

Capítulo 7

Técnicas de modulación digital pasabanda

- Modulación de amplitud (ASK)
- Modulación de fase (PSK)
- Modulación de amplitud en cuadratura (QAM)
- Modulación en frecuencia (FSK)
- Modulación de fase continua (CPM)
- Transmisión digital en canales con desvanecimiento y multicamino

En los dos últimos capítulos se describieron métodos de modulación digital para canales banda base, o sea, canales con una característica en frecuencia pasabajos. Las señales moduladas digitalmente con características espectrales pasabajos pueden transmitirse directamente a través de esos canales sin la necesidad de traslación en frecuencia.

Existen varios canales de comunicaciones, incluyendo canales telefónicos, canales de radio y canales satelitales que permiten el pasaje de señales con una banda de frecuencias que está lejos del origen. Tales canales se denominan *pasabanda*. Como en el caso de la transmisión analógica, la información digital puede transmitirse a través de esos canales utilizando una portadora sinusoidal que es modulada por la secuencia de información en amplitud, fase o frecuencia, o alguna combinación de amplitud y fase. El efecto de asociar la señal de información a uno o más parámetros de la sinusoidal es desplazar el contenido de frecuencias de la señal transmitida a una banda de frecuencias apropiada que pasará por el canal.

En este capítulo se estudiarán varios métodos de modulación de portadora, incluyendo modulación de amplitud, fase, combinación de amplitud y fase, y modulación de frecuencia. Se considerará las características espectrales de esas señales de portadora modulada, se estudiará su demodulación óptima cuando está corrompida por AWGN y se evaluará el desempeño en términos de probabilidad de error.

7.1 Modulación de amplitud (ASK)

Del Capítulo 5 es posible recordar que para PAM digital banda base, las formas de onda de señal son

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

donde $g_T(t)$ es la respuesta impulsiva del filtro de transmisión cuya forma determina las características espectrales de la señal transmitida y A_m es la amplitud de señal que toma valores discretos dados por

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

Como mostrado en el Capítulo anterior el espectro de las señales banda base está contenido en la banda de frecuencias $|f| \leq W$, donde W es el ancho de banda de $|G_T(f)|^2$.

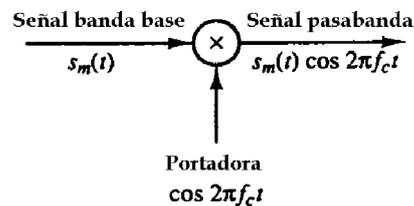


Figura 7.1: Modulación de amplitud de una portadora sinusoidal mediante la señal PAM digital banda base.

Para transmitir la señal digital a través de un canal pasabanda mediante modulación de amplitud las señales banda base $s_m(t)$, $m = 1, 2, \dots, M$ se multiplican por una portadora sinusoidal $\cos 2\pi f_c t$, como mostrado en la figura 7.1, donde f_c es la frecuencia de la portadora ($f_c < W$) y corresponde a la frecuencia central de la banda pasante del canal. De esta forma, las señales transmitidas pueden expresarse como

$$u_m(t) = A_m g_T(t) \cos 2\pi f_c t, \quad m = 1, 2, \dots, M \quad (7.1)$$

Como discutido previamente en el Capítulo 2, la modulación de amplitud de la portadora $\cos 2\pi f_c t$ mediante las señales banda base, $s_m(t) = A_m g_T(t)$, desplaza el espectro de la señal banda base por una cantidad igual a f_c y de esta forma, coloca a la señal en la banda pasante del canal. Recordando que la transformada de Fourier de la portadora es $[\delta(f - f_c) + \delta(f + f_c)]/2$, y como la multiplicación de dos señales en el dominio tiempo corresponde a una convolución de sus espectros en el dominio frecuencia, el espectro de la señal de amplitud modulada dado por (7.1) será

$$U_m(f) = \frac{A_m}{2} [G_T(f - f_c) + G_T(f + f_c)]$$

De esta forma el espectro de la señal banda base $s_m(t) = A_m g_T(t)$ se desplaza en frecuencia en la frecuencia de la portadora f_c . El resultado es una señal AM DSB-SC, como ilustrado en la figura 7.2.

La ilustración de las características espectrales de la señal modulada en amplitud anterior tiene en cuenta la densidad espectral de potencia de una secuencia de símbolos transmitidos. De esa forma, si

$$v(t) = \sum_{-\infty}^{\infty} a_n g_T(t - nT) \quad (7.2)$$

es la señal banda base PAM y $\{a_n\}$ es la secuencia de valores de amplitud, la densidad espectral de potencia de la señal de amplitud modulada

$$u(t) = v(t) \cos 2\pi f_c t \quad (7.3)$$

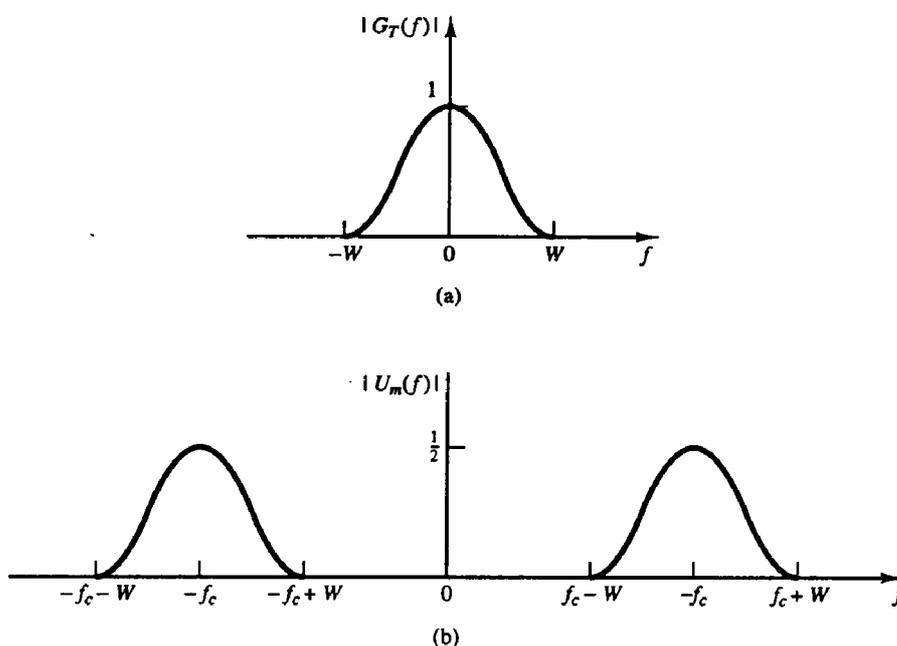


Figura 7.2: Espectro de (a) una señal banda base y (b) una señal modulada en amplitud.

será

$$S_U(f) = \frac{1}{4}[S_V(f - f_c) + S_V(f + f_c)]$$

donde $S_V(f)$ es la densidad espectral de potencia de la señal banda base previamente discutida en el Capítulo anterior, y $S_U(f)$ es la densidad espectral de potencia de la señal pasabanda $u(t)$.

La señal modulada $u(t)$ es una señal de amplitud modulada de doble banda lateral de portadora suprimida (AM-DBL). La banda lateral superior de $u(t)$ está formada por el contenido de frecuencias de $u(t)$ para $|f| > f_c$, o sea, para $f_c < |f| \leq f_c + W$. La banda lateral inferior de $u(t)$ está formada por el contenido de frecuencias para $|f| < f_c$, o sea, para $f_c - W \leq |f| < f_c$. En consecuencia, la señal de amplitud modulada DBL ocupa un ancho de banda de canal de $2W$, el cual es dos veces el ancho de banda requerido para transmitir la señal banda base.

La energía de las señales pasabanda $u_m(t)$, $m = 1, 2, \dots, M$, dadas por (7.1), está definida por

$$\begin{aligned} \mathcal{E}_m &= \int_{-\infty}^{\infty} u_m^2(t) dt = \int_{-\infty}^{\infty} A_m^2 g_T^2(t) \cos^2 2\pi f_c t dt \\ &= \frac{A_m^2}{2} \int_{-\infty}^{\infty} g_T^2(t) dt + \frac{A_m^2}{2} \int_{-\infty}^{\infty} g_T^2(t) \cos 4\pi f_c t dt \end{aligned}$$

Notar que cuando $f_c \gg W$, el término $\int_{-\infty}^{\infty} g_T^2(t) \cos 4\pi f_c t dt$ involucra la integración del producto de una función lentamente variable, $g_T^2(t)$, con un término sinusoidal rápidamente variante, $\cos 4\pi f_c t$, de forma que aquel término es prácticamente cero. En consecuencia

$$\mathcal{E}_m = \frac{A_m^2}{2} \int_{-\infty}^{\infty} g_T^2(t) dt = \frac{A_m^2}{2} \mathcal{E}_g \quad (7.4)$$

donde \mathcal{E}_g es la energía en el pulso de señal $g_T(t)$. De esta forma, se mostró que la energía en la señal pasabanda es la mitad de la energía de la señal banda base. El factor de escala $1/2$ se debe a la componente portadora $\cos 2\pi f_c t$, que tiene una energía promedio de $1/2$.

Cuando la forma del pulso transmitido $g_T(t)$ es rectangular, o sea

$$g_T(t) = \begin{cases} \sqrt{\frac{\mathcal{E}_g}{T}} & 0 \leq t \leq T \\ 0 & \forall f \end{cases}$$

la señal de amplitud modulada se denomina usualmente *llaveado de desplazamiento de amplitud* (Amplitud-shift keying, ASK).

Finalmente, notar que llevar las señales banda base $s_m(t)$ sobre la amplitud de la portadora $\cos 2\pi f_c t$ no cambia la representación geométrica básica de la señal PAM digital. Las señales $u_m(t)$ pueden expresarse como

$$u_m(t) = s_m \psi(t)$$

donde la señal básica está definida por

$$\psi(t) = \sqrt{\frac{2}{\mathcal{E}_g}} g_T(t) \cos 2\pi f_c t \quad (7.5)$$

y

$$s_m = \sqrt{\frac{\mathcal{E}_g}{2}} A_m, \quad m = 1, 2, \dots, M \quad (7.6)$$

Notar que el único cambio en la representación geométrica, comparando con las señales banda base, es el factor escalar $\sqrt{2}$, que aparece en (7.5) y (7.6)

7.1.1 Demodulación y detección de PAM

La demodulación de una señal PAM digital pasabanda puede lograrse de varias formas mediante correlación o filtrado acoplado. Con propósitos ilustrativos se considerará un demodulador de tipo correlador.

Es posible suponer que la señal transmitida es

$$u_m(t) = A_m g_T(t) \cos 2\pi f_c t, \quad 0 \leq t \leq T \quad (7.7)$$

La señal recibida puede expresarse como

$$r(t) = A_m g_T(t) \cos 2\pi f_c t + n(t), \quad 0 \leq t \leq T \quad (7.8)$$

donde $n(t)$ es el proceso de ruido en la banda pasante representado por

$$n(t) = n_c(t) \cos 2\pi f_c t - n_s(t) \sin 2\pi f_c t$$

Mediante la correlación cruzada entre la señal recibida $r(t)$ con la función base $\psi(t)$, como mostrado en la figura 7.3 se obtiene la salida

$$\int_{-\infty}^{\infty} r(t) \psi(t) dt = A_m \sqrt{\frac{2}{\mathcal{E}_g}} \int_{-\infty}^{\infty} g_T^2(t) \cos^2 2\pi f_c t + \int_{-\infty}^{\infty} n(t) \psi(t) dt \quad (7.9)$$

$$= A_m \sqrt{\frac{\mathcal{E}_g}{2}} + n \quad (7.10)$$

donde n representa la componente de ruido aditivo a la salida del correlador.

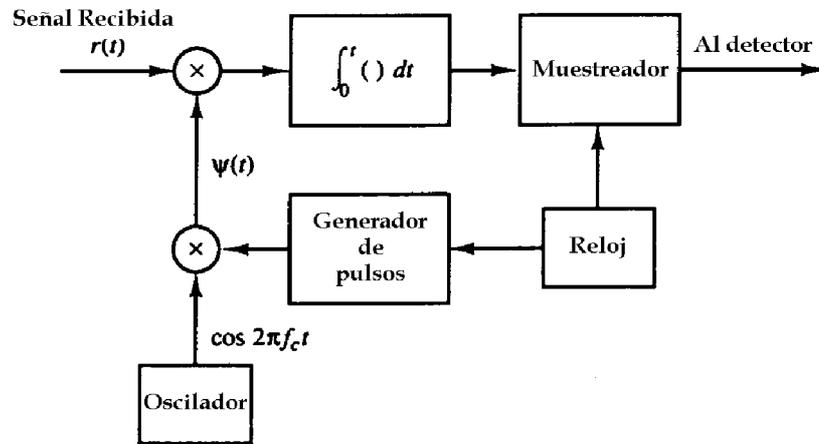


Figura 7.3: Demodulación de una señal PAM digital.

Recuperación de la fase de la portadora

En el desarrollo anterior se supuso que la función $\psi(t)$ está perfectamente sincronizada con la componente de señal en $r(t)$ tanto en tiempo como en fase de la portadora. En la práctica sin embargo, esas condiciones ideales no se cumplen debido el retardo de propagación encontrado al transmitir una señal a través del canal y debido a la deriva de fase y frecuencia que ocurren en cualquier oscilador práctico asociado a la portadora $\cos 2\pi f_c t$. En consecuencia, es necesario generar una portadora de fase coherente en el receptor para realizar la demodulación de la señal recibida.

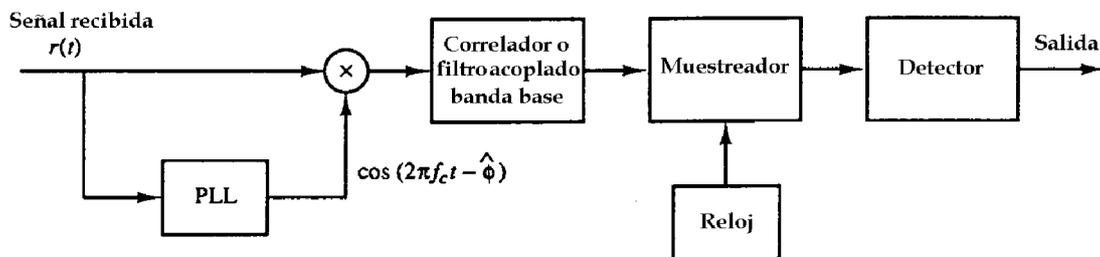


Figura 7.4: Demodulación de una señal de portadora modulada en amplitud.

Como la señal de mensaje dada por (7.2) es generalmente de media cero, la señal de amplitud modulada DBL $u(t)$ dada por (7.3) tiene potencia promedio cero alrededor de $f = f_c$. Por ello, no es posible estimar la fase de la portadora directamente de $r(t)$. Sin embargo, elevando al cuadrado a $r(t)$ es posible generar una componente de frecuencia en $f = 2f_c$ que no tenga potencia promedio nula. Esta componente puede obtenerse luego mediante un filtro pasabanda de banda estrecha sintonizado en $f = 2f_c$, cuya salida puede enviarse a un PLL. Un diagrama en bloques funcional del receptor que emplea un PLL para estimar la fase de la portadora se ilustra en la figura 7.4.

El *Lazo de Costas* sirve como un método alternativo para estimar la fase de la portadora a partir de la señal recibida $r(t)$. Las estimaciones con el Lazo de Costas y un PLL producen estimaciones comparables en calidad en presencia de ruido aditivo en el canal.

Demodulación y detección

Como alternativa a la demodulación banda base es posible realizar correlación cruzada o filtrado acoplado en la banda pasante o en alguna frecuencia intermedia conveniente. En particular, un correlador pasabanda puede utilizarse para multiplicar la señal recibida $r(t)$ por la portadora de amplitud modulada

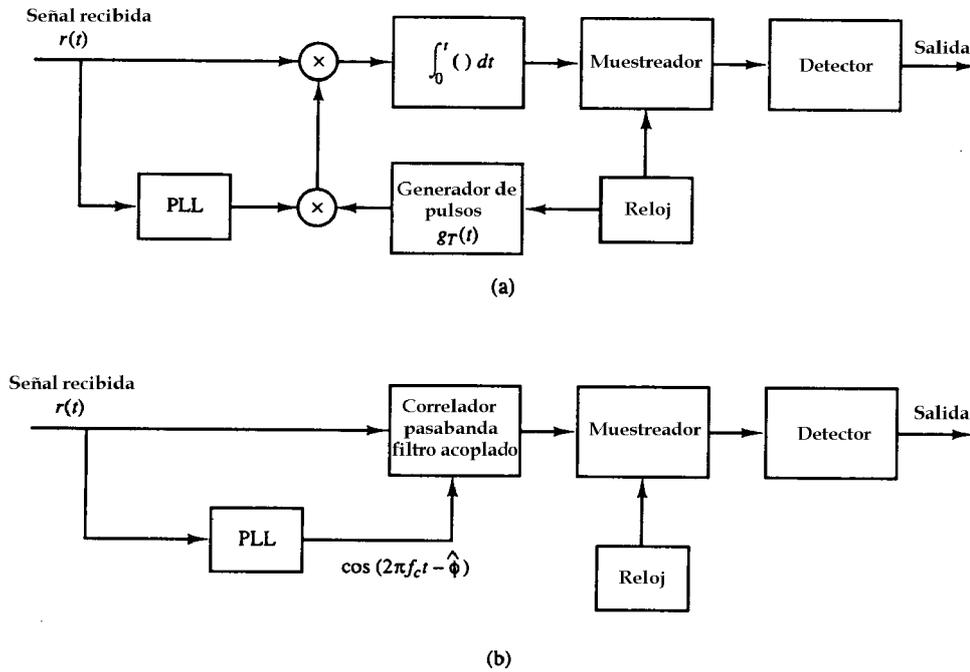


Figura 7.5: Demodulación pasabanda de una señal PAM digital mediante: (a) correlación pasabanda y (b) filtrado acoplado pasabanda.

$g_T(t) \cos(2\pi f_c t - \hat{\phi})$, donde $\cos(2\pi f_c t - \hat{\phi})$ es la salida del PLL. La señal producto se integra en un intervalo de señalización T , la salida del integrador se muestra en $t = T$ y la muestra se conduce a través del detector. Si se utiliza un filtro acoplado en lugar de un correlador, la respuesta impulsiva del filtro será $g_T(T-t) \cos[2\pi f_c(T-t) - \hat{\phi}]$. Un diagrama en bloques funcional para esos demoduladores se muestra en la figura 7.5.

Para describir la operación del detector óptimo para el caso de un canal AWGN, es posible suponer que la señal transmitida es $u_m(t) = s_m \psi(t)$, donde $\psi(t)$ y s_m están dadas por (7.5) y (7.6), respectivamente. Además, es posible suponer que la estimación de la fase de la portadora es $\hat{\phi}(t) = \phi(t)$, donde $\phi(t)$ es la fase de la portadora. Luego, la entrada al detector será

$$r = \int_{-\infty}^{\infty} r(t) \psi(t) dt = s_m \int_{-\infty}^{\infty} \psi^2(t) dt + \int_{-\infty}^{\infty} n(t) \psi^2(t) dt \quad (7.11)$$

$$= A_m \sqrt{\frac{\mathcal{E}_g}{2}} + n \quad (7.12)$$

donde $n(t)$ es una función muestra del proceso de ruido pasabanda.

Para mensajes equiprobables, el detector óptimo basa su decisión en las métricas de distancia dadas por

$$D(r, s_m) = (r - s_m)^2, \quad m = 1, 2, \dots, M$$

o equivalentemente, en las métricas de correlación dadas por

$$C(r, s_m) = 2rs_m - s_m^2$$

como discutido previamente en el Capítulo 5.

Recuperación del temporizado de símbolos

Varios métodos de recuperación del temporizado de símbolos se describieron en el Capítulo anterior. Esos métodos se aplican tanto a sistemas banda base como a sistemas pasabanda. Debido a que cualquier señal modulada puede convertirse en una señal banda base mediante una translación en frecuencia simple, la temporización de símbolos puede recuperarse de la señal recibida después de la conversión de frecuencia a banda base.

En muchos sistemas de comunicaciones modernos la señal recibida se procesa (demodula) digitalmente después que ha sido muestreada a la frecuencia de Nyquist o más rápido. En tal caso la recuperación de la fase de la portadora y el temporizado de símbolos se lleva a cabo mediante operaciones de procesamiento de señal sobre las muestras de señal. De esta forma se utiliza frecuentemente un PLL digital para el primer caso y un lazo de recuperación de reloj para el segundo, como descrito en el Capítulo anterior.

7.1.2 Probabilidad de error para PAM en un canal AWGN

El desempeño en términos de probabilidad de error de los demoduladores descritos anteriormente para un canal AWGN es exactamente el mismo que para transmisión PAM banda base a través de un canal AWGN, tanto con la demodulación mediante un correlador o filtro acoplado. Se discutirá específicamente este aspecto calculando la SNR a la salida del correlador mostrado en la figura 7.4. Para ello se asumirá que la entrada al detector está dada por (7.12). El pulso transmitido $g_T(t)$ se supone limitado en frecuencia a W Hz, o sea, $G(f) = 0$ para $|f| > W$.

Debido a que el proceso de ruido es Gaussiano de media cero se tiene que

$$E[n] = \int_{-\infty}^{\infty} E[n(t)]\psi(t)dt = 0$$

La varianza de una muestra de ruido será

$$\begin{aligned} \sigma_n^2 &= E \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n(t)n(\tau)\psi(t)\psi(\tau)dt d\tau \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E[n(t)n(\tau)]\psi(t)\psi(\tau)dt d\tau \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_n(t-\tau)\psi(t)\psi(\tau)dt d\tau = \int_{-\infty}^{\infty} \psi(t)dt \left[\int_{-\infty}^{\infty} \Psi(f)S_n(f)e^{j2\pi ft}df \right] \\ &= \int_{-\infty}^{\infty} |\Psi(f)|^2 S_n(f)df \end{aligned} \quad (7.13)$$

donde $\Psi(f)$ es la transformada de Fourier de $\psi(t)$ y $S_n(f)$ es la densidad espectral de potencia del ruido aditivo. De (7.5) se obtiene que

$$\Psi(f) = \frac{1}{\sqrt{2\mathcal{E}_g}} [G_T(f-f_c) + G_T(f+f_c)] \quad (7.14)$$

Además, la densidad espectral de potencia del proceso de ruido de ancho de banda limitado será

$$S_n(f) = \begin{cases} \frac{N_0}{2}, & |f-f_c| \leq W \\ 0, & \forall f \end{cases} \quad (7.15)$$

Sustituyendo (7.14) y (7.15) en (7.13) se obtiene

$$\begin{aligned} \sigma_n^2 &= \frac{N_0}{4\mathcal{E}_g} \left[\int_{-f_c-W}^{-f_c+W} |G_T(f-f_c)|^2 df + \int_{-f_c-W}^{-f_c+W} |G_T(f+f_c)|^2 df \right] \\ &= \frac{N_0}{2\mathcal{E}_g} \int_{-W}^W |G_T(f)|^2 df = \frac{N_0}{2} \end{aligned}$$

Seguendo el desarrollo del Capítulo 5, la probabilidad de un error de símbolo será

$$P_M = \frac{M-1}{M} P \left(|r - s_m| > \sqrt{\mathcal{E}_g/2} \right) = \frac{2(M-1)}{M} Q \left(\sqrt{\frac{\mathcal{E}_g}{2}} \right) \quad (7.16)$$

En base a (7.4) es posible relacionar la energía promedio transmitida con \mathcal{E}_g , tal que

$$\mathcal{E}_{av} = P_{av}T = \sum_{m=1}^M \mathcal{E}_m = \frac{\mathcal{E}_g}{2} \sum_{m=1}^M (2m-1-M)^2 = \frac{M^2-1}{6} \mathcal{E}_g$$

De donde

$$\mathcal{E}_g = \frac{6P_{av}T}{M^2-1} \quad (7.17)$$

Sustituyendo (7.17) en (7.16) permite obtener la forma deseada de la probabilidad de error para PAM M -ária, o sea

$$P_M = \frac{2(M-1)}{M} Q \left(\sqrt{\frac{6P_{av}T}{(M^2-1)N_0}} \right) \quad (7.18)$$

7.1.3 Demodulación con distorsión en el canal

En muchos sistemas de comunicaciones prácticos el canal distorsiona la señal transmitida. El tipo más común de distorsión de canal encontrada en la práctica es la distorsión lineal (distorsión de amplitud y fase), discutida en el Capítulo anterior. Este tipo de distorsión resulta en interferencia intersímbolos (ISI) que degrada el desempeño del sistema de comunicaciones.

La ISI puede reducirse mediante el diseño apropiado de los filtros de transmisión y recepción. La metodología descrita en el Capítulo anterior para el diseño de los filtros de transmisión y recepción para un canal con distorsión lineal puede aplicarse directamente convirtiendo el canal pasabanda en un canal banda base equivalente. Específicamente, es posible suponer que el canal pasabanda está caracterizado mediante la respuesta en frecuencia $C_{pb}(f)$ y sea f_c la frecuencia central de la banda pasante. Sea $C(f)$ la respuesta en frecuencia de un canal banda base equivalente (pasabajos), definida como

$$C(f - f_c) = \begin{cases} 2C_{pb}(f), & f > f_c \\ 0, & f < f_c \end{cases}$$

Como $C_{pb}^*(-f) = C_{pb}(f)$, se obtiene que

$$C^*(-f - f_c) = \begin{cases} 0, & f > -f_c \\ 2C_{pb}^*(-f), & f < -f_c \end{cases}$$

En consecuencia, la respuesta en frecuencia $C_{pb}(f)$ del canal pasabanda puede expresarse como

$$C_{pb}(f) = \frac{1}{2} [C(f - f_c) + C^*(-f - f_c)] \quad (7.19)$$

donde $C(f)$ es la respuesta en frecuencia del canal banda base equivalente. Esta relación se ilustra en la figura 7.6. La relación correspondiente entre las respuestas impulsivas de los canales pasabanda y banda base puede obtenerse de (7.19). Si $c_{pb}(t)$ es la respuesta impulsiva del canal pasabanda y $c(t)$ es la del canal banda base, entonces¹

¹En general, $c_{pb}(t)$ es real, mientras que la respuesta impulsiva del canal banda base $c(t)$ puede ser compleja. Si la magnitud $|C_{pb}(f)|$ es par alrededor de $f = f_c$ y la fase $\Theta(f) = \angle C_{pb}(f)$ es impar alrededor de $f = f_c$, entonces $C_{pb}^*(-f) = C_{pb}(f)$ y $c(t)$ resultará real.

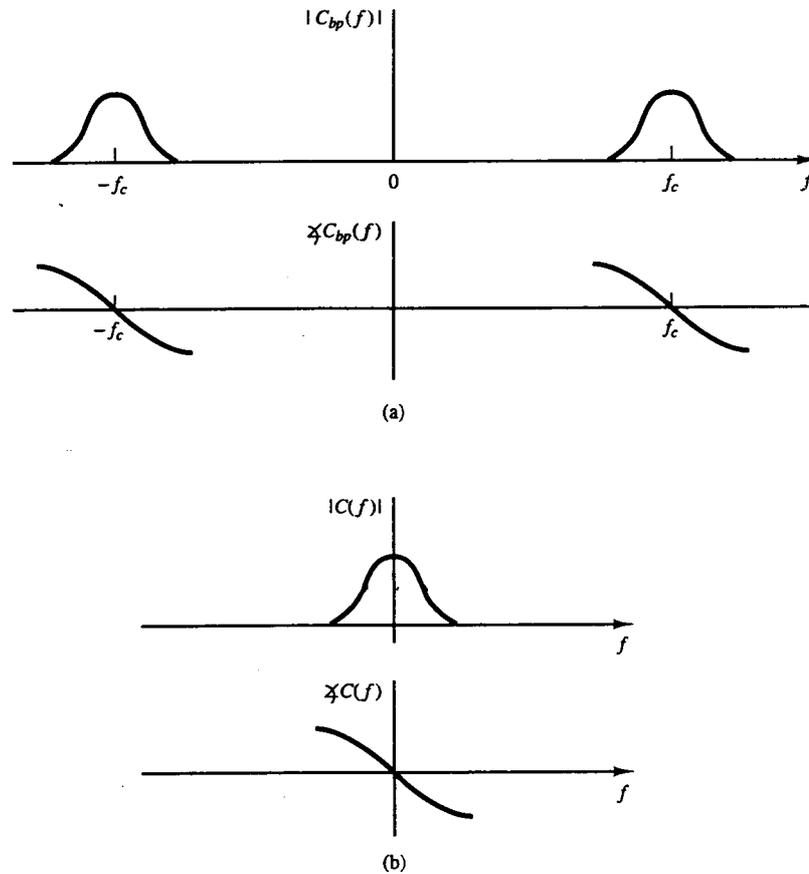


Figura 7.6: Respuesta de amplitud y fase de un canal pasabanda (a) y su canal banda base equivalente (b).

$$c_{pb}(t) = \text{Re}[c(t)e^{j2\pi f_c t}]$$

Una vez que se ha determinado la respuesta en frecuencia $C(f)$ del canal banda base equivalente como discutido anteriormente puede diseñarse los filtros óptimos de transmisión y recepción siguiendo un procedimiento similar al del Capítulo anterior, el cual se aplica, obviamente, a un canal banda base.

Cuando la característica de respuesta en frecuencia del canal $C_{pb}(f)$ es desconocida a priori como en el caso, por ejemplo, de canales telefónicos, la metodología descrita en el Capítulo anterior no es más aplicable. En tal caso será necesario recurrir a un ecualizador adaptivo para compensar la distorsión del canal. La aproximación básica en este caso es diseñar los filtros de transmisión y recepción $G_T(f)$ y $G_R(f)$ para ISI nula o para ISI controlada (señales de respuesta parcial), basándose en la suposición que el canal es ideal. Por ejemplo, es posible seleccionar $G_T(f)$ y $G_R(f)$ para satisfacer la condición

$$|G_T(f)||G_R(f)| = X_{rc}(f)$$

la cual resulta en ISI nula cuando el canal banda base equivalente es ideal. Luego, si se transmite un pulso aislado $g_T(t)$, la componente de señal a la salida del filtro de recepción tendrá la característica de respuesta en frecuencia $G_T(f)G_R(f)C(f)$, donde $C(f)$ representa la respuesta en frecuencia del canal banda base equivalente no ideal. La ISI resultante del canal no ideal puede compensarse utilizando un ecualizador lineal adaptivo, con respuesta en frecuencia $G_E(f)$. De esta forma, $G_E(f)$ sirve para compensar las características no ideales del canal $C(f)$. Esta es básicamente la solución descrita en el Capítulo anterior para el caso de ecualizar adaptivamente un canal banda base.

En un sistema que utiliza modulación de portadora el ecualizador adaptivo puede implementarse como pasabanda o banda base. La figura 7.7 ilustra las dos configuraciones. Las dos implementaciones son funcionalmente equivalentes.

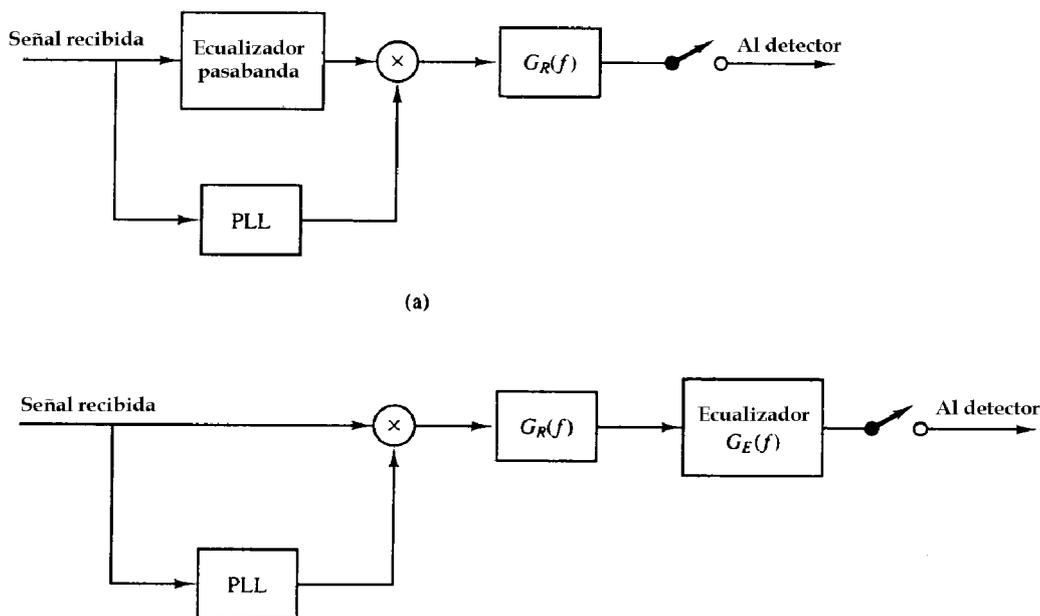


Figura 7.7: (a) Ecualización pasabanda, (b) ecualización banda base.

Finalmente, es posible agregar que podría utilizarse tanto un ecualizador con realimentación de decisión (DFE) o el detector de secuencias de ML (algoritmo de Viterbi) como alternativa a un ecualizador lineal para combatir ISI debido a la distorsión del canal. Estos dos ecualizadores no lineales producen un desempeño significativamente mejor comparado con el obtenido con el ecualizador lineal en un canal con distorsión severa.

7.2 Modulación de fase (PSK)

En la modulación de fase la información transmitida sobre un canal de comunicaciones está asociada a la fase de la portadora. Dado que el rango de la fase de la portadora es $0 \leq \theta < 2\pi$, el rango de fases de portadora utilizadas para transmitir información digital a través de modulación digital de fase es $\theta_k = 2\pi k/M$, $k = 0, 1, \dots, M-1$. De esta forma, para el caso de modulación de fase binaria ($M = 2$), las dos fases de la portadora serán $\theta_0 = 0$ y $\theta_1 = \pi$. Para modulación de fase M -ária, $M = 2^k$ donde k es el número de bits de información por símbolo transmitido.

