情報処理学会論文誌 コンピューティングシステム Vol.5 No.5 55–65 (Oct. 2012)

Starving Writerの解消によるLogTMの高速化

江藤 正通[1] 堀場 匠一朗[1] 浅井 宏樹[1,†1] 津邑 公暁[1,a)] 松尾 啓志[1]

受付日 2012年3月28日, 採録日 2012年4月27日

概要：マルチコア環境における並列プログラミングでは，メモリアクセスの調停には一般にロックが用いられてきた．しかしロックを使用する場合，デッドロックの発生や並列性の低下などの問題がある．そこでロックを用いない並行性制御機構としてLogTMが提案されている．LogTMではpossible_cycleというフラグを用いて競合を解決する．しかし，この競合解決手法ではstarving writerが発生し，長期にわたるストールや競合の繰返しにより性能が大きく低下してしまう．そこで本稿では，starving writerの解決手法を提案する．提案手法の有効性を検証するためにシミュレーションによる評価を行った結果，既存のLogTMに比べて最大で18.7%，平均で6.6%の性能向上が得られることを確認した．

キーワード：ハードウェア・トランザクショナル・メモリ，マルチスレッド，スケジューリング，メモリアクセス競合

A Speed-up Technique for LogTM by Relieving Starving Writers

Masamichi Eto[1] Shoichiro Horiba[1] Hiroki Asai[1,†1]
Tomoaki Tsumura[1,a)] Hiroshi Matsuo[1]

Received: March 28, 2012, Accepted: April 27, 2012

Abstract: Lock-based thread synchronization techniques are commonly used in parallel programming on multi-core processors. However, lock can cause deadlocks and poor scalabilities. Hence, LogTM has been proposed and studied for lock-free synchronization. To solve conflicts, LogTM uses a flag called 'possible_cycle.' However, the performance can decline with some conflict patterns. This paper proposes a method for dynamically changing the priority of threads to solve the conflict patterns. Our model reduces the number of aborts and recurrence of aborts. The result of the experiment shows that the proposed method improves the performance 18.7% in maximum and 6.6% in average.

Keywords: hardware transactional memory, multithreading, scheduling, memory access conflicts

1. はじめに

現在一般的となったマルチコア環境では，複数のプロセッサ・コア間で単一アドレス空間が共有されるプログラミングモデルが多く用いられる．このようなプログラミングモデルでは，共有リソースに対する競合を解決する必要があり，その排他制御機構として一般にロックが用いられてきた．しかしロックを用いた場合，デッドロックの発生やロック操作のオーバヘッド増大にともなう並列性の低下などの問題が起こりうる．さらに，各プログラムに最適なロックの粒度を設定することは困難であるため，プログラマにとって必ずしも利用しやすいものではない．そこで，ロックを用いない並行性制御機構としてトランザクショナル・メモリ [1] が提案されている．

トランザクショナル・メモリのハードウェアによる一実装であるLogTM [2] では，クリティカルセクションを含む一連の命令列として定義されるトランザクションが投機的に実行される．そして，処理のアトミシティを保つために，あるトランザクションで発生したメモリアクセスが他のトランザクションで発生したメモリアクセスと競合するか検査される．競合が検出された場合はトランザクションをストールさせるが，複数のトランザクションにおいてストールが発生するとデッドロックとなる可能性があるため，い

1 名古屋工業大学
Nagoya Institute of Technology, Nagoya, Aichi 466–8555, Japan
†1 現在，株式会社デンソー
Presently with DENSO CORPORATION
a) tsumura@nitech.ac.jp

© 2012 Information Processing Society of Japan

ずれかのトランザクションをアボートさせる．

この際 LogTM では，possible_cycle と呼ばれるフラグを用いてアボート対象を選択しているが，この方法では starving writer と呼ばれるトランザクションが発生するような競合パターンにおいて性能が大きく低下してしまう．したがって，本稿ではこの starving writer の発生に着目し，これによる影響を抑制する手法を提案する．

2. 背景

本章では，**トランザクショナル・メモリ**（**Transactional Memory**，以下 **TM**）の基本概念および TM のハードウェア実装（HTM）の 1 つである LogTM について説明する．

2.1 TM の基本概念

TM における**トランザクション**は，クリティカルセクションを含む一連の命令列として定義され，以下の 2 つの性質を満たす．

シリアライザビリティ（直列可能性）： 並行実行されたトランザクションの実行結果は，当該トランザクションを直列に実行した場合と同じであり，すべてのスレッドにおいて同一の順序で観測される．

アトミシティ（不可分性）： トランザクションはその操作が完全に実行されるか，もしくはまったく実行されないかのいずれかでなければならず，各トランザクション内における操作はトランザクションの終了と同時に観測される．そのため，操作の途中経過が他のスレッドから観測されることはない．

以上の性質を保証するため，TM はトランザクション内のメモリアクセスを監視する．そして，あるトランザクション内でアクセスされたメモリアドレスと他のトランザクション内でアクセスされたメモリアドレスが同一であった場合，これを競合として検出する．競合を検出した場合は，片方のトランザクションの実行を中断する．これを**ストール**という．さらに，複数のトランザクションがストールした場合で，デッドロックが発生したと判断された場合，片方のトランザクションの実行結果を破棄する**アボート**を行う．そしてトランザクション開始時点の状態を復元し，トランザクションを再実行する．一方でトランザクションの終了まで競合が発生しない場合は，トランザクション内で実行された結果をメモリに反映させる**コミット**を行う．

TM はこのように動作することで，競合が発生しない限りトランザクションを並列に実行することができ，ロックを用いる場合よりもプログラムの並行性が向上する．また，プログラマはロックの粒度を考慮する必要がなくなり，容易に並列プログラムを構築できる．

2.2 HTM の設計方式

トランザクション内のアトミシティを保つために，TM ではトランザクション内で実行される命令におけるメモリアクセス競合の有無を検出する必要がある（conflict detection）．この競合検出は，そのタイミングによって以下の 2 つの方式に大別される．

Eager Conflict Detection： トランザクション内でメモリアクセスが発生した時点で，そのアクセスに関する競合が存在しないか検査する．

Lazy Conflict Detection： トランザクションがコミットしようとした時点で，そのトランザクション内で行われたアクセスに関して競合が発生していないか検査する．

また，TM におけるトランザクションの投機実行では，実行結果が破棄される可能性があるため，アクセスするデータのバージョン管理（version management）が必要である．これも以下の 2 つの方式に大別される．

Eager Version Management： 書き換え前の古い値を別領域にバックアップし，新しい値をメモリに上書きする．コミットはバックアップを破棄するだけなので高速に行えるが，アボート時にはバックアップされた値をメモリにリストアする必要がある．

Lazy Version Management： 書き換え前の古い値をメモリに残し，新しい値を別領域に登録する．アボートは高速に行えるが，コミット時にメモリへの値のコピーが必要となるため，オーバヘッドが大きくなる．

