(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2020-141402
(P2020-141402A)
(43)公開日　令和2年9月3日(2020.9.3)

(51)Int.Cl.		F I			テーマコード（参考）
H03M 7/30	(2006.01)	H03M	7/30	Z	5B033
H03M 7/40	(2006.01)	H03M	7/40		5B160
G06F 12/04	(2006.01)	G06F	12/04	530	5J064
G06F 9/34	(2006.01)	G06F	9/34	350A	

審査請求 未請求　請求項の数 33　ＯＬ　外国語出願　（全 29 頁）

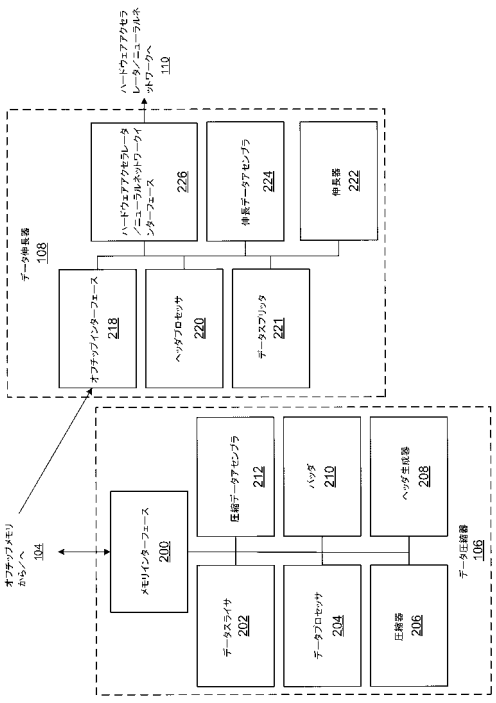
(21)出願番号　特願2020-30252(P2020-30252)
(22)出願日　令和2年2月26日(2020.2.26)
(31)優先権主張番号　16/288026
(32)優先日　平成31年2月27日(2019.2.27)
(33)優先権主張国・地域又は機関
米国(US)

(71)出願人　519410758
モヴィディウス　リミテッド
オランダ国　スヒップホル　レイク　１１１９，カプローニ　ラーン　３７
(74)代理人　100107766
弁理士　伊東　忠重
(74)代理人　100070150
弁理士　伊東　忠彦
(74)代理人　100091214
弁理士　大貫　進介
(72)発明者　ヴァレンティナ　リゴ
アイルランド国　ダブリン　８，ディーＯ８エックス４９ダブリュ，フォークナーズ　テラス　１９

最終頁に続く

(54)【発明の名称】データを圧縮する方法および装置

(57)【要約】　（修正有）
【課題】データを圧縮する方法、装置、システムおよび製造物品を提供する。
【解決手段】装置は、複数のデータブロックのうちの第１のブロックに対する第１の圧縮技術を、第１のブロックの第１の特性に基づいて選択し、複数のデータブロックのうちの第２のブロックに対する第２の圧縮技術を、第２のブロックの第２の特性に基づいて選択するプロセッサと、第１の圧縮技術を使用して第１のブロックを圧縮して第１の圧縮データブロックを生成する第１の圧縮器と、第２の圧縮技術を使用して第２のブロックを圧縮して第２の圧縮データブロックを生成する第２の圧縮器と、第１の圧縮技術を特定する第１のヘッダおよび第２の圧縮技術を特定する第２のヘッダを生成するヘッダ生成器と、オフチップメモリに記憶される、第１のヘッダを伴う第１の圧縮データブロックおよび第２のヘッダを伴う第２の圧縮データブロックを送信するインターフェースと、を含む。
【選択図】図２



【特許請求の範囲】
【請求項１】
データを記憶するオフチップメモリと、
データセットを複数のデータブロックに分割するデータスライサと、
データプロセッサであって、
前記複数のデータブロックのうちの第１のブロックに対する第１の圧縮技術を、該第１のブロックの第１の特性に基づいて選択し、
前記複数のデータブロックのうちの第２のブロックに対する第２の圧縮技術を、該第２のブロックの第２の特性に基づいて選択する、データプロセッサと、
前記第１の圧縮技術を使用して前記第１のブロックを圧縮して第１の圧縮データブロックを生成する第１の圧縮器と、
前記第２の圧縮技術を使用して前記第２のブロックを圧縮して第２の圧縮データブロックを生成する第２の圧縮器と、
前記第１の圧縮技術を特定する第１のヘッダおよび前記第２の圧縮技術を特定する第２のヘッダを生成するヘッダ生成器と、
前記オフチップメモリに記憶される、前記第１のヘッダを伴う前記第１の圧縮データブロックおよび前記第２のヘッダを伴う前記第２の圧縮データブロックを送信するインターフェースと、を含む装置。
【請求項２】
前記データプロセッサは、前記ブロックのデータが同じ値に対応する場合に、前記第１の圧縮技術を選択し、
前記ブロックの前記データが少なくとも１つの異なる値を含む場合に第２の圧縮技術を選択するものである、請求項１に記載の装置。
【請求項３】
前記第１の圧縮技術は、ハフマン圧縮技術、ＲＬＥ圧縮技術または圧縮技術なしのうちの少なくとも１つである、請求項１に記載の装置。
【請求項４】
前記データプロセッサは、ハフマン圧縮技術が前記圧縮技術なしよりも多くのデータをもたらす場合に、前記第１の圧縮技術を前記圧縮技術なしと選択するものである、請求項３に記載の装置。
【請求項５】
前記第１の圧縮器は、前記第１の圧縮データブロックを第１の圧縮データのサブブロックに分割するものである、請求項１に記載の装置。
【請求項６】
前記第１の圧縮データのサブブロックは、サイズが不均一である、請求項５に記載の装置。
【請求項７】
前記サブブロックにパターンを追加することによって、該パターンで前記サブブロックをパッドする１つ以上のパッダをさらに含む、請求項５に記載の装置。
【請求項８】
前記データブロックは、サイズが均一である、請求項１に記載の装置。
【請求項９】
前記第１の圧縮器および前記第２の圧縮器は、前記第１の圧縮技術および前記第２の圧縮技術をそれぞれのブロックに対して並列に実行するものである、請求項１に記載の装置。
【請求項１０】
前記第１の圧縮器は、選択された前記第１の圧縮技術が圧縮技術なしよりも多くのデータをもたらす場合に、前記第１の圧縮データブロックを破棄し、非圧縮の前記データブロックを保持するものである、請求項１に記載の装置。
【請求項１１】

前記第１のヘッダは、前記圧縮技術なしを特定する、請求項１０に記載の装置。
【請求項１２】
実行されると、少なくとも機械に、
データセットを複数のデータブロックに分割し、
前記複数のデータブロックのうちの第１のブロックに対する第１の圧縮技術を、該第１のブロックの第１の特性に基づいて選択し、
前記複数のデータブロックのうちの第２のブロックに対する第２の圧縮技術を、該第２のブロックの第２の特性に基づいて選択し、
前記第１の圧縮技術を使用して前記第１のブロックを圧縮して第１の圧縮データブロックを生成し、
前記第２の圧縮技術を使用して前記第２のブロックを圧縮して第２の圧縮データブロックを生成し、
前記第１の圧縮技術を特定する第１のヘッダおよび前記第２の圧縮技術を特定する第２のヘッダを生成し、
前記第１のヘッダを伴う前記第１の圧縮データブロックおよび前記第２のヘッダを伴う前記第２の圧縮データブロックをチップメモリに記憶する、
ことをさせる命令を含む非一時的なコンピュータ読み取り可能記憶媒体。
【請求項１３】
前記命令は、前記ブロックのデータが同じ値に対応する場合に、前記機械に前記第１の圧縮技術を選択させ、前記ブロックの前記データが少なくとも１つの異なる値を含む場合に、前記第２の圧縮技術を選択させる、請求項１２に記載のコンピュータ読み取り可能記憶媒体。
【請求項１４】
前記第１の圧縮技術は、ハフマン圧縮技術、ＲＬＥ圧縮技術または圧縮技術なしのうちの少なくとも１つである、請求項１２に記載のコンピュータ読み取り可能記憶媒体。
【請求項１５】
前記命令は、前記機械に、ハフマン圧縮技術が前記圧縮技術なしよりも多くのデータをもたらす場合に、前記第１の圧縮技術を前記圧縮技術なしと選択させる、請求項１４に記載のコンピュータ読み取り可能記憶媒体。
【請求項１６】
前記命令は、前記機械に、前記第１の圧縮データブロックを第１の圧縮データのサブブロックに分割させる、請求項１２に記載のコンピュータ読み取り可能記憶媒体。
【請求項１７】
前記第１の圧縮データのサブブロックは、サイズが不均一である、請求項１６に記載のコンピュータ読み取り可能記憶媒体。
【請求項１８】
前記命令は、前記機械に、前記サブブロックにパターンを追加することによって、該パターンで前記サブブロックをパッドさせる、請求項１６に記載のコンピュータ読み取り可能記憶媒体。
【請求項１９】
前記データブロックは、サイズが均一である、請求項１２に記載のコンピュータ読み取り可能記憶媒体。
【請求項２０】
前記命令は、前記機械に、前記第１の圧縮技術および前記第２の圧縮技術をそれぞれのブロックに対して並列に実行させる、請求項１２に記載のコンピュータ読み取り可能記憶媒体。
【請求項２１】
前記命令は、前記機械に、選択された前記第１の圧縮技術が圧縮技術なしよりも多くのデータをもたらす場合に、前記第１の圧縮データブロックを破棄させ、非圧縮データブロックを保持させる、請求項１２に記載のコンピュータ読み取り可能記憶媒体。

【請求項２２】
前記第１のヘッダは、圧縮技術なしを特定する、請求項２０に記載のコンピュータ読み取り可能記憶媒体。
【請求項２３】
データセットを複数のデータブロックに分割することと、
前記複数のデータブロックのうちの第１のブロックに対する第１の圧縮技術を、該第１のブロックの第１の特性に基づいて選択することと、
前記複数のデータブロックのうちの第２のブロックに対する第２の圧縮技術を、該第２のブロックの第２の特性に基づいて選択することと、
前記第１の圧縮技術を使用して前記第１のブロックを圧縮して第１の圧縮データブロックを生成することと、
前記第２の圧縮技術を使用して前記第２のブロックを圧縮して第２の圧縮データブロックを生成することと、
前記第１の圧縮技術を特定する第１のヘッダおよび前記第２の圧縮技術を特定する第２のヘッダを生成することと、
前記第１のヘッダを伴う前記第１の圧縮データブロックおよび前記第２のヘッダを伴う前記第２の圧縮データブロックをチップメモリに記憶することと、を含む方法。
【請求項２４】
前記第１の圧縮技術は、前記ブロックのデータが同じ値に対応する場合に選択され、前記第２の圧縮技術は、前記ブロックの前記データが少なくとも１つの異なる値を含む場合に選択される、請求項２３に記載の方法。
【請求項２５】
前記第１の圧縮技術は、ハフマン圧縮技術、ＲＬＥ圧縮技術または無圧縮技術なしのうちの少なくとも１つである、請求項２３に記載の方法。
【請求項２６】
前記第１の圧縮技術は、ハフマン圧縮技術が前記圧縮技術なしよりも多くのデータをもたらす場合に、圧縮技術なしと選択される、請求項２５に記載の方法。
【請求項２７】
前記第１の圧縮データブロックを、第１の圧縮データのサブブロックに分割することをさらに含む、請求項２３に記載の方法。
【請求項２８】
前記第１の圧縮データのサブブロックは、サイズが不均一である、請求項２７に記載の方法。
【請求項２９】
前記サブブロックにパターンを追加することによって、該パターンで前記サブブロックをパッドすることをさらに含む、請求項２７に記載の方法。
【請求項３０】
前記データブロックは、サイズが均一である、請求項２３に記載の方法。
【請求項３１】
前記第１の圧縮技術および前記第２の圧縮技術が、それぞれのブロックに対して並列に実行される、請求項２３に記載の方法。
【請求項３２】
選択された前記第１の圧縮技術が、圧縮技術なしよりも多くのデータをもたらす場合に、前記第１の圧縮データブロックを廃棄し、非圧縮データブロックを保持することをさらに含む、請求項２３に記載の方法。
【請求項３３】
前記第１のヘッダは、前記圧縮技術なしを特定する、請求項３２に記載の方法。
【発明の詳細な説明】
【技術分野】
【０００１】

この開示は、概してプロセッサに関し、より詳細には、データを圧縮する方法および装置に関する。

【背景技術】

【０００２】

近年、プロセッサチップ上に（例えば、オンチップで）実装されるハードウェアベースのニューラルネットワークが、種々のデータ処理システムにおいて人気が高まっている。ニューラルネットワークは、機械学習アルゴリズムが連帯して、大規模および／または複雑なデータを処理するためのフレームワークを提供する。いくつかの例では、ニューラルネットワークは、オンチップのニューラルネットワークの利用可能な容量を増大させるために、オフチップメモリ（例えば、プロセッサチップ外に位置するメモリ）に接続されている。

