



UNIVERSIDAD DE A CORUÑA
Departamento de Electrónica y Sistemas

TESIS DOCTORAL

TÉCNICAS AVANZADAS DE PROCESADO DE SEÑAL PARA
SISTEMAS MULTIPLE INPUT MULTIPLE OUTPUT (MIMO)

Autor: MIGUEL GONZÁLEZ LÓPEZ

Director: LUIS CASTEDO RIBAS

Febrero de 2004



UNIVERSIDAD DE A CORUÑA
Departamento de Electrónica y Sistemas

TESIS DOCTORAL

**TÉCNICAS AVANZADAS DE PROCESADO DE SEÑAL PARA
SISTEMAS MULTIPLE INPUT MULTIPLE OUTPUT (MIMO)**

Autor: MIGUEL GONZÁLEZ LÓPEZ

Director: LUIS CASTEDO RIBAS

Febrero de 2004

Dr. Luis Castedo Ribas
Catedrático de Universidad
Departamento de Electrónica y Sistemas
Universidad de A Coruña

CERTIFICA:

Que la memoria titulada “Técnicas Avanzadas de Procesado de Señal para Sistemas Multiple Input Multiple Output (MIMO)”, ha sido realizada por D. Miguel González López bajo mi dirección en el Departamento de Electrónica y Sistemas de la Universidad de A Coruña y concluye la Tesis que presenta para optar al grado de Doctor.

A Coruña, 18 de diciembre de 2003

Fdo: Dr. Luis Castedo Ribas
Director de la Tesis Doctoral

A mi familia

Agradecimientos

En esta sección de agradecimientos merece un lugar privilegiado mi director de tesis, el Dr. Luis Castedo Ribas, porque sin su ayuda hubiese sido imposible realizar este trabajo. Él me enseñó que se deben poner en primer lugar los agradecimientos de tipo técnico pero, en su caso, él sabe que este agradecimiento es tanto a nivel técnico como personal.

Mi más profundo agradecimiento también al Dr. Javier García Frías, de la Universidad de Delaware (USA), por compartir conmigo sus conocimientos e ideas, que sin duda han sido fundamentales para completar esta tesis y que me han abierto nuevos terrenos de investigación para el futuro próximo. Haré honor a su condición de experto en redundancia (codificación) para repetir de modo literal lo de “en su caso, él sabe que este agradecimiento es tanto a nivel técnico como personal”.

También me gustaría hacer llegar mi agradecimiento a mis compañeros del Grupo de Señales y Sistemas de Comunicación (GSSC), Carlos, Adriana, Dani, Joaquín, Paula y Fran. A Mónica, antigua miembro del grupo, hoy abriéndose camino en la Universidad de Nueva York en Stony Brook. A mis colegas del laboratorio 1.2, especialmente a Fran y a Quique. A los miembros del Departamento de Electrónica y Sistemas, por demostrar que aquí, como quien dice en el *Finis Terrae*, se puede realizar una investigación de calidad.

A Rama y a Charo Dopico, por haberme descubierto la belleza de las matemáticas. A Juan M. Vilar y R. Cao por hacer lo mismo con la estadística. A Antonio Yáñez, por sus magistrales clases de sistemas operativos.

A Carlos y a Teresa, por sus lecciones y tutorías de cálculo y, sobre todo, porque fueron los primeros en decirme que *existe* la investigación.

A mis amigos y compañeros de carrera Ronnie, Gustavo y Paula.

A Manuel y Carlos, por los ratos y retos pasados y alcanzados en la puesta en marcha del servidor “esmelle”.

A S. Verdú por sus contribuciones a este campo y por ser español, por su curso sobre *multiuser* en la Carlos III y por la foto que nos sacamos todos después.

To Rachel (o Raquel, ya que es de esas personas con la inmensa fortuna de ser totalmente bilingües), *for her patience and dedication in helping me learn to speak English just three*

months before I moved to the US. For teaching me that there is no 'r' and no "castel" but "casel". However, I will never forgive her for failing to teach me to say "how you doin'?" instead of "how do you do?"

A mis alumnos del Máster en Producción e Xestión Audiovisual (MPXA). A mis alumnos de la Facultad de Informática.

A Pedro, por haber creído en mí.

A Óscar Loureda por sus consejos técnicos, en este caso desde el punto de vista lingüístico y comunicativo, sobre cómo afrontar una tesis. Siento no haber podido aplicarlos con mayor profundidad.

No querría dejar de mencionar a los contribuyentes que, en gran medida, han subvencionado esta tesis a través de los siguientes organismos: la Unión Europea (proyecto ESCORT), la Xunta de Galicia mediante su beca predoctoral y el Ministerio de Educación, Cultura y Deporte mediante su beca FPU. Mi agradecimiento no es óbice para reclamar una mayor atención por parte del ente público a la investigación.

En el terreno de lo personal¹ quisiera expresar mi más profunda gratitud a papá (así, literal). A mamá. Al abuelo. A la abuela.

A Alicia.

A Elena. A Jaime, a quien los que nos dedicamos a la investigación en comunicaciones tenemos que agradecer que se haya dedicado a la farmacia y sea en ese campo donde *arrase*. A Blanca, por sus confusiones lingüísticas y porque, pese a ellas, es capaz de cantar en japonés. A Piki, que sabe hablar élfico. A Ana.

A mis amigos *de siempre* Javi, Santi, Queco, Alberto, Jerónimo (hoy en el INRIA, Francia) y Quique.

A Quique y Héctor (Ness), en memoria de los buenos ratos pasados en el piso durante mi primer año de carrera.

A Carlos Abal, porque es de bien nacidos el ser agradecidos. A Andrea, porque me ha demostrado lo que es la amistad. A Andrés Barbé, aunque no crea en la Universidad. A D. Fer por su lógica borrosa. A Guzmán. A Belén, por ser la única amiga que se levantó a las cinco y media de la mañana, cogió el coche y me fue a despedir al aeropuerto el día que me iba a USA. A Delia. A Lorena. A Pepe. A Rosa. A Paco, que es quien mejor conoce lo que me ha costado acabar esta tesis.

A María.

To Wei whom I have to thank for his help and support during my stay in Delaware, and to the rest of the Chinese, Yujing, Ying and Hanqing.

To Baloo, whom I wish the best and a successful career in the promising field of nanotechnology.

¹Todas las personas aquí citadas saben cuáles son los motivos de mi agradecimiento, por lo que simplemente me limitaré a citar, y de vez en cuando, algún motivo explícito, habitualmente más por su valor anecdótico que real. (*Those persons mentioned hereafter know why I am grateful to them, so I will simply mention their names in most cases, occasionally with an anecdotal reference that may not convey the true value of their contribution.*)

To Father Szupper and Kim for making the St. Thomas More Oratory, a Catholic church located just next to the University of Delaware Campus, a place where people find God. To the rest of the people who attend the Oratory.

To Jim Snow and Fr. Juan. To John and Darlene and their wonderful kids. A Pedro, Rosa y Pedro hijo.

To Chris, Kevin (now with the University of Hawaii), Philip and Hedvig for those nice weekends in Alexandria, Virginia. I hope we can meet again soon. To Richard, my roommate in the US and also a visiting scholar of Dr. García's research group. To Alain (now you can call me "Dr. Lopez" for real) and Tina, his fiancé. May their marriage be happy.

Y, por supuesto, a todas aquellas personas que, de una u otra forma, me han ayudado a realizar esta tesis. Es probable que me deje a alguien así que entonaré un "gracias compañeros y amigos".

Resumen

La utilización de múltiples antenas tanto en transmisión como en recepción sobre un enlace radio conlleva un aumento de la capacidad teórica del canal *Multiple-Input Multiple-Output* (MIMO) resultante. Este hecho abre el camino hacia la obtención de altas eficiencias espectrales sobre el canal radio constituyendo, por tanto, una vía muy prometedora para satisfacer los altos requerimientos de tasa de transferencia de información de la futura Cuarta Generación (4G) de sistemas de comunicaciones móviles. En este trabajo se presentan distintas técnicas encaminadas a un uso efectivo de los canales MIMO, explotando la nueva dimensión espacial asociada a los mismos. Es importante resaltar que el coste de esta dimensión es además mucho menor que el del ancho de banda y/o tiempo.

Primeramente, se exponen los resultados obtenidos por distintas técnicas MIMO en un entorno de propagación real: los túneles subterráneos de una red de transporte metropolitano. La utilidad de estos resultados viene dada por la necesidad de los operadores europeos de transporte urbano de disponer de sistemas de comunicación con alta tasa de transferencia para soportar múltiples aplicaciones (voz, vídeo, señales de control, ...). Además, siendo GSM la interfaz radio adoptada como estándar para los sistemas de comunicación de dichos operadores, todas las técnicas aquí presentadas hacen uso de dicha interfaz. El rendimiento fue evaluado utilizando medidas experimentales de la respuesta al impulso del canal dentro de los túneles del Metro de París, realizadas en el contexto del proyecto de la Unión Europea ESCORT (*Enhanced diversity and Space-time Coding for underground metro and Railway Transmission*).

Las técnicas de procesado Turbo se han mostrado como una excelente vía para conseguir altos rendimientos en términos de la relación señal a ruido necesaria para obtener tasas de error de bit suficientemente bajas para propósitos prácticos. La aplicación del principio Turbo a sistemas MIMO es estudiada en el contexto de la ecualización del canal presente en túneles subterráneos. Extendemos posteriormente dicha técnica al caso más realista en el que el receptor desconoce el canal, abordándolo desde dos perspectivas diferentes: la estimación de canal y la igualación *Decision Feedback*. En ambos casos presentamos técnicas basadas en el principio de Máxima-Verosimilitud (MV) cuyo rendimiento es excelente, sobre todo debido a que se benefician del procesado Turbo.

Por otra parte, los esquemas denominados *Bit-Interleaved Coded Modulation* (BICM) se han mostrado como una alternativa a *Trellis Coded Modulation* (TCM), sobre todo para canales multitrayecto. De nuevo, la aplicación del principio Turbo a la decodificación iterativa de BICM proporcionó ganancias significativas, aportando además a BICM una robustez mayor que la de TCM. Esta robustez ha de entenderse en el sentido de ser una técnica que funciona satisfactoriamente en todos los casos que pueden presentarse en un enlace radio: desde desvanecimiento Rayleigh puro, pasando por el de tipo Rice, hasta el caso extremo de comportamiento AWGN.

En esta tesis se considera la utilización de BICM en sistemas MIMO, con la novedad importante de utilizar un nuevo tipo de código de canal: los códigos LDGM (*Low-Density*

Generator Matrix) sistemáticos concatenados. Estos códigos son una alternativa a los códigos *Low-Density Parity Check* (LDPC) y Turbo, compartiendo con ellos el hecho de rendir de modo cercano al límite teórico de la capacidad de Shannon del canal. La ventaja principal de los códigos LDGM es su reducida complejidad en transmisión. La complejidad del algoritmo *Belief Propagation* (BP) utilizado para su decodificación, también aplicado para los códigos LDPC, es mucho menor que la correspondiente a la decodificación de códigos Turbo.

Summary

Employing multiple antennas at both the transmission and reception sides of a radio link provides an increase in the capacity associated to the resulting Multiple-Input Multiple Output (MIMO) channel. This fact opens up the way towards the achievement of high spectral efficiencies over the radio channel, thus constituting a promising way for satisfying the high information transfer rate requirements of the forthcoming Fourth Generation (4G) of mobile communication systems. In this work several techniques aimed to an effective use of MIMO channels are presented, exploiting the new spatial dimension related to them. Moreover, it is worth emphasizing that the cost of this dimension is much lower than that of bandwidth and/or time.

Firstly, results obtained by several MIMO techniques within a real propagation environment, i.e. the underground tunnels of a metropolitan transport system, are presented. The usefulness of these results comes from the European urban transport operators needs of having high throughput communication systems in order to support several applications (voice, video, control signals, ...). Besides, being GSM the radio interface standard adopted by those operators for their communication systems, all the techniques presented here employ this interface. Performance was evaluated using experimental measurements of the channel impulse response existing inside the tunnels of the Metro of Paris, carried out in the context of the European Union ESCORT (*Enhanced diversity and Space-time Coding for underground metrO and Railway Transmission*) project.

Turbo processing techniques have revealed as an excellent way for achieving high performances in terms of the required signal to noise ratio to obtain sufficiently low bit error rates for practical purposes. Applying the Turbo principle in MIMO systems is studied in the context of equalizing the channel present in underground tunnels. We subsequently extend this technique to a more realistic scenario in which the receiver does not know the channel, tackling it from two different perspectives: channel estimation and Decision Feedback equalization. In both cases we present techniques based on the Maximum Likelihood principle whose performance is excellent, especially because they benefit from the Turbo processing.

On the other hand, the so-called Bit-Interleaved Coded Modulation (BICM) schemes have revealed as an alternative to Trellis Coded Modulation (TCM), especially over multipath channels, which are those present in the radio links over which MIMO systems operate. Again, the application of the Turbo principle to iteratively decoding of BICM produced significant gains, providing BICM with a robustness higher than that of TCM. This robustness has to be understood in the sense of being a technique which performs satisfactorily in all the cases that can arise in a radio link: from Rayleigh fading, going through a Rice one, up to the extreme case of AWGN behaviour.

In this thesis the employment of BICM in MIMO systems is considered, with the important novelty of using a new kind of channel code: concatenated systematic LDGM (Low-Density Generator Matrix) codes. These codes are an alternative to Low-Density Parity Check (LDPC)

and Turbo codes, sharing with them the ability of performing near the theoretical Shannon channel capacity. The main advantage of LDGM codes is their low complexity at transmission. The complexity of the Belief Propagation (BP) algorithm used for their decoding, which is also applied to LDPC codes, is much lower than that corresponding to the decoding of Turbo codes.

Contenido

. Lista de figuras	XXI
1. Introducción	1
1.1. Perspectiva histórica de la codificación de canal	3
1.1.1. Códigos LDPC	4
1.2. Perspectiva histórica de las comunicaciones por radio	8
1.2.1. Técnicas MIMO	9
1.3. Estructura de la tesis	12
2. Modelado de la propagación radio	13
2.1. Modelos de propagación a gran escala	14
2.1.1. Modelo de propagación en espacio libre	14
2.1.2. Principales fenómenos físicos que explican la propagación radio	15
2.1.3. Modelo de propagación de dos rayos	16
2.1.4. Otros modelos de propagación	16
2.2. Propagación multitrayecto a pequeña escala	18
2.2.1. Modelo paso-banda	18
2.2.2. Modelo equivalente banda-base	19
2.3. Caracterización de los canales con multitrayecto	22
2.3.1. Dispersión Doppler (<i>Doppler spread</i>)	24
2.3.2. Tiempo de coherencia (T_c)	24
2.3.3. Dispersión del retardo multitrayecto (<i>Multipath delay spread</i> , T_d)	25
2.3.4. Ancho de banda de coherencia (F_c)	25
2.4. Modelos probabilísticos de canal multitrayecto	26
2.5. Modelo equivalente discreto	28
2.6. Concepto de capacidad de un canal	30
2.7. Capacidad del canal Rayleigh con desvanecimiento plano	34
2.8. Diversidad	37
2.9. Codificación en canales con desvanecimiento	41
2.10. Conclusiones	46

3. Modelado de Sistemas Multiple Input Multiple Output (MIMO)	47
3.1. Modelo de sistema de comunicación MIMO	47
3.2. Capacidad del canal MIMO determinista y plano en frecuencia	50
3.2.1. Canal desconocido para el transmisor	51
3.2.2. Canal conocido por el transmisor	52
3.3. Canal MIMO aleatorio plano en frecuencia	54
3.3.1. Capacidad ergódica	54
3.3.2. Capacidad fuera de servicio (<i>outage</i>)	55
3.3.3. Influencia de la correlación espacial en la capacidad	55
3.4. Canal MIMO selectivo en frecuencia	57
3.4.1. Canal desconocido por el transmisor	57
3.4.2. Canal conocido por el transmisor	58
3.5. Diversidad en sistemas MIMO	59
3.5.1. Canal desconocido por el transmisor	60
3.5.2. Canal conocido por el transmisor	61
3.5.3. Extensión a canales MIMO selectivos en espacio, frecuencia y tiempo	62
3.6. Codificación espacio-temporal	63
3.6.1. Códigos ST <i>trellis</i> (STTC)	65
3.6.2. Códigos ST bloque ortogonales (OSTBC)	67
3.6.3. Multiplexación espacial	68
3.6.4. Ventajas e inconvenientes de los distintos esquemas de codificación ST	71
3.6.5. Codificación espacio-temporal cuando el transmisor conoce el canal	72
3.7. Conclusiones	73
4. Sistemas MIMO de tipo GSM en túneles subterráneos	75
4.1. Modelo de canal	75
4.1.1. Canal con desvanecimiento plano	76
4.1.2. Eliminación de la atenuación longitudinal	77
4.1.3. Distribución estadística de los coeficientes del canal	78
4.1.4. Capacidad del canal MIMO medido experimentalmente	79
4.2. Modelo de señal	82
4.3. Conclusiones	86
5. Técnicas de ecualización sin codificación de canal	87
5.1. Receptores para canales planos en frecuencia	87
5.1.1. Receptor óptimo: criterio Máximo-Verosímil	89
5.1.2. Detector decorrelador (<i>Zero-Forcing</i> , ZF)	89
5.1.3. Detector MMSE	91
5.1.4. Detector de cancelación sucesiva (<i>Successive Cancellation</i> , SC)	92
5.2. Receptores para canales MIMO selectivos en frecuencia	94
5.2.1. Detector óptimo (MV)	95

5.3.	Resultados de simulación	97
5.4.	Conclusiones	97
6.	Receptores para sistemas con codificación de canal	101
6.1.	Modelo del transmisor	101
6.2.	Modelo del receptor	102
6.3.	Receptor basado en detector y decodificador MV con decisiones <i>hard</i>	104
6.4.	Ecualización Turbo	105
6.5.	Resultados de simulación	107
6.6.	Conclusiones	108
7.	Técnicas de ecualización con canal desconocido	109
7.1.	Estimación de canal Máximo-Verosímil	109
7.1.1.	Aplicación a sistemas de tipo GSM con diversidad espacial	111
7.2.	Ecualización <i>Decision Feedback</i> basada en el principio MV	113
7.2.1.	Algoritmo SAGE	114
7.2.2.	Ecualizador DF turbo	116
7.3.	Simulaciones numéricas	117
7.3.1.	Estimación MV-EM para canales MIMO Rayleigh	117
7.3.2.	Estimación MV-EM para canales MIMO GSM en túneles subterráneos	118
7.3.3.	Ecualización DF MV-SAGE	121
7.4.	Conclusiones	122
8.	Códigos LDGM en BICM	125
8.1.	Códigos LDGM para BICM	127
8.1.1.	Eliminación del <i>error floor</i> : códigos LDGM concatenados	128
8.1.2.	<i>Mapping</i> bit a símbolo	128
8.2.	Detección y decodificación iterativas mediante el algoritmo de propagación del conocimiento de Pearl	130
8.2.1.	Nodos bit	132
8.2.2.	Nodos bit de paridad externos	132
8.2.3.	Nodos bit de paridad internos	133
8.2.4.	Nodos símbolo	134
8.2.5.	Consideraciones de implementación	134
8.3.	Simulaciones numéricas	136
8.3.1.	Canal AWGN	136
8.3.2.	Canal Rayleigh	140
8.3.3.	Canal MIMO	141
8.4.	Conclusiones	145

9. Conclusiones y líneas futuras de investigación	147
9.1. Conclusiones	147
9.2. Líneas futuras de investigación	150
9.2.1. Estimación de canal y ecualización <i>Decision Feedback</i>	150
9.2.2. Reducción de la complejidad del <i>demapping</i> mediante detección esférica	150
9.2.3. Reducción de la complejidad del <i>demapping</i> mediante cancelación <i>soft</i>	154
9.2.4. Desglose de los nodos símbolo	155
9.2.5. Transmisión adaptativa utilizando códigos LDGM	156
9.2.6. Códigos LDGM irregulares	156
9.2.7. Aplicación de diagramas EXIT a la corrección del <i>error floor</i>	156
9.2.8. Integración de los módulos del receptor mediante representaciones homogéneas	157
9.2.9. Implementación en tiempo real	158
A. Cálculo del filtro blanqueador en tiempo discreto	159
. Bibliografía	161

Lista de figuras

2.1. Comparación de los canales AWGN y Rayleigh (BPSK) en términos de capacidad y probabilidad de error	36
2.2. Efectos de la diversidad y la codificación sobre la SER en el canal de desvanecimiento Rayleigh	40
3.1. Diagrama de bloques de un sistema de comunicación MIMO	48
3.2. Modelo de sistema MIMO con codificación de canal	64
3.3. Código <i>trellis</i> de cuatro estados para QPSK con $n_t = 2$ antenas transmisoras	66
3.4. Diagrama de bloques del esquema de Multiplexación Espacial de Codificación Horizontal	69
3.5. Diagrama de bloques del esquema de Multiplexación Espacial de Codificación Vertical	70
3.6. Diagrama de bloques del esquema de Multiplexación Espacial de Codificación Diagonal	71
3.7. Transmisión D-BLAST	72
4.1. Respuesta al impulso típica del canal en un túnel subterráneo para distintas distancias emisor-receptor	77
4.2. Eliminación del efecto de atenuación en las medidas experimentales del canal	78
4.3. CDFs de los coeficientes del canal MIMO medido frente a las de un canal Rayleigh	79
4.4. CCDF de la capacidad del canal para distintas formas del túnel (4×4 antenas)	80
4.5. CCDF de la capacidad del canal para distintas formas del túnel (2×2 antenas)	80
4.6. CCDF de la capacidad del canal para distintas formas del túnel (1×1 antenas)	81
4.7. Correlaciones espaciales entre los coeficientes del canal MIMO	82
4.8. Pulsos del formato de modulación GMSK	84
4.9. Aproximación lineal de la GMSK	85
5.1. Diagrama de bloques del ecualizador LSS	90
5.2. Diagrama de bloques del ecualizador SC	93
5.3. Rendimiento del receptor MV	98
5.4. Rendimiento del receptor LSS	98
5.5. Rendimiento del receptor SC	99

6.1. Modelo de transmisor de un sistema MIMO	102
6.2. Diagrama de bloques de un receptor para sistemas MIMO con codificación . . .	103
6.3. Receptor basado en detector y decodificador MV con decisiones <i>hard</i>	104
6.4. Rendimiento del ecualizador Turbo iterativo frente al Viterbi de <i>hard decisions</i> (2×2 antenas)	107
6.5. Rendimiento del ecualizador Turbo iterativo frente al Viterbi de <i>hard decisions</i> (4×4 antenas)	108
7.1. Rendimiento del sistema STC considerado sobre un canal no dispersivo	118
7.2. Rendimiento del sistema STC considerado sobre un canal dispersivo con memoria $m_c = 2$	119
7.3. Error Cuadrático Medio (<i>Mean Square Error</i> ,MSE) para distintas longitudes de la secuencia de entrenamiento	120
7.4. <i>Symbol Error Rate</i> (SER) frente a SNR a la salida del decodificador para distintas longitudes de la secuencia de entrenamiento	120
7.5. Comparación del rendimiento entre MV-EM ($N_t=5$ bits), LS ($N_t=26$ bits) y canal conocido (KC)	121
7.6. Comparación entre el detector MAP óptimo y el ecualizador DF MV-SAGE . . .	122
7.7. Rendimiento del ecualizador DF MV-SAGE en un escenario de ISI severa	123
8.1. Diagrama de bloques del transmisor considerado	128
8.2. Red bayesiana que representa el esquema concatenado de codificación y modulación. $\{U_k\}$ representa los bits de información, $\{C_k^o\}$ los bits externos de paridad, $\{C_k^i\}$ los bits internos de paridad y $\{S_k\}$ los símbolos transmitidos a través del canal.	131
8.3. Cadena de Markov sobre la que se realiza el cálculo de $r_0^{w_j}$	136
8.4. Rendimiento del esquema propuesto de BICM 8PSK basado en códigos LDGM regulares de tasa $R_c = 2/3$ y distintos grados de los nodos bit $X = 3, 4, 5, 6$. En todos los casos la longitud de bloque es $K = 20000$ bits de información. . . .	137
8.5. Rendimiento del esquema propuesto de BICM 8PSK basado en códigos LDGM regulares de tasa $R_c = 2/3$ y distintas longitudes de bloque, $K =$ 10000, 20000, 30000 bits de información. En todos los casos el grado de los nodos bit sistemáticos es $X = 4$	138
8.6. Rendimiento del esquema BICM 8PSK propuesto utilizando códigos LDGM irregulares de tasa $R_c = 2/3$ y distintos grados de nodos bit. La longitud de bloque es $K = 30000$ bits de información.	139

- 8.7. Rendimiento del esquema BICM 8PSK propuesto utilizando: i) códigos LDGM concatenados [(150, X ,30000),(14850,4,30150)] (primera aproximación) con distintos grados de nodos bit para el código externo ($X = 3, 4, 5$) y ii) códigos LDGM concatenados [(150,3,30000),(15000,4,30150)] (segunda aproximación). En todos los casos la longitud de bloque es $K = 30000$ bits de información. 140
- 8.8. Comparación del rendimiento entre BICM utilizando: i) códigos convolucionales con *mappings* Set Partitioning (SP) y Semi-Set Partitioning (SSP), ii) códigos LDPC regulares e irregulares, iii) códigos LDGM concatenados, y iv) Turbo-TCM. La longitud de bloque es $K = 20000$ bits de información para los casos de códigos LDPC y Turbo-TCM, $K = 30000$ bits de información para el código LDGM concatenado, y $K = 4000$ bits de información para el código convolucional. 141
- 8.9. Rendimiento del esquema BICM propuesto para distintos métodos de señalización sobre desvanecimiento Rayleigh. 142
- 8.10. Capacidad y capacidad para entrada restringida de un canal MIMO 2×2 para QPSK, 8PSK y 16QAM 143
- 8.11. Bit Error Rate de BICM basada en códigos LDGM sobre un canal MIMO 2×2 144
- 8.12. Bit Error Rate de BICM basada en códigos LDGM y QPSK para distintos canales MIMO ($2 \times 2, 3 \times 3$ y 4×4). 145

Capítulo 1

Introducción

I was confident I was correct, not only in an intuitive way but in a rigorous way. I knew exactly what I was doing, and it all came out exactly right

—Claude Shannon

An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question

—John Tukey

Uno de los retos más importantes a los que deben enfrentarse los ingenieros de diseño de sistemas de comunicaciones sin hilos es la consecución de velocidades de transferencia cada vez más elevadas. Los actuales sistemas de comunicación inalámbricos de Tercera Generación (3G) [1, 2] deben soportar distintos servicios con tasas de transferencia de 144 Kb/s en redes móviles y hasta 2 Mb/s en redes inalámbricas para entornos de propagación en interiores (*indoor*) o en aplicaciones picocelulares. Las tasas de transferencia que se exigirán a los futuros sistemas, ya denominados como Cuarta Generación (4G), serán previsiblemente mucho más elevadas, en el rango de 20-200 Mb/s [3]. La necesidad de requerir tales tasas de transferencia es debida a que, por una parte, se espera que a nivel mundial el número de abonados a redes de comunicación inalámbricas alcance los mil millones en 2005 y los 1.800 millones en 2010 y, por otra, que esos abonados desearán poder disfrutar de modo interactivo de contenidos de Internet con un alto volumen de información: voz, música y vídeo de alta fidelidad.

En el caso de los sistemas 3G, y de modo mucho más acusado en los 4G, satisfacer esos requerimientos sólo puede hacerse mediante un buen aprovechamiento de los recursos de tiempo y ancho de banda. La potencia transmitida también debe minimizarse, especialmente en el caso de terminales móviles. Las técnicas WCDMA (*Wideband Code Division Multiple Access*) [4] y OFDM (*Orthogonal Frequency Division Multiplexing*) [5] han supuesto un

importante hito en la consecución de las tasas de transferencia demandadas por los sistemas 3G. Estas técnicas también están teniendo una relevancia fuera de toda duda en el camino hacia el cumplimiento de los requerimientos 4G. Aunque existan otras razones que justifiquen su empleo, podemos decir que WCDMA es una técnica efectiva para la multiplexación de varios usuarios simultáneos en una red inalámbrica, i.e., para combatir la Interferencia de Acceso Múltiple, (*MultiAccess Interference*, MAI). Sin duda debemos reconocer aquí el papel fundamental que han tenido los trabajos de S. Verdú en el nacimiento y desarrollo de las técnicas multiusuario [6]. Por otra parte, OFDM es una solución excelente para combatir la Interferencia Inter-Simbólica (*InterSymbol Interference*, ISI) introducida por el canal radio y que permite explotar convenientemente la diversidad espectral ofrecida por el canal.

Existe un consenso por parte de la comunidad internacional en afirmar que la vía para afrontar los retos planteados por la 4G debe pasar por la implantación de tecnologías que utilicen múltiples antenas tanto en transmisión como en recepción. La ganancia en rendimiento obtenida por el empleo de varias antenas en recepción se conoce desde hace mucho tiempo [7], pero el hecho de disponer de múltiples elementos tanto en transmisión como en recepción supone no sólo un aumento del rendimiento sino de la eficiencia espectral. El motivo de ello es que el enlace radio puede ser entonces visto como un canal de múltiples entradas y salidas (*Multiple-Input Multiple-Output*, MIMO) cuya capacidad teórica se ha demostrado que puede llegar a crecer de modo proporcional al número de antenas utilizadas [8]. En esencia, se trata de hacer uso de una nueva dimensión, la espacial, sobre la que transportar información. Disponer de una nueva dimensión en el canal de comunicación abre efectivamente la puerta a obtener tasas de transferencia mayores que las actualmente disponibles. Por otra parte, el coste por uso de esta nueva dimensión espacial es inferior al de ancho de banda y tiempo.

El trabajo de C. Shannon [9] supuso en 1948 el pistoletazo de salida a una carrera en pos de la obtención de sistemas que alcanzasen la máxima tasa soportable por el canal de ruido aditivo blanco gaussiano (*Additive White Gaussian Noise*, AWGN). De igual modo, desde que en 1995 I. E. Telatar descubre la capacidad teórica del canal MIMO [8], se ha realizado un gran esfuerzo para desarrollar técnicas que permitan aprovecharla convenientemente. Mientras que en el caso del canal AWGN se necesitaron cuarenta y cinco años para encontrar una técnica que rindiese próxima a la capacidad del canal [10], la investigación de técnicas MIMO se ha apoyado precisamente en todos los avances realizados en el contexto del canal AWGN [11, 12, 13, 14]. Esto ha dado lugar a que a día de hoy se disponga de métodos prácticos para alcanzar la capacidad del canal MIMO [15]. El trabajo de esta tesis ha ido precisamente encaminado en esa línea de investigación [16, 17]. Sin embargo, quedan todavía numerosos aspectos por resolver, en especial la reducción del alto coste computacional asociado a las técnicas MIMO propuestas hasta la fecha. También es de crucial interés extender los resultados disponibles a entornos de propagación más complejos, a sistemas con mayor número de antenas y, sobre todo, la construcción de prototipos y la realización de pruebas de campo de los métodos ya disponibles a nivel teórico [18, 19, 20, 21, 22]. Este es uno de los objetivos de la presente Tesis Doctoral, en la que se estudia cómo la utilización de múltiples antenas en transmisión y recepción puede ser

una vía para aumentar la eficiencia espectral de los sistemas de comunicación que operan en un entorno de propagación concreto: los túneles subterráneos existentes en una red de transporte metropolitano.

Éste fue el objeto del proyecto de la Unión Europea conocido bajo el nombre de ESCORT (*Enhanced diversity and Space-time Coding for underground metro and Railway Transmission*) [23]. El proyecto surgió fruto de la necesidad de que los operadores de transporte urbano europeos pudiesen disponer de sistemas de comunicación con alta tasa de transferencia para soportar múltiples aplicaciones (voz, vídeo, señales de control, ...). Apoyándose en el hecho de que éstos han adoptado la interfaz radio de GSM [24, 25] como estándar (bajo el nombre de GSM-R) para sus sistemas de comunicación, se abordó la tarea de estudiar si era posible aumentar la eficiencia espectral utilizando GSM-R dentro de túneles. Por una parte, se realizaron medidas experimentales de la respuesta al impulso del canal dentro de los túneles de la red del Metro de París para, posteriormente, analizar el rendimiento que podía obtenerse utilizando técnicas espacio-temporales en ese entorno concreto. La presente Tesis Doctoral aborda este aspecto en los Capítulos 4, 5 y 6.

1.1. Perspectiva histórica de la codificación de canal

Repasemos brevemente los principales hitos históricos del camino hacia el aprovechamiento de la capacidad del canal AWGN. El *teorema de codificación de canal* de Shannon (1948) [9] establecía cuál era la máxima capacidad de transferencia de información libre de errores para una relación señal a ruido (*Signal-to-Noise Ratio*, SNR) dada. Hasta entonces, los ingenieros de comunicaciones pensaban que la probabilidad de error del receptor sólo podía disminuirse aumentando la SNR y/o el ancho de banda. En este sentido, el teorema supuso una aportación extraordinariamente positiva al mundo de las comunicaciones: para una SNR y ancho de banda acotados, era posible transmitir sin errores a tasas de información mayores que cero y hasta el límite marcado por la capacidad del canal. Sin embargo, el denominado *recíproco* del teorema también establecía un límite en la tasa de información soportada por el canal. Considerando que estamos dispuestos a admitir una probabilidad de error no nula pero pequeña, la máxima tasa que podemos utilizar para transmitir datos por el canal es prácticamente igual a la capacidad (libre de errores) del canal.

Shannon probaba su teorema demostrando que del conjunto de los códigos de canal aleatorios debía existir al menos uno cuya probabilidad de error para una SNR y ancho de banda fijos tendiese a cero al aumentar el tamaño de bloque. La relación exacta entre el tamaño de bloque necesario para alcanzar una determinada probabilidad de error no fue encontrada hasta 1968 por Gallager [26], investigador de gran relevancia en la historia de la codificación, como veremos después. De la prueba de Shannon también se deduce que casi todos los códigos aleatorios cumplen la misma propiedad. Esto implica que la construcción de un código para alcanzar la capacidad del canal es extremadamente sencilla: basta escoger las palabras que lo componen de modo aleatorio entre el conjunto de palabras código posibles.

Sin embargo, esta falta de estructura en el código implica que la decodificación del mismo sea computacionalmente inviable.

La búsqueda de códigos con una estructura que permitiese su decodificación en tiempos viables se prolongó, como hemos dicho, durante cuarenta y cinco años. Los primeros códigos prácticos fueron los de Hamming, de tipo bloque, propuestos en 1950 [27]. En 1955 Elias descubrió los códigos convolucionales [28], para los cuales Wozencraft y Reiffen [29, 30], Fano [31] y Massey [32] propusieron distintos algoritmos de decodificación. No obstante, hubo que esperar a 1967 para que Viterbi [33, 34] sugiriese el algoritmo que lleva su nombre para calcular en tiempo viable la estimación Máximo-Verosímil (MV) de la secuencia transmitida. El algoritmo de Viterbi [34] obtiene la secuencia más plausible dado el conjunto de observaciones disponible, pero no minimiza la probabilidad de error de bit. Es sabido que si los errores que se pueden cometer al realizar una decisión tienen el mismo peso (como ocurre en el contexto de las comunicaciones al decidir qué bit fue transmitido) el criterio que minimiza la probabilidad de error es el *Maximum A Posteriori* (MAP). No fue hasta 1974 cuando Bahl, Cocke, Jelinek y Raviv [35] particularizaron el algoritmo de Baum-Welch [36], surgido en los años 60, para realizar de modo eficiente la decodificación MAP de cada bit de una secuencia codificada convolucionalmente. Esta particularización se conoce desde entonces como algoritmo BCJR, en honor a ellos, o como algoritmo *Forward-Backward*. Aunque el algoritmo BCJR es propiamente el que minimiza la probabilidad de error de bit, el hecho de no proporcionar, a pesar de su mayor complejidad, grandes ganancias en la práctica con respecto al algoritmo de Viterbi en términos de tasa de error de bit (*Bit-Error Rate*, BER) hizo que no mereciese especial atención hasta 1993. Fue entonces cuando Berrou *et al.* [10, 37] lo consideraron como el módulo esencial del decodificador iterativo para los códigos Turbo por ellos propuestos. Los códigos Turbo recuperaban la idea inicial de codificación aleatoria de Shannon. En esencia, un código Turbo no es más que dos códigos concatenados mediante un entrelazador que dispersa los bits de modo aleatorio [37]. Las palabras código resultado pueden ser muy largas (no hay más que aumentar la profundidad del entrelazado) y tienen estructura pseudoaleatoria. La gran aportación de Berrou *et al.* no fue tanto el diseño del código como el proponer un decodificador de tipo iterativo cuya salida se aproxima mucho a la del decodificador MAP óptimo.

1.1.1. Códigos LDPC

Curiosamente, el tiempo para encontrar una solución práctica al problema de codificación en el canal AWGN pudo haber sido mucho menor. En 1962 Gallager [38, 39] propuso los códigos *Low-Density Parity Check* (LDPC). Siguiendo la idea de código aleatorio de Shannon, Gallager sugería la utilización de códigos bloque lineales cuya matriz de control de paridad fuese generada aleatoriamente. Imponiendo además que esta matriz fuese de estructura dispersa (*sparse*) se posibilitaba la decodificación mediante un algoritmo subóptimo de complejidad razonable. Al igual que dijimos al referirnos a los códigos Turbo, la gran contribución de Gallager no fue sólo la idea de código LDPC sino el algoritmo que permitía su decodificación. ¿Por qué los códigos LDPC, propuestos tan sólo catorce años después del trabajo de Shannon,

no fueron considerados como una solución al problema?. La razón es tan simple como sorprendente: aunque la matriz de control de paridad fuese dispersa, era necesario calcular la matriz generadora del código para poder obtener en transmisión las palabras código a transmitir. Esta matriz no tenía porqué ser necesariamente dispersa, por lo que se requerían N^2 posiciones de almacenamiento en el transmisor, siendo N la longitud de las palabras código. Esta longitud ha estar en torno a 10000 bits para obtener resultados satisfactorios [40]. En aquel entonces, no era factible disponer de tal cantidad de memoria en el transmisor. Además las curvas de probabilidad de error obtenidas por Gallager [39] a través de simulaciones por computador sólo consideraban un tamaño de bloque alrededor de 1000 bits. El rendimiento de los códigos LDPC para esa longitud de bloque era similar al de los códigos concatenados (Reed-Solomon y convolucionales). A modo de curiosidad, decir que el descubridor de los códigos convolucionales, Elias, fue precisamente el director de la tesis de Gallager sobre los códigos LDPC. Además Gallager agradece explícitamente en el prólogo a los revisores de su tesis, Fano y Wozencraft que, como vimos antes, realizaron contribuciones fundamentales en el campo de los códigos convolucionales. Los requerimientos de memoria de los códigos concatenados eran mucho más asumibles para el estado del desarrollo de la tecnología en aquel entonces. Por ello fueron éstos últimos los que merecieron más atención por parte de la comunidad científica. El propio Gallager, en una comunicación personal a David J. C. MacKay [40], reconoce que en aquel momento se asumía que los códigos concatenados eran superiores desde un punto de práctico a los códigos por él propuestos.

Fue precisamente MacKay, junto a Neal, quien redescubrió los códigos LDPC en los años 90 [41, 42, 40, 43]. En 1995 proponían los denominados códigos MN (MacKay-Neal) [41] cuya estructura se asemejaba mucho a la de los LDPC aunque, según MacKay [40], en aquel momento no eran conscientes de ello. Pronto se dieron cuenta de la existencia del trabajo pionero de Gallager y le hicieron justicia histórica en [42]. Además propusieron como método de decodificación la particularización del algoritmo *Belief Propagation* (BP) de Pearl [44] para hallar la palabra código transmitida de acuerdo al criterio MAP. El algoritmo BP nació en el campo de la Inteligencia Artificial (IA) como una técnica de cálculo de las probabilidades *a posteriori* de las variables aleatorias que forman una red bayesiana (*Bayesian Network*, BN). Una red bayesiana es un formalismo para representar conocimiento que consiste en un grafo cuyos nodos son variables aleatorias y sus conexiones indican dependencias estadísticas entre las variables involucradas. El algoritmo de Pearl obtiene la solución exacta cuando la red carece de ciclos pero, si no es así, los resultados que arroja no gozan de suficiente precisión. Esta razón hizo que su aplicabilidad en el campo de la IA fuese limitada. La red bayesiana asociada a un código LDPC (y, en general, a cualquier código) presenta ciclos. Sin embargo, en el problema de la decodificación el valor *exacto* de las probabilidades tiene poca importancia ya que su sentido es proporcionar una medida para tomar una decisión acerca de qué bit fue transmitido. El rendimiento del algoritmo BP como decodificador resultó excelente [42, 40]. Aparte del redescubrimiento del trabajo de Gallager, la gran contribución de MacKay y Neal fue la aplicación del algoritmo BP, que incluía el algoritmo de decodificación sugerido por

Gallager como caso particular. El posterior trabajo de MacKay pasará a la historia por la gran profundidad con que estudió los códigos de Gallager, dando lugar a una nueva manera de ver la codificación de canal basada en la representación por medio de grafos. Aunque él los estudió con la misma profundidad, los códigos MN han sido totalmente olvidados en favor de los códigos LDPC de Gallager. Los códigos LDPC han sufrido un proceso de mejora desde su formulación original. El aspecto más importante en cuanto a mejora del rendimiento ha sido la consideración de códigos LDPC irregulares [45], que son aquellos en los que los nodos del grafo asociado tienen distintos grados, siendo el *grado* el número de nodos al que está conectado.

Los dos esquemas de codificación que tratan de imitar la idea de Shannon de código aleatorio, los códigos Turbo y LDPC, ponían fin al proceso de búsqueda de técnicas que permitiesen alcanzar la capacidad del canal. Ambos esquemas fueron propuestos con tan sólo dos años de diferencia (1993 los códigos Turbo y 1995 los LDPC). Es sorprendente que MacKay diga en [40] que sólo oyó hablar de los códigos Turbo en la etapa final de su trabajo sobre los códigos LDPC. No obstante, ambos esquemas presentan notables diferencias. En cuanto a su estructura, los códigos LDPC encarnan más fielmente la idea de Shannon de codificación aleatoria, mientras que los códigos Turbo se apoyan en los códigos convolucionales como estructura básica. El hecho de que los códigos Turbo no sean tan *aleatorios* puede explicar que su rendimiento sea superior al de los LDPC, cuando la longitud de bloque es pequeña, mientras que, cuando la longitud de bloque es suficientemente grande, los LDPC (irregulares) obtienen mejores resultados. Otro aspecto importante es que los códigos Turbo presentan un “efecto suelo” (*error-floor*) debido a sus malas propiedades de distancia [46]. No obstante, como este efecto suelo suele situarse en probabilidades de error muy bajas (por debajo de 10^{-5}) su utilización no presenta habitualmente inconvenientes desde el punto de vista práctico. Por el contrario, los códigos LDPC tienen muy buenas propiedades de distancia [47] y, consecuentemente, no exhiben este efecto suelo. Lo que acabamos de decir tiene gran importancia desde un punto de vista conceptual [47] ya que, hasta la llegada de los códigos Turbo, era comúnmente aceptado que el criterio para obtener un buen esquema de codificación era precisamente maximizar la distancia entre las palabras código. Sencillamente, no se contemplaba la posibilidad de que existiese un código con malas propiedades de distancia pero con una caída tan fuerte como presentan los códigos Turbo en valores de SNR próximos a la capacidad del canal.

Relacionado con esto último tenemos el hecho de que hasta la aparición de los códigos Turbo en 1993 se había instalado entre la comunidad investigadora un concepto completamente erróneo, consistente en creer que la tasa de corte (*cutoff rate*) del canal, que siempre es menor que la capacidad de Shannon, era la mayor tasa de información soportable por el canal desde un punto de vista práctico. La definición de tasa de corte del canal permite comparar el comportamiento de canales con valores diferentes de este parámetro con respecto a una técnica de codificación concreta. Desde este punto de vista, es un parámetro útil, aunque al ser su obtención muy compleja incluso para modelos de canal sencillos como el AWGN, su aplicabilidad es limitada [47]. El problema viene dado por el hecho de que la tasa de corte

también mide la máxima tasa de un código convolucional para el cual la carga computacional media del algoritmo de decodificación secuencial está acotada [48]. Es por esta razón por la cual se creía que la máxima tasa alcanzable de modo práctico venía dada por la tasa de corte del canal. Quizá esta razón también explique porqué se dió tan poca credibilidad a los inventores de los códigos Turbo cuando los presentaron por vez primera en un congreso [10]. Se llegó a decir que las simulaciones presentadas tenían que ser a la fuerza erróneas. Tal vez sólo hasta que Hagenauer [49] presentó de forma más didáctica las ideas de Berrou *et al.* la comunidad científica valoró en su justa medida el trabajo de estos últimos.

Los códigos bloque lineales con matriz generadora dispersa (*Low-Density Generator Matrix*, LDGM) cuyos coeficientes son generados de forma aleatoria también merecieron la atención de MacKay [50]. La razón fundamental es clara, como comentamos anteriormente. Aunque la decodificación de códigos LDPC puede hacerse utilizando directamente la matriz de control de paridad, aprovechándose de su condición dispersa para que el coste computacional sea razonable, no ocurre lo mismo con el proceso de codificación. En transmisión ha de utilizarse la matriz generadora del código que no tiene porqué ser dispersa. En [50] se analiza el rendimiento de los códigos LDGM concluyendo que su utilidad se reduce al caso de códigos de alta tasa de codificación ya que, para tasas medias y bajas, estos códigos presentan un efecto suelo similar al observado para los códigos Turbo. No obstante, el rendimiento observado para códigos de tasa alta es formidable.

El problema de reducir la complejidad en transmisión de los códigos LDPC fue abordado con éxito por Richardson y Urbanke [51] en 2001. Alternativamente, un nuevo esquema de codificación denominado *Repeat-Accumulate* (RA) se ha mostrado efectivo en este sentido [52]. Los códigos RA están formados básicamente por un código repetitivo seguido de un acumulador. Esta idea surgió inicialmente como un esquema conceptual simple para analizar los códigos Turbo. Curiosamente, más tarde se comprobó que su rendimiento era bastante bueno. Otra esquema con complejidad en transmisión reducida son los códigos LDGM sistemáticos concatenados [53] cuyo estudio como códigos de canal en estructuras que posibilitan altas eficiencias espectrales es uno de los objetivos de esta tesis. Los códigos LDGM sistemáticos concatenados se basan en corregir el efecto suelo asociado a los códigos LDGM originales mediante un código externo. Este código es, a su vez, un código LDGM sistemático pero de tasa muy alta cuyas prestaciones son, como hemos comentado, excelentes. Esto es posible debido a que el efecto suelo de los códigos LDGM es de distinta naturaleza al de los códigos Turbo y está motivado por la forma concreta en que el algoritmo BP realiza la decodificación [53].

La codificación se centró en el dominio binario hasta los años ochenta y, como hemos dicho, se basaba en obtener códigos con buenas propiedades de distancia. En este caso, la distancia entre dos palabras código en el sentido euclídeo es equivalente a su distancia en el sentido de Hamming. Cuando se requirieron técnicas de señalización espectralmente eficientes, se optó por utilizar un código binario seguido de una asignación (*mapping*) bit a símbolo. No fue hasta 1982 cuando Ungerboeck llamó la atención con respecto a este punto, insistiendo en que si la modulación utilizada no es binaria, la construcción del código ha de hacerse teniendo en

cuenta la distancia euclídea entre las palabras código, ya que un código binario optimizado para el caso binario (distancia de Hamming) puede no serlo en el caso no binario. Sugirió además un esquema de codificación denominado *Trellis Coded Modulation* (TCM) [54, 55] que puede ser visto como la generalización de los códigos convolucionales al caso no binario. Las ganancias obtenidas por este esquema frente al tradicional basado en código binario seguido de asignación bit a símbolo fueron significativas.

Ungerboeck también recalcó otra cuestión. La idea es que cuando los valores que puede tomar la entrada al canal están fijados por la técnica de modulación utilizada, hay un punto en el que la curva de la capacidad empieza a alejarse de la correspondiente a la capacidad del canal cuando se considera que la entrada al canal puede tomar cualquier valor continuo. Además, para modulaciones de alto orden hay una diferencia constante con respecto a la de la capacidad con entrada continua. Esto sugería dos cosas: por una parte, que para una eficiencia espectral dada siempre es conveniente utilizar una modulación con el doble de puntos a la que proporciona exactamente esa eficiencia espectral; la redundancia es utilizada precisamente para obtener una ganancia por codificación; al mismo tiempo, aumentar el tamaño de la constelación más allá de ese punto no posibilita obtener ganancias significativas. Por otra parte, la causa de la diferencia constante que se observa con respecto a la capacidad de entrada no restringida para modulaciones de alto orden es la no gaussianidad (que, como sabemos, es la que maximiza la información mutua en el caso continuo) de la distribución de entrada. Utilizando técnicas que intentan que la distribución de entrada se aproxime a una gaussiana, denominadas de “moldeado” (*shaping*) de la constelación, se puede reducir efectivamente esta diferencia [56].

1.2. Perspectiva histórica de las comunicaciones por radio

El canal radio se conoce desde el siglo XIX y su utilización es la base de los sistemas de telegrafía, radar, radio y televisión. En 1861 Maxwell, estando en el *King's College* de Londres, propuso su teoría acerca de la caracterización matemática de las ondas electromagnéticas, cuya existencia probó Hertz en 1887. El primer detector de señales electromagnéticas fue diseñado por Branly y posteriormente refinado por Righi, Lodge y Popov, hasta que en 1895 Marconi (con tan sólo 21 años de edad) construyó el primer radiotelégrafo. El primer uso civil de la tecnología inalámbrica se produjo en 1921 con la implantación de un sistema inalámbrico móvil para la entrega de coches patrulla del Departamento de Policía de Detroit. Una aportación clave fue la Frecuencia Modulada (FM), realizada por Armstrong en 1933, que hizo posible las comunicaciones por radio de alta calidad. El primer sistema de comunicación radio personal fue el *Personal Correspondence System* introducido por los Bell Labs en 1946. La demanda de sistemas inalámbricos fue creciendo, y ello motivó a AT&T para presentar su IMTS (*Improved Mobile Telephone Service*) basado en tecnología FM y que permitía la conexión con la red pública de telefonía. Otro avance importante fue el concepto de red celular, aportado por Ring (de los Bell Labs.) en 1947, que permitía la reutilización del espectro posibilitando un número mayor de usuarios. Desde el punto de vista del procesado de señal, es importante

resaltar el artículo de Bello de 1963 [57] en el que caracteriza matemáticamente los canales lineales variantes en el tiempo que, en esencia, permiten modelar el canal radio. El primer sistema de telefonía analógica de gran capacidad fue propuesto por AT&T en 1970 y se denominó AMPS (*Advanced Mobile Phone Service*). Fue precisamente en los años 70 cuando se lanzaron los primeros satélites de comunicación digitales. El estándar de telefonía móvil digital GSM (*Global System Mobile*) no fue propuesto hasta 1982 y su lanzamiento comercial no se produjo hasta 1991. En 1997 el número de abonados a redes GSM alcanzaba los 50 millones. El éxito de GSM mostró la gran potencialidad de la acción conjunta de los países europeos en la investigación y el desarrollo.

El canal radio se puede modelar conceptualmente como un canal en el que la amplitud recibida varía de modo aleatorio para cada símbolo transmitido. Su capacidad es, por tanto, también aleatoria. Se puede obtener la media estadística de la capacidad cuando la distribución del canal es conocida. En el caso de una distribución Rayleigh, se comprueba que la capacidad media (denominada también ergódica) no se aleja demasiado de la capacidad del canal AWGN. Al igual que hizo Ungerboeck para el canal AWGN, se puede derivar el criterio de optimalidad de un esquema de codificación para el canal con desvanecimiento Rayleigh que resulta curiosamente ser el de mínima distancia de Hamming (en lugar de distancia euclídea) [48]. Como comentamos al principio de esta introducción, la obtención de altas eficiencias espectrales en un enlace radio pasa por la utilización de varios canales radio en paralelo. Una manera práctica de conseguirlo es la utilización de múltiples antenas tanto en transmisión como en recepción.

El hecho de que el criterio de optimización a seguir en la construcción de técnicas de codificación para el canal Rayleigh sea el de la distancia de Hamming motivó que Zehavi [58] propusiese en 1992 la utilización de un código binario seguido de un entrelazador a nivel de bit en sustitución del usual entrelazador a nivel de símbolo que se utilizaba para decorrelar los errores introducidos por el canal. Posteriormente, se realizaba la asignación bit a símbolo. Este esquema se denominó *Bit-Interleaved Coded Modulation* (BICM) [59] y su rendimiento se mostró especialmente bueno cuando se aplicó el principio Turbo a la decodificación [60, 61, 62, 63, 64]. Esto se conoce como BICM con decodificación iterativa (BICM-ID).

1.2.1. Técnicas MIMO

Desde el informe de Telatar sobre la capacidad teórica del canal MIMO en 1995, los hitos de mayor relevancia han sido los que enumeraremos a continuación. En 1998 Tarokh *et al.* [65] derivan el criterio de optimalidad que debe seguirse para diseñar un esquema de codificación en el caso MIMO. El criterio, resumido en los denominados *criterio del rango* y *criterio del determinante*, puede verse como una generalización del criterio de la distancia de Hamming antes comentado para el canal con desvanecimiento Rayleigh en sistemas de una sola antena (*Single-Input Single-Output*, SISO). En [65] se proponen los denominados códigos *trellis* espacio-temporales (*Space-Time Trellis Codes*, STTC) que pueden interpretarse como

una generalización del concepto de TCM al caso MIMO. Sin embargo, en el caso MIMO aparece un factor que no está presente en el caso SISO. La decodificación de los códigos *trellis* tiene una complejidad que crece de modo exponencial no sólo con el orden de la modulación (como en el caso SISO) sino también con el número de antenas utilizadas. Además, su diseño es totalmente dependiente del número de antenas, constelación y *trellis* considerado. No obstante, son una solución elegante al problema de codificación y su rendimiento se muestra aceptable.

El hecho de que la mayor ganancia se obtenga al maximizar el rango de cualquier matriz formada por la diferencia de dos palabras código (que podemos interpretar como su distancia de Hamming) motivó la búsqueda de esquemas de codificación espacio-temporal con un coste computacional asociado mínimo pero que cumpliesen el criterio del rango. Sin duda, el esquema de Alamouti [66], válido sólo para dos antenas transmisoras, puede considerarse como el precursor de todos ellos. Los códigos espacio-temporales bloque ortogonales (*Orthogonal Space-Time Block*) [67, 68] no son más que una generalización del esquema de Alamouti a sistemas MIMO con mayor número de antenas transmisoras. Del mismo modo, el esquema de Lindskog-Paulraj [69] constituye la generalización del esquema de Alamouti a canales selectivos en frecuencia. Una vez obtenida la denominada “ganancia por diversidad” al satisfacer el criterio del rango, basta concatenar un código externo clásico que proporcione la ganancia por codificación (la correspondiente al criterio del determinante) deseada. Ésta ha sido la filosofía empleada en todos los esquemas de codificación espacio-temporal propuestos hasta hace poco. Hay que destacar aquí el papel fundamental que ha tenido OFDM como manera de convertir el canal selectivo en frecuencia en varios canales planos paralelos ya que ha posibilitado la aplicación relativamente directa de los esquemas de codificación espacio-temporal, inicialmente diseñados para el caso no dispersivo, al caso de canales selectivos.

Un aspecto importante a tener en cuenta es que los códigos espacio-temporales (tanto *trellis* como bloque) extraen la mayor diversidad posible del canal a costa de introducir redundancia. Como hemos comentado, uno de los aspectos que suscitan mayor interés por los sistemas MIMO es la posibilidad de conseguir altas eficiencias espectrales en el uso del canal radio. En este sentido, ha sido muy notable la aportación realizada por los Bell Labs con sus esquemas de transmisión *cruda* (esto es, sin codificación) MIMO BLAST (*Bell Labs Layered Space-Time architecture*). En particular, el esquema D-BLAST (*Diagonal-BLAST*) permite aprovechar toda la diversidad proporcionada por el canal MIMO. Al igual que ocurre con los STTCs, la complejidad asociada a la demodulación óptima de los esquemas BLAST crece de modo exponencial en el número de antenas y el tamaño de la constelación. En realidad, la idea de BLAST es permitir una demodulación subóptima basada en la cancelación sucesiva de los flujos de datos ya detectados a costa de una pérdida en el rendimiento medido en términos de SNR requerida para alcanzar un determinado BER.

La decodificación iterativa de los códigos Turbo pronto se generalizó a la ecualización de canales dispersivos [70] y, en general, a esquemas de codificación basados en la concatenación de varios códigos. En el campo de los sistemas MIMO, esta idea también ha sido aplicada con profusión. En [71] se propone la utilización de un STTC seguido de un código de acumulación

independiente en cada antena, estando separados ambos por un entrelazador. El principio Turbo es aplicado en recepción para la decodificación. De un modo general, se plantea en [12] la aplicación del principio Turbo en un receptor MIMO compuesto de una primera etapa de cancelación sucesiva seguida de la decodificación del código de canal utilizado. Los códigos espacio-temporales Turbo fueron propuestos por vez primera en [72, 73]. Asimismo, en [74] se puede ver otra forma de diseñar códigos ST Turbo, además de un buen resumen comparativo de los esquemas Turbo propuestos hasta entonces. El concepto de BICM-ID también ha sido aplicado con éxito a sistemas MIMO [75, 76, 77].

Otro de los objetivos de la presente Tesis Doctoral es precisamente la aplicación del principio Turbo a sistemas de comunicación MIMO. Por una parte, se estudian las ganancias que pueden obtenerse de la ecualización iterativa MIMO con canal conocido en el caso concreto de entornos de propagación en túneles (Capítulo 6) para posteriormente abordar el caso más realista en el que el canal es desconocido en recepción. Esto se realiza bajo dos perspectivas: la estimación del canal y la ecualización *Decision Feedback* (DFE). En ambos casos se aplica el principio MV para la estimación de los parámetros desconocidos (los coeficientes de desvanecimiento introducidos por el canal y los coeficientes de los filtros del DFE). La aplicación directa del criterio MV se traduce en problemas de optimización para los que no existe solución cerrada por lo que se utiliza el algoritmo EM (*Expectation-Maximization*) [78] para calcular dicha solución de modo iterativo. La principal aportación aquí realizada es la incorporación de estas estructuras dentro de un esquema de Turbo igualación, ya que entonces los estimadores se benefician enormemente de las probabilidades *a posteriori* acerca de los símbolos transmitidos calculadas en el proceso de igualación Turbo. Estas ideas se desarrollan en el Capítulo 7.

Como comentamos para el caso del canal AWGN con respecto al criterio de la distancia, también se produjo una pequeña confusión en el caso MIMO con respecto a los conceptos de ganancia por diversidad y por codificación. Los criterios del rango y del determinante fueron derivados utilizando la expresión de la probabilidad de error dos a dos (*pairwise*) media que es una buena aproximación a la verdadera probabilidad de error en el rango de SNRs en que lo es la cota general de la que proviene: la cota de unión (*union bound*). La cota de unión es muy poco ajustada para valores de SNRs próximos a la capacidad del canal, por lo que los criterios de ella derivados son de poca utilidad en el diseño de esquemas que pretendan rendir de modo próxima a dicha capacidad. Esto es lo que motivó la utilización de códigos Turbo directamente conectados a un multiplexor que reparte los símbolos a través de las antenas transmisoras [74], sin ningún tipo de consideración acerca de los criterios del rango y del determinante. La idea subyacente es que el canal MIMO no es sino otro canal más y si se quiere alcanzar su capacidad no hay más que aplicar la idea de codificación aleatoria de Shannon. Ésta es una de las contribuciones realizadas en esta tesis: la consideración de esquemas BICM basados en códigos LDGM sistemáticos concatenados para sistemas MIMO (Capítulo 8).

Sin embargo, estos esquemas siguen teniendo un problema fundamental: la complejidad del proceso de obtención de las verosimilitudes de los bits transmitidos que necesita el algoritmo

de decodificación. Este problema es análogo al comentado anteriormente referido al proceso de demodulación óptimo de BLAST o la decodificación de los STTCs. En este sentido, la detección esférica [79, 80] originalmente diseñada para la demodulación de constelaciones basadas en celosías (*lattices*) [56, 48] se ha revelado como la manera más efectiva de abordar el problema. La detección esférica es en realidad un algoritmo para realizar la detección MV del símbolo transmitido cuyo número esperado de operaciones es pequeño. El empleo de este algoritmo para la obtención de las verosimilitudes de los bits transmitidos necesarias en esquemas iterativos (Turbo) de decodificación fue el objetivo de numerosos trabajos muy recientes entre los que sin duda destaca el de Hochwald y S. ten Brink [15], donde muestran cómo alcanzar la capacidad del canal en sistemas MIMO de hasta ocho antenas en transmisión utilizando un código Turbo clásico, diseñado para el caso SISO binario.

1.3. Estructura de la tesis

Citamos brevemente la estructura seguida en esta Tesis Doctoral. En el Capítulo 2 se presentan los conceptos básicos de la modelización del canal radio, relevantes desde el punto de vista del diseño de técnicas de codificación, señalización y procesado de señal en el receptor. La generalización de estos conceptos al caso de utilizar múltiples antenas tanto en transmisión como en recepción se realiza en el Capítulo 3. Hacemos especial énfasis, además, en los resultados teóricos acerca de la capacidad del canal MIMO. También se presentan las técnicas más importantes propuestas hasta la fecha para aprovechar de modo efectivo la capacidad del canal MIMO. El Capítulo 4 está dedicado a describir las medidas experimentales de la respuesta al impulso del canal existente en túneles subterráneos, realizadas dentro del contexto del proyecto europeo ESCORT. Otra parte importante de este capítulo está formada por la definición de un modelo de señal MIMO compatible con la interfaz radio GSM. Una vez establecido este modelo de señal, en el Capítulo 5 se describen los receptores más importantes a utilizar en este caso, siempre considerando sistemas sin codificación de canal. Su rendimiento es analizado a través de simulaciones por ordenador que utilizan las medidas experimentales del canal. La extensión a sistemas con codificación de canal se aborda, siguiendo el principio Turbo, en el Capítulo 6, observándose ganancias significativas sobre receptores clásicos basados en el algoritmo de Viterbi. El problema de desconocer el canal en recepción es la idea central del Capítulo 7. Como dijimos, se aborda esta cuestión desde dos perspectivas: la estimación de la matriz del canal y la igualación DF. Ambas estructuras resultan de la aplicación del principio MV y encajan a la perfección dentro de un esquema de igualación Turbo. Finalmente, la utilización de códigos que permitan alcanzar la capacidad del canal, como son los LDGM sistemáticos concatenados es estudiada en el Capítulo 8 en el contexto de una estructura BICM-ID. Como muestran los resultados de simulación, la estructura propuesta es un modo efectivo de aprovechar la gran capacidad proporcionada por el canal MIMO. Por último, el Capítulo 9 está dedicado a conclusiones y a describir posibles líneas futuras de investigación.

Capítulo 2

Modelado de la propagación radio

Make everything as simple as possible, but not simpler

—Albert Einstein

El canal utilizado por los sistemas de comunicaciones vía radio presenta numerosas dificultades a la hora de transmitir información a través de él. A diferencia del canal resultante en comunicaciones por cable, que es estacionario y predecible, el canal radio es extremadamente variante con el tiempo y, por tanto, difícil de predecir. Su análisis es complejo, haciendo de la modelización del canal una de las partes más difíciles del diseño de sistemas de comunicación radio. El canal se suele modelar estadísticamente, apoyándose en medidas experimentales realizadas para un sistema de comunicación y entorno de propagación concretos. De todos modos, hay ciertas características comunes a la mayoría de los entornos de propagación radio, que permiten un cierto grado de generalidad en la caracterización del canal.

Los fenómenos físicos involucrados en la propagación de una onda electromagnética son diversos, pero suelen agruparse en tres clases distintas: *reflexión, difracción y dispersión*. Hay dos tipos de modelos de propagación. El primero incluye aquellos modelos que intentan predecir la potencia media recibida en función de la distancia al transmisor. Son útiles para poder estimar el área de cobertura radio de un transmisor, y se denominan modelos *a gran escala*. Las distancias al transmisor consideradas son del orden de centenares o miles de metros. El segundo tipo de modelos de propagación se denomina *a pequeña escala*. Son modelos cuyo objetivo es predecir la fluctuación de la señal recibida en puntos próximos (i.e., a una separación de unas pocas longitudes de onda de la señal transmitida) a una localización determinada. Al ser la separación de los puntos considerados muy pequeña con respecto a la distancia a este último, la atenuación de la señal recibida causada por la distancia al transmisor es común a todos los puntos considerados.

2.1. Modelos de propagación a gran escala

Comenzaremos describiendo uno de los modelos más simples que se consideran en la práctica: transmisor y receptor están separados a una cierta distancia y no existe ningún objeto que interfiera en la propagación de la señal. Este modelo se conoce como *modelo de propagación en espacio libre*. A continuación describiremos los distintos fenómenos físicos que explican la propagación de una onda electromagnética cuando se considera la existencia de objetos entre transmisor y receptor. Ello nos permitirá considerar modelos más complejos que el de propagación en espacio libre, como es el *modelo de reflexión en el suelo*, también denominado *de dos rayos*, y el *modelo log-normal*.

2.1.1. Modelo de propagación en espacio libre

Este modelo resulta de la aplicación directa de las leyes de Maxwell. Puede demostrarse que si por un conductor lineal circula una corriente sinusoidal $\cos 2\pi ft$ y la longitud del mismo es mucho menor que la longitud de onda de dicha sinusoide, $\lambda = c/f$, donde $c = 3 \cdot 10^8$ m/s es la velocidad de la luz en el vacío, entonces los campos eléctrico y magnético en los puntos del espacio situados a una distancia d del conductor pueden expresarse como la suma de tres componentes denominadas *radiante*, *inductiva* y *electrostática* [81, 82]. Dichas componentes son inversamente proporcionales a d , d^2 y d^3 , respectivamente. Para puntos en la región de campo lejano (*far-field*) del transmisor las componentes inductiva y electrostática pueden despreciarse en comparación con la componente radiante, de modo que la potencia recibida en función de la distancia, $P_r(d)$, viene dada por la *ecuación de Friis en espacio libre*

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \quad (2.1)$$

donde P_t denota la potencia transmitida, G_t la ganancia de la antena transmisora, G_r la ganancia de la antena receptora y L mide las pérdidas del sistema no relacionadas con la propagación. Vemos por tanto que la potencia recibida decrece de modo proporcional al cuadrado de la distancia al emisor. Al no ser válida la Ecuación (2.1) en $d = 0$, resulta más conveniente expresar $P_r(d)$ en referencia a la potencia recibida en un punto situado a una distancia pequeña, d_0 , del transmisor (potencia que puede medirse experimentalmente)

$$P_r(d) = P_r(d_0) \left(\frac{d_0}{d} \right)^2 \quad (2.2)$$

o, expresándolo en decibelios,

$$P_r(d) \text{ (dB)} = P_r(d_0) \text{ (dB)} + 20 \log_{10} \left(\frac{d_0}{d} \right) \quad (2.3)$$

Este modelo supone una situación completamente ideal (no hay ningún objeto entre transmisor y receptor) pero tiene la ventaja de ser extremadamente sencillo. Además, hay

sistemas de comunicación en los que el canal puede aproximarse bien por este modelo, como por ejemplo los sistemas de comunicación vía satélite y los enlaces de microondas en los que el receptor está en la línea de visión (*Line of Sight*, LoS) del transmisor.

2.1.2. Principales fenómenos físicos que explican la propagación radio

Para disponer de modelos de propagación radio más realistas, se debe contemplar la existencia de distintos objetos situados entre el transmisor y el receptor, y estudiar su influencia en la propagación de la señal transmitida. Los tres fenómenos básicos que explican la propagación radio se denominan *reflexión*, *difracción* y *dispersión*. Estos tres fenómenos son la base tanto de los modelos de propagación a gran escala como de los modelos de propagación a pequeña escala. Los describiremos a continuación de modo breve, pudiéndose encontrar más detalles en [81, 82, 83].

El fenómeno de *reflexión* ocurre cuando una onda electromagnética incide en una superficie lisa de grandes dimensiones con respecto a la longitud de dicha onda. Ejemplos de objetos que provocan el fenómeno de reflexión son la superficie terrestre y los edificios. La energía de la onda electromagnética incidente es en parte transmitida al objeto y en parte reflejada, estando relacionadas dichas energías por el *coeficiente de reflexión de Fresnel*. Este coeficiente depende de las propiedades del material, el ángulo de incidencia de la onda, así como de su polarización y frecuencia.

Imaginemos ahora que el camino directo (en la línea de visión) entre el transmisor y receptor está obstruido por un objeto. Si sólo tenemos en cuenta el fenómeno de reflexión para explicar la propagación radio, la comunicación entre transmisor y receptor no sería posible a menos que hubiese otros objetos que permitiesen derivar la existencia de un camino entre transmisor y receptor resultado de múltiples reflexiones. Sin embargo, la realidad no es así. Aún en el caso de que no existan más objetos que aquel que obstruye el camino directo, la propagación radio es posible debido a que cuando la onda transmitida incide en el objeto, se generan ondas secundarias que rodean al objeto y que, finalmente, alcanzan el receptor. Este fenómeno se denomina *difracción*. A frecuencias altas, la energía total difractada, al igual que ocurría con la reflexión, depende de la geometría del objeto, así como de la amplitud, fase y polarización de la onda incidente.

