

二電源電圧を用いた命令発行メモリの低消費電力化手法

辻 寛司

井上 弘士

モシニヤガ ワシリー

福岡大学大学院 工学研究科 電子工学専攻

〒 814-0180 福岡県福岡市城南区七隈 8 丁目 19 番 1 号

E-mail : {tsuji, inoue, vasily}@v.tl.fukuoka-u.ac.jp

あらまし：命令発行メモリ (命令ウィンドウ) の低消費電力化を目的として、適応型命令発行メモリが提案された。プログラムが有する命令レベル並列度に応じて使用可能なエントリ数 (命令発行メモリ・サイズ) を動的に最適化し、負荷容量を削減することで低消費電力化できる。本稿では、更なる低消費電力化を実現するため、二電源電圧を用いた適応型命令発行メモリを提案する。従来の適応型命令発行メモリでは、単一電源電圧が用いられる。これに対し、提案手法では、命令発行メモリ・サイズを縮小した際に低電源電圧を使用する。つまり、命令発行メモリ・サイズの変更に応じて電源電圧も変化させる。CMOS 回路の消費電力は電源電圧の 2 乗に比例するため、低電源電圧化により大幅な消費電力の削減を期待できる。また、命令発行メモリ・サイズを縮小した場合にのみ低電源電圧を用いるため、低電源電圧化に伴う遅延時間オーバーヘッドを隠蔽できる。評価を行った結果、本手法を適用することで、大幅な性能低下を伴うことなく最大 36% の命令発行メモリ消費電力を削減できた。

キーワード 低消費エネルギー, 低消費電力, 命令発行メモリ, スーバスカラ, 動的電源電圧

Reducing Energy Dissipation of Complexity Adaptive Issue Queue by Dual Voltage Supply

Hiroshi Tsuji

Koji Inoue

Vasily G. Moshnyaga

Department of Electronic Engineering, Fukuoka University

8-19-1 Nanakuma, Jonan-ku, Fukuoka-shi, Fukuoka, 814-0180 Japan

E-mail : {tsuji, inoue, vasily}@v.tl.fukuoka-u.ac.jp

Abstract: This paper presents a novel architectural technique to reduce energy dissipation of adaptive issue queue, whose functionality is dynamically adjusted at runtime to match the changing computational demands of instruction stream. In contrast to existing schemes, the technique exploits a new freedom in queue design, namely the voltage per access. Since loading capacitance operated in the adaptive queue varies in time, the clock cycle budget becomes inefficiently exploited. We propose to trade-off the unused cycle time with supply voltage, lowering the voltage level when the queue functionality is reduced and increasing it with the activation of resources in the queue. Experiments show that the approach can save up to 36% of the issue queue energy without large performance and area overhead.

key words low energy, low power, issue queue, superscalar, dynamic voltage

1 はじめに

アウト・オブ・オーダー実行を許すマイクロ・プロセッサでは、発行前の命令を一時格納するために命令発行メモリ(命令ウィンドウ)が搭載される。デコードされた命令は、全ての依存関係や資源競合が解消されるまで(つまり、発行可能となるまで)この命令発行メモリ内に保持される。命令発行メモリ・サイズ(総エントリ数)の拡大によりマイクロ・プロセッサの先見能力が向上し、より高い性能を実現できる。そのため、更なる性能向上を目的として、命令発行メモリ・サイズは年々増加傾向にある [1]。

この命令発行メモリへのアクセスは毎クロック・サイクル発生する。その結果、命令発行メモリの消費電力が、プロセッサ全体の消費電力に大きく影響を与えるようになってきた。例えば、128 エントリを有する命令発行メモリを搭載したマイクロ・プロセッサにおいて、全消費電力の約 25% を命令発行メモリが消費する [2]。したがって、低消費電力マイクロ・プロセッサを実現するためには、命令発行メモリの低消費電力化が極めて重要となる。命令発行メモリの低消費電力化を目的として、適応型命令発行メモリが提案された [3]。プログラムが有する命令レベル並列度に応じて命令発行メモリ・サイズを縮小し、動的に負荷容量を削減することで低消費電力化を実現する。