【図面の簡単な説明】

【０００３】

【図１】例示的なデータ圧縮器およびデータ伸長器を含む図のブロック実装である。

【図２】図１のデータ圧縮器およびデータ伸長器の例示的な実装のブロック図である。

【図３】図１および／または図２の例示的なデータ圧縮器を実装してデータセットを圧縮するために実行され得る例示的な機械読み取り可能命令を表すフローチャートである。

【図４】図１および／または図２の例示的なデータ圧縮器を実装してデータセットを圧縮するために実行され得る例示的な機械読み取り可能命令を表すフローチャートである。

【図５】図１および／または図２の例示的なデータ伸長器を実行して圧縮データセットを伸長するために実行され得る例示的な機械読み取り可能命令を表すフローチャートである。

【図６】図１および／または図２のデータ伸長器によって実行され得る例示的なデータ伸長プロセスを示す。

【図７】図１および／または図２の例示的なデータ圧縮器を実装するために、図３～図４の命令を実行するように構成された例示的な処理プラットフォームのブロック図である。

【図８】図１および／または図２の例示的なデータ伸長器を実装するために、図５の命令を実行するように構成された例示的な処理プラットフォームのブロック図である。

【０００４】

図面は縮尺通りではない。一般に、同一または類似の部分を参照するために図面および添付の発明を実施するための形態の全体を通じて、同一の参照番号を使用する。

【発明を実施するための形態】

【０００５】

パーソナルコンピュータおよび／またはモバイルデバイスを含む典型的なコンピューティングシステムは、先進的な画像処理またはコンピュータビジョンアルゴリズムを採用して、人間の視覚が行うことができるタスクを自動化する。コンピュータビジョンタスクには、デジタル画像の取得、処理、分析、および／または理解が含まれる。このようなタスクは、部分的に、デジタル画像から寸法データを抽出し、数値情報および／または記号情報を生成することを容易にする。コンピュータビジョンアルゴリズムは、数値情報および／または記号情報を使用して、意思決定を行い、そうでなければ、とりわけ３次元（３－Ｄ）姿勢推定、イベント検出、オブジェクト認識、ビデオトラッキング等に関連する動作を実行する。したがって、拡張現実（ＡＲ）、仮想現実（ＶＲ）、ロボット工学および／または他の用途をサポートするためには、このようなタスクを迅速に（例えばリアルタイムまたはほぼリアルタイムで）かつ効率的に実行することが重要である。

【０００６】

高度な画像処理またはコンピュータビジョンアルゴリズムは、ハードウェアアクセラレータおよび／または畳み込みニューラルネットワーク（ＣＮＮまたはＣｏｎｖＮｅｔ）を採用することがある。ハードウェアアクセラレータは、汎用中央処理装置（ＣＰＵ）によって実装されるソフトウェアよりも効率的に機能（例えば、画像の分類、画像のクラスタ化、オブジェクト認識の実行等）を実行するハードウェア構成要素である。ＣＮＮは、画

像の分類、類似性による画像のクラスタ化（例えば、フォトサーチ）、および／または畳み込みを使用した画像内での物体認識の実行に使用できるディープ人工ニューラルネットワークである。本明細書で使用されるように、畳み込みは、一方の関数の形状が他方の関数の形状によってどのように修飾されるかを表現する積分によって、２つの与えられた関数から導出される関数であると定義される。したがって、ＣＮＮは、入力画像内の画像特徴の一致を特定するために、入力画像の画像特徴（例えば、水平線、２次元（２－Ｄ）形状等）に対応する１つ以上のフィルタの出力を通すことによって、入力画像に含まれる顔、個人、道路標識、動物等を特定するために使用され得る。

【０００７】

いくつかのハードウェアアクセラレータおよび／またはＣＮＮは制限されたメモリ容量を有する。したがって、プロセッサチップに配置されたこのようなハードウェアアクセラレータおよび／またはＣＮＮは、メモリ（例えば、ダイナミックランダムアクセスメモリ（ＤＲＡＭ）のようなオフチップメモリ）に接続されてもよく、メモリハードウェアアクセラレータおよび／またはＣＮＮの増加したデータ処理パワーをサポートする容量を増加させる。しかし、オフチップメモリからオンチップハードウェアアクセラレータおよび／またはＣＮＮへのデータの転送は、限定された帯域幅に対応するボトルネック効果を生成し、ハードウェアアクセラレータおよび／またはＣＮＮの処理パワーを制限する。いくつかの技術は、オンチップハードウェアアクセラレータおよび／またはＣＮＮに移動される必要があるデータ量を低減する可逆圧縮技術（例えば、ハフマン圧縮）を使用して、オフチップメモリインターフェースの帯域幅要件を低減する。しかし、このような技術は、いくつかのハードウェアアクセラレータおよび／またはＣＮＮが動作する周波数および／または復号器が動作する必要がある周波数とは互換性がない。

【０００８】

本明細書で開示する例は、ハードウェアアクセラレータおよび／またはＣＮＮの周波数と互換性のあるハードウェア伸長実装を可能にしながら、可逆符号化技術によって提供される帯域幅改善を最適化するためのオフチップデータ圧縮技術を含む。本明細書で開示する例は、オフチップメモリに格納されるデータをある数のデータのブロックに分割（例えば、グループ化）し、各ブロックのデータの特性に基づいてブロックごとに圧縮技術（例えば、ランレングス符号化、一部ハフマン符号化、無符号化、および／または任意の他の可逆圧縮技術）を適用することを含む。

【０００９】

たとえば、ブロック内の全てのデータが同じ文字（例えば、全ての「０」、全ての「１」、全ての「ａ」）に対応する場合、本明細書で開示する実施例はＲＬＥ符号化を使用してブロックを圧縮する。ＲＬＥ符号化は、全てのブロックデータを、元のブロックの文字数の情報を含むブロックの全ての値を表す単一の値に圧縮する可逆データ圧縮である。例えば、ブロックが１６文字のデータブロックであり、各文字が「０」に対応する場合、ＲＬＥ符号化は、データブロックの元のサイズを示す文字（例えば１６）を含む１６個の「０」の文字を表す単一の文字「０」にサブブロックを圧縮する。

【００１０】

ブロック内のデータが全て同じ文字に対応するわけではない場合、本明細書に開示された実施例は、ハフマン技術（例えば、完全ハフマン技術または一部ハフマン技術）を実施してもよい。ハフマン技術は、可変長符号を入力文字に割り当てる可逆データ圧縮／符号化アルゴリズムであり、符号の長さは、対応する入力文字の周波数に基づいている。従って、ハフマン技術は、不均一な長さコードをもたらす。例えば、ブロックの頻度の高い文字は、ブロックの頻度の低い文字よりも短い長さコードに割り当てられる。従って、ハフマン技術は、繰り返し文字を備えたデータブロックのサイズを低減する。ハフマン符号化を適用するために、データセット内の各文字の出現頻度に基づいてハフマンツリーを生成する。データブロックの各文字は、ハフマンツリーに基づくコードに割り当てられる。ほとんどの場合、ハフマン符号化技術を使用してデータを符号化すると、可逆データ圧縮をもたらす。

【００１１】
一部ハフマン符号化技術は、ハフマン符号化技術をプリセットされた文字数のみに適用し、残りの文字は非符号化のままとする。次いで、各文字は、符号化されたあるいは符号化されていないとしてマークまたはタグ付けされる。このようにして、復号時間は短縮され、ハフマン符号化データを復号するために必要とされるハフマンテーブルのサイズは縮小される。ハフマンテーブルのサイズを縮小すると、符号化データと共に復号器に送信される必要のあるデータ量が低減される。追加的に、ハフマンテーブルを縮小すると、復号速度を増加させるハードウェア解決を容易にする。本明細書に開示された実施例は、概して、一部ハフマン技術に対応する。しかし、完全ハフマン技術を代替的に使用することができる。

【００１２】
追加的に、本明細書に開示された例は、ハフマン符号化データブロックを符号化データサブブロックに分割する。このような符号化データサブブロックは、各サブブロックの末端にプリセットされたコードを追加することによってパディングされ、オンチップ復号器は並列復号のためにサブブロックを分割することができ、それによって計算帯域幅を増加させることができる。いくつかの例では、ブロック内に繰り返し文字がほとんどないか、またはまったくないような場合に、一部ハフマン技術は、含まれるヘッダに対応するオーバーヘッド量のために、データブロックのサイズを縮小しないことがある。従って、本明細書で開示された例は、一部ハフマン技術がデータブロックのサイズを縮小しないブロックについては、いかなる圧縮／符号化技術も実行しない。

【００１３】
本明細書に開示された例を用いて、ユーザは、より高い圧縮比を達成する特定のアプリケーションデータプロファイルに合わせた圧縮スキームを調整することができる。追加的に、本明細書に開示された例は、ハードウェアメモリ制限および速度要件と互換性があるように、ハフマン符号化ルックアップテーブルの要件を低減する。追加的に、本明細書に開示された例は、復号および／または符号並列化を可能にし、それによって高い復号帯域幅に対応する。

【００１４】
図１は、例示的なオフチップ構成要素１００および例示的なオンチップ構成要素１０２のブロック図である。例示的なオフチップ構成要素１００は、データ／データセットを記憶する例示的なオフチップメモリ１０４と、ハードウェアベースの処理のために例示的なオンチップ構成要素１０２に送信する前に、例示的なオフチップメモリ１０４に記憶されたデータセットを圧縮するレジ的なデータ圧縮器１０６とを含む。例示的なオンチップ構成要素１０２は、圧縮データセットを伸長する例示的なデータ伸長器１０８と、データセットを処理する例示的なハードウェアアクセラレータ／ニューラルネットワークエンジン１１０とを含む。図１の例示的なブロック図は、オフチップ構成要素１００を含むが、例示的なオフチップメモリ１０４および／または例示的なデータ圧縮器１０６は、例示的なオンチップ構成要素１０２と同じチップに配置されてもよい（例えば、図１の全ての構成要素が同じチップに配置されてもよい）。

【００１５】
図１の例示的なオフチップ構成要素１００は、ハードウェアアクセラレータ／ニューラルネットワーク１１０によって処理されるデータを記憶するオフチップメモリ１０４（例えば、ＤＲＡＭ）を含む。例示的なオフチップメモリ１０４は、ハードウェアアクセラレータ／ニューラルネットワーク１１０によって処理され得る追加データを記憶する追加メモリを提供することによって、例示的なハードウェアアクセラレータ／ニューラルネットワーク１１０のメモリ容量を増加させる。上述のように、オフチップメモリ１０４からハードウェアアクセラレータ／ニューラルネットワーク１１０へ非圧縮データを転送すると、オフチップメモリ１０４（例えばＤＲＡＭ）の帯域幅制限によりボトルネックを生成する。従って、例示的なデータ圧縮器１０６は、オンチップ構成要素１０２に送信する前に、オフチップメモリ１０４に記憶され、かつ／あるいは既に記憶されているデータを圧縮

して、データのサイズを縮小し、ボトルネックの問題を低減および／または除去する。例示的なデータ圧縮器１０６は、記憶されたデータをデータのブロックに分割し、ブロックのデータの特性に基づいて３つの圧縮／符号化技術を使用してデータを圧縮する。追加的に、例示的なデータ圧縮器１０６は、使用される圧縮／符号化技術を特定するブロックのヘッダ（例えば、第１のブロックに対する第１のヘッダ、第２のブロックに対する第２のヘッダ等）を生成し、例示的なデータ伸長器１０８がブロックを伸長するのに使用することができる情報（例えば、ブロックサイズ、符号化テーブル、サブブロックサイズ等）を提供する。このようにして、例示的なデータ伸長器１０８が圧縮データを適切に伸長することができる。圧縮されると、例示的なデータ圧縮器１０６は、例示的なオフチップメモリ１０４に圧縮データを記憶する。このようにして、オフチップメモリ１０４は、例示的なハードウェアアクセラレータ／ニューラルネットワーク１１０によって処理される圧縮データを送信することができる。いくつかの例では、データ圧縮器１０６は、複数のデータブロックを並列に圧縮する複数のプロセッサを含み、それによって計算帯域幅を増加させる。例示的なデータ圧縮器１０６は、図２に関連して以下にさらに説明される。

【００１６】

図１の例示的なオンチップ構成要素１０２は、オフチップ構成要素１００とインターフェースで接続して、オフチップメモリ１０４から圧縮データを受信し、ハードウェアアクセラレータ／ニューラルネットワーク１１０に送信される前に圧縮データを伸長するデータ伸長器１０８を含む。いくつかの例では、データ伸長器１０８はハードウェアアクセラレータである。例示的なデータ伸長器１０８は、圧縮データを受信し、圧縮データのブロックのヘッダを特定して、圧縮技術を決定する。このようにして、データ伸長器１０８は、各ブロックのヘッダで特定された圧縮技術に基づいて圧縮されたブロックを伸長する。伸長されると、例示的なデータ伸長器１０８は、例示的なハードウェアアクセラレータ／ニューラルネットワーク１１０に伸長データを送信する。例示的なハードウェアアクセラレータ／ニューラルネットワーク１１０は、ハードウェアアクセラレータ／ニューラルネットワーク１１０の設定に基づいて伸長データを処理する。例示的なデータ伸長器１０８は、図２に関連して以下にさらに説明する。

【００１７】

図２は、例示的なデータ圧縮器１０６の実装例のブロック図と、図１の例示的なデータ伸長器１０８の実装例のブロック図である。例示的なデータ圧縮器１０６は、例示的なメモリインターフェース２００と、例示的なデータスライサ２０２と、例示的なデータプロセッサ２０４と、例示的な圧縮器２０６と、例示的なヘッダ生成器２０８と、例示的なパッダ２１０と、例示的な圧縮データアセンブラ２１２とを含む。例示的なデータ伸長器１０８は、例示的なオフチップインターフェース２１８と、例示的なヘッダプロセッサ２２０と、例示的なデータスプリッタ２２１と、例示的な伸長器２２２と、例示的な伸長データアセンブラ２２４と、例示的なハードウェアアクセラレータ／ニューラルネットワークインターフェース２２６とを含む。