まず競合検出については，実際に競合が発生してからそれを検出するまでの時間が長くなる lazy 方式では，無駄な実行が増大してしまい効率が悪い．

一方バージョン管理については，eager 方式は，必ず実行されるコミットを高速に行い，必ずしも発生するとは限らないアボート時にコストを払う考え方である．アボートが繰り返し発生してしまうようなプログラムでは不利となる場合もあるが，lazy 方式ではコミットのためのオーバヘッドは削減の余地がほぼないのに対し，eager 方式ではスケジューリングを改善しアボートの発生自体を抑制することで性能を向上させることができる余地が大きいと考えられる．

よって本稿では，eager 方式どうしを組み合わせた実装の 1 つである LogTM をターゲットとする．

2.3 LogTM

本節では，HTM の一種であり本稿で提案する手法のターゲットとなる LogTM について述べる．

2.3.1 データのバージョン管理

我々がターゲットとする LogTM は，仮想メモリ領域を用いることで eager version management を実現している．LogTM は**ログ**と呼ばれる仮想メモリ領域をスレッドごとに割り当て，トランザクション内のストア命令によって上書きされる前の値とそのアドレスをこのログに退避させ

© 2012 Information Processing Society of Japan

る．一方でストア命令の結果はメモリに書き込まれる．

ここで投機実行が失敗した場合はアボートを行い，ログに保存されているすべての値をメモリに書き戻すことでトランザクション開始時点の状態を復元する．一方で，投機実行が成功した場合はコミット操作を行うが，すべての更新はすでにメモリに反映されているため，ログの走査や退避した値の書き戻しなどのメモリアクセスは必要なく，ログの内容を破棄するだけでよい．

2.3.2　競合検出

LogTM は上述のとおり eager version management を採用しているが，この競合検出のためにキャッシュライン上に新しく read ビットおよび write ビットを追加している．この read ビットと write ビットは，トランザクション内で当該キャッシュラインに対するリードアクセスまたはライトアクセスが発生した場合にそれぞれセットされ，トランザクションのコミットおよびアボート時にクリアされる．

LogTM は一貫性モデルにディレクトリベースの Illinois プロトコル [3] を採用し，これを拡張することでトランザクションを実行する他のスレッドとの競合を監視している．競合として検出されるのは以下の 3 パターンのアクセスが行われた場合である．

Read after Write（RaW）：　write ビットがセットされているアドレスに対するリードアクセス

Write after Read（WaR）：　read ビットがセットされているアドレスに対するライトアクセス

Write after Write（WaW）：　write ビットがセットされているアドレスに対するライトアクセス

たとえば，あるスレッドがリード/ライトアクセスを行う場合，トランザクション内の一貫性を保つために，アクセス対象となるラインに他のスレッドがすでにアクセス済みであるかどうかをディレクトリに対して問い合わせる．すでにアクセスされていた場合，コヒーレンスリクエストを当該スレッドに送信する．このリクエストを受信したスレッドは，どのメモリアドレスへのアクセスが行われようとしているのか知ることができ，当該キャッシュライン上の read ビットおよび write ビットを参照することで競合を検出することができる．競合が検出されなかった場合は，リクエスト送信者に対して ACK が返信される．一方で競合が検出された場合は NACK が返信される．NACK を受信したスレッドは競合の発生を知り，競合相手のトランザクションが終了するまで一時的に実行を停止する**ストール**を行う．ストールしているトランザクションは同じアドレスに対するリクエストを送信し続ける．競合相手のトランザクションが終了した場合，そのスレッドから ACK が返信されるため，ストールしているトランザクションは相手の終了を検知して実行を再開できる．

しかし図 **1** で示すように，複数のアドレスで競合が発生（時刻 t3 および t5）するとデッドロック状態に陥る場

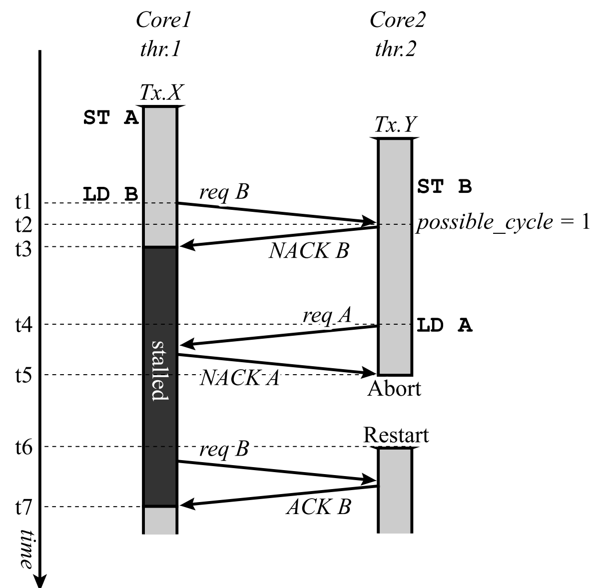

図 1　LogTM におけるトランザクションの競合解決
Fig. 1　Conflict resolution on LogTM.

合がある．この例では，2 つのスレッド *thr.1* と *thr.2* がそれぞれトランザクション *Tx.X* と *Tx.Y* を投機的に実行している．まず，*thr.1* が *Tx.X* の実行を開始した後に *thr.2* が *Tx.Y* の実行を開始する．そして先に *thr.1* が `ST A` を実行し，その後に *thr.2* が `ST B` を実行済みである場合を考える．その後 *thr.1* が `LD B` を実行しようとする際，*thr.1* は他のスレッドに対してアクセスリクエストを送信する（t1）．これを受信した *thr.2* は競合の発生を検知するため NACK を返信し，NACK を受信した *thr.1* はストールする（t3）．なお図中では省略しているが，*thr.1* はアクセスの許可を受けるまで定期的にリクエストを送信している．この後，*thr.2* が `LD A` を実行しようとする（t4）と，*thr.2* は *thr.1* と競合してお互いの実行を制止し合う結果となり，デッドロック状態に陥ってしまう．

そこで LogTM では，Transactional Lock Removal [4] の分散タイムスタンプに倣った方法を採用している．具体的には，まず図 1 の時刻 t2 で示すように，自身より早くトランザクションを開始したスレッドに NACK を返信すると **possible_cycle** と呼ばれるフラグをセットする．そして，このフラグがセットされている状態で，自身よりも早くトランザクションを開始したスレッドから NACK を受信すると，デッドロックを回避するためにアボートする（t5）．こうして，開始時刻の遅いトランザクションが，アボートの対象として選択される．*Tx.Y* をアボートした *thr.2* はトランザクション開始時の状態を復元し，トランザクションを再実行する（t6）．また，*thr.2* が *Tx.Y* をアボートしたため，*thr.1* は `B` にアクセスできるようになり，*Tx.X* のストール状態が解消される（t7）．

2.3.3　競合の抑制

LogTM では，競合の発生を抑制するために **exponential**

© 2012 Information Processing Society of Japan

backoff および **magic waiting** という手法が実装されている．Exponential backoff は，トランザクションをアボート後，再実行開始までの間，ある時間だけ待機する手法である．この待機時間については，アボートが発生するたびに指数関数的に増大させることで，競合の再発を抑制している．なお，この待機時間はコミット時に初期化される．

