電気通信大学 電気通信学部 情報通信工学科

情報通信工学実験 A・B
実験項目 1. 情報通信 — 情報・セキュリティ —
「誤り訂正符号の理解と実装」
— リード・ソロモン符号と最小距離復号の理解と実装 (プログラミング) — [1]

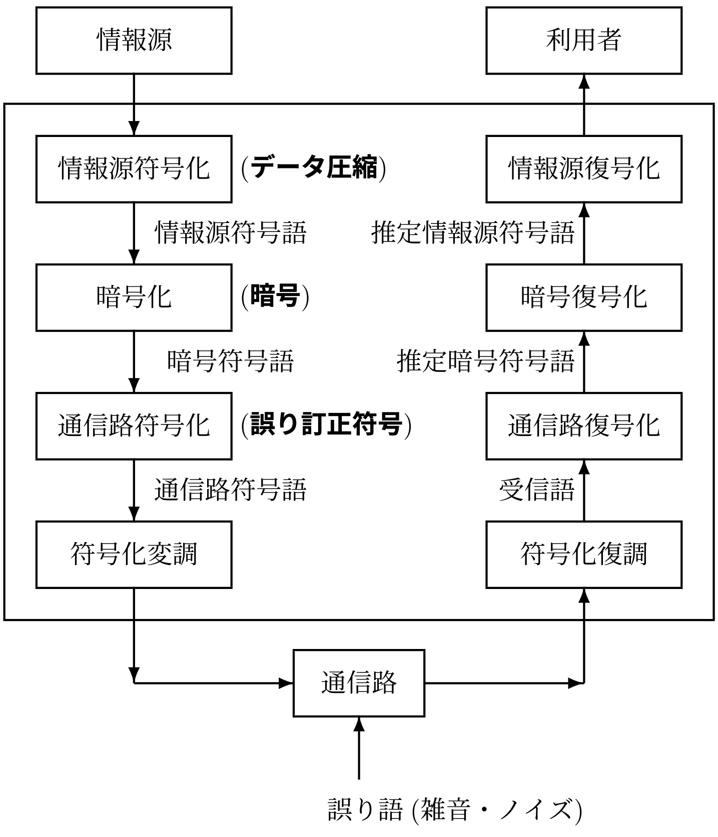
本実験項目「情報・セキュリティ」を履修する学生は、実験担当教員の指示に従って以下の三課題の中から一つの課題を選択し実験を行なう。

本テキストは、課題 **「誤り訂正符号の理解と実装」— リード・ソロモン符号と最小距離復号の理解と実装 (プログラミング) —** に関するテキストである。

1. 「データ圧縮の理解と実装」— ハフマン符号の理解と実装 (プログラミング) —
2. 「暗号化の理解と実装」— RSA 暗号の理解と実装 (プログラミング) —
3. 「誤り訂正符号の理解と実装」— リード・ソロモン符号と最小距離復号の理解と実装 (プログラミング) —

実験項目「情報・セキュリティ」では、下記図に示す 3 種類の**符号化**に関するテーマを扱う。一般に、情報を送信者 (情報源) から受信者 (利用者) に送信する際、次のような 3 種類の符号化を行なう。はじめに、効率性を目的として、**情報源符号化（データ圧縮）**を行なう。次に、安全性を目的として、**暗号化（暗号化）**を行なう。最後に、信頼性を目的として、**通信路符号化（誤り訂正符号）**を行なう。本実験では、いずれかの符号化の具体的手法について理解し、実際に計算機上でプログラムを作成し、実装を行なう。

ディジタル伝送・記憶システム



[1] ©電気通信大学 情報通信工学科 栗原正純 (e-mail: kuri@ice.uec.ac.jp), 2003 – 2009. (2009/3/31/11:21 修正)
本テキストは、http://www.code.ice.uec.ac.jp/kuri/C3/ より入手可能である。
(46: /doc/tex/uec/jikken/2009text/istext2009ecc.tex)

目的 (課題)

1. 誤り訂正符号 (通信路符号化) の一つであるリード・ソロモン (Reed-Solomon) 符号と最小距離復号について調べ、具体的な通信路符号化のための符号化と復号化の方法について理解する．
2. 次に，リード・ソロモン符号に対する最小距離復号を実行するプログラムを作成し，計算機上で実装し、その計算の量（複雑さ）を評価する．具体的には、後に示す 3.1 節の課題 1 を解く。

