

(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2014-236449
(P2014-236449A)

(43) 公開日 平成26年12月15日(2014.12.15)

(51) Int.Cl.
H03M 7/42 (2006.01)

F I
H03M 7/42

テーマコード(参考)
5J064

審査請求 未請求 請求項の数 9 O L (全 25 頁)

(21) 出願番号 特願2013-118356(P2013-118356)
(22) 出願日 平成25年6月4日(2013.6.4)

(出願人による申告)平成24年度、独立行政法人科学技術振興機構、戦略的創造研究推進事業(個人型研究(さきがけ))に関する委託研究、産業技術力強化法第19条の適用を受ける特許出願

(71) 出願人 504171134
国立大学法人 筑波大学
茨城県つくば市天王台一丁目1番1
(71) 出願人 504174135
国立大学法人九州工業大学
福岡県北九州市戸畑区仙水町1番1号
(74) 代理人 100100549
弁理士 川口 嘉之
(74) 代理人 100123319
弁理士 関根 武彦
(72) 発明者 山際 伸一
茨城県つくば市天王台一丁目1番1 国立
大学法人筑波大学内

最終頁に続く

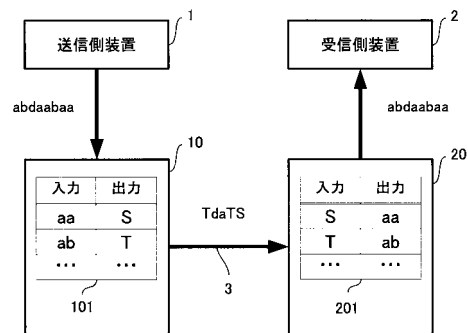
(54) 【発明の名称】 データ圧縮器及びデータ解凍器

(57) 【要約】

【課題】一定化された処理遅延での圧縮及び解凍処理を用いた送受信間における効率的なデータ伝送を行う。

【解決手段】本発明の一側面に係るデータ圧縮器は、複数の固定長のシンボルからなる入力データ列に含まれた、連続する2以上のシンボルが登録されている場合、2以上のシンボルを1つのシンボルに変換する変換部と、変換部で2以上のシンボルが1つのシンボルに変換された場合は、当該1つのシンボルを出力し、そうでない場合は、2以上のシンボルを出力する出力部とを含む。

【選択図】図6



【特許請求の範囲】

【請求項 1】

複数の固定長のシンボルからなる入力データ列に含まれた、連続する 2 以上のシンボルが登録されている場合、前記 2 以上のシンボルを 1 つのシンボルに変換する変換部と、前記変換部で前記 2 以上のシンボルが 1 つのシンボルに変換された場合は、当該 1 つのシンボルを出力し、そうでない場合は、前記 2 以上のシンボルを出力する出力部と、を含むデータ圧縮器。

【請求項 2】

前記出力部の、前記 2 以上のシンボルが 1 つのシンボルに変換された場合の出力は、そうでない場合の出力が取り得ない値となる

10

請求項 1 に記載のデータ圧縮器。

【請求項 3】

前記変換部は、前記 2 以上のシンボルが 1 つのシンボルに変換された場合、当該 1 つのシンボルに変換したことを示すフラグを出力し、そうでない場合、前記 2 以上のシンボルの各々に変換していないことを示すフラグを出力する

請求項 1 又は 2 に記載のデータ圧縮器。

【請求項 4】

前記変換部は、CAM (Content Addressable Memory) である

請求項 1 から 3 のいずれか一項に記載のデータ圧縮器。

【請求項 5】

所定のルールに基づいて、前記入力データ列に含まれる前記シンボル間の区切位置を決定する区切位置決定部

20

をさらに含み、

前記変換部は、前記区切位置を基準として、前記連続する 2 以上のシンボルが登録されているか判断する

請求項 1 から 4 のいずれか一項に記載のデータ圧縮器。

【請求項 6】

前記区切位置決定部は、前記シンボルに対して予め定められた優先度を用いて、連続する 2 つの前記シンボルが示す値の優先度に基づいて、前記区切位置を決定する

請求項 5 に記載のデータ圧縮器。

30

【請求項 7】

請求項 1 から 6 のいずれか一項に記載のデータ圧縮器を複数直列に接続し、

前段のデータ圧縮器の出力を後段のデータ圧縮器に入力する

データ圧縮装置。

【請求項 8】

入力データ列に含まれる固定長のシンボルが、圧縮処理において 2 以上のシンボルから変換されたシンボルとして登録されている場合、当該シンボルを前記 2 以上のシンボルに変換する変換部と、

前記変換部で前記シンボルが前記 2 以上のシンボルに変換された場合は、前記 2 以上のシンボルを出力し、そうでない場合は、前記入力データ列に含まれるシンボルを出力する出力部と、

40

を含むデータ解凍器。

【請求項 9】

前記圧縮処理において 2 以上のシンボルから変換されたシンボルは、アドレスとして登録されており、前記変換部は、前記アドレスに対応付けて保持する 2 以上のシンボルを出力するメモリである

請求項 8 に記載のデータ解凍器。

【発明の詳細な説明】

【技術分野】

【0001】

50

本発明は、データ圧縮器及びデータ解凍器に関する。

【背景技術】

【0002】

近年における、データストリームがネットワーク上を流れる環境下では、データストリームを形成するストリームデータに対するリアルタイム処理を行うため、ストリームデータを送受信する様々なエンティティ間におけるデータ伝送時間の短縮化が求められている。エンティティは、例えば、ネットワークに接続された様々な通信機器（端末装置、中継装置）である。また、データストリームは、通信機器内に搭載されたプロセッサ、LSI（Large Scale Integrated Circuit）、FPGA（Field Programmable Gate Array）のような様々なストリームデータに対する処理を行う電子回路チップ間を流れる。電子回路チップもエンティティの1つであり、エンティティ間の通信は、通信機器間の通信だけでなく、通信機器内部の電子回路チップ間の通信（いわゆる内部通信）を含む。

10

【0003】

近年では、ストリームデータ量が増大する傾向にある。或る量のストリームデータを送信側から受信側へ効率的に伝送する手法として、エンティティ間を結ぶ伝送路の周波数を上げる（伝送帯域を広げる）ことや、エンティティ間を複数の伝送路で結び、ストリームデータを並列に送信することが考えられる。しかしながら、これらの手法は、物理的、周波数的な限界がいずれ来ると考えられている。

【0004】

そこで、送信側エンティティでストリームデータの圧縮を行うことで、送信データ量の減少に伴うデータ伝送時間の短縮化を以て、データ伝送の効率化を図ることが考えられている。例えば、通信機器において、送信データを圧縮装置（LZWやRLEなど複数の圧縮アルゴリズムを含む）に接続し、元の送信データと、複数の圧縮アルゴリズムのそれぞれによる処理を経たデータとのそれぞれのデータサイズを比較し、最も小さいサイズのデータを通信機器から送出する技術がある（例えば、特許文献1）。

20

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2007-65828号公報

【特許文献2】特許第3748003号公報

30

【非特許文献】

【0006】

【非特許文献1】J. Ziv and A. Lempel, "A universal algorithm for sequential data compression." IEEE Transactions on Information Theory, Vol. IT-23, No.3, May 1977, p.337-343

【非特許文献2】T. Welch, "A technique for high-performance data compression." IEEE Computer. 1984, p. 8-19

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、例えば、上記した特許文献1における技術（従来技術）では、ソフトウェア処理による既存の可逆圧縮アルゴリズムの適用を前提としている。このため、以下のような問題があった。

40

【0008】

既存の圧縮アルゴリズムでは、例えば、“abcd・・・”というデータシンボル（以下、単に「シンボル」と表記）列に対して圧縮処理が実行される場合、最初に、シンボル“a”についてルックアップテーブルが検索される。“a”がヒットした場合には、次に、“ab”についてルックアップテーブルが検索される。“ab”がヒットしなかった場合に、データ列“a”が所定の圧縮データ（例えば“x”）に変換される一方で、“ab”に対する変換用のエントリ（例えば“ab” “y”）が新たにルックアップテーブ

50

ルに登録される。

【0009】

上記手法では、検索対象がヒットしなくなるまでルックアップテーブルの検索が繰り返される。従って、ルックアップテーブルの検索回数に応じて、1回の圧縮処理に係る処理ステップが可変となる。また、ストリームデータにおいて、データ列の傾向（或る変換可能なシンボル列の出現頻度）によって、処理ステップ数が左右される。このため、1回の圧縮処理に要する時間が一定とならない。これが第1の問題である。

【0010】

また、上記手法では、ルックアップテーブルの更新によって、圧縮可能なシンボル列が増加する一方で、解凍側に解凍処理用のテーブルエントリ（“y” “ab”）を送らなければ、解凍側で解凍を行うことができない。このとき、圧縮されたデータと解凍処理用データの合計サイズが元のデータのサイズと同等、或いは元のデータのサイズを上回る場合には、伝送路上のデータ量減少に貢献しない。これが、第2の問題である。

【0011】

さらに、上記手法では、ルックアップテーブルにおいて、圧縮されるシンボル列をなすシンボル数は、一定ではなく、どのような長さ（個数）のシンボル列でも登録可能となっている。これが第3の問題である。

【0012】

圧縮処理は、送信側エンティティがストリームデータを受信側エンティティで送信する前に実行される複数の工程（ステップ）の一つとして挿入される。このとき、圧縮処理がボトルネックとなることは回避されるべきである。また、圧縮処理において揺らぎ（ジッタ）を生じさせることも、後段に位置する工程への影響に鑑み回避されるべきである。これらの観点からは、処理ステップ数（処理遅延）が一定とならない第1の問題及び第3の問題を含む従来技術は採用し得ず、また、第2の問題に鑑みても従来技術は採用し得ない。

【0013】

さらに、圧縮処理は、処理速度の観点から、ソフトウェア処理ではなくハードウェア処理にて行われることが好ましいと考えられる。このとき、従来技術で行われる圧縮処理（アルゴリズム）を行うデジタル回路を形成しようとする、処理が複雑なために、回路内にタイミング調整のための様々な遅延を挿入が求められ、結果として回路が複雑化し、回路規模が増大する虞があった。

【0014】

圧縮データを解凍するには、解凍処理として圧縮処理と逆の処理が受信側で実行される。このため、圧縮処理に係る上記の問題点は、そのまま解凍処理に係る問題として提起される。

【0015】

本発明は、上記の事情に鑑みなされたものであり、一定化された処理遅延での圧縮及び解凍処理を用いた送受信間における効率的なデータ伝送を行うことを可能とする技術を提供することを目的とする。

