hazards remaining in conventional multipath execution scheme using confidence counter by 24% in case that multiple paths are executed for 4% of branch instructions. We also propose a method which determines the confidence of a branch prediction from the value obtained by perceptron branch predictor and the value obtained by a confidence counter. The method can reduce control hazards by 18% in case that multiple paths are executed for 10% of branch instructions.

#### 1. はじめに

分岐予測は制御ハザードを回避し、CPUの性能を向上させる有用な手法として現在幅広く使用されている。特に、高精度分岐予測器は近年活発に研究が行われている。しかし、分岐予測器は限られた情報を利用して予測を行うため、予測が困難な分岐命令が存在する。

制御ハザードを回避する別の手法としてマルチパス実行がある。マルチパス実行は分岐命令の分岐成立と不成立の両方の後続パスを同時に実行することでCPUの性能を向上させることができるが、同時実行するパ

ス数は使用できるハードウェア量により制限される。同時実行するパスを減らすためには分岐予測器が予測に失敗しやすい分岐命令のみを実行する必要がある。このため、分岐予測器が予測に失敗しやすい分岐命令を選別する必要がある。

本稿ではマルチパス実行に利用するためのパーセプトロン分岐予測器を用いた2種類の確信度付き予測手法を提案する。パーセプトロン分岐予測器は高精度予測手法の一種であり、複数の重みと呼ばれる値を累算し、予測を行う。この累算値の絶対値を用いて予測の確信度を評価することで、従来法よりも効率よく予測ミスする分岐命令を選別することができる。

#### 2. 関連研究

本章では制御ハザードを回避する手段として用いられる分岐予測器とマルチパス実行の2つの手法を紹介する。

<sup>†</sup> 日本電気株式会社 システム IP コア研究所  
System IP Core Research Laboratories, NEC Corporation  
<sup>††</sup> 電気通信大学 情報工学科  
Department of Computer Science, The University of  
Electro-Communications

## 2.1 分岐予測器

分岐予測とはプロセッサのある分岐命令が分岐する方向をあらかじめ予測することである。プロセッサは予測結果を元に投機的に命令を実行し、パイプラインの充填率を向上させる。これにより制御ハザードを回避し、処理速度を高めることができる。

中でも、パーセプトロン分岐予測器<sup>1)</sup>は履歴の部分的な学習状態を累算によって統合することにより予測を行う。すなわち、重みテーブルから読み出された複数の重みと呼ばれる符号付き整数値を累算する。本論文では分岐を高い精度で予測すると同時に予測の確信度も提供することを考える。パーセプトロンにおける累算値は確信度を与える良い候補となりうる。

パーセプトロン分岐予測器は、分岐方向を予測すべき分岐命令（以下、branch B）のアドレスの他に、過去に実行した分岐命令のアドレスを時系列に並べた実行パス履歴、過去の分岐結果の履歴を時系列順に並べたグローバル履歴、この分岐履歴を対の情報である分岐命令のアドレスを用いて並び替えたローカル履歴などを用いて予測を行う。

branch B のアドレスや実行パス履歴をインデックスとして重みテーブルから重みを読み出す。これらの重みはそれぞれグローバル履歴の特定の場所と 1 対 1 に対応している。グローバル履歴の各要素は、1 ならば分岐成立、-1 ならば分岐不成立を表わす。読み出された重みを  $W_i$ 、対応するグローバル履歴の要素を  $X_i$ 、 $1 \leq i \leq n$ 、とし、次式を計算する。

$$y = W_0 + \sum_{i=1}^n W_i X_i \quad (1)$$

$W_0$  は閾値であり、重みと同様に読み出される。求められた重みの累算値  $y$  の値が正ならば分岐成立として 1 を、負ならば分岐不成立として -1 を出力する。重み  $W_i$  の更新規則は、分岐結果を  $t \in \{1, -1\}$  とすると次式によって表される。通常、 $\alpha = 1$  とする。

$$W_i = W_i + \alpha t X_i \quad (2)$$

これまで様々なパーセプトロン分岐予測器が提案されている<sup>2)3)4)5)</sup>。中でも Compact Perceptron Branch Predictor (CPBP)<sup>6)</sup> は他のパーセプトロン分岐予測器に比べて実装コストが低く、かつ高い予測精度を示す。

図 1 は CPBP の構造図である。CPBP は Stage 1 と Stage 0 の 2 段のパイプライン構造を持つ。Stage 1 では複数の重みテーブルから並列にそれぞれ 1 つの重みを読み出し、対応する GHR ビットの値が -1 ならば値のビット反転を行う。重みは累算され、途中結果

