

## Capítulo 5

# Transmisión digital a través de un canal de ruido blanco gaussiano

- Modulación de pulsos y su representación geométrica
- Receptor óptimo para modulación por pulsos en ruido blanco gaussiano
- Probabilidad de error para señales en ruido blanco gaussiano
- Repetidores regenerativos y análisis de potencia de un enlace

---

En el capítulo anterior se describieron varios métodos para convertir la salida de una fuente de información en una secuencia de dígitos binarios. En este capítulo se considerará la transmisión de la secuencia de información digital sobre canales de comunicaciones caracterizados por ruido blanco gaussiano aditivo (AWGN). El canal AWGN es uno de los de modelos matemáticos más simples para varios canales físicos de comunicaciones, incluyendo los cableados y algunos canales de radio.

El tratamiento se orienta a técnicas de modulación y demodulación digital apropiadas para *canales banda base*, o sea, canales con frecuencias que incluyen usualmente la frecuencia cero. Ejemplo de canales banda base son los canales cableados tal como el cable coaxial y los canales de par trenzado telefónico. Para tales canales la información digital puede transmitirse directamente mediante técnicas de modulación de pulsos banda base, o sea, mediante métodos de modulación que no requieren una portadora sinusoidal para efectuar traslación en frecuencia del espectro de la señal transmitida. En el capítulo 7 se verá como aplicar estos métodos de modulación digital a *canales pasabanda*, o sea, canales que tienen una banda de frecuencias útiles lejos del origen.

Comenzaremos desarrollando una representación geométrica para varios tipos diferentes de señales de modulación por pulsos. Se describirá la demodulación y detección óptima de esas señales. Finalmente, se evaluará la probabilidad de error como una medida del desempeño para varios tipos diferentes de señales de modulación sobre un canal AWGN. Los diferentes métodos de modulación serán comparados sobre la base de sus características de desempeño, sus requerimientos de ancho de banda y sus complejidades de implementación.

Inicialmente no se impondrán restricciones de ancho de banda en el diseño de las señales para modulación digital. Sin embargo, dado que el ancho de banda del canal es un parámetro importante que influye el diseño de la mayoría de los sistemas de comunicaciones, el diseño del modulador y demodulador para canales de ancho de banda limitado se estudiará en el capítulo 6.

## 5.1 Modulación de pulsos y su representación geométrica

Se discutirán en esta sección varios tipos de señales de modulación de pulsos utilizadas para la transmisión de información digital y se desarrollará una representación geométrica de esas señales. Las señales de modulación de pulsos consideradas incluyen: 1) señales de modulación de amplitud de pulsos, 2) señales de modulación de posición de pulsos (ortogonales), 3) señales biortogonales, 4) señales simplex, y 5) señales generadas a partir de secuencias de codificación binaria. Se describirá primero como transportar la información en esos tipos de señales.

### 5.1.1 Señales de modulación de pulsos

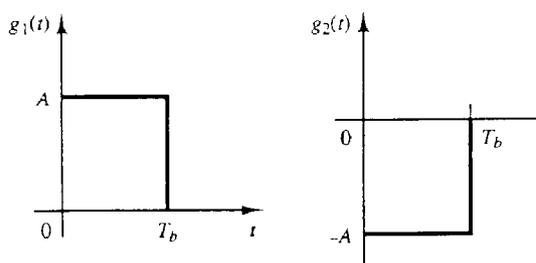


Figura 5.1: Ejemplo de señales para PAM binaria.

En la *modulación de amplitud de pulsos* (PAM), la información es transportada por la amplitud del pulso. Por ejemplo, en PAM binario, el bit de información 1 se representa mediante un pulso de amplitud  $A$  y el bit de información 0 mediante un pulso de amplitud  $-A$ , como ilustrado en la figura 5.1. Los pulsos se transmiten a una velocidad  $R_b = 1/T_b$  bits/seg, donde  $T_b$  se denomina el intervalo de bit. Aunque los pulsos se representan como rectangulares en la práctica el tiempo de subida y bajada no es nulo, de manera que la forma es más suave. El efecto de la forma del pulso sobre las características espectrales de la señal transmitida se considerará en el Capítulo 6.

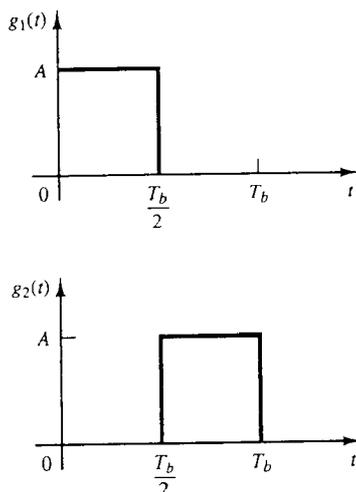


Figura 5.2: Ejemplo de señales para PPM binaria.

En la *modulación de posición de pulsos* (PPM), la información es transportada por el intervalo de tiempo en que aparece el pulso. Por ejemplo, en PPM binaria el intervalo de bit se divide en dos partes de  $T_b/2$  seg. cada una. El bit de información 1 se representa por un pulso de amplitud  $A$  en la primera parte y el bit de información 0 se representa mediante un pulso de amplitud  $A$  en la segunda parte del intervalo de bit, como mostrado en la figura 5.2. La forma del pulso dentro de cada parte no necesita ser rectangular.

Otro método de modulación de pulsos para transmitir información binaria se denomina *llaveado si/no* (OOK, on-off keying). La señal transmitida asociada a un 1 es un pulso de duración  $T_b$ . Si el bit asociado es un 0 no se transmite pulso en el intervalo de duración  $T_b$ .

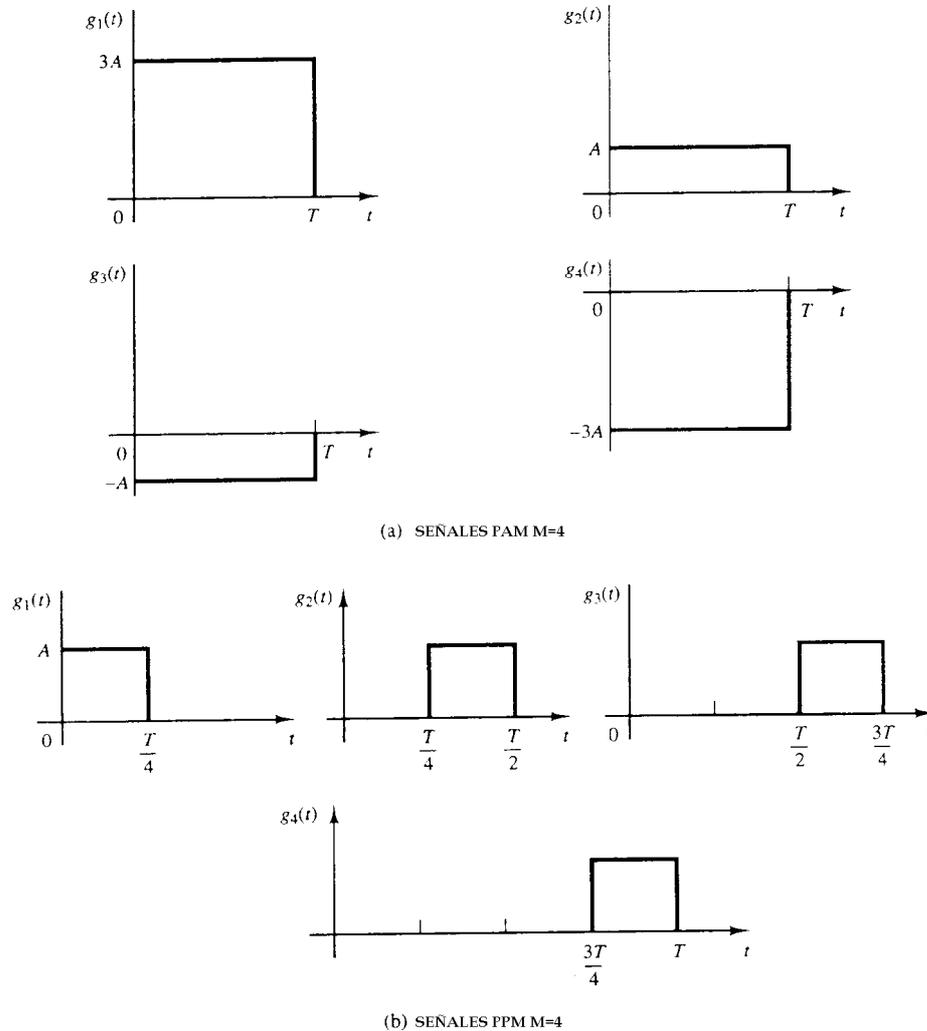


Figura 5.3: Ejemplo de señales PAM y PPM con  $M = 4$ .

La generalización de PAM y PPM para la transmisión de pulsos no binarios (multinivel o  $M$ -ários) es relativamente directa. En lugar de transmitir un bit por vez, la secuencia de información binaria se subdivide en bloques de  $k$  bits, denominados *símbolos*, y cada bloque o símbolo se representa por una de  $M = 2^k$  amplitudes de pulso para PAM, o posiciones de pulso para PPM. De esta forma, con  $k = 2$ , se tiene  $M = 4$  amplitudes de pulso o posiciones. La figura 5.3 ilustra las señales PAM y PPM para  $k = 2$  y  $M = 4$ . Notar que cuando la velocidad de transmisión es  $R_b$ , el *intervalo de símbolo* es

$$T = \frac{k}{R_b} = kT_b$$

como mostrado en la figura 5.4.

Es interesante caracterizar las señales PAM y PPM en términos de sus propiedades básicas. Por ejemplo, las señales PAM  $M$ -árias pueden expresarse como

$$s_m(t) = A_m g_T(t), \quad m = 1, 2, \dots, M, \quad 0 \leq t \leq T \quad (5.1)$$

donde  $g_T(t)$  es un pulso de forma arbitraria, como mostrado por ejemplo en la figura 5.5a. La característica distintiva de las  $M$  señales es la amplitud. Todas las  $M$  señales tienen la misma forma de pulso. Otra

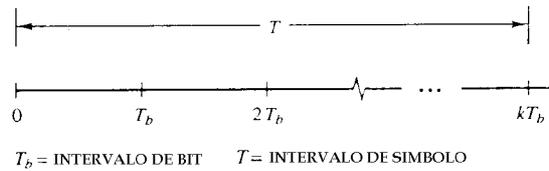


Figura 5.4: Relación entre el intervalo de símbolo y el intervalo de bit.

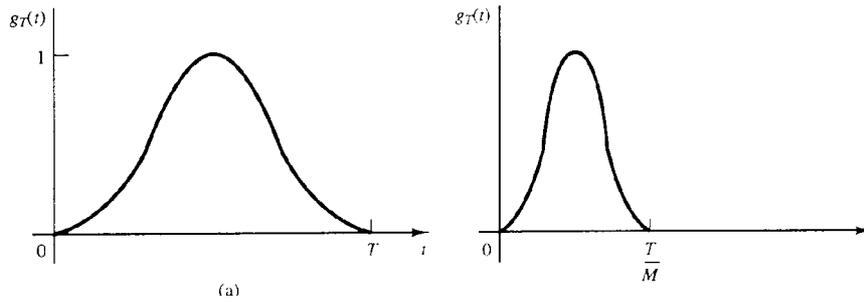


Figura 5.5: Señales pulso para a) PAM y b) PPM.

característica importante de esas señales es su energía. Es posible observar que las señales tienen diferentes energías, o sea,

$$\mathcal{E}_m = \int_0^T s_m^2(t) dt = A_m^2 \int_0^T g_T^2(t) dt, \quad m = 1, 2, \dots, M$$

En el caso de PPM, las señales pueden expresarse como

$$s_m(t) = A g_T(t - (m-1)T/M), \quad m = 1, 2, \dots, M, \quad (m-1)T/M \leq t \leq mT/M \quad (5.2)$$

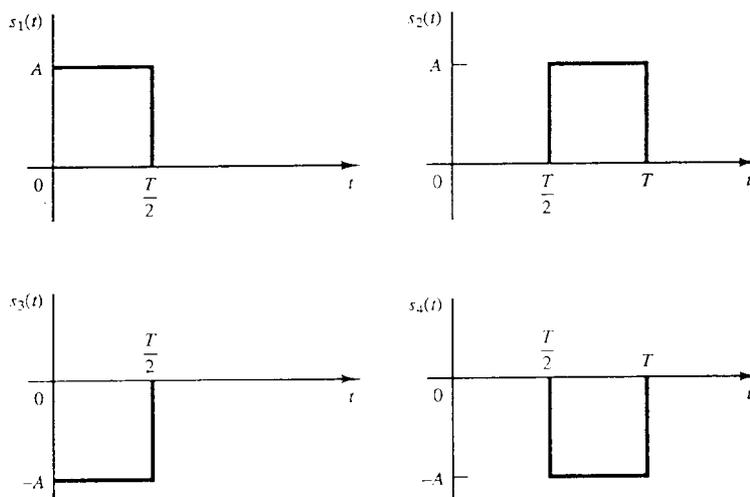
donde  $g_T(t)$  es un pulso de duración  $T/M$  y de forma arbitraria, como mostrado en la figura 5.5b. Una característica distintiva importante de esas formas de onda es que no son solapadas. En consecuencia

$$\int_0^T s_m(t) s_n(t) dt = 0, \quad m \neq n$$

Tales formas de onda se dicen *ortogonales*. Una segunda característica distintiva de las formas de onda de PPM es que todas tienen la misma energía, o sea

$$\begin{aligned} \int_0^T s_m^2(t) dt &= A^2 \int_{(m-1)T/M}^{mT/M} g_T^2(t - (m-1)T/M) dt \\ &= A^2 \int_0^{T/M} g_T^2(t) dt = \mathcal{E}_s, \quad \forall n \end{aligned}$$

Una diferencia importante entre el conjunto de  $M$  señales PAM y el conjunto de  $M$  señales PPM es el ancho de banda de canal requerido para su transmisión. El ancho de banda de canal requerido para transmitir señales PAM está determinado por las características en frecuencia del pulso  $g_T(t)$ . Este pulso básico tiene una duración  $T$ . Por otro lado, el pulso básico  $g_T(t)$  en un conjunto de  $M$  señales PPM tiene duración  $T/M$ . Cualquiera sea la forma elegida del pulso para PAM y PPM y para cualquier definición de ancho de banda empleada, es claro que el espectro del pulso PPM es  $M$  veces más ancho que el de PAM. En consecuencia, las señales PAM no requieren una expansión del ancho de banda de canal cuando  $M$  se incrementa. Por otro lado, las señales PPM requieren un incremento en el ancho de banda del canal con el aumento de  $M$ .

Figura 5.6: Conjunto de señales biortogonales con  $M = 4$ .

PAM y PPM son dos ejemplos de una variedad de diferentes tipos de conjuntos de señales que pueden construirse para la transmisión de información digital sobre canales banda base. Por ejemplo, tomando un conjunto de  $M/2$  señales PPM y construyendo los  $M/2$  pulsos negativos correspondientes, el conjunto combinado de  $M$  señales constituye un conjunto de  $M$  *señales biortogonales*. Un ejemplo para  $M = 4$  se muestra en la figura 5.6. Es fácil concluir que las  $M$  señales tienen igual energía. Además, el ancho de banda de canal requerido para transmitir las  $M$  señales es la mitad del requerido para transmitir  $M$  señales PPM.

Como otro ejemplo, es posible demostrar que de cualquier conjunto de  $M$  señales ortogonales es posible construir otro conjunto de  $M$  señales conocidas como *señales simplex*. A partir de las  $M$  señales ortogonales se resta el promedio de las  $M$  señales. Así

$$s'_m(t) = s_m(t) - \frac{1}{M} \sum_{k=1}^M s_k(t)$$

Luego, no es difícil mostrar que la energía de  $s'_m(t)$  es

$$\mathcal{E}'_s = \int_0^T [s'_m(t)]^2 dt = \left(1 - \frac{1}{M}\right) \mathcal{E}_s$$

y

$$\int_0^T s'_m(t)s'_n(t)dt = -\left(\frac{1}{M-1}\right) \mathcal{E}_s \quad m \neq n$$

donde  $\mathcal{E}_s$  es la energía de cada una de las señales ortogonales y  $\mathcal{E}'_s$  es la energía de cada una de las señales del conjunto simplex. Es posible notar que las señales en el conjunto simplex tienen menor energía que las del conjunto ortogonal. Además, es posible notar que las señales asociadas al conjunto simplex no son ortogonales. En su lugar, tienen una correlación negativa igual para todos los pares de señales. Se ha conjeturado que entre todas las señales multinivel posibles de igual energía  $\mathcal{E}_s$ , el conjunto de señales simplex resulta en la menor probabilidad de error cuando se utiliza para transmitir información sobre un canal AWGN. Esta conjetura no ha sido probada.

Como ejemplo final de la construcción de  $M$  señales, consideremos un conjunto de  $M$  palabras de código binario de la forma

$$\mathbf{c}_m = [c_{m1}, c_{m2}, \dots, c_{mN}], \quad m = 1, 2, \dots, M$$

donde  $c_{mj} = 0$  o  $1$  para todo  $m$  y  $j$ .  $N$  se denomina la *longitud del bloque* o dimensión de las palabras de código. Con  $M$  palabras de código es posible construir  $M$  señales mapeando un bit de código  $c_{mj} = 1$  en un pulso  $g_T(t)$  de duración  $T/N$  y un bit de código  $c_{mj} = 0$  en un pulso negativo  $-g_T(t)$ .

**Ejemplo:** Dadas las palabras de código:

$$c_1 = [1 \ 1 \ 1 \ 1]$$

$$c_2 = [1 \ 1 \ 0 \ 0]$$

$$c_3 = [1 \ 0 \ 1 \ 0]$$

$$c_4 = [0 \ 1 \ 0 \ 1]$$

construir un conjunto de  $M = 4$  señales, como descripto antes, utilizando un pulso rectangular  $g_T(t)$ .

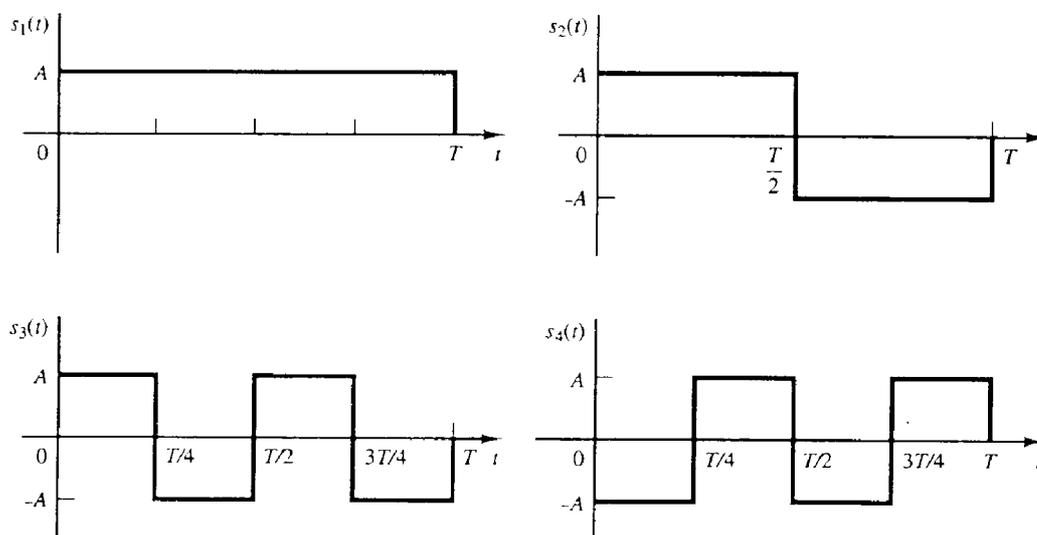


Figura 5.7: Conjunto de  $M = 4$  señales construido a partir de las palabras de código del ejemplo.

Como indicado antes, el bit de código 1 se mapea en un pulso rectangular  $g_T(t)$  de duración  $T/4$  y el bit de código 0 se mapea en un pulso rectangular  $-g_T(t)$ . De esta forma es posible construir las cuatro formas de onda mostradas en la figura 5.7 que corresponden a cuatro palabras de código. Es interesante notar que las primeras tres señales de esa figura son mutuamente ortogonales, pero la cuarta es la negativa de la tercera.

Existen numerosos códigos de bloques binarios diseñados en las pasadas décadas. Varios de esos códigos se describirá en el Capítulo 8. Cualquiera de ellos puede utilizarse para construir señales para transmisión  $M$ -ária sobre un canal banda base.

Además de las señales de modulación de pulsos descriptas, más adelante se discutirán otros tipos de señales banda base útiles para este propósito.

### 5.1.2 Representación geométrica

Para construir un conjunto de señales que forme una base ortonormal es posible utilizar *el procedimiento de Gram-Schmidt*. En esta sección se discutirá una representación geométrica de señales como si fuera puntos en un espacio de señales. Tal representación provee una caracterización compacta del conjunto de señales para transmitir información sobre un canal y simplifica el análisis de su desempeño.

Se asume disponible un conjunto de  $M$  señales  $s_m(t)$ ,  $1 \leq m \leq M$ , las cuales serán utilizadas para transmitir información sobre un canal de comunicaciones. Del conjunto de  $M$  señales, se construye primero

un conjunto de  $N \leq M$  señales ortogonales, donde  $N$  es la dimensión del espacio de señales. Con este propósito se utilizará el procedimiento de ortogonalización de Gram-Schmidt.

**Procedimiento de ortogonalización de Gram-Schmidt.** Se inicia con la primera señal  $s_1(t)$ , la que se supone de energía  $\mathcal{E}_1$ . La primera señal del conjunto ortonormal se construye simplemente como

$$\psi_1(t) = \frac{s_1(t)}{\sqrt{\mathcal{E}_1}}$$

De esta forma  $\psi_1(t)$  es simplemente  $s_1(t)$  normalizada para tener energía unitaria.

La segunda señal se construye de  $s_2(t)$ , calculando primero la proyección de  $s_2(t)$  sobre  $\psi_1(t)$ , la cual es

$$c_{12} = \int_{-\infty}^{\infty} s_2(t)\psi_1(t)dt$$

Luego,  $c_{12}\psi_1(t)$  se sustrae de  $s_2(t)$  para producir

$$d_2(t) = s_2(t) - c_{12}\psi_1(t)$$

Ahora,  $d_2(t)$  es ortogonal a  $\psi_1(t)$ , pero no tiene energía unitaria. Si  $\mathcal{E}_2(t)$  es la energía de  $d_2(t)$ , entonces la señal de energía normalizada que es ortogonal a  $\psi_1(t)$  es

$$\psi_2(t) = \frac{d_2(t)}{\sqrt{\mathcal{E}_2}}$$

En general, la ortogonalización de la  $k$ -ésima función conduce a

$$\psi_k(t) = \frac{d_k(t)}{\sqrt{\mathcal{E}_k}}$$

donde

$$d_k(t) = s_k(t) - \sum_{i=1}^{k-1} c_{ik}\psi_i(t)$$

y

$$c_{ik} = \int_{-\infty}^{\infty} s_k(t)\psi_i(t)dt, \quad i = 1, 2, \dots, k-1$$

De esta forma, el proceso de ortogonalización continúa hasta que todas las  $M$  señales  $\{s_m(t)\}$  se hallan utilizado y se hallan construido  $N \leq M$  señales ortonormales. Las  $N$  señales ortonormales  $\{\psi_n(t)\}$  forman una *base* en el espacio de señales  $N$ -dimensional. La dimensionalidad  $N$  del espacio de señales será igual a  $M$  si las  $M$  señales  $s_i(t)$  son linealmente independientes, o sea, si ninguna de esas señales es una combinación lineal de las otras.

---

**Ejemplo:** Se aplicará el procedimiento de Gram-Schmidt al conjunto de cuatro señales de la figura 5.8a. La señal  $s_1(t)$  tiene energía  $\mathcal{E}_1 = 2$ , tal que  $\psi_1(t) = s_1(t)/\sqrt{2}$ . Entonces es posible observar que  $c_{12} = 0$ , tal que  $\psi_1(t)$  y  $s_2(t)$  son ortogonales. En consecuencia,  $\psi_2(t) = s_2(t)/\sqrt{\mathcal{E}_2} = s_2(t)/\sqrt{2}$ . Para obtener  $\psi_3(t)$  calculamos  $c_{13} = 0$  y  $c_{23} = -\sqrt{2}$ . De esa forma

$$d_3(t) = s_3(t) + \sqrt{2}\psi_2(t)$$

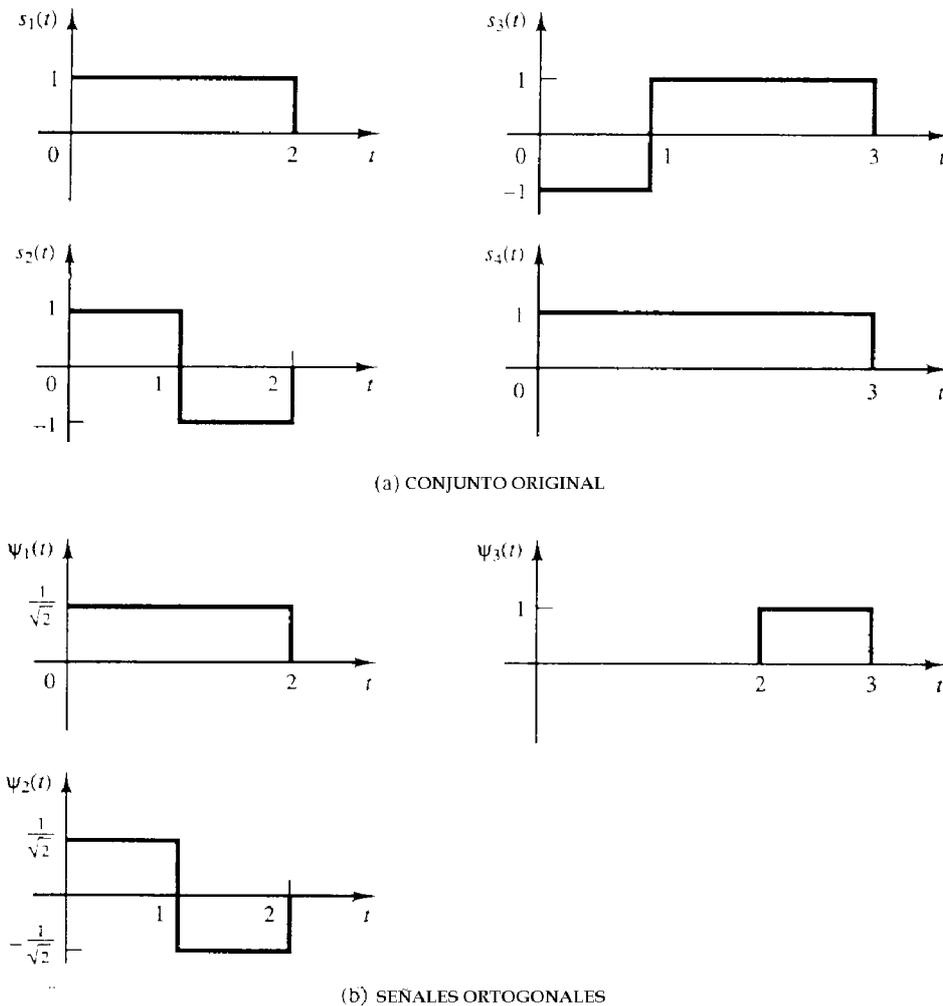


Figura 5.8: Aplicación del procedimiento de ortogonalización de Gram-Schmidt a un conjunto de señales.

