

Design and Evaluation of Shared Memory Architecture for WDM-based λ Computing Environment

WDMに基づく λ コンピューティング環境のための共有メモリアーキテクチャの設計と評価

大阪大学 大学院情報科学研究科
博士前期課程2年 村田研究室
谷口英二

2006年2月17日

修士論文発表会

1

発表内容

- 研究の背景
 - λ コンピューティング環境
- 研究の目的
 - 共有メモリアーキテクチャの設計と評価
- 設計
 - Ring-UMAアーキテクチャ
- モデル化と解析
- 評価
- まとめ

2006年2月17日

修士論文発表会

2

研究の背景

- 現在のグリッドコンピューティングの問題点
 - 分散計算時のデータ共有の際にデータが大規模のためネットワーク上の処理によって性能低下
 - 輻輳、パケット損失、再送
- λ コンピューティング環境の提案
 - WDMを利用してノード間に波長パスを設定
 - 波長パスを利用した高速、高信頼な通信環境を分散計算時のデータ共有に利用

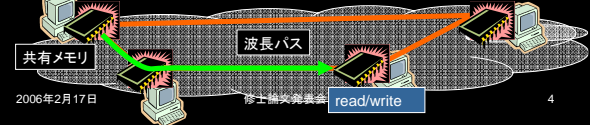
2006年2月17日

修士論文発表会

3

λ コンピューティング環境

- 光ネットワークを利用した分散計算環境
 - 波長パスと共有メモリを直結
 - 波長パスを用いて共有メモリにアクセス
 - ハードウェア分散共有メモリを実現
 - 従来は密結合であった共有メモリシステムを広域に展開することが可能
 - データ共有の際にソケットのオープンなどの処理が不要



2006年2月17日

修士論文発表会

4

研究の目的

- 広域に展開しているハードウェア分散共有メモリの性能調査
- λ コンピューティング環境における共有メモリアーキテクチャの設計と評価
 - 設計
 - 共有メモリアーキテクチャの特徴を決定する要因を検討
 - キャッシュの一貫性保持に焦点を当てた動作設計
 - 評価
 - 計算機ノードのモデル化と解析
 - 計算スループットを指標として評価

2006年2月17日

修士論文発表会

5

設計

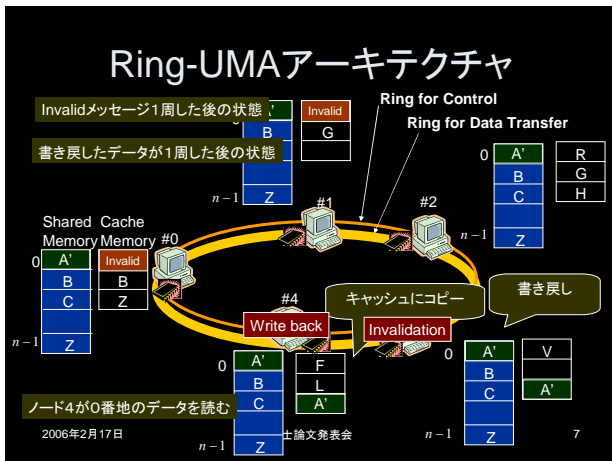
- 以下の事項を検討して設計
 - トポロジ
 - メッシュトポロジ、リングトポロジ
 - メモリアクセスモデル
 - UMA: 任意のアドレスに同一時間でアクセス可能なモデル
 - NUMA: アドレスによってアクセス時間が異なるモデル
 - キャッシュ一貫性制御プロトコル
 - ネットワークの使用頻度が少ない無効化型ライトバックを適用
- 設計したアーキテクチャ
 - Ring-UMAアーキテクチャ
 - Ring-NUMAアーキテクチャ
 - Mesh-NUMAアーキテクチャ

※UMA: Uniform Memory Access
NUMA: Non-Uniform Memory Access

2006年2月17日

修士論文発表会

6



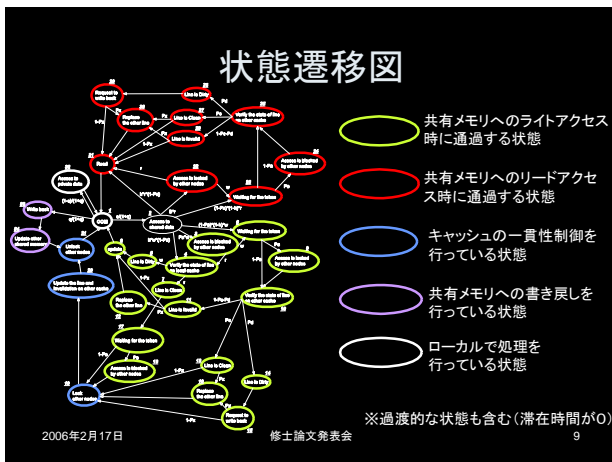
モデル化と解析方法

- 設計したアーキテクチャの動作に基づき状態遷移図を作成
 - 計算機ノードのCPUの状態に着目
 - 計算している状態
 - ローカルメモリにアクセスしている状態
 - キャッシュの一貫性制御を行っている状態 など
- 解析にはセミ・マルコフ過程を利用
 - 状態の滞在時間を任意に設定可能
 - キャッシュ一貫性制御などの複雑なリクエストを含むCPUの処理状態を記述可能

