

令和 6 年 5 月 17 日現在

機関番号：13601

研究種目：若手研究

研究期間：2021～2023

課題番号：21K14160

研究課題名（和文）DNAストレージに対する現代的な符号理論の展開

研究課題名（英文）Applications of modern coding theory to DNA storage

研究代表者

柴田 凌（Shibata, Ryo）

信州大学・工学部・助教

研究者番号：60898401

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究では、DNAストレージでも生じることが知られる挿入、削除、置換（IDS）誤りを訂正するために、低密度パリティ検査（LDPC）符号と確率推論に基づく新しい符号化および復号法を提案した。特に、ID+雑音誤りに対する不規則LDPC符号の設計法や、スライド窓ベースの誤り訂正手法、レーストラックメモリ（RM）の位置誤差を補正する多元LDPC符号化法を開発した。これらの手法により、IDS誤り訂正符号の信頼性・効率性が向上し、実用化への一歩を進めることができた。また、遅延符号化（DC）方式により、計算コストを削減しつつ高い性能の実現が可能となった。

研究成果の学術的意義や社会的意義

本研究は、DNAストレージ技術の信頼性、効率性、安定性を向上させるための新しい基礎理論を提供するものである。学術的には、IDS誤りの数理モデルと通信路容量の計算を通じて理論的境界を明らかにし、LDPC符号と確率推論に基づく革新的な符号化および復号法を提案した。社会的には、次世代ストレージ技術（DNAストレージ、レーストラックメモリ等）の発展とデータ保存の信頼性向上に貢献するものである。これにより、重要なデータの長期保存がより安全かつ確実に行えるようになる。

研究成果の概要（英文）：In this study, we propose a new coding and decoding scheme based on Low-Density Parity-Check (LDPC) codes and probabilistic inference to correct insertion, deletion, and substitution (IDS) errors, which are known to occur in DNA storage. Specifically, we developed a design method for irregular LDPC codes for ID+noise errors, a sliding window-based error correction method, and a nonbinary LDPC coding method to correct position errors in racetrack memory (RM). These methods have improved the reliability and efficiency of IDS error correction codes, taking a step toward practical application. Additionally, the delayed coding (DC) scheme enables high performance while reducing computational costs.

研究分野：符号理論，情報理論，通信理論

キーワード：LDPC符号 空間結合符号 挿入削除誤り DNAストレージ 次世代ストレージ

## 1. 研究開始当初の背景

近年、データ生成速度の飛躍的な増加に伴い、より大容量かつ高信頼性のストレージ技術が求められている。このような背景のもと、DNA (デオキシリボ核酸) を用いたデジタルデータの記録技術、いわゆる DNA ストレージが注目を集めている。DNA ストレージは、データの高密度記録が可能であり、長期保存性に優れるため、次世代のストレージ技術として期待されている。

DNA ストレージの基本的な仕組みは、デジタルデータをアデニン (A)、グアニン (G)、シトシン (C)、チミン (T) の 4 種類の塩基に対応する形で符号化し、その塩基配列を持つ DNA を合成・複製して保存するものである。データの読み出し時には、DNA プールから DNA を抽出し、塩基配列を読み込むことでデジタルデータを再構成する。しかし、DNA の合成・複製過程、保管時、読み込み過程において、IDS (重複・損失・置換) 誤りが発生することは避けられない。

現状、DNA ストレージに対する誤り訂正技術の研究は進展しているが、IDS 誤りを全て同時に扱う方法は未だ確立されていない。従来の研究では、これらの誤りの一部 (例えば、置換誤りのみ) を対象とした訂正方法が提案されているが、DNA ストレージの実用化を目指す上で、IDS 誤り全てに対応する技術の開発は不可欠である。また、IDS 誤りが生じる複数の塩基配列から元のデジタルデータを正確に再構成するための理論的・実践的アプローチが求められている。

本研究では、シャノン限界に近い性能を達成する LDPC (Low-Density Parity-Check) 符号と確率推論に基づく復号法の枠組みを用いて、DNA ストレージで発生する IDS 誤りを効果的に訂正する方法を提案し、その性能を評価することを目指す。

## 2. 研究の目的

本研究の目的は、シャノン限界に近い性能を達成する LDPC (Low-Density Parity-Check) 符号と確率推論に基づく復号法を用いて、DNA ストレージで発生する IDS 誤りを訂正する方法を提案し、その有効性を評価することである。具体的には、以下の 3 つの研究項目を実施する。

