

Instituto Tecnológico de San Luís Potosí

Centro de Telecomunicaciones

Teleproceso y Redes de Computadoras

Representación de la Información

Fís. Jorge Humberto Olivares Vázquez
Centro de Telecomunicaciones
Enero 2007

2. Representación de la Información

2.1 Información y Tipos de Información

2.1.1 Información Textual

2.1.2 Información Digital

2.2 Códigos de Caracteres

2.2.1 Código Baudot

2.2.2 Código ASCII

2.3 Códigos de Línea

2.3.1 Código NRZ-L

2.3.2 Código NRZ-M

2.3.3 Código NRZ-S

2.3.4 Código RZ

2.3.5 Código Bi Fase-L (Manchester)

2.3.6 Código Bi Fase-M

2.3.7 Código Bi Fase-S

2.3.8 Código Manchester Diferencial

2. Representación de la Información

2.1 Información y Tipos de Información

Información es el contenido de un mensaje.

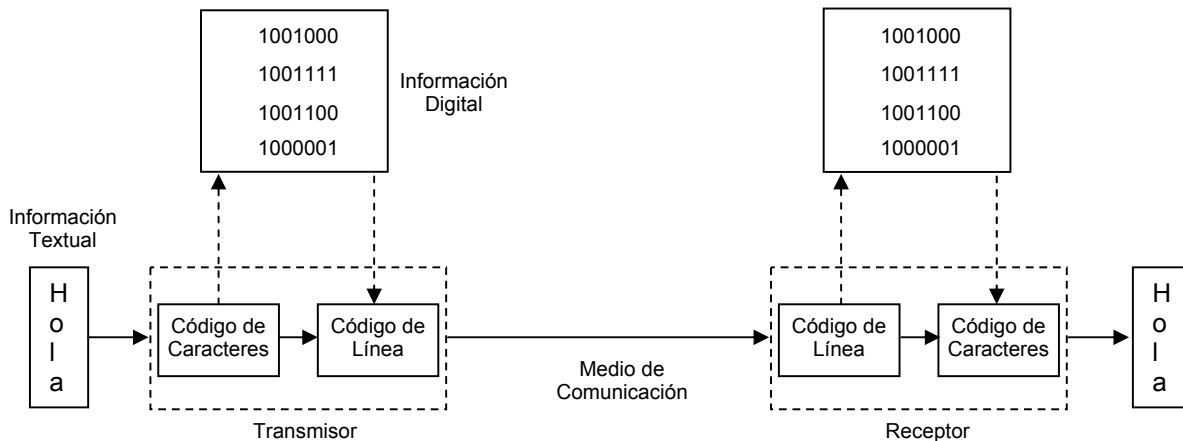
El mensaje lo podemos expresar de diferentes formas:

- por sonidos
- por escritura
- por graficas
- por símbolos

Para transmitirlo debemos elegir la forma más adecuada de expresarlo y representarlo.

Sin embargo, no siempre será tan simple transmitir información tal y como la estamos generando, habrán de producirse algunos cambios para lograrlo.

En Sistemas Digitales, la comunicación se efectúa dando formatos diferentes al mensaje. Estos, formatos son requeridos para representar la información de una manera adecuada en las diferentes etapas de la transmisión.



Para comprender lo anterior, inicialmente consideraremos que la Información es de dos tipos:

Información Textual

Información Digital

2.1.1 Información Textual

La *Información Textual* se define como una secuencia de caracteres bien definidos bajo un alfabeto. Así, la unidad de información, en Información Textual, es el carácter.

Este tipo de información, es transparente para el humano, lo que hace necesario que cualquier proceso de la información se inicie con su representación como información textual.

2.1.2 Información Digital

La información digital se define como una secuencia de símbolos generada bajo patrones bien definidos y utilizadas para la representación de caracteres. Al conjunto de estos patrones, son comúnmente llamados *Códigos de Caracteres*.

En Sistemas Digitales, utilizamos un sistema binario, esto es un sistema de dos símbolos para generar la representación de cualquier mensaje. Estos símbolos, el uno y el cero, son llamados bits. Con estos símbolos, se forman patrones o formatos de bits utilizados para representar caracteres bajo ciertas reglas.

Por ejemplo:

$$A = 1000001$$

Es la representación binaria de un carácter "A" en el Código ASCII.

Con un sistema binario, podemos generar códigos de caracteres, considerando arreglos de símbolos binarios, así como número de bits para la representación de un carácter.