Por último, hay un tercer fenómeno que explica la propagación radio. Cuando una onda electromagnética incide en una superficie rugosa, la energía se refleja en todas las direcciones. Este fenómeno se conoce con el nombre de *dispersión* o *difusión* (del inglés *scattering*).

Estos tres fenómenos son los que contemplan los distintos modelos de propagación, tanto los modelos de propagación a gran escala como los modelos de propagación a pequeña escala. A continuación describiremos el *modelo de propagación de dos rayos*, por su sencillez y amplia utilización en la práctica.

2.1.3. Modelo de propagación de dos rayos

Este modelo se basa en considerar que el único objeto entre transmisor y receptor que interviene en la propagación radio es la superficie terrestre. Por tanto, la señal recibida es la suma de dos componentes: una correspondiente a la componente de línea de visión y otra resultado de la reflexión de la onda transmitida por parte de la superficie terrestre. El resultado de considerar sólo esas dos componentes y aplicar las ecuaciones de Maxwell, permite expresar la potencia recibida a una distancia d del transmisor, $P_r(d)$ como

$$P_r(d) = P_t G_t G_r \frac{h_t^2 h_r^2}{d^4} \quad (2.4)$$

donde h_t y h_r son las alturas a las que están situados transmisor y receptor, respectivamente. Vemos que en este caso la potencia recibida es inversamente proporcional a d^4 , a diferencia de lo que ocurría en el modelo de propagación en espacio libre, en el que lo es sólo a d^2 . Esto quiere decir que la aplicación del modelo de propagación en espacio libre a sistemas de comunicaciones terrestres sobreestima la potencia recibida.

Este modelo es todavía sencillo pero mucho más ajustado a la realidad que el modelo de propagación en espacio libre. Pese a su sencillez, se ha comprobado en la práctica que este modelo puede predecir de forma razonable la potencia de señal recibida sobre distancias de varios kilómetros en sistemas móviles de comunicaciones radio cuyas estaciones base se sitúan en torres altas (por encima de 50 m), así como en entornos urbanos donde existe LoS entre transmisor y receptor [82].

Nótese que el único fenómeno físico (de los tres explicados anteriormente) que tiene en cuenta este modelo es el de reflexión. Otros modelos tienen en cuenta los fenómenos de difracción y difusión. Tenemos por ejemplo, el *modelo de difracción filo de cuchillo* (“*knife-edge*”), o su versión de múltiples objetos, que consideran sólo el fenómeno de difracción, o el *modelo sección radar* (*Radar Cross Section, RCS*), que considera sólo el fenómeno de difusión. Los detalles de estos modelos pueden encontrarse en [82].

2.1.4. Otros modelos de propagación

En la práctica pueden combinarse distintos modelos de propagación para estimar la potencia recibida. Además, para un entorno de propagación concreto, pueden completarse los valores predecidos por los modelos presentados hasta ahora (denominados *analíticos*), con curvas de regresión obtenidas a partir de medidas experimentales que contemplen otros factores de propagación adicionales. Como es lógico, la validez de los modelos de propagación que combinan técnicas analíticas y medidas experimentales está restringida al entorno concreto, así como al rango de frecuencias considerado, donde se realizaron las medidas experimentales. Su extrapolación a otros entornos o rangos de frecuencias es posible pero debe corroborarse con medidas experimentales adicionales.

A pesar de lo anterior debe decirse que sí es posible encontrar una característica común a la mayoría de los entornos de propagación radio. Todos los modelos de propagación, tanto los

analíticos como los experimentales, concluyen en considerar que la potencia recibida decrece de modo logarítmico con la distancia, tanto en exteriores (*outdoor*) como en interiores (*indoor*). Por tanto, en media, la pérdida debida a la propagación (*path loss*), definida como el cociente entre potencia radiada y potencia recibida, puede expresarse en decibelios como

$$\overline{PL} \text{ (dB)} = \overline{PL}(d_0) + 10n \log_{10} \left(\frac{d}{d_0} \right) \quad (2.5)$$

donde n es el exponente de la pérdida debida a la propagación, d_0 es un punto de referencia cercano al transmisor y d es la distancia entre el transmisor y el receptor. Por ejemplo, n es igual a dos y cuatro en los modelos de propagación en espacio libre y de dos rayos, respectivamente.

Sin embargo, la ecuación anterior no tiene en cuenta que el entorno de propagación correspondiente a dos puntos distintos situados a la misma distancia d del transmisor puede ser radicalmente distinto. En general, se ha comprobado también en la práctica que las variaciones de la potencia real recibida con respecto a la potencia media predecida por la ecuación anterior pueden considerarse aleatorias y con una distribución log-normal (es decir, normal si las variables las medimos en decibelios). Por tanto la pérdida *real* debida a la propagación puede expresarse como

$$PL(d) \text{ (dB)} = \overline{PL}(d_0) + 10n \log \left(\frac{d}{d_0} \right) + X_\sigma \quad (2.6)$$

donde X_σ es una variable aleatoria gaussiana de media cero y varianza σ^2 . Esta variable aleatoria describe los efectos aleatorios de apantallamiento (*shadowing*) que ocurren en distintos puntos situados a la misma distancia del transmisor. Por ello este fenómeno también se conoce como *apantallamiento log-normal*. En la práctica, el valor de n y σ puede calcularse mediante regresión lineal a partir de datos experimentales. Un valor típico de σ es 8 dB [84].

Existen modelos más complejos, que combinan las ideas que se han explicado para realizar la predicción de la pérdida debida a la propagación. Una explicación detallada de los mismos está fuera del alcance del presente trabajo. Detalles acerca de estos modelos pueden encontrarse en [81, 82], pero en cualquier caso citaremos los más conocidos. Dentro de los modelos de propagación en exteriores (*outdoor*) cabe destacar los modelos de Longley-Rice, Durkin, Okumura, Hata, Walfisch-Bertoni y COST 231. Los modelos de propagación en interiores (*indoor*) más destacados son: el de *particiones*, el *log-normal* (explicado anteriormente), el *modelo de múltiples puntos de ruptura de Ericsson* y el de *factores de atenuación*.

Por último, no podemos dejar de comentar el hecho de que están surgiendo nuevos métodos para predecir la cobertura radio basados en el uso de modelos de propagación de localización específica (*Site Specific*, SISP) combinados con bases de datos de información geográfica (*Graphical Information Systems*, GIS). Los modelos SISP utilizan el trazado de rayos (*ray tracing*) para modelar de forma determinística el entorno de propagación.

2.2. Propagación multitrayecto a pequeña escala

El objetivo de esta sección es describir la variación de la señal recibida en puntos suficientemente alejados del transmisor y próximos entre sí, de modo que el efecto del desvanecimiento debido a la propagación (predicho por los modelos de propagación a gran escala) puede considerarse común a todos los puntos considerados y por tanto ser despreciado. En este caso el desvanecimiento de la señal recibida no se debe a la propagación, sino a la suma destructiva de distintas réplicas de la señal que llegan al receptor en instantes de tiempo ligeramente diferentes. Cada réplica de la señal recorre un camino diferente para llegar al receptor, siendo el resultado de los fenómenos de reflexión, difracción y difusión de la señal transmitida por parte de los objetos del entorno.

2.2.1. Modelo paso-banda

Consideremos que transmitimos una senoide $\phi(t) = \cos(2\pi ft)$ en un entorno de propagación electromagnético. Asumiremos que tanto el transmisor como el receptor pueden estar en movimiento, así como los objetos del entorno entre ambos. Las ecuaciones de Maxwell nos permiten describir la señal recibida, $x(t)$, como

$$x(t) = \sum_i a_i(f, t) \cos(2\pi f(t - \tau_i(f, t))) \quad (2.7)$$

donde $a_i(f, t)$ y $\tau_i(f, t)$ son la atenuación total y el retardo de propagación en el instante t , respectivamente, entre el transmisor y el receptor a lo largo del trayecto i . La atenuación total es simplemente el producto de los factores de atenuación debidos a los diagramas de radiación de las antenas transmisora y receptora, los factores de atenuación de los distintos objetos del entorno, así como de otros parámetros dependientes de la distancia de dichos objetos al transmisor y al receptor. Haremos la suposición de que tanto los factores de atenuación como los retardos de propagación son independientes de la frecuencia (i.e. $a_i(f, t) \approx a_i(t)$ y $\tau_i(f, t) \approx \tau_i(t)$) para poder invocar el principio de superposición y generalizar la relación de entrada-salida anterior a una entrada arbitraria $s(t)$. Si además incorporamos al modelo la perturbación sobre la señal recibida causada por el ruido térmico en el receptor, que en general es razonable asumir como una componente $w(t)$ de ruido gaussiano aditivo de media cero (AWGN) con densidad espectral de potencia $N_0/2$, i.e., $E[w(\tau)w(\tau + t)] = \frac{N_0}{2}\delta(t)$, tenemos que la señal recibida es

$$\begin{aligned} x(t) &= \sum_i a_i(t)s(t - \tau_i(t)) + w(t) \\ &\triangleq x_s(t) + w(t) \end{aligned} \quad (2.8)$$

Un modo conveniente de representar $x_s(t)$ es mediante un sistema lineal variante en el tiempo (*Linear Time Variant, LTV*)

$$\begin{aligned} x_s(t) &= h(\tau, t) * s(t) \\ &\triangleq \int_{-\infty}^{\infty} h(\tau, t) s(t - \tau) d\tau \end{aligned} \quad (2.9)$$

donde

$$h(\tau, t) = \sum_i a_i(t) \delta(\tau - \tau_i(t)) \quad (2.10)$$

es la respuesta al impulso de dicho sistema. Es interesante resaltar que si tanto el transmisor como el receptor, así como los objetos situados entre los mismos que influyen en la propagación de la señal, son estáticos, la Ecuación (2.9) se reduce a la relación de entrada-salida de un sistema lineal e invariante en el tiempo (*Linear Time Invariant, LTI*)

$$x_s(t) = \int_{-\infty}^{\infty} h(\tau) s(t - \tau) d\tau \triangleq h(t) * s(t) \quad (2.11)$$

Para la respuesta al impulso variante en el tiempo (2.10) podemos definir una respuesta en frecuencia variante en el tiempo

$$H(f, t) = \int_{-\infty}^{\infty} h(\tau, t) e^{-j2\pi f\tau} d\tau = \sum_i a_i(t) e^{-j2\pi f\tau_i(t)} \quad (2.12)$$

En el caso particular de un canal invariante con el tiempo, la expresión anterior se reduce a la respuesta en frecuencia usual. Una manera de interpretar $H(f, t)$ es pensar en un sistema LTI que varía lentamente en el tiempo, t , y que tiene una respuesta en frecuencia $H(f, t)$ para cada instante t .

Como resultado de la movilidad, la transmisión de una senoide de frecuencia f sufre el denominado *desplazamiento Doppler* $D = fv/c$, donde $c = 3 \cdot 10^8$ m/s es la velocidad de la luz y v la velocidad relativa entre transmisor y receptor. Una explicación detallada de este fenómeno puede consultarse en [82]. El desplazamiento Doppler depende tanto de la frecuencia f de la señal como de la velocidad del objeto v . Consecuentemente, tenemos que en la Ecuación (2.9), cada componente de frecuencia de la señal transmitida sufre un desplazamiento Doppler a lo largo del i -ésimo trayecto igual a $D_i(f) = fv_i/c$. La velocidad del receptor relativa a la posición del transmisor y al i -ésimo trayecto, v_i , cumple que $d\tau_i(t)/dt = v_i/c$, por lo que $D_i(f) = f d\tau_i(t)/dt$.

2.2.2. Modelo equivalente banda-base

Consideraremos ahora que la señal real transmitida es paso-banda con frecuencia central f_c y ancho de banda F , con $F < 2f_c$. Para una señal de estas características se puede definir el equivalente banda-base como la señal $s_b(t)$ cuya transformada de Fourier es [85]

$$S_b(f) = \begin{cases} \sqrt{2}S(f + f_c) & f > -f_c + F/2 \\ 0 & f \leq -f_c + F/2 \end{cases} \quad (2.13)$$

Al ser $s(t)$ real, su transformada de Fourier es simétrica y, por tanto, $s_b(t)$ contiene exactamente la misma información que $s(t)$. El factor $\sqrt{2}$ es arbitrario y se ha escogido así para que la energía de $s(t)$ y $s_b(t)$ sean la misma. Nótese que $s_b(t)$ está limitada en banda entre $-F/2$ y $F/2$. Es fácil ver que para reconstruir $s(t)$ a partir de $s_b(t)$ basta tener en cuenta que

$$s(t) = \sqrt{2}\Re[s_b(t)e^{j2\pi f_c t}] \quad (2.14)$$

donde $\Re[\cdot]$ representa la parte real de un número complejo. En adelante, a la parte imaginaria de un número complejo la denotaremos por $\Im[\cdot]$. En términos de señales reales, la señal paso-banda $s(t)$ puede obtenerse modulando $\Re[s_b(t)]$ (denominada componente *en-fase*, “I”) por $\cos(2\pi f_c t)/\sqrt{2}$ e $\Im[s_b(t)]$ (denominada componente *de cuadratura*, “Q”) por $-\sin(2\pi f_c t)/\sqrt{2}$ y sumando ambas. Este proceso de trasladar una señal banda-base a una frecuencia determinada se denomina *up-conversion* y es utilizada en la parte del transmisor de un sistema de comunicaciones radio. La señales $\Re[s_b(t)]$ e $\Im[s_b(t)]$ pueden obtenerse modulando $s(t)$ por $\sqrt{2}\cos(2\pi f_c t)$ y $\sqrt{2}\sin(2\pi f_c t)$, respectivamente, y filtrando con un filtro paso-bajo ideal entre $[-F/2, F/2]$. Este proceso se denomina *down-conversion* y es realizado en recepción. Sustituyendo $s(t) = \sqrt{2}\Re[s_b(t)e^{j2\pi f_c t}]$, $w(t) = \sqrt{2}\Re[w_b(t)e^{j2\pi f_c t}]$ y $x(t) = \sqrt{2}\Re[x_b(t)e^{j2\pi f_c t}]$ en la Ecuación (2.8) tenemos que

$$\Re[x_b(t)e^{j2\pi f_c t}] = \sum_i a_i(t)\Re[s_b(t - \tau_i(t))e^{j2\pi f_c(t - \tau_i(t))}] + \Re[w_b(t)e^{j2\pi f_c t}] \quad (2.15)$$

$$= \Re \left[\left\{ \sum_i a_i(t)s_b(t - \tau_i(t))e^{-j2\pi f_c \tau_i(t)} + w_b(t) \right\} e^{j2\pi f_c t} \right] \quad (2.16)$$

De modo análogo, se puede ver que

$$\Im[x_b(t)e^{j2\pi f_c t}] = \Im \left[\left\{ \sum_i a_i(t)s_b(t - \tau_i(t))e^{-j2\pi f_c \tau_i(t)} + w_b(t) \right\} e^{j2\pi f_c t} \right] \quad (2.17)$$

por lo que el canal banda-base equivalente es

$$\begin{aligned} x_b(t) &= \sum_i a_i^b(t)s_b(t - \tau_i(t)) + w_b(t) \\ &\triangleq x_s^b(t) + w_b(t) \end{aligned} \quad (2.18)$$

donde

$$a_i^b(t) \triangleq a_i(t)e^{-j2\pi f_c \tau_i(t)} \quad (2.19)$$

De nuevo, la relación entre $s_b(t)$ y $x_s^b(t)$ puede expresarse mediante un sistema lineal variante en el tiempo con respuesta al impulso

$$h_b(\tau, t) = \sum_i a_i^b(t)\delta(\tau - \tau_i(t)) \quad (2.20)$$

La expresión anterior es fácil de interpretar en el dominio temporal, y el efecto de la frecuencia de portadora f_c puede verse de modo explícito. La señal banda-base recibida (en ausencia de ruido) es la suma, sobre todos los trayectos, de réplicas retardadas de la señal de entrada. La magnitud del i -ésimo sumando es la magnitud de la respuesta a lo largo de dicho trayecto, cuyo valor cambia lentamente, con cambios significativos que ocurren en el orden de segundos (o incluso mayor). Por el contrario, la fase del i -ésimo sumando cambia en un valor de $\pi/2$, esto es, de modo significativo, cuando el retardo correspondiente a dicho trayecto cambia en $1/(4f_c)$, o equivalentemente, cuando la longitud del trayecto cambia en un cuarto de la longitud de onda, es decir, en $c/(4f_c)$. Si la longitud del trayecto cambia a una velocidad v , el tiempo requerido para dicho cambio de fase es $c/(4f_cv)$. Teniendo en cuenta que el desplazamiento Doppler D en la frecuencia f es fv/c , y asumiendo que $f \approx f_c$ en comunicaciones paso-banda, el tiempo requerido para un cambio de fase de $\pi/2$ es $1/(4D)$. Por ejemplo, para un sistema de comunicaciones móviles en la banda de $f_c = 900$ MHz y un terminal moviéndose a una velocidad $v = 60$ Km/h, esto son 5 ms. Por último, nótese que la transformada de Fourier de $h_b(\tau, t)$ para un tiempo t fijo, $H_b(f, t)$, es igual a $H(f - f_c, t)$, esto es, la respuesta en frecuencia del sistema original (en el instante t) desplazada f_c .

Particularización a un sistema de comunicaciones digital

Las ideas presentadas hasta ahora son válidas independientemente del formato de la señal transmitida. Para particularizar el modelo al caso concreto de un sistema de comunicaciones digital, asumiremos que el equivalente paso-bajo de la señal transmitida se corresponde con un formato de modulación lineal, es decir,

$$s_b(t) = \sum_n s[n]p(t - nT_s) \quad (2.21)$$

donde $s[n]$ es la secuencia de símbolos (en el dominio complejo) a transmitir, T_s es el período de símbolo y $p(t)$ es la forma de pulso utilizada en la modulación¹. El equivalente banda-base de la señal recibida, de acuerdo con la Ecuación (2.18), es

$$x_b(t) = \sum_n s[n] \sum_i a_i^b(t)p(t - nT_s - \tau_i(t)) + w_b(t) \quad (2.22)$$

Obsérvese que la expresión anterior asume que en recepción hay sincronismo perfecto con respecto a la frecuencia de portadora f_c y que se utiliza un filtro paso bajo ideal entre $-F/2$ y $F/2$.

¹Nótese que no hay pérdida de generalidad, ya que cualquiera que sea la señal transmitida, siempre que sea limitada en banda con un ancho de banda $F/2$, puede ser expresada como un formato de modulación lineal con forma de pulso igual al pulso de Nyquist, esto es $p(t) = F \text{sinc}(Ft)$, con $\text{sinc}(t) \triangleq \sin(\pi t)/(\pi t)$, período de símbolo $T_s = 1/F$, y símbolos transmitidos igual a las muestras de la señal en los múltiplos de T_s , es decir, $s[n] \triangleq s_b(t = nT_s)$.

2.3. Caracterización de los canales con multitrayecto

Definiremos a continuación los parámetros más importantes utilizados para describir las características de un canal multitrayecto. Para ello es conveniente expresar (2.21) como

$$\begin{aligned} s_b(t) &= \left(\sum_n s[n] \delta(t - nT_s) \right) * p(t) \\ &\triangleq s_i(t) * p(t) \end{aligned} \quad (2.23)$$

y expresar (2.18) como la relación entre la entrada $s_i(t)$ y la salida $x_b(t)$ de un sistema lineal variante en el tiempo

$$x_b(t) = s_i(t) * h_b(\tau, t) \quad (2.24)$$

donde

$$h_b(\tau, t) = \sum_i a_i(t) e^{-j2\pi f_c \tau_i(t)} p(\tau - \tau_i(t)) \quad (2.25)$$

Es importante resaltar que las definiciones que introduciremos a continuación y que nos permitirán describir las características del canal tienen en cuenta no sólo el canal propiamente dicho, sino el efecto del pulso utilizado por la modulación.

En adelante consideraremos que el canal $h_b(\tau, t)$ es un proceso aleatorio bidimensional en tiempo continuo. En general podemos asumir ciertas condiciones sobre el canal. La primera de ellas es suponer que el canal es *estacionario en sentido amplio* (*Wide Sense Stationary*, WSS). Matemáticamente esta condición quiere decir que la función de autocorrelación temporal del canal, en la variable tiempo t , sólo depende de la diferencia entre instantes de tiempo, es decir

$$R_h(\tau, t) = E\{h_b(\tau, t_0) h_b^*(\tau, t_0 + t)\} = E\{h_b(\tau, 0) h_b(\tau, t)\} \quad (2.26)$$

para cualquier retardo τ e instante t_0 arbitrario. Otra suposición que realizaremos es que los desvanecimientos correspondientes a los distintos objetos causantes del multitrayecto son incorrelados. Esta suposición se denomina *difusión incorrelada* (*Uncorrelated Scattering*, US). Matemáticamente esta condición se expresa como

$$E\{h_b(\tau_1, t) h_b^*(\tau_2, t)\} = 0 \quad \text{si } \tau_1 \neq \tau_2 \quad (2.27)$$

para cualquier instante de tiempo t arbitrario. Cuando un canal cumple tanto la condición de estacionariedad en sentido amplio como la de difusión incorrelada se denota por WSSUS (*Wide Sense Stationary Uncorrelated Scattering*).

Definiremos la *función de difusión Doppler-retardo* [84] como la transformada de Fourier de la función de autocorrelación temporal

$$\varphi_{Do, De}(\tau, \nu) = \int_{-\infty}^{\infty} R_h(\tau, t) e^{-j2\pi \nu t} dt \quad (2.28)$$

Esta función representa la potencia media del canal en función de la frecuencia (denominada en este caso *frecuencia Doppler*) y el retardo. En otras palabras, para un retardo τ fijo, $\varphi_{Do,De}(\tau, \nu)$ es la densidad espectral de potencia del proceso aleatorio en la variable t , $h_b(\tau, t)$. En general, aún bajo la suposición de WSSUS, la función de difusión Doppler-retardo es una función arbitraria tanto de τ como de ν . Sin embargo, puede demostrarse que la condición de estacionariedad en sentido amplio en el dominio temporal t implica incorrelación en el dominio de la frecuencia Doppler, es decir,

$$E\{\varphi_{Do,De}(\tau, \nu_1)\varphi_{Do,De}^*(\tau, \nu_2)\} = 0 \quad \text{si } \nu_1 \neq \nu_2 \quad (2.29)$$

para cualquier retardo τ arbitrario.

Podemos agrupar la información contenida en la función de difusión Doppler-retardo simplemente marginalizando con respecto a la variable retardo τ , dando lugar a la *función densidad espectral de potencia Doppler* [84]

$$\varphi_{Do}(\nu) = \int_0^{\tau_{max}} \varphi_{Do,De}(\tau, \nu) d\tau \quad (2.30)$$

que representa la potencia media del canal sólo en función de la frecuencia Doppler. Esta función permite centrarnos exclusivamente en la variación del canal con respecto a la variable independiente t . Considerando que $\varphi_{Do}(\nu)$ toma valores significativos en el rango de frecuencias $\mathcal{F} = [\nu_c - \nu_{max}, \nu_c + \nu_{max}]$, tenemos que cuanto mayor sea ν_{max} , más rápido variará el canal en el tiempo. De igual modo, marginalizando la función de difusión Doppler-retardo con respecto a la variable frecuencia Doppler ν , obtenemos la *función densidad espectral de potencia retardo* [84]

$$\varphi_{De}(\tau) = \int_{\mathcal{F}} \varphi_{Do,De}(\tau, \nu) d\nu \quad (2.31)$$

que representa la potencia media del canal sólo en función del retardo. Si el intervalo en el que $\varphi_{De}(\tau)$ toma valores significativos lo denotamos por $[0, \tau_{max}]$, tenemos que (en media) la máxima dispersión temporal que introduce el canal es τ_{max} . La función densidad de potencia retardo también recibe el nombre de *perfil de intensidad multitrayecto* (*multipath intensity profile*).

Para un instante de tiempo t fijo, $h_b(\tau, t)$ es un proceso aleatorio en la variable retardo τ . Anteriormente dijimos que es razonable suponer incorrelación entre distintos retardos, de modo que no tiene sentido hablar de función de autocorrelación en la variable retardo τ ni, por tanto, de su transformada de Fourier, tal y como hicimos con la variable tiempo t . Sin embargo, para un instante de tiempo t fijo, ya hemos introducido anteriormente la transformada de Fourier de una realización concreta del proceso $h_b(\tau, t)$, es decir

$$\begin{aligned} H_b(f, t) &\triangleq \int_0^{\tau_{max}} h_b(\tau, t) e^{-j2\pi f\tau} d\tau \\ &= \sum_i a_i(t) e^{-j2\pi f\tau_i(t)} P(f) e^{-j2\pi f\tau_i(t)} \end{aligned} \quad (2.32)$$

donde $P(f)$ es la transformada de Fourier de $p(\tau)$. Se puede demostrar que la suposición de difusión incorrelada en el dominio retardo implica estacionariedad en sentido amplio (WSS) en el dominio de la frecuencia f . Por ello la función de autocorrelación de $H_b(f, t)$ tendrá la forma siguiente

$$R_f(f, t) = E\{H_b(f_0, t)H_b^*(f_0 + f, t)\} = E\{H_b(0, t)H_b^*(f, t)\} \quad (2.33)$$

para cualquier instante t y f_0 arbitrario.

Teniendo en cuenta las definiciones que acabamos de exponer, pasamos a continuación a presentar los parámetros más importantes que caracterizan un canal multitrayecto.

2.3.1. Dispersión Doppler (*Doppler spread*)

Este parámetro mide la dispersión que introduce el canal en el dominio de la frecuencia Doppler. Formalmente, se define como la raíz cuadrada del ancho de banda medio de la densidad espectral de potencia Doppler

$$\nu_{RMS} \triangleq \sqrt{E\{(\nu - \bar{\nu})^2\}} \quad (2.34)$$

donde la operación esperanza se realiza con respecto a la densidad espectral de potencia Doppler normalizada dada por

$$\tilde{\varphi}_{D_o}(\nu) \triangleq \frac{\varphi_{D_o}(\nu)}{\int_{\mathcal{F}} \varphi_{D_o}(\nu) d\nu} \quad (2.35)$$

2.3.2. Tiempo de coherencia (T_c)

Se define como el intervalo temporal sobre el cual $h_b(\tau, t)$ no cambia significativamente en función de t . Anteriormente comentamos que los cambios significativos en $a_i(t)$ se producen en el orden de segundos o incluso más. Sin embargo, los cambios significativos (i.e., $\pi/2$) en la fase de (2.25) se producen cada $1/(4D_i)$, donde D_i es el desplazamiento Doppler del trayecto i . En la mayoría de las aplicaciones prácticas el intervalo temporal para cambios de este tipo es del orden de 10 ms [85]. Los cambios debidos a las réplicas desplazadas de $p(\tau)$ debidos a la variación temporal de $\tau_i(t)$ son proporcionales² a F , mientras que los cambios en la fase son proporcionales a la frecuencia de portadora f_c , que es mucho mayor que F . Tenemos, por tanto, que los cambios más rápidos en los coeficientes del canal, para τ fijo, son debidos a cambios en la fase y que éstos son significativos sobre retardos de $1/(4D_i)$. Por tanto, para τ fijo, el tiempo de coherencia del canal es inversamente proporcional al máximo desplazamiento Doppler (sobre todas las componentes multitrayecto que componen la respuesta al impulso del canal)

$$T_c(\tau) \triangleq \frac{1}{D_{max}(\tau)} \quad (2.36)$$

²Al ser $p(\tau)$ una señal banda-base limitada a $[-F/2, F/2]$, los cambios más rápidos en el dominio temporal serán debidos a una senoide de frecuencia $F/2$, i.e., $\cos(2\pi \frac{F}{2} \tau)$

Esta definición es dependiente de la variable retardo τ . Una definición independiente del retardo se obtiene recurriendo a la función densidad espectral de potencia Doppler del canal. En este caso el tiempo de coherencia se define como el inverso de la dispersión Doppler

$$T_c \triangleq \frac{1}{\nu_{RMS}} \quad (2.37)$$

2.3.3. Dispersión del retardo multitrayecto (*Multipath delay spread, T_d*)

Se define como la diferencia entre el retardo de propagación del trayecto más largo y el del más corto

$$T_d(t) \triangleq \max_i \tau_i(t) - \min_i \tau_i(t) \quad (2.38)$$

que es una definición dependiente de la variable tiempo t . De modo análogo a como hicimos con el tiempo de coherencia, podemos definir T_d independientemente de t utilizando en este caso el perfil de intensidad multitrayecto. De este modo, podemos definir la dispersión del retardo multitrayecto como la raíz cuadrada del retardo cuadrático medio, es decir

$$T_d \triangleq \tau_{RMS} = \sqrt{E\{(\tau - \bar{\tau})^2\}} \quad (2.39)$$

donde el operador esperanza actúa con respecto al perfil de intensidad multitrayecto normalizado dado por

$$\tilde{\varphi}_{De}(\tau) \triangleq \frac{\varphi_{De}(\tau)}{\int_0^{\tau_{max}} \varphi_{De}(\tau) d\tau} \quad (2.40)$$

En la práctica, si una celda o red de área local (*Local Area Network, LAN*) tiene una extensión de pocos kilómetros, es muy poco probable tener trayectos cuyas longitudes difieran en más de 300-600 metros, lo que corresponde a una diferencia en los retardos de propagación de 1-2 μs [85]. Si esto lo comparamos con los períodos de símbolo utilizados en los sistemas prácticos (por ejemplo, 3.7 μs en GSM), vemos que todos los retardos de propagación se extienden tan sólo a lo largo de 2-3 períodos de símbolo.

2.3.4. Ancho de banda de coherencia (F_c)

Se define como el rango de frecuencias a lo largo del cual $H_b(f, t)$ no cambia significativamente en función de t . Consideremos primero un instante de tiempo t fijo. Para un trayecto concreto i , la fase de $H(f, t)$ es lineal en f . La contribución de todos los trayectos en (2.32) resultará en una fase dominada por el término correspondiente a la diferencia de fase entre los trayectos de mayor y menor retardo, i.e., $2\pi f(\max_i \tau_i(t) - \min_i \tau_i(t)) = 2\pi f T_d(t)$. Tenemos por tanto, que para un instante t concreto, el ancho de banda $F_c(t)$ para el cual el canal no cambia significativamente es inversamente proporcional a $T_d(t)$

$$F_c(t) \triangleq \frac{1}{T_d(t)} \quad (2.41)$$

De modo más preciso, se define el ancho de banda de coherencia como el valor de f para el cual $R_f(f, t)$ decae hasta 0.7. Al igual que ocurría con la dispersión del retardo multitrayecto, esta definición del ancho de banda de coherencia es dependiente del instante de tiempo t . Sin embargo, puede demostrarse que, en media, este valor es aproximadamente el inverso de la dispersión del retardo multitrayecto

$$F_c \triangleq \frac{1}{\tau_{RMS}} \quad (2.42)$$

Los parámetros anteriores nos permiten, en cierto modo, clasificar y medir canales que en la realidad pueden ser muy distintos entre sí, estableciendo un marco común sobre el que evaluar el rendimiento de distintos sistemas de comunicación. Aunque a primera vista esto parece muy ventajoso, el problema subyacente es que canales con los mismos valores de esos términos pueden ser en realidad muy diferentes. Hay que resaltar, por tanto, que estos parámetros son útiles especialmente porque permiten la comparación de distintos sistemas de comunicación. Sin embargo, hay que poner especial cuidado si de lo que se trata es de predecir el rendimiento de un sistema de comunicación en un entorno concreto. Se puede ir incluso un paso más allá intentando imponer condiciones adicionales al canal para simplificar todavía más el modelo que lo describe. Desarrollaremos esta idea a continuación.

2.4. Modelos probabilísticos de canal multitrayecto

Al englobar toda la distorsión sobre la señal transmitida causada por el desplazamiento Doppler y el retardo multitrayecto en un modelo probabilístico simple es claro que perdemos precisión pero a cambio el modelo pasa a ser computacionalmente tratable. Es conocido que, en general, los canales presentes en la realidad no se ajustan tan bien a los modelos probabilísticos considerados como ocurre al modelar el ruido térmico por AWGN. Sin embargo, de modo similar a lo que comentamos sobre los parámetros que describen un canal, conviene enfatizar su conveniencia como modelos de referencia sobre el que comparar distintos sistemas y/o técnicas.

Un modelo muy utilizado es el modelo de desvanecimiento plano de Clarke [82]. Esencialmente consiste en modelar la envolvente compleja de la señal recibida como la suma de dos variables gaussianas independientes, una en cada dimensión compleja. La explicación que daremos aquí puede considerarse intuitiva, pero puede encontrarse en [82] un razonamiento basado directamente en el campo eléctrico creado por una antena polarizada verticalmente.

Recordemos que cada coeficiente del canal equivalente banda-base tiene la forma

$$h(\tau, t) = \sum_i a_i(t) e^{-j2\pi f_c \tau_i(t)} p(\tau - \tau_i(t)) \quad (2.43)$$

Cada coeficiente del canal $h(\tau, t)$ resulta de la suma de múltiples réplicas retardadas del pulso utilizado en la modulación. El modelo probabilístico más simple consiste en suponer que existe un gran número de difusores estadísticamente independientes con amplitudes aleatorias en el intervalo temporal correspondiente a un período de símbolo. Considerando que los reflectores

están muy alejados (con respecto a la longitud de onda de la portadora, $\lambda_c = c/f_c$), es razonable asumir que la fase de cada trayecto está uniformemente distribuida entre 0 y 2π y que dichas fases son independientes. Por tanto, cada coeficiente $h(\tau, t)$ puede ser convenientemente modelado como una variable aleatoria (v.a.) que es la suma de un gran número de pequeñas variables aleatorias complejas cada una con fase uniformemente distribuida. Por ello, $\Re[h(\tau, t)]$ es la suma de muchas variables aleatorias reales independientes, por lo que puede ser modelada como una v.a. gaussiana de media cero. De modo similar, puede verse que $\Im[h(\tau, t)e^{j\phi}]$ también es gaussiana con la misma varianza. Una variable aleatoria gaussiana X cuyas componentes real e imaginaria son independientes y de idéntica distribución (i.i.d.) cumple la propiedad de simetría circular, i.e., $e^{j\phi}X$ tiene la misma distribución que X para cualquier ϕ . A la variable aleatoria con estas propiedades la llamaremos gaussiana compleja circularmente simétrica y la denotaremos por $\mathcal{CN}(0, \sigma^2)$, donde $\sigma^2 = E[|X|^2]$. Por tanto, $h(\tau, t)$ es $\mathcal{CN}(0, \sigma_\tau^2)$. Asumiremos también que la varianza σ_τ^2 es función de τ , pero independiente del instante de tiempo t . Este modelo de canal también recibe el nombre de *desvanecimiento Rayleigh*, ya que el módulo de $h(\tau, t)$, al ser $h(\tau, t)$ una variable aleatoria $\mathcal{CN}(0, \sigma_\tau^2)$, sigue una distribución Rayleigh cuya función de densidad de probabilidad (f.d.p.) obedece a la expresión

$$f(x) = \frac{x}{\sigma_\tau^2} \exp\left(-\frac{x^2}{2\sigma_\tau^2}\right) \quad (2.44)$$

Además, si consideramos una matriz \mathbf{H} donde cada elemento h_{ij} es una v.a. $\mathcal{CN}(0, \sigma_{ij}^2)$, y siendo $\mathbf{R} \triangleq E\{\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H\}$, la función generadora de momentos de su norma de Frobenius, i.e., $\|\mathbf{H}\|_F^2 \triangleq \text{Tr}\{\mathbf{H}\mathbf{H}^H\} = \sum_i \sum_j |h_{ij}|^2$, viene dada por

$$\varphi(v) \triangleq E\{e^{-v\|\mathbf{H}\|_F^2}\} \quad (2.45)$$

$$= \frac{1}{\det(\mathbf{I} + v\mathbf{R})} \quad (2.46)$$

$$= \prod_i \frac{1}{1 + v\lambda_i(\mathbf{R})} \quad (2.47)$$

donde \mathbf{I} denota la matriz identidad y $\lambda_i(\mathbf{R})$ es el i -ésimo autovalor de \mathbf{R} siendo la región de convergencia asociada $\Re[v] \geq \max_i -1/\lambda_i(\mathbf{R})$.

Un modelo alternativo, también muy utilizado, se deriva de contemplar en el modelo anterior la existencia de un trayecto de propagación de la señal en la LoS cuyo desvanecimiento correspondiente tiene una magnitud significativa y que se asume conocida. A este camino de propagación se le denomina camino especular. En este caso $h(\tau, t)$, al menos para un valor de τ , se modela como una variable aleatoria gaussiana compleja de media igual a la magnitud del desvanecimiento correspondiente a la componente especular. El módulo de $h(\tau, t)$ en este caso sigue una distribución Rice, cuya f.d.p. está dada por

$$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2 + \mu^2}{2\sigma^2}\right) I_0\left(\frac{x\mu}{\sigma^2}\right) \quad (2.48)$$

donde $\mu = E\{x\}$ e $I_0(\cdot)$ es la función de Bessel modificada de primera clase que se define como

$$I_0(x) \triangleq \frac{1}{2\pi} \int_0^{2\pi} e^{-x \cos \theta} d\theta \quad (2.49)$$

Nótese que la distribución Rice se reduce a una Rayleigh cuando $\mu = 0$. Además, cuando la componente especular es mucho más fuerte que la componente multitrayecto, la distribución Rice es aproximadamente una gaussiana de media μ . Basta considerar que

$$I_0(x) \approx \frac{e^x}{\sqrt{2\pi x}} \quad \text{para } x \rightarrow \infty \quad (2.50)$$

y sustituirlo en la función de densidad Rice para obtener

$$f(x) = \frac{\sqrt{x}}{\sqrt{2\pi\nu\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (2.51)$$

En el caso extremo, el desvanecimiento Rice se puede considerar como determinista, por lo que el canal radio se reduce al canal AWGN.

Es interesante considerar que el comportamiento del canal radio posee dos extremos dados por el canal con desvanecimiento Rayleigh y el canal AWGN, modelándose la situación intermedia mediante un canal con desvanecimiento Rice. El motivo de ello es que si el rendimiento de una determinada técnica es bueno en los dos extremos, será una técnica robusta, mientras que técnicas con rendimiento pobre en alguno de estos extremos adolecerán de una importante falta de robustez. Aunque parezca que estamos explicando lo evidente, es un aspecto muy importante a la hora de analizar las distintas técnicas para canales radio.

2.5. Modelo equivalente discreto

A continuación construiremos el equivalente discreto del modelo continuo hasta ahora considerado. Muestreando a período de símbolo la Ecuación (2.22), se obtiene la señal discreta

$$x[k] = \sum_n s[n] \sum_i a_i^b[k] p[k-n-\tau_i(kT_s)/T_s] + w[k] \quad (2.52)$$

donde $x[k] \triangleq x(t = kT_s)$ y $a_i^b[k]$, $p[k]$ y $w[k]$ se definen de igual modo. Nótese que

$$\Re\{w[k]\} = \int_{-\infty}^{\infty} w(t) \varphi_{k,1}(t) dt \quad (2.53)$$

$$\Im\{w[k]\} = \int_{-\infty}^{\infty} w(t) \varphi_{k,2}(t) dt \quad (2.54)$$

donde

$$\varphi_{1,k}(t) = \sqrt{2}F \cos(2\pi f_c t) \text{sinc}(Ft - mT_s) \quad (2.55)$$

$$\varphi_{2,k}(t) = \sqrt{2}F \sin(2\pi f_c t) \text{sinc}(Ft - mT_s) \quad (2.56)$$

Como $\{\varphi_{k,1}(t), \varphi_{k,2}(t)\}_k$ es un conjunto de funciones ortogonales entre sí y las proyecciones de un proceso de ruido gaussiano blanco sobre funciones ortogonales son independientes, el proceso de ruido $\{w[k]\}$ es temporalmente blanco y además sus componentes real e imaginaria son gaussianas de media cero, varianza $N_0/2$ e independientes entre sí. Por tanto $w[k]$ es $\mathcal{CN}(0, N_0)$.

Introduciendo la notación $m \triangleq k - n$ se tiene

$$x[k] = \sum_m s[k - m] \sum_i a_i^b[k] p[m - \tau_i(kT_s)/T_s] + w[k] \quad (2.57)$$

que puede ser expresada de forma más compacta como

$$x[k] = \sum_m s[k - m] h[m, k] + w[k] \triangleq s[k] * h[m, k] + w[k] \quad (2.58)$$

donde

$$h[m, k] \triangleq \sum_i a_i^b[k] p[m - \tau_i(kT_s)/T_s] \quad (2.59)$$

denota el m -ésimo coeficiente (complejo) del canal en el instante k . Su valor dependerá fundamentalmente de los coeficientes $a_i^b[k]$ correspondientes a los trayectos con retardo $\tau_i(t)$ próximos a kT_s . Si para todo m , el m -ésimo coeficiente del canal no cambia con k , entonces el canal es LTI. Si cada coeficiente cambia lentamente con k , entonces se dice que el canal es lentamente variante en el tiempo. En los sistemas celulares y en la mayoría de los sistemas inalámbricos de interés el canal varía lentamente en el tiempo.

Es conveniente tener en cuenta que, debido al desplazamiento Doppler, el ancho de banda de la salida $x_b(t)$ es generalmente algo mayor que el ancho de banda $F/2$ de la entrada $s_b(t)$, por lo que las muestras $\{y[k]\}$ no contienen toda la información de la señal continua $y_b(t)$. En la práctica este problema es ignorado habitualmente, ya que los filtros de recepción no son lo suficientemente estrechos como para eliminar la parte de la señal recibida fuera de $F/2$, correspondiente a las réplicas debidas al desplazamiento Doppler. Además, los coeficientes del canal cambian lentamente en el tiempo, por lo que se comportan como un filtro LTI sobre los periodos de interés y esencialmente modelan el canal. Además, siempre es conveniente que la frecuencia de muestreo de la entrada y de la salida sean iguales. De otro modo, sería necesario muestrear la salida al doble de la frecuencia de muestreo de la entrada y, por tanto, duplicar el número de muestras a procesar por el receptor.

Atendiendo a la naturaleza bivalente del canal (en las variables retardo y tiempo) hemos descrito los parámetros más importantes que caracterizan a un canal multitrayecto. Con el fin de disponer de modelos de referencia común sobre los que comparar distintos sistemas de comunicación hemos introducido los modelos probabilísticos de desvanecimiento Rayleigh y Rice. Aunque se apoyan en la realidad, las simplificaciones realizadas para su obtención obligan a utilizarlos con cautela si se quieren utilizar como predicción del rendimiento de un sistema de comunicación en un entorno real. Como contrapartida, su sencillez nos permite analizarlos en

mayor profundidad y utilizarlos como modelos de referencia para la comparación de distintos sistemas de comunicación.

El modelo de desvanecimiento Rayleigh tiene una formulación extremadamente sencilla. A primera vista, parece que no difiere demasiado del modelo clásico de canal AWGN. Sin embargo, veremos que el aprovechamiento de la capacidad potencial que nos brinda el canal requiere la utilización de técnicas que explotan un concepto que no existe en el canal AWGN: el concepto de *diversidad*.

Empezaremos analizando cuál es la capacidad teórica del canal Rayleigh, viendo que no se aleja demasiado de la capacidad de un canal AWGN. Sin embargo, si no se utiliza codificación de canal, ilustraremos cómo la probabilidad de error de bit en un canal Rayleigh se aleja muchísimo de la correspondiente a un canal AWGN. Mostraremos cómo técnicas extremadamente sencillas que hacen uso del concepto de *diversidad* permiten reducir dramáticamente la probabilidad de error de bit en el canal Rayleigh. Los grandes beneficios que se obtienen del aprovechamiento de la diversidad son uno de los motivos (junto con el de multiplexación espacial, que describiremos más adelante) que explican el interés suscitado por los sistemas MIMO. Por ello, creemos conveniente una exposición clara del concepto de diversidad.

2.6. Concepto de capacidad de un canal

Un canal discreto es un sistema que acepta como entrada símbolos s pertenecientes a un alfabeto \mathcal{A}_S y produce como salida símbolos x pertenecientes a un alfabeto \mathcal{A}_X , de acuerdo a un conjunto de probabilidades condicionadas

$$p_{ji} \quad i = 1, 2, \dots, N_S \quad j = 1, 2, \dots, N_X \quad (2.60)$$

donde $p_{ji} \triangleq P(x_j | s_i)$ representa la probabilidad de recibir el símbolo x_j dado que se transmitió el símbolo s_i . Es habitual agrupar las probabilidades anteriores en una matriz

$$\mathbf{P} \triangleq \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N_S} \\ p_{21} & p_{22} & \cdots & p_{2N_S} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N_X1} & p_{N_X2} & \cdots & p_{N_XN_S} \end{pmatrix} \quad (2.61)$$

de modo que si \mathbf{p}_S es el vector cuyas componentes son las probabilidades de cada símbolo de entrada (i.e., especifica la distribución de S), entonces se puede calcular la distribución de la salida como

$$\mathbf{p}_X = \mathbf{P}\mathbf{p}_S \quad (2.62)$$

Asumiremos además que el canal no tiene memoria, es decir

$$P(x_1, \dots, x_n | s_1, \dots, s_n) = \prod_{i=1}^n P(x_i | s_i) \quad (2.63)$$

donde s_1, \dots, s_n y x_1, \dots, x_n , con $s_i \in \mathcal{A}_S$ y $x_i \in \mathcal{A}_X$, representan n símbolos consecutivos transmitidos y recibidos, respectivamente.

Un caso particular de canal discreto sin memoria muy utilizado en la práctica es el denominado *canal simétrico*, que es aquel que cumple que cada fila de la matriz \mathbf{P} es una permutación de un conjunto de probabilidades $\{p_j\}_{j=1}^{N_S}$, y cada columna es una permutación de otro conjunto de probabilidades $\{q_j\}_{j=1}^{N_X}$. Si consideramos que $N_S = N_X = 2$ tenemos el canal simétrico binario (*Binary Symmetric Channel*, BSC), caracterizado simplemente por la probabilidad $p = p(1|0) = p(0|1)$. Teniendo en cuenta que $p(0|0) + p(1|0) = 1$ y $p(1|1) + p(0|1) = 1$ su matriz de probabilidades está dada por

$$\mathbf{P} \triangleq \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \quad (2.64)$$

El canal AWGN binario $y = \sqrt{E_s}s + w$, donde $s \in \{-1, 1\}$, E_s es la energía de símbolo y $w \sim N(0, \sigma)$, es equivalente a un BSC variante en el tiempo en el que la probabilidad p varía para cada símbolo transmitido. En esencia, lo que tenemos que hacer es suministrar al detector la información contenida en la muestra y en forma de un valor binario $x \in \{0, 1\} \equiv \mathcal{A}_X$ (el símbolo a la salida del canal) y su correspondiente probabilidad condicionada a cada posible símbolo transmitido $s \in \{-1, +\} \equiv \mathcal{A}_S$. Por ejemplo, fijemos de modo arbitrario la salida del canal al cero binario. La probabilidad de que esa salida del canal provenga del símbolo de entrada $s = -1$ viene dada por la verosimilitud $p(y|s = -1)$, y de que provenga del símbolo $s = +1$ por la verosimilitud $p(y|s = +1)$. A continuación, podemos identificar $p(y|s = -1)$ y $p(y|s = +1)$ con los valores $(1-p)$ y p de un BSC, respectivamente, normalizando estas verosimilitudes para que su suma sea la unidad. Cada uso del canal AWGN binario es, por tanto, equivalente a un BSC en el que la salida es el cero binario y

$$\begin{aligned} p &= \frac{p(y|s = +1)}{p(y|s = -1) + p(y|s = +1)} \\ &= \frac{1}{1 + e^{-\lambda(y)}} \end{aligned} \quad (2.65)$$

donde

$$\lambda(y) \triangleq \log \frac{p(y|s = +1)}{p(y|s = -1)} \quad (2.66)$$

es el cociente de log-verosimilitudes (*Log-Likelihood Ratio*, LLR). Nótese que, por el hecho de fijar arbitrariamente la salida del canal al cero binario, el valor de p puede ser mayor que $1/2$.

Hasta este momento hemos considerado que tanto la entrada como la salida del canal son discretas. También podemos considerar canales en los que tanto la entrada como la salida son continuas, como sucede en el canal AWGN o en el canal Rayleigh, por ejemplo.

La capacidad de Shannon de un canal sin memoria se define como el máximo, sobre todas las posibles distribuciones que puede tomar la entrada, de la información mutua entre la entrada y la salida [9]

$$C \triangleq \max_{f_S} I(S; X) \quad (2.67)$$

donde la información mutua entre S y X se define como

$$I(S; X) \triangleq H(S|X) - H(X) = H(X|S) - H(S) \quad (2.68)$$

y

$$H(S) \triangleq -E_S\{\log_2 f_S(s)\} \quad (2.69)$$

$$H(S|X) \triangleq -E_{S|X}\{\log_2 f_{S|X}(s)\} \quad (2.70)$$

son las definiciones de la *entropía* de S y de la *entropía condicionada* de S a X , respectivamente.

En el caso de un canal discreto simétrico, la distribución de la entrada que maximiza la información mutua es la uniforme y la capacidad del canal es

$$C = \log N_X + \sum_{j=1}^{N_X} p_j \log p_j \quad \text{bit/uso del canal} \quad (2.71)$$

donde $p_j = p(x_j) = \frac{1}{N_S} \sum_{j=1}^{N_S} p_{ji}$ es la suma de los elementos de la j -ésima columna de \mathbf{P} . En el caso del BSC, la expresión de la capacidad se reduce a

$$C = 1 - H_2(p) \quad (2.72)$$

donde $H_2(p) \triangleq -(1-p) \log_2(1-p) - p \log_2 p$ es la *función de entropía binaria*. Anteriormente mostramos la equivalencia entre el canal AWGN binario y un canal BSC variante en el tiempo. Precisamente esta variación temporal no nos permite concluir que la capacidad del canal AWGN binario sea igual a la del BSC, porque esta última es una función de p , que en el caso del canal AWGN binario varía para cada uso del canal.

La capacidad de un canal con entrada y salida continuas debe calcularse siempre imponiendo una restricción de coste sobre las secuencias que pueden transmitirse

$$\frac{1}{N} \sum_{i=1}^N c(s_i) \leq A \quad (2.73)$$

La necesidad de esta restricción es la siguiente. Si se nos permite introducir cualquier valor real en el canal, podríamos transmitir, en un solo uso del canal, una cadena de N dígitos $d_1 d_2 \cdots d_N$ simplemente haciendo que la entrada al canal fuese $s = d_1 d_2 \cdots d_N 00 \cdots 0$. Podríamos aumentar la cantidad de información libre de errores que fluye a través del canal en cada uso del mismo simplemente aumentando N . A su vez, la fiabilidad de la comunicación puede aumentarse sin más que incrementar el número de ceros en la parte final de s . Es por tanto habitual utilizar una restricción en la potencia utilizada, $E\{s^2\} \leq A$.

Una propiedad muy importante es que la distribución de X que maximiza la información mutua $I(X; Y)$, cuando se impone la restricción $E\{X^2\} \leq A$ es la gaussiana. Por tanto, para

canales continuos con restricción de potencia, la capacidad es igual a la información mutua cuando la entrada es gaussiana. Por ejemplo, la capacidad del canal AWGN en tiempo discreto $x = s + w$, donde $w \sim N(0, \sigma_w^2)$,

$$C = \frac{1}{2} \log_2(1 + \rho) \quad (2.74)$$

donde $\rho \triangleq E_s/\sigma_w^2$ la relación señal a ruido. Es fácil extender la expresión anterior a un canal AWGN en tiempo continuo donde la señal de entrada está limitada en banda a $(-B, B)$. Por el teorema del muestreo, se necesitan al menos $2B$ muestras por segundo para representar dicha señal. Cada muestra tiene una varianza σ_s^2 igual a la potencia de la señal P . Asumiendo que el ruido es un proceso gaussiano blanco con densidad espectral de potencia $N_0/2$, las muestras de ruido tomadas cada $1/(2B)$ segundos tendrán una varianza $\sigma^2 = N_0B$. Por tanto, la capacidad del canal AWGN en tiempo continuo se reduce a la del canal AWGN en tiempo discreto sin más que considerar la SNR como $\rho = P/(N_0B)$. Si además queremos expresarla en bit/s, basta tener en cuenta que en cada segundo estamos utilizando $2B$ veces el canal (tenemos $2B$ muestras por segundo)

$$C = B \log_2 \left(1 + \frac{P}{N_0B} \right) \text{ bit/s} \quad (2.75)$$

que, expresada en bit/s/Hz resulta ser

$$C = \log_2(1 + \rho) \text{ bit/s/Hz} \quad (2.76)$$

Asociado al concepto de capacidad del canal existe el *teorema de Shannon de codificación de canal* [9, 56, 47]:

- Para un canal discreto sin memoria, la capacidad del canal tiene la siguiente propiedad: dados cualesquiera $\epsilon > 0$ y $R < C$, existe un código de longitud N acotada con tasa de codificación $\geq R$ y un algoritmo de decodificación, tal que la probabilidad de error de bloque maximal es menor que ϵ .
- (Recíproco) Si se considera aceptable una probabilidad de error de bit p_b , se pueden alcanzar tasas de información menores o iguales que

$$R(p_b) = \frac{C}{1 - H_2(p_b)} \quad (2.77)$$

siendo $H_2(p_b)$ la función de entropía binaria y, en cambio, tasas de información mayores que $R(p_b)$ no son alcanzables.

La extensión del teorema a canales con entradas y salidas continuas se puede probar definiendo una secuencia de canales discretos (para los que es válido el teorema) derivados del canal continuo original, con valores crecientes de entradas y salidas, y probando que el máximo de la información mutua de esos canales tiende a C [47].

El teorema de codificación de canal de Shannon (1948) supuso un importante hito en la historia de las comunicaciones, dando lugar a una disciplina denominada Teoría de la Información. La parte positiva del mismo es que podemos disminuir todo lo que queramos la probabilidad de error sin necesidad de aumentar el ancho de banda ni la SNR, simplemente considerando un código de mayor longitud. La parte negativa del mismo, dada por la parte recíproca, es que introduce un límite en la máxima tasa de información (en bits por uso del canal) que soporta un canal si queremos comunicar mensajes a través de él con probabilidad de error tendiente a cero.

Muchas veces la mejor medida del rendimiento de una determinada técnica de procesamiento de señal y/o codificación es la E_b/N_0 que se necesita para alcanzar un valor de probabilidad de error que se considere esencialmente cero (e.g. 10^{-6}) y compararlo con la E_b/N_0 requerida para alcanzar, según el teorema de Shannon, una tasa de información igual a la eficiencia espectral del sistema.

2.7. Capacidad del canal Rayleigh con desvanecimiento plano

Consideremos el modelo equivalente discreto dado por la Ecuación (2.58) particularizado al caso de canal Rayleigh con desvanecimiento plano en frecuencia

$$x = hs + w \quad (2.78)$$

donde hemos prescindido del índice temporal k por razones de claridad. Consideraremos que el factor h es un proceso aleatorio $\mathcal{CN}(0, 1)$ incorrelado temporalmente. Suponiendo que el transmisor no conoce el canal, es decir, el valor tomado por la variable aleatoria h , la capacidad media del canal es

$$C = E_h \left\{ \frac{1}{2} \log_2(1 + |h|^2 \rho) \right\} \quad (2.79)$$

La Figura 2.1.(a) compara las capacidades del canal AWGN y del Rayleigh. Como se puede observar, la capacidad del canal Rayleigh es menor que la del canal AWGN. Sin embargo, esta diferencia no es muy acusada, siendo por tanto razonable suponer que el rendimiento de un sistema sin codificación, en términos de probabilidad de error en función de la relación señal a ruido en recepción, no difiera demasiado para los casos de canal AWGN y canal Rayleigh. Esta intuición sin embargo es completamente errónea. Consideremos un sistema con señalización BPSK. Para el canal AWGN, la probabilidad de error es [48]

$$P(e) = \frac{1}{2} \operatorname{erfc}(\sqrt{\rho}) \quad (2.80)$$

donde $\operatorname{erfc}(\cdot)$ es la función de error complementario que se define como

$$\operatorname{erfc} \triangleq \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \quad (2.81)$$

La relación entre la función de error complementario y la función de densidad de una v.a. gaussiana de media μ y varianza σ^2 es

$$\frac{1}{\sqrt{2\pi}\sigma} \int_x^\infty e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{2} \operatorname{erfc} \left(\frac{x-\mu}{\sqrt{2}\sigma} \right) \quad (2.82)$$

Particularizando la expresión anterior a $\mu = 0$ y $\sigma = 1$ obtenemos

$$Q(x) = \frac{1}{2} \operatorname{erfc} \left(\frac{x}{\sqrt{2}} \right) \quad (2.83)$$

donde

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt \quad (2.84)$$

Paradójicamente, a la función $Q(x)$ también se le denomina función de error complementario en la literatura. Otra propiedad importante de la función $\operatorname{erfc}(\cdot)$ de la que haremos uso más adelante es

$$\operatorname{erfc}(\sqrt{x}) \leq e^{-x} \quad (2.85)$$

Esta desigualdad se conoce con el nombre de *cota de Chernoff*.

Para obtener la probabilidad de error del canal Rayleigh, consideremos que para una realización concreta del canal, la probabilidad de error es

$$P(e|h) = \frac{1}{2} \operatorname{erfc} \left(\sqrt{|h|^2 \rho} \right) \quad (2.86)$$

La probabilidad de error media se puede calcular promediando la expresión anterior con respecto a la f.d.p. de la variable h . Equivalentemente, podemos promediar con respecto a la f.d.p. de $|h|^2$, que sigue una distribución exponencial de media unidad. En cualquier caso obtenemos [86, 48]

$$\bar{P}(e) = E_h \{P(e|h)\} = \frac{1}{2} \left(1 - \sqrt{\frac{\rho}{1+\rho}} \right) \quad (2.87)$$

que, haciendo uso de la siguiente expresión, válida cuando $x \rightarrow \infty$,

$$1 - \sqrt{\frac{x}{1+x}} \sim \frac{1}{2x} \quad (2.88)$$

podemos aproximar, para SNRs suficientemente altas, por

$$\bar{P}(e) \approx \frac{1}{4\rho} \quad (2.89)$$

Queda por tanto patente la diferencia entre las expresiones de la probabilidad de error correspondientes al canal AWGN y al canal Rayleigh en el caso binario. La Figura 2.1.(b) muestra ambas curvas. Mientras que la probabilidad de error en el canal AWGN decae según la función error complementario, en el canal Rayleigh lo hace tan sólo con el inverso de la relación

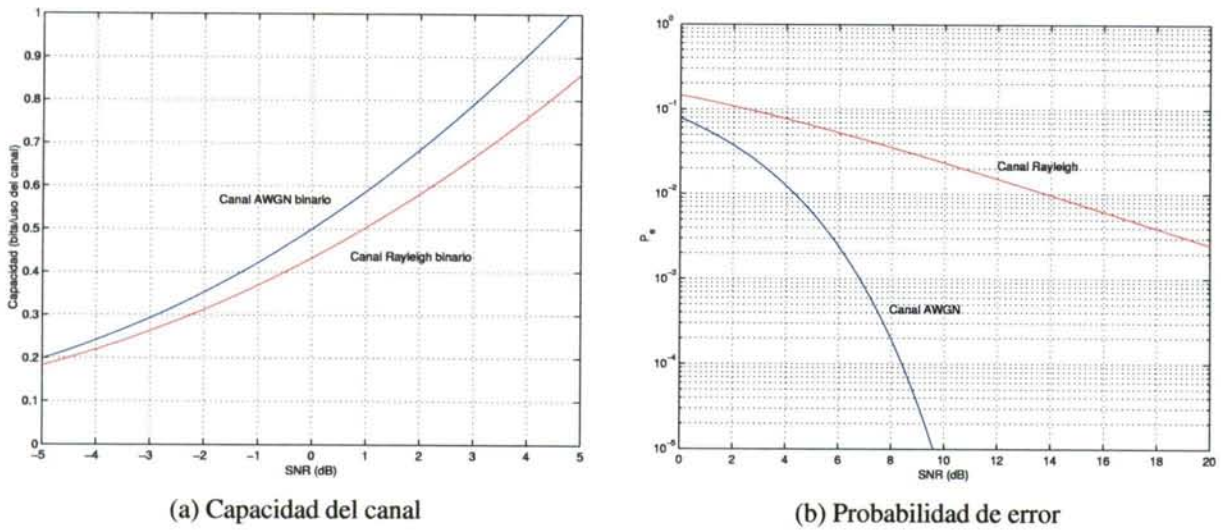


Figura 2.1: Comparación de los canales AWGN y Rayleigh (BPSK) en términos de capacidad y probabilidad de error

señal a ruido. La diferencia en la SNR requerida para alcanzar una determinada probabilidad de error es extremadamente alta. Así, mientras la capacidad asociada a uno y otro canal no presenta excesivas diferencias, la probabilidad de error para un sistema sin codificación correspondiente a uno y otro canal difiere de modo considerable. Expresiones similares pueden obtenerse en el caso de utilizar otros tipos de señalización.

Se plantea aquí la siguiente cuestión. Es conocido que la diferencia entre la SNR requerida para alcanzar una determinada probabilidad de error que podemos considerar como cero en un sistema sin codificación que utilice un método de señalización espectral y computacionalmente eficiente, actuando sobre el canal AWGN, y la SNR necesaria para que la transmisión sin errores sea posible, dada por el teorema de codificación de canal de Shannon, es significativa. Ésta es la razón por la cual la codificación de canal adquiere relevancia en el caso del canal AWGN. El mismo razonamiento se puede aplicar al canal Rayleigh, acentuando que en este caso la diferencia con respecto a la capacidad del canal es mucho más acusada. Esto sugiere que incluso el esquema más simple de codificación, como es un código repetitivo, puede obtener una ganancia de codificación significativa. La razón fundamental que explica esta ganancia es el hecho de que, si se transmite el mismo símbolo varias veces (normalizando convenientemente para que la energía total que utilizamos sea la misma), la SNR efectiva de la que se dispone en recepción para detectar dicho símbolo aumenta porque es poco probable que *todas* las veces que transmitimos dicho símbolo la realización del canal correspondiente sea un factor de atenuación profundo (*deep fade*). Esta idea es lo que se conoce como *diversidad*: la dispersión de la energía correspondiente a un símbolo de modo que experimente varias realizaciones independientes del canal.

La idea de diversidad, conceptualmente sencilla, se ha traducido en la práctica en multitud de esquemas de transmisión/recepción. Cada uno de ellos explota la diversidad de un modo

distinto presentando un grado distinto de complejidad computacional y de utilización de los recursos de tiempo y ancho de banda.

2.8. Diversidad

En esta sección expondremos algunos resultados que nos permitan cuantificar cuál es la ganancia que podemos obtener mediante la idea de diversidad. Recordemos que la diversidad consiste en esparcir la energía correspondiente a la transmisión de un símbolo de modo que en recepción dispongamos de varias réplicas de la misma señal afectadas por realizaciones distintas (independientes) del canal. Las réplicas transmitidas se denominan *ramas de diversidad* (*diversity branches*). Al incrementar el número de ramas de diversidad, la probabilidad de que todas las ramas experimenten simultáneamente un desvanecimiento profundo disminuirá considerablemente. Antes de proseguir, comentemos brevemente los dos tipos de diversidad más utilizados en la actualidad:

- **Diversidad espectral:** consiste en transmitir el mismo símbolo a lo largo de un intervalo de frecuencias múltiplo del ancho de banda de coherencia del canal. Por definición del ancho de banda de coherencia, la señal recibida experimenta desvanecimientos independientes en cada intervalo de frecuencias de longitud igual al ancho de banda de coherencia.
- **Diversidad temporal:** consiste en transmitir el mismo símbolo a lo largo de un intervalo temporal múltiplo del tiempo de coherencia del canal. Por definición del tiempo de coherencia, la señal recibida experimenta desvanecimientos independientes en cada intervalo temporal de duración igual al tiempo de coherencia.

Si atendemos a lo comentado sobre el ancho de banda de coherencia y el tiempo de coherencia del canal vemos que, esencialmente, ambas técnicas repiten la transmisión del mismo símbolo utilizando la variable retardo τ (diversidad espectral) o la variable tiempo t (diversidad temporal). En el capítulo siguiente introduciremos el concepto de diversidad espacial, fruto de considerar múltiples antenas en transmisión, recepción, o ambas. La utilización de varias antenas introduce una nueva dimensión sobre la que el canal varía, y por tanto una nueva dimensión sobre la que aplicar el concepto de diversidad.

Existen múltiples técnicas que implementan el concepto de diversidad espectral, temporal o espacial, o bien una combinación de las tres. En [86] puede encontrarse una completa exposición de técnicas de diversidad básicas que hacen uso de filtros lineales en transmisión y recepción. Estas técnicas, cuando son aplicadas en el dominio temporal o de la frecuencia se denominan de ensanchamiento (espectral o temporal), ya que expanden la respuesta en frecuencia o en tiempo del sistema para obtener diversidad.

Cuantifiquemos en la medida de lo posible la ganancia que puede ser obtenida mediante el uso de diversidad. Consideremos que un determinado símbolo s , perteneciente a una constelación con energía media unitaria, es transmitido a través de M canales Rayleigh

independientes. En recepción las observaciones tendrán la forma

$$x_i = \sqrt{\frac{E_s}{M}} h_i s + w_i \quad i = 1, 2, \dots, M \quad (2.90)$$

donde E_s/M es la energía de símbolo utilizada por el transmisor en cada una de las ramas de diversidad, y_i es la observación ruidosa correspondiente a la i -ésima rama de diversidad, h_i es la realización del canal en dicha rama de diversidad y w_i es el ruido AWGN de varianza σ_w^2 , incorrelado a lo largo de las diferentes ramas de diversidad, i.e., $E\{w_i w_j^*\} = 0$ si $i \neq j$. Recordemos que $h_i \sim \mathcal{CN}(0, 1)$ para cualquier rama de diversidad i . Utilizando notación vectorial podemos escribir la expresión anterior como

$$\mathbf{x} = \sqrt{\frac{E_s}{M}} \mathbf{h} s + \mathbf{w} \quad (2.91)$$

donde $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_M]^T$, $\mathbf{h} = [h_1 \ h_2 \ \dots \ h_M]^T$ y $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_M]^T$. Consideremos que en recepción aplicamos un filtro lineal \mathbf{f} . La salida del filtro la escribiremos como

$$z = \mathbf{f}^H \mathbf{x} \quad (2.92)$$

Dadas múltiples versiones de la misma señal, cada una afectada por un desvanecimiento distinto e independiente, el filtro lineal en recepción que maximiza la SNR a su salida viene dado por $\mathbf{f} = \mathbf{h}$. La aplicación de este filtro a la señal recibida recibe el nombre de *combinación de máxima relación* (*Maximal Ratio Combining*, MRC) [87, 86]. Utilizando MRC, la señal a la salida del filtro es

$$z = \mathbf{h}^H \mathbf{x} = \|\mathbf{h}\|^2 \sqrt{\frac{E_s}{M}} s + \mathbf{h}^H \mathbf{w} \quad (2.93)$$

y la SNR a la salida del filtro es

$$\rho_{out} = \frac{\|\mathbf{h}\|^2}{M} \rho \quad (2.94)$$

siendo $\rho = E_s/\sigma_w^2$ la SNR equivalente a un sistema que no utiliza diversidad. Es importante resaltar, que la Ecuación (2.94) es la SNR a la salida del filtro MRC *para una realización concreta del canal*. Suponiendo que a la salida del filtro MRC se realiza detección MV, la probabilidad de error de símbolo es aproximadamente [87, 88]

$$P(e|h) \approx \frac{\bar{n}_{min}}{2} \operatorname{erfc} \left(\sqrt{\rho_{out} d_{min}^2} \right) \quad (2.95)$$

donde \bar{n}_{min} y d_{min} son el número medio de vecinos más próximos y la mínima distancia euclídea entre elementos de la constelación resultante a la salida del filtro, respectivamente. Sustituyendo (2.94) en (2.95) y haciendo uso de la cota de Chernoff, $\operatorname{erfc}(\sqrt{x}) \leq e^{-x}$, obtenemos que

$$P(e|h) \leq \frac{\bar{n}_{min}}{2} e^{-\|\mathbf{h}\|^2 \frac{\rho d_{min}^2}{M}} \quad (2.96)$$

Al igual que dijimos para la SNR, esta expresión de la probabilidad de error es para una realización concreta del canal. Realizando la aproximación de que \bar{n}_{min} y d_{min} no dependen de la realización del canal, la probabilidad de error media viene dada por (cf. Ecuación (2.47))

$$\begin{aligned}\bar{P}(e) &\leq \frac{\bar{n}_{min}}{2} \prod_{i=1}^M \frac{1}{1 + \frac{\rho d_{min}^2}{M}} \\ &= \frac{\bar{n}_{min}}{2} \left(1 + \frac{\rho d_{min}^2}{M}\right)^{-M}\end{aligned}\quad (2.97)$$

que, para SNR altas, se puede aproximar por

$$\bar{P}(e) \leq \frac{\bar{n}_{min}}{2} \left(\frac{\rho d_{min}^2}{M}\right)^{-M}\quad (2.98)$$

La expresión anterior nos permite ver cuál es la ganancia obtenida por la diversidad, sea ésta del tipo que sea. La Figura 2.2.(a) muestra la probabilidad de error de símbolo (*Symbol Error Rate*, SER) media, $\bar{P}(e)$, para un sistema con señalización BPSK con distinto número de ramas de diversidad ($M = 1, 4, 8, 16$). Es interesante destacar que según la Ecuación (2.98) para $M = 1$ (ausencia de diversidad) la probabilidad de error decrece de modo inversamente proporcional a la SNR, tal y como ya habíamos deducido anteriormente (cf. Ecuación (2.89)). De modo análogo, podemos ver que en el caso de disponer de infinitas ramas de diversidad ($M \rightarrow \infty$), la Ecuación (2.97) se convierte en

$$\bar{P}(e) \leq \frac{\bar{n}_{min}}{2} e^{-\rho d_{min}^2}\quad (2.99)$$

donde hemos echo uso de la propiedad

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n\quad (2.100)$$

Nótese, que como era de esperar, la expresión (2.99) coincide con el resultado de aplicar la cota de Chernoff, $\text{erfc}(\sqrt{x}) \leq e^{-x}$, a la probabilidad de error del canal AWGN binario. Un estudio similar, más preciso al basarse en la expresión exacta de la probabilidad de error, puede encontrarse en [89].