本稿では、更なる低消費電力化を目的として、二電源電圧を用いた適応型命令発行メモリを提案する。従来の適応型命令発行メモリでは単一電源電圧が用いられており、負荷容量を削減することで低消費電力化を実現する。これに対し、提案手法では、命令発行メモリ・サイズを縮小した際に低電源電圧を用いる。つまり、命令発行メモリ・サイズの変更に応じて電源電圧も変化させる。CMOS 回路の消費電力は電源電圧の 2 乗に比例するため、低電源電圧化により大幅な消費電力の削減を期待できる。また、命令発行メモリ・サイズを縮小した場合にのみ低電源電圧を用いるため、低電源電圧化に伴う遅延時間オーバーヘッドを隠蔽できる。

以下、第 2 節では、文献 [3] で提案された適応型命令発行メモリに関して説明する。次に、第 3 節では、更なる低消費電力化を目的として、二電源電圧を用いた適応型命令発行メモリを提案し、その詳細を示す。第 4 節では定性的/定量的評価を行い、最後に第 5 節で簡単にまとめる。

2 適応型命令発行メモリ

命令発行メモリ内には複数のエントリが存在し、デコード後の命令は発行可能となるまでであるエントリに格納される。例えば、総エントリ数が 128 の場合には、最大 128 命令を格納可能となる。命令発行メモリの使用率は、プログラム実行における IPC(Instructions Per Cycle) に依存する。IPC が高い(つまり、命令レベル並列度が高い)場合には、毎クロック・サイクル多くの命令が発行されるため、必要となる命令発行メモリ・エントリ数は少ない。一方、IPC が低い(つまり、命令レベル並列度が低い)場合には発行待ち命令数が増加するため、多くの命令発行メモリ・エントリが必要となる。IPC は実行対象となるプログラム間やプログラム内で変化する。したがって、命令発行メモリに対する性能要求もプログラム実行と伴に変化する。6 種類のベンチマーク・プログラムの実行に関して、命令発行メモリ内エントリの使用頻度を図 1 に示す(これらの結果は、文献 [3] から参照した)。図より、*bzip*、*mcf*、*parser*、*vpar* の実行では、多くのエントリ数を必要とすることが分かる。これに対し、*gcc* や *vortex* では、80% 以上の命令発行はわずか 8 エントリしか必要としていない。

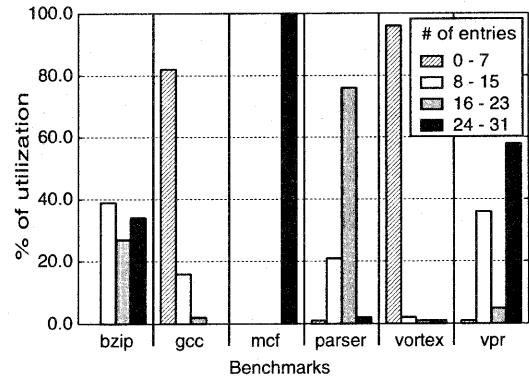


図 1: 命令発行メモリのエントリ使用頻度

そこで、プログラム実行における IPC に応じて命令発行メモリ・エントリ数(命令発行メモリ・サイズ)を動的に最適化する適応型命令発行メモリが提案された [3]。適応型命令発行メモリは分割構造をとる。例えば、図 2 に示すように、命令発行メモリを 4 つのバンクに分割する(バンク 0~3)。そして、命令発行メモリに対する要求性能が低い(つまり、IPC が高い)場合には、少数バンクのみを使用可能とする(例えば、バンク 0 のみ使用可

能)。これにより、命令発行メモリ・アクセスにおける消費電力を削減する（負荷容量を削減）。一方、命令発行メモリに対する要求性能が高い（つまり、IPCが低い）場合には、多数のバンクを使用可能とする（例えば、全てのバンクを使用可能とする）。この場合には、通常の命令発行メモリと同様の電力を消費する。実際には、命令発行メモリを4つのバンクに分割した場合、「バンク0のみ使用可能」、「バンク0と1のみ使用可能」、「バンク0~2のみ使用可能」、「全てのバンクが使用可能」といった4つの動作モードを定義できる。なお、プログラム実行における動的な動作モード決定アルゴリズムに関しては文献 [3] を参照されたい。