【００１８】

図２の例示的なメモリインターフェース２００は、圧縮され、例示的なオンチップデータ伸長器１０８に送信されるデータを図1の例示的なオフチップメモリ１０４から受信する。データ圧縮器１０６がデータを圧縮すると、図２の例示的なメモリインターフェース２００は、例示的なオフチップメモリ１０４に圧縮データを送信し、圧縮データが図１のオンチップ構成要素１０２によって検索され得るようにする。

【００１９】

メモリインターフェース２００が例示的なオフチップメモリ１０４からデータを取得すると、例示的なデータスライサ２０２は、データをデータブロックに分割（例えばスライス）する。いくつかの例では、データスライサ２０２がデータを均一な固定サイズブロックに分割する。追加的に、データスライサ２０２が符号化データ（例えば、一部ハフマン圧縮を使用して符号化されたブロック）を符号化データサブブロックにスライスして、以下にさらに説明するように、例示的なデータ伸長器１０８の並列伸長処理を可能にしても

よい。いくつかの例では、データスライサ２０２は、２つのデータスライサを含む。例えば、元のデータをデータブロックに分割する第１のデータスライサと、ハフマン符号化データを符号化データサブブロックに分割する第２のデータスライサとがある。

【００２０】

図２の例示的なデータプロセッサ２０４は、ブロックのデータの特性に基づいて、データブロックに使用する圧縮技術（例えば、ＲＬＥ符号化、ハフマン符号化、符号化なし、および／または、任意の他の可逆圧縮技術）を決定（例えば、選択）する１つ以上のデータプロセッサである。例えば、データプロセッサ（複数可）２０４がブロックを取得する場合に、データプロセッサ２０４は、ブロックの全てのデータが同じ値および／または文字（例えば、全ての「０」、全て「１」、全て「ａ」等）に対応するかどうかを決定する。データプロセッサ２０４が、ブロックのデータは同じ値／文字に対応すると決定した場合、データプロセッサ２０４は、情報の損失なしにブロックを圧縮するために、ＲＬＥ符号化技術がブロックに適用され得ると決定する。データプロセッサ２０４が、ブロックのデータは異なる値／文字に対応すると決定した場合、データプロセッサ２０４は、（例えば、一部または完全）ハフマン符号化または符号化なしがブロックに適用されるべきであると決定し得る。上述したように、ハフマン符号化技術をいくつかのブロックに適用すると、（例えば、ブロックが繰り返される値／文字をほとんどあるいは全く持たない場合に）より大きなデータ量をもたらすことがある。したがって、いくつかの例において、例示的なデータプロセッサ２０４がデータブロックを処理して、データブロックの一部ハフマン符号化技術は、より小さなあるいはより大きなデータ量をもたらすかどうかを決定してもよい。このような例において、例示的なデータプロセッサ２０４が、データブロックの一部ハフマン符号化が、（例えば、確率的モデリングを使用して）符号化なしよりも多くのデータをもたらすと決定した場合、データプロセッサ２０４は、符号化なしがデータブロックに適用されると決定する。他の例では、圧縮器２０６は、一部ハフマン符号化をデータに適用し、データプロセッサ２０４は、結果が、圧縮技術を適用しない場合（例えば、符号化なし）よりも大きいか小さいかを決定する。データプロセッサ２０４が、結果が圧縮技術を適用しない場合よりも大きいと決定した場合、データプロセッサ２０４は、ハフマン符号化データを破棄し、それによってデータブロックを非符号化のまま（例えば、非圧縮）にする。データプロセッサ２０４は、単一のデータプロセッサであってもよく、複数のデータプロセッサあってもよい。例えば、複数のデータプロセッサ２０４が、データブロックを並列に処理するために利用され得る。

【００２１】

図２の例示的な圧縮器２０６は、データプロセッサ２０４によって選択された技術に基づいて、データブロックに対して圧縮技術（例えば、ＲＬＥ符号化、ハフマン符号化、および／または任意の他の可逆圧縮技術）を実行する。上述のように、データプロセッサ２０４は、ブロックの全てのデータが同じ値および／または文字（例えば、全ての「０」、全ての「１」、全ての「ａ」等）に対応する場合に、ＲＬＥ符号化を選択する。データプロセッサ２０４がデータブロックに対してＲＬＥを選択する場合に、データプロセッサ２０４は、ヘッダ生成器２０８に、ブロックの繰り返し値および／または値の繰り返し数を特定するヘッダまたはデータパケットを生成するように指示する。例えば、ブロックが、各ビットが「１」に対応する８文字ブロックである場合、データプロセッサ２０４は、ヘッダ生成器２０８に、８文字の各々が「１」の値に対応し、「１」が８回繰り返されることの表示を特定するヘッダを生成するように指示する。

【００２２】

図２のデータプロセッサ２０４がハフマン符号化または一部ハフマン符号化を選択する場合に、例示的な圧縮器２０６はハフマン符号化または一部ハフマン符号化をデータブロックに適用する。上述のように、ハフマン符号化を適用するために、ハフマンツリーがデータセットにおける各文字の出現頻度に基づいて生成される。ブロックのデータの各文字は、ハフマンツリーに基づいて符号に割り当てられる。ほとんどの場合、ハフマン符号化技術を使用してデータを符号化すると、可逆データ圧縮をもたらす。一部ハフマン符号化

技術には、ハフマン符号化技術をプリセットされた文字数にだけ適用し、残りの文字は非符号化のままにすることが含まれ、それによって復号時間を短縮し、ハフマン符号化データを復号するのに必要なハフマンテーブルのサイズを縮小する。符号化データ内の文字の前には、その文字が符号化されているかどうかを特定するマーカーが付加される。いくつかの例では、データプロセッサ２０４は、ヘッダオーバヘッドとデータ圧縮との間のトレードオフを含めたハフマン符号化においてどれをプリセットされた文字にするか、かつ／あるいはいくつプリセットされた文字にするかを決定する。例えば、データプロセッサ２０４が、一部ハフマン技術は２３６個の可能な文字のうちの８文字のみに適用され得ると決定してもよい。従って、データプロセッサ２０４は、最も頻繁に繰り返されるデータのうちの８つの文字を決定し、圧縮器２０６は、８つの決定された文字を使用した一部ハフマン符号化を適用するために、８つの決定された文字に対してシンボルを適用する。一部ハフマン符号化技術を適用することにより、より高速な圧縮が可能となり、メモリ１０４は、ハフマン符号化技術を使用して圧縮されたブロックを伸長（例えば、復号）するために使用されるより小さなハフマンテーブルを送信することが可能となる。圧縮器２０６は、単一の圧縮器であってもよく、複数の圧縮器であってもよい。例えば、複数の圧縮器２０６が、データブロックを並列に符号化（例えば、圧縮）するために利用されてもよい。

【００２３】

図２の例示的なヘッダ生成器２０８は、データのブロックを符号化および／または圧縮するために使用される、符号化および／または圧縮タイプ（例えば、ＲＬＥ符号化、ハフマン符号化、符号化なし、および／または、任意の他の可逆圧縮技術）を特定するヘッダおよび／またはデータパケットを生成する。たとえば、ブロックの圧縮タイプがＲＬＥ圧縮である場合、ヘッダ生成器２０８はＲＬＣ圧縮並びにＲＬＥ圧縮に対応する繰り返し文字、およびその文字が繰り返しされる回数を特定するヘッダおよび／またはデータパケットを生成する。ブロックの圧縮タイプがハフマン圧縮または一部ハフマン圧縮である場合、ヘッダ生成器２０８は、ハフマン圧縮および／または一部ハフマン圧縮技術に対応するヘッダおよび／またはデータパケットを生成する。例えば、ヘッダ生成器２０８は、ブロックがハフマン符号化されていること、データを復号するテーブル、符号化データのサイズ、サブブロックのサイズ、および／またはサブブロックが符号化されているかどうかを示す文字マーカーを含むヘッダを生成してもよい。ブロックの圧縮タイプが符号化および／または圧縮されていない場合、例示的なヘッダ生成器２０８は、圧縮なしおよび／または符号化なしに対応すると共に、非符号化のデータのサイズに対応するヘッダおよび／またはデータパケットを生成する。例示的なヘッダ生成器２０８は、符号化または非符号化ブロックにヘッダおよび／またはデータパケットを追加する。このようにして、例示的なデータ伸長器１０８は、各ブロックに対してどの圧縮技術が使用されたかを特定し、圧縮技術に基づいて圧縮ブロックを伸長することができる。ヘッダ生成器２０８は、単一のヘッダ生成器であってもよく、複数のヘッダ生成器であってもよい。例えば、ヘッダ生成器２０８は、並列にデータのブロックに対するヘッダを生成するために利用され得る。

【００２４】

ハフマン符号化または一部ハフマン符号化が、ブロックデータに適用されてハフマン符号化データを生成する場合、図２の例示的なスライサ２０２は、ハフマン符号化データを、Ｎ個の符号化サブブロックに対応するプリセット数の符号化サブブロックにスライス（例えば、グループ化）する。例えば、符号化データが２７ハフマン符号を含み、符号化サブブロックのプリセット数は９である場合、スライサ２０２は、各サブブロックに対して３ハフマン符号をグループ化する。符号化サブブロックのプリセット数は、データ伸長器１０８の並列プロセッサの数に基づいてもよい。このようにして、各並列プロセッサは、各符号化サブブロックまたはＮ個のサブブロックを並列に処理することができる。符号化サブブロックにスライスされると、例示的なパッダ２１０は、サブブロックの終わりを表す定義済みのコードパターンで個々のサブブロックをパッドする。上述するように、ハフマン符号は、長さが可変（例えば、不均一）である。例えば、第１の１６ビットのサブブロックの第１のハフマン符号は８ビットの長さであってもよい一方で、第２の１６ビット

のサブブロックの第２のハフマン符号は１０ビットの長さであってもよい。したがって、１つ以上のハフマン符号をグループ化することは、可変（例えば、不均一）長さ符号化サブブロックに対応する。したがって、パッダ２１０は、サブブロックの終わりに定義済みのコードを追加し、ヘッダ生成器２０８は、ヘッダにパディングされたサブブロックのサイズを含み、例示的なデータ伸長器１０８が、ヘッダで特定されているサイズに基づいて、個々の符号化サブブロックを分割して、並列復号を容易にするようにする。パッダ２１０は、単一のパッダであっても、複数のパッダあってもよい。例えば、パッダ２１０が、並列に符号化サブブロックにパッドするために利用されてもよい。

【００２５】

図２の例示的な圧縮データアセンブラ２１２は、生成されたヘッダと共に順番にパディングされた符号化サブブロックをアセンブルして、データブロックのハフマン符号化を完了する。追加的に、例示的な圧縮データアセンブラ２１２は、符号化（例えば圧縮）または非符号化（例えば非圧縮）データブロックをアセンブルして、例示的なデータ伸長器１０８に送信される圧縮データ全体を生成する。いくつかの例において、データブロックが並列に処理されてもよい。したがって、圧縮データアセンブラ２１２は、データブロックが符号化の前に並べられたのと同じ順序で、符号化ブロックを組み立てることができるようにブロックをモニタする。

【００２６】

図２の例示的なオフチップインターフェース２１８は、例示的なオフチップメモリ１０４から圧縮データを取得する。圧縮データが取得されると、例示的なヘッダプロセッサ２２０は圧縮データを処理して圧縮データのブロックのヘッダを特定する。上述のように、ヘッダは、圧縮データの各ブロックを圧縮するために使用された符号化技術を特定する情報を含む。したがって、ヘッダプロセッサ２２０は、データブロックの伸長のためのヘッダを特定することによって圧縮ブロックを分割することができる。

【００２７】

図２の例示的な伸長器２２２は、対応する各ブロックのヘッド内で特定された対応する圧縮情報に基づいてデータブロックを伸長する。例えば、符号化データの第１のブロックの第１のヘッダがＲＬＥ圧縮に対応する場合、伸長器２２２は第１のヘッダを処理して、どの文字が繰り返されているかと、その文字が何回繰り返されるかを特定し、決定された繰り返された文字の数を伸長データアセンブラ２２４に渡す。回数は、プリセットされていてもよいし、ヘッダ内で特定されていてもよい。符号化データの第２のブロックの第２のヘッダが符号化なしに対応する場合、伸長器２２２は、ヘッダを削除し、そのデータを伸長（例えば復号）せずに非符号化データを伸長データアセンブラに渡す。符号化データの第３のブロックの第３のヘッダが（例えば、一部または完全）ハフマン符号化に対応する場合、伸長器２２２は、ハフマン符号に対応するブロックのヘッダからハフマンテーブルを使用して、符号化データの第３のブロックを伸長する。いくつかの例では、データスプリッタ２２１は、ヘッダ内で特定されるサブブロックサイズに基づいて、圧縮データ圧縮データのブロックをサブブロックに分割する。この点、各サブグループが、複数の伸長器２２２によって並列処理されることができる。伸長されると、伸長器２２２は、伸長サブブロックを例示的な伸長データアセンブラ２２４に渡す。

