DATA COMPRESSION - TEXT, IMAGE, DATA

INTERCHANGE CODES

ERROR DETECTION AND CORRECTION

CODING TECHNIQUES

SHANNON-FANO ENCODING (DIVIDE AND CONQUER)

1. ARRANGE SYMBOLS IN DECREASING ORDER OF PROBABILITY.
2. DIVIDE INTO TWO GROUPS OF EQUAL PROBABILITY EACH.
3. ASSIGN FIRST GROUP 0, THE OTHER 1. APPEND TO ANY DIGITS ALREADY ASSIGNED.
4. FOR EACH GROUP DO 2. (RECURSE)

EXAMPLE

		SYMBOL	PROB.	CODE	LENGTH X PROB
G_{11}	G_1	x_1	1/4	0 0	1/2
G_{12}	G_1	x_2	1/4	0 1	1/2
G_{21}	G_2	x_3	1/8	1 0 0	3/8
G_{21}	G_2	x_4	1/8	1 0 1	3/8
ETC G_{22}	G_2	x_5	1/16	1 1 0 0	1/4
G_{22}	G_2	x_6	1/16	1 1 0 1	1/4
G_{22}	G_2	x_7	1/16	1 1 1 0	1/4
G_{22}	G_2	x_8	1/16	1 1 1 1	1/4
					11/4

AVERAGE LENGTH = 11/4 = 2.75

ENTROPY $H = \Sigma P_I \log P_I = 2.75$

UNCODED LENGTH = 3

ANOTHER EXAMPLE

CONSIDER TWO SYMBOLS: M AND S

$P(M) = 0.8$, $P(S) = 0.2$

ENTROPY $H = -\Sigma P_I \log P_I = 0.72$ BITS/SYMBOL

IF WE CODE M AS 0 AND S AS 1 WE GET 1 BIT/SYMBOL.

HOWEVER, CODE IN TWO-LETTER GROUPS

	P(SYMBOL PAIR)	SHANNON-FANO	LENGTH	AVG. LENGTH
MM	.64	0	1	.64
MS	.16	1 0	2	.32
SM	.16	1 1 0	3	.48
SS	.04	1 1 1	3	.12
				1.56 BIT PER 2 SYMB

AVG. LENGTH $= \frac{1.56}{2} = .78$ BIT/SYMBOL

IF WE FAKE LONGER AND LONGER GROUPS, AVG. LENGTH APPROACHES H!

ENGLISH - UNCODED - 4.75 BITS/SYMBOL
HUFFMAN - 4.2 "
SHANNON ~ 2-3 "

HUFFMAN ENCODING (OVER ALPHABET OF D SYMBOLS)

1. ARRANGE SYMBOLS IN DECREASING ORDER OF PROBABILITY.
2. ASSIGN LAST D SYMBOLS EACH A LETTER OF THE ALPHABET.
3. LUMP THESE SYMBOLS TOGETHER AS A NEW SYMBOL AND PREFIX TO ANY SYMBOLS OF THE ALPHABET ALREADY ASSIGNED.
4. REPEAT AT STEP 1.

EXAMPLE (USING 4-SYMBOL ENCODING ALPHABET)

SYMBOL	PROB				ASSIGN.
			.49 0		
x_1	.18		.18 1	1.0	1
x_2	.17		.17 2		2
		.16	.16 3		
x_3	.16	.16 0			00
x_4	.15	.15 1			01
x_5	.10	.10 2			02
x_6	.08	.08 3			03
x_7	.05 0				30
x_8	.05 1				31
x_9	.04 2				32
x_{10}	.02 3				33

(chars) Text	Distribution (percent) Bl	Let	Dig	Spec Graf	Oth	Line Length No Cod	Run Cod	Bl Cod	Esc Cod
360 asm sou	50	38	4	7	0	31	24	23	17
360 asm lst	69	12	16	2	<1	114	52	49	38
360 obj	20	15	6	2	56	80	67	72	121
PDP asm sou	53	37	4	5	0	35	25	24	18
PDP asm lst	57	17	22	3	0	70	46	43	32
PDP asm obj	<1	<1	<1	<1	99	183	182	185	277
PL1 sou	40	41	2	16	0	38	29	29	22
PL1 lst	64	17	13	4	1	103	54	52	39
SNOBOL sou	43	39	2	15	0	39	31	31	23

