

EDITORIAL

Sumamos número en Buran.

Bienvenidos de nuevo al mundo en que os sumergimos en los secretos de las últimas tecnologías, objetivos desconocidos e iniciativas inimaginables.

Estaremos de acuerdo en la complejidad y la incertidumbre en que el sector de la informática y las telecomunicaciones ha vivido últimamente. Para empezar, se vivió una «crisis» en medio de tanta euforia por las comodidades de los móviles, internet... Una mezcla de dos motivos importantes causó esa citada «crisis»: el déficit de técnicos cualificados y, por otro lado, la alta inversión que han realizado los operadores sin beneficios inmediatos.

En qué quedamos: ¿Esplendor o caída del sector de las comunicaciones?

La revolución que se ha marcado la técnica en cuanto a facilidades para comunicarse hace apuntar más bien por lo primero, pero las noticias sobre los recientes despidos masivos nos hacen alarmarnos, porque de repente se dan cuenta que las «.com» no están lo suficiente de moda produciendo un fracaso de las optimistas expectativas de los dirigentes. De ahí proviene la reestructuración de las empresas que sólo es consecuencia de una mala previsión de sus beneficios y posibilidades, una mala estructuración de la comercialización... Quizá tenga razón Fiorina (máxima responsable de HP) diciendo que ve signos de desaceleración en el mercado, pero su punto de vista se contradice con otros que opinan que nada ha cambiado.

De hecho en estos momentos, parece que los titulares de «crisis» desaparecen, sólo esperamos que dure y que, al mismo tiempo, las empresas sepan hacer llegar a todos la tecnología que se crea cada día, pues aunque falten técnicos, los ya existentes y los estudiantes tenemos muchos proyectos. Por supuesto con el respaldo de los conocimientos y la ilusión de llevarlos a buen fin. Y para mostrarlo, también contamos con Buran.

Marta Ruiz Costa-Jussà

COORDINACIÓN BARCELONA

Miguel Ángel Sastre Serra
José Cástor Vallés Martínez

EDICIÓN BARCELONA

José A. López Salcedo
Miguel Ángel Sastre Serra
Carles Ruiz Floriach
Marc Caballero Gómez
Enric Muntané Calvo
David Garriga Cornelles
José Cástor Vallés Martínez

REVISIÓN

José A. López Salcedo
Miguel Ángel Sastre Serra
José Cástor Vallés Martínez

EDICIÓN GRÁFICA

Daniel Prado Rodríguez

AGRADECIMIENTOS

II. Dir. Juan A. Fernández Rubio,
Ángel Cardama,
IEEE Internacional, Jorge Luis Sánchez Ponz,
Javier Macías Guarasa,
y a los puntos de
distribución en la UPC:
Abacus, CPET, CPDA, Kiosk Campus Nord
y Reprografía Sant Just.

IMPRESIÓN RET, s.a.l.

FOTOMECÁNICA Sistemes d'Edició

DEPÓSITO LEGAL B-19.950-96

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.

$$\begin{array}{c} \notin \alpha \\ \pm \leq \prod \vee \\ \subseteq \sum \Psi \exists \\ \infty \wedge \neg \phi \end{array}$$

UNA INTERPRETACIÓN DE LA CONVOLUCIÓN ÚTIL PARA MULTIPLICAR ENTEROS DE LONGITUD ARBITRARIA CON LA FFT

Enric M. Calvo

*Estudiante de la ETSETB y Miembro de la Rama de
estudiantes del IEEE de Barcelona
enric25@casal.upc.es*

INTRODUCCIÓN

En este artículo se va a intentar presentar una interpretación “evidente” de la convolución (es decir, evidente una vez ha sido explicada).

Después veremos qué utilidad podría tener, si es que tiene alguna, en la multiplicación de números enteros de longitud tan larga como se quiera, gracias a que la convolución de dos secuencias se convierte en un cálculo muy eficiente gracias a la *FFT - Fast Fourier Transform*.

Al mismo tiempo, se intentará presentar las circunstancias que hicieron posible encontrarla, dado que el autor considera interesante conocer de qué modo se producen estos hallazgos.

El único requisito para seguir con éxito este artículo es tener nociones básicas sobre teoría de sistemas lineales (saber qué es una convolución discreta, la *FFT*, ...).

ANTECEDENTES

Hace algún tiempo, mientras ojeaba un libro sobre curiosidades matemáticas vi una pequeña relación que me hizo gracia, de la que luego me acordaría durante algún tiempo. La curiosidad consistía sólo en ver que:

$$\begin{aligned} 11*11 &= 121 \\ 111*111 &= 12321 \\ 1111*1111 &= 1234321 \end{aligned}$$

Por otra parte, al cabo de algunos meses y después de llevar tratando con convoluciones discretas durante algún tiempo, aprecié, de un modo distinto, la convolución entre dos pulsos rectangulares iguales, que da como resultado un pulso triangular (resultado que todo el mundo conoce después de haber hecho una asignatura sobre señales y sistemas digitales o analógicos).

Como se puede observar, parece que tiene que haber alguna relación entre los dos hechos, así que sólo quedaba investigar de qué manera se relacionan la multiplicación de enteros con la convolución.

CONVOLUCIONAR ES MULTIPLICAR COMO SE APRENDE EN E.G.B.

Sólo hace falta pararse unos minutos a ver de qué forma se realiza la convolución cuando se calcula a mano para ver que, cuando realizamos a mano una multiplicación, estamos aplican-

do exactamente el mismo algoritmo (con la única diferencia que lo hacemos en módulo 10). Por ejemplo, si queremos multiplicar 34·67 sólo tenemos que crear dos vectores en orden inverso al visualizado, es decir:

$$x(n) = 4\delta(n) + 3\delta(n-1) = [4, 3]$$

$$h(n) = 7\delta(n) + 6\delta(n-1) = [7, 3]$$

La convolución de ambos vectores da como resultado $y(n) = [28, 45, 18]$. Alguien podría preguntarse en qué se parece este resultado con el esperado (34·67=2278). Pues bien, si tenemos en cuenta que en las secuencias que habíamos creado lo habíamos hecho en módulo 10, veamos qué ocurre si aplicamos ahora módulo 10 llevándonos el acarreo al siguiente dígito de mayor peso, es decir, el de la derecha. Entonces, la secuencia de salida después de llevarnos el acarreo de la primera posición ($28 \bmod 10 = 8$, luego nos llevamos 2 al siguiente) es $y(n) = [8, 47, 18]$. Haciendo lo mismo para el segundo dígito: $y(n) = [8, 7, 22]$. Y, por último, $y(n) = [8, 7, 2, 2]$, que es el número que esperábamos (si, del mismo modo que cuando hemos creado los vectores, leemos éste de derecha a izquierda).

Puede parecer que todo esto es una “idea feliz”, o una simple casualidad, pero ahora veremos que no es así, que tiene una razón de ser.

AHORA CON TRANSFORMADA Z

Seguramente la manera más rigurosa, bonita y fácil de ver que el algoritmo de multiplicar que se aprende con 7 años es el mismo que el que se aplica en la convolución, es que observemos los números como polinomios en s para luego hacer lo mismo con las secuencias transformadas ya en polinomios de variable z^{-1} (después de haber aplicado la Transformada Z). La gracia de multiplicar los polinomios después de transformarlos es que sabemos que esto equivale a convolucionar las secuencias en tiempo, y acabamos de ver que haciéndolo estaremos realizando la multiplicación que deseábamos. Será, por tanto, la manera más “natural” de tratar con los números que queremos multiplicar, ya que multiplicar polinomios sabemos hacerlo.

El número 34 puede verse como un polinomio $x(s) = 3s + 4$ si se evalúa en $s = 10$. El 67 sería: $h(s) = 6s + 7$. Multiplicados ambos polinomios tenemos $y(s) = x(s) \cdot h(s) = 18s^2 + 45s + 28$, que evaluado en $s = 10$ da el resultado esperado: $34 \cdot 67 = 2278$.

Sólo hace falta substituir $s = z'$ en los anteriores polinomios y veremos que conseguimos las transformadas Z de las secuencias de ejemplo que hemos utilizado en el apartado anterior:

$$X(z) = 4 + 3z^{-1}$$

En este caso encontramos que los dos polinomios multiplicados, al dar el mismo resultado que convolucionando las secuencias, resultan:

$$Y(z) = X(z)H(z) = 28 + 45 z^1 + 18 z^2$$

Por supuesto, igual que en el caso de polinomios en s substituyendo $z^1 = 10$ encontraremos el resultado esperado. Pero, del mismo modo que cuando convolucionábamos, si queremos como resultado una secuencia que sea comprensible a primera vista (es decir, con un dígito en cada posición y ordenados de menor a mayor peso según crece n), tenemos que hacer módulo diez en cada coeficiente y llevarnos el acarreo al siguiente dígito de mayor peso. Encontraríamos entonces que el polinomio “maquillado” es:

$$Y'(z) = 8 + 7z^1 + 2z^2 + 2z^3$$

que antitransformado da el esperado:

$$y'(n) = [8, 7, 2, 2]$$

APLICACIONES

Alguien puede preguntar de qué nos va a servir darnos cuenta de todo esto y qué utilidad tiene para multiplicar números. Pues lo bueno de todo este asunto es lo siguiente: por un lado acabamos de demostrar que multiplicar dos números enteros es equivalente a convolucionar dos secuencias adecuadamente creadas. Por otro lado sabemos que convolucionar se hace de la forma más rápida con la FFT (cuando N es un poco grande). Mezclando estas dos ideas surge la de tratar de multiplicar números de longitud arbitraria con este método.

En efecto, dado que $x(n)^*h(n) = y(n)$ equivale a $y(n) = IFFT\{X(k)H(k)\}$ podemos multiplicar dos cifras con este método. Pero, no sólo eso, si pensamos en el problema inverso (lo que sería un problema de *deconvolución*), podemos, además dividir usando este método ($H(k) = Y(k)/X(k)$), aunque ésta última operación conlleva más problemas.

Ejemplo de un multiplicación 568974568712345 · 458796325874125=261043441641038763007803573125 con FFT:

$$x(n) = [5, 4, 3, 2, 1, 7, 8, 6, 5, 4, 7, 9, 8, 6, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$$

$$h(n) = [5, 2, 1, 4, 7, 8, 5, 2, 3, 6, 9, 7, 8, 5, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$$

$$y(n) = x(n) * h(n)$$

$$y(n) = \text{IFFT}\{X(k)Y(k)\}$$

$$y(n) = [25, 30, 28, 40, 63, 119, 141, 125, 134, 183, 271, 321, 328, 320, 352, 380, 402, 406, 376, 342, 308, 287, 282, 263, 224, 159, 102, 49, 20, 0, 0, 0]$$

Haciendo una pasada de acarreos queda:

$$y(n) = [5, 2, 1, 3, 7, 5, 3, 0, 8, 7, 0, 0, 3, 6, 7, 8, 3, 0, 1, 4, 6, 1, 4, 4, 3, 4, 0, 1, 6, 2]$$

Notar que se les ha hecho un *zero-padding* a ambos vectores hasta $N=32$, para así poder hacer cómodamente la IFFT.

PUNTOS FUERTES

Algunos rasgos atractivos de este método son, por ejemplo:

- Realizar FFT's es muy rápido para secuencias con $N = 2^n$, luego se puede conseguir la multiplicación de 2 secuencias larguísimas en muy poco tiempo.
 - Este método es igualmente aplicable para otras bases. Así, es trivial implementarlo en base 16, con lo que el acarreo, por ejemplo, no supondría apenas coste computacional.

- Viendo que es válido para números enteros, también lo es para números reales (de precisión finita). Para ello sólo hay que considerar que la secuencia de salida tiene una coma situada en la posición resultante de sumar el número de decimales de las dos secuencias.

La aplicación más notable para la que pensé que podía llegar a ser útil es en criptografía, dado que los números con los que se trabaja son larguísimos (unos 155 dígitos en decimal equivalen a una clave de 512 bits, aprox.), aunque, dado que la eficiencia de este método crece conforme N aumenta, es posible que sea para un N bastante grande cuando multiplicar usando métodos basados en FFT sea más eficiente que con métodos tradicionales. Tal vez sería útil para la factorización de números enormes en dos primos, aunque como veremos dividir no resulta tan óptimo como multiplicar.

INCONVENIENTES Y OTRAS CONSIDERACIONES A TENER EN CUENTA

La verdad es que para comprobar la utilidad de este método haría falta saber el coste computacional de otros métodos de multiplicación que se utilizan para manejar números de longitud arbitraria, ya sea implementados vía software o vía hardware, costes que en el momento de redactar el artículo el autor desconocía. La *FFT* tiene un coste computacional de $(N/2)\log(N)$ multiplicaciones, con lo que invalidaría la utilidad del método si existe actualmente una implementación con menor coste computacional.

También se presentan varios problemas básicos relacionados con la longitud de las secuencias a multiplicar y la FFT.

· Sabemos que, de hecho, la multiplicación de las FFT's de dos secuencias da como resultado la convolución *circular*, y no la convolución normal. Para solucionar este inconveniente, sólo hace falta llenar las secuencias con 0's (*zero-padding*)

hasta tener 2 secuencias de longitud $N = 2^n$ cada una, donde N sea más grande que la suma de las longitudes en dígitos de las dos cifras a multiplicar.

El mayor inconveniente de este método es que, una vez transformadas las secuencias tenemos que multiplicar (o dividir) punto a punto cada secuencia. Éste es uno de los mayores “handicaps” de este método, ya que requiere la multiplicación de 2 números complejos (donde se necesitan 4 multiplicaciones reales), o la división de 2 números complejos (donde pueden llegar a necesitarse 6 multiplicaciones y 2 divisiones reales, si es que no hay algún método óptimo). De todos modos en el apartado de optimizaciones veremos qué podemos hacer para tener que multiplicar punto a punto.

Otro inconveniente que habría que comprobar si es importante es que, debido a que N puede ser muy grande, no haya un arrastre de errores en la *FFT* y en la *IFFT* que haga que los resultados a partir de una cierta N sean equivocados. Hay que tener en cuenta que cada dígito está representado por un valor de 16 ó 32 bits, por lo que hay que asegurarse que no desborda. El peor caso sería multiplicar dos secuencias de N valores máximos cada una, y el mayor resultado se daría cuando, en la convolución, no hubiese desfase entre ambas (es decir, justo el instante de la convolución en que las dos secuencias “coinciden”). El valor del dígito en ese punto sería de $N \cdot (\text{base}-1)^2$. Podemos calcular entonces fácilmente la N máxima calculable para los casos en que utilicemos enteros de 16, y de 32 bits y usemos base 10 ó 16 (ver tabla adjunta).

	base 10	base 16
int16	809	256
int32	$53 \cdot 10^6$	16777216

Tabla 1: n^o máximo de dígitos

Dado que después de aplicar la *FFT* tenemos una secuencia compleja compuesta por un par de números reales de precisión finita, es posible que al antitransformar los números no sean enteros, con lo que habría que tratar de acotar el error de algún modo.

OPTIMIZACIONES

A la vista de la operación que hay que hacer para calcular la DFT (aunque a la práctica se realizase con la *FFT*)

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi k n}{N}}$$

podemos proponer alguna “optimización”. Por ejemplo para calcular la DFT de $x(n)$. Se puede observar que, como hemos dispuesto un dígito de la cifra en cada posición del vector $x(n)$, en este sumatorio $x(n)$ no tendrá otro valor que uno comprendido en el rango 0-9. Esto da naturalmente pie a que el cálculo de la DFT pueda redu-

cirse a la suma de N referencias a una tabla de $10 \cdot N$ complejos. Esta tabla puede incluso caber en la *caché* de un ordenador, con lo que nos aseguramos una velocidad muy alta de acceso a esta tabla. Naturalmente esta optimización puede ser mayor si descartamos el 0 y nos damos cuenta que con los valores de un $\cos(x)$ evaluado entre $[0, \frac{\pi}{2}]$ tenemos suficiente para representar todos los valores de las funciones trigonométricas que componen la exponencial compleja. Con estas últimas optimizaciones podríamos llegar a necesitar sólo $\frac{9N}{4}$ referencias. El inconveniente de optimizar la *DFT* (no la *FFT*) es que tenemos que hacer $2N^2$ sumas, cosa que no asegura una mejora rotunda del algoritmo en comparación con hacer algunas multiplicaciones.

Otra optimización podría basarse en una de las ventajas antes comentadas. Dado que la base a utilizar en la secuencia es arbitraria, podríamos usar base 16, una más natural para los ordenadores. De este modo el acarreo (dividir módulo 16) pasa a ser una secuencia muy simple de lógica binaria, cosa que parece muy conveniente. Pero aún se puede llevar más allá, de manera que si los dígitos con los que trabajamos son todavía mayores (siempre en base 2^n) podemos ahorrarnos varias multiplicaciones para hacer el mismo cálculo (es como si hubiésemos “concentrado” en menos coeficientes la “señal”).

CONCLUSIONES

Bien, decía al principio del artículo que convolucionar era como multiplicar enteros con el algoritmo que todo niño conoce, aunque, para ser exactos habría que decir que es igual que multiplicar polinomios. Esto es evidente, en cuanto a que convolucionar dos secuencias es lo mismo que multiplicar sus transformadas Z, que no dejan de ser polinomios. Despues hemos aplicado el hecho de que convolucionar se hace más rápido en frecuencia, con lo que nos ahorraremos cálculos. Como puede verse, al fin y al cabo era algo evidente lo que se trataba de hacer notar en este artículo, aunque tal vez no todo el mundo se había dado cuenta de ello.

Futuras líneas de investigación podrían ir encaminadas a estudiar el mismo problema pero con otras transformadas, para ver si es posible ahorrarnos los cálculos en complejos que implica el trabajar con la *FFT*.

También se ha expuesto el hecho de que es muy posible que otros métodos para el cálculo de multiplicaciones y divisiones de números de longitud arbitraria sea más eficiente, o que con este método se produzcan errores a partir de un cierto valor N . Si es así y el método no es válido, esta interpretación de la convolución no pasará más allá de ser una curiosidad bella de admirar.

SISTEMA DE RECONOCIMIENTO DE OBJETOS SOBRE VISION FOVEAL MEDIANTE REALIDAD VIRTUAL



*José Luis Herreros Díaz, Cristina Urdiales García,
Antonio Bandera Rubio*

*Dpto. Tecnología Electrónica, E.T.S.I. Telecomunicación,
Universidad de Málaga, Campus de Teatinos, 29071 Málaga (España)*

Actualmente un número creciente de investigadores ha centrado sus esfuerzos en la robótica, con el objetivo de desarrollar sistemas móviles capaces de desenvolverse en entornos desconocidos, cambiantes y potencialmente hostiles sin necesidad de la intervención directa de un operador humano. Así, cada vez aparecen más aplicaciones de este tipo, como el PIONEER, desarrollado para la limpieza de los restos de Chernobyl, el DANTE II, para el descenso al volcán Mt. Spurr o el SOJOURNER, explorador de Marte.

Uno de los principales problemas de estos sistemas es la necesidad de que se comporten al menos de forma parcialmente autónoma, ya que las condiciones en que desarrollan su trabajo imposibilitan un telecontrol exhaustivo. Para trabajar de esta forma, un agente precisa un gran número de módulos trabajando de forma coordinada. Por citar sólo algunos ejemplos, podrían citarse el módulo para evitar colisiones, el módulo para calcular la posición actual o el módulo para desplazarse a un punto determinado. Un problema adicional radica en el hecho de que, incluso aunque el agente fuese completamente autónomo, en la mayoría de los casos se requiere una comunicación con un sistema remoto para enviar los datos recogidos del entorno en que se opera. Podría pensarse en almacenar localmente dichos datos en el robot para recogerlos cuando éste se recuperare, pero entonces la pérdida del agente, bastante habitual en estos casos, supondría también la pérdida completa de la información. Habitualmente se recurre a la comunicación vía radio de los datos conforme éstos se van adquiriendo. Dichos datos pueden recogerse a partir de diversos sensores, como táctiles, sonar o infrarrojos, pero indudablemente la visión sigue siendo una de las fuentes más ricas de información para el ser humano.

La mayor desventaja de transmitir con imagen digital vía radio es el enorme volumen de información que debe enviarse a través de un medio ruidoso de ancho de banda limitado, lo que reduce drásticamente el número de imágenes transmitidas por segundo y supone pérdida de información en tanto que el vehículo se va desplazando y dejando atrás escenarios que podrían no haber sido capturados. Para solucionar este problema, existen técnicas de compresión que permiten reducir notablemente el volumen de información

de una secuencia de vídeo, como por ejemplo el estándar MPEG. Sin embargo, la mayoría de estas técnicas implican una pérdida de calidad generalizada que podría imposibilitar el procesado de dichas imágenes si los algoritmos de trabajo requieren un alto nivel de detalle. Una segunda alternativa radica en el uso de visión atencional con campos de resolución no uniforme, como se presenta a veces en el reino animal en ejemplares como las águilas o algunos peces. Este tipo de visión consiste en mantener continuamente un área pequeña de muy alta resolución, conocida como fovea, sobre las áreas de interés de la imagen, mientras que el resto de ésta se presenta con un perfil de resolución decreciente en torno a ésta (Tistarelli y Sandini, 1993) (Camacho et al, 1996). En este caso es necesario definir una estrategia para calcular las áreas de interés de la imagen de forma automática o bien posibilitar su definición desde el extremo receptor, pero a cambio se obtienen reducciones muy considerables del volumen de datos sin perder resolución en dichas áreas.

Este trabajo se centra en el desarrollo de un sistema de control para el módulo de visión de un agente de este tipo, prestando especial atención al desarrollo de una interfaz eficaz hombre-máquina que permita una utilización sencilla e intuitiva de la herramienta. El agente presenta un módulo de navegación que reside en un ordenador remoto (Urdiales, 1999). La comunicación entre el ordenador de a bordo y dicha máquina se lleva a cabo mediante un enlace radio y consiste en una secuencia de lecturas de sensores sonar, infrarrojos y de contacto en dirección al módulo de navegación, que devuelve secuencias de comandos a los motores del agente. Al mismo tiempo, el agente está dotado de un módulo de visión binocular que captura mediante un frame-grabber imágenes de resolución homogénea en escala de grises. Estas imágenes deben transmitirse a una máquina cualquiera a través del enlace radio para que el teleoperador pueda observar el entorno sobre el que se encuentra el robot. Para acelerar la comunicación vía radio, dichas imágenes se fovealizan previamente a su envío. Como aplicación en el sistema remoto se ha implementado un reconocedor de objetos, que requiere que el objeto a analizar se presente a alta resolución. Para ello, el teleoperador debe establecer el área de interés en cada momento y



enviar las órdenes pertinentes para arrancar el proceso de reconocimiento. Asimismo, si lo desea, podrá nombrar el objeto reconocido. Como interfaz de usuario, siguiendo las últimas tendencias en este tipo de aplicaciones, se ha recurrido a periféricos de realidad virtual, que permiten que el teleoperador se relacione de forma natural e intuitiva con el software de control.

El conjunto se puede dividir en tres bloques muy diferenciados:

- * Extremo robot, incluyendo los motores, sensores, cámaras y un PC para preprocesado de la información.

- * Vía de comunicación, consistente en un enlace radio Ethernet de bajo ancho de banda.

- * Extremo remoto, donde se reciben las imágenes y las órdenes del teleoperador y se lleva a cabo el proceso de reconocimiento.

Existiría un cuarto bloque, la máquina donde corre el módulo sensomotor del agente, pero no se va a contemplar en este trabajo en tanto que no está relacionado con el ciclo de visión.

1.- ARQUITECTURA FÍSICA

El extremo robot consiste en un agente autónomo móvil modelo Nomad 200 capacitado, entre otras muchas tareas, para la captación de imágenes. El sistema de captura se compone de dos cámaras Hitachi 450 y una placa de adquisición tipo Sensus 700 que permite, si se desea, tratamiento de imágenes de resolución homogénea tanto en color como en escala de grises. La tarjeta capturadora admite hasta un total de 4 DSPs tipo C4X trabajando en paralelo, aunque la arquitectura actual sólo incorpora uno de ellos. El trabajar con un único C44 imposibilita la visión estereóreal, ya que no pueden capturarse dos imágenes simultáneamente.

Además del equipamiento anterior, el Nomad 200 cuenta con un PC tipo Pentium a 133 MHz. El DSP se comunica con el PC mediante un bus ISA de 1 Mbyte/seg de ancho de banda. Dicho PC es el encargado de gestionar todo el sistema autónomo, aunque los sensores se preprocesan mediante microcontroladores tipo HC11 de Motorola para restar carga al sistema.

Para comunicarse con el exterior, el agente hace uso de un enlace radio tipo Ethernet cuya velocidad de transmisión media para el protocolo TCP/IP utilizado es de aproximadamente 70 Kbytes/seg. El otro extremo del enlace está conectado a una red local y, mediante ésta, el agente puede comunicarse con cualquier ordenador conectado a dicha red.

En este caso, el extremo remoto soporte para el procesamiento de las imágenes es un Pentium II a 333

MHz, que cuenta con cuatro puertos serie para poder añadir los periféricos de realidad virtual necesarios, que, en este caso, es un guante tipo 5th DATA GLOBE desarrollado por 5th Dimensions Technologies. Este guante de licra es adaptable a cualquier tamaño y consta de dos tipos de sensores:

- De orientación: mediante un mecanismo interno es capaz de cuantificar el giro del guante. Hay dos tipos de medida: *pitch* (grado de inclinación del brazo) y *roll* (giro lateral de la muñeca), que se representan en la Fig. 1.

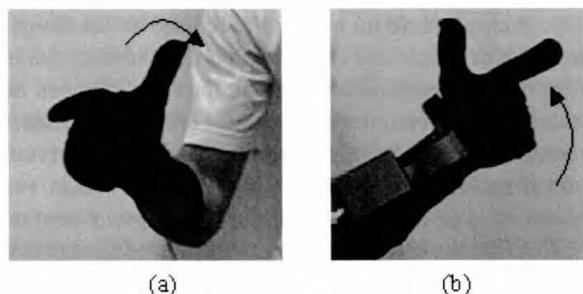


Fig. 1.- a)Roll, y b)Pitch

- De flexión: se usan cables de fibra óptica que se prolongan por cada uno de los dedos. Conforme se flexiona cada uno de ellos, por las características de la fibra, varía el índice de reflexión. A ambos extremos de la fibra se incorporan un fotoemisor y un fotorreceptor respectivamente para medir qué porcentaje de la luz transmitida es recibida, lo que depende directamente del índice de reflexión y, por tanto, del grado de flexión en cada dedo.

La información de ambos tipos de sensores se codifica con 8 bits de resolución por sensor, lo que supone 256 bytes en total, y se envía por un interfaz serie RS-232 a la estación de trabajo encargada de procesar la información. De los 7 sensores disponibles se transmiten hasta 200 muestras por segundo.

Al emplear este tipo de periféricos es recomendable usar gestos simples para disparar las distintas acciones, teniendo en cuenta la flexión o extensión de cada uno de los dedos mediante el uso umbrales. En este trabajo se han desarrollado una serie de mecanismos para perfeccionar estas medidas evitando, en la medida de lo posible, los transitorios asociados a una acción cualquiera. El primero de dichos mecanismos es la implementación de un ciclo de histéresis con el fin de evitar la confusión de distintos tipos de gestos. La consecuencia directa de esto es que para extender un dedo, no bastaría con relajar el miembro, sino que se debe extender completamente. Para flexionarlo es necesario operar de forma similar. Otro mecanismo de mejora a este respecto ha sido la introducción del método matemático «media alfa-recortada», que elimina de las medidas los valores extremos y efectúa la media de los restantes. De esta manera se filtran los datos espúreos que falsean las mediciones.

Debido a la simplicidad del método de medida, diseñado para dar velocidad a proceso de toma de muestras, es necesaria una fase de inicialización que va a ser determinante, ya que a raíz de ésta se fijan los ejes imaginarios de coordenadas sobre los que trabajan el usuario y el sistema.

2.- ESQUEMA GENERAL DEL SISTEMA

Tal como se ha comentado previamente, el sistema se ha dividido en una serie de bloques interconectados. El primero de éstos, que es denominado *extremo robot*, realiza la captación de imágenes, un primer tratamiento de los datos y su empaquetamiento para la transmisión. La *vía de comunicación* comunica el extremo robot y el remoto. Se encarga de transmitir los paquetes de datos hacia el PC, donde serán tratados, y además envía los parámetros de control hacia el robot. La comunicación entre bloques se establece a través de dos sockets: uno para la transmisión de datos y otro para el envío de los parámetros de control. El *extremo remoto* captura la información, monta las imágenes recibidas, crea la estructura de datos necesaria para su posterior estudio, efectúa un tratamiento de éstas y, finalmente, presenta los resultados. Dentro del extremo remoto cabe mencionar varios módulos:

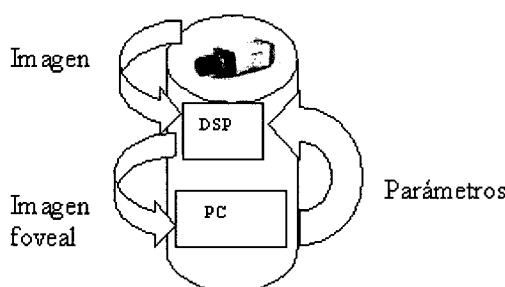


Fig. 2. Esquema de captura de imágenes en el robot

- Tratamiento de la imagen y representación gráfica: incluye en montaje de la imagen, la creación de la estructura de datos y su presentación en pantalla.

- Segmentación y detección de objetos: supone un laborioso tratamiento en la estructura de datos para depurar la información, de manera que se llegue finalmente a la detección de los objetos existentes en la imagen captada.

- Reconocimiento: partiendo de la detección anterior, se conoce el color de cada objeto, lo que permite encajarlo en la zona de máxima resolución de la imagen para analizarlo en detalle y reconocerlo a partir de su contorno.

- Aprendizaje: es la parte encargada de almacenar los datos referentes al reconocimiento e identifica-

ción de los objetos en memoria no volátil. Si se desea, puede asignársele una etiqueta con su nombre. En cualquier caso, la siguiente ocasión en que aparezca un objeto previamente memorizado, éste será reconocido.

- Interfaz guante: es el módulo encargado de captar los gestos y posturas que se adoptan con el guante de realidad virtual e interpretarlos, dotando de significado cada movimiento.

3.- DESARROLLO DEL SISTEMA

3. 1. Extremo robot

I. Captura de imágenes.

Para capturar imágenes el robot hace uso del sistema que se muestra en la Fig. 2. El DSP HEFCG44 (versión del TMS320C40 de Texas Instruments) proporciona tratamiento de vídeo. Aunque dicho DSP está capacitado además para el tratamiento de imagen, éste se efectúa en el PC encargado del control ya que numerosas pruebas mostraron que encargar de dicha tarea al DSP ralentizaba el proceso global. En la captura, el DSP actúa como esclavo del PC, que ordena secuencialmente la captura de imágenes.

La comunicación PC-DSP utiliza un protocolo de comunicaciones llamado *RPC* y, dentro de éste, el modo *call-back*, que permite la ejecución paralela de dos procesos. De esta forma puede llevarse a cabo simultáneamente la captura de la imagen en el DSP, el tratamiento de la misma en el PC y la transmisión de los datos preprocesados. La consecuencia directa de este enfoque fue la duplicación de la velocidad en esta parte del proceso.

II. Tratamiento de la imagen

Para transmitir una secuencia de imágenes a través de un enlace de ancho de banda reducido, tal como se comentó anteriormente, es necesario comprimirlas. La compresión utilizada consiste principalmente en la fovealización y el empaquetamiento de los datos.

La fovealización es la conversión de una imagen de resolución uniforme en una multiresolución. Se basa en definir un área de máximo nivel de detalle denominada fóvea en la zona de interés, mientras que la resolución que va disminuyendo progresivamente al alejarse de ésta. De esta forma, se reduce el volumen de datos de la imagen manteniendo constante el campo de visión y obteniendo un nivel de resolución alto en las áreas importantes. La visión foveal presenta dos geometrías básicas:

- Geometría Logaritmo Polar (Fig. 3.a) (Tistarelli y Sandini, 1993). Similar al mecanismo de visión de



algunos animales. El muestreo que se efectúa es «no uniforme» y los campos receptivos están ubicados en círculos concéntricos. Otra característica es que los sensores no son de tamaño homogéneo. Este mecanismo no se ha empleado en el trabajo actual debido a la necesidad de algoritmos de procesado muy específicos para adecuarse a las estructuras de datos resultantes de un muestreo de este tipo. Además, en la Fig. 3.a puede observarse la existencia de zonas muertas entre el cuadrado de la fóvea y el primer círculo de elementos sensores.

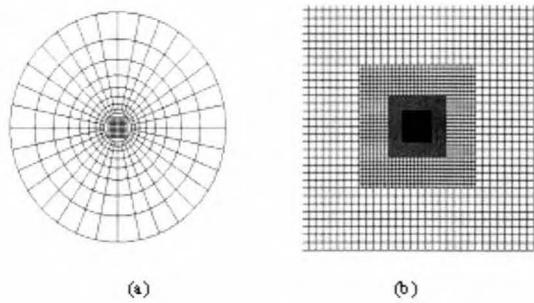


Fig 3. Geometrías foveales: a) Log-Polar; b) cartesiano-exponencial

- Geometría cartesiano-exponencial (Fig. 3.b) (Camacho et al., 1996). Esta geometría presenta discontinuidades en su perfil de resolución. El tamaño de los campos receptivos es siempre un número entero en relación con el tamaño de los sensores de la fóvea.