【００２８】

図２の例示的な伸長データアセンブラ２２４は、復号サブブロックをアセンブルして、データブロックのハフマン復号を完了する。例えば、ハフマン符号化サブブロックを並列に復号する複数の伸長器２２２がある場合、伸長データアセンブラ２２４は、伸長サブブロックを並べて伸長ブロックデータを生成することができるように並列プロセッサをモニタする。追加的に、例示的な伸長データアセンブラ２２４は、復号（例えば伸長）データブロックをアセンブルして、例示的なハードウェアアクセラレータ／ニューラルネットワーク１１０に送信される伸長データ全体を生成する。いくつかの例では、データブロックは並列に処理される（例えば、伸長および／または復号される）。したがって、伸長データアセンブラ２２４は、データブロックが受信されたのと同じ順序で、復号ブロックをア

センブルすることができるようにブロックをモニタする。図２のハードウェアアクセラレータ／ニューラルネットワークインターフェース２２６は、図１のオンチップ構成要素１０２の例示的なハードウェアアクセラレータ／ニューラルネットワーク１１０にフルセットの伸長データを送信する。

【００２９】

図２に、図１の例示的なデータ圧縮器１０６および例示的なデータ伸長器１０８を実装する例示的な方法が示されているが、図２に示された要素、プロセスおよび／またはデバイスのうちの１つ以上が、組み合わされ、分割され、再配置され、省略され、除去され、および／または他の方法で実装されてもよい。さらに、図２の例示的なメモリインターフェース２００、例示的なデータスライサ２０２、例示的なデータプロセッサ２０４、例示的な圧縮器２０６、例示的なヘッダ生成器２０８、例示的なパッダ２１０、例示的な圧縮データアセンブラ２１２、例示的なオフチップインターフェース２１８、例示的なヘッダプロセッサ２２０、例示的なデータスプリッタ２２１、例示的な伸長器２２２、例示的な伸長データアセンブラ２２４、例示的なハードウェアアクセラレータ／ニューラルネットワークインターフェース２２６、および／またはより一般的には、例示的なデータ圧縮器１０６および／または例示的なデータ伸長器１０８が、ハードウェア、ソフトウェア、ファームウェア、および／またはハードウェア、ソフトウェア、および／またはファームウェアの任意の組み合わせによって実装されてもよい。従って、例えば、図２の例示的なメモリインターフェース２００、例示的なデータスライサ２０２、例示的なデータプロセッサ２０４、例示的な圧縮器２０６、例示的なヘッダ生成器２０８、例示的なパッダ２１０、例示的な圧縮データアセンブラ２１２、例示的なオフチップインターフェース２１８、例示的なヘッダプロセッサ２２０、例示的なデータスプリッタ２２１、例示的な伸長器２２２、例示的な伸長データアセンブラ２２４、例示的なハードウェアアクセラレータ／ニューラルネットワークインターフェース２２６、および／またはより一般的には、例示的なデータ圧縮器１０６および／または例示的なデータ伸長器１０８は、１つ以上のアナログまたはデジタル回路、論理回路、プログラマブルプロセッサ、プログラマブルコントローラ、グラフィックス処理ユニット（ＧＰＵ）、デジタル信号プロセッサ（ＤＳＰ）、特定用途向け集積回路（ＡＳＩＣ）、プログラマブル論理デバイス（ＰＬＤ）、および／またはフィールドプログラマブル論理デバイス（ＦＰＬＤ）によって実現することができる。純粋にソフトウェアおよび／またはファームウェアの実装をカバーするために、この特許の装置またはシステムの請求項のいずれかを読むと、図２の例示的なメモリインターフェース２００、例示的なデータスライサ２０２、例示的なデータプロセッサ２０４、例示的な圧縮器２０６、例示的なヘッダ生成器２０８、例示的なパッダ２１０、例示的な圧縮データアセンブラ２１２、例示的なオフチップインターフェース２１８、例示的なヘッダプロセッサ２２０、例示的なデータスプリッタ２２１、例示的な伸長器２２２、例示的な伸長データアセンブラ２２４、例示的なハードウェアアクセラレータ／ニューラルネットワークインターフェース２２６、および／またはより一般的には、例示的なデータ圧縮器１０６および／または例示的なデータ伸長器１０８は、メモリ、デジタル多用途ディスク（ＤＶＤ）、コンパクトディスク（ＣＤ）、ブルーレイ（登録商標）ディスク等、ソフトウェアおよび／またはファームウェアを含む非一時的なコンピュータ読み取り可能記憶デバイスまたは記憶ディスクを含むように本明細書に明確に定義されている。さらに、図２の例示的なデータ圧縮器１０６および／またはデータ伸長器１０８が、図２に示すものに加えて、あるいはその代わりに、１つ以上の要素、プロセスおよび／またはデバイスを含んでもよく、および／または図示の要素、プロセスおよびデバイスのいずれかまたは全てのうちの２つ以上を含んでもよい。本明細書で使用する「通信中」という語句は、その変形を含み、１つ以上の中間構成要素を介した直接的な通信および／または間接的な通信を包含し、直接的な物理的（例えば有線）通信および／または常時通信を必要とせず、むしろ、周期的な間隔、スケジュールされた間隔、非周期的な間隔、および／またはワンタイムイベントにおける選択的な通信を追加的に含む。

【００３０】

図１および／または図２の例示的なデータ圧縮器１０６を実装するための、例示的なハードウェア論理、機械読み取り可能命令、ハードウェア実装状態機械、および／またはそれらの任意の組み合わせを表すフローチャートを図３～４に示す。図１および／または図２の例示的なデータ伸長器１０８を実装するための、例示的なハードウェア論理、機械読み取り可能命令、ハードウェア実装状態機械、および／またはそれらの任意の組み合わせを表すフローチャートを図５に示す。機械読み取り可能命令が、図７および／または図８に関連して下記に論じる例示的なプロセッサプラットフォーム７００、８００に示されるプロセッサ７１２、８１２のようなコンピュータプロセッサによる実行のための実行可能プログラムまたは実行可能プログラムの一部であってもよい。プログラムが、ＣＤ－ＲＯＭ、フロッピーディスク、ハードドライブ、ＤＶＤ、ブルーレイディスク、プロセッサ７１２、８１２に関連付けられたメモリ等の非一時的なコンピュータ読み取り可能な記憶媒体に記憶されたソフトウェアで具体化されてよいが、プログラム全体および／またはその一部は、代替的に、プロセッサ７１２、８１２以外のデバイスによって実行され、かつ／あるいはファームウェアまたは専用ハードウェアで具体化することができる。さらに、例示的なプログラムが図３～図５に示すフローチャートを参照して説明されているが、図２の例示的なデータ圧縮器１０６および／またはデータ伸長器１０８を実装する多くの他の方法が代替的に使用されてもよい。例えば、ブロックの実行順序が変更されてもよく、かつ／あるいは説明したブロックのいくつかが変更、削除、または組み合わせられてもよい。追加的または代替的に、ブロックの任意または全部が、ソフトウェアまたはファームウェアを実行することなく対応する動作を実行するように構成された１つ以上のハードウェア回路（例えば、離散および／または集積アナログおよび／またはデジタル回路、ＦＰＧＡ、ＡＳＩＣ、比較器、演算増幅器（ｏｐ－ａｍｐ）、論理回路など）によって実装されてもよい。

【００３１】

上述したように、図３～図５の例示的なプロセスが、ハードディスクドライブ、フラッシュメモリ、読み出し専用メモリ、コンパクトディスク、デジタル汎用ディスク、キャッシュ、ランダムアクセスメモリ、および／または情報が任意の期間（例えば、長期間、永久的、短時間、一時的なバッファリング、および／または情報のキャッシュ）記憶される任意の他の記憶デバイスまたは記憶ディスク等の非一時的なコンピュータおよび／または機械読み取り可能媒体に記憶される実行可能な命令（例えば、コンピュータおよび／または機械読み取り可能な命令）を使用して実装されてもよい。本明細書で使用される非一時的なコンピュータ読み取り可能媒体という用語は、任意のタイプのコンピュータ読み取り可能記憶デバイスおよび／または記憶ディスクを含み、伝搬信号を排除し、伝送媒体を排除するように明確に定義される。

【００３２】

本明細書において「含む（including）」および「含む（comprising）」（ならびにその全ての形式および時制）は、オープンエンドの用語として使用される。したがって、「含む（include）」または「含む（comprise）」のいずれかの形式（例えば、comprises、includes、comprising、including、having等）を請求項がプリアンブルとして、あるいは任意の種類の請求項の規定内に採用する場合にはいつでも、追加の要素、用語等が、対応する請求項または規定の範囲外に該当することなく存在してよいと理解されるべきである。本明細書で使用される、例えば、請求項のプリアンブルにおいて「少なくとも」という語句が移行句として使用される場合に、「含む（comprising）」および「含む（including）」という語句がオープンエンドであるのと同様に、オープンエンドである。Ａ、Ｂ、および／またはＣのような形式で「および／または」という用語が使用される場合には、（１）Ａのみ、（２）Ｂのみ、（３）Ｃのみ、（４）ＡとＢ、（５）ＡとＣ、（６）ＢとＣ、（７）ＡとＢとＣ等、Ａ，Ｂ，Ｃの任意の組み合わせまたはサブセットを指す。

【００３３】

図３は、図１のオンチップ構成要素１０２に送信される前に、オフチップに記憶されたデータを圧縮する、図１および／または図２のデータ圧縮器１０６の例示的な実装によっ

て実行され得る機械読み取り可能命令を表す例示的なフローチャート３００である。図３のフローチャート３００は、図１および／または図２の例示的なデータ圧縮器１０６に関連して説明されているが、代わりに、他のタイプのデータ圧縮器および／または他のタイプのプロセッサが利用されてもよい。図３の例示的なフローチャート３００は、ＲＬＥ圧縮、ハフマン圧縮、および／または圧縮なしに関連して説明されているが、他の圧縮技術が追加的または代替的に使用されてもよい。

【００３４】

ブロック３０２において、例示的なデータスライサ２０２は、例示的なメモリインターフェースを介して、例示的なオフチップメモリ１０４からデータセットを取得する。データセットは、データセットのハードウェアベースの処理を実行する図１の例示的なオンチップハードウェアアクセラレータ／ニューラルネットワーク１１０によって使用される例示的なオフチップメモリ１０４に記憶されるデータセットである。ブロック３０４において，データスライサ２０２は、データセットをＮ個のデータブロックに分割する。ブロックのサイズおよび／または数が、ユーザおよび／または製造業者の好みに基づいてもよい。Ｎ個のデータブロックは、サイズが均一であってもなくてもよい。ブロック３０６において、例示的なデータ圧縮器１０６は、図４に関連して後述されるように、データブロックを圧縮する。

【００３５】

ブロック３０８において、例示的な圧縮データアセンブラ２１２は、圧縮されたブロックをアセンブルして圧縮データセットを生成する。例えば、データ圧縮器１０６が順序通りにブロックを圧縮する場合、データアセンブラ２１２は、ブロックが圧縮された順序と同じ順序で圧縮ブロックを取り付ける。データ圧縮器１０６がブロックを順序通りではなく、あるいは並列に圧縮する場合、データアセンブラ２１２は、圧縮データブロックをデータセットからのデータと同じ順序で再アセンブルすることができるように圧縮をモニタする。ブロック３１０において、例示的なインターフェース２０２は、圧縮データセットを図のオフチップメモリ１０４に送信し、圧縮データとして記憶される。

【００３６】

図４は、図３に関連して上述したように、データのブロックを圧縮するために、図１および／または図２のデータ圧縮器１０６の例示的な実装によって実行され得る、例示的な機械読み取り可能命令の例示的なフローチャート３０６である。図４のフローチャート３０６は、図１および／または図２の例示的なデータ圧縮器１０６に関連して説明されているが、代わりに、他のタイプのデータ圧縮器および／または他のタイプのプロセッサが使用されてもよい。図４の例示的なフローチャート３０６は、（例えば、並列データスライサ２０２、データプロセッサ２０４、圧縮器２０６、ヘッダ生成器２０８、および／またはパッダ２１０によって）線形および／または並列に、データの複数のブロックに対して実行されてもよい。図４の例示的なフローチャート３０６は、ＲＬＥ圧縮、ハフマン圧縮、および／または、圧縮なしに関連して説明されているが、他の圧縮技術が追加的または代替的に使用されてもよい。

【００３７】

データのＮブロックの各々について（ブロック４０２～４３０）、例示的なデータプロセッサ２０４は、データブロックの全てのデータが同じ値および／または文字に対応するかどうかを決定する（ブロック４０４）。例えば、データブロックが全て「ａ」であるものを含む場合、例示的なデータプロセッサ２０４は、データブロックの全てのデータが同じ文字に対応すると決定する。

【００３８】

例示的なデータプロセッサ２０４が、ブロックの全てのデータが同じ値および／または文字に対応すると決定した場合（ブロック４０４：ＹＥＳ）、例示的な圧縮器２０６は、ブロックのデータを符号化／圧縮するＲＬＥ圧縮技術を実行する（ブロック４０６）。例えば、圧縮器２０６が、ヘッダ生成器２０８に、その文字がデータのブロック内で繰り返される繰り返し文字および／またはその文字の繰り返し数を特定するデータパケットを生