Table 3. Comparison of Remote Job Entry Codes

5. Interchange Codes

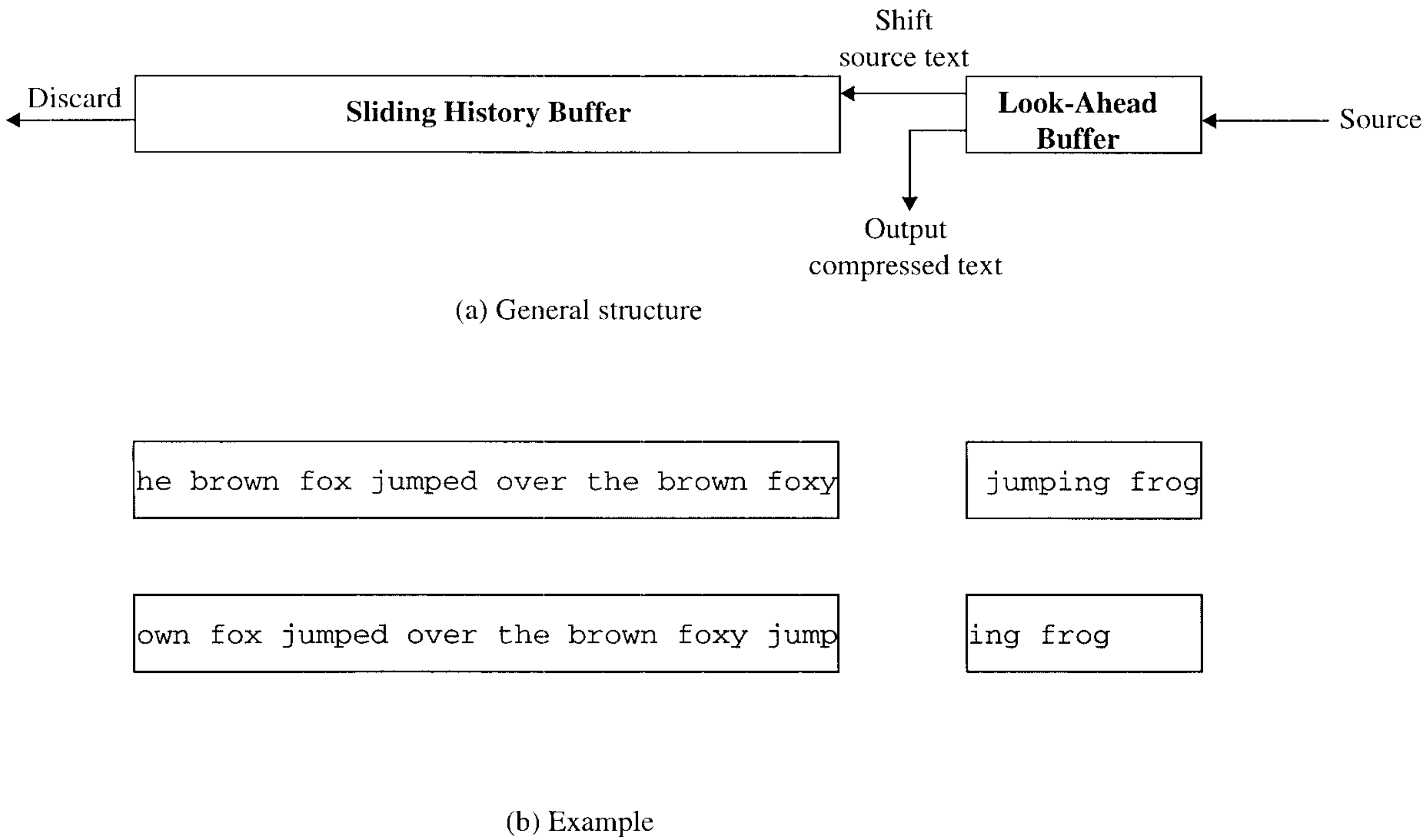


Figure 12.10 LZ77 Scheme