一方 magic waiting は，アボートした後，その競合相手がコミットするまで実行を再開せず待機し続けることで，完全に競合を防ぐ手法である．複数のスレッドと競合した場合には，相手から受信したリクエストまたは NACK から相手のトランザクション開始時刻を取得し，その時刻が一番遅いスレッドがトランザクションをコミットするまで待機する．なお，これらの手法を用いた場合，待機しているスレッドが遊休状態となるため，並列度が低下するという問題がある．

3. 関連研究

トランザクションの途中から再実行することにより，その地点までの再実行を省略する部分ロールバック [5] に関する研究や，適切なスレッド数を動的に設定する研究 [6] など数多くの LogTM に関する研究が行われている．前者の手法を改良した伊藤らの研究 [7] では，競合を起こした命令のプログラムカウンタの値を記憶し，再度そのプログラムカウンタに該当する命令を実行する際，その位置を再実行開始位置（チェックポイント，CP）として設定することで，再実行命令数の削減を可能にしている．また，既存の LogTM では CP の作成数に制限があったが，その制限をなくす手法も提案している．しかし，LogTM で標準的な possible_cycle flag を用いる方法ではなく，競合発生時に即座にトランザクションがアボートする，よりアボートが発生しやすいベースラインモデルに対する高速化で評価している点や，評価に対する考察が少なく，改良モデルのベースとしている Waliullah らの手法 [8] との比較評価もなされていないため，提案手法がどのように性能に寄与したのかが不明である点など，さまざまな問題がある．

一方後者のスレッド数を動的に制御する研究では，競合とトランザクション数の相関関係に着目し，動的にスレッド数を調整することでアボート回数を削減し，高速化を実現している．しかし，提案手法によって発生するオーバヘッドや，実装に必要なハードウェアコストについて評価していない．また，評価に用いたベンチマークプログラムが 1 つのみであり，効果の汎用性も明らかではない．

なお，部分ロールバックを適用した場合，実行再開後，競合しやすいアドレスにアクセスする箇所まで到達するのに要する時間が短縮されるため，LogTM のスケジューリングでは競合がより再発しやすくなる．一方，後者の研究のようにスレッド数を減少させなくても，スケジューリングの改良次第では並列度を高く保つことができる可能性もある．したがって，本稿では，まずはスレッドのスケジューリングに着目することで，従来の問題の解決を目指す．

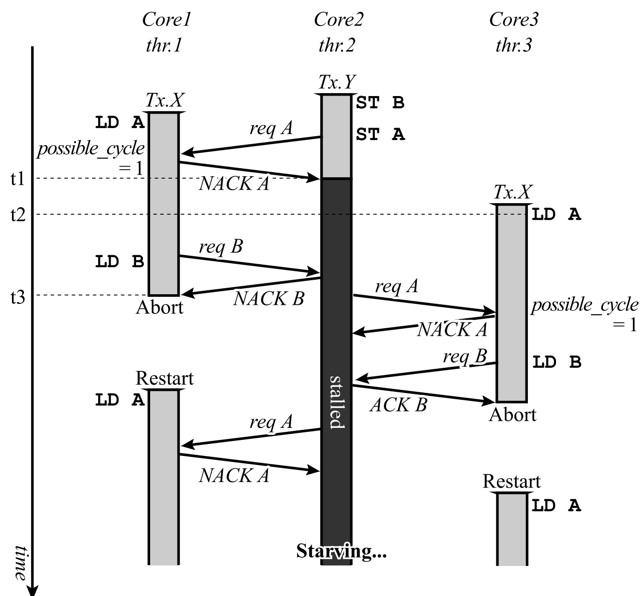

図 2　Starving writer の発生

Fig. 2　An example of a starving writer.

4. Starving Writer 解消手法

本章では，starving writer と呼ばれるトランザクションの発生が既存手法の性能に影響を与えることについて述べ，これを解決する提案手法について説明する．

4.1 既存手法の問題点

LogTM では，ある特定の競合パターンが発生すると著しく性能が悪化する場合がある．その競合パターンの 1 つに大きく関わっているのが **starving writer** [9] と呼ばれるトランザクションの発生である．これは，ストア（ST）を実行するトランザクション（writer）が，ロード（LD）を実行する複数のトランザクション（reader）により妨げられ続けることにより発生する．

いま図 **2** のように，3 つのスレッド（*thr.1*〜*3*）が，それぞれトランザクションを実行する例を考える．なお，*thr.1* および *thr.3* は同じトランザクション *Tx.X* を実行し，*thr.2* は *Tx.X* で LD されるアドレス `A` に対する ST を含む *Tx.Y* を実行するとする．まず *thr.1* が `LD A` を実行済みの状態で，*thr.2* が `ST A` を実行しようとして競合が発生し，*Tx.Y* はストールする（時刻 t1）．この場合，*thr.1* が *Tx.X* をコミットもしくはアボートしない限り *thr.2* は `ST A` を実行できない．次に，*thr.3* が `LD A` を実行しようとするが（t2），これは 2.3.2 項で述べたいずれのアクセスパターンにも該当せず，競合は検出されない．これにより，*thr.1* および *thr.3* が実行中の 2 つの *Tx.X* が，ともにアボートもしくはコミットしない限り *thr.2* は再開することができない状態となる．この後，*thr.1* が *thr.2* と競合して *thr.1* の *Tx.X*

© 2012 Information Processing Society of Japan

がアボートされた場合（t3）でも，*thr.3* がすでにアドレス A にアクセスしているため，*thr.2* は実行を再開できない．そして *thr.1* は *Tx.X* の再実行後，再度 A にアクセスしてしまう．このように，同一アドレスに対する LD を実行するスレッドが複数存在することで，当該アドレスに対する ST を実行しようとしているスレッドが飢餓状態（starving writer）となる．実際には，さらに多くのスレッドが LD を実行している場合が多く，それらすべてのスレッドがトランザクションをアボートあるいはコミットしてリソースを解放しない限り，ST を実行しようとしているスレッドは再開することができない．

この競合パターンは 2.3.3 項で述べた exponential backoff または magic waiting によって対処可能である．しかし，exponential backoff ではアボートしたトランザクションが一時的にしか待機しないため，starving writer が解決されるまでに何度もアボートを繰り返してしまう．これを避けるために backoff 待機時間の増大率を大きくすることも考えられるが，starving writer に関与している reader 数に応じた増大率を設定しなければ，無駄な待ち時間が発生し性能が著しく低下してしまうおそれがある．また，これら reader はアボートなどによりつねにその数が変動するため，reader 数を把握し続けることも一般に容易ではない．一方 magic waiting では，アボートを繰り返していない場合にも待機し続けてしまい並列度が低下してしまう．

そこで本稿では，starving writer によってアボートが繰り返される場合に，アボートしたトランザクションを一時停止させ，starving writer を解消する手法を提案する．

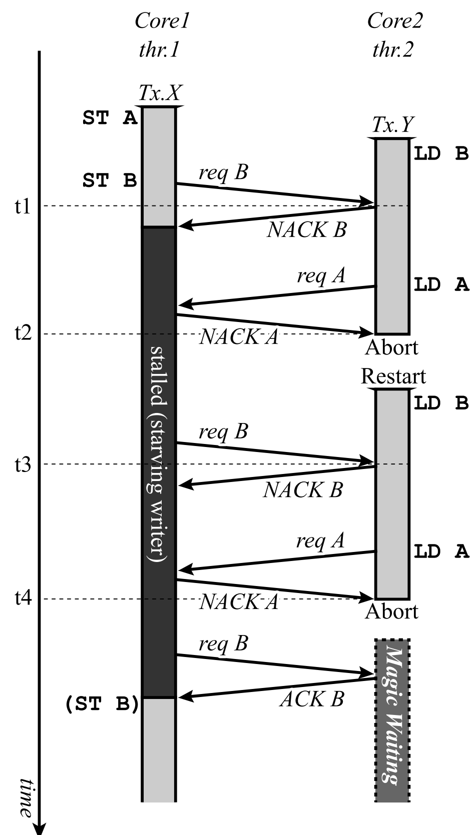

図 3　Starving writer 発生時の動作モデル
Fig. 3　Proposed model with a starving writer.