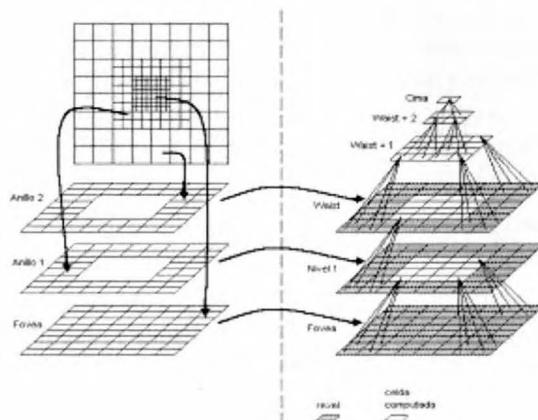


Fig. 4. Polígono foveal

Dado que los algoritmos de visión para imagen uniforme son fácilmente adaptables a este tipo de imágenes, es la que finalmente se empleará en el sistema implementado.

La estructura de datos 3D que permite manejar la geometría cartesiano-exponencial se denomina *polígono foveal* (Fig. 4) y es una estructura compuesta por capas, donde la primera se corresponde con la fóvea y el resto con la combinación de cada nivel de resolución con la capa previa (Bandera, 1994) (Camacho et al., 1996).

Para que dicha capa presente la misma resolución que el anillo de trabajo y, por tanto, pueda combinarse con éste en un nivel de resolución uniforme, la resolución de ésta puede reducirse mediante diezmando, tomando una muestra aleatoriamente de cada conjunto de 2×2 celdas, promediando, llevando a cabo la media aritmética de cada conjunto de 2×2 celdas, o mediante filtros complejos, pero esta última opción ralentiza considerablemente el proceso si se efectúa mediante software. Al existir distintos niveles de resolución en el polígono foveal, se puede trabajar en el más adecuado para cada tarea. Así, para la detección de objetos interesa una capa alta de la estructura de datos, para trabajar con el menor número de celdas posibles. Sin embargo, el reconocimiento, que requiere un buen nivel de detalle, se llevará a cabo en la fóvea.

A pesar de las ventajas mencionadas, la geometría cartesiano-exponencial clásica requeriría movimientos de cámara de alta precisión para centrar el área de máxima resolución en el punto de interés. Dado que serían lentos y costosos, se ha optado por utilizar variaciones de la geometría clásica denominadas *geometrías multirresolución de fóvea desplazable*, evaluándose tres tipos (Fig. 5) (Arrebolá, 1998):

- Básicas: donde el desplazamiento entre niveles constante, así como el número de celdas. La estructura de datos es el *polígono foveal asimétrico*, que es una versión del polígono foveal donde la región de celdas computadas no tiene una ubicación fija (ni centrada como en el caso previo).

- De movimiento generalizado: en este caso el desplazamiento entre niveles de resolución sucesivos no tiene que ser igual. Cada nivel se desplazará los réxeles (denominación que se le da a una celda en las capas de la estructura de datos) necesarios para obte-

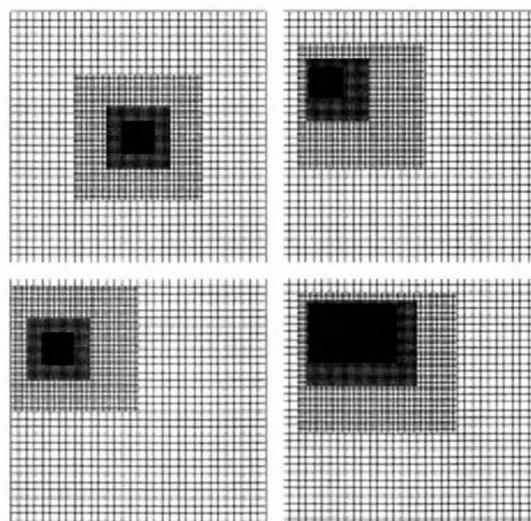


Fig. 5. Geometrías foveales cartesiano-exponentiales: a) de fóvea centrada; b) básicas; c) de movimiento generalizado; d) de tamaño adaptativo

ner la fóvea en la posición deseada. El tamaño de cada nivel de resolución es fijo.

- De fóvea de tamaño adaptativo: aquí el tamaño de la fóvea y, por tanto, del resto de niveles, puede variar y adaptarse al objeto de interés de la escena. Una vez se han determinado las dimensiones de cada nivel de resolución, el desplazamiento entre estos es constante.

Esta última geometría es la empleada finalmente. La motivación esencial es que la fóvea se adapta al objeto a estudiar mejor que en los casos anteriores, con lo que el volumen de datos resultante queda optimizado para cada objeto.

3.2 Transmisión de imágenes

El agente autónomo móvil requiere un interfaz radio para comunicarse sin que su movilidad se encuentre limitada por un cable. El sistema de transmisión aéreo disponible es un enlace radio Ethernet que, tras montar sobre él un protocolo TCP/IP de comunicación, trabaja a una velocidad de aproximadamente 70 KBytes/seg, lo que obliga a reducir al máximo la carga de datos a transmitir para tener una tasa de imágenes por segundo aceptable.

El protocolo de transmisión empleado es de tipo *cliente-servidor*, donde el robot actúa de servidor de imágenes para cualquier usuario que se le conecte, mientras que el PC remoto es el cliente que captura los datos. Así, el servidor envía la imagen fovealizada, y el cliente la captura remitiéndole los parámetros de la próxima ubicación de la fóvea, que el usuario dispondrá mediante el guante de realidad virtual sobre el área que le interese.

3.3. Extremo remoto

I. Segmentación y detección

Una vez recibida la información procedente del extremo robot y antes de procesarla, el receptor debe montar la estructura de datos denominada *polígono foveal asimétrico*, que no es más que una adaptación del *polígono foveal* (Fig. 4) para poder trabajar con la geometría multiresolución de fóvea desplazable y tamaño adaptativo.

La segmentación es un proceso de partición de una imagen en un conjunto de regiones no solapadas cuya unión es la imagen completa. La detección es la determinación de qué regiones de la segmentación corresponden a objetos físicos existentes en la escena. Para ser detectadas como objetos, las regiones han de presentar características homogéneas frente a una aplicación de alto nivel (Haralick y Shapiro, 1992). Una labor esencial en esta parte del procesado es la identificación y etiquetado de todas y cada una de las regiones disponibles.

Dado que la segmentación de imágenes reales es un proceso muy complejo, para llevarlo a cabo se han estudiado diversas alternativas:

- Segmentación por mezclado: consiste en recorrer la imagen píxel a píxel y comparar el valor de brillo de cada punto con el de las regiones existentes. Caso de parecerse a alguna de éstas, se engloba en la región oportuna y en caso contrario, se genera una nueva. Presenta el problema de que no se tiene en cuenta la conectividad entre regiones (Pitas, 1993).

- División y mezclado: sigue una filosofía descendente (*top-down*) y parte de la premisa de que toda imagen es homogénea. Caso de no ser así, la imagen se divide en cuatro subimágenes. El estudio se repite recursivamente dentro de cada subimagen hasta obtener regiones uniformes. Una vez concluido este proceso comienza la fase de mezclado, ya que habrán surgido una infinidad de clases, la mayoría pertenecientes a unas pocas. Se unen aquellas clases en contacto cuyo color sea lo suficientemente parecido. Su mayor desventaja es una excesiva lentitud (Pitas, 1993).

- Segmentación jerárquica multiclasa: que ha sido finalmente seleccionado por adaptarse perfectamente a una estructura del tipo del polígono foveal y porque es el que mayor robustez presenta para imágenes captadas en entornos reales. Es un algoritmo iterativo aplicado al polígono foveal asimétrico que se basa en estabilizar esta estructura nivel a nivel mediante el principio de enlace adaptativo. Así, cada celda del nivel $K-1$ busca la más similar entre las 2×2 celdas situadas inmediatamente sobre ella en el nivel K , denominadas padres. Cuando ha finalizado la este proceso, cada celda padre toma el nivel de gris promedio de las celdas hijas que han enlazado con ella. La estabilización continúa mientras existan celdas padre cuyo nivel de gris cambie (Arrebola, 1998).

Una vez se ha estabilizado toda la estructura de datos, tiene lugar la *generación de clases*. Para ello se elige un nivel que se denomina de propagación. Cada nodo de dicho nivel está asociado a un conjunto de celdas de la base a través de los enlaces establecidos entre niveles. La estabilización adaptativa garantiza que dichos nodos componen una región homogénea, ya que se ha obtenido la agrupación progresiva de los nodos cercanos más similares. Adicionalmente, el proceso aporta un suavizado generalizado y un realce de bordes.

Para localizar las regiones susceptibles de ser un objeto, se calcula su *bounding-box*, que es rectángulo mínimo que engloba a la región a estudiar. Una vez se tienen las *bounding-boxes*, para evaluar potenciales objetos se emplea un criterio de compacidad, que comprueba que las *bounding-boxes* encontradas tie-



nen un área, como máximo, tres veces mayor que la celda asociada en el nivel de propagación. Este criterio permite calcular todas manchas de la imagen que destacan sobre un fondo. Una vez localizadas todas las posibles manchas, es necesario un tratamiento para depurar las detecciones y dejar únicamente las correspondientes a objetos realmente existentes. Se llevan a cabo cuatro mecanismos de depuración:

a) Test de compacidad. Si se detecta más de una clase, se mide el área de cada una. Si el porcentaje de ocupación de la celda es mayor del 80 % la clase se toma como válida y el objeto es detectado. Si no es así se descarta.

b) Fusión de raíces. Se comprueba si alguna raíz potencial es susceptible de unirse con otra (bounding-box solapadas, niveles de gris similares y celdas hijas han de formar una región conectada). Si cumple las condiciones adecuadas se unen en una sola.

c) Test de conectividad. Se recorren las celdas secuencialmente, y las que tengan un nivel de gris similar al del objeto se etiquetan con un número. Pasadas sucesivas asignan la misma etiqueta a las celdas adyacentes. La etiqueta más numerosa es la que se toma como válida.

d) Mecanismo de contraste. Se estudia si el objeto se distingue de su entorno más cercano. Para ello se compara el nivel de gris del objeto con el fondo.

Con todos estos mecanismos se logra tener localizados los posibles objetos englobados en la imagen tratada, listos para el siguiente tratamiento.

II. Reconocimiento.

La detección sirve para señalar posibles objetivos a estudiar, facilitando la labor de búsqueda y, por tanto, economizando recursos. Gracias a este proceso, se obtienen unas zonas señaladas como potenciales objetos. Debido a que el reconocimiento se va a efectuar con exclusividad en la fóvea, el proceso anterior permite orientar dicha zona de máxima resolución, de forma que se centre en un área ya marcada. Todo esto se conoce como *mecanismo atencional* (Bandera et al, 2000). Llegado este punto, el teleoperador es quien decide la zona a estudiar. Para ello hay dos posibilidades:

1. Señalar una zona marcada, lo cual requiere tener activada la detección automática de objetos.
2. Reconocer una zona, independientemente de las detecciones.

Una vez localizada la zona a estudiar tiene lugar la *caracterización* de la imagen contenida en la fóvea. Para ello es imprescindible que el sistema haya detectado un objeto, y que éste se encuentre totalmente

contenido en la zona de máxima resolución. En este procesado se efectúan las operaciones que a continuación se detallan.

En primer lugar se lleva a cabo una binarización objeto-fondo, de manera que se marca con diferente nivel de gris cada una de esas partes. A continuación se aplican unos algoritmos de limpieza para depurar el proceso anterior:

- Dilatación.- Elimina el ruido interior del objeto.

- Erosión.- Es el algoritmo complementario a la dilatación. Aquí se trata de eliminar los puntos externos que, teniendo un color similar al objeto, se encuentran fuera del mismo.

Con todo esto se logra una compactación de colores por zonas y un suavizado del borde. A continuación tiene lugar una nueva segmentación por niveles de gris, de manera que se marca con otro color el contorno del objeto. Finalmente tiene lugar la eliminación de aristas de la periferia.

Tras limpiar y preparar la imagen se lleva a efecto el estudio de la misma, que se inicia con el cálculo del *código cadena* del contorno del objeto (Freeman, 1977). Este código cadena se usa para calcular la curvatura del contorno. Para ello, se emplea el algoritmo de código cadena incremental de longitud variable, que asigna a cada punto del contorno una dirección que permitiría seguir la silueta (Bandera et al, 2000b).

III. Reconocimiento

La función de curvatura obtenida se compara con otros vectores predeterminados que componen una base, la cual ha sido extraída del estudio de unos modelos prefijados. A través de la correlación de la función de curvatura con cada vector de la base, se obtiene un vector de características. Con estos datos se efectúa la clasificación utilizando distancias euclídeas y un umbral: si la distancia el vector obtenido y los previamente almacenados supera dicho umbral se crea una nueva clase. En caso contrario se añade a la clase ya creada más parecida.

IV. Interfaz de usuario

El usuario tiene en la pantalla el interfaz que se presenta en la Fig. 6. En la esquina superior izquierda de la pantalla se puede ver la imagen recibida en tiempo real y a su derecha la versión ya segmentada de ésta, con las detecciones positivas englobadas en su *bounding-box*. La imagen superior derecha corresponde a la fóvea que se capta en cada momento.

La interfaz incluye un guante de realidad virtual que, mediante un conjunto de gestos, permite llevar a cabo las siguientes funciones:

- Desplazamiento de un puntero, que aparece como un punto naranja en la Fig. 6, por la pantalla para seleccionar distintas regiones de la imagen o pulsar botones de la interfaz.

- Centrado de la fóvea en la imagen captada.

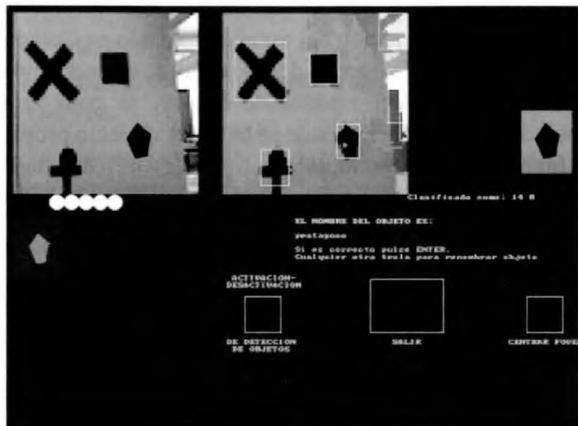


Fig. 6. Pantalla de la interfaz de usuario

- Activación/desactivación de la detección automática de objetos activando con el puntero los botones correspondientes.

- Posicionamiento de la fóvea sobre un objeto cualquiera. Si la detección es automática, basta desplazar el cursor hasta cualquiera de las cajas detectadas y seleccionarla. En caso contrario, hay que arrastrar la fóvea hasta el área deseada y redimensionarla manualmente.

- Solicitud del reconocimiento del objeto contenido en la zona de máxima resolución. En este caso, aparece una imagen en la esquina inferior izquierda, correspondiente a la caracterización de la fóvea: binarización, segmentación y seguimiento del contorno para su reconocimiento. Una vez se ha completado este ciclo con éxito se obtiene el resultado de la clasificación e identificación, donde se indica el nombre que se le asigna a dicho objeto.

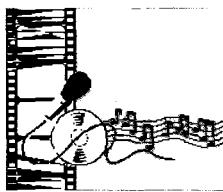
4.- CONCLUSIONES

En este trabajo se presenta la implementación de un sistema de reconocimiento supervisado completo para un agente autónomo móvil que se desenvuelve en entornos cerrados y capta imágenes mediante una cámara analógica y un *frame grabber*. Las principales novedades que presenta el sistema descrito son: i) el empleo de un algoritmo realimentado de detección de objetos, que permite descartar de forma efectiva el

fondo de la imagen, y extraer correctamente los objetos de la imagen; ii) la integración de los algoritmos que tratan con imágenes multirresolución en un sistema complejo de transmisión y compresión de imágenes; y iii) la efectiva caracterización de objetos para su posterior reconocimiento.

REFERENCIAS

- (Tistarelli y Sandini, 1993) Tistarelli, M. y Sandini, G., «On advantage of polar and log-polar mapping for direct estimation of time-to-impact from optical flow», IEEE Trans. on Pattern Analysis and Machine Intell., 15 (4), pp. 401-410, 1993.
- (Camacho et al, 1996) Camacho, P., Arrebola, F. y Sandoval, F., «Shifted fovea multirresolution geometries», IEEE Int. Conf. on Image Processing (ICIP'96), 1, pp. 307-310, Lausana-Suiza, 1996.
- (Urdiales, 1999) Urdiales, C., Arquitectura de control de movimiento y exploración para un agente autónomo, Tesis Doctoral, Dpto. Tecnología Electrónica, Universidad de Málaga, Málaga-España, 1999.
- (Bandera, 1994) Bandera, C., Structures and algorithms for foveal machine vision, Amherst Systems, Technical Report, Buffalo-EEUU, 1994.
- (Arrebola, 1998) Arrebola, F., Sistemas de visión basado en imágenes multirresolución de fóvea desplazable, Tesis Doctoral, Dpto. Tecnología Electrónica, Universidad de Málaga, Málaga-España, 1998.
- (Haralick y Shapiro, 1992) Haralick, R.M. y Shapiro, L.G., Computer and robot vision, Addison-Wesley: New York, 1992.
- (Pitas, 1993) Pitas, I., Digital image processing algorithms, Prentice Hall: New York, 1993.
- (Bandera et al, 2000) Bandera, A., Urdiales, C., Herreros, J.L. y Sandoval, F., «Implementación de un mecanismo atencional sobre un agente autónomo móvil», Actas del XV Simp. Nacional de la Unión Cient. Int. de Radio (URSI'00), pp. 323-324, Zaragoza-España, 2000.
- (Freeman y Davis, 1977) Freeman, H. y Davis, L.S., «A corner finding algorithm for chain-coded curves», IEEE Trans. on Computers, C-26, pp. 297-303, 1977.
- (Bandera et al, 2000b) Bandera, A., Urdiales, C., Arrebola, F. y Sandoval, F., «Corner detection by means of an adaptively estimated curvature function», Electronics Letters, 36 (2), pp. 124-126, 2000.



RECONOCIMIENTO DE VOZ MULTIDIALECTAL ESPAÑA - COLOMBIA

Mónica Caballero Galeote, Asunción Moreno Bilbao

Departamento de Teoría de la Señal y Comunicaciones
Centre de Tecnologies Aplicades al Llenguatge i la Parla (TALP)
Universidad Politécnica de Cataluña
almcg@gps.tsc.upc.es, asuncion@gps.tsc.upc.es

1. INTRODUCCIÓN

El idioma español se encuentra entre los idiomas más hablados en todo el mundo. Se habla en España y en toda Latinoamérica, exceptuando Brasil por más de 300 millones de personas. También es verdad que debido a la gran dispersión geográfica de estas zonas, el español de cada zona difiere y se establecen variedades dialectales.

Es por eso, que al plantearse un sistema de reconocimiento del habla para el castellano, parece interesante no limitarse a un solo dialecto, sino intentar crear un sistema multidialectal.

Más aún si se observa que para el castellano canónico o estándar, el hablado en España, se tienen bases de datos de habla de 1000 y 4000 locutores, que son bastantes recursos. Pero hay muchos dialectos del español de Latinoamérica de las que sólo se tienen pequeñas bases de datos, como es el caso del dialecto hablado en Colombia. Lo que se plantea en este proyecto es el poder aprovechar todos los conocimientos y datos que se puedan extraer de una gran base de datos para mejorar sistemas de reconocimiento del castellano hablado en otras zonas, con otras características fonéticas, con diferencias léxicas, pero a fin de cuentas, el mismo idioma.

Este estudio es una aproximación al sistema multidialectal que se plantea y se trabaja con un solo dialecto latinoamericano, el hablado en Colombia y el castellano hablado en España.

En el capítulo 2 se describirá el sistema de reconocimiento que se ha utilizado, así como también se tratará del tema relacionado con la fonética, la problemática y la resolución de la transcripción para atacar al sistema de reconocimiento. En el capítulo 3 se dan más detalles de las bases de datos que se han utilizado y finalmente en el capítulo 4 se describen los experimentos que se han realizado y los resultados obtenidos. En el capítulo 5 se comentan las conclusiones y algunas líneas de trabajo futuras.

2. SISTEMA DE RECONOCIMIENTO

2.1. Descripción del problema fonético

La representación de las variaciones dialectales es muy importante para el buen funcionamiento de un sistema de reconocimiento. En este caso, nos encontramos

delante de una variación del español hablado en Colombia. Se ha hecho un estudio de las reglas de transcripción grafema-fonema del español de Bogotá, la zona más poblada de Colombia, dónde se habla el dialecto propio de las montañas o andino, así como también del castellano estándar hablado en España [1].

Las transcripciones se han realizado en símbolos SAMPA (Speech Assessment Methods Phonetic Alphabet) y algunos símbolos de SAMPA extendido para representar sonidos típicos colombianos, como el caso de /h/.

Las reglas para la transcripción del colombiano se basan en la transcripción del castellano estándar. Así, respecto a éste caracterizan el dialecto colombiano los siguientes fenómenos:

- Pronunciación de /T/ siempre como /s/, el “seseo”.
- La /L/ siempre se pronuncia como /jj/, el “yeísmo”.
- La velar fricativa /x/ se realiza como la glotal /h/.
- Los sonidos /b/, /d/ y /g/ se pronuncian siempre como consonantes oclusivas excepto cuando se encuentran en posición post-nuclear o en el inicio de una sílaba siguiendo a una vocal. En estos casos se transcriben como los alófonos aproximantes /B/, /D/ y /G/.

Las vocales en castellano representan aproximadamente un 50 % del total de alófonos y éstas no presentan grandes diferencias de pronunciación en los diferentes dialectos.

Dadas estas características se toman estas decisiones a la hora de transcribir:

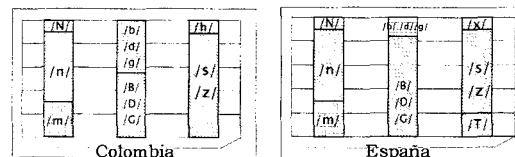


Figura 1. Proporción de los fonemas nasales /m/, /n/ y /N/, de las consonantes /b/, /d/ y /g/ y las aproximantes /B/, /D/ y /G/ y de la presencia de los fonemas /s/ +/z/, /t/, /x/ y /h/ en los dialectos castellano y colombiano.

Para el castellano: En castellano casi no se realiza la /N/. Por lo que se asimila la /N/ a la /n/. Igual pasa con los sonidos /b/, /d/ y /g/ que se asimilan a los alófonos aproximantes /B/, /D/ y /G/. También se considera que los sonidos /L/ y /jj/ están demasiado cercanos y se decide asimilarlos a uno común que se etiqueta como /y/.

Para el colombiano: Se transcribe siempre la /T/, la /L/ y la /x/ como /s/, /j/ y /h/. En cambio se diferencian /b/, /d/ y /g/ de /B/, /D/ y /G/. Y al igual que en el castellano se asimila la /N/ a la /n/.

2.2. Descripción general del sistema

En este estudio se ha trabajado con un sistema de reconocimiento desarrollado por el grupo de procesado del habla del centro de Tecnologies aplicadas al llenguatge i la parla (TALP) de la UPC llamado *RAMSES* [2]. Este sistema se basa en el entrenamiento de modelos ocultos de Markov de unidades fonéticas. Este tipo de sistemas se presentan como unos de los más robustos debido a su capacidad de retener información estadística de los patrones de voz.

La estructura general de este tipo de sistemas es la que se presenta en la figura 2, a continuación:

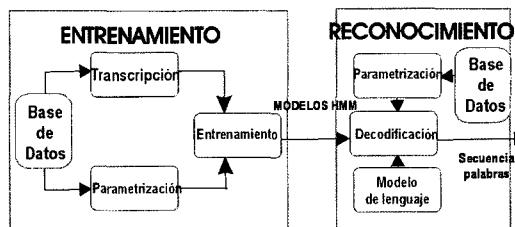


Figura 2. Sistema de reconocimiento del habla.

Ramses está compuesto por una serie de programas que permiten el entrenamiento de modelos de Markov continuos y semicontinuos, (en este estudio se crearan sólo modelos semicontinuos), y el reconocimiento de señal de voz a partir de estos. Ramses permite trabajar con un gran tipo de unidades fonéticas: palabras, fonemas, semifonemas incontextuales y semifonemas por umbral. Para obtener una primera aproximación al programa se trabaja con fonemas. El paso a semifonemas después resulta más fácil.

Debido a que estos programas ofrecen una gran variedad de opciones, existen unos scripts, que facilitan el uso de estos programas en los casos más habituales.

La forma en que se ha utilizado Ramses en este proyecto se explicará a continuación.

- Transcripción :** Creación de ficheros de marcas que contienen la segmentación en unidades fonéticas, ya sean sólo en palabras (es el caso de las señales de test) o en palabras y fonemas (caso de las señales que participarán en el entrenamiento). Los ficheros de marcas se construyen a partir de los ficheros existentes en la base de datos junto a las señales que contienen la transcripción ortográfica de la señal e información del locutor.

Es posible personalizar la transcripción inicial de forma que se adapte mejor a un propósito determinado. Sólo se ha de añadir un diccionario de sustituciones de las unidades fonéticas. Por ejemplo, no se van a crear modelos diferentes de las vocales tónicas respecto a las átonas. Al crear las marcas se sustituyeron todas las

vocales acentuadas por las vocales sin acentuar. De esta forma se pueden asimilar fonemas de los cuales no se quieren calcular modelos.

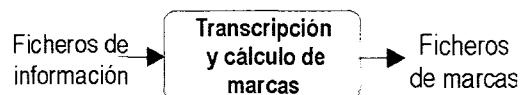


Figura 3. Transcripción.

- Parametrización, cálculo del codebook y cuantificación:** La extracción de características consiste en una estimación espectral de la señal. Se usa una parametrización mel-cepstrum [3]. De esta parametrización se obtienen unos coeficientes que se han de cuantificar vectorialmente, por lo que se necesita calcular un diccionario de vectores significativos o codebook. Se calcula un codebook de 128 coeficientes o vectores diferentes y a continuación se cuantifica.

También se puede realizar la cuantificación con estructura pipeline evitando así almacenar los datos intermedios si ya se tiene el codebook calculado.

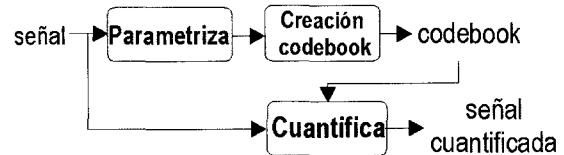


Figura 4. Parametrización, cálculo codebook y cuantificación.

- Entrenamiento de modelos de Markov:** Para evitar que partes de señal donde hay silencios contribuyan al modelo, hay que determinar los tiempos donde se encuentran los silencios, tanto el inicial y el final, como los posibles silencios que se puedan encontrar en medio de la señal. Para ello se utiliza un programa que implementa el algoritmo de Viterbi para reconocimiento del habla, es decir busca el camino más probable. Se le llama con un fichero de arquitectura que le indica el grafo de búsqueda. Según la arquitectura con la que se le llame, varía su funcionalidad. En el entrenamiento de modelos se utilizará para alinear las señales. Se alinean las palabras que forman la señal de forma temporal, permitiendo la inserción de silencios en medio de la señal. Se presentan aquí dos posibles situaciones que varían el esquema en que se entrena los modelos de Markov.

- Si se parte de cero, se crean unos modelos iniciales que contendrán partes de silencio y se alinean las frases. Con dos iteraciones de este proceso, se puede considerar que los silencios ya quedan bien definidos. Finalmente, se desprecian estos modelos y se entrena los modelos finales.



Figura 5. Det. de silencios y entrenamiento hmm's.



2) Si ya se tienen unos modelos iniciales y lo que se pretende es reentrenarlos, entonces basta con un aliñeador inicial de las señales del entrenamiento y entrenar los modelos finales que se utilizarán en el reconocimiento.

- **Modelado del lenguaje:** Es aquí donde se explota el conocimiento sobre el lenguaje. La forma en que utiliza aquí Ramses viene determinada porque la tarea a reconocer es un cuerpo finito de palabras. A nivel léxico, se construye un diccionario con la transcripción de las palabras a reconocer en las unidades fonéticas de las cuales se han creado los modelos de Markov. Así “casa” tiene la representación *k a s a*. En la transcripción no se diferencian las unidades tónicas de las átonas. Como la estructura del mensaje a reconocer es sencilla, con un autómata de estados finitos (FSA, gramáticas N-1 grama) basta para modelar la sintaxis del mensaje. La gramática se genera a partir de un fichero de texto donde han de estar listadas las palabras a reconocer. El script que calcula el diccionario lo puede hacer a partir de un corpus, de un fichero de texto o de una gramática, tal como se ha utilizado.



Figura 6. Modelado del lenguaje.

- **Reconocimiento:** Una vez entrenados los modelos (hmm's) y creada ya una gramática y un diccionario, ya podemos pasar a esta etapa, también denominada decodificación.

En primer lugar se han de cuantificar vectorialmente las señales a reconocer. Después se ha de determinar la secuencia de palabras que con mayor probabilidad ha generado una secuencia de observaciones. Para esta labor se utiliza el algoritmo de Viterbi.. El grafo de búsqueda, o lo que es lo mismo, la arquitectura, se construye de forma jerárquica: en el nivel más alto, la gramática, que indica las palabras a reconocer y las relaciones entre ellas, por debajo se tiene la representación de las palabras en unidades fonéticas y por último, se dispone de hmm's que relacionan las observaciones acústicas con los fonemas.

Ramses, además dispone de scripts que calculan las tasas de aciertos y que muestran que señales no se han reconocido.

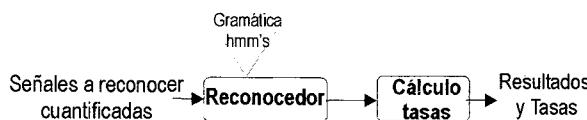


Figura 7. Reconocimiento y evaluación.

3. BASES DE DATOS

Para este estudio se han utilizado 2 bases de datos, la base SpeechDat para castellano y la base de locutores colombianos que se incluye en el proyecto SALA [4].

El proyecto SALA consiste en la creación de bases de datos de los principales dialectos del castellano en Latinoamérica. Los criterios para la creación de estas bases, entre la que se encuentra la del colombiano son los mismos que los que se siguieron para la creación de SpeechDat en España.

Las especificaciones que debían seguir estos corpora son las siguientes:

- Cada locutor pronuncia más de 40 frases, leídas o espontáneas.
- La parte del corpus de habla leída contiene palabras de aplicación, frases con palabras clave, números, dígitos, fechas, expresiones relativas al tiempo, nombres, fechas, empresas, frases y palabras fonéticamente balanceadas.
- La parte del material espontáneo incluye preguntas si/no, ciudades, horas y fechas.

A parte de estas especificaciones más relacionadas con el contenido del corpus existen otras que también se han tenido en cuenta en las dos bases, tales como:

- Factores dialectales
- Características del locutor: edad, sexo
- Características específicas del entorno.

Ambas bases se han grabado sobre la red telefónica fija, mediante una línea RDSI.

Las señales quedan almacenadas en ficheros tras ser muestreadas a 8 khz, cuantificadas con 8 bits y comprimidas con la ley A. Cada señal se guarda en un fichero distinto. A cada fichero le acompaña otro con información del locutor y la transcripción. Esta transcripción se ha hecho con los símbolos SAMPA.

La base SpeechDat está compuesta por grabaciones de 4000 locutores.

La base colombiana englobada dentro del proyecto SALA está compuesta por grabaciones de 1000 locutores.

Ambas bases tienen especificados un conjunto de señales para el test. En el caso de SpeechDat se dedican 500 locutores para el set de test, en el caso de la base colombiana son 200 locutores los que componen este conjunto de test.

4. EXPERIMENTOS Y RESULTADOS

Se han llevado toda una serie de experimentos a fin de mejorar la tasa de reconocimiento tanto en el castellano, como en el colombiano.

4.1. Ficheros entrenamiento y test

Se pretende evaluar el sistema reconociendo palabras de aplicación y dígitos aislados. Para éste propósito se decide entrenar con frases fonéticamente balanceadas y ampliar el entrenamiento con palabras fonéticamente balanceadas.

Entrenamiento. En Colombia se tienen 2700 frases pertenecientes a 800 locutores y 1550 palabras, también pertenecientes a los 800 locutores. En castellano, se dispone de un total de 15156 frases y 9174 palabras pertenecientes a 3500 locutores.

Test. Se diseñan dos tests, uno para el español y otro para el colombiano con las palabras de aplicación y dígitos aislados. En colombiano se tienen un total de 751 ficheros de test de 200 locutores. En Castellano, aunque se dispongan de 500 locutores para realizar el test, sólo se utilizan 200 locutores, al igual que en el caso del colombiano, para así tener unos resultados comparables. En total se dispone de 626 palabras pertenecientes a los 200 locutores.

En total se entrena un total de 26 fonemas en castellano y de 30 en colombiano y se reconocen un total de 30 palabras de aplicación más los 10 dígitos tanto en castellano como en colombiano.

4.2. Experimentos y resultados

En primer lugar, se evaluó cada sistema por separado. Esto es, entrenar y reconocer con castellano por una parte, y por otra, entrenar y reconocer colombiano. Se ha trabajado con fonemas.