成するように指示してもよい。ブロック４０８において、例示的なヘッダ生成器２０８は、データブロックがＲＬＥ圧縮技術を用いて圧縮されたことを特定する圧縮データに含めるためのデータパケットおよび／またはヘッダを生成する。このようにして、例示的なデータ伸長器１０８は、ブロックをＲＬＥ圧縮に対応するものとして特定し、ＲＬＥ伸長技術に基づいてブロックを伸長することができる。

【００３９】

例示的なデータプロセッサ２０４が、ブロックのデータの全てが同じ値および／または文字に対応していないと決定した場合（ブロック４０４：ＮＯ）、例示的な圧縮器２０６は、一部ハフマン符号化技術を実行して、データブロックを符号化／圧縮する（ブロック４１２）。いくつかの例では、圧縮器２０６が、完全ハフマン符号化技術を実行してもよい。上述したように、ハフマン符号化を適用するために、例示的な圧縮器２０６は、データのセットにおける各文字の出現頻度に基づいてハフマンツリーを生成する。圧縮器２０６は、ハフマンツリーに基づいてブロックのデータの各文字にコードを割り当てる。ほとんどの場合、ハフマン符号化技術を用いてデータを符号化すると、無符号化よりもサイズが小さい可逆データ圧縮が生じる。一部ハフマン符号化技術を適用するために、例示的な圧縮器２０６は、ハフマン符号化技術をプリセットされた文字数にのみ適用し、残りの文字を非符号化のまま残し、それによって、符号化および／または復号時間を増加させ、ハフマン符号化データを復号するのに必要なハフマンテーブルのサイズを低減させる。追加的に、例示的な圧縮器２０６が、各文字を符号化または非符号化としてマークを付ける。

【００４０】

ブロック４１４において、例示的なデータプロセッサ２０４は、符号化データおよびヘッダが、非符号化データよりも多くのビットに対応するかどうかを決定する。上述したように、ハフマンまたは一部ハフマン圧縮技術は、ブロック内の文字の繰り返しがほとんどあるいは全くないデータブロックについてはより多くのデータに対応することがある。いくつかの例において、データプロセッサ２０４が、ブロック内のデータの繰り返しに基づいて、ハフマン圧縮または一部ハフマン圧縮が、圧縮を実行する前（例えば、ブロック４１２の前）に、より多くのビットをもたらすかどうかを決定してもよい。

【００４１】

例示的なデータプロセッサ２０４が符号化データおよびヘッダが、非符号化データよりも多くのビットに対応すると決定した場合（ブロック４１４：ＹＥＳ）、例示的なデータプロセッサ２０４は、符号化データを破棄する（ブロック４１６）。ブロック４１８において、例示的なヘッダ生成器２０８は、符号化なしに対応するヘッダを生成する。このようにして、例示的なデータ伸長器１０８は、ブロックのデータが非符号化および／または非圧縮と決定することができる。ブロック４２０において、例示的なヘッダ生成器２０８は、生成されたヘッダを伴ったブロックの非符号化データを例示的な圧縮データアセンブラ２１２に渡す。

【００４２】

例示的なデータプロセッサ２０４が、符号化データは非符号化デーよりも多くのビットに対応していないと決定した場合（ブロック４１４：ＮＯ）、例示的なヘッダ生成器２０８は、ハフマンおよび／または一部ハフマン（ＰＦ）符号化に対応するヘッダを生成する（ブロック４２２）。このようにして、例示的なデータ伸長器１０８は、ハフマンおよび／またはＰＨ圧縮に対応するブロックを特定し、ハフマンおよび／またはＰＨ伸長技術に基づいてブロックを伸長する。ブロック４２４において、例示的なデータスライサ２０２は、符号化データを符号化サブブロックにスライス、分割および／またはグループ化する。サブブロックは、１つ以上のハフマン符号を含んでもよい。サブブロック当たりのハフマンコードの数は、プリセットされており、カスタマイズ可能であり、かつ／あるいはユーザおよび／または製造業者の好みに基づいてもよい。例えば、サブブロック当たりのハフマン符号の数は、例示的なデータ伸長器１０８における並列プロセッサの数に対応してもよく、それによって、データ伸長器１０８の処理能力を最適化する。

【００４３】

ブロック４２４において、例示的なパッダ２１０は、定義済みのパターンを使用して符号化サブブロックをパッドする。上述のように、異なるハフマン符号が、符号化処理プロセスの結果として不均一（例えば、異なる）長さに対応してもよい。例えば、第１のハフマン符号が２ビットのデータであってもよく、第２のハフマン符号が４ビットのデータであってもよい。従って、スライスされたサブブロックのサイズが、不均一な（例えば、異なる）長さに対応することがある。したがって、例示的なパッダ２１０は、サブブロックの終わりに対応して定義済みのパターンで、符号化サブブロックの各々をパッドする。いくつかの例では、パディングは、各可変サイズのサブブロックが同じ長さ（例えば、選択された長さ）であることを保証する。このようにして、例示的なデータ伸長器１０８は、選択されたサブブロック長に基づいてサブブロックを分割することによって、並列伸長のために圧縮サブブロックを分割することができる。いくつかの例では、パディングされたサブブロックは異なる長さである。このような例では、データ伸長器１０８は、ヘッダ内の指定されたパディングに基づいて、符号化ブロックをサブブロックに分割する。ブロック４２８において、例示的なパッダ２１０は、図２の例示的なオンチップ構成要素１０２を送信する前に、ヘッダを伴うパディングされたサブブロックを、アセンブルされるように例示的な圧縮データアセンブラ２１２に渡す。

【００４４】

図５は、図１の例示的なオフチップ構成要素１００から圧縮データを伸長する図１および／または図２のデータ伸長器１０８の例示的な実装によって実行され得る例示的な機械読み取り可能命令を表す例示的なフローチャート５００である。図５のフローチャート５００は、図１および／または図２の例示的なデータ伸長器１０８に関連して説明されているが、代わりに、他のタイプのデータ伸長器および／または他のタイプのプロセッサが利用されてもよい。フローチャート５００が、圧縮ブロックの連続伸長に関連して説明されているが、フローチャートは、各ブロックが並列に処理される並列伸長プロセスに関連して使用されてよい。

【００４５】

ブロック５０２において、例示的なヘッダプロセッサ２２０は、例示的なオフチップインターフェース２１８から圧縮データセットを取得する。圧縮データセットは、本明細書に開示された技術を使用して例示的なデータ圧縮器１０６によって圧縮された例示的なオフチップメモリ１０４に記憶されたデータに対応する。圧縮データブロックの各々について（ブロック５０４～５３０）（例えば、単一の伸長器２２２で順次処理される）、例示的なヘッダプロセッサ２２０は、データの符号化ブロックに対応する利用可能な圧縮ブロックのヘッダを特定する（ブロック５０６）。上述のように、データの圧縮中、各圧縮データブロックは、データブロックを圧縮するために使用される圧縮のタイプを特定するヘッダを含む。したがって、ブロック５０８において、例示的なヘッダプロセッサ２２０は、符号化ブロックのヘッダに基づいて圧縮および／または符号化タイプを決定する。図５の例示的なフローチャート５００は、ＲＬＥ圧縮、ハフマン圧縮、および／または圧縮なしを使用して圧縮されたデータに関連して説明されているが、ヘッダは、追加的および／または代替的な圧縮技術に対応してもよい。このような実施例では、データ伸長器１０８は、このような付加的および／または代替的な圧縮技術に基づいてデータを伸長する。

【００４６】

圧縮データブロックの各々について（ブロック５０８～５３０）（例えば、例示的なデータスプリッタ２２１によって分割され、単一の伸長器２２２によって順次処理される）の各々について、例示的な伸長器２２２は、圧縮タイプがＲＬＥ符号化および／または圧縮に対応するかどうか決定する（ブロック５１０）。例示的な伸長器２２２が、圧縮タイプはＲＬＥ符号化に対応すると決定した場合（ブロック５１０：ＹＥＳ）、例示的な伸長器２２２は、ヘッダ情報に基づいてデータを復号するためにＲＬＥ伸長を実行する（ブロック５１２）。例えば、ヘッダ情報がデータのブロックが１６回繰り返された「ａ」に対応することを特定する場合、伸長器２２２は、データを伸長して１６回繰り返された「ａ」を生成する。ブロック５１４において、例示的な伸長器２２２は、伸長データ（例えば

、上記の例における１６回繰り返された「ａ」）を例示的な伸長データアセンブラ２２４に渡す。

【００４７】

例示的な伸長器２２２が、圧縮タイプはＲＬＥ符号化に対応していないと決定した場合（ブロック５１０：ＮＯ）、例示的な伸長器２２２は、圧縮タイプが非符号化データに対応するかどうかを決定する（ブロック５１６）。例示的な伸長器２２２が圧縮タイプは非符号化データに対応すると決定した場合（ブロック５１６：ＹＥＳ）、例示的な伸長器２２２は、非符号化データを伸長データアセンブラ２２４に渡す。例示的な伸長器２２２が、圧縮タイプは非符号化データに対応していないと決定した場合（ブロック５１６：ＮＯ）、例示的な伸長器２２２は、ヘッダにおいて特定されるサイズのサブブロックに基づいて、ハフマンまたは一部ハフマン符号化データブロックの符号化サブブロックを決定する（ブロック５２０）。上述したように、例示的なデータ圧縮器１０６は、データブロックのサブブロックの終わりを定義済みのエンドパターンでパッドする。このようにして、例示的な伸長器２２２は、ヘッダにおいて特定されたサブブロックサイズに基づいてブロックをサブブロックに分割することができる。

【００４８】

各符号化サブブロックについて（ブロック５２０～５２８）（例えば、例示的なデータスプリッタ２２１によってサブブロックに分割され、単一の伸長器２２２を使用して線形に、あるいは複数の伸長器２２２を使用して並列に処理される）、例示的な伸長器２２２は、符号化サブブロックにブロック（例えば、一部または完全ハフマン）復号を実行する。例えば、伸長器２２２が、伸長データを表す文字にハフマン符号を対応させたハフマンテーブルを利用して、符号化としてマークが付けられた文字を伸長および／または復号し、非符号化としてマークが付けられた文字を変えないままにしてもよい。ブロック５２６において、例示的な伸長器２２２は、復号サブブロックを伸長データアセンブラ２２４に渡す。ブロック５２９において、例示的な伸長データアセンブラ２２４は、サブブロックを伸長ブロックにアセンブルする。

【００４９】

ブロック５３２において、例示的な伸長データアセンブラ２２４は、例示的なオフチップメモリ１０４に記憶されたデータセットに対応する順序で、復号ブロックを伸長データセットにアセンブルする。ブロック５３４において、例示的なハードウェアアクセラレータ／ニューラルネットワークインターフェース２２６は、図１のハードウェアアクセラレータ／ニューラルネットワーク１１０に伸長データセットを送信する。

【００５０】

図６は、図１および／または図２の例示的なデータ圧縮器１０６を使用して、図１の例示的なオフチップ構成要素１００からのデータの例示的なデータ伸長プロセス６００を示す。例示的な伸長プロセス６００は、例示的なデータセット６０１、例示的なデータブロック６０２、例示的な決定ブロック６０４、６１４、例示的なＲＬＥ圧縮６０６、例示的なヘッダ６０８、６１６、６２０、例示的な圧縮データセット６１０、例示的な符号化ブロック６１２、例示的なハフマン圧縮６１３、例示的な符号化データブロック６１６、例示的な非符号化データブロック６１８、例示的な符号化ブロック６２２、例示的なスライシング６２４、例示的な符号化サブブロック６２６、および例示的なパディング６２８を含む。

【００５１】

図６の例示的なデータセット６０１は、例示的なオフチップメモリ１００に記憶されるデータに対応する。例示的なデータスライサ２０２は、例示的なデータブロック６０２を含む複数のデータブロックにデータセット６０１をスライスする。スライスされると、例示的なデータプロセッサ２０４は、決定ブロック６０４においてデータブロック６０２のデータの特性に基づいて符号化技術を選択する。データプロセッサ２０４が、ブロックの全てのデータは同じであると決定した場合、データプロセッサ２０４は、ＲＬＥ圧縮６０６を選択して、データブロック６０２を圧縮する。データプロセッサ２０４が、ブロック

の全てのデータは同じというわけではないと判断した場合、データプロセッサ２０４は、一部ハフマン圧縮（または完全ハフマン圧縮）６１３を選択する。

【００５２】

ＲＬＥ圧縮６０６が選択される場合、例示的なヘッダ生成器２０８は、ＲＬＥ圧縮技術に対応する例示的なヘッダ６０８を生成し、データブロック内の全てのデータを表す値と、ヘッダにおいて特定された繰り返し数とを特定する。従って、例示的なヘッダ６０８は、圧縮データセット６１０内の符号化データブロック６１２として含まれる。一部ハフマン圧縮６１３が選択される場合、圧縮器２０６が一部ハフマン圧縮６１３を実行すると、データプロセッサ２０４は、符号化データおよびヘッダが閾値（例えば、非符号化データブロック６０２に対応するデータ量に対応する閾値）より大きいかどうかを決定する。データプロセッサ２０４が、符号化データおよび対応するヘッダは閾値より大きいと決定する場合、圧縮器２０６は符号化データブロックを破棄し、ヘッダ生成器２０８が符号化なしに対応する例示的なヘッダ６１６を生成し、圧縮データセット６１０の符号化ブロック６１２として非符号化データブロック６０２を伴うヘッダ６１６を含める。