Como  $d_3(t)$  tiene energía unitaria se tiene que  $\psi_3(t) = d_3(t)$ . Finalmente, calculando  $c_{14} = \sqrt{2}$ ,  $c_{24} = 0$  y  $c_{34} = 1$ , de donde se obtiene

$$d_4(t) = s_4(t) - \sqrt{2}\psi_1(t) - \psi_3(t) = 0$$

De esta forma,  $s_4(t)$  es la combinación lineal de  $\psi_1(t)$  y  $\psi_3(t)$ , y en consecuencia, la dimensión del conjunto de señales es  $N = 3$ . Las funciones  $\psi_1(t)$ ,  $\psi_2(t)$  y  $\psi_3(t)$  se muestran en la figura 5.8b.

Una vez que se construyó el conjunto de señales ortogonales  $\{\psi_n(t)\}$  es posible expresar las  $M$  señales  $\{s_m(t)\}$  como combinación lineal de  $\{\psi_n(t)\}$ . En consecuencia es posible escribir

$$s_m(t) = \sum_{n=1}^N s_{mn} \psi_n(t), \quad m = 1, 2, \dots, M \quad (5.3)$$

donde

$$s_{mn} = \int_0^T s_m(t) \psi_n(t) dt$$

y

$$\mathcal{E}_m = \int_{-\infty}^{\infty} s_m^2(t) dt = \sum_{n=1}^N s_{mn}^2$$

En base a la expresión (5.3) cada señal puede representarse mediante el vector

$$\mathbf{s}_m = [s_{m1}, s_{m2}, \dots, s_{mN}]$$

o equivalentemente, como un punto en el espacio  $N$ -dimensional de coordenadas  $\{s_{mi}, i = 1, 2, \dots, N\}$ . La energía de la  $m$ -ésima señal es simplemente el cuadrado de la longitud del vector o, equivalentemente, el cuadrado de la distancia Euclidiana desde el origen al punto en el espacio  $N$ -dimensional. De esta forma, cualquier espacio  $N$ -dimensional puede ser representado geoméricamente como un punto en el espacio de señales barrido por las  $N$  funciones ortogonales  $\{\psi_n(t)\}$ .

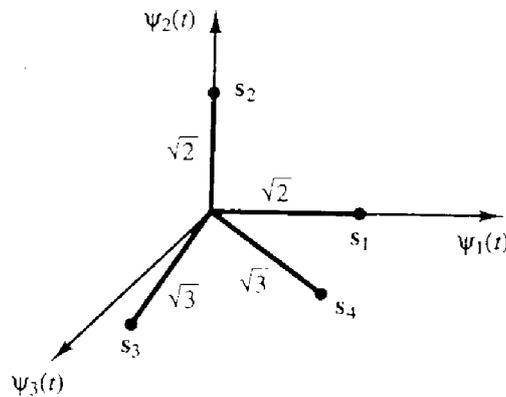


Figura 5.9: Vectores de señal correspondientes a las señales  $s_i(t)$ ,  $i = 1, 2, 3, 4$ .

---

**Ejemplo:** Se determinará la representación vectorial de las cuatro señales mostradas en la figura 5.8b. Dado que la dimensión del espacio es  $N = 3$ , cada señal se describirá mediante tres componentes, las que se obtienen proyectando cada una de las cuatro señales sobre las funciones base ortogonales  $\psi_1(t)$ ,  $\psi_2(t)$  y  $\psi_3(t)$ . De esta forma se obtiene:  $\mathbf{s}_1 = [\sqrt{2}, 0, 0]$ ,  $\mathbf{s}_2 = [0, \sqrt{2}, 0]$ ,  $\mathbf{s}_3 = [0, -\sqrt{2}, 1]$ ,  $\mathbf{s}_4 = [\sqrt{2}, 0, 1]$ . Estos vectores se muestran en la figura 5.9.

---

Finalmente, debería observarse que el conjunto de funciones base  $\{\psi_n(t)\}$  obtenidas por el procedimiento de Gram-Schmidt no es único. Por ejemplo, otro conjunto de funciones base que barre el espacio tridimensional se muestra en la figura 5.10. Para esta base, los vectores de señal son:  $\mathbf{s}_1 = [1, 1, 0]$ ,  $\mathbf{s}_2 = [1, -1, 0]$ ,  $\mathbf{s}_3 = [-1, 1, 1]$ ,  $\mathbf{s}_4 = [1, 1, 1]$ . Es posible notar también que el cambio de las funciones base no cambia las longitudes (o sea, la energía) de los vectores.

### 5.1.3 Representación geométrica de modulación de pulsos multinivel

Se representará un conjunto de señales PAM en forma geométrica. Las señales PAM  $M$ -árias pueden expresarse como sigue

$$s_m(t) = A_m g_T(t), \quad 0 \leq t \leq T, \quad m = 1, 2, \dots, M$$

donde  $M = 2^k$  y  $g_T(t)$  es un pulso de amplitud unitaria (ver figura 5.5a). Las señales PAM  $M$ -árias son unidimensionales, lo cual puede expresarse como

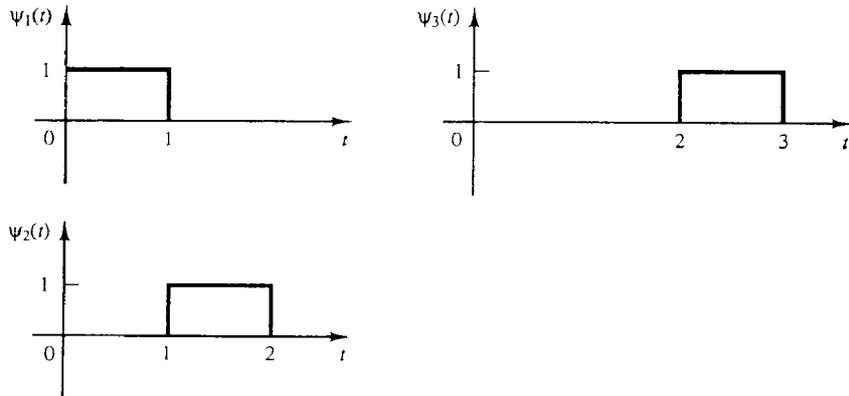


Figura 5.10: Conjunto alternativo de funciones base.

$$s_m(t) = s_m \psi(t), \quad m = 1, 2, \dots, M$$

donde la función base  $\psi(t)$  está definida por

$$\psi(t) = \frac{1}{\sqrt{\mathcal{E}_g}} g_T(t), \quad 0 \leq t \leq T \quad (5.4)$$

$\mathcal{E}_g$  es la energía del pulso  $g_T(t)$ , y los coeficientes (vector unidimensional) son

$$s_m = \sqrt{\mathcal{E}_g} A_m, \quad m = 1, 2, \dots, M$$

La distancia Euclidiana entre dos puntos de señal es

$$d_{mn} = \sqrt{|s_m - s_n|^2} = \sqrt{\mathcal{E}_g (A_m - A_n)^2}$$

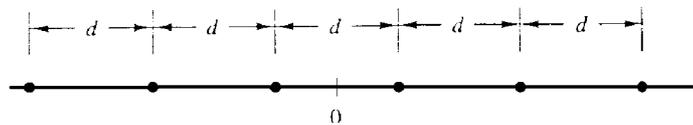


Figura 5.11: Puntos de señal (constelación) para PAM simétrica.

Si elegimos las amplitudes  $\{A_m\}$  de forma que estén simétricamente espaciadas alrededor del origen y equidistantes entre amplitudes de señal adyacentes, se obtienen los puntos de señal de PAM simétrica como mostrado en la figura 5.11.

Es posible observar que las señales PAM tienen diferentes energías. En particular, la energía de la  $m$ -ésima señal es

$$\mathcal{E}_m = s_m^2 = \mathcal{E}_g A_m^2$$

Para señales equiprobables, la energía promedio será

$$\mathcal{E}_{av} = \frac{1}{M} \sum_{m=1}^M \mathcal{E}_m = \frac{\mathcal{E}_g}{M} \sum_{m=1}^M A_m^2$$

Consideremos ahora un conjunto de  $M$  señales PPM como en (5.2), donde  $g(t)$  es un pulso de duración  $T/M$ , como ilustrado previamente en la figura 5.5b. Es necesario recordar que esas señales eran ortogonales y tienen la misma energía  $\mathcal{E}_s$ . En consecuencia, la representación de esas señales requiere  $M$  funciones base. Es posible definir las funciones base como

$$\psi_m(t) = \begin{cases} \frac{1}{\sqrt{\mathcal{E}_s}}g(t - (m-1)T/M), & (m-1)T/M \leq t \leq mT/M \\ 0 & \forall t \end{cases} \quad m = 1, 2, \dots, M.$$

De aquí es posible concluir que las señales PPM  $M$ -árias se representan geoméricamente mediante vectores  $M$ -dimensionales, dados por

$$\begin{aligned} \mathbf{s}_1 &= [\sqrt{\mathcal{E}_s}, 0, 0, \dots, 0] \\ \mathbf{s}_2 &= [0, \sqrt{\mathcal{E}_s}, 0, \dots, 0] \\ &\vdots \\ \mathbf{s}_4 &= [0, 0, 0, \dots, \sqrt{\mathcal{E}_s}] \end{aligned}$$

Claramente estos vectores son ortogonales, o sea,  $(\mathbf{s}_i \cdot \mathbf{s}_j) = 0$  cuando,  $i \neq j$ . Es también interesante notar que los  $M$  vectores de señal son mutuamente equidistantes, o sea

$$d_{mn} = \sqrt{|\mathbf{s}_m - \mathbf{s}_n|^2} = \sqrt{2\mathcal{E}_s}, \quad \text{para todo } m \neq n$$

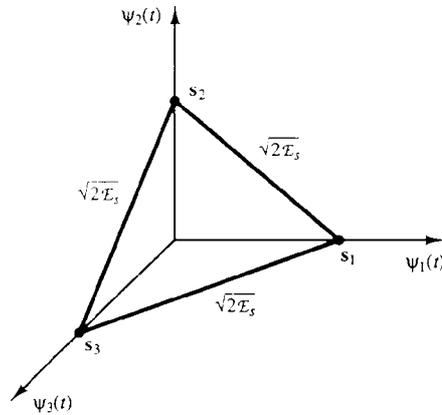


Figura 5.12: Señales ortogonales para  $M = N = 3$ .

En consecuencia, la distancia mínima entre puntos de señal es  $\sqrt{2\mathcal{E}_s}$ . La figura 5.12 muestra un ejemplo de señales ortogonales para  $M = 3$ .

La representación geométrica de señales  $M$ -árias biortogonales  $\{\mathbf{s}_m(t), 1 \leq m \leq M/2, -\mathbf{s}_m(t), M/2+1 \leq m \leq M\}$  es directa. Se tendrán en cuenta  $M/2$  vectores ortogonales en  $N = M/2$  dimensiones y luego se agregarán sus negativos. De esta forma se obtiene

$$\begin{aligned} \mathbf{s}_1 &= [\sqrt{\mathcal{E}_s}, 0, 0, \dots, 0] \\ \mathbf{s}_2 &= [0, \sqrt{\mathcal{E}_s}, 0, \dots, 0] \\ &\vdots \\ \mathbf{s}_{M/2} &= [0, 0, 0, \dots, \sqrt{\mathcal{E}_s}] \\ \mathbf{s}_{M/2+1} &= [-\sqrt{\mathcal{E}_s}, 0, 0, \dots, 0] \\ &\vdots \\ \mathbf{s}_M &= [0, 0, 0, \dots, -\sqrt{\mathcal{E}_s}] \end{aligned}$$

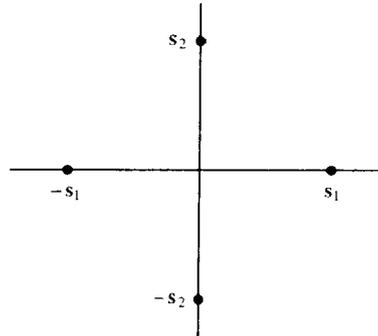


Figura 5.13: Constelación para señales biortogonales con  $M = 4$ .

La figura 5.13 ilustra un conjunto de vectores biortogonales para  $M = 4$ . Notar que la distancia entre cualquier par de vectores de señal es o  $2\sqrt{\mathcal{E}_s}$ , o  $\sqrt{2\mathcal{E}_s}$ . En consecuencia, la distancia mínima entre pares de vectores es  $d_{min} = \sqrt{2\mathcal{E}_s}$ .

La representación geométrica de un conjunto de señales simplex se obtiene restando la media del vector de señales del conjunto de  $M$  vectores ortogonales. De esta forma se tiene

$$\mathbf{s}'_m = \mathbf{s}_m - \frac{1}{M} \sum_{k=1}^M \mathbf{s}_k, \quad m = 1, 2, \dots, M$$

El efecto de restar la media de las señales  $\bar{\mathbf{s}} = \frac{1}{M} \sum_{k=1}^M \mathbf{s}_k$  de cada vector ortogonal es trasladar el origen de las  $M$  señales ortogonales al punto  $\bar{\mathbf{s}}$  y además minimizar la energía en el conjunto de señales  $\{\mathbf{s}'_m\}$ .

Si la energía por señal para las señales ortogonales es  $\mathcal{E}_s = |\mathbf{s}_m|^2$ , entonces la energía para las señales simplex es

$$\mathcal{E}'_s = |\mathbf{s}'_m|^2 = |\mathbf{s}_m - \bar{\mathbf{s}}|^2 = \left(1 - \frac{1}{M}\right) \mathcal{E}_s$$

La distancia entre dos puntos cualquiera no cambia por la traslación del origen, o sea, la distancia entre puntos de señal se mantiene  $d = \sqrt{2\mathcal{E}_s}$ . Finalmente, las  $M$  señales simplex están correlacionadas. El *coeficiente de correlación cruzada* (correlación cruzada normalizada) entre las señales  $m$  y  $n$  es

$$\gamma_{mn} = \frac{\mathbf{s}'_m \cdot \mathbf{s}'_n}{|\mathbf{s}'_m|^2 |\mathbf{s}'_n|^2} = \frac{-1/M}{(1 - 1/M)} = -\frac{1}{M - 1}$$

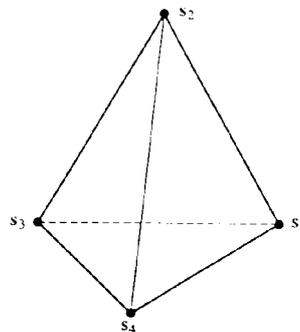


Figura 5.14: Constelación para señales simplex con  $M = 4$ .

En consecuencia, todas las señales tienen la misma correlación entre pares. La figura 5.14 ilustra un conjunto de  $M = 4$  señales simplex.

Finalmente, se considerará la representación geométrica de un conjunto de  $M$  señales generadas mediante un conjunto de  $M$  palabras binarias de la forma

$$\mathbf{c}_m = [c_{m1}, c_{m2}, \dots, c_{mN}], \quad m = 1, 2, \dots, M \quad (5.5)$$

donde  $c_{mj} = 0$  o  $1$  para todo  $m$  y  $j$ , como descrito en la primera sección de este Capítulo. Las  $M$  señales tienen dimensión  $N \leq M$  y se representan geoméricamente en forma vectorial de la forma

$$\mathbf{s}_m = [s_{m1}, s_{m2}, \dots, s_{mN}], \quad m = 1, 2, \dots, M \quad (5.6)$$

donde  $s_{mj} = \pm\sqrt{\mathcal{E}_s/N}$  para todo  $m$  y  $j$ .

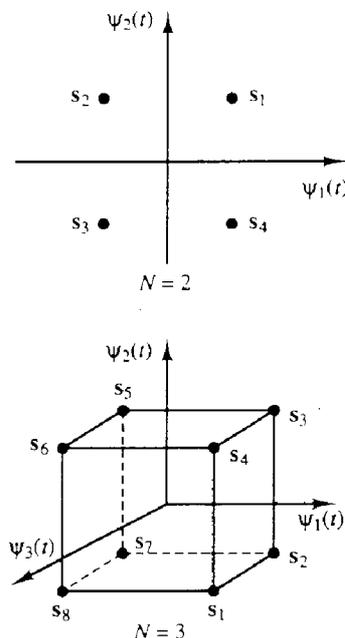


Figura 5.15: Diagrama del espacio de señales para señales generadas a partir de códigos binarios.

Es posible observar que existen  $2^N$  señales posibles que pueden construirse de  $2^N$  palabras de código binario. Es posible seleccionar un subconjunto de  $M < 2^N$  para la transmisión de la información. Es posible observar también que los  $2^N$  puntos de señal posibles corresponden a los vértices de un hipercubo  $N$ -dimensional con centro en el origen. La figura 5.15 ilustra los puntos de señal para  $N = 2$  y  $M = 3$  dimensiones.

Las  $M$  señales construidas de esta forma tienen igual energía  $\mathcal{E}_s$ . El coeficiente de correlación cruzada entre pares de señales depende de como se seleccione las  $M$  señales de entre las  $2^N$  posibles. Este tópico se discutirá en el Capítulo 8. Es aparente que puntos de señal adyacentes tendrán un coeficiente de correlación cruzada  $\gamma = (N - 2)/M$ , y una distancia Euclidiana correspondiente  $d = 2\sqrt{\mathcal{E}_s/N}$ .

#### 5.1.4 Señales y códigos de modulación con memoria

La codificación se utiliza generalmente en las comunicaciones digitales para lograr diferentes tipos de objetivos. Se estudió anteriormente la codificación de fuente como un medio para comprimir la salida de una fuente y representar la salida comprimida mediante una secuencia de dígitos binarios. La codificación se utiliza también para introducir redundancia en los datos del codificador de fuente para corregir errores que puedan ocurrir como resultado del ruido y otras interferencias en la transmisión de los datos a través del canal de comunicaciones. Este tipo de codificación se denomina *codificación de canal* y se discutirá en el Capítulo 8.

La codificación puede utilizarse también en un sistema de comunicaciones digitales para dar forma al espectro de la señal transmitida tal que se adecue a las características espectrales del canal banda base. La codificación que se utiliza para la conformación del espectro se denomina generalmente *código de modulación*, *código de línea*, ó *código de traducción de datos*. Tales códigos introducen generalmente restricciones sobre la *secuencia de bits* entrante al modulador y, de esta forma, introducen memoria en la señal transmitida. Este tipo de códigos se discutirá en esta sección. Se considerará las propiedades de conformación espectral de los códigos de modulación en el próximo Capítulo, donde se discutirá el problema de diseño de señales para canales de ancho de banda limitado.

Los códigos de modulación se emplean usualmente en las grabaciones magnéticas, en grabaciones ópticas, y en las comunicaciones digitales sobre sistemas cableados con el propósito de lograr la conformación espectral de la señal modulada de manera que se adecue a las características de la banda pasante del canal. Se discutirá la grabación magnética como un ejemplo.

En la grabación magnética se encuentra dos problemas básicos. Un problema está relacionado con la densidad de almacenamiento que se utiliza para escribir los datos sobre el medio magnético (disco o cinta). Obviamente se requiere escribir tantos bits cuanto sea posible en una única pista. Sin embargo, existe un límite en lo relativo a cuán próximos los bits sucesivos de una secuencia pueden almacenarse, y este límite está impuesto por el medio.

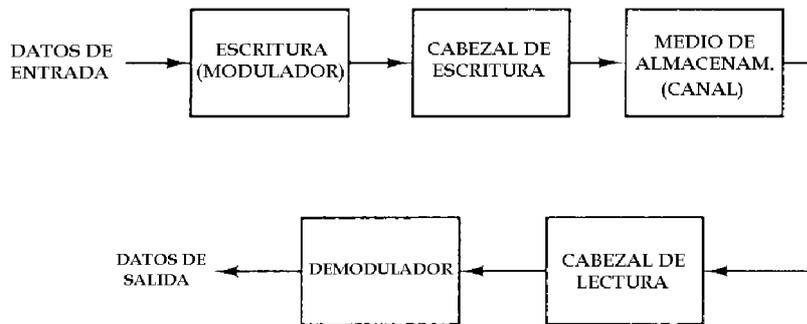


Figura 5.16: Diagrama en bloques de un sistema de lectura escritura de almacenamiento magnético.

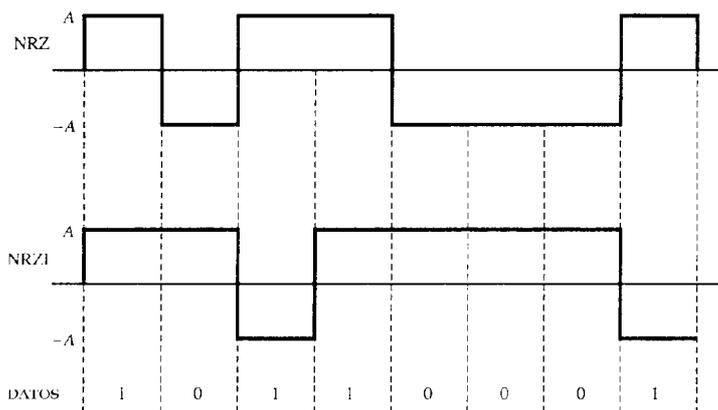


Figura 5.17: Señales NRZ y NRZI.

La figura 5.16 ilustra un diagrama en bloques de un sistema de grabación magnética. La secuencia de datos binarios a ser almacenada se utiliza para generar una corriente de escritura. Esta corriente puede verse como la salida del "modulador". Los métodos más comúnmente utilizados para mapear la secuencia de datos en la señal de corriente de escritura son el *no retorno a cero* (NRZ) y *no retorno a cero invertido* (NRZI). Estas dos señales se ilustran en la figura 5.17. Notar que la señal NRZ es idéntica a PAM binaria en la cual el bit de información 1 se representa mediante un pulso rectangular de amplitud  $A$  y el bit de información 0 se representa mediante un pulso rectangular de amplitud  $-A$ . En contraste, la señal NRZI se

diferencia de la NRZ en que las transiciones de un nivel de amplitud al otro ( $A$  a  $-A$  o  $-A$  a  $A$ ), ocurren solamente cuando el bit de información es un 1. No existen transiciones cuando el bit de información es un 0, o sea, el nivel de amplitud se mantiene igual que el del nivel previo de señal. El pulso de magnitud positiva resultante magnetiza al medio con una polaridad (dirección) y el pulso de magnitud negativa magnetiza al medio con la polaridad (dirección) opuesta.

Dado que la secuencia de datos de entrada es básicamente aleatoria con igual probabilidad de ceros y unos, independientemente de utilizar NRZ o NRZI se encontrarán niveles de transición de  $A$  a  $-A$  o de  $-A$  a  $A$  con probabilidad  $1/2$  para cada bit. La señal leída para una transición positiva ( $-A$  a  $A$ ) es un pulso adecuadamente modelado matemáticamente por

$$p(t) = \frac{1}{1 + (2t/T_{50})^2}$$

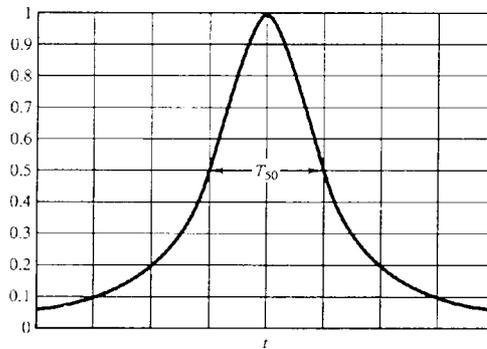


Figura 5.18: Pulso de lectura en un sistema de almacenamiento magnético.

donde  $T_{50}$  se define como el ancho del pulso para un 50 % del nivel de amplitud, como mostrado en la figura 5.18. En forma similar, la señal leída para una transición negativa ( $A$  a  $-A$ ) es el pulso  $-p(t)$ . El valor de  $T_{50}$  está determinado por las características del medio, los cabezales de lectura/escritura, y la distancia del cabezal al medio.