Para el caso del castellano, se partió del mismo número de frases que se disponían en el caso colombiano, para tener un resultado comparable, y posteriormente se fueron añadiendo locutores poco a poco, de modo que se reflejara la mejora al aumentar el material de entrenamiento. Cada vez que se aumentó el número de frases, se usaban como modelos iniciales los que resultaron del anterior entrenamiento. Se pudo comprobar que una vez alcanzados un número suficiente de frases, el sistema no mejoraba notablemente, por lo que se decidió, después de entrenar con 4327 frases añadir a las frases las palabras fonéticamente balanceadas. Esto hizo mejorar la tasa de reconocimiento en más de un punto. Por este motivo se siguió reentrenando el sistema con más locutores, incluyendo además las palabras. Para el test escogido, este incremento de entrenamiento no se notó en el resultado del reconocimiento, quedándose en una tasa de 94,4 %.

En el caso del colombiano, se entrenó con las 2700 frases, dando un resultado de un 94 % de aciertos, resultado muy bueno, considerando el poco número de frases de las que se disponía. Al añadir las palabras, difiriendo del caso del castellano, el sistema no mejoraba. Se decidió alinear las señales con estos últimos modelos y volver a crear modelos nuevos sólo a partir de las frases. De esta manera, la tasa de reconocimiento aumentó, aunque en menor grado, llegando a un 94,14 %.

Por otra parte, se diseñaron una serie de experimentos de reconocimiento cruzado. A la vez que se iban obteniendo modelos nuevos del castellano, se les aplicaba el test colombiano, observándose una buena evolu-

ción. De esta forma, también se comprobó que el sistema, aunque se quedase estancado con el test castellano, mejoraba al aplicarle el colombiano.

Otra cosa a denotar es que los resultados tan parecidos al del sistema entrenado en castellano y reconociendo en el mismo dialecto, indican que los canales son parecidos.

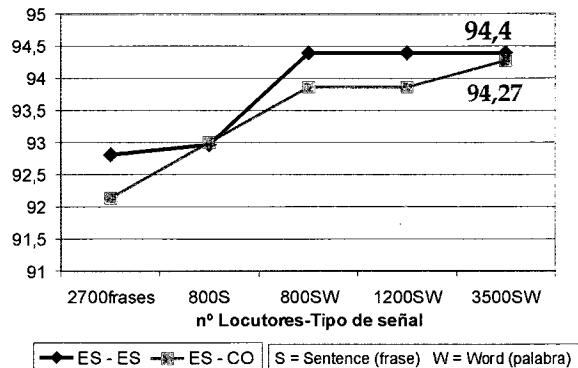


Figura 8. Evolución tasas reconocimiento test castellano ES-ES y test colombiano ES-CO entrenando con base castellana.

5. CONCLUSIONES

Este estudio demuestra que es posible conseguir unas buenas tasas de reconocimiento utilizando los modelos entrenados con la base de datos española. Quedan muchos experimentos por hacer. Los más inmediatos son, trabajar con las dos bases de datos a la vez, trabajar con semifonemas y trabajar más con la transcripción fonética. Sería interesante separar también en castellano los modelos de las oclusivas b, d y g para tener modelos separados para aplicar al sistema el test colombiano.

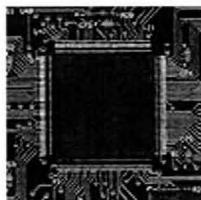
6. AGRADECIMIENTOS

Este trabajo ha sido subvencionado por el proyecto TIC-2000-1005-C03.

7. REFERENCIAS

- [1] A. Moreno, J. B. Mariño. *Speech Dialects: Phonetic Transcription*. International Conference on Spoken Language Processing. Sidney, Australia. Dec, 1998.
- [2] A. Bonafonte, J. B. Mariño, A. Nogueiras, J.A. Rodríguez Fonollosa. *RAMSES: el sistema de reconocimiento del habla continua y gran vocabulario desarrollado por la UPC*, VIII Jornadas de Telecom I+D (TELECOM I+D'98), Madrid.
- [3] J.W. Picone *Signal Modeling Techniques in Speech Recognition*. Proceedings of the IEEE Vol 81 no.9. September 1993
- [4] A. Moreno, H. Höge, J. Koehler, J. B. Mariño. *Project SALA, SpeechDat Across Latin America*. First International Conference on Language Resources and Evaluation, LREC'98, Granada.





BIST: RESPUESTA A LOS NUEVOS DESAFÍOS PARA EL TEST DE SISTEMAS EN CHIP (I)



Jorge-Luis Sánchez-Ponz

Institut für Informatik - Abteilung Rechnerarchitektur
Universität Stuttgart, Deutschland
IEEE Region 8 GOLD Coordinator
E-mail: j.s.ponz@ieee.org

Resumen - Las técnicas BIST posibilitan que un circuito integrado realice su propio test por sí mismo. BIST reduce los costes de test y mantenimiento para sistemas en chip gracias a la eliminación de costosos equipos externos de test y permitiendo la localización de circuitos defectuosos dentro de sistemas. BIST puede llevar a cabo el test a la frecuencia normal de funcionamiento del diseño, lo que es fundamental a la hora de detección de fallos relativos a la temporización. A pesar de estas ventajas, BIST ha tenido y tiene un uso muy limitado en la industria porque añade más área al circuito e incrementa el tiempo de diseño.

Palabras clave - Sistemas en chip, Test, ASIC, Propiedad Intelectual, DFT, ATPG, BIST.

I. INTRODUCCIÓN Y MOTIVACIÓN

Los tradicionales sistemas imbuidos (embedded systems) eran típicamente aplicaciones industriales construidas con un sencillo microcontrolador como principal funcionalidad electrónica. Ahora, estos sistemas están siendo rápidamente introducidos en un amplio abanico de aplicaciones y producidos en inmensas cantidades. Más aún, la complejidad de los sistemas imbuidos se incrementa de forma pareja con los productos electrónicos en general, siguiendo la conocida Ley de Moore (Figura 1).

Considérese lo que está ocurriendo. En lo referente al circuito integrado (chip), el tamaño de los elementos componentes se reduce alrededor de un 30% cada tres años. Pero al mismo tiempo el tamaño del circuito en sí aumenta un 12% al año (Figura 2), mientras que cada año se inician más del 50% de los nuevos diseños. Esta situación ya supone un problema, pues las diferencias en la productividad de los diseños se acrecenta de manera imparable.

Ejemplos de nuevas aplicaciones pueden encontrarse en la industria automovilística, imagen digital, productos de electrónica de consumo, productos industriales, redes de comunicaciones, seguridad, almacenamiento de información y comunicaciones inalámbricas [2].

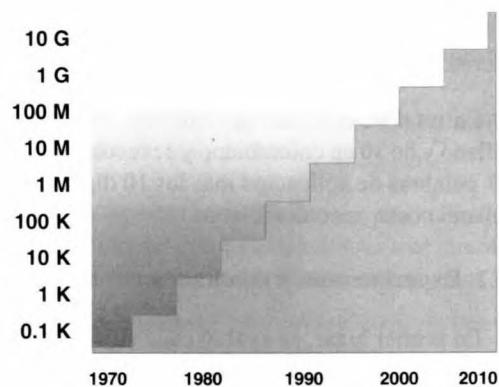


Figura 1 - Ley de Moore aplicada a las memorias (el tamaño se multiplica por cuatro cada tres años).

La gran mayoría de estos productos tienen fuertes restricciones y limitaciones respecto al coste, tamaño, consumo de potencia y rendimiento, por nombrar algunos de los parámetros más importantes con impacto en la competitividad y en el éxito de los diseños. Sin embargo, todos estos parámetros se encuentran correlados con el grado de integración. De ahí, el hecho de que sea tan importante sacar ventaja de las nuevas tecnologías y desarrollar metodologías que soporten la integración.

Los nuevos productos integrados tienden a presentar un ciclo de vida muy corto, a la vez que su complejidad aumenta. Esta circunstancia ya se anticipa como una emergente crisis en el diseño, i.e. la dificultad para enjaezar la complejidad y disminuir el tiempo hasta la comercialización. A fin de afrontar esta situación, un nuevo paradigma de diseño se está desarrollando, principalmente evolucionando desde la composición de diseños a base de bloques constitutivos a muy bajo nivel hacia la reutilización de bloques muy complejos, bien combinando diseños disponibles comercialmente y ASIC (Application Specific Integrated Circuit, Circuito Integrado de Aplicación Específica), o bien usando bloques externos de Propiedad Intelectual (Intellectual Property, IP) con funciones estándar y lógica específica desarrollada en la compañía y componiendo finalmente sistemas en chip (System-on Chip, SoC), véase Figura 3.

El advenimiento de sistemas en chip ha agravado la situación. La complejidad de los dispositivos también se incrementa con la proliferación de diseños basados en SoC. Estos diseños con un elevado número de puertas pueden contener elementos tales como uno o varios microprocesadores, buses, periféricos, secciones basadas en ASIC, software integrado y ya compilado, bloques de propiedad intelectual, bloques reutilizables, diferentes dominios de reloj, diferentes dominios de frecuencia, relojes generados en el propio chip, memorias de diverso tipo y componentes analógicos para el tratamiento de la señal. Los bloques de propiedad intelectual pueden, a su vez, proceder de una gran variedad de empresas, que, a menudo, proporcionan escasos detalles sobre los diseños, lo que complica enormemente el test de los mismos. El nuevo paradigma del diseño resolverá la crisis del diseño, pero nuevas metodologías y herramientas son necesarias para mejorar la productividad del diseñador y del ingeniero en este nuevo contexto.

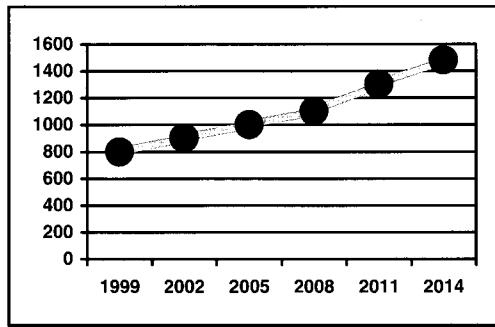


Figura 2 - Tamaño del chip (mm²) en Sistemas en Chip [1].

El problema del test ha crecido, en la misma medida que los circuitos integrados se hacían más rápidos (Figura 4) y complejos (Figura 5). Equipos de test para altas frecuencias cuestan elevadas sumas de dinero (varios millones de dólares), y el acceso físico a la circuitería interna se hace más difícil pues el número de patillas de entrada/salida que estos equipos pueden manejar no crece al mismo ritmo que la lógica contenida en los sistemas en chip.

El desfase creciente entre estos costosos equipos de test y los diseños actuales es alarmante. En la próxima década, la precisión de estos equipos de test pasará de los 200 picosegundos a los 175 picosegundos, según Mentor Graphics [3]. Sin embargo, la frecuencia de reloj de los microprocesadores se incrementará hasta los 3 GHz (330 picosegundos) [3]. Esto provocará alrededor de un 50% de discrepancia entre el rendimiento del chip y las capacidades del equipo automático de test (Automatic Test Equipment, ATE). Hace apenas unos años las imprecisiones eran de apenas un 10%. Más aún, las pérdidas de chips en producción

debido a imprecisiones de los equipos de test podrían llegar a alcanzar el 50% en la próxima década. En los nuevos diseños, el número de chips considerados como no defectuosos podría llegar a ser inferior al 20% de los producidos.

Visto que los equipos externos de test no son capaces de llevar a cabo las labores de test, ¿no sería conveniente transladar algunas de las tareas de test al propio chip? Esta idea es la que ha originado el concepto conocido como diseño para test (Design for Test/Testability, DFT). Un término bastante general, que engloba a todas las técnicas que consideran al dispositivo a la hora de colaborar en la realización del test del mismo.

La nueva forma de construir sistemas embuidos tiene un impacto en las diferentes dimensiones del test. Primero, nuevos protocolos y tecnologías de diseño incorporan igualmente nuevos tipos de posibles fallos en producción y durante operación, y también una diferente distribución de las tasas de fallos para los diversos tipos. Esto reclama un análisis con nuevas herramientas para comprender la implicación sobre las técnicas de test y su efectividad. Existe una necesidad de nuevos medios para evaluar los nuevos métodos de test, que están siendo desarrollados para adaptarse a las nuevas tecnologías y a los requerimientos de las nuevas prácticas de diseño.

En segundo lugar, muchos de los sistemas imbuidos tienen fuertes requerimientos bien en seguridad del producto (por ejemplo, en funciones de control de vehículos como los sistemas de frenado antibloqueo, ABS), o bien en rendimiento y disponibilidad del producto para aumentar la competitividad (véase como ejemplo cualquiera de los productos de electrónica de consumo). Estos requerimientos poseen en común la demanda de auto-test de los propios sistemas en la fase de inicialización y/o durante la operación del mismo (en servicio). Para solucionar esto y todavía ser capaces de manejar la complejidad y adaptarse a los plazos de comercialización del producto, son indispensables métodos estructurados y herramientas de soporte para el diseño de los elementos de auto-test.

En tercer lugar, el diseño con bloques procedentes de distintas fuentes y en diferentes formatos implica que el diseñador, en su labor de ensamblaje de los bloques, posee un conocimiento muy reducido de la implementación detallada de cada uno de ellos. No obstante, es el diseñador el responsable de la creación de un programa de test para el diseño completo, asegurando la calidad en producción e implementando los elementos de auto-test para el sistema final. Para afrontar este reto, el ingeniero demandará capacidades de auto-test en cada uno de los bloques que va a usar, dispondrá de los bloques con los datos de test indispensables, o, en última instancia, poseerá la suficiente

información (procedente del distribuidor del bloque) para generar los datos de test por sí mismo con la ayuda de nuevas herramientas.

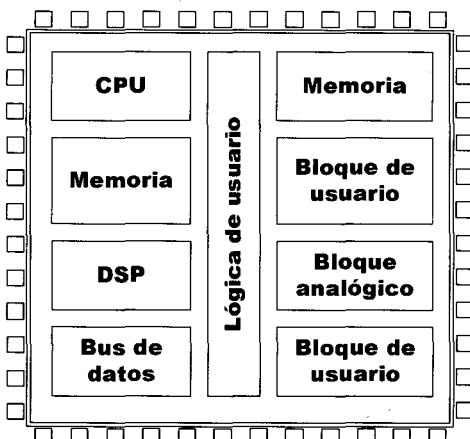
Finalmente, la capacidad para llevar a cabo test de alta calidad en etapas tempranas en el proceso de producción, minimiza la cantidad de basura y residuos generados, lo que es importante a efectos medioambientales.

No cabe duda de que muchos fabricantes de semiconductores son capaces de producir sistemas en chip. Pero, enseguida surge la pregunta: ¿pueden estos fabricantes realizar el test de producción de una manera eficiente y a un precio razonable? Preocupaciones a este respecto han estado suscitando el interés en el auto-test incorporado (Built-In Self-Test, BIST), una propuesta en la que un chip realiza todas las funciones de test por sí mismo.

II. ALTERNATIVAS DISPONIBLES EN DFT PARA LA PROPIEDAD INTELECTUAL: BIST

En estos tiempos de creciente complejidad y de diseños basados en sistemas en chip con varios millones de puertas, junto con equipos de diseño multinacionales y estrategias de fabricación encaminadas a alcanzar el mercado en el menor tiempo posible, es evidente la necesidad de nuevos métodos que aseguren que los nuevos circuitos integrados puedan ser verificados con altos niveles de calidad y con tiempos de test reducidos y económicos.

No obstante, esto demanda que se produzcan en el próximo futuro una serie de cosas. En primer lugar,



BIST es la última de las técnicas DFT que ha aparecido. Nace, en parte, como respuesta al coste de los equipos automáticos de test, una industria cuyos ingresos anuales están creciendo hasta alcanzar los 6.000 millones de dólares. BIST ofrece al usuario una oportunidad para reducir el volumen de los gastos en estos equipos. Mediante el uso de herramientas DFT, es posible que cualquier empresa pueda comprar un equipo de test para cuatro patillas por 50.000 dólares, en lugar de un complejo y costoso equipo por varios millones de dólares (si no se usarán las citadas herramientas).

Algunos observadores sostienen que BIST está cobrando mayor importancia como elemento crítico a la hora de solucionar el enigma del test. Los ingenieros comienzan a darse cuenta de que el test se está convirtiendo en un gran problema. Los grandes equipos de test son ya muy caros y son insuficientes a la hora de realizar el test, por lo que en muchos casos son imprescindibles tres o cuatro equipos. Es evidente, por tanto,

El BIST para lógica es también otra alternativa, aunque para conseguir los niveles de cobertura de fallos requeridos hoy en día se exigen diseños con lógica adicional de inspección para permitir que la circuitería BIST sea realmente efectiva.

Hughes Aircraft ha estado usando técnicas BIST desde 1987. Su principal negocio es la fabricación de sistemas con alta tolerancia a fallos que van en los principales aviones comerciales y en los transbordadores espaciales. Incluso durante las misiones de estos últimos es indispensable la verificación continua del funcionamiento correcto de cada parte, hecho que obviamente no puede realizarse con los equipos tradicionales de test pues no forman parte de ningún transbordador espacial o de avión comercial alguno.

Otras empresas implementaban en el pasado el BIST «a mano», pero en la actualidad usan herramientas comerciales para llevar a cabo estas tareas. En igualdad de condiciones, BIST da mejores niveles de

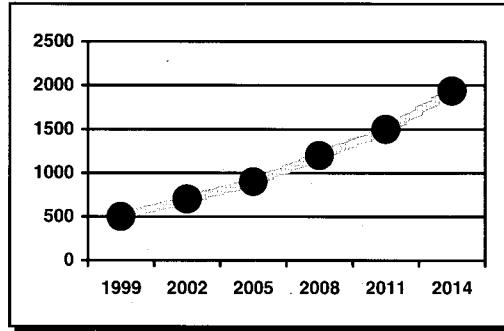


Figura 4 - Frecuencia de funcionamiento (MHz.) de Sistemas en Chip [1].

que la estructura de test tendrá que ir integrada en el propio diseño.

El número de personas con capacidad de decisión en este campo va aumentando progresivamente. Cuando el diseñador se enfrenta a un complicado diseño, tiene que desarrollar métodos para reducir el coste desde la perspectiva del tiempo de test. Uno de los métodos disponibles es BIST: los circuitos realizan su propio test por sí mismo, bien en serie o en paralelo. Incluso algunas compañías usan métodos combinados: BIST y equipos de test para señales analógicas, tales como las señales analógicas procedentes de chips usados en el campo de las comunicaciones.

Las memorias imbuidas, por ejemplo, deben ser consideradas en términos de tamaño y colocación dentro del chip. Pequeños bloques de memoria pueden ser a menudo verificados de manera efectiva usando los recursos del procesador, que se encuentra dentro del mismo chip. Sin embargo, grandes bloques se pueden verificar mucho mejor usando BIST específico para esas memorias.

Algunas compañías dedicadas a la automatización del diseño electrónico (Electronic Design Automation, EDA), están proponiendo BIST como solución a los problemas de testabilidad

calidad que los tradicionales sistemas de test, a un coste inferior.

Sin embargo, la necesidad de complejos equipos de test todavía no ha desaparecido, ya que la gran mayoría de los diseños de los principales suministradores de ASIC no usan BIST. Otras compañías señalan que les gustaría tener herramientas de inserción de BIST mucho más automatizadas para realizar el test tanto a nivel de placa de circuito impreso como a nivel de sistema. Pero esta circunstancia no es un problema. Usando un ASIC o una FPGA (Field Programmable Gate Array), la técnica de BIST puede ser generalizada a nivel de placa de circuito impreso o sistema, permitiendo control y verificación continua. Es sólo cuestión de tiempo.

El hecho de que BIST se haya limitado a aplicaciones críticas o productos con tolerancia a fallos, no debe ser significativo. Es indudable que la llegada de sistemas en silicio va a cambiar esta percepción.

* EDA



Algunas compañías dedicadas a la automatización del diseño electrónico (Electronic Design Automation, EDA), están proponiendo BIST como solución a los problemas de testabilidad. Con esta técnica, la funcionalidad del equipo de test se incorpora, como un elemento más, en el chip durante la fase de diseño. Un equipo de test externo de muy bajo coste genera las señales para iniciar el test (a nivel interno) del chip y, una vez finalizado, leer los resultados del test.

En la actualidad se está poniendo mucho énfasis en la simulación estática y en la simultaneidad de la síntesis y de la disposición física y conexiónado de los diferentes bloques del diseño. Esta síntesis tan compleja, en un sólo paso, parece ser la piedra filosofal del negocio de la industria de EDA. Pero, ¿qué beneficios aporta esta elegante forma de síntesis y obtención del producto para su fabricación masiva, si no puede realizarse un test a un nivel aceptable de calidad para que los clientes lo usen con un cierto nivel de confianza?

DFT y BIST suponen sólo un tres por ciento de los ingresos de la industria de EDA. Y eso, considerando que los costes de desarrollo de test son el treinta y

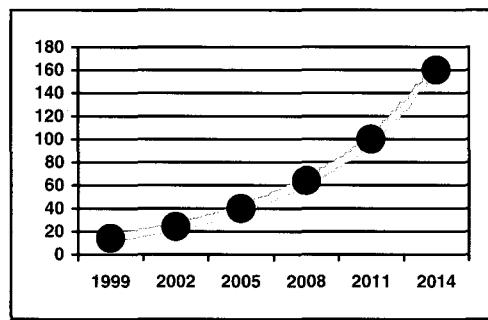


Figura 5 - Número de transistores (millones) en Sistemas en Chip [1].

cinco por ciento (o más) de los costes totales de desarrollo del producto, y que son los que en mayor medida contribuyen al alargamiento de los ciclos de desarrollo y verificación del mismo.

Uno de los máximos exponentes y líderes en el mercado de las herramientas de BIST es LogicVision, Inc. [4]. Según LogicVision: «Con frecuencias de reloj de más de 200 MHz., y en claro proceso de incremento, es muy difícil que los equipos externos de test puedan ser capaces de realizar un test a la frecuencia de funcionamiento del circuito. En consecuencia, la mayoría de los diseñadores se ven obligados a aceptar el test a frecuencias bastante inferiores a las de funcionamiento, lo que permite que ciertas partes defectuosas del diseño pasen inadvertidas».

Intentos para combatir esta indetectabilidad se han llevado a cabo, combinando técnicas de inspec-

ción (scan techniques) con generación automática de patrones de test (ATPG), pero con escaso éxito. LogicVision señala que sus herramientas son capaces de descubrir fallos que ATPG o inspección no descubrirían.

El coste de los equipos automáticos de test externo se está elevando, tal y como se indicó con anterioridad, si bien estos equipos no son capaces de realizar el test de los chips, que cada vez se complican más,

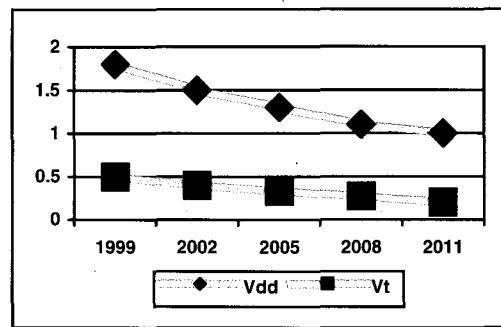


Figura 6 - Tensión de alimentación (voltios) en Sistemas en Chip [6].

tendencia que se incrementará de manera más notable cuando los diseños pasen de tecnologías de 0,18 micras a 0,13 micras.

Las puertas adicionales usadas para el controlador de BIST (de 1.000 a 3.000, según el caso) incluido en el propio chip, se consideran una nimiedad, en una era donde los diseños de centares de miles de puertas son ya algo muy habitual, aduce LogicVision. Asimismo, argumentan que las memorias RAM dinámicas (DRAM) con el tiempo usarán BIST, probablemente a niveles de 256 bits, y motivados por cuestiones de ancho de banda.

Al usar BIST, sus clientes no tendrán que confiar en multimillonarios equipos de test para producciones masivas. LogicVision sostiene que los usuarios pueden usar BIST para labores de mantenimiento, reparación y diagnosis de sistemas, una vez que estos se encuentran en funcionamiento. El mismo test puede ser reutilizado en cualquier momento en la vida del chip, placa o sistema.

El BIST para memorias se ha adoptado mucho más rápidamente que el BIST para lógica, puesto que los diseños de sistemas en chip incluyen ingentes cantidades de megabits de memoria: la necesidad es mucho mayor. Mentor Graphics predice que el BIST para lógica se hará más común en un futuro muy cercano, motivado por el movimiento de las compañías de diseño hacia la propiedad intelectual (IP) [3].

Estas compañías lo que necesitan es alguien y algo que les ayude en la metodología de test. La

densidad, velocidad y características de bajo consumo (Figura 6) de las nuevas tecnologías de fabricación (muy por debajo de las micras) hacen de las técnicas de diseño para test (DFT) un elemento muy importante.

No todas las empresas de EDA se han subido al vagón del BIST. Viewlogic Systems Inc. [5] se ha centrado en la tecnología de inspección y ATPG, que es «una propuesta y aproximación mucho más flexible». Algunas de ellas remarcan que BIST tiene su lugar en el mercado y que está creciendo, pero que se ha limitado a nichos de mercado muy específicos como complejos equipos de conmutación telefónica y ordenadores con tolerancia a fallos, grandes sistemas con requerimientos funcionales para realizar su propio test a la vez que están totalmente operativos y dando servicio. Añaden que es una falacia que por usar BIST en la lógica, los ingenieros no tengan que necesitar equipos de test para altas frecuencias.

El futuro traerá seguramente el desplazamiento del análisis de DFT y BIST y su implementación en las primeras etapas de diseño, comenzando con la comprobación de las reglas de test en el nivel de transferencia de registros (Register Transfer Level, RTL)

III. PRESENTE Y FUTURO

Ciertas, y bien conocidas, configuraciones y circuitos puede causar estragos en el desarrollo del test y de ATPG. Estos incluyen, entre otros y sólo por citar algunos, señales de set/reset asíncronas, relojes generados internamente y bucles de realimentación secuenciales y combinacionales. Y si bien es cierto que estos problemas de testabilidad pueden encontrarse y corregirse en las etapas posteriores a la síntesis, sería mucho más productivo prevenirlos antes de entrar en el proceso de síntesis y sin incurrir en serios riesgos de iteraciones de diseño.

El futuro traerá seguramente el desplazamiento del análisis de DFT y BIST y su implementación en las primeras etapas de diseño, comenzando con la comprobación de las reglas de test en el nivel de transferencia de registros (Register Transfer Level, RTL). Este movimiento a niveles más abstractos del

diseño, aliviará la dificultad de encontrar los problemas relativos a DFT demasiado tarde en el ciclo de diseño del producto y proporcionará a los ingenieros las herramientas necesarias para encontrarlos y prevenirlas, incluso antes de que los ingenieros de test lleguen a descubrirlos.

De forma que la única forma de enfrentarse a los problemas de test en un futuro, que ya está aquí, es abandonar la mentalidad de que el test es algo que se realiza al final y adoptar la mentalidad de incorporarlo a las primeras consideraciones de diseño, donde se plantean cuestiones vitales como tamaño del chip, número de patillas, consumo de potencia y frecuencia de funcionamiento.

IV. CONCLUSIONES

Se ha presentado y descrito la problemática situación existente en la actualidad en lo que respecta al mundo del test en el campo de los circuitos integrados, con especial énfasis en los sistemas en chip. Nuevas técnicas y métodos, especialmente aquellos que se centran en el auto-test, se alzan como una de las más importantes tecnologías para una explotación competitiva de los sistemas en chip en un amplio espectro de productos, basados en sistemas imbuidos.

Este artículo justifica la necesidad de considerar DFT a la hora del diseño de sistemas en chip y, por extensión, de circuitos integrados en general. En el próximo artículo, se presentarán los fundamentos de BIST, tanto para memoria como para lógica, sus principales ventajas e inconvenientes.

AGRADECIMIENTOS

A los ingenieros de ARM Ltd. por su ayuda y consejo.

REFERENCIAS

- [1] SIA, EECA, KSIA, EIAJ, and TSIA, «International Technology Roadmap for Semiconductors». 1998.
- [2] www.arm.com
- [3] www.mentorg.com
- [4] www.logicvision.com
- [5] www.viewlogic.com
- [6] Taur, Y. (T. J. Watson Research Center, IBM), «The Incredible Shrinking Transistor», IEEE Spectrum, July 1999.





SYMBOLATOR: UN SIMULADOR SIMBÓLICO DE CIRCUITOS PARA CALCULADORA

Roberto Pérez Franco

perezfranco@yahoo.com

Rama Estudiantil: Universidad Tecnológica de Panamá

Procedencia: Ciudad de Panamá, Panamá

EXTRACTO

El proyecto Symbolator consiste en crear un simulador simbólico de circuitos lineales poderoso, capaz de correr en una calculadora, para ayudar a los estudiantes de circuitos con la parte matemática del análisis a lo largo de sus diferentes cursos, y permitir la solución numérica y simbólica en corriente directa, corriente alterna, análisis transitorio y dominio de la frecuencia de un circuito lineal cuya descripción sea sencilla y reutilizable. Una versión del Symbolator ya ha sido programada con éxito en el lenguaje BASIC de las calculadoras TI-89.

SOBRE LA NECESIDAD DE UN SIMULADOR SIMBÓLICO

Para el ingeniero, la matemática es - más que un objetivo - una herramienta de trabajo. En el análisis de circuitos en ingeniería, los cálculos matemáticos son un medio para llegar a conocer la respuesta de un circuito dado; no son un propósito en si mismos. Cuando un estudiante se inicia en el análisis de nuevos circuitos, debe enfrentar dos retos. El primero y más importante es comprender a cabalidad la naturaleza del nuevo tipo de circuitos que se le presenta, la lógica que yace detrás de su análisis, y sus aplicaciones en la ingeniería. El segundo es dominar los métodos matemáticos que le permitirán estudiar la respuesta de este tipo de circuitos. Es deseable que el estudiante se concentre en lo primero, dejando lo segundo a una máquina asistente.

Cuando se resuelven circuitos a mano, una infinidad de detalles insignificantes, tales como un signo que se pierde de vista o un punto decimal que se coloca mal, pueden hacer la diferencia entre una respuesta correcta y una incorrecta. Los ingenieros profesionales pueden utilizar en su lugar de trabajo un simulador de circuitos numérico en una computadora personal. La mayoría de los estudiantes no

tienen esta opción en el salón de clases debido a limitantes académicas y económicas. En muchos casos los errores en el análisis de circuitos se deben a una mala manipulación matemática de las fórmulas y los valores, producto de ojos cansados o la presión mental propia de un examen universitario.

A inicios de este año me fijé como meta crear el simulador de circuitos lineales más poderoso de cuantos pueden correr en una calculadora, diseñado de tal forma que ayude al estudiante en el salón de clases con la parte matemática del análisis de circuitos a lo largo de sus diferentes cursos. El simulador debía permitir análisis tanto numérico como simbólico, en corriente directa, corriente alterna, análisis transitorio y dominio de la frecuencia, y aceptar una amplia gama de elementos lineales.

TRABAJOS PREVIOS SOBRE EL TEMA

Un paso preliminar en mi proyecto fue perfeccionar los mejores simuladores de circuitos para calculadoras disponibles en ese momento. Lo hice con dos de ellos.

El primero fue el simulador numérico CSim 2.01 para HP48GY, creado por el ingeniero finés Per Stenius en 1991. Tras varias semanas de consultas y muchas horas de programación, presenté el simulador numérico CSim 3.01. Esta versión es más fácil de usar que la versión anterior. Programé paralelamente dos variantes del CSim: una para estudiantes novatos, llamada Assistant II y otra enfocada al análisis de circuitos de potencia, llamada ZAC 1.4. El algoritmo, por desgracia, no se prestaba para aplicaciones simbólicas en la HP48GX, así que abandoné este curso de acción.

El segundo fue el simulador numérico NFAP para HP48GX., creado por Don Robinsons en 1994. Escribí dos variantes simbólicas de este programa: el Buscador Simbólico de Voltajes (o Symbolic Voltage

Finder) y el Buscador Simbólico de Corrientes (o Symbolic Current Finder). No estuve satisfecho: estos simuladores simbólicos funcionaban bien en corriente alterna, pero tenían demasiadas limitaciones y no llenaban el propósito de ser una herramienta completa para acompañar al estudiante de circuitos a lo largo de sus estudios universitarios.

En ese punto supe que era necesario atacar el problema en una forma diferente, con un método y un simulador totalmente nuevo.

PLANTEAMIENTO: EL PROYECTO SYMBULATOR

Llamé a mi proyecto Simulador de Circuitos Simbólicos, y lo apodé Symbulator. Lo visualicé como una herramienta tan poderosa que no fuese comparable con ningún otro simulador antes que él, usando como plataforma una calculadora en el entorno del salón de clases, con las características ya expuestas arriba.

ALTERNATIVAS CONSIDERADAS

Métodos considerados. En cuanto al método de análisis de circuitos que utilizaría para desarrollar el nuevo algoritmo, consideré dos alternativas: el método del análisis de nodos (simple, modificado o con transformada gyrator) y el método de generación de ecuaciones.

Calculadoras consideradas. En cuanto a la calculadora que serviría de plataforma para desarrollar y ejecutar el programa, consideré como alternativas las calculadoras Hewlett-Packard y Texas Instruments existentes en ese momento.

EVALUACIÓN DE LA VIABILIDAD DE ESTAS ALTERNATIVAS

Evaluación de métodos.

El método del análisis de nodos es poderoso para el análisis de circuitos. Sus ventajas se manifiestan básicamente en la simulación de circuitos numéricos. Sus desventajas son la complejidad de la extracción de respuestas del vector final resultante y el difícil manejo de sus matrices en el caso de circuitos simbólicos.