La representación general de un conjunto de M señales de fase modulada es la siguiente

$$u_m(t) = g_T(t) \cos \left(2\pi f_c t + \frac{2\pi m}{M} \right), \quad m = 0, 1, \dots, M-1 \quad (7.20)$$

donde $g_T(t)$ es el formato de pulso asociado al filtro de transmisión, que determina las características espectrales de la señal transmitida. Esas señales tienen idéntica energía, o sea

$$\begin{aligned} \mathcal{E}_m &= \int_{-\infty}^{\infty} u_m^2(t) dt = \int_{-\infty}^{\infty} g_T^2(t) \cos^2 \left(2\pi f_c t + \frac{2\pi m}{M} \right) dt \\ &= \int_{-\infty}^{\infty} g_T^2(t) dt + \frac{1}{2} \int_{-\infty}^{\infty} g_T^2(t) \cos^2 \left(4\pi f_c t + \frac{4\pi m}{M} \right) dt \\ &= \frac{1}{2} \mathcal{E}_g \equiv \mathcal{E}_s \quad \forall m \end{aligned} \quad (7.21)$$

donde \mathcal{E}_g es la energía del pulso $g_T(t)$ y \mathcal{E}_s es la energía transmitida por símbolo. El término de (7.21) que involucra una componente de frecuencia doble es en promedio cero cuando $f_c \gg W$, donde W es el ancho

de banda de $g_T(t)$. En consecuencia, todas las señales poseen la misma energía. Notar que cuando $g_T(t)$ es un pulso rectangular, o sea está definido como

$$g_T(t) = \sqrt{\frac{2\mathcal{E}_s}{T}}$$

entonces las señales transmitidas serán

$$u_m(t) = \sqrt{\frac{2\mathcal{E}_s}{T}} \cos\left(2\pi f_c t + \frac{2\pi m}{M}\right), \quad m = 0, 1, \dots, M-1 \quad (7.22)$$

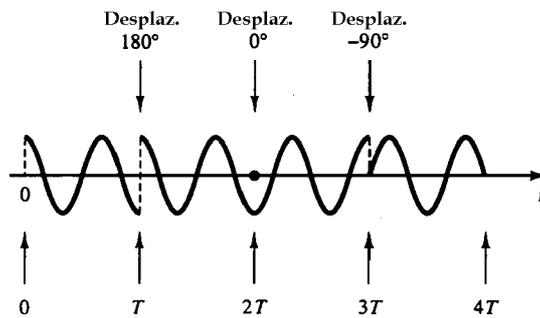


Figura 7.8: Ejemplo de una señal PSK de cuatro fases.

y tendrán una envolvente constante y cambios de fase abruptos al comienzo de cada intervalo de señal. Este tipo de modulación digital de fase se denomina *llaveado por desplazamiento de fase* (Phase-Shift Keying, PSK). La figura 7.8 ilustra una señal PSK de cuatro fases ($M = 4$).

Interpretando el argumento de la función coseno de (7.22) como la suma de dos ángulos, es posible reescribir (7.20) como

$$u_m(t) = g_T(t)A_{mc} \cos 2\pi f_c t - g_T(t)A_{ms} \sin 2\pi f_c t \quad (7.23)$$

donde

$$\begin{aligned} A_{mc} &= \cos 2\pi m/M, & m = 0, 1, \dots, M-1 \\ A_{ms} &= \sin 2\pi m/M, & m = 0, 1, \dots, M-1 \end{aligned} \quad (7.24)$$

De esta forma, una señal de fase modulada puede verse como dos portadoras en cuadratura con amplitudes $g_T(t)A_{mc}$ y $g_T(t)A_{ms}$ como ilustrado en la figura 7.9, las cuales dependerán de la fase transmitida en cada intervalo de señalización.

A partir de (7.23) las señales moduladas en fase pueden representarse geoméricamente como vectores bidimensionales con componentes $\sqrt{\mathcal{E}_s} \cos 2\pi m/M$ y $\sqrt{\mathcal{E}_s} \sin 2\pi m/M$, o sea

$$\mathbf{s}_m = \left(\sqrt{\mathcal{E}_s} \cos 2\pi m/M \quad \sqrt{\mathcal{E}_s} \sin 2\pi m/M \right)$$

Notar que las funciones base ortonormales son $\psi_1(t) = \sqrt{\frac{2}{\mathcal{E}_s}} g_T(t) \cos 2\pi f_c t$ y $\psi_2(t) = -\sqrt{\frac{2}{\mathcal{E}_s}} g_T(t) \sin 2\pi f_c t$. Las constelaciones de puntos de señal para $M = 2, 4, 8$ se ilustran en la figura 7.10. Es posible observar que la modulación de fase binaria es idéntica a PAM binaria.

El mapeo o asignación de k bits de información en $M = 2^k$ fases posibles puede realizarse de dos formas. La asignación preferida es utilizar el *código Gray*, en el cual las fases adyacentes difieren en un único bit

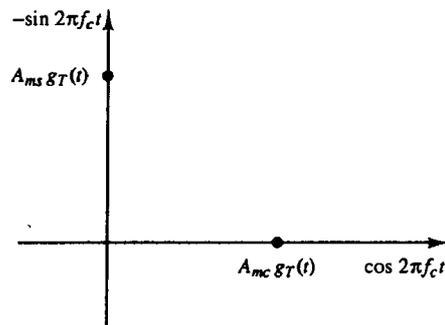


Figura 7.9: Modulación de fase digital vista como dos portadoras en cuadratura moduladas en amplitud.

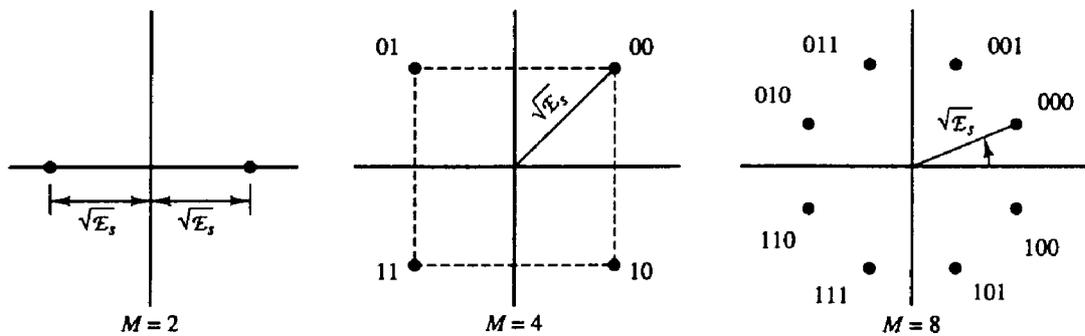


Figura 7.10: Constelaciones PSK.

como ilustrado en la figura 7.10. Debido a que los errores más probables originados por ruido involucran la selección errónea de una fase adyacente a la asociada a la fase transmitida, para el caso del código Gray solo se generará un único bit en error en la secuencia de k bits del símbolo.

Las características espectrales de una señal de fase modulada son similares a las características espectrales de una señal PAM digital. A partir de (7.20) es posible observar que las señales pasabanda $u_m(t)$ pueden expresarse en el intervalo de señalización $0 \leq t \leq T$ como

$$u_m(t) = \text{Re} \left[g_T(t) e^{j2\pi m/M} e^{j2\pi f_c t} \right], \quad m = 0, 1, \dots, M - 1$$

La señal pasabanda correspondiente para una secuencia de información de símbolos puede entonces expresarse como

$$u(t) = \text{Re} \left[v(t) e^{j2\pi f_c t} \right] \tag{7.25}$$

donde, por definición

$$v(t) = \sum_{n=-\infty}^{\infty} e^{j\theta_n} g_T(t - nT)$$

La secuencia transmitida de fases $\{\theta_n\}$ se obtiene mediante el mapeo de la secuencia de símbolos de k bits en las fases de portadora seleccionadas del conjunto de M fases dado por $\{2\pi m/M, m = 0, 1, \dots, M - 1\}$. Definiendo una amplitud compleja A_n como

$$A_n = e^{j\theta_n} = A_{nc} + jA_{ns}$$

es posible obtener una expresión para la señal pasabajos $v(t)$, la cual es idéntica a la expresión para una señal PAM digital (ver Capítulo anterior), excepto que ahora la secuencia $\{A_n\}$ es compleja. Esta diferencia requiere definir la función autocorrelación para la secuencia $\{A_n\}$ como

$$R_A(n) = E[A_k^* A_{n+k}] = E[e^{-j\theta_k} e^{j\theta_{n+k}}] = E[e^{j(\theta_{n+k} - \theta_k)}]$$

Más allá de esta diferencia menor, la derivación de la densidad espectral de potencia de la señal transmitida $u(t)$ dada por (7.25) sigue entonces el procedimiento del Capítulo anterior. De esta forma, se obtiene la función autocorrelación promedio de la señal transmitida como

$$\bar{R}_U(\tau) = \frac{1}{2} \sum_{-\infty}^{\infty} R_A(m) R_g(\tau - mT) \cos 2\pi f_c \tau$$

donde $R_g(\tau)$ es la función autocorrelación temporal del pulso $g_T(t)$, definida por (6.8). La transformada de Fourier de $\bar{R}_U(\tau)$ permite obtener la densidad espectral de potencia asociada dada por

$$S_U(f) = \frac{1}{4} [S_V(f - f_c) + S_V(-f - f_c)] \quad (7.26)$$

donde el espectro de potencia de la señal pasabajos $v(t)$ será

$$S_V(f) = \frac{1}{T} S_A(f) |G_T(f)|^2 \quad (7.27)$$

y

$$S_A(f) = \sum_{n=-\infty}^{\infty} R_A(n) e^{-j2\pi f n T} \quad (7.28)$$

Esto tiene la misma forma que el espectro de potencia promedio obtenido para una señal PAM. La única diferencia es que la función autocorrelación $R_A(n)$ en la señal de fase modulada es ahora de valor complejo.

Ejemplo. Determinar la densidad espectral de potencia de una señal PSK M -ária de envolvente constante cuando la secuencia de información es blanca.

La forma del pulso $g_T(t)$ para PSK es rectangular, en consecuencia

$$|G_T(f)|^2 = 2\mathcal{E}_s T \left[\frac{\sin \pi f T}{\pi f T} \right]^2 = 2\mathcal{E}_s T \text{sinc}^2(fT)$$

Las M posibles fases transmitidas son simétricas y supuestas equiprobables. Por ello,

$$R_A(m) = \begin{cases} \sigma_A^2, & m = 0 \\ 0, & m \neq 0 \end{cases}$$

donde $\sigma_A^2 = 1$. Luego, el espectro de potencia del proceso banda base $V(t)$ será

$$S_V(f) = 2\mathcal{E}_s \left[\frac{\sin \pi f T}{\pi f T} \right]^2 = 2\mathcal{E}_s \text{sinc}^2(fT)$$

y el espectro de potencia de la señal pasabanda será

$$S_U(f) = \frac{\mathcal{E}_s}{2} \left[\frac{\sin \pi(f - f_c)T}{\pi(f - f_c)T} \right]^2 + \frac{\mathcal{E}_s}{2} \left[\frac{\sin \pi(f + f_c)T}{\pi(f + f_c)T} \right]^2$$

Un esquema de $S_U(f)$ se muestra en la figura 7.11, graficada en función de la frecuencia normalizada fT .

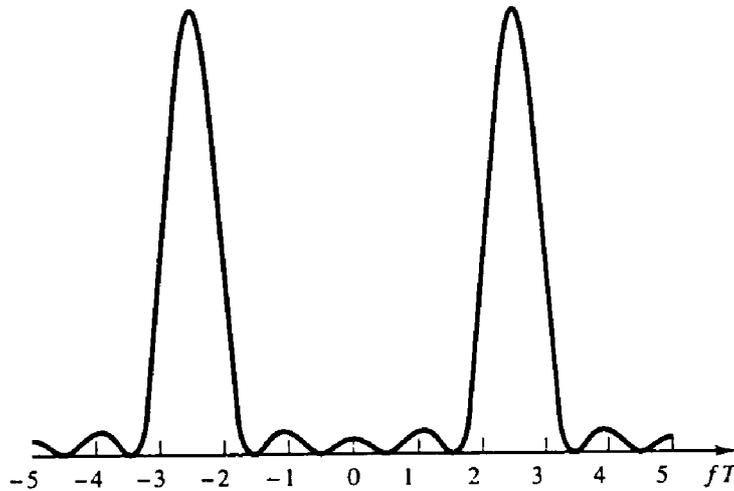


Figura 7.11: Densidad espectral de potencia para PSK M -ária con símbolos de información no correlacionados.

7.2.1 Demodulación y detección de PSK

La señal pasabanda recibida en un intervalo de señalización a partir de un canal AWGN puede expresarse como

$$\begin{aligned} r(t) &= u_m(t) + n(t) \\ &= g_T(t)A_{mc} \cos 2\pi f_c t - g_T(t)A_{ms} \sin 2\pi f_c t + n_c(t) \cos 2\pi f_c t - n_s(t) \sin 2\pi f_c t \end{aligned}$$

donde A_{mc} y A_{ms} están definidas por (7.24). La señal recibida puede correlacionarse con $\psi_1(t) = \sqrt{\frac{2}{\mathcal{E}_s}} g_T(t) \cos 2\pi f_c t$ y $\psi_2(t) = -\sqrt{\frac{2}{\mathcal{E}_s}} g_T(t) \sin 2\pi f_c t$. Las salidas de los dos correladores producen las dos componentes de señal y ruido, las que pueden expresarse como

$$\begin{aligned} \mathbf{r} &= \mathbf{s}_m + \mathbf{n} \\ &= \left(\sqrt{\mathcal{E}_s} \cos 2\pi m/M + n_c \quad \sqrt{\mathcal{E}_s} \sin 2\pi m/M + n_s \right) \end{aligned}$$

donde, por definición

$$\begin{aligned} n_c &= \frac{1}{\sqrt{2\mathcal{E}_g}} \int_{-\infty}^{\infty} g_T(t) n_c(t) dt \\ n_s &= \frac{1}{\sqrt{2\mathcal{E}_g}} \int_{-\infty}^{\infty} g_T(t) n_s(t) dt \end{aligned}$$

Como las componentes de ruido $n_c(t)$ y $n_s(t)$ tienen media cero y no están correlacionadas, entonces $E[n_c] = E[n_s] = 0$ y $E[n_c n_s] = 0$. La varianza de las componentes de ruido será

$$\begin{aligned} E[n_c^2] &= E[n_s^2] = \frac{1}{2\mathcal{E}_g} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_T(t) g_T(\tau) E[n_c(t) n_s(\tau)] dt d\tau \\ &= \frac{1}{2\mathcal{E}_g} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_T(t) g_T(\tau) R_{n_c}(t - \tau) dt d\tau \\ &= \frac{N_0}{2\mathcal{E}_g} \int_{-W}^W |G_T(f)|^2 df = N_0/2 \end{aligned}$$

El detector óptimo proyecta el vector de señal recibido sobre uno de los M posibles vectores de señal transmitidos $\{\mathbf{s}_m\}$ y selecciona aquel correspondiente al valor de mayor proyección. De esta forma se obtienen las métricas de correlación

$$C(\mathbf{r}, \mathbf{s}_m) = \mathbf{r} \cdot \mathbf{s}_m, \quad m = 0, 1, \dots, M - 1$$

Como todas las señales tienen la misma energía, una métrica equivalente para la detección de modulación de fase digital es calcular la fase del vector de señal recibido $\mathbf{r} = (r_1, r_2)$, dada por

$$\Theta_r = \tan^{-1} \frac{r_2}{r_1} \tag{7.29}$$

y seleccionar la señal del conjunto $\{\mathbf{s}_m\}$ cuya fase esté más próxima a Θ_r . En la sección siguiente se evaluará la probabilidad de error basandose en esta métrica.

Es necesario mencionar el problema asociado con el desplazamiento de la fase de la portadora de la señal recibida. En general los osciladores empleados en el transmisor y el receptor no están en sincronismo de fase. En consecuencia, la señal recibida será de la forma

$$r(t) = A_{mc} g_T(t) \cos(2\pi f_c t + \phi) - A_{ms} g_T(t) \sin(2\pi f_c t + \phi) + n(t)$$

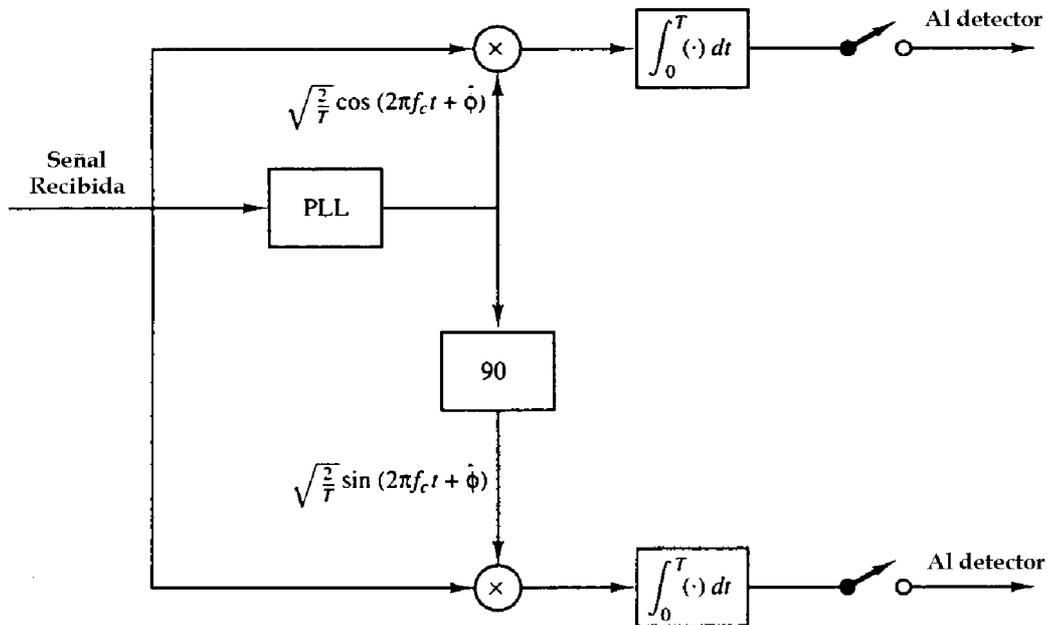


Figura 7.12: Demodulador para señales PSK.

donde ϕ es el desplazamiento de fase de la portadora. Este desplazamiento de fase deberá estimarse en el demodulador, el procedimiento se describirá en secciones siguientes. La estimación de fase obtenida deberá utilizarse en la operación de correlación. Por ello, la señal recibida debe estar correlacionada con las funciones ortogonales $\psi_1(t) = \sqrt{\frac{2}{E_s}} g_T(t) \cos(2\pi f_c t + \hat{\phi})$ y $\psi_2(t) = -\sqrt{\frac{2}{E_s}} g_T(t) \sin(2\pi f_c t + \hat{\phi})$, donde $\hat{\phi}$ es la estimación de la fase de la portadora, como mostrado en la figura 7.12, para el caso en que $g_T(t)$ es un pulso rectangular.

El temporizado de símbolos para una señal modulada en fase digital puede recuperarse de la señal recibida mediante cualquiera de los métodos descritos en el Capítulo anterior, los cuales son apropiados para sincronización de símbolos de la señal analógica recibida. También, como indicado para el caso de PAM, el temporizado de símbolos puede obtenerse procesando una versión muestreada de la señal recibida. En tal caso, el lazo de recuperación del temporizado se implementa en forma digital.

La distorsión de canal es otro problema muy probablemente encontrado en un sistema de comunicaciones práctico al utilizar modulación de fase. En ese caso es posible diseñar filtros de transmisión y recepción óptimos cuando la característica de respuesta en frecuencia del canal es conocida, utilizando la metodología de diseño de PAM pasabanda. Por otro lado, si la respuesta en frecuencia del canal no es conocida a priori puede utilizarse un ecualizador adaptivo para compensar la distorsión del canal.

7.2.2 Probabilidad de error para PSK en un canal AWGN

En esta sección se evaluará la probabilidad de error para modulación de fase M -ária en AWGN con el demodulador y detector óptimos. Se utilizará el detector óptimo basado en la métrica dada por (7.29). Se asumirá que está disponible una estimación perfecta de la fase de la portadora. En consecuencia, el desempeño que se concluirá es para el caso de modulación de fase ideal.

Es posible considerar el caso en el cual la fase de la señal transmitida es $\theta = 0$, correspondiente a la señal $u_0(t)$. De esta forma, el vector de señal transmitida será

$$\mathbf{s}_0 = \begin{pmatrix} \sqrt{\mathcal{E}_s} \\ 0 \end{pmatrix}$$

y las componentes del vector de señal recibida serán

$$\begin{aligned} r_1 &= \sqrt{\mathcal{E}_s} + n_s \\ r_2 &= n_s \end{aligned}$$

Como n_c y n_s son variables aleatorias conjuntamente Gaussianas, se concluye que r_1 y r_2 son variables aleatorias conjuntamente Gaussianas con $E[r_1] = \sqrt{\mathcal{E}_s}$, $E[r_2] = 0$ y $\sigma_{r_1}^2 = \sigma_{r_2}^2 = N_0/2 = \sigma_r^2$. En consecuencia,

$$f_r(r_1, r_2) = \frac{1}{2\pi\sigma_r^2} e^{-[(r_1 - \sqrt{\mathcal{E}_s})^2 + r_2^2]/2\sigma_r^2}$$

La métrica del detector es la fase $\Theta_r = \tan^{-1} r_2/r_1$. La función densidad de probabilidad de Θ_r se obtiene mediante un cambio de variables de (r_1, r_2) a $V = \sqrt{r_1^2 + r_2^2}$ y $\Theta_r = \tan^{-1} r_2/r_1$. Este cambio de variables produce la densidad de probabilidad conjunta

$$f_{V, \Theta_r}(v, \theta_r) = \frac{v}{2\pi\sigma_r^2} e^{-(v^2 + \mathcal{E}_s - 2\sqrt{\mathcal{E}_s}v \cos \theta_r)/2\sigma_r^2}$$

La integración de $f_{V, \Theta_r}(v, \theta_r)$ sobre el rango de v produce $f_{\Theta_r}(\theta_r)$, o sea

$$\begin{aligned} f_{\Theta_r}(\theta_r) &= \int_0^\infty f_{V, \Theta_r}(v, \theta_r) dv \\ &= \frac{1}{2\pi} e^{-2\rho_s \sin^2 \theta_r} \int_0^\infty v e^{-(v - \sqrt{4\rho_s} \cos \theta_r)^2/2} dv \end{aligned}$$

donde, por conveniencia, se define la SNR de símbolo como $\rho_s = \mathcal{E}_s/N_0$. La figura 7.13 ilustra $f_{\Theta_r}(\theta_r)$ para diferentes valores del parámetro de SNR ρ_s , cuando la fase transmitida es cero. Notar que $f_{\Theta_r}(\theta_r)$ se vuelve más pronunciada y estrecha alrededor de $\theta_r = 0$ cuando la SNR ρ_s aumenta.

Cuando se transmite $u_0(t)$ se comete un error de decisión si el ruido hace que la fase caiga fuera del rango $-\pi/M \leq \Theta_r \leq \pi/M$. Por ello, la probabilidad de un error de símbolo será

$$P_M = 1 - \int_{-\pi/M}^{\pi/M} f_{\Theta_r}(\theta) d\theta \quad (7.30)$$

En general, la integral de $f_{\Theta_r}(\theta)$ no tiene una forma cerrada y debe evaluarse numericamente, excepto para $M = 2$ y $M = 4$.

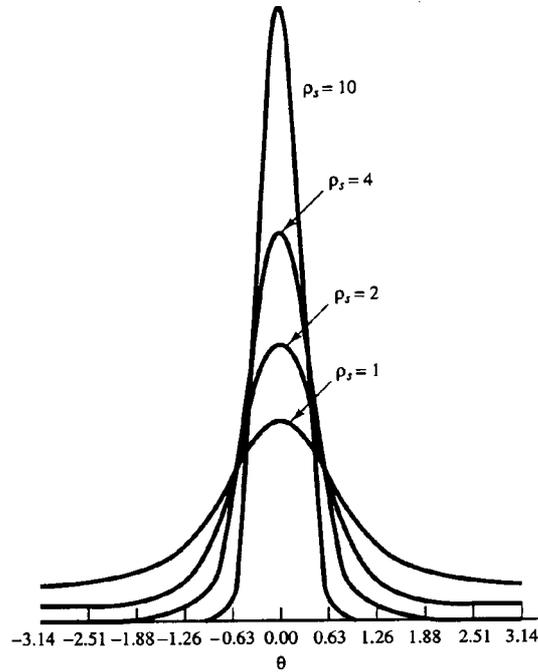


Figura 7.13: Función densidad de probabilidad $p(\theta)$ para $\rho_s = 1, 2, 4, 10$.

Para modulación de fase binaria, las señales $u_0(t)$ y $u_1(t)$ son antipodales y por ello la probabilidad de error será

$$P_2 = Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right) \quad (7.31)$$

Cuando $M = 4$, se tiene el efecto de dos señales de modulación de fase binaria en fase y cuadratura. Con una estimación perfecta de la fase de la portadora no existirá interferencia entre las señales sobre las dos portadoras y en consecuencia la probabilidad de error de bit es idéntica a la de (7.31). Por otro lado, la probabilidad de error de símbolo para $M = 4$ puede determinarse notando que

$$P_c = (1 - P_2)^2 = \left[1 - Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right)\right]^2$$

donde P_c es la probabilidad de decisión correcta para el símbolo de 2 bits. Este resultado se obtiene de la independencia estadística del ruido de las portadoras en cuadratura. En consecuencia, la probabilidad de error de símbolo para $M = 4$ será

$$P_4 = (1 - P_c) = 2Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right) \left[1 - \frac{1}{2}Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right)\right]$$

Para $M > 4$ la probabilidad de error de símbolo P_M se obtiene integrando numericamente (7.30). La figura 7.14 ilustra la probabilidad de error en función de la SNR por bit para $M = 2, 4, 8, 16$ y 32 . Estos gráficos ilustran claramente la penalización en SNR por bit cuando M aumenta más allá de $M = 4$. Por ejemplo, para $P_M = 10^{-5}$, la diferencia entre $M = 4$ y $M = 8$ es aproximadamente 4 dB y la diferencia entre $M = 8$ y $M = 16$ es aproximadamente 5 dB. Para grandes valores de M , duplicar el número de fases requiere un adicional de 6 dB por bit para lograr el mismo desempeño.

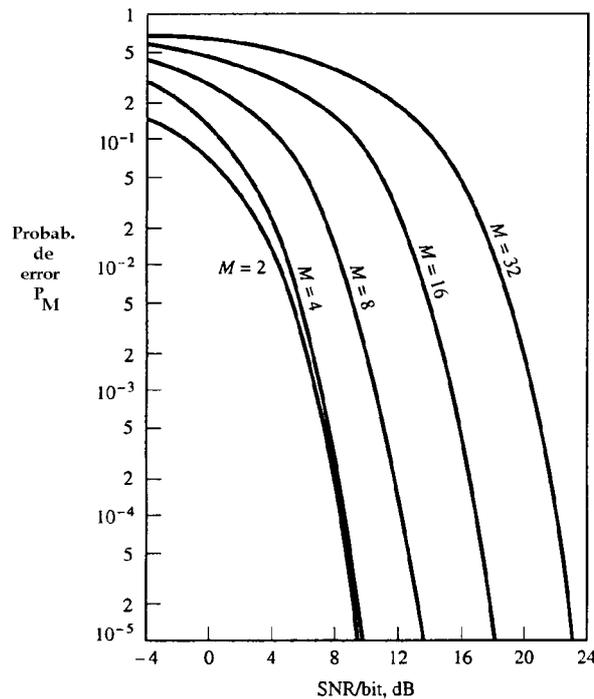


Figura 7.14: Probabilidad de error de símbolo para señales PSK.

Una aproximación a la probabilidad de error para valores grandes de M y para SNR grandes puede obtenerse aproximando primero $f_{\Theta_r}(\theta)$. Para $\mathcal{E}_s/N_0 \gg 1$ y $|\Theta_r| \leq \pi/2$, $f_{\Theta_r}(\theta)$ puede aproximarse como

$$f_{\Theta_r}(\theta_r) \cong \sqrt{\frac{2\rho_s}{\pi}} e^{-2\rho_s \sin^2 \theta_r}$$

Sustituyendo en (7.30) y realizando el cambio de variables de θ_r para $u = 2\sqrt{\rho_s} \sin \theta_r$, se obtiene

$$\begin{aligned} P_M &\cong 1 - \int_{-\pi/M}^{\pi/M} \sqrt{\frac{2\rho_s}{\pi}} \cos \theta_r e^{-2\rho_s \sin^2 \theta_r} d\theta_r \\ &\cong \frac{2}{\sqrt{2\pi}} \int_{\sqrt{2\rho_s} \sin \pi/M}^{\infty} e^{-u^2/2} du \\ &= 2Q\left(\sqrt{2\rho_s} \sin \frac{\pi}{M}\right) = 2Q\left(\sqrt{2k\rho_b} \sin \frac{\pi}{M}\right) \end{aligned}$$

donde $k = \log_2 M$ y $\rho_s = k\rho_b$. Notar que esta aproximación de la probabilidad de error es adecuada para cualquier valor de M . Por ejemplo, cuando $M = 2$ o $M = 4$ se tiene que $P_2 = P_4 = 2Q(\sqrt{2\rho_s})$, lo cual se compara favorablemente (excepto por un factor de 2) con la probabilidad de error exacta dada por (7.31).

La probabilidad de error de bit equivalente para modulación de fase M -ária es bastante tediosa de obtener debido a la dependencia del mapeo de los símbolos de k bits en relación a las fases correspondientes. Cuando se utiliza el código Gray en este mapeo, dos símbolos de k bits correspondientes a fases adyacentes difieren solamente en un bit. Como los errores más probables debido al ruido resultarán en la selección errónea de una fase adyacente en lugar de la fase verdadera, los errores de símbolo de k bits contendrán solo un único bit en error. En consecuencia, la probabilidad de error de bit equivalente para modulación de fase M -ária puede aproximarse adecuadamente por

$$P_b \cong \frac{1}{k} P_M$$

7.2.3 Estimación de la fase de la portadora

Cuando se transmite información digital mediante la modulación de fase digital M -ária de una portadora es posible utilizar un PLL para estimar el desplazamiento de fase de la portadora. Para $M = 2$, el PLL con un elemento no lineal cuadrático o el Lazo de Costas son directamente aplicables.

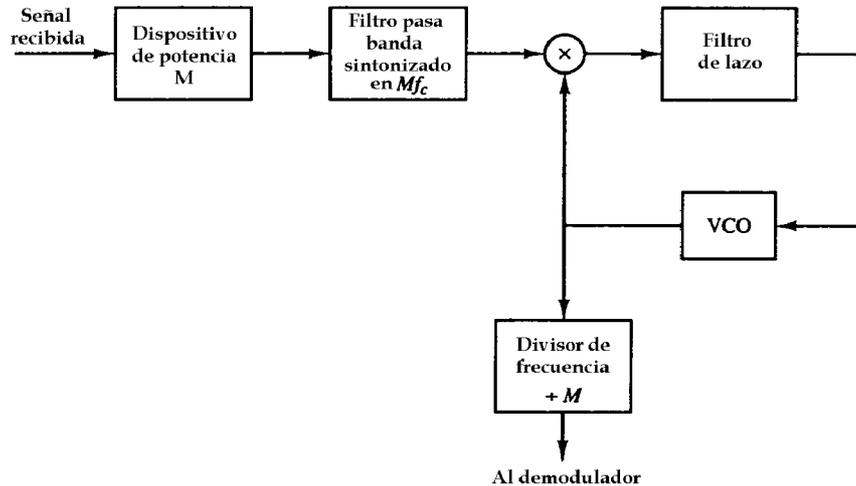


Figura 7.15: Estimación de la fase de la portadora para señales PSK M -árias.

Para $M > 2$, la señal recibida debe elevarse primero a la M -ésima potencia como mostrado en la figura 7.15. De esta forma, si la señal recibida $r(t)$ tiene la forma

$$r(t) = s_m(t) + n(t) = g_T(t) \cos\left(2\pi f_c t + \phi + \frac{2\pi m}{M}\right) + n(t)$$

y se pasa a través de un dispositivo de M -ésima potencia, la señal de salida contendrá armónicas de la portadora f_c . La armónica que se desea seleccionar para llevar al PLL es $\cos(2\pi M f_c t + M\theta)$. Notando que

$$M\theta = M \left(\frac{2\pi m}{M}\right) = 2\pi m = 0 \pmod{2\pi}, \quad m = 1, 2, \dots, M$$

De esta forma, se remueve la información asociada a la M -ésima armónica. El filtro pasabanda sintonizado a la frecuencia $M f_c$ produce la componente de frecuencia deseada $\cos(2\pi M f_c t + M\phi)$. La salida del VCO es $\sin(2\pi M f_c t + M\hat{\phi})$, de forma que su salida se divide en frecuencia por M para producir $\sin(2\pi f_c t + \hat{\phi})$ y se desplaza en fase en $\pi/2$ para producir $\cos(2\pi f_c t + \hat{\phi})$. Estas componentes se introducen luego al demodulador.

Debe notarse que las componentes de portadora así generadas contienen ambigüedad de fase de múltiplos de $2\pi/M$ que resultan de multiplicar la fase de la portadora ϕ por M . Como $M\phi \pmod{2\pi}$ es menor que 2π , dividir el ángulo resultante por M produce una estimación de fase de $|\hat{\phi}| < 2\pi/M$, cuando en realidad el verdadero valor de la fase de la portadora puede exceder esta estimación en múltiplos de $2\pi/M$, o sea, en $2\pi k/M$, para $k = 1, 2, \dots, M - 1$. Estas ambigüedades de fase pueden resolverse codificando en forma diferencial los datos en el transmisor y decodificando diferencialmente en el detector, como se discutirá más adelante.

De la misma forma que en el caso del PLL de potencia dos, el PLL de potencia M opera en presencia de ruido que ha sido amplificado por el dispositivo de M -ésima potencia. La varianza del error de fase resultante en el PLL debido al ruido aditivo puede expresarse como

$$\sigma_{\hat{\phi}}^2 = \frac{1}{\rho_L S_{ML}}$$

donde ρ_L es la SNR del lazo y S_{ML} es la pérdida de potencia de fase M . S_{ML} ha sido evaluada experimentalmente para $M = 4$ y $M = 8$.