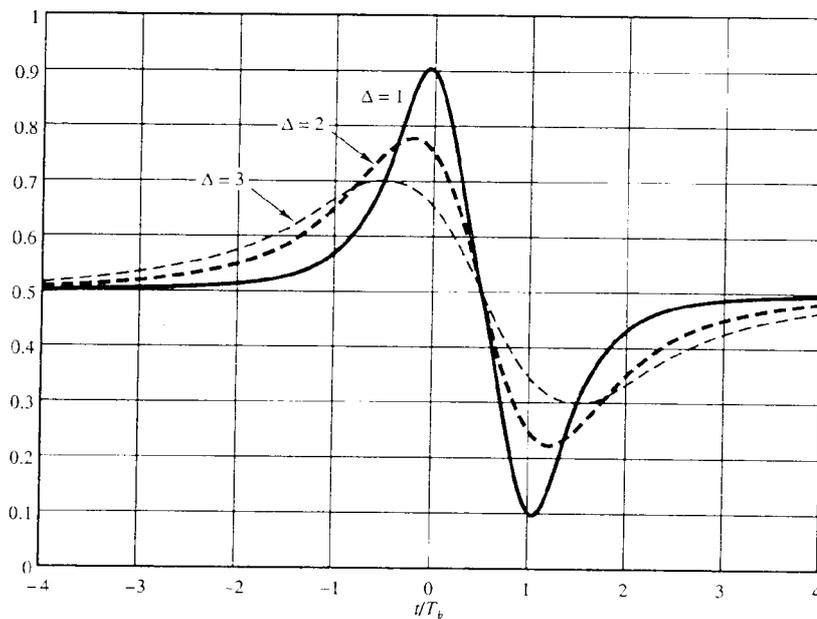


Figura 5.19: Respuesta a una señal pulso de lectura.

Es posible suponer la escritura de una transición positiva seguida de una transición negativa y además variar el intervalo de tiempo entre las dos transiciones, el cual denominamos como  $T_b$  (el intervalo de bit). La figura 5.19 ilustra los pulsos de lectura obtenidos mediante la superposición de  $p(t)$  con  $-p(t - T)$ . El parámetro  $\Delta = T_{50}/T_b$  se define como la *densidad normalizada*. Cuanto más próximas son las transiciones de bit ( $T_b$  pequeño), más grande será el valor de la densidad normalizada y, en consecuencia, mayor será la densidad de almacenamiento. Es posible notar que cuando aumenta  $\Delta$ , la amplitud de los picos de la señal de lectura disminuyen y además esos picos estarán desplazados en tiempo en relación a los instantes deseados. En otras palabras, los pulsos interfieren entre sí, limitando la densidad con la cual es posible escribir.

Este problema sirve como motivación para diseñar códigos de modulación que tomen la secuencia original de datos y la transformen en otra secuencia que resulte en una señal de escritura cuyas transiciones de amplitud estén suficientemente espaciadas. Por ejemplo, si se utiliza NRZI, la secuencia codificada en el modulador debe contener uno o más 0s entre los 1s.

El segundo problema encontrado en la grabación magnética es la necesidad de evitar (o minimizar) el contenido de *cc* en la señal modulada (la corriente de escritura) debido a las características de la respuesta en frecuencia del sistema de lectura y la electrónica asociada. Este requerimiento aparece también en comunicaciones digitales sobre canales cableados. Este problema puede superarse alterando (codificando) la secuencia de datos entrante al modulador. Una clase de códigos que satisfacen esos objetivos son los códigos de modulación descriptos a continuación.

**Códigos con limitación de corrida:** Los códigos que tienen una restricción sobre el número de 0s ó 1s consecutivos en una secuencia se denominan generalmente *códigos con limitación de corrida*. Esos códigos se describen genericamente mediante dos parámetros,  $d$ : el mínimo número de 0s entre dos 1s en una secuencia, y  $\kappa$ : el máximo número de 0s entre dos 1s en una secuencia. Cuando se utilizan con la modulación NRZI, el efecto de colocar  $d$  ceros entre unos sucesivos es dispersar las transiciones de forma de reducir el solapamiento en la respuesta del canal debido justamente a transiciones sucesivas. Colocar un límite superior sobre  $\kappa$  asegura que las transiciones ocurrirán con suficiente frecuencia de forma que la información del temporizado de los símbolos pueda recuperarse de la señal modulada recibida. Los códigos de longitud de corrida limitada se denominan usualmente *códigos  $(d, \kappa)$* .

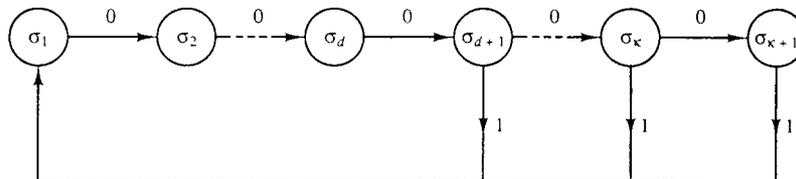


Figura 5.20: Máquina secuencial de estados finitos para un código  $(d, \kappa)$ .

Las restricciones de la secuencia de código  $(d, \kappa)$  puede representarse mediante una máquina secuencial de estados finitos con  $\kappa + 1$  estados,  $\sigma_i$ ,  $1 \leq i \leq \kappa + 1$ , como se muestra en la figura 5.20. Es posible observar que un bit de salida 0 sigue la secuencia desde el estado  $\sigma_i$  al  $\sigma_{i+1}$  con  $i \leq \kappa$ . Un bit de salida 1 sigue la secuencia hacia el estado  $\sigma_1$ . El bit de salida del codificador puede ser un 1 solo cuando la secuencia está en el estado  $\sigma_i$ ,  $d + 1 \leq i \leq \kappa + 1$ . Cuando la secuencia está en el estado  $\sigma_{\kappa+1}$  el bit de salida es siempre 1.

La máquina secuencial de estados finita puede representarse también mediante una *matriz de transición*,  $\mathbf{D}$   $(\kappa + 1) \times (\kappa + 1)$  con elementos  $d_{ij}$ , donde

$$\begin{aligned} d_{i1} &= 1, & i &\geq d + 1 \\ d_{ij} &= 1, & j &= i + 1 \\ d_{ij} &= 0, & &\text{para todo otro } i, j \end{aligned}$$

**Ejemplo:** La matriz de transición de estados para un código  $(d, \kappa) = (1, 3)$  tiene cuatro estados. De la figura 5.20 se obtiene que

$$\mathbf{D} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Un parámetro importante de cualquier código  $(d, \kappa)$  es el número de secuencias de una cierta longitud,  $n$  que satisfacen las restricciones  $(d, \kappa)$ . Cuando  $n$  aumenta, el número de secuencias  $N(n)$  que satisfacen la restricción  $(d, \kappa)$  también aumenta. El número de bits de información que puede representarse unívocamente con  $N(n)$  secuencias es

$$k = \lceil \log_2 N(n) \rceil$$

donde aquí  $\lceil x \rceil$  es el entero más grande contenido en  $x$ . La máxima velocidad del código es entonces  $R_c = k/n$ .

La capacidad de un código  $(d, \kappa)$  está definida como

$$C(d, \kappa) = \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 N(n)$$

Claramente,  $C(d, \kappa)$  es la velocidad máxima posible que puede lograrse con las restricciones  $(d, \kappa)$ . Shannon demostró que la capacidad está dada por

$$C(d, \kappa) = \log_2 \lambda_{max}$$

donde  $\lambda_{max}$  es el mayor autovalor real de la matriz de transición de estados  $\mathbf{D}$ .

**Ejemplo:** La capacidad para un código  $(d, \kappa) = (1, 3)$  puede obtenerse a partir de la matriz de transición de estados. Así

$$\det(\mathbf{D} - \lambda \mathbf{I}) = \det \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 1 & -\lambda & 1 & 0 \\ 1 & 0 & -\lambda & 1 \\ 1 & 0 & 0 & -\lambda \end{bmatrix} = \lambda^4 - \lambda^2 - \lambda - 1 = 0$$

La mayor raíz real de este polinomio es  $\lambda_{max} = 1.4656$ . En consecuencia la capacidad será  $C(1, 3) = \log_2 \lambda_{max} = 0.5515$ .

Las capacidades de  $(d, \kappa)$  para  $0 \leq d \leq 6$  y  $2 \leq \kappa \leq 15$  se muestran en la tabla 5.1. Es posible notar que  $C(d, \kappa) < 1/2$  para  $d \geq 3$  y para cualquier valor de  $\kappa$ . Los códigos más comúnmente utilizados en grabación magnética emplean  $d \leq 2$ , en consecuencia  $R_c$  es al menos  $1/2$ .

Es posible orientar la atención ahora a la construcción de algunos códigos de longitud de corrida limitada. En general, los códigos  $(d, \kappa)$  pueden construirse como códigos de longitud fija o de longitud variable. En un código de longitud fija cada bit o bloque de  $k$  bits se codifica en un bloque de  $n > k$  bits.

En principio, la construcción de un código de longitud fija es directa. Para un bloque dado de longitud  $n$ , seleccionamos un subconjunto de  $2^n$  palabras que satisfagan las restricciones de corrida especificadas. De este subconjunto eliminamos las palabras de código que no satisfacen las restricciones de longitud de corrida cuando se colocan en secuencia. De este forma, se obtiene un conjunto de palabras de código que satisfacen las restricciones y pueden utilizarse en el mapeo de los bits de entrada al codificador. Las operaciones de codificación y decodificación pueden realizarse mediante una tabla.

k	d=0	d=1	d=2	d=3	d=4	d=5	d=6
2	0.8791	0.4057					
3	0.9468	0.5515	0.2878				
4	0.9752	0.6174	0.4057	0.2232			
5	0.9881	0.6509	0.4650	0.3218	0.1823		
6	0.9942	0.6690	0.4979	0.3746	0.2269	0.1542	
7	0.9971	0.6793	0.5174	0.4057	0.3142	0.2281	0.1335
8	0.9986	0.6853	0.5293	0.4251	0.3432	0.2709	0.1993
9	0.9993	0.6888	0.5369	0.4376	0.3620	0.2979	0.2382
10	0.9996	0.6909	0.5418	0.4460	0.3746	0.3158	0.2633
11	0.9998	0.6922	0.5450	0.4516	0.3833	0.3285	0.2804
12	0.9999	0.6930	0.5471	0.4555	0.3894	0.3369	0.2924
13	0.9999	0.6935	0.5485	0.4583	0.3937	0.3432	0.3011
14	0.9999	0.6938	0.5495	0.4602	0.3968	0.3478	0.3074
15	0.9999	0.6939	0.5501	0.4615	0.3991	0.3513	0.3122
$\infty$	1.0000	0.6942	0.5515	0.4650	0.4057	0.3620	0.3282

Tabla 5.1: Capacidad  $C(d, \kappa)$  en función de los parámetros  $d$  y  $\kappa$ .

**Ejemplo:** Se construirá un código  $(0, 2)$  de longitud 3 y se determinará su eficiencia. A partir del listado de todas las palabras de código es posible hallar que las palabras que satisfacen la restricción (0.2) son:  $(0\ 1\ 0)$ ,  $(0\ 1\ 1)$ ,  $(1\ 0\ 1)$ ,  $(1\ 1\ 0)$ , y  $(1\ 1\ 1)$ . Es posible utilizar cuatro de esas palabras para codificar los siguientes pares de bits:  $(00, 01, 10, 11)$ . De esta forma la velocidad será  $k/n = 2/3$  que satisface la restricción  $(0, 2)$ . El código de longitud fija de este ejemplo no es muy eficiente. La capacidad es  $C(0, 2) = 0.8791$ , de forma que la *eficiencia* del código es

$$\text{eficiencia} = \frac{R_c}{C(d, \kappa)} = \frac{2/3}{0.8791} = 0.76$$

Mejores códigos  $(0, 2)$  pueden construirse incrementando la longitud del bloque  $n$ .

En el ejemplo siguiente no se impondrán restricciones en la longitud máxima  $\kappa$ .

**Ejemplo:** Se construirá un código con  $d = 1$  y  $\kappa = \infty$  de longitud  $n = 5$ . En ese caso no se impone restricción sobre el número de ceros consecutivos. Para construir el código se selecciona del conjunto de 32 palabras de código posibles aquellas que satisfacen la restricción  $d = 1$ . Existen ocho de tales palabras, lo cual implica que es posible codificar tres bits de información con cada palabra de código. El código se muestra en la tabla 5.2. Notar que el primer bit de cada palabra de código es un 0, mientras que el último bit puede ser un 0 o un 1. En consecuencia, la restricción  $d = 1$  es satisfecha cuando esas palabras de código están concatenadas. Este código tiene una velocidad  $R_c = 3/5$ . Cuando se compara con la capacidad  $C(1, \infty) = 0.6942$  obtenida de la tabla 5.1, la eficiencia del código es 0.864, la cual es totalmente aceptable.

El método de construcción del código descrito en los ejemplos anteriores produce códigos  $(d, \kappa)$  de longitud fija *independientes de los estados*. Esto significa que las palabras de código de longitud fija pueden concatenarse sin violar la restricción  $(d, \kappa)$ . En general, los códigos  $(d, \kappa)$  de estados independientes y longitud fija requieren longitudes de bloque grandes, excepto como en los casos de los ejemplos anteriores donde  $d$  es pequeño. Códigos de longitudes menores son posibles permitiendo la dependencia de los estados y palabras de código de longitud variable. Más abajo se discutirán códigos para los cuales los bloques de entrada y salida del codificador tienen longitud variable. Para que las palabras del código sean unívocamente decodificables en el receptor, el código de longitud variable debe satisfacer la condición de prefijo, discutida en el Capítulo

Bits entrada	Secuencia salida
000	00000
001	00001
010	00010
011	00100
100	00101
101	01000
110	01001
111	01010

Tabla 5.2: Código de longitud fija,  $d = 1$ ,  $\kappa = \infty$ .

Bits entrada	Secuencia salida
10	1000
11	0100
011	000100
010	001000
000	100100
0011	00100100
0010	00001000

Tabla 5.3: Código de longitud variable,  $(2, 7)$ .

anterior.

---

**Ejemplo:** Se construirá un código de longitud variable con  $d = 0$  y  $\kappa = 2$ . Un código decodificable muy simple para la restricción  $(0, 2)$  es el siguiente:

$$\begin{aligned} 0 &\rightarrow 01 \\ 10 &\rightarrow 10 \\ 11 &\rightarrow 11 \end{aligned}$$

---

El código en el ejemplo anterior tiene un tamaño de bloque de salida fijo pero un tamaño de bloque de entrada variable. En general, ambos bloques de entrada y salida del codificador pueden ser de longitud variable. El siguiente ejemplo ilustra este último caso.

---

**Ejemplo:** Se construirá un código de tamaño de bloque variable de restricción  $(2, 7)$ . La solución a este problema de construcción de código es ciertamente no trivial y no única. Se propuso este ejemplo ya que es ampliamente utilizado por IBM en varios sistemas de almacenamiento de discos. El código se lista en la tabla 5.3. Es posible observar que bloques de entrada de 2, 3, y 4 bits se mapean en bloques de salida de 4, 6, y 8 bits, respectivamente. De esta forma, la velocidad del código es  $R_c = 1/2$ . Dado que esta es la velocidad para todas las palabras del código, este código se denomina un *código de velocidad fija*. La eficiencia es  $0.5/0.5174 = 0.966$ . Notar, además que este código satisface la condición de prefijo.

---

Otro código que ha sido ampliamente utilizado en grabación magnética es el código  $(d, \kappa) = (1, 3)$  de velocidad  $1/2$  mostrado en la tabla 5.4. Es posible observar que cuando el bit de información es un 0, el primer bit de salida es 1 si el bit previo de entrada fué un 0, o un 0 si el bit previo de entrada fué un 1.

Bits entrada	Secuencia salida
0	$x0$
1	01

Tabla 5.4: Codificador (1, 3) para el código Miller.

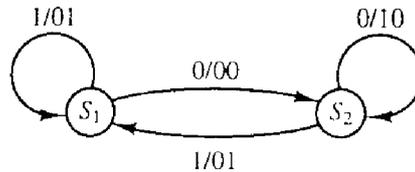


Figura 5.21: Diagrama de estados para un código Miller de  $d = 1$  y  $\kappa = 3$ .

Cuando el bit de información es un 1, la salida del codificador es 01. La decodificación es simple. El primer del bloque de dos bits es redundante y puede descartarse. El segundo es el bit de información. Este código se denomina usualmente *código Miller*. Es posible observar que es un código dependiente de los estados, lo cual se describe mediante el diagrama de estados de la figura 5.21. Existen dos estados  $S_1$  y  $S_2$  con las transiciones mostradas en la figura. Cuando el codificador está en el estado  $S_1$ , un bit de entrada 1 al codificador lo mantiene en  $S_1$  y la salida será 01. Esto se escribe con 1/01. Si el bit de entrada es un 0, el codificador entra al estado  $S_2$  y las salidas serán 00, con la notación 0/00. En forma similar, si el codificador está en  $S_2$ , un bit de entrada 0 no causa transición y la salida es 10. Por otro lado, si el bit de entrada es un 1, el codificador entra al estado  $S_1$  y las salidas serán 01.

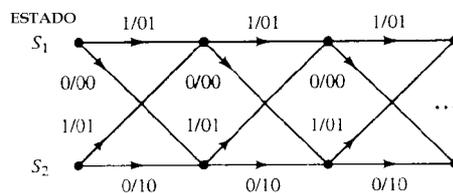


Figura 5.22: Trellis para el código Miller (1, 3).

**Representación trellis de códigos  $(d, \kappa)$  dependientes de los estados** El diagrama de estados provee una representación relativamente compacta de un código dependiente de los estados. Otra forma de describir tales códigos con memoria es por medio de un gráfico denominado *trellis*. Una trellis es un gráfico que ilustra las transiciones de estado como una función del tiempo. El gráfico consiste en un conjunto de nodos representando los estados que caracterizan la memoria en el código en diferentes instantes de tiempo e interconexiones entre pares de nodos que indican las transiciones entre instantes sucesivos de tiempo. Por ejemplo, la figura 5.22 muestra la trellis para un código Miller con  $d = 1$ ,  $\kappa = 3$ , cuyo diagrama de estados se muestra en la figura 5.21.

**Mapeo de los bits codificados en señales** La secuencia de salida de un codificador  $(d, \kappa)$  se mapea en una señal mediante el modulador para después ser transmitida por el canal. Si un dígito binario 1 se mapea en un pulso rectangular de amplitud  $A$  y un dígito binario 0 se mapea en un pulso rectangular de amplitud  $-A$ , el resultado es una señal modulada NRZ codificada  $(d, \kappa)$ . La duración de los pulsos rectangulares es  $T_c = R_c/R_b = R_c T_b$ , donde  $R_b$  es la velocidad de información entrante al codificador,  $T_b$  es el intervalo de bit correspondiente y  $R_c$  es la velocidad del código  $(d, \kappa)$ .

Cuando el código  $(d, \kappa)$  es de longitud fija e independiente de los estados con velocidad  $R_c = k/n$ , es posible considerar que cada bloque de  $n$  bits genera una señal de duración  $nT_c$ . De esta forma, se tendrán  $M = 2^k$  señales, una de las  $2^k$  posibles de bloques de datos de  $k$  bits. Esas señales codificadas tienen la forma general dada por (5.5) y (5.6). En este caso no existe dependencia entre la transmisión de señales

sucesivas.

En contraste a la situación considerada, cuando se utiliza NRZI y/o el código  $(d, \kappa)$  es dependiente de los estados la modulación de señal tiene memoria. Consideraremos primero el efecto de mapear los bits de código en una señal NRZI.

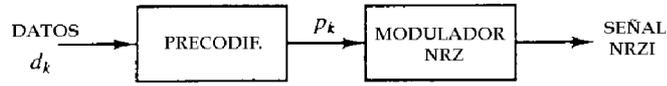


Figura 5.23: Método para generar una señal NRZI usando precodificación.

Una señal modulada NRZI es en sí misma dependiente de los estados. El nivel de amplitud de la señal cambia de su valor actual ( $\pm A$ ) solamente cuando el bit a ser transmitido es un 1. Es interesante notar que la señal NRZI puede verse como una señal NRZ precedida por otra operación de codificación, denominada *precodificación* de una secuencia binaria, como mostrado en la figura 5.23. La operación de precodificación se puede describir mediante

$$p_k = d_k \oplus p_{k-1}$$

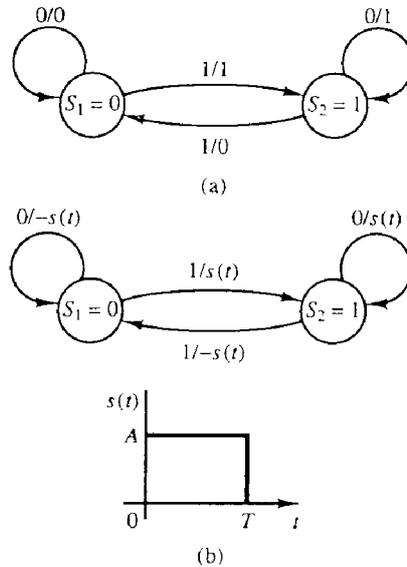


Figura 5.24: Diagram de estados para una señal NRZI.

donde  $\{d_k\}$  es la secuencia binaria entrante al precodificador,  $\{p_k\}$  es la secuencia binaria de salida del precodificador, y  $\oplus$  es la suma modulo 2. Este tipo de codificación se denomina *codificación diferencial*, y está caracterizada por el diagrama de estados mostrado en la figura 5.24a. Luego, la secuencia  $\{p_k\}$  se transmite mediante NRZ. De esta forma, cuando  $p_k = 1$ , la salida del modulador es un pulso rectangular de amplitud  $A$ , y cuando  $p_k = 0$  la salida del modulador es un pulso rectangular de amplitud  $-A$ . Cuando las señales se superponen sobre el diagrama de estados de la figura 5.24a se obtiene el diagrama correspondiente al de la figura 5.24b.

Del diagrama de estados es posible concluir que la codificación diferencial o precodificación introduce memoria en la señal modulada. De la misma forma que en el caso de códigos  $(d, \kappa)$  dependientes de los estados puede usarse un diagrama trellis para ilustrar la dependencia del tiempo de la señal modulada. El diagrama trellis para una señal NRZI se muestra en la figura 5.25. Cuando la salida del codificador  $(d, \kappa)$  dependiente de los estados es seguida por un modulador NRZI es posible combinar los dos diagramas de

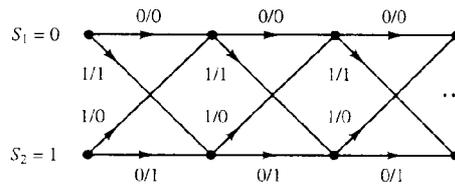


Figura 5.25: Trellis para una señal NRZI.

estado. Una combinación similar puede realizarse con las trellis correspondientes. El ejemplo siguiente ilustra la construcción de un código Miller (1,3) seguido de modulación NRZI.

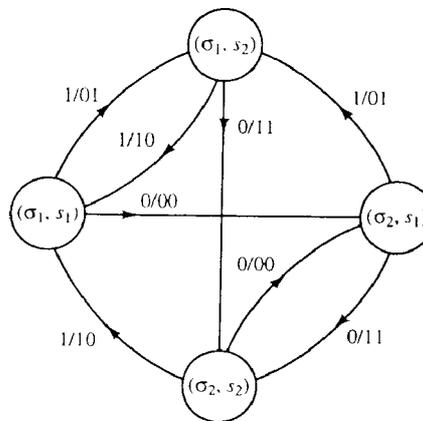


Figura 5.26: Diagrama de estados de un código Miller seguido de un precodificador.

---

**Ejemplo:** Determinar el diagrama de estados de la combinación del código Miller (1,3) seguido por la precodificación inherente a la modulación NRZI. Dado que el código Miller (1,3) tiene dos estados y el precodificador tiene dos estados, el diagrama de estados para el codificador combinado tendrá cuatro estados,  $(S_M, S_N) = (\sigma_1, s_1), (\sigma_1, s_2), (\sigma_2, s_1), (\sigma_2, s_2)$ , donde  $S_M = \{\sigma_1, \sigma_2\}$  representa los dos estados del código Miller y  $S_N = \{s_1, s_2\}$  representan los dos estados del precodificador para NRZI. Por cada bit de entrada al codificador Miller se obtienen dos bits de salida, los cuales luego son precodificados para producir dos bits de salida precodificados. El diagrama de estados resultante se muestra en la figura 5.26, donde el primer bit es el bit de información al codificador Miller y los dos bits siguientes representan la salida correspondiente del precodificador.

---

El diagrama trellis para la secuencia precodificada de Miller puede obtenerse directamente de los diagramas de estados combinados o de una combinación de las trellis de los dos códigos. El resultado de esta combinación es una trellis de cuatro estados, de la cual una etapa se muestra en la figura 5.27.

Es interesante notar que las cuatro señales obtenidas mapeando cada par de bits de la secuencia precodificada de Miller en una señal NRZ son biortogonales. En particular, el par de bits 11 se mapea en la señal  $s_1(t)$  y los bits 01 se mapean en  $s_2(t)$ , como se muestra en la figura 5.28. Luego, los bits 00 en  $-s_1(t)$  y los bits 10 en  $-s_2(t)$ . Dado que  $s_1(t)$  y  $s_2(t)$  son ortogonales, el conjunto de las cuatro señales constituye un conjunto biortogonal de  $M = 4$  señales. En consecuencia, en el diagrama de estados para una secuencia precodificada de Miller es posible sustituir los bits de salida del codificador por las señales del modulador. Este diagrama de estados se ilustra en la figura 5.29, donde los cuatro estados son  $S_i, 1 \leq i \leq 4$ . La señal modulada resultante también se denomina *modulación de retardo*.

Las señales moduladas con memoria tal como NRZI y Miller codificada-NRZI (modulación de retardo) se caracterizan generalmente por una cadena de Markov de  $K$  estados con *probabilidades de estado esta-*

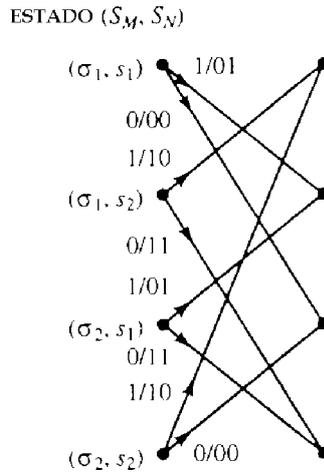


Figura 5.27: Una etapa del diagrama trellis para el código Miller seguido por el precodificador.

