

# 1. 「基礎・境界」が成し遂げたこと、今後に期待できること

## 1-5 符号理論のストレージ技術への応用とその将来

### Application of Information Theory and Coding Theory to Data Storage

鎌部 浩

#### 1. ま え が き

Shannon は 1948 年に発表した「A Mathematical Theory of Communication」<sup>(1)</sup>において、デジタル情報伝送の数学的なモデルを与えた上で、今日のデジタル通信の基礎となる定理を証明している。「情報源符号化定理」は、データを圧縮するときの圧縮率の限界を与えており、「通信路符号化定理」は誤りを幾らでも小さくしながら通信できる伝送率の限界を示している。これらの限界を達成するような符号化方法の発見は、その後の研究の大きな目的の一つとなった。その後の展開や今後の展望などについては、様々な文献で紹介されている。(例えば、本特集の植松氏の記事や文献(2)など。)

Shannon はこの論文<sup>(1)</sup>において、入力できる記号系列に制限があるような通信路を使ってデータを送信するとき、達成可能な伝送率の上界も与えている。

原理的には通信路に入力できる系列を数え上げることで、この上界に幾らでも近い伝送率の符号を構成することは可能である(例えば文献(3)6章)。しかし、実際の通信路が持っている入力制約が与えられたとき、データ、つまり、任意の記号列をその制約を満たすような記号列に変換し、誤りなく元のデータを復号できるような符号化規則を、実用的な規模の電子回路で実現するという問題は、一般には簡単な問題ではない。

通信路符号化定理の場合と同様に、入力制約を持つ通信路に関しても、定理が保証する限界を達成するために、多くの研究が行われてきた。これらの研究によって、理論的な限界に近い符号の構成方法が見いだされ、

実際の伝送路で使用された。本稿では、そうした研究の流れについて述べる。

後半では、今後発展すると思われる二つの研究分野について述べる。デジタルデータの需給バランスが将来崩れる(データの生成量の増加に対して、それを蓄積する装置の全容量が大きく不足する)のではないかと予想されているが<sup>(4)</sup>、このアンバランスを解消する一つの方法として期待されている DNA ストレージについて、その誤り訂正方法の一つと、これまでの符号構成方法との関係(本稿前半の内容との関係)について述べる。

近年、フラッシュメモリなど不揮発性固体メモリのための符号が盛んに研究されているが、その中でも、フラッシュメモリのための WOM 符号は最もよく研究されている。ここでは、最も実用化に近いと思われる WOM 符号の応用について述べる。

#### 2. 入力制約を持つ通信路

##### 2.1 系統的な符号の構成方法

Shannon は、入力制約の例としてモールス信号を考えている<sup>(1)</sup>。モールス信号では、2種類の「スペース」、1記号分のスペースとその2倍の長さのスペースがある。一つのスペースが二つ連続すると2倍の長さのスペースと区別がつかなくなるので、これらのスペースは連続してはならない、という制約がある。モールス信号の記号列は、どのようなデータを符号化したとしても、符号化された系列(送信される系列)がこの制約を満たすように作られている。

ここでは、制約として  $(d, k)$  制約を考える。ある記号  $a$  が  $n$  回連続するとき、 $a^n$  のように書き、これを「 $a$  の連」と呼ぶ。 $A$  を文字の有限集合(有限アルファベット)としたとき、 $A^*$  は、 $A$  の要素から成る有限長の全ての系列(空系列も含む)の集合とする。

鎌部 浩 正員：シニア会員 岐阜大学工学部電気電子・情報工学科  
E-mail kamabe@m.ieice.org  
Hiroshi KAMABE, Senior Member (Faculty of Engineering, Gifu University, Gifu-shi, 501-1193 Japan).  
電子情報通信学会誌 Vol.100 No.6 pp.424-430 2017年6月  
©電子情報通信学会 2017

