

令和 5 年 10 月 27 日現在

機関番号：13701

研究種目：基盤研究(C) (一般)

研究期間：2019～2022

課題番号：19K11822

研究課題名(和文) 新しい高密度記録のための連長制約符号及び組合せ論的な記録符号の理論と応用

研究課題名(英文) Theory of run-length constraint codes and combinatorial recording codes for novel high density recording

研究代表者

鎌部 浩 (KAMABE, Hiroshi)

岐阜大学・工学部・教授

研究者番号：80169614

交付決定額(研究期間全体)：(直接経費) 2,000,000円

研究成果の概要(和文)：記録のための符号を、様々な応用を仮定して行った。結果は五つに大別できる。(1)二次元誤り訂正のためのLDPC符号の構成方法を拡張し、より能力の高い符号の構成方法を与えた。(2)バランス符号のためのDubeらのアルゴリズムを基礎にして、テーブル参照型の符号化を提案し、その冗長度を理論的に示した。(3)MAMR(Microwave Assisted Magnetic Recording)に、マルチユーザ符号と通信路推定の技術を応用した。(4)PAC符号の、高性能な復号法について検討した。(5)分散ストレージにおいて、キャッシングを用いたときのファイルの数を抑えるような符号の構成方法を見出した。

研究成果の学術的意義や社会的意義

現代社会では、利用者が所有する端末の記録装置やクラウドに置かれたデジタルデータが非常に大きな役割を果たしている。インターネット上の様々なツールや最近注目を集めている人工知能なども、大量のデータを自由に扱える環境が基礎となっている。したがって、これらのデータを長期間正しく保管し、すぐに提供できるようにしておくことは社会的に非常に大きな意味がある。ところでデジタルデータは年々指数関数的に増加しており、これらをすべて保管していくためには、記録密度を向上させるための定常的な技術革新が必要となる。本研究の成果は、こうした社会的要請に応えるデジタル記録装置の研究開発において、理論的な基礎を与える。

研究成果の概要(英文)：The process of creating codes for recording was conducted, assuming various applications.

(1) The method for constructing LDPC codes for two-dimensional error correction was expanded, providing a way to construct more capable codes. (2) Based on Dube et al.'s algorithm for balanced codes, a table-lookup type encoding was proposed, and its redundancy was theoretically given. (3) Multi-user codes and communication channel estimation techniques were applied to MAMR (Microwave Assisted Magnetic Recording). (4) We examined a high-performance decoding method for PAC codes. (5) In distributed storage, a method was found for constructing codes to reduce the number of files when using caching.

研究分野：記録のための信号処理

キーワード：バランス符号 有限状態符号 ランク変調符号 二次元誤り訂正 ポーラー符号 多層磁気記録

1. 研究開始当初の背景

ハードディスクやDVDなどの光ディスク、フラッシュメモリ、磁気テープなど全ての記録装置の年間総生産容量は、定常的な技術革新によって毎年およそ1.4倍で増えているが、世界中で生産されるデジタル情報の年間の総量は数年後にはこれを超えると予想されている。従って、今後もこれまで以上の技術革新と新しい記録技術の開発が求められている。

現在、多くの新しい記録技術が研究されているが、その中で本研究では次の三つの記録技術のための符号を対象とする；(1) 符号化技術によるフラッシュメモリなどの固体不揮発性メモリ、(2) DNAストレージ、(3) マイクロ波アシスト磁気記録による3次元記録。これらの記録技術はその原理は異っているが、同様の入力制約を持つ通信路としてモデル化できる場合がある。ここで、入力制約を持つ通信路とは、記録が物理的に困難である、または、読み出し時に非常に多くの誤りを生成するような入力系列もしくは入力パターンがあるような通信路を意味する。

入力制約を持つ記録通信路にデータを記録するためには、記録が困難であるようなパターンを出力しないような符号器が必要になる。そのような入出力関係を与える符号化規則は一般に記録符号と呼ばれる。不揮発性固体メモリや磁気記録装置、光記録、DNAストレージなどのほとんどでは、この符号を用いて、データ(誤り訂正符号化された系列)は以下のように符号化されて記録される。