【００５３】

データプロセッサ２０４が、符号化データおよび対応するヘッダが閾値よりも小さいと決定した場合、例示的なヘッダ生成器２０８は、一部ハフマン圧縮６１３を特定する例示的なヘッダ６２０を生成する。追加的に、例示的なデータスライサ２０２は、例示的なスライシングブロック６２４においてハフマン符号化データ６２２をスライスして、符号化サブブロック６２６を生成する。符号化データサブブロック６２６は、不均一サイズであってもよいため（例えば、各データサブブロック６２６は、異なるサイズに対応してもよい）、例示的なパッダ２１０は、例示的なパディングブロック６２８において選択されたパターンで符号化データサブブロック６２６をパッドする。さらに、ヘッダ生成器２０８は、各サブブロックのサイズをヘッダに追加する。このようにして、例示的なデータ伸長器１０８は、特定されたサイズに基づいて、符号化ブロック６１２を例示的な符号化データサブブロック６２６に分割することができる。パディングされると、ヘッダ６２０、パディングされたサブブロックおよび対応するヘッダは、符号化ブロック６１２として、例示的な圧縮データセット６１０に含まれる。

【００５４】

図７は、図２の例示的なデータ圧縮器１０６を実装する図３～４の命令を実行するように構成された例示的なプロセッサプラットフォーム７００のブロック図である。プロセッサプラットフォーム７００は、例えば、サーバ、パーソナルコンピュータ、ワークステーション、自己学習機械（例えばニューラルネットワーク）、モバイルデバイス（例えば、携帯電話、スマートフォン、ｉＰａｄ（登録商標）のようなタブレット）、または任意の他の種類のコンピューティングデバイスとすることができる。

【００５５】

図示された例のプロセッサプラットフォーム７００は、プロセッサ７１２を含む。図示された例のプロセッサ７１２は、ハードウェアである。例えば、プロセッサ７１２は、１つ以上の集積回路、論理回路、マイクロプロセッサ、ＧＰＵ、ＤＳＰ、または任意の所望のファミリーまたは製造業者のコントローラによって実装されることができる。ハードウェアプロセッサが半導体ベース（例えば、シリコンベース）デバイスであってもよい。この例では、プロセッサは、図２の例示的なメモリインターフェース２００、例示的なデータスライサ２０２、例示的なデータプロセッサ２０４、例示的な圧縮器２０６、例示的なヘッダ生成器２０８、例示的なパッダ２１０および／または例示的な圧縮データアセンブラ２１２を実装する。

【００５６】

図示された例のプロセッサ７１２は、ローカルメモリ７１３（例えば、キャッシュ）を含む。いくつかの例では、ローカルメモリ７１３は、例示的な結果ストレージ２１８を実装する。図示された例のプロセッサ７１２は、バス７１８を介して、揮発性メモリ７１４および不揮発性メモリ７１６を含むメインメモリと通信している。揮発性メモリ７１４が

、ＳＤＲＡＭ（同期ダイナミック・ランダム・アクセス・メモリ）、ＤＲＡＭ（ダイナミック・ランダム・アクセス・メモリ）、ＲＤＲＡＭ（登録商標）（ＲＡＭＢＵＳ（登録商標）ダイナミック・ランダム・アクセス・メモリ）および／または任意の他のタイプのランダムアクセスメモリデバイスによって実装されてもよい。不揮発性メモリ７１６が、フラッシュメモリおよび／または任意の他の所望のタイプのメモリデバイスによって実装されてもよい。メインメモリ７１４、７１６へのアクセスはメモリコントローラによって制御される。

【００５７】

図示された例のプロセッサプラットフォーム７００は、インターフェース回路７２０も含む。インターフェース回路７２０が、Ｅｔｈｅｒｎｅｔ（登録商標）インターフェース、ユニバーサルシリアルバス（ＵＳＢ）、Ｂｌｕｅｔｏｏｔｈ（登録商標）インターフェース、近距離無線通信（ＮＦＣ）インターフェースおよび／またはＰＣＩエクスプレスインターフェース等、任意のタイプのインターフェース規格によって実装されてもよい。

【００５８】

図示された例では、１つ以上の入力デバイス７２２がインターフェース回路７２０に接続されている。入力デバイス７２２は、ユーザがデータおよび／またはコマンドをプロセッサ７１２に入力することを可能にする。入力デバイスは、例えば、オーディオセンサ、マイクロホン、カメラ、（スチルまたはビデオ）キーボード、ボタン、マウス、タッチスクリーン、トラックパッド、トラックボール、アイソポイント、および／または音声認識システムによって実装されることができる。

【００５９】

１つ以上の出力デバイス７２４も、図示された例のインターフェース回路７２０に接続されている。出力デバイス７２４は、例えば、ディスプレイデバイス（例えば、発光ダイオード（ＬＥＤ）、有機発光ダイオード（ＯＬＥＤ）、液晶ディスプレイ（ＬＣＤ）、陰極線管ディスプレイ（ＣＲＴ）、ＩＰＳ（ｉｎ－ｐｌａｃｅ　ｓｗｉｔｃｈｉｎｇ）ディスプレイ、タッチスクリーン等）、触覚出力デバイス、プリンタおよび／またはスピーカによって実装されることができる。したがって、図示された例のインターフェース回路７２０は、典型的には、グラフィックスドライバカード、グラフィックスドライバチップおよび／またはグラフィックスドライバプロセッサを含む。

【００６０】

図示された例のインターフェース回路７２０は、送信機、受信機、トランシーバ、モデム、住宅用ゲートウェイ、無線アクセスポイント、ネットワークインターフェース、ネットワーク７２６を介して外部機械（例えば任意の種類の計算デバイス）とのデータの交換を容易にするネットワークインターフェース等の通信デバイスも含む。通信は、例えば、Ｅｔｈｅｒｎｅｔ接続、デジタル加入者回線（ＳＤＬ）接続、電話回線接続、同軸ケーブルシステム、衛星システム、ラインオブサイト無線システム、セルラー電話システム等を介することができる。

【００６１】

図示された例のプロセッサプラットフォーム７００は、ソフトウェアおよび／またはデータを記憶するための１つ以上のマスストレージデバイス７２８も含む。このようなマスストレージデバイス７２８の例には、フロッピーディスクドライブ、ハードドライブディスク、コンパクトディスクドライブ、ブルーレイディスクドライブ、ＲＡＩＤ（ｒｅｄｕｎｄａｎｔ　ａｒｒａｙ　ｏｆ　ｉｎｄｅｐｅｎｄｅｎｔ　ｄｉｓｋｓ）、およびデジタル多用途ディスク（ＤＶＤ）ドライブが含まれる。

【００６２】

図３～図４の機械実行可能命令７３２は、大容量ストレージ７２デバイス８内、揮発性メモリ７１４内、不揮発性メモリ７１６内、および／またはＣＤ、ＤＶＤ等のリムーバブル非一時的なコンピュータ読み取り可能記憶媒体上に記憶され得る。

【００６３】

図８は、図２の例示的なデータ圧縮器１０６を実施する図５の命令を実行するように構

成された例示的なプロセッサプラットフォーム８００のブロック図である。プロセッサプラットフォーム８００は、例えば、サーバ、パーソナルコンピュータ、ワークステーション、自己学習機械（例えば、ニューラルネットワーク）、モバイルデバイス（例えば、携帯電話、スマートフォン、ｉＰａｄのようなタブレット）、または任意の他のタイプのコンピューティングデバイスとすることができる。

【００６４】

図示された例のプロセッサプラットフォーム８００は、プロセッサ８１２を含む。図示された例のプロセッサ８１２はハードウェアである。例えば、プロセッサ８１２は、１つ以上の集積回路、論理回路、マイクロプロセッサ、ＧＰＵ、ＤＳＰ、または任意の所望のファミリーまたは製造業者のコントローラによって実装されることができる。ハードウェアプロセッサが半導体ベース（例えばシリコンベース）デバイスであってもよい。この例では、プロセッサは、図２の例示的なオフチップインターフェース２１８、例示的なヘッダプロセッサ２２０、例示的なデータスプリッタ２２１、例示的な伸長器２２２、例示的な伸長データアセンブラ２２４、および／または例示的なハードウェアアクセラレータ／ニューラルネットワークインターフェース２２６を実装する。

【００６５】

図示された例のプロセッサ８１２は、ローカルメモリ８１３（例えば、キャッシュ）を含む。いくつかの例では、ローカルメモリ８１３は、例示的な結果ストレージ２１８を実装する。図示された例のプロセッサ８１２は、バス８１８を介して、揮発性メモリ８１４および不揮発性メモリ８１６を含むメインメモリと通信している。揮発性メモリ８１４が、ＳＤＲＡＭ（同期ダイナミック・ランダム・アクセス・メモリ）、ＤＲＡＭ（ダイナミック・ランダム・アクセス・メモリ）、ＲＤＲＡＭ（ＲＡＭＢＵＳダイナミック・ランダム・アクセス・メモリ）および／または任意の他のタイプのランダムアクセスメモリデバイスによって実装されてもよい。不揮発性メモリ８１６が、フラッシュメモリおよび／または任意の他の所望のタイプのメモリデバイスによって実装されてもよい。メインメモリ８１４、８１６へのアクセスは、メモリコントローラによって制御される。

【００６６】

図示された例のプロセッサプラットフォーム８００は、インターフェース回路８２０も含む。インターフェース回路８２０が、Ｅｔｈｅｒｎｅｔインターフェース、ユニバーサルシリアルバス（ＵＳＢ）、Ｂｌｕｅｔｏｏｔｈインターフェース、近距離無線通信（ＮＦＣ）インターフェースおよび／またはＰＣＩエクスプレスインターフェース等、任意のタイプのインターフェース規格によって実装されてもよい。

【００６７】

図示された例では、１つ以上の入力デバイス８２２がインターフェース回路８２０に接続されている。入力デバイス８２２は、ユーザがデータおよび／またはコマンドをプロセッサ８１２に入力することを可能にする。入力デバイスは、例えば、オーディオセンサ、マイクロホン、カメラ（スチルまたはビデオ）、キーボード、ボタン、マウス、タッチスクリーン、トラックパッド、トラックボール、アイソポイント、および／または音声認識システムによって実装されることができる。

【００６８】

１つ以上の出力デバイス８２４も、図示された例のインターフェース回路８２０に接続されている。出力デバイス８２４は、例えば、ディスプレイデバイス（例えば、発光ダイオード（ＬＥＤ）、有機発光ダイオード（ＯＬＥＤ）、液晶ディスプレイ（ＬＣＤ）、陰極線管ディスプレイ（ＣＲＴ）、ＰＳ（ｉｎ－ｐｌａｃｅ　ｓｗｉｔｃｈｉｎｇ）ディスプレイ、タッチスクリーン等）、触覚出力デバイス、プリンタおよび／またはスピーカによって実装されることができる。したがって、図示された例のインターフェース回路８２０は、典型的には、グラフィックスドライバカード、グラフィックスドライバチップおよび／またはグラフィックスドライバプロセッサを含む。

【００６９】

図示された例のインターフェース回路７２０は、送信機、受信機、トランシーバ、モデ

ム、住宅用ゲートウェイ、無線アクセスポイント、ネットワークインターフェース、ネットワーク７２６を介して外部機械（例えば任意の種類の計算デバイス）とのデータの交換を容易にするネットワークインターフェース等の通信デバイスも含む。通信は、例えば、Ｅｔｈｅｒｎｅｔ接続、デジタル加入者回線（ＳＤＬ）接続、電話回線接続、同軸ケーブルシステム、衛星システム、ラインオブサイト無線システム、セルラー電話システム等を介することができる。

【００７０】

図示された例のプロセッサプラットフォーム８００は、ソフトウェアおよび／またはデータを記憶するための１つ以上の大容量ストレージデバイス８２８も含む。このようなマスストレージデバイス７２８の例には、フロッピーディスクドライブ、ハードドライブディスク、コンパクトディスクドライブ、ブルーレイディスクドライブ、ＲＡＩＤ（ｒｅｄｕｎｄａｎｔ ａｒｒａｙ ｏｆ ｉｎｄｅｐｅｎｄｅｎｔ ｄｉｓｋｓ）、およびデジタル多用途ディスク（ＤＶＤ）ドライブが含まれる。

【００７１】

図５の機械実行可能命令８３２は、大容量ストレージデバイス８２８内、揮発性メモリ８１４内、不揮発性メモリ８１６内、および／またはＣＤまたはＤＶＤのようなリムーバブル非一時的なコンピュータ読み取り可能記憶媒体上に記憶され得る。

【００７２】

データを圧縮する例示的な方法、装置、システム、および製造物品が本明細書に開示されている。さらなる実施例およびそれらの組み合わせは、以下を含む。

【００７３】

実施例１は、データを記憶するオフチップメモリと、データセットを複数のデータブロックに分割するデータスライサと、データプロセッサであって、複数のデータブロックのうちの第１のブロックに対する第１の圧縮技術を、第１のブロックの第１の特性に基づいて選択し、複数のデータブロックのうちの第２のブロックに対する第２の圧縮技術を、第２のブロックの第２の特性に基づいて選択する、データプロセッサと、第１の圧縮技術を使用して第１のブロックを圧縮して第１の圧縮データブロックを生成する第１の圧縮器と、第２の圧縮技術を使用して第２のブロックを圧縮して第２の圧縮データブロックを生成する第２の圧縮器と、第１の圧縮技術を特定する第１のヘッダおよび第２の圧縮技術を特定する第２のヘッダを生成するヘッダ生成器と、オフチップメモリに記憶される、第１のヘッダを伴う第１の圧縮データブロックおよび第２のヘッダを伴う第２の圧縮データブロックを送信するインターフェースと、を含む装置を含む。