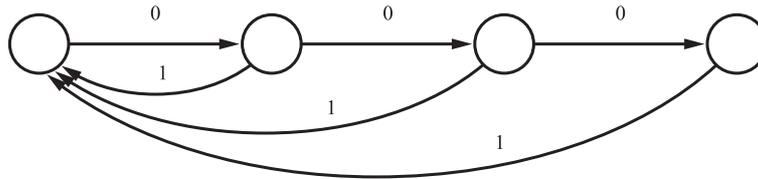


図1 (1,3) 制約を表現するグラフ

[定義1]  $d$  を非負整数,  $k$  を非負整数若しくは  $\infty$  とし,  $d < k$  であるとする. 系列  $\alpha \in \{0, 1\}^*$  に含まれる記号 1 とその次の 1 の間にある 0 の連の長さが必ず  $d$  以上であり,  $\alpha$  に含まれる 0 の連長は全て  $k$  以下であるとき,  $\alpha$  は  $(d, k)$  制約を満たすと言う. この制約は図1のようなラベル付き有限有向グラフ (有限オートマトン) で表現できる.

ある制約を満たすような系列しか入力できない, 若しくは伝送できない通信路を入力制約を持つ通信路という.

磁気記録装置は, 記録するデータを入力と考え読出し信号を出力と考えると, 一種の通信路とみなすことができる. ハードディスクが長手記録方式 (磁化方向がディスク回転する方向と同じ方式. 現在のハードディスクは垂直記録方式を採用している.) を採用していたときには, 磁気記録通信路は, 入力制約として  $(d, k)$  制約を持っていると考えることができた.  $d$  制約と  $k$  制約の物理的な意味は次のとおりである.

$d$  制約: 高い記録密度を達成するために, 物理的な 1 bit を記録するための領域は狭くなっていった. しかし, 一つの磁化反転によって出てくるパルス状の信号の時間的な幅は変化しないので, その幅が物理的な 1 bit の幅よりも相対的に長くなっていった. 隣接する磁化反転はその極性が逆なので, これにより一つの磁化反転が隣接する領域に干渉することになり, 結果的に出力レベルが下がることになる. このような記号間干渉の発生を防ぐために, 磁化反転の間隔がある程度以上長くなることが要請されていた.

$k$  制約: データを再生するときには, クロック (記号周波数) を正確に再生することが重要である. この周波数が不正確であれば, ビットが抜けたり, 意味のないビットが挿入されたりする. 磁気記録では, 磁化反転がないと再生信号のレベルは一定になってしまう. したがって, ハードディスクでは, データを記録している部分でも, クロックに関する情報が取り出せるように, ある程度の間隔 (電子回路が同期を外さない程度の間隔) で磁化反転が生じるようにという要請があった. 物理的な記録方式は異なるが, CD や DVD でも同様の要請がある.

さて, データは一般には任意の系列であるので, 入力制約を満たしているとは限らない. したがって, 入力制約を持つ通信路を介してデータを伝送するときには, まずデータを, 入力制約を満たす系列に変換しなければならない. また, 受信側若しくは再生時には変換された系列から, 若干の誤りが含まれているなどの現実的な仮定の下で元のデータ系列をほぼ正確に復号できなければならない. (通常のシステムでは, この復号を行った後で誤り訂正符号の復号を行う.)

このような変換を行う符号の符号化率は 1 より小さくなければならないが, では, どこまで大きくできるだろうか. それは, 次に定義する制約の容量によって定まる.

[定義2] 制約の容量  $C(S)$  制約  $S$  に対して  $C(S)$  を以下のように定義し, これを  $S$  の容量と呼ぶ.

$$C(S) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log_2 N_S(n)$$

ここで,  $N_S(n)$  は, 制約  $S$  を満たす長さが  $n$  の系列の数である.

$S$  が有限オートマトンで表現できるのであれば, この値は

$$C(S) = \log_2 \lambda(S)$$

で与えられる. ここで  $\lambda(S)$  は, オートマトンを表現する有限有向グラフの遷移行列の最大固有値である.

Shannon は以下の定理を証明している<sup>(1)</sup>.

[定理3]  $R < C(S)$  であれば, 符号化率が  $R$  以下であるような, 制約  $S$  のための符号を構成できる. また,  $S$  のための符号があれば, その符号化率は  $C(S)$  以下でなければならない.

