UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

DISEÑO, METODOLOGÍA DE FABRICACIÓN Y USO DE UNA MEMORIA NAND FLASH COMO TARJETA SIM DE ALTA DENSIDAD

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

INGENIERA EN TELECOMUNICACIONES
E
INGENIERO EN COMPUTACIÓN

P R E S E N T A N

ANIA MADRIGAL REYES
ARTURO MADRIGAL REYES

DIRECTOR DE TESIS: M.I. JUAN CARLOS ROA BEIZA

MÉXICO, D.F. SEPTIEMBRE DEL 2009

Agradecimientos

Gracias a la Universidad Nacional Autónoma de México por darme la posibilidad de ser una profesionista de calidad, llena de ambiciones y colmada de conocimientos. Por siempre PUMA.

Gracias al M.I. Juan Carlos Roa Beiza, nuestro director de Tesis, quien nos apoyó en este último peldaño que nos faltaba, para poder cumplir un sueño que desde algunos años atrás parecía cada vez más lejano.

Gracias al Programa de Apoyo a la Titulación, bajo la coordinación del Ing. Carlos Sánchez, por ser una opción muy valiosa para todos los egresados de las ingenierías.

Gracias a mis maestros de toda la vida, porque cada uno participó en mi desarrollo profesional; sin su apoyo y conocimientos no estaría aquí, terminando un capítulo más de mi vida.

“La mejor manera para predicar tu futuro es crearlo.”

Ania Madrigal Reyes

Agradecimientos

El trabajo de esta tesis no habría sido posible sin la colaboración y el apoyo de muchas personas a las que quisiera expresar mi más profundo agradecimiento.

Quiero dar las gracias al M.I. Juan Carlos Roa Beiza, nuestro director de tesis, por su dedicación y apoyo en la realización de este trabajo.

Debo un reconocimiento especial a Tim Simmons por ser una fuente constante de ayuda, explicaciones y consejos durante la elaboración de este documento.

Quiero expresar mi gratitud a Sandy Supnet por su apoyo y constante ánimo para terminar esta tesis.

Por último, quiero agradecer a mi familia, a mi madre por su apoyo y en particular a mi hermana por haberse decidido hacer esta tesis conmigo. Gracias también a mi querida esposa Suzan por su cariño, incondicional apoyo y paciencia infinita durante el desarrollo de este trabajo.

Arturo Madrigal Reyes

Dedicatorias

Con muchísimo cariño para mi Tata, donde quiera que esté…

A mi mamita, a quien nunca le podré agradecer todos sus consejos, sus desvelos, su apoyo, sus lágrimas y sobre todo.. su amor y amistad…

A Tita, quien siempre ha estado pendiente de nosotros y pide porque nunca nos falte nada en la vida..

A mi hermanita Mariana y mis pequeñines Desi, Allen y Alex, quienes saben que los quiero mucho y espero que lleguen a ser personas de éxito. Acuérdense, siempre sigan adelante..

A mi querido hermanito Puchi, a quien quiero mucho y agradezco todo su apoyo para salir adelante con esta Tesis y con muchas cosas más.. Y por supuesto a Suxy, a quien considero como mi otra hermana..

A Lauris por estar siempre acompañándonos en nuestras vidas..

A mis amigos, que siempre me han ofrecido su apoyo y que me han hecho cada día mejor persona..

Ania Madrigal Reyes

Para Tata

Arturo Madrigal Reyes

Índice General

Índice de Figuras

Índice de Tablas

PREFACIO

Objetivo

El mercado de Memoria Flash, ha crecido a un ritmo extremadamente rápido, especialmente cuando es comparado con otro tipo de semiconductores. Mucho de este crecimiento es causado por el incremento del uso de memorias Flash en reproductores multimedia portátiles, pero también es debido al crecimiento del mercado de teléfonos celulares, ya que solamente este año se esperan vender más de 1.2 billones de celulares por todo el mundo. De acuerdo con el pronóstico de la compañía World Semiconductor Trade Statistics Inc., se espera que los ingresos en el mercado mundial de semiconductores lleguen a los $274.2 billones de dólares este año, siendo $17.6 billones solamente lo correspondiente a Memorias NAND Flash. Esperando así, que para el año 2009 los ingresos tengan un incremento total cercano a los $26.2 billones, aumentándose así en una tasa de crecimiento anual del 32 por ciento desde el 2004.

Debido a la relevancia que conllevan estas cifras, se pretende con este proyecto de tesis, proporcionar precisamente al lector, una explicación de todos los pasos y procesos necesarios en forma teórica y práctica, para la fabricación de una Memoria NAND Flash y una de las posibles aplicaciones dentro del área del conocimiento de la ingeniería, a fin de presentar una idea clara de lo que involucra este crecimiento tan significativo dentro de la industria tecnológica a nivel mundial.

Presentando así, desde cómo se plantea el diseño de la memoria y la concepción de esta, la forma en cómo se preparan los diagramas de componentes o (layout), la jerarquía dentro del diseño, la generación de las simulaciones; cómo se hace el proceso de foto litografía y la alineación de la máscara para la impresión, así como el proceso de grabación de la oblea de silicio, junto con la dilucidación del proceso de pruebas y ensamble, para llegar finalmente al producto que sale al mercado.

Y es precisamente este producto final que hace dirigir nuestra mirada al otro mercado emergente muy importante, el de los teléfonos celulares, lo cuales, como se mencionaba, se están volviendo uno de los más grandes consumidores de Memorias NAND Flash, por lo cual, dentro de este mismo proyecto, se presentará la Tarjeta SIM de alta densidad como una de tantas aplicaciones que se tiene de las Memorias NAND Flash en los teléfonos celulares. Mostrando de forma general, los servicios y contenidos multimedia que actualmente se encuentran comercializándose dentro de los mercados de telefonía de tercera generación, tales como TV móvil, aplicaciones JAVA interactivas de gran capacidad, etc.

Es así, que este proyecto pretende ser una investigación practica de campo en el área de semiconductores y como es el estado del arte para este campo de la ingeniería.

Definición de Problema

La industria de la fabricación de semiconductores a nivel mundial requiere siempre de una innovación tecnológica tanto de sus productos como de sus procesos de diseño y fabricación, lo cual lleva a la constante necesidad del análisis de estos, para lograr así estar a la vanguardia y ser competitivos dentro del mercado internacional.

Para lograr ser uno de los principales productores de semiconductores a nivel mundial es necesario contar con un centro de diseño e investigación, suficiente capital e infraestructura, para poder mantenerse en el mercado cambiante del mundo de los semiconductores.

Debido a que los centros de diseño y manufactura se encuentran fuera del país, no es posible tener un acercamiento a este tipo de tecnología. Presentándose así de esta manera una oportunidad para que el usuario o lector, pueda tener conocimiento de cómo se diseñan y fabrican las memorias NAND Flash.

Por otra parte, el creciente desarrollo de las tecnologías inalámbricas ha promovido un incremento exponencial dentro de los avances de nuevos dispositivos, servicios y aplicaciones para ser utilizados dentro de estas.

Tal es el caso de la telefonía celular, tecnología que fue concebida en un principio estrictamente para la voz y que ha evolucionado tanto en el desarrollo de nuevos servicios: mensajería, transmisión de datos, servicios de información, etc.; como en la manufactura de los dispositivos que permiten el acceso a la red, cuya principal peculiaridad ha sido el disminuir cada vez más, el tamaño y peso, de sus componentes.

Siendo así que la convergencia evolutiva de estos componentes lógicos y físicos ha generado nuevas interfaces óptimas para la distribución de los servicios, la tarjeta NAND Flash de alta densidad, se puede considerar como una clara demostración de esta unión y cuya presencia ya es una realidad en el mercado de la telefonía celular actual.

Metodología

Durante la elaboración de esta tesis expositiva, se presentará en forma documentada una serie de métodos utilizados para la fabricación de una memoria NAND Flash y su aplicación como producto comercial dentro del mercado de las Telecomunicaciones.

Esta investigación contempla la definición de los pasos iniciales dentro del diseño de los componentes, a través de la comprensión de la teoría de los semiconductores y sus características teóricas, asociado al diseño físico a través de medios computacionales para la preparación de los dispositivos. Así como el detallado proceso de fabricación, en el cual se expondrá el método utilizado dentro de una entidad real, lo cual es una

experiencia práctica, que ayuda a enriquecer el conocimiento del desarrollo de estas tecnologías a gran escala.

Así mismo, se expondrá de forma documentada una aplicación desarrollada dentro de la tecnología celular, cuya implementación ya se ha extendido en los mercados evolucionados del sistema GSM, mostrando las diferentes interfaces, características, funcionalidades y peculiaridades de estos dispositivos como medios de desarrollo e implementación de servicios externos. Mostrando además sus ventajas y desventajas dentro de un modelo de negocio actual dentro de los operadores de telefonía celular.

Resultados esperados

Esta tesis espera generar un aporte a los sectores de investigación de tecnología de alto nivel en México que incursionan en el desarrollo de dispositivos y aplicativos basados en soluciones avanzadas de minicomponentes, a fin de promover actividades que orienten e involucren a los interesados para el desarrollo de sus propios productos, pero sobre todo mostrar los elementos críticos que se requieren para la manufactura y lanzamiento a nivel masivo de estos elementos, con el fin de incentivar a la creación de vías de investigación sobre esta rama de la ingeniería.

Por lo que se pretende mostrar las diversas capacidades tanto técnicas, como económicas que se requieren para mantener este tipo de industria de última tecnología, presentando información real y actualizada de una de las industrias a nivel mundial encargada del diseño y fabricación de estos dispositivos electrónicos,

Uno de los propósitos de este proyecto es, acercar al lector a este campo de la ingeniería, a fin de proporcionarle un ambiente real de los procesos actuales de manufactura a nivel mundial en la industria de Semiconductores.

Finalmente, se espera aportar un documento de consulta académica para los alumnos y docentes interesados en el ámbito de la Ingeniería electrónica, de computación y telecomunicaciones.

PARTE I. TEORÍA Y DISEÑO

CAPÍTULO I
ANTECEDENTES

I.1 Introducción

En la actualidad podemos darnos cuenta que la tecnología conforme avanza permite el desarrollo de nuevos dispositivos electrónicos, de menor tamaño, con mayor capacidad tanto en almacenamiento como en funcionalidad, lo cual ha llevado a que sean utilizados en un sin número de aplicaciones que podemos ver y percibir en nuestra vida cotidiana. Para poder tener una idea del alcance de estos dispositivos, pensemos por un momento en los transistores, probablemente la aplicación tecnológica más importante de los semiconductores. Cualquier habitante del mundo moderno se encuentra rodeado por millones de transistores, están en el televisor, en la computadora, en el automóvil, en el teléfono celular. Todo esto debido a la constante investigación, búsqueda y obtención de nuevos materiales semiconductores.

Uno de los avances más importantes de la electrónica ha consistido en la integración de los dispositivos electrónicos. La idea consiste en colocar un mayor número de dispositivos en el mismo espacio, la oblea de Silicio. Esto sólo puede hacerse disminuyendo el tamaño de los dispositivos y de los demás elementos que les acompañan. Por ello aparece la palabra microelectrónica, para indicar que el tamaño está en la escala de la micra, esta continua disminución de tamaño ha llevado a los métodos de procesamiento a escala submicroscópica, lo cual ha conducido al desarrollo de estructuras cuyas dimensiones están en el rango de milésimas a décimas de un micrómetro, es decir dimensiones nanométricas ($1nm = 10^{-9} m$).

La evolución de este proceso puede seguirse en una gráfica denominada ley de Moore, que es una línea recta y que indica que cada 18 meses se dobla el número de

dispositivos. Actualmente el tamaño está ya por debajo de la micra, más exactamente en el rango de la décima de micra. Otro aspecto importante es que no se han encontrado dificultades serias para que los dispositivos sigan funcionando con los mismos tipos de arreglos e ir siguiendo un mismo principio físico. Por supuesto que ha habido que resolver importantes retos tecnológicos, de manera que la tecnología es cada vez más compleja y más cara.

Este proceso se llama miniaturización y no se sabe hasta dónde puede llegar, o si hay algún límite por debajo del cual los dispositivos presenten dificultades insalvables, o bien la tecnología no permita fabricar elementos tan extremadamente pequeños. Una etapa clave en este proceso de miniaturización de semiconductores es la fotolitografía, también denominada como "microlitografía" o "nanolitografía", la cual se refiere a la fabricación de microestructuras con un tamaño de escala que ronda los nanómetros. Esto implica la existencia de patrones litografiados en los que, al menos, una de sus dimensiones longitudinales es del tamaño de átomos individuales y aproximadamente del orden de $10\ nm$.

Para poderse llevar a cabo la transferencia o impresión de imágenes de patrones, a una oblea de silicio, es necesario el uso de foto máscaras. En cada una de estas foto máscaras están contenidos la información de los patrones a imprimir para una sola capa de un diseño, y dependiendo del tipo de chip que se esta fabricando, podríamos estar hablando de arriba de 30 a 40 capas. El producto final producido por estas foto máscaras es un microchip.

En pocos años, el desarrollo de los semiconductores se ha convertido en una industria próspera y base activa de todo el mercado de la electrónica; además, fuera de este campo exclusivo interviene cada vez más en la economía y contribuye a modelar una nueva civilización. Esto acentúa el potente impulso que han dado los transistores a la

Tercera revolución industrial[1]; como se sabe, ésta tiende a remplazar con las máquinas, no solo la fuerza muscular, sino a la inteligencia.

Actualmente existen dos grandes familias de microestructuras que se fabrican de forma industrial: Los circuitos Integrados sobre substrato semiconductor (Circuitos Integrados Monolíticos) y los circuitos Integrados sobre substrato aislante (Circuitos Híbridos).
Las dos técnicas se complementan cada vez más en lugar de hacerse competencia, puesto que se fabrican circuitos integrados complejos, múltiples, en montajes híbridos de chips, etc.

El Circuito Híbrido

En las estructuras híbridas, los elementos activos (transistores) están incorporados en el propio circuito integrado, el cual no contiene componentes pasivos. Por otra parte los circuitos híbridos se dividen en dos subcategorías:

- Circuitos Híbridos de película gruesa, obtenidos por serigrafía
- Circuitos Híbridos de película delgada, obtenidos por evaporación al vacío y pulverización catódica.

En este tipo de estructuras se producen simultáneamente todos los componentes en el curso de un único proceso, y están depositados sobre un substrato de Silicio.

El Circuito Integrado Elemental Monolítico

Destinado a realizar una función determinada, el circuito integrado monolítico constituye un conjunto indivisible de componentes producidos simultáneamente en el curso de un mismo proceso de fabricación. El material de partida es una placa monocristalina de silicio de tipo p, llamada substrato, de $200\ mm$ a $300\ mm$ de diámetro aproximadamente y de $775\ \mu m$ de espesor. Sobre ésta son creados sucesivamente todos los elementos de un cierto número de circuitos idénticos cuyas dimensiones se procura reducir para aprovechar al máximo la superficie de la placa de silicio.

[1] Eva Leticia Orduña Trujillo, Coacciones y oportunidades de la globalización, Publicado por la UNAM, 2006.

La Integración no constituye simplemente un medio, sino que corresponde a necesidades precisas. Los circuitos integrados tienen 4 propiedades esenciales:

- Volumen Pequeño
- Fiabilidad
- Economía
- Rendimiento

I.2 Descripción de actividades y responsabilidades en el área de trabajo

La industria de la fabricación de semiconductores a nivel mundial requiere siempre de una innovación tecnológica tanto de sus productos como de sus procesos de diseño y fabricación, lo cual lleva a la constante necesidad del análisis de estos, para lograr así estar a la vanguardia y ser competitivos dentro del mercado internacional. Una de las áreas más importantes dentro de esta industria es la Ingeniería, para lo cual se hará una descripción de las actividades y responsabilidades que se tienen en el área de trabajo dentro de una compañía dedicada al diseño y manufactura de semiconductores desde hace 30 años.

Dentro de las actividades que realiza el Ingeniero de Diseño Asistido por Computadora *CAD*[2] en el grupo producción de Tapeout (termino común, para indicar la fase final del ciclo de desarrollo de un componente electrónico), es el ser responsable de hacer eficiente el proceso de Tapeout, manteniendo y asegurando la integridad de la información, esto es, garantizando la calidad de los datos contenidos dentro de los archivos.

Durante la fase final del desarrollo de un componente electrónico, es importante que exista y se mantenga un solo flujo en los archivos que contienen el diseño, eliminando de esta forma cualquier tipo de error que pueda afectar o hacer que la información tenga que ser regresada a un proceso anterior, ocasionando un retrazo en la

[2] Diseño asistido por computadora CAD (Computer Aided Design), es el uso de un amplio rango de herramientas computacionales que asisten a ingenieros, arquitectos y a otros profesionales del diseño en sus respectivas actividades.

fabricación de la foto máscara, incrementando de esta manera el tiempo de entrega y su costo. Por lo tanto es necesaria la elaboración e implementación de nuevos procedimientos automatizados basados en la eliminación de fallas comúnmente detectadas para que de esta forma se garantice la integridad de la información. Como un ejemplo de estas fallas, se encuentra la omisión o introducción de parámetros en forma incorrecta; provocando de esta forma errores en la foto mascara y por consiguiente errores durante la impresión sobre el silicio.

Estos procedimientos generalmente son programas o rutinas en lenguaje de programación PERL (Lenguaje Práctico para la Extracción e Informe *Practical Extracting and Reporting Languaje*). Esto es debido a que PERL combina en forma concisa las mejores características de lenguajes como C, sed, awk y sh, pero está enfocado a ser más práctico y fácil de aprender. PERL también es utilizado debido a la ventaja de poder correr sin cambios sobre casi cualquier plataforma utilizada en el área de Tapeout (Windows, Linux, Solaris[3]), lo que convierte a PERL en el lenguaje ideal para desarrollo de prototipos y aplicaciones robustas 100% portables. Otra de las razones por la cual se utiliza este lenguaje de programación es debido a que, PERL es útil en la resolución de cualquier tarea y posee habilidades para integrarse con sistemas operativos, bases de datos, redes, protocolos, ambientes gráficos, otros lenguajes de programación (JAVA, C, etc.). Su versatilidad y eficiencia en el manejo de texto y, específicamente, de "expresiones regulares" no tiene equivalente en ningún otro lenguaje de programación actual.

Otra parte importante dentro de las actividades es el desarrollo de paginas Web dinámicas, utilizando lenguaje PERL CGI (*Common Gateway Interface* o Interfaz de entrada común), para poder de esta forma acceder a los procesos de Tapeout que corren sobre un ambiente de BPM (*Bussines Process Management,* que se conoce como la metodología empresarial, cuyo objetivo es mejorar la eficiencia a través de la gestión sistemática de los procesos de negocio, los cuales deben de ser modelados,

[3] Solaris es un sistema operativo de tipo Unix desarrollado por Sun Microsystems desde 1992 como sucesor de SunOS, que fue la versión desarrollada para sus estaciones de trabajo y servidores.

automatizados, integrados, monitoreados y optimizados de forma continua). Como su nombre sugiere, BPM se enfoca en la administración de los procesos del negocio. A través del modelado de las actividades y procesos puede lograrse un mejor entendimiento del negocio y muchas veces esto representa la oportunidad de mejorarlos.

La automatización de los procesos reduce errores, asegurando que los mismos se comporten siempre de la misma manera, dando elementos que permiten visualizar el estado de los mismos. La administración de los procesos permite asegurar que los mismos se ejecuten eficientemente y que posteriormente, la obtención de información puede ser usada para mejorarlos. Es a través de la información que se obtiene de la ejecución diaria de los procesos, que se puede identificar posibles ineficiencias en los mismos y actuar sobre las mismas para optimizarlos.

Siendo parte importante la automatización de procesos en el área de Tapeout, se puede mencionar como parte de las actividades realizadas la creación de una plataforma de Monitoreo y creación de copias de seguridad, la cual permite el almacenamiento, actualización y notificación tanto de su estado actual de los archivos y discos duros utilizados. Permitiendo de esta forma el acceso a un registro histórico de los archivos utilizados y su ubicación actual.

Otra de las actividades dentro de esta área de trabajo es el desarrollo y mantenimiento de una herramienta para la elaboración y verificación del archivo de instrucciones (*jobdeck*), este archivo es utilizado durante el proceso de fabricación de la foto máscara (técnica parecida a la fotolitografía de circuitos integrados) el cual contiene las posiciones o coordenadas de las estructuras que definen la zona de la exposición que se requieren transferir a la máscara (Figura I.1), este archivo es creado en un formato que puede ser leído por las maquinas encargadas de la escritura de la máscara.

```
SLICE 1,17
*
OPTION PA,M,VA=10
RETICLE
*
*************************************
*  6 INCH MICRON TECH TITLE LOCATIONS  *
*************************************
ORIENT  A,MTITLE,TITLEROT=180,LOC=121000,147000,JUST=L
ORIENT  A,DTITLE,TITLEROT=180,LOC=121000,147000,JUST=L
ORIENT  A,TTITLE,TITLEROT=180,LOC=60000,147000,JUST=L
ORIENT  A,PTITLE,TITLEROT=180,LOC=60000,145000,JUST=L
ORIENT  A,NTITLE,TITLEROT=180,LOC=40000,147000,JUST=L
*************************************
*
CHIP L73X_FRAME_4X,
$          (001,L73X550-A5-BS,AD=0.001,SF=1)
ROWS       76199.98600000/     76311.74600000
*
CHIP L73X_FRAME_LVL2_4X,
$          (050,L73X550-A5-BY,AD=0.01,SF=1)
ROWS       76199.98600000/     76311.74600000
*
CHIP L73X_DIE_4X,
$          (001,L73X55B-05-0F,AD=0.001,SF=1)
ROWS      107921.58400000/     54495.68600000
ROWS      107921.58400000/     97904.31400000
ROWS       44478.41600000/     54495.68600000
ROWS       44478.41600000/     97904.31400000
*
```

Figura I.1 Archivo de instrucciones (jobdeck)

Las figuras impresas en la foto máscara se crean a partir de archivos que contienen los patrones de diseño, en este caso el tipo de archivo de datos al que nos referimos es MEBES (*Moving Electron Beam Exposure* o exposición a base del movimiento de un haz de electrones).

Este tipo de archivo MEBES, contiene los patrones de los diseños que se van a imprimir sobre las obleas de silicio, estos archivos son generados durante la última etapa en el diseño de una foto máscara antes de ser transferido físicamente a una placa de cuarzo. Este proceso o etapa es normalmente conocida con el nombre fractura y preparación de datos (*Fracture*).

Durante este proceso, existen parámetros que son definidos por los diseñadores los cuales deben ser aplicados durante la fractura a los archivos que contienen el diseño

de la foto máscara, por lo cual se debe mantener siempre la integridad tanto en los parámetros como en los archivos que contienen el diseño. Cualquier falla o cálculo aun hablando en diferencia de micras 10^{-6} (0.000001) o nanos 10^{-9} (0.000000001), durante la preparación de los datos puede dar como resultado una placa incorrecta, ya sea que los patrones de diseño se encuentren desfasados con el anterior nivel o simplemente no pueda existir una compatibilidad entre los niveles, perdiendo de esta manera la comunicación o continuidad entre ellos, causando así, el aumento del costo total, ya que la complejidad del diseño de la foto máscara es proporcional al costo de fabricación de la misma. Es necesario también que se realicen rutinas para ser comparados los resultados obtenidos de la fractura, para que se logre así la identificación de cualquier tipo de error que pueda afectar el desempeño de la máscara. Por lo cual se ha requerido el desarrollo y mantenimiento de un programa que permite la creación de los archivos con las instrucciones necesarias para la realización de la fractura, permitiendo de esta forma la reducción y eliminación de los errores ocurridos durante la preparación de los datos.

Otra de las actividades dentro de esta área es la constante reducción de costos y optimización del área de trabajo, se requiere de una evaluación permanente de las aplicaciones y sistemas operativos utilizados, estas acciones han requerido que los programas actualmente instalados en el sistema operativo Solaris se migren paulatinamente a un plataforma Linux, esto es debido a que se han encontrado mayores ventajas al igual que reducción de costos sin la pérdida de calidad e integridad de la información durante la utilización de estas aplicaciones sobre una nueva plataforma. Se requiere hacer pruebas antes y después de haberse realizado la migración. En el caso, en que las aplicaciones presenten faltantes o inconsistencias, es necesario que se analice directamente el código fuente, para identificar las funciones o partes del código que por su anterior definición, necesiten ser remplazadas para poder lograr los mismos resultados, obteniendo al final, nuevas definiciones de funciones y un código actualizado.

PARTE I. TEORÍA Y DISEÑO

CAPÍTULO II
DISEÑO

II.1 Definición y características de los semiconductores

Definición de semiconductores

Los semiconductores son materiales cuya conductividad varía con la temperatura, pudiendo comportarse como conductores o como aislantes, por lo consiguiente se desean obtener variaciones en la conductividad controlables eléctricamente y no a base del uso de temperatura. Para conseguir esto, se introducen átomos de otros elementos en el semiconductor. Estos átomos se llaman impurezas y tras su introducción, el material semiconductor presenta una conductividad controlable eléctricamente.

Existen dos tipos de impurezas, las P y las N, que cambian la conductividad del silicio y determinan el tipo de cristal a fabricar. Por tanto, como hay dos tipos de impurezas habrá dos tipos fundamentales de cristales, cristales de impurezas P y cristales de impurezas tipo N.

El Semiconductor es un material aislador en que el ancho de banda prohibida es menor que $1\,eV$. A fin de precisar nuestra definición de semiconductor recordemos que a temperatura ambiente ($T \approx 300\,K$) la energía térmica transferida a un electrón de la red es del orden de ($kT = 0.025\,eV$), esta energía es suficiente para que una pequeña fracción de los electrones en la banda de valencia pueda “saltar” a la banda desocupada.

Sin embargo, a temperatura nula ningún electrón podrá ocupar la banda superior. Por lo tanto, los semiconductores $T = 0\,K$ son aisladores. Esto nos permite definir los

semiconductores como aisladores de banda prohibida angosta. La Figura II.1 muestra esquemáticamente las diferencias existentes entre las tres clases de sólidos de acuerdo a sus propiedades de transporte de carga.

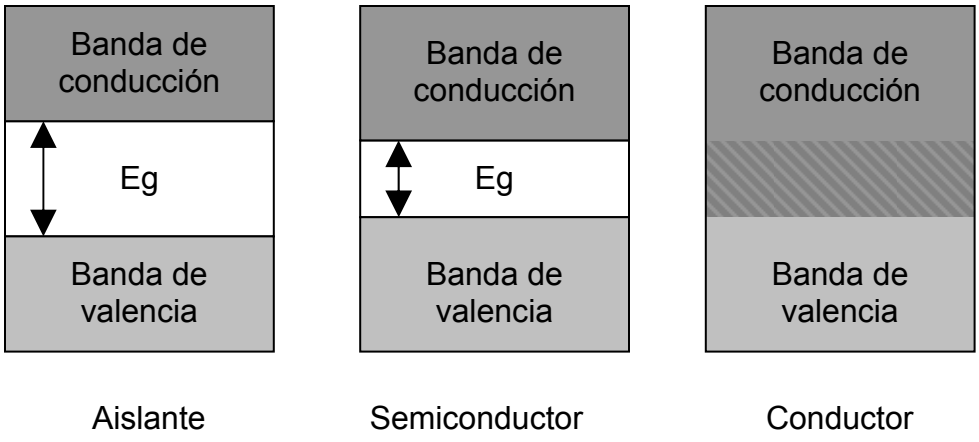


Figura II.1 Metales, semiconductores y aisladores desde el punto de vista de la teoría de bandas.

El material semiconductor más utilizado es el Silicio (*Si*), pero hay otros semiconductores como el Germanio (*Ge*) que también son usados en la fabricación de circuitos. El silicio está presente de manera natural en la arena por lo que se encuentra con abundancia en la naturaleza. Su purificación es relativamente sencilla (alcanzando una pureza del 99,99999%), otra característica importante en el Silicio es que se presta fácilmente a ser oxidado, formándose de esta manera Oxido de Silicio ($Si\,O_2$) y el cual constituye un aislante utilizado en todos los transistores de la tecnología CMOS.

Tipos de semiconductores

Semiconductor elemental

El mejor semiconductor conocido es el elemento Silicio (*Si*). Junto con el Germanio (*Ge*), son el prototipo de una larga clase de semiconductores, teniendo en común estructuras cristalinas.

La estructura cristalina del (*Si*) y el (*Ge*) es igual a la del diamante y al del estaño-α (semiconductor de cero-apertura o también conocido como estaño gris). En este tipo de estructura cada átomo esta rodeado por cuatro átomos formando así un tetraedro. (Figura II.2).

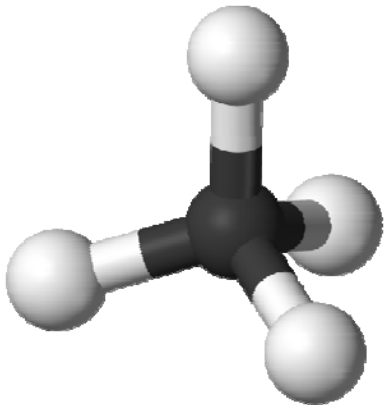


Figura II.2 Estructura en forma de tetraedro.

Este semiconductor con enlace tetraédrico es el pilar de la industria de la electrónica y la piedra angular de la tecnología moderna. Algunos de los elementos de los grupos V y VI de la tabla periódica, como el Fósforo (*P*), Azufre (*S*), Selenio (*Se*) y Telurio (*Te*), también son semiconductores. Los átomos en estos cristales pueden alojar o coordinar tres electrones (*P*), dos electrones (*S*, *Se*, *Te*) o cuatro electrones. Como resultado, estos elementos pueden existir en diversas estructuras cristalinas y también como formadores de vidrio.[1]

Compuesto binario

Los compuestos formados por los elementos de los grupos III y V de la tabla periódica como el *GaAs* (semiconductor binario Galio y Arsénico) tienen propiedades muy similares a la de su contraparte del grupo IV. Al ir de los elementos del grupo IV al los compuestos del III-V, las uniones se vuelven en parte iónicas debido a la transferencia de carga electrónica de los átomos del grupo III a los átomos del grupo V. Los iones

[1] Peter Y. Yu, Manuel Cardona, Fundamentals of Semiconductors: Physics and Materials Properties, Third Edition, Springer, 2001.

producen cambios significativos el las propiedades del semiconductor. Esto incrementa la interacción de Coulombs[2] entre los iones al igual que la energía en la estructura de las bandas de electrones. En los compuestos como el ZnS (Sulfuro de Zinc) la unión de los iones se vuelve más grande e importante. Como resultado, la mayoría de los semiconductores compuestos del grupo II-VI tienen una energía cinética entre las bandas mayores a 1 *eV* (electronvoltio). La excepción de este tipo de compuestos son los que contienen elementos pesados como el Mercurio (Hg).

Óxidos

Aunque la mayoría de los óxidos son buenos aislantes, algunos como el CuO y Cu_2O son mejor conocidos como semiconductores. El oxido cuproso (Cu_2O) se produce como un mineral (Cuprita), una clase de semiconductor y que se han estudiado extensivamente sus propiedades. En general, los óxidos semiconductores no han sido entendidos completamente, solamente se ha considerando el proceso de crecimiento, por lo cual tiene una potencia limitada de aplicaciones en la actualidad. Una excepción es el componente II-VI Oxido de Zinc (ZnO), del cual se han encontrado aplicaciones como transductor y como un ingrediente en las cintas adhesivas. Sin embargo, esta situación ha cambiado con el descubrimiento de la superconductividad en la mayoría de los óxidos de cobre.

Semiconductores en capas

Los compuestos para semiconductores como el yoduro de plomo (PbI_2), Bisulfuro de Molibdeno (MoS_2) y el Seleniuro de Galio ($GaSe$) son característicos por su estructura cristalina en capas. Típicamente los enlaces con las capas son covalentes y mas fuertes que *las fuerzas de van der Waals*[3] entre las capas. Estos semiconductores en capas han sido de interés debido al comportamiento casi bidimensional de los electrones dentro de las capas. También, puede ser modificada la interacción entre

[2] El coulomb (símbolo C), es la unidad derivada del SI para medir la cantidad de carga eléctrica transportada por un flujo de un ampere durante un segundo.

[3] Las fuerzas de van der Waals son fuerzas de estabilización molecular; forman un enlace químico no covalente en el que participan dos tipos de fuerzas o interacciones, las fuerzas de dispersión (que son fuerzas de atracción) y las fuerzas de repulsión entre las capas electrónicas de dos átomos contiguos.

capas para incorporar átomos ajenos entre ellas, este proceso es conocido como intercalación.

Semiconductores Orgánicos

Algunos compuestos orgánicos como el poliacetileno $[(CH_2)_n]$ y el polidiacetileno tienen la característica de ser semiconductores. Aunque los semiconductores orgánicos no son usados aun en dispositivo electrónicos, se mantienen como una gran promesa en el desarrollo de futuras aplicaciones. La ventaja de los semiconductores orgánicos sobre los inorgánicos es la de que pueden ser adaptados a las aplicaciones fácilmente. Por ejemplo, los compuestos que contienen enlaces conjugados como $-C=C-C=$ tienen una larga óptica no lineal por lo tanto pueden tener importantes aplicaciones en la optoelectrónica[4]. Para adaptarse mejor a una aplicación en el espacio entre las bandas de estos compuestos puede ser modificado más fácilmente que los del tipo inorgánico a base de cambios en su formula química.

Semiconductores Magnéticos

La mayoría de los compuestos contienen iones magnéticos como el Europio (*Eu*) y Manganeso (*Mn*), los cuales presentan propiedades magnéticas y semiconductoras, Dependiendo de la cantidad de iones magnéticos concentrados en estas aleaciones, el compuesto puede presentar diferentes tipos de propiedades magnéticas como el ferromagnetismo[5] y antiferromagnetismo. Los semiconductores de aleación magnética que contienen menores concentraciones de iones magnéticos son conocidos como semiconductores magnéticos diluidos.

Otras variedades de semiconductores

Existen muchos semiconductores que no entran en las categorías mencionadas anteriormente. Como por ejemplo, el semiconductor *SbSI* que muestra

[4] La optoelectrónica es el nexo de unión entre los sistemas ópticos y los sistemas electrónicos. Los componentes optoelectrónicos son aquellos cuyo funcionamiento esta relacionado directamente con la luz.

[5] El ferromagnetismo es un fenómeno físico en el que se produce ordenamiento magnético de todos los momentos magnéticos de una muestra, en la misma dirección y sentido.

ferroelectricidad[6] a bajas temperaturas solamente. Compuestos que tengan una formula general del tipo I-III-VI_2 y II-IV-V_2 (como el sulfuro de plata con Galio $AgGaS_2$, interesante por sus propiedades ópticas no lineales, el $CuInSe_2$ usado en celdas solares y el $ZnSiP_2$) se cristalizan en un sistema tetragonal como la calcopirita (Figura II.3).

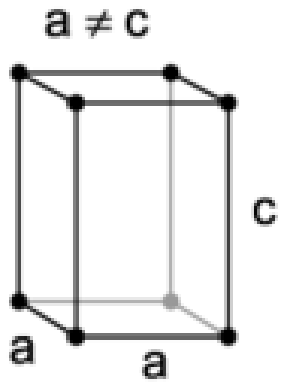


Figura II.3 Representación de un sistema tetragonal

En estos compuestos los enlaces también son tetraédricos y pueden ser considerados análogos a los semiconductores del grupo III-V y II-VI, teniendo una estructura cristalina como la esfalerita o blenda (Figura II.4).

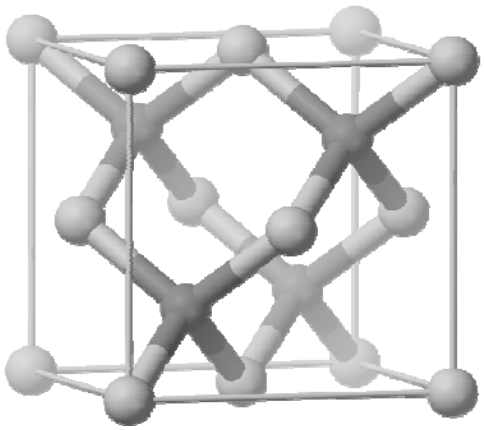


Figura II.4 Estructura tetraédrica de la esfalerita

[6] Es la capacidad de ciertos materiales para retener información en su estructura cristalina, sin necesidad de estar conectados a una fuente de energía. La información es almacenada gracias a la polarización eléctrica que poseen, que puede ser activada externamente por un voltaje, y aún cuando éste sea retirado, la polarización persiste.

II.2 Definición y característica de los VLSI, CMOS

Un circuito integrado IC (*Integrated Circuit*) es un cristal semiconductor se silicio, llamado chip, que contiene los componentes electrónicos para construir compuertas digitales. Las diversas compuertas se interconectan dentro del chip para formar el circuito requerido. El chip se monta en un recipiente de cerámica o plástico, y las conexiones se soldan a terminales externas para formar el circuito integrado. El número de terminales podría variar desde 14 en un paquete de IC pequeño hasta varios miles en los paquetes más grandes. Cada IC tiene una designación numérica impresa en la superficie del paquete, para poder identificarlo. Los fabricantes proporcionan libros de datos, catálogos y sitios Web de Internet que contienen descripciones e información acerca de los IC que producen.

Niveles de integración

Los IC digitales suelen clasificarse según la complejidad de sus circuitos, la cual se mide por el número de compuertas lógicas incluidas en el paquete. La diferenciación entre los chips que tienen pocas compuertas internas y los que tienen cientos de miles de compuertas suelen hacerse diciendo que el paquete es un dispositivo de integración pequeña, mediana, gran o muy grande escala.

Los dispositivos de integración a pequeña escala SSI (*Small-Scale Integration*) contienen varias compuertas independientes en un solo paquete. Las entradas y salidas de las compuertas se conectan directamente a las terminales del paquete. El número de compuertas suele ser menor que 10 y está limitado por el número de terminales que cuenta el IC.

Los dispositivos de integración a mediana escala MSI (*Médium-Scale Integration*) tienen una complejidad de entre 10 y 1000 compuertas en un solo paquete. Por lo regular, efectúan operaciones digitales elementales específicas. Los dispositivos de integración a gran escala LSI (Large-Scale Integration) contienen miles de compuertas

en un solo paquete. Incluyen sistemas digitales como procesadores, chips de memoria y dispositivos de lógica programable.

Los dispositivos de integración a muy grande escala VLSI (*Very Large-Scale Integration*) contienen cientos de miles de compuertas en un solo paquete. Como ejemplo podemos citar las grandes matrices de memoria y los microprocesadores complejos. En virtud de su pequeño tamaño y bajo costo, los dispositivos VLSI han revolucionado la tecnología de diseño de sistemas de cómputo y confieren al diseñador la capacidad de crear estructuras que antes no resultaban económicas construir.

El desarrollo de sistemas VLSI ha progresado históricamente gracias a las innovaciones tecnológicas, A menudo por causas de logros recientes en el proceso de litografía, por componentes semiconductores, mentalización que ha llevado a la introducción de nuevos productos. A la inversa, la demanda en el mercado por productos particulares o específicos tiene una gran influencia enfocada a la investigación de las capacidades de la tecnología necesarias para entrega del producto. Los primeros chips semiconductores contenían sólo un transistor cada uno. A medida que la tecnología de fabricación fue avanzando, se agregaron más y más transistores, y en consecuencia más y más funciones fueron integradas en un mismo chip.

Muchos de los sistemas VLSI convencionales han sido creados en base a una avanzada y especializada tecnología. En contraste con los recientes productos VLSI, tenemos una integración de diversos sistemas que requieren una tecnología y plataforma única. El movimiento detrás de esta tendencia es la demanda de productos electrónicos compactos e inalámbricos, por parte del consumidor al igual que de sectores no comerciales.

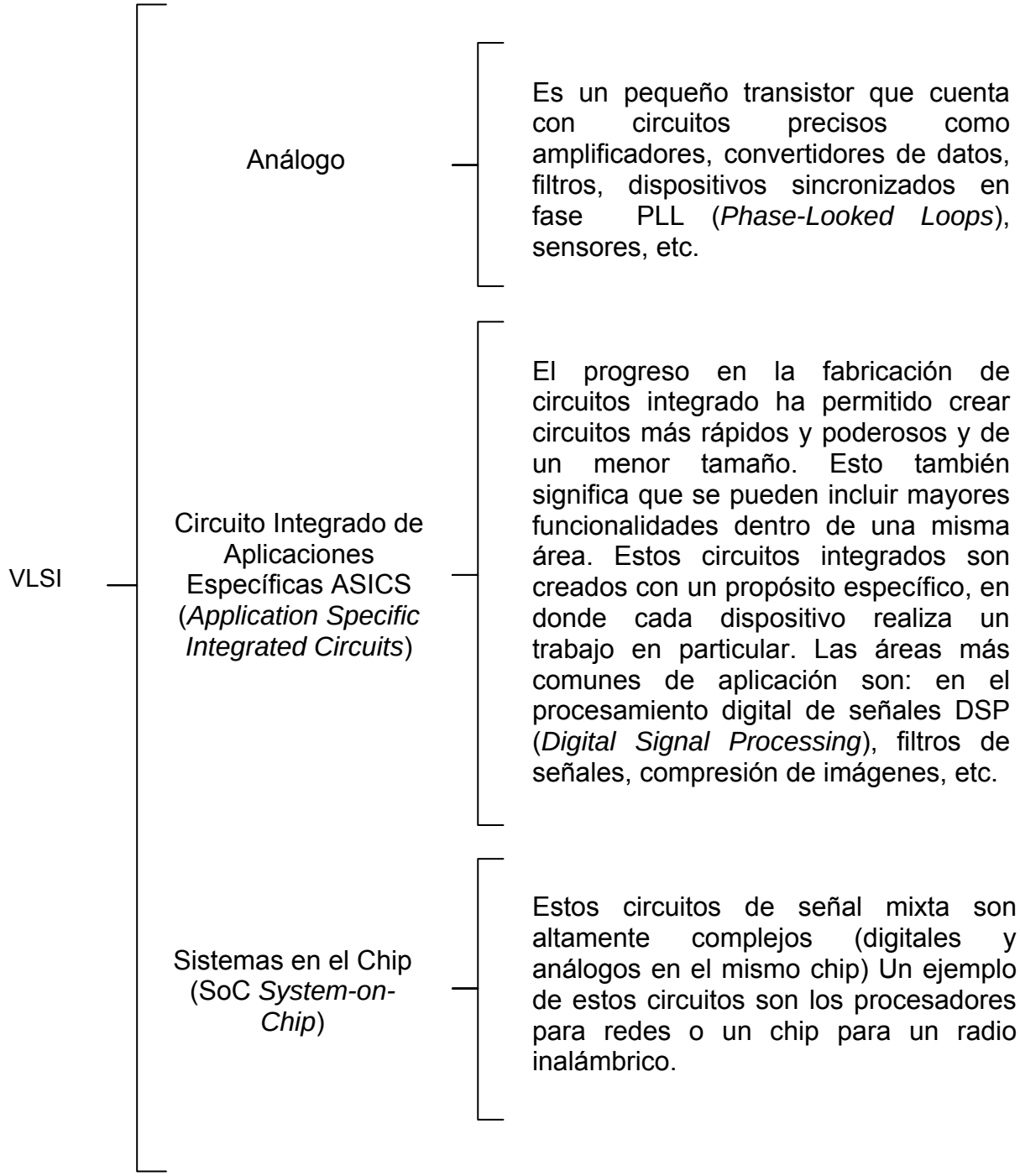
En la (Figura II.5) se ilustra algunos de los sistemas o componentes que juegan un rol importante en este desarrollo. La mayoría de los logros en sistemas VLSI densos, ha sido en base al aumento de escala en el proceso de la oblea de silicio. Al igual que en

la manufactura existe un avance, el costo-rendimiento de los VLSI es superior en la mayoría de las aplicaciones en las cuales se remplazar un chip con un circuito integrado monolítico: bajo costo de empaquetamiento, la disminución en el tamaño de los conectores y reducción en la pérdida de poder de los modules de Entrada/Salida.
Al continuar reduciendo la escala a dimensiones menores de un micrómetro, se produce un incremento mayor en las aplicaciones de los sistemas VLSI integrados, pero también esto ha llevado, a tener una mayor complejidad en interconexión y foto litografía.

Esta evolución ha planteado diferentes preguntas, alguna de estas, como el óptimo nivel de integración: a nivel de paquete o a nivel de chip. Cada una tiene distintas ventajas y algunas deficiencias críticas, como en el costo, calidad y desempeño.

Aunque los Chip con interconexiones a nivel de tarjeta tengan las características de tener bajos costos y grandes volúmenes de producción, los núcleos de sistemas VLSI con integración densa, no pueden proveer adecuadamente un alto desempeño.

Los diseños de VLSI están clasificados en tres categorías:


VLSI
Análogo
Es un pequeño transistor que cuenta con circuitos precisos como amplificadores, convertidores de datos, filtros, dispositivos sincronizados en fase PLL (*Phase-Looked Loops*), sensores, etc.
Circuito Integrado de Aplicaciones Específicas ASICS (*Application Specific Integrated Circuits*)
El progreso en la fabricación de circuitos integrado ha permitido crear circuitos más rápidos y poderosos y de un menor tamaño. Esto también significa que se pueden incluir mayores funcionalidades dentro de una misma área. Estos circuitos integrados son creados con un propósito específico, en donde cada dispositivo realiza un trabajo en particular. Las áreas más comunes de aplicación son: en el procesamiento digital de señales DSP (*Digital Signal Processing*), filtros de señales, compresión de imágenes, etc.
Sistemas en el Chip (SoC *System-on-Chip*)
Estos circuitos de señal mixta son altamente complejos (digitales y análogos en el mismo chip) Un ejemplo de estos circuitos son los procesadores para redes o un chip para un radio inalámbrico.

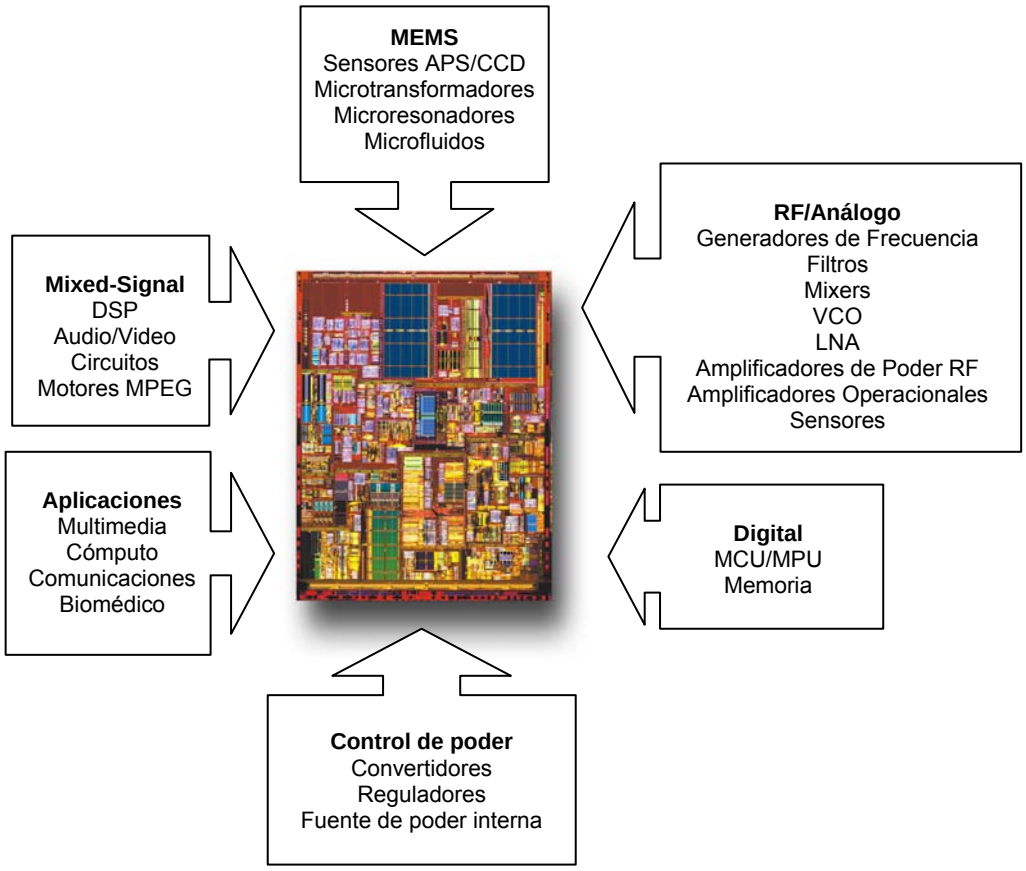


Figura II.5. Cada bloque representa tipos de componentes de un sistema

En la práctica, los sistemas VLSI tienen como adversario el empaquetamiento y la integración a nivel de chip, debido a que durante la implementación se tienen dimensiones compactas y señales de interconexión cortas. También ofrecen un intercambio entre integración monolítica densa y una tecnología de optimización en aplicaciones específicas. Actualmente no es claro el futuro de la evolución de los sistemas VLSI, pero se sabe que seguirán siendo influenciados por el desarrollo de la tecnología.

MOSFET

El Transistor de campo Metal-Oxido-Semiconductor MOSFET (*Metal-Oxide-Semiconductor Field-Efect Transistor*) es un dispositivo de efecto de campo que utiliza un campo eléctrico para crear una canal de conducción. Es un dispositivo más

importante que el JFET ya que la mayor parte de los circuitos integrados digitales se construyen con la tecnología MOS. Existen dos tipos de MOSFET: el de empobrecimiento y el de enriquecimiento.

MOSFET del tipo de empobrecimiento

El MOSFET del tipo de empobrecimiento tiene características muy parecidas a las de la JFET. Su canal puede ser de tipo *n* o de tipo *p*. En el caso del canal tipo *p*, los tipos de portadores y las polaridades de las tensiones e intensidades serán contrarias a la del canal tipo *n*. El la (Figura II.6) se muestra una sección de un MOSFET de empobrecimiento del canal tipo *n*. Partimos de un substrato de silicio ligeramente contaminado de tipo *p*. Se contamina luego la zona del canal para convertirla en región tipo *n*, rotulada n^+, proporcionan dos regiones de gran conductancia que enlazan el canal con los contactos óhmicos conectados a los terminales de fuente y drenaje (sumidero). El electrodo de la compuerta está aislado del canal (de aquí las siglas IGFET *Isolated Gate FET*) por una capa de óxido de silicio, que es aislante. Los tres electrodos son metálicos, normalmente de aluminio.

El substrato suele conectarse a la fuente, como se indica en la (Figura II.7). Si se deja flotante la compuerta entre el drenaje y la fuente se aplica una V_{DS} positiva, la unión p-n existente entre las regiones (*n* n^+) y la región *p* estará polarizada inversamente, con lo que se crea una zona de empobrecimiento, tal como se indica en la (Figura II.7). Así pues, en el funcionamiento normal de un MOSFET, el substrato tiene relativamente poca importancia. En el canal, los electrones se mueven de la fuente al drenaje, dando lugar a una I_D positiva.

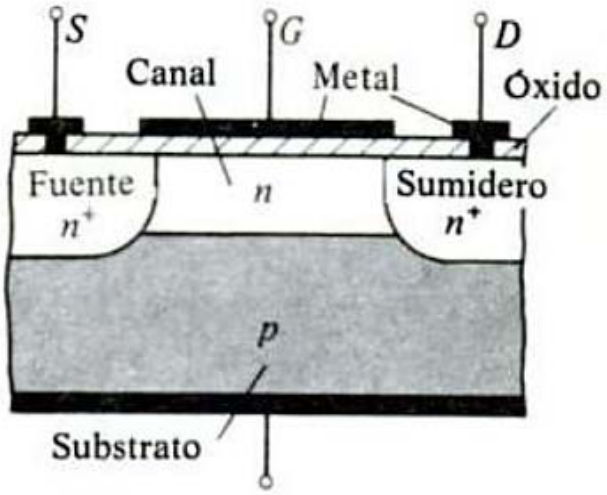


Figura II.6 Estructura básica.

Si hacemos ahora negativa la compuerta y positivo el drenaje, respecto a la fuente, inmediatamente debajo de la compuerta se crea otra región de empobrecimiento a causa del efecto capacitivo entre compuerta y canal, según se ilustra en la (Figura II.8).

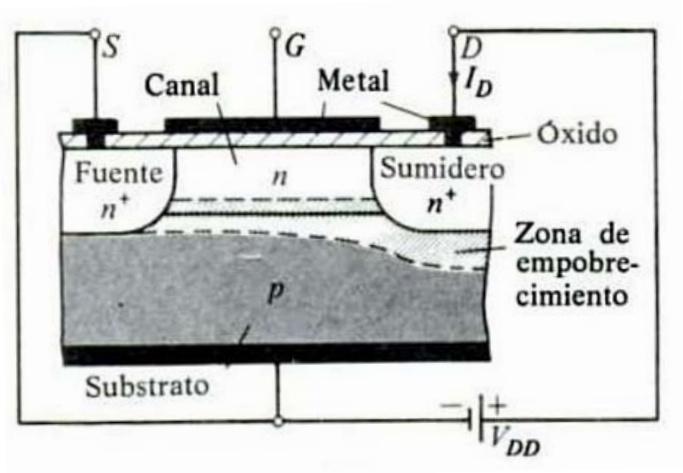


Figura II.7. La zona sombreada representa la zona de empobrecimiento entre las regiones tipo *n* y el substrato.

Al ir haciendo más negativa V_{DS} , se ensancha la zona de empobrecimiento y se estrecha el canal. Un MOSFET de empobrecimiento es análogo a un JFET y para

valores bajos de V_{DS} actúa como resistencia gobernada por tensión. Cuando $|V_{GS}|$ se hace suficientemente grande, el canal se bloquea. Más allá del bloqueo, la intensidad de la corriente del drenaje se mantiene prácticamente inalterada para una amplia gama de valores de V_{DS}.

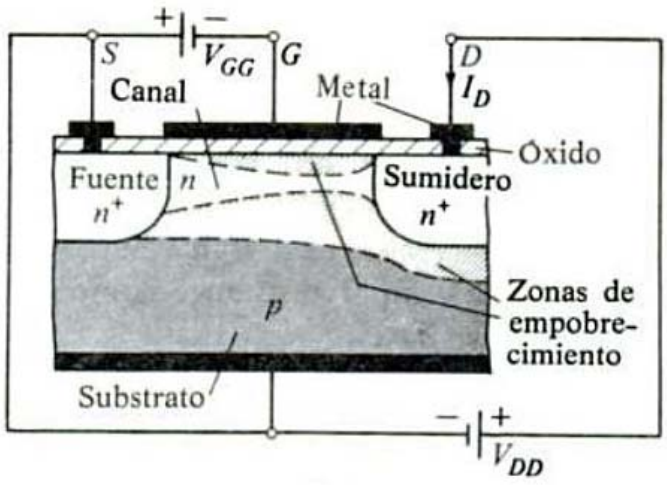
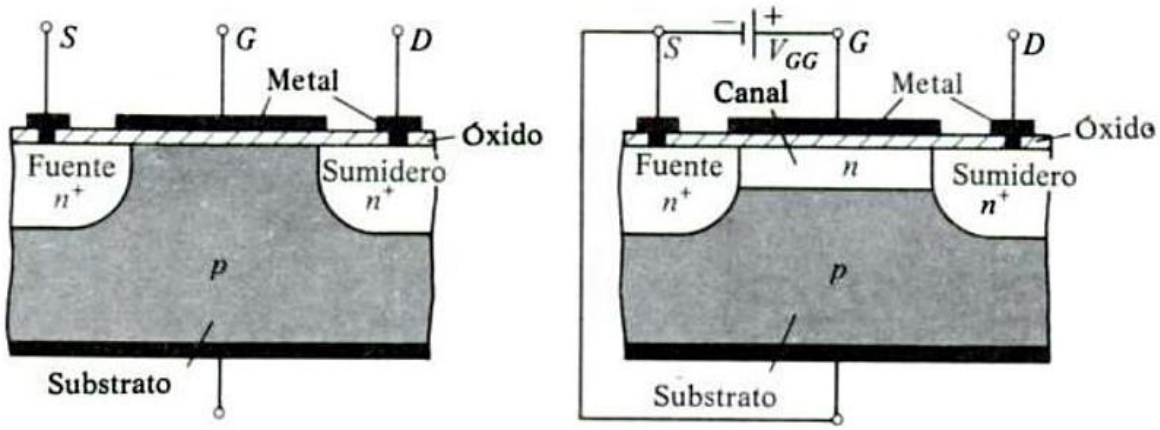


Figura II.8. Zonas del canal y de empobrecimiento: V_{GS} es negativa y V_{DS} es positiva.

MOSFET del tipo de enriquecimiento

El MOSFET del tipo de enriquecimiento se diferencia del de empobrecimiento en que en él no existe la región de tipo n comprendida entre las dos regiones de tipo n^+. En la (Figura II.9) se ha representado una sección de este tipo de MOSFET. Si la compuerta la dejamos flotante y aplicamos una tensión entre el drenaje y la fuente, circulará una corriente de intensidad despreciable ya que una u otra de las regiones n^+ estará polarizada inversamente respecto al substrato y la única corriente que podrá circular será la de saturación inversa de una de las uniones *p*-*n*.



Estructura básica

Formación de un canal de tipo n por una V_{GS} positiva mayor que V_{TH} .

Figura II.9 Sección de un MOSFET de enriquecimiento con canal *n*.

Si se hace la compuerta más positiva respecto al substrato y la fuente, el efecto capacitivo entre la compuerta y el substrato repelerá a los portadores de tipo p alejándolos del área inmediatamente debajo de la compuerta y atraerá hacia ella a los del tipo n. Al ir haciendo V_{GS} gradualmente más positiva, disminuirán gradualmente los portadores de tipo p en esta región y aumentarán los de tipo n. Para un cierto valor de V_{GS} , esta región tendrá más portadores de tipo n que de tipo n. Cuando se alcanza este valor de V_{GS} , se forma un canal, tal como se indica en la (Figura II.9). A este valor de V_{GS} se le da el nombre de tensión umbral y se designa por V_{TH} (TH *Threshold*). Al aumentar V_{GS} aún más, se atraen más portadores tipo n hacia el canal, con lo que se hace más conductor.

Una vez formado el canal, este dispositivo tiene propiedades muy parecidas a las del MOSFET de empobrecimiento. Por ejemplo, cuando se hace el drenaje positivo respecto a la fuente y al substrato, se forma una zona de empobrecimiento entre el área que contiene al canal y la región del drenaje (ambas son de tipo *n*) y el substrato (tipo *p*). De nuevo, el substrato tiene poca importancia para el funcionamiento del dispositivo.

Para $V_{GS} > V_{TH}$ y si se hace aumentar V_{DS} desde cero en sentido positivo, para valores bajos de V_{DS} el canal es, esencialmente, una resistencia. Al aumentar aún más V_{DS} , disminuirá $V_{GD} = V_{GS} - V_{DS}$. Cuando V_{GD} cae por debajo de V_{TH} , se produce el bloqueo e i_D se mantiene esencialmente constante para una amplia gama de valores de V_{DS} .

Un MOSFET de empobrecimiento con canal *n* funcionará también en el modo de enriquecimiento cuando se haga positiva su compuerta respecto a la fuente. Al ir haciendo la compuerta más positiva respecto al substrato, se llevarán más portadores del tipo *n* hacia el canal, elevando así su conductividad. Por lo tanto, un MOSFET de empobrecimiento puede tener su compuerta polarizada positiva o negativamente respecto a la fuente.