4.2　提案モデルとその動作

前節で述べた starving writer の発生を抑制するために，reader が writer を競合相手としてある条件を満たすようなアボートを繰り返す傾向を見せた場合，その reader の実行に magic waiting を適用し，相手 writer を優先的にコミットさせる手法を提案する．

さて，starving writer は WaR 競合パターンが存在する場合に発生する．また，2.3.2 項で述べたように，アボート処理までに少なくとも 2 つのキャッシュラインで競合が発生する．これらをふまえ本節では，starving writer が発生したと判断する条件セットを 3 種提案し，それぞれを用いた場合の動作モデルについて説明する．

モデル 1：同一 writer との競合アドレスの組が一致

あるトランザクションが以下に示す 2 つの条件をともに満たす場合，競合相手を starving writer であると判定し，自身に magic waiting を適用することでその相手 writer を優先的にコミットさせる．

- 条件 I：自身が LD 済みのアドレスに対して他スレッドが ST 要求を発行することにより WaR 競合が発生
- 条件 II：同一の writer との間の競合によって発生した，直近の過去 2 回のアボートに関与したアドレスの組が一致

なお，アボートが発生するためには，2.3.2 項で見たように possible_cycle フラグをセットする原因になった競合と，アボートの直接的な引き金となった競合の 2 つのアドレス競合が必要であるが，条件 II の「アボートに関与したアドレスの組」とは，これら 2 つの競合それぞれにおける対象アドレスの組を指す．

これら 2 つの条件について図 **3** の starving writer が発生する場合の例を用いて説明する．例ではまず *thr.2*（reader）が LD B を実行し，次に *thr.1*（writer）が ST B を実行しようとして競合が発生する（時刻 t1）．これは WaR 競合であるため，条件 I を満たす．その後 *thr.2* が実行する *Tx.Y* は，*thr.1* によって 2 度アボート（t2 および t3）させられており，アボートに関与したアドレスの組合せがともに {B, A} となっている．したがって条件 II を満たす．このように両方の条件を満たした場合，ST を実行しようとしていたトランザクションを starving writer であると見なし，LD を実行していたスレッドに対して 2.3.3 項で述べた magic waiting を有効にすることで，starving writer となっていたトランザクションを優先的に実行させる．

モデル 2：同一 writer との競合アドレスが一致

ある writer トランザクション *Tx.W* が，ある reader トランザクション *Tx.R* にストールさせられて starving 状態

© 2012 Information Processing Society of Japan

にあるとき，*Tx.W* は終始ストールし続けているとは限らず，他の第3者のトランザクションとの競合によりアボートさせられてしまう場合もある．この場合 *Tx.W* は再実行されるが，制御フローの変化などにより，*Tx.R* との再競合の際に，possible_cycle フラグをセットする原因となる競合アドレスが前回とは異なる場合もありうる．このような場合も解決するために2つめのモデルとして，モデル1の条件IIを以下のように緩和したものを考える．

- 条件II′：同一の相手による直近の過去2回のアボートにおいて，そのアボートに直接関与するアクセス対象アドレスが同一

すなわち，possible_cycle フラグをセットする原因となったアクセス対象アドレスの一致を必要としないよう，条件IIを変更する．これと，モデル1の条件Iを併用することで，starving writer を解消する．

図3の例では，*Tx.Y* のアボートはともにアドレス `A` へのアクセス（t2 および t3）を直接的な原因として発生しているため条件II′に該当し，*thr.2* に magic waiting が適用される．

モデル3：任意 writer との競合アドレスが一致

同じアドレスに対する WaR 競合は，異なる複数の writer との間で発生する可能性は少ないと考えられる．したがってモデル2に対し，競合相手を考慮しないように条件を簡略化した拡張モデルを考える．これを実現するため，モデル2の条件II′において，競合相手に関する部分を次のように緩和する．

- 条件II″：競合相手を問わず，直近の過去2回のアボートにおいて，そのアボートに直接関与するアクセス対象アドレスが同一

なおモデル2と同様に，モデル1の条件Iを併用する．

5. 競合の再発検知機構

本章では，前章で述べた提案モデルを実現するにあたり必要なハードウェアおよびその動作モデルについて述べる．

5.1 ハードウェア拡張

4.2節で述べた提案モデルを実現するため，既存の LogTM を拡張して以下に示す3つの機構を各コアに追加する．

WaR flags： 競合パターン WaR 発生の有無を記憶する．自スレッドがすでに LD を実行済みであるアドレスに対し，他スレッドが ST を実行しようとした際の競合発生時にセットする．総スレッド数 n に対して n bit 必要であり，アボートおよびコミット時にクリアされる．

Conflict Table（C-Tbl）： スレッド番号をインデクスとし，そのインデクスに対応するスレッドとの間に起こった直近の競合における，対象アドレスを記憶する表．競合発生時に参照され，現競合の対象アドレスと比較される．アドレスが一致した場合は後述の M-W flags の状態を更新し，一致しない場合は現競合アドレスでエントリを上書きする．提案モデル1は2つのアドレスを条件判定に利用するため，深さ n のこのテーブルを各スレッドに2つ（C-Tbl1，C-Tbl2）保持させ，競合したアドレスに対して先にアクセスしたのが自スレッドである場合は C-Tbl1，競合相手スレッドである場合は C-Tbl2 を用いる．なお，テーブル内の情報はコミット時のみクリアされる．提案モデル2では C-Tbl は1つでよく，提案モデル3では相手スレッドを区別しないため C-Tbl は1つかつ深さ1でよい．

Magic Waiting flags（M-W flags）： Magic waiting を有効にするかどうかを示す $2n$ bit からなるビット列で，各スレッドに対して2ビットずつ使用する．C-Tbl1 で比較したアドレスが一致した場合は1ビット目，C-Tbl2 では2ビット目のビットをセットし，アボート時にこれら2つのビットが両方セットされている場合，magic waiting を有効にする．コミットおよびアボート時にクリアされる．なお提案モデル2および3では，M-W flags は各スレッドに 1 bit でよい．

n スレッドを実行可能な n コア構成のプロセッサの場合，1コアあたりに必要となる WaR flags は n bit，また M-W flags は $2n$ bit であり，あわせて $3n$ bit と少量である．また C-Tbl については，幅 64 bit 深さ n 行の RAM で構成でき，たとえば $n = 32$ では C-Tbl サイズの総和は 16 kB と少量である．