この事実は  $N_S(n)$  が  $N_S(n) \approx 2^{nC(S)}$  と近似できることに対応する. ところで,  $C(S)$  の定義とこの定理の証明に従って, 符号化率が  $C(S)$  に近い符号を, 表引き, 若しくは論理回路で構成しようとする, 一般には符号長をかなり長くしなければならない. 符号長が長くなると

表も論理回路も非常に大きくなる上に、制約によっては収束の速度が非常に遅いこともある（例えば、以下のバランス制約を参照）。したがって、定理の証明の方法をそのまま実際の通信路に応用することは多くの場合に困難である。

IBM ワトソン研究所の Franaszek は、この問題に取り組んでおり、制約が図 1 のように有限有向グラフで表現できるとき、そのグラフを頼りに符号を構成する方法を提案していた。非常に巧妙な方法であり、多くの場合に良い結果を与える方法であったが、どのような制約に対して適用できるのか、その方法を適用可能な制約の種類は明確ではなかった。

この符号構成の問題は、Adler, Coppersmith と Hassner によって、あるクラスの制約に対しては厳密に解かれている<sup>(5)</sup>。

[定理 4] 有限形（後述）の制約  $S$  が与えられたとする。 $m, n$  を正整数としたとき  $n/m \leq C(S)$  であれば、符号化率が  $n/m$  であるような有限状態符号器が存在する。そして、その復号器はブロック写像（以下に定義する）として与えられる。

この定理は、記号力学系（数学の一分野）に関する結果を応用して得られている。

この定理の証明では、構成的に符号器を与えられており、それに従って設計した符号器が実用化された<sup>(5)-(7)</sup>。

[定義 5]  $\mathcal{A}$  を、有限集合とする（以下では、アルファベットと呼ぶ）。 $\mathcal{A}^*$  を、 $\mathcal{A}$  の要素から成る、全ての有限系列の集合とする。空系列も含むとする。 $\mathcal{F} \subseteq \mathcal{A}^*$  を、ある有限系列の集合としたとき、 $S_{\mathcal{F}}$  を以下のように定義する。

$$S_{\mathcal{F}} = \{x \in \mathcal{A}^{\infty} : x \text{ には、} \mathcal{F} \text{ のどの要素も生じない}\}$$

これを記号力学系と呼ぶ。 $\mathcal{F}$  が有限集合であるとき、 $S_{\mathcal{F}}$  は有限形であるという（条件については例えば文献(8)）。有限形力学系は、ある性質を持ったラベル付き有限有向グラフ  $G$  によって表現できる。その性質を持った  $G$  によって定まる力学系を、 $S(G)$  と書く。 $\mathcal{F} = \emptyset$  のとき（何も制約がないとき）、フルシフトと呼ばれる。

[定義 6] ブロック写像

ある力学系  $X$  から別の力学系  $Y$  への写像  $\phi: X \rightarrow Y$  が、次の性質を持つとき、 $\phi$  はブロック写像であるという。ある写像  $f: \mathcal{X}^{j+k+1} \rightarrow \mathcal{Y}$  が存在して、全ての  $i \in \mathbb{Z}$  と  $x \in X$  に対して

$$(\phi(x))_i = f(x_{i-j} \cdots x_i \cdots x_{i+k})$$

が成立する。ここで  $x_i$  は、 $x$  の  $i$  番目の要素を意味している。 $\mathcal{X}$  と  $\mathcal{Y}$  は、それぞれ  $X$  と  $Y$  の記号アルファベットである。

$\phi$  がブロック写像であるということは、 $\phi(x)_i$  は、 $x_i$  を含む、長さが  $j+k+1$  の周辺のパターンから定まり、その規則は場所（添字  $i$ ）には依存しない、ということの意味している。

[定義 7] 力学系  $S$  に対して、 $S = \phi(S_{\mathcal{F}})$  となるような有限集合  $\mathcal{F}$  とブロック写像  $\phi$  があるとき、 $S$  はソフィックであるという。

符号器が有限状態符号であるとき、符号化規則が 1 対 1 でありさえすれば、原理的には、復号規則を定めることができる。しかしながら、ある時点で誤りが入ってくると、それ以降の復号に影響する可能性がある。これは誤り伝搬と呼ばれ、記録符号を構成するときには、一つの誤りが無限に伝搬することがないように構成することが重要であった。