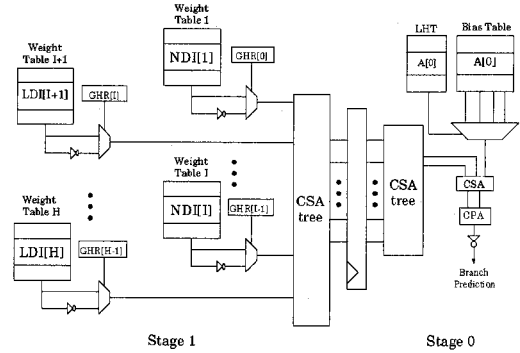


図 1 Compact Perceptron Predictor (CPBP) の構造図

がパイプラインに保存される。

Stage 0 では Bias Table から閾値の読み出しを行う。閾値は候補が全て読み出された後で、Local History Table (LHT) から読み出されたローカル履歴によって 1 つが選択される。この閾値はパイプラインに保存されていた重みの累算の途中結果と加算される。求められた累算値の最上位ビットを反転し、予測結果として出力する。

パーセプトロン分岐予測器やハイブリッド予測<sup>7)</sup>を行う高精度分岐予測器<sup>8)9)10)</sup>は様々な学習状態を分散記憶し、それらを選択統合して予測を行うことで予測精度を向上させている。しかし、この傾向はより複雑な回路を必要とし、実装の妨げとなりうる。また、分岐予測器の予測の基となるグローバル履歴、ローカル履歴、実行パス履歴は CPU が処理する命令の部分的な情報であり、完全な学習を行うことは不可能である。これは、分岐予測器単体では制御ハザードを従来法より大幅に削減することは困難であることを示している。

## 2.2 マルチパス実行

制御ハザードを回避する手法には分岐予測を用いる方法の他にマルチパス実行<sup>11)</sup>がある。分岐命令の分岐成立と不成立の両方の後続パスを実行することにより分岐ミスによるペナルティを回避することができる。しかし、実装面積の制約から、同時実行できるパス数には制限がある。このため、マルチパス実行を採用する場合には、後続パスを同時実行すべき（投機実行した場合にミスする可能性の高い）分岐命令を選別する必要がある。

選別には様々な手法が提案されているが、主要な選別手法の 1 つに予測の確信度を用いる手法<sup>12)13)</sup>がある。予測の確信度が高い分岐命令は投機的実行を行い、確信度の低い分岐命令のみ同時実行を行う。不必要なパス実行を減らすことで、実行できないパスや実行を

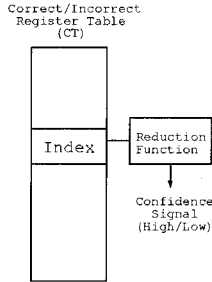


図2 確信度カウンタ (Confidence Counter) の構造図

遅延させるパスが減少し、プロセッサの処理速度を向上させることができる。よって、分岐予測器の予測精度と予測に対する確信度の評価精度が高いほど、マルチパス実行による処理速度の向上度は大きくなる。

予測の確信度を確信度カウンタ (Confidence Counter)<sup>14)</sup>を使用して測定する手法を図2に示す。Correct/Incorrect Register Table (CT) と呼ばれるテーブルから値を読み出し、Reduction Function 部で閾値を超えるか否かを判定する。値が閾値より大きい場合には予測の確かさを保証する。CT から読み出される値の更新手法は、予測が成功した場合には1を加算し、予測が失敗した場合には0にする手法が実験的に効率が良いことが分かっている。

しかし、確信度カウンタは分岐予測器と別に用意しなければならない。使用できる記憶回路の総量を一定とすると、確信度カウンタと分岐予測器でメモリを分割することになり、互いに精度を下げてしまう。

### 3. パーセプトロン分岐予測器の累算値と予測確信度

前章の通り、分岐予測器の実装コストの増加を抑えつつ、単体で制御ハザードの発生を従来法より大幅に削減することは難しい。このため、マルチパス実行と組み合わせるのが効果的であると考えられる。マルチパス実行と高精度分岐予測器との組み合わせは高い効果を生むと予想される。しかし、高精度分岐予測器を用いる場合でも、確信度カウンタのような確信度を測る機構は必要である。確信度カウンタと分岐予測器の両方を用意することは記憶容量の制限が一層深刻となるため効率的とは言えない。この問題を改善するためには、分岐予測器が予測と同時に確信度を提供できることが重要である。

