



UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

**Cuantificación del estado del canal para la
minimización de la potencia en sistemas con
transmisores adaptativos**

Autor:

Antonio García Marqués

Directores:

Prof. Dr. D. Fco. Javier Ramos López

Prof. Dr. D. Georgios B. Giannakis

DEPARTAMENTO DE TEORÍA DE LA SEÑAL Y COMUNICACIONES

Leganés, enero de 2007

TESIS DOCTORAL

Cuantificación del estado del canal para la minimización de la potencia en sistemas con transmisores adaptativos

Autor: Antonio García Marqués

Directores: Prof. Dr. D. Fco. Javier Ramos López
Prof. Dr. D. Georgios B. Giannakis

Firma del Tribunal Calificador:

Firma

Presidente:

Vocal:

Vocal:

Vocal:

Secretario:

Calificación:

Leganés, de de

Resumen

Los sistemas de comunicación operando en canales (móviles) inalámbricos degradan considerablemente sus prestaciones (tasa de transmisión, consumo de potencia o probabilidad de error) debido a los desvanecimientos que sufre el medio inalámbrico. Esta degradación puede mitigarse aprovechando el conocimiento del estado canal (CSI) en el receptor y en el transmisor. El CSI puede ser perfecto si el error de estimación del canal es despreciable o cuantificado si no se conoce el valor exacto del canal sino únicamente la categoría (región) a la que el canal pertenece.

La presente disertación tiene como objetivo diseñar y utilizar de manera óptima la cuantificación del CSI a fin de minimizar la potencia transmitida satisfaciendo requisitos de calidad de servicio en términos tasa de transmisión y probabilidad de error. El escenario básico de aplicación consiste en sistemas de comunicaciones móviles con transmisores adaptativos en los que el receptor tiene conocimiento perfecto del canal mientras que el transmisor, típicamente a través de un canal de control de retorno, únicamente conoce una versión cuantificada del mismo. Las contribuciones de esta tesis cubren sistemas de comunicaciones de distinta complejidad, operando sobre distintos tipos de canales, para uno o múltiples usuarios. Simulaciones numéricas validarán las derivaciones analíticas y mostrarán el significativo ahorro energético cuando se implementan los esquemas propuestos.

Las principales contribuciones de esta tesis pueden resumirse como sigue: (i) se derivan esquemas de cuantificación y de adaptación como la solución de un problema de minimización con ligaduras; (ii) se prueba que los esquemas de cuantificación de canal corresponden al menos a un óptimo local; (iii) se prueba que los esquemas de adaptación son globalmente óptimos; (iv) se demuestra que la complejidad computacional por cada realización del canal requerida para implementar los esquemas adaptativos es moderada o despreciable; (v) se demuestra que las necesidades de envío de información de control del receptor al transmisor son sorprendentemente pequeñas y (vi) se muestra que la potencia consumida por los esquemas adaptativos basados en CSI cuantificado es considerablemente menor que la potencia consumida por sistemas no adaptativos y a su vez que la pérdida de prestaciones de aquellos frente a los que hacen uso de CSI perfecto es pequeña. Estas contribuciones motivan la implantación práctica de esquemas similares a los propuestos en sistemas reales de comunicaciones que, con un incremento moderado de la complejidad, obtienen significativas mejoras en sus prestaciones.

Abstract

Performance of wireless communication systems (quantified in terms of transmission rate, power efficiency or error probability) degrades severely due to fading effects introduced by the shared air-interface. This degradation can be mitigated if channel state information (CSI) is available at the receiver and/or the transmitter side. Full CSI can be assumed when sufficient training allows for essentially perfect channel estimation, while quantized CSI can be used when only the class (cluster or region) that the channel belongs to is known.

This dissertation aims at optimally designing and exploiting quantized CSI to minimize transmit-power under rate and error probability constraints. The setup comprises a wireless communication system with channel-adaptive modulation; where the receiver has available perfect CSI while the transmitter has only quantized CSI. In this context, innovative claims from a high-level vantage point pertain to novel schemes with variable complexity, under different channel types shared by single or multiple users. Simulated numerical tests corroborate the analytical claims and reveal that significant power savings result when implementing the proposed schemes.

Specific contributions can be summarized as follows: (i) the quantization and adaptive transmission parameters can be obtained as the solution of a judiciously formulated constrained minimization problem; (ii) the channel quantization algorithms developed attain at least a local optimum of the objective function; (iii) the adaptive schemes reach the global optimum; (iv) the required computational complexity per channel realization for the adaptive schemes ranks from moderate to negligible; (v) the amount of feedback transmitters need from the receiver is stunningly small; and (vi) power consumed by adaptive schemes based on quantized CSI is significantly less than that required by non-adaptive alternatives, and closely approaches the power savings achieved by the benchmark perfect transmit-CSI design.

These features encourage practical deployment of the proposed schemes that, with moderate computational overhead, offer a significant performance improvement.

*A los que no estabais y ahora me acompañáis,
porque demostráis cuánto se puede aprender de la buena gente que hay en el mundo.*

*A los que estabais y os habéis ido,
porque no sabéis hasta qué punto me habéis marcado.*

*A los que estabais y seguís estando,
porque uno es causa de uno mismo pero consecuencia de lo que le rodea.*

*A los que estabais y os habéis adueñado de un papel protagonista,
porque sin vosotros no puedo explicar la más reciente fase de mi vida.*

*A todos vosotros,
porque a todos os debo tanto.*

Agradecimientos*

Hay tres personas a las que les debo de forma directa la realización de esta tesis doctoral, personas esenciales con las que me encuentro en deuda por todo lo que ella ha supuesto -que mucho ha sido- y que, por consiguiente, a las que debo agradecer su labor de forma explícita: Javier Ramos, Georgios Giannakis y Alfonso Cano. A Javier, además de su papel como co-director de esta tesis, quiero agradecerle la confianza que en mí ha depositado, la libertad que siempre me ha concedido, su sincera preocupación por mi persona y mi trabajo, así como su disposición y su tiempo siempre que se los he requerido. A mi otro co-director, Yorgos, le tengo que agradecer a nivel personal su amabilidad, su cariño y su respeto, presentes de forma permanente desde el primer día que nos conocimos; a nivel académico le reconozco el regalo que supuso ponerme en contacto con un tema de investigación tan estimulante, la inmersión en su grupo de trabajo sin ningún tipo de limitación, su permanente ofrecimiento de ayuda y su continua disposición para explicarme aquello que no entendía (tan temeroso como estaba yo al principio por requerirlas), su ejemplo de cómo enseñanza e investigación pueden ser forma de vida. Y a mi amigo Alfonso, que pese a que no se lo menciona en la portada de este trabajo, ha tenido una influencia sobre el mismo comparable a la de los dos anteriores, le debo dar las gracias por haber compartido esta experiencia conmigo, por haber iniciado el camino y haberlo dejado tan despejado de obstáculos, por haberme ilustrado incluso cuando la ilustración se manifestaba difícil, por haber permitido, en definitiva, que me apoyase de forma tan intensa y tan frecuente en su persona. Sabed pues los tres que, tanto yo como lo bueno que pueda haber en este trabajo, estamos en deuda con vosotros, pero ante todo, sabed que os lo agradecemos.

Y si las líneas anteriores hacen referencia al periodo donde se realizó la labor investigadora de esta tesis, las actuales tienen la intención de expresar mi gratitud a

*Distintas fases del trabajo de investigación que aquí se presenta han sido desarrolladas en colaboración directa con los doctores G. B. Giannakis, F. J. Ramos, X. Wang y F. Digham, y han contado con el apoyo económico de la Universidad Rey Juan Carlos, del Departamento de *Electrical and Computer Engineering* de la Universidad de Minnesota, de la *National Science Foundation* (EE.UU.), de la Comunidad de Madrid y del Ministerio de Educación y Ciencia.

las personas que durante el periodo de redacción de esta tesis me prestaron ayuda en mayor grado del que yo merecía. Queridos Alfonso, Alicia, David y Felipe, gracias por vuestro tiempo.

Los frutos de este trabajo trascienden con mucho el resultado investigador que aquí se presenta, y por lo tanto, mis agradecimientos deben extenderse a un grupo más amplio de personas:

A mis compañeros, repartidos por tantas universidades a ambos lados del Atlántico. Gracias a los profesores y compañeros que me iniciaron en este proceso de en la Universidad Carlos III y que consiguieron hacerlo de una manera amena a la par que fructífera, gracias a mis compañeros (sabéis que sois algo más) de la Universidad Rey Juan Carlos porque vosotros habéis permitido que esto se llevase a cabo y gracias a mis “cercaños” compañeros que he conocido en la Universidad de Minnesota, gracias por acogerme con los brazos abiertos, por tanto cómo me habéis ayudado, por tanto cómo me habéis enseñado.

A mi familia, comenzando por mis padres, pasando por aquellos a los que veo de forma tan frecuente y llegando a los que la distancia obliga a encontrarnos de manera mas esporádica e intensa, debo agradecerle sus constantes muestras de confianza y de apoyo, presentes incluso en los momentos en que alguno de ellos creía que no estaba recorriendo un camino adecuado. Me reconforta poder deciros que el trabajo de estos años ha servido para darme cuenta de cuánto os quiero y de hasta qué punto os necesito.

A mis amigos, a los viejos, a los nuevos, y a los que os habéis revelado (haya o no haya habido después rebelión). Si hubiera que escoger el campo en el que la realización de esta tesis más haya trascendido la faceta investigadora y haya afectado de forma más importante mi vida personal habría que escoger el de la amistad. Por todo lo que me habéis motivado, por tanto cómo he aprendido, por haber conseguido que todo este proceso fuera mucho más rico y sugestivo, por evidenciar cuánto valéis y hasta qué punto los que no creen en vosotros están equivocados, por vuestras debilidades humanas, por hacerme sentir orgulloso de ser vuestro amigo, por todo ello os estoy profundamente agradecido.

Prefacio

Loada sea la ignorancia, loado y necesario el conocimiento.

Tescraso (s. VI-V a.C.)

Loado sea el conocimiento, loada y necesaria la ignorancia.

Setarcos (s. IV-V d.C.)

Escribía yo, hace ahora cuatro años, en el prefacio de mi proyecto fin de carrera que aquel texto representaba el final formal de una aventura, pero también de una experiencia vital, la educativa, que en ese momento concluía en su vertiente oficial. La tozuda realidad se empeña siempre en demostrarnos que los seres humanos, al menos el grupo unipersonal al que yo pertenezco, tiene la capacidad de equivocarse todas las veces en las que es posible hacerlo, y, naturalmente, el prefacio de ese proyecto fue un lugar más en el que la realidad -vanidosa, además de tozuda- quedó satisfecha al demostrar cuánto nos equivocamos.

Pero si pertenezco al grupo de seres humanos que se equivoca sin cesar, creo pertenecer también al grupo de los que algo aprende, y, por consiguiente, no cometeré en esta ocasión el mismo error, no se mantendrá aquí que la realización de esta tesis y el doctorado que la ha precedido supone el principio o el final de nada, tan *sólo* se dirá que han contribuido a un cambio. Creo que la reflexión, el análisis crítico, la distancia son cualidades del científico, como también lo son el coraje y la perseverancia, desconozco hasta qué grado los que pasamos este umbral las poseemos, intuyo que en uno menor que el que deberíamos, pero tengo la certeza de que durante estos años, el contacto con la ciencia y con la investigación nos ayuda a incrementarlo. Así pues, por el trabajo que ha llevado asociado, por la ingente cantidad de nuevos conceptos que he aprendido, por cuánta de mi ignorancia ha puesto de manifiesto, por incrementar mi tolerancia y mi entendimiento, por hasta qué punto mi percepción de tantos conceptos ha cambiado, esta tesis será algo de lo que, cuando mire hacia atrás y juzgue de forma serena estos años, tengo por seguro que me sentiré satisfecho.

Finalizo pidiendo complicidad del lector, y si comencé este prefacio reconociendo una equivocación cometida hace cuatro años, le ruego que me permita concluirlo

con un acierto originado en el mismo tiempo, con la vigencia y reafirmación de una verdad también plasmada en aquellas lejanas líneas: la convicción de que *la necesidad de aprendizaje no acaba nunca*, de que *expira sólo cuando expira nuestra vida*.

Índice general

Resumen	I
Abstract	III
Agradecimientos	v
Prefacio	IX
Índice de figuras	XV
Índice de tablas	XVII
Acrónimos	XIX
Notación	XXIII
1. Introducción	1
1.1. Motivación, objetivos y estado del arte	2
1.2. Estructura y publicaciones científicas asociadas a la tesis	6
2. El canal inalámbrico con desvanecimientos y el uso de la información del estado del canal	13
2.1. El canal inalámbrico	14
2.1.1. Dispersión temporal	15
2.1.2. El efecto Doppler y la dispersión en frecuencia	16

2.1.3. Canales selectivos en tiempo y frecuencia y modelo discreto equivalente	17
2.2. Parámetros básicos y minimización de la potencia en los sistemas de comunicaciones	19
2.2.1. Parámetros básicos	19
2.2.2. Minimización de la potencia	22
2.3. Sistemas adaptativos: información del estado del canal	24
2.3.1. Tipos de CSIT	27
3. Minimización de potencia en sistemas de comunicaciones adapta- tivos con una única fuente de información	31
3.1. Minimización de potencia en WSN con QoS utilizando CSIT	32
3.1.1. Modelo y preámbulos	33
3.1.2. Adaptación óptima basada en F-CSIT	36
3.1.3. Adaptación óptima basada en Q-CSIT	43
3.1.4. Adaptación óptima basada en I-CSIT	51
3.1.5. Simulaciones numéricas	54
3.1.6. Síntesis de resultados	60
3.2. Minimización de potencia en OFDM con QoS instantánea utilizando CSIT	61
3.2.1. Modelo y preámbulos	64
3.2.2. Adaptación óptima basada en P-CSIT	68
3.2.3. Adaptación y cuantificación óptimas basadas en Q-CSIT	71
3.2.4. Reducción de la complejidad de la optimización basada en Q-CSIT	79
3.2.5. Reducción de la tasa de realimentación para Q-CSIT	87
3.2.6. Adaptación óptima basada en S-CSIT	91
3.2.7. Simulaciones numéricas	93
3.2.8. Síntesis de resultados	99
3.3. Consideraciones finales	100
3.4. Apéndices	101

3.4.1.	Expresión analítica de la BER para el caso de Q-CSIT en WSN	101
3.4.2.	Expresión analítica de la derivada de la potencia para el caso de Q-CSIT en WSN	102
3.4.3.	Demostración de la proposición 3.1	103
4.	Minimización de potencia en sistemas de comunicaciones adaptativos con múltiples fuentes de información	105
4.1.	Minimización de potencia en sistemas TDMA con QoS utilizando CSIT	106
4.1.1.	Modelo y preámbulos	108
4.1.2.	Adaptación óptima de recursos basada en P-CSIT	112
4.1.3.	Adaptación de recursos y cuantificación óptimas basadas en Q-CSIT	118
4.1.4.	Simulaciones numéricas	139
4.1.5.	Síntesis de resultados	147
4.2.	Minimización de potencia en sistemas OFDMA con QoS utilizando CSIT	149
4.2.1.	Modelo y preámbulos	151
4.2.2.	Diseño de los modos de transmisión	156
4.2.3.	Diseño del cuantificador	170
4.2.4.	Simulaciones numéricas	178
4.2.5.	Síntesis de resultados	183
4.3.	Consideraciones finales	184
4.4.	Apéndice	186
4.4.1.	Demostración de las ecuaciones (4.21) y (4.22)	186
5.	Conclusiones y líneas futuras de investigación	189
5.1.	Conclusiones	189
5.2.	Líneas futuras de investigación	200
	Bibliografía y referencias	204

Índice de figuras

2.1. Canal AWGN (sin desvanecimientos) frente a canales con desvanecimientos.	15
2.2. Relación entre los canales directo e inverso para sistemas FDD y TDD.	26
2.3. Adquisición del CSIT a través de un canal de realimentación.	27
2.4. Generación del Q-CSI en base a P-CSI para el caso de $B = 4$ bits de realimentación que describen $N = 2^B = 16$ regiones de cuantificación.	28
3.1. Modelo del sistema WSN.	34
3.2. Potencia media transmitida (\bar{p}) frente a tasa media transmitida para diferentes tipos de CSIT ($M = 4, L = 4, N_u = 16$).	55
3.3. Variación del número de bits realimentados para la adaptación basada en Q-CSIT ($M = 4, \tilde{r} = 2,5, L = 4$).	60
3.4. Diagrama de bloques del sistema OFDM.	64
3.5. Interdependencia de los módulos de computación.	88
3.6. Prestaciones de esquemas bajo distintas consideraciones de CSIT y requisitos de BER ($K = 64, \tilde{r} = 25, B = 2$).	96
3.7. Prestaciones de esquemas basados en P-CSIT y Q-CSIT para distintos requisitos de BER ($K = 64, \tilde{r} = 25, B = 2$).	96
4.1. Modelo del sistema TDMA considerado para $M = 3$ usuarios.	109
4.2. Consumo total de potencia para distintos esquemas de asignación de recursos variando la prioridad entre usuarios μ_2/μ_1 (con $\sum_{m=1}^2 \mu_m = 1$) cuando $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1$ <i>bpsb</i> , y $\bar{g}_1 = \bar{g}_2 = 0$ <i>dB</i>	141

4.3. Regiones de potencia para los esquemas basados en P-CSIT y en Q-CSIT (ARCC) ($\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\bar{g}_1 = \bar{g}_2 = 0$ dB).	143
4.4. Asignación óptima de acceso y regiones de cuantificación obtenidas con el algoritmo ARCC. La selección del usuario que accede al canal se indica en la leyenda, las distintas regiones se representan con distintos sombreados y los umbrales de cuantificación señalan con líneas en negrita ($\mu_1 = 2/3$, $\mu_2 = 1/3$, $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1$ bpsb, $\bar{g}_1 = \bar{g}_2 = 0$ dB).	145
4.5. Evolución de la potencia ponderada media del algoritmo ARCC($\mu_1 = 2/3$, $\mu_2 = 1/3$, $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1$ bpsb, $\bar{g}_1 = \bar{g}_2 = 0$ dB).	148
4.6. Modelo del sistema OFDMA.	152
4.7. Cuantificación de las sub-portadoras basada en regiones equiprobables ($K = 64$, $M = 6$, $L = 6$).	177
4.8. Umbrales de cuantificación para los usuarios $m = 1$ y $m = 3$ con ganancias de canal semejantes y distintos requisitos de tasa ($K = 64$, $M = 3$, $\bar{\mathbf{r}}_0 = [20, 40, 60]^T$).	177
4.9. Ganancias de sub-portadora para distintos usuarios ($K = 64$, $M = 3$, $\overline{\text{SNR}} = 0$ dB).	179
4.10. Comparación de la potencia media transmitida (total) para diferentes esquemas de adaptación ($K = 64$, $M = 3$, $\check{\mathbf{r}} = [60, 60, 60]^T$).	180
4.11. Comparación de la potencia media transmitida (por cada usuario) para diferentes esquemas de adaptación ($K = 64$, $M = 3$, $\check{\mathbf{r}} = [20, 40, 60]^T$).	181
4.12. Diferencia de la potencia media transmitida (total) por los diferentes esquemas c.r.a. la adaptación basada en P-CSIT ($K = 64$, $M = 3$, $\check{\mathbf{r}} = [20, 40, 60]^T$).	182
4.13. Variación del número de regiones de cuantificación (bits realimentados) por sub-portadora ($K = 64$, $M = 3$, $\check{\mathbf{r}} = [20, 40, 60]^T$).	183

Índice de tablas

3.1. Potencia media transmitida (dB_W) para sistemas de lazo abierto y sistemas de lazo cerrado según \tilde{r} varía.	56
3.2. Potencia media transmitida (dB_W) para esquemas basados en (F, I, Q, y S)-CSIT. (Caso de referencia: $M = 4$, $\tilde{r} = 2.5$, $\tilde{\epsilon} = 10^{-3}$, $L = 4$, $r_l = [0, 1, 3, 5]$, $N_u = 16$, $E_s/N_0 = 1$; para otros CASOS, se modifica un único parámetro c.r.a. caso de referencia.)	57
3.3. Potencia media (dB_W) y tasa de transmisión por modo AMC (región) y umbrales de cuantificación ($M = 4$, $\tilde{r} = 2.5$, $\tilde{\epsilon} = 10^{-3}$, $L = 4$, $r_l = [0, 1, 3, 5]$, $N_u = 16$).	57
3.4. Potencia media transmitida (dB_W) variando L	59
3.5. Número de iteraciones necesarias para los distintos diseños (métodos) propuestos.	86
3.6. Prestaciones de esquemas basados en SIT-QCSIT y Q-CSIT ($K = 64$, $BER_0 = 10^{-3}$, $\tilde{r} = 20$, $B = 2$).	93
3.7. Prestaciones de esquemas basados en P-CSIT y Q-CSIT ($K = 64$, $BER_0 = 10^{-3}$, $\tilde{r} = 20$, $B = 2$).	94
3.8. Prestaciones de esquemas basados en P-CSIT y Q-CSIT para sub-portadoras con ganancia media uniforme ($K = 64$, $BER_0 = 10^{-3}$, $\tilde{r} = 20$, $B = 2$, $\bar{g}_k = 2$).	95
3.9. Prestaciones de distintos métodos de estimación con interpolación basada en funciones “sinc” para la reducción de la tasa de realimentación ($K = 64$, $\tilde{r} = 20$, $B = 2$, $L = 3$, $\mathbb{E}[h_0 ^2] = 1$, $\mathbb{E}[h_1 ^2] = 0.5$, $\mathbb{E}[h_2 ^2] = 0.25$, $\mathbb{E}[h_3 ^2] = 0.1$).	97

3.10. Variación del número de bits de realimentación utilizando un esquema basado en EP-QCSIT ($K = 64$, $\check{r} = 25$).	98
4.1. Prestaciones del algoritmo ARCC en distintos casos de prueba (en los Casos II y IV se fijó una tasa objetivo de 0.5 bpsb para el usuario 2).	144
4.2. Asignación de potencia y tasa transmitidas ($p_{m,l}$ y $r_{m,l}$) por cada estado de cuantificación resultantes de la aplicación del algoritmo ARCC ($\mu_1 = 2/3$, $\mu_2 = 1/3$, $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1 \text{ bpsb}$, $\bar{g}_1 = \bar{g}_2 = 0 \text{ dB}$).	146
4.3. Potencia ponderada media consumida con ARCC variando el número de bits de realimentación ($\mu_1 = \mu_2 = 1$, $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1 \text{ bpsb}$, $\bar{g}_1 = \bar{g}_2 = 0 \text{ dB}$).	146
4.4. Potencia total transmitida (dB_W) para los esquemas P-CSIT, Q-CSIT2 y Q-CSIT5. (Caso de referencia: $K = 64$, $M = 3$, $\check{\mathbf{r}} = [60, 60, 60]^T$, $\overline{\text{SNR}} = 0 \text{ dB}$; otros <i>casos</i> implica una única variación c.r.a. caso de referencia.)	182

Acrónimos y abreviaturas**

A/D	Analógico/Digital
AMC	<i>Adaptive Modulation and Coding</i>
ARCC	Adaptación de Recursos y Cuantificación Conjunta
AWGN	<i>Average White Gaussian Noise</i>
BER	<i>Bit Error Rate</i>
bpcu	<i>bits per channel use</i>
CDMA	<i>Code Division Multiple Access</i>
cf.	confírmese con
CFI	Centro de fusión de información
CFI	Cuantificación independiente por sub-portadora
CP	<i>Cyclic Prefix</i>
c.r.a.	“con respecto a” o “con respecto al”
CSI	<i>Channel State Information</i>
CSIR	<i>Channel State Information at Receiver</i>
CSIT	<i>Channel State Information at Transmitter</i>
D/A	Digital/Analógico
DFT	<i>Discrete Fourier Transform</i>
DRC	Data Rate Control
E-CSIT	CSIT espacial
E/S	Entrada/Salida
e.g.	<i>exempli gratia</i>
EP	Equi-probable
F-CSIT	<i>Full CSIT</i>
FDD	<i>Frequency Division Duplexing</i>
FDP	Función de Densidad de Probabilidad
FDPA	FDP Acumulada

** A lo largo de esa tesis se mantendrán en su forma original aquellos acrónimos derivados de una expresión inglesa cuyo uso se encuentre extendido en la literatura científica.

HSUPA	<i>High Speed Uplink Packet Access</i>
I-CSIT	<i>Individual CSIT</i>
i.e.	<i>id est</i>
IDFT	<i>Inverse Discrete Fourier Transform</i>
IEEE	<i>Institute of Electrical and Electronic Engineers</i>
i.i.d.	independiente e idénticamente distribuido
ISI	<i>Inter-Symbol Interference</i>
KKT	Karush–Kuhn–Tucker
LMMSE	<i>Linear Minimum Mean Square Error</i>
M-QAM	<i>M-ary Quadrature Amplitude Modulation</i>
MIMO	<i>Multiple-Input Multiple-Output</i>
MIP	<i>Mixed-Integer Programming</i>
MISO	<i>Multiple-Input Single-Output</i>
MMSE	<i>Minimum Mean Square Error</i>
MRC	<i>Maximum Ratio Combining</i>
OFDM	<i>Orthogonal Frequency Division Multiplexing</i>
OFDMA	<i>Orthogonal Frequency Division Multiple Access</i>
OSI	<i>Open Systems Interconnection</i>
P-CSIT	<i>Perfect CSIT</i>
P/S	Paralelo/Serie
PA	Punto de Acceso
PAPR	<i>Peak to Average Power Ratio</i>
PSK	<i>Phase Shift Keying</i>
Q-CSIT	<i>Quantized CSIT</i>
QAM	<i>Quadrature Amplitude Modulation</i>
QoS	<i>Quality of Service</i>
RC	Reducción de la complejidad

S-CSIT	<i>Statistical CSIT</i>
S/P	Serie/Paralelo
SISO	<i>Single-Input Single-Output</i>
spcu	<i>symbols per channel use</i>
s.p.d.g.	sin pérdida de generalidad
s.p.d.o.	sin pérdida de optimalidad
s.y.s.s.	si y sólo si
t.c.c.	también conocido como
TDD	<i>Time Division Duplexing</i>
TDM	<i>Time-Division Multiplexing</i>
TDMA	<i>Time-Division Multiple Access</i>
TF	Transformada de Fourier
UC	Umbrals comunes
UMTS	<i>Universal Mobile Telecommunications System</i>
VA	Variable Aleatoria
WSN	<i>Wireless Sensors Network</i>

Notación

Letras en cursiva denotan valores escalares, letras minúsculas en negrita denotan vectores columna, letras mayúsculas en negrita denotan matrices, letras con tipo caligráfico denotan conjuntos,

$(\cdot)^\dagger$	es el conjugado de un número, vector o matriz,
$(\cdot)^T$	es la transpuesta de una matriz o vector,
$(\cdot)^{\mathcal{H}}$	es la transpuesta conjugada (<i>hermítica</i>) de una matriz o vector,
$[\cdot]_{k,l}$	es el (k, l) -ésimo elemento de una matriz,
$[\cdot]_k$	es el k -ésimo elemento de un vector,
$\lfloor x \rfloor$	es el redondeo al entero más cercano a x ,
$\lfloor x \rfloor, (\lceil x \rceil)$	es el redondeo al entero más cercano menor (mayor) que x ,
$ x $	es el valor absoluto de un escalar x ,
$\ \mathbf{x}\ $	es la norma euclídea,
$d_{ch}(\mathbf{x}_1, \mathbf{x}_2)$	es la distancia cordal entre dos vectores unitarios $\mathbf{x}_1, \mathbf{x}_2$: $d_{ch}(\mathbf{x}_1, \mathbf{x}_2) := (1 - \mathbf{x}_1^T \mathbf{x}_2 ^2)^{\frac{1}{2}}$,
$\mathbf{0}$	es una matriz de ceros,
\mathbf{F}_N	es la matriz DFT normalizada de longitud N : $[\mathbf{F}_N]_{k,l} := N^{-1/2} \exp(-j2\pi(k-1)(l-1)/N)$,
$f_x(x)$	es la FDP de la VA x ,
$F_x(x)$	es la FPDA de la VA x ,
\bar{x}	es el valor medio de la VA x ,
$E_x[y(x)]$	es la esperanza matemática de $y(x)$ respecto a la VA x ,
$\mathcal{CN}(\mu, \sigma^2)$	es una VA compleja Gaussiana de media μ y varianza σ^2 ,
$\mathcal{X} \cup \mathcal{Y} \ (\mathcal{X} \cap \mathcal{Y})$	es la unión (intersección) de los conjuntos \mathcal{X} y \mathcal{Y} ,
$ \mathcal{X} $	es la cardinalidad del conjunto \mathcal{X} ,
\subseteq	es contenido en,
\emptyset	es el conjunto vacío,

\mathbb{R}_0^+	es el espacio de vectores de tamaño N tales que todas sus componentes son reales mayores o iguales que cero,
\arg	es argumento,
$\operatorname{Re}\{\cdot\}$ ($\operatorname{Im}\{\cdot\}$)	es la parte real (imaginaria),
$\min\{\cdot\}$ ($\max\{\cdot\}$)	es el mínimo (máximo) de un conjunto de valores,
s. a	es sujeto a,
\lim	es el límite de una función,
$TF\{\cdot\}$ ($TF^{-1}\{\cdot\}$)	es la transformada directa (inversa) de Fourier,
$DFT\{\cdot\}$ ($IDFT\{\cdot\}$)	es la transformada discreta directa (inversa) de Fourier,
$LHS(x)$ ($RHS(x)$)	es la la igualdad izquierda (derecha) de la ecuación (x),
*	es el operador convolución,
\otimes_N	es el operador convolución circular de longitud N ,
$((x))_N$	es el resto obtenido al dividir x por N ,
$[x]^+$	es $\max\{x, 0\}$,
$\mathbf{I}_{\{x\}}$	es la función indicador: 1 si x es cierto, 0 en caso contrario,
$\operatorname{sinc}(x)$	es la función seno cardinal: $\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$,
$\Gamma(b, x)$	es la función <i>Gamma</i> incompleta: $\Gamma(b, x) := \int_x^\infty t^{b-1} e^{-t} dt$,
$\Gamma(b)$	es la función <i>Gamma</i> : $\Gamma(b) := \Gamma(b, 0)$,
$Q(\cdot)$	es la función <i>Q</i> de Marcum: $Q(x) := 1/\sqrt{2\pi} \int_x^\infty \exp(-t^2/2) dt$.

Capítulo 1

Introducción

*“Estiman algunos los libros por su corpulencia,
como si se escribiesen para ejercitar antes los brazos que los ingenios.
La extensión sola nunca pudo exceder de medianía,
y es plaga de hombres naturales por querer estar en todo no estar en nada.”*

Baltasar Gracián – Oráculo manual y arte de prudencia (s. XVI d.C.)

El objetivo principal de esta tesis consiste en la minimización de la potencia transmitida en sistemas que se comunican a través de canales inalámbricos con desvanecimientos. Con esta finalidad se derivarán esquemas óptimos de adaptación para los transmisores de forma que la potencia transmitida se minimice a la vez que se satisfagan requisitos de tasa de transmisión mínima y probabilidad de error máxima. La adaptación de los transmisores consistirá en el ajuste de parámetros del nivel físico (tasa de transmisión, modulación utilizada y potencia transmitida) y del nivel de enlace (selección de usuario) como función del CSI¹. El mayor énfasis investigador se ha dedicado a sistemas que disponen de un canal de realimentación (control) con tasa limitada a través del cual el receptor envía al transmisor(es)

¹El acrónimo CSI, cuyo uso está extendido en la literatura y que puede traducirse del inglés como *información del estado del canal* o *conocimiento del estado del canal*, se utilizará de forma continuada a lo largo de esta disertación. Cuando se le añada una T para formar CSIT se hará referencia al CSI en el *transmisor* o *transmisores*; de igual modo, cuando se le añada una R formando CSIR se referirá al CSI en el *receptor* o *receptores*. Asimismo, cuando el acrónimo esté precedido por una inicial seguida de un guión estará denotando un tipo específico de CSI. Así por ejemplo, P-CSI hará referencia al conocimiento *perfecto* del estado del canal (el uso de P- se debe al término inglés *perfect*) y Q-CSIT hará referencia al conocimiento *cuantificado* del estado del canal por parte de los transmisores (el uso de Q- se debe al término inglés *quantized*).

CSI. La limitación natural de la tasa del canal de realimentación provoca que el CSI tenga que ser descrito por un número finito de bits y que, por lo tanto, los transmisores sólo puedan adquirir CSI *cuantificado* (Q-). El presente capítulo muestra de una parte las razones que motivan este trabajo, los objetivos a cubrir por esta tesis así como los antecedentes y trabajos previos relacionados con el problema aquí analizado. De otra parte, se describe la estructura de esta tesis y sus principales contribuciones.

1.1. Motivación, objetivos y estado del arte

Las ventajas en movilidad y rapidez de despliegue que ofrecen los sistemas de comunicación inalámbrica han motivado una creciente demanda de estos servicios en los últimos años. Especialmente, se ha impulsado la búsqueda de tecnologías que permitan enlaces de alta capacidad y fiabilidad manteniendo un bajo consumo en diferentes tipos de escenarios [26, 85]. Aunque pertenecientes al ámbito de las comunicaciones digitales y por lo tanto usufructuarias de sus principios básicos, las comunicaciones (móviles) inalámbricas presentan ciertas características propias entre las que destacan: la variabilidad aleatoria del canal debido a la movilidad de los terminales y que se traduce en desvanecimientos de la señal; la escasez del espectro radioeléctrico, en general compartido por diferentes usuarios y potencialmente generador de interferencias, y las limitaciones de consumo, tamaño o coste del terminal. Estas particularidades motivan que al análisis de este tipo de sistemas se realice de forma propia y diferenciada del resto.

Las prestaciones de un sistema de comunicación sobre el medio inalámbrico con desvanecimientos (ya sean medidas en términos de tasa de transmisión, consumo de potencia o probabilidad de error) mejoran si el sistema dispone de CSI. Así por ejemplo, los demoduladores coherentes, que requieren CSIR, son los más extendidos debido a sus buenas prestaciones [72], y la capacidad ergódica del canal (medida de la máxima tasa media de información que se puede transmitir de forma fiable a través de dicho canal) sólo puede alcanzarse si existe CSIT [12, 26, 85].

Si bien en la mayoría de los sistemas de comunicaciones móviles es factible que los terminales dispongan de CSIR *perfecto* (P-), factores como los errores de estimación o los retardos de estimación hacen que la existencia de P-CSIT no sea pragmáticamente viable [38, 63], motivando la utilización de otras alternativas. Para sistemas en los que el CSI es adquirido por el receptor y enviado a los transmisores a través de un canal de retorno con tasa limitada, la existencia de P-CSIT no es siquiera posible; sin embargo, en estos sistemas los terminales sí pueden adquirir Q-CSIT de manera sencilla (la limitación de tasa no permite el envío de los valores analógicos del canal pero sí de versiones que se cuantifican mediante unos pocos bits) y utilizarlo para mejorar las prestaciones del sistema [24, capítulo 13].

Esta tesis se enmarca en el ámbito de las comunicaciones inalámbricas sobre medios con desvanecimientos en los que la configuración de los transmisores puede adaptarse en función de la realización del canal. El principal objetivo consiste en obtener el esquema de adaptación óptimo para los transmisores en base a Q-CSIT que permita *minimizar la potencia transmitida por el sistema satisfaciendo requisitos de calidad de servicio* (o *quality of service –QoS–*) en términos de tasa de transmisión mínima y tasa de error de bit (o *bit error rate –BER–*) máxima. La minimización de la potencia transmitida permite la prolongación del tiempo de vida de los terminales al mismo tiempo que disminuye la potencial interferencia para otros usuarios, aspectos estos cada vez más relevantes en sistemas de comunicaciones móviles. A fin de llevar a cabo esa tarea dos son las preguntas principales que deben contestarse: ¿cuál es la cuantificación óptima del canal?, ¿cuáles son los esquemas de adaptación óptimos basados en los valores cuantificados del canal? Las respuestas de cuestiones están interrelacionadas y para llegar a ellas se requiere la resolución conjunta de ambos problemas.

Tanto la cuantificación del canal como la adaptación de los transmisores en función de la misma dependerá del sistema de comunicaciones considerado. El análisis en esta tesis doctoral se centrará en: (i) sistemas transmitiendo un único flujo de información sobre canales móviles con múltiples entradas (correspondientes a múltiples transmisores) y una única salida o *multi-input single-output* (MISO); (ii)

sistemas de un único usuario transmitiendo sobre canales selectivos en tiempo y frecuencia mediante el uso de *orthogonal frequency division multiplexing* (OFDM); (iii) sistemas de múltiples usuarios operando sobre canales selectivos en el tiempo accediendo al medio compartido utilizando un mecanismo *time division multiple access* (TDMA) y (iv) sistemas de múltiples usuarios operando sobre canales selectivos en el tiempo y frecuencia accediendo al medio compartido utilizando un mecanismo *orthogonal frequency division multiple access* (OFDMA). En base a *técnicas de optimización no lineal*, para cada uno de estos casos se derivará tanto el *cuantificador de canal* que genere el Q-CSIT como los esquemas de *adaptación de los transmisores* que en base al Q-CSIT minimicen la potencia. En todos los casos se hará referencia también a la complejidad del sistema y se analizarán mecanismos que permitan reducirla. Asimismo, para los casos en que no existan, el estudio comenzará derivando los esquemas de adaptación de los transmisores en base a P-CSIT. Además de su uso en sistemas donde el P-CSIT sea factible, la caracterización de estos esquemas ayudará a comprender mejor el funcionamiento de los sistemas adaptativos así como a ponderar, mediante simulación numérica, la diferencia de prestaciones de los esquemas basados en Q-CSIT y P-CSIT.

Las aportaciones de esta tesis se encuentran relacionadas principalmente con la investigación realizada en dos ámbitos que, si bien relacionados, han recibido un tratamiento distinto.

- De un lado el de la *asignación de recursos* –también conocido como (t.c.c.) *adaptación de los modos de transmisión*– que trata de aprovechar el conocimiento perfecto del canal para asignar recursos de forma eficiente. Los trabajos en este ámbito se han centrado masivamente en el uso de P-CSIT. Los resultados obtenidos alcanzan los campos de la teoría de la información [18,29,40,41,84], fundamentos de sistemas de comunicaciones [27,30,69,79] e incluso más recientemente de redes y protocolos de comunicación [39,48]. Los objetivos específicos, aunque dependientes del escenario de aplicación, han estado típicamente relacionados con la maximización de la tasa de informa-

ción (capacidad, tasa de transmisión o flujo de paquetes) y la minimización de la BER. Respecto a la optimalidad de las soluciones y a la capacidad de adaptación de los transmisores, han sido en general variadas. Existen resultados que representan óptimos globales en los que la adaptatividad es total (se adapta tasa, potencia, codificación y se hace para cualquiera de las dimensiones –tiempo, frecuencia– en las que los canales varíen) y que se relacionan con los escenarios más teóricos (e.g., teoría de la información) donde la intención principal suele ser a de encontrar límites fundamentales de prestaciones. Existen asimismo esquemas que perdiendo optimalidad, están orientados a escenarios prácticos de aplicación (i.e., redes de paquetes), donde de hecho no se pretende encontrar esquemas óptimos de adaptación sino simplemente esquemas eficientes de fácil implementación y basados en principios intuitivos) [3].

- De otro lado, el diseño de sistemas Q-CSIT transmitiendo a través de canales con desvanecimientos, orientados principalmente al diseño de un *cuantificador de canal*. Los trabajos existentes en este ámbito se centran primordialmente en sistemas de múltiples entradas y múltiples salidas (o *multi-input multi-output* –MIMO–) en donde la existencia de P-CSIT es complicada (e.g., para el caso de un sistema MIMO de 8 antenas transmisoras y 8 antenas receptoras, la estimación de la matriz de canal exigiría la obtención de 64 valores analógicos complejos que deberían ser conocidos por el transmisor). Para el diseño de sistemas MIMO los parámetros de transmisión a adaptar han sido en general vectores de apuntamiento, matrices de pre-codificación lineal o códigos espacio temporales, pero no modulación o la potencia (mas allá de la distribución intrínseca que producen las propias matrices de pre-codificación). La ventaja de esta simplificación es que permite plantear problemas de cuantificación relacionados con otros problemas clásicos de compactación y cuantificación sin tener que pasar por la formulación de un Lagrangiano (que habitualmente impide la obtención de una

forma cerrada del cuantificador). Las prestaciones de comunicación a optimizar en estos casos han sido típicamente la capacidad instantánea y la relación señal a ruido (o *signal-to-noise-ratio* –SNR–) [44, 46, 66, 93], o la BER [102, 103]. Asimismo, existen también algunos resultados que sí adaptan la tasa de transmisión [45, 94] pero lo hacen en general de forma sub-óptima, sin aprovechar la ergodicidad del canal y sin adaptar la potencia a lo largo del tiempo.

Aunque existen algunos trabajos que intentan combinar ambos ámbitos (véanse, e.g., [19, 37, 47]), ninguno lo ha hecho desde un planteamiento global, donde se intente optimizar tanto el diseño de la cuantificación del canal como de la propia asignación de recursos, ni ha considerado la influencia de estos diseños en el consumo de potencia. Ese será el objetivo principal de esta tesis que, utilizando un planteamiento y una metodología propios y novedosos, intentará también aprovechar parte de los resultados de los trabajos previamente citados.

1.2. Estructura y publicaciones científicas asociadas a la tesis

La presente tesis se articula en torno a cinco capítulos. Tras concluir el que ahora ocupa, en el capítulo 2 se realiza una revisión de conceptos del ámbito de las comunicaciones móviles inalámbricas que resultan de especial relevancia para este trabajo: las características principales de la transmisión inalámbrica, la descripción básica de los modelos de canal utilizados, y el concepto de CSI, haciendo hincapié en sus distintos tipos y en su uso en sistemas de comunicaciones para la mejora de prestaciones. Las principales contribuciones de esta tesis se detallan en los capítulos 3 y 4. Con el objetivo de minimizar la potencia transmitida en base a la adaptación del esquema de transmisión en función del CSI, el en el primero de ellos se derivan esquemas de adaptación del transmisor para sistemas

en los que existe un único flujo de información, mientras que en el segundo se hace lo propio en sistemas de múltiples usuarios que originan múltiples flujos de información independientes. Finalmente, el capítulo 5 cierra esta disertación sintetizando las contribuciones y resultados de esta tesis y reflexionando sobre líneas futuras de investigación relacionadas con el trabajo aquí desarrollado.

A continuación se proporciona una panorámica general de los capítulos 3-5 en la que se detallan las publicaciones científicas asociadas a cada uno de ellos.

Capítulo 3

Este capítulo considera la obtención de estrategias de adaptación en el transmisor en base a CSI que permitan minimización de la potencia transmitida para sistemas en los que existe un único flujo de información. El capítulo comienza analizando el caso de una única fuente de información cuyo flujo de datos se transmite sobre canales móviles con múltiples entradas y una única salida o *multiple-input single-output* (MISO). Este modelo se asocia a una red inalámbrica de sensores o *wireless sensor network* (WSN) con múltiples transmisores y un único receptor. En la segunda sección, para comunicaciones de un único usuario sobre canales doblemente selectivos (i.e., canales selectivos en tiempo y en frecuencia), se analiza la minimización de potencia de un sistema OFDM.

Las contribuciones más relevantes de la primera sección de este capítulo han sido publicadas en un artículo en revista y dos contribuciones a conferencias:

- A. G. Marques, X. Wang y G. B. Giannakis, “Minimizing Transmit-Power for Coherent Communications in Wireless Sensor Networks with Finite-Rate Feedback”, *IEEE Transactions on Signal Processing*. (Enviado para publicación en septiembre de 2006; referencia [57].)
- A. G. Marques, X. Wang y G. B. Giannakis, “Energy Efficient MISO Systems Using Adaptive Modulation and Coding”, *Proc. of 40th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, EE.UU., 29 octubre - 1 noviembre, 2006. (Artículo invitado; referencia [58].)

- A. G. Marques, X. Wang y G. B. Giannakis, “Minimizing Transmit-Power for Coherent Communications in Wireless Sensor Networks using Quantized Channel State Information”, *IEEE Proc. of Intl. Conf. on Acoustics, Speech and Signal Processing*, Honolulu, EE.UU., 16-20 abril , 2007. (Referencia [61].)

Respecto a la minimización de potencia para comunicaciones de un único usuario en sistemas OFDM, parte de los resultados obtenidos han sido publicados en:

- A. G. Marques, F. F. Digham y G. B. Giannakis, “Optimizing Power Efficiency of OFDM Using Quantized Channel State Information”, *IEEE Journal on Selected Areas in Communications*, Vol. 24, No. 8, agosto 2006 pp. 1581 – 1592. (Número especial sobre *Optimización no lineal de sistemas de comunicaciones*; referencia [50].)
- A. G. Marques, F. F. Digham y G. B. Giannakis, “Power-Efficient OFDM via Quantized Channel State Information”, *IEEE Proc. of Intl. Conf. on Communications*, Estambul, Turquía, 11-15 junio, 2006. (Referencia [51].)
- A. G. Marques, F. F. Digham y G. B. Giannakis, “Power-Efficient OFDM with Reduced Complexity and Feedback Overhead”, *IEEE Proc. of Intl. Conf. on Acoustics, Speech and Signal Processing*, Toulouse, Francia, 16-20 mayo, 2006. (Referencia [52].)

Capítulo 4

Este capítulo se dedica al estudio de la minimización de la potencia transmitida por sistemas de múltiples usuarios transmitiendo flujos independientes de información. En primer lugar se analizará la adaptación óptima de los transmisores y la cuantificación de canal para un sistema TDMA. en el que cada usuario puede transmitir de acuerdo a un juego de modos de modulación y codificación adaptativa (*adaptive modulation and coding* –AMC–) predeterminado. En segundo lugar se propondrán esquemas óptimos (más complejos) y sub-óptimos (más sencillos)

para la cuantificación del canal por parte del receptor así como la adaptación de los transmisores en sistemas OFDMA.

Parte de los resultados obtenidos para sistemas TDMA han sido publicados en los artículos:

- X. Wang, A. G. Marques y G. B. Giannakis, “Power-Efficient Resource Allocation and Quantization for TDMA Using Adaptive Transmission and Limited-Rate Feedback”, *IEEE Transactions on Signal Processing*. (Enviado para publicación en mayo de 2006; referencia [89].)
- A. G. Marques, X. Wang y G. B. Giannakis, “Optimizing Energy-Efficient of TDMA with Finite Rate Feedback”, *IEEE Proc. of Intl. Conf. on Acoustics, Speech and Signal Processing*, Honolulu, EE.UU., abril 16-20, 2007. (Referencia [59].)
- A. G. Marques, X. Wang y G. B. Giannakis, “Energy-Efficient TDMA with Quantized Channel State Information”, *Proc. of MILCOM Conf.*, Washington, EE.UU., 23-25 octubre, 2006. (Galardonado con el premio a la mejor presentación –realizada por el tercer autor–; referencia [62].)

Mientras que los resultados correspondientes a la minimización de potencia en sistemas OFDMA han sido publicados en los siguientes artículos [53–55]:

- A. G. Marques, F. F. Digham, G. B. Giannakis y F. J. Ramos, “Power Efficient Wireless OFDMA using Limited-Rate Feedback”, *IEEE Transactions on Wireless Communications* (Enviado para publicación en agosto de 2006; referencia [54]).
- A. G. Marques, G. B. Giannakis, F. F. Digham y F. J. Ramos, “Minimizing Power in Wireless OFDMA with Limited Rate Feedback,” *IEEE Proc. of Wireless and Communications and Networking Conf.*, Hong Kong, Hong Kong, 11-15 marzo, 2007. (Referencia [55].)

- A. G. Marques, G. B. Giannakis, F. F. Digham, and F. J. Ramos, “Reduced-Complexity Power-Efficient Wireless OFDMA using an Equally Probable CSI Quantizer”, *IEEE Proc. of Intl. Conf. on Communications*, Glasgow, Escocia, 24-28 junio, 2007. (Referencia [53].)

Capítulo 5: líneas de investigación relacionadas

A continuación, se enumeran otros resultados que pese a no haberse incluido en el cuerpo de esta tesis, presentan una clara relación con los objetivos de la misma y pueden considerarse un primer paso en distintas direcciones de futuro en las que el trabajo expuesto puede extenderse.

La cuantificación del CSI en el transmisor surge de forma natural a través de la suposición de estimación perfecta del canal en el receptor y de la transmisión de este valor desde el receptor hasta el transmisor a través de un canal de retorno con tasa limitada. No obstante, existen otros escenarios en los que la cuantificación del CSI no se produce sólo en los transmisores, sino también en los receptores. El análisis de este escenario para el caso de múltiples usuarios, así como la obtención de asignación óptima de recursos se considera en:

- A. G. Marqués, G. B. Giannakis and F. J. Ramos, “Optimizing Multiple Access in Wireless Communications Based on Quantized CSI,” *IEEE Transactions on Communications*. (En preparación, fecha estimada de envío en marzo de 2007; referencia [56].)

En sistemas de múltiples usuarios, la adaptación óptima de parámetros del nivel físico (tasa, potencia o modulación) y de enlace (usuario y canal asignado) requiere la integración sobre la función densidad de probabilidad del canal inalámbrico [40, 55,62,84], lo que conlleva la realización de cálculos que en ocasiones pueden resultar computacionalmente costosos. Sin embargo, trabajos existentes que contemplan la asignación de recursos a nivel de enlace muestran la posibilidad de utilizar algoritmos estocásticos que para cada realización del canal calculan la asignación de recursos y actualizan la estimación de los parámetros a largo plazo con un

consumo computacional despreciable. La extensión de los esquemas de adaptación óptima de parámetros de los niveles físico y de enlace obtenidos en esta disertación a algoritmos estocásticos que no necesiten conocer la descripción probabilística del canal resulta de tremendo interés. El marco teórico para atacar este problema así como resultados para el caso de optimización de sistemas TDMA para el enlace ascendente y TDM (*time division multiplexing*) para el enlace descendente en base a conocimiento perfecto y cuantificado del canal se ha publicado en [88]:

- X. Wang, G. B. Giannakis y A. G. Marques, “A Unified Approach to QoS-Guaranteed Scheduling for Channel-Adaptive Wireless Networks”, *Proceedings of IEEE*. (Enviado para publicación en diciembre de 2006; referencia [88].)

Recientes estudios muestran que las redes móviles de telecomunicación actuales infrautilizan el uso del espectro radioeléctrico [4, 31, 64]. A modo ilustrativo, en redes en las que únicamente se permite el acceso a usuarios previamente registrados y se ofrecen garantías de calidad de servicio para todos ellos, el uso del espectro puede llegar a ser menor del 20%. Una posible alternativa para enfrentarse a estas circunstancias consiste en el desarrollo de algoritmos que consideren varias categorías de transmisores en función de sus garantías acceso al canal y de QoS (primarios, secundarios, oportunistas, ...) incrementando así el número de potenciales usuarios del canal. Resultados preliminares de algoritmos adaptativos que asignan recursos de forma óptima y se ajustan al escenario descrito pueden encontrarse en:

- A. G. Marques, X. Wang y G. B. Giannakis, “Distributed Resource Allocation for Cognitive Radios based on Limited-Rate Feedback,” *IEEE Proc. of XV European Signal Processing Conf.*, Poznan, Polonia, Septiembre 3-7, 2007. (Artículo invitado; referencia [60].)

Un análisis más detallado de estas y otras líneas de futuro de este trabajo se proporcionará en el capítulo 5.

Capítulo 2

El canal inalámbrico con desvanecimientos y el uso de la información del estado del canal

Este capítulo se inicia con una revisión de las principales características y modelos de propagación del canal inalámbrico (en general móvil) con desvanecimientos, fenómeno complejo caracterizado por diversos efectos (e.g. efectos Doppler y multi-trayecto) cuyo modelado y caracterización ha captado considerables esfuerzos investigadores [33,74,77]. Posteriormente se realiza una revisión de parámetros de transmisión básicos de un sistema de comunicaciones, haciendo especial hincapié en la potencia transmitida y en su minimización a través de técnicas de optimización, por ser este un objetivo principal de esta tesis. La última parte del capítulo se dedica al concepto de CSI. Tras su presentación, se analizan distintas categorías y su correspondencia con escenarios de transmisión, poniendo especial énfasis en el CSI cuantificado. Tras la revisión de estos conceptos se estará en condiciones de abordar en los capítulos 3 y 4 la derivación y el análisis de los resultados más relevantes de esta disertación.

2.1. El canal inalámbrico

*La intensidad de la onda electromagnética emitida por el transmisor se atenúa con la distancia elevada a un factor que oscila entre 2 para el medio libre y 4-6 en entornos urbanos o interiores [74]. A este hecho, ha de añadirse la posibilidad de que la comunicación no se encuentre en línea directa de visión. En estos escenarios, la frecuencia a la que se está transmitiendo tiene un papel fundamental para definir índices de penetración, reflexiones o sombras electromagnéticas sobre los obstáculos. En este sentido, los efectos más perjudiciales para la propagación inalámbrica son la difracción, reflexión y dispersión de la señal. Como resultado se producen importantes *desvanecimientos aleatorios* de la señal en el receptor.

El efecto denominado *multi-trayecto* aparece cuando en el receptor se obtienen distintas versiones de la señal que han llegado a través de distintos trayectos, con o sin línea de visión directa. Como resultado, a la componente principal de la señal (la correspondiente, en caso de que exista, al rayo directo) se superponen versiones retrasadas de la misma señal causando interferencia entre símbolos o *inter-symbol interference* (ISI). Además, las señales pueden llegar simultáneamente y pueden diferir en amplitud, fase o frecuencia, provocando sumas constructivas o destructivas que atenúan o amplifican aleatoriamente la señal resultante en el receptor. Si además se supone que los terminales o el medio están en movimiento, la señal recibida también varía a lo largo del tiempo, constituyendo el denominado *efecto Doppler*.

Los desvanecimientos aleatorios (conocidos como *fading* en la literatura en lengua inglesa) del medio son una de las mayores diferencias de la transmisión inalámbrica con respecto a (c.r.a.) la transmisión por cable. En general, los efectos del desvanecimiento del canal se clasifican en tres tipos [74]: (i) variaciones a gran escala; (ii) variaciones a media escala modeladas como distribuciones log-normales y (iii) variaciones a pequeña escala, t.c.c. desvanecimientos rápidos, con distribuciones estadísticas Rice (suponiendo rayo directo), Rayleigh (suponiendo

*El autor agradece al Dr. D. Alfonso Cano Pleite su permiso para la reproducción de parte de los contenidos de [71] en la sección 2.1.

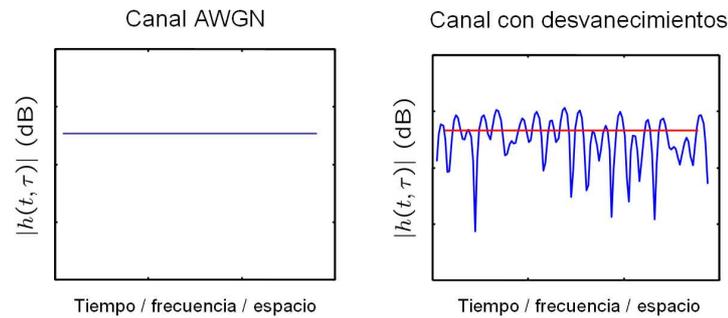


Figura 2.1: Canal AWGN (sin desvanecimientos) frente a canales con desvanecimientos.

la ausencia rayo directo) o en general, Nakagami- m (que engloba las anteriores como casos particulares [74, 77]). Mientras que los efectos (i) y (ii) pueden mitigarse (resolverse) mediante el re-diseño del sistema a largo-medio plazo, este no es el caso para los desvanecimientos rápidos en (iii). Este desafiante escenario, causando además de la degradación más severa de prestaciones en canales inalámbricos móviles, será la principal ocupación a lo largo de esta tesis.

Los parámetros básicos que caracterizan la transmisión inalámbrica sobre entornos con variaciones a pequeña escala son la coherencia y la selectividad. La coherencia describe la longitud del intervalo (espacial, temporal o frecuencial) en el que el valor del canal se puede considerar aproximadamente constante. La selectividad, por otro lado, mide la variabilidad del canal. La figura 2.1 muestra un canal sin desvanecimientos con ruido modelado como *average white gaussian noise* (AWGN) y otro selectivo en tiempo, frecuencia o espacio [74]. El eje vertical representa la amplitud de la respuesta al impulso del canal $h(t; \tau)$ en escala logarítmica. A continuación se repasan los parámetros que caracterizan $h(t; \tau)$: la dispersión temporal que provoca selectividad en frecuencia y la selectividad en el tiempo provocada por el efecto Doppler.

2.1.1. Dispersión temporal

Si se supone que la respuesta al impulso del canal $h(t; \tau)$ es invariante en el tiempo, la respuesta al impulso se escribe $h(t; \tau) = h(\tau)$ y se dice entonces

que el canal es dispersivo en el tiempo [74]. La medida de esta dispersión, τ_d , indica la longitud máxima del proceso $\phi(\tau) = E\{|h(\tau)|^2\}$ que considera todas las realizaciones posibles de $h(\tau)$ [74]. Para una duración y periodo de símbolo dados, el valor de τ_d indica el grado de ISI.

Se puede igualmente apreciar el efecto de la dispersión temporal analizando la respuesta en frecuencia del canal. La transformada de Fourier (TF) de $\phi(\tau)$ define la densidad espectral de potencia del canal, $\Phi(f)$. Sobre esta función, podrá definirse el ancho de banda de coherencia, BW_c , que indica el intervalo de frecuencias en el que se supone la respuesta en frecuencia del canal se puede considerar aproximadamente plana. BW_c está inversamente relacionado con τ_d , es decir

$$BW_c \approx \frac{1}{\tau_d}. \quad (2.1)$$

En general, los sistemas de comunicación de banda ancha pueden superar BW_c y por tanto sufren selectividad en frecuencia.

2.1.2. El efecto Doppler y la dispersión en frecuencia

Supóngase que el canal inalámbrico tiene una respuesta al impulso no dispersiva pero variante en el tiempo $h(t; \tau) = h(t)$. Esta variabilidad viene normalmente provocada por el efecto Doppler. Este tipo de canales se denominan multiplicativos, pues la señal en el receptor es la misma señal de entrada multiplicada por el canal $h(t)$ para cada uno de los instantes de tiempo. Si se introduce un impulso en la frecuencia, este canal esparce esta frecuencia a lo largo de múltiples copias con una máxima dispersión dictada por la frecuencia Doppler [74]. Definiendo BW_d como una medida de esta dispersión máxima, el tiempo de coherencia T_c está inversamente relacionado con BW_d , o lo que es lo mismo

$$T_c \approx \frac{1}{BW_d}. \quad (2.2)$$

Pudiendo interpretarse este resultado como el dual del obtenido para los canales dispersivos en el tiempo.

2.1.3. Canales selectivos en tiempo y frecuencia y modelo discreto equivalente

Sea $h(t; \tau)$ la respuesta al impulso del canal. En un canal dispersivo en tiempo y frecuencia, $h(t; \tau)$ es la superposición retardada de múltiples rayos

$$h(t; \tau) = \sum_p h_p(t) \delta(\tau - \tau_p(t)), \quad (2.3)$$

en donde $h_p(t)$ es igualmente la suma de múltiples sub-rayos con un mismo retardo τ_p . El m -ésimo sub-rayo dentro del p -ésimo rayo tiene amplitud $a_{p,m}(t)$, fase $\phi_{p,m}(t)$ y desplazamiento en frecuencia $f_{p,m}(t)$ debido al efecto Doppler¹ y tiene la forma

$$h_p(t) = \sum_m a_{p,m}(t) e^{j\phi_{p,m}(t)} e^{j2\pi f_{p,m}(t)t}. \quad (2.4)$$

El canal equivalente $c(t; \tau)$ se define como el resultado de convolucionar los filtros en transmisión $p_T(t)$ y en recepción $p_R(t)$ y se escribe como

$$c(t; \tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_T(\tau - u - s) p_R(s) h(t - s; u) ds du. \quad (2.5)$$

Suponiendo que $h(t; \tau)$ permanece invariante durante la duración del filtro en recepción $p_R(t)$, la respuesta equivalente puede reescribirse como

$$c(t; \tau) = \int_{-\infty}^{\infty} h(t; u) \left(\int_{-\infty}^{\infty} p_T(\tau - u - s) p_R(s) ds \right) du. \quad (2.6)$$

Si se define $\varphi(t) = p_T(t) * p_R(t)$, el modelo de canal en (2.6) se reduce a

$$c(t; \tau) = \int_{-\infty}^{\infty} h(t; u) \varphi(\tau - u) du. \quad (2.7)$$

¹Nótese que este desplazamiento en frecuencia podría también deberse a desviaciones en los osciladores del receptor.

Por simplicidad, de aquí en adelante se supondrá que $\tau_p(t)$, $f_{p,m}(t)$ y $a_{p,m}(t)$ permanecen invariantes durante la duración de la transmisión². La expresión (2.7) se puede particularizar para canales únicamente selectivos en tiempo o frecuencia:

- *Canal selectivo en frecuencia:* En este caso $h_p(t)$ en (2.4) es invariante en el tiempo. Fijando $h_p(t) = h_p$, $h(t; \tau)$ en (2.7) se puede escribir como

$$c(t; \tau) = c(\tau) = \sum_p h_p \varphi(\tau - \tau_p), \quad (2.8)$$

que es el conocido perfil de retardo de los canales multi-trayecto [74].

- *Canal selectivo en el tiempo:* En este caso, (2.4) tiene un único rayo τ_0 . Si se supone que $\varphi(\tau - \tau_0) = \delta(\tau - \tau_0)$ (es decir, se satisface el criterio de Nyquist), (2.7) se puede escribir como:

$$c(t; \tau) = c(t) = \sum_m a_m e^{j\phi_m} e^{j2\pi f_m t}. \quad (2.9)$$

Si se define $\beta_m(t) = [-\pi, \pi)$ como una variable aleatoria (VA) con distribución uniforme y se sustituye $f_m = f_{\max} \cos(\beta_m)$, con $f_{\max} := v_{\max} f_c / c$, siendo v_{\max} la velocidad máxima, f_c la frecuencia central y c la velocidad de la luz, la expresión en (2.9) coincide con el modelo Jakes en [33].

Hasta este punto la señal continua $h(t; \tau)$ ha representado la respuesta al impulso del canal. No obstante, los sistemas de comunicaciones trabajan en base a un tiempo de muestreo T_s , por esto resulta conveniente definir la señal discreta

$$h(n; q) := h(t = nT_s; \tau = qT_s), \quad (2.10)$$

con $q = 0, 1, \dots, Q$, $Q := \lceil D_{\max}/T_s \rceil$ y $D_{\max} := \max_p \{\tau_p\}$, que representará el *canal discreto equivalente* y que engloba los efectos de los filtros de transmisión, recepción y muestreo en el receptor. Nótese que con la definición en (2.10), para cada

²La variabilidad temporal de $h_p(t)$ viene ahora dada únicamente por el término $e^{j2\pi f_{p,m} t}$ en (2.4), es decir, las desviaciones en frecuencia.

instante n el canal $h(n; q)$ estará formado por $Q + 1$ coeficientes correspondientes a los distintos rayos recibidos.

2.2. Parámetros básicos y minimización de la potencia en los sistemas de comunicaciones

El primer paso para el análisis de un sistema de comunicaciones, al igual que para cualquier otro sistema de ingeniería, consiste en seleccionar el modelo que lo caracteriza y, en función de su nivel de abstracción, el conjunto de parámetros que describe su funcionamiento. Así pues, una vez expuesto el modelo de canal considerado, resulta necesario aclarar cuál es el grado de abstracción con el que se modelarán los sistemas de comunicaciones que se analicen y, sobre todo, cuáles serán los parámetros fundamentales que describan su comportamiento.

2.2.1. Parámetros básicos

Se supondrá que el transmisor utiliza distintos esquemas para codificar y modular la información a transmitir dando lugar, para el n -ésimo tiempo de muestreo, al símbolo $x(n)$ que s.p.d.g. se asume unitario (i.e. $|x(n)| = 1$). Asimismo, se considera que el transmisor tiene capacidad de escalar este símbolo por un factor $\lambda(n)$ a fin de modificar a potencia (energía) de transmisión. Si la información se transmite por el canal en (2.10) que introduce además un ruido AWGN con media nula y densidad espectral de potencia $N_0/2$ [vatios/hertzio], la n -ésima muestra en el receptor, $y(n)$, será la superposición de los símbolos transmitidos afectados por el canal doblemente selectivo $h(n; q)$ y está dada por

$$y(n) = T_s \sum_{q=0}^Q h(n; q) \lambda(n - q) x(n - q) + v(n), \quad (2.11)$$

donde $v(n)$ representa el ruido, que tiene media nula y varianza N_0T_s . En ocasiones se prefiere la versión escalada de la ecuación anterior y se escribe (véase, e.g., [72])

$$y(n) = \sqrt{T_s} \sum_{q=0}^Q \lambda(n-q)x(n-q)h(n;q) + v(n), \quad (2.12)$$

presentando en esta ocasión el ruido $v(n)$ media nula y varianza N_0 .

Independientemente de la versión elegida, la potencia de transmisión p durante la n -ésima muestra puede calcularse como

$$p(n) = |\lambda(n)x(n)|^2 = |\lambda(n)|^2, \quad (2.13)$$

mientras que la relación señal a ruido (SNR) recibida para el q -ésimo coeficiente resulta

$$\gamma(n;q) = \frac{p(n-q)T_s}{N_0} |h(n;q)|^2. \quad (2.14)$$

adviértase que $h(n;q)$ es en general una VA y que, por lo tanto, la SNR también lo será. En (2.12) es frecuente expresar el valor de la señal recibida $y(n)$ a través del parámetro $\sqrt{E_s(n)} := \lambda(n)\sqrt{T_s}$ de manera que $E_s(n) = |\lambda(n)|^2T_s = p(n)T_s$ puede ser rápidamente interpretado como la *energía por símbolo* y la SNR alternativamente expresada como

$$\gamma(n;q) = \frac{E_s(n-q)}{N_0} |h(n;q)|^2 \quad (2.15)$$

Adviértase que la SNR total asociada al símbolo $x(n)$ dependerá del esquema de detección utilizado y de la forma en que se combinen las contribuciones de los distintos coeficientes $h(n;q)$ $q = 0, 1, \dots, Q$ (suma coherente, *maximum ratio combining*, etc.).

Teniendo en cuenta el modelo anterior, se considerarán como las prestaciones básicas de la transmisión de datos:

- La *tasa binaria de transmisión* r que, medida en términos de bits por segundo y para la muestra n , corresponde al número de bits codificados en $x(n)$

dividido por el tiempo de símbolo T_s . Puesto que el transmisor puede crear el símbolo $x(n)$ de acuerdo a distintos esquemas de modulación y codificación, la tasa transmitida en cada símbolo puede ser diferente. Así por ejemplo, si el transmisor construye $x(n)$ a partir de una modulación *phase shift keying* (PSK) *binaria* a la que añade un codificador convolucional de tasa $1/2$, la tasa de transmisión será $r(n) = \log_2(2)(1/2) = 0.5$ bits por símbolo ($0.5/T_s$ bits por segundo); mientras que si construye $x(n)$ a partir de una modulación PSK *cuaternaria* y no utiliza ningún tipo de codificación la tasa transmitida será $r(n) = \log_2(4) = 2$ bits por símbolo ($2/T_s$ bits por segundo).

- La BER o *tasa de error de bit* ϵ , que corresponde al número de bits erróneos dividido por el número de bits transmitidos. Pese a que la tasa de error de bit puede calcularse de forma exacta para cada n , al ser una VA (nótese que $v(n)$ lo es) suele ser más conveniente considerar su valor esperado. La expresión (función) del valor esperado de ϵ dependerá, de un lado, de los esquemas de codificación y modulación utilizados y, de otro, de la tasa de transmisión de bits y de la SNR recibida. Esta relación entre la tasa de transmisión, la potencia transmitida y el valor del canal a través de la BER jugará un papel fundamental en el diseño del sistema.

Por otro lado los transmisores se caracterizarán además de por la *potencia transmitida* p , por la *complejidad* de los mismos. Básicamente, la *complejidad* describirá la capacidad de adaptación del transmisor recogiendo aspectos tales como la variación del parámetro λ o los esquemas de modulación y codificación que es capaz de implementar.

Como se ha avanzado en la introducción, en el modelo también se tendrán en cuenta la *información de control* enviada y el *acceso al canal* en sistemas de múltiples usuarios. Estos parámetros están claramente relacionados y cualquier diseño que optimice una determinada medida de prestaciones tendrá que tener en cuenta la existencia de una relación de compromiso entre ellos (e.g., dado un punto de operación, el incremento de la tasa requerirá un incremento en la potencia

transmitida, en la complejidad del sistema o en la probabilidad de error).

Concluida la descripción específica de los parámetros más relevantes del sistema, debe cerrarse este apartado retomando la discusión respecto al nivel de abstracción del modelo utilizado. Así, desde un punto de vista más general y tomando como referencia la jerarquía de protocolos de comunicación OSI [80], la presente tesis focalizará su atención en el nivel físico y analizará algún aspecto del nivel de enlace (canal de control y acceso al medio en sistemas de múltiples usuarios). Del mismo modo, debe señalarse que no serán objeto de diseño aspectos tales como la implementación física de los receptores y transmisores (circuitos y componentes hardware), la sincronización o el descubrimiento de terminales; bien porque estén resueltos, bien porque su diseño puede desacoplarse del aquí realizado y resolverse de forma independiente.

2.2.2. Minimización de la potencia

En los diseños clásicos de sistemas de comunicaciones las prestaciones a optimizar habitualmente han sido la tasa de transmisión o en su defecto la probabilidad de error. Tradicionalmente se había considerado que la SNR era suficientemente elevada como para que con una potencia dada se pudiera establecer una comunicación satisfactoria (suficiente) y el objetivo entonces era incrementar al máximo la tasa de transmisión (se hablaba en este caso de *sistemas limitados en tasa*). En esta tesis sin embargo se hace de la minimización de la potencia el objetivo principal, tratando entonces con *sistemas limitados en potencia* asociados a regímenes de baja SNR (la SNR tiene un valor bajo que hace que deba transmitirse la potencia suficiente para que al menos se satisfaga un requisito mínimo de tasa de transmisión). De forma más específica, se considerará la minimización de la potencia transmitida (o de una función de coste relacionada con esta), suponiendo que la complejidad del transmisor estará dada (variará en función del escenario estudiado) y debiendo satisfacerse unos requisitos de tasa de transmisión mínima y probabilidad de error máxima. De esta forma se planteará un problema de

optimización con ligaduras (t.c.c. restricciones) cuya solución podrá encontrarse haciendo uso de un conjunto de técnicas solventes de optimización no lineal que con anterioridad han mostrado su fiabilidad y utilidad en distintos problemas de ingeniería [9, 10, 13]. Nótese que en el ámbito de las comunicaciones el avance de las técnicas de optimización es uno de los factores fundamentales que está influyendo en el abandono gradual de las estrategias de diseño independiente por capas y, como modestamente se hace en esta disertación, en la utilización de técnicas basadas en el principio de *cross-layer design*.

