Capítulo 1

Introducción

La transmisión digital de información introduce una cantidad extraordinaria de ventajas como para dominar crecientemente los sistemas de comunicaciones. En comunicaciones entre computadoras, la información a ser transportada es inherentemente digital, de forma que la transmisión digital es la única alternativa práctica. Pero las comunicaciones entre computadoras son aún solo una pequeña fracción de todas las comunicaciones digitales. Una fracción mucho mayor está orientada a la transmisión de señales inherentemente analógicas, en particular voz e imágenes. Tales señales puede ser transmitidas (y tradicionalmente lo han sido) en forma analógica. ¿Porqué deberían transmitirse en forma digital? No hace mucho tiempo la transmisión digital era considerada ineficiente en términos de ancho de banda y el costo de conversión de analógico a digital y visceversa era un problema central. De cualquier forma cuatro cosas han sucedido para cambiar estos preconceptos:

- La codificación de señales analógicas en forma digital se ha beneficiado de los avances de los algoritmos de compresión, los cuales reducen dramaticamente la velocidad de transmisión requerida para representar una señal de voz o video con alta calidad subjetiva.
- Las técnicas de procesamiento de señal y codificación han incrementado dramaticamente la velocidad de transmisión que es posible alcanzar con un canal físico determinado.
- Los circuitos integrados han reducido considerablemente el costo de realizar funciones complejas de procesamiento de señal y codificación asociadas a la transmisión digital.
- La fibra óptica ha reducido el costo de transmitir a altas velocidades de transmisión sobre largas distancias.

El resultado de esos desarrollos ha sido cambiar la forma de imaginar los sistemas de comunicaciones. Hoy en dia el mayor énfasis asociado a la transmisión digital es frecuentemente el *ancho de banda reducido*, o equivalentemente la mayor capacidad total del sistema alcanzable con la transmisión digital. Por ejemplo, la telefonía celular y la televisión por cable están en proceso de convertirse a transmisión digital, en parte sobre la base de la mayor capacidad del sistema.

1.1 Aplicaciones de comunicaciones digitales

Las comunicaciones digitales son utilizadas por señales que son inherentemente analógicas y contínuas en el tiempo, tales como la voz y las imágenes, y señales que son inherentemente digitales, tales como archivos de texto. Los requerimientos impuestos sobre un sistema de comunicaciones son diferentes en cada caso. Además, las redes de comunicaciones modernas proveen una mezcla de servicios para los que se debe tener en cuenta las demandas de cada tipo de señal y/o sistema.

1.1.1 Transmisión digital de voz y video

Es común en la práctica convertir señales analógicas contínuas en el tiempo a la forma digital para transmisión. Una técnica utilizada comunmente para la transmisión de voz y video es *modulación de pulsos codificada* (PCM), como ilustrado en la figura 1.1. La señal de voz o video contínua en tiempo se muestrea

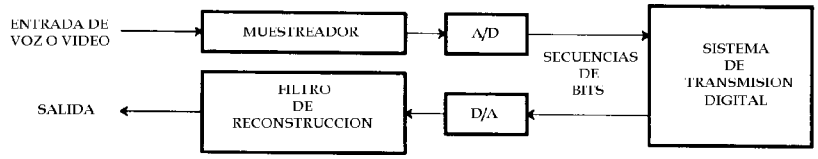

Figura 1.1: Un sistema de transmisión digital utilizado para transmisión PCM.

primero a una taza de f_s Hz (o muestras/seg.), la cual debe ser mayor que dos veces la mayor componente de frecuencia de la señal. Frecuentemente esta operación de muestreo está precedida de un filtro pasabajos, no mostrado, que asegura que la señal es limitada en frecuencia. Cada muestra se convierte luego a una palabra binaria de n bits por medio de un conversor analógico-digital (A/D). La salida es una secuencia de bits con una velocidad de nf_s bits/seg. La secuencia de bits se transmite entonces con un sistema de transmisión digital, y la señal de voz o video se reconstruye con un conversor digital-analógico (D/A) y un filtro de reconstrucción. La técnica de PCM fué patentada en Francia por Reeves al final de los años 30, y fué de las primeras utilizadas en la Segunda Guerra Mundial.

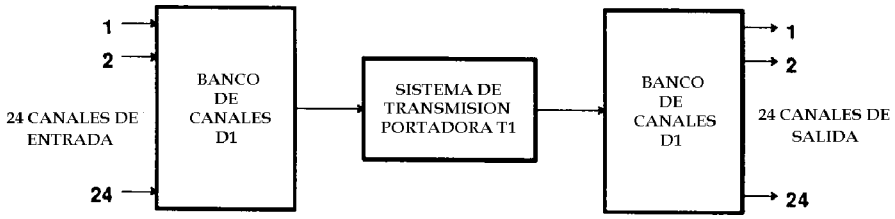

Figura 1.2: El sistema de transmisión T1 con su banco de canales D1 asociado para muestreo, conversión A/D y multiplexado.

Ejemplo: El primer sistema comercial PCM fué el *sistema T1* (americano), ilustrado en la figura 1.2. Su diseño fué completado en los Laboratorios Bell en 1962. Se utiliza aún ampliamente para la transmisión de corta distancia entre centrales de conmutación telefónica en areas metropolitanas. El sistema T1 transmite una secuencia de bits a 1.544 Mb/seg sobre dos pares de cobre, uno por cada dirección de transmisión. De esa capacidad, 1.536 Mb/seg están disponibles para la transmisión de 24 canales de voz con el restante utilizado para funciones de supervisión. Cada canal de voz se muestrea a 8000 muestras/seg y se cuantiza con 8 bits/muestra, lo que determina una velocidad total de 64 kb/seg por canal de voz. El banco de canales también multiplexa o combina las 24 secuencias en conjunto para formar una secuencia de 1536 Mb/seg.

La implementación interna de ese sistema de transmisión digital puede ser complicada. Sin embargo, desde el punto de vista de la transmisión de voz y video, el sistema puede caracterizarse completamente por tan solo cuatro parámetros:

- La *velocidad de transmisión* (bits/seg).
- Los retardos de *propagación* y *procesamiento*.
- La *probabilidad de error* de detección.
- El *error de fase (jitter)*.

La probabilidad de error y el error de fase pueden causar alguna degradación en la calidad de la señal de voz o video recuperada, y un retardo excesivo puede impedir una conversación, de forma que el diseñador

debe controlar esos problemas. Sin embargo, estos problemas son mucho menos complicados que los tipos de distorsión encontrados comunmente en la transmisión analógica.

Otra aplicación de las técnicas de transmisión digital son los *sistemas de almacenamiento o grabación* utilizando medios ópticos y/o magnéticos. En este caso el objetivo no es la transmisión *de un lugar a otro* sino *del momento de grabación al de reproducción.* Esos medios de almacenamiento tienen en general problemas diferentes de los medios de transmisión, pero algunos de los mismos conceptos básicos pueden aplicarse.

1.1.2 Comunicaciones entre computadores

Los datos generados por computadoras para transmisión a otras computadoras o terminales no requieren la conversión mostrada en la figura 1.1. Para la comunicación entre computadoras es común utilizar el término *transmisión de datos* antes que *transmisión digital.* Sin embargo es importante enfatizar que puede utilizarse el mismo sistema de transmisión para PCM y transmisión de datos. En realidad, un objetivo importante de varias alternativas de diseño actuales es combinar voz, video y datos dentro de un ambiente de transmisión *integrada.* Por esa razón nos referiremos al conjunto como *comunicaciones digitales.*

Un sistema de comunicaciones digitales utilizado para la transmisión de datos está caracterizado por su velocidad de transmisión, retardo, probabilidad de error y error de fase, de la misma forma que un sistema utilizado para transmisión PCM. La importancia relativa de esos problemas es sin embargo en general diferente. La transmisión de datos es usualmente mucho más sensible a los errores de transmisión que las señales de voz o video, pero por otro lado es relativamente insensible a los errores de fase. También, los datos requieren solo comunicación esporádica antes que una comunicación contínua. Por ejemplo, un usuario situado en un terminal puede desear que los caracteres individuales sean transmitidos a 9.6 kb/seg, pero los caracteres generados poco frecuentemente requerirán una velocidad de transmisión real de solo 50 bits/seg. Es posible caracterizar esta propiedad más formalmente como una gran diferencia entre las velocidades máxima y promedio.

1.1.3 Redes de telecomunicaciones

Las aplicaciones prácticas requieren más que un simple sistema de comunicaciones digitales enlazado punto a punto. Una *red de comunicaciones digitales* conecta simultaneamente varios usuarios entre sí, y frecuentemente incluye *conmutación*, lo cual permite a los usuarios iniciar conexiones específicas a otros usuarios. Por razones históricas, las redes diseñadas para esos propósitos diferentes tienen terminología diferente: una *red de telefonía* se diseña para manejar transmisión PCM, y una *red de datos* o *red de computadoras* se diseña para transmitir datos. Sin embargo, la terminología se está volviendo rapidamente obsoleta a partir de la introducción de redes diseñadas para todo propósito. Tales redes las llamaremos *redes de telecomunicaciones.* Un buen ejemplo es la *Red Integrada de Servicios Digitales* (ISDN) la cual está siendo instalada actualmente a escala mundial.

Las redes de telecomunicaciones se clasifican frecuentemente de acuerdo a su extensión geográfica. Las *Redes de Area Local* (LAN) están diseñadas para comunicar a velocidades de transmisión muy altas sobre areas geográficas pequeñas (del orden de un km de extensión). La mayoría de las LAN se diseñan primariamente para la comunicación entre computadoras. Las *Redes de Area Metropolitana* (MAN) pueden conceptualizarse en función de una mayor area que su contraparte LAN, y pueden cubrir un area de aproximadamente 50 km utilizando fibras ópticas. Las *Redes de Area Amplia* (WAN), extendidas tanto como la superfície de un país o el globo, utilizan una combinación de cable, radio microondas terrestre, cable submarino, fibra y satelite para cubrir largas distancias. La red telefónica mundial es un ejemplo. Otro ejemplo es *internet*, la cual está orientada principalmente a comunicación entre computadoras.

Las redes que transportan secuencias de bits múltiples también incluyen alguna forma de conmutación. La conmutación posibilita la reconfiguración de conexiones punto a punto en la red y usualmente toma una de dos formas básicas: *conmutación de circuitos* o *conmutación de paquetes.* En la conmutación de circuitos, la red conecta una secuencia de bits de velocidad constante a un destino por un período de tiempo relativamente largo (del orden de minutos o mayor). Este modo aparece en el contexto de redes de voz, donde el circuito se forma para la duración de un llamado telefónico. En la conmutación de paquetes, los datos se encapsulan en paquetes de bits relativamente cortos y se les agrega una dirección de destino. Los paquetes pueden rutearse luego por la red. Una ventaja de la conmutación de paquetes es que la secuencia de bits entre fuente y destino puede tener una velocidad de transmisión variable, esto se consigue generando solo los paquetes necesarios. Esto conduce a una mayor eficiencia para fuentes de datos que tienen una gran dispersión entre las velocidades de transmisión máxima y promedio.

1.2 Comunicaciones analógicas versus comunicaciones digitales

Para comunicaciones de datos no existe otra alternativa práctica que la transmisión digital. Sin embargo, para señales de voz y video existen importantes ventajas y desventajas que caracterizan a la transmisión digital.

Abstracción de la interface: La caracterización simple de un sistema de comunicaciones digitales es una ventaja sobre los sistemas analógicos, donde existen muchas más formas por medio de las cuales la transmisión puede degradarse. Por ejemplo, la relación señal a ruido puede ser pobre, o la señal puede sufrir distorsión de intermodulación de segundo o tercer orden, o interferencia cruzada (crosstalk) entre un usuario y otro, o el sistema puede recortar (saturar) la señal de entrada. En contraste, un sistema de comunicaciones digitales tiene solo tres parámetros: velocidad de transmisión, probabilidad de error de detección y error de fase. Los inconvenientes asociados al medio físico, los cuales pueden ser muy severos, quedarán notablemente ocultos al usuario. Esta propiedad se denomina *abstracción de la interface* del sistema de transmisión. El poder de esta abstracción fué tal vez apreciado por primera vez por C. Shannon en 1948 lo que dió comienzo al campo de la Teoría de la Información. Él demostró que teoricamente no existe ninguna pérdida por definir como una secuencia de bits a la interface entre la señal a ser trasmitida y el sistema de transmisión, tanto para una señal analógica como digital.

Figura 1.3: Una cadena de repetidores regenerativos reduce el efecto de las degradaciones de la cascada regenerando la señal digital en puntos intermedios de la transmisión.

El efecto regenerativo: Consideremos el problema de transportar una secuencia de bits a una gran distancia. La degradación introducida por un medio físico ininterrumpido, tal como un cable o fibra óptica, puede ser inaceptable. La solución digital es colocar a intervalos periódicos en el medio físico *repetidores regenerativos*, como se muestra en la figura 1.3. Cada uno de esos repetidores incluye un receptor, que detecta la secuencia de bits tal como se haría en el destino eventual, y un retransmisor similar al del orígen. Cualquier efecto debido a ruido y distorsión sobre el medio físico han sido *removidos completamente* por el repetidor regenerativo (excepto tal vez por los errores introducidos). En contraste con esto en la transmisión analógica, los repetidores equivalentes consisten basicamente en amplificadores, y el ruido y la distorsión se acumulan cuando la distancia se incrementa.

En la práctica, los sistemas de comunicaciones digitales introducirán algunos errores en la detección de la secuencia de bits. Cuando la probabilidad de error en cada repetidor es baja, la probabilidad total de error para m repetidores es aproximadamente m veces la probabilidad de un solo repetidor. Suponiendo un número máximo de repetidores es posible obtener el requerimiento de desempeño de un único repetidor de la probabilidad total de error objetivo. Es posible lograr entonces el objetivo ajustando el diseño del repetidor, el diseño de la señal, y el espaciamiento entre repetidores.

El mismo efecto regenerativo se aplica para señales de almacenamiento. Consideremos por ejemplo la más alta calidad posible con radio digital. Una justificación de la regeneración en este caso es que cada vez que la señal de audio se copia en un nuevo medio (por ejemplo de una cinta magnética a un disco laser), la señal se regenera, y las degradaciones específicas del medio original son removidas.

Factores económicos: Hemos visto que las comunicaciones digitales tienen poderoras ventajas técnicas. Cuales son los factores económicos relativos de las comunicaciones analógicas y las digitales? Existe un precio a ser pagado por las ventajas técnicas. Sin llegar a un excesivo detalle, una comparación cualitativa puede ser la siguiente:

- El multiplexado y conmutación de señales digitales es de mucho menor costo que para señales analógicas. Esto es particularmente cierto para el multiplexado porque el multiplexado por división en frecuencia de señales analógicas requiere un filtrado complicado.

- Algunos medios económicos, tal como fibra óptica y discos laser, son más adecuados para la transmisión digital que a la analógica.
- Las comunicaciones digitales de señales analógicas tal como voz y video requiere un paso adicional de muestreo y conversión analógico-digital. El costo de esta conversión fué inicialmente un impedimento para la difusión del uso de las comunicaciones digitales, pero con tecnologías de circuitos integrados este costo se está volviendo rapidamente insignificante, particularmente para señales de voz.
- Los repetidores regenerativos para comunicacones digitales son considerablemente más complicados que su contraparte analógica (los cuales son solo amplificadores). Sin embargo, la capacidad de esos sistemas es tan grande que el costo adicionado es insignificante para los usuarios individuales.
- Con las modernas tecnologías de compresión y transmisión, la transmisión PCM de señales analógicas puede lograrse con menor ancho de banda que la transmisión analógica de la misma señal. Esta característica es crítica en la transmisión por radio dado que el espectro de radiofrecuencias es un recurso caro. Esa transmisión tiene importantes ventajas económicas en otros medios, tal como una línea de abonado telefónico o televisión por cable coaxil.
- Las comunicaciones digitales introducen aspectos complicados de sincronización que son notablemente evitados en las comunicaciones analógicas.

Como conclusión es posible decir que tomó un tiempo, alrededor de 20 años, para que las comunicaciones digitales substituyeran casi completamente a sus competidoras analógicas, pero esa revolución está ahora practicamente completada. Este es el resultado de una combinación de factores económicos, avances tecnológicos y demandas de nuevos servicios.

1.3 Elementos de un sistema de comunicaciones digitales

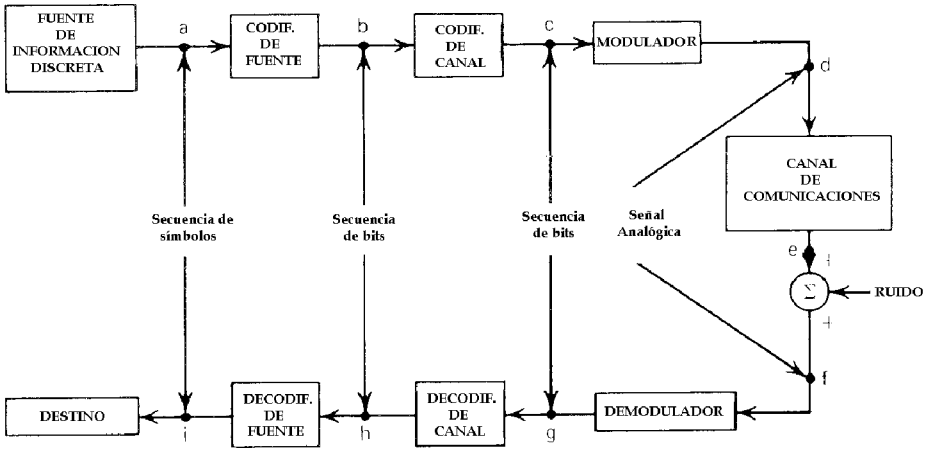

Figura 1.4: Bloques funcionales de un sistema de comunicaciones digitales.

La figura 1.4 esquematiza los elementos funcionales de un sistema de comunicaciones digitales. El propósito del sistema es transmitir los mensages (o secuencias de símbolos) provenientes de la fuente a un punto de destino con la mayor velocidad y precisión posibles. La fuente y el destino están separados fisicamente en el espacio (y/o tiempo) y un canal de comunicaciones los conecta de alguna forma. Los canales aceptan señales eléctricas o electromagnéticas, y su salida es usualmente una versión distorsionada de la entrada debido a la naturaleza no ideal del medio físico que constituye el canal. Además de la dispersión, la señal que transporta información también está corrompida por señales eléctricas impredecibles (ruido) debido a causas naturales o artificiales. La dispersión y el ruido introducen errores en la información que está siendo transmitida y limitan la velocidad a la cual la información puede transportarse. La probabilidad de decodificar incorrectamente un símbolo de mensaje en el receptor se utiliza frecuentemente como una medida del desempeño de los sistemas de comunicaciones digitales. La función principal del codificador, el

modulador, el demodulador y el decodificador es combatir los efectos degradantes del canal sobre la señal y maximizar la velocidad de transporte de la información y su precisión. Con estos aspectos preliminares en mente es posible dar un vistazo general a cada uno de los bloques funcionales del sistema

Fuente de información

Las fuentes de información pueden clasificarse en dos categorias en base a la naturaleza de sus salidas: fuentes analógicas y fuentes discretas. Las fuentes de información analógica, tal como un microfono utilizado para voz, una camara de TV registrando una escena, emiten una o más señales (o funciones del tiempo) de amplitud contínua. La salida de una fuente de información discreta, tal como la de una computadora personal, consiste en una secuencia de símbolos discretos o letras. Una fuente analógica puede transformarse en una discreta a través del proceso de muestreo y cuantización. Las fuentes de información discretas se caracterizan mediante alguno de los siguientes parámetros:

1. Alfabeto de la fuente (símbolos o letras).
2. Velocidad de transmisión.
3. Probabilidad de cada símbolo.
4. Dependencia probabilística de los símbolos.

A partir de estos parámetros es posible construir un modelo probabilístico de la fuente de información y definir la entropía de la fuente (H) y la velocidad de transmisión (R) en bits/símbolo y bits/seg respectivamente.
Para introducir el significado de esas cantidades consideremos una fuente de información discreta primigenia: una teletipo con las 28 letras del alfabeto Español más cuatro caracteres especiales. El alfabeto de la fuente para este ejemplo consiste en 32 símbolos. La velocidad de transmisión se refiere a la velocidad a la cual la teletipo produce caracteres: a los propósitos de esta discusión es posible suponer que la teletipo genera 10 símbolos/seg. Si la teletipo está produciendo secuencias de símbolos en Español es posible deducir que la ocurrencia de una letra particular en la secuencia dependerá de las letras que la preceden. Por ejemplo, la letra E ocurrirá más frecuentemente que la letra Q y la ocurrencia de Q implica que la próxima letra en la secuencia será muy probablemente la letra U, etc. Estas propiedades estructurales de las secuencias de símbolos puede caracterizarse por la probabilidad de ocurrencia de cada símbolo en particular y por las probabilidades condicionales de ocurrencia de los símbolos.
Una propiedad importante de una fuente discreta es su entropía. La entropía de una fuente H, es la información promedio contenida por símbolo en una secuencia de mensaje y tiene unidades de bits/símbolo. En el ejemplo, si se supone que todos los símbolos ocurren con igual probabilidad y en forma estadisticamente independiente en una secuencia de mensaje, entonces la entropía de la fuente es de 5 bits/símbolo (o sea, $2^5 = 32$). Sin embargo, la dependencia estadística de los símbolos en una secuencia y las probabilidades diferentes para cada símbolo reducen considerablemente el contenido de información promedio de los símbolos. Intuitivamente es posible justificar esto último considerando que para la secuencia QUE la letra U transporta poca información debido a que la ocurrencia de la Q implica que la próxima letra en la secuencia debía ser la U.
La velocidad de información de la fuente se define como el producto de la entropía de la fuente y la velocidad de símbolos con unidades de bits/seg. La velocidad de información R representa el número mínimo de bits/seg necesarios en promedio para representar la información proveniente de la fuente discreta. Alternativamente, R representa la velocidad de datos promedio mínima para transportar la información de la fuente al destino.

Codificador / decodificador de fuente

La entrada al codificador de fuente es una secuencia de símbolos con una velocidad de r_s símbolos/seg. El codificador de fuente convierte la secuencia de símbolos en una secuencia binaria de ceros y unos asignado palabras de código a los símbolos de la secuencia de entrada. El camino más simple en que puede llevarse a cabo esta operación es asignar una palabra de código binario de longitud fija a cada símbolo de la secuencia de entrada. Para el ejemplo de la teletipo esto puede hacerse asignando palabras de código de 5 bits de 00000 a 11111 para los 32 símbolos del alfabeto de la fuente y reemplazar cada símbolo en la secuencia de entrada por su palabra de código preasignada. Con una velocidad de 10 símbolos/seg., la velocidad de salida del codificador de fuente será de 50 bits/seg.

La codificación de longitud fija de los símbolos individuales de la salida de la fuente es eficiente solo si los símbolos ocurren con igual probabilidad en una secuencia estadísticamente independiente. En la mayoría de las aplicaciones prácticas los símbolos en una secuencia son estadisticamente dependientes y ocurren con probabilidades diferentes. En esas situaciones el codificador de fuente toma una subsecuencia de dos o más símbolos como un bloque y le asigna palabras de código de longitud variable a cada bloque. El codificador de fuente óptimo se diseña para producir una velocidad de salida que se aproxime a R. Debido a restricciones prácticas, la velocidad de salida del codificador de fuente será mayor que R. Los parámetros importantes de un codificador de fuente son el tamaño del bloque, las longitudes de las palabras de código, la velocidad promedio de salida, y la eficiencia del codificador (o sea, la velocidad de salida comparada con el mínimo lograble R).
En el receptor, el decodificador de fuente convierte la salida binaria del decodificador del canal en una secuencia de símbolos. El decodificador para un sistema que utiliza palabras de código de longitud fija es simple, pero el que utiliza palabras de longitud variable puede ser complejo. Los decodificadores en esos sistemas deben ser capaces de afrontar cierto número de problemas tales como los requerimientos crecientes de memoria y la pérdida de sincronismo debido a errores de transmisión.

Canal

El canal de comunicaciones provee la conexión eléctrica entre la fuente y el destino. El canal puede ser un par trensado, cable coaxil, el espacio libre o fibra óptica. Debido a limitaciones físicas los canales de comunicaciones tienen un ancho de banda finito de B Hz (excepto el caso de fibra óptica, como se discutirá más adelante), y la señal que transporta información cuando viaja por el canal sufre frecuentemente de distorsión de amplitud y fase. Además de la distorsión, la potencia de la señal también disminuye debido a la atenuación del canal. Además, la señal está corrompida por señales eléctricas impredecibles e indeseadas sintetizadas como ruido. Mientras algunos de los efectos degradantes del canal pueden ser removidos o compensados, los efectos del ruido no pueden eliminarse completamente. Desde este punto de vista, el objetivo primario del diseño de un sistema de comunicaciones debería ser la reducción de los efectos nocivos del ruido tanto cuanto sea posible.
Una de los formas para minimizar los efectos del ruido es incrementar la potencia de la señal. Sin embargo, esa potencia no puede incrementarse más allá de ciertos niveles debido a los efectos no lineales dominantes cuando la amplitud de la señal se incrementa. Por esta razón, la relación de señal a ruido (SNR) que puede mantenerse a la salida del canal es un parámetro importante del sistema. Otros parámetros importantes del canal son el ancho de banda utilizable (B), la respuesta de amplitud y fase, y las características estadísticas del ruido.
Si los parámetros del canal de comunicaciones son conocidos, es posible calcular la *capacidad del canal* C, que representa la máxima velocidad de transmisión a la cual es teoricamente posible transmisitir sin errores. Para ciertos tipos de canales de comunicación se ha demostrado que C es igual a $B\log_2(1+SNR)$ bits/seg. La capacidad del canal debe ser mayor que la velocidad de transmisión promedio R de la fuente para transmisión sin errores.

Modulador

El modulador acepta como entrada un flujo de bits y lo convierte en una señal eléctrica adecuada para la transmisión sobre el canal de comunicaciones. La modulación es una de las más poderosas herramientas al alcance del diseñador de sistemas de comunicaciones. Pueden utilizarse efectivamente para minimizar los efectos del ruido en el canal, para adecuar el espectro de frecuencias de la señal transmitida a las características del canal, para proveer la capacidad de multiplexado de varias señales y para superar varias limitaciones de diferentes aspectos de un enlace.
Los parámetros importantes de un modulador son los tipos de formas de onda utilizados, la duración de las formas de onda, el nivel de potencia, y el ancho de banda utilizado. El modulador lleva a cabo estas tareas minimizando los efectos del ruido mediante el uso de gran potencia de señal y ancho de banda, y utilizando las formas de onda de más larga duración. Si bien la utilización de potencia de señal grande y ancho de banda es un método obvio, esos parámetros no pueden incrementarse indefinidamente debido a limitaciones del equipamiento. La utilización de formas de onda de duración más larga para minimizar los efectos del ruido se basa en la conocida *ley de los grandes números* (concepto estadístico a ser discutido en el Capítulo 2), que determina que en términos muy generales, mientras la realización de un experimento aleatório único

puede variar notablemente, el resultado de varias repeticiones del experimento permite mejores predicciones. En comunicaciones digitales, este principio puede utilizarse con ventajas haciendo la duración de las formas de onda de señalización largas. Promediando sobre un tiempo de duración más largo, los efectos del ruido pueden minimizarse.
Para ilustrar este principio, supongamos que la entrada al modulador consiste en ceros y unos ocurriendo a 1 bit/seg. El modulador puede asignar señales una vez cada segundo. Por ejemplo, a las señales $A\cos\omega_1 t$ $(0 \leq t < 1)$ y $A\cos\omega_2 t$ $(0 \leq t < 1)$ puede asignarseles los bits 0 y 1, respectivamente. Notar que la información contenida en el bit de entrada está contenida ahora en la frecuencia de la señal de salida. Para emplear señales de mayor duración, el modulador puede asignar señales una vez cada 4 segundos. Esto requerirá 16 señales, $A\cos\omega_1 t$, $A\cos\omega_2 t$, $\cdots$, $A\cos\omega_{16} t$, cada una de duración 4 segundos; esas 16 señales pueden representar las 16 combinaciones de 4 bits 0000, 0001, $\cdots$, 1111. La limitación de esta técnica es evidente. El número de señales distintas que el modulador tiene que generar (y de esa forma el número de señales que se deberá detectar) aumenta exponencialmente cuando se incrementa la duración. Esto conduce a un incremento en la complejidad del equipamiento y en consecuencia la duración no puede incrementarse indefinidamente.

Demodulador

La modulación debe ser un proceso reversible, y la extracción del mensage de la señal que transporta información producida por el modulador se consigue con el demodulador. Para un tipo específico de modulación, el parámetro más importante del demodulador es el método de demodulación. Existe una gran variedad de técnicas disponibles para demodulación, el procedimiento utilizado determina la complejidad requerida en el equipamiento de recepción y la precisión de la demodulación. Dados el tipo y duración de las señales utilizadas en el modulador, el nivel de potencia en el modulador, las carácterísticas físicas del canal y del ruido, y el tipo de demodulación, es posible obtener relaciones únicas entre la velocidad de transmisión, requerimientos de potencia y ancho de banda, y la probabilidad de decodificar correctamente los bits de información. Una parte considerable de este texto está orientada a la obtención de esas importantes relaciones y su utilización en el diseño de sistemas.
Las características del modulador, el demodulador y el canal establecen una probabilidad promedio de error entre los puntos c y g de la figura 1.4. Frecuentemente la probabilidad de error será mayor que la deseada. Una probabilidad de error menor puede lograrse rediseñando el modulador y el demodulador o utilizando *codificación para control de errores* o codificación de canal.

Codificador / decodificador de canal

La codificación de canal es una metodología práctica para obtener alta confiabilidad y eficiencia de transmisión que solo pueden lograrse de otra manera mediante la utilización de señales de mayor duración en el proceso de modulación/demodulación. A partir de la codificación de fuente, a un grupo de señales analógicas se asocia cierto número de símbolos para transmisión por el canal y el demodulador tiene la tarea de distinguir entre ellas. La codificación de canal, con el propósito de control de errores, se logra adicionando bits extra a la salida del codificador de fuente. A pesar que esos bits extra no transportan información, ellos permiten que el detector encuentre y/o corrija algunos de los errores asociados a los bits que transportan información.

Existen, en principio, dos métodos para llevar a cabo la operación de codificación de canal. En el primer método, denominado método de *codificación en bloques*, el codificador toma un bloque de k bits de información del codificador de fuente y le adiciona r bits de control de error. El número de bits de control de error dependerá del valor de k y de las capacidades de control de error deseadas. En el segundo método, denominado método de *codificación convolucional*, la secuencia de mensaje se codifica en forma contínua entrelazando bits de información y bits de control de error continuamente. Ambos métodos requieren alguna capacidad de almacenamiento y procesamiento de los datos binarios en el transmisor y el receptor.
Los parámetros importantes en el codificador de canal son el método de codificación, la velocidad o eficiencia del codificador (medida como la relación de la velocidad de salida del codificador de fuente y la velocidad de salida del codificador de canal), la capacidad de control de error y la complejidad del codificador.
El decodificador de canal recupera los bits que transportan información a partir de la secuencia binaria codificada. La posible detección y corrección de errores también se realiza a través del decodificador de canal. El decodificador opera o en modo de bloques o en modo secuencial dependiendo del tipo de codificador

utilizado. La complejidad del decodificador y el retardo de tiempo involucrado en la decodificación son los parámetros de diseño importantes.

1.4 Aspectos del modelo de un receptor digital

Las comunicaciones en general serán esencialmente digitales en unos pocos años. La razón para este desarrollo es el progreso realizado en microelectrónica que permite a su vez implementar algoritmos complejos economicamente para lograr velocidades de transmisión próximas a los límites teóricos.

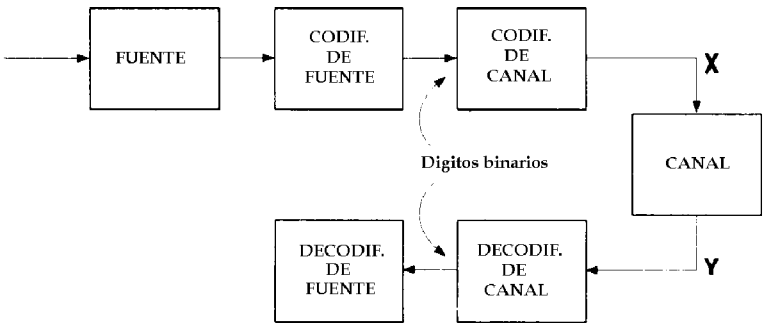

Figura 1.5: Modelo de comunicaciones idealizado basado en Teoría de Información.

El modelo de comunicaciones estudiado en la *Teoría de Información* (Capítulos 4 y 8) que permite cuantificar esos límites se muestra en la figura 1.5. La Teoría de información responde a dos cuestiones fundamentales:

- ¿Cual es la más alta taza de compresión? (respuesta: la entropía H).
- ¿Cual es la máxima velocidad de transmisión? (respuesta: la capacidad del canal C).

La Teoría de Información está asociada a secuencias. Los símbolos de algún alfabeto se mapean en secuencias de símbolos $\boldsymbol{x} = (x_1, \cdots, x_n, \cdots)$ tal que el canal produce la secuencia de salida $\boldsymbol{y} = (y_1, \cdots, y_n, \cdots)$. La secuencia de salida es aleatória, pero tiene una distribución que depende de la secuencia de entrada. De la secuencia de salida intentamos recuperar el mensaje transmitido.

En cualquier sistema de comunicaciones físico se transmite una señal contínua $s(t, \boldsymbol{x})$ asociada a la secuencia $\boldsymbol{x}$ transmitida y no necesariamente la secuencia en si misma. La asignación de un símbolo de la secuencia $\boldsymbol{x}$ a la señal se realiza mediante un modulador. Además de depender de la secuencia $\boldsymbol{x}$, la forma de la señal depende de un conjunto de parámetros $\boldsymbol{\theta} = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_C)$. El subconjunto $\boldsymbol{\theta}_T$ está relacionado con el transmisor y $\boldsymbol{\theta}_C$ son los parámetros del canal. Estos parámetros son en general desconocidos en el receptor. Para poder recupar la secuencia de símbolos $\boldsymbol{x}$, el receptor deberá estimar estos parámetros de la señal recibida. Estas estimaciones se utilizan más tarde como si fueran los valores verdaderos. Aún cuando los parámetros no están estrictamente relacionados con unidades de tiempo hablamos de *detección sincronizada* en abuso del término preciso de sincronización.

En este modelo físico de comunicaciones estamos preocupados con un receptor *interno* y un receptor *externo*, como mostrado en la figura 1.6. La tarea del receptor interno es producir una secuencia $\boldsymbol{y}(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\theta}}_C)$, tal que el *canal sincronizado* tenga una capacidad próxima al límite teórico para ese canal (donde se asume en general que no hay parámetros desconocidos). El receptor externo tiene como tarea decodificar en forma óptima la secuencia transmitida.

En el caso más simple de un canal con ruido aditivo, los parámetros $\boldsymbol{\theta} = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_C)$ pueden suponerse constantes pero desconocidos. El conjunto de parámetros incluye, por ejemplo, la fase θ o una fracción de retardo de tiempo ϵ. En este caso el *estimador de canal* de la figura 1.6 tiene la tarea de estimar un conjunto de parámetros desconocidos en una señal conocida inmersa en ruido. El ajuste de los parámetros se realiza, por ejemplo, mediante el desplazamiento de la fase $\theta(t)$ de un oscilador controlado por tensión (VCO) tal que el error de estimación $\phi(t) = \theta(t) - \hat{\theta}(t)$ sea minimizado.

Este modelo de canal no es aplicable a las comunicaciones móviles, ya que en ese caso los canales son variantes en el tiempo. La tarea del *estimador de canal* consiste en estimar parámetros variantes en el

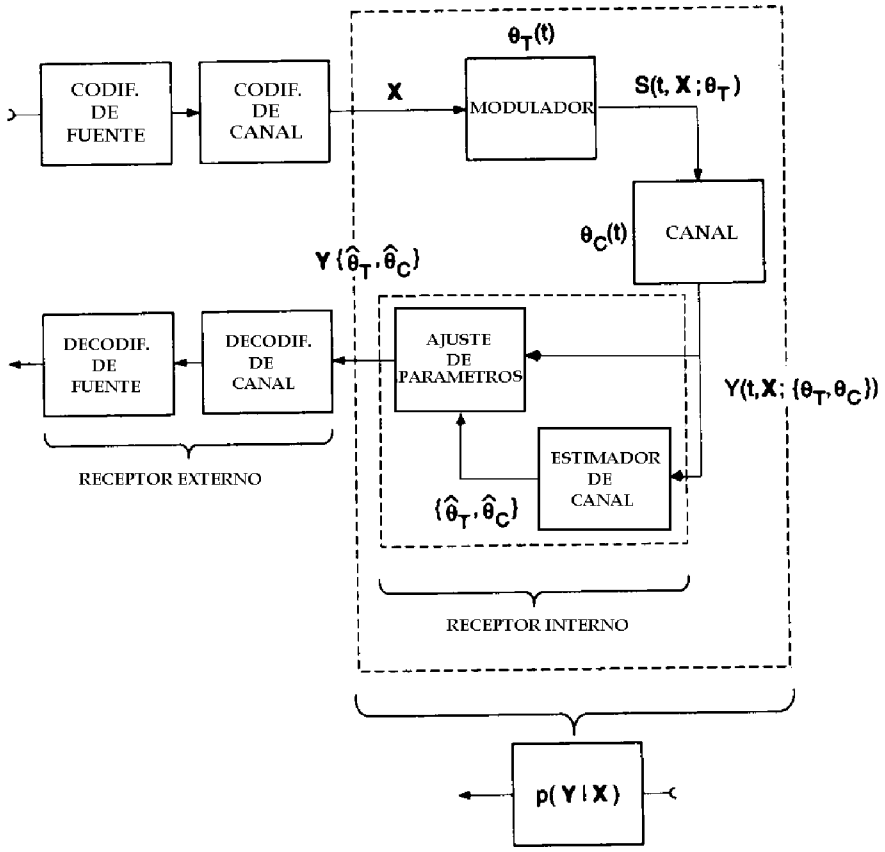

Figura 1.6: Modelo de comunicaciones físico.

tiempo. Matematicamente es necesario resolver el problema de estimar una señal aleatória (la respuesta impulsiva del canal) en un ambiente ruidoso. El bloque de *ajuste de parámetros* realiza tareas tales como el cómputo de los parámetros dependientes del canal: los coeficientes de un *filtro acoplado* o un *ecualizador*, para mencionar algunos ejemplos.

Se mencionó antes que la tarea del receptor interno es producir un buen canal para el decodificador. Para lograr este objetivo es necesario utilizar en forma completa el *contenido de información* de la señal recibida. Dado que estamos interesados en implementaciones digitales del receptor, esto implica que debemos hallar condiciones tales que esta información se obtenga a partir de muestras $\{y(t = kT_s, \boldsymbol{x}, \boldsymbol{\theta})\}$ de la señal contínua recibida (T_s es el período de muestreo). Es posible mostrar que existen condiciones tal que la *discretización temporal* no conduce a pérdida de información. Matematicamente decimos que las muestras $y(kT_s, \boldsymbol{x}, \boldsymbol{\theta})$ representan una *estadística suficiente*.