提出レポートの内容

1. リード・ソロモン符号および誤り訂正符号について調べたことを１ページ以内（A4 サイズ）に簡単にまとめ，記述する．
2. 最小距離復号を実装するに当たり，プログラミングで工夫した点を記述する．
3. プログラムのソースを印刷し，プログラムの各行（あるいは各部分）が何を実行する箇所なのかなどの注釈を記入したものをレポートに添付する．
4. レポートには "参考文献" という節（あるいは項目）を設け，主に参考にした文献を明記する．ここで，文献とは，書籍に限らない．インターネットなどを利用して得られたものも明記すること．その場合，URL を明記する．また，そのページのタイトルがあればそれも明記する．自分のアイデアと他人のアイデアを明確に区別すること．
5. 紙のレポートとは別にソースプログラムをメイルにて担当教員宛に送る．その際，そのファイルのコンパイル方法，実行方法も忘れずにメイルにて送る．

   Subject 欄には半角英数字を用いて "3jikken(name)" と記述し，メイル文章の初めに，氏名と学籍番号を書くこと．ここで、 "name" は各自の氏名のローマ字表記を書くこと。

1　数学的準備（有限体 Z_p）

すべての整数の集合 $\{\dots,-2,-1,0,1,2,\dots\}$ を $\mathbf{Z}$ と記す．その要素 a と b の差 $a-b$ が一つの整数 m で割りきれるとき，a と b は m **を法として互いに合同**であるといい，$a=b \pmod m$ と書く．

素数 p に対し，集合 $\mathbf{Z}_p$ を $\{0,1,2,\dots,p-1\}$ と定義する．このとき，集合 $\mathbf{Z}_p$ 上に p を法とする加法 $+$ と乗法 $\cdot$ の二つの演算を定義することで，集合 $\mathbf{Z}_p$ は，0 を零元，1 を単位元とする p 個の元からなる有限体となる．明示的にはこの有限体を $(\mathbf{Z}_p,+,\cdot,0,1)$ と記すが、本稿では，簡単に $\mathbf{Z}_p$ と書くことにする．

例えば，$p=5$ とすると $\mathbf{Z}_5=\{0,1,2,3,4\}$ の要素 2 と 4 の加法と乗法は，$2+4=1$, $2\cdot 4=3$ となる．

有限体 $\mathbf{Z}_p$ の元 a のべき乗 a^s, $s=1,2,\dots,$ を考える。このとき、$a^s=1$ となる最小の正整数 s の値が $p-1$ となる a を $\mathbf{Z}_p$ の**原始元**という。

例えば，$\mathbf{Z}_5$ の元のうち、0 と 1 は明らかに原始元ではない。2 と 3 は原始元である。すなわち、$2^1=2$, $2^2=4$, $2^3=3$, $2^4=1$ となる。同様に 3 の場合も成り立つ。しかし、4 については、$4^1=4$, $4^2=1$ となり、原始元ではない。

2　誤り訂正符号

2.1　符号、ハミング距離

記号 $\mathbf{Z}_q$ を q 個の元からなる**有限体**とする。そして、0 を零元、1 を単位元とする。有限体 $\mathbf{Z}_p$ 上の n 次元線型空間 $\mathbf{Z}_p^n$ の元を $\boldsymbol{a}=(a_1,\dots,a_n)$ と記し、ベクトルという。

有限体 $\mathbf{Z}_p$ 上の空間 $\mathbf{Z}_p^n$ に対しハミング距離を導入する。二つのベクトル $\boldsymbol{a}=(a_1,\ldots,a_n)$ と $\boldsymbol{b}=(b_1,\ldots,b_n)$ の**ハミング距離** $d(\boldsymbol{a},\boldsymbol{b})$ とは、$a_i \neq b_i$ となる添字 i の個数である。すなわち、$d(\boldsymbol{a},\boldsymbol{b})=|\{i|a_i \neq b_i\}|$ 。ハミング距離は距離の性質を満たす。

同様にして、ハミング重みも導入する。ベクトル $\boldsymbol{a}$ の成分のうち、0 でないものの個数を $w(\boldsymbol{a})$ と記し、ベクトル $\boldsymbol{a}$ の**ハミング重み**という。定義より、$d(\boldsymbol{a},\boldsymbol{0})=w(\boldsymbol{a})$ が成り立つ。