Para el caso de un único usuario (con un único flujo de información), si el vector \mathbf{u} contiene las variables sobre las que se optimiza (variables de diseño) y $p(\mathbf{u})$, $r(\mathbf{u})$ y $\epsilon(\mathbf{u})$ expresan respectivamente la potencia transmitida, tasa transmitida y BER recibida como funciones de \mathbf{u} (cuya forma dependerá naturalmente del sistema específico considerado), el problema de optimización genérico tendrá la forma

$$\left\{ \begin{array}{l} \underset{\mathbf{u}}{\text{mín}} \quad J(p(\mathbf{u})) \\ \text{s. a : C1. } \quad r(\mathbf{u}) \geq \check{r} \\ \quad \quad \quad \text{C2. } \quad \epsilon(\mathbf{u}) \leq \check{\epsilon} \\ \quad \quad \quad \text{C3. } \quad \mathbf{t}(\mathbf{u}) \leq \check{\mathbf{t}}, \end{array} \right. \quad (2.16)$$

donde $J(p)$ es el coste objetivo que será en general una función creciente (lineal o convexa) c.r.a. la potencia transmitida, \check{r} representa el requisito mínimo de tasa transmitida, $\check{\epsilon}$ es el requisito máximo de BER y $\mathbf{t}(\mathbf{u}) \leq \check{\mathbf{t}}$ representa las restricciones de diseño intrínsecas a las características propias al sistema de comunicaciones considerado. Nótese además que para determinados casos no todas las restricciones C1-C3 tienen que escribirse de forma explícita en el problema de optimización sino algunas de ellas se podrán tener en cuenta a través de otros mecanismos (e.g., variables auxiliares, restricción del conjunto de búsqueda, etc.). Por otro lado, dependiendo del problema considerado $p(\mathbf{u})$, $r(\mathbf{u})$ y $\epsilon(\mathbf{u})$ podrán representar cantidades promediadas o instantáneas. Finalmente, \mathbf{u}^* constituirá la solución óptima del problema que, satisfaciendo los requisitos de QoS (C1 y C2) y las restricciones

naturales del problema (C3), minimiza el coste J .

Es conveniente asimismo señalar que la minimización de la potencia transmitida no resulta de interés únicamente en sistemas de baja SNR. La reducción de la energía radiada por un sistema de comunicación inalámbrico conlleva otros efectos que resultan beneficiosos para el establecimiento de comunicaciones: en primer lugar disminuye el consumo de la batería (prolongando así el tiempo de vida de cada terminal y, por ende, el de la red de comunicación) y en segundo lugar reduce la intensidad de la señal radiada (disminuyendo la potencial interferencia para otros usuarios que compartan el medio inalámbrico). La relevancia cada vez mayor de estos aspectos queda recogida en los estándares de comunicaciones inalámbricas que han visto la luz en los últimos años. De esta forma algunas de estas nuevas tecnologías, como las redes de sensores, consideran al ahorro energético como uno de los criterios fundamentales de diseño³ [2]; por otro lado, cada vez más estándares de comunicaciones contemplan redes auto-configurables (*ad-hoc*) que utilizan un ancho de banda compartido y que incluso reconocen el espectro en búsqueda de frecuencias poco utilizadas [1, 4, 31].

2.3. Sistemas adaptativos: información del estado del canal

El *conocimiento del comportamiento del canal* (forma alternativa de referirse al CSI) por parte de un sistema de comunicaciones puede ser aprovechado para mejorar las prestaciones del mismo. Tómese como ejemplo un canal móvil SISO como el descrito en (2.12), con un único coeficiente ($Q = 0$) y energía normalizada (i.e., $\sqrt{E_s} = \lambda\sqrt{T_s} = 1$), entonces la relación entrada-salida puede escribirse como $y(n) = h(n)x(n) + v(n)$ con $x(n)$ representando el símbolo transmitido en el instan-

³Para sistemas en los que el ahorro energético sea el objetivo principal, debe tenerse en cuenta que pueden implantarse mecanismos de reducción de consumo en tareas distintas de la transmisión de información (e.g., encaminamiento, acceso al canal o descubrimiento de terminales). Pese a ello, diversos estudios muestran que la potencia radiada así como los efectos relacionados de forma lineal con esta tienen un elevado impacto en el consumo del transmisor [81] y certifican este criterio como uno de los más importantes para prolongar el tiempo de vida del terminal.

te n , $h(n)$ la realización del canal y $v(n)$ el ruido AWGN. Si el receptor desconoce el valor del canal $h(n)$ (demodulación no coherente) para recuperar $x(n)$ a partir de $y(n)$ dos serán las fuentes de incertidumbre: el valor del canal y la componente de ruido; no obstante si el canal $h(n)$ es conocido por el receptor (demodulación coherente), la recuperación de $x(n)$ a partir de la señal $y(n)$ se vuelve más sencilla y fiable, puesto que la única fuente de incertidumbre se debe a la componente de ruido en $v(n)$. De esta forma la utilización del CSI por parte del receptor permite una mejora considerable en las prestaciones (reducción de la probabilidad de error) de los sistemas inalámbricos [72].

En vista de los potenciales beneficios, los sistemas de comunicaciones implementan mecanismos que permitan el conocimiento del canal. Las técnicas de obtención del canal por parte del receptor son variadas (e.g., uso de pilotos o estimación ciega, etc.) y también lo son sus prestaciones (precisión, fiabilidad). Estas técnicas han sido estudiadas extensamente (véanse, e.g., [26, 34] y referencias) y no son el objeto de este trabajo, por lo que a lo largo de esta disertación no se analizarán los efectos asociados a la obtención del canal por parte del receptor y simplemente se supondrá que el receptor conoce el canal de forma perfecta.

Si los efectos positivos asociados al uso del CSI en el receptor son conocidos y su utilización es generalizada, el uso del CSI por parte del transmisor aunque conocido también hace largo tiempo (el trabajo pionero más significativo del ámbito vio la luz en 1968 de la mano de J. Hayes [30]), está menos extendido. La causa principal de este hecho es que la adaptación del transmisor(es) al CSI requiere transmisores y receptores más complejos. Una de las contribuciones de esta tesis será mostrar que pueden derivarse sencillos esquemas de adaptación que sin pérdida de optimalidad (s.p.d.o.) no requieren ni la utilización de transmisores con gran complejidad hardware ni la presencia de grandes unidades de cómputo en el receptor, pudiendo mejorar así de forma significativa las prestaciones de comunicación sobre canales con desvanecimientos de los sistemas inalámbricos contemporáneos.

El procedimiento de adquisición del CSI por parte del receptor es diferente del requerido por el transmisor. Mientras el receptor tiene acceso a la señal recibida

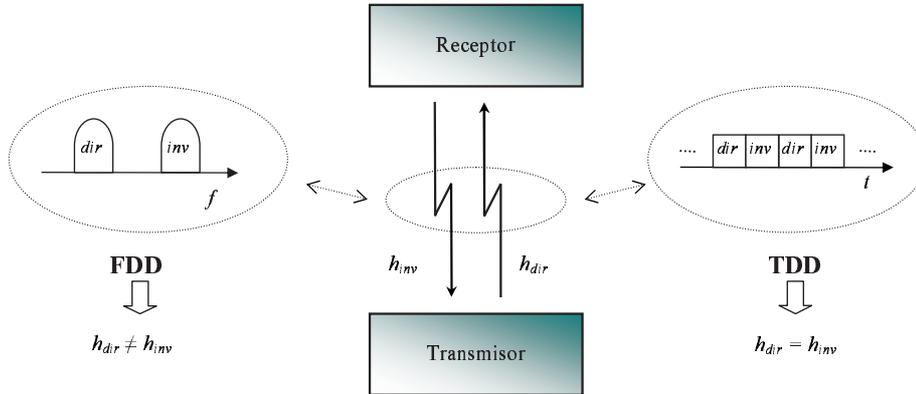


Figura 2.2: Relación entre los canales directo e inverso para sistemas FDD y TDD.

$y(n)$ que de forma natural contiene el efecto del canal $h(n)$, el transmisor no tiene a priori una magnitud a través de la que pueda estimar el canal. La obtención de CSI por parte del transmisor depende básicamente de cómo se separan los canales directo (i.e., el canal entre el transmisor y el receptor) e inverso (i.e., el canal entre el receptor y el transmisor). Para sistemas “duplexados” por división en frecuencia (FDD) el canal directo e inverso utilizan distintas frecuencias portadoras mientras que en sistemas “duplexados” por división en el tiempo (TDD) el canal directo e inverso utilizan la misma frecuencia portadora compartiendo el medio alternando su transmisión en el dominio del tiempo. Puesto que la separación frecuencial en sistemas FDD es significativamente mayor que BW_c los desvanecimientos que experimentan ambos canales están incorrelados, por el contrario, típicamente la separación temporal en sistemas TDD es significativamente menor que T_c lo que implica que los desvanecimientos que experimentan el canal directo e inverso son semejantes (ver figura 2.2). Esto implica que en sistemas TDD el transmisor tenga acceso perfecto al CSI del canal directo aprovechando la estimación natural de su canal inverso; sin embargo, para sistemas FDD la estimación a través del canal inverso no es posible teniendo que ser el receptor el que envíe al transmisor el CSI asociado al canal directo.

El envío del CSI por parte del receptor al transmisor requiere la existencia de

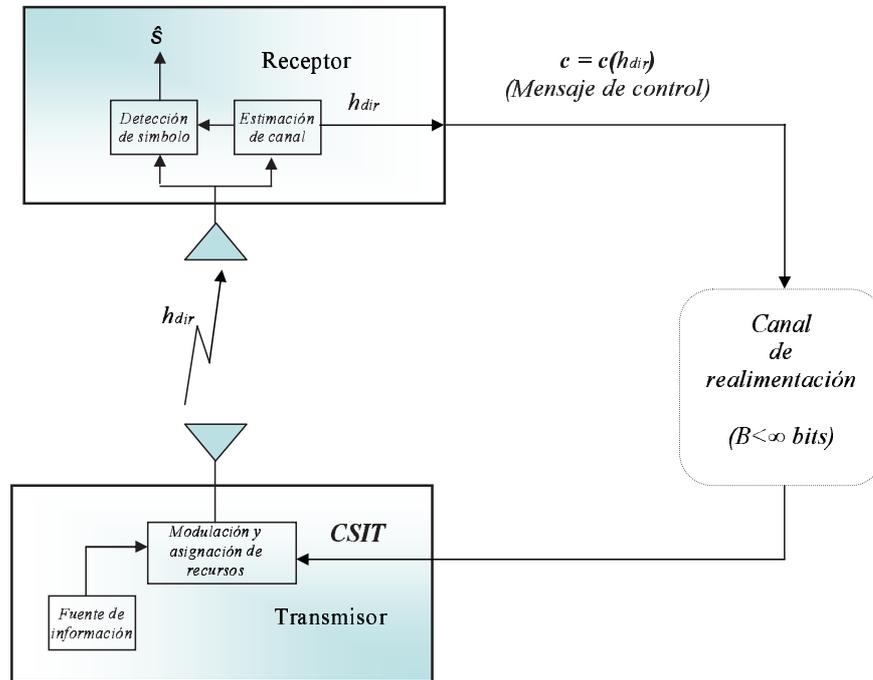


Figura 2.3: Adquisición del CSIT a través de un canal de realimentación.

un canal de realimentación⁴. Los sistemas reales típicamente contemplan este tipo de canales para el envío de señales de control (ver figura 2.3). Habitualmente estos canales presentan una baja tasa de transmisión y una alta fiabilidad (en la práctica pueden considerarse libres de error).

Puesto que la mayoría de sistemas de comunicación utilizan distintas frecuencias de portadora para los canales directo e inverso, la adquisición del CSI en el transmisor estará íntimamente relacionada con las características del canal de realimentación.

2.3.1. Tipos de CSIT

Existen varios criterios para categorizar el CSIT. A efectos de fijar la estructura de adaptación del transmisor pueden distinguirse tres tipos básicos de CSIT diferentes que se enumeran a continuación:

⁴El uso de sistemas realimentados está ampliamente extendido en sistemas electrónicos, de comunicaciones y de control.

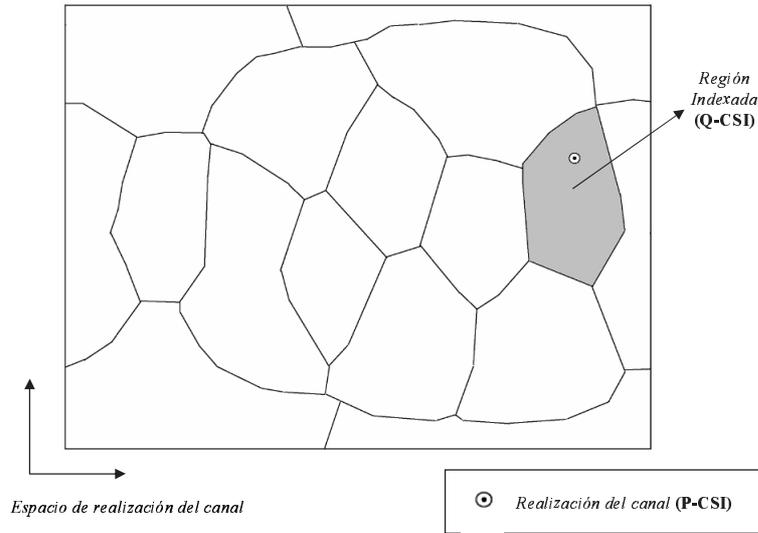


Figura 2.4: Generación del Q-CSI en base a P-CSI para el caso de $B = 4$ bits de realimentación que describen $N = 2^B = 16$ regiones de cuantificación.

- CSIT perfecto (P-CSIT). Este tipo de CSIT corresponde al caso en el que el transmisor conoce de forma instantánea y con un error despreciable cada realización del canal. Como se ha señalado esta suposición puede ser realista para: (i) sistemas TDD o (ii) sistemas FDD en los que la variabilidad temporal del canal sea tan lenta que permita que la limitación de tasa del canal de realimentación no suponga un impedimento para enviar el valor analógico del canal $h(n)$.
- CSIT cuantificado (Q-CSIT). Corresponde al caso en el que el transmisor actualiza de forma instantánea CSIT pero no tiene acceso al valor exacto del canal $h(n)$ sino sólo a la región a la que el canal pertenece. Este categoría se ajusta al caso de sistemas FDD. Para ilustrar la aplicabilidad de este tipo de CSIT considérense el sistema FDD de la figura 2.3 en el que el canal de control tiene una tasa de transmisión máxima de B bits por tiempo de símbolo, está claro entonces que dividiendo el rango de variación del canal en $N = 2^B$ regiones diferentes, para cada símbolo el receptor puede enviar al transmisor el índice de la región a la que la realización del canal pertenece (véase la figura 2.4). Se comprueba de esta manera cómo la limitación de ta-

sa del canal de control conlleva de forma natural a la suposición de Q-CSIT. Por otro lado la instantaneidad del Q-CSIT se garantiza si el canal presenta una variabilidad limitada, lo que en la práctica se reduce a garantizar que el canal se mantenga dentro de la misma región de cuantificación durante el tiempo de transmisión de dos símbolos consecutivos. Es conveniente señalar que debido a su carácter instantáneo los diseños basados en Q-CSIT pueden responder de forma efectiva a variaciones bruscas en el canal debido a efectos inesperados al mismo tiempo que resultan robustos frente a imperfecciones en la estimación del canal por parte del receptor. Estas características hacen que la suposición de Q-CSIT sea también apropiada para sistemas TDD donde la estimación del canal presente un error que no pueda despreciarse.

- CSIT estadístico (S-CSIT). Para canales cuya variabilidad temporal es comparable al tiempo de símbolo, para transmisores con capacidad de adaptación muy limitada o para receptores para los que la estimación del canal resulta costosa –e.g., sistemas no coherentes o sistemas de múltiples entradas y múltiples salidas (MIMO)–, la actualización instantánea del CSIT puede no resultar factible. Incluso para estos escenarios, puede aprovecharse el CSIT para la mejora de prestaciones del sistema mediante la utilización de información estadística del canal inalámbrico. Distintos diseños basados en S-CSIT han capitalizado el conocimiento de los estadísticos de primer orden [101], de los estadísticos de segundo orden [100] o de la totalidad de la *función densidad de probabilidad* (FDP) del canal [50]. La adaptación en base a S-CSIT precisa menos complejidad en los transmisores (las estadísticas del sistema permanecen constantes a corto y medio plazo) y no requiere necesariamente la realimentación por parte del receptor (la caracterización estadística de canal depende en muchas ocasiones de fenómenos físicos tales como la geometría o frecuencia de transmisión a los que el transmisor puede acceder de forma natural).

Nótese que aunque esta sea la clasificación propuesta, en base a estas clases básicas, y dependiendo de la estructura del sistema en cuestión, pueden definirse nuevos tipos de CSIT.

Capítulo 3

Minimización de potencia en sistemas de comunicaciones adaptativos con una única fuente de información

Este capítulo considera el diseño de estrategias de adaptación en el transmisor en base a CSI que permitan minimizar la potencia transmitida por sistemas en los que existe una única fuente de información. La primera sección del capítulo se dedica al análisis del caso de una única fuente de información cuyo flujo de datos se transmite sobre canales móviles con múltiples entradas y una única salida (MISO), correspondiente a una red inalámbrica de sensores (WSN) con múltiples transmisores y un único receptor. En la segunda de las secciones se investigará la minimización de potencia de un sistema OFDM para comunicaciones de un único usuario transmitiendo sobre canales doblemente (i.e., canales que debidos al multi-trayecto son selectivos en la frecuencia y que también son selectivos en el tiempo). Unos breves comentarios que reflexionan respecto a los resultados obtenidos y una listado de demostraciones contenidas en los apéndices constituyen las secciones que cierran este capítulo.

3.1. Minimización de potencia en WSN con QoS utilizando CSIT

Una WSN comprende un número considerable de dispositivos procesadores de señal distribuidos espacialmente (nodos sensores), cada uno de ellos con baterías no recargables (o de difícil recarga) y, por consiguiente, con limitaciones en sus capacidades de computación y comunicación. Cuando se programan y se coordinan de forma adecuada, los nodos en una WSN pueden cooperar para llevar a cabo avanzadas tareas de procesamiento de señal con una robustez y versatilidad sorprendentes. Esto hace de las WSN una atractiva tecnología de bajo coste para un amplio abanico de aplicaciones relacionadas con la estimación y monitorización remota [5]. Como ya se mencionó en el capítulo introductorio, uno de los objetivos principales en la investigación actual en torno a las WSN es el diseño de dispositivos y algoritmos que lleven a cabo diversas tareas del proceso de transmisión de la información a lo largo de la red entrañando un bajo consumo energético [11]. Distintos algoritmos eficientes desde el punto de vista energético han sido propuestos para aspectos tales como: cobertura de red, acceso a medio, estimación descentralizada y encaminamiento (véase, e.g., [11], [98], [95] y [6]). La estructura de la WSN en muchos de estos trabajos incluye un centro de fusión de información o datos (CFI) con el que los sensores están directamente conectados. Por su versatilidad y la riqueza del modelo matemático que la describe [7], así como por estar ampliamente extendido su uso, esta será también la estructura de WSN que se analizará en esta sección.

El escenario propuesto se modelará como un sistema de comunicaciones MISO en el que las múltiples entradas están distribuidas espacialmente y donde las señales transmitidas por los sensores llegan al CFI de forma coherente¹ [7,8]. Utilizando este modelo, se minimizará la potencia media transmitida por el conjunto de sensores de manera que se satisfagan requisitos de QoS en términos de BER y

¹Para satisfacer los requisitos que la suposición de recepción coherente supone, la red puede utilizar algoritmos de sincronización de baja complejidad como, e.g., el propuesto en [49].

tasa media. Para ello se derivarán esquemas de adaptación de los sensores en base a CSIT en tres situaciones diferentes:

- (I) CSIT *pleno* (F-, debido al término inglés *full*), donde cada transmisor tiene conocimiento perfecto de la realización del canal MISO (i.e., P-CSIT tanto la realización propia como de la del resto de sensores);
- (II) Q-CSIT, donde los sensores únicamente disponen de conocimiento cuantificado sobre la realización del canal MISO; y
- (III) CSIT *individual* (I-), donde cada sensor tiene conocimiento perfecto de su propio enlace con el CFI pero únicamente información cuantificada sobre las realizaciones de los canales del resto de sensores.

En todos estos casos, se derivará las estrategias óptimas de adaptación en términos de modulación y codificación, asignación de potencia y conformación de haz así como cuantificadores de canal óptimos que permitan la presencia de Q-CSIT en los sensores.

El resto de la sección se organiza como sigue: tras comenzar detallando el modelo del sistema y los supuestos de trabajo más relevantes, se derivarán los esquemas de adaptación óptimos basados en F-CSIT. Resuelto este aspecto se pasará a la obtención de los esquemas de adaptación para los casos de Q-CSIT e I-CSIT. La validación de los resultados derivados y la comparación de prestaciones entre los diversos esquemas se realizará en base a simulaciones numéricas que junto a la recopilación de resultados más significativos cerrarán esta sección.

3.1.1. Modelo y preámbulos

Se considerará una WSN donde M sensores con índices $m \in \{1, \dots, M\}$ desean enviar información compartida al CFI (e.g., el valor de una VA a la que hacen seguimiento o datos provenientes de otra fuente que ellos reencaminan) conforme a lo que se refleja en la figura 3.1. Se supondrá que:

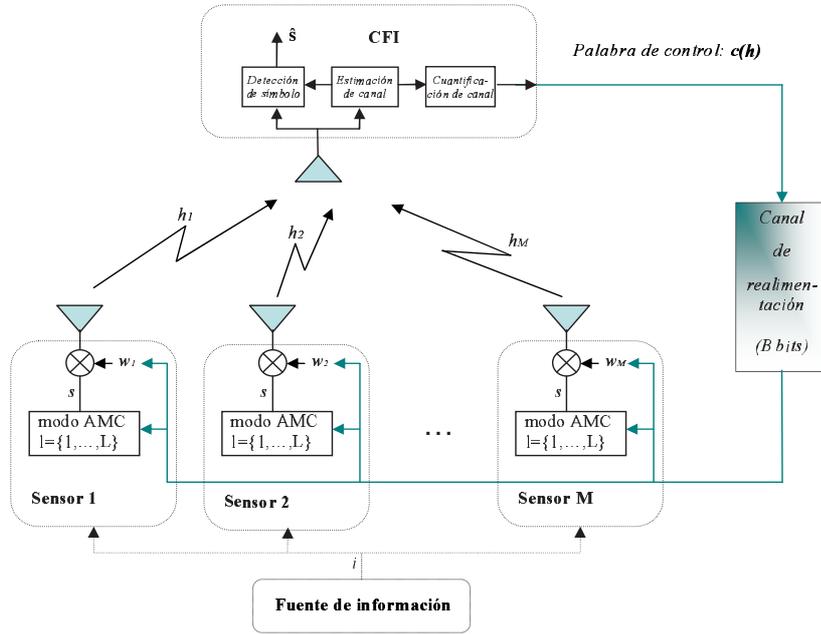


Figura 3.1: Modelo del sistema WSN.

(sp1) la información compartida es común (consistente) en todos los sensores y se recibe de forma coherente en el CFI.

Con $\{h_m\}_{m=1}^M$ representando los coeficientes del canal entre los sensores y el CFI, se supondrá también que:

(sp2) $\{h_m\}_{m=1}^M$ son independientes² e idénticamente distribuidos (i.i.d.) de acuerdo a una distribución compleja gaussiana con media cero y varianza unitaria³, i.e., $h_m \sim \mathcal{CN}(0, 1)$; siendo el proceso aleatorio que describe h_m ergódico;

(sp3) el CFI a través de un canal de control envía a los sensores B bits por cada realización del canal describiendo el CSI. Esta información se recibe en los transmisores libre de error y con un retardo despreciable⁴.

Dado un conjunto de modos de modulación y codificación adaptativa (AMC), suponemos que cada sensor puede implementar un número finito de L modos AMC

²La extensión a canales correlados es posible pero sobrepasa los objetivos de esta disertación.

³Nótese que esta suposición significa que la amplitud de cada uno de los canales sigue una distribución *Rayleigh*.

⁴Estas suposiciones pueden ser fácilmente garantizadas debido a la tasa típicamente baja de los canales de control y al uso de códigos de control de error suficientemente robustos.

indexados mediante $l \in \{1, \dots, L\}$, compactando cada modo una modulación cuya constelación asociada tiene tamaño M_l^r y una codificación de tasa asociada r_c dando lugar a una tasa total de transmisión de $r_l := r_c \log_2(M_l^r)$. A fin de garantizar QoS en la información transmitida, las tasas $\{r_l\}_{l=1}^L$ deben ser recibidas con una BER menor que un nivel $\check{\epsilon}$ pre-establecido. Para mitigar los efectos debidos a los desvanecimientos, los sensores adaptan la amplitud y fase del símbolo a transmitir adaptando así la potencia y la dirección de apuntamiento (conformación de haz) del mismo. Puesto que un sistema MISO tiene ganancia de “multiplexación” uno (véase, e.g., [24, pp. 48]), los sensores transmiten un único símbolo s por cada uso del canal utilizando un modo AMC común. Conforme a este modo AMC, el m -ésimo sensor transmite s multiplicado por un peso complejo (coeficiente de conformación de haz) w_m . Sea $\mathbf{w} := [w_1, \dots, w_M]^T$ el vector que denota la conformación distribuida de haz y el vector $\mathbf{h} := [h_1, \dots, h_M]^T$ la representación del canal MISO. El símbolo recibido por el CFI y puede expresarse como

$$y = \mathbf{w}^T \mathbf{h} s + v := \|\mathbf{w}\|^2 \mathbf{u}^T \mathbf{h} s + v \quad (3.1)$$

donde $\mathbf{u} := \mathbf{w}/\|\mathbf{w}\|$, y v denota el ruido (AWGN) con media cero y varianza N_0 . Téngase en cuenta que, como ya se ha adelantado, a través de la adaptación de la fase y el módulo de \mathbf{w} puede ajustarse no sólo la dirección de apuntamiento del haz de transmisión sino también la potencia transmitida para cada realización del canal \mathbf{h} . Si E_s denota la energía media por símbolo, la potencia transmitida y la relación señal a ruido (SNR) recibida pueden escribirse como

$$p := \|\mathbf{w}\|^2 E_s = \|\mathbf{w}\|^2, \quad (3.2)$$

$$\gamma := |\mathbf{w}^T \mathbf{h}|^2 \frac{E_s}{N_0} = p |\mathbf{u}^T \mathbf{h}|^2, \quad (3.3)$$

donde para obtener la última igualdad en las expresiones (3.2) y (3.3) se ha supuesto sin pérdida de generalidad (s.p.d.g.) que $E_s = N_0 = 1$. Inspeccionando (3.3) se comprueba cómo una vez fijada la dirección de apuntamiento del haz, el

canal vectorial MISO \mathbf{h} en (3.1) queda completamente caracterizado por un canal escalar SISO con ganancia⁵ $g := |\mathbf{u}^T \mathbf{h}|^2$. La SNR recibida por el sistema SISO equivalente puede reescribirse entonces como $\gamma = pg$.

Sea $\mathbf{c} = \mathbf{c}(\mathbf{h})$ la palabra (código) de longitud B bits que representa el Q-CSIT que el CFI realimenta a los sensores de acuerdo a lo expuesto en (sp3). Basándose en $\mathbf{c}(\mathbf{h})$, los sensores adaptan sus parámetros de transmisión a una de las posibles $N = 2^B$ configuraciones pre-diseñadas que especifican la tasa de transmisión (modo AMC) $r = r(\mathbf{c})$, la potencia transmitida $p = p(\mathbf{c})$ y el vector de apuntamiento de haz $\mathbf{u} = \mathbf{u}(\mathbf{c})$.

El objetivo consiste en diseñar de forma óptima el cuantificador de canal que genera el vector $\mathbf{c}(\mathbf{h})$ en base al cual se adaptan $r = r(\mathbf{c})$, $p = p(\mathbf{c})$, y $\mathbf{u} = \mathbf{u}(\mathbf{c})$, de manera que se minimice la potencia media transmitida por la totalidad de los sensores satisfaciendo requisitos de BER y tasa de transmisión media. A fin de llevar a cabo esta tarea, se comenzará encontrando la adaptación óptima en base a F-CSIT, situación que corresponde a considerar $B = \infty$ en (sp3).

3.1.2. Adaptación óptima basada en F-CSIT

En este apartado, se derivarán esquemas (políticas) óptimos de adaptación por parte de los transmisores basados en F-CSIT con la finalidad de obtener intuición y niveles de referencia⁶ para el diseño de los esquemas de adaptación basados en Q-CSIT. De esta manera, dado \mathbf{h} , se desea adaptar la potencia $p(\mathbf{h})$, tasa $r(\mathbf{h})$ y el vector de apuntamiento de haz $\mathbf{u}(\mathbf{h})$ para minimizar la potencia media transmitida sujeta a requisitos de BER máxima ($\check{\epsilon}$) y tasa de transmisión media mínima (\check{r}). Como se mostrará a continuación, la adaptación del apuntamiento de haz puede realizarse de forma independiente a la adaptación de la potencia y tasa transmitida sin pérdida de optimalidad (s.p.d.o.). Esto permite dividir la resolución

⁵Salvo que se indique expresamente lo contrario, a lo largo de esta tesis la ganancia de canal se definirá siempre en términos de potencia (i.e. cuadráticos).

⁶El consumo de potencia cuando la adaptación de los transmisores se realiza en base a F-CSIT es claramente una cota inferior para la adaptación en base a Q-CSIT, puesto que el Q-CSIT tiende asintóticamente a F-CSIT si B tiende a infinito; i.e., $\mathbf{c}(\mathbf{h}) = \mathbf{h}$ si $B = \infty$.

del problema original en dos fases separadas: en primer lugar se derivará el esquema óptimo para la adaptación del apuntamiento de haz $\mathbf{u}^*(\mathbf{h})$; resuelto el primer problema se introducirá en el problema original la expresión obtenida para $\mathbf{u}^*(\mathbf{h})$ y se derivarán las expresiones para la adaptación óptima de potencia $p^*(\mathbf{h})$ y tasa $r^*(\mathbf{h})$ de transmisión.

Apuntamiento de haz distribuido

Observando (3.2) y (3.3) puede concluirse que la selección de \mathbf{u} afecta a la ganancia escalar g . Puesto que para cualquier modo AMC la potencia de transmisión requerida $p = p(\mathbf{h})$ es monótonamente decreciente c.r.a. g (para cualesquiera \check{r} y $\check{\epsilon}$), a fin de minimizar la potencia transmitida, para cada realización del canal \mathbf{h} , el vector $\mathbf{u} = \mathbf{u}(\mathbf{h})$ debe adaptarse de manera que $g = |\mathbf{u}^T(\mathbf{h}) \mathbf{h}|^2$ sea máximo. El vector unitario de orientación de haz que maximiza $|\mathbf{u}^T(\mathbf{h}) \mathbf{h}|^2$ y por consiguiente minimiza la potencia requerida, claramente es

$$\mathbf{u}^*(\mathbf{h}) = \mathbf{h}^\dagger / \|\mathbf{h}\|, \quad (3.4)$$

que depende únicamente de la fase del canal MISO, i.e., $\mathbf{u}^*(\mathbf{h}) = \mathbf{u}^*(\mathbf{h}/\|\mathbf{h}\|)$.

Para poder atacar el problema de adaptación óptima de potencia y tasa es necesario primero caracterizar de forma estadística el comportamiento del canal en (3.1) cuando el apuntamiento de haz se realiza según (3.4). Si $\mathbf{u}^*(\mathbf{h}) = \mathbf{h}^\dagger / \|\mathbf{h}\|$, $\forall \mathbf{h}$, la ganancia escalar del canal SISO equivalente es

$$g = |\mathbf{u}^T(\mathbf{h}) \mathbf{h}|^2 = \|\mathbf{h}\|^2. \quad (3.5)$$

Conforme a lo expresado en (sp2), g sigue entonces una distribución chi-cuadrado con FDP

$$f_g(g) = \frac{g^{M-1} \exp(-g)}{\Gamma(M)} \quad (3.6)$$

donde $\Gamma(b, x) := \int_x^\infty t^{b-1} e^{-t} dt$ es la función *Gamma* incompleta y $\Gamma(b) := \Gamma(b, 0)$. La función de distribución de probabilidad acumulada (FDPA) correspondiente

resulta

$$F_g(g) = \frac{\Gamma(M, g)}{\Gamma(M)}. \quad (3.7)$$

Dada su relevancia en el proceso de diseño, se incide una vez más en que cuando se implementa la adaptación óptima de $\mathbf{u}^*(\mathbf{h})$ descrita en (3.4), el canal MISO en (3.1) se puede caracterizar completamente por un canal SISO equivalente con ganancia escalar g descrita en (3.5). Esto implica que s.p.d.o. encontrar los esquemas de adaptación óptimos $r^*(\mathbf{h})$ y $p^*(\mathbf{h})$ es equivalente a encontrar los esquemas de adaptación óptimos $r^*(g)$ y $p^*(g)$. Téngase en cuenta que puesto que \mathbf{h} (y consecuentemente g) varía de una realización de canal a otra, la tasa de transmisión (a través del modo AMC) y la potencia transmitida serán adaptadas a lo largo del tiempo con el objetivo de minimizar la potencia media transmitida satisfaciendo requisitos de tasa de transmisión media \tilde{r} (como se comprobará en el siguiente apartado, el requisito $\tilde{\epsilon}$ relacionado con la BER será tenido en cuenta de forma automática a través de la relación entre potencia y tasa transmitida).

Adaptación de tasa y potencia de transmisión

Se comienza ordenando los modos AMC de modo que $r_l < r_{l+1} \forall l > 1$ y estableciendo que el primer modo represente una transmisión inactiva con potencia y tasa de transmisión nulas ($r_1 = p_1 = 0$). Si $\epsilon(\cdot)$ denota la función que calcula de forma instantánea (i.e., para cada realización del canal) el valor esperado de la BER, la mínima potencia transmitida que, utilizando el l -ésimo modo AMC, satisface el requisito de BER $\tilde{\epsilon}$ puede ser calculada despejando c.r.a. p_l la ecuación

$$\epsilon(g, p_l, r_l) = \tilde{\epsilon}. \quad (3.8)$$

Para una amplia gama de modos de transmisión, la BER puede ser aproximada de forma precisa como [27]

$$\epsilon(g, p, r) = \kappa_1 \exp(-\kappa_2 gp / (2^r - 1)). \quad (3.9)$$

donde κ_1 y κ_2 son constantes dependientes de los esquemas de modulación y codificación seleccionados.

Observación 3.1 *La expresión en (3.9) modela de forma precisa la BER en modulaciones de amplitud en cuadratura de M^r niveles (M^r -QAM) con $\kappa_1 = 0,2$, $\kappa_2 = 1,5$ y sirve asimismo como una buena aproximación para una amplia variedad de esquemas puesto que acota de forma natural la función $Q(x)$ de Marcum a través de la cota de Chernoff [76]. En lo que al esquema de codificación se refiere, el parámetro $\kappa_2 = 1,5$ puede escalarse fácilmente de modo que se tenga en cuenta la ganancia de codificación del modo AMC considerado. Finalmente, la extensión del trabajo realizado a otras funciones de BER es también posible.*

Sustituyendo (3.9) en (3.8), la potencia requerida por el l -ésimo modo AMC puede ser expresada como

$$p_l(g, r_l, \check{\epsilon}) = \frac{(2^{r_l} - 1) \ln(\kappa_1/\check{\epsilon})}{g \kappa_2}. \quad (3.10)$$

Si el F-CSIT está disponible, (3.10) muestra que la especificación del modo AMC no sólo fija la tasa de transmisión sino también la potencia a transmitir de manera que se satisfaga la BER requerida $\check{\epsilon}$. Además, es sencillo comprobar cómo para p_l y r_l dados, el rango de variación de g puede ser dividido en L intervalos consecutivos $[\tau_l, \tau_{l+1})$ con $\tau_1 = 0$ y $\tau_{L+1} = \infty$, siendo el l -ésimo modo AMC seleccionado si $g \in [\tau_l, \tau_{l+1})$. Recíprocamente, esto significa que una vez los intervalos $[\tau_l, \tau_{l+1})$ están dados $\forall l$, la asignación de tasa y potencia será

$$r(g) = r_l; \quad \text{si } g \in [\tau_l, \tau_{l+1}) \quad (3.11)$$

$$p(g, \check{\epsilon}) = \begin{cases} 0, & g \in [\tau_1, \tau_2) \\ \frac{(2^{r_l} - 1) \ln(\kappa_1/\check{\epsilon})}{g \kappa_2}, & g \in [\tau_l, \tau_{l+1}), l > 1. \end{cases} \quad (3.12)$$

Definiendo $\boldsymbol{\tau} := [\tau_1, \dots, \tau_{L+1}]^T$, (3.11) y (3.12) establecen que para encontrar la asignación óptima de tasa y potencia, únicamente se necesita obtener los um-

brales óptimos $\boldsymbol{\tau}^*$ que son solución del siguiente problema de minimización con restricciones

$$\left\{ \begin{array}{l} \text{mín}_{\boldsymbol{\tau}} \quad \bar{p}, \quad \bar{p} := \sum_{l=1}^L \int_{\tau_l}^{\tau_{l+1}} p_l(g, r_l, \check{\epsilon}) f_g(g) dg \\ \text{s. a :} \quad C1. \quad \sum_{l=1}^L \int_{\tau_l}^{\tau_{l+1}} r_l f_g(g) dg \geq \check{r} \\ \quad \quad \quad C2. \quad \tau_l \leq \tau_{l+1} \quad \forall l \end{array} \right. \quad (3.13)$$

donde la potencia media transmitida \bar{p} en la función objetivo y la tasa de transmisión media en la restricción (ligadura) C1 se calculan como la esperanza de $p(g)$ y $r(g)$ sobre todas las posibles realizaciones de g . Nótese que la introducción del conjunto de ligaduras en C2 asegura la consistencia de los intervalos $[\tau_l, \tau_{l+1})$.

Sea β el multiplicador de Lagrange (positivo o nulo) asociado con la restricción de tasa expresada en C1 y $\boldsymbol{\alpha} := [\alpha_1, \dots, \alpha_L]^T$ los multiplicadores de Lagrange (positivos o nulos) asociados con las L ligaduras en C2. La función dual de Lagrange⁷ para (3.13) puede escribirse como

$$\begin{aligned} \mathcal{L}(\beta, \boldsymbol{\alpha}, \boldsymbol{\tau}) &= \sum_{l=1}^L \int_{\tau_l}^{\tau_{l+1}} p_l(g, r_l, \check{\epsilon}) f_g(g) dg \\ &- \beta \left(\sum_{l=1}^L \int_{\tau_l}^{\tau_{l+1}} r_l f_g(g) dg - \check{r} \right) + \sum_{l=1}^L \alpha_l (\tau_l - \tau_{l+1}). \end{aligned} \quad (3.14)$$

En el óptimo τ_l^* la condición necesaria de Karush-Kuhn-Tucker (KKT) es [13]:

$$\begin{aligned} \frac{\partial \mathcal{L}(\beta^*, \boldsymbol{\alpha}^*, \boldsymbol{\tau}^*)}{\partial \tau_l} &= [p_{l-1}(\tau_l^*, r_{l-1}, \check{\epsilon}) - \beta^* r_{l-1} - p_l(\tau_l^*, r_l, \check{\epsilon}) + \beta^* r_l] f_g(\tau_l^*) \\ &- \alpha_{l-1}^* + \alpha_l^* = 0, \end{aligned} \quad (3.15)$$

que debido a la falta de convexidad de (3.13) es una condición necesaria pero no suficiente para garantizar un óptimo global (i.e., los umbrales puede que representen

⁷A lo largo de la disertación se utilizarán asimismo las expresiones *función dual*, *función de Lagrange* o simplemente *Lagrangiano* para referirse a la *función dual de Lagrange* asociada a un problema de optimización [13].

un mínimo local).

Si todas las ligaduras en $C2$ son débiles, i.e., $\tau_l^* < \tau_{l+1}^*$, entonces $\alpha_l^* = 0$, $\forall l$. En este caso, resolviendo (3.15) c.r.a. τ_l^* se obtiene para $l \in \{2, \dots, L\}$ [cf. (3.10)] que

$$\tau_l^* = \frac{(2^{r_l} - 2^{r_{l-1}}) \ln(\kappa_1/\tilde{\epsilon})}{\beta^*(r_l - r_{l-1}) \kappa_2}. \quad (3.16)$$

La ecuación (3.16) expresa de forma analítica el umbral óptimo τ_l^* como función del multiplicador de Lagrange β^* . Definiendo ahora $f(x) := (2^x) \ln(\kappa_1/\tilde{\epsilon})/(\beta^* \kappa_2)$, la expresión (3.16) puede reescribirse como $\tau_l^* = [f(r_l) - f(r_{l-1})]/(r_l - r_{l-1})$. Puesto que $\partial f(x)/\partial x = 2^x/x^3[\ln^2(2)x^2 - 2\ln(2)x + 2] > 0$, $\forall x > 0$, se comprueba fácilmente que $\tau_l^* < \tau_{l+1}^*$. Por consiguiente, los umbrales solución de (3.16) satisfacen $\tau_l^* < \tau_{l+1}^*$ de forma automática y, por lo tanto, son solución de (3.15).

Con τ_l^* fijado por (3.16), el valor de β^* puede obtenerse satisfaciendo $C1$ a través del siguiente algoritmo.

Algoritmo 3.1 Cuantificación “off-line” (F-CSIT)

(S1.0) Sea ε un pequeño nivel de tolerancia e inicialícese β como un pequeño número positivo aleatorio.

(S1.1) Calcule $\{\tau_l\}_{l=2}^L$ vía (3.16).

(S1.2) Utilizando (3.7), calcule la tasa media como $\bar{r} = \sum_{l=1}^L [F_g(\tau_{l+1}) - F_g(\tau_l)]r_l$, y compruebe la satisfacción de $C1$. Si $|\bar{r} - \check{r}|/\check{r} < \varepsilon$ entonces pare; en otro caso, calcule $\Delta\beta := (\bar{r} - \check{r})c$, actualice el multiplicador en la forma $\beta = \beta + \Delta\beta$, y diríjase al paso (S1.1). El parámetro c que interviene en el cálculo de $\Delta\beta$ es un parámetro de penalización adaptativo que puede ser ajustado (en cada iteración) en función de las necesidades de convergencia⁸.

⁸El método de actualización de β aquí propuesto simplemente realiza una búsqueda por subgradiente basada en el método de los multiplicadores [9].

Una vez que β^* se ha obtenido utilizando el algoritmo 3.1⁹, los valores óptimos $\{\tau_l^*\}_{l=2}^L$ y, por consiguiente, la tasa y potencia óptimas quedan fijadas tras sustituir (3.16) en (3.11) y (3.12).

Realimentación y adaptación de los transmisores en tiempo real

Obtenidos $\{\tau_l^*\}_{l=2}^L$, el siguiente algoritmo resume la asignación de recursos (adaptación de los parámetros de transmisión) que la WSN tiene que ejecutar “on-line” (en tiempo real) para cada realización del canal:

Algoritmo 3.2 Adaptación “on-line” (F-CSIT)

Para cada realización del canal MISO \mathbf{h} :

- (S2.1) *El CFI encuentra el índice $l^*(\mathbf{h}) = l^*(g)$ a partir del intervalo $[\tau_l^*, \tau_{l+1}^*)$ al que la ganancia escalar g pertenece, y envía a los sensores la palabra $\mathbf{c}(\mathbf{h}) = [l^*(\mathbf{h}); \mathbf{h}]$ que contiene el F-CSIT.*
- (S2.2) *Obteniendo los valores de $l^*(\mathbf{h})$ y \mathbf{h} a través de la palabra \mathbf{c} , cada sensor m transmite utilizando el $l^*(\mathbf{h})$ -ésimo modo AMC y el coeficiente óptimo [cf. (3.4) y (3.10)] $w_m^* = \sqrt{p_{l^*(\mathbf{h})}}(g, r_{l^*(\mathbf{h})}, \check{\epsilon}) h_m^\dagger / \|\mathbf{h}\|$.*
-

Nótese cada sensor podría calcular $l^*(\mathbf{h})$ utilizando únicamente la parte de $\mathbf{c}(\mathbf{h})$ correspondiente a \mathbf{h} . Sin embargo, por robustez también se incluye $l^*(\mathbf{h})$ en el mensaje de control. De esta forma, la estructura propuesta para $\mathbf{c}(\mathbf{h})$ disminuye la carga computacional en cada sensor y evita asimismo el envío de $\boldsymbol{\tau}^*$ a los sensores durante la fase inicialización (recuérdese que se requiere el conocimiento de $\boldsymbol{\tau}^*$ para poder determinar la selección óptima del modo AMC y de la asignación de potencia). Con la comprensión e intuición obtenidas gracias al estudio de los esquemas de adaptación basados en F-CSIT, en el siguiente apartado se derivarán los esquemas óptimos de adaptación y cuantificación cuando los sensores disponen únicamente de Q-CSIT.

⁹Recuérdese que el algoritmo se ejecuta de forma “off-line”, i.e. se calcula en base al comportamiento estadístico del canal y no es necesario re-calcularlo en tiempo real para cada realización del mismo.

3.1.3. Adaptación óptima basada en Q-CSIT

En este apartado se derivan esquemas de adaptación y cuantificación eficientes energéticamente cuando el CFI envía Q-CSIT a los sensores. Para cada realización del canal, el CFI cuantifica \mathbf{h} para obtener de forma separada el \mathbf{u} óptimo de acuerdo al cuantificador $Q_u(\cdot)$, y el modo AMC y p óptimos de acuerdo a un cuantificador diferente $Q_l(\cdot)$. Si $\mathbf{c}_u = Q_u(\mathbf{h})$ es el índice del vector de apuntamiento y $\mathbf{c}_l = Q_l(\mathbf{h})$ es el índice del modo de transmisión, el CFI enviará a los usuarios el Q-CSIT a través de la palabra de B bits \mathbf{c} de manera que $\mathbf{c} = [\mathbf{c}_u; \mathbf{c}_l]$, donde $B = B_u + B_t$, con $B_u := \text{length}(\mathbf{c}_u)$ y $B_t := \text{length}(\mathbf{c}_l)$. En base a la palabra \mathbf{c} , los sensores adaptan su modo de transmisión y coeficientes de conformación para minimizar la potencia media total transmitida.

Apuntamiento de haz distribuido

Con sólo B_u bits disponibles, el vector de apuntamiento \mathbf{u} se escoge de un conjunto finito¹⁰ $\mathcal{U} := \{\mathbf{u}_i\}_{i=1}^{N_u}$, donde $N_u = 2^{B_u}$. Al igual que para el caso de F-CSIT, a fin de minimizar la potencia transmitida media, el $\mathbf{u} \in \mathcal{U}$ óptimo debe maximizar la ganancia escalar del canal SISO equivalente en (3.3); i.e.,

$$\mathbf{u}^*(\mathbf{h}) = \arg \max_{\mathbf{u} \in \mathcal{U}} |\mathbf{u}^T \mathbf{h}|^2. \quad (3.17)$$

Para sistemas MISO *empotrados* (t.c.c. *colocados* y que son aquellos donde las múltiples entradas no corresponden a transmisores distribuidos sino a un único transmisor con múltiples antenas) existen en la literatura distintos ejemplos de optimización de constelaciones \mathcal{U} con distintos objetivos [66], [46], [92]. Bajo distintos criterios, las constelaciones óptimas minimizan la correlación máxima entre las palabras que contienen. Utilizando un criterio grassmaniano de compactación de líneas, [46] demuestra además que la minimización de la correlación máxima equivale a la maximización de la distancia cordal mínima, distancia que para dos

¹⁰El conjunto finito de elementos que contenga el representante de todas y cada una de las regiones de cuantificación recibirá el nombre de constelación-código.

vectores complejos unitarios \mathbf{a} y \mathbf{b} se define como [15]

$$d_{ch}(\mathbf{a}, \mathbf{b}) := (1 - |\mathbf{a}^T \mathbf{b}|^2)^{\frac{1}{2}}. \quad (3.18)$$

Dada (3.18), la optimización en (3.17) puede expresarse como

$$\mathbf{u}^*(\mathbf{h}) = \arg \min_{\mathbf{u} \in \mathcal{U}} d_{ch} \left(\mathbf{u}, \frac{\mathbf{h}}{\|\mathbf{h}\|} \right), \quad (3.19)$$

y la constelación óptima \mathcal{U}^* como

$$\mathcal{U}^* = \max_{\{\mathbf{u}_i\}_{i=1}^{N_u}} \min_{\forall i \neq j} d_{ch}(\mathbf{u}_i, \mathbf{u}_j). \quad (3.20)$$

Para distintos valores (tamaños) de M y N_u se encuentran disponibles soluciones numéricas de (3.20) (consúltese, e.g., [43]). Con la constelación óptima de vectores \mathcal{U}^* disponible tanto en el CFI como en los sensores, el Q-CSI y el vector óptimo de apuntamiento se calculan de la forma

$$\mathbf{c}_u = Q_u(\mathbf{h}) := \arg_i \min_{\mathbf{u}_i \in \mathcal{U}^*} \left\{ d_{ch} \left(\mathbf{u}_i, \frac{\mathbf{h}}{\|\mathbf{h}\|} \right) \right\}, \quad (3.21)$$

$$\mathbf{u}^*(\mathbf{h}) = \mathbf{u}_{\mathbf{c}_u}. \quad (3.22)$$

Observación 3.2 *El apuntamiento de haz especificado por (3.21) y (3.22) es óptimo si $N_u \geq M$. Para un análisis detallado de apuntamiento de haz adaptativo basado en Q-CSIT cuando $N_u < M$, puede consultarse [92].*

Trasládese ahora la atención a la caracterización estadística de la ganancia del canal SISO equivalente para el caso de Q-CSIT. Mientras que para el caso de F-CSIT el apuntamiento de haz se realiza de acuerdo a (3.4) y la ganancia del canal SISO equivalente resulta $g = \|\mathbf{h}\|^2$, el apuntamiento de haz óptimo con Q-CSIT se establece en (3.21)-(3.22), presentando entonces el canal SISO equivalente la ganancia $\tilde{g} := g(1 - z)$, donde

$$z := \min_{\mathbf{u} \in \mathcal{U}^*} d_{ch}^2(\mathbf{u}, \mathbf{h}/\|\mathbf{h}\|) = d_{ch}^2(\mathbf{u}^*(\mathbf{h}), \mathbf{h}/\|\mathbf{h}\|) \quad (3.23)$$

puede interpretarse como una reducción de la ganancia debida a la cuantificación. Esta reducción afecta a la SNR instantánea de forma análoga $\gamma := p\tilde{g}$.

En base a la cota de la unión, la FDPA de z se puede acotar superiormente (de forma ajustada) como [103]

$$F_z(z) \leq \tilde{F}_z(z) = \begin{cases} N_u z^{M-1}, & 0 \leq z \leq z_{\text{máx}} \\ 1, & z \geq z_{\text{máx}} \end{cases} \quad (3.24)$$

donde $z_{\text{máx}} := N_u^{-1/(M-1)}$. Puesto que g y z son independientes [cf. (sp2)], utilizando la aproximación $F_z(z) \simeq \tilde{F}_z(z)$, la FDPA de \tilde{g} puede obtenerse como

$$\begin{aligned} F_{\tilde{g}}(x) &= \Pr\{g(1-z) < x\} = \int_{g=0}^{\frac{x}{1-z_{\text{máx}}}} \int_{z=\text{máx}(0,1-x/g)}^{z_{\text{máx}}} f_z(z) dz f_g(g) dg \\ &= \int_{g=0}^x \int_{z=0}^{z_{\text{máx}}} f_z(z) dz f_g(g) dg + \int_{g=x}^{\frac{x}{1-z_{\text{máx}}}} \int_{z=1-x/g}^{z_{\text{máx}}} f_z(z) dz f_g(g) dg \\ &= \int_{g=0}^x [\tilde{F}_z(z_{\text{máx}}) - \tilde{F}_z(0)] f_g(g) dg + \int_{g=x}^{\frac{x}{1-z_{\text{máx}}}} [\tilde{F}_z(z_{\text{máx}}) - \tilde{F}_z(0)] f_g(g) dg \\ &= 1 - \frac{\Gamma(M, x/(1-z_{\text{máx}}))}{\Gamma(M)} - N \exp(-x) \left(1 - \frac{\Gamma(M, x z_{\text{máx}}/(1-z_{\text{máx}}))}{\Gamma(M)} \right). \end{aligned} \quad (3.25)$$

La FDP de \tilde{g} puede obtenerse calculando $\partial F_{\tilde{g}}(\tilde{g})/\partial \tilde{g}$, resultando

$$\begin{aligned} f_{\tilde{g}}(\tilde{g}) &= \frac{1}{\Gamma(M)} \left\{ \exp\left(\frac{-\tilde{g}}{1-z_{\text{máx}}}\right) \frac{\tilde{g}^{M-1}}{(1-z_{\text{máx}})^M} (1 - N_u z_{\text{máx}}^M) \right. \\ &\quad \left. + N_u \exp(-\tilde{g}) \left[\Gamma(M) - \Gamma(M, z_{\text{máx}}\tilde{g}/(1-z_{\text{máx}})) \right] \right\}. \end{aligned} \quad (3.26)$$

La ganancia del canal SISO equivalente \tilde{g} caracteriza por completo el canal MISO cuando se adopta el apuntamiento de haz óptimo establecido en (3.21)-(3.22). Tomando como base las expresiones cerradas (3.25) y (3.26), se seguirá a continuación un camino similar al seguido para el caso de F-CSIT a fin de derivar analíticamente los esquemas de adaptación de potencia y tasa óptimos en base a Q-CSIT.

Adaptación de tasa y potencia de transmisión

Cuando el canal de realimentación (control) tiene tasa de transmisión limitada, el CFI tiene que cuantificar \tilde{g} utilizando un número finito de regiones. Para afrontar este diseño, un primer paso que surge de forma intuitiva para decidir la estructura del cuantificador consiste en identificar cada región de cuantificación con la selección (uso) de un único modo AMC. Así pues, se considerarán L regiones de cuantificación distintas $\{\mathcal{R}_l := [\tilde{\tau}_l, \tilde{\tau}_{l+1})\}_{l=1}^L$, que, como en el caso de F-CSIT, tendrán asociadas un vector de umbrales $\tilde{\boldsymbol{\tau}} := [\tilde{\tau}_1, \dots, \tilde{\tau}_{L+1}]^T$.

El l -ésimo modo de transmisión está caracterizado por la pareja de tasa y potencia de transmisión (r_l, \tilde{p}_l) para la región de cuantificación \mathcal{R}_l . Puesto que r_l se encuentra preestablecido por cada modo AMC, \tilde{p}_l se seleccionará a fin de satisfacer el requisito de BER. La BER esperada $\tilde{\epsilon}_l$ para la región \mathcal{R}_l puede obtenerse de forma sencilla como el número esperado de bits erróneos dividido por el número esperado de bits transmitido; i.e.,

$$\tilde{\epsilon}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, \tilde{p}_l, r_l) := \mathbb{E}_{\tilde{g} \in [\tilde{\tau}_l, \tilde{\tau}_{l+1})} [r_l \epsilon(\tilde{g}, \tilde{p}_l, r_l)] / \mathbb{E}_{\tilde{g} \in [\tilde{\tau}_l, \tilde{\tau}_{l+1})} [r_l]. \quad (3.27)$$

Con el objetivo de satisfacer el nivel global de BER requerido $\check{\epsilon}$, se fijará

$$\tilde{\epsilon}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, \tilde{p}_l, r_l) = \check{\epsilon} \quad \forall l. \quad (3.28)$$

Es fácil comprobar cómo (3.28) tiende a (3.8) cuando $L \rightarrow \infty$ (F-CSIT). Además, la sustitución de (3.27) en (3.28) produce

$$\varphi_\epsilon(\tilde{\tau}_l, \tilde{\tau}_{l+1}, \tilde{p}_l, r_l, \check{\epsilon}) := \int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} \epsilon(\tilde{g}, \tilde{p}_l, r_l) f_{\tilde{g}}(\tilde{g}) d\tilde{g} - \check{\epsilon} \int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} f_{\tilde{g}}(\tilde{g}) d\tilde{g} = 0. \quad (3.29)$$

Utilizando la expresión analítica de φ_ϵ que se deriva en el apéndice 3.4.1, (3.29) puede resolverse c.r.a. \tilde{p}_l . La misma expresión de φ_ϵ puede utilizarse también para obtener $\tilde{\epsilon}_l$ en forma analítica y, por lo tanto, para cuantificar la BER media para cualesquiera regiones $\{\mathcal{R}_l\}$. La resolución de (3.29) c.r.a. \tilde{p}_l puede ser llevada a

cabo fácilmente utilizando una búsqueda uni-dimensional. Con $\tilde{p}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, r_l, \epsilon)$ denotando ese valor, la asignación de tasa y de potencia $\forall \tilde{g}$ puede expresarse como [cf. (3.11) y (3.12)]

$$\tilde{r}(\tilde{g}) = r_l; \quad \text{si } \tilde{g} \in [\tilde{\tau}_l, \tilde{\tau}_{l+1}) \quad (3.30)$$

$$\tilde{p}(\tilde{g}, \epsilon) = \begin{cases} 0, & \tilde{g} \in [\tilde{\tau}_1, \tilde{\tau}_2) \\ \tilde{p}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, r_l, \epsilon), & \tilde{g} \in [\tilde{\tau}_l, \tilde{\tau}_{l+1}), l > 1. \end{cases} \quad (3.31)$$

Llegados a este punto, al igual que para el caso de F-CSIT, derivar la adaptación óptima de tasa y potencia se reduce a encontrar los umbrales óptimos $\tilde{\tau}^*$. El problema de optimización que, basado en Q-CSIT, calcula estos umbrales es:

$$\left\{ \begin{array}{l} \underset{\tilde{\tau}}{\text{mín}} \quad \bar{p}, \quad \bar{p} := \sum_{l=1}^L \tilde{p}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, r_l, \epsilon) \int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} f_{\tilde{g}}(\tilde{g}) d\tilde{g} \\ \text{s. a : } C1. \quad \sum_{l=1}^L r_l \int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} f_{\tilde{g}}(\tilde{g}) d\tilde{g} \geq \tilde{r} \\ C2. \quad \tilde{\tau}_l \leq \tilde{\tau}_{l+1} \quad \forall l \end{array} \right. \quad (3.32)$$

donde tanto la potencia en la función objetivo como la tasa en $C1$ se promedian sobre todas las regiones consideradas para la VA \tilde{g} (estados de cuantificación). Nótese que, al contrario de lo que ocurría para el problema basado en F-CSIT en (3.13), en este caso la potencia a transmitir no varía según la ganancia del canal varía, sino únicamente según el índice de la región varía (i.e., la asignación de potencia es constante por modo AMC) y, por lo tanto la función potencia puede escribirse fuera del operador integral. De esta forma, $\int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} f_{\tilde{g}}(\tilde{g}) d\tilde{g}$ puede ser fácilmente interpretado bien como la probabilidad de que el canal pertenezca a la l -ésima región de cuantificación o bien como la probabilidad de elegir el l -ésimo modo AMC.

A continuación se utilizarán las condiciones KKT (de nuevo de carácter necesario, pero no suficiente, puesto que (3.32) no es convexo) para encontrar $\tilde{\tau}_l^*$. Sea $\tilde{\beta}$ el multiplicador de Lagrange asociado con la restricción de tasa en $C1$ y

$\tilde{\boldsymbol{\alpha}} := [\tilde{\alpha}_2, \dots, \tilde{\alpha}_L]^T$ los multiplicadores de Lagrange correspondientes a $C2$; al igual que en (3.16), con $\tilde{\boldsymbol{\alpha}} = \mathbf{0}$, la condición KKT para el valor óptimo $\tilde{\tau}_l^*$ exige

$$\begin{aligned} \frac{\partial \mathcal{L}(\tilde{\beta}^*, \tilde{\boldsymbol{\tau}}^*)}{\partial \tau_l} &= \left[\tilde{p}_{l-1}(\tilde{\tau}_{l-1}^*, \tilde{\tau}_l^*, r_{l-1}, \tilde{\epsilon}) - \tilde{\beta}^* r_{l-1} - \tilde{p}_l(\tilde{\tau}_l^*, \tilde{\tau}_{l+1}^*, r_l, \tilde{\epsilon}) + \tilde{\beta}^* r_l \right] f_{\tilde{g}}(\tilde{\tau}_l^*) \\ &+ \frac{\partial \tilde{p}_{l-1}}{\partial \tilde{\tau}_l}(\tilde{\tau}_{l-1}^*, \tilde{\tau}_l^*, r_{l-1}, \tilde{\epsilon}) \int_{\tilde{\tau}_{l-1}^*}^{\tilde{\tau}_l^*} f_{\tilde{g}}(\tilde{g}) d\tilde{g} \\ &+ \frac{\partial \tilde{p}_l}{\partial \tilde{\tau}_l}(\tilde{\tau}_l^*, \tilde{\tau}_{l+1}^*, r_l, \tilde{\epsilon}) \int_{\tilde{\tau}_l^*}^{\tilde{\tau}_{l+1}^*} f_{\tilde{g}}(\tilde{g}) d\tilde{g} = 0. \end{aligned} \quad (3.33)$$

Nótese que utilizando la FDP de (3.25) se tiene que $\int_{\tilde{\tau}_l^*}^{\tilde{\tau}_{l+1}^*} f_{\tilde{g}}(\tilde{g}) d\tilde{g} = F_{\tilde{g}}(\tilde{\tau}_{l+1}^*) - F_{\tilde{g}}(\tilde{\tau}_l^*)$.

Puesto que $\tilde{p}_l(\tilde{\tau}_{l-1}^*, \tilde{\tau}_l^*, r_{l-1}, \tilde{\epsilon})$ es una función implícita [cf. (3.29)], para calcular $\partial \tilde{p}_l / \partial \tau_l$ debe utilizarse el teorema de derivación implícita: $d\varphi_\epsilon = \frac{\partial \varphi_\epsilon}{\partial x} dx + \frac{\partial \varphi_\epsilon}{\partial y} \frac{\partial y}{\partial x} dx = 0$, resultando $\frac{\partial y}{\partial x} = -\frac{\partial \varphi_\epsilon / \partial x}{\partial \varphi_\epsilon / \partial y}$. Por lo tanto, $\forall l \in \{2, \dots, L\}$ y $\forall i \in \{1, \dots, L\}$ se tiene

$$\frac{\partial \tilde{p}_i}{\partial \tilde{\tau}_l}(\tilde{\tau}_i, \tilde{\tau}_{i+1}, r_i, \tilde{\epsilon}) = \begin{cases} -\frac{[-\epsilon(\tilde{\tau}_l, \tilde{p}_i, r_l) + \tilde{\epsilon}] f_{\tilde{g}}(\tilde{\tau}_l)}{\int_{\tilde{\tau}_i}^{\tilde{\tau}_{i+1}} [\partial \epsilon(\tilde{g}, \tilde{p}_i, r_l) / \partial p] f_{\tilde{g}}(\tilde{g}) d\tilde{g}}, & i = l; \\ \frac{[-\epsilon(\tilde{\tau}_l, \tilde{p}_i, r_{l-1}) + \tilde{\epsilon}] f_{\tilde{g}}(\tilde{\tau}_l)}{\int_{\tilde{\tau}_i}^{\tilde{\tau}_{i+1}} [\partial \epsilon(\tilde{g}, \tilde{p}_i, r_{l-1}) / \partial p] f_{\tilde{g}}(\tilde{g}) d\tilde{g}}, & i = l - 1; \\ 0, & \text{en otro caso;} \end{cases} \quad (3.34)$$

pudiendo expresarse de forma analítica los denominadores de los términos racionales presentes en (3.34) (consúltese el apéndice 3.4.2).

A diferencia del escenario para F-CSIT, en este caso no puede garantizarse que los umbrales óptimos obtenidos mediante (3.33) satisfagan automáticamente $\tilde{\tau}_l^* < \tilde{\tau}_{l+1}^*$, $\forall l$. Si al resolver (3.33) se obtuviese $\tilde{\tau}_l^* \geq \tilde{\tau}_{l+1}^*$, $C2$ no sería débil, $\tilde{\alpha}_l^* > 0$ y (3.33) perdería validez. Para este caso, puesto que las condiciones KKT también exigen que $\tilde{\alpha}_l^*(\tilde{\tau}_l^* - \tilde{\tau}_{l+1}^*) \leq 0$, la solución óptima requeriría que $\tilde{\tau}_l^* = \tilde{\tau}_{l+1}^*$, y por consiguiente el modo l -ésimo debería eliminarse del conjunto de consideración. Esto supone que cada vez que se utilice (3.33) habrá que comprobar la validez de la solución obtenida. Si al resolver (3.33) se obtiene que $\tilde{\tau}_l^* \geq \tilde{\tau}_{l+1}^*$ para un l determinado, el l -ésimo modo se eliminará del conjunto de AMC activos y se reiniciará la

resolución (3.33) para los modos AMC restantes¹¹. Téngase en cuenta que en este caso el cálculo del $\tilde{\tau}_l^*$ óptimo depende no sólo de $\tilde{\beta}^*$ sino también de los umbrales anterior y ulterior ($\tilde{\tau}_{l-1}^*$ y $\tilde{\tau}_{l+1}^*$). Esta vicisitud hace que no sea posible obtener una expresión cerrada para $\tilde{\tau}_l^*$. Pese a ello, puesto que sí existen expresiones cerradas para todos los términos envueltos en (3.33), $\tilde{\tau}_l^*$ puede obtenerse a través de una búsqueda numérica bi-dimensional con complejidad moderada.

Algoritmo 3.3 Cuantificación “off-line” (Q-CSIT)

(S3.0) Sea ε un pequeño nivel de tolerancia, δ un tamaño de paso pequeño y $\tilde{\tau}_L^{\text{máx}} > 0$ el valor máximo para el umbral de cuantificación más alto (e.g., un valor que haga que la probabilidad de la región más alta sea cercana a 0).

(S3.1) Inicialice $\tilde{\beta}$ como un pequeño número positivo y fije $\tilde{\tau}_L = \tilde{\tau}_L^{\text{máx}}$; calcule entonces $\{\tilde{\tau}_l\}_{l=2}^L$ resolviendo (3.33). Si C2 no se satisface para algún τ_l , fije $\tau_l = \tau_{l+1}$. Si la solución obtenida es válida, diríjase a (S3.2); en caso contrario disminuya $\tilde{\tau}_L = \tilde{\tau}_L - \delta$ y repita (S3.1).

(S3.2) Basándose en la fórmula analítica en (3.25), calcule la tasa de transmisión media $\bar{r} = \sum_{l=1}^L [F_{\tilde{g}}(\tilde{\tau}_{l+1}) - F_{\tilde{g}}(\tilde{\tau}_l)]r_l$. Compruebe si se satisface C1 y en caso de que $|\bar{r} - \tilde{r}|/\tilde{r} < \varepsilon$ finalice; en otro caso, calcule $\Delta\tilde{\beta} := (\bar{r} - \tilde{r})c$ utilizando una pequeña constante adaptativa positiva c , actualice el valor del multiplicador $\tilde{\beta} = \tilde{\beta} + \Delta\tilde{\beta}$, y vuelva a (S3.1).

Nótese que la mayor carga computacional en el algoritmo 3.3 se produce en el paso (S3.1) cuando se calculan los umbrales óptimos a través de una búsqueda bi-dimensional. Una vez que se han calculado los valores óptimos de $\tilde{\beta}^*$ y $\{\tilde{\tau}_l^*\}_{l=2}^L$,

¹¹El método propuesto para resolver la violación de la restricción en C2 es válido para cualquier tipo de problema. Por ello para el resto de escenarios a lo largo de esta disertación en los que se consideren umbrales de cuantificación no se escribirá explícitamente la restricción C2 pero deberá tenerse en mente que si en algún caso $\tilde{\tau}_l^* \geq \tilde{\tau}_{l+1}^*$, los representantes asociados a la región l -ésima deberán ser eliminados.

el cuantificador óptimo $Q_l(\cdot)$ queda definido como

$$\mathbf{c}_l = Q_l(\mathbf{h}) := \arg \left\{ g \left[1 - \min_{\mathbf{u} \in \mathcal{U}} d_{ch}^2(\mathbf{u}, \mathbf{h}/\|\mathbf{h}\|) \right] \in [\tilde{\tau}_l, \tilde{\tau}_{l+1}) \right\}. \quad (3.35)$$

Dado el índice del modo AMC encapsulado en \mathbf{c}_l de acuerdo a (3.35), la adaptación óptima de la potencia y tasa de transmisión se obtienen a partir de (3.30) y (3.31).

Realimentación y adaptación de los transmisores en tiempo real

A fin de llevar a cabo las políticas (esquemas) de adaptación óptimas derivadas para el vector de apuntamiento, la tasa y la potencia, a continuación se describe el algoritmo que el CFI y los sensores tienen que ejecutar en tiempo real para cada realización del canal.

Algoritmo 3.4 Adaptación “on-line” (Q-CSIT)

Para cada realización del canal MISO \mathbf{h} :

(S4.1) El CFI obtiene $\mathbf{c}_u = Q_u(\mathbf{h})$ y $\mathbf{c}_l = Q_l(\mathbf{h})$ utilizando respectivamente (3.21) y (3.35) y envía la agregación de ambas $\mathbf{c} = [\mathbf{c}_u; \mathbf{c}_l]$ a los sensores.

(S4.2) Cada sensor m transmite utilizando la m -ésima componente del vector óptimo de apuntamiento indexado por \mathbf{c}_u y con la potencia y tasa indexadas por \mathbf{c}_l .