Es obvio que la *discretización de amplitud* debe seleccionarse de forma que la distorsión resultante sea despreciable. Esta selección requiere un balance cuidadoso entre dos objetivos conflictivos: aspectos relacionados con desempeño (lo cual requiere alta resolución) y complejidad de la implementación (lo cual crece fuertemente con el incremento del número de niveles de cuantización).

La figura 1.6 sugiere que es posible (facilmente) separar la estimación de los parámetros desconocidos $\boldsymbol{\theta}$ de la estimación de la secuencia de información $\boldsymbol{x}$. Desafortunadamente ese no es el caso en general.

En el caso más general la señal recibida $y(t, \boldsymbol{x}; \boldsymbol{\theta})$ contiene dos secuencias aleatórias $\boldsymbol{x}$ y $\boldsymbol{\theta}$ las cuales no pueden separarse trivialmente y deben estimarse en principio conjuntamente. Por razones de complejidad esto no es posible.

Una forma elegante de salir de este dilema es imponer una *estructura de bloques* (frames) sobre la capa física (en los niveles superiores del modelo OSI de redes de datos la información está siempre estructurada en frames). La estructura de frames permite separar la tarea de estimar los parámetros aleatórios del canal de la detección de la secuencia de datos. Debido a esta separación la complejidad de los algoritmos se reduce considerablemente. En varios casos prácticos es solo esta separación lo que permite algoritmos realizables.

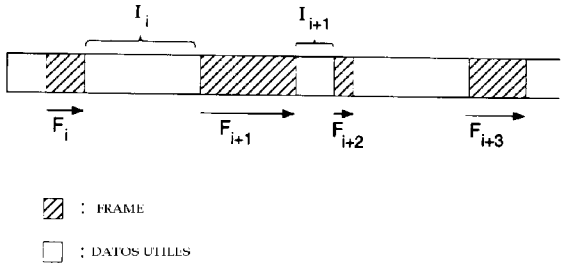

Figura 1.7: Estructura de frames.

La figura 1.7 ilustra el *principio de separación*. El flujo de bits está compuesto por símbolos de frame (sombreado) e información útil. Utilizando los símbolos de frame *conocidos* se estiman los parámetros del canal. Por otro lado, durante la transmisión de los símbolos dato, los parámetros del canal se suponen conocidos y se decodifican los bits de información. Es importante notar que la estructura de frames puede utilizarse para que las características del canal satisfagan dos requeriemientos conflictivos. Por un lado el número de símbolos de frame conocidos debe minimizarse dado que sino la eficiencia del canal disminuye. Por otro lado, la exactitud de la estimación de los parámetros del canal aumenta con el número de símbolos de frame.

Ejemplo 1: Canal invariante en el tiempo. En este caso no es necesaria la estructura de frames para la estimación de los parámetros del canal tal como la fase θ. La dependencia de los datos de la señal recibida puede eliminarse de varias formas.

Ejemplo 2: Canal variante en el tiempo (fading). Dado que el canal es variante en el tiempo, se requiere la estructura de frames. La forma exacta depende de las características del canal.

Es posible resumir las discusiones sobre modelos y límites como sigue. El principio de separación define dos tareas diferentes que se llevan a cabo con el receptor interno y el receptor externo, respectivamente. El *receptor externo* está asociado a un modelo idealizado. La tarea principal de diseño incluye el diseño de codificadores y decodificadores utilizando posiblemente información del canal obtenida con el receptor interno. El canal se supone conocido y se define con la probabilidad $p(\boldsymbol{y}|\boldsymbol{x})$. El *receptor interno* tiene como tarea producir una salida tal que el desempeño del receptor externo sea tan próximo como sea posible al ideal de conocimiento perfecto del canal. El receptor interno óptimo asume el conocimiento perfecto de la secuencia de símbolos transmitida en los segmentos de frame, ver figura 1.7 (una secuencia aleatória desconocida se considera un parámetro no deseado).

El receptor externo está en el domínio de la Teoría de la Información. La Teoría de la Información promete una transmisión libre de errores a una velocidad R_b (bits/seg) casi igual a la capacidad del canal $R_b \leq C$. En un sistema práctico es necesario aceptar una probabilidad de error de transmisión no nula. La probabilidad de error aceptable para una velocidad de transmisión R_b se logra para un ancho de banda B y una relación señal a ruido γ. Los valores numéricos de B y γ dependen de las características del sistema. Para permitir una comparación significativa entre diferentes técnicas de transmisión debemos definir apropiadamente medidas de desempeño normalizadas. Por ejemplo, no es muy significativo comparar la probabilidad de error como una función de la relación señal a ruido (SNR) a menos que esta comparación se realice sobre la base de un ancho de banda similar, o en forma equivalente, una velocidad de transmisión similar. La comparación más compacta está basada en la velocidad de transmisión normalizada $\nu = \frac{R_b}{B}$ (bits/seg/Hz) y la relación señal a ruido γ.

El parámetro ν es una medida de la *eficiencia espectral* dado que relaciona la velocidad de transmisión R_b medida en bits por segundo con el ancho de banda requerido medido en Hz.

El parámetro γ es una medida de *eficiencia de potencia* dado que especifica la potencia relativa requerida

para lograr una taza de error especificada. Ambas medidas son afectadas por la codificación y el receptor interno.

La medida de desempeño del receptor interno es la varianza de un estimador no sesgado. Puede mostrarse que la varianza de cualquier estimador es mayor que una cota fundamental conocida como *Límite de Cramer-Rao.*

El receptor interno ocasiona una pérdida de eficiencia espectral y de potencia Δ debido a la asignación de ancho de banda a la estructura de frames, y ademas una *pérdida de detección* Δ_D debida a una sincronización imperfecta. Δ_D se define como el incremento requerido en la SNR γ, asociado con la reconstrucción imperfecta del conjunto de parámetros $\boldsymbol{\theta}$ (ver figura 1.6) en el receptor en relación al conocimiento completo de esos parámetros para mantener una probabilidad de error especificada.

Tradicionalmente el ingeniero de comunicaciones se orienta al espacio de diseño comprendido por las dimensiones de potencia y ancho de banda. En este espacio de diseño es posible obtener soluciones de compromiso entre eficiencia espectral y eficiencia de potencia.

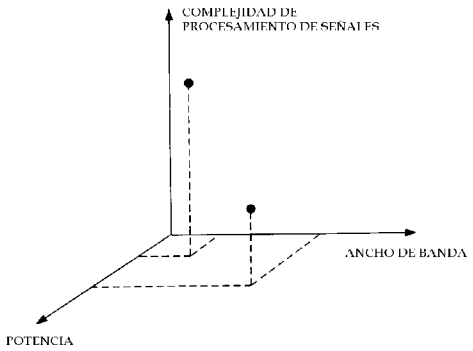

Figura 1.8: Espacio de diseño tridimensional.

El estado del arte de los sistemas de comunicaciones opera próximo a los limites teóricos. Esos sistemas se han vuelto posibles debido al enorme progreso realizado en microelectrónica, que a su vez ha permitido implementar complejos algoritmos de procesamiento digital de señales en forma económica. En consecuencia es posible incluir una tercera dimensión en el espacio de diseño: la complejidad del procesamiento de señales asociado (ver figura 1.8).

La ingeniería de comunicaciones actual tiene la opción de manejar medidas de desempeño físico (potencia y ancho de banda) junto con complejidad de procesamiento de señales. Este hecho real ha revolucionado toda el area de las comunicaciones.

Mientras es posible definir medidas de desempeño claras para el ancho de banda y la potencia, desafortunadamente no existe tal medida universal para la complejidad. Si los algoritmos funcionan sobre un procesador programable, entonces el número de operaciones por unidad de tiempo T (donde $1/T$ es la velocidad a la cual se obtiene el resultado del algoritmo) es un indicador significativo de complejidad. Si se emplea un ASIC (Application-Specific Integrated Circuit) una medida posible de desempeño es el producto AT, o sea la relación entre area de silicio A y la potencia de procesamiento (proporcional a $1/T$).

1.5 Discusión de contenidos

Los contenidos discutidos en los diferentes Capítulos siguen el modelo básico de un canal de comunicaciones digitales y las técnicas de procesamiento, modulación y codificación asociadas para lograr especificaciones de diseño prácticas. La secuencia de presentación contempla, a parte de los Capítulos introductorios, el tratamiento de modelos de canales progresivamente más completos y complejos.

De esta forma, en el Capítulo 2 se incluye una breve discusión de algunos conceptos básicos de señales y sistemas pasabanda y procesos estocásticos, a modo de revisión, aunque abarcando esencialmente todas las herramientas que serán utilizadas a lo largo del texto.

En el Capítulo 3 se incluye una discusión detallada de diversos tipos de canales de comunicaciones donde las técnicas digitales han tenido mayor éxito. El objetivo aquí, más allá de los detalles discutidos, es introducir una idea de la importancia relativa de los parámetros que describen cada canal real. A modo de síntesis, al final de ese Capítulo se incluyen los modelos básicos que se utilizarán en el texto.

La secuencia asociada al modelo básico del canal de comunicaciones comienza en el Capítulo 4, donde se discuten técnicas de codificación de la fuente de información, tanto para el caso sin pérdidas (transmisión de datos), como para el caso de codificación con pérdidas (fundamentalmente PCM). En este Capítulo se discuten límites fundamentales de desempeño de un sistema de comunicaciones, asociados a la representación eficiente de la fuente de información que obviamente influencian al resto del sistema.

En el Capítulo 5 se contempla la discusión de conceptos (señalización, detección) asociados a la forma más simple de modelo de canal. Se introduce algún énfasis particular en una técnica de señalización con memoria que permite introducir, en forma simplificada, varias metodologias asociadas (como el algoritmo de Viterbi).

En el Capítulo 6 se extienden los resultados del Capítulo anterior considerando ahora un canal de ancho de banda limitado y las restricciones que esto impone en un sistema de comunicaciones digitales. En particular se discute el critério de Nyquist para reducir el problema de interferencia intersímbolos y diferentes metodologías asociadas, fundamentalmente a la ecualización de un canal no ideal.

El Capítulo 7 contempla un modelo de canal pasabanda no idealizado que, en principio requiere técnicas de modulación pasabanda. En este caso se discuten técnicas clásicas de modulación en amplitud y ángulo, orientadas a satisfacer el compromiso de eficiencia de ancho de banda - consumo de potencia característico. En particular, se incluye una discusión asociada a canales con desvanecimiento y multicamino que permite analizar el problema de canales variantes en el tiempo. Esto se asocia a formas básicas de solucionar el problema de bajo desempeño en términos de probabilidad de error, en particular algunas técnicas de diversidad.

En el Capítulo 8 se complementa el estudio del desempeño obtenido para un canal específico con técnicas de codificación de canal. En esta discusión se completa la discusión de límites fundamentales de desempeño de un sistema de comunicaciones digitales iniciada en el Capítulo 4 (para obtener una medida de la capacidad de un canal de comunicaciones). La discusión general en este Capítulo está asociada a códigos en bloques, códigos convolucionales y una metodología que combina codificación y modulación.

Finalmente, en el Capítulo 9 se incluye un estudio de técnicas de modulación de espectro disperso, contemplando el caso de modelo de canal más general, o sea, no solo con ruido aditivo si no también con interferencias de tipo causal. La discusión incluye técnicas de espectro disperso de secuencia directa y técnicas de espectro de saltos en frecuencia y sus aplicaciones.

La diversidad y especificidad de herramientas, técnicas y metodologías asociadas a las diversas aplicaciones de las comunicaciones digitales, discutidas en este texto, no pretende ser exhaustiva, aunque si significativa para el apoyo al diseño y especificación de tales sistemas. Como mostrado en el detalle del Capítulo 3, cada canal de comunicaciones requiere técnicas muy particulares, y fundamentalmente nuevas técnologías asociadas a canales de comunicaciones seguramente requerirán cierta complementación. Un caso particular, de suma importancia, es el canales de fibra óptica.

Desde el punto de vista de acceso a la temática en general, se ha diferenciado en el índice los temas de lectura (indicados por (*)) de los temas de estudio específico.

Capítulo 2

Señales pasabanda y procesos estocásticos

2.1 Señales pasabanda y modulación

Las señales pasabanda son de fundamental importancia para comunicaciones digitales sobre canales, como por ejemplo el de radio, donde el espectro de señal debe estar confinado a una banda estrecha de frecuencias. En esta sección definiremos primero un bloque constructivo básico denominado *divisor de fase* para luego desarrollar una representación banda base compleja de cualquier señal pasabanda. Luego se describirán algunos conceptos de muestreo de señales pasabanda y finalmente algunas técnicas de modulación útiles para trasladar el espectro de frecuencia de la señal.

2.1.1 El divisor de fase y una señal analítica

Un *divisor de fase* es un filtro con respuesta impulsiva $\phi(t)$ y función transferencia $\Phi(j\omega)$ tal que

$$\Phi(j\omega) = \begin{cases} 1, & \omega \geq 0, \\ 0, & \omega < 0 \end{cases}$$

Este filtro deja pasar solo frecuencias positivas y rechaza frecuencias negativas. Claramente, como $\Phi(j\omega)$ no tiene simetría compleja conjugada entonces $\phi(t)$ es una respuesta impulsiva compleja. Independientemente de la entrada al divisor de fase, la salida debe tener solo componentes de frecuencia positivos. Una señal con componentes de frecuencia positivos se denomina *señal analítica*. Obviamente, cualquier señal analítica es compleja en el domínio tiempo.

Proximamente relacionada con el divisor de fase está la *Transformada de Hilbert*, un filtro con función transferencia

$$H(j\omega) = -j \operatorname{sign}(\omega)$$

Esta transformada tiene respuesta impulsiva real dado que su función transferencia tiene simetría compleja conjugada. La transformada de Hilbert no modifica la amplitud del espectro de la entrada sino que introduce un desplazamiento de fase de $-\pi$ en todas las frecuencias. Si la entrada a $H(j\omega)$ es $x(t)$, entonces la salida, o sea la transformada de Hilbert de $x(t)$, se notará por $\hat{x}(t)$.

Ejemplo: Si una entrada real al divisor de fase es $x(t)$, entonces la salida será: $1/2[x(t) + j\hat{x}(t)]$. De esta forma, la parte real de la salida está compuesta por la mitad de la entrada y la parte imaginaria por la mitad de la transformada de Hilbert de la entrada.

El divisor de fase y la transformada de Hilbert tienen una discontinuidad de amplitud y fase en cc. En consecuencia son muy dificiles de implementar para frecuencias de bandabase. Sin embargo en muchas

aplicaciones es posible aplicar el divisor de fase a una señal pasabanda lo cual permite una implementación mucho más simple debido a que la función transferencia no es importante en la región de cc.

2.1.2 Representación banda base compleja de una señal pasabanda

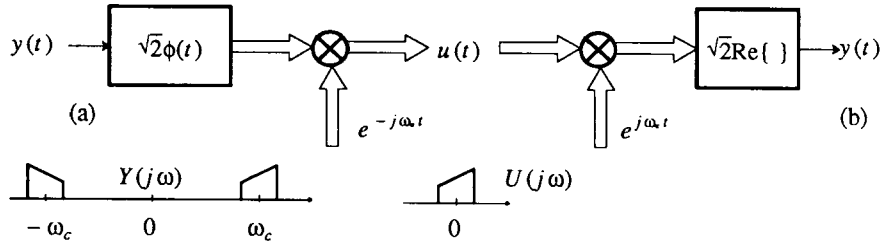

Figura 2.1: Obtención de la representación banda base compleja $u(t)$ a partir de una señal pasabanda $y(t)$. a) Obtención de $u(t)$ de $y(t)$. b) Recuperación de $y(t)$ de $u(t)$.

Supongamos que $y(t)$ es una seña pasabanda real con un espectro centrado en $\omega = \omega_c$. Es posible desarrollar una representación de $y(t)$ en función de una señal banda base compleja $u(t)$, o sea, una señal con su espectro centrado en cc. Consideremos el sistema mostrado en la figura 2.1a. Dado que $y(t)$ es real entonces tiene un espectro concentrado alrededor de $-\omega_c$ y de ω_c. Haciendo pasar $y(t)$ por un divisor de fase, la señal analítica de salida no tendrá términos correspondientes a frecuencias negativas. Los términos remanentes de frecuencias positivas pueden desplazarse a cc multiplicando por una exponencial compleja $e^{-j\omega_c t}$, lo cual produce la representación banda base compleja $u(t)$. Se incluye el factor de escala $\sqrt{2}$ para que ambas señales tengan la misma energía. Matematicamente, la señal compleja banda base puede expresarse por

$$u(t) = \frac{1}{\sqrt{2}}(y(t) + j\hat{y}(t))e^{-j\omega_c t}$$

Como se muestra en la figura 2.1b, la señal original pasabanda puede recuperarse de la representación banda base compleja a través de la ecuación

$$y(t) = \sqrt{2}\, Re\{u(t)e^{j\omega_c t}\} \tag{2.1}$$

Esto puede verificarse facilmente sustituyendo esta ecuación en la anterior. La ecuación de recuperación se denomina *representación canónica* de la señal pasabanda en términos de la señal banda base compleja. Cualquier señal pasabanda compleja puede representarse en esta forma canónica, donde $u(t)$ puede obtenerse como en la figura 2.1a.

2.1.3 Modulación

La forma frecuentemente útil para desplazar el espectro de una seňal se denomina *modulación.*

Ejemplo: Un canal telefónico deja pasar solamente frecuencias en el rango de 300 Hz a 3300 Hz. Cualquier señal transmitida sobre ese canal debe estar limitada en banda, o si no se introducirá distorsión. En forma similar, la radiodifusión AM comercial ocupa frecuencias desde 550 kHz a 1.6 MHz. Una señal de audio está limitada a frecuencias de audio por debajo de 20 kHz. La modulación es necesaria para trasladar la señal de audio a la banda de frecuencias útil para transmisión en AM.

Concretamente, la modulación es opuesta a la representación canónica, en lugar de obtener una representación banda base equivalente de una señal pasabanda, la modulación genera una representación pasabanda de una señal banda base. La representación canónica indica que una señal pasabanda real (necesaria

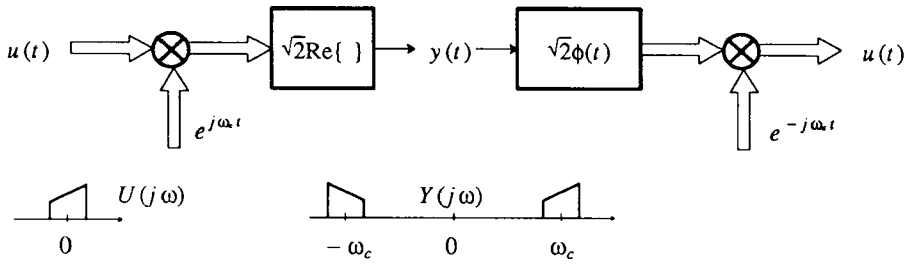

Figura 2.2: Un modulador convierte una señal banda base compleja $u(t)$ en una señal pasabanda real $y(t)$, simplemente revirtiendo la figura 2.1.

para ser transmitida por un medio físico) corresponde en general a una señal banda base compleja. De esta forma, suponemos que la señal banda base $u(t)$ es compleja y se genera la señal modulada de (2.1). El modulador y demodulador resultantes se muestran en la figura 2.2, la cual es el dual de la figura 2.1. Nuevamente, el factor $\sqrt{2}$ se introduce para que la señal modulada $y(t)$ tenga la misma energía que la señal banda base $u(t)$.

Esta representación de modulación es muy general. Todas las técnicas de modulación comunmente usadas pueden representarse en esta forma. Esas técnicas se distinguen por como mapean la señal que lleva información (moduladora) en la señal banda base compleja $u(t)$. Se ilustrará esto con tres técnicas de modulación importantes: AM-DBL, AM-BLU y QAM.

En *modulación de amplitud de doble banda lateral* (AM-DBL), una señal moduladora $a(t)$ se mapea en una señal pasabanda haciendo $u(t) = a(t)$. La señal banda base $u(t)$ es en consecuencia real y tiene un espectro simétrico alrededor de la portadora. La señal pasabanda se representa matematicamente como

$$y(t) = \sqrt{2}\, Re\{a(t)e^{j\omega_c t}\} = \sqrt{2}a(t)\cos(\omega_c t)$$

La señal pasabanda es compleja con simetría conjugada alrededor de la frecuencia portadora ω_c ya que la señal banda base es real.

En el caso de *modulación de amplitud de banda lateral única* (AM-BLU) nuevamente se tiene una señal moduladora real $a(t)$, pero ahora la señal banda base compleja se define como una señal analítica obtenida con un divisor de fase $u(t) = 1/2[a(t) + j\hat{a}(t)]$. Dado que la señal banda base compleja es analítica, la señal pasabanda tiene solo componentes de frecuencia correspondientes a la banda por encima de la portadora. La ventaja de AM-BLU sobre AM-DBL es que para la misma $a(t)$, el ancho de banda de la versión pasabanda de la primera es la mitad que para la segunda. Una desventaja de AM-BLU es el divisor de fase banda base, el cual es difícil de realizar debido a la discontinuidad de fase en cc, a menos que en la señal banda base pudiera despreciarse frecuencias próximas a cc (lo cual sucede para voz telefónica).

La tercera técnica de modulación es *modulación de amplitud en cuadratura* (QAM). En este caso, se tienen dos señales moduladoras reales $a(t)$ y $b(t)$ tal que $u(t) = a(t) + jb(t)$. La señal banda base compleja no es ni analítica ni tiene simetría compleja alrededor de cc. La señal pasabanda tiene en general bandas laterales superior e inferior y no tiene una simetría particular alrededor de la portadora. El término QAM aparece de la representación

$$y(t) = \sqrt{2}\, Re\{(a(t) + jb(t))e^{j\omega_c t}\} = \sqrt{2}a(t)\cos(\omega_c t) - \sqrt{2}b(t)\sin(\omega_c t)$$

En otras palabras, una señal QAM consisten en dos portadoras moduladas independientemente con un desplazamiento de fase relativo de $\pi/2$. Para la misma señal banda base, QAM requiere la misma señal pasabanda relativa a AM-DBL, el cual es el doble que para AM-BLU. Sin embargo, permite transmitir dos señales reales en lugar de una sola. Ofrece la misma eficiencia espectral que AM-BLU, pero sin requerir el complicado divisor de fase de banda base. De esta forma QAM y técnicas de modulación similares se han vuelto las más utilizadas para comunicaciones digitales.

2.1.4 Muestreo de señales pasabanda (*)

Consideremos una señal pasabanda compleja $\tilde{x}(t)$, descripta por

$$x(t) = \sqrt{2}\, Re[\tilde{x}(t)e^{jw_c t}]$$

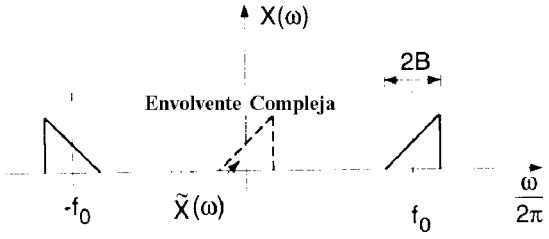

Figura 2.3: Señal pasabanda

Cuando $x(t)$ es estrictamente pasabanda y centrada en w_c, como se muestra en la figura 2.3, entonces $\tilde{x}(t)$ es un proceso pasabajos complejo. Esta puede obtenerse de $x(t)$ efectuando la demodulación

$$\begin{aligned} x(t)e^{-jw_c t} &= \sqrt{2}\, Re[\tilde{x}(t)e^{jw_c t}]e^{-jw_c t} \\ &= \frac{\sqrt{2}}{2}\tilde{x}(t) + \frac{\sqrt{2}}{2}\tilde{x}^*(t)e^{-j2w_c t} \end{aligned}$$

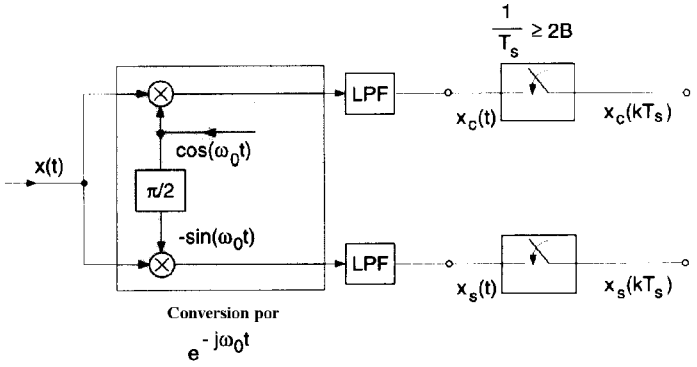

Figura 2.4: Conversión de frecuencia seguida de muestreo.

El término de frecuencia $2w_c$ en esta última ecuación se elimina por medio de un filtro pasabajos como mostrado en la figura 2.4.

La conversión analógico - digital puede trasladarse hacia la entrada $x(t)$ muestreando simultaneamente la señal a la frecuencia de la portadora w_c y trasladandola a la banda base. Si $f_s = 1/T_s$ es la frecuencia de muestreo, entonces

$$x(t)e^{-jw_c t}|_{t=kT_s} = \sqrt{2}\, Re[\tilde{x}(t)e^{jw_c t}]e^{-jw_c t}|_{t=kT_s}$$

Una forma simple de evitar la multiplicación compleja se obtiene si muestreamos la señal a cuatro veces la frecuencia de la portadora f_c, o sea: $f_s = 4f_c$. Dado que $e^{jkw_c T_s} = (j)^k$, en la ecuación anterior se tendrá que

$$(-j)^k x(kT_s) = \frac{\sqrt{2}}{2}\left[\tilde{x}(kT_s) + \tilde{x}^*(kT_s)(-1)^k\right]$$

de donde es posible obtener la *componente en fase* en las muestras pares $(2kT_s)$ y la *componente en cuadratura* en las muestras impares $((2k+1)T_s)$, o sea

$$
\begin{aligned}
(-1)^k x(2kT_s) &= \sqrt{2}\, Re\,[\tilde{x}(2kT_s)] = \sqrt{2}x_c(2kT_s) \\
(-1)^k x((2k+1)T_s) &= -\sqrt{2}\, Im\,[\tilde{x}((2k+1)T_s)] = -\sqrt{2}x_s((2k+1)T_s)
\end{aligned}
$$

Dado que en general se requieren ambas componentes en el mismo instante de muestreo, es posible obtener la señal muestreada en un mismo instante mediante interpolación. Por razones de simetría ambas componentes se interpolan de la siguiente forma:

$$
\begin{aligned}
x'_c(2kT_s) &= x_c(2kT_s - T_s/2) \\
x'_s(2kT_s) &= x_s((2k-1)T_s + T_s/2)
\end{aligned}
$$

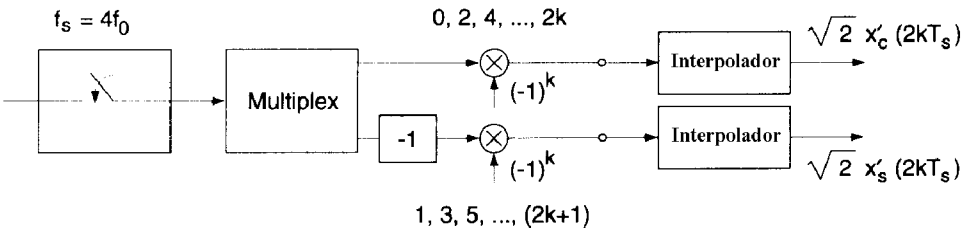

Figura 2.5: Muestreo digital y conversión de frecuencia.

El diagrama funcional de la combinación de muestras y la conversión de frecuencia hacia abajo se ilustran en la figura 2.5. La señal real $x(t)$ se muestrea a cuatro veces la frecuencia f_c. Luego las muestras se dividen a través de un multiplexador, provisto de signos alternantes a la salida, y finalmente se interpola para obtener muestras en el mismo instante de tiempo. Notar que el multiplexador disminuye la frecuencia de muestreo por dos.

El equivalente digital de demodulador de la figura 2.5 es muy atractivo debido a razones de implementación, ya que evita el costo de mezcladores analógicos y filtros pasabajos analógicos a expensas del procesamiento digital en el interpolador. Funcionalmente, el conversor digital provee las dos componentes exactamente en cuadratura dado que el incremento de fase entre dos muestras consecutivas es $w_c T_s = \pi/2$. Un desplazamiento de fase exactamente igual a $\pi/2$ es crucial para una detección óptima. Esto es dificil de lograr con circuitos analógicos.

Una extensión

Si la señal $x(t)$ se convierte a una frecuencia menor por medio de un mezclador analógico, se requiere una frecuencia de muestreo mínima de $f_s \geq 2B$. La desventaja de este conversor digital es que se requiere una frecuencia de muestreo de $4f_c$, grande si comparada con el ancho de banda de la envolvente de $\tilde{x}(t)$ que transporta la información útil. Se discutirá a continuación una generalización de los conceptos básicos anteriores, que mantiene su simplicidad pero utiliza una frecuencia de muestreo menor.

La idea es esencialmente simple. Buscamos una frecuencia de muestreo que produzca incrementos de fase de $\pi/2$, tal que $e^{jk2\pi f_c T_2} = e^{j(\pi/2)k}$. La primera solución a esto es la propuesta anteriormente, $f_c T_s = 1/4$.

La solución general, sin embargo, es

$$
f_c T_s = \pm\frac{1}{4} + N > 0
$$

donde el signo negativo produce disminuciones de fase de $\pi/2$. De esta forma

$$
f_s = \frac{1}{T_s} = \frac{4f_c}{4N \pm 1} \tag{2.2}
$$

o en general

$$f_s = \frac{4f_c}{2n \pm 1}; \quad n \geq 0 \tag{2.3}$$

Basicamente la ecuación (2.2) establece que es necesario saltear N ciclos del fasor $e^{jw_c t}$ antes de obtener la nueva muestra. Hay un máximo para N determinado por el teorema del muestreo. Como se obtienen muestras de las componentes en cuadratura cada $2T_s$ unidades de tiempo, entonces debe ser $1/(2T_s) \geq 2B$, ó

$$f_s \geq 4B$$

Una frecuencia de muestreo menor simplifica el conversor A/D pero requiere un interpolador más sofisticado, este es el compromiso que el diseño permite.

Generalización

Hasta este punto el objetivo ha sido simplificar la implementación digital suponiendo ciertas relaciones entre la frecuencia de la portadora y la frecuencia de muestreo. En algunos casos, puede ser imposible satisfacer tales relaciones exactamente. En esos casos, es deseable elegir f_s de acuerdo a una regla más general. El requerimiento más general para la frecuencia de muestreo está dado por

$$\frac{2f_c + 2B}{N} \leq f_s \leq \frac{2f_c - 2B}{N-1} \tag{2.4}$$

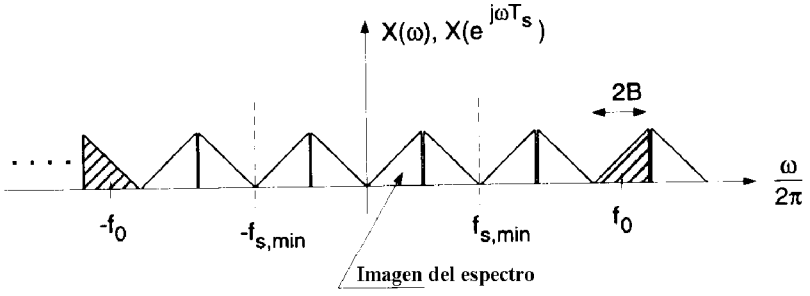

Figura 2.6: Espectro de la señal submuestreada.

para $N = 1, \cdots, N_{max}$ y $N_{max} = int\left(\frac{f_c+B}{2B}\right)$. Notar que la igualdad en el lado derecho de esta ecuación se cumple solo para $N > 1$. Si esa ecuación es satisfecha, el espectro imagen de la señal muestreada $x(kT_s)$ ocupará el espacio de frecuencias entre las dos componentes del espectro original como mostrado en la figura 2.6. La frecuencia de muestreo mínima requerida $f_{s,min}$ (para $N = N_{max}$) es aproximadamente dos veces el ancho de banda $2B$ de la señal real pasabanda. Si $(f_c + B)/2B$ es un valor entero entonces $f_{s,min} = 4B$ se satisface exactamente. En este caso, las mínimas frecuencias de muestreo obtenibles con (2.3) y (2.4) coinciden.

Las imagenes negativas pueden eliminarse por medio de un *transformador de Hilbert*. Un transformador de Hilbert con respuesta en frecuencia dada por

$$H_d(e^{j2\pi fT_s}) = \begin{cases} -j & para\ 0 \leq fT_s < \frac{1}{2} \\ j & para\ \frac{1}{2} \leq fT_s < 1 \end{cases}$$

permite obtener las componentes en cuadratura $x_s(kT_s)$ a partir de $x(kT_s)$ de manera que formen un par transformado de Hilbert. La figura 2.7 ilustra el efecto del transformador de Hilbert digital sobre el espectro.

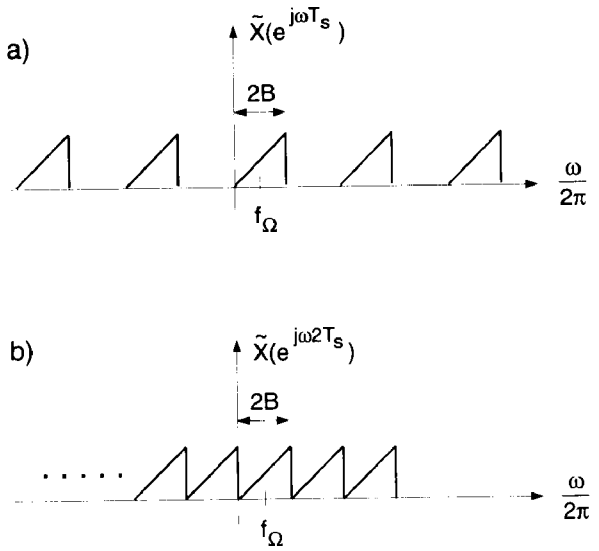

Figura 2.7: a) Espectro de la señal analítica, b) Espectro después de la decimación.

La porción de banda lateral no deseada se elimina obteniendose el espectro de una señal compleja $\tilde{\tilde{x}}(kT_s)$ que es igual a la envolvente de $\tilde{x}(kT_s)$ desplazada del origen en f_Ω.

Dado que el espectro de la señal compleja resultante ocupa solo la mitad del ancho de banda de la señal real original, las secuencias en fase y cuadratura pueden *decimarse* por un factor de 2. La función transferencia del transformador de Hilbert puede aproximarse eficientemente por un filtro FIR con un número de coeficientes impar N_{DHT} cuya respuesta en frecuencia satisfaga

$$\tilde{H}_d(e^{j2\pi fT_s}) = \tilde{H}_d(e^{j(\pi - 2\pi fT_s)})$$

En ese caso, la respuesta impulsiva satisface $\tilde{h}_d(n) = 0$ para n impar.

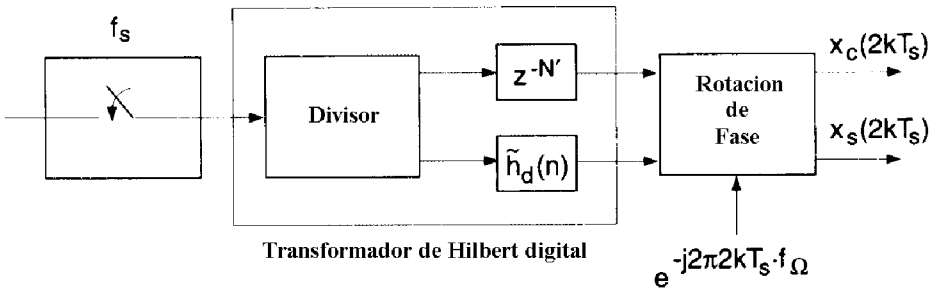

Figura 2.8: Conversión digital usando el Transformador de Hilbert.

La figura 2.8 muestra el diagrama funcional del conversor. Debido a la caracteristica de la respuesta impulsiva del transformador de Hilbert, la decimación puede realizarse a la entrada dividiendo las muestras alternativamente en las ramas de las componentes en fase y cuadratura. La componente en cuadratura se obtiene filtrando con $\tilde{H}_d(e^{j2\pi fT_s})$. Para compensar por la causalidad del filtro FIR, las muestras en fase deben retardarse en $N' = (N_{DHT} - 1)/2$. Finalmente, el desplazamiento al orígen de frecuencia conocido f_Ω, se remueve con una rotación de fase. Esto puede realizarse, por ejemplo, en conjunto con la corrección del error en la frecuencia de una portadora no conocida.

2.2 Procesos estocásticos y ruido

A pesar que los procedimientos de modulación y demodulación son determinísticos, la información a ser transmitida sobre un sistema de comunicaciones, así como el ruido encontrado en el medio físico de transmisión, es aleatória o estocástica. Esos fenómenos no pueden ser predichos a priori exactamente, pero tienen ciertas características predecibles que pueden sintetizarse en un modelo de proceso estocástico.

En esta sección se realizará una revisión de la notación que será utilizada para variables aleatórias y procesos estocásticos, y se cubriran además algunos tópicos que serán particularmente importantes en lo que sigue. Las cadenas de Markov serán especialmente útiles para el modelado de un canal de fibra óptica (procesos de Poisson y ruido shot en particular), y están asociadas también al problema de detección en general (detector de Viterbi).

2.2.1 Variables aleatórias

En lo que sigue se colocará especial énfasis en las combinaciones de variables aleatórias discretas y contínuas usualmente encontradas en comunicaciones digitales.

Notaremos una *variable aleatória* con una mayúscula, tal como X, y una *muestra* de la variable aleatória con una minúscula, x. La variable aleatória es una función compleja o real definida sobre el *espacio muestral* Ω de todas las posibles muestras. Un *evento* E es un conjunto de posibles muestras con una probabilidad asociada $P[E]$, con $0 \leq P[E] \leq 1$. Dado que un evento es un conjunto, es posible definir la unión de dos eventos $E_1 \cup E_2$, o la intersección de eventos $E_1 \cap E_2$. La fórmula básica

$$P[E_1 \cup E_2] = P[E_1] + P[E_2] - P[E_1 \cap E_2]$$

conduce a la muy útil *cota de unión* dada por

$$P[E_1 \cup E_2] \leq P[E_1] + P[E_2]$$

La *función distribución acumulativa* (cdf) de una variable aleatória X es la probabilidad de un evento $X \leq x$, para un x dado, o sea

$$F_X(x) = P[X \leq x]$$

Cuando no introduzca ambigüedad se omitirá el subíndice X, escribiendo la cdf $F(x)$. Para una variable aleatória Y,

$$F_Y(y) = P[Re\{X\} \leq Re\{y\}, Im\{Y\} \leq Im\{y\}]$$

Para una variable aleatória real contínua, la *función densidad de probabilidad* (pdf) $f_X(x)$ se define de forma que para cualquier intervalo $I \subset \boldsymbol{R}$,

$$P[X \in I] = \int_I f_X(x)dx$$

Para el caso de una variable aleatória compleja, I es una región en el plano complejo. Para una variable aleatória X,

$$f_X(x) = \frac{\partial}{\partial x} F_x(x)$$

donde exista la derivada. Se utilizará frecuentemente la derivada en el sentido más amplio, tal que cuando la cdf incluye una función escalón, la pdf correspondiente tiene una función delta de Dirac.