【００７４】

実施例２は、実施例１の装置を含み、データプロセッサは、ブロックのデータが同じ値に対応する場合に、第１の圧縮技術を選択し、ブロックのデータが少なくとも１つの異なる値を含む場合に第２の圧縮技術を選択するものである。

【００７５】

実施例３は、実施例１の装置を含み、第１の圧縮技術は、ハフマン圧縮技術、ＲＬＥ圧縮技術または圧縮技術なしのうちの少なくとも１つである。

【００７６】

実施例４は、実施例３の装置を含み、データプロセッサは、ハフマン圧縮技術が圧縮技術なしよりも多くのデータをもたらす場合に、第１の圧縮技術を圧縮技術なしと選択するものである。

【００７７】

実施例５は、実施例１の装置を含み、圧縮器は、第１の圧縮データブロックを第１の圧縮データのサブブロックに分割するものである。

【００７８】

実施例６は、実施例５の装置を含み、第１の圧縮データのサブブロックは、サイズが不均一である。

【００７９】

実施例７は、実施例５の装置を含み、サブブロックにパターンを追加することによって、パターンでサブブロックをパッドする１つ以上のパッダをさらに含む。

【００８０】

実施例８は、実施例１の装置を含み、データブロックは、サイズが均一である。

【００８１】

実施例９は、実施例１の装置を含み、第１の圧縮器および第２の圧縮器は、第１の圧縮技術および第２の圧縮技術をそれぞれのブロックに対して並列に実行するものである。

【００８２】

実施例１０は、実施例１の装置を含み、圧縮器は、選択された第１の圧縮技術が圧縮技術なしよりも多くのデータをもたらす場合に、第１の圧縮データブロックを破棄し、非圧縮のデータブロックを保持するものである。

【００８３】

実施例１１は、実施例１０の装置を含み、第１のヘッダは、圧縮技術なしを特定する。

【００８４】

実施例１２は、実行されると、少なくとも機械に、データセットを複数のデータブロックに分割し、複数のデータブロックのうちの第１のブロックに対する第１の圧縮技術を、第１のブロックの第１の特性に基づいて選択し、複数のデータブロックのうちの第２のブロックに対する第２の圧縮技術を、第２のブロックの第２の特性に基づいて選択し、第１の圧縮技術を使用して第１のブロックを圧縮して第１の圧縮データブロックを生成し、第２の圧縮技術を使用して第２のブロックを圧縮して第２の圧縮データブロックを生成し、第１の圧縮技術を特定する第１のヘッダおよび第２の圧縮技術を特定する第２のヘッダを生成し、第１のヘッダを伴う第１の圧縮データブロックおよび第２のヘッダを伴う第２の圧縮データブロックをチップメモリに記憶する、ことをさせる命令を含む非一時的なコンピュータ読み取り可能記憶媒体を含む。

【００８５】

実施例１３は、実施例１２のコンピュータ読み取り可能な記憶媒体を含み、命令は、ブロックのデータが同じ値に対応する場合に、機械に第１の圧縮技術を選択させ、ブロックのデータが少なくとも１つの異なる値を含む場合に、第２の圧縮技術を選択させる。

【００８６】

実施例１４は、実施例１２のコンピュータ読み取り可能な記憶媒体を含み、第１の圧縮技術は、ハフマン圧縮技術、ＲＬＥ圧縮技術または圧縮技術なしのうちの少なくとも１つである。

【００８７】

実施例１５は、実施例１４のコンピュータ読み取り可能な記憶媒体を含み、命令は、機械に、ハフマン圧縮技術が圧縮技術なしよりも多くのデータをもたらす場合に、第１の圧縮技術を圧縮技術なしと選択させる。

【００８８】

実施例１６は、実施例１２のコンピュータ読み取り可能な記憶媒体を含み、命令は、機械に、第１の圧縮データブロックを第１の圧縮データのサブブロックに分割させる。

【００８９】

実施例１７は、実施例１６のコンピュータ読み取り可能な記憶媒体を含み、第１の圧縮データのサブブロックは、サイズが不均一である。

【００９０】

実施例１８は、実施例１６のコンピュータ読み取り可能な記憶媒体を含み、命令は、機械に、サブブロックにパターンを追加することによって、該パターンでサブブロックをパッドさせる。

【００９１】

実施例１９は、実施例１２のコンピュータ読み取り可能な記憶媒体を含み、データブロックは、サイズが均一である。

【００９２】

実施例２０は、実施例１２のコンピュータ読み取り可能な記憶媒体を含み、命令は、機械に、第１の圧縮技術および第２の圧縮技術をそれぞれのブロックに対して並列に実行させる。

【００９３】

実施例２１は、実施例１２のコンピュータ読み取り可能な記憶媒体を含み、命令は、機械に、選択された第１の圧縮技術が圧縮技術なしよりも多くのデータをもたらす場合に、第１の圧縮データブロックを破棄させ、非圧縮データブロックを保持させる。

【００９４】

実施例２２は、実施例２０のコンピュータ読み取り可能な記憶媒体を含み、第１のヘッダは、圧縮技術なしを特定する。

【００９５】

実施例２３は、データセットを複数のデータブロックに分割することと、複数のデータブロックのうちの第１のブロックに対する第１の圧縮技術を、第１のブロックの第１の特性に基づいて選択することと、複数のデータブロックのうちの第２のブロックに対する第２の圧縮技術を、第２のブロックの第２の特性に基づいて選択することと、第１の圧縮技術を使用して第１のブロックを圧縮して第１の圧縮データブロックを生成することと、第２の圧縮技術を使用して第２のブロックを圧縮して第２の圧縮データブロックを生成することと、第１の圧縮技術を特定する第１のヘッダおよび第２の圧縮技術を特定する第２のヘッダを生成することと、第１のヘッダを伴う第１の圧縮データブロックおよび第２のヘッダを伴う第２の圧縮データブロックをチップメモリに記憶することと、を含む方法を含む。

【００９６】

実施例２４は、実施例２３の方法を含み、第１の圧縮技術は、ブロックのデータが同じ値に対応する場合に選択され、第２の圧縮技術は、ブロックのデータが少なくとも１つの異なる値を含む場合に選択される。

【００９７】

実施例２５は、実施例２３の装置を含み、第１の圧縮技術は、ハフマン圧縮技術、ＲＬＥ圧縮技術または無圧縮技術なしのうちの少なくとも１つである。

【００９８】

実施例２６は、実施例２５の装置を含み、第１の圧縮技術は、ハフマン圧縮技術が圧縮技術なしよりも多くのデータをもたらす場合に、圧縮技術なしと選択される。

【００９９】

実施例２７は、実施例２３の装置を含み、第１の圧縮データブロックを、第１の圧縮データのサブブロックに分割することをさらに含む。

【０１００】

実施例２８は、実施例２７の装置を含み、第１の圧縮データのサブブロックは、サイズが不均一である。

【０１０１】

実施例２９は、実施例２７の装置を含み、サブブロックにパターンを追加することによって、該パターンでサブブロックをパッドすることをさらに含む。

【０１０２】

実施例３０は、実施例２３の装置を含み、データブロックは、サイズが均一である。

【０１０３】

実施例３１は、実施例２３の装置を含み、第１の圧縮技術および第２の圧縮技術が、それぞれのブロックに対して並列に実行される。

【０１０４】

実施例３２は、実施例２３の装置を含み、選択された第１の圧縮技術が、圧縮技術なしよりも多くのデータをもたらす場合に、第１の圧縮データブロックを廃棄し、非圧縮データブロックを保持することをさらに含む。

【０１０５】

実施例３３は、実施例３２の装置を含み、第１のヘッダは、圧縮技術なしを特定する。

【０１０６】

実施例３４は、実行されると、少なくとも機械に、実行のための第２の命令を配布、設定、アセンブル、インストール、インスタンス化、検索、伸長および復号することのうちの少なくとも１つを実行させる第１の命令と、実行されると、少なくとも機械に、データセットを複数のデータブロックに分割し、複数のデータブロックのうちの第１のブロックに対する第１の圧縮技術を、第１のブロックの第１の特性に基づいて選択し、複数のデータブロックのうちの第２のブロックに対する第２の圧縮技術を、第２のブロックの第２の特性に基づいて選択し、第１の圧縮技術を使用して第１のブロックを圧縮して第１の圧縮データブロックを生成し、第２の圧縮技術を使用して第２のブロックを圧縮して第２の圧縮データブロックを生成し、第１の圧縮技術を特定する第１のヘッダおよび第２の圧縮技術を特定する第２のヘッダを生成し、第１のヘッダを伴う第１の圧縮データブロックおよび第２のヘッダを伴う第２の圧縮データブロックをチップメモリに記憶する、ことをさせる第２の命令とを含む例示的なコンピュータ読み取り可能記憶媒体。

【０１０７】

実施例３５は、実施例３４のコンピュータ読み取り可能記憶媒体を含み、命令は、ブロックのデータが同じ値に対応する場合に、機械に第１の圧縮技術を選択させ、ブロックのデータが少なくとも１つの異なる値を含む場合に、第２の圧縮技術を選択させる。

【０１０８】

実施例３６は、実施例３４のコンピュータ読み取り可能記憶媒体を含み、第１の圧縮技術は、ハフマン圧縮技術、ＲＬＥ圧縮技術または圧縮技術なしのうちの少なくとも１つである。

【０１０９】

実施例３７は、実施例３６のコンピュータ読み取り可能記憶媒体を含み、命令は、機械に、ハフマン圧縮技術が圧縮技術なしよりも多くのデータをもたらす場合に、第１の圧縮技術を圧縮技術なしと選択させる。

【０１１０】

実施例３８は、実施例３４のコンピュータ読み取り可能記憶媒体を含み、命令は、機械に、第１の圧縮データブロックを第１の圧縮データのサブブロックに分割させる。

【０１１１】

実施例３９は、実施例３８のコンピュータ読み取り可能記憶媒体を含み、第１の圧縮データのサブブロックは、サイズが不均一である。

【０１１２】

実施例４０は、実施例３８のコンピュータ読み取り可能記憶媒体を含み、命令は、機械に、サブブロックにパターンを追加することによって、該パターンでサブブロックをパッドさせる。

【０１１３】

実施例４１は、実施例３４のコンピュータ読み取り可能記憶媒体を含み、データブロックは、サイズが均一である。

【０１１４】

実施例４２は、実施例３４のコンピュータ読み取り可能記憶媒体を含み、命令は、機械に、第１の圧縮技術および第２の圧縮技術をそれぞれのブロックに対して並列に実行させる。

【０１１５】

実施例４３は、実施例３４のコンピュータ読み取り可能記憶媒体を含み、命令は、機械に、選択された第１の圧縮技術が圧縮技術なしよりも多くのデータをもたらす場合に、第１の圧縮データブロックを破棄させ、非圧縮データブロックを保持させる。

【０１１６】

実施例４４は、実施例４２のコンピュータ読み取り可能記憶媒体を含み、第１のヘッダは、圧縮技術なしを特定する。

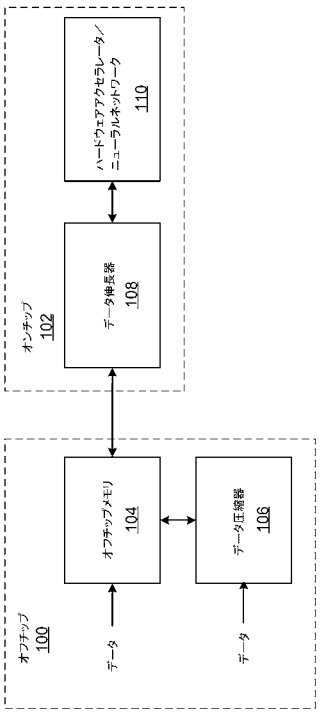
【０１１７】

以上から、オンチップニューラルネットワークおよび／またはハードウェアアクセラレータに送信される前に、オフチップメモリに記憶されたデータを圧縮する例示的な方法、装置および製造物品が開示されていることが理解されるであろう。開示された方法、装置および製造物品は、オンチップのニューラルネットワークおよび／またはハードウェアアクセラレータと共にオフチップメモリの利用を容易にし、メモリ容量を増加させることによって、計算デバイスの効率を改善する。本明細書に開示された例を用いて、ユーザは、より高い圧縮比を達成するために、特定のアプリケーションデータプロファイルに圧縮スキームを仕立てるができる。オフチップ構成要素からオンチップ構成要素にデータを転送する前にデータを圧縮することによって、本明細書に開示された実施例は、圧縮なしでデータを送信することのボトルネック効果を防ぐ、そうでなければ低減する。追加的に、データをデータのブロックおよび／またはサブブロックに分割することによって、本明細書に開示された実施例は、オンチップ構成要素を使用するデータの並列復号を容易にし、それによってオンチップ構成要素の処理帯域幅を増加させる。本明細書に開示された一部ハフマン符号化スキームの例を使用することにより、ハードウェアメモリの制限および速度の要件と互換性があるようにハフマン符号化ルックアップテーブルの要件を低減する。従って、開示された方法、装置、および製造物品は、コンピュータの機能における１つまたは複数の改良に向けられている。