【課題を解決するための手段】

【0016】

本発明の一側面に係るデータ圧縮器は、複数の固定長のシンボルからなる入力データ列に含まれた、連続する2以上のシンボルが登録されている場合、2以上のシンボルを1つのシンボルに変換する変換部と、変換部で2以上のシンボルが1つのシンボルに変換された場合は、当該1つのシンボルを出力し、そうでない場合は、2以上のシンボルを出力する出力部とを含む。

【0017】

また、本発明の他の側面に係るデータ解凍器は、入力データ列に含まれる固定長のシンボルが、圧縮処理において2以上のシンボルから変換されたシンボルとして登録されている場合、当該シンボルを2以上のシンボルに変換する変換部と、変換部でシンボルが2以

10

20

30

40

50

上のシンボルに変換された場合は、2以上のシンボルを出力し、そうでない場合は、入力データ列に含まれるシンボルを出力する出力部とを含む。

【発明の効果】

【0018】

本発明によれば、一定化された処理遅延での圧縮及び解凍処理を用いた送受信間における効率的なデータ伝送を行うことを可能とする技術を提供することができる。

【図面の簡単な説明】

【0019】

【図1】圧縮・解凍器の構成を示す図である。

【図2】データ圧縮器の一例を示す機能ブロック図である。

10

【図3】ルックアップテーブルに登録されるエントリの一例を示す図である。

【図4】データ圧縮器の一例を示す回路構成図である。

【図5】データ解凍器の一例を示す回路構成図である。

【図6】圧縮処理及び解凍処理の具体例を説明するための図である。

【図7】実施形態2に係るデータ圧縮器の一例を示す機能ブロック図である。

【図8】実施形態2に係るデータ圧縮器の一例を示す回路構成図である。

【図9】実施形態2に係るデータ解凍器の一例を示す回路構成図である。

【図10】実施形態3に係るデータ圧縮装置の処理を説明するための図である。

【図11】実施形態3に係るデータ圧縮器の一例を示す機能ブロック図である。

20

【図12】実施形態3に係るデータ解凍装置の処理を説明するための図である。

【図13】実施形態3に係るデータ解凍器の一例を示す回路構成図である。

【図14】実施形態4に係るデータ圧縮器の一例を示す機能ブロック図である。

【図15】実施形態4に係る区切位置決定部の一例を示す回路構成図である。

【図16】エンコーダの入出力を説明するための図である。

【図17】区切位置を決定する処理を説明する図である。

【図18】実施形態4に係る圧縮部の一例を示す回路構成図である。

【図19】データ圧縮器及びデータ解凍器の間で送受信されるデータのフォーマットの一例を説明するための図である。

【発明を実施するための形態】

【0020】

30

以下、本発明の実施の形態について、図面に基づいて説明する。なお、実施の形態は本発明の一例であり、本発明の構成は以下の例には限られない。

【0021】

〔実施形態1〕

図1は、データ圧縮/解凍システムを模式的に示す図である。図1において、データ圧縮/解凍システムは、データ圧縮器10と、データ解凍器20とを含む。

【0022】

データ圧縮器10は、図示しない送信側エンティティが図示しない受信側エンティティへ送信すべきデータ(例えば、ストリームデータ)に対する圧縮処理を行い、圧縮データを出力する。圧縮データは、伝送路3を通過してデータ解凍器20に到達する。データ解凍器20は、解凍処理によって、圧縮データを元のデータに戻す。そして、元のデータが受信側エンティティに供給される。圧縮処理によって、伝送路3へ送られるデータ量が減少することで、或るサイズを有するデータが送信側エンティティから受信側エンティティに受信される迄の時間を、圧縮処理が行われない場合に比べて短縮することができる。

40

【0023】

データ圧縮/解凍システムは、通信機器間の通信に適用されても良く、通信機器内の構成要素(電子回路チップ)間通信(いわゆる内部通信)に適用されても良い。通信機器間の通信のために、データ圧縮器10は、送信側の通信機器に搭載され、データ解凍器20は、受信側の通信機器に搭載されることができる。内部通信のために、データ圧縮器10及びデータ解凍器20は、通信機器や各種の情報処理装置(コンピュータ)内に構成要素

50

の一つとして搭載されることができる。

【0024】

なお、通信機器が双方向通信を行う場合には、送信側及び受信側の通信機器のそれぞれにデータ圧縮器10及びデータ解凍器20が搭載され、上り通信と下り通信とのそれぞれにおいてデータの圧縮/解凍が行われる構成を適用可能である。

【0025】

<データ圧縮器>

図2は、図1に示したデータ圧縮器10の一例を示すブロック図である。データ圧縮器10は、例えば、送信側エンティティ(送信側装置)から入力されるストリームデータに対する可逆圧縮処理(符号化処理とも呼ばれる)を行い、可逆圧縮処理が施されたストリームデータを出力する。データ圧縮器10は、可逆圧縮の手法として、いわゆる文法圧縮を用いた静的符号化を行う。具体的には、データ圧縮器10は、ストリームデータに含まれる所定のシンボル列に対する変換ルールを規定したルックアップテーブルを有し、ルックアップテーブルに従って、所定のデータシンボル列を、当該シンボル列よりサイズの小さい1つのシンボル(圧縮シンボルとも呼ぶ)に圧縮(符号化)する。

10

【0026】

<<ルックアップテーブル>>

図3は、データ圧縮器10が備えるルックアップテーブルの一例を示す。ルックアップテーブルは、「入力(圧縮前)」の2つのシンボル(シンボルペアと称する)と圧縮(符号化)処理によって変換される「出力(圧縮後)」の1つのシンボルとの対応関係が登録された複数のエントリを含む。エントリ数は、ルックアップテーブルを記憶するメモリ101の記憶容量に応じた有限値である。

20

【0027】

図3において、1つの英文字は、1つのデータシンボルを意味する。1シンボルのサイズは固定サイズ(固定長)であり、例えば、1シンボルのサイズは1バイト(8ビット)である。ただし、シンボルのサイズは1バイトには限定されず、所定のサイズを採用することができる。

【0028】

また、図3に示すルックアップテーブルでは、圧縮後のシンボルを便宜的に英字で表しているが、実施形態1では、圧縮後のシンボルの値として、圧縮前のシンボルには存在し得ない値が登録される。換言すれば、圧縮後のシンボルは、圧縮前のシンボルが取り得る値の集合に属さない値である。これによって、データ解凍器20において、圧縮されたシンボルと圧縮されていないシンボルとを区別した解凍処理を行うことができる。

30

【0029】

例えば、ストリームデータが所定の文字コードが直列に並べられたテキストデータである場合には、圧縮前のシンボルとして2文字のペアが登録され、圧縮後のシンボルとしては当該文字コードにおいて未使用のビット列が登録される。また、例えば、ストリームデータが遺伝子データ(「A」、「G」、「T」及び「C」の核酸コード配列)である場合には、ルックアップテーブルの圧縮前の2シンボルとして、「A」、「G」、「T」及び「C」から2文字を取り出した順列が登録され、圧縮後の1シンボルとして、「A」、「G」、「T」及び「C」以外の値が登録される。

40

【0030】

実施形態に係るデータ圧縮器10及びデータ解凍器20には、静的に作成された同内容のエントリ群(シンボルペアリスト)を記憶したルックアップテーブルが備えられる。「静的に」とは、ルックアップテーブルの登録内容が、検索結果に応じて動的に変更(更新)されないことを意味する。但し、ルックアップテーブルの登録内容は、例えば随時の保守作業や、周期的な更新作業によって変更され得る。また、シンボルペアは、ストリームデータにおける出現率が高い順(いわゆる尤度が高い順)で登録される。

【0031】

<<データ圧縮器の構成>>

50

図2に戻ってデータ圧縮器10は、上記したルックアップテーブル(以下、単に「テーブル」と表記することもある)の構成を前提として、テーブルでの検索対象となるシンボルペアを特定するための構成と、シンボルペアがテーブルに登録されているときに、シンボルペアを対応する1シンボル(「圧縮シンボル」と称する)に変換するための構成と、テーブルにおけるシンボルペアがヒットしたか否かに応じて元のシンボルペアと圧縮シンボルとの一方を出力する構成とを備える。

【0032】

具体的には、データ圧縮器10は、ストリームデータのような入力データを保持するラッチ110と、ルックアップテーブルが記憶された変換部に相当するメモリ101(読み書き回路含む)と、入力順にシンボルを出力するシリアライザ(マルチプレクサ: MUX)102と、圧縮シンボル又は元のシンボルペアを出力するための出力部に相当するセレクタ(マルチプレクサ)103と、出力データを保持するラッチ120とを含み、これらが信号線で接続されている。

10

【0033】

ラッチ110は、入力されたストリームデータを一時的に保持するバッファ機能を有する。ストリームデータは有限長のビット列であり、データ圧縮器10において、固定サイズ(1バイト)のシンボル単位で扱われる。ラッチ110は、ストリームデータの先頭から順に、所定タイミングで、シンボルペアを形成する、隣接した2つのシンボル(シンボル1, シンボル2)を並列に出力する。シンボルペアは、メモリ101とシリアライザ102とに入力される。

20

【0034】

メモリ101では、入力されたシンボルペアがテーブルから検索される。検索は、テーブル(図3)の各エントリに登録された“入力(圧縮前)”のシンボルペアと入力されたシンボルペアとのマッチングにより実行される。入力されたシンボルペアと一致する(マッチする)エントリが見つかった(ヒットした)場合には、当該エントリに登録されている“出力(圧縮後)”のシンボル(圧縮シンボル)と、検索がヒットしたことを示すマッチ信号が出力される。圧縮シンボルは、セレクタ12へ入力され、マッチ信号はセレクタ103の制御信号としてセレクタ103に入力される。

【0035】

メモリ101として、例えば連想メモリ(CAM(Content Addressable Memory))を適用することができる。CAMは、入力されたデータワード(データ語)に対応するアドレスを出力することのできる高速検索用のコンピュータメモリである。CAMが適用される場合には、データワードとしてのシンボルペアがCAMに入力され、CAMはエントリがヒットしたときに、データワードに対応するアドレスとして圧縮シンボルを出力するとともに、真(True)信号(“1”信号)を出力する。True信号は、上記したマッチ信号として使用される。一方、エントリがヒットしなかった場合には、CAMからのアドレス(圧縮シンボル)の出力は行われず、偽(False)信号(“0”信号)の出力が行われる。

30

【0036】

シリアライザ(マルチプレクサ)102は、入力されたシンボルペアを、所定順序(シンボル1, シンボル2)でセレクタ103に出力することができる。セレクタ103は、マッチ信号が入力されたとき(True入力)に、元のシンボルペアと圧縮シンボルとのうち、圧縮シンボルを出力する。これに対し、マッチ信号が入力されないとき(False入力)には、シリアライザ102から出力された元のシンボルペアを出力する。セレクタ103の出力は、ラッチ120にて一時的に保持される。このような処理が、ストリームデータの先頭から順に繰り返されることで、ラッチ120には、元のストリームデータに対して圧縮処理が施された圧縮データが蓄積される。圧縮データは適宜のタイミングでラッチ120から出力され、伝送路3を介してデータ解凍器20へ送られる。