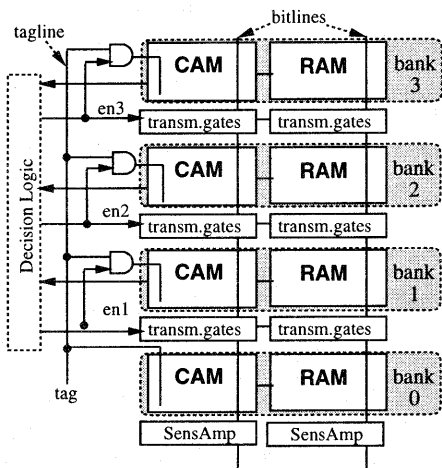


図 2: 適応型命令発行メモリの内部構成

3 二電源電圧を用いた適応型命令発行メモリ

我々は、更なる低消費電力化を目的とし、二電源電圧を用いた適応型命令発行メモリを提案する。適応型命令発行メモリ手法では、単一電源電圧を用いているのに対し、我々は二電源電圧を用いる。適応型命令発行メモリにおいて、命令発行メモリ・サイズを動的に変化させた場合、それに応じて命令発行メモリの遅延時間 (D_{queue}) も変化する。なぜなら、電源電圧を一定と仮定すると、命令発行メモリの遅延時間はそのサイズや幅に比例するからである。つまり、適応型命令発行メモリの遅延時間 (D_{queue}) はアクティブになっているバンク数によって変化する。命令発行メモリのクロック・サイクル時間 (T_{clock}) は、すべてのバンクがアクティブになっているときの遅延時間を基準としており、アクティブになっているバンク数が

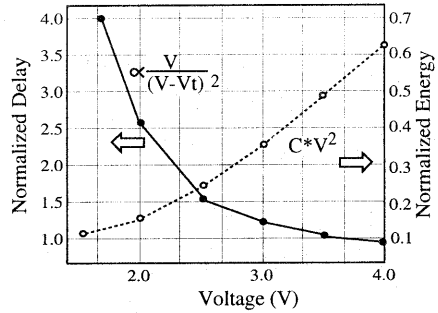


図 3: CMOS 回路における電源電圧と遅延時間のトレードオフ

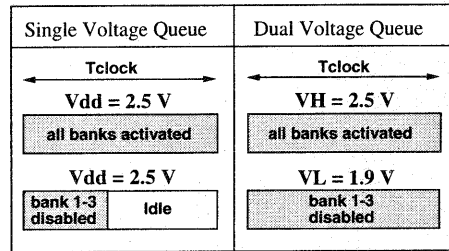


図 4: 提案する手法の基本概念

少ない場合、遅延時間がクロック・サイクル時間を大きく下回る。その結果、図 4 に示すように、何も処理を行わない無駄な時間が発生する（以降、この無駄な時間をアイドル時間 ($T_{idle} = T_{clock} - D_{queue}$) と呼ぶ）。我々は、このアイドル時間を有効に活用することで、命令発行メモリの低消費電力化を行う。

図 3 に示すように、CMOS 回路において、遅延時間は、電源電圧にほぼ反比例するため、電源電圧を下げると遅延時間は増加する。しかしながら、適応型命令発行メモリにおいて、バンク 1~3 がノン・アクティブ (disable) の場合、命令発行メモリの遅延時間 (D_{bank}) はクロック・サイクル時間 (T_{clock}) よりも短い（つまり、アイドル時間が発生する）。したがって、このときのクロック・サイクル時間 (T_{clock}) を満足する範囲内であれば、性能低下を伴うことなく命令発行メモリの電源電圧を下げることが可能となる。

図 5 に二電源電圧を用いた適応型命令発行メモリの内部構成を示す。図 2 に示す適応型命令発行メモリに対して、二つの電源電圧 (V_H と V_L) を切替えるための選択回路 (S) を追加する。ここで、 V_H は通常の電源電圧、 V_L は V_H よりも低い電源電圧とする。この選択回路の