Los circuitos digitales que emplean MOSFETs se dividen en tres categorías:

- P-MOS, en los que sólo se usan MOSFETs de acrecentamiento del canal P.
- N-MOS, en los que sólo se usan MOSFETs de acrecentamiento del canal N.
- CMOS, (MOS complementario), en los que se usan ambos dispositivos de canales P y N.

En los circuitos N-MOS se usa un MOSFET de canal N como un interruptor y también se implementan todas las resistencias usando la resistencia del canal de un MOSFET con la compuerta conectada al drenaje (siempre ENCENDIDO).
Los circuitos P-MOS son similares, pero en ellos se emplean MOSFETs de canal P. La simplicidad de estos circuitos y de sus procesos de manufactura hizo que dominaran los primeros mercados de LSI y VLSI. Las ventajas de velocidad y potencia que ofrecen la tecnología actual de manufactura CMOS han hecho de estos el líder en todos los niveles de integración.

CMOS

En la familia lógica CMOS, el término complementario se refiere a la utilización de dos tipos de transistores en el circuito de salida, en una configuración similar a la tótem-pole de la familia TTL. Se usan conjuntamente MOSFET de canal n (N-MOS) y de canal p (P-MOS) en el mismo circuito, para obtener varias ventajas sobre las familias P-MOS y N-MOS. La tecnología CMOS es ahora la dominante debido a que es más rápida y consume aún menos potencia que las otras familias MOS. Estas ventajas son opacadas un poco por la elevada complejidad del proceso de fabricación del IC y una menor densidad de integración. De este modo, los CMOS todavía no pueden competir con MOS en aplicaciones que requieren lo último en LSI.

La lógica CMOS ha emprendido un crecimiento constante en el área de la MSI, principalmente a expensas de la TTL, con la que compite directamente. El proceso de fabricación de CMOS es más simple que el TTL y tiene una mayor densidad de integración, lo que permite que se tengan más circuitos en un área determinada de sustrato y reduce el costo por función. La gran ventaja de este tipo de componentes es que presentan una alta impedancia (resistencia eléctrica) de entrada consumiendo por lo tanto muy poca corriente eléctrica. El bajo consumo de corriente eléctrica hace a esta tecnología atractiva en aplicaciones en las cuales es imperativo el ahorro de energía (principalmente artículos electrónicos operados con batería portátil desechable).

Puertas lógicas de la familia CMOS

INVERSORES CMOS

Un dispositivo CMOS consiste en distintos dispositivos MOS interconectados para formar funciones lógicas. Los circuitos CMOS combinan transistores P'MOS y N'MOS, cuyos símbolos más comunes son los que se muestran en la (Figura II.10). La circuitería del INVERSOR CMOS básico se muestra en la (Figura II.11). El INVERSOR CMOS tiene dos MOSFET en serie de modo que, el dispositivo con canales *p* tiene su fuente conectada a $+V_{DD}$ (un voltaje positivo) y el dispositivo de canales *n* tiene su fuente conectada a tierra. Las compuertas de los dos dispositivos se interconectan con

una entrada común. Los drenajes de los dos dispositivos se interconectan con la salida común.

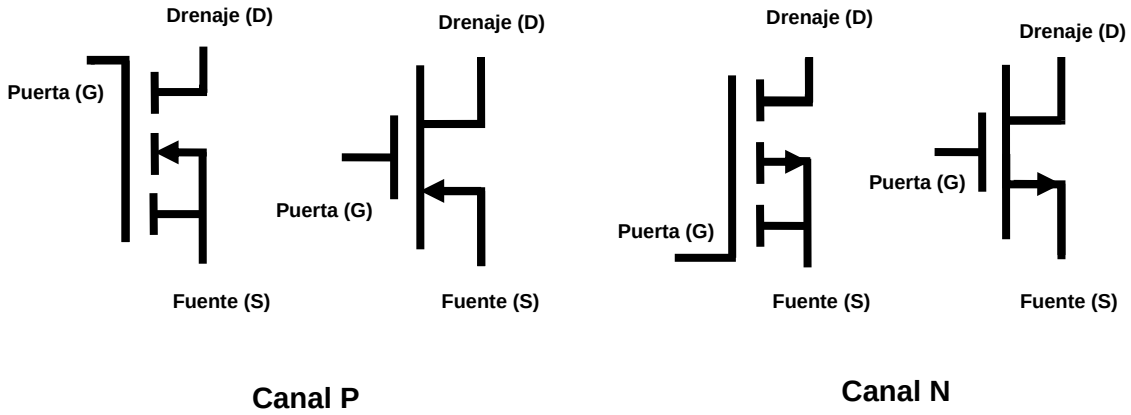


Figura II.10. Símbolos más comunes de los transistores P-MOS y N-MOS

El circuito mostrado en la (Figura II.11) representa un INVERSOR CMOS y está formado por un transistor de canal tipo *p* (Q_{P1}) y otro de canal tipo *n* (Q_{N1}). Los niveles lógicos para CMOS son esencialmente $+V_{DD}$ para 1 lógico y $0\,V$ para el 0 lógico. Consideremos primero el caso donde $A_1 = +V_{DD}$ (la entrada A_1 está en un nivel alto (‘1’)). En está situación, la compuerta de Q_{P1} (canales *p*) está en $0\,V$ en relación con la fuente de Q_{P1}. De este modo, Q_{P1} estará en el estado OFF con $R_{OFF} = 10^{10}\,\Omega$. La compuerta de Q_{N1} (canales *n*) estará en $+V_{DD}$ en relación con su fuente, es decir, transistor Q_{P1} se pone en estado de corte y el transistor Q_{N1} se activa. El resultado es un camino de baja impedancia de tierra a la salida y uno de alta impedancia de V_{DD} a la salida F.

A continuación, consideremos el caso donde $A_1 = 0\,V$ (la entrada A_1 está en nivel bajo (‘0’)). Q_{P1} tiene ahora su compuerta en un potencial negativo en relación con su fuente, en tanto que Q_{N1} tiene $V_{GS} = 0\,V$. De este modo, Q_{P1} estará encendida con

$R_{ON} = 1\,k\Omega$ y Q_{N1} apagada con $R_{OFF} = 10^{10}\,\Omega$, produciendo un F de aproximadamente $+V_{DD}$.

En resumen Q_{P1} se activa y el transistor Q_{N1} se pone en estado de corte. El resultado es un camino de baja impedancia de V_{DD} a la salida F y uno de alta impedancia de tierra a la salida.

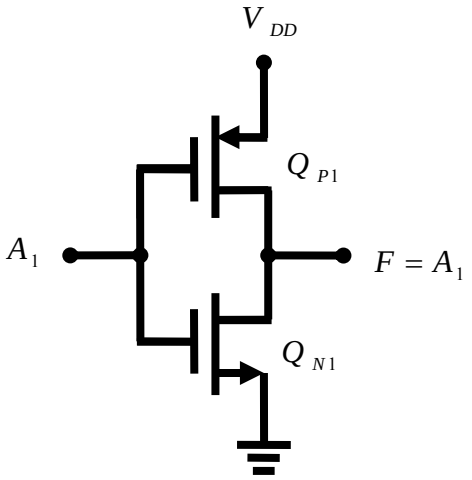


Figura II.11 Esquema del Inversor CMOS

A_1	F
'0'	'1'
'1'	'0'

Tabla II.1 Tabla de estados del INVERSOR CMOS

Como podemos observar, los transistores operan de forma complementaria. Cuando la tensión de entrada se encuentra en alto (1 lógico), el transistor N'MOS entra en estado de conducción y el transistor P'MOS entra en corte, haciendo que la salida quede en bajo (0 lógico). La situación inversa ocurre cuando la tensión se encuentra en bajo. Estos datos de operación se resumen en la (Tabla II.1), donde se muestra que el circuito actúa como un INVERSOR lógico.

COMPUERTA NAND CMOS

Se pueden construir otras funciones lógicas diferentes del INVERSOR básico. La Figura II.12 muestra una compuerta NAND formada por la adición de un MOSFET de canales *p* en paralelo y un MOSFET de canales *n* en serie al INVERSOR básico. Para

analizar este circuito conviene recodar que una entrada de $0V$ enciende el P-MOSPET y apaga el N-MOSFET correspondientes, y viceversa para una entrada $+V_{DD}$. Cuando ambas entradas (A_1 y B_1) están en nivel alto ($+V_{DD}$), hacen que los transistores Q_{P1} y Q_{P2} entren en corte y se encienden ambos N-MOSFET (transistores Q_{N1} y Q_{N2}), con lo cual ofrece una baja resistencia de la terminal de salida a tierra (la salida pasa a bajo (0) a través de Q_{N1} y Q_{N2}). En todas las otras condiciones de entrada, de cuando menos un P-MOSFET estará encendido en tanto que al menos un N-MOSFET estará apagado. Esto produce una salida ALTA (a través de Q_{P1} y Q_{P2}). Las entradas no usadas de una compuerta CMOS no se pueden dejar abiertas, porque la salida resulta ambigua.

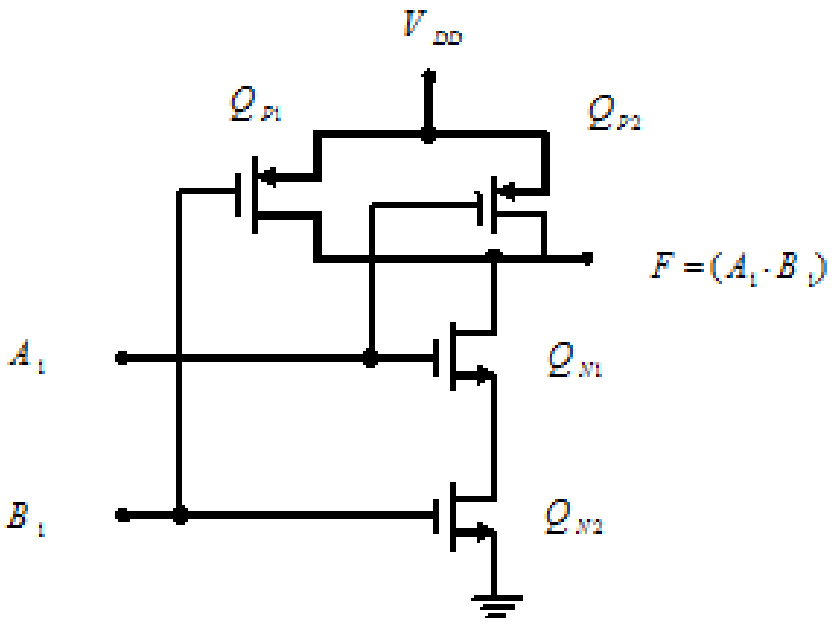


Figura II.12 Esquema de la compuerta NAND CMOS

Cuando sobra alguna entrada de una compuerta CMOS se debe conectar a otra entrada o a uno de los dos terminales de alimentación. Esto también es válido para circuitos secuénciales y demás circuitos CMOS, como por ejemplo, contadores, Flip-Flops, etc. Estos datos de operación se resumen en la (Tabla II.2), donde se muestra que el circuito actúa como una compuerta NAND CMOS.

A_1	B_1	F
‘0’	‘0’	‘1’
‘0’	‘1’	‘1’
‘1’	‘0’	‘1’
‘1’	‘1’	‘0’

Tabla II.2 Tabla de estados de la compuerta NAND CMOS

COMPUERTA NOR CMOS

Una compuerta NOR CMOS se forma agregando un P-MOSFET en serie y un N-MOSFET en paralelo al inversor básico (Figura II.13). Una vez más este circuito se puede analizar entendiendo que un estado BAJO en cualquier entrada enciende P-MOSFET (Q_{P1} y Q_{P2} entran a conducción) y apaga el N-MOSFET (Q_{N1} y Q_{N2} entran a corte) correspondiente. La salida pasa a alto (1) a través de Q_{P1} y Q_{P2}. Las entradas en un estado ALTO, hacen que los transistores Q_{P1} y Q_{P2} entren en corte y ambos transistores Q_{N1} y Q_{N2} en conducción (la salida pasa a bajo (0) a través de Q_{N1} y Q_{N2}). En las parejas de transistores ya sean de canal *n* ó de canal *p*, si cualquier entrada es baja, uno de los transistores entra a corte y otro a conducción. La salida pasa a bajo (0) acoplándose a través de transistores en conducción a tierra.

COMPUERTAS AND Y OR

Las compuertas AND y OR CMOS se pueden formar combinando compuertas NAND y NOR con inversores.

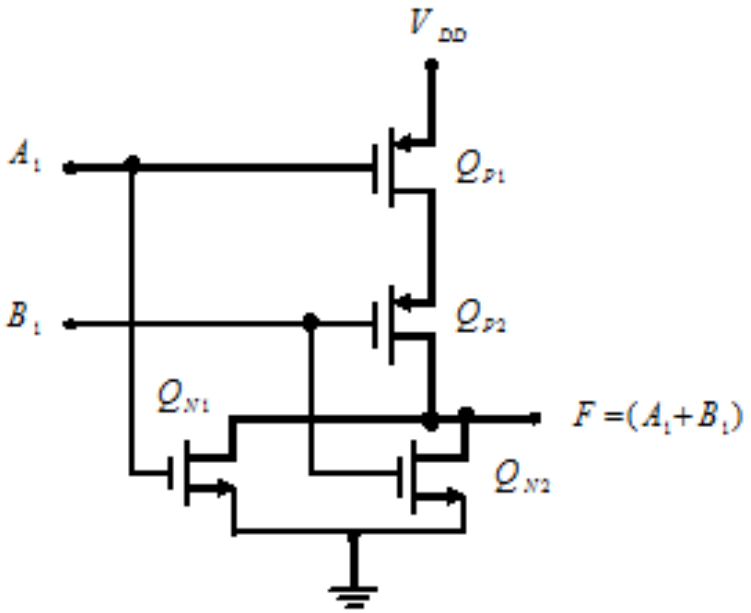


Figura II.13. Esquema de la compuerta NOR CMOS.

A_1	B_1	F
‘0’	‘0’	‘1’
‘0’	‘1’	‘0’
‘1’	‘0’	‘0’
‘1’	‘1’	‘0’

Tabla II.3 Tabla de estados de la compuerta NOR CMOS

II.3 Definición y características del Diseño del Producto

Dentro del diseño de un CMOS, se pueden encontrar una serie de retos que se deben superar a la hora de definir las características requeridas para el nuevo producto. En este apartado se presentaran algunos de estos retos.

Anteriormente se menciono como la Ley de Moore y la teoría del escalamiento como las principales causas que han dado la dirección a la industria de los semiconductores hacia nuevos retos en dimensiones y capacidades. Existen muchas publicaciones en las que se predice la terminación del escalamiento de los CMOS, también es cierto que existe una disminución en el escalamiento tradicional, esto es debido a que no se ha aplicado un escalamiento proporcional a las fuentes de alimentación (voltaje), lo cual no asegura que exista una densidad constante en la alimentación. Cuando el tamaño de los elementos que conforman un CMOS, se aproximan a niveles atómicos y límites mecánicos cuánticos, surgen dos importantes problemas que encabezan la lista para muchos en el área de la tecnología, estos dos problemas son:

- Incremento en la disipación del poder en estado de reposo debido a que:
  - Un aumento en la pérdida en el umbral de corriente.
  - Un crecimiento en la densidad de los dispositivos en un chip.
- Incremento en la variabilidad en las características de los dispositivos.

Si se da una mirada en la historia del desarrollo de los IC, puede uno asegurar que se encontraran soluciones a este tipo de problemas que surgen durante el proceso el diseño de un IC, a través del uso de nuevos materiales, procesos y estructuras de los dispositivos. Se puede predecir, que algunos problemas podrían ser resueltos mediante una cooperación mas cercana entre ingenieros con especialidad en el silicio, diseñadores de IC, diferentes tipos de sistemas y empaquetamiento.

Perdida de Potencia y Potencia Activa

En el mercado de dispositivos móviles de hoy en día, muchas aplicaciones requieren como característica principal tener una larga duración de la batería. Como por ejemplo: computadoras portátiles y teléfonos celulares. Estos últimos, requieren una larga duración de la batería en ambos modos, durante el modo de reserva y el de habla. En la figura II.14 se muestra la típica disipación de la potencia en estos productos.

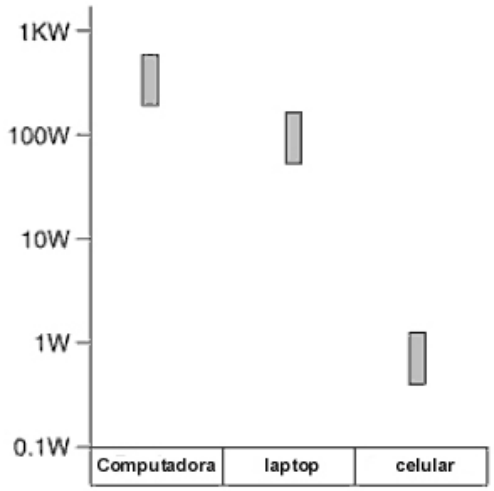


Figura II.14. Comparación en la disipación de poder en computadoras de escritorio, computadoras portátiles y teléfonos celulares.

Para los diseñadores de IC, una de las razones por la cual se ha vuelto más complicada la reducción en la pérdida de potencia, es el hecho de que existe en las últimas tecnologías de procesos nanométricos, un incremento significativo de pérdida en el umbral del transistor. Esto es sumado al incremento esperado en la disipación de potencia activa debido a una frecuencia de operación alta, y el incremento en el número de dispositivos por chip.

En la figura II.15 se muestra el resultado del cálculo en la densidad de potencia activa, causado por el intercambio, en función de la longitud de la compuerta. También es incluido en la grafica el incremento de potencia pasiva, el mayor causante de esto es la corriente túnel en la compuerta.

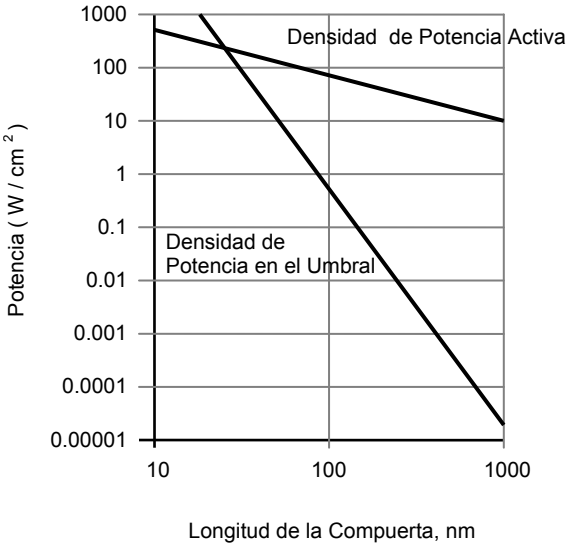


Figura II.15. Rápido Incremento de la potencia en el umbral de las últimas nanotecnologías.

Las dos líneas se intersectan en los 20 nm a temperatura ambiente. La corriente que fluye casi se duplica por cada incremento de 10°C en la unión de temperaturas. La potencia en el umbral igualara la potencia activa cuando la longitud de la compuerta sea de 50 nm. Debido a esto, la industria de los semiconductores siempre esta enfocada en buscar la solución para controlar la disipación de potencia. A continuación de presentan los mayores cambios que la industria ha tenido para el control de la potencia.

1970s-Cambio de dispositivos Bipolares a los tipos MOS.
1980s-Cambio a los dispositivos CMOS.
1990s-Reducción en escala del $V_{DD.}$
2000s-Eficiencia en la potencia, innovaciones en el escalamiento/diseño.
2010s-Nuevos materiales y estructuras en los dispositivos.

Variabilidad en el Proceso

El problema de la variabilidad en el proceso ha estado siempre presente desde el inicio de la industria de semiconductores. Lo que ha cambiado ha sido que las variaciones

son un componente de mucho mayor tamaño en relación a las dimensiones de los procesos en *nano-escala* de hoy en día. Mientras los ingenieros de proceso trabajan en entender y controlar la variabilidad en el proceso, en realidad la tecnología nanométrica debe de ser modelada e incorporada en las herramientas para el diseño. El problema es que la variabilidad puede afectar el desempeño y el rendimiento de los productos de vanguardia. La variabilidad se puede organizar en tres categorías:

- Global, la variación puede ocurrir entre un chip y otro chip, entre una oblea y otra o entre un lote y otro. Este tipo de variación es muy bien identificada por ingenieros con experiencia de fabricación y puede ser controlada comúnmente vía procesos y herramientas de control automatizado.
- Regional, las variaciones de este tipo hacen referencia a las variaciones que existen a través de la oblea o del chip. Si estas variaciones son sistemáticas, estas pueden ser comúnmente corregidas haciendo modificaciones en el proceso. Si estas variaciones ocurren aleatoriamente, pueden convertirse en un reto el tratar de corregirlas. La solución para este tipo de variación es el tener un sistema de monitoreo durante el procesamiento del chip, el cual pruebe y resuelva problemas por si mismo al momento de presentarse.
- Local, este tipo de variaciones pueden ser también sistemáticas o aleatoria.

Ejemplos de variaciones sistemáticas son:

- Superficies no uniformes al término del proceso de Pulido Químico-Mecánico CMP (*Chemical Mechanical Polishing*) debido a la densidad de patrón local.
- Variaciones en la reflectividad a todo lo largo del chip, causadas por la características de la figura. El método de corrección de óptica de proximidad OPC (*Optical Proximity Correction*), a partir del nodo de proceso de 180 nm, ha sido usado para disminuir este tipo de variaciones.

Ejemplos de variaciones locales aleatorias:

- Desde que en la litografía, la sub-longitud de onda ha sido empujada a una resolución de características más pequeñas que los 193 nm de longitud de onda, las impresiones contienen bordes irregulares. En la figura II.16 se muestra el

incremento en el intervalo entre la longitud de onda en la litografía y el tamaño mínimo de impresión. Los bordes irregulares en las líneas impresas son llamados irregularidades del borde de línea LER (*Line Edge Rougness*). Irregularidades a lo ancho de la línea son llamadas LWR (*Line Width Roughness*).

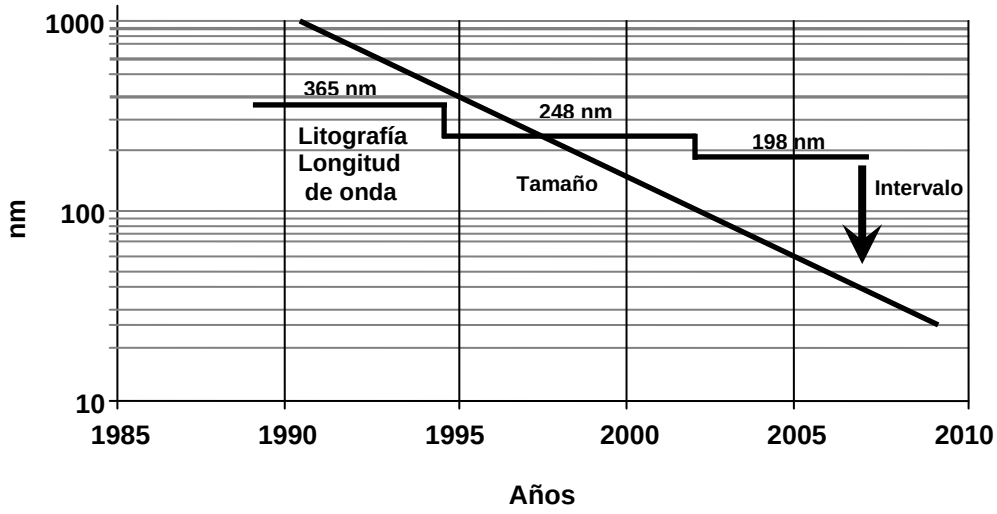


Figura II.16. Incremento en el intervalo de la longitud de onda en la Litografía y el tamaño mínimo de impresión.

A continuación se presentan los mayores desafíos que presenta el diseño de un CMOS:

- Calculo del tiempo para su terminación, este cálculo abarca muchos aspectos en el proceso de diseño y esta contabilizado usando una estimación estática de retardos (STA) que es una parte fundamental de las herramientas de posicionamiento/ruta, pero también puede correr en forma autónoma de medir el tiempo al finalizar los procesos.

  Los componentes primarios incluyen:

  - Un conjunto de tiempos con límite para cubrir todos los aspectos importantes en el diseño, tiempos de respuesta de entrada y salida, retrasos, trayectorias de ciclos múltiples.
  - Para diseños que serán implementados en base a una jerarquía física, el estimado del tiempo del chip debe separarse y limitarse a niveles de bloque. Debido a que requieren a menudo múltiples iteraciones en ciertos bloques.
  - Durante la síntesis lógica, se debe de mantener un estimado adicional, a manera de facilitar la terminación de la síntesis física.
- Integridad de la señal
- Energía Activa
- Fugas
- Implementación de DFM
- Calculo del tiempo estadístico

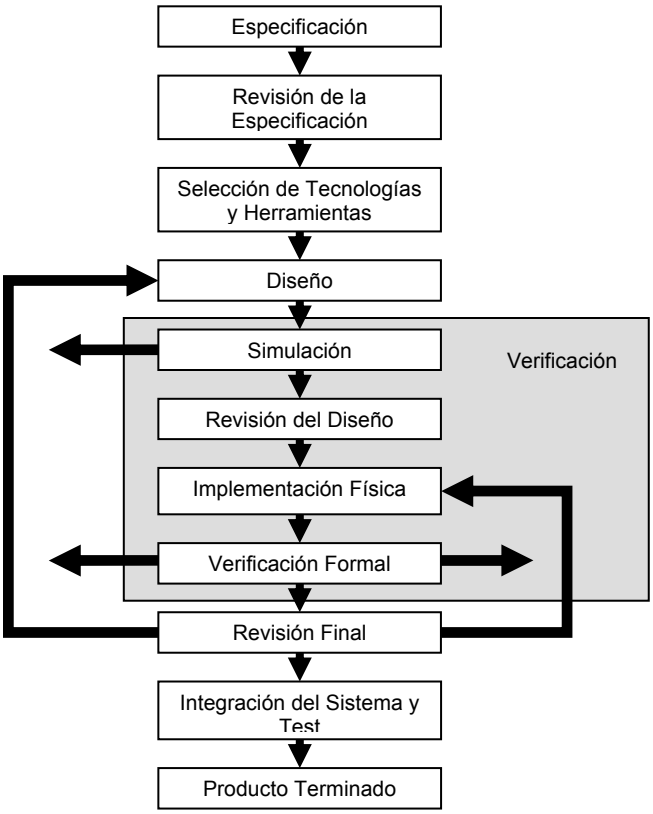


Figura II.17. Metodología Universal de Diseño.

Flujo general del diseño

En la siguiente figura II.17 se presenta en forma general la Metodología Universal de Diseño. A continuación se enlistan tres importantes fases en el desarrollo de un SoC ASIC, también se incluyen las funciones mas importantes dentro de cada una de ellas.

- **Fase del Concepto**
  - Aportación del área de Marketing.
  - Estudio y pruebas de Viabilidad.
  - Pruebas de desempeño, costo y funcionalidad.

- **Fase de Definición**
  - Comportamiento, arquitectura y definición de la estructura.
  - Intercambio en el nivel de modelado TLM (*Transaction Level modeling*) de un sistema y sus especificaciones. La función o el algoritmo comienza ser modelado y puede ser programado en los siguientes lenguajes: C, C++ o SystemC (VHDL[7] y Verilog[8]).
  - División del Hardware/Software. La parte del Harware del Chip se divide en tres categorías: bloque de IP (Propiedad Intelectual) de valor agregado, bloque con especificaciones especiales y Propiedad Intelectual de terceros.
  - Verificación de las funciones y comportamiento de Chip contra las especificaciones y los modelos de sistemas.
  - Definición de la Micro Arquitectura.
  - Codificación en el Nivel de Transferencia de Registros RTL (Register Transfer Level) usando lenguajes como VHDL o Verilog para reproducir el comportamiento.
  - Creación de verificación del entorno y generación de un banco de pruebas.

[7] VHDL es un lenguaje de descripción de hardware de propósito general que se puede utilizar o para describir y simular el funcionamiento de una amplia variedad de sistemas digitales, con un amplio rango de complejidad que va desde unas pocas compuertas lógicas hasta la interconexión de muchos circuitos complejos.

[8] Verilog es un lenguaje de descripción de hardware (Hardware Description Language, HDL) utilizado para describir sistemas digitales, tales como procesadores, memorias o un simple flip-flop, el lenguaje de descripción del hardware puede utilizarse para describir cualquier hardware (digital) a cualquier nivel.

- Verificación de los RTL de PI de terceras personas y bloques con especificaciones especiales en contra de sus especificaciones.
- Generación de código RTL sintetizado.
- Verificación de la equivalencia local del modelo a nivel de compuerta en contra del RTL.

- **Fase de Implementación**
  - De la síntesis del RTL a la conectividad de las compuertas lógicas.
  - Verificación formal del RTL en contra de la conectividad a nivel de compuerta.
  - Realizar un Análisis Estático del tiempo STA (*Static Timing Analysis*) de la conectividad a nivel de compuerta.
  - Hacer pruebas de retroceso a nivel de compuerta.
  - Implementación física.
  - Extracción de parásitos del Resistor-Capacitor y registro del tiempo al término vía STA.
  - Verificación formal de las conexiones en el diagrama de componentes.
  - Control en la verificación física y preparación de los datos.
  - Aprobar las revisiones y continuar con el Tape-Out.

En la siguiente figura II.18 se presenta una versión simplificada del flujo del diseño en la fase de implementación, la mayoría de los diseños digitales contienen algunos bloques Análogos y otros que contienen Propiedad Intelectual.

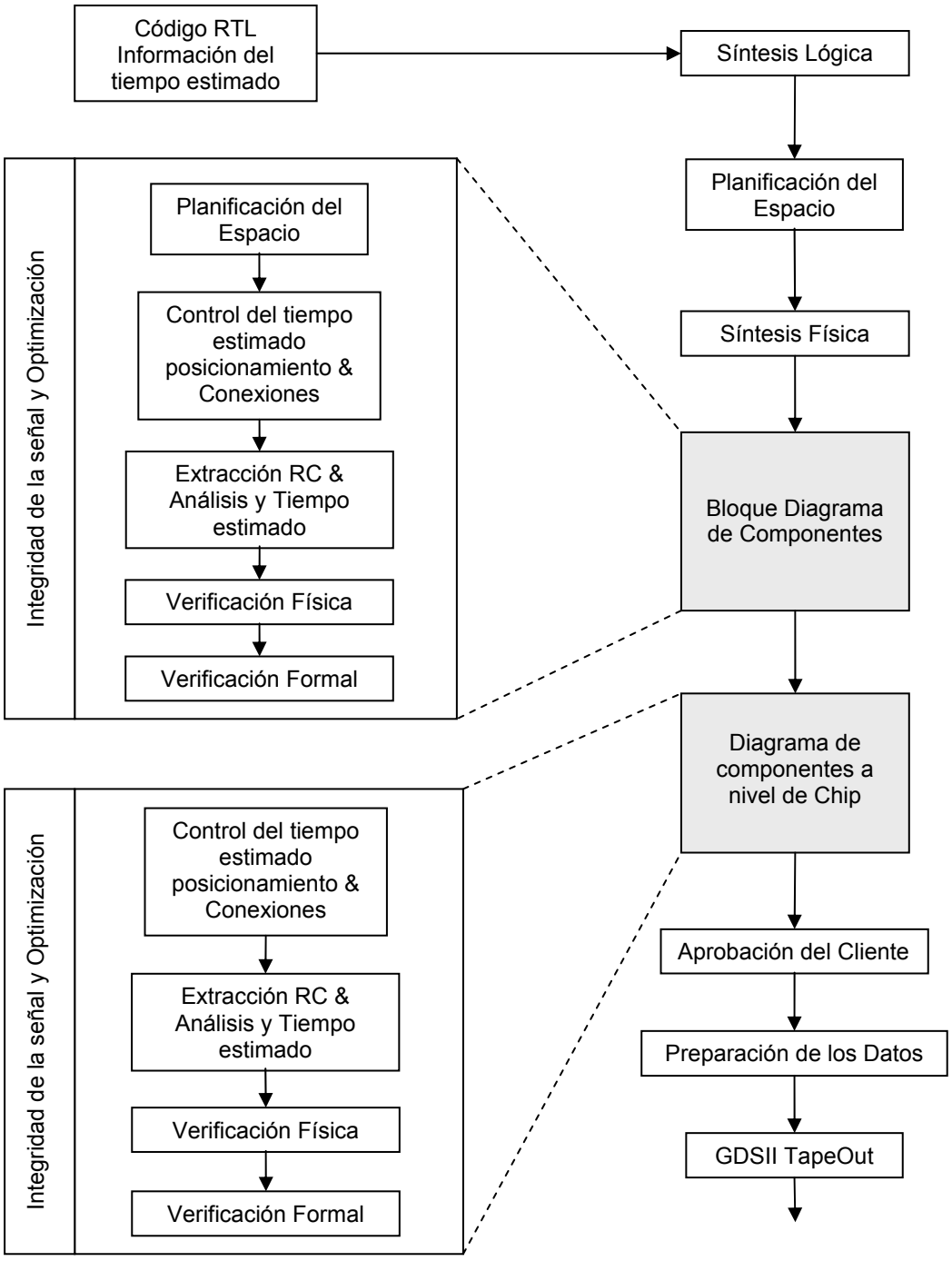


Figura II.18 Flujo del diseño simplificado

A continuación se presenta una breve descripción de las principales actividades dentro de cada uno de los procesos:

- **Síntesis Lógica e Introducción**
  - Convertir a conexiones de compuertas lógicas el comportamiento capturado en RTL.
  - Convertir flip-flops estándar a flip-flops de escaneo.
  - Dividir el diseño en bloques físicos manejables como “instancias” de hasta 500 K cada uno.
  - Aumentar un 10% al tiempo estimado para un diagrama de componentes inicial.
  - Ajustar los tiempos estimados de las siguiente forma:
    - De la entrada al registro (~25%);
    - Del registro a la salida (~25%);
    - De bloque a bloque (~%50).
  - Asignar bloques para manejar baja potencia, ejemplo: compuertas de reloj.
  - Verificar que sea equivalente la funcionalidad del RTL y las conexiones a nivel de compuerta.
  - Elaborar un STA inicial.
- **Planificación del espacio**
  - Elaborar de forma inicial a grandes rasgos una jerarquía de bloques, usando un estimado del tamaño de los bloques.
  - Manejo de la baja potencia:
    - Definir potencia y unidades de voltaje;
    - Calcular el consumo de potencia.
  - Planear la potencia y señal de reloj
  - Colocación Virtual de los elementos a un solo nivel
- **Síntesis Física**
  - Optimizar el diseño en base al estimado del tiempo y localización de los elementos a nivel global.
  - La aplicación de un estimado de tiempo incluye:
    - Entrada de llegada y dirección de celdas.
    - Relojes con introducción de retrasos temporales.

- Manejo de Potencia Baja
  - Incluir compuertas aislantes y/o desplazadores de nivel;
  - Preparar la estimación del tiempo de las librerías para múltiples voltajes
- Llevar acabo STA:
  - Validar el estimado del tiempo usando compuertas aislantes insertadas recientemente
  - Estimar retrasos a la hora del cableado de la colocación de los elementos a nivel global.
- Actualizar el tamaño y localización de los bloques.
- Actualizar la ubicación de la memoria, I/O y de la Propiedad Intelectual.
- Actualizar la distribución de la potencia y el reloj.

- **Diagrama de componentes del Bloque**
  - Aplicar a los bloques digitales un estimado de tiempo dirigido por su posicionamiento.
  - Los bloques pueden incluir como sub-bloques a Macros Análogas y Digitales (bloques de Propiedad Intelectual).
- **Diagrama de componentes del Chip**
  - Una vez que los bloques están terminados, son integrados a la parte alta junto con alguna otra Macro, bloques de Propiedad Intelectual y Memorias.
- **Preparación de los Datos siguiendo los requerimientos establecidos**
  - Metal fill
  - Documentación de las Estructuras
  - Limpieza final y verificación
  - Publicar formalmente la documentación establecida
- **GDSII TapeOut**
  - Subir la base de datos del diseño.
  - Elaborar una solicitud detallada con información requerida para su proceso, con el tamaño de chip y opciones especiales de proceso.
  - Descripción del nivel de máscara, reglas del diseño, modelos usados y cualquier otra instrucción especial.

II.4 Definición y características de Diseño y Layout

Para diseñar un chip se requiere de millones de pasos usando sistemas de cómputo asistidos por computadora (CAD). Millones de elementos del circuito son acomodados, simulados, y nuevamente acomodados en una computadora para lograr el rendimiento optimo del circuito antes de que la primera oblea sea enviada al proceso de fabricación.

Un equipo de desarrollo de procesos determina la mejor metodología de fabricación para el nuevo chip. Un típico proceso de fabricación contiene cerca de 300 pasos individuales durante el proceso. Sumado a esto, cada uno de los pasos es constantemente modificado y mejorado para lograr más pequeños y rápidos circuitos integrados. Diseñar el chip y desarrollar la tecnología necesaria para su proceso puede tomar desde unos pocos meses para modificaciones de un chip ya existente o años si se trata de un nuevo chip.

Fundamentos de la tecnología de fabricación de circuitos integrados CMOS

Los circuitos CMOS son circuitos analógicos, digitales o mixtos configurados a partir de transistores *PMOS* y *NMOS*. En el mercado actual de componentes electrónicos predomina de manera muy destacada la tecnología de circuitos integrados CMOS. Esta tecnología permite la fabricación de circuitos utilizando ambos tipos de transistores sobre un mismo cristal de silicio, y es la tecnología base de la actual microelectrónica o diseño VLSI.

Los transistores *PMOS* deben estar implantados sobre un substrato *N*, y los transistores *NMOS* sobre un substrato *P*. Para poder acomodar ambos tipos de transistores sobre un mismo cristal es preciso crear regiones de suficiente extensión que actúen como substratos, a estas regiones se les acostumbra a denominar pozos (*wells* o *tubs*). Dependiendo de las maneras en como se crean estas regiones dan lugar a tres tipos de tecnología CMOS. En la tecnología denominada de pozo N (N-well), el substrato es de tipo P por lo que acomoda directamente a los transistores *NMOS* y es preciso implantar una región N (pozo N) para acomodar a los transistores *PMOS*. Una tecnología dual de éstas es la tecnología de pozo *P* (P-well) en donde el substrato es

de tipo N y se implanta una región P. La tercera alternativa consiste en implantar los transistores sobre pozos N y P, especialmente creados (tecnología de pozos gemelos, *twin-well*). En esta tecnología se consigue un mayor control de las tensiones umbral de los transistores, tanto *PMOS* como *NMOS*, y reduce las caídas de tensión en los substratos, efecto que puede provocar problemas de *latch-up*.

Fundamento de la fabricación de circuitos integrados

Tal como hemos indicado anteriormente, la fabricación de circuitos integrados actual se basa en una tecnología *planar* que implementa todos los dispositivos del circuito sobre la superficie del cristal (chip, oblea). Estos dispositivos se crean mediante una secuencia de procesos físico-químicos realizados en ambientes libres de partículas contaminantes (cuartos limpios, *clean rooms*) que actúan selectivamente sobre la superficie siguiendo una técnica de máscaras (mask) creadas mediante un procedimiento de fotolitografía y grabado químico (*etching*). Las interconexiones entre estos dispositivos se realizan mediante líneas de metal (con múltiples niveles) que se colocan sobre la superficie mediante procesos de deposición, fotolitografía y grabado químico.

A cada una de las etapas de aplicación de los procesos físico-químicos se les denomina fases del proceso. La definición de las máscaras actuantes en cada una de las fases se realiza mediante lo que se denomina diseño de máscaras o diseño microelectrónica y constituyen el diseño de los circuitos electrónicos a nivel físico a partir del dibujo plano de las capas (*layers*, a partir de las cuales y mediante procedimientos de reducción fotográfica se confeccionan las máscaras del proceso) con ayuda de equipos informáticos para el soporte del diseño (CAD, *Computer Aided Design*).

Relación de máscaras físicas y capas de diseño en una tecnología CMOS *twin-well*

En este apartado procederemos a relacionar las máscaras físicas precisas para la fabricación de circuitos con una tecnología CMOS *twin-well* con un nivel de polisilicio y dos niveles de metalización (en tecnologías actuales el número de niveles de

metalización es superior a dos, aquí se consideran únicamente dos niveles por razones de simplificación de la exposición). Estas máscaras se utilizan para la aplicación selectiva de reactivos en las diversas fases del proceso. El número de pasos o subprocesos de fabricación es superior al número de máscaras, según se expondrá mas adelante. Dichas máscaras, siguiendo un orden de aplicación, son:

- Máscara de implantación del pozo N (*N-well Implant Mask*). Define las zonas sobre las que podrán implementarse transistores PMOS.
- Máscara de implantación del pozo P (*P-well Implant Mask*). Define las zonas sobre las que podrá implementarse transistores *NMOS*. Usualmente esta máscara es complementaria a la anterior, por lo que únicamente es preciso definir una de ellas en la fase de diseño de capas).
- Máscara de área activa (*Active Area Mask*). Define las zonas sobre las que podrán implantarse transistores. Fuera de esta zona aparecerán capas de óxido grueso.
- Máscara de polisilicio (*Polysilicon Mask*). Define las regiones sobre las que discurrirán líneas de polisilicio. En las regiones donde el polisilicio intersecciona con área activa, el polisilicio constituye la puerta de un transistor, depositada sobre óxido fino (óxido de la puerta del MOS).
- Máscara de implantación P+ (*P+ Implant Mask*). Define las regiones sobre las que se difunde o implanta una región tipo P (drenador y surtidor de los transistores PMOS y contactos de polarización del pozo P).
- Máscara de implantación N+ (*N+ Implant Mask*). Define las regiones sobre las que se difunde o implanta una región tipo N (drenador y surtidor de los transistores NMOS y contactos de polarización del pozo N). Esta máscara es complementaria a la máscara anterior.
- Máscara de contactos (*Contact Mask*). Define las perforaciones del óxido por las que el primer nivel de metal contacta a las líneas de polisilicio o la superficie del silicio.
- Máscara de Metal 1 (*Metal 1 Mask*). Definición de las interconexiones de Metal 1

- Máscara de Vías (*Vía Mask*). Define las perforaciones del óxido a través de las cuales el Metal 1 contacta el Metal 2.
- Máscaras de Metal 2 (*Metal 2 Mask*). Define las interconexiones de Metal 2.
- Máscara de pasivación (*Passivation Mask*). Corresponde a la definición de la capa de óxido de protección final del circuito. Esta capa, típicamente, cubre todo el circuito a excepción de los puntos de conexión final con el encapsulado (*pads*).

El diseño microelectrónico corresponde a la definición de estas máscaras. Ello se hace mediante el dibujo de rectángulos (*Manhattan rules*) que constituyen las capas (layers) del diseño y a partir de las cuales se pueden crear las máscaras físicas. En el diseño VLSI de circuitos en la tecnología anterior, estas capas podrían ser 9: Pasivación, Metal 2, Vía, Metal 1, Contactos, Implantación P+ (la Implantación N+ es complementaria), Polisilicio, área activa y Pozo N (el pozo P es complementario). Una característica de una tecnología es la resolución mínima de un dibujo de capas (layout). Esta dimensión, usualmente indicada como λ, tiene fuerte repercusión en las características eléctricas y temporales.

En la Figura II.19 se observa el dibujo (layout) de las capas (layers) que constituyen el diseño de una puerta lógica NAND de dos entradas. También se muestra el aspecto de la sección vertical del circuito físico siguiendo la línea discontinua que atraviesa todos los transistores.

En la figura II.20 se puede observar el layout de una puerta lógica NAND de dos entradas utilizado el programa Microwind, el cual es un programa para realizar layouts de circuitos integrados y simulaciones. Este programa fue principalmente diseñado para aprender a diseñar circuitos CMOS usando escalas de sub-micras.

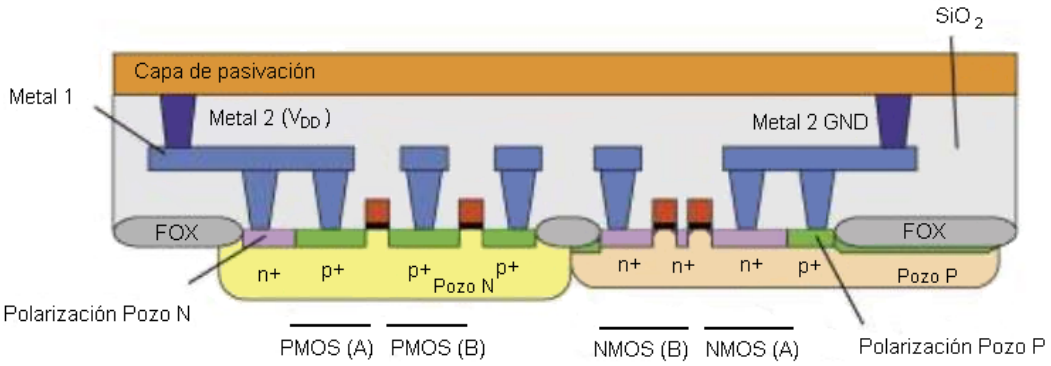


Figura II.19 Dibujo de capas (layout) de una compuerta NAND de 2 entradas.

La Figura II.21 se muestra la representación de la compuerta NAND en 3D, como parte de las funciones de simulación de pasos de procesos del programa Microwind.

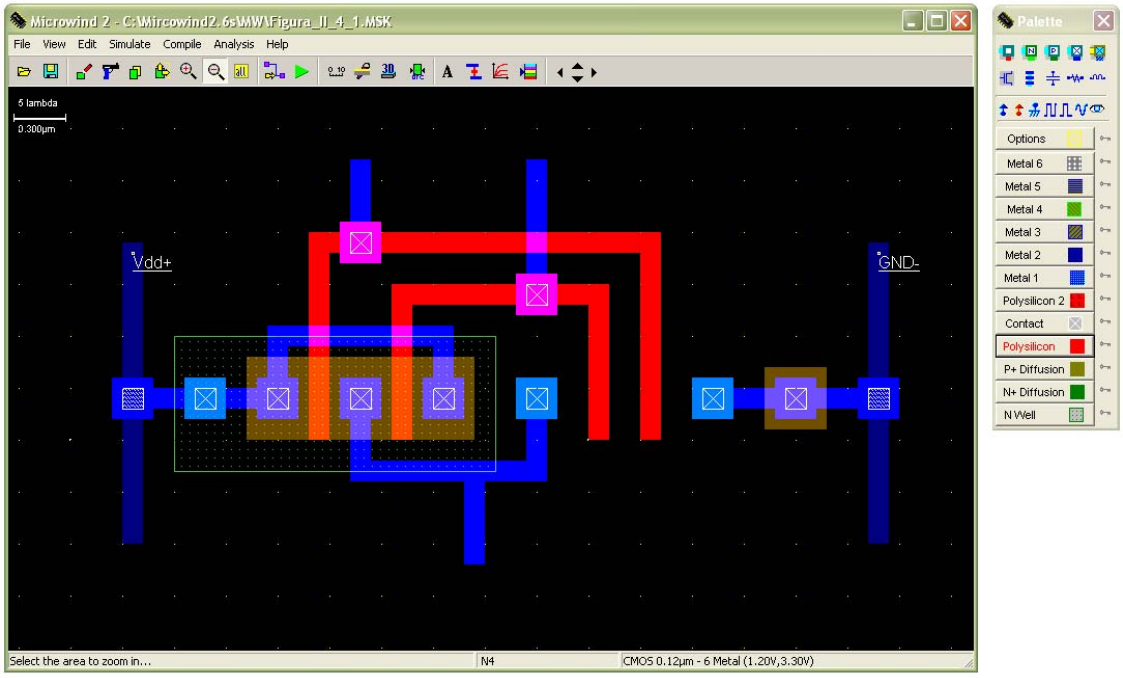


Figura II.20 Layout generado usando el programa Microwind

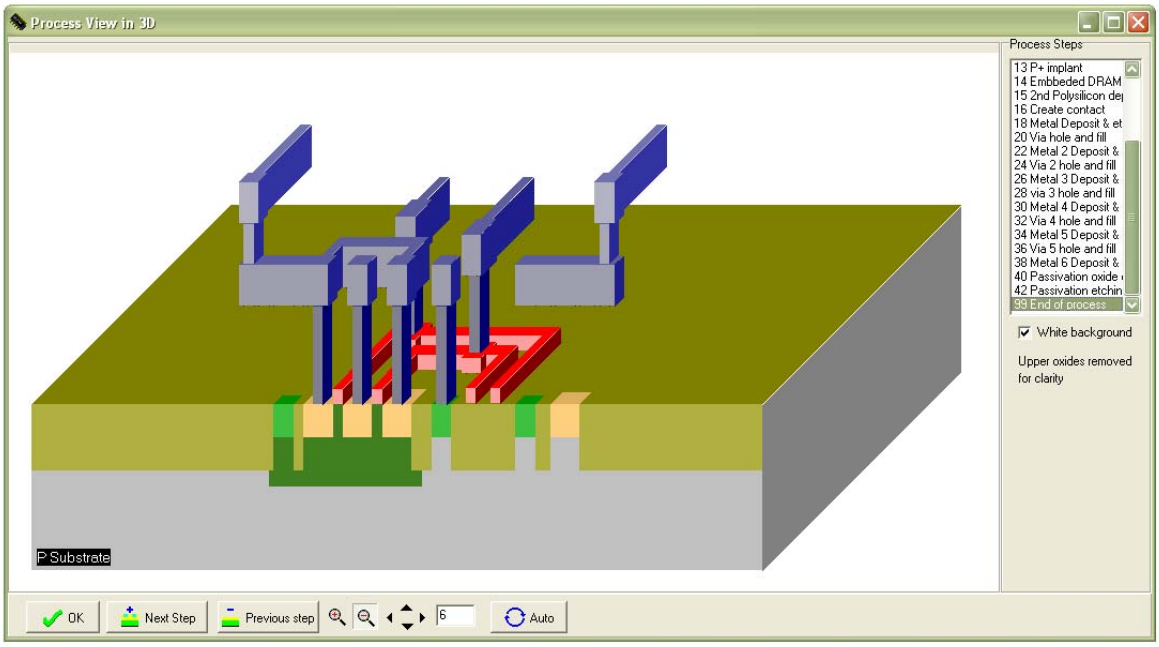


Figura II.21 Simulación de los procesos en 3D utilizando el programa Microwind

Fases del proceso de fabricación

El proceso CMOS que aquí se describe tiene por punto de partida un substrato (oblea) dopado ligeramente tipo P. La superficie se somete a un ambiente rico en oxígeno para crear una capa de óxido (SiO_2) protector. Se procede a un ataque químico mediante un proceso fotolitográfico que utiliza la máscara de Pozo N. Una vez desprotegida las regiones de pozo N, se procede a una implantación de alta energía de arsénico. El resultado de la creación del pozo N, así como la máscara física utilizada (realizada sobre el layout). Figura II.22

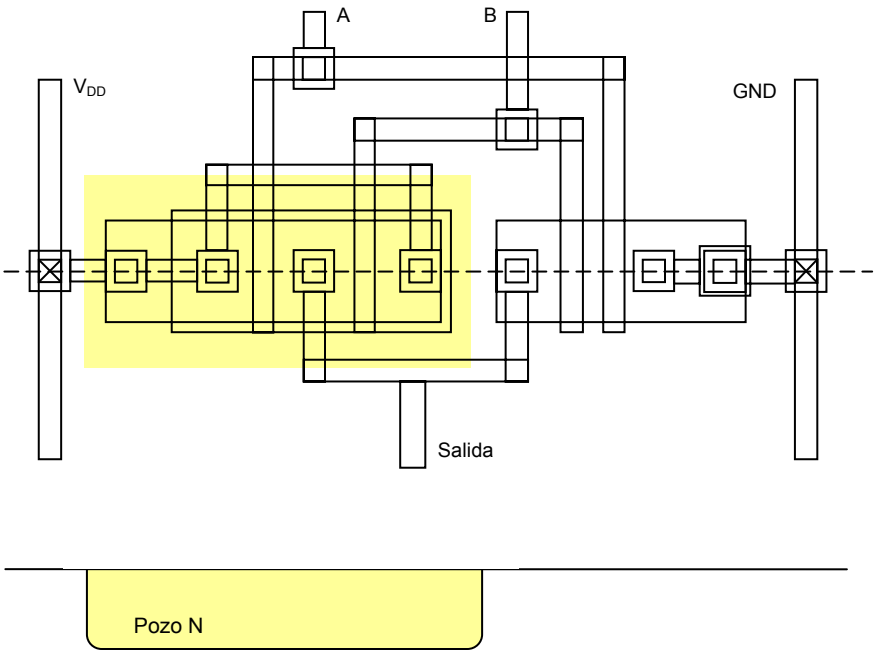


Figura II.22 Implantación de pozos N

Posteriormente se procede a la implantación de boro en la región correspondiente al pozo P, cuya máscara física es el complemento de la anterior. (Figura II.23).

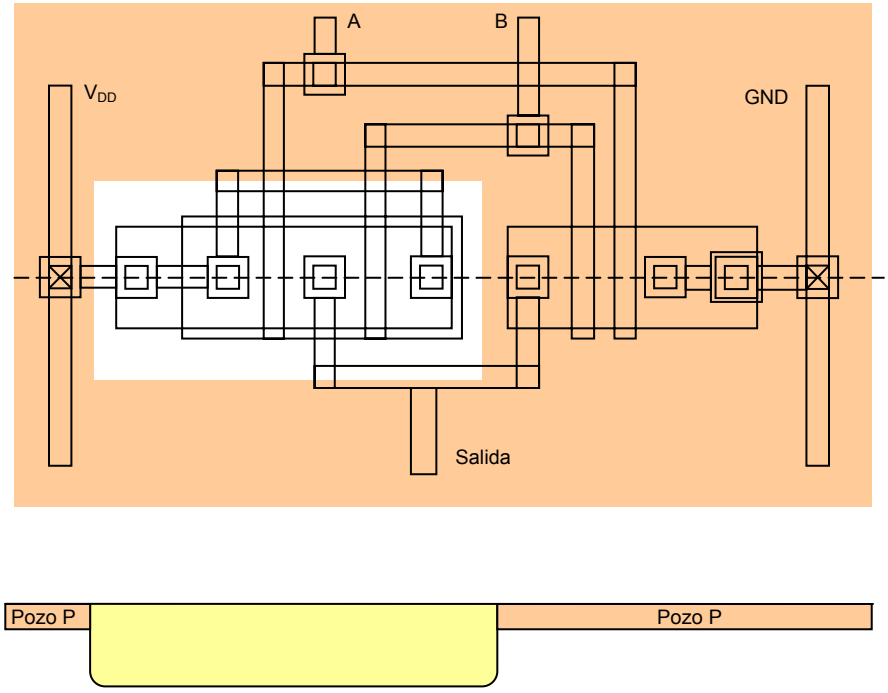


Figura II.23 Implantación de pozos P

A continuación de la creación de los pozos se procede a proteger las regiones activas mediante una máscara de nitruro de silicio (Si_3N_4) que se crea mediante un proceso fotolitográfico utilizando las máscaras de áreas activas.

Tras la creación de la máscara de nitruro se procede a una implantación con boro orientada a formar las regiones *channel-stop* destinadas a delimitar el canal en los transistores NMOS. A continuación se hace crecer térmicamente una capa de óxido grueso (*FOX, Field Oxide*) al mismo tiempo que se provoca la difusión profunda del pozo P (Figura II.24).

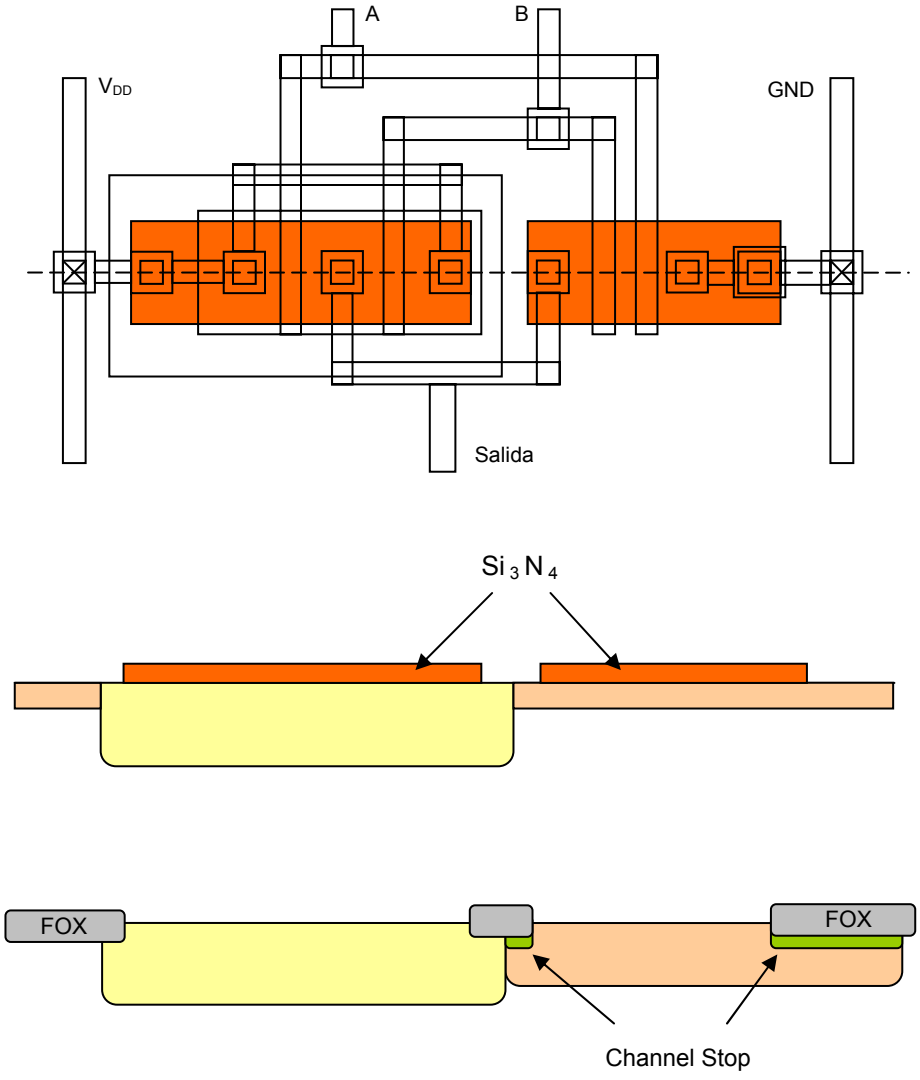
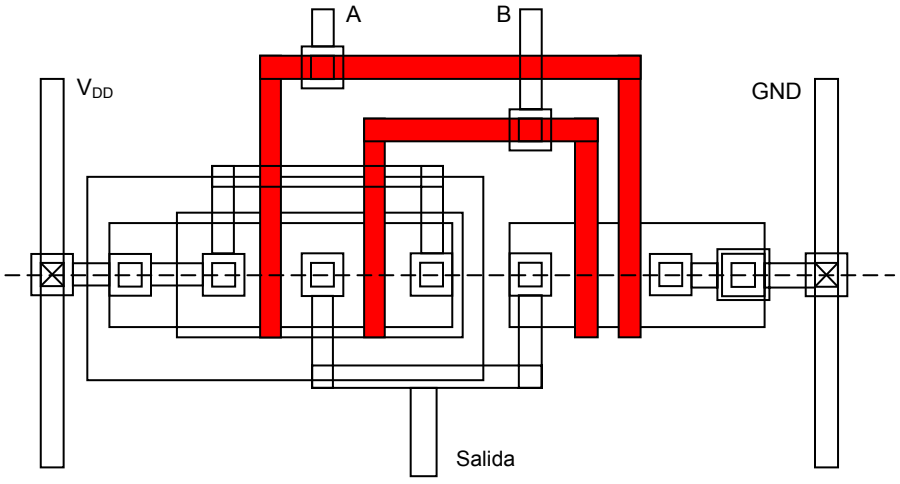


Figura II.24 Creación de máscara de Si_3N_4 correspondiente a las áreas activas. Crecimiento de las regiones óxido grueso (FOX) y de los *channel-stop*

En la siguiente fase se hace crecer una capa de óxido fino sobre las áreas activas. Se deposita una capa global de polisilicio (mediante *CVD*, *chemical vapor deposition*) y se ataca mediante procedimientos fotolitográficos para obtener la información de la capa de polisilicio. Se elimina el óxido fino que no queda cubierto por polisilicio. Esta estructura, polisilicio sobre óxido fino, constituye la puerta de los transistores (tanto PMOS como NMOS). Figura II.25.

