



|              |                                                                             |
|--------------|-----------------------------------------------------------------------------|
| Title        | 大規模分散共有メモリシステムの一貫性制御に関する研究                                                  |
| Author(s)    | 細見, 岳生                                                                      |
| Citation     | 大阪大学, 2022, 博士論文                                                            |
| Version Type | VoR                                                                         |
| URL          | <a href="https://doi.org/10.18910/88156">https://doi.org/10.18910/88156</a> |
| rights       |                                                                             |
| Note         |                                                                             |

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                            |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|
| 氏 名 （ 細見 岳生 ）                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                            |
| 論文題名                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 大規模分散共有メモリシステムの一貫性制御に関する研究 |
| <p>論文内容の要旨</p> <p>高性能な計算機システムの構築において、キャッシュを持ったプロセッサとメモリからなるノードをネットワークで接続した大規模分散共有メモリシステムが注目を集めている。大規模分散共有メモリシステムにおいては、ディレクトリ方式の一貫性制御が用いられる。ディレクトリ方式の一貫性制御の特徴は、メモリにおいてどのノードがデータのコピーを保持しているかを管理するディレクトリをメモリに保持することにある。キャッシュミスが発生すると当該データをメモリに保持しているノードに要求を送り、その要求を受けたノードのメモリはディレクトリを基にそのデータのコピーをキャッシュに保持するノードに要求を転送する処理を行う。</p> <p>大規模分散共有メモリシステムの性能を左右するメモリアクセスレイテンシは次の三点に影響を受ける。一つ目は、あるノードのプロセッサが行った書込みを他のプロセッサに反映するプロトコルとして、Invalidate型やUpdate型等のようなプロトコルを採用するかである。二つ目は、書込み時にはデータを共有しているノード全てに対して要求を伝える必要があるため、その場合の一貫性制御方式である。三つ目は、ノード間の要求やその応答をネットワークを介して送信する必要があるため、ネットワークのレイテンシである。本論文では、一つ目の課題に対して性能評価を行い、どのプロトコルが適切かを示す。また二つ目と三つ目の課題に対してそれらを改善する手法を提案する。また、このディレクトリ方式の一貫性制御の実現においては、安定動作のためにメモリアクセスが有限時間内に完了することを保証することも重要である。本論文では、この課題を実現する手法も提案する。</p> <p>2章では、ディレクトリ方式でのプロトコルの性能評価とその解析を行った。ディレクトリ方式の一貫性制御は、従来評価されてきたスヌープ方式とシステム構成もプロトコルも異なるため、ディレクトリ方式の特性を踏まえて最適なプロトコルを選択することが重要となる。本章では、Invalidate型、Update型、Competitive型の代表的な3種のプロトコルの性能評価を行った。その結果、スヌープ方式での評価結果と異なり、Competitive型がどのワークロードでも良好な性能を発揮することが明らかとなった。また、ディレクトリ方式特有の振舞いである、キャッシュからの読出しあるいは書込み要求を受けたメモリが他のノードに転送する要求の割合や数に強い影響を受けて、プロトコル間の性能差が発生していることを確認した。</p> <p>3章では、共有しているノード数の増加によるメモリアクセスレイテンシの悪化を抑えるディレクトリおよび一貫性制御手法を提案した。ディレクトリはデータのコピーを保持しているノードを管理する。このディレクトリには、システムを構成するノード数よりも少ないビット数で構成して、システムコストを抑える方式が広く用いられている。しかし、不正確な管理となるため、一貫性制御の上では共有しているノード数が実際よりも大幅に増加する問題がある。また、共有しているノードへの要求の送信とそれらのノードからの応答の受信を、ノード数に比例せず一定の時間で実現する一貫制御が不可欠である。本章では、ビットパターン方式のディレクトリを提案する。このビットパターン方式は、共有しているノード数が実際よりも大幅に増加する問題に対して、特に大規模システムの一部のノードを利用してアプリケーションを動作させるケースで、保持している可能性のあるノード数の増加を抑えることができることを目的とする。また、提案するマルチキャスト・ギャザー機構は、このビットパターン方式やその他のCoarse Vector等のディレクトリをマルチキャストの宛先指定や、ギャザー時の待合せパターンに利用可能で、様々なネットワークのトポロジで実現可能な方式である。提案したビットパターン方式とマルチキャスト・ギャザー方式を実装した並列計算機Cenju-4においてメモリアクセスレイテンシの評価を行い、128ノードのシステムにおいても書込みのアクセスレイテンシを一定に抑えることができることを確認した。</p> <p>4章では、近年多数のシステムで採用されている、二つのネットワーク・プレーンを持つシステムを対象に、二重同型ネットワークを提案した。この二重同型ネットワークは、各プレーンをグラフ同型のネットワークで接続することを特徴とする。HypercubeおよびFolded-Hypercubeを題材にして二重同型ネットワークの構成法を示し、またそれらの性能評価を実施した。その結果、適切な同型ネットワークの選択により、同一のネットワークを二重化した場合と比較して、ネットワークコストを増加させることなくレイテンシおよびスループットを改善できることを確認した。</p> <p>最後に、5章では、メモリアクセスが有限時間内に完了することの保障について議論を行った。ディレクトリ方式においては、一貫性制御はノード間で複数回メッセージをやり取りすることによって実現される。この一貫性制御によってデッドロックが発生するため、それを防止することが求められる。また、この一貫制御においては否定応答によって要求を再試行する仕組みが広く採用されており、他の要求に割り込まれて無限に要求の再試行を繰り返し、スタベーションを引き起こす。本章では、メッセージをメモリに退避することで資源の依存関係をなくし、単一チャネルからなるネットワークにも適用可能なデッドロック防止方式を提案する。また、否定応答による一貫制御の再試行をなくし、メモリが一貫性要求を受け付けた順で処理すること、そのために要求をメモリに一時退避する機構を持つことを特徴とする、スタベーション防止方式を提案する。これらの機構を並列計算機Cenju-4に実装し、アプリケーションを実行した評価で、メモリアクセスがスタベーションやデッドロックによってタイムアウトを起こすことなく、安定して動作することを確認した。</p> |                            |

