

EDITORIAL

Ya ha llegado de nuevo el invierno y, con él, una nueva edición de Buran. La gente de la Rama de Estudiantes del IEEE seguimos al pie del cañón: recolectando artículos, paginando, lidiando con la impresión... Y todo para, como si de un hijo se tratase, recibir en el despacho con inmensa alegría el envío de mensajería con los paquetes que contienen nuestra revista recién salida del horno. La verdad es que pocas actividades merecen tanto la pena en la universidad como aquellas que involucran el trabajo en equipo. Participar de una asociación como el IEEE es enriquecedor tanto en el plano formativo como personal. No me equivoco cuando digo sin dudar que los mejores años de la Universidad los he pasado en la Rama.

Ahora que la mayoría de veteranos de la Rama finalizamos nuestros estudios, marchando algunos al extranjero, llega el momento de pasar el testigo a los machot@s (guiño al lector fiel que lee nuestra editorial) que continuarán en la brecha. Aprovechamos entonces estas líneas para despedirnos de nuestros lectores, seguros al cien por cien de dejar la Rama en buenas manos.

Hasta siempre,

Eduard Calvo Page
Coordinador de Buran

COORDINACIÓN BARCELONA

Eduard Calvo

EDICIÓN BARCELONA

Xavier Bielsa
Felipe Calderero
Eduard Calvo
Daniel Camps
José Antonio Casas
Miguel Ángel Fernández
Jordi Gilabert
Héctor Julián
Jose Luis Hernández
Ignasi Pérez
Alfredo C. López Salcedo
Jaume Ribot
Rosa Mª Tarruella

REVISIÓN

Carles Gómez Montenegro
José A. López Salcedo
Miguel Angel Sastre
Josep Paradells

DISEÑO PORTADA

Jordi Gilabert

AGRADECIMIENTOS

Il. Dir. Juan A. Fernández Rubio,
Ángel Cardama, Jose A. Delgado-Penín, Jorge Luis
Sánchez-Ponz y a los puntos de distribución en la
UPC: Abacus, CPET, y Kiosk Campus Nord.

We would also like to thanks Ms. Laura Durrett
(IEEE Student Services Manager), and IEEE
International for their helpful support, encouragement
and financial funding for distributing Buran across
south american Region 9 IEEE Student Branches.

IMPRESIÓN
RET, s.a.l.
FOTOMECAÑICA
Sistemes d'Edició

La organización se reserva el derecho de publicar los artículos. La opinión expresada en los artículos no tiene por qué coincidir con la de la organización.

Agradecemos las colaboraciones hechas desinteresadamente, y a causa de la falta de espacio, pedimos disculpas a todas aquellas personas a las cuales no se les ha publicado su colaboración. Esperamos que en un próximo número tengan cabida.



INTRODUCCIÓN A LOS CÓDIGOS TURBO

Eduard Calvo Page, Jaume Ribot Caner

Estudiantes de la ETSETB y Miembros de
la Rama de Estudiantes del IEEE de Barcelona
{eduard25.jaume25}@casal.upc.es

ABSTRACT

El objetivo de este artículo es dar al lector una visión global de los Códigos Turbo a modo de tutorial. No pretende introducir ninguna innovación en el campo, sino evitar la tarea de tener que recurrir a muchas fuentes demasiado específicas. Se tratarán los aspectos fundamentales de la codificación Turbo: qué es y las bases de su funcionamiento.

0. INTRODUCCIÓN

Desde que *Shannon* inició la Teoría de la Información en 1.948 con su histórico artículo '*A Mathematical Theory of Communication*', la codificación de canal se ha convertido en una rama fundamental de ella. Su objetivo es conseguir fiabilidad en la transmisión de información entre fuente y destino. Es decir, poder reducir tanto como se quiera la probabilidad de error en recepción causada por los efectos negativos del medio físico: canal, interferencias y ruido. Para ello emplea códigos que protegen la información con redundancia para dotarla de mayor robustez.

Los primeros trabajos, a medio camino entre el álgebra abstracta y la ingeniería, se centraron en la búsqueda de códigos bloques llamados 'perfectos', dotados de una estructura algebraica concreta (códigos de *Hamming*, *BCH*, *Reed-Solomon*...). El primer punto y aparte se produjo con el trabajo de *Ungerboeck*, que en 1.982 introdujo los códigos *TCM* ('*Trellis Coded Modulation*') [1]. Su aportación revolucionó la transmisión sobre canales limitados en banda, permitiendo a los sistemas trabajar cercanos al límite de capacidad (el número de bps máximo que podemos transmitir con una probabilidad de error arbitrariamente baja). El mayor progreso desde entonces sin ninguna duda son los *Códigos Turbo*.

Presentados en 1.993 por *Berrou* et.al. [2], representan una innovación total tanto en su construcción (heurística, sin ninguna base matemática) como en su decodificación (iterativa y que recupera conceptos ya presentes en los códigos *LDPC* de *Gallager* [3], 1.963).

Su extraordinario buen comportamiento (rozando el límite de capacidad de *Shannon*) hizo que en un principio se recibieran con escepticismo, que luego tuvo que transformarse en una búsqueda del porqué de su bondad. Actualmente son omnipresentes, pudiéndose encontrar en multitud de aplicaciones de comunicaciones espaciales, pero también en el estándar *UMTS*, *WLAN* y en TV digital.

1. EL CODIFICADOR

El codificador *Turbo* consiste en la concatenación de dos códigos continuos recursivos y sistemáticos, denominados codificadores constituyentes, que operan uno sobre la información a codificar (la parte sistemática) y el otro sobre una versión permutada (entrelazada) de la misma, tal y como muestra la Figura 1.

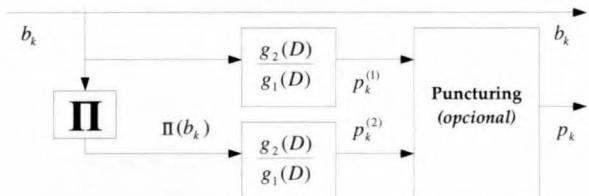


Figura 1. Diagrama de bloques del codificador

Seguidamente se comenta la función de cada bloque del diagrama anterior.

1.1. CODIFICADORES CONSTITUYENTES

A los codificadores continuos que se concatenan en paralelo se les denomina constituyentes, caracterizados por dos polinomios: uno de *feedback* ($g_1(D)$) y otro de *feedforward* ($g_2(D)$). Estos codificadores son recursivos y sistemáticos, siendo su matriz generadora $G_{RS} = [1 \ g_2(D)/g_1(D)]$, de la cual se extrae únicamente la rama de paridad. Su implementación con registros de desplazamiento se muestra en la Figura 2.

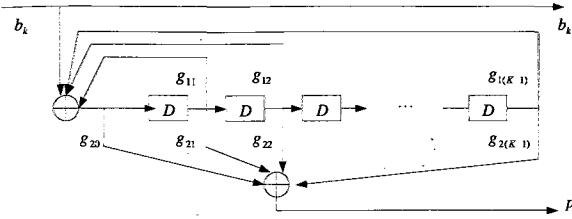


Figura 2. Estructura de un codificador continuo, recursivo y sistemático

Cualquier diseño de un codificador de canal intenta maximizar la distancia mínima entre dos palabras código. Para un código lineal, esta distancia se corresponde con el peso mínimo de entre todas las palabras código distintas de la todo ceros. Como todos los bloques que conforman el codificador Turbo son lineales, este en su conjunto también lo es. Por lo tanto, para maximizar su distancia interesa que las paridades sean de mucho peso. Es por ello que los constituyentes se implementan recursivamente ya que únicamente aquellas secuencias de entrada múltiples de $g_1(D)$ darán una paridad finita (poco peso). En caso contrario, el enrejado nunca converge al estado de reposo y al truncar la paridad a un número de bits igual al tamaño de bloque de la entrada (N) es de esperar que hayan tantos unos como ceros en media (mucho peso). Interesa que haya pocas secuencias divisibles entre y por ello este polinomio se escoge primitivo.

La notación usada en la literatura para describir estos códigos consiste en la representación octal de los dos polinomios. En la Tabla 1 se adjunta un ejemplo.

$g_1(D)$	$1 + D + D^4$	$g_1(D)_{oct} = [11001]_{oct} = 31$	Código
$g_2(D)$	$1 + D^2 + D^3 + D^4$	$g_2(D) = [10111]_{oct} = 27$	(31,27)

Tabla 1. Ejemplo de notación de un código constituyente

1.2. ENTRELAZADOR(INTERLEAVER)

Un entrelazador π es una bloque lineal que mapea una secuencia de N bits de entrada $\mathbf{u} = (u_1, u_2, \dots, u_N)$ en otra secuencia de salida que corresponde a una permutación de la entrada, de manera que para describir el entrelazador empleamos un vector contenido las transposiciones denotadas por $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ de manera que la salida $\pi(\mathbf{u})$ se obtiene como $u'_{\pi(i)} = u_i$ para $i = 1, \dots, N$.

La función del entrelazador es desordenar suficientemente la secuencia de bits de entrada para que si ésta era divisible por $g_1(D)$ a la entrada del

primer constituyente, no lo sea a la entrada del segundo. Así reduciremos la probabilidad de que ambas ramas de paridad tengan un peso pequeño, aumentando la distancia mínima del código y mejorando con ello sus prestaciones. Los más utilizados son los llamados S-random [5][6], aunque debido al excesivo tiempo de cálculo que se requiere para encontrarlos cuando el tamaño de bloque es grande también se usan los entrelazadores rectangulares («block interleavers»). Estos últimos se implementan escribiendo fila a fila los bits en una matriz y leyéndolos columna a columna.

1.3. PUNCTURING

La tasa que ofrece un código turbo sin puncturing es $r=1/3$ (por cada N bits de información obtenemos $3N$ bits de palabra código a la salida). Este overhead, aunque necesario en algunas aplicaciones que exijan mucha robustez, puede ser excesivo en escenarios con restricciones de retardo donde es más apropiada una tasa mayor. Es por ello que puede incluirse un bloque de puncturing que elimine ciertos bits de redundancia. Esta eliminación aunque puede hacerse, por ejemplo, quedándose con los bits pares del primer constituyente y los impares del segundo [1] existen otros métodos más sofisticados para eliminar periódicamente ciertos bits [7][8].

2. DECODIFICACIÓN

La decodificación de los códigos turbo puede que se trate en profundidad en próximos trabajos así que en el presente sólo se pretenden introducir aquellos aspectos más básicos. Se considerará que se usa una modulación BPSK sobre canal AWGN.

Todo algoritmo de decodificación turbo consiste en obtener iterativamente la siguiente métrica para cada bit de información:

$$LLR(b_k) = \log \left(\frac{P(b_k = 1)}{P(b_k = 0)} \right) = \log \left(\frac{P(b_k = 1 | \mathbf{y})}{P(b_k = 0 | \mathbf{y})} \right)$$

Esta métrica se denomina *Log-Likelihood Ratio* y es novedosa en el sentido que no está basada en el criterio *ML* ('Maximum Likelihood') como el algoritmo de Viterbi [9], que es el que tradicionalmente se ha usado para la decodificación de los códigos continuos. Este cambio fue necesario debido a la complejidad de la aplicación de este criterio a los códigos Turbo (se describirían con un enrejado de tamaño prohibitivo), que emplean el algoritmo BCJR [10] y sus modificaciones de menor complejidad, *Log-MAP* y *max-Log-MAP* [11], para aplicar el *MAP* ('Maximum A Posteriori').

La decisión a partir del *LLR* se basa en su signo. Cuando es positivo se decide 1 y 0 en caso contrario. Este modo de proceder constituye lo que se denomina una decodificación *soft* porque el valor absoluto del *LLR* es una medida de la confianza en la decisión. Además, cuantas más iteraciones realicemos para calcular esta métrica con más precisión obtendremos su valor real.

3. SIMULACIONES

Para ilustrar los aspectos más importantes que influyen en el diseño de un código turbo, se ha realizado una implementación. El objetivo es observar la variación de las prestaciones globales (en cuanto a tasa de error de bit, *BER*) del código con la variación de los bloques que lo forman (tamaño de bloque, entrelazador, iteraciones...).

En una primera simulación (Figura 3) hemos comprobado la evolución de la *BER* con el número de iteraciones del decodificador (usando el algoritmo *max-Log-MAP*).

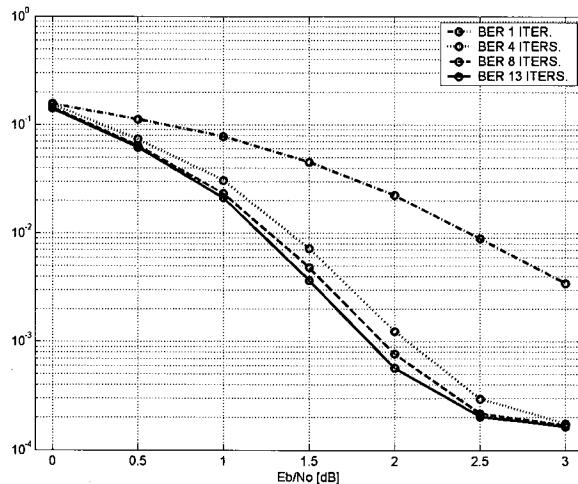


Figura 3. Variación del número de iteraciones en decodificación.

Como es de esperar, al incrementar el número de iteraciones la probabilidad de error de bit disminuye drásticamente, aunque no de forma proporcional. Al pasar de una a cuatro iteraciones, la ganancia de codificación es muy grande, pero una vez que el algoritmo converge el aumento en la capacidad de corrección es menor, no compensando la mejora de la *BER* con el aumento de la complejidad de la decodificación (más iteraciones implica mayor retardo de decodificación). En todas las gráficas siguientes se tomarán 4 iteraciones para decodificar.

La siguiente simulación (Figura 4), pretende mostrar, para un mismo código turbo (tamaño de bloque $N = 200$, 4 iteraciones y entrelazador S-random con

$S = 9$) cómo afecta la elección de los codificadores constituyentes.

Para el primer caso usamos como constituyente el código [7,5] (de polinomios $g_1(D) = 1 + D + D^2$ y $g_2(D) = 1 + D^2$) y para el segundo el [37,21] (con $N=400$ y $S = 9$). En el primer caso, el código constituyente es de cuatro estados mientras que el segundo es de 16. Por ello, esperaríamos que el código [37,21] ofreciera una *BER* mucho mejor. Sin embargo, sólo es primitivo en el primer caso por lo que todo lo que se gana al poner un constituyente cuatro veces más complejo se pierde si no se escogen adecuadamente los polinomios, como se observa en la figura.

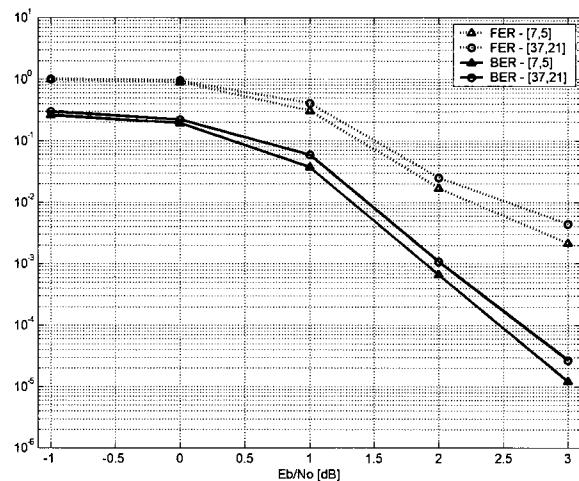


Figura 4. Uso de diferentes polinomios constituyentes.

Una tercera simulación (Figura 5) comprueba la importancia del entrelazador escogido para la codificación. Se han escogido dos de los más habituales, un S-random y un rectangular (denotado como *block*) para un mismo tamaño de bloque $N = 200$. Aunque el tamaño de bloque no es lo suficientemente grande como para apreciar diferencias significativas, sí que se observa que para Eb/No's medias y altas, el S-random supera en un orden de magnitud al rectangular. Esta diferencia se acentuaría aún más si tratásemos con tamaños de bloque del orden de miles de bits. Sin embargo, el tiempo de cálculo y la complejidad del algoritmo de búsqueda de los entrelazadores S-random crece exponencialmente con el tamaño de bloque (aunque se han presentado algunas propuestas para reducir ambos [12][13]) y hace casi imposible atacar el problema si no es usando supercomputadoras.

La cuarta y última simulación (Figura 6) ilustra la influencia del tamaño de bloque.

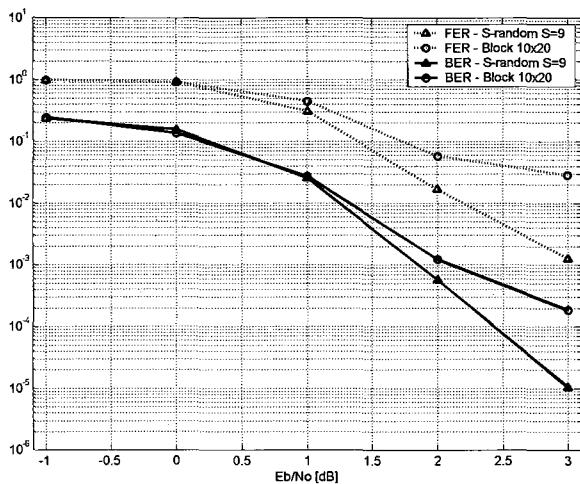


Figura 5. Uso de diferentes entrelazadores.

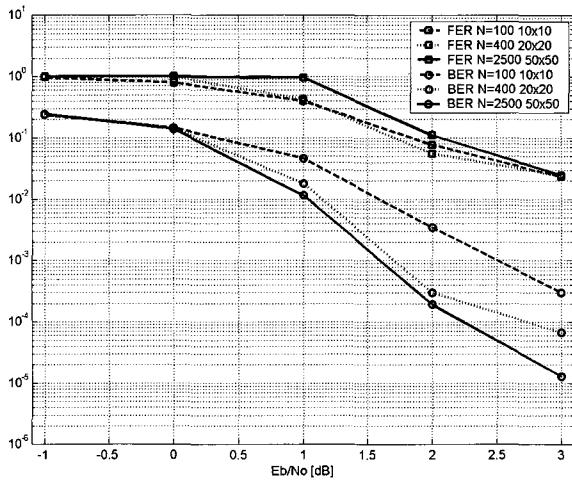


Figura 6. Variación del tamaño de bloque, N.

Entendemos como tamaño de bloque la longitud de la secuencia de entrada que cogemos para pasar a través de los codificadores constituyentes y el entrelazador. Se observa claramente que cuanto mayores es esa longitud, mejor BER obtenemos a la salida. El uso de tramas mayores pretende dotar de capacidad de corrección a bits muy alejados entre sí, haciendo más efectivo el empleo del entrelazador. Sin embargo, y al igual que ocurría con las iteraciones, la ganancia de codificación no es proporcional al tamaño de bloque (al menos si se usan entrelazadores rectangulares).

4. CONCLUSIONES

En el presente artículo, se ha dado una visión global de las características principales de los códigos turbo. En una primera parte, se ha presentado tanto el codificador como el decodificador de forma general. Una vez comentados los parámetros influyentes en las prestaciones del código, hemos realizado una serie de simulaciones para contrastar las previsiones teóricas con los resultados obtenidos a través de la implementación del algoritmo.

Como conclusiones, resaltamos los siguientes puntos:

- Los códigos turbo suponen un punto y a parte en la codificación de canal, sobrepasando en prestaciones cualquier código existente hasta el momento de su aparición.

- Presentan una estructura novedosa pero aplicando conceptos aparecidos hace tiempo (decodificación «soft» iterativa o aleatorización en la secuencia de entrada), que mezclados y aplicados de forma inteligente, forman una estructura que obtiene una calidad excelente.

- Centrándonos en el propio código, apreciamos varios parámetros que influyen directamente sobre su calidad. Vemos que a igualdad de códigos constituyentes, dos codificadores turbo pueden ofrecer prestaciones completamente diferentes dependiendo del entrelazador usado.

- También vemos que el tamaño de la secuencia de entrada es un parámetro muy importante, ya que a mayor longitud de trama, mayor capacidad de corrección. Sin embargo, la elección de este parámetro debe surgir de un compromiso entre BER y retardo, pues éste último aumenta con el tamaño de bloque.

- Por más complejo que sea el algoritmo de búsqueda del entrelazador deseado, una vez encontrado la complejidad de la codificación turbo es parecida.

- Existen gran multitud de alternativas de diseño de entrelazadores [14-19] dependiendo de la aplicación en concreto, estando la mayoría de ellas basadas en métodos heurísticos.

- Según lo visto, no hemos encontrado un métrica única que permita comparar la bondad de distintos entrelazadores.

5. BIBLIOGRAFÍA

- [1] G. Ungerboeck, «Channel coding with multilevel/phase signals», *IEEE Transactions on Information Theory*, vol. IT-28, pp. 55-67, Jan. 1982.
- [2] C. Berrou, A. Glavieux, and P. Thitimajshima, «Near Shannon limit error-correcting coding and decoding: Turbo-codes», *Proc. IEEE Int. Conf. on Comm.*, Geneva, Switzerland, vol. 2, pp. 1064-1070, May 1993.
- [3] R. G. Gallager, «Low Density Parity Check Codes», Monograph, M.I.T. Press, 1963.
- [4] W. E. Ryan, «A Turbo Code Tutorial».

- [5] S. Dolinar, D. Divsalar, «Weight Distributions for Turbo Codes Using Random and Nonrandom Permutations», *TDA Progress Report*, (1999), 42-122.
- [6] D. Divsalar and F. Pollara, «Multiple turbo codes for deep-space communications», *TDA Progress Report* 42-121, Jet Propulsion Laboratory, Pasadena, California, pp. 66-77, May 15, 1995.
- [7] J. Hagenauer, «Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their Applications», *IEEE Transactions on Communications*, vol. 36, no. 4, April 1988.
- [8] T. Ji, W.E. Stark, «Turbo-Coded ARQ Schemes for DS-CDMA Data Networks Over Fading and Shadowing Channels: Throughput, Delay, and Energy Efficiency», *IEEE Journal on Selected Areas in Communication*, vol. 18, no. 8, August 2000.
- [9] A. J. Viterbi, «Error bounds for convolutional codes and an asymptotically optimal decoding algorithm», *IEEE Transactions on Information Theory*, IT-13:260-269, 1967.
- [10] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, «Optimal decoding of linear codes for minimizing symbol error rate», *IEEE Trans. Inf. Theory*, pp. 284-287, Mar. 1974.
- [11] J.P. Woodard and L. Hanzo, «Comparative Study of Turbo Decoding Techniques: An Overview», *IEEE Transactions on Vehicular Technology*, vol. 49, no. 6, Nov. 2000.
- [12] B. G. Lee, S. J. Bae, S. G. Kang, and E. K. Joo, «Design of swap interleaver for turbo codes», *Electronic Letters*, vol. 30, no. 4, pp. 287-288, Feb. 1994.
- [13] B. G. Lee, S. J. Bae, C. K. Jeong, and E. K. Joo, «Performance Analysis of Swap Interleaver for Turbo Codes», *Proc. ICT '2001*.
- [14] H. R. Sadjapour, M. Salehi, N. J. A. Sloane, and G. Nebe, «Interleaver design for short block length Turbo codes», *submitted to ICC2000*, New Orleans, LA.
- [15] J. Hokfelt and T. Maseng, «Methodical interleaver design for turbo codes», *Int. Symposium on Turbo Codes & Related Topics*, pp. 212-215, Brest, France, Sept. 1997.
- [16] J. A. Briffa and V. Buttigieg, «Interleavers for Unpunctured Symmetric Turbo Codes», *Proc. Of the IEEE Int. Symposium on Information Theory*, p.450, Sorrento, Italy, Jun 25-30, 2000.
- [17] M. Breiling, S. Peeters, and J. Huber, «The Class of Double Terminating Turbo Code Interleavers», *IEE Electronic Letters*, no. 5, 99. 389-391.
- [18] M. Breiling and J. Huber, «Interleaver Design for Turbo Codes Using a Backtracking Algorithm», *submitted to the ICC2000*.
- [19] A. Henriksson, J. Hokfelt, O. Edfors, «Evaluation of an Interleaver Design Algorithm for Turbo Codes in UMTS», ETSI SMG2 UMTS-L1 Expert Group Tdoc 419/98.
- [20] L. C. Perez, J. Seghers, and D.J. Costello, «A distance spectrum interpretation of turbo codes», *IEEE Trans. Inform. Theory*, vol. IT-42, no. 6, pp. 1698-1709, Nov. 1996.

6. AUTORES



Eduard Calvo Page es estudiante de 5º curso de Ingeniería Superior de Telecomunicación en la ETSETB (UPC), especialidad en Comunicaciones. Vinculado actualmente al proyecto europeo ROMANTIK a través del departamento de Teoría de la Señal y Comunicaciones, investiga sobre la combinación de esquemas de transmisión cooperativa, técnicas avanzadas de codificación de canal y el uso de diversidad espacial para el aumento de la capacidad en el HSDPA de UMTS. Es miembro de la Rama de Estudiantes del IEEE de Barcelona desde 2001.



Jaume Ribot Caner es Ingeniero Técnico de Telecomunicación por la Escola Universitaria Politècnica de Vilanova i la Geltrú (EUPVG), de la Universitat Politècnica de Catalunya (UPC). Actualmente se encuentra finalizando sus estudios de Ingeniería Superior de Telecomunicaciones en la Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona (ETSETB) también de la UPC, donde a la vez es becario del departamento de Telemática trabajando en temas de comunicaciones móviles y redes Ad-Hoc, dentro del proyecto I2Cat. Es miembro de la Rama de Estudiantes del IEEE de Barcelona desde el presente año.



(LA IMPOSIBILIDAD DE) EL COMPRESOR INFINITO

Pau Garcia i Quiles

Estudiante de Ingeniería de Telecomunicaciones e Ingeniería Informática
Universitat Politècnica de València

pgq@poboxes.com

Desde hace muchos años vemos anuncios de empresas que dicen haber inventado compresores de propiedades asombrosas: ratios de compresión 100 a 1, 1000 a 1, etc. ¿Por qué no los usamos si las ventajas son tan evidentes?

INTRODUCCIÓN

Bit, byte, compresión, tipo de datos... terminología que hace diez años hubiera dejado boquiabierto al más templado es hoy en día de uso corriente. Gracias al avance de la Sociedad de la Información, la informática y su vocabulario se han expandido como la pólvora, a la par que Internet se introducía en miles de hogares y empresas.

Además de hacer la declaración de la Renta y comprar libros, Internet ofrece muchas otras posibilidades. El deseo de transmitir o almacenar cada vez mayor volumen de datos sin cambiar el medio (línea telefónica, tarjeta de memoria, etc.) hace que se trabaje intensamente en el campo de la Teoría de la Información, especialmente, en la compresión de datos.

Un compresor es un algoritmo que disminuye el espacio ocupado por cierta información. Cada año vemos en los medios científicos y especializados (New Scientist, Nature, Byte, ZDNet, etc.) varios anuncios de empresas que aseguran haber superado el límite teórico (límite de Shannon) de compresión de datos. Generalmente, estos anuncios van asociados a peroratas como «necesitamos más dinero para terminar las investigaciones», «tenemos la tecnología pero hace falta un socio interesado en comercializar la tecnología» (*ergo* gastar dinero en la empresa), etc. ¿Le suena de algo?

Un estudiante irlandés de 16 años consiguió engañar a un jurado de 12 personas y ganó el Premio al Joven Científico del Año 2002 con su «invento», un navegador de Internet llamado XWebs, que supuestamente multiplicaba por cuatro la velocidad de navegación. ZeoSync anunciaba en su página nada menos que cinco tecnologías relacionadas con la compresión de datos: codificador (BinaryAccelerator™), compresor (BitPerfect™), secuenciador (Zero Space Tuner™), transformador de dominio (Relational Differentiation Encoding™) y el

conjunto de todo (TunerAccelerator™). Pegasus Technologies afirmaba que podía guardar 1.28 Gigabytes en un disquete de 3.5" (cuya capacidad normal es 1.44 Megabytes) gracias a su tecnología HyperDrive™ (de la que llegaron a obtener una patente –cosa fácil en USA, todo sea dicho–). Web Technologies nos asombraba con su compresor DataFiles/16™, que comprimía cualquier archivo mayor de 64 KBytes a un dieciseisavo de su tamaño. David C. James obtuvo una patente en USA para su tecnología HyperSpace™, que alardeaba de comprimir datos aleatorios. Premier Research Corporation hizo una aparición estelar con su compresor Minc, otro que decía poder comprimir datos aleatorios. Pixelon se aprovechó de la burbuja de las puntocom para vender humo y dilapidar 35 millones de dólares USA de sus inversores, mientras decía haber inventado un compresor de vídeo y audio de propiedades casi mágicas. Todas estas empresas han desaparecido.

Para entrar en materia, veamos un método sencillo de compresión. Si sustituímos las repeticiones de una letra por el número de veces que se repite podríamos comprimir esta secuencia:

AAAAAAAAAAAAAAABAAACDEEEEECCC
CBBBFNTTWCCAJJJLLLLL

como:

15AB3ACD6E4C3BF4TW2C5A3J5L

Lo que antes ocupaba 55 caracteres, ahora ocupa sólo 26: ¡una disminución del 53%!

En el ejemplo anterior hemos usado una versión bastante pedestre de un algoritmo de codificación por longitud de recorrido (Run-Length Encoding, RLE), pero que muestra la idea básica: eliminar aquello que se repite.

Los compresores eliminan la redundancia, posibilitando que una conexión a Internet lenta permita transmitir grandes volúmenes de datos, que se puedan realizar llamadas telefónicas y videoconferencias por Internet, etc. En resumen: permiten exprimir más nuestra capacidad de transmisión y almacenamiento de datos.



CONCEPTOS BÁSICOS DE TEORÍA DE LA INFORMACIÓN

La Teoría de la Información es la rama de la Matemática que estudia qué es la información, cómo medirla, cómo representarla y cómo transmitirla. Este campo es uno de los más jóvenes de la Matemática: su origen es el artículo *Una teoría matemática de la comunicación*, publicado en 1948 por Claude Shannon.

Antes que nada, ¿qué es la información?

La definición formal dice que la «información» de un mensaje es *la medida en que dicho mensaje despeja la incertidumbre sobre algo*. Con esta definición estamos suponiendo que la información se refiere a algo que desconocemos, es decir, algo que desde nuestro punto de vista tiene un componente aleatorio.

En Teoría de la Información se dice que los mensajes proporcionan la información, así que hablaremos indistintamente de «mensaje» o de «información», porque ésta última está asociada al primero. También se dice que los mensajes son emitidos por una *fuente de información*, que se supone aleatoria (así que no podemos predecir qué mensaje va a salir de la fuente). Debido a que la fuente es aleatoria, la Estadística nos permite caracterizarla usando una variable aleatoria X .

La fuente de información puede ser prácticamente cualquier cosa: una persona dictándonos los números que salen del «bombo» de la Lotería, alguien dictándonos nombres que lee al azar de la guía telefónica, un archivo de ordenador, etc. En todos estos casos, podemos reducir los mensajes a caracteres alfanuméricos (texto y números), que en una computadora se codifican como combinaciones de unos y ceros.

Nuestro alfabeto y nuestro sistema numérico tienen un conjunto finito de símbolos (las letras mayúsculas, las letras minúsculas, los números del 0 al 9, etc), pongamos, por ejemplo, que son 500 símbolos (en un caso genérico, hablaríamos de un alfabeto formado por N símbolos). Debido a esto, los valores de la variable aleatoria X (es decir, los mensajes) son también finitos y son exactamente los mismos (si los mensajes son de un símbolo, la fuente podrá emitir como mucho $N=500$ mensajes diferentes).

Matemáticamente, la información de un mensaje x_i (donde i es el número entre 1 y el número total de mensajes posibles N ; en nuestro ejemplo, i varía entre 1 y 500) se caracteriza con un logaritmo en base 2 y la probabilidad de que la fuente X emita ese mensaje x_i :

$$I(x_i) = -\log_2 P(X = x_i)$$

Debido al signo negativo, cuanto más improbable es un mensaje, más información aporta. Por ejemplo, aporta

mucho más información decir «hoy en el Ecuador hace frío» que decir «hoy en el Ecuador hace calor», porque lo primero es mucho más improbable que lo segundo.

Otro concepto de Teoría de Información que necesitaremos es la entropía. La idea es muy similar a la de la Termodinámica: la entropía mide lo desordenada que está la energía, en nuestro caso, lo desordenada que está la información. La función entropía da la cantidad media de información de la fuente de mensajes modelada por la variable X , es decir, la cantidad media de información que nos proporciona un mensaje x_i sobre X :

$$H(X) = \sum_{i=1}^N P(x_i) I(x_i)$$

Como consecuencia de esta definición, cuanto mayor es la entropía de una fuente, más información tiene cada mensaje, es decir, más improbable (difícil de predecir) es el mensaje.

Así pues, cuanto mayor es la entropía de una fuente, más difícil es comprimir los mensajes que emite.

TIPOS DE COMPRESORES: LOSSY Y LOSSLESS

En compresión de datos se distingue entre dos tipos de compresores, los compresores sin pérdida (*lossless*) y los compresores con pérdida (*lossy*).