Si k es el número de bits utilizado para representar un carácter, entonces en el código tendremos

$$M = 2^k$$

arreglos de símbolos o caracteres, de k bits cada uno; y se dice que el Código es de Nivel k .

En la siguiente sección se comentarán dos de los Códigos de Caracteres más conocidos:

BAUDOT

ASCII.

Aunque el Código Baudot está ya fuera de uso, se le considera en estas notas con fines didácticos.

2.2 Códigos de Caracteres

La codificación de caracteres es la representación de caracteres por medio de símbolos binarios (bits) para su procesamiento y/o almacenamiento.

Dicho de otra manera, es la técnica a la que recurrimos para la representación simbólica de la información textual para poderla procesar y/o almacenar en la memoria de una computadora, o bien, transferirla a un dispositivo externo.

Algunos códigos de caracteres utilizados son:

Código Baudot

Código ASCII

Código EBCDIC

2.2.1 Código Baudot

Este código fue creado por el Ingeniero postal Emmil Baudot, en Francia, en el año de 1874.

Se encuentra registrado por CCITT (ITU Telecommunications) como la recomendación Alfabeto Internacional No.2.

Es un código de nivel 5 ($k = 5$) con el cual se pueden generar

$$M = 2^k = 2^5 = 32 \text{ caracteres}$$

Código Baudot					
Alfabeto Internacional No 2					
Binario	LS	FS	Binario	LS	FS
00000	BLK	BLK	10000	T	5
00001	E	3	10001	Z	+
00010	LF	LF	10010	L	)
00011	A	-	10011	"	2
00100	SP	SP	10100	H	Libra
00101	S	'	10101	Y	6
00110	I	8	10110	P	0
00111	U	7	10111	Q	1
01000	<	CR	11000	O	9
01001	D	WAY	11001	B	?
01010	R	4	11010	G	\$
01011	J	BELL	11011	FS	FS
01100	N	,	11100	M	.
01101	F	%	11101	X	/
01110	C	:	11110	V	-
01111	K	(	11111	FS	LS

Como puede observarse, es un código limitado en la generación de caracteres ya que 32, son insuficientes para las exigencias de la representación adecuada de un mensaje alfanumérico.

Para eliminar esta limitante, se generan dos grupos opcionales de caracteres:

Caracteres Letra

Caracteres Figura

seleccionándose cada uno de ellos, según el carácter que se desea generar, por medio de caracteres de desplazamiento: *Letters Shift* (LS) y *Figures Shift* (FS), de esta manera, se tienen 62 caracteres disponibles en el código. Anteponiendo uno de los caracteres de desplazamiento, elegiremos uno de los dos grupos de caracteres que define Baudot.

De la tabla anterior, se puede observar que los caracteres Baudot se pueden clasificar en dos grupos:

Caracteres de Información: Aquellos utilizados para expresar el contenido de un mensaje

Alfabéticos	A, B, C, D,, Z
Numéricos	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
Puntuación	, . : -
Agrupación	(,)
Relación	=, <, >
Operación	+, -, x, /
Especiales	%, \$, Libra

Caracteres de Control: Caracteres utilizados para controlar el formateo de la información en página ó pantalla

BLK
SP
LS
FS
CR

Este código fue ampliamente utilizado en máquinas del Servicio de Telex, ya desaparecidas.

2.2.2 Código ASCII

El *American Standard Code for Information Interchange* es un código de nivel 7 ($k = 7$) en el cual se pueden generar,

$$M = 2^k = 2^7 = 128 \text{ caracteres}$$

Esta reconocido por la CCITT (ITU) bajo la recomendación *Alfabeto Internacional No. 5*.

De los 128 caracteres, 96 de ellos son utilizados como caracteres de información, y los 32 restantes son utilizados como caracteres de control.

Código ASCII American Estándar Code For Information Interchange Alfabeto Internacional No 5								
USB LSB	000	001	010	011	100	101	110	111
0000	NUL	DLE	SP	0	@	P		p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	“	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	\$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	‘	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	-	o	DEL

Los *caracteres de control* se dividen en 4 grupos:

Caracteres Efectores de Formato

Caracteres Delimitadores de Información

Caracteres Controladores de Dispositivo

Caracteres de Control de Transmisión

Caracteres Efectores de Formato

Estos caracteres controlan la distribución física de la información en la página de una impresora ó en la pantalla de una terminal. Son 6 y están localizados en la columna 0 y renglones del 8-13 de la Tabla presentada.