復号規則がブロック写像で与えられれば、誤り伝搬は必ず有限であることが保証できるので、定理 4 によって構成された符号は応用において重要な意味がある。

定理 4 の証明は記号力学系の成果を利用しているが、実はその記号力学系は情報理論と深い関係がある。この点について、次に述べる。

## 2.2 情報理論と記号力学系

Shannon は情報源のエントロピーを定義したが、これは、コロモゴロフによって、測度論的力学系のエントロピーに拡張された<sup>(9)</sup>。二つの測度論的力学系が同形であるかどうかを判定する問題は、この分野の基本的な問題である。エントロピーは、同形性の不変量であることが示され、同形性の判定問題に大きく貢献した。実際、ベルヌーイ形の測度論的力学系に対しては、エントロピーが完全な不変量である、つまり、エントロピーが等しいことと同形であることは同値であることが示されている<sup>(10)</sup>。

このように、情報理論で定義されたエントロピーは、力学系の分野に大きな影響を与えた。逆方向の影響としては、前述の符号構成法がある。この構成方法のアイデアは、整数  $N$  に対してエントロピーが  $\log N$  となるような力学系と、記号数が  $N$  のフルシフトとの間の埋込の問題を解くための証明方法にある。

ある正整数  $N$  に対して、ある有限形の力学系  $S$  のエントロピーが  $\log N$  であったとする。有限形の力学系は、有限有向グラフで表現できるが、Marcus は、エン

トロピーが  $\log N$  であれば、全ての頂点の出次数（頂点から出ていく枝の本数）が  $N$  であるようなグラフで定義される力学系と同形であることを証明した。そのグラフの各頂点から出ている  $N$  個の枝に別々の記号を割り当てることによって、埋込を定義できる<sup>(11)</sup>。その証明では、グラフの状態遷移行列の固有ベクトルを頼りにグラフの変形を行う。これは、状態分割法と呼ばれる。Adler らは、これをエントロピーが  $\log N$  以上の場合に拡張して符号を構成する方法を示した<sup>(5)</sup>。

### 2.3 周波数スペクトルに関する制約

前述の符号構成法の応用例について述べる。

ベースバンド伝送において、AC 結合を使った電子回路で信号処理をする場合には、信号に含まれる直流成分が問題を引き起こす（例えば CD, DVD やイーサネットなど）。そこで、伝送路に信号を送信する直前に、信号が低周波数成分を持たないように符号化することがある。

このような符号を設計するためには、信号が低周波成分を持たない、という条件を明確にする必要がある。1 と -1 を記号とするブロック符号を考えた場合には、各符号語の記号の和が 0 であることが、信号の平均値が 0 であるための十分条件であることはすぐに分かる。実はこれが、周波数スペクトルの離散成分と連続成分が直流において 0 になるための必要十分条件であることが安田らによって示されている<sup>(12), (13)</sup>。吉田らは、この結果を拡張して、記号周波数の有理数倍の周波数において、周波数スペクトルが 0 になるための必要十分条件を与えている<sup>(14)</sup>。これらの結果は、指定された周波成分を含まない信号を生成するための基礎理論となった。ところで、安田らの結果<sup>(12), (13)</sup>は Pierobon によって<sup>(15)</sup>、吉田らの結果<sup>(14)</sup>は、Marcus と Siegel によって再発見されている<sup>(16)</sup>。

CD と DVD では、EMF, EMF+ という符号が使用されている。これは、四つの状態を持つ有限状態符号器である。グラフは、上記の文献の条件を満たすように作成され、ACH の状態分割法を用いて、窓サイズが 2 のスライディングブロック符号器によって復号できるように構成されている（文献(3)11章）。

図 2 は、低周波数成分を持たない信号を生成するグラフであり、有限形力学系ではないソフィックになる。ソフィックに対しては、状態分割法を直接には適用することはできない。ソフィックを内側からエントロピーの意

味で近似することができるので、そのようにして得られた有限形の力学系に対して状態分割法を適用することで、有限状態符号器を構成できる<sup>(17)</sup>。

## 3. DNA メモリ

前章では、有限状態符号の構成法とその応用について述べてきた。ここでは、次世代の応用として、DNA 系列を使った記録装置（以下、DNA メモリと呼ぶ）について述べる。