Ejemplo: Para la siguiente cdf,

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

la pdf consiste exclusivamente de funciones delta de Dirac,

$$f(x) = 0.5\,\delta(x) + 0.5\,\delta(n-1)$$

Tal pdf es característica de una variable aleatória discreta.

Para una variable aleatória discreta X, notaremos la probabilidad de una muestra $x \in \Omega$ como: $p_X(x) = P[X = x]$, donde se omitirá el subíndice cuando no exista ambigüedad. En este caso se habla de *función de masa de probabilidad* (pmf), en lugar de pdf, la que puede escribirse como

$$f_X(x) = \sum_{y \in \Omega_X} p_X(y)\delta(x-y)$$

El *valor esperado* o *media* de X se define como

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx \quad \text{ó} \quad E[X] = \sum_{x \in \Omega} x\, p_X(x)$$

para variables aleatórias contínuas o discretas, respectivamente. Para una variable aleatória compleja Y, es posible integrar sobre el plano complejo

$$E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+jz) f_Y(x+jz)dx\,dz$$

El *Teorema fundamental de la esperanza* establece que si $g(.)$ es cualquier función definida sobre el espacio muestral de X, entonces

$$E[g(X)] = \int_{-\infty}^{\infty} g(X) f_X(x)dx$$

Algunas esperanzas especialmente importantes son la media y la varianza, definidas como

$$\mu = E[X], \quad \sigma_X^2 = E[(X - E[X])^2] = E[X^2] - E[X]^2.$$

Para variables aleatórias complejas la varianza se define similarmente como

$$\sigma_X^2 = E[|X|^2] - |E[X]|^2 = E[XX^*] - E[X](E[X])^*$$

La *cdf conjunta* de dos variables aleatórias X e Y será

$$F_{X,Y}(x,y) = P[X \leq x, Y \leq y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(\alpha,\beta)d\alpha\,d\beta$$

donde $f_{X,Y}(x,y)$ es la *pdf conjunta*. La pdf conjunta puede escribirse en términos de la cdf conjunta como

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$$

donde se ha omitido los subíndices como antes. La *pdf marginal* $f_X(x)$ de una variable aleatória X puede obtenerse de la pdf conjunta como

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

Las variables aleatórias X e Y son *estadísticamente independientes* si para todo intervalo I y J,

$$P[X \in I \cap Y \in J] = P[X \in I] P[Y \in J]$$

lo cual es equivalente a

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \text{ó } F_{X,Y}(x,y) = F_X(x) F_Y(y).$$

Independencia implíca que la *correlación cruzada* es

$$E[XY] = E[X]E[Y]$$

Cuando esta ecuación es satisfecha, las variables aleatórias se dicen *no correlacionadas*. Dos variables aleatórias pueden ser no correlacionadas y no ser independientes.

Función característica

La *función característica* de X se define como

$$\Phi_X(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx$$

para una variable aleatória s. Esta es la transformada de Laplace de $f_X(x)$ evaluada en $x = -s$. Cuando s es real, lo cual es suficiente para las aplicaciones contempladas, la definición anterior se denomina *función generadora de momentos*.

Ejemplo: Definiendo $Z = X + Y$, con X e Y independientes, entonces es fácil mostrar que

$$\Phi_Z(s) = \Phi_X(s) \Phi_Y(s)$$

Ejemplo: En base a la definición es simple mostrar que

$$E[X] = \frac{\partial}{\partial s} \Phi_X(s) \bigg|_{s=0}, \quad E[X^2] = \frac{\partial^2}{\partial s^2} \Phi_X(s) \bigg|_{s=0}$$

La *Cota de Chernoff*, con base en la función generadora de momentos, es muy útil para acotar la cola de probabilidad de una variable aleatória cuando una evaluación exacta es muy complicada.

Ejemplo: a) La probabilidad de un evento $X > x$ está acotada por

$$\begin{aligned} 1 - F_X(x) &= P[X > x] = \int_x^\infty f_X(x)dx = \int_{-\infty}^\infty u(x-t) f_X(t)dt \\ &= \int_{-\infty}^\infty [e^{-s(x-t)} e^{s(x-t)}] u(x-t) f_X(t)dt \\ &\leq \int_{-\infty}^\infty e^{-s(x-t)} f_X(t)dt \\ &\leq e^{-sx} \int_{-\infty}^\infty e^{st} f_X(t)dt = e^{-sx}\Phi_X(s) \end{aligned}$$

($u(t)$ es la función escalón) lo que establece que la cola de la pdf decrece al menos exponencialmente para cualquier distribución para la cual exista la función generadora de momentos.
b) Es posible mostrar por definición además que

$$F_X(x) \leq e^{sx}\Phi_X(-s)$$

para $s \geq 0$.
c) Es posible mostrar que el s que minimiza la cota en a) y b) debe satisfacer

$$x\Phi_X(s) = \frac{\partial \Phi_X(s)}{\partial s}, \quad -x\Phi_X(-s) = \frac{\partial \Phi_X(-s)}{\partial s}$$

Probabilidades Condicionales y Regla de Bayes

La *probabilidad condicional* de que una variable aleatória contínua X esté en el intervalo I, dada Y en el intervalo J, se define para todo J tal que $P[Y \in J] \neq 0$ como

$$P[X \in I | Y \in J] = \frac{P[X \in I \cap Y \in J]}{P[Y \in J]}$$

donde $P[Y \in J]$ se denomina *probabilidad marginal* debido a que en ella no se consideran los posibles efectos de X sobre Y. Para variables aleatórias complejas o vectoriales, I y J son regiones o volumenes antes que intervalos. Si X e Y son independientes, entonces $P[X \in I \cap Y \in J] = P[X \in I]$. La probabilidad conjunta puede escribirse en términos de probabilidad condicional como sigue

$$P[X \in I \cap Y \in J] = P[X \in I | Y \in J] P[Y \in J]$$

Equivalentemente,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

donde $f_Y(y)$ se denomina *densidad marginal*. La *densidad condicional* $f_{X|Y}(x|y)$ está bien definida solamente para y tal que $f_Y(y) \neq 0$. Además, dado que $f_{X,Y}(x,y) = f_{Y,X}(y,x)$, la ecuación anterior implica que

$$f_{X|Y}(x|y) f_Y(y) = f_{Y|X}(y|x) f_X(x)$$

la cual es una forma de la *Regla de Bayes*. Es común en el análisis de sistemas de comunicaciones digitales encontrar variables aleatórias discretas y contínuas en el mismo sistema. En ese caso la ecuación anterior contendrá funciones delta de Dirac.

Ejemplo: Supongamos tener una Y discreta y X contínua, entonces integrando la ecuación anterior sobre intervalos de y, se tiene una forma combinada de la Regla de Bayes dada por

$$f_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)f_X(x)$$

Esto involucra probabilidades y pdf. No tiene funciones delta de Dirac porque X es contínua. Si X es también discreta, entonces se tendrá que

$$p_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)p_X(x)$$

la cual tiene solo probabilidades discretas.

Para el caso de distribuciones discretas, la probabilidad marginal puede escribirse en términos de las probabilidades condicionales como

$$p_Y(y) = \sum_{x\in\Omega} p_{Y|X}(y|x)p_X(x) = \sum_{x\in\Omega} p_{Y,X}(y,x) \tag{2.5}$$

donde Ω es un espacio muestral numerable de X. Esta relación muestra como obtener probabilidades marginales de una variable aleatória dada solamente la probabilidad conjunta, o dadas solamente las probabilidades condicionales y las probabilidad marginales de la otra variable aleatória. Utilizando esta relación es posible escribir la probabilidad condicional de X dado Y en términos de la probabilidad condicional de Y dado X y la probabilidad marginal de X como sigue

$$p_{Y|X}(y|x) = \frac{p_{Y|X}(y|x)p_Y(y)}{\sum_{x\in\Omega_X} p_{Y|X}(y|x)p_X(x)}$$

Esta relación se conoce como *Teorema de Bayes*. El análogo del Teorema de Bayes para variables aleatórias contínuas tiene la forma

$$f_{Y|X}(y|x) = \frac{f_{Y|X}(y|x)f_Y(y)}{\int_{x\in\Omega_X} f_{Y|X}(y|x)f_X(x)dx}$$

Variables aleatórias Gaussianas y el Teorema del Límite Central

Una variable aleatória *Gaussiana* ó *Normal* tiene una pdf dada por

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

donde σ^2 es la varianza y μ es la media. La cdf puede expresarse solo como una integral

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(\alpha-\mu)^2/2\sigma^2} d\alpha$$

para la cual no existe una forma cerrada. Una variable aleatória Gaussiana *estándar* es aquella que tiene media cero y variancia $\sigma^2 = 1$. La *función distribución complementaria* se notará como

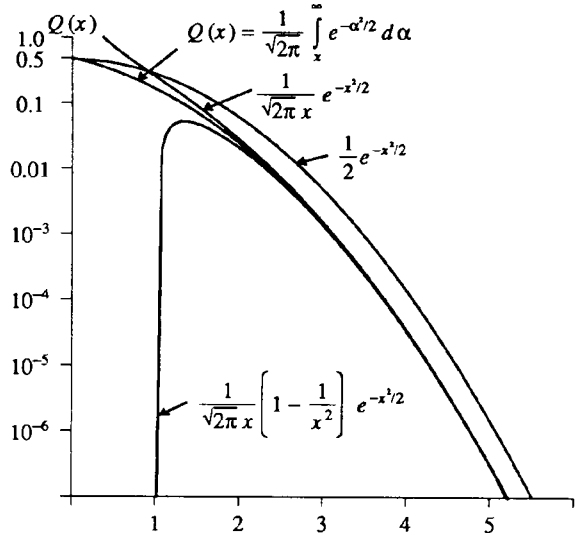

Figura 2.9: Probabilidad $Q(x)$ que una variable aleatória Gaussiana de media cero X exceda a x, en escala logarítmica.

$$Q(x) = P[X > x] = 1 - F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\alpha^2/2} d\alpha$$

lo que representa la integral de la cola de la función densidad de probabilidad Gaussiana. Esta función se ilustra en la figura 2.9, donde se ha utilizado una escala logaritmica para las probabilidades. Esta función está relacionada con la *función error* ($erf(x)$) y la *función error complementaria* ($erfc(x)$) ampliamente tabuladas, de la siguiente forma:

$$Q(x) = \frac{1}{2} erfc\left[\frac{x}{\sqrt{2}}\right] = \frac{1}{2}\left[1 - erf\left[\frac{x}{\sqrt{2}}\right]\right]$$

Ejemplo: Es simple mostrar, cambiando los extremos de integración en la definición, que para una variable aleatória X con media μ y varianza σ^2 se verifica

$$P[X > x] = Q\left[\frac{x - \mu}{\sigma}\right]$$

A pesar que $Q(.)$ puede determinarse solo numericamente o en forma tabulada, es posible obtener valores próximos muy útiles utilizando la Cota de Chernoff.

Ejemplo: La función generadora de momentos de una variable aleatória Gaussiana de media μ y varianza σ^2 es, siguiendo la definición,

$$\log_e \Phi_X(s) = \mu s + \sigma^2 s^2/2$$

En base a este resultado, de la Cota de Chernoff se obtiene que

$$1 - F_X(x) \leq e^{-(x-\mu)^2/2\sigma^2}$$

tal que

$$Q(x) \leq e^{-x^2/2}$$

Otras cotas útiles pueden obtenerse como mostrado en la figura 2.9.

La utilización de la distribución Gaussiana para modelar fenómenos de ruido puede justificarse en base a conceptos físicos mediante el *Teorema del Limite Central*. Este teorema determina que la distribución Gaussiana es un buen modelo para el efecto acumulativo de un gran número de variables aleatórias independientes, independientemente de sus distribuciones individuales. Más precisamente, es posible asumir que $\{Y_i\}$, para $1 \leq i \leq N$ es un conjunto de variables aleatórias estadísticamente independientes con la misma pdf $f_Y(y) = f(y)$ y varianza finita σ^2. O sea, las variables aleatórias son *independientes e indenticamente distribuidas* (i.i.d.). Definiendo además una variable aleatória Z como la suma normalizada de las Y_i, o sea

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Y_i$$

Entonces la función distribución de Z se aproxima a la Gaussiana $(1 - Q(z/\sigma))$ cuando $N \to \infty$. Si cada variable aleatória Y_i representa algún fenómeno físico individual y Z es el efecto acumulativo de esos fenómenos, entonces cuando N se vuelve grande, la distribución de Z se vuelve Gaussiana, independientemente de la distribución de las Y_i.

Ejemplo: Para una combinación lineal de N variables aleatórias Gaussianas de media cero e independientes X_i, cada una con varianza σ_i^2,

$$Z = a_1 X_1 + \cdots + a_N X_N$$

es posible utilizar la función generadora de momentos para mostrar que Z en sí misma es Gaussiana de media cero y varianza:

$$\sigma_Z = (a_1^2 + \cdots + a_N^2)\sigma^2$$

Dos variables aleatórias Gaussianas con media cero y varianza σ^2 son *conjuntamente Gaussianas* si su pdf conjunta es

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{x^2 - 2\rho xy + y^2}{2\sigma^2(1-\rho^2)}}$$

donde ρ se denomina el *coeficiente de correlación*, $\rho = E[XY]/\sigma^2$. Notar que $-1 \leq \rho \leq 1$ y que si X e Y no están correlacionadas $\rho = 0$.

Esta definición puede extenderse para $N > 2$ variables aleatórias conjuntamente Gaussianas. Si un vector aleatório $\boldsymbol{X}$ tiene componentes que son variables aleatórias Gaussianas independientes conjuntamente con la misma varianza σ^2, entonces la pdf conjunta será

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\frac{1}{2\sigma^2}\|\boldsymbol{x}\|^2}$$

donde N es el número de componentes del vector y $\|\boldsymbol{x}\|$ es la norma Euclidiana. Cuando $\boldsymbol{X}$ es un vector complejo con partes real e imaginaria independientes la ecuación anterior aún es válida. Además, como ya se discutió, cualquier combinación lineal de variables aleatórias Gaussianas conjuntas es Gaussiana.

Esto puede generalizarse aún más si consideramos un vector $\boldsymbol{X}$ con M variables aleatórias unicamente con pdf Gaussiana real, tal que la pdf conjunta será

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}_X|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_X)^T \boldsymbol{C}_X{}^{-1}(\boldsymbol{x}-\boldsymbol{m}_X)}$$

donde $\boldsymbol{C}_X = E[(\boldsymbol{x}-\boldsymbol{m}_X)(\boldsymbol{x}-\boldsymbol{m}_X)^T]$ es la *matriz covarianza*, $|\boldsymbol{C}_X|$ es su determinante y $\boldsymbol{m}_X = E[\boldsymbol{X}]$ es el vector media. En el caso especial en que el vector media es cero y los elementos del vector aleatório son independientes con igual varianza, $\boldsymbol{C}_X$ es diagonal y las ecuaciones anteriores son iguales. Una observación de esta última es que la pdf de un vector aleatório Gaussiano está completamente especificada por el vector media y la matriz covarianza. En consecuencia, estos dos conjuntos de parámetros especifican completamente las propiedades estadísticas de un vector aleatório Gaussiano.

Ley de los grandes números. Si se tiene una secuencia de variables aleatórias $(Y_1, Y_2, \cdots, Y_N)$ con las mismas propiedades básicas, entonces el comportamiento del promedio $Z = \frac{1}{n}\sum_{i=1}^{N} Y_i$ es esperable que sea menos aleatório que cada Y_i. La ley de los grandes números y el teorema del límite central son proposiciones precisas de esta idea intuitiva.

La *ley débil de los grandes números* determina que si la secuencia de variables aleatórias $(Y_1, Y_2, \cdots, Y_N)$ no correlacionada con media m_Y y varianza $\sigma_Y^2 < \infty$, entonces para cualquier $\epsilon > 0$, $\lim_{N\to\infty} P(|Z - m_Y| > \epsilon) = 0$, donde $Z = \frac{1}{n}\sum_{i=1}^{N} Y_i$. Esto significa que el promedio converge (en probabilidad) al valor esperado.

El *teorema del límite central* no solo determina que la convergencia del promedio a la media si no que también dá una idea de la distribución del promedio, como ya se discutió anteriormente. Notar que, a pesar que del teorema del límite central es posible concluir que el promedio converge al valor esperado, no es posible concluir que la ley de los grandes números se obtiene del teorema del límite central. Esto se debe a que las hipótesis en general asociadas al teorema del límite central son más específicas. En este caso se requiere que las variables sean independientes, mientras que en la ley de los grandes números se requiere que las variables aleatórias no estén correlacionadas.

2.2.2 Procesos estocásticos

Un proceso estocástico discreto $\{X_k\}$ es una secuencia de variables aleatórias con índices enteros k, y un proceso estocástico contínuo $X(t)$ tiene índices en la variable real t. Notaremos una *realización* o muestra de $\{X_k\}$ o $X(t)$ como la señal determinística $\{x_k\}$ o $x(t)$, respectivamente. Cuando no exista ambigüedad entre una *señal* y una *realización* de la señal se omitirán las llaves $\{.\}$. Cada realización X_k o $X(t)$ puede ser compleja, real o vectorial.

Ejemplo: Un proceso estocástico real $X(t)$ es un *proceso estocástico Gaussiano* si sus muestras $[X(t_1), \cdots, X(t_N)]$ son variables aleatórias conjuntamente Gaussianas para cualquier N y $[t_1, \cdots, t_N]$

Los momentos de primero y segundo orden de un proceso estocástico son: la *media*

$$m_k = E[X_k] \qquad m(t) = E[X(t)]$$

y la *autocorrelación*:

$$R_{k,i} = E[X_k X_i^*] \qquad R_{XX}(t_1, t_2) = E[X(t_1)X^*(t_2)]$$

donde X^* es el complejo conjugado de X.

Ejemplo: Es posible considerar un proceso estocástico Gaussiano real de media cero. Un vector estocástico $\boldsymbol{X}$ puede construirse de un conjunto arbitrário de muestras. Para ese vector, la matriz covarianza puede obtenerse de la definición de autocorrelación anterior. En consecuencia, la pdf conjunta de cualquier conjunto

de muestras puede obtenerse de la función autocorrelación. De esta forma, las propiedades estadísticas de un proceso estocástico Gaussiano real de media cero quedan completamente especificadas a través de su función autocorrelación.

Un proceso estocástico es *estrictamente estacionario* (ESE) si la pdf de cualquier muestra es independiente del índice de la muestra, y la pdf conjunta de cualquier conjunto de muestras depende solo de la diferencia de índices entre muestras, y no de los valores absolutos de los índices de cualquier muestra. Si es *estacionario en sentido amplio* (ESA), la media es independiente del índice, y la autocorrelación depende solamente de la diferencia entre índices d muestras, y no del valor absoluto de los índices. En otras palabras, m_k o $m(t)$ deben ser constantes y $R_{XX}(k,i)$ o $R_{XX}(t_1,t_2)$ deben ser función solo de la diferencia $k-i$ o t_1-t_2. Estacionaridad en sentido estricto implica estacionaridad en sentido amplio, pero no a la inversa a menos que el proceso sea Gaussiano.

Ejemplo: Un proceso estocástico Gaussiano real ESA es también ESE. La función autocorrelación y la media pueden usarse para construir la matriz covarianza para cualquier conjunto de muestras. Como el proceso es ESA, las componentes de la matriz serán independientes del índice absoluto, y dependerán solo de la diferencia de subíndices. En consecuencia, la pdf conjunta de cualquier conjunto de muestras dependerá solo de esas diferencias de índices por lo que, en consecuencia, el proceso es ESE.

La función autocorrelación de un proceso estocástico ESA puede escribirse en términos de la diferencia entre índices, $m=k-i$ ó $\tau=t_1-t_2$, lo que produce la notación simplificada

$$R_X(m)=E[X_{k+m}X_k^*], \qquad R_X(\tau)=E[X(t+\tau)X^*(t)]$$

$R(0)$ es el momento de segundo orden de las muestras

$$R_X(0)=E[|X_k|^2], \qquad R_X(0)=E[|X(t)|^2]$$

y puede interpretarse como la *potencia* del proceso estocástico. La *densidad espectral de potencia* de un proceso estocástico ESA es la transformada de Fourier de la función autocorrelación,

$$S_X(e^{j\omega T})=\sum_{m=-\infty}^{\infty} R_X(m)e^{-j\omega mT}, \qquad S_X(j\omega)=\int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau$$

donde T es el intervalo de muestras del proceso estocástico discreto en tiempo. La potencia, en consecuencia, es la integral de la densidad espectral de potencia,

$$R_X(0)=\frac{T}{2\pi}\int_{m=-\pi/T}^{\pi/T} S_X(e^{j\omega T})d\omega, \qquad R_X(0)=\frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(e^{-j\omega})d\omega$$

La densidad espectral de potencia es real ya que la función autocorrelación es simétrica conjugada, $R_X(m)=R_X^*(-m)$ ó $R_X(\tau)=R_X^*(-\tau)$. Es también no negativa. Además, si X_k ó $X(t)$ es real, entonces la densidad espectral de potencia es simétrica alrededor de $\omega=0$. Es posible escribir también la densidad espectral de potencia como una transformada Z o una transformada de Laplace,

$$S_X(z)=\sum_{m=-\infty}^{\infty} R_X(m)z^{-m}, \qquad S_X(s)=\int_{-\infty}^{\infty} R_X(\tau)e^{-s\tau}d\tau$$

Ejemplo: Considerar un proceso estocástico $\{X_k\}$ donde las muestras X_k son todas variables aleatórias independientes e identicamente distribuidas (i.i.d.), con media cero y varianza σ_x^2. En este caso $R_X(k) = \sigma_x^2 \delta_k$, y la densidad espectral de potencia es una constante, $S_X(e^{j\omega T}) = \sigma_x^2$, independiente de la frecuencia, con potencia $R_X(0) = \sigma_x^2$.

Cualquier proceso de media cero con una densidad espectral de potencia constante se denomina un *proceso estocástico ruido blanco*. Esto puede o no implicar que las muestras del proceso estocástico sean independientes, apesar que en el importante caso Gaussiano lo serán.

Ejemplo: Como en el ejemplo anterior, consideremos un proceso estocástico contínuo $X(t)$ con una función autocorrelación

$$R_X(\tau) = N_0 \delta(\tau)$$

La densidad espectral de potencia de este proceso es constante, $S_X(j\omega) = N_0$, de forma que $\{X(t)\}$ es blanco. La potencia $R_X(0)$ de este proceso blanco contínuo es infinita, de forma que inmediatamente se encuentra dificultades matemáticas que no se encuentra en el caso discreto.

A pesar que el proceso contínuo de ruido blanco del ejemplo anterior conduce a una condición fisicamente no realizable de potencia infinita (o indefinida), el modelo es extremadamente importante. Ya que $R_X(\tau) = 0$ para todo $\tau \neq 0$, esto indica que dos muestras distintas de un proceso contínuo blanco no están correlacionadas. Desafortunadamente esto no tiene ningún significado físico. En consecuencia, muestrear un proceso estocástico contínuo y blanco es un concepto ambiguamente definido. Rigurosamente hablando, un proceso estocástico blanco contínuo varía tan rapidamente que no es posible determinar sus características en el tiempo.
A pesar de esas dificultades matemáticas, el proceso blanco contínuo es muy útil como modelo de ruido, y tiene una densidad espectral de potencia aproximadamente constante sobre un ancho de banda mayor que el ancho de banda del sistema bajo análisis. En ese sistema siempre se limita el ruido para eliminar cualquier componente fuera de la banda de interés. En ese caso, no existe ninguna diferencia si comenzamos considerando un ruido blanco o un modelo más preciso; el resultado será muy aproximadamente el mismo. Pero usando el modelo de ruido blanco se simplifican considerablemente las manipulaciones algebráicas. En lo que sigue se utilizará frecuentemente el modelo de ruido blanco, tomando cuidado siempre de limitar en banda este ruido antes de realizar otras operaciones, tal como el muestreo. Después de la limitación en frecuencia, se obtiene un proceso bien comportado con potencia finita.

Ejemplo: El ruido *térmico* ó *Johnson* de las resistencias eléctricas tiene una densidad espectral de potencia plana hasta aproximadamente 10^{12} Hz., o sea, un ancho de banda mucho mayor que muchos de los sistemas de interés. De esta forma, es posible utilizar adecuadamente el ruido blanco como modelo de este ruido térmico sin comprometer la precisión. El ruido en el modelo a frecuencias mayores que 10^{12} Hz será siempre filtrado en la entrada de nuestro sistema. En contraste, en sistemas ópticos, el ruido térmico es generalmente insignificante a frecuencias ópticas. El ruido térmico se modela como un proceso estocástico Gaussiano, a partir del teorema del límite central, dado que está formado por la superposición de varios eventos independientes (las fluctuaciones térmicas de los electrones individuales).

Correlación cruzada y procesos complejos

Dados dos procesos estocásticos $X(t)$ e $Y(t)$, podemos definir la función *correlación cruzada*

$$R_{XY}(t_1, t_2) = E[X(t_1)Y^*(t_2)]$$

Si $X(t)$ e $Y(t)$ son ESA, entonces serán además *conjuntamente ESA* si $R_{XY}(t_1, t_2)$ es solo función de $(t_1 - t_2)$.

Un proceso estocástico complejo $X(t)$ se define como

$$X(t) = Re\{X(t)\} + j\, Im\{X(t)\}$$

donde $Re\{X(t)\}$ e $im\{X(t)\}$ son procesos reales. La estadística de segundo orden de tal proceso consiste en dos funciones autocorrelación asociadas a las partes real e imaginaria, así como también sus funciones correlación cruzada.

Procesos aleatórios filtrados

Una realización particular x_k o $x(t)$ de un proceso estocástico es una señal y en consecuencia puede filtrarse. Es posible filtrar también el proceso estocástico X_k ó $X(t)$ en si mismo, antes que una realización. Se tendrá entonces un nuevo proceso estocástico con un espacio muestral obtenido aplicando cada elemento del espacio muestral del proceso original a la entrada del filtro.

Ejemplo: Un proceso estocástico Gaussiano filtrado es un proceso estocástico Gaussiano. Intuitivamente, esto es cierto debido a que el filtrado es lineal, y cualquier combinación lineal de variables aleatórias Gaussianas conjuntas es una variable aleatória Gaussiana.

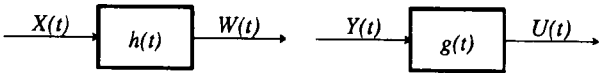

Figura 2.10: Dos sistemas lineales con procesos estocésticos ESA como entrada.

Es posible considerar ahora los dos sistemas contínuos lineales invariantes en el tiempo mostrados en la figura 2.10 con procesos ESA como entradas.

Ejemplo: Es posible mostrar que la salida de un filtro $h(t)$ es ESA, y que su función autocorrelación y densidad espectral de potencia están dadas por

$$R_W(\tau) = h(\tau) * h^*(-\tau) * R_X(\tau), \qquad S_W(j\omega) = S_X(j\omega)|H(j\omega)|^2$$
$$S_W(s) = S_X(s)H(s)H^*(-s^*)$$

Ejemplo: Es posible mostrar que si un proceso discreto ESA se filtra con una respuesta impulsiva h_k, y el resultado es W_k, entonces W_k es ESA y

$$R_W(m) = h_m * h^*_{-m} * R_X(m), \qquad S_W(e^{j\omega T}) = S_X(e^{j\omega T})|H(e^{j\omega T})|^2$$
$$S_W(z) = S_X(z)H(z)H^*(1/z^*)$$

Ejemplo: Un proceso blanco $X(t)$ tiene densidad espectral $S_X(j\omega) = N_0$ constante. Si es filtrado por un filtro pasabajos con función transferencia

$$H(j\omega) = \begin{cases} 1; & |\omega| < 2\pi \times 10^{12} \\ 0; & \forall\omega \end{cases}$$

entonces la densidad espectral de potencia de la salida es:

$$S_W(j\omega) = \begin{cases} N_0; & |\omega| < 2\pi \times 10^{12} \\ 0; & \forall\omega \end{cases} \tag{2.6}$$

lo cual es una aproximación razonable del ruido térmico en una resistencia. Además, dado que el ruido térmico es el efecto acumulativo del movimiento aleatório de un inmenso número de particulas individuales, es posible aplicar el teorema del límite central para argumentar que una muestra de tal ruido térmico debería ser una variable aleatória Gaussiana. De esta forma es posible concluir que el ruido térmico se modela razonablemente como un ruido Gaussiano blanco.

La *densidad espectral cruzada* de dos procesos conjuntamente ESA en las entradas de los filtros de la figura 2.10 están definidas como la transformada de Fourier de la función correlación cruzada,

$$S_{XY}(j\omega) = \int_{-\infty}^{\infty} R_{XY}(\tau)e^{-j\omega\tau}d\tau, \qquad R_{XY}(\tau) = E[X(t+\tau)Y^*(t)]$$

Muestreo de un proceso estocástico

Un proceso estocástico contínuo de potencia finita $X(t)$ puede muestrearse, obteniendose un proceso estocástico discreto $Y_k = X(kT)$. Dado que que esta operación de muestreo se realiza frecuentemente en los sistemas de comunicaciones digitales, es importante relacionar la estadística de los procesos estocásticos contínuos con la de los procesos estocásticos discretos obtenidos muestreandolos. Suponiendo que $X(t)$ es ESA,

$$E[X(kT)X^*(kT)] = R_X(mT),$$

donde $m = k - i$, tal que el proceso muestreado es ESA con autocorrelación igual a la versión muestreada de la autocorrelación $R_X(\tau)$ asociada a la señal contínua. La densidad espectral de potencia de estos procesos están relacionadas por

$$S_Y(e^{j\omega T}) = \frac{1}{T}\sum_{m=-\infty}^{\infty} S_X(j(\omega - m\frac{2\pi}{T}))$$

Como en el caso determinístico, aparece distorsión por solapamiento (aliasing) cuando el ancho de banda es mayor que la mitad de la frecuencia de muestreo, donde el ancho de banda en este caso está definido en términos de la densidad espectral de potencia.

Ejemplo: Consideremos la aproximación al ruido térmico discutida en algún ejemplo anterior. Deseamos determinar cuando las muestras de ese ruido son no correlacionadas. Si lo son, entonces muestrear ruido térmico conduce a un proceso Gaussiano blanco discreto. De (2.6), la función autocorrelación del ruido limitado en banda $W(t)$ es:

$$R_W(\tau) = N_0\frac{B}{\pi}\frac{\sin(Bt)}{Bt}$$

donde $B = 2\pi \times 10^{12}$, ó 1000 GHz. $R_W(\tau)$ tiene cruces por cero en múltiplos de π/B, lo que implica que las muestras del proceso estocástico tomadas en múltiplos de π/B serán no correlacionadas. O sea,

$$R_W(m\frac{\pi}{B}) = E[W(m\pi/B)W(0)] = N_0\frac{B}{\pi}\delta_m$$

Para esas velocidades de muestreo particulares, en consecuencia, las muestras del modelo de ruido térmico corresponden a un proceso de ruido blanco Gaussiano discreto. En la práctica, es improbable muestrear cualquier señal a frecuencias próximas de B/π ó 2000 GHz. Dado que $|R_W(\tau)|$ disminuye cuando τ aumenta, las muestras a cualquier frecuencia de muestreo razonable son *aproximadamente* no correlacionadas.

Utilizando las técnicas discutidas hasta aquí no deberían existir dificultades en considerar sistemas que mezclen procesos estocásticos contínuos y discretos y además señales determinísticas. Sin embargo, existen algunas excepciones que es necesario tener en cuenta. Es posible considerar un proceso estocástico discreto X_k filtrado por un filtro (contínuo) con respuesta impulsiva $h(t)$. La salida puede escribirse

$$Y(t) = \sum_{m=-\infty}^{\infty} X_m h(t - mT) \tag{2.7}$$

De esta forma se obtiene la *Modulación de Amplitud de Pulsos* (PAM), que se estudiará en capítulos siguientes.

Ejemplo: La transmisión de una secuencia discreta de símbolos dato X_m sobre un canal contínuo en el tiempo toma la forma del proceso estocástico que describe PAM. Supongamos que $h(t)$ es como ilustrada en la figura 2.11a y que X_k es una secuencia aleatória con muestras i.i.d. tomando valores ± 1 con igual probabilidad. Una realización posible se muestra en la figura 2.11b. La primera observación importante es que el proceso $Y(t)$ no es ESA ya que $E[Y(t+\tau)Y(t)]$ no es independiente de t. Por ejemplo,

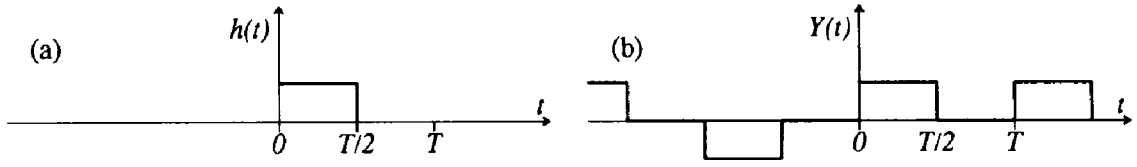

Figura 2.11: a) Ejemplo de una forma de pulso para transmitir bits. b) Un ejemplo de forma de onda para transmitir ese pulso.

$$E[Y(T/4)Y(0)] = E[X_0^2] = 1 \neq E[Y(T)Y(3T/4)] = 0$$

Este proceso en realidad es *cicloestacionario*, una forma más debil de estacionaridad. Dado que este proceso no es ESA, su densidad espectral de potencia no está definida.

Que $Y(t)$ en (2.7) no sea ESA es un inconveniente importante. Un subterfugio simpple permite convertir este proceso en uno ESA. Es posible definir una variable aleatória Θ, uniformemente distribuída en $[0, T]$ e independiente de $\{X_k\}$. Se puede definir luego un nuevo proceso estocástico

$$Z(t) = Y(t + \Theta) = \sum_{m=-\infty}^{\infty} X_m h(t + \Theta - mT)$$

Este proceso tiene una *fase aleatória* constante en el tiempo pero elegida arbitrariamente al inicio de cada intervalo. Fisicamente este nuevo proceso refleja la incertidumbre sobre la fase de la señal. El orígen en el tiempo es por supuesto arbitrário. El proceso así redefinido es ESA, como se discutirá a continuación.

La función autocorrelación del proceso anterior tiene la forma:

$$\begin{aligned} E[Z(t+\tau)Z^*(t)] &= E[Y(t+\Theta+\tau)Y^*(t+\Theta)] \\ &= E[\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} X_m X_n^* h(t+\Theta-mT+\tau)h^*(t+\Theta-nT)] \end{aligned}$$

Suponiendo que se puede intercambiar esperanza y sumatórias, es posible usar la hipótesis que Θ es independiente de X_k para obtener

$$E[Z(t+\tau)Z^*(t)] = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} E[X_m X_n^*]E[h(t+\Theta-mT+\tau)h^*(t+\Theta-nT)] \tag{2.8}$$

El primer valor esperado es simplemente la función autocorrelación $R_X(m-n)$. El segundo valor esperado puede elaborarse usando la definición de esperanza y la pdf uniforme de Θ,

$$E[h(t+\Theta-mT+\tau)h^*(t+\Theta-nT)] = \int_0^T \frac{1}{T} h(t+\theta-mT+\tau)h^*(t+\theta-nT)d\theta$$

Cambiando variables, haciendo $i=m-n$, en (2.8) e intercambiando sumatórias tenemos

$$E[Z(t+\tau)Z^*(t)] = \frac{1}{T}\sum_{i=-\infty}^{\infty} R_X(i) \sum_{n=-\infty}^{\infty}\int_0^T h(t+\theta-(i+n)T+\tau)h^*(t+\theta-nT)d\theta$$

Cambiando variables nuevamente y definiendo $\alpha = t+\theta-nT$, tenemos

$$\begin{aligned} E[Z(t+\tau)Z^*(t)] &= \frac{1}{T}\sum_{i=-\infty}^{\infty} R_X(i) \sum_{n=-\infty}^{\infty}\int_{t-nT}^{t-nT+T} h(\alpha-iT+\tau)h^*(\alpha)d\theta \\ &= \frac{1}{T}\sum_{i=-\infty}^{\infty} R_X(i) \int_{-\infty}^{\infty} h(\alpha-iT+\tau)h^*(\alpha)d\theta \end{aligned} \tag{2.9}$$

donde, la segunda sumatória en la última ecuación es la suma de integrales con limites que pueden extenderse ilimitadamente. Además (2.9) es independiente de t, tal que el proceso $Z(t)$ es ESA. Para obtener la densidad espectral de potencia, se calcula la transformada de Fourier con τ como índice de tiempos

$$S_Z(j\omega) = \frac{1}{T}\sum_{i=-\infty}^{\infty} R_X(i) \int_{-\infty}^{\infty} h^*(\alpha)\left[\int_{-\infty}^{\infty} h(\alpha-iT+\tau)e^{-j\omega\tau}d\tau\right]d\theta$$

La expresión entre corchetes es la transformada de Fourier de $h(t)$ con un desplazamiento de tiempo de $\alpha-iT$, igual a $e^{j\omega(\alpha-iT)}H(j\omega)$. En consecuencia,

$$\begin{aligned} S_Z(j\omega) &= \frac{1}{T}H(j\omega)\sum_{i=-\infty}^{\infty} R_X(i)\left[\int_{-\infty}^{\infty} h^*(\alpha)e^{j\omega(\alpha-iT)}d\theta\right] \\ &= \frac{1}{T}H(j\omega)H^*(j\omega)\sum_{i=-\infty}^{\infty} R_X(i)e^{-j\omega iT} \end{aligned}$$

donde se usó que la expresión entre corchetes es $e^{-j\omega T}H^*(j\omega)$. Finalmente, como la sumátoria restante es la transformada discreta de Fourier $S_X(e^{j\omega T})$ de la función autocorrelación, se obtiene

$$S_Z(j\omega) = \frac{1}{T}|H(j\omega)|^2 S_X(e^{j\omega T})$$

Notar la dependencia en la densidad espectral de potencia del proceso estocástico discreto y la magnitud al cuadrado del espectro del pulso $h(t)$.

Ejemplo: Considerar la transmisión de una secuencia de variables aleatórias X_k con valores ± 1 igualmente probables, que utilizan un pulso de forma $h(t)$. La secuencia X_k es blanca y la varianza unitária, de forma que la densidad espectral de potencia de la secuencia de datos es

$$S_X(e^{j\omega T}) = 1$$

y la densidad espectral de potencia de la señal de fase aleatória transmitida es

$$S_Z(j\omega) = \frac{1}{T}|H(j\omega)|^2$$

Entonces, con una secuencia de datos blanca, la densidad espectral de potencia tiene la forma de la magnitud al cuadrado de la transformada de Fourier del pulso.