また，提案モデル2の場合は4.2節で述べたように，1つのアドレスのみを条件に利用するため，C-Tbl は1つ，M-W flags は 1 bit となる．したがってハードウェアコストは提案モデル1の約半分となる．そして提案モデル3の場合は，4.2節で述べたように，競合相手ごとにアドレスを管理しないため，C-Tbl サイズの総和は 256 B と，ごく小さいものとなる．

5.2 動作モデル

図4の starving writer が発生する場合の例を用いて，*thr.2* における提案モデル1のハードウェア動作を説明する．まず，*thr.2* が `LD B` を実行し，その後に *thr.1* が `ST B` を実行しようとした場合，WaR パターンの競合が検出される（時刻 t1）．したがって *thr.2* では，競合相手のスレッド *thr.1* に対応する WaR flags がセットされ，スレッド番号1をインデクスとして C-Tbl が参照される．なお，今回競合が発生したアドレス `B` に先にアクセスしたのは *thr.2* であるため，C-Tbl1 が参照される．ここでは，C-Tbl1[1] にはアドレスが未登録であるため，`B` が登録される．

次に，*thr.2* が `LD A` を実行し，競合が発生すると（t2），アドレス `A` へ先にアクセスしたのは *thr.1* であるため，先ほどとは別のテーブルである C-Tbl2 が参照される．ここ

© 2012 Information Processing Society of Japan

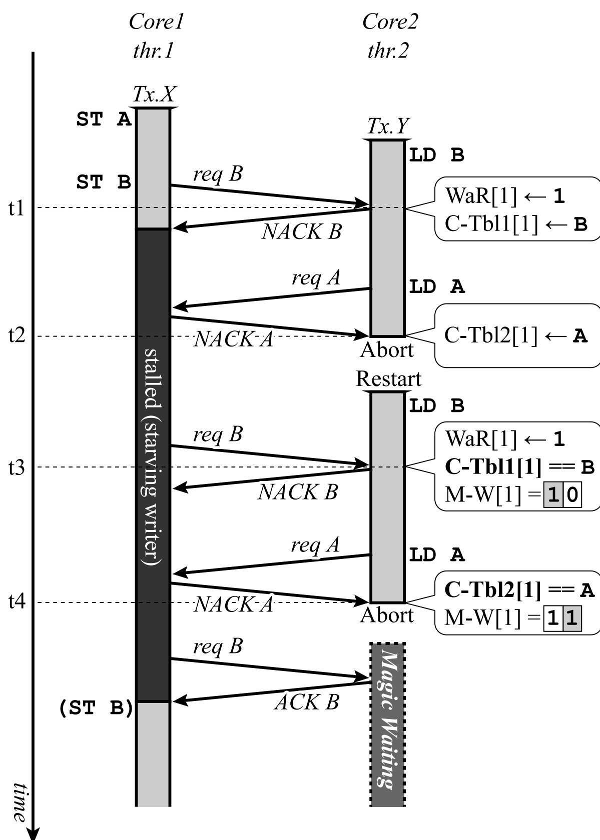

図 4　追加したハードウェアの状態変移
Fig. 4　Behavior of additional hardware units.

で *thr.1* に対応するアドレスは C-Tbl2 にはまだ登録されていないため，A が登録される．そしてデッドロックの発生を検知したことにより，*thr.2* は *Tx.Y* をアボートする．なお，アボート後はすべての競合が解決されるため，WaR flags はリセットされる．

続いて *thr.2* が *Tx.Y* を再実行し，アドレス B で競合すると（t3），時刻 t1 のときと同様に WaR flags がセットされ，B がテーブルに登録されているか参照される．今回はすでに同一のアドレスが登録されているため，M-W flags の左ビットをセットし，結果 10 となる．次に *thr.2* が LD A を実行し，競合が発生すると（t4），時刻 t3 のときと同様に C-Tbl2 が参照される．そして，登録済みのアドレスと今回競合したアドレス A とが一致するため，M-W flags の右ビットをセットする．この結果 M-W flags は 11 と両ビットがセットされている状態になるため，magic waiting が有効となる．

一方提案モデル 2 では，提案モデル 1 と比べ C-Tbl が 1 つ少なく済み，C-Tbl1 への B の登録および参照の処理が省略される．また，M-W flags も 1 ビットとなっており，時刻 t3 における M-W flags に対する処理も省略される．最後に，提案モデル 3 は提案モデル 2 と比べて C-Tbl の構造が異なるが，2 者間によるアボートの繰返し時には提案モデル 2 と同じ動作となる．しかし，異なるスレッドに対して同一のアドレスで競合を繰り返した場合，他の提案手法よりも早く magic waiting が有効となる場合があり，より競合を抑制することが期待できる．このような状況は，ある特定の writer トランザクションを複数のスレッドが並列に実行する場合に考えられる．

表 1　シミュレータ諸元
Table 1　Simulation parameters.

Processor	SPARC V9
number of cores	32 cores
frequency	1 GHz
issue width	single-issue
issue order	in-order
non-memory IPC	1
D1 cache	32 KBytes
ways	4 ways
latency	1 cycle
D2 cache	8 MBytes
ways	8 ways
latency	20 cycles
Memory	4 GBytes
latency	450 cycles
Interconnect network latency	14 cycles

6. 評価

前章で述べた拡張を既存の LogTM に実装し，シミュレーションによる評価を行った．本章では，その評価結果を示し考察する．また，LogTM の他の競合解決手法との比較も行う．

6.1 評価環境

評価には TM の研究で広く用いられている Simics [10] 3.0.31 と GEMS [11] 2.1.1 の組合せを用いた．Simics は機能シミュレーションを行うフルシステムシミュレータであり，GEMS はメモリシステムの詳細なタイミングシミュレーションを担う．プロセッサは 32 コアの SPARC V9 とし，OS は Solaris10 とした．**表 1** に詳細なシミュレーション環境を示す．評価対象のプログラムとしては GEMS 付属 microbench，SPLASH-2 [12]，および STAMP [13] から計 12 個を使用した．各プログラムに与えた入力パラメータを**表 2** に示す．

なお，各コアが 1 スレッドを実行し，プロセッサ全体で 32 スレッドを実行するが，OS 用に 1 コアを使用するとし，31 スレッドによる評価を行った．ただし，STAMP は 2 の冪乗数のスレッドでしか動作しないため，STAMP に限り 16 スレッドで評価した．また，3 章で述べたように，本稿ではまずスレッドのスケジューリングに着目するため，部分ロールバック手法を用いていない．したがって，GEMS 付属の部分ロールバック用テストプログラムである partial rollback は評価対象から除外した．

