

Cell/B.E.とDIMMnetを併用した大容量ボリュームレンダリングの並列処理性能

田邊 昇† 佐々木愛美‡ 中條拓伯* 高田雅美‡ 城 和貴‡

† (株)東芝 ‡ 奈良女子大学 * 東京農工大学

あらまし 我々は数十GBのボリュームデータを実時間で可視化するシステムを提案し、その実現性を検討してきた。ボリュームレンダリングにおける視線方向のメモリアクセスはキャッシュフレンドリーではない。このため、我々はDIMMnet-2などに実装されたメモリ側のgather/scatter機能を、東芝のSpursEngineやCell/B.E.のようなDMAで主記憶をアクセスするCPUと併用し、この問題に対応することを提案している。本報告ではCell/B.E.単体や、それとDIMMnetを組み合わせたシステムにおける並列性能特性をPLAYSTATION®3上で評価した。その結果、SIMD化やマルチコア並列化で性能はスケールすることや、それらによる演算性能改善後もSpursEngineとDIMMnetの間を接続するPCI express x4インタフェースのバンド幅は飽和しないことがわかった。DIMMnetを併用しない場合はボリュームサイズの大規模化によって極端な性能低下が観測された。これに対しDIMMnetを併用すれば、ボリュームサイズの大規模化による性能劣化を抑制できることがわかった。

Performance of Parallel Processing of Volume Rendering for Huge Data with Cell/B.E. and DIMMnet

Noboru Tanabe † Manami Sasaki ‡ Hironori Nakajo* Masami Takada ‡ Kazuki Joe ‡

† Toshiba Corporation ‡ Nara Women's University

* Tokyo University of Agriculture and Technology

Abstract We proposed and investigated the feasibility of the real-time visualization systems for several tens Giga-byte volume data. Memory accesses in the direction of ray for volume rendering are not cache-friendly. Therefore, we proposed the solution for this problem by using memory-side gather/scatter functions implemented on DIMMnet-2 and CPU with DMA-type main memory accessing such as Toshiba SpursEngine and Cell/B.E. In this report, we evaluated the performance of a single Cell/B.E. based visualization system or a parallel visualization system with combination of it and DIMMnet on PLAYSTATION®3. As a result, we confirm the scalability of performance by SIMD processing and multi-core processing. We found that PCI express x4 interface between SpursEngine and DIMMnet does not saturate even though enhancement by using above two parallel processing. Extraordinary performance degradation is observed caused by increasing volume size in the case without DIMMnet. On the other hand, we found that DIMMnet reduces performance degradation caused by increasing volume size.

1 はじめに

スーパーコンピュータや、人工衛星・MRI・油田探索等の各種センサーが生成するデータのサイズは、それらのシステムの高性能化と大規模化を背景に年々飛躍的に増加している。このような大量のデータを分析するために、可視化は非常に有効な手法である。

日本においては、10Peta FLOPS級の次世代スーパーコンピュータの開発が推進されている。2012年に完成予定のこのシステムは、法律により広く共同利用されることが規定されている。このためユーザは、ネットワーク越しに遠隔から利用する形態が主流となると考えられる。

ところが、膨大な計算結果の可視化を遠隔から行う場合、従来はインタラクティブな利用は極めて困難とされている。遠隔ユーザが高性能化するスーパーコンピュータを利用する場合、最もクリティカルな部分は可視化によるデータの分析フェイズとなる。よって、幅広いユーザーからの使い勝手の良い利用のためには、上記の遠隔からの可視化がボトルネックにならないようにする必要がある。

上記の目的のために、数GBから数十GBのボリュームデータを実時間可視化すると同時にH.264のライブストリーミングで配信し、クライアントから対話的に制御できるシステムを提案し、上記のシステムを2年後に構築することを目標とし、そのフィジビリティを検討した。

以上の目標の達成には、複数の高性能計算エンジンと大

容量メモリからなる並列システムが必要になる。なぜならば、単一のGPUに装備されているメモリ容量には限界があり、上記の規模のボリュームデータをそこに保持することは不可能であるためである。ローカルには1GB以下のメモリしか実装できないGPUの現状を踏まえると、ボリュームデータの多くはGPUから見てローカルではなくホストインタフェースの向こう側にある確率が高くなる。よって、演算能力と、メモリ容量と、大容量メモリへのアクセス能力のバランスが重要になる。一方、現状のGPUは消費電力が高く、多数並べるには必ずしも最適ではない。そのホストインタフェースのバンド幅もローカルメモリにデータの大半が配置されることを前提としたバランスになっている。さらにH.264エンコーディングに関してGPUは必ずしも最適ではない。そこで、我々は具体的には通常のGPUではなく、東芝のSpursEngine[12]の並列利用を提案した。