2.2.3 Cadenas de Markov

Un *proceso discreto de Poisson* $\{\Psi_k\}$ es un proceso estocástico que satisface

$$p(\psi_{k+1}|\psi_k, \psi_{k-1}, \cdots) = p(\psi_{k+1}|\psi_k)$$

En otras palabras, si la muestra presente $\Psi_k = \psi_k$ es conocida, la realización futura Ψ_{k+1} es independiente de las realizaciones pasadas Ψ_{k-1}, Ψ_{k-2}, $\cdots$. Cuando se considera el caso particular de un proceso de Markov donde las muestras toman valores de un conjunto discreto y cuantificable Ω_Ψ, el proceso se denomina *cadena de Markov*. En esta sección se considerará usualmente que Ω_Ψ es un conjunto de enteros. Las cadenas de Markov son un modelo útil de una máquina de estados finitos con una entrada aleatória, donde las muestras de la entrada son estadísticamente independientes. Dado que cualquier circuito digital con memoria interna (flip flops, registros o RAM) es una máquina de estados finitos, muchos sistemas de comunicaciones digitales contienen máquinas de estado finitos. Las cadenas de Markov son útiles para modelar la generación de señales para sistemas de comunicaciones digitales con interferencia intersímbolos o codificación convolucional. La teoría de cadenas de Markov también es útil para describir la densidad espectral de potencia de los formatos de señalización usuales. La discusión siguiente utiliza técnicas de transformada Z.

Diagramas de Transición de estados

Consideremos un proceso estocástico Ψ_k (real, complejo o vectorial) cuyas realizaciones son miembros de un conjunto de valores finito o cuantificable Ω_Ψ. El proceso estocástico Ψ_k es una cadena de Markov si se satisface la condición de definición. Dada la muestra presente Ψ_k, la muestra próxima de una cadena de Markov Ψ_{k+1} es independiente de las muestras pasadas Ψ_{k-1}, Ψ_{k-2}, $\cdots$. Además, *todas* las muestras futuras de la cadena de Markov son independientes de las pasadas con el conocimiento de la presente. Para $n > 0$, esto puede escribirse como

$$p(\psi_{k+n}|\psi_k, \psi_{k-1}, \cdots) = p(\psi_{k+n}|\psi_k)$$

Dado que el conocimiento de la muestra presente Ψ_k hace que las muestras pasadas sean irrelevantes, Ψ_k es todo lo que se precisa para predecir el comportamiento futuro de la cadena de Markov. Por esta razón, Ψ_k se denomina el *estado* de la cadena de Markov en el instante k, y Ω_Ψ es el conjunto de todos los estados posibles.

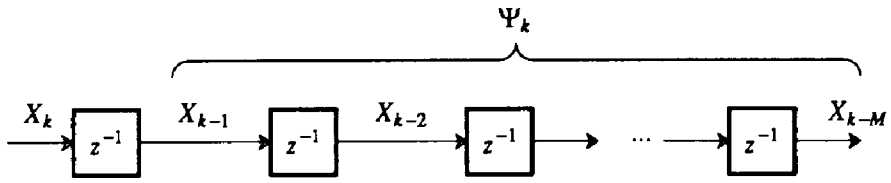

Figura 2.12: Un proceso de registro de desplazamiento con entradas independientes X_k es una cadena de Markov con estados Ψ_k.

Ejemplo: Un proceso denominado *registro de desplazamiento* se muestra en la figura 2.12. Si X_k es independiente de X_{k-M-1}, X_{k-M-2}, $\cdots$ y definimos el vector $\Psi_k = [X_{k-1}, \cdots, X_{k-M}]$, donde M es la memoria del sistema, entonces la propiedad de definición de la cadena de Markov es satisfecha. Esto es debido a que Ψ_{k+1} es función solamente de X_{k+1} y Ψ_k. En consecuencia $\{\Psi_k\}$ es un proceso de Markov vectorial real discreto. Si las entradas X_k tienen valores discretos, entonces esta será una cadena de Markov.

Una cadena de Markov puede describirse graficamente por un *diagrama de transición de estados*. Este gráfico muestra cada estado de la cadena de Markov como un nodo, y muestra también la entrada y la salida o alguna otra propiedad relevante de las transiciones entre estados.

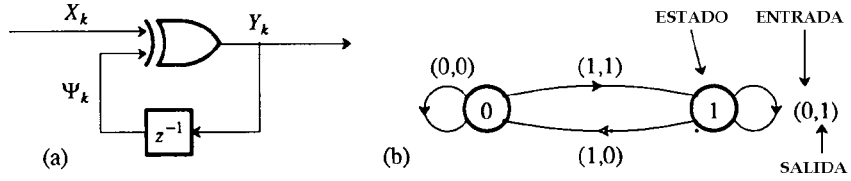

Figura 2.13: a) Un circuito que calcula la paridad de una secuencia de bits X_n. b) El diagrama de transición de estados de la cadena de Markov correspondiente.

Ejemplo: La *paridad* de una secuencia de bits X_k se define como el valor acumulado, módulo 2, de la sumatoria de sus bits, y se calcula mediante el circuito de la figura 2.13a. Para que el proceso de salida $\Psi_k = Y_{k-1}$ sea de Markov es suficiente que los bits de entrada sean independientes. Es fácil ver del diagrama que Ψ_{k+1} depende solo del estado actual Ψ_k y de la entrada actual X_k. Ψ_k tiene un espacio muestral finito $\Omega_\Psi = [0, 1]$, de forma que el verificador de paridad puede representarse por medio de un diagrama de transición de estados como ilustrado en la figura 2.13b, donde los arcos están rotulados con la entrada que estimula la transición del estado y la salida resultante de la transición. Si las probabilidades de transición son independientes del tiempo, los arcos del diagrama de estados pueden rotularse alternativamente con esas probabilidades.

Una cadena de Markov Ψ_k se denomina *homogénea* si la probabilidad condicional $p(\psi_k, |\psi_{k-1})$ no es función de k. La homogeneidad es entonces otra forma de estacionaridad o invarianza en el tiempo. Una cadena de Markov homogénea puede caracterizarse por sus *probabilidades de transición de estados*, lo que puede escribirse como

$$p(j|i) = p_{\Psi_{k+1}|\Psi_k}(j|i)$$

para $i, j \in \Omega_\Psi$.

Ejemplo: Si en el ejemplo anterior los bits entrantes no son solo independientes sino también distribuídos identicamente, entonces la cadena de Markov es homogénea. Si además los bits entrantes son equiprobables, entonces las probabilidades de transición de estados son todas iguales a 0.5.

Es usualmente conveniente definir un proceso de Markov como alguna función real de la trayectoria de estados de la cadena de Markov, como

$$X_k = f(\Psi_k)$$

Este problema se encuentra en el modelado de técnicas de señalización para la transmisión en banda base.

Respuesta transitória de una Cadena de Markov

Para una cadena de Markov homogénea es posible obtener la relación de la evolución de los estados en el tiempo. Usando (2.5) escribimos

$$p_{k+1}(j) = \sum_{i \in \Omega_\Psi} p(j|i) p_k(i) \tag{2.10}$$

para todo $j \in \Omega_\Psi$, donde se ha definido la notación para la probabilidad de estar en el estado i en el instante k como

$$p_k(i) = P\{\Psi_k = i\} \tag{2.11}$$

La nueva notación enfatiza que $p_k(i)$ es una secuencia discreta en el tiempo. En las aplicaciones se desea frecuentemente determinar la probabilidad de estar en cierto estado j en cierto instante de tiempo k, dado el conjunto de probabilidades de estar en los estados en el instante inicial $k = 0$. Esto se logra analizando (2.10), o sea un sistema de ecuaciones a diferencias *invariante en el tiempo*, utilizando técnicas de transformada Z. Si definimos $p_k(j) = 0$ para $k < 0$, entonces la transformada Z de la probabilidad de los estados para el estado j será

$$P_j(z) = \sum_{k=0}^{\infty} p_k(j) z^{-k}$$

En particular, en base a la transformada Z de ambos lados de (2.10) es posible mostrar que

$$P_j(z) = p_0(j) + \sum_{i \in \Omega_\Psi} p(j|i) z^{-1} P_i(z) \tag{2.12}$$

Si existen N estados, (2.12) determina N ecuaciones con N incógnitas $P_j(z)$. Esas ecuaciones pueden resolverse y calcular la transformada Z inversa para determinar la probabilidad de estado $p_k(i)$.

Ejemplo: Continuando con el ejemplo del verificador de paridad, supongamos que el estado inicial es equiprobable, o sea, tal que $p_0(0) = p_0(1) = 0.5$. Supongamos además que los bits de entrada X_k son equiprobables, tal que las probabilidades de transición $p(j|i)$ son todas 0.5. Entonces, de (2.12) se tiene que

$$\begin{aligned} P_0(z) &= 0.5 + 0.5 z^{-1} P_0(z) + 0.5 z^{-1} P_1(z) \\ P_1(z) &= 0.5 + 0.5 z^{-1} P_1(z) + 0.5 z^{-1} P_0(z) \end{aligned}$$

Resolviendo este conjunto de ecuaciones, se tiene que

$$P_0(z) = P_1(z) = \frac{0.5}{1 - z^{-1}}$$

y usando la transformada Z inversa, se tiene

$$p_k(0) = p_k(1) = 0.5\, u_k$$

donde u_k es la función escalón unitário. La cadena tiene en consecuencia igual probabilidad de estar en cualquiera de los dos estados en cualquier instante de tiempo, comenzando en $k = 0$. Una cadena de Markov en la cual las probabilidades de estado son independientes del tiempo se denomina *estacionaria.*

2.2.4 El proceso de Poisson (*)

Existió un tiempo en el que no existía proceso estocástico que pudiera competir con los procesos Gaussianos en su aplicación para el estudio de los sistemas de comunicaciones. Sin embargo, el *proceso de Poisson*, y su generalización el *proceso de nacimientos y muertes*, pueden hoy razonablemente reclamar actualmente esa posición. El problema de aplicación aparece frecuentemente tanto en comunicaciones como en la distribución de tiempos de eventos discretos, tal como en la llegada de mensages a un multiplexador de comunicaciones digitales, o la llegada de fotones en el haz de luz de un detector óptico en un sistema de comunicaciones óptico. El proceso de Poisson modela la más aleatória de las distribuciones y es un excelente modelo para varias de esas situaciones.

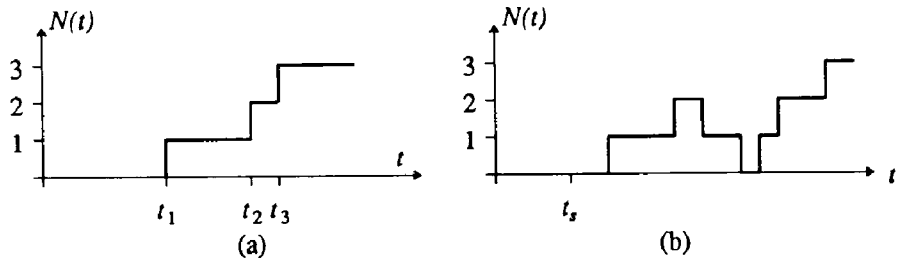

Figura 2.14: Realizaciones típicas de un proceso de conteo $N(t)$. a) Un proceso de conteo monótono creciente. b) Un proceso de conteo con llegadas y partidas, o sea que puede aumentar o disminuir.

Para proceder, es necesario definir la noción de *puntos aleatórios en el tiempo*, donde un punto en el tiempo puede definir la llegada de un mensage de una fuente aleatória (o un fotón) a un fotodetector. Para introducir la notación, definimos t_k como el instante de la k-ésima llegada, donde $t_k \geq t_j$ para $k > j$. Además, se define un proceso estocástico contínuo $N(t)$ que describe el número de llegadas a partir de algún instante inicial t_0. Se denomina $N(t)$ a un *proceso de conteo*, dado que cuenta el número acumulado de puntos aleatórios en el tiempo. $N(t)$ toma valores no negativos enteros, tiene una condición inicial $N(t_0) = 0$, y en cada punto aleatório en el tiempo t_k, $N(t)$ se incrementa en uno. Este proceso de conteo se ilustra en la figura 2.14a, donde los instantes de llegada y el valor del proceso de conteo se muestran para una realización típica.

En algunas situaciones existirán solo llegadas, de forma que el proceso de conteo del tipo mostrado en la figura 2.14a es el modelo apropiado. En otras situaciones existirán tanto llegadas como partidas. Una situación típica es la *cola* ilustrada en la figura 2.15. Podemos definir un proceso de conteo $N(t)$ como la diferencia entre el número acumulado de llegadas y el número acumulado de partidas.

Ejemplo: Considerar un sistema de comunicaciones entre computadoras que almacena los mensajes recibidos en un buffer antes de retransmitirlos a algún otro lugar. $N(t)$ determina la cuenta del número de mensages en el sistema en el instante t. Una realización típica de este proceso se ilustra en la figura 2.14b, donde

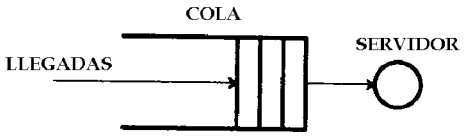

Figura 2.15: Un sistema de cola que modela entre otras cosas el estado de un buffer en un sistema de comunicaciones.

puede notarse que el proceso nunca puede tomar valores debajo de cero (dado que no puede haber partidas si no hay nada en el buffer).

En varias instancias de importancia práctica, la cuenta $N(t)$ en el instante t es todo lo que se necesita para predecir la evolución futura del sistema después del instante t. La forma en la cual el sistema alcanza $N(t)$ es irrelevante en términos de la predicción. Para este caso, el proceso de conteo determina el *estado del sistema* de la misma forma que en las cadenas de Markov de la última sección. En particular, decimos que el sistema está en el estado j en el instante t si $N(t) = j$. Esto es similar a una cadena de Markov con una importante diferencia: una cadena de Markov puede cambiar de estados solamente en puntos discretos en el tiempo. Tal como en las cadenas de Markov, una muestra del proceso de conteo $N(t_0)$ es una variable aleatória discreta. Tal como para las cadenas de Markov de (2.11), se define la probabilidad de estar en el estado j en el instante t como

$$q_j(t) = P[N(t) = j] = P_{N(t)}(j)$$

Esta notación enfatiza que esta probabilidad es una función contínua en el tiempo. La única distinción real entre (2.11) y la ecuación anterior es que esta última está definida para tiempos contínuos y la primera para tiempos discretos.

En las subsecciones siguientes se analizará un proceso de conteo bajo condiciones específicas apropiadas para comunicaciones ópticas.

El proceso de nacimientos y fallecimientos

Los casos de interés están asociados a un proceso general denominado *proceso de nacimientos y fallecimientos*, cuya terminología matemática (ciertamente macabra) resulta del proceso de conteo tanto de llegadas como de partidas. Su análisis se discutirá en esta subsección.

Hasta ahora se ha modelado la evolución del sistema de un estado a otro. En ese sentido la aproximación mediante la cadena de Markov es inapropiada cuando existen muchos estados, dado que la probabilidad de transición entre dos estados cualquiera en cualquier punto en el instante t es muy aproximadamente cero! A pesar que no podemos caracterizar la *probabilidad* de transición, lo que podemos caracterizar es la *velocidad* de transición entre dos estados. Supongamos que para dos estados particulares, la velocidad de transición entre un estado y otro es una constante R. Esto significa que en un instante δt es posible esperar un promedio de $R\,\delta t$ transiciones. Si δt es muy pequeño, entonces $R\,\delta t$ es un número mucho menor que la unidad, y la probabilidad de tener más de una transición en el instante δt es es un número practicamente cero. Bajo esas condiciones, se puede pensar que $R\,\delta t$ es la probabilidad de una transición en el instante δt, y $(1 - R\,\delta t)$ es la probabilidad de que no existan transiciones.

Esta lógica conduce a un diagrama de transición y a un conjunto de ecuaciones diferenciales asociado. El diagrama de transición de la figura 2.16 asocia un nodo a cada estado, y dentro de cada nodo ponemos la probabilidad de estar en ese estado en el instante t, a la cual denominamos $q_j(t)$. Cada transición en el diagrama se rotula con la velocidad a la cual la transición ocurre, donde las velocidades en el caso general se consideran variables en el tiempo (no homogéneas). Cada velocidad se rotula con un subíndice indicando el estado en la cual se origina, donde $\lambda(t)$ es la velocidad para las transiciones correspondientes a los nacimientos o llegadas y $\mu(t)$ corresponde a los fallecimientos o partidas. Reiterando conceptos, la interpretación de esas velocidades es la siguiente: para un intervalo de tiempos muy pequeño δt, la probabilidad de una transición particular es igual a la velocidad por el intervalo de tiempo.

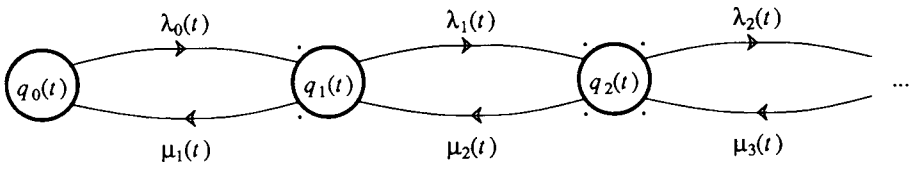

Figura 2.16: Diagrama de transición de estados para un proceso de nacimientos y fallecimientos.

El conjunto de ecuaciones diferenciales que describe la evolución del proceso de nacimientos y fallecimientos es

$$\begin{aligned} \frac{\partial q_j(t)}{\partial t} &= \lambda_{j-1}(t)q_{j-1}(t) + \mu_{j+1}(t)q_{j+1}(t) - (\lambda_j(t) + \mu_j(t))q_j(t), \quad j \geq 0 \\ q_{-1}(t) &= 0 \end{aligned}$$

Estas ecuaciones pueden obtenerse rigurosamente de principios fundamentales, pero para nuestros propósitos son conceptualmente evidentes a partir de consideraciones intuitivas. Las ecuaciones dejan ver que la velocidad de incremento de una probabilidad en el tiempo para el estado j es igual a la velocidad a la cual ocurren las transiciones en ese estado desde los estados $j-1$ y $j+1$ (multiplicada por la probabilidad actual de aquellos estados), menos la velocidad a la cual ocurren las transciones fuera del estado j (multiplicadas por la probabilidad actual del estado j). Es necesario especificar una condición inicial, la cual para nuestros propósitos especifica que el proceso comienza en el estado cero (no existen llegadas) en el instante t_0,

$$q_j(t_0) = \begin{cases} 1, & j = 0 \\ 0, j > 0 \end{cases} \tag{2.13}$$

Las ecuaciones diferenciales de primer orden pueden resolverse para varios casos especiales.

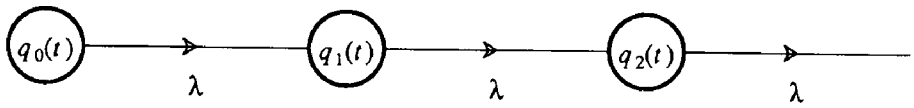

Figura 2.17: Diagrama de transición de estados para un proceso de nacimientos con velocidad constante.

Ejemplo: Considerar el importante caso del *proceso de nacimientos puro*, en el cual $\mu_j(t) = 0$. También suponer que las velocidades de nacimientos son todas iguales a una constante $\lambda_j(t) = \lambda$. El diagrama de transición para este modelo se muestra en la figura 2.17. Este corresponde al importante caso donde la velocidad de llegada no depende del estado del sistema, el caso usual en problemas que encontraremos. De esta forma la ecuación diferencial queda

$$\frac{d\, q_j(t)}{d\, t} + \lambda q_j(t) = \lambda q_{j-1}(t)$$

la cual es una ecuación diferencial simple de primer orden con coeficientes constantes. Suponiendo que la condición inicial es $q_0(0) = 1$, esto implíca que la cuenta inicial en $t = 0$ es 0. Se la puede resolver usando técnicas muy similares a la solución de la cadena de Markov, pero usando transformada de Laplace en lugar de transformada Z. De esa forma, definiendo la transformada de Laplace de la probabilidad de un estado como

$$Q_j(s) = \int_0^\infty q_j(t)e^{-st}dt$$

Tomando la transformada de Laplace de ambos lados de la ecuación diferencial se tiene que

$$sQ_j(s) - q_j(0) + \lambda Q_j(s) = \lambda Q_{j-1}(s)$$

Usando (2.13) con $t_0 = 0$, se tiene que

$$Q_0(s) = \frac{1}{s+\lambda} \qquad Q_j(s) = \frac{\lambda}{s+\lambda} Q_{j-1}(s), \;\; j > 0$$

Este conjunto de ecuaciones recursivo para la transformada de Laplace de la probabilidad de los estados puede resolverse facilmente por iteración como

$$Q_j(s) = \frac{\lambda^j}{(s+\lambda)^{j+1}}$$

y haciendo la transformada de Laplace inversa, encontramos que para $t \geq 0$ y $j \geq 0$ se tiene

$$Pr[N(t) = j] = q_j(t) = \frac{(\lambda\, t)^j}{j!} e^{-\lambda t}$$

Esta se denomina *distribución de Poisson* con parámetro λt. Por esta razón, el proceso de nacimientos puros $N(t)$ que hemos analizado se denomina *proceso de Poisson* con velocidad constante. Se introducirá una generalización para una velocidad variable más adelante.

La distribución de Poisson es muy importante en la teoría de procesos de nacimientos y fallecimientos. Se resumirán sus propiedades en el siguiente ejemplo.

Ejemplo: Considerar una distribución de Poisson con parámetro a, $p_N(k) = e^{-a}\frac{a^k}{k!}$.

- Es posible mostrar (mediante la definición, usando una serie de potencias de e^a y diferenciando dos veces), que la media y varianza de esta distribución son: $E[N] = a$, $Var[N] = a$.
- La función generadora de momentos asociada está dada por

$$\log_e \Phi_N(s) = a(e^s - 1)$$

En el siguiente ejemplo se introducirá una generalización con respecto a la condición inicial.

Ejemplo: Es posible mostrar que si $N(t_0) = k$ (o sea hubo k cuentas hasta el instante t_0), entonces

$$q_j(t) = \frac{(\lambda\,(t - t_0))^{j-k}}{(j-k)!} e^{-\lambda(t-t_0)}, \;\; j \geq k,\; t \geq t_0$$

Este resultado implica que el número de cuentas comenzando en $t = t_0$ tiene una distribución de Poisson con parámetro $\lambda(t - t_0)$, el cual es el número de llegadas esperado desde el instante de tiempo inicial. Además, el índice $j - k$ de la distribución de Poisson es el número de cuentas desde el instante inicial. La conclusión importante es que el número de llegadas en el intervalo que comienza en $t = t_0$ tiene una distribución que no depende de lo que sucede antes de t_0. Esto coincide con la definición rigurosa de un proceso de Markov, y un proceso de conteo de Poisson es en realidad un proceso de Markov. Para ese proceso, el número de llegadas

en el intervalo $[t_0, t]$ es estadísticamente independiente del número de llegadas en cualquier otro intervalo de tiempo no superpuesto. En este sentido el proceso de Poisson es el más aleatório entre todos los procesos monotonos no decrecientes de conteo.

Ejemplo: (Proceso de fallecimientos puro) Para $\lambda_j = 0$, consideremos el caso donde las partidas del sistema son proporcionales al índice del estado, $\mu_j(t) = j\mu$. Este es un modelo apropiado para un sistema en el cual la velocidad de partidas o fallecimientos es proporcional al tamaño de la población, como en una población humana. Además, se supone que el estado inicial en $t = 0$ es n. Es posible mostrar que las probabilidades de estado obedecen una distribución binomial de la forma

$$Pr[N(t) = j] = q_j(t) = \left[\begin{array}{c} n \\ j \end{array} \right] p^j(t)(1 - p(t))^{n-j}, \quad p(t) = e^{-\mu t}$$

Un ejemplo donde ocurren nacimientos y fallecimientos es el *problema de cola*. Para introducirlo se discutirá alguna terminología, particularmente relacionada con comunicaciones digitales. Una cola es un buffer o memoria que almacena mensages. Existirá algún mecanismo que elimina mensajes de la cola, usualmente transmitiendolos a otro lugar. Este mecanismo se denomina el *servidor* de la cola. Se supondrá que el servidor puede acceder a un mensaje de la cola por vez, de forma que si más de un mensaje está siendo procesado (o sea, existen múltiples canales de comunicación para la transmisión de los mensajes), entonces existirá un número equivalente de servidores. Típicamente el buffer contiene espacio para un máximo número de mensajes que esperaran por servicio, y el número de mensajes que pueden esperar en cualquier momento se denomina *número de posiciones de espera*. El *estado del sistema*, que basicamente es un proceso de conteo, es el número de mensajes esperando por servicio más el número de mensajes siendo servidos. Los mensajes llegan a la cola (nacimientos) en instantes aleatórios, y parten de la cola (fallecimientos) debido a la terminación del servicio.

Proceso de Poisson con velocidad de transición variable

En los sistemas de comunicaciones ópticas, el proceso de conteo que determina el número acumulado de tiempos de llegada de los fotones es un proceso de Poisson. El proceso de Poisson es un proceso de nacimientos puro cuya velocidad de llegadas es independiente del estado del sistema, como se discutió anteriormente. En comunicaciones ópticas, la velocidad de llegadas es en realidad dependiente de la señal, de forma que en esta subsección se discutirá ese caso.

El proceso de Poisson con velocidad variante en el tiempo es un proceso de nacimientos puro en el cual la velocidad de llegadas $\lambda(t)$ es independiente del estado del sistema. De esta forma, el sistema está gobernado por una ecuación diferencial de primer orden con coeficientes variantes en el tiempo, de la siguiente forma

$$\frac{dq_j(t)}{dt} + \lambda(t)q_j(t) = \lambda(t)q_{j-1}(t), \quad q_{-1}(t) = 0$$

Se supondrá además que el sistema comienza en el instante t_0 en el estado $j = 0$. Debido a los coeficientes variantes en el tiempo, la transformada de Laplace no es útil y será necesario resolver la ecuación diferencial directamente. Se define

$$\Lambda(t) = \int_{t_0}^{t} \lambda(u)du$$

cuya interpretación es la del promedio del número total de llegadas en el intervalo $[t_0, t]$. Entonces la probabilidad de n llegadas en el intervalo $[t_0, t]$ está gobernada por un proceso de Poisson con parámetro $\Lambda(t)$,

$$q_n(t) = \frac{\Lambda^n(t)}{n!} e^{-\Lambda(t)}$$

Esto se reduce a la solución de aquel ejemplo de la subsección anterior para el caso de velocidad de transición constante. Esta ecuación, además especifica el número $N(t)$ de llegadas durante el intervalo $[t_0, t]$. Este número aleatório de llegadas tiene una distribución de Poisson con parámetro $\Lambda(t)$, de forma que su media y varianza serán

$$E[N(t)] = \Lambda(t), \qquad \sigma_N^2(t) = \Lambda(t)$$

El ruido Shot

En las comunicaciones ópticas, la señal se genera en el fotodetector produciendo impulsos en instantes correspondientes a las llegadas aleatórias de fotones y luego filtrando esos impulsos. Esto se conoce como un *proceso filtrado de Poisson*, o *proceso de ruido shot.*

Si un proceso de Poisson está caracterizado por un conjunto de tiempos de llegada t_k para la llegada k-ésima, y dado un filtro con respuesta impulsiva $h(t)$, entonces el proceso de ruido shot es un proceso estocástico contínuo $X(t)$ con una realización dada por

$$x(t) = \sum_k h(t - t_k)$$

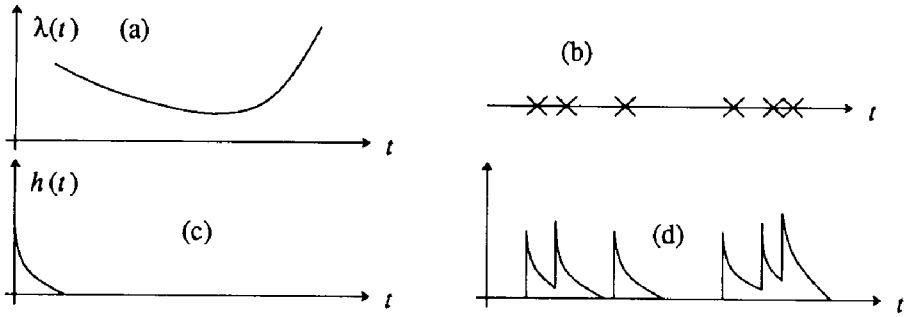

Figura 2.18: Ilustración de un proceso de ruido shot. a) La velocidad promedio de llegadas. b) Los tiempos aleatórios reales de llegada, con llegadas ocurriendo a la velocidad promedio en a). c) La respuesta impulsiva del filtro. d) La realización correspondiente.

Una realización de este proceso se ilustra en la figura 2.18 para una respuesta impulsiva particular. En esta figura se supone cualitativamente que la duración de la respuesta impulsiva es corta si comparada con el tiempo promedio entre llegadas. Si la respuesta impulsiva fuera larga, esto tendría un efecto promedio suavizador sobre la realización.

Es posible mostrar que la función generadora de momentos del proceso de ruido shot en el instante t es

$$\log_e \Phi_{X(t)}(s) = \lambda(t) * (e^{sh(t)} - 1)$$

Ejemplo: Es posible mostrar que el valor medio del ruido shot es la convolución de la respuesta impulsiva del filtro con la velocidad de llegadas,

$$m_X(t) = E[X(t)] = \lambda(t) * h(t)$$

y que la varianza es la convolución del cuadrado de la respuesta impulsiva del filtro con la velocidad de llegadas,

$$\sigma_X^2(t) = E[X^2(t)] - m_X^2(t) = \lambda(t) * h^2(t)$$

Estas relaciones se conocen con el nombre de *Teorema de Campbell.*

El ruido Shot de alta intensidad

Cuando la intensidad del ruido shot es alta, la estadística se vuelve la de un proceso estocástico Gaussiano. La intuición detrás de esto es que $X(t)$ es la suma de un gran número de eventos independientes, y en consecuencia se aproxima a un proceso Gaussiano por el Teorema del Límite Central. Para demostrar esto más formalmente, se verificará que la función generadora de momentos del ruido shot se aproxima a la de un proceso Gaussiano en el límite de alta intensidad.

Para evitar un ruido shot de potencia infinitamente grande cuando crece la intensidad será necesario escalar el tamaño de la respuesta impulsiva $h(t)$. En consecuencia, usaremos una constante de escalamiento β, la que haremos tender a infinito, de forma que

$$\lambda(t) = \beta\lambda_0(t), \qquad h(t) = \frac{1}{\sqrt{\beta}} h_0(t)$$

Con este escalamiento, del teorema de Campbell se tiene que

$$m_X(t) = \sqrt{\beta}\lambda_0(t) * h_0(t), \qquad \sigma_X^2(t) = \lambda_0(t) * h_0^2(t)$$

De esta forma, cuando el factor β crece, la varianza del proceso se mantiene constante y la media crece sin límite. En ese sentido el ruido shot de alta intensidad se aproxima a una señal determinística cuando la intensidad crece.

Solamente dos términos de la función generadora de momentos son importantes cuando crece la constante de escalamiento β.

Ejemplo: Es posible mostrar que para un β grande los únicos términos significativos en la función generadora de momentos son:

$$\log_e \Phi_{X(t)}(s) \cong \sqrt{\beta}\lambda_0(t) * h_0(t) + 0.5s^2\lambda_0(t) * h_0^2(t)$$

Comparando este resultado con la función generadora de momentos para el caso Gaussiano es posible concluir que el ruido shot de alta intensidad es aproximadamente Gaussiano con media y varianza como vistas anteriormente.

Capítulo 3

Modelos de canales de comunicaciones

3.1 Medios físicos y canales

El diseño de los sistemas de comunicaciones digitales depende de las propiedades del canal. El canal es tipicamente parte del sistema de comunicaciones que no es posible alterar. Algunos canales son simplemente un medio físico, como por ejemplo pares de cobre o fibra óptica. Por otro lado, el canal de radio es parte del espectro electromagnético, el cual está dividido por regulaciones gubernamentales en canales limitados en frecuencia de radio que ocupan bandas de frecuencia disjuntas. No discutiremos en general el diseño de *transductores*, tal como antenas, lasers, fotodetectores, etc. y los consideraremos en particular como parte del canal. Algunos canales, particularmente el canal telefónico, son realmente composiciones de varios subsistemas de transmisión. Estos *canales compuestos* determinan sus características de las propiedades de los subsistemas componentes.

Los medios prevalecientes para nuevas instalaciones en el futuro serán la fibra óptica y radio de microondas y probablemente líneas de transmisión sin pérdidas basadas en materiales superconductores. Sin embargo, existe un contínuo interés en lineas de transmisión con pérdidas y canales de voz debido a la existencia de inmensas instalaciones.

3.1.1 Canales compuestos (*)

Es frecuente que varios usuarios compartan un medio común de comunicaciones, por ejemplo por medio de *multiplexado por división en tiempo* (TDM) ó *multiplexado por división en frecuencia* (FDM).

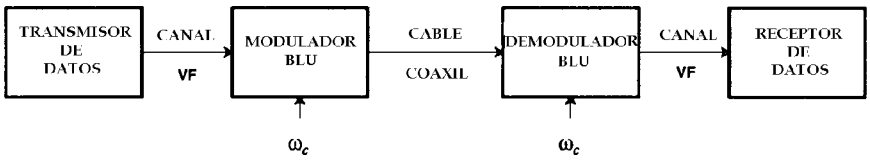

Figura 3.1: Canal de datos compuesto obtenido de un sistema de modulación SSB, con ω_c la frecuencia de la portadora.

Ejemplo 1: Los canales de voz están estrictamente limitados en frecuencia a valores menores de 4 kHz. Un canal banda base adecuado requiere pasar solo frecuencias hasta 4 kHz. Tal canal se obtiene frecuentemente de un medio físico de ancho de banda mucho mayor que se comparte con otros usuarios. Un canal de frecuencias de voz (VF) se obtiene de un cable coaxil usando *modulación de banda lateral única* como mostrado en la figura 3.1. El modulador BLU translada el canal VF a la vecindad de una frecuencia ω_c para transmisión sobre el cable coaxil. Un canal VF puede utilizarse para comunicaciones digitales siempre que la técnica de modulación se adecue a las limitaciones del canal.

El canal del ejemplo anterior es un ejemplo de *canal compuesto*, dado que está formado por varios

subsistemas. Si el canal VF fué diseñado para la transmisión de voz tendrá ciertas características que están más allá del control del diseñador del sistema de comunicaciones digitales. Las características del canal VF en este caso dependen no solo de las propiedades del medio físico sino también del diseño del sistema de modulación BLU.

Los canales compuestos aparecen usualmente en el contexto de un *acceso múltiple a un recurso*, el cual está definido como el acceso a un medio físico por parte de dos o más usuarios independientes. Esto se ilustrará también con un ejemplo.

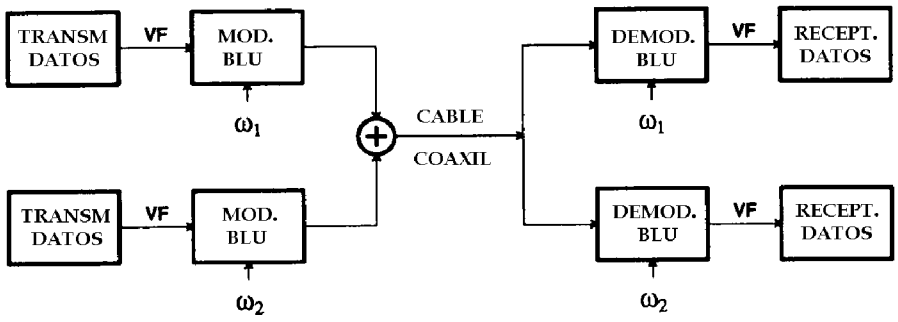

Figura 3.2: Dos canales de datos obtenidos de un único cable coaxil por medio de FDM, donde ω_1 y ω_2 son frecuencias de portadora distintas. TX es el transmisor y RX es el receptor.

Ejemplo 2: La figura 3.2 ilustra la solución FDM utilizando modulación BLU. En este caso se obtienen dos canales VF de un único cable coaxil utilizando modulación BLU. Los dos canales se separan usando dos portadoras diferentes en los moduladores, ω_1 y ω_2, cuyas frecuencias se eligen lo suficientemente apartadas tal que el espectro de los dos canales VF no se solapen. Estos dos canales pueden utilizarse independientemente para comunicaciones digitales.

En realidad FDM es una técnica muy común en redes telefónicas para obtener varios canales VF de un único medio físico tal como cable coaxil o radio microondas (miles antes que dos como en el ejemplo!).

Otro canal compuesto frecuente se ilustra en el siguiente ejemplo.

Figura 3.3: Transmisión de datos sobre un canal VF obtenido de PCM.

Ejemplo 3: Un canal VF obtenido de un sistema de transmisión digital que utiliza PCM se ilustra en la figura 3.3. El sistema PCM muestrea el canal VF a 8 kHz, correspondiente a un máximo ancho de banda de 4 kHz, y luego cuantiza cada muestra con 8 bits. La velocidad de transmisión total para el canal VF codificado por PCM es de 64 kb/seg. El canal VF obtenido puede utilizarse para la transmisión de datos; nuevamente, cualquier técnica de modulación digital puede utilizarse sujeta a las restricciones impuestas por el sistema PCM. La velocidad de transmisión total que puede utilizarse en el canal VF obtenido es menor que 64 kb/seg, en realidad del orden de 20-30 kb/seg. La transmisión directa de la secuencia de bits por el sistema de transmisión digital debería ser más eficiente, pero la situación de la figura 3.3 es aún muy común debido a la presencia de numerosos equipamientos existentes de PCM para la transmisión de voz y el deseo de transmitir datos sobre un canal diseñado inicialmente para voz.

El medio físico y los canales compuestos obtenidos de él imponen restricciones sobre el diseño de un

sistema de comunicaciones digitales. Varias de esas restricciones se mencionarán en las discusiones que siguen para diferentes medios. La naturaleza de esas restricciones usualmente puede clasificarse en dos grandes categorías:

- *Restricción de ancho de banda*: Algunas veces esta aparece en la forma de una atenuación de canal que aumenta gradualmente en altas frecuencias, y otras (particularmente en el caso de canales compuestos) en la forma de restricciones fuertes de ancho de banda.
- *Restricción de potencia transmitida*: Aparece frecuentemente para limitar la interferencia de un sistema de comunicaciones digitales en otro, o impuesta por la incapacidad de un canal compuesto para transmitir un nivel de potencia mayor que cierto umbral, o por una limitación impuesta por la fuente de alimentación del sistema de comunicaciones digitales mismo. Esta restricción de potencia puede aparecer en la forma de una *limitación de potencia pico*, la cual es una restricción sobre el voltage transmitido, o puede aparecer como una *limitación de potencia promedio.*

Ejemplo 4: Un sistema FDM tal como el de la figura 3.2, donde existen tal vez miles de canales multiplexados se diseña bajo suposiciones de potencia promedio de los canales. De esta potencia promedio es posible deducir la potencia total de la señal multiplexada. Esta potencia se ajusta en relación al punto en que los amplificadores del sistema comienzan a volverse no lineales. Si un número significativo de canales VF viola la restricción de potencia promedio, entonces la señal multiplexada sobrecargará los amplificadores y la nolinealidad resultante causará *distorsión de intermodulación* entre canales de VF.

Ejemplo 5: El sistema PCM de la figura 3.3 impone una restricción de ancho de banda sobre la señal de datos, la cual debe ser menor que la mitad de la frecuencia de muestreo. Aparece una restricción de potencia máxima o pico impuesta debido a que el cuantizador del sistema PCM tiene un punto de sobrecarga más allá del cual corta (satura) la señal de entrada.