確信度を提供するためには分岐に対する学習 (適応) が十分進行しているかを知る事が1つの指標となる。学習が十分である予測に対して、学習不十分の予測の

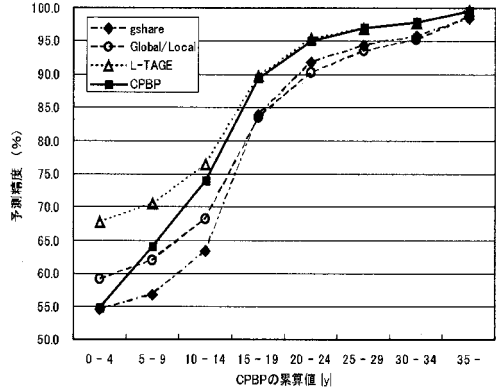


図3 CPBP の累算値  $|y|$  と予測精度

予測精度が低くなることは自然である。学習が難しい分岐もプログラムには存在する。そのような分岐も学習の進行状況を知ることで判別できる。

パーセプトロン分岐予測器では、入力されたグローバル履歴のパターンに対する学習の進行度を表す指標として、予測計算時に求められる累算値の絶対値  $|y|$  がある。  $|y|$  がある閾値 (学習閾値) を超えた場合、学習 (重みの更新) は十分進行したとみなされ、さらなる学習は行われぬ。これは、複数の部分予測を累算によって集計する構造上、過学習を防ぐため元々備えられた機構である。  $|y|$  の値が学習閾値に満たない場合、学習が不十分であり、学習閾値を超えた場合と比較して予測精度が低いことが考えられる。

一方、ハイブリッド予測器は各テーブルで異なる履歴長のグローバル履歴を用いた完全な予測を行い、最適なものをセレクトにより選択する。したがって、過学習に対する備えは無い。よって、学習の進行度を表す指標は得られない。

パーセプトロン分岐予測器の累算値を確信度として用いる妥当性について実験を行った。実験はSPEC INT2000の12種類のベンチマークをSimCoreシミュレータ<sup>15)</sup>で1億命令程度実行した際にトレースしたデータを用いた。パーセプトロン分岐予測器はCPBPを用いた。CPBPの記憶容量は32KB、学習閾値は文献<sup>6)</sup>を参考に20とした。その他の各種パラメータについては、4種類のアドレスをインデックスの生成に使用するCPBP 4x<sup>6)</sup>のものを使用した。

図3の実線はCPBPの  $|y|$  の値と予測精度の関係を示している。図より  $|y|$  の値が小さいほど予測精度が低いことが分かる。また、学習閾値の20未満では予測精度の低下が顕著である。平均の予測精度は97.3%で

あり、この予測精度を超えるのは  $|y|$  の値が閾値の 1.5 倍以上である 30 を超えた場合であった。このことから、パーセプトロン分岐予測器単体による確信度付き予測は可能であると分かった。

本研究はマルチパス実行にパーセプトロン分岐予測器が持つ特性を利用することが本題であるが、予測の確信度が低い場合は他の分岐予測器の予測結果を採用する形のハイブリッド予測器も考えられる。パーセプトロン分岐予測器から得られる確信度をこの形のハイブリッド予測器のセレクトアとして使用する可能性を検討するため、CPBP の累算値と他の分岐予測器の予測精度との関係を実験により求めた。

分岐予測器は gshare 予測器<sup>7)</sup>、Global/Local Perceptron 分岐予測器<sup>1)</sup>、L-TAGE 予測器<sup>8)</sup> を使用した。これらの予測器は CPBP と独立して予測を行う。予測結果を CPBP の累算値の絶対値によってグループ分けし、それぞれの予測精度を求めた。各予測器の使用記憶容量は 32KB とした。各種パラメータは各予測器の引用文献の通りである。結果は図 3 に示してある。図から他の分岐予測器も CPBP の累算値の絶対値  $|y|$  が大きいほど高い予測精度を示している。すなわち、 $|y|$  の値が小さい分岐命令は他の予測器でも予測が難しいということが分かる。以上から、パーセプトロン分岐予測器から得られる確信度をこの形のハイブリッド予測器のセレクトアとして使用する効果はあまりない。

見方を変えれば、パーセプトロン分岐予測器の累算値の絶対値を用いることで、他の分岐予測器では予測が難しい分岐命令を選別することが可能であると言える。パーセプトロン分岐予測器により確信度が低いと判断された分岐命令は、どの分岐予測手法を用いても予測が難しい。したがって、これらの分岐命令による制御ハザードを回避するためには、マルチパス実行は有用であると考えられる。