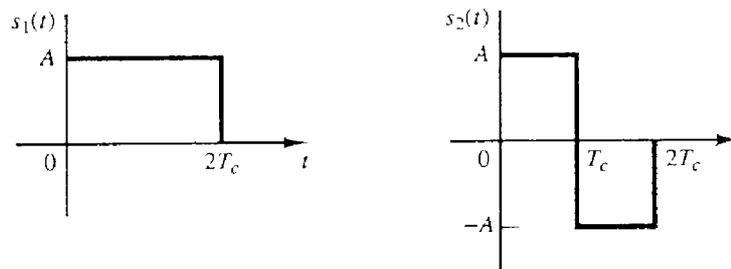


Figura 5.28: Formas de onda para pares de bits Miller codificados.

cionarias  $\{p_i, i = 1, 2, \dots, K\}$ .  $s_j(t)$ , para  $j = 1, 2, \dots, K$ , es la señal asociada con cada transición.  $p_{ij}$  es la probabilidad de que  $s_j(t)$  sea transmitida en un intervalo de señalización determinado después de la transmisión de la señal  $s_i(t)$  en el intervalo de señalización previo. Las probabilidades de transición pueden ordenarse en forma matricial como

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{bmatrix}$$

donde  $\mathbf{P}$  es la *matriz de probabilidades de transición*. Esta matriz se obtiene fácilmente del diagrama de estados y las probabilidades de ocurrencia correspondientes a los bits de entrada (o equivalentemente, las probabilidades de estado estacionario  $\{p_i\}$ ).

Para la señal NRZI con igual probabilidades de estado  $p_1 = p_2 = 1/2$  y el diagrama de estados mostrado en la figura 5.24, la matriz de transición de estados es

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

En forma similar, la matriz de probabilidades de transición para una señal modulada NRZI y codificación Miller con símbolos equiprobables ( $p_1 = p_2 = p_3 = p_4 = 1/4$ ) es

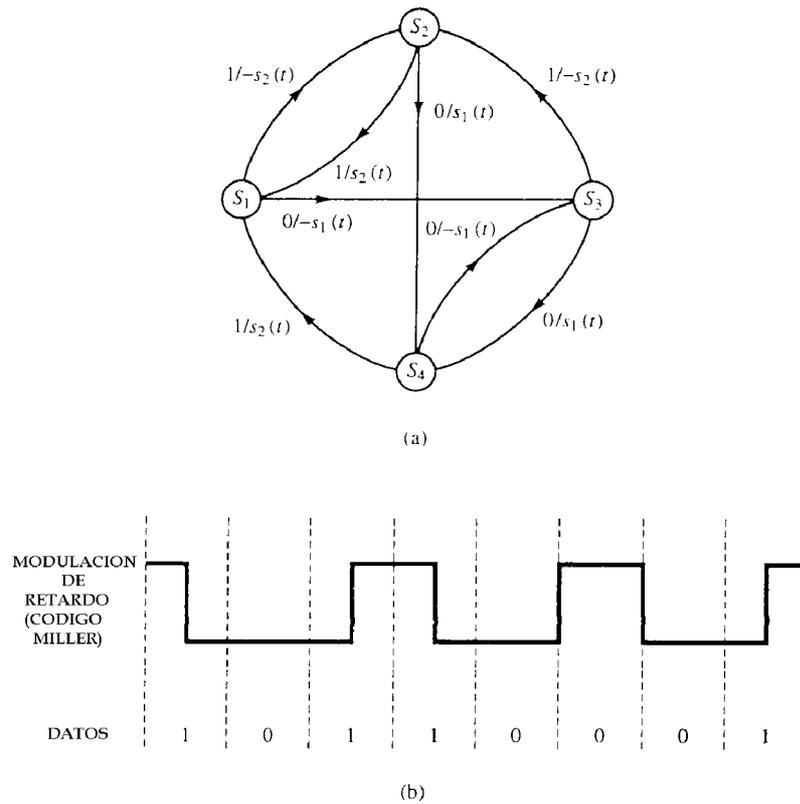


Figura 5.29: Diagrama de estados para la señal Miller precodificada y forma de onda.

$$P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

Como se discutirá en el próximo Capítulo, la matriz de probabilidades de transición es útil para determinar la densidad espectral de potencia de una señal modulada con memoria.

**Códigos de modulación para el sistema de disco compacto.** Varios factores deben ser considerados en el diseño de un código de modulación para el sistema de disco compacto. Una restricción es que la máxima longitud de corrida de ceros debe ser suficientemente pequeña para permitir al sistema sincronizarse a partir de la señal de lectura. Para satisfacer esta restricción se elige  $\kappa = 10$ . Un segundo factor es el contenido de frecuencia de la señal modulada debajo de 20 kHz. En un sistema de CD, los sistemas de servocontrol que mantienen el haz del laser en una pista específica y los circuitos de enfoque son controlados por señales en el rango de 0 a 20 kHz. Para evitar la interferencia con esas señales de control, la información modulada debe estar por encima de ese rango de frecuencias. El código de corrida limitada que ha sido implementado para ese propósito es un código  $(d, \kappa) = (2, 10)$  que resulta en una señal que transporta la información que ocupa el rango de frecuencias de 20 kHz a 1.5 MHz.

El código  $(2, 10)$  seleccionado es de longitud fija y codifica ocho bits de información en 14 bits codificados. Dado que cada muestra de la señal de audio se cuantiza en 16 bits, una muestra de 16 bits se divide en dos bytes y se codifica. Listando todas las palabras de código de 14 bits es posible mostrar que existen 267 palabras codificadas diferentes que satisfacen la restricción  $(d, \kappa)$ . De esas, 256 palabras de código se seleccionan para formar el código. Sin embargo, las restricciones de  $(d, \kappa)$  no son satisfechas cuando las palabras del código se concatenan. Para remediar este problema se incluyen tres bits adicionales a los 14, denominados bits de "mezclado". Los tres bits de mezclado sirven para dos propósitos. Primero, si la restricción  $d$  no es satisfecha en la concatenación, los bits de mezclado se eligen como 0s. Por otro lado, si



## 5.2 Receptor óptimo para modulación por pulsos en ruido blanco gaussiano

Es posible considerar un sistema de comunicaciones que transmite información digital mediante el uso de cualquiera de las señales  $M$ -árias descritas en la sección precedente. De esta forma, la secuencia de entrada al modulador se subdivide en  $k$  bloques o símbolos y cada uno de los  $M = 2^k$  símbolos se asocia a la señal correspondiente del conjunto  $\{s_m(t), m = 1, 2, \dots, M\}$ . Cada señal se transmite dentro del intervalo de símbolo de duración  $T$ . Para ser específico, es posible considerar la transmisión de información sobre el intervalo  $0 \leq t \leq T$ .

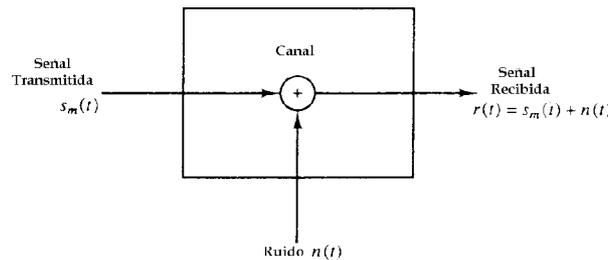


Figura 5.31: Modelo de señal recibida a través de un canal AWGN.

Se supone que el canal corrompe la señal a través de un ruido aditivo blanco Gaussiano (AWGN) como mostrado en la figura 5.31. De esta forma, la señal recibida en el intervalo  $0 \leq t \leq T$  puede expresarse como

$$r(t) = s_m(t) + n(t), \quad 0 \leq t \leq T$$

donde  $n(t)$  es la función muestra del ruido AWGN con densidad espectral de potencia  $S_n(f) = N_0/2$  W/Hz. En base a la observación de  $r(t)$  sobre el intervalo de señal, se desea diseñar un receptor que sea óptimo en el sentido de minimizar la probabilidad de error de detección.

Es conveniente subdividir al receptor en dos partes, el demodulador y el detector. La función del demodulador es convertir la señal recibida  $r(t)$  en un vector  $N$  dimensional  $\mathbf{r} = [r_1, r_2, \dots, r_N]$ , donde  $N$  es la dimensión de las señales transmitidas. La función del detector es decidir cual de las  $M$  posibles señales fue transmitida basandose en la observación de  $\mathbf{r}$ .

En las próximas secciones se describirán dos realizaciones del demodulador. Una se basa en el uso de correladores. La segunda se basa en el uso de filtros acoplados. El detector óptimo que sigue al demodulador se diseña para minimizar la probabilidad de error.

### 5.2.1 Demodulador tipo correlador

Se describirá en esta sección un demodulador de tipo correlador que descompone la señal recibida y el ruido en vectores  $N$  dimensionales. En otras palabras, la señal y el ruido se expanden en una serie de funciones base linealmente ponderadas  $\{\psi_n(t)\}$ . Se supone que las  $N$  funciones base  $\{\psi_n(t)\}$  barren el espacio de señal tal que cada una de las posibles señales transmitidas del conjunto  $\{s_m(t)\}$ ,  $1 \leq m \leq M$  pueden representarse como una combinación lineal de las  $\{\psi_n(t)\}$ . Las funciones  $\{\psi_n(t)\}$  no barren el espacio de ruido. Sin embargo, se mostrará más abajo que los términos de ruido que caen fuera del espacio de señales es irrelevante en lo relativo al problema de detección.

Es posible suponer que la señal recibida  $r(t)$  pasa a través de un banco de correladores en paralelo el cual calcula la proyección de  $r(t)$  sobre las  $N$  funciones base  $\{\psi_n(t)\}$ , como ilustrado en la figura 5.32. De esta forma se tiene que

$$\begin{aligned} \int_0^T r(t)\psi_k(t)dt &= \int_0^T [s_m(t) + n(t)]\psi_k(t)dt \\ r_k &= s_{mk} + n_k, \quad k = 1, 2, \dots, N \end{aligned}$$

donde

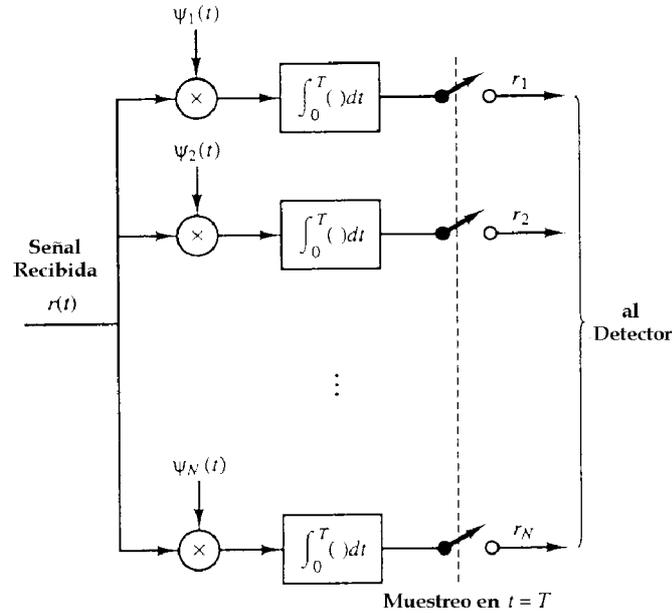


Figura 5.32: Demodulador tipo correlador.

$$s_{mk} = \int_0^T s_m(t) \psi_k(t) dt, \quad k = 1, 2, \dots, N$$

$$n_k = \int_0^T n(t) \psi_k(t) dt, \quad k = 1, 2, \dots, N$$

La señal se representa ahora por el vector  $\mathbf{s}_m$  con componentes  $s_{mk}$ ,  $k = 1, 2, \dots, N$ . Sus valores dependen de cual de las  $M$  señales fué transmitida. Las componentes  $\{n_k\}$  son variables aleatorias que aparecen debido a la presencia del ruido aditivo.

En realidad, es posible expresar la señal recibida  $r(t)$  en el intervalo  $0 \leq t \leq T$  como

$$r(t) = \sum_{k=1}^N s_{mk} \psi_k(t) + \sum_{k=1}^N n_k \psi_k(t) + n'(t)$$

$$= \sum_{k=1}^N r_k \psi_k(t) + n'(t)$$

El término  $n'(t)$ , definido como

$$n'(t) = n(t) - \sum_{k=1}^N n_k \psi_k(t)$$

es un proceso de ruido Gaussiano que representa la diferencia entre el proceso de ruido original  $n(t)$  y la parte que corresponde a la proyección de  $n(t)$  sobre las funciones base  $\{\psi_k(t)\}$ . Mostraremos más adelante que  $n'(t)$  es irrelevante para la decisión de cual señal fué transmitida. Consecuentemente, la decisión puede basarse enteramente en la señal de salida del correlador y la componente de ruido asociada  $r_k = s_{mk} + n_k$ ,  $k = 1, 2, \dots, N$ .

Dado que las señales  $\{s_m(t)\}$  son determinísticas, las señales componentes serán también determinísticas. Las componentes de ruido  $\{n_k\}$  son Gaussianas. Sus valores medios son

$$E[n_k] = \int_0^T E[n(t)]\psi_k(t)dt = 0$$

para todo  $k$ . Sus covarianzas son

$$\begin{aligned} E[n_k n_m] &= \int_0^T \int_0^T E[n(t)n(\tau)]\psi_k(t)\psi_m(\tau)dtd\tau \\ &= \frac{N_0}{2} \int_0^T \int_0^T \delta(t-\tau)\psi_k(t)\psi_m(\tau)dtd\tau \\ &= \frac{N_0}{2} \int_0^T \psi_k(t)\psi_m(\tau)dt \\ &= \frac{N_0}{2} \delta_{mk} \end{aligned}$$

donde  $\delta_{mk} = 1$  cuando  $m = k$  y cero para todo otro valor. En consecuencia, las  $N$  componentes  $\{n_k\}$  son variables aleatorias Gaussianas no correlacionadas de media cero y varianza común  $\sigma_n^2 = N_0/2$ .

Del desarrollo anterior se concluye que la salida del correlador  $\{r_k\}$  asociada a la  $m$ -ésima señal transmitida es una variable aleatoria Gaussiana de media

$$E[r_k] = E[s_{mk} + n_k] = s_{mk}$$

e igual varianza

$$\sigma_r^2 = \sigma_n^2 = N_0/2$$

Dado que las componentes de ruido  $\{n_k\}$  son variables aleatorias AWGN también son estadísticamente independientes. En consecuencia, la salida del correlador  $\{r_k\}$  condicionada por la  $m$ -ésima señal transmitida es una variable aleatoria Gaussiana estadísticamente independiente. Además, las funciones densidad de probabilidad condicionales de las variables aleatorias  $\mathbf{r} = [r_1, r_2, \dots, r_N]$  son simplemente

$$f(\mathbf{r}|\mathbf{s}_m) = \prod_{k=1}^N f(r_k|s_{mk}), \quad m = 1, 2, \dots, M$$

donde

$$f(r_k|s_{mk}) = \frac{1}{\sqrt{\pi N_0}} e^{-(r_k - s_{mk})^2 / N_0}, \quad k = 1, 2, \dots, M$$

Sustituyendo, es posible obtener la pdf condicional conjunta como

$$f(\mathbf{r}|\mathbf{s}_m) = \frac{1}{(\pi N_0)^{N/2}} e^{-\sum_{k=1}^N (r_k - s_{mk})^2 / N_0}, \quad m = 1, 2, \dots, M \quad (5.7)$$

Como aspecto final, es posible mostrar que las salidas de los correladores  $\{r_N\}$  son una *estadística suficiente* para alcanzar una decisión sobre cual de las  $M$  señales fué transmitida, o sea, no se extrae información relevante adicional del proceso remanente  $n'(t)$ . Además,  $n'(t)$  no está correlacionada con las salidas de los correladores, o sea

$$\begin{aligned} E[n'(t)r_k] &= E[n'(t)]s_{mk} + E[n'(t)n_k] \\ &= E[n'(t)n_k] \end{aligned}$$

$$\begin{aligned}
&= E \left\{ \left[ n(t) - \sum_{j=1}^N n_j \psi_j(t) \right] n_k \right\} \\
&= \int_0^T E[n(t)n(\tau)] \psi_k(\tau) d\tau - \sum_{j=1}^N E[n_j n_k] \psi_j(t) \\
&= \frac{N_0}{2} \psi_k(t) - \frac{N_0}{2} \psi_k(t) = 0
\end{aligned}$$

Dado que  $n'(t)$  y  $\{r_k\}$  son Gaussianas y no están correlacionadas, son estadísticamente independientes. En consecuencia,  $n'(t)$  no brinda ninguna información que sea relevante en la decisión de cual señal fué transmitida. Toda la información relevante está contenida en  $\{r_k\}$ . En consecuencia,  $n'(t)$  puede ser ignorada.

**Ejemplo:** Es posible considerar un conjunto de señales PAM  $M$ -árias en las cuales la forma básica del pulso  $g_T(t)$  es rectangular, de amplitud  $a$  y duración  $T$ . El ruido es AWGN de media cero. Se trata de determinar la función base  $\psi(t)$  y la salida de un demodulador tipo correlador.

La energía del pulso rectangular es

$$\mathcal{E}_g = \int_0^T g_T^2(t) dt = \int_0^T a^2 dt = a^2 T$$

Dado que el conjunto de señales PAM tiene dimensión  $n = 1$ , existe una sola función base  $\psi(t)$ . Esta puede obtenerse de (5.4) y está dada por

$$\psi(t) = \frac{1}{\sqrt{a^2 T}} g_T(t) = \begin{cases} \frac{1}{\sqrt{T}}, & 0 \leq t \leq T \\ 0, & \text{para todo otro } t \end{cases}$$

La salida del demodulador tipo correlador es

$$r = \int_0^T r(t) \psi(t) dt = \frac{1}{\sqrt{T}} \int_0^T r(t) dt$$

Es interesante notar que el correlador se convierte en un integrador cuando  $\psi(t)$  es rectangular. Si sustituimos  $r(t)$  en la ecuación anterior se obtiene

$$\begin{aligned}
r &= \frac{1}{\sqrt{T}} \left[ \int_0^T [s_m(t) + n(t)] dt \right] \\
&= \frac{1}{\sqrt{T}} \left[ \int_0^T s_m \psi(t) dt + \int_0^T n(t) dt \right] \\
&= s_m + n
\end{aligned}$$

donde para el término de ruido  $E[n] = 0$  y

$$\begin{aligned}
\sigma_n^2 &= E \left[ \frac{1}{T} \int_0^T \int_0^T n(t) n(\tau) dt d\tau \right] \\
&= \frac{1}{T} \int_0^T \int_0^T E[n(t) n(\tau)] dt d\tau \\
&= \frac{N_0}{T} \int_0^T \int_0^T \delta(t - \tau) dt d\tau = \frac{N_0}{T}
\end{aligned}$$

La función densidad de probabilidad para la salida muestreada es

$$f(r|s_m) = \frac{1}{\sqrt{\pi N_0}} e^{-(r-s_m)^2/N_0}$$

### 5.2.2 Demodulador tipo filtro acoplado

En lugar de usar un banco de  $N$  correladores para generar las variables  $\{r_k\}$  es posible utilizar un banco de  $N$  filtros lineales. Para ser específico, es posible suponer que las respuestas impulsivas de los  $N$  filtros son

$$h_k(t) = \psi_k(T-t), \quad 0 \leq t \leq T$$

donde  $\{\psi_k(t)\}$  son las  $N$  funciones base y  $h_k(t) = 0$  fuera del intervalo  $0 \leq t \leq T$ . Las salidas de esos filtros son

$$\begin{aligned} y_k(t) &= \int_0^t r(\tau) h_k(t-\tau) d\tau \\ &= \int_0^t r(\tau) \psi_k(T-t+\tau) d\tau, \quad k = 1, 2, \dots, N \end{aligned}$$

Si se muestrean las salidas de los filtros en  $t = T$  se obtiene

$$y_k(T) = \int_0^T r(\tau) \psi_k(\tau) d\tau = r_k, \quad k = 1, 2, \dots, N$$

En consecuencia, las salidas muestreadas de los filtros en el instante  $t = T$  son exactamente el conjunto de valores  $\{r_k\}$  obtenidos de los  $N$  correladores lineales.

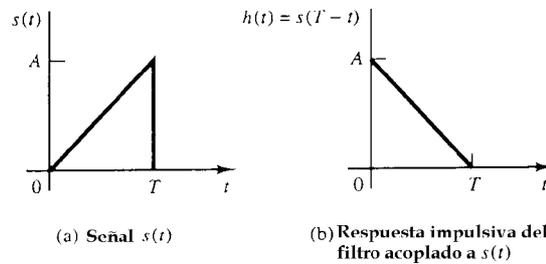


Figura 5.33: Señal  $s(t)$  y filtro acoplado a  $s(t)$ .

Un filtro cuya respuesta impulsiva  $h(t) = s(T-t)$ , donde  $s(t)$  se supone confinada al intervalo  $0 \leq t \leq T$ , se denomina el *filtro acoplado* a la señal  $s(t)$ . Un ejemplo de una señal y su filtro acoplado se muestra en la figura 5.33. La respuesta de  $h(t) = s(T-t)$  a la señal  $s(t)$  es

$$y(t) = \int_0^t s(\tau) s(T-t+\tau) d\tau$$

la cual es básicamente la función autocorrelación temporal de la señal  $s(t)$ . La figura 5.34 ilustra  $y(t)$  para el pulso triangular mostrado en la figura 5.33. Es posible notar que la función autocorrelación de  $y(t)$  es una función par de  $t$  que alcanza un valor máximo en  $\tau = T$ .

En el caso del demodulador descrito antes, los  $N$  filtros están acoplados a las funciones base  $\{\psi(t)\}$ . La figura 5.35 ilustra el demodulador tipo filtro acoplado que genera las variables observadas  $\{r_k\}$ .

**Propiedades del filtro acoplado** Un filtro acoplado tiene propiedades interesantes. Se demostrará la propiedad más importante, la cual puede especificarse como sigue: Si una señal  $s(t)$  se corrompe mediante AWGN, el filtro con respuesta impulsiva acoplada a  $s(t)$  maximiza la relación señal - ruido (SNR) a la salida.

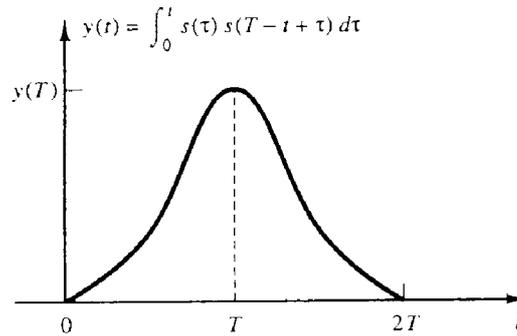


Figura 5.34: La salida del filtro acoplado es la autocorrelación de  $s(t)$ .

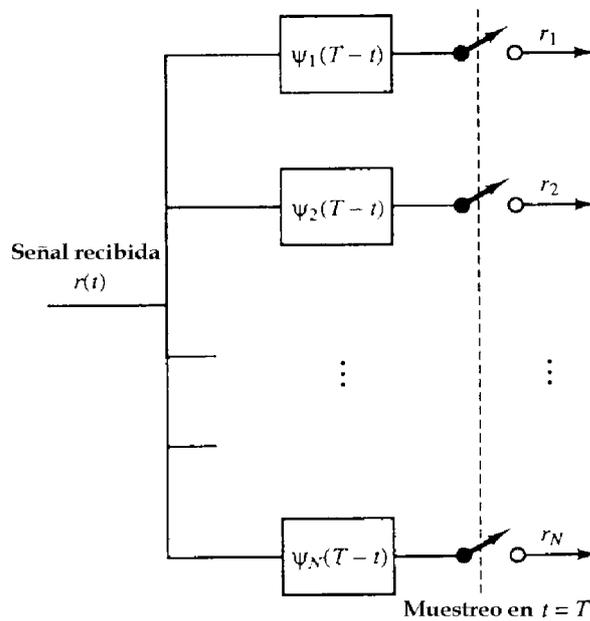


Figura 5.35: Demodulador tipo filtro acoplado.

Para probar esta propiedad es posible suponer que la señal recibida  $r(t)$  consiste de la señal  $s(t)$  y  $n(t)$  AWGN con media cero y  $S_n(f) = N_0/2$  W/Hz. Es posible suponer también que la señal  $r(t)$  pasa por un filtro cuya respuesta al impulso es  $h(t)$ ,  $0 \leq t \leq T$ , y su salida se muestrea en el instante  $t = T$ . La respuesta del filtro a las componentes de señal y ruido será

$$\begin{aligned}
 y(t) &= \int_0^t r(\tau)h(t-\tau)d\tau \\
 &= \int_0^t s(\tau)h(t-\tau)d\tau + \int_0^t n(\tau)h(t-\tau)d\tau
 \end{aligned}$$

En el instante de muestreo  $t = T$ , las componentes de señal y ruido serán

$$\begin{aligned}
 y(T) &= \int_0^T s(\tau)h(T-\tau)d\tau + \int_0^T n(\tau)h(T-\tau)d\tau \\
 &= y_s(T) + y_n(T)
 \end{aligned}$$

donde  $y_s(T)$  representa la componente de señal y  $y_n(T)$  representa la componente de ruido. El problema es seleccionar la respuesta impulsiva del filtro de forma que maximice la SNR definida como

$$\left(\frac{S}{N}\right)_o = \frac{y_s^2(T)}{E[y_n^2(T)]} \quad (5.8)$$

El denominador de (5.8) es simplemente la varianza del término de ruido a la salida del filtro. Es posible evaluar  $E[y_n^2(T)]$  como sigue

$$\begin{aligned} E[y_n^2(T)] &= \int_0^T \int_0^T E[n(\tau)n(t)]h(T-\tau)h(T-t)dtd\tau \\ &= \frac{N_0}{2} \int_0^T \int_0^T \delta(t-\tau)h(T-\tau)h(T-t)dtd\tau \\ &= \frac{N_0}{2} \int_0^T h^2(T-t)dt \end{aligned}$$

Notar que la varianza depende de la densidad espectral de potencia del ruido y la energía de la respuesta impulsiva  $h(t)$ .