また、限られたホストインタフェースバンド幅を有効活用することが望ましい。ボリュームレンダリングにおける視線方向のメモリアクセスは必ずしもキャッシュに適合するものではない。ボリュームレンダリングを行なう際の大容量メモリへのアクセスをアーキテクチャ的にサポートすることで、等価的なアクセス能力を向上できる可能性がある。例えば、不連続メモリアクセスの連続化機能を有する高機能メモリモジュールであるDIMMnet-2[7][8]のようなメモリシステムが有望と考える。よって、本報告では

Gather 機能を有するメモリシステムの有効性について検討する。

以下、本論文では、第2章でメモリ周辺で設計すべき項目を整理する。第3章で可視化システムのメモリアーキテクチャに関する提案方式を紹介する。第4章ではハードウェアシステム構成例について述べる。第5章では評価実験について述べる。第7章で関連研究について述べ、第8章で結論と将来の課題について述べる。

2 可視化装置のメモリ周辺の設計項目

2.1 メモリ容量の確保

スーパーコンピュータや人工衛星等の各種センサーが生産するデータのサイズは、それらのシステムの高性能化と大規模化を背景に年々飛躍的に増加しており、そのデータの可視化には膨大なメモリ容量が必要になってきている。特に多くのユーザから共用される可視化装置では、同時並行で何人かのユーザが利用する可能性も有り、可視化装置に数ユーザ分の高解像度ボリュームデータを保持していないと同時処理ができないことが予想される。

映像化した場合の肉眼での識別可能な解像度や、次世代スーパーコンピュータが出力する生データの格子点数との両面から、現実的にはユーザあたり数 GB(格子点数 $1000 \times 1000 \times 1000$ 程度)から数十 GB 程度(格子点数 $2000 \times 2000 \times 2000$ 程度)のボリュームデータを保持可能とすることが求められると考えられる。

この規模のデータは単一の GPU ボードに装備されているメモリ容量を大幅に超えるため、並列処理が必須となる。また、このクラスの容量を実現する場合ビット単価の高いメモリのみでの構成はシステムコストを上昇させる。

2.2 メモリバンド幅の確保

単一の GPU に装備されているメモリ容量を大幅に超えるボリュームデータのサイズでは、GPU から見てホストインタフェースの向こう側のデータをアクセスする確率が高くなる。さらに、ボリュームレンダリングにおける視線方向のメモリアクセスはキャッシュフレンドリーではない。まず、1回の画面生成において原理的には1回しか各ボリュームデータの格子点はアクセスされないため再利用性がない。つまり、時間的局所性は最悪である。さらに、視線の方向がボリュームデータ配列のデータ並びが連続な方向ではない場合、キャッシュラインサイズより大きなストライドで隔てられた不連続アクセスとなる。つまり、空間的局所性も悪い。よって、容量を拡大するのみでアクセス時の実効バンド幅を考慮しないと、所望の性能が得られない。

3 提案方式

我々は数 GB から数十 GB のボリュームデータを実時間可視化すると同時に、ライブストリーミングで映像を配信し、クライアントから対話的に制御できるシステムを提案した。以下に、そのようなシステムにおけるメモリ周辺に関する提案方式を挙げる。

3.1 DIMMnet-3 によるメモリ容量の強化

ユーザあたり数 GB から数十 GB のボリュームデータを保持可能とする場合、この規模のデータは単一の GPU ボードに装備されているメモリ容量を大幅に超える。このため、別途大容量メモリを GPU ボード外部に設けるか、高価かつ消費電力が大きいハイエンド GPU ボードを大量に

用いて、オンボードの高速メモリのみを寄せ集めてこのような大容量なメモリを実現しなければならない。

コストと消費電力とシステムサイズを節約する観点から、このような用途の大容量メモリの構築において、ネットワークとメモリのエンハンサである DIMMnet-3[9]を用いることを提案する。図1に DIMMnet-3 の写真、図2に DIMMnet-3 のブロック図を示す。4本の独立した DDR2 型のメモリバスを持ち、合計3本の240ピン DIMM スロットと2本の SO-DIMM スロットを有する。現状の DIMMnet-3 の幅は PC の 3.5 インチベイ、長さや高さは 5 インチベイに納まるサイズであり、メモリ容量を1個あたり最大 28GB 拡張することができる。