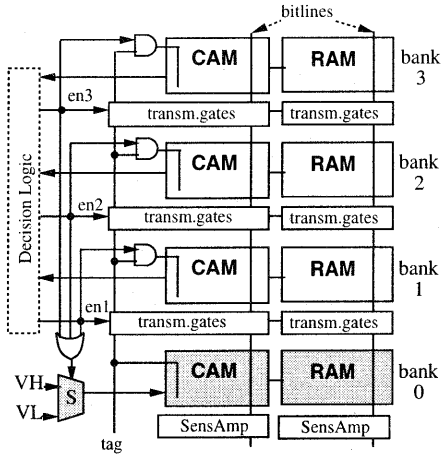


図 5: 二電源電圧を用いた適応型命令発行メモリの内部構成

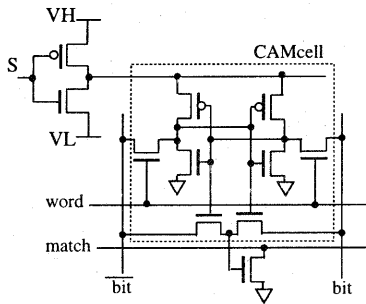


図 6: 選択回路の回路図

構造を図 6 に示す。提案する適応型命令発行メモリは、具体的に次のように動作する。決定回路によりバンク 1~3 が不必要とみなされると、制御信号がトランスミッション・ゲートに送られ、バンク 1~3 がシャットダウンされる。それと同時に、制御信号が選択回路 (S) にも送られ、低い電源電圧 (VL) が選択される。逆に、もしバンク 1~3 が必要になった場合は、同様に選択回路 (S) により、通常の電源電圧 (VH) が選択される。つまり、4 つのバンクのうち、バンク 0 のみを使用する場合には低い電源電圧 (VL) を、バンク 0~3 まで全てを使用する場合には通常の電源電圧 (VH) を使い、これらの 2 つの動作モードを動的に選択する。先に述べたように、低電源電圧化に伴う遅延時間の増加は、アイドル時間により隠蔽することができる。

4 評価

4.1 消費エネルギー・モデル

本評価では、文献 [3] で提案された従来の適応型命令発行メモリ (以降、従来型発行メモリと呼ぶ) を比較対象とし、二電源電圧を用いた適応型命令発行メモリ (以降、提案型発行メモリと呼ぶ) の消費エネルギーを評価する。第 2 節で説明したように、従来型発行メモリでは、4 分割構成を有する場合に 4 つの動作モードを定義できる (つまり、バンク 0 のみが動作するモード、バンク 0~1、バンク 0~2、バンク 0~3 が動作するモード)。よって、従来型発行メモリに本提案手法を適用する場合、各モードそれぞれに対して異なる電源電圧を印加して低消費電力化を実現することが可能である (つまり、4 種類の電源電圧を用いる)。しかしながら、使用する電源電圧数の増加は、電源電圧の変更に伴う消費電力/面積/遅延オーバーヘッドを招く恐れがある。そこで、本評価では、適応型命令発行メモリの動作モードを以下の 2 種類に限定する。

- サイズ縮小モード: バンク 0 のみがアクセス対象となる動作モード。
- 通常モード: 全てのバンクがアクセス対象となるモード。

また、命令発行メモリの総エントリ数は 32 と仮定する。

あるプログラムを実行した場合、従来型発行メモリの消費エネルギー (E_{adp}) は以下の式で近似することができる。

$$E_{adp} = N \times [AR_{b0} \times E_{b0_VH} + (1 - AR_{b0}) \times E_{all_VH}]$$

- N : プログラム実行における総命令発行メモリ・アクセス回数。
- AR_{b0} : 命令発行メモリ・アクセスにおいて、バンク 0 のみがアクセス対象となる確率 (つまり、サイズ縮小モードでのアクセス回数/ N)。
- E_{b0_VH} : サイズ縮小モード・アクセス当たりの平均消費エネルギー。ただし、通常の電源電圧 (VH) を使用。
- E_{all_VH} : 通常モード・アクセス当たりの平均消費エネルギー。通常の電源電圧 (VH) を使用。

これに対し、提案型発行メモリでは、サイズ縮小モードにて動作する場合、より低い電源電圧 VL ($VH > VL$)

を用いる。したがって、提案型発行メモリの消費エネルギー (E_{dual}) は以下ようになる。

$$E_{dual} = N \times [AR_{b0} \times E_{b0_VL} + (1 - AR_{b0}) \times E_{all_VH}]$$

$$E_{b0_VL} = E_{b0_VH} \times VL^2 / VH^2$$

ここで、 E_{b0_VL} は、低電源電圧 (VL) を使用した際の、サイズ縮小モード・アクセスに関する平均消費エネルギーである。なお、本評価においては、消費エネルギー削減率を以下の式で表す。

$$\text{消費エネルギー削減率} = (1 - E_{dual}/E_{adp}) \times 100$$

さらに、実際には、電源電圧の切替えに要する性能/消費エネルギーのオーバーヘッドが生じる。このようなオーバーヘッドは、電源電圧切替え用スイッチやその他周辺回路の実装方法に大きく依存する。そこで本評価では、理想的な場合を想定し、これらのオーバーヘッドは0と仮定した。なお、通常電源電圧 VH は、従来型/提案型発行メモリ共に 2.5V とする。