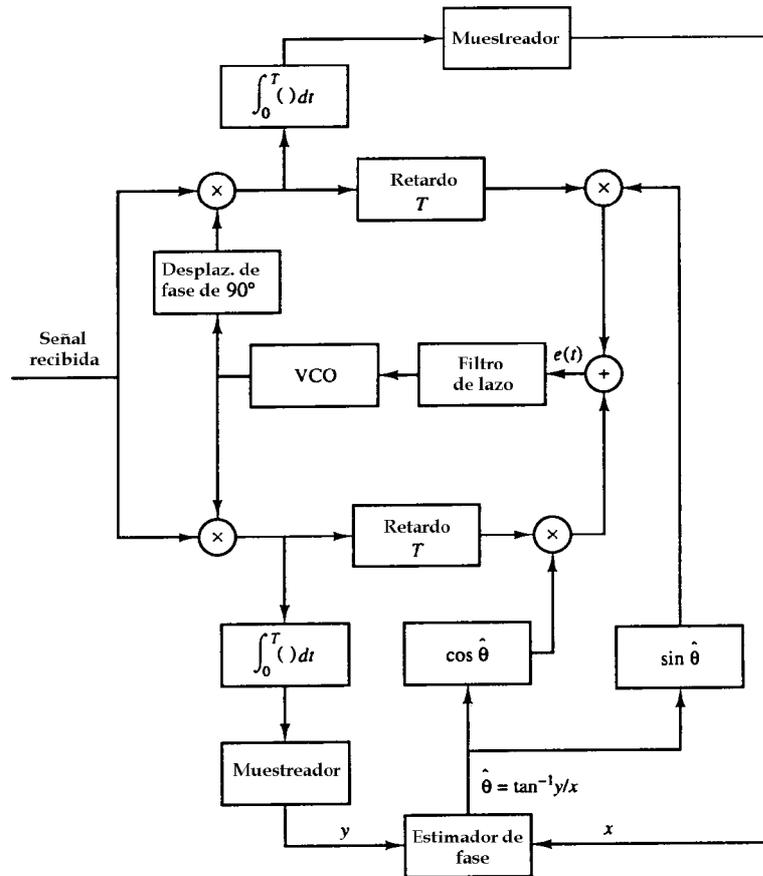


Figura 7.16: Recuperación de la portadora para PSK M -ária usando DFPLL.

Otro método para extraer una estimación de la fase de la portadora $\hat{\phi}$ de la señal recibida para modulación digital de fase M -ária es el PLL con realimentación de decisión (DFPLL), que se ilustra en la figura 7.16. La señal recibida se demodula utilizando dos portadoras en cuadratura para producir $\mathbf{r} = (r_1, r_2)$ en los instantes de muestreo. Se calcula la estimación de fase $\hat{\Theta}_r = \tan^{-1} r_2/r_1$ en el detector y se cuantiza a la más próxima de las M fases transmitidas posibles $\hat{\theta}_m$. Las dos salidas de los multiplicadores en cuadratura se retardan en un intervalo de símbolo T y se multiplican por $\cos \theta_m$ y $-\sin \theta_m$. De esta forma se obtiene

$$\begin{aligned}
 -r(t) \cos(2\pi f_c t + \hat{\phi} \sin \theta_m) &= -\frac{1}{2}[g_T(t) \cos \theta_m + n_c(t)] \sin \theta_m \cos(\phi - \hat{\phi}) \\
 &\quad + \frac{1}{2}[g_T(t) \sin \theta_m + n_s(t)] \sin \theta_m \sin(\phi - \hat{\phi}) \\
 &\quad + \text{términos de frecuencia doble} \\
 -r(t) \sin(2\pi f_c t + \hat{\phi} \cos \theta_m) &= -\frac{1}{2}[g_T(t) \cos \theta_m + n_c(t)] \cos \theta_m \sin(\phi - \hat{\phi}) \\
 &\quad + \frac{1}{2}[g_T(t) \sin \theta_m + n_s(t)] \cos \theta_m \cos(\phi - \hat{\phi}) \\
 &\quad + \text{términos de frecuencia doble}
 \end{aligned}$$

Esas dos señales se suman para generar la señal de error dada por

$$\begin{aligned}
 e(t) &= g_T(t) \sin(\phi - \hat{\phi}) + \frac{1}{2}n_c(t) \sin(\phi - \hat{\phi} - \theta_m) + \frac{1}{2}n_s(t) \cos(\phi - \hat{\phi} - \theta_m) \\
 &\quad + \text{términos de frecuencia doble}
 \end{aligned} \tag{7.32}$$

La señal error es la entrada al filtro de lazo que provee la señal de control para el VCO. Es posible observar que las dos componentes de ruido en cuadratura aparecen en (7.32) como términos aditivos y que además

no existen términos que involucren producto de dos componentes de ruido, de la misma forma que a la salida de un dispositivo de potencia M -ésima. En consecuencia, no existirá pérdida de potencia resultante de operaciones no lineales sobre la señal recibida al DFPLL.

El lazo de seguimiento de realimentación de decisión de fase M -ésima también tiene una ambigüedad de fase de $2\pi/M$, por lo que requerirá la codificación diferencial de la secuencia de información previa a la transmisión y la decodificación diferencial de la secuencia recibida en el detector para recuperar la información.

7.2.4 Modulación de fase diferencial (DPSK)

El desempeño ideal de modulación/demodulación de fase coherente puede alcanzarse muy aproximadamente en sistemas de comunicaciones que transmiten una señal portadora. La componente de señal portadora puede filtrarse a partir de la señal recibida y utilizarse para realizar la demodulación de fase coherente. Sin embargo, cuando no se transmite una portadora separada, el receptor debe estimar la fase de la portadora de la señal recibida. Como discutido en la sección previa, la fase a la salida de un PLL tiene ambigüedades de múltiplos de $2\pi/M$, lo que requiere la codificación diferencial de los datos previa a la modulación. Esta codificación diferencial permite decodificar los datos recibidos en el detector en presencia de ambigüedades de fase.

En codificación diferencial la información está asociada a los desplazamientos de fase relativos al intervalo de señalización previo. Por ejemplo, en modulación de fase binaria el bit de información 1 puede transmitirse desplazando la fase de la portadora en 180° con respecto a la fase de portadora previa, mientras que el bit de información 0 se transmite sin desplazamiento de fase relativo a la fase de la portadora en el intervalo previo. En la modulación de 4 fases, los desplazamientos de fase relativos entre intervalos sucesivos son 0° , 90° , 180° y 270° , correspondientes a los bits de información 00, 01, 11 y 10, respectivamente. La generalización de la codificación diferencial para $M > 4$ es directa. Las señales de fase modulada resultantes de este proceso de codificación se denominan *codificadas diferencialmente*. La codificación se realiza mediante un circuito lógico simple que precede al modulador.

La demodulación y detección de la señal de fase modulada codificada diferencialmente puede realizarse como discutido en la sección precedente utilizando la salida de un PLL para realizar la demodulación. La fase de la señal recibida en el detector $\Theta_r = \tan^{-1} r_2/r_1$ se mapea en la más próxima de las M posibles fases transmitidas $\{\theta_m\}$. Al detector le sigue un comparador de fase que detecta la señal sobre dos intervalos consecutivos para extraer la secuencia de información. De esta forma, las ambigüedades de fase de $2\pi/M$ se vuelven irrelevantes.

La demodulación coherente de señales de fase modulada codificadas diferencialmente resulta en una probabilidad de error mayor que la asociada a la codificación de fase absoluta. Con señales codificadas diferencialmente un error en la fase detectada debido al ruido resultará frecuentemente en errores de decodificación sobre dos intervalos de señalización consecutivos. Ese es particularmente el caso para probabilidades de error por debajo de 10^{-1} . En consecuencia, la probabilidad de error para señales M -árias de fase modulada codificadas diferencialmente es aproximadamente dos veces la probabilidad de error para el caso de codificación absoluta. Sin embargo, un factor de 2 de incremento en la probabilidad de error se traslada en un incremento pequeño de la pérdida de SNR, como puede verificarse en la figura 7.14.

Una señal de fase modulada codificada diferencialmente permite también otro tipo de demodulación que no requiere la estimación de la fase de la portadora. En su lugar, la señal recibida en cualquier intervalo de señalización dado se compara con la fase de la señal recibida del intervalo de señalización precedente. Para llevar a cabo la discusión de este tipo de modulación, es posible suponer que se demodula la señal codificada diferencialmente multiplicando $r(t)$ por $\cos 2\pi f_c t$ y $\sin 2\pi f_c t$ y se integra los dos productos sobre el intervalo T . En el intervalo de señalización k , la salida del demodulador será

$$r_k = \sqrt{\mathcal{E}_s} e^{j(\theta_k - \phi)} + n_k$$

donde θ_k es la fase de la señal transmitida en el intervalo de señalización k , ϕ es la fase de la portadora y $n_k = n_{ck} + jn_{sk}$ es el vector de ruido. En forma similar, el vector de señal recibida a la salida del demodulador en el intervalo de señalización precedente será

$$r_{k-1} = \sqrt{\mathcal{E}_s} e^{j(\theta_{k-1} - \phi)} + n_{k-1}$$

La variable de decisión para el detector de fase es la diferencia de fase entre esos dos números complejos. En forma equivalente, es posible proyectar r_k sobre r_{k-1} y utilizar la fase del número complejo resultante, o sea

$$r_k r_{k-1}^* = \sqrt{\mathcal{E}_s} e^{j(\theta_{k-1} - \theta_k)} + \sqrt{\mathcal{E}_s} e^{j(\theta_k - \phi)} n_{k-1} + \sqrt{\mathcal{E}_s} e^{j(\theta_{k-1} - \phi)} n_k + n_k n_{k-1}^* \quad (7.33)$$

la cual, en ausencia de ruido, produce la diferencia de fase $(\theta_k - \theta_{k-1})$. De esta forma, el valor medio de $r_k r_{k-1}^*$ es independiente de la fase de la portadora. La señalización PSK codificada diferencialmente que se demodula y detecta como discutido anteriormente se denomina *PSK diferencial* (DPSK).

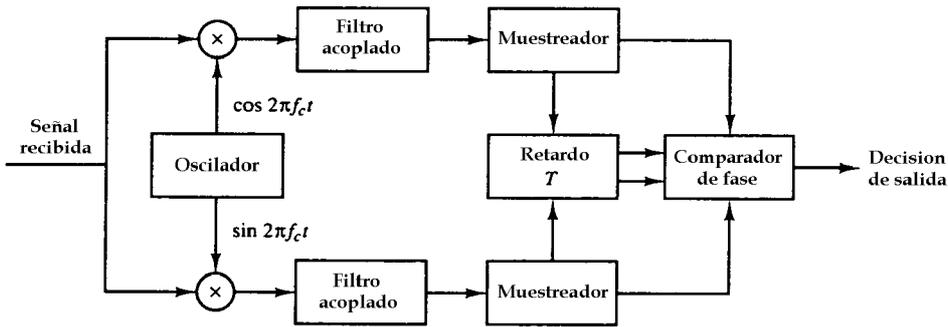


Figura 7.17: Diagrama en bloques de un demodulador DPSK.

La demodulación y detección de DPSK utilizando filtros acoplados se ilustra en la figura 7.17. Si el pulso $g_T(t)$ es rectangular los filtros acoplados pueden sustituirse por integradores.

7.2.5 Probabilidad de error para DPSK en un canal AWGN

Es posible considerar ahora la evaluación del desempeño de probabilidad de error de un demodulador y detector DPSK. La obtención del valor exacto de la probabilidad de error para DPSK M -ária es extremadamente complicada, excepto para $M = 2$. La mayor dificultad se encuentra en la determinación de la función densidad de probabilidad de la fase de la variable aleatoria $r_k r_{k-1}^*$ dada por (7.33). Sin embargo, es posible obtener una aproximación adecuada del desempeño de DSPK como se discutirá a continuación.

Sin pérdida de generalidad es posible suponer que la diferencia de fase $\theta_k - \theta_{k-1} = 0$. Además, los factores exponenciales $e^{-j(\theta_{k-1} - \phi)}$ y $e^{-j(\theta_k - \phi)}$ en (7.33) pueden incluirse en las componentes de ruido Gaussiano n_{k-1} y n_k sin cambiar sus propiedades estadísticas. De esta forma, (7.33) puede expresarse como

$$r_k r_{k-1}^* = \mathcal{E}_s + \sqrt{\mathcal{E}_s} (n_{k-1}^* + n_k) + n_k n_{k-1}^*$$

La complicación para determinar la función densidad de probabilidad de la fase está asociada al término $n_k n_{k-1}^*$. Sin embargo, para las SNR de interés el término $n_k n_{k-1}^*$ es pequeño en relación al término de ruido dominante $\sqrt{\mathcal{E}_s} (n_{k-1}^* + n_k)$. Despreciando el término $n_k n_{k-1}^*$ y normalizando $r_k r_{k-1}^*$ en relación a $\sqrt{\mathcal{E}_s}$, el nuevo conjunto de métricas de decisión será

$$\begin{aligned} x &= \sqrt{\mathcal{E}_s} + \text{Re}(n_{k-1}^* + n_k) \\ y &= \text{Im}(n_{k-1}^* + n_k) \end{aligned} \quad (7.34)$$

donde x e y son variables aleatorias Gaussianas no correlacionadas con varianzas idénticas $\sigma_n^2 = N_0$. La fase será

$$\Theta_r = \tan^{-1} \frac{y}{x}$$

En esta etapa se encuentra un problema similar al resuelto previamente para la demodulación de fase coherente. La única diferencia es que la varianza del ruido es ahora dos veces mayor que en el caso de PSK. De esta forma es posible concluir que el desempeño de DPSK es 3 dB más pobre que el de PSK. Este resultado es relativamente bueno para $M \geq 4$, pero es pesimista para $M = 2$ en el sentido que la pérdida en DPSK binaria relativa a PSK binaria es menor que 3 dB para SNR grandes. Esto último puede verificarse como sigue.

En DPSK binaria, las dos posibles diferencias de fase transmitidas son cero y π radianes. En consecuencia, solo es necesaria la parte real de $r_k r_{k-1}^*$ para recuperar la información. Usando (7.34) es posible expresar la parte real como

$$\text{Re}(r_k r_{k-1}^*) = \frac{1}{2}(r_k r_{k-1}^* + r_k^* r_{k-1})$$

Como la diferencia de fase entre dos intervalos de señalización sucesivos es cero, se incurre en un error si $\text{Re}(r_k r_{k-1}^*) < 0$. La probabilidad que $r_k r_{k-1}^* + r_k^* r_{k-1} < 0$ es un caso especial de la probabilidad de que una forma cuadrática general de variables aleatorias Gaussianas sea menor que cero (ver Capítulo 2). El resultado para la probabilidad de error para DPSK binaria es

$$P_2 = \frac{1}{2}e^{-\rho_b}$$

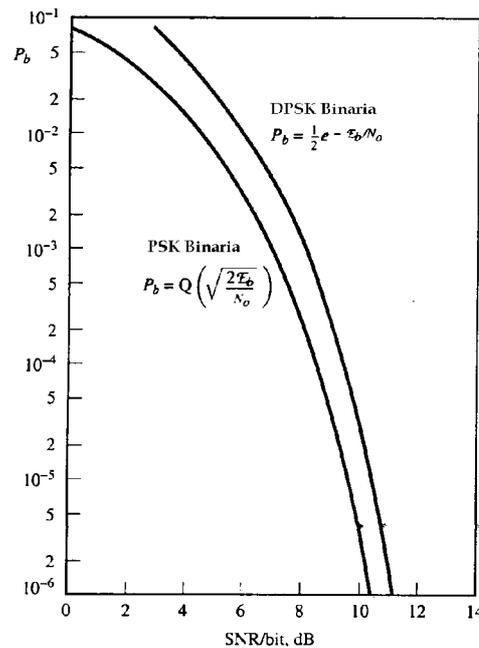


Figura 7.18: Probabilidad de error para PSK binaria y DPSK.

donde $\rho_b = \mathcal{E}_b/N_0$ es la SNR por bit, y se ilustra en la figura 7.18. También se muestra en esa figura la probabilidad de error para PSK binaria. Se observa que a probabilidades de error por debajo de 10^{-4} , la diferencia en SNR entre PSK binaria y DPSK binaria es menor de 1 dB.

7.3 Modulación de amplitud en cuadratura (QAM)

Una señal de amplitud modulada en cuadratura (QAM) utiliza dos portadoras en cuadratura $\cos 2\pi f_c t$ y $\sin 2\pi f_c t$ cada una con amplitud modulada de acuerdo con la secuencia de bits de información. Las señales transmitidas tendrán la forma

$$u_m(t) = A_{mc}g_T(t) \cos 2\pi f_c t + A_{ms}g_T(t) \sin 2\pi f_c t, \quad m = 1, 2, \dots, M \quad (7.35)$$

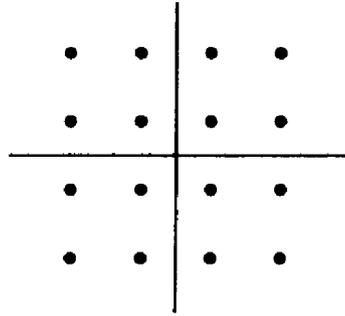


Figura 7.19: Constelación de una señal QAM $M = 16$.

donde A_{mc} y A_{ms} son los conjuntos de niveles de amplitud obtenidos mediante el mapeo de secuencias de k bits en amplitudes de señal. Por ejemplo, la figura 7.19 ilustra una constelación de señales de QAM de 16 puntos obtenido mediante la modulación de amplitud de cada portadora en cuadratura a través de PAM para $M = 4$. En general, cuando dos portadoras en cuadratura son moduladas via PAM se obtiene una constelación rectangular.

En general, QAM puede verse como una forma de combinación de modulación de amplitud y modulación de fase. Las señales QAM transmitidas pueden expresarse como

$$u_{mn}(t) = A_m g_T(t) \cos(2\pi f_c t + \theta_n), \quad m = 1, 2, \dots, M, \quad n = 1, 2, \dots, M \tag{7.36}$$

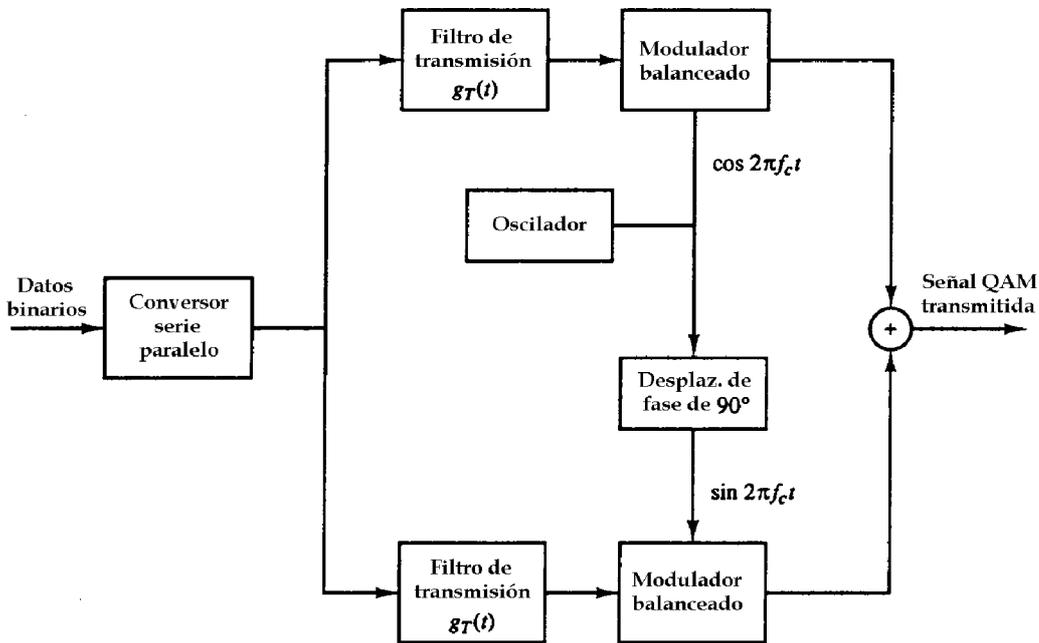


Figura 7.20: Diagrama en bloques funcional de un modulador para QAM.

Si $M_1 = 2^{k_1}$ y $M_2 = 2^{k_2}$, el método de combinación de modulación de amplitud y fase permite la transmisión simultánea de $k_1 + k_2 = \log_2 M_1 M_2$ dígitos binarios a una velocidad de $R_b / (k_1 + k_2)$. La figura 7.20 ilustra un diagrama en bloques funcional de un modulador QAM.

Es claro que la representación geométrica de las señales dadas por (7.35) y (7.36) está asociada a vectores de señal bidimensionales de la forma

$$\mathbf{s}_m = \left(\sqrt{\mathcal{E}_s} A_{mc} \quad \sqrt{\mathcal{E}_s} A_{ms} \right), \quad m = 1, 2, \dots, M$$

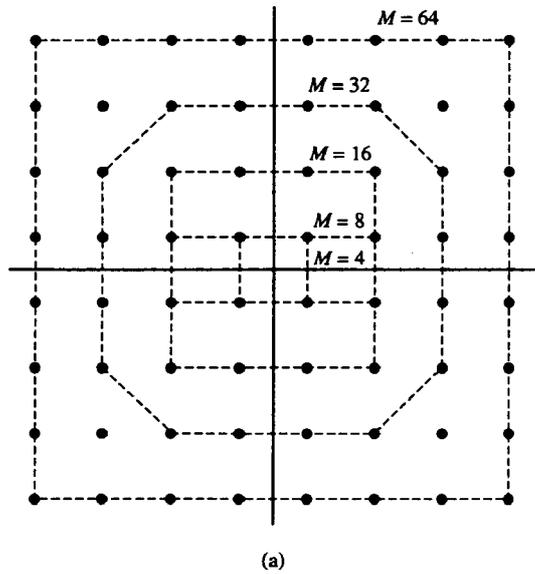


Figura 7.21: (a) Constelaciones rectangulares para QAM.

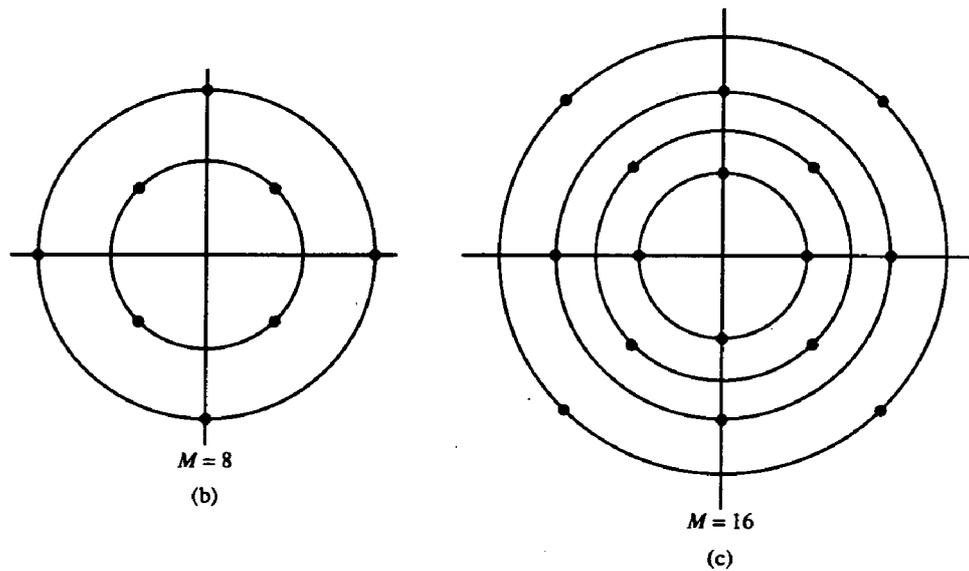


Figura 7.22: (b,c) Ejemplos de diagramas combinados PAM - PSK.

Ejemplos de constelaciones espaciales para QAM se ilustran en la figuras 7.21 y 7.22.

La densidad espectral de potencia de una señal QAM es idéntica en forma a la de una señal de fase modulada, dada por (7.26)-(7.28). La señal pasabanda transmitida $u(t)$ para una secuencia de símbolos de información puede expresarse como en (7.25), donde la señal banda base de información es

$$v(t) = \sum_{n=-\infty}^{\infty} A_n e^{j\theta_n} g_T(t - nT) \tag{7.37}$$

En esta ecuación, A_n es la amplitud y θ_n es la fase de la portadora transmitida. De esta forma, la autocorrelación de la secuencia de amplitudes y fases es

$$R_A(n) = E [A_k e^{-j\theta_k} A_{k+n} e^{j\theta_{k+n}}] = E [A_k A_{k+n} e^{j(\theta_{k+n} - \theta_k)}] \tag{7.38}$$

En el caso especial en que la secuencia de símbolos transmitidos no está correlacionada y tiene media cero, la función autocorrelación de (7.38) se reduce a

$$R_A(n) = \begin{cases} \sigma_A^2, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

donde $\sigma_A^2 = E[A_k^2]$. En este caso la densidad espectral de potencia de la secuencia de entrada es blanca y por ello la densidad espectral de potencia promedio de la señal transmitida depende solamente de las características espectrales del pulso transmitido, o sea, de $|G_T(f)|^2$.

7.3.1 Demodulación y detección de QAM

Es posible suponer que al transmitir la señal a través del canal se introduce un desplazamiento de fase. Además, la señal recibida estará perturbada por ruido aditivo Gaussiano. En consecuencia, $r(t)$ puede escribirse como

$$r(t) = A_{mc}g_T(t) \cos(2\pi f_c t + \phi) + A_{ms}g_T(t) \sin(2\pi f_c t + \phi) + n(t) \tag{7.39}$$

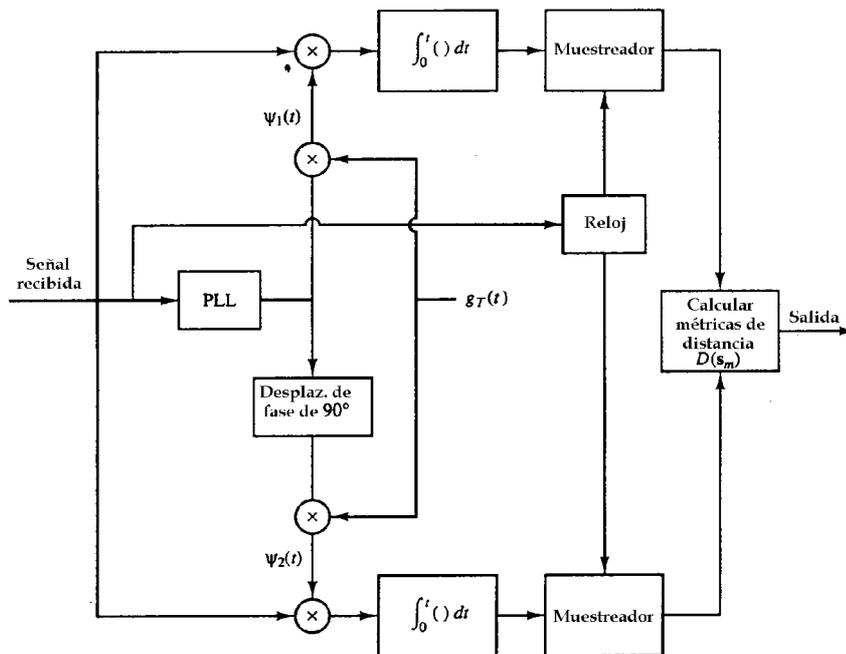


Figura 7.23: Demodulación y detección de señales QAM.

Es posible suponer además que se tiene disponible en el demodulador una estimación de la fase $\hat{\phi}$. Luego, la señal recibida estará correlacionada con las dos funciones base $\psi_1(t) = \sqrt{\frac{2}{\mathcal{E}_s}}g_T(t) \cos(2\pi f_c t + \hat{\phi})$ y $\psi_2(t) = -\sqrt{\frac{2}{\mathcal{E}_s}}g_T(t) \sin(2\pi f_c t + \hat{\phi})$, como ilustrado en la figura 7.23, tal que las salidas de los correladores se muestrean y se pasan al detector.

La entrada al detector consiste en dos muestras de r_1 y r_2 , donde ($\mathcal{E}_s \equiv \mathcal{E}_g/2$)

$$\begin{aligned} r_1 &= A_{mc}\sqrt{\mathcal{E}_s} \cos(\phi - \hat{\phi}) + A_{ms}\sqrt{\mathcal{E}_s} \sin(\phi - \hat{\phi}) + n_c \sin \hat{\phi} - n_s \cos \hat{\phi} \\ r_2 &= A_{mc}\sqrt{\mathcal{E}_s} \sin(\phi - \hat{\phi}) + A_{ms}\sqrt{\mathcal{E}_s} \cos(\phi - \hat{\phi}) + n_c \sin \hat{\phi} - n_s \cos \hat{\phi} \end{aligned}$$

Se observa que el efecto de la estimación de fase imperfecta es doble. Primero, las componentes de señal deseadas r_1 y r_2 están reducidas en amplitud por el factor $\cos(\phi - \hat{\phi})$. A su vez, esto reduce la SNR en

un factor proporcional a $\cos^2(\phi - \hat{\phi})$. Segundo, existe una interferencia de las componentes de señal en cuadratura en la señal deseada. Esta señal de interferencia, escalada por $\sin(\phi - \hat{\phi})$ puede producir una significativa degradación en el desempeño a menos que $\phi - \hat{\phi}$ sea pequeña. Esto sirve para enfatizar la importancia de disponer de una estimación precisa de la fase de la portadora para demodular una señal QAM.

El detector óptimo calcula las métricas de distancia dadas por

$$D(\mathbf{r}, \mathbf{s}_m) = |\mathbf{r} - \mathbf{s}_m|^2, \quad m = 1, 2, \dots, M \quad (7.40)$$

y selecciona la señal correspondiente al menor valor de $D(\mathbf{r}, \mathbf{s}_m)$. Si se utiliza una métrica de correlación en lugar de una métrica de distancia, es importante tener en cuenta que la métrica de correlación debe utilizar una corrección de sesgo debido a que las señales QAM no tienen igual energía.

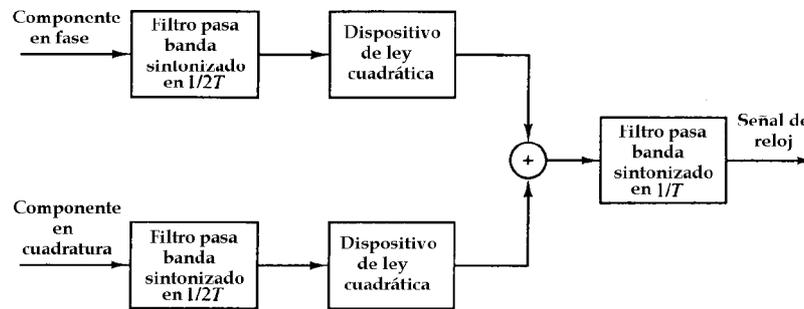


Figura 7.24: Diagrama en bloques del método de recuperación del temporizado para QAM.

El temporizado de símbolos puede extraerse de la señal recibida utilizando alguno de los métodos para sincronización de símbolos descritos en el Capítulo anterior. Los métodos de línea espectral allí descritos han probado ser particularmente efectivos para la recuperación del temporizado de señales QAM. La figura 7.24 ilustra un método de línea espectral basado en el filtrado de una componente de señal de frecuencia $1/2T$ y que eleva la salida al cuadrado para generar una señal sinusoidal a la velocidad de transmisión de símbolos deseada $1/T$. Como la demodulación de la señal de QAM se logra, como discutido anteriormente, mediante la multiplicación de la señal de entrada con las dos componentes en cuadratura de la portadora $\psi_1(t)$ y $\psi_2(t)$, las componentes en fase y cuadratura a la salida de los dos correladores se utilizan como entradas a los filtros pasabanda sintonizados en $1/2T$. Las salidas de los dos filtros se elevan al cuadrado, se suman y luego se filtran mediante un filtro de banda estrecha sintonizado a la frecuencia del reloj $1/T$. De esta forma, se genera una señal sinusoidal adecuada como señal de reloj para muestrear las salidas de los correladores para recuperar la información.

Finalmente, debería considerarse el problema de la distorsión lineal del canal que puede encontrarse en un sistema práctico de comunicaciones que utiliza modulación QAM. Si la característica de respuesta en frecuencia del canal $C_{pb}(f)$ (o la respuesta banda base equivalente) es conocida, es posible utilizar la metodología descrita en el caso de PAM para diseñar los filtros óptimos de transmisión y recepción. Por otro lado, si la respuesta en frecuencia del canal no es conocida a priori puede utilizarse un ecualizador adaptivo para compensar la distorsión del canal. El ecualizador puede implementarse en la banda pasante o en banda base.

7.3.2 Probabilidad de error para QAM en un canal AWGN

Para determinar la probabilidad de error para QAM se debe especificar la constelación de puntos de señal. Es posible comenzar con conjuntos de señales QAM que tienen $M = 4$ puntos. La figura 7.25 ilustra dos conjuntos de señales de cuatro puntos. El primero es una señal modulada de cuatro fases y el segundo es una señal QAM con dos niveles de amplitud A_1 y A_2 y cuatro fases. Como la probabilidad de error está dominada por la distancia mínima entre pares de puntos, es posible imponer la condición $d_{min} = 2A$ para ambas constelaciones de señal y evaluar la potencia promedio transmitida, basándose en la premisa que todos los puntos de señal son igualmente probables. Para la señal de cuatro fases se tendrá que

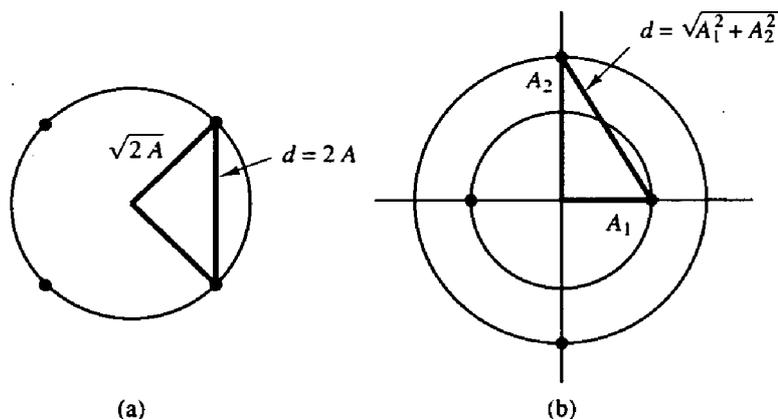


Figura 7.25: Dos constelaciones de señal de cuatro puntos.

$$P_{av} = \frac{1}{4}(4)A^2 = A^2$$

Para la QAM de dos amplitudes y cuatro fases se colocan los puntos de la constelación sobre círculos de radio A y $\sqrt{3}A$. De esa forma, $d_{min} = 2A$ y

$$P_{av} = \frac{1}{4} \left[2 \left(\frac{3}{2} \right) A^2 + 2 \frac{A^2}{2} \right] = A^2$$

que es el mismo resultado que para la constelación de cuatro fases. De esta forma, para todos los propósitos prácticos, el desempeño de error para los dos conjuntos de señales es el mismo. En otras palabras, no existirá ventaja en la elección del conjunto QAM de dos amplitudes en relación al de cuatro fases.