El método de generación de ecuaciones es un método innovador y flexible. Consiste en generar y

resolver ecuaciones que describan los voltajes y corrientes del circuito. Su ventaja principal es la facilidad en la descripción de fuentes dependientes y la exactitud en la simulación tanto simbólica como numérica. Su desventaja es que su programación es mucho más compleja que el método del análisis de nodos.

Evaluación de calculadoras

Los sistemas de álgebra de las calculadoras Hewlett-Packard HP48GX y HP49G no poseen las capacidades de manipulación simbólica y resolución de ecuaciones requeridas por mi simulador. El comando solve() del modelo HP49G es muy débil.

La T189 posee un sistema de álgebra muy poderoso, basado en un programa de computadora de nombre Derive, que permite la manipulación simbólica y el reemplazo automático de expresiones definidas. Posee el comando solve() más poderoso que haya tenido calculadora alguna en la historia.

El comando solve() se usa para resolver ecuaciones no lineales simultáneas.

SELECCIÓN Y JUSTIFICACIÓN

Selección del método. Decidí utilizar el método de la generación de ecuaciones.

Justificación. Es mucho más amigable para el usuario, pues requiere descripciones de circuito mucho más sencillas que el método del análisis de nodos.

Selección de la calculadora. Seleccioné la calculadora T189 para ser la plataforma que daría soporte al simulador.

Justificación. Su software algebraico es el más poderoso. Por ahora, es la única que posee las características necesarias para manejar un programa de esta naturaleza.

PLAN DE ACCIONES

Programé primero una aplicación para análisis de corriente directa con los elementos básicos de circuito: resistor y fuentes independientes. Una vez estuvo lista, me basé en ella para programar las aplicaciones de análisis de corriente alterna, análisis en dominio de la frecuencia y análisis transitorio.



Enriquecí la variedad de elementos aceptados, para incluir inductores, capacitores, amplificadores operacionales, transformadores ideales, inductancias mutuas y bipuertos con los seis parámetros $z/y/h/g/a/b$.

Inicié la programación del Symbulator aprovechando los días libres de la Semana Santa, en abril de 1999. Aunque todavía estoy perfeccionando su código fuente, es claro que ha sido un verdadero éxito.

A continuación, describiré algunas de sus características principales.

SOBRE LA DESCRIPCIÓN DEL CIRCUITO

El circuito con el cual se alimenta el Symbulator se describe usando una notación particular, similar a la del simulador SPICE, con la particular ventaja de ser mucho más sencilla para definir redes bipuertos y fuentes dependientes de corriente y voltaje. Las fuentes dependientes se describen en la misma forma en que lo hacen los libros de texto, sin necesidad de una notación especial. Esto hace que el Symbulator sea más fácil de usar que el SPICE en lo que a fuentes dependientes se refiere.

SOBRE LA SIMULACIÓN DEL CIRCUITO

Para simular un circuito, basta con escribir en la línea de comando de la calculadora el tipo de análisis que se desea, seguido de la descripción del circuito. Existen cuatro análisis disponibles: corriente directa DC, corriente alterna AC, análisis transitorio TR y dominio de la frecuencia FD. En el caso de corriente alterna, se solicita la frecuencia de trabajo, la cual puede dejarse en términos de una variable simbólica.

Internamente, el Symbulator escribirá las ecuaciones que describen el comportamiento de ese circuito, las resolverá y almacenará las respuestas en variables con nombres característicos. Todo análisis da como respuestas expresiones para los voltajes en cada nodo y las corrientes en cada elemento del circuito. En el caso de corriente directa y alterna, el programa también da como respuestas las potencias consumidas por cada elemento del circuito. En caso de dominio de la frecuencia y análisis transitorio, el programa se apoya en las rutinas de transformada de Laplace escritas por el danés Lars Frederiksen, quien es mi amigo personal.

HERRAMIENTAS

El Symbulator posee herramientas que facilitan aún más el trabajo del estudiante. He escrito herramientas para encontrar el equivalente Thévenin de un circuito (ya sea activo o pasivo), los parámetros $z/y/h/g/a/b$ de un circuito de dos puertos, las ganancias de un amplificador y para graficar los diagramas de Bode de amplitud y fase. Las dos primeras trabajan en corriente directa, corriente alterna y dominio de la frecuencia, permitiendo encontrar las funciones de transferencia de un circuito en términos de la frecuencia s . En el área de electrónica, escribí herramientas para simular un transistor de unión bipolar (BJT) en corriente directa, y detectar su región de operación. En el área de potencia, escribí herramientas para encontrar las pérdidas, la regulación y la eficiencia de un transformador real a partir de sus pruebas de corto circuito y circuito abierto, y su carga nominal. Estas herramientas se apoyan en las aplicaciones de simulación del Symbulator para agilizar procedimientos generalmente tediosos a mano. Son muy fáciles de usar.

LIMITACIONES DEL SYMBULATOR

El Symbulator tiene básicamente dos limitaciones: circuitos no lineales y circuitos muy grandes. Existen alternativas para aliviar estas limitaciones. Para la primera, el modelado de los elementos no lineales usando sus equivalentes lineales. Para la segunda, reducir partes del circuito a sus equivalentes en términos de Thévenin o de bipuerto.

EJEMPLOS DE PROBLEMAS RESUELTOS CON EL SYMBULATOR

Presento ahora una selección de cuatro problemas de circuitos lineales que fueron resueltos usando únicamente el Symbulator, con una sola línea de comandos: uno de corriente directa, uno de corriente

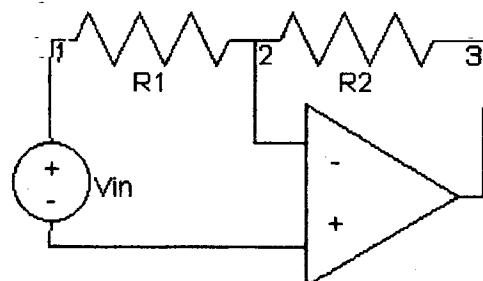


Fig. N°1. Circuito del Ejemplo 1

alterna, uno de análisis transitorio y uno de dominio de la frecuencia.

Ejemplo 1: Corriente directa. Dado el circuito de la figura N° 1, encuentre el voltaje en el nodo 3. Para resolver este problema usando Symbulator, escriba en el área de entrada de su calculadora la siguiente línea de comandos:

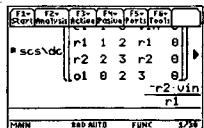
```
scs\dc([e1,1,0,vin,0;r1,1,2,r1,0;r2,2,3,r2,0;o1,0,2,3,0]):v3

```

Fig. N°2. Pantalla de la TI-89 para el Ejemplo 1

La primera parte de esta línea es la orden de realizar un análisis en corriente directa al circuito que se describe entre paréntesis. La segunda es la solicitud del voltaje en el nodo 3. Podemos solicitar el voltaje en cualquier nodo y la corriente y potencia consumida en cualquier elemento del circuito. Nótese la sencilla descripción del circuito (los ceros al final de la descripción de cada elemento son relleno). Mi calculadora TI-89 con 187KB libres en su RAM tomó 44 segundos en resolver este problema. La respuesta se muestra en la pantalla: $(-r_2 \cdot \text{vin})/r_1$. Veamos ahora otro ejemplo.

Ejemplo 2. Corriente alterna. Un circuito consiste de una resistencia en serie con una capacitancia. ¿Cuáles son sus valores si al aplicar un voltaje de 240 V r.m.s. y 200 Hz se obtiene Una corriente de $1.2 + 1.6i$ A? Para resolver este problema usando Symbulator, escriba en el área de entrada de su calculadora la siguiente línea de comandos:

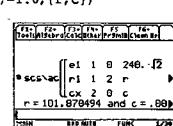
```
scs\ac([e1,1,0,240.*sqrt(2),0;r1,1,2,r,0;cx,2,0,c,0],2*pi*200.):solve(re
al(ir1)=1.2 and imag(ir1)=1.6,(r,c))

```

Fig. N°3. Pantalla de la TI-89 para el Ejemplo 2.

La primera parte de esta línea es la orden de realizar un análisis en corriente alterna al circuito que se describe entre paréntesis. Nótese que se dieron valores simbólicos a la resistencia y la capacitancia. La segunda parte es para solicitar los valores de resistencia y de capacitancia que cumplen con la corriente compleja descrita por el encabezado

del problema. Mi calculadora TI-89 con 187KB libres en su RAM tomó 32 segundos en resolver este problema y dar la respuesta de 102 ohmios y 5.86 microfaradios (debe darse scroll hacia la derecha para ver en la pantalla de la calculadora el resto de la respuesta).

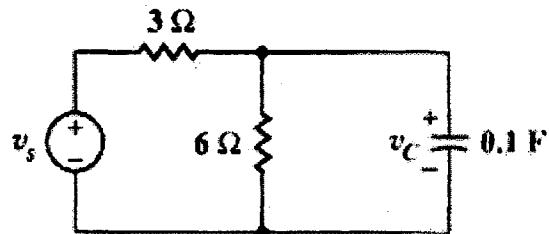


Fig. N°4. Circuito del Ejemplo 3

Ejemplo 3: Análisis transitorio. Dado el circuito de la figura N°4, encuentre una expresión simbólica para el voltaje $vc(t)$ si $vs(9 \cos(t))$ y $vc(0)(2V)$. Para resolver este problema usando Symbulator, escriba en el área de entrada de su calculadora la siguiente linea de comandos:

scs\tr([es,1,0,9*cos(t),0;r3,1,c,3,0;r6,c,0,6,0;cx,c,0,1/10,2]):vc

F1=Start	F2=Analysis	F3=Active	F4=Passive	F5=Ports	F6=Tools
scs\tr	r3 1 c 3 r6 c 0 6 cx c 0 1/10 $-49 \cdot e^{-5 \cdot t} + \frac{75 \cdot \cos(t)}{13} + \frac{1}{13}$				
MAIN	RAD AUTO	FUNC	1/30		

Fig. N°5. Pantalla de la TI-89 para el Ejemplo 3.

La primera parte de esta línea es la orden de realizar un análisis transitorio al circuito que se describe entre paréntesis. Nótese dos cosas: 1) el nodo c tiene una letra por nombre, y 2) la condición inicial del capacitor se colocó en el quinto espacio de su descripción. Este quinto espacio es precisamente para colocar condiciones iniciales de capacitores e inductores en análisis transitorio, y en los demás elementos y análisis suele ser cero. La segunda parte de la línea es para pedir el voltaje en el nodo c, o sea $vc(t)$. Mi calculadora TI-89 con 187KB libres en su RAM tomó 80 segundos en resolver este problema (la respuesta sigue hacia la derecha de la pantalla). El análisis transitorio es el que más tiempo consume. Recuérdese que el simulador nos entrega como respuestas las expresiones de todos los voltajes y Corrientes del circuito, y a cada uno debe aplicarle la



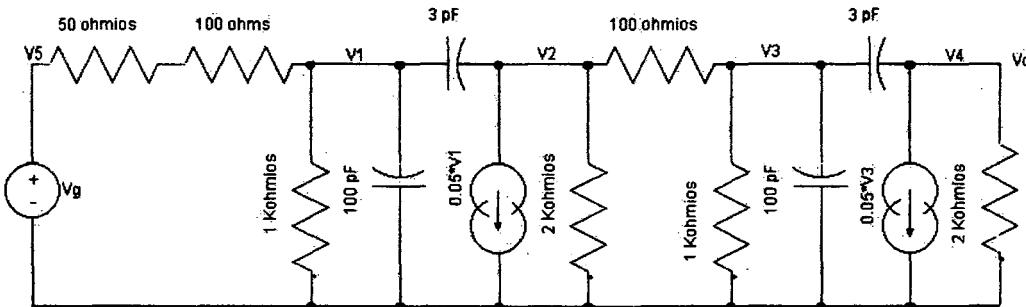


Fig. N°6. Circuito del Ejemplo 4.

transformada inversa de Laplace para llevarla al dominio del tiempo. Por esto tarda más.

Ejemplo 4: Dominio de la frecuencia. Dado el circuito de la figura N°6, grafique el diagrama de ganancia de Bode para V_4/V_5 y frecuencias entre 104 y 1013 rad/seg. Para resolver este problema usando Symbulator, escriba en el área de entrada de su calculadora la siguiente línea de comandos (Figuras 7 y 8).

La primera parte de esta línea es la orden de realizar un análisis en dominio de la frecuencia al circuito que se describe entre paréntesis. Este es un circuito considerablemente grande. La segunda parte es para activar el graficador de Bode del Symbulator. Simular el circuito en mi calculadora TI-89 con 187KB libres en su RAM tomó 4 minutos con 5 segundos. Graficar el diagrama de ganancia de Bode tomó 52 segundos. El tiempo total de trabajo de la calculadora es, por lo tanto, inferior a 5 minutos.

Que sirvan estos problemas como una rápida muestra de lo que puede hacer el Symbulator. En la página del programa en la red Intemet se pueden encontrar alrededor de cuarenta ejemplos y problemas resueltos, cada uno con una característica distintiva.

RESULTADOS HASTA EL PRESENTE

Hasta el presente, los resultados obtenidos por el Symbulator son sumamente satisfactorios.

Personalmente, lo utilzo en mis clases y me siento complacido por la facilidad con que se describe el circuito, la precisión de los resultados y la velocidad del programa. Con el Symbulator, he logrado resolver en un par de minutos algunos problemas de análisis transitorio que a mano habrían tomado un cuarto de hora o más. He aplicado el Symbulator con mucho éxito a problemas de los cursos de Circuitos I, Circuitos II, Circuitos III, Electrónica I, Electrónica II y Conversión de Energía I.

```
scs\fd ([e1,5,0,1.,0;r1,5,1,150.,0;r2,1,0,1000.,0;cc1,1,0,1E-10,0;
cc2,1,2,3E-12,0;jd1,2,0,0.05*v1,0;r3,2,0,2000.,0;r4,2,3,100.,0;
r5,3,0,1000.,0;cc3,3,0,1E-10,0;cc4,3,4,3E-12,0;jd2,4,0,0.05*v3,0;
r6,4,0,2000.,0]):scs\plot()
```

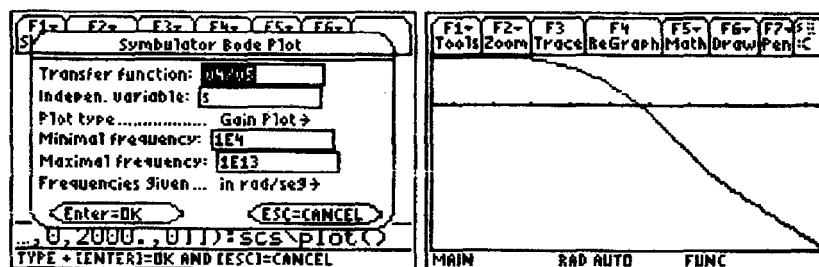


Fig. N°7 y N°8. Pantallas de la TI-89 para el Ejemplo 4.

He recibido muchos comentarios positivos de docenas de personas que utilizan este simulador como una herramienta cada vez más importante en sus estudios. A medio año de haber sido presentado en la red Internet el Symbulator tiene usuarios en todo el mundo, quienes han descargado gratuitamente el programa desde su página oficial. He recibido mensajes de agradecimiento y reportes de funcionamiento de estudiantes y profesionales de la ingeniería eléctrica y áreas afines, provenientes de lugares tan diversos y distantes como Australia, Austria, Bélgica, Brasil, España, Estados Unidos (de los estados de California, Kentucky, Missouri, New York, Ohio, Pennsylvania, Texas), Finlandia, Francia, Gran Bretaña (Oxford), Israel, Italia, Noruega, la República Checa y Suiza, entre otros. Muchos de estos usuarios reportaron estar compartiendo el programa con sus compañeros de clases, creando así una cadena que ha aumentado el número de usuarios que se benefician del Symbulator. Un estudiante de Texas reportó que sólo en su salón de clases, doce estudiantes utilizaban el programa Symbulator para su curso de circuitos.

Quiero citar aquí el más elocuente de los comentarios que he recibido. Lo envió Chris Riegel, un estudiante de ingeniería eléctrica y de computadoras y presidente de la Rama estudiantil del IEEE en la Universidad de Kentucky: "Para los cursos EE 211 y EE 221, el Symbulator en verdad funciona mejor que el SPICE porque se obtienen los resultados simbólicamente".

EVALUACIÓN DE COSTOS VS. BENEFICIOS

La programación del Symbulator para la TI-89 ha tenido un costo relativamente bajo en tiempo y en dinero. Desarrollar el concepto básico me tomó una semana de meditaciones y consultas. Programarlo me ha llevado menos de cien horas de trabajo. Ya me he ahorrado más tiempo que el que invertí, gracias a su ayuda en exámenes y tareas. La inversión que he hecho en calculadoras TI-89 y sus cables para conectarlas con mi computadora es de menos de B/.500.00.

Dado que se ofrece de forma gratuita en la Internet el costo del Symbulator para un estudiante que ya tenga su TI-89 es nulo. Para aprender a usar el Symbulator en su TI-89, el tiempo que necesita un alumno promedio es alrededor de dos horas, que es lo que toma leer y entender a cabalidad la documentación del programa. El tiempo se reduce para estudiantes familiarizados con el SPICE u otro simulador. Según mi propia experiencia puedo decir que en sus primeras semanas con el Symbulator, el estudiante se ahorra más tiempo que el invertido si lo usa para resolver circuitos

en sus clases, tareas y exámenes. Además, reducirá dramáticamente el riesgo de respuestas equivocadas: al programa le basta con que el usuario defina bien el circuito, lo cual es muy fácil con la sencilla notación usada. El Symbulator brinda al estudiante de ingeniería eléctrica y electrónica muchos beneficios, sin gasto en dinero y a un bajo costo en tiempo.

El éxito obtenido por el Symbulator se debe medir en la mejora en el aprendizaje y rendimiento de los muchos estudiantes que ya lo están utilizando alrededor del mundo para enfocarse en el concepto de los circuitos que estudian, en vez de la matemática que involucra su análisis.

Les invito a visitar el sitio en la red del Symbulator (<http://scs.ticalc.org>) y examinar los más de cuarenta ejemplos y problemas resueltos, los halagadores comentarios de los usuarios alrededor del mundo y la descripción completa de la instalación y utilización del Symbulator y todas sus herramientas.

CONCLUSIÓN TÉCNICA

Hemos visto que el proyecto Symbulator se ha llevado a cabo con éxito, y que son muchos los estudiantes que se están beneficiando con él. Ahora sólo resta esperar a depurar cualquier error que pudiese descubrirse en el código. Para concluir, quiero presentar un comentario que Jay Myers, un estudiante de Ingeniería Eléctrica en la Universidad de Texas en Arlington (EUA) y asiduo usuario del Symbulator, hiciera en un grupo de discusión sobre ingeniería eléctrica en la red Internet.

«Este programa es el mejor simulador de circuitos para calculadoras. Si, hay otros, pero ninguno es mejor. Este programa es similar al SPICE en la manera en que se usa, y es incluso más fácil cuando se trata de simulación de fuentes dependientes y bipuertos. Una vez se ha resuelto el circuito, el usuario tiene toda la información que podría querer, voltaje en los nodos, corriente en los elementos, potencia absorbida por los elementos y caída de voltaje en los elementos mediante diferencia de voltajes en los nodos. Resuelve simbólicamente, a diferencia de casi todos los demás simuladores de circuito (incluso para PC). Este programa es, en verdad, el mejor programa para análisis de circuitos que he visto. La resolución de circuitos grandes puede ser lenta y agotar la memoria de la calculadora. Existen alternativas, las cuales incluyen herramientas para partir el circuito en grupos de equivalentes Thévenin o bipuertos. No podría ensalzar este programa lo suficiente. Es difícil de creer que el señor Pérez-Franco permita al mundo usar su fantástico programa libre de costo. ¡excelente trabajo!»





DISEÑO DE UN FILTRO OPTICO PASA BANDA VARIABLE

Eliana Corredor, Tatiana Carrillo, Hans C. Bacares

Rama Estudiantil Santo Tomás de Aquino

Universidad Santo Tomás

Facultad Ingeniería Electrónica,

Calle 116 No. 13B-19 Apartamento 402 Bogotá,

Santa fe de Bogotá Distrito Capital

Colombia

EXTRACTO

Se hace el estudio completo para un filtro de longitudes de onda ópticas utilizando un material dispersante, varias capas refractantes y una cavidad resonante Fabry-Perot. Se pretende determinar las condiciones de diseño y las características finales de este tipo de dispositivo.

INTRODUCCION

En el mundo actual, la optoelectrónica y la fotónica han cobrado relevancias no esperadas en el campo tecnológico. En especial aquellos sistemas que permitan manipular luz.

Esta importancia queda de manifiesto en la implementación de filtros ópticos. Basta una mirada al mercado internacional para descubrir la gran cantidad de aplicaciones que estos dispositivos presentan, en campos tan variados como

En vista de esta importancia, se desea contribuir a este ramo con un dispositivo de separación y selección de ondas electromagnéticas ubicadas en el espectro óptico visible.

El sistema consiste en varias capas de materiales refractantes que en combinación, se encargarán de dispersar un haz de luz, normalizarlo, amplificarlo y seleccionar una longitud de onda específica para su posterior aprovechamiento. Un modelo aproximado del diseño se muestra en la Figura 1.1.

El material ① es un cristal transparente con un coeficiente de refracción dependiente de la longitud de onda, esta variación debiera ser lineal, o aproximarse a una línea recta, con lo cual se facilita el cálculo del ángulo de refracción, un ejemplo de esta variación se muestra en la Figura 1.2.

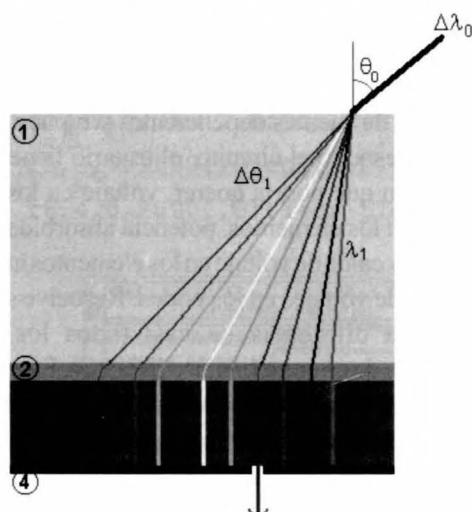


Figura 1.1. Modelo para el filtro en estudio

interferometría, espectro-metría, vídeo, láser, redes y comunicaciones, entre otros.

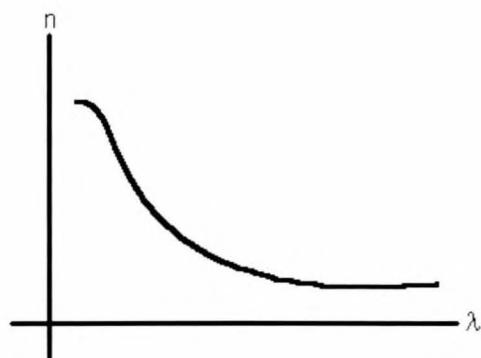


Figura 1.2. Curva típica de $n(l)$

Una vez obtenido el ángulo con el cual será refractada cada longitud de onda, un sencillo cálculo puede informar a qué distancia del origen (x) golpeará dicho rayo. Como se ve en la figura 1.3,

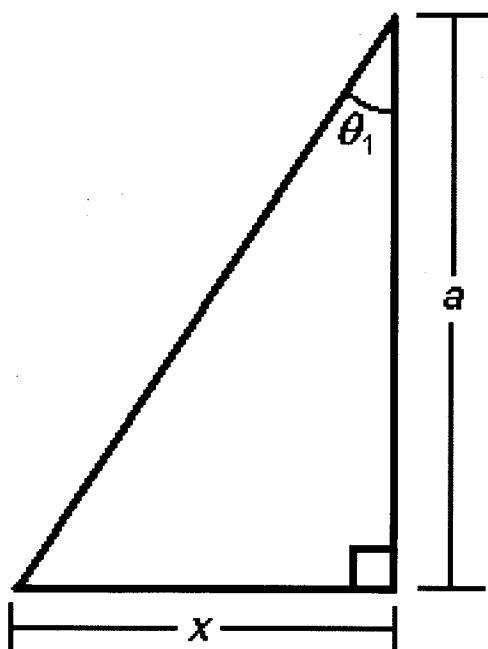


Figura 1.3. Determinación de x

$$x = a \cdot \tan \theta_1 \quad (1.1)$$

El material ② consiste en un conjunto de varias capas de cristal con coeficientes de refracción progresivamente mas altos. Esto con el fin de normalizar cada haz.

De la ley de Snell se sabe que:

$$\frac{\sin \theta_r}{\sin \theta_i} = \frac{n_i}{n_r} \quad (1.2)$$

dada esta relación, es posible calcular una progresión de ángulos a partir del más lejano a la normal (en la Figura 1.1 corresponde al rojo), ya que si este es normalizado, también el menor lo será.

Debemos conocer previamente un material cuya curva en la región del espectro óptico tenga comportamiento lineal

La capa ③ es una cavidad resonante Fabry-Perot construida con un material que responda a la condición de resonancia en cada longitud de onda.

Finalmente, el sustrato ④ es una superficie reflectiva que evitará el escape de las ondas atrapadas en la cavidad resonante, este tiene una ventana de un ancho a determinar, que permitirá el paso de una longitud de onda específica, conformando de este modo un filtro óptico pasa banda variable.

DISEÑO

Para poder determinar como será la variación de n contra λ en la capa ①, se debe conocer previamente un material cuya curva en la región del espectro óptico tenga comportamiento lineal, de este modo se simplifica el control de dispersión angular en éste. La Figura 2.1 muestra una curva idealizada para tal fin.

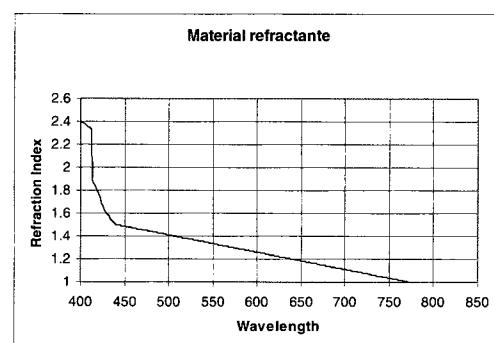


Figura 2.1. Curva de n vs. λ en un cristal

Esta curva puede ser trabajada como lineal en la región 450nm - 750nm del espectro electromagnético. Si se observa la gráfica, se tiene que:

$$n(\lambda=450nm) = 1.48$$

$$n(\lambda=750nm) = 1.03$$

con lo cual se puede deducir que

$$n(\lambda) = 2.15 - 0.0015\lambda \quad (2.1)$$

Sabiendo que el medio exterior es aire ó vacío, se puede plantear la ecuación (1.2) como:

$$\theta_1(\lambda) = \arcsin \left[\frac{\sin \theta_0}{2.15 - 0.0015\lambda} \right] \quad (2.2)$$

la cual permite conocer una expresión del ángulo de refracción con respecto a la longitud de onda y al ángulo de incidencia.

Se debe determinar cuál es el ángulo de incidencia óptimo, para ello se fija en la ecuación (2.2) un λ , y se hace variar θ_0 .

Como se trabajará entre las longitudes de onda para las cuales el cristal se comporta linealmente, se plantean dos ecuaciones; una para 450nm y otra para 750nm.

$$\alpha(r) = \arcsin \left[\frac{\sin \left(\frac{\pi}{r} \right)}{2.15 - 0.0015 \cdot 450\text{nm}} \right] \quad (2.3)$$

y

$$\beta(r) = \arcsin \left[\frac{\sin \left(\frac{\pi}{r} \right)}{2.15 - 0.0015 \cdot 750\text{nm}} \right]. \quad (2.4)$$

Al graficar estas dos expresiones se obtiene la Figura 2.2.

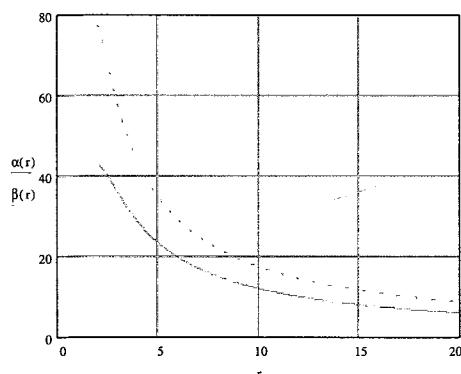


Figura 2.2. Ángulo de refracción contra ángulo de incidencia para las condiciones de frontera.

Teóricamente, $\pi/2$ es el ángulo de incidencia que permite separar más los ángulos de refracción de las diferentes longitudes de onda, pero el haz no puede incidir con esta inclinación. A medida que θ_0 se approxima a la normal, el abanico de colores se estrecha.

El ángulo de incidencia puede ser realmente cualquiera que se necesite, sin embargo, no se pueden separar demasiado los colores ya que se tendrían problemas posteriores para normalizar los haces.

Con este criterio, se escoge un ángulo de incidencia $\theta_0 = 5^\circ$

Con este ángulo de incidencia, la gráfica para la ecuación (2.2) es:

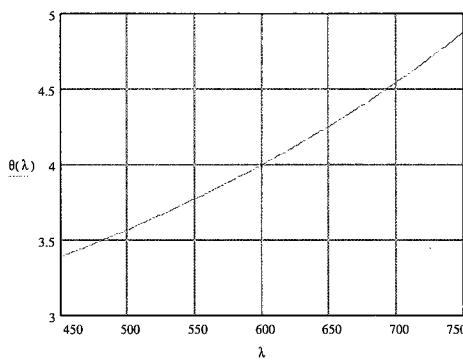


Figura 2.3. Gráfica de la ecuación (2.2) con $\theta_0 = \pi/3$

De la Figura 2.3 se pueden conocer los ángulos de refracción para cada longitud de onda. Esto permite pasar a determinar la separación que tendrán estos haces de luz al llegar al fondo de la capa ①.

El ángulo de incidencia puede ser realmente cualquiera que se necesite

De la ecuación (1.1), se debe determinar el valor para a (Figura 1.3), observando los filtros fabricados en el mercado(1), se puede concluir que los tamaños óptimos para estos están en el orden de las decenas de milímetros, esto permite que el dispositivo tenga aplicaciones en la

optoelectrónica; además de permitir este tamaño una baja perdida por atenuación.

Se escoge una altura $a=20\text{mm}$.

La gráfica para la ecuación (1.1) queda como se muestra en la Figura 2.4.

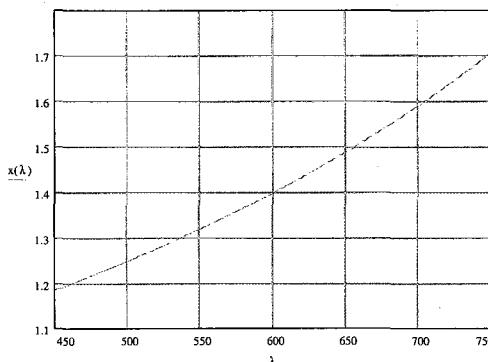


Figura 2.4. $x(\lambda)$ (mm) vs. λ (nm).

Con esto se concluyen las consideraciones para la capa ①, ahora se debe calcular una progresión de placas refractantes con el objeto de normalizar los haces de color.

Los materiales de la capa ② deberían tener cada uno un índice de refracción constante con

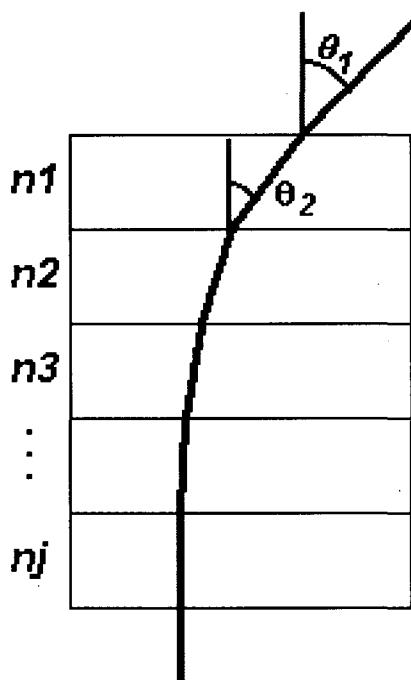


Figura 2.5. Placas refractantes en la capa ②

respecto a la longitud de onda, esto con el objeto de facilitar el cálculo de los ángulos que tomará el haz a corregir. Por esto, el coeficiente n_1 debe ser mayor que el de la capa ① para la menor longitud de onda.

$$n_1 > 1.48$$

$$n_1 = 1.6$$

Se hace que cada capa tenga un n igual a 1.5 veces el coeficiente de la capa superior. De la ecuación (1.2), se obtiene entonces la tabla 2.1

Número de sustrato	n	450nm		750nm	
		θ	$\lambda(\text{nm})$	θ	$\lambda(\text{nm})$
1	1.6	3.1°	720	3.1°	1200
2	2.4	2.1°	1080	2.1°	1800
3	3.6	1.3°	1620	1.3°	2700
4	5.4	0.9°	2430	0.9°	4050
5	8.1	0.6°	3645	0.6°	6075
6	12.1	0.4°	5467	0.4°	9112
7	18.2	0.2°	8201	0.2°	13669
8	27.3	0.1°	12302	0.1°	20503
9	41	0.01°	18453	0.01°	30755
10	61.5	0°	27679	0°	46132

Tabla 2.1 Valores del ángulo refractado (θ) con respecto al coeficiente de refracción (n) en cada capa.