Sustituyendo  $y_s(T)$  e  $E[y_n^2(T)]$  en (5.8) se obtiene la siguiente expresión para la SNR de salida

$$\left(\frac{S}{N}\right)_o = \frac{\left[\int_0^T s(\tau)h(T-\tau)d\tau\right]^2}{\frac{N_0}{2} \int_0^T h^2(T-t)dt} = \frac{\left[\int_0^T h(\tau)s(T-\tau)d\tau\right]^2}{\frac{N_0}{2} \int_0^T h^2(T-t)dt}$$

Dado que el denominador de la SNR depende de la energía en  $h(t)$ , la máxima SNR sobre  $h(t)$  se obtiene maximizando el numerador de  $(S/N)_o$  sujeta a la restricción que el denominador se mantenga constante. La maximización del numerador puede realizarse más fácilmente mediante la desigualdad de Cauchy-Schwarz, la cual especifica en general que si  $g_1(t)$  y  $g_2(t)$  son señales de energía finita, entonces

$$\left[\int_{-\infty}^{\infty} g_1(t)g_2(t)dt\right]^2 \leq \int_{-\infty}^{\infty} g_1^2(t)dt \int_{-\infty}^{\infty} g_2^2(t)dt$$

donde la igualdad es válida cuando  $g_1(t) = Cg_2(t)$ , para cualquier constante arbitraria  $C$ . Haciendo  $g_1(t) = h(t)$  y  $g_2(t) = s(T-t)$ , es claro que la  $(S/N)_o$  es maximizada cuando  $h(t) = Cs(T-t)$ , o sea,  $h(t)$  está acoplada a la señal  $s(t)$ . El factor de escala  $C^2$  desaparece de la expresión de  $(S/N)_o$  debido a que aparece en el numerador y denominador.

La SNR máxima de salida obtenida con el filtro acoplado es

$$\left(\frac{S}{N}\right)_o = \frac{2}{N_0} \int_0^T s^2(t)dt = \frac{2\mathcal{E}_s}{N_0} \quad (5.9)$$

Notar que la SNR de salida del filtro acoplado depende de la energía de  $s(t)$  pero no de las características detalladas de  $s(t)$ . Esta es otra propiedad interesante del filtro acoplado.

**Interpretación en el dominio frecuencia del filtro acoplado.** El filtro acoplado tiene una interesante interpretación en el dominio frecuencia. Dado que  $h(t) = s(T-t)$ , la transformada de Fourier de esta relación es

$$\begin{aligned} H(f) &= \int_0^T s(T-t)e^{-j2\pi ft}dt \\ &= \left[\int_0^T s(\tau)e^{-j2\pi f\tau}d\tau\right] e^{-j2\pi fT} \\ &= S^*(f)e^{-j2\pi fT} \end{aligned}$$

Es posible notar que el filtro acoplado tiene una respuesta en frecuencia que es igual al complejo conjugado del espectro de la señal transmitida multiplicado por el factor  $e^{-j2\pi fT}$ , el cual representa el retardo de muestreo  $T$ . Además,  $|H(f)| = |S(f)|$ , de forma que la magnitud de la respuesta del filtro acoplado es idéntica a la del espectro de la señal. Por otro lado, la fase de  $H(f)$  es la de  $S(f)$  desplazada  $180^\circ$ .

Si ahora la señal  $s(t)$  con espectro  $S(f)$  pasa a través del filtro acoplado, la salida del filtro tiene un espectro  $Y(f) = |S(f)|^2 e^{-j2\pi fT}$ . En consecuencia, la señal de salida será

$$\begin{aligned} y_s(t) &= \int_{-\infty}^{\infty} Y(f) e^{j2\pi f t} df \\ &= \int_{-\infty}^{\infty} |S(f)|^2 e^{-j2\pi f T} e^{j2\pi f t} df \end{aligned}$$

Muestreando la salida del filtro acoplado en  $t = T$  se tiene que

$$\begin{aligned} y_s(T) &= \int_{-\infty}^{\infty} |S(f)|^2 df \\ &= \int_0^T |s(t)|^2 dt = \mathcal{E}_s \end{aligned}$$

donde el último paso se obtuvo por la relación de Parseval.

El ruido a la salida del filtro acoplado está caracterizado por

$$S_0(f) = |H(f)|^2 N_0/2$$

En consecuencia, la potencia total de ruido a la salida del filtro acoplado es

$$\begin{aligned} P_n &= \int_{-\infty}^{\infty} S_0(f) df = \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2 df \\ &= \frac{N_0}{2} \int_{-\infty}^{\infty} |S(f)|^2 df = \frac{\mathcal{E}_s N_0}{2} \end{aligned}$$

La SNR de salida es simplemente la relación de la potencia de señal  $P_s$ , dada por  $P_s = y_s^2(T)$ , a la potencia de ruido  $P_n$ . En consecuencia,

$$\left( \frac{S}{N} \right)_o = \frac{P_s}{P_n} = \frac{\mathcal{E}_s^2}{\mathcal{E}_s N_0/2} = \frac{2\mathcal{E}_s}{N_0}$$

lo cual concuerda con el resultado obtenido en (5.9).

**Ejemplo.** Considerar las señales biortogonales con  $M = 4$  de la figura 5.6 para transmitir información sobre un canal AWGN. Se supone que el ruido tiene media cero y densidad espectral de potencia  $N_0/2$ . Determinar las funciones base para este conjunto de señales, la respuesta impulsiva y las señales de salida de los demoduladores de filtro acoplado cuando la señal transmitida es  $s_1(t)$ .

Las señales biortogonales para  $M = 4$  tienen dimensión  $N = 2$ . En consecuencia, se necesitan dos funciones base para representar las señales. De la figura 5.6 elegimos  $\psi_1(t)$  y  $\psi_2(t)$  como

$$\begin{aligned} \psi_1(t) &= \begin{cases} \sqrt{\frac{2}{T}}, & 0 \leq t \leq \frac{T}{2} \\ 0, & \text{para todo otro } t \end{cases} \\ \psi_2(t) &= \begin{cases} \sqrt{\frac{2}{T}}, & \frac{T}{2} \leq t \leq T \\ 0, & \text{para todo otro } t \end{cases} \end{aligned}$$

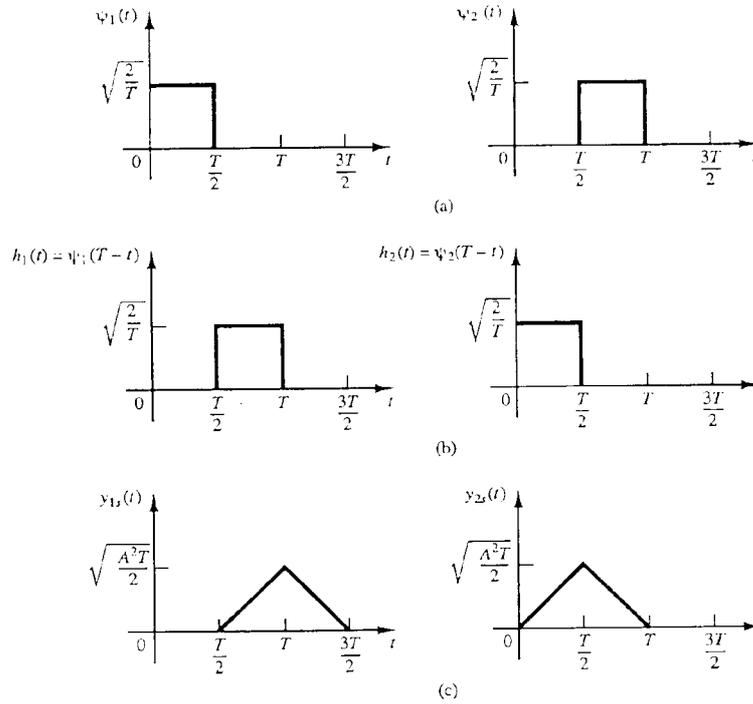


Figura 5.36: Funciones básicas y respuestas del filtro acoplado.

Esas señales se ilustran en la figura 5.36a. Las respuestas impulsivas de los dos filtros acoplados son

$$h_1(t) = \psi_1(T-t) \begin{cases} \sqrt{\frac{2}{T}}, & \frac{T}{2} \leq t \leq T \\ 0, & \text{para todo otro } t \end{cases}$$

$$h_2(t) = \psi_2(T-t) \begin{cases} \sqrt{\frac{2}{T}}, & 0 \leq t \leq T/2 \\ 0, & \text{para todo otro } t \end{cases}$$

y se ilustran en la figura 5.36b.

Si se transmite  $s_1(t)$ , las respuestas sin ruido de los dos filtros acoplados se muestran en la figura 5.36c. Dado que  $y_1(t)$  e  $y_2(t)$  se muestrean en  $t = T$ , se observa que  $y_{1s}(T) = \sqrt{\frac{A^2T}{2}}$  e  $y_{2s}(T) = 0$ . Notar que  $A^2T/2 = \mathcal{E}_s$  es la energía de la señal. En consecuencia, el vector recibido formado por las salidas de los dos filtros acoplados en el instante de muestreo  $t = T$  será

$$\mathbf{r} = [r_1, r_2] = [\sqrt{\mathcal{E}_s} + n_1, n_2]$$

donde  $n_1 = y_{1n}(T)$  y  $n_2 = y_{2n}(T)$  son las componentes de ruido en las salidas de los filtros acoplados dadas por

$$y_{kn}(T) = \int_0^T n(t)\psi_k(t)dt, \quad k = 1, 2$$

Claramente,  $E[n_k] = E[y_{kn}(T)] = 0$ . Sus varianzas serán

$$\begin{aligned} \sigma_n^2 &= E[y_{kn}^2(T)] = \int_0^T \int_0^T E[n(t)n(\tau)]\psi_k(t)\psi_k(\tau)dt d\tau \\ &= \frac{N_0}{2} \int_0^T \int_0^T \delta(t-\tau)\psi_k(t)\psi_k(\tau)dt d\tau \end{aligned}$$

$$= \frac{N_0}{2} \int_0^T \psi_k^2(t) dt = \frac{N_0}{2}$$

Es posible observar también que la  $(S/N)_o$  para el primer filtro acoplado es

$$\left(\frac{S}{N}\right)_o = \frac{(\sqrt{\mathcal{E}_s})^2}{N_0/2} = \frac{2\mathcal{E}_s}{N_0}$$

lo cual concuerda con el resultado previo. Notar también que las cuatro salidas posibles de los dos filtros acoplados, correspondientes a las cuatro señales posibles transmitidas en la figura 5.6 son:  $(r_1, r_2) = (\sqrt{\mathcal{E}_s} + n_1, n_2)$ ,  $(n_1, \sqrt{\mathcal{E}_s} + n_2)$ ,  $(-\sqrt{\mathcal{E}_s} + n_1, n_2)$  y  $(n_1, -\sqrt{\mathcal{E}_s} + n_2)$ .

### 5.2.3 Detector óptimo

En las dos secciones anteriores se demostró que para una señal transmitida sobre un canal AWGN, un demodulador tipo correlador o tipo filtro acoplado produce un vector  $\mathbf{r} = (r_1, r_2, \dots, r_N)$  que contiene toda la información relevante asociada a la señal recibida. En esta sección se describirá la regla óptima de decisión basada en el vector de observación  $\mathbf{r}$ . Para este desarrollo se asumirá que no existe memoria en las señales transmitidas en los intervalos sucesivos.

Se desea diseñar un detector de señal que haga una decisión sobre la señal transmitida en cada intervalo de señal basándose en la observación del vector  $\mathbf{r}$  en cada intervalo tal que la probabilidad de detección correcta sea maximizada. Con este propósito en mente, es posible considerar una regla de decisión basada en el cálculo de la *probabilidad a posteriori*, definida como

$$P(\text{la señal } \mathbf{s}_m \text{ fué transmitida} | \mathbf{r}), \quad m = 1, 2, \dots, M$$

la cual puede abreviarse como  $P(\mathbf{s}_m | \mathbf{r})$ . El criterio de decisión consiste en seleccionar la señal correspondiente al máximo del conjunto de probabilidades a posteriori  $\{P(\mathbf{s}_m | \mathbf{r})\}$ . Más tarde se demostrará que este criterio maximiza la probabilidad de detección correcta y, en consecuencia, minimiza la probabilidad de error. Este criterio de decisión se denomina *criterio de probabilidad a posteriori máxima* (MAP).

Usando la regla de Bayes, la probabilidad a posteriori puede expresarse como

$$P(\mathbf{s}_m | \mathbf{r}) = \frac{f(\mathbf{r} | \mathbf{s}_m)P(\mathbf{s}_m)}{f(\mathbf{r})} \quad (5.10)$$

donde  $f(\mathbf{r} | \mathbf{s}_m)$  es la pdf condicional del vector observado dado  $\mathbf{s}_m$ , y  $P(\mathbf{s}_m)$  es la *probabilidad a priori* de la  $m$ -ésima señal transmitida. El denominador de (5.10) puede expresarse como

$$f(\mathbf{r}) = \sum_{m=1}^M f(\mathbf{r} | \mathbf{s}_m)P(\mathbf{s}_m) \quad (5.11)$$

De (5.10) y (5.11) es posible concluir que el cálculo de la probabilidad a posteriori  $P(\mathbf{s}_m | \mathbf{r})$  requiere el conocimiento de las *probabilidades a priori*  $P(\mathbf{s}_m)$  y las pdf condicionales  $f(\mathbf{r} | \mathbf{s}_m)$ , para  $m = 1, 2, \dots, M$ .

Es posible introducir alguna simplificación en el criterio MAP cuando las  $M$  señales son igualmente probables a priori, o sea,  $P(\mathbf{s}_m) = 1/M$  para todo  $M$ . Además, es posible notar que el denominador de (5.10) es independiente de la señal transmitida. En consecuencia, la regla de decisión basada en hallar la señal que maximiza  $P(\mathbf{s}_m | \mathbf{r})$  es equivalente a hallar la señal que maximiza  $f(\mathbf{r} | \mathbf{s}_m)$ .

La pdf condicional  $f(\mathbf{r} | \mathbf{s}_m)$  o cualquier función monótona de ella se denomina *función verosimilitud*. El criterio de decisión basado en el máximo de  $f(\mathbf{r} | \mathbf{s}_m)$  sobre  $M$  señales se denomina *criterio de máxima verosimilitud* (ML). Es posible observar que un detector basado en el criterio MAP y otro basado en el criterio ML realizan las mismas decisiones siempre que las probabilidades a priori  $P(\mathbf{s}_m)$  sean todas iguales, o sea, las señales  $\{\mathbf{s}_m\}$  sean equiprobables.

En el caso de un canal AWGN, la función verosimilitud  $f(\mathbf{r} | \mathbf{s}_m)$  está definida por (5.7). Para simplificar, es posible trabajar con el logaritmo natural de  $f(\mathbf{r} | \mathbf{s}_m)$ , el cual es una función monótona. De esta forma,

$$\ln f(\mathbf{r}|\mathbf{s}_m) = \frac{-N}{2} \ln(\pi N_0) - \frac{1}{N_0} \sum_{k=1}^N (r_k - s_{mk})^2$$

El máximo de  $\ln f(\mathbf{r}|\mathbf{s}_m)$  sobre  $\mathbf{s}_m$  es equivalente a hallar la señal  $\mathbf{s}_m$  que minimiza la distancia Euclidiana

$$D(\mathbf{r}, \mathbf{s}_m) = \sum_{k=1}^N (r_k - s_{mk})^2$$

Se denominará a  $D(\mathbf{r}, \mathbf{s}_m)$ ,  $m = 1, 2, \dots, M$  la *métrica de distancia*. De esta forma, para el canal AWGN, la regla de decisión basada en el criterio ML se reduce a hallar la señal  $\mathbf{s}_m$  más próxima en distancia al vector de señal recibido  $\mathbf{r}$ . Se denominará a esta regla de decisión como *detección de distancia mínima*.

Otra interpretación de la regla de decisión óptima del criterio ML se obtiene expandiendo la métrica de distancia como

$$\begin{aligned} D(\mathbf{r}, \mathbf{s}_m) &= \sum_{n=1}^N r_n^2 - 2 \sum_{n=1}^N r_n s_{mn} + \sum_{n=1}^N s_{mn}^2 \\ &= |\mathbf{r}|^2 - 2\mathbf{r} \cdot \mathbf{s}_m + |\mathbf{s}_m|^2, \quad m = 1, 2, \dots, M \end{aligned}$$

El término  $|\mathbf{r}|^2$  es común a todas las métricas de decisión y, en consecuencia, puede descartarse en sus cálculos. El resultado es un conjunto de métrica de distancia modificadas dadas por

$$D'(\mathbf{r}, \mathbf{s}_m) = -2\mathbf{r} \cdot \mathbf{s}_m + |\mathbf{s}_m|^2$$

Notar que, seleccionar la señal  $\mathbf{s}_m$  que minimiza  $D'(\mathbf{r}, \mathbf{s}_m)$  es equivalente a seleccionar la señal que maximiza la métrica  $C(\mathbf{r}, \mathbf{s}_m) = -D'(\mathbf{r}, \mathbf{s}_m)$ , o sea

$$C(\mathbf{r}, \mathbf{s}_m) = 2\mathbf{r} \cdot \mathbf{s}_m - |\mathbf{s}_m|^2 \quad (5.12)$$

El término  $\mathbf{r} \cdot \mathbf{s}_m$  representa la proyección del vector de señal recibido sobre cada uno de los  $M$  posible vectores de señal transmitidos. El valor de cada una de esas proyecciones es una medida de la correlación entre el vector recibido y la  $m$ -ésima señal. Por esta razón,  $C(\mathbf{r}, \mathbf{s}_m)$  se denomina la *métrica de correlación* para decidir cual de las  $M$  señales fué transmitida. Finalmente, los términos  $|\mathbf{s}_m|^2 = \mathcal{E}_m$ , puede interpretarse como términos de sesgo que sirven para compensar las diferentes energías del conjunto de señales, como es el caso en PAM. Si todas las señales tienen la misma energía  $|\mathbf{s}_m|^2$  también puede descartarse en el cálculo de las métricas de correlación  $C(\mathbf{r}, \mathbf{s}_m)$ .

En síntesis, se demostró que el detector ML óptimo calcula un conjunto de distancias  $M$ ,  $D(\mathbf{r}, \mathbf{s}_m)$  ó  $D'(\mathbf{r}, \mathbf{s}_m)$  y selecciona la señal correspondiente a la de menor métrica. Equivalentemente, el detector ML óptimo calcula un conjunto de  $M$  métricas de correlación  $C(\mathbf{r}, \mathbf{s}_m)$  y selecciona la señal correspondiente a la mayor métrica de correlación.

En el desarrollo anterior para el detector óptimo se discutió el caso importante en el cual las señales son equiprobables. En este caso, el criterio MAP es equivalente al criterio ML. Sin embargo, cuando las señales no son igualmente probables, el detector MAP óptimo basa su decisión en las probabilidades  $P(\mathbf{s}_m|\mathbf{r})$ ,  $m = 1, 2, \dots, M$ , dadas en (5.10), o equivalentemente en las métricas

$$PM(\mathbf{r}, \mathbf{s}_m) = f(\mathbf{r}|\mathbf{s}_m)P(\mathbf{s}_m) \quad (5.13)$$

---

**Ejemplo** Considerar el caso de señales PAM binarias en las cuales los dos puntos de señal posibles son  $s_1 = -s_2 = \sqrt{\mathcal{E}_b}$ , donde  $\mathcal{E}_b$  es la energía por bit. Las probabilidades a priori son  $P(s_1) = p$  y  $P(s_2) = 1 - p$ . Determinar las métricas para el detector MAP óptimo cuando la señal es transmitida por un canal AWGN.

El vector de señal recibida para PAM binario es

$$r = \pm\sqrt{\mathcal{E}_b} + y_n(T) \quad (5.14)$$

donde  $y_n(T)$  es una variable aleatoria Gaussiana de media cero y varianza  $\sigma_n^2 = N_0/2$ . En consecuencia, las pdf condicionales para las dos señales son

$$\begin{aligned} f(r|s_1) &= \frac{1}{\sqrt{2\pi}\sigma_n} e^{-(r-\sqrt{\mathcal{E}_b})^2/2\sigma_n^2} \\ f(r|s_2) &= \frac{1}{\sqrt{2\pi}\sigma_n} e^{-(r+\sqrt{\mathcal{E}_b})^2/2\sigma_n^2} \end{aligned}$$

Entonces las métricas  $PM(\mathbf{r}, \mathbf{s}_1)$  y  $PM(\mathbf{r}, \mathbf{s}_2)$  definidas en (5.13) son

$$\begin{aligned} PM(\mathbf{r}, \mathbf{s}_1) = pf(r|s_1) &= \frac{p}{\sqrt{2\pi}\sigma_n} e^{-(r-\sqrt{\mathcal{E}_b})^2/2\sigma_n^2} \\ PM(\mathbf{r}, \mathbf{s}_2) = pf(r|s_2) &= \frac{1-p}{\sqrt{2\pi}\sigma_n} e^{-(r+\sqrt{\mathcal{E}_b})^2/2\sigma_n^2} \end{aligned}$$

Si  $PM(\mathbf{r}, \mathbf{s}_1) > PM(\mathbf{r}, \mathbf{s}_2)$ , se selecciona  $s_1$ , de otro modo se selecciona  $s_2$  como la señal transmitida. Esta regla de selección puede expresarse como

$$\frac{PM(\mathbf{r}, \mathbf{s}_1)}{PM(\mathbf{r}, \mathbf{s}_2)} \underset{s_2}{\overset{s_1}{>}} 1$$

Pero

$$\frac{PM(\mathbf{r}, \mathbf{s}_1)}{PM(\mathbf{r}, \mathbf{s}_2)} = \frac{p}{1-p} e^{[(r+\sqrt{\mathcal{E}_b})^2 - (r-\sqrt{\mathcal{E}_b})^2]/2\sigma_n^2} \quad (5.15)$$

de forma que (5.15) puede expresarse como

$$\frac{(r + \sqrt{\mathcal{E}_b})^2 - (r - \sqrt{\mathcal{E}_b})^2}{2\sigma_n^2} \underset{s_2}{\overset{s_1}{>}} \ln \frac{1-p}{p}$$

o equivalentemente

$$\mathcal{E}_b r \underset{s_2}{\overset{s_1}{>}} \frac{\sigma_n^2}{2} \ln \frac{1-p}{p} = \frac{N_0}{4} \ln \frac{1-p}{p}$$

Esta es la forma final del detector óptimo. Con ella se calcula la métrica de correlación  $C(\mathbf{r}, \mathbf{s}_1) = r\sqrt{\mathcal{E}_b}$  y se compara con el umbral  $(N_0/4) \ln(1-p)/p$ .

Es interesante notar que en el caso de probabilidades a priori diferentes es necesario conocer no solamente esas probabilidades sino también el valor de  $N_0$  para calcular el umbral. Cuando  $p = 1/2$  el umbral es cero y ya no se requiere el conocimiento de  $N_0$  en el detector.

Es posible concluir esta sección con la prueba que la regla de decisión basada en el criterio ML minimiza la probabilidad de error cuando las  $M$  señales son equiprobables a priori. Definimos por  $R_m$  la región en el espacio  $N$  dimensional para la cual se decide que la señal  $s_m(t)$  fué transmitida cuando se recibe el vector  $\mathbf{r} = (r_1, r_2, \dots, r_N)$ . La probabilidad de un error en la decisión cuando  $s_m(t)$  fué transmitida es

$$P(e|\mathbf{s}_m) = \int_{R_m^c} f(\mathbf{r}|\mathbf{s}_m) d\mathbf{r}$$

donde  $R_m^c$  es el complemento de  $R_m$ . La probabilidad de error promedio es

$$\begin{aligned} P(e) &= \sum_{m=1}^M \frac{1}{M} P(e|\mathbf{s}_m) \\ &= \sum_{m=1}^M \frac{1}{M} \int_{R_m^c} f(\mathbf{r}|\mathbf{s}_m) d\mathbf{r} \\ &= \sum_{m=1}^M \frac{1}{M} \left[ 1 - \int_{R_m} f(\mathbf{r}|\mathbf{s}_m) d\mathbf{r} \right] \end{aligned}$$

Notar que  $P(e)$  se minimiza seleccionando la señal  $\mathbf{s}_m$  cuya  $f(\mathbf{r}|\mathbf{s}_m)$  es la mayor de las  $f(\mathbf{r}|\mathbf{s}_k)$  para todo  $m \neq k$ . Cuando las  $M$  señales no son equiprobables, la prueba anterior puede generalizarse para mostrar que el criterio MAP minimiza la probabilidad de error promedio.

#### 5.2.4 Detector de secuencias de Máxima Verosimilitud (MLSD)

Cuando la señal no tiene memoria, el detector de símbolo por símbolo descrito en la sección anterior es óptimo en el sentido de minimizar la probabilidad promedio de un error de símbolo. Por otro lado, cuando la señal transmitida tiene memoria, o sea, las señales transmitidas en intervalos de símbolos sucesivos son interdependientes, el detector óptimo es un detector de secuencias de máxima verosimilitud, el que basa sus decisiones en la observación de una secuencia de señales recibidas sobre intervalos sucesivos de señales.