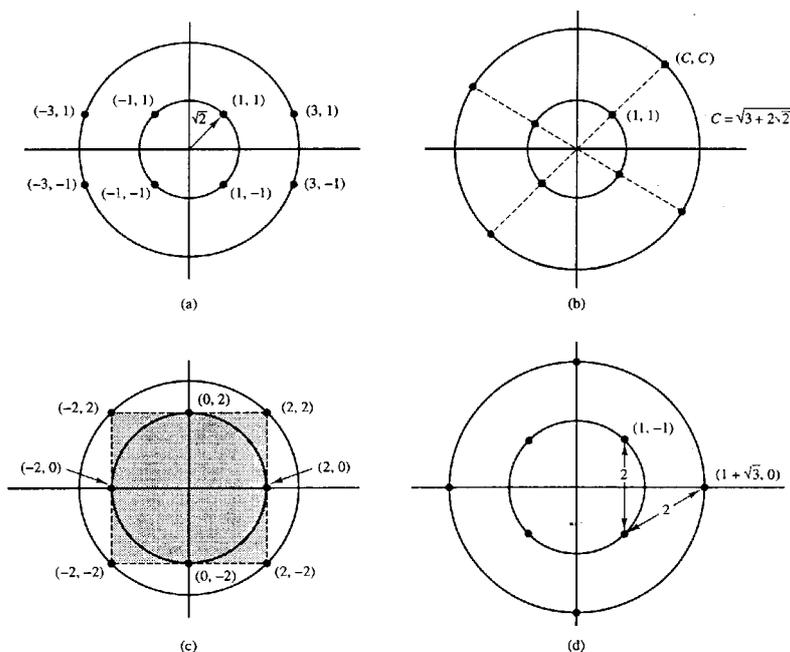


Figura 7.26: Cuatro constelaciones de QAM $M = 8$.

Luego, es posible considerar QAM de $M = 8$. En este caso existen varias constelaciones de señales posibles. Se considerarán las cuatro constelaciones de señal mostradas en la figura 7.26, las cuales consisten

en dos amplitudes que tienen una distancia mínima entre puntos de $2A$. Las coordenadas (A_{mc}, A_{ms}) para cada punto de señal, normalizadas por A , se ilustran también en la figura. Suponiendo que los puntos de señal son igualmente probables, la potencia promedio transmitida será

$$P_{av} = \frac{1}{M} \sum_{m=1}^M \frac{1}{2} (A_{mc}^2 + A_{ms}^2) = \frac{A^2}{2M} \sum_{m=1}^M (a_{mc}^2 + a_{ms}^2)$$

donde (a_{mc}, a_{ms}) son las coordenadas de los puntos de señal, normalizadas por A .

Los dos primeros conjuntos de señal en la figura 7.26 contienen puntos que residen en una grilla rectangular y tienen $P_{av} = 3A^2$. El tercer conjunto de señal requiere una potencia promedio transmitida $P_{av} = 3.41A^2$ y el cuarto requiere $P_{av} = 2.36A^2$. En consecuencia, el cuarto conjunto de señales requiere aproximadamente 1 dB menos de potencia que los dos primeros y 1.6 dB menos que el tercero para lograr la misma probabilidad de error. Esta constelación de señales se conoce como la mejor QAM de 8 puntos debido a que requiere la menor potencia para una distancia mínima dada entre puntos de señal.

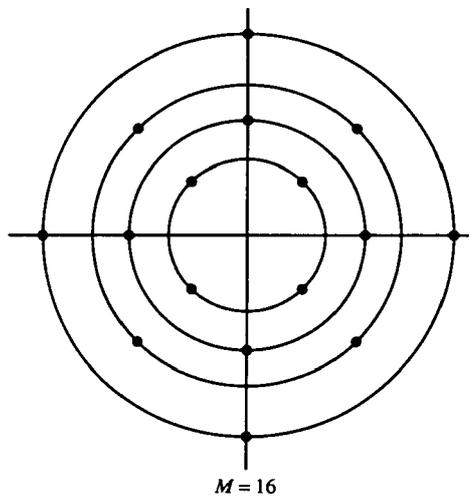


Figura 7.27: Constelaciones QAM circulares de $M = 16$ puntos.

Para $M \geq 16$, existirán muchas más posibilidades para seleccionar los puntos de señal QAM en el espacio bidimensional. Por ejemplo, es posible elegir una constelación circular multi-amplitud para $M = 16$, como mostrado en la figura 7.27. En este caso, los puntos de señal en un nivel de amplitud dado están rotados en fase por $\pi/4$ en relación a los puntos de señal de nivel de amplitud adyacente. Esta constelación QAM 16 es una generalización de la constelación óptima QAM 8. Sin embargo, la constelación QAM 16 no es la mejor constelación para QAM de 16 puntos para un canal AWGN.

Las constelaciones QAM rectangulares tienen una ventaja distintiva al ser fácilmente generables como dos señales PAM asociadas a portadoras en cuadratura. Además, son fácilmente demodulables como discutido previamente. A pesar de no ser las mejores constelaciones para QAM M -aria para $M \geq 16$, la potencia promedio requerida para lograr una distancia mínima dada es solo levemente mayor que la potencia promedio requerida para la constelación de señales de la mejor QAM M -aria. Por esas razones, las señales QAM M -arias rectangulares son las más frecuentemente utilizadas en la práctica.

Para constelaciones de señales en las cuales $M = 2^k$, donde k es par, la constelación QAM es equivalente a dos señales PAM con portadoras en cuadratura, cada una con $\sqrt{M} = 2^{k/2}$ puntos de señal. Como las señales en las componentes en fase y cuadratura están perfectamente separadas mediante detección coherente cuando $\hat{\phi} = \phi$, la probabilidad de error para QAM se determina fácilmente de la probabilidad de error para PAM. Específicamente, la probabilidad de una decisión correcta para un sistema QAM M -ario será

$$P_c = (1 - P_{\sqrt{M}})^2$$

donde $P_{\sqrt{M}}$ es la probabilidad de error de una PAM \sqrt{M} -aria con la mitad de potencia promedio en cada componente en cuadratura del sistema QAM equivalente. Mediante la modificación apropiada de la probabilidad de error para PAM M -aria se obtiene

$$P_{\sqrt{M}} = 2 \left(1 - \frac{1}{\sqrt{M}} \right) Q \left(\sqrt{\frac{3}{M-1} \frac{\mathcal{E}_{av}}{N_0}} \right)$$

donde \mathcal{E}_{av}/N_0 es la SNR promedio por símbolo. De esta forma, la probabilidad de un error de símbolo para QAM M -ária será

$$P_M = 1 - (1 - P_{\sqrt{M}})^2 \quad (7.41)$$

Notar que este resultado es exacto para $M = 2^k$ cuando k es par. Por otro lado, cuando k es impar no existirá un sistema PAM \sqrt{M} -ário equivalente. Esto no es un problema ya que es mucho más simple determinar la probabilidad de error para un conjunto de señales rectangular. Si se utiliza un detector óptimo que basa su decisión en métricas de distancia óptimas dadas por (7.40), es relativamente directo mostrar que la probabilidad de error de símbolo es acotada precisamente en forma superior por

$$\begin{aligned} P_M &\leq 1 - \left[1 - 2Q \left(\sqrt{\frac{3}{M-1} \frac{\mathcal{E}_{av}}{N_0}} \right) \right]^2 \\ &\leq 4Q \left(\sqrt{\frac{3}{M-1} \frac{k\mathcal{E}_{bav}}{N_0}} \right) \end{aligned}$$

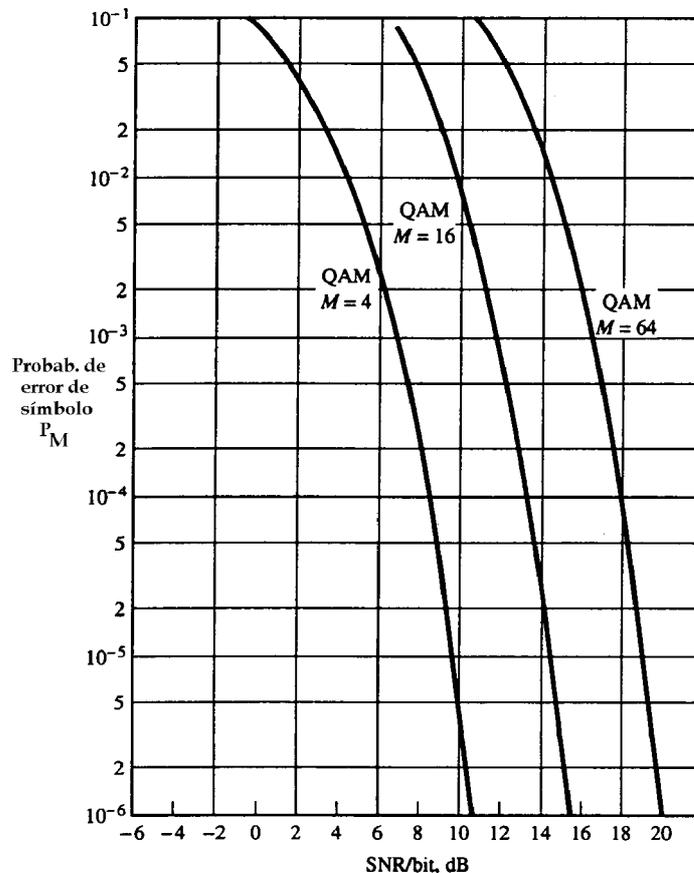


Figura 7.28: Probabilidad de error de símbolo para QAM.

Para cualquier $k \geq 1$, donde \mathcal{E}_{bav}/N_0 es la SNR promedio por bit. La probabilidad de un error de símbolo se grafica en la figura 7.28 en función de la SNR promedio por bit.

M	$10 \log_{10} \mathcal{R}_m$
8	1.65
16	4.20
32	7.02
64	9.95

Tabla 7.1: Ventaja de ANR de QAM M -ária sobre PSK M -ária

M	$10 \log_{10} 2(M-1)/3$
4	3.00
8	6.70
16	10.00
32	13.20
64	16.20
128	19.20

Tabla 7.2: Incremento en potencia promedio (dB) para QAM relativos a $M = 2$.

Es interesante comparar el desempeño de QAM con el de modulación de fase para cualquier tamaño M , debido a que ambos tipos de señales son bidimensionales. Recordar que para la modulación de fase M -ária, la probabilidad de error de símbolo era aproximadamente

$$P_M \cong 2Q\left(\sqrt{2\rho_s} \sin \frac{\pi}{M}\right)$$

donde ρ_s es la SNR por símbolo. Para QAM M -ária es posible usar la expresión de (7.41). Como la probabilidad de error está dominada por el argumento de la función Q , es posible comparar simplemente los argumentos de Q para los dos formatos de señal. De esta forma, la relación de esos dos argumentos es

$$\mathcal{R}_M = \frac{3/(M-1)}{2 \sin^2 \pi/M}$$

Por ejemplo, cuando $M = 4$ se tiene $\mathcal{R}_M = 1$. En consecuencia, PSK 4 y QAM 4 producen un desempeño comparable para la misma SNR por bit. Por otro lado, cuando $M > 4$ entonces $\mathcal{R} > 1$, de forma que QAM M -ária produce mejor desempeño que PSK M -ária. La tabla 7.1 ilustra la ventaja de SNR de QAM sobre PSK para diferentes valores de M . Por ejemplo, se observa que QAM 32 tiene 7 dB de ventaja en SNR sobre PSK 32.

Finalmente, es posible notar que QAM, PAM y PSK son métodos de modulación lineal que permiten incrementar la relación velocidad de transmisión/ancho de banda (R_b/W) mediante el incremento del número de puntos de señal $M = 2^k$ en la constelación, donde k es el número de bits de información asociados a cada símbolo transmitido. La velocidad de incremento de R_b/W es logarítmica en M (lineal en k). Ya se ha discutido que el costo de duplicar el número de fases (incrementar el número de bits por símbolo en un bit) en PSK se aproxima a 6 dB (o en un factor de 4) de potencia transmitida adicional para M grande. Una comparación similar para QAM (lo mismo que para PAM) indica que el incremento en potencia transmitida es de aproximadamente 3 dB por bit/símbolo adicional. En la tabla 7.2 se muestran los valores de $10 \log_{10} 2(M-1)/3$ que representan el incremento en potencia promedio requerido para mantener un nivel de desempeño dado (en relación a $M = 2$) en QAM cuando aumenta el número de puntos de señal en la constelación rectangular. De esta forma se observa que QAM (y PAM) son preferibles a PSK para tamaños de constelación grandes.

7.3.3 Estimación de fase de la portadora

Como se ha discutido anteriormente la demodulación de QAM requiere una portadora en fase con la portadora de la señal recibida. La estimación de la fase de la portadora para QAM puede lograrse por diferentes medios, dependiendo de la constelación de puntos de señal y las relaciones de fase de los diferentes puntos de señal.

Por ejemplo, es posible considerar las constelaciones QAM de 8 puntos de las figuras 7.26c y 7.26d. Los puntos de señal en esas constelaciones toman uno de dos posibles valores de amplitud y ocho posibles fases. Las fases están espaciadas en $\pi/4$. Esta simetría de fase permite utilizar un PLL asociado a la salida de un dispositivo con una ley no lineal de octava potencia que genera una componente de portadora en $8f_c$, donde f_c es la frecuencia de la portadora. De esta forma, puede utilizarse el método ilustrado en la figura 7.15 en cualquier constelación QAM que contenga puntos de señal con fases múltiplo de algún ángulo de fase θ , donde $L\theta = 360^\circ$, para algún L entero.

Otro método para extraer una estimación de la fase de la portadora $\hat{\phi}$ de la señal QAM M -aria recibida es el DFPLL. La idea básica en el DFPLL es estimar la fase de la señal QAM en cada intervalo de señal y remover la modulación de fase de la portadora. El DFPLL puede utilizarse con cualquier señal QAM independientemente de las relaciones de fase entre los puntos de señal.

Para ser específicos, es posible expresar la señal QAM recibida en la forma

$$r(t) = A_m g_T(t) \cos(2\pi f_c t + \theta_n + \phi) + n(t)$$

donde θ_n es la fase de un punto de señal y ϕ es la fase de la portadora. Esta señal se demodula a partir de la correlación cruzada entre $r(t)$ con $\psi_1(t)$ y $\psi_2(t)$. Los valores de las muestras a la salida de los correladores serán

$$\begin{aligned} r_1 &= A_m \sqrt{\mathcal{E}_s} \cos(\theta_n + \phi - \hat{\phi}) + n_c \cos(\theta_n + \phi - \hat{\phi}) - n_s \sin(\theta_n + \phi - \hat{\phi}) \\ r_2 &= A_m \sqrt{\mathcal{E}_s} \sin(\theta_n + \phi - \hat{\phi}) + n_c \sin(\theta_n + \phi - \hat{\phi}) - n_s \cos(\theta_n + \phi - \hat{\phi}) \end{aligned}$$

Suponiendo que el detector, basandose en r_1 y r_2 , ha decidido correctamente sobre el punto de señal transmitido, entonces multiplicando r_1 por $-\sin \theta_n$ y a r_2 por $\cos \theta_n$ se obtiene

$$\begin{aligned} -r_1 \sin \theta_n &= -A_m \sqrt{\mathcal{E}_s} \cos(\theta_n + \phi - \hat{\phi}) \sin \theta_n + \text{componente de ruido} \\ &= A_m \sqrt{\mathcal{E}_s} [-\sin \theta_n \cos \theta_n \cos(\phi - \hat{\phi}) + \sin^2 \theta_n \sin(\phi - \hat{\phi})] + \text{componente de ruido} \\ r_2 \cos \theta_n &= A_m \sqrt{\mathcal{E}_s} \sin(\theta_n + \phi - \hat{\phi}) \cos \theta_n + \text{componente de ruido} \\ &= A_m \sqrt{\mathcal{E}_s} [\sin \theta_n \cos \theta_n \cos(\phi - \hat{\phi}) + \cos^2 \theta_n \sin(\phi - \hat{\phi})] + \text{componente de ruido} \end{aligned}$$

Sumando esos dos términos se obtiene la señal de error $e(t)$ dada por

$$e(t) = e_2 \cos \theta_n - r_1 \sin \theta_n = A_m \sqrt{\mathcal{E}_s} \sin(\phi - \hat{\phi}) + \text{componentes de ruido}$$

Esta señal de error se pasa luego al filtro de lazo que alimenta al VCO. De esta forma, solo se utiliza la fase de la señal QAM para obtener una estimación de la fase de la portadora. En consecuencia, el diagrama en bloques general para DFPLL, mostrado en la figura 7.16, también se aplica a la estimación de la fase de la portadora para una señal QAM M -aria.

Como en el caso de señales moduladas en fase, los métodos descriptos para recuperación de la fase de la portadora introducen ambigüedades de fase. Este problema puede resolverse en general mediante la codificación diferencial de la secuencia de datos a la entrada del modulador.

7.4 Modulación en frecuencia (FSK)

Se han discutido métodos para la transmisión de información digital mediante modulación de la amplitud o la fase de la portadora o una combinación de ambas. La información digital puede transmitirse también mediante la modulación de la frecuencia de la portadora.

Como se observará en la discusión siguiente, la transmisión digital a través de modulación en frecuencia es un método no lineal apropiado para canales que carecen de estabilidad de fase, necesaria para realizar la estimación de la fase de la portadora. En contraste, los métodos de modulación lineal que se han introducido, o sea PAM, PM y QAM, requieren la estimación de la fase de la portadora para realizar detección coherente.

La forma más simple de modulación en frecuencia es *llaveado por desplazamiento de fase* (Frequency-Shift keying, FSK) binario. En FSK binaria se utilizan dos frecuencias diferentes, f_1 y $f_2 = f_1 + \Delta f$, para

transmitir una secuencia de información. La elección de la separación de frecuencias $\Delta f = f_2 - f_1$ se discute a continuación. Las dos señales pueden expresarse de la siguiente forma

$$\begin{aligned} u_1(t) &= \sqrt{\frac{2\mathcal{E}_b}{T_b}} \cos 2\pi f_1 t, & 0 \leq t \leq T_b \\ u_2(t) &= \sqrt{\frac{2\mathcal{E}_b}{T_b}} \cos 2\pi f_2 t, & 0 \leq t \leq T_b \end{aligned} \quad (7.42)$$

donde \mathcal{E}_b es la energía de señal por bit y T_b es la duración del intervalo de bit.

En general, puede utilizarse una FSK M -ária para transmitir un bloque de $k = \log_2 M$ bits por señal. En este caso, las M señales pueden expresarse como

$$u_m(t) = \sqrt{\frac{2\mathcal{E}_s}{T}} \cos(2\pi f_c t + 2\pi m \Delta f t), \quad 0 \leq t \leq T \quad (7.43)$$

donde $\mathcal{E}_s = k\mathcal{E}_b$ es la energía por símbolo, $T = kT_b$ es el intervalo de símbolo y Δf es la separación de frecuencia entre frecuencias sucesivas, o sea, $\Delta f = f_m - f_{m-1}$, para todo $m = 1, 2, \dots, M-1$, donde $f_m = f_c + m\Delta f$.

Notar que las M señales FSK tienen igual energía \mathcal{E}_s . La separación en frecuencia Δf determina el grado por el cual es posible discriminar entre las M posibles señales transmitidas. Como medida de la similitud entre un par de señales es posible utilizar los coeficientes de correlación γ_{mn} , dados por

$$\gamma_{mn} = \frac{1}{\mathcal{E}_s} \int_0^T u_m(t) u_n(t) dt \quad (7.44)$$

Sustituyendo $u_m(t)$ y $u_n(t)$ en (7.44) se obtiene

$$\begin{aligned} \gamma_{mn} &= \frac{1}{\mathcal{E}_s} \int_0^T \frac{2\mathcal{E}_s}{T} \cos(2\pi f_c t + 2\pi m \Delta f t) \cos(2\pi f_c t + 2\pi n \Delta f t) dt \\ &= \frac{1}{T} \int_0^T \cos[2\pi(m-n)\Delta f t] dt + \frac{1}{T} \int_0^T \cos[4\pi f_c t + 2\pi(m+n)\Delta f t] dt \\ &= \frac{\sin 2\pi(m-n)\Delta f T}{2\pi(m-n)\Delta f T} \end{aligned} \quad (7.45)$$

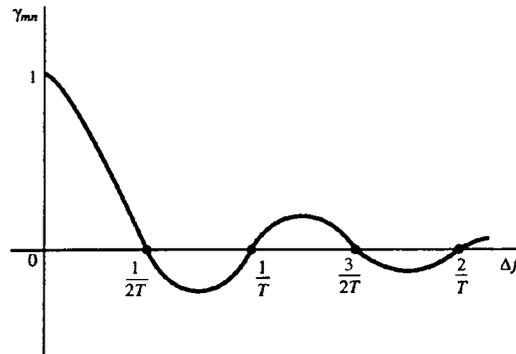


Figura 7.29: Coeficiente de correlación cruzada en función de la separación de frecuencia para señales FSK.

donde la segunda integral se anula cuando $f_c \gg 1/T$. Un gráfico de γ_{mn} en función de la separación en frecuencia Δf se muestra en la figura 7.29. Se observa que las señales son ortogonales cuando Δf es un múltiplo de $1/2T$. Por ello, la separación de frecuencia mínima entre frecuencias sucesivas para ortogonalidad

es $1/2T$. Es posible notar también que el valor mínimo del coeficiente de correlación es $\gamma_{mn} = -0.217$, que ocurre en la separación de frecuencias $\Delta f = 0.715/T$.

Las señales FSK ortogonales M -árias tienen una representación geométrica de M vectores ortogonales, dados por

$$\begin{aligned} \mathbf{s}_1 &= (\sqrt{\mathcal{E}_s}, 0, \dots, 0) \\ \mathbf{s}_2 &= (0, \sqrt{\mathcal{E}_s}, 0, \dots, 0) \\ &\vdots \\ \mathbf{s}_M &= (0, 0, \dots, 0, \sqrt{\mathcal{E}_s}) \end{aligned}$$

donde las funciones base serán $\psi_m(t) = \sqrt{2/T} \cos 2\pi(f_c + m\Delta f)t$. La distancia entre pares de vectores de señal es $d = \sqrt{2\mathcal{E}_s}$ para todo m, n , la cual también es la distancia mínima entre las M señales.

7.4.1 Demodulación y detección de FSK

Es posible suponer que las señales FSK son transmitidas a través de un canal AWGN. Además, se asumirá que cada señal sufre un retardo en la transmisión a través del canal. En consecuencia, la señal filtrada recibida a la entrada del demodulador puede expresarse como

$$r(t) = \sqrt{\frac{2\mathcal{E}_s}{T}} \cos(2\pi f_c t + 2\pi m \Delta f t + \phi_m) + n(t) \quad (7.46)$$

donde ϕ_m es el desplazamiento de fase de la m -ésima señal (debido al retardo de transmisión) y $n(t)$ representa el ruido pasabanda aditivo que puede ser expresado como

$$n(t) = n_c(t) \cos 2\pi f_c t - n_s(t) \sin 2\pi f_c t$$

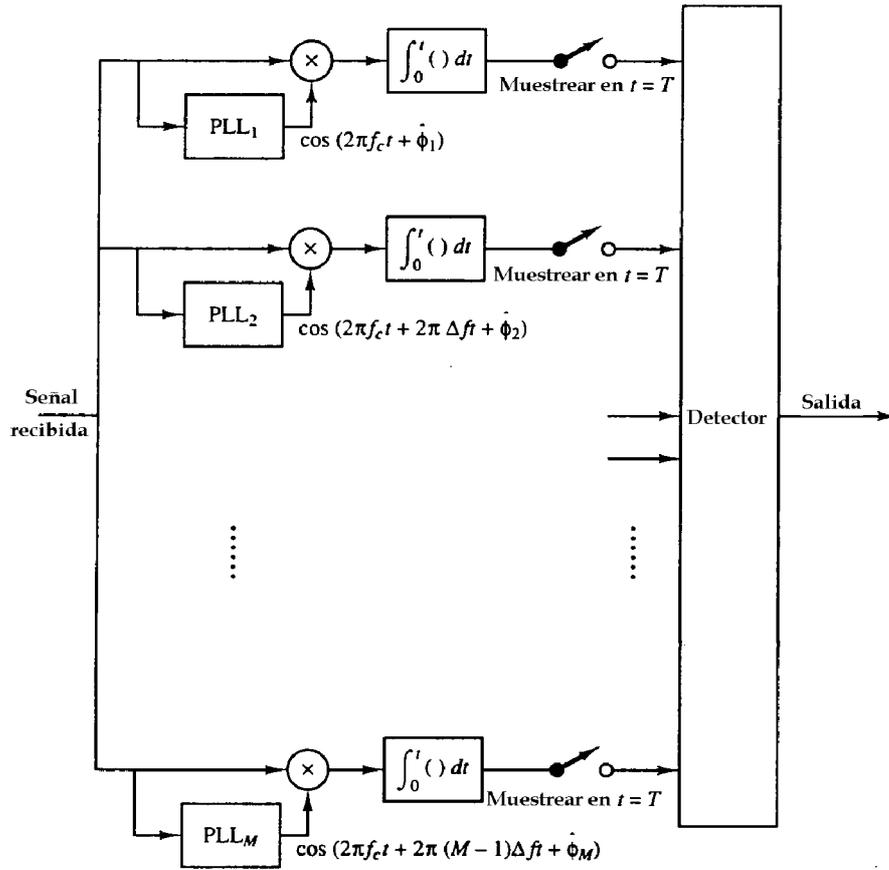
La demodulación y detección de las M señales FSK puede realizarse por dos métodos diferentes. Una aproximación es estimar los M desplazamientos de fase $\{\phi_m\}$ y efectuar una *demodulación y detección coherentes*. Como método alternativo, las fases de las portadoras pueden ignorarse en la demodulación y detección.

En la demodulación de fase coherente la señal recibida $r(t)$ se correlaciona con cada una de las M posibles señales recibidas $\cos(2\pi f_c t + 2\pi m \Delta f t + \hat{\phi}_m)$, $m = 0, 1, \dots, M - 1$ donde $\hat{\phi}_m$ son las estimaciones de fase de las portadoras. Un diagrama en bloques ilustrando este tipo de demodulación se muestra en la figura 7.30. Es interesante notar que cuando $\hat{\phi}_m \neq \phi_m$, la separación de frecuencia requerida para señalización ortogonal en el demodulador es $\Delta f = 1/T$, la cual es dos veces la separación mínima para ortogonalidad cuando $\hat{\phi}_m = \phi_m$.

El requerimiento para estimar las M fases de las portadoras hace que la demodulación coherente de señales FSK sea extremadamente compleja e impráctica, especialmente cuando el número de señales es grande. En consecuencia, no se considerará la detección coherente de señales FSK.

En su lugar se considerará un método para la demodulación y detección que no requiere el conocimiento de las fases de las portadoras. La demodulación puede lograrse como mostrado en la figura 7.31. En este caso hay dos correladores por señal, o un total de $2M$ correladores, en general. La señal recibida se correlaciona con las funciones base (portadoras en cuadratura) $\sqrt{2/T} \cos 2\pi(f_c + m\Delta f)t$ y $\sqrt{2/T} \sin 2\pi(f_c + m\Delta f)t$ para $m = 0, 1, \dots, M - 1$. Las $2M$ salidas de los correladores se muestrean al final del intervalo de señalización y se pasan al detector. De esta forma, si se transmite la m -ésima señal, las $2M$ muestras en el detector pueden expresarse como

$$\begin{aligned} r_{kc} &= \sqrt{\mathcal{E}_s} \left[\frac{\sin 2\pi(k-m)\Delta f T}{2\pi(k-m)\Delta f T} \cos \phi_m - \frac{\cos 2\pi(k-m)\Delta f T - 1}{2\pi(k-m)\Delta f T} \sin \phi_m \right] + n_{kc} \\ r_{ks} &= \sqrt{\mathcal{E}_s} \left[\frac{\cos 2\pi(k-m)\Delta f T - 1}{2\pi(k-m)\Delta f T} \cos \phi_m + \frac{\sin 2\pi(k-m)\Delta f T}{2\pi(k-m)\Delta f T} \sin \phi_m \right] + n_{ks} \end{aligned}$$

Figura 7.30: Demodulación de fase coherente para señales FSK M -arias.

donde n_{kc} y n_{ks} son las componentes de ruido Gaussiano en las muestras de salida.

Es posible observar que cuando $k = m$ los valores de las muestras al detector serán

$$\begin{aligned} r_{mc} &= \sqrt{\mathcal{E}_s} \cos \phi_m + n_{mc} \\ r_{ms} &= \sqrt{\mathcal{E}_s} \sin \phi_m + n_{ms} \end{aligned}$$

Además, se observa que cuando $k \neq m$, las componentes de señal en las muestras r_{kc} y r_{ks} desaparecerán, independientemente de los valores de desplazamiento de fase ϕ_k , previendo que la separación de frecuencia entre muestras sucesivas sea $\Delta f = 1/T$. En tal caso, las otras $2(M-1)$ salidas de los correladores consistirán en ruido solamente, o sea

$$r_{kc} = n_{kc} \quad r_{ks} = n_{ks} \quad k \neq m$$

En el desarrollo siguiente se asumirá que $\Delta f = 1/T$ tal que las señales serán ortogonales.

Es posible mostrar fácilmente que las $2M$ muestras n_{kc} y n_{ks} son variables aleatorias Gaussianas no correlacionadas de media cero y varianza $\sigma^2 = N_0/2$. En consecuencia, la función densidad de probabilidad conjunta para r_{mc} y r_{ms} condicionada sobre ϕ_m será

$$f_{\mathbf{r}_m}(r_{mc}, r_{ms} | \phi_m) = \frac{1}{2\pi\sigma^2} e^{-[(r_{mc} - \sqrt{\mathcal{E}_s} \cos \phi_m)^2 + (r_{ms} - \sqrt{\mathcal{E}_s} \sin \phi_m)^2] / 2\sigma^2} \quad (7.47)$$

y para $m \neq k$ se tiene que

$$f_{\mathbf{r}_k}(r_{kc}, r_{ks}) = \frac{1}{2\pi\sigma^2} e^{-[r_{kc}^2 + r_{ks}^2] / 2\sigma^2} \quad (7.48)$$

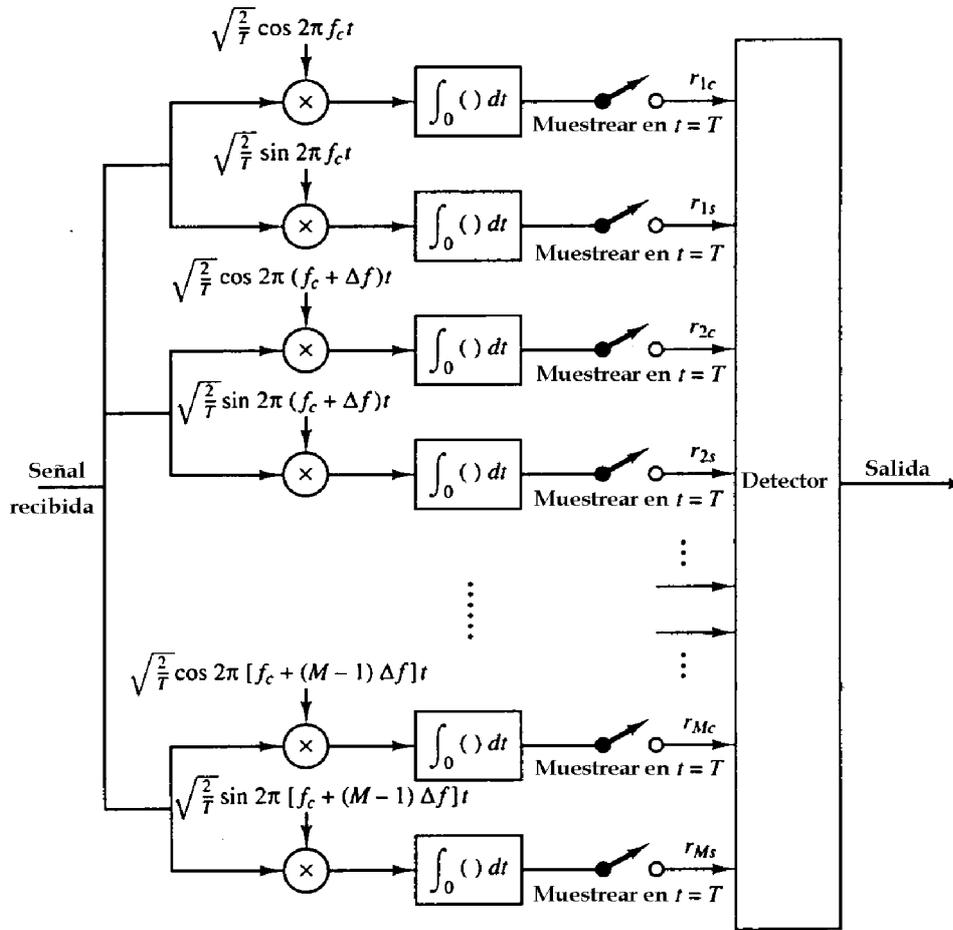


Figura 7.31: Demodulación de señales FSK M -árias para detección no coherente.