#### 4. 累算値を用いた確信度付き予測

パーセプトロン分岐予測器の累算値を利用する確信度付き予測について検討してきたが、問題となるのはその性能である。マルチパス実行において、実装コストの観点から同時実行できるパスの数には制限があるため、処理能力の向上には可能な限り結果的に予測ミスする分岐命令のみを実行させる必要があるのは 2 章で述べた通りである。以下では、実行される分岐命令のうち、閾値を  $T$  として確信度  $C$  が  $|C| < T$  となる分岐命令の割合を分離率と定義する。パーセプトロン分岐予測器単体を選別に用いる場合は重み付き分岐履

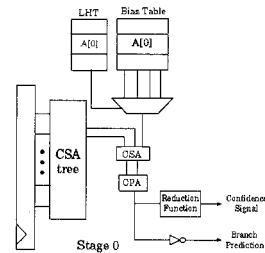


図 4 CPBP の Stage 0 への Reduction Function の追加

歴の累算値の絶対値  $|y|$  の値を確信度とする。確信度カウンタを選別に用いる場合は CT に記憶されている各エントリの値を確信度とする。

また、実際に予測ミスする分岐命令が、確信度によって分離された分岐命令の集合に含まれる割合を、分離効率とする。より低い分離率で高い分離効率を得られる手法が、高性能な確信度付き予測手法であると言える。パーセプトロン分岐予測器単体による分離効率と、確信度カウンタを選別に使用した場合の分離効率を実験により比較した。

実験は 32KB の CPBP と、これに様々な容量の確信度カウンタを追加したものを比較した。CPBP の記憶容量を一定としているのは予測ミス率を同じにすることにより分離効率の比較を単純にするためである。CPBP は確信度を求められるように、累算値の信号線を分配し、追加した端子には確信度カウンタと同様の Reduction Function を追加した。図 4 は実験に使用した CPBP の Stage 0 の構造図である。また、確信度カウンタの CT は 1 エントリ 4 bit とし、記憶容量は 1KB、4KB、16KB の 3 種類を用意した。すなわち、確信度カウンタを利用する場合、分岐予測器と合計した記憶容量は 33KB、36KB、48KB となる。実験環境は 3 章と同様である。実験は CPBP 単体では閾値を 0 から 4 ずつ、確信度カウンタは閾値を 0 から 1 ずつ変化させることにより行った。閾値を 0 とすると予測命令の分離率は 0 である。閾値を上昇させることにより、分岐命令の分離率を増加させることができる。

結果を図 5 に示す。図から、CPBP 単体による分離は分離率が 8~10% より小さい場合、確信度カウンタよりも分離効率が高いことが分かる。しかし、分離率がこれより大きくなると CPBP 単体による分離効率は確信度カウンタを用いた手法よりも劣る。

確信度付き予測を行う場合、低い確信度の分岐命令に対してはその後続パスの同時実行を行い、高い確信度の分岐命令に対しては分岐予測器の予測に基づいて

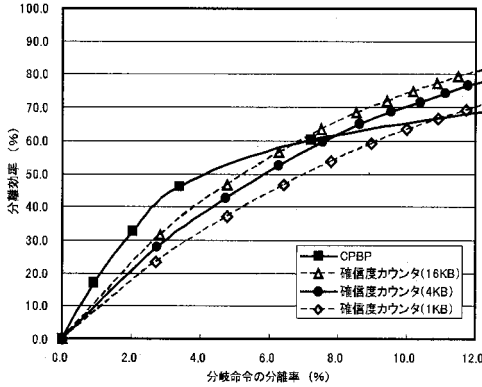


図5 確信度カウンタによる分離とCPBPによる分離の比較

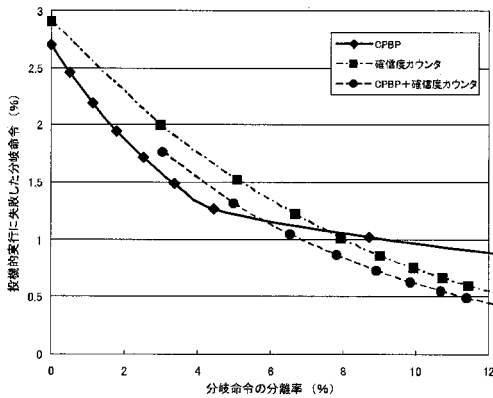


図6 確信度カウンタによる分離とCPBPによる分離の比較

後続パスの片方の投機的実行を行う。確信度付き予測を行う回路全体の記憶容量を32KBとして、分岐命令に対する実行（同時実行と投機的実行の両方）のうち、投機的実行に失敗する割合を測定する。実験は32KBのCPBP単体による確信度付き分岐予測（提案手法）と、16KBのCPBPと1エン트리4bitの16KBのCTによる確信度カウンタのみを使用した確信度付き予測（従来法）に対して行う。実験環境は3章と同様とする。