空間 $\mathbf{Z}_p^n$ の 部分集合 C を $\mathbf{Z}_p$ 上の符号長 n の**符号**という。符号 C の元を**符号語**という。特に、符号 C が空間 $\mathbf{Z}_p^n$ の k 次元線型部分空間 であるとき、C を $\mathbf{Z}_p$ 上の符号長 n, 次元 k の**線型符号**、または (n,k) **線型符号**という。

符号 C の**最小ハミング距離** $d(C)$ を $d(C)=\min\{d(\boldsymbol{a},\boldsymbol{b})|\boldsymbol{a},\boldsymbol{b}\in C \text{ such that } \boldsymbol{a}\neq\boldsymbol{b}\}$ と定義する。同様に、符号 C の**最小ハミング重み** $w(C)$ を $w(C)=\min\{w(\boldsymbol{a})|\boldsymbol{a}\in C \text{ such that } \boldsymbol{a}\neq\boldsymbol{0}\}$ と定義する。定義より、$d(C)=w(C)$ が成り立つ。

2.2 加法的離散通信路

本稿では、情報伝送のための通信路として加法的離散通信路を仮定する。**加法的離散通信路**とは、**送信語**として符号語 $\boldsymbol{c}=(c_1,\ldots,c_n)\in C$ を送信したとき、通信路で発生した雑音を表す**誤り語** $\boldsymbol{e}=(e_1,\ldots,e_n)\in\mathbf{Z}_p^n$ と送信語 $\boldsymbol{c}$ のベクトル和 $\boldsymbol{v}=\boldsymbol{e}+\boldsymbol{c}$ を**受信語**として受信することを仮定する通信路である。通信路で雑音が発生しない場合、誤り語は零ベクトルであり、受信語は送信語と一致する。発生した誤りが t 個の場合、誤り語 $\boldsymbol{e}$ の重みは $w(\boldsymbol{e})=t$ であり、送信語 $\boldsymbol{c}$ と受信語 $\boldsymbol{v}$ のハミング距離は $d(\boldsymbol{c},\boldsymbol{v})=t$ となる。これは $d(\boldsymbol{c},\boldsymbol{v})=d(\boldsymbol{c},\boldsymbol{e}+\boldsymbol{c})=d(\boldsymbol{e},\boldsymbol{0})=w(\boldsymbol{e})$ より明らか。

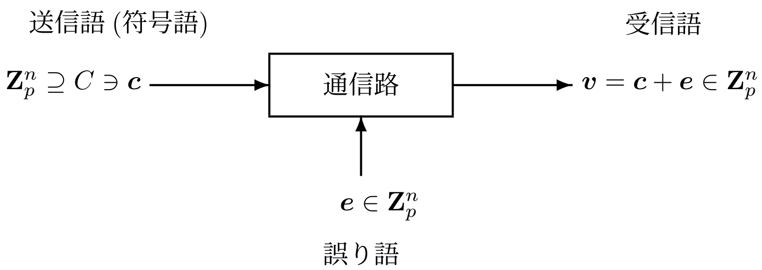


図 1: 加法的離散通信路

2.3 最小距離復号

送信語 $\boldsymbol{c}$ が送信されて受信語 $\boldsymbol{v}$ を受信したとき、$\boldsymbol{v}$ からハミング距離において最も近い符号語を送信語として推定する復号法を**最小距離復号**という。特に、そのような符号語が複数ある場合は、それらすべてを求める。

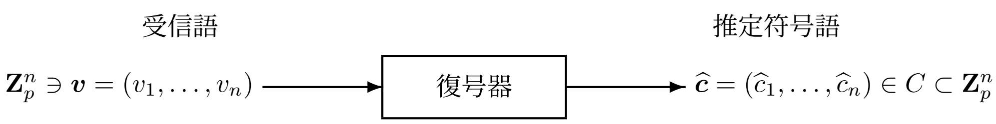


図 2: 通信路復号化

2.4 線型符号

有限体 $\mathbf{Z}_p$ の元を要素にもつ $m\times n$ 行列 $\boldsymbol{H}$ が与えられたとき、線型方程式 $\boldsymbol{H}{}^t\boldsymbol{x}={}^t\boldsymbol{0}$ の解 $\boldsymbol{x}=(x_1,\ldots,x_n)$ の全体は $\mathbf{Z}_p^n$ の中の部分空間になり、行列 $\boldsymbol{H}$ の階数を r とすると、解空間 C

の次元は $k = n - r$ で与えられる。ここで、記号 t はベクトルの転置を表す。C は $\mathbf{Z}_p^n$ の中の部分空間であるから、$\mathbf{Z}_p$ 上の (n, k) 線型符号である。C を線型符号とみなすとき、行列 $\boldsymbol{H}$ を C の**検査行列**という。