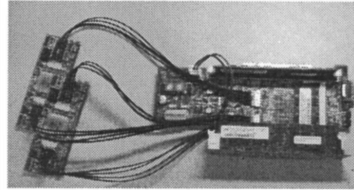


図1: DIMMnet-3 の写真

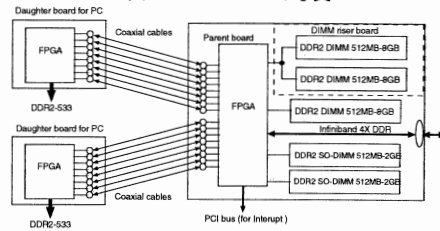


図2: DIMMnet-3 のブロック図

3.2 DIMMnet-3 による実効メモリバンド幅の向上

ボリュームレンダリングにおける視線方向のメモリアクセスはキャッシュフレンドリーではない。この問題の解決のためにも DIMMnet-3 を用いることを提案する。大容量メモリの構成において用いることを提案した DIMMnet-3 は、ベクトル型の不連続アクセスの連続化機能と、多数バンクからなるメモリシステムを有する。

特に計算エンジンとして利用を提案する SpursEngine には Cell/B.E.と同様にキャッシュアーキテクチャではなく DMA で主記憶アクセスを行なう SPE という CPU コアを複数搭載している。図3に示すように DIMMnet-3 によって不連続なアドレスにあるデータを gather して連続化した上で DMA によって SpursEngine 側に取り込む。この手法によって DMA を用いたメモリアクセスの効率化がなされ、不連続アクセスを主体とするアプリケーションにおいて大変有効であることが判っている。[10]

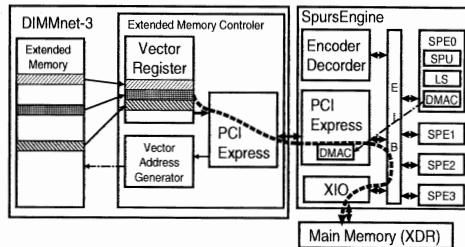


図3: メモリ側の収集機能による DMA パースト長の拡大

4. ハードウェアの概要

4.1 ボード構成

4.1.1 SpursEngine ボード

SpursEngine(SE)ボードは図4に示す CEATEC2007 において展示されたノート PC 向けボードのプロトタイプ、もしくはこれと同等のサイズの基板を用いることを想定する。図4の基板は SpursEngine を1個と XDR DRAM を128MB 搭載している。本アプリケーションで用いられる SE ボードのホストインタフェースは PCI express x4 であることを想定する。

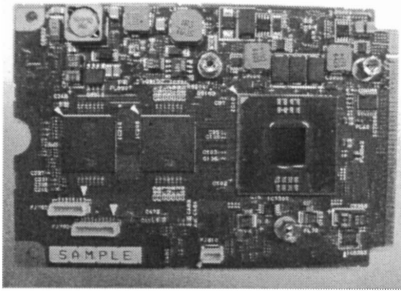


図4: SpursEngine プロトタイプ基板の写真

4.1.2 SE マザーボード

SE マザーボードはスイッチ LSI(PEX8632[13]相当品)と4枚の SE ボードを装着し、PCI express 2.0 x8 のカードエッジと、PCI express 2.0 x8 用コネクタを有する。図5にそのブロック図を示す。後者のコネクタには DIMMnet-3 との間を接続する集合同軸ケーブル用コネクタを有する子基板が装着される。

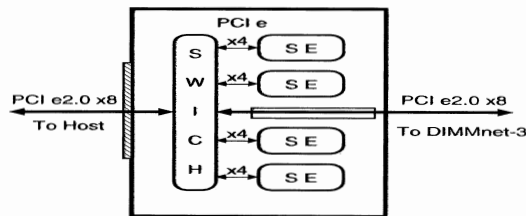


図5: SE マザーボードのブロック図

4.1.3 DIMMnet-3 親基板

DIMMnet-3親基板はDDR2 DIMM スロットに装着可能なネットワークインタフェース兼メモリ容量拡張ボードのメイン部分である。通常、この基板はDDR2 DIMM スロットに装着可能な PC 用子基板2枚と集合同軸ケーブルを介してFPGA の高速シリアルリンク RocketIO に接続して用いる。しかし、本システムでは PC 用子基板は用いず、そのために用いていた RocketIO は SE マザーボードとの接続のために使用される。ネットワークインタフェースはDDRタイプの Infiniband 4X 仕様である。本基板は DIMM を2枚、SO-DIMM を2枚の合計4枚の DIMM を搭載し、ホスト PC の5インチベイに設置されることを想定する。