Nótese que para que este algoritmo pueda ejecutarse, la constelación que contiene los vectores de apuntamiento óptimos (\mathcal{U}^*), debe estar disponible tanto en el CFI como en los sensores. Asimismo, de acuerdo al algoritmo propuesto, en el paso (S4.2) la potencia óptima $\tilde{p}_{\mathbf{c}_l}^*$ correspondiente al índice \mathbf{c}_l es calculada por los sensores mediante la resolución de (3.29). Esto supone que los sensores deben conocer también los umbrales óptimos $\{\tilde{\tau}_l^*\}_{l=2}^L$ (i.e., el CFI debe enviarlos durante la fase preliminar de configuración). A fin de reducir la carga computacional de los sensores, una alternativa consiste en que sea el CFI el que calcule la constelación (conjunto) de potencias $\{\tilde{p}_l^*\}_{l=2}^L$ y envíe después estos valores a los sensores durante

la fase de configuración. En este caso, para cada realización del canal, los sensores simplemente recibirían el índice \mathbf{c}_l e indexarían el elemento correspondiente de la constelación $\{\tilde{p}_l^*\}_{l=2}^L$ almacenada localmente.

3.1.4. Adaptación óptima basada en I-CSIT

Hasta este punto se han derivado los esquemas óptimos de adaptación basados en F-CSIT (donde los sensores conocen de forma perfecta la totalidad del canal vectorial \mathbf{h}) y basados en Q-CSIT (donde los sensores tienen acceso a una información cuantificada del canal vectorial \mathbf{h}). En ambos casos el CSIT debe obtenerse a través de la realimentación proporcionada por el CFI. Sin embargo, de acuerdo a lo analizado en la sección 2.3 para sistemas TDD, el sensor m -ésimo puede obtener h_m de forma natural sin ayuda del CFI (a través de la estimación de su canal inverso). Esta posibilidad motiva el análisis de lo se ha llamado CSIT individual (I-CSIT), donde cada sensor m posee conocimiento perfecto de su propio canal escalar h_m , pero sólo tiene información cuantificada del canal vectorial \mathbf{h} (puesto que debe obtenerla a través del CFI).

Apuntamiento de haz distribuido

Definiendo $\mathbf{v} := \|\mathbf{h}\|\mathbf{u}$ y $\rho := p/\|\mathbf{h}\|^2$, puede escribirse $\mathbf{w} = \sqrt{p}\mathbf{u} = \sqrt{\rho}\mathbf{v}$. Cuando los sensores disponen de F-CSIT, el vector de apuntamiento óptimo $\mathbf{u}^*(\mathbf{h})$ está dado por (3.4) y, por lo tanto, la m -ésima componente del vector $\mathbf{v}^*(\mathbf{h}) = \mathbf{h}^\dagger$ óptimo es $v_m^*(\mathbf{h}) = h_m^\dagger$, que únicamente requiere I-CSIT. La SNR recibida tras implementar la conformación de haz óptima basada en I-CSIT es $\gamma = \rho g^2 = pg$, donde g corresponde a la ganancia del canal SISO equivalente en (3.5), cuyas FDP y FDPA están dadas por (3.6) y (3.7), respectivamente.

Adaptación de tasa y potencia de transmisión

Dado I-CSIT, el vector óptimo de apuntamiento (escalado) es $\mathbf{v}^*(\mathbf{h})$. Para construir el vector de pesos global $\mathbf{w}(\mathbf{h})$, los sensores deberían conocer también la potencia de transmisión (escalada) $\rho(\mathbf{h}) = p(\mathbf{h})/\|\mathbf{h}\|^2$ que requeriría CSI relacionado con la totalidad del canal vectorial \mathbf{h} (o de forma equivalente $\rho(g) := p(g)/g$). Puesto que el canal de realimentación tiene tasa limitada, el CFI tendrá que volver a cuantificar la ganancia del canal g en base a un número finito de regiones. Al igual que para el caso de Q-CSIT, el dominio de g se divide en L regiones de cuantificación distintas $\{\mathcal{R}_l := [\tilde{\tau}_l, \tilde{\tau}_{l+1})\}_{l=1}^L$ de manera que los sensores seleccionan el l -ésimo modo AMC cuando $g \in \mathcal{R}_l$. A diferencia sin embargo del caso de Q-CSIT, ahora el modo l -ésimo queda caracterizado por la pareja tasa-potencia $(r_l, \tilde{\rho}_l)$, donde $\tilde{\rho}_l$ se mantiene constante dentro de cada región y debe ser elegido para satisfacer el requisito de BER $\check{\epsilon}$. Definiendo en este punto

$$\psi_\epsilon(\tilde{\tau}_l, \tilde{\tau}_{l+1}, \tilde{\rho}_l, r_l, \check{\epsilon}) := \int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} \epsilon(g, \tilde{\rho}_l g, r_l) f_g(g) dg - \check{\epsilon} \int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} f_g(g) dg \quad (3.36)$$

y argumentando de forma similar que en (3.27)-(3.35), se deduce que el valor de $\tilde{\rho}_l$ que satisface la BER requerida (que denotaremos como $\tilde{\rho}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, r_l, \check{\epsilon})$) debe ser solución de la ecuación $\psi_\epsilon(\tilde{\tau}_l, \tilde{\tau}_{l+1}, \tilde{\rho}_l, r_l, \check{\epsilon}) = 0$. A diferencia de lo que ocurría para el caso de Q-CSIT en relación a φ_ϵ , no existe ahora una expresión analítica cerrada para ψ_ϵ . Sin embargo, aprovechando el hecho de que $\epsilon(g, \tilde{\rho}_l g, r_l)$ es monótonamente decreciente c.r.a. $\tilde{\rho}_l$, la raíz $\tilde{\rho}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, r_l, \check{\epsilon})$ puede todavía obtenerse de manera eficiente a través de una búsqueda numérica uni-dimensional basada, e.g., en el método de la bisección [65].

De esta manera, puede procederse a formular a la optimización de la adaptación de los transmisores en base a I-CSIT. Dados $\tilde{\rho}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, r_l, \check{\epsilon})$ y la realización del canal MISO \mathbf{h} , la potencia transmitida cuando se elige el l -ésimo modo AMC puede calcularse como $p(\mathbf{h}) = \|\mathbf{h}\|^2 \tilde{\rho}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, r_l, \check{\epsilon}) = g \tilde{\rho}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, r_l, \check{\epsilon}) = p(g)$. A fin de calcular los umbrales de cuantificación $\tilde{\boldsymbol{\tau}}^* := [\tilde{\tau}_1^*, \dots, \tilde{\tau}_{L+1}^*]^T$ que minimicen

la potencia transmitida media, deberá resolverse

$$\left\{ \begin{array}{l} \underset{\tilde{\tau}}{\text{mín}} \quad \bar{p}, \quad \bar{p} := \sum_{l=1}^L \tilde{\rho}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, r_l, \check{\epsilon}) \int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} g f_g(g) dg \\ \text{s. a : C1.} \quad \sum_{l=1}^L r_l \int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} f_g(g) dg \geq \check{r} \\ \text{C2.} \quad \tilde{\tau}_l \leq \tilde{\tau}_{l+1} \quad \forall l. \end{array} \right. \quad (3.37)$$

Si $\tilde{\beta}$ denota el multiplicador de Lagrange asociado con C1 y suponiendo que todas las ligaduras en C2 se cumplen con desigualdad estricta, la condición KKT necesaria para el $\tilde{\tau}_l^*$ óptimo establece

$$\begin{aligned} \frac{\partial \mathcal{L}(\tilde{\beta}^*, \tilde{\tau}^*)}{\partial \tau_l} &= \left[\tilde{\tau}_l^* \tilde{\rho}_{l-1}(\tilde{\tau}_{l-1}^*, \tilde{\tau}_l^*, r_{l-1}, \check{\epsilon}) - \tilde{\beta}^* r_{l-1} - \tilde{\tau}_l^* \tilde{\rho}_l(\tilde{\tau}_l^*, \tilde{\tau}_{l+1}^*, r_l, \check{\epsilon}) + \tilde{\beta}^* r_l \right] f_g(\tilde{\tau}_l^*) \\ &+ \frac{\partial \tilde{\rho}_{l-1}}{\partial \tilde{\tau}_l}(\tilde{\tau}_i, \tilde{\tau}_{i+1}, r_i, \check{\epsilon}) \int_{\tilde{\tau}_{i-1}^*}^{\tilde{\tau}_i^*} g f_g(g) dg + \frac{\partial \tilde{\rho}_l}{\partial \tilde{\tau}_l}(\tilde{\tau}_i, \tilde{\tau}_{i+1}, r_i, \check{\epsilon}) \int_{\tilde{\tau}_i^*}^{\tilde{\tau}_{i+1}^*} g f_g(g) dg = (3.38) \end{aligned}$$

donde $\int_a^b g f_g(g) dg = [\Gamma(M+1, a) - \Gamma(M+1, b)] / \Gamma(M)$; y $\partial \tilde{\rho}_i / \partial \tilde{\tau}_l$, $\forall l \in \{2, \dots, L\}$, $\forall i \in \{1, \dots, L\}$, puede obtenerse mediante derivación implícita como

$$\frac{\partial \tilde{\rho}_i}{\partial \tilde{\tau}_l}(\tilde{\tau}_i, \tilde{\tau}_{i+1}, r_i, \check{\epsilon}) = \begin{cases} -\frac{[-\epsilon(\tilde{\tau}_l, \tilde{\tau}_l \tilde{\rho}_i, r_l) + \check{\epsilon}] f_g(\tilde{\tau}_l)}{\int_{\tilde{\tau}_i}^{\tilde{\tau}_{i+1}} [\partial \epsilon(g, g \tilde{\rho}_i, r_l) / \partial p] g f_g(g) dg}, & i = l \\ \frac{[-\epsilon(\tilde{\tau}_l, \tilde{\tau}_l \tilde{\rho}_i, r_{l-1}) + \check{\epsilon}] f_g(\tilde{\tau}_l)}{\int_{\tilde{\tau}_i}^{\tilde{\tau}_{i+1}} [\partial \epsilon(g, g \tilde{\rho}_i, r_{l-1}) / \partial p] g f_g(g) dg}, & i = l - 1 \\ 0, & \text{en otro caso.} \end{cases} \quad (3.39)$$

Aprovechando la semejanza entre (3.38) y (3.33), podrá derivarse un algoritmo para el cálculo “*off-line*” de $\tilde{\beta}^*$ y $\tilde{\tau}^*$ semejante al algoritmo 3.3.

Realimentación y adaptación de los transmisores en tiempo real

Una vez que se han obtenido los umbrales óptimos $\tilde{\tau}^*$, las $\tilde{\rho}_l(\tilde{\tau}_l^*, \tilde{\tau}_{l+1}^*, r_l, \check{\epsilon})$ correspondientes pueden ser calculadas tanto por el CFI como por los sensores, que tendrán entonces que implementar, para cada realización del canal, el algoritmo de adaptación siguiente:

Algoritmo 3.5 Adaptación “on-line” (I-CSIT)

Para cada realización del canal MISO \mathbf{h} :

(S5.1) *El CFI encuentra $l^*(\mathbf{h}) = l^*(g) = \arg \{g \in [\tilde{\tau}_l^*, \tilde{\tau}_{l+1}^*)\}$ y envía $\mathbf{c} = [l^*(\mathbf{h})]$ a todos los sensores.*

(S5.2) *Cada sensor m transmite el símbolo común s utilizando el $l^*(\mathbf{h})$ -ésimo modo AMC y el coeficiente óptimo $w_m^* = \sqrt{\rho_{l^*(\mathbf{h})} \left(\tilde{\tau}_{l^*(\mathbf{h})}^*, \tilde{\tau}_{l^*(\mathbf{h})+1}^*, r_{l^*(\mathbf{h})}, \check{\epsilon} \right)} h_m^\dagger$.*

Nótese que puesto que el apuntamiento de haz basado en I-CSIT no precisa información por parte del CFI, para implementar el algoritmo 3.5 únicamente se necesitan $B = \log_2(L)$ bits para realimentar el CSIT. Este número puede ser significativamente menor que los $B = \log_2(L + N_u)$ bits que se precisan para enviar la información asociada a los esquemas de adaptación en base a Q-CSIT de la sección 3.1.3, especialmente si se trabaja con elevados valores de N_u (que aparecen típicamente asociados a elevados valores de M).

3.1.5. Simulaciones numéricas

En esta sección, se presentan ejemplos numéricos que muestran el consumo de potencia transmitida por los sensores cuando estos disponen de F-CSIT, Q-CSIT o I-CSIT. S.p.d.g. la energía por símbolo, el ancho de banda y la densidad espectral de potencia del ruido AWGN se eligen para que $E_s/N_0 = 1$. Para realizar los experimentos, se elige un sencillo escenario con cuatro sensores¹² y enlaces con el CFI de acuerdo a (sp1). A menos que se indique explícitamente lo contrario, se supondrá que cada sensor puede transmitir de acuerdo a tres modos activos que implementan modulaciones M^r -arias sin codificar: 2-QAM, 8-QAM y 32-QAM más un modo inactivo; i.e., las tasas de transmisión de los modos AMC son:

¹²Este es un razonable escenario de aplicación para, e.g., una WSN jerárquica estructurada en grupos donde el papel de CFI es desempeñado por un cabeza de grupo y en el que para limitar el consumo energético únicamente se activan (despiertan) unos pocos sensores por grupo.

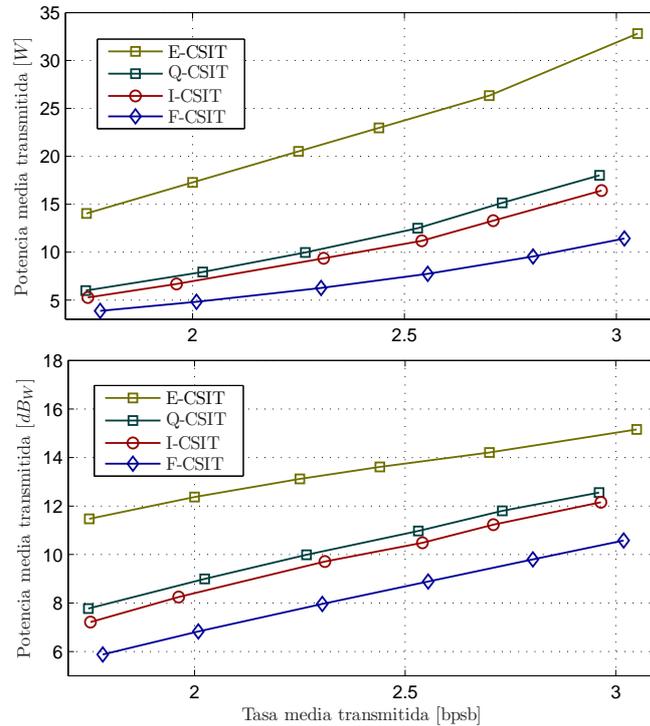


Figura 3.2: Potencia media transmitida (\bar{p}) frente a tasa media transmitida para diferentes tipos de CSIT ($M = 4$, $L = 4$, $N_u = 16$).

$r_l = 0, 1, 3, 5$ bits/simbolo. En todas las simulaciones, se fijará el requisito de BER en $\tilde{\epsilon} = 10^{-3}$ y, para aquellos experimentos en que se utilice, el tamaño de la constelación de los vectores de apuntamiento es $N_u = 16$.

Caso de estudio 1 (*Comparación del consumo de potencia transmitida*): Para distintos requisitos de tasa de transmisión, la figura 3.2 muestra la potencia media transmitida (medida en W –vatios para la gráfica superior– y en dB_W –decibelios c.r.a. 1 watio para la gráfica inferior–) consumida por los esquemas óptimos de adaptación basados en: (i) F-CSIT, (ii) Q-CSIT, (iii) I-CSIT, y (iv) CSIT espacial (E-CSIT). Este cuarto caso, E-CSIT, se considera con fines ilustrativos para comparar el consumo de sensores que, disponiendo de F-CSIT, únicamente apuntan de forma óptima el haz de transmisión (i.e. sincronizan sus transmisiones para maximizar la ganancia del canal) pero no adaptan su potencia y tasa a lo largo del tiempo.

Observando la figura 3.2 puede concluirse que: (i) los dos esquemas de adap-

Tabla 3.1: Potencia media transmitida (dB_W) para sistemas de lazo abierto y sistemas de lazo cerrado según \tilde{r} varía.

\tilde{r}	1.75	2	2.25	2.5	2.75	3
Lazo cerrado (F-CSIT)	5.8	6.8	8.0	8.6	9.8	10.6
Lazo abierto	28.4	29.34	30.1	30.5	31.1	31.8

tación basados en Q-CSIT e I-CSIT presentan un consumo de potencia cercano al consumo óptimo representado por el esquema basado en F-CSIT; (ii) los esquemas basados en Q-CSIT y en I-CSIT se comportan de manera casi idéntica; (iii) la distancia en términos de consumo (medida en dB) entre los esquemas que cuantifican la ganancia del sistema y el esquema óptimo basado en F-CSIT permanece prácticamente constante para distintos requisitos de tasa de transmisión y (iv) las estrategias basadas en Q-CSIT y I-CSIT superan claramente las prestaciones del esquema de adaptación basado en E-CSIT aun cuando este último requiere F-CSIT frente a los dos anteriores que únicamente necesitan la realimentación de unos pocos bits.

A fin de comprobar la relevancia de la realimentación (o alternatively de la adaptación de los transmisores) para el consumo de potencia, se compara el consumo asociado al esquema de adaptación basado en F-CSIT (como representante óptimo de los sistemas realimentados –t.c.c. sistemas de lazo cerrado–) y el consumo asociado a los sistemas sin realimentación –t.c.c. sistemas de lazo abierto–. De acuerdo a los resultados mostrados en la tabla 3.1, la potencia consumida por el sistema de lazo abierto es $20 \sim 25 dB$ más elevada que la consumida por el sistema en lazo cerrado (F-CSIT). Se comprueba así cómo, de acuerdo a lo esperado, la existencia de CSIT disminuye de forma significativa las necesidades de potencia y, por lo tanto, incrementa considerablemente el *tiempo de vida* de la WSN.

Caso de estudio 2 (*Consumo de potencia transmitida en distintos escenarios*):

La tabla 3.2 compara el comportamiento de los esquemas basados en F-CSIT, Q-CSIT, I-CSIT y E-CSIT para un rango variado de parámetros. Las simulaciones numéricas confirman: (i) que la adaptación basada en Q-CSIT e I-CSIT son casi óptimas y (ii) el uso de la adaptación basada en E-CSIT se penaliza con un signi-

Tabla 3.2: Potencia media transmitida (dB_W) para esquemas basados en (F, I, Q, y S)-CSIT. (Caso de referencia: $M = 4$, $\tilde{r} = 2.5$, $\tilde{\epsilon} = 10^{-3}$, $L = 4$, $r_l = [0, 1, 3, 5]$, $N_u = 16$, $E_s/N_0 = 1$; para otros CASOS, se modifica un único parámetro c.r.a. caso de referencia.)

CASOS	F-CSIT	I-CSIT	Q-CSIT	E-CSIT
Referencia	8.6	10.2	10.7	13.8
$M = 6$	5.0	6.5	6.9	11.2
$E_s/N_0 = 3$	3.9	5.4	5.9	9.1
$\tilde{\epsilon} = 10^{-4}$	10.1	11.9	12.1	16.7
$L = 6$	8.5	9.4	9.8	13.8
$r_l = [0, 1, 2, 6, 4]$	8.7	10.2	10.5	13.2

Tabla 3.3: Potencia media (dB_W) y tasa de transmisión por modo AMC (región) y umbrales de cuantificación ($M = 4$, $\tilde{r} = 2.5$, $\tilde{\epsilon} = 10^{-3}$, $L = 4$, $r_l = [0, 1, 3, 5]$, $N_u = 16$).

Modo AMC	F-CSIT			I-CSIT			Q-CSIT		
	$l = 2$	$l = 3$	$l = 4$	$l = 2$	$l = 3$	$l = 4$	$l = 2$	$l = 3$	$l = 4$
\bar{p}_l	4.96	9.57	11.75	4.40	10.64	14.62	7.02	11.45	13.54
r_l	1	3	5	1	3	5	1	3	5
τ_{l+1}	0.8	2.4	9.7	1.7	3.0	5.8	0.6	1.6	5.1
$\Pr\{l^* = l\}$	0.22	0.75	0.01	0.26	0.48	0.17	0.25	0.68	0.05

ficativo incremento del consumo (c.r.a. la adaptación basada en F-CSIT) debido a que este esquema no aprovecha la diversidad temporal ofrecida por el canal.

Caso de estudio 3 (*Caracterización de la solución óptima*): Para mejorar la comprensión del problema, la tabla 3.3 especifica la cuantificación y asignación de recursos óptimas para las tres categorías de CSIT analizadas. Recuérdese que para las soluciones basadas en F-CSIT y I-CSIT los valores de los umbrales se encuentran normalizados respecto al valor medio de la ganancia g mientras que en el caso de Q-CSIT se refieren a $\tilde{g} < g$.

De los resultados mostrados en la tabla puede deducirse que, para todos los sistemas de lazo cerrado, menores tasas de transmisión requieren una menor potencia media transmitida. Asimismo, se observa cómo los esquemas derivados tratan de homogeneizar la potencia por bit para todas las regiones de cuantificación. Se comprueba además cómo, para todos los casos, el modo que se selecciona con una mayor probabilidad corresponde al más cercano al requisito de tasa ($l = 3$ con

$r_3 = 3$, siendo $\check{r} = 2.5$). Este comportamiento es si cabe más pronunciado para el caso de F-CSIT puesto que al permitir el esquema la adaptación de potencia por modo AMC, la necesidad de adaptar la tasa disminuye. Este patrón homogéneo entre los distintos esquemas no se reproduce en el caso de los umbrales óptimos de cuantificación, ya que para los tres casos considerados, los valores resultan bastante diferentes (recuérdese que para juzgar justamente el caso de Q-CSIT, los umbrales deberían expresarse en términos de g cuyos valores son ligeramente mayores que para el caso de \tilde{g} , que es el que se muestra en la tabla 3.3). Estas discrepancias se explican en términos de la diferencia del comportamiento de la SNR recibida en el CFI. Concretamente, la SNR recibida es: (i) constante para el caso de F-CSIT donde $\gamma = p_l(g)g$ con $p_l(g)$ proporcional a $1/g$; (ii) proporcional a \tilde{g} para el caso de Q-CSIT donde $\gamma = \tilde{p}_l\tilde{g}$ con \tilde{p}_l constante; y (iii) proporcional a g^2 para el caso de I-CSIT donde $\gamma = \tilde{\rho}_l g^2$ con $\tilde{\rho}_l$ constante. Este comportamiento también hace que la solución para I-CSIT sea muy sensible a pequeñas ganancias de canal y, por consiguiente, fije el umbral de la primera región activa en un valor relativamente alto.

Caso de estudio 4 (*Efectos asociados a la variación del número regiones*): Hasta ahora se ha comprobado cómo para $L = 4$ (tres modos AMC activos) y $N_u = 16$, los esquemas de adaptación basados en Q-CSIT e I-CSIT presentan un consumo de potencia cercano al consumo mínimo reflejado por la adaptación en base a F-CSIT. Esto ocurre utilizando $\lceil \log_2(4) + \log_2(16) \rceil = 6$ y $\lceil \log_2(4) \rceil = 2$ bits por cada realización del canal para, respectivamente, los casos de Q-CSIT e I-CSIT. A continuación se analiza cómo la variación del número de bits reenviados modifica el comportamiento del sistema.

En primer lugar se estudia el impacto de la variación del número de modos AMC que los sensores implementan. La tabla 3.4 muestra el coste total de potencia para los casos de F-CSIT, I-CSIT y Q-CSIT considerando distintos valores de L . Para $L = 1$, los sensores únicamente implementan un modo AMC por lo que no se requiere realimentación del CFI; mientras que para el resto de casos, los sensores implementan $L - 1$ modos activos más un modo inactivo adicional que requieren

Tabla 3.4: Potencia media transmitida (dB_W) variando L .

L	1	2	4	6	8	∞
F-CSIT	9.8	9.4	8.6	8.5	8.4	8.3
I-CSIT	13.9	11.5	10.2	9.4	8.9	8.3
Q-CSIT	14.2	11.7	10.7	9.8	9.3	8.9

$\lceil \log_2(L) \rceil$ bits de realimentación. Recuérdese que para el diseño basado en Q-CSIT, se había supuesto que se utilizarían $N_u = 16$ vectores de apuntamiento diferentes. Analizando los resultados obtenidos se observa que, según L crece: (i) el consumo de potencia disminuye para todos los casos, (ii) la diferencia de prestaciones entre los esquemas basados en Q-CSIT y I-CSIT y el óptimo basado en F-CSIT también disminuye y (iii) el primer y segundo incremento de L producen el mayor ahorro de potencia. Nótese también que debido al hecho de que $N_u < \infty$ la potencia consumida para el caso de Q-CSIT es siempre superior a la consumida para el caso de F-CSIT incluso cuando $L \rightarrow \infty$.

En segundo lugar, se pondera el efecto de la variación de N_u para el caso de Q-CSIT. En la figura 3.3 se representa la potencia transmitida frente al número de bits dedicados a identificar el vector de apuntamiento ($\log_2(N_u)$). Por propósitos ilustrativos también se representa la potencia transmitida para los casos de I-CSIT y F-CSIT. Al igual que en el caso anterior, se comprueba cómo el consumo de potencia disminuye según N_u crece, mientras que la disminución obtenida por cada bit adicional decrece. Nótese que según $N_u \rightarrow \infty$, $\tilde{g} \rightarrow g$ y a efectos prácticos únicamente se está cuantificando g y la potencia transmitida. Asimismo, se observa que cuando N_u alcanza unos valores elevados, el esquema basado en Q-CSIT mejora al basado en I-CSIT. Como ya se ha mencionado, la divergencia entre ambas soluciones se debe al hecho de que se están cuantificando distintas variables (p y $\rho = p/\|\mathbf{h}\|^2$) y, por lo tanto, los diseños óptimos no coinciden totalmente. De forma intuitiva, la ventaja del caso Q-CSIT podría ser debida a que las variaciones de la SNR recibida son menos pronunciadas en el caso de Q-CSIT (g frente a g^2), comportamiento más cercano al del caso F-CSIT (óptimo global) donde la SNR recibida permanece constante. Sin embargo, resulta conveniente reiterar que la

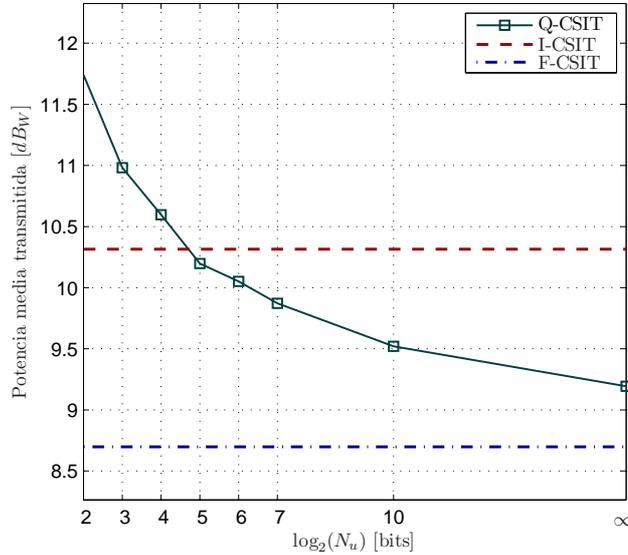


Figura 3.3: Variación del número de bits realimentados para la adaptación basada en Q-CSIT ($M = 4$, $\tilde{r} = 2,5$, $L = 4$).

adaptación basada en Q-CSIT requiere exige una mayor tasa de realimentación que la basada en I-CSIT, especialmente para los casos en que la adaptación basada en Q-CSIT mejora la basada en I-CSIT (i.e., para valores elevados de N_u).

Por último, debe destacarse que, pese a que la teoría efectivamente predecía que asintóticamente las soluciones basadas en Q-CSIT e I-CSIT deberían tender a la solución basada en F-CSIT y, por lo tanto, la diferencia de prestaciones debería desaparecer en el límite ($L, N_u \rightarrow \infty$), las simulaciones numéricas sugieren que incluso con un número relativamente bajo de bits realimentados (e.g., $L = 2^3$ y $N = 2^5$) la distancia es prácticamente nula.

3.1.6. Síntesis de resultados

Para una WSN en la que las comunicaciones se establecen de forma coherente entre los sensores y un CFI, se ha minimizado la potencia media transmitida garantizando requisitos de tasa de transmisión media y BER para escenarios en los que los sensores disponían de CSIT: global (F-), cuantificado (Q-) e individual (I-). A fin de derivar los esquemas de adaptación óptimos, s.p.d.o. se ha separado

el problema original en dos sub-problemas: (i) cuantificación del canal MISO para el apuntamiento de haz, y (ii) cuantificación de la ganancia del canal para la cuantificación y asignación de tasa y potencia. Aprovechando el paralelismo entre el modelo de comunicaciones coherentes de la WSN y el de un sistema MISO distribuido, se han utilizado técnicas de optimización (programación) no lineal para resolver de forma óptima (local o global) los problemas planteados derivando así los esquemas de cuantificación y adaptación que minimizan la potencia y de complejidad computacional moderada. Los resultados numéricos han confirmado que las soluciones obtenidas para los casos de Q-CSIT e I-CSIT presentan un consumo energético sorprendentemente cercano al presentado por el esquema óptimo basado en F-CSIT, mejoran claramente el consumo de un esquema basado en E-CSIT que únicamente aprovecha la diversidad espacial y que requiere F-CSIT y ofrecen un ahorro de potencia muy significativo c.r.a. sistemas de lazo abierto que no utilizan el CSIT.

3.2. Minimización de potencia en OFDM con QoS instantánea utilizando CSIT

El *multiplexado* ortogonal por división en frecuencia (OFDM) es en la actualidad la modulación de referencia para canales selectivos en frecuencia (o canales multi-camino) con limitaciones de ancho de banda. El hecho de que OFDM haya sido la técnica de transmisión adoptada en las tecnologías digitales de bucle de abonado (DSL), en los estándares de difusión de audio y vídeo digital (DAB/DVB) y un multitud de redes de área local, por nombrar algunos [36, 90], no hace sino testimoniar la relevancia y los méritos de esta técnica. Es un hecho constatado que, para estos sistemas, tanto la eficiencia espectral como la robustez frente a errores mejoran de forma clara si existe información del estado del canal en los transmisores (CSIT). Debido a ello, las transmisiones OFDM sobre canales guiados o sobre canales inalámbricos que varían lentamente en el tiempo tradicionalmente

han considerado esquemas adaptativos que, dada una potencia máxima de transmisión, permitieran la maximización de tasa mediante el uso de CSIT de carácter perfecto y determinista. Sin embargo, al igual que en las secciones anteriores, la asimetría entre el canal de transmisión y de recepción (la gran mayoría de los sistemas OFDM son sistemas FDD), los errores de estimación del canal en el receptor y el retardo del canal de retorno hacen que para la mayoría de sistemas inalámbricos resulte imposible que el transmisor disponga de P-CSIT de forma práctica. Este hecho ha motivado el desarrollo de esquemas OFDM adaptativos basados en CSIT estadístico (S-CSIT) o incluso cuantificado (Q-CSIT) donde la descripción de canal se hace en base a un número limitado de bits provenientes del receptor (véase, e.g., [20, 25, 47, 97] y referencias). No obstante, los anteriores trabajos tratan con sistemas limitados por tasa de transmisión fiable, en donde el objetivo principal consiste en maximizar la tasa de transmisión o en minimizar la tasa de error de bit (BER).

Curiosamente, excepto por el caso de [70] donde se utiliza P-CSIT para disminuir el consumo de potencia, no se han dedicado esfuerzos equivalentes para la optimización de las prestaciones de sistemas OFDM de un único usuario limitados por potencia. Este es el tema que se analiza en esta sección, cuyo objetivo principal consiste en derivar esquemas y algoritmos de adaptación para minimizar la potencia instantánea sujeta a requisitos de BER y tasa de transmisión.

Las diferencias con respecto al escenario de la sección anterior son variadas. En primer lugar, desde el punto de vista físico, el modelo de canal es distinto: mientras en el caso anterior se analizaba un canal MISO selectivo en el tiempo y plano en frecuencia, aquí se analizará un canal SISO selectivo en tiempo y en frecuencia. El uso de OFDM transformará el canal doblemente selectivo en un conjunto de canales paralelos con respuesta en frecuencia plana, aumentando los grados de libertad del sistema. En segundo lugar el propio esquema de adaptación varía: mientras en el caso anterior la adaptación de potencia y la tasa se realizaba de manera conjunta relacionándolas a través del requisito de BER, aquí la adaptación se hará de forma independiente y, por lo tanto, habrá que añadir una ligadura que,

de forma explícita, tenga en cuenta el requisito de BER. Finalmente, la tercera diferencia reside en la forma de evaluar las prestaciones del sistema: si en la sección 3.1 tanto la función objetivo (potencia) como la ligadura (tasa) representaban medidas promediadas en el tiempo, en esta sección la potencia, tasa y BER serán promediadas únicamente a través de sub-portadoras y por lo tanto instantáneas en el tiempo (i.e., se minimizará la potencia debiendo satisfacer, para cada símbolo, un requisito de tasa y BER). Se afronta pues un problema más restrictivo que, como se verá, requiere una mayor complejidad para resolverlo.

La presente sección se estructurará como sigue. Tras detallar el modelo y escenario de trabajo, se derivarán en los esquemas de adaptación de potencia y tasa basados en P-CSIT. La función principal de este esquema será facilitar una cota inferior con la que se puedan comparar las prestaciones del resto de esquemas adaptativos derivados. A continuación se derivarán los esquemas basados en Q-CSIT, que son los más relevantes desde el punto de vista práctico. El diseño basado en Q-CSIT se estructurará en dos fases: una fase “off-line” donde se especificará el formato del Q-CSIT a través de la obtención de los umbrales de cuantificación óptimos en base a las estadísticas del canal y una fase “on-line” en la que, basándose en el valor instantáneo del Q-CSIT, se derivan los esquemas óptimos de adaptación de tasa y potencia. A fin de aumentar la aplicabilidad de las técnicas propuestas, el siguiente paso consistirá en encontrar esquemas de cuantificación, adaptación y realimentación que permitan reducir la complejidad global del sistema. Finalizado el análisis para el caso de Q-CSIT, se presentarán los esquemas de adaptación basados en S-CSIT. Finalmente, concluida la presentación de esquemas, las prestaciones de las distintas alternativas se compararán a través de simulaciones numéricas, mientras que una síntesis de resultados a modo de conclusiones servirá para cerrar la sección.

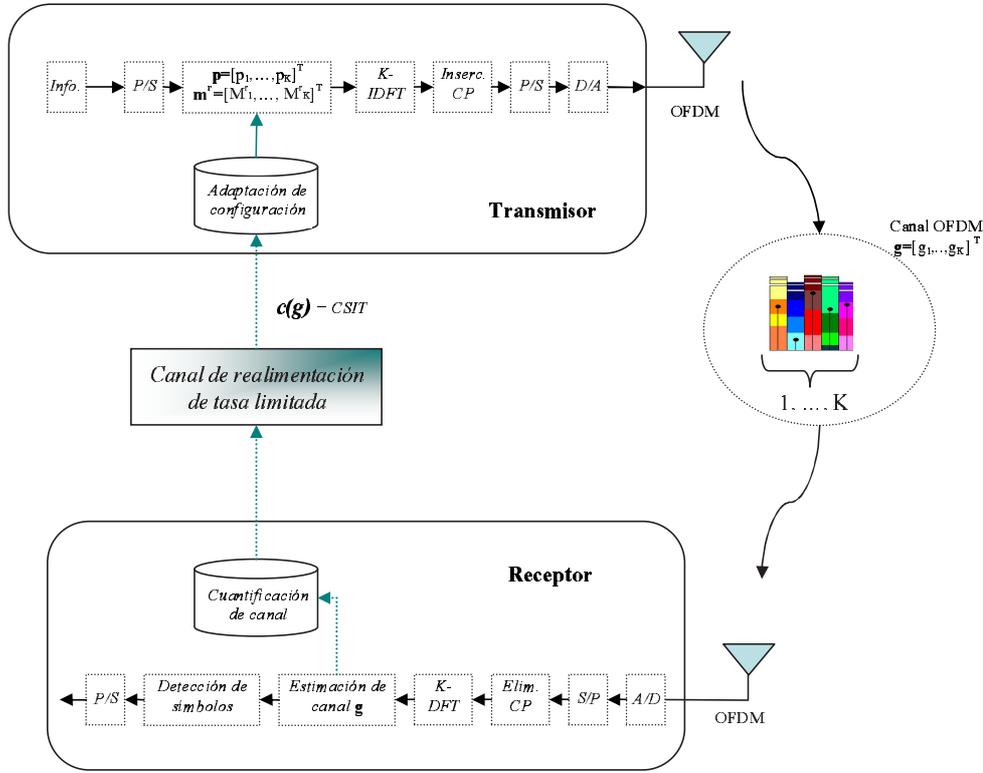


Figura 3.4: Diagrama de bloques del sistema OFDM.

3.2.1. Modelo y preámbulos

De acuerdo a lo representado en la figura 3.4, se considera un sistema de transmisión OFDM operando sobre K sub-portadoras (sub-canales) a través un canal selectivo en frecuencia cuya respuesta al impulso discreta equivalente en banda base para un tiempo determinado está compuesta por los rayos $\{h_q\}_{q=0}^Q$, donde: $Q := \lfloor D_{\text{máx}} BW \rfloor$ denota el orden del canal, $D_{\text{máx}}$ el tiempo máximo de dispersión y BW el ancho de banda de transmisión. El flujo de datos proveniente de la fuente de información será entonces dividido en K flujos paralelos cada uno de los cuales (indexado mediante $k \in \{1, K\}$) se modula para obtener símbolos complejos $\{A_k\}_{k=1}^K$ pertenecientes a una constelación de tamaño M_k^r y se multiplica por una constante determinada a fin de transmitirlo con una potencia p_k . A la aplicación de la inversa de la transformada discreta de Fourier (I-DFT) de K puntos a cada

bloque de K símbolos (cada uno de ellos perteneciente a un sub-flujo distinto) le sigue la inserción de un prefijo cíclico de Q muestras, produciendo así un bloque de $K + Q$ símbolos discretos (t.c.c. símbolo OFDM) que son convertidos en símbolos analógicos y enviados secuencialmente para su transmisión. Estas operaciones junto con su procesado inverso en el receptor (eliminación del CP, cálculo de la DFT, etc.) convierten al canal selectivo en frecuencia debido al multi-camino en un conjunto de K canales planos paralelos con coeficiente complejo de transmisión (consúltese e.g., [90]): $H_k = (1/\sqrt{K}) \sum_{q=0}^{K-1} h_q e^{-j\frac{2\pi}{K}kq}$, donde el valor de K se escoge típicamente para que $K \gg Q$.

La estimación del canal por parte del receptor se realiza en base a símbolos de entrenamiento (pilotos) que se insertan periódicamente (véase, e.g., [67]). Una vez adquirido el canal, el receptor tiene disponible un vector de ganancias de canal normalizadas $\mathbf{g} := [g_1, \dots, g_K]^T$ donde $g_k := |H_k|^2 / \sigma_k^2$ representa la ganancia (medida en términos de potencia) normalizada para la k -ésima sub-portadora y σ_k^2 es la varianza del ruido AWGN presente en la misma que s.p.d.g. se supone unitaria. Cada realización determinista de la VA g_k constituye el P-CSI correspondiente a la k -ésima sub-portadora, mientras que la media estadística $\bar{g}_k := \mathbb{E}_{g_k}[g_k]$ constituye el S-CSI correspondiente. Si bien g_k debe ser conocido en el transmisor en el caso de P-CSIT y \bar{g}_k debe ser conocido en el transmisor en el caso de S-CSIT, en el apartado que describe los esquemas basados en Q-CSIT se mostrará que \bar{g}_k juega también un papel importante en el diseño del cuantificador de canal que generará el Q-CSIT y, por consiguiente, deberá ser conocido en los dos extremos de la cadena de comunicación. En relación a este último punto, es necesario destacar que en este trabajo se considerará un esquema que cuantifica todas y cada una de las sub-portadoras. Esto implicará que, utilizando de nuevo la convexidad de las funciones de BER, el diseño de las regiones de cuantificación $\mathcal{R}_{k,l}$ (con l indicando el índice de la región para la sub-portadora k) corresponderá con la selección de los umbrales $\tau_{k,l}$ correspondientes (i.e., $\mathcal{R}_{k,l} = [\tau_{k,l}, \tau_{k,l+1})$), umbrales que definirán el índice l del intervalo al que la ganancia de la realización de la k -ésima sub-portadora (g_k) pertenece. En base a estas regiones, se podrá definir el

vector aleatorio de bits $\mathbf{c} := [c_1, \dots, c_K]^T$ donde c_k es la representación binaria del índice l para la sub-portadora k . De esta manera, $\mathbf{c} = \mathbf{c}(\mathbf{g})$ constituirá el Q-CSI que, a través del canal de realimentación y para cada realización del canal, el receptor enviará al transmisor a fin de que este pueda adaptar la distribución de bits y potencia a lo largo de las sub-portadoras disponibles.

El objetivo del diseño en esta sección consistirá en, basándose en \mathbf{g} (P-CSIT), en su media $\bar{\mathbf{g}}$ (S-CSIT) o en la cuantificación de las ganancias contenidas en el vector de bits \mathbf{c} (Q-CSIT), escoger vectores de potencia $\mathbf{p} := [p_1 \dots p_K]^T$ y de bits¹³ $\mathbf{m}^r := [M_1^r, \dots, M_K^r]^T$ de manera que se minimice la potencia $p := \sum_{k=1}^K p_k$ o su media \bar{p} cumpliendo requisitos de calidad de servicio consistentes en una tasa de transmisión mínima (\check{r}) y BER máxima ($\check{\epsilon}$) para cada uno de los símbolos OFDM transmitidos $\mathbf{a} := [A_1 \dots A_K]^T$ (i.e., se satisfacen niveles de QoS de forma instantánea). Asimismo, para el caso de Q-CSIT, se elegirá de forma adecuada el cuantificador de canal y se propondrán esquemas flexibles de adaptación de potencia y tasa transmitida que, aunque posiblemente sub-óptimos en lo que a consumo de potencia se refiere, permitan una implementación con complejidad reducida o despreciable.

La consecución de los anteriores objetivos se realizará bajo las siguientes suposiciones:

- (sp1) *Los símbolos A_k pertenecen a constelaciones QAM de tamaño M_k^r ;*
- (sp2) *Los sub-canales H_k permanecen invariantes al menos durante dos símbolos OFDM consecutivos, pueden presentar correlación para distintos k y cada uno de ellos sigue una distribución compleja gaussiana; i.e., la ganancia de potencia de cada sub-portadora, g_k , tiene una FDP exponencial $f_{g_k}(g_k) = (1/\bar{g}_k) \exp(-g_k/\bar{g}_k)$;*
- (sp3) *El canal de realimentación está libre de errores y presenta un retardo despreciable.*

Las suposiciones (sp1)-(sp3) son compartidas por la mayoría de diseños exis-

¹³Nótese que, a diferencia de la sección anterior, aquí no se considerará como variable de adaptación directamente la tasa de transmisión sino el tamaño de la constelación de símbolos con la que se transmite. Como se verá más adelante, este cambio influirá en la técnica de optimización utilizada para resolver el problema.

tentes [20, 25, 36, 47], y típicamente se satisfacen en la práctica. En principio, los resultados derivados aplican a cualquier tipo de FDP, pero (sp2) simplifica las expresiones obtenidas. Como ya se ha comentado en anteriores ocasiones, la invariabilidad expresada en (sp2) hace que los diseños sean tolerantes a pequeños retrasos en la información de realimentación y, aunque en el límite podría resultar restrictiva para diseños basados en P-CSIT, es sin duda razonable para diseños basados en Q-CSIT donde el canal en cada sub-portadora puede variar de un símbolo a otro con tal de que el intervalo de cuantificación permanezca constante. Por último (sp3) es una suposición típica para los canales de realimentación donde la tasa de información es baja y presentan una fuerte protección frente a errores.

Con r denotando la tasa de transmisión (número de bits por símbolo –bloque– OFDM que puede expresarse como $r = \sum_{k=1}^K \log_2 M_k^r$), la BER por símbolo (número esperado de bits erróneos dividido por número de bits transmitido) es

$$\epsilon(\mathbf{g}, \mathbf{p}, \mathbf{m}^r) = \frac{1}{r} \sum_{k=1}^K \log_2 M_k^r \epsilon_k(g_k, p_k, M_k^r) \quad (3.40)$$

donde $\epsilon_k(g_k, p_k, M_k^r) \simeq \kappa_1 \exp(-v_k g_k p_k)$ y, para las constelaciones QAM recogidas bajo (sp1), $\kappa_1 = 0,2$, $v_k := \frac{\kappa_2}{M_k^r - 1}$, $\kappa_2 = 1$ si $M_k^r = 2$ y $\kappa_2 = 1,5$ si $M_k^r \geq 4$. Distintos trabajos ([21, 28]) demuestran que la aproximación de la BER para la k -ésima sub-portadora en base a $\epsilon_k(\cdot)$ es muy precisa y su uso para el diseño y caracterización de sistemas OFDM está ampliamente extendido. Nótese que para un determinado valor de BER para la sub-portadora k -ésima, a través de la inversión de la función $\epsilon_k(\cdot)$, p_k puede expresarse como función de g_k y los valores de κ_1 y v_k de forma similar a lo realizado en (3.10). Esta expresión junto con (3.40) aparecerán de forma explícita en el siguiente apartado, donde se optimiza la asignación de bits y potencia en base a P-CSIT.

3.2.2. Adaptación óptima basada en P-CSIT

En esta apartado, suponiendo que el vector \mathbf{g} es conocido, se optimizará la potencia transmitida por un sistema OFDM. Recuérdesse que si bien el esquema de adaptación que a continuación se derivará es de utilidad para sistemas transmitiendo a través de canales inalámbricos de variabilidad muy lenta o sistemas cableados, al representar \mathbf{g} el CSI más completo (avanzado) posible, sirve sobre todo para acotar el consumo de potencia de los diseños basados en CSIT parcial (bien sea cuantificado o estadístico). Con \mathcal{M} denotando el conjunto finito de posibles valores para \mathbf{m}^r , el problema de optimización basado en P-CSIT puede plantearse como sigue:

$$\left\{ \begin{array}{l} \text{mín}_{\mathbf{m}^r(\mathbf{g}) \in \mathcal{M}, \mathbf{p}(\mathbf{g}) \geq \mathbf{0}} J(\mathbf{p}), \quad J(\mathbf{p}) := \sum_{k=1}^K p_k(\mathbf{g}) \\ \text{s. a : C1. } \sum_{k=1}^K \frac{\log_2 M_k^r}{\tilde{r}} \epsilon_k(g_k, p_k, M_k^r) \leq \check{\epsilon} \\ \text{C2. } \sum_{k=1}^K \log_2 M_k^r \geq \check{r}. \end{array} \right. \quad (3.41)$$

Puesto que las componentes de \mathbf{p} son valores continuos mientras que las de \mathbf{m}^r son valores enteros, el problema de optimización pertenece al tipo de programación entera mixta (o *mixed-integer programming* –MIP–) para cuya resolución se propone proceder en dos fases. En primer lugar se resolverá (3.41) para un \mathbf{m}^r determinado (i.e., se eliminará C2 y el valor óptimo de la asignación de potencia \mathbf{p} obtenido quedará expresado en términos de \mathbf{m}^r), resuelto este primer paso se seleccionará el óptimo global ($\mathbf{p}^*, \mathbf{m}^{r*}$) mediante búsqueda exhaustiva¹⁴ entre todas las combinaciones (finitas) de $\mathbf{m}^r \in \mathcal{M}$. Sustituyendo (3.40) en (3.41), el primer problema puede plantearse como:

$$\left\{ \begin{array}{l} \text{mín}_{\mathbf{p}(\mathbf{g}) \geq \mathbf{0}} J(\mathbf{p}), \quad J(\mathbf{p}) = \sum_{k=1}^K p_k(\mathbf{g}) \\ \text{s. a : C1. } \sum_{k=1}^K \frac{\log_2 M_k^r}{\tilde{r}} \kappa_1 \exp(-\nu_k g_k p_k(\mathbf{g})) \leq \check{\epsilon}. \end{array} \right. \quad (3.42)$$

¹⁴Como se mostrará en breve cuando se presente el algoritmo de adaptación, el tamaño del espacio de búsqueda para \mathbf{m}^r se reducirá drásticamente.

Haciendo uso de las condiciones (necesarias) de optimalidad KKT [32, sección 14.6], la solución óptima expresada en términos de $p_k^*(\mathbf{g})$ puede expresarse como:

$$p_k^*(\mathbf{g}) = \frac{1}{v_k g_k} \left[\ln \left(\frac{\kappa_1 v_k g_k \log_2 M_k^r}{\check{r}} \beta_D \right) \right]^+, \quad (3.43)$$

donde β_D representa el valor del multiplicador de Lagrange cuyo valor se obtiene tras la sustitución de (3.43) en C1 y la satisfacción estricta de la restricción. Puesto que $J(\mathbf{p})$ y C1 son funciones convexas c.r.a. \mathbf{p} , la solución en (3.43) representa un óptimo global. Nótese que para valores de ganancia muy bajos, la asignación óptima exige que se cargue potencia nula en la sub-portadora¹⁵. Sea \mathcal{K}_a entonces el conjunto de índices correspondiente a las K_a ($\leq K$) sub-portadoras activas (i.e., aquellas a las que se asigna una potencia distinta de cero). Sustituyendo (3.43) en C1, se obtiene

$$\beta_D^* = \frac{1}{\check{c}} \sum_{k \in \mathcal{K}_a} \frac{1}{v_k g_k}. \quad (3.44)$$

Una vez resuelta la primera fase, en la segunda debe incorporarse la ligadura C2. La resolución del problema global junto con la reducción de la búsqueda exhaustiva sobre \mathbf{m}^r se describe en el siguiente algoritmo:

Algoritmo 3.6 (P-CSIT)

(S6.1) *Ordene las sub-portadoras de forma descendente en función de sus valores de g_k .*

¹⁵Aunque el hecho de eliminar aquellas sub-portadoras que tengan una ganancia muy pequeña resulte fácil de entender intuitivamente, la explicación matemática de este comportamiento es algo más compleja. La justificación reside en el hecho de que la solución natural asociada a la derivada del Lagrangiano de 3.42 admite valores negativos al mismo tiempo que existe una ligadura implícita que exige $\mathbf{p}(\mathbf{g}) \geq \mathbf{0}$. De forma similar a lo que ocurría en el problema (3.32) para las ligaduras explícitas en C2, la violación aquí de la condición implícita $p_k(\mathbf{g}) \geq 0$ activa inmediatamente el multiplicador de Lagrange implícito asociado (denótese por α_k). Una vez que α_k^* no es nulo y puesto que las condiciones de debilidad complementaria exigen que $\alpha_k^* p_k^*(\mathbf{g}) = 0$, se deduce inmediatamente que $p_k^*(\mathbf{g}) = 0$. El hecho de que resulte imposible que la solución óptima en (3.43) tome un valor negativo se refleja de forma compacta en el uso del operador $[\cdot]^+$. En el siguiente capítulo se resolverán problemas de optimización que deberán tener en cuenta de forma explícita este tipo de ligaduras.

(S6.2) Cree una tabla que contenga todos los vectores \mathbf{m}^r cuyas componentes tengan valores decrecientes y que satisfagan de forma ajustada (i.e., como una igualdad) la restricción C2. Empareje esta tabla con la obtenida en el paso (S6.1) para las ganancias de sub-portadoras (i.e., si $g_{k_1} \geq g_{k_2}$, entonces $M_{k_1}^r \geq M_{k_2}^r$).

(S6.3) Para cada \mathbf{m}^r de la tabla, evalúe en base a (3.43) el vector \mathbf{p}^* y seleccione como solución óptima la pareja $(\mathbf{m}^{r*}, \mathbf{p}^*)$ que entrañe un coste $J(\mathbf{p}) = \sum_{k=1}^K p_k$ mínimo.

El proceso de ordenación en (S6.1) confina drásticamente la búsqueda sobre \mathcal{M} (consúltese [70] para una demostración matemática). La intuición detrás de esta decisión consiste en que, a fin de minimizar $J(\mathbf{p})$, los mayores valores de M_k^r deberán ser asignados a las sub-portadoras con mayores g_k .

Comparando la adaptación del esquema propuesto con el principio de “water-filling” (que maximiza la capacidad sujeta a una restricción de potencia [18] y que dicta que potencia y tasa son valores crecientes c.r.a. la ganancia del canal) se comprueba que para el algoritmo 3.6, aunque parecido, el comportamiento no es idéntico puesto que: (i) M_k^r crece o se mantiene si g_k crece (esto se debe a que M_k^r es una variable entera y el crecimiento de su valor debe ser escalonado) y (ii) para los casos en que el crecimiento de g_k sea lo suficientemente grande como para hacer crecer M_k^r , la potencia p_k crecerá también, pero para los casos en que g_k crece y M_k^r se mantenga constante, la potencia p_k disminuirá [cf. (3.43)].

Observación 3.3 El algoritmo 3.6 presenta una complejidad moderada si K y en especial \tilde{r} toman valores medios-bajos, lo cual resulta razonable para sistemas limitados en potencia donde los requisitos de tasa no suelen ser elevados. En caso contrario, la tabla creada en el paso (S6.2) podría disparar su tamaño y las necesidades de cómputo serían grandes. Aunque no se vayan a derivar en esta tesis, sí conviene señalar que en aquellos casos en los que la capacidad computacional sea un factor limitador y la dimensionalidad del problema sea elevada, el espacio

de búsqueda sobre \mathcal{M} puede confinarse todavía más mediante el uso de esquemas de ordenación –similares a los de los pasos (S6.1) y (S6.2)– que tengan en cuenta no sólo la descripción (tendencia) de primer orden sino la estructura completa de la función BER.

Observación 3.4 *Existen diversas técnicas de MIP que, a diferencia de la aquí propuesta, considerando simultáneamente $C1$ y $C2$ podrían utilizarse para resolver el problema de optimización en (3.41). Sin embargo, la consideración conjunta de $C1$ y $C2$ incrementa la dimensionalidad del problema y la carga computacional. Como ejemplo puede tomarse el algoritmo “branch-and-bound with Lagrangian relaxation” [9] cuya aplicación a distintos problemas de comunicaciones es exitosa. La utilización de esta técnica para el problema en (3.41) requeriría la resolución de un conjunto de sub-problemas que no son convexos, aumentando considerablemente la complejidad c.r.a. la del algoritmo 3.6.*

Observación 3.5 *A pesar de que la relación de potencia de pico a potencia media (PAPR) es un parámetro relevante para el consumo de sistemas OFDM, en los problemas de optimización planteados no se han incluido restricciones relacionadas con PAPR. Esto se debe al hecho de que, existen en la literatura esquemas de pre-distorsión (tales como el algoritmo de “selected mapping” en [73]) que aplicados a cada bloque de símbolos \mathbf{a} satisfacen requisitos de PAPR y que, por lo tanto, no dependen de la asignación posterior de bits y potencia por sub-portadora.*

3.2.3. Adaptación y cuantificación óptimas basadas en Q-CSIT

La minimización de la potencia transmitida por el sistema OFDM en este apartado se realizará no basándose en el vector \mathbf{g} que contiene las realizaciones del canal para las distintas sub-portadoras, sino el vector de bits \mathbf{c} que para la realización considerada contiene los índices de las regiones de cuantificación a las que las ganancias pertenece. En este contexto, se supondrá que:

(sp4) La palabra c_k (k -ésima componente de \mathbf{c}) tiene longitud $B \forall k$; por consiguiente, \mathbf{c} es un vector de tamaño $KB \times 1$ cuyos elementos toman el valor 0, 1.

Estos B bits por sub-portadora indexarán 2^B regiones diferentes $\mathcal{R}_{k,l} := [\tau_{k,l}, \tau_{k,l+1})$ que, en forma de intervalos consecutivos, cubren el espacio de variación de g_k y que por cada sub-portadora $k \in \{1, \dots, K\}$ puede ser equivalentemente representadas por los umbrales de cuantificación $[\tau_{k,1}, \dots, \tau_{k,2^B-1}]^T := \boldsymbol{\tau}_k$. Nótese que el índice $l \in \{0, \dots, 2^B\}$, que $\tau_{k,0} = 0$ y $\tau_{k,2^B} = \infty \forall k$ y que en general $\boldsymbol{\tau}_k \neq \boldsymbol{\tau}_{k'}$.

La minimización de potencia en base a Q-CSIT se realizará en dos fases: (i) obtención de los umbrales de cuantificación óptimos $\boldsymbol{\tau}^* := \{\boldsymbol{\tau}_k^*\}_{k=1}^K$ que se basará y se realizará durante la fase preliminar de establecimiento de la comunicación (i.e., “off-line”); y (ii) derivación de los vectores de adaptación de potencia y tasa óptimos $(\mathbf{p}^*, \mathbf{m}^*)$, esta adaptación se realizará en tiempo real para cada realización del canal (i.e., “on-line”) y tomará como base el vector \mathbf{c} que contendrá el Q-CSI generado de acuerdo al cuantificador obtenido en la fase (i).

Fase “off-line”: diseño del cuantificador

El objetivo de este apartado consiste en derivar umbrales de cuantificación para las VA g_k . El cómputo y valor final de los umbrales dependerá de la medida de distorsión utilizada. Siguiendo la metodología propuesta en esta disertación, puesto que el objetivo final consiste en minimizar la potencia transmitida, una posible forma de diseño consiste en plantear un problema de cuantificación con restricciones donde los umbrales se obtienen como la solución de un problema de minimización (como se verá de forma más detallada en el siguiente capítulo la medida de distorsión en ese caso la constituye la función dual de Lagrange asociada al problema de optimización). No obstante, la función objetivo final que se desea minimizar es una función instantánea mientras que el diseño de los umbrales de cuantificación requiere el uso de una medida ponderada en el espacio de la VA¹⁶, lo que hace que no sea posible aplicar de manera directa el diseño propuesto (nótese

¹⁶La ergodicidad del canal garantiza que el promedio realizado c.r.a. espacio de g_k sea equivalente al realizado c.r.a. tiempo.

que si los umbrales pudieran calcularse para cada realización del canal la expresión óptima de los mismos sería trivial $\tau_{k,l}^* = g_k, \forall l$. Puesto que las prestaciones del cuantificador deben medirse de forma promediada, el cuantificador de canal será diseñado con la finalidad de minimizar la potencia transmitida media $\sum_{k=1}^K \bar{p}_k$ sujeta a restricciones (también medias) de tasa y BER.

A fin de evaluar \bar{p}_k debe considerarse que p_k dependerá de la versión cuantificada de la VA continua g_k que toma 2^B valores diferentes, de esta forma p_k puede entenderse como una VA discreta que puede tomar valores $\{p_{k,l}\}_{l=0}^{2^B-1}$ cada uno de ellos con probabilidad $\Pr(p_k = p_{k,l}) = \Pr(g_k \in \mathcal{R}_{k,l}) := \int_{\tau_{k,l}}^{\tau_{k,l+1}} f_{g_k}(g_k) dg_k = \exp(-\tau_{k,l}/\bar{g}_k) - \exp(-\tau_{k,l+1}/\bar{g}_k)$, donde la última igualdad viene del hecho de que, de acuerdo a (sp2), g_k se distribuye de forma exponencial. En este punto la potencia media puede escribirse como $\bar{p}_k = \sum_{l=0}^{2^B-1} p_{k,l} \Pr(p_k = p_{k,l})$, expresión que muestra claramente la dependencia de la potencia media c.r.a. los umbrales y la FDP del canal y que, por lo tanto, constituye el primer paso para abordar el diseño óptimo del cuantificador. Argumentando de forma similar y teniendo en cuenta que g_k es una VA continua y p_k una VA discreta, la BER media por sub-portadora puede escribirse como: $\bar{\epsilon}_k(\boldsymbol{\tau}_k, \{p_{k,l}\}, M_k^r) = \sum_{l=0}^{2^B-1} \int_{\tau_{k,l}}^{\tau_{k,l+1}} \epsilon_k(g_k, p_{k,l}, M_k^r) f_{g_k}(g_k) dg_k$. Nótese que para la región correspondiente a $l = 0$ se tiene $p_{k,0} = \epsilon_k = 1$.

Utilizando la potencia media como función objetivo e incluyendo la tasa y BER en las restricciones del problema, los umbrales de cuantificación pueden obtenerse como la solución del siguiente problema de minimización con ligaduras:

$$\left\{ \begin{array}{l} \text{mín}_{\tau_{k,l}(B, \bar{\mathbf{g}}) \geq \mathbf{0}} J(\bar{\mathbf{p}}), \quad J(\bar{\mathbf{p}}) := \sum_{k=1}^K \bar{p}_k(B, \bar{\mathbf{g}}) \\ \text{s. a : C1. } \sum_{k=1}^K \frac{\log_2 M_k^r}{\check{r}} \bar{\epsilon}_k(\boldsymbol{\tau}_k, \{p_{k,l}\}, M_k^r) \leq \check{\epsilon} \\ \text{C2. } \sum_{k=1}^K \log_2 M_k^r \geq \check{r}. \end{array} \right. \quad (3.45)$$

Como en el caso de optimización basada en P-CSIT, se resolverá primero (3.45) sin tener en cuenta la restricción C2. Sustituyendo las expresiones de \bar{p}_k y $\bar{\epsilon}_k$ en (3.45) y definiendo $\theta_{k,l} := v_k p_{k,l} + 1/\bar{g}_k$, la función objetivo y la restricción de BER

expresada en C1 para el problema simplificado pueden escribirse como:

$$J(\bar{\mathbf{p}}) = \sum_{k=1}^K \sum_{l=1}^{2^B-1} p_{k,l} \left(e^{-\frac{\tau_{k,l}}{g_k}} - e^{-\frac{\tau_{k,l+1}}{g_k}} \right). \quad (3.46)$$

$$\text{C1.} \quad \sum_{k=1}^K \frac{\kappa_1 \log_2 M_k^r}{\tilde{r} \tilde{g}_k} \sum_{l=1}^{2^B-1} \frac{e^{-\theta_{k,l} \tau_{k,l}} - e^{-\theta_{k,l} \tau_{k,l+1}}}{\theta_{k,l}} \leq \check{\epsilon}. \quad (3.47)$$

Observación 3.6 *Mientras que a efectos de optimalidad las condiciones KKT son necesarias, el carácter de óptimo global requiere también la convexidad de la función objetivo y de las ligaduras [32]. De las tres condiciones exigidas a $J(\bar{\mathbf{p}}) = \sum_{k=1}^K \sum_{l=0}^{2^B-1} p_{k,l} \int_{\tau_{k,l}}^{\tau_{k,l+1}} f_{g_k}(g_k) dg_k$ para ser convexo c.r.a. $p_{k,l}$ y $\tau_{k,l}$ (a saber, $\frac{\partial^2 \bar{P}}{\partial p_{k,l}^2} \geq 0$, $\frac{\partial^2 \bar{P}}{\partial \tau_{k,l}^2} \geq 0$, y $\frac{\partial^2 \bar{P}}{\partial p_{k,l}^2} \frac{\partial^2 \bar{P}}{\partial \tau_{k,l}^2} - \left[\frac{\partial \bar{P}}{\partial p_{k,l} \partial \tau_{k,l}} \right]^2 \geq 0$, [32, apéndice 1]), es sencillo comprobar que $\frac{\partial^2 \bar{P}}{\partial p_{k,l}^2} = 0$ y, por lo tanto, la tercera condición no se satisface. Esto hace que $J(\bar{\mathbf{p}})$ no sea convexo y, por lo tanto, sólo existen garantías de que la solución “off-line” sea un óptimo local.*

Enmarcando el problema de optimización de (3.45) dentro de la teoría clásica de cuantificación, el comportamiento descrito es el típico para cuantificadores vectoriales (consúltese, e.g., [23]) donde la optimalidad global puede únicamente demostrarse para medidas de distorsión uni-modales [83]. No obstante, el comportamiento de los cuantificadores de carácter local es en general satisfactorio, ejerciendo en ocasiones incluso de óptimo global (bien porque es único, bien porque las prestaciones de distintos óptimos locales es semejante).

La función dual de Lagrange (o simplemente Lagrangiano) asociada a la minimización de (3.46) sujeta a (3.47) puede escribirse como

$$\begin{aligned}
 \mathcal{L}(\beta_Q) = & \dots + p_{k,l-1} \int_{\tau_{k,l-1}}^{\tau_{k,l}} f_{g_k}(g_k) dg_k + p_{k,l} \int_{\tau_{k,l}}^{\tau_{k,l+1}} f_{g_k}(g_k) dg_k + \dots \\
 & + \beta_Q \left\{ \sum_k \frac{\kappa_1 \log_2 M_k^r}{\tilde{r}} \left[\dots + \int_{\tau_{k,l-1}}^{\tau_{k,l}} e^{-v_k g_k p_{k,l-1}} f_{g_k}(g_k) dg_k \right. \right. \\
 & \left. \left. + \int_{\tau_{k,l}}^{\tau_{k,l+1}} e^{-v_k g_k p_{k,l}} f_{g_k}(g_k) dg_k + \dots \right] - \check{\epsilon} \right\}, \tag{3.48}
 \end{aligned}$$

τ donde β_Q es el multiplicador de Lagrange asociado a la restricción en C1. Definiendo $\psi(\theta_{k,l}, \tau_{k,l}) := (1 + \theta_{k,l} \tau_{k,l}) e^{-\theta_{k,l} \tau_{k,l}}$, las condiciones KKT correspondientes a $\partial \mathcal{L} / \partial \tau_{k,l} = 0$ y $\partial \mathcal{L} / \partial p_{k,l} = 0$, dictan que (para $k = 1, \dots, K$ y $l = 1, \dots, 2^B - 1$)

$$p_{k,l-1} - p_{k,l} + \beta_Q \frac{\kappa_1}{\tilde{r}} \log_2(M_k^r) [e^{-v_k \tau_{k,l} p_{k,l-1}} - e^{-v_k \tau_{k,l} p_{k,l}}] = 0, \tag{3.49}$$

$$e^{-\frac{\tau_{k,l}}{\bar{g}_k}} - e^{-\frac{\tau_{k,l+1}}{\bar{g}_k}} - \frac{\beta_Q \kappa_1 v_k \log_2 M_k^r}{\tilde{r} \bar{g}_k \theta_{k,l}^2} [\psi(\theta_{k,l}, \tau_{k,l}) - \psi(\theta_{k,l}, \tau_{k,l+1})] = 0. \tag{3.50}$$

Este conjunto de $2(2^B K)$ ecuaciones puede ser representado de forma compacta a través de funciones no lineales cuyos argumentos son las potencias y umbrales de cuantificación. De esta manera $\mathcal{Z}_1(p_{k,l-1}, p_{k,l}, \tau_{k,l}^*) = 0$ representará la ecuación (3.49) y $\mathcal{Z}_2(p_{k,l}, \tau_{k,l}^*, \tau_{k,l+1}^*) = 0$ hará lo propio con (3.50). Utilizando esta notación, los umbrales óptimos $\tau_{k,l}^*$ se obtendrán como resultado de aplicar el siguiente algoritmo:

Algoritmo 3.7 (Q-CSIT)

(S7.1.) Ordene las sub-portadoras de forma decreciente en base a \bar{g}_k y genere una tabla de vectores \mathbf{m}^r ordenados que satisfagan la restricción de tasa de forma análoga a lo realizado en los pasos (S6.1) y (S6.2) del algoritmo 3.6. Finalizadas estas tareas, ejecute el paso (S7.2) para cada entrada \mathbf{m}^r .

(S7.2.) Dado un valor de β_Q ;

- Para $k = 1 : K$
 - Para $\tau_{k,1} = 0 : \tau_{\text{máx}}$, donde $\tau_{\text{máx}}$ se calcula de manera que, e.g., $\Pr(g_k > \tau_{\text{máx}}) = 0,01$;
Dado un valor de $\tau_{k,1}$, resuelva numéricamente (3.49) (e.g., a través de una búsqueda lineal basada en el método de bisección) y obtenga una raíz $p_{k,1}$ (recuerde que $p_{k,0} = 0$). En base al valor obtenido de $p_{k,1}$, resuelva (3.50) c.r.a. $\tau_{k,2}$. De manera análoga, incrementando el índice l y alternando entre (3.49) y (3.50) obtenga $\{p_{k,l}\}$ y $\{\tau_{k,l}\}$ para $l = 1, \dots, 2^B - 1$. Puesto que los valores de potencia y umbrales se calcula sin hacer uso de (3.50) para $l = 2^B - 1$, utilice esta ecuación para comprobar la viabilidad del valor inicial de $\tau_{k,1}$ y, por consiguiente, de la solución asociada calculada.
 - Fin del lazo en $\tau_{k,1}$.
 - Fin del lazo en k .
 - En este punto se tiene una pareja de valores $\boldsymbol{\tau}, \mathbf{p}$ que, para el valor de β_Q considerado, satisfacen las condiciones (3.49) y 3.50). Compruebe si la solución candidata satisfacer de forma ajustada la restricción de BER en (3.47); en caso de que lo haga repita almacene la solución y repita el paso (S7.2) para una nueva entrada \mathbf{m}^r ; en caso de que no lo haga, mantenga \mathbf{m}^r , elimine la solución candidata y modifique el valor de β_Q .

Fin del lazo en β_Q .

(S7.3.) De entre todas las soluciones $\boldsymbol{\tau}$ válidas, escoja como óptima aquella que tenga asociado un coste ($J(\bar{\mathbf{p}})$ en (3.46)) mínimo.

En el paso (S7.1), cada selección de $\{M_k^r\}_{k=1}^{K_a}$ satisface: (i) $K_a \leq K$ y (ii) $M_{k_1}^r \geq M_{k_2}^r$ si $\bar{g}_{k_1} \geq \bar{g}_{k_2}$. Esto significa que el algoritmo 3.7 genera de forma directa umbrales únicamente para las K_a sub-portadoras que son “mejores en

media”. Puesto que el cuantificador de canal se aplicará para un escenario de adaptación instantánea (por cada símbolo), necesitará que los umbrales se definan para todas las sub-portadoras. Los resultados obtenidos al aplicar del algoritmo 3.7 sobre un extenso grupo de casos de estudio, sugieren que para un mismo índice de región l la relación $\tau_{k,l}/\bar{g}_k$ es prácticamente constante para cualquier índice de sub-portadora $k \in \mathcal{K}_a$. De esta forma un sencillo método para extender la definición del cuantificador consiste en considerar los umbrales $\tau_{k',l} = \mathbb{E}_{k \in \mathcal{K}_a}[\tau_{k,l}/\bar{g}_k]\bar{g}_{k'}$ para aquellos $k' \notin \mathcal{K}_a$. Esta interesante característica será además aprovechada en el siguiente apartado para reducir la complejidad asociada al diseño del cuantificador.