El análisis que hemos efectuado muestra cómo la diversidad permite reducir sustancialmente la probabilidad de error, reduciéndola al caso de canal AWGN de modo asintótico en el número de ramas de diversidad. Sin embargo, debemos resaltar que las ideas que hemos presentado hacen uso del código más sencillo posible: la repetición de la transmisión de un mismo símbolo tantas veces como ramas de diversidad dispongamos. Al mismo tiempo, la probabilidad de error asintótica se reduce a la del canal AWGN *sin codificación de canal*, que sabemos que supone estar bien lejos de la capacidad tanto del canal AWGN como del Rayleigh. Parece razonable intuir que introduciendo codificación de canal en el sistema obtendremos, para un grado de diversidad determinado, una ganancia en términos de SNR requerida para alcanzar una probabilidad de error dada. A continuación vamos a cuantificar, en la medida de lo posible,

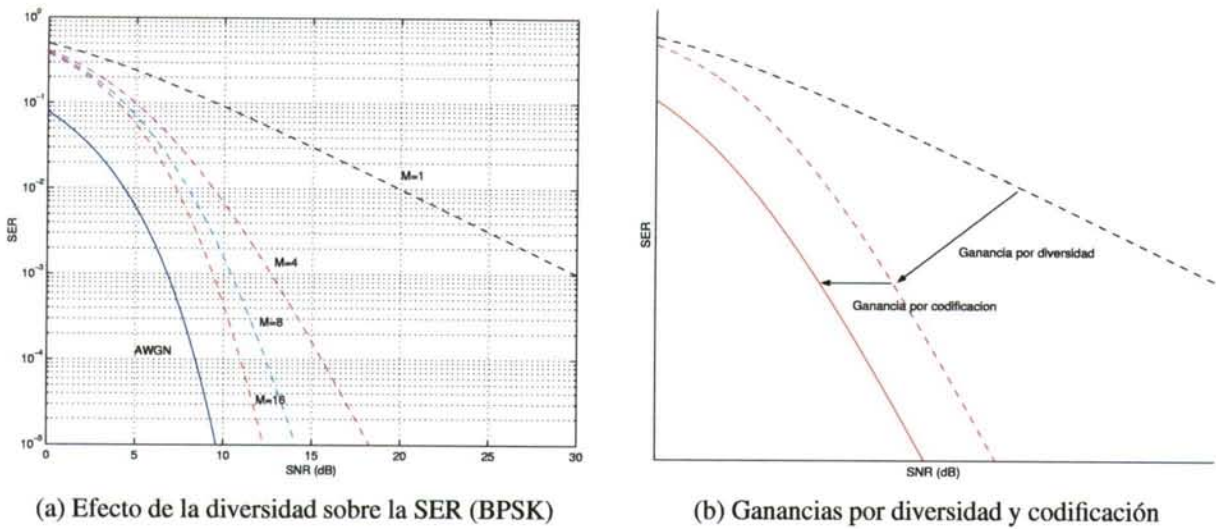


Figura 2.2: Efectos de la diversidad y la codificación sobre la SER en el canal de desvanecimiento Rayleigh

la *ganancia por codificación* que podemos obtener en un canal de desvanecimiento Rayleigh. Para ello, basta tener en cuenta que una manera de ver la codificación de canal es que, para cada símbolo transmitido, no se usa toda la constelación original sino tan sólo un subconjunto de la misma. Esto se traduce en un incremento tanto de la distancia mínima entre elementos de la constelación como en el número medio de vecinos más próximos a un elemento dado de la constelación. Tomemos ahora la expresión (2.98) y denotemos por $\bar{n}_{min,c}$ y $d_{min,c}$ el número medio de vecinos más próximos y la distancia euclídea mínima (media) entre elementos de la constelación para un sistema con codificación. La diferencia en probabilidad de error (en escala logarítmica) entre un sistema con codificación y grado de diversidad M' , y uno sin codificación y grado de diversidad M , será

$$\log P(e)_c - \log P(e) \leq \log \left(\frac{\bar{n}_{min,c}}{\bar{n}_{min}} \right) - M' \log \left(\frac{\rho d_{min,c}^2}{M'} \right) + M \log \left(\frac{\rho d_{min}^2}{M} \right) \quad (2.101)$$

Evaluemos la expresión anterior en el caso de que ambos sistemas tengan el mismo grado de diversidad ($M' = M$)

$$\log P(e)_c - \log P(e) \leq \log \left(\frac{\bar{n}_{min,c}}{\bar{n}_{min}} \right) - M \log \left(\frac{d_{min,c}^2}{d_{min}^2} \right) \quad (2.102)$$

Vemos que la disminución en probabilidad de error es constante con respecto a la SNR. Por tanto, la ganancia de codificación se traduce en un desplazamiento a la izquierda (en escala logarítmica) de la curva de probabilidad de error. En cambio, si consideramos dos sistemas con la misma codificación pero distinto grado de diversidad, la expresión (2.101) se reduce a

$$\log P(e)_c - \log P(e) \leq -(M' - M) \log(\rho) + cte. \quad (2.103)$$

es decir, la disminución en probabilidad de error es lineal (en escala logarítmica) con el incremento de diversidad. En consecuencia, la ganancia por diversidad se refleja en un cambio en la pendiente de caída de la curva de probabilidad de error con respecto a la SNR en dB. Esto hace que sea más importante que, al diseñar una técnica de codificación y señalización, debamos aprovechar al máximo el grado de diversidad disponible, y sólo cuando esta condición se satisface, nos preocupemos de obtener una ganancia por codificación.

2.9. Codificación en canales con desvanecimiento

Consideremos un sistema de comunicación genérico. Denotemos la secuencia transmitida por $\mathbf{s} = (s_1, s_2, \dots, s_N)$, donde s_k es un vector de señales perteneciente a una constelación \mathcal{S} . Estamos considerando aquí que la constelación considerada puede tener varias dimensiones, y los puntos de la constelación en cada dimensión valores reales. El caso de una constelación con coeficientes complejos puede ser englobado en este modelo considerando las partes real e imaginaria como dos dimensiones con coeficientes reales. Denotemos por $\mathbf{x} = (x_1, x_2, \dots, x_N)$ la secuencia recibida. El detector utilizará una determinada regla de decisión para escoger, a partir de la secuencia recibida, cuál fue la secuencia transmitida. La probabilidad de error asociada a cualquier regla de decisión se puede escribir como

$$P(e) = \sum_{\mathbf{s} \in \mathcal{S}^N} P(e|\mathbf{s})P(\mathbf{s}) \quad (2.104)$$

donde $\mathcal{S}^N = \mathcal{S} \times \mathcal{S} \times \dots \times \mathcal{S}$ y considerando que las secuencias transmitidas son equiprobables se tiene que

$$P(e) = \frac{1}{|\mathcal{S}|^N} \sum_{\mathbf{s} \in \mathcal{S}^N} P(e|\mathbf{s}) \quad (2.105)$$

La evaluación exacta de la probabilidad de error es posible sólo en ciertos casos. En general, se debe recurrir a la obtención de cotas superiores y/o inferiores para aproximar su valor. Un método ampliamente utilizado consiste en aprovechar el hecho de que la probabilidad de la unión de varios eventos es menor o igual que la suma de sus probabilidades individuales. Dicha cota se conoce como la cota de unión (*union bound*). Denotando por $P(\mathbf{s} \rightarrow \hat{\mathbf{s}})$ la probabilidad de error dos a dos (*pairwise*), es decir, la probabilidad de decidir en favor de una secuencia $\hat{\mathbf{s}}$ distinta de la realmente transmitida y aplicando la cota de unión tenemos que

$$P(e|\mathbf{s}) \leq \sum_{\hat{\mathbf{s}} \neq \mathbf{s}} P(\mathbf{s} \rightarrow \hat{\mathbf{s}}) \quad (2.106)$$

y sustituyendo (2.106) en (2.105) obtenemos que

$$P(e) \leq \frac{1}{|\mathcal{S}|^N} \sum_{\mathbf{s} \in \mathcal{S}^N} \sum_{\hat{\mathbf{s}} \neq \mathbf{s}} P(\mathbf{s} \rightarrow \hat{\mathbf{s}}) \quad (2.107)$$

La expresión anterior es válida para cualquier regla de decisión, técnica de codificación, constelación y canal. La particularizaremos para los casos de canal AWGN y canal con desvanecimiento asumiendo detección óptima en el sentido MV (equivalente al criterio MAP al ser las secuencias transmitidas equiprobables).

Estudiemos el caso de canal AWGN. Los elementos de la secuencia recibida $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ tendrán la forma

$$\mathbf{x}_k = \mathbf{s}_k + \mathbf{n}_k \quad (2.108)$$

donde \mathbf{n}_k es un vector cuyas componentes son muestras de un proceso de ruido blanco gaussiano con densidad espectral de potencia $N_0/2$. La regla de decisión óptima en el sentido MV es escoger como palabra transmitida aquella que maximiza la verosimilitud de la secuencia observada. Al ser gaussiana la f.d.p. de la secuencia de observaciones, condicionada a la secuencia transmitida, el criterio MV es equivalente a escoger como secuencia transmitida aquella que minimiza la distancia Euclídea a la secuencia observada

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{s} - \mathbf{x}\|^2 \quad (2.109)$$

En este caso las secuencias \mathbf{s} y $\hat{\mathbf{s}}$ de (2.107) pueden verse como puntos de la constelación multidimensional \mathcal{S}^N . La probabilidad de error dos a dos de la regla de decisión MV cuando se utiliza una constelación multidimensional en un canal AWGN es [48]

$$P(\mathbf{s} \rightarrow \hat{\mathbf{s}}) = \frac{1}{2} \operatorname{erfc} \left(\frac{\|\hat{\mathbf{s}} - \mathbf{s}\|}{2\sqrt{N_0}} \right) \quad (2.110)$$

que, sustituida en la expresión general (2.107), da lugar a

$$P(e) \leq \frac{1}{|\mathcal{S}|^N} \sum_{\mathbf{s} \in \mathcal{S}^N} \sum_{\hat{\mathbf{s}} \neq \mathbf{s}} \frac{1}{2} \operatorname{erfc} \left(\frac{\|\hat{\mathbf{s}} - \mathbf{s}\|}{2\sqrt{N_0}} \right) \quad (2.111)$$

Podemos obtener una cota superior menos ajustada, pero más simple, utilizando la cota de Chernoff (2.85)

$$P(e) \leq \frac{1}{2|\mathcal{S}|^N} \sum_{\mathbf{s} \in \mathcal{S}^N} \sum_{\hat{\mathbf{s}} \neq \mathbf{s}} \exp \left(-\frac{\|\hat{\mathbf{s}} - \mathbf{s}\|^2}{4N_0} \right) \quad (2.112)$$

La expresión anterior se denomina cota de unión-Bhattacharyya. La suma más interna en dicha expresión estará dominada por las secuencias codificadas a menor distancia euclídea de \mathbf{s} . Denotando por d_{\min} dicha distancia y por $n_{\min}(\mathbf{s})$ el número de las secuencias separadas a esa distancia tenemos

$$P(e) \lesssim \frac{\bar{n}_{\min}}{2} \exp \left(-\frac{d_{\min}^2}{4N_0} \right) \quad (2.113)$$

donde

$$\bar{n}_{\min} = \frac{1}{|\mathcal{S}|^N} \sum_{\mathbf{s} \in \mathcal{S}^N} n_{\min}(\mathbf{s}) \quad (2.114)$$

es el número medio de vecinos más próximos en la constelación \mathcal{S}^N .

Por tanto, la probabilidad de error en el canal AWGN depende de la distancia euclídea mínima, d_{min} , entre dos palabras del código, además del número de secuencias cuya distancia es d_{min} . Es importante resaltar que la relevancia de d_{min} es mayor que la de n_{min} al aparecer como argumento de una exponencial.

Consideremos que el código de canal es binario, que cada dimensión de la constelación es también binaria, y que la distancia entre los dos puntos de la constelación en cada dimensión es igual, sea cual sea dicha dimensión. Entonces la distancia euclídea entre dos palabras código es equivalente a la distancia de Hamming, definida como el número de posiciones en las que difieren. La probabilidad de error del canal AWGN (2.113) en este caso depende de la mínima distancia de Hamming entre dos palabras código. Sin embargo, en general utilizaremos una constelación no binaria que sea espectralmente eficiente (e.g. M -QAM). Sea cual sea el alfabeto de salida del código, la distancia de Hamming entre dos palabras código ya no es equivalente a la distancia euclídea.

Si la constelación S utilizada es binaria, la distancia Euclídea es proporcional a la distancia de Hamming, por lo que si empleamos un código binario la probabilidad de error depende de la distancia de Hamming entre las palabras del código. Sin embargo, si la constelación empleada no es binaria, la distancia euclídea entre dos secuencias no es en general proporcional a la distancia de Hamming. Subrayaremos aquí el hecho de que la distancia de Hamming entre dos secuencias construidas sobre un alfabeto M -ario sigue siendo exclusivamente el número de posiciones en las que difieren, sin tener ninguna relación con cualquier distancia definida sobre el alfabeto base. Por ejemplo, si consideramos las secuencias (0,7,2,5) y (2,6,2,6), la distancia de Hamming entre ambas es 3, sin importar que la diferencia entre los símbolos que ocupan la primera posición, 0 y 2, sea igual a dos, o que la distancia euclídea entre ambos sea $\sqrt{2}$, suponiendo que el alfabeto etiqueta los puntos de una constelación 8-PSK siguiendo el *mapping* natural. Por tanto, utilizar un esquema consistente en un código binario seguido de una asignación (*mapping*) de bits a símbolos de la constelación utilizada es subóptimo. Por tanto, si se quiere diseñar de modo óptimo un esquema de codificación y modulación utilizando una constelación con dimensiones no binarias, ambos módulos han de diseñarse de modo conjunto para optimizar la distancia euclídea de las palabras código resultantes.

Fue Ungerboeck en 1982 quien desarrolló esta idea, y propuso un método de diseño conjunto de los procesos de codificación y modulación denominado *Trellis Coded Modulation* (TCM). Como su nombre indica, el esquema resultante consiste en un codificador descrito por un *trellis*, con la peculiaridad de que el alfabeto de salida que etiqueta sus ramas está formado por los puntos de la constelación considerada. Ungerboeck formuló reglas empíricas para etiquetar las ramas del *trellis* de modo que se maximizase la mínima distancia euclídea entre cualesquiera dos palabras código. El rendimiento de TCM se mostró sensiblemente superior al tradicional esquema de código binario seguido de *mapping* bit a símbolo.

TCM puede verse como la extensión de los códigos convolucionales a constelaciones no binarias. El alto rendimiento mostrado por los códigos Turbo en el caso binario, también motivó su extensión al caso de constelaciones no binarias mediante la filosofía TCM, dando

lugar a los esquemas Turbo-TCM [90].

Dirijamos ahora nuestro centro de atención al canal con desvanecimiento. La secuencia recibida estará compuesta por elementos de la forma

$$\mathbf{x}_k = h_k \mathbf{s}_k + \mathbf{n}_k \quad (2.115)$$

donde las v.a. h_k son independientes, así como los vectores de ruido gaussiano \mathbf{n}_k y las componentes que los forman. Asumiendo que el receptor conoce el valor tomado por h_k , la regla de decisión MV se reduce en este caso a

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \sum_{k=1}^N \|\mathbf{x}_k - h_k \mathbf{s}_k\|^2 \quad (2.116)$$

Asumiendo desvanecimiento Rayleigh y aplicando la propiedad (2.85), puede demostrarse [48] que la probabilidad de error dos a dos de esta regla de decisión cumple que

$$P(\mathbf{s} \rightarrow \hat{\mathbf{s}}) \leq \prod_{k:\mathbf{s}_k \neq \hat{\mathbf{s}}_k} \frac{1}{1 + \|\hat{\mathbf{s}}_k - \mathbf{s}_k\|^2 / (4N_0)} \quad (2.117)$$

Podemos relajar la cota anterior escribiendo

$$\begin{aligned} P(\mathbf{s} \rightarrow \hat{\mathbf{s}}) &\leq \prod_{k:\mathbf{s}_k \neq \hat{\mathbf{s}}_k} \frac{1}{\|\hat{\mathbf{s}}_k - \mathbf{s}_k\|^2 / (4N_0)} \\ &= \frac{1}{[\delta^2(\mathbf{s}, \hat{\mathbf{s}}) / (4N_0)]^{d_H(\mathbf{s}, \hat{\mathbf{s}})}} \end{aligned} \quad (2.118)$$

donde

$$\delta^2(\mathbf{s}, \hat{\mathbf{s}}) \triangleq \left[\prod_{k:\mathbf{s}_k \neq \hat{\mathbf{s}}_k} \|\hat{\mathbf{s}}_k - \mathbf{s}_k\|^2 \right]^{1/d_H(\mathbf{s}, \hat{\mathbf{s}})} \quad (2.119)$$

es la media geométrica de las distancias euclídeas al cuadrado distintas de cero entre las componentes de \mathbf{s} y $\hat{\mathbf{s}}$. Como hicimos anteriormente en el caso del canal AWGN, asumiremos que la probabilidad de error dos a dos estará dominada por las secuencias con mínima distancia (en este caso distancia de Hamming) a \mathbf{s} .

Comentemos las principales diferencias en la probabilidad de error de los canales AWGN y Rayleigh. En el canal AWGN, el parámetro más importante es la mínima distancia euclídea entre dos secuencias transmitidas. En el canal Rayleigh, el parámetro más importante es la mínima distancia de Hamming entre dos palabras código. El número de pares de secuencias con mínima distancia de Hamming, n_{min} , así como el valor de $\delta^2(\mathbf{s}, \hat{\mathbf{s}})$ de cada una de ellas, juega un papel de mucha menor importancia. Si nos fijamos en la expresión (2.118), y la comparamos con (2.98), vemos que $d_{H,min}$ equivale a la diversidad introducida por el código (*code diversity*). Por tanto, el criterio a seguir en el diseño de esquemas de codificación y modulación para el

canal Rayleigh es completamente distinto al de canales AWGN. La mínima distancia euclídea, que es el parámetro fundamental en el canal AWGN, juega un papel de poca importancia en el canal Rayleigh, donde el parámetro de importancia es la mínima distancia de Hamming. En consecuencia, esquemas optimizados para el canal AWGN (e.g. TCM) pueden ser subóptimos en el canal Rayleigh.

En el caso de los sistemas que utilizan un código y una constelación binarios, el criterio es el mismo tanto para el canal AWGN como para el Rayleigh. Una manera sencilla de construir esquemas de codificación para el canal Rayleigh cuando la constelación es no binaria es lo que se denomina *Bit-Interleaved Coded Modulation* (BICM). La idea es utilizar un código optimizado para el caso binario (máxima distancia de Hamming) seguido de una asignación de bits a símbolos. Ya que estamos asignando varios bits a un símbolo, debemos intentar que los bits en los que difieren dos palabras código no se asignen al mismo símbolo, ya que entonces la distancia de Hamming resultante a nivel de símbolo será menor que la distancia de Hamming original a nivel de bit. Para ello se utiliza un entrelazador a nivel de bit antes del *mapping* bit a símbolo. Además, este entrelazador ya ejerce la función del entrelazador a nivel de símbolo que habitualmente se utiliza en canales con desvanecimiento para decorrelar muestras consecutivas, por lo que no supone un incremento en la complejidad del sistema.

Existe además un aspecto importante en el entrelazador a nivel de bit, que lo diferencia con respecto a un entrelazador a nivel de símbolo. En un canal Rayleigh, la función del entrelazador a nivel de símbolo es, como hemos dicho, decorrelar las muestras recibidas. En un canal AWGN, esto carece de sentido, por lo que utilizar un entrelazador no tiene ningún efecto. Sin embargo, al introducir el entrelazador a nivel de bit entre la salida del codificador y la asignación bit a símbolo, el esquema conjunto puede verse en realidad como un codificador cuyas palabras código tienen una longitud igual a la profundidad del entrelazado y una estructura aparentemente aleatoria (debido a la dispersión introducida por el entrelazador). La decodificación óptima (MV) para este codificador conjunto no es viable, al ser grande la longitud de las palabras código. Por ejemplo, si el código binario utilizado es un codificador convolucional, podríamos describir el esquema de codificación, entrelazado y asignación a símbolos por un trellis. Sin embargo, su complejidad sería tan elevada que haría imposible una decodificación MV. No obstante, precisamente por la presencia del entrelazador veremos en capítulos posteriores que podemos aplicar el principio Turbo para decodificar de modo subóptimo, pero con excelentes resultados, el esquema conjunto. Esto hace de BICM una solución sencilla (aprovecha todo el trabajo realizado en códigos binarios), de bajo coste computacional y, sobre todo, robusta en el sentido de proporcionar excelentes resultados tanto en canal AWGN como Rayleigh. El interés de disponer de una solución que proporcione buenos resultados para los dos casos es clara: en un canal multitrayecto, el canal puede cambiar desde Rayleigh puro a Rice con una fuerte componente especular. En el límite, cuando la componente especular es suficiente grande, el canal Rice se reduce al AWGN.

2.10. Conclusiones

El objetivo del presente capítulo ha sido la descripción del entorno de propagación radio que utilizan los sistemas de comunicación inalámbricos (tanto fijos como móviles). Se han expuesto los tres mecanismos básicos que influyen en la propagación de la señal: reflexión, difracción y difusión. Estos tres mecanismos son la base de los modelos de propagación que podemos clasificar en dos grandes grupos, a gran escala y a pequeña escala, según su rango de aplicabilidad. Los modelos de gran escala se centran en caracterizar la pérdida de potencia de la señal transmitida en función de la distancia al transmisor. Su utilidad es, por ejemplo, la predicción del área de cobertura de una estación base. Los modelos de pequeña escala se centran, en cambio, en describir la variación en la señal recibida a lo largo de puntos cercanos entre sí y alejados del transmisor, de forma que el efecto de atenuación de potencia de la señal sea común a todos ellos. Estos modelos son los que permiten describir el canal de un modo conveniente para el diseñador de técnicas de codificación y/o señalización. El canal resultante de la propagación radio es un canal multitrayecto que, finalmente, se puede expresar como un sistema lineal y variante en el tiempo. Estas condiciones permiten expresar la respuesta al impulso del sistema como una función de dos variables: retardo y tiempo.

Considerando el canal como un proceso doblemente estocástico, se han introducido una serie de funciones que permiten caracterizarlo. Asociadas a las mismas se han definido cuatro parámetros fundamentales: dispersión Doppler, tiempo de coherencia, dispersión multitrayecto y ancho de banda de coherencia.

Finalmente, hemos evaluado la capacidad del canal multitrayecto para el caso particular del canal con desvanecimiento plano Rayleigh y hemos demostrado que es algo menor que la del canal AWGN. Sin embargo, no ocurre lo mismo con la probabilidad de error, lo que nos ha llevado a exponer un concepto clave en los sistemas de comunicación sobre canales multitrayecto: la diversidad. Hemos cuantificado la ganancia que se obtiene por diversidad en términos de la SNR necesaria para alcanzar una determinada probabilidad de error. También hemos cuantificado la ganancia que es posible obtener al utilizar codificación de canal. Hemos visto que, mientras la diversidad provoca un cambio en la pendiente de caída de la curva de probabilidad de error, la codificación sólo permite desplazarla a la izquierda una cantidad constante. Por ello, la ganancia por diversidad es mucho más importante que la ganancia por codificación (en el régimen de alta SNR) en sistemas de comunicación sobre canal multitrayecto.

En el siguiente capítulo consideraremos una nueva forma de obtener diversidad en un sistema de comunicación: la utilización de múltiples antenas. Además, veremos que, a diferencia de la diversidad espectral y/o temporal, la diversidad espacial no sólo reduce la probabilidad de error sino que, cuando se utilizan múltiples antenas tanto en transmisión como en recepción, la capacidad del canal asociado crece de modo significativo, posibilitando altas eficiencias espectrales sin aumento de coste en los recursos de potencia, tiempo o ancho de banda.

Capítulo 3

Modelado de Sistemas Multiple Input Multiple Output (MIMO)

Do not worry about your problems with mathematics, I assure you mine are far greater

—Albert Einstein

En el presente capítulo consideraremos sistemas de comunicación que utilizan múltiples antenas tanto en transmisión como en recepción. El canal resultante es, por tanto, un canal con múltiples entradas y múltiples salidas (*Multiple-Input Multiple-Output*, MIMO). Cada entrada se corresponde con una antena transmisora y cada salida con una antena receptora. Como consecuencia, en el canal se dispone de una nueva dimensión: la espacial. Disponer de la dimensión espacial, además de la temporal, motiva que los sistemas MIMO también se denominen sistemas Espacio-Temporales (*Spatio-Temporal*, ST). Empezaremos extendiendo los modelos expuestos en el capítulo anterior para incorporar la nueva dimensión espacial. A continuación explicaremos las dos consecuencias más importantes que se derivan de disponer de la dimensión espacial: la posibilidad de multiplexación espacial y la ganancia por diversidad espacial. Por último, merecerá también nuestra atención el estudio de la capacidad del canal MIMO.

3.1. Modelo de sistema de comunicación MIMO

Consideremos un sistema de comunicaciones inalámbrico que utiliza múltiples antenas tanto en transmisión como en recepción. La Figura 3.1 muestra un sistema MIMO típico con n_t antenas transmisoras y n_r antenas receptoras. La secuencia de bits de información fuente la denotaremos por $u(k)$, $k = 1, 2, \dots, K$, donde K es el número total de bits que se van a enviar. El módulo de codificación espacio-temporal realiza la tarea de construir, a partir de la secuencia de bits fuente $u(k)$, una secuencia de vectores de bits (o símbolos, según el tipo de código

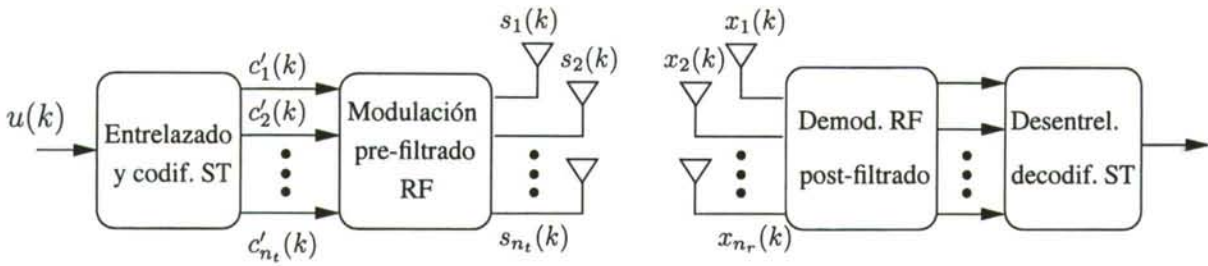


Figura 3.1: Diagrama de bloques de un sistema de comunicación MIMO

de canal utilizado) con una determinada estructura espacio-temporal y que vamos a denotar por $\mathbf{c}(k) = [c_1(k), c_2(k), \dots, c_{n_t}(k)]^T$ con $k = 1, 2, \dots, N$. Este módulo es el equivalente al codificador de canal en sistemas de una sola antena (*Single-Input Single-Output*, SISO) y juega el mismo papel de introducir redundancia para conseguir una mejor utilización de los recursos de potencia y ancho de banda en el canal. La diferencia fundamental con respecto al caso SISO es que ahora la codificación debe tener en cuenta la nueva dimensión espacial para transportar los bits (o símbolos) codificados y, además, un modelo de canal distinto.

Habitualmente la secuencia de vectores $\mathbf{c}(k)$ es entrelazada como paso previo a su transmisión. El entrelazado es ventajoso en los canales con multitrayecto, ya que los desvanecimientos causados por el canal pueden estar correlados. La mayoría de los códigos utilizados en comunicaciones son muy sensibles a la presencia de errores (u observaciones con baja SNR en el caso de procesar directamente éstas) en ráfagas a la entrada del decodificador. Dado un mismo número de errores en un bloque de datos, habitualmente el código es capaz de corregirlos si éstos aparecen dispersos a lo largo del bloque mientras que, si están agrupados en una ráfaga, lo normal es que resulten en una nueva ráfaga de errores a la salida del decodificador. Este hecho se ve de modo muy claro en los códigos convolucionales (y *trellis* en general) en los que la capacidad de corrección de errores está directamente ligada a la duración del mínimo evento de error. Por otra parte, en la práctica el canal es siempre desconocido. La existencia de una correlación temporal en los desvanecimientos se refleja en que las observaciones tomadas en recepción, condicionadas al símbolo transmitido, no son independientes, tal y como ocurre en el canal AWGN. Entrelazando los bits (o símbolos) en transmisión se persigue que varias muestras recibidas de forma consecutiva sean estadísticamente independientes. En sistemas de comunicación ST el entrelazado debe considerar todas las dimensiones disponibles, esto es, tiempo, frecuencia y espacio. Si las observaciones son independientes, en recepción se puede aplicar un algoritmo de decodificación que asuma la hipótesis de independencia lo que nos evita el estimar la función de autocorrelación del canal e incluir esa información estadística en el algoritmo de decodificación.

Es importante resaltar que, en principio, si el canal es AWGN, la utilización de un entrelazador a nivel de símbolo no aporta ninguna ventaja, ya que en este caso las observaciones, condicionadas al símbolo transmitido, ya son independientes entre sí. Posteriormente, sin embargo, veremos las ventajas de realizar un entrelazado a nivel de bit como paso previo

a la asignación de bits a símbolos mediante un esquema denominado *Bit-Interleaved Coded Modulation* (BICM), incluso en el caso de canal AWGN.

Tras los procesos de codificación, entrelazamiento y asignación de bits a símbolos se dispone de una secuencia de vectores de símbolos $c'(k)$ a transmitir. Antes de modular estos símbolos se puede realizar un filtrado espacio-temporal para adaptar la potencia transmitida por cada antena a las características del canal. Evidentemente, esto requiere que el transmisor disponga de algún tipo de información estadística sobre el canal. Los sistemas que incorporan este pre-filtrado en transmisión basándose en información sobre el estado del canal se denominan sistemas con precodificación ST. El disponer de información sobre el estado del canal en transmisión se puede traducir en ganancias significativas, tanto desde un punto de vista teórico ya que se dispone de una mayor capacidad del canal, como a nivel práctico en términos de SNR requerida y complejidad computacional de los algoritmos utilizados.

Finalmente, cada componente de los vectores de símbolos resultantes del pre-filtrado ST es modulada mediante una forma de pulso, trasladada a radiofrecuencia (*Radio Frequency*, RF) y radiada por una de las antenas transmisoras. Posteriormente, las señales transmitidas se propagan por el canal radio, cuyo efecto multitrayecto se traduce en la existencia en cada antena receptora de la suma de múltiples réplicas con diferentes atenuaciones y retardos de todas las señales transmitidas. A la señal RF recibida en cada antena se le suma una componente de ruido térmico. Tras la etapa de traslación a banda-base (*down-conversion*) la mezcla de señales y ruido recibida en cada antena es filtrada mediante el filtro adaptado a la forma de pulso utilizada en la modulación. La señal a la salida del filtro adaptado correspondiente a la j -ésima antena receptora tiene la forma

$$x_j(t) = \sum_{i=1}^{n_t} h_{ji}(\tau, t) * s_i(t) + g_j(t) \quad (3.1)$$

donde $s_i(t) = \sum_{k=1}^N s_i(k)\delta(t - kT_s)$ es la secuencia (en tiempo continuo) de símbolos transmitidos. Agrupando las señales recibidas en un vector $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_{n_r}(t)]^T$, podemos expresar la relación entrada-salida del canal MIMO en tiempo continuo como

$$\mathbf{x}(t) = \mathbf{H}(\tau, t) * \mathbf{s}(t) + \mathbf{g}(t) \quad (3.2)$$

donde

$$\mathbf{H}(\tau, t) = \begin{pmatrix} h_{11}(\tau, t) & h_{12}(\tau, t) & \cdots & h_{1n_t}(\tau, t) \\ h_{21}(\tau, t) & h_{22}(\tau, t) & \cdots & h_{2n_t}(\tau, t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_r1}(\tau, t) & h_{n_r2}(\tau, t) & \cdots & h_{n_rn_t}(\tau, t) \end{pmatrix} \quad (3.3)$$

es la matriz que contiene la respuestas al impulso de los distintos subcanales radio, $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_{n_t}(t)]^T$ es el vector de señales transmitidas y $\mathbf{g}(t) = [g_1(t), g_2(t), \dots, g_{n_r}(t)]^T$ es el vector de ruido.

Considerando que la señal recibida en cada antena es pasada a través del filtro adaptado a la forma de pulso utilizada en la modulación y muestreada a periodo de símbolo, se obtiene

un vector de observaciones $\mathbf{x} = [x_1(k), x_2(k), \dots, x_{n_r}(k)]^T$ que es un conjunto de estadísticos suficientes para la detección de los símbolos transmitidos. Utilizando el modelo discreto equivalente de la sección anterior, tenemos el siguiente modelo discreto equivalente para el caso MIMO

$$\mathbf{x}[k] = \mathbf{H}[m, k] * \mathbf{s}[k] + \mathbf{g}[k] \quad (3.4)$$

donde

$$\mathbf{H}[m, k] = \begin{pmatrix} h_{11}[m, k] & h_{12}[m, k] & \cdots & h_{1n_t}[m, k] \\ h_{21}[m, k] & h_{22}[m, k] & \cdots & h_{2n_t}[m, k] \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_r1}[m, k] & h_{n_r2}[m, k] & \cdots & h_{n_r n_t}[m, k] \end{pmatrix} \quad (3.5)$$

y $\mathbf{s}[k] = [s_1[k], s_2[k], \dots, s_{n_t}[k]]^T$ y $\mathbf{g}[k] = [g_1[k], g_2[k], \dots, g_{n_r}[k]]^T$.

El receptor, a partir de los vectores de muestras $\mathbf{x}[k]$, debe estimar los bits fuente $u(k)$, deshaciendo para ello las distintas operaciones realizadas en transmisión.

3.2. Capacidad del canal MIMO determinista y plano en frecuencia

Teniendo en cuenta el modelo equivalente discreto de un canal MIMO que acabamos de exponer, vamos a proceder a estudiar en esta sección su capacidad de transferencia de información. Empezaremos analizando la situación más sencilla en la que el canal es determinista y plano en frecuencia, i.e., el canal es constante y no varía ni con el tiempo ni con el retardo. Esta situación puede considerarse como la equivalente al canal AWGN para sistemas MIMO.

Supongamos que el canal tiene un ancho de banda de 1 Hz y es plano a lo largo de ese rango de frecuencias. Al ser el ancho de banda unitario, la potencia de cada componente de ruido será $\sigma_n = N_0$. También haremos la suposición de que el período de símbolo es la unidad, $T_s = 1$, de modo que podamos hablar de modo totalmente intercambiable de energía y potencia. Estas suposiciones las mantendremos en adelante, a menos que se especifique lo contrario. La relación entrada-salida del canal es

$$\mathbf{x} = \sqrt{\frac{E_s}{n_t}} \mathbf{H} \mathbf{s} + \mathbf{n} \quad (3.6)$$

donde hemos eliminado el índice k a efectos de claridad en la exposición y normalizado la energía transmitida. Asumiremos que la matriz de autocovarianzas del vector de símbolos transmitidos $\mathbf{R}_{\mathbf{ss}} = E\{\mathbf{s}\mathbf{s}^H\}$ (estamos suponiendo que \mathbf{s} tiene media nula) cumple que $Tr\{\mathbf{R}_{\mathbf{ss}}\} = n_t$. Esta es la restricción de potencia (energía) que asumiremos para poder calcular la capacidad.

Asumiendo que el canal es conocido por el receptor, podemos calcular la capacidad como [8]

$$C = \max_{\mathbf{R}_{ss}: \text{Tr}\{\mathbf{R}_{ss}\}=n_t} \log_2 \det \left(\mathbf{I}_{n_r} + \frac{E_s}{n_t N_0} \mathbf{H} \mathbf{R}_{ss} \mathbf{H}^H \right) \text{ bit/s/Hz} \quad (3.7)$$

3.2.1. Canal desconocido para el transmisor

En este caso el vector \mathbf{s} debe ser escogido de acuerdo a $\mathbf{R}_{ss} = \mathbf{I}_{n_t}$, resultando

$$C = \log_2 \det \left(\mathbf{I}_{n_r} + \frac{E_s}{n_t N_0} \mathbf{H} \mathbf{H}^H \right) \quad (3.8)$$

Utilizando la descomposición en autovalores de $\mathbf{H} \mathbf{H}^H = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H$, podemos expresar la ecuación anterior como

$$C = \log_2 \det \left(\mathbf{I}_{n_r} + \frac{E_s}{n_t N_0} \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H \right) \quad (3.9)$$

Utilizando la identidad $\det(\mathbf{I}_m + \mathbf{A} \mathbf{B}) = \det(\mathbf{I}_n + \mathbf{B} \mathbf{A})$, para matrices $\mathbf{A}_{m \times n}$ y $\mathbf{B}_{n \times m}$, y que $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}_{n_r}$, la ecuación anterior se simplifica a

$$C = \log_2 \det \left(\mathbf{I}_{n_r} + \frac{E_s}{n_t N_0} \mathbf{\Lambda} \right) \quad (3.10)$$

que es equivalente a

$$C = \sum_{i=1}^r \log_2 \left(1 + \frac{E_s}{n_t N_0} \lambda_i \right) \quad (3.11)$$

donde r es el rango de \mathbf{H} y λ_i es el i -ésimo autovalor de $\mathbf{H} \mathbf{H}^H$. En general, además tenemos que si σ_i es el i -ésimo valor singular positivo de \mathbf{H} , entonces el i -ésimo autovalor positivo de $\mathbf{H} \mathbf{H}^H$ es $\lambda_i = \sigma_i^2$. Esta ecuación nos permite ver la capacidad de un canal MIMO como la suma de las capacidades de r canales SISO, cada uno con ganancia de potencia λ_i y potencia transmitida E_s/n_t .

Por tanto, el uso de múltiples antenas en transmisión y recepción en un canal radio equivale a disponer de varios canales escalares (denominados “modas”) entre el transmisor y el receptor.

Una cuestión interesante es ver cuál debe ser la estructura del canal \mathbf{H} de modo que, dada una restricción de potencia sobre el mismo, $\|\mathbf{H}\|_F^2 = \sum_{i=1}^r \lambda_i = \xi$, se maximice la capacidad. Puede demostrarse que, si el número de antenas transmisoras es igual al de receptoras y el canal es de rango completo, $n_t = n_r = r$, el canal debe ser ortogonal, $\mathbf{H} \mathbf{H}^H = \mathbf{H}^H \mathbf{H} = (\xi/r) \mathbf{I}_r$, para maximizar la capacidad, que entonces puede expresarse como

$$C = r \log_2 \left(1 + \frac{\xi E_s}{N_0 r^2} \right) \quad (3.12)$$

Si además los elementos de \mathbf{H} cumplen tienen módulo unidad, $\|h_{ij}\|^2 = 1$, entonces $\|\mathbf{H}\|_F^2 = r^2$ y la capacidad es

$$C = r \log_2 \left(1 + \frac{E_s}{N_0} \right) \quad (3.13)$$

Por lo tanto, la capacidad de un canal MIMO ortogonal de igual número de antenas transmisoras y receptoras $r = n_t = n_r$ es r veces la de un canal AWGN (2.76).

3.2.2. Canal conocido por el transmisor

El conocimiento del canal por parte del transmisor puede conseguirse mediante realimentación (*feedback*) por parte del receptor, o mediante el concepto de reciprocidad si el sistema de comunicación es dúplex. Es por tanto importante ver cómo afecta a la capacidad el hecho de que el transmisor conozca el canal.

Consideremos que se transmite el vector $\tilde{\mathbf{s}}$ de dimensiones $r \times 1$, donde r es el rango de \mathbf{H} . Denotaremos de nuevo la descomposición en valores singulares de \mathbf{H} mediante $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Supondremos que el transmisor multiplica por \mathbf{V} el vector $\tilde{\mathbf{s}}$ antes de transmitirlo, y que el receptor multiplica el vector recibido por \mathbf{U}^H . Tenemos por tanto que la relación entrada-salida total es

$$\begin{aligned}\tilde{\mathbf{y}} &= \sqrt{\frac{E_s}{n_t}} \mathbf{U}^H \mathbf{H} \mathbf{V} \tilde{\mathbf{s}} + \mathbf{U}^H \mathbf{n} \\ &= \sqrt{\frac{E_s}{n_t}} \mathbf{\Sigma} \tilde{\mathbf{s}} + \tilde{\mathbf{n}}\end{aligned}\quad (3.14)$$

donde $\tilde{\mathbf{y}}$ es el vector de señales recibidas transformado, de dimensiones $r \times 1$, y $\tilde{\mathbf{n}}$ es un vector AWGN de dimensiones $r \times 1$ con matriz de autocorrelación $E\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\} = N_0 \mathbf{I}_r$. El vector $\tilde{\mathbf{s}}$ debe satisfacer $E\{\tilde{\mathbf{s}}\tilde{\mathbf{s}}^H\} = n_t$ para restringir la energía total transmitida. La ecuación anterior muestra cómo, conociendo el canal en transmisión, la matriz \mathbf{H} puede ser descompuesta explícitamente en r canales SISO paralelos satisfaciendo

$$\tilde{y}_i = \sqrt{\frac{E_s}{n_t}} \sqrt{\lambda_i} \tilde{s}_i + \tilde{n}_i \quad i = 1, 2, \dots, r \quad (3.15)$$

La capacidad del canal MIMO será la suma de las capacidades de esos r canales SISO paralelos

$$C = \sum_{i=1}^r \log_2 \det \left(1 + \frac{E_s \gamma_i}{n_t N_0} \lambda_i \right) \quad (3.16)$$

donde $\gamma_i = E\{|s_i|^2\}$ es la energía transmitida a través del i -ésimo subcanal. Estas energías deben cumplir $\sum_{i=1}^r \gamma_i = n_t$. Por tanto, la capacidad total depende de los valores γ_i . El transmisor debe utilizar los valores óptimos para maximizar la capacidad total del canal MIMO. Tenemos por tanto que la capacidad máxima es

$$C = \max_{\{\gamma_i\}_{i=1}^r: \sum_{i=1}^r \gamma_i = n_t} \sum_{i=1}^r \log_2 \det \left(1 + \frac{E_s \gamma_i}{n_t N_0} \lambda_i \right) \quad (3.17)$$

Al ser la expresión anterior cóncava en las variables γ_i , puede ser resuelta utilizando métodos lagrangianos. La política óptima de asignación de energía cumple

$$\gamma_i^{opt} = \left(\mu - \frac{n_t N_0}{E_s \lambda_i} \right)^+ \quad i = 1, 2, \dots, r \quad (3.18)$$

donde μ es una constante y

$$(x)^+ \triangleq \begin{cases} x & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad (3.19)$$

así como la restricción de energía antes comentada, $\sum_{i=1}^r \gamma_i^{opt} = n_t$. El cálculo de μ y los valores γ_i^{opt} se puede realizar mediante el algoritmo *waterpouring*, que describiremos brevemente a continuación.

Algoritmo *waterpouring*: repetir el siguiente proceso, donde p denota el número de iteración, hasta que ninguno de los valores γ_i^{opt} sea negativo:

- Cálculo de la constante μ :

$$\mu = \frac{n_t}{r-p+1} \left(1 + \frac{N_0}{E_s} \sum_{i=1}^{r-p+1} \frac{1}{\lambda_i} \right) \quad (3.20)$$

- Cálculo de los valores γ_i^{opt} :

$$\gamma_i^{opt} = \mu - \frac{n_t N_0}{E_s \lambda_i} \quad i = 1, 2, \dots, r-p+1 \quad (3.21)$$

- Si la energía asignada al canal con menor ganancia es negativa, $\gamma_{r-p+1} < 0$, se descarta esa asignación fijando $\gamma_{r-p+1}^{opt} = 0$.

La asignación de la energía utilizada en cada subcanal que produce el algoritmo *waterpouring* consiste en utilizar sólo algunas modas del canal, aquellas más favorables, dejando de transmitir por aquellas por debajo de cierto umbral de SNR, dado por la constante μ . De ahí que esta política de asignación también se denomine oportunista.

La capacidad del canal MIMO cuando el canal es conocido en transmisión es necesariamente igual o mayor que la capacidad cuando el canal no es conocido en transmisión. El conocimiento de γ_i^{opt} es equivalente al conocimiento de \mathbf{R}_{ss} en la ecuación (3.7). Esto es debido a que, teniendo en cuenta que $\mathbf{s} = \mathbf{V}\bar{\mathbf{s}}$, la matriz de autocovarianzas óptima está dada por

$$\mathbf{R}_{ss}^{opt} = \mathbf{V} \mathbf{R}_{\bar{s}\bar{s}}^{opt} \mathbf{V}^H \quad (3.22)$$

donde

$$\mathbf{R}_{\bar{s}\bar{s}}^{opt} = \text{diag}\{\gamma_1^{opt}, \gamma_2^{opt}, \dots, \gamma_r^{opt}\} \quad (3.23)$$

Por tanto, si el canal es conocido en transmisión, podemos maximizar la capacidad del canal asignando la potencia transmitida en cada moda del canal de acuerdo a \mathbf{R}_{ss}^{opt} .

Una vez analizada la capacidad del canal MIMO determinista plano en frecuencia, estamos preparados para analizar la capacidad del canal MIMO cuando los coeficientes de la matriz del canal son variables aleatorias. Como antes dijimos, estamos siguiendo un proceso similar al que utilizamos en el caso SISO: primero analizamos la capacidad del canal AWGN para después analizar la del canal con desvanecimiento.

3.3. Canal MIMO aleatorio plano en frecuencia

Consideraremos ahora que los elementos de \mathbf{H} son variables aleatorias, normalizadas de modo que $E\{|h_{ij}|^2\} = 1$. Cada realización del canal tendrá una capacidad asociada dada por la Ecuación (3.7).

Veamos primero cuál es el valor asintótico de la capacidad cuando incrementamos el número de antenas (tanto en transmisión como en recepción). Utilizaremos en adelante la notación \mathbf{H}_w para referirnos a un canal en el que los coeficientes son $\mathcal{CN}(0, 1)$ y además independientes entre sí. Para este canal y fijando $n_t = n_r = M$ tenemos que, por la Ley de los Grandes Números,

$$\frac{1}{M} \mathbf{H}_w \mathbf{H}_w^H \rightarrow \mathbf{I}_M \quad \text{si } M \rightarrow \infty \quad (3.24)$$

Por tanto, la capacidad de este canal MIMO cuando el transmisor desconoce el canal es, asintóticamente,

$$C^\infty \rightarrow M \log_2(1 + \rho) \quad (3.25)$$

Por tanto, la capacidad asintótica (en el número de antenas) del canal MIMO espacialmente blanco se convierte en determinista y crece de modo lineal con M . Por cada 3 dB de incremento en la SNR, se obtiene un incremento de M bit/s/Hz de la capacidad, en contraste con el 1 bit/s/Hz de incremento que se obtiene en el canal SISO.

En general, estaremos interesados en sistemas con un número de antenas no muy elevado, por lo que nos interesa estudiar la capacidad cuando el número de antenas es reducido. Al ser el canal \mathbf{H} aleatorio, su capacidad es también una variable aleatoria, por lo que necesitamos conocer su función de distribución para especificarla por completo. Sin embargo, en la práctica nos interesa comparar canales distintos por medio de parámetros sencillos, así que habitualmente se utilizan dos definiciones de la capacidad de un canal aleatorio: una, dada por su media, denominada *capacidad ergódica*, y otra, por un percentil p (%) determinado (el valor de C para el cual su CDF es igual a p (%)), denominada *capacidad fuera de servicio (outage)*.

3.3.1. Capacidad ergódica

La capacidad ergódica del canal es simplemente la media de la capacidad del canal

$$\bar{C} = E \left\{ \max_{\mathbf{R}_{ss}: \text{Tr}\{\mathbf{R}_{ss}\} = n_t} \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{n_t} \mathbf{H} \mathbf{R}_{ss} \mathbf{H}^H \right) \right\} \quad \text{bit/s/Hz} \quad (3.26)$$

que, particularizada a los casos de canal desconocido y conocido por el transmisor, respectivamente, es

$$\bar{C} = E \left\{ \sum_{i=1}^r \log_2 \left(1 + \frac{\rho}{n_t} \lambda_i \right) \right\} \quad (3.27)$$

y

$$\bar{C} = E \left\{ \sum_{i=1}^r \log_2 \left(1 + \frac{\rho}{n_t} \gamma_i^{opt} \lambda_i \right) \right\} \quad (3.28)$$

En [84] puede verse cómo la capacidad ergódica de un canal \mathbf{H}_w cuando el canal es conocido por el transmisor es siempre mayor que cuando lo desconoce. De todos modos, esta diferencia sólo es significativa para SNRs bajas. Esto es debido a que cuando la SNR es alta, la capacidad del canal MIMO puede aproximarse por

$$C \approx \max_{\mathbf{R}_{ss}: Tr\{\mathbf{R}_{ss}\}=n_t} \log_2 \det(\mathbf{R}_{ss}) + \log_2 \det\left(\frac{\rho}{M} \mathbf{H}_w \mathbf{H}_w^H\right) \quad (3.29)$$

Cuando el canal es conocido en transmisión, el término $\det(\mathbf{R}_{ss})$ es maximizado, sujeto a la restricción $Tr\{\mathbf{R}_{ss}\} = M$, cuando $\mathbf{R}_{ss} = \mathbf{I}_M$, que es también la matriz de autocovarianzas óptima en el caso de canal desconocido. Por tanto, la solución del algoritmo *waterpouring* se aproxima a la estrategia uniforme de asignación de potencia a medida que crece la SNR para el canal MIMO espacialmente blanco con $n_t = n_r = M$. La diferencia en capacidad entre canal conocido y desconocido incrementa con M , pero se reduce para SNRs suficientemente altas.

3.3.2. Capacidad fuera de servicio (*outage*)

En este caso la capacidad se calcula simplemente utilizando la CDF de la capacidad del canal MIMO. Puede verse en [84] que la capacidad fuera de servicio al 10 % de un canal \mathbf{H}_w crece con el número de antenas utilizadas. Esto es coherente con el comportamiento asintótico de la capacidad que comentamos anteriormente.

Recordemos que la capacidad de fuera-de-servicio es útil para caracterizar un canal que es desconocido en transmisión y aleatorio, pero que se mantiene constante a lo largo de un bloque de símbolos. En este caso, la capacidad de Shannon del canal, en el sentido de la máxima tasa de información que permite una comunicación libre de errores, es cero. Esto es debido a que, sea cual sea la tasa de información utilizada en transmisión, al ser ésta constante, siempre habrá bloques de datos para los cuales la capacidad asociada a la realización del canal en ese bloque sea menor que la tasa de información utilizada, y por tanto, la probabilidad de error distinta de cero.

Tal y como ocurre con la capacidad del canal determinista, el conocimiento del canal por parte del transmisor mejora la capacidad fuera de servicio.

Por último, comentaremos el efecto que tiene sobre la capacidad la correlación espacial de los coeficientes del canal. El efecto sobre la capacidad de otras características del canal, como por ejemplo la existencia de una componente determinista (desvanecimiento Rice), puede consultarse en [84].

3.3.3. Influencia de la correlación espacial en la capacidad

El hecho de que los coeficientes del canal presenten correlaciones espaciales puede modelarse convenientemente escribiendo la matriz del canal como

$$\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2} \quad (3.30)$$

donde las matrices \mathbf{R}_r y \mathbf{R}_t son matrices hermíticas semidefinidas positivas que especifican la correlación espacial en transmisión y recepción, respectivamente. Su relación con la matriz de autocorrelación del canal

$$\mathbf{R} = E\{\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H\} \quad (3.31)$$

viene dada por

$$\mathbf{R} = \mathbf{R}_t^T \otimes \mathbf{R}_r \quad (3.32)$$

donde \otimes denota el producto de Kronecker. Nótese que \mathbf{H}_w es de rango completo con probabilidad 1, mientras que cuando existe correlación espacial, el rango de \mathbf{H} está restringido por $\min(r(\mathbf{R}_r), r(\mathbf{R}_t))$, donde $r(\mathbf{A})$ denota el rango de \mathbf{A} . Asumiremos además que \mathbf{R}_r y \mathbf{R}_t están normalizadas de modo que $[\mathbf{R}_r]_{ii} = 1, i = 1, 2, \dots, n_r$, y $[\mathbf{R}_t]_{jj} = 1, j = 1, 2, \dots, n_t$, por lo que $E\{|h_{ij}|^2\} = 1$. Sustituyendo (3.30) en la expresión de la capacidad del canal MIMO (3.7) obtenemos

$$C = \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{n_t} \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t \mathbf{H}_w^H \mathbf{R}_r^{H/2} \right) \quad (3.33)$$

Supondremos además que $n_t = n_r = M$ y que las matrices de correlación espacial en transmisión y recepción son de rango completo. Para SNRs suficientemente altas podemos aproximar la expresión anterior por

$$C \approx \log_2 \det \left(\frac{\rho}{M} \mathbf{H}_w \mathbf{H}_w^H \right) + \log_2 \det(\mathbf{R}_r) + \log_2 \det(\mathbf{R}_t) \quad (3.34)$$

A partir de la expresión anterior se ve que el impacto de \mathbf{R}_r y \mathbf{R}_t en la capacidad del canal MIMO es idéntico. Examinemos ahora las condiciones sobre \mathbf{R}_r (sobre \mathbf{R}_t serán análogas) que maximizan la capacidad. Los autovalores de \mathbf{R}_r , $\lambda_i(\mathbf{R}_r), i = 1, 2, \dots, M$ cumplen la restricción $\sum_{i=1}^M \lambda_i(\mathbf{R}_r) = M$. Utilizando la desigualdad entre la media aritmética y geométrica obtenemos

$$\prod_{i=1}^M \lambda_i(\mathbf{R}_r) \leq 1 \quad (3.35)$$

Ahora bien, $\prod_{i=1}^M \lambda_i(\mathbf{R}_r) = \det(\mathbf{R}_r)$, por lo que $\log_2 \det(\mathbf{R}_r) \leq 0$. Además, la igualdad se alcanza cuando todos los autovalores de \mathbf{R}_r son iguales a la unidad, lo que implica que $\mathbf{R}_r = \mathbf{I}_M$. Por lo tanto, podemos concluir que, para SNRs suficientemente altas, la existencia de correlaciones espaciales entre los coeficientes del canal siempre supone un detrimento sobre la capacidad del canal MIMO. La pérdida en capacidad con respecto al caso de no existir correlación espacial viene dada precisamente por el término $\log_2 \det(\mathbf{R}_r) + \log_2 \det(\mathbf{R}_t)$.

En el caso extremo, cuando alguna de las matrices de correlación espacial es de rango unidad, el canal \mathbf{H} también es de rango unidad. El canal tiene entonces una sola moda, siendo equivalente a un canal SISO y el incremento en capacidad tan sólo logarítmico en $n_t n_r$ en lugar de lineal en $\min(n_t, n_r)$ como ocurre en el canal \mathbf{H}_w [91].

3.4. Canal MIMO selectivo en frecuencia

Calcularemos la capacidad del canal MIMO selectivo en frecuencia como la suma de las capacidades de N canales MIMO planos en frecuencia. Para ello, simplemente dividiremos el rango de frecuencias de interés, que consideraremos de 1 Hz como hemos hecho hasta ahora, en N subrangos sobre los cuales podemos considerar el desvanecimiento del canal como plano. Sea el i -ésimo subcanal \mathbf{H}_i . La relación entrada-salida de este subcanal es

$$\mathbf{y}_i = \sqrt{\frac{E_s}{n_t}} \mathbf{H}_i \mathbf{s}_i + \mathbf{n}_i \quad (3.36)$$

donde \mathbf{y}_i , de dimensiones $n_r \times 1$, es el vector de señales recibidas, \mathbf{s}_i , de dimensiones $n_t \times 1$, es el vector de señales transmitidas y \mathbf{n}_i es el vector de ruido, de dimensiones $n_r \times 1$. La relación entrada-salida global se puede escribir entonces como

$$\mathcal{Y} = \sqrt{E_s n_t} \mathcal{H} \mathcal{S} + \mathcal{N} \quad (3.37)$$

donde $\mathcal{Y} = [\mathbf{y}_1^T \mathbf{y}_2^T \cdots \mathbf{y}_N^T]^T$ tiene dimensiones $n_r N \times 1$, $\mathcal{S} = [\mathbf{s}_1^T \mathbf{s}_2^T \cdots \mathbf{s}_N^T]^T$ tiene dimensiones $n_t N \times 1$, $\mathcal{N} = [\mathbf{n}_1^T \mathbf{n}_2^T \cdots \mathbf{n}_N^T]^T$ tiene dimensiones $n_r N \times 1$ y \mathcal{H} es una matriz diagonal a bloques de tamaño $n_r N \times n_t N$, siendo las matrices \mathbf{H}_i los bloques que componen su diagonal. A la matriz de covarianzas de \mathcal{S} , $\mathbf{R}_{\mathcal{S}\mathcal{S}} = E\{\mathcal{S}\mathcal{S}^H\}$ se le impone la restricción habitual de potencia $Tr\{\mathbf{R}_{\mathcal{S}\mathcal{S}}\} = N n_t$, de modo que la energía total transmitida sea E_s . Sustituyendo estos valores en la expresión de la capacidad del canal MIMO obtenemos

$$C = \frac{1}{N} \max_{\mathbf{R}_{\mathcal{S}\mathcal{S}}: Tr\{\mathbf{R}_{\mathcal{S}\mathcal{S}}\} = N n_t} \log_2 \det \left(\mathbf{I}_{n_r N} + \frac{E_s}{n_t N_0} \mathcal{H} \mathbf{R}_{\mathcal{S}\mathcal{S}} \mathcal{H}^H \right) \quad (3.38)$$

Analizamos ahora, como hicimos en el caso del canal plano en frecuencia, los casos de canal desconocido y conocido por el transmisor.

3.4.1. Canal desconocido por el transmisor

Al igual que en el caso del canal plano en frecuencia, la matriz de autocovarianzas de los símbolos transmitidos que maximiza la capacidad es $\mathbf{R}_{\mathcal{S}\mathcal{S}} = \mathbf{I}_{n_t N}$. Esto equivale a distribuir uniformemente la potencia transmitida a lo largo de las antenas transmisoras y de los distintos N subcanales en frecuencia. La capacidad resultante es

$$C = \frac{1}{N} \sum_{i=1}^N \log_2 \det \left(\mathbf{I}_{n_r} + \frac{E_s}{n_t N_0} \mathbf{H}_i \mathbf{H}_i^H \right) \quad (3.39)$$

Evidentemente, si la respuesta en frecuencia del canal es plana, $\mathbf{H}_i = \mathbf{H}$, la expresión anterior se reduce a la de un canal plano en frecuencia (3.8). Alternativamente, podemos utilizar la expresión de la capacidad en función de los autovalores de $\mathbf{H}_i \mathbf{H}_i^H$ (3.11)

$$C = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^r \log_2 \left(1 + \frac{E_s}{n_t N_0} \lambda_j(\mathbf{H}_i) \right) \quad (3.40)$$

Utilizando lo que vimos en el caso de canal plano en frecuencia, la capacidad del canal H_i se maximiza cuando esta matriz es ortogonal, por lo que la capacidad máxima del canal selectivo en frecuencia es la misma que la del canal no selectivo en frecuencia.

Analicemos ahora la capacidad cuando el canal es aleatorio. Primero un comentario sobre su comportamiento asintótico en N . Si el ancho de banda de coherencia del canal es $1/N$, todas las matrices H_i serán independientes y tendrán la misma distribución y, por la Ley de los Grandes Números, $C \rightarrow C^\infty$ si $N \rightarrow \infty$, es decir, que la capacidad de cualquier realización del canal tiende a un valor fijo. Por otra parte, la capacidad ergódica del canal MIMO selectivo en frecuencia vendrá dada simplemente por la aplicación del operador esperanza sobre cualquiera de las expresiones (3.39) y (3.40). De modo similar podemos se puede calcular la capacidad de fuera-de-servicio. La capacidad de fuera-de-servicio del canal selectivo en frecuencia es mayor que la del canal con desvanecimiento plano para valores pequeños de la probabilidad de fuera-de-servicio, debido a que la diversidad espectral se traduce en una CDF de la capacidad con una pendiente más brusca. Como hemos comentado, asintóticamente en N , la capacidad de fuera-de-servicio para cualquier realización del canal tiende a un valor fijo. En otras palabras, la capacidad deja de ser aleatoria. Se puede demostrar, que bajo condiciones de regularidad suficientes, el valor al que tiende es precisamente la capacidad ergódica.

3.4.2. Canal conocido por el transmisor

Anteriormente vimos cómo el algoritmo de *waterpouring* nos permite obtener la distribución de energía óptima para el canal con desvanecimiento plano. Lo mismo se puede decir para el canal selectivo en frecuencia, con la salvedad de que en este caso la potencia se ha de distribuir tanto en espacio como en frecuencia. Esta forma de *waterpouring* se conoce como *waterpouring* espacio-espectral [11]. Nótese que el *waterpouring* sólo está definido para canales ortogonales. Para ello podemos utilizar OFDM (*Orthogonal Frequency Division Multiplexing*) para ortogonalizar la dispersión del retardo, en otras palabras, la ISI del canal, y la descomposición en modas para ortogonalizar el canal MIMO.

Si el canal es conocido tanto en transmisión como en recepción, el canal global H puede descomponerse en $r = \text{rango}(H)$ modas espacio-espectrales. La capacidad del canal vendrá dada por

$$C = \frac{1}{N} \max_{\{\gamma_i\}_{i=1}^r: \sum_{i=1}^r \gamma_i = n_t} \sum_{i=1}^r \log_2 \det \left(1 + \frac{E_s \gamma_i}{n_t N_0} \lambda_i(\mathcal{H}\mathcal{H}^H) \right) \quad (3.41)$$

donde $\lambda_i(\mathcal{H}\mathcal{H}^H)$, $i = 1, 2, \dots, r$, son los autovalores positivos de la matriz $\mathcal{H}\mathcal{H}^H$ y γ_i es la energía utilizada en la i -ésima moda espacio-espectral.

Del mismo modo que hicimos para el caso de canal desconocido para el transmisor, se pueden definir las capacidades ergódicas y de fuera-de-servicio cuando el canal es conocido en transmisión.

Después de exponer las expresiones de la capacidad del canal de los casos más importantes,

pasamos a estudiar cómo influye la diversidad obtenida por el uso de múltiples antenas sobre la probabilidad de error.

3.5. Diversidad en sistemas MIMO

La diversidad que se puede obtener en un sistema MIMO proviene del mismo hecho en el que se apoya el concepto de diversidad en sistemas de una sola antena: la independencia entre distintas realizaciones del canal. Como vimos entonces, la independencia entre distintas realizaciones del canal viene dictada por el ancho de banda de coherencia y el tiempo de coherencia. Del mismo modo, tenemos una distancia de coherencia en el dominio espacial.

Antes de proseguir, es justo enfatizar el hecho de que el uso de múltiples antenas no produce ningún efecto *mágico* sobre los sistemas de comunicación. Hemos visto que la capacidad del canal MIMO aumenta con el número de antenas utilizadas, pero del mismo modo la capacidad del canal Rayleigh se define por unidad de ancho de banda, y evidentemente si aumentamos el ancho de banda utilizado disponemos de mayor capacidad. En el fondo, lo único que ocurre es que, además de las dimensiones tiempo y ancho de banda, disponemos de una dimensión adicional, la espacial, sobre la que aplicar técnicas de codificación, diversidad, etc.

Sin embargo, desde un punto de vista práctico, la tres dimensiones no son igual de importantes, o en otras palabras, no tienen el mismo coste. La dimensión espectral quizá es la que mayor coste tiene. El espectro radioeléctrico es limitado y la reserva de un rango de frecuencias determinado para una aplicación tiene un coste muy elevado. En segundo lugar, la dimensión tiempo es importante en aplicaciones que necesitan un tiempo de respuesta rápido. En cuanto a la dimensión espacial, el único coste es el del *hardware* asociado a cada elemento receptor... De ahí deriva la importancia de los sistemas MIMO, en que hacen uso efectivo de una dimensión cuyo coste es mucho inferior a las de ancho de banda y tiempo.

Hay además otro aspecto muy importante, que esencialmente viene dado por el hecho de que en cada antena receptora recibimos toda la potencia que transmitimos (atenuada por el factor correspondiente). Recordemos que cuando utilizamos técnicas de diversidad en sistemas SISO la energía transmitida ha de distribuirse a lo largo de las distintas ramas de diversidad disponibles. Considérese lo que pasa en un sistema SIMO: la energía transmitida correspondiente a un símbolo se recibe en cada antena receptora. En este caso cada antena receptora actúa como una rama de diversidad. Este incremento en la SNR media recibida, fruto de utilizar varias antenas, es lo que se denomina ganancia de *array*. Por tanto, en los sistemas con múltiples antenas, además de poder explotar el concepto de diversidad, disponemos de una ganancia a ningún coste, la ganancia de *array*.

Una exposición didáctica de la diversidad espacial requiere estudiar primero los casos SIMO y MISO. Dicha exposición puede encontrarse en [84]. Nosotros pasaremos a analizar directamente el caso MIMO, distinguiendo los casos de canal desconocido por el transmisor y canal conocido en transmisión, del mismo modo que hicimos al estudiar la capacidad del canal.

3.5.1. Canal desconocido por el transmisor

Explicaremos aquí una técnica sencilla para explotar la diversidad inherente al canal MIMO cuando el transmisor no conoce el canal, conocida como esquema de Alamouti. La técnica consiste simplemente en, dados dos símbolos a transmitir s_1 y s_2 , se transmite en el primer período de símbolo el vector $\mathbf{s} = [s_1 \ s_2]^T$, y en el segundo período de símbolo el vector $[-s_2^* \ s_1^*]^T$. La tasa de información que estamos utilizando es $1/2$. Estamos asumiendo que el canal no cambia durante la transmisión de esos dos vectores. Los vectores de señales recibidas en el primer y segundo período símbolo son

$$\mathbf{y}_1 = \sqrt{\frac{E_s}{2}} \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} \quad (3.42)$$

$$\mathbf{y}_2 = \sqrt{\frac{E_s}{2}} \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} -s_2^* \\ s_1^* \end{pmatrix} + \begin{pmatrix} n_3 \\ n_4 \end{pmatrix} \quad (3.43)$$

donde $n_i, i = 1, 2, 3, 4$, son muestras de ruido $\mathcal{CN}(0, N_0)$. El receptor ahora construye el vector

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2^* \end{pmatrix} \quad (3.44)$$

Utilizando las propiedades de los número complejos, podemos expresar el vector anterior como

$$\mathbf{y} = \sqrt{\frac{E_s}{2}} \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ h_{12}^* & -h_{11}^* \\ h_{22}^* & -h_{21}^* \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \\ n_3^* \\ n_4^* \end{pmatrix} \quad (3.45)$$

$$\triangleq \sqrt{\frac{E_s}{2}} \mathbf{H}_{ef} \mathbf{s} + \mathbf{n} \quad (3.46)$$

Además, la matriz \mathbf{H}_{ef} es ortogonal, $\mathbf{H}_{ef}^H \mathbf{H}_{ef} = \|\mathbf{H}\|_F^2 \mathbf{I}_2$, independientemente de la realización del canal. Si ahora multiplicamos el vector \mathbf{y} por \mathbf{H}_{ef}^H obtenemos

$$\mathbf{z} \triangleq \mathbf{H}_{ef}^H \mathbf{y} \quad (3.47)$$

$$= \sqrt{\frac{E_s}{2}} \|\mathbf{H}\|_F^2 \mathbf{I}_2 \mathbf{s} + \tilde{\mathbf{n}} \quad (3.48)$$

donde $E\{\tilde{\mathbf{n}}\} = \mathbf{0}_{21}$ y $E\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\} = \|\mathbf{H}_F\|^2 N_0 \mathbf{I}_2$. Por tanto, el canal efectivo para cada uno de los símbolos transmitidos es

$$z_i = \sqrt{\frac{E_s}{2}} \|\mathbf{H}\|_F^2 s_i + \tilde{n}_i \quad i = 1, 2 \quad (3.49)$$

y la SNR correspondiente es

$$\eta = \frac{\|\mathbf{H}_F\|^2 \rho}{2} \quad (3.50)$$

En el caso de que el canal sea \mathbf{H}_w , la probabilidad de error de símbolo media está acotada por

$$\bar{P}(e) \leq \bar{N}_e \left(\frac{\rho d_{min}^2}{8} \right)^{-4} \quad (3.51)$$

Por tanto, podemos decir que el esquema de Alamouti extrae un orden de diversidad igual a $n_t n_r$, cuando el canal es desconocido en transmisión. Como $E\{\|\mathbf{H}_w\|^2\} = 4$, la SNR media en recepción es

$$\bar{\eta} = 2\rho \quad (3.52)$$

por lo que el esquema de Alamouti en este caso obtiene una ganancia de array igual al número de antenas receptoras. Es decir, sólo obtiene la mitad de la ganancia de array potencial, $n_t n_r$.