データ \Rightarrow 誤り訂正符号化 \Rightarrow 記録符号化 \Rightarrow 入力制約を持つ通信路 読み出し時には、この逆の順序で復号される。本研究で対象とするのは、上の記録符号の部分である。この符号化技術によって、記録装置が適切に動作することを保証できる。以下では、対象とする記録通信路について説明する。

入力制約(連長制約やバランス制約など) 連長制約や零スペクトル制約は、入力制約の典型例である。これらは、これまで磁気記録や光記録などの記録通信路のための符号として研究され、目的に応じて様々な構成方法が提案されてきている。しかしながら、高い符号化率でいくつかの入力制約を同時に満たすような符号を、現実の装置で実現可能な程度の規模の符号化規則で構成するための一般的な理論は、まだ知られていない。DNAストレージや相変化メモリでは、まさにこうした規模の大きな符号を必要としているため、連長制約のための、規模の大きな記録符号を構成する方法の研究が必要とされている。

2次元および3次元磁気記録 記録密度を向上させるために、磁気記録のトラック間間隔は狭められており、瓦書き記録では0になっている。このような状況では、磁気記録は実質的に二次元的に行われると仮定して信号処理をすることが必要になってくる。また、これによって二次元の誤りに対応した処理も必要になってくる。磁気記録の記録密度をさらに向上させるため、マイクロ波を使って読み書きを行う記録方式が研究されており、実用化間近となっている。この技術を使うと、媒体表面だけではなく、一つのディスクを層に分けて記録できる、つまり3次元的な磁気記録が可能になる。もし、各層の情報を同時に読み出す(一括読み出し)ことができれば、読み出し速度も向上する。これを可能にするために、マルチユーザー符号、WOM符号もしくはPRIO符号のような符号が応用できる。一方で、磁気記録には入力制約があるので、一括読み出しが可能で入力制約を満たす符号を構成することが必要になる。

2. 研究の目的

本研究の目的は、記録のための信号処理における誤り訂正を目的として、通信で使用される誤り訂正方式を適用することと、入力制約のある記録通信路のための新しい制約符号の開発である。具体的には、二次元誤り訂正符号とPolar符号の性能向上、MAMARへの多重アクセス通信方式の応用、DNAストレージのための符号の開発、バランス符号に関する研究を目的とした。

二次元誤り訂正符号について：5Gのような、超高速通信システムと高解像度な放送システムが実現されているため、超大容量な記録システムの必要性は増加し続けている。Woodらは、非常に高い記録密度を達成するために、二次元磁気記録(TDMR)の概念を提案した。TDMRの読み出しヘッドのサイズは、記録トラックの幅より大きく、隣接トラックの間にギャップがないため、複数のトラックの信号を同時に検出しなければならない。従って、記録媒体の傷などは、二次元状の誤りを引き起こす。二次元的に信号処理を行うことで誤り訂正の性能を上げることはできるが、その処理を実装するのは容易ではない。これまで幾つかの二次元誤り訂正符号が提案されている、一般化BPアルゴリ

ズムを適用を目的として、本研究では LDPC 符号を対象とした。

Polar 符号の性能向上について：NAND 型フラッシュメモリは、高密度化によって誤り率が高くなっている。このため、誤り訂正機構の導入が必須となっている。ここでは、速度や遅延を犠牲にすることなく、高い誤り訂正能力を示す符号として Polar 符号を考え、その復号の性能向上を目的とした。

MAMR への多重アクセス通信方式の応用について：社会のデータストレージに対する要求を満たすために、多くの技術が研究されており、それらの中の 하나가、MAMR(Microwave-Assisted Magnetic Recording) である。MAMR は、マイクロ波を補助的に用いて、保磁力の高い磁性媒体でも、小さいヘッドで生成される弱い磁界で記録することを可能にする方式である。この方式の副産物として、複数のマイクロ波を用い、各周波数に応答する磁性材料を層状に重ねることによって多層磁気記録(3次元磁気記録)が可能になる。この方式で複数の層にデジタルデータを記録した場合、各層の信号を別々に取り出すのではなく、同時に読み出すことによって、記録装置としての速度を高速化することが可能になると考えられる。ここに、多重アクセス通信方式の技術が適用できると考えられ、その可能性を探ることを目的とする。