A continuación se implantan o difunden las regiones P+ y N+ con sus respectivas máscaras (Figura II.26 y II.27). La puerta polisilicio actúa como máscara física, no permitiendo la implantación a través de ella, logrando así un efecto de alineación.
Una vez formados todos los transistores se cubre toda la oblea con una capa gruesa de óxido depositado (*LTO*, *low temperature oxide*). Se procede a un ataque selectivo (máscara de contactos) para obtener perforaciones controladas que permitan establecer los contactos (Figura II.28).

Posteriormente se cubre la oblea con una capa de aluminio y se ataca mediante un proceso fotolitográfico (máscara de Metal 1) para obtener el trazado del primer nivel de metalización (Figura II.29). Se procede paralelamente para la realización de las vías y la segunda metalización (Figura II.30 y II.31).



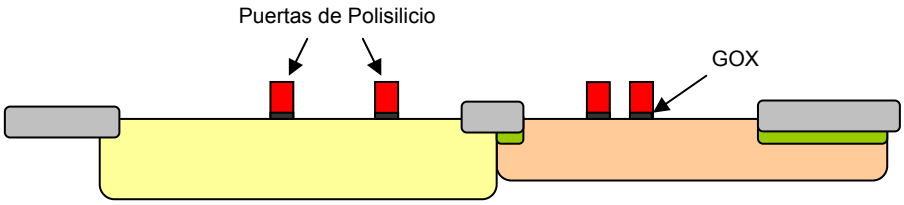


Figura II.25 Formación de las puertas de polisilicio

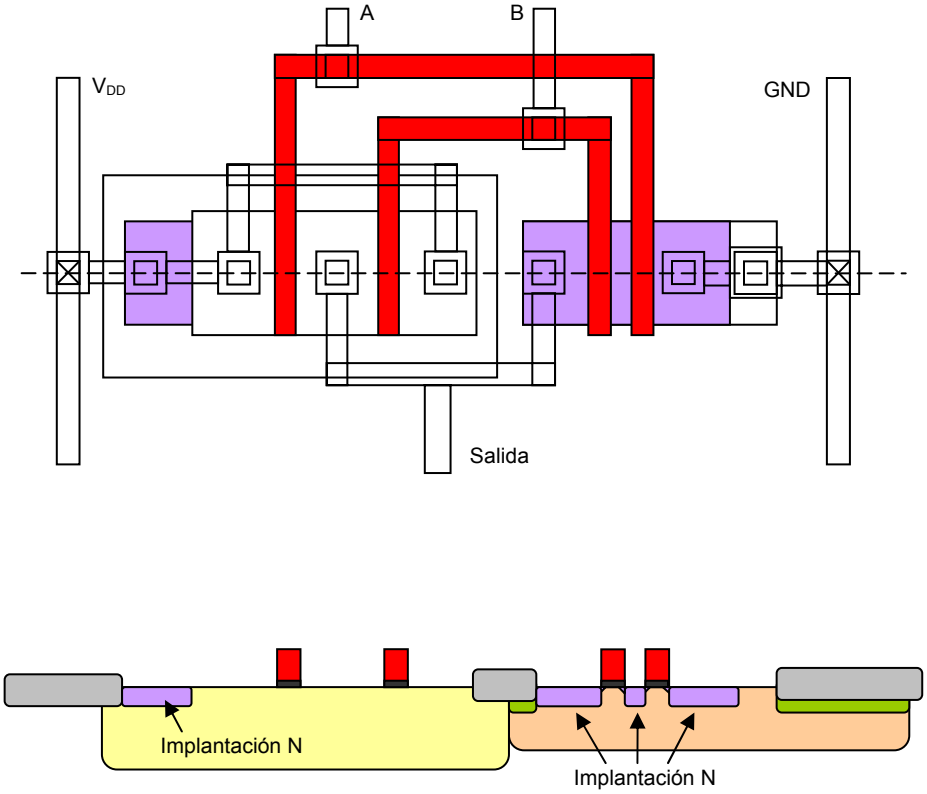


Figura II.26 Implantaciones *N+*

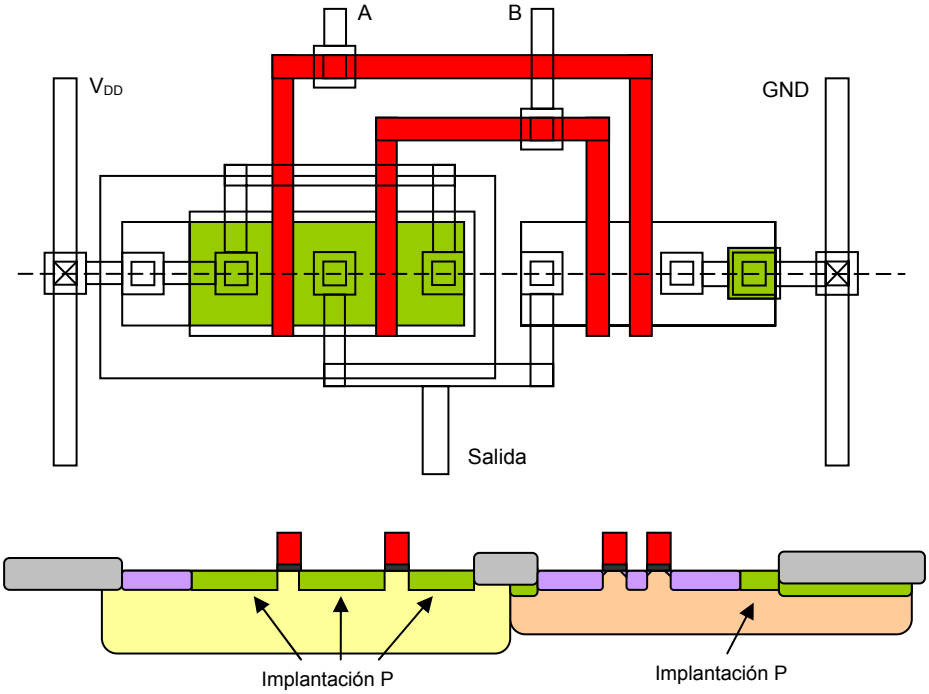


Figura II.27 Implantaciones *P+*

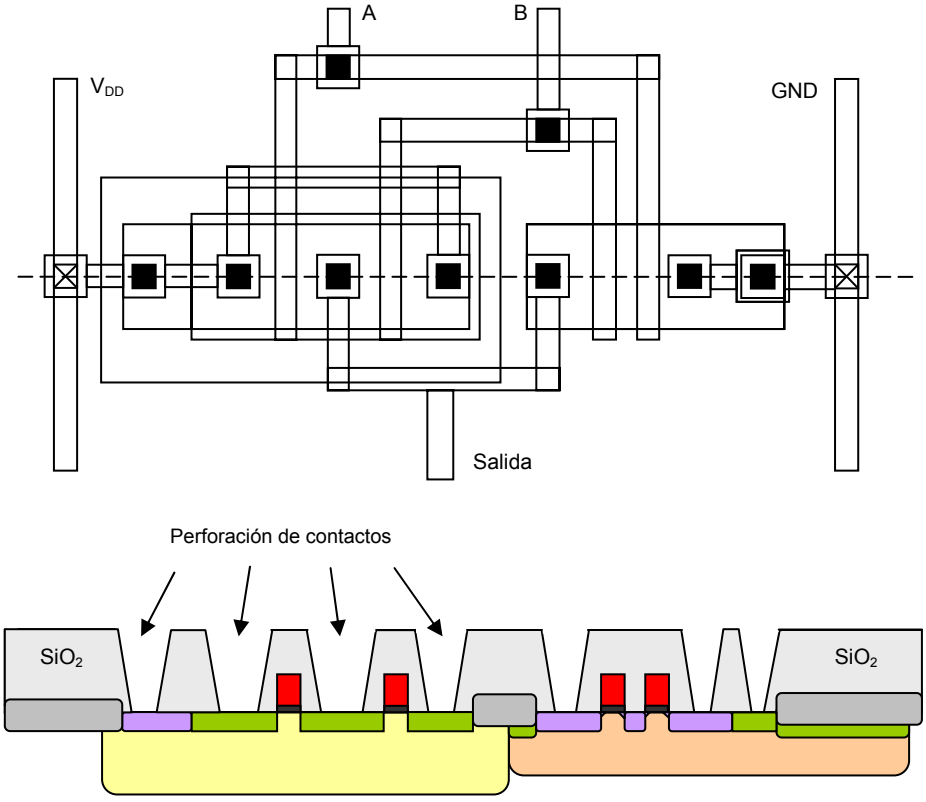


Figura II.28 Perforaciones en el óxido para establecer contactos

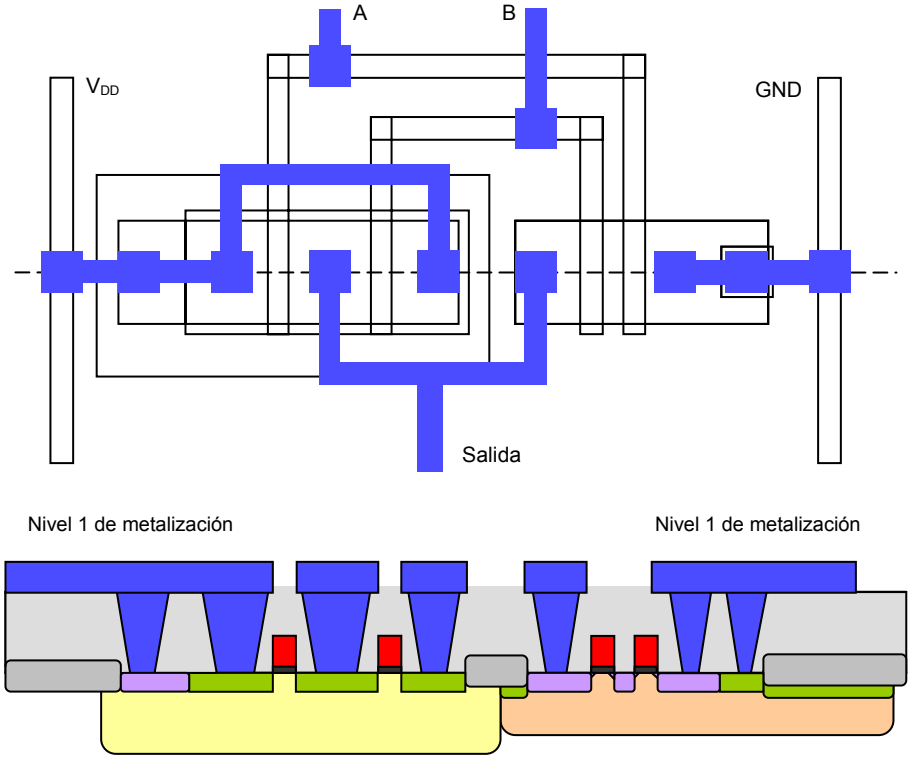


Figura II.29 Primer nivel de metalización

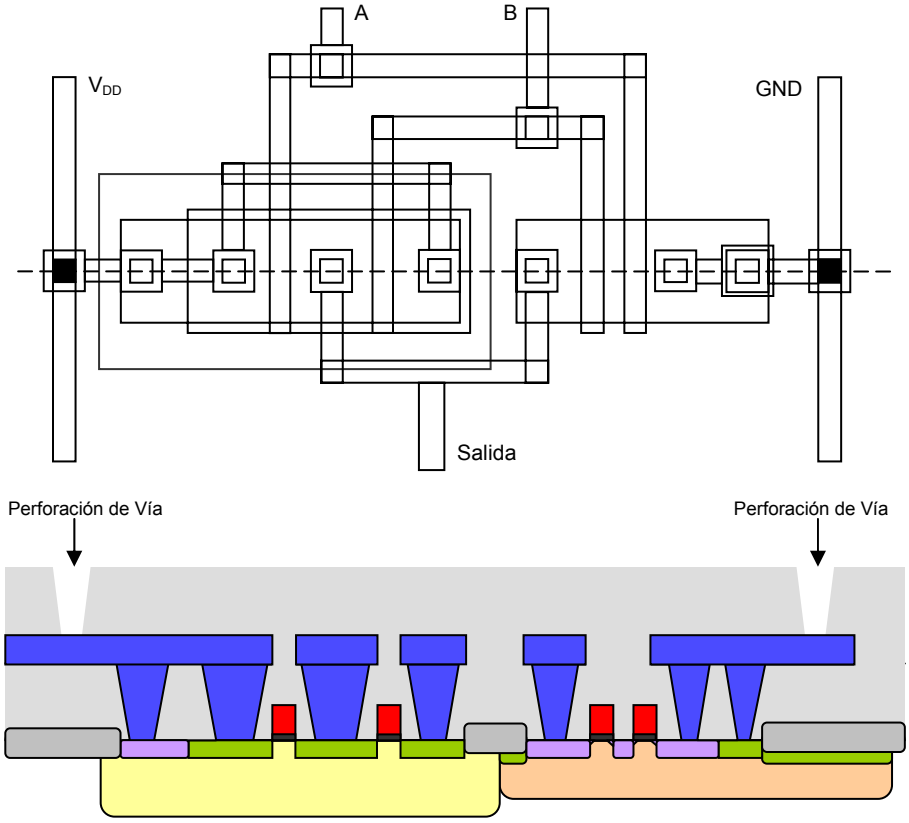


Figura II.30 Perforaciones de Vía

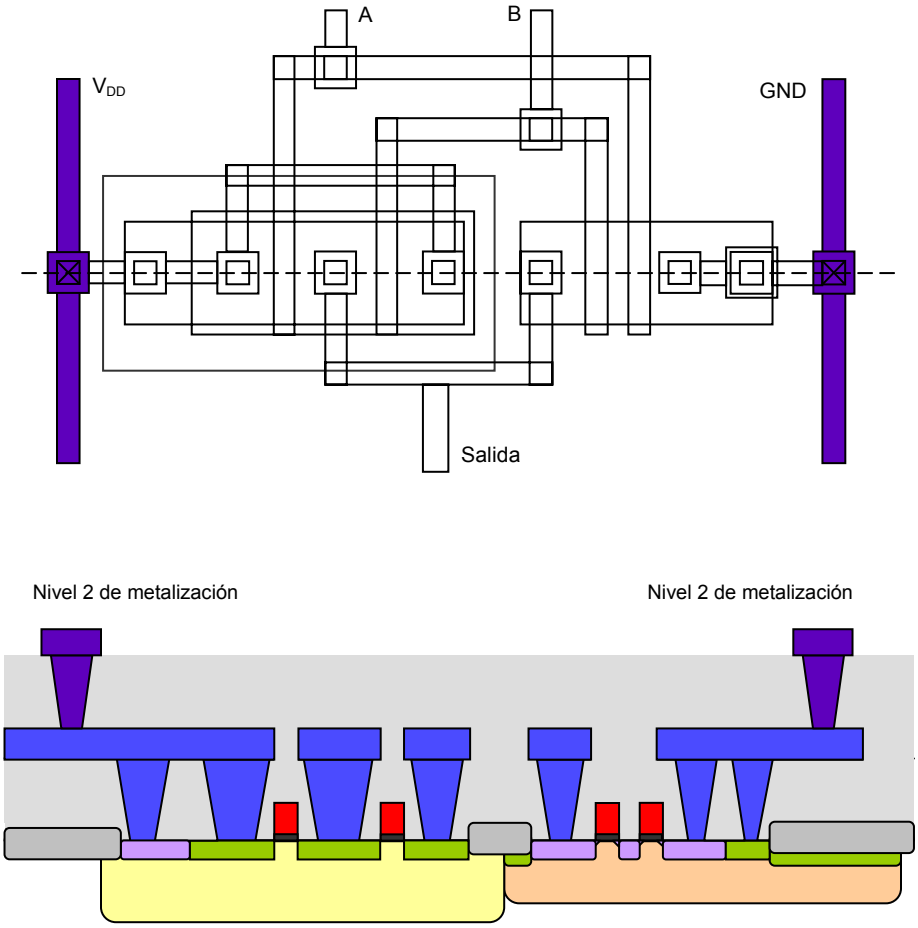


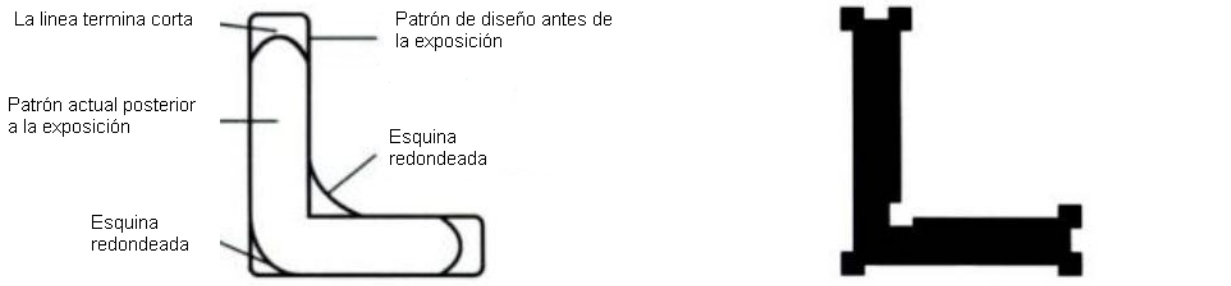
Figura II.31 Segundo Nivel de metalización

Finalmente se aplica la protección de pasivazo (máscara de pasivación) que corresponde a la capa de óxido superior, como se muestra en la Figura II.19.

II.5 Tape-Out (OPC, FILL, Preparación de Datos y Fractura)

OPC

La base teórica de la litografía de proyección óptica es la óptica difractiva, u óptica de difracción limitada. La difracción de imágenes puede causar perdidas de componentes durante la transmisión en alta frecuencia de una onda óptica y también durante la transformación en el sistema de lentes. La consecuencia de perdida en alta frecuencia es tener una imagen borrosa. Mirando de cerca individualmente a cada patrón, la perdida de nitidez en la imagen es reflejada por el redondeo de esquinas o de ángulos en un patrón cuadrado o de líneas que terminan recortadas. Esto es el llamado efecto de óptica de proximidad y se ilustra en la figura II.32. El efecto de óptica de proximidad usualmente no importa cuando el tamaño del patrón es grande.



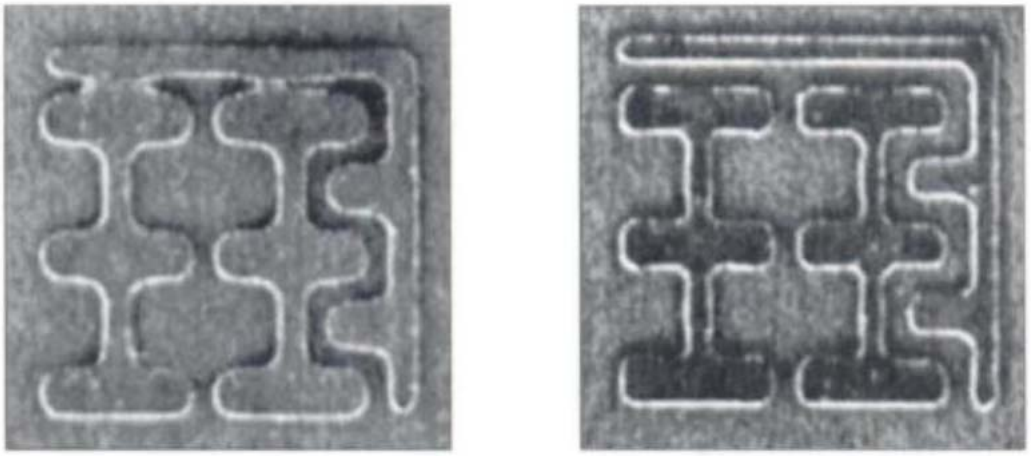
Efecto óptico de proximidad Corrección óptica de proximidad en el diseño

Figura II.32 Efecto óptico de proximidad y la corrección en el esquema.

Hoy en día en la industria de los IC, los diseños de patrones son muy pequeños con un incremento rápido en la densidad de transistores, los tamaños de los patrones pueden ser comparables a las variaciones del patrón causados por el efecto de óptica de proximidad, esto ha llamado la atención en los años recientes. Los efectos de la óptica de proximidad deben ser corregidos, en orden de mantener la fidelidad en los diseños de los IC.

El método básico del la corrección de óptica de Proximidad OPC (*Optical Proximity Correction*) que es distorsionar deliberadamente el patrón original para obtener un balance en la intensidad de la imagen. Por ejemplo, las regiones con una sobre exposición pueden ser reducidas eliminando partes del patrón y las regiones sin tanta exposición pueden ser aumentadas agregando algunas características extras. En la figura II.32 se muestra el patrón de diseño posterior al tratamiento de OPC.

Aunque los diseños con OPC luzcan diferentes de los diseños originales, todas esas figuras añadidas son de muy poca resolución, esto significa que no son resolubles en las imágenes ópticas, por lo tanto no habrá un cambio drástico en la apariencia del diseño original. Al contrario, las características agregadas harán que la imagen sea mas parecida al diseño original. En la figura II.33 se compara la imagen actual de la foto resina, de un diseño con y sin OPC. Las mejoras en la fidelidad de la imagen son muy notables con OPC.



Sin el proceso de OPC Con el proceso de OPC

Figura II.33 Comparación entre dos imágenes de foto resina, con y sin el proceso de OPC.

Para poder implementar OPC, el diagrama de componentes del diseño del IC debe de estar posprocesado (que no tenga modificaciones). Por lo general existen dos métodos: uno es basado en reglas, y el otro es basado en un modelo. El proceso de OPC basado

en reglas tiene la ventaja de tener correcciones rápidas debido a que los patrones son modificados de acuerdo a un conjunto de reglas. Sin embargo, la precisión es pobre debido al número limitado de reglas, las cuales no pueden cubrir todas las situaciones posibles, (gran variedad en la forma de los patrones); no se debe de pensar que estas reglas han sido creadas sin ningún motivo, ya que son derivadas de la experiencia y simulaciones en computadora de imágenes aéreas de formas típicas de patrones. En la figura II.34 se muestra un ejemplo del proceso de OPC aplicado a un diseño.

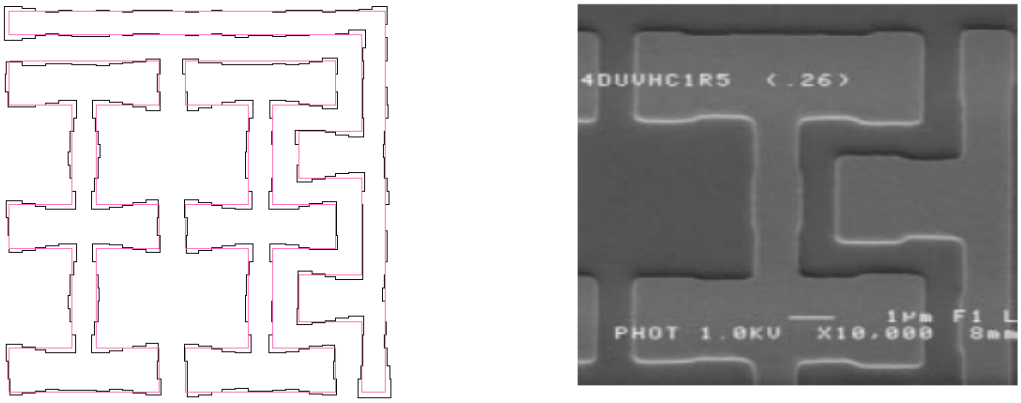


Figura II.34 Proceso de OPC aplicado a un diseño.

El método basado en un modelo comienza primero en calcular la intensidad de todos los patrones en la imagen, con una ecuación de óptica de imágenes para encontrar aquellos patrones que se encuentran expuestos y sobrexpuestos. El proceso OPC es aplicado para balancear la intensidad de la imagen, esto es, añadiendo o substrayendo tentativamente características sobre cada uno de los patrones. El proceso de análisis y corrección es repetido varias veces, y durante cada iteración, son modificadas las características, hasta lograr una distribución satisfactoria en la intensidad de la imagen.
El método basado en un modelo es obviamente mas preciso, pero consume una mayor cantidad de tiempo durante su proceso. La tendencia en los años recientes, es combinar los dos métodos, por ejemplo, las reglas son aplicadas primero a correcciones comunes, y posteriormente los modelos analíticos son usados para analizar el anterior resultado y llevar acabo, ajustes más finos a fin de alcanzar un resultado satisfactorio. En la figura II.35 se muestra en forma sencilla la diferencia de aplicar el proceso de OPC.

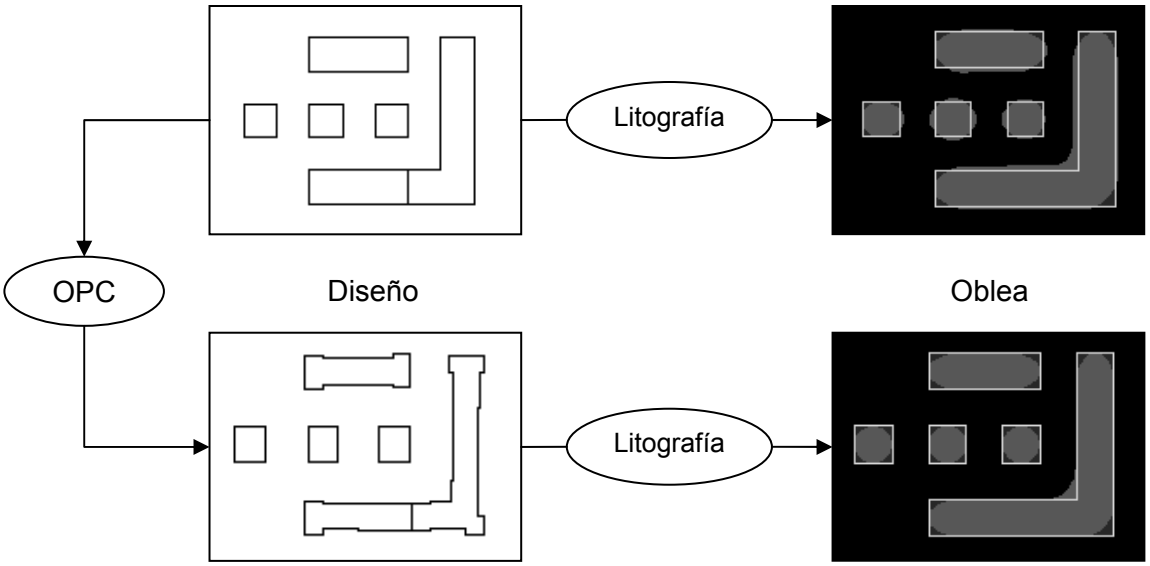


Figura II.35 Proceso de OPC

FILL

Los circuitos integrados están compuestos típicamente por estructuras formadas a base de capas, las cuales contienen diferentes materiales conductores, aisladores y otros. Estos materiales están estructurados en base a un proceso de fabricación en dimensión horizontal que transfiere los patrones definidos de los diseños físicos o layout. Más adelante, el proceso de diseño de los circuitos integrados generalmente emplea varias reglas para asegurar una densidad uniforme y la integridad de señal requerida.

Las multicapas de interconexión en un circuito integrado permiten a varios transistores ser conectados para completar un circuito. En las capas de metal de un circuito integrado, existen algunas áreas con una alta densidad de interconexiones y otras con muy poca. Ciertos pasos durante el proceso de fabricación, como el proceso de CMP usado para pulir capas intermedias de dieléctricos, tiene efectos variables en el dispositivo y sus propiedades de interconexión dependiendo esto de las características

propias del diseño. Para poder hacer estos efectos uniformes y predecibles, el diseño en si debe hacerse uniforme con respecto a ciertos parámetros de densidad. Algunos métodos tradicionales para lograr la uniformidad incluyen inserción (“filling”) o eliminación parcial de características en el diseño. La uniformidad en el proceso de CMP depende de la similitud en las características de la capas de interconexión por debajo de una capa de dieléctrico para evitar irregularidades. El modelado de Metal-fill es el proceso de rellenar largas áreas abiertas en cada capa de metal siguiendo un patrón, el cual puede estar atada o dejado en forma flotante, para compensar las variaciones de produzcan los patrones del diseño. La capa de Metal-fill atada esta conectada a la tierra o a la alimentación. La capa de Metal-fill flotante no se encuentra conectada a tierra ni a la alimentación. Existe un problema con las geometrías flotantes de metal-fill, los valores capacitivos no son conocidos y podrían asociarse con las líneas de señales de la capa de arriba o la de abajo.

Otro de los retos en el diseño de circuitos integrados implica fallas causadas por problemas en la integridad de la señal, como la caída del voltaje que es causada por la resistencia del cableado y la corriente consumida por la red de energía y la de tierra. Si la resistencia en el cable es muy grande o la corriente en la celda es mayor que la producida, puede ocurrir una caída de voltaje inesperada, causando que el voltaje que alimenta a la celda afectada sea menor al requerido, provocando que el tamaño de la compuerta sea mayor, creando retrasos en la señal, los cuales podrían causar una atenuación de temporización en recorrido de la señal así como también un retardo temporal. En el peor caso, la caída del voltaje podría ser suficientemente grande para provocar que el transistor falle en cambiar correctamente, causando una falla en el chip. En la mayoría de los flujos convencionales de diseño de circuitos integrados, el verificar la integridad de la señal es realizada como una actividad post-layout. Al intentar analizar y corregir estos problemas, a menudo resulta en: iteraciones costosas y consumo del tiempo del diseño, tiempo de entrega no cumplido, un funcionamiento reducido de producto e incluso con un mayor tamaño de chip con un rendimiento pobre durante su fabricación.

Fractura de Datos

Se le conoce como fractura al proceso de convertir datos en forma de polígonos a formatos (GDS2 o OASIS) que utilizan las maquinas para fabricar foto máscaras para litografía. Se ha comenzado a incrementar la importancia de optimizar el tiempo de la elaboración de la máscara y el control de las dimensiones críticas (CD).

Las foto-máscaras avanzadas o de alta complejidad, son grabadas utilizando VSB (*Variable Shaped Beam*) o sistema de litografía a base de emisión de electrones. Las maquinas para la impresión de foto máscaras, obtienen los patrones de diseño a partir de simples figuras como triángulos y rectángulos, para esto, se utilizo el formato posterior a la fractura, que básicamente, es descomponer la información en simples polígonos. Cuando se tienen modelos o figuras complejas, casi siempre hay una correspondencia de 1:1 entre figura y la exposición de la placa. El tiempo de escritura de las maquinas es importante debido a que se encuentra directamente relacionado al numero de figuras obtenidas durante la fractura y a su costo de fabricación. Teniendo en cuenta los altos costos, es necesario durante el ciclo de vida de la foto máscara, el aumento en el rendimiento de su uso. Es deseable que el software que se utiliza para la fractura genere la mínima cantidad de figuras necesaria para la representación del patrón o diseño a imprimir, para de esta forma reducir el tiempo de escritura.

Sin embargo, las figuras que están compuestas de dos o más disparos o exposiciones son un elemento a considerar dentro de la calidad de la fractura, estadísticamente su mayoría gustan de tener un tamaño mayor y una posición variable a diferencia de las figuras que solo están compuestas de un solo disparo o exposición. Esto es debido a la adición de errores causados por el tamaño y posición de las exposiciones individuales, por lo cual es importante controlar la cantidad total de exposiciones y la forma en como son ordenadas, para de esta manera obtener un mejor control en las dimensiones críticas.

PARTE II. MANUFACTURA

CAPÍTULO III
FABRICACIÓN DE SEMICONDUCTORES

III.1 Introducción

Aunque desde las primeras décadas del siglo XX se usaron los materiales semiconductores, fue el invento del transistor, en 1948, lo que preparó la escena de lo que llegaría a ser uno de los más grandes avances tecnológicos en toda la historia. La microelectrónica ha jugado un papel cada vez mayor en nuestras vidas, porque la tecnología de los circuitos integrados llegó a ser la base de las calculadoras, de los controles electrodomésticos, de sistemas de información, telecomunicaciones, controles de automóviles, teléfonos celulares, autómatas, viajes espaciales, proyectiles militares y computadoras personales.

Las ventajas principales de los circuitos integrados actuales son su tamaño y costo pequeños. Al ir avanzando la tecnología de fabricación, ha disminuido el tamaño de los dispositivos; en consecuencia, se pueden poner más componentes en un chip, pequeña pieza de material semiconductor sobre la que se fabrica el circuito. Además, el procesamiento en masa y la automatización de procesos han contribuido a reducir el costo de cada circuito terminado. Entre los componentes manufacturados están los transistores, diodos, resistores y capacitares.

Los tamaños de los chips que se producen hoy en día van de 3mm x 3mm hasta más de 50mm x 50mm. Antes, no se podían fabricar más de 100 componentes en un solo chip; sin embargo, la nueva tecnología permite alcanzar densidades del orden de 100 millones de componentes por chip. A esta magnitud de integrados se le ha llamado integración a muy grande escala (VLSI *very large scale integration*). Algunos de los circuitos integrados más avanzados pueden contener más de 100 millones de dispositivos.

Debido a la escala tan diminuta de los componentes microelectrónicos, toda su fabricación se debe hacer en un ambiente extremadamente limpio. Para este fin se usan recintos limpios, donde se permite cierta cantidad máxima de partículas mayores de 0.5 μm por pie cúbico. La mayor parte de los recintos limpios modernos son instalaciones de la clase 1 (una partícula por pie cúbico) hasta la clase 10 (diez partículas por pie cúbico). En comparación, la magnitud de la contaminación en los hospitales modernos es del orden de 10,000 partículas por pie cúbico.

En este capítulo se describen los procesos actuales para fabricar dispositivos microelectrónicos y circuitos integrados, siguiendo el esquema de la figura III.1, la cual se refiere a los pasos principales en la fabricación de una memoria Flash NAND. Introduciéndose así, primeramente las propiedades de las foto máscaras, y posteriormente será discutido cada uno de los principales pasos de su fabricación. Por último serán descritas las tendencias y expectativas de la industria microelectrónica.

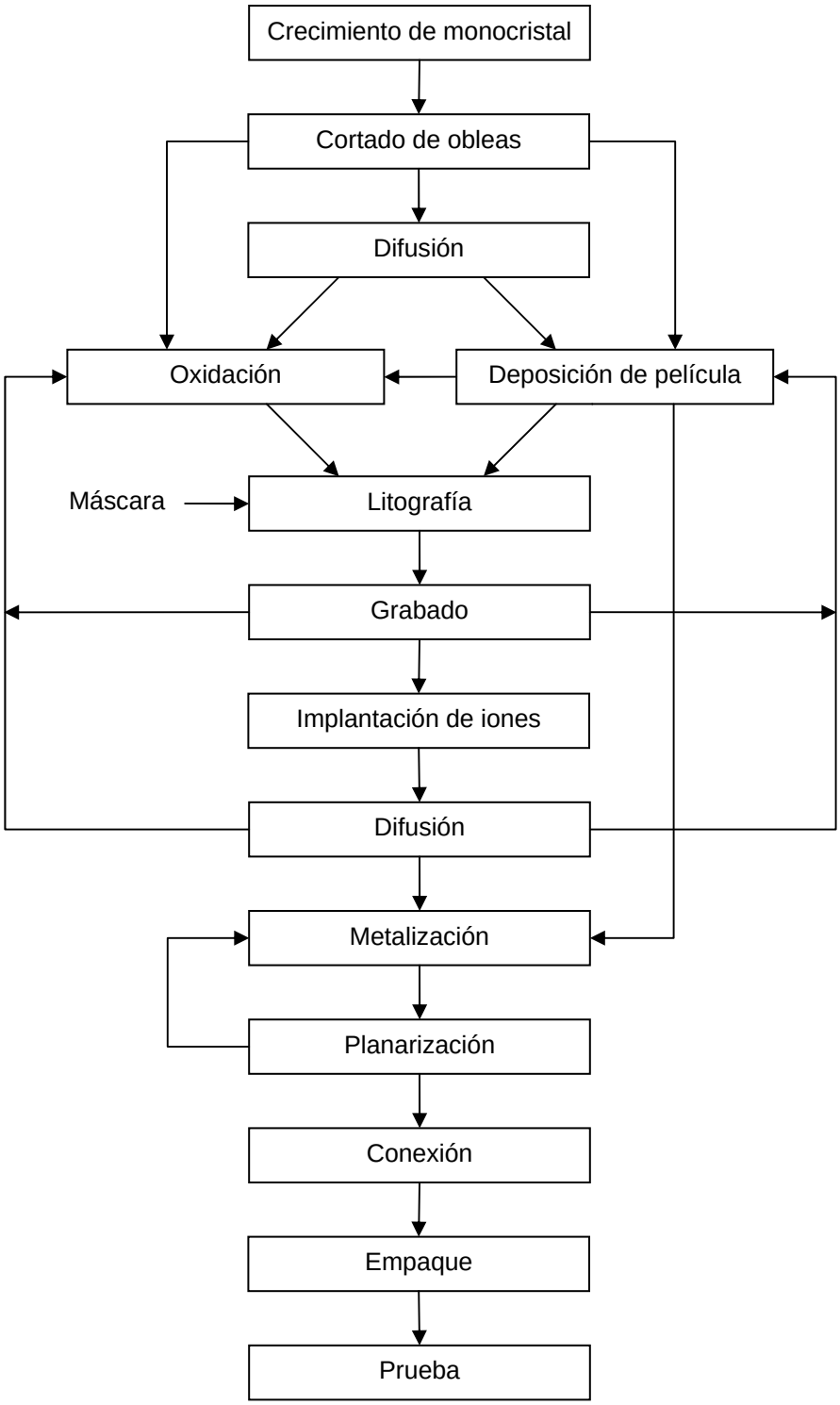


Figura III.1 Secuencia general de fabricación de circuitos integrados.

III.2 Definición y característica de la foto máscara

Definición de la foto máscara

Para fabricar un circuito integrado, se requiere realizar diferentes procesos tecnológicos en partes específicas de la superficie del semiconductor. La selección de estas partes o zonas se realiza mediante la fotolitografía con ayuda de las máscaras. El número de máscaras a utilizar puede ser mayor de 30, según la tecnología de fabricación. El proceso de diseño de la máscara en la actualidad es totalmente asistido por computadora (CAD *Computer Aided Design*), y el resultado final del mismo se presenta en forma de varios archivos que definen la zona de la exposición y proporcionan información necesaria para controlar automáticamente al primero de los equipos a utilizar durante la fabricación de las máscaras (archivo de instrucciones o *jobdeck*) Figura III.2.

Las figuras impresas en la foto máscara se crean a partir de archivos conteniendo los patrones de diseño, en este caso el tipo de archivo de datos al que nos referimos es MEBES (por sus siglas en ingles “Moving Electron Beam Exposure”). Los datos en los archivos se basan en las aportaciones específicas de los diseñadores para definir la lógica funcional, los cuales subsecuentemente son modificados por procesos particulares de los semiconductores y por los proveedores de los equipos a usar.

Las especificaciones en el circuito creadas por el diseñador, se utilizan como una entrada para los algoritmos que definen los detalles estructurales de la foto máscara. Para esto es necesario utilizar un programa de trascripción de tipo convencional, capaz de manipular y convertir los datos (fractura), como un ejemplo seria el Sistema de trascripción Asistido por computadora CATS (*Computer Aided Transcription System*) de la compañía Transcription Enterprises, Ltd., los datos del diseño tienen la característica de ser elaborados en varios niveles, siguiendo así una jerarquía, estos datos son sometidos a un algoritmo el cual define las estructuras de la foto máscara.

```
SLICE 1,17
*
OPTION PA,M,VA=10
RETICLE
*
****************************************
*  6 INCH MICRON TECH TITLE LOCATIONS  *
****************************************
ORIENT  A,MTITLE,TITLEROT=180,LOC=121000,147000,JUST=L
ORIENT  A,DTITLE,TITLEROT=180,LOC=121000,147000,JUST=L
ORIENT  A,TTITLE,TITLEROT=180,LOC=60000,147000,JUST=L
ORIENT  A,PTITLE,TITLEROT=180,LOC=60000,145000,JUST=L
ORIENT  A,NTITLE,TITLEROT=180,LOC=40000,147000,JUST=L
****************************************
*
CHIP L73X_FRAME_4X,
$          (001,L73X55O-A5-BS,AD=0.001,SF=1)
ROWS       76199.98600000/     76311.74600000
*
CHIP L73X_FRAME_LVL2_4X,
$          (050,L73X55O-A5-BY,AD=0.01,SF=1)
ROWS       76199.98600000/     76311.74600000
*
CHIP L73X_DIE_4X,
$          (001,L73X55B-O5-0F,AD=0.001,SF=1)
ROWS      107921.58400000/     54495.68600000
ROWS      107921.58400000/     97904.31400000
ROWS       44478.41600000/     54495.68600000
ROWS       44478.41600000/     97904.31400000
*
CHIP L73X_DIE_SUPPORT_4X,
$          (099,L73X55B-O5-0H,AD=0.001,SF=1)
ROWS      107921.58400000/     54495.68600000
ROWS      107921.58400000/     97904.31400000
ROWS       44478.41600000/     54495.68600000
ROWS       44478.41600000/     97904.31400000
*
CHIP L73X_DIE_LVL2_4X,
$          (050,L73X55B-O5-0V,AD=0.01,SF=1)
ROWS      107921.58400000/     54495.68600000
ROWS      107921.58400000/     97904.31400000
ROWS       44478.41600000/     54495.68600000
ROWS       44478.41600000/     97904.31400000
*
CHIP FID_4X,
$          (001,L73X55B-A5-BZ,AD=0.05,SF=1)
ROWS       76200.00000000/     76200.00000000
*
END
```

Figura III.2 Ejemplo de un archivo de instrucciones

Las máscaras para fotolitografía óptica, son generalmente un soporte de vidrio plano o de cuarzo fundido, de un alto grado de planitud, recubierta de una o varias capas para

grabar las figuras de la topología que posteriormente serán transferidas a la oblea. Estas capas pueden ser:

- Emulsión fotográfica; similar a una película fotográfica de alta resolución, que se sensibiliza y revela en la forma convencional. Permite alcanzar una resolución de hasta 5 μm; presentan problemas con la definición de los bordes.
- Metálica; la capa metálica puede ser generalmente de cromo, de óxido de hierro o de polisilicio, y está recubierta con una capa de foto resina sensible a la luz ultravioleta que se expone y revela en las regiones donde se desea eliminar para conformar las figuras de la topología, dejando abiertas ventanas. A través de estas ventanas se ataca la capa metálica, transfiriendo así las figuras de la foto resina a la capa metálica en cuestión. Estas capas son opacas a la luz ultravioleta y en general también al visible, con la excepción de las capas de óxido de hierro que son semitransparentes a la luz visible. Con cualquiera de ellas se pueden obtener resoluciones submicrométricas.

Para fabricar las máscaras utilizadas en la fotolitografía óptica hay que pasar por varias de las siguientes etapas según el proceso de fabricación de máscaras que se utilice:

- Fabricación de máscaras de primera reducción;
- Repetición y reducción de los motivos;
- Fabricación de máscaras patrones;
- Fabricación de máscaras de trabajo;

Métodos de fabricación de máscaras de primera y segunda reducción

Hay dos métodos básicos de preparación de las foto máscaras, el de reducción fotográfica y el que se basa en el uso de un equipo llamado generador de patrones.

Método de reducción fotográfica

Este método surgió en los años 70, y actualmente no tiene utilización en la industria. Una vez diseñado el circuito que se desea fabricar, la topología final se prepara en un formato con las coordenadas de cada figura que se desea transferir a la máscara, de

manera que pueda ser entendido por una maquina de transferencia de patrones. Esta maquina es semejante a una mesa de dibujo, que controla el desplazamiento mediante motores de paso de una cuchilla de corte, en las direcciones “x” y “y”, así como el giro a determinado ángulo. El equipo de transferencia realiza el corte sobre un acetato especial llamado *mylar*, el cual está recubierto de una película delgada de color rojo, opaca a la luz ultravioleta (UV). Como resultado final, la cuchilla define mediante el corte de la película roja superficial del *mylar*, las figuras geométricas de que consta la topología final del diseño. Al terminar los cortes, la película roja se despega del acetato en las regiones donde deben quedar ventanas. Las dimensiones de las figuras que se transfieren al *mylar* se aumentan generalmente unas 250 veces con respecto a las dimensiones finales que se desea que tengan en la máscara. Una vez preparado el *mylar*, se pasa al proceso denominado de primera reducción, que consiste en obtener una fotografía reducida, por ejemplo, 10 veces (10x) del *mylar*. A esta fotografía se le llama máscara de primera reducción. Para esto se requieren equipos fotográficos de mucha resolución y precisión, así como de una pantalla de dimensiones de más de un metro por lado, con una iluminación muy potente y uniforme en toda su área.

Segunda reducción y repetición

Una vez obtenida la máscara de primera reducción, se pasa a un nuevo proceso de fotografía, para lograr esta vez, una reducción que en total sea igual al aumento con que se definieron las figuras en el *mylar*. En el ejemplo puesto, esta reducción sería de 25x. La imagen reducida de la máscara de primera reducción se proyecta sobre una nueva placa fotográfica, que una vez terminada se llamará máscara patrón. Se repite tantas veces como quepa a lo largo de los ejes “x” y “y”, dejando un espacio fijo entre cada proyección que se denomina calle y que sirve para realizar el corte de los circuitos en la oblea durante el proceso de su independización antes del corte y soldadura. Este proceso de reducción fotográfica, exposición y repetición del motivo para generar la máscara patrón, se realiza en un equipo llamado foto repetidor o “step *and repeat*”.

Método de fabricación de retículos y máscara patrón a través de generador de patrones

Este método también consta de 2 etapas: la preparación de la máscara y la preparación de la máscara patrón. En este caso, el soporte magnético controla a un equipo llamado generador de patrones, el cual consta de una rendija con ancho y largo, así como de su posición angular, ajustables. La luz que atraviesa por la rendija, se proyecta y expone ciertas partes de la superficie de una placa fotográfica, que recibe el nombre de foto máscara o retículo, donde las figuras tienen dimensiones mayores, por ejemplo de 10x, 5x o 4x, respecto a las dimensiones finales que se desean en la máscara patrón. Cada figura se define mediante la exposición de rectángulos yuxtapuestos, según se muestra en la Fig. III.3.

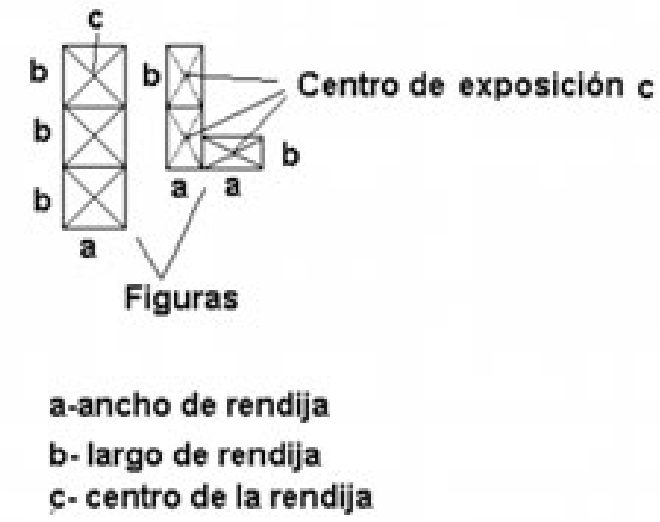


Figura III.3 Conformación de una figura por la exposición de un generador de patrones, mediante la yuxtaposición de rectángulos expuestos a través de una rendija de ancho a, largo b y centro situado en la posición c.

Para fabricar la foto máscara**,** la placa de vidrio cubierta por una emulsión fotográfica o por una capa metálica cubierta a su vez por foto resina, se va moviendo con respecto al haz de luz que va sensibilizando la foto resina que la recubre. Según la complejidad del

circuito, la placa de vidrio puede quedar dividida en millones de rectángulos que forman las figuras, por lo que el tiempo que se demora en posicionar el centro de la rendija, definir su ancho y largo, y efectuar la exposición, puede tomar en suma varias horas de trabajo. En el caso en que se utilicen placas cubiertas por una emulsión fotográfica, el tiempo de sensibilización es del orden de milésimas de segundo, mientras que en el caso de placas con recubrimiento metálico, cubierto a su vez de foto resina, se requieren décimas de segundo. Para una línea cuya dimensión final en la máscara se desea que sea de 1 μm, en la máscara de 10X se trazará una línea de 10 μm, y en el de 5X es de 5 μm. Estas máscaras se utilizan para fabricar las máscaras patrón usando la reducción fotográfica y repetición del motivo por medio del foto repetidor, que se describió en el párrafo anterior. Las dimensiones de las máscaras patrón fueron creciendo desde unos 50 mm por lado, hasta unos 150 mm por lado. Sin embargo, dada la complejidad que surgió al tener que realizar estas reducciones fotográficas en áreas tan grandes con un mínimo de distorsión, se regresó de nuevo a que la región útil de trabajo de las máscaras fuera de unos 75 mm de lado, donde se realiza la reducción fotográfica del retículo hasta alcanzar las dimensiones finales. Estas máscaras, con los motivos localizados en un área pequeña al centro, se utilizan para exponer obleas de mucho mayor diámetro, mediante las alineadoras de exposición y repetición, donde los motivos contenidos en la máscara se exponen y repiten, y a veces también se reducen una vez mas, directamente sobre la oblea (*step on the wafer*).

Máscaras de Trabajo

Las máscaras de trabajo, se producen mediante copias por contacto de las máscaras patrón. Esto es necesario ya que las máscaras de trabajo se dañan con el uso, por lo que tienen que ser sustituidas frecuentemente en el proceso productivo. La frecuencia de cambio e inspección depende del material de las máscaras y del tipo de alineadora utilizada. Cuando se hace la fotolitografía por contacto se deben cambiar cada 25 alineaciones las de emulsión y cada 50 a 100 las máscaras metálicas. Si se usan alineadoras de aproximación el número posible de veces a utilizar aumenta. En el caso de usar alineadoras de proyección, prácticamente las máscaras no se dañan durante el proceso de fotolitografía. En la fabricación de las máscaras patrón y las de trabajo, los

procesos se hacen en forma individual y no en grupos. Para el devastado del cromo, por ejemplo, se puede utilizar el ataque húmedo o el seco. El proceso de fabricación de un juego de máscaras patrón constituye una de las etapas más costosas en la realización de un circuito integrado.

Orientación de la foto máscara

En la figura III.5 se muestra el esquema de layout de una foto máscara de 6 pulgadas. Esta figura nos muestra una vista de todos los patrones posibles a transferir de la máscara. El equipo (o equipos) en el cual la foto máscara va a ser usada, determinan cuales patrones deben de ser incluidos en la máscara, mostrando la posición de cada uno de los patrones a transferir. El punto S es el centro del substrato. Esta definido como 76 mm ± 0.5 mm del lado izquierdo y del lado de abajo del substrato. El punto R es el centro del layout de la máscara, y debe de ser posicionado dentro de una ventana de 1.00 mm^2 alrededor del centro del substrato (S). La posición de cada patrón es definida por la posición del centro del patrón con referencia al centro del layout de la máscara. En la figura III.4 se muestra la relación entre el sistema de coordenadas de la oblea y de la foto máscara.

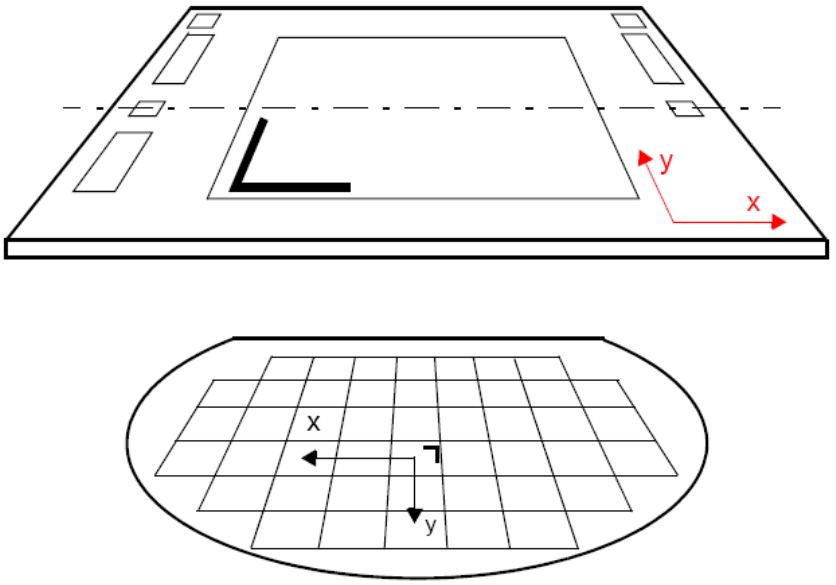
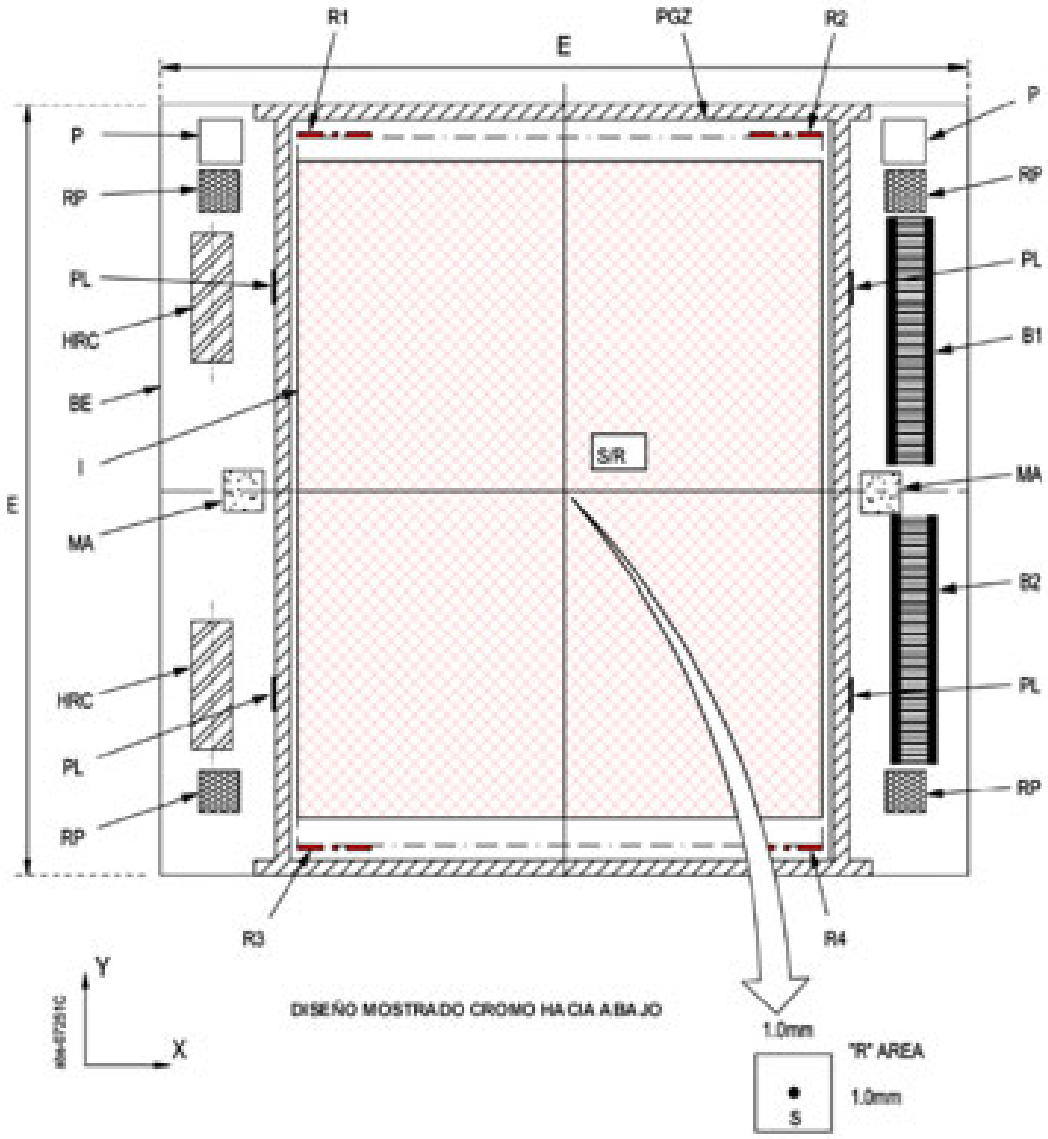


Figura III.4 Relación del sistema de coordenadas de la oblea y de la foto máscara



B1	Área del código de barras	BE	Borde biselado
B2	Área adicional para un código de barras de 24 caracteres (opcional)	PL	Línea de la posición del Pellicle
E	Largo del borde de la máscara (6” = 152.4 mm)	R	Centro del diseño de la máscara
HRC	Código legible	RP	Marca del área de separación
I	Campo de la imagen	S	Centro del substrato
MA	Marca de alineamiento de la máscara (solamente PAS 5500)	PGZ	Zona del pegamento del Pellicle
P	Marca de pre-alineamiento de la máscara	R1-R4	Marca TIS (*Tool Induced Shift*)

Figura III.5 Layout del esquema de una foto máscara de reducción 4X para sistema de paso y escaneo

Marcas de pre-alineamiento de la máscara

Las marcas de pre-alineamiento de la máscara (P) es un diseño con figura de estrella (Figura III.6) que se encuentra dos veces en la máscara. Estas marcas son utilizadas por el subsistema de manejo para pre-alinear la máscara sobre la mesa de exposición. Estas marcas deben de ser transparentes sobre un fondo oscuro (cromo). El borde de cromo de estas marcas, debe de ser un área de por lo menos de 8.5 mm x 8.5 mm centrada alrededor de cada marca de pre-alineamiento. La cruz en el centro puede ser usada para inspección con un equipo de medición óptica.