Ejemplo 6: Los repetidores regenerativos de la figura 1.3 se alimentan usualmente colocando un alto voltage en la terminación del sistema y luego colocando los repetidores en serie (como guirnaldas de Navidad!). Una fracción del voltage total se asocia entonces con cada repetidor. El consumo de potencia de los repetidores está limitado por el voltage aplicado, la pérdida óhmica del cable, y el número de repetidores. Esto introduce una restricción de potencia promedio sobre la señal transmitida en cada repetidor. En la práctica existe también una restricción de potencia pico debido a que en general no se desea generar una señal de voltage mayor que la de caída en cada repetidor. Un factor adicional que limita la potencia transmitida para sistemas cableados es la interferencia recíproca (crosstalk o acoplamiento) en otros sistemas de comunicaciones en el mismo cable multipar.

Ejemplo 7: La fibra óptica se vuelve significativamente no lineal cuando la potencia de entrada excede valores de un mW. Así en varias aplicaciones existe un limite práctico sobre la potencia promedio transmitida.

Además de imponer restricciones sobre la señal transmitida, el medio o el canal compuesto introduce *desajustes* que limitan la velocidad a la cual es posible utilizar las comunicaciones. Se discutirán varios ejemplos a continuación.

3.1.2 Líneas de transmisión (*)

Uno de los medios más comunes para transmisión de datos en el pasado ha sido la línea de transmisión compuesta por un par de conductores o un cable coaxil. El cable coaxil se utiliza comunmente para comunicaciones digitales dentro de edificios, tal como en LAN, y para facilidades de alta capacidad y larga distancia en la red telefónica. Los pares de conductores se utilizan mucho más extensivamente, primariamente para el entroncamiento asociado a sistemas de conmutación en distancias cortas de la red telefónica

de área metropolitana. El espaciamiento entre repetidores regenerativos es tipicamente de alrededor de 1.5 km, con velocidades de transmisión en el rango de 1.5 a 6 Mb/seg, en el caso del par de conductores, hasta 270 a 400 Mb/seg, en el caso de cable coaxil. Además, los pares de conductores se utilizan para la conexión del aparato telefónico a la central de conmutación, y a pesar que esta conexión es primariamente para la transmisión de voz existe una gran actividad para convertir el mismo medio para la transmisión digital a 144 kb/seg o mayores velocidades usando ISDN. Esta aplicación se denomina *Línea de Abonado Digital* (Digital subscriber Line, DSL) y permite distancias de transmisión sin repetidores de hasta unos 4-5 kms.

Revisión de teoría de lineas de transmisión

Una *línea de transmisión uniforme* es un par conductor con sección uniforme. Puede consistir de un par de conductores *trensados* (par trensado) o un cable con un conductor externo cilíndrico alrededor de otro (cable coaxil). A pesar que los detalles de las características del cable dependen de la geometría de su sección, la teoría básica que permite explicar su desempeño no.

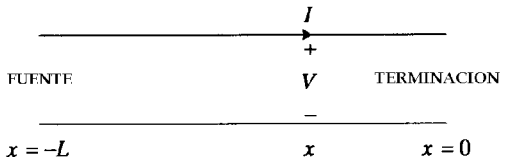

Figura 3.4: Una línea de transmisión uniforme, donde x es la distancia a lo largo de la línea.

Una línea de transmisión uniforme puede representarse por un par de conductores como mostrado en la figura 3.4. La terminación de la línea sobre la derecha se define con $x = 0$, y la fuente a la izquierda con $x = -L$, donde x es la distancia sobre la línea y L es su longitud. Se supone que la línea se excita con una exponencial compleja con frecuencia ω, entonces el voltage y corriente a lo largo de la línea serán función de la frecuencia ω y la distancia x. El voltage y la corriente en un punto x serán:

$$V(x,\omega) = V(x)e^{j\omega t} \qquad I(x,\omega) = I(x)e^{j\omega t}$$

donde $V(x)$ e $I(x)$ son números complejos que representan la amplitud y la fase de la exponencial compleja a la distancia x.

El voltage y la corriente a lo largo de la línea están formados por dos ondas de propagación, una de la fuente a la terminación y la otra en sentido contrário. La primera se denomina *onda fuente* y la segunda *onda reflejada*. Los voltages y corrientes totales serán

$$V(x) = V_+e^{-\gamma x} + V_-e^{\gamma x} \qquad I(x) = \frac{1}{Z_0}(V_+e^{-\gamma x} - V_-e^{\gamma x}) \tag{3.1}$$

donde V_+ y V_- corresponden a las ondas fuente y reflejada, respectivamente. La impedancia compleja Z_0 se denomina *impedancia característica* de la línea de transmisión, dado que es igual a la relación entre el voltage y la corriente en cualquier punto de la línea (independientemente de x) tanto para la onda fuente como para la reflejada. También, γ se denomina *constante de propagación*, y puede escribirse como

$$\gamma = \alpha + j\,\beta$$

donde α es la *constante de atenuación* (con unidades de nepers / unidad de distancia) y β es la *constante de fase* (con unidades de radianes / unidad de distancia).

Tres aspectos distinguen la onda fuente de la reflejada:

- La amplitud y fase asociadas a V_+ y V_- son diferentes.
- La corrientes asociadas tienen direcciones opuestas.
- El signo de las exponentes es diferente.

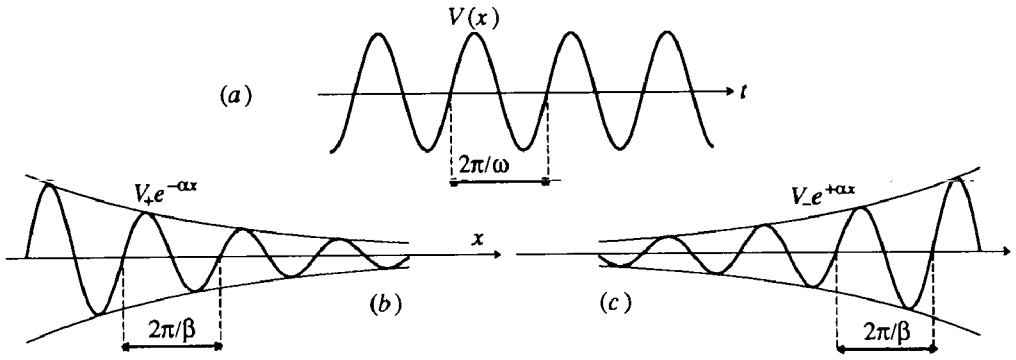

Figura 3.5: El voltage sobre una línea de transmisión uniforme. a) El voltage de un punto de la línea en función del tiempo. b) El voltage en función de la distancia para una onda fuente en un instante de tiempo fijo. c) el caso anterior (b) repetido para la onda reflejada.

La tercera diferencia se ilustra en la figura 3.5. La dependencia del tiempo para ambas ondas en cualquier punto de la línea se muestra en la figura 3.5a. Para un instante de tiempo fijo $t = t_0$, la onda fuente mostrada en la figura 3.5b en función de x está dada por

$$V(x) = V_+ e^{-\alpha x} e^{-j\beta x}$$

cuya amplitud disminuye con la distancia x. La *longitud de onda*, o distancia entre valores nulos es $2\pi/\beta$. El desplazamiento de fase de la exponencial compleja con frecuencia ω para una línea de longitud L es βL radianes, lo cual corresponde a $\beta L/2\pi$ ciclos. Este desplazamiento de fase representa un retardo de propagación asociado a una sinusoide en la línea. Para obtener el tamaño del retardo, y dado que cada ciclo corresponde a $2\pi/\omega$ seg. de la figura 3.5a, es posible tener en cuenta que el retardo total de la exponencial compleja es

$$\frac{\beta L}{2\pi}\text{ciclos} \cdot \frac{2\pi}{\omega}\,\frac{\text{seg}}{\text{ciclos}} = \frac{\beta}{\omega} L \text{ seg.}$$

La velocidad de propagación de la onda sobre la línea de transmisión está entonces relacionada con la frecuencia y la constante de fase por

$$\nu = \frac{\omega}{\beta}$$

Dado que α es siempre mayor que cero, la magnitud de la onda es también una exponencial decreciente con la distancia de la forma $e^{-\alpha x}$. Esto implíca que en cualquier frecuencia la atenuación de la línea en *decibeles* (dB) es proporcional a su longitud. De esta forma tenemos una atenuación de potencia en dB dada por

$$\gamma_0\, L = 20\,\alpha\, \log_{10}\left(\frac{P_T}{P_R}\right)$$

donde $\gamma_0 = 20\alpha \log_{10} e$ es la atenuación en dB por unidad de distancia. Dado que α depende de la frecuencia, también γ_0 dependerá.

En forma similar, en la figura 3.5c se muestra la amplitud de la onda reflejada en función de la distancia a lo largo de la línea. Esta onda es también una exponencial decreciente con la distancia en la dirección de propagación.

En base a estas relaciones es posible determinar el voltage a lo largo de la línea para cualquier impedancia de fuente y terminación simplemente apareando las condiciones de borde.

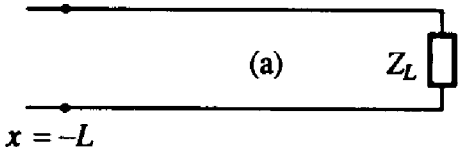

Figura 3.6: Una línea de transmisión terminada a) Sin terminación de fuente, b) Con una terminación de fuente.

Ejemplo 1: Una línea de transmisión terminada en una impedancia Z_L se muestra en la figura 3.6a. Cual es el tamaño relativo de las ondas incidente y reflejada en la terminación? Esta cantidad se denomina *coeficiente de reflexión de voltage*, Γ. La condición de borde es que Z_L es la relación del voltage a la corriente en $x = 0$, tal que de 3.1

$$Z_L = Z_0 \frac{V_+ + V_-}{V_+ - V_-}, \qquad \Gamma = \frac{V_-}{V_+} = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{3.2}$$

Son de interés algunos casos especiales. Cuando la impedancia de carga es igual a la impedancia característica, $Z_L = Z_0$, entonces el coeficiente de reflexión es $\Gamma = 0$. Cuando la línea está abierta $Z_L = \infty$, entonces $\Gamma = 1$ indicando que el voltage reflejado es el mismo que el incidente en el punto de circuito abierto. Finalmente, cuando la línea está cortocircuitada, $Z_L = 0$, $\Gamma = -1$, indicando que el voltage reflejado es el negativo del voltage incidente.

Ejemplo 2: Cual es la impedancia de entrada a la línea de transmisión terminada de la figura 3.6a en función de su longitud? Teniendo en cuenta la relación del voltage a la corriente de la ecuación (3.1) y además (3.2), se obtiene

$$\frac{V(x)}{I(x)} = Z_0 \frac{e^{-\gamma x} + \Gamma e^{\gamma x}}{e^{-\gamma x} - \Gamma e^{\gamma x}}, \qquad Z_{in} = \frac{V(-L)}{I(-L)} = Z_0 \frac{1 + \Gamma e^{-2\gamma L}}{1 - \Gamma e^{-2\gamma L}}$$

Cuando la línea está terminada por su impedancia característica, $\Gamma = 0$, de forma que la impedancia de entrada es igual a la impedancia característica.

Ejemplo 3: Cual es la función transferencia de voltage de la fuente V_{in} a la carga para la línea terminada de la figura 3.6b con una impedancia de fuente Z_S? Escribiendo la ecuación de voltages para la fuente se tiene que

$$V(-L) = V_{in} - I(-L)\, Z_S$$

y teniendo en cuenta que

$$V(-L) = V_+(e^{-\gamma L} + \Gamma e^{-\gamma L}) \qquad I(-L) = \frac{V_+}{Z_0}(e^{-\gamma L} - \Gamma e^{-\gamma L})$$

de donde es posible obtener V_+, $(-L)$ e $I(-L)$. Finalmente, el voltage de salida es

$$V(0) = V_+(1 + \Gamma)$$

tal que

$$\frac{V(0)}{V_{in}} = \frac{Z_0(1+\Gamma)}{(Z_0+Z_S)e^{\gamma L}+\Gamma(Z_0-Z_S)e^{-\gamma L}}$$

Cuando la impedancia de la fuente es igual a la impedancia característica $Z_L = Z_0$, esto implíca que

$$\frac{V(0)}{V_{in}} = \frac{1+\Gamma}{2}e^{-\gamma L}$$

Cuando la línea está terminada por su impedancia característica, $\Gamma = 0$ y la función transferencia consiste en la atenuación y desplazamiento de fase de la línea. Cuando la línea está cortocircuitada, la función transferencia es cero, ya que $\Gamma = -1$. Que sucede cuando la línea está a circuito abierto?

Figura 3.7: Red de dos puertas con la definición de los voltages y corrientes para la matriz cadena.

Las líneas de transmisión se analizan frecuentemente usando el concepto de *matriz cadena*. Una *red de dos accesos*, como ilustrada en la figura 3.7, tiene accesos de entrada y salida para los cuales las corrientes son complementarias. La matriz cadena relaciona el voltage y la corriente de entrada con el voltage y la corriente de salida por

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}$$

donde A, B, C y D son cantidades complejas dependientes de la frecuencia. La matriz cadena caracteriza la función transferencia de la red de dos accesos y puede utilizarse para analizar conexiones de doble acceso (tal como líneas de transmisión, y en consecuencia sirve para analizar líneas de transmisión no uniformes). Su importancia se pone de manifiesto teniendo en cuenta que la conexión en serie de dos redes de doble acceso está caracterizada por una matriz cadena que se obtiene del producto de la asociada con la primera red por la de la segunda.

La matriz cadena de una línea de transmisión uniforme se calcula facilmente, lo que permite disponer de una forma sistemática para analizar combinaciones de líneas de transmisión con diferentes parámetros característicos.

En particular, la matriz cadena para una línea de transmisión uniforme está dada por

$$\begin{bmatrix} \cosh(\gamma L) & Z_0 \sinh(\gamma L) \\ \frac{\sinh(\gamma L)}{Z_0} & \cosh(\gamma L) \end{bmatrix}$$

Constantes primárias del cable

La impedancia característica y la constante de propagación se denominan *parámetros secundarios* de la línea porque no están relacionados directamente con los parámetros físicos. Un modelo simple para una sección de línea de transmisión se muestra en la figura 3.8. Este modelo se describe en términos de cuatro parámetros: la conductancia G en $ohms^{-1}$ / unidad de longitud, la capacidad C en faradios / unidad de longitud, la inductancia L en henrios / unidad de longitud y la resistencia R en ohms / unidad de longitud. Todos estos parámetros son funciones de la frecuencia en general y difieren para diferentes tipos de secciones (por ejemplo: par trensado vs. cable coaxil), y se determinan experimentalmente para cada tipo de cable.

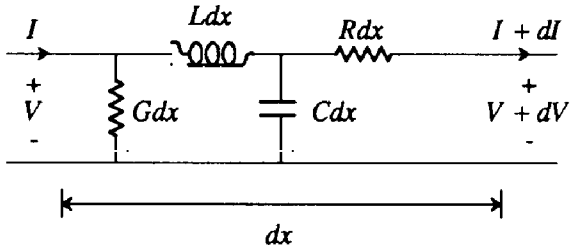

Figura 3.8: Modelo de parámetros distribuidos para una sección breve de línea de transmisión.

Este modelo de parámetros distribuidos se convierte en un modelo exacto de la línea de transmisión cuando $dx \to 0$, y es muy útil ya que indica cantidades de interpretación física directa. Esos parámetros se denominan *constantes primárias* de la línea de transmisión. Los parámetros secundarios pueden calcularse directamente en términos de las constantes primárias como

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}, \qquad \gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{3.3}$$

Ejemplo 1: Una *línea de transmisión sin pérdidas* no tiene los dos elementos disipativos, resistencia y conductancia. Algunos materiales superconductores recientes prometen ser una buena aproximación a este ideal. Los parámetros secundarios en este caso son

$$Z_0 = \sqrt{\frac{L}{C}}, \qquad \gamma = j\omega\sqrt{LC}$$

La impedancia característica de una línea de transmisión sin pérdidas es real y en consecuencia resistiva. Una terminación puramente resistiva se utiliza frecuentemente como aproximación razonable a la impedancia característica *con* pérdidas a pesar que la impedancia característica aumenta a bajas frecuencias e incluye un componente reactivo capacitivo. La constante de propagación es imaginária, y como $\alpha = 0$ la línea no tendrá atenuación como era de esperar. La velocidad de propagación de la línea de transmisión con pérdidas es

$$\nu = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \tag{3.4}$$

Las constantes primárias de cables reales dependen de varios factores tales como la geometría y el material utilizado en la aislación. Para el caso de pares trensados, la capacidad es independiente de la frecuencia para el rango de frecuencias de interés (0.0515 μF/km es un valor típico), la conductancia es despreciablemente pequeña, la inductancia es una función lentamente variable de la frecuencia decreciendo alrededor de 0.62 MH / Km, en bajas frecuencias a alrededor de 70 % ese valor a altas frecuencias, y la resistencia es proporcional a la raíz cuadrada de la frecuencia para altas frecuencias debido al *efecto pelicular* (la tendencia de la corriente a fluir próxima a la superfície del conductor, incrementando la resistencia).

Ejemplo 2: Cual es la velocidad de propagación en un par trensado? De (3.4)

$$\nu = \frac{1}{\sqrt{(0.0515\,10^{-6})(0.62\,10^{-3})}} = 1.76\,10^5\ km/seg$$

Dado que la velocidad de la luz en el espacio libre es de $3\,10^5$ km/seg, la velocidad sobre la línea es un poco mayor que la mitad de la velocidad de la luz. El retardo es de alrededor de 5.65 μseg / km. Esta aproximación es válida para pares trensados en la práctica para frecuencias donde $R << \omega L$.

El cable coaxil es popular en aplicaciones de más altas frecuencias debido basicamente a que el conductor externo evita la radiación hacia el exterior y además protege contra interferencias externas. En bajas frecuencias esta protección no es efectiva y en consecuencia el cable coaxil no tiene ninguna ventaja sobre el más económico par trensado. En términos de las constantes primárias, la principal diferencia entre el cable coaxil y el par trenzado es que la inductancia en el primer caso es independiente de la frecuencia.

Ejemplo 3: Un modelo más preciso que el del ejemplo 1 para par trensado o cable coaxil podría suponer que solo G es cero. En ese caso la constante de propagación de (3.3) tomará la forma

$$\alpha = \omega\sqrt{\frac{LC}{2}}\left\{(1+\frac{R^2}{\omega^2L^2})^{1/2}-1\right\}^{1/2}, \qquad \beta = \omega\sqrt{\frac{LC}{2}}\left\{(1+\frac{R^2}{\omega^2L^2})^{1/2}+1\right\}^{1/2}$$

y en frecuencias donde $R << \omega L$ se tiene que

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}} \text{ (nepers/unid. long.)}, \qquad \beta = \omega\sqrt{LC}$$

De esto es posible concluir que (3.4) es aún válida en este rango de frecuencias. Dado que en altas frecuencias R aumenta con el cuadrado de la frecuencia, la constante de atenuación en dB tendrá la misma dependencia. De esto se concluye que la atenuación de la línea en dB en altas frecuencias es proporcional a la raíz cuadrada de la frecuencia.

Ejemplo 4: Si la atenuación de un cable es de 40 dB a 1 MHz, cuál es la atenuación aproximada a 4 MHz? La respuesta es 40 dB por la raíz cuadrada de 2, u 80 dB.

El resultado del ejemplo 3 sugiere que la constante de propagación es proporcional a la frecuencia. Este *modelo de fase lineal* sugiere que la línea introduce, además de la atenuación, un retardo constante para todas las frecuencias. Sin embargo, un modelo más refinado de la constante de propagación mostraría que existe un término adicional en la constante de fase proporcional a la raíz cuadrada de la frecuencia. Esto implica que las diferentes componentes de frecuencia de un pulso introducido en la línea llegarán a la terminación con diferentes retardos. Tanto la atenuación dependiente de la frecuencia como el retardo de grupo producen *dispersión* sobre la línea de transmisión, de forma de distorsionar en tiempo un pulso transmitido. La atenuación causa dispersión debido al efecto de limitar en banda el pulso, y el retardo produce dispersión debido a que las diferentes componentes de frecuencia llegan con diferentes retardos.

Discontinuidades de impedancia

En la teoría discutida hasta aqui se consideró una única línea de transmisión uniforme. En la práctica es común encontrar líneas con diferentes diámetros. Estos cambios de diámetro no afectan la transmisión materialmente, excepto por introducir pequeñas discontinuidades de impedancia, lo cual puede resultar en pequeñas reflexiones. Un problema más serio para la transmisión digital en la línea de abonado entre la central de conmutación y el usuario es la *derivación* (bridged tap), o sea un elemento para acoplar al usuario (normalmente a circuito abierto) a la línea principal.

Acoplamiento (Crosstalk)

Un aspecto importante en el diseño de sistemas de comunicaciones digitales usando una línea de transmisión como medio físico es el *rango* ó *distancia* que puede ser alcanzado entre repetidores regenerativos. Este rango

está limitado generalmente por la ganancia de alta frecuencia que puede introducirse en la ecualización del receptor para compensar la atenuación de la línea. Esta ganancia amplifica el ruido y las señales de interferencia que puedan estar presentes, haciendo que la señal se deteriore cuando la distancia aumenta. Los ruidos e interferencias más importantes son: el ruido térmico (debido al movimiento aleatório de los electrones), el ruido impulsivo (debido a a la conmutación) y el acoplamiento (crosstalk) entre líneas. El acoplamiento, y la interferencia debido a fuentes externas tal como líneas de alimentación, puede minimizarse utilizando una *transmisión balanceada*, en la cual la señal se transmite y se recibe como una diferencia de voltage en la línea. Esto ayuda debido a que la interferencia externa se acopla en forma aproximadamente similar en los dos conductores y en consecuencia es aproximadamente cancelada cuando en el receptor se tiene en cuenta la diferencia de voltage. Una forma común de lograr la transmisión balanceada es a través del uso de *transformadores de acoplamiento* del transmisor y el receptor. Además, esto permite una protección adicional de la electrónica debido a potenciales extraños tales como rayos, etc.

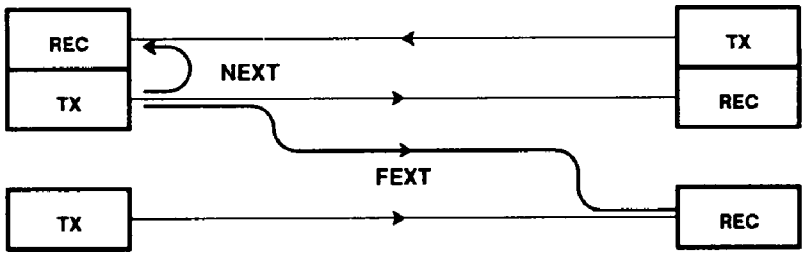

Figura 3.9: Ilustración de dos tipos de acoplamiento: acoplamiento de terminación lejana (FEXT) y acoplamiento de terminación cercana (NEXT).

Existen dos mecanismos básicos de acoplamiento, el *acoplamiento de terminación cercana* (NEXT) y el *acoplamiento de terminación lejana* (FEXT), tal como se ilustra en la figura 3.9. El NEXT representa un acoplamiento de un transmisor local en un receptor local, y experimenta una atenuación precisamente modelada por

$$|H_{NEXT}(j\omega)|^2 = K_{NEXT}|\omega|^{1.5}$$

donde $H_{NEXT}(j\omega)$ es la función transferencia que experimenta el crosstalk. El FEXT representa el acoplamiento de un transmisor local en un receptor remoto, con una atenuación dada por

$$|H_{FEXT}(j\omega)|^2 = K_{FEXT}|C(j\omega)|^2|\omega|^2$$

donde $C(j\omega)$ es la atenuación de la línea. Cuando ambas están presentes el NEXT domina al FEXT debido a que esta última experimenta la atenuación de la longitud total de la línea (además de la pérdida de acoplamiento). Ambas formas de crosstalk experimentan menos atenuación a mayores frecuencias, de forma que en un contexto limitado por el crosstalk es ventajoso minimizar el ancho de banda requerido para la transmisión.

3.1.3 Fibra óptica (*)

El cable de fibra óptica es capaz de transmitir luz en largas distancias con gran ancho de banda y baja atenuación. Carece de sensibilidad a interferencias externas, y ofrece inmunidad de intercepción por medios externos y se obtiene además a partir de materiales baratos y abundantes. Es dificil imaginar un medio más adecuado para las comunicaciones digitales!

La utilización de una guía de onda dieléctrica óptica para las comunicaciones de alto desempeño fué sugerida por Kao y Hockham en 1965. Alrededor de 1986 este medio fué adecuadamente desarrollado y reemplazó rapidamente los pares de conductores y el coaxil en muchas instalaciones cableadas. Este medio permite un gran ancho de banda con un costo modesto de forma que puede reemplazar varios de los usos actuales de transmisión satelital y de radio. De esta forma, las comunicaciones digitales sobre pares de conductores y coaxil parece que están limitadas a aplicaciones que utilizan las facilidades de transmisión existentes (como es el caso de DSL).

La transmisión digital por satélite y radio estará limitada a aplicaciones especiales que hacen uso de sus propiedades particulares. Por ejemplo, las radiocomunicaciones serán indispensables en situaciones donde uno o ambos terminales sean móviles, por ejemplo en telefónia digital móvil o en comunicaciones espaciales. La radiocomunicación es también excelente para proveer una cobertura a grandes distancias donde el ancho de banda requerido sea modesto y la instalación del cable de fibra óptica no pueda ser justificada. Además, los satélites tienen capacidades únicas para ciertos tipos de situaciones de acceso múltiple a un recurso que cubren una amplia región geográfica.

Guía de onda por fibra óptica

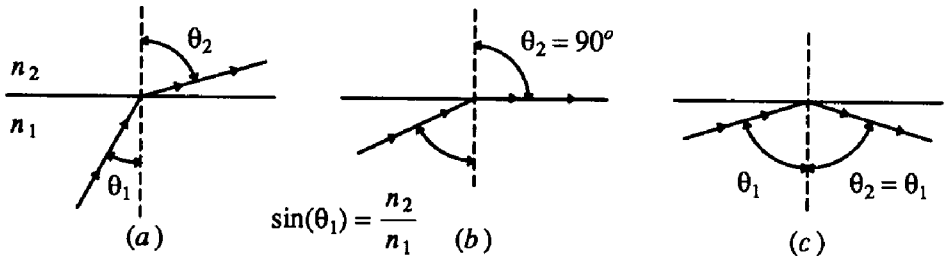

Figura 3.10: Ilustración de la Ley de Snell y la reflexión interna total. a) Definición de los ángulos de incidencia θ_1 y θ_2. El ángulo de refracción es mayor si $n_1 > n_2$. b) El ángulo crítico de incidencia para el cual el ángulo de refracción es de 90^o. c) La reflexión total ocurre para ángulos de incidencia mayores que el ángulo crítico.

El principio de guía de onda por fibra óptica puede entenderse del concepto de *reflexión interna total*, mostrado en la figura 3.10a. Una onda de luz, representada por un haz único, incide sobre el límite entre dos materiales tal que el ángulo de incidencia es θ_1 y el ángulo de refracción es θ_2. Definimos un *haz* como el camino central que un rayo lentamente divergente toma cuando pasa a través del sistema. Tal rayo tiene un diámetro grande con respecto a la longitud de onda de forma que es válida la aproximación de onda plana. Supongamos que el índice de refracción n_1 en el material incidente es mayor que el índice de refracción del medio de refracción n_2, o sea $n_1 > n_2$. Entonces la Ley de Snell predice que

$$\frac{\sin(\theta_1)}{\sin(\theta_2)} = \frac{n_2}{n_1} < 1$$

El ángulo de refracción es mayor que el ángulo de incidencia. En la figura 3.10b se muestra el caso de un ángulo crítico de incidencia donde el ángulo de refracción es de 90^o, tal que la luz es refractada a lo largo del límite entre los materiales. Esto corresponde al ángulo de incidencia *crítico* dado por

$$\sin(\theta_1) = \frac{n_2}{n_1}.$$

Para ángulos mayores que este existe reflexión interna total como se ilustra en la figura 3.10c, donde el ángulo de reflexión es siempre igual al ángulo de incidencia.

Este principio puede ser explotado en un *guía de onda por fibra óptica* como la ilustrada en la figura 3.11. Los materiales del *cuerpo* y el *revestimiento* son plásticos traslúcidos, los cuales transmiten la luz con poca atenuación, mientras que la *vaina* es un material plástico opáco que no sirve para otro propósito que el de otorgar rigidez, absorber cualquier luz que pueda escapar, y prevenir cualquier interferencia de luz externa. El cuerpo tiene mayor índice de refracción que el revestimiento, con el resultado que los haces incidentes con pequeño ángulo de incidencia se capturan por reflexión interna total. Esto se ilustra en la figura 3.12, donde un haz de luz incidente al final de la fibra es capturado por reflexión interna total siempre que el ángulo de incidencia θ_1 sea menor que el ángulo crítico. El modelo de haces predice que la luz rebotará una y otra vez confinandose a la guía de onda hasta emerger en la terminación. Además, es obvio que la longitud del camino de un haz, y en consecuencia el tiempo de transito, es función del ángulo incidente del haz.

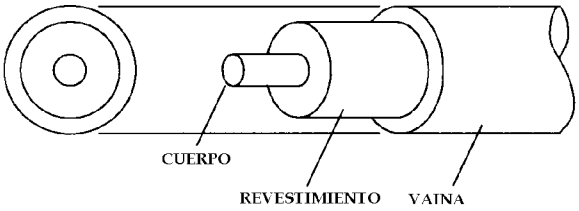

Figura 3.11: Una guía de onda de fibra óptica. El cuerpo y el revestimiento sirven para confinar la luz incidente en ángulos de incidencia estrechos, mientras que la vaina opáca sirve para dar estabilidad mecánica y prevenir acoplamientos e interferencias.

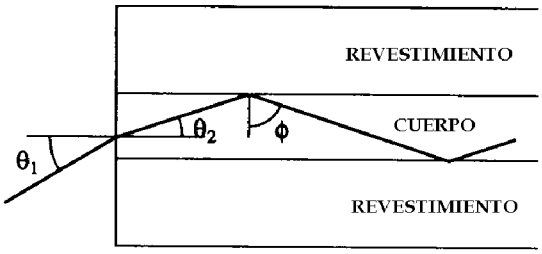

Figura 3.12: Modelo de haces de la propagación de la luz en una guía de onda óptica mediante reflexión interna total. Se muestra una sección de la guía de onda de fibra a lo largo del eje de simetría, con un haz de luz incidente de ángulo θ_1 que pasa a través del eje de la fibra.

Esta variación en tiempo de tránsito para diferentes haces pone en evidencia el *ensanchamiento de pulso* que a su vez limita la velocidad de transmisión de pulsos que puede ser utilizada o la distancia a la que puede transmitirse o ambos. El ensanchamiento de pulso puede reducirse modificando el diseño de la fibra, y especificamente utilizando una *fibra de índice gradual*, en la cual el índice de refracción varia continuamente con la dimensión radial.

Este modelo de haces introduce alguna idea del comportamiento de la luz en una guía de onda de fibra óptica. Por ejemplo, este modelo predice correctamente que existirá mayor ensanchamiento de pulso cuando la diferencia de índices entre cuerpo y revestimiento es mayor. Sin embargo, este modelo no es adecuado para dar una descripción precisa dado que en la práctica las dimensiones radiales de la fibra son del orden de la longitud de onda de la luz. Por ejemplo, el modelo de haces de luz predice que existirá una franja contínua de ángulos para la cual la luz rebotará entre cuerpo y revestimiento indefinidamente. Un modelo más refinado utiliza las ecuaciones de Maxwell para predecir el comportamiento de la luz en la guía de onda. De él se podría obtener que existen solo un número fínito y discreto de valores de ángulos con los cuales la luz se propaga por la fibra indefinidamente. Cada uno de esos ángulos corresponde a un *modo* de propagación, similar a los modos en una guía de onda metálica transportando radiación de microondas. Cuando el radio del cuerpo es varias veces mayor que la longitud de onda de propagación de la luz, existirán varios modos, esto se denomina *fibra multimodo*. Cuando el radio del cuerpo es reducido, menor cantidad de modos pueden introducirse, hasta que un radio del cuerpo del orden de la longitud de onda solo tiene un modo de propagación. Esta se denomina *fibra monomodo*. Para una fibra monomodo el modelo de haces es deficiente ya que su precisión depende de sus grandes dimensiones físicas en relación a la longitud de onda. En realidad, en la fibra monomodo la luz no está confinada al cuerpo, sino que una fracción significativa de la potencia se propaga en el revestimiento. Cuando el radio del cuerpo se vuelve más pequeño más potencia se transportará por este revestimiento.

Por varias razones que se discutirán, la capacidad de transmisión de la fibra monomodo es mayor. Sin embargo, es más dificil su manejo con baja atenuación y también falla para capturar la luz en ángulos de incidencia grandes en relación a los que capturaría una fibra multimodo, lo que hace dificil manejar una potencia óptica especificada. Teniendo en cuenta su mayor capacidad, existe una tendencia al uso exclusivo de fibra óptica monomodo en nuevas instalaciones, a pesar que la fibra multimodo ha sido extensivamente utilizada en el pasado. En la discusión siguiente se enfatizarán las propiedades de la fibra monomodo.

Se discutirán a continuación los factores que limitan el ancho de banda o la velocidad de transmisión a través de una fibra de longitud especificada. Los factores importantes son:

- *Atenuación del material*, asociada a la potencia de señal que resulta inevitable cuando la luz viaja por la guía de onda de fibra óptica. Existen cuatro fuentes de esta atenuación en una fibra monomodo: dispersión de la luz por un plástico de estructura molecular no homogénea, absorción de la luz por impurezas en el plástico, atenuación en los conectores, y pérdidas introducidas al curvar la fibra. Generalmente esas pérdidas estarán afectadas por la longitud de onda de la luz, la cual afecta la distribución de potencia entre el cuerpo y el revestimiento así como también por mecanismos de dispersión y absorción. El efecto de esos mecanismos de atenuación es que la atenuación de potencia de señal en dB es proporcional a la longitud de la fibra. En consecuencia, para una línea de longitud L, si la pérdida en dB/km es γ_0, la pérdida total de la fibra es $\gamma_0 L$ y en consecuencia la relación de potencia transmitida P_T a la recibida P_R será

$$\gamma_0\, L = 10\, \log_{10} \frac{P_T}{P_R}, \quad \text{ó} \quad P_R = P_T\, 10^{-\frac{\gamma_0 L}{10}}$$

  Esta dependencia exponencial de la atenuación en función de la longitud es la misma que para las líneas de transmisión discutidas anteriormente.

- *Dispersión de modo*, o la diferencia en velocidad de grupo entre diferentes modos, que resulta en el ensanchamiento de un pulso introducido en la fibra. Este ansanchamiento de pulso determina una interferencia entre pulsos sucesivos a ser transmitidos, denominada *interferencia intersímbolos*. Dado que este ensanchamiento del pulso aumenta con la longitud de la fibra, esta dispersión limitará la distancia entre repetidores regenerativos. Una ventaja significativa de las fibras monomodo es que la dispersión de modo no existe ya que existe un único modo.

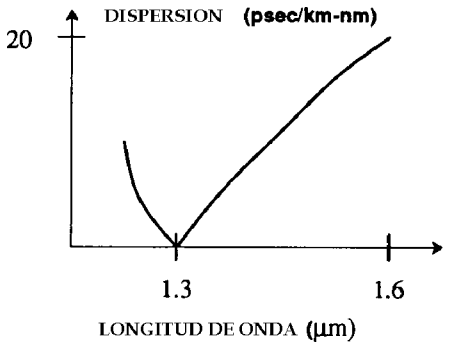

Figura 3.13: Dispersión cromática típica en una fibra de sílica. Se muestra la magnitud de la dispersión; la dirección de la dispersión en realidad se invierte en los cruces por cero.

- *Dispersión cromática o del material* es causada por las diferencias en la velocidad de propagación a diferentes longitudes de onda. Para longitud de onda del orden del infrarrojo y más largas, las longitudes más cortas llegan antes que las longitudes más largas, pero existe un punto de quiebre a alrededor de 1.3 μm más allá del cual las longitudes más largas llegan primero. Dado que las fuentes ópticas tienen ancho de banda no nulo, denominado *ancho de línea*, y la modulación de la señal incrementa el ancho de banda óptico aún más, la dispersión del material causará interferencia intersímbolos y limitará la distancia entre repetidores regenerativos. La dispersión del material es cualitativamente similar a la dispersión que ocurre en las líneas de transmisión debido a la atenuación dependiente de la frecuencia. La dispersión total se expresa usualmente en picosegundos de dispersión de pulso / GHz de ancho de banda de la fuente / km. de distancia, con valores típicos en el rango de 0.15 en la región de 1.3-1.6 μm de mínima atenuación. Esto es sumamente importante ya que la dispersión es practicamente cero en esta longitud de onda. Una curva típica de la magnitud de la dispersión cromática en función de la longitud de onda se muestra en la figura 3.13. La dispersión cromática puede ser despreciablemente pequeña sobre un rango de longitudes de onda relativamente grande. Además, la frecuencia de este cero en la dispersión cromática puede ser desplazada a través del diseño de la guía de onda para coincidir con la longitud de onda de atenuación mínima.

Con estos desperfectos en mente, es posible discutir los límites prácticos y fundamentales sobre la capacidad de información de una fibra. El limite fundamental sobre atenuación se debe a la dispersión intrínseca del material plástico de la fibra: conocido como *dispersión de Rayleigh*, y es similar a la dispersión en la atmósfera de la tierra que resulta en el color azul del cielo. La atenuación de dispersión disminuye rapidamente con la longitud de onda (en términos de la cuarta potencia), por lo que en consecuencia es ventajoso elegir longitudes de onda largas. La atenuación debido a la absorción intrínseca es despreciable, pero para determinadas longitudes de onda es posible observar experimentalmente atenuaciones grandes debido a impurezas. Particularmente importantes son los radicales de hydroxyl (OH), los cuales absorben a una longitud de onda de 2.73 μm y sus armónicas. En grandes longitudes de onda existe absorción infrarroja asociada fundamentalmente con el plástico, la cual crece abruptamente a partir de 1.6 μm.

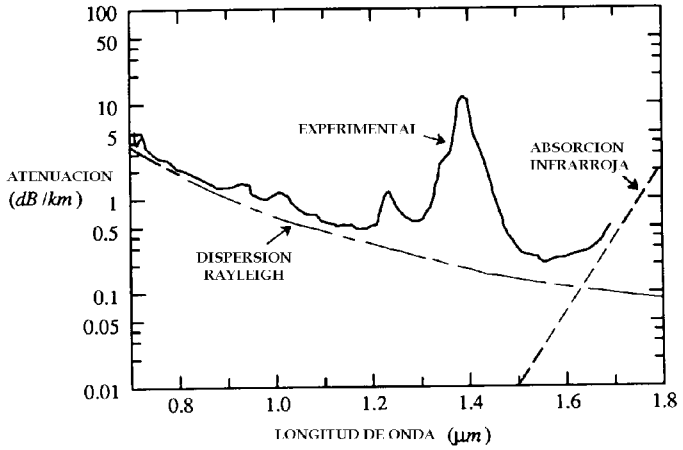

Figura 3.14: Atenuación del espectro medida para una fibra monomodo de pérdida ultrabaja de germaniosilicato junto con la pérdida debido a los efectos intrínsecos del material.

