

ゲノム情報学アプリケーションの 専用プロセッサ協調型並列処理

●中西 恒夫¹⁾ ◆小板 隆浩²⁾

1)奈良先端科学技術大学院大学情報科学研究科(現・九州大学大学院システム情報科学研究院) 2)大阪産業大学工学部(現・同志社大学工学部)

〈研究の目的と進め方〉

超大規模データを対象に組合せ最適化問題を適用するゲノム情報学アプリケーションを高速化するもっとも安価な解決策は、廉価なPCをネットワークで接続したPCクラスタである。しかしながら、PCに搭載されるプロセッサは汎用プロセッサであり、あらゆるアプリケーションを平均的に高速に実行できるように設計されている。そこで我々は特定のゲノム情報学アプリケーションを、ハードウェアによって高速に処理する専用プロセッサを、PCクラスタと併用することで、ゲノム情報学アプリケーションの一層の高速化を図る。

〈研究開始時の研究計画〉

1. ゲノム情報学アプリケーションの解析
2. 専用プロセッサの設計・開発
3. 専用プロセッサ協調型並列アルゴリズムの研究・開発

〈研究期間の成果〉

1. ゲノム情報学アプリケーションの解析
ゲノム情報学アプリケーションのうち、たんぱく質構造予測手法である3D-1D法、ホモロジー解析手法であるNeedleman-Wunsch法、Smith-Waterman法など、動的計画法を使用するアプリケーションの高速化を図るべく、これらのアプリケーションの動的計画法処理部をFPGAによる専用プロセッサにより、低価格で高速化することを検討した。

2. 専用プロセッサの設計・開発

動的計画法をハードウェア的に高速処理するFPGAベースのデータフロー型専用プロセッサを設計、実装した。アフィンギャップに対応したNeedleman-Wunsch法、Smith-Waterman法を当面の対象アプリケーションとして選択した。動的計画法をハードウェア実装する際にもっとも簡潔な実装として、1回の配列要素のスコア計算を1PEで行うアーキテクチャを採用することが考えられるが、その場合、動的計画法のアルゴリズム上、ほとんどのPEが休眠状態になる。本研究では、直線状に並べたPEに検索対象配列情報とスコア計算値を順次伝播させるアーキテクチャを採用することにより、休眠PEをなくしてチップ単位面積あたりの性能を高め、高並列処理を実現した。

選択したアプリケーションのうち、Smith-Waterman法をハードウェア実装した専用プロセッサを設計し、評価を行った。当該専用プロセッサは、10万ゲートのFPGAを2つ、ならびに4MバイトのSRAMを搭載する市販のPCIボード上に実装された。検索対象となる配列情報はPCからPCIボード上のSRAMに転送される。FPGA上に実装された専用プロセッサは、SRAMから配列情報を読み出し、動的計画法によるスコア計算を行い、計算結果をSRAMに書き込み、計算完了通知のためPCに割込みをかける。スコア計算を行うPEはFPGA上に最大128個実装でき、これら全てのPEは並列に動作する。

実機での評価の結果、128個のPEを搭載する当該専用プロセッサにより、Pentium III 1GHzで純ソフトウェア的にSmith-Waterman法を実行する場合と比べて、339.6倍の高速化を確認した。

3. 専用プロセッサ協調型並列アルゴリズムの研究・開発
動的計画法を適用する大規模な配列を分割し、専用プロセッサにその処理を割り当てる分割方式を開発した。

〈国内外での成果の位置づけ〉

FPGAを用いてホモロジー解析をする専用計算機はすでに商用化されているが、本研究は市販のFPGA搭載PCIボードとPCとの連携処理により高速処理を実現するものであり、システム構築に要する費用(1式30万円程度)は極めて小さい。研究実施当時、国内では、シストリックアレイ方式によって動的計画法をハードウェア実現した、FPGA搭載PCIボードによるホモロジー解析システムの開発例があった。動的計画法をシストリックアレイで処理する場合、動的計画法のアルゴリズム上、ほとんどの演算素子が休眠状態になる。本研究では、データフロー型アーキテクチャを採用することにより、休眠演算素子をなくしてチップ単位面積あたりの性能を高めて高並列処理を実現すると同時に、さらにアフィンギャップにも対応した。

〈達成できなかったこと、予想外の困難、その理由〉

専用プロセッサについては設計を完了し、FPGA搭載PCIボードに実装し、Smith-Waterman法のハードウェア計算による高速化を行った。研究期間内に、動的計画法を用いる他のゲノム情報学アプリケーションのハードウェア実装、ならびにPCクラスタとの併用による高速化は実現できなかった。

〈研究期間の全成果公表リスト〉

1. T. Mizuhara, T. Nakanishi, and A. Fukuda: A Data-Flow Genome Array Processor RING to Accelerate Homology Analysis, Proc. 2002 Int. Conf. on Parallel and Distributed Processing Techniques and Applications (PDPTA2002), Vol.III, pp.1193-1199, 2002.
2. 水原, 中西, 福田: FPGAによるゲノムシーケンス解析専用プロセッサの設計, 情報処理学会研究報告, 2001-ARC-145, pp.79-84, 2001年.