多倍長精度浮動小数点回路の性能評価

Performance evaluation of high-precision floating-point units

中里直人
会津大学

１．研究目的

Cygnus システムで採用された FPGA Stratix10 には単精度(IEEE 2008 binary32)浮動小数点加算・乗算の演算器マクロが搭載され、それでパイプライン演算を実現することで高性能かつ GPU と比べて低消費電力で数値シミュレーションが可能となった。一方で、他の演算精度の浮動小数点演算のためには、ALM と DSP を組み合わせることで演算回路を実装する必要がある。我々はこれまで独自に可変精度浮動小数点演算回路を生成するソフトウエア VeRB を開発した。 我々の浮動小数点演算回路生成コード VeRB では、特に binary128 を対象としてできるだけ ALM と DSP の消費量を少なくかつ高性能(200 - 250MHz)を目標として演算回路の最適化をおこなった。VeRB の生成する演算回路は、Stratix10 の特徴的なアーキテクチャに特化した最適化をおこなっていないため、本プロジェクトではより動作周波数の高い多倍長精度浮動小数点回路を生成するよう VeRB を改良し、アプリケーションでの性能評価をおこなう。

２．研究成果の内容

今年度のプロジェクトでは、VeRB により binary128 演算回路を生成し、重力多体問題と行列積の性能評価をおこなった。いずれの実装にも OpenCL カーネルを利用し、VeRB で生成した回路をライブラリとして呼び出す方法で実装した。

重力多体問題では、逆数平方根の演算が必要であり、25 ビット精度の初期値を計算する回路とニュートン法の組み合わせにより実装した。結果として粒子間相互作用あたりの演算数は 30 演算となった。Cygnus システムの FPGA では、演算パイプライン 1 要素の場合の演算性能は 4.45 GFLOPS で動作周波数は 223MHz となった。一方最大まで回路を実装した場合、パイプラインは 12 要素まで実装可能であり、演算性能は 38.2GFLOPS で動作周波数は 159MHz となった。この時の logic utilization は 84％であり、配線資源不足から動作周波数が低下したと考えられる。

行列積の実装では、binary128 の加算と乗算回路を接続し、さらにブロッキングによる性能向上を図った。ブロッキングを 1x1, 2x2, 4x4 とした場合の論理合成結果をテーブルに示す。

Blocking	1x1	2x2	4x4
Logic utilization	178,409 (19%)	194,129 (21%)	372,714 (58%)
ALUTs	120797	148511	387233
Registers	235,123	293,933	631,465
DSP blocks	21 (1%)	137 (2%)	1042 (18%)
Memory bits	7,176,392 (4%)	9,310,408 (5%)	14,845,640 (6%)
RAM blocks	667 (7%)	777 (7%)	1,280 (11%)
Fmax (MHz)	278.16	248.5	218.1

4x4 の場合、演算器総数は 128 のため理論性能は約 28GFLOPS である。しかし、Cygnus システムで性能を計測すると行列サイズが十分大きい場合でも約 6.5GFLOPS の性能となった。

３．学際共同利用が果たした役割と意義

binary128 フォーマットに対応した演算処理は、POWER アーキテクチャを除いてハードウエアを搭載した CPU はなく、Cygnus システムによる大規模な演算回路の性能評価が必須であり、今回の共同利用により Stratix10 FPGA での OpenCL によるアプリケーション性能評価ができた。

４．今後の展望

実際のアプリケーションで利用するためには、行列積での性能乖離の原因究明が必要である。他にファインマンループ積分の性能評価と、Cygnus システムの複数の FPGA を直接接続することでパイプライン回路として実現することは未着手のため今後の課題となる。

成果発表

(1) 学会発表　日本応用数理学会 2019 年度年会「多倍長精度浮動小数点演算の高速化手法」OS, 「多倍長精度浮動小数点回路によるパイプライン型アクセラレータの性能評価」, 中里直人, 台坂博, 石川正, 2019 年 9 月 5 日

使用計算機	使用計算機に○	配分リソース※	
		当初配分	追加配分
Cygnus	○	5000	
Oakforest-PACS			
※配分リソースについてはノード時間積をご記入ください。			