

高速ストレージクラスメモリを用いた 極低消費電力ヘテロジニアス分散ストレージサーバシステムの研究開発

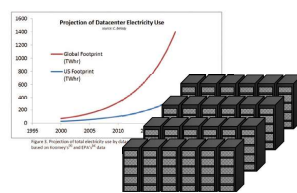
委託先 東京大学、東京工業大学、富士通株式会社、日本電気株式会社

本研究の応用

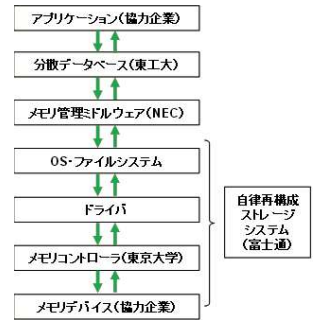
リアルタイム応答の多様なアプリケーション



増大するデータセンターの電力を1/10に削減



研究実施体制



自律再構成
ストレージ
システム
(富士通)

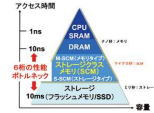
多くの分野・階層を融合する、
ソフト・ハード統合チーム

本研究で開発したヘテロ・ストレージシステム

ストレージクラスメモリを使った性能10倍、電力1/10倍のデータセンタースケール・ストレージと、メモリ構成を自動最適化するインテリジェントなストレージシステムを開発

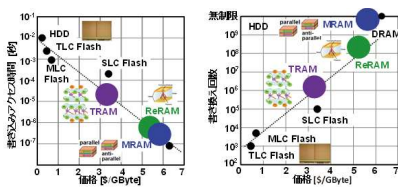
ストレージクラスメモリの活用

MRAM・ReRAMなど、DRAMよりも大容量でフラッシュメモリ・HDDよりも高速なストレージクラスメモリを活用

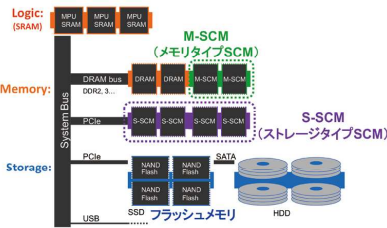


メモリのトレードオフ

メモリの容量・性能・電力・信頼性にはトレードオフが存在
本研究では異種のメモリを最適に組み合わせる(ヘテロメモリ)



開発したヘテロストレージシステム



研究成果の要旨

OS/FS/Driverレイヤの遅延・電力x1/10を確認(富士通)

Application Application アプリ層

ファイルシステム 制御ソフトウェア OS層

Flash S-SCM/Flash ハード

ヘテロ分散データベース(全体)で性能x10、電力x1/10を確認(東工大)

性能測定 検証 測定 測定 ネットワーク

分散データベース

メモリデバイスのヘテロ化で性能x10、電力x1/10を確認(中央大)

メモリデバイス

個別(メモリ、インターコネク、OS)及び、全体システムで性能x10、電力x1/10を確認

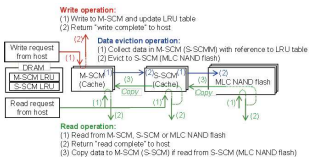
インターコネクの遅延・電力x1/10を確認(NEC)

インターコネク

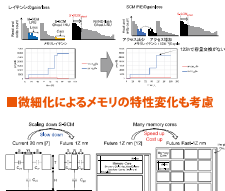
研究成果の詳細

実施項目1 半導体メモリシステム(東京大学)

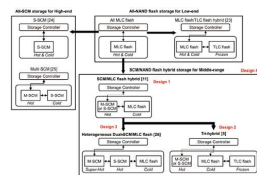
■Tri-hybridストレージを提案
アクセス頻度が多いSuper Hot/HotデータはM-SCM/S-SCMに記憶
アクセス頻度が少ないColdデータはFlashメモリに記憶