Dadas las $2M$ variables aleatorias observadas $\{r_{kc}, r_{ks}, k = 1, 2, \dots, M\}$, el detector óptimo selecciona la señal que corresponde al máximo de las probabilidades a posteriori, o sea

$$P[s_m \text{ fué trasmitido} | \mathbf{r}] \equiv P(s_m | \mathbf{r}), \quad m = 1, 2, \dots, M \tag{7.49}$$

donde \mathbf{r} es el vector de $2M$ dimensiones con elementos $\{r_{kc}, r_{ks}, k = 1, 2, \dots, M\}$.

Detector óptimo para FSK binaria

Se obtendrá la forma para el detector óptimo no coherente para el caso FSK binario. La generalización para el caso de FSK M -ario es directa.

En FSK ortogonal binaria, las dos probabilidades a posteriori serán

$$P(s_1 | \mathbf{r}) = \frac{f_{\mathbf{r}}(\mathbf{r} | s_1) P(s_1)}{f_{\mathbf{r}}(\mathbf{r})}$$

$$P(s_2 | \mathbf{r}) = \frac{f_{\mathbf{r}}(\mathbf{r} | s_2) P(s_2)}{f_{\mathbf{r}}(\mathbf{r})}$$

y, de esta forma, la regla de detección óptima puede expresarse como

$$P(\mathbf{s}_1|\mathbf{r}) \begin{matrix} s_1 \\ > \\ < \\ s_2 \end{matrix} P(\mathbf{s}_2|\mathbf{r})$$

o equivalentemente

$$\frac{f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_1)P(\mathbf{s}_1)}{f_{\mathbf{r}}(\mathbf{r})} \begin{matrix} s_1 \\ > \\ < \\ s_2 \end{matrix} \frac{f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_2)P(\mathbf{s}_2)}{f_{\mathbf{r}}(\mathbf{r})} \quad (7.50)$$

donde \mathbf{r} es el vector de cuatro dimensiones $\mathbf{r} = (r_{1c}, r_{1s}, r_{2c}, r_{2s})$. La relación en (7.50) se puede simplificar de la siguiente forma

$$\frac{f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_1)}{f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_2)} \begin{matrix} s_1 \\ > \\ < \\ s_2 \end{matrix} \frac{P(\mathbf{s}_2)}{P(\mathbf{s}_1)} \quad (7.51)$$

La relación de funciones densidad de probabilidad en el lado izquierdo de (7.51) es la relación de verosimilitud, la cual puede escribirse como

$$\Lambda(\mathbf{r}) = \frac{f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_2)}{f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_1)} \quad (7.52)$$

El lado derecho de (7.51) es la relación de dos probabilidades a priori, las cuales toman el valor de unidad cuando las dos señales son igualmente probables.

Las funciones densidad de probabilidad $f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_1)$ y $f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_2)$ en la relación de verosimilitud pueden expresarse como

$$\begin{aligned} f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_1) &= f_{\mathbf{r}_2}(r_{2c}, r_{2s}) \int_0^{2\pi} f_{\mathbf{r}_1}(r_{1c}, r_{1s}|\phi_1) f_{\phi_1}(\phi_1) d\phi_1 \\ f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_2) &= f_{\mathbf{r}_1}(r_{1c}, r_{1s}) \int_0^{2\pi} f_{\mathbf{r}_2}(r_{2c}, r_{2s}|\phi_2) f_{\phi_2}(\phi_2) d\phi_2 \end{aligned}$$

donde $f_{\mathbf{r}_m}(r_{mc}, r_{ms}|\phi_m)$ y $f_{\mathbf{r}_k}(r_{kc}, r_{ks})$, $m \neq k$ están dadas por (7.47) y (7.48), respectivamente. Las fases ϕ_1 y ϕ_2 se eliminan simplemente promediando $f_{\mathbf{r}_m}(r_{mc}, r_{ms}|\phi_m)$.

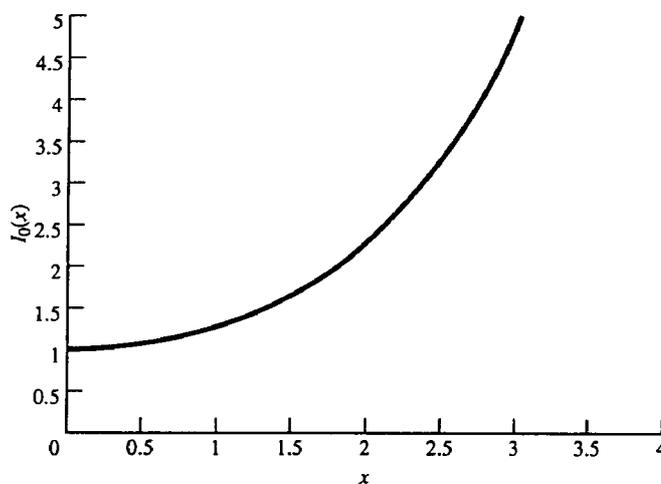
Una función densidad de probabilidad uniforme para ϕ_m representa la mayor ignorancia en relación a las fases de las portadoras. Esto se denomina la *función densidad de probabilidad menos favorable* para ϕ_m . Con $f_{\phi_m}(\phi_m) = 1/2\pi$, $0 \leq \phi_m < 2\pi$, sustituida en las integrales anteriores se obtiene

$$\frac{1}{2\pi} \int_0^{2\pi} f_{\mathbf{r}_m}(r_{mc}, r_{ms}|\phi_m) d\phi_m = \frac{1}{2\pi\sigma^2} e^{-(r_{mc}^2 + r_{ms}^2 + \mathcal{E}_s)/2\sigma^2} \frac{1}{2\pi} \int_0^{2\pi} e^{\sqrt{\mathcal{E}_s}(r_{mc} \cos \phi_m + r_{ms} \sin \phi_m)/\sigma^2} d\phi_m$$

pero

$$\frac{1}{2\pi} \int_0^{2\pi} e^{\sqrt{\mathcal{E}_s}(r_{mc} \cos \phi_m + r_{ms} \sin \phi_m)/\sigma^2} d\phi_m = I_0 \left(\frac{\sqrt{\mathcal{E}_s}(r_{mc}^2 + r_{ms}^2)}{\sigma^2} \right)$$

donde $I_0(x)$ es la función de Bessel modificada de orden cero. Esta es una función monotonamente creciente de su argumento, como ilustrado en la figura 7.32. $I_0(x)$ tiene la siguiente expansión en series de potencias

Figura 7.32: Gráfico de $I_0(x)$.

$$I_0(x) = \sum_{k=0}^{\infty} \frac{x^{2k}}{2^{2k}(k!)^2}$$

A partir de las ecuaciones anteriores es posible obtener la relación de verosimilitud $\Lambda(\mathbf{r})$ en la forma

$$\Lambda(\mathbf{r}) = \frac{I_0\left(\frac{\sqrt{\mathcal{E}_s(r_{1c}^2 + r_{1s}^2)}}{\sigma^2}\right)}{I_0\left(\frac{\sqrt{\mathcal{E}_s(r_{2c}^2 + r_{2s}^2)}}{\sigma^2}\right)} \begin{matrix} > & \frac{s_1}{f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_2)} \\ < & \frac{s_2}{f_{\mathbf{r}}(\mathbf{r}|\mathbf{s}_1)} \end{matrix} \quad (7.53)$$

De esta forma, el detector óptimo calcula las dos envolventes $r_1 = \sqrt{r_{1c}^2 + r_{1s}^2}$ y $r_2 = \sqrt{r_{2c}^2 + r_{2s}^2}$ y los valores correspondientes de la función de Bessel $I_0\left(\frac{\sqrt{\mathcal{E}_s r_1^2}}{\sigma^2}\right)$ y $I_0\left(\frac{\sqrt{\mathcal{E}_s r_2^2}}{\sigma^2}\right)$ para formar la relación de verosimilitud. Es posible observar que este cálculo requiere el conocimiento de la varianza del ruido σ^2 y la energía de la señal \mathcal{E}_s . La relación de verosimilitud se compara luego con el umbral $P(\mathbf{s}_2)/P(\mathbf{s}_1)$ para determinar cual señal fué transmitida.

Una simplificación significativa en la implementación del detector óptimo ocurre cuando las dos señales son equiprobables. En tal caso el umbral es la unidad y debido a la monotonicidad de la función de Bessel, la regla del detector óptimo se reduce a

$$\sqrt{r_{1c}^2 + r_{1s}^2} \begin{matrix} > & \frac{s_1}{s_2} \\ < & \sqrt{r_{2c}^2 + r_{2s}^2} \end{matrix}$$

De esta forma, el detector óptimo basa su decisión en las dos envolventes $r_1 = \sqrt{r_{1c}^2 + r_{1s}^2}$ y $r_2 = \sqrt{r_{2c}^2 + r_{2s}^2}$ y de esta forma se lo denomina *detector de envolvente*.

Es posible observar que el cálculo de las envolventes de las muestras de señal recibida a la salida del demodulador hace que las fases de las portadoras $\{\phi_m\}$ sean irrelevantes en la decisión de cual señal fué transmitida. Equivalentemente, la decisión puede basarse en el cálculo de las envolventes al cuadrado r_1^2 y r_2^2 , en cuyo caso el detector se denomina *detector de ley cuadrática*. La figura 7.33 muestra un diagrama en bloques del demodulador y el detector de ley cuadrática.

La generalización del demodulador y detector óptimos para señales FSK M -árias ortogonales es directa. Como ilustrado en la figura 7.31, la salida del demodulador óptimo en el instante de muestreo consiste en

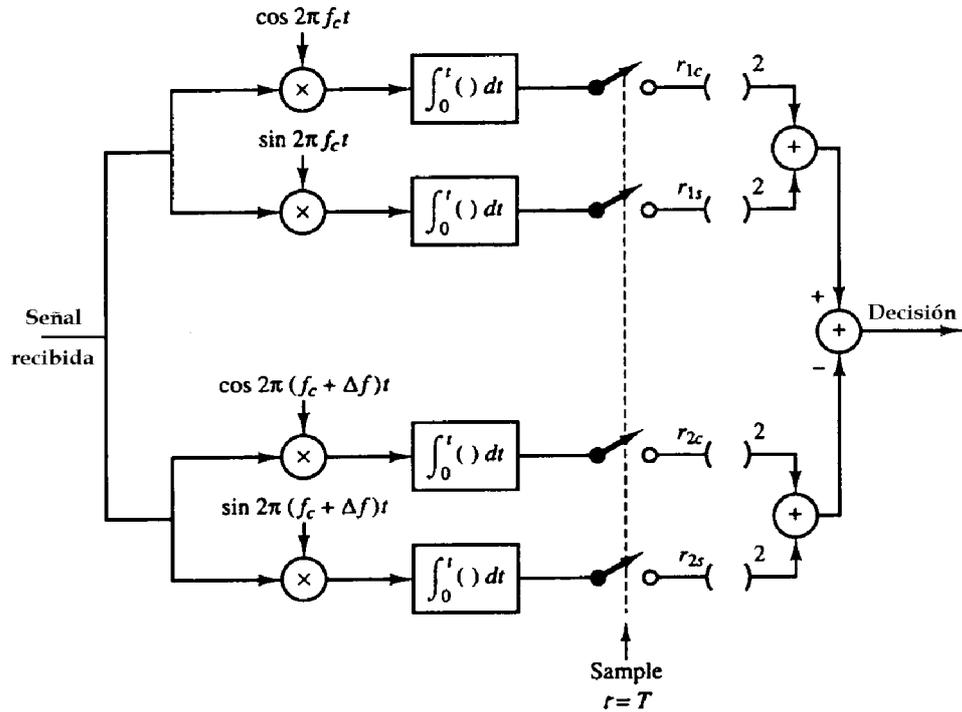


Figura 7.33: Demodulación y detección de ley cuadrática de señales FSK binarias.

un vector de $2M$ componentes $(r_{1c}, r_{1s}, r_{2c}, r_{2s}, \dots, r_{Mc}, r_{Ms})$. Entonces el detector óptimo no coherente calcula las M envolventes como

$$r_m = \sqrt{r_{mc}^2 + r_{ms}^2}, \quad m = 1, 2, \dots, M$$

De esta forma, las fases de portadora de las señales recibidas se vuelven irrelevantes para la decisión sobre cual señal fué transmitida. Cuando todas las señales son igualmente probables de ser transmitidas, el detector óptimo selecciona la señal correspondiente a la mayor envolvente (o envolvente al cuadrado). En el caso de señales transmitidas no igualmente probables, el detector óptimo debe calcular las M probabilidades a posteriori en (7.49) y luego seleccionar la señal correspondiente a la mayor probabilidad a posteriori.

7.4.2 Probabilidad de error para detección de FSK no coherente

Es posible considerar ahora señales FSK M -ária ortogonales que son detectadas no coherentemente, con la suposición que las M señales son igualmente probables a priori y que $s_1(t)$ fué transmitida en el intervalo $0 \leq t \leq T$.

Las M métricas de decisión en el detector son las M envolventes

$$r_m = \sqrt{r_{mc}^2 + r_{ms}^2}, \quad m = 1, 2, \dots, M$$

donde

$$r_{1c} = \sqrt{\mathcal{E}_s} \cos \phi_m + n_{1c} \quad r_{1s} = \sqrt{\mathcal{E}_s} \sin \phi_m + n_{1s}$$

y

$$r_{mc} = n_{mc}, \quad r_{ms} = n_{ms}, \quad m = 2, 3, \dots, M$$

Las componentes de ruido aditivo $\{n_{mc}\}$ y $\{n_{ms}\}$ son variables aleatorias mutuamente independientes Gaussianas de media cero y varianza $\sigma^2 = N_0/2$. De esta forma las funciones densidad de probabilidad de las variables aleatorias a la entrada del detector serán

$$f_{\mathbf{r}_1}(r_{1c}, r_{1s}) = \frac{1}{2\pi\sigma^2} e^{-(r_{1c}^2 + r_{1s}^2 + \mathcal{E}_s)/2\sigma^2} I_0 \left(\frac{\sqrt{\mathcal{E}_s(r_{1c}^2 + r_{1s}^2)}}{\sigma^2} \right) \quad (7.54)$$

$$f_{\mathbf{r}_m}(r_{mc}, r_{ms}) = \frac{1}{2\pi\sigma^2} e^{-[r_{mc}^2 + r_{ms}^2]/2\sigma^2}, \quad m = 2, 3, \dots, M \quad (7.55)$$

Es posible considerar el siguiente cambio de variables en estas ecuaciones

$$R_m = \frac{\sqrt{r_{mc}^2 + r_{ms}^2}}{\sigma} \quad \Theta_m = \tan^{-1} \frac{r_{ms}}{r_{mc}}$$

Claramente, $r_{mc} = \sigma R_m \cos \Theta_m$ y $r_{ms} = \sigma R_m \sin \Theta_m$. El Jacobiano de esta transformación será

$$|\mathbf{J}| = \begin{vmatrix} \sigma \cos \Theta_m & \sigma \sin \Theta_m \\ -\sigma R_m \sin \Theta_m & \sigma R_m \cos \Theta_m \end{vmatrix} = \sigma^2 R_m$$

En consecuencia

$$f_{R_1\Theta_1}(R_1, \Theta_1) = \frac{R_1}{2\pi} e^{-(R_1^2 + 2\mathcal{E}_s/N_0)/2} I_0 \left(\sqrt{\frac{2\mathcal{E}_s}{N_0}} R_1 \right)$$

$$f_{R_m\Theta_m}(R_m, \Theta_m) = \frac{R_m}{2\pi} e^{-R_m^2/2}, \quad m = 2, 3, \dots, M$$

Finalmente, promediando $f_{R_m\Theta_m}(R_m, \Theta_m)$ sobre Θ_m desaparece el factor 2π en estas últimas ecuaciones. De esta forma, es posible concluir que R_1 tiene una densidad de probabilidad Riciana y R_m , $m = 2, 3, \dots, M$ cada una tiene densidad de probabilidad de Rayleigh.

La probabilidad de decisión correcta es simplemente la probabilidad que $R_1 > R_2$ y $R_1 > R_3, \dots$ y $R_1 > R_M$. En consecuencia

$$P_c = P(R_1 > R_2, R_1 > R_3, \dots, R_1 > R_M)$$

$$= \int_0^\infty P(R_2 < R_1, R_3 < R_1, \dots, R_M < R_1 | R_1 = x) f_{R_1}(x) dx$$

Como las variables aleatorias R_m , $m = 2, 3, \dots, M$ son estadísticamente i.i.d. (independientes e igualmente distribuidas), la probabilidad conjunta en la ecuación anterior condicionada sobre R_1 se factoriza en un producto de $M - 1$ términos idénticos. O sea,

$$P_c = \int_0^\infty [P(R_2 < R_1 | R_1 = x)]^{M-1} f_{R_1}(x) dx \quad (7.56)$$

donde $P(R_2 < R_1 | R_1 = x) = \int_0^x f_{R_2}(r_2) dr_2 = 1 - e^{-x^2/2}$, tal que su $M - 1$ -ésima potencia puede expresarse como

$$\left[1 - e^{-x^2/2}\right]^{M-1} = \sum_{n=0}^{M-1} (-1)^n \binom{M-1}{n} e^{-nx^2/2}$$

tal que sustituyendo en (7.56) e integrando sobre x , produce una probabilidad de decisión correcta dada por

$$P_c = \sum_{n=0}^{M-1} (-1)^{n+1} \binom{M-1}{n} \frac{1}{n+1} e^{-n\rho_s^2/(n+1)}$$

donde $\rho_s = \mathcal{E}_s/N_0$ es la SNR por símbolo. Luego, la probabilidad de un error de símbolo ($P_M = 1 - P_c$) será

$$P_M = \sum_{n=1}^{M-1} (-1)^{n+1} \binom{M-1}{n} \frac{1}{n+1} e^{-nk\rho_b^2/(n+1)} \quad (7.57)$$

donde $\rho_b = \mathcal{E}_b/N_0$ es la SNR por bit.

Para FSK binaria ($M = 2$), (7.57) se reduce a

$$P_2 = \frac{1}{2} e^{-\rho_s/(n+1)}$$

Es posible observar que el desempeño de FSK no coherente es 3 dB peor que DPSK binaria.

Para $M > 2$ es posible calcular la probabilidad de un bit de error haciendo uso de la relación

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

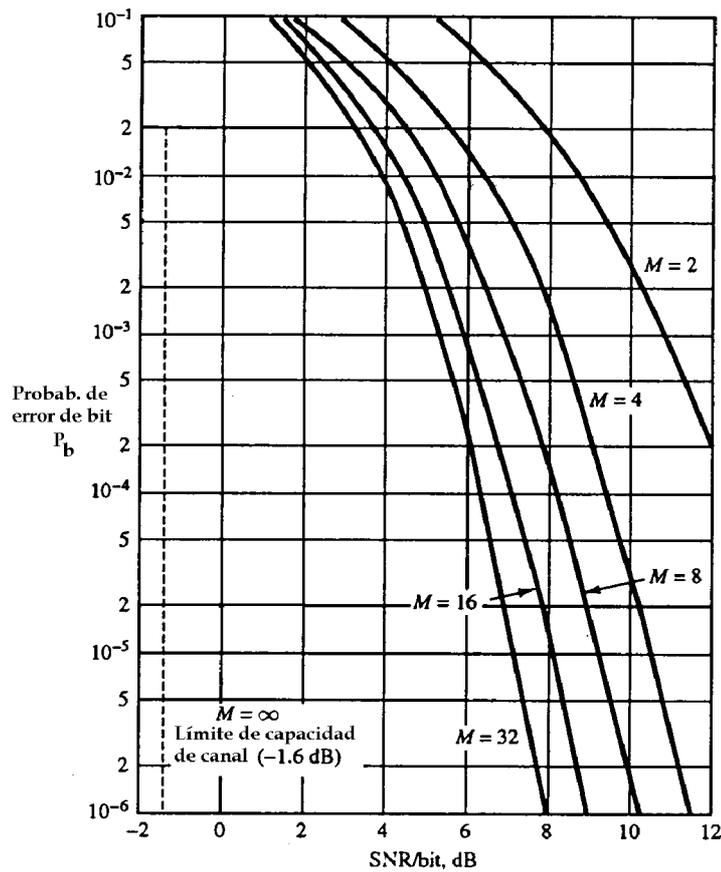


Figura 7.34: Probabilidad de error de bit para detección no coherente de señales FSK ortogonales.

la cual se obtuvo en el Capítulo anterior. La figura 7.34 muestra la probabilidad de error de bit en función de la SNR por bit ρ_b para $M = 2, 4, 8, 16$ y 32 . De la misma forma que en el caso de detección coherente de señales ortogonales M -árias es posible observar que para cualquier probabilidad de error especificada la SNR por bit disminuye cuando M aumenta. En el Capítulo siguiente se mostrará que, en el límite para $M \rightarrow \infty$ (ó $k = \log_2 M \rightarrow \infty$), la probabilidad de error de bit P_b puede hacerse arbitrariamente pequeña previendo que la SNR por bit sea mayor que el límite de Shannon de -1.6 dB. El costo por incrementar M es el ancho de banda requerido para transmitir las señales. Como la separación entre frecuencias adyacentes es

$\Delta f = 1/T$ para señales ortogonales, el ancho de banda requerido para M señales será $W = M\Delta f = M/T$. También, la velocidad de transmisión es $R_b = k/T$, donde $k = \log_2 M$. Entonces, la relación velocidad de transmisión/ancho de banda será

$$\frac{R_b}{W} = \frac{\log_2 M}{M}$$

De esta forma se puede concluir que $R_b/W \rightarrow 0$ para $M \rightarrow \infty$.

7.4.3 FSK de fase continua (CPFSK)

Normalmente las señales FSK descritas anteriormente pueden generarse a partir de $M = 2^k$ osciladores separados, sintonizados en las frecuencias deseadas $f_c + m\Delta f \equiv f_m$, y seleccionando una de las M frecuencias de acuerdo a un símbolo de k bits particular que va a ser transmitido en el intervalo de señal. Sin embargo, el cambio abrupto de una salida de un oscilador a otro en intervalos de señalización sucesivos resulta en un espectro de señal de lóbulos laterales significativos con respecto a la banda central y que decaen lentamente con la separación de frecuencias. En consecuencia, este método es poco eficiente desde el punto de vista del ancho de banda.

Para evitar la utilización de señales que tienen lóbulos laterales grandes es posible usar la señal que transporta la información para modular en fase una portadora simple cuya fase cambie de forma continua. La señal de frecuencia modulada resultante tiene fase continua y por ello se denomina *FSK de fase continua* (CPFSK).

Para representar una señal FSK es posible comenzar con una señal PAM de la forma

$$v(t) = \sum_n a_n g_T(t - nT)$$

donde las amplitudes se obtienen mapeando bloques de k bits de dígitos binarios de la secuencia de información en niveles de amplitud $\pm 1, \pm 3, \dots, \pm(M-1)$, y $g_T(t)$ es un pulso rectangular de amplitud $1/2T$ y duración T . La señal $v(t)$ se utiliza para modular en frecuencia la portadora. En consecuencia, la señal de frecuencia modulada será

$$u(t) = \sqrt{\frac{2\mathcal{E}_s}{T}} \cos[2\pi f_c t + 4\pi T f_d \int_{-\infty}^t v(\tau) d\tau + \phi_0]$$

donde f_d es la *desviación de frecuencia pico* y ϕ_0 es una fase inicial arbitraria asociada a la portadora. Notar que la frecuencia instantánea de la portadora es $f_c + 2T f_d v(t)$.

Observar que, a pesar que $v(t)$ contiene discontinuidades, la integral de $v(t)$ es continua. Es posible escribir la fase de la portadora como

$$\theta(t; \mathbf{a}) = 4\pi T f_d \int_{-\infty}^t v(\tau) d\tau \quad (7.58)$$

donde \mathbf{a} es la secuencia de amplitudes de señal. Como $\theta(t; \mathbf{a})$ es una función continua de t , se tiene entonces una señal continua.

La fase de la portadora en el intervalo $nT \leq t \leq (n+1)T$ se determina integrando (7.58). De esta forma

$$\theta(t; \mathbf{a}) = 2\pi T f_d \sum_{k=-\infty}^{n-1} a_k + 2\pi(t - nT) f_d a_n = \theta_n + 2\pi h a_n q(t - nT) \quad (7.59)$$

donde h , θ_n y $q(t)$ se definen como

$$h = 2f_d T$$

$$\theta_n = \pi h \sum_{k=-\infty}^{n-1} a_k$$

$$q(t) = \begin{cases} 0, & t < 0 \\ t/2T, & 0 \leq t \leq T \\ 1/2, & t > T \end{cases}$$

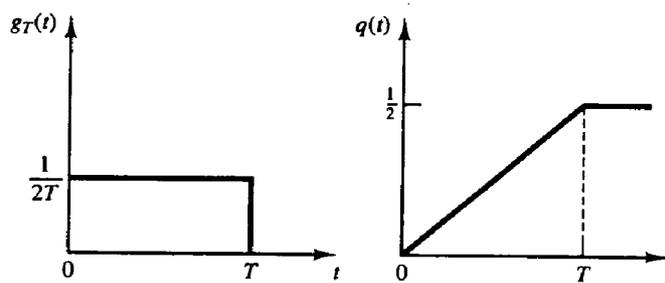


Figura 7.35: La señal pulso $g_T(t)$ y su integral $q(t)$.

El parámetro h se denomina el *índice de modulación*. Es posible observar que θ_n representa la acumulación de fase (memoria) de todos los símbolos hasta el instante $(n - 1)T$. La señal $q(t)$ es simplemente la integral del pulso rectangular, como ilustrado en la figura 7.35.

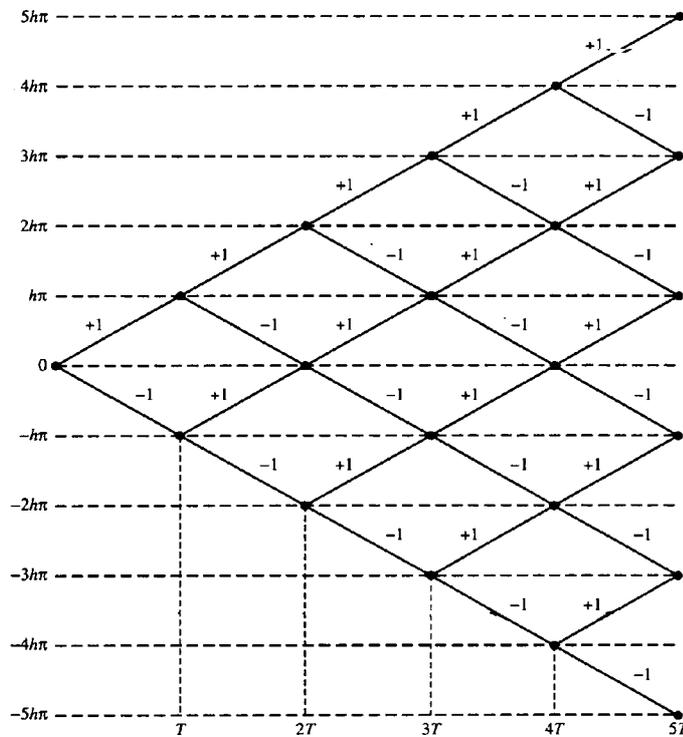


Figura 7.36: Trayectorias de fase para CPFSK binaria.

Es instructivo graficar el conjunto de todas las trayectorias de fase $\theta(t; \mathbf{a})$ generadas por todos los valores posibles de la secuencia de información $\{a_n\}$. Por ejemplo se muestra en la figura 7.36, con símbolos binarios $a_n = \pm 1$, el conjunto de las trayectorias de fase comenzando en el instante $t = 0$. Por comparación, las trayectorias de fase para CPFSK cuaternaria ($a_n = \pm 1, \pm 3$) se ilustran en la figura 7.37. Esos diagramas

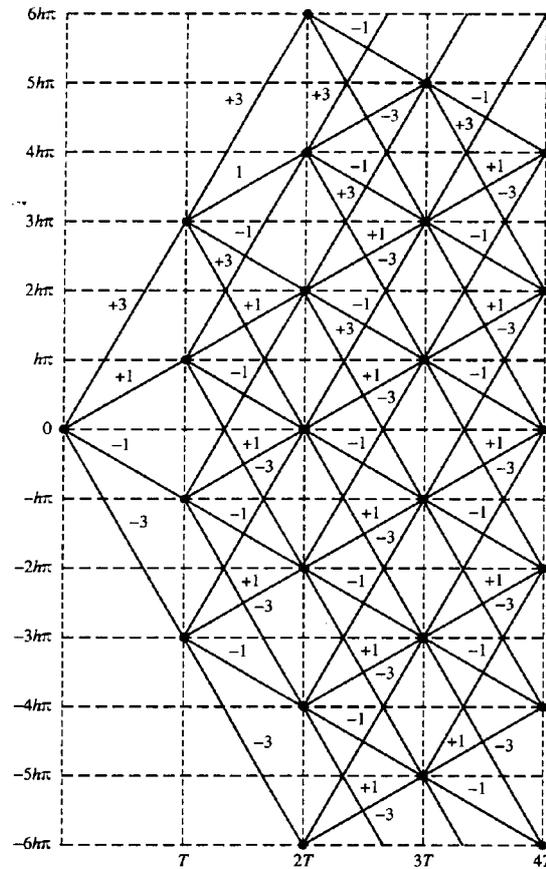


Figura 7.37: Trayectorias de fase para CPFSK cuaternaria.

de fase se denominan *árboles de fase*. Es posible observar que los árboles de fase son lineales a tramos como consecuencia de que el pulso $g_T(t)$ es rectangular. Trayectorias de fase y árboles de fase más suaves pueden obtenerse utilizando pulsos que no contengan discontinuidades.

Los árboles de fase mostrados en esas figuras crecen con el tiempo. Sin embargo, la fase de la portadora es única solo en el rango de $\theta = 0$ a $\theta = 2\pi$ o, equivalentemente, de $\theta = -\pi$ a $\theta = +\pi$. Cuando las trayectorias se grafican módulo 2π , o sea en el rango $(-\pi, \pi)$, el árbol de fases colapsa en una estructura denominada *trellis de fase*. Una forma de ver la trellis de fase en función del tiempo es graficar las dos componentes en cuadratura $x_c(t; \mathbf{a}) = \cos \theta(t; \mathbf{a})$ y $x_s(t; \mathbf{a}) = \sin \theta(t; \mathbf{a})$ como coordenadas (x, y) y el tiempo como tercera dimensión. Como $x^2 + y^2 = 1$ en cualquier instante de tiempo, el gráfico tridimensional generado por $x_c(t; \mathbf{a})$ y $x_s(t; \mathbf{a})$ aparece como una trayectoria sobre la superficie de un cilindro de radio unitario.

Representaciones más simples de las trayectorias de fase pueden obtenerse mostrando solo los valores terminales de la fase de la señal en los instantes de tiempo $t = nT$. En este caso, el índice de modulación h se limitará a un valor racional. En particular, es posible suponer que $h = m/p$, donde m y p son enteros primos entre sí. Luego, en los instantes de tiempo $t = nT$, los estados de fase terminales para m par serán

$$\Theta_s = \left\{ 0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \dots, \frac{(p-1)\pi m}{p} \right\} \tag{7.60}$$

y para m impar serán

$$\Theta_s = \left\{ 0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \dots, \frac{(2p-1)\pi m}{p} \right\} \tag{7.61}$$

De esta forma, existirán p estados de fase terminales cuando m es par y $2p$ estados de fase terminales cuando m es impar. Por ejemplo, CPFSK binaria con $h = 1/2$ tiene cuatro estados de fase terminales. La *trellis*

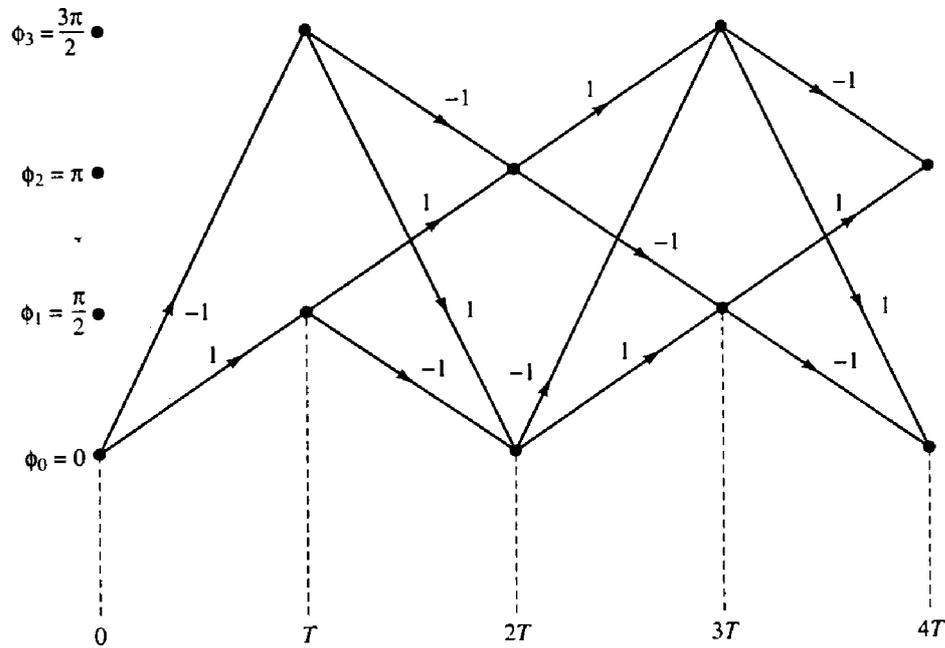


Figura 7.38: Trellis de estados para CPFSK binaria con $h = 1/2$.

de estados para esta señal se ilustra en la figura 7.38. Es necesario enfatizar que las transiciones de fase de un estado a otro no son trayectorias de fase verdaderas. Ellas representan transiciones de fase a los estados terminales en los instantes de tiempo $t = nT$.

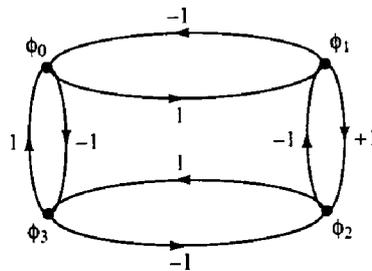


Figura 7.39: Diagrama de estados para CPFSK binaria con $h = 1/2$.