結果を図6に示す。図は、全ての分岐命令に対する実行のうち投機的実行に失敗した割合が、分離率を変えた時にどう変化するかを表す。菱形は提案手法、正方形は従来法を表す。図より、分離率が7.5%以下の時、CPBP単体による確信度付き分岐予測は確信度カウンタを使用したCPBPよりも高い性能を持つことが分かる。提案手法は、投機的実行に失敗した割合が1.5%の時、後続パスの同時実行を35.7%削減できる。また、提案手法は、従来法で発生する制御ハザードを、分離率を2%とした場合は18%、4%とした場合は24%、削減できる。

ドを、分離率を2%とした場合は18%、4%とした場合は24%、削減できる。

マルチパス実行は実行すべきパスが増える必要となるハードウェア量も増加するため、分離率が大きいと必要となるハードウェア量が上昇する。従って、分離率の小さいところで分離効率が高いパーセプトロン分岐予測器単体による確信度付き予測の利用価値は高いと言える。

## 5. パーセプトロン分岐予測器と確信度カウンタによる複合確信度付き予測

前章のように、確信度カウンタの代わりにパーセプトロン分岐予測器の累算値の絶対値 $|y|$ を使用する手法は、同時実行する分岐命令が全体の8~10%程度以下の限定されたマルチパス実行機構を持つ場合に有効である。しかし、使用できるハードウェア量に制限が少ない場合、より多数のパスの投機的実行も考えられる。そのため、パーセプトロン分岐予測器単体では分離効率が低い分離率が10%以上の領域でも、分離効率を向上させる目的で提案手法に確信度カウンタを加えることを考える。確信度カウンタを使用する場合には必ず分岐予測器が必要になるが、分岐予測器に提案手法を用いれば実装コストの増加を抑えられる可能性がある。2種類の確信度を統合する最も単純な方法は、2種類の確信度の両方が閾値を超えた場合に予測が確かであるとするのである。この手法（複合確信度付き分岐予測）の有用性を実験により検討する。

実験は16KBのCPBPと16KBのCTによる複合確信度付き分岐予測を32KBのCPBP単体による確信度付き分岐予測、16KBのCPBPと16KBのCTによる確信度カウンタのみを使用した確信度付き予測と比較した。実験環境は3章と同様とする。また、複合予測でのCPBPの確信度閾値は10とした。この値は図5においてCPBP単体で効果的な範囲から選択した。

結果は図6に丸のプロットで示してある。CPBPと確信度カウンタの複合確信度付き予測を用いる場合、確信度カウンタによる単一の確信度付き予測を使用した場合よりも投機的実行に失敗した分岐命令の割合が低くなった。図より、投機的実行に失敗した分岐命令の割合が1.0%の時、複合確信度付き分岐予測は確信度カウンタのみを使用した場合と比較してマルチパス実行を14.4%削減することができた。また、複合確信度付き分岐予測は、確信度カウンタのみを使用した場合に発生する制御ハザードを、分離率を8%とした場合は16%、10%とした場合は18%削減できた。CPBP

単体の確信度付き予測と比較した場合、同時実行を行う分岐命令の割合が5.7%を超える場合は複合確信度付き分岐予測の方が投機的実行に失敗した分岐命令の割合が低くなっている。

以上の結果より、マルチパス実行機構による同時実行パス数の制限が少ない場合にはCPBPと確信度カウンタによる複合確信度付き分岐予測が有効な手法であることが分かった。

## 6. 実装コストの考察

本章では提案手法の実装コストに関する考察を行う。

パーセプトロン分岐予測器単体による確信度付き分岐予測器を行う場合、パーセプトロン分岐予測器に追加すべきハードウェアは確信度閾値を超えたかを判定する回路、すなわち確信度カウンタのReduction Functionが必要となるだけである。また、確信度付きパーセプトロン分岐予測器の確信度の学習は従来のパーセプトロン分岐予測器の学習回路で行えばよいため、追加のコストは発生しない。また、確信度カウンタを使用する従来法と比べ、CTとその更新回路は不要である。

