

C-004

ライトバックキャッシュにおけるメインメモリアクセスレイテンシの隠蔽 Main Memory Access Latency Tolerance in Write Back Cache

下村 佳生[†]
Yoshio Shimomura

小林 良太郎[†]
Ryotaro Kobayashi

1. はじめに

近年のコンピュータシステムではプロセッサの計算速度が著しく向上し、主記憶へのアクセス速度との間に大きな隔たりが生じている。この問題を解決するために、プロセッサは主記憶よりもアクセスの早いキャッシュを持っており、使用頻度の高いデータをキャッシュ上にコピーすることで全体の性能向上を図っている。

ここでキャッシュ上のデータに対する書き込みについて議論を行う。キャッシュは主記憶のコピーのため、整合性を保つためにはキャッシュと主記憶の両方についてデータの更新を行う必要がある。常に整合性を保つ手法としてライトスルー方式があるが、この手法の場合、速度の遅い主記憶へアクセスするため全体の性能が低下してしまう。キャッシュへのもう一つの書き込み手法であるライトバック方式の場合、キャッシュ上のデータに更新があったことを示す印を付けて主記憶への書き込みを行わないので高速なデータ書き込みが可能となる。ただしこの方式の場合、コンテキストスイッチ発生時などキャッシュ上のデータを破棄する段階で更新のあったデータ調べて主記憶への書き戻しを行うという動作が必要となる。

そこで本稿ではライトバック方式を拡張し、キャッシュから主記憶への書き戻しを定期的に行う機構を提案する。これによりキャッシュ破棄時における主記憶書き戻し遅延を抑制でき、ライトバック方式における性能の向上が見込める。

2. キャッシュ書き込み方式

キャッシュの代表的な書き込み方式であるライトスルー方式とライトバック方式について説明を行う。

ライトスルー方式

ライトスルー方式は図1に示すように、書き込み時にキャッシュと主記憶の両方のデータを更新する手法である。この方式は常に最新のデータが主記憶に書き込まれるため、同期を考慮する必要がない、動作が単純なので制御が行いやすいなどの利点が挙げられるが、主記憶への書き込み待ち時間が問題になるという欠点がある。

ライトバック方式

ライトバック方式は図2に示すように、書き込み時に

キャッシュのみを更新し、主記憶の更新を行わない手法である。この方式の場合、主記憶とキャッシュの整合性が保たれていない期間が存在する。そこでキャッシュ上のデータが更新されていることを示すために、キャッシュライン上にダーティビットと呼ばれる更新があったことを示す印を用意しておき、データの更新が行われた場合はこの値をセットすることにより、主記憶とキャッシュとの同期判定を行う。一般にダーティビットがセットされて主記憶と内容の整合性が取れていない状態をダーティと表現し、逆に整合性が保たれている状態をクリーンと表現する。

キャッシュ上へデータを読み込む際や、コンテキストスイッチ発生時などにキャッシュ上のデータを破棄する場合にはダーティビットを確認し、ダーティラインを主記憶に書き戻す動作が必要となる。この方式はデータの書き戻しを一括で行うため、ライトスルーで問題となった同期待ち時間は軽減できるが、ダーティビットの追加やチェックなどのため機構が複雑になる上、キャッシュ破棄時におけるダーティラインの書き戻しというオーバーヘッドが生じる欠点がある。

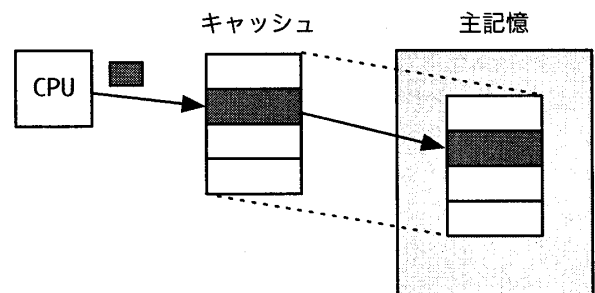


図1 ライトスルー方式

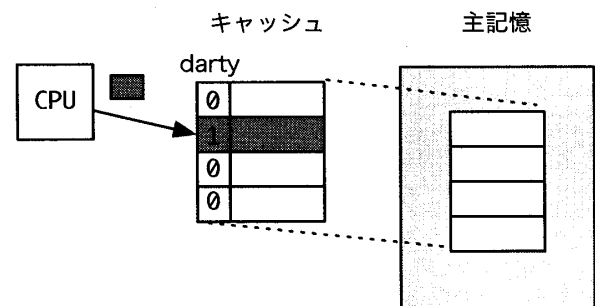


図2 ライトバック方式

[†] 豊橋技術科学大学大学院工学研究科
Graduate School of Engineering, Toyohashi University
of Technology