40

【0037】

なお、メモリ101は、CAMとRAM(Random Access Memory)との組合せであってもよい。この場合、CAMが出力するRAMのアドレスに圧縮シンボルが記憶され、当該

50

アドレスの圧縮シンボルがセレクタ 1 2 に供給される。R A M は D R A M でも S R A M でも良いが、C A M との連動に鑑み、高速動作が可能なものを選択するのが好ましい。

【 0 0 3 8 】

図 2 に示したデータ圧縮器 1 0 において、メモリ 1 0 1 における動作は、「テーブル検索 ヒット判定 圧縮シンボル出力 True/False 信号出力」のシーケンシャルな処理である。そして、テーブル検索は、シンボルペアと各エントリに格納されたシンボルペアとの 1 回のマッチングであり、従来技術のような、検索対象（入力シンボル列）を変更しての再度の検索処理は行われない。換言すれば、1 回目の検索でエントリがヒットしたときに、ラッチ 1 1 0 から次のシンボルを得て再度の検索を行うことがない。従って、テーブル検索の所要時間（処理遅延）は、最長でも、シンボルペアと全エントリのシンボルペアとのマッチングを行う時間内に収まる。

10

【 0 0 3 9 】

ヒット判定、圧縮シンボル出力、及び True/False 信号出力の処理遅延はほぼ固定である。従ってメモリ 1 0 1（C A M）における処理遅延が固定的になる。従って、メモリ 1 0 1 からの圧縮シンボル及び True/False 信号の出力タイミングを固定化することができる。また、シリアライザ 1 0 2 及びセレクタ 1 0 3 での処理開始タイミングは、メモリ 1 0 1 からの圧縮シンボル及び True/False 信号の出力タイミングに依存する。

【 0 0 4 0 】

圧縮シンボル及び True/False 信号の出力タイミングが固定されることで、シリアライザ 1 0 2 及びセレクタ 1 0 3 の処理開始タイミングも固定化できる。そして、シリアライザ 1 0 2 及びセレクタ 1 0 3 の処理遅延はほぼ一定である。従って、或るシンボルペアがラッチ 1 1 0 から出力されてから、圧縮シンボル又は元のシンボルペアがラッチ 1 2 0 に入力されるまでの所要時間（処理遅延）を固定化することができる。

20

【 0 0 4 1 】

これによって、図 2 に示したデータ圧縮器 1 0 をハードウェア（デジタル回路）で実現する（実装する）場合におけるタイミング調整が容易となり、デジタル回路の簡易化を図ることができる。

【 0 0 4 2 】

図 4 は、図 2 に示したデータ圧縮器 1 0 のデジタル回路で構成する場合の一例を示す。データ圧縮器 1 0 は、ルックアップテーブルを記憶する C A M 1 0 1 と、ストリームデータをシンボル単位で保持し、クロック信号に従って伝送する D - F F（D 型フリップフロップ）1 1 1 ~ D - F F 1 1 3 と、セレクタ 1 0 3 と、セレクタ 1 0 3 に出力されたシンボルを保持する D - F F 1 2 1 とを備えている。これらの各構成要素には、共通のクロック信号が供給され、クロック信号に従ったタイミングで同期した動作を行う。

30

【 0 0 4 3 】

ストリームデータは、シンボル単位で、D - F F 1 1 1 に入力される。次のシンボルが D - F F 1 1 1 に入力される毎に、先に入力されたシンボルは、次段の D - F F 1 1 2 や D - F F 1 1 3 へ進んでいく。

【 0 0 4 4 】

一方、C A M 1 0 1 には、イネーブル（有効化）信号又はディスエイブル（無効化）信号が交互に規則的に入力される。例えば、D - F F 1 1 2 から 1 番目のシンボルが出力され、D - F F 1 1 1 から 2 番目のシンボルが出力されるタイミングで、C A M 1 0 1 へイネーブル信号が入力される。イネーブル信号に基づき、C A M 1 0 1 は、1 番目と 2 番目のシンボルペアを用いてルックアップテーブルを検索する（すなわち、圧縮を試みる）。

40

【 0 0 4 5 】

次に、D - F F 1 1 2 から 2 番目のシンボルが出力され、D - F F 1 1 1 から 3 番目のシンボルが出力されるタイミングでは、C A M 1 0 1 へディスエイブル信号が入力される。C A M 1 0 1 は 2 番目と 3 番目のシンボルペアについては圧縮を試みない。さらに、D - F F 1 1 2 から 3 番目のシンボルが出力され、D - F F 1 1 1 から 4 番目のシンボルが出力されるタイミングでは、C A M 1 0 1 へイネーブル信号が入力され、3 番目と 4 番目

50

のシンボルペアを用いて圧縮を試みる。実施形態 1 では、ストリームデータの先頭から順にシンボルを 2 つずつペアにして、圧縮を試みる。

【 0 0 4 6 】

セレクタ 1 0 3 では、1 番目のシンボルが D - F F 1 1 3 に保持されている間に、圧縮シンボルの入力（ヒットした場合）と、マッチ信号の入力とが行われる。マッチ信号の入力時には、D - F F 1 2 1 に対する圧縮シンボルがセレクタ 1 0 3 から D - F F 1 2 1 に入力される。一方、マッチ信号が入力されない場合には、セレクタ 1 0 3 は、1 番目のシンボルと 2 番目のシンボルとを所定タイミングで D - F F 1 2 1 に入力する。

【 0 0 4 7 】

その後、3 番目と 4 番目、5 番目と 6 番目のようなシンボルペアに関して、上記した 1 番目及び 2 番目のシンボルに対して説明した動作が繰り返し行われる。このように、直列接続された複数の D - F F 上を、各シンボルが進んでいく間に、C A M による圧縮処理（圧縮シンボルへの変換）が行われる簡易な回路構成を採用することができる。

10

【 0 0 4 8 】

なお、図 2 に図示した、隣接するシンボル（シンボルペア）を伝送する並列な経路及びシリアライザ 1 0 2 を例示したが、図 4 に示すような、シンボルペアを直列に伝送する D - F F の多段回路を採用することで、並列な経路及びシリアライザ 1 0 2 を備えた場合と同様の（等価な）構成を採ることができる。

【 0 0 4 9 】

以上のようなデータ圧縮器 1 0 によれば、シンボルペアを 1 つの圧縮シンボルに置き換えた分だけ、伝送路 3 を流れるデータ量を削減することができる。また、シンボル（すなわち、処理単位のデータサイズ）を固定長とすること等によって、データ圧縮器 1 0 における圧縮処理が簡素化され、且つ処理遅延が固定化されることで、大幅な遅延や揺らぎが生じるのを回避できる。このようにして、効率的なデータ伝送を行うことが可能になる。さらに、例えば、図 4 に示したようなパイプライン処理を行うハードウェアを容易に形成できる。

20

【 0 0 5 0 】

< データ解凍器 >

次に、図 1 に示したデータ解凍器 2 0 について説明する。図 5 は、データ解凍器 2 0 の一例を示すブロック図である。データ解凍器 2 0 は、圧縮シンボルと圧縮前の（元の）シンボルペアとを対応付けて記憶するメモリ 2 0 1（読み書き回路含む）と、圧縮前のシンボルペアを 2 つのシンボルに分離する分離器（デマルチプレクサ：D M U X）2 0 2 と、2 つのシンボルを元のストリームデータの順に出力するシリアライザ 2 0 3 と、シリアライザ 2 0 3 を制御する信号を出力する 1 ビットカウンタ 2 0 4 と、当該データ解凍器 2 0 に入力されたシンボル又は復号されたシンボルペアを選択的に出力するセレクタ 2 0 5 とを有し、これらが信号線で接続されている。データ解凍器 2 0 は、所定のクロックに従って同期した動作を行う。

30

【 0 0 5 1 】

まず、伝送路 3 からデータ解凍器 2 0 に入力されるストリームデータ（シンボル列）は、ラッチ 2 0 0 にて一時的に保持され、所定のタイミングでシンボル毎に出力される。出力されたシンボルは、データ解凍器 2 0 のメモリ 2 0 1 とセレクタ 2 0 5 とに入力される。

40

【 0 0 5 2 】

メモリ 2 0 1 は、例えば R A M のような記憶装置である。メモリ 2 0 1 には、データ圧縮器 1 0 のメモリ 1 0 1 と対応するシンボルの組み合わせを登録しておく。すなわち、メモリ 2 0 1 は、図 3 に示したようなルックアップテーブルを記憶する。但し、メモリ 2 0 1 に記憶されるルックアップテーブルには、メモリ 1 0 1 に記憶されるルックアップテーブルの入力の値及び出力の値を入れ替えたエントリが登録される。例えば、メモリ 2 0 1 に記憶されるルックアップテーブルの入力の値には、データ圧縮機 1 0 の C A M が出力するアドレスが登録される。

50

【 0 0 5 3 】

また、メモリ 2 0 1 は、テーブルの入力フィールドに入力されたシンボルを含むエントリが登録されている場合（すなわち、入力されたシンボルが圧縮シンボルである場合）には、当該エントリに登録された元の（圧縮前の）シンボルペアに置換して分離器 2 0 2 に出力する。

【 0 0 5 4 】

分離器 2 0 2 は、シンボルペアを 2 つのシンボルに分割し、シリアライザ 2 0 3 に供給する。シリアライザ 2 0 3 は、シンボルペアをなす 2 つのシンボル（シンボル 1，シンボル 2）を、本来の並び順でセクタ 2 0 5 へ出力する。データ解凍器 2 0 は、所定のタイミングで“ 1 ”又は“ 0 ”を出力する 1 ビットカウンタ 2 0 4 を有している。例えば、1 ビットカウンタ 2 0 4 が“ 1 ”を出力しているときに、シリアライザ 2 0 3 からシンボル 1 が出力される。これに対し、1 ビットカウンタ 2 0 4 が“ 0 ”を出力しているときに、シリアライザ 2 0 3 からシンボル 2 が出力される。カウンタ値とシンボルとの対応関係は逆でも良い。なお、メモリ 2 0 1 と 2 0 5 との経路は、図 4 に示したような D - F F の直列回路に置換することができる。