El esquema de Alamouti se puede utilizar en configuraciones de dos antenas transmisoras y un número arbitrario de antenas receptoras, obteniendo siempre diversidad completa (ganancia por diversidad igual a $n_t n_r = 2n_r$) y una ganancia de array igual a n_r . Se pueden construir esquemas análogos al de Alamouti para un mayor número de antenas transmisoras. Estos esquemas se denominan OSTBC (*Orthogonal Space-Time Block Codes*).

3.5.2. Canal conocido por el transmisor

Cuando el canal es conocido en transmisión, la diversidad espacial puede aprovecharse utilizando una técnica denominada transmisión de automodas dominantes. El mismo símbolo se transmite por cada antena multiplicado por un factor distinto. La señal recibida puede escribirse como

$$\mathbf{y} = \sqrt{\frac{E_s}{n_t}} \mathbf{H} \mathbf{w} s + \mathbf{n} \quad (3.53)$$

donde \mathbf{w} es un vector de dimensiones $n_t \times 1$ que pondera la energía transmitida por cada antena. El vector \mathbf{w} debe cumplir $\|\mathbf{w}\|_F^2 = n_t$ para mantener constante la energía transmitida. En recepción se combinan linealmente las muestras obtenidas

$$z \triangleq \mathbf{g}^H \mathbf{y} \quad (3.54)$$

donde \mathbf{g} es un vector de pesos complejos de dimensiones $n_r \times 1$. La SNR en z es

$$\eta = \frac{\|\mathbf{g}^H \mathbf{H} \mathbf{w}\|_F^2}{n_t \|\mathbf{g}\|_F^2} \rho \quad (3.55)$$

por lo que la maximizar la SNR es equivalente a maximizar $\|\mathbf{g}^H \mathbf{H} \mathbf{w}\|_F^2 / \|\mathbf{g}\|_F^2$. Utilizando la descomposición en valores singulares $\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$, puede verse que η es máxima cuando $\mathbf{w} / \sqrt{n_t}$ y \mathbf{g} son iguales al vector singular entrada y salida, respectivamente, correspondientes al máximo valor singular σ_{max} de \mathbf{H} . En este caso podemos escribir la relación entrada-salida efectiva del canal como

$$z = \sqrt{E_s} \sigma_{max} s + n \quad (3.56)$$

donde n es ruido $\mathcal{CN}(0, N_0)$. La SNR efectiva es

$$\eta = \sigma_{max}^2 \rho = \lambda_{max} \rho \quad (3.57)$$

donde λ_{max} es el autovalor positivo máximo de $\mathbf{H}\mathbf{H}^H$ y en la segunda igualdad se ha utilizado el hecho de que $\lambda_i = \sigma_i^2$, $i = 1, 2, \dots, r$. Por tanto, la ganancia de array en el esquema de transmisión automoda dominante es $E\{\lambda_{max}\}$. Podemos acotar el valor de λ_{max} teniendo en cuenta que $\|\mathbf{H}\|_F^2 = \text{Tr}\{\mathbf{H}\mathbf{H}^H\} = \sum_{i=1}^r \lambda_i$

$$\frac{\|\mathbf{H}\|_F^2}{r} \leq \lambda_{max} \leq \|\mathbf{H}\|_F^2 \quad (3.58)$$

donde $r = \text{rango}(\mathbf{H})$. Por tanto, la SNR media, $\bar{\eta} = E\{\lambda_{max}\}\rho$ puede ser acotada del siguiente modo

$$\frac{E\{\|\mathbf{H}\|_F^2\}}{r} \rho \leq \bar{\eta} \leq E\{\|\mathbf{H}\|_F^2\} \rho \quad (3.59)$$

$$\frac{n_t n_r}{r} \rho \leq \bar{\eta} \leq n_t n_r \rho \quad (3.60)$$

En el caso de $\mathbf{H} = \mathbf{H}_w$, $r = \min(n_r, n_t)$, que sustituyéndolo en la ecuación anterior resulta

$$\max(n_t, n_r) \rho \leq \bar{\eta} \leq n_t n_r \rho \quad (3.61)$$

La ecuación anterior nos muestra cómo la ganancia de array cuando el canal es conocido por el transmisor es siempre mayor o igual que cuando el transmisor desconoce lo desconoce.

Utilizando una derivación de la expresión de la probabilidad de error de símbolo media similar a la explicada anteriormente, es decir, utilizando la cota de unión y la de Chernoff, y aplicando la expresión (3.61) se obtiene, para SNRs altas,

$$\bar{n}_{min} \left(\frac{\rho d_{min}^2}{4} \right)^{-n_t n_r} \leq \bar{P}(e) \leq \bar{n}_{min} \left(\frac{\rho d_{min}^2}{4 \min(n_t, n_r)} \right)^{-n_t n_r} \quad (3.62)$$

De la ecuación anterior podemos concluir que la curvatura de $\bar{P}(e)$ en función de la SNR tiene una magnitud $n_t n_r$ (en escala logarítmica). Por tanto, decimos que la técnica de transmisión de automoda dominante extrae un orden de diversidad completo, igual a $n_t n_r$.

También el esquema de Alamouti presentado anteriormente extraía diversidad completa, pero la ventaja del esquema de transmisión de automoda dominante es que su ganancia de array es como mínimo igual a la del esquema de Alamouti, pudiendo ser mayor, incluso hasta el límite de ganancia de array completa, $n_t n_r$.

3.5.3. Extensión a canales MIMO selectivos en espacio, frecuencia y tiempo

Hasta ahora hemos considerado que el canal MIMO era plano en frecuencia y no consideramos la posibilidad de aprovechar la diversidad en la dimensión tiempo. La diversidad

disponible en el canal depende de las dimensiones de las palabras código que utilicemos y de los parámetros de coherencia. Las dimensiones de las palabras código son: número de antenas transmisoras, n_t , número de antenas receptoras, n_r , duración, T , y ancho de banda, B . Los parámetros de coherencia del canal son el tiempo de coherencia, T_c , el ancho de banda de coherencia, B_c y la distancia de coherencia, D_c . El número de ramas de diversidad independientes (la diversidad disponible) que puede aprovechar una palabra código en la dimensión tiempo es T/T_c . Análogamente, para la dimensión frecuencia es B/B_c y para la espacial $n_t n_r$, suponiendo que las antenas transmisoras y receptoras están colocadas convenientemente para experimentar desvanecimientos independientes. Por tanto la diversidad disponible es

$$\text{diversidad disponible} = \frac{T}{T_c} \frac{B}{B_c} n_t n_r \quad (3.63)$$

El transmisor y el receptor deben diseñarse para poder extraer la máxima diversidad, ya que es el factor determinante tanto de la capacidad del canal como de la probabilidad de error.

3.6. Codificación espacio-temporal

Consideremos la arquitectura de codificación de la Figura 3.2. Un bloque de qK bits es la entrada a un codificador que realiza las funciones de codificación temporal, entrelazado y asignación (*mapping*) bit a símbolo. El orden de la modulación utilizada lo denotamos por q . En el proceso se generan $q(N - K)$ bits de paridad y se producen N símbolos como salida. Esos N símbolos son la entrada a un codificador espacio-temporal que produce $n_t T - N$ símbolos de paridad y empaqueta los $n_t T$ símbolos resultantes en una matriz de dimensiones $n_t \times T$, que denominaremos palabra código, donde T es el número de períodos de símbolo utilizados para transmitir la palabra código completa. La tasa de información a través del canal es qK/T bits por uso del canal, valor que no debe exceder la capacidad del canal si queremos una transmisión libre de errores. Podemos escribir la tasa de información haciendo explícitas las tasas de codificación de los dos codificadores involucrados

$$\frac{qK}{T} = q r_t r_s \quad (3.64)$$

donde $r_t = K/N$ es la tasa de codificación del código temporal (externo) y $r_s = N/T$ es la tasa de codificación del código espacio-temporal (interno), definida en este caso como el número medio de símbolos transmitidos por cada uso del canal. Por ejemplo, si todas las antenas transmiten el mismo símbolo, $r_s = 1$. Si no se generan símbolos de paridad y simplemente se distribuyen los símbolos a la entrada a lo largo de las antenas transmisoras, $r_s = n_t$.

El entrelazado se utiliza para dispersar los errores en ráfaga debidos al desvanecimiento a lo largo de las palabras código para mejorar el rendimiento de los algoritmos de decodificación. Además del desvanecimiento selectivo en tiempo y en frecuencia, el entrelazado debe considerar la selectividad en la dimensión espacial. Los códigos espacio-temporales que

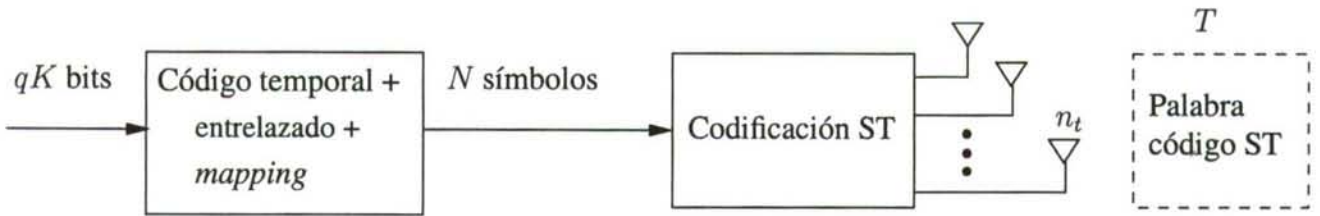


Figura 3.2: Modelo de sistema MIMO con codificación de canal

explicaremos tienen una forma inherente de entrelazado espacial que les permite extraer diversidad espacial completa, $n_t n_r$.

En adelante supondremos que el canal es plano en frecuencia. La extensión de los conceptos que vamos a presentar a canales selectivos en frecuencia puede encontrarse en la literatura [84].

Sea $\mathbf{S} = [s[1]s[2] \cdots s[T]]$ una palabra código, donde $s[k]$ es el vector de símbolos transmitidos en el k -ésimo período de símbolo. Consideraremos que el canal permanece constante a lo largo de la transmisión de cada palabra código (desvanecimiento cuasi-estático). La palabra código recibida la escribiremos como

$$\mathbf{Y} = \sqrt{\frac{E_s}{n_t}} \mathbf{H} \mathbf{S} + \mathbf{N} \quad (3.65)$$

donde $\mathbf{Y} = [y[1]y[2] \cdots y[T]]$ y $\mathbf{N} = [n[1]n[2] \cdots n[T]]$ son matrices de dimensiones $n_r \times T$. La regla de decisión óptima en el sentido MV viene dada por

$$\hat{\mathbf{S}} = \arg \min_{\mathbf{S}} \|\mathbf{Y} - \mathbf{H} \mathbf{S}\|_F^2 \quad (3.66)$$

$$= \arg \min_{\mathbf{S}} \sum_{k=1}^T \|y[k] - \sqrt{\frac{E_s}{n_t}} \mathbf{H} s[k]\|_F^2 \quad (3.67)$$

Utilizaremos un razonamiento similar al empleado previamente en otros casos, consistente en asumir SNR suficientemente alta y aplicar la cota de unión para podernos centrar exclusivamente en la probabilidad de error dos a dos

$$\begin{aligned} P(\mathbf{S}_i \rightarrow \mathbf{S}_j) &= Q \left(\sqrt{\frac{E_s \|\mathbf{H}(\mathbf{S}_i - \mathbf{S}_j)\|_F^2}{2n_t N_0}} \right) \\ &= Q \left(\sqrt{\frac{\rho \|\mathbf{H} \mathbf{E}_{ij}\|_F^2}{2n_t}} \right) \end{aligned} \quad (3.68)$$

donde estamos suponiendo que el receptor conoce el canal y $\mathbf{E}_{ij} \triangleq \mathbf{S}_i - \mathbf{S}_j$ es la matriz diferencia de las palabras código i y j , de dimensiones $n_t \times T$, y $\rho = E_s/N_0$ es la SNR. Aplicando la cota de Chernoff obtenemos

$$P(\mathbf{S}_i \rightarrow \mathbf{S}_j) \leq e^{-\frac{\rho \|\mathbf{H} \mathbf{E}_{ij}\|_F^2}{4n_t}} \quad (3.69)$$

Esta es la probabilidad de error dos a dos para una realización concreta del canal. Su media en el caso $\mathbf{H} = \mathbf{H}_w$ es

$$\bar{P}(\mathbf{S}_i \rightarrow \mathbf{S}_j) \leq \left(\frac{1}{\det \left(\mathbf{I}_{n_t} + \frac{\rho}{4n_t} \mathbf{E}_{ij} \mathbf{E}_{ij}^H \right)} \right)^{n_r} \quad (3.70)$$

que, utilizando la descomposición en autovalores de $\mathbf{G}_{ij} \triangleq \mathbf{E}_{ij} \mathbf{E}_{ij}^H$, podemos escribir como

$$\bar{P}(\mathbf{S}_i \rightarrow \mathbf{S}_j) \leq \prod_{k=1}^{r(\mathbf{G}_{ij})} \left(\frac{1}{1 + \rho \lambda_k(\mathbf{G}_{ij}) / (4n_t)} \right)^{n_r} \quad (3.71)$$

que, cuando la SNR es muy alta ($\rho \gg 1$), podemos simplificar por

$$\bar{P}(\mathbf{S}_i \rightarrow \mathbf{S}_j) \leq \frac{1}{\left(\prod_{k=1}^{r(\mathbf{G}_{ij})} \lambda_k(\mathbf{G}_{ij}) \right)^{n_r}} \left(\frac{\rho}{4n_t} \right)^{-r(\mathbf{G}_{ij})n_r} \quad (3.72)$$

La expresión anterior es la base de los criterios *del rango* y del *determinante* para diseñar códigos ST para el canal cuasi-estático cuando el transmisor no conoce el canal.

- Criterio del rango: busca la optimización de la diversidad espacial extraída por el código ST. El diseño del código debe ser tal que la matriz $\mathbf{G}_{ij} \triangleq (\mathbf{S}_i - \mathbf{S}_j)(\mathbf{S}_i - \mathbf{S}_j)^H$ sea de rango completo, $r(\mathbf{G}_{ij}) = n_t$.
- Criterio del determinante: busca la optimización de la ganancia por codificación del código ST. El diseño del código debe ser tal que $\det(\mathbf{G}_{ij})$ sea máximo.

Como ya hemos explicado en otras ocasiones, la contribución más importante en la expresión de la probabilidad de error dos a dos media viene dada por el exponente de la misma, que depende del rango de \mathbf{G}_{ij} . Es por ello por lo que cualquier código ST, en el régimen de SNRs altas, debe tener una matriz \mathbf{G}_{ij} de rango completo, o en otras palabras, cumplir el criterio del rango. Sólo una vez garantizada esa condición, el diseño del código puede centrarse en maximizar la ganancia de codificación (criterio del determinante).

Los criterios del rango y del determinante han sido extensivamente aplicados para la construcción de códigos espacio-temporales. Consideraremos aquí códigos ST *trellis* y códigos ST bloque ortogonales. Adicionalmente, estudiaremos el caso de que no se añada ninguna redundancia espacial, denominado multiplexación espacial. Finalmente, realizaremos algunos comentarios sobre las ventajas e inconvenientes de los distintos esquemas de codificación espacio-temporal.

3.6.1. Códigos ST *trellis* (STTC)

Los códigos ST *trellis* son la extensión natural de TCM a canales MIMO. La salida del codificador en cada período de símbolo está determinada por la etiqueta de la rama del trellis

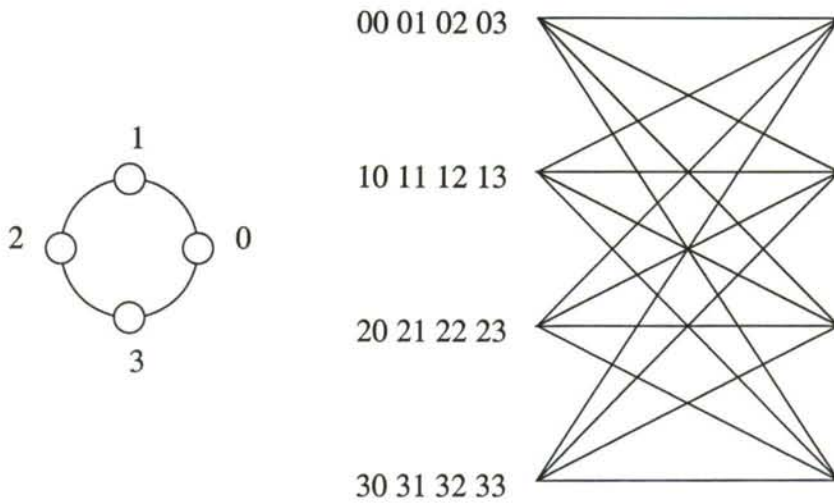


Figura 3.3: Código *trellis* de cuatro estados para QPSK con $n_t = 2$ antenas transmisoras

que se haya seleccionado, en función del estado actual y de los bits de entrada. A modo de ejemplo, se muestra en la Figura 3.3 el diagrama *trellis* de un STTC de cuatro estados para QPSK y $n_t = 2$ antenas transmisoras. La única peculiaridad del *trellis* es que las etiquetas son vectores de símbolos, como también lo es la observación ruidosa proveniente del canal que usa en el proceso de decodificación. En general, el diseño de STTCs que satisfagan el criterio del rango y maximicen el criterio del determinante es totalmente dependiente de los parámetros del sistema, es decir, número de antenas transmisoras y receptoras y el *trellis* y la constelación utilizados. En este sentido, Hammons [92] derivó el equivalente del criterio del rango no sobre el alfabeto de la constelación, sino sobre su representación en binario para constelaciones BPSK y QPSK, en función de la asignación bit a símbolo utilizada. Este criterio es útil para establecer la unión entre la multitud existente de códigos algebraicos y el criterio del rango que ha de cumplir un código espacio-temporal. Es interesante observar que el criterio binario obtenido es suficiente pero no necesario. Por ejemplo, el esquema de Alamouti extrae diversidad completa (aunque no aporta ninguna ganancia de codificación), pero sin embargo su representación en binario no cumple el criterio binario del rango cuando se utiliza señalización BPSK.

Los códigos ST *trellis* obtenidos hasta la fecha han sido difíciles de diseñar, y específicos para el valor de los parámetros (número de antenas, *trellis* y constelación) elegido. De todos modos se han propuesto esquemas que escalan los códigos existentes a un número mayor de antenas, consistentes en realizar codificación multinivel. Sin embargo, la mayor desventaja de los STTCs es la complejidad asociada al proceso de decodificación: el decodificador MV tiene una complejidad que crece exponencialmente en el número de antenas transmisoras, el orden de la modulación y la memoria del *trellis*.

Su elevada complejidad, junto con el hecho de que la ganancia asociada a un código espacio-temporal proviene fundamentalmente de que extraiga diversidad completa, ha motivado el interés por diseñar códigos ST de baja complejidad de codificación que extraigan diversidad

completa, aunque su ganancia por codificación sea mucho menor (incluso nula) que la de los STTCs. La idea básica es obtener esquemas como el de Alamouti, pero para $n_t > 2$, que extraigan diversidad completa. La generalización del esquema de Alamouti a $n_t > 2$ está representada por los denominados códigos ST bloque ortogonales (OSTBC).

3.6.2. Códigos ST bloque ortogonales (OSTBC)

Analizaremos el rendimiento del esquema de Alamouti utilizando los criterio del rango y del determinante. Cualquier palabra código transmitida i es de la forma

$$\mathbf{S}_i = \begin{bmatrix} s_1(i) & -s_2^*(i) \\ s_2(i) & s_1^*(i) \end{bmatrix} \quad (3.73)$$

La diferencia entre cualesquiera palabras código i y j será por tanto de la forma

$$\mathbf{E}_{ij} = \begin{bmatrix} e_1(i, j) & -e_2^*(i, j) \\ e_2(i, j) & e_1^*(i, j) \end{bmatrix} \quad (3.74)$$

donde $e_k(i, j) = s_k(i) - s_k(j)$, $k = 1, 2$. La matriz \mathbf{E}_{ij} es una matriz ortogonal con dos autovalores positivos de igual magnitud. Es, por tanto, de rango completo, por lo que también lo es \mathbf{G}_{ij} , siendo los dos autovalores de esta última matriz iguales al cuadrado de los dos autovalores de \mathbf{E}_{ij} . Tenemos por tanto que el esquema de Alamouti extrae diversidad espacial completa. Además, la estructura de las palabras código es tal que, independientemente de la realización del canal, el canal efectivo es ortogonal, desacoplando por tanto la dimensión espacio, posibilitando una detección sencilla simplemente considerando los canales escalares

$$y_i = \sqrt{\frac{E_s}{2}} \|\mathbf{H}\|_F^2 s_i + n_i \quad i = 1, 2 \quad (3.75)$$

donde n_i es ruido $\mathcal{CN}(0, \|\mathbf{H}\|_F^2 N_0)$.

La generalización del esquema de Alamouti a $n_t > 2$ es lo que se conoce como códigos ST bloque ortogonales (OSTBC). Es posible diseñar OSTBCs para cualquier número de antenas transmisoras si la constelación utilizada es real [84], utilizando la solución al problema de Hurtwiz-Radon. Un ejemplo de diseño ortogonal para $n_t = 4$ y constelación real es el siguiente

$$\mathbf{S} = \begin{bmatrix} s_1 & -s_2 & -s_3 & -s_4 \\ s_2 & s_1 & s_4 & -s_3 \\ s_3 & -s_4 & s_1 & s_2 \\ s_4 & s_3 & -s_2 & s_1 \end{bmatrix} \quad (3.76)$$

Utilizando el hecho de que los autovalores de una matriz ortogonal son iguales en magnitud, la particularización de la Ecuación (3.72) a un OSTBC resulta

$$\bar{P}(\mathbf{S}_i \rightarrow \mathbf{S}_j) \leq \left(\frac{n_t}{\|\mathbf{E}_{ij}\|_F^2} \right)^{n_t n_r} \left(\frac{\rho}{4n_t} \right)^{-n_t n_r} \quad (3.77)$$

por lo que cualquier OSTBC extrae diversidad espacial completa $n_t n_r$, y además permite la detección desacoplada de los flujos de símbolos transmitidos por cada antena. Los OSTBCs descritos hasta ahora tienen tasa de codificación espacial igual a uno.

Desgraciadamente, en el caso de constelaciones complejas, no existen diseños ortogonales con tasa de codificación espacial uno para $n_t > 2$ [84]. Sin embargo, existen diseños ortogonales para cualquier número de antenas y tasas de codificación espacial menores o iguales que $1/2$. También ha sido posible obtener diseños ortogonales para $n_t = 3, 4$ de tasa de codificación mayor que $1/2$.

El hecho de utilizar OSTBCs que permitan una detección sencilla es la razón última de la separación de los bloques de codificación espectro-temporal y espacial en la Figura 3.2. Se trata de desacoplar las dimensiones tiempo y frecuencia de la dimensión espacio. Utilizando un OSTBC como código ST garantizamos el aprovechamiento de toda la diversidad disponible en el canal MIMO, que es el factor determinante en la probabilidad de error, al tiempo que mantenemos una complejidad muy reducida del módulo de decodificación ST. Con ello aseguramos que la pendiente de la curva de probabilidad de error sea la máxima posible, y dejamos la tarea de desplazarla hacia la izquierda a la ganancia de codificación que pueda proporcionar el código espectro-temporal. Por descontado que la ganancia de codificación de un STTC es mayor que la de un OSTBC aislado, pero dicha ganancia de codificación puede ser igualmente obtenida, a un coste computacional global mucho menor, por el código espectro-temporal.

3.6.3. Multiplexación espacial

Los códigos ST considerados hasta ahora tenían una tasa de codificación espacial $0 < r_s \leq 1$. En otras palabras, en cada uso del canal MIMO se transmite como mucho un solo símbolo de información, considerando como símbolo de información un símbolo a la entrada del codificador ST. La idea subyacente a la utilización de estas tasas de codificación espacial es utilizar toda la diversidad de transmisión posible, n_t , para cada símbolo de información. Sin embargo, se pueden considerar tasas de codificación espacial intermedias, $1 < r_s \leq n_t$.

El caso extremo, en el que la tasa de codificación espacial es igual a n_t , se corresponde con una tasa de codificación uno, desde el punto de vista de la cantidad de información fuente transmitida. En este caso el módulo de codificación no añade redundancia alguna, y lo que realiza es simplemente la multiplexación espacial de los símbolos fuente. Lo único que ha de considerarse es cuál es la mejor política de asignación de los símbolos de entrada a las distintas antenas transmisoras.

Como en el caso de un código ST general, el receptor MV escoge como vector transmitido aquel que minimiza la distancia euclídea al vector recibido

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|_F^2 \quad (3.78)$$

Considerando que las palabras código ahora son simplemente los vectores \mathbf{s} transmitidos, la

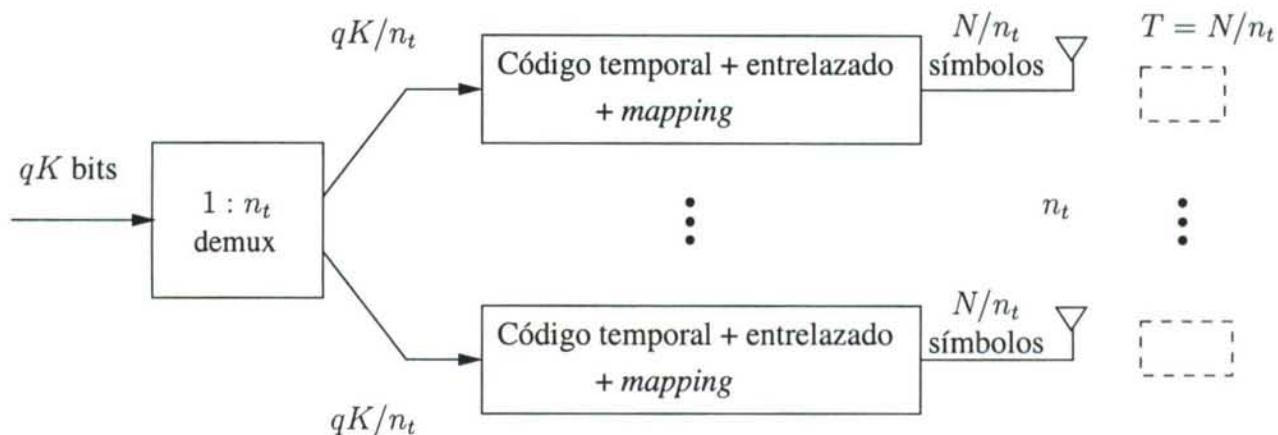


Figura 3.4: Diagrama de bloques del esquema de Multiplexación Espacial de Codificación Horizontal

probabilidad de error (3.72) dos a dos media en el caso $\mathbf{H} = \mathbf{H}_w$ es

$$\bar{P}(s_i \rightarrow s_j) \leq \frac{1}{\lambda(\mathbf{G}_{ij})^{n_r}} \left(\frac{\rho}{4n_t} \right)^{-n_r} \quad (3.79)$$

donde $\lambda(\mathbf{G}_{ij}) = (\mathbf{s}_i - \mathbf{s}_j)(\mathbf{s}_i - \mathbf{s}_j)^H$. La diversidad que extrae este esquema es n_r , ya que al haber considerado como palabra código sólo el vector transmitido en un período de símbolo, el rango de \mathbf{G}_{ij} es uno.

Veamos ahora las tres técnicas más populares de multiplexación espacial.

Codificación horizontal (CH)

Consiste en demultiplexar el flujo de bits original en tantos flujos como antenas y realizar codificación temporal de modo independiente en cada uno de ellos. La Figura 3.4 muestra un diagrama de este esquema de transmisión. La tasa de codificación espacial es n_t . La tasa de señalización es qr_t/n_t bits por uso del canal, donde r_t es la tasa de codificación de cada código temporal. La diversidad que extrae este esquema es como mucho n_r (la diversidad en recepción), ya que estamos transmitiendo cada símbolo fuente por sólo una de las antenas transmisoras. A cambio, este esquema permite el uso de técnicas subóptimas en recepción de baja complejidad computacional, como es el filtrado lineal para separar las señales transmitidas, o la ecualización *Decision Feedback* (DF). Además, es posible extraer una ganancia de array completa, n_r .

Codificación vertical (CV)

La mejora natural del esquema anterior consiste precisamente en dispersar la energía correspondiente a cada bit fuente a lo largo del mayor número posible de antenas. Introduciendo un código temporal, un entrelazador y realizando la asignación bit a símbolo de modo previo a la

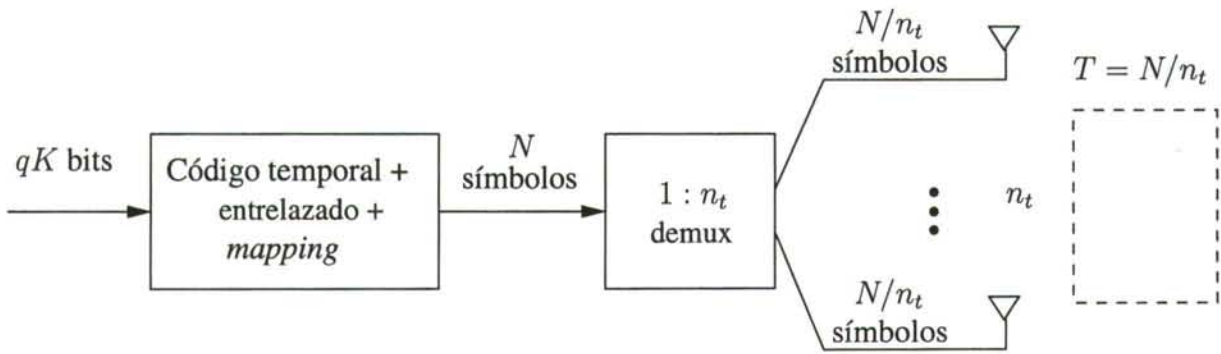


Figura 3.5: Diagrama de bloques del esquema de Multiplexación Espacial de Codificación Vertical

demultiplexación, estamos asignando los distintos símbolos codificados correspondientes a un solo bit fuente a lo largo de desvanecimientos independientes. Potencialmente, podemos extraer una ganancia por diversidad máxima, n_t . La Figura 3.5 muestra un diagrama de este esquema de transmisión. Este esquema ya lo hemos mencionado anteriormente, y si se utiliza un código temporal suficientemente potente, permite alcanzar la capacidad del canal. El inconveniente de este esquema es, como dijimos, la complejidad asociada al proceso de obtención de las verosimilitudes de cada bit codificado, exponencial en el número de antenas transmisoras y el orden de la constelación utilizada.

Codificación Diagonal (CD)

La Figura 3.6 muestra un diagrama de este esquema de transmisión. Podemos considerar este esquema como una combinación de los anteriores. Primero se realiza codificación horizontal para después asignar los flujos de datos a las antenas transmisoras de modo cíclico. Por ejemplo, si en el primer período de símbolo el vector de símbolos transmitidos es $[s_1 s_2 \cdots s_{n_t}]^T$, donde s_i denota el i -ésimo flujo de símbolos y la antena a la que se asigna viene dada por su posición en el vector, en el segundo período de símbolo el vector será $[s_{n_t} s_1 s_2 \cdots s_{n_t-1}]^T$, en el tercero $[s_{n_t-1} s_{n_t} s_1 s_2 \cdots s_{n_t-2}]^T$, y así sucesivamente. De este modo estamos utilizando todas las antenas, y por tanto todas las ramas de diversidad espacial en transmisión, para transmitir cada flujo de símbolos codificados. Por lo tanto, potencialmente podemos obtener diversidad espacial en transmisión completa. El modo de alcanzarla sería utilizando codificación óptima (gaussiana con longitud de bloque infinita) en cada flujo de símbolos. De todos modos, parece razonable suponer que si utilizamos un código temporal que rinda cercano a la capacidad del canal Rayleigh, este esquema nos permita obtener muy buen rendimiento, con la ventaja de que repartimos la complejidad de codificación y decodificación en los n_t flujos transmitidos, además de hacer más viable el uso de técnicas de detección iterativa y/o cancelación sucesiva en recepción.

La técnica D-BLAST (*Diagonal - Bell Labs Layered Space-Time Architecture*) consiste

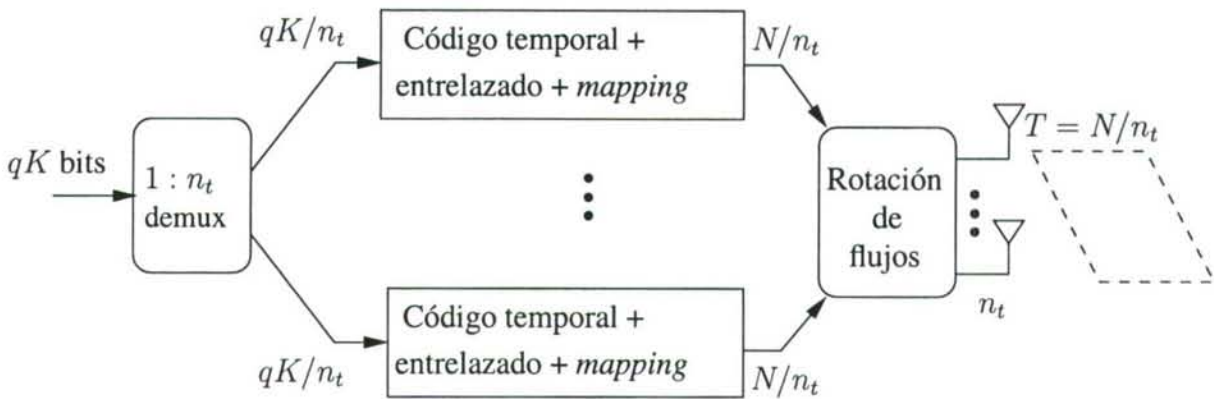


Figura 3.6: Diagrama de bloques del esquema de Multiplexación Espacial de Codificación Diagonal

precisamente en llevar a cabo CD, con la particularidad de que en el primer período de símbolo sólo se utiliza la primera antena para transmitir, dejando inactivas el resto, en el segundo se utilizan sólo la primera y la segunda, en el tercero las tres primeras, y así sucesivamente. Sólo a partir del n_t -ésimo período de símbolo se utilizan todas las antenas. La Figura 3.7, donde los números hacen referencia a los flujos de datos originales, ilustra esta idea. Este pequeño desperdicio de recursos permite detectar los flujos de datos mediante cancelación sucesiva de los símbolos ya detectados.

3.6.4. Ventajas e inconvenientes de los distintos esquemas de codificación ST

Parece que entonces la combinación de códigos espectro-temporales y OSTBCs es la solución definitiva en el campo de los transmisores MIMO. Sin embargo, debemos decir que lo es sólo en ciertos casos: en aquellos derivados de las suposición fundamental realizada al derivar los criterios del rango y del determinante: la cota de unión. Recordemos que esta cota es ajustada para SNRs altas, por lo que aquellos códigos que rinden cerca de la capacidad del canal no caen dentro de su ámbito de aplicación. Cuando hablamos de este tipo de códigos debemos recordar el teorema de codificación de canal de Shannon, que establece la capacidad utilizando el conjunto de los códigos generados aleatoriamente. Esta aleatoriedad es lo que intentan imitar los códigos Turbo, LDPC, LDGM, etc., aunque en el fondo tengan una cierta estructura que permite una decodificación en tiempo viable. Considerando que el canal MIMO no es más que otro tipo de canal, la utilización de un código de los tipos mencionados debería rendir de modo próximo a la capacidad del canal. Además, de este modo no es necesario distinguir entre código espectro-temporal y espacial, porque es un solo código el que se encarga de distribuir, del modo más aleatorio posible, los bits fuente a lo largo de las palabra código.

¿Cuál es entonces la desventaja de utilizar códigos que rindan cercanos a la capacidad del

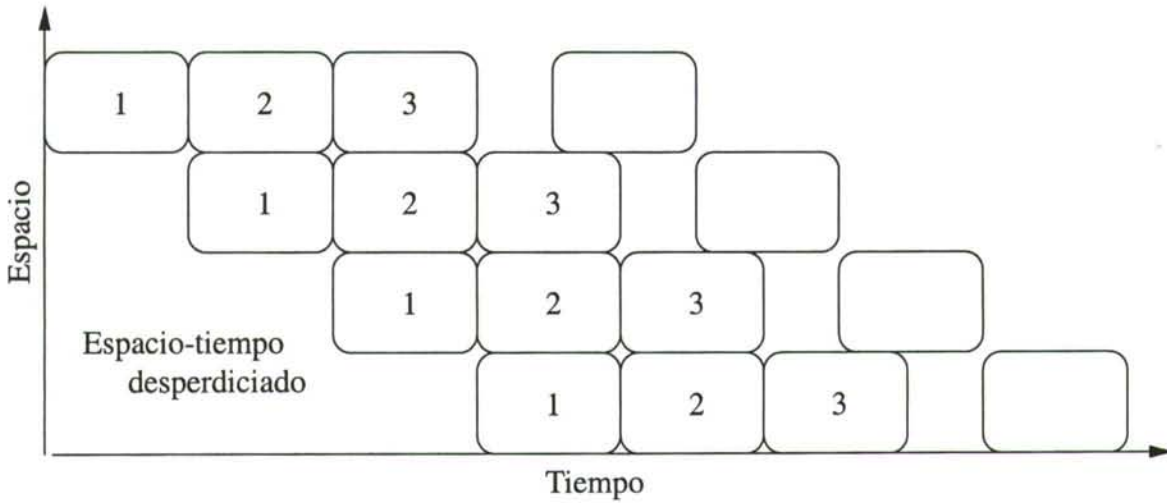


Figura 3.7: Transmisión D-BLAST

canal?. En primer lugar, la longitud de las palabras código. Como también se aprecia en el caso de canales SISO, las longitudes que deben tener las palabras generadas por estos códigos para un buen rendimiento pueden ser excesivamente largas en ciertas aplicaciones. De ahí que pueda ser preferible utilizar un código *trellis* o bloque. En segundo lugar, comparten con los STTCs el hecho de que el cálculo de las verosimilitudes de las observaciones del canal es exponencial tanto en el número de antenas transmisoras como en el orden de la modulación. Sin embargo, este inconveniente puede ser efectivamente resuelto utilizando estrategias subóptimas de cálculo de las verosimilitudes, como la propuesta por ten Brink basada en detección esférica. De hecho, el propio ten Brink muestra cómo es posible obtener un rendimiento cercano a la capacidad del canal MIMO utilizando un código Turbo clásico para un número de antenas elevado, y con una complejidad asumible. También es posible considerar el uso de un OSTBC precisamente para reducir la complejidad del proceso de detección, a costa de una pérdida significativa en la tasa de transferencia del transmisor, pero no es necesario desde un punto de vista de reducción de la probabilidad de error.

3.6.5. Codificación espacio-temporal cuando el transmisor conoce el canal

El conocimiento del canal por parte del transmisor, no sólo se traduce en una capacidad del canal mayor para SNRs bajas, sino que simplifica notablemente el diseño de un sistema de comunicación y, sobre todo, permite obtener rendimientos excelentes con soluciones de bajo coste computacional.

El esquema de transmisión y recepción a seguir cuando el transmisor conoce perfectamente el canal es el explicado cuando hablamos de la capacidad del canal. El caso de conocimiento imperfecto, más realista, es tema de constante investigación en la actualidad. Por ejemplo,

podemos suponer que el transmisor conoce la matriz de correlación del desvanecimiento en transmisión, \mathbf{R}_t , pero no la realización concreta del canal. En este caso la matriz óptima de asignación de potencias es todavía una cuestión sin resolver, aunque se hayan hecho algunos esfuerzos en esta línea. Una solución subóptima es la basada en lo que se denomina *waterpouring* estocástico que consiste en maximizar una expresión aproximada de la capacidad ergódica, y se reduce a realizar *waterpouring* sobre los autovalores de \mathbf{R}_t en lugar de sobre los de \mathbf{H} . La capacidad del canal cuando se utiliza este esquema es muy superior al caso de canal desconocido en transmisión, aunque evidentemente, es inferior al caso de conocimiento perfecto. Otras técnicas de interés, tanto para canal conocido totalmente como parcialmente, consideran el caso en el que la longitud de las palabras código utilizadas es pequeña.

3.7. Conclusiones

En el presente capítulo hemos estudiado los aspectos fundamentales del modelado de los sistemas MIMO resultantes de la utilización de varias antenas tanto en transmisión como en recepción. Tras la caracterización del canal radio realizada en el Capítulo 3, la obtención del modelo de señal, tanto continuo como discreto, se ha realizado de modo inmediato. En esencia, hemos modelado el canal MIMO expresando su relación entrada-salida de modo vectorial, agrupando los coeficientes de desvanecimiento en la denominada matriz del canal.

Una vez establecido el modelo de señal, hemos analizado cuál es la capacidad teórica del enlace MIMO distinguiendo básicamente los casos correspondientes a cuando el canal es o no conocido en transmisión. Si el transmisor no conoce el canal, la estrategia de asignación de potencia a cada antena transmisora ha de ser uniforme. Cuando sí lo conoce, la asignación óptima puede realizarse mediante el algoritmo *waterpouring* para obtener, especialmente en el régimen de baja SNR, un valor mayor de capacidad que en el caso de canal desconocido. La consecuencia más importante que se extrae es que, cuando los coeficientes del canal son estadísticamente independientes, la capacidad del canal crece aproximadamente de modo lineal con el mínimo número de antenas utilizadas en transmisión y recepción. Esta es una de las razones fundamentales de la atención que han suscitado los sistemas MIMO. También hemos visto cómo obtener expresiones de la capacidad cuando el canal es selectivo en frecuencia.

Después de analizar las ganancias potenciales que se pueden obtener de los sistemas MIMO, hemos pasado a estudiar otro de los aspectos fundamentales que explican la conveniencia de los sistemas MIMO: el concepto de diversidad espacial. En esencia, la utilización de múltiples antenas proporciona nuevas ramas de diversidad cuyo aprovechamiento redundará en una reducción significativa de la probabilidad de error. La ventaja de la diversidad espacial proporcionada por los sistemas MIMO es obtenida mediante la replicación de un recurso (las antenas transmisoras y/o receptoras) con un coste asociado mucho más bajo que los de tiempo y ancho de banda en los que se basan las técnicas de obtención de diversidad clásicas. De nuevo hemos distinguido los casos de canal conocido o no en transmisión e introducido las técnicas respectivas que permiten extraer la diversidad espacial existente: el esquema de Alamouti y el

ya mencionado algoritmo *waterpouring*.

Además de la ganancia por diversidad, ya comentamos anteriormente la posibilidad de obtener una ganancia por codificación, aunque su efecto en la disminución de la probabilidad de error sea mucho menos acusado que el proporcionado por el aprovechamiento efectivo de la diversidad. Se han introducido los criterios del *rango* y del *determinante* cuya maximización ha de perseguir cualquier técnica de codificación espacio-temporal. Este criterio se puede interpretar como la generalización del criterio de la distancia euclídea en el canal AWGN, y los códigos *trellis* espacio-temporales (STTC) como la generalización de TCM al caso MIMO.

La gran complejidad asociada al proceso de decodificación de los STTCs, que crece exponencialmente con el número de antenas transmisoras, el orden de la modulación y el *trellis* utilizados, ha motivado la introducción de códigos que se centran en simplemente obtener ganancia por diversidad, aunque su ganancia por codificación sea escasa o prácticamente nula, a cambio de una complejidad del algoritmo de decodificación creciente tan sólo de modo lineal en el número de antenas transmisoras. La ganancia por codificación puede ser obtenida mediante la utilización de un código de canal clásico, como por ejemplo un código Turbo.

Hemos contemplado también el caso en el que lo que nos interesa es maximizar la tasa de transferencia del sistema y, por tanto, no realizar ningún tipo de codificación espacio-temporal. Las técnicas BLAST (*Bell Labs Layered Space-Time architecture*) son una solución efectiva en este caso. En concreto, la técnica D-BLAST (*Diagonal-BLAST*) permite el aprovechamiento de la diversidad espacial proporcionada por el canal y la aplicación de un algoritmo de detección de los símbolos basado en la cancelación sucesiva de los flujos de datos ya detectados.

Por último, hemos comentado las ventajas e inconvenientes de los distintos sistemas de codificación espacio-temporal, llamando la atención sobre el hecho de que los criterios del rango y del determinante son sólo válidos en el régimen de altas SNRs. En el caso de SNRs bajas es suficiente la aplicación de la idea de codificación aleatoria de Shannon. Pero este punto será precisamente el objeto del Capítulo 8.

Capítulo 4

Sistemas MIMO de tipo GSM en túneles subterráneos

All explorers are seeking something they have lost. It is seldom that they find it, and more seldom still that the attainment brings them greater happiness than the quest

—Arthur C. Clarke

Presentaremos un esquema de transmisión/recepción MIMO como medio de aumentar la eficiencia espectral en un entorno de aplicación concreto: la transmisión digital inalámbrica en entornos de túneles subterráneos. El esquema propuesto presenta como ventaja fundamental la compatibilidad a nivel de interfaz radio con el estándar de comunicaciones móviles más ampliamente implantado, el estándar *Global System Mobile* (GSM). Los sistemas de comunicación radio europeos en ferrocarriles han adoptado desde 1993 la interfaz radio GSM (GSM-R) y además se ha reservado para ellos una banda de frecuencias concreta, denominada la banda “R” (876-880 MHz y 921-925 MHz). Debido a la naturaleza conservadora de este mercado, es razonable suponer que los sistemas de comunicación radio para ferrocarriles emplearán GSM-R en los próximos años.

Una manera de incrementar la capacidad de las interfaces GSM en dichos entornos es emplear múltiples antenas tanto en transmisión como en recepción, de modo que el canal resultante entre transmisor y receptor sea MIMO, y utilizar técnicas de codificación, señalización, ecualización y decodificación apropiadas a dicho modelo de canal.

4.1. Modelo de canal

En entornos de propagación de túneles subterráneos es razonable suponer un modelo de propagación multitrayecto (*multipath*), debido a que la propagación radio se ve severamente afectada por las numerosas reflexiones en el techo y las paredes de los túneles y estaciones. La

Teoría de las Imágenes (*Image Theory*) es el método más ampliamente usado para modelar la propagación de ondas de alta frecuencia en entornos complejos. Aunque el principio en el que se basa es bastante simple, de todos modos asume que todos los posibles trayectos entre las antenas transmisoras y receptoras pueden ser determinados. Estos trayectos se pueden obtener de modo sencillo si el túnel considerado es recto y su sección rectangular. En [93] se muestra cómo la aplicación del principio de *Image Theory* a un túnel de estas características resulta en modelar el canal como de desvanecimiento plano cuando se considera el período de símbolo utilizado en GSM ($3.7 \mu\text{s}$) [24].

En la realidad, un túnel puede diferir significativamente del modelo simplificado de túnel recto y sección rectangular. Es por ello que, en el contexto del proyecto europeo ESCORT (*Enhanced diversity and Space-time Coding for underground metro and Railway Transmission*), se llevaron a cabo medidas experimentales de la respuesta al impulso del canal en túneles del Metro de París [23]. Las medidas que aquí consideraremos fueron realizadas en la estación “Porte des Lilas” y contemplaban, por una parte, un túnel que empezaba en línea recta para continuar con una curva suave y, por otra, un túnel formado únicamente por una larga curva. A estos dos túneles los denominaremos en adelante *one way* y *two way*, respectivamente. El equipo de medida consistía en cuatro antenas de tipo bocina localizadas en el andén de la estación (a una altura de 2 m) y cuatro antenas de tipo parche colocadas en el parabrisas del tren. Las respuestas al impulso (en el dominio complejo) fueron medidas con un *channel sounder* de 35 MHz de ancho de banda, activando sucesivamente las antenas transmisoras. Las medidas se tomaron parando el tren aproximadamente cada dos metros. El canal puede considerarse estacionario, al no haber ningún objeto móvil entre transmisor y receptor.

4.1.1. Canal con desvanecimiento plano

La primera conclusión obtenida fue que al igual que en el modelo teórico resultante de aplicar *Image Theory* a un túnel recto y de sección rectangular, era razonable asumir el desvanecimiento como plano en frecuencia. La Figura 4.1 muestra dos ejemplos de respuesta al impulso del canal típicas en el túnel *one way* a distancias de 75 m y 350 m entre el transmisor y el receptor. En el primer caso el receptor está situado en la línea de visión (*Line Of Sight*, LoS) del transmisor mientras que en el segundo el receptor está situado después de la curva que enmascara la señal transmitida. El eje vertical tiene una escala lineal arbitraria mientras que la escala horizontal es de 200 ns por división. Es evidente que la duración de la respuesta al impulso del canal es considerablemente menor que el período de símbolo utilizado en GSM ($3.7 \mu\text{s}$). Este comportamiento del canal es razonable teniendo en cuenta que, debido a las características físicas de los túneles, los retardos relativos entre distintas componentes multitrayecto son muy pequeños.

Al ser el canal de desvanecimiento plano, cada respuesta al impulso $h_{li}(t)$, correspondiente a las características (en el dominio temporal) del canal entre la i -ésima antena transmisora y la l -ésima antena receptora, se reduce a un número complejo igual al valor máximo de la respuesta al impulso. El primer problema a solucionar antes de poder interpretar los datos medidos es

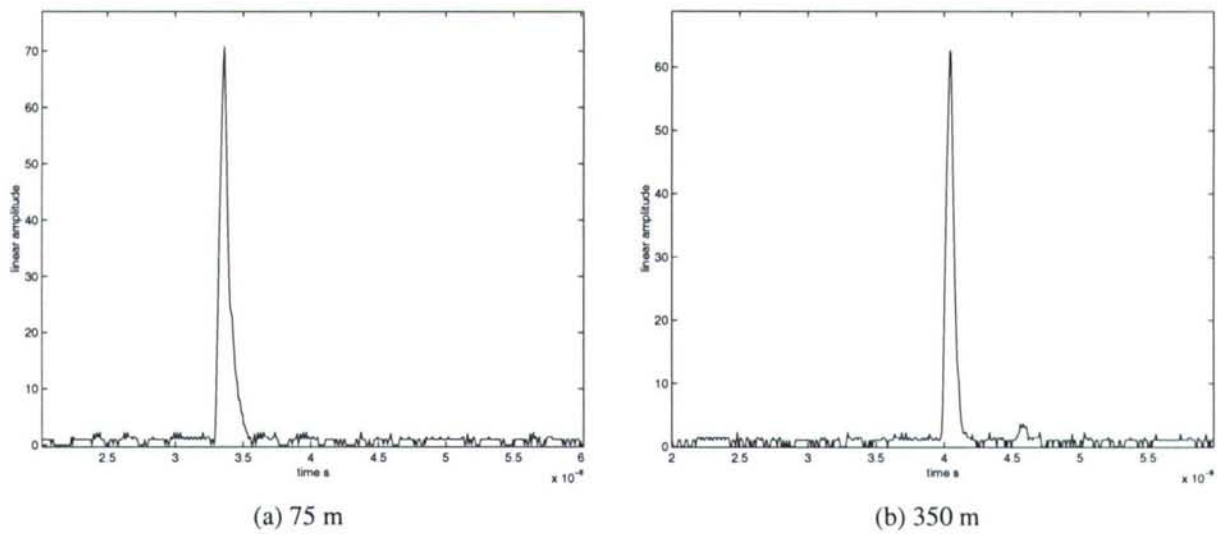


Figura 4.1: Respuesta al impulso típica del canal en un túnel subterráneo para distintas distancias emisor-receptor

substraer el efecto de la atenuación longitudinal. En otras palabras, debemos aplicar un modelo de propagación a gran escala (cf. Capítulo 2) y substraer su efecto sobre las medidas para que éstas puedan describirse mediante un modelo de propagación a pequeña escala, que es el tipo de modelo que interesa desde el punto de vista del procesado de señal.

4.1.2. Eliminación de la atenuación longitudinal

Estimaremos la potencia del canal en cada punto de medición i simplemente como el cuadrado del decaimiento correspondiente medido en dicho punto, i.e., $p_i = h_i^2$. La atenuación en túneles es muy acusada (del orden o mayor que 10 dB/100 m). Teniendo en cuenta que en un túnel la potencia recibida está relacionada de modo exponencial con la distancia al emisor [94, 21], y expresando la potencia en decibelios, tenemos que

$$P_i \text{ (dB)} = 10 \log_{10} p_i = d_i \log_{10} a + d_0 \triangleq kd_i + d_0 \quad i = 1, 2, \dots, n_m \quad (4.1)$$

donde P_i es el valor de potencia estimado (en dB) en el punto d_i , d_0 es la potencia en el primer punto de medida y n_m es el número total de medidas. Podemos obtener una nueva secuencia de valores de decaimiento en la que se haya eliminado el efecto de la atenuación debida a la propagación, i.e., $E\{10 \log_{10} \tilde{p}_i\} = E\{10 \log_{10} \tilde{h}_i^2\} = 0$. Para ello basta ajustar los valores de potencia estimados a lo largo de los puntos en los que se han realizado las mediciones mediante la regresión $\{d, P \text{ (dB)}\}$, es decir,

$$P_i \text{ (dB)} = \hat{k}d_i + \hat{d}_0 + \epsilon_i \quad i = 1, 2, \dots, n_m \quad (4.2)$$

donde ϵ_i es la desviación entre el valor observado y el predicho por el modelo en el punto i . Para definir una nueva secuencia $\{\tilde{h}_i\}$ que no esté afectada por la atenuación basta tener en cuenta

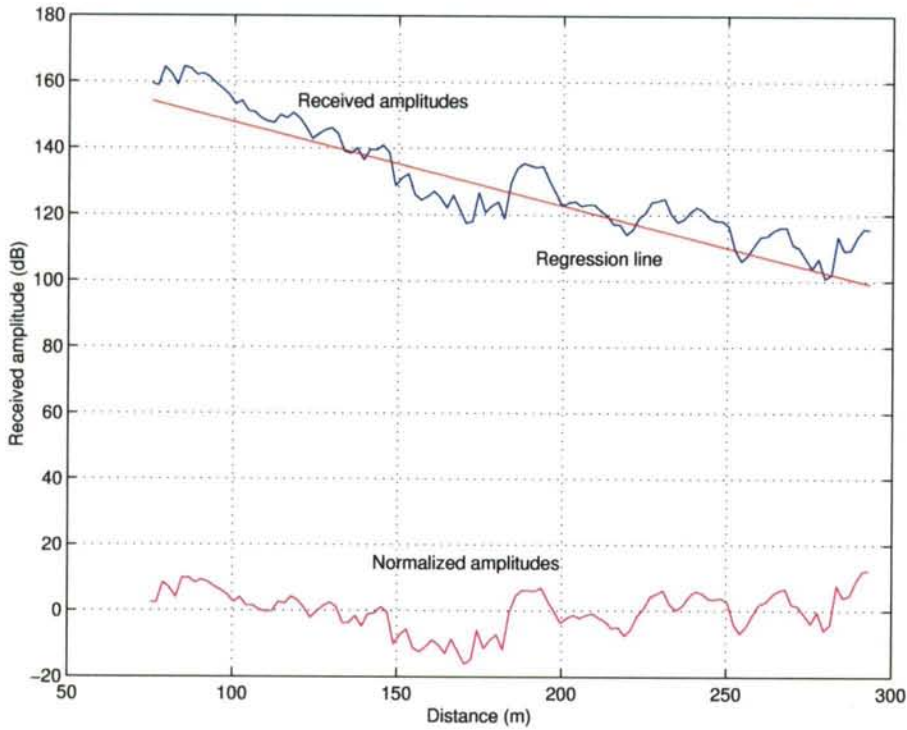


Figura 4.2: Eliminación del efecto de atenuación en las medidas experimentales del canal

que $E[\epsilon_i] = 0$, y como

$$\begin{aligned} \epsilon_i &= 10 \log_{10} p_i - (\hat{k}d_i + \hat{d}_0) = 10 \log_{10} h_i^2 - 10 \log_{10} 10^{\frac{\hat{k}d_i + \hat{d}_0}{10}} \\ &= 10 \log_{10} \frac{h_i^2}{10^{\frac{\hat{k}d_i + \hat{d}_0}{10}}} = 10 \log_{10} \left(\frac{h_i}{\sqrt{10^{\frac{\hat{k}d_i + \hat{d}_0}{10}}}} \right)^2 \end{aligned} \quad (4.3)$$

podemos definir los nuevos valores del decaimiento mediante

$$\tilde{h}_i \triangleq \frac{h_i}{\sqrt{10^{\frac{\hat{k}d_i + \hat{d}_0}{10}}}} \quad (4.4)$$

A modo de ejemplo, la Figura 4.1.2 muestra las medidas de las amplitudes del canal en un túnel subterráneo, tomadas en un intervalo de 75 a 295 m de distancia al emisor, la recta de regresión, y las amplitudes finales obtenidas.

4.1.3. Distribución estadística de los coeficientes del canal

Hasta el momento no hemos dicho nada acerca de la caracterización estadística de los coeficientes del canal. Un sencillo análisis justificará la posibilidad de asumir cada coeficiente como una muestra de una variable aleatoria gaussiana en variable compleja de media cero.

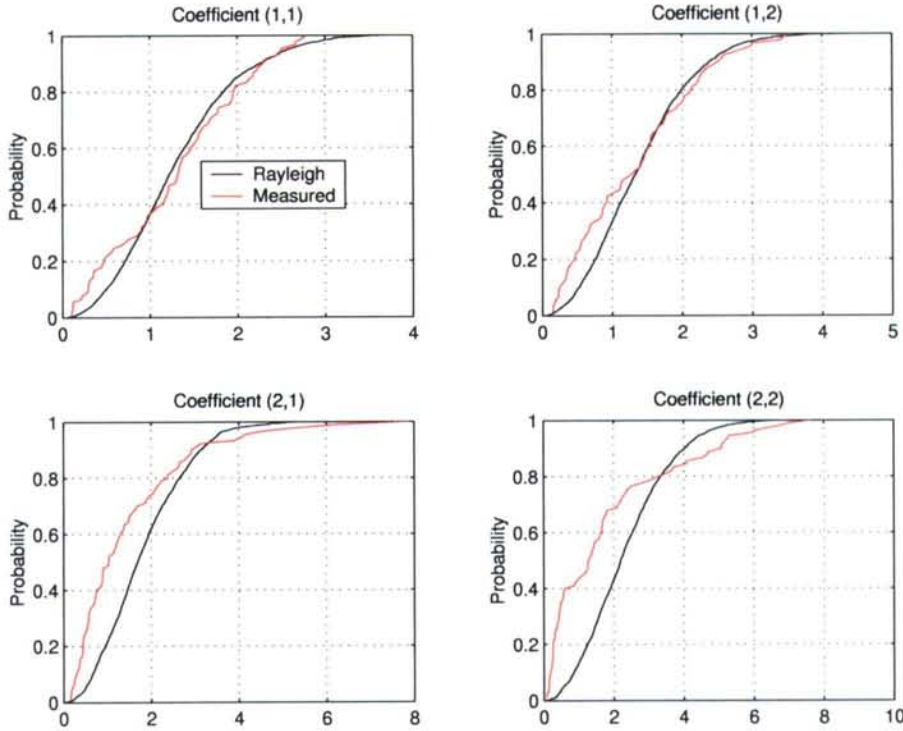


Figura 4.3: CDFs de los coeficientes del canal MIMO medido frente a las de un canal Rayleigh

La Figura 4.1.3 muestra la función de densidad complementaria (*Cumulative Density Function*, CDF) correspondiente a una distribución Rayleigh frente al módulo de cuatro coeficientes del canal MIMO medido experimentalmente. Aunque podría hacerse un estudio estadístico más refinado [94], parece razonable asumir que, en general, el módulo de cada coeficiente del canal MIMO real sigue una distribución Rayleigh. La suposición es razonable si se tiene en cuenta el gran número de multitrayectos que se producen en la propagación de la señal dentro de un túnel ya que, por la Ley de los Grandes Números, provocará que los coeficientes del canal tiendan a seguir una distribución gaussiana en variable compleja y de media cero.

4.1.4. Capacidad del canal MIMO medido experimentalmente

En el proyecto ESCORT se llevaron a cabo numerosas campañas de medida para determinar la posición óptima de las antenas en los *arrays* de transmisión y recepción, de modo que se minimizase la correlación espacial entre los distintos subtrayectos. Es claro que las antenas parche de recepción deben ser colocadas lo más alejadas posible, i.e., situando cada una de ellas en una de las esquinas del parabrisas del tren. Por otra parte, parece que las antenas bocina utilizadas en transmisión deben ser colocadas de modo perpendicular a la vía del tren para obtener la máxima diversidad espacial. Debido al ancho del andén de la estación la máxima separación entre las antenas transmisoras es de 1 m, i.e., 3λ , siendo λ la longitud de la onda

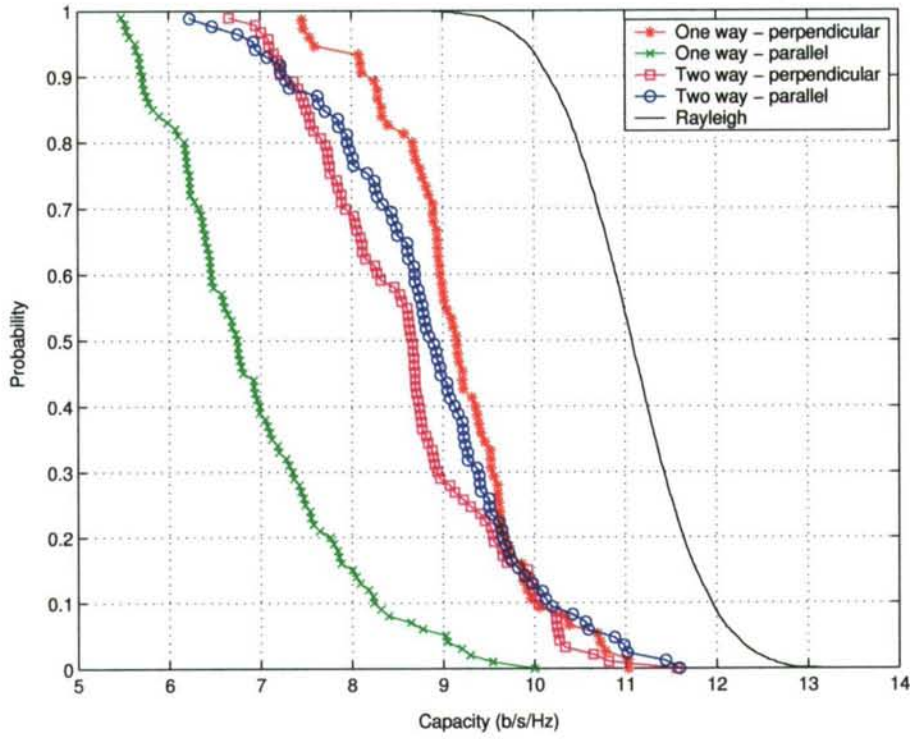


Figura 4.4: CCDF de la capacidad del canal para distintas formas del túnel (4×4 antenas)

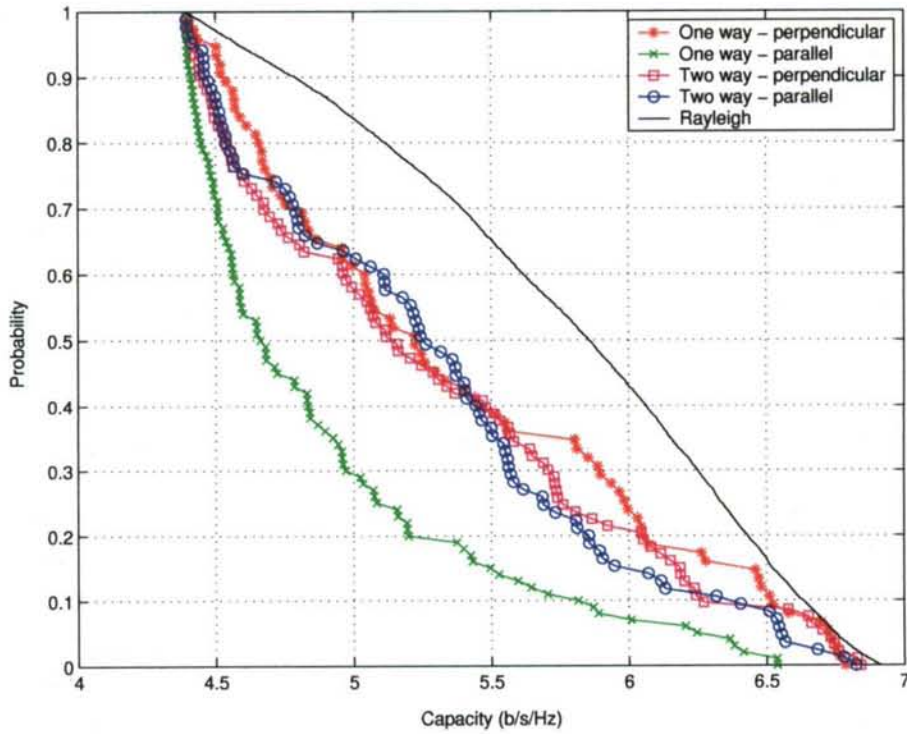


Figura 4.5: CCDF de la capacidad del canal para distintas formas del túnel (2×2 antenas)

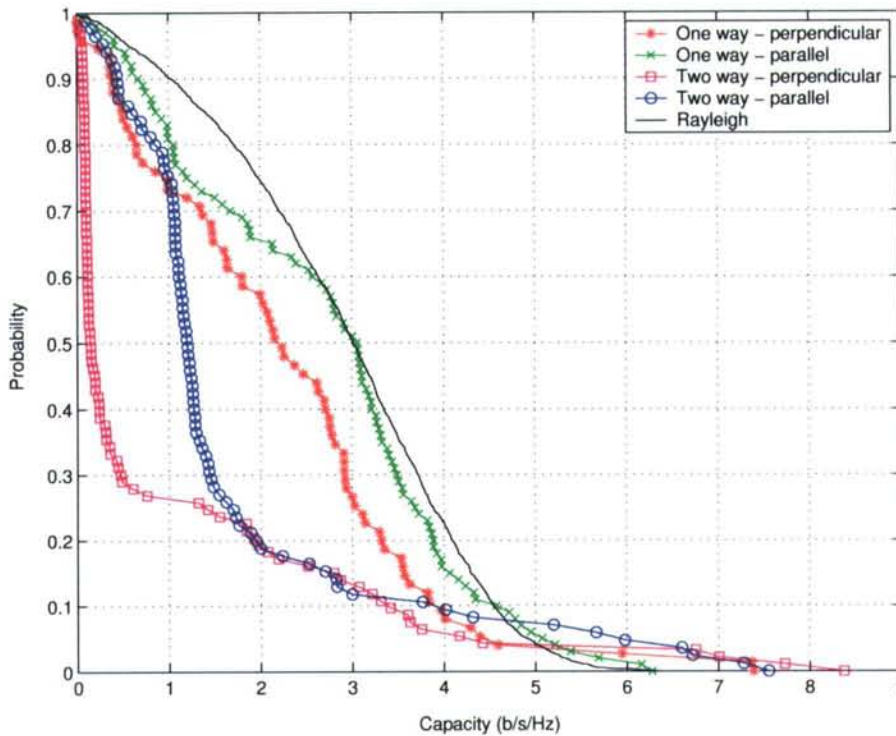


Figura 4.6: CCDF de la capacidad del canal para distintas formas del túnel (1×1 antenas)

utilizada. Para validar esta suposición, presentaremos aquí los resultados correspondientes utilizando una configuración diferente, en la que las antenas transmisoras se colocan de modo paralelo a la vía del tren.

La correlación espacial tiene un fuerte impacto en la capacidad del canal como vimos en el Capítulo 3. Las Figuras 4.4, 4.5 y 4.6 muestran la función complementaria de densidad acumulada (*Complementary CDF*, CCDF) de la capacidad (cf. Ecuación (3.8)) de los canales 4×4 , 2×2 y 1×1 (respectivamente) medidos para una $\text{SNR} = 10$ dB. Cada matriz de canal en el caso 2×2 resulta de considerar los coeficientes de la matriz 4×4 correspondientes a las antenas transmisoras y receptoras más alejadas. En las figuras también se muestra la función de densidad complementaria de la capacidad de un canal Rayleigh incorrelado.