本システムへの適用においては既存の DIMMnet-3 ボード上のFPGA に以下の2つの改良を加える。

- (1)RocketIO を PCI express x8 エンドポイント仕様で動作させる
- (2)他の DIMMnet-3 へのパケットを Infiniband ポートにル

ーティングする回路を入れる

4.2 結合方法

PCI express スイッチを活用して、ホスト~SpursEngine 間結合網、ホスト~DIMMnet-3 間結合網、SpursEngine~SpursEngine 間結合網を構築する。その際、リーフノード間の通信をルートへのトラフィックに影響を与えずに行なえ、かつ、1本の2.0 x8 仕様の Up リンク、1本の2.0 x8 仕様の Down リンク、4本の x4 仕様の Down リンクを構成できるスイッチを用いる。例えば市販 LSI としては、PLX 社の PEX8632[13]などが上記の仕様を満たす。Up リンクをホスト PC に、2.0 x8 仕様の Down リンクを DIMMnet-3 に、4本の x4 仕様の Down リンクを4個の SpursEngine に接続することでリーフノードである SpursEngine と DIMMnet-3 間の通信を効率的に行なえる低コストな結合網が構築可能であると考えられる。このような構成では4本の PCI express 2.0 x8 以上のスロットを有する PC に16個の SpursEngine と4個の DIMMnet-3 を接続することが可能になる。

このようなノード PC を4台接続することで、64個の SpursEngine と16個の DIMMnet-3 を接続することが可能になる。各 DIMMnet-3 には1~2GB の DIMM を4本搭載することで、合計64~128GB の大容量メモリを実現できる。DIMMnet-3 からは各々1本の DDR タイプの Infiniband 4X ポートがあり、これらは1台の Infiniband スイッチで接続することで DIMMnet-3~DIMMnet-3 間結合網とする。この Infiniband スイッチには、元となるボリュームデータを保持しているデータベースサーバも接続する。

以上のシステム全体の接続方法を図6に示す。

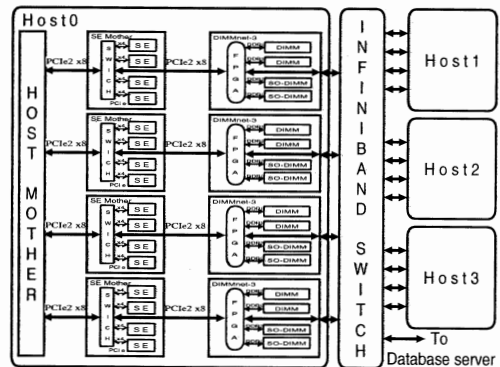


図6: 全体構成のブロック図

5. 評価実験

本章では、物量や遅延時間と性能の関係や、バンド幅に関する知見を得ることを目標に、簡素なボリュームレンダリングプログラムのプロトタイプを作成し、その上での性能評価実験を行った。

5.1 実験環境

実験環境を表1に示す。最終的なターゲットは前述のように SpursEngine を用いたシステムであるが、SpursEngine の環境は現段階では我々は入手できていない。一方、PLAYSTATION®3 に用いられているのは Cell/B.E.である。表2に Cell/B.E.は SpursEngine の仕様上の主な違いを示す。Cell/B.E.はレイアウトや周波数が異なるものの SpursEngine と機能的には同等の SPU を4個有する。この

ため、予備性能評価実験として実施される PLAYSTATION®3 上の測定は、SpursEngine ベースのシステム上の性能予測を行うための参考になると考えられる。

表 1 実験環境

機器	PlayStation®3(Cell B.E. 3.2GHz)
OS	Fedora 7
ソフトウェア開発環境	IBM Cell Broadband Engine SDK 3.0.0-1.0 libspe2-2.2.0 ppu-gcc-4.1.1 spu-gcc-4.1.1
時間計測	SPU Decrementer(精度 12.5ns)