10

【 0 0 5 5 】

セクタ 2 0 5 は、メモリ 2 0 1 から入力されたシンボルペア、又はラッチから出力されたシンボルとの一方を出力する。図 5 に示す例では、データ解凍器 2 0 は、メモリ 2 0 1 から分離器 2 0 2 へのデータ出力を検知する検知回路 2 0 5 A を備えている。検知回路 2 0 5 A は、メモリ 2 0 1 から読み出されるデータの信号線を監視しており、所定の監視タイミングにおいて、データの読み出しを検知した場合に、セクタ 2 0 5 がシリアライザ 2 0 3 からのシンボルペアを出力するための制御信号（True 信号に相当）をセクタ 2 0 5 に供給する。これに対し、所定の監視タイミングでデータ出力が検知されないときには、セクタ 2 0 5 がラッチ 2 0 0 からのシンボルを出力するための制御信号（False 信号に相当）をセクタ 2 0 5 に与える。

20

【 0 0 5 6 】

以上の構成によって、データ解凍器 2 0 は、ルックアップテーブルを用いて、圧縮シンボルを、元のシンボルペアに変換（復元）することができる。そして、解凍処理の処理遅延を固定的にすることができる。

【 0 0 5 7 】

< データ圧縮 / 解凍システムの動作 >

図 6 を用いて圧縮処理及び解凍処理の全体を説明する。図 6 において、データ圧縮 / 解凍システムは、送信側エンティティとしての送信側装置 1 と、データ圧縮器 1 0 と、受信側エンティティとしての受信側装置 2 と、データ解凍器 2 0 とを含む。送信側装置 1 及び受信側装置 2 のそれぞれは、例えば、通信機能を有する通信機器（端末装置、中継装置）であり、データ圧縮 / 解凍システムは、通信機器間の通信（「機器間通信」と称する）で転送されるデータを圧縮及び解凍する。

30

【 0 0 5 8 】

また、送信側装置 1 及び受信側装置 2 は、例えば、通信機器に搭載されたプロセッサや L S I，A S I C，プログラマブルロジックデバイス（P L D（例えば F P G A））のような電子回路チップであり、データ圧縮 / 解凍システムは、チップ間の通信（いわゆる内部通信）で伝送されるデータを圧縮及び解凍する。伝送路 3 は、機器間通信の場合、機器間を有線又は無線で接続する通信回線であり、内部通信では、チップ間を接続する信号線である。

40

【 0 0 5 9 】

メモリ 1 0 1 には、図 3 に示したようなルックアップテーブルが登録され、メモリ 2 0 1 には入出力の値が逆のテーブルが登録されている。最初に、送信側装置 1 からデータ圧縮器 1 0 へ、送信対象のストリームデータが出力される。ここで、ストリームデータが「 a b d a a b a a」（右側が先頭）のようなシンボル列を含む場合を仮定する。

【 0 0 6 0 】

50

データ圧縮器 10 は、メモリ 101 に登録されている圧縮前のシンボルペアを圧縮シンボルに置換して伝送路 3 へ出力する。ストリームデータ中の上記シンボル列は、ルックアップテーブルに基づいて、圧縮データ「T d a T S」に変換される。一方、データ解凍器 20 は、伝送路 3 を介して到着したストリームデータ中の圧縮シンボルを、圧縮前のシンボルペアに変換する。図 6 の例では、メモリ 201 が図示のようなテーブルを記憶しており、データ解凍器 20 での解凍処理によって、ストリーム中の圧縮データ「T d a T S」は、元の「a b d a a b a a」と変換（復元）される。

【0061】

〔実施形態 2〕

以下、本発明の実施形態 2 について説明する。実施形態 2 は実施形態 1 と共通する構成を含むので、共通する構成については同一の符号を付して説明を省略し、主として相違点について説明する。

【0062】

実施形態 1 では、変換後のシンボルとして、変換前のシンボルが取り得る値の集合に属さない値をルックアップテーブルに登録していた。しかしながら、ストリームデータがバイナリデータである場合のように、変換前のシンボルがあらゆるビット列を取り得る場合では、圧縮シンボルとして変換されるシンボルの値が、ストリームデータに含まれないことは保証できない。すなわち、ストリームデータに含まれるシンボル値と、圧縮シンボルの値とが同じになることが起こり得る。

【0063】

実施形態 2 では、各シンボルに対し、シンボルが圧縮シンボルか否かを示すフラグ（識別子）を設定する。データ圧縮器は、シンボルと当該シンボルに対応するフラグとを出力し、データ解凍器は、フラグが対応シンボルが圧縮シンボルであることを示す場合に、対応シンボルを変換前のシンボルに置換する。フラグは、例えば 1 ビットのデータ（シンボルに付加される識別子ビットとの意で「付加ビット」とも呼ぶ）であり、例えば、ビット値“1”のときに圧縮シンボルであることを示し、ビット値“0”のときに非圧縮シンボルであることを示す。

【0064】

なお、シンボル及びフラグのそれぞれは、ストリームデータ中の順序を保つことにより、先頭から処理すれば自ずと対応付けられる。このようにすれば、データ解凍器において圧縮されたシンボルと圧縮されていないシンボルとを区別して解凍処理を行うことができる。

【0065】

図 7 は、実施形態 2 に係るデータ圧縮器の一例（データ圧縮器 10 a）を示すブロック図である。図 2 に示したデータ圧縮器 10 との相違は、メモリ 101 a から出力されるマッチ信号が、フラグとして扱われる点にある。すなわち、メモリ 101 a は、シンボルペアがテーブルにてヒットしたときに、フラグ“1”を出力し、ヒットしなかったときにフラグ“0”を出力する。フラグ“1”を示す信号は、実施形態 1 におけるマッチ信号としてセレクタ 103 の動作制御に使用される。

【0066】

図 8 は、データ圧縮器 10 a の機能を実現するデジタル回路構成の一例を示す。図 4 に示したデータ圧縮器 10 との相違点は、CAM 101 a から出力される True / False 信号をラッチする D - FF 122 を備え、D - FF 122 からの出力値が、フラグ（付加ビット）として出力される。

【0067】

図 9 は、実施形態 2 に係るデータ解凍器（データ解凍器 20 a）の一例を示すブロック図である。図 5 に示すデータ解凍器 20 との相違点は、データ圧縮器 10 a 側から受信されたフラグ（付加ビット）がセレクタ 205 の制御信号として使用される点にある。データ圧縮器 20 a で生成される各付加ビット（フラグ）は、シンボル数に応じたビット数のビット列として、データ解凍器 20 a に伝達される。データ解凍器 20 a では、ラッチ 2

10

20

30

40

50

00からのシンボルの出力に応じて、対応するフラグ（付加ビット）がセレクタ205に入力される。これによって、フラグ値“1”の場合に、メモリ201から出力された復元後のシンボルペアをセレクタ205から出力することができる。逆に、フラグ値“0”の時に、ラッチ200から出力されたシンボルをセレクタ205から出力することができる。

【0068】

以上のように、実施形態2では、データ圧縮器10aにて、シンボル毎に圧縮/非圧縮を示すビット（フラグ）が生成され、データ解凍器20aへ送信される。これによって、データ解凍器20aは、フラグ値に応じた解凍処理を実行することが可能となる。したがって、例えばバイナリデータが圧縮される場合であっても、データ解凍器20aは、圧縮シンボルと非圧縮シンボルとを区別した解凍処理を実行できる。実施形態2に係る圧縮・解凍処理では、圧縮対象のデータの属性を問わない点で、実施形態1よりも汎用的である。

10

【0069】

〔実施形態3〕

以下、本発明の実施形態3について説明する。実施形態3は実施形態1及び2と共通する構成を含むので、共通する構成については同一の符号を付して説明を省略し、主として相違点について説明する。

【0070】

実施形態3では、実施形態2で説明したデータ圧縮器10a及びデータ解凍器20aをそれぞれ複数段直列に接続することで圧縮率を高めるようにした例を示す。図10は、データ圧縮器10bを多段接続した装置（データ圧縮装置とも呼ぶ）の一例を示す。図10に示すデータ圧縮装置は、直列に接続された4つのデータ圧縮器10bから形成される。但し、データ圧縮装置を形成するデータ圧縮器の数は適宜選択可能である。

20

【0071】

データ圧縮装置に含まれる複数のデータ圧縮器10b1～データ圧縮器10b4を、伝送路3から遠い順に1段目、2段目、・・・4段目と呼ぶ。1段目のデータ圧縮器10b1は、圧縮処理の結果として得られた圧縮シンボルを含むデータ列（圧縮データ（1））と、圧縮データ（1）を形成する各シンボルに対応するフラグ（付加ビット）が、シンボル順と一致する順序で並べられたビット列（第1フラグ列と称する）とが出力される。

30

【0072】

2段目のデータ圧縮器10b2では、データ圧縮器10b1からの圧縮データ（1）に対する圧縮処理によって得られた圧縮データ（圧縮データ（2））と、圧縮データ（2）に対応するフラグ群で形成されたビット列（第2フラグ列と称する）と、第1フラグ列とが出力される。

【0073】

3段目のデータ圧縮器10b3では、データ圧縮器10b2からの圧縮データ（2）に対する圧縮処理によって得られた圧縮データ（圧縮データ（3））と、圧縮データ（3）に対応するフラグ群で形成されたビット列（第3フラグ列と称する）と、第1フラグ列と、第2フラグ列とが出力される。

40

【0074】

4段目（図10の例における最終段）のデータ圧縮器10b4では、データ圧縮器10b3からの圧縮データ（3）に対する圧縮処理によって得られた圧縮データ（圧縮データ（4））と、圧縮データ（4）に対応するフラグ群で形成されたビット列（第4フラグ列と称する）と、第1～第3フラグ列とが出力される。これらの圧縮データ（4）及び第1～第4フラグ列が、伝送路を介してデータ解凍器側へ送信される。

【0075】

図11は、図10に示したデータ圧縮器10b2～10b4の構成例の説明図である。データ圧縮器10b1は、実施形態2で説明したデータ圧縮器10a（図7）と同じ構成を有するので説明を省略する。図11に図示された構成は、最終段（4段目）のデータ圧

50

縮器 10b4 が備える構成を示す。

【0076】

図 11 において、データ圧縮器 10b4 は、データ圧縮器 10a が備える構成に加えて、データ圧縮器 10b4 の入力ポートと出力ポートとを結ぶ 3 つの信号線 131 ~ 133 を含んでいる。信号線 131 は、上記した第 1 フラグ列の伝送用に使用され、信号線 132 は、第 2 フラグ列の伝送用に使用され、信号線 133 は、第 3 フラグ列の伝送用に使用される。信号線 134 には、第 4 フラグ列が出力される。

【0077】

3 段目のデータ圧縮器 10b3 は、図 11 の構成を以下のように変形した構成を持つ。すなわち、図 11 に図示した信号線 133 が省略される。図 11 に図示した信号線 134 が信号線 133 として機能する。また、2 段目のデータ圧縮器 10b2 では、図 11 に図示した信号線 133 及び 132 が省略される。そして、信号線 134 が信号線 132 として扱われる。なお、データ圧縮器 10b1 に関しては、図 7 に示した付加ビットを出力する信号線が、信号線 131 に相当する。