Son designados por:

0/08 FE0 BS Back Space

0/09 FE1 HT Horizontal Tabulation

0/10 FE2 LF Line Feed

0/11 FE3 VT Vertical Tabulation

0/12 FE4 FF Form Feed

0/13 FE5 CR Carriage Return

Caracteres Delimitadores de Información

Son utilizados para delimitar lógicamente la información y facilitar el manejo de registros efectuados internamente por la computadora. Estos caracteres delimitan la información en

Unidad de Información

Registro = Conjunto de Unidades

Grupo = Conjunto de Registros

Archivo = Conjunto de Grupos

Son 4, y están localizados en la columna 1 y en los renglones del 12 al 15 de la tabla. Se designan por:

1/12 IS4 FS File Separator

1/13 IS3 GS Group Separator

1/14 IS2 RS Record Separator

1/15 IS1 US Unit Separator

Caracteres Controladores de Dispositivos

Son utilizados para controlar las funciones físicas de un dispositivo conectado a una terminal.

Son 4, y están localizados en la columna 1 y en los renglones del 1 al 4 de la tabla. Se designan por:

1/01 DC1 Activa el dispositivo

1/02 DC2 Desactiva el dispositivo

1/03 DC3 Activa la impresión de una pantalla

1/04 DC4 Desactiva teclado

Caracteres de Control de Transmisión

Estos caracteres, se utilizan para formatear el mensaje en tramas fácilmente reconocibles y operables para el receptor y también, para controlar el flujo de información en una transmisión.

Los caracteres de Control de Transmisión en ASCII, desempeñan un papel importante en la segmentación y bloqueo de un mensaje. Esto es, si el mensaje es una secuencia de bits muy grande, es conveniente dividirla en paquetes de información más pequeños llamados bloques (Segmentación). Estos bloques de información facilitan su procesamiento y transmisión en cualquier sistema considerándolos inclusive, como unidades para la detección y corrección de errores.

Esto nos permitirá:

Transmitir bloques uno por uno

Omitir la transmisión de un determinado bloque

Utilizar el bloque como una unidad para la detección y corrección de errores

El mensaje deberá contener, además del mensaje, la información necesaria para que llegue a su destino, y el destinatario reconozca quien lo envió. Esta información es incorporada al inicio del mensaje y se le conoce con el nombre de Encabezado (Header).

Además de la información necesaria:

Dirección de la terminal origen

Dirección de la terminal destino

el encabezado puede contener información de tipo administrativa como: identificación de la persona que envía el mensaje, identificación de la persona que recibe el mensaje, fecha en que se originó el mensaje, hora en que se originó el mensaje, identificación de la línea portadora y algún código de seguridad.

Como aspectos importantes del encabezado debe señalarse que, el encabezado no es relevante para la información, sólo define la trayectoria origen-destino; por otro lado, el encabezado puede omitirse dependiendo de las características del hardware y del software utilizados, así como del número de terminales a emplearse.

De acuerdo a lo anterior, veremos ahora cual es el papel que desempeñan los caracteres de control de transmisión en el formateo y transmisión del mensaje. Estos caracteres se localizan en las columnas 0 y 1 de la tabla dada.

Se designan por:

0/01 TC1 SOH Start of Header

0/02 TC2 STX Start of Text

0/03 TC3 ETX End of Text

0/04 TC4 EOT End of Transmission

0/05 TC5 ENQ Enquiry

0/06 TC6 ACK Acknowledge

1/00 TC7 DLE Data Line Escape

1/05 TC8 NAK Negative Acknowledge

1/06 TC9 SYN Synchronous

1/07 TC10 ETB End of transmission block

TC9 SYN.- Es un caracter de sincronía. Es utilizado para establecer un sincronismo entre transmisor y receptor, en ausencia de cualquier otro caracter, en un sistema de transmisión síncrona.

TC5 ENQ.- Es un caracter de petición de transmisión. Es utilizado por transmisor para solicitar respuesta del receptor. Se puede utilizar para iniciar una transmisión, o bien después de un EOT.

TC6 ACK.- Es un caracter de confirmación de recepción. Es utilizado por el receptor para confirmar la recepción de la señal del transmisor.

TC8 NAK.- Es un caracter de confirmación de no recepción. Es utilizado por el receptor para indicar que no se recibe adecuadamente la señal del transmisor.

TC1 SOH.- Es un caracter delimitador de inicio del encabezado. Se utiliza para indicarle al receptor que llega información relativa a las trayectorias del mensaje (Dirección origen-destino).