DNA メモリは、僅か 1 グラムの DNA に 1 ZByte ( $10^{21}$  Byte,  $10^9$  TByte) のデータを記録することが可能であると考えられており、デジタル記録容量不足問題に対する、一つの答えになるのではないかと期待されている<sup>(18)</sup>。

また、数千年前に氷漬けになった DNA を現代の技術で解析できることが示しているように、DNA メモリの寿命は非常に長いと期待されている。

DNA メモリは大きな期待を受けてはいるが、ランダムアクセスの困難さ、記録と読出しのスピードの問題など、実用化のために解決すべき問題は多い<sup>(19)</sup>。また、DNA の操作の信頼性が余り高くないことと、その誤りの発生の仕方が、これまでの記録装置と全く違っているため、誤り訂正技術も一から構築し直さなければならないという問題もある。

以下では、二重化誤りという DNA メモリ特有の誤りに注目し、そのための誤り訂正符号が、 $(d, k)$  制約を用いて表現できることを示す。

### 3.1 二重化誤り

$\epsilon$  は長さが 0 の系列（空列）を意味し、任意の系列  $\mathbf{y}$  に対して、 $|\mathbf{x}|$  はその長さを意味する。系列  $\mathbf{x} = x_1 x_2 \cdots x_n$  に対して、長さ  $t$  の接頭辞  $\text{Pref}_t$  と長さ  $t$  の接尾辞  $\text{Suff}_t$  は以下のように定義される。

$$\text{Pref}_t(\mathbf{x}) = x_1 x_2 \cdots x_t, \quad \text{Suff}_t(\mathbf{x}) = x_{n-t+1} x_{n-t+2} \cdots x_n$$

系列  $\mathbf{x} = x_1 x_2 \cdots x_n$  が  $\mathbf{x} = \mathbf{a} \cdot \mathbf{b} \cdot \mathbf{c}$  と分解できるとき、系列  $\mathbf{y} = \mathbf{a} \cdot \mathbf{b} \cdot \mathbf{b} \cdot \mathbf{c}$  は、 $\mathbf{x}$  から  $\mathbf{b}$  の二重化誤りによって得られる、若しくは、 $\mathbf{x}$  から長さ  $k$  の二重化誤りによって得られる、という。ここで  $k = |\mathbf{b}|$  である。

二重化誤りの関数を次のように定義する： $i$  と  $k$  を正

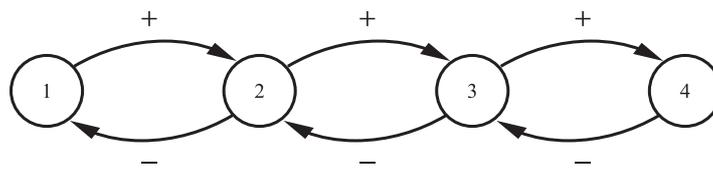


図 2 低周波成分を持たない信号を生成するグラフ

整数とする.  $T_{i,k}: \Sigma^* \rightarrow \Sigma^*$  は

$$T_{i,k}(\mathbf{x}) = \begin{cases} uvvw & \mathbf{x} = uvw, |u|=i \text{ かつ } |v|=k \text{ のとき} \\ \mathbf{x} & \text{その他の場合} \end{cases}$$

DNA ストレージのための誤り訂正符号を考えるために, 誤り訂正符号を以下のように定義する.

[定義 8] 長さ  $k$  の二重化誤りのための  $(n, M; t)_k$  符号  $C$  とは, 以下の条件を満たす符号である.

1.  $C \in \Sigma^n$ ,
2.  $|C| = M$ ,
3. 全ての  $\mathbf{x}, \mathbf{y} \in C$  に対して  $\mathbf{x}$  と  $\mathbf{y}$  に, 長さが  $k$  の二重化誤りをどのように  $t$  回適用したとしても, 同じ系列になることはない.

もし  $C$  が, 任意の  $t$  に対して  $(n, M; t)_k$  符号であるならば,  $(n, M; *)_k$  符号であるという.