(n, k) 線型符号 C の k 個の一次独立なベクトルを $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k$ とする。それらは k 次元空間 C の基底をなし、C はこれらのベクトルによって生成される。これら k 個の行ベクトルを並べてできる $k \times n$ 行列 $\boldsymbol{G}$ を符号 C の**生成行列**という。明らかに、符号 C の検査行列と生成行列の間には、$\boldsymbol{H}^t\boldsymbol{G} = {}^t\boldsymbol{0}$ が成り立つ。

(n, k) 線型符号の場合、生成行列 $\boldsymbol{G}$ を用いることで情報源符号語（メッセージ）$\boldsymbol{m} = (m_1, \ldots, m_k) \in \mathbf{Z}_p^k$ から通信路符号語 $\boldsymbol{c} = (c_1, \ldots, c_n) \in C$ を $\boldsymbol{c} = \boldsymbol{m}\boldsymbol{G}$ により得ることができる。

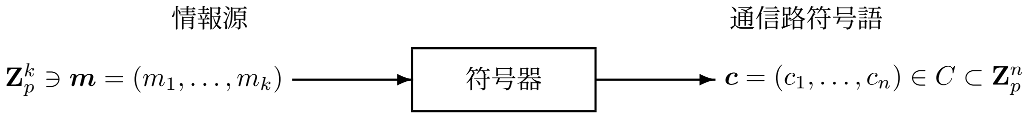


図 3: 通信路符号化

2.5 Z_p 上の Reed-Solomon 符号

有限体 $\mathbf{Z}_p$ の元を要素にもつ $(p-1) \times (d-1)$ 行列 $\boldsymbol{H}$ を次のように定義する.

$$\boldsymbol{H} = \begin{bmatrix} (a^0)^1 & (a^1)^1 & (a^2)^1 & \cdots & (a^{p-2})^1 \\ (a^0)^2 & (a^1)^2 & (a^2)^2 & \cdots & (a^{p-2})^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (a^0)^{d-1} & (a^1)^{d-1} & (a^2)^{d-1} & \cdots & (a^{p-2})^{d-1} \end{bmatrix} \tag{1}$$

ただし、a は有限体 $\mathbf{Z}_p$ の 原始元である。このとき，行列 $\boldsymbol{H}$ を検査行列とする $\mathbf{Z}_p$ 上の Reed-Solomon 符号 (RS 符号) C は、

$$C = \{\boldsymbol{c} = (c_1, \ldots, c_{p-1}) \in \mathbf{Z}_p^{p-1} | \boldsymbol{H}^t(c_1, \ldots, c_{p-1}) = {}^t(0, \ldots, 0)\} \tag{2}$$

として与えられる。Vandermonde 行列の性質より、検査行列 $\boldsymbol{H}$ の階数は $d-1$ であり、任意の $d-1$ 個の列ベクトルは線型独立である。それゆえ、C は $(p-1, p-d)$RS 符号となり、最小ハミング距離は d となる。

RS 符号の場合、検査行列 $\boldsymbol{H}$ が与えられたとき、通常の行列の基本行操作を行なうことで、$[\boldsymbol{I}_{d-1}\boldsymbol{P}]$ という形の検査行列を得ることができる。ここで、行に関する操作のみであることに注意する。この行列を標準形検査行列という。ここで、$\boldsymbol{I}_{d-1}$ は $(d-1) \times (d-1)$ 単位行列、$\boldsymbol{P}$ は $(d-1) \times (p-d)$ 行列である。このとき、符号 C の生成行列 $\boldsymbol{G}$ は $(p-d) \times (p-1)$ 行列 $[-{}^t\boldsymbol{P}\boldsymbol{I}_{p-d}]$ として与えられる。この行列を標準形生成行列という。そこで、RS 符号 C は $\boldsymbol{G}$ を用いて

$$C = \{\boldsymbol{m}\boldsymbol{G} | \boldsymbol{m} = (m_1, \ldots, m_k) \in \mathbf{Z}_p^k\} \tag{3}$$

と定義することもできる。

例えば、$(p, d) = (5, 3)$ と設定すると、$\mathbf{Z}_5$ 上の $(4, 2)$RS 符号 C を構成することができる。2 は $\mathbf{Z}_5$ の原始元であるから検査行列は

$$\boldsymbol{H} = \begin{bmatrix} (2^0)^1 & (2^1)^1 & (2^2)^1 & (2^3)^1 \\ (2^0)^2 & (2^1)^2 & (2^2)^2 & (2^3)^2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 1 & 4 & 1 & 4 \end{bmatrix} \tag{4}$$