Una representación alternativa a la trellis de estados es el *diagrama de estados*, el cual también ilustra las transiciones de estados en los instantes de tiempo $t = nT$. Esta última es una representación aún más compacta de la señal CPFSK. En el diagrama de estados solamente se representan los estados de fase terminales posibles y sus transiciones. El tiempo no aparece explícitamente como una variable. Por ejemplo, el diagrama de estados para la señal CPFSK con $h = 1/2$ se muestra en la figura 7.39.

Debe enfatizarse que una señal CPFSK no puede representarse mediante puntos discretos en un espacio de señales como en el caso de PAM, PSK o QAM debido a que la fase de la portadora es variante en el tiempo. En cambio, la señal de amplitud constante CPFSK puede representarse en un espacio bidimensional mediante un círculo, donde los puntos sobre el círculo representan la amplitud y trayectoria de fase de la portadora en función del tiempo. Por ejemplo, la figura 7.40 ilustra los diagramas de espacio de señales para CPFSK binaria con $h = 1/2$ y $h = 1/4$. Los cuatro puntos en $\theta = 0, \pi/2, \pi, 3\pi/2$ y $\theta = 0, \pm\pi/4, \pm\pi/2, \pm3\pi/4$ para $h = 1/2$ y $h = 1/4$, respectivamente representan los estados de fase terminales previamente mostrados en el diagrama de estados.

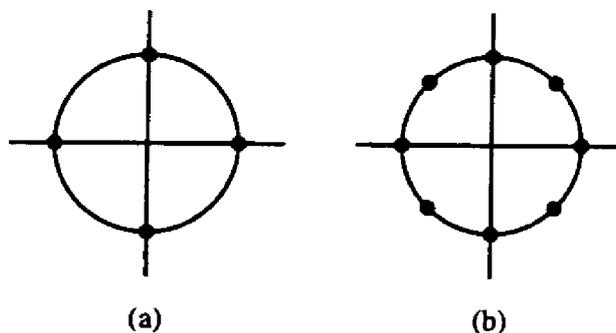


Figura 7.40: Constelaciones para FSK binaria con (a) $h = 1/2$ y (b) $h = 1/4$.

Llaveado de desplazamiento mínimo (MSK)

MSK es una forma especial de CPFSK en la cual el índice de modulación es $h = 1/2$. De esta forma la fase de la portadora para la señal MSE será

$$\begin{aligned}\theta(t; \mathbf{a}) &= \frac{\pi}{2} \sum_{k=-\infty}^{n-1} a_k + \pi a_n q(t - nT) \\ &= \theta_n + \frac{\pi}{2} \left(\frac{t - nT_b}{T_b} \right) a_n, \quad nT_b \leq t \leq (n+1)T_b\end{aligned}$$

la que se obtiene de (7.59). La señal de portadora modulada correspondiente será

$$\begin{aligned}u(t) &= \sqrt{\frac{2\mathcal{E}_b}{T_b}} \cos \left[2\pi f_c t + \theta_n + \frac{\pi}{2} \left(\frac{t - nT_b}{T_b} \right) a_n \right] \\ &= \sqrt{\frac{2\mathcal{E}_b}{T_b}} \cos \left[2\pi \left(f_c + \frac{1}{4T_b} a_n \right) t - \frac{n\pi}{2} a_n + \theta_n \right]\end{aligned}$$

Esta expresión indica que la señal MSK (CPFSK binaria) es básicamente una senoide consistente en una de dos posibles frecuencias en el intervalo $nT_b \leq t \leq (n+1)T_b$, o sea

$$f_1 = f_c - \frac{1}{4T_b} \quad f_2 = f_c + \frac{1}{4T_b}$$

En consecuencia, las dos señales sinusoidales pueden expresarse como

$$u_i(t) = \sqrt{\frac{2\mathcal{E}_b}{T_b}} \cos \left[2\pi f_i t + \theta_n + \frac{n\pi}{2} (-1)^{i-1} \right], \quad i = 1, 2$$

La separación de frecuencia es $\Delta f = f_2 - f_1 = 1/2T_b$. Recordar que esta es la separación de frecuencia mínima que garantiza la ortogonalidad de las dos sinusoidales previendo que las señales se detecten coherentemente. Esto explica porque a una señal binaria CPFSK con $h = 1/2$ se la denomina MSK. Notar que la fase de la portadora en el intervalo de señalización n es el estado de fase de la señal que resulta en la continuidad de la fase entre dos intervalos adyacentes.

Es interesante demostrar que MSK es también una forma de PSK de cuatro fases. Para probar este punto, es posible comenzar con una señal PSK de cuatro fases, la cual tiene la forma

$$u(t) = \sqrt{\frac{2\mathcal{E}_b}{T_b}} \left\{ \left[\sum_{n=-\infty}^{\infty} a_{2n} g_T(t - 2T_b) \right] \cos 2\pi f_c t + \left[\sum_{n=-\infty}^{\infty} a_{2n+1} g_T(t - 2T_b - T_b) \right] \sin 2\pi f_c t \right\}$$

donde $g_T(t)$ es un pulso sinusoidal definido como

$$g_T(t) = \begin{cases} \sin \frac{\pi t}{2T_b}, & 0 \leq t \leq 2T_b \\ 0, & \text{para todo otro } t \end{cases}$$

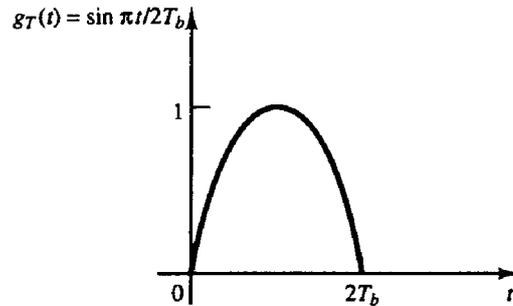


Figura 7.41: Forma de pulso sinusoidal.

y se ilustra en la figura 7.41. Primero, es posible observar que la señal PSK de cuatro fases consiste en dos portadoras en cuadratura, $\cos 2\pi f_c t$ y $\sin 2\pi f_c t$, las cuales están moduladas en amplitud a una velocidad de un bit por intervalo de $2T_b$. Los bits de información pares $\{a_{2n}\}$ se transmiten modulando la portadora coseno, mientras que los bits de información impares $\{a_{2n+1}\}$ se transmiten modulando la portadora seno. Notar que la modulación de las dos portadoras en cuadratura está desplazada en tiempo por T_b y que la velocidad de transmisión para cada portadora es $1/2T_b$. Este tipo de modulación de cuatro fases se denomina *PSK en cuadratura offset* (OQPSK).

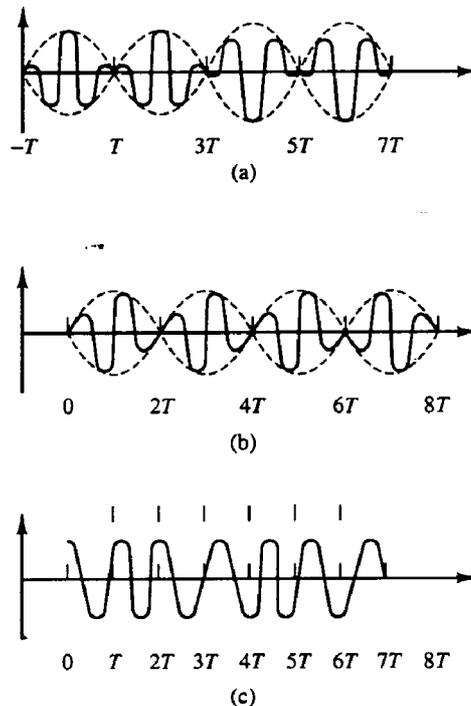


Figura 7.42: Representación de la señal MSK como una forma de dos señales OQPSK, cada una con envoltura sinusoidal. (a) Componente de señal en fase, (b) componente de señal en cuadratura y (c) señal MSK (a+b).

La figura 7.42 ilustra la señal OQPSK en términos de las dos señales PSK binarias moduladas en cuadratura desplazadas. La suma correspondiente de las dos señales en cuadratura es una señal FSK de fase continua de amplitud constante.

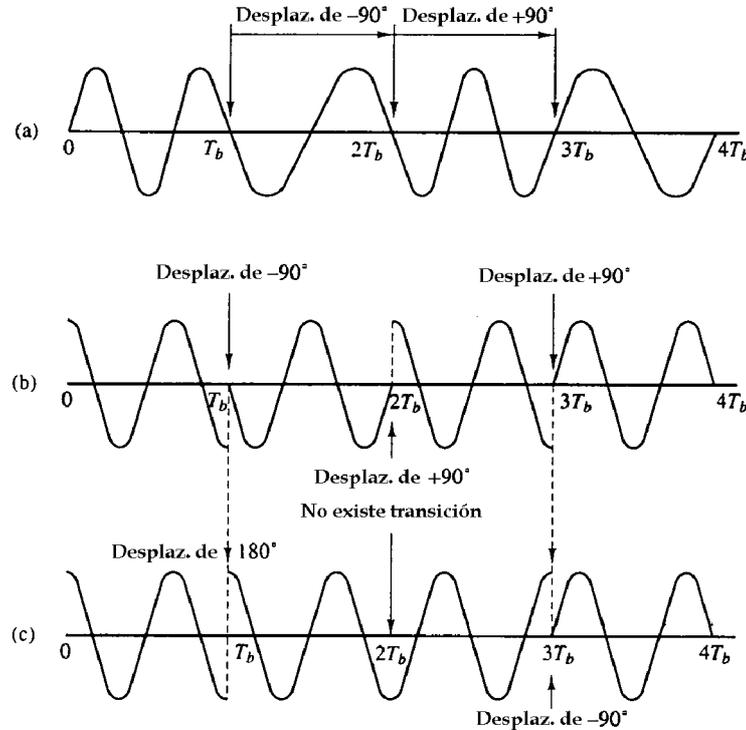


Figura 7.43: Formas de onda para (a) MSK, (b) OQPSK con pulso rectangular y (c) QPSK convencional.

Es interesante también comparar las formas de onda de MSK con las de OQPSK en las cuales el pulso $g_T(t)$ es rectangular para $0 \leq t \leq 2T_b$ y también con las de QPSK convencional, para el mismo pulso banda base. Las tres técnicas de modulación logran la misma velocidad de transmisión. La señal MSK es de fase continua. La señal OQPSK con un pulso banda base rectangular es básicamente dos señales PSK binarias para las cuales las transiciones de fase están desplazadas en tiempo en T_b seg. En consecuencia, esta señal contiene saltos de fase en $\pm 90^\circ$, y que puede ocurrir tan frecuentemente como cada T_b seg. Por otro lado, en QPSK convencional con envolvente constante, uno o ambos símbolos de información pueden causar transiciones de fase cada $2T_b$ seg. Esos saltos de fase pueden ser de $\pm 180^\circ$ ó $\pm 90^\circ$. Una ilustración de los tres tipos de señales PSK de cuatro fases se muestra en la figura 7.43.

A continuación se describirán las características espectrales de señales CPFSK y se las comparará con otras señales QPSK. La demodulación y detección de CPFSK se describirá más adelante cuando se discuta la demodulación de una clase más general de señales de fase continua.

7.4.4 Características espectrales de CPFSK

En esta sección se considerarán las características espectrales de señales CPFSK y se discutirán algunos resultados relacionados con su densidad espectral de potencia.

Una señal CPFSK puede expresarse en la siguiente forma general

$$\begin{aligned}
 u(t; \mathbf{a}) &= \sqrt{\frac{2\mathcal{E}_s}{T}} \cos[2\pi f_c t + \theta(t; \mathbf{a})] \\
 &= \operatorname{Re} \left\{ \sqrt{\frac{2\mathcal{E}_s}{T}} e^{j[2\pi f_c t + \theta(t; \mathbf{a})]} \right\}
 \end{aligned}$$

donde la fase de la portadora $\theta(t; \mathbf{a})$ está dada por (7.59). Dado que la exponencial compleja $e^{j2\pi f_c t}$ sirve

como una traslación en frecuencia de una señal bandabase compleja $s(t) = \sqrt{\frac{2\mathcal{E}_s}{T}} e^{j\theta(t; \mathbf{a})}$, es suficiente enfocar la atención sobre las características espectrales de la señal que transporta información $s(t)$.

A diferencia del cálculo de la densidad espectral de potencia para los métodos de modulación lineal, tal como PAM, PSK y QAM, para los cuales es relativamente simple, el cálculo de la densidad espectral de potencia de una señal CPFSK es mucho más tedioso y complicado. Las dificultades están asociadas a la memoria de la señal CPFSK y a la relación exponencial entre $s(t)$ y $\theta(t; \mathbf{a})$. Independientemente de eso, puede utilizarse el procedimiento general para la obtención de la densidad espectral de potencia de PAM discutido en el Capítulo anterior.

El procedimiento general comienza con el cálculo de la función autocorrelación del proceso estocástico $s(t)$, $R_s(t + \tau, t)$. Como en el caso de los métodos de modulación lineal, el proceso estocástico $s(t)$ es cicloestacionario y, debido a ello la función autocorrelación es periódica con período T , o sea, $R_s(t + T + \tau, t + T) = R_s(t + \tau, t)$. Promediando $R_s(t + \tau, t)$ en un período es posible eliminar la dependencia de t y de esta forma obtener la función autocorrelación promedio $\bar{R}_s(\tau) = \frac{1}{T} \int_0^T R_s(t + \tau, t) dt$. Finalmente, se calcula la transformada de Fourier de $\bar{R}_s(\tau)$ para obtener la densidad espectral de potencia $S_s(f)$. La densidad espectral de potencia de una señal real $u(t; \mathbf{a})$ se puede obtener trasladando $S_s(f)$ en frecuencia mediante una portadora en f_c . De esta forma se obtiene

$$S_U(f) = \frac{1}{4} [S_s(f - f_c) + S_s(f + f_c)]$$

En el caso de una señal CPFSK este procedimiento produce la siguiente expresión para la densidad espectral de potencia

$$S_s(f) = T \left[\frac{1}{M} \sum_{n=1}^M A_n^2(f) + \frac{2}{M^2} \sum_{n=1}^M \sum_{m=1}^M B_{nm}(f) A_n(f) A_m(f) \right] \quad (7.62)$$

donde

$$\begin{aligned} A_n(f) &= \frac{\sin \pi [fT - (2n - 1 - M)h/2]}{\pi [fT - (2n - 1 - M)h/2]} = \text{sinc} \left(fT - \frac{h(2n - 1 - M)}{2} \right) \\ B_{nm}(f) &= \frac{\cos(2\pi fT - \alpha_{nm}) - \beta \cos \alpha_{nm}}{1 + \beta^2 - 2\beta \cos 2\pi fT} \\ \alpha_{nm} &= \pi h(m + n - 1 - M) \\ \beta &= \frac{\sin M\pi h}{M \sin \pi h} \end{aligned}$$

La densidad espectral de potencia de CPFSK para $M = 2$ se muestra en las figuras 7.44 y 7.45 en función de la frecuencia normalizada fT , con el índice de modulación $h = 2f_d T$ como parámetro. Notar que en esos gráficos se muestra solo la mitad del espectro porque el mismo es simétrico en frecuencia. El origen $fT = 0$ corresponde a la frecuencia de la portadora f_c en el espectro de la señal real.

Esos gráficos muestran que el espectro de una señal CPFSK es relativamente suave y bien confinado para $h < 1$. Cuando h se aproxima a uno, el espectro se vuelve muy abrupto, y para $h = 1$ donde $|\beta| = 1$ ocurren impulsos en M frecuencias. Cuando $h > 1$ el espectro se vuelve mucho más disperso. En sistemas de comunicaciones que utilizan CPFSK el índice de modulación se selecciona para conservar ancho de banda, de forma que $h < 1$.

El caso especial de CPFSK binaria con $h = 1/2$ (ó $f_d = 1/4T_s$) y $\beta = 0$ corresponde a MSK. En este caso, la densidad espectral de potencia obtenida de las ecuación (7.62) será

$$S_s(f) = \frac{32\mathcal{E}_s}{\pi^2} \left[\frac{\cos 2\pi fT}{1 - 16f^2 T^2} \right]^2 \quad (7.63)$$

En contraste, la densidad espectral de potencia de OQPSK con un pulso rectangular $g_T(t)$ de duración $2T_b$ será

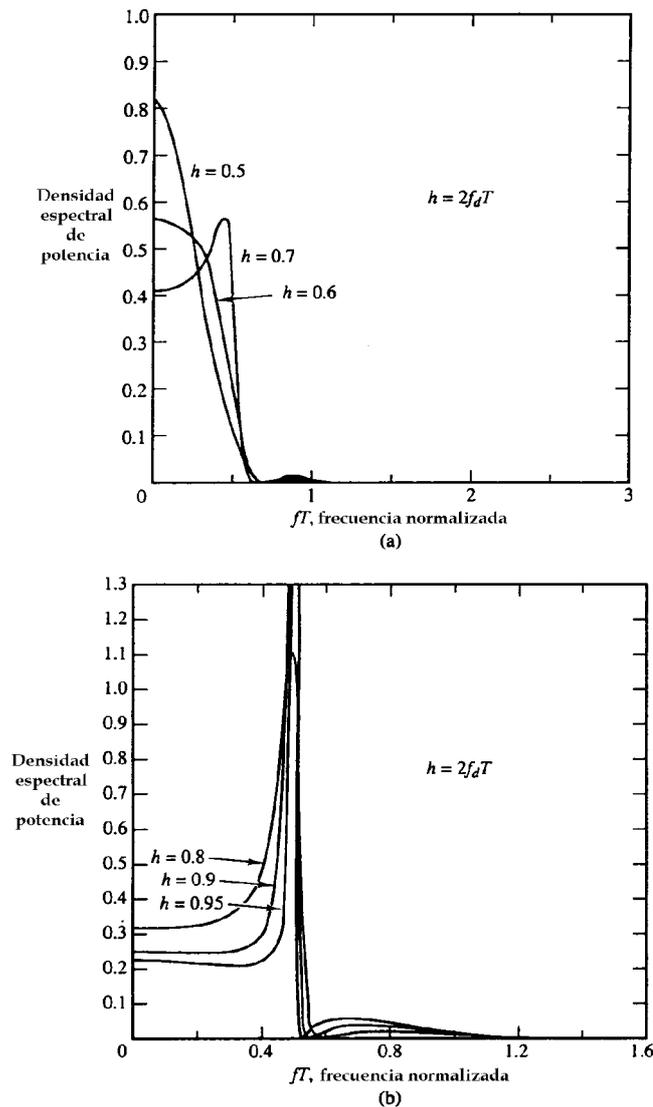


Figura 7.44: Densidad espectral de potencia de CPFSK binaria.

$$S_S(f) = 4\mathcal{E}_s \left[\frac{\sin 2\pi f T_b}{2\pi f T_b} \right]^2 \quad (7.64)$$

Las densidades espectrales de potencia en (7.63) y (7.64) se ilustran en la figura 7.46. Notar que el lóbulo principal de MSK es 50% más ancho que el de OQPSK. Sin embargo, los lóbulos laterales de MSK decrecen considerablemente más rápido. En consecuencia, MSK es significativamente más eficiente en términos de ancho de banda que OQPSK.

7.5 Modulación de fase continua (CPM) (*)

Los métodos de modulación PAM, PSK y QAM descriptos en las secciones anteriores son básicamente métodos de modulación sin memoria. En otras palabras, la amplitud transmitida en PAM, la fase de la portadora transmitida en modulación de fase digital y la combinación de fase y amplitud de la portadora transmitida en QAM son estadísticamente independientes en intervalos sucesivos de señalización, a menos que el canal introduzca memoria (interferencia intersímbolo).

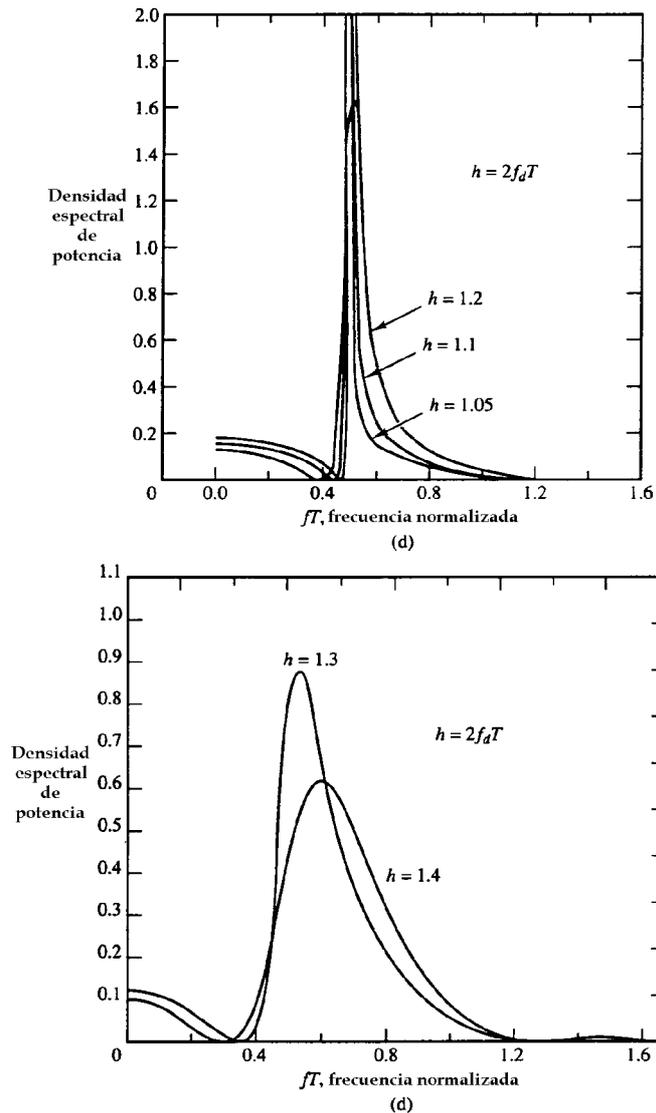


Figura 7.45: Densidad espectral de potencia de CPFSK binaria.

En contraste CPFSK es un método de modulación con memoria. La memoria resulta de la continuidad de fase de la portadora transmitida desde un intervalo de señalización hasta el próximo. Cuando la fase se expresa en la forma (7.59), CPFSK se vuelve un caso especial de una clase general de señales moduladas de fase continua (CPM) en la cual la fase de la portadora está definida por

$$\theta(t; \mathbf{a}) = 2\pi \sum_{k=-\infty}^n a_k h_k q(t - kT), \quad nT \leq t \leq (n+1)T \quad (7.65)$$

donde $\{a_k\}$ es la secuencia M -ária de símbolos de información con valores $\pm 1, \pm 3, \dots, \pm(M-1)$, $\{h_k\}$ es una secuencia de índices de modulación y $q(t)$ es alguna forma de onda normalizada. Recordar que para CPFSK, $q(t) = t/2T$ para $0 \leq t \leq T$, $q(t) = 0$ para $t < 0$ y $q(t) = 1/2$ para $t > T$.

Cuando $h_k = h$ para todo k , el índice de modulación es fijo para todos los símbolos. Cuando el índice de modulación varía de un símbolo a otro, la señal CPM se denomina *multi h*. En tal caso, los $\{h_k\}$ se eligen de forma de variar según un patrón cíclico a través del conjunto de índices.

La forma de onda $q(t)$ es la integral del pulso $g_T(t)$ de forma arbitraria, o sea

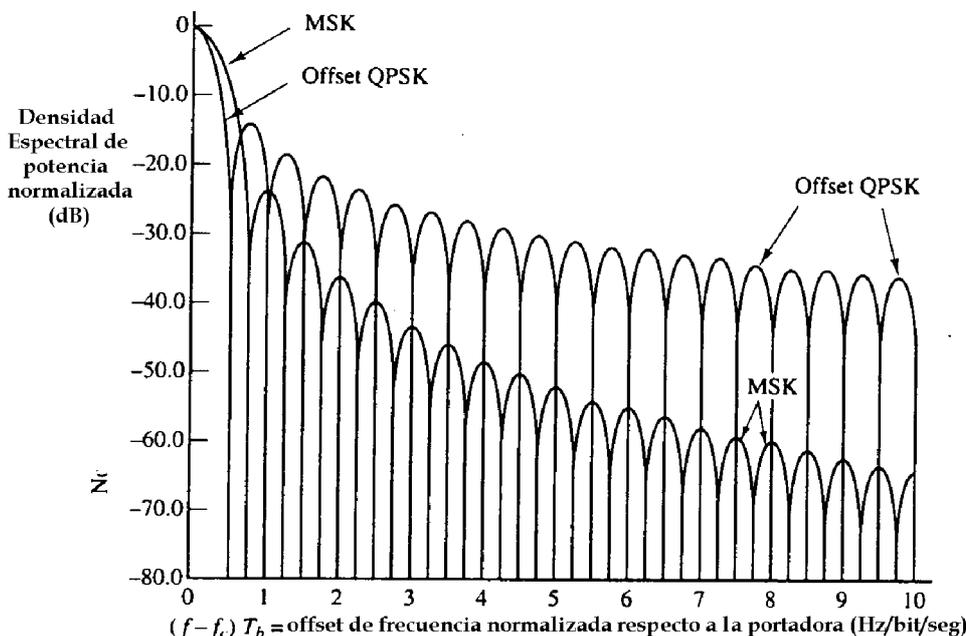


Figura 7.46: Densidad espectral de potencia de MSK y QPSK.

$$q(t) = \int_0^t g_T(\tau) d\tau$$

Si $g_T(t) = 0$ para $t > T$, la señal CPM se denomina una *señal CPM de respuesta completa*. Si $g_T(t) \neq 0$ para $t > T$, la señal CPM se denomina una *señal CPM de respuesta parcial*. En la figura 7.47 se ilustran varias formas de pulso para $g_T(t)$ y la señal $q(t)$ correspondiente. Es evidente que existirá un número infinito de señales CPM que pueden obtenerse seleccionando diferentes formas de pulso para $g_T(t)$ o variando el índice de modulación h o el número de símbolos M .

En general, las señales CPM no pueden representarse mediante puntos discretos en un espacio de señales como en el caso de PAM, PSK y QAM debido a que la fase de la portadora es variante en el tiempo. Como se discutió en el caso de señales CPFSK es posible utilizar trayectorias de fase (o árboles de fase) para ilustrar la fase de señal en función del tiempo. Alternativamente, es posible utilizar una trellis de estados o un diagrama de estados para ilustrar los estados de fase terminales y las transiciones de fase de un estado a otro. Finalmente, como en el caso de CPFSK una señal CPM se puede representar en el espacio de señales mediante un círculo, donde los puntos sobre el círculo corresponden a la combinación de amplitud y fase de la portadora en función del tiempo. Los estados de fase terminales se identifican usualmente como puntos discretos sobre el círculo. Por ejemplo, la figura 7.48 ilustra los diagramas de espacio de señales para CPFSK con $h = 1/4, 1/3$ y $2/3$. Notar que la longitud de las trayectorias de fase (longitudes de los arcos) entre estados de fase terminales aumenta con un incremento en h . Un incremento en h también resulta en un incremento del ancho de banda de la señal, como se discutió anteriormente.

7.5.1 Demodulación y detección de CPM (*)

La señal CPM transmitida puede expresarse como

$$u(t) = \sqrt{\frac{2\mathcal{E}_s}{T}} \cos[2\pi f_c t + \theta(t; \mathbf{a})]$$

donde $\theta(t; \mathbf{a})$ es la fase de la portadora. La señal recibida filtrada por un canal con AWGN es

$$r(t) = u(t) + n(t)$$

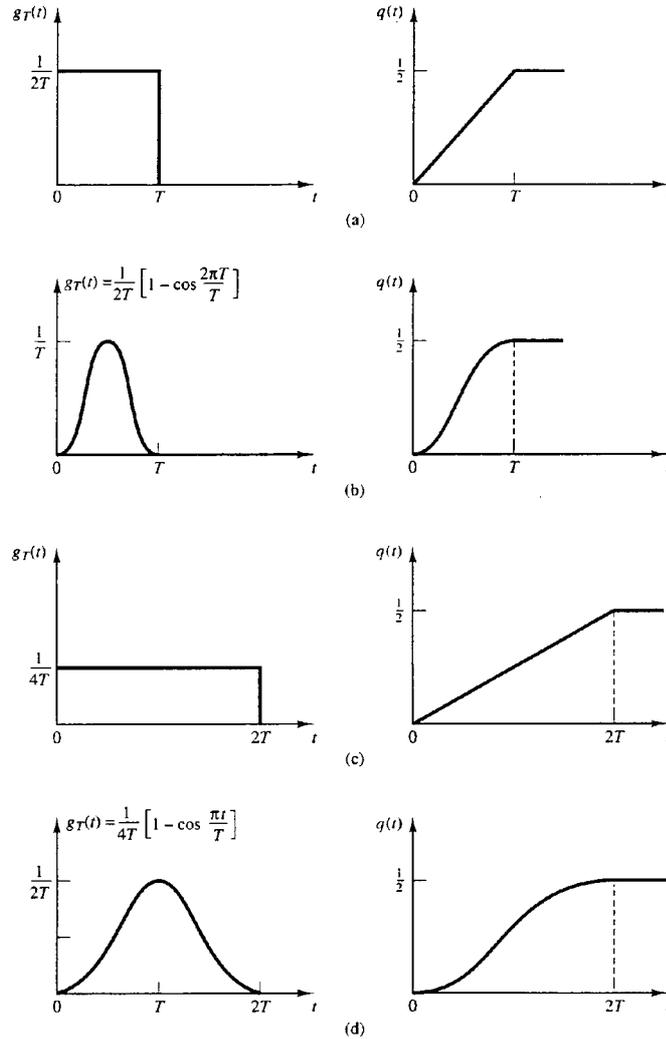


Figura 7.47: Formas de pulso para CPM, (a,b) respuesta completa, (c,d) respuesta parcial.

donde

$$n(t) = n_c(t) \cos 2\pi f_c t - n_s(t) \sin 2\pi f_c t$$

El receptor óptimo para esta señal consiste en un correlador seguido de un detector de secuencias ML que busca los caminos con distancia Euclidiana mínima a través de la trellis de estados. Un método eficiente para realizar esta búsqueda es el algoritmo de Viterbi. Se discutirá los aspectos relacionados con la estructura general de la trellis de estados para CPM y luego se describirá el cálculo de las métricas asociadas.

La fase de la portadora para una señal CPM con índice de modulación fijo puede expresarse como

$$\begin{aligned} \theta(t; \mathbf{a}) &= 2h\pi \sum_{k=-\infty}^n a_k q(t - kT), \quad nT \leq t \leq (n+1)T \\ &= \pi h \sum_{k=-\infty}^{n-L} a_k + 2\pi h \sum_{k=n-L+1}^n a_k q(t - kT) \\ &= \theta_n + \phi(t; \mathbf{a}), \quad nT \leq t \leq (n+1)T \end{aligned}$$

donde por definición

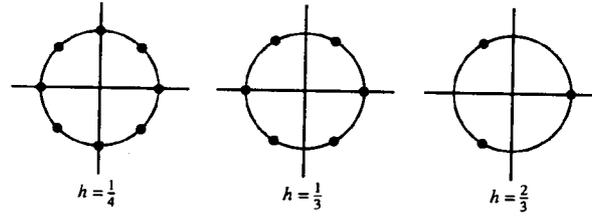


Figura 7.48: Constelaciones para CPFSK.

$$\theta_n = \pi h \sum_{k=-\infty}^{n-L} a_k \quad (7.66)$$

$$\phi(t; \mathbf{a}) = 2\pi h \sum_{k=n-L+1}^n a_k q(t - kT) \quad (7.67)$$

y $q(t) = 0$ para $t < 0$, $q(t) = 1/2$ para $t \geq LT$, y

$$q(t) = \int_0^t g_T(\tau) d\tau$$

donde L es un entero positivo. El pulso de señal será $g_T(t) = 0$ para $t < 0$ y $t \geq LT$. Para $L = 1$ se tiene una señal CPM de respuesta completa, y para $L > 1$ se tiene una señal de respuesta parcial. Notar que θ_n representa la fase de la portadora en $t = nT$ y $\phi(t; \mathbf{a})$ representa la acumulación de fase adicional en $nT \leq t \leq (n+1)T$.

Cuando h es racional, o sea $h = m/p$ donde m y p son enteros primos, las fases de la señal CPM en $t = nT$ pueden representarse mediante una trellis. El número de estados de fase terminales para $L = 1$ es p para m par y $2p$ para m impar, como indicado previamente con (7.60) y (7.61). Sin embargo, para $L > 1$ se tiene estados adicionales de fase debido a la característica de respuesta parcial del pulso de señal $g_T(t)$. Esos estados de fase terminales adicionales pueden determinarse expresando la fase $\phi(t; \mathbf{a})$ en (7.67) como

$$\phi(t; \mathbf{a}) = 2\pi h \sum_{k=n-L+1}^{n-1} a_k q(t - kT) + 2\pi h a_n q(t - nT) \quad (7.68)$$

El primer término del lado derecho de (7.68) es función de los símbolos de información $(a_{n-1}, a_{n-2}, \dots, a_{n-L+1})$, denominado *vector de estados correlativos*, y representa la contribución de fase de los pulsos que no han alcanzado su valor final. El segundo término del lado derecho de (7.68) representa la contribución de fase debido al símbolo más reciente a_n . En consecuencia, el estado de la señal CPM en $t = nT$ puede expresarse como la combinación del estado de fase θ_n y el vector de estados correlativos $(a_{n-1}, a_{n-2}, \dots, a_{n-L+1})$, lo cual se notará por

$$S_n = \{\theta_n, a_{n-1}, a_{n-2}, \dots, a_{n-L+1}\}$$

para el caso de una señal de respuesta parcial de longitud LT . De esta forma el número de estados será

$$N_s = \begin{cases} pM^{L-1} & m \text{ par} \\ 2pM^{L-1} & m \text{ impar} \end{cases}$$

donde $h = m/p$.

