

課題名 (和文) 様々な演算形式による行列積の性能評価

課題名 (英文) Evaluation of Matrix Multiplication in Various Arithmetic Formats

代表者氏名 中里直人
所属 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(binary32)浮動小数点加算・乗算の演算器が搭載され、これにより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。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ェア VerB を開発し、多倍長精度浮動小数点演算回路(binary128 相当)を VHDL で実装した。

本研究では、これまでの研究に基づき、binary128 演算でのアプリケーションで利用した場合の現実的な性能評価と誤差評価をおこなう。これに加えて、可変精度浮動小数点フォーマットである POSIT による行列乗算の性能評価も行う。POSIT は、指数部を可変エンコーディングで符号化することで数値の絶対値が 1.0 に近い場合には仮数部が伸張する。32 ビット POSIT の場合、binary32 と比べて、最大 8 ビット仮数部が拡張されるため、より高精度な計算が可能である。本研究では、これまであまり評価されていない 64 ビット POSIT による行列積の性能評価をおこなう。

2. 研究成果の内容

(1) binary128 フォーマットの性能評価

これまで我々が実装を進めてきた binary128 行列乗算 FPGA アクセラレータの応用的な評価のために Cygnus に搭載される Stratix 10 を利用した。過年度研究により、我々のアクセラレータ実装は行列の 1 辺のサイズが 20000 以上となる問題において有効であることが分かっており、Cygnus の各ノードが約 192GB の大容量な宿主メモリを有していることから、宿主メモリに載せられる大規模行列の積に対して、我々のアクセラレータ実装の評価を進めた。

宿主に対して、FPGA ボードのメモリ容量は 32GB と小さいため、宿主側で入力行列をブロック分割し、FPGA ボードのメモリ容量に収まる行列サイズの問題を FPGA で複数回解いて集約する方法を宿主側で実装し、236MHz で動作する我々の FPGA アクセラレータ実装を用いてベンチマークを実行した。1 台の Stratix10 を用いたとき、60000×60000 の行列積で 60GFlops となり、我々が算出したピーク性能の 98%を達成した。本結果は RECONF 研究会および応用数学会年会で発表した。

さらに、Cygnus を利用して、ブロック分割した行列の各計算を複数の FPGA に割り当

てて並列計算をするために、MPI によるホストプログラムの拡張を行った。Cygnus が 2024 年度末で稼働終了したため、チューニング途中の予備評価に留まっているが、分割した行列を 2 台の Stratix10 に分散して解かせることで、16000×16000 の行列積の演算性能が 106GFlops に到達しており、複数の FPGA による並列計算の効果を確認した。

(2) POSIT フォーマットの性能評価

64 ビット POSIT の加算器と乗算器をパイプライン化し、FPGA でシストリックアレイとして実装することで、64 ビット POSIT による行列積を高速化した。Cygnus 上で論理合成と動作確認をおこなった。より高性能な FPGA である Altera 社 Agilex FPGA ボードで、シストリックアレイの PE 数が $8 \times 16 = 128$ 個の場合、行列積回路は 388.95MHz で動作し、ロジックセルは全体の 85%、DSP ブロックは 23%を消費した。64 ビット POSIT の行列積については、GPU での実装との性能評価を行い、また LU 分解および Cholesky 分解の性能評価と誤差評価をおこなった。以上の結果は、HPC 研究会にて発表した。

3. 学際共同利用プログラムが果たした役割と意義

FPGA ボード単体では Cygnus に搭載されている Stratix10 の性能はより最新の FPGA よりも劣っているが、今回の共同利用により、他にはない複数の FPGA ボードを利用できる環境で、binary128 による 2 台の並列計算を実証できた。また、64 ビット POSIT でも同様の高速化と検証が実施できた。

4. 今後の展望

複数の FPGA を利用した binary128 での行列計算の高速化は、FPGA を GPU に置き換えることで他の計算機システムでも実行が可能のため、今後はそれによる性能評価とアプリケーションの高性能化を目指す。

5. 成果発表

- (1) 学術論文
- (2) 学会発表

「FPGA による Binary128 行列乗算とアプリケーションの高速化」

河野郁也, 中里直人, 中田真秀

RECONF/SWoPP2024, 2024, 8 月 7 - 9 日, 徳島市

「Stratix10 FPGA による大規模な四倍精度行列乗算の性能評価」

河野郁也

日本応用数理学会 2024 年度年会 [正会員 OS] 多様な演算手法の高速化と応用, 2024, 9 月 14 - 16 日, 京都市

「様々な浮動小数点演算形式の評価プラットフォームとしての FPGA と GPU の比較」

中里直人, 河野郁也, 中田真秀

情報処理学会 第 198 回 HPC 研究会, 2025, 3 月 17 - 19 日, 札幌市

(3) その他

使用計算機	使用計算機に ○	配分リソース※		
		当初配分	移行*	追加配分
Cygnus	○	4000		
Pegasus				
Wisteria/BDEC-01				
※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。 *バジェット移行を行った場合、「+2000」「-1000」のように記入				