Una curva de atenuación para una fibra real se muestra en la figura 3.14. Notar las curvas de atenuación para los dos efectos intrínsecos que podrían estar presentes en un material ideal, la dispersión Rayleigh y la absorción infrarroja, y picos de absorción adicional en 0.95, 1.25 y 1.39 μm debido a impurezas OH. Las más bajas atenuaciones están en aproximadamente 1.3 y 1.5 μm, y esas son las longitudes de onda a las cuales el sistema opera con el desempeño más alto. La atenuación es tan baja como 0.2 dB/km, lo que implica potencialmente en un espaciamiento entre repetidores para sistemas de comunicaciones digitales con fibra óptica mucho más largo que para par trensado y cable coaxil. En la figura 3.15 se muestra una curva de atenuación en función de la frecuencia para distintos tipos de cable y fibra óptica mostrando que en esta última la atenuación es mucho menor.

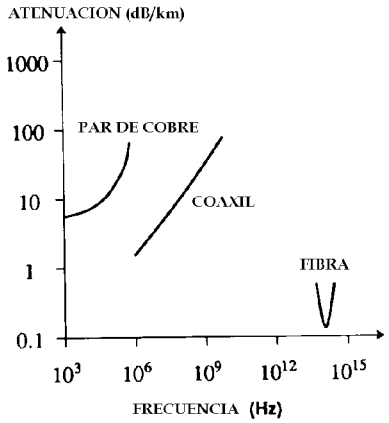

Figura 3.15: Atenuación en función de la frecuencia para cables y fibra. La banda de frecuencia sobre la cual la atenuación de la fibra es menor que 1 dB/km es mayor que 10^{14} Hz.

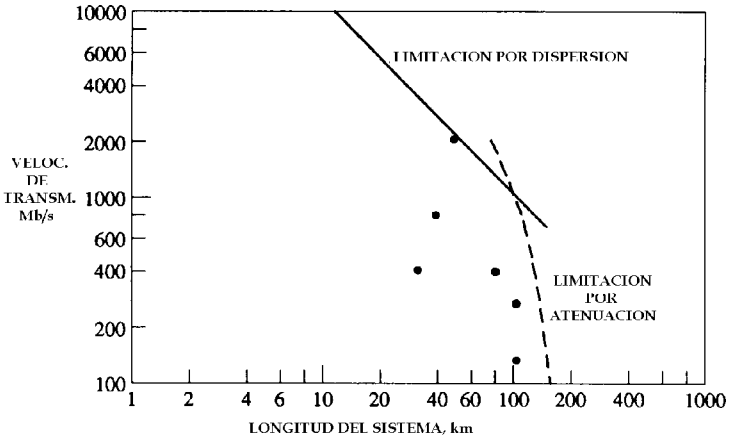

Figura 3.16: Compromiso entre distancia y velocidad para una fibra monomodo con un conjunto particular de especificaciones. Los puntos representan el desempeño real de sistemas ensayados.

La atenuación de una fibra por unidad de distancia es mucho más importante en la determinación de la distancia entre repetidores que la velocidad de transmisión de la información que se transmite. Esto se ilustra para una fibra monomodo en la figura 3.16, donde es posible visualizar una región de límite impuesta por la atenuación, donde la curva del espaciamiento entre repetidores en función de la velocidad de transmisión es relativamente plana. Sin embargo, si incrementamos la velocidad de transmisión es posible aproximarse eventualmente a una región donde el espaciamiento entre repetidores está limitado por la dispersión (dispersión de modos en fibra multimodo y dispersión cromática en fibra monomodo). La magnitud de esta última puede cuantificarse simplemente considerando la transformada de Fourier de un pulso transmitido y en particular su ancho de banda W. La dispersión del pulso será proporcional al espaciamiento entre repetidores L y el ancho de banda W, con una constante de proporcionalidad D. De esta forma, si requerimos que esta dispersión sea menor que la mitad del ancho del pulso a una velocidad de R pulsos por segundo, se tiene que

$$D\,L\,W < \frac{1}{2R}$$

El ancho de banda de la fuente W depende del *ancho de línea*, o ancho de banda intrínseco en ausencia de modulación, y también de la modulación. Dado que un ancho de línea no nulo incrementará el ancho de banda y en consecuencia la dispersión cromática, es posible hacer una interpretación de los limites fundamentales suponiendo ancho de línea cero. Es posible mostrar que el ancho de banda debido a modulación es aproximadamente igual a la velocidad de transmisión $W \cong R$, y de esta forma la ecuación anterior se convierte en

$$R^2\,L < \frac{1}{2D}$$

Esta ecuación implica que en la región donde la dispersión es limitante, el espaciamiento entre repetidores L debe disminuir rapidamente cuando la velocidad de transmisión es aumentada, como mostrado en la figura 3.16.

Como resultado de esas consideraciones, la primera generación (1980) de sistemas de transmisión por fibra óptica utilizaba fibra multimodo con una longitud de onda dealrededor de 0.8 μm y lograba velocidades de hasta 150 Mb/seg. La dispersión Rayleigh es de alrededor de 2 dB/km a esta longitud de onda, y la distancia entre repetidores regenerativos era de 5 a 10 km. Una segunda generación de sistemas (1985) se movió hacia fibras monomodo y longitudes de onda de alrededor de 1.3 μm, donde la dispersión Rayleigh es de alrededor de 0.2 dB/km (y las atenuaciones prácticas están más en el orden de 0.3 dB/km).

La longitud finita de fibras manufacturadas y las consideraciones de instalación del sistema determinan el tipo de conectores de la fibra. Estos presentan complicados problemas de alineamiento, tanto más complicados para fibras monomodo debido al cuerpo más pequeño. En la práctica pérdidas de 0.1 a 0.2 dB pueden obtenerse aún con fibras monomodo. Dado que debe preverse en cualquier instalación quiebres accidentales, el empalme subsiguiente será requerido en numerosos puntos, de forma que pérdidas por conectores y empalmes serán las pérdidas dominantes que limitan el espaciamiento entre repetidores.

Las pérdidas por curvatura se deben a diferentes velocidades de propagación requeridas por el radio externo en relación al interno de la curvatura. Cuando el radio de curvatura disminuye, eventualmente la luz en el radio externo debe viajar más rápido que la velocidad de la luz, lo cual es obviamente imposible. Lo que sucede en realidad es que ocurre una atenuación significativa debido a la pérdida de potencia confinada. Existe generalmente un compromiso entre pérdidas por curvatura y pérdidas por empalme en fibras monomodo dado que las pérdidas por curvatura se minimizan confinando la mayoría de la potencia en el cuerpo, pero esto hace que la alineación de un empalme sea más crítica.

En la figura 3.16 se cuantifica el compromiso entre distancia máxima y velocidad de transmisión para una fibra monomodo para un conjunto particular de suposiciones (los valores numéricos dependerán de esas suposiciones). En velocidades de transmisión debajo de un Gb/seg la distancia está limitada por atenuación y sensibilidad del receptor. En este rango la distancia disminuye cuando la velocidad de transmisión aumenta dado que la sensibilidad del receptor disminuye. A mayores velocidades, el ensanchamiento de pulso limita la distancia antes que la atenuación se vuelva importante. La capacidad total de un sistema de fibra óptica se mide mejor por medio de una figura de mérito igual al producto de la velocidad de transmisión y la distancia entre repetidores, medida en Gb-km/seg. Algunos sistemas comerciales actuales logran velocidades del orden de 100 a 1000 Gb-km/seg.

Fuentes

A pesar que la transmisión por fibra óptica utiliza la energía de la luz para transportar los bits de información, con la tecnología actual las señales son generadas y manipuladas electricamente. Esto implica una conversión eléctrica a óptica a la entrada de la fibra y una conversión óptica a eléctrica a la salida.

Existen dos fuentes de luz para sistemas de comunicaciones digitales por fibra: el *diodo emisor de luz* (LED) y el *laser semiconductor de injección*. Como el laser semiconductor es más importante para sistemas de alta capacidad se discutirá con más detalle.

En contraste al LED, la salida del laser es *coherente*, o aproximadamente confinada a una sola frecuencia. Como la salida del laser no tiene exactamente ancho de banda nulo es necesario un diseño cuidadoso para que el ancho de línea sea pequeño en relación al ancho de banda de la señal, esto que se consigue con una estructura denominada de *realimentación distribuida*. De esta forma son factíbles esquemas de modulación y demodulación coherentes, a pesar que la mayoría de los sistemas comerciales utilizada modulación de intensidad (no coherentes).

La salida del laser puede acoplarse a una fibra monomodo con una eficiencia muy alta (alrededor de 3 dB de pérdidas), y puede tener potencias en el rango de 0 a 10 mW, con 1 mW (0 dBm) como valor típico. Dado que emite un haz más estrecho que el LED, el laser es necesario para fibras monomodo excepto para distancias cortas.

La luz de salida del laser depende de la temperatura, por lo que es necesario monitorearla y controlar la corriente de alimentación utilizando un circuito realimentado. No existen muchas alternativas para aumentar la potencia a ser provista por la fibra debido a los efectos no lineales que aparecen, a menos que sea posible hallar formas de compensar o evitar dichos fenómenos.

Fotodetectores

La energía óptica de la salida de la fibra se convierte en una señal eléctrica por medio de un *fotodetector*. Existen dos tipos de fotodetectores: el *fotodiodo PIN*, más usando para velocidades hasta 100 Mb/seg, y el *fotodiodo de avalancha* (APD), más utilizado por encima de 1 Gb/seg.

Una sección del fotodiodo PIN se muestra en la figura 3.17. Este diodo tiene una región intrínseca (región no dopada no típica de los diodos) entre las regiones de silício n y p. Los fotones de la señal óptica recibida son absorbidos y se crean pares hueco-electrón. Si el diodo se polariza en forma inversa, existirá un campo inverso a lo largo de la región de deplexión (que incluye la porción intrínseca) que separará los huecos de los electrones y conducirá a estos a los contactos, creando una corriente proporcional a la potencia óptica incidente. El objetivo de la región intrínseca es alargar la región de deplexión, lo que permite aumentar la fracción de fotones incidentes que se convertirán en corriente (los portadores creados fuera de la región de deplexión, o más allá de la distancia de difusión de la región de deplexión, se recombinarán con alta probabilidad antes que cualquier corriente pueda generarse).

La fracción de fotones incidentes convertida en portadores que alcanzan los electrodos se denomina *eficiencia cuántica* del detector, η. Dada una eficiencia cuántica es posible predecir la corriente generada como una función de la potencia óptica incidente. La energía de un fotón es $h\nu$, donde h es la constante de

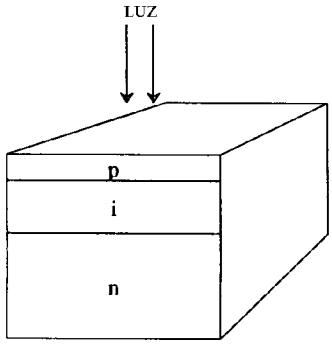

Figura 3.17: Sección de un diodo PIN. El electrodo de conexión a las regiones n ó p crea un diodo polarizado en forma inversa.

Planck (6.6×10^{-34} Joule-seg) y ν es la frecuencia óptica, relacionada con la longitud de onda λ por: $\nu\lambda = c$, donde c es la velocidad de la luz (3×10^8 m/seg). Si la potencia óptica incidente es de P watts, entonces el número de fotones por segundo es $P/h\nu$, y si una fracción de η de esos fotones genera un electrón con carga q (1.6×10^{-19} Coulombs), la corrientes total será

$$i = \eta q \frac{P}{h\nu}$$

Ejemplo: Para una longitud de onda de 1.5 μm y una eficiencia cuántica de 1, cual es la relación entre la corriente de salida y la potencia de entrada del fotodiodo PIN? En base a las relaciones anteriores:

$$\frac{i}{P} = \frac{q}{h\nu} = \frac{q\lambda}{hc} = \frac{1.6\,10^{-19}\,1.5\,10^{-6}}{6.6\,10^{-34}\,3.0\,10^{8}} = 1.21\,amps/watt$$

Si la potencia óptica incidente es 1 nanoWatt, la corriente máxima obtenida de un fotodiodo PIN es de 1.21 nanoamp.

Tanto para fotodiodos PIN como para todos los fotodetectores en general, existe un compromiso entre eficiencia cuántica y velocidad. Valores de eficiencia cuántica cercanos a la unidad pueden alcanzarse con fotodiodos PIN, pero esto requiere una región intrínseca de absorción grande. Esto implica en un campo eléctrico más pequeño (con una velocidad de los portadores menor) y una mayor distancia de difusión y en consecuencia una respuesta más lenta a una entrada óptica. Mayores velocidades resultan inevitablemente en una sensibilidad menor.

Dado que es difícil procesar electronicamente pequeñas corrientes sin introducir un ruido térmico significativo, es deseable incrementar la corriente de salida del diodo antes de la amplificación. Este es el objetivo del APD que posee ganancia interna, generando más de un par hueco-electrón por fotón incidente. En forma similar al fotodiodo PIN, el APD también es un diodo polarizado en inversa, pero la diferencia es que el voltage inverso es lo suficientemente grande como para que los portadores, liberados por un fotón y separados por el campo eléctrico, tengan suficiente energía para colisionar con los átomos del cristal semiconductor. Estas colisiones ionizan los átomos, generando pares hueco-electrón adicionales. Un costo de este mecanismo de ganancia es un ancho de banda inherente menor. Otro costo del APD es la naturaleza probabilística del número de portadores secundarios generados: a mayor ganancia en el APD se obtiene una mayor fluctuación en la corriente obtenida para una determinada potencia óptica. Además, el ancho de banda del dispositivo disminuye con el incremento de la ganancia, dado que el proceso de avalancha toma cierto tiempo para establecerse.

Ambos fotodiodos PIN y APD exiben una corriente pequeña que fluye en ausencia de luz incidente debido a portadores de excitación térmica. Esta corriente se denomina *corriente de oscuridad* por razones obvias y representa una señal de ruido en relación a la señal a ser detectada.

Modelo para la recepción con fibra

La salida para un detector de fibra óptica puede obtenerse en base a un modelo estadístico basado en procesos de Poisson y ruido shot. Esta señal tiene características bastante diferentes a la de otros medios de interés dado que las fluctuaciones cuánticas en la señal son importantes. Dado que la señal en si misma posee fluctuaciones cuánticas podemos considerarlas como de un tipo de *ruido multiplicativo.*

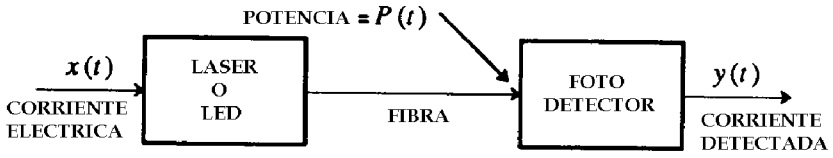

Figura 3.18: Elementos de un sistema de fibra óptica con detección directa.

En los sistemas comerciales se utiliza un modo de *detección directa*, como ilustrado en la figura 3.18. En ese modo, la intensidad o potencia de la luz se modula directamente con una fuente eléctrica (señal dato) y el fotodetector convierte esta potencia en otra señal eléctrica. Si la corriente de entrada a la fuente es $x(t)$, entonces la potencia de salida de la fuente es proporcional a $x(t)$.

Dos aspectos negativos ocurren sobre la potencia introducida al propagarse por la fibra. Primero la atenuación, lo que reduce la potencia de señal en el detector. Segundo, la dispersión debido a la dispersión cromática (y dispersión de modo en una fibra multimodo), lo cual puede modelarse por medio de una operación de filtrado lineal. Si $g(t)$ es la respuesta impulsiva del filtro de dispersión equivalente, incluyendo la atenuación, la potencia recibida en el detector será

$$P(t) = x(t) * g(t)$$

La situación es un poco más complicada en la conversión final a una corriente eléctrica en el fotodetector dado que los efectos cuánticos son importantes. La luz incidente consiste en fotones discretos que son convertidos a pares hueco-electrón en el detector. De esta forma, la corriente generada consiste en paquetes discretos de carga generada en puntos discretos en el tiempo. Intuitivamente es posible esperar que los tiempos de llegada de los paquetes de carga formen un proceso de Poisson (ver sección 2.2.4) dado que no existe razón para esperar que los tiempos entre llegadas de fotones dependan entre sí. En realidad esto puede predecirse por la teoría cuántica. Si $h(t)$ es la respuesta del circuito fotodetector a un único fotoelectrón entonces la salida de la corriente detectada $y(t)$ es un proceso de Poisson filtrado dado por

$$Y(t) = \sum_m h(t - t_m)$$

donde t_m son los tiempos de llegada de Poisson. Los tiempos de llegada de Poisson están caracterizados por la velocidad de llegadas proporcional a la potencia incidente,

$$Y(t) = \frac{\eta}{h\nu} P(t) + \lambda_0$$

donde η es la eficiencia cuántica y λ_0 es la corriente de oscuridad. El valor esperado de la corriente detectada será

$$E[Y(t)] = \lambda(t) * h(t) = \frac{\eta}{h\nu} x(t) * g(t) * h(t) + \lambda_0 H(0)$$

De esta forma, la relación entrada-salida del canal está caracterizada, con respecto al valor medio de corriente de salida del detector, por la convolución de dos filtros: la dispersión de la fibra y la respuesta del circuito del detector. Obviamente, existirán fluctuaciones alrededor del valor medio que deberán ser también

caracterizadas. Este modelo lineal simple para el canal es notablemente preciso a menos que la potencia óptica introducida sea lo suficientemente alta como para excitar efectos no lineales en la fibra y el detector.

Fotodetector de avalancha

En el caso de un APD es necesario modificar este modelo adicionando al modelo filtrado de proceso de Poisson el multiplicador aleatório del proceso de avalancha, de la forma:

$$Y(t) = \sum_m g_m h(t - t_m)$$

cuya estadística está caracterizada por

$$\begin{aligned} m_Y(t) &= E[G_m]\lambda(t) * h(t) \\ \sigma_Y^2(t) &= E[G_m^2]\lambda(t) * h^2(t) \end{aligned}$$

Si definimos la media y el segundo momento de la ganancia de avalancha como

$$\overline{G} = E[G_m], \qquad \overline{[G^2]} = E[G_m^2]$$

y en base a la estadística de Poisson es posible concluir que el efecto de la ganancia de avalancha es multiplicar el valor medio del proceso aleatório recibido por $\overline{G}$ y la varianza por $\overline{G^2}$.

Si el proceso de avalancha fuera determinístico, o sea, exactamente $\overline{G}$ electrones secundarios fueran generados por cada fotoelectrón primário, entonces el segundo momento debería coincidir con el valor medio cuadrático, $\overline{G^2} = \overline{G}^2$. El efecto de la aleatoridad del proceso de multiplicación se traduce en un momento de segundo orden más grande por un factor F_G mayor que la unidad, $\overline{G^2} = F_G\overline{G}^2$. Este factor se denomina *factor de ruido en exceso.* En realidad, un análisis detallado de la física del APD determina el siguiente resultado:

$$F_G = k\overline{G} + (2 - \frac{1}{\overline{G}})(1 - k)$$

donde $0 \leq k \leq 1$ es un parámetro bajo control del proyectista denominado *relación de portadores ionizados.* Notar que cuando $k \to 1$, $F_G \to \overline{G}$, o sea que el factor de ruido en exceso es aproximadamente igual a la ganancia de avalancha. Esto indica que la aleatoridad se vuelve mayor cuando es mayor la ganancia de avalancha. Por otro lado, cuando $k \to 0$, $F_G \to 2$ para $\overline{G}$ grande, o sea que el factor de ruido en exceso es aproximadamente independiente de la ganancia de avalancha. Finalmente, cuando $\overline{G} = 1$ (no existe ganancia de avalancha), $F_G = 1$ y no existe ruido en exceso. Este es el caso del fotodiodo PIN.

La fibra y el ruido térmico del preamplificador

Cualquier sistema físico a temperatura no nula experimenta ruido debido al movimiento térmico de los electrones y la fibra óptica no es una excepción. El estudio teórico del ruido térmico se aproxima frecuentemente por ruido blanco Gaussiano. La propiedad Gaussiana es un resultado del teorema del límite central y también de que el ruido térmico proviene de la superposición de varias acciones independientes. La propiedad de ruido blanco no puede extenderse obviamente a infinitas frecuencias dado que de esa forma la potencia total debería ser infinita. En realidad este ruido se considera blanco hasta frecuencias del orden de los 300 GHz aproximadamente. El resultado es que el ruido térmico tiene una potencia disponible (en una carga con impedancia acoplada) por Hz de

$$N(\nu) = \frac{h\nu}{e^{h\nu/kT_n} - 1}$$

donde h es la constante de Planck, k es la constante de Boltzmann ($1.38\,10^{-23}$ Joules por grado Kelvin), y T_n es la temperatura en grados Kelvin. Considerandola como una densidad espectral, es necesario dividirla por dos.

En frecuencias por debajo de las de microondas, el exponente en la ecuación anterior es muy pequeño, y si aproximamos e^x por $1 + x$ se tiene que el espectro es aproximadamente blanco,

$$N(\nu) = kT_n$$

lo que corresponde a una densidad espectral de potencia $N_0 = \frac{kT_n}{2}$. Sin embargo, en altas frecuencias, este espectro se aproxima a cero exponencialmente, lo que produce una potencia total finita.

Existen dos fuentes posibles de ruido térmico: a la entrada del detector y en el preamplificador del receptor. A la entrada del detector solo es relevante el ruido térmico a las frecuencias ópticas (el detector no responde a frecuencias más bajas) y en esas frecuencias el ruido térmico es despreciable.

Ejemplo: A la temperatura ambiente kT_n es 4×10^{-21} Joules. A 1 GHz, o frecuencias de microondas, $h\nu$ es aproximadamente 10^{-24} Joules, y estamos bien en el régimen donde el espectro es plano. Sin embargo, con una longitud de onda correspondiente a $\nu = 3 \times 10^{14}$ Hz, $h\nu$ es aproximadamente 2×10^{-19} Joules, y $\frac{h\nu}{kT_n}$ es alrededor de 50. De esta forma, el ruido térmico es mucho menor que kT_n a esas frecuencias. En general el ruido térmico a frecuencias ópticas es despreciable en sistemas de fibra óptica donde las longitudes de onda son menores que 2 μm.

Dado que el nivel de la señal es muy bajo a la salida del detector de la figura 3.18, debemos amplificar la señal usando un preamplificador como la primera etapa del receptor. El ruido térmico introducido en el preamplificador es una fuente significativa de ruido, y en realidad es la fuente de ruido predominante en varios sistemas ópticos. Dado que la señal en este punto es la forma de onda digital banda base, ella ocupa un ancho de banda que puede extenderse posiblemente hasta las frecuencias de microondas pero no hasta frecuencias ópticas, de aquí la importancia del ruido térmico.

Un estudio pormenorizado permitiría concluir que el ruido térmico es la razón primária para considerar preferentemente el uso del detector APD en lugar del fotodiodo PIN.

3.1.4 Radio microondas (*)

El término *radio* se utiliza para referirse a la transmisión electromagnética a través del espacio libre a frecuencias de microondas y por debajo. Existen varias applicaciones de transmisión digital que utilizan este medio, primariamente a frecuencias de microondas, un conjunto representativo del cual incluye *radio digital terreste punto a punto*, *radio digital móvil*, *comunicaciones digitales satelitales* y *comunicaciones digitales espaciales*.

Los sistemas de radio digital terrestre utilizan antenas de microondas tipo bocina colocadas en torres para extender el horizonte de alcance e incrementar el espaciamiento entre antenas. Este medio ha sido utilizado en el pasado principalmente para la transmisión analógica (usando FM y más recientemente BLU), pero en años recientes ha sido gradualmente convertida a transmisión digital debido a la demanda creciente de servicios de datos.

Ejemplo 1: En norteamérica existe una asignación de frecuencia para radio telefonía digital centrada en 2, 4, 6, 8 y 11 GHz. En los Estados Unidos existían alrededor de 10000 enlaces de radio en 1986, incluyendo una red de costa a costa de 4 GHz.

Una aplicación relacionada es la radio móvil digital.

Ejemplo 2: En Estados Unidos la banda de frecuencia de 806 a 947 MHz está asignada a servicios de radio móvil terrestre. Esta banda es utilizada por la *radio móvil celular*, en la cual un area geográfica se

divide en un conjunto de celdas, cada una de las cuales tiene su propia antena base fija omnidireccional para transmisión y recepción. Cuando un vehículo pasa a través de las celdas, la antena base asociada se conmuta automaticamente a la más próxima a la trayectoria del mismo. Una ventaja de este concepto es que pueden acomodarse teléfonos móviles adicionales disminuyendo el tamaño de las celdas e introduciendo antenas base adicionales.

Los satélites se utilizan para las comunicaciones de larga distancia entre dos antenas terrestres, donde el satelite usualmente actúa como un repetidor no regenerativo. O sea, el satélite simplemente recibe una señal de la antena terrestre transmisora, la amplifica, y la transmite de vuelta hacia otra antena receptora terrestre. Los canales satelitales ofrecen una alternativa excelente a la fibra o el cable para la transmisión sobre grandes distancias, y particularmente en destinos dispersos donde el tráfico total es pequeño. Los satélites tienen también la poderosa característica de proveer un medio natural de acceso múltiple, el cual es invalorable para comunicaciones de acceso aleatório entre varios usuarios. Una limitación del enlace satelital es su limitada potencia disponible para transmisión, dado que la potencia se obtiene a partir de energía solar o fuentes descartables. Además, la configuración de los vehículos de lanzamiento usualmente limita el tamaño de las antenas de recepción y transmisión (las cuales son usualmente la misma). La mayoría de los satelites de comunicaciones se colocan en orbita sincrónica, de forma que aparecen como estacionarios en relación a un punto en la tierra. Esto simplifica enormemente el problema de direccionamiento de la antena y disponibilidad del satélite.

En comunicaciones espaciales, el objetivo es transmitir datos y recibirlos desde una plataforma que esté a gran distancia de la tierra. Esta aplicación incluye las particularidades de las comunicaciones satelitales y las móviles, ya que el vehículo está usualmente en movimiento. Como en el caso de satélite, el tamaño de la antena y las disponibilidades de potencia en el vehículo espacial son limitados.

Con la excepción de los problemas de propagación multicamino en enlaces terrestres, el canal de transmisión de microondas es relativamente simple. Existe una atenuación introducida en el medio debido a la dispersión de la energía, donde esta atenuación es independiente de la frecuencia, y el ruido térmico introducido en la antena y en los amplificadores del receptor. Esos aspectos del canal se discutirán en los siguientes parágrafos.

Antenas y transmisión por microondas

La propagación por microondas en espacio libre es simple ya que se debe a la atenuación asociada a la dispersión de la radiación. La atenuación varía tan lentamente con la frecuencia que puede considerarse virtualmente constante dentro del ancho de banda de la señal. Es posible considerar primero una *antena isotrópica*, o sea, una que irradia igual potencia en todas las direcciones. Se supone que la potencia total irradiada es P_T watts y que a una distancia de d metros de esta antena transmisora existe una antena receptora con un area de A_R metros^2. Luego, la máxima potencia que la antena receptora puede capturar es la potencia transmitida por la relación de A_R al área de la esfera de radio d, o sea, $4\pi d^2$. Existen dos factores que modifican este resultado. Primero, la antena transmisora puede diseñarse concentrando la energía irradiada hacia la antena receptora. Esto adiciona a la potencia recibida un factor G_T denominado *ganancia de la antena transmisora*. El segundo factor es la *eficiencia de antena* η_R de la antena receptora, un número próximo pero menor que uno, esto indica que la antena receptora no captura toda la radiación electromagnética incidente en ella. De esta forma la potencia recibida será

$$P_R = P_T \frac{A_R}{4\pi d^2} G_T \eta_R$$

A frecuencias de microondas se utilizan tipicamente antenas de apertura, tal como las bocinas o las parabólicas, y para ellas la ganancia de antena alcanzable es

$$G = \frac{4\pi A}{\lambda^2} \eta$$

donde A es el área de la antena, λ es la longitud de onda de transmisión, y η es un factor de eficiencia. Esta expresión se aplica tanto a la antena de recepción como a la de transmisión, donde debe sustituirse el área y la eficiencia apropiados. Esta expresión es razonablemente intuitiva ya que dice que la ganancia de antena es

proporcional al cuadrado de la relación de la dimensión de la antena y la longitud de onda. De esta forma, el tamaño de la antena transmisora en relación con la longitud de onda es todo lo que hay que tener en cuenta en la directividad o ganancia de la antena. Esta ganancia de antena aumenta con la frecuencia para un área de antena especificada, de forma que altas frecuencias tienen la ventaja de requerir menores antenas para una ganancia determinada. La eficiencia de una antena está tipicamente en el rango de 50 a 75 por ciento para un reflector parabólico y es tan alta como el 90 por ciento para una antena bocina.

Una forma alternativa de la ecuación de potencia recibida puede obtenerse sustituyendo el área en la ecuación de la antena receptora en términos de su ganancia tal que

$$P_R = G_T G_R \left[\frac{\lambda}{4\pi d}\right]^2$$

que se conoce como *ecuación de Friis*. El término entre corchetes se denomina *pérdidas de camino*, mientras que G_T y G_R sintetizan los efectos de las dos antenas. A pesar que esta pérdida es función de la frecuencia, la potencia real sobre el ancho de banda de señal no varía apreciablemente cuando el ancho de banda es muy pequeño en relación a la frecuencia central de la señal modulada. La ecuación de Friis no tiene en cuenta otras fuentes de pérdida tal como atenuación por lluvia o mal direccionamiento de las antenas.

La aplicación de esas relaciones a una configuración particular puede determinar la potencia recibida y los factores que contribuyen a la pérdida de potencia de señal. Este proceso se conoce como *balance de potencias del enlace*, como ilustrado por el siguiente ejemplo.

Ejemplo 1: Determinar la potencia recibida por el enlace de un satélite sincrónico a una antena terrestre, con los siguientes parámetros: distancia 40000 Km, potencia transmitida por el satelite 2 watts, ganancia de la antena transmisora 17 dB, área de la antena receptora 10 metros^2 con eficiencia perfecta, frecuencia 11 GHz. La longitud de onda será:

$$\lambda = \frac{c}{\nu} = \frac{3 \times 10^8}{11 \times 10^9} = 27.3\, mm$$

La ganancia de la antena receptora será

$$10 \log_{10} G_R = 10 \log_{10} \frac{4\pi 10}{(27.3 \times 10^{-2})^2} = 52.2$$

la pérdida de camino es

$$10 \log_{10} \left[\frac{\lambda}{4\pi d}\right]^2 = 20 \log_{10} \left[\frac{27.3 \times 10^{-3}}{4\pi 4 \times 10^7}\right] = -205.3\, dB$$

Finalmente, es posible calcular la potencia recibida, expresada en dBw (decibeles relativos a 1 watt) y teniendo en cuenta que la potencia transmitida es de 3 dBw se tiene que

$$10 \log_{10} P_R = 3dBw + 17 + 52.3 - 205.3 = -133\, dBw.$$

Ejemplo 2: La misión espacial a Mercurio Mariner 10 de 1974 utilizó una potencia transmitida de 16.8 watts y frecuencia de 2.3 GHz. El diámetro de la antena transmisora en la espacionave era de 1.35 m. con eficiencia 0.54, lo que resulta en una ganancia de antena de 27.6 dB. El diámetro de la antena receptora terrestre era de 64 m., con una eficiencia de 0.575, lo que produce una ganancia de antena de 61.4 dB. La distancia de la espacionave a la tierra es de 1.6 $\times 10^{11}$ m, para una pérdida de camino de 263.8 dB. Finalmente la potencia recibida era de

$$10 \log_{10} P_R = 10 \log_{10} 16.8 + 27.6 + 61.4 - 263.8 = -162.6\, dBw.$$

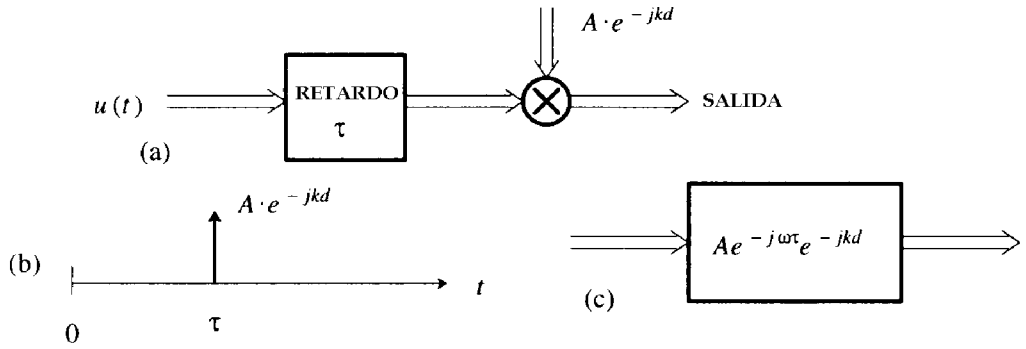

Figura 3.19: Canal banda base equivalente complejo para la propagación en el espacio libre con atenuación A a una distancia d. a) Sistema equivalente, b) respuesta impulsiva equivalente y c) función transferencia banda base equivalente.

Los dos efectos dominantes en la propagación de microondas son la atenuación y el retardo. Es de interés determinar estos efectos sobre una señal pasabanda en general. Supondremos una atenuación A, una distancia de propagación d y una velocidad de propagación c. El retardo que experimenta la señal es $\tau = d/c$, y dada la representación pasabanda de la forma de la figura 2.2, la salida del canal es

$$\sqrt{2}A\ Re\{u(t-\tau)e^{j\omega_c(t-\tau)}\} = \sqrt{2}A\ Re\{u(t-\tau)e^{-jkd}e^{j\omega_c t}\}$$

donde $k = \frac{\omega_c \tau}{d} = \frac{\omega_c}{c} = \frac{2\pi}{\lambda}$ se denomina *constante de propagación*. El canal banda base complejo equivalente se muestra en la figura 3.19 y caracteriza el efecto de propagación sobre la señal banda base compleja. No es sorprendente que la señal banda base tenga un retardo τ, el mismo que la señal pasabanda. Además, existe un desplazamiento de fase $kd = 2\pi d/\lambda$ radianes, o 2π radianes por cada longitud de onda sobre la distancia de propagación. La respuesta impulsiva compleja equivalente de la propagación es un impulso con retardo τ y area $A\,e^{-jkd}$, y la función transferencia equivalente es $A\,e^{-j\omega\tau}e^{-jkd}$, que resulta en una función lineal de la frecuencia debida al retardo. Para el caso de receptores móviles el efecto de pequeños cambios en d sobre la respuesta del canal banda base es particularmente significativo. Este efecto es dramaticamente más pronunciado para el desplazamiento de fase que para el retardo.

Ejemplo 3: Para una frecuencia de portadora de 1 GHz (típica para radio móvil), la constante de propagación es $k = \omega_c/c = 21$ radianes/m. De esta forma, un desplazamiento de fase de $\pi/2 = 1.57$ radianes, muy significativo para los demoduladores, ocurrirá con cada 7.4 cm de cambio de la distancia de propagación. En contraste, el retardo de propagación cambia en 3.3 nanoseg por cada metro de cambio en la distancia de propagación. En relación con los anchos de banda de banda base típicos esto es poco significativo. Por ejemplo, a 1 MHz, el cambio en el desplazamiento de fase debido a este cambio de retardo es de solo $\omega\tau = 2\pi\ 0.0033$, o aproximadamente 1 grado.

Ruido en amplificadores de microondas

En un enlace de radio el ruido entra al receptor a través de la antena y como fuente interna de ruido en el receptor. Ya se ha discutido que ambas fuentes de ruido son Gaussianas y pueden considerarse como blancas hasta las frecuencias de microondas. El ruido blanco queda completamente especificado por la densidad espectral N_0. Sin embargo, en la transmisión de radio es común expresar esta densidad espectral en términos de un parámetro equivalente, la *temperatura de ruido* expresada en grados Kelvin. La costumbre de especificar la temperatura de ruido proviene de que T_n es razonable en magnitud, del orden de las decenas o centenas de grados, mientras que N_0 es un número muy pequeño. Notar sin embargo que el ruido térmico total en algún punto del sistema puede ser la superposición de varias fuentes de ruido térmico a diferentes temperaturas. De esta forma, la temperatura de ruido es meramente una especificación conveniente de la

potencia de ruido, y no es necesariamente igual a la temperatura física de cualquier parte del sistema! Por ejemplo, si se amplifica el ruido se incrementa la temperatura de ruido sin afectar la temperatura física de la fuente que generó el ruido.

Existen dos fuentes de ruido, el ruido incidente sobre la antena y el ruido introducido internamente en el receptor. El ruido incidente sobre la antena depende de la temperatura efectiva de ruido en la dirección en que está apuntada la antena. Por ejemplo, el sol tiene una temperatura efectiva de ruido más alta que la de la atmósfera. El ruido interno del receptor depende de su diseño y sofisticación. Es costumbre referir todas las temperaturas de ruido a la entrada del receptor (la antena), y definir una temperatura equivalente de ruido en ese punto. Dado que las etapas de un receptor tienen generalmente grandes ganancias, el ruido interno del receptor tiene usualmente una temperatura de ruido mucho menor cuando es referido a la entrada. Esas temperaturas de ruido del receptor van de cuatro grados Kelvin para amplificadores maser sobreenfriados hasta 70 a 200 grados Kelvin para receptores sin enfriamiento físico.

Ejemplo: Continuando con el ejemplo de la misión Mariner 10, la temperatura efectiva de ruido de la antena y el receptor era de 13.5 grados Kelvin. Se utilizaba una velocidad de transmisión de 117.6 kb/seg. Cual es la relación de señal a ruido (SNR) suponiendo un ancho de banda del sistema de la mitad de la velocidad de transmisión, o sea 58.8 kHz? (en otros capítulos se estudiará que este es el mínimo ancho de banda posible para transmisión binaria). La potencia total de ruido dentro del ancho de banda del receptor debería ser

$$P_n = kT_nB = 1.38\,10^{-23}\,13.5\,58.8\,10^3 = -169.6dBw$$

La SNR será entonces:

$$SNR = -162.6 + 169.6 = 7.0\,dB$$

En la práctica el ancho de banda de ruido será más grande que este, y la SNR será menor, tal vez por un par de dB. Esta SNR deberá soportar la transmisión de datos aunque a una taza de error relativamente pobre. En este caso las técnicas de codificación de canal que se estudiarán más adelante compensarán este recurso limitado.

La SNR como calculada en este ejemplo es una cantidad muy útil ya que no se alterará por la gran ganancia introducida en el receptor (señal y ruido serán afectadas de la misma forma).