Es posible considerar como ejemplo la señal NRZI descrita anteriormente. Su memoria está caracterizada por la trellis mostrada en la figura 5.25. La señal transmitida en cada intervalo es PAM binaria. En consecuencia, existen dos señales transmitidas posibles correspondientes a los puntos de señal  $s_2 = -s_1 = \sqrt{\mathcal{E}_b}$ , donde  $\mathcal{E}_b$  es la energía por bit. Como fué mostrado también, mediante un ejemplo, la salida del demodulador del tipo correlador o filtro acoplado para PAM binaria en el  $k$ -ésimo intervalo puede expresarse como (ver (5.14))

$$r_k = \pm \sqrt{\mathcal{E}_b} + n_k$$

donde  $n_k$  es una variable aleatoria Gaussiana de media cero con varianza  $\sigma_n^2 = N_0/2$ . En consecuencia, las pdf condicionales para las dos señales son

$$\begin{aligned} f(r|s_1) &= \frac{1}{\sqrt{2\pi}\sigma_n} e^{-(r_k - \sqrt{\mathcal{E}_b})^2 / 2\sigma_n^2} \\ f(r|s_2) &= \frac{1}{\sqrt{2\pi}\sigma_n} e^{-(r_k + \sqrt{\mathcal{E}_b})^2 / 2\sigma_n^2} \end{aligned}$$

Es posible suponer ahora la observación de la secuencia de salida de los filtros acoplados  $r_1, r_2, \dots, r_K$ . Dado que el ruido del canal se supone blanco y Gaussiano y  $\psi(t-iT)$ ,  $\psi(t-jT)$  para  $i \neq j$  son ortogonales, es posible concluir que  $E(n_i n_j) = 0$ , para  $i \neq j$ . En consecuencia, la secuencia de ruido  $n_1, n_2, \dots, n_K$  también es blanca. De esta forma, para cualquier secuencia transmitida  $\mathbf{s}^{(m)}$ , la pdf conjunta de  $r_1, r_2, \dots, r_K$  puede expresarse como el producto de  $K$  pdf marginales, o sea

$$\begin{aligned} p(r_1, r_2, \dots, r_K | \mathbf{s}^{(m)}) &= \prod_{k=1}^K p(r_k | s_k^{(m)}) \\ &= \prod_{k=1}^K \frac{1}{\sqrt{2\pi}\sigma_n} e^{-(r_k - s_k^{(m)})^2 / 2\sigma_n^2} \\ &= \left( \frac{1}{\sqrt{2\pi}\sigma_n} \right)^K e^{-\sum_{k=1}^K (r_k - s_k^{(m)})^2 / 2\sigma_n^2} \end{aligned} \quad (5.16)$$

donde  $s_k = \sqrt{\mathcal{E}_b}$  ó  $s_k = -\sqrt{\mathcal{E}_b}$ , Luego, dada la secuencia recibida  $r_1, r_2, \dots, r_K$  a la salida del demodulador tipo correlador o filtro acoplado, el detector determina la secuencia  $\mathbf{s}^{(m)} = [s_1^{(m)}, s_2^{(m)}, \dots, s_K^{(m)}]$  que maximiza la pdf condicional  $p(r_1, r_2, \dots, r_K | \mathbf{s}^{(m)})$ . Este detector se denomina *detector de secuencia de máxima verosimilitud* (MLSD).

En base al logaritmo de (5.16), y despreciando los términos que son independientes de  $(r_1, r_2, \dots, r_K)$ , es posible hallar que el MLSD selecciona la secuencia  $\mathbf{s}^{(m)}$  que minimiza la *métrica de distancia Euclidiana* dada por

$$D(\mathbf{r}, \mathbf{s}^{(m)}) = \sum_{k=1}^K \left( r_k - s_k^{(m)} \right)^2$$

En la búsqueda a través de la trellis por la secuencia que minimiza la distancia Euclidiana  $D(\mathbf{r}, \mathbf{s}^{(m)})$  puede parecer que es necesario calcular la distancia  $D(\mathbf{r}, \mathbf{s}^{(m)})$  para cada secuencia posible. Para el ejemplo NRZI que emplea modulación binaria, el número total de caminos es  $2^K$ , donde  $K$  es el número de salidas obtenidas del demodulador. Sin embargo, por suerte ese no es el caso. Es posible reducir el número de secuencias en la búsqueda sobre la trellis mediante el uso del *algoritmo de Viterbi* para eliminar secuencias mientras se reciben nuevos datos del demodulador.

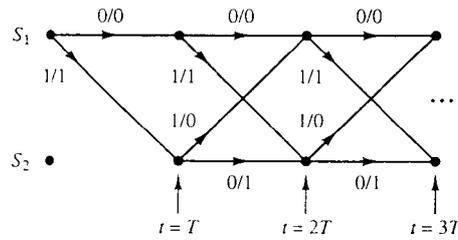


Figura 5.37: Trellis para la señal NRZI con estado inicial  $s_1$ .

El algoritmo de Viterbi es una búsqueda secuencial sobre la trellis mediante MLSD. Se describirá en el Capítulo 8 cuando se discutan técnicas de codificación de canal. Se describirá aquí su aplicación en el contexto de la señal NRZI. Se supone que el proceso de búsqueda comienza en el estado  $S_1$ . La trellis correspondiente se muestra en la figura 5.37.

En el instante  $t = T$ , se recibe  $r_1 = s_1^{(m)} + n_1$  del demodulador, y en  $t = 2T$  se recibe  $r_2 = s_2^{(m)} + n_2$ . Dado que la memoria de la señal es de un bit, lo cual se notará por  $L = 1$ , es posible observar que la trellis alcanza su forma regular (estado estacionario) después de la primera transición. De esta forma, habiendo recibido  $r_2$  en  $t = 2T$ , es posible observar que existen dos caminos de señal entrando en cada uno de los dos nodos y dos caminos de señal saliendo de cada nodo. Los dos caminos entrando al nodo  $S_1$  en  $t = 2T$  corresponden a los bits de información  $(0, 0)$  y  $(1, 1)$ , o equivalentemente, a los puntos de señal  $(-\sqrt{\mathcal{E}_b}, -\sqrt{\mathcal{E}_b})$  y  $(\sqrt{\mathcal{E}_b}, -\sqrt{\mathcal{E}_b})$ , respectivamente. Los dos caminos entrando al nodo  $S_2$  en  $t = 2T$  corresponden a los bits de información  $(0, 1)$  y  $(1, 0)$ , o equivalentemente, a los puntos de señal  $(-\sqrt{\mathcal{E}_b}, \sqrt{\mathcal{E}_b})$  y  $(\sqrt{\mathcal{E}_b}, \sqrt{\mathcal{E}_b})$ , respectivamente.

Para los dos caminos entrando al nodo  $S_1$  es posible calcular las métricas de distancia Euclidiana dadas por

$$\mu_2(0, 0) = (r_1 + \sqrt{\mathcal{E}_b})^2 + (r_2 + \sqrt{\mathcal{E}_b})^2 \quad (5.17)$$

$$\mu_1(1, 1) = (r_1 - \sqrt{\mathcal{E}_b})^2 + (r_2 - \sqrt{\mathcal{E}_b})^2 \quad (5.18)$$

$$\mu_2(0, 1) = (r_1 + \sqrt{\mathcal{E}_b})^2 + (r_2 + \sqrt{\mathcal{E}_b})^2 \quad (5.19)$$

utilizando las salidas  $r_1$  y  $r_2$  del modulator. El algoritmo de Viterbi compara esas dos métricas y descarta el camino que tiene el mayor valor (la mayor distancia). El camino restante con la menor métrica se guarda y se denomina *superviviente* en  $t = 2T$ . La eliminación de uno de los dos caminos puede hacerse sin comprometer la optimalidad de la búsqueda sobre la trellis ya que cualquier extensión del camino con la mayor distancia más allá de  $t = 2T$  siempre tendrá una métrica mayor que la asociada añadiendo al del superviviente el camino más allá de  $t = 2T$ .

En forma similar, para los dos caminos entrando al nodo  $S_2$  es posible calcular las métricas de distancia Euclidianas dadas por

$$\begin{aligned}\mu_2(0,1) &= (r_1 + \sqrt{\mathcal{E}_b})^2 + (r_2 - \sqrt{\mathcal{E}_b})^2 \\ \mu_1(1,0) &= (r_1 - \sqrt{\mathcal{E}_b})^2 + (r_2 - \sqrt{\mathcal{E}_b})^2\end{aligned}$$

utilizando las salidas  $r_1$  y  $r_2$  del modulador. Se compara las dos métricas y se descarta el camino que tiene el mayor valor. De esta forma, en  $t = 2T$  se ha dejado dos caminos supervivientes, uno en el nodo  $S_1$  y otro en el nodo  $S_2$  y sus correspondientes métricas. Los caminos de señal en los nodos  $S_1$  y  $S_2$  se extienden luego a lo largo de los dos caminos supervivientes.

Habiendo recibido  $r_3$  en  $t = 3T$ , es posible calcular las métricas de los dos caminos entrando al estado  $S_1$ . Supongamos que los supervivientes en  $t = 2T$  son los caminos  $(0,0)$  en  $S_1$  y  $(0,1)$  en  $S_2$ . Luego, las dos métricas de los caminos entrantes en  $S_1$  en  $t = 3T$  serán

$$\begin{aligned}\mu_3(0,0,0) &= \mu_2(0,0) + (r_3 + \sqrt{\mathcal{E}_b})^2 \\ \mu_3(0,1,1) &= \mu_2(0,1) + (r_3 + \sqrt{\mathcal{E}_b})^2\end{aligned}$$

Estas dos métricas se comparan y el camino con mayor valor se elimina. En forma similar, las métricas de los caminos entrantes en  $S_2$  en  $t = 3T$  serán

$$\begin{aligned}\mu_3(0,0,1) &= \mu_2(0,0) + (r_3 - \sqrt{\mathcal{E}_b})^2 \\ \mu_3(0,1,0) &= \mu_2(0,1) + (r_3 - \sqrt{\mathcal{E}_b})^2\end{aligned}$$

Estas dos métricas se comparan y el camino con mayor valor se elimina.

Este proceso continúa con cada muestra de señal recibida por el demodulador. Así, el algoritmo de Viterbi calcula dos métricas para los dos caminos de señal que entran a un nodo en cada etapa de la búsqueda sobre la trellis y elimina uno de los caminos en cada nodo. Los dos caminos supervivientes se extienden luego hacia adelante al próximo estado. En consecuencia, el número de caminos buscados sobre la trellis se reduce por un factor de dos en cada etapa.

Es relativamente simple generalizar la búsqueda sobre la trellis realizada por el algoritmo de Viterbi para modulación  $M$ -ária. Por ejemplo, la modulación de retardo emplea  $M = 4$  señales y está caracterizada por la trellis de cuatro estados mostrada en la figura 5.27. Es posible observar que cada estado tiene dos caminos de señal entrantes a cada nodo y dos caminos salientes en cada estado. La memoria de la señal es  $L = 1$ . En consecuencia, el algoritmo de Viterbi tendrá cuatro supervivientes en cada etapa y sus métricas correspondientes. En cada nodo se calculan dos métricas correspondientes a los dos caminos entrantes y uno de los caminos de señal entrante al nodo se elimina en cada estado de la trellis. De esta forma, el algoritmo de Viterbi minimiza el número de caminos de búsqueda sobre la trellis para realizar MLSD.

De la descripción discutida del algoritmo de Viterbi no queda claro como se realizan las decisiones sobre los símbolos de información detectados dadas las secuencias supervivientes. Si se ha avanzado hasta alguna etapa, por ejemplo  $K$ , con  $K \gg L$ , en la trellis, y comparamos las secuencias supervivientes, se hallará que con probabilidad próxima a uno todas esas secuencias supervivientes serán idénticas en  $K - 5L$  posiciones de bit (o símbolo). En una implementación práctica del algoritmo de Viterbi, las decisiones sobre cada bit (o símbolo) de información se toman después de un retardo de  $5L$  bits (o símbolos), y de esta forma, las secuencias supervivientes se truncan a los  $5L$  bits (o símbolos) más recientes. De esta forma, se evita un retardo variable en la detección del bit o símbolo. La pérdida de desempeño resultante del procedimiento de detección subóptimo es despreciable si el retardo es al menos de  $5L$ .

---

**Ejemplo:** Describir la regla de decisión para detectar la secuencia de datos en una señal NRZI con el algoritmo de Viterbi teniendo un retardo de  $5L$  bits.

La trellis para una señal NRZI se muestra en la figura 5.25. En este caso,  $L = 1$ , de donde el retardo en la detección es de 5 bits. De esta forma, en  $t = 6T$ , se tendrán dos secuencias supervivientes, una por cada uno de los estados, y sus métricas correspondientes  $\mu_6(b_1, b_2, b_3, b_4, b_5, b_6)$  y  $\mu_6(b'_1, b'_2, b'_3, b'_4, b'_5, b'_6)$ . En esta etapa, con probabilidad próxima a uno, el bit  $b_1$  será el mismo que  $b'_1$ , o sea, ambas secuencias supervivientes

tendrán una primer rama común. Luego se elimina el primer bit de las dos secuencias supervivientes. En  $t = 7T$ , las métricas  $\mu_7(b_2, b_3, b_4, b_5, b_6, b_7)$  y  $\mu_7(b'_2, b'_3, b'_4, b'_5, b'_6, b'_7)$  se usarán para determinar la decisión sobre el bit  $b_2$ . Este proceso continúa en cada etapa de la búsqueda sobre la trellis para la secuencia de distancia mínima. De esta forma el retardo de detección queda fijo en 5 bits.

## 5.3 Probabilidad de error para señales en ruido blanco gaussiano

En esta sección se evaluará la probabilidad de error para las diferentes señales descritas en la primera sección de este Capítulo.

### 5.3.1 Probabilidad de error para modulación binaria

Es posible considerar señales PAM binarias,  $s_1(t) = g_T(t)$  y  $s_2(t) = -g_T(t)$ , con  $g_T(t)$  un pulso arbitrario no nulo en el intervalo  $0 \leq t \leq T_b$  y cero fuera de él.

Dado que  $s_1(t) = -s_2(t)$ , estas señales se dicen *antipodales*. La energía en el pulso  $g_T(t)$  es  $\mathcal{E}_b$ . Como ya fué discutido, las señales PAM son unidimensionales, y entonces, su representación geométrica es simplemente un vector unidimensional  $s_1 = \sqrt{\mathcal{E}_b}$  ó  $s_2 = -\sqrt{\mathcal{E}_b}$ .

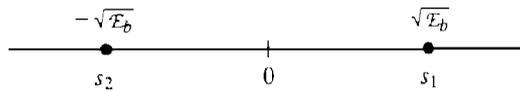


Figura 5.38: Puntos de señal para señales binarias antipodales.

Es posible suponer que las dos señales son equiprobables y que  $s_1(t)$  fué transmitida. Luego, la señal recibida por el demodulador será

$$r = s_1 + n = \sqrt{\mathcal{E}_b} + n$$

donde  $n$  representa la componente AWGN que tiene media cero y varianza  $\sigma_n^2 = N_0/2$ . En este caso, la regla de decisión basada en la métrica de correlación dada por la ecuación (5.12) compara  $r$  con el umbral cero. Si  $r > 0$ , la decisión se realiza en favor de  $s_1(t)$  y si  $r < 0$  la decisión es que  $s_2(t)$  fué transmitida. Claramente las dos pdf condicionales de  $r$  son

$$f(r|s_1) = \frac{1}{\sqrt{\pi N_0}} e^{-(r-\sqrt{\mathcal{E}_b})^2/N_0}$$

$$f(r|s_2) = \frac{1}{\sqrt{\pi N_0}} e^{-(r+\sqrt{\mathcal{E}_b})^2/N_0}$$

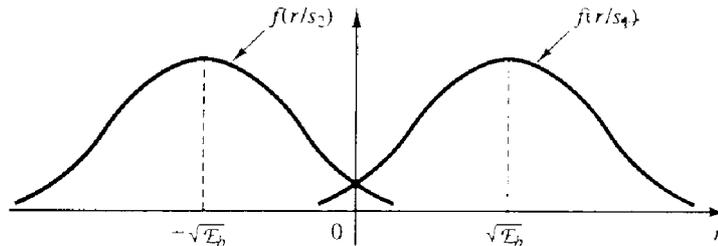


Figura 5.39: Funciones densidad de probabilidad condicionales de dos señales.

Estas dos pdf condicionales se muestran en la figura 5.39. Cuando  $s_1(t)$  fué transmitida, la probabilidad de error es simplemente la probabilidad que  $r < 0$ , o sea

$$\begin{aligned}
 P(e|s_1) &= \int_{-\infty}^0 f(r|s_1) dr = \frac{1}{\sqrt{\pi N_0}} \int_{-\infty}^0 e^{-(r-\sqrt{\mathcal{E}_b})^2} dr \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{2\mathcal{E}_b}/N_0} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2\mathcal{E}_b}/N_0}^{\infty} e^{-x^2/2} dx \\
 &= Q \left[ \sqrt{\frac{2\mathcal{E}_b}{N_0}} \right]
 \end{aligned}$$

donde  $Q(x)$  es la función introducida en el Capítulo 2. En forma similar, si suponemos que  $s_2(t)$  fué transmitida,  $r = -\sqrt{\mathcal{E}_b} + n$  y la probabilidad que  $r > 0$  es también  $P(e|s_2) = Q \left( \sqrt{\frac{2\mathcal{E}_b}{N_0}} \right)$ . Dado que las señales  $s_1(t)$  y  $s_2(t)$  son igualmente probables, la probabilidad de error promedio es

$$P_b = \frac{1}{2}P(e|s_1) + \frac{1}{2}P(e|s_2) = Q \left[ \sqrt{\frac{2\mathcal{E}_b}{N_0}} \right]$$

Es importante notar las características de esta medida de desempeño. Primero, se puede notar que la probabilidad de error depende solo de la relación  $\mathcal{E}_b/N_0$  y no de cualquier otra característica de las señales o el ruido. Segundo, se puede notar que  $2\mathcal{E}_b/N_0$  es también la SNR de salida del demodulador. La relación  $\mathcal{E}_b/N_0$  se denomina en general *relación señal - ruido*.

Es posible notar también que la probabilidad de error puede expresarse en términos de la distancia entre las señales  $s_1$  y  $s_2$ . De la figura 5.38 se observa que las señales están separadas por la distancia  $d_{12} = 2\sqrt{\mathcal{E}_b}$ . Sustituyendo  $\mathcal{E}_b = d_{12}^2/4$  en la ecuación anterior se obtiene

$$P_b = Q \left[ \sqrt{\frac{d_{12}^2}{2N_0}} \right]$$

Esta expresión ilustra la dependencia de la probabilidad de error de la distancia entre los puntos de señal.

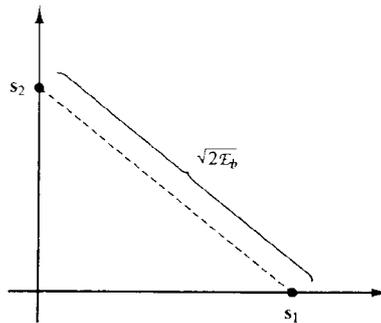


Figura 5.40: Puntos de señal para señales ortogonales binarias.

Es posible evaluar ahora la probabilidad de error para señales binarias ortogonales. (PPM binaria es un ejemplo de señalización ortogonal binaria). En este caso, los vectores de señal  $s_1$  y  $s_2$  son bidimensionales, como mostrado en la figura 5.40, y pueden expresarse como

$$s_1 = (\sqrt{\mathcal{E}_b}, 0) \quad s_2 = (0, \sqrt{\mathcal{E}_b})$$

donde  $\mathcal{E}_b$  es la energía de cada señal. Notar que la distancia entre puntos de señal es  $d_{12} = \sqrt{2\mathcal{E}_b}$ . Para evaluar la probabilidad de error, se supone en principio que  $s_1$  fué transmitida. Luego, el vector recibido a la salida del demodulador es

$$\mathbf{r} = [\sqrt{\mathcal{E}_b} + n_1, n_2]$$

Es posible ahora sustituir  $\mathbf{r}$  en la métrica de correlación de (5.12) para obtener  $C(\mathbf{r}, \mathbf{s}_1)$  y  $C(\mathbf{r}, \mathbf{s}_2)$ . Entonces la probabilidad de error es la probabilidad que  $C(\mathbf{r}, \mathbf{s}_2) > C(\mathbf{r}, \mathbf{s}_1)$ . De esta forma

$$P(\epsilon|\mathbf{s}_1) = P[C(\mathbf{r}, \mathbf{s}_2) > C(\mathbf{r}, \mathbf{s}_1)] = P[n_2 - n_1 > \sqrt{\mathcal{E}_b}]$$

Dado que  $n_1$  y  $n_2$  son variables aleatorias AWGN de media cero y varianza  $N_0/2$ , la variable aleatoria  $x = n_2 - n_1$  también es Gaussiana, de media cero y varianza  $N_0$ . De esta forma,

$$\begin{aligned} P[n_2 - n_1 > \sqrt{\mathcal{E}_b}] &= \frac{1}{\sqrt{2\pi N_0}} \int_{\sqrt{\mathcal{E}_b}}^{\infty} e^{-x^2/2N_0} dx = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\mathcal{E}_b/N_0}}^{\infty} e^{-x^2/2} dx \\ &= Q \left[ \sqrt{\frac{\mathcal{E}_b}{N_0}} \right] \end{aligned}$$

Debido a la simetría, si se supone que  $\mathbf{s}_2$  fué transmitida se obtiene la misma probabilidad de error. En consecuencia, la probabilidad de error promedio para las señales binarias ortogonales es

$$P_b = Q \left[ \sqrt{\frac{\mathcal{E}_b}{N_0}} \right] \quad (5.20)$$

Si se compara la probabilidad de error para señales antipodales con la de señales ortogonales es posible hallar que estas últimas requieren un factor de dos de incremento en la energía para lograr el mismo desempeño. Como  $10 \log_{10} 2 = 3$  dB, entonces es posible decir que las señales ortogonales son 3 dB más pobres que las antipodales. La diferencia de 3 dB se debe simplemente a la distancia entre los puntos de señal,  $d_{12}^2 = 2\mathcal{E}_b$  para señales ortogonales y  $d_{12}^2 = 4\mathcal{E}_b$  para las antipodales.

La probabilidad de error en función de  $10 \log_{10} \mathcal{E}_b/N_0$  para esos dos tipos de señales se muestra en la figura 5.41. Como se observa en esta figura, para cualquier probabilidad de error, el valor de  $\mathcal{E}_b/N_0$  requerido por las señales ortogonales es 3 dB mayor que el requerido por las señales antipodales.

### 5.3.2 Probabilidad de error para modulación multinivel

**Señales PAM multinivel:** Teniendo en cuenta que este tipo de señales se representa geoméricamente mediante  $M$  puntos en un espacio unidimensional con valores  $s_m = \sqrt{\mathcal{E}_g} A_m$ ,  $m = 1, 2, \dots, M$ , donde  $\mathcal{E}_g$  es la energía del pulso básico  $g_T(t)$ , los valores de amplitud pueden expresarse como

$$A_m = (2m - 1 - M), \quad m = 1, 2, \dots, M$$

donde la distancia entre puntos adyacentes es  $2\sqrt{\mathcal{E}_g}$ . Como indicado previamente, las señales PAM tienen diferente energía  $\{\mathcal{E}_m\}$ . La energía promedio es

$$\begin{aligned} \mathcal{E}_{av} &= \frac{1}{M} \sum_{m=1}^M \mathcal{E}_m = \frac{\mathcal{E}_g}{M} \sum_{m=1}^M (2m - 1 - M)^2 \\ &= \frac{\mathcal{E}_g}{M} \frac{M(M^2 - 1)}{3} = \left( \frac{M^2 - 1}{3} \right) \mathcal{E}_g \end{aligned}$$

En forma equivalente, es posible caracterizar esas señales en términos de su potencia promedio como sigue

$$P_{av} = \frac{\mathcal{E}_{av}}{T} = \left( \frac{M^2 - 1}{3} \right) \frac{\mathcal{E}_g}{T}$$

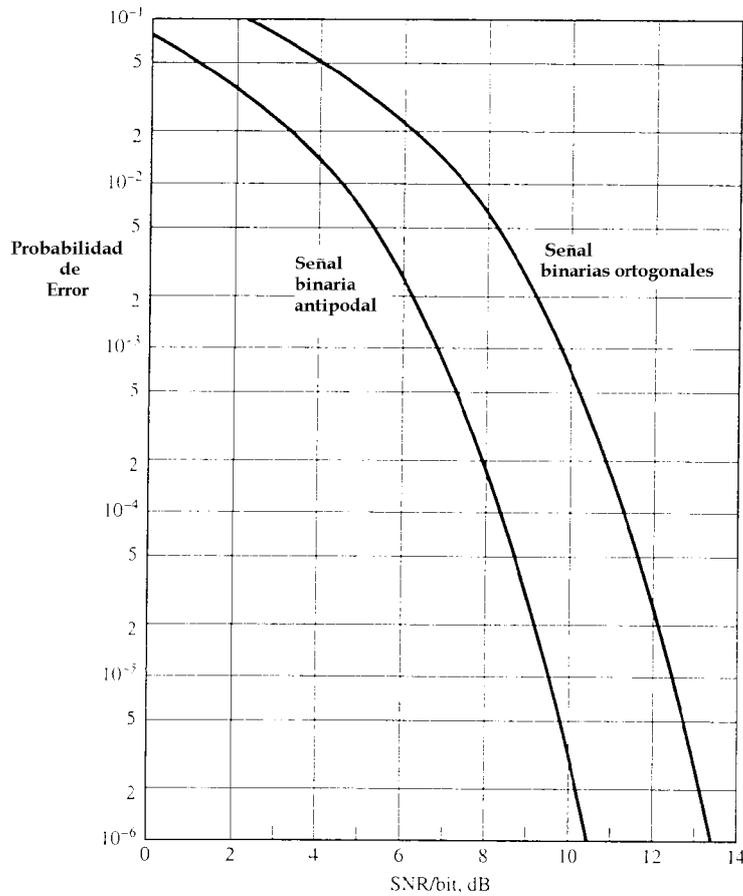


Figura 5.41: Probabilidad de error para señales binarias.