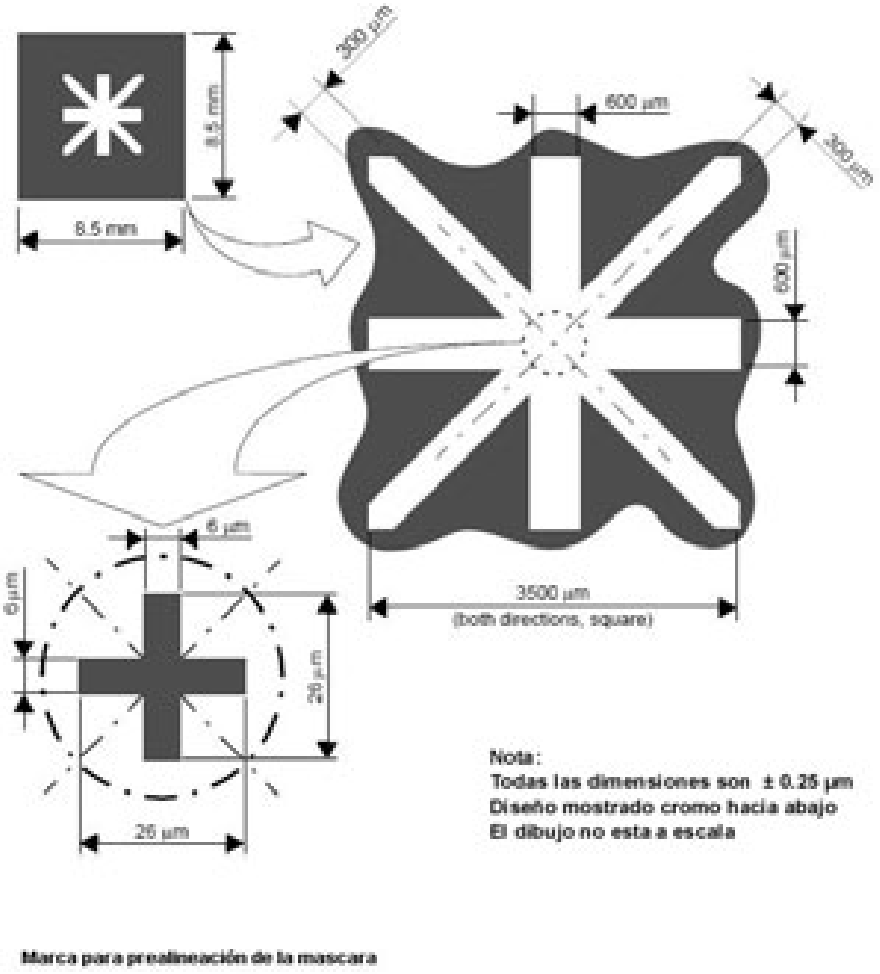


Figura III.6 Marcas de pre-alineamiento en la máscara

Debido a la importancia de la foto máscara dentro del proceso de manufactura, por lo general las empresas hacen el esfuerzo de fabricar sus foto máscaras en sus propias instalaciones, para de esta manera reducir los costos de fabricación. Cabe mencionar que dentro de la fabricación de semiconductores a nivel mundial, existen solo algunas compañías que pueden fabricar ciertos tipos de foto máscaras de alta tecnología o conocidas como HIGH-TECH, este tipo de foto máscaras son utilizas para la elaboración de ciertos niveles en la arquitectura del semiconductor, y que debido a la complejidad en su diseño es necesaria la elaboración de este tipo de placas. Debido a esto es necesario eliminar los errores y reducir el tiempo al mínimo del proceso de fractura y preparación de datos, para poder así, ser enviados al fabricante de foto máscaras lo mas pronto posible ya que en la fabrica (en donde se elaboran los semiconductores) las obleas se han empezado ha preparar para poder así tener de forma sincronizada la entrega de la foto máscara ya terminada y las obleas esperando ya en el proceso de foto litografía. (Tabla III.1).

Cautivas[1]	Comerciales
Micron Technology, Inc. Photronics, Inc. http://www.mpmask.com	Dai Nippon Printing Co Ltd. http://www.dnp.co.jp
Intel Mask Operations Intel Corporation, IMO http://www.intel.com	Toppan Printing Co Ltd http://www.toppan.co.jp
SAMSUNG Semiconductor http://www.samsung.com	Photronics, Inc. http://www.photronics.com
Hynix Semiconductor Inc. http://www.hynix.com	
AMTC Advanced Micro Devices, Inc. Qimonda AG Toppan Photomasks, Inc. http://www.amtc-dresden.com	
Taiwan Semiconductor Manufacturing Co. Ltd http://www.tsmc.com	

Tabla III.1. Fabricas de Foto Máscaras HIGH-TECH a nivel mundial

[1] El tipo de fabrica cautiva se refiere a que la compañía tiene el equipo necesario para elabora sus propias foto máscaras.

III.3 Procesos en la fabricación

El primer paso de la producción de un Circuito Integrado es la obtención de una oblea de material semiconductor con estructura cristalina. Los semiconductores más importantes para la fabricación tanto de dispositivos discretos como de circuitos integrados son, con diferencia, el Silicio (Si) y el Arseniuro de Galio (GaAs). Los procesos que se siguen para conseguir una oblea semiconductora a partir de la materia prima son los siguientes (figura III.7):

- Purificación del substrato mediante tratamiento químico
- Crecimiento en volumen del cristal
- Corte, limpiado y pulido de obleas.

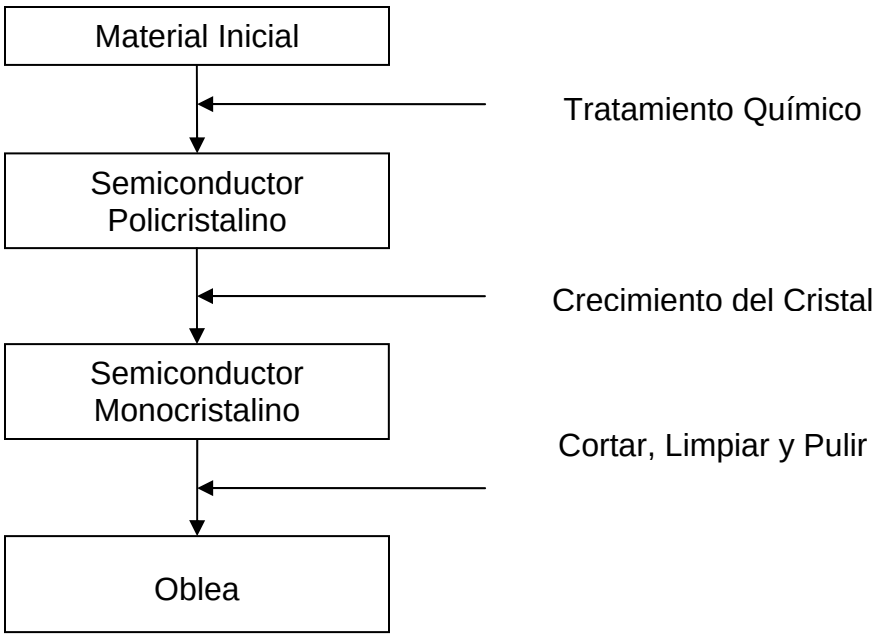


Figura III.7 Pasos para la fabricación de una oblea

III.3.1Preparación del substrato

Purificación del substrato (obtención de Si puro)

Para la obtención de Si puro se parte de la cuarcita (SiO_2) (forma relativamente pura de arena). Esta se coloca en un horno junto con varias formas de carbón (hulla, coque, astillas de madera), dando lugar a la reacción siguiente:

$$SiC\ (sólido)\ +\ SiO_2\ (Sólido)\ \rightarrow\ Si\ (Sólido)\ +\ SiO\ (Gas)\ +\ CO\ (Gas)$$

Esta reacción produce Silicio Metalúrgico (MGS) con una pureza del 98%. Este silicio no es todavía lo suficientemente puro para poder utilizarlo en la fabricación de circuitos electrónicos. Por tanto es necesario un proceso de purificación. Para llevar a cabo tal proceso, el silicio es pulverizado y tratado con cloruro de hidrógeno para obtener Triclorosilano (Cl_3HSi), de acuerdo con la reacción:

$$SiO_2\ (Sólido)\ +\ 3HCl\ (Gas)\ \rightarrow\ Cl_3HSi\ (Gas)\ +\ H_2\ (Gas)$$

A temperatura ambiente el Triclorosilano es un líquido. La destilación fraccionada de este líquido permite eliminar las impurezas indeseadas. A continuación, el Triclorosilano se reduce con hidrógeno para obtener Silicio Electrónico EGS (*Electronic Grade Silicon*):

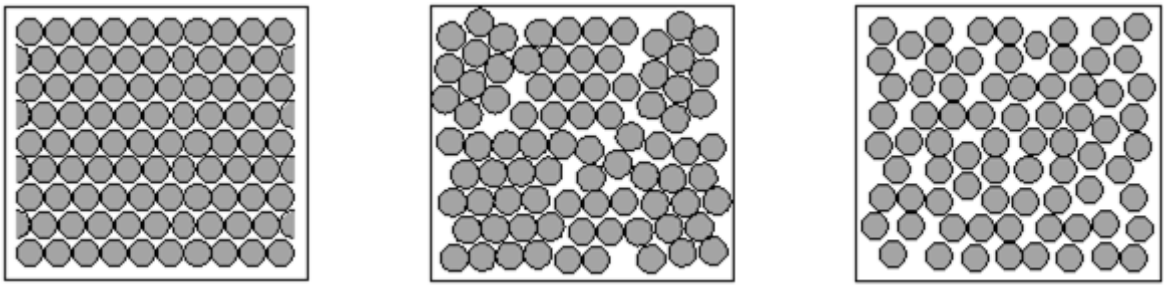
$$Cl_3HSi\ (Gas)\ +\ H_2\ (Gas)\ \rightarrow\ Si\ (Gas)\ +\ 3HCl\ (Gas)$$

Esta reacción tiene lugar en un reactor que contiene una barra de silicio caliente que sirve para que el silicio electrónico se deposite sobre ella. El EGS es un silicio policristalino de alta pureza (concentración de impurezas en una parte por mil millones) y es el elemento de partida para crear silicio monocristalino.

Clasificación un material sólido según su ordenación atómica:

- Estructura cristalina
- Estructura policristalina
- Estructura amorfa

En la figura III.8 la estructura cristalina y la amorfa son ilustradas con una vista microscópica de sus átomos, mientras que la estructura policristalina se muestra de una forma más macroscópica con sus pequeños cristales con distinta orientación pegados unos con otros.



Estructura cristalina Estructura policristalina Estructura amorfa

Figura III.8 Tipos de Estructuras

Crecimiento en volumen

Una vez que se ha conseguido silicio de alta pureza o EGS (Electronic Grade Silicon) para la fabricación de un circuito integrado se requiere Silicio con estructura cristalina. Para conseguir un cristal de Si se pueden utilizar varias técnicas. Las más importantes son:

- El método de Czochralski
- El método de Zona Flotante

Método de Czochralski

El método de Czochralski es el método empleado en el 90% de los casos para obtener silicio monocristalino a partir de silicio policristalino (EGS). Este método utiliza para el crecimiento de cristales un aparato denominado “puller”, que consta de tres componentes principales como se muestra en la Figura III.9 y III.10.

- Un horno, que incluye un crisol de sílice fundida (SiO_2), un soporte de grafito, un mecanismo de rotación (en el sentido de las agujas del reloj) un calentador y una fuente de alimentación.
- Mecanismo de crecimiento del cristal, que incluye un soporte para la semilla (muestra patrón del cristal que se pretende crecer) y un mecanismo de rotación (en el sentido contrario al de las agujas del reloj).
- Mecanismo del control de ambiente. Incluye una fuente gaseosa (argón por ejemplo), un mecanismo para controlar el flujo gaseoso y un sistema de vaciado.

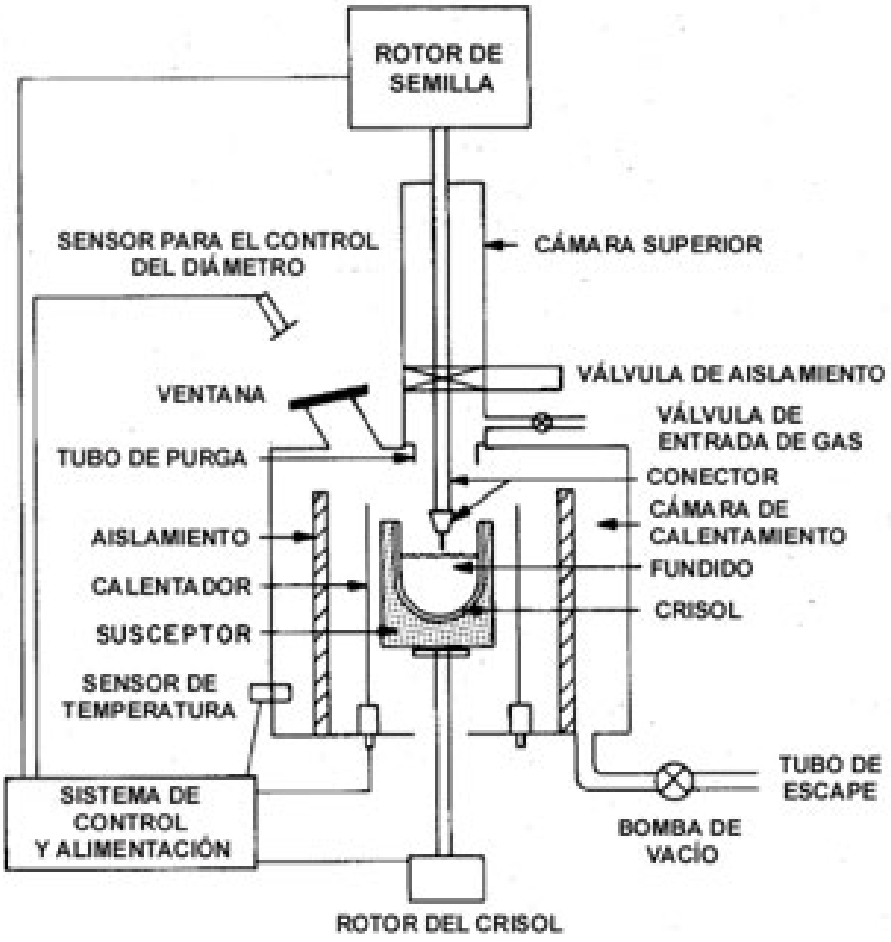


Figura III.9 Esquema de un Puller

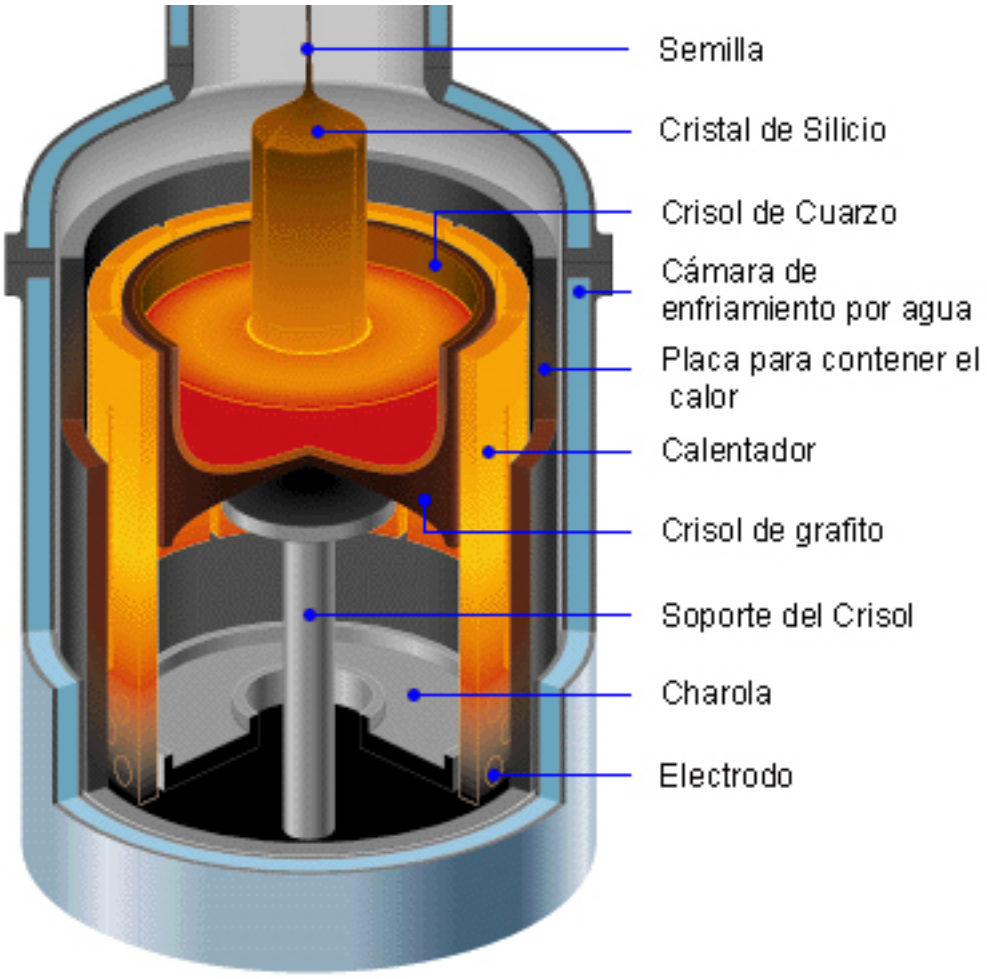
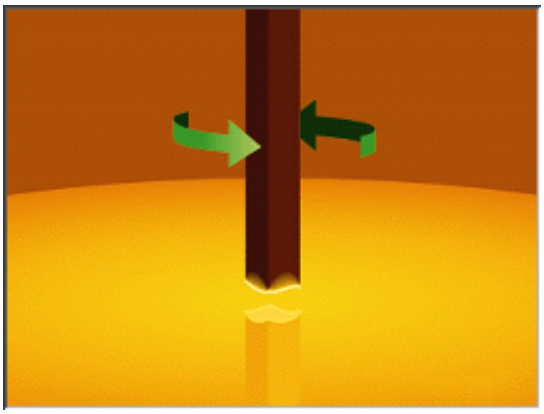
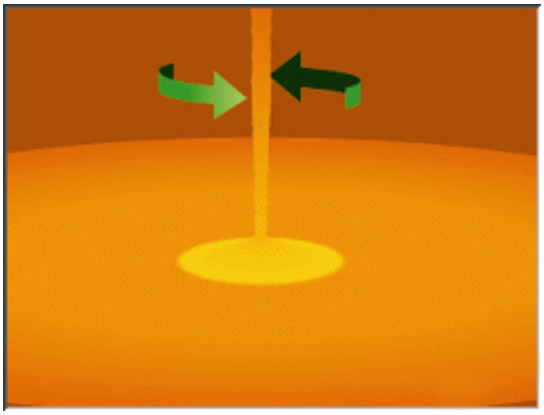
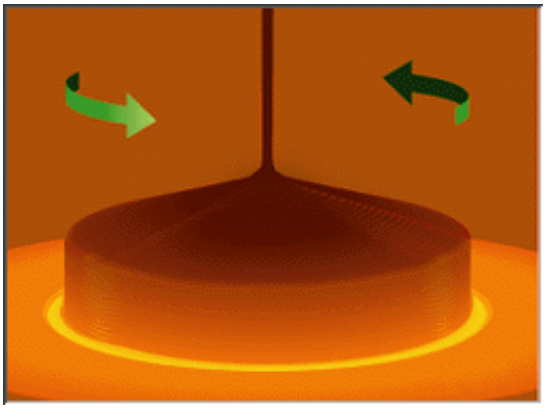


Figura III.3.10 Esquema de un Puller (2)

El proceso de crecimiento se detalla a continuación en la Tabla III.2.

El silicio policristalino (EGS) se coloca en el crisol y el horno se calienta a una temperatura superior a la de fusión del silicio obteniéndose el material fundido (MELT).	
Se suspende sobre el crisol una muestra pequeña del tipo de cristal que se quiere crecer.	
Se introduce la semilla en el fundido, parte de la misma se funde, pero la punta de la misma aún toca a la superficie del líquido.	
Se levanta lentamente la semilla. El progresivo enfriamiento en la interface sólido-líquido proporciona silicio monocristalino con la misma orientación cristalina que la semilla pero de mayor diámetro.	

En este proceso se gira tanto la semilla como el crisol en sentido contrario. Controlando cuidadosamente la temperatura, la velocidad de elevación y rotación de la semilla y la velocidad de rotación del crisol, se mantiene un diámetro preciso de la barra de cristal.

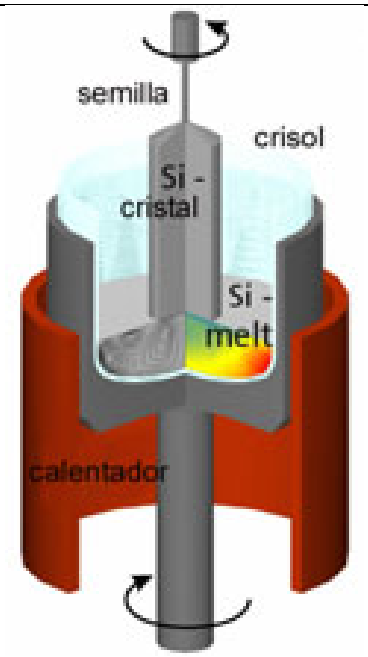


Tabla III.2 Proceso de crecimiento usando el método de Czochralski.

Mientras los lingotes son estirados, se refrescan para que adquiera un estado sólido. La longitud del lingote vendrá determinada por la cantidad de silicio fundido que hay en el crisol. Algunos ejemplos se muestran en la figura III.11.

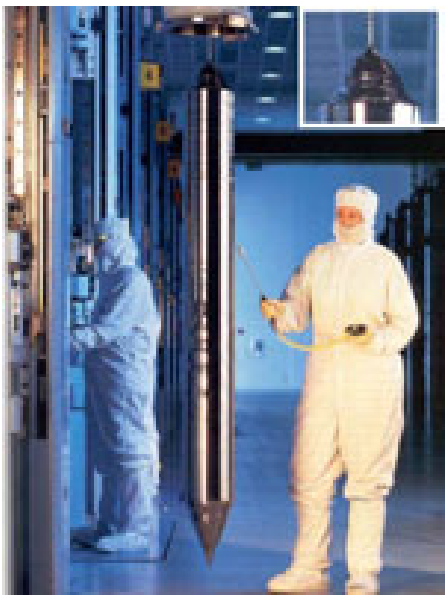


Figura III.11 Lingotes de Si crecidos por el método de Czochralski

En este proceso se añaden la cantidad de impurezas necesarias para formar un semiconductor tipo N o P con el dopado deseado. Normalmente la concentración de impurezas es de $10^{15} cm^{-3}$. Para conseguir esta concentración se incorpora cuidadosamente una pequeña cantidad de dopante por ejemplo Fósforo (para conseguir semiconductor tipo N) o Boro (para tipo P) al Silicio fundido.

Efecto de segregación:

- La concentración de dopante del silicio una vez que se solidifica es siempre inferior a la del silicio fundido.
- Esta segregación causa que la concentración del dopante aumente a medida que la barra de cristal crece.
- La concentración de impurezas es menor en el lado de la semilla que en el otro extremo.
- También se tiene un pequeño gradiente de concentración a lo largo del radio de la barra de cristal.

El Silicio fabricado por el método de Czochralski contiene una considerable cantidad de oxigeno, debido a la disolución del crisol de Sílice (SiO_2). Este oxigeno no es perjudicial para el silicio de baja resistividad usado en un circuito integrado, además puede controlar el movimiento accidental de impurezas metálicas. Sin embargo para aplicaciones de alta potencia donde se necesita Si con alta resistividad este oxigeno es un problema. En estos casos se usa el método de Zona Flotante.

Método de Zona Flotante

El método Zona Flotante se utiliza para crecer silicio monocristalino con concentración de impurezas más bajas que las normalmente obtenidas por el método de Czochralski.

Pasos que se realizan en el método de zona flotante:

- El proceso parte de un cilindro de silicio policristalino.
- Se sostiene verticalmente y se conecta uno de sus extremos a la semilla.

- Una pequeña zona del cristal se funde mediante un calentador por radio frecuencia que se desplaza a lo largo de todo el cristal desde la semilla.
- El Si fundido es retenido por la tensión superficial entre ambas caras del Si sólido.
- Cuando la zona flotante se desplaza hacia arriba, el silicio monocristalino se solidifica en el extremo inferior de la zona flotante y crece como una extensión de la semilla.

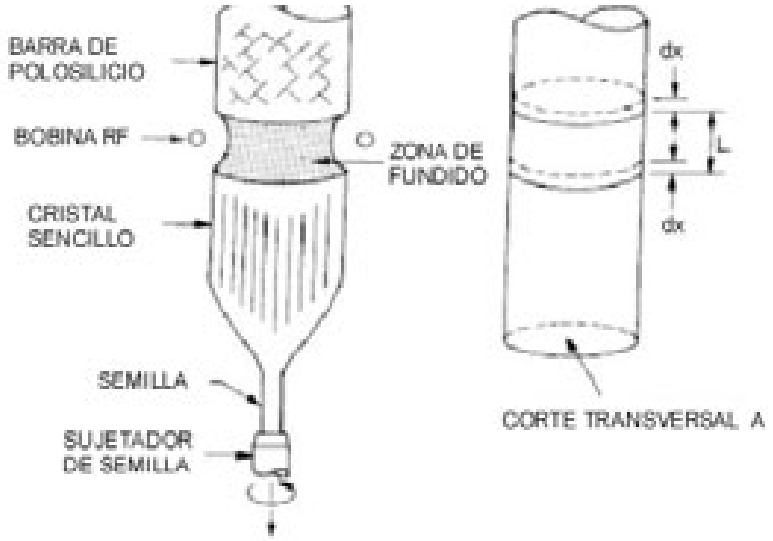


Figura III.12 Método de Zona flotante para el crecimiento de un cristal

Mediante este proceso de "float zone" pueden obtenerse materiales con resistividades más altas que mediante el método de Czochralski. Además, como no se necesita crisol, no existe, como en el caso anterior, posible contaminación desde el crisol. En la figura III.12 se muestra el Método de Zona Flotante.

Corte limpiado y pulido

Después de crecido el cristal la primera operación a realizar es quitar los extremos del lingote, tanto el de la semilla, como el último extremo crecido

La operación siguiente es desgastar la superficie hasta que quede definido el diámetro del lingote.

A continuación, y paralela a la generatriz del cilindro se hacen unas marcas planas para especificar la orientación del cristal y el tipo de conductividad del material. La Figura III.13 muestra las marcas realizadas y el significado de éstas.

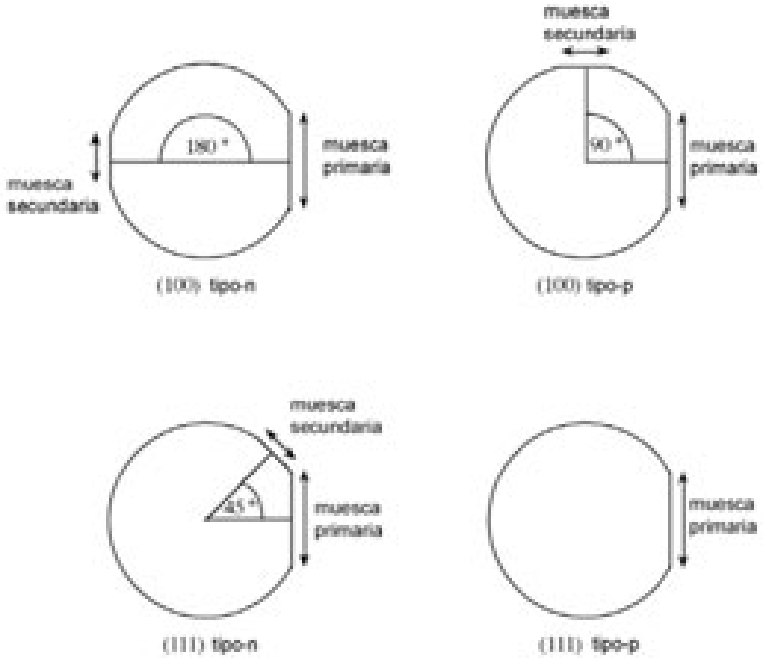


Figura III.13 Marcas para señalar la orientación del cristal.

Una vez realizadas estas operaciones, el lingote está preparado para ser cortado en obleas. El corte en obleas se suele realizar con una sierra de filo de diamante circular que corta por su parte interior.

Cortadas las obleas, se someten a un proceso de esmerilado, las dos caras de estas son tratadas con una mezcla de Al2O3 y glicerina para producir una superficie plana homogénea con un error de 2 μm. Esta operación daña y contamina la superficie y bordes de la oblea. Para reparar estos daños, las obleas son limpiadas mediante ataques químicos (Limpieza RCA).

El paso final en la obtención de las obleas es el pulido Figura III.14, cuyo propósito es obtener una superficie plana dónde puedan definirse los detalles de los dispositivos electrónicos.

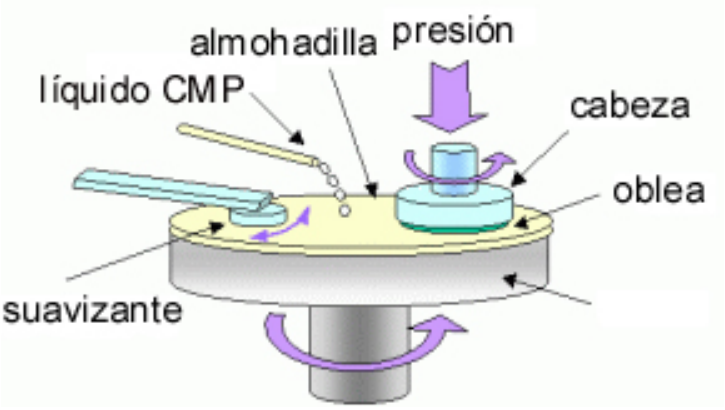


Figura III.14 Procesos de pulido

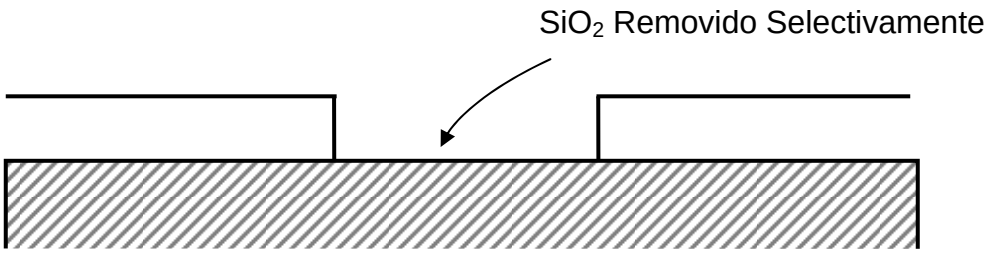
Proceso planar del silicio

Además de las propiedades semiconductoras del silicio, la razón principal que ha llevado al silicio a ser el material más utilizado para la fabricación de circuitos integrados, es la habilidad de formar sobre él una capa de óxido estable, de buena calidad y con magníficas propiedades aislantes.

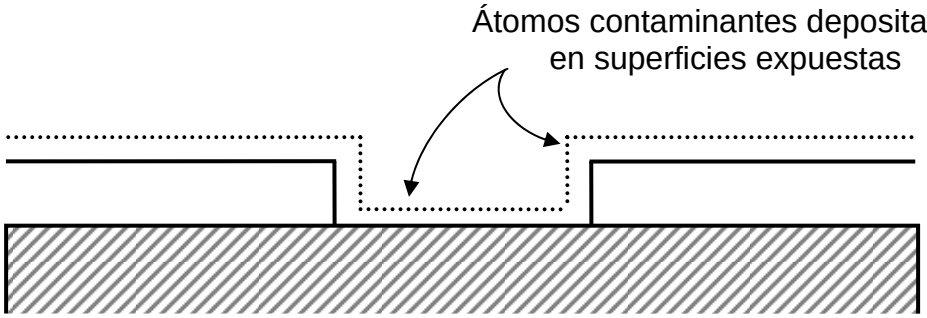
Esta capacidad, que no se consigue con cualquier combinación aislante-semiconductor, hace posible la introducción de cantidades controladas de dopantes en áreas selectivas del substrato. La habilidad de dopar selectivamente regiones de la oblea, es la clave para la producción de arreglos densos de dispositivos en circuitos integrados. Esta habilidad se basa en dos propiedades químicas del sistema Si-SiO_2 :

- Grabado selectivo. Es posible utilizar diferentes agentes (físicos o químicos) que atacan sólo a uno de los dos materiales. Por ejemplo el ácido fluorhídrico disuelve el SiO2 pero no el Si.
- Protección contra la difusión de impurezas**.** Las capas de óxido crecidas encima del silicio evitan que los átomos de impurezas del dopante se difundan por el interior del silicio.

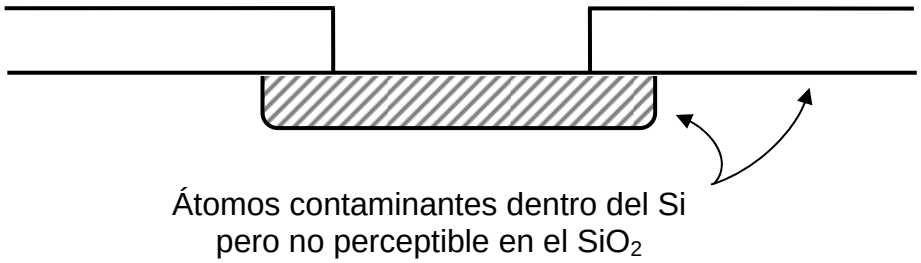
Formación del óxido



Grabado y eliminación selectiva del óxido.



Deposición de dopantes cerca de la superficie de la oblea.



Difusión de las impurezas en las áreas no protegidas del silicio.

Tabla III.3 Tecnología planar del silicio

Estas dos propiedades hacen posible la introducción de átomos de dopante únicamente en las áreas del silicio que no han sido cubiertas por SiO_2. Las zonas cubiertas o protegidas se definen cuidadosamente usando películas de polímeros fotosensibles que son sensibilizados usando máscaras fotográficas y algún medio de iluminación. El polímero sensibilizado protege el SiO_2 del ataque del ácido fluorhídrico. De esta forma se consigue abrir en el óxido ventanas que dejan al descubierto zonas del silicio cristalino.

Cuando la muestra es colocada en un ambiente en el que se depositan átomos en la superficie de la oblea, estos átomos entrarán únicamente en el silicio no protegido, con lo que se consigue dopar selectivamente la oblea. Los pasos más importantes de la tecnología planar del silicio se muestran en la Tabla III.3.

La repetición de estos procesos constituye la tecnología planar del silicio. Una ventaja importante de este proceso es que cada paso de fabricación se aplica a toda la oblea. Por lo tanto, es posible hacer e interconectar una gran cantidad de dispositivos con gran precisión para construir circuitos integrados.

Proceso de Oxidación térmica

Una vez limpiada la superficie de la oblea se somete a un proceso de oxidación para crear una capa de oxido en su superficie. Esta capa protege en primer lugar la superficie de impurezas y sirve además de máscara en el proceso de difusión posterior. Se puede formar bien por oxidación térmica o por deposición. En el caso de la deposición ambos elementos Si y O_2 se dirigen a la superficie de la oblea y reaccionan allí formando una capa SiO_2.

En el proceso de oxidación térmica se produce una reacción entre los átomos de silicio de la superficie de la oblea y oxigeno dentro de un horno a alta temperatura. El oxido creado por oxidación térmica tiene mucha mas calidad que el que se obtiene mediante deposición. Aunque su estructura es amorfa, tiene una relación métrica firme casi

perfecta y esta fuertemente unido al silicio. Además la interface SiO_2– Si tiene muy buenas propiedades eléctricas. Para crear el oxido térmico las obleas de silicio se montan en un carrete de cuarzo y este se mete dentro de un tubo de cuarzo situado dentro de un horno de apertura cilíndrica calentado por resistencia. La figura III.15 muestra el montaje básico para un proceso de oxidación térmica.

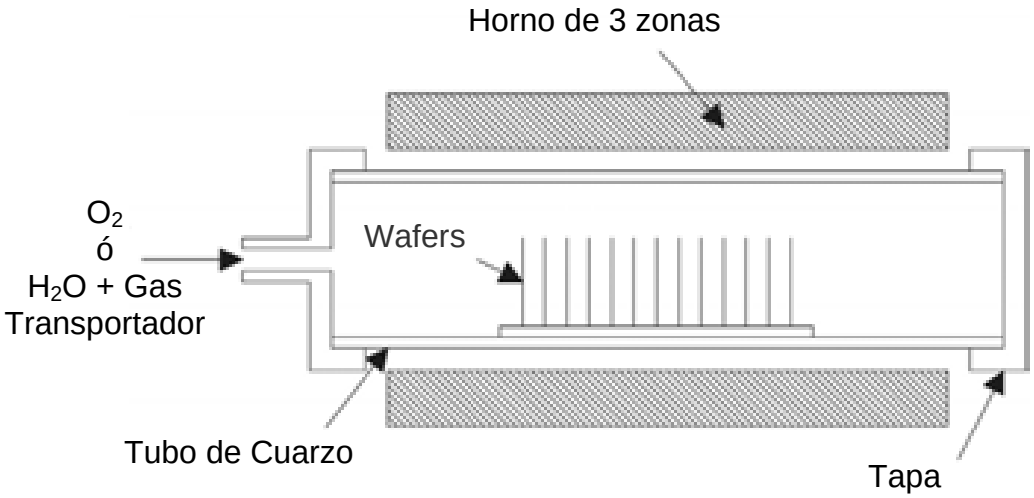


Figura III.15 Proceso de oxidación térmica

El rango de temperaturas a la que se produce la reacción esta comprendido entre los 850 y 1100ºC. El silicio no se funde hasta los 1412ºC. La temperatura de oxidación se mantiene bastante por debajo para evitar la generación de defectos en el cristal y el movimiento de los dopantes añadidos anteriormente. Además del soporte, el tubo de cuarzo y otros elementos del horno se empiezan a reblandecer y degradar a partir de los 1150ºC. En la figura III.16 se muestra la superficie de la oblea posterior al proceso de Oxidación.

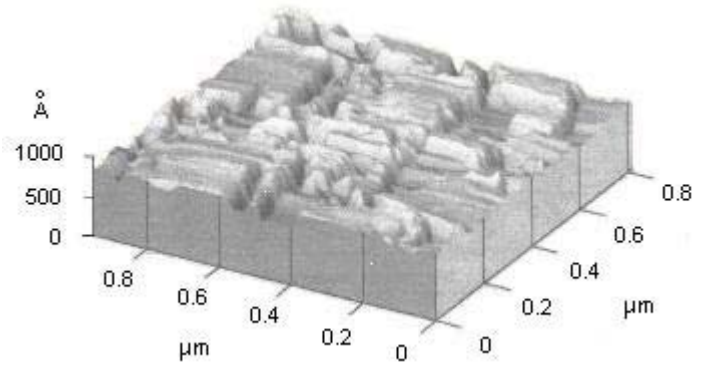


Figura III.16 Superficie de la oblea posterior al Proceso de Oxidación Térmica a 900ºC en O_2

La oxidación Térmica puede ser de dos tipos:

- Oxidación húmeda
- Oxidación seca

En la oxidación húmeda se introduce vapor de agua en el horno la reacción que sucede es la siguiente:

$$Si\ (Sólido)\ +\ 2H_2O\ (Gas)\ \rightarrow\ SiO_2\ (Sólido)$$

El caso de la oxidación seca se introduce gas de oxigeno puro. La reacción que se produce es la siguiente:

$$Si\ (Sólido)\ +\ O_2\ (Gas)\ \rightarrow\ SiO_2\ (Sólido)\ +\ 2H_2\ (Gas)$$

La oxidación húmeda es más rápida y se utiliza para crear óxidos gruesos. Con la oxidación seca se consiguen óxidos de mayor calidad pero esta técnica no es apropiada para la creación de óxidos gruesos ya que se puede producir una redistribución de las impurezas introducidas en los anteriores procesos. En la figura III.17 se muestran dos tipos de hornos (vertical y horizontal) para el proceso de Oxidación.

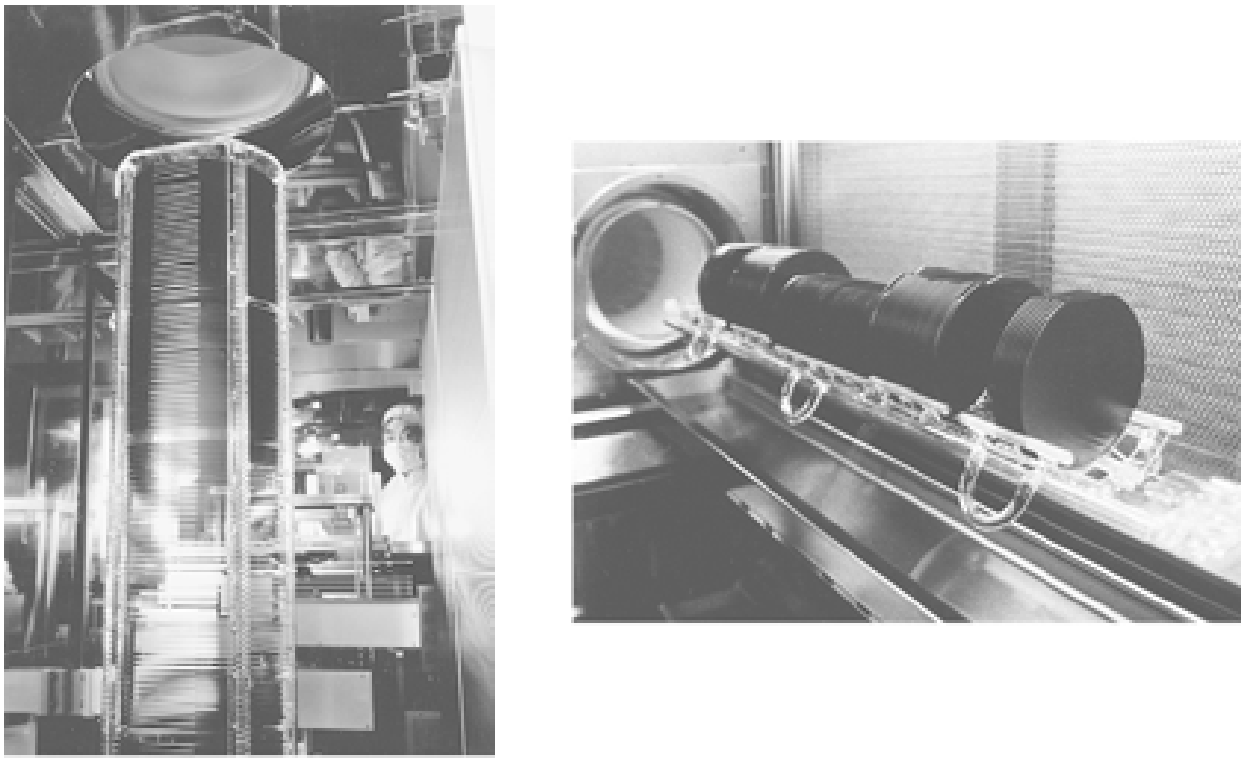


Figura III.17 Horno vertical y horizontal

La oxidación se produce en la interface Si-SiO_2 por tanto el oxigeno tiene que difundirse a través del oxido hasta la interface para reaccionar allí con el Silicio.

- Cuando el espesor del oxido formado es pequeño el crecimiento del oxido esta limitado por la reacción en la interface Si-SiO_2 en este caso el espesor varia linealmente con el tiempo.
- Cuando el espesor del oxido es grande, la velocidad de crecimiento vendrá limitada por la difusión de las especies oxidantes en este caso el espesor del oxido es proporcional a la raíz cuadrada del tiempo.

Durante el proceso de oxidación parte de la capa de Si se consume de forma que la interface Si-SiO_2 se introduce en el Si (figura III.18). Por cada micra de oxido crecido se consume 0.44 micras de Si.

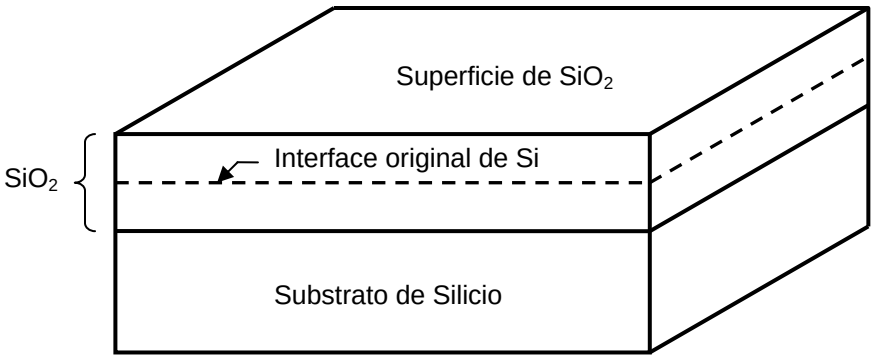


Figura III.18 Consumo de Si durante el proceso de oxidación

III.3.2 Foto Litografía

Proceso de litografía y grabado

Una vez creada la capa de aislante SiO_2 sobre la oblea, parte de ella debe ser eliminada selectivamente en aquellos sitios en los que deben introducirse los átomos de dopante. El grabado selectivo se realiza generalmente mediante el uso de un material sensible a la luz denominado fotoresist o foto resina. Para ello, la oblea oxidada se cubre en primer lugar por una capa de foto resina. A continuación se recubre está con un negativo fotográfico parcialmente transparente denominado máscara o foto máscara (figura III.19).

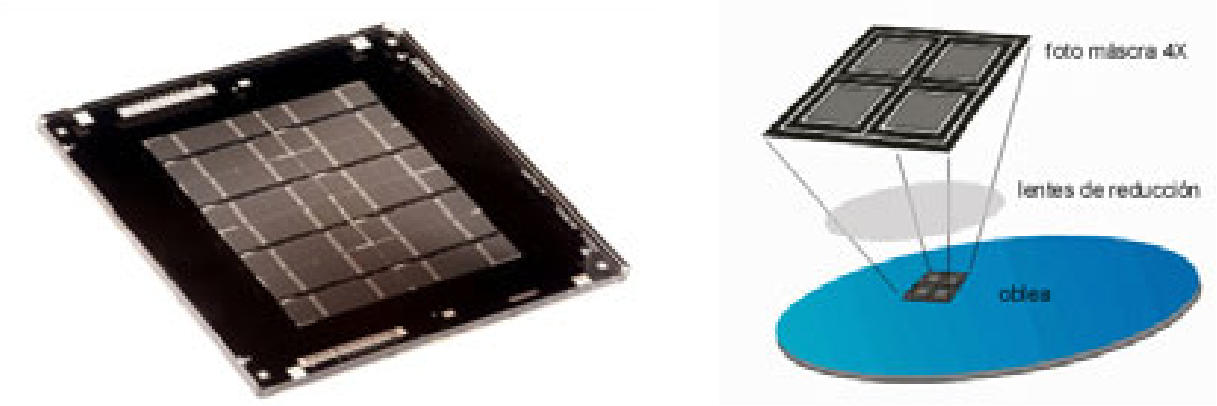


Figura III.19 Foto máscara

La luz ultravioleta cambia la estructura de la foto resina: las moléculas de la foto resina para negativos se unen entre si (polimerizan) en las regiones expuestas a la luz. Por el contrario, en el caso de foto resinas para positivos, los enlaces entre las moléculas se rompen al iluminarse, permaneciendo polimerizadas el resto. Las partes no iluminadas de la foto resina no se ven afectadas. Una vez convenientemente alineada la máscara se ilumina con luz ultravioleta.

Las áreas no polimerizadas de la foto resina se disuelven selectivamente usando por ejemplo tricloroetileno. De esta forma las zonas polimerizadas, resistentes al ataque del ácido quedan protegiendo al SiO_2.

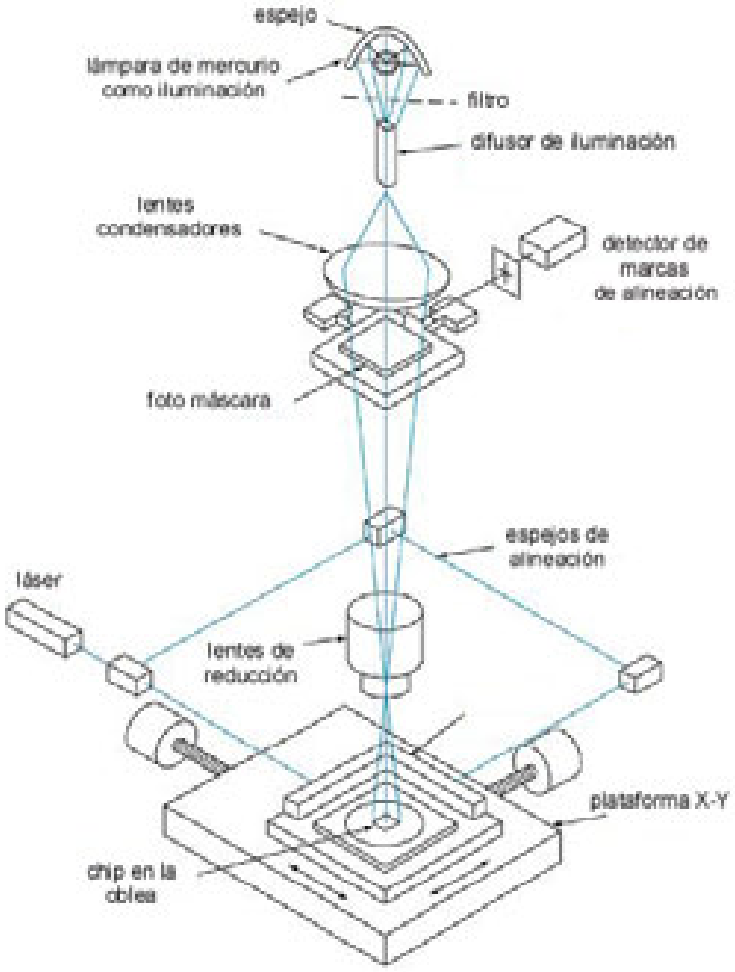


Figura III.20 Esquema de un sistema de alineamiento

El método más utilizado en la actualidad en fotolitografía es la proyección. El patrón de la máscara es proyectado directamente sobre la superficie de la oblea mediante una máquina denominada escáner o stepper (Figura III.21 y Figura III.22). Las funcionalidades del stepper/scanner son similares a las de un proyector. La luz procede de una lámpara de arco de mercurio o de un láser excímero focalizado a través de un complejo sistema de lentes sobre la máscara, que contiene la imagen deseada.

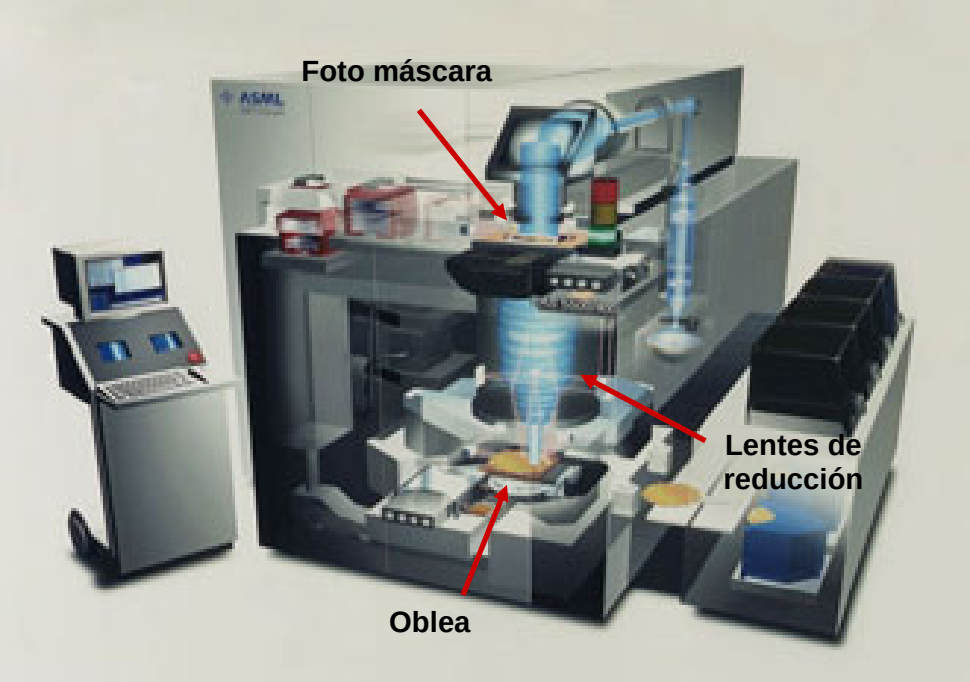
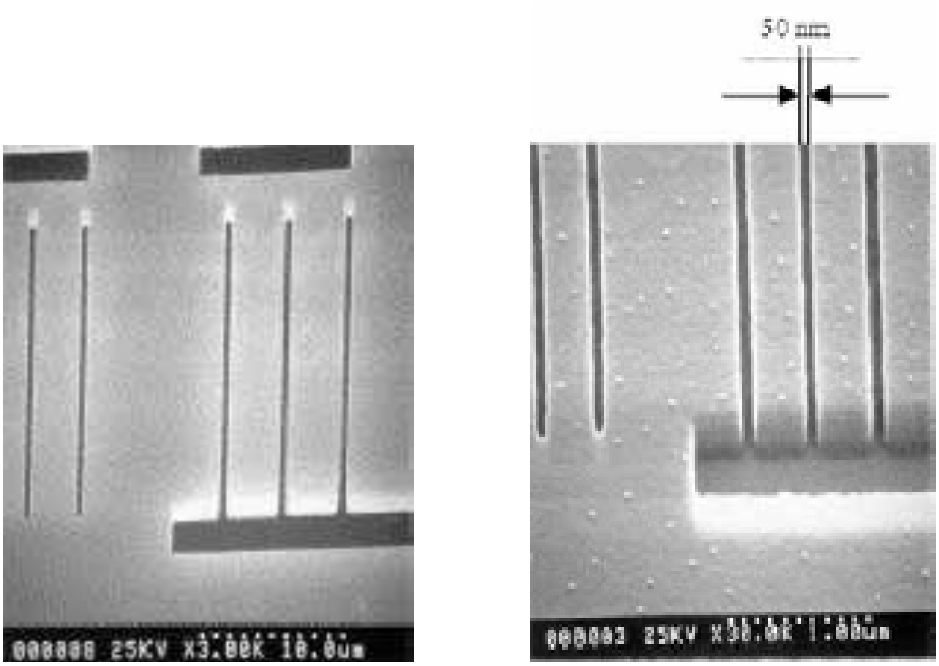


Figura III.21 Scanner 5500 compañía ASML

Figura III.22 TWINSCAN de la compañía ASML

La luz pasa a través de la máscara y se focaliza sobre la superficie de la oblea mediante un sistema de lentes de reducción. El sistema de reducción puede variar según el diseño pero suele ser bastante usual un orden de magnitud en la reducción de 4X y 5X. Cuando la imagen es proyectada sobre la oblea, el material foto sensible actúa sólo a ciertos rangos de longitudes de onda, lo que causa que las regiones expuestas cambien sus propiedades físico-químicas (figura III.23).Generalmente se cambia la acidez del sustrato de la resina, haciendo que sea más ácido o alcalino que la parte no expuesta. Si la región expuesta es más ácida se dice que es una resina positiva, mientras que es negativa si es más alcalina (figura III.24). La resistencia es "revelada" por exposición a una solución alcalina que elimina las partes expuestas de la resina (en el caso de una foto resina positiva) o no expuesta (foto resistencia negativa).



Patrones en la Máscara Patrones en la oblea

Figura III.23 Patrones de la foto máscara proyectados en la oblea

La capacidad para imprimir imágenes claras depende de la longitud de onda empleada en la proyección. Las fuentes de luz actuales emplean longitudes de onda en el rango del ultravioleta profundo (DUV), es decir, de longitudes de onda que varían entre los 248

y 193 nanómetros. Estas longitudes de onda permiten una capacidad de discernimiento de detalles de como máximo 50 nanómetros. Para reducir este límite de imprimación por debajo de los 50 nm se necesitan de otras técnicas basadas en luz de 193 nm, así como técnicas de inmersión en líquidos (litografía por inmersión).

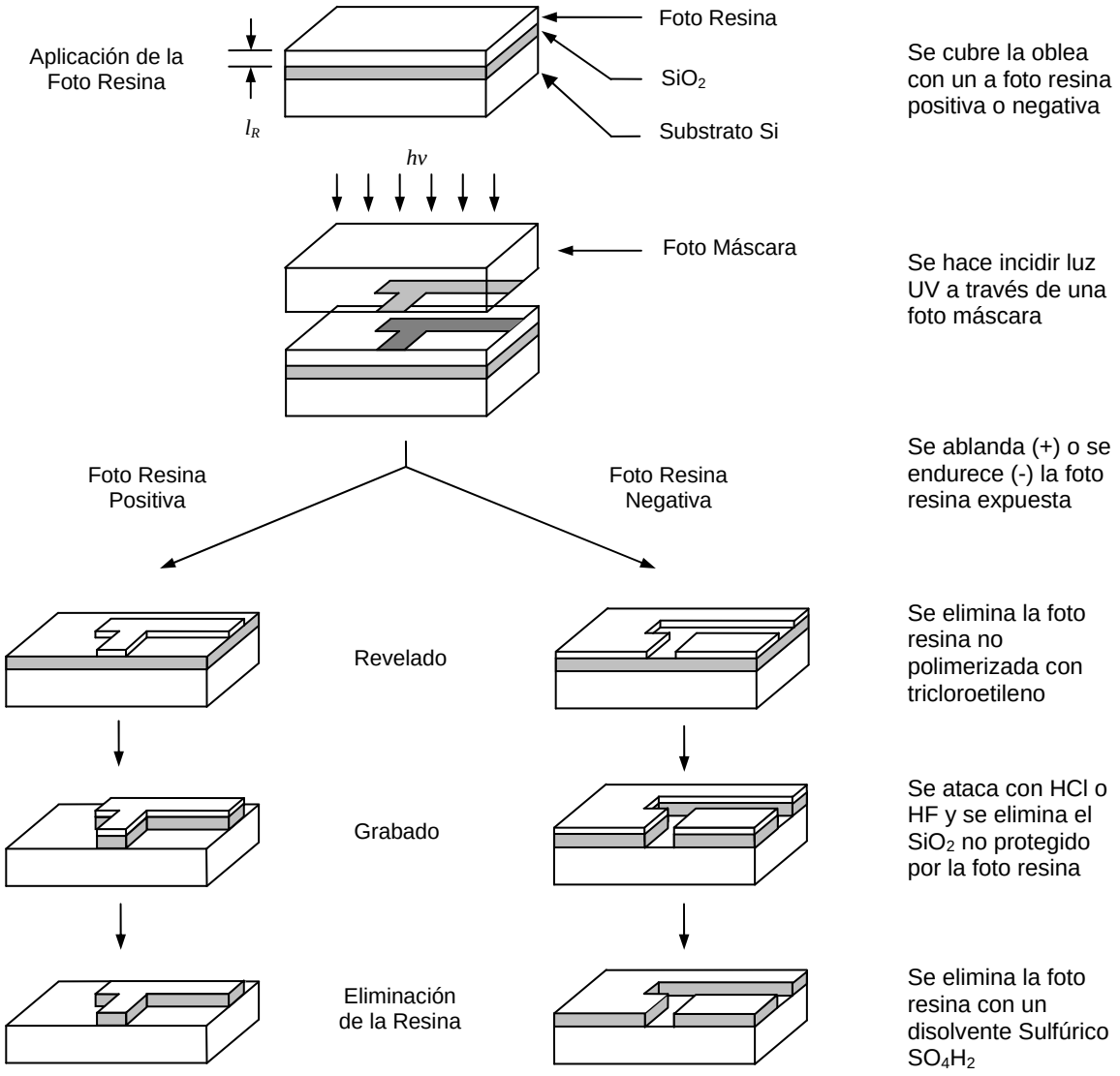


Figura III.24 Proceso de transferencia de patrones usando foto litografía.