Mascaras de emisión

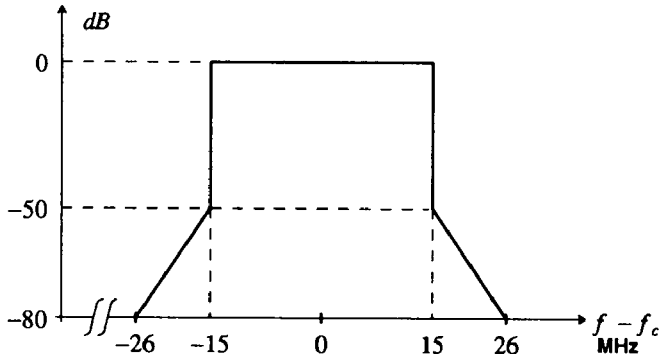

Figura 3.20: Máscara de emisión espectral referenciada a un ancho de banda de canal nominal de 30 MHz. La coordenada vertical es el espectro de potencia transmitida referenciada a una potencia de una portadora no modulada. La señal del usuario debe estar por debajo de esta máscara (aplicable en EEUU).

Un canal de radio no introduce en sí mismo ninguna restricción significativa sobre el ancho de banda que es posible utilizar para comunicaciones digitales. Además, el canal de espacio libre introduce una atenuación

con dependencia lentamente variable con la frecuencia (debida a la ganancia de antena). De esta forma, no existe nada inherente al canal que introduzca una motivación significativa para ser espectralmente eficiente. Es ahí donde aparecen las regulaciones estatales. Dado que el espectro es un recurso escaso, el mismo es asignado cuidadosamente a usuarios individuales. A diferencia de la fibra óptica, donde diferentes usuarios pueden instalar su propia fibra, todos nosotros debemos compartir el mismo contexto radial. Para prevenir una interferencia significativa entre usuarios, por regulación se especifican las *máscaras de emisión espectral.* Un ejemplo de máscara se muestra en la figura 3.20. En este caso, la agencia regulatoria ha asignado un ancho de banda nominal de 30 MHz centrado en f_c a un usuario particular, pero por razones prácticas ha permitido al usuario transmitir una pequeña cantidad de potencia (más abajo de 50 dB) fuera de esa banda.

Esta máscara se introduce colocando un filtro de frecuencia de corte abrupta en el transmisor de radio. Dado que este filtro se impone por restricciones externas, es natural incluirlo como parte del canal (obviamente que este razonamiento es una simplificación ya que los requerimientos del filtro dependen del espectro de la señal de entrada al filtro). Desde esta perspectiva, el canal de radio microondas tiene bandas laterales muy abruptas, en contraste con los medios que se analizaron anteriormente los cuales tienen a lo sumo un incremento gradual de la atenuación con la frecuencia. Esta característica es compartida por los canales voz para datos de la próxima sección, y por esta razón se utilizan técnicas de modulación similares en ambos medios físicos.

Desvanecimiento multicamino

Se a discutido como puede determinarse el balance de potencias para un enlace de radio. El cálculo realizado supone en su mayor parte idealizaciones, mientras que en la práctica debería incluirse *márgenes de sistema* adicionales para tener en cuenta situaciones imprevisibles o circunstanciales. Por ejemplo, en altas frecuencias existirá una *atenuación por lluvia* adicional en la antena receptora durante tormentas. En sistemas de microondas terrestres existe una fuente adicional de atenuación importante que debe tenerse en cuenta: el *desvanecimiento multicamino.* La atenuación por lluvia y el desvanecimiento multicamino resultan en una atenuación en el camino de la señal que varía con la frecuencia. Una diferencia significativa es que entre ambas atenuaciones es que el desvanecimiento multicamino resulta en una gran atenuación dependiente de la frecuencia para una señal de ancho de banda estrecho. Este fenómeno se conoce como *desvanecimiento selectivo.*

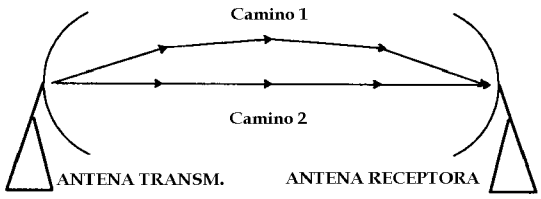

Figura 3.21: Ilustración de dos caminos entre las antenas de radio de transmisión y recepción. El desvanecimiento resulta cuando los dos caminos tienen diferentes retardos de propagación.

El mecanismo para el desvanecimiento multicamino se muestra en la figura 3.21, y es muy similar a la distorsión de modo en las fibras ópticas multimodo y a la distorsión introducida por los puentes de adaptación en pares trenzados, excepto por el hecho de ser variante en el tiempo. La atmósfera no es homogénea a la radiación electromagnética debido a las variaciones espaciales en temperatura, presión, humedad y turbulencia. Esta no homogeneidad resulta en variaciones en el índice de refracción, lo que resulta en posiblemente dos o más haces electromagnéticos que viajan del transmisor al receptor. Otra fuente de multicamino es la reflexión de ondas de radio debida a obstáculos, tal como edificios. La longitud efectiva de los caminos puede ser diferente para diferentes caminos y en general interferirán entre sí dado que el receptor percibirá solo la suma de las señales.

Es posible determinar el efecto del desvanecimiento multicamino sobre una señal pasabanda usando la respuesta compleja banda base equivalente para un único haz y aplicando superposición. Si suponemos dos haces con atenuaciones A_1 y A_2 y distancias de propagación d_1 y d_2, correspondientes a retardos de propagación $\tau_1 = d_1/c$ y $\tau_2 = d_2/c$, es posible definir dos parámetros $\Delta d = d_1 - d_2$ y $\Delta\tau = \tau_1 - \tau_2$. Entonces por superposición la función transferencia equivalente compleja banda base del canal será

$$A_1 e^{-j\omega\tau_1} e^{-jkd_1} + A_2 e^{-j\omega\tau_2} e^{-jkd_2} = A_1 e^{-j\omega\tau_1} e^{-jkd_1} (1 + \frac{A_2}{A_1} e^{j\omega\Delta\tau} e^{jk\Delta d})$$

Los primeros términos tienen un desplazamiento de fase constante y lineal debido al retardo τ_1, idéntico al primer haz. El término entre paréntesis es importante, porque muestra una complicada dependencia de la frecuencia debido a interferencia constructiva y destructiva de las dos señales en el receptor.

El parámetro crítico importante es $\Delta\tau$, denominado *dispersión de retardo*. Pueden distinguirse dos casos distintos. El primero ocurre cuando, para las frecuencias de banda base de interés, $|\omega\Delta\tau| << \pi$, tal que la dependencia de la frecuencia del segundo término no es significativa. Esto se denomina *modelo de banda estrecha*. Para este caso la propagación de dos haces es similar a un único camino, lo que resulta en un retardo (un desplazamiento de fase lineal con la frecuencia) más un desplazamiento de fase constante. El caso contrário se denomina *modelo de banda ancha* y resulta en una dependencia con la frecuencia más complicada debido a interferencia constructiva y destructiva.

Ejemplo 1: Se define la transición entre los modelos de banda estrecha y banda ancha como una dispersión de retardo tal que $|\omega\Delta\tau| = 0.01\pi$ (1.8 grados) a la más alta frecuencia de la banda base de interés. En forma equivalente, es posible esperar que $f = 1/200\Delta\tau$ sea la más alta frecuencia. Entonces si la dispersión de retardo es de 1 nanoseg, los canales banda base con un ancho de banda menor que 5 MHz se consideran de banda estrecha y anchos de banda mayores a 5 MHz (especialmente aquellos significativamente más grandes) se consideran de banda ancha. Si la dispersión de retardo aumenta en 100 nanoseg el canal de banda estrecha tiene un ancho de banda menor a 50 kHz siguiendo este critério. Notar que todo lo que hay que tener en cuenta es la dispersión de retardo, y no el valor absoluto del retardo ni la frecuencia de la portadora. Notar también que la señal pasabanda real tiene un ancho de banda el doble de la señal banda base equivalente compleja.

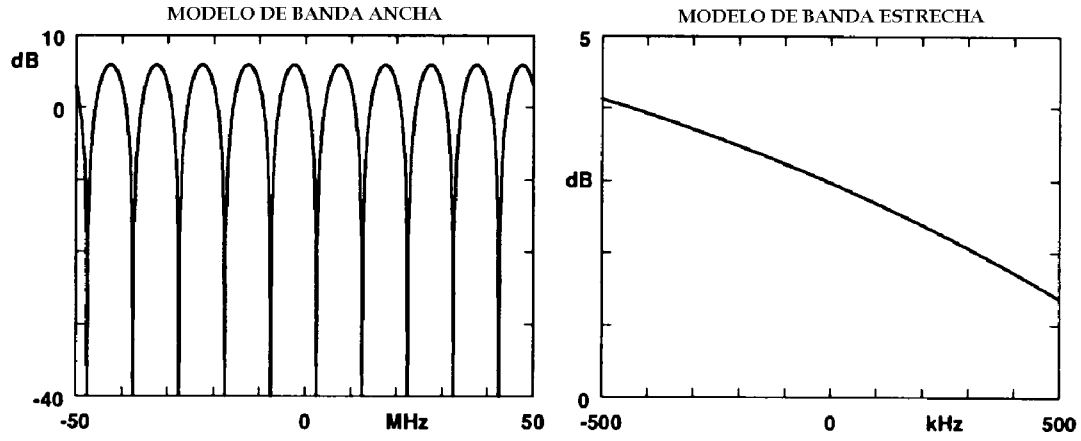

Figura 3.22: Respuesta de amplitud del canal banda base complejo sobre rango de frecuencias amplio y un rango de frecuencias estrecho para el modelo de dos caminos con $\rho = 0.99$.

Ejemplo 2: Para el caso de dos haces, el cuadrado de la magnitud de la respuesta en frecuencia para el término de interés dependiente de la frecuencia es

$$|1 + \rho e^{j\omega\Delta\tau}|^2 = 1 + |\rho|^2 + 2\, Re\{\rho e^{j\omega\Delta\tau}\}$$

para alguna constante compleja ρ. Es posible elejir una dispersión de retardo de 10 nanoseg (una situación típica de peor caso en un contexto urbano) y un valor razonablemente grande $|\rho| = 0.99$. Esto se ilustra en

la figura 3.22 en un rango de ± 50 MHz, para un modelo de banda ancha y un modelo de banda angosta. Es posible notar las atenuaciones abruptas debido a interferencia destructiva en algunas frecuencias, acentuadas por que los dos haces tienen aproximadamente la misma amplitud. Es posible notar también que la ganancia es de cerca de 6 dB a algunas frecuencias debido a interferencia constructiva. El modelo de banda estrecha se graficó sobre un rango de ± 500 kHz, el cual por el critério anterior es un modelo de banda estrecha. En este caso la respuesta del canal varía solo un par de dB sobre todo el rango de frecuencias.

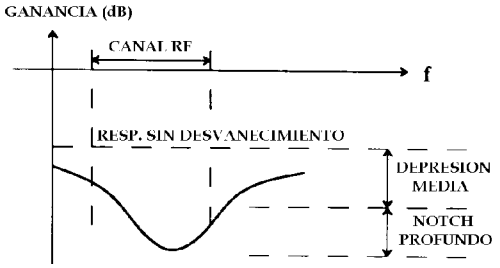

Figura 3.23: Notch típico selectivo en frecuencia debido al desvanecimiento con su terminología. El impacto sobre el canal depende fuertemente de la ubicación del notch en relación al ancho de banda del canal.

El modelo de dos haces, usualmente adecuado para sistemas de microondas terrestres, sugiere que el desvanecimiento puede resultar en un cambio de ganancia monótono a lo largo del canal o en atenuaciones abruptas (notch) en la respuesta del canal dentro del ancho de banda de interés. Una respuesta típica de un canal con desvanecimiento se muestra en la figura 3.23, donde los parámetros típicos que caracterizan el desvanecimiento son identificados.

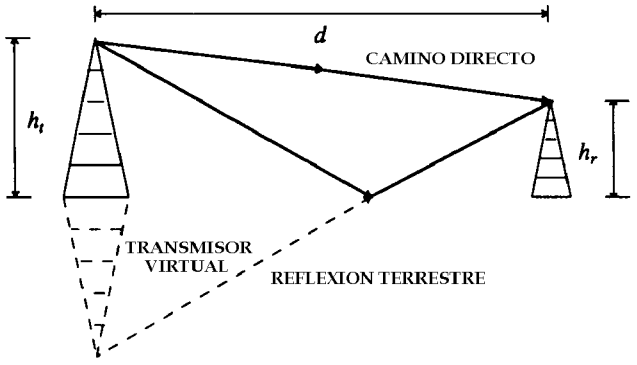

Figura 3.24: La atenuación de un sistema de microondas terrestre se incrementa debido a la reflexión en tierra. Existirán reflexiones en tierra de todos los puntos entre las dos antenas, pero se muestra una reflexión simple como si la tierra fuera un espejo perfecto. La altura de la antena de transmisión es h_t y la de recepción es h_r. La distancia entre antenas es d.

En la sección de antenas y transmisión de microondas se mostró que la pérdida de potencia en la radio propagación en el espacio libre obedece una relación de ley cuadrática, o sea, la potencia recibida disminuye con d^{-2}, o el camino de pérdidas en dB aumenta como $20 \log_{10} d$. Para la transmisión de microondas terrestre, el camino de pérdidas aumenta más rapidamente que en el espacio libre, tipicamente como d^{-4} o $40 \log_{10} d$ en dB. Esto puede explicarse usando el modelo simple de la figura 3.24. Aún para antenas con alta direccionalidad, para un valor de d grande existirá una reflexión substancial en la tierra que interferirá en la antena receptora. Tipicamente la tierra es próxima a un cortocircuito para ángulos oblícuos de incidencia a frecuencias de microondas, lo que implica un coeficiente de reflexión próximo a -1 (tal que los campos eléctricos incidente y reflejado suman cero).

Ejemplo 3: Para la geometría de la figura 3.24, consideremos solo la reflexión resultante si la tierra actúa

como un espejo perfecto y que los haces directo e indirecto sufren pérdidas de espacio libre. Suponemos la distancia entre antenas mucho mayor que las alturas de las antenas. Es posible mostrar que la pérdida de potencia neta resultante es aproximadamente

$$\frac{P_R}{P_T} = \left[\frac{P_R}{P_T}\right]_{\text{espacio libre}} \left[\frac{4\pi h_t h_r}{\lambda d}\right]^2$$

y en consecuencia el efecto de la reflexión es una interferencia destructiva que aumenta el camino de pérdidas en otro factor de d^{-2} addicional al de la pérdida de espacio libre.

Es posible notar intuitivamente que es ventajoso tener antenas altas (la pérdida disminuye con el cuadrado de las alturas de antena).

Aún cuando el transmisor y el receptor están en posiciones fijas uno respecto al otro, el desvanecimiento es un fenómeno variante en el tiempo para distancias grandes (30 km a mayores) debido a los fenomenos atmosféricos. La profundidad y la duración de los desvanecimientos son de considerable importancia para los diseñadores de sistemas de radio. Afortunadamente, se ha observado que cuanto más profundo es el desvanecimiento con menor frecuencia ocurre y de menor duración será cuando ocurre. También, la severidad del desvanecimiento aumenta cuando la distancia entre antenas aumenta o la frecuencia de la portadora aumenta. El desvanecimiento puede mitigarse también utilizando *técnicas de diversidad*, en las cuales dos o más canales independientes se combinan. La filosofía en este caso es que solo uno de esos canales a la vez se verá afectado por el desvanecimiento.

Radio móvil

Uno de los usos más llamativos de la radiotransmisión es para comunicaciones con personas o vehículos en movimiento. Para este tipo de comunicaciones no existe realmente alternativa a la radiotransmisión, excepto por el infrarrojo, el cual no trabaja bien en exteriores. La radio móvil exibe algunas características que son diferentes de la transmisión punto a punto. Primero, las antenas deben ser generalmente omnidireccionales y de esa forma exiben mucho menor ganancia de antena. Segundo, pueden existir obstáculos para la propagación directa, que causan el *efecto de ensombrecimiento* que resulta en grandes variaciones de la potencia de señal recibida en función de la ubicación. Tercero, la aplicación más común es en áreas urbanas, donde existen muchas alternativas de reflexiones múltiples, y el modelo de dos haces deja usualmente de ser preciso. Cuarto, el usuario está frecuentemente en movimiento, lo que resulta en variaciones extremas en las condiciones de transmisión aún sobre distancias cortas y además aparece un desplazamiento Doppler en la frecuencia de la portadora.

El modelo de dos haces puede extenderse facilmente a un modelo de M caminos usando nuevamente superposición. En este caso, la salida banda base compleja del canal es

$$\sum_{i=1}^{M} A_i u(t-\tau_i) e^{-jkd_i}$$

donde A_i son los coeficientes reales de atenuación, d_i es la longitud del i-ésimo camino, y τ_i es el retardo de propagación del i-ésimo camino. Puede existir un camino dominante cuyo coeficiente de atenuación obedece una ley de cuarta potencia con la distancia, pero los otros coeficientes dependerán de los coeficientes de reflexión de los caminos indirectos y en consecuencia se relacionarán en forma complicada con la posición. Además, debido a los efectos de ensombrecimiento puede que no exista un camino dominante. Por ejemplo, si el receptor móvil está localizado detrás de un edificio, las ondas de radio sufrirán una pérdida por difracción. Esta pérdida de ensombrecimiento varía tipicamente en forma remarcable sobre una distancia de decenas a centenas de metros. Si se promedia la potencia recibida sobre un área del orden de 1 km^2, sería posible verificar la atenuación de cuarta potencia con la distancia, pero si se promedia sobre un área del orden de 1 m^2 se verá una fluctuación adicional con la posición debido al ensombrecimiento. El ensombrecimiento ocurre frecuentemente como el resultado de una distribución *log-normal* en la potencia promedio local recibida, o sea, la potencia expresada en dB tiene distribución Gaussiana. La desviación estándar de la potencia expresada en dB es rigurosamente 4 dB para áreas urbanas típicas.

Cuando se examina la potencia recibida en forma local, no promediando sobre un área, es posible percibir grandes fluctuaciones debido al desvanecimiento multicamino. Para un vehículo en movimiento, son frecuentes desvanecimientos de 40 dB o mayores por debajo del nivel promedio local, con mínimos sucesivos ocurriendo a cada mitad de longitud de onda (o fracción de metro a frecuencias de microondas). De esta forma, el movimiento del vehículo introduce una dimensión totalmente nueva en relación al desvanecimiento que experimenta un sistema punto a punto, donde las fluctuaciones son mucho más lentas. Esta fluctuación rápida se conoce como *desvanecimiento Rayleigh* debido a que la distribución de la envolvente de la portadora recibida frecuentemente obedece una distribución Rayleigh.

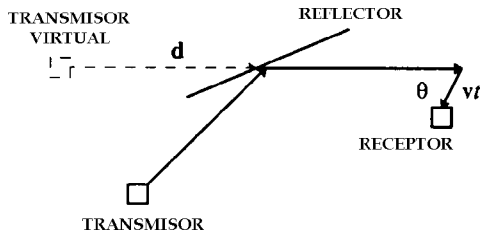

Figura 3.25: Trayectoria del movimiento de un vehículo con velocidad uniforme, relativa al camino de propagación que incluye una reflexión.

Para entender el desvanecimiento Rayleigh, se debe examinar el efecto del movimiento del vehículo que resulta en una fase de la portadora recibida variante en el tiempo. Como antes, esto puede comprenderse considerando un único haz (camino) y aplicando luego superposición a los caminos múltiples. La geometría de un camino simple se ilustra en la figura 3.25, donde se incluye una reflexión entre el transmisor y el receptor. Como mostrado, puede definirse detrás del reflector un transmisor virtual con una propagación lineal al receptor. $\boldsymbol{d}$ es un vector del transmisor virtual al receptor en el instante $t = 0$, $\boldsymbol{v}$ es el vector de velocidad para el vehículo en el instante $t = 0$, θ es el ángulo entre $\boldsymbol{d}$ y $\boldsymbol{v}$, o el ángulo de incidencia del camino de propagación relativo a la velocidad del vehículo. La distancia escalar inicial y la velocidad serán $d = \|\boldsymbol{d}\|$ y $\nu = \|\boldsymbol{v}\|$. El vector del transmisor al receptor es $\boldsymbol{d} + \boldsymbol{v} \cdot \boldsymbol{t}$, y la distancia de propagación como una función del tiempo es

$$\|\boldsymbol{d} + \boldsymbol{v} \cdot \boldsymbol{t}\| = \left(d^2 + \nu^2 t^2 + 2 < \boldsymbol{d}, \boldsymbol{v} > t\right)^{1/2}$$

donde el producto interno es $< \boldsymbol{d}, \boldsymbol{v} >= d\nu \cdot \cos\theta$. Esta distancia no cambia linealmente con la frecuencia, pero puede aproximarse por una función lineal del tiempo.

Ejemplo: Es posible mostrar que la ecuación anterior puede aproximarse en forma suficientemente precisa por $d + \nu t \cdot \cos\theta$. Por ejemplo, si $d = 1$ km y $\nu = 30$m/seg (aproximadamente 100 km/h) entonces la aproximación se verifica para $t << 66$ seg.

La escala de tiempos sobre la cual es válida la aproximación lineal de la distancia es suficientemente grande relativa a las fluctuaciones significativas de la fase de la portadora y en consecuencia es seguro asumir que la distancia al receptor cambia con $\nu \cdot \cos\theta \cdot t$. Este cambio en distancia tiene una pendiente $+\nu$ cuando el receptor se mueve directamente desde el transmisor, $-\nu$ cuando se está moviendo hacia el transmisor, y cero cuando el receptor se está moviendo orthogonal al transmisor.

Con este resultado geométrico básico a mano, la señal banda base compleja recibida será

$$A \cdot Re\{u(t - \frac{d}{c} - \frac{\nu}{c}\cos\theta\, t)e^{-jkd}e^{-jk\nu\cos\theta\, t}e^{j\omega_c t}\}$$

Es posible observar aquí varios efectos de propagación. Primero, la señal banda base $u(t)$ está retrasada por una cantidad variante en el tiempo debido a la distancia de propagación cambiante. Este efecto es generalmente poco significativo en las frecuencias de banda base de interés. Segundo, Existe un desplazamiento

de fase estático e^{-jkd} debido a la distancia de propagación en $t = 0$. Tercero, y más interesante, existe un desplazamiento de fase que varía linealmente con el tiempo. En efecto, es un corrimiento de frecuencia, conocido como *corrimiento Doppler*. La frecuencia de la portadora se desplaza de ω_c a $\omega_c - \omega_d$, tal que la frecuencia Doppler es

$$\omega_d = kd\cos\theta = \frac{2\pi\nu}{\lambda}\cos\theta$$

Cuando el receptor se mueve alejandose del transmisor el corrimiento Doppler es negativo y, por el contrário, cuando se acerca el corrimiento será positivo.

Ejemplo: Si la velocidad del vehículo es $\nu = 30$ m/seg (100 km/hr) y la frecuencia de la portadora es 1 GHz ($\lambda = 0.3$ m), entonces el corrimiento Doppler máximo es $f_d = \nu/\lambda = 100$ Hz. Esto permite concluir que en forma relativa a la frecuencia de la portadora, el corrimiento Doppler es tipicamente muy pequeño, pero relativo a las frecuencias de banda base puede ser relativamente grande. Es posible notar también que para una velocidad constante del vehículo el corrimiento Doppler se vuelve mayor cuando se incrementa la frecuencia de la portadora.

Además de afectar la distancia de propagación y el ángulo de incidencia, la reflexión en la figura 3.25 afecta también la constante de atenuación e introduce un desplazamiento de fase desconocido debido al coeficiente de reflexión.

El corrimiento Doppler en sí mismo no es un aspecto muy importante, dado que resulta en un desplazamiento de la frecuencia de la portadora que puede no ser conocida muy precisamente en ese punto. El efecto más substantivo ocurre cuando existen dos o más caminos, cada uno con diferente corrimiento Doppler debidos a sus ángulos de incidencia en el receptor diferentes. Si la dispersión de retardo de los diferentes caminos es pequeña es posible suponer un modelo de banda estrecha, o sea, los diferentes retardos de las réplicas de señal banda base $u(t)$ no son significativos para las frecuencias de banda base de interés. La superposición resultante de diferentes corrimientos Doppler puede resultar en fluctuaciones rápidas de fase y amplitud. Por ejemplo, para un conjunto de caminos con amplitud A_i, suponiendo retardos $\tau_i = \tau$ iguales para todos los caminos (lo que justamente representa al modelo de banda estrecha), retardos de fase ϕ_i en el instante cero, máximo corrimiento Doppler ω_d y ángulos de incidencia θ_i, la señal banda base compleja recibida será

$$\sqrt{2}\cdot\{\sum_i A_i u(t-\tau_i)e^{j(\phi_i-\omega_d\cos\theta_i t)}\} = \sqrt{2}\cdot\{u(t-\tau)r(t)\}$$

donde

$$r(t) = \sum_i A_i e^{j(\phi_i-\omega_d\cos\theta_i t)}$$

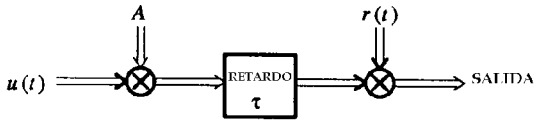

Figura 3.26: Modelo para el canal banda base con un receptor en movimiento a velocidad uniforme.

El modelo equivalente banda base complejo se muestra en la figura 3.26. El efecto es la multiplicación por una onda compleja $r(t)$. Es instructivo graficar esta onda para un par de casos. Por ejemplo, en la figura 3.27 se muestra el efecto de sumar dos portadoras con un corriemiento Doppler relativo de 100 Hz. El resultado es desvanecimientos a intervalos de 10 mseg, siendo este el recíproco del corrimiento Doppler. Aparecen saltos de fase muy rápidos y periódicos, correspondientes precisamente a las veces en que hay desvanecimientos de

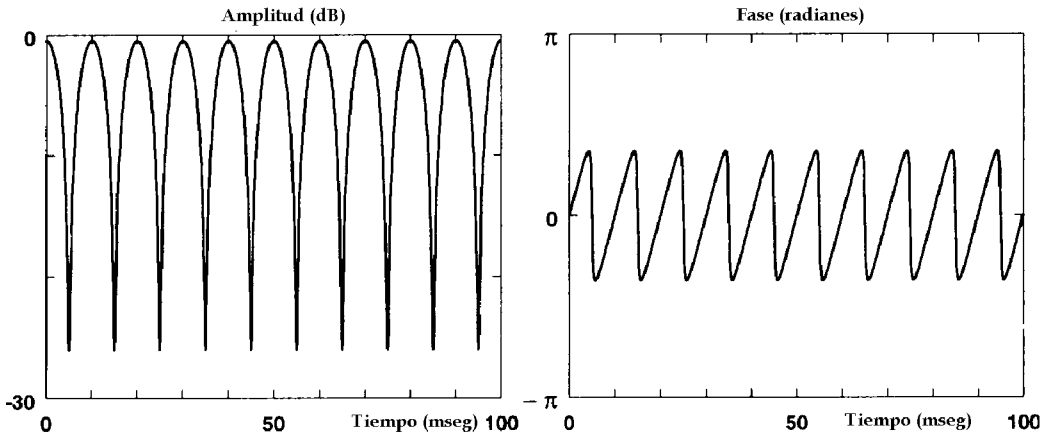

Figura 3.27: Amplitud y fase de $r(t)$ resultantes de la superposición de dos caminos con desplazamiento Doppler de 0 a 100 Hz, con $A_1 = 1$ y $A_2 = 0.9$.

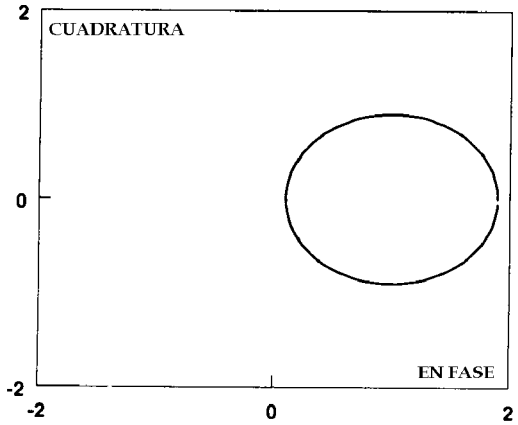

Figura 3.28: Gráfico polar de la trayectoria de $r(t)$ para el mismo caso de la figura 3.27.

amplitud grandes. Este efecto se explica mediante la figura 3.28, la cual muestra un gráfico polar. La onda $r(t)$ sigue una trayectoria circular con una velocidad angular constante. Los desvanecimientos de amplitud ocurren cuando la trayectoria está próxima al orígen, lo cual coincide con el instante en que la fase cambia más rapidamente.

Esos gráficos se repiten para 40 señales llegando desde direcciones uniformemente espaciadas en las figuras 3.29 y 3.30. Mientras el resultado es cualitativamente similar, la trayectoria es mucho más complicada y con forma aleatória. Nuevamente existen desvanecimientos de amplitud profundos ocasionalmente los cuales coinciden con rápidas variaciones de fase. La escala de tiempos de esos desvanecimientos rápidos es nuevamente del orden de 10 mseg, el cual es el recíproco del máximo corrimiento Doppler. Esto corresponde también al instante en que el receptor recorre la mitad de la longitud de onda de la frecuencia de la portadora.

Los cambios caóticos en amplitud y fase con el tiempo mostrados en la figura 3.29 pueden caracterizarse estadísticamente empleando el teorema del limite central. Volviendo al modelo de la figura 3.26, es posible modelar la señal multiplicativa $r(t)$ como un proceso estocástico $R(t)$. Examinando este proceso en algún instante t_0, la fase del i-ésimo camino será $\xi_i = \phi_i - \omega_d \cos\theta_i \, t_0$. Dado que las fases ϕ_i son funciones muy sensibles a la posición inicial, es razonable suponer que los ξ_i son variables aleatórias *independientes e identicamente distribuidas* (i.i.d.) sobre el intervalo $[0, 2\pi]$. Escribiendo las partes real e imaginaria independientemente,

$$Re\{R(t_0)\} = \sum_i A_i \cos\xi_i, \qquad Im\{R(t_0)\} = \sum_i A_i \sin\xi_i$$

donde cada término es la suma de variables aleatórias independientes. Por el teorema del límite central

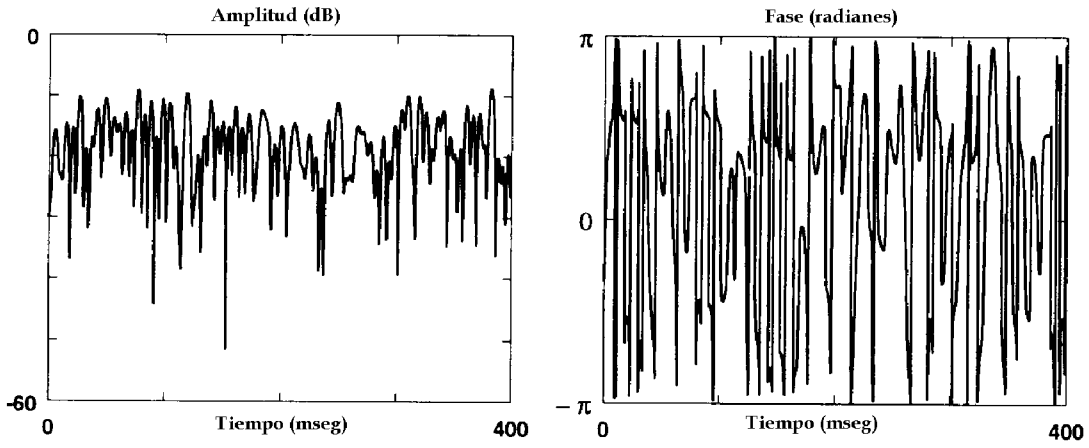

Figura 3.29: a) amplitud y b) fase para la superposición de 40 señales llegando en ángulos uniformemente distribuidos, cada una con la misma amplitud y fases aleatórias.

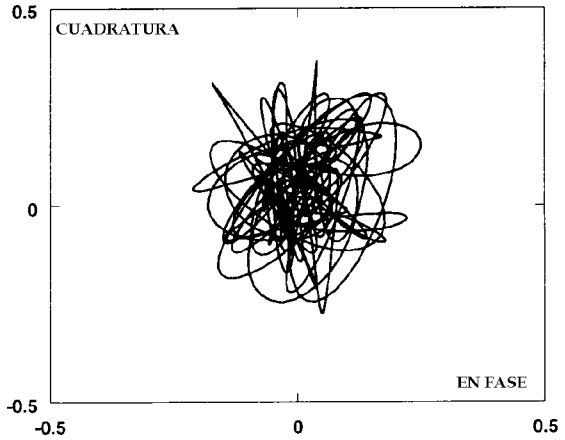

Figura 3.30: Gráfico polar para el mismo caso de la figura 3.29.

(Capítulo 2), cuando el número de términos aumenta, $Re\{R(t_0)\}$ y $Im\{R(t_0)\}$ tendrán una distribución Gaussiana, y en consecuencia $R(t)$ será una variable aleatória Gaussiana compleja.

Ejemplo: Es posible mostrar que, con la suposición que ξ_i son variables aleatórias i.i.d, entonces

$$E[Re\{R(t_0)\}] = E[Im\{R(t_0)\}] = \sigma^2$$
$$\sigma^2 = \tfrac{1}{2}\textstyle\sum_i A_i^2$$
$$E[Re\{R(t_0)\}Im\{R(t_0)\}] = 0$$

Cuando

$$R(t_0) = Re^{j\Theta}$$

es una variable aleatória Gaussiana compleja con partes real e imaginaria i.i.d, entonces R es una variable aleatória Rayleigh, cuya función densidad de probabilidad está dada por

$$f_R(r) \begin{cases} \frac{r}{\sigma} e^{-\frac{r^2}{2\sigma}}, & r \geq 0 \\ 0, & r < 0 \end{cases}$$

y la fase Θ tendrá una distribución uniforme en $[0, 2\pi]$. La amplitud R es la envolvente de la portadora recibida y Θ es su fase, de lo que puede decirse que la envolvente tendrá distribución Rayleigh y su fase distribución uniforme.

La conclusión es que cuando se transmite una portadora contínua, la señal recibida $R(t)$ es adecuadamente aproximada por un proceso Gaussiano complejo. El espectro de potencia de esa proceso puede calcularse si hacemos suposiciones adicionales sobre la distribución de potencia recibida en función del ángulo. Esto se debe a que la frecuencia de las componentes que llegan al receptor dependen directamente del coseno del ángulo de llegada. Es posible suponer que $R(t)$ tiene espectro de potencia $S_R(jw)$. La contribución a $R(t)$ que llega en un ángulo θ está relacionada con la frecuencia $[\omega_c - k\nu, \omega_c + k\nu]$. En particular, la potencia total recibida en la banda $[\omega_0, \omega_c + k\nu]$ corresponde a ángulos de llegada en el rango

$$k\nu \cdot \cos\theta + \omega_c \geq \omega_o, \quad \text{ó} \quad \theta| \leq \theta_0 = \cos^{-1}\left[\frac{\omega_0 - \omega_c}{k\nu}\right]$$

Si se supone por ejemplo, que la potencia total recibida llega uniformemente sobre todos los ángulos $\theta \leq \pi$, entonces la porción de potencia que llega en la banda $[\omega_0, \omega_c + k\nu]$ debe ser $P\theta_0/\pi$. De esta forma,

$$\int_{\omega_o}^{\omega_c + k\nu} S_R(j\omega)\frac{d\omega}{2\pi} = \frac{P}{\pi}\cos^{-1}\left[\frac{\omega_0 - \omega_c}{k\nu}\right]$$

y diferenciando ambos lados con respecto a ω_0, el espectro de potencia es

$$S_R(j\omega) = \frac{2P}{\sqrt{(k\nu)^2 - (\omega - \omega_c)^2}}, \quad |\omega - \omega_c| \leq k\nu,$$

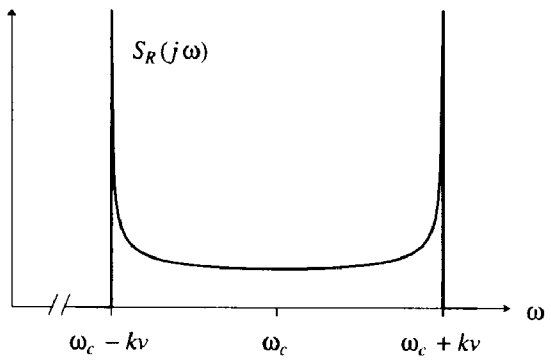

Figura 3.31: Espectro de potencia Doppler de la portadora recibida por un vehículo viajando a velocidad ν, suponiendo que la potencia de la señal recibida se distribuye uniformemente sobre todos los ángulos de llegada.

y cero para todo otro valor (obviamente el espectro es simétrico alrededor de $\omega = 0$). Este espectro de potencia se muestra en la figura 3.31, donde se observa que la potencia está concentrada en la región de frecuencias $\omega_c \pm k\nu$. Una realización del proceso estocástico con este espectro de potencia puede interpretarse como una versión aleatória de la señal determinística $\cos(\omega_c t)\cos(k\nu t)$, dado que esta última tiene una transformada de Fourier compuesta por funciones delta en $\omega_c \pm k\nu$. Esta señal de AM-BLU es la portadora multiplicada por una envolvente con cruces por cero periódicos (desvanecimientos) espaciados en intervalos de $\pi/k\nu = \lambda/2\nu$ seg. Este es justamente el tiempo que le lleva al vehículo para desplazarse la mitad de longitud de onda. De esta forma, temporalmente, el desvanecimiento Rayleigh ocurre frecuentemente cada

mitad de longitud de onda cuando la potencia está uniformemente distribuida sobre todos los ángulos de llegada.

Ejemplo: Si la velocidad del vehículo es de 100 km/hr y la frecuencia de la portadora es de 1 GHz, la máxima frecuencia Doppler es aproximadamente 100 Hz. Esto significa que los caminos individuales que van al receptor pueden tener corrimientos Doppler del orden de $\pm$ 100 Hz, o que el ancho de banda de la señal pasabanda aumenta 200 Hz debido al movimiento del vehículo. La longitud de onda es alrededor de 0.3 m, tal que el tiempo que toma para que el vehículo se desplaze media longitud de onda es

$$t = \frac{0.15\, m}{30\ \text{m/seg}} = 5\ \text{mseg}$$

Entonces, es posible esperar desvanecimientos significativos cada 5 mseg, el cual es el recíproco del rango de corrimientos Doppler de 200 Hz.

El modelo de la figura 3.26 y la obtención del desvanecimiento Rayleigh suponían un modelo de banda estrecha, o sea, la dispersión de retardos pequeña con respecto a la recíproca del ancho de banda, o equivalentemente que los retardos τ_i son idénticos para todos los caminos. En consecuencia, el modelo debe ser modificado para tener en cuenta un modelo de banda ancha cuando justamente la señal es de banda suficientemente amplia. Usualmente esto se maneja como sigue. Primero, se tiene en cuenta que cualquier reflexión, tal como las asociadas con edificios altos, es una complicada superposición de reflexiones múltiples de forma que la dispersión de retardo asociada es lo suficientemente pequeña como para obedecer el modelo de banda estrecha. De esta forma, esta reflexión simple puede representarse por un modelo de banda estrecha de desvanecimiento Rayleigh con un retardo τ_1 asociado. Si existe una segunda reflexión con retardo significativamente diferente, ella puede representarse con otro modelo de desvanecimiento Rayleigh con retardo $\tau_2 \neq \tau_1$. El modelo de banda ancha se obtiene de la superposición de esos modelos de banda estrecha.