© 2012 Information Processing Society of Japan

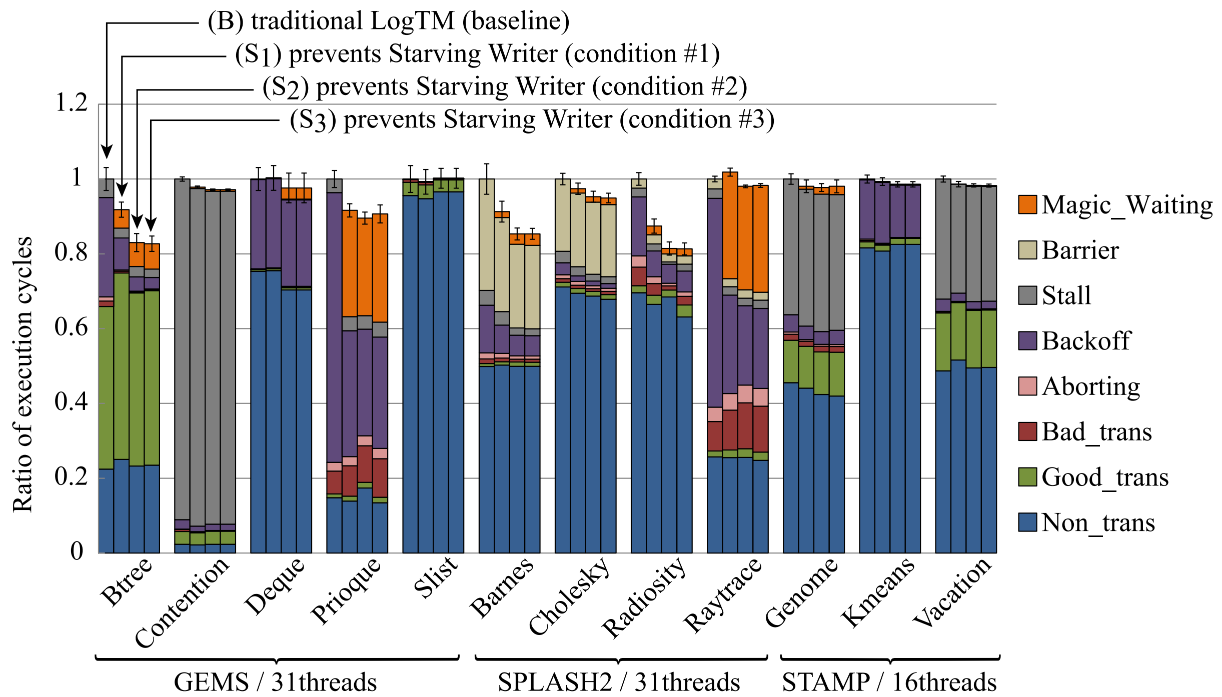

図 5　実行サイクル数比（GEMS，SPLASH-2，STAMP ベンチマーク）

Fig. 5　Ratio of execution cycles w/ GEMS, SPLASH-2 and STAMP benchmark suites.

表 2　ベンチマークプログラムの入力パラメータ

Table 2　Input parameters.

GEMS	
Btree	priv-alloc-20pct
Contention	config 1
Deque	1024ops 32bkoff
Prioque	8192ops
Slist	500ops 64len
SPLASH-2	
Barnes	512BODIES
Cholesky	tk14.0
Radiosity	-p 31 -batch
Raytrace	teapot
STAMP	
Genome	-g256 -s16 -n16384 -t16
Kmeans	-m40 -n40 -t0.05 -p16
Vacation	-n2 -q90 -u98 -r16384 -t4096 -c16

表 3　実行サイクル数の削減率

Table 3　Reduced execution cycles.

		GEMS	SPLASH-2	STAMP	all
(S_1)	平均	3.9%	5.7%	1.3%	3.9%
	最大	8.4%	12.6%	1.9%	12.6%
(S_2)	平均	6.7%	10.2%	1.8%	6.7%
	最大	17.0%	18.6%	2.3%	18.6%
(S_3)	平均	6.6%	10.3%	1.7%	6.6%
	最大	17.3%	18.7%	1.9%	18.7%

表 4　アボート発生回数の削減率

Table 4　Reduced aborts.

		GEMS	SPLASH-2	STAMP	all
(S_1)	平均	37.1%	25.5%	40.0%	34.2%
	最大	76.2%	45.7%	67.7%	76.2%
(S_2)	平均	46.6%	44.7%	47.9%	46.3%
	最大	86.8%	67.1%	72.9%	86.8%
(S_3)	平均	46.1%	45.4%	47.6%	46.3%
	最大	86.6%	67.4%	72.9%	86.6%

6.2　評価結果

図 5 および**表 3** に実行サイクル数比，**表 4** にアボート回数の削減率を示す．図 5 のグラフは左から順にそれぞれ

(B)　既存の LogTM．

(S_1)　提案モデル 1：同一相手との WaR 競合による直近 2 回のアボートにおいて，そのアボートに関わったアドレスの組が一致した場合に magic waiting を有効にするモデル．

(S_2)　提案モデル 2：同一相手との WaR 競合による直近 2 回のアボートにおいて，そのアボートの直接の原因となったアドレスが一致した場合に magic waiting を有効にするモデル．

(S_3)　提案モデル 3：競合相手を問わず，直近 2 回の WaR によるアボートにおいて，そのアボートの直接の原因となったアドレスが一致した場合に magic waiting を有効にするモデル．

の実行サイクル数比を表しており，既存手法 (B) の実行サイクル数を 1 として正規化している．(S_1)～(S_3) はそれぞれ，4 章で述べた 3 つのモデルに対応している．また，凡例は内訳を示しており，Non_trans はトランザクショ

© 2012 Information Processing Society of Japan

ン外，Good_trans，Bad_trans はそれぞれ結果的にコミット/アボートされたトランザクション内，Aborting，Stall，MagicWaiting，Barrier，Backoff はそれぞれ，アボート，ストール，magic waiting，バリア同期，exponential backoff に要したサイクル数である．

なお，フルシステムシミュレータ上でマルチスレッドを用いた動作のシミュレーションを行うには，性能のばらつきを考慮しなければならない [14]．したがって，各評価対象につき試行を 10 回繰り返し，得られた結果から 95%の信頼区間も求めた．信頼区間はグラフ中にエラーバーで表している．

実行サイクル数に関しては，評価に使用したベンチマークプログラムの多くは，本提案手法が解決すべき対象とした競合の再発，およびそれにともなうアボートの頻発を含んでいたため，提案手法によりこれを解決することで性能が向上した．ただし，Slist に関しては，競合の繰返しがほとんど発生しないプログラムであるため，既存モデルとほぼ同等の結果となっている．一方アボートの発生回数に関しては，表 4 から分かるように大きく削減できており，提案手法が非常に有効に働いていることが分かる．

プログラムを個別に見ると，まず Contention，Deque，Genome，Kmeans，Vacation では，ほぼすべての手法で提案手法によりわずかに高速化している．これは主にアボートの抑制によるもので，アボート回数は既存手法 (B) に対し最大 72.9% (Kmeans)，最低でも 15.1% (Deque) 削減されている．また，全実行サイクルに占める magic waiting の割合は，たとえば Kmeans では 0.1%以下となっており，本提案によって新たに加えられた待機処理が短時間で済んでいることが分かる．しかしこれらのプログラムでは，元来アボートが実行サイクルに与える影響は小さかったため，高速化率は小さくなっている．

