

課題名 (タイトル) :

格子量子色力学クォークソルバーの FX100 向け最適化

利用者氏名 : 石川健一

所属 : 計算科学研究機構 連続系場の理論研究チーム

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

格子量子色力学(格子 QCD)は素粒子理論の強い力に関する力学(量子色力学 QCD)を第一原理から計算することのできる理論である。強い力は陽子や中性子、中間子などの性質とともに素粒子標準模型のクォークセクターの力学も記述している。格子量子色力学では連続的な 4 次元時空を格子に離散化することで量子力学の経路積分をモンテカルロ法により数値計算することが可能な体系となっている。物理量の期待値の計算のためのこのモンテカルロ法ではクォークの運動方程式(離散化された時空上の差分方程式)を解くところ(クォークソルバー)が最も時間のかかるところとなっている。格子量子色力学全体の数値計算がうまく進むためには、この運動方程式を効率よく高速に解くことが必要である。クォークソルバーの最適化・高速化は新しい計算機が登場するたびに行わなくてはならない。現在、主流の格子 QCD 計算の計算資源量は 100~300 TFlops*Years のものでスーパーコンピュータを用いて行われている。

京コンピュータも戦略分野5における格子 QCD を用いる課題で使用されており、そのコードは京コンピュータ用に最適化されたものを用いている。またそのコードは他計算機センターの FX10 でも利用されている。

現在国内では理研も含めて FX100 世代の計算機の導入が進んできているため、FX100 向けの高効率コードを提供することで分野全体の計算機資源の有効利用に寄与することができる。

本課題では、京コンピュータ用に開発された計算コードに FX100 のアーキテクチャに即した最適化を行うことを目的とする。

2. 具体的な利用内容、計算方法

最適化対象のクォークソルバーは、格子 QCD の Wilson/Clover フェルミオン運動方程式から導出される 4 次元時空格子に対する大規模な連立方程式を解

くことに対応する。連立方程式の係数行列は大規模疎行列であり、いわゆるステンシル系の問題である。解法としては Krylov 部分空間反復法の BiCGStab 法を用いており、数値精度はメモリバンド幅やキャッシュ効率を有効利用できるように単精度を用いている。

京向けの計算コードでのクォークソルバーは、京の 8 コアの有効利用と C 言語の SIMD 組み込み関数 (intrinsics) を用いた倍精度 2 成分の SIMD 最適化が施してある。その性能はノードあたり 4 次元格子サイズ $6^3 \times 12$ の時、疎行列ベクトル積演算部の演算効率は約 50%、ソルバー全体の演算効率は約 26% である。

一方 FX100 では、コア数が 16 コアをひとまとめにしたもの(CMG)が 2 つ実装された 32 コア(2CMG)であり、1CMG あたり到一个のアシスタントコアが設けられている。さらに、SIMD 長さは倍精度が長さ 4、単精度が長さ 8 に拡張されている。

連立方程式で解くべきデータは 4 次元格子上の格子点に 12 個の複素数データを置いたものに対応する。演算効率はメモリバンド幅に律速されている。並列化は 4 次元の格子点を分割することで行う。

京におけるコードでは、メモリバンド幅削減のためメモリ上のデータには単精度を用いているが、単精度 SIMD は利用できないため、データを演算処理する段階で単精度データをレジスタ上で倍精度に変換した後、倍精度の 2-SIMD 演算を行うようにしている。また、8 コアの並列化はコアごとにメモリアクセスが連続になるように、格子点データレイアウトを工夫している。

FX100 においては 16 コア \times 2 が使用でき、単精度の 8 SIMD 長の演算ができるようになったため、大幅なデータレイアウトの変更が必要であると考えられる。

HOKUSAI システムでは以上のこと考慮し FX100 向けのアーキテクチャに合わせたデータレイアウトの変更を行った。HOKUSAI システム使用開始段階では C 言語の SIMD 組み込み関数は提供されていなかったため、FORTRAN90 言語による記述に変更した。

3. 結果

1 プロセスあたりの 4 次元格子点数を T,Z,Y,X 方向に対してそれぞれ(NT,NZ,NY,NX)とする。DO ループは時間 T 方向が一番内側になるようにし、この T 方向を 8SIMD 長でベクトル化できるように目指した。Z,Y,X 方向は 16 コアでさらに分割する方針をとりいくつかのパターンを評価した。FX100 の 1 ノードには 1CMG たり 1MPI プロセスを割り当て(1 ノードに MPI プロセスを 2 つ起動)、1 ノードの 32 コア(2CMG)を使うこととした。評価には FX100 システムの提供する PA 情報によるプロファイリングと、コンパイラによって出力されるアセンブラコードによる SIMD 化の確認を繰り返しチューニングを行なった。