Se dispone un ancho de 1mm para la capa ②, lo que quiere decir que para cada sustrato en esta capa, se tiene un ancho de 0.1mm

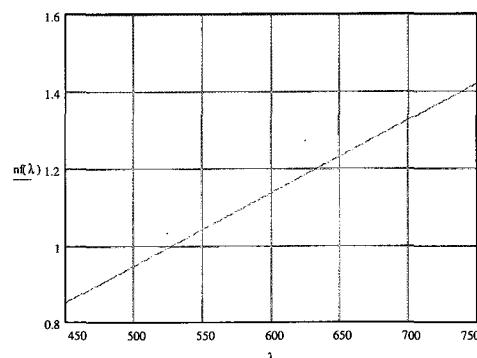


Figura 2.6. Característica de n_j vs. λ



Ahora se introducen los haces de luz a una cavidad Fabry-Perot, la cual se encargará de amplificar cada longitud de onda hasta recuperar la energía perdida por cada color en el paso por el filtro.

La condición de resonancia en una cavidad de este tipo esta dada por

$$n_f \cdot L = N \left(\frac{\lambda_{10} n_{10}}{2} \right) \quad (2.5)$$

donde L es el alto de la cavidad, n_f es el índice de refracción en el material de la cavidad, λ_{10} es la longitud de onda en el último sustrato de la capa ②, y N es un número entero iterable que permite que se cumpla la condición.

Se escoge N=1, y L=1mm, entonces se obtiene una función para n_f

$$n_f(\lambda_{10}) = \frac{\lambda_{10} n_{10}}{2 \times 10^6} \quad (2.6)$$

La Figura 2.6 muestra la característica de índice de refracción contra longitud de onda requeridas para el cristal que compone la cavidad Fabry-Perot. Si el material responde de este modo, se producirá interferencia constructiva para todos los colores.

La capa ④ es una superficie reflectora móvil. Se debe conocer entonces en qué posición (x) va a incidir cada longitud de onda con el objeto de calibrar el dispositivo.

Por efecto de la capa ②, todos los haces sufrirán un desplazamiento de su posición original hacia la izquierda.

Para determinar este desplazamiento se utiliza la ecuación (2.7).

$$\Delta x = \sum_{n=1}^{10} \tan \theta_n \quad (2.7)$$

de donde se obtiene que $\Delta x=0.151\text{mm}$.

Se hace con esto una corrección a la ecuación (1.1), quedando:

$$x = a \cdot \tan \theta_1 + \Delta x \quad (2.8)$$

Con el valor de x se ajusta un tornillo micrométrico para mover la capa ④ y seleccionar de este modo la longitud de onda deseada.

Desafortunadamente conocer las características ópticas de un material no es posible por cuanto los fabricantes no las proveen, dificultando así el trabajo del diseñador.

Si se desea una muy alta selectividad de color (obtener idealmente una sola longitud de onda), la ventana de selección debería tener un ancho de 1.66nm, para otras aplicaciones, el ancho de esta ventana se deja a juicio del usuario.

También existe la posibilidad de obtener simultáneamente varias longitudes de onda distintas, si se colocan en la capa ④ varias ventanas controlables (tipo iris), que se abrirían y cerrarían mediante mandos electrónicos.

CARACTERISTICAS TECNICAS

Característica	Valor
Ancho de banda	450nm-750nm
Alto	22mm
Ancho	1mm
Espesor	1mm
Ángulo de incidencia	5°

Nota: Para que el dispositivo funcione adecuadamente, se hace necesario que el medio exterior sea aire, o que posea un índice de refracción $n=n_0$

CONSTRUCCION DE LOS MATERIALES

Al construir el dispositivo se deben tener en cuenta los materiales existentes en el mercado que sean adecuados para el diseño, de tal suerte que las características técnicas del filtro mantengan márgenes de error muy bajos.

Desafortunadamente conocer las características ópticas de un material no es posible por cuanto los fabricantes no las proveen, dificultando así el trabajo del diseñador. Por lo tanto en esta sección se hacen consideraciones muy generales con respecto a los materiales con los que se debería fabricar el filtro.

Un candidato para constituir la primera capa es el Beta-BBO (BaB_2O_4), el cual tiene un comportamiento como el descrito en la Figura 4.1.

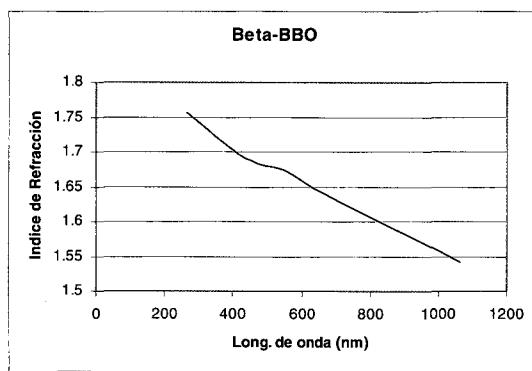


Figura 4.1. Característica de n vs. λ para Beta-BBO

Estas características presentan sin embargo un inconveniente de fabricación, puesto que la pendiente de descenso del índice de refracción al aumentar la longitud de onda es muy baja, proveyendo un abanico muy estrecho para el haz de luz refractado. Esto dificulta la tarea de selección posterior del color deseado, ya que se encuentran demasiado cerca uno de los otros.

Al realizar una línea de tendencia para la gráfica 4.1, es posible encontrar la ecuación que describe el comportamiento del material:

$$n(\lambda) = 1.8161 - 0.0003\lambda \quad (4.1)$$

quedando la ecuación (2.2) para este material expresada de la siguiente manera.

$$\theta_1(\lambda) = \operatorname{sen}^{-1} \left[\frac{\operatorname{sen} \theta_0}{1.8161 - 0.0003\lambda} \right] \quad (4.2)$$

La consideración del ángulo de incidencia se mantiene igual para este caso, de este modo se gráfica el ángulo refractado contra la longitud de onda.

La Figura 4.2 confirma lo anteriormente expuesto; la diferencia entre el ángulo mayor y el menor es muy pequeña.

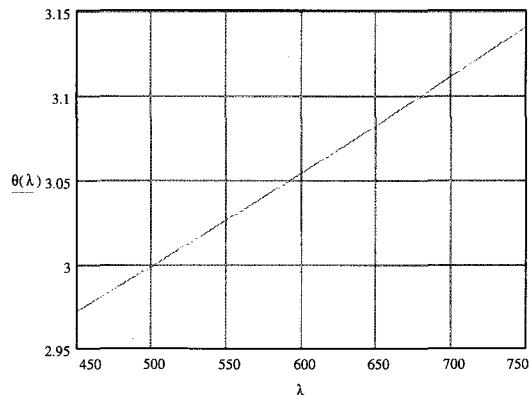


Figura 4.2 Relación del ángulo de refracción para cada λ .

En la Figura 4.3 se determina la separación de los haces de luz en el fondo de la primera capa (Ecuación 1.1).

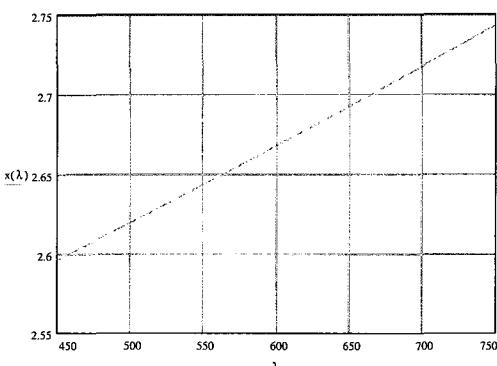


Figura 4.3. Distancia x entre los haces de color.



Se observa que la separación entre la menor longitud de onda y la mayor es de tan solo unas cuantas décimas de milímetro; esto presenta un considerable problema de construcción que puede ser solucionado aumentando el alto de la primera capa, o aumentando el ángulo de incidencia. La primera solución hará que las dimensiones del filtro crezcan de manera notable, llegando a ser un dispositivo poco manejable en sistemas pequeños; mientras que la segunda generará problemas posteriores para la normalización de los haces de luz refractados.

Se observa que para la segunda capa se requieren cristales con índices de refracción muy altos (hasta 60.1)

Se observa que para la segunda capa se requieren cristales con índices de refracción muy altos (hasta 60.1). A pesar de que estos cristales pueden ser fabricados, no son transparentes al espectro óptico, por tanto se debe evitar su utilización en el dispositivo.

Una posible solución a este inconveniente es tomar un cristal transparente y doparlo con moléculas metálicas, de este modo aumentará su índice de refracción, pero se mantendrá su transparencia.

CONCLUSIONES

Este filtro tiene aplicaciones en optoelectrónica, como un chip óptico; al disponerse de dimensiones mayores (lo cual es posible aumentando el alto de la capa ①), puede ser utilizado en combinación con una película fotográfica, para espectrografía.

Una de las mayores ventajas de este diseño es que la cavidad Fabry-Perot se encarga de

reponer las pérdidas ocasionadas en el proceso de dispersión y normalización de la onda.

El diseño es acomodable a una amplia variedad de características de materiales de tal suerte que sea versátil y sencillo de montar con familias de cristales debidamente escogidas.

El diseño es acomodable a una amplia variedad de características de materiales, debido a que ha sido planteado de una manera muy general, de tal suerte que sea versátil y sencillo de montar con familias de cristales debidamente escogidas.

La calibración del dispositivo es muy delicada, ya que se requiere un tornillo micrométrico de alta precisión para la correcta selección de longitudes de onda. Una segunda opción al respecto es utilizar un micromotor para la colocación de la ultima capa; aunque esta especificación aumentaría ostensiblemente el costo del dispositivo.

REFERENCIAS

- [1] CFHT: Canada - France - Hawai Telescope.
www.cfht.hawai.edu
- [2] Cleveland Crystals.
www.clevelandcrystals.com
- [3] Brown, earle. «Optical Instruments». Chemical Publishing Co. Inc. 1945
- [4] Feynman, Richard. «Lectures on Physics». Addison Wesley. 1987
- [5] Lipson, S.G. «Optical Physics». Cambridge University Press. 1995
- [6] Solymar, L. «Lectures on the Electrical Properites of Materials». Oxford Science Publications. 1993



DISEÑO DE UN MODULO PARA EXPERIMENTOS CON LA TARJETA DE EVALUACION DE MOTOROLA «MC68HC11EVB»

Erick Altamiranda

Rama Estudiantil del IEEE de la Universidad Tecnológica de Panamá

Ciudad de Panamá, Panamá

erick@depsa.com

RESUMEN

Este trabajo presenta los criterios de diseño del hardware de un módulo para experimentos con la tarjeta de evaluación MC68HC11EVB de Motorola. El módulo tiene como objetivo resolver el problema de construir experimentos en protoboard, evitando así los errores de hardware y reducir el tiempo de desarrollo de aplicaciones en los laboratorios. El módulo presentado poseerá dispositivos utilizados actualmente en la industria, estrechando la brecha entre ésta y los estudiantes. Entre estos dispositivos estarán: pantalla LCD, teclado, relay, optoacopladores, entre otros. Este módulo será de un tamaño reducido, y su consumo de potencia será bajo.

I. INTRODUCCION

En los sistemas educativos universitarios avanzados como el de los Estados Unidos, Japón, Australia, etc., no se pierde tiempo armando grandes experimentos y aplicaciones en protoboard, porque poseen tarjetas que están compuestas por los dispositivos electrónicos necesarios para algunas aplicaciones. Luego en las clases de licenciatura en ingeniería, simplemente los estudiantes aprenden a programar el dispositivo, a entender el diagrama esquemático de periféricos, y crean las aplicaciones directamente con la tarjeta proporcionada por el laboratorio de la Universidad.

El microcontrolador 68HC11 está catalogado como uno de los más didácticos en el mercado, por su facilidad de programación y sus diversos periféricos embebidos.

Los microcontroladores más famosos en las universidades extranjeras, son el 8751 y el 68HC11. Estos dos son los más utilizados en tarjetas de desarrollo y de aplicaciones directas, así como para adquisición de datos y control. El microcontrolador 68HC11 está catalogado

como uno de los más didácticos en el mercado, por su facilidad de programación y sus diversos periféricos embebidos. Existen pocos equipos comerciales de bajo costo que enfoquen en forma didáctica y aplicada el campo de los microcontroladores. Los microcontroladores, por lo general, no se abordan en los cursos, y si se hace, se pierde demasiado tiempo en desarrollar aplicaciones.

La compañía Motorola, líder en ventas de microcontroladores, donó hace unos años a la Facultad de Ingeniería Eléctrica de la Universidad Tecnológica de Panamá, unas tarjetas de evaluación basadas en su microcontrolador más famoso: el 68HC11. Esta tarjeta intenta reducir el tiempo de desarrollo de aplicaciones brindando un sistema ya armado, y permite aprender a programar el 68HC11 y hacer experimentos. Pero tiene una desventaja, estas tarjetas son para evaluación de sistemas, por tal motivo no tienen teclado, ni pantalla de cristal líquido, y las salidas están dispuestas de tal forma que hay que armar todo el sistema de interfaces en protoboard antes de realizar la aplicación. Si la aplicación es pequeña y sencilla, esta bien; pero si la aplicación es más compleja, armarla en protoboard puede acarrear errores diversos. Lo anterior es una desventaja porque hay que acordarse que las horas de laboratorio constan de 135 minutos, 1 vez a la semana. Esto nos da un total de aproximadamente 30 horas reloj (asumiendo un semestre de 14 semanas activas, lo cual es muy difícil), para aprender a usar la tarjeta de evaluación, programar, armar cada uno de los experimentos en protoboard, hacer los informes, y armar también en protoboard el proyecto final de laboratorio.

La idea de diseñar y construir una tarjeta con los elementos necesarios para experimentar, surge de la necesidad de solucionar el problema descrito anteriormente. Los objetivos de diseño, breve descripción de la MC68HC11EVB [1], los criterios de diseño, y costos, se presentan en el contexto del trabajo.

II. OBJETIVOS DE DISEÑO

* Tamaño. Se quiere una tarjeta que ocupe el mínimo espacio posible, de manera que los estudiantes puedan llevarlo sin problemas a cualquier parte, inclusive a la comodidad de su hogar.



* Dispositivos. La tarjeta debe contener distintos dispositivos que le permitan al estudiante ingresar en el competitivo campo de la industria. Entre estos dispositivos están la pantalla de cristal líquido, el teclado, los relay, los optoacopladores, etc. En la Tabla 1 mostramos las áreas de aplicación de estos dispositivos, notando la gran utilidad de los mismos.

* Programación. La tarjeta debe ser de fácil acceso. Es decir, se debe accesar sin problemas mediante programación.

* Rango de Aplicaciones. Esta tarjeta debe permitir al usuario realizar aplicaciones desde encender un LED, hasta adquisición de datos y control de motores.

Esta tarjeta trabajará casi en su totalidad con tecnología CMOS, lo cual garantiza un consumo de potencia bastante bajo.

* Consumo de Energía. Esta es una característica que nunca debe faltar en ningún sistema. Esta tarjeta trabajará casi en su totalidad con tecnología CMOS, lo cual garantiza un consumo de potencia bastante bajo. Además debe tener conectores para alimentación externa de voltaje.

Tabla 1. Campos de Aplicación de los Componentes a utilizar.

Dispositivo	Área de Aplicación
Relays	Industria, Automóvil, Consumo
Optoacopladores	Industria, Automóvil, Consumo
Teclado	Industria, Médica, Automóvil, Consumo
Pantalla de Cristal Líquido	Industria, Médica, Automóvil, Consumo
Transistores de Potencia	Industria, Automóvil, Consumo

* Precio. Se requiere que el proyecto sea de bajo costo, para poder lograr el patrocinio de alguna entidad para la realización en serie de algunas tarjetas.

* Construcción. Se desea que el proyecto sea fácil de construir, sin importar la experiencia del usuario en soldadura, manejo de componentes electrónicos, etc. El circuito impreso debe estar debidamente señalizado, de manera que el usuario pueda encontrar allí y en el manual, la información necesaria para armarlo. También debe ser de fácil instalación y acoplamiento a la tarjeta de la Motorola.

III. TARJETA DE EVALUACION MC68HC11EVB

En la Figura 1 se muestra una fotografía de la tarjeta de evaluación MC68HC11EVB de Motorola. Como se puede apreciar, no contiene los elementos que deseamos

colocar en la tarjeta para experimentos. El bus de expansión que aparece a la izquierda de la Figura 1 (línea negra vertical), es el que se utilizará para realizar la interface entre esta tarjeta y la de experimentos.

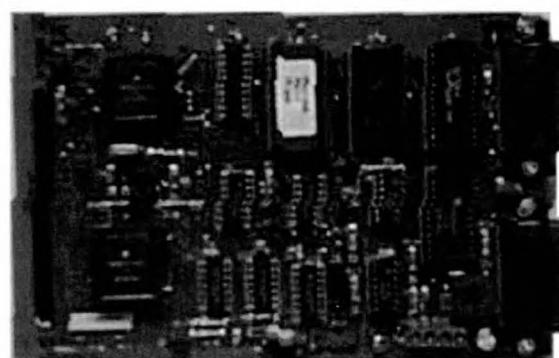


Figura 1. Tarjeta de Evaluación MC68HC11EVB.

En la Figura 2 se muestra la disposición de pines este bus de expansión. En éste se ponen a disposición del usuario todas las líneas de los puertos (A, B, C, D, y E), las líneas de interrupciones, algunos pines de voltaje y tierra, además de las señales de control. Un bus idéntico debe colocarse en la tarjeta de experimentos. Existe la necesidad de construir un cable para realizar la conexión, y Motorola hace una sugerencia al respecto, la cual se muestra en la Figura 2-3 del Manual del Usuario de la MC68HC11-EVB [1].

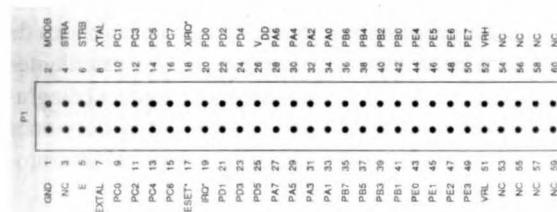


Figura 2. Disposición de los Pines del bus de Expansión de la MC68HC11EVB.

Otro aspecto importante de la tarjeta de evaluación, es el mapa de memoria y disposición de los periféricos que allí se encuentran. En la Tabla 2 se muestran las localidades de memoria utilizadas por la MC68HC11EVB y, aquellas disponibles para la Tarjeta de Experimentos. Resaltada aparece la localidad de memoria que va de \$1800 a \$3FFF, ya que ésta es la que se utilizará para las funciones de teclado y pantalla de cristal líquido.

Los programas que escriba el usuario se almacenarán en las localidades desde \$C000 hasta \$DFFF (8 Kbytes), y si desea más memoria, las localidades \$6000 a \$7FFF (8 Kbytes) están disponibles para eso.

Tabla 2. Mapa de Memoria de la Tarjeta de Evaluación MC68HC11EVB.

FUNCION	DIRECCION
RAM Interno (Reservado)	\$0000 - \$00FF
Disponible	\$0100 - \$0FFF
PRU ¹ + Registros	\$1000 - \$17FF
Disponible (Tarjeta Exper.)	\$1800 - \$3FFF
Flip-Flop Decodificador	\$4000 - \$5FFF
8K de RAM Opcionales	\$6000 - \$7FFF
Disponible	\$8000 - \$97FF
Terminal Serial 6850	\$9800 - \$9FFF
Disponible	\$A000 - \$B5FF
Memoria EEPROM	\$B600 - \$B7FF
Disponible	\$B800 - \$BFFF
RAM del Usuario	\$C000 - \$DFFF
Programa Monitor BUFFALO	\$E000 - \$FFFF

¹ PRU: Port Replacement Unit. Unidad de Reemplazo de Puertos. Se utiliza cuando el 68HC11 está en modo expandido, para devolver los puerto B y C. El dispositivo es el MC68HC24.

Como se mencionó anteriormente, al usuario están disponibles los siguientes puertos de entrada salida:

* Puerto A. Este posee algunos 4 pines para salida solamente, 3 sólo para entrada, y 1 pin bidireccional. Comparte las funciones del sistema de temporización.

* Puerto B. Es un puerto de 8 bit sólo de salida, proporcionado por el PRU MC68HC24.

* Puerto C. Es un puerto bidireccional de 8 bits, también proporcionado por el PRU MC68HC24.

* Puerto D. Es un puerto de 6 bits bidireccional, que comparte las funciones del sistema de comunicación serial. En nuestro caso, quedan disponibles sólo 4 pines.

* Puerto E. Es un puerto de 8 bits para entradas solamente, que comparte las funciones con el sistema de análogo digital.

El diseño de esta tarjeta se basó en las características de cada uno de los puerto que proporcionaba el bus de expansión de la MC68HC11EVB.

Tomando en cuenta todas las características de los puertos que están a disposición del usuario, se procedió a realizar el diseño de la tarjeta de experimentación, el cual se describe en los siguientes puntos.

IV. DISEÑO DE LA TARJETA DE EXPERIMENTOS

En las secciones siguientes se describe el diagrama de bloques del sistema, la selección de los componentes, y se muestra el diagrama de esquemático del circuito. El diseño de esta tarjeta se basó en las características de cada uno de los puerto que proporcionaba el bus de expansión de la MC68HC11EVB.

1. Diagrama de Bloques

Ya que la tarjeta debía proporcionar al usuario un medio de poder experimentar con diferentes dispositivos de uso común en las aplicaciones industriales, ésta contendría relés, optoacopladores, transistores de potencia, entradas y salidas digitales, acceso al convertidor análogo-digital, teclado y pantalla de cristal líquido [2]. Todos estos elementos tienen que estar conectados a los puertos de microcontrolador, que puedan realizar la función deseada, dependiendo de las características del puerto (por ejemplo, el puerto B sólo puede tener dispositivos de salida). Como no se cuentan con muchos puertos, el teclado y la pantalla de cristal líquido estarían conectadas al sistema utilizando un decodificador de direcciones, y aprovechando de que no todas las direcciones de memoria de la MC68HC11EVB están siendo utilizadas (ver Tabla 2).

El diagrama de bloques que surgió de todas esas condiciones de diseño se muestra en la Figura 3.

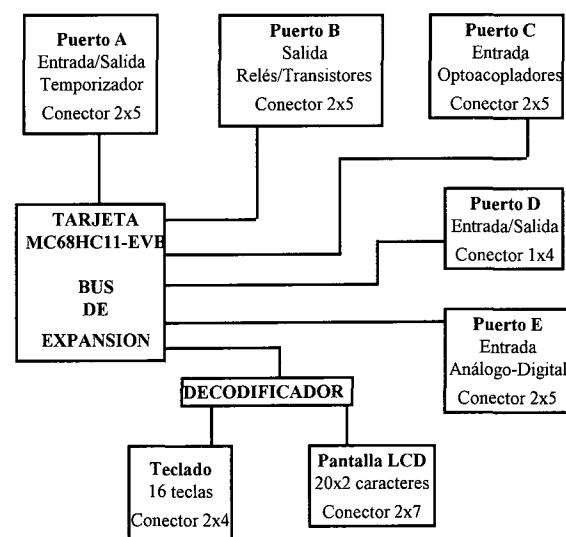


Figura 3. Diagrama de Bloques de la Tarjeta para Experimentación.

Como se puede observar en la Figura 3, cada puerto tendrá una labor a realizar con algún dispositivo electrónico de uso común, y que cumple con las características de uso mencionadas en la Tabla 1. El proceso de selección



de cada componentes electrónicos mencionado en el diagrama de bloques, se realizó con el fin de escoger aquellos que mejor cumplieran con los objetivos mencionados al inicio del trabajo (precio, tamaño, aplicación, etc.). Este proceso se describe en el siguiente punto.

Para la selección de componentes se utilizaron diferentes criterios, los cuales nos aseguraría de cumplir con los requisitos básicos del sistema.

2. Selección de los Componentes

Para la selección de componentes se utilizaron diferentes criterios, los cuales nos aseguraría de cumplir con los requisitos básicos del sistema. Una breve descripción de estos criterios se presenta a continuación.

Para el decodificador de direcciones, se quería que fuera de tecnología CMOS, y por lo menos admitiera 4 dispositivos. Al momento de diseñar el proyecto, el autor poseía decodificadores de 3 a 8 líneas, y siendo estos de bajo costo, se decidió por ellos. Así que el decodificador a usar es el 74HC138, dejando para otras aplicaciones las líneas sobrantes.

Para la pantalla de cristal líquido, se deseaba que tuviera al menos 16x2 líneas, pero como la diferencia de precios entre ésta y una de 20x2 no era mucha, se decidió usar la de 20x2 caracteres. Con relación al teclado, se tomó la decisión de utilizar un controlador de teclado de 16 teclas (74C922), porque proporciona teclas suficientes para realizar experimentos (10 números y algunos caracteres).

Las entradas optoaisladas (u optoacopladas), sirven para realizar conexiones entre el microcontrolador y otro dispositivo con voltaje distinto a 5V o para aislar al microcontrolador de ruido eléctrico. Por tal razón había que escoger optoacopladores que tuvieran una buena capacidad de aislamiento. Por esas razones y por su precio se escogieron optoacopladores fototransistores Darlington de 7.5 KV de aislamiento. Su identificación es 4N32.

Por último, con respecto a los transistores de potencia y los relés, tenían que tener buenas características de voltaje y corriente. Los relés se escogieron de 125VAC/100VDC a 2A del tipo DIP (parecido a un circuito integrado, sólo que más alto). Los transistores de potencia se escogieron Darlington NPN de 500mA y 50V.

Con esta breve explicación se la selección de los componentes más importantes del proyecto, podemos pasar a ver el diagrama esquemático del mismo.

3. Diagrama Esquemático

El diagrama esquemáticos de este proyecto, es un plano que ocupa alrededor de dos hojas, ya que contiene todas las sesenta (60) conexiones que posee el bus de expansión de la tarjeta MC68HC11EVB. Por esa razón, este diagrama no aparece en este trabajo. Aunque se intentó reducir en fotocopia, se pierde mucha resolución y no se aprecian bien las conexiones. El autor considera que con la descripción del diagrama de bloques y los demás, se tiene una idea bastante clara de lo que se propone.

V. OPERACION Y PROGRAMACION DE LA TARJETA PARA EXPERIMENTACION

La tarjeta en cuestión opera simplemente cuando el usuario la conecta y la accesa por medio de la MC68HC11-EVB. La tarjeta de experimentación será dependiente del sistema que contenga a el microcontrolador. Para su programación se utiliza el mismo lenguaje y herramientas que se utilizan para usar la MC68HC11-EVB, con la única diferencia que hay que conocer las direcciones del mapa de memoria en las cuales están instalados el teclado y la pantalla de cristal líquido. Las direcciones de los puertos A, B, C, D, y E son las mismas que aparecen en el Manual del Usuario de la MC68HC11-EVB [1]. Las direcciones de memoria que hay que destacar para la programación del teclado y la pantalla de cristal líquido son: \$1880 para el teclado, y \$1890 para la pantalla (ver Tabla 2). Esto quiere decir que al igual que se accesa un puerto del microcontrolador, se puede accesar el teclado o la pantalla.

La tarjeta en cuestión opera simplemente cuando el usuario la conecta y la accesa por medio de la MC68HC11-EVB.

Para utilizar la tarjeta para experimentación con la MC68HC11-EVB, sólo conecta esta última como usualmente se hace, y luego se conecta la tarjeta de experimen-

LDAA #\$FF	: cargo el acumulador A con 00 (00000000 en binario)
STAA #\$1004	; coloca el acumulador en el puerto B. Esto apagaría los motores.
LDAA #\$00	; cargo el acumulador A con 00
STAA #\$1007	; coloca a el puerto C como entrada
OTRO LDAA #\$1003	; lee el puerto C. Esto verifica el estado de los interruptores.
STAA #\$1004	; enciende el motor correspondiente
BRA OTRO	; repite el procedimiento
END	; finaliza el programa.

tación al bus de expansión. Cuando todo esto esté listo, será cuestión de programar y ya.

Un ejemplo de lo que sería la utilización de esta tarjeta es el siguiente: supongamos que tenemos los optoacopladores conectados a unos interruptores, y en el puerto B (puerto de salida) tenemos motores, con el siguiente fragmento de código lograríamos que al prenderse un interruptor, el correspondiente motor arranque [2].

Con este ejemplo se quiere mostrar la versatilidad que brinda la tarjeta propuesta, y lo fácil y rápido que se puede crear un sistema con este proyecto.

En cuanto a los circuitos impresos, hay una compañía canadiense llamada Alberta Printed Circuit Boards, que puede hacer 6 de estas tarjetas por alrededor de \$150.00.

VI. ANALISIS DE COSTOS

Uno de los objetivos del proyecto era lograr que éste tuviera un bajo costo. La Tabla 3 muestra la lista de los componentes y su precio aproximado en el mercado [3][4].

En cuanto a los circuitos impresos, hay una compañía canadiense llamada Alberta Printed Circuit Boards, que puede hacer 6 de estas tarjetas por alrededor de \$150.00. Hay que ver que sólo hay 6 tarjetas MC68HC11-EVB en la Universidad.

Tabla 3. Precios de los Componentes Electrónicos.			
DISPOSITIVO	CANTIDAD	PRECIO	TOTAL
Transistores Darlington	4	0.17	0.68
Optoacopladores	8	0.39	3.12
Pantalla LCD	1	22.90	22.90
Controlador de teclado	1	5.95	5.95
Relés	4	2.25	9.00
Sip de Resistencias 10K	2	0.25	0.50
Resistores de 4.7K (ciento)	1	0.89	0.89
Conectores de 20 contactos	1	0.75	0.75
Decodificador 74HC138	1	0.45	0.45
TOTAL		\$44.24	

En base a estos precios, podemos establecer un total de \$416.00 para la confección de 6 tarjetas con todos los accesorios. Estos precios son tomados de catálogos de compañías de Estados Unidos.

VII. CONCLUSIONES Y RECOMENDACIONES

Basado en todo lo que se puede ver y aprender a través de la INTERNET, puede concluir entre otras cosas que debemos como Universidad ponernos al día, e incentivar este tipo de proyectos. El precio final para realizar 6 tarjetas de estas, no es caro, y si lo vemos del punto de vista del avance tecnológico y del aprendizaje que se puede obtener, menos caro es. ¿CUANTO CUESTA LA EDUCACION?. Además el proyecto contiene los dispositivos básicos de control que se utilizan en el campo industrial, para instrumentación y control.

Se espera que el proyecto funcione perfectamente, ya que la parte más difícil ya está hecha, la cual es la que tiene que ver con el microcontrolador. En lo que respecta a la programación, como mostró el ejemplo, será sumamente fácil y rápido crear sistema con diferentes dispositivos. Si seguimos haciendo cosas como ésta, compañías como Motorola se seguirán interesando en donar sistemas de evaluación.

Con este proyecto el tiempo de aprendizaje y construcción de experimentos se reduce aproximadamente en más de un 50%, pudiendo utilizar las tres (3) primeras

En lo que respecta a la programación, será sumamente fácil y rápido crear sistema con diferentes dispositivos.

semanas para aprender a programar, y las 11 que restan seguir programando, sin casi armar nada en protoboard.

El autor considera que el proyecto expuesto aún puede pulirse más, con un poco de tiempo, asesoramiento y presupuesto, de manera que puede hacerse intercambio con otras universidades.

Uno de los siguientes pasos con este proyecto es su proyección en INTERNET, para recibir retroalimentación de diferentes formas de pensamiento alrededor del mundo.

VIII. REFERENCIAS

- [1] MC68HC11EVB Evaluation Board User's Manual, Motorola. 1986.
- [2] Michael Kheir. The M68HC11 Microcontroller. Application in Control, Instrumentation, and Communication, First Edition. Prentice-Hall, 1997.
- [3] JAMECO Electronics Components Catalog. February-April 199.
- [4] DIGI-KEY Catalog. April-June 1999.





III CONGRESO NACIONAL DE RAMAS DE ESTUDIANTES DEL IEEE

Rama de Estudiantes del IEEE de Málaga

ieee@etsit.uma.es



1. - INTRODUCCIÓN

Del 7 al 9 de marzo de 2001 tuvo lugar en la ETSI de Telecomunicación de la Universidad de Málaga el Tercer Congreso Nacional de Ramas de Estudiantes del IEEE, recogiendo así el testigo de los previamente celebrados en Barcelona y en Valencia.

Había si cabe más expectación ya que en el último año ha aumentado el número de Ramas existentes en España. Así, junto a las ya conocidas Ramas de Madrid, Barcelona, Valencia y Málaga, participaron las recién creadas Ramas de la Universidad Carlos III de Madrid y de la Universidad Rovira y Virgili. Asimismo, asistió un representante de la ya virtualmente constituida Rama de la Universidad Complutense de Madrid. Sólo hubo una ausencia, la de la Rama de Estudiantes del IEEE de Las Palmas de Gran Canaria.

2. - INAUGURACIÓN

Desde la tarde-noche del 6 de marzo, cuando llegó D. Jesús Brezmes, Presidente de la Rama de Estudiantes del IEEE de la Universidad Rovira y Virgili, empezaron a llegar los representantes de las diferentes Ramas. El grueso de asistentes llegó a Málaga la mañana del 7 de marzo. De esta forma, a las 14:00 todos, salvo Dña. Ana García Armada, Consejera de la Rama de Estudiantes del IEEE de la Universidad Carlos III de Madrid, y los representantes de la Rama de Barcelona, almorcamos en compañía de nuestro Consejero, D. Francisco Sandoval Hernández, y del Director de nuestra Escuela, D. Antonio Puerta Notario.

