

報道関係者各位

国立大学法人筑波大学  
国立研究開発法人科学技術振興機構 (JST)

## 流れるデータをリアルタイムに圧縮・解凍する新技術を開発

映像など無限に流れるデータを圧縮器に一度通すだけで、頻出するデータのパターンを自動的に見つけて圧縮・完全に復元（解凍）可能な新技術を開発しました。従来の技術より10～30%圧縮率が高く、次世代データ伝送方式としてIoTやクラウドなどに用いれば、消費電力を大幅に削減できます。

IoT（モノのインターネット）やクラウドなど現代社会に欠かせない情報通信技術では、映像やセンサなどのデジタルデータを計算し、伝送する速度がサービスの質を決めます。例えば、監視カメラ映像による人物特定では、映像の伝送、人工知能（AI）での認識、結果を利用者に返すといった一連の流れを高速化する技術が求められます。実はこの過程で時間がかかるのは、AIによるデータ計算よりもデータの移動（通信）です。このため、元データの情報の高精細さを損なわずにデータ量を可能な限り削減し、通信時間を減らす手法の開発が望まれています。実現すれば、AIの計算に必要なデータをより短時間で伝送できます。さらに、5G/6Gなどの無線通信やインターネットなど広域通信に流れるデータ量を減らせるため、世界規模での消費電力の削減につながります。

従来のデータ圧縮技術では、データを有限なある程度のまとまりとして記録し、それを圧縮していたため、データを記録するメモリーやデータを処理するプロセッサが不可欠でした。このため、終わりなく連綿と流れるデータ（データストリーム）を圧縮する場合、圧縮機構は大がかりになる一方、リアルタイムの圧縮はできないという問題がありました。

本研究では、データストリームを圧縮器に一度通すだけで、頻出するデータパターンを自動的に探し出し、最少で1ビットにまで圧縮できる新技術を開発しました。しかも、どのようなデータでもリアルタイムに完全に元に戻すことができます。これまでの技術では、一つの単位データ（シンボル）を1ビットにまで圧縮することはできましたが、新方式では複数の単位データをまとめて1ビットにまで圧縮できます。これにより、従来方式に比べ圧縮効率が10～30%向上しました。

また、今回の新技術では、プロセッサやメモリーなどを利用することなく、高速でコンパクトなデータ圧縮機能のハードウェア化が容易にできるという特色があります。半導体チップに今回の成果を適用し、AIに実装すれば、データ量の削減によるさらなる高速化、データの移動経路における省電力化が図れ、Society 5.0を実現する持続可能なブレークスルー技術となることが期待されます。

研究代表者

筑波大学システム情報系

山際 伸一 准教授

## 研究の背景

これまでのロスレスデータ圧縮技法（圧縮後のデータを解凍すると完全に元に戻る圧縮方式）は無遠（ここでは、終わりの定めがないという意味）のデータストリーム<sup>注1</sup>を扱うことができませんでした。さらに、ソフトウェアでの実装方式が主であり、ハードウェアでの実装が難しく、高速な伝送路への対応が困難でした。

本研究の先行研究では、データ単位（シンボル）<sup>注2</sup>を二つから一つに変換して圧縮する技術（LCA-DLT）を提案し、無遠のデータストリームをメモリ無しで扱えるようにできました。例えば、ABABCD<sub>CD</sub>...という文字の列はストリームに出現後に AB→X、CD→Y と置き換えられ、ABXCDY...と最高で50%までの圧縮ができます。さらに圧縮器を数珠つなぎにして1/2、1/4...と圧縮できるユニバーサル性がありましたが、シンボルの大きさ（ビット幅）以下には圧縮できませんでした。

その後、無遠のデータストリームのリアルタイムデータの複雑性（エントロピー）<sup>注3</sup>を自動的に把握するエントロピー計算<sup>注4</sup>を発見し、上の例ではAB11CD11...とシンボルを1ビットにまで圧縮できる方式ASE Codingを開発しました。ASE CodingはLCA-DLTに対してハードウェア量を90%削減できる一方で、ユニバーサル性がなく、複数のシンボルからなる頻出データパターンを1ビットにすることができないため、圧縮効率が頭打ちになる問題を抱えていました。

## 研究内容と成果

本研究ではASE Codingで1ビットに圧縮可能なエントロピー把握能力と、LCA-DLTの持つユニバーサル性を併せ持つ新ロスレスデータ圧縮方式Universal ASE Codingを開発しました。無遠のデータストリームを扱うことができ、一度のデータの読み込みで複数のシンボルからなる頻出データパターンを1ビットにまで圧縮可能な新手法です。また、その仕組みをアルゴリズム化し、圧縮器・解凍器として任意のデータを扱える方法を確立しました。