## 論文審査の結果の要旨及び担当者

| 氏 名 ( 細見 岳生 ) |     |     |       |
|---------------|-----|-----|-------|
|               | (職) | 氏   | 名     |
| 論文審査担<br>当者   | 主査  | 教授  | 下條 真司 |
|               | 副査  | 教授  | 藤原 融  |
|               | 副査  | 教授  | 鬼塚 真  |
|               | 副査  | 教授  | 原 隆浩  |
|               | 副査  | 教授  | 松下 康之 |
|               | 副査  | 教授  | 春本 要  |
|               | 副査  | 准教授 | 伊達 進  |

## 論文審査の結果の要旨

大規模分散共有メモリスシステムの性能を左右するメモリアクセスレイテンシは次の三点に影響を受ける。一つ目は、あるノードのプロセッサが行った書込みを他のプロセッサに反映するプロトコルとして、Invalidate型やUpdate型等のようなプロトコルを採用するかである。二つ目は、書込み時にはデータを共有しているノード全てに対して要求を伝える必要があるため、その場合の一貫性制御方式である。三つ目は、ノード間の要求やその応答をネットワークを介して送信する必要があるため、ネットワークのレイテンシである。提出された論文では、一つ目の課題に対して性能評価を行い、どのプロトコルが適切かを示している。また二つ目と三つ目の課題に対してそれらを改善する手法を提案している。また、このディレクトリ方式の一貫性制御の実現においては、安定動作のためにメモリアクセスが有限時間内に完了することを保証することも重要であり、本論文では、この課題を実現する手法も提案している。2章では、ディレクトリ方式でのプロトコルの性能評価とその解析を行っている。ディレクトリ方式の一貫性制御は、従来評価されてきたスヌープ方式とシステム構成もプロトコルも異なるため、ディレクトリ方式の特性を踏まえて最適なプロトコルを選択することが重要となる。本章では、Invalidate型、Update型、Competitive型の代表的な3種のプロトコルの性能評価を行い、その結果、スヌープ方式での評価結果と異なり、Competitive型がどのワークロードでも良好な性能を発揮することを明らかにしている。3章では、共有しているノード数の増加によるメモリアクセスレイテンシの悪化を抑えるディレクトリおよび一貫性制御手法を提案している。本章では、ビットパターン方式のディレクトリを提案し、提案したビットパターン方式とマルチキャスト・ギャザー方式を実装した並列計算機Cenju-4においてメモリアクセスレイテンシの評価を行い、128ノードのシステムにおいても書込みのアクセスレイテンシを一定に抑えることができることを確認している。4章では、近年多数のシステムで採用されている、二つのネットワーク・プレーンを持つシステムを対象に、二重同型ネットワークを提案している。この二重同型ネットワークは、各プレーンをグラフ同型のネットワークで接続することの特徴とする。HypercubeおよびFolded-Hypercubeを題材にして二重同型ネットワークの構成法を示し、またそれらの性能評価を実施し、適切な同型ネットワークの選択により、同一のネットワークを二重化した場合と比較して、ネットワークコストを増加させることなくレイテンシおよびスループットを改善できることを確認している。最後に、5章では、メモリアクセスが有限時間内に完了することの保障について議論を行い、メッセージをメモリに退避することで資源の依存関係をなくし、単一チャネルからなるネットワークにも適用可能なデッドロック防止方式を提案している。また、否定応答による一貫制御の再試行をなくし、メモリが一貫性要求を受け付けた順で処理すること、そのために要求をメモリに一時退避する機構を持つことを特徴とする、スタベーション防止方式を提案している。これらの機構を並列計算機Cenju-4に実装し、アプリケーションを実行した評価で、メモリアクセスがスタベーションやデッドロックによってタイムアウトを起こすことなく、安定して動作することを確認している。

本論文は、高性能な大規模システムの構築において重要となる、キャッシュを持ったプロセッサとメモリからなるノードをネットワークで接続した大規模分散共有メモリスシステムにおいて、性能を左右するメモリアクセスレイテンシに影響を与えるプロトコル、一貫性制御、レイテンシに関して、新たな方式を提案するとともに性能評価を行い有効性を示しており、計算機アーキテクチャ技術の向上に大きく貢献し、重要な成果を挙げた研究として、情報科学に寄与するところが大きい。よって、本論文は博士（情報科学）の学位論文として価値のあるものと認める。