- (1) DNA ストレージで生じる IDS 誤りの確率的な通信路モデルとその通信路容量の導出
- (2) DNA ストレージ通信路に対する符号化法と確率推論に基づく復号法の提案
- (3) DNA ストレージの効率性・信頼性・安定性の向上

## 3. 研究の方法

### (1) 確率的通信路モデルの構築：

DNA ストレージにおける IDS 誤りを数学的なモデルに落とし込み、入出力の塩基配列の関係を確率的通信路モデルとして定義する。このモデルに基づき通信路容量を計算し、IDS 誤りのもとで DNA ストレージが保証する信頼性と効率性の理論的境界を明らかにする。

### (2) 符号化法と復号法の開発：

LDPC 符号と確率推論に基づく復号法を提案し、その有効性を示す。符号化法では符号語を塩基配列に変換し、復号法では IDS 誤りが生じた複数の塩基配列から元の符号語を再構成する方法を考える。

### (3) 効率性・信頼性・安定性の向上：

効率性と信頼性を高める符号構造を提案し、その最適化手法を検討する。また、DNA ストレージの安定性を高めるための符号化・復号法を実現し、信頼性・効率性・安定性の関係を解明する。

## 4. 研究成果

### (a) ID + 雑音誤りに対する LDPC 符号設計：

ID+雑音誤りを含むチャンネルに対する不規則 LDPC 符号の設計方法を提案し、異なる雑音レベルに対して同時に次数分布を得る方法を開発した。これにより、様々な雑音レベルでの復号性能を調査することが容易になる。設計された符号は、設計された雑音レベルにおいて対称情報レート (SIR) に接近する性能を示し、異なる雑音レベルでは符号構造が異なることを確認した。また、ID-AWGN チャンネルに対するシミュレーション結果を通じて、設計された符号が優れたビット誤り率 (BER) 性能を発揮することを示した。この研究は、同期誤りを伴う雑音チャンネルに対する符号化手法の改善に貢献し、実用的なデータストレージや通信システムにおける信頼性向上に寄与する。

### (b) ID 誤りに対するスライド窓同期・復号法：

ID 誤りを伴うチャンネルに対する空間結合低密度パリティ検査 (SC-LDPC) 符号のためのスライド窓ベースの誤り訂正手法を提案した。提案手法は、低遅延のウィンドウデコード (WDD) に基づき、チャンネルトレリスと LDPC 符号グラフを部分的に結合したジョイントファクターグラフ上で反復的にデコードを行うものである。本研究では、いくつかの SC-LDPC プロトグラフの漸近復号性能を検討し、更に、有限長 SC-LDPC 符号を用いることで、低遅延かつ低遅延で大きな性能向

上が得られることを示した。特に、提案手法が従来のフラッディングベースのデコード (FDD) と比較して、ビット誤り率 (BER) の低減とデコード遅延の削減において優れた性能を発揮することを確認した。本研究は、ID 誤りを含むチャネルに対する実用的な SC-LDPC 符号化手法の開発に寄与するものである。

(c) レーストラックメモリ (RM) に対するシンボル同期と多元 LDPC 符号化法：

RM の位置誤差 (PE) を補正するために、多元 LDPC 符号とシンボル単位検出 (SW 検出) を用いた新しい符号化方式を提案している。RM においては、PE によって ID 誤りが発生する。従来の手法は、2 元 LDPC (BI-LDPC) 符号とビット単位検出 (BW 検出) を用いていたが、短いブロックサイズでの性能が不十分であった。提案手法では、RM の特殊な特性を活用し、検出器とデコーダの間に短いサイクルが生成されないようにして、有限長の性能を向上させ、収束速度を速めている。提案手法の有効性は、理論解析およびシミュレーションによって確認されており、従来の BI-LDPC 符号と BW 検出に基づく方式よりも優れた BER 性能と高速な収束を示した。また、EXIT チャート解析を用いて、提案手法に適した NB-LDPC 符号の設計も行っている。この研究は、PE による ID 誤りを補正するための RM の信頼性向上に貢献するものである。