Parece claro que la capacidad del canal incrementa con el número de antenas. Por ejemplo, considerando una probabilidad de fuera de servicio (*outage*) de 0.5 y la configuración *one way - parallel*, las capacidades de los canales 4×4 , 2×2 y 1×1 son, respectivamente, 6.7, 4.6 y 3 (bit/s/Hz). Nótese que para otras configuraciones las ganancias de capacidad obtenidas al incrementar el número de antenas son aún mayores.

Es interesante notar que la configuración *one way-parallel* exhibe la máxima capacidad en el caso de 1×1 antenas, mientras que presenta el peor valor de capacidad para los casos de 4×4 y 2×2 antenas. Esto es debido a la presencia de fuertes correlaciones entre los coeficientes de la matriz de canal. La Figura 4.7 muestra la correlación observada entre el primer coeficiente, h_{11} , y los otros coeficientes de la matriz de canal para las configuraciones *one way*. Es claro

que la configuración *one way-parallel* exhibe una correlación espacial mayor que la *one way-perpendicular*. La consideración de otras configuraciones arrojó las mismas conclusiones.

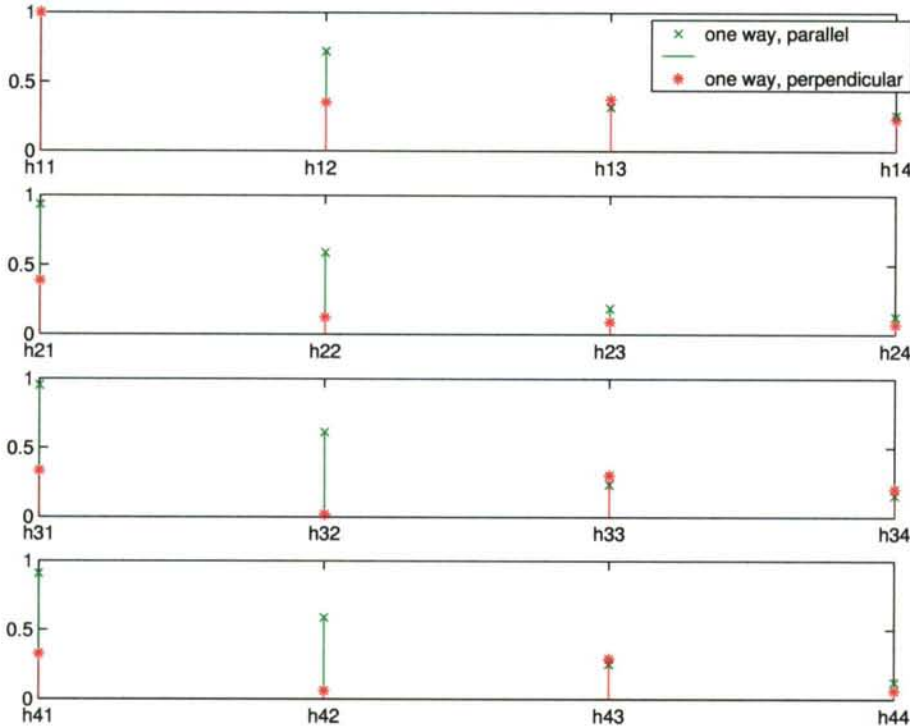


Figura 4.7: Correlaciones espaciales entre los coeficientes del canal MIMO

De la discusión anterior podemos concluir que si el túnel presenta curvas largas, la posición de las antenas transmisoras no es importante ya que es entonces la geometría de dichas curvas la que aporta diversidad espacial al canal. En cambio, la posición de las antenas transmisoras es crucial si consideramos túneles rectos (*one way*). En este caso, las medidas experimentales revelan que la mejor opción es colocar las antenas perpendiculares a la vía del tren.

4.2. Modelo de señal

En GSM el formato de modulación utilizado es GMSK (*Gaussian Minimum Shift Keying*), que es una modulación de fase continua (*Continuous Phase Modulation*, CPM) de respuesta parcial con índice $h = 1/2$. El equivalente banda-base de la señal radiada por la i -ésima antena se puede representar de la siguiente manera

$$s_i(t; b_i(n)) = \sqrt{\frac{2E_b}{T}} \exp \left[j\pi \sum_{n=-\infty}^{\infty} b_i(k)q(t - kT) \right] \quad (4.5)$$

donde E_b es la energía de bit, T el período de símbolo, $b_i(n)$ la secuencia de bits a modular y

$$q(t) = \int_{-\infty}^t g(\tau) d\tau \quad (4.6)$$

donde $g(t)$ es la convolución entre un pulso de forma gaussiana y un pulso rectangular centrado en el origen [25, 95], es decir,

$$g(t) = u(t) * \text{rect}\left(\frac{t}{T}\right) \quad (4.7)$$

donde

$$\text{rect}\left(\frac{t}{T}\right) = \begin{cases} \frac{1}{2T} & |t| \leq \frac{T}{2} \\ 0 & \text{en otro caso} \end{cases} \quad (4.8)$$

$$u(t) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left\{-\frac{1}{2}\left(\frac{t}{\sigma_u}\right)^2\right\} \quad (4.9)$$

con

$$\sigma_u = \frac{\sqrt{\log 2}}{2\pi B} \quad (4.10)$$

siendo B el ancho de banda a 3 dB de $u(t)$. Alternativamente, es posible derivar una expresión cerrada para $g(t)$ dada por [96, 97, 98]

$$g(t) = \frac{1}{2T} \left[Q\left(\frac{t - \frac{T}{2}}{\sigma_u}\right) - Q\left(\frac{t + \frac{T}{2}}{\sigma_u}\right) \right] \quad (4.11)$$

donde

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\frac{\tau^2}{2}} d\tau \quad (4.12)$$

es la función de error complementario gaussiana. Con el fin de simplificar el análisis subsiguiente, redefiniremos $g(t)$ como $g(t - p/2T)$ de modo que esté limitada al intervalo $[0, pT]$, siendo p el número de períodos de símbolo donde la señal toma valores significativos. Para GSM ($B = 0,3$ dB), es razonable fijar $p = 3$ [97], como se puede ver en las figuras 4.8.(a) y 4.8.(b), que muestran las versiones de $g(t)$ y $q(t)$ con el desplazamiento temporal apropiado.

Es conocido que cualquier modulación CPM puede expresarse mediante la expansión de Laurent [96], consistente en la suma de 2^{p-1} señales PAM, donde p es la memoria de la modulación. Se puede demostrar [98] que en el caso de GMSK la primera componente PAM contiene el 99,63 % de la energía total de la señal [97, 98]. Es por ello que la señal radiada por la i -ésima antena transmisora en el período $[(k-1)T, kT]$, $s_i(t; b_i(n))$, puede aproximarse con bastante precisión por la siguiente expresión

$$s_i(t; b_i(n)) \approx \sqrt{\frac{2E_b}{T}} \sum_{n=-\infty}^{\infty} a_i(n) h(t - nT) \quad (4.13)$$

donde $b_i(n)$ es la secuencia binaria de información, E_b es la energía de bit y T el período de símbolo. Los símbolos transmitidos, $a_i(n)$ vienen dados por la relación recursiva $a_i(n) =$

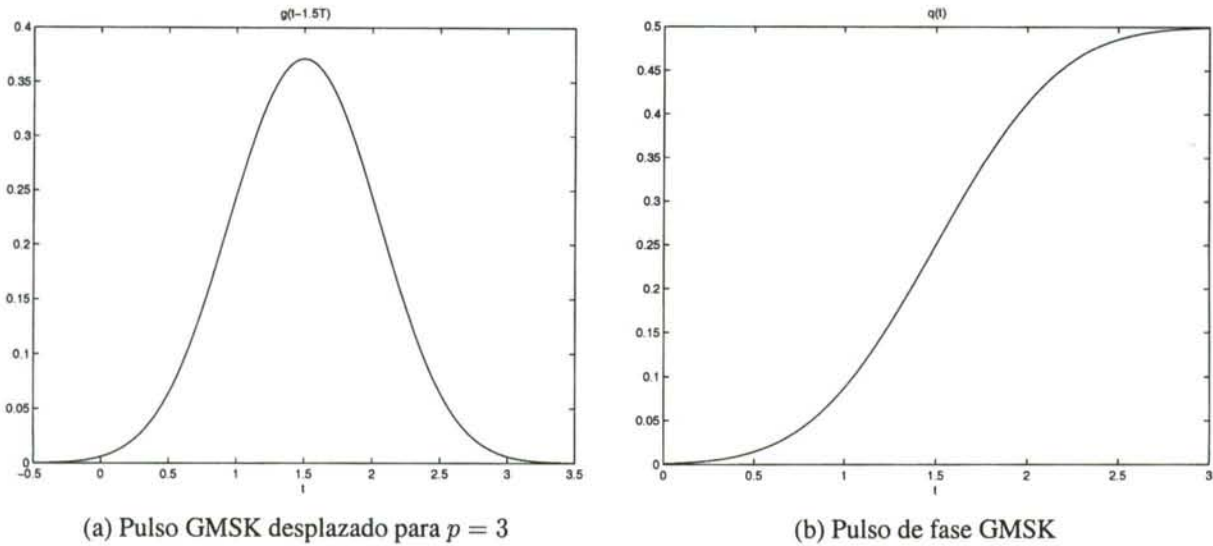


Figura 4.8: Pulsos del formato de modulación GSMK

$ja_i(n-1)b_i(n)$. Es importante subrayar que los símbolos transmitidos pertenecen a una constelación QPSK, $a_i(n) \in \{1, j, -1, -j\}$, son incorrelados y tienen varianza unidad [96]. Con el fin de simplificar el proceso de detección en el receptor, asumiremos que se emplea un precodificador diferencial como paso previo a la modulación, $d_i(k) = b_i(k-1)b_i(k)$, ya que entonces tendremos que $a_i(k) = ja_i(k-1)d_i(k) = j^k b_i(k)$. La forma de pulso de respuesta parcial $h(t)$ tiene por expresión $h(t) = C(t-3T)C(t-2T)C(t-T)$ con $C(t) = \cos(\pi q(|t|))$. La Figura 4.9.(a) muestra que $h(t)$ toma valores significativos sobre el intervalo $[0,5T, 3,5T]$ ya que las formas de pulso verdadera y la de la aproximación lineal de la GSMK están desplazadas entre sí medio período de símbolo.

Considerando que el canal de transmisión dentro de un túnel subterráneo se ve afectado por desvanecimiento multirayecto plano, la señal recibida en la l -ésima antena es

$$x_l(t) = \sum_{i=1}^{n_t} h_{li} s_i(t; b_i(n)) + n_l(t), \quad l = 1, 2, \dots, n_r \quad (4.14)$$

donde h_{li} es el coeficiente de desvanecimiento correspondiente al subcanal existente entre la i -ésima antena transmisora y la l -ésima antena receptora. La componente de ruido $n_l(t)$ se modela como un proceso estocástico en tiempo continuo de variable compleja, gaussiano y blanco con densidad espectral de potencia $N_0/2$ por dimensión.

Las señales $x_l(t)$, $l = 1, 2, \dots, n_r$ son pasadas a través de un banco de filtros adaptados a la respuesta al impulso utilizada en el modulador, $h(t)$, y muestreadas a período de símbolo para producir un conjunto de estadísticos suficientes con el que detectar los símbolos transmitidos. La salida del filtro adaptado viene dada por

$$r_l(t) = a_l(t) * h(t) * h^*(-t) + n_l(t) * h^*(-t) = a_l(t) * R_h(t) + v_l(t) \quad (4.15)$$

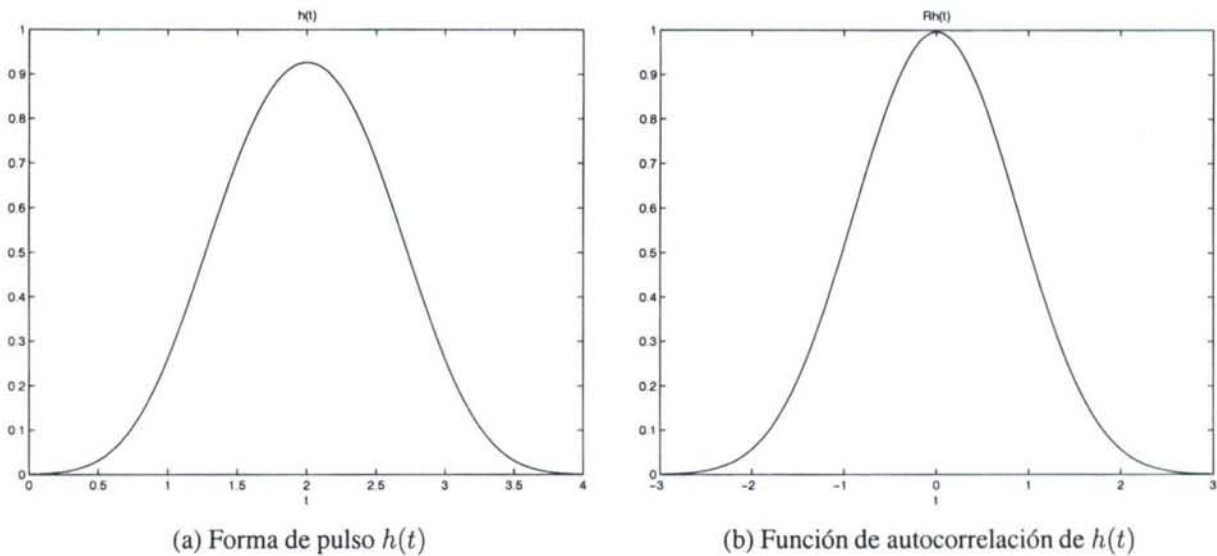


Figura 4.9: Aproximación lineal de la GMSK

donde

$$a_l(t) = \sqrt{\frac{2E_b}{T}} \sum_{k=-\infty}^{\infty} a_l(k)\delta(t - kT) \tag{4.16}$$

y $R_h(t)$ (ver Figura 4.9.(b)) denota la función de autocorrelación de $h(t)$. Tras la operación de muestreo se tiene que

$$r_l(k) \equiv r_l(t = kT) = a_l(k) * R_h(k) + v_l(k) \tag{4.17}$$

donde la función de autocorrelación de $v_l(k)$ es $R_v(k) = N_0R_h(k)$. Claramente, el ruido $v_l(k)$ es coloreado ya que $h(t)$ no cumple la condición de *zero-ISI*. No obstante, como es preferible que el proceso de ecualización se efectúe suponiendo ruido blanco, colocaremos tras el muestreador un filtro blanqueador en tiempo discreto (*discrete-time whitening filter*) [87, 99] cuya expresión en el dominio \mathbf{Z} es

$$W(z) = \frac{1}{F^*(z^{-1})} \tag{4.18}$$

donde $F^*(z^{-1})$ proviene de la factorización de la transformada \mathbf{Z} función de autocorrelación $R_h(z) = F(z) F^*(z^{-1})$. Esta expresión del filtro blanqueador consigue que la respuesta global del sistema sea $F(z)$. En el Apéndice A demostramos que el polimio de fase mínima $F(z)$ viene dado por

$$F(z) = \sqrt{\frac{r_2}{\rho_1\rho_2}}(1 - \rho_1z^{-1})(1 - \rho_2z^{-1}) = 0,8053 + 0,5853z^{-1} + 0,0704z^{-2} \tag{4.19}$$

donde $\rho_1 = -0,1522$, $\rho_2 = -0,5746$ y $r_2 = R_h(-2)$.

Además, la rotación j^k inducida por la modulación GSMK es compensada multiplicando la señal recibida por j^{-k} , resultando en la siguiente expresión para las observaciones

$$x_l(k) = \sum_{i=1}^{n_t} h_{li} \sum_{m=0}^{p-1} f(m)b_i(k-m) + g_l(k) = \sum_{i=1}^{n_t} h_{li}s_i(k) + g_l(k) \quad (4.20)$$

donde $g_l(k)$ es AWGN con varianza σ_g^2 y $f(m) = [0,8053 - 0,5853j - 0,0704]$ es la respuesta al impulso equivalente en tiempo discreto que incorpora los filtros transmisor, receptor y blanqueador, y la operación de desrotación. Usando notación vectorial las muestras tras esta última operación pueden ser expresadas mediante

$$\mathbf{x}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{g}(k) \quad (4.21)$$

donde $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \cdots \ x_{n_r}(k)]^T$, $\mathbf{s}(k) = [s_1(k) \ s_2(k) \ \cdots \ s_{n_t}(k)]^T$, y

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n_t} \\ h_{21} & h_{22} & \cdots & h_{2n_t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_r1} & h_{n_r2} & \cdots & h_{n_rn_t} \end{bmatrix} \quad (4.22)$$

que se corresponde con el modelo de señal de referencia (cf. Ecuación (3.4)).

4.3. Conclusiones

En este capítulo hemos abordado el estudio del canal radio existente en los túneles subterráneos, basándonos en las medidas experimentales de la respuesta al impulso del canal realizadas en el Metro de París dentro del contexto del proyecto europeo ESCORT (*Enhanced diversity and Space-time Coding for underground metro and Railway Transmission*). Las consecuencias extraídas pueden resumirse en que el canal puede considerarse como plano en frecuencia y con desvanecimiento Rayleigh pero presentando fuertes correlaciones espaciales. Esto hace que la capacidad del canal MIMO resultante no sea tan alta como en el caso de incorrelación espacial, lo que tendrá consecuencias sobre el rendimiento de los receptores espacio-temporales, tema del siguiente capítulo.

Por otra parte, también hemos presentado un modelo de sistema MIMO que utiliza la interfaz radio GSM. El interés del mismo radica en que esta interfaz es la adoptada por los operadores europeos de transporte urbano. Hemos visto cómo la modulación GMSK utilizada en GSM es una modulación no lineal pero que admite una aproximación lineal de bastante precisión, lo cual nos ha permitido expresar el modelo de señal correspondiente a este sistema MIMO de la misma forma que en el caso general de sistemas MIMO presentada en el capítulo anterior.

Capítulo 5

Técnicas de ecualización sin codificación de canal

All people dream, but not equally. Those who dream by night in the dusty recesses of their mind, wake in the morning to find that it was vanity. But the dreamers of the day are dangerous people, for they dream their dreams with open eyes, and make them come true

—T. E. Lawrence (“Lawrence de Arabia”)

El receptor debe estimar la secuencia de bits fuente transmitida utilizando el conjunto de observaciones y el conocimiento de la distribución estadística de las variables que componen el modelo. En adelante consideraremos el canal como un parámetro del modelo, conocido o no por el receptor, en lugar de caracterizarlo mediante una variable aleatoria. En general es razonable realizar esta suposición debido al hecho de que los modelos de propagación de desvanecimiento Rayleigh, Rice, etc., son demasiado simplistas y, por tanto, pueden ajustarse poco a la realidad. Estando interesados no sólo en estudiar el rendimiento de los receptores propuestos bajo el modelo utilizado, sino en que sean robustos cuando el canal no sigue exactamente las distribuciones comentadas, considerar el canal como parámetro en lugar de como variable aleatoria es bastante sensato. Este capítulo se centrará en los receptores que asumen el canal como conocido. El caso en el que el receptor no conozca el canal es abordado en el capítulo siguiente.

5.1. Receptores para canales planos en frecuencia

En este caso los bits fuente son demultiplexados en n_t flujos. En transmisión cada flujo es modulado independientemente del resto y transmitido por una de las antenas.

Consideraremos dos formatos de modulación. El primero corresponde a las técnicas de modulación sin memoria en las que los símbolos transmitidos por la i -ésima antena tienen la

forma

$$s_i(k) = \text{map}(b_i((k-1)M_c + 1), b_i((k-1)M_c + 2), \dots, b_i(kM_c)) \quad (5.1)$$

con $k = 1, 2, \dots, N/M_c$, siendo N el número de bits de la trama, M_c el número de bits por símbolo y “map” la función de asignación (*mapping*) bit-a-símbolo. Ejemplos de este tipo de modulación son PAM, PSK o QAM. Como vemos, en este tipo de modulaciones existe una correspondencia uno a uno entre cada posible símbolo transmitido en un período de símbolo y los bits fuente asignados a ese período. Esto hace que sea suficiente realizar la detección en el dominio de los símbolos. Por tanto, para este tipo de modulaciones los estimadores que consideraremos producen estimaciones de los símbolos transmitidos, en lugar de estimaciones de los bits fuente.

El otro formato de modulación que contemplaremos son las modulaciones con memoria y redundancia, en las que los símbolos transmitidos por la i -ésima antena tienen la forma

$$s_i(k) = \text{map}(b_i(k - m_s + 1), b_i(k - m_s + 2), \dots, b_i(k)) \quad (5.2)$$

con $k = 1, 2, \dots, N + m_s - 1$, siendo m_s la memoria inducida por la modulación. Los bits iniciales, denominados preámbulo, $b_i(-m_s + 1), \dots, b_i(-1)$, y los bits finales, $b_i(N + 1), \dots, b_i(N + m_s - 1)$, se suponen conocidos. La relación entre la cantidad de información a la entrada y a la salida del modulador, que podemos denominar como la tasa de información de la modulación, es $R_s = N/((N + m_s - 1)m_s) \approx 1/m_s$. Ejemplos de este tipo de modulación son la modulación duobinaria y las modulaciones de fase continua (*Continuous Phase Modulation*, CPM), como la *Gaussian Minimum Shift Keying* (GMSK) utilizada en el estándar GSM. En estas modulaciones ya no existe una correspondencia uno-a-uno entre el bit k -ésimo y el símbolo transmitido en el k -ésimo período, por lo que es necesario plantear los estimadores directamente sobre la secuencia de bits fuente original.

Consideremos el modelo habitual correspondiente a un sistema MIMO con canal plano en frecuencia

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{s}(k) + \mathbf{n}(k) \quad k = 1, 2, \dots, T \quad (5.3)$$

Las muestras $\mathbf{y}(k)$ son independientes, condicionadas al vector de símbolos transmitido. La única observación que contiene información acerca del vector $\mathbf{s}(k)$ es precisamente $\mathbf{y}(k)$, por lo que podemos basar nuestra decisión acerca de cuál fue el vector $\mathbf{s}(k)$, $k = 1, 2, \dots, T$ simplemente en $\mathbf{y}(k)$. Asumiendo que los símbolos transmitidos son equiprobables, el receptor debe tomar la decisión basándose en el valor de la observación $\mathbf{y}(k)$ y la f.d.p. de este vector condicionada a cada uno de los posibles vectores $\mathbf{s}(k)$ transmitidos.

Veamos entonces los receptores más utilizados, cada uno de ellos con diferente complejidad computacional y rendimiento. En adelante, suprimiremos el índice k por claridad en la exposición. Estudiaremos primero el caso de modulación sin memoria para después extender los receptores a modulaciones con memoria y redundancia.

5.1.1. Receptor óptimo: criterio Máximo-Verosímil

La regla de decisión que minimiza la probabilidad de error, cuando el coste de los errores que pueden cometerse es el mismo, es la dada por el criterio MAP. El regla de decisión MAP escoge como vector transmitido aquel que maximiza la f.d.p. de $s(k)$ a posteriori. Cuando los vectores de símbolos transmitidos son equiprobables, el criterio MAP es equivalente al MV, por lo que el receptor óptimo es

$$\hat{s} = \arg \max_{\mathbf{s}} f_{\mathbf{y}|\mathbf{s}}(\mathbf{y}) \quad (5.4)$$

que, dado que los elementos del vector de ruido \mathbf{n} son $\mathcal{CN}(0, N_0)$ independientes, es equivalente a

$$\hat{s} = \arg \min_{\mathbf{s}} \left\| \mathbf{y} - \sqrt{\frac{E_s}{n_t}} \mathbf{H}\mathbf{s} \right\|^2 \quad (5.5)$$

El inconveniente de esta regla de decisión es que el número de métricas a evaluar es $2^{n_t M_c}$, siendo M_c el orden de la modulación. El producto $\mathbf{H}\mathbf{s}$ requiere $n_r n_t$ multiplicaciones y $n_r(n_t - 1)$ sumas, pero sólo es necesario evaluarlo una vez para toda la secuencia transmitida. El cálculo de la norma al cuadrado consume n_r multiplicaciones y $2n_r - 1$ sumas. Por tanto, el coste computacional asociado a la detección de una trama es $T n_r 2^{n_t M_c} + n_r n_t 2^{n_t M_c}$ multiplicaciones y $T(2n_r - 1)2^{n_t M_c} + n_r(n_t - 1)2^{n_t M_c}$ sumas. Es decir, el coste para la detección de cada vector de símbolos transmitido es de orden exponencial en el número de antenas transmisoras y el orden de la modulación. Visto de otro modo, el coste computacional es de un orden igual al número de métricas a evaluar, $2^{n_t M_c}$. Aunque para valores pequeños de n_t y M_c el coste computacional es asumible, en general no es así. Considérese, por ejemplo, un sistema con $n_t = 4$ antenas transmisoras y modulación 16-QAM. El número de métricas a evaluar, para cada vector de símbolos transmitido, es en este caso $2^{4 \cdot 4} = 2^{16} = 65536$.

El coste computacional medio de la búsqueda del vector transmitido con mejor métrica puede reducirse imponiendo un orden adecuado a dicha búsqueda. Esta idea es la que explota el receptor de detección esférica. La idea básica de la detección esférica es considerar una hiperesfera centrada en el vector recibido y evaluar las distancias al vector recibido sólo de los puntos de la constelación interiores a dicha hiperesfera. El radio de la hiperesfera debe ser suficientemente grande para que haya al menos un punto de la constelación interior a la hiperesfera y suficientemente pequeño para que el número de puntos de la constelación contenidos en la hiperesfera no sea excesivamente grande. La detección esférica fue planteada como una manera de implementar la detección MV en constelaciones multidimensionales, y lo cierto es que su extensión a canales MIMO es totalmente trivial.

5.1.2. Detector decorrelador (*Zero-Forcing*, ZF)

En sistemas de comunicación en general, no sólo en sistemas MIMO, la complejidad del receptor óptimo MV no siempre es asumible, de modo que siempre ha habido un interés en diseñar receptores basados en la combinación de filtros lineales. Empezaremos estudiando el caso más simple, en el que el receptor consiste en un único filtro lineal. Denominaremos

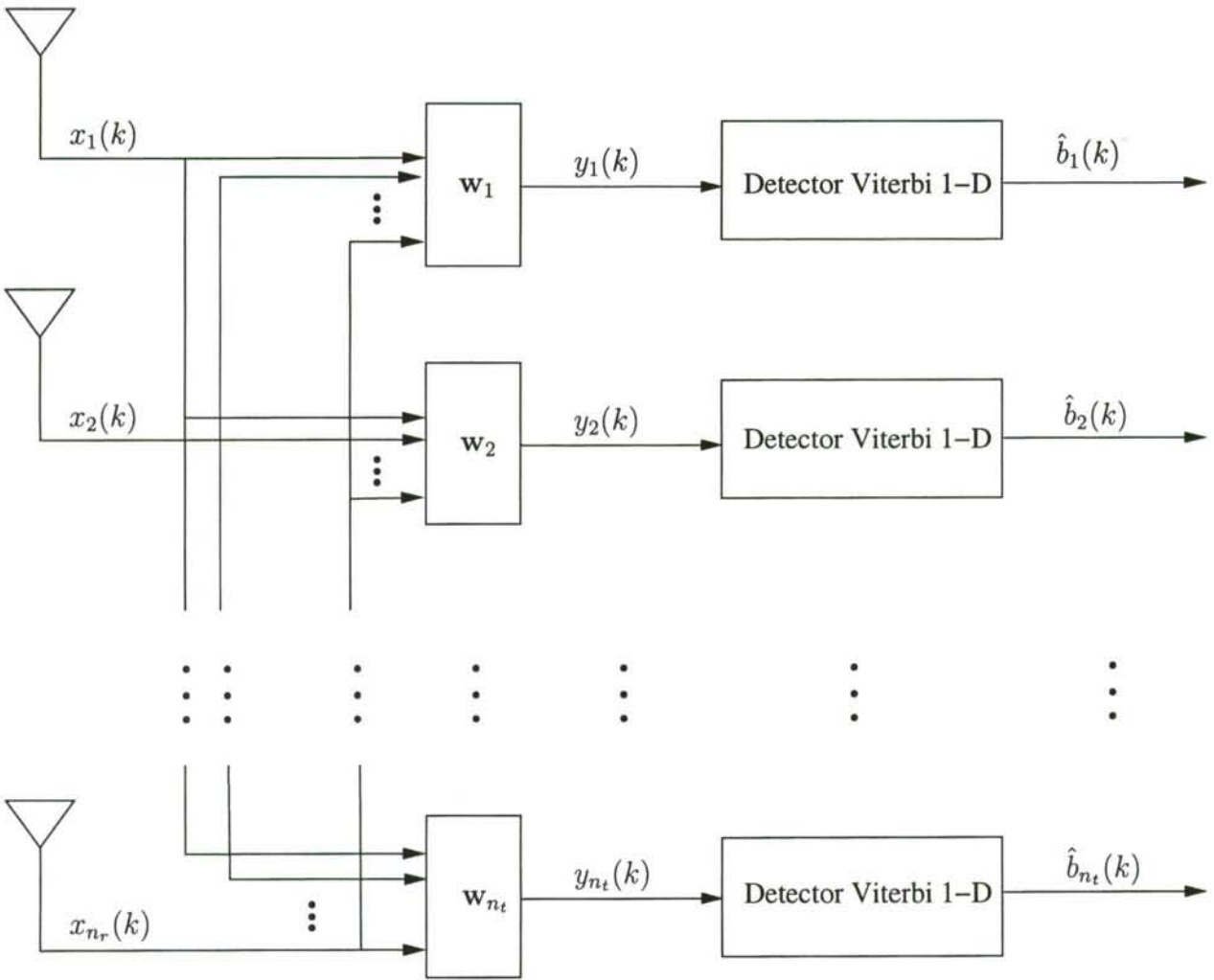


Figura 5.1: Diagrama de bloques del ecualizador LSS

a este esquema “Separación Lineal de Señales” (*Linear Signal Separation, LSS*). La Figura 5.1 muestra un esquema general de este tipo de receptor. Los coeficientes del filtro han de seleccionarse según un cierto criterio de optimización. En esta sección nos centraremos en el criterio de decorrelación (*Zero-Forcing, ZF*) mientras que en la siguiente sección estudiaremos el criterio de error cuadrático medio mínimo (*Minimum Mean Square Error, MMSE*).

El receptor decorrelador consiste en una matriz de filtrado que simplemente invierte el canal

$$\mathbf{W}_{ZF} \triangleq \sqrt{\frac{n_t}{E_s}} \mathbf{H}^{-1} \tag{5.6}$$

donde \mathbf{H}^{-1} es la matriz pseudo-inversa de \mathbf{H} . La salida del filtro decorrelador es

$$\mathbf{z} = \mathbf{H}^{-1}\mathbf{H}\mathbf{s} + \sqrt{\frac{n_t}{E_s}}\mathbf{H}^{-1}\mathbf{n} \quad (5.7)$$

$$= \mathbf{s} + \sqrt{\frac{n_t}{E_s}}\mathbf{H}^{-1}\mathbf{n} \quad (5.8)$$

El receptor decorrelador convierte el canal MIMO en n_t subcanales escalares paralelos, consiguiendo la eliminación completa de la interferencia co-canal. El inconveniente de este receptor es la amplificación de la potencia del ruido a la salida del filtro, además de que las componentes de ruido resultantes están correladas. Veamos cómo utilizando otro criterio para seleccionar los coeficientes del filtro podemos reducir estos efectos negativos.

5.1.3. Detector MMSE

El receptor MMSE representa una solución de compromiso entre la cantidad de interferencia co-canal suprimida y la amplificación del ruido a la salida del filtrado. La formulación del criterio MMSE es la siguiente

$$\mathbf{W}_{MMSE} \triangleq \arg \min_{\mathbf{W}} E\{\|\mathbf{W}\mathbf{y} - \mathbf{s}\|^2\} \quad (5.9)$$

que es un problema de optimización cuadrático que podemos resolver aplicando el principio de ortogonalidad

$$E\{(\mathbf{W}\mathbf{y} - \mathbf{s})\mathbf{y}^H\} = \mathbf{0}_{n_t, n_r} \quad (5.10)$$

resultando la denominada “solución de Wiener”

$$\mathbf{W}_{MMSE} = \mathbf{R}_x^{-1}\mathbf{P} \quad (5.11)$$

donde $\mathbf{R}_x = E[\mathbf{x}(k)\mathbf{x}^H(k)]$ es la matriz de autocorrelación de las observaciones y $\mathbf{P} = E[\mathbf{x}(k)\mathbf{s}^H(k)]$ la matriz de correlaciones cruzadas entre las observaciones y los símbolos transmitidos. Teniendo en cuenta la forma de las observaciones obtenemos

$$\mathbf{R}_x = E[\mathbf{x}(k)\mathbf{x}^H(k)] = E\left[\left(\sqrt{\frac{E_s}{n_t}}\mathbf{H}\mathbf{s}(k) + \mathbf{g}(k)\right)\left(\sqrt{\frac{E_s}{n_t}}\mathbf{H}\mathbf{s}(k) + \mathbf{g}(k)\right)^H\right] = \sigma_s^2\mathbf{H}\mathbf{H}^H + \sigma_g^2\mathbf{I} \quad (5.12)$$

Utilizando un razonamiento similar es fácil ver que $\mathbf{P} = E_s/n_t\mathbf{H}$. Por tanto el sistema de separación obtenido utilizando el criterio MMSE viene dado por

$$\mathbf{W}_{MMSE} = \sqrt{\frac{n_t}{E_s}} \left(\mathbf{H}^H\mathbf{H} + \frac{n_t}{\rho}\mathbf{I}_{n_t} \right)^{-1} \mathbf{H}^H \quad (5.13)$$

Para SNRs bajas, el receptor MMSE se aproxima al filtro adaptado

$$\mathbf{W}_{MMSE} \approx \frac{1}{N_0} \sqrt{\frac{E_s}{n_t}} \mathbf{H}^H \quad (5.14)$$

por lo que maximiza la SNR a su salida, a diferencia del comportamiento del receptor decorrelador, que amplifica la potencia del ruido a su salida. En el caso de SNRs altas, la solución MMSE se aproxima a la de decorrelación

$$\mathbf{W}_{MMSE} \approx \sqrt{\frac{n_t}{E_s}} \mathbf{H}^{-1} \quad (5.15)$$

eliminando por tanto la interferencia co-canal de modo efectivo.

Nótese que en la expresión de la matriz de filtrado \mathbf{W}_{MMSE} es necesario conocer tanto la matriz de canal \mathbf{H} como la potencia del ruido σ_g^2 . En la práctica, \mathbf{W}_{MMSE} puede estimarse haciendo uso de una secuencia de entrenamiento,

$$\hat{\mathbf{W}}_{MMSE} = \hat{\mathbf{R}}_{\mathbf{x}}^{-1} \hat{\mathbf{P}} \quad (5.16)$$

donde

$$\hat{\mathbf{R}}_{\mathbf{x}} = \frac{1}{N_t} \sum_{k=1}^{N_t} \mathbf{x}(k) \mathbf{x}^H(k) \quad (5.17)$$

$$\hat{\mathbf{P}} = \frac{1}{N_t} \sum_{k=1}^{N_t} \mathbf{x}(k) \mathbf{s}^H(k) \quad (5.18)$$

y N_t denota la longitud de la secuencia de entrenamiento.

Tras el proceso de separación es necesario, para cada una de las n_t señales transmitidas, detectar su secuencia de bits $b_i(k)$ asociada. Si la modulación es sin redundancia ni memoria, el detector óptimo vendrá dado por la aplicación de las regiones de decisión de la constelación en cuestión, mientras que si la modulación tiene memoria y redundancia es necesario aplicar el algoritmo de Viterbi para obtener la secuencia de bits más probable. La complejidad del ecualizador LSS es lineal en el número de antenas transmisoras y exponencial en la memoria de la modulación.

5.1.4. Detector de cancelación sucesiva (*Successive Cancellation, SC*)

La idea de cancelación sucesiva consiste en eliminar de la señal recibida las señales que ya han sido detectadas, de modo que la detección de la siguiente señal se efectúe bajo una relación señal a interferencia más ruido (*Signal to Interference plus Noise Ratio, SINR*) mayor. Obviamente, este receptor sufrirá del problema de la propagación de los errores cometidos en flujos de símbolos ya detectados.

El proceso de detección se puede enunciar del siguiente modo:

- Detectar el símbolo transmitido por una de las antenas, que en principio puede ser cualquiera y denotaremos por i_1 , mediante filtrado lineal

$$\hat{s}_{i_1} = \text{sgn}[\mathbf{w}_{i_1} \mathbf{y}] \quad (5.19)$$

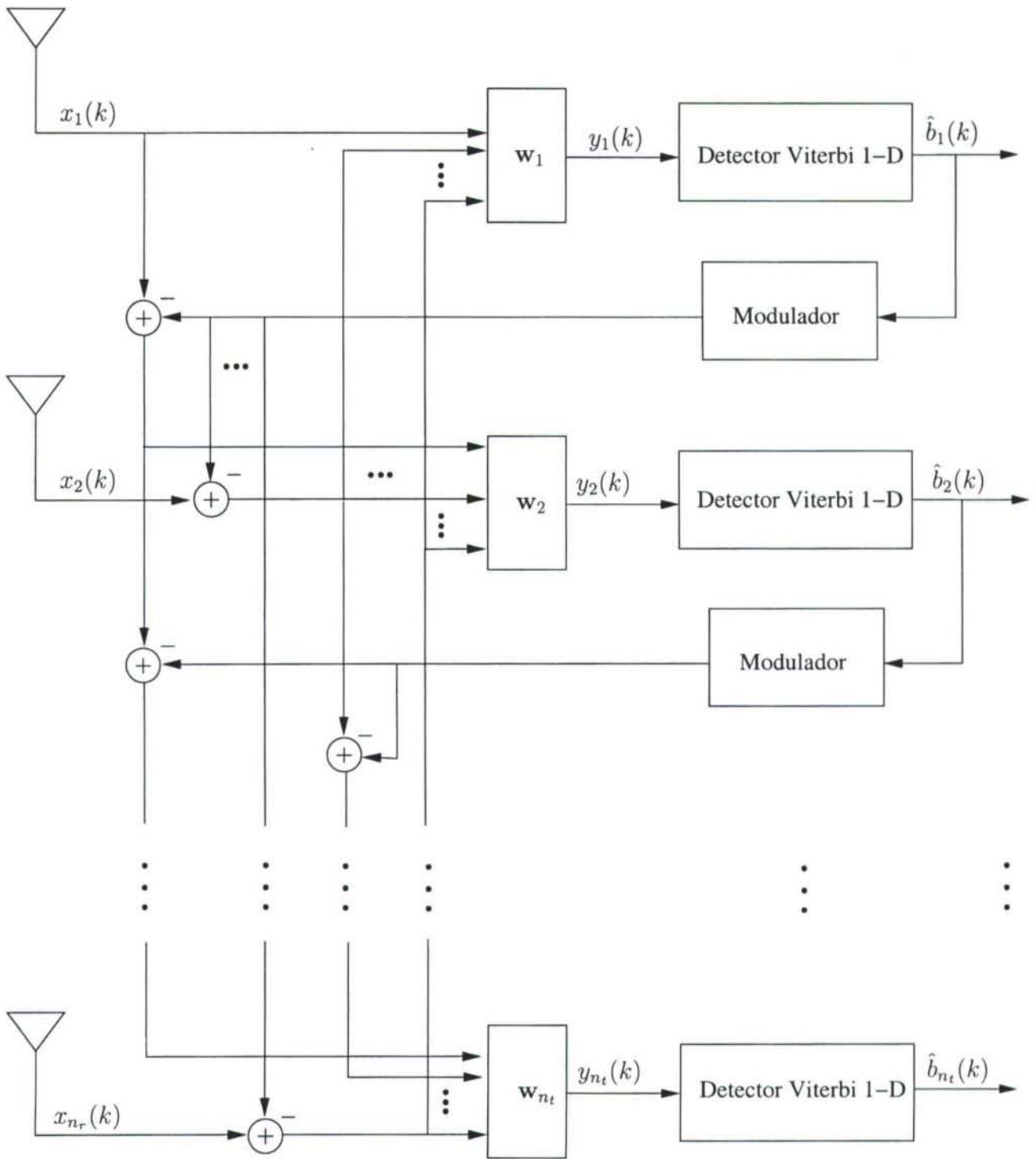


Figura 5.2: Diagrama de bloques del ecualizador SC

donde \mathbf{w}_{i_1} es el filtro lineal (la i_1 -ésima fila de la matriz de filtrado \mathbf{W}) utilizado para la antena i_1 .

- Restar la contribución de dicho símbolo en la señal recibida

$$\mathbf{y}^1 \triangleq \mathbf{y} - \mathbf{h}_{i_1} \hat{s}_{i_1} \quad (5.20)$$

donde \mathbf{h}_{i_1} es la i_1 -ésima columna de la matriz de canal \mathbf{H} .

- Detectar el símbolo transmitido por otra de las antenas, i_2 , utilizando el nuevo vector de observaciones \mathbf{y}^1

$$\hat{s}_{i_2} = \text{sgn}[\mathbf{w}_{i_2}^H \mathbf{y}^1] \quad (5.21)$$

- Restar la contribución de dicho símbolo en la señal recibida

$$\mathbf{y}^2 \triangleq \mathbf{y}^1 - \mathbf{h}_{i_2} \hat{s}_{i_2} \quad (5.22)$$

donde \mathbf{h}_{i_2} es la i_2 -ésima columna de la matriz de canal \mathbf{H} .

- Repetir los pasos anteriores de modo sucesivo para el resto de las antenas.

La matriz de filtrado \mathbf{W} se puede seleccionar de acuerdo al criterio que se prefiera: ZF, MMSE, ...

El orden en que se detectan los símbolos influye bastante el rendimiento del detector SC. Lo mejor es detectar los símbolos en orden decreciente de SINR.

5.2. Receptores para canales MIMO selectivos en frecuencia

Mientras que cuando el canal es plano en frecuencia el receptor debe combatir la interferencia co-canal y el ruido, cuando el canal es selectivo en frecuencia el receptor debe tener también en cuenta la interferencia intersimbólica. En otras palabras, las observaciones en recepción son una mezcla en las dimensiones espacio y tiempo de las señales transmitidas corrompidas por ruido. Podemos formular la relación de entrada-salida del canal como

$$\mathbf{x}(k) = \mathbf{H}(k)\mathbf{z}(k) + \mathbf{g}(k) \quad k = 1, 2, \dots, K \quad (5.23)$$

donde ahora

$$\mathbf{H} \triangleq [\mathbf{H}_{m_c-1} \mathbf{H}_{m_c-2} \cdots \mathbf{H}_0] \quad (5.24)$$

siendo \mathbf{H}_i , $i = 0, 1, \dots, m_c - 1$ una matriz de dimensiones $n_r \times n_t$ que contiene los coeficientes de desvanecimiento del canal que afectan al vector de símbolos $\mathbf{s}(k - i)$. El vector $\mathbf{z}(k)$ tiene la forma $\mathbf{z}(k) = [\mathbf{s}^T(k - m_c + 1) \mathbf{s}^T(k - m_c + 2) \cdots \mathbf{s}^T(k)]^T$. El vector $\mathbf{g}(k)$ tiene dimensiones

$n_r \times 1$ y sus componentes son independientes y representan ruido $\mathcal{CN}(0, N_0)$. Este modelo de señal es equivalente al siguiente

$$\mathbf{x}(k) = \sum_{i=0}^{m_c-1} \mathbf{H}_i(k) \mathbf{s}(k-i) + \mathbf{n}(k) \quad k = 1, 2, \dots, K \quad (5.25)$$

que se parece más al modelo de señal clásico de canal que introduce ISI.

Explicaremos cómo obtener el detector MV en este caso. Los receptores subóptimos basados en filtros lineales de la sección anterior también pueden extenderse a canales MIMO selectivos en frecuencia, pudiéndose consultar los detalles en la bibliografía.

5.2.1. Detector óptimo (MV)

Del mismo modo que hicimos en el caso de canal no dispersivo en tiempo, podemos derivar la regla de decisión óptima acerca de los símbolos transmitidos de utilizando el criterio MV. Ahora las observaciones en cada período de símbolo ya no son condicionalmente independientes con respecto al símbolo transmitido en ese instante, por lo que es necesario plantear el estimador MV de la secuencia transmitida completa utilizando la secuencia completa de observaciones. Para ello escribiremos la relación entre ambas secuencias como

$$\tilde{\mathbf{x}} = \mathcal{H}\tilde{\mathbf{s}} + \tilde{\mathbf{g}} \quad (5.26)$$

donde $\tilde{\mathbf{g}} = [\mathbf{g}^T(1) \cdots \mathbf{g}^T(M)]^T$ es un vector de coeficientes $\mathcal{CN}(0, N_0)$ independientes con dimensiones $Mn_r \times 1$, $\tilde{\mathbf{x}} = [\mathbf{x}^T(1) \mathbf{x}^T(2) \cdots \mathbf{x}^T(M)]^T$ es la trama de observaciones, con las mismas dimensiones,

$$\mathcal{H} = \begin{bmatrix} \mathbf{H}_{m_c-1}(1) & \mathbf{H}_{m_c-2}(1) & \cdots & \mathbf{H}_0(1) & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{m_c-1}(2) & \mathbf{H}_{m_c-2}(2) & \cdots & \mathbf{H}_0(2) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_{m_c-1}(M) & \mathbf{H}_{m_c-2}(M) & \cdots & \mathbf{H}_0(M) \end{bmatrix} \quad (5.27)$$

es la matriz del canal MIMO global, con dimensiones $n_r M \times M n_t$, y $\tilde{\mathbf{s}} = [\mathbf{s}^T(1) \cdots \mathbf{s}^T(M)]^T$ es la trama de símbolos transmitidos.

Asumiendo que los bits de entrada al modulador son equiprobables el decodificador MV es equivalente al MAP y viene dado por

$$\begin{aligned} \hat{\mathbf{b}} &= \arg \max_{\mathbf{b}} \left\{ \log \left[f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}};\mathcal{H}}(\tilde{\mathbf{x}}) \right] \right\} \\ &= \arg \min_{\mathbf{b}} \left\{ \theta_{MV}(\tilde{\mathbf{b}}) = \|\tilde{\mathbf{x}} - \mathcal{H}\tilde{\mathbf{s}}\|^2 \right\} \end{aligned} \quad (5.28)$$

donde

$$f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}}(\tilde{\mathbf{x}}) = \frac{1}{\pi^{n_r M} \sigma_g^{n_r M}} \exp \left[-\frac{1}{\sigma_g^2} \|\tilde{\mathbf{x}} - \mathcal{H}\tilde{\mathbf{s}}\|^2 \right] \quad (5.29)$$

es la función de densidad de probabilidad (*Probability Density Function*, pdf) condicional de la trama de observaciones, $\tilde{\mathbf{x}}$. La función $\theta_{MV}(\tilde{\mathbf{b}})$ es la métrica MV que resulta de desprestigiar en $\log \left[f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}},\mathcal{H}}(\tilde{\mathbf{x}}) \right]$ aquellos términos que no dependen de $\tilde{\mathbf{b}}$. El problema (5.28) se puede resolver de manera eficiente mediante el algoritmo de Viterbi si se dispone de una descomposición aditiva de la métrica $\theta_{MV}(\tilde{\mathbf{b}})$. Teniendo en cuenta la definición de \mathcal{H} , $\tilde{\mathbf{x}}$ y $\tilde{\mathbf{s}}$, podemos reescribir $\|\tilde{\mathbf{x}} - \mathcal{H}\tilde{\mathbf{s}}\|^2$ como una suma de términos cuadráticos

$$\theta_{MV} = \|\tilde{\mathbf{x}} - \mathcal{H}\tilde{\mathbf{s}}\|^2 = \sum_{k=1}^M \|\mathbf{x}(k) - \mathbf{H}(k)\mathbf{z}(k)\|^2 \quad (5.30)$$

lo que nos da la descomposición aditiva deseada. La métrica asociada a cada etapa del algoritmo de Viterbi es, por tanto,

$$\lambda_{MV} = \|\mathbf{x}(k) - \mathbf{H}(k)\mathbf{z}(k)\|^2 \quad (5.31)$$

El algoritmo de Viterbi recorre el *trellis* que representa la relación entre los bits a la entrada del modulador y los vectores de símbolos recibidos en ausencia de ruido, dados por el producto $\mathbf{H}(k)\mathbf{z}(k)$. En cada etapa del *trellis* calcula, para cada estado, la métrica asociada a cada posible vector $\mathbf{z}(k)$ de símbolos transmitido. Recordemos que este vector resulta de apilar los vectores de símbolos $\mathbf{s}(k - m_c + 1)$, $\mathbf{s}(k - m_c + 2)$, ..., $\mathbf{s}(k)$. Cabe distinguir aquí dos casos distintos: si la modulación es sin memoria ni redundancia, entonces hay una correspondencia uno-a-uno entre el símbolo $s_i(k)$ y los bits $b_i((k - 1)M_c + 1)$, $b_i((k - 1)M_c + 2)$, ..., $b_i(kM_c)$, con lo que cada estado del *trellis* se corresponde con cada posible $(n_t(m_c - 1)M_c)$ -upla $[\mathbf{b}^T((k - m_c)M_c + 1), \mathbf{b}^T((k - m_c)M_c + 2), \dots, \mathbf{b}^T((k - m_c + 1)M_c), \mathbf{b}^T((k - m_c + 1)M_c + 1), \dots, \mathbf{b}^T(kM_c)]^T$ de bits transmitidos en los m_c períodos de símbolo anteriores. Se tiene entonces que el número de estados en cada etapa del *trellis* es $2^{n_t(m_c - 1)M_c}$. En cambio, si la modulación tiene memoria, cada símbolo $s_i(k)$ depende, además del bit actual, de los m_s bits anteriores. El *trellis* a considerar debe describir conjuntamente tanto la ISI controlada introducida por la modulación como la ISI introducida por el canal. Habitualmente a esta representación conjunta se le denomina *supertrellis*. Cada estado del *supertrellis* se corresponde con cada posible $n_t(m_c - 1)(m_s - 1)$ -upla de bits transmitidos en los m_c períodos de símbolo anteriores. Se tiene entonces que el número de estados en cada etapa del *supertrellis* es $2^{n_t(m_c - 1)(m_s - 1)}$.

Como en el caso de canal plano en frecuencia, la limitación principal de la detección MV es su elevada complejidad. El número de estados del *trellis* es exponencial en el número de antenas transmisoras y en la ISI, tanto la inducida por la modulación como la introducida por el canal. Al utilizar una representación en *trellis* de la relación entrada-salida del canal podemos aplicar todas las técnicas disponibles para la reducción de la complejidad del algoritmo de Viterbi disponibles en la bibliografía, como por ejemplo, considerar un conjunto reducido de estados del *trellis* en lugar de el conjunto completo de estados.

5.3. Resultados de simulación

Presentamos en esta sección la evaluación del rendimiento de los receptores propuestos para multiplexación espacial, en el contexto del sistema MIMO de tipo GSM expuesto en el Capítulo 4. Como vimos entonces, medidas experimentales revelaron que las condiciones de propagación dentro de un túnel subterráneo resultan en desvanecimiento plano en frecuencia. Para poder predecir el rendimiento real de las distintas técnicas de recepción expuestas, utilizaremos las matrices del canal medidas experimentalmente.

Los receptores que analizaremos son el detector MV y los ecualizadores LSS y SC. El formato de modulación utilizado es GMSK, que como vimos se puede aproximar de modo preciso por una modulación lineal de respuesta parcial, que se ajusta al modelo de modulación con memoria y redundancia que contemplamos al describir los distintos receptores. Nos centraremos en el caso de un sistema de 4×4 antenas al ser el que mayores ganancias de capacidad aporta y, por tanto, mayores eficiencias espectrales soporta. Las Figuras 5.3, 5.4 y 5.5 muestran el *Bit-Error-Rate* (BER) obtenido con respecto a la relación señal a ruido en recepción para los receptores MV, LSS y SC. Las grandes diferencias existentes entre estos métodos de ecualización se pueden apreciar claramente: LSS es el peor método, SC proporciona una ganancia significativa y el receptor MV es el que mejor rendimiento obtiene. Nótese que para canales Rayleigh incorrelados espacialmente la diferencia en rendimiento de estos esquemas es mucho menor. Es también interesante observar que la configuración que resulta en peor rendimiento es la *one way - parallel*, que a su vez es la que presentaba menor valor de capacidad (cf. Capítulo 4, Sección 4.1.4). Los resultados obtenidos están en concordancia con el hecho de que el rendimiento de los ecualizadores lineales y *Decision Feedback* (DF) en sistemas SISO se ve severamente degradado cuando el canal presenta malas características espectrales [87].

5.4. Conclusiones

Se han presentado en este capítulo receptores adecuados para sistemas ST de máxima eficiencia espectral, es decir, sin codificación de canal. Se ha contemplado en particular el caso de canal MIMO no selectivo en frecuencia pero en el que la modulación utilizada es de respuesta parcial, tal y como ocurre con el formato de modulación GMSK empleado en el estándar GSM.

Hemos expuesto primero el receptor óptimo (MV) para después estudiar dos estructuras subóptimas de menor complejidad: el receptor de separación lineal de señales (*Linear Signal Separation*, LSS) y el receptor de Cancelación Sucesiva (*Successive Cancellation*, SC). También se ha extendido el receptor MV al caso más general de canal dispersivo.

Finalmente, hemos mostrado, mediante simulaciones por computador, el rendimiento de los distintos receptores considerando la respuesta al impulso del canal MIMO presente en los túneles del Metro de París.

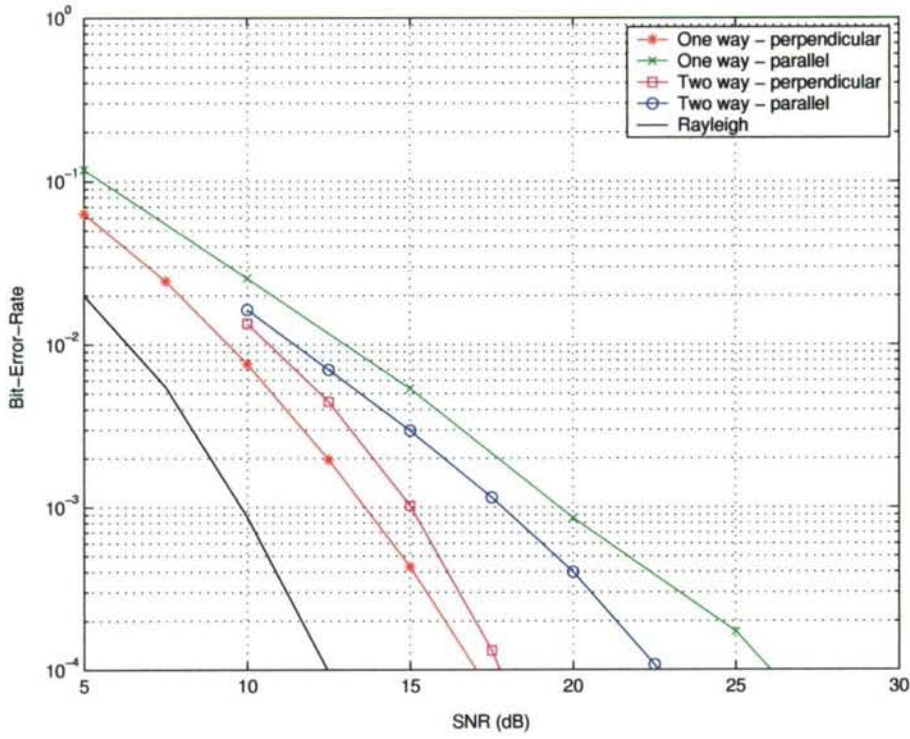


Figura 5.3: Rendimiento del receptor MV

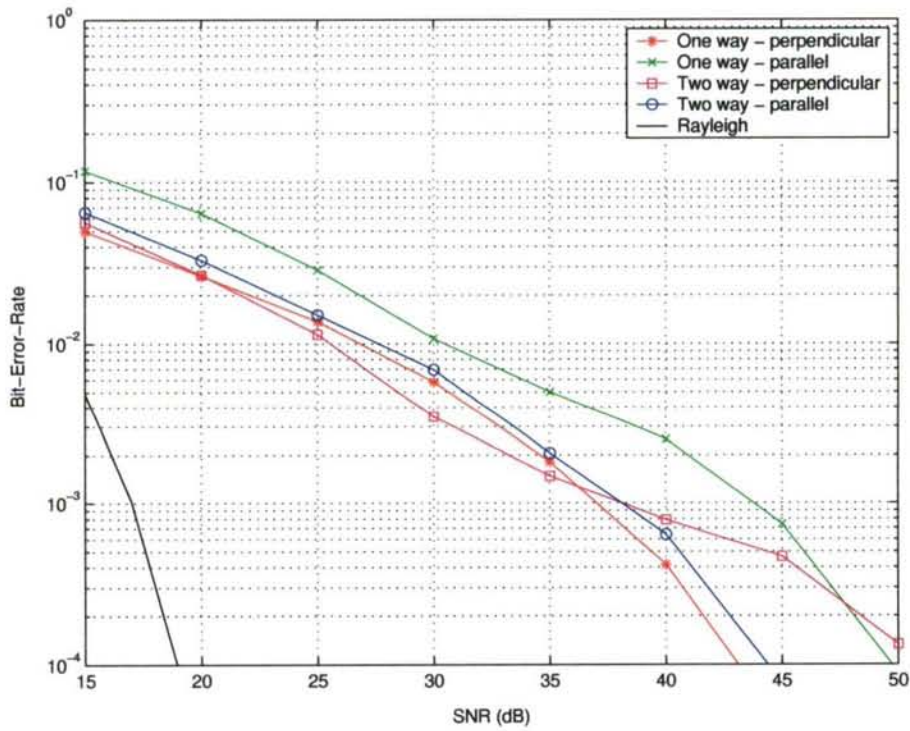


Figura 5.4: Rendimiento del receptor LSS

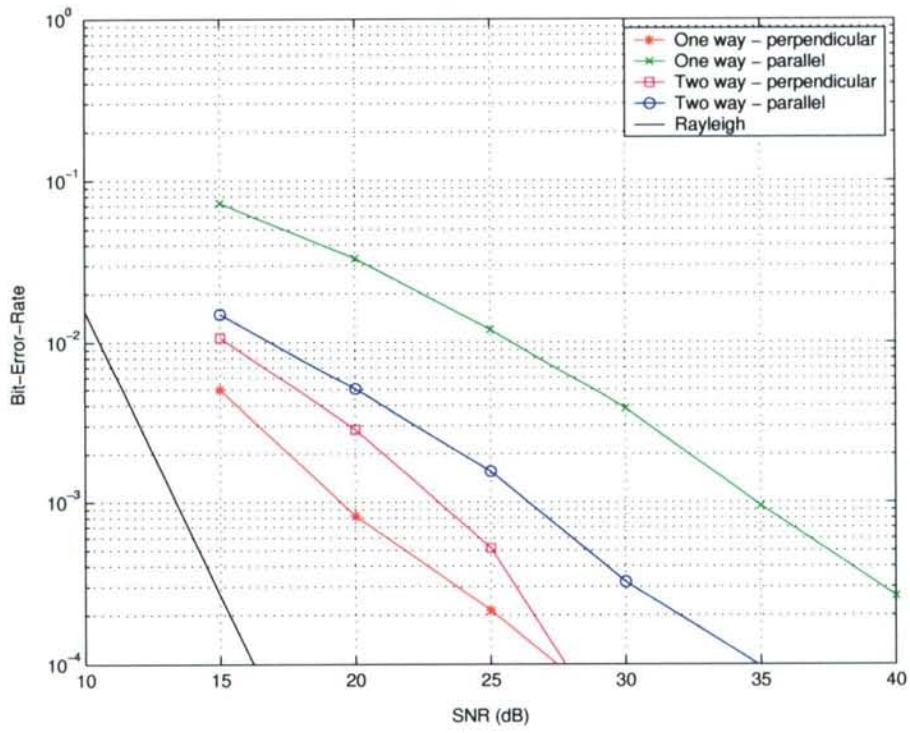


Figura 5.5: Rendimiento del receptor SC

Capítulo 6

Receptores para sistemas con codificación de canal

Nada en el mundo es tan poderoso como una idea cuyo momento ha llegado

—Víctor Hugo

Presentamos en este capítulo un esquema MIMO basado en la interfaz radio de GSM que incorpora codificación espacio-temporal. Consiste en utilizar un código ST convolucional para BPSK, seguido de un entrelazador y modulador independientes para cada antena. De este modo se respeta la estructura del código ST para que proporcione la máxima diversidad. El entrelazador decorrela los errores y además nos permite aplicar el principio Turbo en recepción para obtener un alto rendimiento. El rendimiento de este esquema utilizando medidas reales del canal MIMO existente en los túneles del Metro de París se contrasta con el proporcionado por una estructura clásica de igualación-decodificación basada en algoritmos de Viterbi con decisiones *hard*.

6.1. Modelo del transmisor

La Figura 6.1 muestra el diagrama de bloques del transmisor de un sistema MIMO. La secuencia de bits de información $u(k)$, $k = 1, 2, \dots, K$, donde K es el número total de bits a transmitir, sirve como entrada a un codificador de canal que produce la secuencia de vectores de bits $c(k) = [c_1(k) \ c_2(k) \ \dots \ c_{n_t}(k)]^T$ con $k = 1, 2, \dots, N$ y donde n_t es el número de antenas transmisoras. La tasa de codificación del código es $R_c = K/(Nn_t)$. La secuencia $c(k)$ es entrelazada componente a componente resultando en la secuencia $b(k) = [b_1(k) \ b_2(k) \ \dots \ b_{n_t}(k)]^T$. Cada secuencia de bits $b_i(k)$, $k = 1, 2, \dots, N$ es modulada independientemente del resto, convertida a paso banda (*up-converted*) e introducida en el canal por medio de la i -ésima antena. El formato de modulación que consideraremos es el utilizado

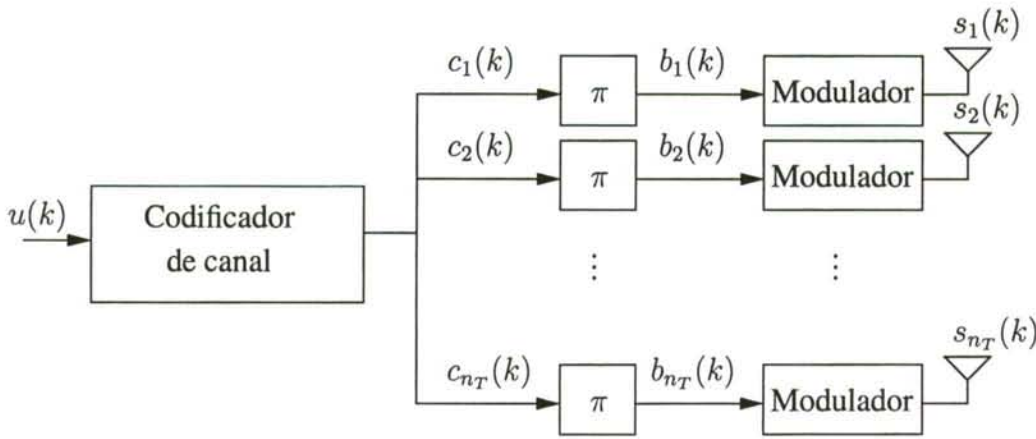


Figura 6.1: Modelo de transmisor de un sistema MIMO

en GSM, denominado GMSK (*Gaussian Minimum Shift Keying*). Este formato de modulación admite una aproximación bastante buena en forma de una modulación lineal de respuesta parcial (cf. Capítulo 4). Como hemos dicho anteriormente, el entrelazado y modulación independiente en cada antena transmisora garantiza aprovechar toda la diversidad proporcionada por el código ST BPSK utilizado [100].

Asumiremos que el canal es plano en frecuencia tal como ocurre en los túneles, por lo que tras las etapas de conversión a banda-base (*down-conversion*), filtrado adaptado a la forma de pulso utilizada en la modulación y muestreo a periodo de símbolo, podemos escribir el conjunto de observaciones de acuerdo a la Ecuación (4.21).

6.2. Modelo del receptor

El diagrama de bloques del receptor se muestra en la Figura 6.2. La tarea del ecualizador *Maximum A Posteriori* (MAP), desde una perspectiva general, es estimar las probabilidades *a posteriori* de los vectores de bits transmitidos, $\mathbf{b}(k)$, dado el conjunto de observaciones disponibles, $P[\mathbf{b}(k)|\tilde{\mathbf{x}}]$, donde $\tilde{\mathbf{x}} = [\mathbf{x}^T(1) \mathbf{x}^T(2) \dots \mathbf{x}^T(M)]^T$. Por razones de implementación, es habitual trabajar con el logaritmo de las probabilidades. En adelante utilizaremos la notación $L[x] = \log P[x]$. Si el ecualizador produce decisiones *soft*, entonces la log-probabilidad calculada de que el vector transmitido hubiese tomado un valor concreto \mathbf{b}_j , tiene la forma $L[\mathbf{b}(k) = \mathbf{b}_j|\tilde{\mathbf{x}}] \in [0, 1]$ para todo j , si produce decisiones *hard*, entonces $L[\mathbf{b}(k) = \mathbf{b}_j|\tilde{\mathbf{x}}] = 0$ sólo para un valor concreto de j y $L[\mathbf{b}(k) = \mathbf{b}_{j'}|\tilde{\mathbf{x}}] = -\infty$ para $j' \neq j$.

Restando a las log-probabilidades *a posteriori* $L[\mathbf{b}(k)|\tilde{\mathbf{x}}] = c + L[\tilde{\mathbf{x}}|\mathbf{b}(k)] + L[\mathbf{b}(k)]$, donde $c = L[\tilde{\mathbf{x}}]$, el término $L[\mathbf{b}(k)]$, se obtienen las medidas de verosimilitud (*likelihood*), $\lambda[\mathbf{b}(k)] = c + L[\tilde{\mathbf{x}}|\mathbf{b}(k)]$, a utilizar por el decodificador. Obviamente se permuta su orden mediante un desentrelazador, de modo que el decodificador las reciba en el orden adecuado, esto es, $\lambda[\mathbf{c}(k)] = \pi(\lambda[\mathbf{b}(k)])$. Asumiendo que los bits fuente $u(k)$ son equiprobables, el decodificador

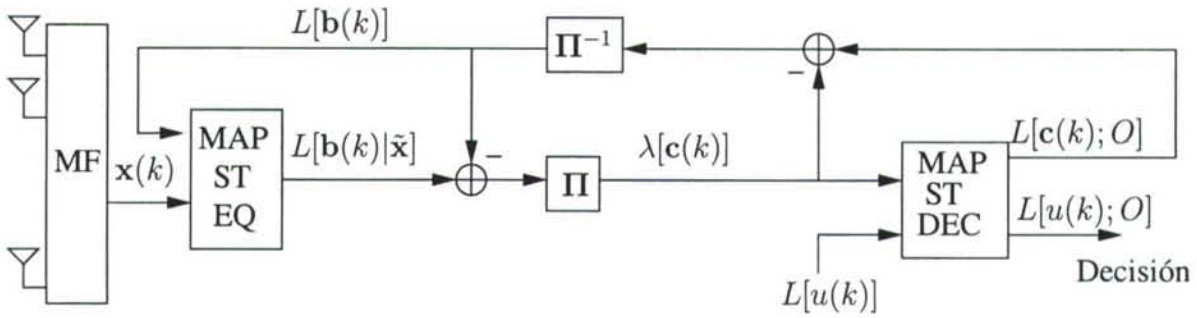


Figura 6.2: Diagrama de bloques de un receptor para sistemas MIMO con codificación

estima las log-probabilidades *a posteriori* de los bits fuente, $L[u(k)|\tilde{\mathbf{x}}]$, a partir de las cuales se toma una decisión según la acostumbrada regla $\hat{u}(k) = 0$ si $L[u(k) = 0|\tilde{\mathbf{x}}] > L[u(k) = 1|\tilde{\mathbf{x}}]$, y $\hat{u}(k) = 1$ en caso contrario. Es importante notar que el decodificador asume independencia entre las verosimilitudes $\lambda[\mathbf{c}(k)]$. Nótese que aún en el caso de decisiones *soft* el método es subóptimo cuando el canal es dispersivo, ya que entonces no existe independencia real entre distintas verosimilitudes, simplemente es “simulada” por el proceso de entrelazamiento. De ahí que sea conveniente denotar a las log-probabilidades *a posteriori* obtenidas por el decodificador sin hacer referencia a las observaciones originales. Las log-probabilidades *a posteriori* de los bits fuente y de los vectores $\mathbf{c}(k)$ se denotan por $L[u(k); O]$ y $L[\mathbf{c}(k); O]$, respectivamente.

En receptores con realimentación han de calcularse también las log-probabilidades *a posteriori* de los vectores de bits codificados, $L[\mathbf{c}(k); O]$, a partir de las cuales se pasan al ecualizador (en el orden apropiado, utilizando para ello un entrelazador) nuevas estimaciones de las log-probabilidades *a priori* $L[\mathbf{b}(k)]$. De modo análogo a como ocurría en el ecualizador, esto se hace restando a $L[\mathbf{c}(k); O] \approx L[\mathbf{c}(k)|\tilde{\mathbf{x}}] = \mathbf{c} + L[\tilde{\mathbf{x}}|\mathbf{c}(k)] + L[\mathbf{c}(k)]$ el término $L[\tilde{\mathbf{x}}|\mathbf{c}(k)]$.