En el paso (S7.2), dado un valor inicial de β_Q , existen diversos métodos que pueden utilizarse para actualizar el valor del multiplicador β_Q [9, capítulo 4]. Escogiendo, como ya se hizo en la sección anterior, el método de los multiplicadores, la actualización para cada iteración en el paso (S7.2) se calcula como $\beta_Q^{(i+1)} = \beta_Q^{(i)} + c^{(i)} \left[\sum_{k=1}^K \frac{\log_2 M_k^r}{\tilde{\gamma}} \bar{\epsilon}_k(\boldsymbol{\tau}_k, \{p_{k,l}\}, M_k^r) - \tilde{\epsilon} \right]$ donde i denota el índice de iteración y c es un parámetro de penalización que puede ser adaptado en función de las necesidades de convergencia.

Nótese asimismo que aunque el problema inicial expresado en (3.49) y (3.50) consistía en resolver un sistema de $2(2^B K)$ ecuaciones no lineales de forma numérica, el método de resolución propuesto en el algoritmo 3.7 reduce esta tarea a K búsquedas unidimensionales cada una de las cuales requiere la realización secuencial de otras 2^{B+1} búsquedas unidimensionales, lo que supone una disminución drástica de la complejidad computacional requerida.

Observación 3.7 *Además de los umbrales, el algoritmo 3.7 genera como resultado los parámetros $(\mathbf{m}^r, \mathcal{K}_a, \beta_Q)$ que minimizan la potencia media satisfaciendo requisitos de BER media. Los valores $(\mathbf{m}^r, \mathcal{K}_a, \beta_Q)$ pueden utilizarse en algoritmos que, aunque sub-óptimos, reduzcan considerablemente la complejidad asociada a los esquemas de adaptación “on-line”. Estos algoritmos serán analizados en el apartado 3.2.4, no obstante, primero deberá derivarse el esquema óptimo de adaptación “on-line” que, tomando como entrada la palabra $\mathbf{c} = \mathbf{c}(\mathbf{g})$ construida de acuer-*

do a los umbrales aquí derivados, genera los vectores de tasa y potencia $(\mathbf{m}^r, \mathbf{p})$ que minimizan la potencia a la par que satisfacen requisitos de QoS de forma instantánea.

Fase “on-line”: adaptación de tasa y potencia

Con los parámetros del cuantificador (umbrales y por consiguiente regiones) dados, el receptor estima cada realización \mathbf{g} y la cuantifica para obtener el vector de bits \mathbf{c} . Recibida \mathbf{c} (Q-CSIT), el transmisor deberá calcular de forma “on-line” la pareja $(\mathbf{m}^r, \mathbf{p})$ que, para el símbolo considerado, minimiza la potencia satisfaciendo restricciones de tasa y de lo que se llamará BER condicionada. Definiendo $\theta_k := v_k p_k + 1/\bar{g}_k$, esta última podrá expresarse como:

$$\bar{\epsilon}_{k|\mathbf{c}} := \frac{\int_{\tau_{k,l}}^{\tau_{k,l+1}} \epsilon_k(p_k, g_k) f_{g_k}(g_k) dg_k}{\text{Pr}(g_k \in \mathcal{R}_{k,l})} \quad (3.51)$$

$$= \frac{\kappa_1 (e^{-\theta_k \tau_{k,l}} - e^{-\theta_k \tau_{k,l+1}})}{\bar{g}_k \theta_k (e^{-\tau_{k,l}/\bar{g}_k} - e^{-\tau_{k,l+1}/\bar{g}_k})}; \quad (3.52)$$

i.e., $\bar{\epsilon}_{k|\mathbf{c}}$ corresponde al valor esperado de la BER para el símbolo OFDM *dado el hecho de que* (condicionado a que) la versión cuantificada del canal \mathbf{g} es \mathbf{c} (el numerador de (3.51) promedia la BER sobre todas las posibles realizaciones del canal que pertenecen a la región $\mathcal{R}_{k,l}$, mientras que el denominador no es más que la probabilidad de que la realización del canal pertenezca a $\mathcal{R}_{k,l}$). En base a la BER condicionada, el problema de optimización del que se derivan los esquemas óptimos de adaptación puede formularse como:

$$\left\{ \begin{array}{l} \text{mín}_{\mathbf{m}^r(\mathbf{c}) \in \mathcal{M}, \mathbf{p}(\mathbf{c}) \geq \mathbf{0}} J(\mathbf{p}), \quad J(\mathbf{p}) := \sum_{k=1}^K p_k(\mathbf{c}) \\ \text{s. a : C1. } \sum_{k=1}^K \frac{\log_2 M_k^r}{\tilde{r}} \bar{\epsilon}_{k|\mathbf{c}}(\tau_{k,l}, \tau_{k,l+1}, p_k, M_k^r | \mathbf{c}) \leq \check{\epsilon} \\ \text{C2. } \sum_{k=1}^K \log_2 M_k^r \geq \check{r}. \end{array} \right. \quad (3.53)$$

Comparando (3.53) con (3.45), se deduce que, aparte del hecho de que el diseño del cuantificador esté dado, la estructura de ambos problemas es semejante. Puesto que el valor escalar del denominador en (3.51) simplemente escala el valor del

multiplicador β_Q , una versión simplificada del algoritmo 3.7 que no necesita buscar sobre el valor de los umbrales y que no tiene en cuenta (3.49) puede aplicarse también para resolver (3.53). Por otro lado, debido a que el objetivo aquí consiste en encontrar los valores óptimos de $(\mathbf{m}^r, \mathbf{p})$ para cada realización del canal (al igual que en el algoritmo 3.6 y al contrario que en el algoritmo 3.7 cuyo objetivo era el cálculo de umbrales) se mostrará que el problema es convexo y que, por lo tanto, las condiciones KKT dan lugar a un óptimo global. A fin de probar convexidad, puesto que no existe dependencia cruzada entre la potencia de distintas sub-portadoras, basta con demostrar que $\partial^2 J(\mathbf{p})/\partial p_k^2 \geq 0$ y $\partial^2 \bar{\epsilon}_{|c}/\partial p_k^2 \geq 0$. La satisfacción de la primera condición es trivial, mientras que la de la segunda se verifica de la forma:

$$\frac{\partial^2 \bar{\epsilon}_{|c}}{\partial p_k^2} = \frac{\kappa_1 \log_2 M_k^r}{\check{r} (e^{-\tau_{k,l}/\bar{g}_k} - e^{-\tau_{k,l+1}/\bar{g}_k})} \int_{\tau_{k,l}}^{\tau_{k,l+1}} (v_k g_k)^2 e^{-\theta_k g_k} dg_k \geq 0, \quad (3.54)$$

donde la última desigualdad se debe al hecho de que $\tau_{k,l} < \tau_{k,l+1}$ y el integrando sea mayor o igual que cero.

3.2.4. Reducción de la complejidad de la optimización basada en Q-CSIT

Cuando se calcula el diseño óptimo del cuantificador, en cada iteración del algoritmo 3.7 (correspondiente a cada posible valor ordenado de \mathbf{m}^r y un valor de β_Q), debe encontrarse el valor de $2^B - 1$ umbrales $\tau_{k,l}$ y $2^B - 1$ niveles de potencia $p_{k,l}$ por cada sub-portadora, lo que supone la búsqueda de $2K(2^B - 1)$ incógnitas para cada par (\mathbf{m}^r, β_Q) . A pesar de que, como ya se mencionó tras la presentación del algoritmo 3.7, se pueda aprovechar la estructura de la solución para no recurrir a una búsqueda exhaustiva, para valores elevados de K , B y \check{r} la complejidad computacional necesaria puede ser muy elevada. Este comportamiento motiva el desarrollo en este apartado de técnicas que, si bien sub-óptimas, reducen la complejidad asociada a los algoritmos de cuantificación y adaptación.

Alternativas para la reducción de la complejidad

Un aspecto clave para la reducción de la complejidad computacional consiste en disminuir el número de incógnitas. La idea para conseguir este objetivo consiste en expresar los niveles de potencia $p_{k,l}$ en términos de los umbrales de cuantificación $\tau_{k,l}$ (lo que, como primera consecuencia, reduce el número de incógnitas a la mitad). Para obtener dicha relación, se aprovechará la estructura de la asignación óptima de potencia $p_k^*(g_k)$ para el caso de P-CSIT [cf., (3.43)] y se analizarán cuatro *alternativas* diferentes para la *reducción de la complejidad* (RC):

RC1. Estimar la ganancia cuantificada mediante $\hat{g}_{k,l} = (\tau_{k,l} + \tau_{k,l+1})/2$ y utilizar (3.43) para obtener la potencia como $p_{k,l} = p_k(\hat{g}_{k,l})$.

RC2. Utilizar la FDP de para estimar la ganancia de la k -ésima sub-portadora como

$$\begin{aligned} \hat{g}_{k,l} &= \mathbb{E}_{g_k}[\hat{g}_k|l] = \mathbb{E}_{g_k}[\hat{g}_k|\tau_{k,l} < g_k < \tau_{k,l+1}] = \frac{1}{e^{-\theta_{k,l}/\bar{g}_k} - e^{-\theta_{k,l+1}/\bar{g}_k}} \\ &\times [\tau_{k,l}e^{-\tau_{k,l}/\bar{g}_k} - \tau_{k,l+1}e^{-\tau_{k,l+1}/\bar{g}_k}\bar{g}_k(e^{-\tau_{k,l}/\bar{g}_k} - e^{-\tau_{k,l+1}/\bar{g}_k})], \end{aligned} \quad (3.55)$$

y emplear nuevamente (3.43) para obtener los niveles de la potencia $p_{k,l} = p_k(\hat{g}_{k,l})$.

RC3. Evitar la estimación del valor cuantificado de g_k estimando directamente el valor cuantificado de la potencia (véase [19])

$$p_{k,l} = \frac{1}{2}p_k(g_k)|_{g_k=\tau_{k,l}} + \frac{1}{2}p_k(g_k)|_{g_k=\tau_{k,l+1}}. \quad (3.56)$$

RC4. Utilizar, como en RC2, la FDP de la ganancia para estimar el valor cuantificado de la potencia de una forma más precisa: $p_{k,l} = \mathbb{E}_{g_k}[p_k(g_k)|\tau_{k,l} < g_k < \tau_{k,l+1}]$, que puede ser calculado empleando (3.43) e integración numérica.

Entre estas cuatro alternativas, RC1 (punto medio de la región de cuantificación) es la más simple, pero como confirmarán las simulaciones numéricas, también es

la que conduce a los diseños de mayor consumo de potencia. Por otro lado, los diseños basados en RC4 (potencia esperada) son ligeramente mejores desde el punto de vista de consumo que los basados en RC2 y RC3 pero entrañan una complejidad considerablemente mayor y no tienen la ventaja añadida de presentar una expresión analítica de $p_{k,l}$ en términos de $\tau_{k,l}$. Finalmente, la superioridad de RC3 (punto medio de la región de potencia) frente a R2 (ganancia esperada) hace que sea RC3 la alternativa más prometedora cuando se tiene en cuenta conjuntamente la complejidad y el consumo de potencia. Por esta causa, de ahora en adelante se utilizará RC3 como alternativa por defecto para relacionar $p_{k,l}$ y $\tau_{k,l}$. Sustituyendo (3.43) en (3.56), $p_{k,l} = p_{k,l}(\tau_{k,l}, \tau_{k,l+1})$ puede escribirse como

$$p_{k,l} = \frac{1}{2v_k} \left\{ \frac{[\ln(v_k \tau_{k,l} \log_2(M_k^r) \beta_1)]^+}{\tau_{k,l}} + \frac{[\ln(v_k \tau_{k,l+1} \log_2(M_k^r) \beta_1)]^+}{\tau_{k,l+1}} \right\}, \quad (3.57)$$

donde β_1 es una constante (que estará relacionada con la satisfacción de la restricción de BER) que sustituye a β_D y que, como se verá más tarde, deberá ser calculada de forma conjunta con el multiplicador de Lagrange β_Q correspondiente al algoritmo original de optimización para Q-CSIT. Basándose en (3.57), se simplificará la fase “off-line” (optimizando únicamente c.r.a. $\tau_{k,l}$) así como la fase “on-line” (reduciendo el problema de optimización a una búsqueda unidimensional sobre β_1).

Cuantificadores de complejidad reducida

En este punto se propondrán dos cuantificadores de canal sub-óptimos que reducirán la complejidad de la fase asociada al diseño del cuantificador.

A) Umbrales Comunes (UC) para todas las sub-portadoras

En este punto, para simplificar el diseño del cuantificador, se elimina la dependencia del umbral c.r.a. a la sub-portadora, es decir, se elimina el sub-índice k y se utiliza $\tau_{k,l} = \tau_l$ para todos los valores de k . Este diseño reduce la dimensionalidad del espacio de búsqueda por un factor K (lo que hace que por cada iteración del algoritmo 3.7 se tengan que calcular únicamente $2^B - 1$ incógnitas). La apli-

cación directa de las condiciones KKT sobre el problema simplificado permite la obtención de los $2^B - 1$ umbrales comunes (UC) óptimos como la solución de las siguientes ecuaciones (para $l = 1, \dots, 2^B - 1$):

$$\begin{aligned} & \sum_{k=1}^K \rho_{k,l}^{(j)} (e^{-\tau_l/\bar{g}_k} - e^{-\tau_{l+1}/\bar{g}_k}) - \frac{p_{k,l}}{\bar{g}_k} e^{-\tau_l/\bar{g}_k} + \rho_{k,l-1}^{(l)} (e^{-\tau_{l-1}/\bar{g}_k} - e^{-\tau_l/\bar{g}_k}) + \frac{p_{k,l-1}}{\bar{g}_k} e^{-\tau_l/\bar{g}_k} \\ & - \frac{\beta_{UC}}{\check{r}} \sum_{k=1}^K \frac{\kappa_1 \log_2 M_k^r}{\bar{g}_k} \left\{ \frac{\rho_{k,l}^{(l)} v_k (e^{\theta_{k,l}\tau_l} - e^{\theta_{k,l}\tau_{l+1}})}{\theta_{k,l}^2} + \frac{(\theta_{k,l} - \tau_l v_k \rho_{k,l}^{(l)}) e^{\theta_{k,l}\tau_l}}{\theta_{k,l}} \right. \\ & + \frac{\tau_{l+1} v_k \rho_{k,l}^{(l)} e^{\theta_{k,l}\tau_{l+1}}}{\theta_{k,l}} + \frac{\rho_{k,l-1}^{(l)} v_k (e^{\theta_{k,l-1}\tau_{l-1}} - e^{\theta_{k,l-1}\tau_l})}{\theta_{k,l-1}^2} \\ & \left. - \frac{(\theta_{k,l-1} - \tau_l v_k \rho_{k,l-1}^{(l)}) e^{\theta_{k,l-1}\tau_l}}{\theta_{k,l-1}} + \frac{\tau_{l-1} v_k \rho_{k,l-1}^{(l)} e^{\theta_{k,l-1}\tau_{l-1}}}{\theta_{k,l-1}} \right\} = 0, \end{aligned} \quad (3.58)$$

donde β_{UC} es el multiplicador de Lagrange asociado al requisito de BER y $\rho_{k,i}^{(l)}$ se define como

$$\rho_{k,i}^{(l)} := \frac{\partial p_{k,i}}{\partial \tau_l} = \begin{cases} \frac{1 - \ln(v_k \tau_l \log_2(M_k^r) \beta_1)}{2 v_k \tau_l^2}; & \text{si } (i = l, l-1) \text{ y } (\beta_1 > \frac{1}{v_k \log_2(M_k^r) \tau_{k,l}}), \\ 0; & \text{en otro caso.} \end{cases} \quad (3.59)$$

Las iteraciones se ejecutan de forma similar al algoritmo 3.7 con la única excepción de que se necesitan dos lazos de repetición para la búsqueda de β_{UC} y β_1 (al estar ambos parámetros relacionados con la satisfacción del requisito de BER existe también la posibilidad de realizar una búsqueda conjunta). Tras encontrar la solución de los umbrales óptimos y del parámetro β_1 , los distintos $p_{k,l}$ pueden obtenerse utilizando (3.57). A pesar de que el esquema UC-QCSIT es sencillo, puede darse el caso de que no aproveche de forma eficiente la variabilidad estadística de las ganancias de sub-portadora. En el punto siguiente se propondrá un diseño para el cuantificador que, siendo todavía más simple, sí es capaz de explotar la diversidad estadística de las sub-portadoras.

B) Cuantificación independiente por sub-portadora (CIS)

Tras finalizar la descripción del algoritmo 3.7, se había hecho mención a que los valores de umbrales óptimos para distintos casos de estudio satisfacían la condición

heurística $\tau_{k,l} \simeq \vartheta_l \bar{g}_k$, done la constante ϑ_l dependía del índice de la región pero era independiente de la sub-portadora considerada. El objetivo de esta sección consistirá en proponer y caracterizar diseños de cuantificadores de canal que satisfagan esta condición, es decir que sigan un principio de *cuantificación independiente por sub-portadora* (CIS). Estos diseños presentan además la ventaja añadida sobre el cuantificador de UC de considerar de forma natural la diferencia estadística entre sub-portadoras, puesto que portadoras con distintas ganancias medias producirán distintos umbrales de cuantificación. Dos serán los diseños basados en el principio de CIS que se propondrán: uno de ellos impondrá regiones de cuantificación equi-probables (EP), mientras que el otro minimizará el error cuadrático medio (o *minimum mean square error* –MMSE–) asociado a la cuantificación bien de las ganancias por sub-portadora (diseño al que se referirá como cuantificador Lloyd- g_k) o bien de la potencia transmitida por sub-portadora (diseño al que se referirá como cuantificador Lloyd- p_k).

(B.i) *Cuantificador EP* : En este caso, los umbrales $\tau_{k,l}$ se diseñan de forma que $\Pr(g_k \in \mathcal{R}_{k,l}) = \int_{\tau_{k,l}}^{\tau_{k,l+1}} f_{g_k}(g_k) dg_k = 1/2^B \forall k, l$. Considerando la FDP de la ganancia descrita en (sp2) esta condición de diseño conduce a la siguiente expresión analítica para los umbrales

$$\tau_{k,l} = \bar{g}_k \ln \left(\frac{2^B}{2^B - l} \right) \quad (3.60)$$

que curiosamente, al igual que la solución óptima, satisface la condición $\tau_{k,l} \simeq \vartheta_l \bar{g}_k$ lo cual puede interpretarse como un indicador de la bondad del diseño propuesto¹⁷. Asimismo, la expresión analítica de los umbrales en (3.60) disminuye también la complejidad computacional durante la fase de adaptación “on-line” ya que la restricción de BER en (3.53) puede escribirse como

$$\bar{\epsilon}_{EP|c} = \frac{2^B \kappa_1}{\check{r}} \sum_{k=1}^K \frac{\log_2 M_k^r}{\bar{g}_k \theta_{k,l}} \left[\left(\frac{2^B - l}{2^B} \right)^{\bar{g}_k \theta_k} - \left(\frac{2^B - l - 1}{2^B} \right)^{\bar{g}_k \theta_k} \right]. \quad (3.61)$$

¹⁷Nótese además que el diseño EP es el que maximiza la entropía de la información c_k enviada a través canal de realimentación, puesto que hace que todas las posibles realizaciones de c_k tengan igual probabilidad [18].

(B.ii) *Cuantificador escalar de Lloyd* : En este caso, se adopta una estrategia de cuantificación por sub-portadora basada en el algoritmo escalar de Lloyd [23, 42], que para una VA x con FDP $f_x(x)$ utilizando 2^B niveles de cuantificación genera $\{\hat{x}_l\}_{l=0}^{2^B-1}$ minimizando la medida de distorsión

$$\mathbb{E}[(x - \hat{x}_l)^2]. \quad (3.62)$$

Los valores cuantificados \hat{x}_l así como los umbrales de cuantificación τ_l , se obtienen resolviendo iterativamente de forma alternativa las ecuaciones:

$$\tau_l = \frac{1}{2}(\hat{x}_l + \hat{x}_{l+1}) \quad (3.63)$$

$$\hat{x}_l = \frac{\int_{\tau_l}^{\tau_{l+1}} x f_x(x) dx}{\int_{\tau_l}^{\tau_{l+1}} f_x(x) dx}. \quad (3.64)$$

Al igual que las técnicas de reducción de la complejidad, el algoritmo de Lloyd puede aplicarse bien sobre la variable $x = g_k$ o bien sobre la variable $x = p_k$ (alternativas que serán comparadas en el apartado de simulaciones numéricas y a las que se referirá como cuantificadores Lloyd- g_k o Lloyd- p_k).

Observación 3.8 *El diseño de los cuantificadores basados en CIS se tiene en cuenta durante la fase de adaptación “on-line” que emplea (3.57) para obtener $(\mathbf{m}^r, \mathbf{p})$. Aunque los cuantificadores basados en CIS traten (consideren) a las sub-portadoras de forma individualizada, el algoritmo de adaptación “on-line” las tiene en cuenta de forma conjunta a través de las restricciones de tasa y BER, por lo que durante esta fase sí se aprovecha la diversidad estadística (debido al multi-camino) contenida en las sub-portadoras.*

Reducción de la complejidad de los esquemas de adaptación

Como se mencionó en la presentación de las alternativas de reducción de complejidad, (3.57) puede utilizarse en la fase “on-line” para simplificar la búsqueda necesaria para satisfacer C1 en (3.53) para cada \mathbf{m}^r . En la versión óptima de

la adaptación “on-line” del algoritmo 3.7, para cada valor de β_Q se realizan K_a búsquedas que, resolviendo (3.50), determinan la asignación de potencia, debiendo después realizar una búsqueda sobre el valor de β_Q hasta que la restricción C1 se satisface de forma. Sin embargo, si se utiliza (3.57), dado únicamente debe realizarse una búsqueda unidimensional sobre el parámetro β_1 para satisfacer C1, lo que supone una reducción drástica de la complejidad requerida.

Asimismo, como ya se mencionó en la observación 3.7, la versión “off-line” del algoritmo 3.7 genera como resultado valores de $(\mathbf{m}^r, \mathcal{K}_a, \beta_Q)$ sub-óptimos. Estos valores pueden utilizarse también para reducir la complejidad computacional durante la fase “on-line” de adaptación. En base a ellos, el transmisor puede adoptar una estrategia de adaptación basada en los siguientes modos (ordenados en orden decreciente de complejidad y, por lo tanto, en orden creciente en términos de consumo de potencia asociado):

Tx A : elegir de forma óptima el conjunto de sub-portadoras activas \mathcal{K}_a , la asignación de tasa \mathbf{m}^r y el valor del multiplicador de Lagrange, β_Q , obteniendo así el óptimo global $\mathbf{p}^*(\mathbf{c})$.

Tx B : elegir de forma óptima \mathcal{K}_a y β_Q , mientras que se utiliza un vector \mathbf{m}^r fijo (i.e., no se adapta “on-line” la tasa de transmisión por sub-portadora sino únicamente el conjunto \mathcal{K}_a que contiene las mejores K_a sub-portadoras).

Tx C : elegir de forma óptima \mathcal{K}_a , mientras que se utilizan valores fijos de \mathbf{m}^r y β_Q .

Tx D : elegir de forma óptima β_Q , mientras que se utilizan valores fijos de \mathbf{m}^r y \mathcal{K}_a (i.e., las sub-portadoras con mejores prestaciones medias).

La selección entre TxA-TxD será decisión del diseñador del sistema en función de las limitaciones específicas (consumo y complejidad) del mismo.

Tabla 3.5: Número de iteraciones necesarias para los distintos diseños (métodos) propuestos.

Fase	Método	Número de iteraciones sin RCx	Número de iteraciones con RCx
“Off-line” (Cuantif.)	Alg. 3.7 (“off”)	$2^B K_a N_M I_\beta N_\tau (I_P + I_\tau)$	–
	UC	$2^B K_a N_M I_\beta N_\tau (I_P + \frac{I_\tau}{K_a})$	$2^B N_M I_\beta N_\tau I_\tau$
	CIS-EP	$[2^B K_a,$ $2^B K_a (N_M I_\beta I_P + 1)]$	$[2^B K_a,$ $2^B K_a (N_M I_\beta + 1)]$
	CIS-Lloyd	$[2^B K_a I_L,$ $2^B K_a (N_M I_\beta I_P + I_L)]$	$[2^B K_a N_{Ll},$ $2^B K_a (N_M I_\beta + I_L)]$
“On-line” (Adapt.)	Alg. 3.7 (“on”)	$K_a N_M I_\beta I_P$	–
	Tx A	$K_a N_M I_\beta I_P$	$K_a N_M I_\beta$
	Tx B	$K_a I_\beta I_P$	$K_a I_\beta$
	Tx C	$K_a I_P$	K_a
	Tx D	$I_\beta I_P$	I_β

Análisis comparativo de complejidad

Sea N_M el número de posibles candidatos \mathbf{m}^r que satisfagan el requisito de tasa de transmisión, N_τ el número de valores candidatos para el primer umbral en cada sub-portadora, I_β el número de iteraciones necesarias hasta encontrar el valor del multiplicador de Lagrange que satisface de forma ajustada la restricción de BER, I_τ el número de iteraciones necesarias para realizar la búsqueda uni-dimensional que resuelve \mathcal{Z}_1 c.r.a. $\tau_{k,l}$, I_P el número de iteraciones necesarias para realizar la búsqueda uni-dimensional que resuelve \mathcal{Z}_2 c.r.a. $p_{k,l}$ e I_L el número de iteraciones necesarias para que el algoritmo de Lloyd converja [cf. (3.63) y (3.64)]. La tabla 3.5 resume la complejidad computacional asociada a los distintos diseños (métodos) propuestos, diferenciando explícitamente entre los que se ejecutan en tiempo real (“on-line”) y lo que no (“off-line”) y enfatizando la reducción potencial asociada al uso de las alternativas RC propuestas. Como ejemplo ilustrativo, tómesese el caso de UC implementando alternativas de RC (de entre las posibles RCx, concretamente se implementa RC3), donde se comprueba cómo la complejidad de la fase “off-line” se reduce considerablemente. Para el caso de CIS, los umbrales pueden ser obtenidos en solitario o en conjunción con $(\mathbf{m}^r, \mathcal{K}_a, \beta_Q)$ (para poder utilizar después estos

valores en la fase “on-line” en caso de que se estime necesario), para contemplar esta disyuntiva se ha introducido en la tabla 3.5 la notación $[x, y]$ representando x el número de iteraciones cuando únicamente se está interesado en el valor de los umbrales de cuantificación y representando y el número de iteraciones necesario cuando quieren obtenerse también los valores promedio de $(\mathbf{m}^r, \mathbf{p}, \beta_Q)$ para que un transmisor operando en los modos Tx-B, Tx-C or Tx-D pueda utilizarlos durante la fase de adaptación “on-line”. Por último, pero sin embargo probablemente uno de los resultados más significativos de este análisis, la tabla 3.5 muestra que la complejidad de la fase de adaptación puede resultar tan simple como seleccionar el conjunto de sub-portadoras activas o simplemente realizar una búsqueda numérica uni-dimensional sobre el parámetro β_1 .

Hasta este punto se han venido utilizando B bits de realimentación por cada sub-portadora, lo que hace que el canal de realimentación deba soportar una tasa de transmisión de KB bits por cada realización del canal. Este número puede resultar prohibitivo, especialmente para grandes valores de K y canales con escaso tiempo de coherencia (i.e., de alta variabilidad). Para poder manejar estos casos, en el siguiente apartado, se aprovechará la dependencia estadística a lo largo de sub-portadoras (especialmente marcada cuando $K \gg Q$, puesto que es Q la variable que fija la riqueza estadística del canal) y se introducirán técnicas de procesamiento de señal que permitirán reducir la tasa de realimentación del sistema.

3.2.5. Reducción de la tasa de realimentación para Q-CSIT

Puesto que los coeficientes de cálculo de la DFT son linealmente independientes, se puede comprobar fácilmente que $Q + 1$ valores complejos de la DFT pueden identificar con exactitud $Q + 1$ coeficientes complejos h_q . Basado en este hecho, en el apéndice 3.4.3 se muestra que:

Proposición 3.1 $(2Q + 1)B$ bits son suficientes para cuantificar el vector de ganancias de sub-portadora \mathbf{g} .

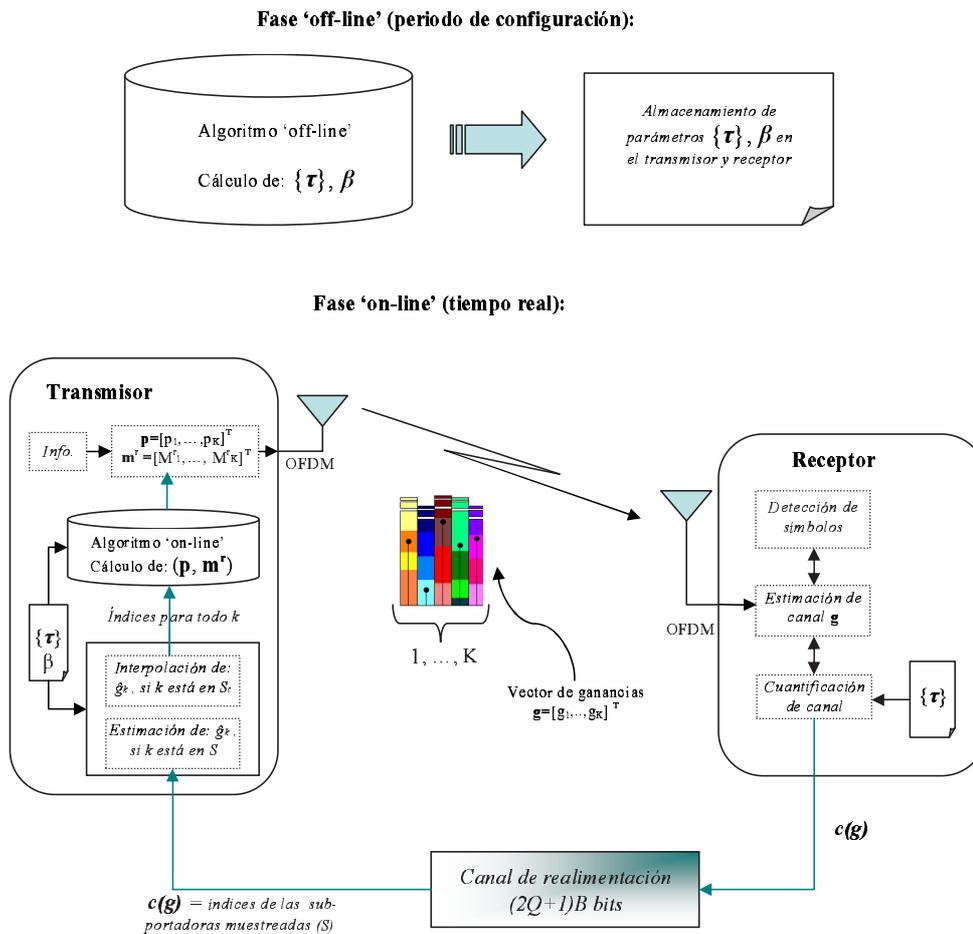


Figura 3.5: Interdependencia de los módulos de computación.

Para $K \gg Q$ (situación habitual en los sistemas OFDM), el número de bits de realimentación se reduce considerablemente: desde KB hasta $(2Q + 1)B$. Puesto que $2Q + 1$ componentes de \mathbf{g} son suficientes para identificar el vector \mathbf{g} en su conjunto, se utilizará \mathcal{S} para denotar el conjunto de $2Q + 1$ índices cuyos valores de canal se muestrean (índices correspondientes al muestreo equi-espaciado de las K sub-portadoras) y \mathcal{S}_c su complemento. Dado c_k , en primer lugar se estimará \hat{g}_k para $k \in \mathcal{S}$ y después se interpolarán los valores de esta estimación para encontrar \hat{g}_k para $k \in \mathcal{S}_c$. Nótese que cualquiera de los diseños “off-line” descritos en la sección precedente puede utilizarse para cuantificar los valores g_k para $k \in \mathcal{S}$ y formar c_k en el receptor. A continuación, el transmisor deberá ejecutar de manera “on-line” las mencionadas tareas de estimación e interpolación y, finalmente, adaptará la tasa y potencia a transmitir en base a cualquiera de los algoritmos “on-line” analizados en el apartado anterior. La figura 3.5 representa los diferentes módulos computacionales junto con sus interacciones. La primera tarea del módulo de estimación e interpolación (la estimación de la ganancia de sub-portadoras muestreadas), puede llevarse a cabo utilizando cualquiera de las cuatro opciones RC presentadas en el apartado anterior. En RC3 y RC4, se estima primero la potencia a transmitir, que se utiliza a continuación para estimar g_k para $k \in \mathcal{S}$. Tomando RC3 como ejemplo, conociendo los umbrales de cuantificación así como el valor “off-line” de los parámetros v_k y β_1 (que se obtienen al calcular el diseño óptimo del cuantificador), la estimación \hat{g}_k puede calcularse a través de (3.57) resolviendo

$$\frac{[\ln(v_k \hat{g}_k \log_2(M_k^r) \beta_1)]^+}{\hat{g}_k} = \frac{[\ln(v_k \tau_{k,l} \log_2(M_k^r) \beta_1)]^+}{2\tau_{k,l}} + \frac{[\ln(v_k \tau_{k,l+1} \log_2(M_k^r) \beta_1)]^+}{2\tau_{k,l+1}}, \quad k \in \mathcal{S}. \quad (3.65)$$

Las prestaciones de los diversos métodos de estimación serán comparadas en el apartado 3.2.7.

Dado \hat{g}_k para $k \in \mathcal{S}$, el siguiente paso consiste en interpolar y estimar g_k para $k \in \mathcal{S}_c$. Si bien el abanico de esquemas de interpolación es amplio (e.g., MMSE o LMMSE, consúltese [34] para un estudio más detallado), análisis exploratorios basados en simulaciones numéricas sugieren que, aunque con prestaciones parecidas, el interpolador que conduce a un menor consumo de potencia es el que interpola las muestras en base a funciones seno cardinal (de aquí en adelante “interpolador sinc”) cuyo uso está ampliamente extendido en la recuperación de señales de banda limitada muestreadas periódicamente [68] y que ya se ha utilizado con anterioridad para sistemas OFDM [67].

Definiendo el conjunto de sub-portadoras disponibles como $\mathcal{S} := \{\lfloor K/(4Q + 1) + nK/(2Q + 1) \rfloor | n = 0, \dots, 2Q\}$, el “interpolador sinc” aplicado a las sub-portadoras pertenecientes a \mathcal{S}_c produce

$$\hat{g}_k = \sum_{k' \in \mathcal{S}} \hat{g}_{k'} \text{sinc} \frac{(k - k')(2Q + 1)}{K}, \quad k \in \mathcal{S}_c. \quad (3.66)$$

Téngase en cuenta que, una vez estimados e interpolados, el transmisor conoce los valores de $\hat{g}_k \forall k$. En base a esta información, el receptor podría suponer que $g_k = \hat{g}_k$ y utilizar así el esquema de adaptación para el caso de P-CSIT (algoritmo 3.6), no obstante el algoritmo 3.6 no tiene en cuenta los errores presentes en \hat{g}_k y la solución obtenida no es capaz de satisfacer los requisitos de BER exigidos. Sin embargo, si en lugar de suponer que $g_k = \hat{g}_k$ el transmisor cuantifica los valores estimados de \hat{g}_k y utiliza el esquema de adaptación para el caso de Q-CSIT (algoritmo 3.7), como este esquema sí tiene en cuenta de forma inherente el error cuantificación¹⁸. los resultados obtenidos minimizan la potencia transmitida sin violar los requisitos de QoS. Un esquema detallado que ilustra este procedimiento y representa las diferencias de operación entre la fase “off-line” y la fase “on-line” se muestra en la figura 3.5.

¹⁸Nótese que la cuantificación \hat{g}_k presenta una doble fuente de error incertidumbre: la asociada al proceso de interpolación y la asociada al proceso de cuantificación. No obstante al ser el error de interpolación menor que el de cuantificación el efecto preponderante se debe al segundo, que está recogido de forma óptima en el diseño del algoritmo 3.7

Hasta este punto, se ha explorado la utilización de P-CSIT y Q-CSIT para disminuir la potencia transmitida por un sistema OFDM. Para complementar este análisis, queda por investigar la utilización de esquemas de adaptación basados en CSIT estadístico (S-CSIT). Sacrificando las prestaciones en términos de consumo de potencia, estos esquemas llevarán asociada la menor complejidad de cómputo y necesitarán menor información de realimentación.

3.2.6. Adaptación óptima basada en S-CSIT

Basándose en la FDP del canal (que bajo (sp2) queda totalmente definida por la ganancia media \bar{g}_k), el objetivo en esta sección consiste en encontrar los vectores $(\mathbf{m}^r, \mathbf{p})$ tasa (modulación) y potencia *fixos* (i.e., vectores que no varían en el tiempo a menos que las ganancias medias lo hagan) que satisfaciendo los requisitos de QoS minimicen la potencia transmitida. Esta minimización se realizará para dos escenarios diferentes: en el primero de ellos se resolverá el problema adaptando tasa \mathbf{m}^r y potencia \mathbf{p} , mientras que en el segundo caso se considerará que la asignación de potencia no varía por sub-portadora y se optimizará únicamente la tasa de transmisión \mathbf{m}^r .

Adaptación de tasa y potencia

El problema de optimización que devuelve los esquemas de adaptación de tasa y potencia óptimos puede formularse como:

$$\left\{ \begin{array}{l} \text{mín}_{\mathbf{m}^r(\bar{\mathbf{g}}) \in \mathcal{M}, \bar{\mathbf{p}}(\bar{\mathbf{g}}) \geq \mathbf{0}} J(\bar{\mathbf{p}}), \quad J(\bar{\mathbf{p}}) = \sum_{k=1}^K \bar{p}_k(\bar{\mathbf{g}}) \\ \text{s.a : C1. } \sum_{k=1}^K \frac{\log_2 M_k^r}{\check{r}} \bar{\epsilon}_k(\bar{p}_k, M_k^r, \bar{g}_k) \leq \check{\epsilon} \\ \text{C2. } \sum_{k=1}^K \log_2 M_k^r \geq \check{r}. \end{array} \right. \quad (3.67)$$

donde tanto \mathbf{m}^r como $\bar{\mathbf{p}}$ dependen claramente de $\bar{\mathbf{g}}$ (i.e., S-CSIT). Al calcular el promediado en C1 sobre todos los posibles valores de g_k se obtiene: $\bar{\epsilon}_k(\bar{p}_k, M_k^r) = \int_0^\infty e^{-v_k \bar{p}_k g_k} f_{g_k}(g_k) dg_k = \frac{\kappa_1}{1+v_k \bar{p}_k \bar{g}_k}$. Relegando temporalmente C2 y al igual que para los casos anteriores, aplicando las condiciones KKT al problema simplificado

se obtiene

$$\bar{p}_k^*(\bar{\mathbf{g}}) = \left[\beta_S \sqrt{\frac{\kappa_1 \log_2 M_k^r}{\check{r} v_k \bar{g}_k}} - \frac{1}{v_k \bar{g}_k} \right]^+, \quad (3.68)$$

donde $\beta_S = \frac{1}{\epsilon} \sum_{k=1}^K \sqrt{\frac{\kappa_1 \log_2 M_k^r}{\check{r} v_k \bar{g}_k}}$. Es sencillo demostrar que $\frac{\partial^2 J}{\partial \bar{p}_k^2} \geq 0$ y $\frac{\partial^2 C1}{\partial \bar{p}_k^2} \geq 0$, lo que hace que el problema de optimización sea convexo y convierte a las condiciones necesarias derivadas (i.e., (3.68) $\forall k$) en condiciones suficientes para que la solución obtenida constituya un mínimo global. De forma similar a lo realizado en las secciones previas, (3.68) debe calcularse para todas las combinaciones ordenadas de \mathbf{m}^r que satisfagan la restricción de tasa \check{r} , eligiendo como solución final aquella que requiera menos potencia.

Adaptación de tasa con potencia uniforme

En último lugar se estudia un esquema de adaptación en base a S-CSIT en el que la asignación de potencia sobre las K_a sub-portadoras activas es uniforme y, por lo tanto, únicamente se adaptan el valor de dicha potencia $\bar{p}^u > 0$ y la asignación de bits a través de sub-portadoras. Utilizando la función indicador $\mathbf{I}_{\{\cdot\}}$ se define el conjunto de posibles vectores de potencia transmitida como $\mathcal{P} := \{[\bar{p}_0 \dots \bar{p}_K]^T \mid \bar{p}_k = \bar{p}^u \cdot \mathbf{I}_{\{k \in \mathcal{K}_a\}}\}$, estando entonces el espacio de búsqueda de (3.67) restringido a $\mathbf{m}^r \in \mathcal{M}$ y $\mathbf{p} \in \mathcal{P}$. Puesto que no es posible obtener una solución analítica, el problema se optimiza para todos los valores ordenados de \mathbf{m}^r que cumplan C2 en forma de igualdad, para a continuación buscar numéricamente el nivel de \bar{p}^u que satisfaga ajustadamente C1. Por último, se elige el par óptimo $(\mathbf{m}^r, \mathbf{p})$ con el que se alcance la mínima $\bar{p}^u K_a$. Nótese que puesto que \bar{p}^u es constante sobre las sub-portadoras, también podría haberse implementado un algoritmo más sofisticado para la la distribución que considerara una asignación *avariciosa* de los bits siguiendo las líneas de [17, capítulo 16], seguido de una búsqueda numérica para encontrar el valor de \bar{p}^u . (La idea básica de este tipo de algoritmos consiste en asignar los bits de forma secuencial y de manera que cada uno de

Tabla 3.6: Prestaciones de esquemas basados en SIT-QCSIT y Q-CSIT ($K = 64$, $BER_0 = 10^{-3}$, $\tilde{r} = 20$, $B = 2$).

Esquema	Estrategias de Adaptación	Potencia Transmitida [dB_W]	BER
Q-CSIT	Tx-A	14.2	$1.20 \cdot 10^{-3}$
	Tx-B	14.6	$1.10 \cdot 10^{-3}$
EP-QCSIT	Tx-A	14.5	$1.15 \cdot 10^{-3}$
	Tx-B	14.9	$1.15 \cdot 10^{-3}$
Lloyd- g_k	Tx-A	14.8	$1.05 \cdot 10^{-3}$
	Tx-B	15.4	$1.05 \cdot 10^{-3}$
Lloyd- P_k	Tx-A	15.0	$1.00 \cdot 10^{-3}$
	Tx-B	15.6	$0.90 \cdot 10^{-3}$

ellos quede asignado a la sub-portadora que minimice el coste añadido asociado a su transmisión. Ejemplos de asignaciones avariciosas en el entorno multi-usuario serán analizados en el siguiente capítulo).

3.2.7. Simulaciones numéricas

A fin de analizar de forma numérica los esquemas propuestos, se considerará como caso de referencia un sistema OFDM adaptativo con $K = 64$ sub-portadoras, $B = 2$ bits de realimentación por sub-portadora (en los casos en que se utilice Q-CSIT), transmitiendo sobre canales con desvanecimientos con ganancias medias $\{\bar{g}_k\}_{k=1}^K$ que varían sobre un rango de 10 dB , satisfaciendo un nivel máximo de BER de $\tilde{\epsilon} = 10^{-3}$ y tasa mínima de $\tilde{r} = 20$ bits por símbolo OFDM.

Caso de estudio 1 (Estrategias de adaptación y alternativas para Q-CSIT): En la tabla 3.6 se comparan los cuantificadores de canal óptimos y los de complejidad reducida, a saber, EP-QCSIT, esquemas Lloyd- g_k y Lloyd- p_k , en estrategias de adaptación Tx-A y Tx-B. Como era de esperar, Tx-A supera en todos los casos a Tx-B. Menos esperado era el hecho de que todos los cuantificadores de complejidad reducida sufren pérdidas mínimas (como mucho 1 dB) de eficiencia en el consumo c.r.a. los óptimos. En particular, dentro de los cuantificadores sub-

Tabla 3.7: Prestaciones de esquemas basados en P-CSIT y Q-CSIT ($K = 64$, $BER_0 = 10^{-3}$, $\tilde{r} = 20$, $B = 2$).

Esquema	Estrategias de Adaptación	Potencia Transmitida [dB_W]	BER
P-CSIT	–	12.6	$1.01 \cdot 10^{-3}$
Q-CSIT	Tx-A	13.4	$0.95 \cdot 10^{-3}$
	Tx-B	14.2	$0.93 \cdot 10^{-3}$
	Tx-C	14.9	$0.90 \cdot 10^{-3}$
	Tx-D	15.0	$0.85 \cdot 10^{-3}$
UC-QCSIT	Tx-A	15.2	$0.95 \cdot 10^{-3}$
	Tx-B	15.8	$0.80 \cdot 10^{-3}$
	Tx-C	16.8	$0.70 \cdot 10^{-3}$
	Tx-D	16.6	$0.75 \cdot 10^{-3}$
EP-QCSIT	Tx-A	13.7	$1.10 \cdot 10^{-3}$
	Tx-B	14.5	$1.00 \cdot 10^{-3}$
	Tx-C	15.4	$1.20 \cdot 10^{-3}$
	Tx-D	15.6	$0.90 \cdot 10^{-3}$

óptimos, EP-QCSIT es el que mejores prestaciones presenta (una diferencia algo menor de 0.5 dB c.r.a. óptimo). Por esta razón, se tomará este cuantificador como el representativo de esta clase de ahora en adelante.

Caso de estudio 2 (P-CSIT frente a Q-CSIT sub-portadoras con distintas ganancias medias y sub-portadoras con ganancias medias homogéneas): La tabla 3.7 lista el consumo añadido de potencia cuando se utilizan los esquemas basados en Q-CSIT en lugar de los basados en P-CSIT. Para este estudio, se han analizado tres situaciones distintas (Q-CSIT óptimo, UC-QCSIT, y EP-QCSIT) considerando todas las diferentes opciones de adaptación planteadas para el transmisor (TxA, TxB, TxC y TxD). Como se puede observar, las prestaciones de las cuatro estrategias TxA-D se distancian en no más de 2 dB . Asimismo, EP-QCSIT muestra una eficiencia de consumo cercana a la del óptimo (ligeramente superior a 1 dB c.r.a. a la del P-CSIT) cuando se emplea un esquema de transmisión totalmente adaptativo (i.e., Tx-A). Finalmente, respecto a las pequeñas variaciones de la BER alrededor del valor objetivo $\tilde{\epsilon} = 10^{-3}$, debe comentarse que se deben a que

Tabla 3.8: Prestaciones de esquemas basados en P-CSIT y Q-CSIT para sub-portadoras con ganancia media uniforme ($K = 64, BER_0 = 10^{-3}, \tilde{r} = 20, B = 2, \bar{g}_k = 2$).

Esquema	Estrategias de Adaptación	Potencia Transmitida [dB_W]	BER
P-CSIT	–	13.3	$1.05 \cdot 10^{-3}$
Q-CSIT	Tx-A	14.2	$1.05 \cdot 10^{-3}$
	Tx-B	14.2	$1.02 \cdot 10^{-3}$
	Tx-C	16.8	$1.02 \cdot 10^{-3}$
	Tx-D	17.1	$0.90 \cdot 10^{-3}$
UC-QCSIT	Tx-A	14.1	$1.10 \cdot 10^{-3}$
	Tx-B	14.1	$1.15 \cdot 10^{-3}$
	Tx-C	16.9	$1.15 \cdot 10^{-3}$
	Tx-D	17.0	$0.85 \cdot 10^{-3}$
EP-QCSIT	Tx-A	14.6	$1.01 \cdot 10^{-3}$
	Tx-B	14.7	$1.01 \cdot 10^{-3}$
	Tx-C	17.1	$1.01 \cdot 10^{-3}$
	Tx-D	17.1	$0.95 \cdot 10^{-3}$

las simulaciones han sido realizadas con un número finito de muestras y no a que los esquemas propuestos no sean capaces de satisfacer con garantías ϵ .

La tabla 3.8 es la homóloga de la tabla 3.7 cuando el valor \bar{g}_k es el mismo $\forall k$. Como se podía suponer intuitivamente, en este caso el esquema óptimo Q-CSIT presenta unas prestaciones idénticas a UC-QCSIT, si bien ambos algoritmos utilizan umbrales distintos (recuérdese que los métodos de cálculo no eran los mismos). Este resultado pone de manifiesto que la solución óptima puede no ser única. Además, la equivalencia práctica de Tx-A y Tx-B certifica el funcionamiento (casi)óptimo de una asignación de tasa bit sobre las sub-portadoras no adaptativa cuando estas son estadísticamente idénticas.

Caso de estudio 3 (Consumo de P-CSIT, Q-CSIT y S-CSIT para distintos valores de BER): Para distintos requisitos de BER, en la parte superior de la figura 3.6 se representa el consumo de potencia por los esquemas de S-CSIT y PU-SCSIT. Se puede apreciar que, el diseño S-CSIT óptimo (que tiene la libertad de asignar potencia distinta a lo largo de sub-portadoras) presenta una eficiencia

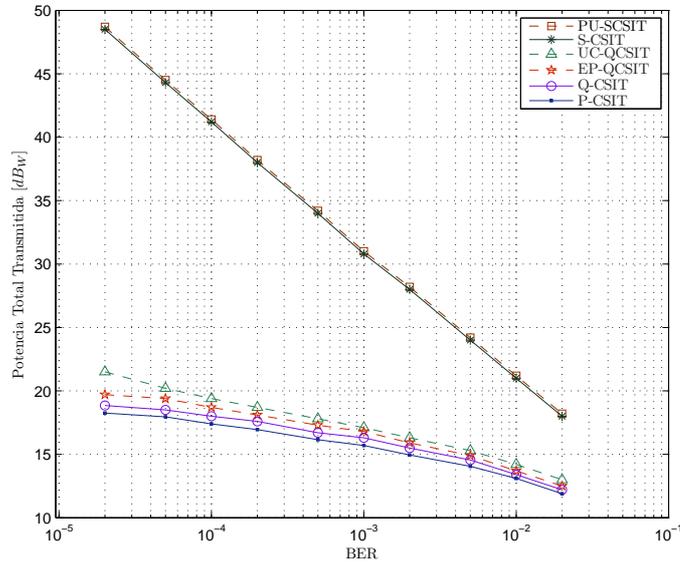


Figura 3.6: Prestaciones de esquemas bajo distintas consideraciones de CSIT y requisitos de BER ($K = 64$, $\tilde{r} = 25$, $B = 2$).

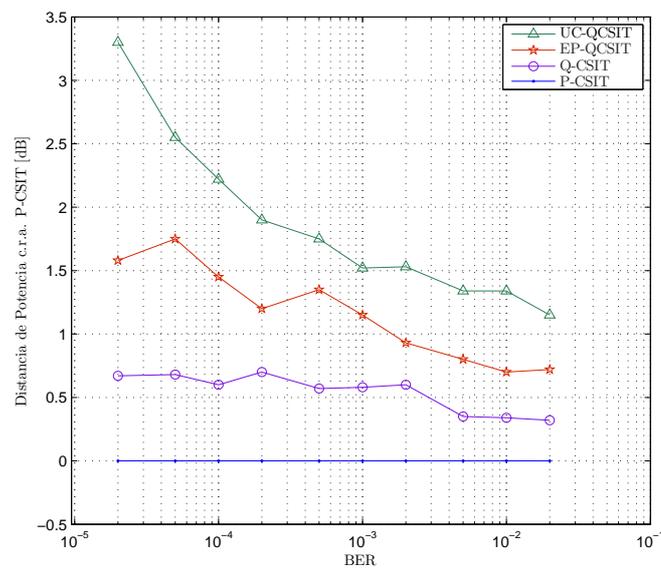


Figura 3.7: Prestaciones de esquemas basados en P-CSIT y Q-CSIT para distintos requisitos de BER ($K = 64$, $\tilde{r} = 25$, $B = 2$).

Tabla 3.9: Prestaciones de distintos métodos de estimación con interpolación basada en funciones “sinc” para la reducción de la tasa de realimentación ($K = 64$, $\tilde{r} = 20$, $B = 2$, $L = 3$, $\mathbb{E}[|h_0|^2] = 1$, $\mathbb{E}[|h_1|^2] = 0.5$, $\mathbb{E}[|h_2|^2] = 0.25$, $\mathbb{E}[|h_3|^2] = 0.1$).

Esquema	Bits realimentados	Potencia Media Total [dB_W]	BER
P-CSIT	∞	12.3	$1.00 \cdot 10^{-3}$
Q-CSIT	KB	13.2	$0.95 \cdot 10^{-3}$
RC1	$(2Q + 1)B$	13.4	$1.35 \cdot 10^{-3}$
RC2	$(2Q + 1)B$	13.5	$0.99 \cdot 10^{-3}$
RC3	$(2Q + 1)B$	14.1	$0.55 \cdot 10^{-3}$
RC4	$(2Q + 1)B$	13.4	$0.95 \cdot 10^{-3}$

de consumo similar a la del diseño basado en PU-SCSIT lo que, debido a su mayor sencillez, favorece la utilización de este último en escenarios prácticos. Haciendo ya referencia al resto de la figura 3.6, se observa asimismo cómo el diseño basado en Q-CSIT, comparado con los diseños S-CSIT, presenta un ahorro en la potencia consumida de aproximadamente 8, 15, y 24 dB para $\tilde{\epsilon} = 10^{-2}$, 10^{-3} , y 10^{-4} , respectivamente. Comparado con el caso en el que no existe CSIT, este ahorro es significativamente mayor puesto que el propio esquema S-CSIT presenta una ventaja c.r.a. a los sistemas no realimentados (sistemas de lazo abierto) de uno o dos órdenes de magnitud (ente 15 y los 25 dB). Ampliando los resultados anteriores –como efectivamente se hace en la figura 3.7– para los esquemas P-CSIT y Q-CSIT, se comprueba cómo el esquema óptimo Q-CSIT se acerca sorprendentemente (entorno a 1 dB) al esquema P-CSIT. De hecho, el consumo asociado al esquema EP-QCSIT se encuentra dentro del margen de 1 dB con respecto al consumo del diseño óptimo basado en Q-CSIT. Debe destacarse también que el esquema UC-QCSIT es el que presenta un mayor consumo, especialmente para valores de BER muy pequeños.

Caso de estudio 4 (Reducción de la tasa de realimentación y alternativas para la estimación de canal): En este apartado se estudia la posible merma de prestaciones asociada a la reducción del número de bits realimentados mediante la implementación de los esquemas propuestos en el apartado 3.2.5 (recuérdese que la idea básica

Tabla 3.10: Variación del número de bits de realimentación utilizando un esquema basado en EP-QCSIT ($K = 64$, $\tilde{r} = 25$).

B	1	2	3	4	∞
Potencia media total [dB_W]	14.2	13.3	12.8	12.7	12.0
BER	$0.85 \cdot 10^{-3}$	$1.00 \cdot 10^{-3}$	$0.95 \cdot 10^{-3}$	$0.99 \cdot 10^{-3}$	$1.05 \cdot 10^{-3}$

consistía en disminuir el número de sub-portadoras cuyas ganancias se cuantifican y obtener después las ganancias restantes mediante interpolación). En la tabla 3.9 se muestran la potencia y la BER obtenidas cuando el proceso de estimación de las sub-portadoras se realiza en base a las alternativas RC1-RC4 (la interpolación se realiza en base a funciones “sinc” de acuerdo a (3.66)) y comparándolas con aquellas obtenidas para los esquemas óptimos basados en P-CSIT y en Q-CSIT (este último sin que se reduzca la tasa de realimentación). La precisión en los distintos pasos de la estimación e interpolación correspondientes a las distintas alternativas RC se manifiestan en las fluctuaciones de la BER alrededor de su valor nominal. Gracias a su simplicidad y robustez frente a los errores de estimación, el método RC2 con estimación de la amplitud es preferible cuando la región de cuantificación viene ya dada¹⁹. Nótese además, la pérdida insignificante de 0.3 dB en la potencia total media relativa a la realimentación por sub-portadora Q-CSIT, y la pérdida mínima de 1 dB en comparación con el esquema P-CSIT.

Caso de estudio 5 (Número de regiones de cuantificación): En este último caso se investiga los efectos asociados a la variación del número de bits de realimentación en el esquema EP-QCSIT. Los resultados de la tabla 3.10 muestran que la ganancia de potencia aumenta de forma perceptible cuando B aumenta de 1 a 2 bits, si bien la ganancia relativa por cada incremento adicional en B disminuye. También se puede observar que $B = 1$ conlleva solamente una pérdida de potencia de 2 dB comparado con el caso P-CSIT, el cual se corresponde con $B = \infty$.

¹⁹No debe confundirse la ventaja del método RC2 en el proceso de estimación, interpolación y posterior cuantificación con la ventaja del método RC3 para llevar a cabo los esquemas de complejidad reducida en el apartado 3.2.4.

3.2.8. Síntesis de resultados

Bajo restricciones de tasa y probabilidad de error, se ha optimizado la potencia transmitida por un sistema OFDM inalámbrico con desvanecimientos mediante varios esquemas de CSIT. Se ha prestado especial atención al diseño Q-CSIT, esquema con un alto potencial práctico debido a que el receptor sólo necesita realimentar al transmisor un número limitado de bits que indexan la región de cuantificación en la que el sub-canal subyacente se encuentra. Los transceptores optimizados resultantes presentan un gran potencial puesto que ahorran 15 dB de potencia con respecto a S-CSIT (el doble como mucho comparado con OFDM con no-CSIT), y se encuentran dentro del margen de 1 dB del punto de referencia de los diseños P-CSIT, si bien estos últimos son más adecuados para canales guiados o para canales inalámbricos de baja variabilidad. Para elegir una forma sensata de Q-CSIT, se han diseñado cuantificadores de canal óptimos “off-line” que utilizando la descripción estadística del canal seleccionan los umbrales de cuantificación de las sub-portadoras. Se han propuesto, además, cuantificadores de complejidad reducida, siendo el más prometedor (debido a su sencillez y a la cercanía al consumo de potencia óptimo) aquel para el que los umbrales se seleccionan independientemente por sub-portadora y se calculan de manera se generen regiones de cuantificación equi-probables (EP) para cada una de ellas.

Una vez realizado el diseño del esquema Q-CSIT, se han desarrollado también algoritmos óptimos (más complejos) y algoritmos sub-óptimos (más sencillos) para la adaptación “on-line” de potencia y asignación de bits. La aplicación de estos métodos a transmisores OFDM con grados variables de adaptación (e.g., que solamente dejen variar el conjunto de sub-portadoras activas o las modulaciones a utilizar) permite alcanzar un compromiso flexible entre ahorro de potencia y facilidades en implementación y funcionamiento “on-line” de complejidad limitada. Para hacer aun más atractivo el funcionamiento de los diseños basados en Q-CSIT, se han empleado técnicas de interpolación con el fin de obtener esquemas de realimentación de tasa reducida los cuales únicamente hacen uso de Q-CSIT en un

número pequeño de sub-portadoras (tantas como el doble de la longitud del canal discreto equivalente). Interesante hubiera sido asimismo la caracterización de las prestaciones de sistemas que realimentaran un número de bits diferente por sub-portadora. Finalmente, debe destacarse que en las simulaciones realizadas se ha analizado el efecto asociado a variar el número de bits de realimentación en las prestaciones del sistema. Los resultados han puesto de manifiesto que cuantificadores de baja complejidad (EP) con un bit de realimentación por sub-portadora e interpolación no requieren para escenarios típicos de aplicación más de 7-11 bits de realimentación y no exceden los 2 *dB* de pérdidas en potencia c.r.a. consumo de esquemas óptimos basados en P-CSIT.

3.3. Consideraciones finales

Este capítulo ha supuesto la primera toma de contacto con el diseño de sistemas de comunicaciones relativamente sofisticados que hacen uso de Q-CSIT. Curiosamente, a pesar de tratarse de sistemas a priori muy distintos (una WSN que comunica de forma coherente con un centro de fusión de datos y un sistema OFDM), las soluciones obtenidas dan lugar a la reflexión respecto a distintos aspectos comunes.

En los dos escenarios estudiados la alternativa de adaptación en base a Q-CSIT es de sencilla implementación y ofrece prestaciones en términos de consumo muy cercanos a la adaptación óptima basada en P-CSIT.

También para los dos sistemas la mayor desventaja se encuentra en la complejidad computacional necesaria para resolver el cuantificador de canal, que en ambos casos presenta carácter de óptimo local. Asimismo debe mencionarse que esta complejidad es sustancialmente mayor en el caso de la optimización del sistema OFDM, en gran parte porque la garantía de niveles instantáneos de QoS complica la resolución del problema (tanto en la fase “on-line” como, especialmente, en la “off-line”). Esto ha motivado el análisis de cuantificadores sub-óptimos de menor complejidad cuyas prestaciones han resultado ser cercanas a las de los

diseños óptimos.

Asimismo, se ha comprobado también cómo una adecuada formulación del problema puede simplificar la resolución del mismo (una de las alternativas de mayor interés a consistido en relacionar tasa y potencia de transmisión a partir de la función de BER) y cómo la estructura de adaptación en base a P-CSIT ayuda a decidir el diseño del cuantificador de canal para cada caso.

Finalmente debe destacarse que en ambos casos las necesidades de tasa de realimentación son extremadamente bajas y que, pese a que el diseño del cuantificador de canal fuera considerablemente distinto, en las dos ocasiones la idoneidad de la solución propuesta se ha visto refrendada por la escasa diferencia de prestaciones con respecto a la solución óptima basada en P-CSIT.

3.4. Apéndices

3.4.1. Expresión analítica de la BER para el caso de Q-CSIT en WSN

Encontrar una expresión analítica cerrada para $\varphi_\epsilon(\tilde{\tau}_l, \tilde{\tau}_{l+1}, \tilde{p}_l, r_l, \tilde{\epsilon})$ exige el cálculo explícito de las dos integrales presentes en (3.29). Utilizando la FDA de g , para la segunda de las integrales se tiene que $\int_{\tilde{\tau}_l}^{\tilde{\tau}_{l+1}} f_{\tilde{g}}(\tilde{g})d\tilde{g} = \tilde{F}_{\tilde{g}}(\tilde{\tau}_{l+1}) - \tilde{F}_{\tilde{g}}(\tilde{\tau}_l)$. La primera integral en (3.29) requiere calcular de forma analítica $\tilde{\Phi}(p, r, x) := \int \epsilon(\tilde{g}, p, r) f_{\tilde{g}}(\tilde{g})d\tilde{g} \Big|_{\tilde{g}=x}$. Apoyándose en (3.9) con $b := \kappa_2 p / (2^r - 1)$, la expresión anterior puede reescribirse como $\tilde{\Phi}(p, r, x) = \kappa_1 \int \exp(b, \tilde{g}) f_{\tilde{g}}(\tilde{g})d\tilde{g} \Big|_{\tilde{g}=x}$, que tras

algunas tediosas manipulaciones matemáticas queda [cf. (3.26)]

$$\begin{aligned}
& \int \exp(-b\tilde{g}) f_{\tilde{g}}(\tilde{g}) d\tilde{g} \\
&= \int \frac{1}{\Gamma(M)} \left\{ \exp\left(-\frac{1+b(1-z_{\text{máx}})}{1-z_{\text{máx}}}\tilde{g}\right) \frac{\tilde{g}^{M-1}}{(1-z_{\text{máx}})^M} (1-N_u z_{\text{máx}}^M) \right. \\
&\quad \left. + N_u \exp(-(1+b)\tilde{g}) \left[\Gamma(M) - \Gamma(M, z_{\text{máx}}\tilde{g}/(1-z_{\text{máx}})) \right] \right\} d\tilde{g} \\
&= \frac{-1}{\Gamma(M)[1+b]} \left\{ \frac{(1+b)(1-N_u z_{\text{máx}}^M) + N_u z_{\text{máx}}^M}{[1+b(1-z_{\text{máx}})]^M} \Gamma\left(M, \tilde{g} \frac{1+b(1-z_{\text{máx}})}{1-z_{\text{máx}}}\right) \right. \\
&\quad \left. + \frac{N_u \exp(-(1+b)\tilde{g})}{1+b} [\Gamma(M) - \Gamma(M, \tilde{g} z_{\text{máx}}/(1-z_{\text{máx}}))] \right\}. \quad (3.69)
\end{aligned}$$

Basándose en (3.69), puede escribirse (3.29) en forma cerrada como

$$\varphi_{\epsilon}(\tilde{\tau}_l, \tilde{\tau}_{l+1}, \tilde{p}_l, r_l, \tilde{\epsilon}) = \tilde{\Phi}(\tilde{p}_l, r_l, \tilde{\tau}_{l+1}) - \tilde{\Phi}(\tilde{p}_l, r_l, \tilde{\tau}_l) - \tilde{\epsilon}[\tilde{F}_{\tilde{g}}(\tilde{\tau}_{l+1}) - \tilde{F}_{\tilde{g}}(\tilde{\tau}_l)]. \quad (3.70)$$

Por último, utilizando (3.28) y (3.70) se concluye que

$$\tilde{\epsilon}_l(\tilde{\tau}_l, \tilde{\tau}_{l+1}, \tilde{p}_l, r_l) = \frac{\tilde{\Phi}(\tilde{p}_l, r_l, \tilde{\tau}_{l+1}) - \tilde{\Phi}(\tilde{p}_l, r_l, \tilde{\tau}_l)}{\tilde{F}_{\tilde{g}}(\tilde{\tau}_{l+1}) - \tilde{F}_{\tilde{g}}(\tilde{\tau}_l)}, \quad (3.71)$$

expresión que cuantifica de forma analítica el valor de la BER media cuando se implementa el esquema de adaptación y cuantificación propuesto.

3.4.2. Expresión analítica de la derivada de la potencia para el caso de Q-CSIT en WSN

Para la FPD de \tilde{g} en (3.26), puede escribirse

$$\int_{\tilde{\tau}_i}^{\tilde{\tau}_{i+1}} [\partial_{\epsilon}(\tilde{g}, p, r)/\partial p] f_{\tilde{g}}(\tilde{g}) d\tilde{g} = \tilde{\xi}(p, r, \tilde{\tau}_{i+1}) - \tilde{\xi}(p, r, \tilde{\tau}_i). \quad (3.72)$$

Derivando (3.9) c.r.a. p , (3.72) puede reescribirse como

$$\tilde{\xi}(p, r, x) := \int [\partial_{\epsilon}(\tilde{g}, p, r)/\partial p] f_{\tilde{g}}(\tilde{g}) d\tilde{g} \Big|_{\tilde{g}=x} := \frac{-\kappa_1 \kappa_2}{2^r - 1} \tilde{I}_{\xi} \left(\frac{\kappa_2 p}{2^r - 1}, \tilde{g} \right) \Big|_{\tilde{g}=x} \quad (3.73)$$

donde $\tilde{I}_\xi(b, g)$ puede expresarse de forma cerrada como

$$\begin{aligned}
\tilde{I}_\xi(b, \tilde{g}) &:= \int \tilde{g} \exp(-b\tilde{g}) f_{\tilde{g}}(\tilde{g}) d\tilde{g} \\
&= \int \frac{1}{\Gamma(M)} \left\{ \exp\left(-\frac{1+b(1-z_{\text{máx}})}{1-z_{\text{máx}}}\tilde{g}\right) \frac{\tilde{g}^M}{(1-z_{\text{máx}})^M} (1-N_u z_{\text{máx}}^M) \right. \\
&\quad \left. + N_u \exp(-(1+b)\tilde{g}) \left[\Gamma(M) - \Gamma(M, z_{\text{máx}}\tilde{g}/(1-z_{\text{máx}})) \right] \tilde{g} \right\} d\tilde{g} \\
&= \frac{-1}{\Gamma(M)} \left\{ (1-z_{\text{máx}}) \frac{1-N_u z_{\text{máx}}^M}{[1+b(1-z_{\text{máx}})]^{M+1}} \Gamma\left(M+1, \tilde{g} \frac{1+b(1-z_{\text{máx}})}{1-z_{\text{máx}}}\right) \right. \\
&\quad \left. + \frac{N_u \exp(-(1+b)\tilde{g})(1+(1+b)\tilde{g})}{(1+b)^2} [\Gamma(M) - \Gamma(M, \tilde{g} z_{\text{máx}}/(1-z_{\text{máx}}))] \right. \\
&\quad \left. + \frac{N_u z_{\text{máx}}^M}{(1+b)^2 [1+b(1-z_{\text{máx}})]^{M+1}} \left[1+b(1-z_{\text{máx}}) \Gamma\left(M, \tilde{g} \frac{1+b(1-z_{\text{máx}})}{1-z_{\text{máx}}}\right) \right. \right. \\
&\quad \left. \left. \times (1+b)(1-z_{\text{máx}}) \Gamma\left(M+1, \tilde{g} \frac{1+b(1-z_{\text{máx}})}{1-z_{\text{máx}}}\right) \right] \right\}. \tag{3.74}
\end{aligned}$$

3.4.3. Demostración de la proposición 3.1

La información digital enviada al transmisor es una versión cuantificada del parámetro real g_k . Sea DFT_K la DFT de longitud K y $\delta[n]$ el impulso unitario de tiempo discreto. Los valores de los coeficientes de canal y de las ganancias de sub-portadora pueden representarse como secuencias de longitud K de manera que $h[k] = \sum_{u=0}^{K-1} h_u \delta[k-u]$ y $g[k] = \sum_{u=0}^{K-1} g_{u+1} \delta[k-u]$. Recordando que $\sigma_k^2 = 1$, $g[k]$ puede expresarse como

$$\begin{aligned}
g[k] &= |H[k]|^2 = H[k]H^\dagger[k] \\
&= \text{DFT}_K \{h[n] \otimes_K h^\dagger[((-n))_K]\} \\
&= \text{DFT}_K \left\{ \sum_{q=0}^{K-1} h[q] h^\dagger[((n+q))_K] \right\} \\
&:= \text{DFT}_K \{\phi_{hh}[n]\}, \tag{3.75}
\end{aligned}$$

donde \otimes_N denota la convolución circular de longitud K y $((x))_N$ representa el resto obtenido al dividir x por K (t.c.c. x módulo K). La estructura de la correlación discreta $\phi_{hh}[n]$ es la siguiente (exceptuando el caso trivial de $K < 2Q + 1$ donde

se retorna al caso de realimentación por cada sub-portadora)

$$\phi_{hh}[n] = \begin{cases} \sum_{q=0}^Q |h[q]|^2, & n = 0; \\ \sum_{q=0}^Q h[q]h^\dagger[n+q], & n = 1, \dots, Q; \\ \phi_{hh}^\dagger[K-n], & n = K-Q, \dots, K-1; \\ 0, & \text{en otro caso.} \end{cases} \quad (3.76)$$

Utilizando la estructura de $\phi_{hh}[n]$ junto con (3.75) y definiendo las secuencias $\phi_{hh,R} := \text{Re}\{\phi_{hh}[n]\}$ y $\phi_{hh,I} := \text{Im}\{\phi_{hh}[n]\}$, cualquier ganancia g_k puede expresarse como

$$g_k = \phi_{hh}[0] + 2 \sum_{n=1}^Q \left[\phi_{hh,R}[n] \cos \frac{2\pi(k-1)n}{K} + \phi_{hh,I}[n] \sin \frac{2\pi(k-1)n}{K} \right]. \quad (3.77)$$

Por lo tanto, $2Q + 1$ g_k coeficientes bastan para describir todas las muestras de $\phi_{hh}[n]$. Este resultado implica que todas las ganancias g_k pueden calcularse si $2Q + 1$ de ellas son conocidas. Esto supone que, utilizando técnicas de estimación, interpolación y cuantificación, la realimentación de $(2Q + 1)B$ bits sea suficiente para describir la totalidad del vector de cuantificación \mathbf{c} .