DNA ストレージについて：新しい原理に基いた記録装置として、DNA を用いた記録装置の研究が活発になってきている。従来の記憶媒体の密度を大きく上回っているおり、40 万年前の生物の DNA から情報を読み取ることが出来るように、DNA は適切に保存されれば従来の記憶媒体に比べて耐久性が非常に高い。しかし、データ蓄積のために DNA 鎖を利用するには、様々な制約がある。それらの制約を満たしながら誤り訂正が可能な符号化方法を構成することを目的とする。

Lehmer 符号を用いたバランス符号：バランス制約は重要な入力制約であり、これまで非常に多くの方法が提案されてきたが、Lehmer 符号を用いた手法は近年提案されたばかりであり、その解析は重要である。本研究では、この手法の解析を目的とする。

3. 研究の方法

研究は、計算機シミュレーションと理論的な解析とを組み合わせで行った。具体的な方法は、サブテーマ毎に述べる。

4. 研究成果

4.1 二次元誤り訂正符号

Matcha らは、大きな二次元バースト消失誤りを訂正可能な LDPC 符号を提案し、そのバースト消失誤り訂正能力の下界を示している。本研究では、誤り訂正能力が向上するようにその符号を修正した。

定理 1 C を、ビット位置で考えたサイズが $cp \times hp \times wp$ の立方体を使って、ここで述べたようにしてえられるパリティ検査行列によって定まる符号とする。ただし、 $c \geq p^3 + p^2 + p$ とする。この符号は、サイズが $p \times (2p - 1)$ のバースト消失誤りを訂正できる。

この符号の符号化率は、

$$\frac{hwp - c}{hwp}$$

であり、これは上から

$$\frac{hwp - (p^3 + p^2 + p)}{hwp}$$

で抑えられる。 hw を大きくしていくと、符号化率は 1 に漸近する。しかし、 p が変化しなければ、訂正可能な誤りのサイズの上界は変化しない。

4.2 Polar 符号の性能の向上

Polar 符号は 2009 年に Arikan によって提案された符号で、通信路分極とよばれる考えに基づいている。この考え方に基いた方法で符号化を行うと、符号長を長くすることによって二元対称無記憶通信路において通信路容量を達成できることが証明されている。しかし、符号長が短い場合、これまでの復号手法では性能が低下するという問題があった。

この問題を解決するために、2019 年に Arikan によって畳み込み符号化と Polar 符号化を組み合わせた PAC(Polarization-Adjusted Convolutional) 符号が提案されている。短い符号長 $N = 128$ において Polar 符号よりも良い性能を達成できることが示されている。

PAC 符号の復号には、Fano 復号が用いられている。Polar 符号にも Fano 復号が適用されており、誤り検出を行う巡回冗長検査 (CRC : Cyclic Redundancy Check) を導入した Polar 符号では、Fano 復号に SCF(Successive Cancellation Flip) 復号の考えを取り入れた SCF-Fano 復号が提案されている。SCF-Fano 復号は CRC で誤りを検出しなくなるか、最大復号試行回数 T_{max} まで 1 ビットを反転した Fano 復号を繰り返し実行する手法である。

ここでは、PAC 符号の復号に CRC を導入し、SCF-Fano 復号を行うことで更なる改善を目指して計算機実験を行った。

4. 3 MAMAR への多重アクセス通信方式の応用

MAMR による多層化で実現される記録再生系は、通信モデルとしては、フェージングの影響がある、加法的な多重接続通信路に対応する。これは、複数の信号が各々別々に増幅されたり減衰させられたりしたあとで、アナログ信号として単に加え合わさる、という通信路である。

MAMR で多層化した場合には、層によって再生信号の強度は違ってくるものと思われる。これは、通信のモデルではフェージング係数が二つの信号に対応する通信路で異っていることを意味する。したがって、MAMR による高速な多層記録の可能性を検討するためには、フェージング係数の変化の、通信路の性能への影響を調べる必要がある。