La diferencia es evidente: cuando comprimimos un fichero de ordenador con un compresor *lossless*, al descomprimirlo obtenemos exactamente el mismo fichero que teníamos en origen, porque el compresor no produce pérdida de información. Con un compresor *lossy*, en cambio, es imposible recuperar el fichero tal y como era en origen, porque el compresor ha producido pérdida de información. Como contrapartida, los compresores *lossy* alcanzan niveles de compresión mucho mayores que los compresores *lossless* (a costa de la calidad del fichero obtenido).

Hay aplicaciones en las que es imprescindible usar compresores *lossless*, por ejemplo, en un documento de un procesador de texto: si usáramos un compresor *lossy*, perderíamos información, y al decomprimirla, podríamos encontrarnos con que nos faltan frases, párrafos o imágenes.

En otras aplicaciones, en cambio, se puede usar sin problemas un compresor *lossy*. Los ejemplos más claros son la voz y la imagen. Como ni nuestro oído ni nuestro ojo son perfectos, una disminución moderada de la calidad es inapreciable. Los conocidos formatos MP3, MPEG, AVI, etc. son formatos de compresión *lossy*: se consigue que el audio o el vídeo ocupen poco espacio a costa de disminuir su calidad.

Después de ver en qué consiste la compresión lossy y la compresión lossless, podemos intuir que:

- La compresión lossless, al ser sin pérdida de información, tiene un límite máximo de compresión
 - La compresión lossy, al ser con pérdida de información, no tiene un límite máximo de compresión: podemos comprimir tanto como queramos. Obviamente, cuanto más comprimamos, menos información tendremos (es decir, peor será la calidad final obtenida). En el caso límite, tenemos un 100% de compresión, o lo que es lo mismo, no tendremos nada de información.

Por supuesto, siempre que se anuncia uno de estos «compresores mágicos» se habla de compresores lossless, que son los que presentan mayores dificultades. Conseguir un compresor lossy con un nivel de compresión 1000 a 1 es fácil (eso sí, la calidad se resentirá).

En adelante, hablaremos sólo de compresión lossless, que es la que nos interesa en este artículo.

SUSTITUCIÓN DE CÓDIGOS

Las ideas básicas que debemos tener en mente son las siguientes:

- Queremos conservar toda la información
 - La compresión lossless es la única que conserva toda la información
 - Los compresores lossless actúan eliminando la redundancia

En la introducción ya vimos una forma de eliminar redundancia: los algoritmos RLE. Veamos ahora otra forma: los algoritmos de sustitución de códigos (normalmente se usan los dos conjuntamente: RLE y sustitución de códigos).

La sustitución de códigos para comprimir una cadena C (formada por varios mensajes enlazados uno a continuación de otro) consiste en cambiar la forma en que se representa esa cadena. A la hora de decidir cómo se hará el cambio, lo fundamental es la probabilidad con que se presenta parte de la cadena (cada mensaje).

Veamoslo con un ejemplo. Imaginemos una fuente que emite 5 mensajes: M1, M2, M3, M4 y M5, con probabilidades 35%, 25%, 20%, 12% y 8% y longitudes (en caracteres) 10, 7, 6, 4 y 3.

Si dejamos que la fuente emita 1000 mensajes, teniendo en cuenta las probabilidades y las longitudes de los mensajes, la cadena emitida ocupará 7170 caracteres

$$\frac{35x1000x10}{100} + \frac{25x1000x7}{100} + \frac{20x1000x6}{100} + \frac{12x1000x4}{100} + \frac{8x1000x3}{100} = 7170$$

Como el mensaje M1 (representado como 1111111111) aparece mucho (35% de las veces) y es muy largo (10 caracteres) y el mensaje M5 (representado como 555) aparece poco (8% de las veces) y es corto (3 caracteres) una forma sencilla de comprimir consiste en intercambiar las representaciones de los mensajes:

Al hacer esta «compresión», la cadena de 1000 mensajes tiene una longitud de 5280 caracteres (hemos conseguido más de un 26% de compresión):

5555552222224444111111113333335555
55222222222222225555533333555333333
222222555551111111122222223333335
554444222222333335552222222222224
444444333332222222222333335551
111111115551111111115555533333222
222244445555...

(ahora los 555 representan a M1, no a M5)

Podemos hacer lo mismo otra vez, intercambiando las representaciones de los mensajes M2 y M4:

Msj	Representación								Prob.
M1	5	5	5						35%
M2	4	4	4	4					25%
M3	3	3	3	3	3	3			20%
M4	2	2	2	2	2	2	2		12%
M5	1	1	1	1	1	1	1	1	8%

Tras esta nueva vuelta de tuerca, la cadena de 1000 mensajes ocupa sólo 4890 caracteres (hemos conseguido un 32% de compresión respecto a la cadena original).

El código más conocido, el Morse, es un código de este tipo. La letra más frecuente, la *e*, se representa por un punto (es decir, la letra más frecuente es la más corta). Por desgracia para el español, la *e* es la más frecuente en inglés, mientras que la más frecuente en español es la *a*. Algo similar ocurre con la *W*, codificada como — (es decir, relativamente corta, lo que tiene sentido en inglés, pero ningún sentido en español).

Ahora que ya sabemos «grosso modo» cómo funciona la compresión por sustitución de códigos, vemos que unos mensajes disminuyen el espacio que ocupan (M1 y M2), otros lo aumentan (M4 y M5) y algunos no lo varían (M3). Como los mensajes que ocupan más aparecen poco, y los que ocupan menos aparecen mucho, el total del espacio ocupado es menor tras la compresión.

Pero, ¿qué ocurre cuando ya no se pueden hacer más cambios de representación? Dicho de otro modo, ¿qué ocurre si la información es completamente aleatoria (todos los mensajes tienen exactamente la misma probabilidad de aparición)? En este caso, no se puede comprimir la cadena, porque al hacer nuestros «intercambios de representaciones» no ganamos nada.

La conclusión obvia es que llega un momento en que se ha eliminado toda la redundancia (utilizando RLE) y ya no se pueden hacer intercambios de representaciones, así que no se puede comprimir más. Hemos llegado al «límite de compresión», que viene dado por la entropía de la cadena (o fichero) que deseamos comprimir, y cualquier intento de «recomprimir» la información ya comprimida será infructuoso.

Veamos un ejemplo sencillo de límite de compresión, con los mismos mensajes que antes, pero cambiando la probabilidad de aparición: ahora todos los mensajes son equiprobables (aparecen las mismas veces):

Msj	Representación								Prob.
M1	1	1	1	1	1	1	1	1	20%
M2	2	2	2	2	2	2	2		20%
M3	3	3	3	3	3	3			20%
M4	4	4	4	4					20%
M5	5	5	5						20%

Una cadena formada por 1000 mensajes tiene ahora 6000 caracteres. Intercambiemos las representaciones de M1 y M5:

Msj	Representación								Prob.
M1	5	5	5						20%
M2	2	2	2	2	2	2	2		20%
M3	3	3	3	3	3	3			20%
M4	4	4	4	4					20%
M5	1	1	1	1	1	1	1	1	20%

La cadena formada por 1000 mensajes vuelve a tener exactamente 6000 caracteres. No hemos conseguido nada. El lector puede comprobar que tampoco se gana nada

Antes de RLE	Después de RLE
555	?
4444	?
333333	?
222222	?
1111111111	?
	35
	44
	63
	72
	101

intercambiando las representaciones de M2 y M4, o haciendo cualquier otro cambio de representación. Lo único que quedaría por hacer es aplicar un algoritmo RLE para conseguir algo parecido a esto:

TRANSFORMACIONES

Las Matemáticas y la Física suelen recurrir a cambios de dominio (por ejemplo, pasar de dominio tiempo a dominio frecuencia) para facilitar determinadas operaciones.

Podríamos preguntarnos si haciendo algún procesado, tal como una Transformada de Fourier, una Transformada de Haar, o cualquier otro procesado de señal) a la cadena sin comprimir (preprocesado) o a la cadena comprimida (postprocesado) sería posible alcanzar una mayor compresión. Es decir, ¿cambia el límite de compresión al cambiar el dominio en que representamos la información?

La respuesta es «no». La razón es que la Teoría de la Información se basa en la probabilidad, y el modelo probabilístico ya incluye todas las posibles transformaciones que se puedan hacer, porque al cambiar

de dominio, lo único que cambiamos es el modelo probabilístico que usamos. Veamos la demostración formal (no es necesario entender la demostración para seguir leyendo el artículo; puede pasarlo por alto si lo desea). Dado un modelo probabilístico P y una función reversible M, siempre existe un modelo probabilístico Q tal que $Q(M(C)) = P(C)$, para cualquier cadena C.

Como M es reversible, para cualesquier cadenas C y D (tales que $C \ll D$), se cumple que $M(C) \ll M(D)$.

Podemos definir Q(C) tal que para cualquier cadena C, $Q(M(C)) = P(C)$, es decir, dos modelos probabilísticos diferentes (uno en cada dominio) nos dan el mismo resultado. Así pues, el cambio de dominio no mejora la entropía de la fuente y por tanto, no mejora la compresión.

CONCLUSIÓN

Como hemos visto a lo largo del artículo, llega un momento en que seguir aplicando algoritmos de compresión (RLE, sustitución de códigos, etc.) no disminuye el tamaño final, sino que lo aumenta: hemos llegado al límite de compresión. Esto se puede ver incluso de forma intuitiva: si siempre se pudiese seguir comprimiendo, llegaría un momento en que cualquier cadena original se reduciría a un único

AAAAA	→ DF	→ A
BBBBBBBBBBB	→ HRFD	→ B
CCCCCC	→ CFD	→ D
ADVCEFDD	→ CDF	→ D
AGHBCED	→ AS	→ A
ACVWREYIPGDD	→ WECB	→ B

A → ¿ DF → AAAA ?

¿ AS → AGHBCED ?

B → ¿ HRFD → BBBBBBBBBB ?

¿ WECB → ACVWREYIPGDD ?

carácter. ¿Cómo saber a cuál de las infinitas cadenas originales corresponde ése carácter? (es decir, a qué se tiene que descomprimir ése carácter)

Aunque las tecnologías de compresión usadas hoy en día no siempre son óptimas, cualquier anuncio de tecnología de compresión con límites de compresión muy superiores a los actuales es muy probablemente falso.

En cualquier caso, si una tecnología de compresión lossless que afirma superar el límite de Shannon, no hay duda de que mentirá, así que no vale la pena perder tiempo en ella. Ahora ya podemos contestar a la pregunta que se planteaba al inicio del artículo: no usamos esos compresores fantásticos porque no existen (ni existirán nunca: la analogía más clara sería una máquina de movimiento perpetuo).

Aún así, el lector osado puede aceptar el reto de Mike Goldman: 5000 dólares USA para quien consiga comprimir datos aleatorios, desafiando así el Teorema de la Complejidad de Kolmogorov (si se intenta comprimir un fichero de datos aleatorios, el tamaño del fichero comprimido más el tamaño del compresor siempre será mayor que el tamaño del fichero sin comprimir).

REFERENCIAS BIBLIOGRÁFICAS

[1] «Comunicación de datos»
J. R. Vidal Català, J. Martínez Zaldívar
Editorial de la Universidad Politécnica de Valencia, 1995

[2] comp.compression FAQ
Jean-Luc Gailly
<http://www.faqs.org/faqs/compression-faq/>

[3] There's no magic
Charles Bloom
<http://www.cbloom.com/news/nomagic.html>

[4] Information content and compression limit FAQ
Glyph
<http://www.geocities.com/CollegePark/9315/infofaq.htm>

ZeoSync, BinaryAccelerator, BitPerfect, Zero Space Tuner, Relational Differentiation Encoding y TunerAccelerator son marcas registradas de ZeoSync Corporation

Web Technologies y DataFiles/16 son marcas registradas de Web Technologies

Premier Research Corporation y Minc son marcas registradas de Premier Research Corporation.

Pegasus Technologies y HyperDrive son marcas registradas de Pegasus Technologies.

HyperSpace es una marca registrada de David C. James.

Pixelon es una marca registrada de Pixelon Corporation.

Agradecimientos especiales a Félix Ares y Jean René Moreau por sus valiosos consejos y correcciones.

AUTOR



Pau Garcia i Quiles, nacido en Elche (Alicante) en 1979. Ingeniero Técnico de Telecomunicación, especialidad de Telemática, por la Universidad Politécnica de Valencia. Actualmente haciendo en la UPV 2º de Ingeniería Informática y el Proyecto Final de Carrera de Ingeniería Superior de Telecomunicación ("Evaluación del algoritmo Rijndael para la generación de números pseudoaleatorios"). Intereses principales: Unix, seguridad informática y divulgación científica (dentro de la cual se enmarca este artículo).





SIMULACIÓN ELECTROMAGNÉTICA FD-TD: UNA AYUDA AL DISEÑO EN COMPATIBILIDAD ELECTROMAGNÉTICA

Alex Vidal, Gabriel Anzaldi, Ferran Silva

Grupo de Compatibilidad Electromagnética.

Dept. de Ingeniería Electrónica. UPC.

<http://www.upc.es/web/gcem/>

alex@eel.upc.es, gabriel@eel.upc.es ferran@eel.upc.es

INTRODUCCIÓN

El algoritmo FD-TD (Finite Difference Time Domain) es un método de simulación de campos electromagnéticos. La técnica FD-TD, y sus variantes como XFDTD, están presentes desde hace tiempo en muchas aplicaciones y puede compararse en madurez a técnicas ya consolidadas como el método de los momentos u otros algoritmos de elementos finitos [1].

La simulación ha dejado de ser algo exclusivo de los trabajos de investigación para pasar a ser una herramienta dentro de los procesos de diseño de un producto. De entre las opciones de simulación (resolución analítica, métodos numéricos y sistemas expertos), los métodos numéricos permiten una mayor flexibilidad. La potencia bruta de cálculo y el hecho de que no necesita resultados previos como los sistemas expertos, convierte la simulación EM en la herramienta más idónea al atacar problemas nuevos y de gran complejidad.

A lo largo del artículo veremos una introducción al algoritmo FD-TD. Se intentará dar una perspectiva del estado actual de la simulación y su aplicación en diseños reales.

EJEMPLOS DE APLICACIONES CON FD-TD.

La importancia de la simulación es cada vez mayor y con ello se toma más en serio su utilización. Hay dos claros ejemplos que lo corroboran: cada vez hay más productos comerciales y cada vez aparecen más resultados en publicaciones técnicas compartiendo protagonismo con resultados experimentales.

Dado su carácter temporal, FD-TD permite el cálculo de transitorios, resolución de problemas de scattering, propagación,... El buen comportamiento de FD-TD a alta frecuencia le permite estar presente en aplicaciones a EMC, análisis y síntesis de antenas, cálculo de Radar Cross Section (RCS), interacción de campos de telefonía móvil en seres vivos, diseño microelectrónico,...

Esta última es un área en el que las medidas son en ocasiones muy complicadas. La verificación cualitativa de las soluciones al diseño es un campo importante de aplicación de la simulación. Las fórmulas asociadas a los parámetros de capacidad e inductancia de las interconexiones suelen deducirse para estructuras aisladas y pueden ser muy imprecisas cuando se trabaja con estructuras más complejas. La simulación también es un recurso en este caso ya que puede evaluar la estructura completa. No solo permite ver los acoplamientos sino también las distribuciones de corriente que son importantes en la problemática actual de disipación de calor. Las figuras siguientes muestran un ejemplo. En este caso lo que se analiza es la distribución de la corriente en un integrado LM7171, montado sobre un PCB.

En la figura 1 podemos ver el modelo del interconexión de salida del integrado que está conectado sobre un PCB. En la figura 2 se observan los resultados obtenidos, donde las zonas más claras (verdes y rojas) corresponden a una densidad de corriente mayor. Además este ejemplo también muestra otra potencialidad de estos métodos: la utilización de diversas técnicas de simulación simultáneas. El *core* del circuito (que aparece en el centro de la figura 1) corresponde a un modelo PSPICE del circuito integrado, al mismo tiempo en la simulación EM se definen puertos de entrada en los que se calcula V o I. Estos parámetros se introducen como excitación en un nodo de un circuito SPICE que se simula al mismo tiempo de modo que esta excitación produce una respuesta en el circuito que puede utilizarse como condición de contorno en otro punto del modelo físico.

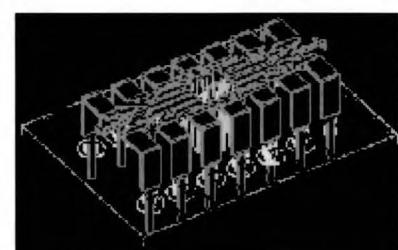


Figura 1. Modelo de un circuito integrado LM7171 sobre una PCB.

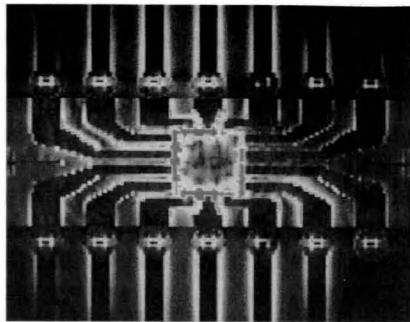


Figura 2: Distribución de corrientes en la estructura anterior.

Las técnicas de FD-TD también están siendo aplicadas a problemas de muy alta frecuencia [2]. La aplicación de FD-TD ya no solo corresponde al dominio radioeléctrico sino incluso se utiliza a problemas en el dominio óptico [3] asociados por ejemplo a la dispersión de la luz en materiales o en LASERS.

Otro ejemplo destacable fue el proyecto realizado para SAAB en aviática sobre el estudio de las corrientes en el fuselaje debidos a la caída de rayos durante el vuelo. El modelo ocupa un volumen de 30x30x15 m y se ha discretizado a un tamaño de celda de 2.5cm. Estos valores implican un número de celdas que supera los 1000 millones. En total se realizaron 2500 pasos de simulación con un timestep de 83.3ps. La precisión y el volumen de celdas requiere un total de 64Gbytes de RAM. El tiempo de simulación necesario fue de 66 min en un paralelizador de 125 procesadores RS/600 a 160MHz. Los valores de corriente obtenidos y el campo magnético radiado se muestran en la figura 3.

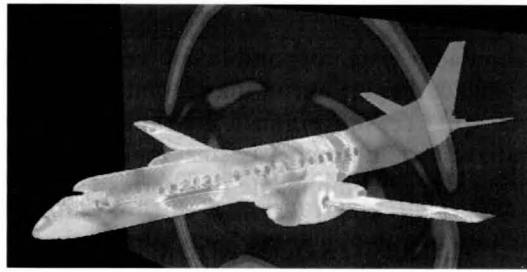


Figura 3. La simulación permite estudiar las corrientes de un rayo circulando por el fuselaje.

REVISIÓN DEL MÉTODO NUMÉRICO

La revisión que se da en este punto solo pretende ser una pincelada de los fundamentos matemáticos de FD-TD [4].

Las ecuaciones de Maxwell en caso de no existir fuentes de campo eléctrico ni magnético pueden reducirse a:

$$\nabla \times H = J + \epsilon \frac{\partial E}{\partial t} \quad (1)$$

$$\nabla \times E = -\mu \frac{\partial H}{\partial t} \quad (2)$$

donde ϵ es la permitividad eléctrica y μ la permeabilidad magnética. Para un medio con conductividad no infinita podemos expresar: $J = \sigma E$. El desarrollo de estas ecuaciones nos da un sistema acoplado en que E depende las componentes de H y viceversa. De forma cualitativa, podemos decir que FD-TD substituye los operadores derivada por incrementos, pasándose así al dominio discreto.

$$\frac{\partial H_x}{\partial t} = \frac{1}{\mu} \left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y} \right) \quad (3a)$$

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu} \left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} \right) \quad (3b)$$

$$\frac{\partial H_z}{\partial t} = \frac{1}{\mu} \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right) \quad (3c)$$

$$\frac{\partial E_x}{\partial t} = \frac{1}{\epsilon} \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - \sigma E_x \right) \quad (3d)$$

$$\frac{\partial E_y}{\partial t} = \frac{1}{\epsilon} \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - \sigma E_y \right) \quad (3e)$$

$$\frac{\partial E_z}{\partial t} = \frac{1}{\epsilon} \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - \sigma E_z \right) \quad (3f)$$

Para solventar este problema de interdependencia de E y H , Yee propuso en 1966 [5] una discretización diferente para calcular el campo eléctrico y magnético. Esta nueva discretización se da tanto en el dominio temporal como espacial. Yee definió lo que se conoce como celda Yee. La figura 4 representa los puntos en los que se calcula campo eléctrico y magnético.

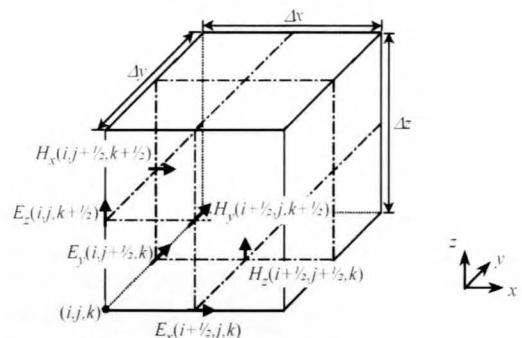


Figura 4. Celda Yee.

Cada punto en el que se calcula E está rodeado de 4 puntos en los que se conoce H , por lo que de se puede obtener el

valor de campo tanto por las expresiones en forma diferencial como a través de su forma integral. Considerando el campo E en el paso de simulación n, el algoritmo realiza el cálculo de los valores de campo magnético para un instante de tiempo entre dos pasos de simulación n+1/2. De forma similar, se realiza el cálculo del campo E en el instante n+1 a partir de los valores de H calculados anteriormente. Esto demuestra una importante característica del algoritmo: su natural paralelización. Dado que el cálculo solo depende de los valores de E o H anteriores, este se puede realizar de forma paralela para diversos puntos.

El algoritmo permite la definición de las características de cada celda por separado (e, m y s). Incluso pueden introducirse materiales anisótropos [6].

Para que el algoritmo converja se han de considerar unas condiciones tanto para los incrementos espaciales como temporales. Por lo que respecta a las condiciones espaciales, existe la recomendación de no hacer la celda mayor de 1/20 parte de la mínima longitud de onda presente en el problema. En muchas ocasiones esta cota es mucho menos restrictiva que las propias condiciones de resolución que se desean para el modelo a discretizar. Otro aspecto importante en FD-TD es la interrelación entre los incrementos temporales y espaciales: una discretización fina obliga a un cálculo con incrementos de tiempo pequeños. Esto es debido a la velocidad de propagación de los campos y su efecto en las celdas vecinas. Las condiciones anteriores se resumen en la condición de Courant, que garantiza la convergencia:

$$\Delta t \leq \frac{1}{v_{\max} \sqrt{\left(\frac{1}{\Delta x}\right)^2 + \left(\frac{1}{\Delta y}\right)^2 + \left(\frac{1}{\Delta z}\right)^2}} \quad (4)$$

donde Dt será el incremento de tiempo (discretización temporal). Dx, Dy, Dz corresponden a la resolución de la discretización espacial. Y por último v_{\max} será la máxima velocidad de propagación de la luz en el modelo y que dependerá de las características físicas de los materiales utilizados.

Por último hay que destacar la problemática de las Absorbing Boundary Conditions (ABC). En el modelo deben definirse unas condiciones de contorno en los límites del mismo. Si no fuese así existirían reflexiones de potencia que no aparecerían en el caso real para propagación en espacio libre. Las ABC tienen la función de realizar una adaptación de las impedancias de los materiales del modelo y el medio ficticio correspondiente a su extensión. Principalmente se utilizan dos técnicas: las condiciones de absorción de MUR [7], y las PML (Perfectly Matched Layers) [8].

Por último comentar que de la misma forma que FD-TD es muy utilizado a alta frecuencia, una limitación importante del método son las simulaciones a baja frecuencia. La resolución temporal (que es función de la espacial) puede ser muy pequeña requiriendo tiempos de simulación demasiado largos.

UNA HERRAMIENTA DE SIMULACIÓN: LC.

LC nació en el seno de Silicon Graphics con el objetivo de tener una herramienta FD-TD propia que pudiese servir de ayuda en el diseño de sus supercomputadores. En este aspecto la simulación puede ayudar a obtener resultados cualitativos sobre las distribuciones de corriente y acoplamiento en modelos sencillos del interconexión interno.

LC es un simulador lineal que permite la caracterización de materiales isótropos a partir de su permitividad eléctrica y permeabilidad magnética. Permite también introducir pérdidas en el modelo a partir de la conductividad eléctrica del material. Incluso en la última versión se puede definir la conductividad magnética que permite modelar materiales que disipan la energía del campo magnético en forma de calor.

Otro de los refinamientos importantes que tiene el método FD-TD es la posibilidad de una discretización no uniforme del modelo permitiendo así mejorar la resolución del modelo en zonas donde es necesario, mitigando el problema de la memoria excesiva. En LC esta opción aparece bajo el nombre de *submesh*.

El simulador está disponible en versiones para paralelizadores pero existe una versión para LINUX, de la que recientemente ha aparecido una actualización que permite sistemas multiprocesador. El programa se compone de tres partes:

-**EL PRE-PROCESADOR:** es un editor simple que permite la definición de la geometría así como de las características de los materiales; las fuentes de excitación y los puntos (o superficies) de prueba de campos E-H, tensiones o corrientes. También permite definir parámetros de la simulación como el método de truncado del dominio de simulación o el tiempo de la misma.

-**EL SOLVER DEL SIMULADOR.** Esta parte realiza la discretización del modelo e implementa el algoritmo computacional de FD-TD.

-**EL POST-PROCESADOR:** permite la visualización de los resultados. Además permite realizar operaciones con estos para obtener información adicional como diagramas de radiación o distribuciones de líneas equipotenciales. También

permite la obtención de los parámetros S y las capacidades e inductancia de estructuras.

Algunos productos de simulación comercial no desarrollan mucho el primer y tercer componente. Son más flexibles en estos aspectos permitiendo importar los modelos de formatos estándar como AUTOACAD en lugar de crear su propio formato. De esta forma pueden adaptarse a las herramientas ya disponibles en el entorno de trabajo. Un ejemplo de estas herramientas de pre y post-procesado es el GID [9], desarrollado en el CIMNE, el Centro Internacional de Métodos Numéricos en Ingeniería de la UPC. Lo mismo ocurre para las herramientas de post-procesado como MATLAB. Incluso existen solvers para complementar a este último. Pero LC, al combinar las tres partes, resulta una herramienta muy intuitiva y completa para iniciarse en la simulación.

Actualmente LC ha pasado a ser de CRAY RESEARCH INC al igual que Silicon Graphics. Se puede obtener más información sobre él y sobre como conseguir la distribución gratuita en <http://lc.cray.com>.

En la figura 5 aparece una aplicación extraída de esta propia web en la que se estudia la distribución de corrientes en un circuito integrado.

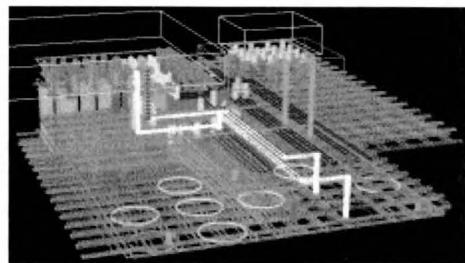


Figura 5. modelo de un circuito integrado sobre el que se obtienen densidades de corriente.

Es importante ver que una empresa como Silicon Graphics considerara la simulación como una herramienta más en el desarrollo de sus productos.

Por último, destacar que dentro del Grupo de Compatibilidad Electromagnética (GCEM) de la UPC, se han desarrollado diversos productos de instrumentación para medidas de EMC y que han tenido aportaciones de la simulación en su diseño. En concreto, se ha desarrollado una sonda para la medida en campo próximo [10] [11] y actualmente se trabaja en obtener una sonda isotrópica. Antes de construir los prototipos la simulación se utilizó para obtener diferentes parámetros de la sonda: impedancia de entrada, sensibilidad en frecuencia o diagrama de recepción. En la figura 6 se puede comparar el modelo real y el de simulación.

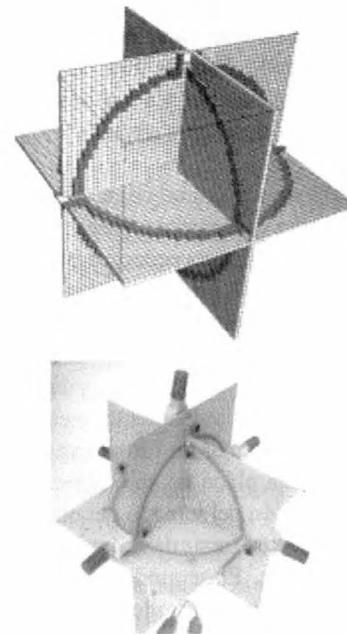


Figura 6: Arriba, el modelo simulado de la sonda isotrópica. Abajo, el modelo real montado en PCB.

UNA APLICACIÓN PRÁCTICA: GENERACIÓN DE UNA ONDA PLANA.

La onda plana es un valioso recurso a la hora evaluar la respuesta de un elemento al efecto de un campo lejano. La obtención del diagrama de radiación, la sensibilidad, los ajustes en las impedancias de carga,... de una antena o la RCS son parámetros que pueden obtenerse a partir de la iluminación de la estructura con una onda plana. Pueden considerarse dos caminos en su implementación en simuladores: la generación numérica y el modelado completo de un sistema real que la genere.

La generación numérica de una onda plana no es nueva en los algoritmos de FD-TD. Por cuestiones de seguridad nacional, el simulador LC solo permite esta opción en distribuciones para su uso dentro de EEUU, ya que a partir de simulación podrían obtenerse firmas Radar por ejemplo. Por este motivo, para poder utilizar ondas planas en LC en otros países se debe generar la onda mediante la simulación completa de un sistema que radie.

Una de las posibilidades sería la utilización de una antena patrón como elemento radiante, a una distancia suficiente del elemento a ensayar para considerar campo lejano. Como patrón puede usarse un dipolo resonante. Esta solución presenta diversos inconvenientes:

-Problemas con el método. La respuesta del dipolo tiene una fuerte dependencia de la frecuencia. En si esto no es un problema en la simulación, ya que puede corregirse calibrando. Pero es más razonable considerar un método cuya excitación sea uniforme en la banda de frecuencias a considerar.

-Problemas de simulación. La separación entre las antenas emisora y receptora presenta varios problemas. En primer lugar, la distancia necesaria puede ser grande en términos de número de celdas por lo que el consumo de memoria necesaria puede dispararse. Esto se agrava si la resolución espacial necesaria para el objeto a ensayar tiene que ser pequeña. Por otra parte, y dado que el tiempo de simulación es aproximadamente lineal con el tamaño, éste se puede incrementar de forma exagerada. Además, conseguir una onda plana no solo depende de la antena emisora sino también del tamaño del objeto a iluminar y sobre el que se desea un frente de onda plano (equipotencial) en todos sus puntos. El segundo problema es que esta situación corresponde a un modelo abierto. Esto requiere de condiciones de contorno absorbentes que permitan la simulación sin retorno de potencia al modelo. Realmente la adaptación nunca es perfecta y pueden aparecer reflexiones. Para evitar este efecto se recomienda dejar un espacio de 1 o 2 longitudes de onda entre el modelo y el límite de este en todas las direcciones, cosa que agrava el problema de la memoria ya comentado. Si no se tienen en cuenta todos los aspectos anteriores, los resultados que se obtienen divergen en el caso de que la potencia reflejada sea grande o el espacio libre alrededor de los elementos sea insuficiente.

En nuestro caso, el modelo escogido para la generación es la estructura TEM [12]. El funcionamiento de una celda TEM puede verse de modo intuitivo igual que el funcionamiento de un cable coaxial. El cable coaxial se comportará como una línea de transmisión sobre la que se propaga una onda electromagnética. El ensanchamiento de este coaxial y su conductor central permiten crear un volumen en el que se tiene un campo aproximadamente uniforme y líneas de campo paralelas. En el caso de la celda TEM ambos requerimientos (memoria, y las condiciones de absorción ABC's) se relajan. Por lo que respecta a la potencia, la TEM es una estructura completamente cerrada que confina el campo generado. De esta forma se elimina la necesidad de aire circundante reduciéndose la memoria necesaria. En el caso de la generación de onda a partir de un dipolo existe además la restricción del tamaño del DUT. Si el objeto es muy grande, es necesario alejarlo de la antena patrón para asegurar el frente de onda. En el caso de la TEM, si el DUT cabe dentro de la zona de campo uniforme, la memoria necesaria es independiente del tamaño de este, ya que el volumen total se mantiene constante.

En la figura 7 se representa el modelo de celda con un corte transversal que permite ver el objeto a ensayar

colocado en su interior. El modelo consta de tres componentes: el coaxial del generador, el coaxial de la carga, y la parte central de transmisión de la potencia. Cada uno de estos tres elementos puede variar la respuesta frecuencial de la celda.

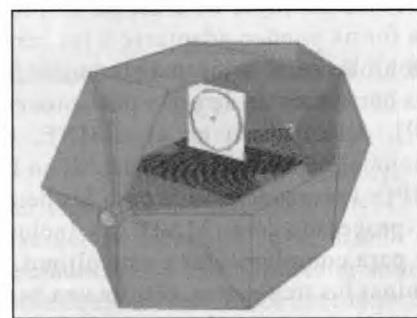


Figura 7. Sección de la celda TEM utilizada. En su interior se coloca el elemento a iluminar (en este caso una sonda de campo cercano).

La modificación de la geometría de un coaxial para obtener la TEM conlleva modificaciones en la distribución y propagación del campo. Esto se mejora si tanto la impedancia del generador como la carga coinciden con la impedancia característica de la línea. En caso contrario las reflexiones que se produzcan en la carga pueden variar su respuesta frecuencial y crear ondas estacionarias.