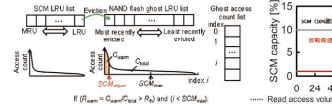
■メモリの信頼性も考慮し容量を自動調整



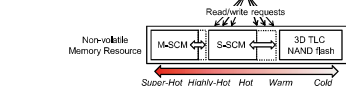
■Step-by-Stepのヘテロストレージの設計手法を確立



■ワークロードに応じメモリ容量を自動調整

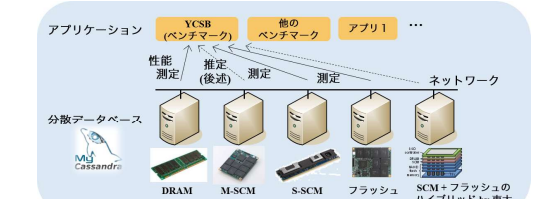


■従来のラックスケールからデータセンタースケールに拡張したテナントに対応



実施項目2 分散データベース(東京工業大学)

■分散ストレージとしての統合
任意の組み合わせ(ハード、ストレージエンジン、均質 vs. 不均質)を構成可能
ハード: SCMカード (by 富士通), 既存のフラッシュ等
ストレージエンジン: Apache Cassandra, MySQL, Redis 等



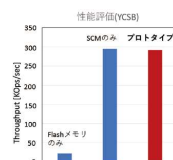
■性能の推定
問題: 将来のデバイスは現時点では存在しない
最新のデバイスは数が揃わない
解決: 回帰分析で性能を推定する手法を開発

実施項目3 自律再構成ストレージシステム(富士通)

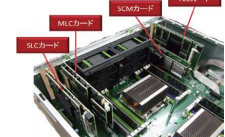
■ヘテロジニアス分散ストレージサーバシステムのプロトタイプ
Flashメモリ, SCMを制御するソフトウェア
最先端の市販デバイス、独自開発デバイスを同一APIでアクセス



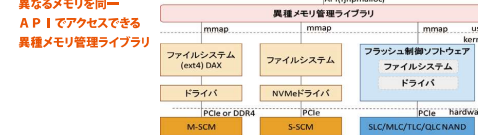
■プロトタイプでの性能評価



■試作したプロトタイプ



■インターフェースの異なるメモリを同一APIでアクセスできる異種メモリ管理ライブラリ



実施項目4 メモリ管理ミドルウェア(NEC)

■分散key-valueストレージ型の低遅延なインターコネクションの提案
①複数のOS/CM+GPUデバイス間でCPUやOSを経由せず直接データを接続
②1回のアクセスでデータを読み書き可能なデータ管理ミドルウェアによって目標の10usecデータアクセスを達成
キャッシュストレージとGPU直結バーストI/OのユースケースをPoC実証

アプリケーションからネットワークを通じた1回のデータ読み書きにかかる時間

データアクセス × データ管理

従来: 1回の読み書きにかかる時間比較

提案: 1回の読み書きにかかる時間比較

データ管理Middle Ware Stack

User-space: Application, KVS API, Consistent Hashing, Device mapping

Kernel-space: Intel SPDK User-mode Driver

M-SCM (MRAM) デバイスを用いた実証の評価結果

5.31usec

100us

100ns

10ns

1ns

事業者からのメッセージ

- 汎用的な性能向上から特化型による性能向上への流れができつつありますが、一般のアプリケーション開発者に各種特化デバイスに応じたコードを書いてもらうことは難しくなっています。このような状況の中、今後のシステムの性能向上とコスト削減は、特化型デバイスを統合するソフトウェア技術が中心になると考えられています。
- 今回研究開発した技術は、特化型デバイスの製品化を促している事業者様、先端アプリケーションの性能向上やコストに課題をお持ちの事業者様に貢献できると考えており、一緒にビジネスの開拓・拡大を進めていきたいと考えています。

竹内 健
東京大学大学院 工学系研究科
電気系工学専攻
E-mail: takeuchi@co-design.t.u-tokyo.ac.jp

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)の委託業務(JPNP16007)の結果得られたものです。