その結果、通信の多重化の方法が適用できる可能性があることがわかった。

4. 4 DNA ストレージのための符号

DNA 鎖を用いた装置のための符号化は、化学的な理由により、いくつかの制約を満たすように行う必要がある。そうした制約を満たす符号化の方法の 1 つに、CGK 法と呼ばれる方法がある。この方法では制約を満たしながら符号化を行うことが出来るが、誤りを検出できない場合が多く、訂正は一般には不可能である。これに対して本研究では、制約を満たしながら誤りの訂正が可能な符号を提案した。

DNA ストレージでは A (アデニン), C (シトシン), G (グアニン), T (チミン) の 4 種類のシンボル (塩基) の配列によって情報が保存されるが、その物理的な特性と記録方法によって、以下の誤りが支配的である。

- シンボル置換 (塩基置換) : DNA 鎖のあるシンボルが別のシンボルへと変化してしまう誤り;
- シンボル挿入 (塩基挿入) : DNA 鎖にシンボルが挿入されてしまう誤り;
- シンボル削除 (塩基削除) : DNA 鎖のあるシンボルが削除されてしまう誤り。

二進系列を DNA 鎖へ符号化する際、塩基の種類は四つなので、数の上では 2 ビットを一つの塩基に対応させることができる。例えば、A を 00 に、C を 01 に、G を 10 に、T を 11 に変換すればよい。しかしながら、DNA ストレージでは、物理的もしくは化学的な理由により、上に述べた誤りが高い頻度で生じるような DNA 系列がある。例えば、同じ塩基が多数個連続する系列、あるいは C の個数と G の個数の差が大きい系列では、DNA ストレージの記録再生の過程において、上記の誤りが発生する確率が大きくなる。このため、1 つのシンボルに 2 ビットのデータを保存するという単純な方法は、DNA ストレージでは使用できない。

DNA ストレージでは、様々な方法でデータを DNA 鎖に変換したあと、PCR(Polymerase Chain Reaction, ポリメラーゼ連鎖反応) などを用いて記録再生を行う。

○ DNA ストレージのための有限状態符号

DNA ストレージで生じる誤りを検出・訂正するために、様々な符号化方法が提案されており、有限状態機械を使って、ブロック単位で符号化することで誤りの検出及び訂正を行いつつ、符号化率をあまり下げない方法も提案されている。本節では、この符号器の誤り検出能力及び訂正能力について考える。この符号は対象とする 3 種類の誤りを完全に検出するものではなく、一定の割合で誤りを見逃すことがある。この意味では、この符号は一般的な意味では誤り訂正符号とは言えないが、誤り訂正能力を持ちつつ、符号化率が極端に小さくなることはないという長所がある。

以下では、 \mathbb{N} は自然数の集合とする。

ブロック長 (状態を遷移するときに出る系列の長さ) を $4n, n \in \mathbb{N}$ とする。

- (A1) 4種類のシンボルをそれぞれ n 個 ($n \in \mathbb{N}$) ずつ含んだ、長さが $4n$ の系列を 1 つのブロックとする。
- (A2) 符号器に状態を持たせ、複数の符号化規則を用いて符号化を行う (変換表の作成方法は後述)。
- (A3) 1 つのブロックを出力するたびに (情報系列に対応したブロックに変換するたびに)、次の状態が決まる (有限状態符号)。
- (A4) 各記号に対して、符号化して得られた系列の各ブロックの先頭に生じる回数を数え、その回数の偶奇に関する情報を、最後に別途符号化する。

誤りの検出及び訂正を可能にするために、変換表は以下の条件を満たすように構成する。

○ 変換表が満たすべき条件

- (B1) 変換表の中の各 DNA 系列 (ブロック) は、各シンボルを同じ個数含んでいる。
- (B2) 変換表の中の各 DNA 系列は、すべての変換表の中で一回だけ表われる。
- (B3) ある状態に対応した変換表において、同じ記号が先頭にあるブロックに対しては、それらのブロックを出力したあとで到達する状態が異っている。