En la creación del modelo se ha tenido en cuenta que deben minimizarse estos efectos. En lo que respecta a los coaxiales de entrada y de terminación, inicialmente se consideró una fuente de tensión y una carga que unen el conector central y la carcasa por un único punto. Esta asimetría también introduce variaciones en la distribución de los campos. Se ha optado por poner dos fuentes de igual amplitud en paralelo (ya que ambas conectan el conductor central y la carcasa exterior) como excitación. Este problema de simetría también se traslada a la carga donde se han colocado 4 cargas de 200W en paralelo para obtener una impedancia de 50W. Las cargas se distribuyen formando una estrella entre el conductor central y la carcasa exterior. La figura 8 muestra ambos coaxiales.

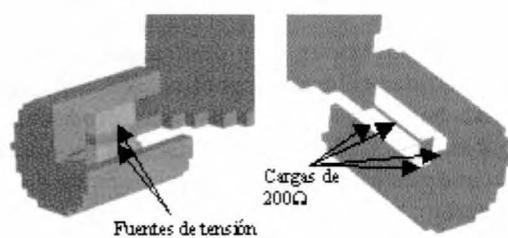


Figura 8. Cortes de los coaxiales de excitación (izq) y de la carga (dcha).

Una vez construida la celda, se pasó a verificar su funcionamiento. Para ello se utilizó un pulso con una respuesta plana en la zona del espectro de interés. Una

matriz de puntos de prueba en el centro y laterales de la celda permite determinar el volumen útil en la que puede considerarse una onda plana. Finalmente se obtuvo un volumen de 10x10x10 cm dentro de la celda en la que el campo puede considerarse dentro de una dispersión de $\pm 5\%$, y que puede observarse en la figura 9.

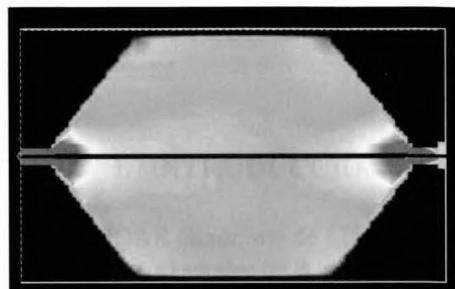


Figura 9: Distribución del campo eléctrico en un plano transversal de la celda.

La respuesta de la celda puede considerarse plana hasta 300 MHz. Este valor depende del tamaño de la celda: si la celda es menor, su frecuencia de corte es mayor pero disminuye su volumen útil.

Los resultados obtenidos, figura 10, demuestran un buen comportamiento de la celda. La componente principal (componente z) difiere en más de un orden de magnitud de las componentes residuales de campo en cuadratura (componente y) y transversales al frente de onda (componente x).

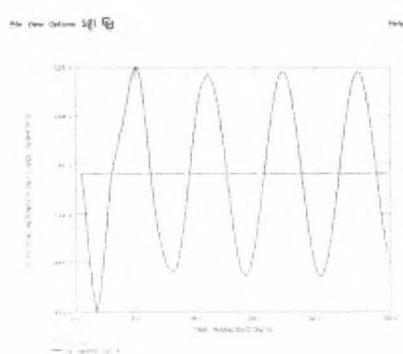


Figura 10: Respuesta a campo E de la TEM. La componente principal es claramente superior a las componentes residuales.

La memoria necesaria se acerca a los 70Mb para un modelo que tiene del orden de 1 millón de celdas. Este valor se ha obtenido discretizando la mayor parte del modelo a 5 mm y utilizando submeshes (zonas de discretización más fina) de 2 mm en el volumen útil mencionado anteriormente. El timestep necesario es de 3.3 ps siendo el tiempo de simulación total de 43ns. Cada una de las simulaciones lanzadas se completa en un tiempo aproximado de 8 horas. Es

importante destacar que se utilizó en la simulación un ordenador personal, un PIII a 766MHz y 1Gb de RAM.

CONCLUSIONES

La simulación de campos electromagnéticos es una técnica cada vez más usada y ya no pertenece solo al dominio de la investigación. Puede ser importante en la evaluación cualitativa de cambios en un diseño, reduciendo así el número de prototipos e incluso el *Time To Market*. Aun así los aspectos más cuantitativos aún están en una fase de verificación: los diseñadores extraen de la simulación «reglas de diseño» mas que datos numéricos.

Actualmente estas herramientas ya no se relegan a grandes corporaciones provistas de recursos como supercomputadores. Un ordenador personal puede llegar a ser una plataforma suficiente para realizar proyectos de simulación permitiendo a las pequeñas empresas el acceso a estas técnicas.

LC es una herramienta perfecta para realizar los primeros en aspectos de simulación. También adolece de diversas limitaciones, que en algunos casos pueden ser superadas como se ha visto con la necesidad de generar una onda plana y la construcción de la celda TEM.

Por último destacar la gente, tanto de desarrollo como de investigación, implicada en estos temas que demuestran las posibilidades y la confianza que se ha puesto en este nuevo recurso que es la simulación EM.

REFERENCIAS

- [1] Q. Li, Y. Chen, and D. Ge, «Comparison Study of the PSTD and FDTD Methods for Scattering Analysis», *Microwave Opt. Technol. Lett.*, Vol. 25, no. 3, pp. 220-226, Mayo 2000.
- [2] Young-Seek Chung; Joo-Hiuk Son; Changyul Cheon; Song-Yop Hahn; «Propagation characteristics of free-space terahertz electromagnetic pulses», *Lasers and Electro-Optics*, 1999. CLEO/Pacific Rim '99. The Pacific Rim Conference on , Vol. 4 , pp 1111 -1112 ,1999
- [3] Yamauchi, J.; Kanbara, H.; Nakano, H.; «Analysis of optical waveguides with high-reflection coatings using the FD-TD method» *IEEE Photonics Technology Letters* , Vol. 10, pp 111-113, Jan 1998.
- [4] A. Taflove «Advanced in Computational Electrodynamics, The finite-Difference Time-Domain Method», Artech House, 1998.

- [5] Yee, K. S., «Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media» IEEE Trans. Antennas and Propagation, Vol. 14, pp 302-307, 1966
- [6] Materdey, T.; Gonzalez, S.; Gomez, R.; Garcia, B, «A new vision of numerical methods for the solution of Maxwell's equations related to the FD-TD method; application to general anisotropic media» . Computation in Electromagnetics, 1994. Second International Conference on , pp 138 –141, Apr 1994.
- [7] G.Mur, «ABC for the finite difference approximations of the time-domain electromagnetic field equations», IEEE trans. EMC, Vol. 23, pp 377-382, Nov 1981.
- [8] J.P. Berenguer, «A PML for the absorption of electromagnetics waves», J Computational Phys., Vol. 114, pp 185-200, 1994.
- [9] Ramón Ribó Rodríguez «Desarrollo de un sistema integrado para tratamiento de geometría y generación de malla y datos para el análisis por el método de los elementos finitos». Tesis doctoral. Departamento de resistencia de materiales y estructuras de la Universitat Politècnica de Catalunya. Febrero de 2002.
- [10] Quilez, M.; Silva, F.; Martin, A.; Fontanilles, J.; Riu, P.; «Radiation pattern of a PCB based probe intended for near field measurements», Electromagnetic Compatibility, 1999 IEEE International Symposium on , Vol. 2, pp 879-881, 1999.
- [11] Silva, F.; Sanchez, F.; Riu, P.J.; Pallas-Areny, R.; «Low-cost near-field probe for simultaneous E and H measurement with analog optical link», Electromagnetic Compatibility, 1997. IEEE 1997 International Symposium on , pp 533-536, 18-22 Aug 1997.
- [12] Crawford, M.L «The generation of standard EM field using TEM transmission cells», IEEE Trans on EMC-16, pp 189.



REMOTE SENSING MEDIANTE LIDAR EN LA UPC

**F. Rocadenbosch, M. Sicard, A. Comerón, A. Rodríguez, C. Muñoz,
M.A. López, D. García, R. Agishev**

Dep. Teoría de la Señal y Comunicaciones (TSC), UPC. ETSETB, Módulo D4, Desp. 016. Email:
roca@tsc.upc.es; msicard@tsc.upc.es; comeron@tsc.upc.es

1. INTRODUCCIÓN

Las técnicas LIDAR (acrónimo de LIgth Detection And Ranging) permiten el sondeo de la atmósfera (aunque no necesariamente limitado a éste) mediante haces láser.

El sondeo remoto de la atmósfera mediante haces de luz (grandes reflectores) antecede la propia invención del láser. Fue, sin embargo, el nacimiento de este último en los años 1960, el que, ofreciendo anchos espectrales muy reducidos (< 0.01 nm) y una directividad sin precedentes, supuso un gran avance para estas técnicas de teledetección activa. Así, en 1962, Fiocco y Smullin [1] rebotaron por primera vez un haz láser en la Luna: Había nacido el LIDAR... Ese mismo año, McClung y Hellwarth inventaban el Q-switch, tecnología que permitía generar pulsos de luz muy cortos (≈ 10 ns) y con elevadas energías de pico (≈ 1 J), lo cuál se traduce en potencias pulsadas del orden de 100 MW. Posteriormente, en 1973, ya se tiene constancia de un primer lidar utilizando un láser semiconductor de AsGa [2].

En la actualidad, los sistemas lidar (Fig. 1) representan la contrapartida más cercana a los radares de microondas, con la salvedad que la radiación de RF se ha sustituido por otra en las bandas ópticas y que las resoluciones espaciales resultan muy superiores, del orden de unos pocos metros. El espectro de longitudes de onda de trabajo puede comprender desde el ultravioleta (UV) hasta el infrarrojo (IR).

Como una realidad imparable, los sistemas lidar están empezando a reemplazar los sensores *in situ* (p.ej., sondas locales tales como anemómetros, higrómetros y termómetros) y radiosondas (instrumentos embarcados en globos sondas controlados por radio) [3][4].

Las aplicaciones de los sistemas lidar son enormes y comprenden desde ambiciosos proyectos espaciales como LAWS (Lidar Atmospheric Wind Sounder, NASA) y ALADIN (Atmospheric LASer Doppler INstrument, ESA) orientados a realizar mapas de viento a escala mundial, hasta sistemas terrestres y embarcados (diversos proyectos en el NOAA (National Oceanic and Atmospheric Administration) y en la UE fruto de diversos Programas Marco de I+D e iniciativas privadas) que abarcan tanto aplicaciones meteorológicas como medioambientales:



Figura 1. Primera estación lidar UPC, operación nocturna (1996, V. Sec.2.1)

Las primeras utilizan los lidares para derivar la densidad de aerosoles, velocidad y dirección del viento (el estado del arte actual permite alcances en torno a 20 km de altura con resoluciones mejores de 3 m/s) [5], concentración de especies químicas (con lo que ello supone desde el punto de vista regulador), variables físicas (en particular, perfiles de temperatura hasta unos 100 km de altura con precisiones de 1 K) y estudios ceilometría (altura y espesor de nubes, por la importancia de éstas en el balance radiativo tierra-atmósfera).

En el campo medioambiental, los sistemas lidar permiten detectar y monitorizar la dispersión de aerosoles (Fig. 2) y efluentes contaminantes como O_3 , SO_2 , NOX (óxidos de nitrógeno), CO, CO_2 , H_2S , CH_4 y C_6H_6 , entre otros, emitidos desde plantas industriales y químicas hasta el receptor crítico, a distancias típicas de 5 km, con sensibilidades de unas pocas ppb -partes por billón- y resoluciones de 10 m. Los lidares embarcados en barcos y helicópteros también ofrecen interesantes aplicaciones de batimetría láser (estudio del fondo marino) en regiones costeras.

En condiciones climáticas adversas como las que se dan bajo intensa niebla o nubes bajas, los lidares no pueden operar debido a la elevada extinción atmosférica. Sin embargo, cuando éstos se combinan con instrumentación «cooperativa» como redes de satélites, radares y sensores locales, es posible implementar redes de observación a escalas incluso continentales (con resoluciones y

sensibilidades muy notables) y acercarse a lo que viene en llamarse «all weather sounding» (sondeo en «casi» cualquier condición climática) [6].

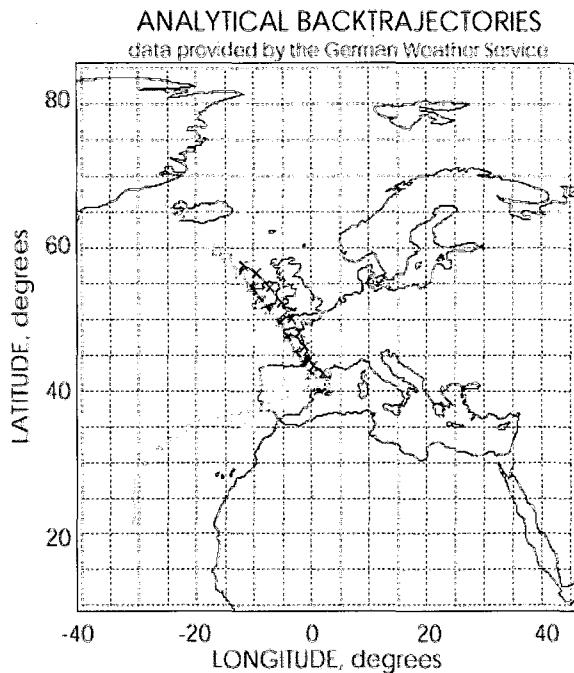


Figura 2. Mapa de trayectorias representativo de los distintos flujos de aerosoles y sus orígenes con destino Barcelona. Cada marca representa un retroceso de 6h (hasta 96 h en total) y cada color un origen o especie diferenciada (cortesía del German Weather Service).

2.SISTEMAS LIDAR

Los radares láser atmosféricos o lidar se basan en la interacción entre el haz láser emitido (generalmente pulsado) y los aerosoles y moléculas atmosféricos. Esta interacción se haya sujeta a las leyes de dispersión o *scattering* (Rayleigh, Mie y Raman, principalmente) y de absorción, de forma que esta interacción se manifiesta en una extinción o pérdida exponencial de energía conforme el haz láser se propaga por la atmósfera de acuerdo con la ley de Beer (también conocida como ley de Bouger). Así, la intensidad de un haz propagándose en la dirección \mathbf{r} en un medio inhomogéneo entre $r=0$ y $r=R$ viene dada por

$$\frac{I(\lambda)}{I_0} = T(\lambda, R) = \exp\left[-\int_0^R \alpha(r, \lambda) dr\right] \quad (1)$$

donde I_0 es la intensidad inicial en $r=0$, I es la intensidad en $r=R$, λ es la longitud de onda de trabajo, α es el coeficiente de extinción atmosférico [km^{-1}] y $T(\lambda, R)$ es la transmitancia en $[0, R]$. Es evidente que la extinción atmosférica condiciona los parámetros de diseño del lidar en un doble sentido:

Por un lado, es necesario aprovechar las ventanas espectrales de transmisión de la atmósfera a fin de que el haz láser no se vea muy atenuado (ventanas en 0.4-0.7 μm (visible), 0.7-1.5 (infrarrojo cercano, NIR), y ventanas en 3-5 μm y 9-13 μm), lo cuál condiciona a su vez la disponibilidad de láseres y tecnologías adecuadas para cada aplicación lidar, en particular, que sean seguros ocularmente ($>1.4 \mu\text{m}$, como es el caso de los láseres a 1.55 μm). En recepción, los lidares utilizan un telescopio a modo de «antena óptica» seguida de un sistema optoelectrónico de detección, el cuál componen fotodetectores que también deberán ser adecuados a la banda o subbanda de interés (PMTs (típ. 0.2-0.7 μm), PINs y APDs (típ. $\lambda > 0.8 \mu\text{m}$), VAPDs, detectores térmicos, etc.).

Por otro, hay que notar que la extinción es el resultado combinado de tres procesos que actúan simultáneamente en la atmósfera: a) absorción de energía por parte de las moléculas presentes, b) dispersión o scattering molecular y, c) dispersión por parte de partículas (aerosoles) en suspensión. Formalmente,

$$\alpha = \alpha_{g,abs} + \alpha_{g,sca} + \alpha_{p,sca} \quad (2)$$

donde el subíndice «g» indica gases (componente molecular), «p» partículas (aerosoles), «abs» absorción y «sca» scattering. En consecuencia, la longitud de onda de operación del instrumento lidar debe elegirse adecuadamente para magnificar la interacción óptica de interés. Además, esta interacción podrá ser elástica o inelástica, según que la longitud de onda en recepción coincida o no (caso del scattering Raman) con la del láser emisor.

Con estas premisas, para un sistema lidar elástico o *backscatter lidar*, la ecuación lidar toma la forma [7]

$$P(\lambda, R) = \frac{K(\lambda)}{R^2} \beta(\lambda, R) T(\lambda, R)^2 \xi(R) \quad (3)$$

donde $P(R)$ es la potencia recibida en función de la distancia, $T(\lambda, R)$ es la transmitancia atmosférica (Eq.(3)) en el camino de ida del haz láser (el cuadrado asimila los caminos de ida y vuelta), $\beta(\lambda, R)$ es el coeficiente de retrodispersión (o *backscatter*) atmosférico [$\text{km}^{-1}\text{sr}^{-1}$], que no es más que una «reflectividad» a la longitud de onda λ (específicamente, la sección recta de retrodispersión por unidad de volumen y ángulo sólido), $K(\lambda)$ es un factor de calibración función de la energía emitida y área equivalente en recepción y, $\xi(R)$ es un factor de corrección que tiene en cuenta el solapamiento óptico en función de la distancia R entre la sección recta iluminada por el láser y la verdaderamente captada por el telescopio dentro de su campo de visión (V. detalle en Fig. 3).

2.1. Sistemas elásticos

Como se ha dicho, en un sistema elástico o *backscatter lidar*, las longitudes de onda emitida y recibida coinciden, de forma que la interacción entre el haz láser y la atmósfera es relativamente fuerte. Si bien este hecho no permite identificar especies químicas, el retorno recibido en condiciones de buena visibilidad es básicamente proporcional a la mezcla de aerosoles y partículas en suspensión, lo cuál es con todo suficiente para, mediante una arquitectura relativamente poco compleja, invertir parámetros medioambientales básicos tales como la altura de la capa de mezcla (o PBL, una capa formada por aerosoles bien mezclados que se extiende desde el suelo hasta unos 2-3 km de altura y muy influenciada por los ciclos diurnos/nocturnos, vientos y fenómenos de transporte como, por ejemplo, las inyecciones de masas debidas a incendios y a invasiones de polvo en suspensión), concentración/flujo de aerosoles en efluentes tales como chimeneas, intensidad de lluvia, nieve, etc. [8].

La Fig. 3 muestra la arquitectura de los diferentes subsistemas de emisión y recepción partiendo de la que ya en 1996 fue la primera estación lidar fija de la UPC. Estos se distribuían entre la cúpula y la sala de control (localizada en la tercera planta del edificio D3). El sistema emisor estaba basado en un láser de estado sólido (Nd:YAG, sala de control) que emitía pulsos de 10 ns de anchura y unos 400 mJ de energía a 532 nm (verde) con una frecuencia de 10 Hz (obsérvese que estas especificaciones se traducen en potencias de pico de 40 MW; el haz láser podía verse fácilmente desde la Pza. M^a Cristina en la Av. Diagonal de Barcelona). El sistema receptor propiamente dicho se localizaba en la cúpula y estaba formado por un telescopio Schmidt-Cassegrain de 20 cm de apertura y 2 m de distancia focal y un receptor optoelectrónico que, ensamblado en el plano focal de aquél y en sustitución su ocular, transducía las señales ópticas recibidas (Eq.(3)) en eléctricas. La cúpula también albergaba diversa instrumentación de polarización y monitorización. Finalmente, la sala de control alojaba, además de la fuente láser emisora, el equipo de adquisición y control, que digitalizaba las señales recibidas a 20 Msps (esto se traduce en resoluciones espaciales tan impresionantes como 7.5 m) y controlaba diversos subsistemas, el propio láser y un generador de retardo a modo de unidad de sincronismo. El haz láser se dirigía verticalmente mediante un espejo deflecto en la misma sala y el sistema tenía un alcance máximo (SNR(R_{\max})=1) de unos 15 km en condiciones de buena visibilidad ($\alpha \approx 0.01 \text{ km}^{-1}$) y 5 s de integración de pulsos. El sistema de control (implementado en LabView[®]) ya permitía extraer semicuantitativamente las características atmosféricas más significativas en tiempo real a partir de un simple procesado de la forma $R^2P(R)$ sobre la Eq.(3), como se muestra en la Fig. 4. A diferencia de los sistemas elásticos más clásicos como el descrito, que utilizan pulsos muy cortos de alta energía para interrogar la atmósfera, la tendencia tecnológica

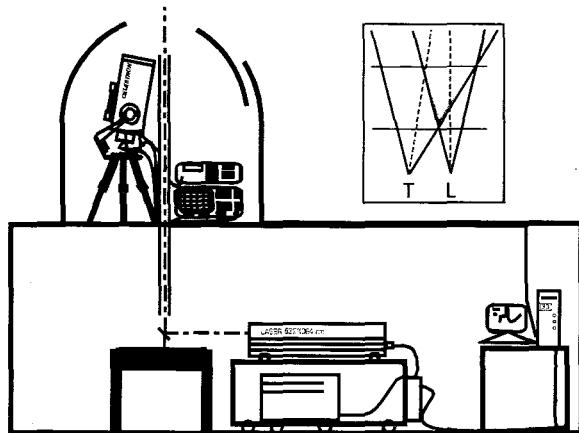


Figura 3. Esquema de la estación lidar elástica de la UPC (1996) mostrada en la Fig. 1 (y V. Sec.2.1).

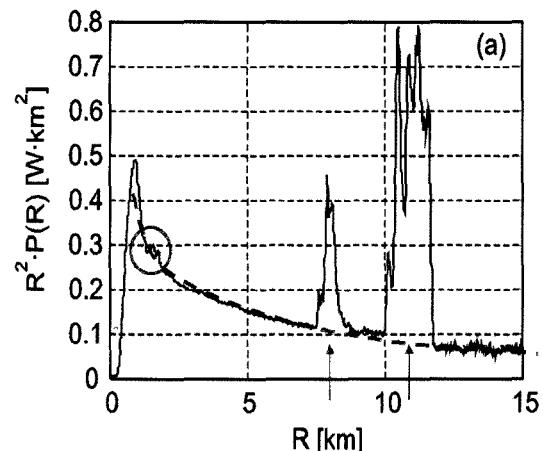


Figura 4. Ejemplo de una señal lidar corregida en distancia. En trazo discontinuo, retorno Rayleigh debido a la concentración de moléculas en la atmósfera. Las flechas indican la presencia de dos capas de nubes localizadas entre 7.5-9 km y 10-12 km, respectivamente, y el círculo una inhomogeneidad en la capa de mezcla debida a la advección de aerosoles a unos 2 km de altura.

actual es hacia lo que se conoce como *microlidares*, sistemas pseudoaleatorios de baja potencia, que básicamente modulan la portadora óptica mediante secuencias pseudoaleatorias con determinadas propiedades de ortogonalidad y que merced a este esquema de modulación evitan tener que utilizar potencias pico tan elevadas en emisión [9]. Esta nueva concepción y tecnologías electrónicas basadas en ASICs permiten diseñar sistemas de bajo coste, de menor volumen y *eye-safe* (seguros ocularmente), basados enteramente en tecnologías semiconductoras. La Fig. 5 muestra una escena obtenida con un lidar elástico de este tipo, de forma que cada columna de pixels no es más que un registro de datos como el de la Fig. 4, orientado verticalmente y procesado con un mapa de color apropiado.

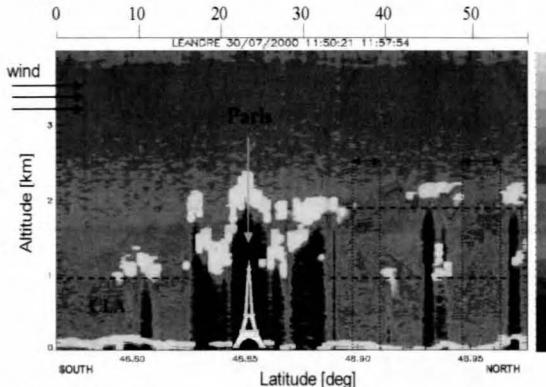


Fig. 5 Mapa de concentración de aerosoles (altura vs. latitud) en el área metropolitana de París sondeada mediante un microlidar backscatter embarcado en avión. La escala de colores representa la concentración de aerosoles (a.u.). Alrededor de la ciudad, la contaminación proporciona retornos máximos (en color blanco).

2.2 Medida de campos de viento

El principio operacional que utilizan estos sistemas lidar es la medida de la velocidad radial del viento a partir de los retornos que proporcionan las partículas atmosféricas en suspensión (polvo, polen, etc., típicamente de unas 2-3 μm de diámetro) y que actúan como «trazadores».

Puesto que las medidas que proporcionan los lidares de viento están resueltas en distancia (esto es, p.ej. cada cincuentena de metros se invierte el vector velocidad del viento), estos lidares encuentran aplicaciones no sólo en el campo meteorológico y medioambiental, en este último caso como predictores de los flujos de transporte de polución o como feedback de modelos software de mesoscala (una de las principales líneas de investigación de los centros de modelización medioambiental), sino también en los aeropuertos, como sistemas de alerta temprana de fenómenos de cizalladura de viento, cuyos efectos son potencialmente catastróficos en las maniobras de aproximación a tierra de los aviones.

La arquitectura de los lidares de viento comprende tanto sistemas coherentes como incoherentes para medir el desplazamiento Doppler,

$$f_d = -\frac{2v_r}{\lambda} \quad (4)$$

donde v_r es la velocidad de las partículas dispersoras, λ es la longitud de onda y f_d es el desplazamiento Doppler.

Como se sabe, la detección coherente implica la mezcla de la señal láser recibida de la atmósfera con la del propio láser emisor (oscilador local) a la misma frecuencia (detección homodina) u otra desplazada un cierto offset (mediante moduladores acustoópticos) respecto de la original, dando lugar entonces a una frecuencia intermedia

(detección heterodina), que facilita la información de sentido de la velocidad radial, v_r .

En principio, en 1964, Foreman y Huffaker [10] realizaron las primeras medidas Doppler de viento experimentales mediante un láser de CO₂ ($\lambda=10.59\text{ }\mu\text{m}$) pero no es hasta 1992 cuando aparece el primer sistema de Doppler lidar comercial capaz de medir velocidades y direcciones de viento entre 30 y 3000 m con precisiones de 1 m/s, resoluciones espaciales de 150 m y 5 minutos de integración [5]. Con todo, hoy por hoy, cabe decir que se trata de sistemas muy caros y con un know-how procedente de centros de investigación gubernamentales.

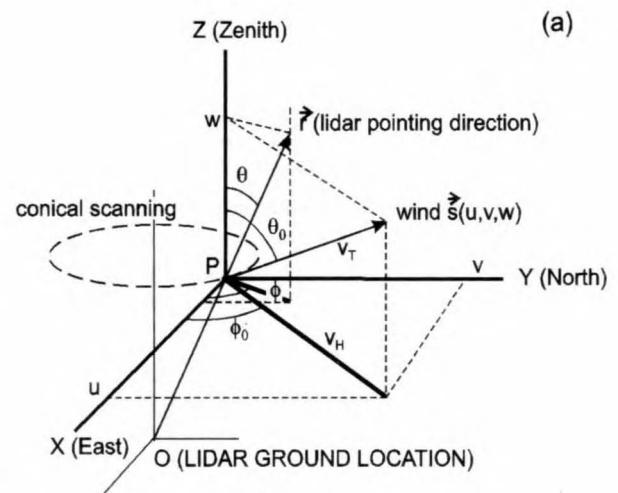


Figura 6. Ejemplo de operación de un lidar Doppler de medida de velocidad del viento. El sistema de localiza en O y realiza un scan cónico de la atmósfera de forma que en $t=t_0$ apunta según r . El vector velocidad del viento es s . Puesto que en cada instante de tiempo, la velocidad radial medida por el sistema no es más que la proyección de s sobre r , es posible determinar las tres componentes de velocidad del viento (adaptado de [6]).

Los sistemas coherentes suelen operar en la banda de 9-10 μm (láser de CO₂) y en menor grado en la banda «eye-safe» de 2 μm (láseres Tm:YAG) ya que las longitudes de onda más largas permiten mayores purezas espectrales y se reducen los efectos de la turbulencia atmosférica (básicamente, turbulencia fruto de las fluctuaciones del índice de refracción que se traducen en pérdidas de la coherencia de la señal y, consiguientemente, en menores eficiencias de mezcla en el sistema Doppler).

La Fig. 6 muestra una técnica de escaneado cónico de la atmósfera que permite recuperar las tres componentes de velocidad del viento y la Fig. 7 un mapa de la velocidad radial con indicación de sentido.

El procesado de la señal Doppler es complejo y se ve especialmente dificultado por el hecho que, a diferencia de lo que ocurre en radares de medida de viento (de microondas), que utilizan frecuencias órdenes de magnitud por debajo de las ópticas, no es posible determinar el

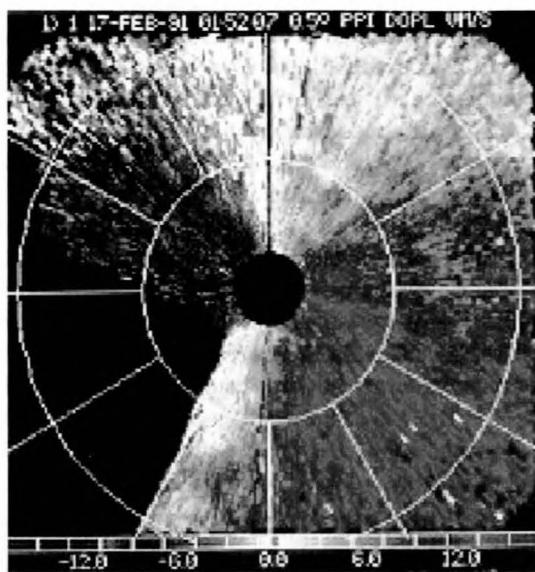


Figura 7. Ejemplo de un mapa PPI de velocidad del viento. El lidar se localiza en el origen. La escala de color representa velocidades radiales desde -12 m/s (entrantes) hasta +12 m/s (salientes) [11].

desplazamiento Doppler f_d a partir de la variación de fase ϕ entre pulsos recibidos ($f_d = \frac{1}{2\pi} \frac{d\phi}{dt}$) sino que es preciso

determinarla a partir de la información contenida de cada pulso individual, aislado. Esto es debido a que la coherencia temporal (el tiempo de correlación que tardan las partículas atmosféricas dispersoras en reorganizarse a partir de un estado de fase definida cuando fueron excitadas por la onda incidente) a frecuencias ópticas es comparativamente mucho menor. Como resultado, para invertir la velocidad es preciso utilizar técnicas avanzadas de estimación espectral como el periodograma o estimadores de covarianza, entre otros [6].

Existen también sistemas que permiten detectar el desplazamiento Doppler mediante técnicas de detección directa o interferométrica. Entre ellas cabe citar la *edge technique* (ET) [12] y la *fringe-technique* (FT) [13], respectivamente. La primera utiliza un interferómetro Fabry-Pérot a modo de «filtro paso banda» de forma que, cualitativamente, cuando el retorno Doppler se posiciona en el HHFW (half-height full width) de la curva de transmisión del filtro (banda lateral), las variaciones de frecuencia se traducen en variaciones de amplitud, construyendo de esta forma el transductor básico f_d -amplitud (algo similar al principio de los primeros detectores de FM haciendo un paralelismo con el mundo de la RF). El grupo lidar de TSC se encuentra actualmente trabajando en esta técnica de detección. La segunda técnica mencionada (FT) utiliza patrones de anillos interferentes como técnica básica de detección una reminiscencia de los conocidos anillos de Newton (...).

Finalmente, también es posible realizar medidas de campos de viento mediante sistemas lidar elásticos incoherentes como el descrito en la Sec.2.1, aunque con menores resoluciones, utilizando técnicas de correlación cruzada a partir de las inhomogeneidades detectadas en dos o más escenas adquiridas sucesivamente (a modo de «fotos» o «snapshots») [14].