表 2 Cell/B.E.と SpursEngine の仕様上の主な違い

	Cell/B.E.	SpursEngine
PPU	あり	なし
SPU	8 個	4 個
エンコード	なし	H.264, MPEG2
主記憶バンド幅	25.6GB/s	12.8GB/s
周波数	3.2GHz	1.5GHz
入出力 I/F	独自 (FlexIO)	PCI express x4
入出力バンド幅	76.8GB/s	1GB/s

5.2 評価アプリケーション

最終的なターゲットは、SpursEngine に直接接続される形で実装される外部メモリである XDR DRAM ベースのオンボードメモリ上に入りきらない数 GB から数十 GB クラスのボリュウムデータに対するレンダリングを行うことにある。本実験はそのような状況を想定して、ボリュウムデータは全て SpursEngine とは別に設置された DIMMnet-3 上に置かれているものとする。

DIMMnet と SpursEngine の組み合わせというこれまでの実行環境とは全く異なる環境での実現を目標としており、従来環境上での最適化手法は必ずしも適合するとは限らない。また、その環境上でのアルゴリズムの最適化自体が今後の研究対象でもある。このため、測定に用いるボリュウムレンダリングプログラムは現段階では極力シンプルなものを用いる。

ボリュウムレンダリングでは RayCasting 法[1]が通常用いられる。RayCasting 法では、視点からスクリーンの各ピクセルにベクトルを伸ばし、そのベクトル上にボリュウムデータのサンプリングポイントを取り、不透明度の累積計算を行う。本予備評価実験で使用しているプログラムは簡素化のためこれとは少し異なり、スクリーンに垂直なベクトルを各ピクセルから伸ばして計算している。つまり、厳密な RayCasting 法ではないが、不連続アクセスとなる不透明度の累積計算の部分に関しては RayCasting 法と同じである。よって今回の実験の範囲においては問題ないと考えている。

ベースとして使用しているボリュウムレンダリングプログラムのフローは以下の通りである。

Step1

ボリュウムデータを読み込み配列 alpha に格納

Step2

スクリーンの左上の座標

スクリーンを右方向に走査するベクトル

スクリーンを下方向に走査するベクトル

スクリーンに垂直なベクトル などの設定

Step3

スクリーンのピクセルに垂直な Ray を設定

Step4

Ray を伸ばしてサンプリングポイントを取り、ボリュウムデータ内にあるかどうか判定

Step5

ボリュウムデータ内であれば、alpha 値から輝度値を累積計算(alpha ブレンディング)する

Step4 に戻る

Step6

Ray を伸ばし終わったら求めた輝度値を配列に保持

Step3 に戻る

Step7

計算結果を PPM 形式の画像として出力

2008 年 10 月の HPC 研究会論文[16]で用いたプログラムでは上記のフローのプログラムをベースに Cell/B.E.上に以下の三つの手法での移植を行った。

- (1) alpha 値すべてを最初にローカルメモリ (LS)へ転送後、計算
 - (2) alpha 値を使用するとき使用する要素のデータだけを単純 DMA で LS へ転送し、計算
 - (3) あらかじめ使用する要素のアドレスを計算して配列に格納し、DMA リスト[14]転送で LS へ転送後、計算
- 本報告では上記の(3)のプログラムをベースに改良したプログラムで評価を行う。

5.3 SIMD 化とマルチコア化の効果

5.3.1 方法

2008 年 10 月の HPC 研究会論文[16]で用いたプログラムをベースにより高性能化を目指してプログラムの SIMD 化と、複数のコアを用いた並列化を同時に行ったプログラムを作成し、実行時間の変化を観察する。さらにボリュウムサイズを 16~256 の範囲で変化させた場合の実行時間の変化を測定する。

5.3.2 結果と考察

SIMD 化によって実行時間が 1.52 分の 1 に短縮した。これに伴い総実行時間に対するメモリアクセス時間の比率が上昇した。さらにマルチコア化の併用によっても実行時間が概ねスケールしており、SpursEngine で実質的な処理に用いることが可能な 3 個の SPU を用いた時に 2.28 分の 1 に短縮した。これに伴いメモリアクセスの実行時間における比率がさらに上昇した。