### 3.2 $(n, M; t)$ 符号と連長制約

次の写像を定義する.

[定義 9]  $\phi_k: \mathbb{Z}_q^{\geq k} \rightarrow \mathbb{Z}_q^k$  を, 次式で定義する.

$$\phi_k(\mathbf{x}) = (\text{Pref}_k(\mathbf{x}), \text{Suff}_{|\mathbf{x}|-k}(\mathbf{x}) - \text{Pref}_{|\mathbf{x}|-k}(\mathbf{x})).$$

ここで,  $\mathbb{Z}_q^{\geq k} = \bigcup_{l=k}^{\infty} \mathbb{Z}_q^l$  である.

詳細は省略するが,  $\phi$  によって, 二重化誤りと 0 の連長との間に, 一対一の対応を付けることができる.

Jain らは, 以下の定理 10 を証明している<sup>(20)</sup>. まず, 符号  $C_1$  を次のように構成する.

符号の構成 A: 正整数  $n \geq k \geq 1$  に対して,

$$C_1 = \bigcup_{i=0}^{\lfloor n/k \rfloor - 1} \{ \phi_k^{-1}(y, z0^k) | \phi_k^{-1}(y, z) \in \text{Irr}_k(n-k) \}.$$

このとき, 次の定理が成立する.

[定理 10] 文献(21)上で定義した符号  $C_1$  は,  $(n, M; *)_k$  符号である. ここで

$$M = \sum_{i=0}^{\lfloor n/k \rfloor - 1} q^k M_{\text{RLL}_q(0, k-1)}(n - (i+1)k)$$

である. ただし,  $M_{\text{RLL}_q(0, k-1)}(m)$  は,  $\mathbb{Z}_q$  上の  $(0, k-1)$ -制約を満たす, 長さが  $m$  の系列の数である.

DNA 系列を工学的に利用するための符号理論の研究が進められているが, この結果のようにこれまでの  $(d, k)$  制約などに帰着できるものもあれば, 全く新しい体系を必要とするものもある<sup>(22)</sup>.

## 4. フラッシュメモリと WOM 符号

### 4.1 WOM 符号

Rivest と Shamir は, 書換えができない記録媒体を使って, 複数回の記録を行うための符号 (WOM 符号) を提案した<sup>(23)</sup>. フラッシュメモリは, 書換え回数が有限な記録媒体であるため, WOM 符号を, フラッシュメモリにうまく適用できる. WOM 符号については様々な研究が行われているが, 実際に適用するためには, 現在の方式を大きく変更する必要があるため, 実用化には至っていない.

Sharon らは複数の電荷レベルを認識可能なフラッシュメモリ (MLC や TLC) の信号処理に WOM 符号の考え方を適用することで, セルの電圧を効率的に読み出せることを示した<sup>(24)</sup>. ここでは, その考え方について述べる. この方法は, 現在のフラッシュメモリへの書込み規則を若干変更するだけで実装でき, メリットも大きい. WOM 符号が実用化される最初の例になる可能性が高いと思われる<sup>(25)</sup>.

### 4.2 ランダム入出力符号

TLC (Triple Level Cell) では, 一つのセルに 3 ビットのデータ,  $b_0 b_1 b_2$ ,  $b_i \in \{0, 1\}$  を記録することができる. 各ビットは独立したページのデータとして扱われる. つまり,  $b_i$  はページ  $i$  のデータとなり, 各ページの内容は独立にアクセスされるとする.

電圧レベルとビットパターンと対応表は, 表 1 に与えられている. 電圧が, 参照電圧と比較して高いか低いかを確認することでビットパターンを読み出す. 具体的には, ページ 0 の情報を読み出すには電圧 4 より高いかどうかを確認すればよい. ページ 1 の情報を読み出すときには, 電圧が 2 と 5 の間にあるかどうかを確認する, などである. このようにすると, ページ 1 の情報を読み出すためには 2 回の電圧の比較が, ページ 2 の場合には 4 回の比較が必要になる. これが MLC の速度を遅くして

表 1 TLC のレベルとパターンの対応

レベル	0	1	2	3	4	5	6	7
ページ 2	1	0	0	1	1	0	0	1
ページ 1	1	1	0	0	0	0	1	1
ページ 0	1	1	1	1	0	0	0	0