2.3. Sistemas Raman, DIAL y otros

Existen por lo menos dos grandes categorías de sistemas lidar basados en técnicas inelásticas o absorción con importantes aplicaciones prácticas:

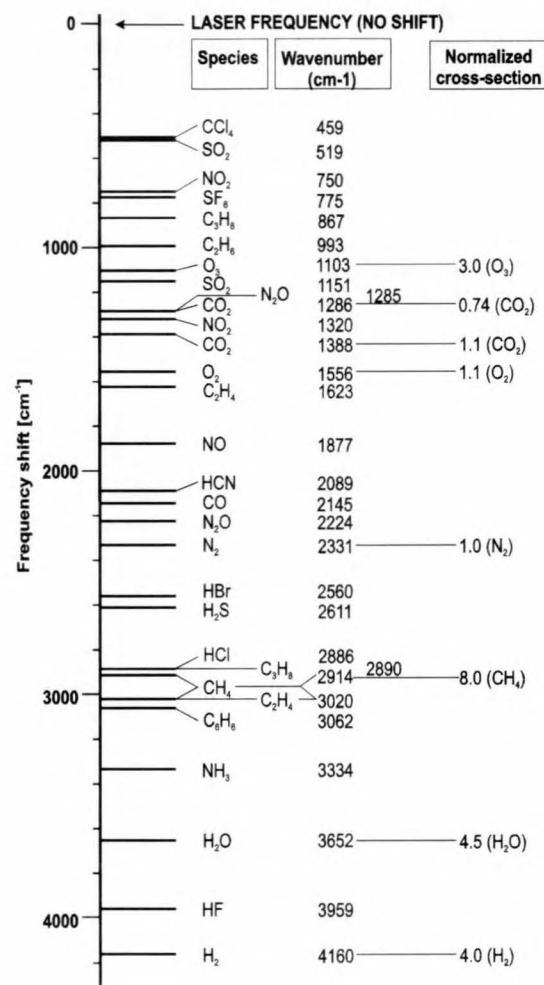


Figura 8. Desplazamientos frecuenciales por efecto Raman de diversas especies químicas típicamente presentes en una atmósfera contaminada y secciones rectas normalizadas a la de N₂ (adaptado de [16]).

a) *Sistemas lidar Raman*, que formalmente son sondas espectroscópicas de moléculas y átomos que permiten identificar el tipo de especie química detectada [15], perfiles de temperatura y de concentración de vapor de agua. La potencialidad de estos sistemas desde el punto de vista de monitorización y detección de fugas en plantas químicas y petroquímicas es pues enorme. A diferencia

de lo que ocurría con los sistemas elásticos en que la longitud de onda en recepción era la misma que en emisión, el campo dispersado por efecto de la interacción Raman con la atmósfera, se produce a una longitud de onda *desplazada* κ ($\kappa=1/\lambda$ [cm⁻¹], «wavenumber» en Fig. 8) con respecto de la emitida, de forma que este desplazamiento es propio de cada molécula o átomo, de ahí las enormes posibilidades de esta técnica para la identificación molecular de compuestos químicos.

b) *Sistemas DIAL (Differential Absorption Lidar)*. Como su nombre sugiere, el principio de medida es de naturaleza diferencial, de forma que utilizan dos (o más) longitudes de onda sintonizables, una absorbida por el gas o especie química a detectar y la otra (habitualmente próxima en longitud de onda) no [17]. La principal ventaja de estos lidares frente a los Raman yace en su gran sensibilidad (del orden de 10^4 - 10^5 veces mayor en el caso de la detección de CO) y en el hecho que el procedimiento de medida diferencial tiende a cancelar errores instrumentales. Los alcances típicos son de algunos kilómetros y, desde 1990 se existen equipos comerciales a bordo de furgonetas, que realizan medidas en terminales de gas y petrolíferos por diversos puntos de Europa.

3. SISTEMA LIDAR 3D DE LA UPC

Desde 1993, se han desarrollado dos grandes sistemas lidar en la UPC. El primero de ellos, un sistema lidar elástico (*o backscatter*) operando a 532 nm a modo de estación fija, ya ha sido presentado en la Fig. 1 y Fig. 3 y Sec.2.1.

El segundo, cuya construcción se inició en 1997 y que parte de una concepción modular a fin de hacerlo fácilmente transportable en diversas campañas de medidas por Europa, tiene como objetivo desarrollar un sistema lidar multilongitud de onda, dos de ellas elásticas (V. Sec.2.1) y una tercera (en recepción) inelástica Raman (V. Sec.2.3)[18]. Además, el sistema cuenta con capacidad para escanear en tres dimensiones (barridos en acimut y elevación con resoluciones de 30'' de arco) a partir de un editor vectorial de movimientos (en desarrollo). La iniciativa es fruto del V Programa Marco de Investigación y Desarrollo Tecnológico de UE (en el marco del proyecto europeo EARLINET) y de diversos proyectos de ámbito nacional y autonómico (CICYT, CIRIT, V. Sec.5).

Es evidente que cualquier proyecto como EARLINET [8], en el que participan más de una veintena de estaciones lidar operando coordinadamente en toda Europa (un mínimo de tres medidas regulares por semana), debe orientarse a la creación de valores añadidos en forma de desarrollo de curva de experiencia tecnológica y consolidación de equipos de trabajo pluridisciplinares en tecnologías punta, máxime en un marco globalizador e

integrador como el actual. En esta línea, las principales orientaciones del lidar-Raman 3D en desarrollo son hacia el campo meteorológico y de monitorización medioambiental (Sec.1), en particular, el establecimiento de una base de datos estadística a escala continental, que cuantificará las emisiones biogénicas y antropogénicas de aerosoles y su impacto ambiental.

La arquitectura del sistema (actualmente en su fase final de desarrollo) se ilustra en la Fig. 10 y ha reunido un amplio equipo investigador del grupo EEF (Enginyeria Electromagnètica i Fotònica, TSC) y la colaboración de más de una treintena de proyectos fin de carrera (PFC).

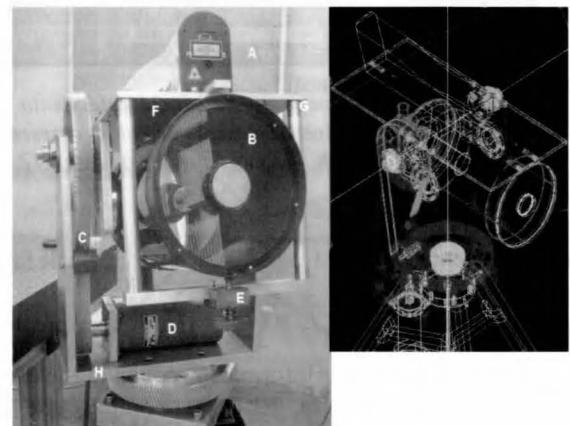


Figura 9. Vista frontal del actual sistema lidar-Raman 3D de la UPC.

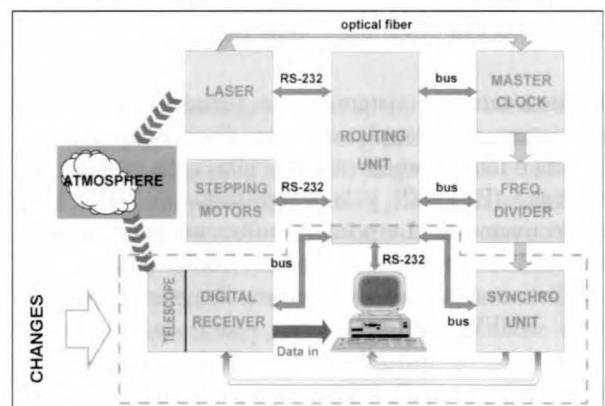


Figura 10. Arquitectura del sistema lidar-Raman 3D de la UPC. Basada en un sistema de control distribuido (multiprocesador), integra básicamente dos receptores digitales (canales elásticos de recepción), un contador fotónico (para el canal Raman), motores paso a paso y un complejo sistema de control, enruteamiento y monitorización de los diversos subsistemas.

4. BIBLIOGRAFÍA

- [1] Smullin, L. D.; Fiocco, G. Optical echoes from the moon. *Nature* **1962**, *194*, 1267-1267.
- [2] Brown, R. T. A New Lidar for Meteorological Application. *J. Appl. Meteorol.* **1973**, *12*, 698-708.
- [3] Collis R.T.H.; Russell, P.B. Lidar Measurement of Particles and Gases by Elastic Backscattering and Differential Absorption. In *Laser Monitoring of the Atmosphere*; Hinkley, E. D., Ed.; Springer-Verlag: New York, 1976; Chap. 4, 71-102.
- [4] Measures, R. M. Interaction and Propagation of Radiation. In *Laser Remote Sensing: Fundamentals and Applications*; Krieger Publishing Co.: Malabar, Florida, 1992; 138-145.
- [5] Roux, R. Cooperative ventures monitor atmospheric conditions. *Laser Focus World* **1994**, *30* (8), S7-S9.
- [6] Clifford, S. T.; Kaimal, J. C.; Lataitis, R. J.; Strauch, R. G. Ground-Based Remote Profiling in Atmospheric Studies: An Overview. *Proc. IEEE* **1994**, *82* (3), 313-355.
- [7] Collis, R.T.H. Lidar: a new atmospheric probe. *Q. J. R. Meteorol. Soc.* **1966**, *92*, 220-230.
- [8] Bösenberg, J.; Ansmann, A.; Baldasano, J. M.; Balis, D.; Böckmann, C.; Calpini, B.; Chaikovsky, A.; Flamant, P.; Hägård, A.; Mitev, V.; Papayannis, A.; Pelon, J.; Resendes, D.; Schneider, J.; Spinelli, N.; Trickl, T.; Vaughan, G.; Visconti, G.; Wiegner, M. EARLINET: A European aerosol research lidar network. In Proc. 20th International Laser Radar Conference (IPSL), Vichy (France), July 10-14, 2000; Pelon, J., Dabas, A., Loth, C., Eds.; IPSL (Institut Pierre Simon Laplace): 2000; Session 2 (Aerosol Properties and Radiation Budget (ARB)), paper 4.
- [9] Takeuchi, N.; Sugimoto, N.; Baba, H.; Sakurai, K. Random modulation CW lidar. *Applied Optics* **1983**, *22* (9), 1382-1386.
- [10] Huffaker, R. M. CO₂ laser Doppler systems for the measurements of atmospheric winds and turbulence. *Atmospheric Technology Journal* **1974-1975**, winter, (issue 6), 71-76.
- [11] Darby, L.S., Neff, W.D., Banta, R.M. «Multiscale analysis of a meso-beta frontal passage in the complex terrain of the Colorado Front Range.» *Mon. Wea. Rev.* **127**, 2062-2081 (1999).
- [12] McKay, J. A. Modelling of direct detection Doppler wind lidar: I. The edge technique. *Appl. Opt.* **1998**, *37* (27), 6480-6486.
- [13] McKay, J. A. Modelling of direct detection Doppler wind lidar: II. The fringe technique. *Applied Optics* **1998**, *37* (27), 6487-6493.
- [14] Eloranta, E. W.; King, J. M.; Weinman, J. A. The determination of wind speeds in the boundary layer by monostatic lidar. *J. Appl. Meteorol.* **1975**, *14*, 1485-1489.
- [15] Inaba, H.; Kobayashi, T. Laser-Raman Radar. *Opto-electronics* **1972**, *4*, 101-123.
- [16] Inaba, H. Detection of Atoms and Molecules by Raman Scattering and Resonance Fluorescence. In *Laser Monitoring of the Atmosphere*, Hinkley, E. D., Ed.; Springer-Verlag: New York, 1976; Chap. 5, 153-236.
- [17] Rothe, K. W.; Brinkman U.; et al. Remote Sensing of NO_x Emission from a Chemical Factory by the Differential Absorption Technique. *Appl. Phys.* **1974**, *4*, 181-182.
- [18] Rocadenbosch, F. et al. 3D Scanning Portable Backscatter Lidar Platform for Atmospheric Remote Sensing: Performance and Architecture overview. In *Remote Sensing of Clouds and the Atmosphere V*, EUROPTO 2000, Barcelona, Sept. 26-28, 2000; Russell, J.E., Schäfer, K., Lado-Bordowsky, O., Eds.; SPIE: Bellingham, Washington, 2001; Proc. 4168, pp. 158-169.

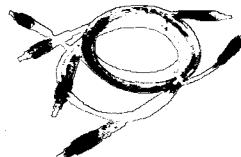
5. AGRADECIMIENTOS

Este trabajo ha sido posible gracias a la financiación de los siguientes entes: CICYT (Comisión Interministerial de Ciencia y Tecnología) bajo los proyectos TIC 431/93, AMB96-1144-C02-C01, TIC99-1050-C03-01, REN2000-1907-CE y REN2000-1754-C02-02/CLI, Ministerio de Educación y Cultura bajo la acción Hispano-Francesa HF1997-0212, CIRIT (Comisión Interdepartamental de Investigación e Innovación Tecnológica) y Dep. de Medio Ambiente de la Generalitat de Catalunya bajo el proyecto IMMPACTE y Comisión Europea bajo el proyecto EVR1-CT-1999-40003 EARLINET.

AUTORES



Francesc Rocadenbosch nació en Barcelona en 1967. Ingeniero Superior en Telecomunicación (1991) y Doctor Ingeniero (1996) por la Universidad Politécnica de Cataluña inició su carrera profesional en 1991, en la Universidad de Las Palmas de Gran Canaria, donde colaboró en el diseño de sistemas de alta frecuencia y microondas, y posteriormente, en 1993, la continuó como profesor del Departamento de Teoría de la Señal y Comunicaciones (UPC), de la que actualmente es profesor Titular de Universidad desde 1997. Premio al mejor Proyecto de Investigación en Cataluña (ACET, 1999) y máster en Administración y Dirección de Empresas (MBA) por la Universidad de Barcelona (2001), desde 1993 ha desarrollado una intensa actividad investigadora en el campo de los radares láser y la optoelectrónica a nivel internacional, donde ha participado en más de una decena de proyectos. Miembro de la Optical Society of America, es autor de numerosas publicaciones internacionales y revisor de diversas revistas de ingeniería óptica y procesado de la señal. Sus intereses actuales comprenden el diseño de sistemas de radar láser (lidar) de baja energía, receptores de altas prestaciones y técnicas de procesado en teledetección láser.



TÉCNICAS DE ETIQUETADO DE SEÑALES ÓPTICAS EN REDES DE COMMUTACION DE RÁFAGAS

J.J.Vegas Olmos, I.Tafur Monroy, A.M.J Koonen

COBRA Institute, Eindhoven University of Technology
P. O. Box 513, 5600 MB Eindhoven, The Netherlands

E-mail: j.j.vegas@tue.nl

RESUMEN

En este artículo presentamos una revisión de los asuntos más importantes relacionados con las redes de commutación de ráfagas de señales ópticas etiquetadas (LOBS – Labeled optical burst switched) y las tecnologías que permitirán en un futuro disponer de redes totalmente ópticas. La commutación de ráfagas de señales ópticas etiquetadas permite de un modo rápido y con un mecanismo eficiente de enrutamiento, el direccionamiento de paquetes/ráfagas sobre redes multiplexadas por división de onda (WDM – Wavelength division multiplexing). Por otro lado, y debido a que se unifica el algoritmo de enrutamiento y tiene una baja latencia, esta tecnología es fácilmente escalable hasta rangos de Terabits.

1. INTRODUCCIÓN

En las redes LOBS, las ráfagas de datos se crean a partir de juntar varios paquetes que han ingresado en los nodos de acceso y que tienen un mismo destino o bien una misma clase de servicio (CoS – Class of service). Para cada ráfaga datos, una etiqueta corta y de longitud constante, se crea y se asigna a ese bloque de datos. Esta etiqueta se usará en los routers internos de la red para dirigir correctamente el paquete a través de la red [1]. De este modo, la decisión de enrutamiento no va a venir dada por el reconocimiento y lectura de la información de los paquetes, sino por la lectura y procesado de estas pequeñas etiquetas ópticas, obteniendo por lo tanto una mejora en términos de latencia y de dimensionamiento de la cabecera de enrutamiento. Estos factores hacen que el enrutamiento de paquetes se simplifique y se pueda reescalar el sistema hasta tasas de bit del orden de Terabits [1]. En la Figura 1 podemos ver el aspecto global de la red.

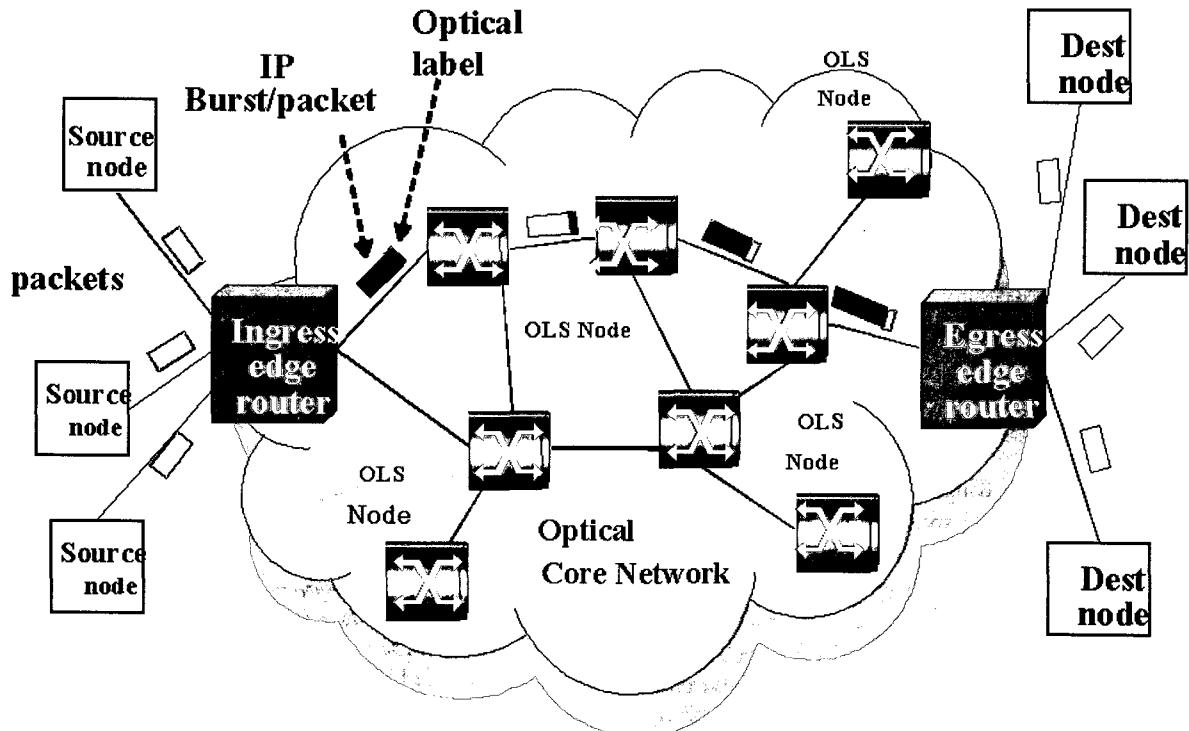
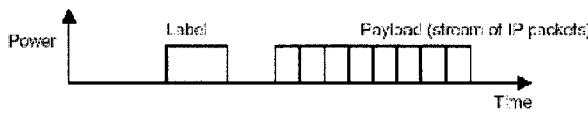


Figura 1. Arquitectura de una red de commutación de ráfagas de señales etiquetadas

El presente artículo pretende describir las técnicas que implementan etiquetado óptico de ráfagas IP. Varias técnicas han sido propuestas hasta la fecha para etiquetar paquetes ópticos o ráfagas de paquetes, pero nos centraremos en las cinco básicas para redes con múltiples longitudes de onda de las cuales se derivan el resto: multiplexación por división de tiempo (TDM), multiplexación por división de código óptico (OCDM), multiplexación por subportadora (SCM), modulaciones ortogonales y etiquetado por multiplexación por división de onda (WDM). En los primeros cuatro métodos, la etiqueta es asignada al paquete de información en el mismo canal espectral, mientras que en el quinto método, un canal diferente es el que se utiliza para transportar la etiqueta.

2. TÉCNICAS DE ETIQUETADO DE SEÑALES ÓPTICAS

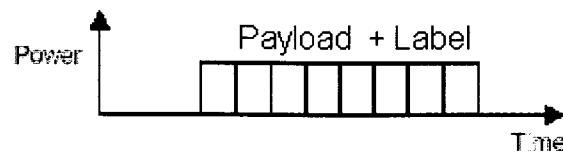
ETIQUETADO TDM : En el etiquetado TDM (Multiplexación por división de tiempo), también conocido como etiquetado en serie, la información de la etiqueta se junta en el dominio temporal, poniéndola delante del propio paquete de información. Este paquete de datos con su correspondiente etiqueta se codifica en la misma longitud de onda. Por lo tanto, datos y etiqueta quedan codificados en la misma longitud de onda. Eso hace que deban utilizarse bandas de guarda y bits de sincronización.



Las **ventajas** de esta técnica vienen dadas por el echo de trasmisitir toda la información por un mismo canal, de modo que se simplifica el posterior algoritmo de enrutamiento en los nodos. Como **desventaja** debemos argumentar el estricto nivel de sincronización que debemos asegurar tanto en el proceso de inserción de etiquetas como en el proceso de borrado y reescritura de etiquetas. Hay que notar también que la tasa de bit usada para la etiqueta puede ser la misma que para la información (TDM sincrónico), o menor (TDM asincrónico). Este segundo caso es el que mayor ventajas ofrece puesto que nos permite utilizar equipo electrónico de baja velocidad para el procesado de las etiquetas.

ETIQUETADO OCDM: El etiquetado OCDM (multiplexado por división de código óptico) ha sido también propuesto como técnica de etiquetado en redes ópticas por varios autores [3][4]. La etiqueta se inserta mezclando la información con un código específico que contiene la información de la etiqueta. Pese a que el OCDM es una de las técnicas de etiquetado que permite el reconocimiento de etiquetas para el enrutado, sin necesidad de mirar tablas de operaciones, su implementación es complicada. Si por ejemplo un canal

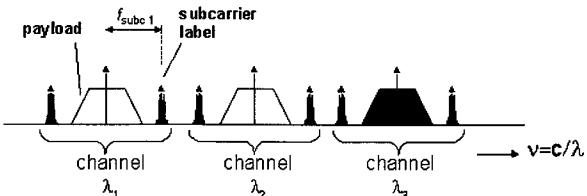
soporta N OCDM códigos, se requiere un banco de N autocorreladores ópticos por canal. De todos modos, el etiquetado OCDM ofrece la posibilidad de ser combinado con WDM (WDM con sub-bandas).



Por tanto, como **ventajas**, podemos nombrar el echo de que tanto la información como la etiqueta estan en el mismo canal, y por lo tanto se facilita la reserva de rutas en los nodos.

Como **desventajas**, nombraremos el echo de que el tamaño de la información final que se transmite es bastante superior a la que se debería transmitir si enviasemos por separado información y etiqueta (debido al efecto del código de secuencia).

ETIQUETADO SCM : Con el etiquetado SCM (Multiplexado con subportadoras), la información de la etiqueta es modulada en una subportadora, que se coloca en el mismo canal óptico, separada del espectro en banda base de la información. En un sistema con varios canales, la frecuencia de la subportadora se puede configurar de modo que sea la misma para todos, de modo que el proceso de detección se simplifica.

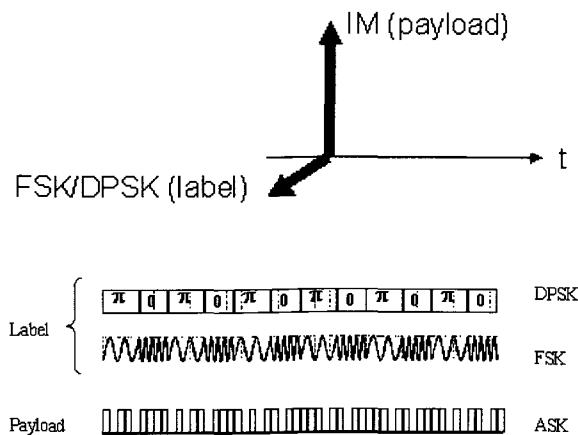


Las **ventajas** del etiquetado SCM van relacionadas con el echo de mandar por un mismo canal la información y la etiqueta. Debido a esto, no se requieren grandes niveles de sincronización entre dicha etiqueta y la información enviada. Otro punto interesante, es que para una detección óptica directa mediante fotodiodos, las diferentes subportadoras pueden ser detectadas sin necesidad de demultiplexar ópticamente la señal.

Algunas **desventajas** vienen dadas por la posibilidad de sufrir desapariciones de la subportadora por culpa de la dispersión de la fibra. Por otro lado, las no-linealidades pueden causar distorsiones por intermodulación, causando interferencias en otros canales. Ademas, para tasas de bits de la información altas, las subportadoras necesitan estar colocadas a alta frecuencia, lo que requiere equipos electrónicos sofisticados, y lo que puede hacer que rebasemos en espacio mínimo entre canales.



ETIQUETADO ORTOGONAL: La información de la etiqueta se inserta en la fase o la frecuencia de la portadora de la información, la cual está modulada en amplitud. De este modo aprovechamos la ortogonalidad de la fase, la amplitud y la frecuencia de las señales, transmitiendo por separado información en cada una de estas variables.

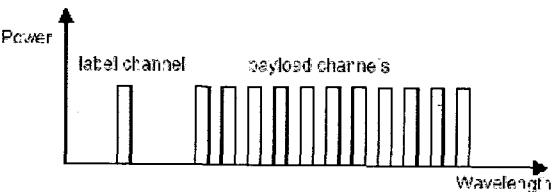


Las **ventajas** que comporta esta técnica están relacionadas principalmente con el hecho de que la etiqueta y la información viajan juntas, en la misma longitud de onda (y temporalmente en el mismo instante), con lo que simplificamos los mecanismos de lectura y reserva de los routers. Por otro lado, tanto la etiqueta como la información pueden separarse sin grandes restricciones a nivel de sincronización, la cual solo se requiere a nivel de paquete (y no de bit). A nivel de espectro, este no se incrementa al insertar la información. Por último, no se necesita ningún mecanismo de delimitación en el momento de borrar y reescribir la nueva etiqueta en los nodos internos.

Como **desventajas** debemos resaltar el hecho de que este tipo de formato es susceptible de sufrir crosstalk. Es decir, interferencias entre la información y la etiqueta introducidas por efecto del chirp del conversor de longitud de onda utilizado en el intercambio de etiquetas y/o interferencias de la etiqueta a la información debido a conversión de FM a IM (debido por ejemplo a efectos de dispersión o interfereometría en los links de fibra).

ETIQUETADO WDM: En el etiquetado WDM (Multiplexación por división de longitud de onda), las etiquetas de cada canal pueden multiplexarse y mandarse vía un canal separado dedicado específicamente para este propósito. Esto significa por lo tanto que debe existir un sistema de sincronización bastante estricto, pues de otro modo no va a ser posible identificar qué etiqueta corresponde a cada canal. La dispersión cromática tiene aquí un papel importante, pues este efecto introduce diferencias en las velocidades de grupo entre los canales que contienen información, y el que contiene las etiquetas.

Una **ventaja** de esta aproximación es que solo el canal común de las etiquetas debe ser inspeccionado para el correcto enrutado.



Como **desventaja**, podemos volver a remarcar el tema de la sincronización, ya sea en el aspecto de detectar correctamente qué etiqueta corresponde con cada paquete de información como a nivel de realizar el borrado y la inserción de las etiquetas en los nodos intermedios.

En la Tabla 1 se presenta un estudio comparativo de las técnicas de etiquetado de señales ópticas (página siguiente).

3. CONCLUSIONES

La próxima generación de redes ópticas con etiquetado, conocidas como LOBS, prometen subministrar grandes anchos de banda aprovechando el dominio óptico en el que trabajan, y al hecho de utilizar técnicas de etiquetado, que permiten reducir considerablemente el tiempo de procesado en los nodos intermedios a la vez que la conversión O/E/O (y por tanto, depender de equipos electrónicos que por definición, admiten tasas de bit menores). Varias técnicas de etiquetado óptico han sido revisadas en este artículo, dando especial énfasis a sus ventajas y puntos débiles de cara a aplicaciones en redes LOBS.

AGRADECIMIENTOS

Este trabajo se ha realizado bajo el marco del proyecto europeo IST-STOLAS, que está parcialmente financiado por el Programa IST de la Comunidad Europea. Los autores quieren mostrar su reconocimiento por las contribuciones de sus colegas en el consorcio STOLAS.

REFERENCIAS

- [1] Chunming Qiao, «Labeled Optical Burst Switching for IP over WDM Integration», IEEE Communication Magazine, September 2000, pp.104-114.
- [2] A. Banerjee, et al., «Generalized Multiprotocol Label Switching: An Overview of Routing and Management Enhancements», IEEE Communication Magazine, Jan. 2001, pp.144-150.
- [3] Y. G. en, Y. Zhang, L. K. Chen, On Architecture and Limitations of Optical Multiprotocol Label Switching (MPLS) Networks Using Optical-Orthogonal-Code(OCC)/Wavelength Label, OTF, Vol. 8, pp. 43-70, 2002
- [4] Hideyuki Sotobayashi ; Wataru Chujo ; Ken-ichi Kitayama, Photonic Gateway: Multiplexing Format Conversions of OCDM-to-WDM and WDM-to-OCDM at 40 Gbit/s (4 x10 Gbit/s) , Journal of Lightwave Technology, accepted for future publication , 2002, pp. 1

AUTORES



J.J. Vegas Olmos recibio el titulo de Ingeniero Tecnico en Telecomunicaciones y el titulo de Ingeniero Electronico por la UPC en el 2001 y el 2003 respectivamente. Actualmente realiza sus estudios de Doctorado en el Grupo de Comunicaciones Opticas de la Eindhoven University of Technology.



I. Tafur Monroy recibio el titulo de Ingeniero en Telecomunicaciones Multicanal por el Bonch-Bruevitch Institute of Communications, St.Petersburgo, Rusia, en 1992. Posee el titulo de Licenciado en Tecnologia por el Royal Institute of Technology, Estocolmo, Suecia, y en 1999 recibio el Doctorado en Ingenieria Electrica por la Eindhoven University of Technology, Paises Bajos.



A.M.J. Koonen sirvio desde 1979 al 2000 como miembro del cuerpo tecnico y manager de Bell Laboratories, Lucent Technologies, en los Paises Bajos. Desde 1991 al 2000, fue profesor asociado en Redes Fotonicas en la University of Twente, Paises Bajos, y desde el 2001, profesor titular en Redes de Comunicaciones de Banda Ancha, en el Grupo de Comunicaciones Opticas de la Eindhoven University of Technology, Paises Bajos. Fue el primer empleado de Lucent Technologies en Europa en conseguir el Bells Labs Fellow.

	Etiquetado Ortogonal	Etiquetado SCM	Etiquetado sincrono TDM	Etiquetado Asincrono TDM	Etiquetado OCDM	Etiquetado WDM
Sincronización de los datos con la etiqueta	No estricto. A nivel de paquete	No estricto. A nivel de paquete	Estricto. A nivel de bit	No estricto. A nivel de paquete	Estricto. A nivel de bit	No estricto. A nivel de paquete
Ancho de banda del canal *	Tasa de bit de la Información + Espaciado de la FSK	Frecuencia de la subportadora mas alta	Tasa de bit de la Información+ tasa de bit de la etiqueta	Tasa de bit de la información	Multiplo de la tasa de bit de la información	Tasa de bit de la información
Tasa de bit de los datos *	Tasa de linea	Tasa de linea	Tasa de linea - tasa de etiqueta	Se reduce cuando la etiqueta acorta	Fraccion de la tasa de linea	Tasa de linea
Lectura de etiqueta	Demultiplexado de todos los λ -canales	Detección OE para todos los λ -canales, no λ -demultiplexado	Demultiplexado de todos los λ -canales + electronica alta velocidad	Demultiplexado de todos los λ -canales + comunicacion optica+ electronica baja velocidad	Demultiplexado de todos los λ -canales	Demultiplexado de las etiquetas en el λ -canal comun
Borrado de etiqueta	λ - demultiplexado + λ -conversion (XGM o XPM)	λ - demultiplexado + XGM λ -conversion, o filtro optico notch	λ - demultiplexado + conversion OE de la informacion y de la etiqueta	Separacion de la etiqueta de la data mediante comunicacion optica lenta	λ - demultiplexado + conversion OE de la informacion y la etiqueta + decodificacion	Conversion OE + demultiplexado en tiempo de las etiquetas en el λ -canal comun
Insercion de etiquetas	Por modulacion FSK del laser sintonizable o modulacion DPSK externa	Mediante modulador externo	Conversion E/O del paquete + nueva etiqueta	Multiplexado de la informacion y la etiqueta mediante commutador optico	Conversion OE y decodificado del paquete, +codificado del nuevo paquete	Conversion OE + multiplexado temporal de las nuevas etiquetas en el λ -canal comun
Transmision	Conversion IM-a-FM	Desvanecimiento de las subportadoras	Alineacion de la informacion y las etiquetas	Grandes bandas de guarda entre la informacion y la etiqueta	Alta tasa de linea	Dispersión cromatica

Tabla 1. Comparativa de las técnicas de etiquetado de señales ópticas. *Omitiendo bandas de guarda, y asumiendo que la tasa de bit de la etiqueta es mucho mas pequeña que la tasa de los datos.



METAMATERIALES EN MICROONDAS Y ANTENAS

J. M. González-Arbesú, E. Úbeda y J. Romeu

Universitat Politècnica de Catalunya, Campus
Nord, Mòdul D3, tlf: (34) 934016849

jmgonzalez@tsc.upc.es, uheda@tsc.upc.es,
romeu@tsc.upc.es

Abstract - Desde la década de los '80 una gran variedad de materiales artificiales (no existentes en la naturaleza) son intensivamente estudiados por sus inusitadas y sorprendentes propiedades electromagnéticas (como partes reales de permitividad eléctrica y permeabilidad magnética efectivas negativas, refracciones anómalas, no causalidad e inversión del desplazamiento Doppler, entre otras). Las investigaciones realizadas en algunos de ellos han conducido ya a nuevos diseños de antenas. En otros, con propiedades aún más sorprendentes y controvertidas, las investigaciones se encuentran todavía en un estadio incipiente y tan sólo auguran prometedores desarrollos que revolucionarán el mundo de las microondas y las antenas.