10

【0078】

図 12 は、データ解凍器 20b を複数接続したデータ解凍装置の一例を示す。図 12 に示すデータ解凍装置は、図 10 に示したデータ圧縮装置に対応するものであり、4 段のデータ解凍器 20b から形成されている。伝送路 3 から遠い順に 1 段目、2 段目、・・・n (n = 4、n は自然数) 段目と呼ぶ。

【0079】

4 段目のデータ解凍器 20b4 は、伝送路 3 を介して、圧縮データ (4) 及び第 1 ~ 第 4 フラグ列を受信する。図 12 の例でも、シンボルの経路を太い矢印で示し、付加ビット (フラグ列) を細い矢印で示している。データ解凍器 20b4 は、第 4 フラグ列に基づいて圧縮データ (4) の解凍処理を実行し、圧縮データ (3) と第 1 ~ 第 3 フラグ列を出力する。データ解凍器 20b3 は、第 3 フラグ列に基づいて圧縮データ (3) の解凍処理を実行し、圧縮データ (2) と第 1 及び第 2 フラグ列を出力する。データ解凍器 20b2 は、第 2 フラグ列に基づいて圧縮データ (2) の解凍処理を実行し、圧縮データ (1) と第 1 フラグ列とを出力する。最後に、データ解凍器 20b1 が、第 1 フラグ列に基づいて圧縮データ (1) の解凍処理を実行する。これによって、データ解凍器 20b1 (データ解凍装置) から元のストリームデータが出力される。

20

30

【0080】

図 13 は、実施形態 3 に係るデータ解凍器 20b (20b2 ~ 20b4) の構成例を説明するブロック図である。図 13 に示すデータ解凍器 20b の構成は、図 12 に示したデータ解凍装置の 1 段目に位置するデータ解凍器 20b4 の構成を示す。データ解凍器 20b4 は、データ解凍器 20b の構成 (図 9) に加えて、データ圧縮器 10b1 ~ データ圧縮器 10b3 が出力した第 1 ~ 第 3 フラグ列をデータ解凍器 20b1 ~ データ解凍器 20b3 へ伝送するための信号線 211 ~ 213 を有する。なお、ラッチ 200 には、圧縮データ (4) が入力され、信号線 214 には第 4 フラグ列が入力される。

【0081】

データ解凍器 20b3 は、図 13 に図示する構成から、信号線 213 を省略し、信号線 214 が 213 とされた構成を備える。データ解凍器 20b2 は、図 13 に図示する構成から、信号線 213 及び 212 を省略し、信号線 214 が 212 とされた構成を備える。そして、データ解凍器 20b1 は、図 9 と同じ構成を有し、付加ビットとして第 1 フラグ列が入力される。このように、ある段のデータ圧縮器によって圧縮された圧縮シンボルを、対応する段のデータ解凍器によって解凍するため、伝送路へ送信されるデータ量を削減しつつ、オリジナルデータに復号することができる。

40

【0082】

また、前段のデータ圧縮器の出力を、後段のデータ圧縮器がさらに圧縮するため、複数段のデータ圧縮器を経て伝送路 3 へ出力される転送データは、1 段のデータ圧縮器を用いた場合よりも圧縮率が高くなる。例えば、実施形態 2 の場合、仮にすべてのシンボルペア

50

を圧縮シンボルに置き換えたとしても、付加ビット（フラグ）の分だけデータは増加するため、圧縮率は50%以下にはなり得ない。複数段のデータ圧縮器を用いる場合、データサイズの観点では、1段増加するごとに圧縮率が向上し得る。また、複数段のデータ圧縮器又はデータ解凍器にわたってパイプライン処理を行うことができるため、処理速度の点でも有利である。

【0083】

なお、実施形態3では、実施形態2のデータ圧縮器10a及びデータ解凍器20aをそれぞれ複数段に接続したが、実施形態1に示したデータ圧縮器10及びデータ解凍器20をそれぞれ複数段接続するようにしてもよい。

【0084】

〔実施形態4〕

以下、本発明の実施形態4について説明する。実施形態4は実施形態1～3と共通する構成を含むので、共通する構成については同一の符号を付して説明を省略し、主として相違点について説明する。

【0085】

実施形態1～3では、ストリームデータの先頭から順に、2つのシンボルをペアにして圧縮するか否かを判断した。ここで、データストリームを2シンボルより長い或る長さの単位で見た場合、仮にデータストリームの異なる位置に同一の或る長さのシンボル列が出現する場合であっても、シンボルペアの区切り目が前後にずれた場合は、別のシンボルペアの列として処理される。

【0086】

ここで、例えばCAMに記憶されるルックアップテーブルに登録できるエントリの数には限界があるところ、何らかのルールに基づいて、同一のシンボル列は同一のシンボルペアに区切ることができれば、より少ないエントリで同一の或る長さのシンボル列を圧縮できる。すなわち、全体として圧縮率の向上が期待できる。実施形態4では、シンボル間の区切位置を決定する回路を含むデータ圧縮器を用いて、決定された各区切位置の先頭から2つずつのシンボルを組み合わせてシンボルペアにする。

【0087】

図14に、本実施の形態に係るデータ圧縮器10cのブロック図を示す。図14のデータ圧縮器10cは、入力データとしてのストリームデータを一時的に記憶するラッチ105と、メモリ101aと、シリアライザ102と、セクタ103と、区切位置決定回路104とを有し、これらが信号線で接続されている。データ圧縮器10cは、図11に示したデータ圧縮器10bに、区切位置決定回路104を加えた構成になっている。ここでは、図11に示したデータ圧縮器10bとの差異を中心に説明する。

【0088】

区切位置決定回路104は、シンボル列の入力を受け、所定の条件に基づいてシンボル間の区切位置を決定する。実施形態4では、予め定められた優先度に基づいて隣接する4つのシンボルを比較し、区切位置を決定する。また、区切位置決定回路104は、入力されたシンボル列の順にシンボルを出力するとともに、区切位置を示す信号を出力する。また、シリアライザ13は、シンボルペアを、入力されたシンボル列と同じ順序で出力する。なお、実施形態4では、便宜的に、区切位置決定回路104までの部分を区切位置決定部と呼び、これ以降の部分を圧縮部と呼ぶ。圧縮部は、データ圧縮器10bと同様の構成を備える。

【0089】

実施形態4では、区切位置決定回路104は、例えば、シンボルが示す値の大小関係を用いて、予め定められた、連結する2シンボルの大小関係の優先度に基づき、シンボル間の区切位置を決定する。具体的には、区切位置決定回路104は、シンボルが示すビット列を数値として扱い、隣接する2つのシンボルの大小を比較する。そして、数値が増加し続けている単調増加区間（「増加列」とも呼ぶ）、数値が減少し続けている単調減少区間（「減少列」とも呼ぶ）、又は数値が等しい区間（「等価列」とも呼ぶ）に分割し、区間

10

20

30

40

50

の境界を区切位置とする。このとき、例えば「等価列」、「増加列」、「減少列」の順に優先順位を定め、区間の境界に位置するシンボルを優先順位の高い区間に組み入れるようにしてもよい。なお、区間に1つのシンボルしか含まれない場合は、例えば同様の優先順位に基づいて前後の区間に組み入れるようにしてもよい。

【0090】

以上をまとめると、本実施の形態では、次の2点のルールが予め定められている。

- (1) 文字の大小関係： $a < b < c < d \dots$
- (2) 区切目の優先順位：等価列 > 増加列 > 減少列

【0091】

このようなルールによれば、隣接する4つのシンボルを比較することによって区切位置を決定することができる。換言すれば、前後のシンボルを4つ保持するだけで当該区切位置決定処理を実行することができる。このため、回路構成の増大を抑えることができる。

10

【0092】

次に、区切位置決定回路104を含むデータ圧縮器およびこれに対応するデータ解凍器の実装例を説明する。図15は、図14に示した区切位置決定回路104の一例を示す回路構成図である。図15に示す区切位置決定回路104は、ストリームデータをシンボル単位で保持し、クロック信号に従って伝送するD-FF141a~141cと、2シンボルの大小関係と比較器142a~142cと、比較結果を保持するQ-FF143a~143dと、比較結果に基づいて圧縮部のCAM101を有効化するタイミング(すなわち、ルックアップテーブルを検索するタイミング)を示す信号を出力するエンコーダ144とを有する。

20

【0093】

D-FF141a、D-FF141b、及びD-FF141cは、直列に圧縮部と送信側装置1との間を接続し、送信側装置1が出力したストリームデータをシンボル単位で圧縮部へ伝送する。比較器142aは、D-FF141aの出力とD-FF141bの出力と(すなわち、ある時点における、ストリームデータ中の隣接する1番目及び2番目のシンボル)を受け取り、シンボルの示す数値の関係を「増加(<)」、「等価(=)」又は「減少(>)」のいずれかで出力する。なお、「増加」、「等価」又は「減少」を示す信号として、例えばそれぞれ「1」、「2」又は「3」のような識別子を出力する。同様に、比較器142bは、D-FF141bの出力とD-FF141cの出力と(同様に、隣接する2番目及び3番目のシンボル)を受け取り、「増加」、「等価」又は「減少」のいずれかを出力する。また、比較器142cは、D-FF141cの出力と、区切位置決定部への新たな入力と(同様に、隣接する3番目及び4番目のシンボル)を受け取り、「増加」、「等価」又は「減少」のいずれかを出力する。なお、長さが1シンボルの区間を前後いずれかの区間に組み入れるようにしてもよい。

30

【0094】

比較器142a~比較器142cの出力は、それぞれQ-FF143a~Q-FF143cに保持される。さらに、Q-FF143a~Q-FF143cに保持された値および後述するQ-FF143dに保持される値は、エンコーダ144に入力される。Q-FF143dの入力はエンコーダ144の出力と接続され、1クロック前にエンコーダ144が出力した信号が保持されている。なお、エンコーダ144の出力が「TRUE」のタイミングで、後述する圧縮部ではシンボルペアをメモリから検索し、圧縮を試みる。

40

【0095】

図16に、エンコーダ144の入力と出力との組合せを示す。図16の表は、「入力」の列(「1・2番目」、「2・3番目」、「3・4番目」及び「前回の出力」と、「出力」の列を有している。入力の各列は、対応するシンボルを比較する比較器における大小関係を示している。「1・2番目」の列は、Q-FF143aに保持される大小関係に相当する。「2・3番目」の列は、Q-FF143bに保持される大小関係に相当する。「3・4番目」の列は、Q-FF143cに保持される大小関係に相当する。そして、エンコーダ144は、各比較器における大小関係が一致するレコードの「出力」列に記載され

