

論文内容の要旨

博士論文題目

Performance Evaluation of a 3D-Stencil Library for Distributed Memory
Array Accelerators

(分散メモリアレイアクセラレータ向け3次元ステンシルライブラリの性能評価)

氏名 稲垣 慶和

本論文では、内部に小規模なメモリを分散配置してデータ移動を最小限に抑えたメモリ分散型レイ型アクセラレータ (Energy-aware Multimode Accelerator eXtension : EMAX) をターゲットとした 3D ステンシル演算ライブラリの命令、データ配置最適化手法と、その性能評価結果について述べる。

本研究では、画像処理、3次元シミュレーション、科学技術演算などの演算カーネルとして用いられるステンシル演算を EMAX において高速かつ容易に実現可能となるライブラリを提案した。通常 EMAX のような CGRA にてステンシル演算を実行するには、ハードウェア構造を意識したプログラミング設計や専用アセンブラなどによる命令写像の記述が必要となるが、本ライブラリを用いることで C 言語の関数呼び出しのように 1 行記述するのみで EMAX へのステンシル演算の命令写像が可能となる。

また本ライブラリは EMAX の持つ分散メモリと FIFO の共通データパス、命令シフトによるデータ再利用性を最大限利用し、データ転送量を最大限削減することで高速化を実現する。またライブラリによるループ自動展開による命令並列マッピング手法によりデータ転送量をさらに削減する。

提案ライブラリによる命令並列マッピング手法適用時の三次元ステンシル演算実行時間を測定し、測定データとハンドチューンで命令マッピングした実行時間を比較した結果、最大 90% 実行時間を削減できることを示した。また汎用プロセッサによる三次元ステンシル演算の実行時間と比較し、提案ライブラリを使用した EMAX では実行時間を 23% 削減できることを示した。

(論文審査結果の要旨) (A4 1枚 1、200字程度)

実用的な数値計算に頻出するステンシル計算は、座標空間上は隣接する入力データ群が、メモリ空間上は離散することから、空間的局所性に乏しく、商用汎用計算機に搭載されるキャッシュメモリの効果が出にくいことで知られる。一方、離散的に配置されたデータの再利用性を高め、メモリバンド幅を最大限に活用できる計算基盤として、CGRA (Coarse Grained Reconfigurable Architecture) が有望であるものの、CGRAのプログラミングにはハードウェアの構成や動作に関する十分な知識が必要であり、一般のプログラマが利用するためには、パラメタによりプログラマが必要とする計算パターンにハードウェアの動作を一致させることができるライブラリの提供が不可欠である。

本研究は、局所メモリを備える演算ユニットを格子状に配置して、局所メモリの再利用を行うことができるメモリ分散型アレイ型アクセラレータ (Energy-aware Multimode Accelerator eXtension : EMAX) を想定した、3次元ステンシル計算ライブラリの構成方法および性能評価をまとめたものである。最終試験では、以下の主要な成果について報告があった。

【1】3次元ステンシル計算の次数をパラメタ化し、X軸、Y軸方向の再利用を最大限に行うとともに、さらにZ軸方向の並列写像も併用することにより、次数が高い場合でも局所メモリの理想的な再利用を行うことができる高効率な基本写像方法を考案した。

【2】汎用CPUと性能比較した結果、次数=1の場合は、汎用CPUのほうが高速であったのに対し、より実用的なステンシル計算である次数=3の場合には、EMAXのほうが高速であることを示した。

【3】GPUと性能比較した結果、次数=1および次数=3のいずれについてもGPUのほうが高速であったものの、EMAXのコア数をGPUと同等に仮定した見積もりでは、次数=3の場合の性能はGPUの2倍に達した。また、GPUのメモリバンド幅が12GB/secであるのに対し、EMAXは300MB/secであることから、メモリバンド幅あたりの性能が極めて高いことも明らかになった。

以上、本論文は学術上、實際上寄与するところが少なくない。よって、本論文は博士(工学)の学位論文として価値あるものと認める。