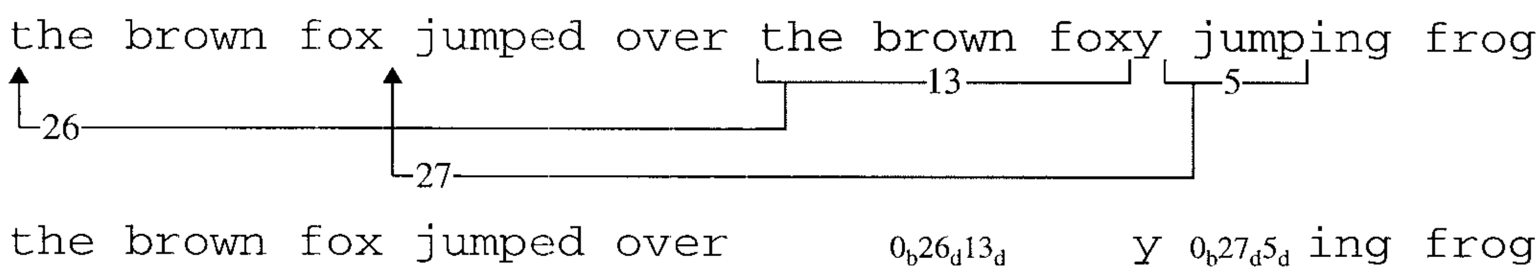
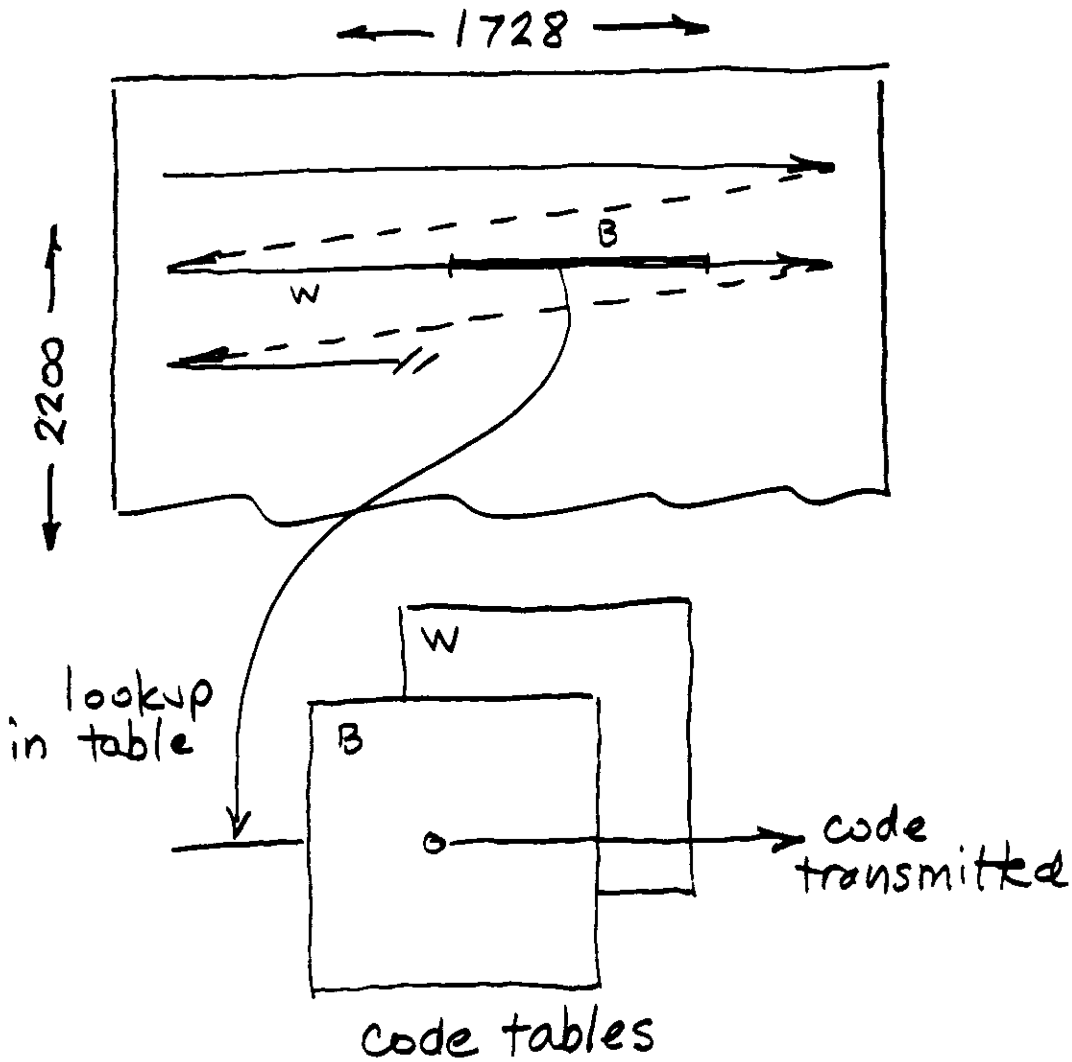
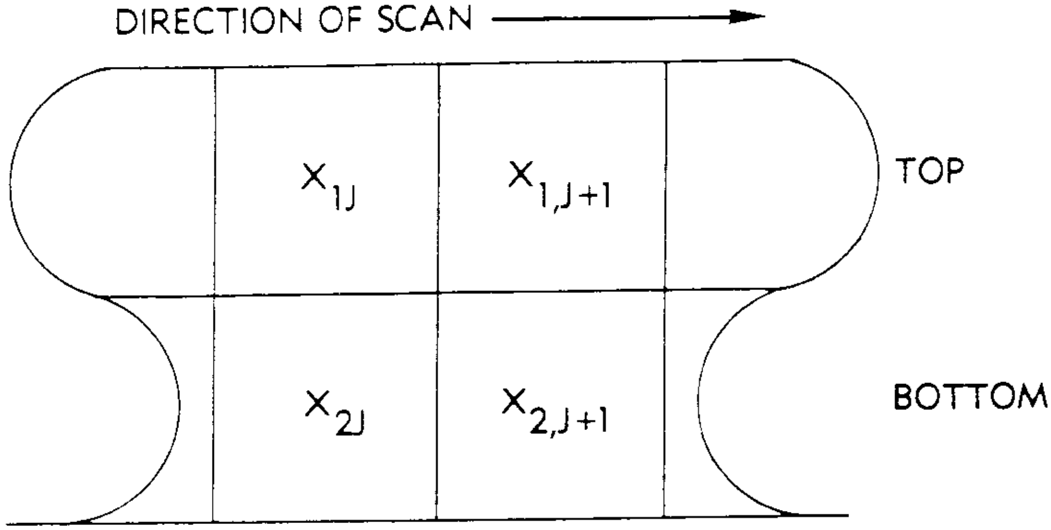