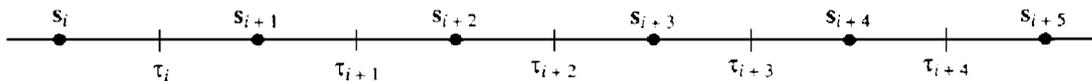


Figura 5.42: Ubicación de umbrales en los puntos medios de niveles de amplitud sucesivos.

La probabilidad de error promedio para PAM  $M$ -ária puede determinarse de la regla de decisión que maximiza la métrica de correlación dada por (5.12). Equivalentemente, el detector compara la salida del demodulador  $r$  con el conjunto de  $M - 1$  umbrales colocados en el punto medio de los sucesivos niveles de amplitud, como mostrado en la figura 5.42. Así, una decisión se hace en favor de un nivel de amplitud siempre que esté más próximo de  $r$ .

La ubicación de los umbrales como mostrado en la figura 5.42 ayuda en la evaluación de la probabilidad de error. Es posible notar que si se transmite el  $m$ -ésimo nivel de amplitud la salida del demodulador es

$$r = s_m + n = \sqrt{\mathcal{E}_g} A_m + n$$

donde la variable de ruido  $n$  tiene media cero y varianza  $N_0/2$ . Sobre la base que todos los niveles de amplitud tienen igual probabilidad a priori, la probabilidad promedio de error de símbolo es simplemente la probabilidad que la variable de ruido  $n$  exceda en magnitud la mitad de la distancia entre niveles. Sin embargo, cuando uno de los dos niveles  $\pm(M - 1)$  halla sido transmitido, un error ocurrirá en una dirección solamente. Así se tiene

$$P_M = \frac{M - 1}{M} P(|r - s_m| > \sqrt{\mathcal{E}_g}) = \frac{M - 1}{M} \frac{2}{\sqrt{\pi N_0}} \int_{\sqrt{\mathcal{E}_g}}^{\infty} e^{-x^2/N_0} dx$$

$$\begin{aligned}
&= \frac{M-1}{M} \frac{2}{\sqrt{2\pi}} \int_{\sqrt{2\mathcal{E}_g}}^{\infty} e^{-x^2/2} dx \\
&= \frac{2(M-1)}{M} Q \left[ \sqrt{\frac{2\mathcal{E}_g}{N_0}} \right]
\end{aligned}$$

Esta probabilidad de error también puede expresarse en términos de la potencia promedio transmitida, de forma que teniendo en cuenta que  $\mathcal{E}_g = \frac{3}{M^2-1} P_{av} T$ , sustituyendo en la ecuación anterior se obtiene

$$P_M = \frac{2(M-1)}{M} Q \left[ \sqrt{\frac{6P_{av}T}{(M^2-1)N_0}} \right]$$

o equivalentemente

$$P_M = \frac{2(M-1)}{M} Q \left[ \sqrt{\frac{6\mathcal{E}_{av}}{(M^2-1)N_0}} \right]$$

Al graficar la probabilidad de error de símbolo para señales  $M$ -árias, como en el caso de PAM multinivel, es costumbre utilizar la SNR por bit como parámetro básico. Dado que  $T = kT_b$  y  $k = \log_2 M$ , la ecuación anterior puede expresarse como

$$P_M = \frac{2(M-1)}{M} Q \left[ \sqrt{\frac{6(\log_2 M)\mathcal{E}_{bav}}{(M^2-1)N_0}} \right]$$

donde  $\mathcal{E}_{bav} = P_{av}T_b$  es la energía promedio por bit y  $\mathcal{E}_{bav}/N_0$  es la SNR promedio por bit. La figura 5.43 ilustra la probabilidad de error de símbolo como una función de  $10 \log_{10} \mathcal{E}_{bav}/N_0$  con  $M$  como parámetro. Notar que el caso  $M = 2$  corresponde a la probabilidad de error de las señales binarias antipodales. Es posible observar también que la SNR por bit aumenta en 4 dB por cada factor de dos de incremento en  $M$ . Para  $M$  grande, la SNR por bit adicional requerida para aumentar  $M$  en un factor de dos se aproxima a 6 dB.

**Señales Ortogonales Multinivel:** Como ya fuera discutido, PPM es un método de modulación que resulta en señales ortogonales cuyo espacio de representación está dado por (5.5). Notar que todas las señales tienen la misma energía  $\mathcal{E}_s$ .

Para señales ortogonales de igual energía, el detector óptimo selecciona la señal resultante de la mayor correlación cruzada entre el vector recibido  $\mathbf{r}$  y cada una de los  $M$  posibles vectores transmitidos  $\{\mathbf{s}_m\}$ , o sea

$$C(\mathbf{r}, \mathbf{s}_m) = \mathbf{r} \cdot \mathbf{s}_m = \sum_{k=1}^M r_k s_{mk}, \quad m = 1, 2, \dots, M$$

Para evaluar la probabilidad de error, se supone que se transmitió la señal  $\mathbf{s}_1$ . Entonces el vector de señal recibido será

$$\mathbf{r} = [\sqrt{\mathcal{E}_s} + n_1, n_2, \dots, n_M]$$

donde  $n_1, n_2, \dots, n_M$  son variables aleatorias Gaussianas estadísticamente independientes, de media cero y varianza  $\sigma_n^2 = N_0/2$ . En este caso, las  $M$  salidas del demodulador serán

$$\begin{aligned}
C(\mathbf{r}, \mathbf{s}_1) &= \sqrt{\mathcal{E}_s}(\sqrt{\mathcal{E}_s} + n_1) \\
C(\mathbf{r}, \mathbf{s}_2) &= \sqrt{\mathcal{E}_s}n_2 \\
&\vdots \\
C(\mathbf{r}, \mathbf{s}_M) &= \sqrt{\mathcal{E}_s}n_M
\end{aligned}$$

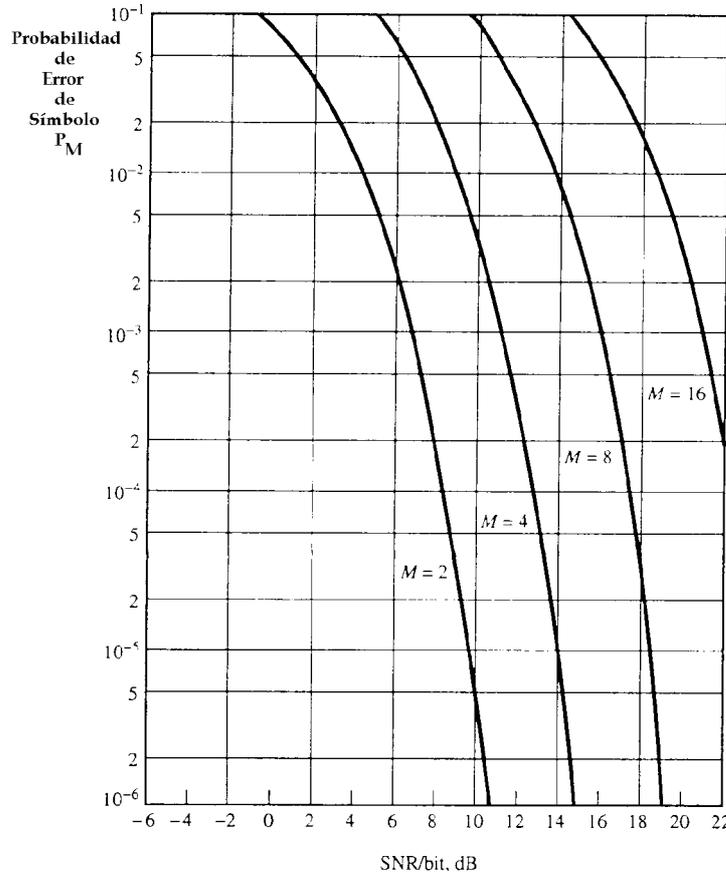


Figura 5.43: Probabilidad de un error de símbolo en PAM.

Notar que el factor de escala  $\mathcal{E}_s$  puede eliminarse de las salidas del demodulador dividiéndolas por  $\sqrt{\mathcal{E}_s}$ . Entonces, con esta normalización, la pdf de la primer salida del demodulador es

$$f_{r_1}(x_1) = \frac{1}{\sqrt{\pi N_0}} e^{-(x_1 - \sqrt{\mathcal{E}_s})^2 / N_0} \quad (5.21)$$

y las pdf de las  $M - 1$  salidas restantes será

$$f_{r_m}(x_m) = \frac{1}{\sqrt{\pi N_0}} e^{-x_m^2 / N_0}, \quad m = 2, 3, \dots, M$$

Es matemáticamente conveniente obtener primero la probabilidad de detección correcta, o sea, es la probabilidad que  $r_1$  sea mayor que cualquiera de las  $M - 1$  restantes salidas  $n_2, n_3, \dots, n_M$ . Esta probabilidad puede expresarse como

$$P_c = \int_{-\infty}^{\infty} P(n_2 < r_1, n_3 < r_1, \dots, n_M < r_1 | r_1) f(r_1) dr_1$$

donde  $P(n_2 < r_1, n_3 < r_1, \dots, n_M < r_1 | r_1)$  es la probabilidad conjunta que  $n_2, n_3, \dots, n_M$  sean todas menores que  $r_1$  condicionadas sobre  $r_1$ . Luego, esta probabilidad conjunta puede promediarse sobre todo  $r_1$ . Dado que los  $\{r_m\}$  son estadísticamente independientes, la probabilidad conjunta se obtiene como el producto de  $M - 1$  probabilidades marginales de la forma

$$P(n_m < r_1 | r_1) = \int_{-\infty}^{r_1} f_{r_m}(x_m) dx_m, \quad m = 2, 3, \dots, M$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{2r_1^2/N_0}} e^{-x^2/2} dx \\
&= 1 - Q \left[ \sqrt{\frac{2r_1^2}{N_0}} \right]
\end{aligned}$$

Estas probabilidades son idénticas para  $m = 2, 3, \dots, M$ , y en consecuencia, la probabilidad conjunta en consideración es simplemente el resultado anterior elevado a la  $(M - 1)$ -ésima potencia. Así, la probabilidad de decisión correcta es

$$P_c = \int_{-\infty}^{\infty} \left( 1 - Q \left[ \sqrt{\frac{2r_1^2}{N_0}} \right] \right)^{M-1} f(r_1) dr_1$$

y la probabilidad de un error de símbolo es

$$P_M = 1 - P_c \quad (5.22)$$

donde  $P_c = \int_{-\infty}^{\infty} \left\{ 1 - \left( Q \left[ \sqrt{x^2} \right] \right)^{M-1} \right\} e^{-(x - \sqrt{2\mathcal{E}_s/N_0})^2/2} dx$ . La misma expresión para la probabilidad de error se obtiene cuando cualquiera de las otras  $M - 1$  señales se transmite. Dado que las  $M$  señales son igualmente probables, la expresión (5.22) es la probabilidad promedio de error por símbolo. Esta expresión puede evaluarse numericamente.

Para comparar el desempeño de las diferentes técnicas de modulación es deseable tener la probabilidad de error expresada en términos de la SNR por bit,  $\mathcal{E}_b/N_0$ , en lugar de la SNR por símbolo  $\mathcal{E}_s = k\mathcal{E}_b$ .

A veces es deseable convertir la probabilidad de error de símbolo en la probabilidad equivalente de error de bit. Para señales ortogonales equiprobables todos los símbolos son equiprobables y ocurren con probabilidad

$$\frac{P_M}{M-1} = \frac{P_M}{2^k-1}$$

Además, existen  $\binom{k}{n}$  formas en los cuales  $n$  bits de  $k$  pueden estar en error. De esta forma, el número promedio de errores de bit para un símbolo de  $k$  bits es

$$\sum_{n=1}^k n \binom{k}{n} \frac{P_M}{2^k-1} = k \frac{2^{k-1}}{2^k-1} P_M$$

y la probabilidad promedio de error de bit es este resultado dividido por  $k$ , el número de bits por símbolo. Así

$$P_b = \frac{2^{k-1}}{2^k-1} P_M \cong \frac{P_M}{2} \text{ si } k \gg 1$$

Los gráficos de la probabilidad de error de bit en función de la SNR por bit se muestran en la figura 5.44 para  $M = 2, 4, 8, 16, 32$  y  $64$ . Esta figura ilustra que incrementando el número de señales es posible reducir la SNR por bit requerida para lograr una probabilidad de error de bit. Por ejemplo, para lograr una  $P_b = 10^{-5}$ , la SNR por bit es un poco mayor que 12 dB para  $M = 2$ , pero incrementando a  $M = 64$  ( $k = 6$  bits/símbolo), la SNR por bit requerida es aproximadamente 6 dB. Así, se obtienen ahorros de 6 dB (o un factor de reducción de 4) en la potencia transmitida necesaria para lograr la probabilidad de error especificada. ¿Cuál es la SNR por bit mínima requerida para lograr una probabilidad de error arbitrariamente pequeña cuando  $M$  toma valores infinitamente grandes? Esta pregunta se responderá a continuación.

**Una cota sobre la probabilidad de error.** Se investigará el efecto de incrementar  $M$  sobre la probabilidad de error para señales ortogonales. Para simplificar el desarrollo matemático, primero se obtendrá una cota

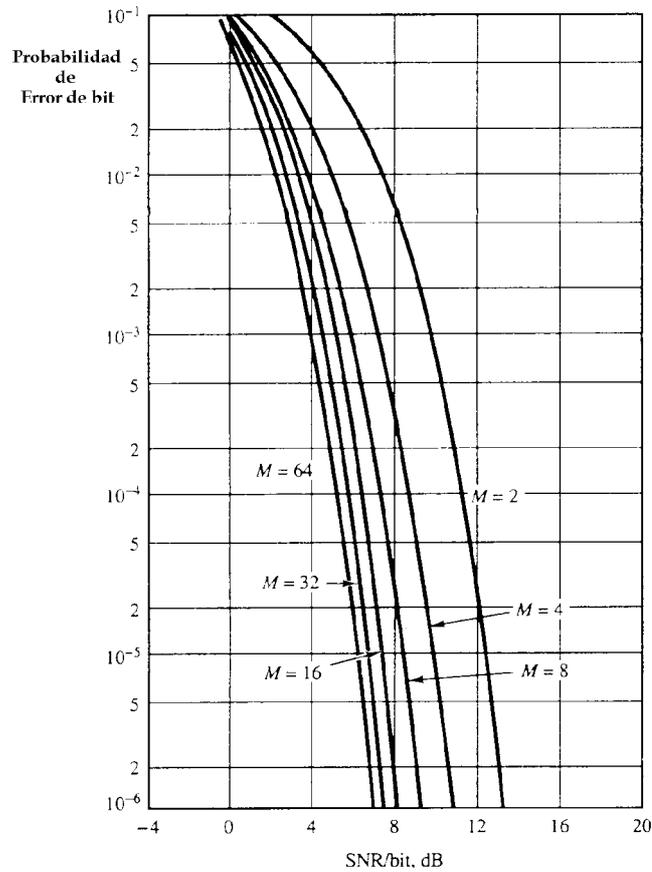


Figura 5.44: Probabilidad de error de bit para detección coherente de señales ortogonales.

superior sobre la probabilidad de error de símbolo la cual es mucho más simple que la forma exacta dada por (5.22).

Teniendo en cuenta la probabilidad de error para señales ortogonales binarias, dada por (5.20), y si se interpreta el detector para  $M$  señales ortogonales como uno que toma  $M - 1$  decisiones binarias entre la salida del correlador  $C(\mathbf{r}, \mathbf{s}_1)$ , que contiene la señal, y las restantes  $M - 1$  salidas de correladores  $C(\mathbf{r}, \mathbf{s}_m)$ ,  $m = 2, 3, \dots, M$ , la probabilidad de error está acotada por encima por la *cota de unión* de los  $M - 1$  eventos. O sea, si  $E_i$  representa el evento  $C(\mathbf{r}, \mathbf{s}_m) > C(\mathbf{r}, \mathbf{s}_1)$  para  $m \neq 1$ , entonces se tiene que  $P_M = P(\cup_{i=1}^n E_i) \leq \sum_{i=1}^n P(E_i)$ . De esta forma,

$$P_M \leq (M - 1)P_2 = (M - 1)Q \left[ \sqrt{\mathcal{E}_s/N_0} \right] < MQ \left[ \sqrt{\mathcal{E}_s/N_0} \right]$$

Esta cota puede simplificarse aún más con una cota superior de  $Q \left[ \sqrt{\mathcal{E}_s/N_0} \right]$ , ya que  $Q \left[ \sqrt{\mathcal{E}_s/N_0} \right] < e^{-\mathcal{E}_s/2N_0}$ . Así,

$$P_M < M e^{-\mathcal{E}_s/2N_0} = 2^k e^{k\mathcal{E}_b/2N_0} = e^{-k(\mathcal{E}_b/N_0 - 2 \ln 2)/2} \quad (5.23)$$

Cuando  $k \rightarrow \infty$ , o equivalentemente, cuando  $m \rightarrow \infty$ , la probabilidad de error se aproxima a cero exponencialmente, siempre que  $\mathcal{E}_b/N_0$  sea mayor que  $2 \ln 2$ , o sea

$$\frac{\mathcal{E}_b}{N_0} > 2 \ln 2 = 1.39 \quad (1.42 \text{ dB})$$

Esta cota superior sobre la probabilidad de error de (5.23) implica que siempre que la SNR  $> 1.42$  dB, es posible lograr una  $P_m$  arbitrariamente baja. Sin embargo, esta cota de unión no es muy precisa para SNR

suficientemente bajas debido a que la cota superior de la función  $Q$  no lo es. En realidad, mediante técnicas más elaboradas es posible mostrar que la cota anterior es suficientemente precisa para  $\mathcal{E}_b/N_0 > 4 \ln 2$ . Para  $\mathcal{E}_b/N_0 < 4 \ln 2$ , una cota más precisa para  $P_M$  es la siguiente

$$P_M < 2e^{-k(\sqrt{\mathcal{E}_b/N_0} - \sqrt{\ln 2})^2}$$

En consecuencia,  $P_M \rightarrow 0$  para  $k \rightarrow \infty$ , siempre que

$$\frac{\mathcal{E}_b}{N_0} > \ln 2 = 0.693 \quad (-1.6\text{dB})$$

De esta forma,  $-1.6$  dB es la SNR mínima requerida por bit para lograr una probabilidad de error arbitrariamente pequeña en el límite cuando  $k \rightarrow \infty$  ( $M \rightarrow \infty$ ). Esta SNR mínima por bit ( $-1.6$  db) se denomina *límite de Shannon* para un canal AWGN.

**Señales biortogonales.** Como se discutió anteriormente, un conjunto de señales biortogonales se construye de  $M/2$  señales ortogonales incluyendo los negativos de las señales ortogonales. Así, se logra una reducción en la complejidad del demodulador ya que, mientras este conjunto utiliza  $M/2$  correladores, el conjunto ortogonal completo utiliza  $M$  correladores.

Para evaluar la probabilidad de error para un detector óptimo, es posible suponer inicialmente que fué transmitida la señal  $s_1(t)$  correspondiente al vector  $\mathbf{s}_1 = (\sqrt{\mathcal{E}_s}, 0, 0, \dots, 0)$ . Luego, el vector de señal recibido será

$$\mathbf{r} = (\sqrt{\mathcal{E}_s} + n_1, n_2, \dots, n_{M/2})$$

donde los  $\{n_m\}$  son variables aleatorias idénticamente Gaussianas, estadísticamente independientes entre sí, de media cero y varianza  $\sigma_n^2 = N_0/2$ . El detector óptimo decide en favor de la señal correspondiente a la de mayor amplitud de salida de los correladores, dada por

$$C(\mathbf{r}, \mathbf{s}_m) = \mathbf{r} \cdot \mathbf{s}_m = \sum_{k=1}^{M/2} r_k s_{mk}, \quad m = 1, 2, \dots, M/2$$

mientras que el signo del mayor término se utiliza para decidir cual de  $s_m(t)$  ó  $-s_m(t)$  fué transmitida. De acuerdo a esta regla de decisión, la probabilidad de una decisión correcta es igual a la probabilidad que  $r_1 = \sqrt{\mathcal{E}_s} + n_1 > 0$  y  $r_1$  excede  $|r_m| = |n_m|$  para  $m = 2, 3, \dots, M/2$ . Pero

$$P(|n_m| < r_1 | r_1 > 0) = \frac{1}{\sqrt{\pi N_0}} \int_{-r_1}^{r_1} e^{-x^2/N_0} dx = \frac{1}{\sqrt{2\pi}} \int_{-r_1/\sqrt{N_0/2}}^{r_1/\sqrt{N_0/2}} e^{-x^2/2} dx$$

Luego, la probabilidad de una decisión correcta es

$$P_c = \int_0^\infty \left[ \frac{1}{\sqrt{2\pi}} \int_{-r_1/\sqrt{N_0/2}}^{r_1/\sqrt{N_0/2}} e^{-x^2/2} dx \right]^{\frac{M}{2}-1} f(r_1) dr_1$$

en la cual sustituyendo  $f(r_1)$  se obtiene

$$P_c = \int_{-\sqrt{2\mathcal{E}_s/N_0}}^\infty \left[ \frac{1}{\sqrt{2\pi}} \int_{-(v+\sqrt{2\mathcal{E}_s/N_0})}^{v+\sqrt{2\mathcal{E}_s/N_0}} e^{-x^2/2} dx \right]^{\frac{M}{2}-1} e^{-v^2/2} dv$$

donde se utilizó la pdf de  $r_1$  dada por (5.21). Finalmente la probabilidad de un error de símbolo  $P_M = 1 - P_c$ .

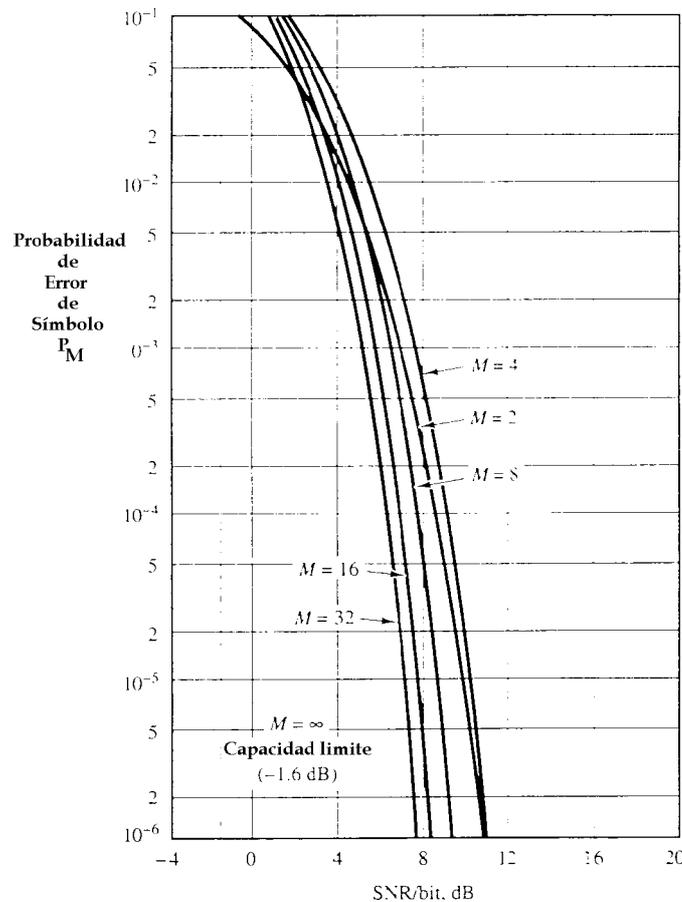


Figura 5.45: Probabilidad de error de símbolo para señales biortogonales.

$P_c$ , y en consecuencia  $P_M$ , puede evaluarse numericamente para diferentes valores de  $M$  a partir de esta última ecuación. El gráfico de la figura 5.45 ilustra  $P_m$  como una función de  $\mathcal{E}_b/N_0$ , donde  $\mathcal{E}_s = k\mathcal{E}_b$ , para  $M = 2, 4, 8, 16$  y  $32$ . Se observa que este gráfico es similar al de las señales ortogonales. Sin embargo, en este caso la probabilidad de error para  $M = 4$  es mayor que para  $M = 2$ . Esto se debe a que en esta figura se graficó la probabilidad de error de símbolo  $P_M$ . Si graficáramos la probabilidad de error de bit equivalente se podría verificar que los gráficos para  $M = 2$  y  $M = 4$  coinciden. Como en el caso de las señales ortogonales, cuando  $M \rightarrow \infty$  (ó  $k \rightarrow \infty$ ), la SNR por bit requerida para alcanzar una probabilidad de error arbitrariamente baja es  $-1.6$  dB, o sea el límite de Shannon.

**Señales simplex** Teniendo en cuenta la primera sección de este Capítulo, las señales simplex son un conjunto de  $M$  señales igualmente correlacionadas con coeficiente de correlación cruzada  $\gamma_{mn} = -1/(M-1)$ . Esas señales tienen la misma separación mínima de  $\sqrt{2\mathcal{E}_s}$  entre puntos de señal adyacentes en un espacio  $M$  dimensional como señales ortogonales. Se logra esta separación mutua con una energía transmitida de  $\mathcal{E}_s(M-1)/M$ , la cual es menor que la requerida para señales ortogonales en un factor de  $(M-1)/M$ . En consecuencia, la probabilidad de error para señales simplex es idéntica a la probabilidad de error para señales ortogonales, pero su desempeño se logra con un ahorro de

$$10 \log(1 - \gamma_{mn}) = 10 \log \frac{M}{M-1} \quad \text{dB}$$

en la SNR.