A continuación, a las 16:00 tuvo lugar en la Sala de Juntas el acto de inauguración del Congreso en el que tuvieron cabida una palabras de introducción por parte del Director de la Escuela, nuestro Consejero y nuestro Presidente. Finalizado la presentación, tuvo lugar una visita por la Escuela guiada por el Director, nuestro Consejero, en calidad de Director del Departamento de Tecnología Electrónica, y por el Catedrático D. Carlos Camacho Peñalosa, Director del Departamento de Ingeniería de Comunicaciones. Durante el recorrido, los asistentes pudieron ver los laboratorios tanto docentes como de I+D

a la vez que fueron informados de las líneas de investigación seguidas en nuestra Escuela.

3. - ACTIVIDADES DESARROLLADAS

La misma tarde en que comenzó el Congreso tuvo lugar una sesión en la que se resumió la evolución de los acuerdos adoptados el año anterior en Valencia. Se siguió advirtiendo como principal problema la falta de una completa y fluida comunicación entre todas las Ramas a la vez que se volvió a establecer un compromiso en firme para reactivarla mediante medios como la lista de correo electrónico "ramashispania".

Al día siguiente, por la mañana, tuvo lugar una sesión en la que se debatieron las distintas formas que tienen las Ramas para poder financiarse. Acabada esta sesión, los representantes de cada Rama ofrecieron una presentación sobre las mismas.

Posteriormente, a primera hora de la tarde, el Dr. D. Fabián Arrebola Pérez, Profesor del Departamento de Tecnología Electrónica, nos ofreció una Conferencia sobre Visión Foveal, una de las principales líneas de investigación de su departamento. Acabada la conferencia, volvimos a reunirnos esta vez para debatir sobre posible actividades que podríamos realizar todos en conjunto. Durante la sesión se rechazó por unanimidad la organización del SBC 2002, idea que había surgido la semana anterior en Sevilla con motivo de la reunión de la Junta Directiva de la Región 8 del IEEE. De igual modo, se decidió que la próxima sede del Congreso sería Madrid y que sería organizado por la Rama de la UPM.

A lo largo de la tarde llegó a Málaga D. Juan Jesús Rodríguez Yubero, recientemente nombrado responsable de actividades de estudiantes de la Sección Española y también tuvimos que despedir a los representantes de la Rama de la Universidad Carlos III de Madrid y de la Universidad Rovira i Virgili.

El viernes por la mañana nos dirigimos al Parque Tecnológico de Andalucía (PTA) con el fin de realizar una visita. En primer lugar, fuimos recibidos en la empresa Atlinks, Joint Venture de Alcatel y Thomson Multimedia,

dedicada a la fabricación de teléfonos como el de sobra conocido “Domo”. El ingeniero D. Francisco Juárez, antiguo profesor de la Escuela, nos guió por las instalaciones de la empresa. Acabada esta visita, fuimos recibidos por D. Felipe Romera Lubias, Director General del PTA. Tras ofrecernos una presentación del Parque, fuimos invitados a almorzar en la Sede Social del PTA. Durante el transcurso de la presentación se incorporó al Congreso Dña. Magdalena Salazar Palma, Presidenta de la Sección Española del IEEE, y finalizando el almuerzo, también se incorporaron D. Javier Macías Guarasa, Consejero de la Rama de Estudiantes del IEEE de Madrid, D. Jens Hannemann, editor de la “R8 Student Newsletter” y su esposa, Dña. Regina Hannemann. A continuación, nos dirigimos a Cetecom, empresa dedicada a un amplio espectro de actividades relacionadas con la Ingeniería de Telecomunicación. Fuimos recibidos por D. Joaquín Torrecilla Torregrosa, Director de Marketing de la empresa y al mismo tiempo profesor de nuestra Escuela. Tuvimos oportunidad de contemplar todos los medios con los que Cetecom cuenta, destacando por encima de todos la gigantesca cámara semi-anecoica que emplean en ensayos de compatibilidad electromagnética.

Finalizada la visita al PTA, volvimos a la Escuela,

donde tuvo lugar una sesión de clausura en la que a los que no habían estado presentes previamente se les hizo una recapitulación a modo de resumen de lo debatido durante el Congreso.

Para poner el broche de oro al Congreso y, en la línea de las noches anteriores, en las que salimos todos a cenar fuera y tomarnos una copa con el fin de estrechar lazos, tuvo lugar una cena de despedida en el Parador Nacional de Gibralfaro, junto al Castillo, desde donde se divisa una impresionante panorámica de nuestra ciudad.

4. - CONCLUSIONES Y AGRADECIMIENTOS

Nos gustaría resaltar aquí el grado de satisfacción que tenemos en Málaga al haber organizado este Tercer Congreso, actividad que creemos que está ya plenamente consolidada y que esperamos que tenga lugar en múltiples ocasiones posteriores.

Asimismo nos gustaría aquí agradecer públicamente a todas las personas que hicieron posibles estos tres días en Málaga y en especial a todos los participantes que nos honraron con su visita.



Imagen de los asistentes al Congreso de Málaga, tras la cena de clausura en el Parador de Gibralfaro.





NETEXPERT: SISTEMA BASADO EN INTELIGENCIA ARTIFICIAL PARA LA LOCALIZACIÓN DE EXPERTOS

Josep M. PUJOL SERRA y Ramon SANGÜESA SOLÉ

Josep M. Pujol es Ingeniero en Informática por la FIB-UPC y estudiante del doctorando de Inteligencia Artificial del LSI-UPC.

Ramon Sangüesa Solé es Doctor por la UPC en Inteligencia Artificial

e-mail: {jmpujol,sanguesa}@lsi.upc.es

URL: <http://www.lsi.upc.es/~{jmpujol,sanguesa}>

1.- ABSTRACT

Localizar las competencias y conocimientos de los miembros de una comunidad es un problema clave para mejorar la gestión del conocimiento en un grupo. En la actual sociedad, inundada por grandes cantidades de información, las personas que tengan conocimientos en un tema concreto, (también llamadas expertos), pueden ayudar a ahorrar recursos y esfuerzos. Encontrar el experto adecuado y saber cómo poder llegar a él son problemas que siempre han existido pero, debido a las nuevas tecnologías de la información y la comunicación, el problema está adquiriendo mas y mas importancia ya que las nuevas tecnologías ocultan gran parte del conocimiento sobre el resto de nuestra comunidad (también llamado entorno social).

NetExpert es un sistema basado en agentes autónomos que permite buscar a un experto y descubrir cómo podemos acceder a él a través de dos interfaces de usuario: o bien con un *web browser* (*Applet*), o bien con un teléfono móvil WAP. El sistema NetExpert muestra: 1) quién tiene el conocimiento necesario y 2) el mejor camino para poder acceder al experto mediante la creación de una red social (*Social Network*). Saber que miembro de la comunidad tiene conocimientos sobre un tema es vital pero también es vital saber quien conoce a quien para poder acceder a un determinado experto. Por lo tanto NetExpert construye no solo una Red Social sino también una Red de Conocimientos que descubren el entramado oculto de una comunidad.

2.- INTRODUCCIÓN

Localizar quién tiene el conocimiento necesario para realizar una determinada tarea es un problema de importancia vital. En el futuro será un factor crítico, a causa del aumento de grupos de trabajo distribuidos propiciado por las nuevas tecnologías de la información y las comunicaciones.

Para poder ser más eficientes es básico gestionar el conocimiento que se genera así como su posterior reutilización. En cualquier entorno los problemas a resolver tienen similitud a problemas resueltos anteriormente, poder recuperar el conocimiento que se generó durante la resolución del problema anterior puede ser de ayuda para

resolver el problema actual. La buena gestión del conocimiento permite también dar soporte a la decisión, puesto que cualquier decisión que se tome en un tema debe estar respaldada por un conocimiento en dicho tema. La disciplina conocida como *Knowledge Management* [1,2,3] (Gestión de Conocimiento) se encarga de los métodos y tecnologías necesarias para este tipo de situaciones.

Para una buena gestión del conocimiento no basta con crear una cantidad ingente de documentación. Ya existen unos vastos repositorios de información y muchas maneras de acceder a ellos. Pero ¿cuál es la pieza de información clave que nos permitirá resolver nuestro problema? El tiempo y el esfuerzo son recursos limitados que no pueden ser malgastados analizando información irrelevante. Entendemos por información relevante aquella que puede ayudarnos a resolver un problema o tomar la decisión correcta. Se debe primar el concepto de *calidad* por encima de la *cantidad* de información. ¿Cómo podemos discernir la calidad de la información o del conocimiento si no tenemos un buen conocimiento del dominio? Precisamente los expertos tienen un gran conocimiento de su dominio ganado con el tiempo y la práctica: son capaces de filtrar la información obtenida tras una fase de búsqueda y acopio, encontrando la que necesitan para resolver el problema. De momento solo disponemos de expertos humanos pero todo cambia.

Un experto es alguien con un gran conocimiento y experiencia sobre un determinado dominio. Cabe remarcar que un experto no tiene porqué ser doctor ni ingeniero, (por ejemplo, un experto puede ser aquel que sabe desatascar la fotocopiadora). La utilidad de los expertos para resolver problemas es evidente. ¿Qué es más eficiente en términos de rendimiento, buscar en un gran repositorio de información o bien hablar un rato con un experto? ¿Que decisión da mas confianza, aquella tomada solamente por nosotros o aquella tomada con la colaboración de un experto? [4]

Podríamos pensar que los expertos no son indispensables ya que existen grandes repositorios de información y conocimientos que pueden ayudarnos a resolver los problemas que surgen, pero vivimos en un mundo muy competitivo y de recursos limitados, donde aparte de tener que resolver problemas también se deben resolver antes que el resto de competidores.



LOCALIZAR UN EXPERTO

El objetivo de resolución de un problema se transforma en el de encontrar un experto que lo puede resolver o ayudar a resolverlo pero, ¿cómo encontramos a un experto?

Primero se debe remarcar que un experto es una persona reconocida como tal dentro de una comunidad [5]. ¿Es útil la información de la comunidad para localizar expertos? Vamos a intentar mostrar que, en efecto, lo es y cómo las tecnologías de la información y la Inteligencia Artificial pueden ayudar a usar esta información de la comunidad para localizar expertos.

La utilización del término *localizar* en lugar de *buscar* no es arbitraria. Con una pequeña licencia lingüística tomamos el término *localizar* como la composición de dos operaciones. Una es *buscar*, buscar a alguien que tiene el conocimiento suficiente para resolver un problema. La segunda es *saber cómo llegar al él*. El problema de localizar un experto se puede descomponer en dos partes:

- Encontrar quien tiene el conocimiento necesario. *Who knows what?* [6]
- Saber como podemos llegar a contactar con el experto. *Who knows who knows what?* [6]

Cuando tenemos un problema pensamos en quién es la mejor persona para preguntarle sobre cómo solucionarlo. Si no conocemos a nadie entonces preguntamos a nuestros conocidos; de hecho estamos trasladando el problema de localizar un experto a nuestros conocidos. Supongamos que finalmente tenemos éxito y damos con alguien que nos puede ayudar ¿Cómo podemos pedirle ayuda si seguramente no lo conocemos o no tenemos la confianza suficiente? Para salvar este escollo las personas utilizamos las *cadenas de referencias* [7]. No accedemos directamente a la persona sino que utilizamos a otras personas para que nos presenten. Por ejemplo, si queremos acceder a una persona *X* puede que no la conozcamos pero esa persona puede tener un conocido *Y* que a su vez sea conocido nuestro. En tal caso utilizaremos a *Y* para que nos presente y así obtener la atención de *X*.

Para resolver el problema de encontrar un experto y saber como llegar a él hay que recurrir a dos conceptos clave.

- **Knowledge Network (Red de Conocimientos):** Es un grafo que relaciona personas que comparten conocimiento. Las personas son nodos. Una arista del grafo significa que las dos personas relacionadas tienen conocimientos similares. Esta es la red que utilizamos para saber quien sabe de algún tema. El concepto de Knowledge Network fue

utilizado por Noshir Contractor y Ann Bishop en el sistema IKNOW [6], desarrollado en la Universidad de Illinois.

- **Social Network (Red Social):** Es un grafo que relaciona personas que se conocen. Las personas son nodos. Una arista del grafo significa que las dos personas relacionadas se conocen aunque no comparten conocimientos. El concepto de Social Network fue utilizado por Henry Kautz en el sistema Referral Web [7], desarrollado por el departamento de investigación de ATT.

Toda comunidad tiene estos dos tipos de redes. Ahora bien, una persona aislada sólo tiene una visión parcial y sesgada de las estructuras sociales y de los conocimientos de los miembros de la comunidad a la que pertenece. Una persona sólo conoce bien la información que le atañe a si mismo. Y seguramente también tendrá un buen conocimiento de las personas con las que tiene una relación muy fuerte. Actualmente las tecnologías de la información y la comunicación favorecen el trabajo distribuido, los nuevos canales de comunicación como el email, videoconferencias, videochats, etc. permiten comunicarnos eliminando las barreras físicas pero a cambio nos ocultan el entorno de nuestro interlocutor. Gran parte del conocimiento que nosotros tenemos de los miembros de la comunidad real se extrae de la observación su entorno, este entorno se ve reducido en los sistemas virtuales

3.- DESCRIPCIÓN DEL SISTEMA NETEXPERT

A fin de resolver el problema de la localización de expertos en una comunidad virtual y hacer más eficiente la tarea de localizarlos se ha desarrollado el sistema NetExpert.

NetExpert es un sistema, basado en varios agentes autónomos [8,9], que permite localizar expertos dentro de una comunidad formada por investigadores (el concepto de agente será descrito posteriormente en la arquitectura). Esta comunidad de investigadores está compuesta por diversas personas de diferentes campos de investigación como Agentes, Seguridad, Banda Ancha, etc.

Cabe resaltar que en esta comunidad los miembros son expertos en sus correspondientes áreas de trabajo.

NetExpert permite encontrar qué miembros de la comunidad tienen conocimiento sobre un determinado tema y también cómo podemos llegar a ellos a través de la Red Social de la comunidad. El sistema NetExpert se integra dentro de otro sistema llamado *Colaboratorio* [10,11] desarrollado por la UPC dentro del proyecto de Internet2 Catalunya (i2CAT) [12].

La Figura 1 muestra la interficie de usuario Applet del sistema NetExpert.

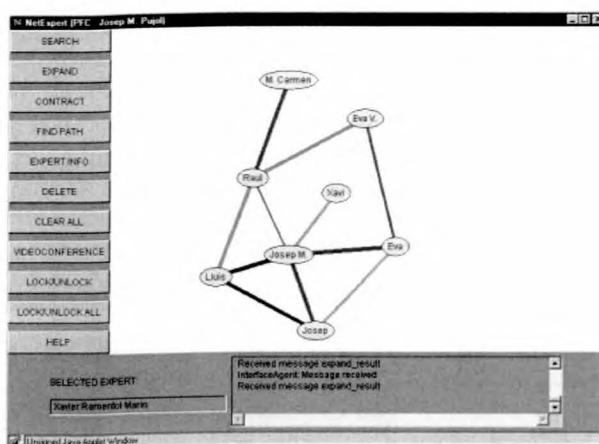


Figura 1: Pantalla principal de la interficie de usuario en Applet.

En la pantalla de la Figura 1 podemos observar que hay diversos expertos, las relaciones entre los cuales pueden ser de diversos tipos y tener diferentes importancias. Vemos también en la Figura 2 unas pantallas de la interficie de usuario WAP. Esta interficie de usuario se ha debido simplificar al máximo debido a las limitaciones de los dispositivos WAP.



Figura 2: Diversas pantallas de la interficie de usuario WAP

La mejor manera para mostrar las capacidades del sistema NetExpert es describir brevemente sus funcionalidades.

FUNCIONALIDADES DE NETEXPERT

a) Buscador

NetExpert permite realizar búsquedas de un experto utilizando palabras de la misma forma que en un motor de búsqueda convencional como Google, Altavista, etc. Existen cuatro tipos de búsqueda:

- **Name:** Búsqueda por el nombre o apellido de los expertos. Los resultados son expertos.
- **Knowledge profile:** Búsqueda por el perfil de conocimientos. Todos los expertos, que son miembros de la comunidad, tienen un perfil de conocimientos, generado automáticamente por el sistema. Utilizan-

do técnicas de aprendizaje automático. Los resultados de la búsqueda son expertos.

- **Web:** Búsqueda por el contenido de las webs personales de los expertos de la comunidad. Los resultados de la búsqueda son expertos.
- **Resources:** Búsqueda por el contenido de las webs personales. Los resultados son recursos web. Un recurso web es cualquier fichero accesible a través de la web. Los recursos web que el sistema NetExpert devuelve como resultado de una búsqueda no se limitan a páginas web como la gran mayoría de motores de búsqueda, también pueden ser documentos pdf, doc, archivos comprimidos, ficheros de vídeo, de audio, etc. Así pues, podemos realizar una búsqueda y que el resultado sea una imagen jpg o bien un documento word. Por poner un ejemplo, el segundo resultado por importancia de la búsqueda «mapa campus nord» es una imagen jpg que contiene un mapa de los edificios de la UPC cercanos a la diagonal incluido el *Campus Nord*.

b) Obtener información del experto

Podemos obtener información sobre los miembros de la comunidad. Esta información puede ser desde información personal como email, nombre, numero de teléfono hasta información de los documentos que el experto ha publicado en la comunidad.

c) Comunicarse con un experto

NetExpert permite ponerse en contacto con el experto sin tener que abandonar el sistema. Podemos contactar con el experto vía email, videoconferencia de altas prestaciones. Finalmente también podemos contactar con el experto con una llamada telefónica desde nuestro móvil en caso de utilizar la interficie de usuario WAP.

d) Explore the Knowledge and Social Networks

NetExpert permite navegar por las redes de conocimientos y social mediante dos simples operaciones: expandir y contraer. Cuando seleccionemos un experto podemos ver con qué miembros de la comunidad comparte conocimientos (*Knowledge Network*) y a quién conoce (*Social Network*). Gracias a las redes sociales podemos descubrir información sobre los miembros de la comunidad que estaba oculta para nosotros. Sacar provecho del conocimiento descubriendo depende de nosotros.

e) Encontrar el camino óptimo

El camino óptimo entre dos expertos es la mejor cadena de referencia para ir de un experto hacia cual-



quier otro utilizando las relaciones sociales. Muchas veces cuando encontramos un experto que se ajusta a nuestras necesidades pero puede que no le conozcamos y no sepamos como llegar a él. Utilizando la red social podemos descubrir si tenemos algún conocido en común que nos pueda presentar. Una manera más directa de descubrir este camino es utilizando esta funcionalidad del NetExpert.

La optimalidad del camino viene calculada por un heurístico que considera la distancia del camino, (es decir, el numero de pasos que debemos hacer), y la importancia de las relaciones entre los miembros del camino. El camino óptimo resultante debe minimizar la distancia del camino y maximizar el peso de las relaciones de los integrantes del mismo.

Las comunidades de personas tienden a tener una topología característica denominada *Small Worlds* [13,14]. Una de las características de esta topología es que tiene una distancia media entre nodos pequeña, es decir, el numero de pasos que se deben hacer para ir de un nodo (experto) a cualquier otro es pequeño.

Veamos a continuación algunas pantallas de la interficie de usuario Applet del sistema:

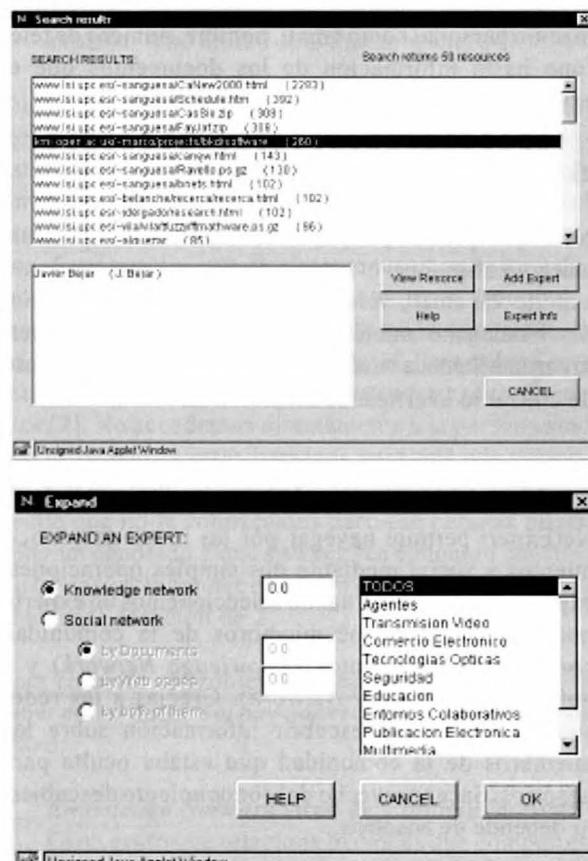


Figura 3: En la primera pantalla podemos ver el resultado de una búsqueda por recursos (observar que el 3er resultado es un .zip). En la segunda pantalla, el diálogo de la operación expand.

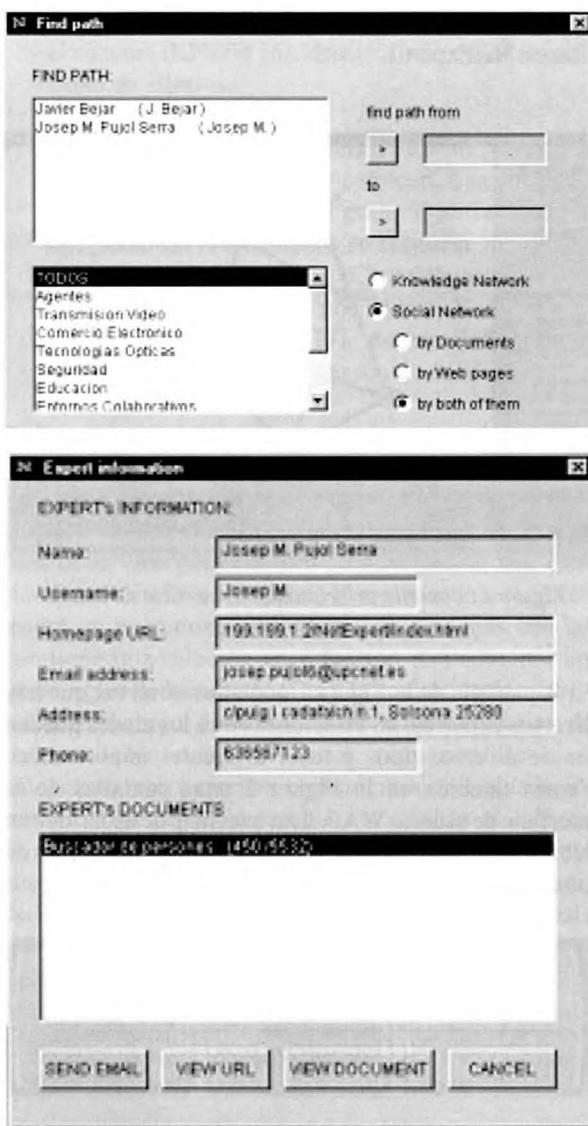


Figura 4: En la primera pantalla podemos ver el diálogo para encontrar el camino óptimo. En la siguiente, el diálogo donde se puede obtener información personal sobre un experto.

4.- CONOCIMIENTO GENERADO POR NETEXPERT

Hasta ahora se ha mostrado como el sistema NetExpert puede ayudar a resolver el problema de la localización de expertos. Para realizar su tarea NetExpert necesita diversas fuentes de información a partir de las cuales ha de generar el conocimiento. Todo este proceso de extracción de conocimiento a partir de información se realiza de forma implícita, sin la intervención de ninguna persona. Para conseguirlo se han utilizado técnicas y métodos de inteligencia artificial y aprendizaje automático.

Parte del conocimiento que usa NetExpert se genera en cooperación con el sistema Colaboratorio, los dos sistemas generan conocimiento con técnicas avanzadas de inteligencia artificial. Veamos pues el conocimiento que

crean conjuntamente el sistema NetExpert y el Colaboratorio.

- **Knowledge profile:** Perfil de conocimiento. Para modelar el perfil de un experto, que a su vez es miembro de la comunidad de investigadores, se ha usado la técnica del *Vector Space Model* con TF-IDF [15,16]. Esta técnica se basa en tener un vector de palabras relevantes que modelan el experto. La relevancia de una palabra viene marcada por el TF-IDF (*Term Frequency x Inverse Document Frequency*). Las palabras que se usan para modelar el conocimiento de un experto se extraen de diversas fuentes de información:
- **Documentos que el experto ha publicado en la comunidad.**
- **Documentos que el experto lee a través del sistema.** Una vez se ha leído un documento el usuario debe valorarlo; de esta forma se puede considerar el interés que dicho documento tenía para el experto.
- **Intereses declarados:** A fin de crear un primer perfil de conocimientos el experto debe declarar sus intereses. Esta parte es explícita y sólo sirve para inicializar el perfil de conocimiento, ya que el perfil se construye a medida que el experto lee y publica dentro de la comunidad.

Análisis de la web personal: Otra forma de representar los conocimientos y los intereses de un miembro de la comunidad es a través del análisis de su web personal. Cada página o recurso tiene un modelo construido a partir de palabras con su TFIDF. Las páginas web personales normalmente representan el conocimiento y los intereses de sus autores, más aún cuando los autores son investigadores. Cabe recordar que la comunidad donde trabaja NetExpert está formada por investigadores. Para analizar la web NetExpert utiliza el sistema WebMining [17]. WebMining es un sistema genérico e independiente que analiza la web y crea modelos de los recursos analizados.

- **Knowledge Network:** La red de conocimientos es un grafo donde los nodos son expertos y las aristas indican similitud de conocimientos. Las aristas pueden tener un peso en el intervalo (0..1] que indica la medida de similitud del conocimiento de los dos expertos. Para calcular la similitud de conocimientos de dos expertos se utiliza el coseno similitud [18], que es el producto escalar entre los dos vectores de palabras que modelan el conocimiento del experto.
- **Social Network:** La red social es también un grafo. Una arista indica que los dos expertos se conocen: esto no quiere decir que sean amigos, simplemente se conocen el uno al otro. La arista también tiene un

peso en el intervalo (0..1] que indica el grado de conocimiento mutuo que se tiene, por ejemplo: dos personas que van a comer juntos se conocen aún cuando sus intereses y conocimientos sean distintos.

Para construir la red social hemos creado algunos heurísticos que son capaces de evaluar el nivel de relación entre dos personas de la comunidad. Estos heurísticos consideran los siguientes conceptos para construir la red social:

- **Autoría de los documentos publicados:** Los coautores de un documento normalmente se encuentran entre ambos.
- **Relaciones entre las webs personales:** La web personal es una fuente de información muy útil para conocer la relación con otros miembros de la comunidad. En las páginas web personales se pueden encontrar referencias a otras web personales de otros miembros de la comunidad. Estas referencias pueden ser:
 - Direcciones de email de otros miembros de la comunidad.
 - Links hacia las páginas personales de otros miembros de la comunidad.
 - Links hacia recursos que están dentro de las webs personales. En estos casos se considera la profundidad de los recursos en el árbol que forma una web personal.

Las relaciones entre webs personales y la co-autoría de documentos son los criterios usados para construir la red social. Estos criterios son dependientes de la comunidad: no es lo mismo una comunidad de investigadores que una comunidad de cinéfilos. Los investigadores son gente que publica un gran número de documentos, papers, reports y que normalmente tienen una buena web personal donde describen sus áreas de investigación. Cada tipo de comunidad puede requerir diferentes heurísticos que trabajen sobre los conceptos relevantes para esa comunidad concreta.

- **Camino óptimo:** El camino óptimo es el mejor camino para ir de un miembro de la comunidad hacia otro utilizando las relaciones sociales. El camino óptimo no tiene por qué ser el camino más corto. La optimalidad se calcula mediante un heurístico que considera dos conceptos:
- **Longitud del camino:** Es la distancia entre los dos nodos extremos del camino, también se puede ver como el número de saltos que se debe hacer para ir de un extremo al otro.
- **Pesos de las relaciones:** El camino tendrá mayor credibilidad si los miembros que lo forman tienen relaciones fuertes.



5.- ARQUITECTURA INTERNA DEL NETEXPERT

NetExpert es un sistema formado por diversos agentes autónomos, y por lo tanto también es un sistema distribuido. Las entidades software, sean agentes o no, pueden estar en distintas máquinas que a su vez pueden estar en distintas plataformas.

El término *agente* ha sido sobreutilizado. Por este motivo la definición del concepto de agente ha generado mucha polémica existen diversas definiciones del concepto agente [17]. Aun que no existe una definición universalmente aceptada se ha llegado a unos puntos de acuerdo para toda la comunidad de investigadores.

Los agentes son sistemas computacionales que:

- Son persistentes.
- Habitán en un entorno dinámico durante su existencia.
- Interactúan con el entorno. Pueden captar información del entorno mediante sensores. También puede realizar acciones que modifiquen este entorno.
- Tienen metas o objetivos que tienen que cumplir o intentarlo.
- Toman decisiones de forma autónoma.

La definición de agente que acabamos de realizar es conocida como la concepción débil de agentes, el sistema NetExpert usa esta. La concepción fuerte de los agentes les atribuye otras propiedades que ya son consideradas como noción mentales: creencias, intenciones, obligaciones. En la concepción fuerte de agentes estos deben tener actitudes emocionales, *Emotional Agents* [8].

La Figura 5 muestra como es la arquitectura interna del sistema NetExpert.

Podemos observar que hay diversos componentes en el sistema NetExpert. Los tipos de comunicación entre estos componentes se muestran en el diagrama de la Figura 6.

- Comunicación a través de ACL (*Agent Communication Language*). Esta es la comunicación entre las entidades software que son agentes autónomos.
- - - → Comunicación entre personas y componentes software a través de una interficie de usuario (UI).
- → Comunicación entre entidades software agentes y los componentes software no agentes. Esta comunicación se realiza invocando métodos (programación convencional).

Figura 6: Tipos de comunicación.

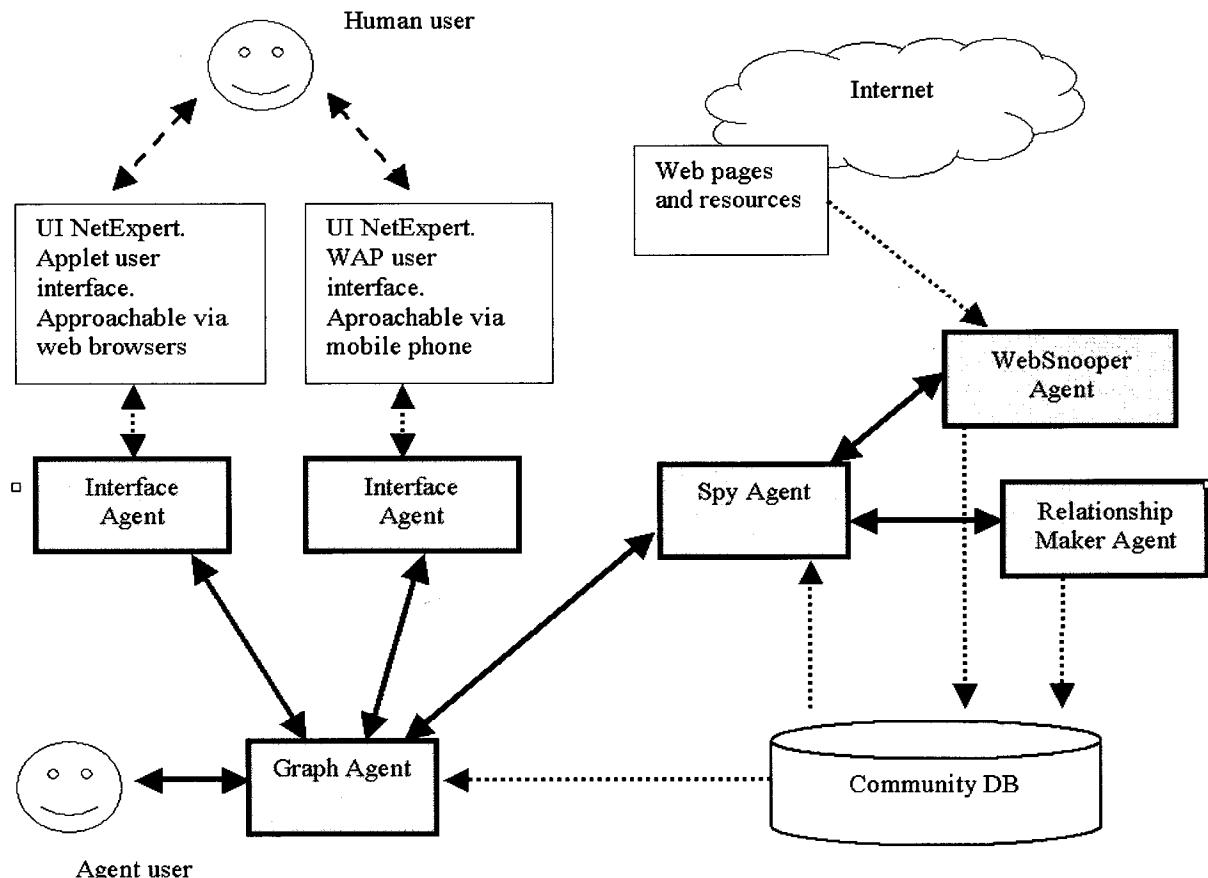


Figura 5: Arquitectura interna del sistema

En la Figura 5 se observa que el sistema NetExpert puede tener dos tipos de usuarios:

- **Usuarios humanos:** Este tipo de usuarios tiene dos vías para acceder a NetExpert, a través de Applets o bien a través de páginas WML. Todas las formas de acceder a NetExpert usan programas cliente ligeros, un navegador o un teléfono móvil WAP, respectivamente.
- **Usuarios agentes:** Este tipo de usuarios son agentes inteligentes que circulan por la red. Estos pueden obtener el conocimiento y la información generada y almacenada por el sistema a través del Graph Agent. Los agentes exteriores que hacen de usuarios del sistema deberán hablar el mismo lenguaje que el Graph Agent.