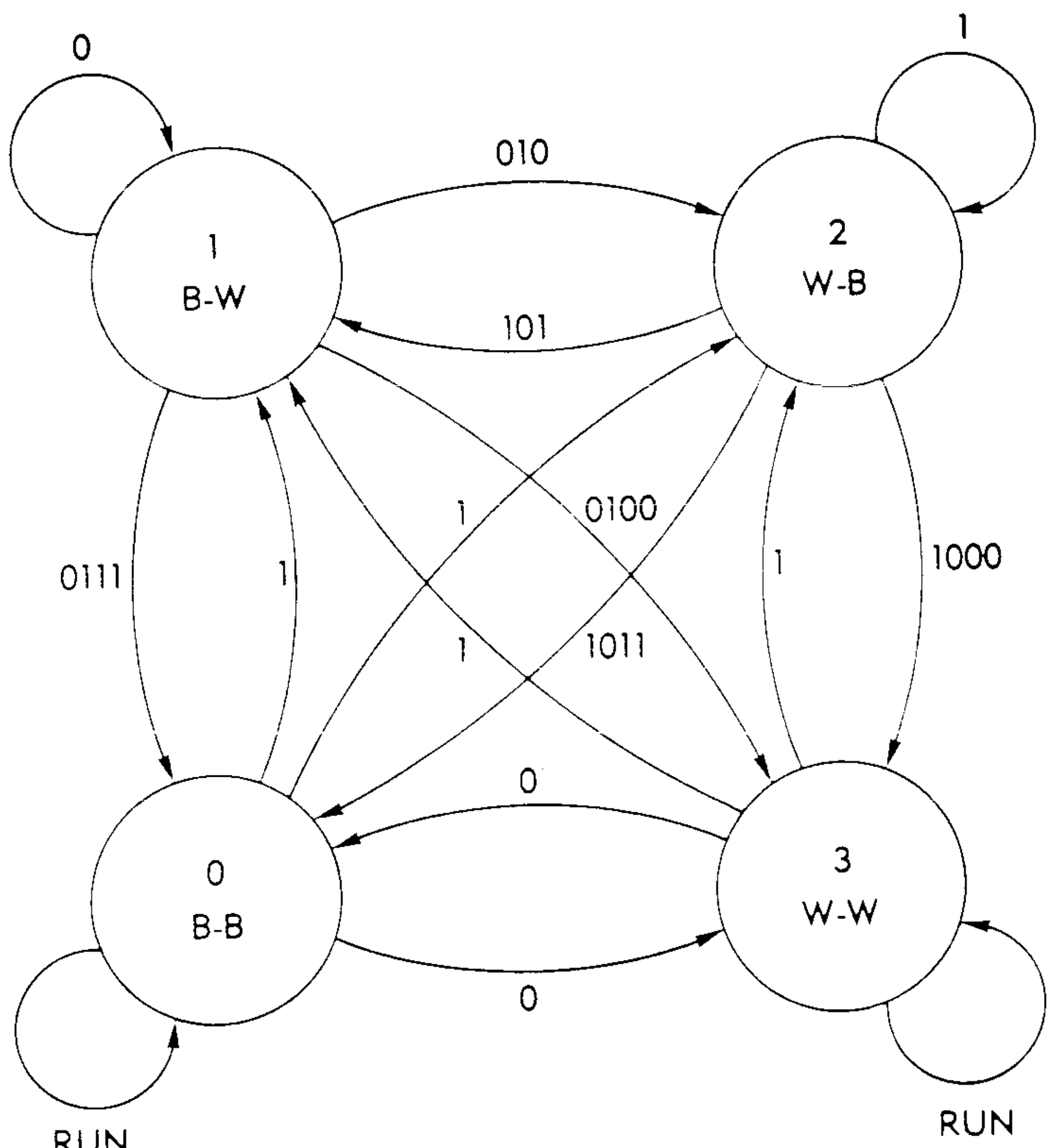
Figure 12.9 Example of LZ77 Scheme



T.4 Compression

Facsimile Codes





Rapifax Codes

Figure 2 - NFSA Model of Encoding Algorithm

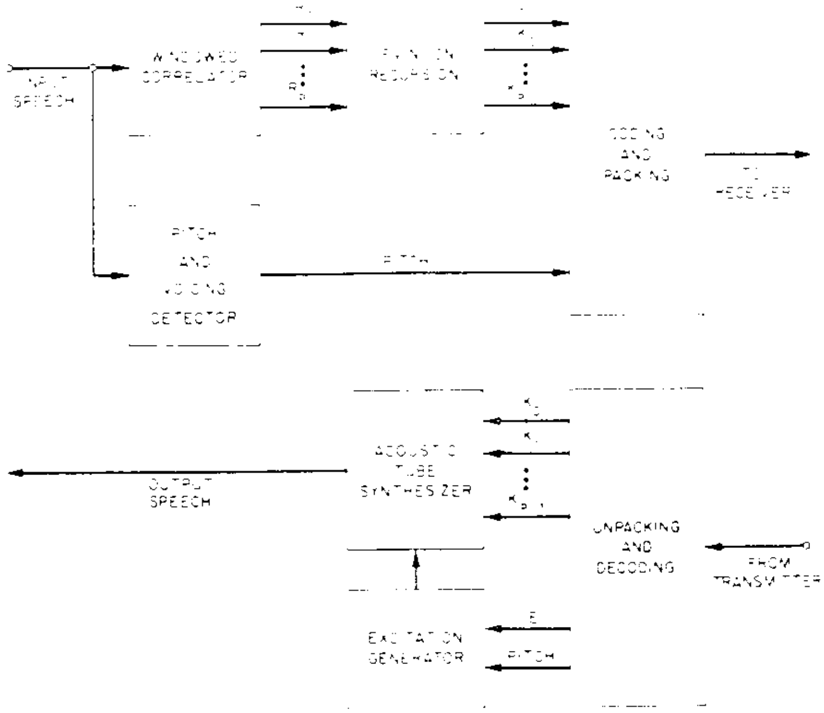


Fig. 6. The LPC vocoder algorithm.

LPCM data consists of bit-serial synchronous stream at 2400 bps. The bit stream is divided into 49-bit frames which encode the LPCM parameters for 20-ms intervals of the speech signal. The 49 bits are broken up into 13 variable-sized data words, each of which is transmitted least-significant-bit first:

```
Bit 0:        Toggling bit
Bits 1..7:    Pitch, or = 0 if unvoiced
Bits 8..12:   Energy
Bits 13..18: K0         Bits 19..24: K1
Bits 25..28: K2         Bits 29..32: K3
Bits 33..35: K4         Bits 36..38: K5
Bits 39..41: K6         Bits 42..44: K7
Bits 45..46: K8         Bits 47..48: K9
```

Each LPCM frame is packed into a parcel of 8 octets in the following format:

```
 | Octet 1                         | Octet 0
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| T |               Repeat Count                                |
+---------------------------------+-------------------------------+

 | Octet 3                         | Octet 2
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|  K0 low   |      Energy       |          Pitch                 |
+---------------------------------+-------------------------------+

 | Octet 5                         | Octet 4
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|       K3      |      K2       |            K1            | K0 hi |
+---------------------------------+-------------------------------+

 | Octet 7                         | Octet 6
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|  K9   |  K8   |     K7    |    K6     |    K5     |     K4     |
+---------------------------------+-------------------------------+
```