

論文の内容の要旨

論文題目

Efficient Polar and LDPC Coding for

Asymmetric Channels and Sources

(非対称な通信路と情報源の効率的な

polar 符号化および LDPC 符号化)

氏名 本多 淳也

情報理論における最も基本的な問題として、ノイズを含む通信路を通して正しく情報を伝送する通信路符号化と、与えられたデータをより短いデータ系列に変換する情報源符号化がある。また、情報源符号化はデータの損失が許されるかどうかで無歪み符号化（無歪み圧縮）と有歪み符号化（有歪み圧縮）に分類される。ここで無歪み圧縮については理論限界を達成可能かつ高速な符号化方式が早くから知られていたが、通信路符号化および有歪み圧縮において理論限界に近い性能をもつ実用的な符号化方式の構成は現在においても重要な研究課題となっている。

本論文ではこれら通信路符号化および有歪み圧縮について考える。これらの問題においては、対称モデル、すなわち対称通信路の符号化および一様情報源の対称な歪み基準における符号化については多くの研究がなされており、polar 符号や LDPC 符号といった線形符号により有限の符号長で理論限界に近い性能を達成できることが知られている。一方、現実には用いられる通信路や情報源は必ずしも対称とは限らないが、非対称モデルに対する研究はそれほど多くない。対称でない通信路や情報源の符号化においては、理論限界を達成するために符号語の各シンボルが従うべき理想的な頻度分布が一様分布とならない。一般に線形符号においては各シンボルが一様に出現するため、非対称モデルに対して素朴に線形符号を適用した場合には理論限界を達成することができない。このような非対称モデルの符号化を線形符号の枠組みで行う方式としては Gallager の多元符号を用いる方法が主に知られているが、これは理論限界を漸近的には達成可能であるものの、現実的には計算量等の問題からあまり実用的ではない。

そこで本論文では非対称モデルにおける通信路符号化と有歪み圧縮について、polar 符号または LDPC 符号を用いた新たな符号化方式を提案し、理論限界を漸近的に達成可能であることを示

した. 提案方式では一様でない符号語を生成する際に, **Gallager** の方法のかわりに無歪み圧縮を応用した方法を用いる. これは, 無歪み圧縮が一様でない生起確率をもつ系列を一様な系列へと変換する操作だとみなせることから無歪み圧縮の逆操作を用いれば一様でない符号語を生成できるはずというアイディアに基づく. 同じアイディアを用いた符号は **Miyake-Muramatsu** によって提案されているが, 彼らの方法が NP 完全問題の厳密計算を前提としているのに対し, 提案方式では有限の符号長で理論限界に近い性能を現実的な計算量で達成できる.

まず, 第 2 章では通信路符号化および有歪み圧縮の理論限界を示すとともに, 非対称モデルの符号化問題を対称モデルの符号化に帰着させる従来手法として **Gallager** の方法を紹介した. また, 無歪み圧縮において知られている効率的な符号化法についても併せて紹介した.

第 3 章では **polar** 符号を用いた通信路符号と有歪み符号の提案を行った. **Polar** 符号は通信路分極とよばれる現象を用いた符号化方式として 2008 年に **Arikan** によって提案された. これは多項式時間で 2 値対称通信路の通信路容量を漸近的に達成できることが理論的に示された最初の符号として近年盛んに研究がなされており, その漸近最適性は後に一般の通信路の符号化および有歪み圧縮に対しても拡張されている. ここで非対称モデルにおける通信路符号化および有歪み圧縮については **Gallager** の方法により最適性が示されているが, これは大きなアルファベットサイズの多元 **polar** 符号を必要とするために, 計算量が非常に大きくなる問題があった. そこで本論文では **polar** 符号による無歪み圧縮を応用した符号化方式を新たに提案し, その漸近最適性を示した.

Polar 符号による無歪み圧縮では, 一様でない頻度分布をもつ n ビット情報源系列 $X = (X_1, X_2, \dots, X_n)$ を **polar** 符号の生成行列 G によって $U = GX$ と可逆変換する. ここで生成行列 G の構造から, 変換後の系列 $U = (U_1, U_2, \dots, U_n)$ の各シンボル U_i は (A) 以前のビット U_1, \dots, U_{i-1} を見れば U_i の情報が復元できる, (B) 以前のビット U_1, \dots, U_{i-1} を見たうえでも U_i は一様ランダムである (すなわち U_i の情報は全く分からない), という 2 パターンのみに分かれることが示される. したがって, グループ (B) に属するビットについてのみ情報を送信すれば U 全体が復元でき, そこから G の逆変換により X が復元できる. すなわち, グループ (A) に属するビットの数だけ系列を圧縮できることになる.