となる。このとき、行基本操作を行なうことで標準形検査行列は

$$\begin{bmatrix} 2 & 4 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 & 3 \\ 1 & 4 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & 1 & 3 \end{bmatrix} \tag{5}$$

となる。$\boldsymbol{P} = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}$ より、$-{}^t\boldsymbol{P} = \begin{bmatrix} 3 & 4 \\ 3 & 2 \end{bmatrix}$. したがって、標準形生成行列 $\boldsymbol{G}$ は

$$\boldsymbol{G} = \begin{bmatrix} 3 & 4 & 1 & 0 \\ 3 & 2 & 0 & 1 \end{bmatrix} \tag{6}$$

となる。実際に、符号 C は以下のようになる。

$$C = \left\{ \begin{array}{lllll} (0000), & (3201), & (1402), & (4103), & (2304), \\ (3410), & (1111), & (4312), & (2013), & (0214), \\ (1320), & (4021), & (2222), & (0423), & (3124), \\ (4230), & (2431), & (0132), & (3333), & (1034), \\ (2140), & (0341), & (3042), & (1243), & (4444) \end{array} \right\} \tag{7}$$

3　課題

3.1　課題 1(必修)：Z_p 上の Reed-Solomon 符号の最小距離復号

パラメータ p, d を $(p, d) = (11, 6)$ と設定し、$\mathbf{Z}_{11}$ 上の $(10, 5)$ Reed-Solomon 符号 C を考える。この符号 C に対する最小距離復号を実行するプログラムを作成せよ。（もし、$p = 11$ が難しいようであれば $(p, d) = (7, 4)$ としてもよい。）

具体的には、$\mathbf{Z}_{11}^{10}$ の任意のベクトル $\boldsymbol{v}$ を選び、受信語とする。この受信語 $\boldsymbol{v}$ からハミング距離で最も近い符号語をすべて求めるプログラムを作成せよ。

プログラムの入出力：

入力	：受信語 $\boldsymbol{v}$.
出力	：$\boldsymbol{v}$ から最も近いすべての符号語。

3.1.1　復号例

$\mathbf{Z}_5$ 上の $(4, 2)$RS 符号 C を考える。

- 受信語 $\boldsymbol{v}$ が (0000) のとき、$\boldsymbol{v}$ に最も近い符号語は受信語自身の (0000) のみである。したがって、受信語から最も近い符号語とのハミング距離は 0 である。

入力	：(0000)
出力	：(0000)

- 受信語 $\boldsymbol{v}$ が (1000) のとき、$\boldsymbol{v}$ に最も近い符号語は (0000) のみである。したがって、受信語から最も近い符号語とのハミング距離は 1 である。

入力	：(1000)
出力	：(0000)

- 受信語 $\boldsymbol{v}$ が (1100) のとき、$\boldsymbol{v}$ に最も近い符号語は (0000), (1320), (2140), (1111), (1402), (4103) のみである。したがって、受信語から最も近い符号語とのハミング距離は 2 である。

入力	：(1100)
出力	：(0000), (1320), (2140), (1111), (1402), (4103)

3.2 発展課題 1(選択)：Z_p 上の Reed-Solomon 符号の最小距離復号（効率化）

一般に、$\mathbf{Z}_p$ 上の (n,k) 線型符号の場合、最小距離復号を行なうにはすべての符号語を対象に受信語とのハミング距離を調べる必要がある。その符号語の数は p^k 個である。例えば、課題 1 の場合、その数は $11^5 = 161051$ となる。p の値が大きくなると現実的な時間では、処理できない可能性がある。そこで、効率良く最小距離復号を実行する方法を提案し、プログラムを作成せよ。

3.3 発展課題 2(選択)

課題 1 または 発展課題 1 にて作成したプログラムに対し、パラメータ p または d を任意に設定できるように修正せよ。

参考文献

[1] 水野弘文, 情報代数の基礎，森北出版，1980.

[2] 平沢茂一、西島利尚, 符号理論入門，培風館，1999.

[3] Joern Justesen, Tom Hoeholdt, A course in error-correcting codes，European Mathematical Society，2004.