実行時間の短縮に伴って、必要とされるバンド幅はその比率分だけ上昇したことになる。SIMD 化と 3 個の SPU のマルチコア並列化によって合計で 3.47 倍のメモリバンド幅が必要になったことになる。SpursEngine の場合は Cell/B.E.に比べて周波数が約半分になっているため要求バンド幅はその分軽減される。以下では SpursEngine と Cell/B.E.は周波数に比例した性能比になると仮定する。こうして、トータルで以下の式で示されるバンド幅が SpursEngine の 3 個の SPU によるマルチコア並列化をした場合に要求されることが予測される。ここで、104MB/s は文献[16]のプログラムが必要としたバンド幅である。

$$104\text{MB/s} \times 1.52 \times 2.28 \times 1.5\text{GHz}/3.2\text{GHz} = 169\text{MB/s}$$

これは PCI express x4 のバンド幅(片方向 1GB/s)よりも小さいため、まだ SpursEngine の PCI express x4 インタフェースがボトルネックにはならないことを意味する。

使用コア数とボリュウムデータサイズを変化させた場合

の上記プログラムの実行時間の変化を図7に示す。

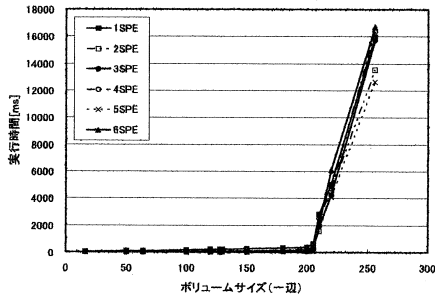


図7 PS3のみの場合の総実行時間

ボリュームサイズについては後述の DIMMnet を併用せず単純に PLAYSTATION3 のみを用いて実行している場合はボリュームデータの一面のサイズが200を超えたあたりから急激(概ね1000倍程度)性能が劣化する。この理由は主記憶の不足によって実行中に TLB ミスやページフォルトが多発する(スラッシングする)ようになり、HDD を記憶領域として使いながらの処理になるためと考えられる。

5.4 DIMMnet を併用する効果

5.4.1 方法

SPU ごとに8枚の512バイトのサイズを有するバッファ(配列)を主記憶に確保し、DIMMnet がプリフェッチスレッドの要求に応じて不連続アクセスを行って512バイトの連続領域に Gather し、プリフェッチスレッドがその領域から主記憶にコピーさせることを想定する。ワークスレッドは前述のプログラムにおいて DMA リスト転送をしていた部分を、上記バッファへのバースト DMA リードによってローカルメモリに整理したデータ列を取り込み、前節のプログラムと同様の処理を行う。本実験ではプリフェッチが十分なタイミングで行われ、ワークスレッドがバースト DMA を起動するまでにはプリフェッチが完了しているものと仮定し、その場合の処理時間を測定する。DMA の連続化の効果と、アクセス領域の削減に伴うスラッシング回避効果の二つの効果を分離すべく、前節の測定プログラム上でリストアクセス DMA を単純に連続アクセス DMA に書き換えたものの実行時間も併せて測定する。

5.4.2 結果と考察

図8にリストアクセスDMAを単純に連続アクセスDMAに書き換えたものの実行時間、図9に同様の場合のDMA転送時間を示す。

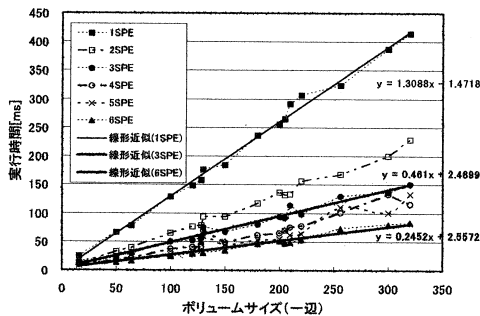


図8 リストアクセスDMAを単純に連続アクセスDMAに書き換えた場合の総実行時間

図8に示されるように、DMAの連続化は高い効果を示し、前節の図9の実行時間よりは実行時間が大幅に短縮していることがわかる。この範囲での特性は線形近似で近似可能といえる範囲に見受けられる。ただし、図9に示されるように、DMA転送部分のみの実行時間を観察するとスラッシングの兆候が観察されており、ボリュームデータの一面のサイズが200を超えたあたりから総実行時間に若干しか影響を与えない程度でのスラッシングと考えられる線形近似の直線から大きく外れた性能劣化があることがわかる。図8のボリュームサイズをこれ以上大きくすると総実行時間においても線形近似の直線から大きく外れる可能性が高いと考えられる。図7に比べ図8でページフォルトの影響が少ないのはメモリアクセス回数自体が連続化によって大幅に減少したためであると考えられる。なお、この測定環境ではボリュームサイズは320より大きくできなかった。