50

た信号を出力する。

【0096】

1行目は、比較器142bに入力される2シンボルが等しく、エンコーダ144からの1クロック前の出力がFALSEの場合、前後のシンボルにかかわらず、当該クロックでのエンコーダ144からの出力がTRUEになることを表す。なお、入力列の「1・2番目」及び「3・4番目」に記載された「ANY」は、「<」、「>」及び「=」のいずれでもよいことを示している。2行目、3行目は、それぞれ、単調増加区間、単調減少区間であって、エンコーダ144からの1クロック前の出力がFALSEの場合に、当該クロックでのエンコーダ144からの出力がTRUEになることを表す。すなわち、等価区間、単調増加区間、単調減少区間においては、エンコーダ144の出力は1シンボルごとにTRUEとFALSEを繰り返し、区切位置決定部における2番目及び3番目のシンボルを、圧縮部においてシンボルペアとしてルックアップテーブルを検索して変換を試みる。また、4行目は、比較器142bに入力される2シンボルが単調増加区間の終わりに相当し、エンコーダ144からの1クロック前の出力がFALSEの場合、当該クロックでのエンコーダ144からの出力がTRUEになることを表す。5行目は、比較器142bに入力される2シンボルが単調増加区間の始まりに相当し、エンコーダ144からの1クロック前の出力がFALSEの場合、当該クロックでのエンコーダ144からの出力がTRUEになることを表す。6行目は、比較器142bに入力される2シンボルが、2つの単調減少区間の境界に位置する、長さが2シンボルの単調増加区間に相当し、エンコーダ144からの1クロック前の出力がFALSEの場合、当該クロックでのエンコーダ144からの出力がTRUEになることを表す。

10

20

【0097】

図16の例では、区切り目に当たるシンボルを組み入れる先を、「等価列>増加列>減少列」の優先順位に基づいて決定している。すなわち、1行目からわかるように、前後のシンボルとの大小に関係なく、等価列が最も優先的に連結され、区間を形成する。また、4~6行目からわかるように、減少列よりも増加列が優先して連結され、区間を形成する。このようなルールは一例であり、所定のルールに従って区切位置を決定すれば、同一のシンボル列を同一のシンボルペアに区切ることができる。

【0098】

図17を用いて、区切位置を決定する処理を説明する。例えば、図3に示したルックアップテーブルを用いて、図17のシンボル列1を圧縮する場合を説明する。

30

【0099】

シンボル列1には「bcdeacbdddabcd a a a d c」という長いシンボル列が2つ含まれている(図17:シンボル列1の下線部)。実施形態4では、まずシンボル列の先頭を最初の区切位置として、シンボル列の先頭から2シンボルをシンボルペアにしてルックアップテーブルを検索する(図17:ステップS1)。ここで、S1に示すように、ルックアップテーブルにシンボルペアが登録されていなければ、変換はされない。また、区間の境界となる区切位置があれば、区切位置の次から2シンボルをシンボルペアとしてルックアップテーブルを検索する(図17:S2)。ここで、S2に示すように、ルックアップテーブルにシンボルペアが登録されていれば、変換シンボルに置換される。なお、図17の「cc」は等価区間、「bcde」は増加区間であり、これらの境界が区切位置となる。以降も同様にして、区切位置を検出しつつ、各区間の先頭からシンボルペアの変換を行う(図17:S3)。なお、図17では、これ以降の処理の説明は省略する。

40

【0100】

シンボル列1は、S1~S3の処理によって、図17のシンボル列2に示すような区切りが挿入された状態になる。シンボル列2が図3に示した内容のルックアップテーブルに基づいて変換されると、図17のシンボル列3で示すような圧縮データになる。なお、図17では便宜的に区切り記号を記載しているが、ストリームデータに区切り記号を示すシンボルやビットが追加されるわけではない。

【0101】

50

シンボル列 3 で、シンボル列 1 に含まれていた 2 つの「 b c d e a c b d d d a b c d a a a d c 」は、それぞれ「 V Z | U | b | Y d | T W | S a | X 」及び「 T W e | U b | Y d | T W | S a | X 」に変換されている。このように、複数の区間にわたってシンボル列が一致する場合、シンボル列が一致する複数の区間のうち、両端の区間以外は同じシンボルペアが生成される。図 17 の例では、シンボル列 2 中の 2 つのシンボル列において、最初の区間のシンボル列のそれぞれは、「 c c 」と「 a b c d e 」であり、両者は異なっている。同様に、最後の区間のシンボル列のそれぞれは、「 d c a 」と「 d c c 」とであり、両者は異なっている。これに対し、これらの最初及び最後のシンボル列で挟まれた区間のシンボル列（シンボル 2 の斜体参照）は一致する。よって、当該挟まれた区間は、同一のロジックで変換されることになる（シンボル 3 の斜体参照）。

10

【 0 1 0 2 】

一方、図 3 に示したルックアップテーブルを用いて、シンボル列 1 を先頭から 2 シンボルずつシンボルペアにして圧縮した場合（すなわち、実施形態 1 ~ 3 の場合）、図 17 のシンボル列 4 になる。

【 0 1 0 3 】

シンボル列 4 は、シンボル列 3 よりも 3 シンボル分だけデータ量が多く、すなわち、圧縮率が悪い。換言すれば、同程度の圧縮率を達成するためには、ルックアップテーブルにより多くのエントリが必要になる。このように、実施形態 4 によれば、より少ないエントリ数で同一の長いシンボル列を圧縮できるため、全体として圧縮率を向上させることができる。

20

【 0 1 0 4 】

実施形態 4 における圧縮部の回路構成は、図 4、図 8 に示したものをそれぞれ採用することができる。より詳細には、図 18 に示すようにエンコーダの出力（CAM を有効化するための信号）は、タイミング合わせのディレイ回路（D - F F 1 1 4 ~ D - F F 1 1 6）を經由して CAM へ入力される。また、実施形態 4 においては、例えば、図 5、図 9、図 13 と同様のデータ解凍器を採用することができる。

【 0 1 0 5 】

実施形態 4 に示したデータ圧縮器及びデータ解凍器は、実施形態 1 ~ 3 の少なくとも一部と組み合わせることもできる。すなわち、付加ビット（フラグ）を用いずに、ルックアップテーブルには圧縮後のシンボルとして圧縮前のシンボルが取り得ない値を登録しておいてもよい。また、実施形態 4 に示すデータ圧縮器及びデータ解凍器をそれぞれ複数段接続するようにしてもよい。

30

【 0 1 0 6 】

また、区切位置の決め方は上記の方法には限定されない。例えば、L C A（Lowest Common Ancestor：最少共通祖先）と呼ばれるアルゴリズムを用いて、効率のよいシンボルペアを特定して区切位置を決定してもよいし、その他の方法を用いてもよい。

【 0 1 0 7 】

< 変形例 >

上記の実施形態では、2 つのシンボルを 1 つのシンボルに圧縮したが、圧縮前のシンボルの組合せは 2 つには限られない。複数のシンボルを、それ未満の数のシンボルに置き換えるという構成であれば、データ圧縮器として機能する。ただし、ルックアップテーブルのエントリとストリームデータのシンボルペアとの一致率を向上させるという観点や、データの記憶容量を抑えるといった観点からは、2 つのシンボルを 1 つのシンボルに圧縮するという態様が好ましいといえる。

40

【 0 1 0 8 】

また、上記の実施形態では、1 番目のシンボル及び 2 番目のシンボルのシンボルペアがルックアップテーブルに登録されていない場合、次の処理対象は 3 番目のシンボル及び 4 番目のシンボルとし、これらのシンボルペアがルックアップテーブルに登録されているか判断していた。ここで、1 番目のシンボル及び 2 番目のシンボルのシンボルペアがルックアップテーブルに登録されていない場合、次の処理対象を 2 番目のシンボル及び 3 番目の

50

シンボルのシンボルペアとしてもよい。このようにすれば、圧縮率の向上が期待できる。

【0109】

ルックアップテーブルのエントリは、圧縮対象のデータにおけるシンボルの出現尤度に基づいて変更するようにしてもよい。例えば、メモリがシンボルペアとエントリとの一致率を算出し、一致率が所定の閾値以下に下がった場合、ルックアップテーブルのエントリを更新するようにしてもよい。このとき、ルックアップテーブルのエントリは、送信側装置1及び受信側装置2に接続された管理装置が配布するようにしてもよい。例えば、圧縮対象のデータが、いわゆるミニプログのタイムラインである場合、時間の経過と共に投稿される話題が変化し、これに伴い圧縮対象として好ましいシンボルペアも変化する。シンボルペアとエントリとの一致率が所定の閾値以下になった場合、ルックアップテーブルのエントリを更新することで、圧縮率の低下を避けることができる。

10

【0110】

本発明に係るデータ圧縮器及びデータ解凍器は、例えば、ネットワークに接続されたコンピュータのネットワークインターフェイス（ネットワークカード）に設けてもよいし、複数のプロセッサを接続するバスの2点間に設けてもよい。ネットワークインターフェイスに設ける場合は、例えばイーサネット（登録商標）においてフレーム（MTU：Max Transmission Unit、パケットとも呼ぶ）に分割（フラグメント）する前の段階に設ける。なお、フレーム内に、例えばシンボル列（圧縮データ）及び各段の付加ビット列（フラグ列）のそれぞれを識別可能な識別子を追加してもよい。識別子は、例えば、圧縮データ及び各フラグ列がフレーム上にマッピングされた際の位置情報のような、受信側で、多重された圧縮データ及びフラグ列を切り分け可能とする情報を含む。

20

【0111】

また、本発明に係るデータ圧縮器及びデータ解凍器は、いわゆるビッグデータの転送に適用してもよいし、特定の装置間でのデータのバックアップに適用してもよい。また、例えば、コンピュータウィルスのパターンデータとネットワークを流れるデータとをそれぞれ圧縮した状態で比較するような場合も、処理速度の向上が期待できる。

【0112】

ところで、上述した実施形態2では、データ圧縮器からは、シンボル列（圧縮データ）とフラグ列との2つのデータ系列が出力される。また、実施形態3では、シンボル列（圧縮データ）に加え、データ圧縮器の段数に応じた2以上のフラグ列が出力される。データ圧縮器から出力される圧縮データは、上記したように、伝送路3を介してデータ解凍器へ伝送される。このとき、データ圧縮器とデータ解凍器との間が並列な専用回線で接続される構成が採られる場合には、データ圧縮器が備える各出力ポートからの出力（各データ系列）をデータ解凍器に備えられた、各データ系列に対応する入力ポートに接続することができる。

30

【0113】

これに対し、データ圧縮器からの出力が、所定の伝送媒体（例えばMACフレーム）にマッピングされて伝送路3上を伝送される場合がある。このとき、圧縮データ及びフラグ列の伝送媒体へのマッピングは、データ圧縮器よりも下位に位置するレイヤのプロトコルに従う。この結果、データ解凍器側の下位レイヤが、圧縮データとフラグ列とが直列に接続されたデータ列を、データ解凍器の位置するレイヤ（「解凍レイヤ」と称する）に引き渡すことが起こり得る。この場合、解凍レイヤにおいて、圧縮データ及びフラグ列を適正なデータ解凍器の入力ポートに接続するために、圧縮データとフラグ列とを切り分ける（分離する）ことが要求される。このため、例えば以下のような構成が採られる。