Dentro de la litografía de semiconductores existe el problema de la aproximación a la dimensión mínima de los dispositivos con la longitud de onda de la luz utilizada en la

exposición óptica para sensibilizar la foto resina, el cual puede producir que los fenómenos de difracción puedan limitar la resolución del método (tamaño mínimo que puede distinguirse). Para evitar esta limitación se han propuesto técnicas alternativas:

A continuación se muestra en la siguiente Tabla III.4, las ventajas e inconvenientes en el uso de litografía por haz de electrones y la litografía por rayos x.

Litografía por haz de electrones	Un chorro de electrones energéticos se dirige sobre la fotorresistencia que queda sensibilizada. En vez de sensibilizar todos los patrones a la vez, se van "dibujando" uno a uno los distintas partes del circuito integrado, por lo que no es necesaria ninguna máscara.	Ventaja	Se consigue una resolución mucho mayor que cualquier dimensión del circuito integrado.
		Inconveniente	Proceso lento, puesto que hay que grabar uno a uno los diferentes partes del circuito integrado.
Litografía por rayos x	Un haz de rayos X se hace pasar por una máscara para sensibilizar selectivamente la fotorresistencia. Al igual que la fotolitografía convencional, la litografía con rayos X permite grabar varios patrones de forma simultánea	Ventajas	se consigue una mejor resolución, y por lo tanto unos dispositivos de menor tamaño ya que la longitud de onda es mucho más pequeña
		Inconvenientes	Las máscaras son difíciles de fabricar y que además la utilización de rayos X puede dañar las partes activas de los dispositivos.

Tabla III.4, las ventajas e inconvenientes en el uso de litografía por haz de electrones y la litografía por rayos x

III.3.3 ETCH (mojado y seco)

Proceso de Grabado

Consiste en eliminar la parte de SiO_2 no protegida para abrir las ventanas deseadas en el óxido, que dejen a la vista el substrato de silicio. Para eliminar la parte de óxido no protegido puede usarse un baño de ácido fluorhídrico que ataca al dióxido de silicio no protegido, pero no ataca al silicio. Existen dos tipos de grabado:

- Grabado Húmedo o químico
- Grabado seco o por plasma

Grabado Húmedo o químico

Características:

- Baño de ácido fluorhídrico o clorhídrico que ataca SiO_2 no protegido
- Gran selectividad: eliminan la capa de óxido produciendo un ataque muy pequeño sobre los materiales subyacentes
- Ataque isotrópico igual en todas las direcciones no sólo se ataca hacia abajo sin que también se ataca lateralmente por debajo del protector.

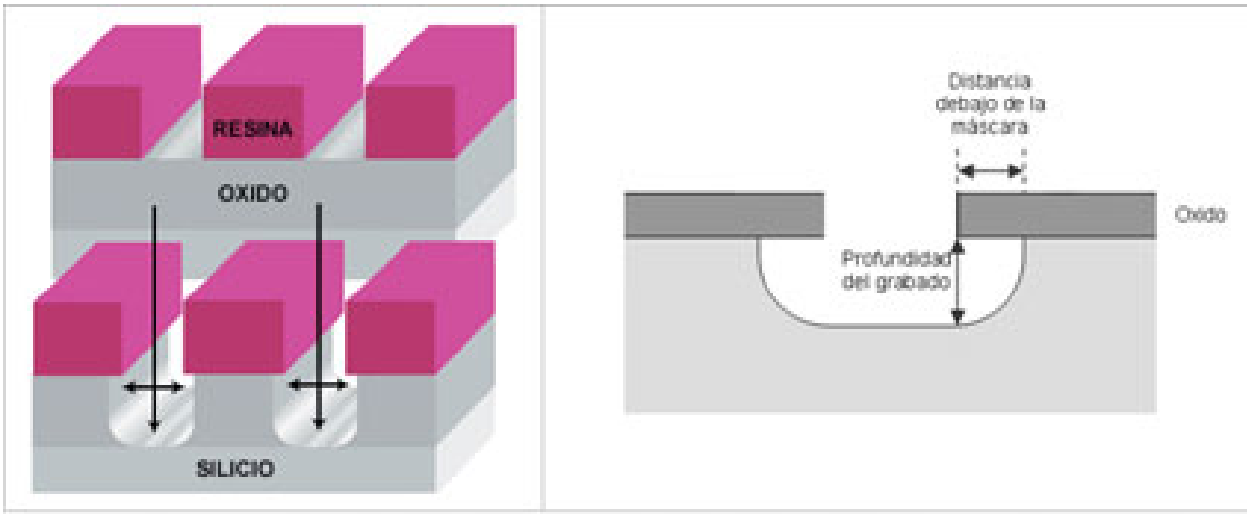


Figura III.25 Grabado húmedo o mojado

Grabado seco o por plasma

Características:

- Se usa un plasma con un gas ionizado
- Grabado físico, químico o combinado
- Obtenemos las capas perfectamente definidas
- Ataque anisótropo: ataque únicamente en la dirección vertical

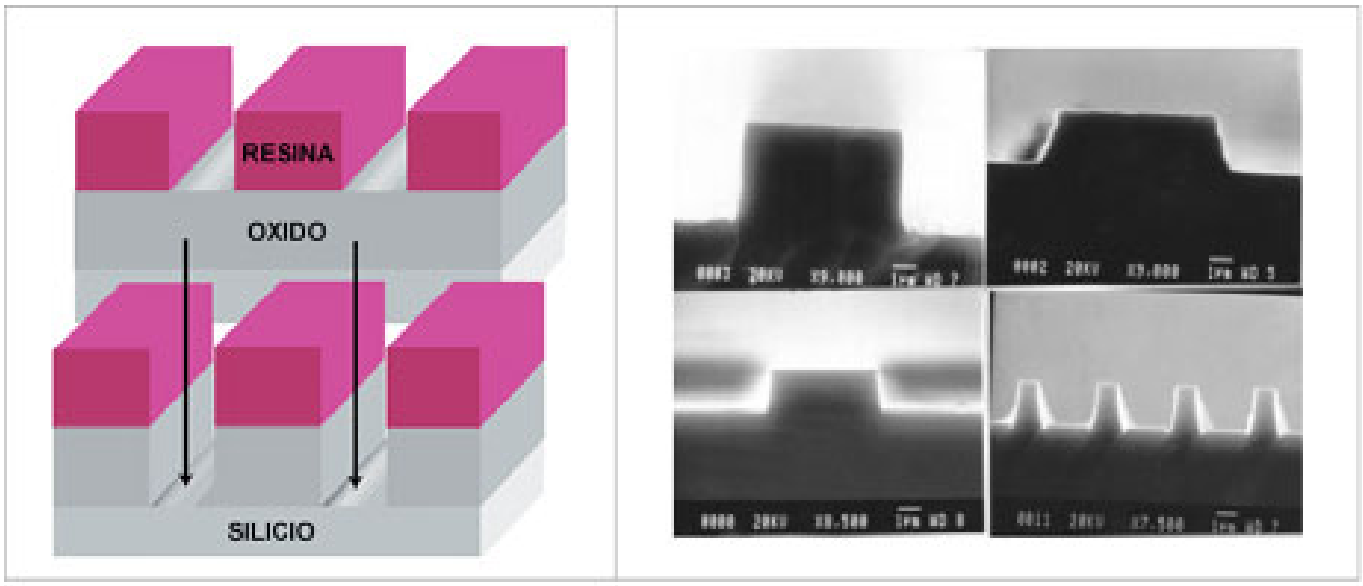


Figura III.26 Grabado seco

Entre los grabados por plasma podemos distinguir tres tipos dependiendo del tipo de mecanismo físico o químico del proceso:

- Grabado por plasma puro
- Grabado por haz de iones RIBE
- Grabado por iones reactivos RIE

En los tres casos, la oblea se expone a un plasma, que consiste en un gas parcial o totalmente ionizado compuesto de iones, electrones y neutrones. El plasma se produce cuando un campo eléctrico de suficiente magnitud se aplica al gas, causando la ionización de las moléculas o átomos del gas.

Grabado por iones reactivos (RIE), en este caso se combina el grabado físico con el químico, además del efecto físico del bombardeo de iones, las moléculas ionizadas reaccionan químicamente con el material que debe ser atacado, con lo que se consigue una mayor selectividad. El proceso de acción de grabado RIE queda esquematizado en la figura III.27 siguiente:

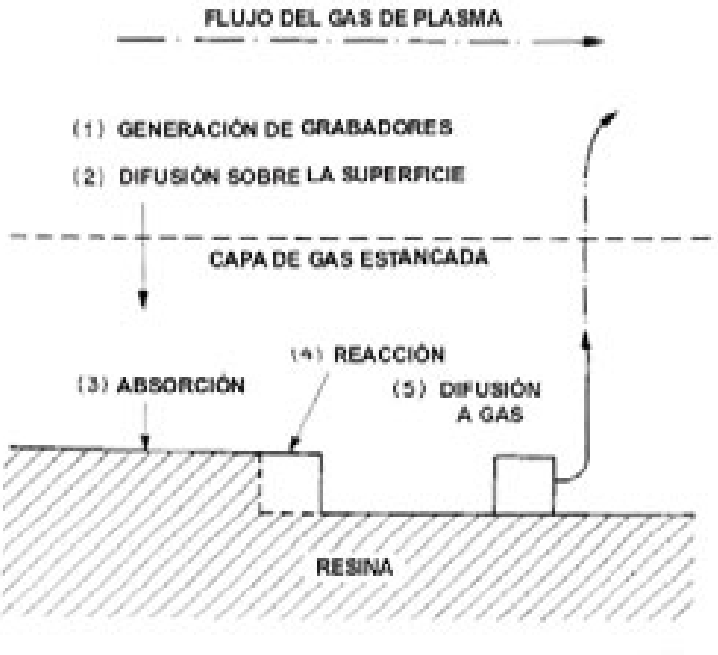


Figura III.27 Proceso de acción de grabado RIE

(1) El proceso comienza con la formación de los reactivos

(2) Los reactivos son transportados por difusión a través de una capa gaseosa de estaño hacia la superficie.

(3) La superficie adsorbe a los reactivos.

(4) Se produce la reacción química de los reactivos con la especies de la superficie, junto con efectos físicos (bombardeo iónico).

(5) Los materiales resultados de la reacción química o bombardeo físico se desprenden de la superficie y son eliminados por un sistema de vacío.

Después de haberse realizado el grabado se elimina la capa de foto resina con un disolvente orgánico por ejemplo Sulfúrico SO_4H_2.

Tabla III.5 Capas, procesos y materiales

Layers	Termal Oxidation	Chemical Vapor deposition	Evaporation	Sputtering	Electroplating
Insulators	Silicon dioxide	Silicon dioxide		Silicon dioxide	
		Silicon nitride		Silicon monoxide	
Semiconductors		Eptaxial silicon			
		Polycrystalline silicon			
Conductors			Aluminum	Aluminum	Gold
			Aluminum alloys	Alluminum alloys	Copper
			Nichrome	Tungsten	
			Gold	Titanium	
				Molybdenum	

III.3.4 Implantación

Impurificación (adición de dopantes)

Una vez abiertas las ventanas en el oxido se realiza el proceso de impurificación mediante la adición de dopantes. Con estas técnicas se puede dopar selectivamente el substrato del semiconductor y producir regiones tipo P ó tipo N según convenga. Existen dos métodos para impurificar el substrato:

- Difusión
- Implantación iónica

Difusión

Para producir la difusión de impurezas en el interior del semiconductor, se colocan las obleas en el interior de un horno a través del cual se hace pasar un gas inerte que contenga el dopante deseado. El sistema es similar al empleado en la oxidación térmica. Los rangos de temperatura van entre 800ºC y 1200ºC. En el caso del silicio tipo P, el dopante más usual es el Boro y para conseguir un semiconductor tipo N se usa el Arsénico y Fósforo. La introducción de estos dopantes puede hacerse de diferentes formas a partir de fuentes sólidas, líquidas y gaseosas. Generalmente el material elegido es transportado hasta la superficie del semiconductor por un gas inerte Nitrógeno (N_2). En la siguiente figura III.28 se muestra el esquema de un sistema para la difusión de impurezas.

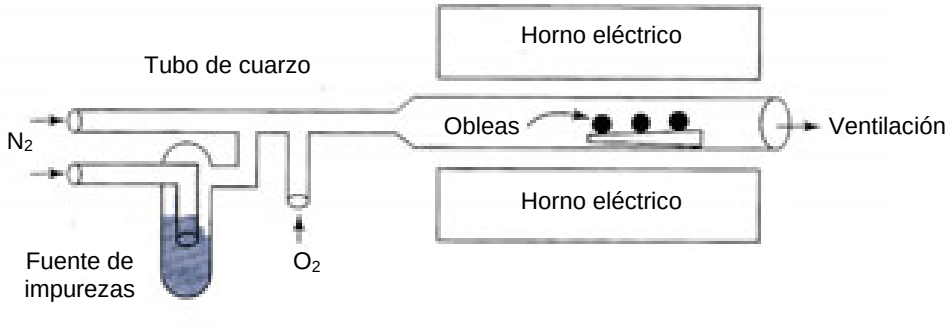


Figura III.28 Esquema de un sistema para la difusión de impurezas

Se puede distinguir entre dos formas al realizar la difusión:

- Con fuente limitada: cuando se mantiene la misma concentración de impurezas durante el proceso
- Con fuente ilimitada: se parte de una concentración inicial y no se añaden mas dopantes

Normalmente en el proceso de difusión se usan los dos métodos uno seguido del otro. La profundidad de la difusión dependerá del tiempo y de la temperatura del proceso (figura III.29). La concentración de dopante disminuye monótonamente a medida que se aleja de la superficie. La técnica de difusión tiene el problema de que las impurezas se difunden lateralmente.

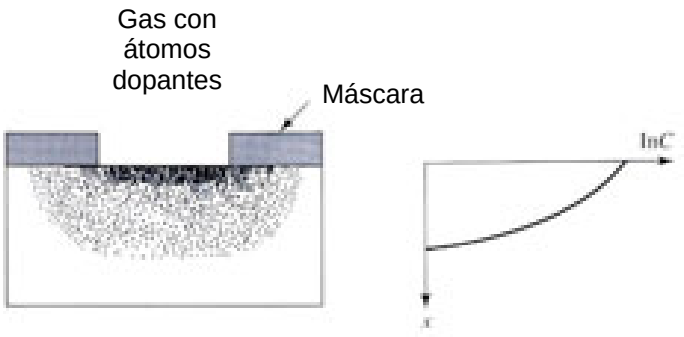


Figura III.29 Difusión

Implantación iónica

En esta técnica, los átomos del dopante son implantados en el interior del semiconductor por medio de haces iónicos de alta energía (figura III.30). El perfil del dopado tiene un máximo en el interior del semiconductor, y está determinado por la masa de los iones y la energía con que se hacen incidir los mismos sobre la superficie semiconductora. Las ventajas de la implantación iónica sobre la difusión son: tener el control preciso de la cantidad de dopantes introducidos, la reproducción de los perfiles de impurezas y una menor temperatura de proceso. Al introducir las impurezas se producen daños en el cristal, por lo cual posteriormente, se somete la oblea a un proceso de recocido “annealing” para una reordenación del cristal con las nuevas impurezas.

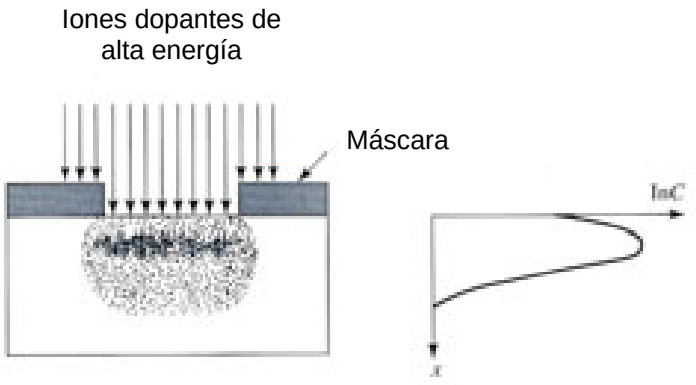


Figura III.30 Implantación Iónica

Formación de capas delgadas (Deposiciones y Epitaxia)

En el proceso de fabricación es necesario la formación de películas delgadas de distintos materiales: Óxidos, Polisilicio, metales, silicio amorfo y silicio cristalino.

En el caso del Crecimiento Epitaxial el material depositado formara un cristal siguiendo la estructura del substrato. Cuando se habla de deposición el material que se precipita no forma una estructura cristalina. Este es el caso de una capa aislante, polisilicio, metales y del silicio amorfo. En la figura III.31, se muestra la operación en la formación de capas sobre la oblea de silicio.

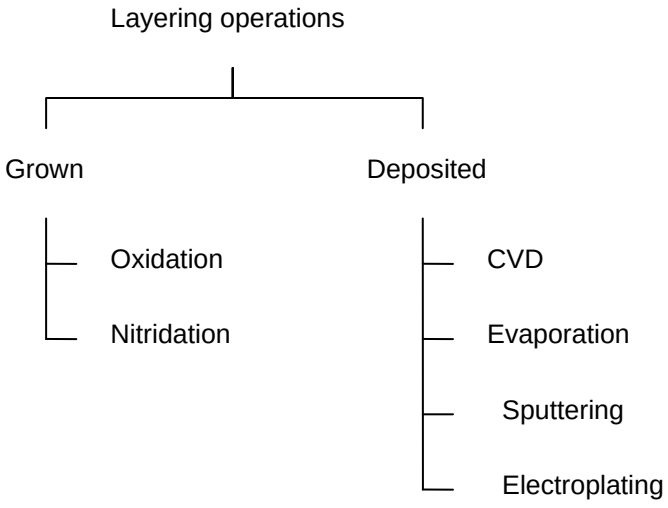


Figura III.31 Operación de formación de capas.

Epitaxia

La epitaxia es uno de los procesos más importantes en la fabricación de un circuito integrado. Es la base sobre la cual muchos dispositivos son construidos. La epitaxia es simplemente un crecimiento de un cristal sobre un substrato con estructura cristalina. Se pueden distinguir dos tipos de epitaxia.

- La homoepitaxia, cuando los materiales del substrato y la capa que se crece son del mismo material
- La heteroepitaxia, cuando el material del substrato y capa son de diferente material.

En principio es mucho mas sencillo conseguir capas epitaxiales de buena calidad si el substrato y la capa son iguales, homoepitaxia, ya que en este caso no existirán diferencias en la forma de la red cristalina como sucede en los procesos de heteroepitaxia. En el caso de crecimiento de silicio sobre el substrato, el proceso de epitaxia aporta grandes ventajas frente a otros procesos de fabricación; es un método mucho más sencillo para controlar el espesor, la concentración y perfil de dopado de la capa. Por ejemplo, se puede crear una capa de Silicio poco dopado sobre un substrato altamente dopado.

Deposiciones

El método más usado para realizar deposiciones es el Deposición Química de Vapor CVD (*Chemical Vapour Deposition*). En este caso se consigue la formación de una película sólida sobre un substrato mediante la reacción de reactivos químicos en fase de vapor de los componentes. La reacción química de los gases puede producirse en la superficie de la oblea o muy cerca (reacción heterogenia) o alejada de la superficie (reacción homogénea). El uso de reacciones heterogénea es más aconsejable, esto es a que existe una buena calidad en la creación de películas sobre la superficie de la oblea. Las reacciones homogéneas no son aconsejables, debido a que pueden crear películas de poca adherencia con baja densidad y defectos.

La estructura de la película dependerá del substrato sobre la cual se deposita (amorfa o cristalina) y de las condiciones del proceso temperatura presión del gas etc. La reacción de deposición esta provocada por el calentamiento del substrato pero la energía en el sistema se puede introducir generando un plasma dentro de la cámara.

Clases de deposiciones:

- Deposición Física de Vapor PVD (*Physical Vapour Deposition*)
  - vaporización térmica (Evaporative)
  - Deposición epitaxial por haces moleculares MBE (*Molecular Beam Epitaxy*)
  - Dispersión (*Sputtering*)

- Deposición Química de Vapor CVD (*Chemical Vapour Deposition*)
  - Presión Atmosférica APCVD (*Atmospheric-Pressure CVD*)
  - Presión Reducida LPCVD (*Low Pressure CVD*)
  - Asistido por Plasma PECVD (*Plasma Enhancement CVD*)

El proceso de PVD es estrictamente un proceso físico que implica la deposición de un recubrimiento mediante condensación sobre un sustrato, desde la fase de vapor. En comparación, la deposición química de vapor CVD, implica la intervención entre una mezcla de gases y la superficie de un sustrato calentado, provocando la descomposición química de algunas partes del gas y la formación de una película sólida en el sustrato.

Vaporización Térmica

La separación térmica es el método físico más sencillo para la producción de capas en vacío. El material de vaporización es vaporizado mediante el calentamiento de navecillas, filamentos o recipientes cónicos hechos de W, Mo o Ta, esto mediante flujo de corriente eléctrica (calentamiento por resistencia) (figura III.32). Normalmente el cambio de material sólido a gaseoso no ocurre de manera directa (sublimación), si no a través de la etapa intermedia de la fase liquida. Habiendo una distancia suficientemente

grande entre la fuente de vaporización y el sustrato, la fuente aparece como fuente puntual. A medida que los haces de vapor se dispersen sin efecto recíproco depende de la longitud libre media de trayectoria de las partículas. Dicha longitud será mayor mientras menor sea la presión, esto es, mientras mejor sea el vacío del recipiente.

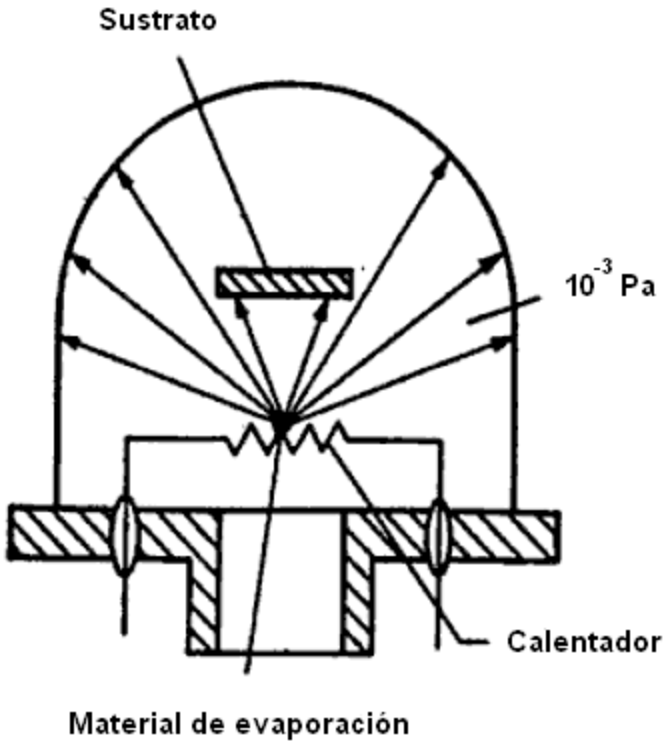


Figura III.32 Esquema de un equipo de evaporación

Deposición epitaxial por haces moleculares

El termino MBE (*Molecular Beam Epitaxy*), describe el crecimiento epitaxial de laminas delgadas de semiconductores compuestos mediante la reacción de haces moleculares térmicos de los elementos constituyentes con la superficie de un sustrato cristalino, mantenido a temperatura adecuada y en condiciones de ultravacío. Este método consiste en la evaporación de fuentes sólidas de manera que se producen haces moleculares y se dirigen sobre un sustrato caliente sobre el cual se deposita el material (figura III.33).

Es una técnica habitual en el crecimiento de hetero estructuras de semiconductores por la gran perfección cristalina que alcanza. Los haces moleculares inciden sobre un sustrato y diversas reacciones químicas ocasionan la deposición de monocapas sucesivas. Mediante el adecuado control de las especies químicas de los haces se puede variar la composición de las capas epitaxiales. Esta técnica, se debe de realizar en condiciones de ultravacío, ya que cualquier impureza dañaría la capa. Además, los materiales deben de ser de gran pureza. Dentro de las ventajas de este método se encuentran las siguientes:

- Control de espesores con muy alta precisión
- Baja temperatura de crecimiento
- Posibilidad de fabricar estructuras complicadas
-

Inconvenientes:

- Se deben de realizar en cámaras de vacío
- Existe una complejidad de instalación y uso
- Alto costo

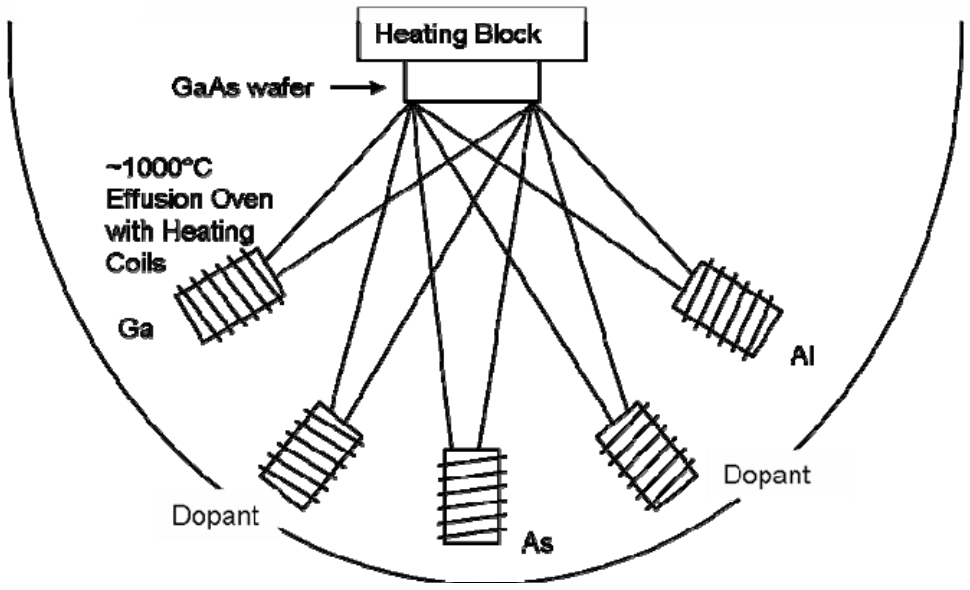


Figura III.33 Deposición epitaxial por haces moleculares

Presión Atmosférica APCVD

La técnica de APCVD es la más simple ya que no requiere el uso de vacío. Los gases se introducen en el reactor una vez caliente, normalmente produciendo una sobre presión con objeto de evitar el reflujo del aire de la atmósfera por la boca de salida de los gases (figura III.34).

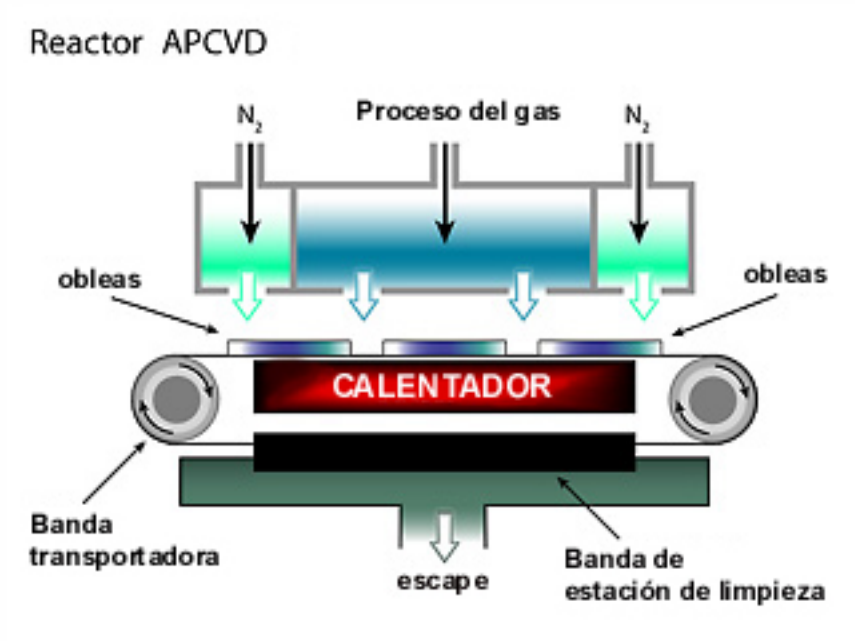


Figura III.34 Presión atmosférica APCVD

Previamente es preciso hacer un purgado de la atmósfera de aire del reactor mediante un barrido con algún gas inerte. Al no requerir equipo de evacuación, el equipamiento necesario para esta técnica se reduce al reactor con el horno correspondiente y al sistema de entrada y de control del flujo de gases. La velocidad de reacción puede ser bastante elevada sobre todo si la reacción ocurre a temperaturas altas, por lo que esta técnica se usa a menudo cuando se pretende obtener películas gruesas de un material. Para este tipo de aplicaciones tiene el inconveniente, sin embargo, de que la alta presión de los gases favorece la reacción en fase homogénea, produciendo una concentración elevada de partículas sobre el substrato y, a su vez, defectos en el recubrimiento. La homogeneidad del espesor también puede ser un problema, sobre todo en los puntos de difícil acceso a los gases reactivos.

Presión Reducida LPCVD

Durante los procesos de formación de películas delgadas por CVD, la velocidad de deposición de las películas se encuentra determinada fundamentalmente por la difusión a través de la capa que rodea el sustrato (capa límite) y la reacción en superficie (figura III.35).

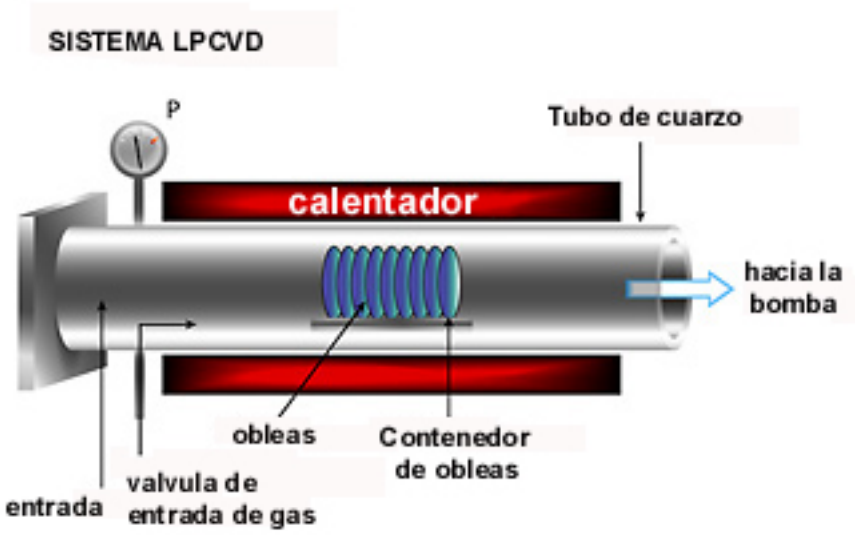


Figura III.35 Presión reducida LPCVD

En la técnica de APCVD ambos procesos presentan velocidades similares. Sin embargo, para una temperatura dada, cuando se baja la presión la velocidad de difusión aumenta notablemente por lo que es la reacción en la superficie la que determina la cinética de la reacción de formación de la película. Al mismo tiempo mejora también la uniformidad del espesor de las capas. De esta forma, es posible un control más preciso del proceso mediante la elección adecuada de los parámetros experimentales. El trabajo a presión reducida permite además la deposición de un gran número de muestras en un solo experimento, colocando las muestras muy próximas entre sí, sin pérdida de la homogeneidad de espesor. Esto hace que la técnica de CVD a baja presión resulte muy económica, y de ahí que se haya extendido en la industria electrónica en procesos de deposición de materiales aislantes, de silicio amorfo y policristalino y de metales refractarios y siliciuros.

Asistido por Plasma PECVD

En recubrimientos del tipo PECVD, el desarrollo de las distintas variantes técnicas se ha centrado en una buena difusión de los gases y el sistema generador de plasma. En los reactores industriales, el plasma es generado por RF (Radio Frecuencia) o por diferencias de potencial eléctrico, de polaridad fija o variable (plasma pulsante) (figura III.36).

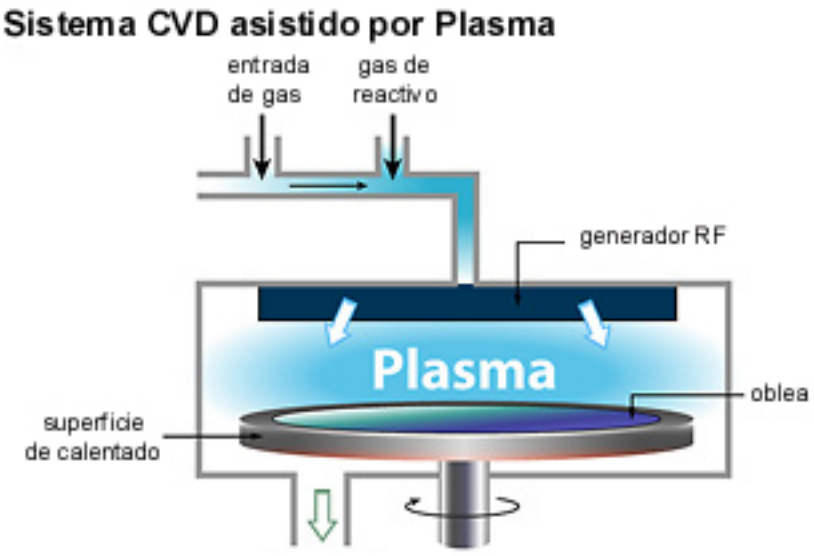


Figura III.36 Asistido por plasma PECVD

El plasma generado por RF permite recubrir materiales no conductores. Los gases reactivos y las piezas son calentados a la temperatura del proceso tras la realización del vacío previo. Se forma el compuesto y se depositan sobre las obleas. En piezas conductoras se focalizan los iones mediante un campo eléctrico generado por una diferencia de potencial negativo (bias) o variable. En algunos casos para conseguir la homogeneidad del recubrimiento en todas las piezas, la carga gira sobre un sistema de traslación y rotación planetario. Acabado el proceso, las piezas se enfrían en atmósfera inerte.

III.3.5 CMP

Pulido Mecánico Químico CMP (*Chemical Mechanical Planarization*) es un proceso especializado, el cual remueve material no deseado y crea una superficie plana en la oblea. Esta superficie plana permite extender la capacidad de procesamiento al reducir la topografía variable que existe en las capas. Durante el proceso de fabricación ciertos pasos de grabado en seco asistido con plasma han sido remplazados con mecanismo de pulido usados en el proceso de CMP. Figura III.37.

Este proceso es el responsable de alisar una capa desigual de material o de quitar una película de la superficie de la oblea. Por ejemplo, una capa gruesa de BPSG es depositada en la oblea antes de la foto resina de la máscara 41. Debido a las geometrías muy juntas de los contactos auto alineados (SACO), la capa de BPSG debe ser extremadamente plana (planar) para asegurar la consistencia del grabado del diseño a través de la oblea. Es también esencial tener una superficie lisa antes de que la capa de metal sea aplicada en los niveles de máscara 71 y 72. El propósito del proceso del CMP y del tungsteno se diferencia muy poco del proceso descrito anteriormente. El titanio, el nitruro del Ti (CVD) y el tungsteno se depositan sobre la oblea para llenar el nivel 60 (contactos). Después de que se formen los contactos con el tungsteno, el exceso del material en la superficie de la oblea debe ser eliminado ya sea por grabado seco o por el proceso de CMP.

El proceso de CMP con el tungsteno es responsable de pulir todo el exceso de material (Ti, TiN y Tungsteno) sobre la superficie de la oblea sin dañar el conector de tungsteno. Se debe de mantener un control riguroso sobre este proceso para asegurarse, de que la mezcla no quede atrapada por debajo y por dentro de los huecos formados en los contactos del tungsteno llamados (“cerraduras"). Los contenedores de celdas de los tipos de partes también utilizan esta tecnología para ayudar a construir el nodo del almacenaje del capacitor. El proceso de CMP pule y elimina el material que se encuentra en la parte baja de la celda-placa sobre la superficie de la oblea dejando solamente el material que se encuentra dentro del contacto en la máscara 42.

Para comenzar el proceso de pulido, las obleas entrantes que se encuentran dentro de su contenedor, se colocan boca abajo en el elevador de envío. Se toman dos obleas del contenedor y se colocan en el punto inicial de carga del equipo. El puente que contiene dos grandes rotores (dependiendo del equipo) se mueve sobre el punto de carga. Los rotores recogen las obleas de la parte trasera creando un vacío. Posteriormente el puente se posiciona sobre la placa giratoria, la cual tiene una superficie áspera para realizar el pulido primario.

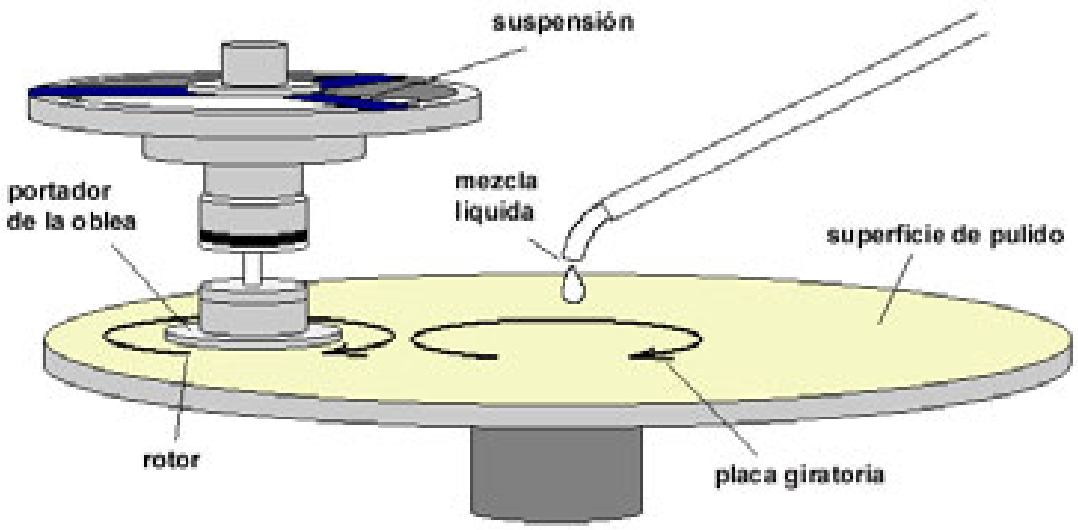


Figura III.37 Proceso de CMP

Los rotores hacen girar las obleas al mismo tiempo que las presionan boca abajo sobre la superficie giratoria que es rociada constantemente con una mezcla liquida (solución que contiene gránulos de silicio). El puente se mueve hacia adelante y hacia atrás sobre la superficie giratoria para producir un grabado mecánico. La mezcla afecta químicamente al material en la superficie de la oblea para permitir que se obtenga una superficie lisa durante este proceso mecánico.

Las obleas entonces pasan a través de un proceso secundario de pulido, para eliminar surcos o rasguños producidos sobre la superficie. Cuando las obleas han terminado con el proceso de pulido, se rocían y se limpian antes de ser sumergidas en agua dentro de los contenedores procesados. Para eliminar cualquier mezcla residual sobre la oblea, existen dos paso de limpieza importantes (limpieza por HF y/o limpieza por

tallado) que son realizados posteriormente al proceso del CMP. Las obleas contaminadas con mezcla seca en la superficie o en la parte de contacto son desechadas. Dentro de este proceso de tratamiento mojado, se debe de realizar una inspección que compruebe la existencia de mezcla residual y la medición topográfica de la superficie de la oblea. La uniformidad del proceso de CMP es crucial para evitar cortocircuitos entre las capas.

III.3.6 Metal

Este proceso forma los “caminos” en, a través, y fuera del chip, creando las trayectorias de comunicación para que el semiconductor funcione. Las interconexiones entre los dispositivos activos son estructuras muy importantes en circuitos integrados. Las líneas de la interconexión están exclusivamente hechas (a excepción de algunas interconexiones locales donde el silicio policristalino puede ser utilizado) de metal o compuestos de metal intercalados en capas con dióxido de silicio, o un dieléctrico de más baja permitividad. Una capa de metal, que consiste de varios elementos tales como aluminio, cobre, tungsteno y titanio, se deposita sobre la oblea usando procesos químicos o físicos de deposición de vapor.

Originalmente, eran utilizados los métodos de evaporación, pero estos métodos han sido desplazados en las tecnologías modernas. Esto fue resultado de la existencia de problemas durante la formación de la aleación y la cobertura de bajo nivel, que previene que los métodos de la evaporación puedan fabricar buenos contactos en líneas.

La oblea ya terminada simula la forma en como es construido un edificio, formada por varias regiones dopadas de silicio y de capas adicionales con diseños de aisladores y de conductores. Cuando el proceso de fabricación esta terminado, las obleas son enviadas a PROBE, en donde cada chip recibe las primeras pruebas de funcionalidad.

III.4 Back-End

III.4.1Probe

En el área de Probe, se realiza la primera de varias pruebas que cada componente debe pasar antes de ser enviado a un cliente. Para probar el funcionamiento del chip, se coloca una tarjeta de prueba sobre cada oblea (Figura III.38). Los contactos de la tarjeta de prueba se colocan sobre la superficie de la oblea justo en área correspondiente de prueba de cada chip.

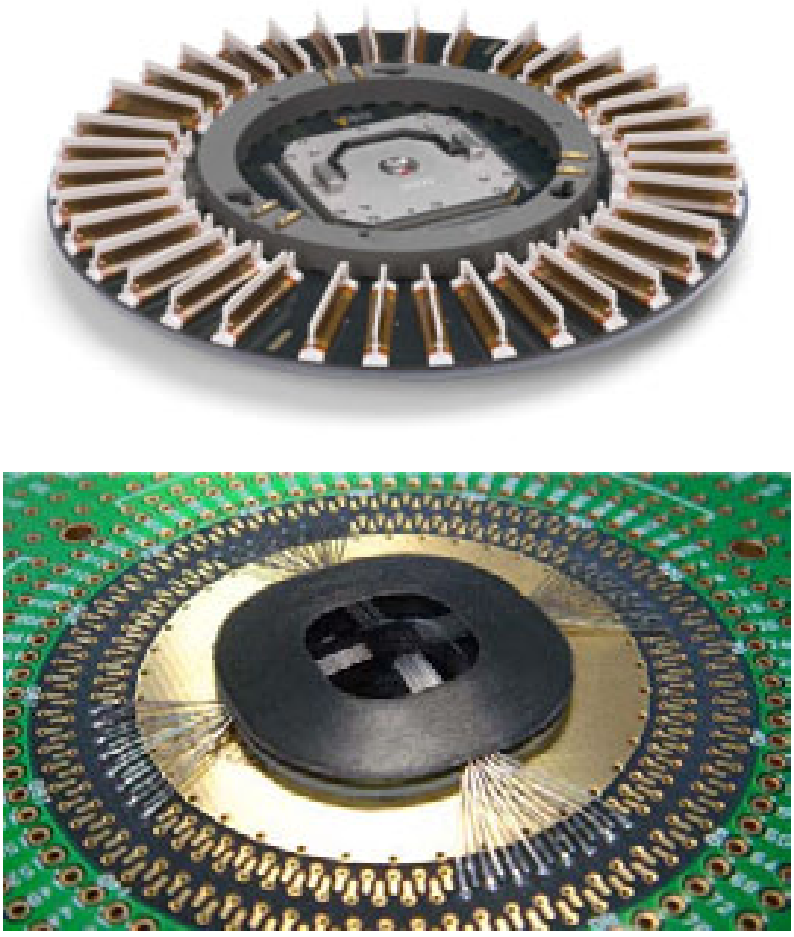


Figura III.38 Tarjetas de prueba

Estos contactos permiten a un equipo controlado por una computadora comunicarse con el chip para verificar que trabaje correctamente. La información obtenida es almacenada en una base de datos. De la cual posteriormente se genera en la computadora un mapa de la oblea mostrando el resultado de cada uno de los chips.

En el área de prueba se realizan también las pruebas paramétricas, las cuales permiten definir las variaciones en el proceso de fabricación y para comprobar la confiabilidad de cada una de las piezas. Este tipo de pruebas permiten el ahorro en el costo de empaquetamiento (packaging) y en Test, ayudan a aumentar la funcionalidad permitiendo la reparación de algunos chips que no funcionan y generar datos importantes para el área de diseño, ingenieros de producto y análisis de la funcionalidad. Después de que las obleas son probadas, se envían a área de ensamblado. En algunos casos, los chips y las obleas son clasificados y enviados a clientes externos que empaquetan el chip siguiendo sus propias especificaciones.

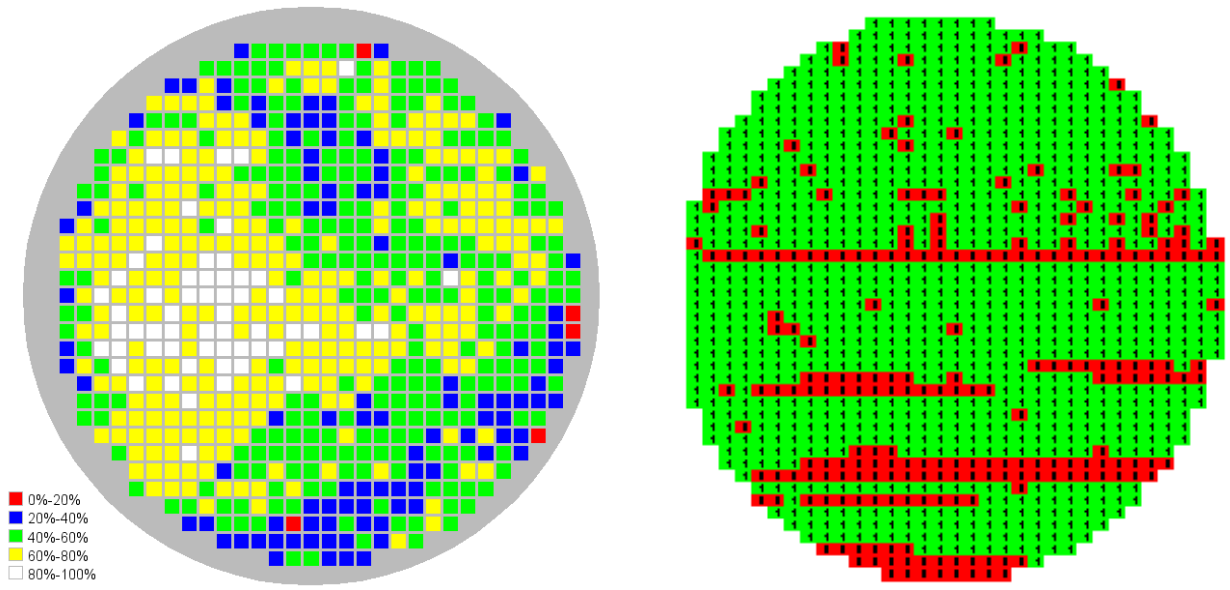


Figura III.39 Mapas de la oblea

III.4.2 Assembly

Pasos de ensamblado

En el área de ensamblado, se cortan las obleas en donde los chips buenos que son separan del resto son empaquetados y preparados para la prueba final.

Backgrind

El grueso de la oblea se reduce puliendo la cara inferior, esto es para proporcionar una superficie limpia, uniforme y un grueso específico del producto (Figura III.40). El proceso de Backgrind es capaz de reducir aproximadamente el grueso de cada oblea por debajo de las 75 micras -- el grueso promedio de un cabello humano. Los procesos de Backgrind y CMP son similares, pero la diferencia esta en que material es eliminando de la parte trasera de la oblea.

Figura III.40 Proceso de Backgrind

Montaje de la oblea

En proceso de montaje se adhiere la oblea por la parte posterior a un marco de metal utilizando una cinta adhesiva sensible a los rayos ultravioleta (Figura III.41). Esta cinta mantiene la oblea y los chips en el lugar durante los procesos subsecuentes. Esta cinta sostiene al chip en su lugar durante el proceso de corte, esto es para asegurarse de que solamente sea cortada la línea del marco del chip.

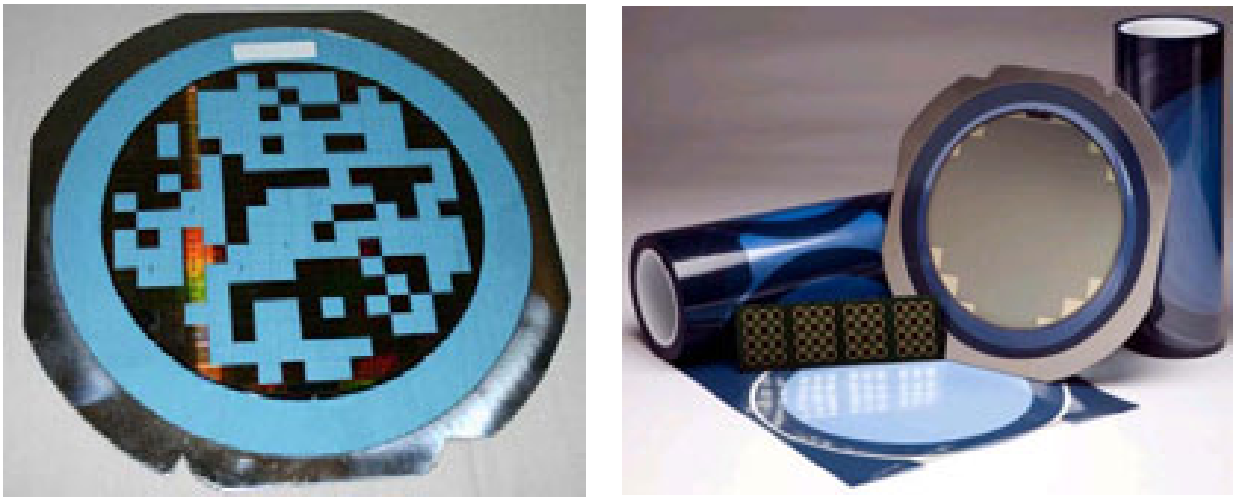


Figura III.41 Ejemplo del montaje de la oblea

Corte de la Oblea

En este proceso de corte de la oblea, se utiliza una sierra de diamante de un grosor aproximadamente al de un cabello humano. La sierra gira a 55,000 revoluciones por minuto y corta a una velocidad de 8cm por segundo (Figura III.42). Durante el proceso de corte, se rocía agua en la sierra y la oblea para mantener la temperatura baja. Después de que se corta la oblea, se lavan con agua a alta presión.

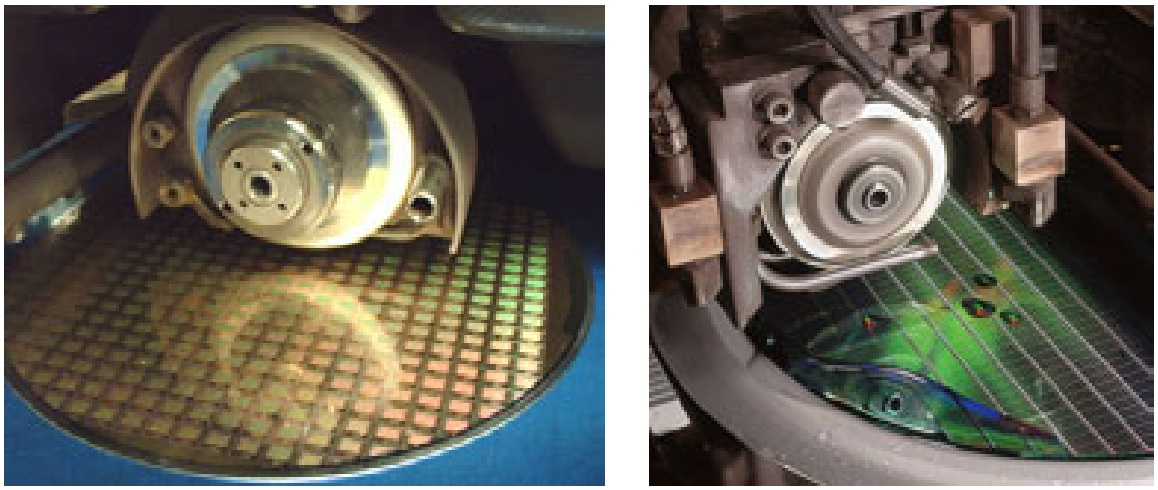


Figura III.42 Corte de la oblea

Tratamiento UV

Después de haber sido cortada la oblea, se expone a la luz ultravioleta de 5 a 10 segundos. Durante la exposición, ocurre una reacción química, y la adherencia de la cinta disminuye, lo cual permite separar más fácilmente el chip de la cinta adhesiva durante el proceso de fijación de chip.

Fijación del Chip

Usando el mapa de la oblea creado en Probe, se identifican los chips que funcionan, se separan de la oblea y se colocan en un marco a una velocidad de hasta 2,500 chips por hora. Para separar el chip de la cinta, pequeñas agujas empujan el chip por debajo de la cinta mientras que un brazo mecánico, (utilizando una bomba de vacío) despega el chip de la cinta (Figura III.43). El chip es adherido al marco y calentado en un horno.

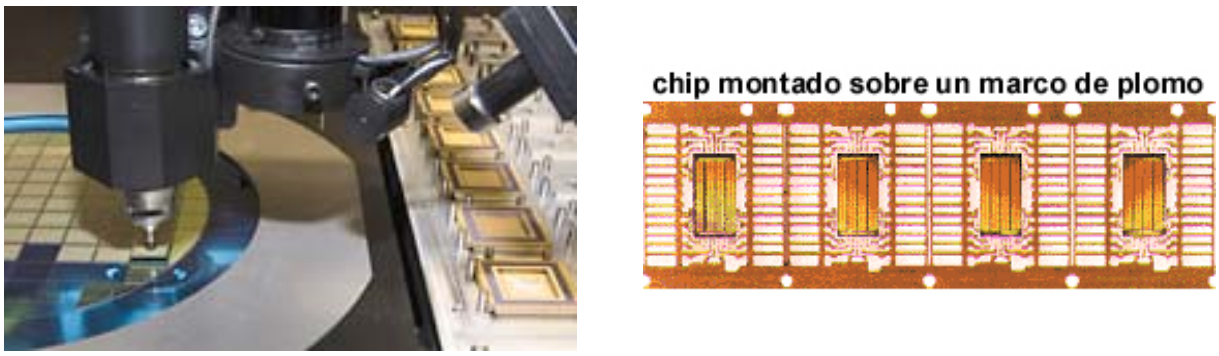


Figura III.43 Montaje de los chips en un marco de plomo

Soldadura de hilo

En el proceso de soldadura de hilo, se utiliza alambre de oro 99.999 por ciento puro y de un grosor menor al del cabello humano (Figura III.44). Este alambre proporciona la comunicación (conexión del circuito) entre el chip y otros componentes eléctricos. Otra técnica utilizada en este proceso, es la unión por vibración ultrasónica, que combina energía, calor, y fuerza ultrasónica capaz de interconectar los contactos del chip con los contactos del marco.

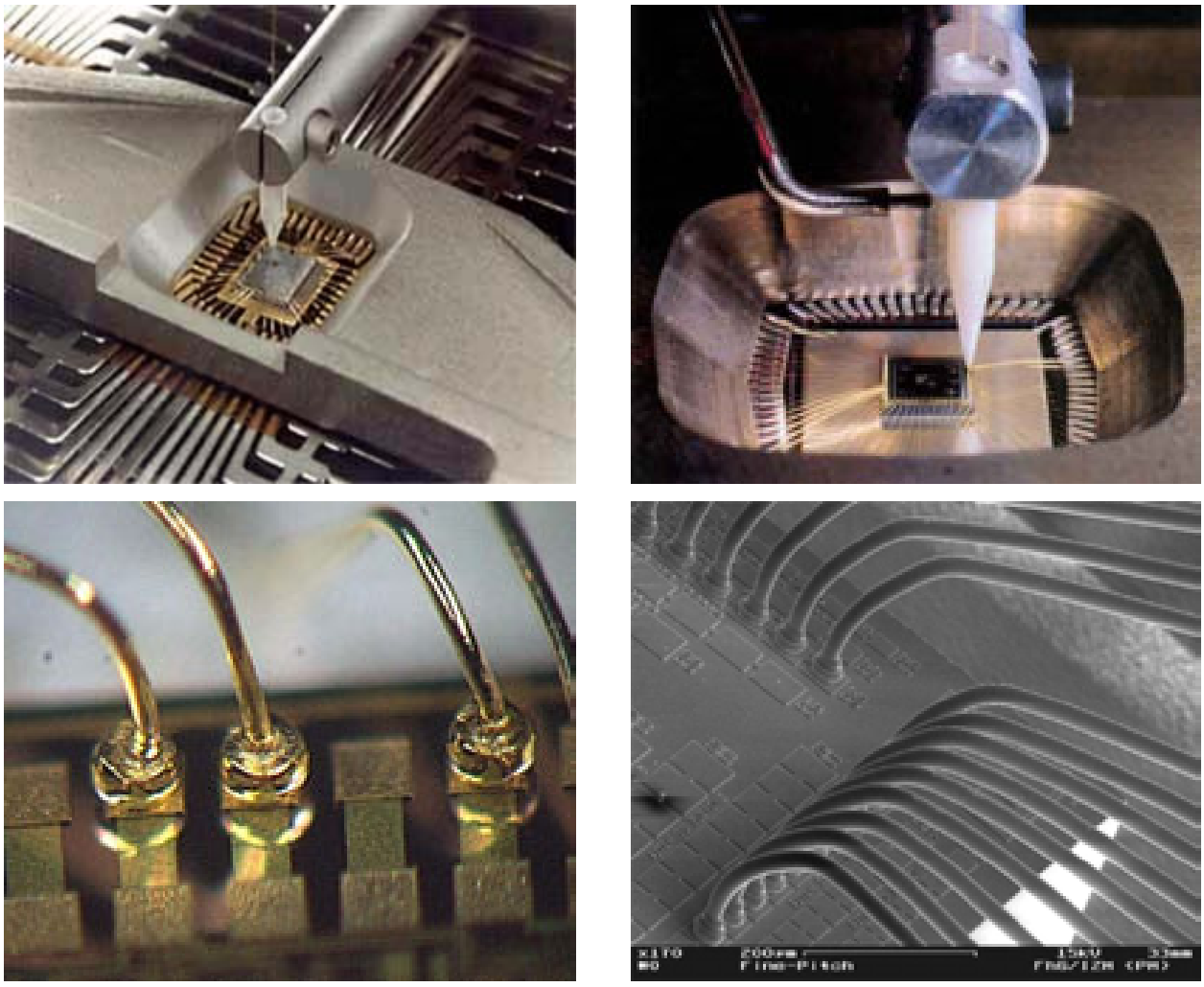


Figura III.44 Ejemplos de soldadura con hilos de oro

Encapsulado

El objetivo esencial del encapsulado es acomodar un circuito integrado en una montura que cumpla los requisitos eléctricos, térmicos, químicos y físicos asociados a la aplicación del circuito integrado (Figura III.45 y Figura III.46). Las monturas más extendidas son la de conductores radiales, la montura plana y la doble en línea (DIP). Las monturas de conductores radiales se fabrican casi todas de Kovar, que es una aleación de hierro, níquel y cobalto, con sellos de vidrio duro y conductores de Kovar. Las monturas planas tienen un marco de conductores metálicos, por lo general de una aleación de aluminio combinada con componentes cerámicos, de vidrio y metálicos.

Las monturas dobles en línea son las más corrientes y a menudo utilizan cerámica o plásticos moldeados. Las monturas de plástico moldeado para semiconductores se producen sobre todo por dos procesos diferentes—moldeo por transferencia y moldeo por inyección.