複素数ベクトルは、カラー3 とスピン4の自由度が 4 次元格子上にのっているが、時間方向長さ NT をベクトル長8で割り、長さ8の実数ベクトルをカラー成分(長さ3)とスピン成分(長さ4)の組にし、それを 2 つ並べて複素数を表現するようにした (y = type su3_spinor; real(4) :: yr(8,3,4); real(4) :: yi(8,3,4); end type)。次に時間 T のブロック番号、Z 格子番号、Y 格子番号、X 格子番号の順でアクセスするようにした (yb(NT/8,NZ,NY,NX)%y)。再内側のベクトル長 8 に対する DO ループが SIMD 化対象のループとなる。SIMD レジスタには、カラーとスピンの固定された実部または虚部の時間長さ 8 のデータが乗る。カラーやスピンの添え字に対する演算はレジスタ間の SIMD 演算に置き換わるようになっている。

つぎに 16 コアへの割り当てであるが、最終的に (NY,NX) の部分を (4,4) で分割し、(y(16)%yb(NT/8,NZ,NY/4,NX/4)) の様に 16 コアの コア番号の添え字が最外側に来るようにした。コア番号に対するループは OpenMP の並列化対象となる。以上の様なデータレイアウトの変更、及びアシスタント コアを用いた非同期通信の促進 (FX100 固有の通信促進関数使用) を施した結果、表 1 の性能が得られた。

測定した格子サイズと MPI 並列度は以下の通りである。全体 4 次元格子サイズ:24³x96, MPI 並列分割:4 次元 2x2x2x2, 1 MPI プロセスあたり格子サイズ:12³x48、である。16MPI プロセスを FX100 の 8 ノードに配置した。1MPI プロセスの担当する格子はさらに T 方向に 2 分割されており、これを 1 ブロックとする。1 ブロックの格

子サイズは 12³x24 である。

表1の区間名の説明は以下の通りである。「ブロック内差分」とはこの 1 ブロック内に作用するステンシル計算の部分を表す。「ブロック間差分」は、ブロック間を跨ぐステンシル計算部分を表す。ここではノード間通信が

区間	性能 [GFlops/MPI] (効率)	L1D ビジ ー率	L2D ビジ ー率	メモリ ビジ ー率
単精度ソルバ ー全体	56.3 (11.1%)	37%	21%	32%
ブロック内差 分演算	122.7 (24.3%)	65%	26%	39%
ブロック間差 分前処理演算	5.0 (1.0%)	23%	4%	5%
ブロック間差 分後処理演算	18.4 (3.6%)	26%	19%	26%

表 1 ソルバーの各区間の性能とキャッシュ・メモリビジョー率

必要となるため、MPI 通信の送信用のバッファにデータを集める前処理部「ブロック間差分前処理演算」と、受信バッファからデータを取り出して差分計算を行なう後処理の部分「ブロック間差分後処理演算」に分かれている (MPI 通信時間は含まれない)。「単精度ソルバー全体」は本チューニング対象の単精度 BiCGStab ソルバー全体をさす(通信時間も含まれる)。表にあるように通信を必要としないブロック内差分演算はチューニングの結果 1MPI プロセスあたり 122.7GFlops(効率 24.3%)の性能が得られた。なおこの部分の浮動小数点演算に対する SIMD 化率は、ほぼ 100%となっている。効率は倍精度のピーク性能との比である。単精度ではさらに倍のピーク性能であるので単性能効率は表の値の半分となる。ステンシル計算部の格子点あたりの演算量と必要データ量は約 1.5Flop/B である。FX100 の 1CMG あたりのメモリバンド幅を 480GB/s/2=240GB/s とすると、メモリバンド幅律速では理想性能は 375GFlops と見積もられる。理想値に比べると約 3 分の 1 の性能である。表にあるようにキャッシュとメモリの利用効率(ビジョー率)が低いことがその原因と考えられる。ステンシル計算ループ内へのソフトウェアプリフェッチの導入のチューニングが必要と考えられる。

4. まとめ

本課題では、格子 QCD 計算で最も時間がかかるクオークソルバーの FX100 への最適化について行った。本年度は、メモリ上のデータレイアウトを変更し演算部についての SIMD 化率をあげることで前述の様な性能

平成 27 年度 利用報告書

が得られた。しかしながら理想的なメモリバンド幅の利用ができていないことが判明しており、現在ソフトウェアフェッチの導入による最適化に取り組んでいるところである。