TC2 STX.- Es un caracter delimitador de inicio para un bloque de información. Se utiliza para indicarle al receptor que inicia el contenido del mensaje.

TC3 ETX.- Es un caracter delimitador de texto de mensaje. Se utiliza para indicarle al receptor que ha finalizado la transmisión de un mensaje, esto es, de un conjunto de bloques.

TC4 EOT.- Es un caracter delimitador de la transmisión. Es utilizado para indicarle al receptor que ha concluido la transmisión de uno ó más mensajes. Habilita la desconexión.

TC7 DLE.- Es un caracter utilizado para cambiar el significado de una secuencia de caracteres. Solamente son validos en secuencias DLE.

Caracteres Efectores de Formato				Caracteres Delimitadores de Información				Caracteres Controladores de Dispositivo			Caracteres de Control de Transmisión			
0/08	FE0	BS	Back Space	1/12	IS4	FS	File Separator	1/01	DC1	Activa Dispositivo	0/01	TC1	SOH	Start of Header
0/09	FE1	HT	Horizontal Tabulation	1/13	IS3	GS	Group Separator	1/02	DC2	Desactiva Dispositivo	0/02	TC2	STX	Start of Text
0/10	FE2	LF	Line Feed	1/14	IS2	RS	Record Separator	1/03	DC3	Activa Impresión de Pantalla	0/03	TC3	ETX	End of Text
0/11	FE3	VT	Vertical Tabulation	1/15	IS1	US	Unit Separator	1/04	DC4	Desactiva Teclado	0/04	TC4	EOT	End of Transmission
0/12	FE4	FF	Form Feed								0/05	TC5	ENQ	Enquiry
0/13	FE5	CR	Carriage Return								0/06	TC6	ACK	Acknowledge
											1/00	TC7	DLE	Data Line Escape
											1/05	TC8	NAK	Negative Acknowledge
											1/06	TC9	SYN	Synchronous
											1/07	TC10	ETB	End of transmission Block

2.3 Códigos de Línea

Como ya se comentó, la información sufre transformaciones para transmitirse. Es representada primero en forma transparente y natural por caracteres. Luego estos, son representados por símbolos binarios denominados bits. Este proceso es llamado Codificación de Caracteres.

Ya para transmitirse al medio de comunicación, estos símbolos binarios o bits, son representados por señales eléctricas discretas. Este proceso es llamado Codificación de Línea.

En todos los códigos de línea establecidos y utilizados para transmitir una señal digital, se utilizan funciones discretas de $V(t)$, esto es, definidas sólo en dos valores posibles para $V(t)$.

Así, la información se representa por dos *niveles de voltaje*:

Un nivel alto de $V(t)$, $V(t) = V(\text{constante})$

Un nivel bajo de $V(t)$, $V(t) = 0$

o bien, por *transiciones* entre estos dos niveles.

Para facilitar la implementación de los códigos, se define un periodo de bit constante T , intervalo de tiempo en el cual debe generarse eléctricamente la representación de un bit utilizando ya sea un nivel de voltaje de la señal y/o una transición.

Aunque existe una amplia gama de códigos de línea, en este trabajo solamente comentaremos 8 de ellos pertenecientes a los grupos NRZ, RZ y Bi - ϕ .

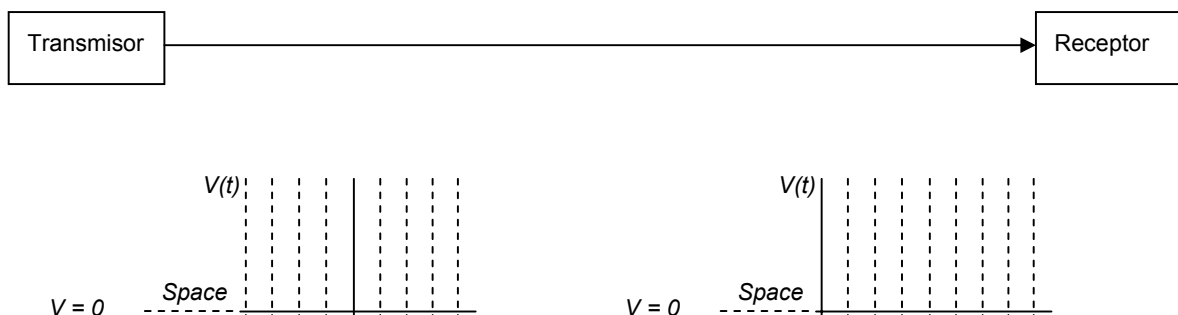
Antes de describir cada uno de los Códigos de Línea de las familias mencionadas, es conveniente considerar que una Línea de Datos, antes de transmitir información, se encuentra en *estado ocioso*.