表2 WOM符号の例

データ	1回目	2回目
00	000	111
01	100	011
10	010	101
11	001	110

表3 WOM符号を利用した符号

2回目の情報	1回目の情報			
	00	01	10	11
00	111	211	121	112
01	011	200	021	012
10	101	201	020	102
11	110	210	120	002

いる原因の一つになっている。

表2に、RivestとShamirが示した、0から1への変化が1回しかできない三つのセルを使って、2ビットの情報を2回書くためのWOM符号の符号化規則を示している。この表に基づいて、0,1,2と三つの電圧レベルを区別できるセルを三つ使って、2ビットの情報を二つ書くことを考える。一つ目の情報は、WOM符号の1回目の規則で、二つ目の情報は、WOM符号の2回目の規則で求め、それらを足し算した値をセルに書く。ただし、一つ目と二つ目の情報が同じ場合には、一つ目の情報を2倍した値を書くとする。例えば、一つ目の情報が01で、二つ目の情報が10であれば、 $100+101=201$ が書き込まれる。この規則を表にしたのが、表3である。

セルに注入する電荷を計算するときに、 $1+0$ という計算はないので、以下の規則で復号できる。

一つ目の情報の復号:2を1に、1を0に変換して、WOM符号の規則で復号する。

二つ目の情報の復号:2を1に変換して、WOM符号の規則から復号する。WOM符号の2回目の符号語でない場合には、1回目の符号化の規則から探す。

一般に $t$ 回書き込み可能なWOM符号を使用したときには、各セルについて1回の電圧比較を行うだけで、 $j$ 番目のページの情報を読み出すことができる。

Yaakobiらはこのアイデアを発展させて、より符号化率の高い符号化方法を提案している<sup>(25)</sup>。Sekiyaらは、これと同じアイデアを多重アクセス通信路に対して適用している<sup>(26)</sup>。

## 5. おわりに

シャノンの論文が発表されて以降、二元対称通信路やガウス通信路など、一般的な通信路に対する誤り訂正符号の分野は大きく発展し、また、通信以外の様々な分野にも応用されるようになってきている。一方で、技術の進歩に伴って様々な性質を持った新しい伝送路が今後も出現すると予想される。そうした伝送路を有効に使用するために、伝送路の性質に合わせた符号を構成するための理論、特に有限状態符号の理論は、今後も発展していくものと思われる。

本稿中に述べたように、記録装置の総容量は今のままでは大きく不足するので、高密度化と大容量化は喫緊の課題である。このため、近い将来、二次元記録や三次元記録が実用化されていくと思われる。しかしながら、二次元以上の制約については、その容量(エントロピー)を計算する一般的な方法はいまだに知られていない。情報理論、符号理論がこれまでに果たしてきた役割を、次の100年でも果たそうとするならば、この問題は解くべき最も重要な問題の一つであると思われる。