次に，Btree，Prioque，Barnes，Radiosity については，アボート回数の削減による Bad_trans や Aborting サイクルの減少，競合自体の削減による Stall サイクルの減少，アボートの繰返しを抑制したことによる Backoff サイクルの減少などにより大きく高速化しており，提案手法の有効性が確認できた．なおこれらのうち Prioque については Magic_Waiting が比較的大きいことから，writer が早期にコミットできたというよりむしろ，magic waiting によって exponential backoff よりも適切な待ち時間が reader に設定された結果であると考えられる．しかし残りの 3 プログラムについては Magic_Waiting はあまり多くを占めておらず，writer が従来手法より早期にコミットに至ったことの効果が大きいと考えられる．これら 3 プログラムについて，ストールサイクル数の中でも特に，writer がストアを実行できずにストールさせられているサイクル数が (B) と (S_3) でどう異なっているかを調査した結果，**表 5** に示すように (S_3) では (B) に対し大きく削減できており，本提案手法の有効性が確認できた．

表 5　ストア待ちストールのサイクル数の平均値
Table 5　Average stall cycles for stores.

	(B)	(S_3)	reduced
Btree	248,269	110,422	55.5%
Barnes	252,265	119,949	52.5%
Radiosity	239,285	173,493	27.5%

なお，これらの中でも Btree は最も starving writer の影響を受けるプログラムであるが，starving writer 発生時のアボート抑制および Backoff 削減も高速化に寄与している．Btree において，手法 (S_2) および (S_3) のアボート回数を調査したところ，既存手法に対し 86.8%も削減できていることが確認できた．また，最長のアボート繰返し回数についても，約 1/4 程度に削減されていた．

また，Barnes の場合，提案モデル (S_2)，(S_3) は，既存モデル (B) に対し Barrier を約 25%削減している．これは，既存モデルでは特定のトランザクションがアボートの繰返しにより遅延することで，バリア同期において他のトランザクションを長期間待機させてしまっていた状況を，提案モデルで解決できたためであると考えられる．

提案手法による効果が最も大きかった Radiosity については，提案手法により半数以上のスレッドが一時的に magic waiting を有効にする状況が見られた．これはすなわち，非常に競合を起こしやすい特徴を持つスレッドが存在し，提案手法によりそのスレッドをコミットまで優先的に進行させることで，既存モデルで発生していたアボートの頻発を抑制したと考えられる．

一方，Prioque や Raytrace に見られる傾向として，アボート回数は減少しているものの Bad_trans サイクルが増加してしまっていることがあげられる．既存手法ではトランザクション開始直後にアボートする状況が頻発しており，個々のアボートで計上される Bad_trans も小さいものであったが，提案手法では，トランザクション中のより多くの命令を実行した後にアボートする場合があり，アボート回数は少ないものの個々のアボートで計上される Bad_trans サイクルが大きくなったためであると考えられる．

結果を総合すると，手法 (S_2)，(S_3) は (S_1) よりも高い性能を示しており，possible_cycle フラグをセットする原因となった競合アドレスを考慮せず，同アドレス競合によるアボートの繰返しを抑制することが重要であることが分かった．また，(S_2) と (S_3) には有意な差はなく，競合相手スレッド別に競合アドレスを記憶する必要性は低いと考えられることから，ハードウェアコストも軽量である手法 (S_3) が最も優れているといえる．

6.3　他の競合解決方式との比較

LogTM で用いられる代表的な競合解決手法は，2.3.2 項

© 2012 Information Processing Society of Japan

表 6　アボートおよびストールに関わるサイクル数の比較
Table 6　Execution cycles for aborts and stalls normalized to the total cycles of (B).

benchmark	program	(B)	(T)	(S_3)
GEMS	Btree	34.1%	9.7%	5.8%
	Contention	94.2%	193.9%	90.9%
	Deque	24.2%	17.2%	23.7%
	Prioque	84.2%	75.7%	46.8%
	Slist	0.9%	19.7%	0.4%
SPLASH-2	Barnes	19.5%	9.6%	9.0%
	Cholesky	8.3%	1.9%	4.8%
	Radiosity	26.1%	5.4%	10.9%
	Raytrace	70.1%	70.7%	40.6%
STAMP	Genome	43.1%	234.4%	42.2%
	Kmeans	16.8%	7.2%	14.4%
	Vacation	35.8%	425.9%	33.1%

で述べたように possible_cycle フラグを用いてデッドロックを検出するものであり，本稿の提案手法もこれをベースとしている．一方，他の選択肢として，possible_cycle フラグを用いず，キャッシュリクエスト送信側トランザクションの開始時刻（タイムスタンプ）が受信側トランザクションの開始時刻よりも早かった場合に，即座に受信側トランザクションをアボートさせる方法がある．これを timestamp 方式と呼ぶ．

この方式は，本稿で提案した手法と同様にトランザクションの starvation に着目し，それを抑制する意図を持ったものであるが，必要以上にアボートが発生してしまうことで逆に性能に悪影響を及ぼす場合も多いと考えられる．

そこでまず，timestamp 方式のモデル（以下 **(T)** とする）について総実行サイクル数を調査し，possible_cycle フラグを用いる既存の LogTM モデル (B) と比較したところ，一部性能向上するプログラムも存在したものの，(B) に対して全プログラムの幾何平均で約 14.8%，算術平均で約 43.9%の性能低下を引き起こすことを確認した．

次にこの理由を調査するため，全体性能に悪影響を及ぼす要素であるアボートおよびストールに関わるサイクル数が，(B) と (T) でどのように異なっているかを調査した．結果を**表 6** に示す．ここではまず，アボートに関わる Bad_trans，Aborting，Backoff とストールに関わる Stall の総和を，性能に悪影響を及ぼすサイクル数であると定義し，(B) においてこのサイクル数が総実行サイクル数に占める割合を表の (B) の列に示している．次に，timestamp 方式によって各ベンチマークプログラムを実行した結果から，上記 4 項目に要したサイクル数を (B) の総実行サイクル数で正規化した値を，(T) の列に示している．同様に，提案モデル 3 の結果も (S_3) の列に示した．

まず Btree, Barnes, Radiosity においては，(T) は (B) に対し，これらのサイクル数が (B) の総実行サイクル数比で約 10%以上減少しており，timestamp 方式が possible_cycle 方式よりも有利に働くプログラムが存在することが分かる．一方で，Contention で約 2 倍，Genome で約 5 倍，Vacation においては約 12 倍にまで膨れ上がってしまっており，これが原因でこれらは総実行サイクル数でも (B) に対し，それぞれ約 2 倍，3 倍，4.5 倍を要していた．このように timestamp 方式では性能が著しく低下してしまうプログラムが存在することも分かった．