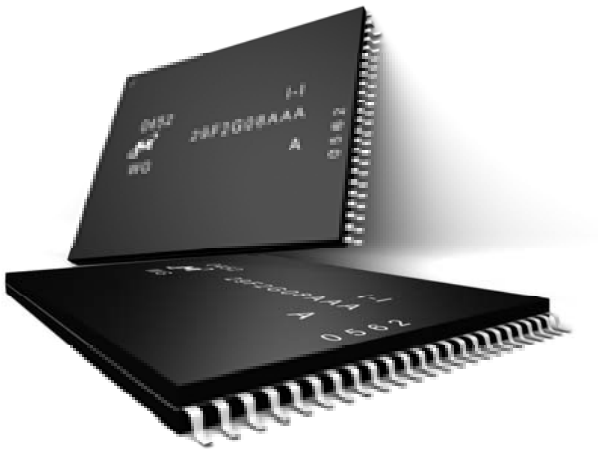


Figura III.45 Ejemplo de encapsulado

El moldeo por transferencia es el método de encapsulado en plástico predominante. En él, los chips se montan sobre marcos de conductores sin cortar y después se cargan en moldes por lotes. Porciones en polvo o bolas de compuestos de plástico termoendurecible para moldeo se funden en una olla caliente y después son impulsados (transferidos) a presión hasta los moldes cargados. Los sistemas de formar porciones en polvo o bolas con compuestos de plástico para moldeo pueden utilizarse con epoxia, silicona o silicona/resinas epoxídicas. El sistema suele consistir en una mezcla de:

- resinas termoendurecibles—epoxia, silicona o silicona/epoxia;
- endurecedores—novolacas epoxídicas y anhídridos epoxídicos;
- sustancias de relleno— dióxido de silicio (SiO_2) amalgamado en sílice o cristalino y alúmina (Al_2O_3), por lo general en la proporción de 50-70 % en peso;
- pirorretardante—trióxido de antimonio (Sb_2O_3) por lo general en la proporción de 1-5 % en peso.

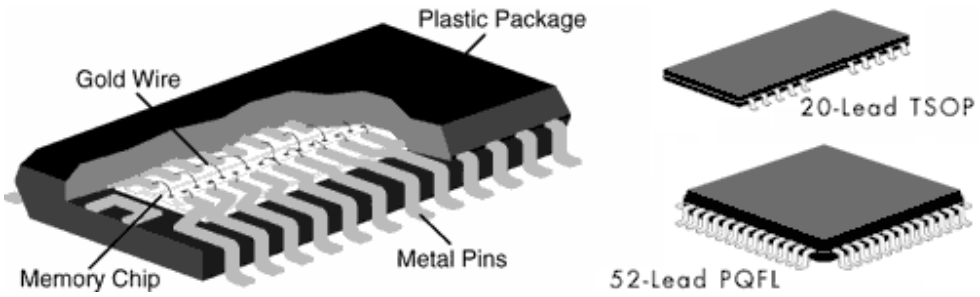


Figura III.46 Ejemplo de encapsulado

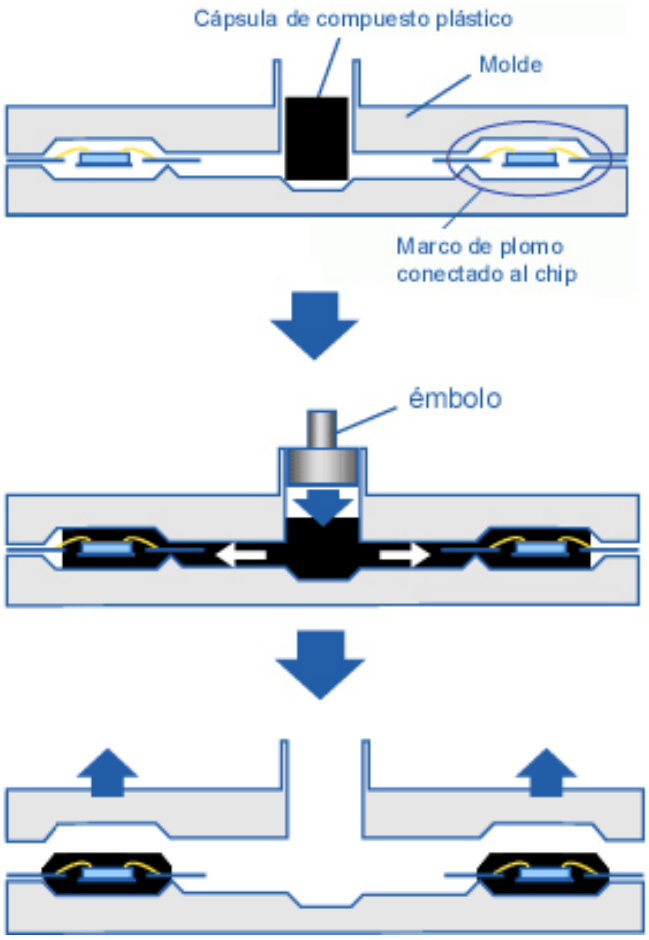


Figura III.47 Proceso de encapsulado

El moldeo por inyección emplea un compuesto termoplástico o termoendurecible para moldeo que se calienta hasta su punto de fusión en una botella a temperatura controlada y se hace pasar a presión por una boquilla hasta el molde. La resina se solidifica en seguida, se abre el molde y la montura encapsulada sale expulsada (Figura III.47). En el moldeo por inyección se utiliza una extensa variedad de compuestos de plástico. Las resinas epoxídicas y a base de sulfuro de polifenileno (PPS) son las últimas sustancias que se han incorporado al encapsulado de semiconductores. Los encapsulados finales de dispositivos semiconductores de silicio se clasifican en función de su resistencia a las fugas o capacidad de aislar el circuito integrado de su medio ambiente. Se distingue entre el sellado hermético (estanco al aire) y el no hermético.

Terminación de plomo o solder-ball

El producto ahora debe pasar por un proceso de fijación a una placa o al proceso de solder-ball. En el proceso de fijación a una placa, el metal expuesto en el marco de plomo se cubre con una capa de metal conductora. Mientras que están sumergidos en una solución de 95% de estaño y 5% de plomo, los marcos se cargan de energía para atraer los iones del estaño y del plomo. Esto da como resultado la creación de una capa uniforme que aumenta la conductividad, protegiendo al plomo de la corrosión y proporcionando un acabado limpio y parejo de la superficie de la placa o modulo.

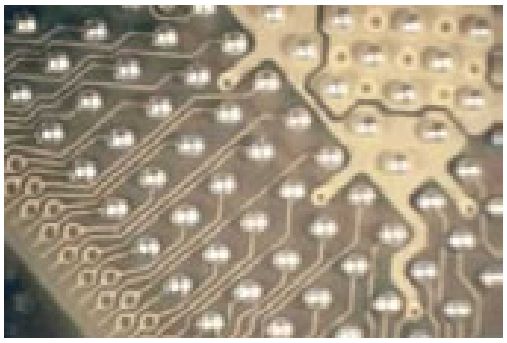


Figura III.48 Solder-ball

Durante el proceso de solder-ball, las esferas son colocadas en pequeñas placas de oro situadas en el substrato (Figura III.48). Cuando se le aplica calor a la pieza, las pequeñas esferas se adhieren a placa. Las terminales de plomo o las esferas proporcionan la interconexión final entre el componente y el producto la aplicación final (Figura III.49).

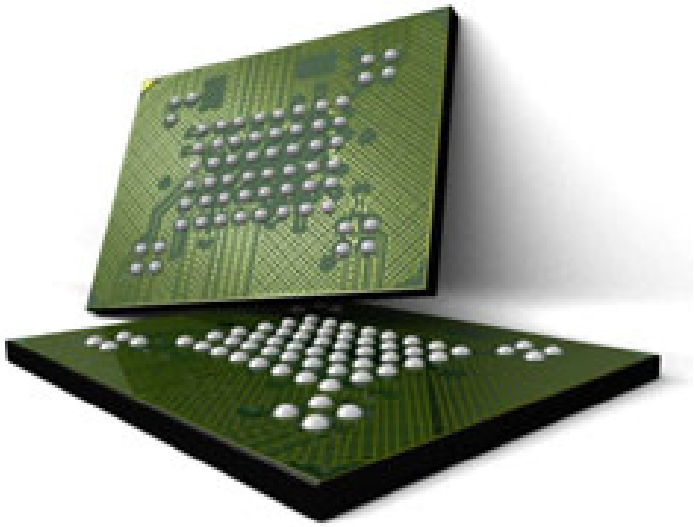


Figura III.49 Ejemplo de un producto utilizando terminales del tipo solder-ball

Terminado y forma

Después de que el marco de plomo es recubierto, el equipo utilizado en el proceso de terminado y forma, corta los paquetes individuales de los marcos y forma las terminales en cada dispositivo. A cada uno los chips ya empaquetados, se les realiza una prueba en busca de cortocircuitos y posteriormente son colocados en charolas o tubos, siendo clasificados en buenos o defectuosos.

Igualmente los chips del tipo solder-ball, después de que han sido soldadas las esferas, cada uno de ellos son sometidos apruebas en busca de cortocircuitos, se examinan las dimensiones del empaquetado y la presencia de todas las esferas de contacto. Posteriormente son clasificadas y colocadas en charolas de la misma forma anteriormente descrita.

III.4.3 Test y Producto final

Test

Es la última etapa en el proceso de manufactura. En el área de Test, se verifica la funcionalidad y velocidad de cada uno de los chips para garantizar que sean productos de alta calidad y que cumplan con los estándares requeridos por la industria. Durante este proceso el producto es marcado, inspeccionado (defectos físicos) y empaquetado antes de ser enviado. En el área de Test varios procesos son utilizados para verificar que los chips terminados cumplan con el rendimiento especificado en el diseño. Esto asegura el envió de productos de calidad a los clientes.

Procesos Eléctricos

Calentamiento

Los chips se prueban bajo diversos voltajes y temperaturas para determinar su calidad, funcionalidad a largo plazo y para eliminar chips potencialmente defectuosos. Esto ayuda a garantizar la funcionalidad de los chips que el cliente recibe y de reducir el tiempo del proceso de prueba y costos de producción.

Test área

En el área de Test, a cada chip se le realiza una prueba de funcionalidad, de velocidad de almacenamiento y lectura de información.

Procesos Post-Eléctricos

Marcado

Cada chip se marca con información que lo identifica. Los chips pasan debajo de un rayo láser, que graba la numeración del componente y la fecha sobre la superficie del chip.

Escaneado

Cada chip es escaneado para verificar que el paquete cumpla con los estándares de la industria. Lentes o rayos láser examinan los chips para identificar defectos.

Empaquetado

Los chips son preparados para el envío según las necesidades del cliente. Por ejemplo, los clientes que ponen los chips en tarjetas o módulos prefieren que el producto se coloque en paquetes equidistantes en una cinta y posteriormente enrollada en un carrete, esto es para facilitar el montaje automatizado del producto. Otros clientes prefieren que los chips sean empaquetados en tubos o charolas antiestáticas.

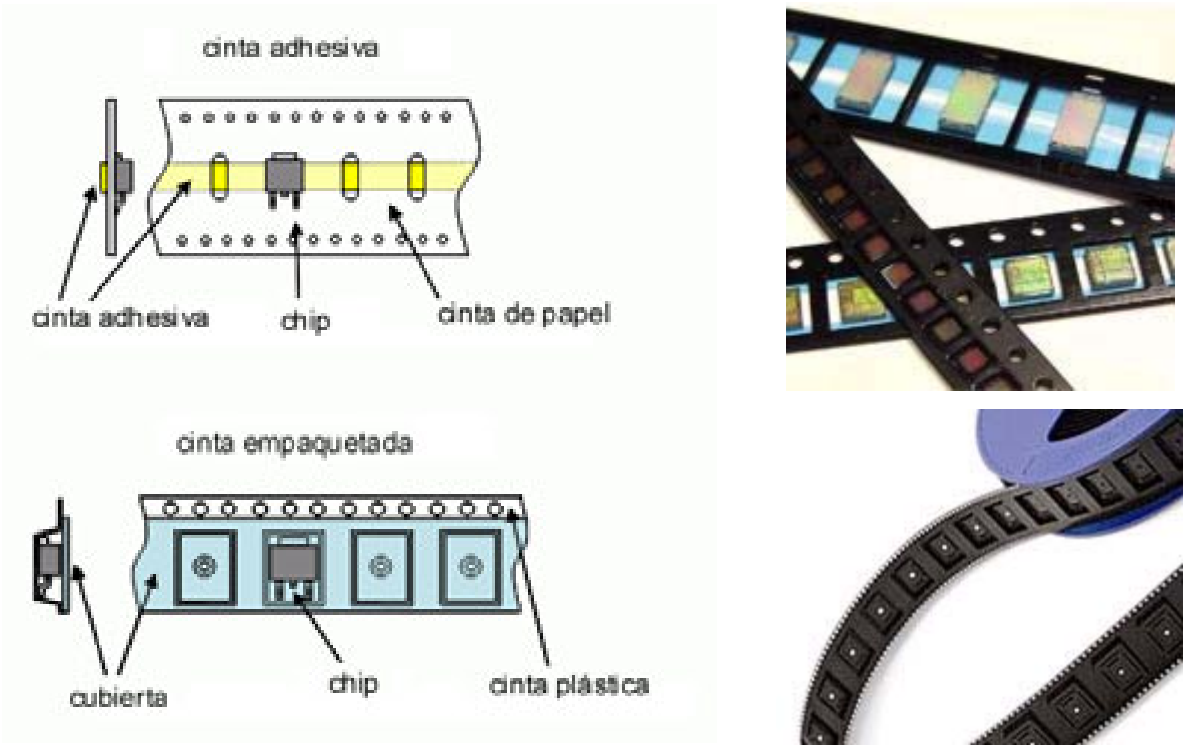


Figura III.50 Ejemplo de empaquetado en carrete

PARTE III. APLICACIÓN

CAPÍTULO IV
ANTECEDENTES DE LA TELEFONÍA CELULAR

Las comunicaciones celulares han experimentado un crecimiento explosivo en las últimas dos décadas. Hoy en día, millones de personas alrededor del mundo usan teléfonos celulares, los cuales permiten llamar o recibir una llamada en casi cualquier parte del mundo. La comunicación celular es soportada por una infraestructura llamada red celular, la cual integra teléfonos celulares dentro de la red de telefonía pública conmutada.

Los sistemas celulares han evolucionado en el tiempo por generaciones y cada una se caracteriza principalmente por la forma en como el dispositivo móvil accede a la red celular (esto es, técnica de acceso al medio compartido) y por las aplicaciones que ofrece a los usuarios la operadora celular (por ejemplo, servicios de voz, mensajería, Internet, etc.). Siendo así que la primera generación desarrollada (1G) fue puramente analógica y a fin de incrementar la capacidad de la red y añadir más usuarios, las tecnologías TDMA (**Time Division Multiple Access**) y CDMA (**Code Division Multiple Access**) se usaron en la segunda generación (2G). Ya con tecnologías digitales, la voz podía ser digitalizada y con ello podía ser codificada y cifrada, lo cual se aprovecha para que la tercera generación (3G) integre en los teléfonos celulares la transmisión de datos de alta velocidad mediante la conmutación de paquetes, además de la transmisión de voz por conmutación de circuitos. En la actualidad los operadores más avanzados se encuentran en proceso de migración de sus redes tecnológicas de tercera generación, y ya se habla del desarrollo de la generación 4G.

IV.1 Conceptos básicos

Red celular

Una red celular proporciona a los teléfonos celulares o también llamados estaciones móviles (MS -**Mobile Stations**), el acceso inalámbrico a la red publica telefónica conmutada (PSTN - **Public Switched Telephone Network)**.

El área de cobertura de la red celular se divide en muchas áreas pequeñas, llamadas células o celdas, las cuales son atendidas por estaciones base (BS – **Base Station**). Las BS son fijas y están conectadas a la oficina de telefonía móvil conmutada (MTSO- **Mobile Telephone Switching Office**), también conocida como centro de conmutación móvil (MSC- **Mobile Switching Center**). Un MTSO está a cargo de varias estaciones base y alternadamente está conectado a la PSTN y a través de este el vínculo, las estaciones móviles (tales como teléfonos celulares), pueden comunicarse con teléfonos convencionales.

Por otra parte, tanto las estaciones base, como las estaciones móviles están equipados con un transceiver (transmisor-receptor) para proveer una comunicación full duplex. La Figura IV.1 ilustra una típica red celular.

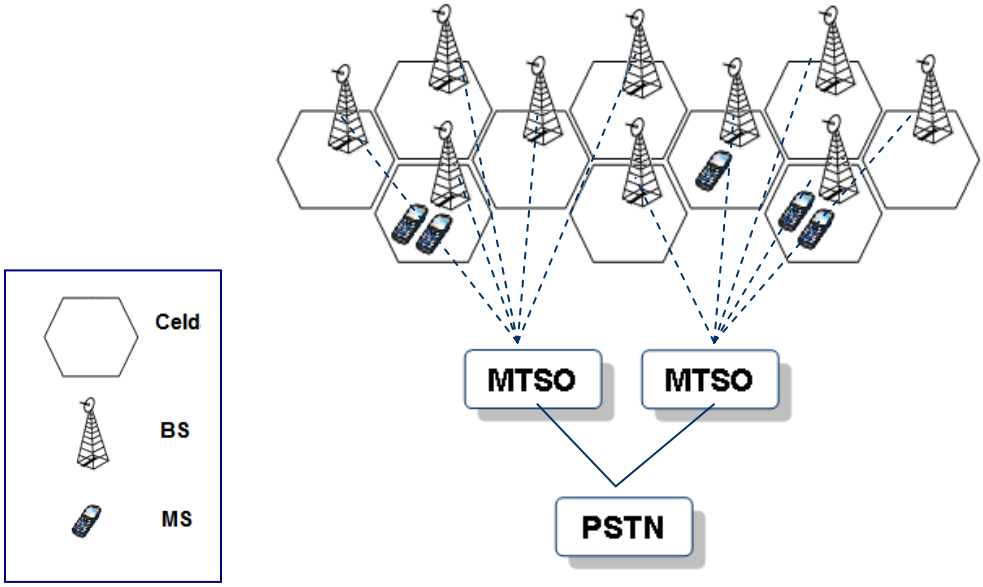


Figura IV.1. Red celular

El espectro de frecuencia asignado para las comunicaciones es limitado, por lo que el éxito de la red celular, es principalmente los conceptos de reutilización de frecuencia y división de celdas, como se muestra en la Figura IV.2.

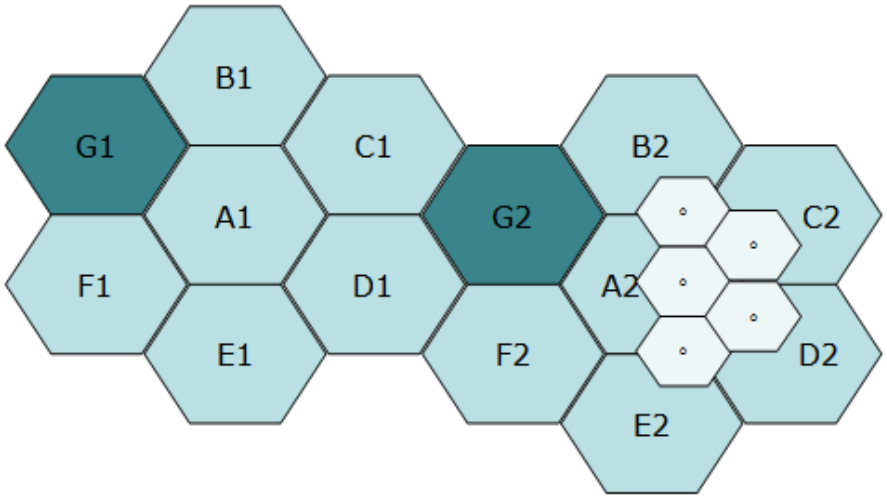


Figura IV.2. Reuso de frecuencias

Esta es la razón por la cual el área de cobertura se divide en celdas (A1, B1, A2, B2,…), cada una atendida por una BS. A cada celda se le asigna un conjunto de canales que son ortogonales al resto de canales en celdas adyacentes. Esta ortogonalidad entre canales garantiza que celdas adyacentes no causen interferencia entre sí (interferencia de radio co-canal). La manera como se obtiene esta ortogonalidad puede ser por: división de frecuencia, división de tiempo, o códigos de dispersión ortogonales (principio de Espectro Amplio- **Spread Spectrum**).

El número total de canales en una celda es proporcional al espectro radioeléctrico asignado al sistema y por tanto es finito. Entonces en algún punto del proceso de asignación de canales a celdas estos se acaban, y si el terreno no ha sido totalmente cubierto, se deben crear nuevas celdas que usen canales que han sido utilizados en celdas lejanas. Esta técnica, llamada reuso de frecuencias o reuso de códigos, se ilustra en la Figura IV.2 con las celdas A1y A2, las cuales usan el mismo conjunto de canales al igual que B1 y B2, C1 y C2 y así sucesivamente. La separación geográfica entre estas celdas debe garantizar que las señales de interferencia que percibe una de

las celdas proveniente de la otra hayan sido suficientemente atenuadas de modo que la interferencia co-canal sea mínima.

La división en celdas (**cell splitting**) más pequeñas permite que el sistema aumente su capacidad cuando la densidad de usuarios aumenta tanto que no puede ser satisfecha con las celdas originalmente planeadas. En esta técnica, el tamaño de la celda se reduce disminuyendo el tamaño de la antena y la potencia del transmisor. En la Figura IV.2 este proceso se muestra con las celdas pequeñas. Suponiendo que en la región de las celdas A2, C2 y D2 aumenta la densidad de usuarios y estas celdas no pueden atender esta demanda porque los canales asignados a las mismas están siendo usados por otros usuarios, la división de celdas creando las celdas pequeñas, permite atender un mayor número de usuarios ya que al cubrir una menor área ahora se deben atender menos usuarios. En realidad, las celdas más grandes no son sustituidas totalmente por celdas más pequeñas. Es por ello que existen celdas de diversos tamaños (ejemplo: pico, micro y macro celdas) y pueden coexistir en un área. Esto permite que los suscriptores de alta velocidad, utilicen celdas más grandes, que reducen el **handover** (Sistema utilizado para transferir servicio de una BS a otra cuando la calidad del enlace es insuficiente, con lo cual se garantiza el servicio cuando un MS se traslada a lo largo de su zona de cobertura). Como se muestra en la Figura IV.3.

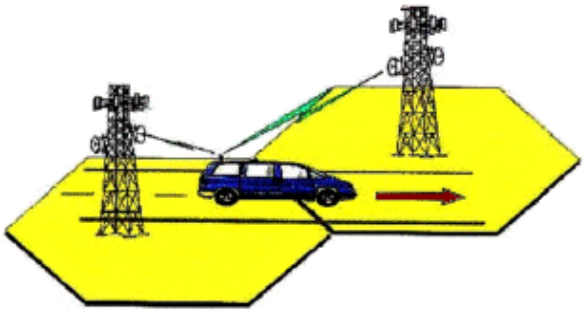


Figura IV.3 Manejo de Handover. Continuidad en cambio de celda

El seccionamiento es otra técnica para incrementar la capacidad de red. En el seccionamiento, el tamaño de la celda sigue siendo igual, pero una celda es dividida en varios sectores usando varias antenas direccionales en la BS, en lugar de una sola

antena omnidireccional. Típicamente una celda es divida en 3 sectores de 120° o seis sectores de 60°. La interferencia de radio co-canal se reduce dividiendo una celda en sectores, lo cual reduce el número de celdas en un grupo. Por lo tanto la capacidad de la red se incrementa.

Tecnologías de acceso celular

Dentro de la celda con cobertura de una BS, hay varias estaciones móviles que requieren comunicarse con la BS. Esas estaciones móviles deben compartir la interfaz de aire de manera ordenada, de modo que ninguna estación móvil dentro de la celda se interfiera. Los métodos para que los MSs compartan la interfaz de aire de una forma ordenada se especifica como tecnologías de acceso celular y en la actualidad existen tres tecnologías comúnmente usadas:

- Acceso Múltiple por División de Frecuencia (FDMA – Frecuency Division Multiple Access)
- Acceso Múltiple por División de Tiempo (TDMA – Time Division Multiple Access)
- Acceso Múltiple por División de Código (CDMA, Code Division Multiple Access)

La diferencia primordial entre estas tecnologías yace en el método de acceso, el cual varía entre frecuencia, tiempo y códigos.

La primera parte de los nombres de las tecnologías (Acceso múltiple) significa que más de un usuario (múltiple) puede usar (accesar) cada celda.

La tecnología FDMA separa el espectro en distintos canales de voz, al separar el ancho de banda en pedazos (frecuencias) uniformes, como se muestra en la Figura IV.4. Estos canales están separados y no interfieren con ningún otro.

La tecnología FDMA es mayormente utilizada para la transmisión analógica. Esta tecnología no es recomendada para transmisiones digitales, aun cuando es capaz de llevar información digital. FDMA es usado en el sistema AMPS (Advanced Mobile Phone System). El sistema AMPS utiliza un total de 40MHz en el espectro de 800MHz (825-845 MHz y 870-890 MHz). En AMPS, cada canal tiene un ancho de banda de

30kHz, lo cual genera cerca de 1332 canales. AMPS utiliza FDD (frequency division multiplexing) multiplexacion por división de frecuencia. Basado en el patrón de reuso de celdas, alrededor de 45MSs dentro de una celda pueden comunicarse con la BS simultáneamente.

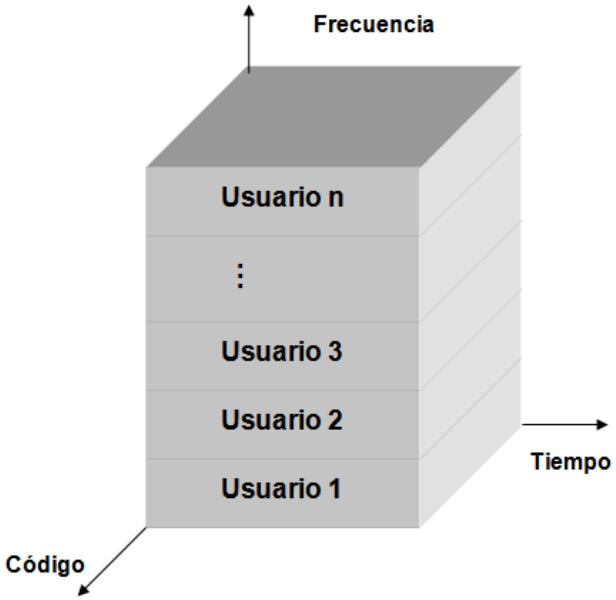


Figura IV.4. FDMA (Acceso múltiple por división de frecuencia)

La tecnología TDMA permite que múltiples estaciones móviles compartan el mismo canal. En TDMA, el tiempo se secciona en ranuras, en cada ranura de tiempo (time slot), solo un MS es permitido para utilizar el canal compartido para transmitir y recibir. Las MSs toman su turno para transmitir o recibir en sus ranuras asignadas en un proceso round-robin. Aunque se comparta el canal, ninguna interferencia puede presentarse entre ellos, ya que solo una MSs puede usar el canal en un tiempo a la vez. La figura IV.5 ilustra el concepto de TDMA.

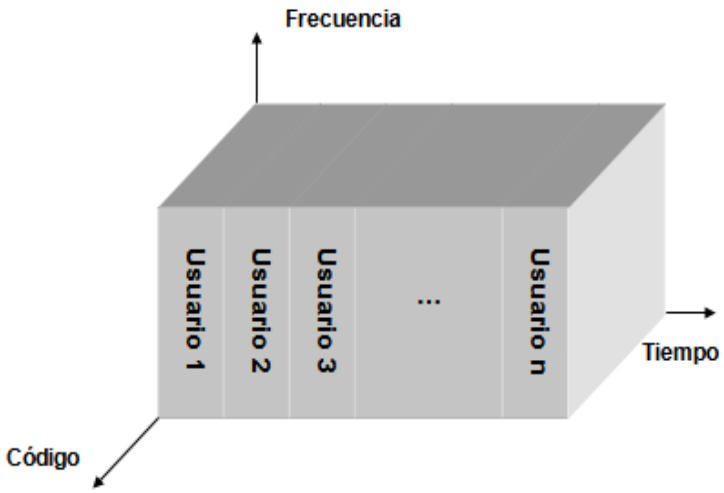


Figura IV.5. TDMA (Acceso múltiple por división de tiempo)

La tecnología TDMA comprime las conversaciones (digitales), y las envía cada una utilizando la señal de radio por un tercio de tiempo solamente. La compresión de la señal de voz es posible debido a que la información digital puede ser reducida de tamaño por ser información binaria. Debido a esta compresión, la tecnología TDMA tiene tres veces la capacidad de un sistema analógico que utiliza el mismo número de canales.

En CDMA, múltiples estaciones móviles comparten el mismo ancho de banda del espectro. En lugar de la asignación de ranuras de tiempo como TDMA, a cada MS se le asigna una secuencia de código único. La señal de cada estación móvil se extiende sobre todo el ancho de banda por la secuencia de código. En el receptor, ese mismo código único es usado para recuperar la señal. Sin embargo el espectro de radiofrecuencia es compartido y ninguna interferencia puede presentarse porque la secuencia de los códigos usados por las estaciones móviles presentan una distribución ortogonal. La Figura IV.6 ilustra el concepto de CDMA. Usando la tecnología CDMA es posible comprimir entre 8 y 10 llamadas digitales para que estas ocupen el mismo espacio que ocuparía una llamada en el sistema analógico.

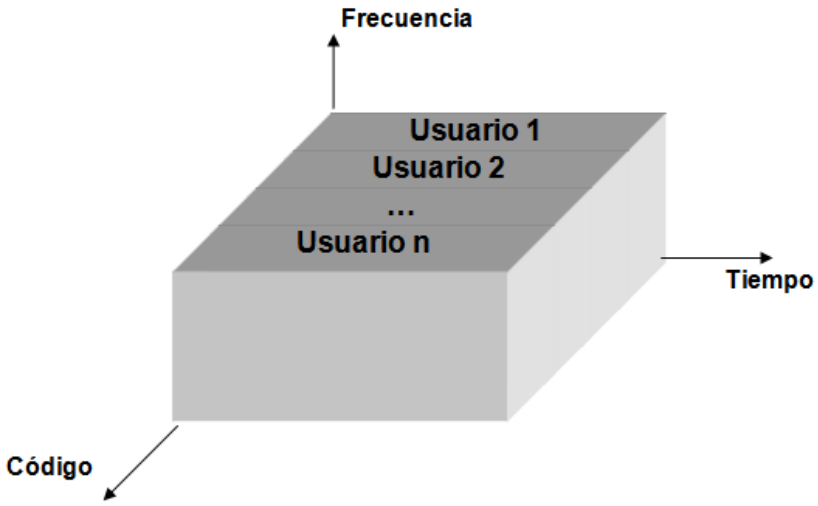


Figura IV.6. CDMA (Acceso múltiple por división de código)

IV.2 Arquitectura GSM

La red GSM (Sistema Global de comunicaciones Móviles) fue a comienzos del siglo XXI el estándar más usado por las operadoras en el mundo. Se denominó estándar de segunda generación (2G) porque a diferencia de la primera generación, las comunicaciones se producían de un modo completamente digital. En 1982 fue estandarizado por primera vez y se le denominó "**Groupe Special Mobile**" y en 1991 se convirtió en un estándar internacional llamado "Sistema Global de Comunicaciones Móviles".

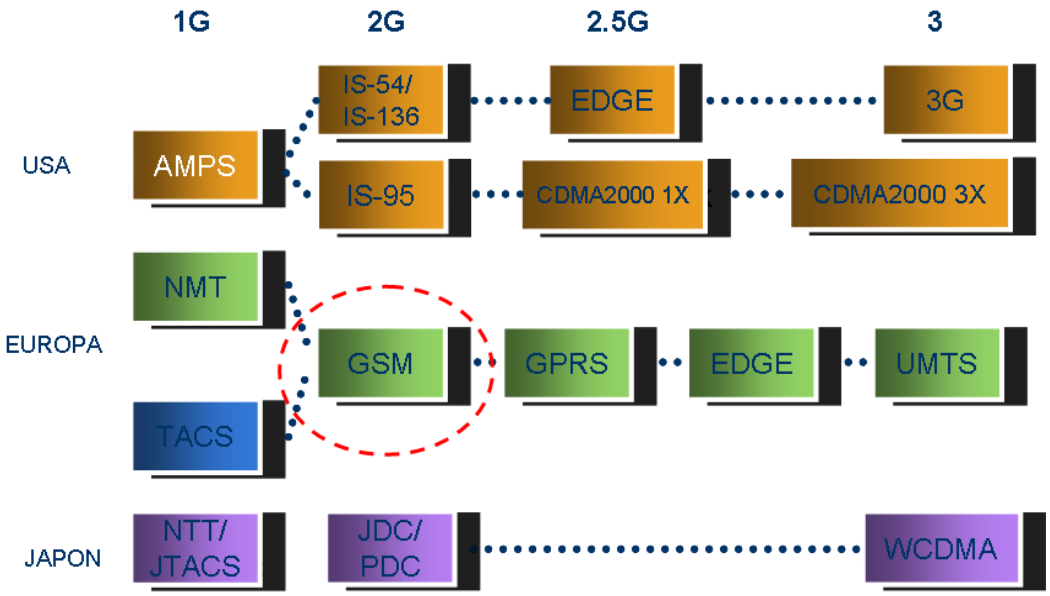


Figura IV.7. Evolución de las redes celulares

En esta arquitectura (Figura IV.8) una estación móvil se comunica con el sistema de estación base (BSS- **Base Station Subsytem**) a través de la interfaz de radio y el BSS está conectado al subsistema de red y conmutación (NSS – **Network and Switching Subsystem**) comunicándose por medio de un sistema de conmutación móvil (MSC).

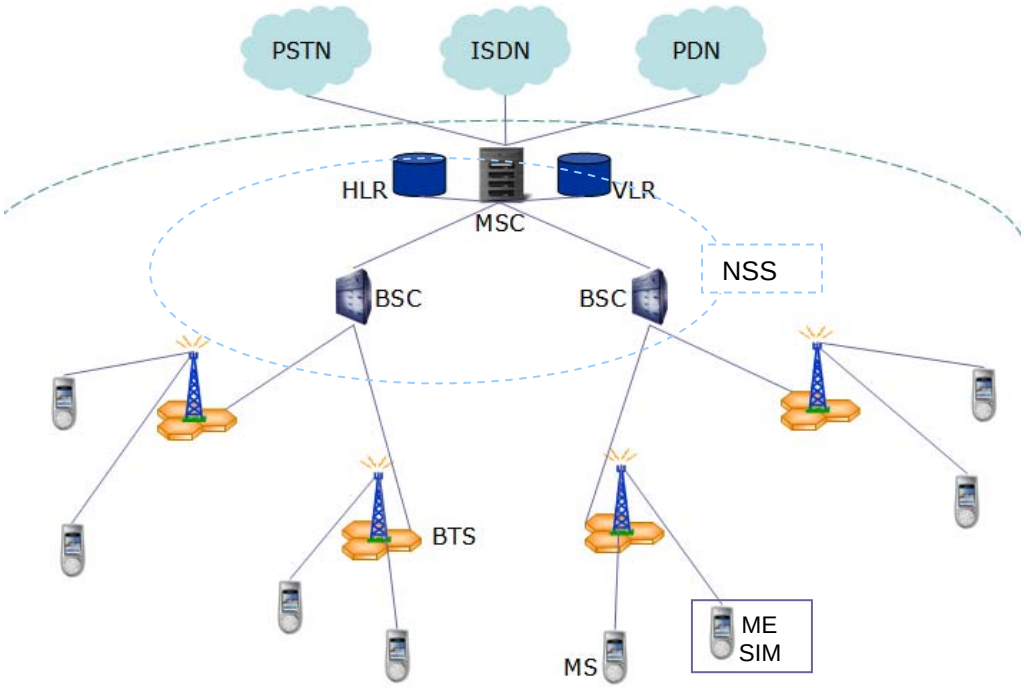


Figura IV.8. Arquitectura de red GSM

- La Estación Móvil (MS)

Está compuesta por dos partes: El módulo de identificación del suscriptor (SIM- **Subscriber Identity Module**) y el equipo móvil (ME - **Mobile Equipment**). La SIM permite identificar de manera única al usuario y al equipo móvil.

Las terminales se identifican por medio de un número único de identificación de 15 dígitos denominado IMEI (Identificador Internacional de Equipos Móviles – **International Mobile Equipment Identity**). Cada tarjeta SIM posee un número de identificación único y secreto denominado IMSI (Identificador Internacional de Abonados Móviles- **International Mobile Suscriber Identity**). Este código se puede proteger con una clave de 4 dígitos llamada código PIN.

La comunicación entre una estación móvil una estación base se producen a través de un vínculo de radio, por lo general denominado interfaz de aire (o interfaz Um).

- El Sistema de Estación Base (BSS)

Este subsistema conecta la estación móvil con el subsistema de red y conmutación.
Consta de dos partes:

- Estación Base -**Base Transceiver Station** (BTS)
- Controlador de Estaciones Base- **Base Station Controller** (BSC)

La BTS está compuesta por transmisores, receptores, equipo específico a la interfaz de radio para poder establecer contacto con las estaciones móviles. Una parte importante de la BTS es el TRAU (Transcoder/Rate Adapter Unit) que realiza la codificación y decodificación de la voz, y el control en las transmisiones de datos. El controlador de la estación base está a cargo de las funciones de conmutación en el BSS. El BSC puede conectarse a varios BTS’s, estas comunicaciones se hacen empleando el protocolo ISDN (Red Digital de Servicios Integrados - **Integrated Services Digital Network**).

- El subsistema de Red y Conmutación

Este subsistema soporta las funciones de conmutación y el manejo de la movilidad y de los perfiles de los usuarios.

La función básica de conmutación es realizada por el Centro de conmutación móvil (**Mobile Switching Center** MSC) a través de la Interfaz de aire. El cual está conectado físicamente a los controladores de estaciones base. El MSC pertenece a un Subsistema de conmutación de red (NSS) que gestiona las identidades de los usuarios, su ubicación y el establecimiento de comunicaciones con otros usuarios. La localización de una estación móvil está mantenida en un sistema jerárquico de dos niveles, que emplea diversas bases de datos que proporcionan funciones adicionales:

- Registro de ubicación de origen. **Location Registrer** (HLR): es una base de datos que contiene la información de los abonados registrados (posición geográfica, información administrativa, información del perfil, etc.) dentro de la zona del conmutador (MSC).
- Registro de ubicación de visitante. **Visitor Location Register** (VLR): es una base de datos que contiene información de usuarios que no son abonados locales. El VLR recupera los datos de un usuario nuevo del HLR de la zona de abonado del usuario. Los datos se conservan mientras el usuario está dentro de la zona y se eliminan en cuanto la abandona o después de un periodo de inactividad prolongado.
- Registro de Identificación del Equipo. **Equipment Identity Register** (EIR). Es una base de datos que contiene la lista de terminales móviles.
- Centro de autenticación. **Authentication Center** (AUC): verifica las identidades de los usuarios.

- Interfaz de radio

GSM utiliza dos bandas de 25 MHz para transmitir y para recibir. La banda de 890-915 MHz se usa para las transmisiones desde la MS hasta el BTS (uplink) y la banda de 935-960 MHz se usa para las transmisiones entre el BTS y la MS (downlink). GSM usa FDD (Duplex por división de Frecuencia) para proporcionar a las estaciones base y a

los usuarios un acceso múltiple. Las bandas de frecuencias superiores e inferiores se dividen en canales de 200 KHz llamados ARFCN (**Absolute Radio Frequency Channel Number** ó Números de Canales de Radio Frecuencia Absolutos). El ARFCN denota un par de canales "uplink" y "downlink" separados por 45 MHz y cada canal es compartido en el tiempo por hasta 8 usuarios usando TDMA.

Gracias a la estructura del paquete se pueden definir varios canales lógicos:

- Canales de Tráfico (Traffic Channels, TCHs), diseñados para transportar información del usuario (voz o datos).
    - Full Rate Traffic Channel (TCH/F) que proporciona transmisión de voz a 13 kbps o de datos a 6 kbps
    - Half Rate Traffic Channel (TCH/H) que proporciona transmisión de voz a 7 kbps o de datos a 3.6 kbps.
- Canales de Control (Control Channels, CHS) y Canales de Broadcast (Broadcast Channels BCH) que se encargan de transportar información de control:
    - Frecuency Correction Channel (FCCH) y Synchronization Channel (SCH)
    - Broadcast Control Channel (BCCH)
    - Paging Channel (PCH)
    - Access Grant Channel (AGCH)
    - Random Access Channel (RACH)

Todos estos canales son comunes, ya que son enviados por la red a todos los móviles. Adicional a estos canales, GSM soporta canales de control dedicados para ser utilizados por una estación móvil específica:

- Dedicated Control Channel (SDCCH) es empleado únicamente para señalización y mensajes cortos.
- Slow Associated Control Channel (SACCH) empleado para procedimientos que no son urgentes, principalmente para información del estado de la señal.
- Fast Associated Control Channel (FACCH) empleado para procedimientos como el establecimiento de las llamadas, autenticación de los usuarios, proceso de handover.

- Funcionalidad de la Red GSM

Localización de una estación móvil

El proceso de registro de un móvil, se realiza en 5 pasos:

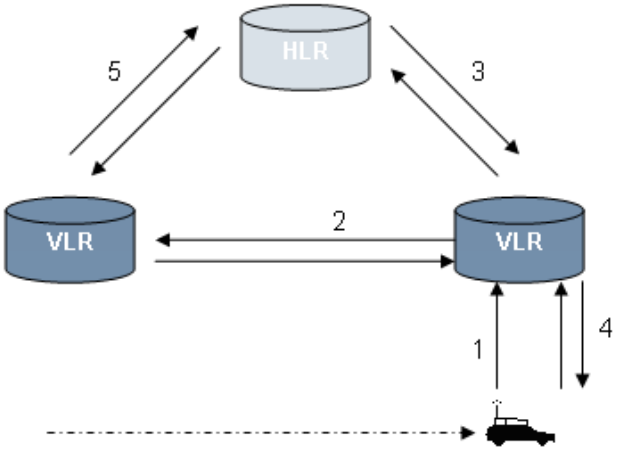


Figura IV.9. Localización de un MS

1. Cuando un móvil entra en el rango de una nueva celda, escucha el canal BCCH, detectando que ha entrado en el área de una nueva celda, el móvil le manda a la estación base su Temporary Mobile Suscriber Identity (TMSI) y la identificación de la localización anterior.
2. Del TMSI y de la identificación de la localización anterior, el VLR obtiene del antiguo VLR el Interntational Mobile Subscriber Identity (IMSI) del móvil y los parámetros de autenticación.
3. El nuevo VLR envía un mensaje al HLR para actualizar la ubicación del móvil, el HLR responde con la información necesaria par manejar la llamada.
4. El VLR genera un nuevo TMSI y se lo envía al móvil, el cual confirma la recepción.
5. Después del paso 3, el HLR envía un mensaje al antiguo VLR, el cual cancela el registro y le confirma la cancelación al HLR.

Control de llamadas

Al llamar a un usuario GSM, el siguiente método se utiliza para enrutar la llamada:

1. Al marcar el Mobile Station ISDN Number (MSISDN). La llamada llega al Interrogating Exchange (INTX) mas cercano, que es un switch con capacidad de preguntarle al HLR la localización actual del subscriptor, para saber como enrutar la comunicación. El HLR le pide al VLR el número roaming del móvil.
2. El VLR le manda el Mobile Station Roaming Number (MSRN) al INTX por medio del HLR.
3. El INTX usa el MSRN para enrutar la llamada hasta la estación móvil

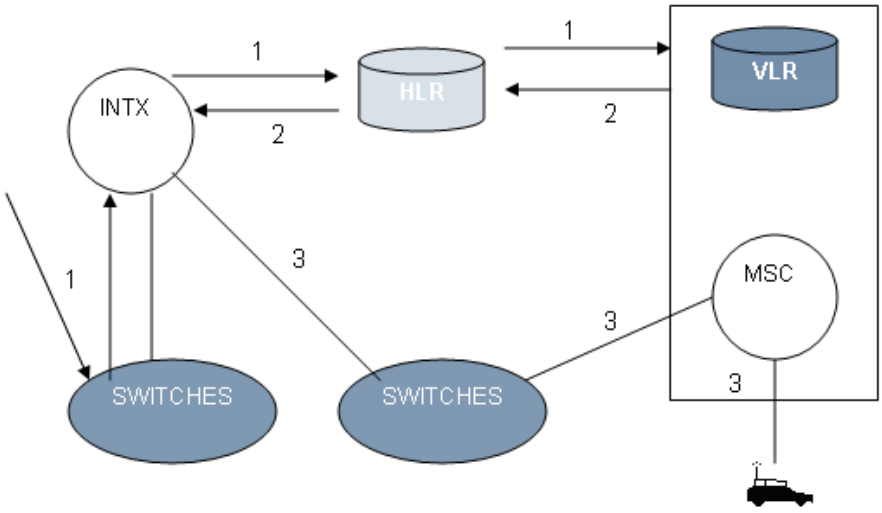


Figura IV.10. Control de llamada

IV.3 La Tarjeta SIM

Desde la realización inicial de GSM (GSM 11.11)[1], cada teléfono GSM lleva en su interior una tarjeta inteligente denominada “Subscriber Identity Module” (SIM), un módulo de seguridad que proporciona al usuario la autenticación necesaria para acceder a la red además de funciones relacionadas con el cifrado de las comunicaciones de voz. Figura IV.11.

Figura IV.11 Tarjetas SIM

La tecnología de tarjetas inteligentes proporciona al módulo SIM la seguridad necesaria para realizar estas tareas y permite un almacenamiento seguro de información confidencial. El SIM contiene datos necesarios para el funcionamiento del servicio GSM (la clave de autenticación del usuario, la clave de cifrado, información de localización...) y datos del subscriptor del servicio (números de marcación abreviada, mensajes cortos...). Estos datos son protegidos físicamente ante intentos de extracción o manipulación por el hardware de la tarjeta.

Una de las ventajas de las tarjetas inteligentes es la portabilidad. El SIM es una especie de llave; una vez extraída del terminal móvil, éste no puede usarse salvo para llamadas de emergencia (si lo permite la red). La posibilidad de extraer el SIM presenta ventajas para el usuario. La información que contiene la tarjeta puede ser fácilmente transferida a otro terminal, permitiendo a los usuarios de GSM utilizar cualquier teléfono compatible sin notificarlo a la red.

[1] GSM 11.11:”Digital cellular telecommunications system (Phase2+); Specification of the Subscriber Identity Module-Mobile Equipment(SIM-ME) interface”.

El papel del SIM definido en GSM 11.11 es esencialmente pasivo; las funciones de autenticación y seguridad las dirigen red y terminal móvil, y la tarjeta SIM se limita a seguir sus instrucciones ejecutando los comandos que se le ordena. Posteriormente a la definición de este funcionamiento básico del SIM, en 1994, el SMG91[2] comienza el trabajo de estandarización de la transferencia de datos hacia el SIM a través del aire y de un comportamiento proactivo del SIM, que puede así iniciar acciones que se ejecutan en el terminal móvil. Figura IV.12.

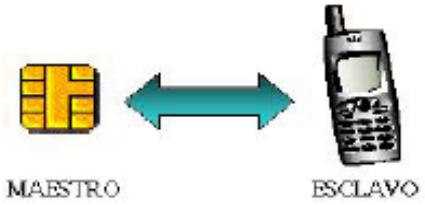


Figura IV.12. SIM Proactiva: la tarjeta SIM le ordena al terminal llevar a cabo determinadas tareas.

Estas funciones forman la base de un nuevo estándar, el **SIM Application Toolkit** (SAT- Kit de herramientas para desarrollo de aplicaciones basadas en SIM), GSM 11.14[3], aprobado en 1996, que pretendía extender el ámbito del SIM más allá de GSM aprovechando la tecnología emergente de tarjetas multiaplicación, que permiten que en una misma tarjeta inteligente coexistan varias aplicaciones.

SAT incrementa el potencial de las tarjetas SIM, permitiendo a los operadores de red desarrollar aplicaciones competitivas que residan dentro de la tarjeta. Así mismo se definen los mecanismos de seguridad necesarios para poder implementar servicios con fuertes restricciones de seguridad, como email, comercio móvil y operaciones bancarias. SAT es una facilidad opcional y está separada de la funcionalidad de GSM; la tarjeta contiene una aplicación específica que lleva a cabo las funciones de GSM y las demás aplicaciones se comunican con el exterior a través ella. De esta forma, la única aplicación que la red “ve” en la tarjeta es la aplicación GSM.

[2] Special Mobile Group 9
[3] GSM 11.14: ”Digital cellular telecommunications system (Phase2+); Specification of the SIM Application Toolkit for the Subscriber Identity Module - Mobile Equipment (SIM-ME) interface”.

IV.3.1 Estructura y características

Tarjetas inteligentes

Del mismo tamaño que una tarjeta de crédito, una tarjeta inteligente contiene un circuito integrado en su cuerpo de plástico que la convierte en un ordenador portable. Al contrario de las tarjetas de banda magnética, las tarjetas inteligentes tienen capacidad de procesar datos y proporcionar protección física (hardware) de los datos que almacenan. Las tarjetas inteligentes pueden transferir datos a través de contactos en su superficie o, en las llamadas tarjetas inteligentes sin contactos, mediante campos electromagnéticos. Figura IV.13.

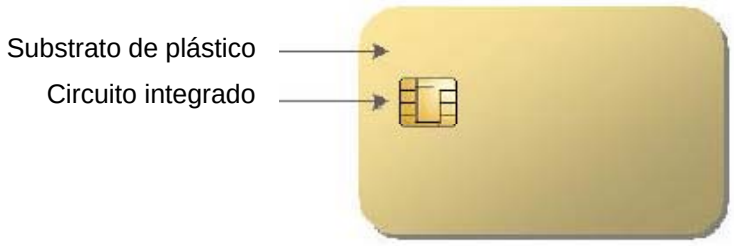


Figura IV.13 Ejemplo de tarjeta inteligente.

La tecnología de banda magnética padece de una importante debilidad, los datos almacenados en la banda pueden ser leídos, borrados o modificados por cualquiera que posea el dispositivo de lectura/escritura apropiado. Por esta causa, no es un medio apropiado para almacenar datos confidenciales y tienen que emplearse medios adicionales para asegurar la confidencialidad e integridad de los datos. La mayoría de los sistemas que emplean tarjetas de banda magnética están conectados “en-línea” a un ordenador, lo que genera unos costos de transmisión considerables.

Gracias a que las tarjetas inteligentes proporcionan capacidad de procesamiento y protección física y lógica de los datos que almacenan (además de poseer una capacidad de almacenamiento muy superior a las tarjetas de banda magnética), superan las debilidades, ofreciendo la posibilidad de que las transacciones a través de las tarjetas se ejecuten sin la conexión con un ordenador y sin que peligre la seguridad del sistema.

La idea de incorporar un circuito integrado en una tarjeta de plástico fue introducida por primera vez en 1968 por dos inventores alemanes, Jurgen Dethloff y Helmut Grotrupp. Sin embargo, no fue hasta principios de los años 80 cuando se realizaron las primeras pruebas comerciales con tarjetas inteligentes; entre 1982 y 1984, en Francia y Alemania, se utilizaron con éxito como tarjetas bancarias y tarjetas prepago de telefonía.

Con los avances en la tecnología de circuitos integrados y criptografía, las tarjetas inteligentes han evolucionado para ser más potentes, seguras y baratas. Por su seguridad y portabilidad, son un dispositivo idóneo para el almacenamiento y procesamiento de información confidencial.

Las tarjetas inteligentes suelen utilizarse para implementar módulos de seguridad y sus principales aplicaciones se encuentran en los sectores bancarios, de telefonía móvil y de comercio electrónico, aunque las podemos encontrar en muchas otras aplicaciones, como las relacionadas con el control de acceso, registros médicos, etc.

En 1987 se publicó el primer estándar para la industria de tarjetas inteligentes, ISO-7816, con el que se intentaba solucionar el problema de interoperabilidad de las tarjetas inteligentes. Por medio de este estándar se establece la forma y dimensiones de las tarjetas, el significado y localización de los contactos del circuito integrado y el protocolo de comunicación de la tarjeta.

- Tipos de tarjetas inteligentes

La comunicación de una tarjeta inteligente (Figura IV.14) con el exterior puede producirse a través de contactos en el chip o, a través de una antena, por ondas electromagnéticas. Se distinguen así dos tipos de tarjetas inteligentes según tengan o no contactos, aunque existen tarjetas híbridas que pueden comunicarse utilizando los dos medios.

Figura IV.14 Tarjeta inteligente. Smart card

Los contactos de las tarjetas inteligentes son una de las más frecuentes causas de fallo por uso o mala conexión, mucho más que el número de ciclos de escritura de la memoria de la tarjeta (de la memoria EEPROM). Además, ya que los contactos están directamente conectados al circuito integrado, existe el peligro de que se produzcan descargas electrostáticas que puedan dañar el circuito integrado.

Las tarjetas inteligentes sin contactos superan esta limitación: no tienen contactos que fallen por el uso, y son más fáciles de usar, ya que no necesitan ser insertadas con una orientación determinada en un dispositivo lector como se muestra en la Figura IV.15. Sin embargo, las tarjetas inteligentes sin contactos deben estar lo suficientemente cerca del dispositivo lector y, como la tarjeta puede moverse rápido, sólo puede transmitirse una cantidad limitada de datos, debido a la corta duración de las transacciones. Además, la potencia para el funcionamiento del microcontrolador debe ser transmitida como una onda electromagnética, lo que limita el consumo máximo del microprocesador y, por lo tanto, sus capacidades.

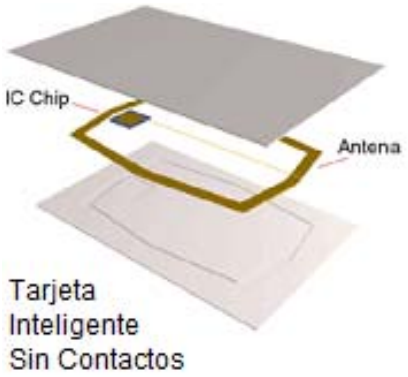


Figura IV.15. Tarjeta inteligente sin contactos

Las tarjetas inteligentes sin contactos se suelen utilizar en situaciones que requieren transacciones rápidas y cuando sólo se necesita intercambiar un número limitado de datos, como por ejemplo en sistemas de control de acceso a edificios.

Debido a que combinan tecnología analógica y digital, es más difícil la integración en las tarjetas inteligentes sin contactos y, actualmente, resultan más caras que las tarjetas con contactos.

Dentro de la tecnología de Tarjetas inteligentes podemos encontrar dos tipos, las llamadas Tarjetas de Proximidad (**Proximity Cards**) como la mostrada en la Figura IV.16.

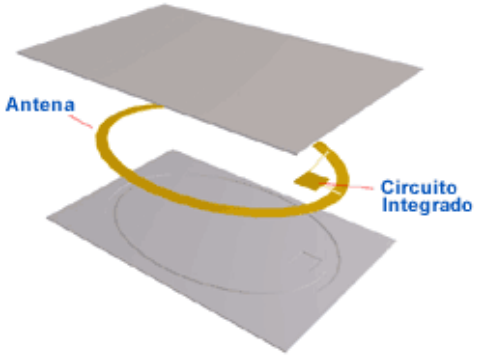


Figura IV.16. Tarjeta de proximidad. Proximity card

Y las tarjetas Híbridas (**Hybrid Cards**), las cuales como su nombre lo indica, contienen tecnología de tarjeta sin contacto y tecnología de Chip de contacto, como se muestra en la Figura IV.17.

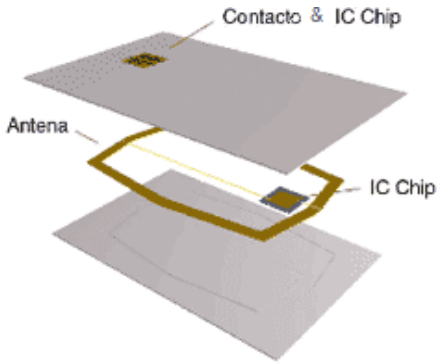


Figura IV.17. Tarjeta híbrida. Hybrid card

- Arquitectura

Una tarjeta inteligente almacena y procesa información a través del microcontrolador embebido en su cuerpo de plástico. El microcontrolador de una tarjeta inteligente está compuesto por un microprocesador y tres tipos de memoria: **Read Only Memory** (ROM), **Electrical Erasable Programmable Read Only Memory** (EEPROM) y **Random Access Memory** (RAM). Como se muestra en la Figura IV.18.

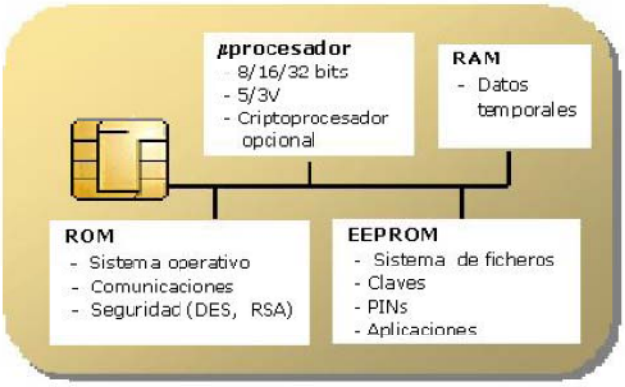


Figura IV.18 Componentes de una tarjeta inteligente

El microprocesador y la memoria están fabricados sobre el mismo chip, lo cual hace difícil y caro interceptar las señales que se intercambian entre el procesador y la memoria, proporcionando así una alta seguridad física de los datos almacenados en la memoria.

- El microprocesador de las tarjetas inteligentes puede ser de 8 bits, 16 y 32 bits. Dado que los chips de las tarjetas inteligentes suelen diseñarse para aplicaciones de seguridad, algunos chips incluyen un coprocesador criptográfico que aumenta la velocidad de las operaciones de criptografía.
- La memoria ROM se utiliza para almacenar los programas permanentes de la tarjeta que se escriben en la memoria durante la fase de producción de la tarjeta: el sistema operativo, datos y aplicaciones de usuario fijas. No se necesita alimentación para mantener los datos en este tipo de memoria y no se puede escribir en ella después de la fabricación.

- La memoria EEPROM, al igual que la memoria ROM, preserva el contenido cuando la alimentación se apaga, por lo que es utilizada como almacenamiento permanente. Pero, a diferencia de la ROM, el contenido de la memoria EEPROM puede ser modificado durante el uso normal de la tarjeta (sería el equivalente al disco duro de una PC). Las aplicaciones de usuario pueden ser escritas en la EEPROM después de la fabricación de la tarjeta.
- La memoria EEPROM se caracteriza por el número de ciclos de escritura que soporta durante la vida de la tarjeta, el periodo de retención de datos y el tiempo de acceso. En la mayoría de las tarjetas inteligentes la EEPROM puede aceptar al menos 100000 ciclos de escritura y puede retener datos durante 10 años. La lectura en una EEPROM es tan rápida como en la RAM, pero la escritura es 1000 veces más lenta.
- La memoria RAM se utiliza para el almacenamiento temporal de datos. La RAM es una memoria no persistente, es decir, su contenido no se mantiene cuando se quita la alimentación. La RAM puede ser accedida un número ilimitado de veces y no posee las restricciones de la EEPROM.