Capítulo 4

Minimización de potencia en sistemas de comunicaciones adaptativos con múltiples fuentes de información

Este capítulo se dedica al estudio de la minimización de la potencia transmitida por sistemas multi-usuario con flujos de información individuales garantizando niveles de calidad de servicio.

La optimización de sistemas de múltiples usuarios involucra no sólo parámetros del nivel físico (tales como potencia, tasa, modulación, BER), sino también del nivel de enlace (prioridad entre usuarios, acceso al medio). Mientras el diseño tradicional a través de capas (niveles) obliga a encontrar los valores de estos parámetros de forma independiente para cada capa, la consideración de los nuevos parámetros en problemas de optimización semejantes a los planteados en el capítulo anterior, permite la resolución conjunta de todos los parámetros y la obtención de soluciones óptimas de carácter global. Esta alternativa, que apenas presenta una penalización en términos de complejidad y alcanza prestaciones de carácter óptimo, proporciona además de forma natural esquemas óptimos de acceso al medio

que garantizan de forma estricta la equidad entre usuarios.

Como consecuencia de los motivos presentados, en este capítulo se analizan, en primer lugar, la adaptación óptima de los transmisores y la cuantificación de canal para un sistema TDMA (canales selectivos en el tiempo) en el que cada usuario puede transmitir de acuerdo a un juego de modos de modulación y codificación adaptativa (AMC) predeterminado y, en segundo lugar, la cuantificación del canal por parte del receptor así como la adaptación por parte de los transmisores (potencia, tasa y sub-portadoras) para un sistema OFDMA (canales selectivos en tiempo y en frecuencia).

4.1. Minimización de potencia en sistemas TDMA con QoS utilizando CSIT

Con la eficiencia energética emergiendo como un aspecto crucial en el diseño de sistemas inalámbricos comerciales y tácticos (militares, emergencia, etc.), la gestión y adaptación de recursos ha atraído una creciente atención tanto en canales aditivos gaussianos (AWGN) [35,86,99] como en canales con desvanecimientos con acceso múltiple por división en el tiempo (TDMA) [22,96]. La adaptación óptima de recursos para canales inalámbricos móviles ha sido estudiada en [29,40,41,84], mientras que esquemas de adaptación eficientes energéticamente para sistemas TDMA han sido investigados desde la perspectiva de la teoría de la información [87]. Los citados estudios trabajan bajo el supuesto de que tanto los transmisores como los receptores conocen de forma perfecta el canal (P-CSI), ofreciendo así límites fundamentales (al trabajar con códigos que alcanzan capacidad y utilizando P-CSI) para las prestaciones de cualesquiera otros sistemas.

El uso de técnicas FDD y de canales de control, así como la existencia de errores de estimación del canal, de pequeñas variabilidades en el retardo [38,63] o de limitación de tasa de realimentación, no son ajenas a los sistemas TDMA que operan sobre canales selectivos en el tiempo. Esto motiva que, al igual que en capítulos

anteriores, se analice la gestión y adaptación de recursos en sistemas TDMA en base a conocimiento cuantificado del canal (Q-CSI) con desvanecimientos.

De acuerdo a lo anunciado en la introducción, en esta sección se analizan sistemas TDMA inalámbricos limitados por potencia donde, explotando el Q-CSIT, los usuarios transmiten hacia un punto de acceso (PA) a través de una única antena utilizando modos de codificación y modulación adaptativa (AMC) que permiten satisfacer niveles de BER y tasa individuales para cada usuario (los detalles del modelo se describen en el apartado 4.1.1). En sintonía con el *modus operandis* de esta disertación, en la fase de diseño se atacan dos problemas principales: la cuantificación de canal para formar el Q-CSIT y la adaptación de los terminales usuarios al Q-CSIT. El diseño se deriva de forma conjunta mediante la formulación de problemas de minimización de la potencia media transmitida con restricciones de BER, tasa y acceso al canal. Además de los propios algoritmos de cuantificación y adaptación, se proporciona una visión del funcionamiento de este tipo de algoritmos y de los límites de prestaciones de sistemas TDMA limitados en potencia con realimentación cuantificada. El marco teórico propuesto incorpora prioridad relativa entre usuarios con garantías de *justicia social*¹, el uso de modos AMC y la adaptación de potencia. Puesto que la separación entre canales ascendente (de los usuarios al PA) y descendente (del PA a los usuarios) basada en TDD no es infrecuente en sistemas TDMA, el modelo teórico se aplica en profundidad para caracterizar la solución basada en P-CSIT (apartado 4.1.2). Este esquema se utiliza también para inicializar la solución basada en Q-CSIT que se desarrolla en el apartado 4.1.3, así como para obtener un nivel asintótico de referencia para las prestaciones del diseño de realimentación limitada. Una vez que la constelación con los niveles cuantificados de potencia está diseñada, la asignación de acceso de los usuarios y los esquemas de adaptación de potencia y modo de transmisión se resuelven de manera computacionalmente eficiente y con garantías de convergencia

¹El concepto de *justicia social*, término tomado de la literatura económica, aparece asociado a problemas de optimización en distintos ámbitos –financiero, industrial, comunicaciones–, en los que debe distribuirse una serie de recursos limitados de manera que maximizando (minimizando) una función de utilidad (coste) global –t.c.c. utilidad social– se garantice a cada usuario unos niveles mínimos de acceso al recurso.

al óptimo global mediante algoritmos de optimización convexa.

El punto clave del apartado 4.1.3 (y por ende de la sección) es el desarrollo de un algoritmo iterativo (basado en el principio de descenso coordinado por bloques) que resuelve de forma conjunta la cuantificación y la adaptación. La resolución conjunta se realizará de forma “off-line” y tendrá las ventajas añadidas de generar una adaptación “on-line” de baja complejidad y de presentar unas necesidades (tasas) de realimentación sorprendentemente pequeñas. Las simulaciones numéricas del apartado 4.1.4 comparan las prestaciones del esquema Q-CSIT propuesto con las de otros esquemas heurísticos existentes así como con sus límites asintóticos representados por el esquema P-CSIT. La recopilación de resultados que finaliza esta sección incide en la utilidad tanto práctica (considerable ahorro de potencia, tasa de realimentación pequeña, complejidad computacional “off-line” acotada y complejidad “on-line” despreciable) como teórica (otros esquemas de tasa de realimentación limitada que asignen recursos deberían comparar sus prestaciones con las del aquí derivado) de los esquemas propuestos.

4.1.1. Modelo y preámbulos

Considérense M terminales transmisores (usuarios) comunicándose con un PA a través de canales inalámbricos con desvanecimientos en un entorno TDMA (enlace ascendente). El sistema trabajará bajo las siguientes condiciones de operación:

(sp1) Cada usuario (indexado vía $m \in \{1, \dots, M\}$) transmite utilizando un conjunto finito de modos AMC (indexados vía $l \in \{1, \dots, L_m\}$) cada uno de ellos con una tasa de transmisión asociada $r_{m,l}$.

(sp2) Los coeficientes del canal $\{h_m\}_{m=1}^M$ son planos en frecuencia (i.e., escalares para cada usuario m), permanecen constantes durante el tiempo de duración de un intervalo de transmisión T_i pero pueden variar de un intervalo a otro (modelo de desvanecimiento por bloques). Compactando las ganancias del canal (medidas en términos de potencia) en un vector de tamaño $M \times 1$ se obtiene $\mathbf{g} := [|h_1|^2, \dots, |h_M|^2]^T$, que es ergódico con FDPA $F_{\mathbf{g}}(\mathbf{g})$ supuesta conocida (si,

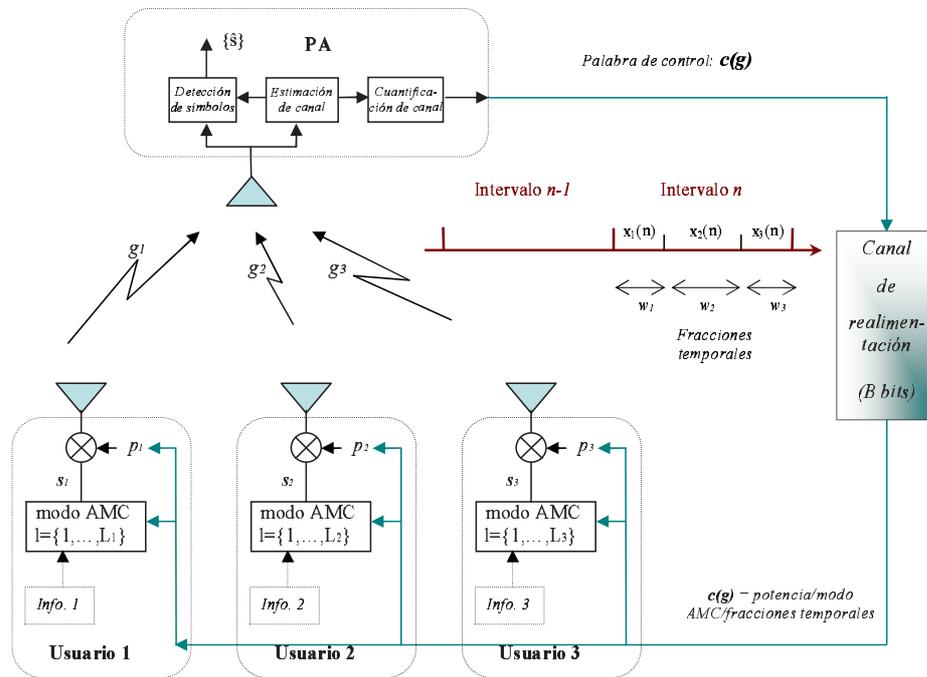


Figura 4.1: Modelo del sistema TDMA considerado para $M = 3$ usuarios.

como se hizo en el capítulo anterior, se tomase el ejemplo de canales Rayleigh, $\{h_m\}_{m=1}^M$ serían gaussianas complejas de media cero).

Todos los usuarios pueden transmitir (sin solaparse) durante un intervalo (bloque) utilizándolo durante una fracción de la duración del mismo $\{w_m(\mathbf{g})\}_{m=1}^M$ (la duración absoluta del intervalo de transmisión T_i dependerá de la duración de la realización del canal \mathbf{g} ; i.e., con el tiempo de coherencia T_c). Si se supone, s.p.d.g., que cada intervalo de transmisión tiene duración unitaria, claramente $\sum_{m=1}^M w_m(\mathbf{g}) \in [0, 1]$. Nótese que la condición anterior permite a todos los usuarios, o en el extremo opuesto a ningún usuario, transmitir durante un intervalo dado. Además, si $r_m(\mathbf{g})$ denota la tasa de transmisión (medida en bits/segundo/Hertzio), la tasa transmitida por el m -ésimo usuario durante el intervalo considerado es $r_m(\mathbf{g})w_m(\mathbf{g})$; del modo semejante $p_m(\mathbf{g})w_m(\mathbf{g})$ será la potencia transmitida por el usuario m durante dicho intervalo.

De acuerdo a (sp1) y del mismo modo que en la sección 3.1, el usuario m puede elegir en cada intervalo una modulación con tasa $r_{m,l}^{(\text{mod})}$ (e.g., 16-QAM)

junto con una codificación de canal con tasa $r_{m,l}^{(\text{cod})}$ (e.g., un código convolucional de tasa 1/2) para transmitir conforme al l -ésimo modo AMC con tasa $r_m(\mathbf{g}) = r_{m,l} := r_{m,l}^{(\text{mod})} r_{m,l}^{(\text{cod})}$. Aparte de con las tasas pre-especificadas por los modos AMC $r_{m,l}$ (que como se deduce por el uso del sub-índice m pueden ser diferentes para cada usuario), es también posible para cada terminal transmitir con una tasa resultante de una combinación lineal de los valores pre-especificados a través de la fragmentación de su propio intervalo de transmisión. Así por ejemplo, utilizando el modo l durante una porción $\theta_{m,l}$ de su propio intervalo w_m y el modo $l + 1$ durante la porción restante $(1 - \theta_{m,l})$ del tiempo, el usuario m puede transmitir durante un bloque con tasa $r_m(\mathbf{g})w_m(\mathbf{g}) = \tilde{w}_{m,l}(\mathbf{g})r_{m,l} + \tilde{w}_{m,l+1}(\mathbf{g})r_{m,l+1}$, donde $\tilde{w}_{m,l}(\mathbf{g}) := \theta_{m,l}w_m(\mathbf{g})$ y $\tilde{w}_{m,l+1}(\mathbf{g}) := (1 - \theta_{m,l})w_m(\mathbf{g})$. En general, el usuario m puede transmitir con tasa

$$r_m(\mathbf{g})w_m(\mathbf{g}) = \sum_{l=0}^{L_m} \tilde{w}_{m,l}(\mathbf{g})r_{m,l}, \quad (4.1)$$

haciendo notar que el primer modo (denotado en este caso con $l = 0$) corresponde a una tasa de transmisión $r_{m,0} = 0$ (en cuyo caso el usuario renuncia a utilizar el canal) y el último modo ($l = L_m$) corresponde a la máxima tasa de transmisión (r_{m,L_m}) con la que el m -ésimo usuario puede transmitir.

A fin de satisfacer requisitos de QoS *específicos para cada usuario*, las comunicaciones del sistema TDMA deben cumplir restricciones de tasa de transmisión *media* y BER *media*. La tasa media del usuario m está dada por

$$\bar{r}_m := \mathbb{E}_{\mathbf{g}}[r_m(\mathbf{g})w_m(\mathbf{g})] := \int_{\mathbf{g}} r_m(\mathbf{g})w_m(\mathbf{g})dF_{\mathbf{g}}(\mathbf{g}) \quad (4.2)$$

y su valor debe estar por encima del nivel promedio pre-especificado; i.e., $\bar{r}_m \geq \check{r}_m$ con $\check{r}_m \leq r_{m,L_m} \forall m$.

Del mismo modo, el requisito de BER media deberá satisfacer $\bar{\epsilon}_m \in [0, \check{\epsilon}_m]$ $\forall m = 1, \dots, M$, donde

$$\epsilon_m = \epsilon_m(g_m p_m, r_m) \quad (4.3)$$

representa la función de la BER instantánea de forma que depende de la tasa transmitida r_m y de la potencia recibida $g_m p_m$ (al igual que en anteriores ocasiones y s.p.d.g. se fija a uno el valor de la varianza del ruido AWGN en el receptor). De forma general ϵ_m será una combinación lineal de las BER instantáneas $\epsilon_{m,l}$ asociadas a los distintos modos l utilizados para transmitir r_m .

La relación entre tasa transmitida, potencia transmitida y BER expresada en (4.3) jugará un papel instrumental para lograr el *objetivo* de minimización de la *potencia media ponderada* \bar{P} dados los requisitos $\{\check{r}_m, \check{\epsilon}_m\}_{m=1}^M$ y la disponibilidad de la información del estado del canal en los transmisores (CSIT). La ponderación se realizará considerando los pesos $[\mu_1 \cdots \mu_M]^T := \boldsymbol{\mu}$ que expresan la prioridad relativa entre usuarios (e.g., si $\mu_1/\mu_2 = 2$, en el problema de minimización el consumo de potencia del usuario 1 será dos veces más importante que el del usuario 2), resultando el objetivo a minimizar

$$\bar{P} = \sum_{m=1}^M \mu_m \bar{p}_m := \sum_{m=1}^M \mu_m \mathbb{E}_{\mathbf{g}} [p_m(\mathbf{g}) w_m(\mathbf{g})] = \sum_{m=1}^M \mu_m \int_{\mathbf{g}} p_m(\mathbf{g}) w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}). \quad (4.4)$$

Las dos únicas formas de CSIT consideradas son perfecto y cuantificado. El Q-CSIT llega al transmisor a través del canal de realimentación en forma de la palabra $\mathbf{c} = \mathbf{c}(\mathbf{g})$. Pese a que la descripción precisa del cuantificador de canal y de la formación de la palabra $\mathbf{c}(\mathbf{g})$ se realiza en el apartado 4.1.3, a fin de cerrar la descripción del modelo del sistema deben adelantarse ciertos aspectos. Asumiendo que cada usuario utilizará durante un periodo de transmisión un único modo AMC, si g_m pertenece a la región de cuantificación $\mathcal{R}_{m,l}$ en la que el usuario m puede utilizar el modo l , para describir (i.e., indexar) la región de cuantificación para todos los usuarios la palabra $\mathbf{c}(\mathbf{g})$ debería contener al menos $\lceil \sum_{m=1}^M \log_2 L_m \rceil$ bits (i.e., $\prod_{m=1}^M L_m$ estados de cuantificación). En apartados posteriores se comprobará que la utilización de un único modo AMC no penaliza el diseño del cuantificador y que el número de bits de la palabra $\mathbf{c}(\mathbf{g})$ es considerablemente menor que el inicialmente esperado.

Puesto que la equidad (o alternativamente la justicia social) es un aspecto de

gran relevancia en sistemas de comunicaciones con múltiples usuarios, debe cerrarse este apartado remarcando que el planteamiento propuesto garantiza *justicia* en la asignación de recursos a través de dos mecanismos diferentes: (i) el vector de *prioridades* $\boldsymbol{\mu}$ y (ii) las restricciones de *tasa* media mínima $\{\check{r}_m\}_{m=1}^M$ para cada *usuario*.

4.1.2. Adaptación óptima de recursos basada en P-CSIT

En este apartado se derivan esquemas de asignación de recursos que minimizan el coste de potencia para TDMA en base a AMC y P-CSIT. La motivación de este estudio es triple: (i) para el uso de los esquemas derivados en sistemas TDD que de forma realista pueden adquirir P-CSIT; (ii) para orientar el diseño del esquema de Q-CSIT apoyándose en la estructura de la solución basada en P-CSIT y (iii) para juzgar la calidad de prestaciones por parte del esquema de Q-CSIT comparándolo con la cota inferior que representa el consumo del esquema de P-CSIT.

Dada la disponibilidad de P-CSIT (i.e., suponiendo \mathbf{g} conocido), la tasa media del usuario m -ésimo puede expresarse en términos de las tasas fijas asociadas a los modos AMC $\{r_{m,l}\}$ y a las fracciones desconocidas $\{\tilde{w}_{m,l}\}$ [cf. (4.1) y (4.2)] $\bar{r}_m = \mathbb{E}_{\mathbf{g}}[\sum_{l=0}^{L_m} \tilde{w}_{m,l}(\mathbf{g})r_{m,l}]$. De la misma manera, dados g_m y $\check{\epsilon}_m$, para cada tasa r_m puede resolverse (4.3) c.r.a. la potencia transmitida para obtener $p_m = (1/g_m)\epsilon_m^{-1}(\check{\epsilon}_m, r_m)$ (como ya se hizo en (3.10) bajo la premisa de utilizar un único modo de transmisión). Mas aún, puesto que ϵ_m es una función convexa y monótona, cualquier tasa que, como en (4.1), se exprese a través una combinación lineal de tasas $\{r_{m,l}\}$, origina también la misma combinación lineal de las potencias transmitidas ($\{p_{m,l}(g_m)\}$) asociadas a cada uno de los modos AMC que satisfacen el requisito de BER $\check{\epsilon}_m$ para una g_m dada. Por consiguiente, la potencia transmitida por el usuario m durante la realización \mathbf{g} puede expresarse en términos de las potencias $\{p_{m,l}(g_m)\}$ y de las fracciones desconocidas $\{\tilde{w}_{m,l}\}$ como $p_m(\mathbf{g}) = \sum_{l=0}^{L_m} \tilde{w}_{m,l}(\mathbf{g})p_{m,l}(g_m)$, teniendo entonces $\bar{p}_m = \mathbb{E}_{\mathbf{g}}[p_m(\mathbf{g})] = \mathbb{E}_{\mathbf{g}}[\sum_{l=0}^{L_m} \tilde{w}_{m,l}(\mathbf{g})p_{m,l}(g_m)]$.

De acuerdo a lo anterior, tanto la potencia como la tasa transmitidas por usuario pueden expresarse como una función de $\{\tilde{w}_{m,l}(\mathbf{g})\}$ y se encuentran acopladas de forma que el requisito de BER esperada se satisface para cada realización \mathbf{g} . De este modo, el esquema de adaptación que en base a P-CSIT (i.e., dado \mathbf{g}) que minimiza la potencia transmitida satisfaciendo requisitos de QoS se reduce a encontrar el vector $\tilde{\mathbf{w}}(\mathbf{g})$ que contiene las fracciones de transmisión $\{\tilde{w}_{m,l}(\mathbf{g}), m = 1, \dots, M, l = 1, \dots, L_m\}$ como solución del problema de optimización con restricciones

$$\left\{ \begin{array}{l} \min_{\tilde{\mathbf{w}}(\mathbf{g})} \quad \bar{P}, \quad \bar{P} = \sum_{m=1}^M \mu_m \mathbb{E}_{\mathbf{g}} \left[\sum_{l=0}^{L_m} \tilde{w}_{m,l}(\mathbf{g}) p_{m,l}(g_m) \right] \\ \text{s. a:} \quad C1. \mathbb{E}_{\mathbf{g}} \left[\sum_{l=0}^{L_m} \tilde{w}_{m,l}(\mathbf{g}) r_{m,l} \right] \geq \check{r}_m, \quad m = 1, \dots, M, \\ \quad \quad C2. 0 \leq \sum_{m=1}^M \sum_{l=0}^{L_m} \tilde{w}_{m,l}(\mathbf{g}) \leq 1, \quad \forall \mathbf{g}. \end{array} \right. \quad (4.5)$$

Relegando de forma temporal la restricción C2, la función dual Lagrangiana del problema de optimización puede escribirse como $\mathcal{L}^{P_w}(\boldsymbol{\beta}^{P_w^*}) = \sum_{m=1}^M \sum_{l=0}^{L_m} \mathbb{E}_{\mathbf{g}} [\varphi_{m,l}^{P_w}(\mathbf{g}, \beta_m^{P_w^*}) \tilde{w}_{m,l}(\mathbf{g})]$, donde la función de coste *instantáneo* $\varphi_{m,l}^{P_w}(\mathbf{g}, \beta_m^{P_w^*}) := \mu_m p_{m,l}(g_m) - \beta_m^{P_w^*} r_{m,l}$ depende de la m -ésima componente del vector óptimo de multiplicadores de Lagrange $\boldsymbol{\beta}^{P_w^*} := [\beta_1^{P_w^*} \dots \beta_M^{P_w^*}]^T$ correspondiente a la m -ésima restricción de tasa en C2.² Si se definen el índice de usuario y el índice del modo para los que esta función de coste se minimiza como

$$(m^*, l_{m^*}^*) := \arg \min_{(m,l)} \varphi_{m,l}^{P_w}(\mathbf{g}, \beta_m^{P_w^*}), \quad \varphi_{m^*,l_{m^*}^*}^{P_w}(\mathbf{g}, \beta_{m^*}^{P_w^*}) := \mu_{m^*} p_{m^*,l_{m^*}^*}(g_{m^*}) - \beta_{m^*}^{P_w^*} r_{m^*,l_{m^*}^*}, \quad (4.6)$$

se deduce rápidamente que $\sum_{m=1}^M \sum_{l=0}^{L_m} \varphi_{m,l}^{P_w}(\mathbf{g}, \beta_m^{P_w^*}) \tilde{w}_{m,l}(\mathbf{g}) \geq \varphi_{m^*,l_{m^*}^*}^{P_w}(\mathbf{g}, \beta_{m^*}^{P_w^*}) \sum_{m=1}^M \sum_{l=0}^{L_m} \tilde{w}_{m,l}(\mathbf{g}) \forall \mathbf{g}$. Pero para $\varphi_{m^*,l_{m^*}^*}^{P_w}(\mathbf{g}, \beta_{m^*}^{P_w^*}) \neq 0$ esta última cota inferior se satisface como una igualdad si el intervalo de transmisión se asigna completamente al usuario m^* y este transmite con su $l_{m^*}^*$ -ésimo modo AMC. De esta forma, para la

²A lo largo de la sección β (respectivamente χ) denotará el multiplicador de Lagrange asociado a la ligadura de tasa media (BER). El superíndice P_w^* (también utilizado para las funciones de coste φ) indicará que el valor corresponde a la solución P-CSIT y se utiliza para obtener el valor óptimo de w . Del mismo modo, se escribirá, e.g, $\beta^{Q_p^*}$ cuando se compute la asignación óptima de potencia p en el caso Q-CSIT.

asignación óptima de fracciones de transmisión por realización \mathbf{g} (i.e., adaptación óptima en base a P-CSIT) se tiene:

$$\begin{aligned} \text{Si } \varphi_{m^*,l_m^*}^{P_w}(\mathbf{g}, \beta_m^{P_w^*}) \neq 0, \text{ entonces } \quad & \tilde{w}_{m^*,l_m^*}^*(\mathbf{g}) = 1 \quad \text{y} \\ & \tilde{w}_{m,l}^*(\mathbf{g}) = 0 \quad \text{para } (m,l) \neq (m^*,l_m^*); \end{aligned} \quad (4.7)$$

o el caso extremo (de resolución trivial) en el que todos los usuarios deben renunciar a la transmisión:

$$\text{Si } \varphi_{m^*,l_m^*}^{P_w}(\mathbf{g}, \beta_m^{P_w^*}) = 0, \text{ entonces } \quad \tilde{w}_{m,l}^*(\mathbf{g}) = 0 \quad \forall m = 1, \dots, M, l = 1, \dots, L_m. \quad (4.8)$$

La asignación óptima en (4.7) claramente satisface $\sum_{l=0}^{L_m} \tilde{w}_{m,l}(\mathbf{g}) = 1$, del mismo modo que la de (4.8) hace que esta suma sea 0. Puesto que bien una u otra son óptimas para cada realización de \mathbf{g} y satisfacen la restricción C2, son la solución óptima de (4.5) siempre y cuando se obtenga un método para calcular el valor óptimo del vector de multiplicadores de Lagrange $\beta^{P_w^*}$ que satisfaga el conjunto de restricciones en C1. Nótese que en general $\varphi_{m,l}^{P_w}(\mathbf{g}, \beta_m^{P_w^*}) \leq 0 \quad \forall m$ puesto que en caso contrario la asignación extrema de renuncia global a la transmisión estaría continuamente activa. Con la asignación óptima descrita en (4.7) el valor óptimo de la tasa (potencia) media puede calcularse tras retener únicamente el sumando r_{m^*,l_m^*} (respectivamente $p_{m^*,l_m^*}(g_m)$) en (4.5) e integrar después sobre la FDP de \mathbf{g} .

La relación convexa entre potencia y tasa en (4.5) hace que para encontrar $\beta_m^{P_w^*}$ haya que satisfacer las ligaduras en C1 de forma ajustada (i.e., $\bar{r}_m = \check{r}_m$). Más concretamente, si el argumento (i) denota el índice de iteración y $\delta_\beta > 0$ un tamaño de paso, $\beta_m^{P_w^*}$ se encontrará al converger la siguiente recursión

$$\beta_m^{P_w}(i+1) = [\beta_m^{P_w}(i) + \delta_\beta(\check{r}_m - \mathbb{E}_{\mathbf{g}}[r_m(i; \mathbf{g})\tilde{w}_{m,l}(i; \mathbf{g})])]^+, \quad (4.9)$$

donde $[x]^+$ asegura que el multiplicador de Lagrange nunca sea negativo. La recur-

sión en (4.9) representa una actualización por sub-gradiente [13], cuya convergencia al *único* óptimo global $\beta_m^{P_w^*} \forall m$ se produce de forma rápida y garantizada desde cualquier punto inicial debido a la aludida convexidad de (4.5).

En la práctica, si la integración para resolver $\mathbb{E}_{\mathbf{g}}$ en (4.9) resulta complicada (analítica o numéricamente), su cálculo puede ser reemplazado por $N^{-1} \sum_{n=1}^N r_m(i; \mathbf{g}^{(n)}) \tilde{w}_{m,l}(i; \mathbf{g}^{(n)})$, donde $\{\mathbf{g}^{(n)}\}_{n=1}^N$ son realizaciones de \mathbf{g} generadas de acuerdo a $F_{\mathbf{g}}(\mathbf{g})$, lo que resulta particularmente sencillo cuando los canales de los distintos usuarios están incorrelados (caso típico) o cuando $\{h_m\}_{m=1}^M$ aun correlados, son gaussianas complejas, puesto que la asignación óptima en (4.7) o (4.8) establece que a lo sumo una pareja usuario-modo (m, l) será distinta de cero para cada muestra $\mathbf{g}^{(n)}$. Nótese que el valor de $\beta_m^{P_w}(i-1)$ se utiliza ahora en (4.6) para calcular el usuario $m^*(i; \mathbf{g}^{(n)})$ y el modo $l_{m^*}^*(i; \mathbf{g}^{(n)})$ ganadores para la realización $\mathbf{g}^{(n)}$, basados en los cuales se encuentra $r_m(i; \mathbf{g}^{(n)}) = r_{m^*(i; \mathbf{g}^{(n)}), l_{m^*}^*(i; \mathbf{g}^{(n)})}$ y $\tilde{w}_{m,l}(i; \mathbf{g}^{(n)}) = 1$ para $m = m^*(i; \mathbf{g}^{(n)})$ y $l = l_{m^*}^*(i; \mathbf{g}^{(n)})$, siendo cero en otro caso. Como criterio de parada para las iteraciones en (4.9), se comprueba si la diferencia $|\mathcal{L}^{P_w}(\beta^{P_w}(i+1)) - \mathcal{L}^{P_w}(\beta^{P_w}(i))| / |\mathcal{L}^{P_w}(\beta^{P_w}(i+1))|$ medida en términos de la función dual (definida justo después de (4.5)) cae por debajo de un nivel de tolerancia preseleccionado, en cuyo caso se devuelve $\beta_m^{P_w^*} = \beta_m^{P_w}(i+1) \forall m$.

Antes de resumir el funcionamiento del esquema de adaptación de los usuarios basado en P-CSIT, es conveniente prestar atención a qué parámetros y fases del algoritmo deben ser ejecutadas “off-line” durante la fase de inicial del sistema y qué parámetros y cálculos deben ejecutarse de forma “on-line” en tiempo real. Uno de los puntos clave reside en el carácter estadístico de las restricciones $C1$ en (4.9), que al estar expresadas en términos de valor esperado, su satisfacción requiere únicamente conocimiento de la FDPA $F_{\mathbf{g}}(\mathbf{g})$, lo que hace que los valores óptimos de los multiplicadores de Lagrange puedan ser calculados “off-line” utilizando información estadística a largo plazo del canal inalámbrico (e.g., media, varianza) que permitan caracterizar su FDPA. Sorprendentemente, puesto que el usuario m^* conoce $\beta_m^{P_w^*}$ de la fase “off-line” y g_{m^*} a través de la estimación del canal inverso, si el PA difundiese a los usuarios la palabra $\mathbf{c}(m^*)$ indexando el

usuario óptimo (que calcula haciendo uso de $\beta^{P_w^*}$ y \mathbf{g}) no habría necesidad de que los usuarios conocieran la totalidad del vector \mathbf{g} sino únicamente su propio canal (situación que correspondería al CSIT individual analizado en el apartado 3.1.4). A continuación el usuario ganador utilizará su modo óptimo $l_{m^*}^*$ (cuyo índice bien puede ser calculado por él mismo según (4.6) o bien puede ser enviado también por el PA, no habiendo entonces necesidad de que los usuarios conozcan el valor $\beta_m^{P_w^*}$ lo que hace que sea una alternativa si cabe más práctica) para transmitir con tasa $r_{m^*,l_{m^*}^*}$ y potencia $p_{m^*,l_{m^*}^*}(g_m) = (1/g_m) \epsilon_m^{-1}(\check{\epsilon}_m, r_{m^*,l_{m^*}^*})$. Esta necesidad mínima de tasa de realimentación, así como la sencillez en el esquema de adaptación se deben a la formulación del problema con requisitos de tasa media (frente a la formulación con requisitos instantáneos como, e.g., los considerados en la sección 3.2). Asimismo, el uso de modos AMC permite que, estando disponible P-CSIT, los requisitos para el valor esperado de la BER se satisfagan para cada realización del canal (y por tanto también en forma promediada) no teniendo que incluirlos así de forma explícita en el problema de optimización. Finalmente, puesto que las restricciones de tasa están desacopladas entre usuarios, una vez que se determina el usuario ganador este no necesita conocer el valor de \mathbf{g} sino únicamente el de su propia realización g_{m^*} .³ En suma, se ha establecido que para el problema basado en P-CSIT:

Proposición 4.1 *Bajo (sp1) y (sp2), la minimización de la potencia media ponderada transmitida satisfaciendo requisitos de BER y tasa media se reduce al problema de optimización con ligaduras presentado en (4.5) con variables de optimización $\tilde{w}_{m,l}$, problema que es convexo y tiene un único mínimo global. La solución corresponde a una asignación avariciosa [cf. (4.6) y (4.7) o (4.8)] donde a lo sumo el usuario con mínimo coste m^* transmitirá durante toda la duración del intervalo (bloque) de acceso con el modo AMC de mínimo coste $l_{m^*}^*$ y potencia mínima que adaptada al P-CSIT g_{m^*} satisfaga el requisito de BER $\check{\epsilon}_{m^*}$, i.e., $p_{m^*,l_{m^*}^*}(g_{m^*}, \check{\epsilon}_{m^*})$. Los multiplicadores de Lagrange $\{\beta_m^{P_w^*}\}_{m=1}^M$ que se necesitan pa-*

³Este comportamiento contrasta con el obtenido para la sección 3.1 donde las prestaciones basadas en CSIT pleno (\mathbf{g}) y CSIT individual (g_m y $\mathbf{c}(\mathbf{g})$) eran diferentes.

ra encontrar el usuario de mínimo coste pueden calcularse en el PA de forma “off-line” utilizando la FDPA del canal; mientras que durante la fase “on-line” los usuarios únicamente necesitan recibir desde el PA el mensaje de realimentación $\mathbf{c}(\mathbf{g}) = [\mathbf{c}^T(m^*), \mathbf{c}^T(l_{m^*}^*)]^T$, en base al cual el m^* -ésimo usuario transmitirá con la tasa y potencia correspondientes mientras el resto de usuarios permanecerá silente.

Nótese que pese a que el modelo propuesto permitía a los usuarios compartir el intervalo de transmisión y a que cada uno de ellos utilizase distintos modos de transmisión, la asignación óptima derivada es un esquema avaricioso (t.c.c. oportunista) en el que el acceso al canal (recurso) lo realiza el usuario que para esa realización entraña un menor coste, transmitiendo además con único modo AMC. Respecto al coste $\varphi_{m,l}^{P_w}(\mathbf{g}, \beta_m^{P_w^*})$, puede interpretarse como un agregado de costes (o coste neto) compuesto en este caso por dos términos distintos: por un lado la potencia transmitida ponderada por la prioridad del usuario (coste) y por otro lado y de signo opuesto la tasa transmitida ponderada por el precio óptimo de la tasa expresado en términos de potencia y recogido por el multiplicador $\beta_m^{P_w^*}$ (utilidad o beneficio). Si para una realización dada existieran múltiples usuarios que alcanzasen el mínimo, cualquier distribución aleatoria del acceso entre los mismos (renunciando obligatoriamente a la transmisión, claro está, aquellos que no alcanzan el mínimo) sería óptima. Téngase también en cuenta que conforme mayor es la realización del canal para el usuario ganador g_{m^*} mayor es la tasa transmitida $r_{m^*}(g_{m^*})$ que puede tolerarse satisfaciendo $\check{\epsilon}$. En el extremo opuesto, si todos los usuarios experimentan un desvanecimiento severo ($g_m \approx 0 \forall m$) se tendrá que $l = 0 \forall m$ y $r_{m^*}(g_{m^*}) = 0$ en cuyo caso todos los usuarios deben renunciar a acceder al canal [cf. (4.8)]. Las anteriores observaciones muestran que la asignación óptima descrita por la proposición 4.1 sigue el espíritu de de la asignación óptima dictada por el “water-filling” en entornos multi-usuario.

Observación 4.1 *La asignación basada en el principio de el ganador consigue todo⁴ no es ajena ni a los problemas de distribución de recursos entre usuarios en*

⁴Expresión también común para referirse al comportamiento *oportunista* o *avaricioso*.

general ni a la gestión específica de sistemas comunicación (e.g., maximización de la capacidad en sistemas multi-usuario de acceso ortogonal con códigos de infinita longitud y restricciones de potencia media [40]). Sin embargo, no resultaba sencillo intuir (prever) que este fuera también el caso también para el problema aquí formulado. El hecho de que la optimización de sistemas TDMA con modos AMC tenga una sencilla asignación oportunista con una baja tasa de realimentación respalda la aplicación práctica de este tipo de sistemas.

4.1.3. Adaptación de recursos y cuantificación óptimas basadas en Q-CSIT

En esta sección se analiza el problema de minimización de potencia con garantías de QoS para un sistema TDMA en base a modos AMC y Q-CSIT, un escenario particularmente adecuado para sistemas FDD limitados por potencia. Cuando las regiones de cuantificación $\mathcal{R} := \{\mathcal{R}_{m,l}, \forall(m,l)\}$ están diseñadas a priori, los esquemas en esta sección encuentran la asignación de usuario, modo y potencia globalmente óptima. Asimismo, para el caso general en el que las regiones son desconocidas, en este apartado se propone un algoritmo sistemático de descenso coordinado por bloques que optimiza de forma conjunta la asignación de recursos (adaptación de los transmisores) y la cuantificación del canal definiendo así por completo el esquema Q-CSIT a utilizar.

Antes de plantear el problema, resulta instructivo señalar las mayores diferencias entre el sistema Q-CSIT a diseñar y el sistema P-CSIT ya analizado. Estas diferencias aluden básicamente a dos facetas diferentes:

(d1) Puesto que la palabra $\mathbf{c}(\mathbf{g})$ que describe el Q-CSIT contiene un número finito B de bits, únicamente existe un número finito de configuraciones (específicamente 2^B) a las que los terminales pueden adaptarse.

(d2) La ausencia del conocimiento de \mathbf{g} hace que el sencillo acoplamiento de potencia y tasa a través del requisito de BER (a través de (4.3)) ya no sea posible. Este hecho tiene tres implicaciones fundamentales: (i) debe incorporarse al pro-

blema de optimización un nuevo conjunto de variables $\mathbf{p} := [\mathbf{p}_1^T \cdots \mathbf{p}_M^T]^T$, donde $\{\mathbf{p}_m := [p_{m,1} \cdots p_{m,L_m}]^T\}_{m=1}^M$; (ii) deben tenerse en cuenta de forma explícita un conjunto adicional de multiplicadores de Lagrange asociado a las restricciones de BER esperada y (iii) no resulta posible expresar la potencia transmitida en términos de las fracciones $\{\tilde{w}_{m,l}(\mathbf{g})\}$ que permitían combinaciones lineales entre modos, lo que implica que en esta sección únicamente se utilizarán las fracciones originales de acceso TDMA $w_m(\mathbf{g})$ por usuario m que formarán el vector de acceso $\mathbf{w}(\mathbf{g}) := [w_1(\mathbf{g}) \cdots w_M(\mathbf{g})]^T$.

Puesto que un cuantificador es esencialmente un clasificador, para definir el Q-CSIT habrá que especificar su entrada, salida, tipo y número de clases y su relación con el problema que se analiza. Esta caracterización puede sintetizarse como sigue:

(sp3) *Para cada terminal m se consideran L_m clases, tantas como modos AMC; la entrada del cuantificador es \mathbf{g} y la salida es la palabra $\mathbf{c}(l_m^*; \mathbf{g})$ que indexa el modo l_m^* ; si $\mathbf{g} \in \mathcal{R}_{m,l_m^*}$ y el terminal m es seleccionado por el esquema de asignación durante una fracción $0 \leq w_m \leq 1$ transmitirá con tasa $r_m = r_{m,l_m^*}$ y potencia $p_m = p_{m,l_m^*}$ ambos supuestos constantes a lo largo de la región \mathcal{R}_{m,l_m^*} . Si no se especifican a priori, las regiones $\mathcal{R}_{m,l}$ serán también optimizadas. Al igual que en el caso de P-CSIT se obtendrá que la asignación óptima del sistema es oportunista y, por tanto, la palabra $\mathbf{c}(\mathbf{g})$ que representa el Q-CSIT de todo el sistema deberá contener el índice del terminal seleccionado $\mathbf{c}(m^*; \mathbf{g})$ y su modo de transmisión $\mathbf{c}(l_{m^*}^*; \mathbf{g})$.*

La principal característica del diseño del cuantificador propuesto en (sp3) consiste en acoplar cada modo AMC con una región de cuantificación y, por consiguiente, con una única potencia de transmisión, esta decisión está avalada por el hecho de que la solución basada en P-CSIT pese a permitirlo a priori selecciona un único modo AMC de transmisión. Este acoplamiento facilitará la formulación convexa del problema y reducirá las necesidades de realimentación del sistema, haciendo que la palabra $\mathbf{c}(\mathbf{g}) := [\mathbf{c}^T(m^*; \mathbf{g}), \mathbf{c}^T(l_{m^*}^*; \mathbf{g})]^T$ conteniendo el Q-CSIT transporte únicamente $B = \lceil \log_2 M + \log_2(\max_m L_m) \rceil$ bits.

También contribuye a la reducción de la sobrecarga el hecho de que la cuanti-

ficación en (sp3) esté desacoplada para un mismo usuario. En cierto modo podría decirse que se tienen M problemas de cuantificación cada uno de ellos involucrando L_m clases. Las regiones de cuantificación por usuario no se solapan (intersecan) entre distintos modos pero para un mismo modo sí se solapan entre usuarios; i.e., si \cap (\cup) denota la intersección (unión) de conjuntos, se tiene $\mathcal{R}_{m,l} \cap \mathcal{R}_{m,l'} = \emptyset$ para $l \neq l'$, pero $\mathcal{R}_{m,l} \cap \mathcal{R}_{m',l} \neq \emptyset$ para $m \neq m'$ y, claramente, $\cup_{l=0}^{L_m} \mathcal{R}_{m,l} = \mathcal{G} \subseteq \mathbb{R}_0^{+M} \forall m$, donde \mathcal{G} denota el dominio de \mathbf{g} que es el sub-espacio de vectores con componentes reales positivas o nulas \mathbb{R}_0^{+M} . Si las regiones de cuantificación estuvieran acopladas entre usuarios el problema de cuantificación sería único, con regiones que no se solapan, pero con un número considerablemente mayor de clases ($\prod_{m=1}^M L_m$).

Habiendo descrito la forma del Q-CSIT, se procede a expresar la potencia, tasa y BER medias así como las variables envueltas en los problemas de optimización que se analizarán. Puesto que, de acuerdo a lo recalado en (d1), el número de configuraciones de transmisión para cada usuario es finito, la potencia transmitida por el m -ésimo usuario $p_m(\mathbf{c}(\mathbf{g}))$ por cada realización del canal \mathbf{g} es una VA discreta; mientras que $0 \leq w_m(\mathbf{g}) \leq 1$ es una variable continua⁵. De esta manera, $\bar{p}_m := \mathbb{E}_{\mathbf{g}}[p_m(\mathbf{c}(\mathbf{g}))w_m(\mathbf{g})] = \int p_m(\mathbf{c}(\mathbf{g}))w_m(\mathbf{g})dF_{\mathbf{g}}(\mathbf{g}) = \sum_{l=0}^{L_m} \int_{\mathcal{R}_{m,l}} p_m(\mathbf{c}(\mathbf{g}))w_m(\mathbf{g})dF_{\mathbf{g}}(\mathbf{g})$, donde la última igualdad se debe a la separación de la integral a lo largo de regiones de cuantificación que no se solapan (intersecan). No obstante, puesto que (sp3) establece que la potencia transmitida por región es constante y se tiene además que $p_{m,0} = 0$, se llega a la expresión

$$\bar{p}_m = \sum_{l=1}^{L_m} p_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g})dF_{\mathbf{g}}(\mathbf{g}) . \quad (4.10)$$

Argumentando de forma semejante puede demostrarse fácilmente que la tasa media

⁵Nótese que si $0 < w_m(\mathbf{g}) < 1$ o lo que es lo mismo, durante el intervalo considerado transmitiera más de un usuario, el sistema debería actualizar el Q-CSIT no al final del intervalo, sino cuando la fracción $w_m(\mathbf{g})$ activa concluyese y debiera transmitir el siguiente usuario.

transmitida puede expresarse como

$$\bar{r}_m = \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) \geq \check{r}_m, \quad m = 1, \dots, M \quad (4.11)$$

donde el término integral puede interpretarse como la probabilidad de que el terminal m utilice el modo l (i.e., r_m tome el valor $r_{m,l}$).

El hecho de que en sistemas adaptativos la tasa de transmisión para cada usuario sea una VA impide que la BER pueda calcularse simplemente integrando (4.3) sobre $F_{\mathbf{g}}(\mathbf{g})$, como se hace cuando la tasa r_m es un valor constante determinista. De esta forma, $\bar{\epsilon}_m$ debe calcularse como la razón entre el número promedio de bits erróneos y el número promedio de bits enviados [cf. (3.40)]. Procediendo de forma análoga a lo que se hizo para (4.10) y (4.11), el numerador de esta razón es $\sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon_{m,l}(g_m p_{m,l}) dF_{\mathbf{g}}(\mathbf{g})$, donde $\epsilon_{m,l}$ es la BER instantánea del usuario m utilizando su l -ésimo modo; mientras que el denominador simplemente es la tasa media en (4.11). De esta manera, el requisito de valor medio de la BER bajo (sp1)-(sp3) puede expresarse como

$$\bar{\epsilon}_m = \frac{\sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon_{m,l}(g_m p_{m,l}) dF_{\mathbf{g}}(\mathbf{g})}{\sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g})} \leq \check{\epsilon}_m, \quad m = 1, \dots, M. \quad (4.12)$$

No obstante, (4.12) puede simplificarse de forma sencilla sustituyendo el denominador por el valor de \check{r}_m en (4.11). Si el problema no fuera convexo c.r.a. $\{r_m\}$ esta simplificación podría dar lugar a una solución más conservadora, puesto que podría darse el caso de que $\bar{r}_m > \check{r}_m$. Tras esta sustitución, el problema de optimización que se pretende resolver puede formularse como:

$$\left\{ \begin{array}{l} \min_{\mathcal{R}, \mathbf{p}, \mathbf{w}(\mathbf{g})} \bar{P}, \quad \bar{P} = \sum_{m=1}^M \mu_m \sum_{l=1}^{L_m} p_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) \\ \text{s.a } C1. \quad \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) \geq \check{r}_m, \quad m = 1, \dots, M; \\ \quad C2. \quad \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon_{m,l}(g_m p_{m,l}) dF_{\mathbf{g}}(\mathbf{g}) / \check{r}_m \leq \check{\epsilon}_m, \quad m = 1, \dots, M; \\ \quad C3. \quad 0 \leq \sum_{m=1}^M w_m(\mathbf{g}) \leq 1, \quad \forall \mathbf{g}. \end{array} \right. \quad (4.13)$$

Claramente, si las regiones \mathcal{R} están dadas a priori (utilizando, e.g., cuantificadores uniformes, Lloyd o de máxima entropía por cada g_m), entonces el conjunto de variables se reduce del original $(\mathcal{R}, \mathbf{p}, \mathbf{w}(\mathbf{g}))$ al simplificado $(\mathbf{p}, \mathbf{w}(\mathbf{g}))$. Puesto que para este nuevo conjunto de variables tanto la función objetivo como las restricciones son convexas respecto a \mathbf{p} y $\mathbf{w}(\mathbf{g})$, el problema de asignación de recursos tiene un óptimo global único.

En las páginas siguientes, se resolverá en primer lugar (4.13) c.r.a. $\mathbf{w}(\mathbf{g})$ cuando \mathcal{R} y \mathbf{p} están dados (sub-problema 1). Parte de la solución de este sub-problema será todo lo que se necesite para establecer la adaptación “on-line” óptima basada en Q-CSIT una vez que tanto la constelación de valores óptimos de potencia \mathbf{p}^* como las regiones óptimas de cuantificación \mathcal{R}^* , ambas calculadas “off-line”, estén disponibles. La construcción de \mathbf{p}^* (sub-problema 2) y \mathcal{R}^* (sub-problema 3) se apoyará también en la solución del esquema óptimo de acceso al medio. Para su optimización conjunta, se seguirá un principio de descenso coordinado por bloques [9] (t.c.c. descenso coordinado iterativo), en el que las variables de optimización se dividen en tres grupos: $(\mathcal{R}^{(t-1)}, \mathbf{p}^{(t-1)}, \mathbf{w}(\mathbf{g})^{(t-1)})$. Durante la (t) -ésima iteración del bloque se considerarán tres sub-problemas distintos, en cada uno de ellos dos de los tres vectores $(\mathcal{R}^{(t-1)}, \mathbf{p}^{(t-1)}, \mathbf{w}(\mathbf{g})^{(t-1)})$ se considerarán dados (tomarán el valor que tuvieron en la $(t-1)$ -ésima iteración) y se utilizarán para encontrar el valor del tercero de forma óptima. El algoritmo se iterará hasta que la solución converja (los detalles y caracterización de este algoritmo se proporcionarán después de haber resuelto cada uno de los sub-problemas).

Sub-problema 1: acceso óptimo

En este punto se resuelve (4.13) para $\mathbf{w}(\mathbf{g})$ cuando \mathcal{R} y \mathbf{p} están dadas. La solución resultante será de utilidad en tres situaciones diferentes: (i) para asignar (adaptar) los recursos de transmisión cuando $\mathcal{R} = \tilde{\mathcal{R}}$ y/o $\mathbf{p} = \tilde{\mathbf{p}}$ están dados por el propio diseño del sistema (e.g., las especificaciones de un estándar determinado o las características de construcción de unos receptores que no permitan ajustar las regiones de cuantificación); (ii) para la fase “on-line” donde los valores $\mathcal{R} = \mathcal{R}^*$ y $\mathbf{p} = \mathbf{p}^*$ se han fijado de acuerdo a sus valores óptimos obtenidos durante la fase “off-line” y (iii) para la fase “off-line” para obtener $\mathbf{w}^{(t)}(\mathbf{g})$ cuando $\mathcal{R} = \mathcal{R}^{(t-1)}$ y $\mathbf{p} = \mathbf{p}^{(t-1)}$ están dados por el resultado de la iteración previa del descenso coordinado por bloques.

Puesto que \mathcal{R} es conocido, para una realización dada $\mathbf{g} \in \mathcal{R}_{m,l}$ el modo $l_m(\mathbf{g})$ con tasa de transmisión $r_{m,l_m(\mathbf{g})}$ y la potencia asociada $p_{m,l_m(\mathbf{g})}$ que el usuario m utilizará en caso de ser elegido son conocidos [cf. (sp3)]. Como en el apartado 4.1.2, el Lagrangiano de este problema de optimización con restricciones puede escribirse como $\mathcal{L}^{Q_w}(\boldsymbol{\beta}^{Q_w^*}, \boldsymbol{\chi}^{Q_w^*}) = \sum_{m=1}^M \sum_{l=1}^{L_m} \varphi_m^{Q_w}(\mathbf{g}, \beta_m^{Q_w^*}, \chi_m^{Q_w^*}) \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g})$, donde la función instantánea de coste

$$\begin{aligned} \varphi_m^{Q_w}(\mathbf{g}, \beta_m^{Q_w^*}, \chi_m^{Q_w^*}) &:= \mu_m p_{m,l_m(\mathbf{g})} - \beta_m^{Q_w^*} r_{m,l_m(\mathbf{g})} \\ &+ \chi_m^{Q_w^*} r_{m,l_m(\mathbf{g})} \epsilon_{m,l_m(\mathbf{g})} (g_m p_{m,l_m(\mathbf{g})}) / \check{r}_m \end{aligned} \quad (4.14)$$

depende el multiplicador de Lagrange óptimo $\beta_m^{Q_w^*}$ ($\chi_m^{Q_w^*}$) correspondiente a la m -ésima restricción de tasa (BER esperada) media. Definiendo el índice del usuario que minimiza este coste como $m^* := \arg \min_m \varphi_m^{Q_w}(\mathbf{g}, \beta_m^{Q_w^*}, \chi_m^{Q_w^*})$ y replicando los argumentos que se siguieron para derivar (4.7) y (4.8) se encuentra que, con $\boldsymbol{\beta}^{Q_w^*} := [\beta_1^{Q_w^*} \dots \beta_M^{Q_w^*}]^T$ y $\boldsymbol{\chi}^{Q_w^*} := [\chi_1^{Q_w^*} \dots \chi_M^{Q_w^*}]^T$ dados, la asignación óptima consiste en:

$$\text{Si } \varphi_{m^*}^{Q_w}(\mathbf{g}, \beta_{m^*}^{Q_w^*}, \chi_{m^*}^{Q_w^*}) \neq 0, \text{ entonces } w_{m^*}^*(\mathbf{g}) = 1 \text{ y } w_m^*(\mathbf{g}) = 0 \text{ para } m \neq m^* ; \quad (4.15)$$

o en el caso extremo en donde ningún usuario accede al canal

$$\text{Si } \varphi_m^{Q_w}(\mathbf{g}, \beta_m^{Q_w^*}, \chi_m^{Q_w^*}) = 0, \text{ entonces } w_m^*(\mathbf{g}) = 0 \quad \forall m = 1, \dots, M. \quad (4.16)$$

Las asignaciones de acceso al canal de (4.15) y (4.16) satisfacen claramente la restricción $\sum_{m=1}^M w_m(\mathbf{g}) \in [0, 1]$ y, puesto que una de ellas será cierta para cualquier \mathbf{g} , minimizan también la potencia media transmitida en (4.13) suponiendo que los valores óptimos de los multiplicadores de Lagrange $\beta^{Q_w^*}$ y $\chi^{Q_w^*}$ son conocidos.

En base a $F_{\mathbf{g}}(\mathbf{g})$, los multiplicadores buscados se calculan “off-line” para satisfacer los requisitos de tasa y BER medias $\forall m = 1, \dots, M$ a través de las siguientes iteraciones por sub-gradiente [cf. (4.9)]

$$\beta_m^{Q_w}(i+1) = \left[\beta_m^{Q_w}(i) + \delta_\beta \left(\check{r}_m - \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m^*(i; \mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) \right) \right]^+ \quad (4.17)$$

$$\chi_m^{Q_w}(i+1) = \left[\chi_m^{Q_w}(i) + \delta_\chi \left(\sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m^*(i; \mathbf{g}) \epsilon_{m,l}(g_m p_{m,l}) dF_{\mathbf{g}}(\mathbf{g}) / \check{r}_m - \check{c}_m \right) \right]^+. \quad (4.18)$$

Como en el apartado 4.1.2, las recursiones (4.17) y (4.18) tienen convergencia garantizada al par óptimo $\beta^{Q_w^*}, \chi^{Q_w^*}$ para cualquier condición inicial puesto que el problema en (4.13) es convexo en las variables $\{w_m(\mathbf{g})\}_{m=1}^M$.

De forma semejante a $\mathbb{E}_{\mathbf{g}}$ en (4.9), las integrales en (4.17) y (4.18) pueden ser reemplazadas en la práctica por sumas sobre las realizaciones $\{\mathbf{g}^{(n)}\}_{n=1}^N$. Durante la iteración (i) , si $g_m^{(n)}$ se encuentra en la región $l(i; g_m^{(n)})$ los multiplicadores $\beta_m^{Q_w}(i-1)$ y $\chi_m^{Q_w}(i-1)$ se utilizan en (4.14) para calcular el usuario ganador $m^*(i; \mathbf{g}^{(n)})$ que con un valor de tasa $r_{m^*(i; \mathbf{g}^{(n)}), l(i; g_m^{(n)})}$ contribuye al valor de la integral (ahora suma) en (4.17) y con un valor de potencia $p_{m^*(i; \mathbf{g}^{(n)}), l(i; g_m^{(n)})}$ contribuye a la integral (ahora suma) en (4.18). Nótese que las dos actualizaciones por sub-gradiente por usuario deben ser ejecutadas (y finalizadas) de forma conjunta. De nuevo, el criterio de

parada para las actualizaciones consiste en comprobar la estabilización del valor de la función dual (definida tras presentar (4.13)) comprobando si $|\mathcal{L}^{Q_w}(\beta^{Q_w}(i+1), \chi^{Q_w}(i+1)) - \mathcal{L}^{Q_w}(\beta^{Q_w}(i), \chi^{Q_w}(i))|/|\mathcal{L}^{Q_w}(\beta^{Q_w}(i+1), \chi^{Q_w}(i+1))|$ cae por debajo de un nivel de tolerancia pre-especificado, devolviendo $\beta_m^{Q_w^*} = \beta_m^{Q_w}(i+1)$ y $\chi_m^{Q_w^*} = \chi_m^{Q_w}(i+1) \forall m$ en caso afirmativo.

Con los multiplicadores $\{\beta_m^{Q_w^*}, \chi_m^{Q_w^*}\}_{m=1}^M$ calculados “off-line” utilizando la FDPA de la ganancia del canal, durante la fase de operación “on-line” para cada realización \mathbf{g} el PA necesita: (i) seleccionar para cada usuario m , a partir del diseño del cuantificador \mathcal{R} y \mathbf{p} , la región $\mathcal{R}_{m,l}$ a la que g_m pertenece y, por consiguiente, la tasa $r_{m,l_m(\mathbf{g})}$ y potencia $p_{m,l_m(\mathbf{g})}$ asociadas y (ii) evaluar el coste en (4.14) para todo $m = 1, \dots, M$ a fin de elegir el terminal ganador m^* (mínimo coste) y realimentar la palabra de control $[\mathbf{c}^T(m^*; \mathbf{g}), \mathbf{c}^T(l_{m^*}; \mathbf{g})]^T$ conteniendo el Q-CSIT de manera que el usuario ganador transmita de acuerdo a la tasa y potencia indexada (recuérdese que cada usuario conoce su propia constelación de potencias y modos AMC) o todos los usuarios renuncien a la transmisión para el caso extremo contemplado e (4.16). En suma, se ha probado que:

Proposición 4.2 *Bajo (sp1)-(sp3), si \mathcal{R} y \mathbf{p} están dados, la asignación óptima de acceso al canal basada en Q-CSIT es globalmente óptima, se describe en (4.15) y (4.16) y depende únicamente del funcional de coste $\varphi_m^{Q_w^*}(\mathbf{g}, \beta_m^{Q_w^*}, \chi_m^{Q_w^*})$ en (4.14). Los vectores $\beta^{Q_w^*}$ y $\chi^{Q_w^*}$ conteniendo los valores óptimos de los multiplicadores de Lagrange se calculan de forma “off-line” en el PA utilizando la FDPA del canal, mientras que durante la fase “on-line” de operación se requiere la realimentación a los usuarios de la palabra $\mathbf{c}(\mathbf{g}) = [\mathbf{c}^T(m^*), \mathbf{c}^T(l_{m^*})]^T$ de $\lceil \log_2 M + \log_2(\max_m L_m) \rceil$ bits desde el PA.*

Aunque con distinta definición de coste, debe señalarse que cuando las regiones de cuantificación y la constelación de potencias de transmisión están especificadas, la asignación óptima de acceso al canal basada en Q-CSIT es, al igual que la basada en P-CSIT, oportunista. El coste negativo (o nulo) $\varphi_m^{Q_w^*}(\mathbf{g}, \beta_m^{Q_w^*}, \chi_m^{Q_w^*})$ puede interpretarse como un indicador de la calidad del enlace para el usuario

m (cuanto más pequeño sea el indicador menor será el coste y, por consiguiente, mejor el enlace) basado en el cual a lo sumo un usuario (el de mínimo coste) puede transmitir durante la duración del intervalo. Si hubiera múltiples usuarios $\{m_j^*\}_{j=1}^J$, $J \leq M$ que presentasen el coste mínimo $\varphi_m^{Q_w^*}$, cualquier distribución del acceso entre ellos (incluida la selección aleatoria de únicamente uno de ellos) sería igualmente óptima. La dependencia del coste de los parámetros de transmisión y de $\beta_m^{Q_w^*}$ y $\chi_m^{Q_w^*}$ refuerza la interpretación de $\varphi_m^{Q_w}$ como un agregado de costes (potencia –coste–, tasa –beneficio– y BER –coste–) ponderados por sus respectivos multiplicadores (precios de equilibrio). Al igual que para el caso de P-CSIT donde todos los usuarios renunciaban a la transmisión cuando $\mathbf{g} \approx \mathbf{0}$, la condición de desvanecimiento severo con Q-CSIT se corresponde con tener $\varphi_m^{Q_w}(\mathbf{g}) \geq 0$, $\forall m$. Puesto que el objetivo que se persigue es la minimización de potencia satisfaciendo requisitos medios de tasa y BER, la decisión de permitir la transmisión durante desvanecimientos profundos únicamente reportaría el envío de pequeña tasa de información con grandes costes de potencia y BER. De esta forma se comprueba cómo la intuición detrás de la asignación de la proposición 4.2 consiste en reservar la potencia a transmitir para mejores realizaciones de canal que entrañan un menor coste (mayor recompensa) favoreciendo así la eficiencia en el consumo.

Observación 4.2 *La asignación óptima de recursos para la solución basada en Q-CSIT (así como a la basada en P-CSIT) es socialmente justa aun cuando para una realización de canal dada el acceso se asigne a un único usuario. La justicia en el diseño del sistema se refleja en los parámetros $\{\check{r}_m\}_{m=1}^M$, $\{\check{\epsilon}_m\}_{m=1}^M$ y $\{\mu_m\}_{m=1}^M$. Si la potencia transmitida resulta especialmente crítica para un usuario m determinado, basta con asignarle un μ_m más pequeño de manera que la solución óptima favorezca la selección de este usuario. Por otro lado, si es la tasa de transmisión la que resulta más importante para el usuario m , simplemente deberá elevar su exigencia de tasa \check{r}_m y de esta manera accederá en más ocasiones al canal incluso si su enlace presenta una peor ganancia. En suma, se comprueba de nuevo cómo la formulación propuesta para el problema hace que la asignación dictada por las proposiciones*

4.1 y 4.2, aunque oportunista para cada realización del canal, sea socialmente justa cuando se promedia en el tiempo.

Sub-problema 2: diseño “off-line” óptimo de la cuantificación de las potencias de transmisión

En este punto, se considera que las regiones de cuantificación \mathcal{R} son conocidas (bien porque estén pre-especificadas por el sistema $\mathcal{R} = \check{\mathcal{R}}$, bien porque se obtengan del anterior bloque de iteración del algoritmo de descenso coordinado como $\mathcal{R} = \mathcal{R}^{(t-1)}$) y que el esquema de acceso al medio $\mathbf{w}(\mathbf{g})$ está dado. El objetivo, pues, es resolver de forma “off-line” (4.13) c.r.a. \mathbf{p} de manera que se obtengan el valor óptimo para la constelación de las potencias de transmisión \mathbf{p}^* , cuando todas las otras variables han sido resueltas (objetivo que, desde el punto de vista del algoritmo de descenso coordinado, servirá para obtener el valor $\mathbf{p}^{(t)}$ para la siguiente iteración).

Puesto que los niveles de transmisión de potencia para el caso de Q-CSIT se optimizan independientemente de los modos de transmisión (que están acoplados con las regiones de cuantificación [cf. (sp3)]), no es necesario considerar en este problema las restricciones de tasa (i.e., la selección de los valores de potencia no influye explícitamente en C1 [cf. (4.13)]). Lo mismo ocurre para las restricciones de acceso al canal puesto que $\mathbf{w}(\mathbf{g})$ está dado (i.e., la selección de los valores de potencia no influye explícitamente en C3 [cf. (4.13)]). Por lo tanto, el problema de diseño óptimo de potencias de transmisión se reduce a

$$\begin{cases} \underset{\mathbf{p}}{\text{mín}} & \sum_{m=1}^M \mu_m \sum_{l=1}^{L_m} p_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) \\ \text{s.a. } C2. & \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon_{m,l}(g_m p_{m,l}) dF_{\mathbf{g}}(\mathbf{g}) / \check{r}_m \leq \check{\epsilon}_m, \quad m = 1, \dots, M. \end{cases} \quad (4.19)$$

Debido a que $\check{\epsilon} \geq 0 \forall m$ y $p_{m,l}$ puede incrementarse arbitrariamente para hacer $\epsilon_{m,l}(g_m p_{m,l})$ arbitrariamente pequeño, es posible satisfacer cualquier requisito de BER pre-establecido y, por consiguiente, el problema de minimización (4.19) tiene siempre solución. Añadido al hecho de que el problema es convexo (puesto que $\epsilon_{m,l}$

es una función convexa c.r.a. la potencia transmitida), supone que la solución de (4.19) será única, de carácter global y podrá obtenerse de forma eficiente.

Para resolver (4.19), se acude de nuevo a la función dual de Lagrange $\mathcal{L}(p_{m,l}, \chi_m^{Q_p})$ y se derivan las condiciones KKT (en este caso suficientes) c.r.a. $p_{m,l}^*$

$$\mu_m \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) + \chi_m^{Q_p^*} (r_{m,l} / \check{r}_m) \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon'_{m,l}(g_m p_{m,l}^*) dF_{\mathbf{g}}(\mathbf{g}) - \alpha_m^{Q_p^*} = 0 \quad (4.20)$$

donde: $\epsilon'_{m,l}$ denota la derivada de la función de BER c.r.a. $p_{m,l}$; $\chi_m^{Q_p^*} \geq 0$ denota el multiplicador óptimo asociado con la m -ésima restricción de BER y $\alpha_m^{Q_p^*} \geq 0$ denota el multiplicador asociado a la (m, l) -ésimo restricción implícita $p_{m,l} \geq 0$. Como ya se ha mencionado, puesto que (4.19) es convexo, la distancia del problema original c.r.a. problema dual se hace cero y las condiciones KKT son suficientes y necesarias para garantizar optimalidad global [13]. De forma interesante, cuando el conjunto $\{w_m(\mathbf{g})\}_{m=1}^M$ está dado, la optimización de la potencia a transmitir está desacoplada entre usuarios. Por lo tanto, resolver (4.19) equivale a resolver M pequeños problemas; i.e., $\forall m$, basta resolver: $\min_{p_m} \mu_m \sum_{l=1}^{L_m} p_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g})$ sujeto a $\sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon_{m,l}(g_m p_{m,l}) dF_{\mathbf{g}}(\mathbf{g}) / \check{r}_m \leq \check{\epsilon}_m$.

Se demuestra además en el apéndice 4.4.1 que: (i) en el óptimo $p_{m,l}^*$ se tiene $\chi_m^{Q_p^*} > 0 \forall m$, lo que implica que todas las restricciones de BER media se satisfacen como igualdades estrictas; (ii) si $\int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) > 0$, entonces $p_{m,l}^* > 0$ en cuyo caso $\alpha_m^{Q_p^*} = 0$ y (iii) si $\int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) = 0$, la potencia óptima de transmisión para la pareja (m, l) es $p_{m,l}^* = 0$, en cuyo caso $\alpha_m^{Q_p^*} > 0$. El caso (iii) puede tratarse simplemente a través de la eliminación de \mathcal{R} de los posibles modos AMC con potencia nula y reformulando (4.19) con un conjunto \mathcal{R} reducido en el que sólo se encuentren los modos AMC activos. Con esta reducción se garantiza que $\alpha_m^{Q_p^*} = 0 \forall m$, pero puesto que $\chi_m^{Q_p^*} > 0 \forall m$, (4.20) puede simplificarse y $p_{m,l}^*$ se obtiene como la raíz de la ecuación no lineal

$$\mathcal{L}'(p_{m,l}, \chi_m^{Q_p}) = \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon'_{m,l}(g_m p_{m,l}) dF_{\mathbf{g}}(\mathbf{g}) + \frac{\mu_m \check{r}_m}{r_{m,l} \chi_m^{Q_p}} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) = 0 \quad (4.21)$$

donde el segundo sumando es conocido si el valor del multiplicador $\chi_m^{Q_p} > 0$ también lo es. El valor óptimo del multiplicador $\chi_m^{Q_p^*}$ puede calcularse mediante la satisfacción ajustada de la restricción de BER; i.e., para cada m debe encontrarse la raíz de la ecuación no lineal

$$f(p_{m,l}(\chi_m^{Q_p})) := \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon_{m,l}[g_m p_{m,l}(\chi_m^{Q_p})] dF_{\mathbf{g}}(\mathbf{g}) / \check{r}_m - \check{\epsilon}_m = 0 \quad (4.22)$$

donde se ha escrito explícitamente $p_{m,l}(\chi_m^{Q_p})$ para remarcar la dependencia de $p_{m,l}$ c.r.a. multiplicador $\chi_m^{Q_p}$ en (4.21). Debido a que $\epsilon_{m,l}$ y $\epsilon'_{m,l}$ son funciones convexas y monótonas c.r.a. $p_{m,l}$ para cualquier $\chi_m^{Q_p}$, también lo son las funciones $\mathcal{L}'(p_{m,l}, \chi_m^{Q_p})$ y $f(p_{m,l}(\chi_m^{Q_p}))$. Esto implica que sus raíces pueden encontrarse de forma eficiente iterando búsquedas unidimensionales por sub-gradiente, cuya convergencia al par único $(p_{m,l}^*, \chi_m^{Q_p^*})$ está garantizada para cualquier inicialización arbitraria.