Este proceso de eliminar la contribución de la log-probabilidad *a priori* $L[\mathbf{b}(k)]$ (en el caso del ecualizador) y de la verosimilitud $L[\tilde{\mathbf{x}}|\mathbf{c}(k)]$ en las estimaciones intercambiadas entre ambos evita que se produzca el fenómeno denominado realimentación positiva (*positive feedback*), es decir, que una misma probabilidad sea utilizada dos veces en una misma etapa del proceso iterativo de ecualización/decodificación. Las probabilidades obtenidas una vez eliminada la contribución de la probabilidad entrante a cualquiera de los módulos se denominan información (o simplemente probabilidad) extrínseca (*extrinsic information*) en la literatura del principio Turbo.

En general, el código de canal utilizado puede ser de tipo muy diverso (bloque, convolucional, etc.). Aquí nos centraremos exclusivamente en el caso de códigos convolucionales, aunque las ideas aquí expuestas son inmediatamente generalizables a otros tipos de métodos de codificación. Recordaremos brevemente que un código convolucional de *constraint length* m_v y tamaño de ventana $m_v k_v$ admite una descripción mediante un *trellis* en el que los estados de la etapa k -ésima tienen la forma $s = (u_1(k - m_v + 1), u_2(k - m_v + 2), \dots, u_{k_v}(k))$, donde se ha realizado el cambio de notación para los bits fuente dado

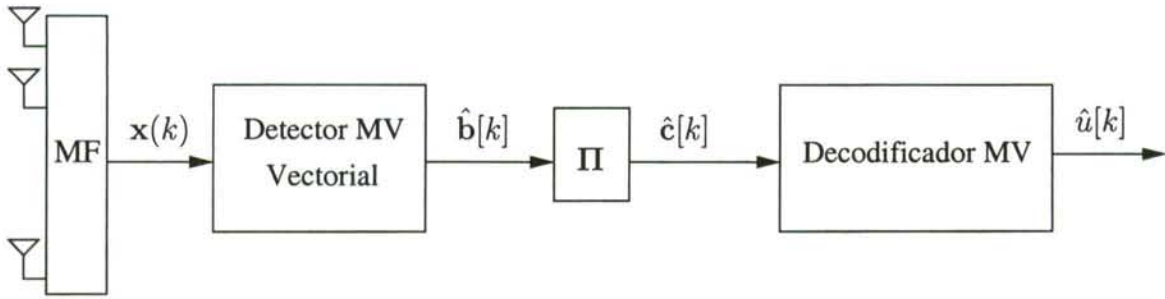


Figura 6.3: Receptor basado en detector y decodificador MV con decisiones *hard*

por $u_i(k) \equiv u((k-1)k_v + i)$. El número de estados es por tanto $2^{(m_v-1)k_v}$. En cada etapa se genera un vector de bits codificados $[c_1(D), \dots, c_{n_v}(D)] = [u_1(D), \dots, u_{k_v}(D)]\mathbf{G}(D)$ donde

$$\mathbf{G}(D) = \begin{bmatrix} g_{11}(D) & g_{12}(D) & \cdots & g_{1n_v}(D) \\ g_{21}(D) & g_{22}(D) & \cdots & g_{2n_v}(D) \\ \vdots & \vdots & \ddots & \vdots \\ g_{k_v1}(D) & g_{k_v2}(D) & \cdots & g_{k_vn_v}(D) \end{bmatrix} \quad (6.1)$$

es la matriz que contiene los polinomios generadores del código convolucional. La razón del código es

$$R_c = \frac{k_v}{n_v \left(1 + \frac{m_v-1}{K}\right)} \quad (6.2)$$

que, cuando $K \gg m_v$, como sucede habitualmente, $R_c \approx k_v/n_v$.

6.3. Receptor basado en detector y decodificador MV con decisiones *hard*

La decodificación de códigos convolucionales se puede realizar eficientemente y de modo óptimo (en el sentido MAP, equivalente al MV al ser los bits fuente equiprobables) mediante el algoritmo de Viterbi. La complejidad del algoritmo es lineal en la longitud del bloque y exponencial en la *constraint length* del código.

La métrica a utilizar en cada etapa es

$$\lambda(\mathbf{c} = c_1, c_2, \dots, c_{n_v}) = D(\mathbf{c}, \mathbf{x}) \quad (6.3)$$

donde \mathbf{c} y \mathbf{x} denotan el vector de bits transmitido y el vector de observaciones recibido, respectivamente, en dicha etapa. La función de distancia $D(\cdot)$ depende del tipo de canal considerado. En el caso que nos ocupa, el vector \mathbf{x} contiene estimaciones *hard* de los bits codificados, con lo que la función de distancia viene dada por la distancia de Hamming entre los dos vectores. La Figura 6.3 muestra el diagrama de bloques correspondiente a esta técnica de igualación-decodificación.

6.4. Ecualización Turbo

El hecho de que el módulo de ecualización produzca decisiones *hard* significa una importante pérdida de información para el proceso de decodificación. La primera solución en esta línea fue aportada por Hagenauer [101], pero sin duda la aplicación del principio Turbo al proceso de ecualización y decodificación ha llevado a sus últimas consecuencias la idea de intercambio de información entre ambos módulos.

Como ya se comentó en la Sección 6.2, la tarea del ecualizador es obtener las log-probabilidades *a posteriori* $L[\mathbf{b}(k)|\tilde{\mathbf{x}}]$. Teniendo en cuenta que el efecto conjunto de la modulación y del canal puede describirse con una máquina de estados finitos (*Finite State Machine*, FSM), la aplicación del algoritmo *Forward-Backward* sobre la misma permite llevar a cabo de modo eficiente el cálculo de las log-probabilidades *a posteriori* deseadas.

Es interesante decir que es posible generalizar el algoritmo a *trellis* en general. Es más, el algoritmo *Forward-Backward*, también denominado BCJR, puede verse como un caso particular del algoritmo de Baum-Welch, que permite calcular las distribuciones que forman un modelo oculto de Markov (*Hidden Markov Model*, HMM). El algoritmo de Baum-Welch fue propuesto por Leonard Baum y Lloyd Welch en los años sesenta [36]. En 1974 Bahl, Cocke, Jelinek y Raviv (BCJR) aplicaron el algoritmo de Baum-Welch al problema de la decodificación de códigos convolucionales.

La FSM que describe el efecto conjunto de la modulación y el canal tiene $2^{n_t(m_c-1)m'_s}$ estados, donde $m'_s = m_s$ ó $m'_s = M_c$ si la modulación es con o sin memoria, respectivamente. Sea $e_k = (s_{k-1}, \mathbf{b}(k), \mathbf{o}(k), s_k)$ una de las posibles $2^{n_t m_c m'_s}$ posibles transiciones de la FSM en la etapa k . Esta transición depende de cuatro parámetros: el estado de partida s_{k-1} , el estado de llegada s_k , el vector de bits de entrada $\mathbf{b}(k)$ y el vector de símbolos de salida $\mathbf{o}(k) = \mathbf{H}\mathbf{z}(k)$. Introduciremos la notación $\mathbf{b}(k) = L_{in}(e_k)$ y $\mathbf{o}(k) = L_{out}(e_k)$ para representar los vectores de entrada y de salida asociados a la transición e_k , respectivamente. Nótese que el vector de salida no depende del estado de llegada, s_k , de modo que cambiaremos ligeramente nuestra notación y escribiremos

$$\mathbf{o}(k) = L_{out}(e_k) = L_{out}(s_{k-1}, \mathbf{b}(k)) = L_{out}(\mathbf{z}(k)) = \mathbf{H}\mathbf{z}(k) \quad (6.4)$$

El primer paso para la aplicación del algoritmo *Forward-Backward* consiste en darse cuenta que

$$L[\mathbf{b}(k)|\tilde{\mathbf{x}}] = L[\mathbf{b}(k), \tilde{\mathbf{x}}] + h_b \quad (6.5)$$

donde h_b es la constante que hace que $P[\mathbf{b}(k)|\tilde{\mathbf{x}}]$ sea una función de masa de probabilidad y

$$L[\mathbf{b}(k), \tilde{\mathbf{x}}] = \log \sum_{e_k: L_{in}(e_k)=\mathbf{b}(k)} \exp L[e_k, \tilde{\mathbf{x}}] \quad (6.6)$$

es la log-probabilidad conjunta de la transición e_k y el conjunto de observaciones disponible, $\tilde{\mathbf{x}}$. Esta log-probabilidad conjunta puede ser expresada mediante

$$L[e_k, \tilde{\mathbf{x}}] = \alpha_{k-1}[s_{k-1}] + \gamma_k[e_k] + \beta_k[s_k] \quad (6.7)$$

donde

$$\alpha_{k-1}[s_{k-1}] = L[s_{k-1}, \tilde{\mathbf{x}}_k^-] \quad (6.8)$$

$$\gamma_k[e_k] = L[\mathbf{b}(k)] + L[\mathbf{x}(k)|\mathbf{o}(k)] \quad (6.9)$$

$$\beta_k[s_k] = L[\tilde{\mathbf{x}}_k^+|s_k] \quad (6.10)$$

con

$$L[\mathbf{x}(k)|\mathbf{o}(k)] = -\frac{1}{\sigma_g^2} \|\mathbf{x}(k) - \mathbf{H}\mathbf{z}(k)\|^2 \quad (6.11)$$

$$\tilde{\mathbf{x}}_k^- = [\mathbf{x}^T(1) \mathbf{x}^T(2) \cdots \mathbf{x}^T(k-1)] \quad (6.12)$$

$$\tilde{\mathbf{x}}_k^+ = [\mathbf{x}^T(k+1) \mathbf{x}^T(k+2) \cdots \mathbf{x}^T(M)] \quad (6.13)$$

Nótese que en la Ecuación (6.11) es necesario conocer la varianza del ruido σ_g^2 . El cálculo de las cantidades $\alpha_{k-1}[s_{k-1}]$, $\gamma_k[e_k]$ y $\beta_k[s_k]$ puede ser llevado a cabo recursivamente mediante una primera recursión hacia adelante (*forward*)

$$\alpha_{k-1}[s_{k-1}] = \log \sum_{\substack{\mathbf{b}^{(k), s_{k-2}: \\ f_{next}(s_{k-2}, \mathbf{b}^{(k-1)})=s_{k-1}}} \exp \{ \alpha_{k-2}[s_{k-2}] + L[\mathbf{b}(k-1)] + L[\mathbf{x}(k)|\mathbf{o}(k)] \} \quad (6.14)$$

con condiciones iniciales $\alpha_1[s_1 = 0] = 0$ y $\alpha_1[s_1 \neq 0] = -\infty$, continuando con una recursión hacia atrás (*backward*)

$$\beta_k[s_k] = \log \sum_{\substack{\mathbf{b}^{(k+1), s_{k+1}: \\ f_{next}(s_k, \mathbf{b}^{(k+1)})=s_{k+1}}} \exp \{ \beta_{k+1}[s_{k+1}] + L[\mathbf{b}(k+1)] + L[\mathbf{x}(k+1)|\mathbf{o}(k+1)] \} \quad (6.15)$$

con condiciones iniciales $\beta_M[s = s_M] = 0$ and $\beta_M[s \neq s_M] = -\infty$. La función $f_{next}(\cdot)$ describe las posibles transiciones de la FSM.

De modo similar, el decodificador tiene que calcular las log-probabilidades *a posteriori* de los bits fuente, $L[u(k); O]$, a partir de sus log-probabilidades *a priori* $L[u(k)] = \log(0,5)$ y de las medidas de log-verosimilitud $\lambda[\mathbf{c}(k)]$ que provienen del ecualizador. El algoritmo BCJR se aplica de modo totalmente análogo al proceso de ecualización, simplemente teniendo en cuenta la FSM asociada al código convolucional. En este caso es necesario calcular también las log-probabilidades *a posteriori* de los bits codificados, $L[\mathbf{c}(k); O]$, mediante

$$L[\mathbf{c}(k); O] = \log \sum_{e_k: L_{out}(e_k)=\mathbf{c}(k)} \exp \{ \alpha_{k-1}[s_{k-1}] + \gamma_k[s_k] + \beta_k[s_k] \}$$

donde la métrica utilizada en cada transición es $\lambda[\mathbf{c}(k)]$. Estas log-probabilidades *a posteriori* son entonces realimentadas al ecualizador para actuar como las log-probabilidades *a priori* $L[\mathbf{b}(k)]$. Como ya se comentó anteriormente, es siempre necesario abstraer la componente *a priori* de las log-probabilidades calculadas antes de introducirlas en el siguiente módulo para evitar el efecto de *positive feedback*.

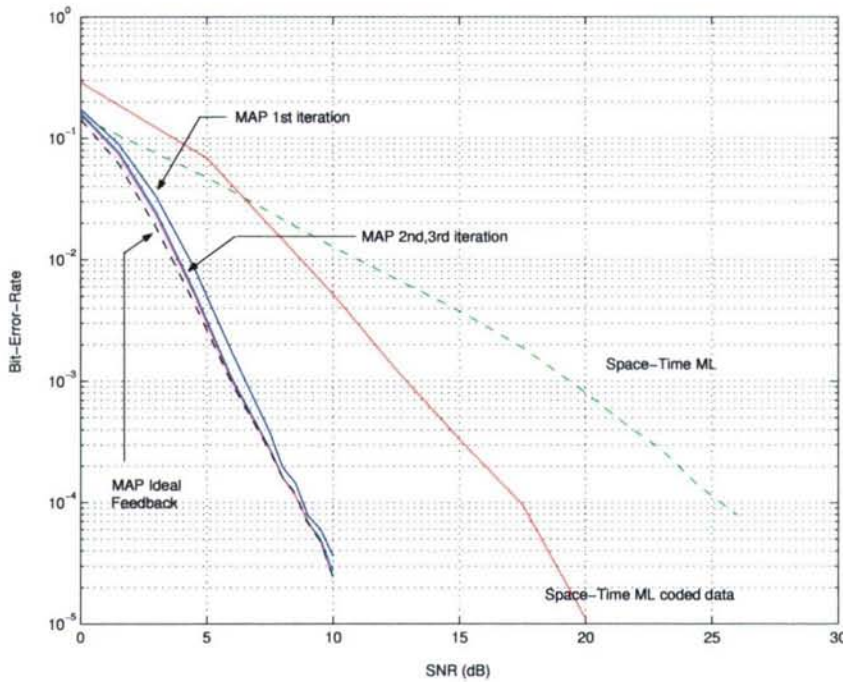


Figura 6.4: Rendimiento del ecualizador Turbo iterativo frente al Viterbi de *hard decisions* (2x2 antenas)

6.5. Resultados de simulación

La Figura 6.5 compara el rendimiento de los receptores aquí presentados para el canal MIMO 2x2 medido experimentalmente en el Metro de París (configuración *one way*). El código de canal utilizado es convolucional binario con matriz generadora $\mathbf{G} = [5, 7]$ (en representación octal) que extrae diversidad completa [92]. Las ventajas de incorporar codificación son patentes, así como la diferencia en rendimiento entre la aproximación clásica basada en algoritmos de Viterbi con decisiones *hard* y el igualador Turbo. La Figura 6.5 muestra el BER frente a la SNR de los mismos esquemas pero para el canal MIMO 4x4. La matriz generadora del código convolucional viene dada en este caso por $\mathbf{G} = [52, 56, 66, 76]$.

Se observa también que, aunque el rendimiento del igualador Turbo es muy bueno en la primera iteración, no se obtiene una mejora significativa al realizar más iteraciones. Algunas razones que justifican este comportamiento del formato de modulación GMSK pueden encontrarse en [102]. Sin embargo, este fenómeno no quiere decir que el esquema no funcione: hemos incluido también en las figuras el BER obtenido por un decodificador MAP ideal, i. e., aquél que permite alimentar el ecualizador con las verdaderas probabilidades *a priori* de los símbolos de entrada al modulador. Nótese que no se aprecia una diferencia sustancial entre las curvas del decodificador MAP ideal y las curvas correspondientes a la segunda y tercera iteraciones.

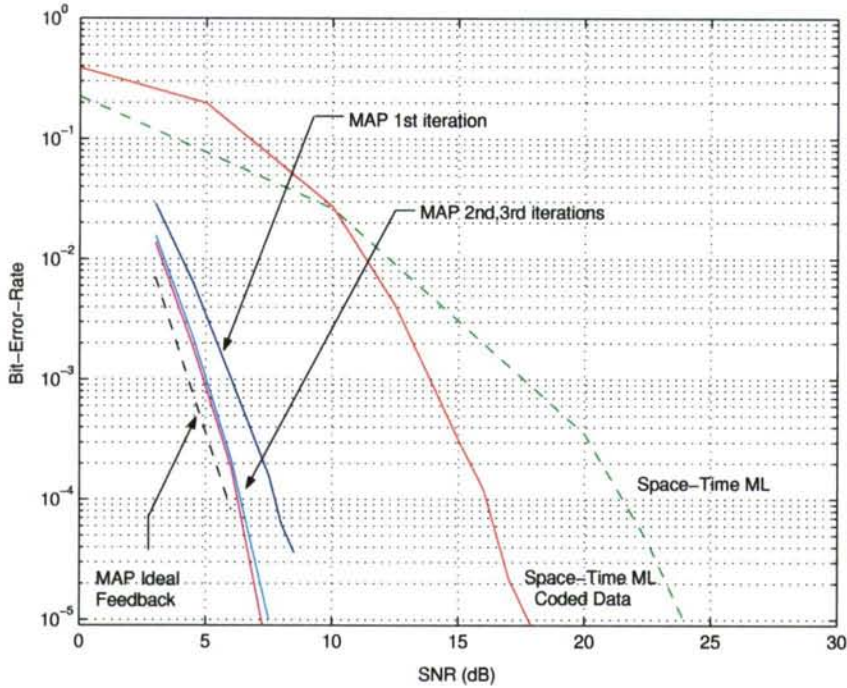


Figura 6.5: Rendimiento del ecualizador Turbo iterativo frente al Viterbi de *hard decisions* (4×4 antenas)

6.6. Conclusiones

En el presente capítulo hemos abordado los sistemas MIMO con codificación de canal. Se ha considerado en primer lugar el esquema clásico consistente en un ecualizador de Máxima Verosimilitud (MV) implementado mediante el algoritmo de Viterbi, cuyas decisiones *hard* acerca de los bits codificados son usadas por el decodificador de canal para estimar la secuencia de bits fuente original.

Es claro que el hecho de tomar una decisión *hard* tras el algoritmo de Viterbi acerca de los bits transmitidos degrada notablemente el rendimiento. Hemos visto cómo utilizando el algoritmo BCJR podemos obtener estimaciones de las probabilidades *a posteriori* de esos bits que pueden ser usadas directamente por el decodificador. Además, aprovechando la existencia de un entrelazador que separa el codificador de canal y el modulador en transmisión, hemos mostrado las ventajas de realizar el proceso de ecualización y decodificación de modo iterativo. Esto es lo que se conoce precisamente como principio Turbo.

Los resultados de simulación utilizando la respuesta al impulso MIMO real correspondiente a los túneles del Metro de París muestran las grandes ventajas de utilizar el principio Turbo en lugar de la aproximación clásica basada en algoritmos de Viterbi con decisiones *hard*.

Capítulo 7

Técnicas de ecualización con canal desconocido

*New ideas pass through three periods: *It can't be done. *It probably can be done, but it's not worth doing *I knew it was a good idea all along!*

—Arthur C. Clarke

El caso de suponer desconocido el canal en el receptor, es solucionado desde dos perspectivas diferentes: o bien se incorpora un módulo de estimación de canal al ecualizador, o bien se construye directamente el ecualizador teniendo en cuenta esta suposición. En este capítulo presentamos nuevas soluciones bajo ambas perspectivas. En ellas hacemos uso del principio de Máxima Verosimilitud (MV) para estimar los parámetros desconocidos del modelo. Dado que los problemas de optimización resultantes carecen de solución cerrada, utilizamos el algoritmo *Expectation-Maximization* (EM) para obtener de modo iterativo el valor de dicha solución. La gran novedad aquí presentada es la incorporación de los procesos de estimación a una estructura de Turbo igualación, ya que entonces los estimadores extraen un alto beneficio de las probabilidades *a posteriori* calculadas en el proceso iterativo Turbo.

7.1. Estimación de canal Máximo-Verosímil

La estimación de canal es una tarea de carácter obligado en cualquier implementación práctica de una estrategia de detección espacio-temporal. El método de estimación de canal aquí presentado aprovecha el hecho de que, en la mayoría de los sistemas prácticos, las tramas de datos transmitidas contienen una secuencia piloto determinista, cuya longitud denotaremos como N_t , conocida en el receptor para servir en la tarea de estimar el canal por parte de este último. Por ejemplo, en GSM esta secuencia tiene una longitud de $N_t = 26$ bits [24]. Sea $\tilde{\mathbf{b}}_f = [\tilde{\mathbf{b}}_t^T \tilde{\mathbf{b}}^T]^T$ la trama de datos, que incluye $\tilde{\mathbf{b}}_t = [\mathbf{b}_t^T(1) \mathbf{b}_t^T(2) \cdots \mathbf{b}_t^T(N_t)]^T$

como secuencia de entrenamiento y $\tilde{\mathbf{b}} = [\mathbf{b}^T(1) \mathbf{b}^T(2) \cdots \mathbf{b}^T(N)]^T$ como secuencia de información. Análogamente, $\tilde{\mathbf{x}}_f = [\tilde{\mathbf{x}}_t^T \tilde{\mathbf{x}}^T]^T$ son las observaciones correspondientes a una trama de datos, donde $\tilde{\mathbf{x}}_t = [\mathbf{x}_t^T(1) \mathbf{x}_t^T(2) \cdots \mathbf{x}_t^T(M_t)]^T$ representa la secuencia piloto y $\tilde{\mathbf{x}} = [\mathbf{x}(1) \mathbf{x}(2) \cdots \mathbf{x}(M)]^T$ corresponde a la secuencia de información. El estimador de Máxima Verosimilitud de la matriz de canal \mathbf{H} es por tanto

$$\hat{\mathbf{H}} = \arg \max_{\mathbf{H}} f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}_t; \mathbf{H}}(\tilde{\mathbf{x}}) \tag{7.1}$$

donde $f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}_t; \mathbf{H}}$ es la pdf de las observaciones condicionada a la información disponible (la secuencia de entrenamiento $\tilde{\mathbf{b}}_t$) y los parámetros a estimar (la matriz de canal \mathbf{H}). Aunque éste es un problema para el que no existe solución cerrada, se puede emplear el algoritmo *Expectation Maximization* (EM) para solucionar (7.1) de modo iterativo. El algoritmo EM se basa en definir un conjunto llamado de “datos completos” formado por las variables observadas y por variables adicionales no observables. En cada iteración del algoritmo se calcula una estimación más refinada promediando la log-verosimilitud del conjunto de datos completos con respecto a la f.d.p. de las variables no observables condicionada al conjunto de observaciones disponible (también denominado conjunto de “datos incompletos”). Usando la terminología del algoritmo EM, definiremos el conjunto de datos completos como la unión de las observaciones y de la secuencia de bits transmitida (que es el conjunto de variables “no observables”), es decir, $\tilde{\mathbf{x}}_e = [\tilde{\mathbf{b}}_f^T \tilde{\mathbf{x}}_f^T]^T$, mientras que las observaciones $\tilde{\mathbf{x}}_f$ son el conjunto de datos incompletos. La relación entre $\tilde{\mathbf{x}}_e$ y $\tilde{\mathbf{x}}_f$ tiene que venir dada por una transformación lineal (posiblemente no invertible), es decir, $\tilde{\mathbf{x}}_f = \mathbf{T}\tilde{\mathbf{x}}_e$. Es fácil comprobar que en nuestro caso esta transformación lineal es $\mathbf{T} = [\mathbf{0}_{n_r(M_t+M) \times n_t(N_t+N)} \quad \mathbf{I}_{n_r(M_t+M)}]$. Teniendo en cuenta las definiciones previas, el estimador del canal en la iteración $i + 1$ -ésima se obtiene solucionando

$$\hat{\mathbf{H}}^{(i+1)} = \arg \max_{\mathbf{H}} E_{\tilde{\mathbf{x}}_e|\tilde{\mathbf{x}}_f, \tilde{\mathbf{b}}_t; \hat{\mathbf{H}}^{(i)}} \left\{ \log f_{\tilde{\mathbf{x}}_e|\tilde{\mathbf{b}}_t; \mathbf{H}}(\tilde{\mathbf{x}}_e) \right\} \tag{7.2}$$

donde $E_f\{\cdot\}$ denota el operador esperanza con respecto a la f.d.p. $f(x)$. Expandiendo la expresión anterior

$$\begin{aligned} \hat{\mathbf{H}}^{(i+1)} &= \arg \max_{\mathbf{H}} E_{\tilde{\mathbf{b}}|\tilde{\mathbf{x}}; \hat{\mathbf{H}}^{(i)}} \left\{ \log \left[f_{\tilde{\mathbf{x}}_f|\tilde{\mathbf{b}}_f; \mathbf{H}}(\tilde{\mathbf{x}}_f) f_{\tilde{\mathbf{b}}}(\tilde{\mathbf{b}}) \right] \right\} \\ &= \arg \max_{\mathbf{H}} E_{\tilde{\mathbf{b}}|\tilde{\mathbf{x}}; \hat{\mathbf{H}}^{(i)}} \left\{ \log \left[f_{\tilde{\mathbf{x}}_t|\tilde{\mathbf{b}}_t; \mathbf{H}}(\tilde{\mathbf{x}}_t) f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}; \mathbf{H}}(\tilde{\mathbf{x}}) \right] \right\} \\ &= \arg \max_{\mathbf{H}} \log f_{\tilde{\mathbf{x}}_t|\tilde{\mathbf{b}}_t; \mathbf{H}}(\tilde{\mathbf{x}}_t) + E_{\tilde{\mathbf{b}}|\tilde{\mathbf{x}}; \hat{\mathbf{H}}^{(i)}} \left\{ \log f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}; \mathbf{H}}(\tilde{\mathbf{x}}) \right\} \end{aligned}$$

donde la última igualdad se apoya en el hecho de que, siempre que supongamos AWGN, la f.d.p. de las observaciones condicionada a los símbolos transmitidos, $f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}; \hat{\mathbf{H}}^{(i)}}$, es gaussiana. Se llega por tanto al siguiente problema de optimización cuadrático

$$\begin{aligned} \hat{\mathbf{H}}^{(i+1)} &= \arg \min_{\mathbf{H}} \sum_{k=1}^{M_t} \|\mathbf{x}_t(k) - \mathbf{H}\mathbf{z}_t(k)\|^2 \\ &\quad + \sum_{k=1}^M E_{\mathbf{z}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}^{(i)}} \left\{ \|\mathbf{x}(k) - \mathbf{H}\mathbf{z}(k)\|^2 \right\} \end{aligned} \tag{7.3}$$

cuya solución cerrada es¹

$$\hat{\mathbf{H}}^{(i+1)} = (\mathbf{R}_{xz,t} + \mathbf{R}_{xz}) \times (\mathbf{R}_{z,t} + \mathbf{R}_z)^{-1} \quad (7.4)$$

donde

$$\mathbf{R}_{xz,t} = \sum_{k=1}^{M_t} \mathbf{x}_t(k) \mathbf{z}_t^H(k) \quad (7.5)$$

$$\mathbf{R}_{z,t} = \sum_{k=1}^{M_t} \mathbf{z}_t(k) \mathbf{z}_t^H(k) \quad (7.6)$$

$$\mathbf{R}_{xz} = \sum_{k=1}^M E_{\mathbf{z}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}_i} \{ \mathbf{x}(k) \mathbf{z}^H(k) \} \quad (7.7)$$

$$\mathbf{R}_z = \sum_{k=1}^M E_{\mathbf{z}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}_i} \{ \mathbf{z}(k) \mathbf{z}^H(k) \} \quad (7.8)$$

Nótese que para calcular (7.7) y (7.8) es necesario conocer la f.m.p. $p_{\mathbf{z}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}^{(i)}}$. Es aquí donde podemos beneficiarnos del proceso de ecualización Turbo ya que

$$L[\mathbf{z}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}^{(i)}] = L[\mathbf{z}(k), \tilde{\mathbf{x}}; \hat{\mathbf{H}}^{(i)}] + h_z = L[e_k, \tilde{\mathbf{x}}] + h_z \quad (7.9)$$

donde h_z es la constante que hace que $p_{\mathbf{z}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}^{(i)}}$ sea una f.m.p. y $L[e_k, \tilde{\mathbf{x}}]$ es la log-probabilidad conjunta de la transición e_k y el conjunto de observaciones disponible. Esta cantidad ya se ha calculado en el proceso de ecualización turbo (ver Ecuación (6.7)). Este hecho hace que el estimador de canal propuesto encaje a la perfección dentro de una estructura de ecualización turbo.

7.1.1. Aplicación a sistemas de tipo GSM con diversidad espacial

Nos centraremos ahora en la aplicación del estimador de canal MV-EM descrito previamente sistemas de tipo GSM con diversidad espacial para sistemas de transporte ferroviario subterráneos. Como ya se explicó anteriormente, las condiciones de propagación en el interior de un túnel subterráneo se pueden modelar convenientemente por un modelo de canal multitrayecto con desvanecimiento plano, ya que el retardo temporal (*delay spread*) es pequeño comparado con el período de símbolo utilizado en GSM. Sin embargo, la modulación GMSK empleada en el estándar GSM introduce ISI controlada, y por tanto se puede

¹Al ser lineal el operador esperanza, la derivación que conduce a la Ecuación (7.4) sigue, paso por paso, el procedimiento de optimización usual de encontrar la estimación de mínimos cuadrados de un sistema lineal dado un conjunto de observaciones ruidosas (cf. [103]). Este procedimiento incluye el cálculo del gradiente respecto a los coeficientes del sistema y la resolución del sistema de ecuaciones resultante de igualar dicho gradiente a cero. Por tanto, la solución de (7.3) es tediosa, ya que hay que calcular la derivada con respecto a cada coeficiente de la matriz \mathbf{H} , pero conceptualmente sencilla.

emplear ecualización espacio-temporal turbo para realizar conjuntamente la ecualización y la decodificación. Mostraremos cómo el estimador de canal propuesto permite reducir la longitud de la secuencia de entrenamiento empleada de los 26 bits del estándar GSM a tan sólo 5 bits, manteniendo el rendimiento muy cercano al óptimo (es decir, a la tasa de error de bit (*Bit-Error-Rate*, BER) obtenida considerando que el canal es perfectamente conocido por el receptor), lo que implica una gran ganancia en la eficiencia total del sistema.

El modelo de las observaciones que representa la Ecuación (4.21) no pone de manifiesto el hecho de que la ISI es introducida por la modulación GMSK en lugar de por el canal multitrayecto. Estimar el canal de acuerdo a la dicha ecuación y la posterior aplicación del método descrito previamente es altamente ineficiente porque se requiere la estimación de un número de parámetros innecesariamente elevado. Además, de este modo no se está teniendo en cuenta el conocimiento de la ISI controlada introducida por el modulador. Por todo ello es conveniente que reescribamos la Ecuación (4.21) del siguiente modo

$$\mathbf{x}(k) = \mathbf{H}_c \mathbf{B}(k) \mathbf{f} + \mathbf{v}(k) \tag{7.10}$$

donde $\mathbf{B}(k) = [\mathbf{b}(k) \ \mathbf{b}(k-1) \ \mathbf{b}(k-2)]$ y $\mathbf{f} = [0,8053 \quad -0,5853j \quad -0,0704]^T$ denota la ISI introducida por la modulación.

La Ecuación (7.10) nos permite formular el problema de estimar exclusivamente los parámetros desconocidos del modelo, es decir, los coeficientes h_{li} del canal. Asumiremos de nuevo que cada trama de datos contiene una secuencia piloto conocida por el receptor de N_t bits de longitud. El estimador de Máxima Verosimilitud del canal viene dado por

$$\hat{\mathbf{H}}_c = \arg \max_{\mathbf{H}_c} f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}_t; \mathbf{H}_c}(\tilde{\mathbf{x}}) \tag{7.11}$$

Este problema no tiene solución cerrada, por lo cual aplicaremos el algoritmo EM de un modo similar al caso general previamente explorado. Definiremos los conjuntos de datos completos e incompletos como $\tilde{\mathbf{x}}_e = [\tilde{\mathbf{b}}_f^T \ \tilde{\mathbf{x}}_f^T]^T$ y $\tilde{\mathbf{x}}_f$, respectivamente. Ambos conjuntos están relacionados por la transformación lineal $\tilde{\mathbf{x}}_f = \mathbf{T} \tilde{\mathbf{x}}_e$ donde $\mathbf{T} = [\mathbf{0}_{n_r(M_t+M) \times n_t(N_t+N)} \ \mathbf{I}_{n_r(M_t+M)}]$. Usando estas definiciones, la estimación i -ésima del canal se calcula usando el algoritmo EM como

$$\hat{\mathbf{H}}_c^{(i+1)} = \arg \max_{\mathbf{H}_c} E_{\tilde{\mathbf{x}}_e|\tilde{\mathbf{x}}_f, \tilde{\mathbf{b}}_t; \hat{\mathbf{H}}_c^{(i)}} \left\{ \log f_{\tilde{\mathbf{x}}_e|\tilde{\mathbf{b}}_t; \mathbf{H}_c}(\tilde{\mathbf{x}}_e) \right\}. \tag{7.12}$$

Mediante manipulaciones similares a las para el caso general, llegamos al siguiente problema de optimización

$$\begin{aligned} \hat{\mathbf{H}}_c^{(i+1)} = \arg \min_{\mathbf{H}_c} & \sum_{k=1}^{M_t} \|\mathbf{x}_t(k) - \mathbf{H}_c \mathbf{B}_t(k) \mathbf{f}\|^2 \\ & + \sum_{k=1}^M E_{\mathbf{b}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}_c^{(i)}} \left\{ \|\mathbf{x}(k) - \mathbf{H}_c \mathbf{B}(k) \mathbf{f}\|^2 \right\} \end{aligned} \tag{7.13}$$

que también es un problema de optimización cuadrático cuya solución es

$$\hat{\mathbf{H}}_c^{(i+1)} = (\mathbf{R}_{x_b,t} + \mathbf{R}_{x_b}) \times (\mathbf{R}_{b,t} + \mathbf{R}_b)^{-1} \tag{7.14}$$

donde

$$\mathbf{R}_{xb,t} = \sum_{k=1}^{M_t} \mathbf{x}_t(k) (\mathbf{B}_t(k) \mathbf{f})^H \quad (7.15)$$

$$\mathbf{R}_{b,t} = \sum_{k=1}^{M_t} (\mathbf{B}_t(k) \mathbf{f}) (\mathbf{B}_t(k) \mathbf{f})^H \quad (7.16)$$

$$\mathbf{R}_{xb} = \sum_{k=1}^M E_{\mathbf{b}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}_{c_i}} \{ \mathbf{x}(k) (\mathbf{B}(k) \mathbf{f})^H \} \quad (7.17)$$

$$\mathbf{R}_b = \sum_{k=1}^M E_{\mathbf{b}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}_{c_i}} \{ (\mathbf{B}(k) \mathbf{f}) (\mathbf{B}(k) \mathbf{f})^H \} \quad (7.18)$$

Aquí necesitamos promediar con respecto a la f.d.p. $f_{\mathbf{B}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}_c^{(i)}}$. De nuevo, podemos beneficiarnos del proceso de ecualización turbo ya que

$$L[\mathbf{B}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}_c^{(i)}] = L[\mathbf{B}(k), \tilde{\mathbf{x}}; \hat{\mathbf{H}}_c^{(i)}] + h_B = L[e_k, \tilde{\mathbf{x}}] + h_B \quad (7.19)$$

donde h_B es la constante que hace que $p_{\mathbf{B}(k)|\tilde{\mathbf{x}}; \hat{\mathbf{H}}_c^{(i)}}$ sea una f.m.p. y $L[e_k, \tilde{\mathbf{x}}]$ es una cantidad previamente calculada en el proceso de ecualización turbo.

7.2. Ecualización *Decision Feedback* basada en el principio MV

Una alternativa al proceso de estimación de canal y detección es la ecualización. Consideremos el siguiente modelo para nuestro ecualizador *Decision Feedback* (DFE)

$$\mathbf{y}(k) = \mathbf{W}^H \mathbf{x}_m(k) + \mathbf{V}^H \hat{\mathbf{s}}_v(k), \quad k = 1, 2, \dots, M \quad (7.20)$$

donde $\mathbf{y}(k) = [y_1(k) \ y_2(k) \ \dots \ y_{n_t}(k)]^T$ es el vector resultante del proceso de filtrado, \mathbf{W} es el filtro lineal hacia adelante, \mathbf{V} el filtro lineal hacia atrás, $\mathbf{x}_m(k) = [\mathbf{x}^T(k) \ \mathbf{x}^T(k+1) \ \dots \ \mathbf{x}^T(k+m_c-1)]^T$ las observaciones que contienen información acerca del vector de símbolos transmitidos $\mathbf{s}(k)$, y $\hat{\mathbf{s}}_v(k) = [\hat{\mathbf{s}}^T(k-m_c+1) \ \hat{\mathbf{s}}^T(k-m_c+2) \ \dots \ \hat{\mathbf{s}}^T(k-1)]^T$ estimaciones de los símbolos que interfieren con el vector de símbolos $\mathbf{s}(k)$.

Para derivar una expresión de los filtros del ecualizador, empezaremos considerando que existen los filtros lineales \mathbf{W}_* y \mathbf{V}_* capaces de eliminar completamente las interferencias de co-canal e intersimbólica. Bajo esta suposición tenemos

$$\mathbf{y}(k) = \mathbf{s}(k) + \mathbf{g}_f(k), \quad k = 1, 2, \dots, M \quad (7.21)$$

donde $\mathbf{g}_f(k)$ es el AWGN del canal. Por razones de simplicidad, supondremos que $E[\mathbf{g}_f(k) \mathbf{g}_f(k-n)^H] = \sigma_f^2 \mathbf{I} \delta(n)$ lo que nos permite escribir la f.d.p. del vector filtrado $\mathbf{y}(k)$

como sigue [104]

$$\begin{aligned} f_{\mathbf{y}(k); \Theta_*}(\mathbf{y}(k)) &= E_{\mathbf{s}(k)} \{ f_{\mathbf{y}(k)|\mathbf{s}(k); \Theta_*}(\mathbf{y}(k)) \} \\ &\propto E_{\mathbf{s}(k)} \left\{ \exp \left[-\frac{1}{\sigma_f^2} \|\mathbf{y}(k) - \mathbf{s}(k)\|^2 \right] \right\} \end{aligned}$$

donde $\Theta_* = [\mathbf{W}_*, \mathbf{V}_*]$ representa los coeficientes óptimos de filtrado. La f.d.p. conjunta de los vectores filtrados $\tilde{\mathbf{y}} = [\mathbf{y}^T(1) \mathbf{y}^T(2) \cdots \mathbf{y}^T(M)]^T$ es igual al producto de las densidades marginales, es decir,

$$\begin{aligned} f_{\tilde{\mathbf{y}}; \Theta_*}(\tilde{\mathbf{y}}) &= \prod_{k=1}^M f_{\mathbf{y}(k); \Theta_*}(\mathbf{y}(k)) \\ &\propto \prod_{k=1}^M E_{\mathbf{s}(k)} \left\{ \exp \left[-\frac{1}{\sigma_f^2} \|\mathbf{y}(k) - \mathbf{s}(k)\|^2 \right] \right\} \end{aligned}$$

Los coeficientes óptimos de filtrado pueden ser estimados ahora de acuerdo al principio de Máxima Verosimilitud

$$\begin{aligned} \hat{\Theta} &= \arg \max_{\Theta_*} \log f_{\tilde{\mathbf{y}}; \Theta_*}(\tilde{\mathbf{y}}) \\ &= \arg \max_{\Theta_*} \sum_{k=1}^M \log E_{\mathbf{s}(k)} \left\{ \exp \left[-\frac{1}{\sigma_f^2} \|\mathbf{y}(k) - \mathbf{s}(k)\|^2 \right] \right\} \end{aligned}$$

Éste es un problema de optimización para el que existe solución en forma cerrada. Mostraremos a continuación como resolverlo utilizando una generalización del algoritmo EM, denominada algoritmo *Space Alternating Generalized EM* (SAGE).

7.2.1. Algoritmo SAGE

El algoritmo SAGE [105] es una variante del algoritmo EM de menor complejidad y mayor velocidad de convergencia, que es especialmente útil cuando el conjunto total de parámetros a estimar puede ser particionado en subconjuntos disjuntos. Haciendo uso de la terminología empleada en el algoritmo EM, definiremos el conjunto de datos completos como $\tilde{\mathbf{y}}_e = [\mathbf{s}^T(1) \mathbf{y}^T(1) \mathbf{s}^T(2) \mathbf{y}^T(2) \cdots \mathbf{s}^T(M) \mathbf{y}^T(M)]^T$, mientras que el conjunto de datos incompletos está formado por las observaciones $\tilde{\mathbf{y}}$.

Antes de continuar con la aplicación del algoritmo SAGE, es necesario mencionar que la varianza del ruido a la salida del filtro, σ_f^2 , debe ser considerada como un parámetro a estimar adicional, ya que es función de \mathbf{W}_* , que es desconocido *a priori*. Por tanto, redefiniremos nuestro conjunto de parámetros a estimar como $\Theta_* = [\mathbf{W}_*, \mathbf{V}_*, \sigma_f^2]$, donde \mathbf{W}_* , \mathbf{V}_* and σ_f^2 se consideran como disjuntos a lo largo del proceso de estimación. De acuerdo a la formulación del

algoritmo SAGE, la estimación $i + 1$ -ésima de \mathbf{W}_* se obtiene resolviendo el siguiente problema de optimización

$$\hat{\mathbf{W}}_{i+1} = \arg \max_{\mathbf{W}_*} E_{\tilde{\mathbf{y}}_e | \tilde{\mathbf{y}}; [\hat{\mathbf{W}}_i, \hat{\mathbf{V}}_i, \hat{\sigma}_{f,i}^2]} \left\{ \log f_{\tilde{\mathbf{y}}_e; [\mathbf{W}_*, \hat{\mathbf{V}}_i, \hat{\sigma}_{f,i}^2]}(\tilde{\mathbf{y}}_e) \right\} \quad (7.22)$$

Después de algunas manipulaciones llegamos a

$$\begin{aligned} \hat{\mathbf{W}}_{i+1} = \arg \min_{\mathbf{W}_*} \sum_{k=1}^M \mathbf{x}_m^H(k) \mathbf{W}_* \mathbf{W}_*^H \mathbf{x}_m(k) + \mathbf{x}_m^H(k) \mathbf{W}_* \hat{\mathbf{V}}_i^H \hat{\mathbf{s}}_v(k) + \hat{\mathbf{s}}_v^H(k) \hat{\mathbf{V}}_i^H \mathbf{W}_*^H \mathbf{x}_m(k) \\ - \mathbf{x}_m^H(k) \mathbf{W}_* E_{ii} \{ \mathbf{s}(k) \} - E_{ii} \{ \mathbf{s}^H(k) \} \mathbf{W}_*^H \mathbf{x}_m(k) \end{aligned} \quad (7.23)$$

donde

$$E_{ij} \{ g(\mathbf{s}(k)) \} = E_{\mathbf{s}(k) | \mathbf{y}(k); [\hat{\mathbf{W}}_i, \hat{\mathbf{V}}_i, \hat{\sigma}_{f,i}^2]} \{ g(\mathbf{s}(k)) \} \quad (7.24)$$

siendo $g(\mathbf{s}(k))$ una función arbitraria de $\mathbf{s}(k)$. Nótese que esta esperanza puede calcularse usando la regla de Bayes como

$$E_{ij} \{ g(\mathbf{s}(k)) \} = \frac{E_{\mathbf{s}(k)} \left\{ g(\mathbf{s}(k)) \exp \left[-\frac{1}{\hat{\sigma}_{f,i}^2} \|\mathbf{y}_{ij}(k) - \mathbf{s}(k)\|^2 \right] \right\}}{E_{\mathbf{s}(k)} \left\{ \exp \left[-\frac{1}{\hat{\sigma}_{f,i}^2} \|\mathbf{y}_{ij}(k) - \mathbf{s}(k)\|^2 \right] \right\}} \quad (7.25)$$

donde

$$\mathbf{y}_{ij}(k) = \hat{\mathbf{W}}_i^H \mathbf{x}_m(k) + \hat{\mathbf{V}}_i^H \hat{\mathbf{s}}_v(k) \quad (7.26)$$

La Ecuación (7.23) representa un problema de optimización cuadrático cuya solución exacta es

$$\hat{\mathbf{W}}_{i+1} = \hat{\mathbf{R}}_x^{-1} (\hat{\mathbf{R}}_{xs} - \hat{\mathbf{R}}_{xv} \hat{\mathbf{V}}_i) \quad (7.27)$$

donde

$$\hat{\mathbf{R}}_x = \sum_{k=1}^M \mathbf{x}_m(k) \mathbf{x}_m^H(k) \quad (7.28)$$

$$\hat{\mathbf{R}}_{xs} = \sum_{k=1}^M \mathbf{x}_m(k) E_{ii} [\mathbf{s}^H(k)] \quad (7.29)$$

$$\hat{\mathbf{R}}_{xv} = \sum_{k=1}^M \mathbf{x}_m(k) \mathbf{s}_v^H(k) \quad (7.30)$$

Una vez que $\hat{\mathbf{W}}_{i+1}$ ha sido calculado, podemos obtener la estimación $i + 1$ -ésima de \mathbf{V}_* resolviendo el problema de optimización siguiente

$$\hat{\mathbf{V}}_{i+1} = \arg \max_{\mathbf{V}} E_{\tilde{\mathbf{y}}_e | \tilde{\mathbf{y}}; [\hat{\mathbf{W}}_{i+1}, \hat{\mathbf{V}}_i, \hat{\sigma}_{f,i}^2]} \left\{ \log f_{\tilde{\mathbf{y}}_e; [\hat{\mathbf{W}}_{i+1}, \mathbf{V}_*, \hat{\sigma}_{f,i}^2]}(\tilde{\mathbf{y}}_e) \right\} \quad (7.31)$$

que, de nuevo, es un problema de optimización cuadrático con la solución exacta siguiente

$$\hat{\mathbf{V}}_{i+1} = \hat{\mathbf{R}}_v^{-1}(\hat{\mathbf{R}}_{vs} - \hat{\mathbf{R}}_{xv}^H \hat{\mathbf{W}}_{i+1}) \tag{7.32}$$

donde

$$\hat{\mathbf{R}}_v = \sum_{k=1}^M \hat{\mathbf{s}}_v(k) \hat{\mathbf{s}}_v^H(k) \tag{7.33}$$

$$\hat{\mathbf{R}}_{vs} = \sum_{k=1}^M \hat{\mathbf{s}}_v(k) E_{i+1i} [\mathbf{s}^H(k)] \tag{7.34}$$

$$\tag{7.35}$$

Finalmente, dados $\hat{\mathbf{W}}_{i+1}$ y $\hat{\mathbf{V}}_{i+1}$, podemos calcular $\hat{\sigma}_{f,i}^2$ como

$$\hat{\sigma}_{f,i}^2 = \arg \max_{\sigma_f^2} E_{\tilde{\mathbf{y}}_e | \tilde{\mathbf{y}}_i; [\hat{\mathbf{W}}_{i+1}, \hat{\mathbf{V}}_{i+1}, \hat{\sigma}_{f,i}^2]} \left\{ \log f_{\tilde{\mathbf{y}}_e; [\hat{\mathbf{W}}_{i+1}, \hat{\mathbf{V}}_{i+1}, \hat{\sigma}_{f,i}^2]}(\tilde{\mathbf{y}}_e) \right\} \tag{7.36}$$

cuya solución es

$$\hat{\sigma}_{f,i}^2 = \frac{1}{Mn_t} \sum_{k=1}^M E_{i+1i+1} \left\{ \|\mathbf{y}_{i+1i+1}(k) - \mathbf{s}(k)\|^2 \right\} \tag{7.37}$$

7.2.2. Ecualizador DF turbo

Como mencionamos en el Capítulo 3, el ecualizador ST MAP debe calcular $L[\mathbf{b}(k)|\tilde{\mathbf{x}}]$. En nuestro DFE calculamos esas log-probabilidades después de suprimir las interferencias co-canal e intersimbólicas de las observaciones, es decir, $L[\mathbf{b}(k)|\tilde{\mathbf{x}}] \approx L[\mathbf{b}(k)|\tilde{\mathbf{y}}]$. Teniendo en cuenta la estructura de $\mathbf{y}(k)$ dada por (7.21) y usando la regla de Bayes, tenemos

$$\begin{aligned} L[\mathbf{s}(k)|\tilde{\mathbf{y}}] &= L[\mathbf{s}(k)|\mathbf{y}(k)] = L[\mathbf{y}(k)|\mathbf{s}(k)] + L[\mathbf{s}(k)] - L[\tilde{\mathbf{y}}] \\ &= -\frac{1}{\sigma_f^2} \|\mathbf{y}(k) - \mathbf{s}(k)\|^2 + L[\mathbf{s}(k)] + h_s \end{aligned} \tag{7.38}$$

donde h_s es la constante que hace que $P[\mathbf{s}(k)|\tilde{\mathbf{y}}]$ sea una f.m.p.. Claramente, la información extrínseca que debe ser comunicada al decodificador de canal es $L_e[\mathbf{s}(k)] = -\frac{1}{\sigma_f^2} \|\mathbf{y}(k) - \mathbf{s}(k)\|^2$.

Comentaremos brevemente cómo la información proporcionada por el decodificador es usada en nuestro DFE. Según la Ecuación (7.25) es claro que los valores *a priori*, $L[\mathbf{s}(k)]$, realimentados desde el decodificador juegan un papel clave en el cálculo de los parámetros del DFE. Por tanto, por cada iteración de decodificación, se pueden realizar una (o más) iteraciones del algoritmo SAGE para obtener estimaciones más refinadas de \mathbf{W}_* , \mathbf{V}_* y σ_f^2 . Es más, el filtro hacia atrás \mathbf{V}_* puede ser alimentado con estimaciones de los símbolos transmitidos provenientes del decodificador de canal, ya que se supone que son más precisas que las obtenidas a la salida del ecualizador. Por ejemplo, si la modulación empleada es BPSK, podemos estimar los símbolos transmitidos mediante $\hat{\mathbf{s}}(k) = \text{sign} [E_{s(k)}\{\mathbf{s}(k)\}]$.

7.3. Simulaciones numéricas

7.3.1. Estimación MV-EM para canales MIMO Rayleigh

La Figura 7.1 muestra el Bit-Error-Rate (BER) a la salida del decodificador obtenido para un sistema con codificación espacio-temporal con dos antenas transmisoras y dos antenas receptoras (2×2) sobre un canal no dispersivo. Los datos son transmitidos en bloques de 218 bits de los cuales la secuencia piloto ocupa $N_t = 10$ bits. Se muestran también, a modo de comparación, las curvas de rendimiento correspondientes al mismo sistema cuando se emplea el estimador de mínimos cuadrados del canal y cuando el canal es perfectamente conocido por el receptor. Nótese que cuando el canal es perfectamente conocido no hay ninguna ganancia por realizar varias iteraciones de decodificación, ya que, al no existir ISI, la primera iteración de decodificación equivale a la detección MAP óptima. Sin embargo, esto no es cierto cuando se utiliza el estimador de canal MV-EM, ya que el canal es reestimado en cada iteración del proceso de ecualización Turbo. El código de canal utilizado es un código convolucional binario espacio-temporal de *rate* $R_c = 1/2$ cuya matriz generadora es $\mathbf{G} = [46, 72]$ en representación octal [92]. El entrelazador utilizado en cada antena transmisora tiene una longitud de 20800 bits. El formato de modulación es BPSK y cada coeficiente del canal se modela como un proceso aleatorio gaussiano complejo, circularmente invariante y de medio cero. También se ha supuesto que los coeficientes del canal están incorrelados tanto espacial como temporalmente y que su varianza es $\sigma_h^2 = 1/2$ por dimensión compleja. La relación señal a ruido se define como

$$\text{SNR} = \frac{E \{ (\mathbf{H}\mathbf{z}(k))^H (\mathbf{H}\mathbf{z}(k)) \}}{E \{ \mathbf{v}^H(k) \mathbf{v}(k) \}} = \frac{\text{Tr} \{ \mathbf{H}\mathbf{H}^H \}}{L\sigma_v^2} \quad (7.39)$$

donde $\text{Tr} \{ \cdot \}$ denota el operador traza. El canal cambia para cada bloque transmitido, permaneciendo constante a lo largo de la transmisión de un bloque. La Figura 7.1 muestra que, incluso siendo muy pobre el resultado de la primera iteración, el estimador de canal MV-EM supera en rendimiento al estimador clásico de mínimos cuadrados a partir de la cuarta iteración de decodificación.

El mal rendimiento obtenido por el estimador MV-EM en la primera iteración proviene del hecho de que el ecualizador Turbo está utilizando una estimación inicial del canal totalmente carente de información. Específicamente, la Ecuación (7.4) se puede ver como un estimador de mínimos cuadrados en el que las matrices de correlación $\mathbf{R}_{xz,t}$ y $\mathbf{R}_{z,t}$ han sido modificadas por la adición de las matrices \mathbf{R}_{xz} and \mathbf{R}_z , respectivamente. En la primera iteración estas matrices son calculadas asumiendo que $p_{\mathbf{z}(k)|\hat{\mathbf{x}};\hat{\mathbf{H}}_i}$ es una f.m.p. uniforme (por tanto independiente de la estimación inicial del canal, $\hat{\mathbf{H}}_0$) en las ecuaciones (7.7) y (7.8). Esto resulta en una degradación del estimador de mínimos cuadrados puro y en un *symbol error rate* muy alto (alrededor de 0.4) tras el proceso de decodificación. Un SER como éste no puede conducir nunca al proceso de ecualización Turbo a la convergencia. Sin embargo, en nuestro caso la convergencia se alcanza porque, en las siguientes iteraciones, se obtiene una mejora sustancial en la estimación del canal proveniente del algoritmo EM (no de la estructura Turbo en sí misma). Nótese que

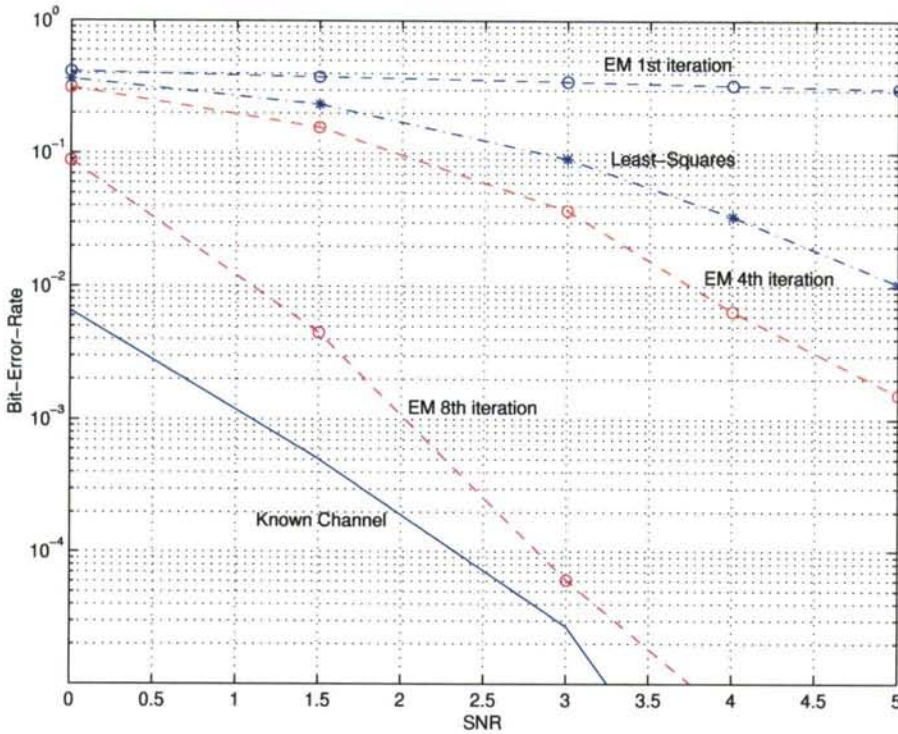


Figura 7.1: Rendimiento del sistema STC considerado sobre un canal no dispersivo

cada iteración del algoritmo EM (Ecuación (7.4)) se realiza sólo después de una iteración completa de ecualización-decodificación. En cualquier caso, una vez que la estimación del canal es suficientemente buena para que la estructura de ecualización Turbo yacza en su región de convergencia, tanto el algoritmo EM como el proceso iterativo Turbo ayudan en la reducción de la tasa de error. La Figura 7.1 también muestra que en la octava iteración el rendimiento es muy próximo al óptimo, es decir, al caso de canal conocido. Sólo 0.5 dB separan ambas curvas para un BER de 10^{-4} .

La Figura 7.2 muestra los resultados (BER tras la decodificación) obtenidos al considerar un canal MIMO dispersivo en el tiempo con memoria $m_c = 2$. Los parámetros de simulación son los mismos que en la figura anterior. En particular, nótese que, de nuevo, cada coeficiente del canal tiene varianza $\sigma_h^2 = 1/2$ por cada dimensión compleja. Es claro que en la cuarta iteración el estimador MV-EM rinde de modo muy similar al método de mínimos cuadrados, el cual no mejora significativamente a lo largo de las iteraciones. En la octava iteración el rendimiento del estimador MV-EM está de nuevo muy próxima al caso de canal conocido.

7.3.2. Estimación MV-EM para canales MIMO GSM en túneles subterráneos

El rendimiento del sistema de transmisión de tipo GSM y recepción Turbo espacio-temporal propuesto para entornos de túneles subterráneos ha sido también evaluado mediante

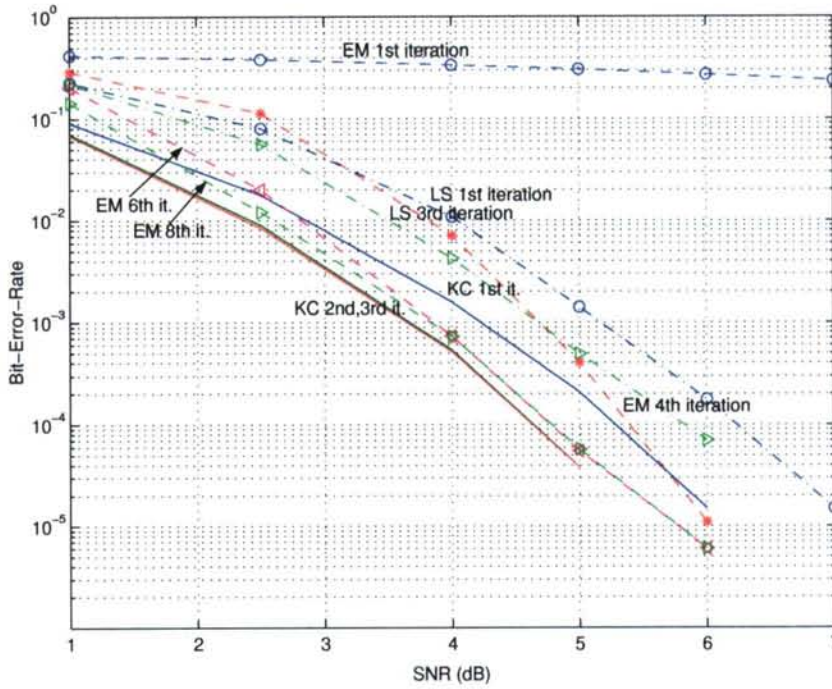


Figura 7.2: Rendimiento del sistema STC considerado sobre un canal dispersivo con memoria $m_c = 2$

simulaciones por computador. Las matrices de canal utilizadas, H_c , son el resultado de las medidas experimentales de la respuesta al impulso del canal MIMO presente en un túnel subterráneo realizadas en el contexto del proyecto europeo ESCORT.

La habilidad de nuestra técnica de estimación de canal para combinar la información determinística de los símbolos piloto y de la información estadística de los símbolos desconocidos proporcionada por el detector Turbo espacio-temporal, nos permite reducir el tamaño de la secuencia de entrenamiento de los sistemas GSM. De hecho, hemos determinado mediante simulaciones por computador la longitud mínima que debe tener la secuencia de entrenamiento para el sistema MIMO de tipo GSM considerado.

La Figura 7.3 muestra el error cuadrático medio (*Mean Square Error*, MSE) de la estimación de canal para distintas longitudes de la secuencia de entrenamiento ($N_t = 4, 5$ y 6 bits). El código de canal utilizado es el mismo que en las simulaciones anteriores. El tamaño de cada entrelazador es de 20800 bits y la longitud de bloque es la del estándar GSM (148 bits). Existe una diferencia significativa en el error de estimación entre usar $N_t = 4$ bits y $N_t = 5$ bits, al tiempo que la diferencia entre $N_t = 5$ y $N_t = 6$ es muy pequeña. Esto apunta a que $N_t = 5$ bits es la longitud mínima que debe tener la secuencia de entrenamiento. Esta suposición puede ser también corroborada a la vista de la Figura 7.4, donde se representa el Symbol Error Rate (SER) a la salida del decodificador frente a la SNR requerida.

A continuación comparamos los resultados obtenidos por el estimador propuesto utilizando una secuencia de entrenamiento de $N_t = 5$ bits y aquéllos obtenidos por el estimador clásico

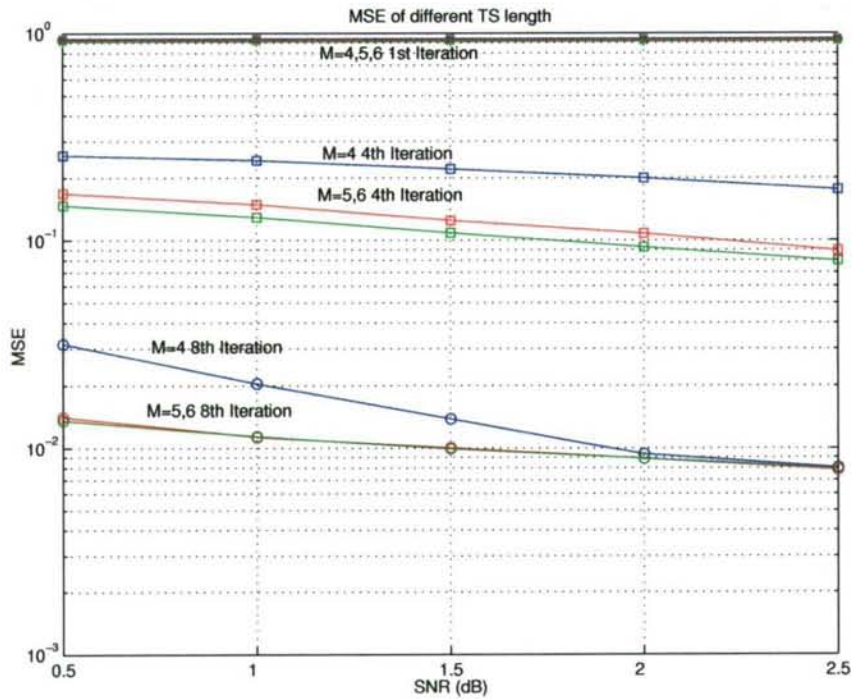


Figura 7.3: Error Cuadrático Medio (*Mean Square Error*,MSE) para distintas longitudes de la secuencia de entrenamiento

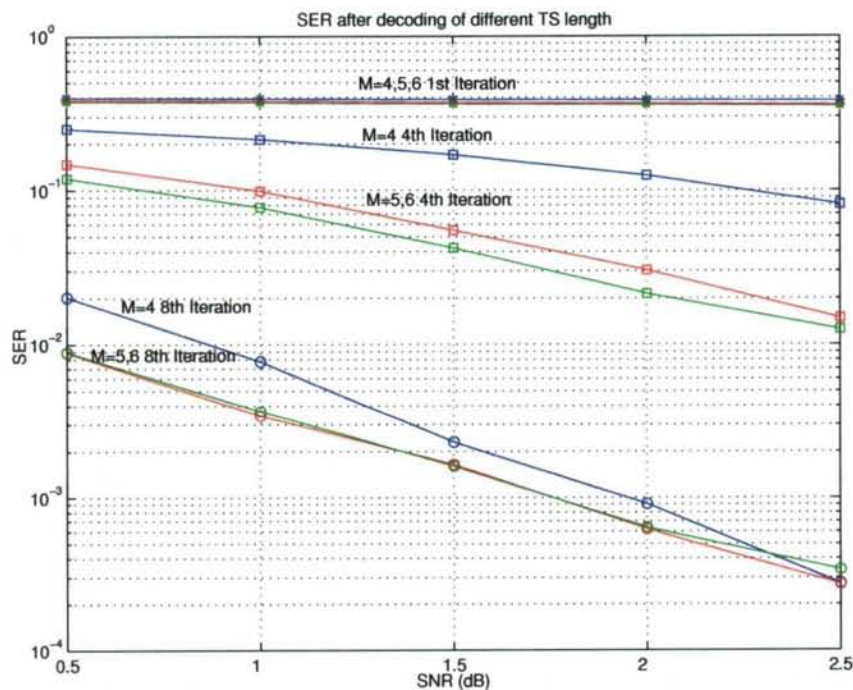


Figura 7.4: *Symbol Error Rate* (SER) frente a SNR a la salida del decodificador para distintas longitudes de la secuencia de entrenamiento

de mínimos cuadrados utilizando una secuencia de entrenamiento de $N_t = 26$ bits (la longitud estándar en GSM). Los resultados obtenidos cuando el receptor conoce perfectamente el canal se muestran también por propósitos de comparación. Tal y como muestra la Figura 7.5 el método MV-EM propuesto con $N_t = 5$ bits rinde mejor que el estimador de mínimos cuadrados con $N_t = 26$ bits a partir de la sexta iteración, alcanzando un rendimiento muy cercano al caso de canal conocido a partir de la séptima iteración.

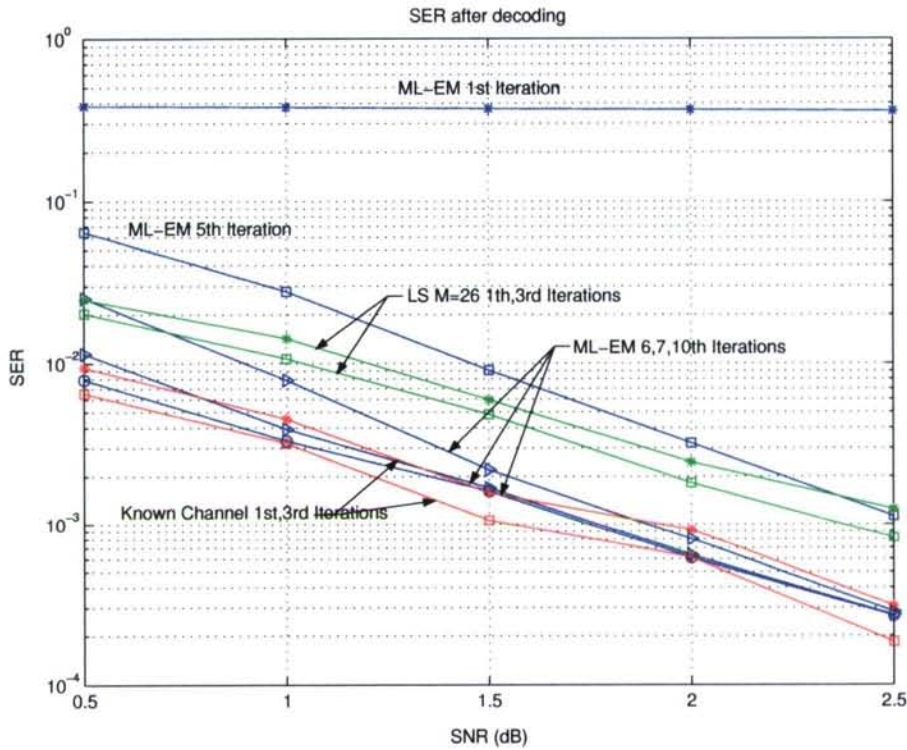


Figura 7.5: Comparación del rendimiento entre MV-EM ($N_t=5$ bits), LS ($N_t=26$ bits) y canal conocido (KC)

7.3.3. Ecuación DF MV-SAGE

Ilustraremos con simulaciones de Monte Carlo el rendimiento del ecualizador espacio-temporal propuesto. La Figura 7.6 muestra una comparación entre los resultados obtenidos por el detector MAP óptimo y los obtenidos por el ecualizador DF MV-SAGE para un canal dispersivo de memoria $m_c = 2$. La modulación escogida es BPSK y el número de antenas transmisoras y receptoras es $n_r = n_t = 2$. El código de canal utilizado es un código convolucional binario espacio-temporal de $rate R_c = 1/2$ con matriz generadora $G = [46, 72]$ (en representación octal) [92]. La longitud del entrelazador es 20800 bits. Los datos son procesados en bloques de longitud 233 bits de los cuales $N_t = 25$ corresponden a una secuencia piloto determinística que es conocida *a priori* por el receptor.