En este artículo se describirá escuetamente la diversidad de materiales artificiales objeto de estudio en ingeniería de microondas y en antenas, algunas de sus propiedades (tanto las verificadas experimentalmente como las teóricas) y sus potenciales aplicaciones para la fabricación de dispositivos. También se describirán las actividades que se están desarrollando en este campo en el Departament de Teoria del Senyal i Comunicacions de la Universitat Politècnica de Catalunya.

1. INTRODUCCIÓN

Desde los tiempos de G. Marconi metales y (aislantes) dieléctricos son los materiales que se vienen empleando en la fabricación de antenas, dispositivos de microondas y circuitos electrónicos. Dada su elevada conductividad eléctrica, la principal función que desarrollan los metales en este tipo de aplicaciones era la conducción y/o almacenamiento de cargas, mientras que los dieléctricos se empleaban inicialmente como elementos estructurales de soporte dada su baja conductividad. La diferencia en las conductividades de metales y dieléctricos se puede explicar sobre la base de la existencia de cargas que pueden moverse libremente en los primeros, mientras que en los segundos están *ligadas* admitiendo tan solo pequeños desplazamientos.

Para el ingeniero-diseñador de dispositivos de microondas y antenas, la conductividad eléctrica s es el parámetro físico de mayor interés al trabajar con metales. La permitividad eléctrica relativa ϵ_r , la permeabilidad magnética relativa m_r , las tangentes de pérdidas eléctricas $tg d_e$ y magnéticas $tg d_m$

y la conductividad son los parámetros requeridos para la realización de un diseño en el que se empleen dieléctricos. Evidentemente, otros parámetros son también de importancia en el momento de realizar un diseño, como por ejemplo el espesor de los metales.

Tanto permitividad como permeabilidad relativas lo son respecto a los valores del vacío, y son representadas por números complejos. Los dieléctricos podrían clasificarse atendiendo al signo de las partes reales de su permitividad y permeabilidad efectivas, tal y como se representa en el sistema de coordenadas de la figura 1. Los materiales dieléctricos convencionales, existentes en la naturaleza, empleados en las tecnologías de las comunicaciones se sitúan en el primer cuadrante. Estos materiales se caracterizan por poseer valores positivos de la parte real de permitividad eléctrica y permeabilidad magnética, y se les conoce como materiales DPS (*Double Positive*) o RHM (*Right-Handed Media*), para los que la energía (definida a través del vector de Poynting, S) se propaga en la misma dirección que las ondas electromagnéticas (definida por el vector número de onda k). Estos materiales se caracterizan por tener un índice de refracción $n(\omega)$ positivo (aunque pueda ser complejo, su parte real es positiva).

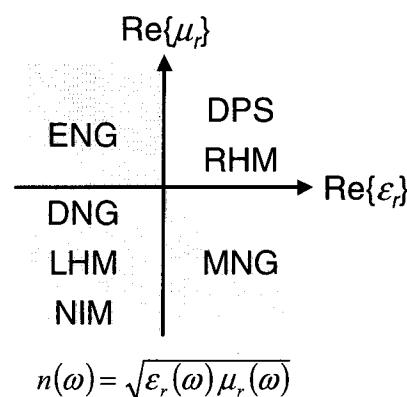


Figura 1. Mapa de los materiales dieléctricos clasificados según el signo de su permitividad y permeabilidad. El significado de las siglas se encuentra en el glosario, al final del artículo.

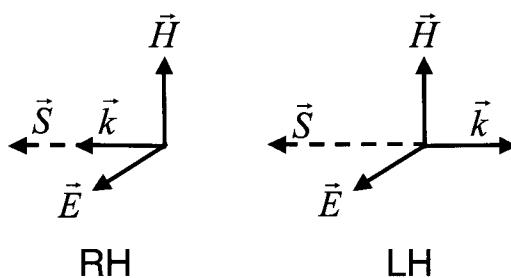
$$n(\omega) = \sqrt{\epsilon_r(\omega) \mu_r(\omega)} \quad (1)$$

No obstante, las propiedades de materiales complejos en los que la parte real de la permitividad y/o permeabilidad

pueden adquirir valores negativos se llevan estudiando desde los años '50 del pasado siglo y son objeto de un gran interés en la actualidad por muchos equipos de investigación, no tan solo por las aparentes peculiaridades de su comportamiento electromagnético sino también por las aplicaciones *tangibles* a las que puede dar lugar.

Así, ya han sido diseñados y fabricados materiales ENG (*Epsilon Negative*) y MNG (*Mu Negative*), pertenecientes al segundo y cuarto cuadrante, respectivamente, en los que tan sólo las partes reales de permitividad o permeabilidad pueden ser negativas (sólo una de ellas y no las dos simultáneamente). Las primeras experiencias con éxito en las que se ha conseguido fabricar un material MNG en una estrecha banda de frecuencia son muy recientes (año 1999).

Los materiales artificiales que en la actualidad son objeto de un estudio más *esforzado*, y provocadores de una gran controversia, son los pertenecientes al tercer cuadrante. Se caracterizan por poseer partes reales de permitividad y permeabilidad simultáneamente negativas, al menos en un estrecho margen de frecuencias, dando lugar a **índices de refracción negativos**. Por ello se les llama materiales DNG (*Double Negative*) o NIM (*Negative Index Media*). Adicionalmente estos materiales también son referidos como LHM (*Left-Handed Media*), pues en ellos **las ondas electromagnéticas se propagan en dirección contraria a la propagación de la energía**, tal como se representa en la figura 2.



$$\vec{k} \times \vec{E} = \omega \mu \vec{H}$$

$$\vec{k} \times \vec{H} = -\omega \epsilon \vec{E}$$

$$\vec{S} = \vec{E} \times \vec{H}$$

Figura 2. Medios RH y LH.

2. METAMATERIALES

Como hemos podido ver, los materiales comentados reciben diversos nombres en función de sus características, si bien algunos de ellos podrían englobarse bajo el epígrafe Metamateriales (MM). En definitiva, Metamateriales serían todos aquellos materiales artificiales que poseen propiedades novedosas y superiores a los materiales que pueden encontrarse en la naturaleza. Todos tienen en común la existencia de ciertas estructuras

periódicas o *celdas básicas* como constituyentes elementales. Atendiendo a esta periodicidad es usual realizar una clasificación de los materiales artificiales empleados en electromagnetismo en términos de sus dimensiones respecto a la longitud de onda. Así, cuando las celdas básicas son mucho menores que la longitud de onda podemos encontrarnos con materiales ENG, MNG o DNG, mientras que si las celdas básicas son del orden de la longitud de onda se habla de estructuras EBG (*Electromagnetic Band-Gap*). En este último caso, cuando el diseño corresponde a frecuencias ópticas se habla de estructuras PBG (*Photonic Band-Gap*).

A. Electromagnetic Band-Gap

Aunque en este artículo no trataremos en detalle sobre materiales EBG (ni PBG) cabe destacar la gran cantidad de aplicaciones sugeridas y llevadas a la práctica con estos materiales. Todas ellas se aprovechan del control que se tiene sobre las frecuencias y números de onda, siendo posible seleccionar las bandas de propagación y no propagación en el sustrato EBG (o PBG) a partir del control de las periodicidades (uni-, bi- o tridimensionales) de las estructuras en espacios que forman el material.

De esta forma, en el diseño de antenas planas, a partir de la realización de perforaciones (de dimensiones y disposiciones específicas) sobre un sustrato dieléctrico convencional, es posible la práctica supresión de las ondas de superficie que se desplazan por el sustrato, tal como se representa en la figura 3, y afectan a los diagramas de radiación y a la eficiencia de radiación. Este mismo diseño reduce los acoplamientos en el caso de agrupaciones formadas por antenas planas. No obstante, no es esta la única posibilidad para la realización de un sustrato EBG para antenas planas. Existen diseños en los que el sustrato está fabricado con postes (cilíndricos) de material dieléctrico de dimensiones y separaciones específicas, e incluso con geometrías «woodpile» (figura 4).

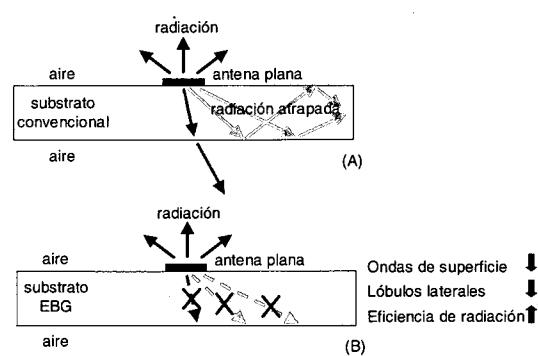


Figura 3. Sección de una antena plana convencional (A) y otra con substrato EBG 3-D (B). Se muestra la propagación de ondas en el interior del sustrato convencional y la práctica eliminación de la propagación en el interior del substrato EBG. Si las antenas fueran tipo parche el substrato EBG podría ser 2-D.

También están bien documentados los desarrollos de filtros, guías de onda, antenas de cavidad, antenas parabólicas y estructuras sintonizables con materiales EBG.

B. Double Negative Media

Los materiales DNG se caracterizan por los valores negativos de las partes reales de permitividad y permeabilidad, simultáneamente en la misma banda de frecuencias, originando valores también negativos para el índice de refracción.

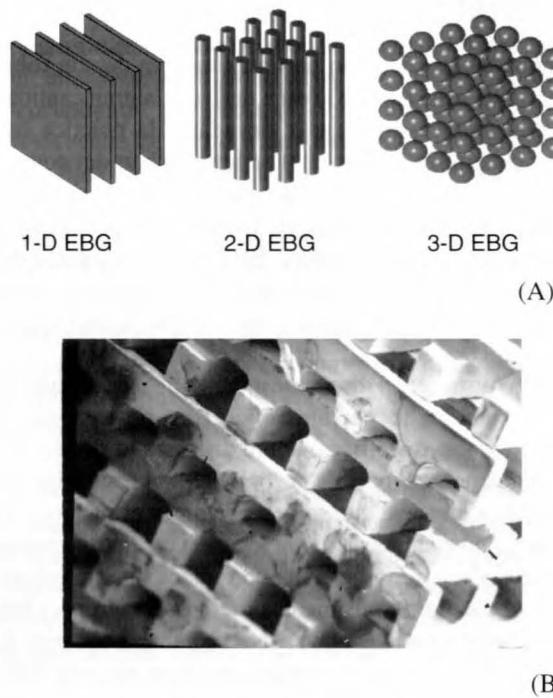


Figura 4. (A) Configuraciones típicas de substratos EBG que reflejan la periodicidad en 1-D, 2-D y 3-D. (fuente: [11, T. Itoh]). (B) Substrato EBG en configuración «woodpile» (fuente: [3]).

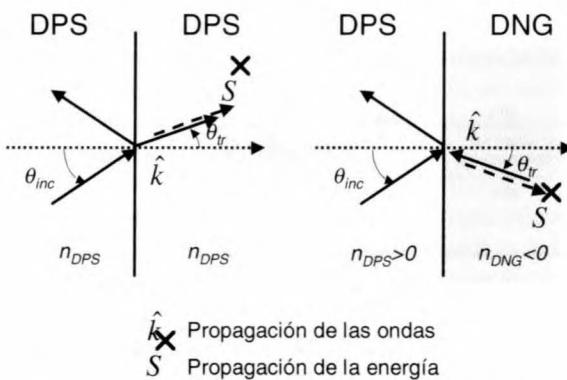
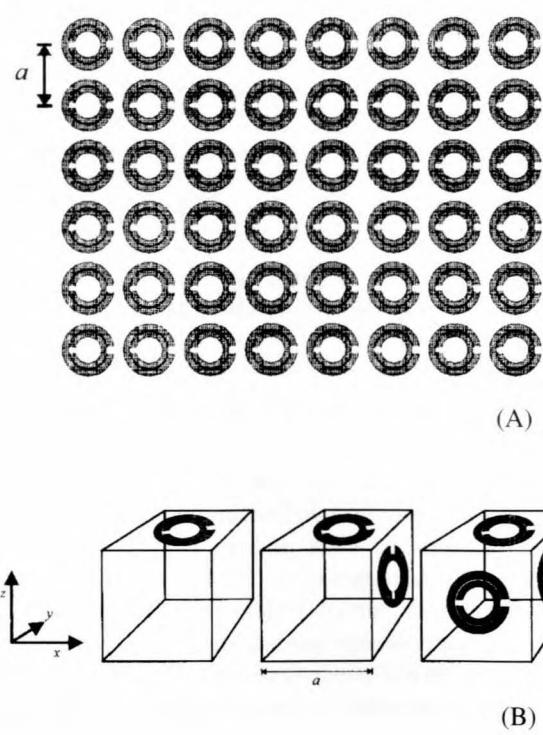


Figura 5. Inversión del ángulo de refracción en un medio DNG respecto a un medio DPS para una onda electromagnética continua.

Las propiedades de estos materiales artificiales fueron estudiadas desde un punto de vista teórico por el físico

ruso Víctor Veselago, ya en 1967. Aunque algunos de sus resultados contienen errores, las principales propiedades de un material DNG son peculiares. Al poner en contacto un material DPS con un material DNG, las ondas continuas que se estuvieran propagando desde el material DPS presentan un ángulo de refracción *anómalo* (de signo opuesto al que tendrían en otros materiales DPS convencionales). Esta anomalía sólo se observa para ondas continuas, no para ondas moduladas (las que transportan información), como puntualiza P. Valanju, de la University of Texas. También es peculiar la inversión del efecto Doppler en el interior de un material DNG, donde las ondas transmitidas desde una fuente sufren una reducción en su frecuencia cuando son observadas por un receptor al que la fuente se aproxima. Lamentablemente estas (y otras) propiedades no pudieron ser demostradas en 1967 debido a la inexistencia de materiales DNG.

No fue hasta el año 1999 cuando Pendry y sus colaboradores, del Imperial College in London, fueron capaces de diseñar un material con una elevada permeabilidad magnética efectiva negativa en un reducido margen de frecuencias a partir de la agrupación periódica de conductores magnéticos (figura 6). A frecuencias inferiores a la de resonancia el comportamiento del material posee una permeabilidad efectiva positiva, y sobre la resonancia una permeabilidad efectiva negativa. El interés de este tipo de estructura radica en la posibilidad de diseñar materiales DNG o NIM a partir de su combinación con materiales con permitividades efectivas negativas (en el mismo margen frecuencial). Estos materiales ya habían sido fabricados con tiras metálicas, presentando un comportamiento equivalente a una permitividad efectiva negativa en una pequeña región del espectro. En la figura 7 se muestran pantallas con estas características.



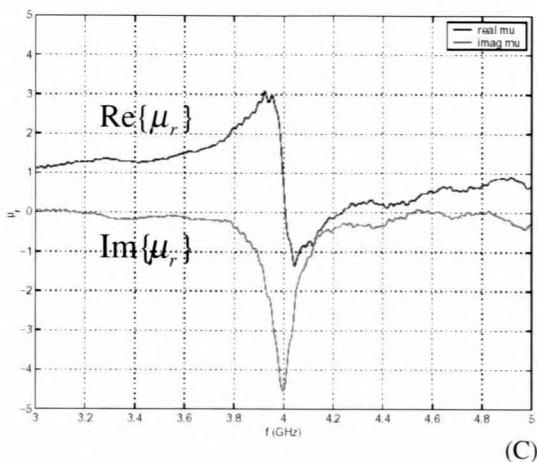
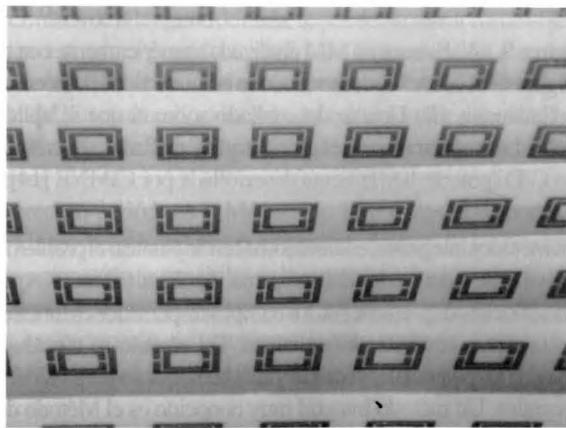


Figura 6. (A) Pantalla MNG formada con SRR (Split Ring Resonators); posee una permeabilidad efectiva negativa sobre la frecuencia de resonancia, aunque presenta un comportamiento anisotrópico (varía con la dirección de llegada de las ondas electromagnéticas) (fuente: [2]). (B) Añadiendo simetrías 3-D a las celdas básicas se consigue incrementar la isotropía del conjunto (fuente: [2]). (C) Partes real e imaginaria de la permeabilidad magnética efectiva de una pantalla de SRR como la de la figura 7, diseñada en la Universidad Politécnica de Cataluña.



(A)



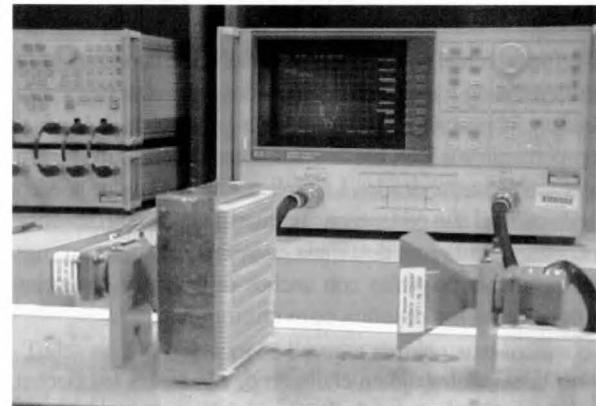
(B)

Figura 7. Ejemplo de pantallas ENG (A) y MNG (B), diseñadas en la Universitat Politècnica de Catalunya (fotografías: J. Giner).

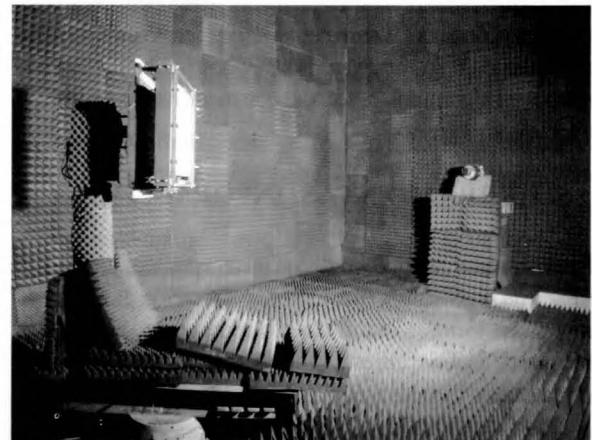
La aparición de material MNG rápidamente (en el año 2000) condujo a la fabricación de la primera estructura DNG por el grupo de D. Smith en la Universidad de California en San Diego.

3. CARACTERIZACIÓN DE METAMATERIALES DNG

La caracterización de materiales DNG requiere la utilización de medidas electromagnéticas en reflexión y en transmisión. El material en estudio se sitúa (dependiendo de su tamaño físico y de la frecuencia de trabajo) en una guía de ondas o en el interior de una cámara anechoica (figura 8). Empleando un analizador de redes vectorial se determinan los parámetros S_{11} y S_{21} en intensidad y fase -es decir, los campos reflejados y transmitidos por el material con relación a los campos incidentes- de la muestra de material en un ancho de banda que incluya la región donde se encuentra la resonancia del material. Con estos datos, y conocido el espesor de la muestra, es posible resolver el sistema de ecuaciones que relaciona los parámetros S de la muestra de material con su permitividad $\epsilon_{eff}(w)$ y permeabilidad efectiva $\mu_{eff}(w)$ en la banda frecuencial de trabajo.



(A)



(B)

Figura 8. (A) Caracterización de pequeñas pantallas MM en laboratorio (fuente: [11, R. W. Ziolkowski]). (B) Caracterización de pantallas DNG en la cámara anechoica de la Universitat Politècnica de Catalunya (fotografía: J. Giner).

Empleando este método se han realizado medidas experimentales que certifican la existencia de regiones en las que ciertos materiales se comportan de forma equivalente a pantallas con permitividad efectiva negativa, y otros materiales con comportamientos equivalentes a pantallas con permeabilidades efectivas negativas. Asimismo, y mediante la combinación de ambas, se ha comprobado (aunque con persistentes reticencias por parte de algunos investigadores) la existencia de estructuras con comportamientos equivalentes a materiales DNG en una banda de frecuencias reducida y con una incidencia electromagnética normal a las pantallas.

4. EXPECTATIVAS

La utilización de estructuras EBG en antenas y microondas es una realidad pues una pléyade de dispositivos han sido ya fabricados, revelando excepcionales características en los diferentes aspectos técnicos en los que han sido evaluados. Las reducciones de tamaño y peso con respecto a diseños más convencionales son notables, aunque por contra suelen requerir procesos de fabricación más complejos. Así, se han fabricado prototipos de, entre otros dispositivos, antenas de tipo parche, acopladores, filtros pasa banda con bandas pasantes espurias de menor intensidad que con diseños convencionales, guías rectangulares TEM, e incluso conductores magnéticos artificiales.

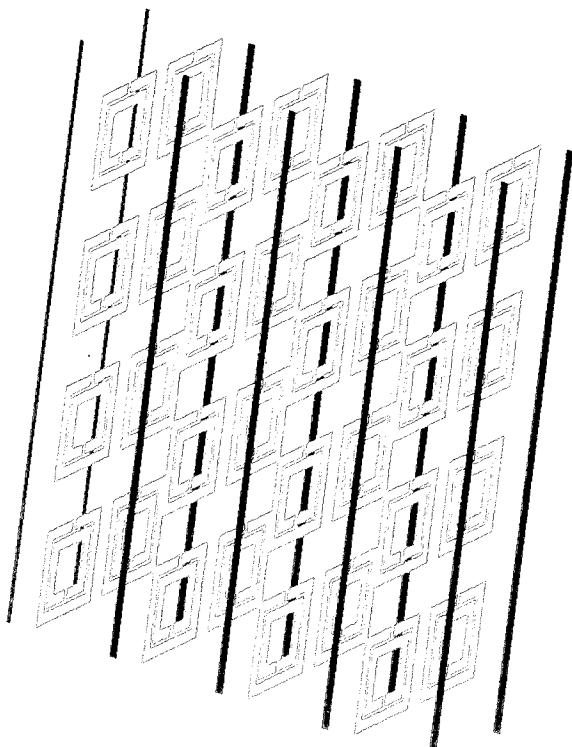
El esfuerzo actual en el campo de los DNG se concentra en la verificación de los efectos demostrados teóricamente y en el diseño y fabricación de metamateriales con anchos de banda y comportamientos DNG en ángulos de incidencias superiores a los actualmente fabricados. Las aplicaciones en las que pueden ser de utilidad aún pertenecen al campo de la ficción científica encontrándose la más destacada en la fabricación de antenas eléctricamente pequeñas con anchos de banda y ganancias superiores a los diseños convencionales.

Otras líneas de trabajo en el diseño de materiales se orientan hacia la creación de materiales DPS artificiales con permitividades y permeabilidades *a medida*. Para ello se combinan adecuadamente materiales con permitividades y permeabilidades existentes y conocidas. Los materiales a emplear y su disposición es determinada, típicamente, mediante algoritmos de optimización (basados en algoritmos genéticos) que requieren de un intenso esfuerzo computacional.

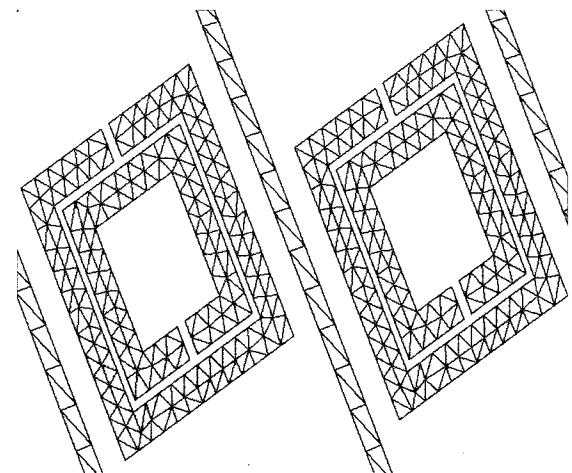
5. HERRAMIENTAS INFORMÁTICAS

El análisis electromagnético de los metamateriales pasa por el mallado de la geometría (figura 9) y la aplicación de alguna de las variadas técnicas numéricas existentes. Las dos familias generales de análisis para este tipo de problemas son los métodos diferenciales o bien los métodos integrales.

Los métodos diferenciales nacen de aproximar las derivadas presentes en las ecuaciones de Maxwell por diferencias finitas en el espacio y en el tiempo. Sin duda, el método diferencial más usado es FDTD (*Finite Differences in Time Domain*), que es el método adecuado para analizar el efecto del objeto bajo la incidencia de pulsos modulados a una cierta frecuencia.



(A)



(B)

Figura 9. (A) Estructura MM analizada numéricamente con el programa FIESTA 3-D desarrollado en la Universidad Politécnica de Catalunya. (B) Detalle del mallado sobre el que se aplica MoM. Las representaciones gráficas y los mallados se realizan con GiD (potente herramienta desarrollada por CIMNE [14]).

Los métodos integrales, en cambio, nacen de plantear el problema equivalente -derivado del teorema de Equivalencia- que desemboca en expresiones de los campos dispersados en función de corrientes equivalentes radiantes. Tal vínculo se efectúa a través de los potenciales vectores y escalares, que son expresiones integrales. Un método integral muy conocido es el Método de Momentos (MoM), el cual se basa en aplicar las condiciones de contorno en la superficie de los objetos mediante el uso de unas funciones base. Normalmente, tales funciones base pueden ser funciones tejido (de sumo interés para la discretización de

estructuras planas y regulares) o bien funciones triángulos (adecuadas para el mallado de una superficie arbitraria) [17].

La elección del uso de FDTD o de MoM depende del problema a analizar. FDTD es adecuado para la obtención de respuestas transitorias o bien para conocer la respuesta bajo la incidencia de un pulso modulado en frecuencia. MoM es una herramienta válida para el análisis a una sola frecuencia y en régimen permanente sinusoidal.

Los métodos numéricos eventualmente requieren de una notable carga computacional en el uso de la CPU y en espacio de memoria. FDTD requiere de operaciones muy básicas (sumas y diferencias) relacionando puntos cercanos físicamente. Comoquiera que hay que procesar un número muy elevado de puntos -especialmente en problemas abiertos- hay que resolver matrices muy dispersas y muy grandes. Para mejorar sus prestaciones es adecuado el uso de muchos procesadores trabajando en paralelo.

MoM requiere la inversión de matrices llenas, que representan todas las interacciones mutuas entre funciones base. Para la mejora de sus prestaciones, es por tanto crítica la aceleración del método de inversión. Básicamente, existen dos grandes técnicas para acelerar los algoritmos iterativos de inversión aplicados a las matrices de MoM. Por un lado, el uso de técnicas de precondicionado; es decir, pre- o post-multiplicar el sistema por una matriz mejor condicionada –y relacionada con la matriz de MoM original– de manera que el número de iteraciones total para converger disminuya drásticamente. Por otro lado, especialmente para problemas con dimensiones eléctricas suficientemente grandes, el cálculo de las interacciones entre funciones base suficientemente alejadas mediante el desarrollo analítico en ondas planas, lo cual supone una disminución en el tiempo de cálculo. Ejemplos de tales técnicas son las técnicas basadas en multipolos como el *Multilevel Fast Multipole Method (MLFMM)* [18].

Para el caso concreto de metamateriales, tales herramientas de análisis numérico -en general farragosas y con apreciable carga computacional- pueden ser simplificadas. A saber, dado que los metamateriales son objetos con variación periódica y grandes en términos de la longitud de onda, se pueden analizar como objetos con estructura periódica e infinitos. Esta asunción simplifica sobremanera el análisis numérico ya sea con FDTD que con MoM. En efecto, ahora la carga computacional reside solamente en el mallado en una de las células de periodicidad. Sin embargo, estas formulaciones -en concreto, FDTD con condiciones de contorno periódicas y Método de Momentos con modos de Floquet- no son realistas pues ignoran el efecto de los bordes de la agrupación y son útiles solamente bajo incidencia de onda plana. Sirven entonces para efectuar una primera aproximación a la descripción del problema y para permitir sopesar si es conveniente el posterior análisis riguroso con FDTD o MoM.

6. CONCLUSIONES

La capacidad de sintetizar materiales con las partes reales de permitividad y/o permeabilidad negativas es un campo de

investigación en el que se están invirtiendo grandes recursos dadas las potenciales aplicaciones en el mundo de las comunicaciones. El diseño de nuevos dispositivos de microondas y la fabricación de antenas con mejores prestaciones y un tamaño más reducido son algunas de las aplicaciones más inmediatas.

El esfuerzo actual en este campo está particularmente concentrado la verificación experimental de los comportamientos DNG observados en algunos materiales (y teóricamente demostradas, aunque centro de una gran controversia en la comunidad científica). Así mismo, la fabricación de materiales con comportamientos DNG en anchos de banda mayores que los actuales y en ángulos de incidencia diferentes del normal son también objeto de estudio.

En la actualidad, miembros del Departament de Teoria de la Senyal i Comunicacions de la Universitat Politècnica de Catalunya están involucrados en un proyecto coordinado con la Universitat Autònoma de Barcelona y con la Universidad Pública de Navarra para el diseño y caracterización de este tipo de materiales. El proyecto pretende no tan solo el desarrollo de herramientas de cálculo informáticas adaptadas a la resolución de problemas electromagnéticos sobre estructuras periódicas, sino también el diseño y fabricación de estructuras MM multibanda. Así mismo se pretende evaluar el potencial real de los materiales DNG para su utilización en el diseño de antenas.

En la actualidad, en el Departamento se está desarrollando una herramienta software propia llamada FIESTA 3-D (*Fast Integral Equation Solver for scatterers and Antennas in 3-D*) que permite el análisis de estructuras metamateriales empleando el método de los momentos. Esta herramienta es capaz de analizar geometrías con un elevado número de incógnitas empleando un ordenador personal.

Todas estas actividades se llevan a término aprovechando la capacidad de cálculo disponible en el Departamento, y a la disponibilidad de laboratorios de radiofrecuencia y de una cámara anechoica de grandes dimensiones que permiten la caracterización electromagnética precisa de antenas y materiales.

GLOSARIO

A continuación figuran, por orden alfabético, algunas de las siglas comúnmente empleadas en la terminología de esta especialidad. No todas ellos han sido referenciadas en el presente texto.

AMC: *Artificial Magnetic Conductor*;

BW: *Backward Wave*;

DNG: *Double Negative*;

DPS: *Double Positive*;

EBG: *Electromagnetic Band-Gap*;

ENG: *Epsilon Negative*;

FDTD: *Finite Differences in Time Domain*

FSS: *Frequency Selective Surface*;

FW: *Forward Wave*;

LHM: *Left-Handed Medium*;

MLFMM: *MultiLevel Fast Multipole Method*;



MM: *MetaMaterial*;
 MoM: *Method of Moments*;
 MNG: *Mu NeGative*;
 NIM: *Negative Index Medium* (medio con índice de refracción negativo);
 PBG: *Photonic Band-Gap*;
 PIM: *Positive Index Medium* (medio con índice de refracción positivo);
 RHM: *Right-Handed Medium*;
 SNG: *Single NeGative*;
 SRR: *Split Ring Resonator*.

REFERENCIAS

- [1] V.G. Veselago: «The electrodynamics of substances with simultaneously negative values of ϵ and μ », *Soviet Physics Uspekhi*, Jan.-Feb. 1968, **10**, (4), pp. 509-514.
- [2] J. B. Pendry, A. J. Holden, D. J. Robbins, and W. J. Stewart: «Magnetism from conductors and enhanced nonlinear phenomena», *IEEE Trans. on Microwave Theory Tech.*, **47**, (11), Nov. 1999, pp. 2075-2084.
- [3] R. Gonzalo, «Study of photonic crystal structures and their application in the field of antennas.» Ph.D. dissertation, Dept. Electric and Electronic Eng., Public Univ. of Navarra, Pamplona, Spain, 2000.
- [4] D. R. Smith, W. J. Padilla, D. C. Vier, S. C. Nemat-Nasser, S. Schultz: «Composite medium with simultaneously negative permeability and permittivity», *Physical Review Letters*, May 2000, **84**, (18), pp. 4184-4187.
- [5] R. A. Shelby, D. R. Smith, S. Schultz: «Experimental verification of a negative index of refraction», *Science*, 6 Apr. 2001, **292**, pp. 77-79.
- [6] R. A. Shelby, D. R. Smith, S. C. Nemat-Nasser, S. Schultz: «Microwave transmission through a two-dimensional, isotropic, left-handed metamaterial», *Applied Physics Letters*, Jan. 2001, **78**, (4), pp. 489-491
- [7] Y. Rahmat-Samii, H. Mossallaei: «Electromagnetic band-gap structures: classification, characterization, and applications», *11th International Conference on Antennas and Propagation*, 17-20 April 2001, pp. 560-564.
- [8] N. García, M. Nieto-Vesperinas: «Is there an experimental verification of a negative index of refraction yet?», *Optics Letters*, June 2002, **27**, (11), pp. 885-887.
- [9] R. W. Ziolkowski: «Double negative metamaterial design, experiments, and applications», *IEEE Antennas Propagat. Society Int. Sym.*, 2002, **2**, pp. 396-399.
- [10] P. Gay-Balmaz, O.J.F. Martin: «Efficient isotropic magnetic resonators», *Applied Physics Letters*, July 2002, **81**, (5), pp. 939-941.
- [11] Lecture notes of 2nd International PhD School on selected topics in applied electromagnetics «Metamaterials and RF microelectromechanical systems», Villa Umbra, Perugia (Italy), Sept. 2003. Prof.: T. Itoh, N. Engheta, R. W. Ziolkowski, A. M. Ionescu, T. Liseck, L. Vietzorrec, L. Buchaillet, P. Blondy, and R. Plana.
- [12] E. Úbeda, A. Heldring, J. Parrón, J. Romeu, J. M. Rius: «Numerical analysis of composite finite arrays of split ring resonators and thin strips», *IEEE Antennas Propagat. Society Int. Symp.*, 2003, 3, pp. 500-503.