De las tres, la memoria ROM es la más barata porque ocupa menos espacio por celda. Una celda de una EEPROM ocupa hasta cuatro veces más que una celda ROM, y una celda RAM ocupa aproximadamente cuatro veces más que una celda EEPROM. Por esta causa, las tarjetas contienen pequeñas cantidades de memoria RAM.

El tamaño concreto de los chips de memoria de una tarjeta inteligente depende de la implementación. En el mercado se encuentran chips de 4K de ROM, 1K de EEPROM y 256 bits de RAM, hasta las tarjetas más avanzadas con memorias de hasta 256K de ROM, 128K de EEPROM y 10K de RAM.

Una tarjeta inteligente posee ocho contactos mecánicos, que le proporcionan alimentación y señal de reloj, permitiéndole recibir y transmitir datos. El estándar ISO

7816-2[4] especifica la posición, el tamaño mínimo y el cometido de los contactos como lo muestra la Figura IV.19:

Figura IV.19 Contactos del chip de una tarjeta inteligente

- ◊ *Vcc* se utiliza para suministrar la alimentación al chip. El voltaje que se aplica es 3 ó 5 voltios, con una desviación máxima del 10 por ciento. En teléfonos móviles generalmente se aplica el voltaje de 3 V.
- ◊ *RST* se utiliza para enviar la señal de “reset” al microprocesador.
- ◊ El microprocesador de la tarjeta inteligente no posee reloj interno. A través del contacto *CLK* se proporciona una señal de reloj externa a partir de la cual se deriva la señal de reloj interno.
- ◊ *GND* es la conexión de masa. (Tierra)
- ◊ El contacto *Vpp* se utiliza en tarjetas antiguas para proporcionar el voltaje necesario para programar la EEPROM. En las tarjetas actuales este contacto no se utiliza porque el voltaje se genera internamente.
- ◊ *I/O* se utiliza para transferir datos entre la tarjeta y el dispositivo lector en modo semi-duplex.
- ◊ Los contactos *RFU* están reservados para usos futuros.

La forma y tamaño de las tarjetas inteligentes están estandarizados por razones de compatibilidad, siendo los formatos más importantes para tarjetas inteligentes el ID-1 y ID-000. Figura IV.20.

[4] ISO/IEC 7816-2: ”Identification cards - Integrated circuit(s) cards with contacts - Part 2: Dimensions and locations of the contacts.”

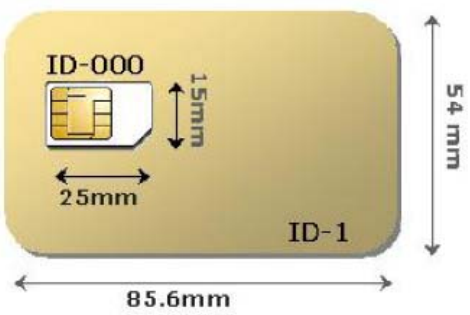


Figura IV.20 Formatos de las tarjetas inteligentes

El formato ID-1 tiene el tamaño de una tarjeta de crédito, para que sea compatible con las tarjetas de banda magnética, y se utiliza en la mayoría de las aplicaciones. El formato ID-000 tiene menores dimensiones y se utiliza en la telefonía celular.

- Protocolo de comunicación

Las tarjetas inteligentes se comunican por medio de una pila de protocolos de comunicación. En el nivel superior se produce la comunicación entre la aplicación de la tarjeta y una aplicación en el exterior. Los comandos y datos intercambiados entre ellas tienen sólo significado para una aplicación particular. En el siguiente nivel se intercambian datos a través de **Application Protocol Data Units** (APDUs Protocolo de Aplicación para Unidades de Datos). El formato de las APDUs está estandarizado, pero el contenido y el significado son específicos de la aplicación. En el nivel inferior encontramos un protocolo de transmisión, T=0 o T=1.

Las tarjetas inteligentes necesitan que un dispositivo externo, generalmente un lector de tarjetas, que les proporcione las señales de reloj y alimentación.

El lector es responsable, de abrir un canal de comunicación entre la aplicación exterior a la tarjeta y el sistema operativo de la tarjeta. El canal de comunicación con la tarjeta es semi-duplex; es decir, los datos pueden ir del lector a la tarjeta o de la tarjeta al lector, pero no en ambas direcciones a la vez.

El protocolo de APDUs, especificado en ISO 7816-4[5], es un protocolo de nivel de aplicación entre la tarjeta inteligente y la aplicación externa. El modelo de comunicación con una tarjeta inteligente es un modelo maestro-esclavo, en el que la tarjeta tiene un comportamiento pasivo (esclavo) esperando a los comandos APDUs de la aplicación externa. Una vez que la tarjeta ha ejecutado la instrucción especificada en el comando, envía el resultado a través de otra APDU.

ISO 7816-4 define dos tipos de APDUs: Comandos APDU (C-APDU), utilizados para transferir datos de la aplicación externa a la tarjeta, y Respuestas APDU (R-APDU), enviadas por la tarjeta para responder a los comandos. Figura IV.21.

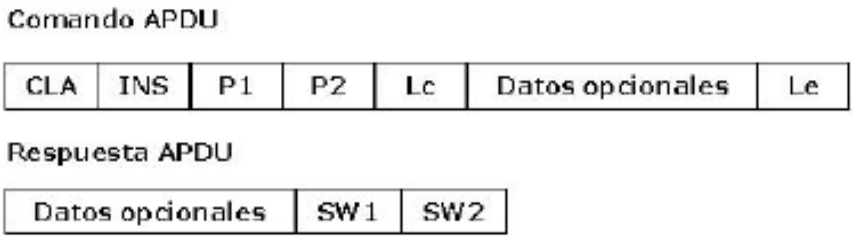


Figura IV.21 Estructura de los comandos y respuestas APDU

Cada C-APDU contiene:

- ◊ Un byte de clase, *CLA*, que identifica la clase de instrucción; por ejemplo, si es una instrucción ISO o es propietaria.
- ◊ Un byte de instrucción, *INS*, que determina el comando que se envía.
- ◊ Un byte de longitud de comando, *Lc*, que especifica la longitud de los datos que se envían a continuación.
- ◊ Datos (opcionales).
- ◊ Un byte de longitud esperada, *Le*, que especifica la longitud prevista de los datos en la R-APDU. Si *Le* es 0x00, se espera que la tarjeta envíe todos los datos disponibles en la respuesta al comando.

[5] ISO/IEC 7816-4: ”Identification cards - Integrated circuit(s) cards with contacts - Part 4: Interindustry commands for interchange.”

Una R-APDU contiene:

- ◊ Datos (opcionales).
- ◊ Dos bytes de estado, SW1 y SW2, que indican el resultado de la operación ejecutada. Por ejemplo, si el contenido de SW1 y SW2 es 0x9000, se indica que el comando fue ejecutado con éxito.

- Sistema operativo

El sistema operativo de una tarjeta inteligente soporta un pequeño conjunto de instrucciones a través del cual se interactúa con la tarjeta. El sistema operativo de una tarjeta puede soportar todas o algunas de estas APDUs, así como nuevas instrucciones que añada el fabricante.

La mayoría de los sistemas operativos implementan también un sistema de archivos basado en ISO 7816-4. Los archivos se organizan jerárquicamente, distinguiéndose tres tipos: Master File (MF-Archivo Maestro), Dedicated File (DF-Archivo Dedicado) y Elementary File (EF-Archivo Elemental). Cada archivo es especificado por un identificador de dos bytes o por un nombre simbólico de hasta 16 bytes.

Tarjetas SIM

GSM introdujo la tecnología de tarjetas inteligentes en el mundo de la telefonía móvil a través del **Subscriber Identity Module** (SIM), un módulo de seguridad que proporciona autenticación del usuario ante la red, flexibilidad en la provisión de servicios, portabilidad de los datos y aplicaciones de usuario y alta seguridad en los servicios de datos a través del móvil, como por ejemplo en las aplicaciones de comercio móvil.

ETSI **European Telecommunications Standards Institute** (Instituto Europeo de Normas para Telecomunicaciones) estandariza la estructura y funciones del SIM a través de cuatro especificaciones:

- GSM 11.11- Especificación de la interfaz entre el SIM y el equipo móvil.

- GSM 11.14 - Especificación de SIM Application Toolkit.
- GSM 03.48[6] - Mecanismos de seguridad para SIM Application Toolkit.
- GSM 03.19[7] - API SIM para la plataforma Java Card.

En la realización inicial de GSM, GSM 11.11, el módulo SIM juega un papel esencialmente pasivo. Las funciones de autenticación y seguridad las dirigen red y terminal móvil, que envían comandos a la tarjeta SIM para que ésta los procese. Posteriormente las funciones del SIM han sido ampliadas respecto a los requerimientos iniciales. En 1996 se aprueba la especificación de SIM Application Toolkit (SAT), GSM 11.14, que estandariza la transferencia de datos hacia el SIM a través del aire y el comportamiento proactivo del SIM, que puede así iniciar acciones que se ejecutan en el teléfono móvil.

SAT permite desarrollar aplicaciones que residen en la tarjeta SIM y proporcionan al usuario servicios de valor añadido más allá del ámbito de los servicios de telefonía, aprovechando las ventajas en seguridad que proporcionan las tarjetas inteligentes: autenticación, confidencialidad e integridad de los datos (por ejemplo claves) o encriptación y firma digital con algoritmos criptográficos como 3DES o RSA.

En GSM 03.48 se definen los mecanismos de seguridad necesarios para poder implementar servicios con fuertes restricciones de seguridad, como email, comercio móvil y operaciones bancarias.

- Estructura lógica del SIM

Muchas de las funciones del SIM necesitan llevarse a cabo en un dispositivo que proporcione una protección física contra ataques a la integridad y confidencialidad de la información, y tenga la capacidad de realizar funciones criptográficas. Estos requisitos

[6] GSM 03.48: “Digital cellular telecommunications system (Phase2+); Security Mechanisms for the SIM Application Toolkit; Stage 2.”
[7] GSM 03.19: “Digital cellular telecommunications system (Phase2+); Subscriber Identity Module Application Programming Interface (SIMAPI); SIM API for *Java CardTM*; Stage 2.”

de seguridad son satisfechos mediante la tecnología de tarjetas inteligentes. Como tarjeta inteligente, el SIM posee un microcontrolador embebido en su cuerpo de plástico, y se comunica con el teléfono móvil, a través de un protocolo de comandos y respuestas.

Para asegurar la interoperabilidad entre el SIM y el ME, independientemente de los fabricantes y operadoras de éstos, GSM 11.11 especifica la interfaz entre el SIM y el ME. En particular estandariza los siguientes aspectos:

- ◊ Requisitos físicos de la tarjeta SIM, señales eléctricas y protocolo de transmisión.
- ◊ Estructura lógica del SIM y contenido de los archivos que se requieren para las aplicaciones de GSM.
- ◊ Comandos que soporta y protocolo de aplicación.
- ◊ Mecanismos de seguridad.

GSM 11.11 especifica dos formatos básicos de tarjetas SIM: el ID-1 y el Plug-in 6. Figura IV.22.

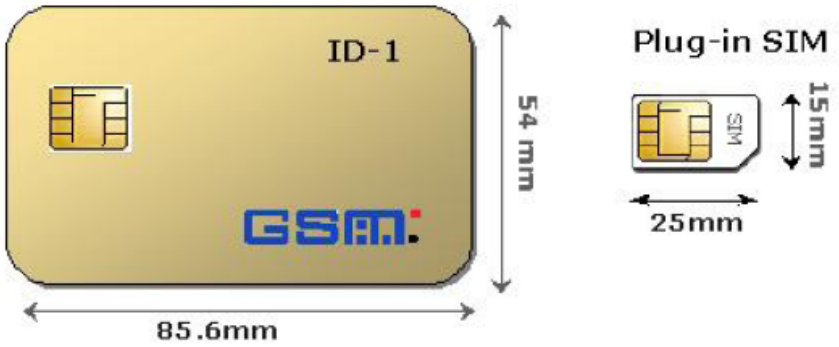


Figura IV.22 Formatos de tarjeta SIM

El ID-1 es el formato tradicionalmente utilizado en las tarjetas de plástico (tarjetas de crédito), y está pensado para teléfonos en los que se cambia el SIM frecuentemente. Sin embargo, el formato actualmente más utilizado en GSM es el SIM plug-in, diseñado para teléfonos que son físicamente muy pequeños y en los que raramente se cambia el

SIM. La única diferencia entre ambos formatos son las dimensiones, ya que son idénticos en términos de sus características físicas y lógicas.

El SIM almacena la información a través de un sistema de archivos jerárquico de cuya raíz, el **Master File** (MF), descienden los demás archivos, **Dedicated Files** (DFs) o **Elementary Files** (EFs). Los archivos de datos o EFs están compuestos por una cabecera y un cuerpo que contiene los datos como se muestra en la Figura IV.23.

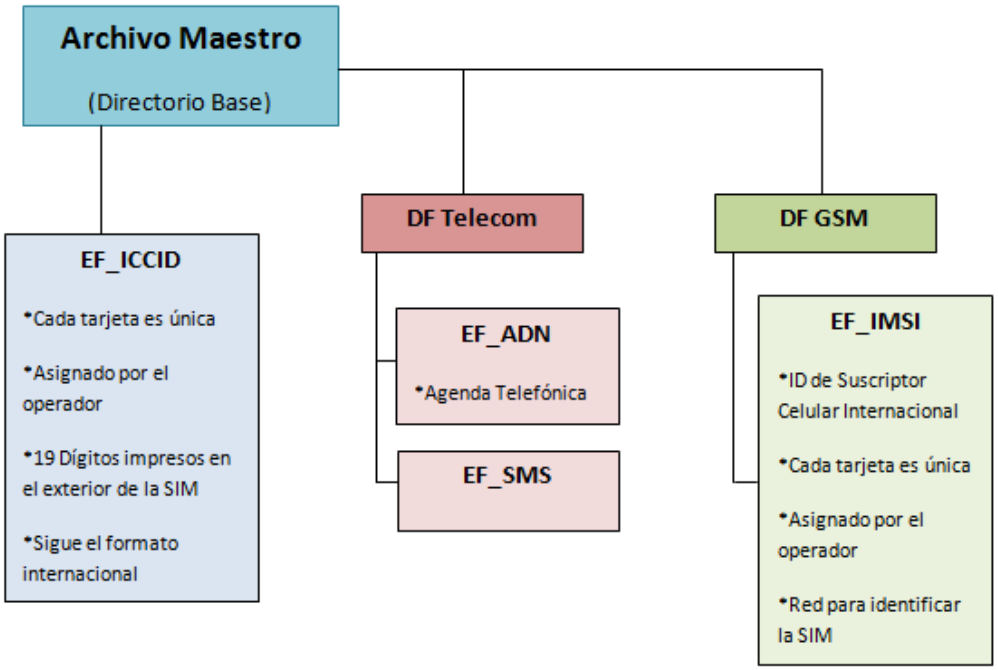


Figura IV.23. Estructura de archivos

En GSM se utilizan EFs de tres tipos: transparente, lineal de longitud fija y cíclico Figura IV.24. Los EFs cíclicos se utilizan para almacenar registros en orden cronológico. Cuando todos los registros han sido utilizados, la siguiente escritura de datos sobrescribirá la información más antigua.

Figura IV.24 Estructura de EFs

En los EFs se encuentra toda la información necesaria para el funcionamiento de la red GSM e información relativa al subscriptor del servicio. Los EFs se agrupan en áreas funcionales organizadas mediante DFs.

IV.3.2 Aplicaciones en la tarjeta SIM

SAT

Desde la primera especificación del SIM en 1991, el desarrollo de la tecnología de circuitos integrados potenció un aumento de la capacidad y de las habilidades de las tarjetas inteligentes, emergiendo en estos años las tarjetas inteligentes multiaplicación. Estas mejoras se reflejaron en un aumento de la funcionalidad del SIM, que llegó con la especificación de SIM **Application Toolkit** (SAT), GSM 11.14.

SAT se apoya en la tecnología de tarjetas inteligentes multiaplicación para proporcionar al usuario de GSM nuevos servicios de valor agregado a través de aplicaciones que residen en la tarjeta SIM. La especificación SAT define los comandos y procedimientos necesarios para que las aplicaciones almacenadas en el SIM puedan interactuar y operar con cualquier ME que soporte esta capacidad.

Los comandos SAT son proactivos, es decir, son comandos que utilizan las aplicaciones en el SIM para ordenar al ME que realice una determinada acción, por ejemplo:

- ◊ Mostrar un texto por la pantalla del teléfono móvil o que suene un tono.
- ◊ Establecer una llamada de voz o datos a un número proporcionado por el SIM.
- ◊ Enviar un mensaje corto.
- ◊ Obtener información local del ME.
- ◊ Comunicarse con tarjetas adicionales.
- ◊ Pedir al ME que lance el visor para una URL determinada.
- ◊ Gestionar los temporizadores del ME.
- ◊ Establecer un canal de comunicación independiente, etc.

SAT es una facilidad opcional que no debe influir en las funciones GSM de la tarjeta SIM. SAT contiene comandos adicionales e independientes a los definidos en GSM 11.11 para la comunicación entre el SIM y el ME. Esto facilita que las nuevas aplicaciones de las operadoras o de terceras partes puedan residir en el SIM con la aplicación GSM. Desde el punto de vista de la red GSM sólo existe una aplicación en la

tarjeta SIM, aquella que lleva a cabo las funciones definidas en GSM 11.11, la aplicación GSM. Las aplicaciones SAT se comunican con el exterior a través de la aplicación GSM y para ello se utilizan comandos GSM 11.11 definidos para este propósito.

El modelo de comunicación con una tarjeta inteligente sigue un esquema maestro-esclavo en el que la tarjeta inteligente tiene un comportamiento pasivo (esclavo), esperando a recibir comandos APDU para ejecutarlos. Es el ME el que inicia la acción enviando una APDU, y el SIM se limita a ejecutar los comandos que le ordenan. Con los comandos proactivos de SAT los papeles se invierten: el SIM se convierte en el maestro y puede enviar comandos al ME para que éste los ejecute.

Con la funcionalidad que le proporciona SAT, el SIM puede realizar el control de las llamadas y de los mensajes cortos, por ejemplo:

- El SIM puede proporcionar un mensaje completo SMS al terminal para que éste lo envíe.
- Puede recibir directamente un mensaje SMS enviado desde la red. Este mensaje puede contener, por ejemplo, una nueva aplicación que se ejecutará en el SIM o datos/comandos para el SAT.
- Puede realizar control de llamada inteligente. Los números de teléfono introducidos por el usuario pasarán a la aplicación SIM antes de ser marcados para que ésta pueda modificarlos convenientemente o incluso bloquearlos.
- Según la información local sobre identidad de celda, estado de llamada, estado de cobertura, etc. las aplicaciones residentes en el SIM pueden, si lo desean, modificar su comportamiento cuando la situación del móvil cambia.

Las aplicaciones SAT suelen diseñarse siguiendo un modelo cliente-servidor. Con los comandos proactivos, la aplicación en el SIM puede establecer un canal de datos con el ME y, a través del ME, con un servidor remoto en la red. El ME permite entonces intercambiar datos entre la aplicación en el SIM y el servidor de forma transparente. Actualmente las comunicaciones entre el cliente (la aplicación en el SIM) y el servidor

se realizan a través de SMS, pero sería posible utilizar otros mecanismos de transporte, como USSD (**Unstructured Supplementary Service Data**- Servicio Suplementarios de Datos no Estructurado) o GPRS (**General Packet Radio Service** - Servicio de Radio por Paquetes).

Las aplicaciones SAT interactúan con el usuario a través del sistema de menús del teléfono móvil. Las aplicaciones pueden añadir una entrada al sistema de menús del móvil para que el usuario seleccione los nuevos servicios.

Cuando el usuario selecciona un servicio SAT en el menú de su teléfono, el ME comunica al SIM la selección y el SIM activa la aplicación correspondiente. La aplicación seleccionada se comunica con el servidor remoto intercambiando la información necesaria para proporcionar el servicio. Si el servicio requiere la intervención del usuario, la aplicación SAT puede ordenarle al ME que pida una entrada al usuario, que le de a elegir entre una lista de opciones o que muestre un texto por pantalla. El ME transmite la respuesta del usuario al SIM para que la interprete.

Para aplicaciones con requisitos altos de seguridad, como email, servicios bancarios, comercio móvil, compraventa o cualquier servicio orientado a transacción, SAT proporciona mecanismos para proteger los mensajes cortos intercambiados entre la aplicación cliente en el SIM y la aplicación servidora en la red. La seguridad en este tipo de servicios implica:

- Autenticación: verificar que la otra entidad en la comunicación es quien asegura ser. En la inicialización de la conexión, el servicio debe asegurar que las dos entidades son auténticas y, posteriormente, debe asegurar que una tercera parte no suplanta a ninguna de las dos partes legítimas.
- Integridad: asegurar que los datos de una comunicación no se alteren, es decir, que los datos recibidos por el receptor coincidan con los transmitidos por el emisor. Y en caso de no ser así, detectar esta modificación.

- Confidencialidad: el intercambio de datos entre las dos partes de la comunicación debe permanecer secreto, de manera que ninguna persona no autorizada pueda acceder a esta información.
- No repudio: prevenir que el emisor o el receptor nieguen un mensaje transmitido. Puede ser de dos tipos:
    - Con prueba de origen: el destinatario tiene garantía de quién es el emisor, de forma que puede probar ante una tercera parte que el mensaje fue enviado por éste.
    - Con prueba de entrega: el emisor tiene la prueba de que los datos han llegado íntegramente al destinatario correcto.

La especificación GSM 03.48 define los métodos para proteger el contenido de los mensajes de aplicación, describiendo formato e implementación de los llamados Paquetes seguros para el servicio de mensajes cortos punto a punto, SMS-PP (SMS Point to Point), y para el servicio de difusión de mensajes cortos, SMS-CB (SMS Cell Broadcasting). GSM 03.48 no especifica el tipo de mecanismo de seguridad (cifrado simétrico o de clave pública, firma digital,...) ni los algoritmos criptográficos a utilizar, por lo que éstos dependerán de la implementación. La seguridad real del sistema estará en función de la robustez de las claves y algoritmos elegidos. El SIM puede encargarse de “almacenar” de forma segura los algoritmos criptográficos y las claves correspondientes en el lado del cliente.

En aplicaciones comerciales que requieren realizar transacciones de forma segura, se utilizan las tarjetas SIM para proporcionar integridad y confidencialidad extremo a extremo en los mensajes y, más importante, para proporcionar firma digital aprobada por el usuario, que autentique al usuario y evite problemas de repudio. En estos sistemas, la tarjeta SIM cifra los mensajes en el lado del cliente. En el lado del servidor, los mensajes son descifrados, por ejemplo, por las instituciones financieras. Todos los mensajes intercambiados pueden ser firmados. El SIM pide al usuario su intervención para aprobar cualquier transacción antes de que se envíe a la institución financiera.

IV.3.3 Evolución de la tarjeta SIM

La primera tarjeta fue lanzada en 1985 por la operadora móvil celular alemana Netz C, la que fue simplemente una tarjeta magnética. La movilidad de la suscripción y el aumento de la seguridad a través de la remoción de la tarjeta fueron las principales ventajas de la introducción de la tarjeta. El número del teléfono y los otros datos necesarios al **billing** estaban relacionados a la tarjeta y no más al aparato celular.

También en 1985, algunos países europeos firmaron un acuerdo para el desarrollo del GSM y un nuevo padrón para uso de tecnología digital. En 1992, la primera red GSM fue lanzada.

Figura IV.25. Evolución del chip

A partir de ese momento, las normas ETSI GSM han definido por el momento tres fases diferenciadas para las tarjetas SIM.

Fase 1 (1991-1993)

- Capacidad de 2-3 kybtes de memoria EEPROM
- Capacidad para presentación de servicios básicos:
  - Efectuar y recibir llamadas;
  - Autenticación;
  - Cifrado;
  - Administración y desbloqueo de claves;
  - Selección de redes preferenciales;
  - Agenda personal, etc.
  - Números abreviados para efectuar llamadas;
  - Solo recepción de SMS

Fase 2 (1994-1997)

- Capacidad de 8 kybtes de memoria EEPROM
- Inclusión de nuevas funcionalidades extra a Fase 1:
  - Informaciones sobre cobranza (Advice of Charge): valor de la llamada, conversión del costo de la llamada para moneda local y su visualización en el display, limite máximo de valor para llamadas;
  - Números fijos para efectuar llamadas;
  - Status de short messages – aviso de memoria disponible;
  - Último número llamado – función de rellamada en el SIM;
  - Elección del idioma preferencial en el SIM;
  - Selección de mensajes por región (**Cell Broadcast**) en donde se encuentra el usuario;
  - Almacenamiento de números telefónicos con más de 20 dígitos;
  - Visualización del proveedor del servicio de telefonía móvil en la pantalla del dispositivo móvil.

Fase 2 + (1997 – Actual)

- Capacidad de 16 y 32 kbytes de memoria EEPROM
- Procesadores de 8 y 32 bits

- 128 bytes en RAM
  - Lanzamiento de la funcionalidad **Over The Air**;
  - Lanzamiento de la funcionalidad y aplicaciones **SIM Tool Kit** ();
  - Números bloqueados para efectuar llamadas;
  - Inclusión de caracteres especiales de idiomas como el chino, japonés, ruso, etc.

PARTE III. APLICACIÓN

CAPÍTULO V

TARJETA SIM DE ALTA DENSIDAD

Las tarjetas de Alta Densidad representan la siguiente generación de tokens portátiles y seguros, para los mercados móviles e inalámbricos. Lo que hace particulares a estas tarjetas, es la cantidad de megabytes de memoria Flash permanente (no-volátil) disponibles en el mismo dispositivo. Esto es una pequeña revolución comparado a las tarjetas EEPROM actuales, las cuales permiten el uso de solamente pocos kilobytes de memoria, tanto para aplicaciones como datos. La memoria flash puede ser accedida vía USB (**Universal Serial Bus**) o vía una interfaz de alta velocidad MMC (**Multi Media Card**- Tarjeta Multimedia). Es así que dos diversos ecosistemas coexisten en el mismo chip, lo que provoca que los aspectos de seguridad sobre estas tarjetas sea particularmente interesante.

V.1 Características de la SIM de Alta densidad

La tarjeta de Alta Densidad (HDSC – **High Density Smart Card**) es una respuesta de la industria, para tratar de disipar algunas limitaciones que han aplicado a las tarjetas inteligentes convencionales, de tal modo desafían la realidad actual, sobre cómo pueden y no pueden ser utilizadas.

Las compañías que conducen el desarrollo de los dispositivos para HDSC incluyendo Renesas[1], Samsung, M-Systems, Atmel[2] y ST[3], junto con los principales distribuidores de las tarjetas inteligentes, tales como Giesecke & Devrient[4], Gemalto[5], Sagem-Orga[6],

[1] Renesas: http://www.renesas.com/
[2] Atmel: http://www.atmel.com/
[3] ST: http://www.st.com/
[4] Giesecke & Devrient: www.gi-de.com
[5] Gemalto: www.gemalto.com
[6] Sagem-Orga: www.sagem-orga.com

Oberthur[7], etc., tratan de dar a este desarrollo una importante oportunidad de negocio. Aunque el tamaño de la memoria sea la característica principal, las implicaciones sobre estos productos van más allá, por lo que el potencial para las tarjetas inteligentes es trascendental en un futuro cercano.

Memoria

De acuerdo a la Ley de Moore (cada 18 meses la potencia de los ordenadores se duplica), esperaríamos que la memoria de las tarjetas inteligentes se duplicaría cada 18 meses. Sin embargo, esto no haría a la tarjeta inteligente relativamente más atractiva como fuente de almacenamiento, pues el resto de las memorias que funcionan actualmente también aumentarían su capacidad basadas en la misma ley. Ahora bien, lo trascendental de la tarjeta HDSC es que el cambio en la capacidad de almacenamiento ha sido radical respecto a una tarjeta SIM normal, ya que el incremento se estima de casi mil veces, lo que ha justificado que los desarrolladores de tarjetas aumenten sus esfuerzos por impulsar esta tecnología, tratando de reducir los costos para que los productos puedan ser fácilmente asequibles al mercado en masa.

La figura V.1 muestra un ejemplo de una tarjeta HDSC de Samsung en un formato SIM, las cuales son las que empezarán a sustituir las tarjetas SIM convencionales dentro de las nuevas redes de tercera generación.

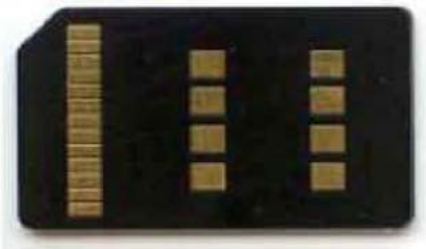


Figura V.1 Samsung HDSC

Los diseños de aplicaciones ya no están basados en tarjetas de 128kb, sino que actualmente ya se consideran tarjetas de 128Mb y de hecho ya existen prototipos de 1Gb. Actualmente el almacenamiento es equivalente a una memoria USB flash, pero se

[7] Oberthur: www.oberthur.com

incluyen las características de seguridad. La razón por la que esto es posible es debido a la combinación de una tarjeta inteligente junto con la tecnología de las memorias flash.

Ahora bien, el almacenamiento de datos “importantes” es un punto muy relevante, si se relaciona con información personal (privada), expedientes de salud o la ultima descarga de música y videoclips. Donde se decida almacenar esta información, se debe tomar en cuenta la seguridad, utilidad, portabilidad y costo de las arquitecturas de almacenamiento. La tabla V.1 resume algunas características de algunas arquitecturas de almacenamiento y sugiere que la HDSC puede llegar a ser una buena opción cuando se requiere de alta seguridad y gestión de la información.

	Memoria interna del Dispositivo Móvil	**Conector Flash**	**Conector Flash con password para acceso**	**Tarjeta Sim Convencional**	**HDSC**
Seguridad	0	0	1	2	2
Capacidad	2	2	2	0	2
Administración	0	0	1	2	2

0= Bajo, 1= Medio, 2= Alto

Tabla V.1 Características de memorias

Velocidad de transferencia

Una memoria grande no puede ser explotada realmente sin una interfaz rápida de transferencia, ya que el tiempo se consumiría en la lectura y escritura de datos. Esto se está tratando dentro de las organizaciones estandarizadoras, donde se proponen dos candidatos. El primero se basa en la tecnología USB y el segundo se basa en la interfaz de tarjetas de memoria (MMC). Hasta este momento no se ha decidido cual será la tecnología estándar a utilizar, ya que por un lado se enfoca la decisión hacia la tecnología USB por el potencial de compatibilidad con las PCs, pero los distribuidores de equipos móviles proponen la solución MMC, debido a que esta interfaz ya se ha desarrollado en equipos anteriores, por lo que se expresa que el desarrollo bajo esta implementación sería mas rápida. Cualquiera que sea la opción final, se observa un aumento drástico en la velocidad respecto a las tarjetas SIM convencionales. En la

tabla V.2, se observa la comparativa en las tasas de transferencia entre las tarjetas SIM convencionales y las HDSC.

	Velocidad I/O (Entrada-Salida)	
	Típica	Máxima
SIM Card	9.6 kbits/s	78 kbits/s
HDSC (USB)		8 Mb/s

Tabla V.2. Comparativa entre SC y HDSC (Velocidad de transmisión)

En este caso, si el CPU en la HDSC funciona más rápido que una tarjeta SIM convencional, entonces más algoritmos complejos pueden ser utilizados, así como más memoria y la velocidad de Entrada/salida estará disponible y no habrá necesidad de utilizar formatos especiales para el almacenamiento de datos seguros. Así mismo, es también posible desafiar la presentación de que un SC (SIM Card) es demasiado lenta para realizar el cifrado y descifrado de servicios de **streaming** (ver u oír archivos directamente sin necesidad de descargar a un dispositivo) de audio y video. Para altas tasas de trasferencia como TV por satélite, la solución fue utilizar SCs para entregar solo claves de sesión dentro del hardware no seguro, lo cual podría hacer frente a las tasas de transferencia requeridas. Esto podría tener un ataque en la seguridad, en cuanto se realizara una redistribución de llaves de sesión. Con HDSC es posible cifrar directamente dentro el dispositivo seguro y así las claves de sesión nunca pasan por un ambiente no seguro.

Velocidad de CPU

Un chip más grande brinda la oportunidad de tener procesadores más rápidos e incrementa la disponibilidad de la memoria RAM, ayudando con almacenamientos temporales. Las frecuencias del reloj del CPU en tarjetas SIM convencionales pueden alcanzar un poco más de 66MHz usando multiplicadores internos de frecuencia a fin de hacer frente a la limitación de la fuente de alimentación de energía. Sin embargo,

Gestión de Derechos Digitales y Almacenamiento de Datos Personales

Para las soluciones de Administración de Derechos Digitales en los datos, la industria requiere una gran cantidad de almacenamiento de datos, acoplando también una solución robusta de seguridad para la administración de la portabilidad de los derechos de los usuarios. Siendo la HDSC un candidato a considerar, ya que no solo tiene una gran capacidad de almacenamiento seguro, sino que también es portable entre dispositivos. Mientras que algunas memorias flash tienen características semejantes, carecen de una presentación centralizada y funcionalidades de administración tal como lo tiene la HDSC que lo hereda de la tarjeta SIM predecesora.

Para los datos personales, tales como información sobre registros de salud, passwords, etc., ha habido muchas discusiones acerca de el trato que ofrece a estos datos las tarjetas inteligentes convencionales, no obstante los debates han sido mayores, cuando se plantea el uso de bases de datos centralizadas que contengan toda esa información. Las HDSC ofrecen almacenamiento seguro para información personal sin exponerla a las vulnerabilidades de una base de datos centralizada, la cual puede ser utilizada solo como alternativa o como complemento de una solución.

Biometría es una clase particular de datos personales que por lo regular causa discusión, sobre donde es el mejor lugar para alojar dicha información, así como el debate sobre donde debe tomar lugar la verificación de esta. Claramente, la memoria de la HDSC es suficiente para no solo almacenar grandes cantidades de información biométrica sino también para alojar los algoritmos utilizados para hacer la verificación, y para cualquier otro certificado Hash o firma digital requerida.

V.2 Estructura y modo de operación

Para el caso de la arquitectura GSM, el chip utilizado se trata de una tarjeta ICC (**Integrated Circuit Card** - Tarjeta de Circuitos Integrados) llamada tarjeta SIM y para el caso UMTS se trata de una UICC (**Universal Integrated Circuit Card** –Tarjeta Universal de Circuitos Integrados) con una aplicación USIM. Tanto la tarjeta SIM como la tarjeta USIM contienen un conjunto de archivos con datos del operador de la red de telefonía celular y del usuario y están dotadas de medios para ejecutar operaciones

asociadas a una serie de comandos que permiten al terminal acceder a esos archivos (leerlos, escribirlos, selecciones, verificar claves de usuario, etc.). Entre los datos de usuario están los que autentifican al usuario ante la red.

Hasta la llegada de UMTS no se solía hacer distinción alguna entre la interfaz física de la aplicación (dependiente de la propia naturaleza de la tarjeta inteligente ICC) y la propia aplicación: ambos eran llamados SIM. La tercera generación de telefonía celular, introdujo la separación de la interfaz física y las aplicaciones. Vea figura V.2:

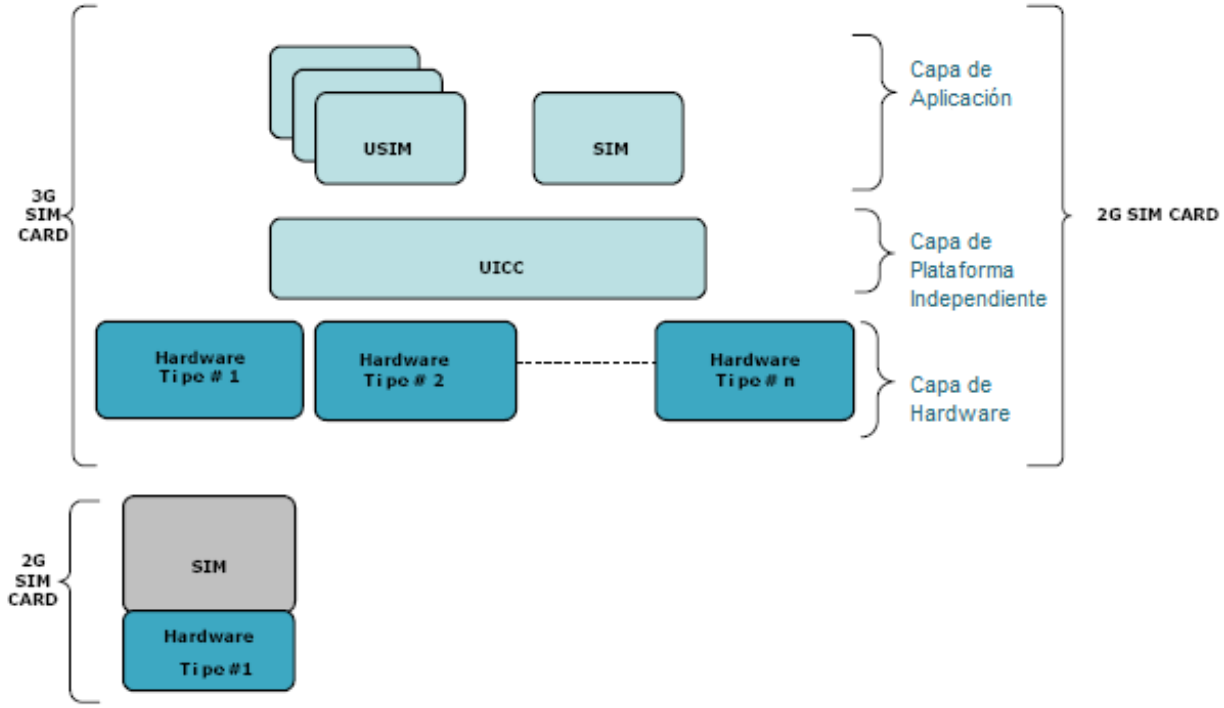


Figura V.2. Representación física y lógica de de una SIM y USIM sobre UICC

La interfaz física recibe el nombre de UICC y se trata de una plataforma en la que pueden convivir varias aplicaciones simultáneamente; entre ellas puede haber aplicaciones SIM y/o USIM, por tanto la aplicación de identificación de usuario propiamente dicha recibe en UMTS el nombre de USIM. De acuerdo a esto, se considera que una tarjeta inteligente es un dispositivo físicamente seguro, es decir, es un dispositivo en el que los datos almacenados se encuentran protegidos frente a ataques de terceros que pretendan leerlos, modificarlos, borrarlos o falsearlos sin permiso del propietario de la información.

Dentro de los estándares desarrollados para las tarjetas USIM, se estipulan las siguientes características de estructura:

Para efectos de homologación de criterio se definen los siguientes aspectos:

- UICC: es una tarjeta removible que contiene un USIM, según lo especificado en 3GPP TS 21.111[8]
- USIM: es una aplicación lógica de 3G en una tarjeta de circuito integrado (IC), la cual interopera con los terminales móviles 3G y provee acceso a los servicios de 3G. Para el acceso a los servicios, un UICC deberá contener un USIM válido, el cual estará presente en todo momento, excepto para las llamadas de emergencia.
- ISIM: es una aplicación lógica residente en el UICC de una tarjeta de circuito integrado (IC) según lo especificado en 3GPP TS 31.103[9]. Esta contiene información para identificar y autenticar al usuario en el IMS (IP Multimedia Subsystem). La aplicación ISIM puede coexistir con la aplicación USIM en el mismo UICC.

Todas las UICC/USIM en el mercado deben estar homologadas bajo las siguientes especificaciones. Vea Tabla V.3 y Figura V.3

3GPP	TS 21.111	Requerimientos de tarjetas USIM y IC
3GPP	TS 31.102	Características de la aplicación de USIM (Universal Subscriber Identity Module)
3GPP	TS 31.111	Aplicación Toolkit para USIM (USAT)
3GPP	TS 31.101	Interfaz terminal -UICC-; Características físicas y lógicas
3GPP	TS 31.103	Módulo de Identidad de servicios ISIM (IP Multimedia Services Identity Module)
3GPP	TS 33.102	Seguridad 3G; Arquitectura de seguridad
3GPP	TS 22.101	Aspectos de servicio; principios de servicio
3GPP	TS 31.101	Sistema de numeración para aplicaciones en telecomunicaciones con tarjetas IC.
ETSI	TS 102.221	Interfaz Terminal- UICC; Características físicas y lógicas

[8] 3GPP TS 21.111. USIM and IC card requirements. www.3gpp.org/ftp/Specs/html-info/21111.htm

[9] 3GPP TS 31.103. Characteristics of the IP Multimedia Services Identity Module (ISIM) application. www.3gpp.org/ftp/specs/html-info/31103.htm

ETSI	TS 102.222	Comandos administrativos para aplicaciones en telecomunicaciones
ISO/IEC	7816-3	Señales eléctricas y protocolos de transmisión
GSM	11.12	Interfaz 3V SIM/ME
GSM	11.18	Reserved for 1.9V SIM

Tabla V.3 Estándares UICC/USIM

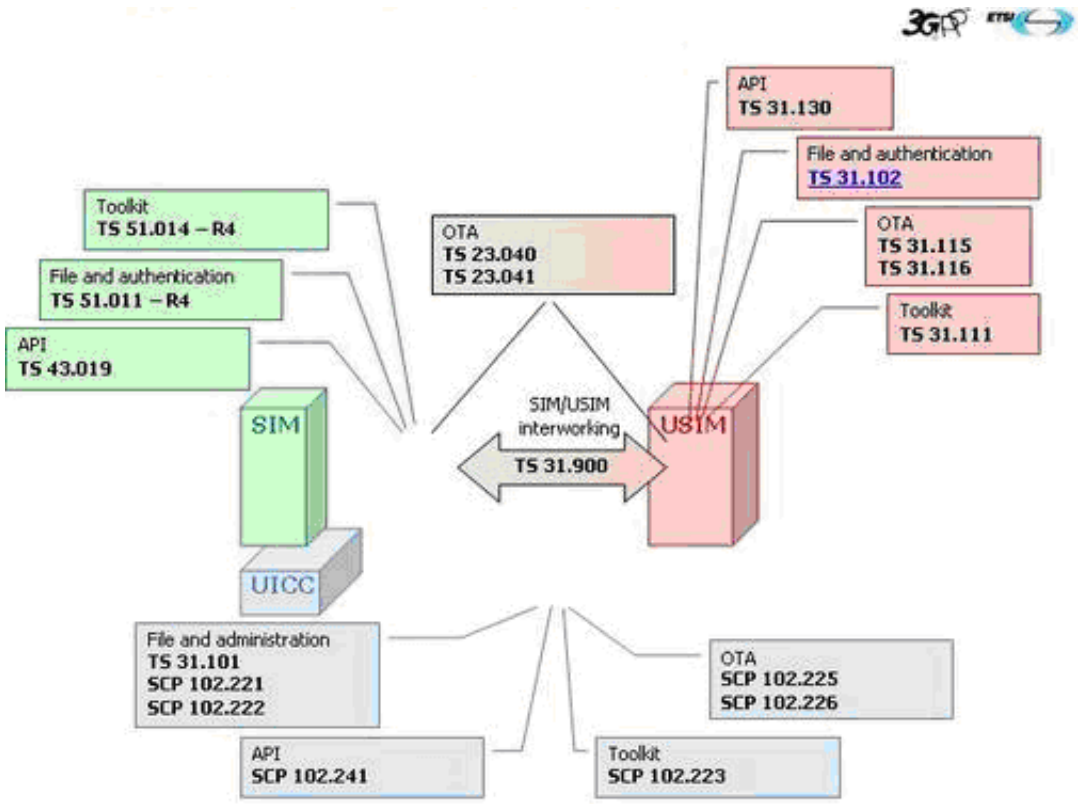


Figura V.3. Relación de los estándares para SIM y USIM

Características de Seguridad y Autenticación

Todas las tarjetas UICC deben emplear como algoritmo de autenticación A3/A8 GSM en COMP 128. Versión 3. De manera que se garantice la seguridad y confidencialidad de la información. Esto es, los sistemas GSM/UMTS utilizan cuatro algoritmos A3, A5/1, A5/2 y A8. El A3 realiza la autentificación del usuario y evita la clonación de teléfonos. El A5/1 se encarga del cifrado fuerte de la comunicación (cifrado de voz entre el teléfono y la estación base GSM). El A5/2 es similar al anterior, pero se usa para la exportación a ciertos países y es mucho más débil. El A8 se utiliza para la generación de la clave que utilizará posteriormente el A5/1 o A5/2. Los algoritmos A3 y

A8 son funciones unidireccionales "hash" dependientes de la clave. Los algoritmos A3 y A8 de GSM son similares en funcionalidad y se implementan como un único algoritmo denominado COMP128.

Así mismo, como característica de seguridad, las tarjetas UICC deberán proveer autentificación y negociación de claves que garanticen la seguridad y privacidad de la transferencia de datos, tanto en la red como en la provisión de servicios en las comunicaciones de 3G, así como permitir la portabilidad de los servicios del sistema 3G independientemente del terminal del usuario.

Debido a los algoritmos de comunicación, las tarjetas UICC deben autenticarse a sí mismas en la red y viceversa. Y deberán proveer capacidad para autenticar al usuario, para lo cual este deberá poseer un Número Personal de Identificación – PIN (**Personal Identification Number**) de 4 a 8 dígitos decimales (este número depende de los requerimientos finales de la operadora).

Las tarjetas deberán proveer bloqueo de PIN cuando este ha sido digitado erróneamente 3 veces consecutivas (el PIN puede ser modificado por el usuario), así como tener un mecanismo de desbloqueo de PIN usando PUK (**PIN Unblocking Key**) que debe poseer de 4 a 8 números decimales, el cual debe ser bloqueado en cuanto este haya sido digitado erróneamente 10 veces consecutivas. (El PIN y PUK no pueden ser leídos).
Las tarjetas UICC deberán operar en modo dual (GSM/ 3G) para que los usuarios UMTS puedan transitar en la red GSM con base en las especificaciones 3GPP TS 22.101 para la provisión de los servicios GSM.

Toda la información del usuario transferida a la terminal móvil durante la operación con la red deberá ser suprimida de la terminal móvil después de la extracción del UICC, deselección del USIM, desactivación de la terminal, o después de un reset eléctrico del UICC.

El UICC deberá poseer una clave de autenticación K con la red, la cual deberá ser única para cada UICC/USIM, asegurando la información de señalización trasmitida entre los terminales móviles y la red, para lo cual deberá usar Integrity Key (IK), el cual deberá ser almacenado en el USIM.

Características de Memoria

La tarjeta USIM deberá tener la capacidad de implementar en ROM, la funcionalidad SIM dual (doble), para el manejo de dos suscripciones de cliente (MSISDN A y MSISDN B) en caso de que la operadora de telefonía móvil así lo requiera, de acuerdo con la compatibilidad de los comandos de la norma GSM 11.14 fase 2+, release 99.

La vida de la tarjeta deberá estar garantizada contra fallas en el rango de 10^5 reescrituras en EEPROM y se deberá garantizar la integridad de los datos almacenados en EEPROM durante al menos 10 años y deberá soportar el modo **sleep** (dormido).

Las tarjetas UICC deberán tener precargadas las aplicaciones, que se soliciten en el perfil de fabricación de cada entrega y que ya hayan sido licenciadas con el fabricante y deberán estar protegidas contra escrituras y borrados accidentales.

El UICC deberá tener la capacidad de almacenar al menos la siguiente información:

a) Información relativa al UICC:
   - Identificación de la tarjeta (IC **Card Identification)**
   - Lenguajes preferidos
   - Directorio de aplicaciones

b) Información relativa al USIM/ISIM:
   - Información Administrativa sobre el modo de operación del USIM/ISIM
   - Tabla de servicios USIM / ISIM;
   - IMSI;

- Indicación de Lenguaje
- Información de localización
- Cipher key (Kc) y número de secuencia cipher key
- Control de acceso a clases;
- PLMNs prohibidas (**Public Land Mobile Network**-Red Móvil Terrestre Pública)
- Identificación de fase
- **Ciphering Key** para GPRS (Llave de cifrado para GPRS)
- Información de localización GPRS
- Cell Broadcast related information
- Códigos de llamadas de emergencia
- Números de teléfonos (ADN - **Abbreviated Dialing Number-** Números de Marcación Abreviado, FDN – **Fixed Dialing Numbers –** Números de Marcación Fijos, SDN - **Service Dialing Numbers –**Números de Servicio de Marcado)
- Mensajería corta y parámetros relacionados
- Parámetros de capacidad y configuración
- Períodos de búsqueda HPLMN (**Home Public Land Mobile Network**)
- Información BCCH: lista de frecuencias portadoras a ser usadas para la selección de celda.

Adicionalmente el UICC debe manejar y proveer almacenamiento para la siguiente información de acuerdo a los requerimientos de seguridad:

- PIN
- **PIN enabled/disabled indicador** (Indicador de autorización/desautorización de PIN)
- **PIN error counter** (Contador de errores de PIN)
- **Unblock PIN** (Desbloqueo de PIN)
- **Unlock PIN error counter** (Contador de desbloqueos de errores de PIN)
- **Data integrity keys** (Llaves de Integridad de Datos)
- **Subscriber authentication keys** (Llaves de autenticación de suscriptores)

Así mismo, el UICC debe tener soporte para directorio telefónico y detalle de llamadas, basado en las funcionalidades de ADN definidos en GSM 11.11, de manera que se provea al usuario final una base de datos personal para administrar direcciones de correos electrónicos, números telefónicos y números de fax de sus contactos.

- Directorio Telefónico

- Debe poseer dos campos de nombre por entrada, por ejemplo, para dos representaciones diferentes del mismo nombre (por ejemplo, en caracteres japoneses y en caracteres latinos).
- Debe poseer soporte para múltiples números telefónicos por entrada.
- Debe poseer soporte para direcciones de correo electrónico enlazadas con el directorio telefónico.
- Debe poseer soporte para agrupar entradas del directorio telefónico por grupos definidos por el usuario por ejemplo: amigos, trabajo, privado, etc.
- Debe poseer soporte para almacenar al menos 500 entradas, dependiendo de los requerimientos de cada operador este número puede ser variable.

- Detalles de llamadas

El UICC debe poseer soporte para el almacenamiento de los detalles de las llamadas basado en los siguientes atributos:

- Llamadas terminadas: número llamante, fecha y hora, nombre del llamante y estatus de la llamada (ejemplo: contestada, perdida, rechazada), costo y duración.
- Llamadas originadas: número saliente, fecha y hora, nombre del destino, y duración.
- Acumulado de llamadas, separadas por originadas y terminadas.

A continuación en la Figura V.4 se muestra la estructura de bloques de una SIM UICC de Alta densidad

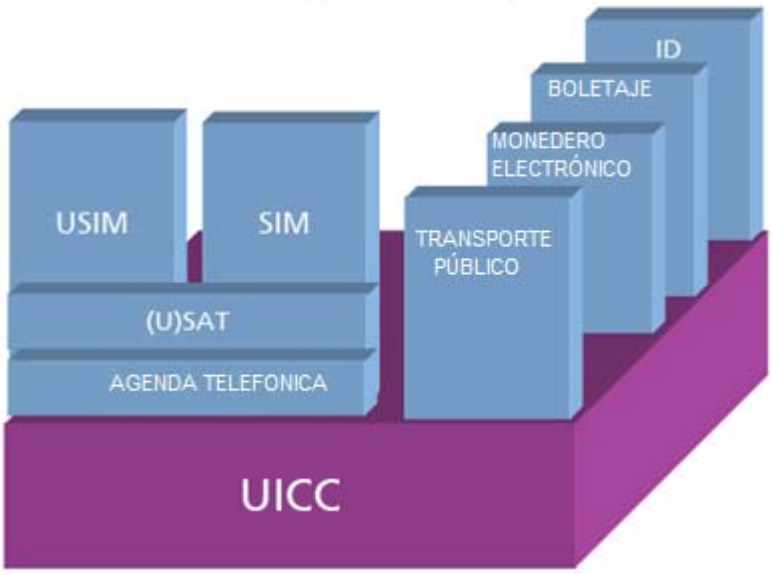


Figura V.4. Estructura de bloques de la tarjeta UICC

Soporte USIM para Aplicaciones

En el campo de la telefonía móvil también se conoce el concepto de SAT (SIM Application Toolkit) en UMTS es USAT (USIM Application Toolkit), que consiste en un conjunto de herramientas para aplicaciones sobre la SIM.

Las aplicaciones Toolkit son una característica opcional tanto de las tarjetas SIM como de las tarjetas UICC (con las aplicaciones correspondientes). Los procedimientos de alto nivel, contenidos y codificación de los comandos, están especificados en la norma GSM 11.14 para GSM y en la norma 3GPP TS 31.111[10] para UMTS.

El UICC debe soportar las características definidas en la especificación USAT (USIM Application Toolkit) definidas en 3GPP TS 31.111, que le permitan interactuar y operar con cualquier terminal que soporte los mecanismos específicos requeridos por la aplicación.

[10] 3GPP TS 31.111. Universal Subscriber Identity Module (USIM) Application Toolkit (USAT). www.3gpp.org/ftp/Specs/html-info/31111.htm

Así mismo debe soportar aplicaciones ISIM para IP Multimedia Services Identity Module (ISIM) application definido en 3GPP 31.103, con la finalidad de asegurar la interoperabilidad entre un ISIM y un terminal independientemente del fabricante, el emisor de la tarjeta o el operador.

Es deseable que SIM cards tengan en máscara o ROM, las implementaciones para soporte de estándares de facto, como:

- SIM Java[11]™ (2.5G)
    - SIM File System (3GPP TS 11.11 / GSM 11.11 Specifications)
    - SIM Toolkit (3GPP TS 11.14 / GSM 11.14 Specifications)
    - SIM API (3GPP TS 43.019 / GSM 03.19 Specifications)
    - SIM Remote Management (3GPP TS 23.048 / GSM 03.48 Specifications)

- SmartTrust[12] WIB™
    - SmartTrust WIB™, version 1.3 – Implementation Specification
    - SmartTrust WIB™, version 1.3 – Client-Server protocol
    - SmartTrust WIB™, version 1.3 – PAD Data Management Plug-in for SmartTrust
- S@T Java™
    - SIMalliance[13] S@T Byte Code (S@T 01.00 Specification)
    - SIMalliance S@T Markup Language (S@T 01.10 Specification)
    - SIMalliance S@T Session Protocol (S@T 01.20 Specification)
    - SIMalliance S@T Administrative Commands (S@T 01.21 Specification)
    - SIMalliance S@T Operational Commands (S@T 01.22 Specification)
    - SIMalliance S@T Push Commands (S@T 01.23 Specification)
    - SIMalliance S@T Browser Behaviour Guidelines (S@T 01.50 Specification)

[11] SIM Java. http://java.sun.com/javacard/
[12] SmartTrust. www.smarttrust.com
[13] SIMalliance. www.simalliance.org

También se conoce el concepto de OTA (Over The Air), que se refiere al acceso remoto a la tarjeta SIM (o USIM). Inicialmente, cuando las tarjetas salían al mercado, el operador no podía modificar nada en ellas, estaban fuera de si alcance. Sin embargo, parecía interesante poder modificar el contenido de algunos archivos en la tarjeta, modificar el perfil de personalización y cargar o modificar aplicaciones Toolkit una vez que la tarjeta estaba en posesión del cliente.

Por tanto, los fabricantes de tarjetas empezaron a incorporar sistemas OTA (de acceso remoto) que permitían gestionar los contenidos de la tarjeta mediante mensajes cortos especiales. Cada fabricante disponía de una solución propietaria incompatible con las de otros fabricantes. Posteriormente, se han generado especificaciones estándar para realizar este tipo de modificaciones remotas OTA[14] . En la actualidad se está trabajando en la definición de estándares que permitan emplear otros tipos de portadoras, no solo mensajes cortos, para hacer comunicaciones más rápidas y flexibles, estas portadoras pueden ser GPRS (General Packet Radio Service), Bluetooth, etc.

V.3 Evolución de la tarjeta SIM de Alta Densidad

Diseñada como un mecanismo de identificación del usuario, las tarjetas SIM han sido una piedra angular dentro de los servicios de la red GSM, desde el principio de la telefonía móvil digital. Esencialmente un tipo de tarjeta inteligente, similar a las utilizadas en las tarjetas de crédito, las tarjetas SIM se han diseñado para extender los servicios básicos tales como el almacenamiento de contactos y mensajes cortos SMS, presentación del menú de servicios SIM de los operadores celulares y comunicación con la red vía mensajes SMS. Algunos de los servicios se configuran de acuerdo al operador móvil, como se muestra en la Figura V.5.

[14] 3GPP 23.048. Security mechanisms for the (U)SIM application toolkit; Stage 2. www.3gpp.org/ftp/specs/html-info/23048.htm

Figura V.5. Menú de servicios SIM de Telcel (Perfil 2 y 2.5 GSM)

Con el lanzamiento de las redes 3G, la SIM se ha transformado en USIM, que se ha diseñado para alojar múltiples e independientes aplicaciones de terceros. Sin embargo, en los últimos años las USIMs han sido rezagadas, detrás del desarrollo de los sistemas operativos de los equipos y los módulos de memoria removibles, en términos de la velocidad de procesamiento, capacidad de memoria, velocidad de comunicación y la interfaz de usuario para la entrega de información, sin embargo has seguido evolucionado poco a poco.

A las tarjetas SIMs de Alta Capacidad, se les ha calificado como MegaSIMs, SuperSIMs, SIMs de Alta Densidad, etc., dependiendo de los proveedores, esta nueva generación de tarjetas inteligentes ofrece entre 4MB hasta 1GB de memoria, un incremento masivo de más de 1000 veces en lo referente a la capacidad de almacenamiento comparadas a las primeras tarjetas SIM. Las tarjetas SIM de Alta Densidad, brindan varias nuevas aplicaciones, tales como almacenamiento y distribución segura del contenido. El paradigma de las tarjetas de Alta Densidad ha sido el avance en la tecnología de fabricación de memorias, iniciada por los proveedores, tales como Spansion, M-Systems, Atmel y Samsung, así como algunos fabricantes de tarjetas como Gemalto (Axalto + Gemplus), Giesecke & Devrient, Oberthur y Sagem Orga. En los inicios del 2006 se empezaron a distribuir las primeras tarjetas de Alta Densidad y en el 2007 alcanzaron una mayor distribución en los operadores europeos.

Algunos de los operadores europeos más importantes como grupo Orange, han empezado a promover el paradigma de las tarjetas de Alta Capacidad, instigados por la

amenaza inminente del Internet abierto y la convergencia entre las Tecnologías de la Información y las Telecomunicaciones.

2006 fue el año en el que el paradigma de las SIMs de Alta Capacidad apareció dentro del mercado celular. Virtualmente todos los proveedores importantes, anunciaron en el Congreso Mundial de 3GSM del mes de Febrero, un producto de nueva-generación, basados en la SIM. Gemplus en ese momento anunciaba ".SIM" (dot SIM), Axalto "U2 SIM", Oberthur anunció su tarjeta "GIGantIC", Giesecke & Devrient su tarjeta "GalaxSIM" y Sagem Orga "SIMply XXL". En ese momento Spansion, un fabricante líder de Memorias Flash, anunció su desarrollo de una nueva clase de memoria Flash segura para SIMs, con capacidad de almacenamiento de 64MB hasta 256 MB y finalmente M-Systems, también productor de memorias flash, desde ese momento, anunció varias sociedades y colaboraciones que consideran extender el rol de la SIM más allá del tradicional papel basado en la red GSM.