4.2 定性的評価

第3節では、命令発行メモリを4バンクに分割した場合を想定して提案手法を説明した(総エントリ数 = 8 エントリ \times 4バンク)。しかしながら、実際には、命令発行メモリを2バンク構成とすることも可能である(総エントリ数 = 16 エントリ \times 2バンク)。本節では、4分割構成ならびに2分割構成の命令発行メモリに関して、提案手法の有効性を定性的に評価する。

第4.1節で示した消費エネルギー式に基づき、提案手法における消費エネルギー削減率を求めた。その結果を図7に示す。提案型発行メモリにおいては、サイズ縮小モード時のアクセス時間が通常モード時のアクセス時間を越えない範囲内で、低電源電圧 VL を決定しなければならない。そこで、4分割構成に関しては単純な回路モデルを作成し、サイズ縮小モードならびに通常モードにおける命令発行メモリ・アクセス時間を spice シミュレーションにより求めた。そして、これらの結果に基づき、 $VL=1.9v$ と仮定した¹。また、2分割構成に関しては、4分割構成における電源電圧値を参考にし、 $VL=2.2v$ と仮定した。

図7で示すように、60%程度の命令発行メモリ・アクセスがサイズ縮小モードで実行される場合には、提案手

¹ 具体的には、文献 [4] を参考に、サイズ縮小モード動作時、ならびに、通常動作モード時における命令発行メモリの負荷容量を求めた。そして、この負荷容量の充放電に用いる時間を各モードにおけるアクセス時間と仮定した。

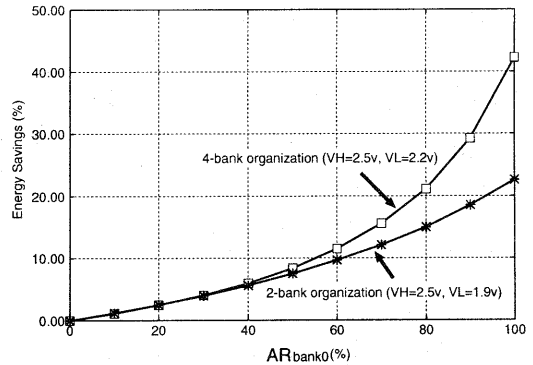


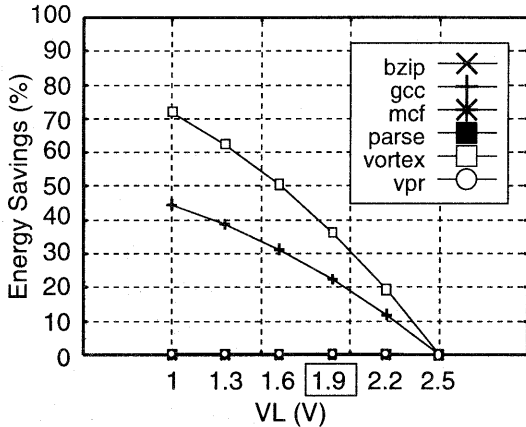
図7: 提案型発行メモリの消費エネルギー削減率(定性的評価)

法を用いることで10%以上の消費エネルギーを削減できる。また、全ての命令発行メモリ・アクセスがサイズ縮小モードで実行される理想的な場合では、2分割構成で最大42%、4分割構成で最大22%消費エネルギーを削減できることが分かった。