Universal ASE Codingでは、ASE Codingのモジュールをデータの流れるパイプラインに複数配置します（参考図）。パイプラインは圧縮データとオリジナルシンボルのものを配置し、前者は圧縮されたデータ、後者は入力されたシンボルが、入力毎にそれぞれに流れていきます。ASE Codingのモジュールは1個以上のシンボルを担当するようにし、パイプラインに先行するシンボル列が変換テーブルにマッチしたときに、先行するオリジナルシンボルを無効化し、圧縮データに変換してパイプラインに流れていきます。このように、入力を止めることなく次々と圧縮処理を行っていく制御を発見しました。圧縮処理は、変換テーブルにヒットしたシンボルパターンのインデックスに変換し、データ量を削減します。さらに、テーブルのエントリの使用数を元にしたエントロピー計算によってリアルタイムにシンボルパターンの出現頻度を予測し、さらに圧縮データのビット数を最少で1ビットにまで減らします。圧縮データがパイプラインの出口まで達すると、圧縮データが有効であれば、それを出力する選択処理を行います。

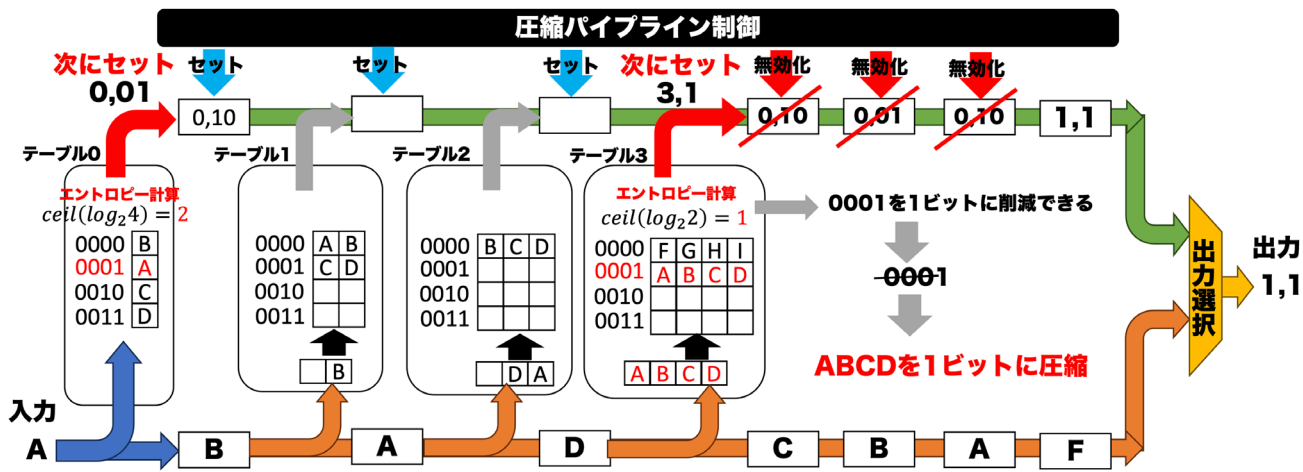
Universal ASE Codingの仕組みは半導体などの集積回路への実装が容易であることが特徴です。データの滞りが一切なく長いデータの列を1ビットにまで削減できるため、超高速通信網や大容量のストレージにおいてデータ移動時間を短縮し、これにより消費電力を削減します。

## 今後の展開

Universal ASE Codingは従来のASE Codingに比べ10~30%ほど圧縮効率を改善します。しかし、圧縮効率を高めるファクターの解明が十分ではないため、情報理論の学問体系から科学的に探求していきます。また、ハードウェア化においても可能な限りコンパクトで高速動作する構成を追求し、IoTやクラ

ウドへの応用をめざします。さらに、省電力効率や性能向上率についても、人工知能アプリケーションを中心に実験し、社会実装を進めていきます。

参考図



本研究で開発した Universal ASE Coding の圧縮器の動作例

圧縮器には二つのパイプライン（次々とバケツリレーのように出力に向かって流れていくデータの通り道）があり、圧縮データ（図の上側）とオリジナルデータ（図の下側）が伝搬していく。オリジナルデータが流れていく過程で、異なる長さのデータパターンを扱う圧縮器がパイプラインに配置されており、圧縮データに変換するためのテーブルをもっている。流れるオリジナルデータを蓄積しながら対応する長さのデータパターンをテーブルに登録していく。