本論文ではこの **polar** 符号による無歪み圧縮を用いて通信路符号を構成した. 提案手法においては, 一様系列をグループ (B) に属するビットに割り当てることにより系列 U を生成する. **Polar** 符号の性質から, グループ (A) に属するビットの値を適切に決めることにより, 理想的なシンボル分布をもつ符号語 $X = UG^{-1}$ を構成できる. ここで通信路符号化において通信路入力 X に対する出力 Y を考えた場合, グループ (B) に属するビット U_i はさらに (B1) 以前のビット U_1, \dots, U_{i-1} を見たうえでも U_i は一様ランダムだが, 通信路出力 Y を併せて見れば U_i の情報を復元できる, (B2) 以前のビット U_1, \dots, U_{i-1} と通信路出力 Y を見たうえでも U_i は一様ランダム, という 2 パターンに分極する. ここでグループ (B1) に属するビットに一様なメッセージを割り当てることにより, 入力系列 X の理想的な頻度分布を崩さず, かつ通信路出力からメッセージが復元することが可能な通信路符号を構成できることを本論文では示した. また, 有歪み圧縮の場合にも通信路符号の符号化と復号を入れ替えることにより漸近最適な符号を構成することができることを併せて示した.

第 4 章では **low density parity check (LDPC)** 符号を用いた通信路符号と有歪み符号の提案を

行った。LDPC 符号は疎なパリティ検査行列を用いて定義される線形符号であり、belief propagation (BP) と呼ばれる復号法により符号長に対して線形時間で理論限界に近い性能を達成できることが実験的に知られている。

LDPC 符号が通信路容量を達成可能であることは理論的にも証明がなされ、またその最適性は有歪み圧縮にも拡張されている。ただし、LDPC 符号の最適性においては NP 完全な復号問題を解くことを仮定しており、多項式時間で最適性を保証できるような復号アルゴリズムは一般には知られていない。ここで対称通信路については BP 復号を用いることにより polar 符号に比べて実験的に良い性能を達成できることが知られているが、非対称モデルの場合には Gallager の方法と Miyake-Muramatsu の方法いずれを用いた場合でも多項式時間アルゴリズムの性能が著しく悪化する問題があった。

そこで本論文では無歪み圧縮を用いるアイデアを Miyake-Muramatsu とは異なる形で適用した符号化方式を提案した。Miyake-Muramatsu の方式は LDPC 符号による固定長符号を無歪み圧縮に用いるものであるが、固定長無歪み圧縮は有限の符号長では元の系列を正しく復号できない場合があり、さらに NP 完全な復号問題を厳密に解く必要がある。そこで提案手法では可変長符号である算術符号または homophonic 符号を無歪み圧縮に用いる。これらの可変長符号の適用にあたっては符号語中の各シンボルの生起確率を計算する必要があるが、これは BP を用いることで効率的に計算できる。さらに、計算した確率が近似値であった場合も可変長符号では符号化率がわずかに悪化するのみで、即座に復号失敗とはならないため近似アルゴリズムとの相性が良い。まず第 4 章では、符号化・復号における計算が厳密にできた場合に理論限界を達成可能であることを証明した。

次に第 5 章では、第 4 章で提案した LDPC 符号による通信路符号および有歪み符号に対する実用的なアルゴリズムの提案を行い、シミュレーションによりそれらの性能を確認した。提案方式における通信路符号の復号および有歪み符号の符号化における主要な計算は、与えられた系列に対して事後確率が最大となる符号語を求める符号語推定問題と、符号語の各シンボルの周辺確率の計算の 2 つからなる。まず 5.2 節では BP を用いて周辺確率の近似計算を行う効率的なアルゴリズムの提案を行い、厳密計算を行った場合とほぼ同等の符号化率が達成できることを確認した。次に 5.3 節と 5.4 節では符号語推定の問題に対する近似アルゴリズムの提案を行った。この問題は通信路符号化と有歪み圧縮で共通の形の最適化問題となるが、近似アルゴリズムに求められる性質は両者で大きく異なる。まず 5.3 節では有歪み圧縮における符号語推定の問題に対して BP の改良版である RBP (Reinforced Belief Propagation) とマトロイド最適化を組み合わせたアルゴリズムを提案した。次に 5.4 節では通信路符号化における符号語推定の問題を線形計画法を用いて解く手法 (LP 復号) を紹介し、また多元 LDPC 符号に対して LP 復号を高速に行う手法を新たに提案した。

最後に第 6 章では本論文の成果を総括し、また Gallager の方法と無歪み圧縮による方法の誤り指数の比較といった問題を今後の課題として述べた。