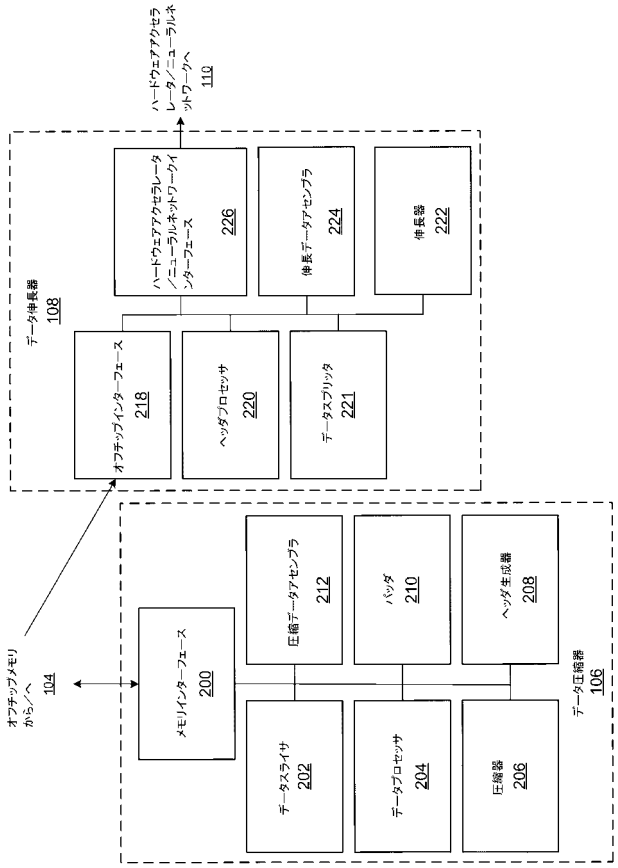
【０１１８】

一定の例示的な方法、装置および製造物品が本明細書に開示されているが、本特許がカバーする範囲はそれに限定されない。それどころか、この特許は、この特許の特許請求の範囲内にある全ての方法、装置および製造物品をカバーする。

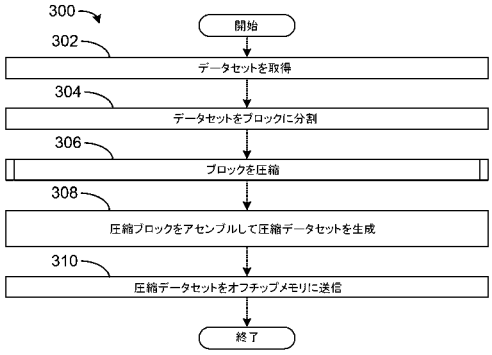
【図１】



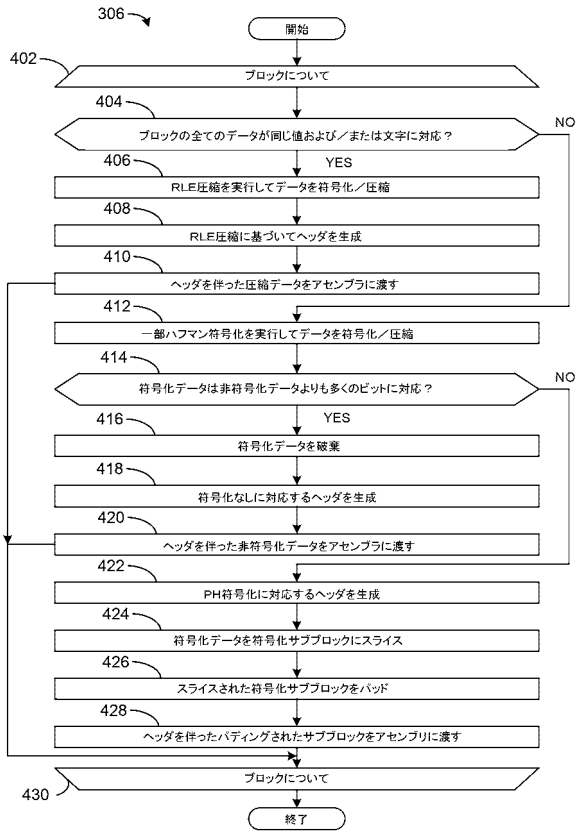
【図２】



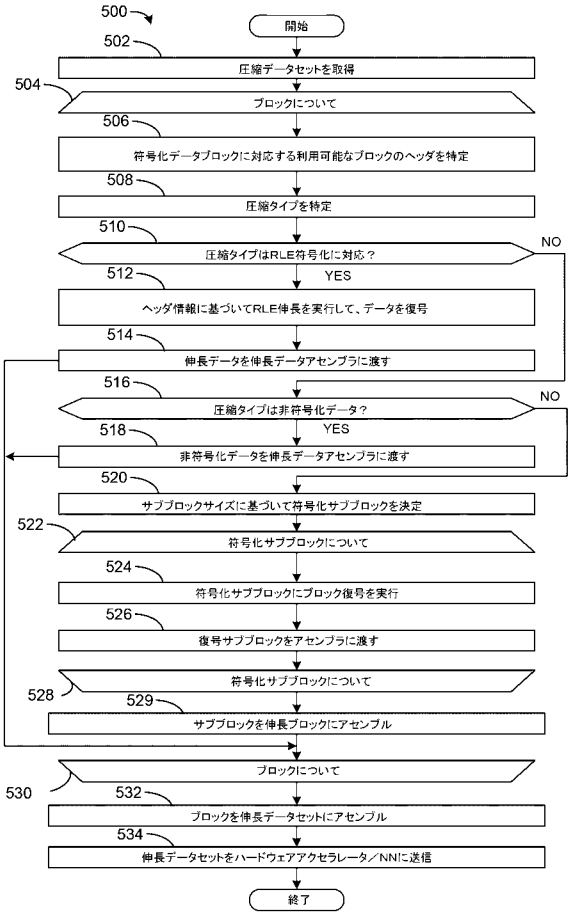
【図３】


300
開始
302
データセットを取得
304
データセットをブロックに分割
306
ブロックを圧縮
308
圧縮ブロックをアセンブルして圧縮データセットを生成
310
圧縮データセットをオフチップメモリに送信
終了

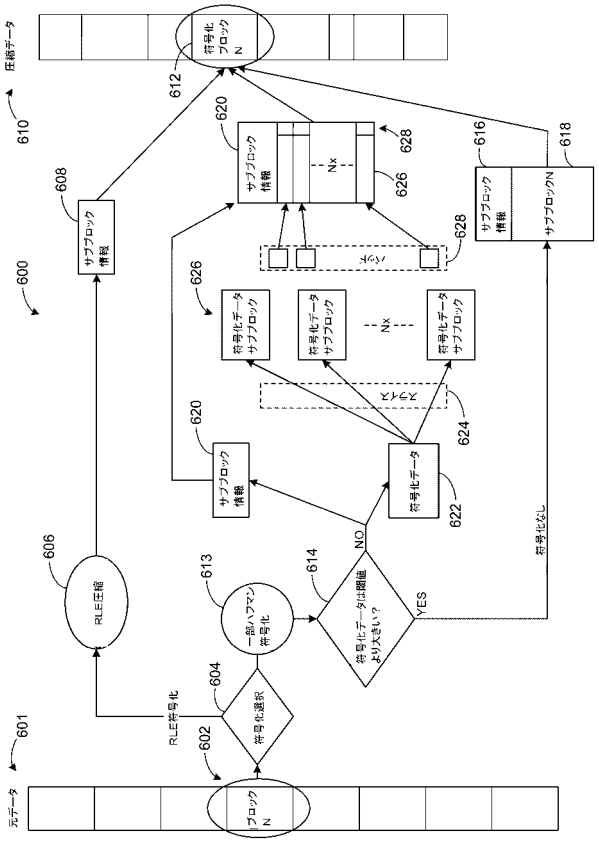
【図４】


306
開始
402
ブロックについて
404
ブロックの全てのデータが同じ値および／または文字に対応？
NO
YES
406
RLE圧縮を実行してデータを符号化／圧縮
408
RLE圧縮に基づいてヘッダを生成
410
ヘッダを伴った圧縮データをアセンブラに渡す
412
一部ハフマン符号化を実行してデータを符号化／圧縮
414
符号化データは非符号化データよりも多くのビットに対応？
NO
YES
416
符号化データを破棄
418
符号化なしに対応するヘッダを生成
420
ヘッダを伴った非符号化データをアセンブラに渡す
422
PH符号化に対応するヘッダを生成
424
符号化データを符号化サブブロックにスライス
426
スライスされた符号化サブブロックをパッド
428
ヘッダを伴ったパディングされたサブブロックをアセンブリに渡す
ブロックについて
430
終了

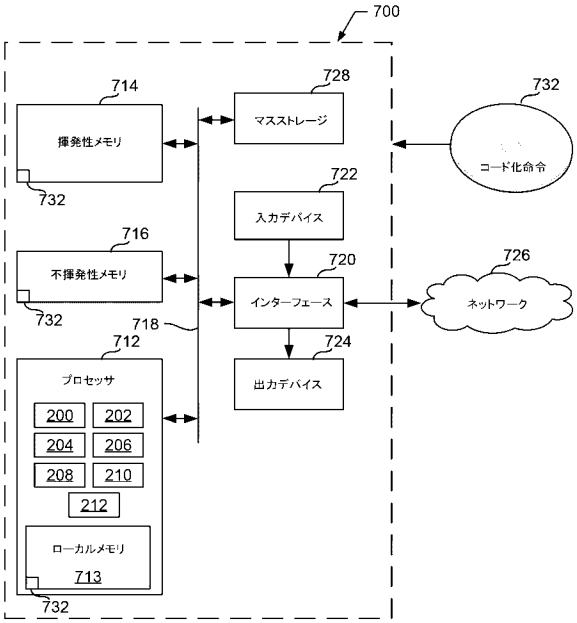
【図５】


500
開始
502
圧縮データセットを取得
504
ブロックについて
506
符号化データブロックに対応する利用可能なブロックのヘッダを特定
508
圧縮タイプを特定
510
圧縮タイプはRLE符号化に対応？
NO
512
YES
ヘッダ情報に基づいてRLE伸長を実行して、データを復号
514
伸長データを伸長データアセンブラに渡す
516
圧縮タイプは非符号化データ？
NO
518
YES
非符号化データを伸長データアセンブラに渡す
520
サブブロックサイズに基づいて符号化サブブロックを決定
522
符号化サブブロックについて
524
符号化サブブロックにブロック復号を実行
526
復号サブブロックをアセンブラに渡す
528
符号化サブブロックについて
529
サブブロックを伸長ブロックにアセンブル
ブロックについて
530
532
ブロックを伸長データセットにアセンブル
534
伸長データセットをハードウェアアクセラレータ／NNIに送信
終了

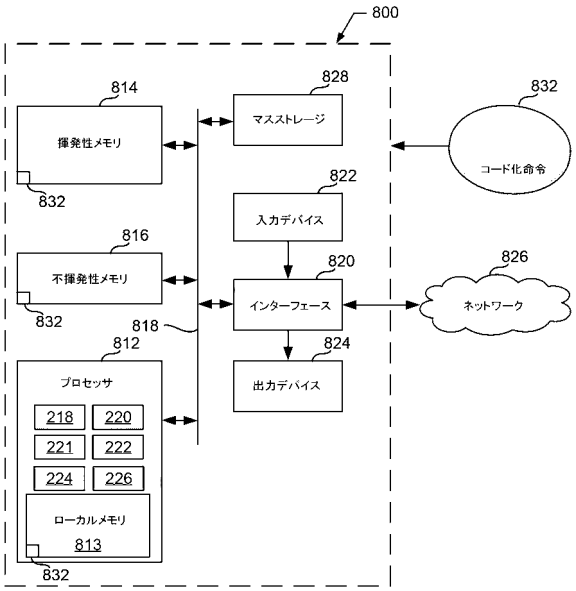
【図６】


圧縮データ
符号化ブロックN
612
610
サブブロック情報
620
628
626
616
618
サブブロックN
608
600
符号化データサブブロック
624
622
606
RLE圧縮
613
一部ハフマン符号化
614
符号化データは閾値よりは大きい？
NO
YES
符号化なし
604
符号化選択
RLE符号化
602
601
元データ
ブロックN

【図７】


700
714
揮発性メモリ
732
728
マスストレージ
732
コード化命令
722
入力デバイス
716
不揮発性メモリ
732
720
インターフェース
726
ネットワーク
718
712
プロセッサ
200
202
204
206
208
210
212
ローカルメモリ
713
732
724
出力デバイス

【図８】


800
814
揮発性メモリ
832
828
マスストレージ
832
コード化命令
822
入力デバイス
816
不揮発性メモリ
832
820
インターフェース
826
ネットワーク
818
812
プロセッサ
218
220
221
222
224
226
ローカルメモリ
813
832
824
出力デバイス

フロントページの続き

(72)発明者　デヴィッド　バーナード
アイルランド国　キルデア　キルカレン，ニコラスタウン，プリムローズ　コテージ
(72)発明者　ピーター　マクグリン
アイルランド国　ダブリン，　ディー１４　イー２７５，ラスファーナム，バートン　ドライヴ　８３

Ｆターム(参考) 5B033 DB00
5B160 DA08
5J064 AA01　BA08　BA09　BC02　BD02

【外国語明細書】
2020141402000001.pdf