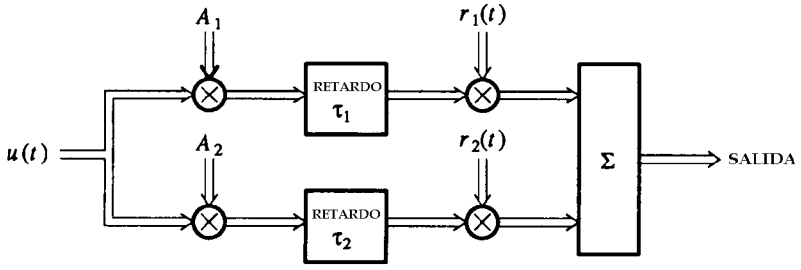

Figura 3.32: Modelo de dos caminos de banda ancha, donde cada camino se supone con desvanecimiento Rayleigh independiente.

Un modelo de banda ancha de dos caminos se ilustra en la figura 3.32. La señal banda base compleja $u(t)$ experimenta los dos retardos τ_1 yτ_2 y las dos salidas retrasadas se multiplican por procesos Gaussianos complejos independientes $r_1(t)$ y $r_2(t)$. Cada camino tiene asociado también una atenuación A_i y un desplazamiento estático de fase que puede asociarse a $r_1(t)$ y $r_2(t)$. Este modelo de banda ancha puede generalizarse facilmente a un número arbitrário de caminos.

3.1.5 Canales telefónicos (*)

Muchos lugares en el mundo pueden ser alcanzados debido a la red telefónica pública, por lo que el canal de ancho de banda de voz es un vehículo casi universal para las comunicaciones de datos. El diseño de modems digitales para canales telefónicos es realmente un desafío debido a que el canal fué diseñado primariamente para voz y los desacoples que no son serios para voz pueden serlo para señales asociadas a datos. El canal telefónico es un ejemplo básico de un canal compuesto consistente en varios medios tal como pares de conductores, canales satelitales, cable coaxil, enlaces terrestres de microonda y fibra óptica. Aún más importante que el medio son los varios sistemas de modulación construidos sobre ellos, tal como PCM y BLU. Las características del canal varian ampliamente dependiendo de la conexión particular. Es útil discutir esas

Imperfección	**Nivel**
Atenuación de un tono de 1004 Hz	27 dB
Relación señal a ruido *C filtrado*	20 dB
Distorsión de 2da. armónica	34 dB
Distorsión de 3ra. armónica	33 dB
Offset de frecuencia	3 Hz
Jitter de fase pico a pico (2-300 Hz)	20 grados
Jitter de fase pico a pico (20-300 Hz)	13 grados
Ruido impulsivo (umbral -4 dB)	4 por minuto
Saltos de fase (umbral de 20°)	1 por minuto
Retardo completo (sin satelites)	50 mseg

Tabla 3.1: Imperfecciones típicas de peor caso para canales telefónicos.

características, no solo debido a la importancia de este canal particular, sino también porque es posible encontrar desacoples que pueden occurrir en otras situaciones.

Medidas del desempeño de canales telefónicos

Debido a la amplia variedad de conexiones posibles, no existe un modelo analítico simple del canal telefónico. Los diseñadores de modems se refieren en general a caracterizaciones estadísticas de los circuitos telefónicos. El diseñador necesita determinar el porcentaje aceptable de conexiones telefónicas sobre la cual el modem funcionará y luego hallar los parámetros de umbral que son satisfechos o excedidos por ese porcentaje de canales. Los umbrales resultantes pueden ser muy sensibles en relación al porcentaje.

Ejemplo: De acuerdo a una investigación, el 99 % de los canales punto a punto atenúan un tono de 1004 Hz en 27 dB o menos. Además, el 99.9 % de los canales atenúan el mismo tono en 40 dB o menos. Para tener el cubrimiento de ese 0.9 % extra, deberá tolerarse una pérdida de 13 dB.

En la tabla 3.1 se muestran parámetros típicos de peor caso para algunas de las imperfecciones del canal. El porcentaje de canales telefónicos que excede este desempeño es rigurosamente el 99 %. La distorsión lineal es la principal distorsión que no se encuentra en la tabla dado que es dificil sintetizarla. Esa distorsión se discutirá a continuación, además de las otras imperfecciones del canal.

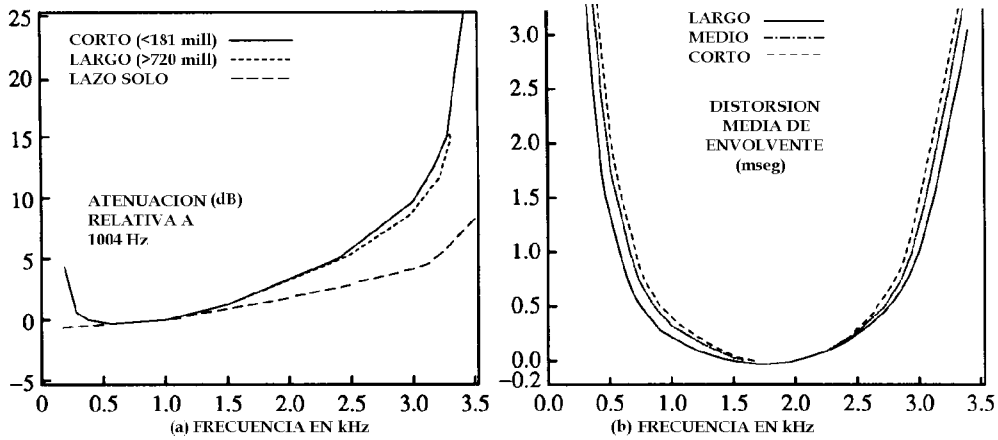

Figura 3.33: Atenuación (a) y distorsión de la envolvente de retardo de un canal telefónico típico en función de la frecuencia. La atenuación se obtiene en forma relativa aun tono de 1004 Hz, y la distorsión de la envolvente de retardo relativa a 1704 Hz, donde está próxima a su valor mínimo.

- **Distorsión lineal**: La respuesta en frecuencia de un canal telefónico puede aproximarse por una función transferencia lineal $B(j\omega)$, un filtro pasabanda desde 300 a 3300 Hz. Este ancho de banda se elige para dar una calidad de voz aceptable en la red y es forzada por los filtros pasabanda de los sistemas de modulación analógica y digital utilizados en la red. Una función transferencia típica de un canal telefónico se ilustra en la figura 3.33, donde se utilizó la siguiente terminología. La distorsión de amplitud, la magnitud de la respuesta en frecuencia, se grafica como una atenuación (ó pérdida) en función de la frecuencia. La distorsión de amplitud se sintetiza frecuentemente como un conjunto de mediciones de *pendiente de distorsión*, los cuales muestran características como las de la figura 3.33a. Una medida típica de pendiente de distorsión es la peor de las dos diferencias siguientes, 1) la atenuación en 404 Hz menos la atenuación en 1004 Hz, y 2) la atenuación en 2804 Hz menos la atenuación en 1004 Hz. Para el 99 % de las conexiones telefónicas, ese número es menor que 9 dB. Existen varias formas alternativas de caracterizar la distorsión de pendiente pero son en general más complicadas de utilizar en la práctica.

  Un aspecto interesante es que la atenuación de la figura 3.33a es casi precisamente la atenuación típica de la línea de abonado combinada con la respuesta en frecuencia de los filtros en los moduladores PCM de la figura 3.3, sugiriendo que esas son las fuentes dominantes de atenuación dependiente de la frecuencia.

  La distorsión de fase, o sea la desviación en relación a fase lineal de la respuesta de fase de $B(j\omega)$ se describe tradicionalmente como *distorsión del retardo de envolvente*. El *retardo de envolvente* se define como el valor negativo de la derivada de la fase de la señal recibida con respecto a la frecuencia, y en consecuencia mide la desviación de la respuesta de fase lineal. La distorsión del retardo de envolvente se describe por un conjunto de mediciones en forma similar a como se caracteriza la distorsión de pendiente.

  La atenuación completa del canal se mide típicamente en 1004 Hz, donde justamente la atenuación es usualmente próxima al mínimo. La atenuación es típicamente de 6 dB entre un conmutador local y otro, a lo cual se suma la pérdida en la línea de abonado en cada extremo.

- **Fuentes de ruido**: Además de la atenuación, existe ruido en el canal de banda vocal asociado a cuatro fuentes principales: *cuantización*, *ruido térmico*, *acoplamiento* y *ruido impulsivo*. Estas fuentes se discutirán en orden de importancia decreciente.

  El acoplamiento (crosstalk), ya discutido anteriormente, es un tipo de imperfección más severa para voz que para datos, de forma que ha sido notablemente reducido en la red. La inducción de tonos interferentes de 50 Hz y sus armónicos de las líneas de alimentación es más significativa. El *ruido impulsivo* consiste en picos grandes y espontáneos de corta duración y se mide contando el número de veces que el ruido excede un umbral determinado. El ruido impulsivo se debe a conmutadores electromecánicos en la red, tal como conmutadores telefónicos y telefonos de discado decádico. El ruido impulsivo no está bien caracterizado y aún así el diseño de modems no están influenciados por su presencia.

  La fuente dominante de ruido es el *error de cuantización* introducido por los sistemas PCM, como el de la figura 3.3. El error de cuantización es la consecuencia de usar un número limitado de bits para representar cada muestra en el sistema PCM. A pesar que el error de cuantización depende en forma determinística de la señal, la aleatoridad de esta le dá una característica semejante a un ruido. Tiene un espectro de potencia aproximadamente blanco y el nivel de ruido se mide usualmente por la *relación señal a ruido de cuantización* (SQNR). La SQNR para un único cuantizador se muestra en la figura 3.34. Es posible notar que sobre un rango de entrada de 30 dB (-40 a -10 dBm0) la SQNR varía solo unos 6 dB (de 33 a 39 dB). Esta SQNR casi constante implica que la potencia del error de cuantización varía casi en proporción directa con la potencia de la señal, o sea, no es constante independientemente de la señal como el ruido térmico. Para los modems más rápidos puede ser la imperfección dominante. Para modems de baja velocidad se puede aproximar por ruido blanco gaussiano.

  En la tabla 3.1, el ruido se denomina *ruido C filtrado*, lo cual hace referencia a un filtro particular (pasabanda de banda estrecha) aplicado sobre el ruido previo a la medición de la potencia. Este filtro se elige sobre la base de efectos subjetivos de la voz, y si el ruido es blanco no tiene efectos más allá de un offset fijo en la medición de la potencia. La potencia del error de cuantización se mide aplicando un *tono de mantenimiento*, usualmente en 1004 Hz, y filtrandolo con un filtro elimina banda previo a la medición del error de cuantización remanente con un mensaje C filtrado.

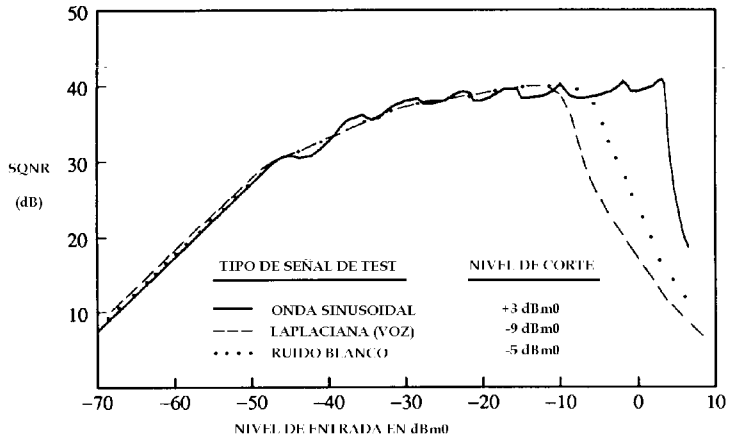

Figura 3.34: SQNR en función de la potencia de entrada absoluta para tres tipos diferentes de entrada. Las entradas gaussiana y laplaciana son aleatórias, con la última como una buena aproximación a la distribución de muestras de voz.

- **Distorsión no lineal**: La distorsión no lineal se debe a imperfecciones en amplificadores y también a errores en los conversores A/D y D/A. Como es de bajo nivel, la distorsión no lineal solo es un problema significativo en modems sofisticados de altas velocidades.

- **Desplazamiento de frecuencia**: El desplazamiento en frecuencia es particular de los canales telefónicos y canales con desplazamiento Doppler. Si la entrada al canal es $x(t)$, con transformada de Fourier $X(j\omega)$, y el canal tiene un desplazamiento de ω_0 radianes, y no existen otras imperfecciones, entonces la salida del canal tendrá una transformada de Fourier

$$Y(j\omega) = \begin{cases} X(j\omega - j\omega_0) & \text{para } \omega > 0 \\ X(j\omega + j\omega_0) & \text{para } \omega < 0 \end{cases}$$

Este pequeño desplazamiento en el espectro de la señal tiene implicaciones importantes para la recuperación de la portadora y el cancelamiento de eco.

Ejemplo: Una señal pasabanda puede expresarse como

$$x(t) = Re\{s(t)e^{j\omega_c t}\}$$

donde $s(t)$ es una señal de datos banda base compleja y ω_c es la frecuencia de la portadora. El efecto del desplazamiento de fase en el canal sobre la señal recibida será

$$y(t) = Re\{s(t)e^{j(\omega_c - \omega_0)t}\}$$

donde se supone que $\omega_0 > 0$ y $s(t)$ es limitada en frecuencia tal que $S(j\omega) = 0$ para $|\omega| > |\omega_c|$.

El desplazamiento en frecuencia es una consecuencia de utilizar frecuencias levemente diferentes para modular y demodular señales de banda lateral única en las facilidades de transmisión analógica (ver figura 3.1). Esto se permite porque no tiene efectos perceptibles en la calidad de la voz, y puede compensarse mediante un diseño adecuado del modem VF de transmisión de datos.

- **Error de fase (jitter)**: El error de fase en canales telefónicos es una consecuencia de la sensibilidad de los osciladores, utilizados para la generación de portadoras en sistemas BLU (ver figura 3.1), asociada a las fluctuaciones de la fuente de alimentación. Dado que las fluctuaciones de la alimentación son frecuentemente de 50 Hz o sus armónicas, las mayores componentes de error de fase están frecuentemente a esas frecuencias. El error de fase se mide observando la desviación de los cruces por cero de un tono de 1004 Hz de su posición nominal en el tiempo.

  El error de fase puede verse como una generalización del desplazamiento (offset) de frecuencia. Si el error de fase de un canal es $\theta(t)$, el efecto sobre la señal transmitida es otra señal recibida como

$$y(t) = Re\{s(t)e^{j(\omega_c t+\theta(t))}\}$$

  Un error de fase de $\theta(t) = \omega_0 t$ es un desplazamiento en frecuencia. Es común que $\theta(t)$ tenga componentes oscilatorias de la frecuencia de línea (50 Hz) o sus armónicas. Si solo se demodula esta señal utilizando la portadora $e^{j\omega_c t}$, se recuperará una señal banda base distorsionada $s(t)e^{j\theta(t)}$ en lugar de $s(t)$. Para reducir esta distorsión es común en la recuperación de la portadora incluir algoritmos para hacer un seguimiento y remover el error de fase indeseado.

  Un salto de fase es un cambio abrupto en la fase nominal de una señal sinusoidal recibida que dura al menos 4 mseg. Es poco lo que puede hacerse para defender un modem contra esta degradación, pero debe tenerse en cuenta en el diseño de los circuitos de recuperación de portadora.

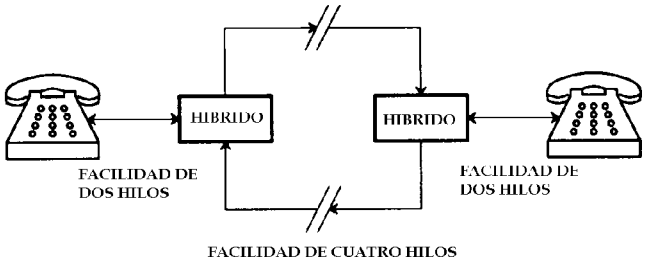

Figura 3.35: Un canal telefónico simplificado, mostrando el lazo local de dos hilos y la facilidad de transmisión de cuatro hilos.

- **Retardo y eco**: El *retardo* y el *eco* son las imperfecciones en los canales telefónicos que se considerarán para finalizar. Un canal telefónico simplificado se muestra en la figura 3.35. La *línea de abonado* o *lazo local*, o sea el par trensado que conecta la central de conmutación con el usuario, se utiliza para la transmisión en ambas direcciones. Ambas señales comparten el mismo par trensado. En la central de conmutación existe un circuito, denominado el *híbrido*, que separa las dos direcciones de transmisión. En facilidades de larga distancia se tienen *cuatro hilos*, lo que significa que las dos direcciones están separadas fisicamente.

  Una implementación posible del circuito híbrido se muestra en la figura 3.36. La señal del otro lado de la facilidad de dos hilos se alimenta por los conectores de recepción. La señal transmitida aparece en el transformador como un divisor de voltage con impedancias R y Z_0, tal que esta última es la impedancia de la facilidad de dos hilos. Es posible cancelar esta desviación indeseada construyendo otro divisor de voltage con una *impedancia de balance* Z_B. Cuando $Z_B = Z_0$, la pérdida de los hilos del transmisor a los del receptor es infinita. En la práctica se utiliza una impedancia fija Z_B de compromiso y una componente de la señal recibida (A) se fuga hacia (B), con una atenuación tan pequeña como 6 a 10 dB, debido a la variación de impedancia de la facilidad de dos hilos.

  En la figura 3.37 se ilustran la señal y dos tipos de caminos de eco para la configuración de la figura 3.35. Un *eco* se define como una componente de señal que ha seguido algún camino diferente del *camino de voz del locutor*. El *eco del locutor* es la señal que se fuga a través del híbrido de la terminación lejana y vuelve a quien lo envio (el locutor). El *eco del escucha* es la componente del eco del locutor que se fuga a través de la terminación cercana del híbrido y vuelve nuevamente al escucha. Este eco es similar al de propagación multicamino en canales de radio. La longitud del canal telefónico determina el retardo de eco completo. Los ecos desde la terminación cercana de la conexión van desde 0 a 2 mseg

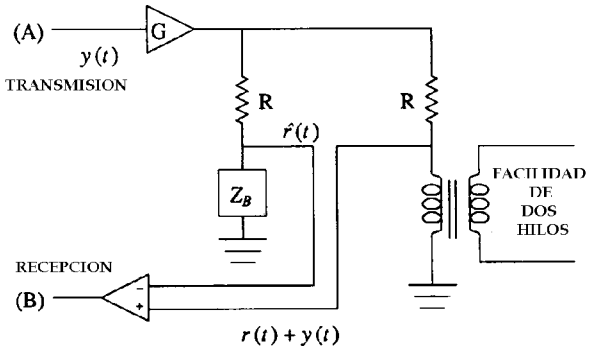

Figura 3.36: Un híbrido electrónico. Para evitar la fuga de la señal recibida (A) en el camino de transmisión (B), la impedancia Z_B debe acoplarse exactamente con la impedancia del transformador y la facilidad de dos hilos.

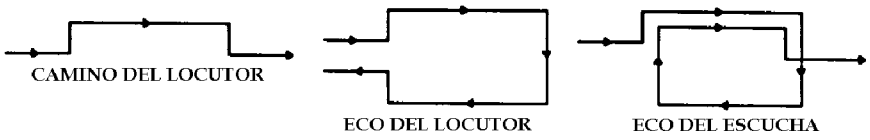

Figura 3.37: Tres formas posibles de caminos de señal en un canal telefónico simplificado con una conversión de dos a cuatro hilos en cada terminación.

de retardo, mientras que los ecos de terminación lejana pueden tener un retardo completo de 10 a 60 mseg para facilidades terrestres, y hasta 600 mseg en conexiones satelitales.

Para reducir los efectos del eco en la calidad de la voz, coexisten varias estratégias en la red. El efecto de cada estratégia sobre las señales de datos es diferente. Para retardos cortos, la pérdida se adiciona en el camino de voz del locutor, lo cual es ventajoso ya que los ecos experimentan esta pérdida más de una vez. Esta pérdida, más la pérdida de la línea de abonado en cada terminación, es la fuente de atenuación que debe tenerse en cuenta en la transmisión de datos. Puede ser tan alta como 40 dB (en 1004 Hz). Para retardos más largos existen elementos conocidos como *supresores de eco* ó *canceladores de eco* que se adicionan a la conexión.

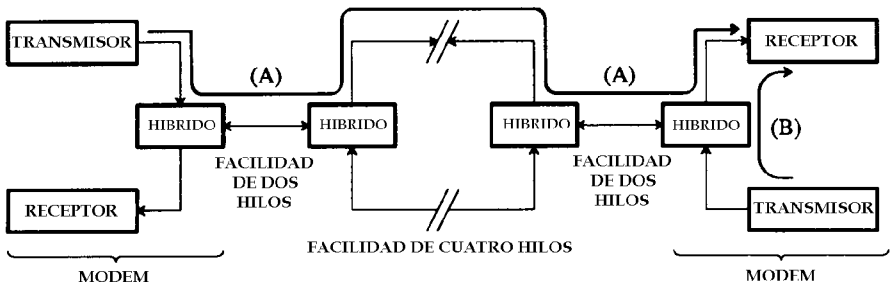

Figura 3.38: Dos modems conectados sobre un canal telefónico simplificado. El receptor de la derecha debe conseguir distinguir la señal deseada (A) de la señal de fuga de su propio transmisor (B).

Un modem *full duplex* (FDX) transmite y recibe sobre el mismo canal telefónico. Ese modem requiere una conversión interna de dos a cuatro hilos, como se muestra en la figura 3.38. Debido al balance imperfecto de impedancias de los híbridos, algo de la señal transmitida hace eco en el receptor e interfiere con la señal débil de datos de la terminación lejana. Esta pérdida del híbrido puede ser tan baja como 6 dB, y la señal recibida puede haber experimentado tanto como 40 dB de pérdida, tal que la señal de la terminación lejana podría estar tan bajo como 34 dB por debajo del eco.

velocidad (b/seg)	**Velocidad por símbolo**	**Método duplex**	**Estándar CCITT**	**modulación**
≤ 300	≤ 300	full (FDM)	V.21	2-FSK
1200	1200	half	V.23	2-FSK
1200	600	full (FDM)	V.22	4-FSK
2400	1200	half	V.26	2-FSK
2400	600	full (FDM)	V.22bis	16-QAM
2400	1200	full (EC)	V.26ter	4-PSK
4800	1600	half	V.27	8-PSK
4800	2400	full (EC)	V.32	4-QPSK
9600	2400	half	V.29	16-AM/PM
9600	2400	full (EC)	V.32	32-QAM+TC
14400	2400	full (EC)	V.32bis	128-QAM+TC
≤ 28800	≤ 3429	full (EC)	V.fast(V.34)	1024-QAM+TC

Tabla 3.2: La columna duplex indica cuando un único canal es compartido en ambas direcciones para transmisión o deben usarse canales separados en cada dirección (half). Para modems full duplex, también indica donde se utiliza multiplexado por división en frecuencia (FDM) ó cancelamiento de eco (EC) como técnica de acceso múltiple. La columna CCITT identifica el estándar internacional que se aplica a este tipo de transmisión. Finalmente, la columna modulación identifica el tipo de modulación a ser discutida en próximos capítulos. Los números indican el número de símblos en el alfabeto. TC en V.32 y V.34 refiere a codificación *Trellis*.

Capacidad del canal comparada con modems prácticos

Una estimación rigurosa de la capacidad de un canal de banda vocal telefónica indica unos 30000 b/seg. Las velocidades alcanzables en modems de datos estandarizados se sintetizan en la tabla 3.2. Velocidades tan altas como 28800 b/seg son alcanzables y aún mayores solo en una pequeña fracción de las conexiones posibles. Además, varios de los modems de alta velocidad estandarizados se utilizan exclusivamente con líneas especiales, las cuales pueden condicionarse adecuadamente para garantizar la calidad de la transmisión.

3.1.6 Canales de almacenamiento magnético (*)

Las comunicaciones digitales no se utilizan solo para comunicaciones a distancia (de un lugar a otro), sino también para las comunicaciones en el tiempo (de un instante a otro). Esta última aplicación se denomina *almacenamiento digital* o *grabación*, y se logra usualmente a través de un medio magnético en la forma de cinta o disco. Más recientemente, particularmente en el contexto de aplicaciones de lectura solamente, se ha utilizado también un medio de almacenamiento óptico.

Ejemplo 1: El disco compacto ROM, una mejora de una tecnología similar de audio, permite el almacenamiento de más de 700 Mbytes de datos en un disco plástico de 12 cm de diámetro. Los bits son almacenados como pequeñas cavidades en una superficie y son leídos girando el disco apuntando un diodo laser sobre la superfície y detectando la luz reflejada con un sensor óptico.

El almacenamiento digital se utiliza extensivamente en sistemas de cómputo, pero está siendo utilizado crecientemente además como medio de almacenamiento de música o voz.

Ejemplo 2: El sistema de audio digital de disco compacto, el cual es un gran éxito comercial, almacena música digitalmente utilizando una tecnología similar a la del disco compacto ROM. La música es convertida a la forma digital usando 16 bits/muestra a una frecuencia de muestreo de 44.1 kHz en dos canales, para lograr una velocidad de transmisión total de 1.4 Mb/seg. Es posible grabar alrededor de unos 70 minutos

de material sobre un único disco.

Ejemplo 3: La grabación digital se utiliza en sistemas de voz de almacenamiento y reproducción, los cuales son esencialmente el reemplazo funcional de los sistemas de respuesta automática telefónica, excepto porque sirven para un gran número de clientes.

La grabación digital ofrece algunas de las ventajas del almacenamiento analógico, discutidas antes para trasmisión. La principal ventaja nuevamente es el *efecto regenerativo*, en el cual la grabación no se deteriora con el tiempo (excepto por la introducción de errores aleatórios que pueden ser eliminados mediante técnicas de codificación) o con grabaciones múltiples o regrabaciones. Una ventaja adicional es la compatibilidad de la grabación digital con el procesamiento digital de la señal, lo cual ofrece poderosas alternativas.

La cinta magnética o disco pueden considerarse como medios de transmisión de la misma forma que otros medios como cables o fibras. Se discutirá a seguir algunas propiedades de este medio.

Escritura y lectura

En el proceso de escritura, se genera un campo magnético en un electromagneto, denominado *cabezal*, cuando pasa a alta velocidad sobre un medio ferromagnético orientando la orientación de la magnetización a lo largo de un surco en la cercanía del medio magnético sobre el disco o cinta. En la lectura, cuando el patrón magnético orientado pasa bajo el mismo cabezal produce un voltage que puede sensarse y amplificarse.

Existen dos tipos básicos de grabación. La *grabación por saturación* se utiliza casi siempre para grabación digital en la cual la magnetización es saturada en una dirección o la otra. De esta forma, en la grabación por saturación el medio se utiliza solo para transmisión binaria, o sea, se permiten solo dos niveles. Esto contrasta con medios cableados y coaxiales en los cuales se contempla la transmisión multinivel. La otra forma de grabación magnética es *grabación con polarización a.c.*, en la cual la señal se acompaña con una sinusoide de polarización mucho más fuerte y de mayor frecuencia con el propósito de linealizar el canal. Este tipo de grabación es utilizado necesariamente en la grabación analógica, cuando la linealidad es importante, pero no ha sido aplicada a la grabación digital por el deterioro en la relación señal a ruido y porque la grabación por saturación es más adecuada para técnicas binarias de modulación y demodulación.

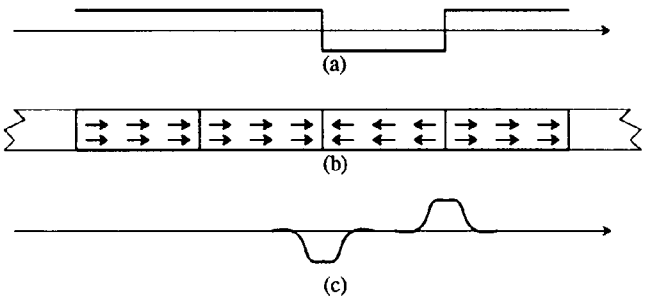

Figura 3.39: Ilustración de la grabación magnética. a) La forma de onda NRZ aplicada al cabezal de grabación es la secuencia "1101". La abcisa es tiempo, y es también proporcional a la distancia entre el medio por la velocidad constante del cabezal. b) La magnetización de un surco después de la magnetización por saturación. c) el voltage en la bobina del cabezal de lectura correspondiente a la posición del cabezal de lectura, el cual a una velocidad constante es lo mismo que tiempo.

El proceso de grabación magnética se ilustra cualitativamente en la figura 3.39. Para la grabación por saturación, el voltage aplicado al cabezal de escritura asume un valor positivo y un valor negativo correspondientes a las dos direcciones de magnetización deseada. En la figura 3.39a se supone aplicada al cabezal de escritura una onda cuadrada correspondiente a la secuencia binaria "1101". La forma de onda correspondiente a esa secuencia de bits se denomina *no retorno a cero* (NRZ). La secuencia de bits se graba sobre surcos lineales (cintas) o circulares (discos) sobre el medio magnético; un surco se muestra en la figura 3.39b. Es posible notar las dos direcciones de magnetización, esquematicamente indicadas por flechas. El voltage en el cabezal de lectura (que fisicamente es el mismo que el de escritura) durante una operación de lectura se muestra en la figura 3.39c. Mientras la magnetización es constante no se induce voltage en la

bobina del cabezal de lectura, pero en base a un cambio en la magnetización se inducirá un voltage (recordar que el voltage inducido en una bobina es proporcional a la derivada del campo magnético). La polaridad de ese voltage está determinada por la *dirección de cambio* en la magnetización.

Este proceso magnético de escritura-lectura puede verse como un canal de comunicaciones si se observa solo las formas de onda de los voltages de entrada y salida en la figura 3.39a y 3.39b y se ignora el medio físico de la figura 3.39b. Ambas formas de onda representan señales en el tiempo, tal como en comunicaciones, a pesar que existe un retardo de tiempo insignificante e indeterminado entre las operaciones de lectura y escritura. Visto como un canal de comunicaciones, es posible notar que el canal de grabación magnética de la figura 3.39 incluye inherentemente una operación de diferenciación. Otra forma de ver esto es teniendo en cuenta que el canal es sensible solo a *transiciones* en la forma de onda de entrada antes que a su polaridad. En consecuencia, desde el punto de vista de comunicaciones digitales, se desea mapear los bits de entrada en transiciones en la forma de onda de entrada antes que en polaridad absoluta.

Linealidad del canal magnético

El canal magnético puede hacerse lineal en cierto sentido como explicado a continuación. Esta linealidad es una característica muy deseable, porque simplificará enormemente el diseño del sistema.

La perspectiva de la figura 3.39 es sobresimplificada ya que supone que la magnetización es en una dirección u otra. En realidad, el medio magnético contiene una multiplicidad de finas particulas y cada partícula debe además magnetizarse en una dirección u otra. La magnetización total neta puede suponer casi un contínuo de valores, dependiendo del número de particulas magnetizadas en cada dirección. Desafortunadamente este contínuo de magnetización depende en forma no lineal del campo magnético aplicado y sufre de histéresis, y en consecuencia el proceso de escritura es altamente no lineal. Por otro lado, el proceso de lectura es muy lineal, ya que el voltage inducido sobre el cabezal de lectura es una función lineal de la magnetización.

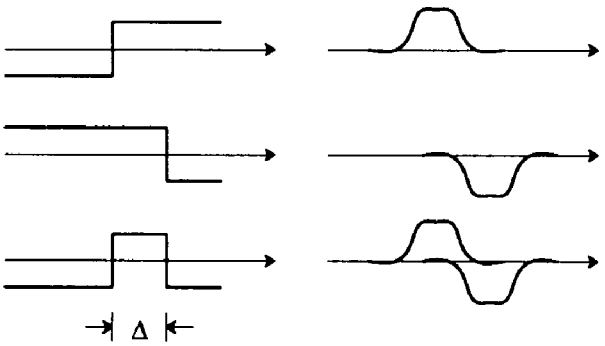

Figura 3.40: Superposición en el proceso de lectura de la grabación magnética.

Si el campo aplicado en el cabezal de lectura es suficientemente fuerte y se mantiene lo suficiente como para que el medio sea totalmente saturado, entonces la salida del cabezal de lectura muestra una forma de superposición. Esto se debe a que la saturación destruye la memoria de la histéresis. Esta forma de superposición se ilustra en la figura 3.40. Si la respuesta a una transición positiva en el instante t es $h(t)$, y la respuesta a una transición negativa en el instante $t + \Delta$ es $-h(t + \Delta)$, entonces la respuesta a una transición positiva seguida de una negativa obedece a superposición y es

$$h(t) - h(t + \Delta)$$

Esto es cierto con gran precisión siempre que el tiempo entre transiciones Δ sea mayor que cierto umbral Δ_0. Este umbral está determinado por el tiempo necesario para alcanzar una saturación total del medio en una dirección o la otra luego de la última transición, y depende del diseño del cabezal de escritura así como también del medio.

3.2 Modelos simplificados de canales de comunicaciones

En el diseño de sistemas de comunicaciones para transmitir información a través de canales físicos es conveniente contruir modelos matemáticos que reflejen las características más importantes del medio de transmisión. De esta forma, el modelo del canal se utilizará en el diseño del codificador del canal y el modulador en el transmisor y el demodulador y el decodificador de canal en el receptor. A continuación se introduce una breve descripción de los modelos de canal que se utilizan frecuentemente para caracterizar varios de los canales físicos encontrados en la práctica.

3.2.1 Canal de ruido aditivo

El modelo matemático más simple de un canal de comunicaciones es el canal de ruido aditivo, ilustrado en la figura 3.41. En este modelo, la señal transmitida $s(t)$ está corrompida por un proceso estocástico de ruido aditivo $n(t)$. Físicamente, el proceso de ruido aditivo puede aparecer debido a componentes electrónicos y amplificadores en el receptor del sistema de comunicaciones o debido a interferencia encontrada en la transmisión como en el caso de transmisión de señales de radio.

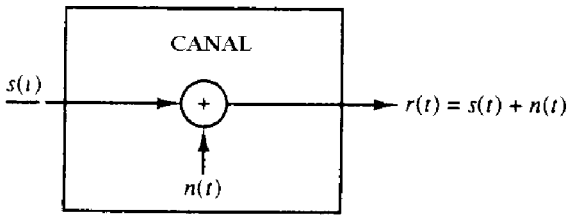

Figura 3.41: Canal de ruido aditivo.

Si el ruido es introducido primariamente por componentes electrónicos y amplificadores en el receptor, puede caracterizarse como ruido térmico. Este tipo de ruido está caracterizado estadísticamente por un *proceso de ruido Gaussiano*. De esta forma, el modelo matemático resultante para el canal se denomina usualmente *canal de ruido aditivo blanco Gaussiano* (AWGN). Como este canal se aplica a una amplia variedad de canales de comunicaciones físicos y debido a su tratabilidad matemática, es el modelo de canal predominante en el análisis y diseño de sistemas de comunicaciones que se discutirá en nuestro contexto. La atenuación del canal se incorpora facilmente en el modelo. Cuando la señal transmitida sufre atenuación en la transmisión a través del canal, la señal recibida será

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

donde α es en general un factor de atenuación.

3.2.2 Canal de filtro lineal invariante en el tiempo

En algunos canales físicos tal como canales telefónicos cableados, se utilizan filtros para asegurar que las señales transmitidas no exceden determinadas restricciones de ancho de banda y de esta forma no interfieran con otras señales similares. Tales canales están generalmente caracterizados matematicamente como canales con ruido aditivo y filtrado lineal, como ilustrado en la figura 3.42. De esta forma, si la entrada al canal es la señal $s(t)$, la salida del canal será

$$r(t) = \int_{-\infty}^{\infty} h(\tau)s(t-\tau)d\tau + n(t)$$

donde $h(t)$ es la respuesta impulsiva del filtro lineal.

3.2.3 Canal de filtro lineal variante en el tiempo

Otros canales físicos, como por ejemplo el caso de canales de radio de propagación ionosférica, resultan en una señal transmitida con propagación multicamino variante en el tiempo lo cual obviamente puede

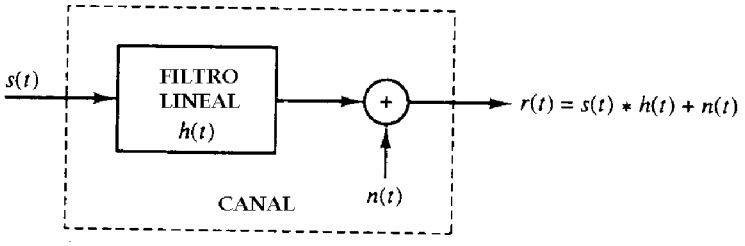

Figura 3.42: Canal de filtrado lineal con ruido aditivo.

caracterizarse matematicamente mediante un canal variante en el tiempo con respuesta impulsiva $h(\tau;t)$, tal que $h(\tau;t)$ es la respuesta del canal en el instante t debido a un impulso aplicado en el instante $t-\tau$. Así, τ representa la "historia" (tiempo transcurrido). El canal de filtro lineal variante en el tiempo con ruido aditivo se ilustra en la figura 3.43. Para una señal de entrada $s(t)$, la señal de salida del canal será

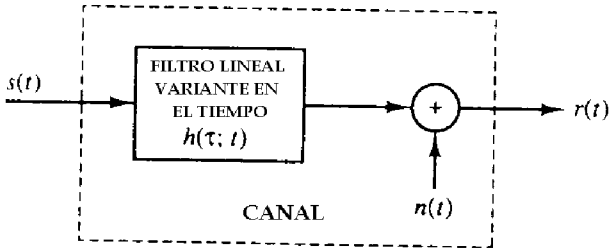

Figura 3.43: Canal de filtrado lineal variante en el tiempo con ruido aditivo.

$$r(t) = \int_{-\infty}^{\infty} h(\tau;t)s(t-\tau)d\tau + n(t)$$

Un modelo adecuado para la propagación de señales multicamino a través de canales físicos, tal como transmisión ionosférica (frecuencias por debajo de 30 MHz) y canales de radio celulares, es un caso especial de la ecuación anterior en el cual la respuesta impulsiva variante en el tiempo tiene la forma

$$h(\tau;t) = \sum_{k=1}^{L} \alpha_k(t)\delta(\tau-\tau_k)$$

donde $\alpha_k(t)$ representan los factores de atenuación posiblemente variantes en el tiempo asociados a los L caminos de propagación. Sustituyendo esta ecuación en la anterior, la señal recibida tendrá la forma

$$r(t) = \sum_{k=1}^{L} \alpha_k(t)s(t-\tau_k) + n(t)$$

En consecuencia, la señal recibida consite en L componentes multicamino, cada unos de ellos atenuado en $\{a_k(t)\}$ y retrasado $\{\tau_k\}$.

Los tres modelos matemáticos descriptos caracterizan adecuadamente a la gran mayoría de los canales físicos encontrados en la práctica y discutidos anteriormente. Estos tres modelos de canal serán utilizados en los Capítulos siguientes para el diseño y análisis de sistemas de comunicaciones digitales.