De forma más específica, considérese la iteración i -ésima, sean δ_p y δ_χ pequeños valores positivos y supóngase que $\chi_m(i)$ es conocido. Utilizando este valor en (4.21) en lugar de $\chi_m^{Q_p^*}$, puede iterarse la recursión

$$p_{m,l}(j_i + 1) = [p_{m,l}(j_i) + \delta_p \mathcal{L}'(p_{m,l}(j_i), \chi_m(i))]^+ \quad (4.23)$$

hasta que converja y fijar entonces $p_{m,l}(i)$ al valor de convergencia. En base a $p_{m,l}(i)$, se actualiza el valor del multiplicador de manera que (4.22) se satisfaga de forma ajustada como

$$\chi_m(i + 1) = [\chi_m(i) + \delta_\chi f(p_{m,l}(\chi_m(i)))]^+ \quad (4.24)$$

y retornar entonces a (4.23) con $i + 1$ para encontrar $p_{m,l}(i + 1)$ hasta que el proceso converja.

El criterio de parada para estas dos recursiones anidadas es semejante a los de apartados anteriores. También como se ha comentado con anterioridad, las integrales envueltas en (4.21) y (4.22) pueden ser reemplazadas en la práctica

por promedios sobre realizaciones numéricas del canal generadas a partir de la FDPA $F_{\mathbf{g}}(\mathbf{g})$ (nótese que puesto que este cálculo se realiza de forma “off-line” puede ser tan preciso como el sistema lo requiera). Téngase en cuenta también que, especialmente para el caso de la potencia óptima donde los distintos usuarios están desacoplados, para el cálculo de los valores óptimos de la solución, en lugar de realizar las actualizaciones por sub-gradiente en (4.23) y (4.24), pueden llevarse a cabo sencillas búsquedas uni-dimensionales, e.g. a través del método de bisección, que presentan una mayor velocidad de convergencia (geométrica para el caso de la bisección) sin tener que elegir un valor adecuado para los parámetros (pasos) δ_p y δ_χ .

Resumiendo los resultados derivados en este punto se tiene que:

Proposición 4.3 *Bajo (sp1)-(sp3), si \mathcal{R} y $\mathbf{w}(\mathbf{g})$ son conocidos, los valores correspondientes a la asignación óptima de potencia basada en Q-CSIT están dados por el mínimo global de (4.19). La potencia cuantificada óptima $p_{m,l}^*$ para cada pareja (m, l) es: bien $p_{m,l}^* = 0$ o bien la única raíz positiva de (4.21) que se calcula “off-line” de forma eficiente (junto con el valor óptimo del multiplicador de Lagrange asociado) utilizando dos búsquedas uni-dimensionales anidadas.*

Para cada usuario m , la proposición 4.3 dicta lo que podría entenderse como una asignación de potencia en base a un principio de “water-filling” que en lugar de considerar los valores de las ganancia tiene en cuenta los valores de los índices de las regiones de cuantificación $\{\mathcal{R}_{m,l}\}_{l=1}^{L_m}$. Una vez más, debido a la convexidad y monotonía de $\epsilon_{m,l}(g_m p_{m,l})$, la asignación óptima de potencia $p_{m,l}^*$: (i) crece según $r_{m,l}$ aumenta [cf. (4.21)] y (ii) decrece según $\check{\epsilon}_m$ aumenta [cf. (4.22)].

Sub-problema 3: diseño “off-line” del cuantificador de canal

En este sub-problema se consideran dados la asignación de usuarios al canal $\mathbf{w}(\mathbf{g})$ y los niveles cuantificados de potencia \mathbf{p} (calculados de acuerdo a las proposiciones 4.2 y 4.3) y el objetivo consiste en minimizar (4.13) c.r.a. \mathcal{R} . Evidentemente, este paso no es necesario cuando las regiones de cuantificación están

pre-especificadas (pre-diseñadas) como $\mathcal{R} = \check{\mathcal{R}}$. Por lo tanto, este sub-problema juega únicamente un papel en el algoritmo de descenso coordinado por bloques para calcular $\mathcal{R} = \mathcal{R}^{(t)}$ cuando $\mathbf{w}(\mathbf{g}) = \mathbf{w}^{(t-1)}(\mathbf{g})$ y $\mathbf{p} = \mathbf{p}^{(t-1)}$ son conocidos como parte del resultado de la iteración previa. Una vez alcanzada la convergencia del algoritmo de descenso, la solución (calculada “off-line”) de este sub-problema generará las regiones de cuantificación óptimas \mathcal{R}^* que se utilizarán durante la fase “on-line”.

Debido a que $\mathbf{w}(\mathbf{g})$ está dado, las restricciones de acceso al canal de (4.13) no aparecen en este problema. Por otro lado, puesto que (sp3) hace que los M problemas de cuantificación estén desacoplados entre usuarios, el problema de optimización correspondiente para cada usuario m es

$$\left\{ \begin{array}{l} \text{mín}_{\{\mathcal{R}_{m,l}\}_{l=1}^{L_m}} \mu_m \sum_{l=1}^{L_m} p_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) \\ \text{s. a } C1. \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) \geq \check{r}_m; \\ C2. \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon_{m,l}(g_m p_{m,l}) dF_{\mathbf{g}}(\mathbf{g}) / \check{r}_m \leq \check{\epsilon}_m. \end{array} \right. \quad (4.25)$$

Nótese que las regiones de cuantificación definen regiones de integración en (4.25). Puesto que (4.25) es generalmente no convexo c.r.a. estos límites de integración, al igual que en otros muchos problemas en los que diseña un cuantificador, garantizar un diseño óptimo de carácter global no resulta en general posible y la obtención de este diseño a través del uso de algoritmos eficientes es complicada.

Pese a ello como criterio de diseño se seguirá el criterio clásico de formulación del problema dual a través del Lagrangiano $\mathcal{L}(\mathcal{R}_{m,l}, \beta_m^{Q_{\mathcal{R}^*}}, \chi_m^{Q_{\mathcal{R}^*}}) = \sum_{l=1}^{L_m} \varphi_{m,l}^{Q_{\mathcal{R}}} (g_m) \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g})$, donde $\beta_m^{Q_{\mathcal{R}^*}}$ y $\chi_m^{Q_{\mathcal{R}^*}}$ denotan, respectivamente, los valores óptimos de los multiplicadores de Lagrange correspondientes a las restricciones de tasa y BER medias y la función *instantánea* de coste está dada por

$$\varphi_{m,l}^{Q_{\mathcal{R}}} (g_m) := \mu_m p_{m,l} - \beta_m^{Q_{\mathcal{R}^*}} r_{m,l} + \chi_m^{Q_{\mathcal{R}^*}} r_{m,l} \epsilon_{m,l}(g_m p_{m,l}) / \check{r}_m \quad \text{for } l = 1, \dots, L_m \quad (4.26)$$

con $\varphi_{m,0}^{Q_{\mathcal{R}}}(g_m) := 0$. Para encontrar las regiones que minimicen este coste, cada realización \mathbf{g} debe asignarse a $\mathcal{R}_{m,l}^*$ si y sólo si su m -ésima componente satisface $\varphi_{m,l}^{Q_{\mathcal{R}}}(g_m) \leq \varphi_{m,l'}^{Q_{\mathcal{R}}}(g_m) \forall l' \neq l, l' = 0, 1, \dots, L_m$. Esta condición genera de forma automática las regiones de cuantificación para el usuario m que optimizan (4.25) como

$$\mathcal{R}_{m,l}^* = \left\{ \mathbf{g} : \varphi_{m,l}^{Q_{\mathcal{R}}}(g_m) \leq \varphi_{m,l'}^{Q_{\mathcal{R}}}(g_m); \forall l' \neq l, l' \in \{0, 1, \dots, L_m\} \right\}. \quad (4.27)$$

Estas regiones pueden construirse tan pronto como los valores óptimos de los multiplicadores $(\beta_m^{Q_{\mathcal{R}^*}}, \chi_m^{Q_{\mathcal{R}^*}})$ involucrados en el coste sean conocidos. Para el caso no trivial en el que $\beta_m^{Q_{\mathcal{R}^*}} > 0$ y $\chi_m^{Q_{\mathcal{R}^*}} > 0$, las restricciones en el punto óptimo se satisfacen de forma ajustada (i.e., como una igualdad) y los multiplicadores de Lagrange óptimos pueden obtenerse como las soluciones de las siguientes ecuaciones no lineales⁶:

$$\mathcal{L}'_{\beta}(\beta_m^{Q_{\mathcal{R}}}, \chi_m^{Q_{\mathcal{R}}}) = \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}(\beta_m^{Q_{\mathcal{R}}}, \chi_m^{Q_{\mathcal{R}}})} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) - \check{r}_m = 0 \quad (4.28)$$

$$\mathcal{L}'_{\chi}(\beta_m^{Q_{\mathcal{R}}}, \chi_m^{Q_{\mathcal{R}}}) = \sum_{l=1}^{L_m} r_{m,l} \int_{\mathcal{R}_{m,l}(\beta_m^{Q_{\mathcal{R}}}, \chi_m^{Q_{\mathcal{R}}})} w_m(\mathbf{g}) \epsilon_{m,l}(g_m p_{m,l}) dF_{\mathbf{g}}(\mathbf{g}) / \check{r}_m - \check{\epsilon}_m = 0 \quad (4.29)$$

donde se ha indicado de forma explícita en los límites de integración la dependencia de las regiones de cuantificación de los valores de los multiplicadores de Lagrange. Para resolver estas ecuaciones podrían utilizarse búsquedas anidadas por sub-gradiente como se hizo en (4.23) y (4.24); sin embargo, puesto que en esta ocasión las funciones no lineales no son convexas, las búsquedas por sub-gradiente únicamente pueden garantizar la convergencia a un óptimo (posiblemente sin carácter único) local que además podría depender del punto inicial.

⁶Por simplicidad en la notación, para el cálculo de los valores de los multiplicadores se utilizará \mathcal{L}'_x para denotar la derivada parcial de \mathcal{L} c.r.a. x .

Por esta razón, resulta más adecuado en esta ocasión encontrar los valores óptimos de (4.28) y (4.29) mediante una búsqueda exhaustiva sobre el espacio bi-dimensional de $(\beta_m^{Q\mathcal{R}}, \chi_m^{Q\mathcal{R}})$ (para lo cual, para cada pareja de multiplicadores candidata, deberán evaluarse las integrales involucradas (4.28) y (4.29) de forma analítica o de forma numérica según el método propuesto para los casos previos). Aun cuando una búsqueda bi-dimensional no sea tan eficiente como una búsqueda por sub-gradiente o una búsqueda uni-dimensional, resulta todavía manejable puesto que se realiza de forma “off-line” y genera una solución óptima sin tener que preocuparse de la inicialización. Además, tan pronto como se encuentran multiplicadores para los que $|\mathcal{L}'_\beta| < \varepsilon$ y $|\mathcal{L}'_\chi| < \varepsilon$ para un nivel de tolerancia ε , la búsqueda se detiene. (Vale la pena destacar que, de forma inesperada, la implementación práctica de esta búsqueda concluye típicamente en un número reducido de iteraciones.) En este punto puede volverse la mirada atrás, hacia (sp3) y reflexionar sobre la importancia de desacoplar el diseño de las regiones de optimización entre usuarios. Si la optimización de $\mathcal{R}_{m,l}$ se hubiera realizado a lo largo de todos los pares (m, l) , para obtener los valores óptimos de los vectores conteniendo los multiplicadores de Lagrange se hubiera necesitado una búsqueda sobre un espacio $2M$ -dimensional, una tarea computacionalmente mucho más costosa que las M búsquedas bi-dimensionales obtenidas con el auspicio de (sp3).

La recapitulación de resultados básicos de este punto se hace en la siguiente proposición:

Proposición 4.4 *Bajo (sp1)-(sp3), si $\mathbf{w}(\mathbf{g})$ y \mathbf{p} son conocidos, las L_m regiones de cuantificación que optimizan (4.25) están dadas por (4.27) para cada usuario $m = 1, \dots, M$. El cómputo de los multiplicadores de Lagrange óptimos $(\beta_m^{Q\mathcal{R}^*}, \chi_m^{Q\mathcal{R}^*})$ requeridos para describir las regiones óptimas exige la resolución de (4.28) y (4.29) para cada usuario, lo que implica la realización “off-line” de M búsquedas exhaustivas bi-dimensionales.*

Pese a que en principio la expresión para las regiones era más genérica, las simulaciones numéricas muestran que en todos los casos analizados la forma de

las regiones puede describirse en base a intervalos independientes y consecutivos; i.e., en función de umbrales $\tau_{m,l}$ de forma que $\{\mathbf{g} \in \mathcal{R}_{m,l} \equiv g_m \in [\tau_{m,l}, \tau_{m,l+1})\}_{l=0}^{L_m}$. También es interesante el hecho de que la proposición 4.4 implemente el cuantificador del canal asignando cada valor de g_m a la región que minimiza el Lagrangiano de (4.25). Intuitivamente, $\beta_m^{Q_{\mathcal{R}^*}}$ puede interpretarse como la utilidad de cada bit transmitido, mientras que $\chi_m^{Q_{\mathcal{R}^*}}$ representaría el coste de cada bit erróneo recibido. Con esta interpretación, el cuantificador asigna g_m a la región $\mathcal{R}_{m,l}$ que presenta una menor contribución al coste (o agregado de costes) promedio cuyo valor para la realización actual representa $\varphi_{m,l}^{Q_{\mathcal{R}}}(g_m)$. Nótese también que si $\mathbf{g} \in \mathcal{R}_{m,0}^*$ el usuario m -ésimo debería renunciar al acceso al canal puesto que el coste de asignar g_m a una región activa ($l_m > 0$) es mayor que el coste de asignarlo a la región inactiva $l_m = 0$ que, evidentemente, incurre en un coste cero.

Asignación de recursos y cuantificación

A continuación se combinan los resultados derivados en los puntos anteriores (sub-problemas 1-3) para resolver el problema de asignación de recursos y cuantificación conjunta (ARCC) presentado en (4.13). El algoritmo ARCC propuesto utiliza una estrategia de descenso coordinado por bloques que minimiza la función objetivo global $J(\mathbf{w}(\mathbf{g}), \mathbf{p}, \mathcal{R}) := \sum_{m=1}^M \mu_m \sum_{l=1}^{L_m} p_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g})$. La idea básica consiste en dividir las variables en tres conjuntos diferentes $(\mathbf{w}(\mathbf{g}), \mathbf{p}, \mathcal{R})$, estos tres conjuntos originan tres problemas de optimización diferentes (con el mismo objetivo y restricciones que el original pero un número menor de variables de optimización). Cada uno de los problemas debe resolverse c.r.a. conjunto de variables que lo ha originado y suponiendo que los conjuntos de variables restantes están dados. Una vez que los tres problemas están resueltos (proposiciones 4.2, 4.3 y 4.4), tan sólo queda acoplar todas las soluciones a través de una estrategia iterativa de convergencia garantizada resumida a continuación:

Algoritmo 4.1 ARCC

- (AR0.) *Inicialización: Genere regiones iniciales de cuantificación $\mathcal{R}^{(0)}$ y vectores de potencia transmitida $\mathbf{p}^{(0)}$ que satisfagan las restricciones en (4.13) (ayúdese para ello de $F_{\mathbf{g}}(\mathbf{g})$). Seleccione un nivel de tolerancia $\varepsilon > 0$, inicialice el valor de la función objetivo como $J^{(0)} = \infty$ y fije el índice de iteración como $t = 1$.*
- (AR1.) *Dados $\mathcal{R}^{(t-1)}$ y $\mathbf{p}^{(t-1)}$, obtenga $\mathbf{w}^{(t)}(\mathbf{g})$ utilizando la proposición 4.2.*
- (AR2.) *Dados $\mathbf{w}^{(t-1)}(\mathbf{g})$ y $\mathcal{R}^{(t-1)}$, obtenga $\mathbf{p}^{(t)}$ utilizando la proposición 4.3.*
- (AR3.) *Dados $\mathbf{p}^{(t-1)}$ y $\mathbf{w}^{(t-1)}(\mathbf{g})$, obtenga $\mathcal{R}^{(t)}$ utilizando la proposición 4.4.*
- (AR4.) *Criterio de parada: Calcule el objetivo $J^{(t)}$ en base a $\mathcal{R}^{(t)}$, $\mathbf{p}^{(t)}$ y $\mathbf{w}^{(t)}(\mathbf{g})$. Si $|(J^{(t)} - J^{(t-1)})/J^{(t)}| < \varepsilon$, devuelva la (t) -ésima asignación de recursos y cuantificación y finalice. En caso contrario, incremente t en una unidad y vuelva a (A1).*
-

Para cada uno de los cuatro pasos (AR1)-(AR4) se garantiza que el valor del objetivo global J no aumenta, por lo que se comprueba fácilmente que el algoritmo ARCC tiene convergencia asegurada en un número finito de iteraciones al menos a un punto estacionario. Sin embargo, las proposiciones 4.2 y 4.4 muestran que en cada iteración encuentran analíticamente (aunque no siempre en forma cerrada) el mínimo global c.r.a. la variable que optimizan. Esto hace que cuando las regiones de cuantificación están impuestas por el diseño del sistema, la iteración coordinada de los pasos (AR1) y (AR2) converge al óptimo global. La única razón por la que el descenso coordinado por bloques del algoritmo ARCC puede no alcanzar el óptimo global⁷ se debe a que el problema en (4.13) es convexo c.r.a. $\mathbf{w}(\mathbf{g})$ y \mathbf{p} pero no lo es c.r.a. las regiones de cuantificación \mathcal{R} . Como ya se ha comentado en secciones anteriores, la falta de convexidad es un rasgo asociado a la gran mayoría de esquemas de cuantificación especialmente si la cuantificación es vectorial y,

⁷El descenso coordinado por bloques converge de forma garantizada al óptimo global si el problema inicial es convexo c.r.a. las variables implicadas [9].

así por ejemplo, el ampliamente utilizado –también en esta tesis, véase (3.63) y (3.64)– algoritmo de Lloyd [42] es otro cuantificador que comparte este rasgo⁸. Sin embargo, certificando en parte el uso extendido de estas y similares (Lloyd) técnicas de cuantificación, las simulaciones numéricas confirmarán que el consumo de potencia asociado al esquema de Q-CSIT basado en el algoritmo ARCC es muy cercano al asociado al esquema de P-CSIT y por ser este último una cota inferior para cualquier diseño basado en Q-CSIT, corrobora la hipótesis de la (casi-)optimalidad del esquema propuesto.

Un aspecto clave para poder garantizar la (casi-)optimalidad del diseño pero sobre todo para facilitar su velocidad de convergencia es la elección de las variables iniciales $\mathcal{R}^{(0)}$ y $\mathbf{p}^{(0)}$ necesarias para ejecutar por primera vez el paso (AR1). Para juzgar la importancia de este aspecto, téngase en cuenta que si el algoritmo ARCC se inicializa de forma arbitraria, no sólo la velocidad de convergencia puede verse afectada, sino que el algoritmo puede incluso ser incapaz de encontrar una solución factible que satisfaga las restricciones del problema. Reconocida pues la importancia de inicializar el algoritmo ARCC con un conjunto de variables que represente un punto factible de (4.13) (i.e., que, con independencia del consumo de potencia asociado, satisfaga las restricciones), el método propuesto consiste en utilizar algunas de las variables de la solución basada en P-CSIT para inicializar el algoritmo. Pese a que en primera instancia esto pudiera parecer extraño puesto que la solución del caso P-CSIT utiliza el conocimiento de la realización \mathbf{g} , que no está disponible para los usuarios en el caso de Q-CSIT, téngase en cuenta que la mayor parte de la solución dictada por la proposición 4.1 se calcula “off-line” y, por consiguiente, no se necesita conocer \mathbf{g} sino únicamente su FDPA.

Las consideraciones anteriores motivan el reemplazo del coste $\varphi_{m,l}^{Q\mathcal{R}}(g_m)$ en (4.27) por el coste basado en P-CSIT $\varphi_{m,l}^{P_w}(g_m, \beta_m^{P_w^*}) := \mu_m p_{m,l}(g_m) - \beta_m^{P_w^*} r_{m,l}$ en (4.6) e inicializar las regiones de cuantificación por para cada usuario m utili-

⁸En la sección siguiente se relacionará de forma analítica del algoritmo de Lloyd con los problemas de cuantificación con restricciones como los tratados en esta disertación.

zando

$$\mathcal{R}_{m,l}^{(0)} = \{ \mathbf{g} : \varphi_{m,l}^{P_w}(g_m, \beta_m^{P_w^*}) \leq \varphi_{m,l'}^{P_w}(g_m, \beta_m^{P_w^*}); \forall l' \neq l, l' = 0, 1, \dots, L_m \} . \quad (4.30)$$

Como se ha comentado con anterioridad, el cálculo de $\beta_m^{P_w^*}$ requiere únicamente conocimiento de la FDPA de la ganancia del canal; además, el otro parámetro que se necesita para la evaluación de $\varphi_{m,l}^{P_w}(g_m, \beta_m^{P_w^*})$, la potencia transmitida para cada para cada g_m , se calcula simplemente invirtiendo (4.3) de la forma $p_{m,l} = (1/g_m)\epsilon_m^{-1}(\check{\epsilon}_m, r_{m,l})$, donde el valor $r_{m,l}$ del modo AMC es conocido.

Por otro lado, para inicializar el vector de niveles de potencia transmitida, se toma un criterio conservador que garantice la sobre-satisfacción del requisito de BER calculando $p_{m,l}^{(0)} \forall m = 1, \dots, M, l = 1, \dots, L_m$ como

$$p_{m,l}^{(0)} = \max_{\mathbf{g} \in \mathcal{R}_{m,l}^{(0)}} (1/g_m)\epsilon_m^{-1}(\check{\epsilon}_m, r_{m,l}) . \quad (4.31)$$

Puesto que todas las cantidades de la solución basada en P-CSIT se han calculado para satisfacer el requisito de tasa de transmisión y la asignación en (4.31) claramente satisface el requisito de BER, la inicialización basada en (4.30) y (4.31) genera variables factibles.

Una vez clarificada la inicialización en el paso (AR0), esta sección se cerrará con una serie de consideraciones prácticas respecto a la operación “off-line” y “on-line” del algoritmo ARCC propuesto. Para problemas con estructura parecida a los involucrados en los pasos (AR1) y (AR2), los avances en optimización convexa han permitido que puedan resolverse de forma sencilla y rápida⁹. Por lo que el único impedimento a una rápida convergencia se encuentra en el paso (AR3), no obstante el esquema de cuantificación propuesto reduce la carga computacional de este paso a búsquedas bi-dimensionales, por lo que la complejidad del mismo está tam-

⁹De hecho, las herramientas de optimización desarrolladas permiten resolver problemas de una dimensionalidad mucho mayor que la esperada en un sistema TDMA (véase, e.g., [13, capítulo 1] donde se manejan problemas con más de 100 variables con una complejidad moderada y convergiendo en un número medio de entre 10 – 100 iteraciones –búsquedas por sub-gradiente–).

bien limitada. Esto hace que, como corroborarán las simulaciones numéricas, la complejidad computacional durante la fase “off-line” (sólo habría que ejecutar de nuevo el algoritmo cuando las estadísticas a largo plazo del canal variaran) no suponga un problema para el funcionamiento del sistema.

Respecto a la operación “on-line” (para cada realización del canal) del esquema, recuérdese que únicamente involucra a una parte de los cálculos asociados al paso (AR1). De esta manera, conocidos los valores de \mathcal{R}^* , \mathbf{p}^* , $\beta^{Q_w^*}$ y $\chi^{Q_w^*}$ (todos calculados “off-line”), el PA puede encontrar el modo y el usuario que transmitan de forma más eficiente a través del canal y enviar (realimentar) a los usuarios la palabra $\mathbf{c}(\mathbf{g})$ correspondiente, conteniendo únicamente $\lceil \log_2 M + \log_2(\max_m L_m) \rceil$ bits por cada realización del canal [cf. proposición 4.2]. Puesto que los usuarios conocen los niveles de potencia cuantificada \mathbf{p}^* , una vez que el PA designe al usuario ganador (excepto para el caso de $\mathbf{g} \simeq \mathbf{0}$), este podrá transmitir de acuerdo al índice indicado (potencia y modo AMC) durante todo el intervalo de transmisión para el que la realización del canal se mantenga vigente.

Observación 4.3 *A fin de evaluar los méritos prácticos del algoritmo ARCC considérese un ejemplo con $M = 8 - 16$ usuarios TDMA activos, cada uno de ellos implementando $L_m = 6$ modos AMC diferentes, tal y como se contempla en las especificaciones del sistema IEEE 802.16. De acuerdo al algoritmo presentado, durante la operación en tiempo real, el punto de acceso únicamente debe enviar a los usuarios 7 bits por cada realización. La palabra de realimentación se puede incluir en el mensaje UL-MAP que se encapsula en un paquete que se envía a los usuarios con información relativa a la siguiente trama de transmisión; véase [3, sección 6.2.7]. El funcionamiento de sistemas basados en Q-CSIT está también estandarizado en otras tecnologías, como por ejemplo, a través del canal de control de tasa de datos (DRC) en CDMA 2000 1xEV-DO o vía el canal indicador de calidad (CQI) en el modo HSUPA de UMTS [78]. En suma, las necesidades de tasa de realimentación por parte del algoritmo ARCC propuesto son ciertamente admisibles en la mayoría de los sistemas prácticos.*

4.1.4. Simulaciones numéricas

En esta sección se evaluarán las prestaciones del algoritmo ARCC en un escenario con un canal TDMA, desvanecimiento plano Rayleigh y dos usuarios ($M = 2$). Debe observarse que al tratarse TDMA de un esquema ortogonal, el número de usuarios no es, desde un punto de vista analítico, un parámetro determinante en el escenario que se plantea. Sin embargo, el paradigma de dos usuarios resulta de gran interés puesto que posibilita la representación gráfica de: (i) la región de potencia [87] que muestra las limitaciones en términos de consumo basada en Q-CSIT (para el algoritmo ARCC) y (ii) las regiones de cuantificación y la asignación de usuarios; ayudando así a la mejor comprensión de los resultados obtenidos.

El sistema se caracteriza además por un ancho de banda de BW y una densidad bilateral de potencia N_0 unitarias. Los coeficientes de desvanecimiento g_m , $m = 1, 2$, tienen media \bar{g}_m y se consideran completamente incorrelados. A menos que se indique lo contrario, se supondrá que los usuarios admiten tres modos de modulación de amplitud en cuadratura: 2-QAM, 8-QAM y 32-QAM; en consecuencia, las tasas AMC son: $r_{m,l} = 1, 3, 5$ bits por símbolo (bpsb). En el caso propuesto, la BER instantánea puede aproximarse por [cf. (3.9)]

$$\epsilon_{m,l}(\gamma) = 0.2e^{-\gamma/[1.5(2^{r_{m,l}}-1)]}. \quad (4.32)$$

con $\gamma = pg$. En todas las simulaciones realizadas, la BER media tiene un valor establecido de $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$.

Caso de estudio 1 (Comparación de consumo para los esquemas basados en P-CSIT y Q-CSIT): Suponiendo la utilización de P-CSIT o Q-CSIT, y $\bar{g}_m = 0$ dB para $m = 1, 2$, se probará el esquema de asignación de recursos basado en P-CSIT propuesto en el apartado 4.1.2 así como la implementación del algoritmo ARCC basada en Q-CSIT y propuesta en el apartado 4.1.3. Para evaluar la influencia de la cuantificación y adaptación óptima de potencia, se simulará también un esquema de cuantificación y adaptación heurístico basado en Q-CSIT, en el que se asignan idénticas fracciones de tiempo a los usuarios y en la que cada terminal

transmite con potencia fija independientemente de la tasa AMC empleada. En función de la potencia de transmisión específica de cada usuario, el punto de acceso selecciona el modo AMC de forma independiente para cada usuario de modo que la BER instantánea sea menor o igual que el umbral establecido. Esta cuantificación es claramente conservadora puesto que, salvo en el límite inferior de las regiones de cuantificación, la BER siempre es menor que el nivel máximo tolerado. Con esta cuantificación, se selecciona la potencia fija de transmisión de cada terminal de modo que se satisfaga la tasa media requerida. Dada la simplicidad del proceso descrito, el uso de la este esquema heurístico de adaptación y cuantificación está ampliamente extendido en sistemas reales con transmisiones adaptativas (e.g., CDMA2000 1xEVDO y HSUPA UMTS). En primer lugar se considerarán los requisitos de tasa media individual: $\check{r}_1 = 1 \text{ bpsb}$ y $\check{r}_2 = 1 \text{ bpsb}$. Para distintos niveles de prioridad entre usuarios, la parte superior de la figura 4.2 representa el consumo total de potencia ponderada para los cuatro esquemas; mientras que en la parte inferior de la figura 4.2, con el objetivo de medir el *precio* pagado al limitar la tasa de realimentación, se representa la pérdida de prestaciones de los dos sistemas TDMA basados en Q-CSIT c.r.a. sistema basado en P-CSIT. Puede observarse que: (i) el algoritmo ARCC supera al método heurístico basado en Q-CSIT (ahorrando hasta 6 dB) y (ii) la diferencia entre las soluciones ARCC y P-CSIT es muy pequeña. Puesto que la solución P-CSIT acota inferiormente al resto de las alternativas basadas en Q-CSIT, estos resultados certifican que el diseño de descenso coordinado por bloques propuesto en esta sección presenta unas prestaciones próximas a las del óptimo P-CSIT y, por consiguiente, a efectos prácticos puede representar un óptimo de facto para sistemas basados en Q-CSIT y modos AMC.

Caso de estudio 2 (Límites en el consumo – región de potencia): Tomando como base la región de capacidad derivada en estudios de teoría de la información, en [87] se define la región de potencia suponiendo P-CSIT y esquemas que alcanzan capacidad. Aquí se extenderá este concepto al esquema propuesto y se utilizará para ponderar de una forma más completa la diferencia de prestaciones entre los esque-

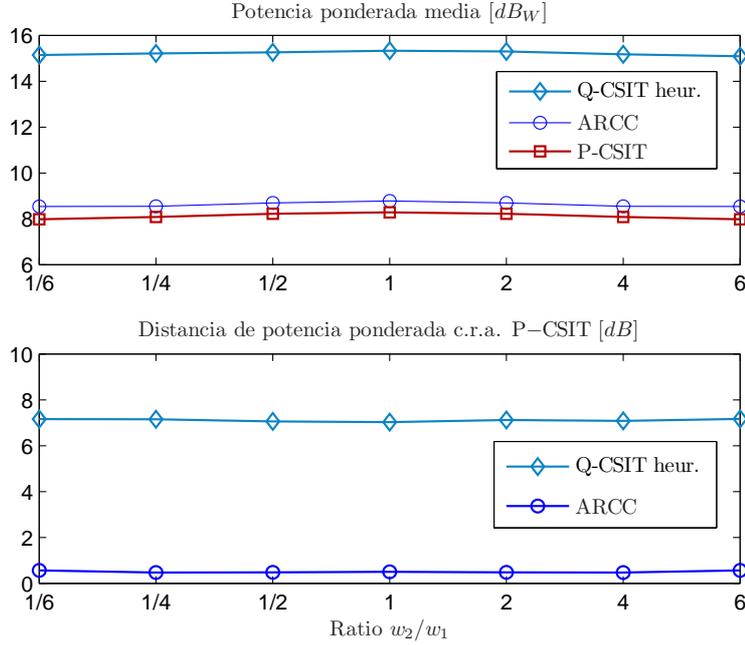


Figura 4.2: Consumo total de potencia para distintos esquemas de asignación de recursos variando la prioridad entre usuarios μ_2/μ_1 (con $\sum_{m=1}^2 \mu_m = 1$) cuando $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1$ bpsb, y $\bar{g}_1 = \bar{g}_2 = 0$ dB.

mas P-CSIT y Q-CSIT. De este modo, definiendo el vector de potencias medias de transmisión por usuario como $\bar{\mathbf{p}} := [\bar{p}_1, \dots, \bar{p}_M]^T$ y con $\tilde{\mathcal{F}}$ denotando el conjunto de todos los vectores $(\mathcal{R}, \mathbf{p}, \mathbf{w}(\mathbf{g}))$ que satisfacen las restricciones de tasa y BER (4.11) y (4.12), la región de potencia para el caso de Q-CSIT se define como

$$\bar{\mathcal{P}}(\check{\mathbf{r}}, \check{\epsilon}) = \bigcup_{(\mathcal{R}, \mathbf{p}, \boldsymbol{\tau}(\mathbf{g})) \in \tilde{\mathcal{F}}} \bar{\mathcal{P}}_{TD}(\mathcal{R}, \mathbf{p}, \boldsymbol{\tau}(\mathbf{g})), \quad (4.33)$$

donde la región de potencia media alcanzable para una tripleta específica de regiones, potencia cuantificada y acceso $(\mathcal{R}, \boldsymbol{\pi}, \boldsymbol{\tau}(\mathbf{g}))$ está dada por

$$\bar{\mathcal{P}}_{TD}(\mathcal{R}, \mathbf{p}, \mathbf{w}(\mathbf{g})) = \left\{ \bar{\mathbf{p}} : \bar{p}_m \geq \sum_{m=1}^{L_m} p_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}), \quad \forall m \right\}. \quad (4.34)$$

Nótese que dada una tripleta $(\mathcal{R}, \mathbf{p}, \boldsymbol{\tau}(\mathbf{g})) \in \tilde{\mathcal{F}}$ factible (i.e., que no viole restricciones del problema), los requisitos de tasa y BER (4.11) y (4.12) pueden satisfacerse utilizando la media de potencias de transmisión

$\sum_{l=1}^{L_m} p_{m,l} \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}), \forall m$. Claramente, si se utilizan potencias de transmisión $\bar{p}_m \geq \sum_{l=1}^{L_m} p_{m,l} \int_{\mathcal{R}_{m,l}} \tau_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}), \forall m$, se satisfacen (posiblemente con creces) los requisitos de tasa y BER y, por consiguiente, el vector de potencia así definido pertenecería a la región definida en (4.34). Los límites de la región en (4.33) (t.c.c. puntos de contorno o puntos de frontera) pueden obtenerse resolviendo (4.13) para todos los posibles valores $\boldsymbol{\mu} \geq \mathbf{0}$. En el caso de $M = 2$ usuarios, se han realizado experimentos para dos conjuntos diferentes de requisitos de tasa media: i) $\tilde{r}_1 = 1 \text{ bpsb}$, $\tilde{r}_2 = 1 \text{ bpsb}$, y ii) $\tilde{r}_1 = 1 \text{ bpsb}$, $\tilde{r}_2 = 0.5 \text{ bpsb}$. En la figura 4.3 se muestran las regiones de potencia de los canales TDMA con desvanecimiento Rayleigh obtenidas a través de la solución P-CSIT y de ARCC basada en Q-CSIT, donde los ejes de abscisas y ordenadas corresponden a la potencia de transmisión media para el primer y el segundo usuario (\bar{p}_1 y \bar{p}_2 respectivamente). Las “líneas” de la figura representan los límites de las regiones de potencia, mientras que las respectivas regiones son las áreas situadas en la zona superior derecha c.r.a. sus contornos asociados [cf. (4.34)]. Nótese que utilizando como potencias medias de transmisión (\bar{p}_1, \bar{p}_2) la coordenadas de un punto que se encuentre dentro de la región de potencia, siempre podrán satisfacerse los requisitos de tasa y BER utilizando un determinado esquema de acceso (posiblemente sin necesidad de ser óptimo). Claramente, cualquier solución óptima que corresponda a la minimización de la potencia ponderada constituirá un punto de la frontera de la región. Obsérvese asimismo que cuando los requisitos de tasa son idénticos, las regiones de potencia son simétricas c.r.a. la bisectriz $\bar{p}_1 = \bar{p}_2$ (regiones III y IV); simetría que desaparece para el caso $\tilde{r}_2 = \tilde{r}_1/2$. Como ya se ha comentado con anterioridad el esquema P-CSIT puede interpretarse como el caso límite del esquema Q-CSIT cuando la tasa de realimentación tiende a infinito (con más información disponible en los transmisores, estos pueden actuar de una manera más inteligente y, por consiguiente, utilizar de forma más eficiente los recursos). Esto justifica el hecho de que en la figura las regiones Q-CSIT estén siempre contenidas por las respectivas regiones P-CSIT (la región I contiene a II y la región III a la IV). Debe recordarse además que las regiones Q-CSIT de la figura 4.3 son en realidad

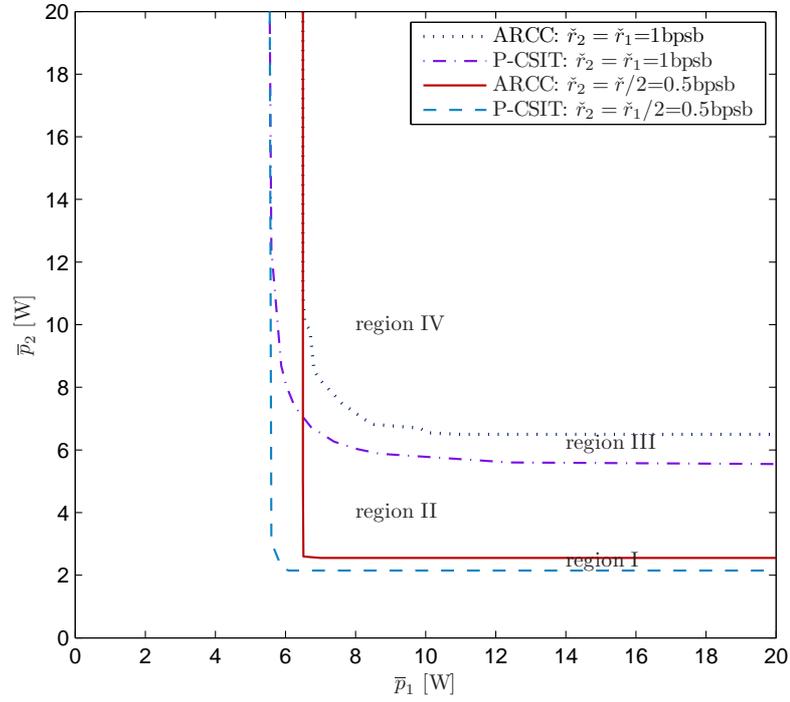


Figura 4.3: Regiones de potencia para los esquemas basados en P-CSIT y en Q-CSIT (ARCC) ($\tilde{\epsilon}_1 = \tilde{\epsilon}_2 = 10^{-3}$, $\bar{g}_1 = \bar{g}_2 = 0$ dB).

estimaciones conservadoras (puesto que no se ha podido demostrar la optimalidad global del algoritmo propuesto). En cualquier caso, se hace ver que las regiones Q-CSIT están muy próximas a las regiones P-CSIT, reafirmando así la eficiencia en consumo del algoritmo ARCC propuesto.

Por otro lado, en la tabla I se recogen resultados numéricos que describen el comportamiento del algoritmo ARCC en distintos casos. Desde el punto de vista de consumo total, se muestra que para diferentes escenarios de simulación el algoritmo de descenso coordinado por bloques propuesto para resolver el problema basado en Q-CSIT presenta siempre una eficiencia energética próxima a la óptima conseguida por el algoritmo basado en P-CSIT. Asimismo, los resultados testimonian que se satisfacen de manera ajustada todas las restricciones, que variando los valores del vector μ los usuarios con una mayor prioridad efectivamente se ven favorecidos y que el incremento del número de usuarios no supone una diferencia significativa en lo que al comportamiento general del algoritmo se refiere.

Tabla 4.1: Prestaciones del algoritmo ARCC en distintos casos de prueba (en los Casos II y IV se fijó una tasa objetivo de 0.5 *bpsb* para el usuario 2).

Caso	Usuario (m)	μ_m	$\bar{g}_m[dB]$	$\check{r}_m[bpsb]$	$\check{\epsilon}_m$	$\bar{p}_m[dBw]$	P-CSIT $\bar{P}_m[dBw]$
I	1	1	0	0.99	10^{-3}	8.75	8.21
	2	1	0	1.00	10^{-3}	8.80	8.21
II	1	1	0	1.00	10^{-3}	8.13	7.75
	2	1	0	0.50	10^{-3}	4.15	3.80
III	1	1	3	1.00	10^{-3}	6.60	6.03
	2	1	0	1.00	10^{-3}	8.58	8.03
IV	1	1	3	1.00	10^{-3}	6.64	5.93
	2	1	0	0.50	10^{-3}	3.52	3.31
V	1	4/3	0	0.99	10^{-3}	8.46	7.88
	2	2/3	0	0.99	10^{-3}	9.07	8.32
VI	1	4/3	0	1.00	10^{-3}	9.26	8.64
	2	4/3	0	0.49	10^{-3}	5.27	4.74
	3	2/3	0	0.99	10^{-3}	10.25	9.71
	4	2/3	0	0.50	10^{-3}	6.47	5.95

Caso de estudio 3 (Caracterización de la solución Q-CSIT): Para tener una mejor perspectiva, se estudiará con más detalle el algoritmo ARCC para $\check{r}_1 = \check{r}_2 = 1$ *bpsb* y $\mu_1/\mu_2 = 2$. En la tabla 4.2 se listan los valores cuantificados para las asignaciones de potencia y tasa, mientras que en la figura 4.4 se muestran las regiones de cuantificación y la asignación de usuarios donde los distintos sombreados junto con los símbolos “•” y “×” representan la región y el usuario seleccionado para acceder al canal. A partir de la tabla 4.2 se deduce que $p_{m,l_1} > p_{m,l_2} \forall l_1 > l_2$, lo cual indica que, conforme a lo anunciado tras presentar la proposición 4.3, comportamientos de primer orden presentes en la asignación de recursos del principio clásico de “water-filling” son compartidos con los esquemas de P-CSIT y Q-CSIT derivados en esta sección. A su vez, cuando el canal es más fiable (mayor ganancia media), pueden lograrse mayores tasas de transmisión con un menor coste de potencia. La figura 4.4 confirma que, al menos para el caso de estudio simulado, las regiones óptimas de cuantificación $\{\mathcal{R}_{m,l}^*\}_{l=1}^3$, $m = 1, 2$, en lo que a g_m se refiere, corresponden a intervalos consecutivos (no solapados) y, por lo tanto, pueden ser igualmente

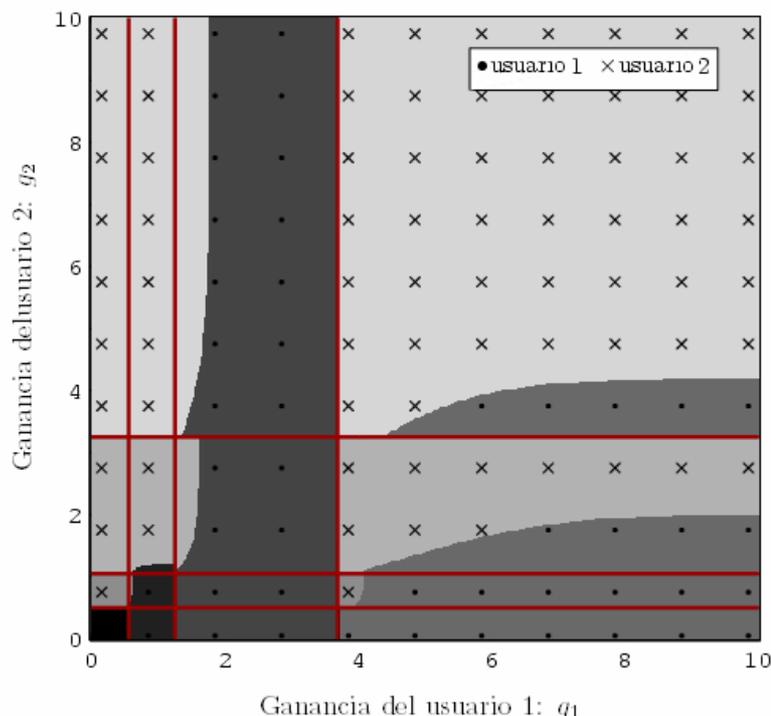


Figura 4.4: Asignación óptima de acceso y regiones de cuantificación obtenidas con el algoritmo ARCC. La selección del usuario que accede al canal se indica en la leyenda, las distintas regiones se representan con distintos sombreados y los umbrales de cuantificación señalan con líneas en negra ($\mu_1 = 2/3$, $\mu_2 = 1/3$, $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1$ bpsb, $\bar{g}_1 = \bar{g}_2 = 0$ dB).

descritas por un conjunto de umbrales $\{\tau_{m,l}^*\}$ señalados en la figura con líneas sólidas. Esto implica que una cuantificación más sencilla basada en umbrales de cuantificación puede dar lugar a prestaciones muy cercanas (si no equivalentes) a las aquí obtenidas.

Caso de estudio 4 (Variación de la tasa de realimentación): Se ha visto que con tres modos AMC, el algoritmo ARCC proporciona una eficiencia energética próxima a la de la solución óptima para el esquema P-CSIT. Para lograrlo, se necesitan $\lceil \log_2(2) + \log_2(4) \rceil = 3$ bits de Q-CSIT por intervalo de transmisión¹⁰. A continuación se muestra cómo afecta el número de bits de realimentación al rendimiento de ARCC. Para $\check{r}_1 = \check{r}_2 = 1$ bpsb y $\mu_1/\mu_2 = 1$, en la tabla 4.3 se recoge el coste total de potencia de transmisión media para un canal TDMA con desvanecimiento plano Rayleigh para un número variable de bits de realimentación. En el caso de

¹⁰Como se verá en la sección siguiente, este número puede reducirse ligeramente.

Tabla 4.2: Asignación de potencia y tasa transmitidas ($p_{m,l}$ y $r_{m,l}$) por cada estado de cuantificación resultantes de la aplicación del algoritmo ARCC ($\mu_1 = 2/3$, $\mu_2 = 1/3$, $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1$ bpsb, $\bar{g}_1 = \bar{g}_2 = 0$ dB).

	Usuario 1			Usuario 2		
Región de cuantificación	$\mathcal{R}_{1,1}$	$\mathcal{R}_{1,2}$	$\mathcal{R}_{1,3}$	$\mathcal{R}_{2,1}$	$\mathcal{R}_{2,2}$	$\mathcal{R}_{2,3}$
Potencia de transmisión $p_{m,l}$ [dBw]	8.56	13.23	15.60	8.99	13.84	16.29
Tasa $r_{m,l}$ [bits/sym]	1	3	5	1	3	5

Tabla 4.3: Potencia ponderada media consumida con ARCC variando el número de bits de realimentación ($\mu_1 = \mu_2 = 1$, $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1$ bpsb, $\bar{g}_1 = \bar{g}_2 = 0$ dB).

Algoritmo	ARCC	ARCC	ARCC	ARCC	P-CSIT
# de bits	1	2	3	4	∞
Potencia Media [dBw]	23.05	11.98	8.52	8.43	8.10

un solo bit, la información retroalimentada sólo indica la selección de usuario; una vez seleccionado el terminal, este transmite con independencia del valor de \mathbf{g} . En la figura 4.4 se ilustra que la región en las dos componentes de \mathbf{g} presentan un desvanecimiento profundo, ambos usuarios deben aplazar sus transmisiones. Aunque la región en este caso sea pequeña, se necesita mucha potencia (23.05 dBw) para compensar estos “malos” canales. Al aumentar el número de bits realimentados, también aumenta el número de modos AMC activos para cada usuario. Como se puede ver en la tabla 4.3, sorprende el hecho de que con sólo dos bits de realimentación (un usuario que dispone de dos modos activos y otro de sólo uno), el algoritmo ARCC proporciona un coste de potencia medio no muy distinto de la solución P-CSIT. Esto revela que el esquema de asignación de usuario al canal (i.e., de acceso al medio) juega un papel principal para la eficiencia del consumo potencia. Los resultados numéricos revelan también que unos pocos (2-4) modos AMC por usuario, y por lo tanto pocos bits de realimentación, bastan para cercar la distancia entre Q-CSIT y P-CSIT.

Caso de estudio 5 (Convergencia del algoritmo ARCC): En la figura 4.5 se

muestra la convergencia de ARCC, con la potencia ponderada total media (coste global) variando según varía el índice de iteración referido a cada uno de los pasos internos del algoritmo ARCC (se recuerda que para cada iteración-bloque (t), el algoritmo ARCC implementa tres pasos internos: ($AR1$), ($AR2$) y ($AR3$), de esta manera el primer punto corresponderá a al paso ($AR1$) para $t = 1$ y, e.g., el quinto a ($AR2$) para $t = 2$). Se observa que el algoritmo ARCC converge tras un número pequeño de iteraciones ($t = 5$ iteraciones bloque que suponen $(5 \times 3) = 15$ pasos internos para un total de alrededor de 4000 iteraciones). Las pequeñas variaciones que se producen a lo largo de la curva se deben a la resolución finita de las integraciones numéricas involucradas. Otra observación interesante es que incluso el primer paso de iteración interno devuelve una solución aceptable, lo que indica la calidad de la inicialización¹¹ en el paso ($AR0$). A pesar de que el algoritmo ARCC se ejecuta “off-line” basado en el comportamiento a largo plazo del canal (i.e., si este no cambia no hay que volver a ejecutar el algoritmo), la figura 4.5 demuestra además que la convergencia del mismo se produce en un pequeño número de iteraciones, por lo que la complejidad computacional asociada es limitada.

4.1.5. Síntesis de resultados

En esta sección se ha propuesto un marco teórico para minimizar la potencia de transmisión media ponderada sujeta a restricciones de tasa media y BER en un sistema TDMA, donde los transmisores disponen de un conjunto de modos AMC y ajustan su configuración de transmisión según la realimentación -de tasa limitada- que reciben por parte del punto de acceso. Cuando se dispone de P-CSIT, los procedimientos de selección y asignación de recursos han resultado ser oportunistas (a lo sumo un usuario transmite por trama), característica compartida con sistemas de acceso ortogonal que utilizan códigos de longitud infinita alcanzando capacidad. La incorporación de restricciones de QoS de carácter promedio (en lugar de

¹¹Aunque no presentados en esta disertación, algoritmos que aprovechan esta característica han sido propuestos en [59].

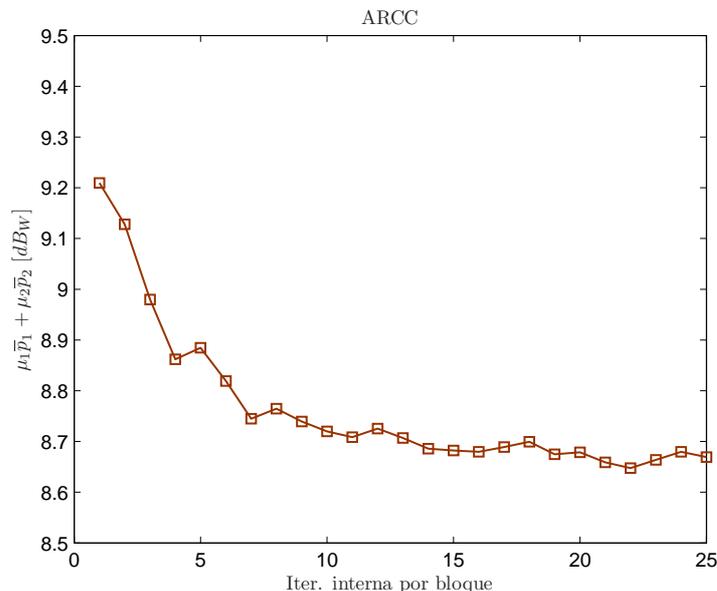


Figura 4.5: Evolución de la potencia ponderada media del algoritmo ARCC($\mu_1 = 2/3$, $\mu_2 = 1/3$, $\check{\epsilon}_1 = \check{\epsilon}_2 = 10^{-3}$, $\check{r}_1 = \check{r}_2 = 1$ bpsb, $\bar{g}_1 = \bar{g}_2 = 0$ dB).

instantáneo) en la formulación del problema P-CSIT ha permitido la obtención “off-line” de eficientes algoritmos que convergen a la solución óptima global, presentando unas necesidades mínimas de tasa de realimentación y una complejidad “on-line” despreciable.

Interesante es asimismo el hecho de que, cuando se derivan los esquemas de acceso al medio para el caso de sistemas basados en Q-CSIT, la política de acceso sea también oportunista. Para el caso en el que las regiones de cuantificación estén preespecificadas por el diseño del sistema, se han formulado y resuelto analíticamente problemas de optimización convexa que son globalmente óptimos. Para el caso en el que las regiones no están diseñadas a priori se ha propuesto un algoritmo de asignación de recursos y cuantificación conjunta (ARCC) que de forma simultánea resuelve la cuantificación del canal y la adaptación al CSIT. El diseño propuesto ha acoplado las regiones de cuantificación con los modos de transmisión de forma individual para cada usuario, lo que ha contribuido a reducir la complejidad y la necesidad de realimentación proveniente del punto de acceso.

El problema de ARCC ha sido resuelto mediante el uso de una estrategia de descenso iterativo coordinado por bloques que asegura la convergencia al menos a

un óptimo local (recuérdese que esto es lo mejor que se puede garantizar cuando se trata con problemas no convexos de cuantificación). Necesitando únicamente recursiones en base a métodos de gradiente y búsquedas bi-dimensionales, todas ellas realizadas “off-line”, el algoritmo propuesto tiene una complejidad abordable. Más relevante es el hecho de que la asignación “on-line” (oportunista) de recursos requiera un número extremadamente pequeño de bits por cada palabra de control (del orden de 4 – 8 por cada realización del canal para sistemas TDMA con 32 – 64 usuarios, cada uno de ellos implementando 4 – 8 modos AMC).

Finalmente se han realizado simulaciones para comparar la selección de usuarios, la asignación de recursos, y los esquemas ARCC basados en Q-CSIT que se han propuesto en esta sección, frente a una alternativa heurística y frente al nivel de referencia marcado por el esquema P-CSIT. Los resultados de las pruebas confirmaron lo anunciado por las soluciones analíticas y demostraron un ahorro considerable en la potencia transmitida cuando se comparaban con métodos sub-óptimos existentes. Los resultados también sugieren que el algoritmo ARCC basado en Q-CSIT tiene un gran potencial de desarrollo práctico, puesto que se adapta a las especificaciones de los estándares de acceso actuales y con una fase de operación “on-line” de complejidad muy reducida puede alcanzar prestaciones sorprendentemente cercanas a las del esquema óptimo basado en P-CSIT.

4.2. Minimización de potencia en sistemas OFDMA con QoS utilizando CSIT

Al igual que en el capítulo anterior, tras haber analizado la cuantificación y adaptación de los transmisores en canales selectivos en el tiempo y planos en frecuencia, en esta sección se abordará el diseño óptimo de la cuantificación de canal y de los esquemas de adaptación de los transmisores cuando múltiples usuarios quieren comunicarse con un punto de acceso a través de un canal *doblemente selectivo*. Como ya se mencionó en la sección 3.2, OFDM es una tecnología de transmisión

idónea cuando el canal presenta selectividad en frecuencia, puesto que de manera sencilla y eficiente transforma el canal selectivo en un conjunto de canales planos ortogonales. Asimismo, cuando se trabaja en escenarios con múltiples usuarios, OFDM presenta la ventaja añadida de dividir el canal inicial en distintos canales independientes, facilitando así la compartición (acceso) del mismo entre distintos usuarios. Cuando OFDM se utiliza no sólo como modulación de transmisión sino también como método de acceso para que puedan transmitir distintos usuarios se habla de sistemas OFDMA.

De esta manera, de acuerdo a lo indicado en el título de la sección, el objetivo aquí perseguido consiste en minimizar la potencia transmitida por un sistema OFDMA cuanto existen requisitos individuales de tasa de transmisión y BER media.

La versatilidad de OFDM unida al gran despliegue de las redes de acceso múltiple ha provocado que OFDMA sea una de las tecnologías con mayor implantación y, por consiguiente, que más atención investigadora ha recibido en los últimos tiempos. La asignación de recursos en OFDMA minimizando la potencia y satisfaciendo requisitos de tasa y BER por símbolo suponiendo P-CSIT fue estudiada por primera vez en [91]. La metodología de diseño propuesta por sus autores ha dado origen a la publicación de variaciones del algoritmo de [91] que, utilizando P-CSIT, se aplican tanto a sistemas limitados en tasa (las más) como a sistemas limitados en potencia. Asimismo, durante los últimos dos años se han venido realizando considerables esfuerzos en la obtención de esquemas de adaptación basados en P-CSIT que apliquen a escenarios de comunicación OFDM donde los usuarios que comparten el medio no pretenden comunicarse con un punto de acceso (comunicaciones jerárquicas) sino que quieren comunicarse arbitrariamente entre cualesquiera de ellos (comunicaciones distribuidas). Prometedores resultados se han obtenido en [75, 82], donde la aplicación de algoritmos de “water-filling” iterativo permiten alcanzar soluciones de carácter óptimo. La adaptación en base a Q-CSIT en sistemas OFDM tampoco ha sido ajena al esfuerzo investigador. La optimización de tasa o de potencia por cada símbolo en base a estructuras fijas de cuantificación ha sido analizada en [14], [37]. No obstante, los trabajos existentes

utilizan criterios de diseño que, forzando *a priori* esquemas de cuantificación o de adaptación aparentemente sencillos, sacrifican la optimalidad de la solución obtenida. A diferencia de estos trabajos, en esta sección se adapta de forma conjunta potencia, tasa y asignación de sub-portadoras en base a Q-CSIT para minimizar la potencia media transmitida satisfaciendo requisitos de tasa y BER media. Se realizará un especial esfuerzo en plantear problemas de optimización que, *a posteriori*, generen esquemas de asignación de recursos con complejidad computacional despreciable. Por último, pero no menos importante, el diseño aquí propuesto no considera fijas las regiones de cuantificación, optimizando también el cuantificador de canal.

El resto de la sección se organiza como sigue. Tras introducir el modelo y las suposiciones de trabajo básicas, se derivan esquemas de adaptación al Q-CSIT de carácter óptimo y sub-óptimo. Una vez caracterizados estos esquemas se procede primero al diseño óptimo del cuantificador de canal, prestando a continuación especial interés a un sencillo esquema de cuantificación que con una reducida complejidad, presenta resultados semejantes al óptimo. Las simulaciones numéricas comparando las prestaciones de los esquemas derivados y validando el análisis teórico realizado junto con la recopilación de resultados cierran esta sección.

4.2.1. Modelo y preámbulos

Se considera un sistema OFDMA como el representado en la figura 4.6) donde M usuarios, indexados mediante $m \in \{1, \dots, M\}$, compartiendo K sub-portadoras (canales), indexadas mediante $k \in \{1, \dots, K\}$ intentan comunicarse de forma individual con un punto de acceso (PA). La potencia y tasa *instantáneas* (por símbolo) que el usuario m transmite en la sub-portadora k se denotan como $p_{k,m}$ y $r_{k,m}$. Con estos valores como componentes, se forman las matrices \mathbf{P} y \mathbf{R} de tamaño $K \times M$ que contienen la potencia y tasa transmitida por cualquier usuario en cualquier sub-portadora de manera que si $[\cdot]_{k,m}$ denota el (k, m) -ésimo elemento de una matriz, entonces $[\mathbf{P}]_{k,m} := p_{k,m}$ y $[\mathbf{R}]_{k,m} := r_{k,m}$. Durante el tiempo de

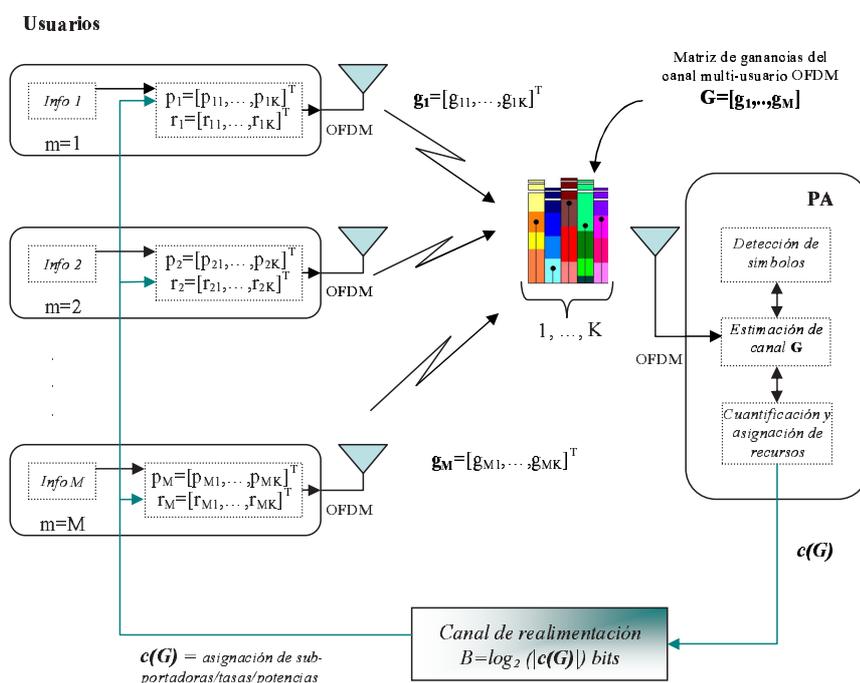


Figura 4.6: Modelo del sistema OFDMA.

duración de una realización del canal se considera un mecanismo de compartición del acceso a cada sub-portadora mediante la distribución del tiempo de la vigencia de la realización¹². Este método de compartición temporal de sub-portadoras se describe mediante la matriz de acceso \mathbf{W} de tamaño $K \times M$ cuyo (k, m) -ésimo elemento $w_{k,m}$ representa la fracción temporal durante la cual el m -ésimo usuario utiliza la k -ésima sub-portadora. Claramente, al igual que en la sección anterior: $\mathbf{W} \geq \mathbf{0}$ y $\sum_{m=1}^M w_{k,m} \leq 1, \forall k$, mientras que la potencia y tasa efectiva transmitidas durante el tiempo de coherencia del canal (o de forma equivalente, entre dos actualizaciones consecutivas del CSIT) por el usuario m en la sub-portadora k -ésima del usuario m son respectivamente $p_{k,m}w_{k,m}$ y $r_{k,m}w_{k,m}$.

La respuesta discreta equivalente en banda base del canal de cada usuario es

¹²La compartición de la sub-portadora puede realizarse a través de cualquier mecanismo de carácter ortogonal. La selección específica de un esquema de compartición en base a división en el tiempo se debe a la familiarización del lector con la variable w presentada en el apartado 4.1.1 y porque, en sintonía con lo establecido en la proposición 4.2, se comprobará más adelante que la asignación óptima corresponde a un esquema oportunista en la que un único usuario accede a la totalidad de la sub-portadora y, por consiguiente, no existe necesidad de que se utilice.

$\mathbf{h}_m := [h_{m,0}, \dots, h_{m,Q_m}]^T$, donde [cf. (2.10)]: $Q_m := \lfloor D_{m,\text{máx}}/T_s \rfloor$ denota el orden de canal (número de caminos adicionales), $D_{m,\text{máx}}$ el retardo de dispersión máximo, T_s el tiempo de muestreo y $Q_{\text{máx}} := \max_{m \in \{1, M\}} Q_m$ (recuérdese que en OFDM típicamente $K \gg Q_{\text{máx}}$). Para obtener una formulación más compacta, las M respuestas al impulso vectoriales se apilarán en la matriz $\mathbf{H} := [\mathbf{h}_1, \dots, \mathbf{h}_M]$ de tamaño $K \times M$, donde la longitud de cada columna se ha incrementado hasta alcanzar un valor K mediante la inserción del número correspondiente de ceros.

Con referencia al caso de un único usuario (véase la figura 3.4), cada transmisor toma un bloque con K símbolos a transmitir y tras colocarlos en un vector columna los multiplica por la inversa de la matriz de la DFT de tamaño $K \times K$ ($\mathbf{F}_K^{\text{mathcal{H}}}$) para insertar a continuación de un prefijo cíclico de Q muestras y producir así un bloque de $K + Q$ símbolos discretos (t.c.c. símbolo OFDM) que son convertidos en símbolos analógicos y enviados secuencialmente para su transmisión¹³. Estas operaciones junto con su procesado inverso en el receptor (eliminación del CP, cálculo de la DFT, etc.) convierten al canal selectivo en frecuencia de cada uno de los usuarios en un conjunto de K canales planos, cada uno de ellos con coeficiente de desvanecimiento dado por la respuesta en frecuencia del canal del usuario evaluado en la sub-portadora en cuestión. De esta manera, en base a la matriz multi-usuario de canal (temporal) \mathbf{H} podrá definirse entonces la matriz multi-usuario de canal (frecuencial) $\tilde{\mathbf{H}} := (1/\sqrt{K})\mathbf{F}_K\mathbf{H}$ de tamaño $K \times M$, cuya columna m -ésima contendrá los coeficientes de canal correspondientes a las K sub-portadoras del usuario m .

A través de la matriz multi-usuario de canal $\tilde{\mathbf{H}}$ (estimada a través de pilotos), el receptor conoce la matriz de ganancia (en términos de potencia) \mathbf{G} , donde $[\mathbf{G}]_{k,m} := |[\tilde{\mathbf{H}}]_{k,m}|^2/\sigma_{k,m}^2$, con $\sigma_{k,m}^2$ denotando la varianza del ruido AWGN en el receptor que, s.p.d.g., se supone unitaria. Se utilizará $g_{k,m} := [\mathbf{G}]_{k,m}$ para denotar el valor instantáneo de la ganancia de la k -ésima sub-portadora del m -ésimo usuario; de manera análoga, definiendo $\bar{\mathbf{G}} := \mathbb{E}_{\mathbf{G}}[\mathbf{G}]$, $\bar{g}_{k,m} := [\bar{\mathbf{G}}]_{k,m}$ denotará la ganancia

¹³Nótese que si el usuario no accede a todas las sub-portadoras –y por consiguiente transmite un menor número de símbolos– la simple puesta a cero de filas de la matriz $\mathbf{F}_K^{\text{mathcal{H}}}$ hace que los puntos del símbolo OFDM asociados a las sub-portadoras no utilizadas sean nulos.

media de la k -ésima sub-portadora del m -ésimo usuario. En base al conocimiento de la realización \mathbf{G} (P-CSI), el PA encontrará el acceso óptimo de los usuarios a las distintas sub-portadoras y asignará las cargas adecuadas de tasa y potencia para cada usuario en función de las regiones a las que los elementos de \mathbf{G} pertenecen. Utilizando los índices de estas regiones, el PA realimenta a los usuarios el vector de bits $\mathbf{c} = \mathbf{c}(\mathbf{G})$ (Q-CSI) en base al cual los usuarios adaptarán sus modos (configuración) de transmisión (i.e., potencia, tasa y sub-portadoras) de acuerdo a un conjunto finito de configuraciones.

Los *objetivos* principales para el sistema analizado son: (OB.1) diseñar un cuantificador de canal que genere \mathbf{c} y (OB.2) dado \mathbf{c} , encontrar matrices de transmisión \mathbf{P} , \mathbf{R} , y \mathbf{W} que, satisfaciendo los requisitos promedio de QoS, minimicen la potencia transmitida. Al igual que para el caso de TDMA, la solución simultánea del problema es difícil de manejar en términos computacionales. Por esta razón el diseño del sistema se abordará en dos etapas. En primer lugar se fijará (se considerará dado) el diseño del cuantificador de canal y en función de su forma y la del Q-CSIT que este genera, se diseñarán modos (configuraciones) de transmisión y esquemas de adaptación para las matrices \mathbf{P} , \mathbf{R} , y \mathbf{W} de manera que la potencia media ponderada \bar{P} se minimiza bajo requisitos promedios de tasa de transmisión de bit, $\check{\mathbf{r}} := [\check{r}_1, \dots, \check{r}_M]^T$, y BER, $\check{\boldsymbol{\epsilon}} := [\check{\epsilon}_1, \dots, \check{\epsilon}_M]^T$, independientes para cada usuario. Al igual que en la sección anterior, la ponderación se deberá a la existencia de distintos coeficientes de prioridad entre usuarios $\boldsymbol{\mu} := [\mu_1, \dots, \mu_M]^T$ en base a los cuales se calcula $\bar{P} := \sum_{m=1}^M \sum_{k=1}^K \mathbb{E}_{\mathbf{G}} [p_{k,m}(\mathbf{G}) w_{k,m}(\mathbf{G})]$. En la segunda etapa, se diseñará cuidadosamente el cuantificador de canal de manera que se obtenga un compromiso entre la complejidad de implementación, la tasa de realimentación exigida y la eficiencia de prestaciones. Estas tareas se realizarán bajo la cobertura de las siguientes suposiciones:

(sp1) *Los canales de los distintos usuarios están incorrelados; i.e., no existe correlación entre las columnas de \mathbf{G} .*

(sp2) *Las sub-portadoras de un mismo usuario pueden estar correladas, y se suponen distribuidas de acuerdo a una gaussiana compleja; i.e., $g_{k,m}$ sigue una dis-*

tribución exponencial $f_{g_{k,m}}(g_{k,m}) = (1/\bar{g}_{k,m})\exp(-g_{k,m}/\bar{g}_{k,m})$.

(sp3) El estado de las sub-portadoras (región de cuantificación a la que pertenecen) se mantiene constante durante al menos dos símbolos OFDM.

(sp4) El canal de realimentación está libre de errores y presenta un retardo despreciable.

(sp5) Los símbolos transmitidos pertenecen a modulaciones cuya BER instantánea puede aproximarse por

$$\epsilon(p_{k,m}, g_{k,m}, r_{k,m}) \simeq \kappa_1 \exp(-p_{k,m} \kappa_2 g_{k,m} / (2^{r_{k,m}} - 1)). \quad (4.35)$$

Puesto que los usuarios están en general suficientemente separados (sp1) suele ser cierta para la práctica totalidad de los sistemas; como ya se ha comentado (sp2) corresponde un modelo de desvanecimientos con amplitudes siguiendo distribuciones Rayleigh, caso típico de comunicaciones con ausencia de rayo directo (generalizaciones a otras distribuciones de canal también son posibles), (sp3) y (sp4) son suposiciones (condiciones de operación) típicas para el trabajo con sistemas Q-CSIT, semejantes a las realizadas para resolver problemas previos y (sp5) se ha supuesto teniendo en cuenta la versatilidad mostrada por parte de (4.35) en secciones anteriores así como su extendido uso en el diseño de sistemas OFDM. Nótese que a diferencia del caso de asignación TDMA analizado en la sección anterior, en donde el marco teórico propuesto no padecía las limitaciones impuestas por (sp1), (sp2) y (sp5), aquí se restringe la generalidad del modelo. No obstante, como se verá más adelante, estas tres suposiciones contribuirán a una reducción significativa de la complejidad del sistema (especialmente la asociada a los cómputos “off-line”, aspecto ciertamente relevante en sistemas OFDMA que *grosso modo* incrementan en un factor K los cómputos necesarios con respecto a los de un sistema TDMA), sin, por otro lado, suponer una pérdida excesiva de generalidad, puesto todas ellas son típicamente válidas en escenarios prácticos de aplicación.