[13] E. Úbeda, A. Heldring, J. Parrón, J. Romeu, J. M. Rius: «Radiation pattern of composite finite arrays of conducting elements with an exciting elementary dipole», *IEEE Antennas Propagat. Society Int. Symp.*, 2003, 3, pp. 379-382.

[14] Sobre GiD (interface de modelado gráfico): <http://gid.cimne.upc.es/>

[15] R.W. Ziolkowski, «Double Negative Metamaterial: Physics, design, and experiments», <http://online.kitp.ucsb.edu/online/q02/ziolkowski/>

[16] P. Valanju, «Left Handed Media (LHM) Righted», <http://www.utexas.edu/research/cemd/nim/Intro.html>

[17] S.M. Rao, D.R. Wilton, and A. W. Glisson: «Electromagnetic scattering by surfaces of arbitrary shape», *IEEE Trans. Antennas Propagat.*, **30**, May 1982, pp. 409-418.

[18] R. Coifman, V. Rohklin, and S. Wandzura: «The fast multipole method for the wave equation: a pedestrian description», *IEEE Antennas Propagat. Mag.*, **35**, pp. 7-12, 1993.

AUTORES



José María González, Ingeniero de Telecomunicación y Doctor Ingeniero de Telecomunicación por la UPC, ha desempeñado diversos puestos técnicos en empresas del sector de las telecomunicaciones y ha sido profesor en la Universidad de Zaragoza. En la actualidad es investigador contratado a través del Programa Ramón y Cajal del Ministerio de Ciencia y Tecnología en su convocatoria de 2001. Su actividad investigadora y docente se desarrolla en el Departamento de Teoría de la Señal y Comunicaciones de la UPC. Sus principales temas de trabajo son el análisis, diseño y caracterización de antenas en sistemas de medidas convencionales y a través de termografías infrarrojas.



Eduard Úbeda, es investigador Ramón Cajal en el Departamento de Teoría del Señal i Comunicacions (UPC). Tiene experiencia investigadora en el Joint Research Center, en Ispra, Italia, centro dependiente de la Unión Europea, en los años 1996 y 1999. En los años 2002 y 2003 trabajó como Visiting researcher en el Electromagnetic and Communication Lab de Pennsylvania State University (PSU). Sus áreas de interés principales son la computación numérica de dispersión y radiación electromagnética usando ecuaciones integrales.



Jordi Romeu, Ingeniero de Telecomunicación y Doctor Ingeniero de Telecomunicación por la UPC. Es Catedrático del Departamento de Teoría de la Señal y Comunicaciones y profesor en la ETSETB. Sus líneas de investigación son el análisis, diseño y medida de antenas. Por sus trabajos de investigación ha recibido el Gran Premio IST otorgado por la Comisión Europea, el premio Ciutat de Barcelona y la Distinción de la Generalitat para la Promoción de la Investigación. Es coautor de más de 30 artículos en revistas internacionales y del libro *Antenas*.



SONIDO VIRTUAL O CÓMO MOVER UNA FUENTE MONOFÓNICA POR EL ESPACIO

Màrius Valles Segura, Néstor Andreu Ramírez, Enric Monte Moreno

marius_valles@menta.net, nestor_andreu@yahoo.es, enric@gps.tsc.upc.es

ABSTRACT

En este artículo se presenta un posible modelo capaz de sintetizar sonido virtual a partir de una fuente monofónica manteniendo la ilusión de espacialidad. Este modelo se basa en una descripción simplificada en el dominio temporal del comportamiento de las ondas sonoras (teniendo en cuenta el filtrado que el cuerpo realiza sobre ellas) y en las características más relevantes de la percepción humana concernientes a su localización (difracción, retardo y reflexión del sonido).

1. INTRODUCCIÓN

La reproducción del sonido en 3D está incrementando su importancia dentro del mundo científico, comercial y del ocio, por lo que la demanda de sistemas capaces de operar en tiempo real aumenta cada vez más.

El oído humano es capaz de situar en el espacio una fuente de sonido. Para lograrlo tiene en cuenta las reflexiones y el filtrado que la configuración anatómica de cada oyente realiza. La función de transferencia que relaciona el sonido que llega al tímpano con el emitido por la fuente se denomina HRTF (*Head Related Transfer Function*). Ésta puede ser sintetizada desarrollando un modelo estructural relativamente simple cuyos parámetros están estrechamente relacionados con las características fisiológicas de cada individuo.

La correcta simulación de la HRTF permite crear la ilusión de especialidad en el cerebro, permitiendo que pueda localizar la procedencia de una onda sonora determinada.

2. PERCEPCIÓN ESPACIAL DEL SONIDO

Cuando un sonido proviene de algún lugar del espacio, una persona es capaz de localizar su procedencia en azimut, elevación y distancia. Sin duda, la estimación de su posición en azimut es mucho más precisa que el resto. La *Teoría Duplex* de Lord Rayleigh da cuenta de los principales factores que intervienen. Principalmente se habla de la diferencia de tiempo interaural o ITD (*interaural time difference*) y de la diferencia de nivel interaural o ILD (*interaural level difference*). Antes de definirlos, es

necesario establecer un sistema de coordenadas que permita especificar la posición de la fuente respecto al oyente: un sistema esférico centrado en la cabeza del oyente. La figura 2.1 muestra gráficamente su comportamiento:

2.1 Localización en el plano de azimut

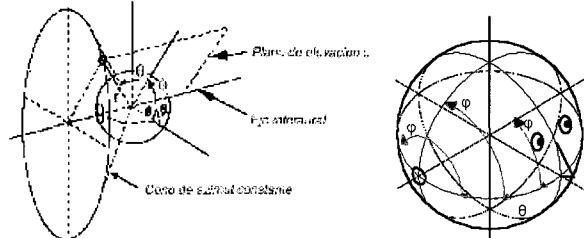


Figura 2.1. Sistema de coordenadas polar interaural. Se puede observar que la distinción entre delante y detrás se realiza mediante la elevación (j) y no azimut (q).

Por lo general, el sonido llega en momentos diferentes a cada oreja ya que debe recorrer caminos distintos, esta diferencia de tiempos es el ITD y se puede calcular de manera geométrica de la siguiente manera (Figura 2.2):

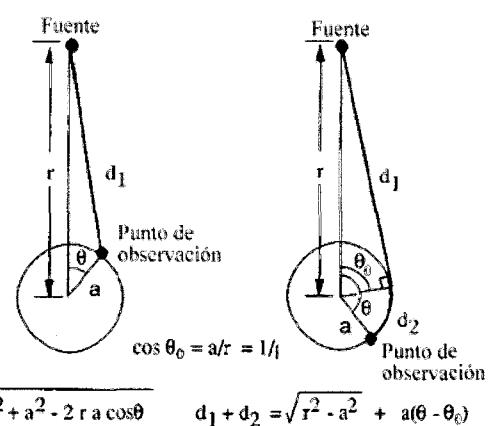


Figura 2.2. El sonido puede llegar de forma directa al punto de observación (d_1) o de manera indirecta ($d_1 + d_2$).

Así pues, esta diferencia de tiempo $\Delta\tau$ viene dada por:



$$\Delta\tau = \frac{c\Delta t}{2\pi a} = \begin{cases} \frac{1}{2\pi} (\sqrt{\rho^2 - 2\rho \cos(\theta)} + 1 - \rho) & \text{para } 0 \leq \theta \leq \theta_o \\ \frac{1}{2\pi} (\theta - \theta_o + \sqrt{\rho^2 - 1 - \rho}) & \text{para } \theta_o \leq \theta \leq \pi \end{cases} \quad (1)$$

donde,

$$\rho = r/a \quad (2)$$

$$\theta_o = \sin^{-1}(1/\rho) \quad \text{para } \rho \geq 1$$

Existe, también, una aproximación del ITD muy realista calculada de manera empírica. Su valor oscila entre dos valores diferentes, dependiendo si es a bajas frecuencias o a altas, de la siguiente manera (siendo k el número de onda ()):

$$ITD_{baja} = \frac{3a}{c} \cdot \sin\theta, \quad \text{para } ka \ll 1 \quad (3)$$

$$ITD_{alta} = \frac{2a}{c} \cdot \sin\theta, \quad \text{para } ka \gg 1 \quad (4)$$

Estos dos valores de ITD son sólo valores asintóticos, en realidad el ITD tiene una pendiente decreciente en función de la frecuencia. La figura 2.3 muestra cómo varía el ITD real según la frecuencia para una fuente situada a $\theta = 40^\circ$ y dichos valores asintóticos.

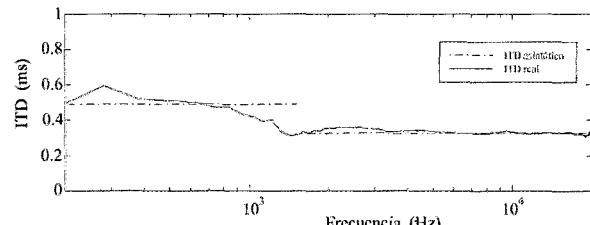


Figura 2.3. Variación del ITD según la frecuencia con cabeza esférica.

La cabeza produce una zona de sombreado acústico en el lado opuesto al que le llega el sonido directo, es decir, existe una diferencia en el nivel sonoro de las señales en ambos oídos (ILD). Lord Rayleigh solucionó la ecuación de onda en una esfera, de esta manera se puede determinar exactamente la medida de esta difracción producida para un frente incidente de ondas planas (fuente situada en el infinito). Como es de esperar, el ILD es altamente dependiente de la frecuencia. En la figura 2.4 se muestra la respuesta de una esfera rígida de radio a según la solución de Rayleigh. En ella se observa la relación existente entre la intensidad de la onda en el punto de incidencia perpendicular y la intensidad en un punto situado a un ángulo β de éste.

Obviamente, el considerar la cabeza como una esfera es una aproximación que permite calcular rápidamente el ILD. En realidad, en la cabeza se producen numerosas reflexiones y difracciones que, en altas frecuencias, provocan un ILD aún mayor. Además, el ILD no sólo depende del ángulo y de la frecuencia sino también de la distancia a la que se encuentra

la fuente, sobretodo para distancias inferiores a 1m., Donde el ILD tiende a ser mayor y se comporta de un modo ligeramente distinto.

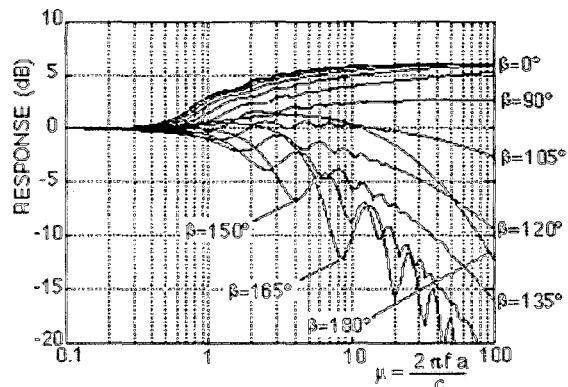


Figura 2.4. Respuesta de una esfera rígida de radio a para el cálculo del ILD.

Según la teoría Duplex, el ILD y el ITD son fenómenos complementarios. A frecuencias bajas ($=1,5$ KHz) la diferencia de tiempos aporta información importante para la localización mientras que la diferencia de nivel es casi despreciable. Por otro lado, a partir de 3 KHz, es más importante el ILD que el ITD. Es por este motivo que el sistema de localización tiende a ser pobre cuando se intenta localizar un sonido que sólo tenga componentes frecuenciales en el margen de 1,5 KHz a 3 KHz.

2.2 Localización en el plano de elevación

Si se mantiene la aproximación de una cabeza esférica ideal, el ITD y el ILD no varían en función de la elevación. Sin embargo, el oído externo sí produce un filtrado que depende de la elevación, por lo que resulta de especial importancia para detectar la elevación de la fuente. Su configuración anatómica lo hace funcionar como una antena receptora, con cavidades resonantes que amplifican unas determinadas frecuencias y configuraciones geométricas que atenuan otras. Su respuesta depende del ángulo de incidencia del sonido, tal y como se muestra en la figura 2.5.

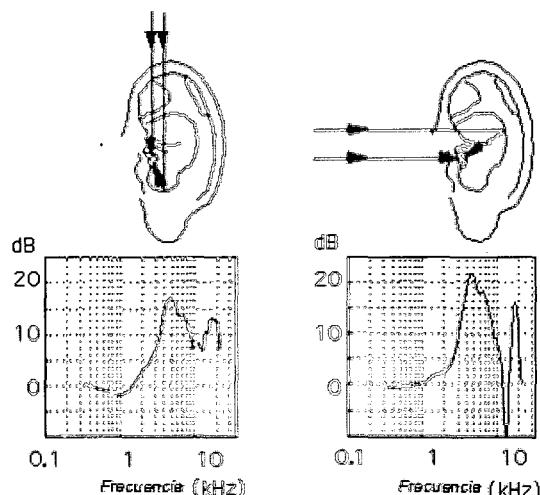


Figura 2.5. Respuesta frecuencial del oído externo con diferentes ángulos de incidencia.

Como se puede observar en la figura anterior, el sonido entra en el conducto auditivo de dos maneras diferentes: por el camino directo sin retardo y a través de reflexiones en el oído externo retardadas. A frecuencias bajas, básicamente, la energía de los diferentes caminos recorridos se suma en fase. Sin embargo, a altas frecuencias las señales que llegan retardadas no tienen por qué llegar en fase con la señal del camino directo y, por lo tanto, se pueden producir interferencias destructivas. La interferencia destructiva máxima tiene lugar cuando la distancia que recorre el sonido reflejado es: $d = \lambda/2$, siendo λ la longitud de onda, y generalmente se produce entre 6KHz y 16KHz, dependiendo de la morfología del oído del oyente. Además, cabe destacar que este pico mínimo es mayor cuando el sonido incide frontalmente que cuando lo hace desde arriba y, como la distancia d cambia en función del ángulo de incidencia, la posición de éste varía en función de la elevación. Y es precisamente esta variación de la frecuencia en función de la dirección la que el cerebro es capaz de reconocer para localizar la fuente sonora. Además, para una fuente situada en el plano de $\theta=0^\circ$ (plano vertical que divide la cabeza en dos mitades iguales), tanto el ILD como el ITD son nulos, de forma que ésta es la única forma de localizar la fuente.

2.3 Localización en distancia

La localización en distancia no depende de unos parámetros tan fácilmente caracterizables como el ITD o el ILD. Es necesario el uso de diversas técnicas como, por ejemplo, añadir algún eco y cola reverberante al sonido tratado (de forma que la impresión es mucho más natural, puesto que cualquier sonido que se escucha en la realidad presenta estos fenómenos) o modificar el volumen al que la fuente suena.

Hay que tener en cuenta que, cuando una fuente suena muy cercana a un oído, proporcionalmente existe una distancia bastante mayor que la separa del otro. Esto provoca que exista un ILD mucho mayor del que causaría la cabeza por sí sola. Este incremento de ILD es más independiente de la frecuencia, pues está causado principalmente por la atenuación en distancia, y es determinante para localizar fuentes cercanas al oyente, especialmente para rangos inferiores a un metro.

Otro fenómeno útil es el propio movimiento de la cabeza para ayudar a determinar la distancia. Si la fuente está en un punto cercano, la posición en azimut en la que se localiza variará según se mueva el oyente. En cambio, si la fuente se halla lejos, su azimut apenas variará. Es probable que el cerebro use este fenómeno, conocido como paralaje, para precisar mejor la distancia a la que se encuentra la fuente.

3. SÍNTESIS DE SONIDO VIRTUAL

El modelo que se propone es un modelo simple de procesado de señal capaz de sintetizar sonido binaural a

partir de una fuente monoaural. Éste mantiene una correspondencia unívoca con los hombros, la cabeza y el oído externo del oyente, cada elemento aporta una característica temporal única a la respuesta impulsional global. El principal motivo por el que se presenta el modelo en el dominio temporal es porque el sonido llega al oído interno a través de múltiples caminos. Señales que llegan por caminos de longitud diferente interactúan de una manera mucho más clara en el dominio temporal que en el frecuencial. El esquema del modelo general se presenta en la figura 3.1. Principalmente, el modelo de la cabeza y los hombros proporcionan información para localizar en azimut y en rango, generando el ITD e ILD necesarios para cada posición que se procese, mientras que el modelo del oído externo permite la localización en elevación. A continuación se detallará individualmente cada uno de sus componentes.

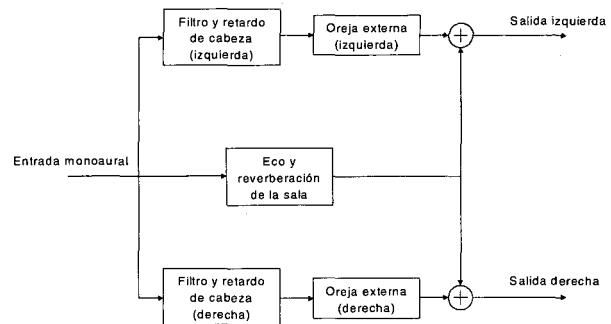


Figura 3.1. Componentes del modelo

3.1 Modelado de la cabeza

Es posible sintetizar el ILD con una función simple de un polo y un cero, consiguiendo así modelar el efecto de sombreado acústico que causa la cabeza. La siguiente ecuación muestra dicho filtro de sombreado acústico:

$$H_{sa}(\omega, \beta) = \frac{1 + j \frac{\alpha(\beta)\omega}{2\omega_0}}{1 + j \frac{\omega}{2\omega_0}} \quad \text{para } 0 \leq \alpha(\beta) \leq 2 \quad (5)$$

con valor de la constante $\omega_0 = c/a$.

La frecuencia normalizada referida a la constante ω_0 es $\mu = \omega_0 a / c = 1$. El coeficiente α depende del ángulo de incidencia β , controlando así la posición donde se encuentra el cero. Si $\alpha=2$ existe un pico a altas frecuencias de 6dB, mientras que para $\alpha<1$ hay una zona de corte. Para coincidir con el patrón de las curvas mostradas en la figura 2.4, α debe depender de β de la siguiente manera:

$$\alpha(\beta) = \left(1 + \frac{\alpha_{\min}}{2}\right) + \left(1 - \frac{\alpha_{\min}}{2}\right) \cos\left(\frac{\beta}{\beta_{\min}} 180^\circ\right) \quad (6)$$



donde las constantes que producen una buena aproximación son:

$$\alpha_{\min} = 0.1$$

$$\beta_{\min} = 150^\circ$$

Por otro lado, para sintetizar el ITD basta con modelar un elemento retardador correspondiente con la ecuación 2. De esta manera, si se encadenan ambos filtros se obtiene la aproximación $\hat{H}(\infty, \mu, \beta)$.

Además, nótese que H_{sa} introduce el siguiente retardo de grupo:

$$\tau_g = \frac{1 - \alpha(\beta)}{2\omega_o} = \frac{a}{2c}(1 - \alpha(\beta)) \quad (7)$$

el cual, aumenta el tiempo $\Delta\tau$ para frecuencias bajas ($\mu < 1$), de forma que lo acerca más a los resultados observados por Khun [3]. De hecho, para $\beta=0^\circ$ este retardo de grupo añade exactamente un retardo adicional del 50%. La figura 3.2 muestra la respuesta frecuencial del filtro H_{sa} y la respuesta temporal del modelo retardador. Se puede observar como las curvas de respuesta frecuencial se aproximan de manera satisfactoria a las teóricas mostradas en la figura 3.2:

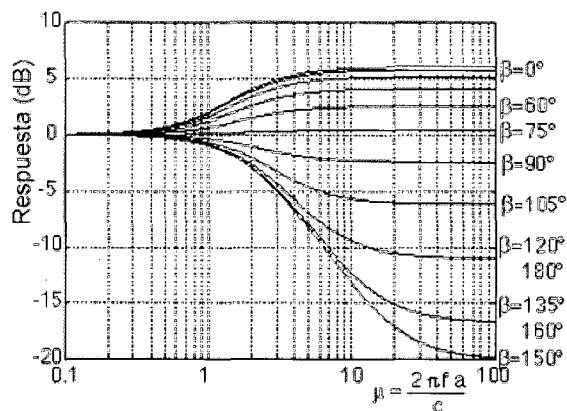


Figura 3.2. Respuesta frecuencial del filtro de un polo y un cero que aproxima la respuesta de una esfera rígida.

3.2 Modelado del oído externo

El oído externo se simula mediante un modelo de propagación multicamino que simula los eventos observados en diversos experimentos, cuyo esquema se puede observar en la figura 3.3:

Realizando tests auditivos se ha podido determinar que cinco eventos son suficientes para este modelo y que su amplitud se puede aproximar como constante sea cual sea la dirección de la que la que provenga el sonido. Lo único

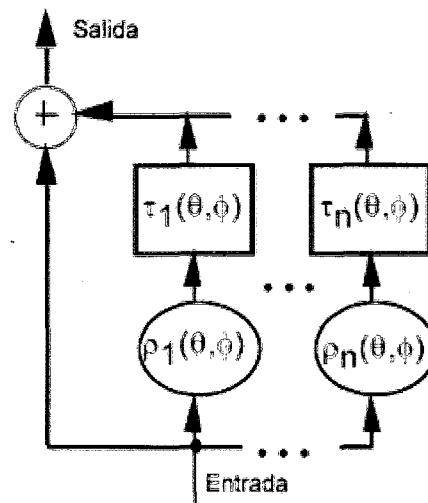


Figura 3.3. Esquema del modelo del oído externo. Donde ρ_k y τ_k son, respectivamente la amplitud y el retardo de los eventos.

que varía son sus retardos, que se pueden aproximar por la siguiente ecuación:

$$t_k(q, f) = A_k \cos(q/2) \sin(D_k(90^\circ - f \cos q)) + B_k \quad (8)$$

Se da una cierta variación entre distintos individuos en la morfología de la oreja externa. Para adaptarse a ellas solamente hace falta variar los parámetros D_k . A continuación se presenta una tabla con los valores de los parámetros y las amplitudes para los cinco eventos y para un individuo concreto. A_k y B_k se dan en muestras para una frecuencia de muestreo de 44100 Hz:

k	ρ_k	A_k	B_k	D_k
1	0.5	1	2	1
2	-1	5	4	0.5
3	0.5	5	7	0.5
4	-0.25	5	11	0.5
5	0.25	5	13	0.5

3.3 Eco y reverberación de la sala

Para poder externalizar bien el sonido y así, localizar en rango, es necesario crear la impresión de encontrarse en una sala. Para ello, se deben generar ecos y reverberaciones del sonido original y añadirlos al sonido ya tratado. Se ha encontrado que un modelo rápido, pero suficientemente preciso, consiste en crear un filtro que genere un solo eco y que, además, disponga de unos cuantos lazos de realimentación para generar la reverberación. El número de coeficientes, así como su valor y retardo, dependerán del tipo de sala o efecto que se quiera conseguir (de forma aproximada), y la mejor manera de encontrarlos es de forma empírica. Cuantos más lazos de realimentación se usen, más rica resulta la reverberación, aunque con cinco o seis ya es posible crear reverberaciones convincentes.

Para hacer la reverberación más realista, hay que hacer un filtrado paso-bajo a la señal antes de que pase por los lazos, de forma que se simula el hecho de que los materiales absorben mejor las altas frecuencias que las bajas.

4. IMPLEMENTACIÓN DEL MODELO

Hasta este punto, se ha presentado el modelo de manera teórica. Sin embargo, si se desea implementarlo sobre algún lenguaje de programación se deben tener diversas consideraciones en cuenta. La principal diferencia radica en la posibilidad de realizar las simulaciones con fuentes dinámicas, es decir, de definir una trayectoria sobre la cual se debe mover la fuente. La figura 4.1 muestra el esquema para una implementación práctica del modelo:

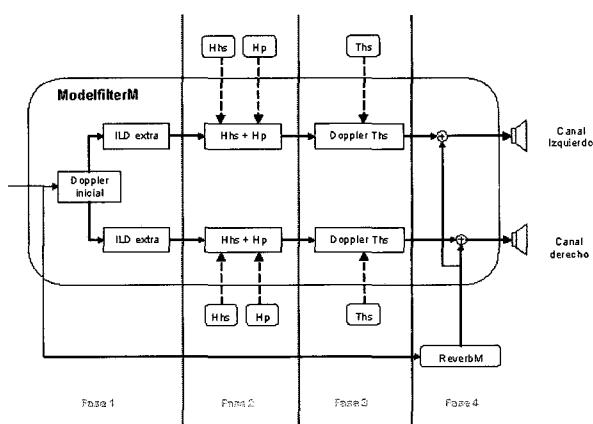


Figura 4.1. Modelo para la implementación práctica del sistema.

La señal de entrada se divide en bloques para acelerar el procesado, puesto que hay muchos cálculos que sólo es necesario realizarlos una vez por bloque (por ejemplo, los coeficientes de los filtros o el cálculo de los retardos). Cuanto mayores son los bloques, mayor es la rapidez de procesado, pero también es mayor la latencia del sistema y menor su precisión espacial. El compromiso, sin embargo, es fácil de encontrar, puesto que la eficiencia prácticamente no mejora para bloques mayores de 300 muestras, y con este tamaño tanto la latencia como la resolución espacial son totalmente aceptables.

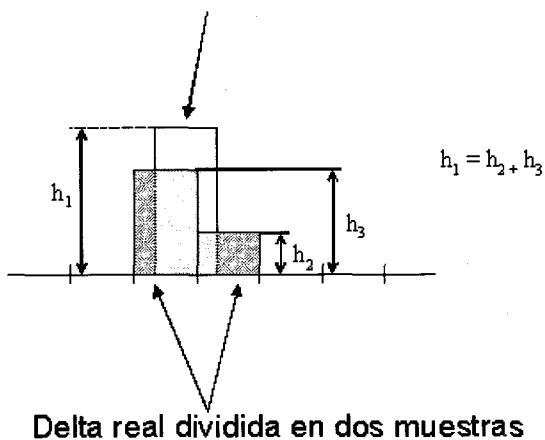
Como se ve, el procesado se realiza en cuatro fases encadenadas, que a continuación se explican más detalladamente:

Fase 1: Aquí se calcula el efecto Doppler relacionado con el movimiento de la fuente respecto del centro de la cabeza del oyente. También se introduce el ILD extra que aparece cuando la fuente se encuentra cerca del oyente. El efecto Doppler se calcula mediante repetición o supresión de muestras, según convenga, para obtener mayor velocidad. Otros

sistemas más precisos que evitan efectos de aliasing se han descartado por el enorme incremento de tiempo de cálculo que representan.

Fase 2: En esta fase se realiza el filtrado correspondiente tanto a la sombra acústica como a las orejas exteriores. El primero (Hhs) es de implementación inmediata. Sin embargo, la realización del segundo (Hp) tiene la dificultad de que los eventos de la respuesta impulsional no caen normalmente justo en el tiempo de inicio de una muestra. Para subsanar esto, se divide un evento en dos muestras de la respuesta impulsional de forma proporcional, tal y como indica la figura 4.2

Posición teórica de la delta



Delta real dividida en dos muestras

Figura 4.2. Separación de un evento (delta) en dos muestras cuando su inicio no coincide exactamente con el de ninguna muestra.

Fase 3: Aquí se generan los retardos entre canales derecho e izquierdo de forma que se adapten al movimiento de la fuente. Para ello es necesario volver a aplicar efectos Doppler que, aunque sus efectos en frecuencia son prácticamente inaudibles, comprimen o expanden las señales para que se adapten a dichos retardos.

Fase 4: En esta última fase se añade la salida del módulo de reverberación, obteniendo la salida binaural.

Este sistema está pensado para que su salida se conecte directamente a unos auriculares, puesto que las señales calculadas para los canales derecho e izquierdo son exactamente las que deben recibir las respectivas orejas. Es posible, sin embargo, obtener el mismo resultado con dos altavoces, aunque el sistema se complica porque el altavoz de un lado tiene que emitir a la vez su señal y la que debe cancelar la proveniente del otro altavoz. Este sistema tendría el problema añadido de que dicha cancelación sólo se obtendría en una zona muy pequeña de la sala (figura 4.3):

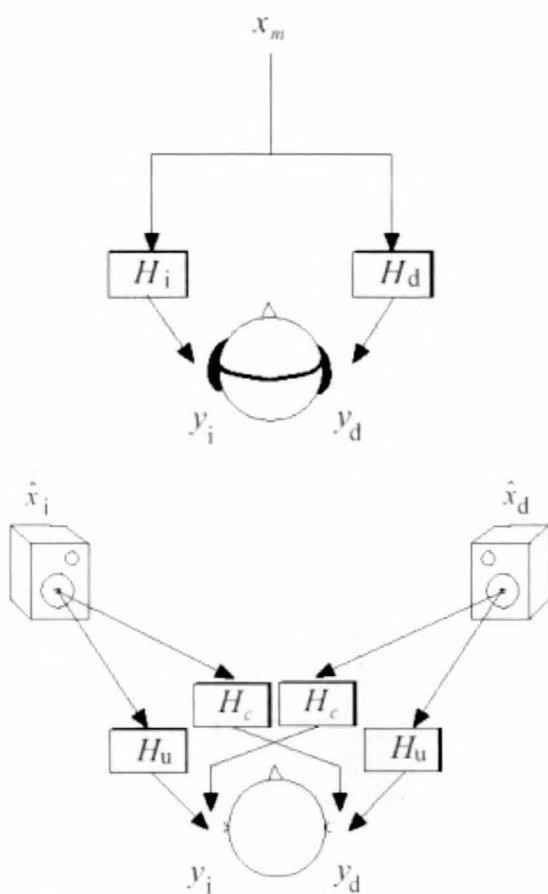


Figura 4.3. Métodos de reproducción: con auriculares (filtrado independiente de la señal) o con dos altavoces (filtrado útil de la señal y filtrado cancelador de crosstalk H_c).

5. CONCLUSIONES

Se ha presentado un modelo simple de procesado de señal para las síntesis de sonido virtual. Dicho modelo, contiene componentes que procesan, de manera independiente, los elementos que hacen posible la localización de una fuente en el espacio: el ILD y ITD para la localización en azimut, el modelado del oído externo para la elevación y la reverberación para el rango. Los diferentes experimentos acústicos realizados han demostrado que con este modelo se puede simular de manera convincente la percepción espacial del sonido. Además, la simplicidad de los filtros usados y el planteamiento en si del modelo, lo convierten en un sistema ideal para una implementación a tiempo real.

Otra ventaja añadida de este sistema es que se puede adaptar a cada usuario, puesto que permite personalizar los parámetros más relevantes para la percepción (los referidos, fundamentalmente, a la configuración anatómica del oído externo). De este modo, se incrementa la sensación derealismo.

En definitiva, la sencillez, la eficiencia y su versatilidad hacen que el sistema presentado en este artículo resulte óptimo para usos tanto científicos como comerciales, ya que permite una simulación rápida y lo suficientemente precisa de sonido 3D.

6. REFERENCIAS

- [1] Brown, C. Phillip & Duda, Richard O.; «A Structural Model For Binaural Sound Synthesis», 1998.
- [2] Brown, C. Phillip & Duda, Richard O.; «An Efficient HRTF Model For 3-D Sound», 1997.
- [3] Kuhn, G. F.; «Model for the interaural time differences in the azimuthal plane,» *J. Acoust. Soc. Am.*, vol. 62, no. 1, pp. 157-167, July 1977.
- [4] Hupaniemi, Jyri; «Virtual Acoustic and 3D sound in multimedia signal processing», Helsinki University of Technology (Laboratory of Acoustics and Audio Signal Processing), Report 53, 1999
- [5] Algazi, V. Ralph & Duda, Richard O.; «Approximating the head-related transfer function using simple geometric models of the head and torso», *J. Acoust. Soc. Am.*, vol. 112 no. 1 , Nov 1992.
- [6] Algazi, V. Ralph et al.; «Structural Composition and decomposition of HRTFs», IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001.
- [7] Algazi, V. Ralph et al.; «The CIPIC HRTF database», IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001
- [8] Hartmann, William M.; «How we localize sound», Physics Today, Npv 1999.

AUTORES



Marius Valles Segura. Desde 1998 ha estado cursando Ingeniería Superior de Telecomunicaciones, y actualmente está realizando el Proyecto de Fin de Carrera. Ha participado en varias ediciones del Concurso Nacional de robots de la UPC. Ha realizado un curso de acústica y ya se encuentra asociado al colegio de ingenieros de telecomunicaciones.



Néstor Andreu Ramírez. Desde 1998 ha estado cursando Ingeniería Superior de Telecomunicaciones, y actualmente está realizando el Proyecto de Fin de Carrera. Ha participado en varias ediciones del Concurso Nacional de robots de la UPC. Ha realizado un curso de acústica y ya se encuentra asociado al colegio de ingenieros de telecomunicaciones.