複合確信度付き分岐予測を使用する場合、確信度カウンタを使用する従来法に比べて増加するのはパーセプトロン分岐予測器のReduction Function回路と確信度信号の統合に用いるAND回路である。この手法の使用が想定されるのは実装コストの制限が少ない場合であり、この実装コストの増加は許容される範囲内であると考えられる。

## 7. まとめ

本稿ではパーセプトロン分岐予測器の累算値の絶対値を利用した予測確信度付き分岐予測機構を提案した。提案手法は確信度カウンタを用いる従来法よりも高い分離効率を実現することが分かった。後続パスの同時実行を行う分岐予測を全体の4%とした場合、従来法で発生する制御ハザードの24%を削減することができる。

また、パーセプトロン分岐予測器と確信度カウンタの複合確信度付き分岐予測を使用することにより、同時実行を行う分岐命令を全体の10%とした場合、従来法で発生する制御ハザードの18%を削減することができる。この手法はより多数のパスを同時実行できるアーキテクチャに対して有効である。

**謝辞** AlphaプロセッサシミュレータSimCoreを提供していただいた東京工業大学大学院情報理工学研究所の吉瀬謙二講師に感謝する。本研究は、一部、日本学

術振興会科学研究費補助金(基盤研究(C)(2)18500048)による。

## 参考文献

- 1) D. A. Jimenez, and C. Lin : Dynamic Branch Prediction with Perceptrons, *Proc. the Seventh International Symposium on High-Performance Computer Architecture (HPCA'01)*, pp.197-206 (2001)
- 2) D. A. Jimenez : Fast Path-Based Neural Branch Prediction, *Proc. the 36th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO-36)*, pp.243-252 (2003)
- 3) D. A. Jimenez: Piecewise Linear Branch Prediction, *Proc. the 32nd Annual International Symposium on Computer Architecture (ISCA'05)*, pp.382-393 (2005)
- 4) 石井康雄, 平木敬 : 実行パス履歴情報を利用した分岐予測手法, 情報処理学会論文誌 : コンピューティングシステム, 47 巻, SIG3(ACS13) 号, pp.58-72 (2006)
- 5) Y. Ninomiya and K. Abe : A3PBP: A Path Traced Perceptron Branch Predictor Using Local History for Weight Selection, *The Journal of Instruction-Level Parallelism*, Vol.9 (2007)
- 6) 二ノ宮康之, 阿部公輝 : CPBP: 実行パス履歴を有効に利用する低コスト高精度パーセプトロン分岐予測器, 情報処理学会論文誌 : コンピューティングシステム, Vol.1, No.1, pp.1-9 (2008)
- 7) S. McFarling: Combining Branch Predictions, WRL Technical Note TN-36 (1993)
- 8) A. Seznec : The L-TAGE Branch Predictor, *The Journal of Instruction-Level Parallelism*, Vol.9 (2007).
- 9) Y. Ishii : Fused Two-Level Branch Prediction with Ahead Calculation, *The Journal of Instruction-Level Parallelism*, Vol.9 (2007)
- 10) H. Gao and H. Zhou : PMPM: Prediction by Combining Multiple Partial Matches, *The Journal of Instruction-Level Parallelism*, Vol.9 (2007)
- 11) A. K. Uht and V. Sindagi : Disjoint Eager Execution: An Optimal Form of Speculative Execution. *Proc. of 28th Annual International Symposium on Microarchitecture (MICRO-28)*, pp.313-325 (1995)
- 12) S. Wallace, B. Calder, and D. M. Tullsen : Threaded multiple path execution, *Proc. 25th Annual International Symposium on Computer Architecture (ISCA'98)*, pp.238.249 (1998)
- 13) 高木特通, 平木敬 : 高度パイプライン化プロセッサ上のトレースに基づく複数パス実行方式, 電子情報通信学会技術研究報告, Vol. 101, No. 2, pp.73-80 (2001)
- 14) E. Jacobsen, E. Rotenberg, and J. E. Smith : Assigning Confidence to Conditional Branch Predictions. *Proc. of 29th Annual International Symposium on Microarchitecture (MICRO-29)*, pp.142-152 (1996)
- 15) 吉瀬謙二, 片桐孝洋, 本多弘樹, 弓場敏嗣 : SimCore/Alpha Functional Simulator の設計と実装, 電子情報通信学会論文誌, Vol. J88-D-I, No.2, pp.143-154 (2005)