ASCHII

ASCII(アスキー) とは、American Standard Code for Information Interchange(情報交換用米国標準符号) の略。その中身は、1963 年にアメリカ規格協会 (ANSI: American National Standard Institute) が定めた、7/8 ビット英数字のコード体系の 1 つ。7 ビット版は、128 種類のローマ字、数字、記号、制御コードで構成されている。実際にはコンピュータは 1 文字を 8 ビット (1 バイト) で表現するため、256 種類の文字を扱うことができるが、ASCII が定めていない 128 文字分の拡張領域には、コンピューターメーカーや国によって異なる文字が収録されている。日本では、拡張領域にカナ文字を収録したコード体系が JIS X 0201 として規格化されている。

下記に、7 ビットコードを上位 3 ビットと下位 4 ビットに分けて 16 進表示したものを示す。たとえば、大文字の「Z」は、16 進表示では 5A、2 進表示では 1011010 となる。

ASCHII

下位 \ 上位	0	1	2	3	4	5	6	7
0	NUL	DLE	SP	0	@	P	‘	p
1	SOH	DC1	!	1	A	Q	a	q
2	STX	DC2	”	2	B	R	b	r
3	ETX	DC3	#	3	C	S	c	s
4	EOT	DC4	$	4	D	T	d	t
5	ENQ	NAC	%	5	E	U	e	u
6	ACK	SYN	&	6	F	V	f	v
7	BEL	ETB	’	7	G	W	g	w
8	BS	CAN	(	8	H	X	h	x
9	HT	EM	)	9	I	Y	i	y
A	LF/NL	SUB	*	:	J	Z	j	z
B	VT	ESC	+	;	K	[	k	{
C	FF	FS	,	<	L	\	l	|
D	CR	GS	-	=	M	]	m	}
E	SO	RS	.	>	N	ˆ	n	˜
F	SI	US	/	?	O	_	o	DEL

[2]©電気通信大学 情報通信工学科 栗原正純 (e-mail: kuri@ice.uec.ac.jp), 2005.(2005/05/10)
本資料は参考文献欄に挙げた文献をもとに作成したものである。

制御符号の説明

制御符号	コード	制御符号名	意味
NUL	00	Null	空
SOH	01	Start of Heading	ヘディング開始
STX	02	Start of Text	テキスト開始
ETX	03	End of Text	テキスト終了
EOT	04	End of Transmission	伝送終了
ENQ	05	Enquiry	問い合わせ
ACK	06	Acknowledge	肯定応答
BEL	07	Bell	ベル
BS	08	Backspace	バックスペース（1 文字後退する）
HT	09	Horizontal Tablation	水平タブ
LF/NL	0A	Line Feed/New Line	改行／復行（復帰・改行）
VT	0B	Vertical Tablation	垂直タブ
FF	0C	Form Feed	改頁
CR	0D	Carriage Return	復帰
SO	0E	Shift Out	シフト・アウト
SI	0F	Shift In	シフト・イン
DLE	10	Data Link Escape	データ・リンクでの拡張
DC1	11	Device Control 1(X-ON)	装置制御 1（送信を開始する要求に使用）
DC2	12	Device Control 2	装置制御 2
DC3	13	Device Control 3(X-OFF)	装置制御 3（送信を止める要求に使用）
DC4	14	Device Control 4	装置制御 4
NAC	15	Negative Acknowledge	否定応答
SYN	16	Synchronous Idle	同期文字
ETB	17	End of Transmission Block	伝送ブロック終了
CAN	18	Cancel	取り消し
EM	19	End of Medium	媒体終端
SUB	1A	Substitute Character	（CP/M でファイルのデータ終了記号に使用している）
ESC	1B	Escape	拡張（画面やグラフィックなどの制御コードの拡張に使用している）
FS	1C	File Separator	ファイル・セパレイタ
GS	1D	Group Separator	グループ・セパレイタ
RS	1E	Record Separator	レコード・セパレイタ
US	1F	Unit Separator	ユニット・セパレイタ
SP	20	Space	空白、ブランク、スペース
DEL	7F	Delete	抹消

参考文献

[1] `http://e-words.jp/w/ASCII.html`

[2] `http://www.psl.ne.jp/perl/pdojo00b.html`

[3] `http://www.komonet.ne.jp/~perl/chap13.htm`

[4] 椋田 (むくだ) 實, はじめての C (改訂第三版), 技術評論社, 1995(平成 7 年).

[5] 平林雅英，ANSI C 言語辞典 (初版 第 10 刷発行), 技術評論社, 2000(平成 12 年).