Enric Monte Moreno. Se graduó y se doctoró en Ingeniería de Telecomunicación por la Universidad Politécnica de Catalunya (UPC) en 1987 y 1992 respectivamente. Vinculado al departamento de Teoría de Señal y Comunicaciones desde el año 1989, actualmente ejerce el cargo de profesor titular. Su interés actual se centra en el tratamiento de señal de voz.

APLICACIÓN DE MÉTODOS DE COMPRESIÓN DE CABECERAS TCP/IP COMO OPTIMIZACIÓN DE LAS COMUNICACIONES MÓVILES



Anna Calveras Augé, María Luisa Catalán Cid

Grupo de Comunicaciones Móviles
Departamento de Ingeniería Telemática (ENTEL)
Universidad Politécnica de Cataluña (UPC)
Barcelona, España

Abstract- El protocolo TCP (Transmission Control Protocol) está implementado para actuar en redes fijas, donde las incidencias en la comunicación pueden interpretarse como problemas de congestión. En entornos móviles, en cambio, los problemas son debidos a una alta tasa de error en bit (BER); la utilización del protocolo TCP en este tipo de entornos afecta negativamente a la eficiencia de la transmisión. En este artículo se presenta una solución para optimizar la comunicación TCP: la compresión de cabeceras; concretamente se expone la necesidad de un algoritmo específico para comunicaciones móviles y se valora la propuesta de compresión de cabeceras de la IETF.

1. INTRODUCCIÓN

Con el auge de las tecnologías móviles en las últimas décadas, ha aparecido también la necesidad de que el usuario pueda comunicarse a cualquier hora y desde cualquier parte, ya sea mediante servicios de voz o de transferencia de datos (e-mail, acceso a Internet, MMS). La aparición de los servicios de 2.5G y 3G, tales como GPRS o UMTS, ofrecen la posibilidad de proporcionar al usuario acceso a todos estos recursos con las ventajas de la tecnología inalámbrica y con el único requerimiento de que exista cobertura.



Figura 1. Esquema de una conexión a Internet mediante un terminal móvil

Internet está basada en la arquitectura de protocolos TCP/IP. Estos están diseñados para actuar de forma eficiente en redes fijas, donde la tasa de error en bit (BER) es baja y los principales problemas son causados por la congestión de la red. Las comunicaciones móviles, en cambio, se realizan a través de enlaces inalámbricos, caracterizados por una alta BER. Especialmente, en condiciones de movilidad de terminales y reselección de celdas, pueden produ-

cirse ráfagas de error pronunciadas que afectan negativamente al rendimiento de los protocolos TCP/IP [3].

Por otra parte, los retardos extremo-extremo (RTT)¹ en una comunicación móvil toman valores muy superiores a los obtenidos en redes fijas (entre centenares de milisegundos en sistemas 3G y entre 1 y 3 segundos en comunicaciones GPRS) [15]. Además, los sistemas móviles permiten combatir la alta probabilidad en bit mediante la actuación de mecanismos de recuperación de errores a nivel de enlace RLC/LLC (por ejemplo, después de que se haya llevado a cabo una reselección de celda o después de una interrupción en la comunicación); esto, unido al hecho de que durante la transmisión pueden limitarse la cantidad de recursos (slots) disponibles, puede provocar variaciones importantes del retardo entre paquetes consecutivos. Por lo tanto, se provocará un funcionamiento poco eficiente del mecanismo habitual de TCP [14] para el cálculo del temporizador de retransmisión (RTO).

La Figura 2 ejemplifica ambos aspectos. Se muestran los resultados de la captura (en el terminal móvil) de una transmisión realizada en una situación de movilidad. Tal como puede observarse, debido al paso de los paquetes a través del enlace inalámbrico, se producen dos fenómenos importantes a destacar. Por una parte cortes o desconexiones en los que se pierden ráfagas de paquetes, por otra, retardos variables y de una duración considerable.

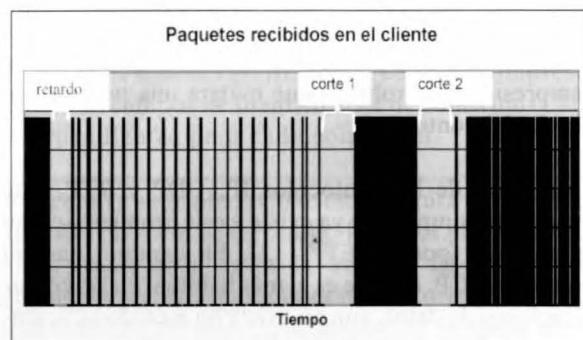


Figura 2. Fragmento de una transmisión. Escenario móvil

¹ Se denomina retardo extremo-extremo ó Round Trip Time al tiempo transcurrido entre que se envía un paquete y se recibe la confirmación.

Desde el punto de vista del usuario, otra característica negativa de este tipo de comunicaciones es que se llevan a cabo en enlaces de baja velocidad. Actualmente en Internet y redes IP se utiliza la pila de protocolos TCP/IP en su versión 4. En estos sistemas se envía en cada paquete un mínimo de 40 bytes de cabecera a nivel TCP/IP. Sin embargo, para poder dar servicio a más usuarios y facilitar la movilidad, en un futuro será necesaria la utilización del protocolo IPv6 [9] que entre otras características dispone de un campo de direcciones mayor (128 bits frente a los 32 de IPv4), por lo que la cabecera TCP/IP se verá incrementada. Este hecho reducirá el tamaño del campo disponible para enviar información y se hará aún más plausible la sensación de baja velocidad y la consecuente insatisfacción del usuario. Por todo esto se justifica la necesidad de métodos adicionales que permitan optimizar el protocolo TCP para obtener un mejor rendimiento en entornos móviles.

Si se analizan los campos de las cabeceras de los principales protocolos de nivel de transporte (TCP/IP, UDP, RTP), puede observarse que para una transmisión concreta, gran parte de los campos de las cabeceras se mantienen constantes, son deducibles o varían siguiendo un cierto patrón. Una manera de reducir la influencia de la cabecera en el tamaño total del paquete sería la utilización de algoritmos de compresión: el compresor únicamente envía los campos de cabecera necesarios para que el descompresor pueda reconstruir la cabecera original a partir de ese paquete y de un cierto contexto². Así se aumentaría el throughput de la transmisión. Este hecho mejoraría la apreciación del usuario sobre el rendimiento de las comunicaciones móviles de baja velocidad.

No obstante, la mayoría de los algoritmos de compresión existentes [7, 10], aunque fueron ideados para enlaces de baja velocidad no son directamente aplicables a las comunicaciones móviles. Esto es debido, sobretodo, a que no están suficientemente adecuados para soportar altas BER. En el caso de pérdida de un solo paquete, se pierde el contexto y se descartan muchos paquetes en el descompresor antes de volver a recuperarlo. Por lo tanto, sería necesario un método de compresión más robusto que evitara una pérdida tan rápida del contexto.

En el caso de los protocolos IP, UDP y RTP, este método de compresión ya existe y está propuesto como un estándar por el IETF [12]. En cambio, para el protocolo TCP, aunque es el más habitual en las comunicaciones de datos, aun no existe un método suficientemente eficiente y robusto de compresión.

El grupo de trabajo ROHC de la IETF [18] en el área de transporte está elaborando una nueva propuesta de compresión, el [17].

En el artículo se analiza la viabilidad de optimizar las comunicaciones móviles mediante métodos de compresión de cabeceras. En concreto, se detallarán los problemas de los mecanismos de compresión de la cabecera TCP/IP actuales [7, 10] y se presentarán posibles soluciones. En particular nos centraremos en la propuesta que actualmente desarrolla el grupo ROHC de IETF [17], basada en el protocolo de compresión robusta ROHC [12].

El artículo se estructura de la siguiente forma. En las secciones 2 y 3 analizaremos el funcionamiento y los problemas que presentan los métodos actuales de compresión de cabeceras TCP: Van Jacobson e IPHC, en enlaces móviles. En la sección 4 presentaremos las características y las mejoras respecto a los protocolos de otras propuestas anteriores, en concreto la del grupo de trabajo ROHC del IETF para un protocolo de compresión robusta [17]. Finalmente, en la sección 5 concluiremos por qué creamos que la compresión de cabeceras, y en concreto la propuesta [17], podría ser una buena opción para la optimización de las comunicaciones móviles de 2.5G y 3G.

2. COMPRESIÓN VAN JACOBSON

El protocolo de compresión Van Jacobson [7] fue ideado para comunicaciones unidireccionales mediante el protocolo TCP/IPv4 en enlaces fijos de baja velocidad (PPP, SLIP). Basa su funcionamiento en mantener, para cada conexión, una cabecera base (o contexto) común en compresor y descompresor. En el compresor la cabecera base se actualiza con la información de cabecera de cada nuevo paquete enviado; en el descompresor ésta se inicializa cuando se envía un paquete sin comprimir y se actualiza mediante la información que se transmite en cada paquete comprimido.

La cabecera comprimida incluye como mínimo un primer byte a modo de máscara que indica los campos de la cabecera que se transmiten a continuación, y el checksum TCP de la cabecera original, necesario para la detección de errores en la descompresión³. El resto de los campos mostrados en la Figura 3 se transmiten sólo si varían respecto a los campos de la cabecera base. Para reducir al máximo el tamaño de la cabecera comprimida, en vez del valor real del campo del paquete a comprimir, se transmite la diferencia entre éste y el almacenado en la cabecera base del compresor.

² Se denomina contexto a toda la información necesaria (normalmente se trata de los campos de cabecera de un paquete anterior) para recuperar la cabecera original de un paquete comprimido

³ Se considera que un paquete se ha descomprimido correctamente si el CRC de la cabecera descomprimida coincide con el Checksum TCP de la cabecera original.

En el caso de que varíen otros campos no contemplados en la estructura de la cabecera comprimida o que se produzca una retransmisión, debe volver a enviarse un paquete sin comprimir para actualizar la cabecera base del descompresor.

C	I	P	S	A	W	U
C: Identificador de conexión						
Checksum TCP (2 bytes)						
U: Puntero Urgente						
W: Δ Window						
A: Δ ACK						
S: Δ Secuencia						
I: Δ Identificador IP						

Figura 3 Cabecera comprimida (VJacob)

Al tratarse de una comunicación unidireccional, la compresión Van Jacobson hace uso del mecanismo de reconocimientos de TCP para sincronizar la cabecera base en caso de pérdida o error de los paquetes. Si se pierde un paquete o se descarta por ser erróneo, la cabecera base en el descompresor no podrá actualizarse y, por tanto, diferirá de la del compresor que sí se habrá actualizado. Al estar los campos codificados de forma incremental, la cabecera del próximo paquete comprimido no podrá reconstruirse correctamente, ya que el Checksum TCP enviado será diferente del calculado en el descompresor. Por lo tanto, se descartará el paquete y el destino no enviará ningún ACK al emisor. Así sucederá con todos los paquetes que lleguen al descompresor, hasta que expire el temporizador en el emisor y se reenvíe el paquete erróneo o perdido (sin comprimir) y se vuelva a sincronizar el contexto del descompresor.

Esta política de recuperación de errores presenta tres grandes problemas en las comunicaciones móviles:

- En primer lugar, como los RTTs pueden llegar a ser muy altos, aunque el ancho de banda sea reducido, el tamaño de la ventana de transmisión puede ser elevado. Por lo tanto, el descompresor ya habrá descartado gran cantidad de paquetes, antes de que se haya logrado actualizar el contexto, y deberá retransmitirse toda la ventana de transmisión y la eficiencia de la transmisión se verá afectada.
- En segundo lugar, cada vez que expire el RTO del emisor, éste, suponiendo que se trata de un problema de congestión, utilizará el mecanismo

de Slow Start y consecuentemente cerrará la ventana de transmisión. Teniendo en cuenta que, a diferencia de las redes fijas, el problema principal de las comunicaciones inalámbricas es la alta tasa de error y no la congestión, se estaría aplicando una política de recuperación inadecuada que muy lejos de ayudar, ralentizaría aún más la transmisión.

- Por último, debe destacarse también que el protocolo de compresión Van Jacobson no contempla la utilización de opciones TCP, como por ejemplo SACK [6] y Timestamp [8]. En comunicaciones móviles es recomendable la utilización de la opción SACK [13], por lo tanto parece adecuado que el protocolo de compresión contemple la utilización, como mínimo de esta opción, y que sea capaz de comprimirla de forma eficaz.

En [3, 1] se muestra que, en un entorno inalámbrico, el throughput obtenido en conexiones que utilizan compresión Van Jacobson es menor que en conexiones en las que no se aplica compresión y por lo tanto no se considera como una buena opción para comunicaciones móviles. [4] presenta posibles soluciones para adaptar el protocolo Van Jacobson a entornos con altas BER. Para ello propone la implementación de algoritmos que aceleren la recuperación en caso de pérdida de sincronismo entre el compresor y el descompresor; además expone la necesidad de adaptar el protocolo de Van Jacobson a comunicaciones IPv6. Ambos puntos se incluyen en el método de compresión IPHC [10].

3. COMPRESIÓN IPHC

IPHC [10] es un método de compresión pensado no sólo como posible solución a los inconvenientes indicados en el apartado anterior para comunicaciones inalámbricas, sino también como alternativa para la compresión de cabeceras que no estén basadas en TCP.⁴ Como método pensado para las comunicaciones móviles multiacceso, en el descompresor existirán campos de Identificador de conexión (CID) independientes para cada compresor.

Tal como puede observarse en la Figura 4, la estructura del paquete comprimido IPHC es prácticamente idéntica a la definida en [7]. El campo del identificador de conexión es obligatorio para evitar confusiones en el descompresor. El paquete comprimido contempla la utilización de opciones TCP,

⁴ El protocolo IPHC puede comprimir cabeceras UDP y cabeceras IP fragmentadas, por ejemplo. No contempla, en cambio, la compresión de cabeceras RTP, cosa que ha originado la aparición de otros protocolos de compresión [7] que no trataremos aquí

que se incluyen sin comprimir en el campo O de la cabecera comprimida. Por otra parte, también pueden transmitirse los posibles cambios en los bits reservados (TCP) o en el campo de tipo de servicio (IPv4/IPv6). En el campo Random, en el caso de comprimir subcabeceras, se incluyen los campos variables.

C: Identificador de conexión							
R	O	I	P	S	A	W	U
Checksum TCP (2 bytes)							
Campo Random							
R: Reserved (0-5) + TOS (6-7)							
U: Puntero Urgente							
W: Δ Window							
A: Δ ACK							
S: Δ Secuencia							
I: Identificador IP							
O : Opciones TCP							

Figura 4 Cabecera comprimida (IPHC)

En lo referente a la recuperación después de errores, se proponen dos formas de recuperar el contexto más rápidamente [4]:

- a) Algoritmo TWICE: En las comunicaciones «bulk data», por ejemplo el ftp, la variación de los campos de la cabecera comprimida (Delta Seq, Delta ACK) suele seguir un mismo patrón. En caso de error el descompresor podrá aplicar a la cabecera base hasta dos veces el incremento indicado en la cabecera del paquete recibido para intentar recuperar el contexto localmente. Si aun así la descompresión es incorrecta, se recurrirá a b) o se esperará a que el contexto se actualice tal como en [7].
- b) Petición de cabeceras: Si la descompresión del paquete recibido es incorrecta, el descompresor, puede solicitar al compresor que le envíe un paquete sin comprimir o comprimido de manera no diferencial para solventarlos.

Aunque estas soluciones representan una mejora del throughput respecto a la compresión Van Jacobson [4], no son totalmente eficientes en presencia de errores. El algoritmo TWICE presenta inconvenientes a la hora de comprimir los reconocimientos propios del algoritmo TCP (debido a la existencia de «delayed ACKs») y tampoco funcionará en servicios habituales como la descarga de páginas web o correo electrónico, debido a que en ellos es frecuente la variación del

tamaño de los segmentos de datos. Además, tampoco será posible la recuperación si se producen errores en más de un paquete consecutivo. Por otra parte, mediante la petición de cabecera, como mínimo se tardará un RTT en corregir el contexto en el descompresor, y por lo tanto aun se perderán muchos paquetes antes de que esto suceda.

Por lo tanto, parece necesario un método que evite las situaciones de inconsistencia⁵ en el descompresor. En la siguiente sección, se exponen diferentes propuestas ideadas con este fin.

4. OTRAS PROPUESTAS

Entre las posibles opciones de compresión que se centran en evitar la pérdida de contexto se encuentran [5, 2]. Estas propuestas basan su funcionamiento en variaciones del protocolo Van Jacobson; además el grupo de trabajo ROHC de la IETF en el área de transporte desarrolla [17] como una nueva alternativa a [7].

En [7] los campos de cada paquete se comprimen tomando como referencia el paquete anterior, por lo que si se pierde un paquete debe retransmitirse toda la ventana de transmisión.

En [5], en cambio, se mantiene un contexto al que se refieren todas las cabeceras comprimidas (cabecera base). De esta manera, aunque se pierda un paquete comprimido, no peligrará la consistencia del descompresor y sólo deberá retransmitirse el paquete perdido. Esta cabecera base sólo variará si se envía un paquete descomprimido; por ejemplo una retransmisión. La pérdida de uno de estos paquetes supondría la inconsistencia del descompresor. Por lo tanto, aunque la posibilidad de pérdida de contexto se ha minimizado, aun podría producirse.

El principal inconveniente de [5] es que si en la conexión no se producen errores durante un periodo de tiempo elevado, no se actualiza el contexto. Entonces se necesitará una cabecera cada vez mayor para poder codificar todos los campos y se aprovechará ancho de banda.

La eficiencia mejora en enlaces sin errores si se introduce un valor umbral para los campos de las cabeceras comprimidas [2]. Si cualquier campo del paquete comprimido sobrepasa este umbral, se deberá actualizar el contexto. Esta actualización se indicará activando el primer byte de la máscara del paquete comprimido [7] en el que se ha sobrepasado el valor umbral y se realizará a partir de la información del paquete anterior que debe poder almacenarse en compresor y descompresor.

⁵ Consideramos que hay inconsistencia en el descompresor cuando su contexto no coincide con el del compresor y por lo tanto, se descomprimen erróneamente los paquetes y, por lo tanto, se descartan.

Destacamos que, debido a que se utiliza el primer byte de la máscara, no es posible la adaptación inmediata de este método de compresión para poder utilizar el protocolo IPv6.

La propuesta actual del IETF, [17], pretende elaborar un método de compresión transparente, eficiente y robusto indicado para las comunicaciones móviles y, en general, para cualquier tipo de enlaces caracterizados por altos RTTs y altas tasas de pérdidas. Para ello, adopta un funcionamiento y una estructura de paquetes basados en el protocolo ROHC [12] definido para IP/UDP y RTP:

a) Los campos de cabecera se clasifican en tres tipos según la probabilidad de que varíen durante una conexión:

- Información estática: Presenta un valor constante (direcciones IP, puertos).
- Información dinámica: Campos que suelen ser constantes durante la conexión pero que pueden cambiar de valor circunstancialmente.
- Información comprimible: Campos cuyo valor puede variar en cada paquete y que pueden incluirse en la cabecera comprimida.

b) Tanto el compresor como el descompresor van pasando por distintos estados⁶ de funcionamiento durante la comunicación. La transición entre estados se producirá debido a información de realimentación en caso de enlaces bidireccionales, o a la expiración de temporizadores. Los estados del compresor indican la cantidad de información de la cabecera base de la que dispone el descompresor; sólo en el último estado el compresor puede enviar paquetes comprimidos. En el resto de estados, se enviará información de refresco del contexto.

Al tratarse de un método de compresión robusta, el descompresor trabajará mayoritariamente en el estado superior (que indica que el contexto está sincronizado); únicamente volverá a estados inferiores en caso de que se sobrepase un límite de paquetes erróneamente descomprimidos.

Un ejemplo de funcionamiento, suponiendo una estructura de 2 estados para el compresor (INICIO y COMPRESION) y 3 para el descompresor (NO CONTEXTO, CONTEXTO ESTÁTICO, CONTEXTO COMPLETO), sería el siguiente:

Primeramente, el compresor enviará los campos estáticos (direcciones IP, puertos TCP...) y esperará a recibir un reconocimiento del receptor. A partir de este reconocimiento, el compresor interpretará que la cabecera base del descompresor contiene la información estática

correcta (el estado del compresor será CONTEXTO ESTÁTICO) y enviará un nuevo paquete con información dinámica. Una vez que el descompresor haya confirmado que dispone de toda la información el contexto, tanto dinámica como estática, y que por lo tanto, se encuentra en el estado CONTEXTO COMPLETO, el compresor cambiará al estado COMPRESION y podrá enviar paquetes con los campos de cabecera comprimidos.

- c) La estructura de los paquetes utilizados para iniciar/refrescar el contexto es la misma que la definida en [12], por lo tanto se utilizarán paquetes especiales dependiendo de si se trata de una actualización del contexto estático o dinámico.
- d) La estructura del paquete comprimido está aún por definir, pero especialmente de cara a la aplicación del protocolo a enlaces móviles debe contemplar la compresión de:
- Las opciones TPC: SACK y Timestamp
 - Los campos especificados en IPHC
 - El protocolo IPv6 (incluyendo opciones y extensiones)
 - Cabeceras que contengan subcabeceras IPSEC

En la Tabla 1 se muestra que la eficiencia conseguida en cuanto al tamaño de la cabecera ROHC-TCP es comparable al resto de propuestas de compresión.

Método de compresión	Sin compresión	Van Jacobson, [5],[2]	IPHC	ROHC
Tamaño mínimo de la cabecera comprimida	40 bytes (TCP/IPv4) 60 bytes (TCP/IPv6)	3 – 5 bytes (TCP/IPv4)	4 – 5 bytes (TCP/IPv4/6)	< 4 bytes ⁷ (TCP/IPv4/6)

Tabla 1. Comparación de los tamaños de cabecera para los distintos métodos de compresión

Para evitar la pérdida de contexto y proporcionar un método de compresión robusto, se propone que el compresor guarde una ventana de m posibles contextos que podría haber en el descompresor. Se comprimirán los paquetes de tal manera que el paquete resultante pueda descomprimirse siempre en el destino sea cual sea el contexto del descompresor.

⁷ El formato del paquete comprimido no está definido aun. Según [19], el tamaño de la cabecera ROHC-TCP debe ser menor o igual a la de IPHC



Cada vez que se reciba comprobación de que un paquete ha sido descomprimido en el destino, ya sea por realimentación o por el mecanismo implícito de TCP, los contextos asociados a ese paquete y a los anteriores serán borrados del compresor.

Esta técnica, aunque robusta, provoca un aumento del tamaño indeseado del paquete comprimido. Para solventar este efecto, se utilizará la codificación «Window LSB» para los campos de cabecera: se parte del hecho de que los campos de cabeceras variarán entre unos límites establecidos de tal manera que sólo es necesario enviar los k últimos bits significativos para poder recuperar el valor original. En el caso de comunicaciones en las que el tamaño del campo de datos se mantiene constante ($SEQ[ACK] = m*Payload + n$); se minimizará el tamaño de cabecera enviando solamente m y aumentando así la eficiencia de la compresión.

Por último, cabe destacar que la propuesta [17] contempla un método de optimización de conexiones breves. Debido a que es necesario un tiempo considerable para inicializar el contexto en el descompresor, la mayoría de los paquetes enviados en conexiones de corta duración no estarán comprimidos y por lo tanto no se aprovecharán las ventajas del método de compresión de cabeceras. La descarga http de una página web, por ejemplo, se lleva a cabo mediante la apertura de distintas conexiones por las que se transmite una pequeña cantidad de datos. Estas conexiones difieren en el puerto TCP de origen pero coinciden en el valor del resto de los campos estáticos de cabecera (direcciones IP, puerto de destino...). [17] permite replicar en el descompresor [16] los valores de los campos comunes del contexto de este tipo de conexiones, de manera que sólo sea necesario inicializar los campos diferentes. De esta manera, puede minimizarse el número de paquetes no comprimidos que deben enviarse para iniciar el contexto y se mejora así la eficiencia de las transmisiones.

5. CONCLUSIONES

En los apartados anteriores se han analizado los inconvenientes de los protocolos de compresión existentes. Aunque existen diferentes propuestas para mejorar la eficiencia obtenida con el método de compresión Van Jacobson en enlaces de baja velocidad con altas BER y RTTs; en la práctica, si se utiliza compresión, el algoritmo implementado sigue siendo [7]. Este hecho, probablemente, se debe a que ninguna de las propuestas puede considerarse como una solución suficientemente global a los problemas de este tipo de enlaces.

El principal problema que presentan [7] y IPHC es que el tiempo de recuperación en el caso de pérdida de contexto es demasiado grande y no puede evitarse,

se, mientras tanto, el descarte de paquetes en el descompresor. Aunque las propuestas expuestas en la sección anterior no solucionan que el contexto pueda recuperarse con más rapidez en caso de pérdidas de sincronismo, minimizan el riesgo de que esto suceda. Sin embargo, la robustez conseguida con [5, 2] presenta otros inconvenientes. En [5] la eficiencia de la transmisión se ve afectada en términos de ancho de banda debido al incremento del tamaño de las cabeceras comprimidas. Por otra parte, se considera necesario que un método de compresión orientado a comunicaciones móviles pueda comprimir cabeceras IPv6; sin embargo [2] no contempla esta opción por defecto.

La propuesta [17] se presenta como una solución eficiente, tanto en robustez (mediante la utilización de una ventana de contextos) como en términos de ancho de banda, ya que se requiere un tamaño de cabeceras igual o menor que el definido en IPHC. Por otra parte, aunque no se contempla la compresión de todas las opciones TCP, se especifica que las opciones SACK y Timestamp, recomendadas en [13], sí que deben poder comprimirse; por lo tanto, considerando que el resto de las opciones no se utilizarán con tanta frecuencia, se concluye que la propuesta [17] es suficientemente adecuada en este aspecto.

Por último, aunque la propuesta nos parece acertada y, concretamente, una buena solución para las comunicaciones móviles, como nota negativa, destacamos que el hecho de utilizar este método basado en estados podría implicar un coste de procesado destacable, tanto en recursos como en tiempo. Los terminales móviles presentan limitaciones de proceso, por lo que la realización de un número elevado de instrucciones a la hora de comprimir/descomprimir supondría un aumento no deseable del retardo extremo-extremo. Por lo tanto, es un aspecto que debería tenerse en cuenta a la hora de implementar la solución [17].

6. AGRADECIMIENTOS

Parte de este trabajo ha estado financiado por el proyecto TIC2000-1041-C03-01.

7. REFERENCIAS

- [1] A. Calveras. J. Paradells. "Performance Optimisation Evaluation of TCP/IP Over Wireless Networks". IEEE International Performance, Computing and Communications Conference. Phoenix/Tempe (Arizona). USA. Febrero 1998.
- [2] Calveras. M. Arnau. J. Paradells. "A controlled Overhead for TCP/IP Header Compression Algorithm

- over Wireless Links". The 11th International Conference on Wireless Communications (Wireless'99). Calgary. Alberta. Canada. Julio 1999.
- [3] A. Calveras. R. Rebollo. J. Paradells. "La Arquitectura TCP/IP en Entornos Móviles". Telecom I+D '97. Madrid. Octubre 1997
- [4] Degermark, M et al, "Low Loss TCP/IP Header Compression for Wireless Networks". Mobicom96
- [5] Perkins, Stephen J., Mutka, Matt W., "Dependency Removal for Transport Protocol Header Compression over Noisy Channels", IEEE, 1997
- [6] M. Mathis et al, "TCP Selective Acknowledgement options", Octubre 1996, RCF2018
- [7] Jacobson, V., "Compressing TCP/IP Headers for Low-Speed Serial Links", Febrero 1990. RFC1144
- [8] V. Jacobson et al, "TCP Extensions for High Performance", Mayo 1992 RFC1323
- [9] S. Deering, R. Hinden , "Internet Protocol, Version 6 (IPv6). Specification", Diciembre 1998 RFC2460
- [10] Degermark, M., Nordgren, B. and S. Pink, "IP Header Compression", Febrero 1999. RFC2507
- [11] Casner, S. y V. Jacobson, "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links", Febrero 1999. RFC2508
- [12] Carsten Bormann, et. al., "Robust Header Compression (ROHC)", Julio 2001. RFC3095
- [13] H.Inamura y G. Montenegro et al, "TCPo v e r second (2.5G) and Third (3G) Generation Wireless Networks", Febrero 2002. RFC3481.
- [14] Joan Postel, "Transport Control Protocol", Septiembre 1981 RFC793
- [15] RIU253 IST-2001-36510 D3.1 "Protocol options and their relation with the wireless packet protocols" <http://www-riu253.upc.es/>
- [16] Pelletier, G., "Robust Header Compression (ROHC): Context replication for ROHC profiles", Internet Draft (work in progress), <draft-pelletier-rohc-context-replication-00.txt>, Mayo 2003
- [17] Ghyslain Pelletier et al. "RObust Header Compression (ROHC):TCP/IP Profile (ROHC-TCP)", Internet Draft (work in progress), <draft-ietf-rohc-tcp-04.txt>, Marzo 2003
- [18] <http://www.ietf.org/html/rohc-charter.html>
- [19] M.West y S. McLann "TCP/IP Field behaviour", Internet Draft (work in progress), <draft-ietf-rohc-tcp-field-behaviour-02.txt>, Marzo 2003
- [20] Lars-Erik Jonsson, "Requirements on ROHC TCP/IP Header Compression", Internet Draft (work in progress), <draft-ietf-rohc-tcp-requirements-06.txt>, Junio 2003

8. ACRÓNIMOS

BER	Bit Error Rate
GPRS	General Packet Radio System
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPHC	IP Header Compression
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
MMS	Multimedia Messaging Service
RFC	Request For Comment
RLC	Radio Link Control
ROHC	Robust Header Compression
RTO	Retransmission TimeOut
RTP	Real Time Protocol
RTT	Round Trip Time
SACK	Selective ACKnowledgement
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UMTS	Universal Mobile
	Telecommunications System

AUTORES



María Luisa Catalán Cid. Ingeniera de Telecomunicaciones por la ETSETB desde el año 2003. Realizó el proyecto sobre la aplicación de mecanismos de mejora del comportamiento del protocolo TCP en entornos GPRS. Actualmente realiza el doctorado en el Departamento de Telemática de la UPC y participa en el proyecto de promoción tecnológica de UMTS Servicios UMTS y técnicas avanzadas de localización.



Anna Calveras-Augé. Dr. Ingeniero de Telecomunicación por la Universidad Politécnica de Catalunya (UPC). Actualmente es profesor Titular de Universidad en dicha universidad, en el Departamento de Ingeniería Telemática. Es experta en optimización de los protocolos de Internet en entornos de red heterogéneos, especialmente el protocolo TCP. Ha estado involucrada en diversos proyectos. Actualmente está trabajando en el proyecto europeo ELIN (The Electronic Newspaper Initiative), el proyecto europeo RIU253 (Recommendations for Internet Usage over 2.5 and 3G networks), y el proyecto español Estudio de las Alternativas para ofrecer servicios multimedia con calidad de servicio extremo a extremo en una red IP integrada con servicios móviles (CICYT TIC2000-1041-C03-01). Ha publicado diversos trabajos en conferencias nacionales e internacionales.





MODELADO SEMÁNTICO Y UBICUO DEL COMERCIO ELECTRÓNICO MEDIANTE WEB SERVICES

Juan Carlos Peláez López, Jorge Mata Díaz

Departamento de Ingeniería Telemática. Universitat Politècnica de Catalunya
Jorge Girona 1 y 3. Campus Nord, Mód C3, UPC. 08034 Barcelona
Teléfono: 934 01 60 28 Fax: 934 01 59 81

E-mail: {jpelaez, jmata}@entel.upc.es

Abstract.- La demanda de interoperabilidad semántica y ubicuidad tecnológica en aplicaciones de comercio electrónico tiene un crecimiento exponencial. Añadir propiedades inteligentes a la Web actual, ya es una realidad a través del desarrollo e implementación de Web Services basados en estándares de fácil gestión y cada vez más estables, cumpliendo con los requerimientos básicos de Calidad de Servicio (QoS). Este trabajo presenta un análisis de la tecnología de los Web Services con el principal objetivo de aportar valor añadido a aplicaciones de Comercio Electrónico; así como, exponer una visión general del estado actual de los Web Services y sus contribuciones a las necesidades de las empresas en general.

1. INTRODUCCIÓN

Actualmente, una de las áreas de mayor crecimiento e interés dentro de las tecnologías de la información es la relacionada con los Web Services. El respaldo que ha tenido y tiene esta tecnología por parte de las principales empresas tecnológicas [1], tales como: IBM, Microsoft, Bea System, Sun, HP, Verisign, entre otras, fortalece la definición continua de estándares para Web Services, así como, su adaptación dentro de las herramientas de desarrollo comercial propuestas.

Por otra parte, el desarrollo del Comercio Electrónico se ha visto limitado, en términos técnicos, por la carencia de una tecnología que aporte una verdadera interoperabilidad a transacciones comerciales entre empresas sobre redes de comunicación, concretamente Internet.

El presente trabajo tiene como finalidad presentar un análisis sobre la aplicación de los Web Services en el campo del Comercio Electrónico, vinculando su aportación dentro de características semánticas [5] y ubicuas que fortalezcan conceptos sobre interoperabilidad e integración, llevadas a cabo por empresas, iniciadas en su entorno interno hasta lograr su extrapolación al entorno externo; es decir, implementar aplicaciones de Comercio Electrónico fiables, seguras, escalables y, considerando también el aspecto económico, aplicaciones rentables.

