

論文の内容の要旨

論文題目 **Improving Cache Performance by Exploiting Redundancy, Temporal Affinity, and Reuse of Data**

(データの冗長性、時間的親和性、再利用を用いたキャッシュの性能向上)

氏名 高木 将通

近年、急速に拡大したプロセッサと DRAM の間の速度差は、プログラム実行速度を規定する主要な原因の 1 つである。プロセッサと DRAM 間の速度差を埋めるキャッシュメモリを用いたプロセッサにおいても、超高速プロセッサではキャッシュミスが数百プロセッササイクルの待ち時間を引き起こす。このため、キャッシュミスの削減がコンピュータシステムの実行性能の向上に必須である。キャッシュミスは、(1) 置き換えの判断を誤ったために生じるキャッシュミス、(2) メモリブロックに対する初めての参照によるキャッシュミスや、キャッシュの容量が足りないために起こるキャッシュミスなど、置き換えの判断が正しくても起こるキャッシュミス、の二つに大別される。本論文はそれぞれの種類のキャッシュミスを減らす方法を提案する。

あるデータと別のデータが近い時間に参照されるとき、それらは時間的親和性を持つと言う。このデータの時間的親和性と、データの冗長性を利用して、(2)のキャッシュミスを削減する。数値計算プログラムに関しては、コンパイラはデータの時間的親和性を解析することができる。一方で、再帰的構造体(**Recursive Data Structures, RDS**)を用いるプログラムにおけるデータの時間的親和性を解析することはコンパイラには難しい。**RDS** は非数値計算プログラムでは広く用いられている。そこで **RDS** によって起こされるキャッシュミスを減らすことを目的として、**Field Array Compression Technique (FACT)**と呼ぶ、ソフトウェアとハードウェアによる方法を提案する。**FACT** は、時間的親和性を持つデータを、メモリ上での配置を変換することによってメモリの連続領域に集める。その後ポインタフィールドと整数フィールドを圧縮する。結果として、1回のキャッシュフィルが時間的親和性を持つより多くのデータをキャッシュに運ぶことができるようになる。さらに、圧縮によって実効的なキャッシュ容量が増大する。**FACT** を実行駆動シミュレーションによって評価する。ポインタを頻繁に用いる **Olden** ベンチマークの 8 個のプログラムにおいて、**FACT** はキャッシュミスによる待ち時間を平均 41.6%削減する。また、速度を平均 37.4%向上する。

次にキャッシュ置き換えアルゴリズムの性能を向上することにより、(1)によるキャッシュミスを削減する。そこでプロセッサのセットアソシアティブ 2 次キャッシュのための新たな置き換えアルゴリズム、**Inter-Reference Gap Distribution Replacement**

(IGDR) を提案する。IGDR は、メモリブロックに重みを付け、置き換えの際は、最小重みのメモリブロックを選択する。メモリブロックが次に参照される時刻が分かっているとすると、その時刻と現在の時刻との差の逆数が、理想的なメモリブロックの重みである。あるメモリブロックが参照された時刻において、前回参照されてから経過した時間を参照間隔と呼ぶ。提案手法では理想的な重みを参照間隔の逆数を使って近似する。そのために、各メモリブロックの参照間隔の確率分布は一定であると想定する。そして、参照間隔の逆数の統計を取ることによって、参照間隔の逆数の分布を推定する。この分布から参照間隔の逆数の期待値を推定し、この推定値をメモリブロックの重みとする。統計は、参照回数、参照間隔によってメモリブロックをクラスに分類し、クラスごとに取る。SPEC CPU2000 の 10 個のプログラムにおいて、提案手法は最大 48.8%、平均 17.0% のミス数の削減、最大 39.6%、平均 9.1% の速度向上を示す。