論文の内容の要旨

論文題目　クラスタ化プロセッサにおける分散投機メモリフォワーディング手法の研究

氏　　名　　入江 英嗣

情報処理の中核となるマイクロプロセッサには常に性能向上が期待されている。マイクロプロセッサの構成や動作を設計するマイクロアーキテクチャ研究の分野では、動作周波数の向上と、1 サイクルあたりの処理の並列度の双方のバランスを取りながら、全体の性能を向上させてきた。今後プロセッサ設計を取り巻く状況を概観すると、Simultaneous Multi Threading 技術やメディア演算など、広い並列実行幅の利用が見込まれる一方、配線遅延の影響が深刻になることが予測されており、高い動作周波数と広い実行幅を両立させる、効率的な設計が求められている。

このような視点から、近年、実行部分を複数の軽量な実行 node に分割する、“clustered microarchitecture” が研究者の注目を集めている。命令実行において node を跨ぐ通信が発生する場合、通信遅延が IPC(instruction per cycle:実行並列度)上のペナルティとなるが、テーブルの分散化やフォワーディングパスの短縮により、集中型構成に較べて動作周波数を高速化することができる。このため、クラスタ化による周波数向上の効果が、IPC 低下によって相殺されなければ、クラスタ構成を採用する利点がある。

本論文はこの clustered microarchitecture のキャッシュシステムに着目し、問題点の解析と、改善方法の提案を行う。評価のために、高い動作周波数向けにセッティングされた clustered super scalar モデルを策定し、トレースシミュレータに実装した。

clustered architecture のキャッシュには、アクセス遅延を増加させる要因がいくつか存在する。実行 node とキャッシュ間の通信遅延、キャッシュテーブルを参照するためのアクセス遅延などは主に配線遅延に起因しており、クラスタ化によって配線遅延の影響を軽減している実行部分との間で速度の乖離が生じてくる。また複数の node がキャッシュを共有している構成では、キャッシュポートの競合を回避するための調停が必要となるが、通信レイテンシが無視できないほど大きいため、調停にかかる遅延も大きなものとなる。これらの理由から、先行する clustered architecture 研究モデルで見られる集中型のキャッシュモデルでは、アクセス遅延が著しく増加することが予想される。一方、既存の分散キャッシュモデルでは、個々のキャッシュを小容量にしてアクセス遅延を軽減させることができるものの、コンシステンシを保つための通信オーバヘッドが大きくなってしまう。また node 間でアクセスが干渉しない、アドレスによる分割法では、キャッシュを小容量にでき、コ

ンシステンシのオーバヘッドも回避できるが、フロントエンド処理時にどの node のキャッシュを使用するか予測して実行 node を決定する必要があるため、予測器の精度の低さが問題となる。

node 内に小さなバッファを設け、アクセスを node 内のみに限定すれば、アクセス遅延やポート競合の影響を軽減することができる。本研究では、このような限定されたバッファの利用法として、メモリ依存予測に基づく投機データフォワーディングに適用することを提案する。一般にストア命令とロード命令のメモリ依存関係には局所性があることが知られており、依存のあったストア命令とロード命令の PC を記憶しておくことにより、高い確率でストア-ロードのメモリ依存関係を予測することができる。提案手法では、フロントエンド処理でこの依存予測情報をマージし、依存関係にあるストア命令とロード命令が同じ node で実行されるようにステアリング処理を行う。実行 node 内には小容量のバッファが追加され、ストア命令の動的な ID とストア値がペアで保持される。ロード命令はこの小容量バッファを、親と予測されたストア命令の ID によって参照することで、値を投機的に得ることができる。この機構を採用することで、正しく依存予測を行えたロード命令は、長いキャッシュアクセスレイテンシを回避することができ、後続する命令を滞らせずに実行させることができる。

この手法の利点は、node 内に設けるバッファが非常にシンプルで高速なこと、小エントリで高い効果を得ることが期待できること、親のストアを監視することにより、正確に後続命令の予測 wakeup ができること、などが挙げられる。一方、バッファの利用が依存予測手法の適用率に依存すること、データ投機であるため、キャッシュへのチェックロードが必要であること、ミスフォワードをしてしまった場合はパイプラインフラッシュが必要であることなどが欠点である。

シミュレータにより、まずメモリ依存予測手法の適用率について傾向を調べ、次に、バッファ容量と IPC 向上の相関について調べた。これらの評価を通して、本手法では非常に少ないエントリ数のバッファにより大きな効果を得ることができることを示す。