En el siguiente apartado se presentan conceptos relacionados a los Web Services; así como, se analiza su estado actual, describiendo los principales estándares propuestos hasta la fecha. En la tercera sección se tratan los aspectos claves de los Web Services con los que se intenta alcanzar el fortalecimiento del Comercio Electrónico respecto a su desarrollo e implantación. Los temas relacionados a Calidad de Servicio (QoS) que garanticen Servicios Web fiables, se abordan en el cuarto apartado. Posteriormente, en la quinta sección se presentan diferentes áreas de aplicación de los Web Services, adicionales al eCommerce. En la sexta sección, se presenta un análisis de la relación de los Web Services con otras tecnologías. Los temas abiertos, desarrollos en curso, nuevas definiciones de estándares, así como, el futuro de los Web Services se trata en la séptima sección. Finalmente, en la octava sección presentamos nuestras conclusiones.

2. WEB SERVICES Y ESTÁNDARES RELEVANTES

Tratar de conceptualizar la tecnología de los Web Services implica dividir su definición en dos partes:

1) Desde el punto de vista tecnológico, es decir, su contribución al avance y consolidación dentro de las tecnologías de la información, se considera a los Web Services como un nuevo tipo de aplicaciones Web. Aplicaciones de carácter modular, que cumplen tareas específicas y a las que se pueden acceder a través de protocolos abiertos, tales como: HTTP, TCP/IP; es decir, estándares Internet ampliamente aceptados. Con lo cual, los Web Services logran alcanzar la interoperabilidad en entornos heterogéneos, sin preocuparse por la interacción sobre diferentes sistemas operativos, lenguajes de programación, modelos de objetos y datos, y redes de transporte utilizadas.

2) Desde el punto de vista de las empresas, los Web Services son considerados un paso más en el mundo de la integración de aplicaciones y una puerta hacia nuevos modelos de negocio. En definitiva, una interconexión sin precedentes en la empresa, que involucra reducción del time-to-market, mejora del tiempo del retorno de la

inversión (ROI, Return on Investment), y reutilización de aplicaciones heredadas (existentes), todo ello procesado de forma automática.

Considerando los enunciados anteriores, no se puede hablar de una revolución tecnológica sino más bien de una evolución dentro de la integración de aplicaciones.

En el caso de los servicios telemáticos, la evolución va desde los entornos centralizados hasta los distribuidos; en lo que respecta a las aplicaciones, en un principio se manejaban formatos de datos sencillos (texto) hasta llegar a los cada vez más demandados formatos de información multimedia (audio/video). Además, se debe considerar la evolución de las redes y sus componentes que se presentaban inicialmente como islas propietarias hasta lograr la integración de redes y consolidar la Internet, protocolos abiertos de comunicación, amplia escalabilidad, entorno distribuido y acceso universal.

El desarrollo de un entorno distribuido [2] comienza por los años 80 dentro del Entorno de Computación Distribuida, DCE (Distributed Computing Environment) [28], y la estandarización de la Llamada a Procedimientos Remotos, RPC (Remote Procedure Call), fallido intento debido al poco soporte de la industria y a la carencia de control de mensajes. Posteriormente, las Aplicaciones Orientadas a Objetos adquirieron protagonismo debido a la reusabilidad de código y al sencillo mantenimiento que conllevan. El avance de la Computación Distribuida continuó con la aparición de arquitecturas como CORBA (Common Object Request Broker Architecture) [27] y DCOM (Distributed Component Object Model) [29], que no han logrado convencer debido a la falta de escalabilidad, por la complejidad de implementación y por la dependencia en componentes propios a su infraestructura (ej. brokers).

La consolidación del Internet llevó a los entes reguladores a focalizar su atención en el desarrollo de estándares que exploten la sencillez y la facilidad de acceso a la Web. Así nacieron, definiciones de nuevos lenguajes que minimizan la utilidad del hasta ahora predominante HTML (Hyper Text Markup Language); en este caso, el XML (eXtensible Markup Language) [19], que es un metalenguaje universal que facilita la comunicación entre diferentes sistemas de computación. Su principal ventaja, es que a diferencia del HTML, el cual pone más énfasis en la presentación de la información, el XML separa la presentación de los contenidos, característica por la cual su adopción se ha extendido a entornos diversos como las comunicaciones móviles, Web inteligente, Internet dinámica, etc.

Por otro lado, el SOAP (Simple Object Access Protocol) [20] es un protocolo ligero de comunicación basado en XML utilizado para el intercambio de información a través

de Internet. Utiliza XML para describir las propiedades, métodos y eventos que admite un objeto; así como, la forma en que las aplicaciones pueden invocarlos.

Dentro de los protocolos de descripción, se definió el WSDL (Web Services Description Language) [21], que es un lenguaje que describe interfaces de Web Services y que también se encarga de especificar la interacción entre ellos.

Finalmente, con respecto a la estandarización del registro de servicios, se definió el UDDI (Universal Definition, Discovery and Integration) [22], directorio que intenta ser utilizado para publicar y descubrir Web Services en Internet.

Los esfuerzos de definición de estándares han sido dirigidos por tres organismos oficiales: W3C (World Wide Web Consortium) [16], OASIS (Organization for the Advancement of Structured Information Standards) [26], y finalmente, el IEFT (Internet Engineering Task Force) [30]. En abril de 2001, en el WSWS (W3C Workshop on Web Services) [2], planearon el futuro de los Web Services, bajo la idea de agrupar todos los estándares bajo el W3C. Como resultado se fortaleció la adopción de los protocolos SOAP, WSDL y UDDI, cuyos roles y funciones se presentan en la figura 1, y en la figura 2 se presenta de forma simplificada las transacciones en una comunicación de Web Services. En febrero de 2002 se crea el WS-I (Web Services Interoperability Organization) [31] para la promoción de los estándares de Web Services (trabajo conjunto de W3C y OASIS), cuya idea base es no crear nuevos estándares, sino en su lugar, ensamblar los existentes nacidos de dichos organismos de estandarización.

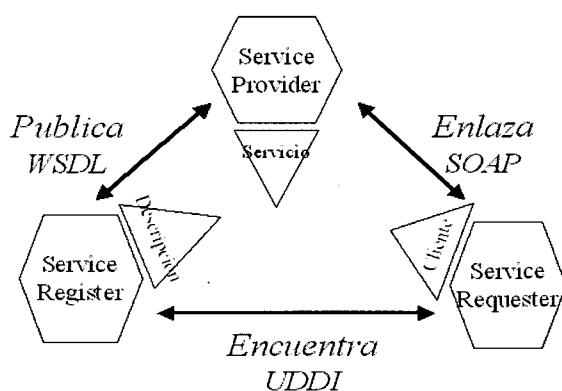


Figura 1. Roles y funciones de los Web Services

Con respecto al modelo en que se basan los Web Services se debe destacar al SOA (Services-Oriented Architecture) [24], que es un intento de la ingeniería de software para fomentar la reutilización de componentes de software.



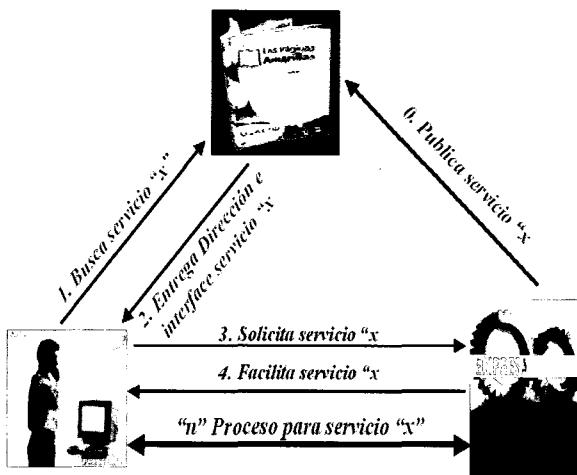


Figura 2. Transacciones de Web Services

La tecnología de los Web Services está también subordinada a la estrategia empresarial [13] que incluye una reducción del tiempo de desarrollo de aplicaciones, así como, del coste de implementación; por lo cual, es necesario la definición de estándares que soporten soluciones empresariales [7] para integrar aplicaciones, tales como: EAI (Enterprise Application Integration) [14], SCM (Supply Chain Management) y CRM (Customer Relationship Management). Entre dichos estándares, que soportan la gestión del proceso de negocio se encuentran: WSCI (Web Services Choreography Interface), WSIF (Web Services Invocation Framework), WSFL (Web Services Flow Language), WORF (Web Services Object Runtime Framework), BPEL4WS (Business Process Execution Language for Web Services), ebXML (Electronic Business XML Initiative), XLANG (Web Services for Business Process Design) [15] [16] [17] [18].

Además, se debe destacar la iniciativa UBL (Universal Business Language) [5] que define documentos de negocio basados en XML con el fin de reemplazar al EDI (Electrónica Data Interchange) [23]. El UBL tiene como objetivo desarrollar una librería estándar de documentos de negocio XML (órdenes de compra, factura, etc.) basándose en la modificación de librerías ya existentes y la aportación de nuevas definiciones. Es de resaltar que UBL intenta ser un estándar internacional para el Comercio Electrónico, de libre acceso y disponibilidad para todos sin necesidad de una licencia.

3. APORTE DE LOS WEB SERVICES AL COMERCIO ELECTRÓNICO

La red Internet para el mundo empresarial ha sido la fuente de múltiples oportunidades de negocio, su amplio despliegue, acceso universal y protocolos estables de comunicación han sido las bases para su hegemonía dentro de las redes telemáticas de cobertura global.

Sin embargo, a pesar de su gran aceptación, Internet presenta ciertas carencias que limitan las estrategias de las empresas para llevar a cabo proyectos de Comercio Electrónico sobre Internet.

Así tenemos, que la Web puede mejorarse en dos aspectos:

1) Procesos manuales.

Debido a los procesos manuales, es decir, la interacción humana, se introduce retardo y errores que de forma automática no se tendría. Por lo cual, no se logra el objetivo de disponer de una Web «sin navegador» (Browserless Web) [14], también conocida como Internet Dinámica o Comercio Electrónico Dinámico.

2) Integración multifuente.

Recuperación de diferentes fuentes (sitios Web) de forma automática y dinámica en tiempo real.

Los Web Services aportan soluciones a dichas carencias, siendo una alternativa fácil y económica para poner en marcha un proyecto de Comercio Electrónico.

Al tener como base el lenguaje de comunicación XML, los Web Services contribuyen con dos aspectos claves para el Comercio Electrónico:

1) Interoperabilidad Semántica

2) Ubicuidad Tecnológica

La Interoperabilidad Semántica [5] cubre el problema de reunir sistemas de información heterogéneos y distribuidos, considerando la próxima generación de la Web: Web Semántica. Es decir, considera no sólo la estructura de la información, sino también, la representación del conocimiento, el significado del concepto de un dato. Bajo este contexto se desarrolla el trabajo del W3C Semantic Activity [32], cuyo objetivo de representar de forma abstracta los datos en la Web, lo logra usando la definición RDF (Resource Description Framework) [33], una estructura de metadatos que provee interoperabilidad entre aplicaciones. La RDF se encarga de describir la información por atributos, creando perfiles, a partir de los cuales se facilita la recuperación de información en Internet.

Respecto al tema de la ubicuidad tecnológica [3] [9], se trata de proveer un acceso «everytime and everywhere» a los servicios que forman parte de la aplicación de Comercio Electrónico, basando sus propiedades de accesibilidad en estándares como XML, SOAP, WSDL y UDDI, definiciones abiertas que constituyen el núcleo de los Web Services. Además, se puede también expresar la ubicuidad tecnológica en términos del dispositivo electrónico (PDA, PC, móvil, etc.) utilizado para interactuar con los servicios de la Web.

Finalmente, considerando que el EDI fue el proyecto pionero en temas de Comercio Electrónico, son deducibles las ventajas que los Web Services, basados en los estándares

mencionados en el apartado anterior, aportarían respecto al alto coste de implementación del EDI, el cual no es viable para pequeñas y medianas empresas, llevando a la práctica un cambio en la forma en que diseñamos y construimos aplicaciones de Comercio Electrónico.

4. CALIDAD DE SERVICIO EN WEB SERVICES

Las garantías de Calidad de Servicio (QoS, Quality of Service) [4] de los Web Services están expresadas en términos de seguridad, fiabilidad y gestión.

La autenticación a través de certificados digitales, y la confidencialidad basada en métodos de encriptación y firma digital, son las bases para garantizar seguridad en los Web Services. Al mismo tiempo, se requiere una completa definición de políticas de autenticación y control de acceso, tanto en los Web Services, como en los roles (ver figura 1) que participan en la comunicación de dichos servicios.

La entrega garantizada, el no repudio y la entrega única son parámetros por los cuales se mide la fiabilidad de los Web Services.

Los procesos de gestión de los servicios incluyen tareas de exploración, monitorización, identificación y auditoria. Este punto es una de las bases para la adaptación de Web Services en aplicaciones de Comercio Electrónico, al facilitar técnicas de fácil gestión extremo a extremo entre el proveedor y el receptor del servicio.

Bajo los términos mencionados anteriormente, WSPL (Web Services Policy Language) [26], define los casos en que se debe negociar una variedad de políticas dentro de la arquitectura de los Web Services, como son las siguientes:

- o política de seguridad criptográfica
- o política de autenticación
- o política de autorización
- o política de privacidad
- o política de mensajería fiable
- o política de transmisión

Dichas políticas se negocian bajo el fundamento del XACML (eXtensible Access Control Markup Language) [26], el cual define un vocabulario específico sobre temas, derechos, objetos y condiciones, que constituye información esencial en temas del e-commerce actual.

Por otro lado, la segmentación de la información proporcionada por la RDF, contribuye a la disponibilidad de información común entre varios sectores empresariales, a través de los cuales, es posible integrar información básica (core-data), y además ampliar su contenido, con la

definición de extensiones pertenecientes a un sector en particular. El soporte de diferentes contextos de negocio, mejora la gestión de los servicios e incrementa su utilidad y aplicación en entornos reales.

Actualmente, la seguridad y fiabilidad son los ejes de desarrollo de estándares y políticas para lograr un mayor despliegue y aceptación de los Web Services; así como, para conseguir soluciones empresariales.

5. ÁREAS DE APLICACIÓN DE LOS WEB SERVICES

Sin duda alguna, el Comercio Electrónico, en todas sus variantes B2B, B2C, B2G, P2P, etc., se ve fortalecido con el surgimiento de los Web Services. Sin embargo, este valor añadido se extiende a otras aplicaciones, tales como: integración interna de sistemas y aplicaciones departamentales de una empresa, servicios de directorio, políticas de encaminamiento, sistemas de código abierto, reutilización de sistemas ya existentes, adaptabilidad a nuevas necesidades de negocio sin entorpecer las funcionalidades preexistentes, modelo de negocio de software «pay-as-you-go» (pago asociado al uso de un servicio), entornos móviles (WML, Wireless Markup Language), servidores de aplicaciones, software colaborativo, búsqueda de información distribuida y computación bajo demanda.

En definitiva, todas aquellas aplicaciones de negocio con altas necesidades de interactividad, integración, acoplamiento, combinación, control de granularidad fina, sensibles al tiempo, de alta complejidad en desarrollo e implantación, suministros de servicios de infraestructura para otras aplicaciones Web o Web Services, alta escalabilidad, y requerimientos de balanceo de carga.

6. TECNOLOGÍAS PARALELAS A LOS SERVICIOS WEB

El grado de adaptabilidad de los Web Services a entornos heterogéneos, se extiende también a su compatibilidad y relación que guarda con otras tecnologías. En este apartado, analizaremos brevemente 5 tecnologías con las que los Web Services tienen relación.

1. OSS (Open Source Software)

La comunidad OSS [34] ve a los Web Services como base para la próxima generación Web para intereses comerciales; es decir, que los Web Services se enmarcarían dentro de los intereses de terceros. Para evitar ello, a pesar de las herramientas de desarrollo existentes (ej. .NET de Microsoft, SUN ONE de SUN, Websphere de IBM) que tienen un perfil propietario,



podemos aprovechar su madurez en la integración de los estándares SOAP, UDDI y WSDL. Además, se puede llevar a cabo de forma particular el diseño y desarrollo de Web Services basándonos en las especificaciones de los protocolos antes mencionados, sin necesidad de recurrir a las herramientas de desarrollo comerciales, aunque ello conlleve una mayor especialización de parte del desarrollador de Web Services.

2. Web Inteligente (Web Semántica)

La Web Inteligente [32] abarca al conjunto de tecnologías que mejoran la Web. Un trabajo que se está llevando a cabo y que es una de las bases de este artículo, es la Web Semántica (SemWeb, Semantic Web), la cual intenta aportar a los Web Services las ventajas de un sistema integrado de gestión de conocimientos.

Entre los avances que se han alcanzado entre Web Services y SemWeb se destacan:

- Asignación de URIs (Uniform Resource Identifier)
- Gestión de metadatos de Web Services estructurados
- Uso de RDF sobre protocolos SOAP (RDF es la clave de la tecnología SemWeb).

3. MDA (Model Driven Architecture)

OMG's MDA [27] trata de acercarse al desarrollo a gran escala de software. En esta área se ha efectuado la implementación de WSDL y SOAP combinados para modelar MDA.

4. Grid

Grid [25] es una tecnología que tiene como objetivo la creación de un supercomputador virtual donde los ciclos de procesamiento y otros recursos puedan ser compartidos para llevar a cabo diferentes tareas.

Globus Project (Grupo de desarrollo abierto para computación Grid) [25] aporta mecanismos para ofrecer Grid a través de Web Services. Por ejemplo, la especificación OGSA (Open Grid Services Architecture) [35] ha sido diseñada para facilitar a los Web Services el acceso a los recursos Grid. Su objetivo final es el eBusiness-on-demand.

La integración de OGSA y Web Services dará soporte a temas como seguridad, flujos de trabajo (trabajo colaborativo), base de datos, sistemas de archivos, directorios y redes.

5. Computación Alternativa - Peer-to-Peer Computing

Desde proyectos como Jini y Jxta (P2P, Peer to Peer, always.on connectivity) [36], hasta aplicaciones WAP, la computación alternativa

ofrece sistemas estándares para la integración y comunicación en dispositivos como impresoras de alta tecnología así como PDAs. Web Services aportaría los beneficios de la interoperabilidad y convergencia tecnológica, reduciendo los costes y aumentando las opciones de utilidad («utility computing»). Los sistemas peer-to-peer son actualmente uno de los campos de mayor investigación, el alto grado de aceptación por parte de usuarios a aplicaciones de intercambio de ficheros como Napster, Gnutella, Kazaa, nos hacen apostar por una convergencia de tecnologías: Web Services, Sistemas peer-to-peer y Computación Gris.

7. TEMAS ABIERTOS

El esfuerzo que se está llevando a cabo dentro del área de los Web Services deja todavía aspectos no resueltos completamente. A continuación se resumen los más relevantes:

1. Seguridad

Se trata de garantizar Web Services fiables basados en una seguridad sobre estándares XML. Dentro de los estándares propuestos se detallan los siguientes:

- a) XML Signature: promovido por el W3C y el IETF, su objetivo es garantizar la integridad de datos y cubrir características de autenticación (tanto del mensaje como del firmante).
- b) W3C's XML Encryption: proporciona confidencialidad de datos usando técnicas de encriptación.
- c) Ws-security: promovido por OASIS, constituye un mecanismo de seguridad incluido dentro de mensajes SOAP, soporta característica de integridad, confidencialidad y autenticación del mensaje.
- d) Definición de SAML (Security Assertion Markup Language), promovido por OASIS, provee un concepto a las aplicaciones de socios comerciales para compartir autenticación de usuario e información de autorización. Aporta el soporte de SSO (Single Sing On), cuyo estado actual es el basado en el uso de cookies en comunicaciones http, pero que con el futuro SAML los datos de las cookies irán dentro del documento XML, alcanzando la interoperabilidad de SSO.
- e) Políticas de Control de Acceso definidos en WSPL y requeridas en XACML. Políticas de autenticación y control de acceso bajo un vocabulario específico para: temas, derechos, objetos y condiciones, que constituyen la información esencial en temas de Comercio Electrónico actual para políticas de autorización.

2. Soporte Multimedia

Se ha propuesto el SMIL [37] (Synchronized Multimedia Integration Language), un lenguaje basado en XML para describir el diseño (formato) y sincronización de aplicaciones multimedia. Constituye un sofisticado desarrollo multimedia para: 1) añadir comentarios de audio a imágenes y texto, 2) animar presentaciones de diapositivas que cambian dinámicamente, 3) añadir controles on-screen que permitan a los usuario parar e iniciar una presentación, y 4) crear productos que integren el audio, video, animación y texto.

En conclusión, con el SMIL se busca un desarrollo de contenido dinámico basado en el ensamblaje de componentes individuales multimedia.

3. Especificaciones que den soporte a temas de Calidad de Servicio (QoS).

- a) ws-routing: especificación de un protocolo de enrutamiento asincrónico para mensajes SOAP sobre una variedad de protocolos de transporte como el http, TCP y UDP.
- b) ws-coordination: especificación que abarca el control del flujo de trabajo (workflow)
- c) ws-transaccion: especificación para gestionar las transacciones involucradas dentro del proceso de negocio.
- d) ws-reliableMessaging: aporta un protocolo capaz de detectar los mensajes no recibidos y los duplicados, así como procesar los recibidos en el orden que fueron enviados.
- e) ws-addressing: identifica e intercambia referencias en los puntos finales de los servicios Web; además, define un conjunto de cabeceras de información de mensajes

4. Especificaciones que definan los temas relacionados a la economía: modelo de coste de servicios, niveles de servicio, niveles de usuario, etc.

8. CONCLUSIONES

Tal como se aprecia, el entorno de desarrollo de los Web Services está aún fortaleciéndose, sin embargo, su crecimiento exponencial debido al gran apoyo de la industria del software así como del sector empresarial, pronostica un excelente futuro para su despliegue y consolidación.

Es comprensible que los Servicios Web sean actualmente más populares en entornos Intranet que Internet, debido al riesgo que conlleva apostar

por una nueva tecnología que no es aún lo suficientemente madura. Posterior al proceso de integración interna (Intranet) de aplicaciones de la empresa, se continuará con la integración «semi-externa» que involucrará la integración con la cadena de valor del negocio (socios, distribuidores, clientes). Finalmente, se apostará por una integración externa, es decir, lograr un alcance global de Servicios Web para beneficio de la empresa.

La existencia de planes pilotos, basados en los estándares XML, SOAP, WSDL y UDDI, aportan la experiencia y retroalimentación tanto del desarrollo como implantación de esta nueva tecnología, donde se abarcan temas de Calidad en Servicio (QoS) de gran importancia expresados en términos de gestión, seguridad y fiabilidad.

La relación entre Web Services, Peer-to-Peer (P2P) y tecnología Grid aportarán mayor rentabilidad a proyectos de Comercio Electrónico, pues además de asegurar aplicaciones semánticas y ubicuas, proveerán soluciones escalables que requieran gran capacidad de procesamiento computacional, así como, recursos de almacenamiento.

Finalmente, la utilización dinámica de recursos, su publicación, descubrimiento, acceso universal, nos asegura el alcance de la era de la Computación bajo demanda, en entornos heterogéneos y de forma inteligente, garantizando una gestión dinámica y flexible del uso de los Web Services.

REFERENCIAS

- [1] C.Mohan, IBM Almaden Research Center. «Dynamic e-Business: Trends in Web Services». (http://www.almaden.ibm.com/u/mohan/WebServices_TES2002.pdf).
- [2] Uche Ogbuji. «The Past, Present and Future of Web Services» (<http://www.xml.com/pub/r/1377>).
- [3] A White Paper from Flamenco Networks. «Web Services and the need for Web Service Networks». (https://www.flamenconetworks.net/Pages/Help/Web_Services_White_Paper.pdf).
- [4] V. Cardellini, E. Casalicchio, University of Rome Tor Vergata; M. Colajanni, University of Modena. «A performance Study of Distributed architectures for the Quality of Web Services». (January 2001 IEEE).
- [5] David Burdett, Commerce One. «Avoiding EDI's Mistakes With Web Services. Semantic Interoperability» (EAI Journal, December 2002).
- [6] B. Lublinsky and M. Farrell Jr. «10 Misconceptions About Web Services». (EAI Journal, February 2003).



- [7] Andre den Haan. «Web Services: A Sea Change Taking Place». (EAJ Journal, February 2003).
- [8] Romin Irani, InSyns Information Systems, Inc. California. «Versioning of Web Services. Solving the Problem of Maintenance».(<http://www.webservicesarchitect.com/content/articles/irani04print.asp>).
- [9] Joe McKendrick. «Web Services at the FrontEnd».(<http://www.webservices.org/index.php/article/articlestatic/845/1/2/>).
- [10] Dr. Bob Sutor, Director of Web Services Strategy, IBM. «My wish list for Web Services in 2003». (<http://zdnet.com.com/2102-1107-980511.html>, January 2003).
- [11] F. Pires, M. Benevides and M. Mattoso, computer Science Department, COPPE – Federal university of Rio de Janeiro. «Building Reliable Web Services Compositions»(<http://www.jeckle.de/ws-rsd/program/pires.pdf>).
- [12] M. Pierce and G. Fox, Community Grid Labs, Indiana University; C. Youn, Syracuse University; S. Mock and K. Mueller, University of California at San Diego; O. Baloy, Florida State University. «Interoperable Web Services for Computational Portals». (2002 IEEE).
- [13] F. Leymann, D. Roller and M. Schmidt. «Web services and Business Process Management». (IBM System Journal, vol. 41, No 2, 2002).
- [14] M. Venugopal and Kupper. CGI. «EAI and Web Services – A Simple Guide». (<http://www.capescience.com/articles/content/WebServicesandEAI.pdf>).
- [15] The Web Services Community Portal: <http://www.webservices.org/>.
- [16] World Wide Web Consortium: <http://www.w3.org/>.
- [17] Web Services Activity: <http://www.w3.org/2002/ws/>
- [18] Web Services Tutorials: http://www.xml.com/pub/rw/Web_Services_Tutorials/.
- [19] XML Development, XML Resources, XML Specifications: <http://www.xml.com/>
- [20] Simple Object Access Protocol (SOAP 1.2): <http://www.w3.org/TR/soap12-part1/>
- [21] Web Services Description Language (WSDLv.1.1): <http://www.w3.org/TR/WSDL.html>.
- [22] Universal, Description, Discovery and Integration (UDDIv.3): <http://www.udii.org/>.
- [23] Electronic Data Interchange (EDI): <http://www.xmledi-group.org/xmledigroup/ediindex.htm>.
- [24] Sonic Software White Papers: «Understanding Services-Oriented Architectures (SOA)». <http://www.sonicsoftware.com/>.
- [25] The Globus Project: <http://www.globus.org/>
- [26] OASIS «Organization for the Advancement of Structured Information Standards»: <http://www.oasis-open.org/home/index.php>
- [27] CORBA «Common Object Request Broker Architecture»: <http://www.cs.wustl.edu/~schmidt/corba.html>
- [28] DCE «Distributed Computing Environment»: <http://www.opengroup.org/dce/>
- [29] DCOM «Distributed Component Object Model»: http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dndcom/html/msdn_dcomarch.asp
- [30] IETF «Internet Engineering Task Force»: <http://www.ietf.org/>
- [31] WS-I «Web Services interoperability»: <http://www.ws-i.org/>
- [32] Semantic Web: <http://www.w3.org/2001/sw/>
- [33] RDF «Resource Description Framework»: <http://www.w3.org/TR/REC-rdf-syntax/>
- [34] OSS «Open Source Software»: <http://www.oss.net/extr/page/>
- [35] OGSA «Open Grid Service Architecture»: <http://www.globus.org/ogsa/>
- [36] Sing Li. JXTA Peer-to-Peer Computing with Java. ISBN 1-861006-35-7. 2001 Wrox Press Ltd.
- [37] SMIL «Synchronized Multimedia Integration Language»: <http://www.w3.org/TR/REC-smil/>

AUTORES



Juan Carlos Peláez López, es estudiante de Doctorado de la Universidad Politécnica de Cataluña, en el Departamento de Ingeniería Telemática, donde se encuentra vinculado al grupo de investigación de Servicios Telemáticos desde 2001. Recibió el grado de Ingeniero en Sistemas Computacionales de la Universidad Católica de Santiago de Guayaquil, Ecuador, en 1998. Entre sus campos de investigación se encuentran el desarrollo y propuesta de servicios distribuidos multimedia, tecnología peer-to-peer, Web Services, Grid Computing, y comercio electrónico. Actualmente se encuentra trabajando en su proyecto de tesis sobre "Video Streaming Peer-to-Peer with QoS for the recovery and delivery".



Jorge Mata Díaz, es Doctor Ingeniero en Telecomunicación y profesor del Departamento de Ingeniería Telemática de la Universidad Politécnica de Cataluña. Su trabajo de investigación se enmarca dentro del grupo "Servicios Telemáticos" del mismo departamento y su actividad docente se desarrolla en la ETS de Ingeniería de Telecomunicación de Barcelona. Entre sus campos de investigación se encuentra el desarrollo y propuesta de servicios telemáticos de redes de comunicación terrestres, sin hilos y VSAT: servicios audiovisuales, sistemas distribuidos multimedia, servicios corporativos en entornos intranet e Internet, y servicios de seguridad y comercio electrónico. Los resultados de sus trabajos se han publicado en numerosas revistas y congresos tanto nacionales como internacionales (<http://sertel.upc.es>).

WHEN OPPORTUNITIES GOT SMALLER IN USA



Xavier Anguera Miró

xanguera@gps.tsc.upc.es

If we made a poll 3 years ago to see which country people thought it was «the land of opportunities», for sure the majority of them would answer that the United States of America were. I've been living in the USA most of these last 3 years and I've realized, now that I just came back, that probably at the present time the answer wouldn't be so clear. In these lines I pretend to explain you my journey through this «land», and how it turned out to be very special to me.

These last years the USA have been in the news, mostly for sad reasons. As you might recall, problems began with the presidential reelection process in 2000, where the World saw how the most industrialized country in the world could suffer core organization problems. But it wasn't until the terrorist attack to the World Trade Center that the Americans themselves realized that there were people that hated them. This was such astonishing news that many Americans have not still recovered from. It all got worse with the «explosion» of the Technological Bubble by the end of 2001. By then, many Americans lost their jobs and many looked back to the recession period with fear that it could come back. Finally, during 2002 and 2003 we had to get used to news about war, first in Afghanistan and then in Iraq, trying to erase all these people that USA thought that they were dangerous to them.

I have always wanted to work in the United States. Since early ages I've been reading how people could achieve their dreams, everything was possible in USA. That was why I wanted to be part of that «American Dream». I studied Telecommunications Engineering in UPC and I was lucky to be offered an internship in the US to do my final degree thesis. I didn't doubt and quickly jumped into the plane.

This internship was in a prestigious research laboratory, in Santa Barbara, California. A town located about 90 miles (150 Km) north of Los Angeles, in the coast. It is a very pretty town, between two interesting attraction poles, the wealthy living of Montecito and the typical student life from UCSB (University of California in Santa Barbara). Soon after I begun the internship I was offered to remain in the company after finishing my degree, which was fulfilling my dream. Such offer was implying the request for a working visa, that wasn't supposed to be a problem by then, with so much work to do and the economy going so well.

But it was during a short visit in Spain to present my thesis by September 2001, when the Twin towers collapsed in New York, and with them arrived the end of a period. I came back to my workplace and all plans remained the same, the visa was in progress and, even though Americans were not the same after 9-11, it all seemed to me as if nothing had changed in the other coast of the States. But with the end of the year there came very bad economic results and some companies began to lay off thousands of people, being California the State that suffered the most. I kept

believing that my position was ensured until that came to me. My engineering status offer was turned down, therefore I was technically «laid off».

I was luckier than most people because I could finally keep my engineering position in the company. Some friends of mine were not so lucky and had to go, and some other interns that wanted to stay in the company after their internship, couldn't. In general, during all 2002 and part of 2003 many companies applied a «freezing» strategy in the Human Resources departments. There was in some cases a lot of work to be done, and a big need for more people to be hired, but many managers were scared to see that companies next door were closing down. Therefore the Unemployment rate went up steeply and stayed there almost until nowadays. One area that was not affected at all was the military related jobs, but in these cases the job description was asking for US citizens only, thus not serving as a relieve for all those engineers like me that had moved to California during those crazy technological years.

Despite the troubles, I was able to work in the company in interesting areas within the speech processing area. My work went into topics such as Text-to-Speech synthesis, Automatic Speaker Verification, speech databases creation and robotics.

After 2.5 years I returned to Spain on September 2003. The overall experience was very fulfilling. I am very happy I took the decision to go to another country to work because it gave me a totally new perspective of the way things can be done. I realized there is not a «good» and a «bad» way of doing things when it comes to differences in cultures, there is simply a «different» way.

I encourage everybody to go to another country and merge with another culture for a reasonable period of time (tourism is nice but it only shows us the façade of a country). If problems arise, one just have to keep trying until they get better. I found my personal experience to be a life changing one.

AUTOR



Xavier Anguera Miró. Nacido en Tarragona en 1978. Ingeniero superior de Telecomunicación, graduado en 2001. Màster Europeu en Llenguatge i Parla (MELP), 2001. Durante los últimos años de la carrera se especializó en el tratamiento de la señal de la voz. Ha trabajado durante dos años y medio en el Panasonic Speech Technology Laboratory, en USA. Actualmente es estudiante de doctorado en el departamento de Teoría de la Señal y Comunicaciones de la UPC.