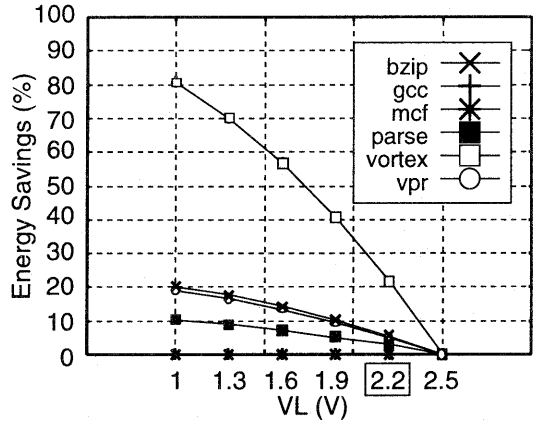
4.3 定量的評価

第4.2節においては、2分割構成および4分割構成に関して、サイズ縮小モード時に印加される低電源電圧 VL をそれぞれ 2.2v および 1.9v と仮定した。しかしながら、 VL は命令発行メモリの設計技術、回路技術、プロセス技術等、様々な要素技術に依存する。そこで、本節では、第4.1節で示した消費エネルギー式、ならびに、文献 [3] で示されたシミュレーション結果(図1)に基づき、 VL をパラメータとした場合の消費エネルギー削減効果を定量的に評価する。

低電源電圧 VL を変化させた際の消費エネルギー削減率を図8に示す。まず、4分割構成に関して議論する。図8(A)で示すように、*gcc* や *vortex* に関しては、 VL の削減と共に高い消費エネルギー削減率を達成している。これに対し、その他のプログラムに関しては、消費エネルギー削減効果を得ることができなかった。これは、図1から分かるように、*gcc* や *vortex* では使用される命令発行メモリ・エントリ数が少ない(8エントリ以下)のに対し、その他のプログラムでは多くのエントリを必要としているためである。一方、2分割構成の場合、4分割構成の場合と比較して、全てのプログラムにおいて消費エネルギー削減率を向上している。これは、命令発行メモリを(4分割ではなく)2分割することでバンク内エントリ数が2倍となり、その結果、サイズ縮小モードでのア



(A) 4-bank organization



(B) 2-bank organization

図 8: 提案型発行メモリの消費エネルギー削減率 (定量的評価)

クセス回数が増加したためである。

しかしながら、2分割構成においては、1バンク当たりの負荷容量が増大するため、4分割構成のVLほどは低電源電圧化できない。第4.2節で示したように、2分割構成におけるVLを2.2v、4分割構成におけるVLを1.9vと仮定した場合、図8より以下のことが分かる。

- 4分割構成 (VL = 1.9v) をとった場合、gccで約23%、vortexで約36%の消費エネルギー削減を達成している。これに対し、その他のプログラムに関しては殆んど消費エネルギーを削減できていない。
- 2分割構成 (VL = 2.2v) をとった場合、バンク当たりの負荷容量が増加するため、第3節で説明したアイドル時間が短くなり、4分割構成ほどは低電源電圧化できなくなる。その結果、gccとvortex共に約20%の消費エネルギー削減である。

以上より、命令発行メモリの使用効率が低い(つまり、命令レベル並列度が高い)場合にはメモリ分割数を増やすことにより高い消費エネルギー削減効果を得ることができる。一方、命令発行メモリの使用効率が低い(つまり、命令レベル並列度が低い)プログラムに関しては、メモリ分割数の削減が有効であると考えられる。

5 おわりに

本稿では、適応型命令発行メモリの更なる低消費電力化を目的として、二電源電圧を用いた命令発行メモリの

低消費電力化手法を提案した。従来の単一電源電圧を用いる手法に比べて、二電源電圧を適用することにより、命令発行メモリの消費電力を最大36%削減できることを示した。本稿では、電源電圧を切替える際に生じる性能/消費エネルギーオーバーヘッドを0として評価を行った。今後、これらのオーバーヘッドを考慮したより詳細な評価を行う予定である。

謝辞

日頃から御討論頂く、福岡大学モシニヤガ研究室の諸氏に感謝します。

参考文献

- [1] R.Kessler, "The Alpha 21264 microprocessor", *IEEE Micro*, 19(2), pp.24-36, March/April 1999.
- [2] D.Folegnani and A.Gonzalez, "Reducing Power Consumption of the Issue Logic", *Proc. ISCA Workshop on Complexity-Effective Design*, June 2000.
- [3] A. Buyuktosunoglu, S. Schuster, D. Brooks, P. Bose, P. Cook, D. Albonese, "An Adaptive Issue Queue for Reduced Power at High Performance", *Workshop on Power Aware Systems, held at 9-th Int. Conf. on Architectural Support for Programming Languages and Operating Systems*, November 2000.
- [4] S.Palacharla, N.P.Jouppi, J.E.Smith, "Quantifying the complexity of Superscalar Processors", Technical Report CS.TR-96-1328, University of Wisconsin-Madison, Nov. 1996.