## 文 献

- (1) C. Shannon, "The mathematical theory of communication," Bell Syst. Tech. J., vol. 27, no. 3, pp. 379-423; 623-656, 1948.
- (2) 進化する符号理論, 萩原 学(編), 日本評論社, 2016.
- (3) K.A.S. Immink, Codes for Mass Data Storage Systems, 2nd ed., Shannon Foundation Publishers, 2004.
- (4) 服部正勝, 鈴木 博, 菅谷誠一, "HDD, ODD, 及びSSDの技術動向," 東芝レビュー, vol. 66, no. 8, pp. 30-35, 2011.
- (5) R. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes," IEEE Trans. Inf. Theory, vol. IT-29, no. 1, pp. 5-22, Jan. 1983.
- (6) K.A.S. Immink, P.H. Siegel, and J.K. Wolf, "Codes for digital recorders," IEEE Trans. Inf. Theory, vol. IT-44, pp. 2260-2299, 1998.
- (7) B. Marcus, P.H. Siegel, and J.K. Wolf, "Finite-state modulation codes for data storage," IEEE J. Sel. Areas. Commun., vol. 10, no. 1, pp. 5-37, 1992.
- (8) B. Marcus and D. Lind, Symbolic dynamics and coding, Cambridge, 1995.
- (9) A.N. Kolmogorov, "A new metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces," Dokl. Akad. Nauk SSSR, vol. 119, no. 5, pp. 861-864, 1958.
- (10) D. Ornstein, "The impact of Roy Adler's work on symbolic dynamics and applications to data storage Bernoulli shifts with the same entropy are isomorphic," Advances in Math, vol. 4, no. 3, pp. 337-352, 1970.
- (11) B. Marcus, "Factors and extensions of full shifts," Monats. fur Math., vol. 88, no. 3, pp. 239-247, 1979.
- (12) H. Yasuda and H. Inose, "A necessary condition for constructing the balanced codes," (in Japanese) Trans. IECE Japan, vol. 55-A, pp. 194-195, 1972.
- (13) 安田 浩, 猪瀬 博, "ループ和零遷移図を用いた多値平衡符号の一般的構成法," 信学論(A), vol. J54-A, no. 9, pp. 506-513, Sept. 1971.
- (14) S. Yoshida and S. Yajima, "On the relation between an encoding automaton and the power spectrum of its output sequence," IECE Trans., vol. E59, no. 5, pp. 1-7, May 1976.
- (15) G. Pierobon, "Codes for zero spectral density at zero frequency," IEEE Trans. Inf. Theory, vol. IT-30, no. 2, pp. 435-439, March 1984.
- (16) B. Marcus and P. Siegel, "On codes with spectral nulls at rational submultiples of the symbol frequency," IEEE Trans. Inf. Theory, vol.

- IT-33, no. 4, pp. 557-568, July 1987.
- (17) B. Marcus, "Sofic systems and encoding data," IEEE Trans. Inf. Theory, vol. IT-31, no. 3, pp. 366-377, May 1985.
- (18) N. Yachie, K. Sekiyama, J. Sugahara, Y. Ohashi, and M. Tomita, "Alignment-based approach for durable data storage into living organisms," Biotechnol. Prog., vol. 23, no. 2, pp. 501-505, 2007.
- (19) S.M. Hossein Tabatabaei Yazdi, Y. Yuan, J. Ma, H. Zhao, and O. Milenkovic, "A rewritable, random-access DNA-based storage system," Scientific Reports, vol. 5, Aug. 2015, <http://www.nature.com/articles/srep14138>
- (20) S. Jain, F. Farnoud, M. Schwartz, and J. Bruck, Duplication-Correcting Codes for Data Storage in the DNA of Living Organisms, 2016, <http://authors.library.caltech.edu/63940/1/etr131.pdf>
- (21) S. Jain, F. Farnoud, M. Schwartz, and J. Bruck, "Duplication-correcting codes for data storage in the DNA of living organisms," Proc. ISIT 2016, pp. 1028-1032, 2016.
- (22) O. Milenkovic, "Constrained coding for context-free languages with applications to genetic sequence modelling," Proc. ISIT2007, pp. 1686-1690, Nice, 2007.
- (23) R. Rivest and A. Shamir, "How to reuse a "write-once" memory," Inf. Control, vol. 55, no. 1, pp. 1-19, 1982.
- (24) E. Sharon and L. Alrod, "Coding scheme for optimizing random I/O performance," in Non-Volatile Memories Workshop, San Diego, April 2013, <http://arxiv.org/abs/1202.6481>
- (25) E. Yaakobi and R. Motwani, "Construction of random input-output codes with moderate block lengths," IEEE Trans. Commun., vol. 64, no. 5, pp. 1819-1828, 2016.
- (26) R. Sekiya and B.M. Kurkoski, "Applying write-once memory codes to binary symmetric asymmetric multiple access channels," IEICE Trans. Fundamentals, vol. E99-A, no. 12, pp. 2202-2210, Dec. 2016.

(平成 29 年 1 月 25 日 受付)



かまべ ひろし  
鎌部 浩 (正員：シニア会員)

昭 57 豊橋技科大・工・情報卒, 昭 59 同大学院修士課程了. 同年三重大助手. 平 10 岐阜大講師. 現在, 同大学教授. 博士 (工学). 記録のための符号に関する研究に従事.