La transición de estados de S_n a S_{n-1} puede determinarse por el efecto del nuevo símbolo a_n en el intervalo de tiempo $nT \leq t \leq (n+1)T$. En general se tendrá que

$$S_{n+1} = \{\theta_n, a_n, a_{n-1}, \dots, a_{n-L+2}\}$$

donde

$$\theta_{n+1} = \theta_n + \pi h a_{n-L+1}$$

Ejemplo. Determinar los estados terminales para una señal binaria con índice de modulación $h = 3/4$ y un pulso de respuesta parcial de $L = 2$. Graficar la trellis de estados.

Como $p = 4$, m es impar y $M = 2$ se tiene que $N_s = 16$ estados de fase. Los $2p$ estados de fase correspondientes a θ_n serán

$$\Theta_s = \left\{ 0, \pm \frac{\pi}{4}, \pm \frac{\pi}{2}, \pm \frac{3\pi}{4}, \pi \right\}$$

Por cada uno de esos estados de fase existirán dos estados asociados a la memoria de la señal de respuesta parcial. De esta forma, los 16 estados $S_n = (\theta_n, a_{n-1})$ serán

$$\begin{aligned} &(0, +1), (0, -1), (\pi/4, 1), (\pi/4, -1), (-\pi/4, 1), (-\pi/4, -1), \\ &(\pi/2, 1), (\pi/2, -1), (-\pi/2, 1), (-\pi/2, -1), (3\pi/4, 1), (3\pi/4, -1), \\ &(-3\pi/4, 1), (-3\pi/4, -1), (\pi, 1), (\pi, -1) \end{aligned}$$

Suponiendo que el sistema está en el estado de fase $\theta_n = 3\pi/4$ y $a_{n-1} = -1$. Entonces $S_n = (3\pi/4, -1)$ y

$$\theta_{n+1} = \theta_n + \pi h a_{n-1} = 3\pi/4 + (3\pi/4)(-1) = 0$$

En consecuencia, $S_{n+1} = (\theta_{n+1}, a_n) = (0, a_n)$. Si $a_n = 1$, entonces $S_{n+1} = (0, 1)$. Si $a_n = -1$, entonces $S_{n+1} = (0, -1)$. La trellis de estados se ilustra en la figura 7.49. En la figura 7.50 se ilustra un camino a través de la trellis de estados correspondiente a la secuencia de datos $(1, -1, -1, -1, 1, 1)$.

Se discutirá ahora aspectos relacionados con la demodulación y detección de señales CPM. Como todas las formas de onda tienen la misma energía es posible basar las decisiones del detector en métricas de correlación. La correlación cruzada entre la señal recibida $r(t)$ y cada una de las posibles señales transmitidas produce la siguiente métrica

$$\begin{aligned} C_n(\mathbf{a}, \mathbf{r}) &= \int_{-\infty}^{(n+1)T} r(t) \cos[2\pi f_c t + \theta(t; \mathbf{a})] dt \\ &= C_{n-1}(\mathbf{a}, \mathbf{r}) + \int_{nT}^{(n+1)T} r(t) \cos[2\pi f_c t + \theta_n + \phi(t; \mathbf{a})] dt \end{aligned}$$

En el detector de secuencias ML, el término $C_{n-1}(\mathbf{a}, \mathbf{r})$ representa la métrica de las secuencias supervivientes hasta el instante nT . El segundo término del lado derecho de la ecuación anterior, $C_n(\theta_n; \mathbf{a}, \mathbf{r})$, representa los incrementos adicionales en la métrica de correlación debidos a la señal recibida en el intervalo $nT \leq t \leq (n+1)T$. Es posible observar que existen M^L secuencias posibles $\mathbf{a} = (a_n, a_{n-1}, \dots, a_{n-L+1})$ de símbolos y p o $2p$ estados de fase posibles $\{\theta_n\}$. En consecuencia, existirán pM^L o $2pM^L$ valores diferentes de la métrica de correlación $C_n(\theta_n; \mathbf{a}, \mathbf{r})$ a ser calculados en cada intervalo, de forma que cada valor se utiliza para incrementar la métrica correspondiente a las secuencias supervivientes de los intervalos de señalización previos. Un diagrama en bloques general que ilustra el procesamiento de $C_n(\theta_n; \mathbf{a}, \mathbf{r})$ para el detector de Viterbi se muestra en la figura 7.51.

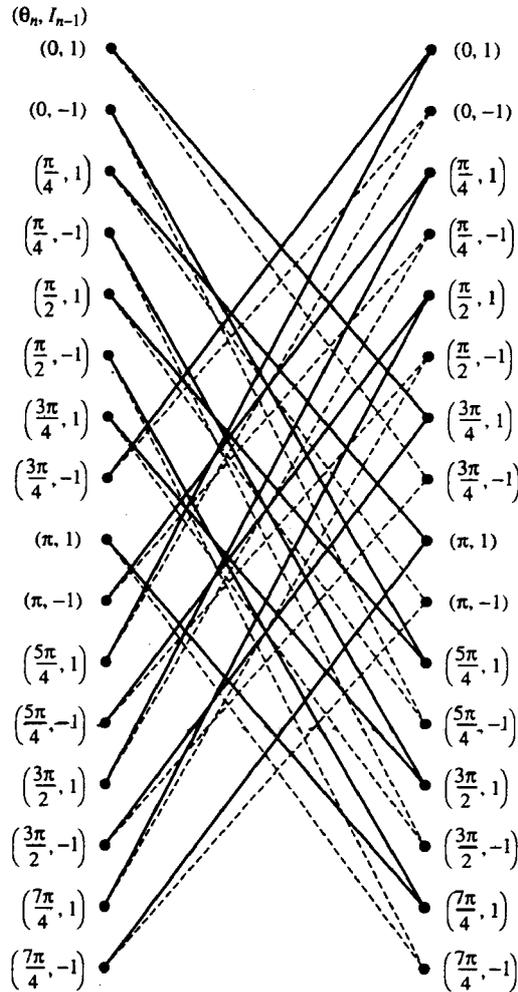


Figura 7.49: Trellis de estados para CPM de respuesta parcial ($L = 2$) con $h = 3/4$.

El número de secuencias supervivientes en cada estado del proceso de búsqueda de Viterbi en el detector de secuencias ML es pM^{L-1} o $2pM^{L-1}$. Por cada secuencia superviviente se tienen M nuevos incrementos de métrica de correlación $C_n(\theta_n; \mathbf{a}, \mathbf{r})$ que se suman a las métricas existentes para producir pM^L o $2pM^L$ secuencias con sus correspondientes métricas. Sin embargo, este número se reduce a pM^{L-1} o $2pM^{L-1}$ supervivientes con sus correspondientes métricas seleccionando la más probable de las M secuencias convergiendo a cada nodo de la trellis y descartando las otras $M - 1$ secuencias.

7.5.2 Desempeño de CPM en un canal AWGN (*)

Para la determinación de la probabilidad de error asociada al detector de Viterbi para señales CPM es necesario determinar la distancia Euclidiana mínima de caminos a través de la trellis que se separan en un nodo y vuelven a convergir posteriormente en el mismo nodo. La distancia entre dos caminos a través de la trellis dependerá de la diferencia de fase entre los caminos, como se discutirá a continuación.

Es posible suponer dos señales $u_i(t)$ y $u_j(t)$ correspondientes a las trayectorias de fase $\theta(t; \mathbf{a}_i)$ y $\theta(t; \mathbf{a}_j)$. Las secuencias \mathbf{a}_i y \mathbf{a}_j deben diferir en el primer símbolo. Luego, la distancia Euclidiana al cuadrado entre las dos señales en un intervalo de longitud NT , donde $1/T$ es la velocidad de transmisión de símbolos, se define como

$$d_{ij}^2 = \int_0^{NT} [u_i(t) - u_j(t)]^2 dt = \int_0^{NT} u_i(t)^2 dt + \int_0^{NT} u_j(t)^2 dt - 2 \int_0^{NT} u_i(t)u_j(t) dt$$

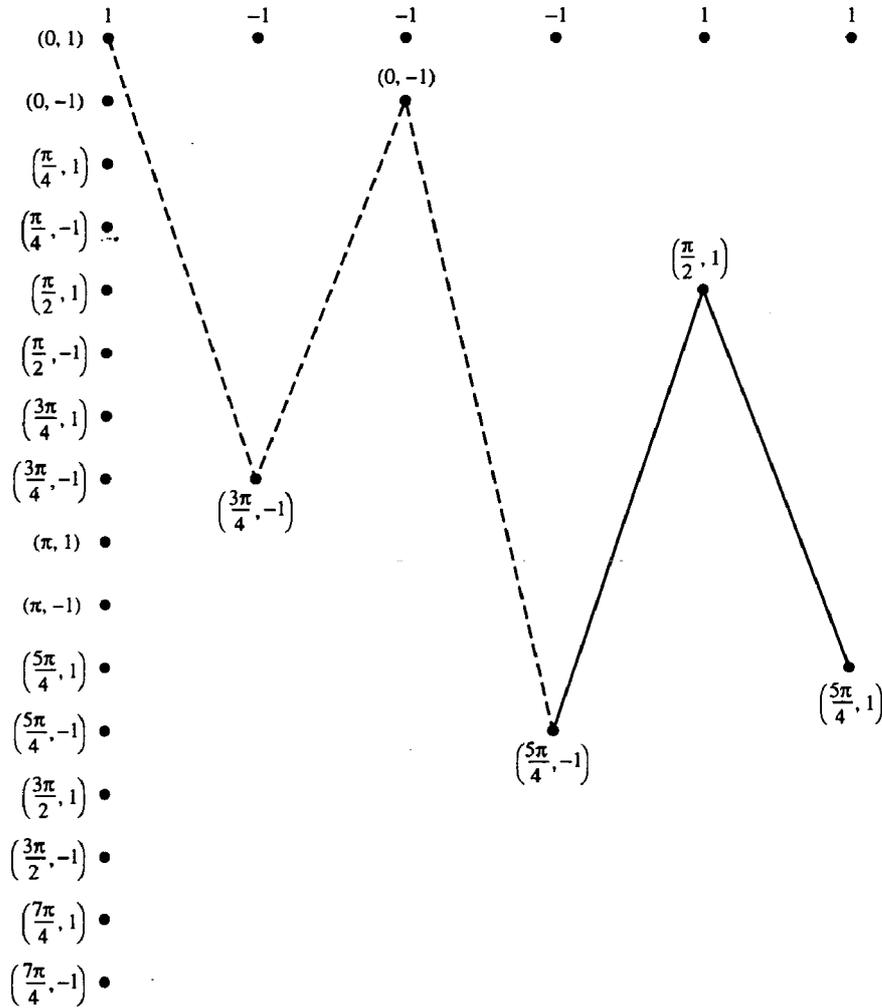


Figura 7.50: Un camino de señal a través de la trellis.

$$\begin{aligned}
 &= 2N\mathcal{E}_s - 2 \left(\frac{2\mathcal{E}_s}{T} \right) \int_0^{NT} \cos[2\pi f_c t + \theta(t; \mathbf{a}_i)] \cos[2\pi f_c t + \theta(t; \mathbf{a}_j)] dt \\
 &= \left(\frac{2\mathcal{E}_s}{T} \right) \int_0^{NT} \{1 - \cos[\theta(t; \mathbf{a}_i) - \theta(t; \mathbf{a}_j)]\} dt
 \end{aligned}$$

Como la energía por símbolo está relacionada con la energía por bit $\mathcal{E}_s = \mathcal{E}_b \log_2 M$, la distancia Euclidiana al cuadrado d_{ij}^2 puede expresarse también como

$$d_{ij}^2 = \left(\frac{2\mathcal{E}_b \log_2 M}{T} \right) \int_0^{NT} \{1 - \cos[\theta(t; \mathbf{a}_i) - \theta(t; \mathbf{a}_j)]\} dt$$

La probabilidad de error para el detector de Viterbi está dominada por el término correspondiente a la distancia Euclidiana mínima, definida como

$$d_{min}^2 = \lim_{N \rightarrow \infty} \min_{i,j} [d_{i,j}^2]$$

Luego, la probabilidad de error de símbolo puede aproximarse como

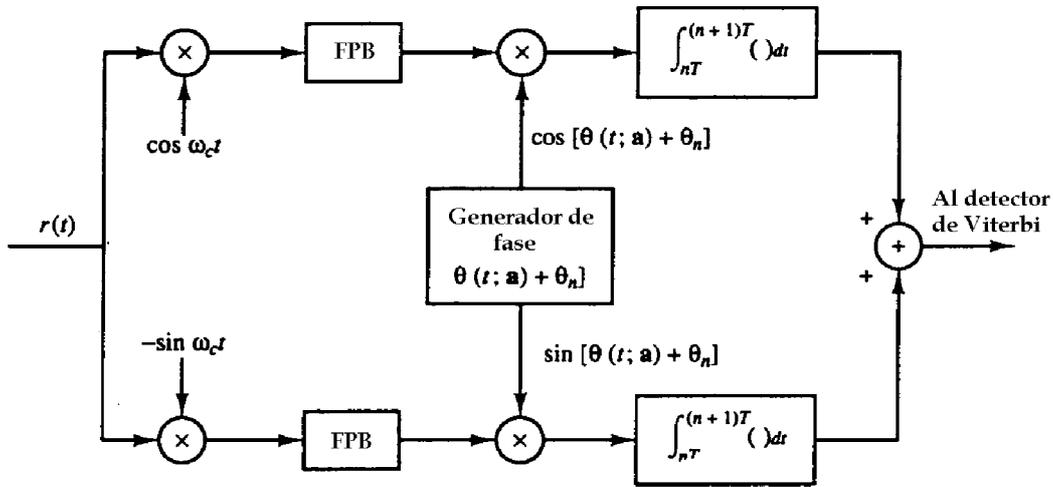


Figura 7.51: Cálculo de la métrica $C_n(\theta_n; \mathbf{a}, \mathbf{r})$.

$$P_M \approx N_{d \min} Q \left(\sqrt{\frac{d_{\min}^2}{2N_0}} \right) \quad (7.69)$$

donde $N_{d \min}$ es el número de caminos de distancia mínima. Para PSK binaria convencional sin memoria $N = 1$ y $d_{\min}^2 = d_{12}^2 = 2\mathcal{E}_b$. De esta forma (7.69) concuerda con los resultados previos.

En general d_{\min}^2 es una función del índice de modulación h , el número de símbolos M y la forma del pulso $g_T(t)$. La elección del índice de modulación h y la forma del pulso tienen un impacto significativo en la determinación del ancho de banda ocupado por la señal transmitida. La clase de pulsos de coseno elevado, de la forma

$$g_T(t) = \begin{cases} \frac{1}{2LT} \left(1 - \cos \frac{2\pi t}{LT} \right), & 0 \leq t \leq LT \\ 0, & \text{para todo otro } t \end{cases} \quad (7.70)$$

es especialmente adecuada para el diseño de señales CPM de ancho de banda eficiente.

El valor de d_{\min}^2 ha sido evaluado para una amplia variedad de señales CPM, incluyendo de respuesta completa y de respuesta parcial. Por ejemplo, la figura 7.52 ilustra el valor de d_{\min}^2 normalizada por $4\mathcal{E}_b$ en función del producto tiempo - ancho de banda $2WT_b$, donde W está definido por el ancho de banda del 99 % de potencia. La señal MSK sirve como referencia (0 dB) porque resulta en $d_{\min}^2 = 4\mathcal{E}_b$. La notación LRC indica que $g_T(t)$ es un pulso de coseno elevado con $L = 3, 4$ y 6 , según definido por (7.70). A lo largo de cualquier curva en la figura el ancho de banda W aumenta cuando aumenta el índice de modulación h , lo mismo que la SNR. Además, la eficiencia de ancho de banda aumenta con el incremento del tamaño del símbolo M para una forma del pulso coseno elevado fija. Se puede observar en esta figura que es posible ganar varios decibeles utilizando señales de respuesta parcial y alfabetos de símbolos grandes. El principal precio pagado por esta ganancia de desempeño es el incremento exponencial de complejidad computacional en la implementación del detector de Viterbi para llevar a cabo la búsqueda del camino más probable a través de la trellis.

Los resultados de desempeño de la figura 7.52 ilustran que mediante CPM de respuesta parcial y pulso coseno elevado para $M = 4$ se puede obtener una ganancia relativa de 3-4 dB en relación a MSK sin incremento relativo en ancho de banda. A pesar que esos resultados se han obtenido para pulsos coseno elevado, ganancias similares podrían obtenerse con otras señales de respuesta parcial. Es posible enfatizar que esta ganancia en SNR se logra introduciendo memoria en la modulación de la señal y utilizando esta memoria en la detección de la señal.

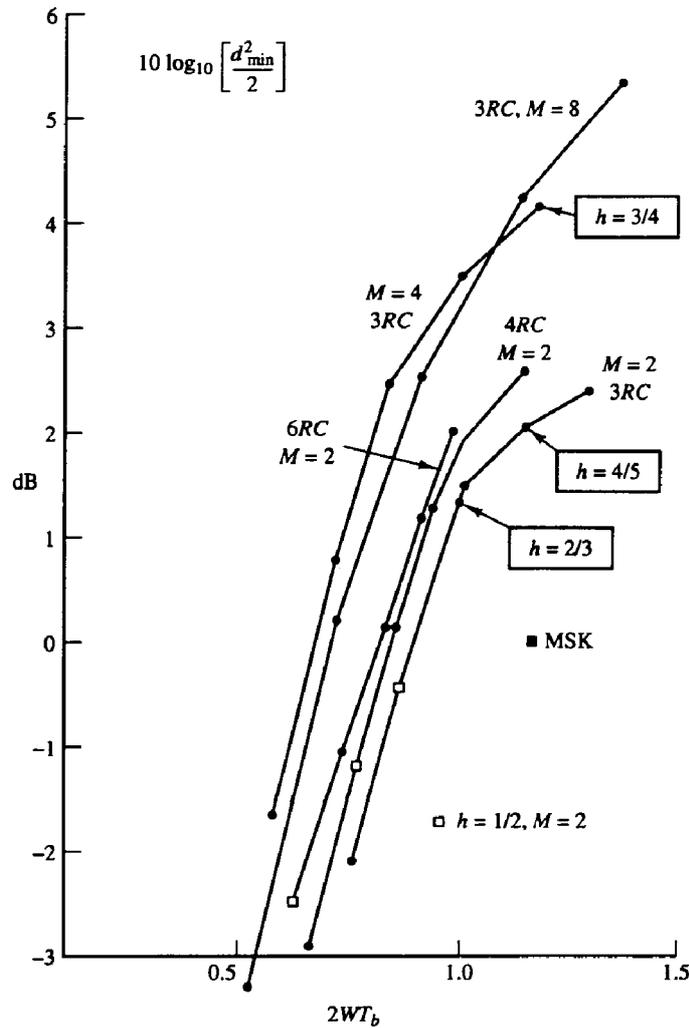


Figura 7.52: Compromiso potencia - ancho de banda para señales CPM de respuesta parcial con pulsos coseno elevado; W es el ancho de banda del 99 % de potencia.

7.5.3 Características espectrales de señales CPM (*)

Se discutió anteriormente las características espectrales de señales CPFSK, que son un caso especial de señales CPM. El procedimiento general para determinar la densidad espectral de potencia de una señal CPM es básicamente el mismo que el descrito para señales CPFSK. Como CPM es más general, sin embargo, la obtención de la función autocorrelación promedio y la densidad espectral de potencia es mucho más complicada y no se discutirá aquí.

Se describirá brevemente algunos aspectos generales sobre las características espectrales de las señales CPM. En general el ancho de banda ocupado por una señal CPM depende de la elección del índice de modulación h , la forma del pulso $g_T(t)$ y el número de señales M . Como se vió para el caso de señales CPFSK, valores pequeños de h resultan en señales con un ancho de banda relativamente pequeño. Este es el caso también para señales CPM.

El uso de pulsos suaves como el de coseno elevado de (7.70) donde $L = 1$ para respuesta completa y $L > 1$ para respuesta parcial, resulta en general en menor ancho de banda y en consecuencia en mayor eficiencia de ancho de banda que la utilización de pulsos rectangulares. Por ejemplo, la figura 7.53 ilustra la densidad espectral de potencia para CPM binaria con diferentes pulsos coseno elevado de respuesta parcial y $h = 1/2$. Por comparación se muestra también el espectro de CPFSK binaria con $h = 1/2$ (MSK). Es posible notar que cuando L aumenta, el pulso $g_T(t)$ se vuelve aún más suave y de esta forma se reduce el ancho de banda de la señal.

El efecto de variar el índice de modulación de una señal CPM se ilustra en la figura 7.54 para el caso

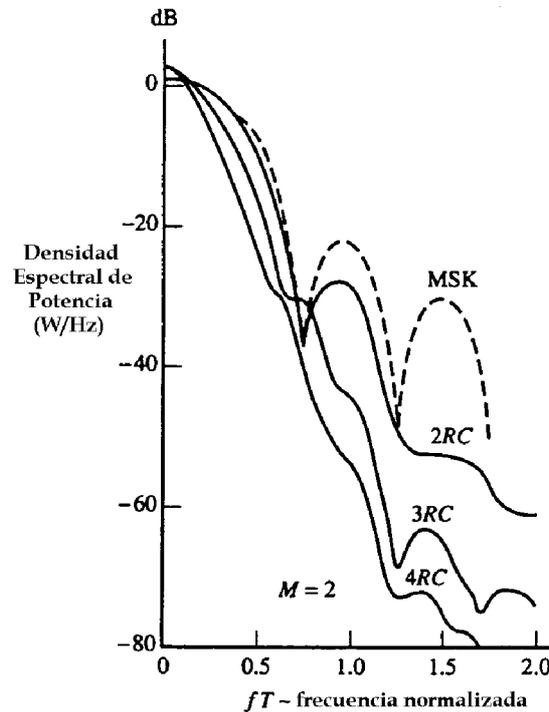


Figura 7.53: Densidad espectral de potencia para CPM binaria con $h = 1/2$ y diferentes formas de pulso.

de $M = 4$ y un pulso coseno elevado de la forma dada por (7.70) con $L = 3$. Notar que esas características espectrales son similares a las ilustradas previamente en la sección donde se discutió CPFSK, excepto que ahora se tiene características espectrales más estrechas debido a la forma del pulso coseno elevado más suave.

7.6 Transmisión con desvanecimiento y multicamino (*)

En este Capítulo y los previos se describieron metodologías de modulación y demodulación para la transmisión de información sobre dos tipos de canales, un canal AWGN y un canal de filtrado lineal. Esos canales son apropiados para una gran variedad de canales físicos. Sin embargo, no son adecuados para describir la transmisión de señales por canales de radio donde las características de transmisión cambian en el tiempo. En tales casos es necesario desarrollar modelos matemáticos más generales que caractericen ese comportamiento variante en el tiempo. Se discutirán algunos ejemplos de canales de comunicaciones que requieren un tipo diferente de modelo de canal (ver Capítulo 3).

Transmisión via propagación ionosférica en la banda HF

De la discusión en el Capítulo 3 es posible recordar la propagación de señales transmitidas en la banda de frecuencia HF, producida mediante la refracción en la ionósfera consistente en varias capas de partículas cargadas en altitudes de 50 a 400 km sobre la superficie de la tierra. Como consecuencia de esas capas de la ionósfera la señal llega al receptor a través de diferentes caminos de propagación con diferentes retardos. Esas componentes de señal se denominan *componentes multicamino*. Esas componentes de señal multicamino tienen generalmente diferentes desplazamientos de fase y, en consecuencia, pueden sumarse destructivamente resultando en un fenómeno denominado *desvanecimiento de señal*. De esta forma, el desvanecimiento es un resultado de la propagación multicamino de la señal. Para caracterizar el comportamiento de este tipo de canal es necesario adoptar un modelo de respuesta impulsiva variante en el tiempo.

Transmisión celular móvil

En la transmisión de radio celular móvil, entre la estación base y un automóvil, la señal transmitida de la estación base al automóvil se refleja usualmente en edificios, colinas u otras construcciones circundantes. En

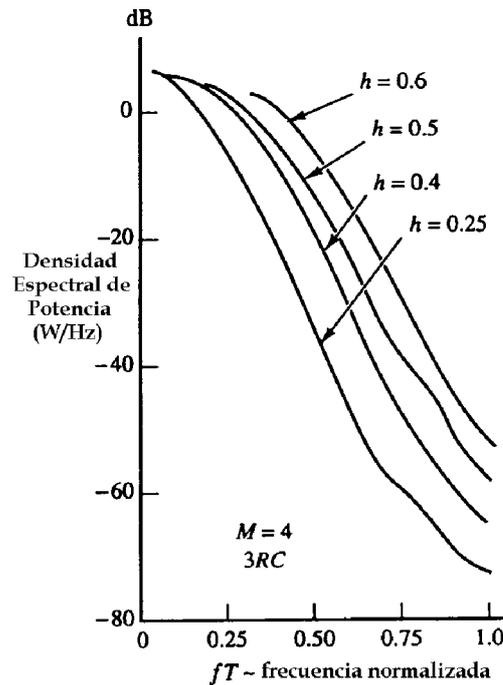


Figura 7.54: Densidad espectral de potencia para CPM cuaternaria para diferentes índices de modulación (3RC).

consecuencia, en el receptor se observan caminos de propagación múltiples con diferentes retardos. De esta forma, la señal recibida tiene características similares a aquellas asociadas a la propagación ionosférica. Lo mismo sucede para la transmisión desde un automóvil a la estación base. Además, la velocidad a la que el automóvil se está desplazando resulta también en desplazamientos en frecuencia de las varias componentes de frecuencia de la señal, estos se denominan *desplazamientos Doppler*.

Transmisión de radio microondas de enlace visual

En la transmisión de señales de radio de enlace visual las antenas de transmisión y recepción se montan generalmente sobre torres para evitar obstrucciones tal como edificios y colinas en el camino de propagación. Sin embargo, como no puede eliminarse totalmente esas obstrucciones es probable que las señales se reflejen, aún en el suelo. Esto es especialmente un problema sobre todo en condiciones meteorológicas severas. En ese caso existe una componente de señal recibida que llega al receptor via el camino directo y un conjunto de caminos secundarios que se reflejan en el terreno. Estos últimos llegan al receptor con diferentes retardos y constituyen propagación multicamino. En general en la transmisión de enlace visual por microondas se utiliza antenas de ancho de haz relativamente estrecho para reducir la ocurrencia de reflexiones secundarias. Independientemente de ello, algunas reflexiones secundarias de señal se observan frecuentemente en la práctica.

Comunicaciones de radio de avión a avión

En las comunicaciones radiales entre dos aviones es posible que componentes secundarias de señal sean recibidas a partir de reflexiones en el suelo. Este es el caso cuando se utiliza antenas omnidireccionales. El conjunto de componentes de señal reflejadas a partir del suelo llegan generalmente al receptor con diferentes retardos y diferentes atenuaciones. Además, el movimiento relativo de los aviones produce desplazamientos de frecuencia Doppler en las diferentes componentes de señal. En varios aspectos, esta situación es similar a la de comunicaciones celulares móviles.

Los cuatro canales descriptos aquí brevemente pueden caracterizarse en general como sistemas lineales

con respuestas impulsivas variantes en el tiempo. Como es generalmente difícil, si no imposible, caracterizar los efectos microscópicos de la transmisión de señal sobre los canales descritos en forma determinística, es lógico que se adopte una caracterización estadística.

7.6.1 Modelo para canales con multicamino (*)

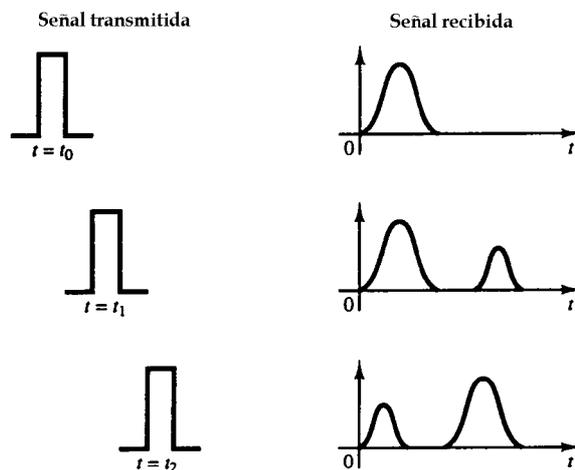


Figura 7.55: Ilustración de las características de respuesta de un canal variante en el tiempo.

Como es posible notar, existen dos características diferentes en los tipos de canales de radio discutidos. Una característica es que la señal transmitida llega al receptor via múltiples caminos de propagación, cada uno con un retardo de tiempo asociado. Por ejemplo, si se transmite un pulso extremadamente corto, idealmente un impulso, la respuesta del canal debido a dispersores múltiples (tal como las partículas ionizadas de la atmósfera) podría aparecer como mostrado en la figura 7.55. Como la señal recibida se dispersa en el tiempo debido a los dispersores múltiples con diferentes retardos, se dice que el canal es *dispersivo en tiempo*.

Una segunda característica de los tipos de canales de radio discutidos está relacionada con las variaciones en el tiempo de la estructura del medio. Como resultado de tales variaciones en el tiempo, la respuesta del canal a cualquier señal transmitida a través de él cambiará en el tiempo. En consecuencia, si se repite el experimento con un pulso corto exhaustivamente será posible observar cambios en la señal recibida debidos a cambios en el medio físico. Tales cambios incluyen variaciones en los retardos relativos de las señales de los dispersores múltiples. De esta forma la señal recibida podría parecer como ilustrada en la figura 7.55. En otras palabras, la respuesta impulsiva del canal es variante en el tiempo. En general, las variaciones en el tiempo en la señal recibida parecen impredecibles para el usuario del canal. Esto conduce a caracterizar al canal multicamino variante en el tiempo estadísticamente.

Para obtener una descripción estadística del canal, es posible considerar la transmisión de una portadora no modulada $c(t) = A \cos 2\pi f_c t$. La señal recibida en ausencia de ruido puede expresarse como

$$x(t) = A \sum_n \alpha_n(t) \cos[2\pi f_c(t - \tau_n(t))] = A \operatorname{Re} \sum_n \alpha_n(t) e^{-j2\pi f_c \tau_n(t)} e^{j2\pi f_c t}$$

donde $\alpha_n(t)$ es la atenuación variante en el tiempo asociada al camino de propagación n y $\tau_n(t)$ es el retardo de propagación correspondiente. La señal compleja

$$z(t) = \sum_n \alpha_n(t) e^{-j2\pi f_c \tau_n(t)} = \sum_n \alpha_n(t) e^{-j\phi_n(t)}$$

representa la respuesta del canal a la exponencial compleja $e^{j2\pi f_c t}$. Notar que, a pesar que la entrada al canal es una señal monocromática, o sea una señal de una única frecuencia, la salida del canal consiste en una señal que contiene varias componentes de frecuencia diferentes. Esas nuevas componentes se generan como

resultado de las variaciones en el tiempo en la respuesta del canal. El ancho de banda de $z(t)$ se denomina *dispersión de frecuencia Doppler* del canal, B_d . Esta cantidad es una medida de cuán rápidamente varía la señal $z(t)$ en el tiempo. Si $z(t)$ varía lentamente, la dispersión de frecuencia Doppler es relativamente pequeña, mientras que si $z(t)$ varía rápidamente, la dispersión de frecuencia Doppler es grande.

Es posible interpretar la señal compleja recibida $z(t)$ como la suma de cierto número de vectores (fasores), cada uno de los cuales tiene una amplitud variante en el tiempo $\alpha_n(t)$ y fase $\phi_n(t)$. En general, se requieren grandes cambios dinámicos en el medio físico para lograr un gran cambio en $\{\alpha_n(t)\}$. Por otro lado, las fases $\{\phi_n(t)\}$ cambiarán en 2π radianes siempre que $\{\tau_n(t)\}$ cambie en $1/f_c$. Pero $1/f_c$ es un valor pequeño y, en consecuencia, las fases $\{\phi_n(t)\}$ cambiarán en 2π o más radianes con cambios relativamente pequeños en las características del medio. Es posible esperar también que los retardos $\{\tau_n(t)\}$ asociados a los diferentes caminos de señal cambien a diferentes velocidades y de forma impredecible (aleatoria). Esto implica que la señal compleja $z(t)$ puede modelarse mediante un proceso estocástico. Cuando exista un gran número de caminos de propagación de señal puede aplicarse el teorema del límite central. De esta forma $z(t)$ puede modelarse como un proceso Gaussiano complejo.

El modelo de propagación multicamino del canal, asociado a la señal recibida $x(t)$, o equivalentemente $z(t)$, resulta en el desvanecimiento de la señal. El fenómeno de desvanecimiento es resultado primariamente de los factores de fase variantes en el tiempo $\{\phi_n(t)\}$. A veces los vectores complejos $z(t)$ se suman destructivamente para reducir el nivel de potencia de la señal recibida. Otras veces los vectores $z(t)$ se suman constructivamente y de esta forma producen una señal mayor. Las variaciones de amplitud en la señal recibida debidas a la propagación multicamino variante en el tiempo del canal se denominan usualmente *desvanecimiento de señal*.

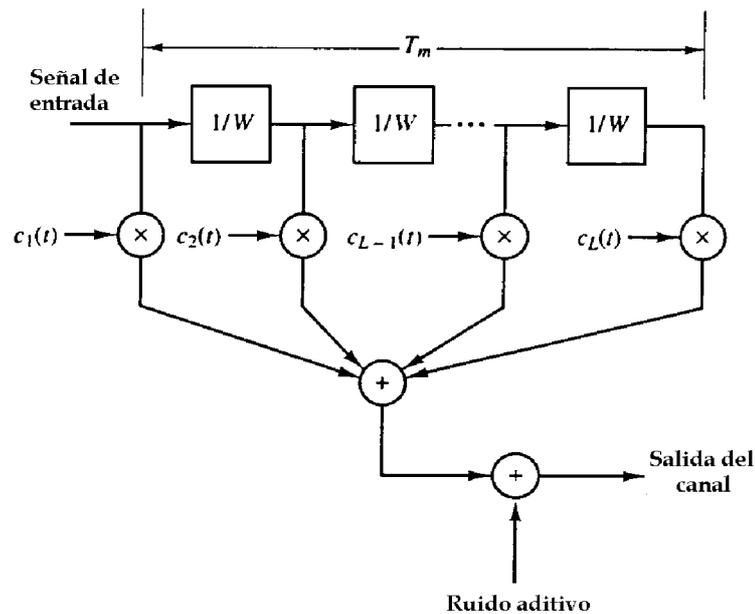


Figura 7.56: Modelo para un canal multicamino variante en el tiempo.