Los componentes grises en el diagrama de la arquitectura (Figura 5) son agentes software. NetExpert es un sistema multi-agente formado por cinco tipos de agentes. Veamos una breve descripción:

- **Interface Agent:** Este agente es un adaptador entre la interficie de usuario y el Graph Agent. Cualquier usuario humano que quiera entrar en el sistema a través de la UI necesita una adaptador. Cada usuario dentro del sistema utiliza una UI. Cada UI necesita un Interface Agent para comunicarse con el Graph Agent. Tanto el UI como el Interface Agent se ejecutan en el ordenador del cliente así que el sistema no se sobrecarga con los procesos de los usuarios.
- **Graph Agent:** Este agente encapsula todo el conocimiento y la información del sistema NetExpert y es el encargado de servir este conocimiento a aquel que se lo pida.
- **Spy Agent:** Este agente es el encargado de buscar cambios en la base de datos de la comunidad y actuar en caso que se hayan producido alteraciones.
- **RelationshipMaker Agent:** Este agente se encarga de construir y mantener las relaciones entre los miembros de la comunidad, las redes.
- **WebSnooper Agent:** Este agente se encarga de analizar las webs personales de los miembros de la comunidad. Para realizar esta tarea utiliza el sistema WebMining [17].

NOTAS DE DISEÑO

Los agentes han sido desarrollados con JATLite [18] pero como lenguaje de comunicación entre agentes no se ha utilizado el KQML que es el lenguaje estándar en JATLite

sino el FIPA ACL [19]. La fusión de JATLite con FIPA ACL ha sido realizada por el EPFL [20]. Todos los agentes y los componentes software no agentes han sido programados en Java a fin de asegurar la portabilidad entre diferentes plataformas. El sistema gestor de base de datos de la comunidad es MySQL, existen implementaciones de MySQL para las plataformas más utilizadas: Unix, Linux, Windows.

6.- EXPERIMENTACIÓN

NetExpert es un sistema que trabaja conjuntamente con el Colaboratorio, un sistema colaborativo que requiere muchos usuarios y la participación de estos para funcionar correctamente. Por este motivo se ha probado NetExpert con los miembros del Departamento de *Llenguatges i Sistemes Informàtics* (LSI) de la *Universitat Politècnica de Catalunya* (UPC). Este test solo muestra la *Social Network* (Red Social) generada a través del análisis de las webs personales ya que es información completamente independiente del Colaboratorio. Para probar todas las funcionalidades del sistema se han creado juegos de prueba que han sido superados con excelentes resultados.

El juego de pruebas del LSI tiene información sobre los 170 miembros del departamento. De los 170 miembros 104 tienen web personal y el resto no la tienen o tienen la web personal por defecto (*impersonal page*). La muestra tiene 104 individuos. La figura 7 muestra un gráfico con el tamaño de las webs personales. Para cada individuo de la muestra tiene el tamaño (es decir, el numero de recursos) de su web personal. Los individuos están ordenados por tamaños de la web personal.

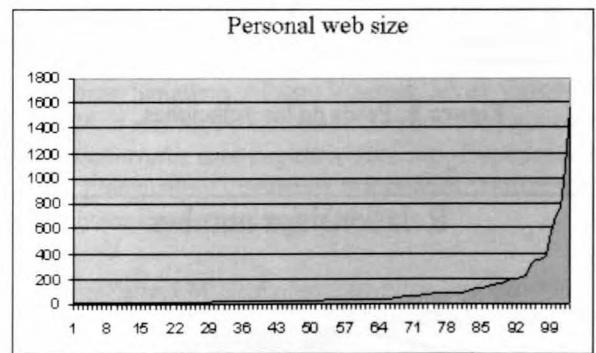


Figura 7: Tamaño de las webs personales

Podemos observar que existe una gran variabilidad, la mayoría de los miembros de esta comunidad tiene webs personales pequeñas y hay un pequeño grupo que tiene webs personales muy grandes. La frontera entre una web

personal grande y pequeña es muy difusa pero se podría considerar que una web personal de tamaño medio tiene aproximadamente cien recursos (esto no significa cien páginas, sino cien recursos modelizados por el WebMining ya sean páginas, ficheros, links, emails, etc.). Al existir tanta variabilidad la mediana nos aporta más información. El 50% de la muestra tiene webs personales de menos de 24 recursos: es decir, muy pequeña. Aun así en los siguientes gráficos podemos ver como el sistema extrae conocimiento del análisis de las webs personales.

NetExpert ha encontrado 190 relaciones sociales entre los 104 miembros del departamento que tienen web personal. El peso de las relaciones está en el intervalo (0..1]. La Figura 8 muestra el peso de las relaciones en cuartiles. La relación mayoritaria es *low-medium*, entre [0.25..0.5], en un 44%, las relaciones débiles (*weak*) y *high-medium* están en el 23% y el 21% respectivamente. En la Figura 9 se muestran el número de relaciones por individuo. La gran mayoría, un 69% están poco relacionados (entre 1 y 3 relaciones). Las siguientes clases, *medium related* y *high related* tienen el 12% y el 11% de los individuos de la muestra,

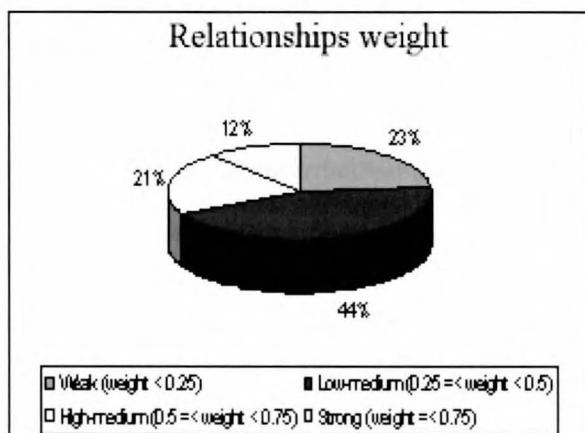


Figura 8: Pesos de las relaciones

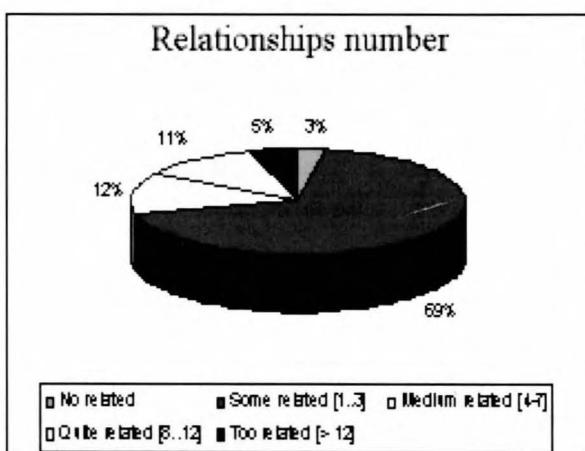


Figura 9: Relaciones por individuo

respectivamente. Solo el 3% de individuos han quedado aislados si ninguna relación. Así pues el sistema encuentra relaciones en el 97% de los individuos que tienen web personal. La clase mayoritaria es la *somewhat related* (poco relacionada) pero tenemos que recordar que esta no es la red social completa, falta tener en cuenta la co-autoría de documentos, que corresponde a datos existentes en el colaboratorio donde no están todos los miembros del LSI, sino otra comunidad distinta.

Hasta ahora hemos visto que el sistema era capaz de encontrar relaciones aun cuando las webs personales no fueran muy completas. Tenemos que analizar la calidad de las relaciones encontradas, es decir, si las relaciones se corresponden con la realidad. Para ello se ha pedido la opinión de algunos de los miembros de LSI. La opinión de estos fue muy favorable. Parece que las relaciones encontradas se corresponden con la realidad pero que el sistema no encuentra todas las relaciones posibles. Esto se debe en parte a que no se puede emplear en esta comunidad el criterio de co-autoría de documentos. La Figura 10 muestra un fragmento de la Red Social de los miembros de LSI.

Este fragmento de la Red Social se ha generado expandiendo a partir de los expertos R.Sangüesa y U.Cortés. Se ha realizado una expansión de estos expertos y para cada experto relacionado con los dos iniciales se ha realizado otra expansión. Así pues, cualquier persona de este grafo tiene una distancia menor igual a tres respecto cualquiera de los dos expertos iniciales.

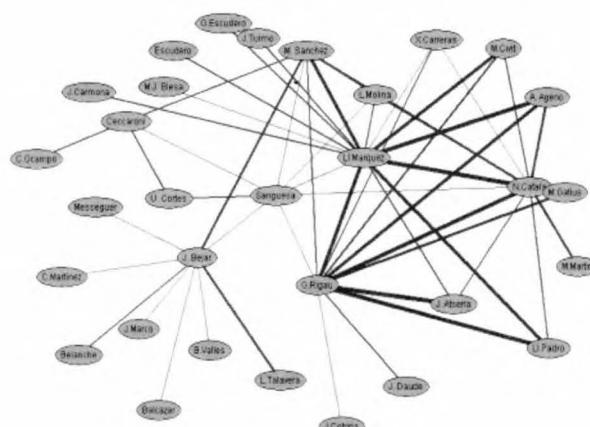


Figura 10: Fragmento de la Social Network de LSI

7.- CONCLUSIONES

En la sociedad del conocimiento la gestión de éste es un tema crítico. Saber quien sabe sobre un tema y saber cómo poder acceder a él es un gran problema que cada día que pasa se vuelve más importante.

El sistema NetExpert es un sistema que permite a un usuario de una comunidad extraer información y conocimiento de los otros miembros de la comunidad, información que estaba oculta a causa de la percepción parcial que cada uno tenemos del resto de la comunidad.

NetExpert cumple los siguientes objetivos:

- Permite buscar qué miembro de la comunidad tiene un determinado conocimiento por diferentes vías.
- Permite buscar por recursos que pertenecen a los miembros de la comunidad. Estos recursos pueden ser cualquier tipo de fichero ya sea página web, fichero de audio, vídeo, imágenes, etc. Y todo ello a partir de las palabras como en un buscador convencional.
- Permite conocer cuál es la mejor forma de acceder a otros miembros de la comunidad.
- Permite conocer con qué personas de la misma comunidad compartimos conocimientos e intereses.
- A fin de hacer mas cómoda la interacción entre los miembros de la comunidad NetExpert permite utilizar email, videoconferencia y llamadas telefónicas para comunicarse con los otros miembros de la comunidad sin tener que abandonar el sistema.

Sistemas como NetExpert o sus predecesores (IKNOW y ReferralWeb), nos ayudan a superar la visión parcial de la comunidad para poder aprovechar todo su potencial.

8.- REFERENCIAS

- [1] Drucker, P.F. Prusak, L. *Working Knowledge*. Boston: Harvard Bussiness Review Press. (1998).
- [2] Glance, N. and Arregui, D. and Dardenne, M. Knowledge Pump:supporting the Flow and Use of Knowledge. In Borghoff, U.M. and Pareschi, R.(Eds.), *Information Technology Support for Knowledge Management* . pp. 35-50. Springer-Verlag. Berlin. 1998.
- [3] Borghoff, U.M. and Pareschi, R.(Eds.), *Information Technology Support for Knowledge Management* Springer-Verlag. Berlin.1998.
- [4] Adriana S. Vivacqua. *Agents for Expertise Location*. Proceedings of the 1999 AAAI Spring Symposium on Intelligent Agents in Cyberspace. Stanford, California (March 1999).
- [5] Wenger, E. *Communities of Practice: Learning, Meaning, Identity* Cambridge University Press; (2nd Edition) 1999.
- [6] Contractor, N.S., O'Keefe, B.J., Jones, P.M. *IKNOW: Inquiring Knowledge Networks On the Web*. Computer software. University of Illinois. (1997).
- [7] Henry Kautz, Bart Selman. Mehul Shah. *The Hidden Web*. AI Magazine vol. 18 (Summer 1997).
- [8] Pattie Maes. *Artificial Life Meets Entertainment: Life like Autonomous Agents*. Communications of the ACM. (1995).
- [9] Bates, J. *The role of emotion in believable agents*. Communications of the ACM, 37(7):122-125. (1994).
- [10] Barrio, I., Pujol, J.M., Sangüesa, R., Vázquez, A. *An Agent-Based Collaborative Environment*. Universitat Politècnica de Catalunya, Software Department, 2001 (forthcoming).
- [11] Sangüesa, R., Vázquez, A., Barrio, I., Pujol, J.M. *Agents for Collaborative Support: the I2CAT experience*. To be presented at the SIG meeting on intelligent information agents. Agentlink, Amsterdam 23rd February 2001.
- [12] i2CAT, Internet 2 a Catalunya. Disponible a <http://www.canet.upc.es/i2-cat/>
- [13] Watts D.J. and Strogatz, S.H. *Collective dynamics of 'small-world' networks*. Nature 393, 1998.
- [14] Adamic, Lada. *The Small World Web*. Xerox Palo Alto Research Center.
- [15] Gerald Salton and Michael J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [16] Ramon Sangüesa, Alberto Vázquez, Javier Vázquez. *ACE: A multiagent recommender system using mixed collaborative and cognitive filtering*. Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya. (1999).
- [17] Pujol Serra, J.M. *NetExpert: Localitzador d'expertesa*. Final Project Thesis. Software Department. Technical University of Catalonia. December 2000.
- [18] Java Agent Template, Lite. Stanford University. Available on <http://java.stanford.edu>
- [19] FIPA. Foundation Intelligent Physical Agents. Available on <http://www.fipa.org/>
- [20] The JATLITE-ACL Package. Available on <http://liawww.epfl.ch/~calisti/ACL-LITE/>



WLANS (WIRELESS LOCAL AREA NETWORKS): DESCRIPCIÓN Y APLICACIONES

*Miquel Mujal Colilles, Francisco Álvarez Vicente,
José Cástor Vallés Martínez¹*

Estudiantes de la ETS de Ingenieros de Telecomunicaciones de Barcelona.

(1) Miembro de la Rama del IEEE de Barcelona

miquel.mujal@terra.es, falv@unoinfo.com, josecastor@ieee.org



1.- ABSTRACT

Llevamos ya bastante tiempo trabajando con redes de área local (LAN), y hasta ahora se ha conseguido que la velocidad de transmisión de estas redes sea bastante alta comparándola con otras redes de nuestro alrededor. Pero como no, la tecnología sigue avanzando, y ahora lo que se pretende es facilitar la movilidad a los usuarios de estas redes, es decir cada vez la tecnología sin hilos se va apoderando de las comunicaciones, tanto en voz (desde hace ya algunos años) como en datos. En este aspecto las Wireless Local Area Networks (WLAN) ponen de manifiesto que realmente se pueden conseguir tasas de transmisión altas sin cables.

2.-FUNDAMENTOS Y CARACTERÍSTICAS DE LAS LAN

Es la más simple de todas las conexiones que existen, ya que solo enlaza computadoras de un área pequeña como un edificio u oficina, así mismo, una LAN puede estar conectada con otras LAN a cualquier distancia por medio de línea telefónica y ondas de radio.

Pueden ser desde 2 computadoras, hasta cientos de ellas. Todas se conectan entre sí por varios medios y topología, a la computadora(s) que se encarga de llevar el control de la red es llamada «servidor» y a las computadoras que dependen del servidor, se les llama «nodos» o «estaciones de trabajo».

Los nodos de una red pueden ser PC's que cuentan con su propio CPU, disco duro y software y tienen la capacidad de conectarse a la red en un momento dado; o pueden ser PC's sin CPU o disco duro y son llamadas «terminales tontos», las cuales tienen que estar conectadas a la red para su funcionamiento.

Las LANs son capaces de transmitir datos a velocidades muy rápidas, algunas inclusive más rápido que por línea telefónica; pero las distancias son limitadas.

Asimismo, en este tipo de red, tenemos tres sistemas de colocar las computadoras que se van a conectar en red, estos tres sistemas son:

Igual a Igual: Cada estación de trabajo puede compartir alguno, todos o ninguno de sus recursos con las demás estaciones de trabajo.

Recursos Compartidos: Uno o mas servidores centralizados envían y reciben ficheros, y contienen los recursos de las estaciones de trabajo en uso. Las estaciones de trabajo no pueden acceder a los recursos de las restantes estaciones, por lo que deben realizar ellas mismas todos los procesos.

Cliente-Servidor: Reparte una aplicación entre el cliente (estaciones de trabajo) y los componentes del servidor. El cliente de la aplicación acepta las entradas del usuario, las prepara para el servidor y le envía una solicitud. El servidor recibe las solicitudes de los clientes, las procesa y facilita el servicio solicitado al cliente. Entonces, el cliente presenta los datos u otros resultados al usuario por medio de su propia interfaz.

3.- INTRODUCCIÓN A LAS WLAN'S

3.1.- Introducción

En este apartado podemos empezar analizando las tecnologías propuestas para las WLAN's. Básicamente las WLAN trabajan con técnicas de espectro ensanchado, y entre estas técnicas podemos destacar las de FHSS (Frequency Hopping Spread Spectrum) y DSSS (Direct Sequence Spread Spectrum). Esta segunda técnica se utiliza ya en la telefonía móvil de 2^a generación estadounidense, y se piensa que será la base de UMTS (3^a generación de telefonía móvil). Si bien estas técnicas presentan varias ventajas sobre sus competidoras, el mayor problema que tienen es que la capacidad de la «celda» no vendrá condicionada por las interferencias entre otras frecuencias, sino que vendrá limitada por la cantidad de usuarios la potencia que éstos transmitan. Por tanto será necesario un estricto control de la potencia emitida por cada usuario para poder mantener la capacidad del sistema, como se verá más adelante.

Hoy en día la mayoría de las WLAN móviles trabajan en el margen de 1Mbps hasta 2Mbps. Aunque se ha conseguido que algunas WLAN «fijas» (no móviles) alcancen velocidades de hasta 10Mbps. Aunque como se puede intuir los más usuales son los que llegan a 2Mbps,

los cuales se construyen, básicamente, con dos técnicas para la conexión de sus terminales: mediante infrarrojos, o bien, mediante enlaces de radiofrecuencia. Aunque son velocidades más o menos aceptables, se está investigando en estos campos para conseguir mejorar, y por tanto aumentar, estas velocidades.

Para cerrar esta visión global de las WLAN, solo nos queda hablar de los protocolos de acceso al medio. El que se está utilizando en la actualidad es el CSMA/CA, pero ahora solo veremos algunas de las ventajas e inconvenientes del protocolo CSMA en general.

-Ventajas:

- * Es simple y robusto
- * Tiene tiempos de acceso rápido cuando hay poco tráfico
- * No necesita una función de coordinación

-Inconvenientes:

- * La máxima capacidad que soporta está limitada a una fracción de la velocidad de los datos en aplicaciones reales
- * La eficiencia disminuye drásticamente para tasa de tráfico altas.

Algunos de los tipos que se utilizan en las WLAN pueden ser:



Figura 1. Red peer-to-peer

La más básica se da entre dos ordenadores equipados con tarjetas adaptadoras para WLAN, de modo que pueden poner en funcionamiento una red independiente siempre que estén dentro del área que cubre cada uno. Esto es llamado red de igual a igual. (Figura 1)

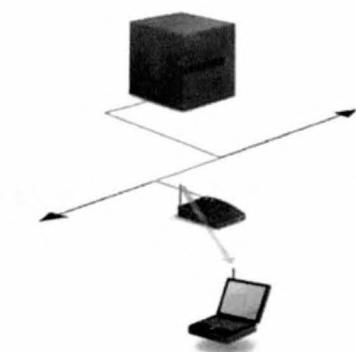


Figura 2. Cliente y punto de acceso

Instalando un Punto de Acceso (APs) se puede doblar el rango al cuál los dispositivos pueden comunicarse, pues actúan como repetidores (Figura 2).

Los puntos de acceso tienen un rango finito, del orden de 150m en lugares cerrados y 300m en zonas abiertas. En zonas grandes como por ejemplo un campus universitario o un edificio es probablemente necesario más de un punto de acceso. La meta es cubrir el área con células que solapen sus áreas de modo que los clientes puedan moverse sin cortes entre un grupo de puntos de acceso. Esto es llamado «roaming» (Figura 3).

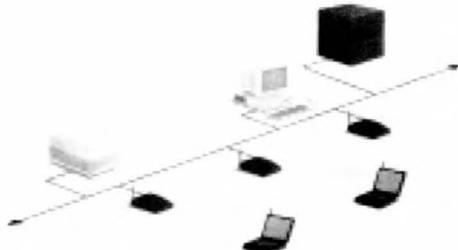


Figura 3. Múltiples puntos de acceso y «roaming».

Para resolver problemas particulares de topología, el diseñador de la red puede elegir usar un Punto de Extensión (EPs) para aumentar el número de puntos de acceso a la red, de modo que funcionan como tales pero no están enganchados a la red cableada como los puntos de acceso (Figura 4).

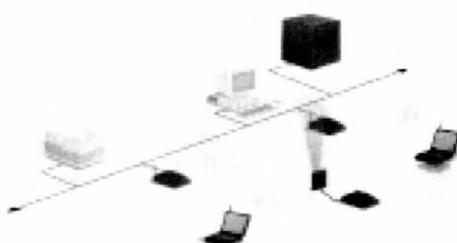


Figura 4. Uso de un punto de extensión.

Uno de los últimos componentes a considerar en el equipo de una WLAN es la antena direccional. Por ejemplo: se quiere una LAN sin cable a otro edificio a 1Km de distancia. Una solución puede ser instalar una antena en cada edificio con línea de visión directa. La antena del primer edificio está conectada a la red cableada mediante un punto de acceso. Igualmente en el segundo edificio se conecta un punto de acceso, lo cuál permite una conexión sin cable en esta aplicación (Figura 5).

Después de esta visión general de lo que son las WLAN, pasamos ahora a estudiar y discutir las dos técnicas de espectro ensanchado comentadas anterior-

mente. Empezamos con FHSS para después continuar con DSSS y poder comparar estas dos técnicas.



Figura 5. Utilización de antenas direccionales.

3.2.-Frequency Hopping Spread Spectrum

En este apartado veremos las propiedades más significativas de la técnica de espectro ensanchado mediante saltos de frecuencia (FHSS), como pueden ser: diseño del radioenlace, especificaciones del emisor y receptor, etc..

Antes de empezar a considerar los aspectos más interesantes de ésta técnica cabe destacar que los enlaces vía radio dentro de un edificio son bastante complicados, y se tienen que considerar bastantes atenuación y rebotes debidos a muebles, paredes, etc.. Es curioso observar que algunas paredes no ofrecen resistencia a ondas con frecuencias menores a 10GHz, mientras que las atenuaciones que producen otras clases de paredes son inviables para sistemas de comunicaciones. Para poder caracterizar estos desvanecimientos se ha propuesto el siguiente modelo de decaimiento de la potencia según la distancia:

- > Hasta los tres primeros metros, la potencia disminuye según la raíz de cuadrada de la distancia.
- > A partir de 3 metros, en un medio sin paredes que se interpongan a la transmisión, esta potencia disminuye con la raíz tercera de la distancia.
- > A partir de 3 metros pero la transmisión la interrumpen distintas paredes, la disminución depende de la raíz cuarta de la distancia.
- > A través de paredes externas esta atenuación aumenta con la raíz sexta de la distancia.

Por tanto, se puede concluir que la zona de cobertura de los puntos de acceso es de unos 50 metros.

Pero estas limitaciones no son siempre malas, es decir, que es posible que tengamos que utilizar diferentes puntos de acceso para cada planta de un edificio, pero esto significa que podríamos hacer un buen reuso de los canales para diferentes plantas del edificio.

Otro problema a considerar es la propagación multicamino y su efecto sobre la velocidad de transmisión de los datos. Como ocurre en los radioenlaces externos, en las comunicaciones sin cables internas, la propagación multicamino supone un pequeño problema para los recep-

tores. Aunque se puede ver que en los entornos externos es bastante más crítica que en los internos. Es decir, esta propagación produce interferencia intersimbólica si el retardo de propagación de las señales rebotadas es mayor que el tiempo de un bit. En el caso de comunicaciones externas es lo que ocurre en la gran mayoría de señales de banda ancha (tiempos de bit muy pequeños y caminos de propagación bastante grandes), mientras que en transmisiones internas estos caminos de propagación suelen ser pequeños, por tanto el retardo que sufren las señales rebotadas suele ser bastante más pequeño que el tiempo de transmisión de un bit, y a consecuencia de esto, en vez de producir esta interferencia, lo que se consigue es que la señal recibida tenga una potencia bastante mayor que si solo hubiera un solo camino de propagación. Por tanto, no se puede considerar como un problema, sino más bien como una ventaja, ya que la sensibilidad de los receptores puede ser peor que en el exterior, y por tanto estos receptores pueden ser más baratos.

En conclusión, para velocidades de 1Mbps hasta 2Mbps esta propagación multicamino supone una mejora para la transmisión, pero a velocidades de 2 a 4Mbps esta propagación multicamino empieza a producir ya ISI.

3.4.-Selección de los conjuntos de salto en frecuencia.

Según el estándar FCC parte 15.247 tenemos una banda de 83MHz a 2.4GHz de portadora. Si el ancho de banda máximo de un señal es de 1MHz, en esta banda podremos colocar 79 slots de frecuencia, teniendo en cuenta los tiempos de guarda entre canales. Y a partir de estos recursos tenemos que planificar los conjuntos de salto en frecuencia. Para el diseño de estos modelos se tendrán que tener en cuenta ciertas limitaciones, por ejemplo:

- * Tener la mínima interferencia posible entre conjuntos usados en:
 - Redes diferentes
 - Celdas diferentes en la misma red
- * Evitar dos modelos en la misma frecuencia y al mismo tiempo
 - * Minimizar dos conjuntos en la misma frecuencia en tiempos consecutivos
 - * Maximizar la separación entre frecuencias en tiempos consecutivos para reducir ráfagas de errores debidas a interferencias.

El sistema FHSS es complicado de entender ya que estamos acostumbrados a considerar como un canal de usuario a un slot de frecuencia o bien de tiempo, y en este caso un canal para un usuario es un conjunto de saltos de frecuencias. Es decir, al igual que en FDMA cada slot de frecuencia se considera un canal, en FHSS un conjunto de frecuencias se consideran como un solo canal, más concretamente, una combinación de este conjunto de frecuen-

cias es lo que forma un canal. Ahora bien, uno podría considerar que cada combinación de frecuencias podría corresponder con un canal, pero tenemos que tener presente siempre las limitaciones expuestas anteriormente. Por ello para saber cuales son los canales de que disponemos tenemos que hacer una buena selección del conjunto de combinaciones (de frecuencias) posibles.

Esto lo podemos ver con un sencillo ejemplo. Si tenemos seis portadoras consecutivas para combinar, el conjunto de permutaciones posibles es $6!$ (teóricamente), esto es 720 posibles combinaciones:

F1 F2 F3 F4 F5 F6 —> Permutación 1 (usuario 1)
F1 F2 F3 F4 F6 F5 —> Permutación 2 (usuario 2)
F1 F2 F3 F5 F4 F6 —> Permutación 3 (usuario 3)
Etc...

Pero como se puede observar, hay conjuntos de estas permutaciones que utilizan una misma frecuencia al mismo tiempo, y esto es intolerable. Si se estudia más a fondo se puede observar que solo 6 de estas 720 combinaciones no utilizan al mismo tiempo, y son:

F1 F2 F3 F4 F5 F6
F2 F3 F4 F5 F6 F1
F3 F4 F5 F6 F1 F2
F4 F5 F6 F1 F2 F3
F5 F6 F1 F2 F3 F4
F6 F1 F2 F3 F4 F5

Estamos de acuerdo que no utilizan la misma frecuencia al mismo tiempo, pero lo que se observa es que utilizan portadoras adyacentes en el mismo intervalo de tiempo y esto supondría tener en los receptores filtros muy selectivos, y aún así no se garantizaría una buena recepción.

Por tanto, para evitar esta situación tenemos que eliminar las filas 2, 4, 6. Y nos quedaría una planificación como:

F1 F2 F3 F4 F5 F6
F3 F4 F5 F6 F1 F2
F5 F6 F1 F2 F3 F4

Así pues de 720 permutaciones posibles nos quedamos sólo con 3. Según esto, con el ancho de banda disponibles que tenemos en FHSS, tenemos 79 portadoras para combinar, y a partir de aquí se pueden obtener un máximo de 20 canales disponibles, que es precisamente la cantidad de canales que propone es estándar IEEE 802.11 aunque la mayoría de los fabricantes ofrecen equipos con sólo 15 canales.

3.5.- Direct Squence Spread Spectrum

Este sistema se ha diseñado para trabajar en la banda ISM de los 900MHz. Pero hay otra variante que permite

trabajar en la banda ISM de los 2.4GHz para aquellos países que no tengan disponible la banda anterior. La máxima potencia permitida para los sistemas DSSS en los Estados Unidos es de 1 Watio, pero la potencia máxima permitida en otros países para la banda de 2.4GHz es de 100 mW. El diseño de estos sistemas utiliza 250 mW en USA, y 100 mW en el resto de países.

La tasa de transmisión de los datos en este sistema es de 2Mbps, aunque según el medio de transmisión, esta tasa puede caer a 1Mbps. La tasa de error que se necesita en estos sistemas es de 10^{-5} para una E_b/N_0 de 16 dB. Con estos requerimientos la disponibilidad del canal es del 99.5%.

Esta tasa de 2Mbps tiene que ser ensanchada por una factor de al menos 11, para satisfacer la parte 15 del FCC. Además se necesitaría una tasa de chip de 22Mcps, que si utilizamos una modulación en cuadratura, podemos trabajar a una tasa de 11Mcps. El ancho de banda requerido para estas señales es de 22MHz. Con este ancho de banda, el número máximo de canales independientes en la banda de 900MHz es uno, mientras que en la banda de los 2.4GHz es de tres, en contraposición con los 15 a 20 canales de FHSS. Por esto los comerciantes se inclinan por el sistema FHSS en vez de DSSS.

Una de las principales características de este sistema es el entrelazado de los datos, para evitar que se produzcan ráfagas de errores. También se necesita una secuencia para aleatorizar, y ensanchar la secuencia de datos, y por tanto que en el receptor se tendrá que conocer esta secuencia para poder demodular estos datos.

3.7.- Interconexión con redes troncales.

En este apartado estudiaremos la movilidad de las WLAN dentro de una misma red, y también la movilidad entre otras redes, incluyendo una pequeña introducción a MobileIP.

3.7.1.- Movilidad dentro de la misma red.

En este apartado estudiaremos la movilidad de los nodos dentro de una misma red que tiene cobertura gracias a varios puntos de acceso. Cada uno de estos puntos de acceso contienen tres componentes:

- * Una interfaz WLAN para comunicarse con los nodos en su área de servicio.
- * Una tarjeta para el interfaz con la red troncal, por ejemplo una Ethernet.
- * Un bridge de la capa MAC para filtrar el tráfico entre las subredes sin cables y entre la troncal.

Este filtrado es esencial para prevenir todo el tráfico en la red troncal para que no sea repetido en la subred inalámbrica. Para ello utilizamos este bridge, que funcio-



na de la misma manera que los que se ponen en redes alámbricas normales.

La función de estos «bridge» es aislar el tráfico entre los interfaces a los que están conectados. Cada vez que pasa un paquete por ellos, éstos van formando una tabla de los nodos que están en la red y el segmento de la red en el que están. Pero estas entradas que están en las tablas no se guardan para siempre, ya que los nodos hay veces que pueden estar apagados, o se pueden mover, y enviar paquetes hacia estos nodos puede ser inútil. Por eso, los «bridge» tienen un reloj que cada cierto tiempo va actualizando las tablas, quitando aquellos nodos que ya no están en esta red. Estos tiempos de actualización en las redes de cable suelen ser de unas horas, en cambio en redes inalámbricas deben ser más pequeños, de orden de unos pocos minutos, ya que los cambios de posición, y las desconexiones de los nodos, son bastante más frecuentes que en redes con cable.

Otras funciones de estos «bridge» son:

- * Almacenar la información de diferentes redes que tiene diferentes velocidades de transmisión.
- * Cambiar el formato de las tramas entre redes incompatibles.
- * Añadir o borrar campos dentro de trama.

Dentro de la familia de estándares 802 podemos destacar los que nos interesan para nuestras aplicaciones. Por ejemplo, el 802.2 es donde el LLC (Logical Link Control) funciona tal como lo hemos explicado. Mientras que el 802.11, es el estándar que se refiere a las WLAN. Dentro del 802 podemos distinguir dos protocolos de la capa MAC:

- * «Spanning tree bridges»
- * «Source routing bridges»

Tanto uno, como el otro tipo de estos «bridges» construyen tablas para almacenar la información de la red, como ya hemos comentado anteriormente. La mayor diferencia entre ellos, desde un punto de vista de implementación es, que el segundo tipo de bridge coloca muchas más carga en los nodos. Por tanto la mayoría de los «bridges» utilizan la primera opción.

El primer tipo de bridge funciona de la siguiente manera: cada nodo crea una base de datos en la que están todos los nodos de la red y el mejor camino para alcanzarlos. La topología de la red no está en ningún nodo, sino que con el conjunto de todos los nodos esta topología va cogiendo forma. La forma de crear esta topología se basa en que cada bridge hace un broadcast de su identidad, y de todos los nodos que él conoce. Y con un algoritmo distribuido se selecciona un bridge raíz y un árbol que alcance a todos los demás nodos, así se minimiza el número de saltos desde cada bridge hasta la raíz. Este árbol se va actualizando continuamente.