### 5.3.3 Probabilidad de error para MLSD

En la discusión de MLSD, la detección de secuencias involucra la comparación de caminos múltiples de señal que convergen en cada nodo de una trellis. Cuando el ruido aditivo es suficientemente grande para provocar

errores en el proceso de detección, los errores involucran generalmente caminos de señal de dos o más bits o símbolos de información. La selección del camino de señal incorrecto en un nodo se denomina *evento de error*. En general, un evento de error comienza en un punto donde el camino superviviente diverge del camino correcto y termina cuando el camino superviviente vuelve a convergir con el camino correcto.

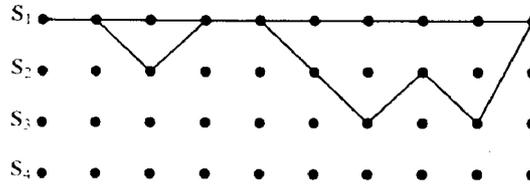


Figura 5.46: Ilustración de dos eventos de error en una búsqueda sobre la trellis.

La figura 5.46 ilustra dos caminos a través de una trellis de cuatro estados. El camino superior representa el camino correcto, el cual se supone es el camino de todos cero, y el camino inferior representa el camino seleccionado por el MLSD. Esta figura muestra dos eventos de error, uno de longitud dos y el otro de longitud cinco. Es necesario notar que dentro de cada evento de error algunos bits o símbolos pueden ser correctos mientras otros pueden ser errados. Independientemente de eso, la probabilidad de un error (de bit o símbolo) puede obtenerse ponderando adecuadamente la probabilidad de cada evento de error con la fracción correspondiente de símbolos incorrectos.

En general, existen muchos (posiblemente infinitos) eventos de error distintos, cada uno con su propia probabilidad de ocurrencia. Por esta razón es extremadamente difícil obtener una expresión exacta para la probabilidad de error. En su lugar, es usual recurrir a cotas superiores (de unión) sobre el desempeño de error.

La aproximación más simple a la probabilidad de error se obtiene considerando eventos de error distintos y sus distancias Euclidianas correspondientes. Luego, la probabilidad de error se estima adecuadamente (acotada en forma superior) mediante la suma ponderada de las probabilidades de error de los distintos eventos de error, donde los factores de ponderación son el número de eventos de error distintos en las distancias Euclidianas correspondientes, o sea

$$P_e \leq \sum_i N_{d_i} Q \left[ \sqrt{\frac{d_i^2}{2N_0}} \right]$$

donde  $d_i^2$  es la distancia Euclidiana del  $i$ -ésimo camino relativo al camino correcto, y  $N_{d_i}$  es el número de caminos a distancia  $d_i$ .

Cuando las distancias  $\{d_i\}$  de los eventos de error varían mucho con  $i$ , es posible despreciar aquellos términos con grandes distancias en favor de los términos con menores distancias. Por supuesto, la aproximación más simple a la probabilidad de error es utilizar solo el término dominante en la suma de la ecuación anterior, correspondiente a la distancia Euclidiana mínima,  $d_{min}$ , o sea

$$P_e \cong N_{d_{min}} Q \left[ \sqrt{\frac{d_{min}^2}{2N_0}} \right]$$

donde  $N_{d_{min}}$  es el número de eventos de error con distancia  $d_{min}$ .

Para demostrar su cómputo, es posible considerar la trellis para la señal NRZI. La figura 5.47 ilustra la trellis e indica dos eventos de error distintos. El camino superior representa el camino correcto. Se evaluará la probabilidad de los dos eventos de error. La probabilidad del evento de error de longitud dos es simplemente la probabilidad que  $\mu_2(1,1) < \mu_2(0,0)$ , donde  $\mu_2(1,1)$  y  $\mu_2(0,0)$  está dadas por (5.18) y  $r_1 = -\sqrt{\mathcal{E}_b} + n_1$  y  $r_2 = -\sqrt{\mathcal{E}_b} + n_2$ . En consecuencia,

$$P[\mu_2(1,1) < \mu_2(0,0)] = P(n_1 > \sqrt{\mathcal{E}_b})$$

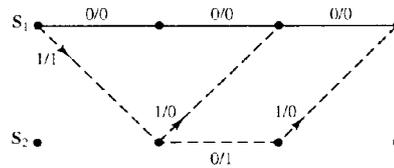


Figura 5.47: Dos eventos de error en una trellis para una señal NRZI.

$$\begin{aligned}
 &= \frac{1}{\sqrt{2\pi}\sigma_n} \int_{\sqrt{\mathcal{E}_b}}^{\infty} e^{-x^2/2\sigma_n^2} dx \\
 &= Q \left[ \sqrt{\frac{2\mathcal{E}_b}{N_0}} \right]
 \end{aligned}$$

donde  $\sigma_n^2 = N_0/2$ . Es interesante notar que esta es exactamente la misma probabilidad de error de un detector de símbolos que ignora la memoria inherente a la señal NRZI. Así, la señal NRZI es un ejemplo de una señal con memoria en la cual la distancia Euclidiana para un camino de longitud dos no se incrementa en relación a la distancia Euclidiana de un camino de longitud uno. Como consecuencia, no existe ventaja en el desempeño obtenido al efectuar MLSD comparada con la detección por símbolo.

La probabilidad de un evento de error de longitud tres es la probabilidad que  $\mu_2(1, 0, 1) < \mu_2(0, 0, 0)$ , donde

$$\begin{aligned}
 \mu_3(0, 0, 0) &= (r_1 + \sqrt{\mathcal{E}_b})^2 + (r_2 + \sqrt{\mathcal{E}_b})^2 + (r_3 + \sqrt{\mathcal{E}_b})^2 \\
 \mu_3(1, 0, 1) &= (r_1 - \sqrt{\mathcal{E}_b})^2 + (r_2 - \sqrt{\mathcal{E}_b})^2 + (r_3 + \sqrt{\mathcal{E}_b})^2
 \end{aligned}$$

donde  $r_i = -\sqrt{\mathcal{E}_b} + n_i$ ,  $i = 1, 2, 3$ . En consecuencia,

$$\begin{aligned}
 P[n_1 + n_2 > 2\sqrt{\mathcal{E}_b}] &= \frac{1}{\sqrt{2\pi}} \int_{\sqrt{4\mathcal{E}_b/N_0}}^{\infty} e^{-x^2/2} dx \\
 &= Q \left[ \sqrt{\frac{4\mathcal{E}_b}{N_0}} \right]
 \end{aligned}$$

Se puede observar que este evento de error ocurre con una probabilidad de error significativamente menor si comparado con el evento de longitud dos.

En general, eventos de error largos tendrán una gran distancia Euclidiana entre ellos y el camino correcto, y de esta forma, sus probabilidades de ocurrencia serán significativamente menores. En consecuencia, es posible concluir que en general, la probabilidad de error para el MLSD estará dominada por los eventos de error con la distancia Euclidiana mínima  $d_{min}$ .

### 5.3.4 Comparaciones

Los métodos de modulación digital descritos en este Capítulo pueden ser comparados en diferentes formas. Por ejemplo, es posible compararlos sobre la base de la SNR requerida para lograr una probabilidad de error especificada. Sin embargo, tal comparación puede no ser significativa a menos que sea realizada sobre la base de alguna restricción tal como una velocidad de transmisión fija.

Es posible suponer que la velocidad de transmisión en bits es fija  $R_b$ , y considerar el ancho de banda de canal requerido para transmitir las diferentes señales. Si empleamos PAM  $M$ -ária, donde  $M = 2^k$ , el intervalo de símbolo es  $T = k/R_b$ . Un pulso de señal de duración  $T$  tiene una característica espectral de ancho  $1/2T$  (aproximadamente). De esta forma, el ancho de banda de canal requerido para transmitir una señal PAM  $M$ -ária es

$$W = R_b/2k = R_b/2 \log_2 M \quad Hz$$

En PPM, el intervalo de símbolo  $T$  se subdivide en  $M$  subintervalos de duración  $T/M$  y se transmiten pulsos de ancho  $T/M$  en los subintervalos correspondientes. En consecuencia, el espectro de cada pulso tiene aproximadamente un ancho  $M/2T$ . Así, el ancho de banda de canal requerido para transmitir señales PPM es

$$W = \frac{M}{2T} = \frac{M}{2(k/R_b)} = \frac{MR_b}{2 \log_2 M} \text{ Hz}$$

Las señales biortogonales y simplex se obtienen a partir de relaciones similares a PPM (ortogonal). En el caso de las señales biortogonales, el ancho de banda requerido es la mitad del asociado a las señales ortogonales.

Una comparación compacta y significativa de esos métodos de modulación es aquella basada en la velocidad normalizada  $R_b/W$  (bits por segundo por Hertz de ancho de banda) en función de la SNR por bit ( $\mathcal{E}_b/N_0$ ) requerida para lograr una probabilidad de error especificada. Para señales PAM y ortogonales, se tiene que

$$\text{PAM: } \frac{R_b}{W} = 2 \log_2 M$$

$$\text{PPM (ortogonal): } \frac{R_b}{W} = \frac{2 \log_2 M}{M}$$

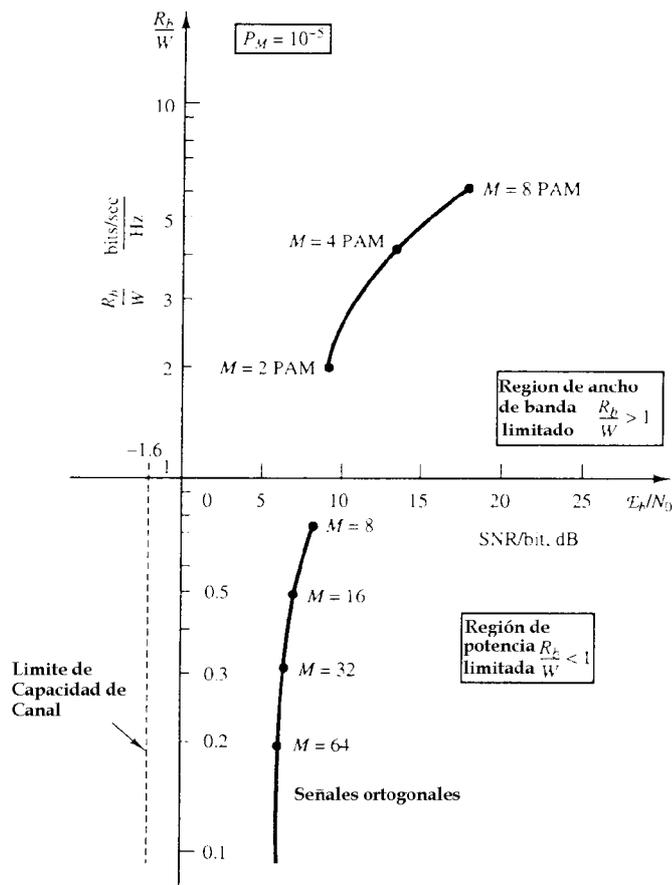


Figura 5.48: Comparación de varios métodos de modulación para una probabilidad de error  $10^{-5}$ .

La figura 5.48 ilustra el gráfico de  $R_b/W$  versus SNR por bit para señales PAM y señales ortogonales para el caso en el que la probabilidad de error de símbolo es  $P_M = 10^{-5}$ . Se observa que en el caso de PAM, incrementar el número de amplitudes  $M$  resulta en una mayor  $R_b/W$ . Sin embargo, el costo de lograr esta mayor relación es incrementar la SNR por bit. En consecuencia, la modulación PAM  $M$ -ária es apropiada para canales de comunicación que están limitados en ancho de banda, donde se desea una  $R_b/W > 1$  y

donde existe una SNR suficiente para soportar los niveles de amplitud múltiples. Los canales telefónicos son ejemplos de tales canales de ancho de banda limitado.

En contraste, las señales ortogonales  $M$ -árias producen una  $R_b/W \leq 1$ . Cuando  $M$  aumenta,  $R_b/W$  disminuye debido al incremento en el ancho de banda de canal requerido. Sin embargo, la SNR por bit requerida para lograr una probabilidad de error (en este caso  $P_M = 10^{-5}$ ) disminuye cuando  $M$  aumenta. En consecuencia, las señales ortogonales  $M$ -árias son apropiadas para canales de potencia limitada que tienen un ancho de banda suficientemente grande para acomodar suficiente número de señales. En este caso, cuando  $M \rightarrow \infty$  la probabilidad de error puede hacerse tan pequeña como deseado, previendo que  $\mathcal{E}_b/N_0 > 0.693$  (-1.6 dB). Esta es la mínima SNR por bit requerida para lograr una transmisión confiable en el límite cuando el ancho de banda del canal tiende a infinito y la correspondiente  $R_b/W \rightarrow 0$ .

## 5.4 Repetidores regenerativos y análisis de potencia de un enlace (\*)

Se ha observado que en la transmisión de señales digitales a través de un canal AWGN el desempeño del sistema de comunicaciones, medido en términos de la probabilidad de error, depende de la SNR recibida,  $\mathcal{E}_b/N_0$ , donde  $\mathcal{E}_b$  es la energía transmitida por bit y  $N_0/2$  es la densidad espectral de potencia del ruido aditivo. En consecuencia, el ruido aditivo limitará el desempeño del sistema de comunicaciones.

Además del ruido aditivo, otro factor que afecta el desempeño de un sistema de comunicaciones es la atenuación del canal. Como indicado anteriormente todos los canales físicos, cableados o inalámbricos tienen pérdidas. De esta forma, la señal se atenúa a través del canal. El modelo matemático simple con atenuación es útil ahora para el diseño del sistema de comunicaciones digitales. En consecuencia, si la señal transmitida es  $s(t)$ , la señal recibida será

$$r(t) = \alpha s(t) + n(t)$$

Entonces, si la energía en la señal transmitida es  $\mathcal{E}_b$ , la energía en la señal recibida será  $\alpha\mathcal{E}_b$ . En consecuencia, la señal recibida tendrá una relación señal a ruido  $\alpha\mathcal{E}_b/N_0$ . Como en el caso de los sistemas de comunicaciones analógicos, el efecto de la atenuación de la señal en comunicaciones digitales es reducir la energía de la señal recibida y de esta forma hacer que el sistema de comunicaciones sea más vulnerable al ruido aditivo.

Recordar que en sistemas de comunicaciones analógicos, se utilizan amplificadores denominados repetidores para mantener la potencia de la señal en la transmisión a través del canal. Sin embargo, cada amplificador también actúa sobre el ruido en el sistema. En contraste, los sistemas de comunicaciones digitales permiten detectar y regenerar una señal limpia (libre de ruido) en la transmisión a través del canal. Tales dispositivos, denominados *repetidores regenerativos*, se utilizan frecuentemente en canales cableados y de fibra óptica.

### 5.4.1 Repetidores regenerativos (\*)

La primera etapa de un repetidor regenerativo consiste en un demodulador/detector que demodula y detecta la secuencia de información digital transmitida enviada por el repetidor precedente. Una vez detectada la secuencia pasa a la etapa de transmisión del repetidor que mapea la secuencia en señales que son transmitidas al próximo repetidor. Este tipo de repetidor se denomina repetidor regenerativo.

Como se regenera una señal libre de ruido en cada repetidor, el ruido aditivo no se acumula. Sin embargo, cuando existen errores en el detector de un repetidor, los errores se propagarán hacia adelante a los repetidores siguientes en el canal. Para evaluar el efecto de los errores sobre el desempeño del sistema completo es posible suponer que la modulación es PAM binaria, tal que la probabilidad de un bit de error para un salto (transmisión entre dos repetidores) es

$$P_2 = Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right)$$

Como los errores ocurren con baja probabilidad es posible ignorar la probabilidad que cualquier bit sea detectado incorrectamente más de una vez en la transmisión a través del canal con  $K$  repetidores. En

consecuencia, el número de errores aumentará linealmente con el número de repetidores regenerativos en el canal y por ello la probabilidad de error total puede aproximarse como

$$P_b \cong KQ \left( \sqrt{\frac{2\mathcal{E}_b}{N_0}} \right) \quad (5.24)$$

En contraste, la utilización de  $K$  repetidores analógicos en el canal reduce la SNR recibida en  $K$ , y en consecuencia, la probabilidad de error de bit será

$$P_b \cong Q \left( \sqrt{\frac{2\mathcal{E}_b}{KN_0}} \right) \quad (5.25)$$

Claramente, para el mismo desempeño en términos de probabilidad de error, el uso de repetidores regenerativos resultará en ahorros significativos en potencia transmitida si comparados con repetidores analógicos. De esta forma, los repetidores regenerativos son preferibles en sistemas de comunicaciones digitales. Sin embargo, en canales cableados telefónicos que se utilizan para transmitir señales analógicas y digitales se emplean generalmente repetidores analógicos.

---

**Ejemplo.** Un sistema de comunicaciones digitales binario transmite datos sobre un canal cableado de 1000 km de longitud. Se utilizan repetidores cada 10 km para compensar el efecto de la atenuación del canal. Determinar la  $\mathcal{E}_b/N_0$  requerida para lograr una probabilidad de un bit de error de  $10^{-5}$  si: a) se utilizan repetidores analógicos y b) se utilizan repetidores regenerativos.

El número de repetidores utilizados en el sistema es  $K = 100$ . Si se utilizan repetidores regenerativos, de (5.24) se obtiene que  $\mathcal{E}_b/N_0$  es aproximadamente 11.3 dB. Si se utiliza repetidores analógicos entonces la  $\mathcal{E}_b/N_0$  obtenida de (5.25) es 18.3 dB, o aproximadamente 70 veces la potencia transmitida requerida para el sistema de comunicaciones digitales.

---

#### 5.4.2 Balance de potencia para canales de radio (\*)

En el diseño de sistemas de comunicaciones de radio que transmiten sobre canales de microondas de enlace visual y canales satelitales se debe considerar también el efecto de las características de la antena para determinar la SNR en el receptor requerida para lograr un nivel especificado de desempeño. El procedimiento de diseño del sistema se describe a continuación.

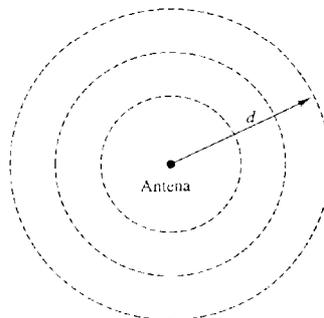


Figura 5.49: Antena que irradia isotrópicamente en el espacio libre.

Es posible suponer que una antena radía isotrópicamente en el espacio libre a un nivel de potencia de  $P_T$  watts, como ilustrado en la figura 5.49. La densidad de potencia a una distancia  $d$  de la antena es  $P_T/4\pi d^2$  w/m<sup>2</sup>. Si la antena transmisora tiene directividad en una dirección determinada, la densidad de potencia en esa dirección aumenta por un factor denominado la *ganancia de antena*  $G_T$ . Luego, la densidad de potencia a una distancia  $d$  será  $P_T G_T/4\pi d^2$  w/m<sup>2</sup>. El producto  $P_T G_T$  se denomina usualmente *potencia isotrópica*

*efectiva radiada* (EIRP), la cual es básicamente la potencia radiada relativa a una antena isotrópica para la cual  $G_T = 1$ .

Una antena receptora orientada en la dirección de la potencia irradiada absorbe una porción de potencia que es proporcional al área de sección cruzada. De esta forma, la potencia extraída recibida por la antena puede expresarse como

$$P_R = \frac{P_T G_T A_R}{4\pi d^2} \quad (5.26)$$

donde  $A_R$  es el *área efectiva de la antena*. La relación básica entre la ganancia de la antena  $G_R$  y su área efectiva, obtenida de la teoría básica de electromagnetismo, es

$$A_R = \frac{G_R \lambda^2}{4\pi} \quad m^2 \quad (5.27)$$

donde  $\lambda$  es la longitud de onda de la señal transmitida. Sustituyendo  $A_R$  de (5.27) en (5.26) se obtiene la expresión para la potencia recibida dada por

$$P_R = \frac{P_T G_T G_R}{(4\pi d/\lambda)^2} \quad (5.28)$$

El factor  $\mathcal{L}_s = (4\pi d/\lambda)^2$  es la pérdida de atenuación en el espacio libre. Otras pérdidas que puedan ser encontradas, tales como las atmosféricas, se incluyen en un factor de pérdidas adicional  $\mathcal{L}_a$ . En consecuencia, la potencia recibida puede expresarse como

$$P_R = \frac{P_T G_T G_R}{\mathcal{L}_s \mathcal{L}_a} \quad (5.29)$$

o equivalentemente

$$P_R|_{dB} = P_T|_{dB} + G_T|_{dB} + G_R|_{dB} - \mathcal{L}_s|_{dB} - \mathcal{L}_a|_{dB}$$

El área efectiva de una antena depende generalmente de la longitud de onda  $\lambda$  de la potencia radiada y las dimensiones físicas de la antena. Por ejemplo, una antena parabólica de diámetro  $D$  tiene un área efectiva de  $A_R = \eta \pi D^2/4$ , donde  $\pi D^2/4$  es el área física y  $\eta$  es el *factor de eficiencia de iluminación*, que está típicamente en el rango de  $0.5 \leq \eta \leq 0.6$ . De esta forma, la ganancia de antena para una antena parabólica de diámetro  $D$  será

$$G_R = \eta \left( \frac{\pi D}{\lambda} \right)^2$$

Como un segundo ejemplo, una antena bocina de área física  $A$  tiene un factor de eficiencia de 0.8, un área efectiva  $A_R = 0.8A$ , y una ganancia de

$$G_R = \frac{10A}{\lambda^2}$$

Otro parámetro que está relacionado con la ganancia (directividad) de una antena es su *ancho de haz*,  $\Theta_B$ , ilustrado en la figura 5.50. Usualmente, el ancho de haz se mide como el ancho de -3 dB del patrón de la antena. Por ejemplo, el ancho de -3 dB para una antena parabólica es aproximadamente

$$\Theta_B \cong 70\lambda/D \quad \text{grados}$$

tal que  $G_T$  es inversamente proporcional a  $\Theta_B^2$ . De esta forma, un aumento de ancho del haz por un factor de dos, obtenido duplicando el diámetro, aumenta la ganancia de antena por un factor de cuatro (6 dB).

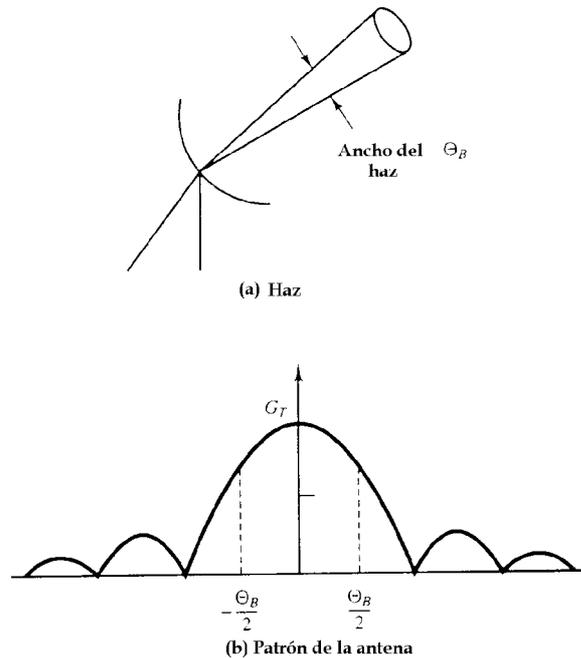


Figura 5.50: Una antena de haz estrecho y su patrón de radiación.

**Ejemplo.** Un satélite en órbita geosincrónica (36000 km sobre la superficie de la tierra) radia 100 watts de potencia (20 dBW). La antena transmisora tiene una ganancia de 18 dB, tal que la EIRP = 38 dBW. La estación terrestre utiliza una antena parabólica de 3 m. y la frecuencia de trabajo es de 4 GHz. Determinar la potencia recibida.

La longitud de onda es  $\lambda = 0.75$  m. De forma que la atenuación de espacio libre será

$$\mathcal{L}_s|_{dB} = 20 \log \left( \frac{4\pi d}{\lambda} \right) = 195.6 \text{ dB}$$

Suponiendo que  $\eta = 0.5$ , la ganancia de la antena será de 39 dB. Dado que no existe otra pérdida

$$P_R|_{dB} = 20 + 18 + 39 - 195.6 = -118.6 \text{ dBW}$$

o equivalentemente  $P_R = 2.5 \times 10^{-11}$  watts.

Es posible llevar el cálculo un paso más allá relacionando la SNR requerida para lograr un nivel de desempeño a la  $P_R$  obtenida. Como

$$\frac{\mathcal{E}_b}{N_0} = \frac{T_b P_R}{N_0} = \frac{1}{R_b} \frac{P_R}{N_0}$$

de donde se concluye que

$$\frac{P_R}{N_0} = R_b \left( \frac{\mathcal{E}_b}{N_0} \right)_{rec}$$

donde  $(\mathcal{E}_b/N_0)_{rec}$  es la SNR requerida por bit para lograr el desempeño deseado. La relación de la ecuación anterior permite determinar la velocidad de transmisión  $R_b$ , o sea

$$10 \log_{10} R_b = \left( \frac{P_R}{N_0} \right)_{dB} - \log_{10} \left( \frac{\mathcal{E}_b}{N_0} \right)_{rec} \quad (5.30)$$

---

**Ejemplo.** Si  $(\mathcal{E}_b/N_0)_{rec} = 10$  dB, determinar la velocidad de transmisión para el sistema de comunicaciones satelital del ejemplo anterior. Suponer que la primera etapa del receptor tiene  $N_0 = 4.1 \times 10^{-21}$  W/Hz, o equivalentemente, -203.9 dBW/Hz.

Luego,

$$\left(\frac{P_R}{N_0}\right)_{dB} = -118.6 + 203.9 = 85.3 \text{ dB} - \text{Hz}$$

De (5.30) se obtiene que

$$10 \log_{10} R_b = 85.3 - 10 = 75.3$$

o equivalentemente,

$$R_b = 33.9 \times 10^6 \text{ b/seg}$$

---