Los coeficientes del canal siguen el modelo Rayleigh espacialmente incorrelado. El canal cambia en cada bloque transmitido, permaneciendo constante a lo largo de la transmisión de un bloque. Hemos considerado que por cada iteración de decodificación se lleva a cabo una sola iteración del algoritmo SAGE para obtener una versión más refinada de los parámetros del ecualizador. La Figura 7.6 muestra cómo el Bit-Error-Rate disminuye a lo largo de las iteraciones de decodificación cuando se emplea el ecualizador DF MV-SAGE. Nótese que en este caso la detección MAP óptima no produce una mejora en el BER a partir de la tercera iteración. En la iteración duodécima del ecualizador propuesto, la diferencia con respecto al detector MAP óptimo está en torno a 2.1 dB para un BER de 10^{-3} y 2.7 dB para un BER de 10^{-4} .

La Figura 7.7 muestra el rendimiento del ecualizador DF MV-SAGE en un escenario de ISI más severa donde la memoria introducida por el canal es $m_c = 5$. Los datos son procesados en bloques de 248 bits de los cuales la secuencia piloto ocupa $N_t = 40$ bits. El resto de los parámetros de simulación son los mismos que se usaron en la Figura 7.6.

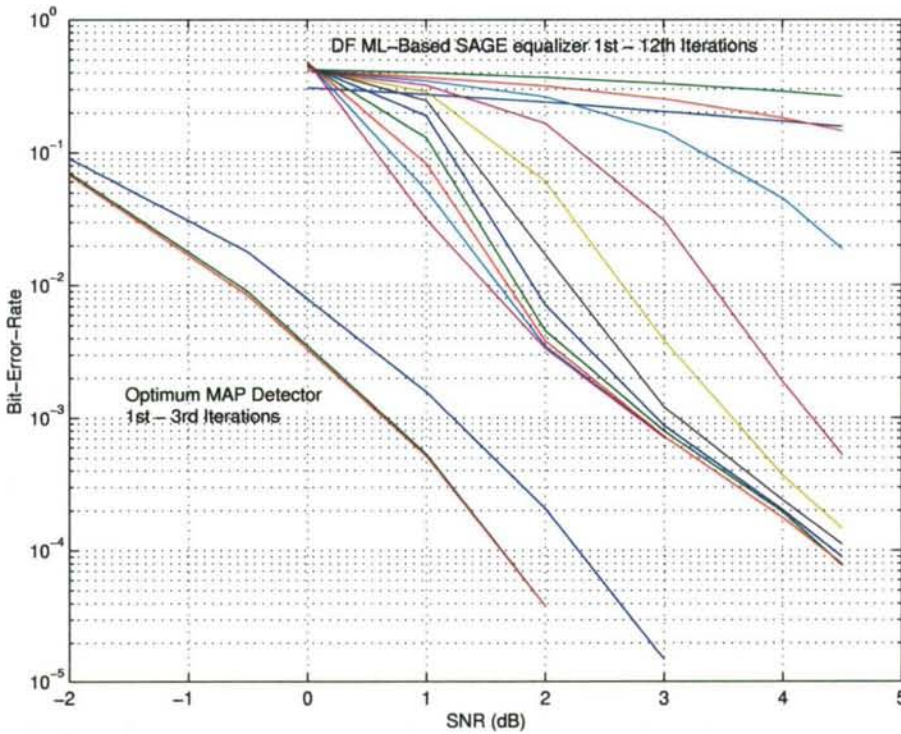


Figura 7.6: Comparación entre el detector MAP óptimo y el ecualizador DF MV-SAGE

7.4. Conclusiones

En sistemas prácticos el receptor no conoce el canal, por lo que es necesario estimarlo o bien compensarlo escogiendo adecuadamente los parámetros que configuran el ecualizador. Hemos

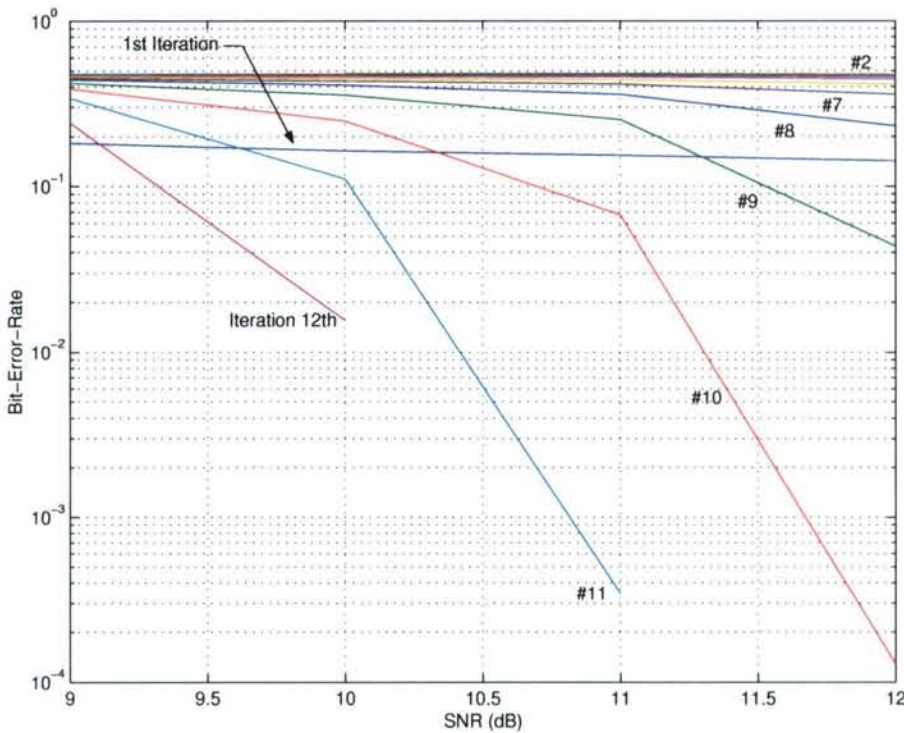


Figura 7.7: Rendimiento del ecualizador DF MV-SAGE en un escenario de ISI severa

contemplado estas dos perspectivas en este capítulo.

En primer lugar hemos propuesto realizar la estimación del canal MIMO de acuerdo al criterio Máximo-Verosímil (MV). Subyacente a esta idea está el hecho de que consideramos el canal como un parámetro fijo, pero desconocido en el receptor, en lugar de como una variable aleatoria tal y como habíamos contemplado en el modelo de señal introducido en el Capítulo 3. Esto es debido a que pretendemos que el receptor sea robusto y el asumir una determinada distribución para el canal puede conducir a que la estimación falle cuando el canal real no se ajusta bien a esa distribución. Esto es especialmente importante ya que, como comentamos anteriormente, los modelos probabilísticos de canal (Rayleigh, Rice, ...) son demasiado simplistas y no siempre se ajustan bien al comportamiento real del canal.

La aplicación directa del principio MV para estimar el canal MIMO conduce a un problema de optimización cuya solución no tiene forma cerrada. Esto nos ha llevado a aplicar el algoritmo EM (*Expectation-Maximization*) para obtener de forma iterativa dicha solución. Hemos visto cómo cada paso del algoritmo EM requiere la resolución de un problema de optimización cuadrático, cuya solución sí se puede obtener en forma cerrada. Esta última expresión requiere el cálculo de esperanzas matemáticas con respecto a la función de masa de probabilidad (f.m.p) *a posteriori* de los símbolos transmitidos. El cálculo de esta f.m.p. se puede efectuar mediante la aplicación del algoritmo *Forward-Backward*, por lo que lo natural es beneficiarse de la utilización de un ecualizador MAP Turbo en el que esta misma f.m.p. ha de obtenerse para detectar los símbolos transmitidos.

La potencia del método de estimación propuesto ha sido ilustrada mediante simulaciones por computador utilizando el formato de modulación GMSK y medidas experimentales del canal en túneles subterráneos. En este caso concreto, la utilización de esta técnica de estimación nos permite obtener resultados próximos al caso de suponer el canal conocido considerando una secuencia de entrenamiento de tan sólo 5 bits. Se ha mostrado también que este rendimiento es superior al correspondiente a técnicas de estimación usuales (eg. *Least-Squares*) que utilizan secuencias de entrenamiento mucho más largas (de 26 bits tal y como contempla el estándar GSM). En resumen, nuestro estimador permite un incremento del 14% de la tasa de transferencia del sistema.

El inconveniente del método aquí presentado es su alta complejidad computacional, aunque éste es un problema más bien inherente a los sistemas MIMO, como ya hemos comentado, y se está en la vía de desarrollar métodos efectivos para reducir esta complejidad.

Alternativamente, hemos abordado el problema de que el receptor desconozca el canal MIMO mediante un ecualizador *Decision Feedback* (DFE). Los coeficientes del DFE usualmente son calculados de acuerdo al criterio de minimización del error cuadrático medio (*Minimum Mean Square Error*, MMSE). Aquí se calculan los coeficientes de los filtros involucrados utilizando el criterio MV y aplicando el algoritmo EM para el cálculo de la solución. De modo análogo a como ocurría con el estimador de canal MV, las expresiones de los coeficientes dependen de la f.m.p. *a posteriori* de los bits transmitidos, por lo que podemos beneficiarnos de un esquema de igualación-decodificación Turbo iterativo en el que esas probabilidades *a posteriori* son calculadas por el decodificador MAP del código de canal utilizado.

El coste computacional del DFE aquí propuesto es mucho menor que el de un ecualizador MAP, mientras que los resultados de simulación muestran que su rendimiento, para un número suficiente de iteraciones, es tan sólo un poco peor (2 dB de diferencia en términos de SNR requerida para un BER de 10^{-4}).

Capítulo 8

Códigos LDGM en BICM

Los imposibles plausibles son preferibles a las posibilidades no convincentes

—Aristóteles

Desde 1982, *Trellis Coded Modulation* (TCM) ha sido considerado como el mejor procedimiento para diseñar esquemas de codificación y modulación espectralmente eficientes, ya que trata de modo conjunto la optimización de los bloques de codificación y modulación. Esto redundaba en un rendimiento considerablemente superior al del esquema tradicional consistente en utilizar un código de canal binario seguido de un *mapping* de bits a símbolos. El rendimiento de TCM es bueno tanto en canales AWGN como en canales con desvanecimiento. En este último caso, para incorporar diversidad temporal, se utiliza un entrelazador a nivel de símbolo en transmisión, de modo que símbolos consecutivos en la secuencia codificada original experimenten desvanecimientos independientes.

En 1992 Zehavi [58] sugirió que se podía obtener diversidad temporal sobre canales con desvanecimiento utilizando un codificador de canal seguido de un entrelazador a nivel de bit y de un *mapper* bit a símbolo. Denominó a *Bit-Interleaved Coded Modulation* (BICM) a dicho esquema de codificación/modulación. Contrariamente al caso de TCM, un esquema BICM no puede ser diseñado mediante la optimización conjunta del sistema de codificación/modulación como un todo, ya que las palabras código resultantes del proceso de entrelazamiento a nivel de bit son extremadamente largas. En cualquier caso, la utilización de códigos convolucionales estándar en un sistema BICM presentaba un rendimiento razonable sobre canales con desvanecimiento. Sin embargo, para el caso de canal AWGN (donde no se puede obtener ninguna ganancia del proceso de entrelazamiento) el rendimiento del esquema de Zehavi era peor que el de TCM.

Desde 1993, los códigos Turbo [10] se han mostrado como un esquema de codificación capaz de alcanzar la capacidad teórica del canal AWGN. Se pueden encontrar dos maneras diferentes de aplicar el así llamado “principio Turbo” a la transmisión espectralmente eficiente. Turbo TCM, introducida por vez primera en [106] y refinada posteriormente en [90], es la

generalización natural de los códigos Turbo binarios estándar a métodos de señalización de mayor dimensionalidad. La idea es construir el esquema turbo utilizando como componentes códigos TCM, en lugar de códigos convolucionales. La otra aproximación es BICM con decodificación iterativa. Debido a que el *mapping* de bits a símbolos es susceptible de ser visto como un código interno de razón unidad con entradas binarias y salidas complejas, se puede aplicar a un esquema BICM un proceso de decodificación iterativo consistente en calcular de modo separado las probabilidades *a posteriori* de los símbolos codificados tanto en el *demapper* como en el decodificador.

Para canales de desvanecimiento Rayleigh, la decodificación iterativa de esquemas BICM (BICM-ID) basados en códigos convolucionales [64] resulta en un rendimiento cercano al de los sistemas Turbo-TCM, aunque la diferencia se hace mayor cuando se consideran canales AWGN. Esta degradación del rendimiento es compensada por una menor complejidad de decodificación y una estructura más sencilla en transmisión: la codificación se realiza por medio de un código estándar, optimizado exclusivamente para el caso binario. En [107], se utilizaron códigos *Low Density Parity Check* (LDPC) como componentes de BICM se analizó el rendimiento de tanto de códigos LDPC regulares como irregulares¹. Los esquemas basados en códigos LDPC irregulares presentaban un rendimiento muy superior al de BICM-ID basada en códigos convolucionales, presentando el mismo rendimiento que Turbo-TCM, pero con la ventaja de no presentar ningún *error floor* y tener una menor complejidad tanto de codificación como de decodificación.

Desde el informe de Telatar de 1995 [8] sobre la capacidad de los canales MIMO resultantes del uso de múltiples antenas tanto en transmisión como en recepción, diferentes técnicas de señalización han sido propuestas para adecuarse a las características del canal MIMO. La mayoría de ellas se centran en permitir una detección muy sencilla de los símbolos en recepción [65]. Sin embargo, es necesario incorporar un código de canal externo para obtener un rendimiento lo más próximo posible a la capacidad del canal. Una manera es diseñar conjuntamente los esquemas de codificación y señalización, de modo análogo a TCM en sistemas de una sola antena. Los códigos *trellis* espacio-temporales [65] son una instancia de esta enfoque del problema. El inconveniente de estos esquemas es que son diseñados específicamente para cada caso particular (en cuanto al *trellis* y método de señalización empleados) y su complejidad de decodificación es exponencial en el número de antenas transmisoras.

Los discusión previa indica que la aplicación del concepto de BICM con decodificación iterativa a sistemas MIMO constituye una idea muy atractiva. Esto fue propuesto por vez primera por Tonello en [75], bajo el término de BICM espacio-temporal (Space-Time BICM, STBICM), utilizando códigos convolucionales. El mayor inconveniente de BICM-ID en sistemas MIMO es que la complejidad del proceso de *demapping* es exponencial en el número de antenas transmisoras. Esto puede ser efectivamente superado mediante el uso de potentes técnicas subóptimas de *demapping*, como la presentada en [15], en la que los autores muestran

¹Estos últimos eran construidos mediante el proceso de optimización denominado *Differential Evolution*

la capacidad de un sistema MIMO BICM-ID, utilizando códigos Turbo como códigos de canal, para rendir muy próximamente al límite de la capacidad del canal para un amplio rango de técnicas de señalización y número de antenas transmisoras y receptoras.

Consideraremos aquí esquemas BICM que utilizan una concatenación de dos códigos bloque lineales con matriz generada *sparse* denominados códigos *Low-Density Generator Matrix* (LDGM) [53]. La idea de utilizar códigos LDGM para BICM ya fue explorada por Cheng y McEliece en [50], pero sólo para códigos con alta razón de codificación, que no presentan *error floor* y por tanto no necesitan de concatenación. Abordaremos aquí el uso de códigos LDGM desde una perspectiva general, permitiendo códigos de tasa más baja eliminando su *error floor* asociado mediante un esquema concatenado. La motivación para usar códigos LDGM es diversa: primero, queremos investigar si, como se mostró en [107] para códigos LDPC, el rendimiento de los esquemas BICM puede ser mejorado al utilizar un código de canal más potente que uno convolucional. Segundo, la utilización de códigos LDGM en lugar de códigos convolucionales redundante en una menor complejidad de decodificación, ya que el algoritmo *Forward/Backward* requerido en cada iteración de decodificación es sustituido por operaciones más simples. Es importante señalar que, al contrario de lo que ocurre con los códigos LDPC en general, la complejidad de codificación de los códigos LDGM es bastante baja.

8.1. Códigos LDGM para BICM

Consideraremos el caso de códigos LDGM sistemáticos, que son códigos lineales con una matriz generadora *sparse*, $\mathbf{G}_{K \times N} = [\mathbf{I}_K \mathbf{P}_{K \times L}]$, donde $N = K + L$ y $\mathbf{P}_{K \times L}$ es una matriz *sparse*. Dada una palabra fuente $\mathbf{u} = (u_1, u_2, \dots, u_K)$, los bits de paridad $\mathbf{c} = (c_1, c_2, \dots, c_L)$ se obtienen como $\mathbf{c} = \mathbf{uP}$. Nótese que los códigos LDGM son un caso particular de los códigos LDPC estándar, ya que es obvio que sus matrices de control de paridad son también *sparse*. Como se indicó anteriormente, la ventaja de los códigos LDGM sobre los códigos LDPC es reside en que su complejidad en codificación es muy baja.

La decodificación de los códigos LDGM se lleva a cabo representándolos gráficamente mediante una red bayesiana y aplicando el algoritmo de propagación del conocimiento de Pearl [44] (*Belief Propagation*, BP) sobre el grafo resultante. Específicamente, cada nodo del grafo representa una variable aleatoria que describe o bien un bit sistemático o bien un bit de paridad. Las conexiones entre los nodos están determinadas por la matriz generadora \mathbf{G} . Dado que una conexión sólo es posible entre un nodo bit sistemático y un nodo bit de paridad, el grafo es bipartito. Utilizaremos la notación (L, X, K) para indicar que el número de bits de paridad y de entrada es, respectivamente, L y K , y que el grado de los nodos bit sistemáticos (el número de nodos bits de paridad que dependen de un nodo bit sistemático) es X .

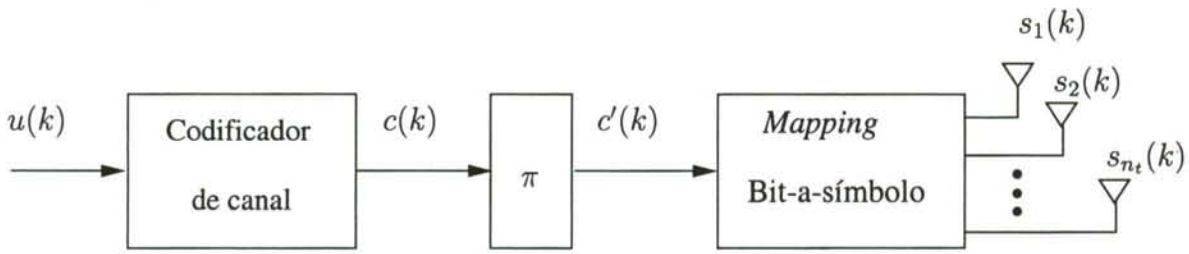


Figura 8.1: Diagrama de bloques del transmisor considerado

8.1.1. Eliminación del *error floor*: códigos LDGM concatenados

Excepto en el caso de códigos de tasa muy alta [50], los códigos LDGM presentan *error floors* al ser usados como códigos de canal estándar. Sin embargo, estos *error floors* pueden ser sustancialmente reducidos (y prácticamente eliminados) si se utiliza un esquema concatenado muy simple [53]. La idea básica es utilizar un código externo adicional, concatenado en serie, que permita corregir los pocos errores que permanezcan tras la decodificación del código interno. Nótese que cualquier clase de código podría utilizarse como código externo, pero aquí supondremos que se trata otro código LDGM sistemático. La tasa del código externo puede ser extremadamente alto, ya que el número de errores a corregir es muy pequeño. La salida del código externo $(u_1, \dots, u_K, c_1^o, \dots, c_{L_o}^o)$ es introducida en el código interno, que produce la secuencia de bits que servirán como entrada al *mapper* siguiente

$$(c_1, c_2, \dots, c_{N_i}) = (u_1, \dots, u_K, c_1^o, \dots, c_{L_o}^o, c_1^i, \dots, c_{L_i}^i)$$

Denotaremos por $[(L_o, X_o, K_o), (L_i, X_i, K_i)]$ el código resultante de la concatenación de un código LDGM (L_o, X_o, K_o) externo y un código LDGM (L_i, X_i, K_i) interno.

Para conservar fija la tasa resultante, tenemos que ajustar las tasas de los códigos interno y externo. Consideraremos que no se realiza ningún tipo de *puncturing* de los bits sistemáticos, así que estaremos transmitiendo K_o bits sistemáticos y $L_o + L_i$ bits de paridad. La tasa global viene dada por $R_c = K_o / (K_o + L_o + L_i)$. Podemos construir el código interno modificando ligeramente el número de bits sistemáticos y de paridad de un código LDGM simple (no concatenado) que presente buenas propiedades de *threshold* y *error floor*. En otras palabras, si sabemos que un código LDGM (L, X, K) simple presenta un buen rendimiento, podremos utilizar como código interno en un esquema concatenado un código LDGM $(L - L_o, X, K + L_o)$ simple, donde L_o es pequeño, y un código LDGM $(L_o, X_o, K_o = K)$ como código externo.

8.1.2. Mapping bit a símbolo

La Figura 8.1 muestra el diagrama de bloques del transmisor considerado en este capítulo, que difiere ligeramente del modelo de transmisor considerado hasta ahora.

Dado un bloque de $N_i = K_o + L_o + L_i$ bits a la salida de la etapa de codificación, consideraremos que la asignación de bits a símbolos se realiza de modo pseudoaleatorio,

mediante un entrelazador. Denotaremos por $(c'_1, c'_2, \dots, c'_{N_i}) = \pi(c_1, c_2, \dots, c_{N_i})$ la secuencia de bits codificada tras el proceso de entrelazamiento. El *mapping* al símbolo transmitido $s(k) = [s_1(k), s_2(k), \dots, s_{n_t}(k)]^T$ se realiza entonces de acuerdo a una regla de asignación dada

$$s(k) = \text{map}(c'(k)) \quad k = 1, 2, \dots, N_i/(n_t M_c)$$

donde

$$c'(k) = (c'_{n_t M_c(k-1)+1}, c'_{n_t M_c(k-1)+2}, \dots, c'_{n_t M_c(k-1)+n_t M_c})$$

y n_t es el número de antenas transmisoras y M_c denota el número de bits transportados por cada símbolo en la modulación empleada. Por ejemplo, $M_c = 2, 3, 4$ bits para QPSK, 8PSK y 16QAM, respectivamente. En particular, para cada subbloque de $n_t M_c$ bits se puede realizar el *mapping* asignando los primeros M_c bits a la primera antena transmisora, los siguientes M_c bits a la segunda, y así sucesivamente. Para cada antena transmisora se puede utilizar un *mapping* habitual, como Gray, Set Partitioning o Semi-Set Partitioning.

Es interesante señalar que, al contrario de lo que ocurre en los esquemas BICM basados en códigos convolucionales, cuando se utilizan códigos LDGM el *mapping* Gray es el que produce los mejores resultados. Una explicación intuitiva es que los códigos LDGM presentan una transición muy abrupta entre la región de convergencia (con algo de *error floor*) y la región de no-convergencia (con un número elevado de errores), mientras que un código convolucional tiene una curva de probabilidad de error mucho más suave. Por tanto, es razonable esperar que en los códigos LDGM el factor más importante con respecto a la convergencia sea recibir “un buen empuje en la buena dirección” proveniente del *demapping* en la primera iteración de decodificación. En otras palabras, un “empuje” que conduzca al código LDGM a su región de convergencia (ya que si el “empuje” no es suficientemente bueno, el código LDGM permanecerá estancado en la región de no-convergencia y, contrariamente a un código convolucional, la información será pasada al *demapper* en la siguiente iteración no mejorará en absoluto). Obviamente, el *mapping* Gray es el más adecuado para proporcionar este “empuje”. Para corroborar esta interpretación, decodificamos el esquema BICM propuesto sin realizar iteración alguna entre el *demapper* y el decoder (es decir, el *demapper* se utiliza sólo una vez y a partir de entonces se decodifica el código LDGM utilizando siempre la misma información proveniente del *demapper*). Los resultados fueron prácticamente los mismos (sólo se degradaba el rendimiento 0.05 dB) a los obtenidos iterando entre el *demapper* y el decoder. Nótese que en [107] se obtuvo una conclusión idéntica con respecto a BICM utilizando códigos LDPC.

Las señales recibidas pueden escribirse como

$$\mathbf{x}(k) = \mathbf{H}(k)\mathbf{s}(k) + \mathbf{g}(k) \quad k = 1, 2, \dots, N_i/(n_t M_c) \quad (8.1)$$

donde $\mathbf{H}(k)$ es la matriz, de tamaño $n_r \times n_t$, que modela el canal. Supondremos que cada coeficiente complejo de la matriz de canal es circularmente invariante y de distribución gaussiana (estando por tanto su módulo distribuido según una Rayleigh) con varianza unidad. Los coeficientes del canal son independientes tanto en la dimensión espacial como en la temporal. Las componentes del vector de ruido $\mathbf{g}(k)$ son tanto espacial como temporalmente

incorrelados, circularmente invariantes, y de distribución gaussiana con varianza $N_0/2$ por dimensión compleja.

Como se muestra en la Figura 8.2, el esquema conjunto de modulación y codificación puede ser convenientemente representado mediante una red bayesiana. Cada nodo de la red representa una variable aleatoria y la conexión entre dos nodos significa que existe una dependencia estadística entre ambos. En nuestro caso, las dependencias entre nodos son determinísticas, ya que reflejan o bien restricciones de paridad o el *mapping* entre bits y símbolos. Ambos casos son contemplados simplemente definiendo las correspondientes f.m.p.s como funciones indicadoras.

8.2. Detección y decodificación iterativas

En esta sección describiremos cómo aplicar el algoritmo de propagación del conocimiento de Pearl (*Belief Propagation*, BP) [44] sobre el grafo que representa el esquema de BICM con la concatenación de dos códigos LDGM. El objetivo del proceso de decodificación es calcular el denominado *belief* de cada nodo bit de información, denotado por $BEL(u_k)$. Si la red bayesiana careciese de ciclos², $BEL(u_k)$ sería igual a $P[u_k|e]$, la probabilidad *a posteriori* de que el nodo bit U_k sea igual a u_k dada la *evidencia* (es decir, las observaciones y las restricciones probabilísticas contenidas en la red bayesiana) de la que se dispone. Una vez que se ha calculado $BEL(u_k)$, se toma una decisión de acuerdo a la regla usual

$$\hat{u}_k = \begin{cases} 0 & \text{if } BEL(u_k = 0) > BEL(u_k = 1) \\ 1 & \text{if } BEL(u_k = 0) \leq BEL(u_k = 1). \end{cases} \quad (8.2)$$

En general, para un nodo dado U , el *belief* de $U = u$ se calcula como

$$BEL(u) \stackrel{\alpha}{=} \lambda(u)\pi(u), \quad (8.3)$$

donde $\stackrel{\alpha}{=}$ representa la igualdad salvo una constante multiplicativa y $\lambda(u) = P[e_U^-|u]$ y $\pi(u) = P[u|e_U^+]$ son, respectivamente, la verosimilitud de $U = u$ dada la evidencia contenida en el árbol bajo el nodo U , y la probabilidad *a posteriori* de $U = u$ dada la evidencia contenida en la red sobre el nodo U . Denotaremos por Y_i , $i = 1, \dots, m$, y W_j , $j = 1, \dots, n$, los hijos y los padres del nodo U , respectivamente. Dados los mensajes recibidos desde sus hijos, $\lambda_{Y_i}(u)$, $i = 1, \dots, m$, y desde sus padres, $\pi_U(w_j)$, $j = 1, \dots, n$, se pueden calcular $\lambda(u)$ y $\pi(u)$ como

$$\lambda(u) = \prod_{i=1}^m \lambda_{Y_i}(u) \quad (8.4)$$

$$\pi(u) = \sum_{\mathbf{w}} P[u|\mathbf{w}] \prod_{j=1}^n \pi_U(w_j), \quad (8.5)$$

²En nuestro caso, como ocurre en general con los esquemas de codificación Turbo y LDPC, la red efectivamente presenta ciclos, pero, como veremos en el apartado de simulaciones numéricas, el rendimiento resultante es razonable.

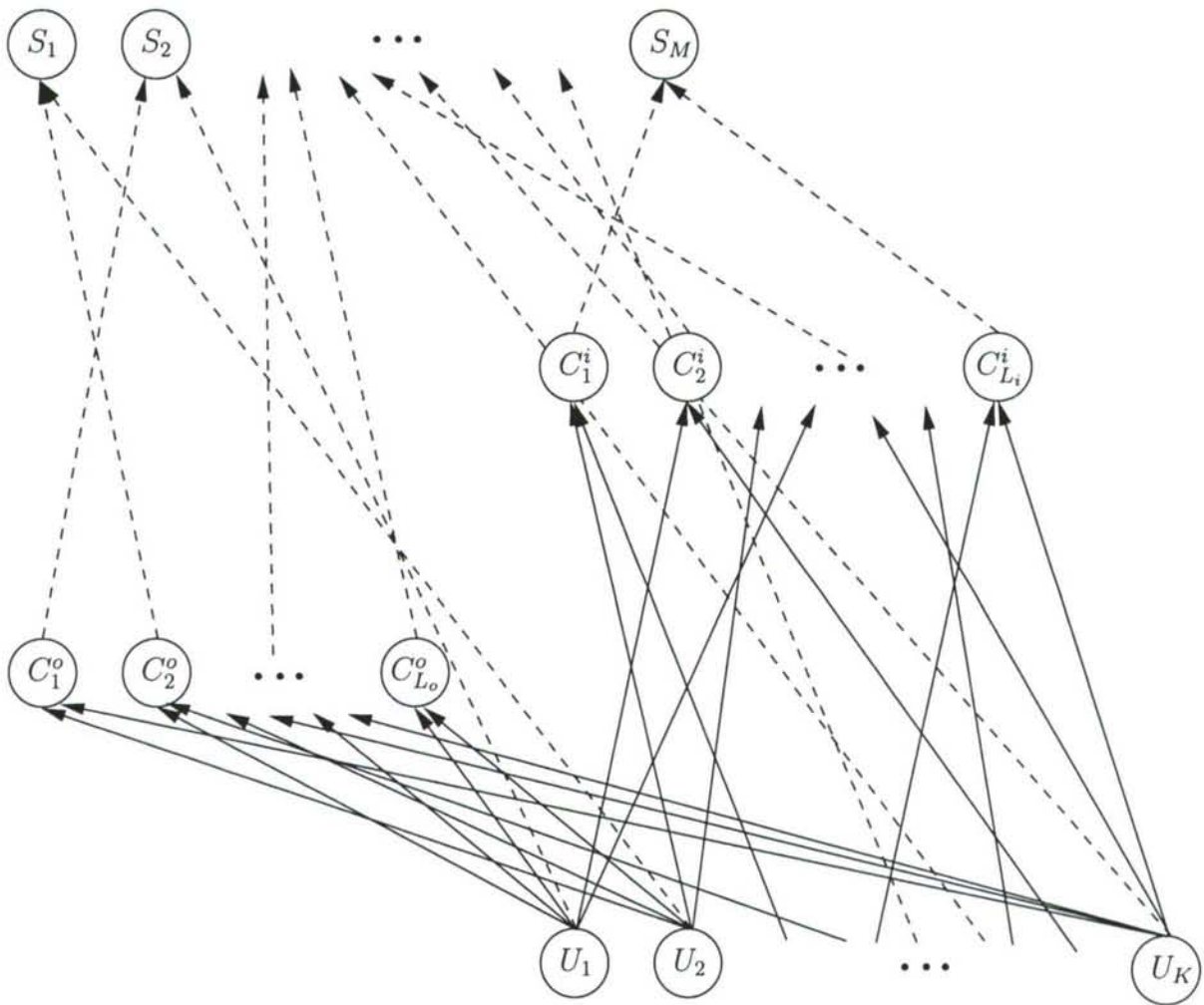


Figura 8.2: Red bayesiana que representa el esquema concatenado de codificación y modulación. $\{U_k\}$ representa los bits de información, $\{C_k^o\}$ los bits externos de paridad, $\{C_k^i\}$ los bits internos de paridad y $\{S_k\}$ los símbolos transmitidos a través del canal.

donde $\mathbf{w} = (w_1, \dots, w_n)$ denota los valores tomados por los padres de U .

El nodo U entrega mensajes tanto a sus hijos como a sus padres. El mensaje $\lambda_U(w_j) = P[e_U \setminus e_U^{W_j} | w_j]$ entregado por el nodo U a su padre $W_j = w_j$ es la verosimilitud de $W_j = w_j$ dada toda la evidencia observada por el nodo U excepto aquella proveniente de W_j , y se calcula como

$$\lambda_U(w_j) \stackrel{\alpha}{=} \sum_u \lambda(u) \sum_{\mathbf{w} \setminus w_j} P[u | \mathbf{w}] \prod_{l \neq j} \pi_U(w_l), \quad (8.6)$$

donde $\mathbf{w} \setminus w_j$ denota el vector \mathbf{w} sin su componente j -ésima. De modo similar, el mensaje entregado por U a su hijo Y_i , $\pi_{Y_i}(u) = P[Y_i | u, e_U \setminus e_U^{Y_i}]$, es la probabilidad *a posteriori* de Y_i dado $U = u$ y toda la evidencia observada por el nodo U salvo aquella proveniente de Y_i , y se calcula como

$$\pi_{Y_i}(u) \stackrel{\alpha}{=} \frac{\lambda(u)}{\lambda_{Y_i}(u)} \pi(u). \quad (8.7)$$

La particularización de las anteriores ecuaciones al esquema propuesto produce las siguientes expresiones:

8.2.1. Nodos bit

$$BEL(u_k) \stackrel{\alpha}{=} \lambda(u_k) \pi(u_k) \stackrel{\alpha}{=} \lambda(u_k), \quad (8.8)$$

done la última igualdad resulta de considerar los bits de información como equiprobables, es decir, $\pi(u_k) = P[u_k] = 0,5$ y

$$\lambda(u_k) = \lambda_{S_k}(u_k) \prod_{j=1}^{m_o} \lambda_{C_j^o}(u_k) \prod_{j=1}^{m_i} \lambda_{C_j^i}(u_k). \quad (8.9)$$

Los mensajes entregados por U_k a sus hijos (nótese que U_k no tiene ningún nodo padre) vienen dados por

$$\pi_{S_k}(u_k) \stackrel{\alpha}{=} \prod_{j=1}^{m_o} \lambda_{C_j^o}(u_k) \prod_{j=1}^{m_i} \lambda_{C_j^i}(u_k) \quad (8.10)$$

$$\pi_{C_j^o}(u_k) \stackrel{\alpha}{=} \lambda_{S_k}(u_k) \prod_{l \neq j} \lambda_{C_l^o}(u_k) \prod_{j=1}^{m_i} \lambda_{C_j^i}(u_k) \quad (8.11)$$

$$\pi_{C_j^i}(u_k) \stackrel{\alpha}{=} \lambda_{S_k}(u_k) \prod_{j=1}^{m_o} \lambda_{C_j^o}(u_k) \prod_{l \neq j} \lambda_{C_l^i}(u_k). \quad (8.12)$$

8.2.2. Nodos bit de paridad externos

$$BEL(c_k^o) \stackrel{\alpha}{=} \lambda(c_k^o) \pi(c_k^o), \quad (8.13)$$

donde

$$\lambda(c_k^o) = \lambda_{S_k}(c_k^o) \prod_{j=1}^{m_i} \lambda_{C_j^i}(c_k^o) \quad (8.14)$$

$$\pi(c_k^o) = \sum_{\substack{\mathbf{u}: \\ \sum_i u_i = c_k^o}} \prod_{j=1}^{N_i} \pi_{C_k^o}(u_j). \quad (8.15)$$

Los mensajes entregados por un nodo bit de paridad externo a sus padres están dados por

$$\lambda_{C_k^o}(u_j) \stackrel{\alpha}{=} \sum_{c_k^o} \lambda(c_k^o) \sum_{\substack{\mathbf{u} \setminus u_j: \\ \sum_i u_i = c_k^o}} \prod_{l \neq j} \pi_{C_k^o}(u_l), \quad (8.16)$$

y aquéllos que entrega a sus hijos por

$$\pi_{S_j}(c_k^o) \stackrel{\alpha}{=} \prod_{l=1}^{m_i} \lambda_{C_l^i}(c_k^o) \pi(c_k^o) \quad (8.17)$$

$$\pi_{C_j^i}(c_k^o) \stackrel{\alpha}{=} \lambda_{S_k}(c_k^o) \prod_{l \neq j} \lambda_{C_l^i}(c_k^o) \pi(c_k^o). \quad (8.18)$$

8.2.3. Nodos bit de paridad internos

$$BEL(c_k^i) \stackrel{\alpha}{=} \lambda(c_k^i) \pi(c_k^i), \quad (8.19)$$

donde

$$\lambda(c_k^i) = \lambda_{S_k}(c_k^i) \quad (8.20)$$

$$\pi(c_k^o) = \sum_{\substack{\mathbf{w}: \\ \sum_i w_i = c_k^i}} \prod_{j=1}^{n_i} \pi_{C_k^i}(w_j), \quad (8.21)$$

donde w_j es el valor tomado por o bien un nodo bit o bien por un nodo bit de paridad externo. Los mensajes entregados por un nodo bit de paridad interno a sus padres son

$$\lambda_{C_k^i}(w_j) \stackrel{\alpha}{=} \sum_{c_k^i} \lambda(c_k^i) \sum_{\substack{\mathbf{w} \setminus w_j: \\ \sum_i w_i = c_k^i}} \prod_{l \neq j} \pi_{C_k^i}(w_l), \quad (8.22)$$

y aquéllos entregados a sus hijos son

$$\pi_{S_k}(c_k^i) \stackrel{\alpha}{=} \pi(c_k^i). \quad (8.23)$$

8.2.4. Nodos símbolo

$$BEL(s_k) \stackrel{\alpha}{=} \lambda(s_k)\pi(s_k), \quad (8.24)$$

donde $\lambda(s_k) = P[X_k|s_k] \propto \exp(-\|\mathbf{x}(k) - \mathbf{H}(k)\mathbf{s}(k)\|^2/N_0)$ es la verosimilitud de $S_k = s_k$ dada la observación ruidosa proveniente del canal MIMO, y

$$\pi(s_k) = \sum_{\mathbf{w}:\text{map}(\mathbf{w})=s_k} \prod_{j=1}^{n_s} \pi_{S_k}(w_j), \quad (8.25)$$

Los mensajes entregados por un nodo símbolo a sus padres son

$$\lambda_{S_k}(w_j) \stackrel{\alpha}{=} \sum_{s_k} \lambda(s_k) \sum_{\mathbf{w} \setminus w_j:\text{map}(\mathbf{w})=s_k} \prod_{l \neq j} \pi_{S_k}(w_l). \quad (8.26)$$

Se pueden utilizar diferentes planificaciones para la actualización de los mensajes. Sin embargo, hemos corroborado que la mayoría de ellos alcanzan un rendimiento similar. En adelante supondremos que la planificación contempla la actualización sucesiva de los nodos bit de paridad internos, externos, y sistemáticos.

8.2.5. Consideraciones de implementación

Es importante indicar algunos aspectos relativos a la implementación de las expresiones anteriores. En particular, el cálculo de $\pi(c_k^o)$, $\pi(c_k^i)$, $\lambda_{C_k^o}(u_j)$ y $\lambda_{C_k^i}(w_j)$ puede simplificarse de forma notable. Explicaremos cómo implementar el cálculo de $\pi(c_k^i)$ y $\lambda_{C_k^i}(w_j)$, siendo el cálculo de $\pi(c_k^o)$ y $\lambda_{C_k^o}(u_j)$ totalmente análogo. Primero expresaremos la Ecuación (8.22) como

$$\lambda_{C_k^i}(w_j) \stackrel{\alpha}{=} \sum_{c_k^i} \lambda(c_k^i) r_{c_k^i}^{w_j}, \quad (8.27)$$

donde

$$r_{c_k^i}^{w_j} = \sum_{\substack{\mathbf{w} \setminus w_j: \\ \sum_l w_l = c_k^i}} \prod_{l \neq j} \pi_{C_k^i}(w_l) \quad (8.28)$$

Recordemos que $\pi_{C_k^i}(w_l)$ se define como la probabilidad *a posteriori* de $W_l = w_l$ dada toda la evidencia contenida en sus padres e hijos salvo la correspondiente al nodo C_k . Denotando a esta probabilidad como $P[w_l|\tilde{w}_l]$, donde \tilde{w}_l representa una observación *ficticia* de W_l (la evidencia

contenida en W_l), tenemos que

$$r_{c_k}^{w_j} = \sum_{\substack{\mathbf{w} \setminus w_j: \\ \sum_l w_l = c_k^i}} \prod_{l \neq j} P[w_l | \tilde{w}_l] \quad (8.29)$$

$$= \sum_{\substack{\mathbf{w} \setminus w_j: \\ \sum_l w_l = c_k^i}} P[\mathbf{w} \setminus w_j | \tilde{\mathbf{w}} \setminus \tilde{w}_j] \quad (8.30)$$

$$= \sum_{\mathbf{w} \setminus w_j} P[c_k | \mathbf{w} \setminus w_j] P[\mathbf{w} \setminus w_j | \tilde{\mathbf{w}} \setminus \tilde{w}_j] \quad (8.31)$$

$$= P[c_k | \tilde{\mathbf{w}} \setminus \tilde{w}_j] \quad (8.32)$$

donde $\tilde{\mathbf{w}} = [\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_{n_i}]$. Teniendo en cuenta que

$$r_1^{w_j} = P[1 | \tilde{\mathbf{w}} \setminus \tilde{w}_j] = 1 - P[0 | \tilde{\mathbf{w}} \setminus \tilde{w}_j] = 1 - r_0^{w_j} \quad (8.33)$$

es suficiente calcular

$$r_0^{w_j} = P[0 | \tilde{\mathbf{w}} \setminus \tilde{w}_j] \quad (8.34)$$

Definiendo

$$S_j^- = \sum_{l=1}^{j-1} W_l \quad (8.35)$$

$$S_j^+ = \sum_{l=j+1}^{n_i} W_l \quad (8.36)$$

$$\tilde{\mathbf{w}}_j^- = [\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_{j-1}] \quad (8.37)$$

$$\tilde{\mathbf{w}}_j^+ = [\tilde{w}_{j+1}, \tilde{w}_{j+2}, \dots, \tilde{w}_{N_i}] \quad (8.38)$$

podemos escribir (8.34) como

$$r_0^{w_j} = P[S_j^- + w_j + S_j^+ = 0 | w_j, \tilde{\mathbf{w}}_j^-, \tilde{\mathbf{w}}_j^+] \quad (8.39)$$

que es conveniente expresar mediante

$$r_0^{w_j} = \sum_{s_j^-} P[s_j^- + w_j + S_j^+ = 0 | w_j, \tilde{\mathbf{w}}_j^-, \tilde{\mathbf{w}}_j^+, s_j^-] P[s_j^- | w_j, \tilde{\mathbf{w}}_j^-, \tilde{\mathbf{w}}_j^+] \quad (8.40)$$

$$= \sum_{s_j^-} P[S_j^+ = s_{j+1}^- | \tilde{\mathbf{w}}_j^+] P[s_j^- | \tilde{\mathbf{w}}_j^-] \quad (8.41)$$

donde en la última igualdad se ha tenido en cuenta que S_j^- y S_j^+ dependen sólo de $\tilde{\mathbf{w}}_j^-$ y $\tilde{\mathbf{w}}_j^+$, respectivamente. Introduciendo una notación similar a la del algoritmo *Forward/Backward*

$$\alpha(s_j^-) = P[s_j^- | \tilde{\mathbf{w}}_j^-] \quad (8.42)$$

$$\beta(s_j^+) = P[s_j^+ | \tilde{\mathbf{w}}_j^+] \quad (8.43)$$

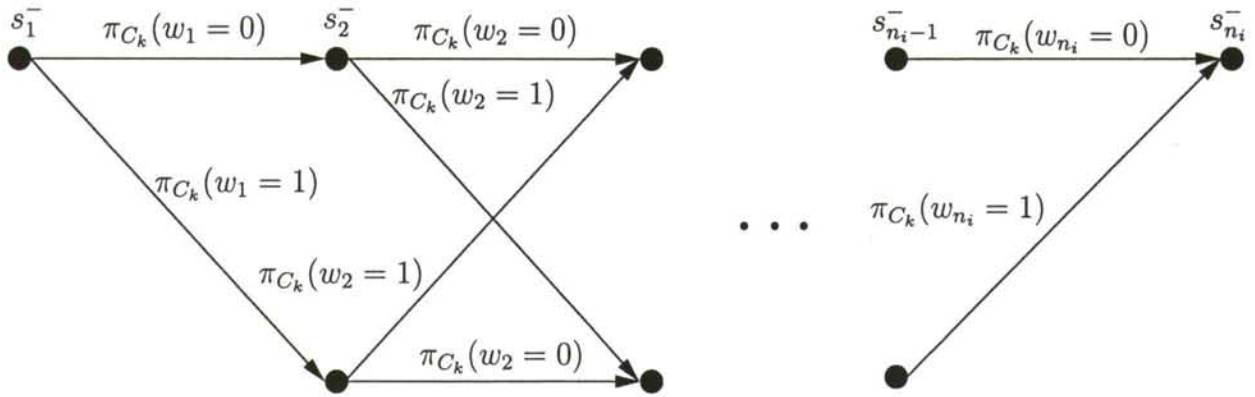


Figura 8.3: Cadena de Markov sobre la que se realiza el cálculo de $r_0^{w_j}$

se puede comprobar que $\alpha(s_j^-)$ y $\beta(s_j^+)$ satisfacen las siguientes recursiones

$$\alpha(s_j^-) = \sum_{w_{j-1}} \alpha(s_{j-1}^-) P[w_{j-1} | \tilde{w}_{j-1}] = \sum_{w_{j-1}} \alpha(s_{j-1}^-) \pi_{C_k^i}(w_{j-1}) \quad (8.44)$$

$$\beta(s_j^+) = \sum_{w_{j+1}} \beta(s_{j+1}^+) P[w_{j+1} | \tilde{w}_{j+1}] = \sum_{w_{j+1}} \beta(s_{j+1}^+) \pi_{C_k^i}(w_{j+1}) \quad (8.45)$$

con condiciones iniciales

$$\alpha(s_1^- = 0) = 1 \quad (8.46)$$

$$\alpha(s_1^- = 1) = 0 \quad (8.47)$$

$$\beta(s_{n_i}^+ = 0) = 1 \quad (8.48)$$

$$\beta(s_{n_i}^+ = 1) = 0 \quad (8.49)$$

8.3. Simulaciones numéricas

8.3.1. Canal AWGN

Consideraremos aquí el sistema BICM propuesto utilizando una modulación 8PSK sobre canal AWGN. La tasa de los códigos considerados es $R_c = 2/3$.

La curva de BER correspondiente a un código LDGM presenta tanto un *threshold* como un *error floor*, que dependen del grado de los nodos bit sistemáticos del código [53]. De todos los códigos LDGM regulares (aquéllos en los cuales los nodos bit sistemáticos tienen el mismo grado), nos interesan aquellos con mejor *threshold* pero con un *error floor* suficientemente bajo para ser corregido por un código externo de alta tasa. La Figura 8.4 muestra las curvas de BER correspondientes al sistema BICM propuesto utilizando códigos LDGM (10000, X, 20000), con $X = 3, 4, 5, 6$, regulares. Nótese que la capacidad (cuando la distribución de la entrada se

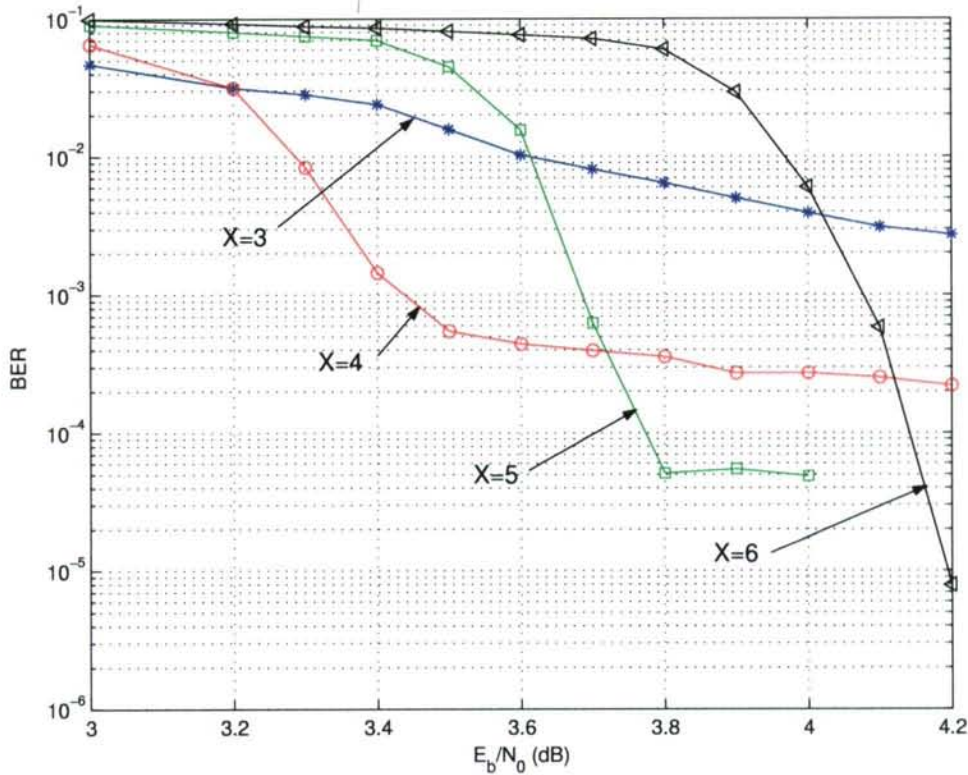


Figura 8.4: Rendimiento del esquema propuesto de BICM 8PSK basado en códigos LDGM regulares de tasa $R_c = 2/3$ y distintos grados de los nodos bit $X = 3, 4, 5, 6$. En todos los casos la longitud de bloque es $K = 20000$ bits de información.

restringe a la modulación utilizada) está situada en $E_b/N_0 = 2,8$ dB. El *error floor* para $X = 3$ es inaceptablemente alto ($10^{-2} - 10^{-3}$). Se observa un buen balance entre el *threshold* (en $E_b/N_0 = 3,2$ dB) y el *error floor* (10^{-4}) para $X = 4$. Los códigos de grado mayor ($X = 5, 6$) exhiben *error floors* más bajos pero *thresholds* más altos.

A continuación nos centraremos en el impacto de la longitud de bloque en el rendimiento. La Figura 8.5 muestra el rendimiento obtenido por el esquema BICM propuesto cuando los códigos LDGM tienen un grado de los nodos bit sistemáticos $X = 4$ y se consideran distintas longitudes de bloque, $K = 10000, 20000$ y 30000 bits de información. La Figura 8.5 muestra una clara diferencia entre los casos de $K = 10000$ bits y $K = 20000$ bits, mientras que incrementar la longitud de bloque hasta $K = 30000$ bits de información no proporciona una mejora significativa.

Antes de introducir un código externo para eliminar el *error floor*, examinaremos si la utilización de una clase sencilla de códigos LDGM irregulares (aquellos códigos con grado de nodo bit sistemáticos entre $X = 3$ y $X = 4$) puede proporcionar alguna ventaja en el esquema BICM propuesto respecto al *threshold* o al *error floor*. Para ello, en la Figura 8.6 se muestran las curvas de BER correspondientes a diferentes códigos LDGM irregulares con X variando desde 3,6 to 4,2. Un grado de $X = x_e.x_f$ significa que algunos nodos bit tienen grado igual a x_e

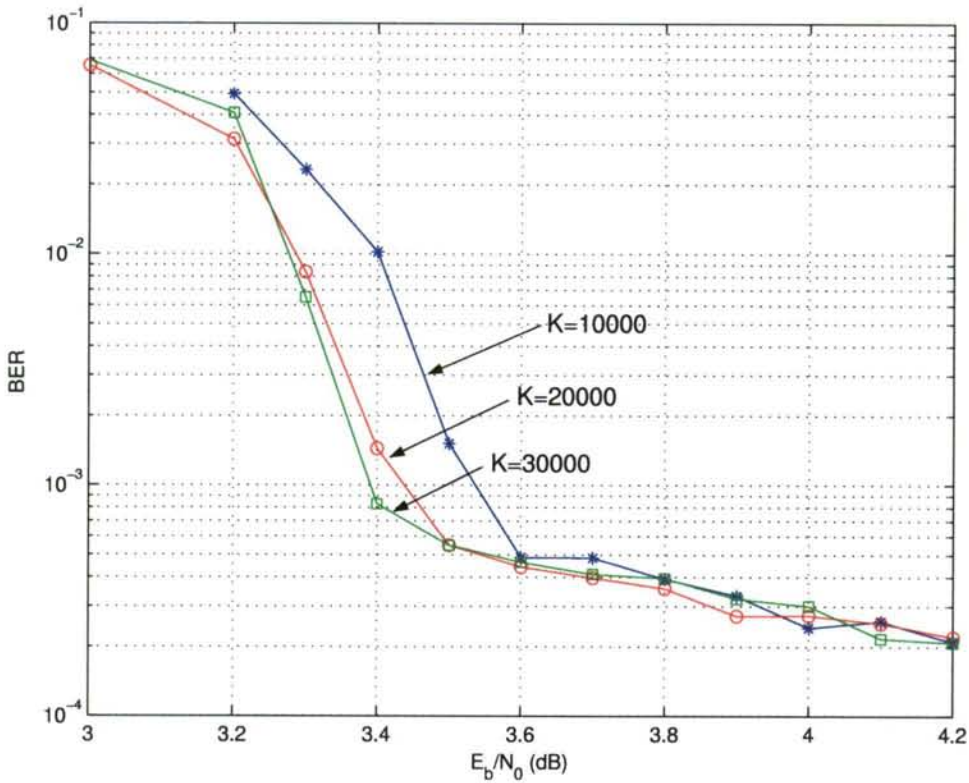


Figura 8.5: Rendimiento del esquema propuesto de BICM 8PSK basado en códigos LDGM regulares de tasa $R_c = 2/3$ y distintas longitudes de bloque, $K = 10000, 20000, 30000$ bits de información. En todos los casos el grado de los nodos bit sistemáticos es $X = 4$.

y el resto igual a $x_e + 1$. Nótese que la fracción de nodos bit sistemáticos con grado x_e verifica la restricción $KX = LY$, donde Y es el grado de los nodos bit de paridad. Debido a que no parece que se puedan obtener ganancias significativas derivadas del uso de estos códigos LDGM irregulares sencillos, nos centraremos de nuevo en códigos LDGM regulares con grado de nodo bit sistemático $X = 4$ para la construcción del esquema concatenado.

La Figura 8.7 muestra los resultados obtenidos por el esquema BICM propuesto utilizando diferentes configuraciones (correspondientes a distintos valores del grado de nodo bit sistemático del código externo) del esquema concatenado resultante de la primera y segunda aproximación descritos anteriormente. Como se aprecia en la figura, el uso de un grado de nodo bit sistemático muy bajo para el código externo es suficiente para eliminar el *error floor*. Aunque el esquema concatenado resultante de la segunda aproximación presenta una complejidad de decodificación ligeramente superior, no muestra una mejora en el rendimiento.

Finalmente, la Figura 8.8 compara el rendimiento del esquema propuesto con el de otras aproximaciones recientes al problema de la transmisión espectralmente eficiente sobre canales AWGN. Los resultados de BICM basada en códigos convolucionales están tomados de [64], los de Turbo-TCM son los mejores de los presentados en [90], y los de BICM basada en códigos LDPC son los correspondientes a [107]. Los esquemas BICM basados en códigos de

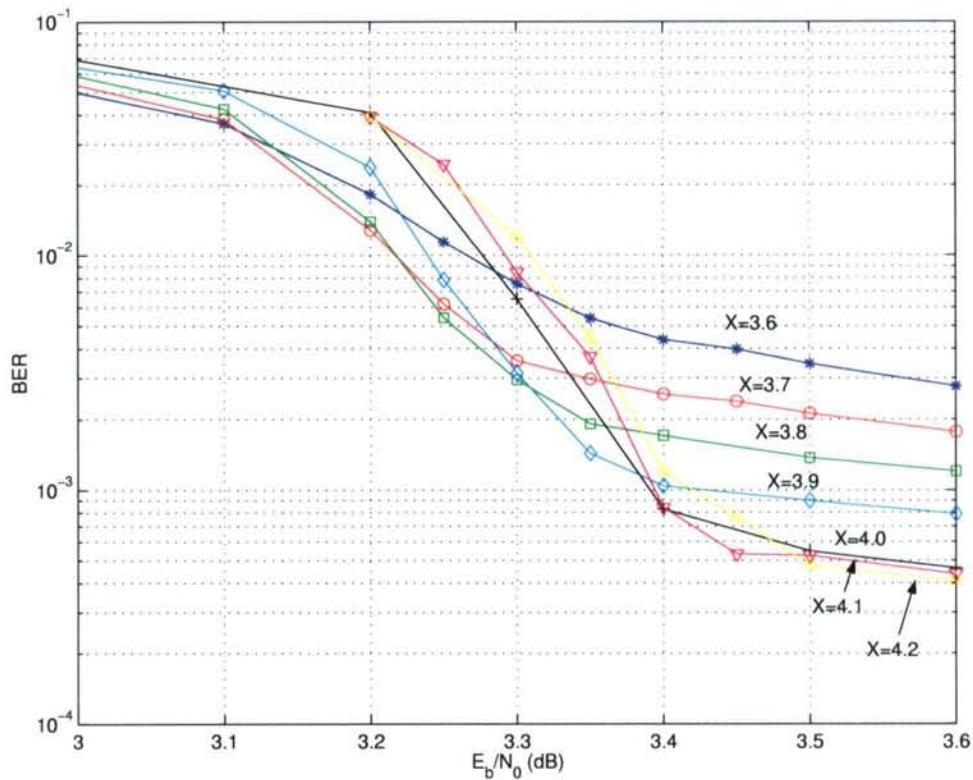


Figura 8.6: Rendimiento del esquema BICM 8PSK propuesto utilizando códigos LDGM irregulares de tasa $R_c = 2/3$ y distintos grados de nodos bit. La longitud de bloque es $K = 30000$ bits de información.

low-density (LDPC y LDGM) utilizan *mapping* Gray. La longitud de bloque para BICM basada en códigos LDGM concatenados es $K = 30000$ bits de información, para BICM basada en códigos LDPC y Turbo-TCM es $K = 20000$ bits de información, y para BICM basada en códigos convolucionales es $K = 4000$ bits de información. Aunque este último valor es mucho menor que los de las otras aproximaciones, los resultados de [64] indican que no se puede esperar una mejora significativa en el rendimiento por el hecho de aumentar la longitud de bloque más allá de $K = 4000$ bits de información. El rendimiento de BICM basada en códigos convolucionales, tanto para el *mapping* Set Partitioning como para el Semi-Set Partitioning, está muy lejos de la de BICM basada en códigos *low-density* (LDPC y LDGM) y de la de Turbo-TCM. El rendimiento de este método cuando se utiliza *mapping* Gray no se muestra ya que es incluso peor.

Es obvio que los mejores resultados en la Figura 8.8 se obtienen por el esquema BICM basado en códigos LDPC irregulares, ya que presenta el mismo BER (10^{-5}) que Turbo-TCM para una $E_b/N_0 = 3,2$ dB y no exhibe ningún *error floor* para valores más altos de la E_b/N_0 . Es más, su complejidad tanto de codificación como de decodificación son menores. Sin embargo, nótese que el rendimiento se ve sustancialmente degradado (alrededor de 0,7 dB para un BER= 10^{-5}) cuando se utilizan códigos LDPC regulares. El rendimiento del esquema

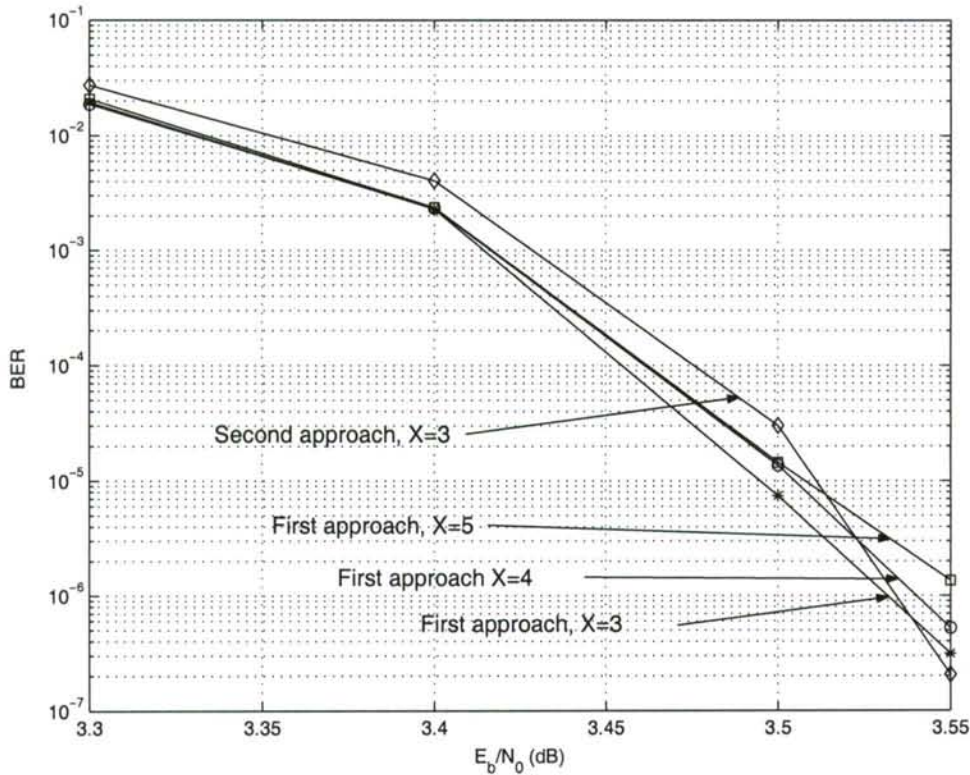


Figura 8.7: Rendimiento del esquema BICM 8PSK propuesto utilizando: i) códigos LDGM concatenados $[(150, X, 30000), (14850, 4, 30150)]$ (primera aproximación) con distintos grados de nodos bit para el código externo ($X = 3, 4, 5$) y ii) códigos LDGM concatenados $[(150, 3, 30000), (15000, 4, 30150)]$ (segunda aproximación). En todos los casos la longitud de bloque es $K = 30000$ bits de información.

BICM propuesto está entre el de los códigos LDPC irregulares y el de los códigos LDPC regulares ($E_b/N_0 = 3,5$ dB para un $BER=10^{-5}$ y $E_b/N_0 = 3,525$ dB para un $BER=10^{-6}$). El esquema propuesto tiene dos ventajas importantes: primera, la complejidad de codificación es menor que la de los códigos LDPC, mientras que su complejidad de decodificación no se ve aumentada de modo significativo. Segunda, los códigos LDGM utilizados en nuestro esquema BICM son regulares y fueron generados de modo pseudoaleatorio, es decir, no se utilizó ningún procedimiento de optimización para construirlos.

8.3.2. Canal Rayleigh

Presentamos a continuación los resultados de nuestro esquema para distintos métodos de señalización sobre un canal con desvanecimiento Rayleigh. La Figura 8.9 muestra el rendimiento de BICM basada en códigos LDGM concatenados para diferentes técnicas de modulación. Para propósitos de comparación, también se muestra el límite de la capacidad del canal (cuando la distribución de entrada se restringe a la modulación empleada). Aunque la

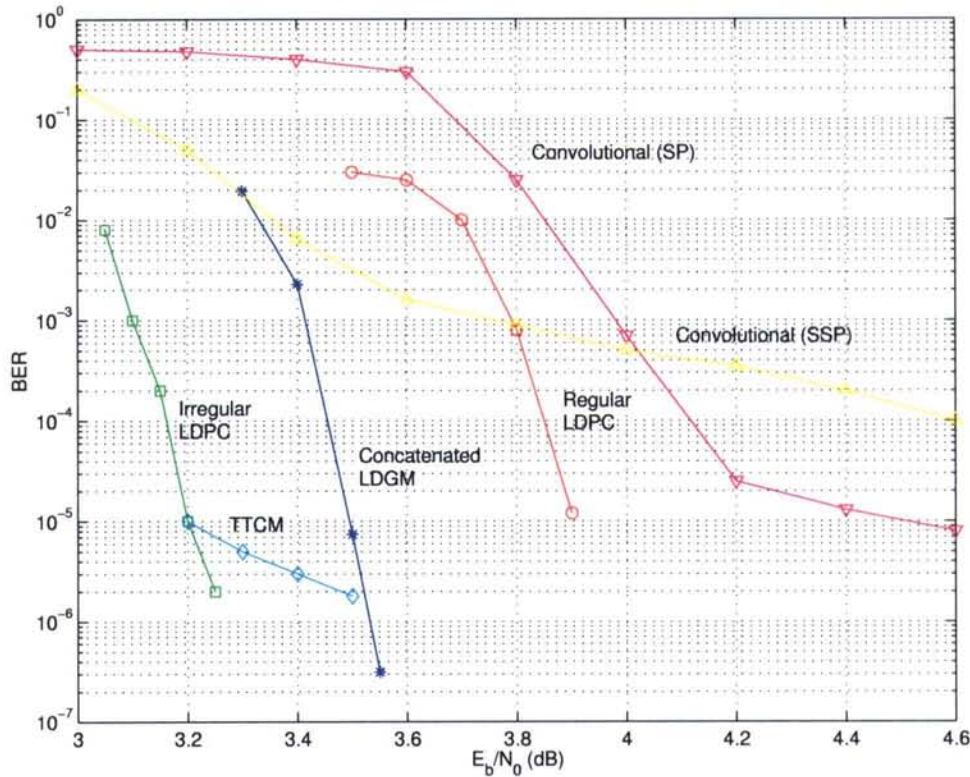


Figura 8.8: Comparación del rendimiento entre BICM utilizando: i) códigos convolucionales con *mappings* Set Partitioning (SP) y Semi-Set Partitioning (SSP), ii) códigos LDPC regulares e irregulares, iii) códigos LDGM concatenados, y iv) Turbo-TCM. La longitud de bloque es $K = 20000$ bits de información para los casos de códigos LDPC y Turbo-TCM, $K = 30000$ bits de información para el código LDGM concatenado, y $K = 4000$ bits de información para el código convolucional.

diferencia con respecto a la capacidad es más amplia que en el caso del canal AWGN, todavía es aceptable y muy similar a presentada en [15], correspondiente a BICM basada en códigos Turbo.

8.3.3. Canal MIMO

Como ya vimos en el Capítulo 3, la capacidad del canal MIMO es

$$C \text{ (bit/s/Hz)} = \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\text{SNR}}{n_t} \mathbf{H}\mathbf{H}^H \right) \quad (8.50)$$

Es importante señalar que la ecuación anterior no hace ninguna suposición sobre la distribución de los símbolos de entrada al canal, $\mathbf{s}(k)$, de modo que la maximización de la información mutua entre la entrada y la salida del canal se realiza sobre todas las posibles distribuciones de entrada. Es conocido que la distribución de entrada que maximiza la información mutua es la gaussiana.

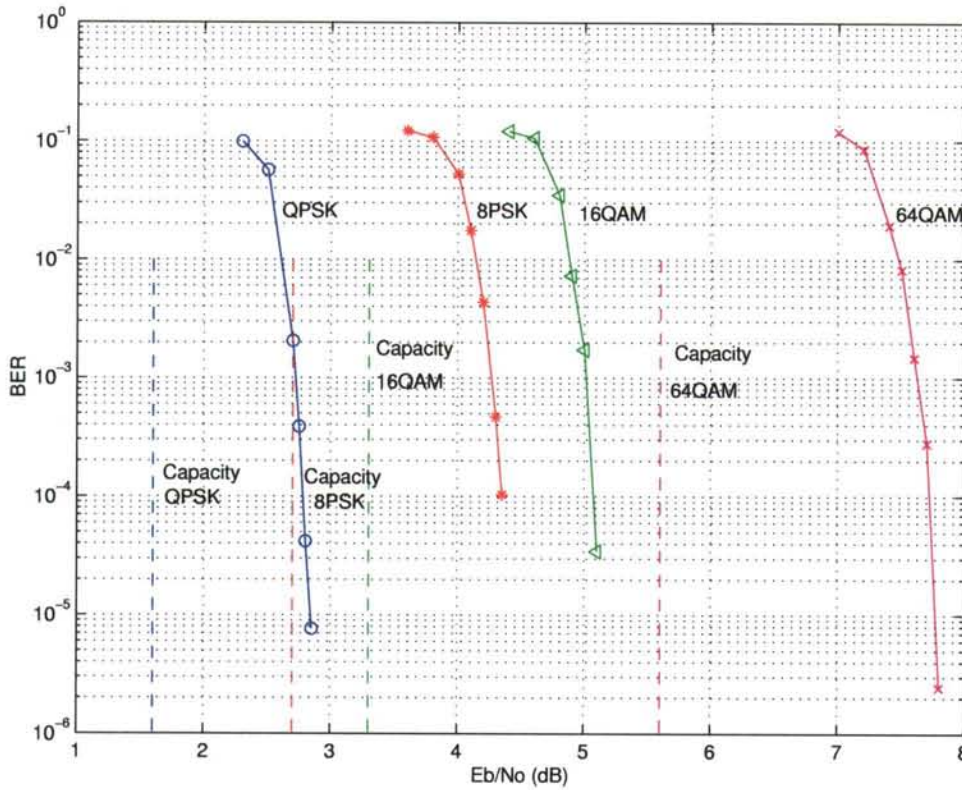


Figura 8.9: Rendimiento del esquema BICM propuesto para distintos métodos de señalización sobre desvanecimiento Rayleigh.