$$P_i = \frac{\eta_i \pi_i}{\sum_{j=1}^K \eta_j \pi_j}$$

P_i : 状態 i の定常状態確率
 η_i : 状態 i の滞在時間
 π_i : 離散マルコフ連鎖時の状態 i の定常状態確率
 K : 状態数

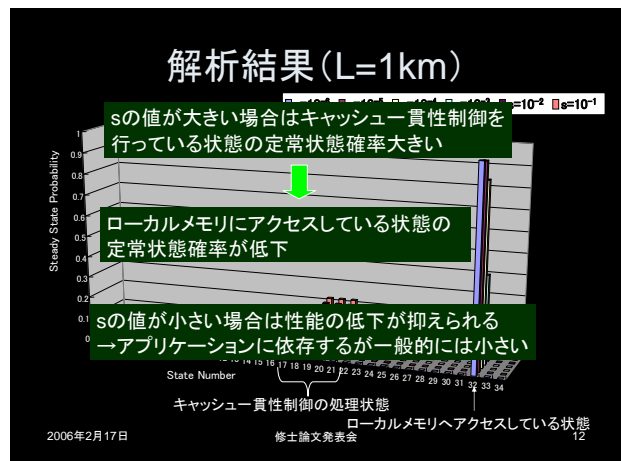
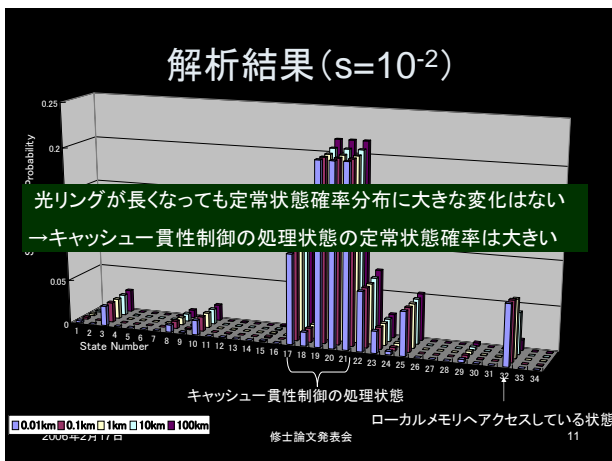
2006年2月17日 修士論文発表会 8



評価のための数値例

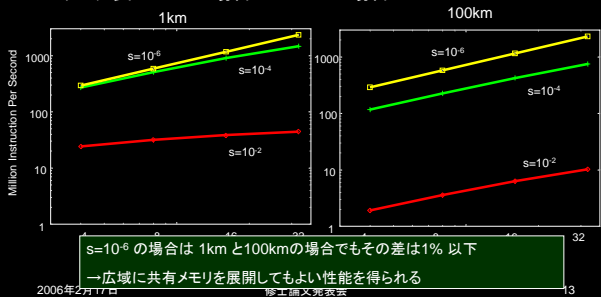
- 計算機ノードのスペック
 - CPU-キャッシュ間の遅延 10ns
 - キャッシュメモリ間の遅延 1us
 - キャッシュ容量 1024 KB
 - キャッシュラインサイズ 4 KB
- ネットワーク
 - インターフェイスでの処理遅延 3 us
 - 帯域 10 Gbps
 - 伝播遅延 5 us/km
- その他、必要となるパラメータ
 - キャッシュヒット率 0.95
 - CPUが発行する命令頻度
 - LOAD命令 15%, STORE命令5%, その他(レジスタ間演算など) 80%
 - ノード数 16
- 解析に際し、変化したパラメータ
 - 光リングの長さ L km (0.01km~100km)
 - 共有メモリへアクセスする割合 s ($10^{-1} \sim 10^{-9}$)

2006年2月17日 修士論文発表会 10



計算スループット (MIPS)

- 得られた定常状態確率を利用して計算スループットを導出
- リング長が1kmの場合と100kmの場合



まとめ

- まとめ
 - 共有メモリアーキテクチャの設計と評価
 - Ring-UMAアーキテクチャの場合
 - 広域に展開した場合も共有メモリのアクセス頻度が小さい場合は十分な性能が得られる
 - 共有メモリへのアクセス頻度は一般的には小さい
 - 他アーキテクチャでも同様の結果が得られた
- 今後の課題
 - 使用可能な波長数を考慮したアーキテクチャの提案

2006年2月17日

修士論文発表会

14