3. 提案手法

本稿ではライトバック方式におけるダーティライン書き戻し時のオーバーヘッドを削減することで、全体の性能向上を行うことを目的とする。

上記で説明したように、ライトバックキャッシュの場合、プログラム実行中の書き込み時間は緩和できるが、コンテキストスイッチなどのキャッシュ破棄時にダーティラインの書き戻しオーバーヘッドが発生する。この時にダーティライン数が少ないほど遅延時間が減ることは自明であるが、一度ダーティになったラインは同じライン上に新しいデータが読み込まれない限りダーティなままであり、全体のチェック、書き戻しが行われるまではクリーンにならない。

しかしながら使用されるキャッシュ範囲は空間的、時間的な局所性を持つことが知られており、一定期間アクセスがないキャッシュに関しては今後もアクセスされないと考えられる。そこで本手法では一定時間アクセスがないダーティラインをバックグラウンドで書き戻し、クリーンにすることでコンテキストスイッチ発生時のアクセス時間を隠蔽する手法を提案する。なお、同じように定期的な書き戻しすることでダーティラインの数を抑える機構はすでに提案されている [1] が、本稿ではより単純な機構で実現することを目指す。

提案する機構を図3に示す。キャッシュラインには通常のタグ、データフィールド、ダーティビットに加えて、最終アクセスからのカウンタを用意する。このカウンタは一定サイクルごとにインクリメントされ、ラインにアクセスがあるとリセットされる。カウンタが閾値を超えるとそのラインは使われなくなったと見なし、ダーティラインを主記憶へと書き戻す。これはカウンタを追加するだけのシンプルな機構で実現でき、キャッシュ破棄時の大量なダーティビットのチェック、ラインの書き戻しを抑制するのに有効であると考えられる。

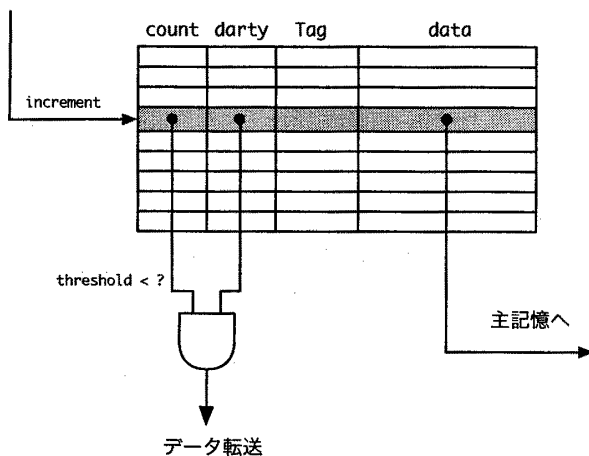


図3 提案機構

4. まとめ

ライトバック方式において、コンテキストスイッチ発生時に問題となる主記憶への書き戻しを軽減する手法について提案を行った。本手法は通常の機構に参照間隔測定用のカウンタを付けるだけのシンプルな構成で実現できる。

現在、コンテキストスイッチ発生時に書き戻すダーティライン数の測定中である。また、本機構を用いた場合、通常のライトバック方式に比べて書き戻し回数が増えるためバスの占有率が増加するという懸念点がある。この問題はシングルコアの場合は、性能に然したる影響を与えないと考えられるが、複数コアの場合は他コアのアクセス時間に影響を与える可能性がある。そこで、複数コア上で本機構を実行した際にバス占有率の増加が性能にどの程度影響を与えるのかを調査する予定である。

謝辞

本研究の一部は、文部科学省科学研究費補助金若手研究(B) 課題番号 21700057、柏森情報科学振興財団研究助成、及び、中山隼雄科学技術文化財団研究助成の支援により行った。

参考文献

- [1] 森眞一郎, 大森洋一, 中島浩, 富田眞治, “セルフ・クリーンアップ型ライトバック・キャッシュの提案,” 情報処理学会研究報告 1993-ARC-100, pp.9-16, 1993.