

高速情報圧縮復元ハードウェアアルゴリズムに関する研究

阿曾弘具

東北大学大学院工学研究科
電気・通信工学専攻

今日の広範囲へのコンピュータの普及およびコンピュータネットワークの発展により、多種多様かつ大量の情報が相互に送受信され、蓄積、管理・運用されている。この状況において、情報をより効率的に伝送、蓄積する有効な技術として、データ圧縮技術が重要となっている。符号化/復号化(圧縮/復元)する既存の手法は、利用者の明示的な指示によるオフラインの実行であるが、大量の情報やオンラインの情報に対しては、さらなる高速性が要求される。この高速化はVLSI/ULSI技術を応用したデータ符号化/復号化の専用ハードウェアの構築で可能となる。専用ハードウェアにより、コンピュータと通信線・外部記憶装置の間で自動的に圧縮/復元することが可能になり、有用性が増す。

本研究では、多様な形式の情報に対して高い圧縮率を実現するデータ符号化/復号化手法の一つであるLempel-Ziv符号化法について、その一つであるLZ77符号化/復号化法のハードウェアアルゴリズムを提案し、並列処理アーキテクチャPAHL (Parallel Architecture for High-speed Lempel-ziv data coding/decoding)として実現した。このアーキテクチャは、符号化を実現するPAHL-Cおよび復号化を実現するPAHL-Dからなっている。PAHL-Cでは、LZ77符号化法において最も計算時間を要する最長一致記号系列探索部の冗長な比較演算をできるだけ削減する工夫をした。すなわち、最長一致記号系列探索部を各位置からの一致長を判定する部分と、各位置での一致長から最長一致系列を決定する部分とに分割し、それぞれを並列に同期動作させ、より高いスルー・レートの符号化処理を実現している。本研究では、提案したアーキテクチャに関して、高位論理合成システムPARTHENON (NTT情報通信網研究所)を用いて論理設計、動作シミュレーション、論理合成等を行ない、その実現性を検証し、性能を定量的に評価した。LZ77符号化による生成符号の統計的特徴を調査して、現実のデータに対するPAHL-Cのスルー・レートを定量的に評価し、約20~25 MByte/secというスルー・レートが得られることを確かめた。

LZ77符号化法は、より高い効率を目指した種々の改良がなされている。その一つにLZSS符号化法があり、LZ77符号表現の冗長部分を削減することにより高圧縮率を実現するものである。LZSS符号化は、実際に使用されているデータ圧縮アプリケーションソフトウェアのlhaやgzip等で使われている。LZSS符号によるデータ圧縮はLZ77と同様に可逆圧縮で、他のデータ圧縮法との組み合わせにより様々なデータに対し大きな効果をもたらしている。そこで、本研究の次のステップとして、LZSS符号の高速符号化/復号化を実現する高速LZSS符号化/復号化並列処理アーキテクチャ-PAHL-LZSSを新しく考案した。PAHL-LZSSは、本研究で提案したPAHLの基本アーキテクチャを拡張することにより実現している。PAHLと同様に、PARTHENONでの論理設計・動作検証・論理合成を行ない、PAHL-LZSSの予測性能を求め、PAHLとほぼ同様の性能が実現可能であることを確認した。

さらに、PAHLおよびPAHL-LZSSを実際にハードウェアとして実装するため、モジュール化による実装手法を考案した。復号化部は、回路規模的には、実用的な参照部バッファサイズに対して、参照部バッファ以外は全て一つのモジュールに収まる。符号化部は、一致記号系列を探索するモジュールModule-1(参照部バッファを含む、約1万ゲート、入出力約100ビット)、最長一致記号系列データを選択するモジュールModule-2(約2千ゲート、入出力約220ビット)、カウンタ・データ遅延用バッファモジュールModule-3(小規模)の3つのモジュールで構成できることを示している。

本研究およびその後の研究において、藤岡豊太氏(現在、岩手大学)が大きな貢献をしている。ここに記して協力に感謝する。