Un estado ocioso es un estado de voltaje constante, ya sea un voltaje alto o bien, un voltaje bajo. Se reconocen técnicamente dos estados ociosos para una línea de datos:

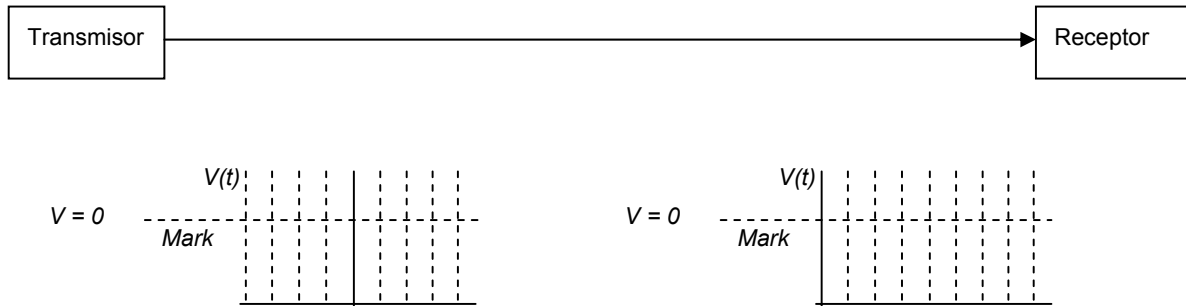
Estado Ocioso de Espacio *Space*

Estado Ocioso de Marca *Mark*

Un estado ocioso de espacio, *Space*, es un estado de voltaje bajo ($V = 0$ volts), como se muestra la figura siguiente.



Un estado ocioso de marca, *Mark*, es un estado de voltaje alto ($V = 5$ volts), como se muestra la figura siguiente.

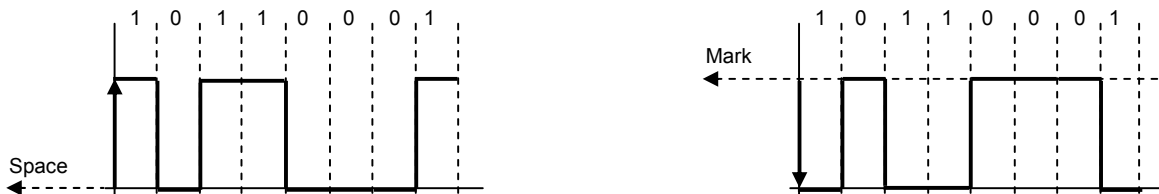


El transmisor le comunica al receptor que se inicia el envío de información, cuando rompe el estado ocioso utilizando una transición

2.3.1 Código NRZ-L

Non Return to Zero -Level

En este código, un voltaje alto es utilizado para representar un nivel lógico alto, esto es un 1; y un voltaje de cero volts es utilizado para representar un nivel lógico bajo, esto es un 0.

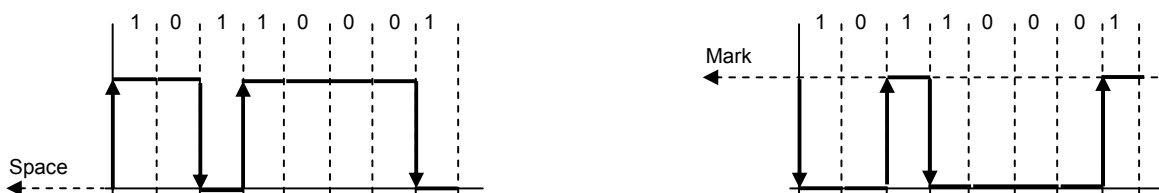


Este código es utilizado para la representación eléctrica de bits en el bus de una computadora, y en puertos paralelos.

2.3.2 Código NRZ-M

Non return to Zero-Mark

En este código, una transición al inicio de un intervalo o período de bit, representará un 1; la ausencia de esta transición, representará un 0.



2.3.3 Código NRZ-S

Non return to Zero-Space

En este código, una transición al inicio de un intervalo o período de bit, representará un 0; la ausencia de esta transición, representará un 1.