4.2.2. Diseño de los modos de transmisión

Para el diseño de los modos de transmisión se supondrá además que: **(sp6)** *Para cada realización del canal, la ganancia $g_{k,m}$ pertenece a una de las $L_{k,m}$ regiones disjuntas $\{\mathcal{R}_{k,m|l}\}_{l=1}^{L_{k,m}}$.*

De acuerdo a lo anunciado, en este apartado se considerará que el conjunto $\{\mathcal{R}_{k,m|l}\}_{l=1}^{L_{k,m}}$ es fijo y conocido. La selección de $\{\mathcal{R}_{k,m|l}\}_{l=1}^{L_{k,m}}$ será parte del problema asociado al diseño del cuantificador que se abordará en el apartado 4.2.3. De acuerdo a lo que ocurría en casos anteriores tanto el diseño del cuantificador de canal como gran parte del diseño de los modos de transmisión requerirá únicamente de cómputos “off-line” y se derivará de nuevo un simple algoritmo “on-line” que de forma óptima resolverá la adaptación de los transmisores a cada realización del canal mediante la asignación de las matrices \mathbf{P} , \mathbf{R} y \mathbf{W} .

Formulación del problema

Dado (sp6), sea $\mathcal{R}_{k,m|l} := \{\mathbf{G} : g_{k,m} \in \mathcal{R}_{k,m|l}\}$ el conjunto de matrices \mathbf{G} para el que $g_{k,m}$ pertenece a la región $\mathcal{R}_{k,m|l}$. Asimismo, sean¹⁴ $p_{k,m|l}$ y $r_{k,m|l}$ la potencia y tasa que el usuario m transmitirá de forma *instantánea* en la sub-portadora k cuando $\mathbf{G} \in \mathcal{R}_{k,m|l}$. Recuérdese que $w_{k,m}(\mathbf{G}) \leq 1$ y, por lo tanto, el valor *efectivo* (o promedio) potencia y tasa que el usuario m transmitirá durante el tiempo de duración de la realización \mathbf{G} será $p_{k,m|l}w_{k,m}(\mathbf{G})$ y $r_{k,m|l}w_{k,m}(\mathbf{G})$.

El propósito del diseño es la minimización de la potencia media transmitida $\mathbb{E}_{\mathbf{G}}[p_{k,m|l(\mathbf{G})}w_{k,m}(\mathbf{G})]$ a lo largo de todas las sub-portadoras y usuarios garantizando requisitos promedios de tasa y BER para cada usuario. De forma explícita, se quiere que para usuario dado el promedio de la de la suma de las tasas transmitidas por todas las sub-portadoras alcance un nivel mínimo pre-especificado, i.e., $\sum_{k=1}^K \mathbb{E}_{\mathbf{G}}[r_{k,m|l(\mathbf{G})}w_{k,m}(\mathbf{G})] \geq \check{r}_m$. En lo que a la BER respecta, el número medio de bits erróneos esperado $\sum_{k=1}^K \mathbb{E}_{\mathbf{G}}[r_{k,m|l(\mathbf{G})}w_{k,m}(\mathbf{G})\epsilon(p_{k,m|l(\mathbf{G})}, g_{k,m}, r_{k,m|l(\mathbf{G})})]$ dividido por el número me-

¹⁴El sub-índice que aparece aquí escrito como l se escribirá de forma explícita como $l(\mathbf{G})$ en los lugares en donde esta dependencia deba ser enfatizada.

dio de bits transmitidos $\sum_{k=1}^K \mathbb{E}_{\mathbf{G}}[r_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G})]$ correspondientes al usuario m deberá mantenerse por debajo del nivel máximo pre-especificado por $\check{\epsilon}_m$. Sin embargo puesto que esta ligadura acopla las variables $p_{k,m|l(\mathbf{G})}$ y $r_{k,m|l(\mathbf{G})}$ para todas las $\sum_{k=1}^K L_{k,m}$ regiones, resulta más conveniente simplificar ligeramente la formulación del requisito de BER y, de forma similar a (3.27), limitar el valor esperado de la BER para cada una de las regiones, estableciendo así $\check{\epsilon}_m$ como límite superior para $\mathbb{E}_{\mathbf{G} \in \mathcal{R}_{k,m|l}}[r_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G}) \epsilon(p_{k,m|l(\mathbf{G})}, g_{k,m}, r_{k,m|l(\mathbf{G})})] / \mathbb{E}_{\mathbf{G} \in \mathcal{R}_{k,m|l}}[r_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G})] \forall k, m, l$.

En base a las anteriores expresiones, la minimización de la potencia media ponderada que transmite el sistema garantizando calidad de servicio individual queda:

$$\left\{ \begin{array}{l} \text{mín}_{\mathbf{P}(\mathbf{G}) \geq \mathbf{0}, \mathbf{R}(\mathbf{G}) \geq \mathbf{0}, \mathbf{W}(\mathbf{G}) \geq \mathbf{0}} \bar{P}, \quad \bar{P} := \sum_{m=1}^M \mu_m \sum_{k=1}^K \mathbb{E}_{\mathbf{G}}[p_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G})] \\ \text{s. a : } C1. \sum_{k=1}^K \mathbb{E}_{\mathbf{G}}[r_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G})] \geq \check{r}_m, \quad \forall m, \\ C2. \frac{\mathbb{E}_{\mathbf{G} \in \mathcal{R}_{k,m|l}} \left[r_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G}) \kappa_1 \exp \left(-\frac{p_{k,m|l(\mathbf{G})} \kappa_2 g_{k,m}}{2^{r_{k,m|l(\mathbf{G})}} - 1} \right) \right]}{\mathbb{E}_{\mathbf{G} \in \mathcal{R}_{k,m|l}}[r_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G})]} \leq \check{\epsilon}_m, \quad \forall k, m, l, \\ C3. \sum_{m=1}^M w_{k,m}(\mathbf{G}) \leq 1, \quad \forall k, \mathbf{G}, \end{array} \right. \quad (4.36)$$

donde para C2 se ha utilizado la expresión en (4.35) y a través de la tercera restricción se fuerza a que el uso temporal de una sub-portadora por parte de los usuarios no exceda de uno.

Para finalizar el manejo de (4.36) se definirán las fracciones de acceso promedio por región $\bar{w}_{k,m|l} := \int_{\mathbf{G} \in \mathcal{R}_{k,m|l}} w_{k,m}(\mathbf{G}) f_{\mathbf{G}}(\mathbf{G}) d\mathbf{G}$, junto con las variables de potencia y tasa asociadas $\bar{p}_{k,m|l} := \int_{\mathbf{G} \in \mathcal{R}_{k,m|l}} p_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G}) f_{\mathbf{G}}(\mathbf{G}) d\mathbf{G} = p_{k,m|l} \bar{w}_{k,m|l}$ y $\bar{r}_{k,m|l} := r_{k,m|l} \bar{w}_{k,m|l}$. Recuérdese que $p_{k,m|l}$ y $r_{k,m|l}$ son región-dependientes y que, por lo tanto, representan magnitudes deterministas sobre la región que indexan.

Con este cambio de variables, (4.36) puede reescribirse como:

$$\left\{ \begin{array}{l} \text{mín}_{\bar{p}_{k,m|l}, \bar{r}_{k,m|l}, w_{k,m}(\mathbf{G})} \bar{P}, \quad \bar{P} := \sum_{m=1}^M \mu_m \sum_{k=1}^K \sum_{l=1}^{L_{k,m}} \bar{p}_{k,m|l} \\ \text{s. to : } C1. \quad - \sum_{k=1}^K \sum_{l=1}^{L_{k,m}} \bar{r}_{k,m|l} + \check{r}_m \leq 0, \quad \forall m, \\ \quad C2. \quad \mathbb{E}_{\mathbf{G} \in \mathcal{R}_{k,m|l}} \left[w_{k,m}(\mathbf{G}) \kappa_1 \exp \left(- \frac{\bar{p}_{k,m|l}(\mathbf{G}) \kappa_2 g_{k,m}}{\frac{\bar{r}_{k,m|l}(\mathbf{G})}{2 \frac{w_{k,m|l}}{-1}}} \right) \right] - \check{\epsilon}_m \bar{w}_{k,m|l} \leq 0, \quad \forall k, m, l, \\ \quad C3. \quad \sum_{m=1}^M w_{k,m}(\mathbf{G}) - 1 \leq 0, \quad \forall k, \mathbf{G} \quad C4. \quad - \bar{p}_{k,m|l} \leq 0, \quad \forall k, m, l, \\ \quad C5. \quad - \bar{r}_{k,m|l} \leq 0, \quad \forall k, m, l, \quad C6. \quad - w_{k,m}(\mathbf{G}) \leq 0, \quad \forall k, m, \mathbf{G}, \end{array} \right. \quad (4.37)$$

Pese a que no pueda garantizarse que la optimización conjunta en (4.37) sea convexa c.r.a. a las variables $\bar{p}_{k,m|l}$, $\bar{r}_{k,m|l}$ y $w_{k,m}$, en el punto siguiente se propondrá una simplificación que garantiza la convexidad del problema y que por consiguiente, garantiza la convergencia al mínimo global de forma eficiente.

El objetivo en (4.37) es la minimización de la potencia ponderada *media* transmitida *sobre todas las posibles realizaciones del canal*. Sin embargo, las ligaduras tienen distinto carácter estadístico (instantáneo frente a promedio) e involucran distintos tipos de CSI: *C1* es un requisito promedio; *C2* se refiere a un promedio sobre una región de cuantificación; *C3* deber ser satisfecho para cada realización del canal y *C4* – *C6* toman como base distintos tipos de CSI en función de la naturaleza de la variable cuyo valor restringen. En el siguiente punto se derivarán las condiciones KKT asociadas con (4.37). Estas no sólo conducirán a las expresiones que dicten la construcción de los modos de transmisión óptimos, sino que también proveerán entendimiento e intuición de la estructura de los esquemas de asignación eficiente de recursos¹⁵.

Antes de cerrar este punto debe señalarse que la observación 3.5 relativa a la utilización de esquemas de pre-distorsión para reducir el PAPR de cada uno de los usuarios aplica también a sistemas OFDMA.

¹⁵El papel clave que desempeñan en esta sección las condiciones KKT es la causa por la que en esta ocasión en (4.37) se han escrito en formato *estándar* ($C(\mathbf{x}) \leq 0$ con $C(\mathbf{x})$ denotando la restricción c.r.a. las variables de optimización \mathbf{x}) y de forma explícita (*C4-C6*) todas las restricciones del problema de optimización; véase, e.g., [13]

Condiciones KKT: esquemas óptimos de adaptación

Sean β_m^r , $\beta_{k,m|l}^\epsilon$, β_k^w , $\alpha_{k,m|l}^p$, $\alpha_{k,m|l}^r$, $\alpha_{k,m}^w$ los multiplicadores de Lagrange (positivos o nulos) asociados respectivamente con las restricciones C1-C6. Si las derivadas de la función dual de Lagrange de (4.37) c.r.a. $\bar{p}_{k,m|l}$ y $\bar{r}_{k,m|l}$ se igualan a cero en el punto óptimo, se obtienen, tras la realización de algunas manipulaciones matemáticas, la condición KKT

$$\begin{aligned} & \int_{\mathbf{G} \in \mathcal{R}_{k,m|l}} w_{k,m}^*(\mathbf{G}) \kappa_2 g_{k,m} \left(\frac{\beta_m^{r*} + \alpha_{k,m|l}^{r*}}{(\mu_m - \alpha_{k,m|l}^{p*}) \frac{\bar{p}_{k,m|l}^*}{\bar{w}_{k,m|l}} \ln(2)} - 1 \right) \\ & \times \kappa_1 \exp \left(- \frac{\beta_m^{r*} + \alpha_{k,m|l}^{r*} - (\mu_m - \alpha_{k,m|l}^{p*}) \frac{\bar{p}_{k,m|l}^*}{\bar{w}_{k,m|l}} \ln(2)}{(\mu_m - \alpha_{k,m|l}^{p*}) \ln(2)} \kappa_2 g_{k,m} \right) f_{\mathbf{G}}(\mathbf{G}) d\mathbf{G} \\ & = \frac{\mu_m - \alpha_{k,m|l}^{p*}}{\beta_{k,m|l}^{\epsilon*}}, \end{aligned} \quad (4.38)$$

proveniente de la derivada del Lagrangiano de (4.37) c.r.a. $\bar{p}_{k,m|l}$ y su homóloga para el caso de $\bar{r}_{k,m|l}$

$$\frac{\bar{r}_{k,m|l}^*}{\bar{w}_{k,m|l}} = \log_2 \left(\frac{\beta_m^{r*} + \alpha_{k,m|l}^{r*}}{\beta_m^{r*} + \alpha_{k,m|l}^{r*} - (\mu_m - \alpha_{k,m|l}^{p*}) \frac{\bar{p}_{k,m|l}^*}{\bar{w}_{k,m|l}} \ln(2)} \right). \quad (4.39)$$

Las condiciones KKT para C4 y C5 también establecen $\bar{p}_{k,m|l}^* \alpha_{k,m|l}^{p*} = 0$ y $\bar{r}_{k,m|l}^* \alpha_{k,m|l}^{r*} = 0$ [9]. Estas ecuaciones implican que $\bar{p}_{k,m|l}^* > 0$ si y sólo si (s.y.s.s.) $\alpha_{k,m|l}^{p*} = 0$ y $\bar{r}_{k,m|l}^* > 0$ s.y.s.s. $\alpha_{k,m|l}^{r*} = 0$. Cuando $\alpha_{k,m|l}^{p*} \neq 0$ y/o $\alpha_{k,m|l}^{r*} \neq 0$, entonces $\bar{p}_{k,m|l}^* = \bar{r}_{k,m|l}^* = 0$ y, por consiguiente, la región $\mathcal{R}_{k,m|l}$ es inactiva en el sentido de que no afecta a la asignación de recursos. Por otro lado, haciendo $\alpha_{k,m|l}^{p*} = \alpha_{k,m|l}^{r*} = 0$ en (4.38) y (4.39), se tiene que $\bar{p}_{k,m|l}^*/\bar{w}_{k,m|l} < \beta_m^{r*}/(\ln(2)\mu_m)$ debe ser cierto para que la región $\mathcal{R}_{k,m|l}$ esté activa. Intuitivamente, si la ganancia del canal en la región $\mathcal{R}_{k,m|l}$ es tan pobre que para satisfacer el requisito de BER la potencia necesaria excede el precio óptimo representado por $\beta_m^{r*}/(\ln(2)\mu_m)$, entonces la asignación óptima de potencia y tasa para esa región es cero.

Suponiendo que $\mathcal{R}_{k,m|l}$ es activa y derivando el Lagrangiano de (4.37) c.r.a.

$w_{k,m}(\mathbf{G})$, se obtiene para el óptimo que

$$\left\{ \frac{\mu_m \bar{p}_{k,m|l}^*}{\bar{w}_{k,m|l}} - \beta_m^{r^*} \frac{\bar{r}_{k,m|l}^*}{\bar{w}_{k,m|l}} + \beta_{k,m|l}^{\epsilon^*} \left[\kappa_1 \exp \left(-\frac{\frac{\bar{p}_{k,m|l}^*}{\bar{w}_{k,m|l}} \kappa_2 g_{k,m}}{\frac{\bar{r}_{k,m|l}^*}{\bar{w}_{k,m|l}} - 1} \right) - \check{\epsilon}_m \right] \right\} f_{\mathbf{G}}(\mathbf{G}) + \beta_k^{w^*}(\mathbf{G}) - \alpha_{k,m}^{w^*}(\mathbf{G}) = 0, \quad (4.40)$$

donde se ha hecho la dependencia de β_k^w c.r.a. \mathbf{G} explícita. (Recuérdese que $\beta_m^r \neq \beta_m^r(\mathbf{G})$ puesto que β_m^r está asociado con una restricción de un valor promedio.) Recuérdese también que si $\bar{w}_{k,m|l} \neq 0$, entonces $\bar{p}_{k,m|l}^*/\bar{w}_{k,m|l} := p_{k,m|l}^*$ y $\bar{r}_{k,m|l}^*/\bar{w}_{k,m|l} := r_{k,m|l}^*$ en (4.38)-(4.40). Para valorar correctamente las implicaciones de (4.40), defínase

$$\varphi_{k,m}(\mathbf{G}) := \mu_m p_{k,m|l}^*(\mathbf{G}) - \beta_m^{r^*} r_{k,m|l}^*(\mathbf{G}) + \beta_{k,m|l}^{\epsilon^*} \left[\kappa_1 \exp \left(-\frac{p_{k,m|l}^*(\mathbf{G}) \kappa_2 g_{k,m}}{r_{k,m|l}^*(\mathbf{G}) - 1} \right) - \check{\epsilon}_m \right], \quad (4.41)$$

para representar el *coste neto* cuando el usuario m utiliza la sub-portadora k [cf. (4.6) y (4.14)]. En base a esta definición (4.40) puede reescribirse como

$$\varphi_{k,m}(\mathbf{G}) f_{\mathbf{G}}(\mathbf{G}) + \beta_k^{w^*}(\mathbf{G}) - \alpha_{k,m}^{w^*}(\mathbf{G}) = 0, \quad \forall \mathbf{G}, \quad \forall m \in \{1, \dots, M\}. \quad (4.42)$$

Es conveniente hacer notar de forma expresa que: (i) la parte izquierda de la igualdad en (4.42) (que se denotará como $LHS(4.42)$) no depende de forma explícita de $w_{k,m}^*(\mathbf{G})$ sino únicamente de forma implícita a través de los multiplicadores $\beta_k^{w^*}(\mathbf{G})$ y $\alpha_{k,m}^{w^*}(\mathbf{G})$; (ii) el multiplicador $\beta_k^{w^*}(\mathbf{G})$ es común $\forall m$ y (iii) para una misma sub-portadora k y una realización \mathbf{G} dada, el coste neto $\varphi_{k,m}(\mathbf{G})$ está fijo y, en general, es distinto para cada usuario m . Es también importante tener en cuenta que, para cada k , la condición KKT correspondiente a $C6$ establece

$$w_{k,m}^*(\mathbf{G}) \alpha_{k,m}^{w^*}(\mathbf{G}) = 0, \quad \forall \mathbf{G}, \quad \forall m \in \{1, \dots, M\}. \quad (4.43)$$

Puesto que $\varphi_{k,m}(\mathbf{G})$ es fijo para un \mathbf{G} dado [cf. (4.41)], para cada sub-portadora k (4.42) representa un sistema indeterminado de M ecuaciones con $M+1$ incógnitas,

a saber, $\beta_k^{w^*}(\mathbf{G})$ y $\{\alpha_{k,m}^{w^*}(\mathbf{G})\}_{m=1}^M$. Esto hace que con \mathbf{G} y k fijos, $\alpha_{k,m}^{w^*}(\mathbf{G}) = 0$ para no más de un m , puesto que en otro caso el sistema de M ecuaciones se vuelve sobre-determinando y (4.42) no tiene solución (i.e., no se puede satisfacer para todos y cada uno de los valores de m). Por otro lado, si $\alpha_{k,m}^{w^*}(\mathbf{G}) \neq 0 \forall m$, la condición (4.43) implica que $w_{k,m}^*(\mathbf{G}) = 0 \forall m$ y la sub-portadora k se rechaza $\forall m$ (ningún usuario transmite potencia a través de ella). De esta manera, puesto que la solución exacta de (4.42) requiere a lo sumo un $\alpha_{k,m}^{w^*}$ nulo y al menos un $\alpha_{k,m}^{w^*}$ nulo (para evitar la situación no deseada de que $\alpha_{k,m}^{w^*}(\mathbf{G}) \neq 0 \forall m$), se deduce que $\alpha_{k,m}^{w^*}(\mathbf{G}) = 0$ para exactamente un único usuario m por sub-portadora k y realización de canal \mathbf{G} . En otras palabras, el esquema de acceso óptimo permite únicamente a un usuario m_k transmitir a través de la k -ésima sub-portadora. Como $\beta_k^{w^*}(\mathbf{G})[\sum_{m=1}^M w_{k,m}^*(\mathbf{G}) - 1] = 0$ y $\beta_k^{w^*}(\mathbf{G}) \neq 0$, esto supone además que $w_{k,m_k}^*(\mathbf{G}) = 1$ y $w_{k,m}^*(\mathbf{G}) = 0$ para $m \neq m_k$, es decir que el usuario que trasmite utiliza la sub-portadora en su totalidad (como se recordará este comportamiento oportunista se había encontrado también para el caso de TDMA). La siguiente proposición detalla el mecanismo por el cual se determina el usuario m_k que de forma óptima debe acceder a la sub-portadora k .

Proposición 4.5 *El usuario óptimo m_k que debe acceder a la k -ésima sub-portadora es aquel cuyo coste neto para la sub-portadora sea mínimo, i.e., $m_k = \arg \min_m \{\varphi_{k,m}(\mathbf{G})\}_{m=1}^M$.*

Demostración: Supóngase que m_k es el usuario candidato a utilizar la sub-portadora k , i.e., aquel para el que $w_{k,m_k}^*(\mathbf{G}) = 1$ y $\alpha_{k,m_k}^{w^*}(\mathbf{G}) = 0$. Para este usuario, (4.42) implica que $\beta_k^{w^*}(\mathbf{G}) = -\varphi_{k,m_k}(\mathbf{G})f_{\mathbf{G}}(\mathbf{G})$. Aplicando ahora (4.42) para otro usuario $m'_k \neq m_k$ se tiene que $\varphi_{k,m'_k}(\mathbf{G})f_{\mathbf{G}}(\mathbf{G}) + \beta_k^{w^*}(\mathbf{G}) - \alpha_{k,m'_k}^{w^*}(\mathbf{G}) = [\varphi_{k,m'_k}(\mathbf{G}) - \varphi_{k,m_k}(\mathbf{G})]f_{\mathbf{G}}(\mathbf{G}) - \alpha_{k,m'_k}^{w^*}(\mathbf{G}) = 0$. La satisfacción de la condición anterior exige $\varphi_{k,m'_k}(\mathbf{G}) \geq \varphi_{k,m_k}(\mathbf{G})$, puesto que $\alpha_{k,m'_k}^{w^*}(\mathbf{G}) \geq 0$; es decir $\varphi_{k,m_k}(\mathbf{G}) = \min_m \varphi_{k,m}(\mathbf{G})$. □

Nótese que si más de un usuario presenta el valor mínimo de $\{\varphi_{k,m}\}_{m=1}^M$ cual-

quier división arbitraria del acceso a la sub-portadora k entre estos usuarios es óptima (al igual que se hizo en secciones anteriores, simplemente puede elegirse uno de ellos de forma aleatoria). Por último, si existe una realización \mathbf{G} tal que $\varphi_{k,m}(\mathbf{G}) > 0 \forall m$, entonces, puesto que $\beta_k^{w*}(\mathbf{G}) \geq 0$ se deduce de (4.42) que $\alpha_{k,m}^{w*}(\mathbf{G}) \neq 0, \forall m$ (la solución óptima no asignará esta sub-portadora k a ningún usuario). Así pues, introduciendo un usuario virtual (ficticio) representado por $m = 0$ de manera que $\varphi_{k,0}(\mathbf{G}) = 0 \forall k, \mathbf{G}$, el valor de $w_{k,m}^*(\mathbf{G})$ puede expresarse en forma compacta utilizando la función indicador como

$$w_{k,m}^*(\mathbf{G}) = \mathbf{I}_{\{m = \arg \min_{m'} \{\varphi_{k,m'}(\mathbf{G})\}_{m'=0}^M\}}. \quad (4.44)$$

Observación 4.4 *A partir de la inspección de (4.38), (4.39) y (4.44), puede deducirse que: (i) el único acoplamiento entre sub-portadoras es a través del multiplicador β_m^r (i.e., dado $\beta_m^r \forall m$, la asignación de tasa y potencia en cada sub-portadora puede realizarse de forma independiente); (ii) dado β_m^r , la asignación óptima de tasa y potencia para el usuario m no depende de las asignaciones en otras sub-portadoras y (iii) para una sub-portadora k , la asignación óptima de usuarios equivale a satisfacer conjuntamente (4.42) $\forall m$.*

Al igual que para el caso de un sistema TDMA, el hecho de que el acceso óptimo sea oportunista para cada realización del canal no significa que en promedio exista un acaparamiento de recursos por parte de un único usuario. Por el contrario, la formulación del problema propuesta garantiza una distribución equitativa de recursos de acuerdo a lo expuesto en la observación 4.2. Conviene asimismo destacar que, a diferencia del sistema anterior, puesto que en un sistema OFDMA son varios los recursos (sub-portadoras) los que se comparten, durante una misma realización del canal potencialmente serán varios los usuarios que envíen simultáneamente información a través del mismo. (Argumentando en términos de diversidad, es sencillo comprobar cómo cuanto menor sea la correlación entre distintas sub-portadoras, menor será también la correlación entre los valores de $\varphi_{k,m}(\mathbf{G})$ y mayor será el número de usuarios que accedan al canal.)

Nótese que tanto la asignación de acceso oportunista, como la definición de un funcional de coste que indica el precio de acceso al canal por parte de cada usuario (mediante la ponderación simultánea de potencia y tasa transmitidas y BER cometida) obtenidas en esta sección para sistemas OFDMA estaban presentes también en la sección anterior cuando se trabajaba con sistemas TDMA. La estructura de ambos problemas hace pensar además que la asignación oportunista sería una característica común de los sistemas multi-usuario con acceso ortogonal y con QoS y Q-CSIT adecuadamente diseñado. Por último, debe reseñarse que, si bien de resultado semejante, el camino para llegar a este resultado ha sido diferente en ambos casos. La causa principal se debe a que al incremento de dimensionalidad del problema OFDMA (K canales simultáneos y optimización simultánea sobre potencia y tasa) hace por un lado que el método de demostración elegido en la sección anterior, aunque intuitivo, resulte de más compleja aplicación a este caso, mientras que, por otro lado, la existencia de condiciones de optimización relacionadas con la tasa de transmisión provoca que el tratamiento de las condiciones KKT (método utilizado en este caso para derivar el esquema de acceso) tenga que ser en esta ocasión más cauteloso y detallado.

Hasta este punto se han obtenido las condiciones que los modos de transmisión y los esquemas de adaptación óptimos deben satisfacer. A continuación se realiza la descripción genérica de los pasos de un algoritmo que haciendo uso de estas condiciones y de las propiedades recogidas en la observación 4.4 puede ser ejecutado para obtener de forma sistemática los valores de la solución óptima.

Algoritmo 4.2 Asignación de recursos genérica

(S2.0) Sea δ un pequeño número positivo y β^r el vector formado por $\{\beta_m^r\}_{m=1}^M$.

Comience con un $\beta^r \geq \mathbf{0}$ arbitrario.

(S2.1) Para cada sub-portadora k :

(S2.1.1) Fije $\beta_{k,m|l}^\epsilon \geq 0 \forall m, l$ de forma arbitraria.

(S2.1.2) Inicialice $p_{k,m|l} = \bar{p}_{k,m|l} / \bar{w}_{k,m}$ de manera que $0 < p_{k,m|l} <$

$$\beta_m^r / \ln(2), \forall m, l.$$

(S2.1.3) Para $\alpha_{k,m|l}^p = 0$ y $\alpha_{k,m|l}^r = 0$, utilice (4.39) para obtener $r_{k,m|l} = \bar{r}_{k,m|l} / \bar{w}_{k,m} \forall m, l$.

(S2.1.4) Calcule $w_{k,m}(\mathbf{G}) \forall m$ de acuerdo a (4.44).

(S2.1.5) Verifique (4.38) $\forall m, l$. Si $|LHS(4.38) - 1/\beta_{k,m|l}^\epsilon| < \delta\beta_{k,m|l}^\epsilon \forall m, l$ vaya a (S2.1.6); en caso contrario aumente $p_{k,m|l}$ si (m, l) es tal que $LHS(4.38) > 1/\beta_{k,m|l}^\epsilon$; disminuya $p_{k,m|l}$ si (m, l) es tal que $LHS(4.38) < 1/\beta_{k,m|l}^\epsilon$, y vuelva a (S2.1.3).

(S2.1.6) Verifique la restricción C2 en (4.37) $\forall m, l$. Si $|C2| < \delta\check{\epsilon}_m \forall m, l$, cambie a la siguiente sub-portadora $k+1$ y vuelva a (S2.1); en caso contrario aumente $\beta_{k,m|l}^\epsilon$ si (m, l) es tal que $C2 > 0$; disminuya $\beta_{k,m|l}^\epsilon$ si (m, l) es tal que $C2 < 0$ y vuelva a (S2.1.3).

(S2.2) Verifique la restricción C1 en (4.37) $\forall m$. Si $|C1| < \delta\check{r}_m \forall m$ entonces pare; en caso contrario aumente β_m^r si m es tal que $C1 > 0$; disminuya β_m^r si m es tal que $C1 < 0$, y vuelva a (S2.1).

La convergencia y la calidad de la solución resultantes de este algoritmo dependerán claramente de cómo se realicen las búsquedas sobre las distintas variables (e.g., qué esquemas se utilizan para incrementar o disminuir $p_{k,m|l}$, $\beta_{k,m|l}^\epsilon$ y β_m^r en los pasos (S2.1.5), (S2.1.6), y (S2.2), cuáles son los tamaños de los pasos, cuáles los niveles de tolerancia, etc.). Las razones fundamentales por las que no se ha proporcionado un mayor nivel de detalle respecto a estas búsquedas han sido: (i) se pretendía ilustrar que, además de caracterizar la solución óptima, en base a las condiciones KKT podía resolverse además el problema de optimización en su totalidad y (ii) para una implementación en sistemas reales se promueve el uso de un algoritmo simplificado que se describirá en el siguiente punto.

Observación 4.5 A diferencia de lo realizado para el caso de un único usuario transmitiendo sobre OFDM, en lugar de utilizar tasas de transmisión discretas, la optimización en esta sección se ha realizado suponiendo que las tasas de transmi-

sión pueden tomar valores continuos. La complejidad hardware para implementar modulaciones de tasa continua (a través de constelaciones no rectangulares y esquemas de codificación) es más elevada que la necesitada para modulaciones de tasa discreta [28]. Por otro lado la consideración de tasa continua permite una formulación más clara y allana el camino para la obtención de expresiones analíticas que facilitan el entendimiento del problema y proveen guías de diseño aplicables a ambos (posiblemente también a otros) escenarios. Finalmente, debe tenerse en cuenta que para prestaciones promediadas, los esquemas que, derivados suponiendo tasa continua, se transforman de manera óptima para su utilización como esquemas de asignación de tasas discretas (véase, e.g., [36]), se comportan de forma parecida.

Asignación de recursos de complejidad reducida

En lugar de minimizar \bar{P} , en este punto se minimizará \bar{P}^+ , una cota superior de \bar{P}^* . La minimización de una cota superior es una técnica comúnmente utilizada cuando la optimización directa sobre el objetivo original resulta compleja [9]. Para este fin, denote ϵ_P^{-1} la función inversa involucrada cuando se resuelve (4.35) c.r.a. $p_{k,m}$ y represente $g_{k,m|l}^{\min} := \min\{g_{k,m} \mid g_{k,m} \in \mathcal{R}_{k,m|l}\}$ la ganancia mínima (i.e., la peor) de la región $\mathcal{R}_{k,m|l}$. Se puede entonces acotar superiormente \bar{P} en (4.36) utilizando $\bar{P}^+ := \sum_{m=1}^M \mu_m \sum_{k=1}^K \mathbb{E}_{\mathbf{G}}[p_{k,m|l}^+(\mathbf{G}) w_{k,m}(\mathbf{G})]$, donde

$$p_{k,m|l}^+(\mathbf{G}) := \epsilon_P^{-1}(r_{k,m|l}(\mathbf{G}), g_{k,m|l}^{\min}(\mathbf{G}), \check{\epsilon}_m). \quad (4.45)$$

Los parámetros de diseño obtenidos utilizando esta técnica se identificarán con el uso del super-índice “+”. La cota superior en (4.45) ofrece las siguientes ventajas: (i) la restricción C2 se satisface automáticamente lo que hace que $\beta_{k,m|l}^\epsilon = 0$ y (ii) se establece una relación directa entre $p_{k,m|l}$ y $r_{k,m|l}$ a través de ϵ_P^{-1} y por lo tanto, sólo existe necesidad de optimizar sobre una de las variables, reduciendo así la dimensionalidad del problema.

Eliminando las variables de potencia, el nuevo problema puede formularse

como:

$$\left\{ \begin{array}{l} \text{mín}_{\mathbf{R}^+(\mathbf{G}) \geq \mathbf{0}, \mathbf{W}^+(\mathbf{G}) \geq \mathbf{0}} \bar{P}^+, \\ \bar{P}^+ := \sum_{m=1}^M \mu_m \sum_{k=1}^K \mathbb{E}_{\mathbf{G}}[\epsilon_P^{-1}(r_{k,m|l}^+, g_{k,m|l}^{\text{mín}}, \check{\epsilon}_m) w_{k,m}^+(\mathbf{G})] \\ \text{s. a : } \sum_{k=1}^K \mathbb{E}_{\mathbf{G}}[r_{k,m|l}^+ w_{k,m}^+(\mathbf{G})] \geq \check{r}_m, \quad \forall m; \\ \sum_{m=1}^M w_{k,m}^+(\mathbf{G}) \leq 1, \quad \forall k. \end{array} \right. \quad (4.46)$$

Utilizando $\bar{r}_{k,m|l}^+ := \mathbb{E}_{\mathbf{G}}[r_{k,m|l}^+ w_{k,m}^+(\mathbf{G})]$ y siguiendo pasos similares a aquellos en [91, ecuación (11)], la optimización en (4.46) puede demostrarse convexa.

Este diseño (basado en un principio de *peor escenario posible* para derivar la cota de la función objetivo) obtendrá $w_{k,m}^{+*}$ y $r_{k,m|l}^{+*}$ como solución de (4.46). Concretamente, definiendo $\kappa_{3,m} := \mu_m \kappa_2^{-1} \ln(\kappa_1 / \check{\epsilon}_m)$ la resolución de (4.46) da lugar a

$$r_{k,m|l}^{+*} = \log_2 \left(\frac{\beta_m^{r_{k,m|l}^{+*}} g_{k,m|l}^{\text{mín}}}{\ln(2) \kappa_{3,m}} \right), \quad (4.47)$$

$$\varphi_{k,m}^+(\mathbf{G}) = \frac{(2^{r_{k,m|l}^{+*}} - 1) \kappa_{3,m}}{g_{k,m|l}^{\text{mín}}} - \beta_m^{r_{k,m|l}^{+*}} r_{k,m|l}^{+*}, \quad \varphi_{k,0}^+(\mathbf{G}) = 0 \quad (4.48)$$

$$w_{k,m}^{+*}(\mathbf{G}) = \mathbf{I}_{\{m = \arg \min_{m'} \{\varphi_{k,m'}^+(\mathbf{G})\}_{m'=0}^M\}}, \quad (4.49)$$

donde la asignación de acceso dada por (4.48) y (4.49) sigue el mismo principio que el de la solución óptima para el problema original en (4.41) y (4.44) y descrito por la proposición 4.5. Además, debe hacerse notar que debido a la expresión logarítmica de ϵ_P^{-1} [cf. (4.35)], la asignación de tasa evoca la asignación clásica del algoritmo de “water-filling”.

Sin embargo, una vez obtenidos $\mathbf{R}^{+*}(\mathbf{G})$ (o de forma precisa $\mathbf{R}^{+*}(\mathbf{c})$, puesto que la variación c.r.a. \mathbf{G} se debe en (4.47) a $l(\mathbf{G})$) y $\mathbf{W}^{+*}(\mathbf{G})$ no hay necesidad de sustituir estos valores en (4.45) para calcular la carga final de potencia $\mathbf{P}^{+*}(\mathbf{c})$ en base a la cota superior utilizada, sino que, como alternativa, los valores finales de $\mathbf{R}^{+*}(\mathbf{c})$ y $\mathbf{W}^{+*}(\mathbf{G})$ pueden utilizarse para satisfacer de forma ajustada los requisitos de BER de la restricción C2 en (4.37), lo cual claramente generará una

potencia de transmisión menor que la basada en la cota superior. Sea $\mathbf{P}^{+++}(\mathbf{c})$ la matriz que denota el valor exacto de potencia que se obtiene tras sustituir los $\mathbf{R}^{++}(\mathbf{c})$ y $\mathbf{W}^{++}(\mathbf{G})$, obtenidos según (4.47) y (4.49), en la restricción C2 en (4.37). La valor genérico $p_{k,m|l}^{+++}$ se calcula entonces como la solución de

$$\int_{\mathbf{G} \in \mathcal{R}_{k,m|l}} w_{k,m}^{++}(\mathbf{G}) \kappa_1 e^{-\frac{p_{k,m|l}^{+++}(\mathbf{G})^{\kappa_2 g_{k,m}}}{r_{k,m|l}^{++}(\mathbf{G}) - 1}} f_{\mathbf{G}}(\mathbf{G}) d\mathbf{G} = \check{\epsilon}_m \bar{w}_{k,m|l}^+. \quad (4.50)$$

Aunque $p_{k,m|l}^{+++}$ en (4.50) no puede encontrarse en forma cerrada, la monotonía de la función exponencial permite obtener su valor a través de una sencilla búsqueda uni-dimensional utilizando el método de bisección. La calidad del diseño de los modos de transmisión y de los esquemas de adaptación obtenidos con esta alternativa de complejidad reducida será evaluada a través de simulaciones numéricas en el apartado 4.2.4. El siguiente algoritmo resume los pasos principales para la resolución del conjunto de ecuaciones descrito por (4.47), (4.49) y (4.50).

Algoritmo 4.3 Asignación de recursos simplificada

- (S3.0) Fije un pequeño número positivo δ e inicialice el vector $\boldsymbol{\beta}^r \geq \mathbf{0}$ de forma arbitraria.
- (S3.1) Para cada combinación (k, m, l) por iteración:
- (S3.1.1) Utilice β_m^r para determinar $r_{k,m|l}^+$ a través de (4.47).
 - (S3.1.2) Utilice β_m^r para determinar $w_{k,m}^+(\mathbf{G})$ a través de (4.49).
- (S3.2) Verifique la restricción de tasa en (4.46) $\forall m$; si $|\sum_{k=1}^K \mathbb{E}_{\mathbf{G}}[r_{k,m|l}^+ w_{k,m}^+(\mathbf{G})] - \check{r}_m| < \delta \check{r}_m \forall m$, entonces vaya (S3.3); en caso contrario, aumente β_m^r para los usuarios m cuya tasa media sea menor que \check{r}_m ; disminuya β_m^r para los usuarios m cuya tasa media sea mayor que \check{r}_m ; y vaya a (S3.1).
- (S3.3) Una vez que se obtengan $\mathbf{R}^{++}(\mathbf{c})$, $\boldsymbol{\beta}^{r++}$, $\mathbf{W}^{++}(\mathbf{G})$, utilice (4.50) para calcular la potencia finalmente asignada.

La técnica para garantizar la convexidad del problema de optimización no sólo reduce la complejidad de la estructura de la solución sino que también posibilita el uso de métodos eficientes para la búsqueda de las variables involucradas en el cálculo de la solución. Especialmente amplio es el abanico para la búsqueda sobre los multiplicadores de Lagrange β^r (véase, e.g., [9]). Así por ejemplo, pueden inicializarse todas las componentes de β^r como un pequeño número positivo y actualizar después cada componente β_m^r de forma independiente $\forall m$ fijando $[\beta^r]_{m'}, \forall m' \neq m$ al valor de la previa iteración. La adaptación de cada β_m^r puede realizarse entonces usando una búsqueda uni-dimensional hasta que la restricción de tasa para el m -ésimo usuario se satisface de forma ajustada. Este simple algoritmo tiene convergencia garantizada y facilita la computación *distribuida* para cada usuario.

Estructura de la palabra de realimentación

Dado el diseño del cuantificador, se han derivado esquemas de asignación de recursos para la adaptación de tasa, potencia y acceso a las sub-portadoras por parte de los transmisores. De acuerdo a la figura 4.6, una vez que el cuantificador y la estrategia de asignación de recursos están diseñados, para cada realización \mathbf{G} el PA cuantifica y realimenta a los usuarios una palabra (vector de bits) que identifica el acceso de los usuarios a las distintas sub-portadoras así como la región a la que los canales asociados pertenecen. A través de esta forma de Q-CSIT, cada usuario es informado sobre su propio subconjunto de sub-portadoras (en caso de que acceda a alguna) y en base a los índices de región notificados escoge los niveles de potencia y tasa de transmisión de acuerdo a constelaciones (tablas) que se le han hecho llegar a cada uno de los transmisores durante la fase de inicialización del sistema tras haberse ejecutado el algoritmo 4.3. La proposición siguiente describe la construcción (y por ende la estructura) de la palabra de realimentación $\mathbf{c}^*(\mathbf{G})$.

Proposición 4.6 *Dado el diseño del cuantificador y el diseño de los modos de*

transmisión $(\mathbf{P}^{+*}(\mathbf{c}), \mathbf{R}^{+*}(\mathbf{c}), \mathbf{W}^{+*}(\boldsymbol{\beta}^{r+*}))$ generados por el algoritmo 4.3, el PA realimenta a los usuarios la palabra $\mathbf{c}^*(\mathbf{G}) = [\mathbf{c}_1^*(\mathbf{G}), \dots, \mathbf{c}_K^*(\mathbf{G})]$ que indica la asignación óptima de recursos (adaptación de los transmisores) para el estado actual del canal, donde $\mathbf{c}_k^*(\mathbf{G}) = [m_k^*(\mathbf{G}), l_k^*(\mathbf{G})]^T$ se construye $\forall k$ de acuerdo a:

1. $m_k^*(\mathbf{G}) = \arg \min_m \{\varphi_{k,m}^+(\mathbf{G}, \mathbf{P}^{+*}, \mathbf{R}^{+*}, \boldsymbol{\beta}^{r+*})\}_{m=1}^M$ (seleccione aleatoriamente cualquier usuario m_k^* cuando existan múltiples mínimos) y
2. $l_k^*(\mathbf{G}) = \{ l \mid \mathbf{G} \in \mathcal{R}_{k,m_k^*(\mathbf{G}),l}, \quad l = 1, \dots, L_k \}$.

Nótese que la proposición 4.6 también rige para el algoritmo 1 sin más que eliminar el super-índice “+” de los parámetros de diseño correspondientes.

Observación 4.6 *La estructura de la palabra $\mathbf{c}^*(\mathbf{G})$ en la proposición 4.6 codifica información asociada a cada sub-portadora (a saber, el usuario asignado y la región correspondiente) que el envío de la información de transmisión asociada a cada usuario independiente (i.e., el conjunto de sub-portadoras que puede utilizar y los índices de la región de pertenencia para cada una de ellas). Puesto que en cada sub-portadora se consideran $L_k - 1$ regiones activas y una región inactiva (o de renuncia a la transmisión), puede reducirse el número de bits realimentados codificando exclusivamente las regiones activas. Únicamente cuando para una misma sub-portadora la ganancia de todos los usuarios pertenece a la región inactiva correspondiente, deberá codificarse un mensaje de rechazo de la sub-portadora. Esto puede realizarse de manera sencilla mediante el indexado de un usuario virtual (e.g., $m = 0$) con una única región. Considerando todos estos índices, la longitud de la palabra de realimentación $\mathbf{c}^*(\mathbf{G})$ será $\left\lceil \sum_{k=1}^K \log_2 \left(\sum_{m=1}^M (L_{k,m} - 1) + K \right) \right\rceil$ bits.*

Nótese que particularizando para $K = 1$ la estructura de la palabra \mathbf{c} puede aplicarse también para el esquema de Q-CSIT para sistemas TDMA analizado en el apartado 4.1.3, lo que potencialmente puede contribuir a disminuir la tasa de realimentación requerida [cf. proposición 4.2].

Se concluye este apartado enfatizando que puesto que se ha formulado el problema de optimización de manera que las variables $\mathbf{P}(\mathbf{c})$ y $\mathbf{R}(\mathbf{c})$ en (4.36) y (4.46) están involucradas únicamente en expresiones promediadas en el tiempo, las soluciones óptimas $\mathbf{P}^{+*}(\mathbf{c})$ y $\mathbf{R}^{+*}(\mathbf{c})$ se calculan de forma “off-line” y únicamente el acceso de los usuarios al canal (involucrado en restricciones instantáneas) y el índice de las regiones de cuantificación de estas matrices deben ser calculados (obtenidos) de forma “on-line”.

4.2.3. Diseño del cuantificador

En la sección anterior se resolvió el objetivo (OB.2) derivando esquemas óptimos y sub-óptimos de asignación de sub-portadoras, tasa y potencia suponiendo unas regiones de cuantificación $\mathcal{R}_{k,m|l}$ dadas. En esta sección se abordará (OB.1) derivando un esquema iterativo óptimo para la selección de $\mathcal{R}_{k,m|l}^*$, así como un esquema alternativo sub-óptimo que simplifica considerablemente el diseño del cuantificador.

Cuantificación óptima

Además de $\mathbf{P}^*(\mathbf{c})$, $\mathbf{R}^*(\mathbf{c})$, y $\mathbf{W}^*(\mathbf{G})$, en esta sección se considerarán $\{\mathcal{R}_{k,m|l}^*\}_{l=1}^{L_{k,m}}$ como variables $\forall(k, m)$ para la optimización de \bar{P} en (4.36). Antes de proceder con este problema de optimización conjunta, ha de recordarse que: (i) la asignación de recursos óptima del apartado 4.2.2 puede descomponerse para cada usuario m y sub-portadora k ; y (ii) $\mathcal{R}_{k,m|l}$ representa una región de cuantificación de una única variable $g_{k,m}$. De hecho, $\{\mathcal{R}_{k,m|l}^*\}_{l=1}^{L_{k,m}}$ se puede representar por un conjunto de umbrales equivalentes $\{\tau_{k,m|l}^*\}_{l=1}^{L_{k,m}+1}$, con $\tau_{k,m|1}^* = 0$ y $\tau_{k,m|L_{k,m}+1}^* = \infty \forall(k, m)$. En otras palabras, gracias a (i) y (ii) el problema de cuantificación vectorial se reduce a KM problemas de cuantificación escalar.

Para determinar $\{\tau_{k,m|l}^*\}_{l=2}^{L_{k,m}} \forall k, m$ se comienza por la función dual de Lagrange \mathcal{L} de (4.36). Si se expresa esta última en términos de los umbrales y si se define

$\mathcal{G}_{k,m}\{x, y\} := \{\mathbf{G} : g_{k,m} \in [x, y]\}$, se puede escribir

$$\begin{aligned} \mathcal{L} := & \sum_{k=1}^K \sum_{m=1}^M \sum_{l=1}^{L_{k,m}} \int_{\mathbf{G} \in \mathcal{G}_{k,m}\{\tau_{k,m|l}, \tau_{k,m|l+1}\}} \left[p_{k,m|l}(\mathbf{G}) - \beta_m^r r_{k,m|l}(\mathbf{G}) \right. \\ & \left. + \beta_{k,m|l}^{\epsilon*} \kappa_1 e^{-\frac{p_{k,m|l}(\mathbf{G})^{\kappa_2} g_{k,m}}{2^{\tau_{k,m|l}(\mathbf{G})-1}}} - \beta_{k,m|l}^{\epsilon} \check{\epsilon}_m \right] w_{k,m}(\mathbf{G}) f_{\mathbf{G}}(\mathbf{G}) d\mathbf{G}, \end{aligned} \quad (4.51)$$

en donde, por claridad, se han omitido las contribuciones de las ligaduras que no dependen de $\{\tau_{k,m|l}\}_{l=1}^{L_{k,m}+1}$.

Las condiciones KKT necesarias después de fijar $\partial \mathcal{L} / \partial \tau_{k,m|l}^* = 0$, son

$$\begin{aligned} \mu_m p_{k,m|l-1} - \beta_m^r r_{k,m|l-1} + \beta_{k,m|l-1}^{\epsilon} \kappa_1 e^{-\frac{p_{k,m|l-1}^{\kappa_2} \tau_{k,m|l}^*}{2^{\tau_{k,m|l-1}-1}}} \\ = -\mu_m p_{k,m|l} + \beta_m^r r_{k,m|l} - \beta_{k,m|l}^{\epsilon} \kappa_1 e^{-\frac{p_{k,m|l}^{\kappa_2} \tau_{k,m|l}^*}{2^{\tau_{k,m|l}-1}}}. \end{aligned} \quad (4.52)$$

y pueden resolverse para $\tau_{k,m|l}^*$ mediante búsqueda uni-dimensional que encuentre los umbrales deseados $\forall(k, m)$.

Puesto que (4.52) contiene $\{\tau_{k,m|l}^*\}_{l=2}^{L_{k,m}}$, $\{p_{k,m|l}^*\}_{l=1}^{L_{k,m}}$ y $\{r_{k,m|l}^*\}_{l=1}^{L_{k,m}}$, para cada iteración interna del paso (S.1.1) en el algoritmo 4.2, deberá resolverse no sólo para $\{p_{k,m|l}^*\}_{l=1}^{L_{k,m}}$, $\{r_{k,m|l}^*\}_{l=1}^{L_{k,m}}$, y $w_{k,m}^*(\mathbf{G}) \forall m$ (como hasta ahora), sino también para $\{\tau_{k,m|l}^*\}_{l=2}^{L_{k,m}} \forall m$ a través de (4.52). Esto exige la modificación de los pasos (S2.1.1) y (S2.1.6) del algoritmo 4.2 e incorporar un paso adicional (S2.1.7) como sigue:

(S2.1.1) Además de $\beta_{k,m|l}^{\epsilon} \forall m, l$, inicialice también $\{\tau_{k,m|l}\}_{l=2}^{L_{k,m}} \forall m$.

(S2.1.6) Si se satisface C2 en lugar de ir a (S2.1), vaya al nuevo paso (S2.1.7).

(S2.1.7) Dado $\{p_{k,m|l}, r_{k,m|l}, \beta_{k,m|l}^{\epsilon}\}_{l=1}^{L_{k,m}}$, $w_{k,m}(\mathbf{G})$ y $\beta_m^r \forall m$ en (S2.1.6), actualice los valores de $\{\tau_{k,m|l}\}_{l=2}^{L_{k,m}} \forall m$ como en (4.52). Si el cambio c.r.a. los valores previos de $\{\tau_{k,m|l}\}_{l=2}^{L_{k,m}}$ es menor que δ , entonces vuelva a (S2.1); en caso contrario vuelva a (S2.1.3).

Con esta ampliación del algoritmo 4.2, los valores $\{p_{k,m|l}^*\}_{l=1}^{L_{k,m}}$, $\{r_{k,m|l}^*\}_{l=1}^{L_{k,m}}$ y $w_{k,m}^*(\mathbf{G})$, y $\{\tau_{k,m|l}^*\}_{l=2}^{L_{k,m}} \forall m$ deberán encontrarse iterativamente (para cada valor de β^r y sub-portadora k).

La condición de optimalidad en (4.52) se refiere al problema de optimización en (4.36). De forma análoga, para la asignación de recursos simplificada, la optimización en (4.46) puede llevarse incluyendo los umbrales en el conjunto de variables de optimización. Las condiciones KKT asociadas a los valores óptimos de los umbrales en este caso son

$$\begin{aligned} \mu_m \frac{2^{r_{k,m|l}^+} - 1}{\tau_{k,m|l-1}^*} \kappa_{3,m} - \beta_m^{r+} r_{k,m|l}^+ &= \mu_m \frac{2^{r_{k,m|l+1}^+} - 1}{\tau_{k,m|l}^*} \kappa_{3,m} - \beta_m^{r+} r_{k,m|l+1}^+ \\ &+ \frac{2^{r_{k,m|l+1}^+} - 1}{(\tau_{k,m|l}^*)^2} \kappa_{3,m} A_{k,m|l}, \end{aligned} \quad (4.53)$$

en donde $A_{k,m|l} := \frac{\int_{\mathbf{G} \in \mathcal{G}_{k,m} \{\tau_{k,m|l}^*, \tau_{k,m|l+1}^*\}} w_{k,m}(\mathbf{G}) f_{\mathbf{G}}(\mathbf{G}) d\mathbf{G}}{\int_{\mathbf{G} \in \mathcal{G}_{k,m} \{\tau_{k,m|l}^*\}} w_{k,m}(\mathbf{G}) f_{\mathbf{G}}(\mathbf{G}) d\mathbf{G}}$, y $\mathcal{G}_{k,m}\{x\} := \{\mathbf{G} : g_{k,m} = x\}$.

Nótese que la diferencia principal entre (4.52) y (4.53) es que en (4.46) se tiene $g_{k,m|l}^{\min} = \tau_{k,m|l}$. Esto implica que $p_{k,m|l+1}^+$ es una función explícita de $\tau_{k,m|l}$ y consecuentemente contribuye a la condición de optimalidad a través de $A_{k,m|l}$.

En el algoritmo 4.3, el paso (S3.0) ha de modificarse y el paso adicional (S3.1.3) debe ser incluido:

(S3.0) Junto con δ y β^r , inicialice también $\{\tau_{k,m|l}\}_{l=2}^{L_{k,m}} \forall k, m$.

(S3.1.3) Utilice $\{r_{k,m|l}^+\}_{l=1}^{L_{k,m}}$, $w_{k,m}^+(\mathbf{G})$ y $\beta_m^{r+} \forall m$, para actualizar los valores de $\{\tau_{k,m|l}\}_{l=2}^{L_{k,m}} \forall m$ como en (4.53). Si el cambio c.r.a. los valores previos de $\{\tau_{k,m|l}\}_{l=2}^{L_{k,m}}$ es menor que $\delta \forall m, l$, vuelva a (S3.1); en caso contrario, vuelva a (S3.1.1).

Aunque tanto (4.52) como (4.53) devuelven cuantificadores que alcanzan óptimos locales, no se puede garantizar optimalidad global debido a la falta de convexidad c.r.a. $\{\tau_{k,m|l}\}$. Por otro lado, es evidente que las versiones modificadas de los algoritmos 4.2 y 4.3 son más complejas que sus versiones iniciales. Por esta razón,

en la siguiente sección se presenta el diseño de un cuantificador no-iterativo que da lugar a una cuantificación del canal simple y efectiva. Antes de proceder a presentar dicho cuantificador, este apartado termina con una interesante observación que conecta la solución propuesta con la teoría clásica de cuantificación.

Observación 4.7 *Puesto que se está tratando con un problema de minimización con ligaduras, el Lagrangiano en (4.51) se puede interpretar como la métrica de distorsión media asociada con el diseño del cuantificador diseñado [23]. En este contexto, se pueden encontrar los umbrales de cuantificación y palabras-código de la constelación minimizando el coste medio \mathcal{L} a través del algoritmo de Lloyd [42]. Como se vio en el apartado 3.2.4 para el caso clásico del error cuadrático medio, dicho algoritmo minimiza una distorsión media iterando a través de dos pasos genéricos: (L1) dadas las regiones, encuentra las palabras-código óptimas asociadas (condición del centroide) y (L2) dadas las palabras-código actualizadas, encuentra las regiones óptimas correspondientes (regla del vecino más próximo). Nótese que desde el punto de vista de la optimización, (L1) equivale a fijar $\partial\mathcal{L}/\partial p_{k,m|l}$, $\partial\mathcal{L}/\partial r_{k,m|l}$, y $\partial\mathcal{L}/\partial w_{k,m}$ igual a cero ($p_{k,m|l}$, $r_{k,m|l}$, $w_{k,m}$ representa los centroides), mientras que (L2) es equivalente a $\partial\mathcal{L}/\partial\tau_{k,m|l} = 0$ ($\tau_{k,m|l}$ caracterizan las regiones). Para lanzar el algoritmo de Lloyd solamente se necesita definir una métrica para la distorsión, $d^{\mathcal{L}}$, como una función de \mathbf{G} . Si \mathbf{c} es una palabra-código que de la Proposición 2 (es decir, \mathbf{c} indexa $p_{k,m|l}$, $r_{k,m|l}$, $w_{k,m}$, $\forall k, m$), podemos eventualmente expresar $d^{\mathcal{L}}$ como*

$$d^{\mathcal{L}}(\mathbf{G}, \mathbf{c}(\mathbf{G})) := \sum_{m=1}^M \sum_{k=1}^K \left(\mu_m p_{k,m|l(\mathbf{G})} - \beta_m^{r*} r_{k,m|l(\mathbf{G})} + \beta_{k,m|l}^{\epsilon*} \kappa_1 e^{-\frac{p_{k,m|l(\mathbf{G})} \kappa_2 g_{k,m}}{2^{r_{k,m|l(\mathbf{G})} - 1}}} \right) w_{k,m}(\mathbf{G}). \quad (4.54)$$

Si se compara (4.54) con (4.52), se puede interpretar $\tau_{k,m|l}^*$ en (4.52) como la solución a la condición KKT aplicada en (4.36), o como el punto cuya distancia $d^{\mathcal{L}}$ del centroide ($p_{k,m|l}$, $r_{k,m|l}$, $w_{k,m}$) es igual a la distancia desde el centroide vecino ($p_{k,m|l-1}$, $r_{k,m|l-1}$, $w_{k,m}$).

Al igual que el algoritmo ARCC propuesto en el apartado 4.1.3, el algoritmo de Lloyd sigue un principio de descenso coordinado por bloques [9], converge a un óptimo local y, para métricas de distorsión unimodales, genera el cuantificador óptimo de carácter global [83].

Cuantificador de región equi-probable

La métrica (función de coste) para calcular $\tau_{k,m|l}$ en el punto anterior se basaba la minimización restringida de \bar{P} . En este punto primero se resolverá el problema de asignación de recursos suponiendo la disponibilidad de CSI sin cuantificación (es decir, $L_{k,m} \rightarrow \infty$), y subsecuentemente se calcula $\tau_{k,m|l}$ para satisfacer

$$\int_{\tau_{k,m|l}}^{\tau_{k,m|l+1}} \Pr(w_{k,m} = 1, g_{k,m}) dg_{k,m} = \int_{\tau_{k,m|l}}^{\tau_{k,m|L_{k,m}+1}} \Pr(w_{k,m} = 1, g_{k,m}) dg_{k,m} / L_{k,m}, \quad (4.55)$$

con $\tau_{k,m|1} = 0$ y $\tau_{k,m|L_{k,m}+1} = \infty$. Si la expresión (o el cómputo) de las probabilidades conjuntas es conocida, la resolución de (4.55) da lugar a umbrales $\{\tau_{k,m|l}\}_{l=1}^{L_{k,m}}$ por sub-portadora k y usuario m que dividen la probabilidad conjunta $\Pr(w_{k,m} = 1, g_{k,m})$ en regiones de igual área; de ahí la denominación *cuantificador de región equi-probable*. De forma intuitiva, este cuantificador intenta maximizar la entropía de la información de realimentación asociada a cada una de las sub-portadoras.

Para evaluar $\Pr(w_{k,m} = 1, g_{k,m})$ (expresión necesaria en (4.55)), se aplica la regla de Bayes para reescribir la probabilidad conjunta como $\Pr(w_{k,m} = 1|g_{k,m})f_{g_{k,m}}(g_{k,m})$, recordando que $f_{g_{k,i}}(g_{k,m})$ se supone conocida en (sp2). Para obtener $\Pr(w_{k,m} = 1|g_{k,m})$, se necesitará resolver el problema de asignación óptima de recursos sin suponer cuantificación. Para resolver este problema se aprovecharán los resultados obtenidos para (4.46). De esta forma, es claro que cuando $L_{k,m} \rightarrow \infty$, se tiene que $g_{\min} \rightarrow g$ en (4.48). Si $\varphi_{k,m}^\infty$ y $\beta_m^{r\infty*}$ denotan, respectivamente, el indicador del coste y el multiplicador de Lagrange cuando $L_{k,m} \rightarrow \infty$,

se puede obtener [cf. (4.48)]

$$\varphi_{k,m}^\infty(g_{k,m}) = \left[\frac{\beta_m^{r_{\infty^*}}}{\ln(2)} - \frac{\kappa_{3,m}}{g_{k,m}} \right]^+ - \beta_m^{r_{\infty^*}} \left[\log_2 \left(\frac{g_{k,m} \beta_m^{r_{\infty^*}}}{(\kappa_{3,m} \ln(2))} \right) \right]^+. \quad (4.56)$$

Puesto que (4.49) establece que $w_{k,m}(\mathbf{G}) = \mathbf{I}_{\{m=\arg \min_{m'} \{\varphi_{k,m'}^\infty(\mathbf{G})\}_{m'=0}^M\}}$, si $\varphi_{k,m}^\infty(g_{k,m}) < \varphi_{k,0}^\infty = 0$, entonces $g_{k,m} > \ln(2)\kappa_{3,m}/\beta_m^{r_{\infty^*}}$ es una condición necesaria para que el usuario m esté activo. Teniendo también en cuenta que los canales de usuarios distintos están incorrelados, [cf. (sp1)],

$$\Pr(w_{k,m} = 1 | g_{k,m}) = \mathbf{I}_{\{g_{k,m} > \frac{\ln(2)\kappa_{3,m}}{\beta_m^{r_{\infty^*}}}\}} \prod_{\nu=1, \nu \neq m}^M \Pr(\varphi_{k,m}^\infty < \varphi_{k,\nu}^\infty | g_{k,m}). \quad (4.57)$$

Es interesante observar que para los usuarios activos se tiene $\varphi_{k,m}^\infty(g_{k,m}) = \mathbf{I}_{\{g_{k,m} > \frac{\ln(2)\kappa_{3,m}}{\beta_m^{r_{\infty^*}}}\}} \left(\frac{\beta_m^{r_{\infty^*}}}{\ln(2)} - \frac{\kappa_{3,m}}{g_{k,m}} - \beta_m^{r_{\infty^*}} \log_2 \left(\frac{g_{k,m} \beta_m^{r_{\infty^*}}}{(\ln(2)\kappa_{3,m})} \right) \right)$, y por lo tanto $\frac{\partial \varphi_{k,m}^\infty(g_{k,m})}{\partial g_{k,m}} = \mathbf{I}_{\{g_{k,m} > \frac{\ln(2)\kappa_{3,m}}{\beta_m^{r_{\infty^*}}}\}} \left(\frac{\kappa_{3,m}}{g_{k,m}} - \frac{\beta_m^{r_{\infty^*}}}{\ln(2)} \right) \frac{\kappa_{3,m}}{g_{k,m}} \leq 0$, lo cual implica que $\varphi_{k,m}^\infty(g_{k,m})$ decrece monótonamente. Por consiguiente, pueden encontrarse ganancias de canal únicas $\gamma_{k,\nu}$, $\forall \nu \neq m$, tales que

$$\varphi_{k,m}^\infty(g_{k,m}) = \varphi_{k,\nu}^\infty(\gamma_{k,\nu}). \quad (4.58)$$

Claramente, $\varphi_{k,m}^\infty(g_{k,m}) \leq \varphi_{k,\nu}^\infty(g_{k,\nu})$ si $g_{k,\nu} \in [0, \gamma_{k,\nu}]$, y $\varphi_{k,m}^\infty(g_{k,m}) > \varphi_{k,\nu}^\infty(g_{k,\nu})$ si $g_{k,\nu} \in (\gamma_{k,\nu}, \infty)$. Y en consecuencia, $\Pr(\varphi_{k,m}^\infty < \varphi_{k,\nu}^\infty | g_{k,m}) = \Pr(g_{k,\nu} < \gamma_{k,\nu} | g_{k,m})$.

Si se excluye $\varphi_{k,m}^\infty(g_{k,m}) > 0$, $w_{k,m} = 0$ tal y como se discutió previamente. Resolviendo (4.58) c.r.a. $\gamma_{k,\nu}$ da lugar a

$$\gamma_{k,\nu}(g_{k,m}) = - \frac{\kappa_{3,\nu} \ln(2) / \beta_\nu^{r_{\infty^*}}}{f^W \left[-2^{-\frac{\kappa_{3,m}}{g_{k,m} \beta_\nu^{r_{\infty^*}}}} e^{-1} \left(\frac{e \ln(2) \kappa_{3,m}}{g_{k,m} \beta_m^{r_{\infty^*}}} \right)^{\frac{\beta_m^{r_{\infty^*}}}{\beta_\nu^{r_{\infty^*}}}} \right]}, \quad (4.59)$$

en donde $f^W[x] = y$ es la función de Lambert real f^W que resuelve la ecuación $ye^y = x$ para $-1 \leq y \leq 0$ y $-1/e \leq x \leq 0$ [16].

Utilizando (4.57)-(4.59), la probabilidad $\Pr(w_{k,m} = 1, g_{k,m})$ en (4.55) puede

expresarse como

$$\Pr(w_{k,m} = 1, g_{k,m}) = \mathbf{I}_{\{g_{k,m} > \frac{\ln(2)\kappa_{3,\nu}}{\beta_m^{r_{\infty}^*}}\}} \frac{e^{-g_{k,m}/\bar{g}_{k,m}}}{\bar{g}_{k,m}} \prod_{\nu=1, \nu \neq m}^M (1 - e^{-\gamma_{k,\nu}(g_{k,m})/\bar{g}_{k,\nu}}) \quad (4.60)$$

Puesto que (4.60) depende de $\{\beta_\nu^{r_{\infty}^*}\}_{\nu=1}^M$, para poder obtener los umbrales habrá que resolver la asignación óptima de potencia cuando $L_{k,m} \rightarrow \infty$. Los umbrales $\{\tau_{k,m|l}\}_{l=2}^{L_{k,m}}$ se obtienen después resolviendo (4.55) mediante una sencilla búsqueda uni-dimensional. Nótese también que se puede tomar ventaja de la condición $\varphi_{k,m}^\infty(g_{k,m}) < 0$, estableciendo $\tau_{k,m,2} \geq \kappa_{3,m} \ln(2)/\beta_m^{r_{\infty}^*}$.

En la figura 4.7 se ilustra un ejemplo del cuantificador propuesto. Para una sub-portadora k dada, se muestran en tres gráficas distintas $\Pr(w_{k,m} = 1|g_{k,m})$, $f_{g_{k,m}}(g_{k,m})$ y $\Pr(w_{k,m} = 1, g_{k,m})$ frente a $g_{k,m}/\bar{g}_{k,m}$ para $M = 6$, $L_{k,m} = 6$ y $\bar{g}_{k,m}$ y \check{r}_m iguales $\forall m$. El primer gráfico en esta figura, $\Pr(w_{k,m} = 1|g_{k,m})$, revela que cuanto mejor sea el canal, más probable es que el usuario correspondiente sea seleccionado. Si esta observación se considera conjuntamente con el comportamiento exponencial de $f_{g_{k,m}}(g_{k,m})$ plasmado en el segundo gráfico, se obtiene de manera natural la forma característica de campana de la FDP conjunta $\Pr(w_{k,m} = 1, g_{k,m})$ así como los umbrales cuantificación asociados calculados de acuerdo a (4.55) que se muestran en el tercer gráfico (tras multiplicación punto a punto de las funciones de los dos primeros gráficos).

La figura 4.8 representa $\Pr(w_{k,m} = 1, g_{k,m})$ junto con las regiones de cuantificación para dos de los usuarios de un escenario con $M=3$, $\check{\mathbf{r}} = [20, 40, 60]^T$ y $\boldsymbol{\mu} = [1, 1, 1]^T$. En particular, se consideran los usuarios $m = 1$ y $m = 3$ con $\bar{g}_{k,1} = \bar{g}_{k,3}$ y $\check{r}_1/\check{r}_3 = 1/3$. De la observación de la figura se puede comprobar cómo la probabilidad de elegir al usuario 3 es considerablemente mayor que la de elegir al usuario 1, concordando con el hecho de que el usuario 3 tenga un requisito de tasa mayor. Asimismo, los umbrales de cuantificación para el usuario 3 son más pequeños que los del usuario 1, proporcionando así al usuario 3 una ventaja para acceder al canal incluso cuando este es más pobre, mientras que el usuario 1 limita su acceso a realizaciones de canal con ganancia más favorable.

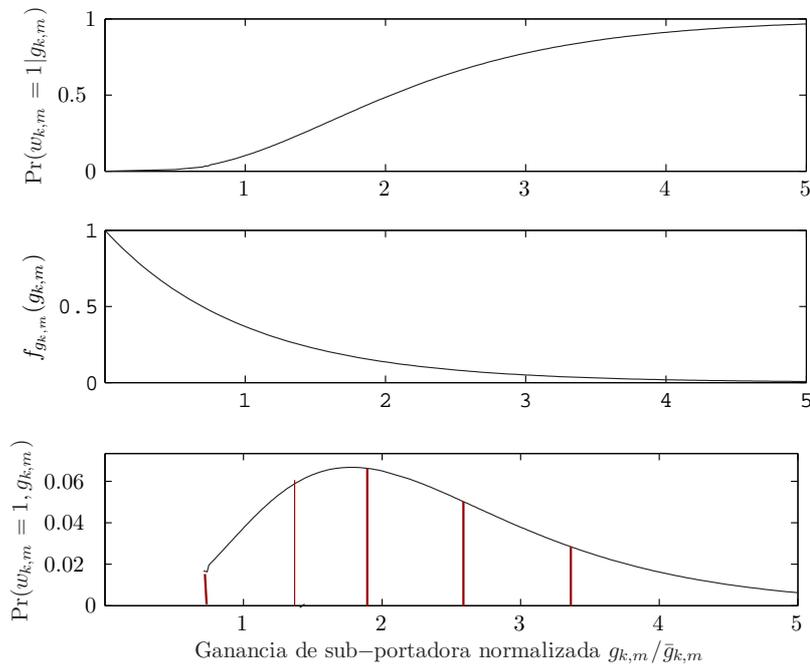


Figura 4.7: Cuantificación de las sub-portadoras basada en regiones equi-probables ($K = 64$, $M = 6$, $L = 6$).