40

【0114】

図19に、圧縮データとフラグ列とが直列に接続されたデータ列（ビット列）を、データ解凍器においてデータ系列ごとに分離可能なデータフォーマットの一例を示す。図11及び図13に示したデータ圧縮器10b、データ解凍器20bを例に説明すると、データ圧縮器10dから出力される圧縮データ（4）及び第1～第4フラグ列の5つのデータ系列は、例えば、図示しないパラレル/シリアル変換器で、所定順序（例えば、圧縮データ

50

、第1フラグ列、第2フラグ列、第3フラグ列、第4フラグ列)を有する1つのシリアルなデータ列に変換される。データ列は、1つのデータブロックとして扱われる。さらに、各データ系列のビット数が図示しないカウンタで計数され、計数結果に基づくヘッダが生成される。ヘッダは、例えばヘッダサイズと、データブロック中における各データ系列のサイズとを少なくとも含む。ヘッダ中のヘッダサイズ及びデータ系列のサイズは、例えば、固定長ビットで表現される(これにより、ヘッダサイズは固定長となる。なお、ヘッダサイズが固定長(解凍側にて既知)のとき、ヘッダにヘッダサイズを含めるのを省略できる)。このようなヘッダが、データブロックの前段にセットされて、下位レイヤに渡される。なお、ヘッダには、データサイズの代わりにオフセット(各データ系列の開始位置)及びデータブロックのサイズを記憶させるようにしてもよい。

10

【0115】

データ解凍器20b(図14)側では、データ解凍器の前段に、図示しないヘッダ解析回路が置かれる。ヘッダ解析回路は、下位レイヤからヘッダ及びデータブロックからなるデータ列(図19)を受け取り、ヘッダに格納された情報(ヘッダサイズ、各データ系列のサイズ)を参照して、データブロックを5つのデータ系列に分離する。図19の例では、ヘッダは固定長であり、ヘッダに後続するデータブロックの開始位置は決まっている。また、第1フラグ列の開始位置は、データブロックの先頭から圧縮データのサイズだけ後方になる。同様に、第2~第4フラグ列の開始位置も、圧縮データのサイズ及び第1~第3フラグ列のサイズを用いて求めることができる。また、データブロック全体のサイズは、圧縮データのサイズ及び第1~第4フラグ列のサイズの合計になる。このように、データ解凍器側は、各データ系列のオフセット及びサイズを求めることができ、データブロックを各データ系列に分離することができる。分離されたデータ系列の一つ(圧縮データ(4))は、ラッチ200へ入力(接続)される。また、第4フラグ列は、セレクタ205aに適宜のタイミングでフラグ値を供給するためのバッファ(図示せず)に蓄積される。第1~第3フラグ列は、後段に位置する他のデータ解凍器へ送られる。以上のように、解凍側で複数のデータ系列を分離するためのプロトコルが、圧縮側と解凍側との間で決められる。

20

【0116】

以上説明した実施形態1~4の構成は、適宜組み合わせることができる。

【符号の説明】

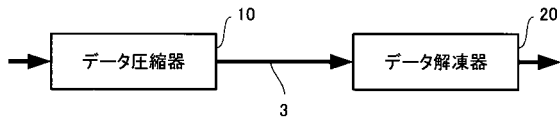
30

【0117】

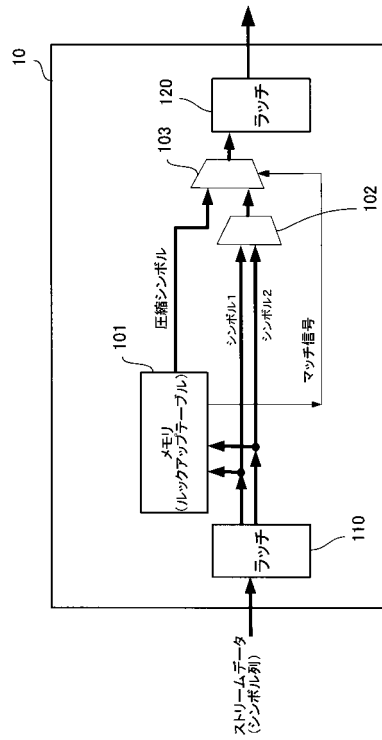
10, 10a, 10b データ圧縮器
 101, 101a メモリ(CAM)
 102 シリアライザ(マルチプレクサ)
 103 セレクタ(マルチプレクサ)
 104 区切位置決定回路
 20, 20a, 20b データ解凍器
 201 メモリ
 202 分離器(デマルチプレクサ)
 203, 205 セレクタ
 204 1ビットカウンタ
 3 伝送路