Un modelo general para un canal multicamino variante en el tiempo se ilustra en la figura 7.56. El modelo de canal consiste de una línea de retardos uniformemente espaciada. El espaciamiento entre retardos es $1/W$, donde W es el ancho de banda de la señal transmitida a través del canal. Los coeficientes $\{c_n \equiv \alpha_n(t)e^{j\phi_n(t)}\}$ se modelan usualmente como procesos Gaussianos complejos mutuamente decorrelacionados. La longitud de la línea de retardos corresponde a la cantidad de dispersión temporal en el canal multicamino, que usualmente se denomina *dispersión multicamino*, $T_m = L/W$, donde L representa el número máximo posible de componentes de señal multicamino.

Ejemplo 1. Determinar un modelo de canal apropiado para la propagación ionosférica de dos caminos, donde el retardo de tiempo relativo entre los dos caminos de señal es 1 msec y el ancho de banda de la señal

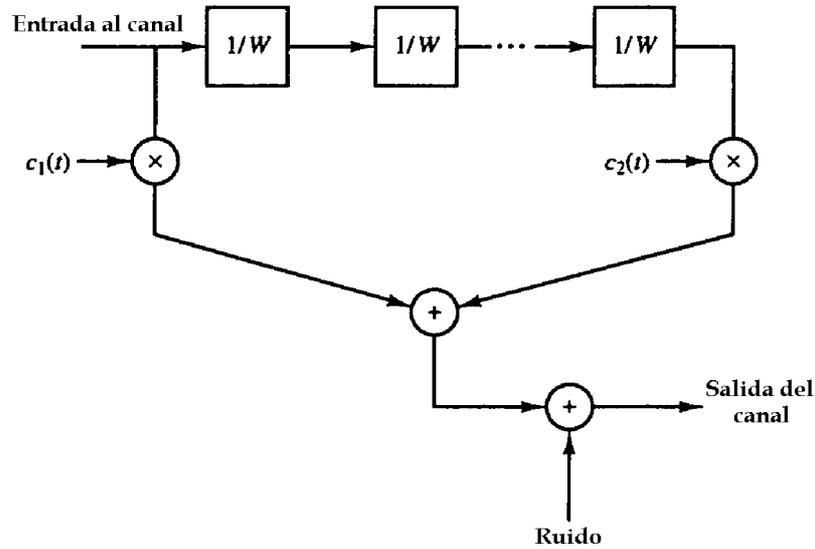


Figura 7.57: Modelo de canal de dos caminos.

transmitida es de 10 kHz.

Una señal de 10 kHz provee una resolución en tiempo de $1/W = 0.1$ mseg. Como el retardo relativo de tiempo entre los caminos de señal recibidos es 1 mseg. el modelo de línea de retardos incluye 10 retardos con solo el primer y último coeficientes variantes en el tiempo $c_1(t)$ y $c_2(t)$ no nulos, como se ilustra en la figura 7.57. Como $c_1(t)$ y $c_2(t)$ representan la respuesta de señal a un gran número de partículas ionizadas de dos regiones diferentes de la ionósfera es posible caracterizarlos como procesos estocásticos Gaussianos no correlacionados complejos. La velocidad de variación de los coeficientes determinará el valor de la dispersión Doppler para cada camino.

Ejemplo 2. Determinar un modelo de canal apropiado para un enlace entre aviones en el cual existe un camino directo de señal y una propagación secundaria resultante de la dispersión de señal debido a la superficie circundante. El camino secundario tiene un retardo de propagación de $\tau_0 = 10 \mu\text{seg}$ relativo al retardo de propagación del camino directo. El ancho de banda de señal es $W = 100$ kHz.

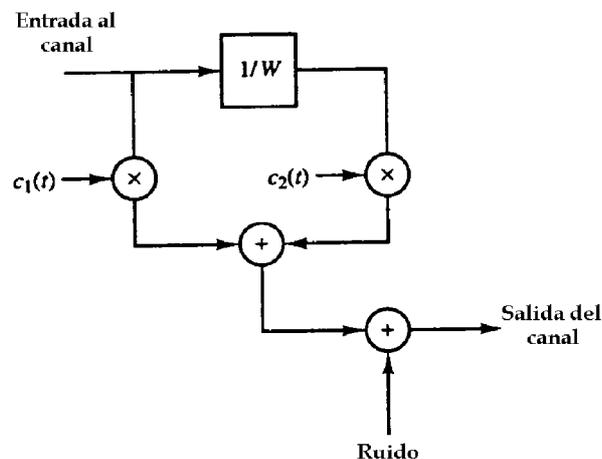


Figura 7.58: Modelo de canal de dos caminos.

Una señal de 100 kHz provee una resolución en tiempo de $1/W = 1 \mu\text{seg}$. En consecuencia, el camino de señal secundario es resoluble debido a que el retardo de tiempo relativo es $10 \mu\text{seg}$. En este caso, un modelo apropiado de canal tiene dos coeficientes $c_1 = \alpha_1 e^{j\phi_1}$ y $c_2(t) = \alpha_2(t) e^{j\phi_2(t)}$, como mostrado en la figura 7.58. Notar que el camino directo se modela con una atenuación fija (invariante en el tiempo) y una fase fija. Por otro lado, el camino secundario que resulta del terreno se modela con atenuación y fase variantes en el tiempo, debido a que el terreno cambia en el tiempo por el movimiento de los aviones. En este caso es apropiado modelar a $c_2(t)$ como un proceso estocástico Gaussiano complejo. La velocidad de variación del coeficiente $c_2(t)$ determina el valor de la dispersión de frecuencia Doppler para este camino. Notar que en ambos caminos de señal el directo y el reflejado existirán desplazamiento de frecuencia Doppler resultantes del citado movimiento de los aviones.

Ejemplo 3. Determinar el modelo apropiado de canal para el enlace entre aviones del ejemplo anterior pero ahora suponiendo que el ancho de banda de la señal transmitida es 10 kHz.

Una señal de 10 kHz provee una resolución en tiempo de $100 \mu\text{seg}$. Como el retardo relativo entre las dos componentes de señal es de $10 \mu\text{seg}$ los dos caminos de señal no pueden resolverse en tiempo. En consecuencia, el canal aparece como un único camino con coeficiente $c(t)$ que incluye el camino directo y el secundario. Como el camino secundario resulta de la dispersión de señal a través de un gran número de dispersores es posible caracterizar $c(t)$ como un proceso estocástico Gaussiano complejo con valor medio $\alpha_1 e^{j\phi_1}$, asociado al camino directo.

En los modelos de canal de estos ejemplos los coeficientes en el modelo de línea de retardos fueron caracterizados por un proceso estocástico Gaussiano complejo. Es posible expresar cada coeficiente como

$$c(t) = c_r(t) + jc_i(t)$$

donde $c_r(t)$ y $c_i(t)$ representan procesos estocásticos Gaussianos reales. Se asume que $c_r(t)$ y $c_i(t)$ son estacionarios y estadísticamente independientes. Esta suposición se mantiene en general para el modelo de línea de retardos de canales físicos.

Es posible también expresar $c(t) = v(t)e^{j\phi(t)}$, donde $v(t) = \sqrt{c_r^2(t) + c_i^2(t)}$ y $\phi(t) = \tan^{-1} c_i(t)/c_r(t)$. En esta representación, si $c_r(t)$ y $c_i(t)$ son Gaussianos de media cero, la amplitud $v(t)$ está caracterizada estadísticamente por una función densidad de probabilidad Rayleigh y $\phi(t)$ estará uniformemente distribuida en el intervalo $[0, 2\pi)$. De esta forma el canal se denomina *canal con desvanecimiento Rayleigh*. Por otro lado, si $c_r(t)$ y $c_i(t)$ son Gaussianas con media no nula, como en el enlace de comunicaciones entre aviones, la amplitud $v(t)$ estará caracterizada por una densidad de probabilidad Rice y la fase $\phi(t)$ tampoco tendrá media cero. En este caso el canal se denomina *canal con desvanecimiento Rice*.

Además de la dispersión (en tiempo) multicamino T_m y la dispersión (en frecuencia) Doppler B_d , existen otros dos parámetros útiles para caracterizar canales con multicamino y desvanecimiento. Un parámetro es el inverso de la dispersión Doppler. Esta cantidad es una medida del intervalo de tiempo sobre el cual las características del canal cambiarán muy poco. Este parámetro se denomina el *tiempo de coherencia* y se define como

$$T_{tc} = \frac{1}{2B_d}$$

El otro parámetro útil es el inverso de la dispersión multicamino, que tiene unidades de frecuencia. Esta cantidad es una medida del ancho de banda sobre el cual las características del canal $v(t)$ y $\phi(t)$ estarán altamente correlacionadas. En otras palabras, todas las componentes de frecuencia de una señal dentro de este ancho de banda se desvanecerán simultáneamente. Este parámetro se denomina *ancho de banda de coherencia* y está definido por

$$B_{ac} = \frac{1}{2T_m}$$

Canal	T_m seg.	B_d Hz.
Propagación ionosférica de onda corta (HF)	10^{-3} - 10^{-2}	10^{-1} - 1
Dispersión ionosférica (VHF)	10^{-4}	10
Dispersión troposférica (SHF)	10^{-6}	10
Luna llena	10^{-2}	10

Tabla 7.3: Dispersión multicamino y dispersión Doppler para varios canales multicamino variantes en el tiempo.

Ejemplo 4. Considerar el enlace de comunicaciones del segundo ejemplo. Determinar el ancho de banda de coherencia de ese canal.

La dispersión multicamino para este canal es $T_m = 10 \mu\text{seg}$. Entonces, el ancho de banda de coherencia es de 50 kHz. Dentro de este ancho de banda todas las componentes de frecuencia de la señal estarán afectadas en forma similar por el canal. Por ejemplo, todas los componentes de frecuencia de una señal que caen dentro del ancho de banda de coherencia se desvanecerán simultáneamente.

El primero, segundo y último ejemplos proveen alguna indicación del rol que juega el ancho de banda de la señal en relación con las características del modelo del canal. Por ejemplo, si la señal transmitida tiene un ancho de banda W que es mayor que el ancho de banda de coherencia B_{ac} del canal, como en el caso del primer y segundo ejemplos, las componentes multicamino son resolubles, y como las componentes de frecuencia de la señal transmitida están separadas más que B_{ac} , las mismas serán afectadas por el canal de forma diferente. En tal caso, se dice que el canal es *selectivo en frecuencia*. Por otro lado, si el ancho de banda de la señal W es menor que B_{ac} como en el caso del tercer ejemplo, todas las componentes de frecuencia de la señal estarán afectadas en forma similar por el canal en cualquier instante de tiempo. En tal caso se dice que el canal es *no selectivo en frecuencia*.

A continuación se describe como pueden usarse esos parámetros del canal para el diseño de métodos de modulación para la transmisión de información. La tabla 7.3 muestra los valores de esos parámetros para diferentes señales multicamino.

7.6.2 Señales para canales con desvanecimiento y multicamino (*)

Existen diferentes aproximaciones al diseño de señales para transmitir información digital a través de canales con desvanecimiento y multicamino. A pesar que en principio cualquier método de modulación (PAM, QAM, PM, FSK y CPM) es aplicable, el desvanecimiento de señal origina generalmente grandes fluctuaciones en la amplitud de la señal recibida y en consecuencia es extremadamente difícil discriminar entre niveles de amplitud diferentes en la señal recibida. Por esta razón se evita generalmente PAM y QAM en este tipo de canales. Esta discusión se enfocará la atención en PSK y FSK.

Como se indicó en la sección previa, el ancho de banda de canal disponible juega un papel importante en el diseño de la modulación. Si el ancho de banda asignado al usuario excede el ancho de banda de coherencia B_{ac} del canal se tienen básicamente dos opciones. Una opción es transmitir la información sobre una única portadora sinusoidal, usando una señal que ocupe el ancho de banda completo del canal. En este caso las componentes de señal multicamino son resolubles dentro de una resolución en tiempo de $1/W$. De esta forma, los caminos de propagación de señal que difieren en tiempo de llegada al receptor mayor o igual $1/W$ serán distingibles como componentes de señal separadas. El número de componentes de señal diferentes depende de las características del canal. Por ejemplo, en el caso del enlace de comunicaciones entre aviones es posible esperar el camino directo y el camino de reflexión terrestre. En la transmisión mediante propagación ionosférica en la banda de frecuencias de HF, el número de componentes de señal recibidas dependerá del número de capas reflectoras en la atmósfera y de la posibilidad de caminos llegando por saltos (rebotes) múltiples.

Observar que cuando $W > B_{ac}$, la señal recibida puede sufrir interferencia intersímbolo a menos que la velocidad de transmisión de símbolos se elija significativamente menor que la dispersión multicamino T_m . Recordar que la velocidad de transmisión de símbolos $1/T$ satisface la condición $1/T \leq 2W$. Por ejemplo, si

se escoge $1/T = W$, entonces como $T_m > 1/W$ se concluye que $T_m > T$. En este caso, el receptor encuentra ISI y debe utilizar algún tipo de ecualización, como discutido en el Capítulo anterior. Por otro lado, si la velocidad de transmisión de símbolos se escoge muy pequeña, o sea $T \gg T_m$, entonces la ISI se vuelve insignificante y no se requiere ecualización. Ese tipo de diseño se realiza usualmente para señales de espectro disperso, que se discutirá en el Capítulo 9.

Ejemplo. Suponer la utilización de un canal HF con un ancho de banda nominal de 3200 Hz para la transmisión de información digital a una velocidad de a) 4800 bits/seg o b) 20 bits/seg. La dispersión multicamino del canal es $T_m = 5$ mseg. Especificar un método de modulación para lograr tales velocidades e indicar cuando es necesario un ecualizador en el receptor para combatir la interferencia intersímbolo.

Para el sistema de 4800 bits/seg es posible escoger una velocidad de 2400 símbolos/seg y modulación de cuatro fases (PSK o DPSK). Una velocidad de 2400 puede asignarse a un canal de 3200 Hz. de ancho de banda, existiendo un exceso de ancho de banda de 800 Hz. Puede utilizarse un filtro de transmisión de coseno elevado para obtener la forma de señal deseada. La velocidad de $1/T = 2400$ implica que la señal recibida tendrá ISI ($T_m \gg T$), lo cual indica la necesidad de utilizar un ecualizador para reducir sus efectos sobre el desempeño.

Para el sistema de 20 bits/seg es posible utilizar nuevamente la modulación de cuatro fases (PSK o DPSK). De esta forma, la velocidad de transmisión de símbolos será de 10 símbolos/seg y entonces $T = 100$ mseg. Como $T \gg T_m$, el efecto de la dispersión multicamino es despreciable y no será necesaria ecualización en el receptor. Para utilizar el ancho de banda completo del canal es posible utilizar una señal de espectro disperso como se discutirá en el Capítulo 9.

La segunda opción cuando el ancho de banda de canal disponible excede el ancho de banda de coherencia $W > B_{ac}$ es utilizar transmisión multiportadora. En este método el ancho de banda de canal disponible se subdivide en N subcanales tal que el ancho de banda de cada uno sea mucho menor que B_{ac} . El objetivo primario de esta subdivisión del ancho de banda del canal es transmitir información en cada subcanal a una velocidad $T_{sc} \gg T_m$. Con N subcanales la velocidad de transmisión de símbolos total será N/T_{sc} .

Ejemplo. Considerar un canal HF con un ancho de banda nominal de 3200 Hz y una dispersión multicamino de $T_m = 1$ mseg. Diseñar un modulador multiportadora que lograr una velocidad de transmisión de 4800 bits/seg.

Es posible seleccionar el número N de subportadoras tan grande como se quiera para lograr la condición deseada $T_{sc} \gg T_m$. Sin embargo, la complejidad del demodulador aumenta (linealmente) con N y el retardo de demodulación para brindar la información al usuario también aumentará (linealmente) con N . Por ello, es deseable mantener N tan pequeño como sea posible. Suponiendo la elección de N de forma que $T_{sc} = 100$ mseg. Entonces cada subcanal puede ser tan estrecho como $W_{ac} \cong 1/T_{sc} = 10$ Hz. Notar que $W_{sc} \ll B_{ac} = 500$ Hz como deseado. Si se utiliza modulación de cuatro fases (PSK o DPSK) para cada subcanal se puede lograr una velocidad de transmisión de 20 bits/seg por subcanal. Con $N = 240$ subcanales se logra la velocidad deseada de 4800 bits/seg.

De las dos opciones descriptas, el sistema de subportadoras múltiples no requiere ecualización y puede ser preferible desde el punto de vista de la reducción de la complejidad del receptor, especialmente cuando la dispersión multicamino del canal es grande, requiriendo la utilización de un ecualizador muy largo en el receptor. Por otro lado, si la dispersión multicamino del canal es relativamente pequeña y la ISI está confinada a pocos símbolos (2 o 3 por ejemplo), la complejidad del ecualizador también será pequeña. En tal caso, el método de diseño por multiportadora puede que no sea atractivo.

En la discusión anterior se supuso que el ancho de banda disponible de canal excedía el ancho de banda de coherencia B_{ac} del canal. Por otro lado, si el ancho de banda del canal es mucho menor que B_{ac} , la alternativa de multiportadoras no tiene el menor interés. Puede diseñarse un sistema de portadora única que utilice el ancho de banda completo, con una velocidad de transmisión de símbolos de $W \leq 1/T \leq 2W$. En este caso $T \gg T_m$, tal que el efecto de la ISI sobre el desempeño del sistema será despreciable.

A continuación se evaluará el desempeño del demodulador y detector para las modulaciones PSK, DPSK y FSK en un canal Rayleigh bajo la condición que la ISI sea despreciable.

7.6.3 Desempeño en canales con desvanecimiento Rayleigh (*)

En esta sección se determinará la probabilidad de error en un receptor de un sistema de comunicaciones digital binario que transmite información a través de un canal con desvanecimiento Rayleigh. El ancho de banda de la señal W se supone mucho menor que el ancho de banda de coherencia B_{ac} del canal, como en el tercer ejemplo de la sección anterior. Como las componentes de multicamino no son resolubles, el canal no es selectivo en frecuencia y, por ello, la respuesta impulsiva del canal se representa como

$$h(\tau; t) = \alpha(t)\delta(\tau - \tau_0(t))$$

donde $\alpha(t)$ tiene densidad de probabilidad Rayleigh en cualquier instante de tiempo.

Se supone que las variaciones en el tiempo de $\alpha(t)$ y $\tau_0(t)$ son muy lentas comparadas con el intervalo de símbolo, de forma que dentro del intervalo de tiempo $0 \leq t \leq T$, la respuesta impulsiva del canal es constante, o sea

$$h(\tau; t) \equiv h(\tau) = \alpha\delta(\tau - \tau_0)$$

donde la amplitud tiene densidad de probabilidad Rayleigh, o sea

$$p(\alpha) = \begin{cases} \frac{\alpha}{\sigma_\alpha^2} e^{-\alpha^2/2\sigma_\alpha^2}, & \alpha > 0 \\ 0, & \text{para todo otro } \alpha \end{cases}$$

y $\sigma_\alpha^2 = E(\alpha^2)/2$.

Suponiendo ahora la utilización de señales binarias antipodales para la modulación PSK, las dos señales posibles serán

$$u_m(t) = \sqrt{\frac{2\mathcal{E}_b}{T}} \cos(2\pi f_c t + m\pi) + n(t), \quad m = 0, 1$$

La señal recibida en el intervalo $0 \leq t \leq T$ será

$$r(t) = \alpha \sqrt{\frac{2\mathcal{E}_b}{T}} \cos(2\pi f_c t + m\pi + \phi) + n(t)$$

donde ϕ es el desplazamiento de fase de la portadora. Suponiendo que ϕ es conocida en el receptor, debe correlacionarse con

$$\psi(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t + \phi), \quad 0 \leq t \leq T$$

Tal que la entrada al detector en los instantes de muestreo será

$$r = \alpha \sqrt{\mathcal{E}_b} \cos m\pi + n, \quad m = 0, 1$$

Para un valor de α fijo, la probabilidad de error tendrá la forma familiar

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

Es posible interpretar $P_2(\alpha)$ como una probabilidad de error condicional para un valor dado de atenuación del canal α . Para determinar la probabilidad de error promedio para todos los valores posibles de α se calcula la integral

$$P_2 = \int_0^{\infty} P_2(\alpha) p(\alpha) d\alpha \quad (7.71)$$

donde $p(\alpha)$ es la densidad de probabilidad Rayleigh. Esta integral tiene una expresión simple en forma cerrada dada por

$$P_2 = \frac{1}{2} \left[1 - \sqrt{\frac{\bar{\rho}}{1 + \bar{\rho}}} \right]$$

donde, por definición

$$\bar{\rho} = \frac{\mathcal{E}_b}{N_0} E(\alpha^2) \quad (7.72)$$

tal que, $\bar{\rho}$ es la SNR/bit promedio y $E(\alpha^2) = 2\sigma_\alpha$.

Si las señales binarias son ortogonales, como en el caso de FSK ortogonal, las dos señales transmitidas posibles serán

$$u_m(t) = \sqrt{\frac{2\mathcal{E}_b}{N_0}} \cos \left[2\pi \left(f_c + \frac{m}{2T} t \right) \right], \quad m = 0, 1$$

y la señal recibida será

$$r(t) = \alpha \sqrt{\frac{2\mathcal{E}_b}{T}} \cos \left[2\pi \left(f_c + \frac{m}{2T} t + \phi \right) \right] + n(t)$$

En este caso, la señal recibida se correlaciona con las señales

$$\begin{aligned} \psi_1(t) &= \sqrt{\frac{2}{T}} \cos[2\pi f_c t + \phi] \\ \psi_2(t) &= \sqrt{\frac{2}{T}} \cos[2\pi(f_c + (1/2T)t + \phi)] \end{aligned}$$

Si $m = 0$, por ejemplo, las salidas de los dos correladores serán

$$\begin{aligned} r_1 &= \alpha \sqrt{\mathcal{E}_b} + n_1 \\ r_2 &= n_2 \end{aligned}$$

donde n_1 y n_2 son las componentes de ruido aditivo a la salida de los dos correladores. De esta forma, la probabilidad de error es simplemente la probabilidad que $r_2 > r_1$. Como las señales son ortogonales, la probabilidad de error para un valor fijo de α tiene la forma familiar

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

Como en el caso de señales antipodales, la probabilidad de error para todos los valores de α se determina calculando la integral de (7.71). De esta forma se obtiene

$$P_2 = \frac{1}{2} \left[1 - \sqrt{\frac{\bar{\rho}}{2 + \bar{\rho}}} \right]$$

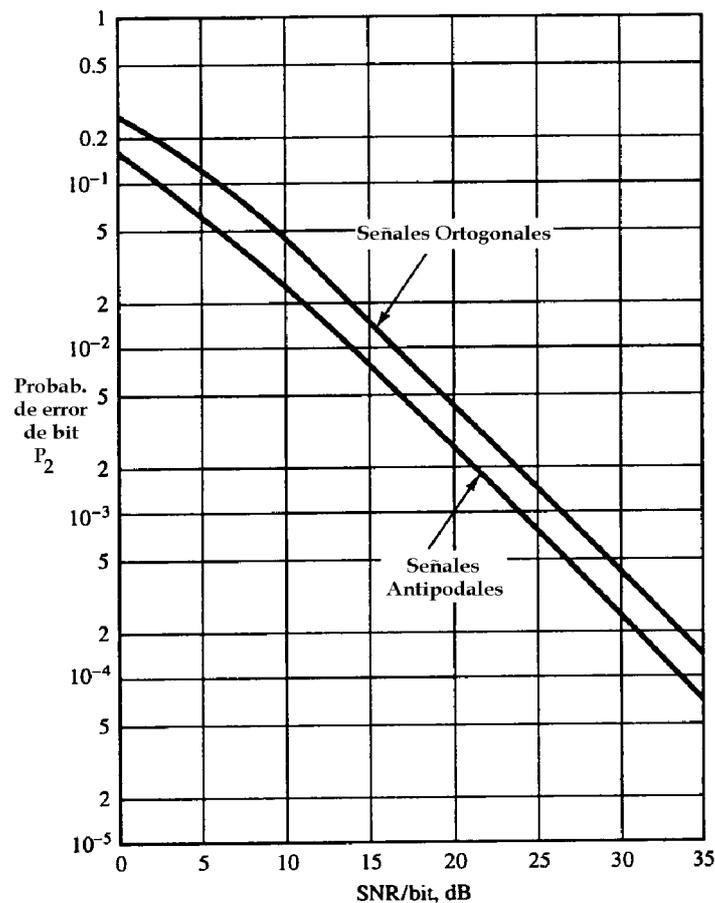


Figura 7.59: Desempeño de la señalización binaria en un canal con desvanecimiento Rayleigh.

donde $\bar{\rho}_b$ es la SNR/bit promedio como definida en (7.72).

La figura 7.59 ilustra la probabilidad de error promedio para señales binarias antipodales y ortogonales. El aspecto sorprendente de esos gráficos es el lento decaimiento de la probabilidad de error en función de la SNR. En realidad, para grandes valores de $\bar{\rho}_b$ la probabilidad de error para señales binarias es

$$P_2 \approx \frac{1}{4\bar{\rho}_b}, \quad \text{señales antipodales}$$

$$P_2 \approx \frac{1}{2\bar{\rho}_b}, \quad \text{señales ortogonales}$$

de forma que la probabilidad de error en ambos casos disminuye solo inversamente con la SNR. Esto contrasta con la disminución exponencial en el caso de un canal AWGN. Es posible notar también que la diferencia en SNR entre señales antipodales (PSK binaria) y señales ortogonales (FSK binaria) es de 3 dB.

Otros dos tipos de demodulación de señales son DPSK y FSK no coherente. La probabilidad de error promedio en estos casos es

$$P_2 \approx \frac{1}{2(1 + \bar{\rho}_b)}, \quad \text{DPSK}$$

$$P_2 \approx \frac{1}{2 + \bar{\rho}_b}, \quad \text{FSK no coherente}$$

Mejora de desempeño a través de una señal de diversidad

El problema básico de las comunicaciones digitales a través de un canal con desvanecimiento es que un gran número de errores ocurren cuando la atenuación del canal es grande, o sea, cuando el canal está en un

desvanecimiento profundo. Si es posible proveer al receptor de dos o más réplicas de la misma información de señal transmitida a través de canales con desvanecimiento independientes, la probabilidad que todas las componentes de señal se desvanezcan simultáneamente se reduce considerablemente. Si p es la probabilidad que cualquier señal se desvanezca en algún valor crítico, entonces p^D es la probabilidad que D réplicas con desvanecimiento independiente de la misma señal se desvanezcan debajo del valor crítico. Existen varias formas de proveer al receptor con D réplicas con desvanecimiento independiente de la misma señal que transporta información.

Un método para lograr D versiones con desvanecimiento independiente de la misma señal que transporta información es transmitir sobre D frecuencias portadoras, donde la separación entre portadoras sucesivas es igual o mayor que el ancho de banda de coherencia B_{ac} del canal. Este método se denomina *diversidad en frecuencia*.

Un segundo método para lograr D versiones con desvanecimiento independiente de la misma señal que transporta información es transmitir la misma información en D diferentes bloques de tiempo, donde la separación entre bloques sucesivos es igual o mayor que el tiempo de coherencia del canal T_{tc} . Este método se denomina *diversidad en tiempo*.

Otro método comúnmente utilizado para lograr diversidad es mediante la utilización de antenas de recepción múltiples, pero solo una antena transmisora. Las antenas receptoras deben estar espaciadas lo suficiente tal que las componentes multicamino en la señal tengan caminos de propagación suficientemente diferentes. Usualmente se requiere una separación de 10 o más longitudes de onda entre dos antenas receptoras para obtener señales que desvanezcan independientemente.

Existen otras técnicas de diversidad en transmisión y recepción que han sido investigadas en la práctica, tal como diversidad de ángulo de llegada y diversidad de polarización. Sin embargo, esos métodos no son tan ampliamente utilizados como los discutidos anteriormente.

Como la información se transmite al receptor mediante D canales con desvanecimiento independiente, existen varias formas en que el receptor puede extraer la información transmitida de la señal recibida. El método más simple es que el receptor monitoree el nivel de potencia recibida en las D señales recibidas y seleccione para demodulación aquellas de mayor nivel. En general, esta solución resulta en la conmutación frecuente de una señal a otra. Una leve modificación que conduce a una implementación más simple es utilizar una señal para demodulación y detección siempre que el nivel de potencia recibido supere un umbral preestablecido. Cuando la señal cae debajo del umbral se conmuta al canal que tiene el mayor nivel de potencia. Este método de combinación de señal y detección se denomina *combinación de selección*.

Existen varios métodos más sofisticados para combinar las señales con desvanecimiento independiente. Uno que es apropiado para demodulación coherente y detección requiere que el detector estime y corrija los diferentes desplazamiento de fase de cada una de las D señales después de la demodulación. Luego, las señales de fase corregida a la salida de los D demoduladores se suman y se conducen al detector. Este tipo de combinación de señales se denomina *combinación de igual ganancia*. Si, además, se estima el nivel de potencia de cada una de las D señales recibidas y las salidas de los demoduladores de fase corregida se ponderan en proporción al nivel de potencia correspondiente y luego se conducen al detector, el combinador se denomina *combinador de relación máxima*. Por otro lado, si se utiliza señales ortogonales para transmitir información a través de los D canales con desvanecimiento independiente, el receptor puede emplear demodulación no coherente. En ese caso las salidas de los D demoduladores pueden elevarse al cuadrado, sumarse y luego conducirse al detector. Este combinador se denomina *combinador de ley cuadrática*.

Todos esos tipos de métodos de combinación conducen a características de desempeño que resultan en una probabilidad de error que se comporta como $K_D/\bar{\rho}_b^D$ donde K_D es una constante que depende de D y $\bar{\rho}_b$ es la SNR promedio. De esta forma, se logra una disminución exponencial de la probabilidad de error. La probabilidad de error para señales antipodales con combinación de relación máxima tiene la forma siguiente

$$P_2 \approx \frac{K_D}{(4\bar{\rho}_b)^D}, \quad \bar{\rho}_b \gg 1$$

donde K_D se define como

$$K_D = \frac{(2D-1)!}{D!(D-1)!}$$

Para señales ortogonales con combinación de ley cuadrática, la probabilidad de error tiene la siguiente forma asintótica

$$P_2 \approx \frac{K_D}{\bar{\rho}_b^D}, \quad \bar{\rho}_b \gg 1$$

Finalmente, para DPSK con combinación de igual ganancia, la probabilidad de error tiene la siguiente forma asintótica

$$P_2 \approx \frac{K_D}{(2\bar{\rho}_b)^D}, \quad \bar{\rho}_b \gg 1$$

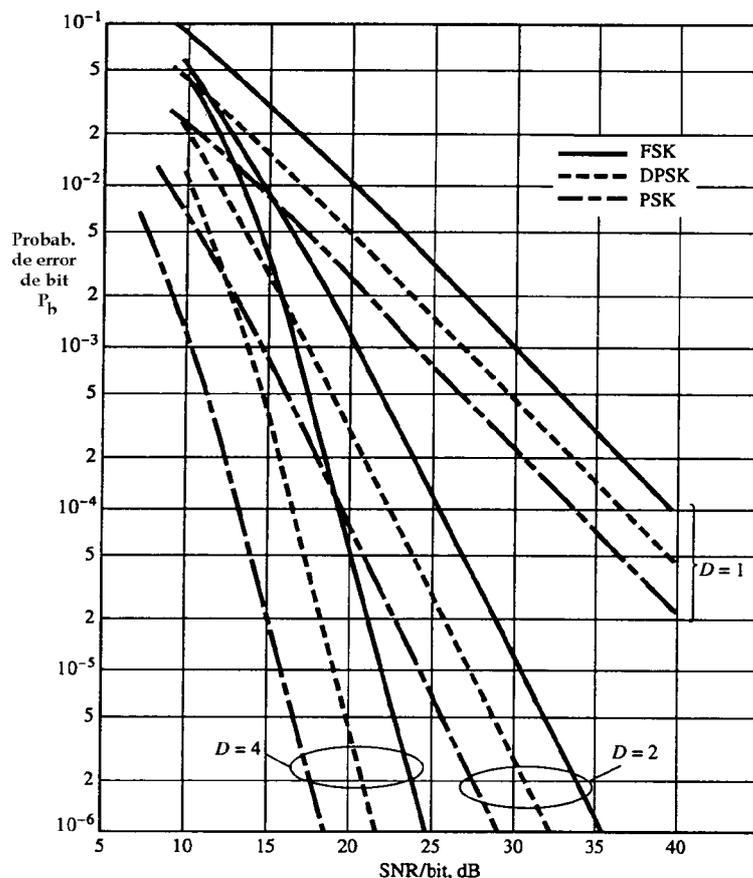


Figura 7.60: Desempeño de señalización binaria con diversidad.

Esas probabilidades de error se grafican en la figura 7.60 para $D = 1, 2, 4$. Es evidente que se ha logrado una gran reducción en SNR/bit al tener $D = 2$ (diversidad dual) comparando con no diversidad. Una mayor reducción en SNR se logra incrementando el orden de diversidad a $D = 4$, aunque la ganancia adicional de $D = 2$ para $D = 4$ es menor que la de ir de $D = 1$ para $D = 2$. Más allá de $D = 4$, la reducción adicional en SNR no es significativa. Por ello es necesario cuestionar las ventajas de desempeño al incrementar la diversidad más allá de $D = 4$ cuando el costo del incremento es utilizar un ancho de banda de canal adicional.

En conclusión se demostró que el uso eficiente de la potencia transmitida en un canal Rayleigh puede lograrse mediante alguna forma de diversidad para proveer al receptor con varias señales con desvanecimiento independiente todas transportando la misma información. Los tipos de diversidad discutidos (tiempo, frecuencia, etc.) constituyen una forma de codificación de canal usualmente denominada *codificación por repetición*. En el Capítulo siguiente se describirán técnicas de codificación de canal que son considerablemente más eficiente en términos de ancho de banda que la codificación de repetición y son generalmente más atractivas para comunicaciones digitales a través de un canal Rayleigh.