El otro tipo de bridge depende bastante más del nodo fuente, el cual guarda la tabla de donde están los otros nodos de la red, y construye una vista general de la topología de la red. También incluye la ruta que el paquete tiene que poner en su cabecera, por tanto, y como estos cálculos son costosos y se tienen que hacer en el nodo, se cargan demasiado los nodos y esto hace que los recursos del nodo sean menores que el anterior tipo de nodo.

3.7.2.- Movilidad entre diferentes redes y MobileIP.

Definiremos los parámetros más interesantes de la movilidad entre redes, e introduciremos algunos conceptos de MobileIP.

Las metas importantes de MobileIP son:

- * La movilidad se tiene que conseguir en la capa de red
- * Las capas más altas no se tienen que enterar de esta movilidad
 - * Las aplicaciones no tienen porque cambiar.
 - * Las infraestructuras de los routers no móviles no se ven afectados
 - * Los terminales no móviles tampoco se tienen que ver afectados.
 - * Varias operaciones ocurren continuamente a través de varias redes
 - * La seguridad es tan buena como en las redes actuales.

La implementación de la movilidad en las redes tiene que ser gradual, y además no se podrá molestar a aquellos nodos que no deseen incorporar movilidad. Ya que queremos minimizar el impacto de la movilidad en Internet y en sus nodos, se tiene que conseguir lo que se ha propuesto arriba, es decir, que esta movilidad sólo afecte al nivel de red, mientras que las capas más altas no se tienen por que enterar de esta movilidad.

El principal objetivo de MobileIP consiste en proporcionar movilidad a un terminal pero siempre conservando su dirección IP. En la actualidad la dirección IP está asociada a un lugar fijo dentro de una red. Por tanto lo que se pretende hacer con MobileIP es proporcionar un número de direcciones IP para que se muevan junto con los terminales móviles. Para ello MobileIP configura un camino para encontrar donde se encuentran estas direcciones MobileIP en cualquier instante de tiempo, y así poder enviarles los paquetes que van dirigidos hacia estas estaciones sin perturbar el comportamiento de Internet.

La operación de MobileIP es bastante simple. Se consideran dos opciones, con naturaleza similar a los dos tipos de «bridges» considerados anteriormente. Una opción es la que hace trabajar a los routers, y la otra propuesta es la que se implementa en los nodos. La que elegimos aquí es la primera de las dos opciones propuestas. Los paquetes

destinados hacia la estación con MobileIP son dirigidos en una primera instancia hacia la red donde debería estar. Si este nodo no está allí, el router encapsula este paquete en otro y lo manda hacia la dirección donde se encuentra en estos momentos.

Así pues, un terminal consigue una dirección MobileIP de un conjunto nuevo de direcciones IP. Esta dirección está asociada con el nodo de la red «original». Esta asignación se comporta de la misma manera que un teléfono móvil con un número asociado a su HLR (Home Location Register). Cuando este usuario sale de su zona «local» tiene que conseguir otra dirección temporal mientras visita esta zona «extranjera». Esta situación también se parece al registro de un usuario en un VLR (Visitor Location Register) cuando se aleja de la estación base a la que «pertenece». Hay que destacar que la dirección IP que tiene en un principio es permanente, mientras que la segunda dirección IP que consigue para poder moverse por la red es temporal, y solo la utiliza mientras está fuera de su red.

El software que permite el desarrollo de MobileIP se denomina agents «agentes», que se dividen, según la función que hacen dentro de MobileIP, en Foreign Agent y Home Agent. El Home Agent es el que redirige los paquetes que le envían a la estación móvil hacia el nuevo lugar de emplazamiento, mientras que el Foreign Agent es el encargado de proporcionar a esta estación la nueva dirección IP temporal para que el Home Agent sea capaz de encontrar a esta estación. También se encarga de buscar a la estación cuando se le envía un paquete. Es decir el Home Agent envía el paquete hacia el Foreign Agent y éste se lo envía a la estación. Típicamente estas funciones las implementan los routers de las redes de área local (o cualquier otro tipo de redes) donde se encontraba la estación, de la red donde estará después de desplazarse.

4.- CONCLUSIONES

La mayoría de las futuras aplicaciones en Wireless LANs serán para ordenadores móviles, aunque, al mismo tiempo, las aplicaciones que mantengan el cable seguirán siendo un importante nicho de mercado. El coste de ambos tipos de adaptadores es similar, pero no todos los fabricantes dan un adaptador que vaya bien para aparatos portátiles, porque sólo quieren estar dentro del nicho de mercado de redes con cables. Otros dan las dos soluciones, un carta de interfaz con la red (Card Interface Network) al igual que soluciones PCMCIA; en el futuro los proveedores fabricarán básicamente soluciones PCMCIA. Por lo tanto, los ordenadores de mesa tendrán que soportar tal interfaz (aunque para los que no lo tengan hay un adaptador entre en interfaz físico NIC y el PCMCIA).

En cuanto a la velocidad de datos que proporcione la red, ésta tendrá basarse en que los usuarios no vean diferencia alguna entre las redes con cables y las redes inalámbricas; no obstante, actualmente esto no está lo suficientemente avanzado y las WLAN, como se ha visto, aún no pueden competir con el ancho de banda de las redes con cables.

Los sistemas de infrarrojos permitirían implementaciones más baratas, pero su cobertura tiende a ser menor, usualmente, requieren visión directa entre el ‘punto de acceso’ y el móvil. Debido a eso, un mismo espacio necesitaría de más ‘puntos de acceso infrarrojos’ que en una red normal, por lo tanto, estos tendrían que ser lo suficientemente baratos y su instalación suficientemente simple para hacer que todo el sistema fuera más eficiente en cuanto a costes que un sistema a banda de radio.

Es importante también saber con qué red de transporte se quiere conexión y con cuáles sistemas operativos de red, para así establecer una red de área local inalámbrica eficiente. Por ahora el tema es bastante claro y la mayoría de implementaciones son Ethernet, aunque haya apuestas para sistemas Token Ring y una gran predisposición para utilizar estructuras ATM, como muestra el caso de la aplicación WAND.

Un tema muy importante en todas las aplicaciones de las WLANs es la seguridad, que ahora ya es un hecho disponible mediante la administración de passwords y claves de acceso, aunque las opciones de seguridad que existen en este tipo de redes no son lo suficientemente satisfactorias como tendrían que ser para tener un producto realmente eficiente.

No obstante, al final el parámetro más determinante en el uso de una tecnología u otra es su precio y sus funcionalidades relativas entre ellas considerando el entorno donde se quiera implementar. Por eso, los precios han hecho que las aplicaciones iniciales de WLAN se pongan en mercados verticales y no en horizontales, que son una gran cantidad y no podrían asumir tal coste. Aún así, el hecho de tener un estándar siempre facilitará que los precios tiendan a bajar.

5.-BIBLIOGRAFÍA

- [1] www.wlana.com
- [2] “CAPACITY OF A WIRELESS LAN WITH VOICE AND DATA SERVICES”, Ali Zahedi and Kevin Pahlavan, Fellow, IEEE, *IEEE TRANSACTIONS ON COMMUNICATIONS*, VOL. 48, NO. 7, JULY 2000.
- [3] “A UNIFIED WIRELESS LAN ARCHITECTURE FOR REAL-TIME AND NON-REAL-TIME COMMUNICATION SERVICES”, Sunghyun Choi, Member, IEEE and Kang G. Shin, Fellow, IEEE, *IEEE/ACM TRANSACTIONS ON NETWORKING*, VOL. 8, NO. 1, FEBRUARY 2000.
- [4] “ISSUES IN HIGH-SPEED WLAN”, David J. Skellern, Alex C. K. Lam, Neil Weste. *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, Volume: 6, 2000 Page(s): 3698 -3701.
- [5] “A HIGH-SPEED WIRELESS LAN” Skellern, D.J.; Lee, L.H.C.; McDermott T.; Weste, N.H.E.; Dalton, J.; Graham, J.; Wong, T.; Myles, A.F.; Percival, T.M.P. Ryan, P.J. *IEEE Micro*, Volume: 17 Issue: 1, Jan.-Feb. 1997 Page(s): 40-47.
- [6] “NEW HIGH-RATE WIRELESS LAN STANDARDS” van Nee, R.; Awater, G.; Morikura, M.; Takanashi, H.; Webster, M.; Halford, K. *W IEEE Communications Magazine*, Volume: 37 Issue: 12, Dec. 1999 Page(s): 82-88
- [7] “IEEE 802.11 WIRELESS LAN: CAN WE USE IT FOR MULTIMEDIA?” Andren, C. *IEEE Multimedia*, Volume: 5 Issue: 2, April-June 1998 Page(s): 84-89
- [8] www.meetel.com





¿ES QUE TIENE LA CIENCIA ALGÚN SENTIDO?

Fortanet, J.

Estudiante de Filosofía de la Universidad de Barcelona

ximofortanet@ole.com

“Todo nuestro razonamiento consiste en ceder al sentimiento”

PASCAL

“La electrodinámica cuántica describe a la Naturaleza como entidad absurda para el sentido común, y se halla plenamente de acuerdo con la experimentación. Espero, pues, que puedan aceptar a la Naturaleza como es: absurda”

R. FEYNMAN

Este artículo tiene la intención de contribuir de alguna manera a replantear algunos problemas que, desde los años sesenta, han ido anidando en diversas disciplinas como la física, la sociología, la psicología y la filosofía. Dichos problemas se refieren fundamentalmente al relativismo epistemológico, es decir, al convencimiento de vivir sin verdad.

Desde todas estas disciplinas se ha llegado en las últimas décadas a situaciones que dinamitan el concepto clásico de verdad. De este modo, la corriente del postmodernismo en los años setenta comenzó a gritar a favor de la ausencia de verdad. Quizá de una manera irresponsable, aunque fundamentada. El que las ciencias humanas carecieran de verdad las acercaba al arte (que no tiene verdad y ha sabido vivir sin ella). Así, hubo todo un movimiento (psicoanálisis, filosofía, física) que pregona ba el carácter creador y caótico del hombre en todas sus disciplinas, como en un neo-romanticismo.

El blanco de este movimiento conocido con el nombre de postmodernismo fue aquella disciplina en dónde era más complicado desprenderse del concepto de verdad. Sin duda, esta disciplina era la física. En los años setenta se dio un increíble esfuerzo intelectual por socavar los fundamentos de la física y en definitiva, mostrar que la verdad que esta disciplina creía poseer no era tal verdad. Todo este movimiento que los físicos observaban atónitos dio pie al llamado relativismo científico. Este artículo va encaminado a esbozar los caracteres y principales argumentos del relativismo científico, contraponerlos al científico clásico y así poder valorar estas dos tendon-

cias que han dominado muchos de los debates de los últimos tiempos.

Para este objetivo creo que, puesto que es sobradamente conocido el punto de vista científico, se deben exponer primero las críticas más acertadas que lleva a cabo el relativismo científico–o epistemológico (episteme significa conocimiento científico en griego). Con todo esto queda claro que el presente artículo es simplemente una exposición crítica de ambas posiciones, sin optar religiosamente por ninguna de ellas.

1-CIENCIA Y VERDAD

Cualquier persona sabe lo que es la verdad. No creo que sean necesarios elevadas argumentaciones filosóficas en torno al concepto de verdad. Tomemos, así, la verdad como aquello que es cierto, que tiene una evidencia tal que no podemos dudar de ella. Así es como tomó Descartes el concepto de verdad. Pero, una vez dicho esto, nos encontramos que hay muy pocas cosas de las que tenemos evidencia absoluta. Es decir, podemos dudar de muchas cosas. De esta manera, se hace necesaria la prueba. Y aquí es en dónde la física se hace fuerte. No considero admisible la duda de los sentidos. Si lo que vemos no es verdad, es al menos la única verdad a la que tenemos acceso, y por tanto la única verdad que hay que tener en cuenta. La verdad es, de este modo, algo que tiene que ver con los hechos que percibimos y que requiere una prueba en estos mismos hechos. Lo que la física llama una prueba empírica.[1]

La física trabaja con hechos empíricos y con pruebas empíricas que validan una determinada teoría que tiene como características la regularidad y la previsión. De este modo, resulta extraño de qué manera se puede venir abajo esta noción de verdad. Resulta muy extraño, y más aún cuando la física realiza previsiones que se cumplen, como la teoría de la gravedad o el movimiento uniformemente acelerado.

Pero estos temas son de la mecánica newtoniana, del paradigma newtoniano. Efectivamente, la teoría de la gravedad prevé que un lápiz caiga al suelo, y gracias al movimiento uniformemente acelerado podemos saber que llegaremos en veinte minutos al bar de turno. Pero las

cosas se van complicando a medida que nos enfrentamos a problemas más complicados, tales como la mecánica cuántica o la meteorología, en dónde las previsiones –lo que consolida la redondez de la verdad – ya tienen un carácter meramente estadístico (teoría del caos).

Y es que si hablamos de una verdad estadística todo se complica extraordiariamente. Si me cae una maceta en la cabeza mientras paseo por la calle, se me podrá decir que la verdad de este hecho es una ley probabilística que viene caracterizada por el peso de la maceta, el estado del balcón en dónde se apoya la maceta, la probabilidad que tenía yo de pasar por esa calle de camino al bar, etc... Pero para una persona de una cultura distinta esto no tendría nada que ver con la verdad. Se preguntaría el porqué me cae una maceta a mí, quizás si había cometido algún pecado, quizás si me habían maldecido de alguna manera....

Con esto quiero decir que la verdad y la ciencia tenían en la mecánica newtoniana una relación mucho más estrecha que hoy en día. Quizás porque toda la mecánica newtoniana estaba salvaguardada por la idea de Dios (Newton admitía que no sabía lo que era la gravedad, que era algo producido por Dios, efecto del éter).[2]

Aún así, la fragilidad de la verdad estadística no justifica los ataques a la ciencia. Se precisan de mayores argumentos, más profundos y sutiles. Así, el esfuerzo posmoderno por socavar esta vieja idea de verdad no es trivial y tiene a sus espaldas un riguroso anclaje político, ético e intelectual. Podríamos exponer la crisis de esta vieja idea de verdad en cuatro puntos principales: la inducción, la subdeterminación, la incommesurabilidad de los paradigmas y la sociología científica.

A partir de estos temas se eleva toda una crítica a la verdad científica que ha sido convenientemente tenida en cuenta, explícitamente o no, por numerosos físicos como Feynman, Bohr, Schrödinger, Sokal, Bricmont, etc...

a) El primero de estos ataques es el relativo a la inducción. Aunque fue llevado a cabo por el filósofo D.Hume, ha sido retomado por epistemólogos como Popper (amigo personal de Einstein). La inducción es aquella manera de razonar por la cual, a partir de un hecho, se infiere el resto. Es decir, si vemos que un lápiz cae, podemos decir que todos los lápices caen. Esta manera de conocer no es aceptada en la lógica deductiva, y apoyándose en eso y en Hume, Popper dirá que nunca podemos verificar una teoría científica por medio de la inducción, que nunca podemos probar todos y cada uno de los casos.

De esta manera, la teoría de la gravedad nunca quedará rigurosamente demostrada, pues no se deduce, sino que se infiere: “*¿Está racionalmente justificado razonar partiendo de casos con los que ya hemos experimentado para llegar a otros casos en los que carecemos de la menor experiencia? La reiterada respuesta de*

Hume es: no, no está justificado (...) y en mi opinión, la respuesta de Hume a este problema es correcta” (K.Popper, ‘La lógica de la investigación científica’).

Popper salvará el problema diciendo que , puesto que no se pueden verificar las teorías científicas que se apoyan en la inducción, tendrán que falsarse. Es decir, no podemos decir que la teoría de la gravedad es verdad porque inducimos el hecho de que el lápiz va a caer; pero sí que podemos decir que la teoría de la gravedad es verdad si no encontramos un modo de probar que es falsa. De falsarla. Esto trae un número increíble de problemas, tanto al tema de la verdad de la ciencia , como al modo de trabajo, pues por un lado puede ocurrir que una teoría sea parcialmente cierta aunque falle una predicción (pueden haber numerosos factores que la hagan fallar), y por otro lado ,existen teorías científicas tan complicadas que el hecho de su falsación es prácticamente imposible.

b) El segundo de los ataques a la verdad científica es la subdeterminación. De acuerdo con la tesis de Duhem-Quine (dos filósofos de la ciencia y del lenguaje) las teorías están subdeterminadas por hechos. Hay un número de hechos finito, en cambio hay un número infinito de teorías. Por lo tanto , ocurre que hay infinitas teorías válidas para un hecho concreto, con lo cual la elección de tal o cual teoría no corresponde tanto a la verdad científica como a la creencia. Esto es absurdo con ciertas teorías científicas contundentes, como el movimiento uniformemente acelerado o la circulación de la sangre, pero en el caso del origen de las especies o ‘la unificación de la relatividad general con la física de las partículas elementales’ la cuestión presenta serias dudas acerca del estatuto de la verdad científica, pues hay varias teorías simultáneas que explican dichos hechos.[3]

A parte del tema de la inducción y la subdeterminación, se han vertido ataques mucho más contundentes sobre la verdad científica. Estos ataques mucho más devastadores tienen consecuencias no solamente científicas, sino que van encaminados hacia consecuencias políticas y sociológicas. Por ello, por su radicalidad, las siguientes posturas son ejemplos por antonomasia de relativismo epistémico, que si bien puede parecer en un principio que dotan al hombre de una capacidad creadora del mundo que le rodea, pueden caer –en determinados desarrollos- en un relativismo absoluto en el que se deje toda verdad en manos de las instancias del poder, es decir, en un relativismo que anule toda posición crítica.

2.-POSICIONES ANTI-CIENTÍFICAS

Hemos visto, al tratar el tema de la verdad científica, que la ciencia caracterizaba la verdad como aquello que se prueba empíricamente (hecho) y que, por tanto, es regular (da lugar a leyes). El resto es , para la ciencia, creencia. Y las creencias no entran dentro del campo científico, pues corresponden a mitos, ficciones, narraciones, etc... Al mismo tiempo, también hemos visto que esta verdad



consistente se iba licuando, evaporando, hasta ser una verdad mucho más escurridiza de lo esperado. Aún así, todavía se mantenía la separación entre verdad científica y creencia. La tarea de las últimas posiciones anti-científicas se sitúa en esta distinción verdad-creencia. Y la subvierte. Siguiendo el modelo dadaísta, surrealista y situacionista, estas posiciones se encarnizan en darle la vuelta a esta dicotomía tan asumida por los científicos para demostrar, en última instancia, que ,aún partiendo de esa misma distinción, la verdad científica no se corresponde a lo que se llamaba tradicionalmente verdad, sino que se aproxima a lo denominado creencia. En términos posmodernos: que la ciencia es una ficción[4], una narración que carece de objetividad indiscutible y que es , además, una producción cultural.

Con el objetivo de observar someramente el desplazamiento que sufre la verdad científica hasta devenir creencia, tomaremos los razonamientos que llevan a cabo algunas de las posiciones anti-científicas que mayor repercusión han tendido.

La primera de estas posiciones beligerantes viene representada por la figura y obra de Kuhn. A él pertenece la teoría citadísima de la incomensurabilidad de paradigmas. El razonamiento de Kuhn[5] , sintetizado, vendría a ser el siguiente: el grueso de la actividad científica, se desarrolla en el interior de paradigmas, que definen el tipo de problemas que hay que estudiar, los criterios con los que evaluar una solución y los experimentos que se consideran aceptables. Pero estos paradigmas científicos, de vez en cuando entran en crisis, y entonces se asiste a un cambio de paradigma científico. Por ejemplo , el paradigma newtoniano supuso la desaparición del aristotélico, y la teoría cuántica ha destrozado el paradigma de Newton.

Pero esto tan solo es la premisa (A) de la que parte Kuhn. Premisa, por otra parte, que ha sido casi totalmente aceptada por los físicos y teóricos de la ciencia[6]. Kuhn, desde esta premisa, llega a la conclusión que es el centro de su teoría : la incomensurabilidad de los paradigmas (conclusión B).

En efecto, cada paradigma se sustituye por otro cuando entra en crisis. Pero es un craso error creer que la sustitución de paradigmas (Ej. de Aristóteles a Newton) se debe a que surge una nueva teoría que es en algún sentido superior. Recordemos que, previa a la aparición del nuevo paradigma, el antiguo paradigma entra en crisis. Y esta crisis del paradigma antiguo no tiene nada que ver con motivos científicos, sino que corresponde a lo que Foucault llama '*cambio en las condiciones de posibilidad de las estructuras*'[7]. O dicho de manera un tanto más gruesa, que los cambios de paradigma corresponden a cambios profundos en las estructuras sociales, míticas, psicológicas, económicas...cuya causa viene dada directamente por un cambio en las relaciones de poder[8].

Es cierto que no se puede explicar el cambio de paradigma aristotélico al paradigma newtoniano sin acudir a las convulsiones sociales, culturales y económicas que produjo la gestación del movimiento racionalista. De esta manera, y siguiendo el razonamiento de Kuhn, no se puede decir que un paradigma fuese mejor que otro, puesto que cada uno, mientras no permaneciese en crisis, resolvía las inquietudes científicas y los problemas derivados de las estructuras de cada época. Con todo esto, Kuhn concluye (B), que los paradigmas no se pueden comparar entre sí, son incomensurables:

'A la vista de las pruebas, los químicos no podían aceptar sin más la teoría de Dalton, ya que una buena parte de ellas seguían siendo negativas. Sin embargo, una vez aceptada, tuvieron que forzar a la naturaleza para que se adaptara a ella, un proceso que, en este caso, aún se prolongó casi durante toda una generación, después de lo cual, cambió incluso el porcentaje de los compuestos más conocidos. Los propios datos habían cambiado. Éste es el último de los sentidos en que podemos afirmar que , después de una revolución, los científicos trabajan en un mundo diferente [en un paradigma incommensurable al anterior]' (Kuhn, 'la estructura de las revoluciones científicas'ed. FCE).

Este es el primer paso que aproxima la verdad científica a creencia, pues la hace depender de estructuras no científicas, que bien pueden ser míticas , como en el caso de Newton (cambio de paradigma relacionado, entre otras cosas con un cambio profundo en la teología, en la que se debía albergar , por primera vez en la historia, una subjetividad moderna, es decir, un sujeto tal y como lo entendemos hoy en día).

A partir de Kuhn y apoyándose en él han ido surgiendo nuevos ataques contra esta idea de verdad científica. Tal es el caso de Feyerabend, quien amparándose en esta idea de la incomensurabilidad de los paradigmas, y en concreto en el hecho de que cada paradigma surgía de una construcción cultural o mítica (ej: copérnico y su adoración religiosa al sol), llegaba a la conclusión que toda verdad científica es una narración, surgida de mitos y articulada razonablemente, pero que en última instancia no existen reglas ni leyes objetivas ni reglas universales : '*la idea de que la ciencia puede y debe organizarse a tenor de unas leyes fijas y universales es, a la vez, utópica y perniciosa*'. Y , por tanto, la ciencia es un sistema de articulaciones razonables en el que 'todo vale', puesto que no existe una teoría definitiva, y las incompletas se apoyan sobre construcciones culturales arbitrarias.

Pero, después de Kuhn, el ataque más encarnizado viene dado por el programa de sociología de la ciencia, a principios de los setenta, llevado a cabo por D.Bloor y B.Barnes. Es este el último paso en la dicotomía verdad-creencia, y con él se cumple el desplazamiento que lleva la verdad científica al estatuto de creencia.

Amparándose en un análisis de la sociología de la ciencia, y partiendo de un relativismo epistemológico consistente en la afirmación de que no hay verdades sino creencias, relegan a la ciencia a un singular estatuto. La ciencia vendría dada por las preferencias de unas creencias sobre otras, y lejos de ostentar la categoría de verdad, la ciencia procuraría unas pruebas para las creencias razonables preferidas.

Pero es que aún así, el lápiz cae, tardo veinte minutos exactos en llegar al bar, la tierra es –casi– redonda, el vaso de whisky es tubular, y todo esto tiene una regularidad, cada día es así...

Aquí se toma toda la tradición pragmatista (Dewey, Pierce), de manera que una creencia es preferida por el hecho de que es mejor para nuestros intereses. Hay que decir que esta corriente al contrario de las anteriores, (que intentaban dar cuenta de los cambios de paradigmas), se preocupa de porqué es aceptado tal o cual método y en consiguiente , tal o cual regla.

Un ejemplo famoso es el tema del Challenger, en el que la NASA invertía en lanzaderas reutilizables, reparándolas , a pesar de que era menor el coste de fabricar una nueva. De esta manera, se ve que la ciencia está subordinada a un sistema de preferencias que le marca las reglas, a unas creencias que interesan.

'Según nuestro postulado de equivalencia, todas las creencias están en un plano de igualdad en lo que se refiere a las causas de su credibilidad. Eso no significa que todas las creencias sean igualmente verdaderas o igualmente falsas, sino que, independientemente de su veracidad o de su falsedad, se debe considerar su credibilidad como igualmente problemática. La postura que vamos a defender es que, sin excepción, la incidencia de todas las creencias reclama una investigación empírica y que hay que dar cuenta de ellas hallando las causas específicas, locales, de dicha credibilidad. Esto significa que, independientemente de si el sociólogo evalúa una creencia como verdadera o racional, o como falsa e irracional, debe indagar más en las causas de su credibilidad' (Barnes y Bloor).

Barnes y Bloor toman , de esta forma, un programa de trabajo en el que , partiendo de que todo es creencia, se debe investigar porqué tal o cual creencia es aceptada como verdad. Y la razón de esta aceptación vendrá dada por factores de preferencia que no tendrán nada que ver con aquello que hemos llamado al principio de este artículo algo así como verdad.

El desplazamiento ya se ha efectuado. La verdad científica es una creencia aceptada por oscuras razones,

razones que van cambiando y dan lugar a nuevas creencias que modifican el concepto de verdad. Así, las creencias científicas son mecanismos o dispositivos articulados de manera complicadísima que , a fin de cuentas, resolverían problemas que interesan a esas oscuras razones que mueven los sistemas de creencias.

Pero, ¿es la ley de la gravedad una creencia aceptada de esta manera tan sospechosa?. En complicadas teorías podríamos ver, gracias a todos los ataques que se han formulado, que la verdad como tal no es tan clara y que quizás nos hallaríamos en un lugar en el que la ciencia moderna utiliza ficciones con el fin de ir esclareciendo las teorías más oscuras. Pero en teorías tan asumidas como la ley de la gravedad, se nos hace difícil aceptar todo este relativismo, aún cuando sabemos que nuestras verdades más íntimas dependen en última instancia de deseos, ya sean inconscientes, ya sean impuestos por el poder; que nuestra estructura humana viene atravesada por un torrente de pulsiones, nuestras y que se nos imponen , que somos eso y poco más, que nuestras creencias responden a lo que deseamos o a lo que quieren que deseemos, a aquello que muchas veces se nos escapa. Pero es que aún así, el lápiz cae, tardo veinte minutos exactos en llegar al bar, la tierra es –casi– redonda, el vaso de whisky es tubular, y todo esto tiene una regularidad, cada día es así...En un palabra, que es dudoso que se puedan tomar las verdades científicas por meras creencias, aunque sí que se puedan subvertir ambas para denunciar cierto hegemonismo científico y hacerle ver que las reglas no son eternas , que cambian, que dependen de numerosísimos factores extra-empíricos y extra-científicos.

3.-A MODO DE CONCLUSION

Tras esta exposición, es innegable que la verdad científica se ha cuestionado seriamente. Pero al mismo tiempo, también es innegable el hecho de que existen determinadas verdades empíricas a las que no se puede renunciar, que ya forman parte del sentido común, de nuestra visión del mundo.

La verdad científica no es una mera creencia. Pero hay que distinguir entre verdad y sentido. Toda verdad encierra, ella misma, un sentido. Y es el sentido lo que mueve la verdad, lo que la hace cambiar, como bien vería Nietzsche[9]. El sentido es parte inseparable de cierta verdad, y todas las posiciones anti-científicas no separan la verdad de su sentido.

Dicir, como hace la sociología crítica, que solamente hay creencias, que nuestro conocimiento del mundo viene dado por creencias que se sitúan al mismo nivel –que valen lo mismo- puede dar lugar a desarrollos peligrosos. Se anula toda posición crítica, pues si todo son creencias, desde una posición relativista no se puede deslegitimar una forma de poder, no se puede acusar al ejército mexicano de torturar y vilipendiar los derechos del pueblo indígena de Chiapas.

No podemos aducir porqué hay que defender libertades básicas del hombre, porqué la ciencia debe dar



ET FALTA ALGUNA *BURAN*?

BURAN 1

EVOLUCIÓN DE LA TV (I), Pedro Vicente
LASER, GUIES ÓPTIQUES I SOLITONS, Lluís Torner
FRANCESC SALVÀ I CAMPILLO, A. Elias et al.
PROCESADO DE IMAGEN, F. Marqués

BURAN 3

UNAS PALABRAS SOBRE REALIDAD VIRTUAL, Paul Derek
EL RADAR: DE DÓNDE VIENE Y HACIA DÓNDE VA, A. Broquetas
COM CONSTRUIR-NOS UN ROBOT, Antoni Ferraté

BURAN 5

LA NOSTRA LÒGICA ÉS DIFUSA, Ramir De Porrata-Dòria et al.
SISTEMAS LIDAR PARA CONTROL DE TRÁFICO, A. Comerón et al.
INTRODUCCIÓN A LOS FRACTALES, Carlos Alonso
LA TEORIA DE LA RELATIVITAT, Jordi Cosp



BURAN 7

WAVELETS, I. Bañó
ROBOTS AUTÓNOMOS, J.M. Molina et al.
SISTEMAS DE TELEFONÍA POR SATÉLITE, F. Estebaranz

BURAN 9

SISTEMAS MÓVILES DE 3^a GENERACIÓN: CDMA, Pilar Díaz
XARxes NEURONALS, Enric Claverol
CODIFICACIÓN DE CANAL, E. Monroy

BURAN 11

RECONOCIMIENTO DE CARAS, L. Lorente
MIR: LA GRAN DESCONOCIDA, D. González
SOUND BLASTER COMO INSTRUMENTO VIRTUAL, M. Vargas

BURAN 13

ESPECIAL EXPO-ESA

ANTENAS FRACTALES, Carlos Prieto et al.
MOBILE IP, Miquel Oliver et al.
COMPATIBILIDAD ELECTROMAGNÉTICA, Ramón Pallás et al.
SENSORES INTELIGENTES, Angel A. Custodio

Vols tenir la col.lecció completa de *BURAN*?

Si et fa falta algun exemplar i no saps on trobar-lo,
passa't pel despaxt de la Branca

Mòdul B5-S104 (Casal)
Campus Nord de la UPC

o envia'ns un e-mail a:

buran@ieee.upc.es
ieee@ieee.upc.es

BURAN 2

LA BIOINGENIERÍA EN LA ESCUELA, Ramon Pallàs
EVOLUCIÓN DE LA TELEVISIÓN II, Pedro Vicente
MÓDEMOS: QUÈ HEM DE CONÉIXER SI N'HEM DE TRIAR UN?, J. Serrat

BURAN 4

LA TELEFONÍA MÓVIL EN ESPAÑA, J. Ignacio Mena
LA CODIFICACIÓN DE LA INFORMACIÓN
EN EL CD-DA, Jaume Llardén
INTRODUCCIÓ A LA RADIOASTRONOMIA, Xavier Condeminas
TELEVISIÓN AMATEUR. PREMISAS BÁSICAS, Magí Casamitjana

BURAN 6

¿QUÉ ES LA CRIPTOGRAFÍA?, Ana María Peñas
INTRODUCCIÓN A LOS SERVICIOS DE INTERNET, A. Antón
PRINCIPIOS DEL GPS, Luis Isidro Ramírez García
LES AUTOPISTES DE LA
INFORMACIÓ, Jordi Domingo

BURAN 8

SISTEMAS DE COMUNICACIONES MÓVILES,
Ramón Agustí
MPEG-4, J.R. Casas, F. Marquès y P. Salembier
METABALLS PARA MODELADO 3D, Guillermo Diez
LOS "TURBO CÓDIGOS", J.A. Delgado-Penín et al.

BURAN 10

LA INGENIERÍA ELECTRÓNICA Y LA MEDICINA, R. Pallàs
COMPUTACIÓN CUÁNTICA, J. Tejada et al.
MÓDEMOS A 56 Kbps, Xavier Hesselbach

BURAN 2

PEDRO DUQUE, EL PRIMER ESPAÑOL
EN EL ESPACIO, Daniel Gonzalez
APLICACIÓN DE WAVELETS A LA
COMPRESIÓN DE IMÁGENES, Carlos López
LANZADERA BURAN, UNA GRAN DESCONOCIDA, Daniel Prado

BURAN 14

MP3, SONIDO DIGITAL AL ALCANCE DE TODOS David Rincón
ADSL, Xavier Lago
SENSORES INTELIGENTES, Angel Custodio et al.
TEOREMA DE GÖDEL, Carles Ruiz

BURAN A INTERNET

La difusió de matèries relacionades amb les telecomunicacions, la informàtica i l'enginyeria han estat el primer objectiu de BURAN des del seu naixement, el març de 1993. La Branca d'Estudiants de l'IEEE de Barcelona renova i actualitza el format electrònic de la revista. Podeu trobar-la a la URL:

<http://citel.upc.es/users/buran>
<http://ieee.upc.es>