40

【 図 1 】



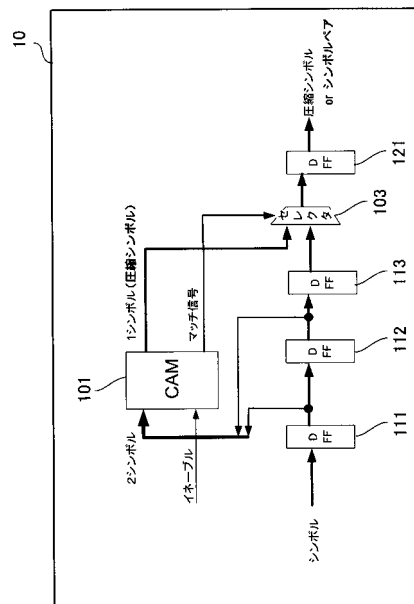
【 図 2 】



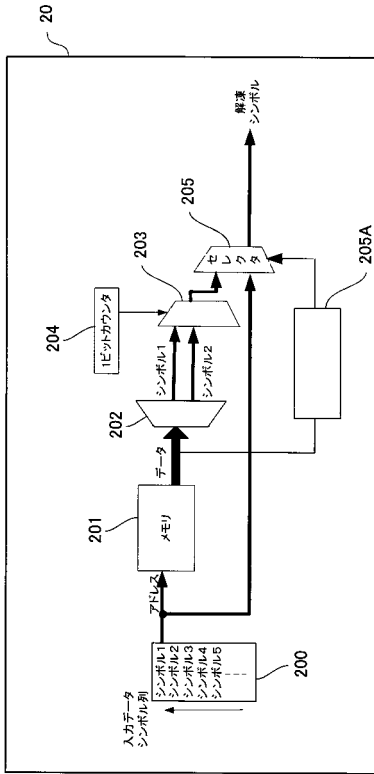
【 図 3 】

入力	出力
aa	S
ab	T
ac	U
bc	V
cd	W
dc	X
dd	Y
de	Z
⋮	⋮

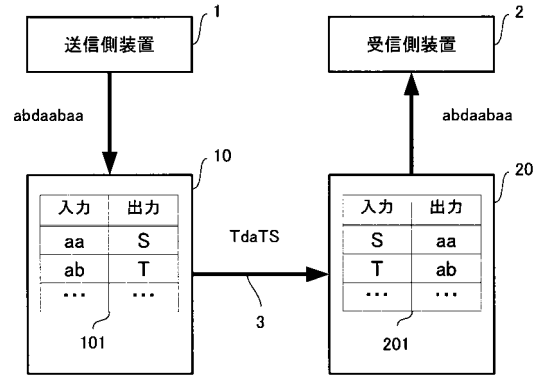
【 図 4 】



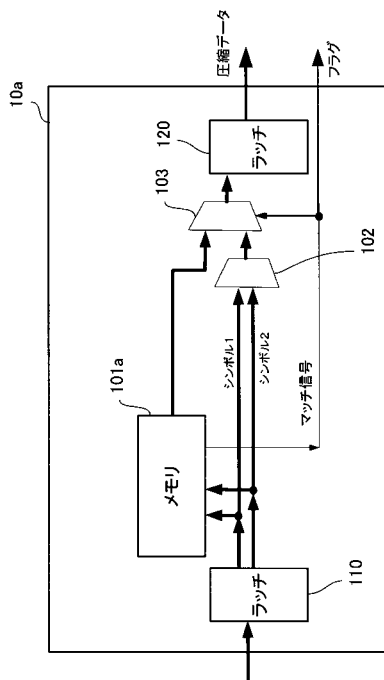
【 図 5 】



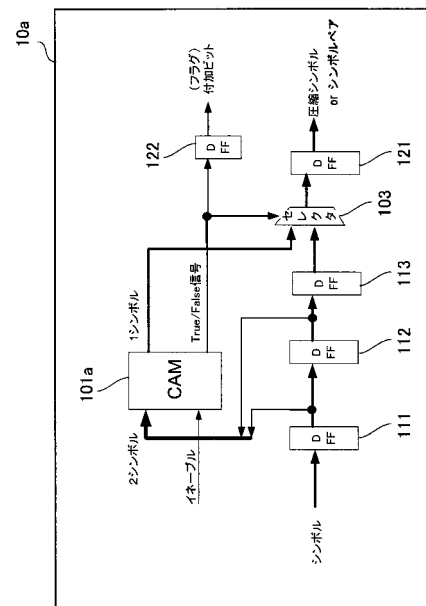
【 図 6 】



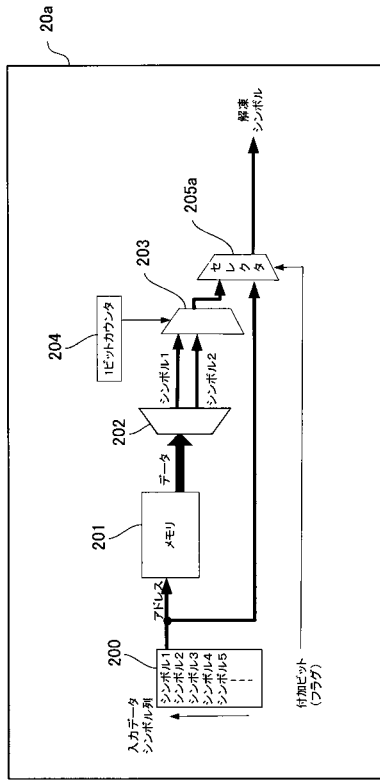
【 図 7 】



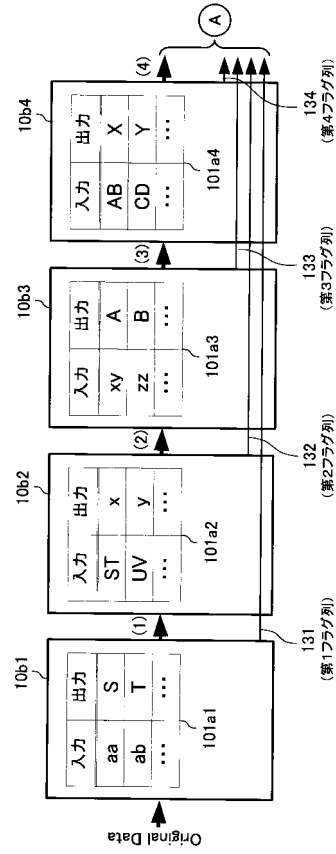
【 図 8 】



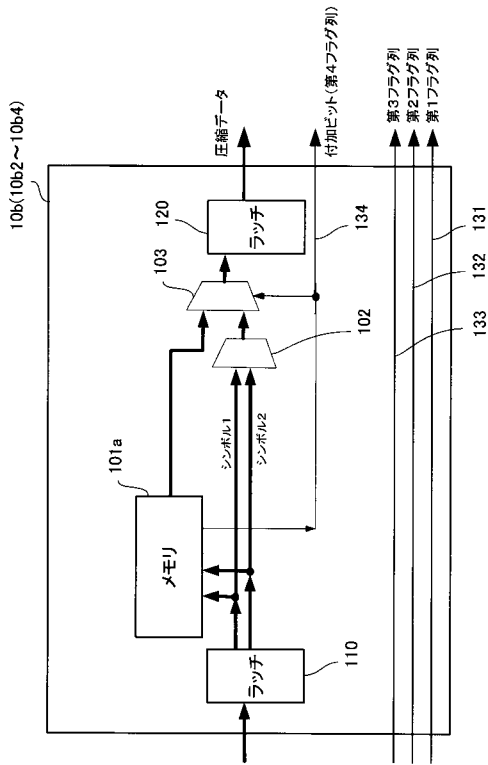
【 図 9 】



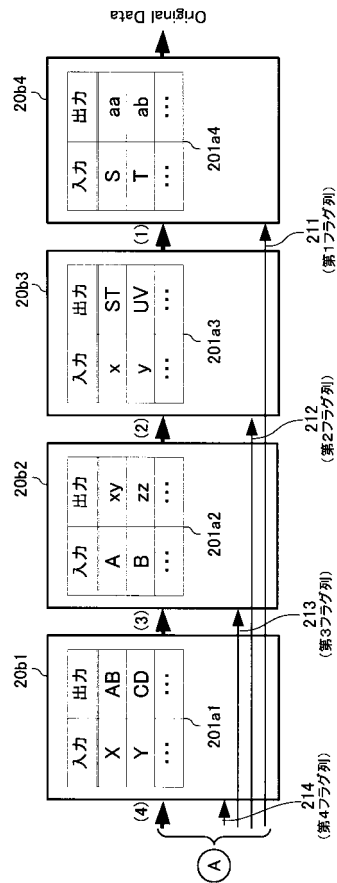
【 図 10 】



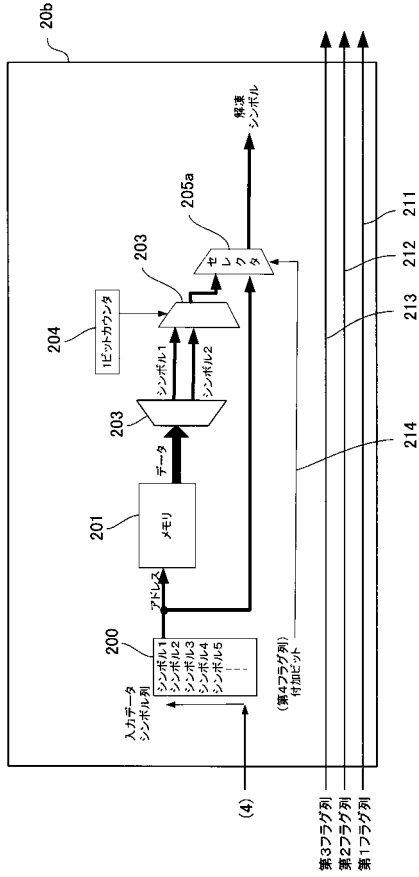
【 図 11 】



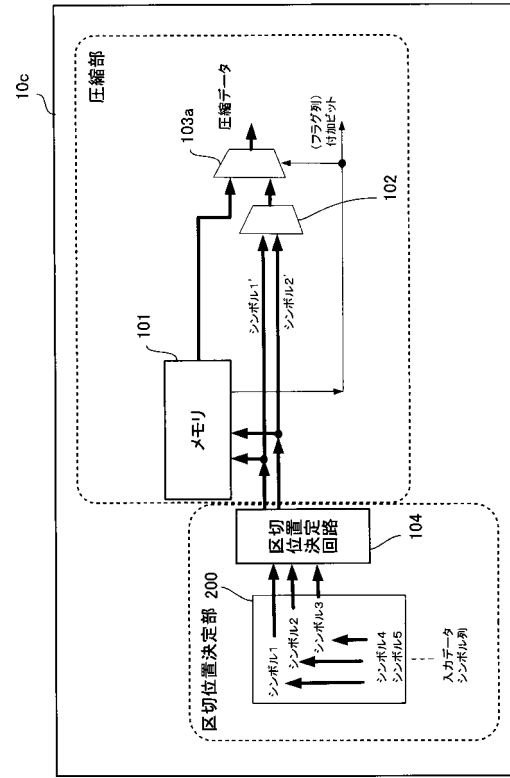
【 図 12 】



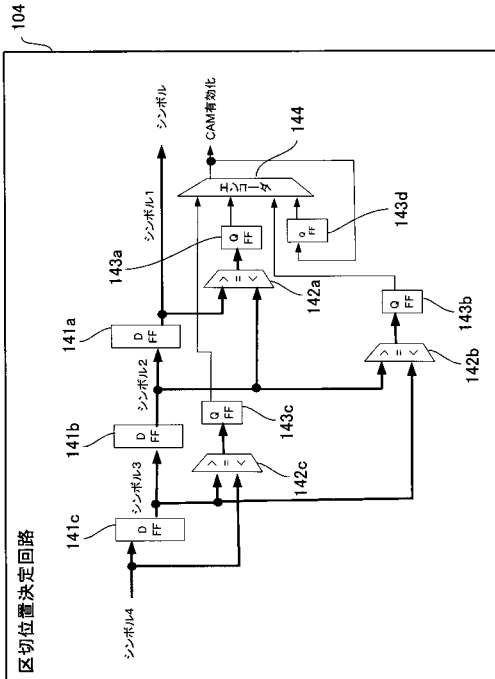
【図 13】



【図 14】



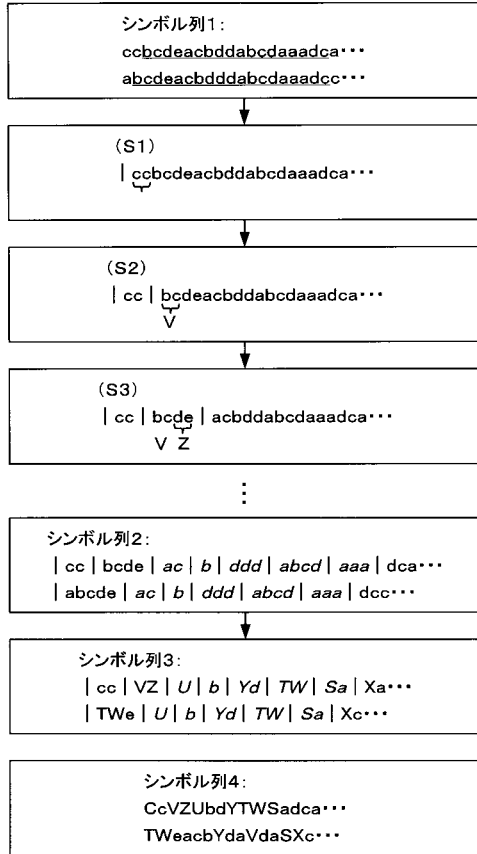
【図 15】



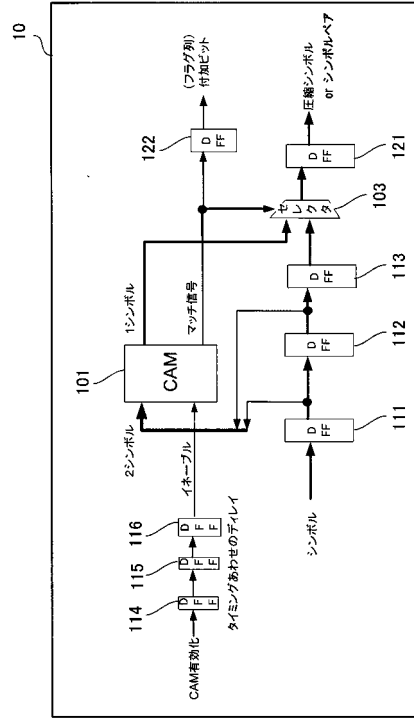
【図 16】

入力				前回の出力	出力
1・2番目	2・3番目	3・4番目			
ANY	=	ANY	FALSE	TRUE	
<	<	<	FALSE	TRUE	
>	>	>	FALSE	TRUE	
<	<	>	FALSE	TRUE	
>	<	<	FALSE	TRUE	
>	<	>	FALSE	TRUE	
(上記以外の組合せ)					FALSE

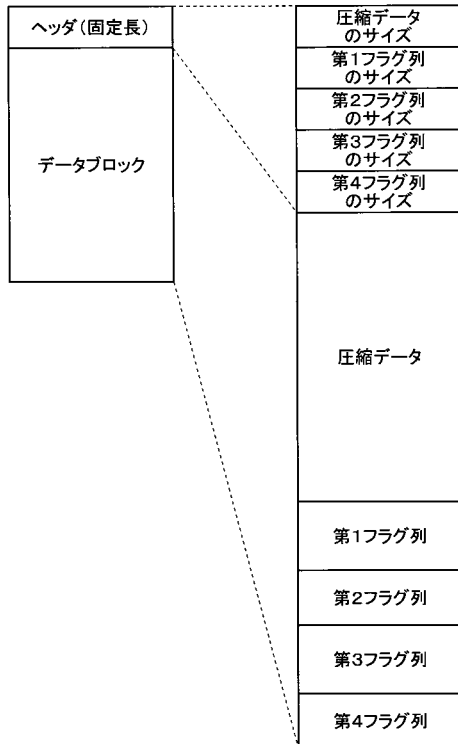
【 図 1 7 】



【 図 1 8 】



【 図 1 9 】



フロントページの続き

(72)発明者 坂本 比呂志

福岡県北九州市戸畑区仙水町 1 - 1 国立大学法人九州工業大学内

Fターム(参考) 5J064 AA03 BA09 BB13 BC01 BC25