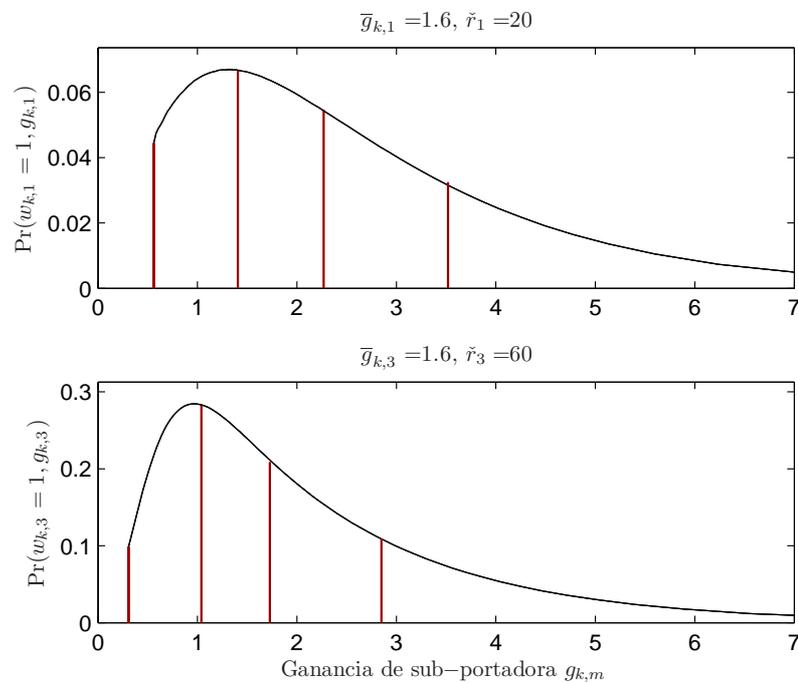


Figura 4.8: Umbrales de cuantificación para los usuarios $m = 1$ y $m = 3$ con ganancias de canal semejantes y distintos requisitos de tasa ($K = 64$, $M = 3$, $\bar{\mathbf{r}}_0 = [20, 40, 60]^T$).

Para concluir esta apartado, se resumen los aspectos de complejidad que conllevan la asignación de recursos, (OB.2); así como el diseño del cuantificador, (OB.1). Por un lado, las versiones modificadas de los algoritmos del apartado 4.2.2 que consideran (4.52) y (4.53) resuelven los problemas (OB.2)/(OB.1) de forma iterativa. Por otro lado, con el método de diseño del cuantificador basado en (4.55), el cálculo de $\tau_{k,m|l}$ debe llevarse a cabo únicamente una vez (resolviendo (OB.1)); con estos umbrales disponibles, la asignación de recursos puede obtenerse con facilidad a través de la versión inicial del algoritmo 4.3, de menor complejidad, descrito en el apartado 4.2.2 (resolviendo (OB.2)). Los resultados numéricos mostrarán que este diseño, de menor complejidad y no iterativo, tiene un consumo de potencia similar a aquel que exhibe el algoritmo óptimo, que resuelve conjunta e iterativamente el problema de asignación de recursos y cuantificación.

4.2.4. Simulaciones numéricas

Para caracterizar de forma numérica la eficiencia en potencia de los diseños aquí propuestos, se considerará un sistema OFDMA adaptativo con $M = 3$ usuarios con prioridades semejantes ($\mu_m = 1 \forall m$), $K = 64$ sub-portadoras, potencia de ruido por usuario y sub-portadora de 0 dB_W , $L_{k,m} = 5$ regiones (es decir, 4 regiones activas) por sub-portadora, $\check{\epsilon}_m = \bar{\epsilon}_0 = 10^{-3} \forall m$ y dos posibles requisitos para la tasa: $\check{\mathbf{r}} = [60, 60, 60]^T$ y $\check{\mathbf{r}} = [20, 40, 60]^T$. El canal simulado consistirá en tres rayos Rayleigh incorrelados para cada usuario con una relación señal a ruido media ($\overline{\text{SNR}}$) de 0 dB , los valores de las ganancias de los tres usuarios para cada una de las sub-portadoras correspondientes a este escenario se representan en la figura 4.9.

Caso de estudio 1 (*Comparación de los esquemas de asignación*): Para distintos valores de la $\overline{\text{SNR}}$ e igual valor para la restricción sobre la tasa media de los usuarios, la figura 4.10 compara la potencia media *total* de cuatro esquemas de asignación de potencia media en transmisión basados en: (i) el óptimo representado por P-CSIT, (ii) el algoritmo 4.2 con cuantificación óptima iterativa (que

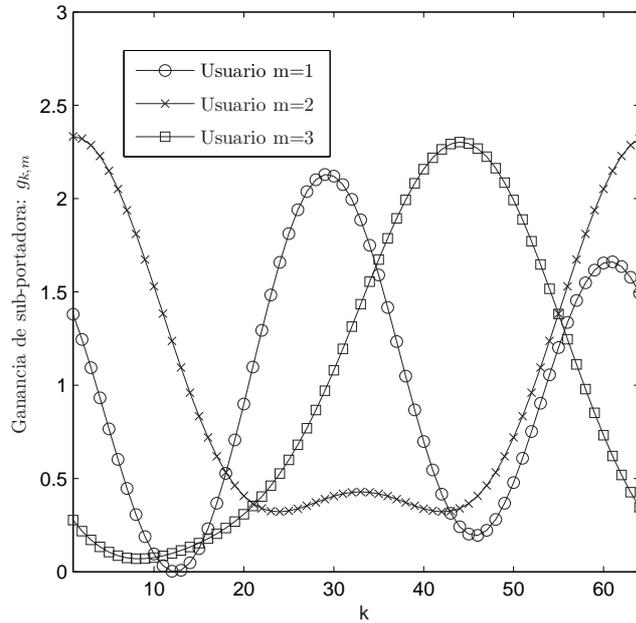


Figura 4.9: Ganancias de sub-portadora para distintos usuarios ($K = 64$, $M = 3$, $\overline{\text{SNR}} = 0 \text{ dB}$).

se denotará como Q-CSIT1), (iii) el algoritmo 4.3 (usando \bar{P}^{++}) y regiones de cuantificación dadas por (4.55) (que se denotará como Q-CSIT2) y (iv) el algoritmo 4.3 (empleando \bar{P}^+) y regiones de cuantificación dadas por (4.55) (que se denotará como Q-CSIT3). Se puede observar que las prestaciones de los esquemas Q-CSIT1, Q-CSIT2 y P-CSIT son similares y que la ganancia c.r.a. esquema Q-CSIT3 es relativamente pequeña (menos de 2 dB). Para explicar la apenas perceptible diferencia entre Q-CSIT1 y Q-CSIT2 debe repararse en que el pequeño margen entre Q-CSIT2 y Q-CSIT3 significa que la cota superior de transmisión de potencia en (4.45) es bastante ajustada. Esto implica que cuando la cota en (4.45) se utiliza sólo para asignación de recursos, la asignación resulta muy parecida a la óptimo y, por lo tanto, al realizar la carga de potencia final de forma exacta, el consumo de potencia resultante es similar. Además, estos resultados sirven también para validar el sencillo diseño del cuantificador basado en regiones equiprobables para cada portadora (recuérdese que los resultados en el apartado 3.2.7 ya apuntaron que para el caso de OFDM con un único usuario un cuantificador

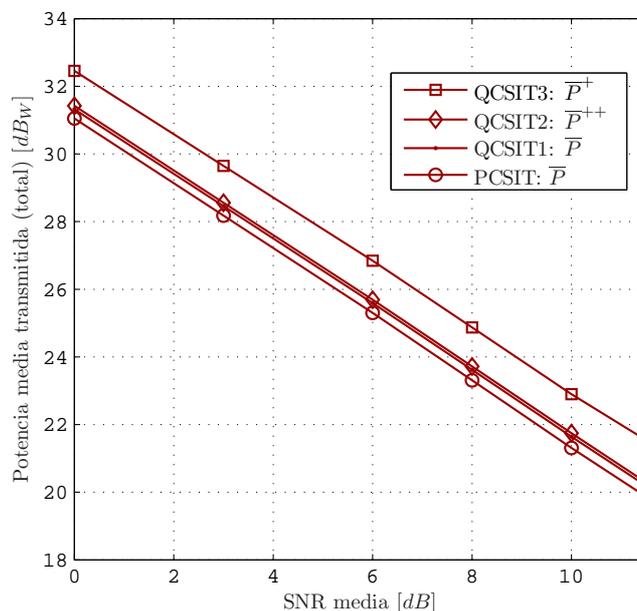


Figura 4.10: Comparación de la potencia media transmitida (total) para diferentes esquemas de adaptación ($K = 64$, $M = 3$, $\bar{\mathbf{r}} = [60, 60, 60]^T$).

basado en el mismo principio presentaba prestaciones cercanas a las del óptimo). En lo que a la complejidad se refiere, los esquemas Q-CSIT2 y Q-CSIT3 balancean de forma óptima su complejidad y prestaciones c.r.a. las de sus competidores. En concreto, Q-CSIT2 da lugar a prestaciones similares a las del esquema P-CSIT, a pesar de que, de acuerdo a lo mostrado en (4.50), sólo conlleva búsquedas unidimensionales. Por otro lado, Q-CSIT3 da lugar a expresiones cerradas y sólo sufre pérdidas de 2 dB en comparación con el esquema P-CSIT. Debido a la simplicidad y casi-optimalidad del consumo de potencia del esquema Q-CSIT2, éste será el esquema que analice en los siguientes casos de estudio.

Para el caso en el que los usuarios presentan distintos requisitos de tasa, la figura 4.11 representa la potencia media transmitida por cada usuario como función de la SNR recibida. Los resultados de la figura 4.11 validan y corroboran los análisis realizados en relación a la figura 4.10.

Caso de estudio 2 (*Comparación de Q-CSIT2 con otros esquemas alternativos basados en Q-CSIT*): En este caso se comparan las prestaciones del esquema Q-

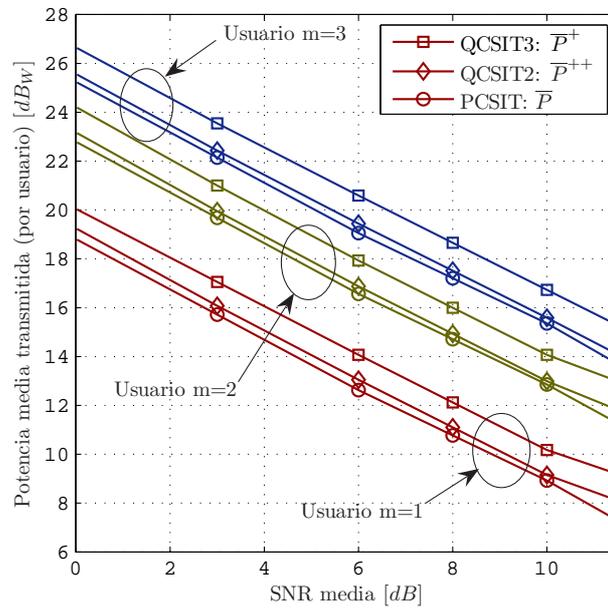


Figura 4.11: Comparación de la potencia media transmitida (por cada usuario) para diferentes esquemas de adaptación ($K = 64$, $M = 3$, $\bar{\mathbf{r}} = [20, 40, 60]^T$).

CSIT2 con las de otros esquemas alternativos basados en Q-CSIT. Estos esquemas incluyen: (i) asignación de recursos óptima con umbrales de cuantificación tomados de forma arbitraria (que se denotará como Q-CSIT4), (ii) asignación óptima de tasa y potencia cuando las sub-portadoras se asignan de forma fija (que se denotará como Q-CSIT5) y (iii) Q-CSIT5 con un único bit de realimentación (que se denotará como Q-CSIT6). La figura 4.12 muestra la pérdida relativa de potencia de estos esquemas con respecto al esquema P-CSIT. Como se puede ver, mientras que Q-CSIT2 alcanza prestaciones similares a P-CSIT, como se afirmó con anterioridad, los esquemas Q-CSIT4-6 muestran, respectivamente, un consumo adicional de potencia de más de 4, 7 y 11 dB.

Caso de estudio 3 (*Variaciones del escenario de simulación*): En la tabla 4.4 se muestran valores numéricos que comparan las prestaciones de P-CSIT, Q-CSIT2 y Q-CSIT5 variando (uno por uno) los valores un amplio rango de parámetros. La observación más relevante es que los resultados numéricos respaldan las conclusiones previas, a saber: (i) la casi-optimalidad de Q-CSIT2 y (ii) la pérdida de

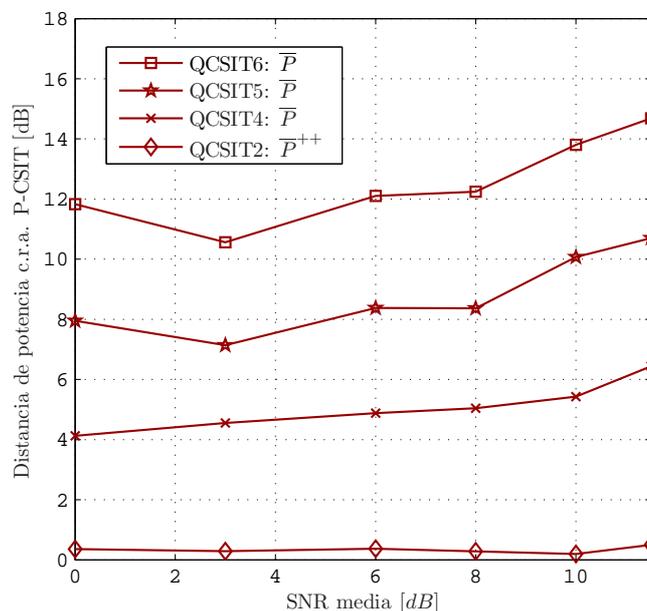


Figura 4.12: Diferencia de la potencia media transmitida (total) por los diferentes esquemas c.r.a. la adaptación basada en P-CSIT ($K = 64$, $M = 3$, $\bar{\mathbf{r}} = [20, 40, 60]^T$).

prestaciones que exhiben los ejemplos heurísticos como el Q-CSIT5.

Caso de estudio 4 (*Número de regiones de cuantificación*): Finalmente, la figura 4.13 muestra la potencia media transmitida frente al número de regiones activas por cada sub-portadora para tres usuarios con distintos requisitos en términos de tasa media. Se recuerda que el número de regiones activas es igual a $L_{k,m} - 1$; e.g., $L_{k,m} = 2$ implica una única región activa y una región en la que se renuncia a la transmisión. Los resultados de la simulación mostrados en esta figura demuestran

Tabla 4.4: Potencia total transmitida (dB_W) para los esquemas P-CSIT, Q-CSIT2 y Q-CSIT5. (Caso de referencia: $K = 64$, $M = 3$, $\bar{\mathbf{r}} = [60, 60, 60]^T$, $\overline{\text{SNR}} = 0$ dB; otros *casos* implica una única variación c.r.a. caso de referencia.)

Caso	Q-CSIT5: \bar{P}	Q-CSIT2: \bar{P}^{++}	P-CSIT: \bar{P}
Caso de referencia	38.9	31.6	31.2
$\check{\epsilon} = 10^{-4}$	40.8	33.3	32.8
$\bar{\mathbf{r}} = [30, 30, 30]^T$	31.4	26.7	26.3
$K = 128$	34.8	30.1	29.2
$M = 6$	46.8	39.9	39.2

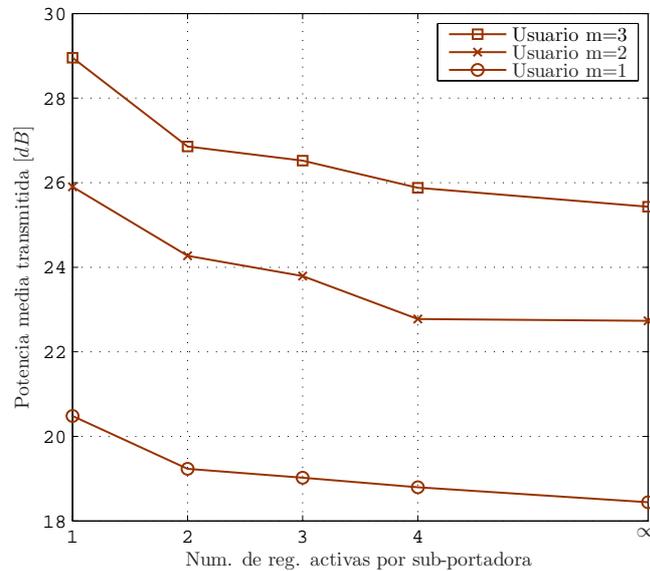


Figura 4.13: Variación del número de regiones de cuantificación (bits realimentados) por sub-portadora ($K = 64$, $M = 3$, $\mathbf{\check{r}} = [20, 40, 60]^T$).

que la optimización conjunta de asignación de recursos y regiones de cuantificación da lugar a una pérdida de potencia no mayor a 3-5 dB con respecto al caso P-CSIT ($L_{k,m} = \infty$). Como es natural, la diferencia se reduce a medida que el número de regiones aumenta (siendo la reducción asociada al incremento de una nueva región cada vez menor), alcanzando una pérdida de potencia de aproximadamente sólo 0.5 dB para el caso de cuatro regiones activas.

4.2.5. Síntesis de resultados

En esta sección se han derivado esquemas de cuantificación de canal y de adaptación de los transmisores para sistemas OFDMA que, en base a Q-CSIT, minimizan la potencia transmitida satisfaciendo requisitos individuales de tasa y BER media. Además de las soluciones óptimas, para impulsar su posible implementación práctica, se han desarrollado diseños sub-óptimos de menor complejidad. La versión más compleja requiere la optimización conjunta de potencia, tasa y sub-portadoras para cada usuario además de las regiones de cuantificación. La solución propuesta para reducir la complejidad se basa en separar la asignación de recursos

del diseño del cuantificador. Dado el diseño de un cuantificador, el problema de asignación de recursos se ha simplificado todavía más mediante la consideración de una cota superior para la potencia asignada. Mientras que esta cota superior ha facilitado el proceso de optimización y la obtención de esquemas de asignación de recursos en forma cerrada, los valores finales de potencia transmitida se han calculado de forma exacta. El resultado de este proceso es un sencillo esquema (protocolo) de asignación de recursos en el que tanto los posibles valores de tasa y potencia a transmitir por un usuario están disponibles en el transmisor (y en el punto de acceso) almacenadas en una tabla y sólo la asignación de sub-portadoras y el índice que indexa los valores de la tabla a utilizar deben ser calculados de forma “on-line”. Con respecto al diseño “off-line” del cuantificador, además de haber propuesto un diseño iterativo óptimo y de haberlo relacionado con el cuantificador de Lloyd, se ha derivado un esquema simple de cuantificación que, en base a la solución P-CSIT, calcula los umbrales de cuantificación de forma que para un usuario y una sub-portadora dados genera regiones de cuantificación que resultan en un acceso al canal equi-probable para cada una de las regiones diseñadas.

Finalmente, existen distintas direcciones de interés para expandir el trabajo aquí presentado, destacando la investigación respecto a la reducción de la cantidad de información a realimentar aprovechando la correlación entre sub-portadoras. Dos posibles vías para llevar a cabo esta reducción de forma sencilla consistirían en agruparlas e indexar de forma común cada uno de los grupos o mediante la inclusión de técnicas de cuantificación siguiendo la línea de [37].

4.3. Consideraciones finales

Antes de cerrar este capítulo deben realizarse una serie de consideraciones que ayudarán a juzgar de forma más adecuada los resultados obtenidos.

En primer lugar, se ha reafirmado el comportamiento de los sistemas de adaptación en términos de consumo observado en el capítulo anterior. Los esquemas

propuestos basados en Q-CSIT presentan un consumo muy cercano al de la adaptación óptima basada en P-CSIT y sustancialmente menor que el de otras alternativas existentes.

Asimismo la formulación del problema ha jugado un importante papel para la obtención de la solución. Especialmente importantes han sido en esta ocasión la consideración de QoS promedio y la propuesta de esquemas de codificación de canal independiente por cada usuario.

Los mecanismos óptimos de adaptación para los sistemas multi-usuario han resultado ser sencillos esquemas de acceso oportunistas donde de forma intuitiva se tiene en cuenta el coste neto del acceso de cada usuario al canal. Además, por construcción, la solución óptima considera equidad y justicia social en la asignación de recursos, respetando las demandas individuales de cada uno de ellos así como las potenciales diferencias de prioridad.

Interesante resulta asimismo la vinculación de los esquemas óptimos de cuantificación con el algoritmo de Lloyd. Habiéndose comprobado cómo si se define como función de distorsión el Lagrangiano del problema de minimización, el diseño del cuantificador puede realizarse en base las reglas iterativas básicas de *condición de centroide* y *vecino más próximo*.

Pese a que los estudios realizados se han centrado en el análisis del envío de información desde los usuarios al punto de acceso (enlace ascendente o problema de acceso), la extensión de los resultados al envío de información desde el punto de acceso a los usuarios (enlace descendente o problema de difusión) resulta también de interés. Las diferencias principales residen en que: (i) en el problema de difusión la restricción de tasa suele una restricción común sobre la suma total de tasas individuales (lo que implicaría que el conjunto de M multiplicadores asociado a las restricciones individuales de tasa se convirtiera en un único multiplicador), (ii) la cuantificación del canal debe ser realizada en este caso por lo usuarios, que únicamente conocen su canal, pero no el canal de sus vecinos (esto provocará que en la evaluación del coste neto no aparezca el canal sino el índice de la región) y (iii) la tasa de realimentación requerida será mayor, puesto que el punto de

acceso precisará recopilar la información de todos los usuarios (una configuración alternativa que reduciría la realimentación necesaria, especialmente si los usuarios están próximos entre sí pero lejos del punto de acceso podría consistir en que fueran los propios usuarios los que, intercambiando información, realizaran la decisión óptima de acceso y se la notificasen al punto de acceso). No obstante, teniendo en cuenta las anteriores diferencias, la estructura de cuantificación y adaptación sería por lo demás similar a la estudiada, obteniéndose una asignación de recursos íntimamente relacionada con la dictada por el algoritmo 4.3.

4.4. Apéndice

4.4.1. Demostración de las ecuaciones (4.21) y (4.22)

Primero se demostrará que $\chi_m^{Q_p^*} \neq 0, \forall m$. Para intentar llegar a una contradicción, se supone que $\chi_m^{Q_p^*} = 0$ para un m dado y se deduce que [cf. (4.20)]

$$\mu_m \int_{\mathcal{R}_{m,l}} w_m(\mathbf{h}) dF_{\mathbf{g}}(\mathbf{g}) = \alpha_{m,l}^{Q_p^*}, \quad l \in \{1, \dots, L_m\}. \quad (4.61)$$

Puesto que $\forall l, \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) > 0$, se tiene que $\forall l, \alpha_{m,l}^{Q_p^*} > 0$. Utilizando la propiedad de *debilidad complementaria*¹⁶ para $\alpha_{m,l}^{Q_p^*}$ y la restricción $p_{m,l}^* \geq 0$, se tiene que $p_{m,l}^* = 0, \forall l$. La BER media correspondiente al usuario m resulta entonces 0.5 lo que está en clara contradicción con la satisfacción del requisito $\check{\epsilon}_m$. Por lo tanto, se tiene que $\chi_m^{Q_p^*} > 0 \forall m$. Dirigiendo ahora la atención a (4.20) y utilizando de nuevo la propiedad de debilidad complementaria, se tiene que bien $p_{m,l}^* = 0$, o bien, si $p_{m,l}^* > 0$ (y por lo tanto $\alpha_{m,l}^{Q_p^*} = 0$),

$$\mu_m \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) dF_{\mathbf{g}}(\mathbf{g}) + \chi_m^{Q_p^*} r_{m,l} / \check{r}_m \int_{\mathcal{R}_{m,l}} w_m(\mathbf{g}) \epsilon'_{m,l}(g_m p_{m,l}^*) dF_{\mathbf{g}}(\mathbf{g}) = 0, \quad (4.62)$$

¹⁶El multiplicador de Lagrange asociado a una restricción es mayor que cero si la restricción se satisface con igualdad y cero en caso contrario [13]

que conduce directamente a (4.21). Puesto que $\chi_m^{Q_p^*} > 0, \forall m$, la debilidad complementaria entre $\chi_m^{Q_p^*}$ y las restricciones de BER implican que el valor medio de BER satisface la restricción de forma ajustada (i.e., como una igualdad) tal y como se escribe en (4.22).

Capítulo 5

Conclusiones y líneas futuras de investigación

5.1. Conclusiones

Para sistemas de comunicaciones operando sobre canales inalámbricos con desvanecimientos se han derivado esquemas de adaptación de los transmisores que, haciendo uso del *conocimiento del estado del canal* (CSI), minimizan la potencia transmitida satisfaciendo requisitos de tasa de transmisión y probabilidad de error recibida. El uso del CSI por parte del transmisor en cualquiera de sus vertientes (perfecto, cuantificado, individual, estadístico) mejora sustancialmente las prestaciones del sistema, la contribución de esta tesis ha consistido en diseñar y utilizar diversos tipos de CSI en distintos escenarios de comunicación. El mayor énfasis investigador se ha dedicado a sistemas que disponen de un canal de realimentación (control) con tasa limitada a través del cual el receptor envía a los transmisores CSI. La limitación natural de la tasa del canal de realimentación provoca que el CSI tenga que ser descrito por un número finito de bits y que, por lo tanto, los *transmisores* sólo puedan adquirir *CSI cuantificado* (Q-CSIT). El diseño y uso de Q-CSIT en este escenario requiere la resolución de dos problemas relacionados: (i) diseñar un cuantificador de canal que para cada estado del canal codifique el

Q-CSIT correspondiente y (ii) diseñar esquemas de adaptación de los terminales al Q-CSIT generado.

A lo largo de las secciones de esta tesis se ha abordado el diseño óptimo de los esquemas de cuantificación y adaptación para distintos escenarios: (i) sistemas de un único usuario con canales selectivos en tiempo y en frecuencia, (ii) redes de sensores donde los transmisores distribuidos se comunican de forma coherente con un centro de fusión de datos, (iii) sistemas de acceso múltiple donde los usuarios transmiten a través de canales selectivos en el tiempo y (iv) sistemas donde el acceso múltiple se realiza por división ortogonal en frecuencia y se transmite a través de canales selectivos en tiempo y frecuencia. La amplia gama de escenarios analizados ha permitido que, dentro de un mismo marco teórico (el de la optimización de sistemas en base a Q-CSIT) y con un mismo objetivo (la minimización de potencia garantizando *calidad de servicio* –QoS–), se hayan planteado distintas formulaciones del problema, analizado la relevancia de las suposiciones de trabajo básicas y sopesado las distintas alternativas de diseño, caracterizando siempre su comportamiento a través del análisis de las prestaciones obtenidas. La *intención* perseguida mediante esta metodología era *doble*: por un lado se pretendía proporcionar soluciones para problemas significativos (i.e., escenarios reales de comunicación) de manera que los diseños derivados tuvieran una *aplicación práctica*; por otro lado se pretendía adquirir una visión global sobre la utilidad, prestaciones y *aspectos clave del diseño* de este tipo de sistemas.

El marco teórico utilizado para la derivación de los esquemas de adaptación ha sido el de la optimización no lineal. En general se han formulado problemas de minimización (convexa o no) con restricciones. En la fase inicial de análisis se ha procedido a caracterizar de *forma analítica* la solución óptima y, allá donde el análisis matemático se detenía, se pasaba a una segunda fase en la se aprovechaban la caracterización de la solución y las herramientas numéricas de optimización para encontrar *algoritmos eficientes* que permitieran alcanzar el valor final de las variables a optimizar. De esta manera no sólo se ha facilitado la convergencia más rápida y eficiente de los algoritmos numéricos que generan los valores óptimos de

los parámetros del sistema (uno de los objetivos principales), sino que también se ha logrado adquirir entendimiento e intuición analítica respecto a los principios subyacentes que rigen el comportamiento de estos sistemas. Finalmente, debe señalarse que los recientes y dramáticos avances en el campo de la optimización no lineal (especialmente en su vertiente convexa) han tenido también un reflejo en esta tesis. De un lado en los métodos de búsqueda numérica de los parámetros de diseño (sin olvidar que la exhaustividad en este campo no era ese el objetivo de esta tesis); de otro en la propia formulación de los problemas que, siguiendo el principio de *cross-layer design*, se han planteado y resuelto sin simplificaciones mayúsculas en los que a su modelado teórico se refiere.

De acuerdo con la metodología y objetivos propuestos, se recopilan a continuación conclusiones relativas al diseño y uso de la *cuantificación del estado de canal para la minimización de la potencia en sistemas con transmisores adaptativos* derivadas de la realización de esta tesis doctoral.

Respecto al modo de operación

El hecho de que los esquemas de cuantificación y adaptación se obtengan como la solución de un problema de minimización con restricciones hace que el modo de operación del sistema quede determinado por la propia estructura de la solución del problema de minimización. De esta forma la operación del sistema se divide en dos fases:

- Una fase “off-line” que se realiza durante la inicialización del sistema y en la que se calculan los parámetros *estáticos* que intervienen en la solución óptima (diseño del cuantificador de canal, valores óptimos de los parámetros de transmisión para cada una de las regiones) en función de los requisitos de QoS y el comportamiento estadístico del canal. Esta fase no debe volver a calcularse a menos que las estadísticas del canal (i.e. comportamiento a largo plazo) o estructura del mismo (cambio de las restricciones del problema) se modifiquen.

- Una fase “on-line”, que se lleva a cabo de forma simultánea al proceso de comunicación de la información, en la que para cada realización del canal se decide la configuración básica que el transmisor debe utilizar (potencia, tasa, codificación, sub-portadora,...). Puesto que esta fase se realiza en tiempo real, lleva asociada una carga computacional muy ligera que, dependiendo de los casos, puede tratarse de la indexación de unos valores predeterminados de una tabla (juego de modos), la evaluación sencilla de una función (determinación el usuario que accede de forma óptima al canal) o de algún algoritmo de pequeña complejidad (asignación de tasa y potencia en un sistema OFDM con QoS instantánea).

Se ha comprobado también cómo el diseño de ambas fases está relacionado y cómo, aunque ejecutadas en tiempos (momentos) diferentes, el lugar de ejecución sí debe ser compartido. De forma más específica, la solución de la fase “off-line” tiene que ser en general conocida tanto por el receptor (que es el que cuantifica el canal) como por el transmisor (que es el que adapta su comportamiento de acuerdo a los valores cuantificados de potencia y tasa o bien los calcula en función de la estructura del cuantificador). Desde un punto de vista práctico, para los escenarios estudiados, la alternativa más factible es que sea el propio receptor el que calcule la solución y la notifique al transmisor(es) durante la fase de inicialización. No obstante, sobre el papel el algoritmo de optimización podría ejecutarse en el transmisor y notificar este resultado al receptor, podría ejecutarse en ambos transceptores, o podría ejecutarlo un tercer actor (e.g. centro de recogida de datos de la red) y notificar después la solución al transmisor y al receptor. Por otro lado la fase “on-line” debe ejecutarse en ambos extremos de la cadena de comunicación: el receptor, conociendo de forma perfecta el canal, debe generar el Q-CSIT de acuerdo al esquema derivado, mientras que el transmisor, en función del Q-CSIT recibido proveniente del receptor adaptará su configuración de transmisión.

Finalmente, el análisis de optimización realizado, también ha permitido llegar a conclusiones relevantes con respecto a la (sub)optimalidad de los modos de

operación del sistema. Un primer hecho significativo es que en todos los escenarios de trabajo sólo se ha podido garantizar optimalidad de carácter local para el cuantificador. Este hecho se debe a que la definición del cuantificador interviene típicamente en la limitación de regiones de integración lo que en general da lugar a problemas que no son convexos. Se ha relacionado además la estructura de los cuantificadores propuestos con la de los cuantificadores vectoriales en general y con la cuantificación vectorial con restricciones en particular y se ha destacado que esta falta de convexidad es una característica compartida en la mayoría de problemas de cuantificación. A diferencia del problema anterior, para todos los casos analizados la asignación óptima de recursos (o, equivalentemente, la adaptación óptima del canal), que se calcula en base a parámetros de la fase “off-line” y a la realización actual del canal, tiene carácter de óptimo global. Dos consecuencias directas de estos hechos son que, para todos los casos: (i) no se ha podido garantizar la optimalidad global del diseño conjunto del sistema y (ii) si el cuantificador de canal está dado o diseñado a priori, los esquemas de asignación de recursos presentan carácter óptimo.

Respecto a la complejidad

Algunos de los aspectos relacionados con la complejidad que han quedado de manifiesto tras realizar esta tesis han sido ya objeto de comentario en esta sección. No obstante, dos son los aspectos que, aunque de forma breve, merecen un mayor detalle:

- La *complejidad computacional* se asocia principalmente a los algoritmos de optimización y depende radicalmente de la fase de ejecución considerada. Se ha comprobado cómo el diseño de los parámetros asociados a la fase “off-line” (cuantificador) conlleva una complejidad computacional media-alta (el nivel específico depende del escenario en cuestión y de la sofisticación de las técnicas de optimización y búsqueda numérica realizada), mientras que los esquemas de adaptación en el transmisor y de asignación de recursos

tienen asociada una complejidad computacional despreciable. Asimismo, se ha comprobado cómo la formulación del problema tiene un fuerte impacto en la complejidad de la solución final y cómo, en general, la exigencia de niveles de QoS en términos promedio reduce la complejidad computacional requerida con respecto a la satisfacción de QoS en términos instantáneos. Finalmente, para aquellos sistemas en los que la complejidad computacional sea un factor limitador, se han desarrollado en varios casos alternativas de complejidad computacional reducida tanto para la fase de asignación de recursos como especialmente para el diseño del cuantificador.

- La *complejidad hardware* requerida para que los transceptores pueden implementar las técnicas propuestas, sin haber sido una variable a optimizar, ha tenido una significativa influencia en las decisiones de diseño. De esta manera, con el objetivo de obtener unos esquemas de adaptación de complejidad moderada: (i) los mecanismos de adaptación en los transmisores propuestos *a priori* son de sencilla implementación (número finito de modos AMC, modulaciones QAM), donde en general la potencia es el único parámetro del transmisor cuya adaptación es libre; (ii) la implementación física de los cuantificadores propuestos, aunque sofisticada en su estructura, no es complicada (por lo general cuantificación escalar sobre las ganancias del canal) y (iii) los esquemas de adaptación derivados no requieren una complejidad elevada (pequeñas tasas de realimentación, asignación de usuarios oportunista). Todos estos aspectos contribuyen a facilitar la implantación de este tipo de técnicas en sistemas reales que típicamente no pueden tolerar los diseños óptimos de otro tipo de esquemas (códigos de longitud infinita para alcanzar capacidad, decodificación iterativa entre usuarios, etc.).

Respecto a las prestaciones

El objetivo nominal de esta tesis era la minimización de potencia satisfaciendo niveles de QoS, por lo que al analizar sus prestaciones deberán tenerse en cuenta

dos aspectos: (i) si efectivamente se han satisfecho los niveles de QoS exigidos y (ii) hasta qué punto se ha podido reducir el nivel de potencia. La respuesta a la *primera de las cuestiones* es sencilla, puesto que, al resolver el diseño a través de técnicas de optimización, para todos los casos el problema considerado ha sido factible y las restricciones de QoS se han satisfecho de forma ajustada. La respuesta a la *segunda de las cuestiones* debe realizarse comparando el consumo de potencia obtenido para los esquemas propuestos con los obtenidos por otros esquemas. Siguiendo esta metodología puede afirmarse que de la realización de esta tesis puede concluirse que el consumo de los esquemas de cuantificación y adaptación al Q-CSIT propuestos:

- Presenta una diferencia de consumo pequeña con respecto al consumo asociado a los esquemas en base a CSIT *perfecto* (P-CSIT).
- Presenta una ventaja significativa con respecto a otros esquemas Q-CSIT heurísticos que si bien cuantifican el canal, no abordan el problema de diseño óptimo ni explotan adecuadamente las características del canal con desvanecimientos.
- Presenta una diferencia de al menos un orden de magnitud con respecto a esquemas que hacen uso de conocimiento *estadístico* del canal (S-CSIT).
- Presenta una diferencia de consumo de varios órdenes de magnitud con respecto a sistemas que no aprovechan el CSIT. La causa fundamental de esta diferencia se debe a la dificultad en explotar la diversidad del canal y, por consiguiente, a la imposibilidad de reacción frente a desvanecimientos severos.

Como es natural el valor concreto de la distancia entre unos y otros esquemas dependerá del escenario de aplicación específico, lo que hace que las diferencias cuantitativas, si bien consistentes, deban tomarse con cierto nivel de precaución. Incluso así, el hecho de que los comportamientos anteriores se observen de forma reiterada sí permite concluir la presencia de una diferencia cualitativa entre las

prestaciones de los esquemas basados en Q-CSIT propuestos y aquellos con los que se compara.

Finalmente, pese a que no se haya podido demostrar la optimalidad global de los diseños propuestos para el diseño y utilización del Q-CSIT (es decir, que no se haya podido demostrar que los esquemas derivados sean los mejores esquemas Q-CSIT posibles), el hecho de que el consumo de potencia obtenido se encuentre tan próximo al de los esquemas basados en P-CSIT (que claramente son una cota inferior para el consumo de cualquier esquema basado en Q-CSIT), hace que no pueda descartarse la posibilidad de que los esquemas propuestos sean de facto óptimos (casi-)globales y, lo que es más importante, certifica el valor de la metodología propuesta para la obtención de los esquemas de cuantificación y adaptación.

Respecto a las guías de diseño

El interés por conocer los principios que rigen el comportamiento de sistemas adaptativos en base a Q-CSIT para utilizarlos en su diseño puede darse en dos tipos de situaciones: (i) aquellas en la que, de forma similar a lo realizado en esta disertación, se quiera derivar de forma teórica esquemas de cuantificación y adaptación en base Q-CSIT para escenarios o aplicaciones distintas a las consideradas y (ii) aquellas en las que el interés se focalice en el diseño físico de sistemas que puedan utilizar de forma práctica este tipo de técnicas. A continuación se sintetizan los resultados principales que podrían resultar de interés para estas situaciones.

La asignación óptima de recursos basada en Q-CSIT sigue principios intuitivos que pueden aprovecharse para su diseño. Los ejemplos que corroboran esta afirmación han sido variados: el uso de modos de transmisión nulos que permite desechar aquellas realizaciones del canal que presentan un desvanecimiento profundo, la dependencia de los esquemas de adaptación con respecto a los requisitos específicos QoS (utilizará peores valores de ganancia si los requisitos de tasa son elevados, desechará modos de transmisión de tasa elevada si el requisito de BER es estricto) o la evaluación ponderada de los distintos parámetros de transmisión

(potencia, tasa y BER) en la asignación de acceso de los usuarios al canal.

La formulación adecuada de los problemas de optimización contribuye a la obtención de soluciones de mayor calidad. Especialmente adecuada para este problema es la formulación con requisitos promedios de QoS que: permite adaptaciones “on-line” de complejidad despreciable, facilita el diseño del cuantificador y presenta un nivel de consumo de potencia menor que el de su contrapartida instantánea (puesto que estos últimos no tienen capacidad de explotar la diversidad temporal ofertada por el canal). Igualmente importante es la elección del formato del cuantificador, habiéndose comprobado cómo, por ejemplo, para los casos de múltiples usuarios con flujos de información individuales la elección de un cuantificador que desacopla los usuarios simplifica de forma considerable no sólo los valores óptimos del propio cuantificador sino la propia asignación de recursos.

La mayor complejidad computacional para el diseño del sistema se asocia a la resolución del codificador óptimo. No obstante, la aplicación de diseños de cuantificadores sub-óptimos *razonables* (i.e., que consideren la estructura interna del problema y/o que optimicen prestaciones secundarias relacionadas con la función objetivo), seguida de una adaptación de recursos que tenga en cuenta la forma específica del cuantificador, reduce de forma contundente la complejidad computacional del diseño y supone una pérdida pequeña en términos de consumo final de potencia.

Las tasas de realimentación requeridas para notificar el Q-CSIT no son por lo general elevadas. Se ha comprobado además como aun cuando la descripción del Q-CSIT no es precisa, las prestaciones del sistema no sufren en exceso, mientras que cuando se utilizan otros tipos de CSIT que, si bien precisos, pierden su carácter instantáneo (i.e., no se actualizan por cada realización del canal) las prestaciones del sistema sí se ven seriamente afectadas. Esto lleva a pensar que ante sistemas con tasa de realimentación fija, resulta preferible actualizar frecuentemente el Q-CSIT aunque este sea de peor calidad que aumentar el tiempo entre actualizaciones a fin de enviar una versión de Q-CSIT más precisa.

Por otro lado, no se puede dejar de hacer mención a ciertas *particularidades*

negativas que dependiendo del escenario de aplicación pueden estar *asociadas al uso Q-CSIT*.

- Desde un punto de vista de las prestaciones de consumo y de sus diferencias con respecto a las de sistemas que utilizan otro tipo de CSIT, deben considerarse dos aspectos principales. En primer lugar, respecto a las ventajas de los sistemas de Q-CSIT frente a los basados en S-CSIT, debe remarcar-se el hecho de que se han considerado canales con valores de SNR recibida medios-bajos (que es en los que la minimización de la potencia está más justificada), sin embargo, según se incrementa la SNR del sistema la asignación de recursos (especialmente de la potencia) tiende a ser más uniforme y las ventajas de la adaptación instantánea frente a los esquemas que hacen uso de conocimiento estadístico se reducen (intuitivamente los desvanecimientos profundos del canal tienden a desaparecer y lo que hace que la pérdida de capacidad de reacción frente a ellos no sea tan relevante). En segundo lugar, podrían encontrarse casos en los que la cercanía del consumo entre los esquemas P-CSIT y Q-CSIT no fuera tan significativa. De forma intuitiva este divergencia se justificaría en el hecho de que aquí el CSIT se ha utilizado para gestionar de forma más eficiente los recursos, o alternativamente, para el *beneficio propio del transmisor*, mientras que existen otros sistemas de comunicación (e.g., un sistema de acceso MIMO con múltiples usuarios) donde más importante que la gestión del CSIT en beneficio propio es la gestión el CSIT que *evite el daño ajeno* (en el ejemplo mencionado, cada transmisor no transmite en el sub-espacio que maximice su señal sino que debe transmitir en un sub-espacio de canal ortogonal a los sub-espacios del resto de transmisores para evitar así la interferencia). En este tipo de escenarios, pequeños errores en el P-CSIT sí pueden resultar en grave perjuicio de las prestaciones teóricas del sistema.
- Desde el punto de vista de la complejidad computacional de los esquemas de cuantificación, debe tenerse en cuenta que existen escenarios en los que

una variabilidad excesiva de las condiciones de diseño del sistema (cambios permanentes en los niveles de QoS, entrada y salida continua de nuevos usuarios) podría requerir el cálculo de la fase “off-line” de forma continuada, lo que desde el punto de vista de complejidad computacional puede resultar una desventaja (nótese que aunque esta variabilidad excesiva exija también volver a calcular las asignaciones en base a P-CSIT o S-CSIT, la complejidad computacional de sus soluciones es menor).

Finalmente, se cerrará esta sección con una reflexión sobre el diseño y uso del Q-CSIT en sistemas de comunicaciones. Láncese la mirada atrás y retómese el análisis de algunos de los escenarios estudiados. Para el caso de un único usuario, el valor del canal que se ha cuantificado no ha sido el del coeficiente complejo de desvanecimiento, sino el de su ganancia de potencia (el cuadrado de su amplitud) lo que se ha cuantificado. Para el caso de una red de sensores con múltiples transmisores se ha comprobado cómo en lugar de cuantificar cada uno de los coeficientes individuales del canal o de las ganancias de los mismos resulta más adecuado considerar la realización conjunta del sistema como un vector para el que se cuantifica de un lado su fase y de otro su módulo. No obstante, para la misma configuración física del medio, cuando los usuarios tienen flujos de información independientes (TDMA) la cuantificación del canal escogida ha consistido en la cuantificación individual de cada una de las ganancias. Y si bien puede pensarse que en todos los casos la estructura de cuantificación ha sido fijada *a priori* (que no los valores finales -umbrales, constelaciones-código- que se resolvían a través del problema de optimización), no es menos cierto que en todos los casos la mínima diferencia de prestaciones con respecto a los esquemas sin cuantificación han refrendado la decisión tomada. Retómese ahora el caso del sistema TDMA, si bien es cierto que el receptor potencialmente podría cuantificar todas y cada una de las ganancias de todos los enlaces, no es esa la información que finalmente envía a los usuarios y no lo es porque, para acceder de forma óptima al medio los usuarios, no necesitan conocer todas y cada una de las realizaciones de los otros usuarios sino únicamente

si van a acceder al canal y, sólo en el caso de que efectivamente lo hagan, cuál es la cuantificación de su ganancia (no necesitan nunca saber la de los otros y en la mayor parte de las ocasiones ni siquiera la suya).

Estos ejemplos ilustran el aspecto más relevante cuando se trabaja con sistemas adaptativos. Las prestaciones finales del sistema no dependen directamente del canal, sino de variables relacionadas con el mismo, y en la identificación de estas variables (estado) la estructura del transmisor y su capacidad de adaptación juegan un papel esencial. La definición en sí misma de lo que CSI significa en cada caso es el aspecto fundamental cuando se trabaja con sistemas adaptativos, el elemento en base al cual la formulación se articula. No es lo relevante el canal, sino el uso del estado del mismo. No es, pues, lo relevante la cuantificación del canal, sino el uso y la cuantificación del estado del mismo.

5.2. Líneas futuras de investigación

Como resultado del proceso de realización y de las conclusiones de esta tesis doctoral surgen distintas líneas de investigación futuras que pueden extender el trabajo aquí iniciado. Como ya se puso de manifiesto en el capítulo introductorio para algunas de estas líneas que se mencionan a continuación ya se han obtenido resultados preliminares que avalan su viabilidad y su potencial interés.

- La cuantificación del CSI en el transmisor surge de forma natural a través de la suposición de estimación perfecta del canal en el receptor y de la transmisión de este valor desde el receptor hasta el transmisor a través de un canal de retorno con tasa limitada. No obstante, existen otros escenarios en los que la cuantificación del CSI no se produce sólo en los transmisores, sino también en los receptores (e.g., cuando los receptores están distribuidos y, por lo tanto, ninguno de ellos conoce la totalidad del canal o cuando la unidad que resuelve el problema de asignación de recursos y que genera la palabra de control conteniendo el Q-CSI no es el receptor). El análisis de este nuevo paradigma ha sido una de las líneas de crecimiento que se iniciaron

de forma simultánea a la finalización de esta tesis y habiéndose obtenido ya resultados preliminares para la asignación óptima en sistemas de múltiples usuarios que transmiten sobre un conjunto de canales ortogonales con desvanecimientos [56]. La profundización en los resultados preliminares obtenidos y su extensión a otros sistemas de comunicación resulta de interés.

- Puesto que el canal de realimentación presentaba una tasa de transmisión pequeña y se ha propuesto el uso de esquemas de codificación para la protección frente a errores, todos los diseños derivados han supuesto la integridad de los datos del canal de realimentación o, equivalentemente, la adquisición de Q-CSIT sin existencia de errores. No obstante, la transmisión en entornos especialmente adversos (grandes retardos de comunicación, desvanecimientos severos o fuertes interferencias) puede provocar que este principio de diseño no se cumpla. La modificación de la formulación propuesta (primero) y de los esquemas derivados (después) para considerar Q-CSIT potencialmente errónea resulta así una legítima futura línea de investigación.
- Como se mencionó en la introducción, la maximización de la tasa de transmisión en base a CSI es un objetivo que ha recibido notable atención investigador, no obstante la adaptación en base a Q-CSIT diseñando de forma conjunta la cuantificación del estado del canal y los esquemas de adaptación no está todavía resuelta. A este respecto, debe señalarse que, pese a ser la minimización de la potencia transmitida una motivación fundamental de este trabajo, la mayor parte de los resultados obtenidos son válidos también si el problema a resolver es el de maximización de la tasa de transmisión. La reformulación de los problemas de optimización requeriría únicamente el intercambio de la función objetivo (hasta ahora potencia, en el futuro tasa) con la restricción de hasta ahora tasa, en el futuro potencia. Aplicando además los principios de dualidad la estructura de la solución sería la misma, si bien los algoritmos numéricos de resolución deberían incorporar algunos cambios. Asimismo, habría que tener en cuenta que en la mayoría de los

sistemas analizados, la tasa no era una magnitud continua y la potencia sí lo era. Esto podría implicar que para algún caso, dos valores distintos de potencia generen una misma tasa de transmisión óptima, lo que implica que la restricción de potencia no tendría porqué ser necesariamente ajustada.

- En sistemas de múltiples usuarios, la adaptación óptima de parámetros del nivel físico (tasa, potencia o modulación) y de enlace (usuario y canal asignado) requiere la integración sobre la función densidad de probabilidad del canal inalámbrico [40,54,84], lo que conlleva la realización de cálculos que en ocasiones pueden resultar computacionalmente costosos. Sin embargo, trabajos que contemplan la asignación de recursos en redes de paquetes muestran la posibilidad de utilizar algoritmos estocásticos que calculan para cada realización del canal la asignación de recursos con un consumo computacional despreciable. La extensión de nuestros resultados de adaptación óptima de parámetros de los niveles físico y de enlace a algoritmos estocásticos que no necesiten conocer la descripción probabilística del canal resulta de tremendo interés. El marco teórico para abordar este problema así como resultados para el caso de optimización de sistemas TDMA (enlace ascendente) y TDM (enlace descendente) en base a conocimiento perfecto y cuantificado del canal se han propuesto en: [88].
- Recientes estudios muestran que las redes móviles de telecomunicación actuales infrautilizan el uso del espectro radioeléctrico [31,64]. A modo ilustrativo, en redes en las que únicamente se permite el acceso a usuarios previamente registrados y se ofrecen garantías de calidad de servicio para todos ellos, el uso del espectro puede llegar a ser menor del 20 %. En otras palabras, midiendo el uso del canal en términos temporales, el canal se encuentra libre durante el 80 % del tiempo. Una posible alternativa para enfrentarse a estas circunstancias consiste en el desarrollo de algoritmos que consideren dos o más categorías de usuarios. Un extremo estaría representado por los usuarios que bajo cualquier circunstancia acceden al canal con garantías de servicio y

el otro por usuarios *oportunistas* que sólo pueden hacer uso del canal cuando este se encuentra libre (el resto de categorías podrían representar situaciones intermedias entre las dos anteriormente descritas). Resultados preliminares de algoritmos adaptativos que asignan recursos de forma óptima y se ajustan al escenario descrito pueden encontrarse en: [60]

- En esta tesis la cuantificación del estado del canal y de las variables de transmisión se ha realizado con el objeto de maximizar la calidad de las prestaciones de un sistema de comunicaciones adaptativo. No obstante, existen escenarios de aplicación con objetivos distintos a la transmisión de información (e.g., estimación o control) en donde la *adaptación de los dispositivos* a distintas variables resulta beneficiosa para el sistema. La cuantificación de las variables de *estado* y de diseño para este tipo de sistemas puede asimismo resultar de interés (tómense como ejemplos los casos donde los sensores y actuadores estén distribuidos o donde los dispositivos adaptativos pueda únicamente adoptar un conjunto finito de modos). La extensión y adaptación de las técnicas de optimización y de los principios de diseño utilizados en esta tesis a estos nuevos escenarios resulta una línea de crecimiento a largo plazo que, si bien ambiciosa, presenta gran interés.

Y esto, porque unos saben la causa y otros no.

Pues los expertos saben el qué, pero no el porqué.

Aquellos [los sabios], en cambio, conocen el porqué y la causa.

Aristóteles – Metafísica (s. IV a.C.)

Bibliografía

- [1] “IEEE Std 802.15.1-2005 - IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements. - Part 15.1: Wireless medium access control (MAC) and physical layer (PHY) specifications for wireless personal area networks (WPANs),”
- [2] “IEEE Std 802.15.4-2006 - IEEE Standard for Information technology- Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs),”
- [3] “IEEE Std 802.16-2002. WG - IEEE Standard for Information technology- Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements Part 16: Air interface for fixed broadband wireless access systems,”
- [4] “FCC Report and Order, FCC-05-57A1 – Facilitating Opportunities for Flexible, Efficient, and Reliable Spectrum Use Employing Cognitive Radio Technologies,” Mar. 2005.
- [5] I. F. Akyildiz, W. Su, Y. Sankarsubramaniam, and E. Cayirci, “Wireless sensor networks: A survey,” *Computer Networks*, vol. 38, pp. 393–422, Mar. 2002.
- [6] J. N. Al-Karaki and A. E. Kamal, “Routing techniques in wireless sensor networks: A survey,” *IEEE Wireless Communications Magazine*, vol. 11, pp. 6–28, Dec. 2004.
- [7] W. U. Bajwa, A. M. Sayeed, and R. Nowak, “Matched source-channel communication for field estimation in wireless sensor networks,” in *Proc. Intl.*

- Conf. on Inform. Processing in Sensor Networks*, Los Angeles, CA, pp. 332–339, Apr. 2005.
- [8] G. Barriac, R. Mudumbai, and U. Madhow, “Distributed beamforming for information transfer in sensor networks,” in *Proc. of Intl. Symposium on Information Processing in Sensor Networks*, Berkeley, CA, pp. 81–88, Apr. 2004.
- [9] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, second ed., 1999.
- [10] D. Bertsekas and R. Gallager, *Dynamic Programming and Optimal Control (Volumes 1 and 2)*. Prentice-Hall, second ed., 1992.
- [11] M. Bhardwaj, T. Garnett, and A. P. Chandrakasan, “Upper bounds on the lifetime of sensor networks,” in *Proc. of Intl. Conf. on Commun.*, vol. 3, Helsinki, Finland, pp. 785–790, Jun. 2001.
- [12] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: Information-theoretic and communications aspects,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [14] M. Cho, W. Seo, Y. Kim, and D. Hong, “A Joint Feedback Reduction Scheme Using Delta Modulation for Dynamic Channel Allocation in OFDMA Systems,” in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, Berlin, Germany, pp. 2747–2750, Sep. 2005.
- [15] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, “Packing lines, planes, etc.: Packings in Grassmannian space,” *Experimental Mathematics*, vol. 5, pp. 139–159, Apr. 1996.
- [16] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, “On the Lambert W function,” *Advances in Computational Mathematics*, vol. 5, pp. 329–359, Dec. 1996.
- [17] T. Cormen, C. Leiserson, and R. Rivest, *Introduction to Algorithms*. MIT Press, 1990.

- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York:Wiley, 1991.
- [19] F. F. Digham and M. S. Alouini, "Diversity combining with discrete power loading over fading channels," in *Proc. of Wireless Comm. and Networking Conf.*, Atlanta, GA, pp. 328–332, Mar. 2004.
- [20] F. F. Digham and M. O. Hasna, "Performance of OFDM with M-QAM modulation and optimal loading over Rayleigh fading channels," in *Proc. of IEEE Veh. Tech. Conf.*, Los Angeles, CA, pp. 479–483, Sep. 2004.
- [21] F. F. Digham and M. -S. Alouini, "Variable-Rate Variable-Power Hybrid M-FSK M-QAM for Fading Channels," in *Proc. of IEEE Veh. Tech. Conf.*, vol. 3, Orlando, FL, pp. 1512–1516, Oct. 2003.
- [22] A. Fu, E. Modiano, and J. Tsitsiklis, "Optimal energy allocation for delay-constrained data transmission over a time-varying channel," in *Proc. of INFOCOM Conf.*, vol. 2, San Francisco, CA, pp. 1095–1105, Mar. 30 - Apr. 4 2003.
- [23] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.
- [24] G. B. Giannakis, Z. Liu, X. Ma, and S. Zhou, *Space-Time Coding for Broadband Wireless Communications*. John Wiley and Sons, 2006.
- [25] L. Goldfeld, V. Lyandres, and D. Wulich, "Minimum BER power loading for OFDM in fading channel," *IEEE Transactions on Communications*, vol. 49, pp. 14–18, Jan. 2001.
- [26] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [27] A. J. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," *IEEE Transactions on Communications*, vol. 46, pp. 595–602, May 1998.
- [28] A. J. Goldsmith and S. G. Chua, "Variable-rate variable-power M-QAM for fading channels," *IEEE Transactions on Communications*, vol. 45, pp. 1218–1230, Oct. 1997.
- [29] S. V. Hanly and D. Tse, "Multiaccess fading channels–Part II: Delay-limited capacities," *IEEE Transactions on Information Theory*, vol. 44, pp. 2816–2831, Nov. 1998.

- [30] J. F. Hayes, “Adaptive feedback communications,” *IEEE Transactions on Communications*, vol. 16, pp. 29–34, Feb. 1968.
- [31] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 201–220, Feb. 2005.
- [32] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*. Holden-Day, Inc., 1986.
- [33] W. C. Jakes, *Microwave Mobile Communications*. Wiley, 1974.
- [34] S. M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*. Prentice Hall, 1993.
- [35] M. A. Khojastepour and A. Sabharwal, “Delay-constrained scheduling: Power efficiency, filter design, and bounds,” in *Proc. of INFOCOM Conf.*, vol. 3, Hong Kong, China, pp. 1938–1949, Mar. 2004.
- [36] B. S. Krongold, K. Ramchandran, and D. L. Jones, “Computationally efficient optimal power allocation algorithms for multicarrier communication systems,” *IEEE Transactions on Communications*, vol. 48, pp. 23–27, Jan. 2000.
- [37] J. H. Kwon, D. Rhee, I. M. Byun, Y. Whang, and K. S. Kim, “Adaptive Modulation Technique with Partial CQI for Multiuser OFDMA Systems,” in *Proc. of Intl. Conf. on Advanced Communication Technology*, vol. 2, Phoenix Park, Korea, pp. 1283–1286, Feb. 2006.
- [38] A. Lapidoth and S. Shamai, “Fading channels: how perfect need ‘perfect side information’ be?,” *IEEE Transactions on Information Theory*, vol. 48, pp. 1118–1134, May 2002.
- [39] J. W. Lee, M. Chiang, and R. A. Calderbank, “Optimal MAC design based on utility maximization: Reverse and forward engineering,” in *Proc. of IEEE INFOCOM Conf.*
- [40] L. Li and A. J. Goldsmith, “Capacity and optimal resource allocation for fading broadcast channels—Part I:,” *IEEE Transactions on Information Theory*, vol. 47, pp. 1083–1102, Mar. 2001.

- [41] L. Li and A. J. Goldsmith, “Capacity and optimal resource allocation for fading broadcast channels—Part II: Outage capacity,” *IEEE Transactions on Information Theory*, vol. 47, pp. 1103–1127, Mar. 2001.
- [42] S. Lloyd, “Least squares quantization in PCM,” *IEEE Transactions on Information Theory*, vol. 28, pp. 129–137, Mar. 1982.
- [43] D. J. Love, “Grassmannian Subspace Packing,” (disponible en <http://cobweb.ecn.purdue.edu/~djlove/grass.html>).
- [44] D. J. Love and R. W. Heath, “Limited feedback unitary precoding for spatial multiplexing systems,” *IEEE Transactions on Information Theory*, vol. 51, pp. 2967 – 2976, Aug. 2005.
- [45] D. J. Love and R. W. Heath, “Multimode precoding for MIMO wireless systems,” *IEEE Transactions on Signal Processing*, vol. 53, pp. 3674 – 3687, Oct. 2005.
- [46] D. J. Love, J. R. W. Heath, and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *IEEE Transactions on Information Theory*, vol. 49, pp. 2735–2747, Oct. 2003.
- [47] D. J. Love and R. W. Heath Jr., “Limited feedback power loading for OFDM,” in *Proc. of Military Comm. Conf.*, Monterey, CA, pp. 71–77, Oct. 2004.
- [48] S. H. Low and D. E. Lapsley, “Optimization flow control, I: Basic algorithm and convergence,” *IEEE/ACM Transactions on Networking*, vol. 7, pp. 861 – 874, Dec. 1999.
- [49] X. Luo and G. B. Giannakis, “Raise your voice at a proper pace to synchronize in multiple ad hoc piconets,” *IEEE Transactions on Signal Processing*, vol. 55, pp. 267–278, Jan. 2007.
- [50] A. G. Marques, F. F. Digham, and G. B. Giannakis, “Optimizing Power Efficiency of OFDM Using Quantized Channel State Information,” *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 1581–1592, Aug. 2006.
- [51] A. G. Marques, F. F. Digham, and G. B. Giannakis, “Power-Efficient OFDM via Quantized Channel State Information,” in *IEEE Proc. of Intl. Conf. on Commun.*, Istanbul, Turkey, Jun. 11-15 2006.

- [52] A. G. Marques, F. F. Digham, and G. B. Giannakis, "Power-Efficient OFDM with Reduced Complexity and Feedback Overhead," in *IEEE Proc. of Intl. Conf. on Acoustics, Speech and Signal Processing*, Toulouse, France, May 16-20 2006.
- [53] A. G. Marques, F. F. Digham, G. B. Giannakis, and F. J. Ramos, "Reduced-Complexity Power-Efficient Wireless OFDMA using an Equally Probable CSI Quantizer," in *IEEE Proc. of Intl. Conf. on Commun.*, Glasgow, Scotland, Jun. 24–28 2007.
- [54] A. G. Marques, G. B. Giannakis, F. F. Digham, and F. J. Ramos, "Power Efficient Wireless OFDMA using Limited-Rate Feedback," *IEEE Transactions on Wireless Communications*. (Enviado.).
- [55] A. G. Marques, G. B. Giannakis, F. F. Digham, and F. J. Ramos, "Minimizing Power in Wireless OFDMA with Limited Rate Feedback," in *IEEE Proc. of Wireless Comm. and Networking Conf.*, Hong Kong, Hong Kong, Mar. 11–15 2007.
- [56] A. G. Marques, G. B. Giannakis, and F. J. Ramos, "Optimizing Multiple Access in Wireless Communications Based on Quantized CSI," *IEEE Transactions on Communications*. (Enviado.).
- [57] A. G. Marques, X. Wang, and G. B. Giannakis, "Minimizing Transmit-Power for Coherent Communications in Wireless Sensor Networks with Finite-Rate Feedback," *IEEE Transactions on Signal Processing*. (Enviado.).
- [58] A. G. Marques, X. Wang, and G. B. Giannakis, "Energy Efficient MIMO Systems Using Adaptive Modulation and Coding," in *Proc. of 40th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 29–Nov. 1 2006.
- [59] A. G. Marques, X. Wang, and G. B. Giannakis, "Energy-Efficient TDMA with Quantized Channel State Information," in *Proc. of MILCOM Conf.*, Washington, DC, Oct. 23–25 2006.
- [60] A. G. Marques, X. Wang, and G. B. Giannakis, "Distributed Resource Allocation for Cognitive Radios based on Limited-Rate Feedback," in *IEEE Proc. of XV European Signal Processing Conf.*, Poznan, Poland, Sep. 3–7 2007.

- [61] A. G. Marques, X. Wang, and G. B. Giannakis, “Minimizing Transmit-Power for Coherent Communications in Wireless Sensor Networks using Quantized Channel State Information,” in *IEEE Proc. of Intl. Conf. on Acoustics, Speech and Signal Processing*, Honolulu, HI, Apr. 16–20 2007.
- [62] A. G. Marques, X. Wang, and G. B. Giannakis, “Optimizing Energy-Efficient of TDMA with Finite Rate Feedback,” in *IEEE Proc. of Intl. Conf. on Acoustics, Speech and Signal Processing*, Honolulu, HI, Apr. 16–20 2007.
- [63] M. Medard, “The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *IEEE Transactions on Information Theory*, vol. 46, pp. 933–946, May 2000.
- [64] I. J. Mitola and G. Q. Maguire, “Cognitive radio: making software radios more personal,” *IEEE Personal Communications*, vol. 6, pp. 13–18, Aug 1999.
- [65] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall Publishers, 2000.
- [66] K. Mukkavilli, E. E. A. Sabharwal, and B. Aazhang, “On beamforming with finite-rate feedback in multiple-antenna systems,” *IEEE Transactions on Information Theory*, vol. 49, pp. 2562–2579, Oct. 2003.
- [67] R. Nilsson, O. Edfors, M. Sandell, and P. O. Boerjesson, “An analysis of two-dimensional pilot-symbol assisted modulation for OFDM,” in *Proc. of Intl. Conf. on Personal Wireless Comm.*, Bombay, India, pp. 71–74, Dec. 1997.
- [68] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [69] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, “Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints,” *IEEE Transactions on Signal Processing*, vol. 52, pp. 1179 – 1197, May 2004.
- [70] L. Piazzo, “Fast algorithm for power and bit allocation in OFDM systems,” *IEE Electronic Letters*, vol. 35, pp. 2173–2174, Dec. 1999.
- [71] A. C. Pleite, *Técnicas de máxima diversidad para modulaciones diferenciales en canales inalámbricos con desvanecimientos*. PhD thesis, Universidad Carlos III de Madrid, Jun. 2006.

- [72] J. G. Proakis, *Digital Communications*. McGraw-Hill Higher Education, fourth ed., 2001.
- [73] H. Qian, C. Xiao, N. Chen, and G. T. Zhou, “Dynamic selected mapping for OFDM,” in *Proc. of Intl. Conf. on Acoustics, Speech and Signal Processing*, Philadelphia, PA, pp. 325–328, Mar. 2005.
- [74] T. S. Rappaport, *Wireless Communications*. Prentice Hall, 1996.
- [75] G. Scutari, D. P. Palomar, and S. Barbarossa, “Simultaneous Iterative Water-Filling for Gaussian Frequency-Selective Interference Channels,” in *Proc. of IEEE International Symposium on Information Theory*.
- [76] K. Simon and M. Alouini, “Exponential-Type Bounds on the Generalized Marcum Q-Function Application to Error Probability Analysis over Fading Channels,” *IEEE Transactions on Communications*, vol. 48, Mar. 2000.
- [77] M. K. Simon and M.-S. Alouini, *Digital Communications over Fading Channels*. Wiley-Interscience, 2000.
- [78] D. Soldani, M. Li, and R. Cuny, *QoS and QoE Management in UMTS Cellular Systems*. John Wiley and Sons, 2006.
- [79] G. L. Stuber, J. W. Mark, and I. F. Blake, “An adaptive rate algorithm for FH/BFSK signaling,” *IEEE Transactions on Communications*, vol. 36, pp. 1269 – 1276, Dec. 1988.
- [80] A. S. Tanenbaum, *Computer Networks*. Prentice Hall, 3rd ed., 1996.
- [81] Q. Tang, L. Yang, G. B. Giannakis, and T. Qin, “Battery Power Efficiency of PPM and FSK in Wireless Sensor Networks,” *IEEE Transactions on Wireless Communications*, vol. vol. 6, Jan. 2007.
- [82] Q. Tang, L. Yang, G. B. Giannakis, and T. Qin, “Distributed multiuser power control for digital subscriber lines,” *IEEE Journal on Selected Areas in Communications*, vol. vol. 20, pp. 1105–1115, Jun. 2002.
- [83] A. Trushkin, “Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions,” *IEEE Transactions on Information Theory*, vol. vol. 28, pp. 187 – 198, Mar. 1982.

- [84] D. Tse and S. V. Hanly, “Multiaccess fading channels—Part I: Polymatroid structure, optimal resource allocation and throughput capacities,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2796–2815, Nov. 1998.
- [85] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [86] E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal, “Energy-efficient packet transmission over a wireless link,” *IEEE/ACM Transactions on Networking*, vol. 10, pp. 487–499, Aug. 2002.
- [87] X. Wang and G. B. Giannakis, “Energy-efficient resource allocation in TDMA over fading channels,” in *Proc. of the Intl. Symp. on Info. Theory*, Seattle, Washington, pp. 1505–1509, Jul. 2006.
- [88] X. Wang, G. B. Giannakis, and A. G. Marques, “A Unified Approach to QoS-Guaranteed Scheduling for Channel-Adaptive Wireless Networks,” *Proceedings of the IEEE*. (Enviado.).
- [89] X. Wang, A. G. Marques, and G. B. Giannakis, “Power-Efficient Resource Allocation and Quantization for TDMA Using Adaptive Transmission and Limited-Rate Feedback,” *IEEE Transactions on Signal Processing*. (Enviado.).
- [90] Z. Wang and G. B. Giannakis, “Wireless Multicarrier Communications: Where Fourier Meets Shannon,” *IEEE Signal Process. Mag.*, vol. 17, pp. 29–48, May 2000.
- [91] C. Wong, R. Cheng, K. Lataief, and R. Murch, “Multiuser OFDM with adaptive subcarrier, bit, and power allocation,” *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1747–1758, Oct. 1999.
- [92] J. Wu, S. Zhou, Z. Wang, and M. Doroslovacki, “Quantifying the performance gain of direction feedback in a MISO system,” in *Proc. of Conf. on Info. Sciences and Systems*, Princeton, NJ, Mar. 2006.
- [93] P. Xia and G. B. Giannakis, “Design and analysis of transmit-beamforming based on limited-rate feedback,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 1853–1863, May 2006.

-
- [94] P. Xia, S. Zhou, and G. B. Giannakis, “Multi-antenna adaptive modulation with beamforming based on bandwidth-constrained feedback,” *IEEE Transactions on Communications*, vol. 53, pp. 526–535, Mar. 2005.
- [95] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, “Power scheduling of universal decentralized estimation in sensor networks,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 413–422, Feb. 2006.
- [96] Y. Yao and G. B. Giannakis, “Energy-efficient scheduling for wireless sensor networks,” *IEEE Transactions on Communications*, vol. 53, pp. 1333–1342, Aug. 2005.
- [97] Y. Yao and G. B. Giannakis, “Rate-Maximizing Power Allocation in OFDM Based on Partial Channel Knowledge,” *IEEE Transactions on Wireless Communications*, vol. 4, pp. 1073–1083, May 2005.
- [98] W. Ye, J. Heidemann, and D. Estrin, “An energy-efficient MAC protocol for wireless sensor networks,” in *Proc. of 21st Annual Joint Conf. IEEE Computer Commun. Societies*, vol. 3, New York, pp. 1567–1576, Jun. 2002.
- [99] M. Zafer and E. Modiano, “A calculus approach to minimum energy transmission policies with quality of service guarantees,” in *Proc. of INFOCOM Conf.*, vol. 1, Miami, FL, pp. 548–559, Mar. 2005.
- [100] S. Zhou and G. B. Giannakis, “Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations,” *IEEE Transactions on Signal Processing*, vol. vol. 49, pp. 1673 – 1690, July 2003.
- [101] S. Zhou and G. B. Giannakis, “Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback,” *IEEE Transactions on Signal Processing*, vol. vol. 50, pp. 2599 – 2613, Oct. 2002.
- [102] S. Zhou and B. Li, “BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 1653–1665, May 2006.
- [103] S. Zhou, Z. Wang, and G. B. Giannakis, “Quantifying the power-loss when transmit-beamforming relies on finite-rate feedback,” *IEEE Transactions on Wireless Communications*, vol. 4, pp. 1948–1957, Jul. 2005.