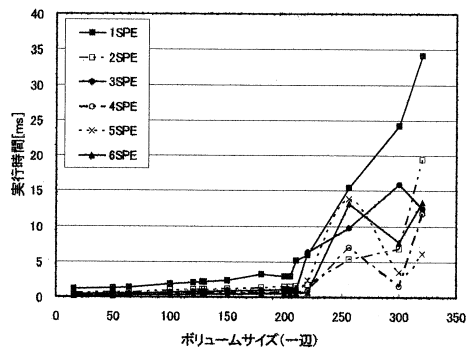


図9 リストアクセスDMAを単純に連続アクセスDMAに置き換えた場合のDMA転送時間

これに対して図10に示されるように、DIMMnetを併用して主記憶上の小容量バッファへのプリフェッチを行う場合は、サイズを300まで増やしても線形近似の直線から外れることなくボリュームサイズに比例した実行時間で実行できることがわかった。図示していないがDMA転送時間だけのグラフでも上記と同じ傾向であった。つまり、DIMMnet併用でスラッシングと考えられる性能劣化は全く観測されなくなった。よって、DMAの連続化による高い加速率を維持したまま、スラッシングによる性能劣化を回避してボリュームサイズのスケールビリティを提案方式が維持できることが示されている。

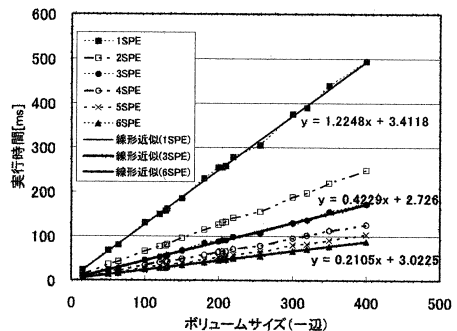


図10 DIMMnetでリストアクセスDMAを主記憶上のバッファへの連続アクセスDMA化した場合の総実行時間

6 関連研究

Castanie[6]は数十から数百 GB class のボリュームデータを扱うことができるメモリシステムの architecture を発表している。可視化における read only な性質を前提に coherence 維持のオーバーヘッドを排除した簡素な分散共有メモリを構築する点は我々の研究と共通である。しかし、ディスクアクセスを遠隔メモリアccessに置き換えることが主眼になっており、brick と呼ばれる固定サイズを有する一種のページをベースにしたメモリ管理が行われる。これはハードウェアとして COTS ベースの PC クラスタを用いており、DIMMnet-3 のような不連続アクセスの連続化による効率化は行わない。

一方、Muraki[5]は COTS の GPU を装着した PC クラスタに部分映像を合成する機能を有する専用ネットワークを付加することで、scaleable な volume rendering が可能であることを示した。本研究はこれとは全く異なるアプローチによって、使用する PC の個数やシステム全体のコストと消費電力の削減を目指すものである。専用合成ハードウェアにおいては画像の動的な解像度変更は困難であるが、我々の提案する方式は専用合成ハードウェアを持たないため、より柔軟性がある。

7 まとめ

本報告では、東芝の SpursEngine や Cell/B.E. のような DMA で主記憶をアクセスする CPU の併用における並列性能特性を PLAYSTATION®3 上で評価した。その結果、SIMD 化やマルチコア並列化で性能はスケールすることや、それらによる演算性能改善後も SpursEngine と DIMMnet の間を接続する PCI express x4 インタフェースのバンド幅は飽和しない見込みであることがわかった。つまり、DIMMnet 側の不連続アクセススループットを十分に確保できれば、遅延を隠蔽するのに十分な枚数のバッファを主記憶上に準備し、リストベクトル型のプリフェッチを DIMMnet に発行することで、Cell/B.E. が使う前までに上記バッファにデータをプリフェッチできる可能性が高いことを意味する。

DIMMnet を用いない素の PLAYSTATION®3 のままではボリュームデータの一边のサイズが 200 を超えると千倍におよぶ顕著な性能低下が発生することも観測した。提案手法は DIMMnet を併用することでメモリ容量を拡張し、プリフェッチを行う。常時プリフェッチが成功している場合、ボリュームサイズを大きくしたとしてもページフォルトはほとんど無い。このため、DIMMnet を併用すればボリュームデータの一边が 400 でも、高々それに比例した時間で実行できることがわかった。