これらの条件は、誤り検出と訂正に利用される。

これらの条件を満たすように変換規則を構成し、これまでの研究結果と比較して十分な性能を持つとともに、符号化率を上げることができることを示した。

4. 5 Lehmer 符号を用いたバランス符号

バランス符号は、光ファイバー通信や、光、もしくは磁気記録など、様々な分野で使用されている。二元バランス符号からなる信号は、低周波成分をわずかしか含んでいない。よって、バランス符号は DC-フリー符号、もしくは零スペクトル符号とも呼ばれる。この性質は、上に述べた応用では有用である。アルファベットを $\{0, 1\}$ とした二元バランス符号は、符号語の中に含まれる 0 と 1 の数が同じである。有限状態零スペクトル符号もほとんど同じ条件を満たす。

D. Knuth は、単純であるが非常に興味深いバランス符号化を与えた。

Dubé と Mechqrane は、Lehmer 符号を利用して、Pacman algorithm と呼ばれる、非常に興味深い符号化方法を与えた。このアルゴリズムは、本質的には、長い入力と出力の対応関係に依存した、有限状態機械である。このアルゴリズムは、各出力ブロックがバランスしているので、零スペクトル符号のための必要十分条件を満たす。Dubé らは、計算機実験によって、バランス符号の理論的な限界をほぼ満す、彼らのアルゴリズムのあるバージョンを与えた。

本研究では、Pacman algorithm の中心的な要素である、二つの写像 H_1 と H_2 を与え、それらの冗長度が $(1/2) \log n$ と $\log n$ であることを示した。

一つ目の結果は、すべてのバランス系列を Pacman アルゴリズムを使用することで、数え上げ符号化を行った。その結果、冗長は $(1/2) \log n$ の符号を構成できた。

二つ目の結果として、以下の結果を示した。この冗長度は、 $\log n$ である。

定理 2 n をブロック長とし、ある正整数 k に対して、 $n = 2^k$ であると仮定する。 H_2 によって、表参照符号化法によって埋め込まれるビットの数 J は

$$J = n - \log n.$$

本研究の結果を踏まえて、符号化の複雑と冗長度のトレードオフを考慮した方法を開発できる可能性がある。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Hiroshi Kamabe, Shan Lu	4. 巻 2019
2. 論文標題 Burst Erasure Correction by 2D LDPC Codes	5. 発行年 2019年
3. 雑誌名 Proceedings of 2019 IEEE Global Communications Conference (GLOBECOM)	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/GLOBECOM38437.2019.9013593	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Hiroshi Kamabe, Takumi Marumo and Shan Lu	4. 巻 2022
2. 論文標題 Balanced coding with Lehmer code	5. 発行年 2022年
3. 雑誌名 Proceedings of ISITA(International Symposium on Information theory and its Applications)	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 Hiroshi Kamabe, Takumi Marumo, Shan Lu
2. 発表標題 Balanced Coding with Lehmer Code
3. 学会等名 第44回情報理論とその応用シンポジウム
4. 発表年 2021年

1. 発表者名 丸茂 巧, 鎌部 浩, 路サン
2. 発表標題 DNAストレージのための有限状態符号における誤りの検出及び訂正能力に関する研究
3. 学会等名 電子情報通信学会 磁気記録・情報ストレージ研究会
4. 発表年 2021年

1. 発表者名 神野寛和、鎌部浩、路サン
2. 発表標題 分割可能な構造の拡張に基づいた符号化キャッシング
3. 学会等名 電子情報通信学会情報理論研究会
4. 発表年 2021年

1. 発表者名 諫山 茉奈 , 鎌部 浩 , 路 サン
2. 発表標題 フェージングを考慮した再生系の性能評価
3. 学会等名 電子情報通信学会 マグネテックス研究会
4. 発表年 2020年

1. 発表者名 小川 響生 , 鎌部 浩 , 路 サン
2. 発表標題 PAC符号のSCF-Fano復号を用いた性能改善
3. 学会等名 電子情報通信学会情報理論研究会
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------