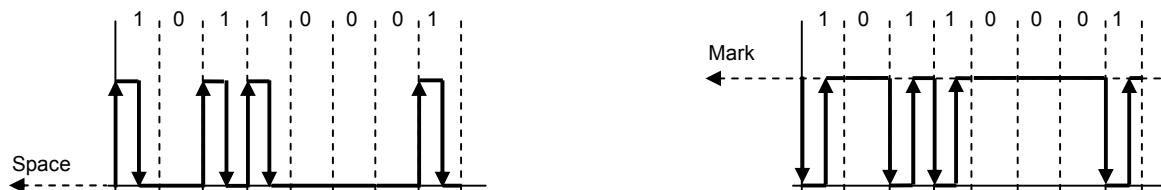


Como puede observarse, este código no se recomienda en sistemas que utilicen formatos de bits que inicien con “unos”.

2.3.4 Código RZ

Return to Zero

En este código, un 1 esta representado por la presencia de un pulso en la primera mitad del período del bit; la ausencia de este pulso nos representa un 0.



2.3.5 Bi-Φ-L

Biphase-L (Manchester)

En este código, el período del bit es dividido en dos partes iguales; si en la mitad del período del bit se presenta una transición de un voltaje alto a un voltaje bajo, esto representa un 1; si en la mitad del período del bit se presenta una transición de un voltaje bajo a un voltaje alto, esto representará un 0.



Este código es utilizado por la representación eléctrica de bits en adaptadores de red del estándar Ethernet y su compatible IEEE 802.3.

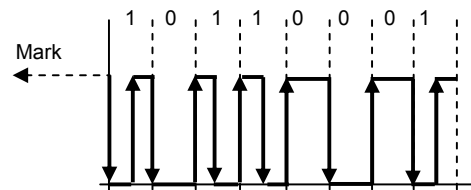
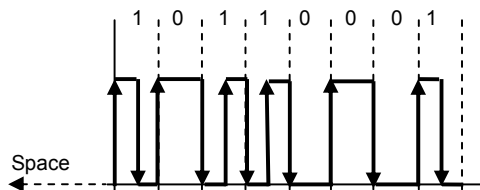
Como puede observarse, el estado ocioso de *Mark* de este código, no se recomienda en sistemas que utilicen formatos de bits que inicien con “unos”.

2.3.6 Bi- Φ -M

Biphase-M

En este código siempre deberá existir una transición al inicio de cada intervalo o período de bit.

Un 1 lo representará la presencia de una transición a la mitad del intervalo; un 0 lo representará la ausencia de una transición a la mitad del intervalo.

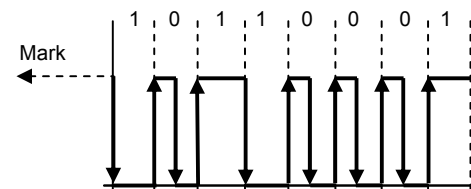
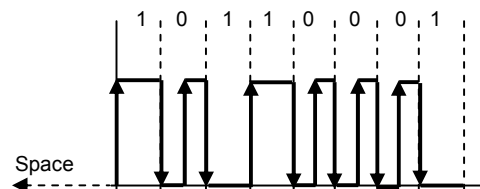


2.4.7 Bi- Φ -S

Biphase-S

En este código siempre deberá existir una transición al inicio de cada intervalo o período de bit.

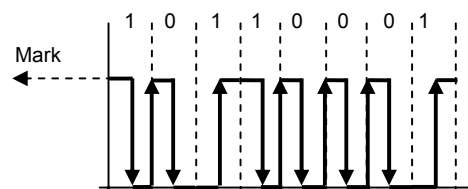
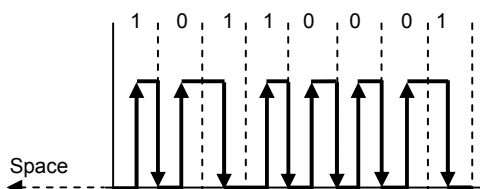
Un 1 lo representará la ausencia de una transición a la mitad del intervalo; un 0 lo representará la presencia de una transición a la mitad del intervalo.



2.4.8 Manchester Diferencial

En este código, siempre deberá existir una transición a la mitad del período del bit.

La ausencia de una transición al inicio del período del bit nos representará un 1; la presencia de una transición al inicio del período del bit nos representará un 0.



Como puede observarse, este código no se recomienda en sistemas que utilicen formatos de bits que inicien con “unos”.