本報告では ERT などの本システムでも大きな有効性が期待できる最適化は行っていない。ERT などの最適化技術を提案システムに向けて実際に適用したアプリケーションレベルでの実性能評価や、それに基づくハードウェア構成の最適化は今後の課題である。また、開発予算に目処がつかない SpursEngine ベースの環境での評価や、ハードウェアシステムの試作も行なう予定である。

謝辞

本研究の一部(DIMMnet-3 の開発)は総務省戦略的情報通信研究開発推進制度(SCOPE)の一環として行われたものである。

参考文献

- [1] M. Levoy : "Display of surface from volume data", IEEE Computer Graphics and Applications, Vol.8, No.3, pp.29-37 (May 1988)
- [2] M. Levoy : "Efficient Ray tracing of volume data", ACM Trans. Graphics, Vol.9, No.3, pp.245-261 (1990)
- [3] W.M. Hsu : "Segmented ray casting for data parallel volume rendering", Proceedings of 1993 Parallel Rendering Symposium (PRS'93), pp.7-14 (Oct.1993)
- [4] T. Porter and T. Duff. : "Compositing Digital Images" ACM Computer Graphics (SIG-GRAPH'84), Vol.18, No.3, pp.253-259 (1984)
- [5] S. Muraki, M. Ogata, K. Ma, K. Koshizuka, K. Kajihara, X. Liu, Y. Nagano and K. Shimokawa : "Next generation visual supercomputing using PC clusters with volume graphics hardware devices", Proceedings of Supercomputing Conference (2001)
- [6] Laurent Castanie, Christophe Mion, Xavier Cavin and Bruno Levy : "Distributed Shared Memory for Roaming Large Volumes", IEEE Trans. on Visualization and Computer Graphics, Vol.12, No.5, pp.1299-1306 (2006)
- [7] 田邊, 安藤, 箱崎, 土肥, 中條, 天野 : "プリフェッチ機能を有するメモリモジュールによる PC 上での間接参照の高速化", 情報処理学会論文誌コンピュータシステム, Vol. 46, No. SIG12 (ACS11), pp. 1-12 (Aug. 2005).
- [8] 田邊, 羅, 中條, 箱崎, 安藤, 土肥, 宮代, 北村, 天野: プリフェッチ機能を有するメモリモジュールによる等間隔アクセスの高速化, ハイパフォーマンスコンピューティングと計算科学シンポジウム(HPCS2006), p.55-62 (Jan. 2006).
- [9] 田邊, 北村, 宮部, 宮代, 天野, 羅, 中條 : "主記憶以外に大容量メモリを有するメモリ/ネットワークアーキテクチャ", 情報処理学会計算機アーキテクチャ研究会, 2007-ARC-172, pp.157-162 (Mar. 2007)
- [10] 田邊, 太田, 金, 中條 : "DMA で主記憶をアクセスする CPU における不連続アクセスの連続化", 第7回情報科学技術フォーラム FIT2008, 3L-4 (Sep. 2008)
- [11] InfiniBand Trade Association, <http://www.infinibandta.org/>
- [12] 東芝 : "メディアストリーミングプロセッサ SpursEngineTM SE1000 のサンプル出荷開始について", http://www.toshiba.co.jp/about/press/2008_04/pr_i0801.htm
- [13] PLX Technology : "PCI Express 2.0 Switches - PCIe ExpressLane I/O Interconnect", <http://www.plxtech.com/products/expresslane/gen2.asp>
- [14] M. Kistler, M. Perrone, and F. Petrini: "Cell Multiprocessor Communication Network: Built for Speed" IEEE Micro, vol. 26(3) pp. 10-23, (May-June 2006)
- [15] 田邊, 佐々木, 中條, 城: 大容量データ向け対話的実時間遠隔可視化装置の実現性検討, 電子情報通信学会コンピュータシステム研究会 2008-CPSY-18, pp.43-48, (2008,8)
- [16] 佐々木, 田邊, 中條, 高田, 城 : "Cell/B.E. と DIMMnet を併用した大容量ボリュームレンダリングの予備評価", 情報処理学会ハイパフォーマンスコンピューティング研究会 2008-HPC-117, pp.43-48, (2008.10)