一方，提案モデル (S_3) では，(T) に有利となっていた Btree，Barnes，Radiosity で (T) と同等もしくはそれ以上の削減が実現できていることが分かる．また，(T) では大きく悪化してしまっていた Contention，Genome，Vacation においても，わずかながら (B) よりも改善されていることが分かる．さらに，(T) ではあまり改善できていない Prioque や Raytrace においても (S_3) は大きく改善できており，提案モデルの優位性が確認できた．

7. おわりに

本稿では LogTM を拡張し，starving writer に対処するための競合抑制手法および競合の再発抑制手法を提案した．

シミュレーションにより GEMS 付属 microbench，SPLASH-2，および STAMP ベンチマークを用いて評価した結果，提案手法は競合再発に起因するアボートの繰返しを抑制することで，結果的にアボートによって破棄されてしまう実行サイクルや，再実行までの backoff サイクルを削減することを確認した．その結果，既存の LogTM に比べてアボート発生回数を最大 86.6%削減することに成功し，実行サイクル数でも最大で 18.7%，平均で 6.6%の高速化を実現した．

なお，本稿では競合パターンの 1 つである starving writer の影響に着目したが，LogTM には著しく性能を低下させてしまう競合パターンが他にも複数存在する．特に，今回の評価結果からストールサイクルや backoff サイクルの占める割合が多いプログラムが存在していることから，今後これらの要因を調査し，対処方法を検討していきたい．また，magic waiting やストール時における遊休状態コアを有効活用する手法を検討することも今後の課題である．

参考文献

[1] Herlihy, M. and Moss, J.E.B.: Transactional Memory: Architectural Support for Lock-Free Data Structures, *Proc. 20th Annual Int'l Symp. on Computer Architecture*, pp.289–300 (1993).

[2] Moore, K.E., Bobba, J., Moravan, M.J., Hill, M.D. and Wood, D.A.: LogTM: Log-based Transactional Memory, *Proc. 12th Int'l Symp. on High-Performance Computer Architecture*, pp.254–265 (2006).

[3] Censier, L.M. and Feautrier, P.: A New Solution to Coherence Problems in Multicache Systems, *IEEE Trans. Comput.*, Vol.C-27, No.12, pp.1112–1118 (1978).

© 2012 Information Processing Society of Japan

[4] Rajwar, R. and Goodman, J.R.: Transactional Lock-Free Execution of Lock-Based Programs, *Proc. 10th Symp. on Architectural Support for Programming Languages and Operating Systems*, pp.5–17 (2002).

[5] Moravan, M.J., Bobba, J., Moore, K.E., Yen, L., Hill, M.D., Liblit, B., Swift, M.M. and Wood, D.A.: Supporting Nested Transactional Memory in LogTM, *Proc. 12th Int'l Conf. on Architectural Support for Programming Languages and Operating Systems* (*ASPLOS*), pp.1–12 (2006).

[6] 武田　進，島崎慶太，井上弘士，村上和彰：トランザクショナルメモリにおける並列実行トランザクション数動的制御法の提案とその評価，信学技報，Vol.108, No.ICD-28, pp.81–86 (2008).

[7] 伊藤悠二，塩谷亮太，五島正裕，坂井修一：最適なロールバック・ポイントを選択するトランザクショナル・メモリ，先進的計算基盤システムシンポジウム SACSIS2011 論文集，pp.324–331 (2011).

[8] Waliullah, M.M. and Stenstrom, P.: Intermediate Checkpointing with Conflicting Access Prediction in Transactional Memory Systems, *Proc. Int'l Symp. on Parallel and Distributed Processing* (*IPDPS*), pp.1–11 (2008).

[9] Bobba, J., Moore, K.E., Volos, H., Yen, L., Hill, M.D., Swift, M.M. and Wood, D.A.: Performance Pathologies in Hardware Transactional Memory, *Proc. 34th Annual Int'l Symp. on Computer Architecture* (*ISCA'07*), pp.81–91 (2007).

[10] Magnusson, P.S., Christensson, M., Eskilson, J., Forsgren, D., Hållberg, G., Högberg, J., Larsson, F., Moestedt, A. and Werner, B.: Simics: A Full System Simulation Platform, *Computer*, Vol.35, No.2, pp.50–58 (2002).

[11] Martin, M.M.K., Sorin, D.J., Beckmann, B.M., Marty, M.R., Xu, M., Alameldeen, A.R., Moore, K.E., Hill, M.D. and Wood, D.A.: Multifacet's General Execution-driven Multiprocessor Simulator (GEMS) Toolset, *ACM SIGARCH Computer Architecture News*, Vol.33, No.4, pp.92–99 (2005).

[12] Woo, S.C., Ohara, M., Torrie, E., Singh, J.P. and Gupta, A.: The SPLASH-2 Programs: Characterization and Methodological Considerations, *Proc. 22nd Annual Int'l. Symp. on Computer Architecture* (*ISCA'95*), pp.24–36 (1995).

[13] Minh, C.C., Chung, J., Kozyrakis, C. and Olukotun, K.: STAMP: Stanford Transactional Applications for Multi-Processing, *Proc. IEEE Int'l Symp. on Workload Characterization* (*IISWC'08*) (2008).

[14] Alameldeen, A.R. and Wood, D.A.: Variability in Architectural Simulations of Multi-Threaded Workloads, *Proc. 9th Int'l Symp. on High-Performance Computer Architecture* (*HPCA'03*), pp.7–18 (2003).

江藤 正通　（学生会員）

1987 年生．2011 年名古屋工業大学工学部情報工学科卒業．現在，同大学大学院工学研究科創成シミュレーション工学専攻博士前期課程在籍．計算機アーキテクチャ，並列処理等に興味を持つ．

堀場 匠一朗　（学生会員）

1989 年生．2012 年名古屋工業大学工学部情報工学科卒業．現在，同大学大学院工学研究科創成シミュレーション工学専攻博士前期課程在籍．計算機アーキテクチャ，マルチコア・プロセッサ等に興味を持つ．

浅井 宏樹

1988 年生．2010 年名古屋工業大学工学部情報工学科卒業．2012 年同大学大学院工学研究科創成シミュレーション工学専攻博士前期課程修了．同年（株）デンソー入社．計算機アーキテクチャ，並列処理等に興味を持つ．

津邑 公暁　（正会員）

1973 年生．1996 年京都大学工学部情報工学科卒業．1998 年同大学大学院工学研究科情報工学専攻修士課程修了．2001 年同大学院情報学研究科博士後期課程学修認定退学．同年同大学院経済学研究科助手．2004 年豊橋技術科学大学工学部助手．2006 年名古屋工業大学大学院工学研究科助教授．2007 年同准教授．博士（情報学）．プロセッサアーキテクチャ，並列処理，脳型情報処理等に関する研究に従事．ACM，IEEE-CS 各会員．

松尾 啓志　（正会員）

1960 年生．1985 年名古屋工業大学大学院修士課程修了．1989 年同大学院博士後期課程修了．同年同大学工学部電気情報工学科助手．1993 年同講師．1995 年同助教授．2003 年同教授．2004 年同大学工学部情報工学科教授．工学博士．計算機工学，分散協調システムに関する研究に従事．IEEE，人工知能学会，電子情報通信学会各会員．

© 2012 Information Processing Society of Japan