(d) ID + 雑音誤りに対する遅延符号化に基づく符号化法：

ID+雑音誤りを伴うチャネルに対する新しい符号化方式である遅延符号化 (DC) 方式を提案した。提案方式は、遅延符号化と非反復的な検出・復号戦略を採用し、複数の符号語の送信を効率的に管理するものである。DC 方式では、チャネル入力系列が複数の符号語のサブブロックで構成され、受信側では最大事後確率 (MAP) 検出により符号語を順次推定し、事前に復号された符号語が検出を支援する。これにより、SIR に近づく性能を達成する。さらに、LDPC 符号と組み合わせることで、DC 方式の漸近的および有限長の性能が優れていることを確認した。提案方式は従来の反復検出・復号方式に比べて複雑性が低く、計算コストを削減することができる。本研究は、ID+雑音誤りを含むチャネルに対する信頼性の高い符号化方式の実現に貢献するものである。

## 5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 4件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 R. Shibata and H. Kaneko	4. 巻 Jan. 26
2. 論文標題 Shift-Interleave Coding for DNA-Based Storage: Correction of IDS Errors and Sequence Losses	5. 発行年 2024年
3. 雑誌名 arXiv [cs.IT]	6. 最初と最後の頁 1-6
掲載論文のDOI（デジタルオブジェクト識別子） 10.48550/arXiv.2401.14594	査読の有無 無
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 Shibata Ryo, Yashima Hiroyuki	4. 巻 11
2. 論文標題 Windowed-based synchronization error-correction for spatially coupled codes	5. 発行年 2022年
3. 雑誌名 IEICE Communications Express	6. 最初と最後の頁 697 ~ 702
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/comex.2022XBL0121	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Shibata Ryo, Yashima Hiroyuki	4. 巻 59
2. 論文標題 Symbol-Level Detection With Matched Non-Binary LDPC Codes for Position Errors in Racetrack Memories	5. 発行年 2023年
3. 雑誌名 IEEE Transactions on Magnetics	6. 最初と最後の頁 1 ~ 9
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TMAG.2022.3214932	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 Ryo Shibata, Hiroyuki Yashima	4. 巻 E105.A
2. 論文標題 Design and Performance of Low-Density Parity-Check Codes for Noisy Channels with Synchronization Errors	5. 発行年 2022年
3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	6. 最初と最後の頁 63 ~ 67
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.2021EAL2013	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 Jumpei Usui, Ryo Shibata, Hiroyuki Yashima	4. 巻 11
2. 論文標題 Deep-learning-aided design of LDPC coding scheme for two-user Gaussian multiple access channels	5. 発行年 2022年
3. 雑誌名 IEICE Communications Express	6. 最初と最後の頁 111 ~ 116
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/comex.2021XBL0200	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9件 (うち招待講演 1件 / うち国際学会 4件)

1. 発表者名 M. Kawasumi, H. Yashima, R. Shibata
2. 発表標題 LDPC coded DBICM system for multiple access channels
3. 学会等名 in Proc. International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP) (国際学会)
4. 発表年 2024年

1. 発表者名 S. Takie, R. Shibata, and H. Yashima
2. 発表標題 A novel energy efficient clustering using a sleep awake algorithm for energy harvesting wireless sensor networks
3. 学会等名 in Proc. International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP) (国際学会)
4. 発表年 2023年

1. 発表者名 川隅匡登, 柴田 凌, 八嶋弘幸
2. 発表標題 マルチアクセス通信路に対するLDPC符号化DBICMシステムの設計と解析
3. 学会等名 第46回情報理論とその応用シンポジウム
4. 発表年 2023年

1. 発表者名 R. Shibata
2. 発表標題 Improved HICR scheme for noisy insertion and deletion channels: dynamic scheduling and unequal error protection
3. 学会等名 The 46th Symposium on Information Theory and its Applications
4. 発表年 2023年

1. 発表者名 川隅匡登, 柴田 凌, 八嶋弘幸
2. 発表標題 マルチアクセス通信路に対するLDPC符号化DBICMシステム
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

1. 発表者名 Shibata Ryo
2. 発表標題 Modern Channel Coding for Synchronization Errors
3. 学会等名 2022 International Conference on Emerging Technologies for Communications (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 Shibata Ryo, Yashima Hiroyuki
2. 発表標題 Delayed Coding Scheme for Channels with Insertion, Deletion, and Substitution Errors
3. 学会等名 2022 IEEE Global Communications Conference (国際学会)
4. 発表年 2022年

1. 発表者名 樋口 智大, 柴田 凌, 八嶋 弘幸
2. 発表標題 非正則Multi-Kernel Polar符号の提案と性能調査
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 平川竜之, 柴田 凌, 八嶋 弘幸
2. 発表標題 FSK信号を入出力としたAND/OR切替可能な全光論理回路
3. 学会等名 2022年電子情報通信学会総合大会
4. 発表年 2022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