Todos los esfuerzos realizados desde ese momento, han superado los límites del dominio de los proveedores de tarjetas SIM, por ejemplo, Intel en el año 2007 anunció una iniciativa en colaboración de la Asociación de GSM, para equipar futuros equipos de cómputo con una ranura (**slot**) para tarjeta SIM. La iniciativa, ha sido conducida por la convergencia entre las Tecnologías de la Información (TI) y las Telecomunicaciones, apuntando a la tarjeta SIM, como el vehículo de autenticación para el acceso a los datos dentro de las redes inalámbricas como 2G, 3G y WiFi. Esta iniciativa está siendo promovida también por grandes operadores celulares, como Orange, Vodafone y Cingular Wireless.

Así mismo, Nokia ha adicionado desde 2007 un Perfil de Acceso vía Bluetooth como característica estándar de sus Smartphones basados en la Serie 60 con software de 3º Edición (los equipos incluyen las serie E de Nokia, el Nokia N71, N80, N91, N92 , N95 y el Nokia 3250). Esto permite que en las terminales, algunas aplicaciones se basen en la identificación de los datos del usuario a través de la SIM Card, como la aplicación utilizada en los modelos de auto Volkswagen, en donde el usuario se identifica y el equipo instalado en el automóvil recibe temporalmente esos datos para generar un

Perfil. Nokia, así mismo, ha incluido una implementación del JSP-177 como una característica estándar de la Serie 60 3º Edición, como un facilitador crítico de seguridad para aplicaciones, basado en SIMs de Alta Densidad.

Del lado de los operadores, a favor del paradigma de las tarjetas de Alta Densidad, Orange en Noviembre del 2005, anunció una colaboración con M-Systems, Oberthur y LG Electronics para el lanzamiento mundial de la primera Sim Card de Alta Densidad de 512 MB, comercializada en 2006 en Francia, inicialmente disponible en el modelo de teléfono LG U8210 con una interfaz de alta velocidad[15]. En la Tabla V.4 se resume todo lo anterior:

Proveedor	Fecha	Lanzamiento
Oberthur	Enero 2005	Oberthur anuncia GIGantIC, una SIM de 128MB que ofrece funcionales de cifrado avanzado para la protección del contenido digital, permitiendo almacenamiento y acceso seguro a los archivos multimedia, a juegos y a ajustes personales.
Axalto	Noviembre 2005	Introduce su SIM U2, ofreciendo características tales como manejo de derechos digitales de contenido y hosteo de blogs de Internet en la SIM.
Gemplus	Febrero 2006	Gemplus lanza .SIM, su línea de tarjetas SIM de alta capacidad, que facilita a los operadores, el ofrecer servicios multimedia a través del móvil, web, línea fija y de Canales de TV.
Sagem Orga	Febrero 2006	Sagem Orga introduce su tarjeta SIM de alta capacidad SIMply XXL, que ofrece a los operadores nuevas capacidades para la personalización y segmentación de los equipos móviles.
Giesecke & Devrient	Febrero 2006	Anuncia la GalaxSIM, que tiene entre 64MB y 512 MB de memoria para contenidos y aplicaciones e incluye su propio servidor WEB.

[15] En el congreso 3GSM del 2006, Orange demostró con el equipo LG U8210 una transferencia de datos de la SIM a una PC, con una tasa de transferencia de 250Kbps, 20 veces más rápido que una tarjeta SIM normal.

M-Systems	Octubre 2006	M-Systems y su subsidiaria Microelectrónica anunciaron la disponibilidad comercial de la MegaSIM de 1GB.
Intel	Febrero 2006	Intel en colaboración con GSMA, anunció su iniciativa de equipamiento para futuros equipos de cómputo con una ranura de SIM, previendo este como vehículo de autenticación para las redes de datos inalámbricas.
Spansion	Febrero 2007	Anuncia el desarrollo de una nueva clase de memoria Flash segura, sobre un chip SIM, con capacidad de almacenamiento de 64MB a 256MB.
Oberthur	Febrero 2007	Lanza su oferta de TV Móvil sobre su tarjeta SIMphonIC™
Giesecke & Devrient	Febrero 2007	Lanza GalaxSIM de 1GB incorporando gráficos, logos, juegos y una aplicación de libreta de direcciones llamada GalaxSIM Sync.

Tabla V.4 Importantes lanzamientos de tarjetas y aplicativos para SIMs de Alta Densidad.

Sagem Orga SIMply XXL

Oberthur GIGAntIC

Gemplus .SIM

Gieseck & Devrient GalaxSIM

Axalto U2SIM

Figura V.6. Primeras SIMs de Alta Densidad

Los fabricantes de las tarjetas SIM de Alta Densidad utilizan la ROM para el almacenamiento de las funciones del sistema operativo y de la seguridad de la tarjeta y la memoria Flash se utiliza para las variaciones dependientes de cada operador como aplicaciones, archivos y deltas del sistema operativo. El almacenamiento de las Flash puede ser re-escrito mucho más rápida y fácilmente, aun cuando ya este en venta, es decir puede ser modificado durante todo el tiempo de vida de la SIM. Esto contrasta con el almacenamiento ROM, que requiere un tiempo de proceso en las máscaras. Es por ello que la mayoría de los fabricantes de tarjetas están cambiando el almacenamiento ROM por Flash, incluso para las tarjetas SIM ordinarias, debido a la flexibilidad de uso de los datos, ya que los costos son similares.

La capacidad de memoria creciente de las tarjetas SIM de Alta Densidad se ha hecho posible debido a los avances en el almacenamiento de las memorias Flash. El primer chip Flash, fue tipo NOR y fue introducido por Intel en 1988, seguido por la NAND Flash de Samsung y de Toshiba en 1989. La NAND Flash es más rápida respecto a los tiempos de borrado y escritura, así mismo, tiene mayor densidad y menor costo por bit que una NOR Flash. Sin embargo, permite solamente el acceso secuencial a los datos (en comparación con el acceso aleatorio). La NAND Flash es la tecnología de memoria detrás de las tarjetas removibles como las SD y las tarjetas HC SIM, sin embargo ya se están introduciendo nuevas tecnologías, por ejemplo la tecnología MirrorBit ORNAND de Spansion para las tarjetas SIM de Alta densidad, que combinan los beneficios de la NAND y la NOR.

V.4 Beneficios de la SIM de Alta Densidad

La llegada de las SIMs de Alta Densidad marca una evolución en la tecnología de la SIM. La limitación en la capacidad de almacenamiento, ya no es un obstáculo para el lanzamiento de servicios multimedia a través de la SIM. Las tarjetas SIM de Alta Densidad se diseñan para incorporar 64MB a 1GB de memoria Flash, un incremento de más de 1000 veces en la capacidad de almacenamiento comparadas con las SIMs

tradicionales. Las Sims de Alta Densidad tienen como característica un canal de comunicación de alta velocidad hacia el equipo terminal en forma de USB o interfaz MMC y en algunos casos tienen un protocolo de comunicación de alta velocidad con la red, llamado Bearer Independent Protocol (BIP- Protocolo Independiente del portador). Así mismo, las tarjetas SIM de Alta Densidad puede incorporar características de seguridad avanzadas como funciones de cifrado asimétrico, arquitectura de núcleo sencillo (single-core), etc.

Las tarjetas SIM de Alta Densidad conservan compatibilidad en los estándares 3GPP, pero permiten la introducción de una nueva gama de aplicaciones avanzadas. Además de permitir el almacenamiento portable de contenido multimedia como videos, imágenes y música, las tarjetas SIM de Alta densidad están bien adaptadas para proporcionar seguridad a los contenidos almacenados a través del manejo de DRM, certificados digitales y tockens, en una forma centralizada, por lo cual tanto el contenido como los derechos pueden exportarse en diversos equipos terminales.
Siendo que las SIMs de Alta Densidad, además de estar desarrolladas de forma principal para el almacenamiento, tienen desarrollos de seguridad, estos pueden utilizarse dentro de aplicaciones en ambientes corporativos donde las capacidades de cifrado, como la generación de un password único one-time password (OTP) tienen interés primario.

Las tarjetas SIM de Alta Densidad también hacen frente a desafíos importantes: Costo-beneficio, la capacidad adicional impone un aumento significativo en el costo de las tarjetas, que se espera que tenga un precio promedio de $10 USD en un plazo mediano. Y se estima que el costo de las SIMs de Alta Densidad será aproximadamente 30% mayor que el costo del Silicio (y por tanto del costo de las tarjetas de memoria SD y MMC). El incremento es debido al costo agregado del sistema operativo en la tarjeta, la incorporación de aplicaciones y el costo del mantenimiento para la compatibilidad de aplicaciones y estándares anteriores. Y más importante, por lo tanto para la viabilidad de las SIM de Alta Densidad es el costo del

silicio (para memoria y procesador) se mantenga en un rango aceptable en rangos de 12 a 24 meses.

La complejidad es otro factor que hacer frente, debido a que estas tarjetas cuentan con un nuevo sistema operativo desarrolladas para abastecer la capacidad creciente de la SIM. La interoperabilidad con el equipo terminal sigue siendo un asunto complicado, dado que pocos modelos de equipos apoyan las interfaces de red de alta velocidad de la SIM como el protocolo BIP e interfaces de comunicación con el equipo como la USB y la MMC.

Las tarjetas SIM de Alta densidad, también hacen frente a la competencia con las tarjetas de almacenamiento removibles como la MMC, SD y MemoryStick. Los almacenamientos removibles solo están disponibles para un pequeño subconjunto de equipos terminales, en 2004, alrededor del 9.2% de las ventas de teléfonos tenían al menos un tipo de ranura para memoria. Este porcentaje se elevó en el 2005 con el 13.4% y se espera que para el 2010 el 65% de los terminales tengan como característica al menos una ranura para memoria flash[16]. Sin embargo, en los últimos meses, se ha observando una promoción inminente de las tarjetas removibles (MMC y SD), de parte de algunos fabricantes de equipos.

Los Sistemas Operativos abiertos en los teléfonos tienen ya interfaces programadas para el acceso al contenido almacenado sobre las memorias removibles, contrariamente a la situación que acontece con las tarjetas SIM (aunque las tarjetas de Alta Densidad tratan esta limitación a través de su diseño). Además, los estándares de las tarjetas de memoria se están desarrollando para competir en DRM (Digital Right Management), VPN (Virtual Private Network) y otras aplicaciones que requieren almacenamiento seguro. En Marzo del 2004, la Asociación de MMC anunció la formación de un equipo de trabajo para agilizar la adopción de SecureMMC V2.0 como la base para la especificación de OMA DRM 2.0 para el almacenamiento seguro de los medios y para las aplicaciones VPN que requieran un almacenamiento seguro de los Token PKI. Samsung ofrece su desarrollo SecureMMC basado en los estándares OMA,

[16] De acuerdo a la empresa Informa Telecoms

mientras que 4cEntity[17] ofrece un esquema basado en la protección de los contenidos multimedia registrado (CPRM – Content Protection for Recordable Media) utilizado en las tarjetas SD.

Así mismo, es importante notar que la capacidad de remover instantáneamente las tarjetas SD, MMC y Memory Stick les brindan un rol complementario al tomado por la SIM que debe ser removida pero con el compromiso de deshabilitar las conexiones de red. En última instancia, la SIM y las tarjetas seguras removibles pueden trabajar no en competencia, sino como complementos, interactuando con los teléfonos y la red en muchas de las aplicaciones cliente-servidor como el DRM. En otro escenario potencial, las tarjetas SD/MMC pueden ser usadas para un ultra almacenamiento no seguro (arriba de 1GB) mientras que la SIM de Alta Densidad puede ser usada como alto almacenaje seguro (hasta 1GB). No obstante, has muchos factores (a menudo políticos) que influencian el papel futuro de las tarjetas SD/MMC contra las tarjetas SIM de Alta densidad, por lo que el resultado sigue siendo incierto.

Por último, las memorias de Alta Densidad compiten con las memorias insertadas directamente en los dispositivos móviles para aplicaciones de almacenamiento de datos de usuarios (aunque no aplica para personalización de menús). La mayoría de los teléfonos de alto nivel (como smartphones) vienen con la memoria Flash arriba de 16MB. Pero se espera que para el 2010 la capacidad mínima sea de 64MB.

[17] 4cEntity: http://www.4centity.com/

PARTE III. APLICACIÓN

CAPÍTULO VI
APLICACIÓN DE LA SIM DE ALTA DENSIDAD

Las tarjetas SIM nunca fueron diseñadas como plataformas independientes de distribución de servicios. En lugar de ello, su papel ha sido siempre un facilitador, es decir, una pieza dentro del rompecabezas en los servicios punto a punto de los operadores móviles. Como tal, el éxito de la SIM ha sido dependiente de la colaboración de tres entidades: el fabricante de la tarjeta SIM, el operador móvil y el fabricante de equipos móviles.

El operador móvil ha sido siempre la fuerza primaria detrás de la evolución de la SIM, ya que la tarjeta SIM es la manifestación física de un suscriptor. La SIM es también un canal de distribución de los servicios de información enteramente proporcionados por el operador móvil.

Los fabricantes de tarjetas SIM han sido naturalmente incentivados a promover la evolución tecnológica de la SIM. Los fabricantes de memorias y micro controladores para la SIM son una parte fundamental en la cadena de valor, pero son poco influyentes en lo que se refiere a como manejar el mercado, ya que estos requerimientos siempre han estado basados en los requisitos de los proveedores de tarjetas.

Los fabricantes de equipos, por otra parte, no han sido totalmente beneficiados por el adelanto en el desarrollo de la SIM. Esto es porque, tradicionalmente, la SIM ha estado firmemente en control del operador móvil y ha sido usada para desarrollar aquellos servicios de los operadores, que entran en conflicto con la agenda de los fabricantes de equipos. Como resultado, los fabricantes no han sido motivados para invertir en la convención de estándares para sus equipos móviles, dando por resultado una amplia variación en la implementación de la SIM en los equipos móviles. Y en años recientes

el duelo por definir servicios finales para el usuario, entre el operador y el fabricante de equipos, se ha venido abajo, dejando el control al operador móvil.

Paradójicamente, un nuevo jugador se ha unido a la cadena de valor de la SIM para desarrollar el papel que desempeña esta. SmartTrust, pionera del navegador basado en SIM, en los inicios del Internet móvil, antes de que el GPRS y WAP se convirtieran en una parte integral de los equipos móviles, lanzó en 1999, una especificación inalámbrica de Navegación sobre Internet (Wireless Internet Browsing- WIB), en la cual se disponía de un sistema propietario de protocolos, de comandos y de guías para desarrollar un micro navegador basado en la SIM. SmartTrust tuvo éxito, haciendo a WIB[1] una característica "estándar" en las tarjetas SIM. De acuerdo a SmartTrust, la tecnología WIB está integrada actualmente en más de 200 millones de tarjetas SIM. Su última versión, WIB 1.3 ha sido especificada como característica mandatoria para la mayoría de los operadores GSM. Las herramientas de la SIM Alliance (Sim Alliance's Toolbox –S@T) es una iniciativa de especificación similar, no obstante esta ha sido rechazada por la mayoría de los fabricantes de tarjetas SIM, que ha logrado solamente una pequeña parte del mercado de los micro navegadores.

Finalmente, los proveedores de contenido han tomado interés en el papel de la SIM, en particular, debido a su papel dentro de la protección del contenido y DRM. No obstante, como la SIM está enteramente dentro del control del operador, es poco probable que los proveedores de contenido tomen un papel directo en la cadena de valor de la SIM.
Ellos pretenden, no obstante, tener un papel al menos indirecto, insistiendo para que los operadores móviles manejen dentro de sus acuerdos de nivel de servicio, un control para la distribución de contenido, sin importar cuales componentes tecnológicos deseen utilizar, ya que las pérdidas por distribución ilegal de contenido, son la mayoría de las veces absorbidas por los proveedores de contenido.

[1] WIB. Wireless Internet Browser. Mobile Browser

La cadena de valor está formada entre los siguientes jugadores. El siguiente diagrama de la figura VI.1 representa las relaciones de compra-venta y la influencia entre los jugadores del ecosistema.

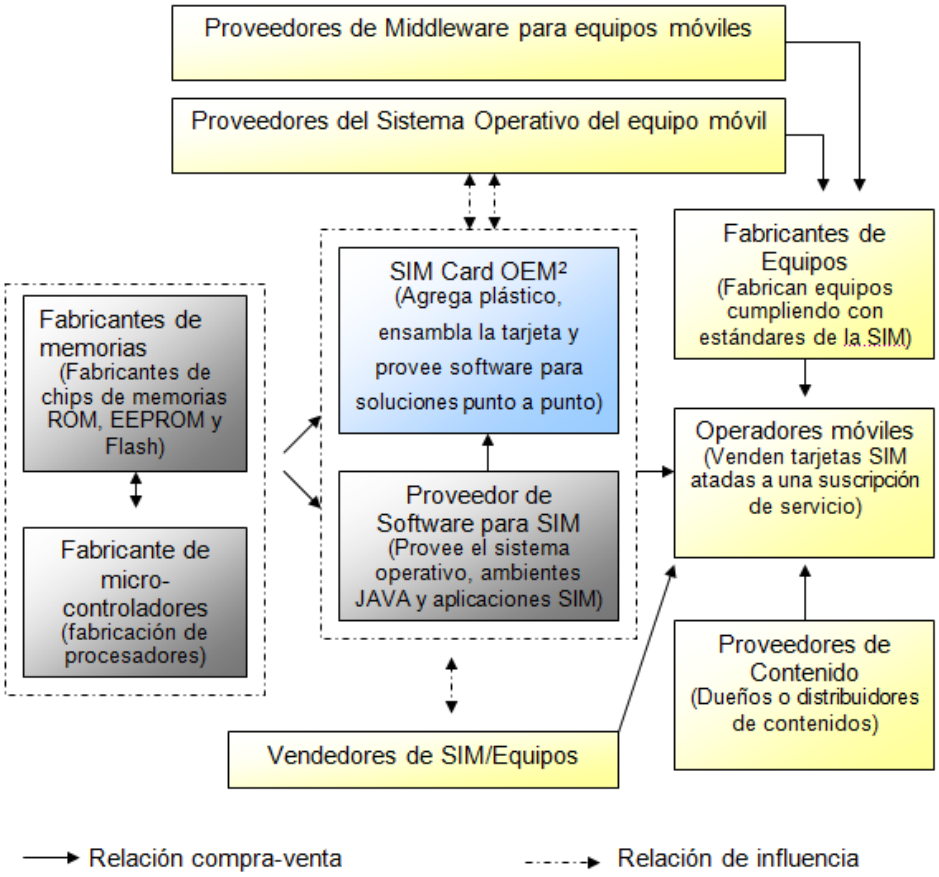


Figura VI.1 Cadena de valor dentro del ecosistema de la tarjeta SIM

Dentro del diagrama, el papel del proveedor del software de la SIM, abarca el Sistema Operativo, los proveedores de JavaCard, proveedores de aplicaciones SIM y proveedores de software intermediario.

VI.1 Ecosistema de los servicios móviles

Los operadores móviles tienen la mayor ganancia dentro del mercado de la SIMs de alta capacidad. En primer lugar, cuando se utilizan como medio de almacenamiento central de los archivos multimedia de los usuarios, las SIMs de alta capacidad pueden influenciar la decisión de la compra del equipo a favor de la marca del operador y proporcionar un enlace con el operador durante el tiempo de vida del suscriptor

[2] Original Equipement Manufacturer- Fabricante Original del Equipo.

reduciendo el churn (abandono de clientes). En segundo lugar, las SIMs de alta capacidad pueden actuar como medio de distribución eficaz de los servicios modificados especialmente para los operadores, de acuerdo a sus requisitos, ya que es logísticamente más fácil modificar las tarjetas SIM, que los propios equipos celulares. Finalmente, las SIMs de alta capacidad son un juego estratégico de largo plazo para los operadores móviles, protegiendo el valor del operador en cuanto aparezca el Internet abierto y la convergencia de las Tecnologías de la Información y los sectores de Telecomunicaciones, ya que el operador tendrá el poder de permitir o censurar los accesos de sus suscriptores a estas tecnologías a través de las especificaciones de las SIMs.

Al mismo tiempo, la carencia de una masa crítica de equipos que apoyen la interfaz USB o MMC y el alto costo inicial de las tarjetas SIM de alta capacidad puede hacer que algunos operadores vacilen sobre la adopción del paradigma de las SIMs de alta capacidad. Incluso para los operadores más innovadores, se requerirá una ayuda extra para la modificación de los equipos, lo que implicaría un costo adicional al operador. Generalmente, los operadores alientan a hacer mejoras en la interoperabilidad entre los equipos y la SIM, pero en este caso, pueden tomar un papel conservador y no promover a las tarjetas de alta densidad, con el fin de evitar costos de inversión adicional.

Mientras que la energía de la industria ahora se reclina con los operadores, los fabricantes continuarán buscando las características y nuevos servicios de los equipos que agreguen valor y los distingan de los dispositivos de los competidores. En consecuencia, los fabricantes de equipos GSM quienes tradicionalmente han mantenidos una relación estrecha con el operador, están adaptando sus equipos con el fin de soportar las características avanzadas de las tarjetas de alta capacidad. Mientras que otros fabricantes como Nokia han mostrado apoyo también a las implementaciones de JAVA, como en su modelo S60 3era edición. Tales movimientos, contribuyen indudablemente a la aparición de las tarjetas de alta capacidad.

Los proveedores de contenido están naturalmente, entusiasmados de ver la aparición de los mecanismos end-to-end para asegurar y controlar la distribución del contenido. En este caso, las tarjetas de alta densidad pueden jugar un papel importante en este respecto, actuando como un almacenaje seguro, para los contenidos y para las reglas de control de los derechos de esos contenidos.

Por otra parte, las tarjetas de alta densidad brindan varias ventajas en comparación a las tarjetas de almacenamiento removible. Las SIMs combinan portabilidad dentro de los equipos, aunque está atada a la identidad de un suscriptor y por lo tanto no puede ser compartida prácticamente. Además, las tarjetas SIM se restringen en el mundo móvil, al contrario de las tarjetas removibles, que se pueden transferir a una PC, un hecho que aumenta la protección del contenido a beneficio del proveedor de contenidos móvil. En general, la ubicuidad y la seguridad de las tarjetas de alta densidad, cuando están combinadas con un modelo compatible de utilidad, puede ser una alternativa segura para la distribución del contenido.

Análisis FODA (Fortalezas, Oportunidades, Debilidades y Amenazas) – SIMs de Alta Densidad.

Fortalezas:

- Almacenamiento portable de contenido multimedia como videos, imágenes y música, cuando muchos de los usuarios están familiarizados con el uso de la SIM.
- Almacenamiento central de DRM (certificados, derechos) y contenidos, certificados.
- Bien adaptado para el operador – modificaciones y personalización de acuerdo a requisitos particulares.
- Diseñado para almacenamiento seguro de multimedia y cumple con estándares de seguridad de tarjetas inteligentes.

- Las tarjetas de alta densidad encontrarán aplicaciones en ambientes corporativos donde la seguridad del almacenamiento y las capacidades de cifrado son de interés primario.
- Las tarjetas de alta densidad pueden actuar como reemplazo de tarjetas de memoria removibles y como costosos equipos con memoria abundante.

Oportunidades:

- Las tarjetas SIMs de alta densidad se beneficiaran de la fuerte influencia que los operadores móviles tienen en la industria.
- La fusión de Axalto y Gemplus (marcas líder), generó que la industria se alineara con la agenda y desarrollos marcados por ellos.
- Los fabricantes de equipos como Nokia, están lanzando equipos con soporte JAVA. Estos aprovisionaran la demanda de las tarjetas de alta densidad para adaptarse a los clientes corporativos.

Debilidades:

- La adopción de las tarjetas en la industria, depende de la masa crítica de los equipos que apoyen la interfaz USB o MMC.
- El costo de la SIM es probablemente lo doble, debido a la memoria adicional.
- Inadecuado para procesos intensivos de uso.
- La interfaz de usuario es menos atractiva que la interfaz de usuario de los equipos.

Amenazas:

- Las tarjetas de alta densidad tienen una oportunidad relativamente estrecha dentro de la competencia contra tecnologías de reemplazo, tales como tarjetas removibles de almacenamiento.
- Las nuevas soluciones de funcionalidad y seguridad en los equipos móviles (por ejemplo escáner de huellas dactilares, GPS, antivirus) no implican a la SIM, lo cual puede proliferar.

VI.2 Servicios y contenidos multimedia basados en SIM de alta densidad

Habiendo analizado el mercado y los desafíos que enfrentan las tarjetas de alta capacidad, el foco de la discusión en este momento se centrará en los mercados dominantes donde las tarjetas de alta densidad deberán tener impacto, específicamente dentro de mercados como la TV móvil y Juegos Móviles, cuyo desarrollo en el mercado masivo es de los más poderosos y crecientes. Así mismo, se evaluara el uso de las tarjetas de alta densidad de acuerdo a su capacidad de facilitador, como almacenamiento de datos multimedia, distribución de servicios y gestión de derechos digitales.

VI.2.1 Mobile TV

Se dijo que en 2004, cuando apareció 3G (Tercera Generación de la Telefonía celular), la primera noción de esta generación se manifestó en la video-llamada, la siguiente en las tarjetas de datos de alta velocidad y desde 2006 se manifestó en la TV móvil. De hecho, la TV móvil ha estado generando interés en muchos sectores en las industrias móviles y de entretenimiento desde entonces, por lo que muchos desarrolladores han buscado diseñar la mejor aplicación que ayude al desarrollo de este servicio, por lo que los desarrolladores de SIMs no se quedaron atrás y han propuesto el ecosistema ideal para que las tarjetas sean un componentes estratégico.

Hay un gran número de estándares para la TV móvil, como DVB-H[3], DBM, ISDB[4] y Mediaflo[5], tecnologías que introducen diferentes formas de difusión, modelos de negocios y puntos de control dentro de la cadena de valor. Un punto diferenciador para

[3] DVB-H. (*Digital Video Broadcasting Handheld*). constituye una plataforma de difusión IP orientada a terminales portátiles que combina la compresión de video y el sistema de transmisión de DVB-T, estándar utilizado por la TDT (*Televisión Digital Terrestre*).
[4] ISDB. (Integrated Services Digital Broadcasting) Estadandar japones para la televisión digital y radio digital
[5] MediafFLO (Forward Link Only) es una tecnología desarrollada por la empresa Qualcomm para la radiodifusión de televisión móvil a dispositivos portátiles y que se utiliza tan sólo en Estados Unidos. Esta tecnología permite la radiodifusión de canales en tiempo real, en tiempo no real, audio o transmisiones de datos IP.

cada tecnología que relaciona el modelo de negocio propuesto a cada uno, es si la red de datos móvil puede ser usada para entrega de contenido de TV y/o la aplicación agrega interactividad a la programación de la TV. Lo cual atrae a algunos de los jugadores más importantes dentro de la industria del entretenimiento como Disney, Time Warner y Sony Entertainment, hacia la industria móvil.

El Ecosistema de la TV Móvil

Muy parecido al ecosistema de TV digital terrestre, la cadena de valor de la TV móvil consiste en cuatro áreas de calor básicas, como se muestra en la figura VI.2:

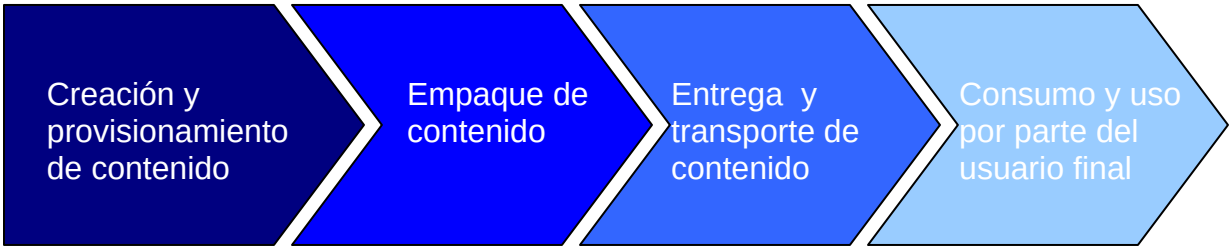


Figura VI.2 Cadena de valor de la TV móvil

La creación de contenido y el aprovisionamiento se refieren a la creación y distribución del contenido original, el cual es provisto por los estudios cinematográficos, sello discográfico, distribuidores de señales y los abastecedores de marcas de propiedad intelectual.

El empaquetado de contenido se refiere a la adición y publicación de contenido para la distribución, que es proporcionada por los distribuidores de señal, proveedores de canales de televisión, operadores de cable y satélite.

La entrega y transporte de contenidos se refiere a la entrega física del contenido sobre una red del operador móvil.

La pieza final de la cadena de valor se refiere al usuario final y a su consumo del contenido de TV.

La fortaleza de los operadores móviles, proveer de un canal de vuelta para la interactividad y con ello entablar una relación con el usuario final, donde son responsables del aprovisionamiento del servicio, de la facturación y del cuidado del cliente. Además, los operadores móviles tienen a menudo una influencia fuerte en especificaciones del equipo móvil, la distribución y el subsidio de los equipos móviles. Están también, bien posicionados para proporcionar servicios de localización y DRM. Por otra parte, los operadores móviles están confiados en los dueños y proveedores del contenido, para la administración de contenido original de marca, lo cual es lo que genera la atracción de los usuarios.

Un punto clave en la TV móvil es como los servicios puedes ser cobrados. Los escenarios más usados son pay-per-view (paga por ver) y modelos de suscripción, los cuales son fáciles de implementar en sociedad con los operadores móviles quienes ya tienen una relación de facturación con el usuario final.

Los operadores móviles son responsables de implementar las ventajas de las tarjetas de alta densidad en usos de TV móvil. Los primero dos jugadores dentro de la cadena de valor (por ejemplo. Creación de contenido y empaque de contenido) no están anticipados a tener un interés muy grande en la SIM de alta capacidad- como jugadores impondrán muy probablemente las condiciones de autorización del contenido, con las cuales los operadores móviles deben conformarse. El uso del SIM para las aplicaciones de la TV móvil, incumben por lo tanto a la última instancia de la cadena: a los operadores móviles.

Rol de la SIM

En primer lugar, el rol de las tarjetas de alta densidad puede parecer periférico al servicio de TV móvil, perteneciendo solamente a la autenticación del suscriptor. Sin embargo, el valor que ofrecen al mercado de TV móvil las tarjetas de alta densidad se

aplican a varias áreas, como DRM, aprovisionamiento de servicio y gestión de servicio. Específicamente las tarjetas de lata densidad se pueden utilizar:

- Almacenamiento seguro de los derechos de los contenidos de TV como parte de un esquema punto-a-punto de DRM y permite que estos sean manejados por el operador móvil mediante OTA. Esto ofrece portabilidad de derechos de contenido cuando el usuario cambia de diferente equipo, aunque no se extiende el caso, cuando un usuario cambia por otro operador. Esta tecnología está siendo utilizada desde 2005, por T-Mobile en la República Checa, donde utiliza la SIM como elemento llave para el acceso seguro de los derechos dentro de su servicio interactivo “TV en tu bolsillo”, basado en tecnología DVB-H.
- Las tarjetas de alta densidad pueden utilizar un algoritmo para generar la llave del contenido, para la protección de este. Las tarjetas de alta densidad se colocaron para entregar esta funcionalidad comparada al software del equipo,
- Aprovisionamiento y manejo de características del servicio, como características de canales de datos interactivos. Las tarjetas SIM pueden fácilmente modificarse para diversos segmentos de usuarios finales, con ello se hace un vehículo bien adaptado para almacenar los ajustes del servicio de TV móvil.
- Aprovisionamiento, almacenamiento y gestión vía OTA de la Guía de Programación Electrónica (EPG-Electronic Program Guide) y de los datos del perfil del usuario relacionados con el servicio. Una vez almacenado en la tarjeta SIM, el EPG puede ser fácilmente personalizados vía aire y en consistencia con la moda de los equipos móviles.

Naturalmente, las tarjetas de alta densidad necesitan ser adaptadas para acomodar los escenarios de TV móvil mencionados. Eso es esencialmente que la tarjeta de alta densidad soporte protocolos de comunicación de alta velocidad con el equipo móvil mediante USB o MMC, ya que los protocolos tradicionales son notoriamente lentos para la mayoría de las aplicaciones de datos. Además, para que la tarjeta SIM actúe como agente seguro, se requiere establecer un enlace seguro en el agente del aparato

de lectura de medios y gestionar el agente residente en el equipo, para poder entregar una solución combinada de seguridad con hardware y software.

Hay también algunas trampas en el uso de la SIM de alta densidad en el uso de la TV móvil. En primer lugar, la gestión del contenido basado en la SIM como el servicio de programas y la guía EGP, requerirá el uso de una infraestructura propietaria de SIM OTA a los operadores. En la práctica, se considera que los ajustes de los perfiles de los usuarios y del uso de guías EGP serán residentes en el equipo, mientras que el servicio que contiene la información estará almacenado en la tarjeta SIM.

Análisis FODA- SIMs de Alta Densidad como facilitador de la solución de TV móvil.

Fortalezas:

- Bien adaptado para almacenar derechos de contenido para programación de TV.
- Puede ser adaptado fácilmente para generar llaves de descifrado del contenido.
- Adaptado para el aprovisionamiento y la gestión de los ajustes del servicio.
- Adaptado para el aprovisionamiento personalizado de Guías Electrónicas de Programación.

Oportunidades

- Incrementa la seguridad y la fiabilidad del uso de aplicaciones de TV móvil.
- Aumenta los puntos de control del operador en la cadena de valor de los servicios de datos móviles.

Debilidades:

- Requiere la adopción consistente del protocolo para la comunicación entre equipo-SIM (USB o MMC).
- Requiere canales de comunicación seguros entre la aplicación de TV móvil en el equipo y la aplicación relacionada en la SIM de Alta densidad, que alternadamente requiere una combinación de hardware y de software (esto todavía está en la etapa pre-comercial).

Amenazas:

- La TV móvil sufre ya de un proceso divergente de casos de negocio y fragmentación de tecnología. Esto todavía tiene que ser probado como un servicio generador de ganancias en un caso de negocios en que ganen todas las partes involucradas.

VI.2.2 Juegos Móviles

Los juegos móviles constituyen una de las fuentes de ingresos principales para los operadores móviles, después de la voz y de la mensajería. Un problema grave en la industria de los juegos móviles es la variedad en las plataformas de los equipos móviles, en las especificaciones del equipo, en los requerimientos del operador y en la diversidad de canales e idiomas que deben ser soportados. El mercado de juegos móviles es también uno donde los márgenes son pequeños, por lo tanto la experimentación en nuevas tecnologías como el uso de la SIM de alta densidad en el uso de juegos móviles requiere una inversión en desarrollo y una gestión fuerte para conducirla y que se materialice. El rol de la SIM de alta densidad es probablemente ser marginal en las aplicaciones de los juegos móviles.

La cadena de valor de los juegos móviles se ha alargado considerablemente durante los últimos tres años. Los protagonistas en la industria de juegos móviles incluyendo a los operadores móviles, editores de juegos de video, marcas, editores de juegos móviles, desarrolladores de juegos y proveedores de la solución para la gestión del servicio.

Los editores de los juegos de video siguen una ruta indirecta en el mercado, mediante el licenciamiento de su propiedad intelectual hacia los editores de juegos móviles o desarrolladores. Muchas marcas y compañías de medios también han escogido un acercamiento indirecto en el mercado de los juegos móviles. Los operadores siguen siendo el canal por el cual la mayoría de los juegos móviles son vendidos al consumidor y es por lo tanto un enlace principal dentro de la cadena de valor.

Rol de la SIM

Actualmente, el papel de las tarjetas SIM en el mercado móvil es extrínseco. Los juegos basado en la SIM y en SMS estuvieron en boga del 2001 al 2003, pero su renombre se ha opacado, pues los juegos basados en la plataforma que otorga el equipo han gestionado la entrega de una experiencia rica al usuario, después de que la tarjeta SIM nunca estuvo diseñada como plataforma de juegos, tal como algunos equipos móviles si lo están. Recientemente, los servicios de publicación basados en SIM y las aplicaciones en micro-navegadores han jugado un papel cercano dentro de los juegos móviles y en servicios móviles en general, proporcionando una lista personalizada de servicios que pueden ser restaurados vía aire.

En la nueva era de las tarjetas de alta densidad, los juegos móviles continúan siendo servicios circundantes relacionados a la SIM debido a:

- Las tarjetas de alta densidad se pueden utilizar para almacenar el perfil del usuario, las puntuaciones y los ajustes del juego. Aunque tal concepto se haya demostrado en el pasado, no hay razón por la que los datos relacionados al juego no puedan continuar residiendo en la memoria del equipo móvil, particularmente cuando los desarrolladores de los juegos tienen poco incentivo (financiero, técnico, o de otra clase) para cambiar esta situación. Adicionalmente, el ciclo vital de un juego descargado es mucho más corto que la vida útil del equipo móvil, lo que implica que la SIM de alta densidad es poco probable para ser empleada para almacenar datos relacionados a un juego y ser transportado a un nuevo equipo móvil.
- Las tarjetas de alta densidad puede ser utilizadas como mecanismo de distribución, como en el caso de aplicaciones personalizadas por el operador. No obstante, en la práctica los juegos de gama alta, tales como Handy Games' Townsmen, EA FIFA World Cup, vienen con cerca de 300 variaciones para diferentes equipos, adicionalmente a estas variaciones se requiere cumplir con diversos requerimientos por parte del operador y de lenguajes. En la práctica, para empacar todas las variaciones de un juego dentro de un paquete de distribución vía SIM de alta densidad, serían necesarios varios centenares de

megabytes de almacenamiento, que es claramente no factible, incluso con las tarjetas de alta densidad. Además, el tiempo de vida de un equipo es totalmente diferente al tiempo de vida de la SIM, lo cual implica que los desarrolladores de juegos no podrían enviar su juego en una tarjeta SIM y asegurarse que cumpla con los equipos que se distribuyen en los siguientes 12 a 24 meses.

- Las tarjetas de alta densidad podrían ofrecer características de seguridad para los juegos móviles, pero nuevamente esto no es algo demandado por los desarrolladores de juegos móviles. No obstante, a largo plazo, asumiendo que los derechos digitales prevalezcan basados en la SIM, puede ser admisible que tales mecanismos de DRM también serán empleados para proteger los intereses de los dueños de los derechos dentro de la distribución de los juegos móviles.
- Podría ser discutido que las tarjetas de alta densidad podrían utilizarse para facilitar los micro-pagos para ofrecer compras extras, como niveles extra en los juegos, un nuevo personaje, etc. No obstante, como este tipo de facturación esta basado mediante pagos directos en la red o SMS Premium, los micro pagos basados en la SIM no brindarían ningún beneficio a los juegos móviles.

Finalmente, los desarrolladores de juegos móviles están haciendo frente a los pequeños marines y están diversificando su base de usuarios a través de sus territorios, con el fin de sostener el crecimiento en su negocio. Como resultado, ellos están renuentes a invertir en nuevas plataformas de desarrollo como la SIM de alta densidad, sin tener clara un beneficio financiero o soporte del operador.

Análisis FODA – SIMs de Alta Densidad como facilitador en aplicaciones de juegos móviles.

Fortalezas:

- Utilizados como facilitador para descubrir juegos residentes en el portal del operador.
- Uso limitado como medio de distribución del servicio para promover el uso de juegos.

Oportunidades:

- Las tarjetas de alta densidad pueden estimular su potencial futuro en esquemas de DRM basados en la SIM para ampliar su papel en juegos móviles.

Debilidades

- Opción no muy efectiva como medio de distribución o como almacenamiento de características relacionadas al juego.

Amenazas:

- Ninguna- el juego y las tarjetas de alta densidad se relacionan solo de forma externa.

VI.2.2 Almacenamiento seguro (DRM) para contenido multimedia

Almacenamiento de contenido multimedia

Hay una gran discusión que rodea al uso de la tarjeta SIM de alta densidad como medio de almacenamiento de los contenidos multimedia de los usuarios, como videos, imágenes y música. Algunos operadores móviles, especialmente europeos, han lanzado programas, en donde venden equipos móviles con poca memoria interna (gama de teléfonos baja, con aproximadamente 32 MB de memoria), insertando tarjetas de alta densidad de 128 MB y han tenido resultados favorables respecto a las ventas de este tipo de modelos, es por ello que estos operadores han seguido desarrollando programas a favor de las tarjetas de alta densidad.

Sin embargo, al mismo tiempo se debe observar, que para transferir archivos multimedia de la tarjeta SIM de alta densidad hacia una computadora, con el fin de archivarlos o compartirlos, es necesario que el usuario tenga comprensión de estas tecnologías, por lo que el éxito de las tarjetas SIM como medio de almacenamiento

dependerá del segmento de usuarios finales a los que la SIM sea impulsado. Como se muestra en la figura VI.3.

Figura VI. 3 Portabilidad de archivos, información, derechos digitales mediante la HD-SIM

DRM (Digital Rights Management) Gestión de Derechos Digitales

Digital Rights Management es un mecanismo para hacer cumplir la protección de los intereses de los dueños de los contenidos, pues el contenido se utiliza y se comparte a través de usuarios, dispositivo y redes. Brindando la facilidad del almacenamiento y de la portabilidad seguras de las tarjetas SIM. Las tarjetas SIM de alta densidad son utilizadas para almacenar los derechos en algunos esquemas de DRM.

Las tarjetas de alta densidad se diseñan para los equipos móviles de gama alta porque permiten un número de aplicaciones multimedia que son claves para el desarrollo del mercado del contenido. La colocación del DRM dentro del controlador de la tarjeta inteligente y los archivos multimedia dentro de la memoria Flash de la tarjeta de alta densidad, aseguran la seguridad del contenido en el contenedor, la HD-SIM. Esto provee al usuario la portabilidad necesaria para transferir archivos multimedia protegidos desde un equipo móvil a otro.

Un nivel adicional de protección de contenido se alcanza con la capacidad de la tarjeta de Alta Densidad de generar llaves cifradas para el contenido utilizando un algoritmo

interno. Por tanto, a diferencia de las tarjetas de memoria removibles, cuyo diseño primario es el almacenamiento en lugar de la seguridad, la HD-SIM aumenta las ventajas tradicionales de la tarjeta SIM –tarjeta inteligente con alta seguridad y resistencia a ser alterado- para beneficio de usuarios, operadores y proveedores de contenido. De acuerdo a las diferentes tecnologías de DRM que existe, como se muestra en la Figura VI.4.

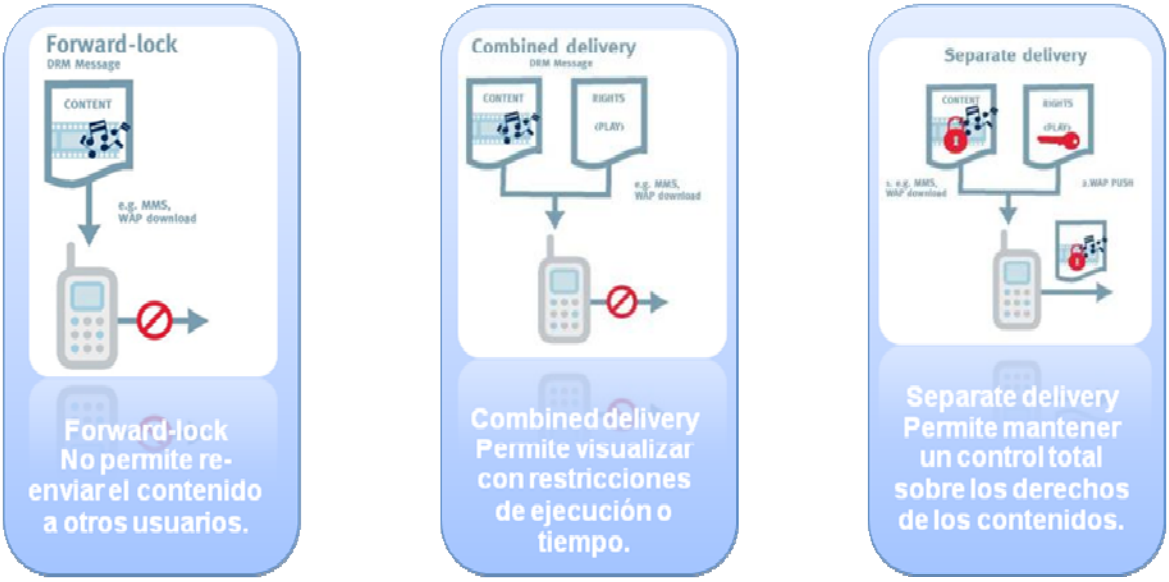


Figura VI. 4. Tecnologías de DRM

VI.2.3 Personalización de dispositivos

La personalización de los dispositivos es un término que ha estado en boga desde 2004 y es probable que permanezca mucho tiempo. La personalización del equipo de acuerdo a requisitos particulares es cuando la apariencia (llamada look & feel) del dispositivo está adaptada para satisfacer requerimientos del operador. Dentro del

contexto de la SIM, la personalización del equipo para cubrir a una audiencia específica es particularmente importante. Refiriéndose este tipo de personalización como personalización dirigida. Un ejemplo, es un equipo con una aplicación de compra dirigida a una audiencia de usuarios femeninos.

En otro extremo también puede existir una personalización directa del usuario y esto sucede cuando el usuario modifica la apariencia del equipo utilizando por ejemplo imágenes de fondo, sonidos de alarma, etc. En particular interés dentro del contexto de la SIM está la personalización dirigida, es decir, el acto iniciado por la operadora para modificar el contenido o las aplicaciones para acomodar mejor a cada usuario de forma individual. Un ejemplo es un portal WAP[6] que muestre los menús más frecuentemente visitados en la parte inicial del menú.

La personalización dirigida es de importancia para la operadora debido a varias razones. En primer lugar, hacen frente a otros operadores mediante la competencia de numerosas marcas. Hacen frente a la limitación de tener una sola marca, mediante la segmentación de los clientes, mediante la personalización de los servicios y del equipo para agrupar a los clientes dentro de un nicho. En segundo lugar, la pobre accesibilidad y descubrimiento del contenido son importantes razones para tomar en consideración servicios de datos basados en la personalización dirigida para mejorar los aspectos de una experiencia del usuario para mantener contenidos relevantes y aplicaciones de personalización en el perfil del usuario final.

Las tarjetas de alta densidad ofrecen varias ventajas para la personalización dirigida. Los operadores GSM han estado utilizando por algún tiempo la personalización basados en la SIM para adaptar servicios y equipos a geografías diferentes, segmentos de usuarios (por ejemplo usuarios corporativos y masivos) y distribución de canales (por ejemplo menú de descargas de contenido).

[6] WAP. Wireless Application Protocol o WAP (protocolo de aplicaciones inalámbricas) es un estándar abierto internacional para aplicaciones que utilizan las comunicaciones inalámbricas, p.ej. acceso a servicios de Internet desde un teléfono móvil.

Las tarjetas de alta densidad pueden ofrecer un importante empuje a las operaciones de personalización, permitiendo que los operadores empaqueten los servicios y ajustes dentro de la tarjeta SIM, ya que requiere un cambio mínimo en el proceso de negocio. El uso de las tarjetas de alta densidad en los panoramas anteriormente mencionados, generan un tiempo de lanzamiento al mercado muy corto, desde que ocurre la personalización en la SIM independientemente de la personalización del equipo, los cuales tienen un proceso muy complejo y tienen mucho desperdicio de tiempo.

Así mismo, las tarjetas de alta densidad ofrecen ventajas cuando viene una personalización dirigida. El perfil del usuario guardado en la tarjeta puede abarcar una gama de preferencias de uso tales como esquemas y colores del fondo, artistas favoritos para sonidos de alarma (tonos) y el contenido recientemente visto. Esta información puede ser reutilizada inmediatamente por aplicaciones dentro del equipo móvil sin requerir ninguna comunicación con la red. Así mismo, la información puede aplicarse dentro de distintos equipos, persistiendo a través del tiempo de vida de la suscripción, de modo que el usuario pueda llevar su perfil con él, incluso cuando cambie el equipo móvil (por ejemplo cuando los usuarios cambian su equipo por un modelo más actual de acuerdo a las tendencias).

Las ventajas de la personalización dirigida permiten nuevos escenarios de lealtad de usuarios, ya que el usuario al querer cambiar de operadora, toma el riesgo de perder su información, ya que el operador principal invalida a la SIM y por lo tanto los ajustes personales se podrían perder. Por lo tanto, en este caso, las capacidades de la SIM actúan como desaliento para disuadir al usuario final para salirse de la operadora.

En términos de tecnología, el uso de las tarjetas Flash de alta densidad (en comparación con las tarjetas ROM) puede utilizarse para el almacenamiento de aplicaciones personalizadas y para almacenar características solicitadas por el operador móvil. Los fabricantes de tarjetas de alta densidad utilizan la ROM para almacenar las funciones del sistema operativo y de la seguridad de la tarjeta (que raramente cambia), y el almacenamiento Flash se utiliza para las variaciones

dependientes del operador tales como aplicaciones, archivos y deltas del sistema operativo. El almacenamiento basado en Flash puede ser reescrito mucho más rápidamente y fácilmente, en comparación con el almacenamiento de la ROM que requiere de tiempo de consumo en el proceso de la mascara. Con las tarjetas SIM de alta densidad usando el almacenamiento Flash, los archivos de personalización del operador pueden ser provisionados en diferentes tiempos: en el Punto de Venta (vía una terminal lectora de tarjetas), vía OTA en el punto cuando la tarjeta SIM se inserta en el equipo, vía OTA durante el tiempo de vida de la SIM, o cuando se reciclan los stocks de las tarjetas SIM y reponerlas con tarjetas SIM con nuevo contenido. Como resultado, los fabricantes importantes de tarjetas SIM, están cambiando el almacenamiento ROM por el almacenamiento basado en Flash, inclusive para las tarjetas ordinarias con baja capacidad.

Conclusiones

Esta tesis de licenciatura proporciona un ambiente real de los procesos actuales de manufactura a nivel mundial en la industria de Semiconductores, dando forma a un documento de consulta partiendo de una investigación de práctica de campo en el área de manufactura de memorias.

En el actual trabajo, se presentó de manera general la metodología requerida en el proceso de elaboración y uso de una memoria NAND Flash, dentro del segmento de telefonía celular. Ofreciendo a los lectores un documento útil de consulta, en donde pueden encontrar datos con referencias exactas, ya que se han tratado de identificar escenarios que solamente en libros especializados se podrían encontrar.

En la primera parte de este trabajo de investigación se presentaron los procesos y sus características durante el diseño de una memoria, demostrando la importancia de los procesos de corrección óptica para la eliminación de errores de impresión sobre las obleas, los cuales pueden generar un mal funcionamiento en el dispositivo.

La posibilidad de tener procesos de mas o siguiendo un orden diferente es muy común, en esta investigación de campo se presento el método utilizado dentro de una entidad real, documentando las características generales de cada uno de los procesos.

Así mismo, dentro del contexto del trabajo, se pudo comprobar que desafortunadamente, la tecnología requerida para la manufactura de estos dispositivos (memorias NAND Flash), está muy lejos de implementarse en nuestro país, debido a la elaborada implementación tecnológica que se requiere y a la alta competitividad de empresas extranjeras que sostienen tecnología de punto.

Por otra parte, se observa que debido a la rapidez en la evolución de las tecnologías (en este caso, inalámbricas particularmente), es cada vez más importante que los

involucrados en tecnología, fortalezcan sus habilidades y conocimientos. Por lo que este proyecto presenta un escenario real de lo que sucede dentro de las comunicaciones y su sinergia con la electrónica y la computación.

Actualmente en México, las dos empresas más fuertes en telefonía celular, están basadas en la tecnología GSM, por lo que el uso de la tarjeta SIM, es imprescindible. Sin embargo, por los altos costos que trae el tener una tarjeta con alta capacidad de memoria, las operadoras telefónicas, aún no se arriesgan en incursionar en muchos servicios que utilicen estas tecnologías, debido a que el mercado aún no está preparado para utilizar nuevos servicios. Probablemente en el transcurso de un par de años, la influencia de mercados europeos y orientales ayude a impulsar estos servicios especiales.

Bibliografía

[1] Rubio Sola, José Antonio. *Diseño de circuitos y sistemas integrados*, Ediciones UPC, 2003.

[2] Saburo Nonogaki, Takumi Ueno, Toshio Ito, *Microlithography Fundamentals in Semiconductor Devices and Fabrication Technology*, CRC Press, 1998.

[3] Robert Doering, Yoshio Nishi, *Handbook of Semiconductor Manufacturing Technology*, CRC, Press, 2008.

[4] Mark Burns, Gordon W. Roberts, *An Introduction to Mixed-signal IC Test and Measurement*, Oxford University Press US, 2001.

[5] Raj Karamchedu, Varada Raj Karamchedu, *It's Not about the Technology: Developing the Craft of Thinking for a High Technology Corporation*, Springer, 2004.

[6] Kendall Ling-Chiao Su, *Introducción al estudio de circuitos*, Reverté, 1979.

[7] Zheng Cui, *Micro-nanofabrication: Technologies and Applications*, Springer, 2005

[8] Gary S. May, Costas J. Spanos, *Fundamentals of semiconductor manufacturing and process control*, Wiley-IEEE Press, 2006.

[9] Robert L Boylestad, Louis Nashelsky, *Electronic devices and circuit theory*, Prentice Hall, 2005.

[10] John Paul Uyemura, *Chip design for submicron VLSI: CMOS layout and simulation*, CL-Engineering, 2005.

[11] Peter Van Zant, *Microchip fabrication: a practical guide to semiconductor processing*, McGraw-Hill Professional, 2004.

[12] R. Jacob Baker, *CMOS circuit design, layout, and simulation*, Wiley-IEEE Press, 2007.

[13] Kevin F Brennan, *Introduction to semiconductor devices: for computing and telecommunications applications*, Cambridge University Press, 2005.

[14] Benjamin G. Eynon, Jr. Banqiu Wu, *Photomask Fabrication Technology*, McGraw-Hill Professional, 2005.

[15] James R. Sheats, Bruce W. Smith, *Microlithography: Science and Technology*, CRC Press, 1998.

[16] D.L. Schilling, C. Belove, Circuitos *electrónicos discretos e integrados*, McGraw-Hill, 1993.

[17] Ronald J. Tocci, *Sistemas Digitales: Principios y Aplicaciones*, Pearson Educación, 2003.

[18] Zhang, Jingyuan. *Location management in cellular networks. Handbook of wireless networks and mobile computing*. John Wiley & Sons, 2002.

[19] Mare, Renzo. *Tecnologías de Banda Angosta*. Universidad Nacional el Rosario. 2003.

[20] Dominguez Sánchez, Juan José. *Telefonía móvil digital GSM*. Revista: Anales de mecánica y electricidad. España. 2000.

[21] Hillebrand, Friedhelm. *GSM and UMTS. The Creation of Global Mobile Communication*. John Wiley & Sons. 2001.

[22] Pooters, Ivo. *An Approach to full User Data Integrity Protection in UMTS Access Networks*. University of Twente Paper. 2006.

[23] Dahle, Peter. *Architecture for generic service development on mobile handsets*. Master's Thesis in Computer Science. University of Tromso.2007.

[24] Barthe, Gilles & Dufay, Guillaume. *Formal Methods for Smartcard Security*. University of Ottawa. 2005

[25] Dohmen, Jon Robert. *UMTS Authentication and Key Agreement*. Agder University College. Norway. 2001.

[26] Giesecke & Devrient. *Cards for Telecommunications Whitepaper*. Alemania. 2008.

[27] Giesecke & Devrient. *UniverSIM®, Securing the 3rd generation Whitepaper*. Alemania. 2004.

[28] Smart Trust. *SIM- The basis for Mobile Value Added Services. Whitepaper*. 2002.

[29] Web ProForum Tutorials. *Global System for Mobile Communications (GSM)*. The International Engineering Consortium. 2002.

[30] 3GPP TS 21.111. *Technical Specification Group Terminal; USIM and IC card requirements (Release 1999).* 3GPP. 1999

[31] 3GPP 23.048. *Security mechanisms for the (U)SIM application toolkit; Stage 2*. 2005.

[32] 3GPP TS 31.111. *Universal Subscriber Identity Module (USIM) Application Toolkit (USAT)*. 2007.

[33] 3GPP TS 31.103. *Characteristics of the IP Multimedia Services Identity Module (ISIM) application*. 2008.

[34] GSM 02.19. *Digital cellular telecommunications system (Phase 2+, Release 98): Subscriber Identity Module Application Programming Interface (SIM API); Service description*. ETSI, Francia, 1999.

[35] GSM 03.19. *Digital cellular telecommunications system (Phase 2+); Subscriber Indentity Module Application Programming Interface (SIM API); SIM API for Java Card* ^{TM}. ETSI. Francia. 1999.

[36] GSM 03.48. *Digital cellular telecommunications system (Phase2+); Security Mechanisms for the SIM Application Toolkit; Stage 2*.

[37] GSM 11.11. *Digital cellular telecommunications system (Phase2+); Specification of the Subscriber Identity Module-Mobile Equipment(SIM-ME) interface*.

[38] GSM 11.14. *Digital cellular telecommunications system (Phase2+); Specification of the SIM Application Toolkit for the Subscriber Identity Module - Mobile Equipment (SIM-ME) interface*.

[39] ISO/IEC 7816-4. *Identification cards - Integrated circuit(s) cards with contacts - Part 4: Interindustry commands for interchange*.

[40] ISO/IEC 7816-2. *Identification cards - Integrated circuit(s) cards with contacts - Part 2: Dimensions and locations of the contacts*.

[41] http://emes2.wikispaces.com/Semiconductores *(6-Oct-2008)*

[42] http://www.cardtechnology.com/article.html?id=200603015RXWB9XC *(11-Oct-2008)*

[43] http://www.electronicafacil.net/tutoriales/Principios-Basicos-Materiales-Semiconductores.php *(15-Oct-2008)*

[44] http://agamenon.uniandes.edu.co/~revista/articulos/redes_moviles/rm.html *(26-Oct-2008)*

[45] http://www.eetindia.co.in/ART_8800489141_1800006_NT_80daa54e.HTM *(3-Nov-2008)*

[46] http://www.oberthur.com *(4-Nov-2008)*

[47] http://www.smarttrust.com *(4-Nov-2008)*

[48] http://www.gi-de.com *(4-Nov-2008)*

[49] http://www.gemalto.com *(4-Nov-2008)*

[50] http://www.sagem-orga.com *(5-Nov-2008)*

[51] http://java.sun.com/javacard/ *(14-Nov-2008)*

[52] http://www.4centity.com/ *(27-Nov-2008)*

[53] http://www.smartcard.co.uk/NOLARCH/2007/February/130207.html *(5-Dic-2008)*

[54] http://www.smartcardalliance.org/articles/2007/02/12/giesecke-devrient-supplying-teliasonera-with-1-gigabyte-sim-cards *(6-Dic-2008)*

[55] http://www.gsmworld.com *(6-Dic-2008)*

[56] http://www.teleco.com.br/es/tutoriais/es_tutorialsim/pagina_1.asp *(7-Ene-2009)*

[57] http://www.dte.us.es/ing_inf/tec_comp/Tc/Temario/Tema4/t4.pdf *(12-Ene-2009)*

[58] http://www.simalliance.org *(12-Ene-2009)*

[59] http://www.ceqna.com/mobile-phones-plans/1653-cell-phones-4.html *(13-Ene-2009)*

[60] http://www.renesas.com/ *(15-Ene-2009)*

[61] http://www.atmel.com/ *(17-Ene-2009)*

[62] http://www.st.com/ *(21-Ene-2009)*