Sin embargo, en la práctica utilizaremos una técnica concreta de modulación, por lo que la distribución estadística de los símbolos de entrada al canal vendrá determinada por las características de dicha modulación. En este caso es conveniente definir la capacidad para entrada restringida (*constrained input*) simplemente como la información mutua entre la entrada y la salida

$$I(\mathbf{s}; \mathbf{x}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{s}) \quad (8.51)$$

donde $H(\cdot) = -E \log p(\cdot)$ es la función entropía. Teniendo en cuenta que $H(\mathbf{x}|\mathbf{s}) = n_r \log_2(\pi N_0 e)$ para el canal gaussiano dado por la Ecuación (8.1) para cualquier constelación de símbolos, es suficiente calcular³

$$H(\mathbf{x}) = -E \log \left(\frac{1}{2^{n_t M_c} (2\pi\sigma^2)^{n_r}} \sum_{\mathbf{s}} \exp \left[-\frac{1}{N_0} \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2 \right] \right)$$

donde la esperanza es sobre las tres fuentes de aleatoriedad, es decir, \mathbf{H} , \mathbf{s} y η . La Figura 8.10 compara la capacidad suponiendo entrada continua con distribución gaussiana con la capacidad para entrada restringida (para QPSK, 8PSK y 16QAM) (8.50) para un canal MIMO 2×2 .

³Para aproximar numéricamente esta esperanza se pueden utilizar métodos de muestreo (Monte Carlo)

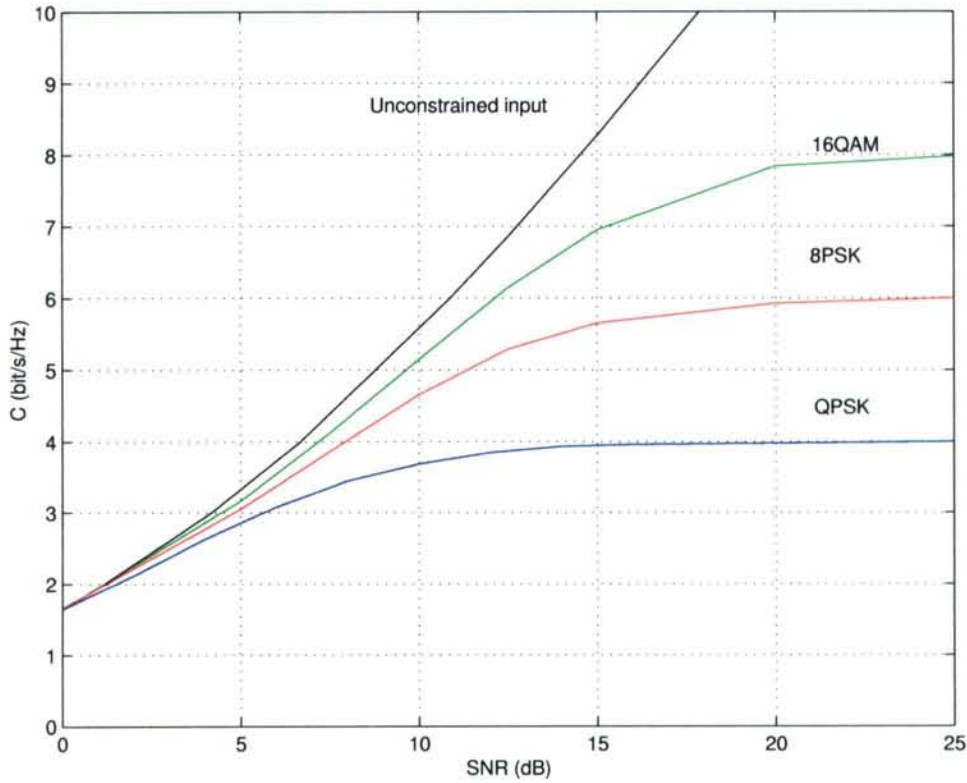


Figura 8.10: Capacidad y capacidad para entrada restringida de un canal MIMO 2×2 para QPSK, 8PSK y 16QAM

La Figura 8.11 muestra el rendimiento del esquema BICM-ID propuesto para un sistema MIMO con dos antenas transmisoras y dos antenas receptoras (2×2). Consideramos un canal de desvanecimiento Rayleigh y conocimiento perfecto del estado del canal (CSI) en recepción. El código utilizado es un código LDGM concatenado $[(500, 4, 15000), (14500, 6, 15500)]$ de tasa $R_c = 1/2$, siendo por tanto la longitud de bloque $N_i = 30000$ bits, donde $K_o = 15000$ son bits sistemáticos y $L_o + L_i = 500 + 14500 = 15000$ son bits de paridad. Se utiliza un entrelazador pseudoaleatorio de longitud igual a la longitud de bloque entre la salida del codificador LDGM concatenado y el *mapper* bit a símbolo. Para cada subbloque de $n_t M_c$ bits asignamos los primeros M_c bits a la primera antena transmisora, los siguientes M_c bits a la segunda, y así sucesivamente. Para cada antena transmisora utilizamos etiquetado Gray, que proporciona un “empuje” suficientemente alto en la primera iteración⁴ para llevar al decodificador LDGM a la región de convergencia. Se realiza una iteración sobre los nodos símbolo (que forman el módulo de *demapping* en nuestro esquema) seguida de diez iteraciones sobre el resto del grafo, que representa el código LDGM concatenado. El algoritmo se para cuando produce la misma secuencia de bits decodificados a lo largo de tres iteraciones sucesivas.

⁴Este “empuje” puede ser medido en términos de la información mutua entre los L-valores *a priori* de los bits codificados y los L-valores extrínsecos *a posteriori* a la salida del módulo de *demapping*, tal y como propone ten

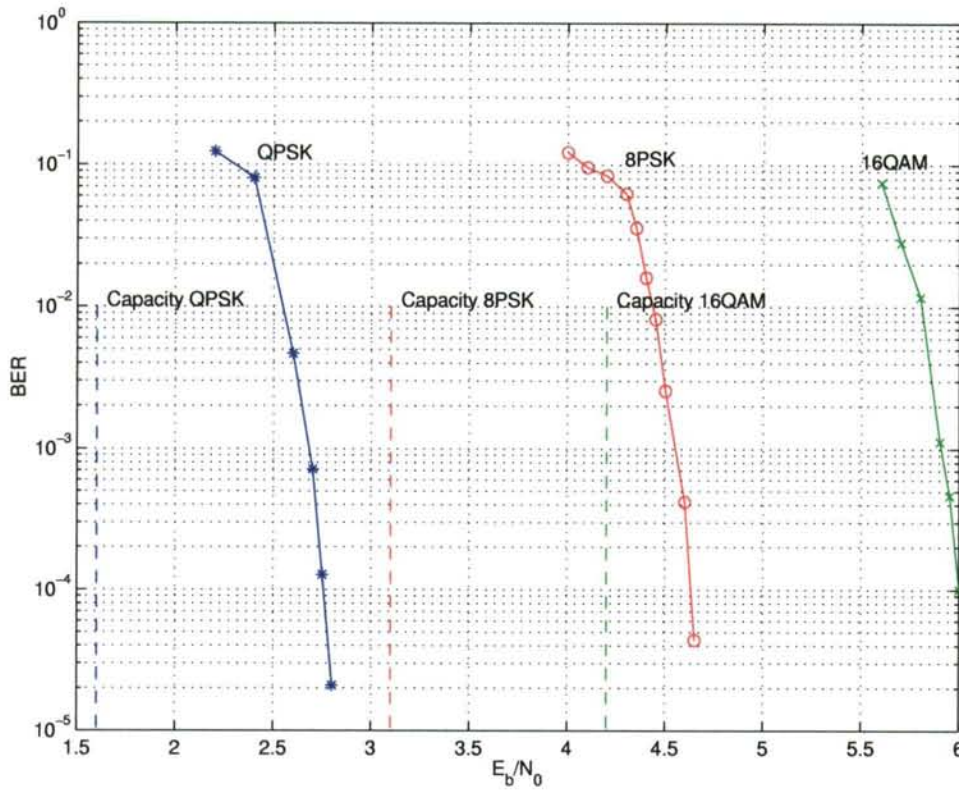


Figura 8.11: Bit Error Rate de BICM basada en códigos LDGM sobre un canal MIMO 2×2

En la Figura 8.11 también se muestra la E_b/N_0 necesaria para alcanzar la capacidad (con distribución de entrada restringida) del canal [15] en bit/s/Hz igual a la tasa de información correspondiente al uso de QPSK, 8PSK y 16QAM, es decir, 2, 3 y 4 bit/s/Hz, respectivamente. Es claro a partir de la Figura 8.11 que el esquema propuesto de BICM-ID basado en códigos LDGM concatenados rinde cerca a la capacidad del canal para las distintas técnicas de modulación consideradas. La diferencia con respecto a la capacidad está en torno a 1.2, 1.6 y 1.9 dB para QPSK, 8PSK y 16QAM, respectivamente.

También estudiaremos el efecto sobre el rendimiento que tiene incrementar el número de antenas transmisoras y receptoras para una modulación dada, en concreto para QPSK. La Figura 8.12 representa el BER frente a la E_b/N_0 obtenido utilizando QPSK para distintas configuraciones MIMO: 2×2 , 3×3 and 4×4 . El resto de los parámetros de simulación son los mismos que antes. La diferencia en rendimiento está alrededor de 0.05 dB y no permite concluir una pérdida de rendimiento al incrementar el número de antenas. Esto resulta en una ganancia en cuanto a *throughput*, ya que se está incrementando el número de antenas (lo cual incrementa la capacidad del canal MIMO, permitiendo utilizar una tasa más alta) sin un incremento de la E_b/N_0 requerida para una decodificación exitosa en recepción.

Nótese que la complejidad del proceso de *demapping* (8.26) es exponencial tanto en

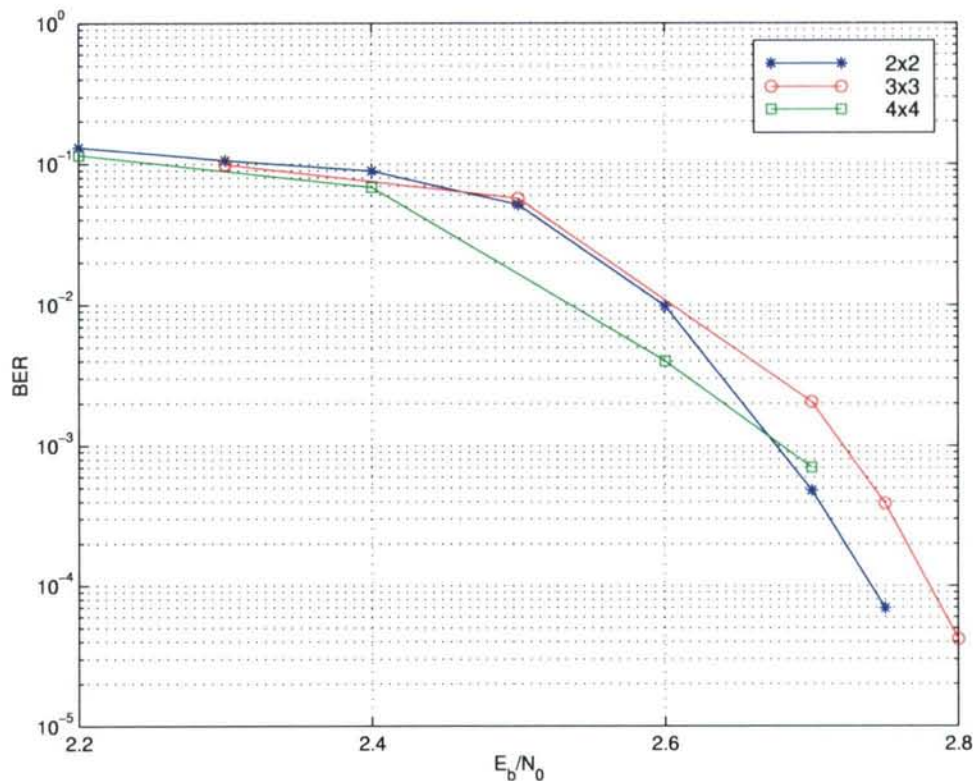


Figura 8.12: Bit Error Rate de BICM basada en códigos LDGM y QPSK para distintos canales MIMO (2×2 , 3×3 y 4×4).

el número de antenas transmisoras, n_t , como en el orden de la modulación, M_c . Como se indicó anteriormente, una manera de superar esta limitación es utilizar esquemas subóptimos de *demapping*.

8.4. Conclusiones

Este capítulo ha estado dedicado a la utilización de códigos *Low-Density Generator Matrix*, LDGM) sistemáticos concatenados como códigos de canal en esquemas *Bit-Interleaved Coded Modulation* (BICM). Tras exponer las ideas fundamentales acerca de códigos LDGM, hemos representado gráficamente el esquema de codificación y modulación mediante una red bayesiana. Esta representación nos ha permitido utilizar el algoritmo *Belief Propagation* de Pearl en la tarea de detectar la secuencia de bits fuente transmitida. Hemos analizado el rendimiento de distintos códigos concatenados en el caso del canal AWGN para obtener la mejor combinación en cuanto al grado de sus nodos y sus tasas de codificación respectivas, elegidas siempre bajo la restricción de mantener una tasa de codificación conjunta fija.

El rendimiento obtenido se ha comparado con el correspondiente a las técnicas utilizadas más recientemente para obtener altas eficiencias espectrales (BICM con codificación

códigos LDPC, con la ventajas que tiene nuestro esquema en cuanto a baja complejidad y a que tan sólo se han contemplado códigos regulares.

Posteriormente, hemos extendido el esquema BICM basado en códigos LDGM concatenados a canales Rayleigh SISO y MIMO. En ambos casos los resultados de simulación muestran cómo este esquema permite obtener un rendimiento próximo al límite teórico de la capacidad del canal, para distintos números de antenas y formatos de modulación.

Capítulo 9

Conclusiones y líneas futuras de investigación

Las ciencias matemáticas en particular exhiben orden, simetría y limitación: y éstas son las mayores formas de la belleza

—Aristóteles (384-322 A.C.) “Metaphysica”

Hoy en día, no es sólo que nuestros reyes no sepan matemáticas, es que nuestros filósofos no saben matemáticas

—J. Robert Oppenheimer (1904-1967)

9.1. Conclusiones

El objetivo de esta tesis ha sido la investigación de distintas técnicas de procesamiento de señal y codificación especialmente concebidas para sistemas Multiple-Input Multiple-Output (MIMO). Estos sistemas se caracterizan por el uso de múltiples antenas tanto en transmisión como en recepción de un canal con multitrayecto que resulta de un entorno de propagación electromagnético. El interés de las técnicas propuestas proviene, en última instancia, de la gran capacidad asociada al canal MIMO considerado. Dicha capacidad crece de modo directamente proporcional al mínimo del número de antenas transmisoras y receptoras.

Más en detalle, se ha propuesto un esquema MIMO de comunicaciones inalámbricas para el entorno de propagación presente en los túneles subterráneos. La elevada capacidad asociada a los canales MIMO permitiría la transmisión a gran velocidad de flujos de datos sin incurrir para ello en un incremento del ancho de banda utilizado y/o la potencia requerida. El hecho de disponer de un enlace de elevada tasa de bit tiene múltiples aplicaciones en entornos de este

tipo. Entre ellas cabe citar la transmisión de señales de control, imágenes y vídeo para los trenes de los metros.

El esquema propuesto utiliza la misma señalización que el estándar de comunicaciones celular GSM (*Global System Mobile*) de uso generalizado en Europa como estándar de segunda generación de telefonía móvil. Los sistemas europeos de comunicación radio en ferrocarriles han adoptado desde 1993 la interfaz radio GSM (GSM-R), reservándose además para ellos una banda de frecuencias concreta, denominada la banda "R" (876-880 MHz y 921-925 MHz). Debido a la naturaleza conservadora de este mercado, es razonable suponer que los sistemas de comunicación radio para ferrocarriles emplearán GSM-R en los próximos años haciendo, por tanto, más atractivo el esquema MIMO de comunicación radio propuesto.

El análisis del rendimiento de un sistema de comunicación en un entorno de propagación multitrayecto suele hacerse asumiendo un determinado modelo probabilístico para la distorsión introducida por el canal sobre la señal transmitida. Los modelos probabilísticos son útiles para poder comparar técnicas diferentes sin necesidad de disponer de medidas reales de la respuesta al impulso del canal. Sin embargo, dada la complejidad de los fenómenos físicos causantes de la distorsión de la señal transmitida en un entorno multitrayecto, los modelos probabilísticos utilizados habitualmente no se ajustan suficientemente bien a la realidad que describen. Dicho de otra manera, canales muy diferentes en la realidad pueden representarse exactamente con los mismos valores de los parámetros del modelo probabilístico utilizado. Considerando que lo que estamos haciendo al utilizar un modelo probabilístico es agrupar en clases la totalidad de las realizaciones posibles del canal, podemos decir que los modelos probabilísticos habitualmente empleados tienen una granularidad demasiado gruesa (i.e., cada clase agrupa demasiados elementos). Este defecto no es demasiado importante a la hora de comparar dos técnicas diferentes de procesado y/o codificación, permitiendo disponer de un modelo matemático de suficiente sencillez como para poder ser analizado y/o simulado. Por el contrario, si nuestro interés es poder predecir el comportamiento de una técnica determinada, las simplificaciones realizadas al enunciar el modelo probabilístico tienen un impacto mucho mayor. Al contrario de lo que ocurre con la modelización del ruido térmico como un proceso aditivo gaussiano y blanco (AWGN), los modelos probabilísticos utilizados para describir el canal no reflejan con suficiente precisión los procesos físicos que describen.

Por todo ello, el esquema MIMO de tipo GSM propuesto ha sido evaluado utilizando medidas experimentales del canal real presente en los túneles subterráneos. Se emplearon las medidas de la respuesta al impulso del canal presente en los túneles de la red de Metro de París, realizadas dentro del contexto del proyecto europeo ESCORT (*Enhanced diversity and Space-time Coding for metrO and Railway Transmission*). La diferencia más importante que presentaban las medidas reales con respecto a los modelos probabilísticos utilizados habitualmente por la comunidad científica era la presencia de fuertes correlaciones espaciales entre los coeficientes del canal que influyen negativamente en la capacidad del canal.

Se evaluaron técnicas ya clásicas en el contexto de los sistemas de comunicación espacio-temporales. Primeramente, se consideraron las técnicas sin codificación de canal más

representativas: Máxima Verosimilitud (MV), Separación Lineal de Señales (*Linear Signal Separation*, LSS) y Cancelación Sucesiva (*Successive Cancellation*, SC). Pese a ser el canal medido experimentalmente plano en frecuencia, la modulación GMSK utilizada en el estándar GSM introduce Interferencia InterSimbólica (*InterSymbol Interference*, ISI) controlada, por lo que todos los ecualizadores contemplados debían compensar dicha ISI. Los ecualizadores espacio-temporales considerados se pueden ordenar en cuanto a rendimiento y complejidad computacional del siguiente modo: MV, SC y LSS (de mayor a menor complejidad y rendimiento). Las diferencias de rendimiento observadas fueron muy significativas, siendo el ecualizador MV el único que proporcionaba un rendimiento adecuado medido en términos de la relación señal a ruido necesaria para alcanzar una determinada tasa de error de bit.

A continuación se consideraron esquemas que tuviesen en cuenta la existencia de un codificador de canal que incorpore redundancia como paso previo a la transmisión. En primer lugar se estudió el clásico esquema de un ecualizador MV (implementado mediante el algoritmo de Viterbi) que produce decisiones *hard* seguido de un decodificador de canal de Viterbi. El código de canal considerado fue un código convolucional espacio-temporal. Este esquema es ampliamente superado por un esquema iterativo que utilice decisiones *soft* basado en el principio Turbo. Aunque las técnicas consideradas no son una aportación de esta tesis doctoral, sí lo es el esquema de tipo GSM planteado y, sobre todo, los resultados de rendimiento obtenidos utilizando medidas experimentales reales. Además, la Turbo-ecualización en sistemas MIMO es algo de mucha actualidad.

En sistemas prácticos, donde el canal existente entre transmisor y receptor es desconocido, es necesario incorporar un módulo de estimación de canal al proceso de ecualización, o bien disponer de un ecualizador cuyos coeficientes sean ajustados mediante un proceso de entrenamiento, ya sea ciego o semi-ciego. En ese sentido, son dos las aportaciones realizadas en esta tesis doctoral: por una parte, un módulo de estimación de canal y, por otra, un ecualizador lineal. En ambos casos se utiliza el principio MV, bien para obtener la estimación de la matriz de canal, bien para ajustar los coeficientes del filtro lineal que actúa como ecualizador. Ya que la aplicación del principio MV resulta en un problema de optimización para el que no existe solución en forma cerrada, se utiliza el algoritmo *Expectation Maximization* (EM) para obtener de forma iterativa la solución de dicho problema. En ambos casos, cada iteración del algoritmo EM consiste en resolver un problema de optimización cuadrático, para el que sí existe solución cerrada. Además, ambos esquemas se benefician de las probabilidades *a posteriori* obtenidas por el ecualizador Turbo, haciendo que su rendimiento sea muy superior a esquemas resultantes de la aplicación del criterio de mínimos cuadrados (*Least Squares*, LS). Se muestra una posible aplicación de esta ganancia en rendimiento para el sistema de comunicación MIMO de tipo GSM para entornos de túneles subterráneos propuesto: el estimador de canal MV permite reducir la longitud de la secuencia de entrenamiento requerida de los 26 bits contemplados en el estándar GSM a tan sólo 5 bits, con el incremento en tasa de transferencia que ello supone.

Por último, se han extendido los resultados existentes sobre códigos *Low-Density Generator Matrix* (LDGM) a sistemas de tipo *Bit-Interleaved Coded Modulation* (BICM), para canales

AWGN, Rayleigh y MIMO. Los resultados obtenidos son comparables a los ya obtenidos para códigos Turbo y *Low-Density Parity-Check Matrix* (LDPC), pero con la ventaja de la menor complejidad de codificación (lineal) asociada a los códigos LDGM. El mayor problema es el crecimiento exponencial de la complejidad del módulo de *demapping* en el número de antenas transmisoras y el tamaño de la constelación utilizada.

9.2. Líneas futuras de investigación

Estructuraremos esta sección de acuerdo a los distintos tópicos tratados en esta tesis.

9.2.1. Estimación de canal y ecualización *Decision Feedback*

En cuanto a la estimación de canal MV-EM y la ecualización lineal basada en el mismo principio, hay que resaltar que su rendimiento es muy cercano al óptimo (es decir, al caso de suponer el canal conocido), por lo que poco puede hacerse en este sentido. Sí podría plantearse el estudio de la reducción de la complejidad asociada a dichos algoritmos.

Pero sin lugar a dudas, el problema más interesante a abordar en este sentido es el bajo rendimiento de dichos algoritmos cuando el canal varía en el tiempo, y hay una correlación temporal y/o espacial en los coeficientes del canal. En esta línea, están siendo exploradas por la comunidad científica alternativas basadas en modelos ocultos de Markov (*Hidden Markov Model*, HMM). La idea es aproximar la variación del canal en tiempo continuo mediante un modelo discreto basado en HMMs.

Además, del mismo modo que hemos presentado aquí esquemas iterativos Turbo adecuados a la situación de canal desconocido en el receptor, se puede aplicar el mismo principio para incluir en el proceso iterativo el módulo de sincronización.

9.2.2. Reducción de la complejidad del *demapping* mediante detección esférica

En el apartado de códigos LDGM para BICM en sistemas MIMO, se está actualmente abordando la aplicación de un método subóptimo de *demapping* basado en la detección esférica. De modo breve, podemos decir que el método consiste en seleccionar, mediante detección esférica, aquellos candidatos con mayor verosimilitud de entre todos los posibles vectores transmitidos para, a continuación efectuar el *demapping* utilizando sólo esos candidatos cuyo número, evidentemente, es mucho menor que el total de posibles vectores transmitidos.

La detección esférica es posible cuando la constelación está basada en una celosía (*lattice*). Una celosía (*lattice*) Λ en el espacio euclídeo \mathbb{R}^n se define como un conjunto infinito de vectores n -dimensionales cerrado con respecto a la suma y la multiplicación por enteros. Estas propiedades hacen que una celosía sea un grupo. En consecuencia, la celosía tiene la misma estructura independientemente del punto en el que se sitúe el observador: cada punto de la

celosía tiene el mismo número de puntos vecinos en cada posible distancia. Además las regiones de decisión correspondientes a un detector de distancia mínima (denominadas *Regiones de Voronoi*) son congruentes y crean un mosaico de \mathbb{R}^n . Una base de una celosía es un conjunto de vectores $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ de \mathbb{R}^m linealmente independientes tal que cualquier elemento de la celosía puede ser expresado como una combinación lineal, con coeficientes enteros, de los elementos de la base. Utilizando una base de la celosía, podemos expresar cualquier elemento \mathbf{s} de la misma como $\mathbf{s} = \mathbf{A}\mathbf{u}$, donde $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_m]$ es la matriz que contiene los vectores de la base (matriz generadora) y \mathbf{u} es un vector de enteros que contiene las coordenadas de \mathbf{s} en la base \mathbf{A} . Se dice que dos celosías son equivalentes si una de ellas se puede obtener mediante rotación, reflexión o escalado de la otra.

Una constelación de celosía $C(\Lambda, \mathcal{R}) \triangleq \Lambda \cap \mathcal{R}$ es el conjunto finito de puntos de la celosía Λ interiores a la región \mathcal{R} del espacio \mathbb{R}^n . Un ejemplo de constelación de celosía es la correspondiente a la modulación M -PAM

$$\frac{d_0}{2} \{\pm 1, \pm 3, \dots, \pm(M-1)\} \quad (9.1)$$

donde M es un número par y d_0 es la distancia entre cualesquiera dos puntos de la constelación. Esta constelación la podemos denotar por $C(d_0\mathbb{Z} + 1/2, (d_0/2)[-M, M])$.

Pasemos entonces a describir el algoritmo de detección esférica. Sea la señal transmitida $\mathbf{s} = \mathbf{A}\mathbf{u}$, donde \mathbf{A} es la matriz generadora de la celosía y \mathbf{u} las coordenadas enteras de la señal transmitida en la base \mathbf{A} . En el caso de un canal AWGN, la señal recibida es

$$\mathbf{x} = \mathbf{s} + \mathbf{n} \quad (9.2)$$

El detector óptimo (MV) es equivalente a escoger como vector transmitido el más próximo en distancia euclídea al vector recibido

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s} \in \Lambda} \|\mathbf{x} - \mathbf{s}\|^2 = \arg \max_{\mathbf{w} \in \mathbf{x} - \Lambda} \|\mathbf{w}\|^2 \quad (9.3)$$

lo que quiere decir que el problema es encontrar el vector $\mathbf{w} \triangleq \mathbf{x} - \mathbf{s}$ con menor módulo. El algoritmo de detección esférica consiste en fijar un radio C y buscar dentro de la esfera centrada en el origen y radio C el vector \mathbf{w} con menor módulo. Es decir, limitamos la búsqueda a los vectores que cumplen

$$\|\mathbf{w}\|^2 \leq C \quad (9.4)$$

Escogiendo bien el radio C podemos limitar efectivamente el número de vectores \mathbf{w} a evaluar. Si el radio es demasiado grande, el número de vectores a evaluar puede hacer inviable la búsqueda. Si el radio es demasiado pequeño, la esfera no contendrá ningún \mathbf{w} cuyo vector \mathbf{s} correspondiente pertenezca a la constelación utilizada. En este caso, el algoritmo puede o bien volver a empezar escogiendo un radio mayor, o simplemente declarar un fallo de detección (*erasure*). Otro aspecto clave del algoritmo de detección esférica está en calcular de modo eficiente las condiciones que deben cumplir los puntos interiores a la esfera considerada. A continuación explicaremos el procedimiento habitual.

Dado que la base \mathbf{A} también expande el espacio \mathbb{R}^n , podemos escribir $\mathbf{w} = \mathbf{A}\tilde{\mathbf{w}}$, es decir, expresamos el vector \mathbf{w} en coordenadas (reales) de la base \mathbf{A} . Tenemos entonces que

$$\mathbf{w} \triangleq \mathbf{x} - \mathbf{s} \quad (9.5)$$

$$\mathbf{A}\tilde{\mathbf{w}} = \mathbf{A}\tilde{\mathbf{x}} - \mathbf{A}\mathbf{u} \quad (9.6)$$

$$\tilde{\mathbf{w}} = \tilde{\mathbf{x}} - \mathbf{u} \quad (9.7)$$

donde \tilde{x}_i , $i = 1, 2, \dots, n$ son las coordenadas reales del vector recibido \mathbf{x} en la base \mathbf{A} . Escribamos ahora la norma cuadrática del vector \mathbf{w} utilizando las coordenadas anteriores

$$\|\mathbf{w}\|^2 = (\mathbf{A}\tilde{\mathbf{w}})^T(\mathbf{A}\tilde{\mathbf{w}}) = \tilde{\mathbf{w}}^T \mathbf{A}^T \mathbf{A} \tilde{\mathbf{w}} \leq C \quad (9.8)$$

La matriz $\mathbf{G} = \mathbf{A}^T \mathbf{A}$ es una matriz de Gram¹, por lo que podemos utilizar su factorización de Cholesky $\mathbf{G} = \mathbf{U}^T \mathbf{U}$, siendo \mathbf{U} una matriz triangular superior, para escribir

$$\|\mathbf{w}\|^2 = \tilde{\mathbf{w}}^T \mathbf{U}^T \mathbf{U} \tilde{\mathbf{w}} \quad (9.9)$$

$$= \|\mathbf{U}\tilde{\mathbf{w}}\|^2 \quad (9.10)$$

$$= \sum_{i=1}^n \left(u_{ii}\tilde{w}_i + \sum_{j=i+1}^n r_{ij}\tilde{w}_j \right)^2 \leq C \quad (9.11)$$

Introduciendo la notación $q_{ii} \triangleq r_{ii}^2$ y $q_{ij} \triangleq r_{ij}/r_{ii}$ para $i = 1, 2, \dots, n$, $j = i+1, 2, \dots, n$ podemos expresar la condición anterior como

$$\|\mathbf{w}\|^2 = \sum_{i=1}^n q_{ii} \left(\tilde{w}_i + \sum_{j=i+1}^n q_{ij}\tilde{w}_j \right)^2 \leq C \quad (9.12)$$

Dado que $\tilde{\mathbf{w}} = \tilde{\mathbf{x}} - \mathbf{u}$, la condición (9.12) la podemos expresar con respecto a las coordenadas enteras de los posibles vectores transmitidos como

$$\sum_{i=1}^n q_{ii} \left((\tilde{x}_i - u_i) + \sum_{j=i+1}^n q_{ij}(\tilde{x}_j - u_j) \right)^2 \leq C \quad (9.13)$$

Los términos de la suma anterior son todos positivos, ya que son el cuadrado de cada una de las coordenadas \tilde{w}_i , por lo que el algoritmo de detección esférica efectúa la búsqueda de los puntos interiores a la esfera de modo recursivo del siguiente modo: primero selecciona un valor (entero) para la coordenada u_n de aquellos incluidos en la constelación utilizada que cumpla

$$q_{nn}(\tilde{x}_n - u_n)^2 \leq C \quad (9.14)$$

¹Dada una matriz de vectores de tamaño $n \times 1$, $\mathbf{A} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$, la matriz de Gram de \mathbf{A} se define como la matriz de los productos escalares de dichos vectores, es decir, $\mathbf{G}_{ij} = \mathbf{v}_i^T \mathbf{v}_j = \mathbf{A}^T \mathbf{A}$. Cualquier matriz de Gram es definida positiva, por lo que siempre admite la descomposición de Cholesky.

que al ser u_n entero es equivalente a

$$\left[-\sqrt{\frac{C}{q_{nn}}} + \tilde{x}_n \right] \leq u_n \leq \left[\sqrt{\frac{C}{q_{nn}}} + \tilde{x}_n \right] \quad (9.15)$$

y después evalúa las consecuencias de esta elección sobre los posibles valores que puede escoger para la coordenada u_{n-1} . Nótese la gran diferencia entre este proceso y la cancelación sucesiva, que en este punto ya realiza una decisión definitiva sobre u_n . Para ver la consecuencia de dicha elección escoge un valor para u_{n-1} entre aquellos enteros que cumplen

$$\left[-\sqrt{\frac{C - q_{nn}\tilde{w}_n^2}{q_{n-1,n-1}}} + \tilde{x}_{n-1} + q_{n-1,n}\tilde{w}_n \right] \leq u_{n-1} \leq \left[\sqrt{\frac{C - q_{nn}\tilde{w}_n^2}{q_{n-1,n-1}}} + \tilde{x}_{n-1} + q_{n-1,n}\tilde{w}_n \right] \quad (9.16)$$

Si el rango anterior no incluye ningún valor u_{n-1} de los contemplados en la constelación, el algoritmo vuelve atrás (*back-tracking*) y selecciona otro valor para u_n . En caso contrario, escoge uno de esos posibles u_{n-1} y pasa a evaluar las consecuencias de esa decisión sobre u_{n-2} . El proceso continúa hasta que se dispone de un candidato $\mathbf{u} = [u_1 \ u_2 \ \dots \ u_n]^T$, interior a la esfera y perteneciente a la constelación. Definiendo

$$S_i \triangleq S_i(\tilde{w}_{i+1}, \tilde{w}_{i+2}, \dots, \tilde{w}_n) \triangleq \tilde{x}_i + \sum_{l=i+1}^n q_{il}\tilde{w}_l \quad (9.17)$$

$$T_{i-1} \triangleq T_{i-1}(\tilde{w}_i, \tilde{w}_{i+1}, \dots, \tilde{w}_n) \triangleq C - \sum_{l=i}^n q_{ll} \left(\tilde{w}_l + \sum_{j=l+1}^n q_{lj}\tilde{w}_j \right)^2 \quad (9.18)$$

$$= T_i - q_{ii}(S_i - u_i)^2 \quad (9.19)$$

podemos expresar la condición a evaluar en la etapa i -ésima como

$$\left[-\sqrt{\frac{1}{q_{ii}}T_i} + S_i \right] \leq u_i \leq \left[\sqrt{\frac{1}{q_{ii}}T_i} + S_i \right] \quad (9.20)$$

Además, una vez que se ha hallado un vector candidato, su distancia al vector recibido se puede calcular como

$$\|\mathbf{x} - \mathbf{A}\mathbf{u}\|^2 = C - T_1 + q_{11}(S_1 - u_1)^2 \quad (9.21)$$

Una vez hallado un vector candidato el algoritmo fija un nuevo radio C' igual a la distancia de dicho candidato al vector recibido, y procede a la búsqueda de un nuevo candidato. El proceso se repite hasta que todos los posibles vectores hayan sido evaluados.

Aplicación a sistemas MIMO

Consideremos el modelo de señal para un canal MIMO plano en frecuencia

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (9.22)$$

Si \mathbf{s} es un vector perteneciente a una constelación de celosía, tendremos $\mathbf{s} = \mathbf{A}\mathbf{u}$, donde \mathbf{A} es la matriz generadora de la celosía y \mathbf{u} las coordenadas enteras de \mathbf{s} en esa base. Consideraremos valores reales, representando el caso de valores complejos como el valor real de las partes real e imaginaria. Así, el modelo anterior lo podemos escribir en el dominio real como

$$\tilde{\mathbf{x}} = \tilde{\mathbf{H}}\tilde{\mathbf{s}} + \tilde{\mathbf{n}} \quad (9.23)$$

donde $\tilde{\mathbf{x}} \triangleq [\Re(\mathbf{x}^T) \Im(\mathbf{x}^T)]^T$, $\tilde{\mathbf{s}} \triangleq [\Re(\mathbf{s}^T) \Im(\mathbf{s}^T)]^T$, $\tilde{\mathbf{n}} \triangleq [\Re(\mathbf{n}^T) \Im(\mathbf{n}^T)]^T$ y

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \Re(\mathbf{H}) & \Im(\mathbf{H}) \\ -\Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix} \quad (9.24)$$

La matriz $\tilde{\mathbf{H}}$ habitualmente será de rango completo, por lo que podemos considerar $\tilde{\mathbf{A}} \triangleq \tilde{\mathbf{H}}\mathbf{A}$ como matriz generadora de la celosía y obtener el modelo de señal

$$\tilde{\mathbf{x}} = \tilde{\mathbf{A}}\mathbf{u} + \tilde{\mathbf{n}} \quad (9.25)$$

sobre el cual podemos aplicar el algoritmo de detección esférica.

9.2.3. Reducción de la complejidad del *demapping* mediante cancelación *soft*

Una alternativa al algoritmo de *demapping* basado en detección esférica es lo que se conoce como cancelación *soft*. El mensaje que debe entregar el nodo símbolo $\mathbf{s}(k)$ a su nodo padre W_j viene dado por la Ecuación (8.26). En adelante prescindiremos del índice k a efectos de claridad en la exposición. Recordemos que el problema de la complejidad viene dado porque el número de valores posibles que puede tomar s_k en el primer sumatorio es $2^{n_t M_c}$ que, en muchos casos, es demasiado elevado. Una alternativa a la solución basada en detección esférica consiste en, primeramente, recurrir a la definición original del mensaje, i.e. $\lambda_{\mathbf{s}}(w_j) \triangleq P[\mathbf{e}_{\mathbf{s}} \setminus e_{\mathbf{s}}^{W_j} | w_j]$. En el caso del canal MIMO, esta expresión es equivalente a la verosimilitud $\lambda_{\mathbf{s}}(w_j) = P[\mathbf{x} | w_j]$. La Ecuación (8.26) calcula el valor exacto de esta verosimilitud mediante la marginalización con respecto a todos los posibles valores que puede tomar \mathbf{s} . Sin embargo, es claro que podemos aproximar dicha verosimilitud del siguiente modo

$$P[\mathbf{x} | w_j] \approx P[\mathbf{x} | w_j, \hat{\mathbf{w}}_{(j)}] \quad (9.26)$$

donde $\hat{\mathbf{w}}_{(j)} \triangleq [\hat{w}_1, \dots, \hat{w}_{j-1}, \hat{w}_{j+1}, \dots, \hat{w}_{n_t M_c}]^T$ son estimaciones del valor tomado por dichos nodos. Podemos utilizar como estimador la esperanza con respecto a la f.d.p. *a posteriori* calculada en el proceso de decodificación, i.e.

$$\hat{w}_l = E_{w_l | \mathbf{e}_{W_l} \setminus e_{W_l}^{s_k}} \{w_l\} = \sum_{w_l} w_l \pi_{\mathbf{s}}(w_l) \quad (9.27)$$

Asumiremos que la asignación bit a símbolo se realiza de modo lineal, i.e. $\mathbf{s} = \mathbf{A}[w_1, w_2, \dots, w_{n_t M_c}]^T$, tal y como ocurre en las constelaciones basadas en celosías. De este modo podemos escribir

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{g} \quad (9.28)$$

$$= \mathbf{HA}[\hat{w}_1, \dots, \hat{w}_{j-1}, w_j, \hat{w}_{j+1}, \dots, \hat{w}_{n_t M_c}]^T + \mathbf{g} \quad (9.29)$$

que, denotando por \mathbf{h}_j^a a la j -ésima columna de la matriz \mathbf{HA} y por $\mathbf{HA}_{(j)}$ a la matriz que resulta de suprimir dicha columna, se puede expresar como

$$\mathbf{x} = \mathbf{h}_j^a w_j + \mathbf{HA}_{(j)} \hat{\mathbf{w}}_{(j)} + \mathbf{g} \quad (9.30)$$

Utilizando la notación $\mathbf{w}_{(j)} = [w_1, w_2, \dots, w_{j-1}, w_{j+1}, \dots, w_{n_t M_c}]^T$, la verosimilitud (9.26) se puede calcular como

$$P[\mathbf{x}|w_j, \hat{\mathbf{w}}_{(j)}] = P[\mathbf{x} - \mathbf{h}_j^a w_j - \mathbf{HA}_{(j)} \hat{\mathbf{w}}_{(j)}] \quad (9.31)$$

$$= P[\mathbf{HA}_{(j)}(\mathbf{w}_{(j)} - \hat{\mathbf{w}}_{(j)}) + \mathbf{g}] \quad (9.32)$$

$$\triangleq P[\mathbf{HA}_{(j)}\mathbf{z} + \mathbf{g}] \quad (9.33)$$

Finalmente, asumiendo que el estimador $\hat{\mathbf{w}}_{(j)}$ es insesgado y que los residuos \mathbf{z} son gaussianos de media cero y con matriz de covarianzas \mathbf{C}_z , finalmente tenemos que

$$P[\mathbf{x}|w_j, \hat{\mathbf{w}}_{(j)}] \triangleq \exp\{-\tilde{\mathbf{x}}^H \mathbf{C}_{\tilde{\mathbf{x}}}^{-1} \tilde{\mathbf{x}}\} \quad (9.34)$$

donde $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{h}_j^a w_j - \mathbf{HA}_{(j)} \hat{\mathbf{w}}_{(j)}$ y $\mathbf{C}_{\tilde{\mathbf{x}}} = \mathbf{HA}_{(j)} \mathbf{C}_z (\mathbf{HA}_{(j)})^H + \sigma_g^2 \mathbf{I}$. Obviamente, en la práctica la matriz de covarianzas \mathbf{C}_z es desconocida y debe ser estimada.

9.2.4. Desglose de los nodos símbolo

Hasta ahora hemos considerado que los nodos símbolo en la red bayesiana que representa el esquema conjunto de codificación y transmisión MIMO estaban constituídos por vectores \mathbf{s} . Desglosando cada uno de estos vectores en sus correspondientes componentes escalares, obtenemos un grafo equivalente pero con un número mayor de nodos símbolo. Aunque este grafo representa la misma información, posibilita que el algoritmo BP actúe sobre una red de granularidad más fina. Por una parte, se intuye que siguiendo un orden de activación de los nodos de la red adecuado, el coste computacional medio para alcanzar la convergencia sea menor. Además, sobre este esquema se pueden aplicar los dos métodos subóptimos que acabamos de describir con una granularidad más fina. Esto posibilita soluciones de compromiso entre coste computacional y rendimiento que utilicen para algunos nodos el método óptimo de cálculo de los mensajes mientras que para otros se utilice uno subóptimo. En resumen, un mayor desglose de la red bayesiana nos proporciona la ventaja de construir esquemas de decodificación más flexibles.

9.2.5. Transmisión adaptativa utilizando códigos LDGM

La capacidad del canal cuando el transmisor conoce el canal es mayor que cuando lo desconoce, especialmente en el régimen de baja SNR. Recientemente, las técnicas que consideran que el transmisor dispone de información acerca del estado del canal han recibido gran atención, tanto en el contexto de los sistemas SISO como MIMO. En función de las características del canal, el transmisor puede ajustar la potencia utilizada en cada portadora y/o modo del canal (en adelante utilizaremos el término “subcanal” para referirnos a cualquiera de estos dos términos). Además, se puede ajustar el número de bits asignados a cada subcanal seleccionando modulaciones con diferentes eficiencias espectrales.

En cuanto a la codificación de canal, es lógico utilizar en cada subcanal una tasa de codificación diferente, de acuerdo a su capacidad teórica (que depende directamente de la SNR existente en el mismo). Un ejemplo de códigos de tasa ajustable son los códigos *Rate Compatible Punctured Convolutional* (RCPC). Además de que los códigos LDGM operan en valores de SNR cercanos al límite teórico marcado por la capacidad, su estructura permite un *puncturing* muy fino, pudiéndose conseguir prácticamente cualquier tasa de codificación. Esto supone una notable diferencia respecto a lo que ocurre con los códigos RCPC, cuya conjunto de tasas posibles es de granularidad muy gruesa. Por esta razón, es de esperar que la utilización de códigos LDGM con tasa variable proporcione mejor rendimiento que el obtenido con códigos RCPC o Turbo.

9.2.6. Códigos LDGM irregulares

Como ya hemos comentado, se pueden obtener grandes ganancias si se consideran códigos LDPC irregulares, es decir, aquéllos en los que el grado de un nodo, i.e. el número de nodos al que está conectado, es distinto para los diferentes nodos que componen el grafo. La construcción de códigos irregulares requiere resolver el problema de optimización relativo a qué proporción de nodos poseen un grado determinado. En este sentido, es de esperar que la consideración de códigos LDGM irregulares también aporte ganancias significativas. Además, es posible aprovechar todo el trabajo realizado en esta línea con códigos LDPC, eso sí, teniendo en mente la diferente estructura de los códigos LDGM.

La optimización del grado de los nodos en un código LDGM irregular es también una vía prometedora para la obtención de códigos de longitud de bloque pequeña, restricción necesaria en determinadas aplicaciones que no toleran las longitudes de bloque consideradas en este trabajo.

9.2.7. Aplicación de diagramas EXIT a la corrección del *error floor*

Los diagramas EXIT (*EXtrinsic Information Transfer*) fueron introducidos por S. ten Brink para el análisis de la convergencia de los métodos de procesado iterativo Turbo. Posteriormente también ha sido aplicado al diseño del módulo de asignación bit a símbolo en esquemas BICM.

Esencialmente, el uso de diagramas EXIT permite el estudio y optimización de los módulos que componen un esquema iterativo de modo independiente. Una aplicación muy sugerente de estos diagramas es la optimización de las tasas de los códigos interno y externo de un esquema LDGM concatenado. En el presente trabajo las tasas de dichos códigos componentes se ha realizado de modo empírico, probando diferentes tasas hasta conseguir el rendimiento deseado. Los diagramas EXIT constituyen una vía muy prometedora para realizar esta tarea de modo sistemático y riguroso.

Creemos importante comentar otra razón que justifica la optimización de las tasas de los códigos componente. A lo largo del presente trabajo hemos observado que el rendimiento (en cuanto a umbral y efecto suelo) de los códigos LDGM no depende tanto del canal utilizado (AWGN, Rayleigh SISO o MIMO) como de la tasa global del codificador. Una razón intuitiva es que al rendir estos códigos cercanos al límite marcado por la capacidad y los canales considerados ajustados al modelo de canal contemplado en el teorema de Shannon, la estructura pseudoaleatoria de estos códigos hace que su comportamiento sea muy parecido al de los códigos bloque aleatorios considerados por dicho teorema. Por tanto, es muy posible que un código optimizado de una determinada tasa se comporte de modo igualmente satisfactorio sobre distintos modelos de canal, lo cual, en definitiva, simplifica el proceso de búsqueda de códigos óptimos.

9.2.8. Integración de los módulos del receptor mediante representaciones homogéneas

Hemos visto cómo el formalismo de red bayesiana nos permite no sólo representar el esquema de codificación sino también la parte del transmisor dedicada a la asignación bit a símbolo, de modo que en recepción el algoritmo BP actúa para resolver de modo conjunto tanto la demodulación como la decodificación. Además, recientemente se ha introducido un formalismo derivado del concepto de red bayesiana denominado grafo de factores (*factor graph*). Esto sugiere que la vía para alcanzar esquemas más potentes que los actuales y, sobre todo, de modo más sistemático, es la integración de otros algoritmos (estimación de canal, sincronización, ...) en dicho formalismo. El mayor problema a resolver es que, mientras que en los casos aquí analizados todas las variables de la red son discretas, muchas de esas técnicas operan sobre variables aleatorias continuas.

A primera vista, existe la posibilidad de tratar las variables aleatorias continuas como nodos cuyo mensaje de salida es precisamente la estimación MAP (o MV) de dicha variable. Nótese que eso es lo que se realiza precisamente en el estimador de canal MV-EM presentado en esta tesis. Sin embargo, lo comentado con respecto a la estimación de canales variantes en el tiempo mediante HMMs parece una vía más adecuada. Se trata, en el fondo, de discretizar el espacio continuo sobre el que están definidas dichas variables.

Disponer de un formalismo de representación común tiene múltiples ventajas. Por una parte, se gana en cuanto a comprensión de qué es lo que hace, en última instancia, un determinado algoritmo de procesado y, aspecto no menos importante, qué aproximaciones realiza con

respecto al método óptimo. Estas dos razones permiten tener una idea, digamos “a primera vista”, de cómo se va a comportar el algoritmo (en cuanto a rendimiento y complejidad computacional) con respecto a otros *definidos utilizando el mismo formalismo*. Por otra parte, introduce una notación homogénea, reduciendo el tiempo de comprensión del algoritmo por parte de la comunidad científica. Por último (aunque posiblemente existan más ventajas que las aquí comentadas), facilita el paso del plano teórico a la implementación real.

9.2.9. Implementación en tiempo real

Una diferencia fundamental de los códigos LDPC y LDGM con respecto a los Turbo es que el algoritmo de decodificación es altamente paralelizable. Hoy en día, las tecnologías mayormente utilizadas para la implementación en tiempo real de algoritmos de procesado de señal y/o decodificación son los procesadores digitales de señal (*Digital Signal Processors*, DSPs) y los *arrays* de puertas programables (*Field Programmable Gate Arrays*, FPGAs). En esta línea, hemos ya realizado trabajos tanto en FPGAs [109] como en DSPs. En este último caso, los algoritmos implementados fueron precisamente los expuestos en esta tesis para el caso de sistemas MIMO sin codificación de canal [110].

Fruto de nuestra experiencia en los mencionados trabajos, podemos decir que, mientras los DSPs permiten obtener un alto rendimiento en aquellos algoritmos en los que se hace uso intensivo del producto escalar de vectores, su rendimiento se ve severamente afectado cuando la mayor parte de las operaciones son de salto y almacenamiento. Nótese que este caso es precisamente el del algoritmo *Forward/Backward* utilizado por los decodificadores Turbo. Este hecho, unido a los prometedores resultados obtenidos al implementar dicho algoritmo mediante tecnología VLSI (*Very Large Scale Integration*) analógica, sugieren que las FPGAs son una tecnología mucho más adecuada. Por otra parte, la estructura inherentemente paralela del algoritmo BP (más si cabe al saber que, en general, el esquema de planificación influye poco en el rendimiento) encaja a la perfección con el esquema FPGA, ya que pueden replicarse en *hardware* tantos nodos de la red bayesiana considerada como se desee, ejecutándose el procesamiento en cada uno de ellos de modo totalmente paralelo.

Apéndice A

Cálculo del filtro blanqueador en tiempo discreto

It is a good thing for an uneducated man to read books of quotations

—Sir Winston Churchill

Primero, es importante señalar que existen 2^p elecciones de $F(z)$ que satisfacen la factorización deseada $R_h(z) = F(z)F^*(z^{-1})$. Las distintas elecciones resultan en filtros $1/F^*(z^{-1})$ con idéntica magnitud pero distinta respuesta en fase. Una posible elección es escoger $F^*(z^{-1})$ de modo que sea de fase mínima, esto es, con todas sus raíces dentro del círculo unidad. De este modo, $1/F^*(z^{-1})$ es un sistema discreto realizable, causal y estable. El problema de esta elección es que la respuesta global resultante $F(z)$ es de fase máxima y anticausal, y la ISI resultante difícil de compensar. Para superar esta limitación, escogeremos $F^*(z^{-1})$ de fase máxima y por tanto el filtro blanqueador $1/F^*(z^{-1})$ será estable sólo si es considerado anticausal. De todos modos, un filtro anticausal siempre puede ser implementado si se introduce un retardo suficiente. La ventaja de esta aproximación es que ahora la respuesta global $F(z)$ del sistema es causal y de fase mínima.

Considerando que $R_h(t)$ toma valores significativos sólo sobre el intervalo $[-2, 2]$ (cf. Figura 4.9.(b)) tenemos que

$$R_h(k) = [r_{-2} \ r_{-1} \ r_0 \ r_1 \ r_2] = [r_2^* \ r_1^* \ r_0 \ r_1 \ r_2] = [0,0567 \ 0,5127 \ 0,9963 \ 0,5127 \ 0,0567]$$

Como se mencionó anteriormente, nótese que $h(t)$ contiene el 99,63 % de la energía total de la verdadera modulación GMSK porque $R_h(0) = 0,9963$. La transformada Z de $R_h(k)$ es

$$R_h(z) = r_{-2}z^2 + r_{-1}z + r_0 + r_1z^{-1} + r_2z^{-2}$$

que podemos expresar como

$$R_h(z) = r_2(z - 1/\rho_1^*)(z - 1/\rho_2^*)(1 - \rho_1z^{-1})(1 - \rho_2z^{-1})$$

Forzando que $|\rho_1|, |\rho_2| \leq 1$ para que el filtro blanqueador resultante exista y sea estable tenemos que $\rho_1 = -0,1522$ and $\rho_2 = -0,5746$. Teniendo en cuenta que ρ_1 and ρ_2 son números reales llegamos a que

$$F(z) = \sqrt{\frac{r_2}{\rho_1 \rho_2}} (1 - \rho_1 z^{-1})(1 - \rho_2 z^{-1})$$

y por tanto el filtro blanqueador viene dado por

$$W(z) = \frac{1}{F^*(z^{-1})} = \frac{\sqrt{\frac{\rho_1 \rho_2}{r_2}}}{(1 - \rho_1 z)(1 - \rho_2 z)}$$

cuya transformada **Z** inversa es

$$w(k) = \{w_k\}_{k=-\infty}^0 = \left\{ \frac{\sqrt{\frac{\rho_1 \rho_2}{r_2}}}{\rho_2 - \rho_1} (\rho_2^{-k+1} - \rho_1^{-k+1}) \right\}$$

Ya que $\{|w_k|\}_{k=0}^{-\infty}$ es una serie estrictamente decreciente, podemos considerar sólo los primeros coeficientes w_k significativos. Teniendo en cuenta que $|w_{-20}| < 10^{-4}$ podemos implementar $w(k)$ como un filtro FIR anticausal

$$\begin{aligned} w(k) &\approx [w_{-19} \ w_{-18} \ \cdots \ w_{-1} \ w_0] = \\ &= [0,0001 \ -0,0001 \ 0,0002 \ -0,0004 \ 0,0007 \ -0,0013 \ 0,0022 \ -0,0038 \ 0,0066 \\ &\quad -0,0115 \ 0,0201 \ -0,0349 \ 0,0608 \ -0,1058 \ 0,1839 \ 0,3189 \ 0,5473 \ -0,9025 \ 1,2417] \end{aligned}$$

Bibliografía

- [1] "Third generation partnership project website." [Online]. Available: <http://www.3gpp.org>
- [2] "Third generation partnership project-2 website." [Online]. Available: <http://www.3gpp2.org>
- [3] R. Esmailzadeh, M. Nakagawa, and A. Jones, "TDD-CDMA for the 4th Generation of wireless communication," *IEEE Wireless Commun. Mag.*, vol. 10, no. 4, pp. 8–15, Aug. 2003.
- [4] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications," *IEEE Signal Processing Mag.*, vol. 17, no. 3, pp. 29–48, May 2000.
- [5] J. Heiskala and J. Terry, *OFDM Wireless LANs: A Theoretical and Practical Guide*. SAMS Publishing., 2002.
- [6] S. Verdú, *Multiuser detection*. New York: Cambridge University Press, 1998.
- [7] W. Jakes, *Microwave Mobile Communication*. New York, NY: Wiley, 1974.
- [8] Í. E. Telatar, "Capacity of multi-antenna Gaussian channels," Bell Laboratories, Lucent Technologies, Tech. Rep., Oct. 1995, published in *European Transactions on Telecommunications*, Vol. 10, No. 6, pp. 585-595, Nov/Dec 1999.
- [9] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, jul-oct 1948.
- [10] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. ICC 93*, May 1993, pp. 1064–1070.
- [11] G. Raleigh and J. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–466, Mar. 1998.
- [12] S. L. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1347–1359, Aug. 2000.
- [13] C. Luschi, M. Sandell, P. Strauch, J.-J. Wu, C. Ilas, P.-W. Ong, R. Baeriswyl, F. Battaglia, S. Karageorgis, and R.-H. Yan, "Advanced signal-processing algorithms for energy-efficient wireless communications," *Proc. IEEE*, vol. 88, no. 10, pp. 1633–1650, Oct. 2000.
- [14] T. H. Liew and L. Hanzo, "Space-time codes and concatenated channel codes for wireless communications," *IEEE Proceedings*, vol. 90, no. 2, pp. 187–219, Feb. 2002.

- [15] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.
- [16] M. González-López, L. Castedo, and J. García-Frías, "BICM for MIMO systems using Low-Density Generator Matrix (LDGM) codes," in *Proc. ICASSP 2004 (to appear)*, Montreal, Canada, May 2004.
- [17] F. J. Vázquez-Araújo, M. González-López, L. Castedo, and J. García-Frías, "BICM for MIMO channels using LDGM codes and Sphere Detection," in *Proc. SPAWC 2004 (submitted)*, Lisbon, Portugal, July 2004.
- [18] D. Gesbert, D. shan Shiu, P. J. Smith, and A.Ñaguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.
- [19] G. J. Foschini, D. Chizhik, M. J. Gans, C. Papadias, and R. A. Valenzuela, "Analysis and performance of some basic space-time architectures," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 303–320, Apr. 2003.
- [20] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in Manhattan," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 321–331, Apr. 2003.
- [21] M. Lienard, P. Degauque, J. Baudet, and D. Degardin, "Investigation on MIMO channels in subway tunnels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 332–339, Apr. 2003.
- [22] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. W. Wolniansky, "Correlation analysis based on MIMO channel measurements in an indoor environment," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 713–720, June 2003.
- [23] ESCORT, "Enhanced diversity and Space-time Coding for underground metro and Railway Transmission," Final Report D 6021, Oct. 2002.
- [24] A. Mehrotra, *GSM System Engineering*. Artech House Publishers, 1997.
- [25] J. M. H. Rábanos, *Comunicaciones Móviles GSM*. Fundación Airtel, 1999.
- [26] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [27] R. Hamming, "Error detecting and error correcting codes," *Bell Syst. Tech. J.*, vol. 29, pp. 147–160, 1950.
- [28] P. Elias, "Coding for noisy channels," *IRE Nat. Conv. Rec.*, pp. 37–47, 1955.
- [29] J. Wozencraft, "Sequential decoding for reliable communication," *IRE Nat. Conv. Rec.*, vol. 5, no. 2, pp. 11–25, 1957.
- [30] J. Wozencraft and B. Reiffen, *Sequential Decoding*. Cambridge, MA: MIT Press, 1961.
- [31] R. Fano, "A heuristic discussion of probabilistic coding," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 64–74, Apr. 1963.

- [32] J. Massey, *Threshold decoding*. Cambridge, MA: MIT Press, 1963.
- [33] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, Apr. 1967.
- [34] G. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–277, Mar. 1973.
- [35] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimum decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.
- [36] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Stat.*, vol. 37, pp. 1554–1563, 1966.
- [37] C. Heegard and S. B. Wicker, *Turbo Coding*. Kluwer Academic Publishers, 1999.
- [38] R. G. Gallager, "Low-Density Parity Check codes," *IRE Trans. Inform. Theory*, vol. 39, pp. 37–45, Jan. 1962.
- [39] ———, *Low-Density Parity Check codes*. Cambridge, MA: MIT Press, 1963.
- [40] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.
- [41] D. J. C. MacKay and R. M. Neal, "Good codes based on very sparse matrices," in *Cryptography and Coding 5th IMA Conf., Lectures Notes in Computer Science*, no. 1025, Berlin, Germany: Springer, May 1995, pp. 100–111.
- [42] ———, "Near Shannon limit performance of Low-Density Parity Check codes," *Electron. Lett.*, vol. 32, no. 18, pp. 1645–1646, Aug. 1996.
- [43] D. J. C. MacKay, "Correction: Errata for "Good error-correcting codes based on very sparse matrices"," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, p. 2101, July 2001.
- [44] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- [45] T. J. Richardson, M. A. Shkrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular Low-Density Parity Check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
- [46] G. Battail, "A conceptual framework for understanding turbo codes," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 245–254, Feb. 1998.
- [47] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
- [48] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*. Kluwer Academic / Plenum Publishers, 1999.
- [49] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

- [50] J.-F. Cheng and R. J. McEliece, "Some high-rate near capacity codecs for the Gaussian channel," in *Proc. 34th Allerton Conference on Communications, Control and Computing*, Oct. 1996.
- [51] T. J. Richardson and R. L. Urbanke, "Efficient encoding of Low-Density Parity Check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 638–656, Feb. 2001.
- [52] D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for 'turbo-like' codes," in *Proc. 36th Ann. Allerton Conf. on Commun., Control and Computing*, Sept. 1998, pp. 201–210.
- [53] J. Garcia-Frias and W. Zhong, "Approaching Shannon performance by iterative decoding of linear codes with Low-Density Generator Matrix," *IEEE Commun. Lett.*, vol. 7, no. 6, pp. 266–268, June 2003.
- [54] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 56–67, Jan. 1982.
- [55] E. Biglieri, D. Divsalar, P. McLane, and M. Simon, *Introduction to Trellis-Coded Modulation with Applications*. MacMillan Publishing Co., 1991.
- [56] S. Verdú, "Fifty years of Shannon theory," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2057–2078, oct 1998.
- [57] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 360–393, 1963.
- [58] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–884, May 1992.
- [59] G. Caire, G. Taricco, and E. Biglieri, "Bit-Interleaved Coded Modulation," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 927–946, may 1998.
- [60] X. Li and J. A. Ritcey, "Bit-Interleaved Coded Modulation with Iterative Decoding," *IEEE Commun. Lett.*, vol. 50, no. 8, pp. 1250–1257, Aug. 2002.
- [61] —, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE J. Select. Areas Commun.*, vol. 17, no. 4, pp. 715–724, Apr. 1999.
- [62] —, "Bit-Interleaved Coded Modulation with Iterative Decoding," in *Proc. ICC'99*, June 1999, pp. 858–863.
- [63] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 944–957, May 2001.
- [64] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8PSK signaling," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1250–1257, Aug. 2002.
- [65] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

- [66] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
- [67] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Select. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.
- [68] ———, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.
- [69] E. Lindskog and A. Paulraj, "A transmit diversity scheme for channels with intersymbol interference," in *Proc. ICC*, vol. 1, New Orleans, LA, 2000, pp. 307–311.
- [70] C. Douillard, A. Picart, P. Didier, M. Jézèchel, C. Berrou, and A. Glavieux, "Iterative correction of Intersymbol Interference: Turbo-Equalization," *European Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, Sept. 1995.
- [71] X. Lin and R. S. Blum, "Improved space-time codes using serial concatenation," *IEEE Commun. Lett.*, vol. 4, no. 7, pp. 221–223, July 2000.
- [72] H.-J. Su and E. Geraniotis, "Space-time turbo codes with full antenna diversity," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 47–57, Jan. 2001.
- [73] Y. Liu, M. P. Fitz, and O. Y. Takeshita, "Full rate space-time turbo codes," *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 969–980, May 2001.
- [74] A. Stefanov and T. M. Duman, "Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches, and practical considerations," *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 958–968, May 2001.
- [75] A. M. Tonello, "Space-time Bit-Interleaved Coded Modulation with an Iterative Decoding strategy," in *Proc. IEEE VTC Fall 2000*, Boston, Sept. 2000.
- [76] ———, "Space-Time Bit-Interleaved Coded Modulation over frequency selective fading channels with Iterative Decoding," in *Proc. IEEE GLOBECOM'00*, vol. 3, San Francisco, CA, Dec. 2000, pp. 1616–1620.
- [77] ———, "MIMO MAP equalization and turbo decoding in interleaved space-time coded systems," *IEEE Trans. Commun.*, vol. 51, no. 2, pp. 155–160, Feb. 2003.
- [78] G. J. McLachlan and T. Krishnan, *The EM algorithm and extensions*. John Wiley and Sons, 2000.
- [79] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1639–1642, July 1999.
- [80] O. Damen, A. Chkeif, and J. Belfiore, "Lattice code decoder for space-time codes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 161–163, May 2000.
- [81] J. M. H. Rábanos, *Transmisión por radio. Segunda edición*. Editorial Centro de Estudios Ramón Areces, 1995.

- [82] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.
- [83] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks*. Prentice Hall, 2002.
- [84] A. Paulraj, R. N̄abar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.
- [85] D. N̄. C. Tse, "Fundamentals of wireless communications," Notes of the Fall EE290S Course, University of California at Berkeley. Dept. of Electrical Engineering and Computer Sciences, 2002. [Online]. Available: <http://inst.eecs.berkeley.edu/ee290s/fa02/>
- [86] H. V. Poor and G. W. Wornell, *Wireless Communications*. Prentice Hall Signal Processing Series, 1998.
- [87] J. G. Proakis, *Digital Communications*, 3rd ed. McGraw-Hill, Inc., 1995.
- [88] A. Goldsmith, "Wireless communications," Notes of the Fall 2003 EE359 Course, Stanford University, 2003. [Online]. Available: <http://wsl.stanford.edu/andrea/EE359Reader/Reader03.pdf>
- [89] J. Ventura-Travest, G. Caire, E. Biglieri, and G. Taricco, "Impact of diversity reception on fading channels with coded modulation. Part I: coherent detection," *IEEE Trans. Commun.*, vol. 45, no. 5, pp. 676–686, may 1997.
- [90] C. Fragouli and R. D. Wesel, "Turbo-encoder design for symbol-interleaved parallel concatenated Trellis-Coded Modulation," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 425–435, Mar. 2001.
- [91] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Trans. Commun.*, vol. 48, Mar. 2000.
- [92] A. R. Hammons and H. El Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Trans. Commun.*, vol. 46, no. 2, pp. 524–542, Mar. 2000.
- [93] J. Baudet, M. González-López, D. Degardin, A. Dapena, M. Lienard, L. Castedo, and P. Degauque, "Performance of space time coding in subway tunnel environments," in *Proc. IEE Technical Seminar on MIMO Communications Systems: from Concept to Implementation*, London, UK, Dec. 2001, pp. 2/1–2/6.
- [94] M. Lienard and P. Degauque, "Propagation in wide tunnels at 2 GHz: a statistical analysis," *IEEE Trans. Veh. Technol.*, vol. 47, no. 4, pp. 1322–1328, 1998.
- [95] J. D. Laster, "Robust GMSK demodulation using demodulation diversity and BER estimation," Ph.D. dissertation, Faculty of the Virginia Polytechnic Institute and State University, 1997. [Online]. Available: <http://scholar.lib.vt.edu/theses/public/etd-5941513972900/etd-title.html>
- [96] A. N̄. D'Andrea and U. Mengali, *Synchronization Techniques for Digital Receivers*. Plenum Press, 1997.
- [97] N. Al-Dhahir and G. Saulnier, "A high-performance reduced-complexity GMSK demodulator," GE Corporate Research and Development, Tech. Rep. 96CRD107, 1996. [Online]. Available: http://www.crd.ge.com/crd_reports/pdf/1996crd107.pdf

- [98] —, “A high-performance reduced-complexity GMSK demodulator,” *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1409–1412, Nov. 1998.
- [99] J. Kurzweil, *An Introduction to Digital Communications*. John Wiley and Sons, 2000.
- [100] X. Zang and M. P. Fitz, “Space-time code design with Continuous Phase Modulation,” *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 783–792, June 2003.
- [101] J. Hagenauer and P. Hoeher, “A Viterbi algorithm with soft-decision outputs and its applications,” in *Proc. GLOBECOM’89*, Nov. 1989, pp. 1680–1686.
- [102] V. Franz, *Turbo-detection for GSM-systems - Channel estimation, equalization and decoding*. PhD Thesis, Univ. of Munich, Mar. 2000.
- [103] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. Prentice-Hall, 1992.
- [104] M. F. Bugallo, J. Míguez, and L. Castedo, “Decision-Feedback semiblind channel equalization in space-time coded systems,” in *Proc. ICASSP 2002*, Orlando, FL, May 2002.
- [105] J. A. Fessler and A. O. Hero, “Space-Alternating Generalized Expectation-Maximization algorithm,” *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.
- [106] P. Roberston and T. Wörz, “Bandwidth-efficient Turbo Trellis-Coded Modulation using punctured component codes,” *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 206–218, Feb. 1998.
- [107] R. Narayanaswami, “Coded modulation with Low Density Parity Check codes,” Master’s thesis, Texas A&M University, College Station, TX, June 2001. [Online]. Available: <http://ee.tamu.edu/krishna/PAPERS/ravithesis.ps>
- [108] S. ten Brink, “Convergence of iterative decoding,” *IEEE Elec. Lett.*, vol. 35, no. 10, pp. 806–808, May 1999.
- [109] M. G. López, “Diseño de dispositivos PCI con lógica programable FPGA,” Master’s thesis, Universidade da Coruña, A Coruña, Spain, July 2000.
- [110] M. G. López, A. D. Janeiro, and C. E. Cascón, “Receptores para sistemas GSM con diversidad espacial: Diseño, simulación e implementación,” in *Proc. XVII Simposium Nacional de la URSI*, Alcalá de Henares, Madrid, Spain, Sept. 2002.