入力データがテーブルに存在している場合は圧縮が可能であり、例えばテーブル0ではAが格納されているインデックス0001を求める。このテーブルの使用数は4であるから、エントロピー計算によって、2が求まり、0001を01の2ビットに削減し、次の入力タイミングで圧縮データとしてセットする。2個以上のシンボル列を担当するテーブル1, 2, 3はそれぞれ、2, 3, 4個のデータが蓄積したらテーブルを検索し、一致したデータ列が格納されているかをチェックする。格納されていた場合（例えばテーブル3）、ABCDが格納されているインデックスは0001であり、エントロピー計算から、1ビットに削減され、ABCDが1ビットの1に圧縮できることを示している。このとき、パイプラインに先行する圧縮データと矛盾が発生するため、先行する三つの圧縮データを無効化している。(3, 1)という圧縮データはパイプラインを伝って出力されていく。

パイプラインの最後では出力選択があり、圧縮データがセットされているときはそれを出力する。圧縮データの出力は、圧縮されたテーブル番号と圧縮データのペアで構成され、例えば、最終段にある1,1が出力されている。これはテーブル1で圧縮されていて、そのデータが1ビットの1であることを示している。出力されたデータを先頭から順に同様な操作であるが逆順に適用することで、データの入力を止めることなく、長いデータパターンに一致させながらそのパターンを最少で1ビットに圧縮できている。

## 用語解説

### 注1) データストリーム

温度計やGPS、カメラ映像といった連綿と流れるデータを出力するデータ形式で、データの終わりが定義できないデータ。対する概念は、楽曲や映像作品など、終わりが定まっているデータ。従来の圧縮技法はデータの始めから終わりまでのデータを調べて、頻出するデータパターンを調べているが、データストリームは終わりが無いためデータの頻出度合いを調べられなかった。

### 注2) シンボル

本研究では、単位データ量のことをシンボルと呼ぶ。例えば、映像などは8ビットを単位として、赤・緑・青の要素がストリームとしてカメラから出力される。この8ビットの単位のデータをシンボルと呼んでいる。

### 注3) データの複雑性 (エントロピー)

データが複雑である、という抽象的な情報量を数値化したもの。例えば、ABAB というシンボルの列は2種類からなるので、エントロピーは2と表される。2進数の数値ではAを0、Bを1とすれば0101と表せるので、4ビットで表すことができる。AABCDの場合、エントロピーは4(2進数では2ビット)となり、Aを00、Bを01、Cを10、Dを11とすれば、000001101100と12ビットで表せる。データ列全てのエントロピーを把握できれば、例えばシンボルが8ビットであれば前者は1/8に、後者は1/4に圧縮できる。

### 注4) エントロピー計算

変換テーブルで扱うシンボルパターンが現れる頻度を予測する数値計算で、最低限のインデックスのビット数が圧縮データに必要なことを示す。本研究の圧縮器ではデータストリームに現れるシンボルをテーブルに登録しながら、テーブルに既に登録されているときにはそのインデックスに変換して圧縮する。このテーブルの使用数の対数が圧縮対象のデータを読み込んだ瞬間でのデータの複雑さを表すため、この計算による数値のビット数にインデックスの桁数を削減できることを発見した。

## 研究資金

本研究は、JST さきがけ(「IoTが拓く未来」領域、JPMJPR203A)、JST AIP 加速課題(JPMJCR24U4)、科研費(23H03359)による研究プロジェクトの一環として実施されました。

## 掲載論文

【題名】 Universal Adaptive Stream-based Entropy Coding

(ユニバーサル性をもつ適応的なストリーム指向のエントロピー符号化)

【著者名】 S. Yamagiwa and T.Kato

【掲載誌】 *IEEE Access*

【掲載日】 2024年7月16日(オンライン先行公開)

【DOI】 DOI: 10.1109/ACCESS.2024.3429389

**問合わせ先**

**【研究に関すること】**

山際 伸一（やまぎわ しんいち）

筑波大学システム情報系 准教授

個人URL: <https://www.cs.tsukuba.ac.jp/~yamagiwa/>

研究室URL: <https://www.edge.cs.tsukuba.ac.jp/>

**【取材・報道に関すること】**

筑波大学広報局

TEL: 029-853-2040

E-mail: [kohositu@un.tsukuba.ac.jp](mailto:kohositu@un.tsukuba.ac.jp)

科学技術振興機構広報課

TEL: 03-5214-8404

E-mail: [jstkoho@jst.go.jp](mailto:jstkoho@jst.go.jp)

**【JST事業に関すること】**

前田 さち子（まえだ さちこ）

科学技術振興機構戦略研究推進部 ICTグループ

TEL: 03-3512-3526 Fax : 03-3222-2066

E-mail: [presto@jst.go.jp](mailto:presto@jst.go.jp)